nature portfolio

https://doi.org/10.1038/s41567-021-01232-0

Supplementary information

Sample-efficient learning of interacting quantum systems

In the format provided by the authors and unedited

Supplementary Information for "Sample-efficient learning of interacting quantum systems"

Anurag Anshu* Srinivasan Arunachalam† Tomotaka Kuwahara‡ Mehdi Soleimanifar§

March 23, 2021

In this Supplementary Information, we provide the detailed proofs and analysis of the results presented in the main text. We first begin with giving an overview of the structure of this document, summarizing the subject of each section.

- **Preliminaries (Section 1)**: We gather the notations, definitions, facts, and well-known technical tools that are frequently used in the subsequent sections. This includes
  - Quantum belief propagation (Section 1.3)
  - Change in the spectrum after applying local operators (Section 1.4)
  - Local reduction of global operators (Section 1.5)
- **A lower bound on the sample complexity (Section 2)**: The proof of the $\sqrt{n}$ lower bound on the sample complexity of learning quantum Hamiltonians is given.
- **Efficient sample complexity assuming strong convexity (Section 3)**: We show how assuming the strong convexity of the log-partition function of quantum systems (whose proof is given in Section 4) implies an efficient sample complexity for the quantum Hamiltonian learning. This is done in two parts:
  - Sufficient statistics (Section 3.1)
  - Stochastic convex optimization (Section 3.2)
- **Proof of strong convexity of log-partition function (Section 4)**: Our main technical contribution is presented in depth here. The proof the strong convexity is divided into four steps. These steps are first briefly reviewed in Section 4.2, where an intuitive and broad perspective of them is provided. A diagram in Figure 1 depicts the theorems and lemmas used in each of these four steps and the connection between them.

*Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA. anuraganshu@berkeley.edu

†IBM T. J. Watson Research Center, Yorktown Heights, NY, USA Srinivasan.Arunachalam@ibm.com

‡Mathematical Science Team, RIKEN Center for Advanced Intelligence Project (AIP), Japan and Interdisciplinary Theoretical & Mathematical Sciences Program (iTHEMS) RIKEN, Japan. tomotaka.kuwahara@riken.jp

§Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA, USA. mehdis@mit.edu

– Outline of the proof of the quantum case (Section 4.2)
– Step 1: Relating the Hessian to a variance (Section 4.3)
– Step 2: From global to local variance (Section 4.5)
– Step 3: Bounding the effect of local unitaries (Section 4.6)
– Step 4: Reduction to infinite temperature variance (Section 4.7)
– Final step: Putting things together (Section 4.8)

1 Preliminaries

1.1 Some mathematical facts

Here we summarize some of the basic mathematical facts used in the proof. Let A, B be arbitrary operator. The *operator norm* of A which is its largest singular value is denoted by $\|A\|$. The minimum singular value of A (or equivalently the minimum eigenvalue when A is Hermitian) is denoted by $\sigma_{\min}(A)$. We also often use the *Frobenius norm* $\|A\|_F := \sqrt{\mathrm{tr}[A^\dagger A]}$ and more generally the Hilbert-Schmidt inner product between A, B defined by $\mathrm{tr}[A^\dagger B]$. Additionally using Hölder's inequality we have,

$$\|AB\|_F = \sqrt{\mathrm{tr}(B^\dagger A A^\dagger B)} \leq \sqrt{\|B\|^2 \mathrm{tr}(AA^\dagger)} = \|B\| \cdot \|A\|_F. \tag{1}$$

We define the von Neumann entropy of a quantum state ρ by $S(\rho) = -\mathrm{tr}[\rho \log \rho]$ and the relative entropy between two states ρ_1 and ρ_2 by $S(\rho_1 \| \rho_2) = -\mathrm{tr}[\rho_1 \log \rho_2] - S(\rho_1)$.

The gradient of a real function $f : \mathbb{R}^m \mapsto \mathbb{R}$ is denoted by $\nabla f(x)$ and its *Hessian* (second derivative) matrix by $\nabla^2 f(x)$. The entries of the Hessian matrix are given by $\frac{\partial^2}{\partial x_i \partial x_j} f(x)$. The minimum eigenvalue of the Hessian of $f(x)$ is shown by $\sigma_{\min}(\nabla^2 f(x))$.

We write $A \succeq 0$ to represent a *positive semi-definite* (PSD) operator A, one such example of a PSD operator is the Hessian matrix $\nabla^2 f(x)$ of a convex function $f(x)$.

We will use the standard notations $\mathcal{O}(\cdot)$, $\Theta(\cdot)$, and $\Omega(\cdot)$ to specify the scaling of various parameters. In particular, $\Theta(1)$ will denote a constant.

For convenience, we will also gather a collection of infinite sums over exponentials. For $t > 0$, let

$$\Gamma(t) := \int_0^\infty x^{t-1} e^{-x} dx = \frac{1}{t} \int_0^\infty e^{-x} d\left(x^t\right) = \frac{1}{t} \int_0^\infty e^{-y^{\frac{1}{t}}} dy$$

be the *gamma function*. It holds that $\Gamma(t) \leq t^t$. This can be used to simplify several summations that we encounter later. Finally, we collect a few useful summations that we use in our proofs in the following fact.

Fact 1. *Let $a, c > 0$, $1 \geq p > 0$ be reals and b be a positive integer. Then*

1) $\sum_{j=0}^\infty e^{-cj} \leq \frac{e^c}{c}$.

2) $\sum_{j=0}^\infty j^b e^{-cj^p} \leq \frac{2}{p} \cdot \left(\frac{b+1}{cp}\right)^{\frac{b+1}{p}}$.

3) $\sum_{j=0}^{\infty} e^{-c(a+j)^p} \le e^{-\frac{c}{2}a^p}\left(1+\frac{1}{p}\left(\frac{2}{cp}\right)^{\frac{1}{p}}\right)$.

Proof. The first summation follows from

$$\sum_{j=0}^{\infty} e^{-cj} = \frac{1}{1-e^{-c}} = \frac{e^c}{e^c-1} \le \frac{e^c}{c}.$$

For the second sum, notice that the function $t^b e^{-ct^p}$ achieves the maximum at $t^* = \left(\frac{b}{cp}\right)^{\frac{1}{p}}$. Then

$$\begin{aligned}
\sum_{j=0}^{\infty} j^b e^{-cj^p} &\le t^* \left(t^*\right)^b e^{-c(t^*)^p} + \int_0^\infty t^b e^{-ct^p} dt \\
&= \left(\frac{b}{cp}\right)^{\frac{b+1}{p}} e^{-\frac{b}{p}} + \frac{1}{(b+1)c^{\frac{b+1}{p}}} \int_0^\infty e^{-y^{\frac{p}{b+1}}} dy \\
&= \left(\frac{b}{e^{\frac{b}{b+1}} cp}\right)^{\frac{b+1}{p}} + \frac{1}{pc^{\frac{b+1}{p}}} \Gamma\left(\frac{b+1}{p}\right) \\
&\le \left(\frac{b}{e^{\frac{b}{b+1}} cp}\right)^{\frac{b+1}{p}} + \frac{1}{pc^{\frac{b+1}{p}}} \left(\frac{b+1}{p}\right)^{\frac{b+1}{p}} \le \frac{2}{p} \cdot \left(\frac{b+1}{cp}\right)^{\frac{b+1}{p}}.
\end{aligned}$$

For the third sum, we will use the identity

$$(a+j)^p \ge \frac{1}{2}\left(a^p + j^p\right).$$

Now, consider the following chain of inequalities and change of variables:

$$\begin{aligned}
\sum_{j=0}^{\infty} e^{-c(a+j)^p} &\le e^{-\frac{c}{2}a^p} \sum_{\ell=0}^{\infty} e^{-\frac{c}{2}\ell^p} \\
&\le e^{-\frac{c}{2}a^p} \left(1 + \int_0^\infty e^{-\frac{c}{2}t^p} dt\right) \\
&= e^{-\frac{c}{2}a^p} \left(1 + \frac{2^{\frac{1}{p}}}{c^{\frac{1}{p}}} \int_0^\infty e^{-y^p} dy\right) \\
&= e^{-\frac{c}{2}a^p} \left(1 + \frac{2^{\frac{1}{p}}}{pc^{\frac{1}{p}}} \Gamma\left(\frac{1}{p}\right)\right) \le e^{-\frac{c}{2}a^p} \left(1 + \frac{1}{pc^{\frac{1}{p}}} \left(\frac{2}{p}\right)^{\frac{1}{p}}\right).
\end{aligned}$$

This completes the proof. $\square$

1.2 Local Hamiltonians and quantum Gibbs states

Local Hamiltonians. In this work, we focus on Hamiltonians that are *geometrically local*. To describe this notion more precisely, we consider a D-dimensional lattice $\Lambda \subset \mathbb{Z}^D$ that contains n sites with a d-dimensional qudit (spin) on each site. Now the interactions terms in the Hamiltonian act on a constant number of qudits that are in the neighborhood of each other (in the lattice). We denote the dimension of the Hilbert space associated to the lattice Λ by $\mathcal{D}_\Lambda$. The Hamiltonian of this system is

$$H = \sum_{X \subset \Lambda} H_X.$$

Each term H_X acts only on the sites in X and X is restricted to be a connected set with respect to Λ. We call the Hamiltonian κ-local when the support of all the local terms H_X is $|X| \leq \kappa$. We also define the Hamiltonian restricted to a region $A \subseteq \Lambda$ by $H_A = \sum_{X \subseteq A} H_X$.

Let $B(r, i) := \{j \in \Lambda | \mathrm{dist}(i, j) \leq r\}$ denote a ball (under the Manhattan distance on the lattice) of size r centered at site i. For a given connected set $X \in \Lambda$, let $\mathrm{diam}(X) := \max\{\mathrm{dist}(i, j) : i, j \in X\}$ denote the *diameter* of this set, $X^c := \Lambda \setminus X$ denote the complement of this set and ∂X denote its boundary. Given two sets $X, Y \in \Lambda$, we define $\mathrm{dist}(X, Y) := \min (\mathrm{dist}(i, j) : i \in X, j \in Y)$.

In order to describe our Hamiltonians, we consider an orthogonal Hermitian basis for the space of operators acting on each qudit. For instance, for qubits, this basis consists of the Pauli operators. By decomposing each local term H_X in terms of the tensor product of such basis operators, we find the following canonical form for the Hamiltonian H:

Definition 2 (Canonical representation for κ-local Hamiltonians). *A κ-local Hamiltonian H on a lattice Λ can be written as a sum of m Hermitian operators E_ℓ each having a connected support with respect to Λ and acting non-trivially on at most κ qudits. That is,*

$$H = \sum_{\ell=1}^{m} \mu_\ell E_\ell. \tag{2}$$

where $\mu_\ell \in \mathbb{R}$ and we assume $\|E_\ell\| \leq 1$, $\mathrm{tr}[E_\ell^2] = \mathcal{D}_\Lambda$, $E_\ell^\dagger = E_\ell$ for $\ell \in [m]$, and

$$\mathrm{tr}[E_k E_\ell] = 0 \quad \textit{for } k \neq \ell. \tag{3}$$

Since we assume H is geometrically local (i.e. it is defined over a lattice), it holds that m the number of local terms in (2) satisfies $m = \mathcal{O}(|\Lambda|) = \mathcal{O}(n)$.

As discussed earlier, we extensively use the notion of quasi-local operators, which we now formally define.

Definition 3 (Quasi-local operators). *An operator A is said to be* (τ, a_1, a_2, ζ)-quasi-local *if it can be written as*

$$\begin{aligned} A &= \sum_{\ell=1}^{n} g_\ell \bar{A}_\ell \quad \text{with} \quad g_\ell \leq a_1 \cdot \exp(-a_2 \ell^\tau), \\ \bar{A}_\ell &= \sum_{|Z|=\ell} a_Z, \quad \max_{i \in \Lambda} \left(\sum_{Z : Z \ni i} \|a_Z\| \right) \leq \zeta, \end{aligned} \tag{4}$$

where the sets $Z \subset \Lambda$ are restricted to be balls.[1]

[1]The assumption that Z is a ball suffices for us. Our results on quasi-local operators also generalize to the case where Z is an arbitrary *regular* shape, for example when the radii of the balls inscribing and inscribed by Z are of constant proportion to each other.

Although local operators are morally a special case of quasi-local operators (when $\tau = \infty$), we will reserve the above notation for operators with $\tau \leq 1$. A useful tool for analyzing quasi-locality is the Lieb-Robinson bound, which shows a light-cone like behavior of the time evolution operator.

Fact 4 (Lieb-Robinson bound [LR72], [NS09]). *Let A, B be operators supported on regions X, Y of the D dimensional lattice Λ respectively. Also, let H be a geometrically local Hamiltonian. There exist $\mathcal{O}(1)$ constants v_{LR}, f, c that only depend on the details of the Hamiltonian such that*

$$\|[e^{iHt}Ae^{-iHt}, B]\| \leq f\|A\|\|B\| \cdot \min\left(|\partial X|, |\partial Y|\right) \cdot \min\left(e^{c(v_{\mathrm{LR}}|t| - \mathrm{dist}(X,Y))}, 1\right).$$

Gibbs states. At an *inverse-temperature* β, a quantum many-body system with the Hamiltonian $H(\mu)$ is in the Gibbs (thermal) state if it is given by

$$\rho_\beta(\mu) = \frac{e^{-\beta H(\mu)}}{\mathrm{tr}[e^{-\beta H(\mu)}]}. \tag{5}$$

The partition function of this system is defined by $Z_\beta(\mu) = \mathrm{tr}[e^{-\beta H(\mu)}]$.

Remark 5. *In our notation, we sometimes drop the dependency of the partition function or the Gibbs state on μ. We also often simply use the term local Hamiltonian H or quasi-local operator A when referring to Definition 2 and Definition 3.*

1.3 Quantum belief propagation

Earlier we saw that we could express the Gibbs state of a Hamiltonian H by Eq. (5). Suppose we alter this Hamiltonian by adding a term V such that

$$H(s) = H + sV, \quad s \in [0, 1]. \tag{6}$$

How does the Gibbs state associated with this Hamiltonian change? If the new term V commutes with the Hamiltonian H, i.e., $[H, V] = 0$, then the derivative of the Gibbs state of $H(s)$ is given by

$$\frac{d}{ds}e^{-\beta H(s)} = -\beta e^{-\beta H(s)}V = -\frac{\beta}{2}\left\{e^{-\beta H(s)}, V\right\}, \tag{7}$$

where $\{e^{-\beta H(s)}, V\} = e^{-\beta H(s)}V + Ve^{-\beta H(s)}$ denotes the anti-commutator. In the non-commuting case though, finding this derivative is more complicated. The *quantum belief propagation* is a framework developed in [Has07, Kim17, KB19] for finding such derivatives in a way that reflects the locality of the system.

Definition 6 (Quantum belief propagation operator). *For every $s \in [0, 1]$, $\beta \in \mathbb{R}$, define $H(s) = H + sV$ where $V = \sum_{j,k} V_{j,k}|j\rangle\langle k|$ is a Hermitian operator. Also let $f_\beta(t)$ be a function whose Fourier transform is*

$$\tilde{f}_\beta(\omega) = \frac{\tanh(\beta\omega/2)}{\beta\omega/2}, \tag{8}$$

i.e., $f_\beta(t) = \frac{1}{2\pi}\int d\omega \tilde{f}_\beta(\omega)e^{i\omega t}$. The quantum belief propagation operator $\Phi_{H(s)}(V)$ is defined by

$$\Phi_{H(s)}(V) = \int_{-\infty}^{\infty} dt f_\beta(t)\, e^{-iH(s)t}\, V\, e^{iH(s)t}.$$

Equivalently, in the energy basis of $H(s) = \sum_j \mathcal{E}_j(s) \, |j\rangle\langle j|$, we can write

$$\Phi_{H(s)}(V) = \sum_{j,k} |j\rangle\langle k| \, V_{j,k} \, \tilde{f}_\beta(\mathcal{E}_j(s) - \mathcal{E}_k(s)). \tag{9}$$

Proposition 7 (cf. [Has07]). *In the same setup as Definition 6, it holds that*

$$\frac{d}{ds} e^{-\beta H(s)} = -\frac{\beta}{2} \left\{ e^{-\beta H(s)}, \Phi_{H(s)}(V) \right\}. \tag{10}$$

1.4 Change in the spectrum after applying local operators

For a Hamiltonian H, let $P^H_{\le x}$ and $P^H_{\ge y}$ be projection operators onto the eigenspaces of H whose energies are in $\le x$ and $\ge y$, respectively (we use similar notation $P^A_{\le x}$, $P^A_{\ge y}$ for the quasi-local operator A). Consider a quantum state $|\psi\rangle$ in the low-energy part of the spectrum such that $P^H_{\le x}|\psi\rangle = |\psi\rangle$. Suppose this states $|\psi\rangle$ is perturbed by applying a local operator O_X on a subset $X \subset \Lambda$ of its qudits. Intuitively, we expect that the operator O_X only affects the energy of $|\psi\rangle$ up to $\mathcal{O}(|X|)$, i.e., $\|P^H_{\ge y} O_X |\psi\rangle\| \approx 0$ for $y \gg x + |X|$. A simple example is when $|\psi\rangle$ is the eigenstate of a classical spin system. By applying a local operation that flips the spins in a small region X, the energy changes at most by $\mathcal{O}(|X|)$. The following lemma rigorously formulates the same classical intuition for quantum Hamiltonians

Lemma 8 (Theorem 2.1 of [AKL16]). *Let H be an arbitrary κ-local operator such that*

$$H = \sum_{|Z| \le \kappa} h_Z, \quad \max_{i \in \Lambda} \sum_{Z : Z \ni i} \|h_Z\| \le g. \tag{11}$$

Then, for an arbitrary operator O_X which is supported on $X \subseteq \Lambda$, the operator norm of $P^H_{\ge y} O_X P^H_{\le x}$ is upper-bounded by

$$\|P^H_{\ge y} O_X P^H_{\le x}\| \le \|O_X\| \cdot \exp\big(-\frac{1}{2g\kappa}(y - x - 2g|X|) \big). \tag{12}$$

In our analysis, we need an different version of this lemma for *quasi-local* operators instead of κ-local operators. The new lemma will play a central role in lower-bounding the variance of quasi-local operators. The proof follows by the analysis of a certain moment function (as opposed to the moment generating function in [AKL16]). Due to formal similarities between the proofs, we defer the proof of the next lemma to Appendix 5.3.

Lemma 9 (Variation of [AKL16] for quasi-local operators). *Let A be a $(\tau, a_1, a_2, 1)$-quasi-local operator, as given in Eq. (4), with $\tau \le 1$. For an arbitrary operator O_X supported on a subset $X \subseteq \Lambda$ with $|X| = k_0$ and $\|O_X\| = 1$, we have*

$$\|P^A_{\ge x+y} O_X P^A_{\le x}\| \le c_5 \cdot k_0 \exp\Big(-(\lambda_1 y / k_0)^{1/\tau_1} \Big), \tag{13}$$

where $\tau_1 := \frac{2}{\tau} - 1$, c_5 and λ_1 are constants depending on a_1 and a_2 as $c_5 \propto a_2^{2/\tau}$ and $\lambda_1 \propto a_2^{-2/\tau}$ respectively.

1.5 Local reduction of global operators

An important notion in our proofs will be a reduction of a global operator to a local one, which has influence on a site i. Fix a subset $Z \subseteq \Lambda$ and an operator O supported on Z. Define

$$O_{(i)} := O - \mathrm{tr}_i[O] \otimes \frac{\mathbb{1}_i}{d}, \tag{14}$$

where operator $\mathbb{1}_i$ is the identity operator on the ith site, d is the local dimension, tr_i is the partial trace operation with respect to the site i. Note that $O_{(i)}$ removes all the terms in O that do not act on the ith site. This can be explicitly seen by introducing a basis $\{E_Y^\alpha\}_{\alpha\in\mathbb{N},Y\subseteq Z}$ of Hermitian operators, where Y labels the support of E_Y^α and α labels several possible operators on the same support. We can assume that $\mathrm{tr}[(E_Y^\alpha)^2] = \mathcal{D}_\Lambda, \mathrm{tr}_i[E_Y^\alpha] = 0$ for every $i \in Y$, and the orthogonality condition $\mathrm{tr}[E_Y^\alpha E_{Y'}^{\alpha'}] = 0$ holds if $\alpha \neq \alpha'$ or $Y \neq Y'$. These conditions are satisfied by the appropriately normalized Pauli operators. Expand

$$O = \sum_{\alpha,Y} g_{\alpha,Y} E_Y^\alpha.$$

Then

$$O_{(i)} = \sum_{\alpha,Y} g_{\alpha,Y} E_Y^\alpha - \sum_{\alpha,Y} g_{\alpha,Y} \mathrm{tr}_i[E_Y^\alpha] \otimes \frac{\mathbb{1}_i}{d} = \sum_{\alpha,Y:Y\ni i} g_{\alpha,Y} E_Y^\alpha.$$

Thus, $O_{(i)}$ is an operator derived from O, by removing all E_Y^α which act as identity on i. The following claim shows that the Frobenius norm of a typical $O_{(i)}$ is not much small in comparison to the Frobenius norm of O.

Claim 10. *For every operator O and $O_{(i)}$ defined in Eq. (14), it holds that*

$$\max_{i\in Z} \|O_{(i)}\|_F^2 \geq \frac{1}{|Z|} \sum_{i\in Z} \|O_{(i)}\|_F^2 \geq \frac{1}{|Z|} \|O\|_F^2. \tag{15}$$

Proof. Using the identities $\mathrm{tr}[E_Y^\alpha E_{Y'}^{\alpha'}] = 0$ and $\mathrm{tr}[(E_Y^\alpha)^2] = \mathcal{D}_\Lambda$, we have

$$\|O\|_F^2 = \mathcal{D}_\Lambda \sum_{Y,\alpha} g_{\alpha,Y}^2 \leq \mathcal{D}_\Lambda \sum_{i\in Z} \sum_{\alpha,Y:Y\ni i} g_{\alpha,Y}^2 = \sum_{i\in Z} \|O_{(i)}\|_F^2,$$

where the inequality comes from the fact that O is supported on the subset Z (i.e., $Y \subseteq Z$). This completes the proof. □

2 A lower bound on the sample complexity

In this section, we prove a lower bound on the sample complexity of the Hamiltonian Learning Problem as claimed in the main text. For convenience of the reader, we restate the precise definition of the Hamiltonian Learning Problem before proceeding to the proof of the lower bound.

Consider a κ local Hamiltonian $H(\mu) = \sum_{\ell=1}^{m} \mu_\ell E_\ell$ (defined in Section 1.2) that acts on n qudits and consists of m geometrically-local terms that satisfy $\max_{\ell \in [m]} |\mu_\ell| \leq 1$. In the Hamiltonian Learning Problem problem, we are given N copies of the Gibbs state of this Hamiltonian

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

at a fixed inverse-temperature β. The goal of a learning algorithm is to produce an estimate $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_m)$ of the coefficients μ_k that satisfy the following: with probability at least $1 - \mathrm{err}$, the algorithm outputs $\hat{\mu}$ such that

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

where $\|\mu - \hat{\mu}\|_2 = \left(\sum_{\ell=1}^{m} |\mu_\ell - \hat{\mu}_\ell|^2\right)^{\frac{1}{2}}$ is the ℓ_2-norm difference between the coefficient vectors μ and $\hat{\mu}$.

Theorem 11. *The number of copies N of the Gibbs state needed to solve the Hamiltonian Learning Problem and outputs a $\hat{\mu}$ satisfying $\|\hat{\mu} - \mu\|_2 \leq \varepsilon$ with probability $1 - \mathrm{err}$ is lower bounded by*

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

Proof. In order to prove the lower bound, we consider learning the parameters $\mu \in \mathbb{R}^m$ of the following class of one-local Hamiltonians on m qubits:

$$H(\mu) = \sum_{i=1}^{m} \mu_i |1\rangle\langle 1|_i,$$

where $|1\rangle\langle 1|_i$ is projection onto a basis $|1\rangle$ on the ith qudit. Let $T_m : \{\mu \in \mathbb{R}_+^m : \sum_i \mu_i^2 \leq 100\varepsilon^2\}$ be an orthant of the hypersphere of radius θ in $\mathbb{R}_+^m$. We have the following claim.

Claim 12. *There exists a collection of 2^m points in T_m, such that the ℓ_2 distance between each pair is $\geq \varepsilon$.*

Proof. Pick 2^m points uniformly at random in T_m. By union bound, the probability that at least one pair is at a distance of at most ε is at most $(2^m)^2$ times the probability that a fixed pair of points is at a distance of at most ε. But the latter probability is upper bounded by the ratio between the volume of a hypersphere of radius ε and the volume of T_m, which is $\frac{\varepsilon^m}{(10\varepsilon)^m/2^m} = \frac{1}{5^m}$. Since $(2^m)^2 \frac{1}{5^m} < 1$, the claim concludes. □

Let these set of 2^m points be S. For some temperature $\beta > 0$ and unknown $\mu \in S$, suppose $\mathcal{A}$ is an algorithm that is given N copies of $\rho_\beta(\mu)$ and, with probability $1 - \mathrm{err}$, outputs μ' satisfying $\|\mu' - \mu\|_2 \leq \varepsilon$. We now use $\mathcal{A}$ to assign the estimated $\hat{\mu}$ to exactly one of the parameters μ. Once the learning algorithm obtains an output μ', we can find the closest point in S (in ℓ_2 distance) as our estimate of μ, breaking ties arbitrarily. With probability $1 - \mathrm{err}$, the closest $\mu \in S$ to μ' is the correct μ since by the construction of S, $\|\mu' - \mu\|_2 \leq \varepsilon$. Thus, the algorithm $\mathcal{A}$ can be used to solve the problem of estimating the parameters μ themselves (not only approximating it). We furthermore show that the number of samples required to estimate $\mu \in S$ is large using lower bounds in the quantum state discrimination. We will directly use the lower bound from [HKK08] (as given in

[HW12]). Before we plug in their formula, we need to bound on the spectral norm of $\rho_\beta(\mu)$ for an arbitrary $\mu \in S$ (denoted by $\|\rho_\beta(\mu)\|$). That is,

$$
\begin{aligned}
\max_{\mu\in S}\{2^m \|\rho_\beta(\mu)\|\} &= \max_{\mu\in S} 2^m \left(\bigotimes_{i=1}^m \left\| \frac{1}{1+e^{-\beta\mu_i}}|0\rangle\langle 0| + \frac{e^{-\beta\mu_i}}{1+e^{-\beta\mu_i}}|1\rangle\langle 1| \right\|\right) \\
&= \max_{\mu\in S}\left(\bigotimes_{i=1}^m \left|\frac{2}{1+e^{-\beta\mu_i}}\right|\right) \\
&= \max_{\mu\in S}\left(\bigotimes_{i=1}^m \left|\frac{2e^{\beta\mu_i}}{e^{\beta\mu_i}+1}\right|\right) \\
&\le \max_{\mu\in S}\left(\bigotimes_{i=1}^m \left|\frac{2e^{\beta\mu_i}}{2}\right|\right) = \max_{\mu\in S}\left(e^{\beta\sum_{i=1}^m \mu_i}\right) \le e^{\beta\sqrt{m}\sqrt{\sum_{i=1}^m \mu_i^2}} \le e^{\beta\sqrt{m}\cdot 10\varepsilon},
\end{aligned}
$$

since $\sum_i \mu_i^2 \le 100\varepsilon^2$ for all $i \in S$. Thus, the lower bound for state identification of $\{H(\mu) : \mu \in S\}$ in [HW12, Equation 2] (cf. [HKK08] for the original statement) implies that

$$
N \ge \frac{\log|S| + \log(1-\mathrm{err})}{\log\left(\max_{\mu\in S}\{2^m\|\rho(\mu)_\beta\|\}\right)} = \frac{m\log 2 + \log(1-\mathrm{err})}{10\sqrt{m}\beta\varepsilon} = \Omega\Big(\frac{\sqrt{m} + \log(1-\mathrm{err})}{\varepsilon\beta}\Big).
$$

This establishes the lower bound. □

3 Efficient sample complexity assuming strong convexity

In this section, we provide more details on how one can find rigorous upper bounds on the sample complexity of the Hamiltonian Learning Problem using the notions of sufficient statistics and the strong convexity from the fields of stochastic convex optimization and statistics. Before introducing these concepts in detail, let us briefly review similar problems and some of the previous approaches to solve the Hamiltonian Learning Problem and why they do not immediately result in a provably efficient bound on the required number of samples for this task.

In "shadow tomography" [Aar18a, AR19, BKL+19], the goal is the following: given a set of observables $E_1, \dots, E_m$, how many copies of an unknown state ρ are sufficient to estimate approximate $\mathrm{tr}(\rho E_i)$ for all $i \in [m]$. We remark that the Hamiltonian Learning Problem pursues a different task than the shadow tomography problem. Our goal is to estimate the Hamiltonian (i.e. the coefficients μ_ℓ) within some given error bound, which is a stronger requirement than just approximating the expectation values of $\rho = e^{-\beta H}$ on a set of observables.

In another line of work, [BAL19, EHF19, QR19] considered learning an unknown Hamiltonian using local measurements by phrasing this question as solving a system of linear equations. Consequently, their complexity depends inverse polynomially on the "spectral gap" of the matrix in the linear system. It is not known how to obtain a non-trivial bound on this gap. There have also been several other heuristic and numerical evidence for learning an unknown classical and quantum Hamiltonian [WGFC14b, WGFC14a, VMN+19, AAR+18b] which we do not discuss here.

3.1 Sufficient statistics

As discussed in the main text, local expectations of the Gibbs states can be used to uniquely specify these states. This provides us with "sufficient statistics" for learning the Hamiltonians from ρ_β.

This observation is formalized in the following result:

Proposition 13. *Consider the following two Gibbs states*

$$\rho_\beta(\mu) = \frac{e^{-\beta \sum_\ell \mu_\ell E_\ell}}{\mathrm{tr}[e^{-\beta \sum_\ell \mu_\ell E_\ell}]}, \quad \rho_\beta(\lambda) = \frac{e^{-\beta \sum_\ell \lambda_\ell E_\ell}}{\mathrm{tr}[e^{-\beta \sum_\ell \lambda_\ell E_\ell}]} \tag{16}$$

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

Proof. We consider the relative entropy between $\rho_\beta(\lambda)$ and the Gibbs state $\rho_\beta(\mu)$. We have

$$\begin{aligned}
S\left(\rho_\beta(\mu)\|\rho_\beta(\lambda)\right) &= \mathrm{tr}\left[\rho_\beta(\mu)\left(\log \rho_\beta(\mu) - \log \rho_\beta(\lambda)\right)\right] \\
&= -S(\rho_\beta(\mu)) + \beta \cdot \mathrm{tr}\left[\rho_\beta(\mu) \sum_\ell \lambda_\ell E_\ell\right] + \log Z(\lambda) && (17) \\
&\overset{(1)}{=} -S(\rho_\beta(\mu)) + \beta \sum_\ell \lambda_\ell \mathrm{tr}[\rho_\beta(\lambda)E_\ell] + \log Z(\lambda) && (18) \\
&= -S(\rho_\beta(\mu)) + S(\rho_\beta(\lambda)) \\
&\overset{(2)}{\geq} 0, && (19)
\end{aligned}$$

where (1) follows because $\mathrm{tr}[\rho_\beta(\mu)E_\ell] = \mathrm{tr}[\rho_\beta(\lambda)E_\ell]$ for all $\ell \in [m]$ and (2) used the positivity of relative entropy. Similarly, we can exchange the role of $\rho(\mu)$ and $\rho(\lambda)$ in (19) and get

$$S\left(\rho_\beta(\lambda)\|\rho_\beta(\mu)\right) = -S(\rho_\beta(\lambda)) + S(\rho_\beta(\mu)) \geq 0. \tag{20}$$

Combining these bounds imply $S(\rho_\beta(\mu)) = S(\rho_\beta(\lambda))$ and hence from Eq. (19), we get $S(\rho_\beta(\mu)\|\rho_\beta(\lambda)) = 0$. It is known that the relative entropy of two distribution is zero only when $\rho_\beta(\mu) = \rho_\beta(\lambda)$. Hence, we also have $\log \rho_\beta(\mu) = \log \rho_\beta(\lambda)$ or equivalently up to an additive term $\sum_{\ell=1}^m \mu_\ell E_\ell = \sum_{\ell=1}^m \lambda_\ell E_\ell$. Since the operators E_ℓ form an orthogonal basis (see Eq. (3)), we see that $\lambda_\ell = \mu_\ell$ for all $\ell \in [m]$. □

Remark 14. *When the local expectations of the two Gibbs states only approximately match, i.e.,*

$$|\mathrm{tr}[\rho_\beta(\mu)E_\ell] - \mathrm{tr}[\rho_\beta(\lambda)E_\ell]| \leq \delta$$

for $\ell \in [m]$*, then a similar argument to* (19) *shows that* $S(\rho_\beta(\mu)\|\rho_\beta(\lambda)) \leq \mathcal{O}(m\delta)$*. By applying Pinsker's inequality*[2]*, we get* $\|\rho_\beta(\mu) - \rho_\beta(\lambda)\|_1^2 \leq \mathcal{O}(m\delta) = \mathcal{O}(n\delta)$*, where we used the fact that for geometrically-local Hamiltonians* $m = \mathcal{O}(n)$ *(see below Definition 2).*

Proposition 13 tells us that if we have a Gibbs state that matches the local expectation values to the unknown Gibbs state than uniquely identifies the unknown Hamiltonian. However, a natural question is, if there an algorithm that finds the Gibbs state given its local expectations. A solution to this question is to use the *maximum entropy optimization* [Jay57, BKL+19]. This is also closely related to the problem of *maximum likelihood estimation* that is used often to analyze the sample complexity of statistical problems. Formally, this problem can be expressed as follows:

[2] Pinsker's inequality states that for two density matrices ρ, σ, we have $\|\rho - \sigma\|_1^2 \leq 2\ln 2 \cdot S(\rho\|\sigma)$.

$$
\begin{aligned}
\max_{\rho} \quad & S(\rho) \\
\text{s.t.} \quad & \mathrm{tr}[\rho E_\ell] = e_\ell, \quad \forall \ell \in [m] \\
& \rho > 0, \quad \mathrm{tr}[\rho] = 1.
\end{aligned} \tag{21}
$$

where $S(\rho) = -\mathrm{tr}[\rho \log \rho]$ is the *von Neumann entropy* of ρ.

Below, we will be mainly concerned with the convex dual of this optimization which directly produces the interaction coefficients μ. Understanding the effect of statistical errors on the dual optimization program (and in more generality, stochastic convex optimization) will be the main the subject of the next few sections.

3.2 Stochastic convex optimization

Suppose we want to solve the optimization

$$\max_{x \in \mathbb{R}^m} f(x)$$

for a function $f : \mathbb{R}^m \to \mathbb{R}$ which is of the form $f(x) = \mathbb{E}_{y \sim \mathcal{D}}[g(x, y)]$. Here $g(x, y)$ is some convex function and the expectation $\mathbb{E}_{y \sim \mathcal{D}}$ is taken with respect to an unknown distribution $\mathcal{D}$. Algorithms for this maximization problem are based on obtaining i.i.d. *samples* y drawn from the distribution $\mathcal{D}$. In practice, we can only receive finite samples $y_1, y_2, \ldots, y_\ell$ from such a distribution. Hence, instead of the original optimization, we solve an *empirical* version

$$\max_{x \in \mathbb{R}^m} \frac{1}{\ell} \sum_{k=1}^{\ell} g(x, y_k).$$

The natural question therefore is: How many samples ℓ do we need to guarantee the output of the empirical optimization is close to the original solution? One answer to this problem relies on a property of the objective function known as *strong convexity*.

Definition 15. *Consider a convex function $f : \mathbb{R}^m \mapsto \mathbb{R}$ with gradient $\nabla f(x)$ and Hessian $\nabla^2 f(x)$ at x. The function f is said to be α-strongly convex in its domain if it is differentiable and for all x, y, and*

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

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

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

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

Next, we discuss how the framework of convex optimization and in particular strong convexity, can be applied to the Hamiltonian Learning Problem. To this end, we define the following optimization problems based on the convex dual of the optimization (21)

Definition 16 (Optimization program for learning the Hamiltonian). *We denote the objective function in the* Hamiltonian Learning Problem *and its approximate version by $L(\lambda)$ and $\hat{L}(\lambda)$ respectively, i.e.,*

$$L(\lambda) = \log Z_\beta(\lambda) + \beta \cdot \sum_{\ell=1}^{m} \lambda_\ell e_\ell, \quad \hat{L}(\lambda) = \log Z_\beta(\lambda) + \beta \cdot \sum_{\ell=1}^{m} \lambda_\ell \hat{e}_\ell, \tag{22}$$

where the partition function is given by $Z_\beta(\lambda) = \operatorname{tr}\left(e^{-\beta \sum_{\ell=1}^m \lambda_\ell E_\ell}\right)$. The parameters of the Hamiltonian that we intend to learn are $\mu = \arg\min_{\lambda \in \mathbb{R}^m : \|\lambda\| \le 1} L(\lambda)$. As before, we also define the empirical version of this optimization by

$$\hat{\mu} = \underset{\lambda \in \mathbb{R}^m : \|\lambda\| \le 1}{\arg\min} \hat{L}(\lambda). \tag{23}$$

We prove later in Lemma 27 that $\log Z_\beta(\lambda)$ is a convex function in parameters λ and thus, the optimization in (23) is a convex program whose solution can be in principle algorithmically found. In this work, we do not constraint ourselves with the running time of solving (23). We instead obtain sample complexity bounds as formulated more formally in the next theorem.

Theorem 17 (Error in $\boldsymbol{\mu}$ from error in local expectations $\boldsymbol{e_\ell}$ assuming strong convexity). *Let $\delta, \alpha > 0$. Suppose the local expectations e_ℓ are determined up to error δ, i.e., $|e_\ell - \hat{e}_\ell| \le \delta$ for all $\ell \in [m]$. Additionally, assume the strong convexity condition $\sigma_{\min}(\nabla^2 \log Z(\lambda)) \ge \alpha$ holds for $\|\lambda\| \le 1$. Then the optimal solution to the program (23) satisfies*

$$\|\mu - \hat{\mu}\|_2 \le \frac{2\beta\sqrt{m}\delta}{\alpha}. \tag{24}$$

Proof. From the definition of $\hat{\mu}$ as the optimal solution of $\hat{L}$ in (23), we see that $\hat{L}(\hat{\mu}) \le \hat{L}(\mu)$. Thus, we get

$$\log Z_\beta(\hat{\mu}) + \beta \cdot \sum_{\ell=1}^{m} \hat{\mu}_\ell \hat{e}_\ell \le \log Z_\beta(\mu) + \beta \cdot \sum_{\ell=1}^{m} \mu_\ell \hat{e}_\ell.$$

or equivalently,

$$\log Z_\beta(\hat{\mu}) \le \log Z_\beta(\mu) + \beta \cdot \sum_{\ell=1}^{m} (\mu_\ell - \hat{\mu}_\ell)\hat{e}_\ell. \tag{25}$$

We show later in Lemma 26 that for every $\ell \in [m]$, we have $\frac{\partial}{\partial \mu_\ell} \log Z_\beta(\mu) = -\beta e_\ell$.[3] This along with the assumption $\sigma_{\min}(\nabla^2 \log Z(\mu)) \ge \alpha$ in the theorem statement, implies that for every μ' with $\|\mu'\| \le 1$

$$\log Z_\beta(\mu') \ge \log Z_\beta(\mu) - \beta \cdot \sum_{\ell=1}^{m} (\mu'_\ell - \mu_\ell)e_\ell + \frac{1}{2}\alpha\|\mu' - \mu\|_2^2. \tag{26}$$

Hence, by choosing $\mu' = \hat{\mu}$ and combining (26) and (25), we get

$$\log Z_\beta(\mu) - \beta \cdot \sum_{\ell=1}^{m} (\hat{\mu}_\ell - \mu_\ell)e_\ell + \frac{1}{2}\alpha\|\hat{\mu} - \mu\|_2^2 \le \log Z_\beta(\mu) + \beta \cdot \sum_{\ell=1}^{m} (\mu_\ell - \hat{\mu}_\ell)\hat{e}_\ell$$

[3] In particular, see Eq. (38), where we showed $\frac{\partial}{\partial \mu_\ell} \log Z_\beta(\mu) = -\beta \cdot \operatorname{tr}[E_\ell \rho_\beta(\mu)] = -\beta e_\ell$.

which further implies that

$$\frac{1}{2}\alpha\|\hat{\mu}-\mu\|_2^2 \leq \beta \cdot \sum_{\ell=1}^{m}(\hat{\mu}_\ell-\mu_\ell)(e_\ell-\hat{e}_\ell),$$
$$\leq \beta \cdot \|\hat{\mu}-\mu\|_2 \cdot \|\hat{e}-e\|_2.$$

Hence, we have

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

□

Remark 18. *We note that the bound (24) in Theorem 17 can be also derived when other methods are used instead of the maximum entropy optimization or its dual in Definition 16. Suppose we use an alternative approach, which could be an approximate version of the maximum entropy method, a heuristic quantum algorithm that variationally prepares the Gibbs state, or one that infers the interactions in a different way without relying on the thermal averages. As long as the interaction coefficients inferred via such method (denoted by $\tilde{\mu}$) satisfy $\|\tilde{\mu}\| \leq 1$ and their corresponding Gibbs state possesses local expectations $\tilde{e}_k$ within δ distance of the original values, a similar expression as in (24) bounds the error of that approach as well.*

More formally strong convexity leads to Equation 26, which says

$$\log Z_\beta(\tilde{\mu}) \geq \log Z_\beta(\mu) - \beta \cdot \sum_{\ell=1}^{m}(\tilde{\mu}_\ell-\mu_\ell)e_\ell + \frac{1}{2}\alpha\|\tilde{\mu}-\mu\|_2^2.$$

Switching μ and $\tilde{\mu}$ as well as e_ℓ and $\tilde{e}_\ell$, we have

$$\log Z_\beta(\mu) \geq \log Z_\beta(\tilde{\mu}) - \beta \cdot \sum_{\ell=1}^{m}(\mu_\ell-\tilde{\mu}_\ell)\tilde{e}_\ell + \frac{1}{2}\alpha\|\tilde{\mu}-\mu\|_2^2.$$

Adding these equations and cancelling some terms, we obtain

$$0 \geq \beta \cdot \sum_{\ell=1}^{m}(\mu_\ell-\tilde{\mu}_\ell)(e_\ell-\tilde{e}_\ell) + \alpha\|\tilde{\mu}-\mu\|_2^2.$$

This leads to the inequality (using Cauchy-Schwartz)

$$\alpha\|\tilde{\mu}-\mu\|_2^2 \leq \beta \cdot \|\tilde{\mu}-\mu\|_2 \cdot \|\tilde{e}-e\|_2,$$

which is equivalent to

$$\|\tilde{\mu}-\mu\|_2 \leq \frac{\beta}{\alpha}\|\tilde{e}-e\|_2.$$

Finally, we obtain the following upper bound on the sample complexity assuming the log-partition function is strongly convex.

Corollary 19 (Sample complexity from strong convexity). *Under the same conditions as in Theorem 17, the number of copies of the Gibbs state ρ_β that suffice to solve the Hamiltonian Learning Problem and outputs a $\hat{\mu}$ satisfying $\|\hat{\mu}-\mu\|_2 \leq \varepsilon$ with probability $1-\mathrm{err}$ is*

$$N = \mathcal{O}\left(\frac{\beta^2 2^{\mathcal{O}(\kappa)}}{\alpha^2\varepsilon^2} m \log\frac{m}{\mathrm{err}}\right).$$

Proof. First observe that, using Theorem 17, as long as the error in estimating the local expectations e_ℓ are

$$\delta \leq \frac{\alpha\varepsilon}{2\beta\sqrt{m}}, \tag{27}$$

we estimate the coefficients μ by $\hat{\mu}$ such that $\|\hat{\mu} - \mu\|_2 \leq \varepsilon$. The local expectations e_ℓ can be estimated in various ways. One method considered in [CW20, BMBO20] is to group the operators E_ℓ into sets of mutually commuting observables and simultaneously measure them at once. Alternatively, we can use the recent procedure in [**?**, Theorem 1] based on a variant of shadow tomography. In either case, the number of copies of the state needed to find all the local expectations with accuracy δ and success probability $1 - \mathrm{err}$ is

$$N = \mathcal{O}\left(\frac{2^{\mathcal{O}(\kappa)}}{\delta^2}\log\frac{m}{\mathrm{err}}\right),$$

where recall that κ is the locality of the Hamiltonian. Plugging in Eq. (27) gives us the final bound

$$N = \mathcal{O}\left(\frac{\beta^2 2^{\mathcal{O}(\kappa)}}{\alpha^2\varepsilon^2} m \log\frac{m}{\mathrm{err}}\right).$$

□

Remark 20 (Final upper bound on the sample complexity). *In the next section, we prove that the α parameter in Corollary 19 is $\alpha \leq e^{-\Theta(\beta^c)} \cdot \frac{\beta^{c'}}{m}$. This implies the claimed sample complexity for the Hamiltonian Learning Problem in the main text, i.e.*

$$N = \mathcal{O}\left(\frac{e^{\mathcal{O}(\beta^c)}}{\beta^{\tilde{c}}\varepsilon^2} \cdot m^3 \cdot \log\frac{m}{\mathrm{err}}\right). \tag{28}$$

Notice that although we assume in the main that err*, the probability of error, is a constant (e.g.,* 0.01*), the dependency on* err *in* (28) *only appears in the logarithm and can be suitably scaled.*

4 Proof of strong convexity of log-partition function

We now state our main technical contribution which proves the strong convexity of the logarithm of the partition function.

Theorem 21. *Let $H = \sum_{\ell=1}^{m} \mu_\ell E_\ell$ be a κ-local Hamiltonian over a finite dimensional lattice. For a given inverse-temperature β, there are constants $c, c' > 3$ depending on the geometric properties of the lattice such that the following strong convexity property (see Definition 15) for $\log Z_\beta(\mu)$ holds*

$$\sigma_{\min}(\nabla^2 \log Z_\beta(\mu)) \geq e^{-\Theta(\beta^c)} \cdot \frac{\beta^{c'}}{m}. \tag{29}$$

In other words[4]*, for every vector $v \in \mathbb{R}^m$ we have $v^\top \cdot \nabla^2 \log Z_\beta(\mu) \cdot v \geq \beta^{c'} e^{-\Theta(\beta^c)}/m \cdot \|v\|_2^2$.*

The proof of Theorem 21 is explained in multiple steps. In the next few sections we first review the proof of this statement for the classical Hamiltonians. Then, we give a brief overview of the main steps in the proof of the quantum case.

[4] In future sections, we often use this characterization in terms of vectors v to obtain proper bounds on the minimum eigenvalue $\sigma_{\min}$.

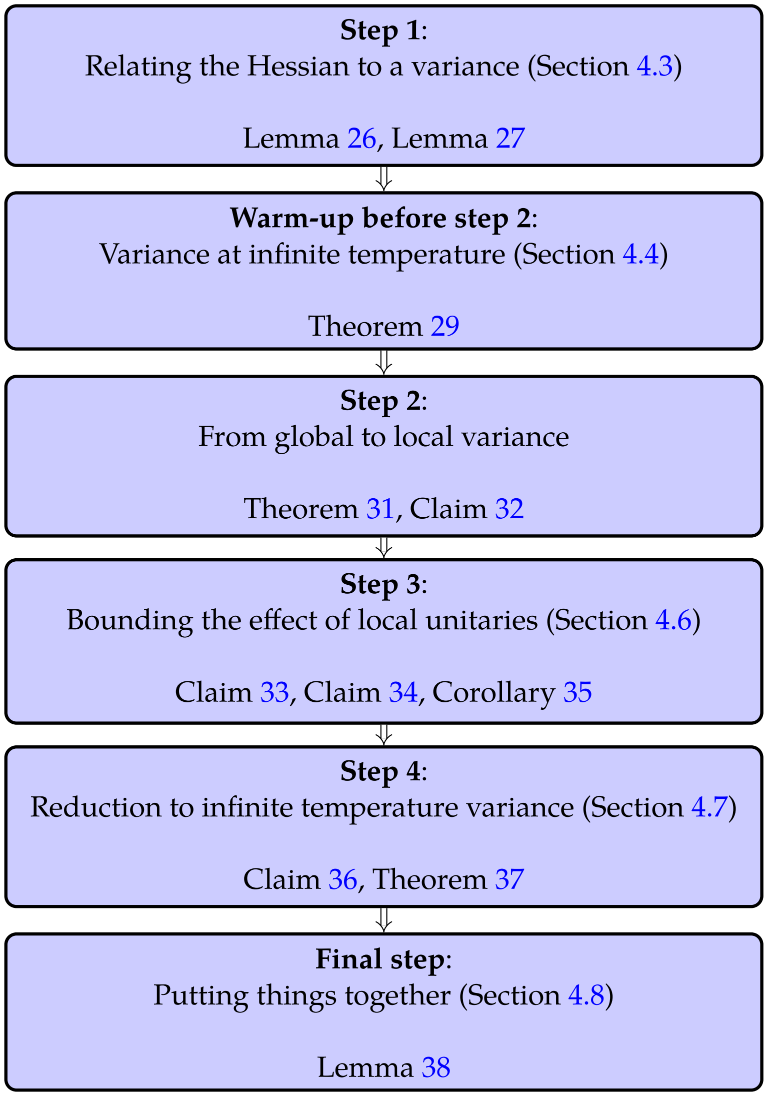


Figure 1: Flow of the argument in the proof of Theorem 21.

4.1 Review of the proof for classical systems

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

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

$$\frac{\partial^2}{\partial\mu_i\partial\mu_j}\Big[\log Z_\beta(\mu)\Big] = \mathrm{Cov}[E_i, E_j] \tag{30}$$

where Cov is the covariance function which is defined as $\mathrm{Cov}[E_i, E_j] = \langle E_iE_j\rangle - \langle E_i\rangle\langle E_j\rangle$ with the expectation taken with respect to the Gibbs distribution (i.e., $\langle E\rangle = \mathrm{tr}[E\cdot\rho_\beta(\mu)]$). To prove the strong convexity of the log-partition function at a constant β (defined in Definition 15), using (30) it suffices to show that for every vector v, we have

$$\sum_{i,j} v_iv_j\frac{\partial^2}{\partial\mu_i\partial\mu_j}\log Z_\beta(\mu) = \mathrm{Var}\left[\sum_{\ell=1}^m v_\ell E_\ell\right] \geq \Theta(1)\cdot\sum_{\ell=1}^m v_\ell^2. \tag{31}$$

Although the operator $\sum_\ell v_\ell E_\ell$ is a local Hamiltonian, note the mismatch between this operator and the original Hamiltonian in the Gibbs state $\sum_{\ell=1}^m \mu_\ell E_\ell$. Note that the inequality (31) is stronger than our main technical contribution in the quantum setting (i.e., for the case of quantum partition functions, we proved the analogue of inequality (31) when $\Theta(1)$ replaced by $\Theta(1/m)$). Before proving Eq. (31), we remark that an *upper bound* of $\mathrm{Var}[\sum_{\ell=1}^m v_\ell E_\ell] \leq \mathcal{O}(1)\|v\|_2^2$ is known in literature, under various conditions like the decay of correlations both in classical and quantum settings [Ara69, Gro79, PY95, Uel04, KGK+14, FU15]. This upper bound intuitively makes sense because the variance of the thermal state of a Hamiltonian and other local observables are expected to be *extensive*, i.e., they scale with the number of particles (spins) or norm of the Hamiltonian, which is replaced by $\|v\|_2^2$ in our setup. However, in the classical Hamiltonian learning problem, we are interested in obtaining a *lower bound* on the variance. To this end, a crucial property of the (classical) Gibbs distributions that allows us to prove the inequality (31) is the conditional independence or the *Markov property* of classical systems.

Definition 22 (Markov property). *Suppose the interaction graph is decomposed into three disjoint regions A, B, and C such that region B "shields" A from C, i.e., the vertices in region A are not connected to those in C. Then, conditioned on the sites in region B, the distribution of sites in A is independent of those in C. This is often conveniently expressed in terms of the conditional mutual information by $I(A:C|B) = 0$.*

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

$$\begin{aligned}
\mathrm{Var}\left[\sum_{\ell=1}^{m} v_\ell E_\ell\right] &\overset{(1)}{\geq} \mathbb{E}_{s_B}\left[\mathrm{Var}\left[\sum_{\ell=1}^{m} v_\ell E_\ell \,\Big|\, s_B\right]\right] \\
&\overset{(2)}{=} \sum_{x\in A} \mathbb{E}_{s_B}\left[\mathrm{Var}\left[\sum_{\ell: E_\ell \text{ acts on } x} v_\ell E_\ell \,\Big|\, s_B\right]\right] \\
&\overset{(3)}{\geq} \Theta(1)\cdot \sum_{\ell=1}^{m} v_\ell^2,
\end{aligned} \tag{32}$$

where inequality (1) follows from the law of total variance, equality (2) can be justified as follows: by construction, the local terms E_ℓ either completely lie inside region B or intersect with only one of the sites in region A. In the former, the local conditional variance $\mathrm{Var}\,[E_\ell\,|s_B]$ vanishes, whereas in the latter, the interaction terms E_ℓ that act on different sites $x \in A$ become uncorrelated and the global variance decomposes into a sum of local variance. Finally, inequality (3) is derived by noticing that at any constant inverse-temperature β, the local variance is lower bounded by a constant that scales as $e^{-\Theta(\beta)}$. By carefully choosing the partitions A and B such that $|A| = \mathcal{O}(n)$, we can make sure that the variance in inequality (2) is a constant fraction of the $\sum_{\ell=1}^{m} v_\ell^2$ as in (32) (see [Mon15, VMLC16] for details).

4.2 Outline of the proof of the quantum case

If we try to directly quantize the proof strategy of the classical case in the previous section, we immediately face several issues. In what follows we describe the challenges in obtaining a quantum proof along with our techniques to overcome them which allow us to establish Theorem 21. The content of our proof of the quantum case is divided into four steps whose overview is given below.

4.2.1 Overview of step 1: Relating the Hessian to a variance

The first problem is that we cannot simply express the entries of the Hessian matrix $\nabla^2 \log Z_\beta(\mu)$ in terms of $\mathrm{Cov}[E_i, E_j]$ as in (30). This expression in (30) only holds for Hamiltonians with *commuting* terms, i.e., $[E_i, E_j] = 0$ for all $i, j \in [m]$. The Hessian for the non-commuting Hamiltonians takes a complicated form (see Lemma 26 for the full expression) that makes its analysis difficult. Our first contribution is to recover a similar result to (31) in the quantum case by showing that, for every v, we can still *lower bound* $v^\top \cdot \nabla^2 \log Z_\beta(\mu)\cdot v$ by the variance of a suitably defined *quasi-local* operator. We later define what we mean by "quasi-local" more formally (see Definition 3 in the body), but for now one can assume such an operator is, up to some small error, sum of local terms.

Lemma 23 (A lower bound on $v^\top \cdot \nabla^2 \log Z_\beta \cdot v$). *For any vector $v \in \mathbb{R}^m$, define a quasi-local operator $\widetilde{W} = \sum_{\ell=1}^{m} v_\ell \widetilde{E}_\ell$, where the operators $\widetilde{E}_\ell$ are defined by*

$$\widetilde{E}_\ell = \int_{-\infty}^{\infty} f_\beta(t)\, e^{-iHt}\, E_\ell\, e^{iHt} dt. \tag{33}$$

Here $f_\beta(t) = \frac{2}{\beta\pi}\log\frac{e^{\pi|t|/\beta}+1}{e^{\pi|t|/\beta}-1}$ is defined such that $f_\beta(t)$ scales as $\frac{1}{\beta}e^{-\pi|t|/\beta}$ for large t and $f_\beta(t) \propto \log(1/t)$

for $t \to +0$. Then

$$\sum_{i,j} v_i v_j \frac{\partial^2}{\partial \mu_i \partial \mu_j} \log Z_\beta(\mu) \geq \beta^2 \operatorname{Var}[\widetilde{W}]. \tag{34}$$

This implies the bound $\sigma_{\min}(\log Z_\beta(\mu)) \geq \beta^2 \operatorname{Var}[\widetilde{W}]$ on the the minimum eigenvalue of $\log Z_\beta(\mu)$.

4.2.2 Overview of step 2: From global to local variance

As a result of Lemma 23, we see that from here onwards, it suffices to lower bound the variance of the quasi-local operator $\widetilde{W} = \sum_{\ell=1}^m v_\ell \tilde{E}_\ell$. One may expect the same strategy based on the Markov property in (32) yields the desired lower bound. Unfortunately, it is known that a natural extension of this property to the quantum case, expressed in terms of the *quantum conditional mutual information* (qCMI), does not hold. In particular, example Hamiltonians are constructed in [LP08] such that for a tri-partition A, B, C as in Definition 22, their Gibbs states have non-zero qCMI, i.e., $I(A : C|B) > 0$. Nevertheless, it is *conjectured* that an *approximate* version of this property can be recovered i.e., $I(A : C|B) \leq e^{-\Omega(\mathrm{dist}(A,C))}$. In other words, the approximate property claims that qCMI is exponentially small in the *width* of the shielding region B. Thus far, this conjecture has been proved only at sufficiently high temperatures [KKBa20] and on 1D chains [KB19]. Even assuming this conjecture is true, we currently do not know how to recover the argument in (32). Given this issue we ask,

Can we obtain an unconditional lower bound on the variance of a quasi-local observable at any inverse-temperature β without assuming quantum conditional independence?

Our next contribution is to give an affirmative answer to this question. To achieve this, we modify the classical strategy as explained below.

One ingredient in the classical proof is to lower bound the global variance $\operatorname{Var}[\sum_\ell v_\ell E_\ell]$ by sum of local conditional variances $\operatorname{Var}[E_\ell | s_B]$ as in (32). We prove a similar but slightly weaker result in the quantum regime. To simplify our discussion, let us ignore the fact that $\widetilde{W} = \sum_\ell v_\ell \tilde{E}_\ell$ is a quasi-local operator and view it as (strictly) local. Consider a special case in which v is such that the operator $\widetilde{W}$ is supported on a small number of sites. For instance, it could be that $v_1 > 0$ while $v_2, \ldots, v_m = 0$. Then the variance $\operatorname{Var}[\widetilde{W}]$ can be easily related to the local variance $\operatorname{Var}[E_1]$ and since $E_1^2 = \mathbb{1}$, $|\mathrm{tr}[E_1 \rho_\beta]| < 1$, we get

$$\operatorname{Var}[\widetilde{W}] = v_1^2 \cdot \left(\mathrm{tr}[E_1^2 \rho_\beta] - \mathrm{tr}[E_1 \rho_\beta]^2\right) \geq \Theta(1) \cdot v_1^2$$

We show that even in the general case, where $v_1, \ldots, v_m$ are all non-zero, we can still relate $\operatorname{Var}[\widetilde{W}]$ to the variance of a local operator supported on a constant region. Compared to the classical case in (32), where the lower bound on $\operatorname{Var}[W]$ includes a sum of $\mathcal{O}(m)$ local terms, our reduction to a *single* local variance costs “an extra factor of m” in our strong convexity bound.

Our reduction to local variance is based on the following observation. By applying Haar-random local unitaries on a site i, we can remove all the terms of the operator $\widetilde{W}$ except those that act on an arbitrary qudit at the site (see also Section 1.5). We denote the remainder terms by $\widetilde{W}_{(i)}$ defined via

$$\widetilde{W}_{(i)} = \widetilde{W} - \mathbb{E}_{U_i \sim \mathrm{Haar}}[U_i^\dagger \widetilde{W} U_i].$$

By using the triangle inequality this relation implies

$$\mathrm{Var}[\widetilde{W}] \geq \frac{1}{2}\mathrm{tr}[\widetilde{W}^2_{(i)}\rho_\beta] - \mathbb{E}_{U_i}\left[\mathrm{tr}[\widetilde{W}^2 \cdot U_i\rho_\beta U_i^\dagger]\right]. \tag{35}$$

Hence, if we could carefully analyze the effect of the term $\mathbb{E}_{U_i}[\mathrm{tr}[\widetilde{W}^2 \cdot U_i\rho_\beta U_i^\dagger]]$, this will allow us to relate the global variance $\mathrm{Var}[\widetilde{W}]$ to the local variance $\mathrm{tr}[\widetilde{W}^2_{(i)}\rho_\beta]$. We discuss this next.

4.2.3 Overviwe of step 3: Bounding the effect of local unitaries

While applying the above reduction helps us to go to an easier local problem, we need to deal with the changes in the spectrum of the Gibbs state due to applying the random local unitaries U_i. Could it be that the unitaries U_i severely change the spectral relation between $\widetilde{W}$ and ρ_β? We show that this is not the case, relying on the facts: (1) local unitaries cannot mix up subspaces of $\widetilde{W}$ and H that are energetically far away and (2) the weight given by the Gibbs state ρ_β to nearby subspaces of H are very similar at small β. Thus, (1) allows us to focus on the subspaces that are close in energy and (2) shows that similar weights of these subspaces do not change the variance by much. In summary, we prove:

Proposition 24 (Invariance under local unitaries, informal). *Let U_X be a local unitary operator acting on region X that has a constant size. There exists a constant $c \leq 1$ such that*

$$\mathrm{tr}\left[\widetilde{W}^2 \cdot U_X\rho_\beta U_X^\dagger\right] \lesssim \left(\mathrm{Var}[\widetilde{W}]\right)^c. \tag{36}$$

When combined with (35), the inequality (36) implies the following loosely stated local lower bound on the global variance:

$$\mathrm{Var}[\widetilde{W}] \gtrsim \left(\mathrm{tr}\left[\widetilde{W}^2_{(i)}\rho_\beta\right]\right)^{\frac{1}{c}}.$$

4.2.4 Overview of step 4: Reduction to infinite temperature variance

With this reduction, it remains to find a constant lower bound on $\mathrm{tr}[\widetilde{W}^2_{(i)}\rho_\beta]$. This can be done, again, by applying a local unitary U. Roughly speaking, we use this unitary to perform a "change of basis" that relates the local variance at finite temperature to its infinite-temperature version. The spectrum of ρ_β majorizes the maximally mixed state η. Hence, by applying a local unitary, we can rearrange the eigenvalues of $\widetilde{W}^2_{(i)}$ in the same order as that of ρ_β such that when applied to both ρ_β and η, we have $\mathrm{tr}[\widetilde{W}^2_{(i)}U\rho_\beta U^\dagger] \geq \mathrm{tr}[\widetilde{W}^2_{(i)}\eta]$. Formally, we show that

Proposition 25 (Lower bound on the local variance, informal). *There exists a unitary U supported on $\mathcal{O}(1)$ sites such that*

$$\mathrm{tr}\left[\widetilde{W}^2_{(i)}U\rho_\beta U^\dagger\right] \geq \mathrm{tr}\left[\widetilde{W}^2_{(i)}\eta\right],$$

where η is the maximally mixed state or the infinite temperature Gibbs state.

In summary, starting from (35) and following Proposition 24 and Proposition 25, the lower bound on the global variance takes the following local form:

$$\mathrm{tr}\left[\widetilde{W}^2\rho_\beta\right] \geq \left(\mathrm{tr}\left[\widetilde{W}^2_{(i)}\eta\right]\right)^{\Theta(1)}.$$

Lower bounding the quantity $\mathrm{tr}\left[\widetilde{W}_{(i)}^2\eta\right]$ by a constant is now an easier task, which we explain in more details later in Lemma 38 and Theorem 29.

4.3 Step 1: Relating the Hessian to a variance

We begin our detailed proofs with finding an expression for the Hessian of $\log Z_\beta(\lambda)$ in terms of the variance of a quasi-local operator.

Lemma 26. *For every vector $v \in \mathbb{R}^m$, define the local operator $W_v = \sum_{i=1}^m v_i E_i$ (for notational convenience, later on we stop subscripting W by v). The Hessian $\nabla^2 \log Z_\beta(\lambda)$ satisfies*

$$v^\top \cdot \left(\nabla^2 \log Z_\beta(\lambda)\right) \cdot v = \frac{\beta^2}{2}\mathrm{tr}\Big[\big\{W_v, \Phi_{H(\lambda)}(W_v)\big\}\rho_\beta(\lambda)\Big] - \beta^2\big(\mathrm{tr}\left[W_v\rho_\beta(\lambda)\right]\big)^2, \tag{37}$$

Proof. Since the terms in the Hamiltonian are non-commuting, we use Proposition 7 to find the derivatives of $\log Z_\beta(\lambda)$. We get

$$\begin{aligned}
\frac{\partial}{\partial\lambda_j}\log Z_\beta(\lambda) &= \frac{1}{Z_\beta(\lambda)}\mathrm{tr}\left[-\frac{\beta}{2}\left\{e^{-\beta H(\lambda)}, \Phi_{H(\lambda)}(E_j)\right\}\right] \\
&= \frac{-\beta}{Z_\beta(\lambda)}\mathrm{tr}\left[e^{-\beta H(\lambda)}\int_{-\infty}^{\infty} dt f_\beta(t) e^{-iH(\lambda)t}\, E_j\, e^{iH(\lambda)t}\right] \\
&= -\beta\ \mathrm{tr}\left[E_j \frac{e^{-\beta H(\lambda)}}{Z_\beta(\lambda)}\right],
\end{aligned} \tag{38}$$

where the second equality used the definition of the quantum belief propagation operator $\Phi_{H(\lambda)}(E_j) = \int_{-\infty}^{\infty} dt f_\beta(t)\ e^{-iH(\lambda)t}\ E_j\ e^{iH(\lambda)t}$ with f_β as given in Definition 6. The third equality used the fact that $e^{iH(\lambda)t}$ commutes with $e^{-\beta H(\lambda)}$. Similarly, we have

$$\begin{aligned}
\frac{\partial^2}{\partial\lambda_k\partial\lambda_j}\log Z_\beta(\lambda) &= -\beta\ \mathrm{tr}\left[E_j \cdot \frac{\partial}{\partial\lambda_k}\left(\frac{e^{-\beta H(\lambda)}}{Z_\beta(\lambda)}\right)\right] \\
&= -\beta\ \mathrm{tr}\left[E_j \cdot \frac{1}{Z_\beta(\lambda)}\frac{\partial}{\partial\lambda_k}\left(e^{-\beta H(\lambda)}\right)\right] + \beta\mathrm{tr}\left[E_j \cdot \frac{e^{-\beta H(\lambda)}}{Z_\beta(\lambda)}\right]\cdot\frac{1}{Z_\beta(\lambda)}\frac{\partial}{\partial\lambda_k}Z_\beta(\lambda) \\
&= \frac{\beta^2}{2}\mathrm{tr}\left[E_j\cdot\{\rho_\beta(\lambda), \Phi_{H(\lambda)}(E_k)\}\right] - \beta^2\ \mathrm{tr}[E_k\rho_\beta(\lambda)]\ \mathrm{tr}[E_j\rho_\beta(\lambda)] \\
&= \frac{\beta^2}{2}\mathrm{tr}\left[\{E_j, \Phi_{H(\lambda)}(E_k)\}\cdot\rho_\beta(\lambda)\right] - \beta^2\ \mathrm{tr}[E_k\rho_\beta(\lambda)]\ \mathrm{tr}[E_j\rho_\beta(\lambda)].
\end{aligned}$$

One can see from this equation that $\nabla^2 \log Z_\beta(H)$ is a symmetric real matrix, [5] and hence its eigenvectors have real entries. Finally, we get

$$\begin{aligned}
v^\top \cdot \left(\nabla^2 \log Z_\beta(\lambda)\right)\cdot v &= \sum_{j,k} v_j v_k \frac{\partial^2}{\partial\lambda_k\partial\lambda_j}\log Z_\beta(\lambda) \\
&= \frac{\beta^2}{2}\mathrm{tr}\Big[\big\{W_v, \Phi_{H(\lambda)}(W_v)\big\}\rho_\beta(\lambda)\Big] - \big(\beta\mathrm{tr}\left[W_v\rho_\beta(\lambda)\right]\big)^2.
\end{aligned} \tag{39}$$

□

[5] The terms $\mathrm{tr}[E_k\rho_\beta(\lambda)]$ and $\mathrm{tr}[E_j\rho_\beta(\lambda)]$ are real, being expectations of Hermitian matrices. Moreover, $\{E_j, \Phi_{H(\lambda)}(E_k)\}$ is a Hermitian operator, being an anti-commutator of two Hermitian operators. Hence $\mathrm{tr}\left[\{E_j, \Phi_{H(\lambda)}(E_k)\}\,\rho_\beta(\lambda)\right]$ is real too.

The statement of Lemma 26 does not make it clear that the Hessian is a variance of a suitable operator, or even is positive. The following lemma shows how to lower bound the Hessian by a variance of a quasi-local operator. The intuition for the proof arises by writing the Hessian in a manner that makes its positivity clear. This in particular, shows that $\log Z_\beta(\mu)$ is a convex function in parameters μ — we later improve this to being strongly convex.

Lemma 27 (A lower bound on $v^\top \cdot \nabla^2 \log Z_\beta \cdot v$). *For every $v \in \mathbb{R}^m$ and local operator $W_v = \sum_i v_i E_i$, define another quasi-local operator $\widetilde{W}_v$ such that*

$$\widetilde{W}_v = \int_{-\infty}^{\infty} f_\beta(t)\, e^{-iHt}\, W_v\, e^{iHt} dt, \tag{40}$$

where

$$f_\beta(t) = \frac{2}{\beta\pi} \log \frac{e^{\pi|t|/\beta}+1}{e^{\pi|t|/\beta}-1} \tag{41}$$

is defined such that $f_\beta(t)$ scales as $\frac{4}{\beta\pi} e^{-\pi|t|/\beta}$ for large t. It holds that

$$\begin{aligned}&\frac{1}{2}\mathrm{tr}\Big[\big\{W_v, \Phi_{H(\lambda)}(W_v)\big\}\rho_\beta(\lambda)\Big] - \big(\mathrm{tr}\,[W_v\rho_\beta(\lambda)]\,\big)^2 \\ &\qquad \geq \mathrm{tr}\left[(\widetilde{W}_v)^2\rho_\beta(\lambda)\right] - \left(\mathrm{tr}\left[\widetilde{W}_v\rho_\beta(\lambda)\right]\right)^2\end{aligned} \tag{42}$$

This via Lemma 26 implies that $v^\top \cdot \big(\nabla^2 \log Z_\beta(\lambda)\big) \cdot v \geq \beta^2\, \mathrm{Var}[\widetilde{W}_v]$

Remark 28. *For the rest of the paper, we are going to fix an arbitrary $v \in \mathbb{R}^m$, in order to avoid subscripting $W, \widetilde{W}$ by v.*

Proof of Lemma 27. Let us start by proving a simpler version of Eq. (42), where we only show

$$\frac{1}{2}\mathrm{tr}\Big[\big\{W, \Phi_{H(\lambda)}(W)\big\}\rho_\beta(\lambda)\Big] - \big(\mathrm{tr}\,[W\rho_\beta(\lambda)]\,\big)^2 \geq 0. \tag{43}$$

Since v is an arbitrary vector, this shows that, as expected, $\nabla^2 \log Z_\beta(\lambda)$ is a positive semidefinite operator.

Consider the spectral decomposition of the Gibbs state $\rho_\beta(\lambda)$: $\rho_\beta(\lambda) = \sum_j r_j(\lambda)|j\rangle\langle j|$. Then observe that

$$\begin{aligned}&\frac{1}{2}\mathrm{tr}\Big[\big\{W, \Phi_{H(\lambda)}(W)\big\}\rho_\beta(\lambda)\Big] - \big(\mathrm{tr}\,[W\rho_\beta(\lambda)]\,\big)^2 &(44)\\ &= \frac{1}{2}\sum_j r_j(\lambda)\langle j|\{W, \Phi^H(W)\}|j\rangle - \left(\sum_j r_j(\lambda) W_{j,j}\right)^2 \\ &= \frac{1}{2}\sum_{j,k} r_j(\lambda)\left(W_{j,k}\langle k|\Phi^H(W)|j\rangle + \langle j|\Phi^H(W)|k\rangle W_{k,j}\right) - \left(\sum_j r_j(\lambda) W_{j,j}\right)^2 \\ &\overset{(1)}{=} \frac{1}{2}\sum_{j,k} r_j(\lambda)\left(W_{j,k}W_{k,j}\tilde{f}_\beta(\mathcal{E}_k - \mathcal{E}_j) + W_{j,k}W_{k,j}\tilde{f}_\beta(\mathcal{E}_j - \mathcal{E}_k)\right) - \left(\sum_j r_j(\lambda) W_{j,j}\right)^2 \\ &\overset{(2)}{=} \sum_{j,k} r_j(\lambda)|W_{j,k}|^2 \tilde{f}_\beta(|\mathcal{E}_j - \mathcal{E}_k|) - \left(\sum_j r_j(\lambda) W_{j,j}\right)^2. &(45)\end{aligned}$$

In equality (1), we use Definition 6 and in equality (2) we use the facts that W is Hermitian and resp$\tilde{f}_\beta(\omega) = \tilde{f}_\beta(-\omega)$. Since $\tilde{f}_\beta(0) = 1$ and $\tilde{f}_\beta(\omega) > 0$ for all ω, it is now evident from last equation that

$$\operatorname{tr}\left(\frac{1}{2}\{W, \Phi^H(W)\}\rho_\beta\right) - \operatorname{tr}\left(W\rho_\beta\right)^2 \geq \sum_j r_j(\lambda)|W_{j,j}|^2 - \left(\sum_j r_j(\lambda)W_{j,j}\right)^2 \geq 0.$$

We can improve this bound by using the operator $\widetilde{W}$ in (40). The function $f_\beta(t)$ in (40) is chosen carefully such that its Fourier transform satisfies $\tilde{f}_\beta(\omega) = \tilde{f}_\beta(|\omega|)$. Then, we have that

$$\widetilde{W} = \int_{-\infty}^{\infty} f_\beta(t)\, e^{-iHt}\, W\, e^{iHt} dt = \sum_{j,k} |j\rangle\langle k|\, W_{j,k}\, \tilde{f}_\beta(|\mathcal{E}_j - \mathcal{E}_k|). \tag{46}$$

Similar to (45) we get

$$\begin{aligned}
\operatorname{tr}(\widetilde{W}^2\rho_\beta(\lambda)) - [\operatorname{tr}(\widetilde{W}\rho_\beta(\lambda))]^2 &= \sum_{j,k} r_j(\lambda)|W_{j,k}|^2 \tilde{f}_\beta(|\mathcal{E}_j - \mathcal{E}_k|)^2 - \left(\sum_j r_j(\lambda)W_{j,j}\right)^2 \\
&\leq \sum_{j,k} r_j(\lambda)|W_{j,k}|^2 \tilde{f}_\beta(|\mathcal{E}_j - \mathcal{E}_k|) - \left(\sum_j r_j(\lambda)W_{j,j}\right)^2 \\
&= \operatorname{tr}\left(\frac{\{W, \Phi^H(W)\}}{2}\rho_\beta(\lambda)\right) - [\operatorname{tr}(W\rho_\beta(\lambda))]^2,
\end{aligned} \tag{47}$$

where the inequality is derived from $\tilde{f}_\beta(x)^2 \leq \tilde{f}_\beta(x)$ for arbitrary $-\infty < x < \infty$.

□

4.4 Warm-up before step 2: Variance at infinite temperature

In the previous section, we showed how to give a lower bound on the Hessian of the logarithm of the partition function. To be precise, for a vector $\lambda = (\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m$ with $\|\lambda\| \leq 1$, Hamiltonian $H(\lambda) = \sum_i \lambda_i E_i$ and $\rho_\beta(\lambda) = \frac{1}{Z_\beta(\lambda)} e^{-\beta H(\lambda)}$ (where $Z_\beta(\lambda) = \operatorname{tr}(e^{-\beta H(\lambda)})$), we showed in Lemma 27 how to carefully choose a local operator $\widetilde{W}$ such that for every v, we have

$$v^\top \cdot \left(\nabla^2 \log Z_\beta(\lambda)\right) \cdot v \geq \beta^2 \operatorname{Var}[\widetilde{W}]. \tag{48}$$

In the next few sections, we further prove that the variance of $\widetilde{W}$ with respect to $\rho_\beta(\lambda)$ can be bounded from below.

Before looking at how this can be achieved for the highly non-trivial case of finite temperature, we will look at a simpler case of *infinite temperature limit*. As we see later, certain elements of the proof strategy in this case extends to the general case too.

Consider the infinite temperature Gibbs state (i.e., the maximally mixed state) $\eta = \frac{\mathbb{1}_\Lambda}{\mathcal{D}_\Lambda}$. Assuming that the locality of W, namely W is κ-local with $\kappa = \mathcal{O}(1)$, the following theorem holds.

Theorem 29. *For $\widetilde{W}$ as defined in Lemma 27, we have*

$$\operatorname{tr}[(\widetilde{W})^2\eta] - \operatorname{tr}[\widetilde{W}\eta]^2 \geq \frac{\Theta(1)}{(\beta\log(m) + 1)^2} \sum_{i=1}^m v_i^2. \tag{49}$$

The intuition behind the theorem is as follows. In the above statement, if $\widetilde{W}$ is replaced by W, then the lower bound is immediate (see Eq. (50) below). Similarly, if H and W were commuting, then $\widetilde{W}$ would be the same as W and the statement would follow. In order to show (49) for $\widetilde{W}$ in general, we expand it in the energy basis of the Hamiltonian and use the locality of W to bound the contribution of cross terms (using Lemma 8). This accounts for the contributions arising due to non-commutativity of W and H.

Proof of Theorem 29. Recall from Lemma 27 that $W = \sum_i v_i E_i$. We first note that $\mathrm{tr}[\widetilde{W}\eta] = \mathrm{tr}[W\eta] = 0$. From the definition, we have $\mathrm{tr}[(\widetilde{W})^2\eta] = \frac{1}{\mathcal{D}_\Lambda}\|(\widetilde{W})^2\|_F^2$. To begin with, we observe

$$\|W\|_F^2 = \mathcal{D}_\Lambda \sum_{i=1}^{m} v_i^2, \tag{50}$$

which holds since the basis E_i satisfies $\|E_i\|_F^2 = \mathcal{D}_\Lambda$ and $\mathrm{tr}[E_iE_j] = 0$ if $i \neq j$. Define P_s^H as the projection onto the energy range $(s, s+1]$ of H.

$$P_s^H := \sum_{j:\mathcal{E}_j\in(s,s+1]} |j\rangle\langle j|. \tag{51}$$

Using the identity $\sum_s P_s^H = \mathbb{1}_\Lambda$ and the definition of $\widetilde{W}$, let us expand

$$\begin{aligned}
\|W\|_F^2 &= \sum_{s,s'=-\infty}^{\infty} \|P_{s'}^H W P_s^H\|_F^2, \\
\|\widetilde{W}\|_F^2 &= \sum_{s,s'=-\infty}^{\infty} \left\| \int_{-\infty}^{\infty} dt P_{s'}^H f_\beta(t) e^{-iHt} W e^{iHt} P_s^H \right\|_F^2 \qquad (52) \\
&= \sum_{s,s'=-\infty}^{\infty} \left\| \sum_{\substack{j:\mathcal{E}_j\in(s,s+1] \\ k:\mathcal{E}_k\in(s,s+1]}} W_{j,k} \tilde{f}_\beta(|\mathcal{E}_j - \mathcal{E}_k|) P_{s'}^H |j\rangle\langle k| P_s^H \right\|_F^2 . \qquad (53)
\end{aligned}$$

By using the inequality

$$\tilde{f}_\beta(\omega) = \frac{\tanh(\beta\omega/2)}{\beta\omega/2} \geq \frac{1}{\frac{\beta}{2}|\omega| + 1},$$

we have

$$\begin{aligned}
&\left\| \sum_{\substack{j:\mathcal{E}_j\in(s,s+1] \\ k:\mathcal{E}_k\in(s,s+1]}} W_{j,k} \tilde{f}_\beta(|\mathcal{E}_j - \mathcal{E}_k|) P_{s'}^H |j\rangle\langle k| P_s^H \right\|_F^2 \\
&= \sum_{j:\mathcal{E}_j\in(s,s+1]} \sum_{k:\mathcal{E}_j\in(s',s'+1]} [\tilde{f}_\beta(|\mathcal{E}_j - \mathcal{E}_k|)]^2 |W_{j,k}|^2 \\
&\geq \frac{1}{\left(\frac{\beta}{2}(|s-s'|+1)+1\right)^2} \sum_{j:\mathcal{E}_j\in(s,s+1]} \sum_{k:\mathcal{E}_j\in(s',s'+1]} |W_{j,k}|^2 = \frac{\|P_{s'}^H W P_s^H\|_F^2}{\left(\frac{\beta}{2}(|s-s'|+1)+1\right)^2}. \qquad (54)
\end{aligned}$$

Plugging this lower bound in Eq. (53) gives the following lower bound for $\|\widetilde{W}\|_F^2$:

$$\|\widetilde{W}\|_F^2 \geq \sum_{s,s'=-\infty}^{\infty} \frac{\|P_{s'}^H W P_s^H\|_F^2}{\left(\frac{\beta}{2}(|s-s'|+1)+1\right)^2} = \sum_{s_0=-\infty}^{\infty} \sum_{s_1:\frac{s_0+s_1}{2}\in\mathbb{Z}} \frac{\|P_{(s_0+s_1)/2}^H W P_{(s_0-s_1)/2}^H\|_F^2}{\left(\frac{\beta}{2}(|s_1|+1)+1\right)^2}, \tag{55}$$

where we have introduced $s_0 = s + s'$, $s_1 = s - s'$. Let us consider the last expression for a fixed s_0, introducing a cut-off parameter $\bar{s}$ which we fix eventually:

$$\begin{aligned}
&\sum_{s_1:\frac{s_0+s_1}{2}\in\mathbb{Z}} \frac{\|P_{(s_0+s_1)/2}^H W P_{(s_0-s_1)/2}^H\|_F^2}{[\frac{\beta}{2}(|s_1|+1)+1]^2} \\
\geq& \frac{1}{\left(\frac{\beta}{2}(\bar{s}+1)+1\right)^2} \left(\sum_{s_1:\frac{s_0+s_1}{2}\in\mathbb{Z},|s_1|\leq\bar{s}} \|P_{(s_0+2s_1')/2}^H W P_{(s_0-2s_1')/2}^H\|_F^2 \right) \\
=& \frac{1}{\left(\frac{\beta}{2}(\bar{s}+1)+1\right)^2} \left(\sum_{s_1:\frac{s_0+s_1}{2}\in\mathbb{Z}} \|P_{(s_0+s_1)/2}^H W P_{(s_0-s_1)/2}^H\|_F^2 - \sum_{s_1:\frac{s_0+s_1}{2}\in\mathbb{Z},|s_1|>\bar{s}} \|P_{(s_0+s_1)/2}^H W P_{(s_0-s_1)/2}^H\|_F^2 \right).
\end{aligned} \tag{56}$$

By combining the inequalities (55) and (56), we obtain

$$\|\widetilde{W}\|_F^2 \geq \frac{\|W\|_F^2}{\left(\frac{\beta}{2}(\bar{s}+1)+1\right)^2} - \frac{1}{\left(\frac{\beta}{2}(\bar{s}+1)+1\right)^2} \sum_{s_0=-\infty}^{\infty} \sum_{s_1:\frac{s_0+s_1}{2}\in\mathbb{Z},|s_1|>\bar{s}} \|P_{(s_0+s_1)/2}^H W P_{(s_0-s_1)/2}^H\|_F^2. \tag{57}$$

Now, we will estimate the second term in (57). Since the subspaces $P_{(s_0+s_1)/2}^H$ and $P_{(s_0-s_1)/2}^H$ are sufficiently far apart in energy, we can use the exponential concentration on the spectrum [AKL16] (as stated in Lemma 8) to obtain the following: for $W = \sum_i v_i E_i$, we have

$$\begin{aligned}
\|P_{(s_0+s_1)/2}^H W P_{(s_0-s_1)/2}^H\| &\leq \sum_{i=1}^{m} v_i \|P_{(s_0+s_1)/2}^H E_i P_{(s_0-s_1)/2}^H\| \\
&\leq Ce^{-\lambda(|s_1|-1-\kappa)} \sum_{i=1}^{m} |v_i| \leq Cme^{-\lambda(|s_1|-1-\kappa)} \max_i |v_i|.
\end{aligned} \tag{58}$$

where we use the condition that E_i are tensor product of Pauli operators with weight at most κ, and the parameters C and λ are $\mathcal{O}(1)$ constants (see Lemma 8 for their explicit forms). Then, the second term in (57) can be upper-bounded by

$$
\begin{aligned}
&\sum_{s_0=-\infty}^{\infty} \sum_{s_1:\frac{s_0+s_1}{2}\in\mathbb{Z},|s_1|>\bar{s}} \|P^H_{(s_0+s_1)/2} W P^H_{(s_0-s_1)/2}\|_F^2 \\
&\overset{(1)}{\le} \sum_{s_0=-\infty}^{\infty} \sum_{s_1:\frac{s_0+s_1}{2}\in\mathbb{Z},|s_1|>\bar{s}} \|P^H_{(s_0+s_1)/2} W P^H_{(s_0-s_1)/2}\|^2 \cdot \|P^H_{(s_0+s_1)/2}\|_F^2 \\
&\overset{(2)}{\le} \sum_{|s_1|>\bar{s}} \sum_{s_0:\frac{s_0+s_1}{2}\in\mathbb{Z}} \|P^H_{(s_0+s_1)/2}\|_F^2 \cdot C^2 m^2 e^{-2\lambda(|s_1|-1-\kappa)} \max_i v_i^2 \\
&= \mathcal{D}_\Lambda C^2 m^2 e^{2\lambda(1+\kappa)} \max_i v_i^2 \sum_{|s_1|\ge\bar{s}+1} e^{-2\lambda|s_1|} \\
&\overset{(3)}{\le} \mathcal{D}_\Lambda \max_i v_i^2 \frac{C^2 m^2 e^{2\lambda(\kappa+1)}}{2\lambda} e^{-2\lambda\bar{s}} \overset{(4)}{\le} \mathcal{D}_\Lambda \frac{C^2 m^2 e^{2\lambda(\kappa+1)}}{2\lambda} e^{-2\lambda\bar{s}} \left(\sum_i v_i^2\right),
\end{aligned}
\tag{59}
$$

where inequality (1) follows from Eq. (1), (2) follows from Eq. (58), (3) follows from Fact 1 and (4) follows from $\max_i v_i^2 \le \sum_i v_i^2$. Therefore, by applying Eq. (50) and (59) to (57), we arrive at the lower bound as

$$
\|\widetilde{W}\|_F^2 \ge \frac{\mathcal{D}_\Lambda}{[\beta(\bar{s}+1)/2+1]^2} \left(\sum_{i=1}^m v_i^2\right) \left(1 - \frac{C^2 m^2 e^{2\lambda(\kappa+1)}}{2\lambda} e^{-2\lambda\bar{s}}\right). \tag{60}
$$

Since $\lambda, C, \kappa = \mathcal{O}(1)$, by choosing $\bar{s} = \mathcal{O}(\log(m))$, we obtain the main inequality (49). This completes the proof. □ □

4.5 Step 2: From global to local variance

Having seen how to obtain a lower bound for $\widetilde{W}$ at the infinite temperature as in Theorem 29, we move on to the more challenging task of proving a variance lower bound at *finite* temperatures.

A main challenge in bounding the variance of the operator $\widetilde{W}$ is that even though $\widetilde{W}$ is a sum of quasi-local operators, it acts globally on all the sites and hence, its spectral properties might scale badly with the system size. To address this issue, we will follow the strategy introduced in Section 1.5 and reviewed in Section 4.2.2 to reduce the problem to the study of a more local operator. For some $i \in \Lambda$, recall the definition of $\widetilde{W}_{(i)}$ from the subsection 1.5:

$$
\widetilde{W}_{(i)} := \widetilde{W} - \mathrm{tr}_i[\widetilde{W}] \otimes \frac{\mathbb{1}_i}{d} \tag{61}
$$

The operator $\widetilde{W}_{(i)}$ includes essentially the terms in $\widetilde{W}$ that are supported on the ith site. We now show how the variance of the global operator $\widetilde{W}$ can be related to the variance of $\widetilde{W}_{(i)}$.

Remark 30. *Instead of working with $\widetilde{W}$, it is more convenient to define a quasi-local operator A that is proportional to $\widetilde{W} - \mathrm{tr}[\rho_\beta \widetilde{W}]\mathbb{1}$. This allows us to express $\mathrm{Var}[\widetilde{W}]$ in terms of $\mathrm{tr}[A^2\rho_\beta]$. The exact choice of this operator will be made later, but all the results in this and upcoming sections hold generally for a (τ, a_1, a_2, ζ)-quasi-local operator A (see Eq. (4)) where $a_2 = \mathcal{O}(1/\beta), a_1 = \mathcal{O}(1)$ are constants, $\zeta = 1$,*

$\tau \leq 1$, *and* $\mathrm{tr}[A\rho_\beta] = 0$. *We note that the quasi-locality of the opeartor* $\widetilde{W}$ *in terms of the definition 3 is given in Section 5.4.*

We will interchangeably use the Frobenius norm to write the variance of A as

$$\mathrm{Var}[A] = \mathrm{tr}(A^2\rho_\beta) = \|A\sqrt{\rho_\beta}\|_F.$$

Using Haar random unitaries, we obtain the integral representation of $A_{(i)} = A - \mathrm{tr}_i[A] \otimes \frac{\mathbb{1}_i}{d}$ as

$$A_{(i)} = A - \frac{1}{d}[\mathrm{tr}_i(A)] \otimes \mathbb{1}_i = A - \int d\mu(U_i) U_i^\dagger A U_i, \tag{62}$$

where $d\mu(U_i)$ is the Haar measure for unitary operator U_i which acts on the ith site. Using triangle inequality, we have

$$\begin{aligned} \|A_{(i)}\sqrt{\rho_\beta}\|_F &\leq \|A\sqrt{\rho_\beta}\|_F + \int d\mu(U_i)\|U_i^\dagger A U_i \sqrt{\rho_\beta}\|_F \\ &= \|A\sqrt{\rho_\beta}\|_F + \int d\mu(U_i)\|A U_i \sqrt{\rho_\beta}\|_F \end{aligned} \tag{63}$$

This implies

$$\|A_{(i)}\sqrt{\rho_\beta}\|_F^2 \leq \Big(\|A\sqrt{\rho_\beta}\|_F + \int d\mu(U_i)\|AU_i\sqrt{\rho_\beta}\|_F\Big)^2 \tag{64}$$

$$\leq 2\|A\sqrt{\rho_\beta}\|_F^2 + 2\Big(\int d\mu(U_i)\|AU_i\sqrt{\rho_\beta}\|_F\Big)^2. \tag{65}$$

In fact, we need a slightly generalized version of this bound in the forthcoming steps of the proof. This modified version allows us to include the effect of a *further* local unitary U_{X_i} that is applied to the Gibbs state. Here, $X_i := B(R, i)$ is a ball around the site i with a constant radius R that we specify later in Section 4.7. We skip the proof since it is the same as the above argument.

Theorem 31. *Let A be a quasi-local operator as in Remark 30 and $A_{(i)} = A - \mathrm{tr}_i[A] \otimes \frac{\mathbb{1}_i}{d}$ defined as in Section 1.5 for some $i \in \Lambda$. Consider local unitaries U_i and U_{X_i} acting on site i and X_i respectively, where X_i is a ball of constant size at site i . The following generalization of (65) holds:*

$$\|U_{X_i}^\dagger A_{(i)} U_{X_i}\sqrt{\rho_\beta}\|_F^2 \leq 2\|AU_{X_i}\sqrt{\rho_\beta}\|_F^2 + 2\Big(\int d\mu(U_i)\|AU_iU_{X_i}\sqrt{\rho_\beta}\|_F\Big)^2. \tag{66}$$

We now move to another ingredient we need for future steps. Although the operator $A_{(i)}$ is mostly localized around the site i, it is still obtained from a quasi-local operator A and hence it is quasi-local itself. Next claim will approximate $A_{(i)}$ by a *local* operator.

Claim 32. *Consider A which is a (τ, a_1, a_2, ζ)-quasi-local operator (see Eq. (4) and Remark 30) where $a_2 = \mathcal{O}(1/\beta), a_1 = \mathcal{O}(1)$ are constants, $\zeta = 1$, and $\tau \leq 1$. For any $i \in \Lambda$, there exists an operator A_{X_i} supported entirely on X_i, such that*

$$\|A_{(i)} - A_{X_i}\| \leq 2a_1 \cdot \left(1 + a_2^{-\frac{1}{\tau}}\right) \cdot \left(\frac{4}{\tau^2}\right)^{\frac{1}{\tau}} \cdot e^{-\frac{a_2}{2}(R)^\tau}.$$

Proof. Using Eq. (4) and the fact that local operators not containing i in their support are removed, we can write

$$A_{(i)} = \sum_{\substack{k, Z \subseteq \Lambda: \\ |Z| \leq k, Z \ni i}} g_k \left(a_Z - \frac{1}{d} [\mathrm{tr}_i(a_Z)] \otimes \mathbb{1}_i \right),$$

Define

$$A_{X_i} = \sum_{\substack{k, Z \subseteq X_i: \\ |Z| \leq k, Z \ni i}} g_k \left(a_Z - \frac{1}{d} [\mathrm{tr}_i(a_Z)] \otimes \mathbb{1}_i \right).$$

be the desired approximations of $A_{(i)}$ by removing all operators that are not contained within X_i. Observe that

$$\begin{aligned}
\|A_{(i)} - A_{X_i}\| &\leq 2 \sum_{\substack{k, Z \subseteq \Lambda: \\ Z \not\subseteq X_i, |Z| \leq k, Z \ni i}} g_k \|a_Z\| \overset{(1)}{\leq} 2 \sum_{\substack{k, Z \subseteq \Lambda: \\ \mathrm{diam}(Z) \geq R, |Z| \leq k, Z \ni i}} g_k \|a_Z\| \\
&\overset{(2)}{\leq} 2 \sum_{k \geq R} g_k \left(\sum_{Z: Z \ni i} \|a_Z\| \right) \overset{(3)}{\leq} 2\zeta \sum_{k \geq R} g_k \\
&\leq 2\zeta a_1 \sum_{k \geq R} e^{-a_2 k^\tau} \overset{(4)}{\leq} 2\zeta a_1 e^{-\frac{a_2}{2} R^\tau} \left(1 + \tau^{-1} [2/(a_2 \tau)]^{1/\tau} \right) \\
&\leq 2\zeta a_1 \cdot \left(1 + a_2^{-\frac{1}{\tau}} \right) \cdot \left(\frac{2}{\tau} \right)^{\frac{2}{\tau}} \cdot e^{-\frac{a_2}{2} (R)^\tau}.
\end{aligned} \tag{67}$$

For inequality (1), note that since $Z \not\subseteq X_i$ and $Z \ni i$, the diameter of Z (recall that Z is a ball) must be larger than the radius of X_i, which is R. Inequality (2) holds since $k \geq |Z| \geq \mathrm{diam}(Z)$, inequality (3) uses Definition 3 and inequality (4) uses Fact 1. Since $\zeta = 1$ for the given A, this completes the proof. □

4.6 Step 3: Bounding the effect of local unitaries

Recall that we reduced the problem of bounding the variance of A (which as explained in Remark 30) is closely related to our original operator $\widetilde{W}$) to that of the operators $\{A_{(i)}\}_{i \in \Lambda}$. These operators are essentially supported on a small number of sites. However, in this process, we introduced other local unitaries that are applied on the state. In order to handle the action of these unitaries, we will use two claims which show that local unitaries do not make much difference in the relative behavior of spectra of A and H. To elaborate, consider any local operator U_X acting on constant number of sites X on the state ρ_β. It is expected that the quantum state $U_X^\dagger \rho_\beta U_X$ has "similar" spectral properties as ρ_β. So if the eigen-spectrum of the operator A is strongly concentrated for ρ_β, one would expect this behavior to hold even for $U_X^\dagger \rho_\beta U_X$. We make this intuition rigorous in this section.

We begin this by introducing some key quantities that are used repeatedly in the proof. For notational simplicity, let

$$H' := H - \frac{1}{\beta} \log Z_\beta, \tag{68}$$

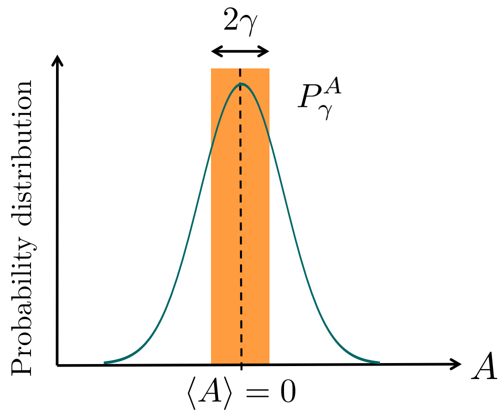


Figure 2: Plot of the probability distribution $\mathrm{tr}[\Pi_\omega \rho_\beta]$, where Π_ω is the projector onto the eigenvectors of A with eigenvalue ω. It is assumed that $\mathrm{tr}[A\rho_\beta] = 0$. For a γ to be chosen in the proof, P_γ^A is the projector onto the subspace of eigenvectors of A with eigenvalue between $[-\gamma, \gamma]$. A lower bound on the variance of A follows if we can show that for a constant γ, the probability mass in the colored range is small (see Equation 71).

which allows us to write $\rho_\beta = e^{-\beta H'}$. As before, we will interchangeably use the Frobenius norm to write

$$\langle A^2 \rangle = \mathrm{tr}(A^2 \rho_\beta) = \|A\sqrt{\rho_\beta}\|_F.$$

We now define the projection operator P_γ^A as follows (see Figure 2):

$$P_\gamma^A := \sum_{\omega \in [-\gamma, \gamma]} \Pi_\omega, \tag{69}$$

where Π_ω is the projector onto the eigenspace of A with eigenvalue ω. We then define δ_γ by

$$\delta_\gamma := 1 - \|P_\gamma^A \sqrt{\rho_\beta}\|_F^2. \tag{70}$$

Using δ_γ, observe that we can lower bound $\langle A^2 \rangle$ as

$$\langle A^2 \rangle = \sum_\omega \omega^2 \langle \omega | \rho_\beta | \omega \rangle \geq \sum_{|\omega| \geq \gamma} \omega^2 \langle \omega | \rho_\beta | \omega \rangle \geq \gamma^2 \sum_{|\omega| \geq \gamma} \langle \omega | \rho_\beta | \omega \rangle \geq \gamma^2 \delta_\gamma. \tag{71}$$

Let $Q_\gamma^A = \mathbb{1} - P_\gamma^A$, then observe that from Eq. (70) that

$$\|P_\gamma^A \sqrt{\rho_\beta} - \sqrt{\rho_\beta}\|_F^2 = \|Q_\gamma^A \sqrt{\rho_\beta}\|_F^2 = \delta_\gamma. \tag{72}$$

Claim 33. *Let c_1, c_2, λ be universal constants. Let $X \subseteq \Lambda$. For every unitary U_X supported on X, we have*

$$\|Q_\gamma^A U_X \sqrt{\rho_\beta}\|_F^2 \leq \exp\left(\lambda |X|\right) \delta_\gamma^{\frac{c_2}{c_2+\beta}}. \tag{73}$$

Let us see a simple application of the claim. It allows us to control the variance of A even after local operations are applied to it. More precisely,

$$\begin{aligned} \|AU_X\sqrt{\rho_\beta}\|_F^2 &= \|AP_\gamma^A U_X \sqrt{\rho_\beta}\|_F^2 + \|A(\mathbb{1} - P_\gamma^A)U_X\sqrt{\rho_\beta}\|_F^2 \\ &\leq \gamma^2 + \|A\|^2 \cdot \|(\mathbb{1} - P_\gamma^A)U_X\sqrt{\rho_\beta}\|_F^2. \end{aligned} \tag{74}$$

By Claim 33, the expression on the second line is upper bounded by $\gamma^2 + \|A\|^2 e^{\Theta(1)|X|}\delta_\gamma^{\Theta(1)/\beta}$. This upper bound on $\|AU_X\sqrt{\rho_\beta}\|_F$ suffices to provide an inverse-polynomial *lower bound* on the variance of A^2, since we can lower bound δ_γ for an appropriate choice of γ. However we now show how one can polynomially improve upon this upper bound (thereby the lower bound on variance) using the following claim. This claim, along the lines of Claim 33, also shows that local unitaries U_X do not change the desired expectation values.

Claim 34. *Let $X \subseteq \Lambda$. For every unitary U_X supported on X, we have*[6]

$$\left\|AQ_\gamma^A U_X\sqrt{\rho_\beta}\right\|_F^2 \leq \frac{1}{\gamma}\cdot \exp\left(\Theta(1)\cdot|X|\right)\delta_\gamma^{\Theta(1)/\beta} + \Theta(1)\cdot|X|^6\cdot\langle A^2\rangle. \tag{75}$$

Proof of both the Claims 33, 34 appear in Section 5.2. An immediate corollary of this claim is the following, that improves upon Eq. (74).

Corollary 35. *Let X be a subset of Λ of size $|X| = \mathcal{O}(1)$. For every unitary U_X supported on X, we have*

$$\|AU_X\sqrt{\rho_\beta}\|_F^2 \leq \gamma^2 + \frac{e^{\Theta(1)\cdot|X|}\delta_\gamma^{\Theta(1)/\beta}}{\gamma} + \Theta(1)|X|^6\langle A^2\rangle. \tag{76}$$

Proof. Similar to Eq. (74), we upper bound $\|AU_X\sqrt{\rho_\beta}\|_F^2$ as

$$\|AU_X\sqrt{\rho_\beta}\|_F^2 = \left\|AP_\gamma^A U_X\sqrt{\rho_\beta}\right\|_F^2 + \left\|AQ_\gamma^A U_X\sqrt{\rho_\beta}\right\|_F^2 \leq \gamma^2 + \left\|AQ_\gamma^A U_X\sqrt{\rho_\beta}\right\|_F^2, \tag{77}$$

since $Q_\gamma^A = \mathbb{1} - P_\gamma^A$. By combining this with Claim 34, the corollary follows. □

4.7 Step 4: Reduction to infinite temperature variance

In Section 4.5, we reduced the variance of a global operator A (which is related to $\widetilde{W}$ via Remark 30) to the quasi-local operators $\{A_{(i)}\}_{i\in\Lambda}$. We now argue that there is some i such that it is possible to bound the *finite* temperature variance of $A_{(i)}$ in terms of its variance at *infinite* temperature. This can be done by applying some local rotations given by a unitary U_{X_i} on the state. The intuition here is that if rotations are allowed, then the eigenvectors of A_{X_i} can be rearranged to yield largest possible variance with ρ_β. This turns out to be larger than the variance with η the infinite temperature state. Define the site i_0 as

$$i_0 := \arg\max_i \|A_{(i)}\sqrt{\eta}\|_F. \tag{78}$$

Claim 36. *Let η be the infinite temperature Gibbs state. There exists a unitary $U_{X_{i_0}}$ supported on X_{i_0} such that*

$$\|U_{X_{i_0}}^\dagger A_{(i_0)} U_{X_{i_0}}\sqrt{\rho_\beta}\|_F \geq \frac{1}{2}\|A_{(i_0)}\sqrt{\eta}\|_F.$$

Proof of Claim 36. We first show the existence of a unitary $U_{X_{i_0}}$ satisfying

$$\|U_{X_{i_0}}^\dagger A_{(i_0)} U_{X_{i_0}}\sqrt{\rho_\beta}\|_F \geq \|A_{(i_0)}\sqrt{\eta}\|_F - 2\|A_{X_{i_0}} - A_{(i_0)}\|. \tag{79}$$

[6] Explicit $\mathcal{O}(1)$ constants that appear in this inequality are made clear in the proof.

Recall that the notation A_{X_i} has been defined in Claim 32. For the proof of the above inequality, we start from the following,

$$\|U_{X_{i_0}}^\dagger A_{(i_0)} U_{X_{i_0}} \sqrt{\rho_\beta}\|_F = \|U_{X_{i_0}}^\dagger \left[(A_{(i_0)} - A_{X_{i_0}}) + A_{X_{i_0}}\right] U_{X_{i_0}} \sqrt{\rho_\beta}\|_F$$
$$\geq \|U_{X_{i_0}}^\dagger A_{X_{i_0}} U_{X_{i_0}} \sqrt{\rho_\beta}\|_F - \|A_{(i_0)} - A_{X_{i_0}}\| \tag{80}$$

and lower-bound the norm of $\|U_{X_{i_0}}^\dagger A_{X_{i_0}} U_{X_{i_0}} \sqrt{\rho_\beta}\|_F$. For this, define

$$\rho_{\beta,X} := \mathrm{tr}_{X^c}(\rho_\beta), \tag{81}$$

where tr_{X^c} is the partial trace operation for the Hilbert space on X^c. We define the spectral decomposition of $A_{X_{i_0}}$ as

$$A_{X_{i_0}} = \sum_{s=1}^{\mathcal{D}_{X_{i_0}}} \varepsilon_s |\varepsilon_s\rangle\langle\varepsilon_s|, \tag{82}$$

where ε_s is ordered as $|\varepsilon_1| \geq |\varepsilon_2| \geq |\varepsilon_3| \geq \cdots$ and $\mathcal{D}_{X_{i_0}}$ is the dimension of the Hilbert space on X_{i_0}. Additionally, define the spectral decomposition of $\rho_{\beta,X_{i_0}}$ as

$$\rho_{\beta,X_{i_0}} = \sum_{s=1}^{\mathcal{D}_{X_{i_0}}} p_s |\mu_s\rangle\langle\mu_s|, \tag{83}$$

where p_s is ordered as $p_1 \geq p_2 \geq p_3 \geq \cdots$ and $|\mu_s\rangle$ is the sth eigenstate of $\rho_{\beta,X_{i_0}}$. We now choose the unitary operator $U_{X_{i_0}}$ such that

$$U_{X_{i_0}} |\mu_s\rangle = |\varepsilon_s\rangle \quad \text{for} \quad s = 1, 2, \ldots, \mathcal{D}_{X_{i_0}} \tag{84}$$

We then obtain

$$U_{X_{i_0}} \rho_{\beta,X_{i_0}} U_{X_{i_0}}^\dagger = \sum_{s=1}^{\mathcal{D}_{X_{i_0}}} p_s |\varepsilon_s\rangle\langle\varepsilon_s|. \tag{85}$$

This implies

$$\|U_{X_{i_0}}^\dagger A_{X_{i_0}} U_{X_{i_0}} \sqrt{\rho_\beta}\|_F^2 = \mathrm{tr}[U_{X_{i_0}}^\dagger A_{X_{i_0}}^2 U_{X_{i_0}} \rho_\beta]$$
$$= \mathrm{tr}[A_{X_{i_0}}^2 U_{X_{i_0}} \rho_\beta U_{X_{i_0}}^\dagger] = \mathrm{tr}_{X_{i_0}}[A_{X_{i_0}}^2 U_{X_{i_0}} \rho_{\beta,X_{i_0}} U_{X_{i_0}}^\dagger]$$
$$= \sum_{s=1}^{\mathcal{D}_{X_{i_0}}} p_s \varepsilon_s^2 \geq \frac{1}{\mathcal{D}_{X_{i_0}}} \sum_{s=1}^{\mathcal{D}_{X_{i_0}}} \varepsilon_s^2 = \|A_{X_{i_0}} \sqrt{\eta}\|_F^2, \tag{86}$$

where the inequality used the fact that p_s, ε_s are given in descending order. Then, the minimization problem of $\sum_s p_s \varepsilon_s$ for $p_s{}_s$ with the constraint $p_1 \geq p_2 \geq p_3 \geq \cdots$ has a solution of $p_1 = p_2 = \cdots = p_{\mathcal{D}_{X_{i_0}}}$. Using the lower bound

$$\|A_{X_{i_0}} \sqrt{\eta}\|_F = \|(A_{X_{i_0}} - A_{(i_0)} + A_{(i_0)})\sqrt{\eta}\|_F \geq \|A_{(i_0)}\sqrt{\eta}\|_F - \|A_{X_{i_0}} - A_{(i_0)}\|, \tag{87}$$

we can reduce inequality (86) to

$$\|U_{X_{i_0}}^\dagger A_{X_{i_0}} U_{X_{i_0}} \sqrt{\rho_\beta}\|_F \geq \|A_{(i_0)}\sqrt{\eta}\|_F - \|A_{X_{i_0}} - A_{(i_0)}\|. \tag{88}$$

By combining the inequalities (80) and (88), we obtain the inequality (79) as follows:

$$\begin{aligned}\|U_{X_{i_0}}^\dagger A_{(i_0)} U_{X_{i_0}} \sqrt{\rho_\beta}\|_F &\geq \|U_{X_{i_0}}^\dagger A_{X_{i_0}} U_{X_{i_0}} \sqrt{\rho_\beta}\|_F - \|A_{(i_0)} - A_{X_{i_0}}\| \\ &\geq \|A_{(i_0)}\sqrt{\eta}\|_F - 2\|A_{(i_0)} - A_{X_{i_0}}\| \,,\end{aligned}$$

To arrive at the final bound we use Claim 32. Fix

$$R = \left\lceil \left(\frac{2}{a_2} \log \frac{8 \cdot 4^{\frac{1}{\tau}} \cdot a_1 \cdot \left(1 + a_2^{-\frac{1}{\tau}}\right)}{\tau^{\frac{2}{\tau}} \|A_{(i_0)}\sqrt{\eta}\|_F} \right)^{\frac{1}{\tau}} \right\rceil$$

in Claim 32. We get

$$\|A_{(i_0)} - A_{X_{i_0}}\| \leq \frac{1}{4}\|A_{(i_0)}\sqrt{\eta}\|_F \tag{89}$$

Substituting $a_2 = \mathcal{O}(1/\beta), a_1 = \mathcal{O}(1), \tau = \mathcal{O}(1)$, we find that we can ensure the condition (89) for

$$R = \mathrm{diam}(X_{i_0}) = \left(\beta \log\left(\frac{1}{\|A_{(i_0)}\sqrt{\eta}\|_F}\right)\right)^{\Theta(1)}. \tag{90}$$

□

The previous claim demonstrates that by applying local unitaries, the finite temperature variance of $A_{(i_0)}$ can be related to $\|A_{(i_0)}\sqrt{\eta}\|_F$. Here, we use the discussion in Section 4.5 and Section 4.6 to prove that the rotated local variance $\|U_{X_{i_0}}^\dagger A_{(i_0)} U_{X_{i_0}} \sqrt{\rho_\beta}\|_F$ in Claim 36 is close to the variance of $A_{(i)}$ and hence, can be related to the variance of the global operator A. By establishing this, we can obtain a lower bound on the variance of A (an in turn $\widetilde{W}$) in terms of the infinite temperature variance $\max_{i\in\Lambda} \|A_{(i)}\sqrt{\eta}\|_F$. This is stated in the following theorem:

Theorem 37. *Let $\beta > 0$, H be a κ-local Hamiltonian on the lattice Λ and $\rho_\beta = \frac{e^{-\beta H}}{\mathrm{tr}(e^{-\beta H})}$. Let A be a (τ, a_1, a_2, ζ)-quasi-local operator (see Eq. (4) and Remark 30) where $a_2 = \mathcal{O}(1/\beta), a_1 = \mathcal{O}(1)$ are constants and we assume $\zeta = 1$, $\tau \leq 1$ and $\mathrm{tr}[A\rho_\beta] = 0$. We have*

$$\langle A^2 \rangle = \mathrm{tr}(A^2\rho_\beta) \geq \left(\max_{i\in\Lambda} \mathrm{tr}[A_{(i)}^2 \eta]\right)^{\beta^{\Theta(1)}}.$$

We remark that the theorem statement above hides several terms that depend on the lattice, such as the lattice dimension, the degree of the graph and the locality of Hamiltonian (which we have fixed to be a constant). Additionally, the assumptions $a_2 = \mathcal{O}(1/\beta), a_1 = \mathcal{O}(1)$ are suitably made for the later application of this bound in Section 4.8, but we note that this theorem also holds for other choices of a_1, a_2, with small modifications to the proof.

Proof. Let $U_{X_{i_0}}$ be the unitary as chosen in Claim 36. Using Theorem 31, we obtain the following upper bound for $U^\dagger_{X_{i_0}} A_{(i_0)} U_{X_{i_0}}$

$$\|U^\dagger_{X_{i_0}} A_{(i_0)} U_{X_{i_0}} \sqrt{\rho_\beta}\|_F^2 \leq \left(\|AU_{X_{i_0}}\sqrt{\rho_\beta}\|_F + \int d\mu(U_{i_0})\|AU_{i_0}U_{X_{i_0}}\sqrt{\rho_\beta}\|_F\right)^2$$
$$\leq 2\|AU_{X_{i_0}}\sqrt{\rho_\beta}\|_F^2 + 2\int d\mu(U_{i_0})\|AU_{i_0}U_{X_{i_0}}\sqrt{\rho_\beta}\|_F^2.$$

Now, we can use Corollary 35 to upper bound $\|AU_{X_{i_0}}\sqrt{\rho_\beta}\|_F^2$ and $\|AU_{i_0}U_{X_{i_0}}\sqrt{\rho_\beta}\|_F$ in the right hand side, which yields

$$\|U^\dagger_{X_{i_0}} A_{(i_0)} U_{X_{i_0}} \sqrt{\rho_\beta}\|_F^2 \leq 4\gamma^2 + \frac{e^{\Theta(1)|X_{i_0}|}\delta_\gamma^{\Theta(1)/\beta}}{\gamma} + \Theta(1)|X_{i_0}|^6\langle A^2\rangle. \tag{91}$$

Using Claim 36, we have

$$\|U^\dagger_{X_{i_0}} A_{(i_0)} U_{X_{i_0}} \sqrt{\rho_\beta}\|_F \geq \frac{1}{2}\|A_{(i_0)}\sqrt{\eta}\|_F. \tag{92}$$

Putting together the upper bound in Eq. (91) and the lower bound in Eq. (92), we have

$$4\gamma^2 + \frac{e^{\Theta(1)|X_{i_0}|}\delta_\gamma^{\Theta(1)/\beta}}{\gamma} + \Theta(1)|X_{i_0}|^6\langle A^2\rangle \geq \frac{1}{4}\|A_{(i_0)}\sqrt{\eta}\|_F^2. \tag{93}$$

By choosing as $\gamma^2 = \|A_{(i_0)}\sqrt{\eta}\|_F^2/32 =: \gamma_0^2$, we obtain

$$\frac{e^{\Theta(1)|X_{i_0}|}\delta_{\gamma_0}^{\Theta(1)/\beta}}{\gamma_0} + \Theta(1)|X_{i_0}|^6\langle A^2\rangle \geq \gamma_0^2. \tag{94}$$

This inequality implies that either

$$\delta_{\gamma_0} \geq \left(\gamma_0^3 e^{-\Theta(1)|X_{i_0}|}\right)^{\beta\cdot\Theta(1)}$$

or

$$\langle A^2\rangle \geq \frac{\Theta(1)\gamma_0^2}{|X_{i_0}|^6}.$$

Combining with Eq. (71), we conclude that

$$\langle A^2\rangle \geq \min\left\{\gamma_0^2\cdot\left(\gamma_0^3 e^{-\Theta(1)|X_{i_0}|}\right)^{\beta\cdot\Theta(1)}, \frac{\Theta(1)\gamma_0^2}{|X_{i_0}|^6}\right\}.$$

Eq. (90) ensures that

$$|X_{i_0}| = \Theta(1)R^D = \beta^{\Theta(1)}\log\left(\frac{1}{\|A_{(i_0)}\sqrt{\eta}\|_F}\right)^{\Theta(1)},$$

where we have used the assumption that lattice dimension D is $\Theta(1)$. Plugging in this expression for $|X_{i_0}|$ with the choice of γ_0, we find

$$\begin{aligned}
\operatorname{tr}(A^2\rho_\beta) = \langle A^2\rangle &\geq \min\left\{\|A_{(i_0)}\sqrt{\eta}\|_F^{\beta\cdot\Theta(1)}\cdot e^{-\beta\Theta(1)|X_{i_0}|}, \frac{\Theta(1)\|A_{(i_0)}\sqrt{\eta}\|_F^{\Theta(1)}}{|X_{i_0}|^6}\right\}\\
&\geq \min\left\{\|A_{(i_0)}\sqrt{\eta}\|_F^{\beta\Theta(1)}, \frac{\Theta(1)\|A_{(i_0)}\sqrt{\eta}\|_F^{\Theta(1)}}{\beta^{\Theta(1)}\log(\frac{1}{\|A_{(i_0)}\sqrt{\eta}\|_F})^{\Theta(1)}}\right\}\\
&\geq \min\left\{\|A_{(i_0)}\sqrt{\eta}\|_F^{\beta\Theta(1)}, \frac{\Theta(1)\|A_{(i_0)}\sqrt{\eta}\|_F^{\Theta(1)}}{\beta^{\Theta(1)}}\right\}\\
&\geq \|A_{(i_0)}\sqrt{\eta}\|_F^{\beta\Theta(1)}.
\end{aligned}$$

Since we chose i_0 in Eq. (78) such that $\|A_{(i_0)}\sqrt{\eta}\|_F = \max_i \|A_{(i)}\sqrt{\eta}\|_F$, this proves the theorem. □

4.8 Final step: Putting things together

We are now ready to apply the results of the past steps to prove Theorem 21. The bound in Theorem 37 in the previous section is stated for a general quasi-local observable A. As discussed in Remark 30, a special choice of A is when it is proportional to $\widetilde{W} - \operatorname{tr}[\rho_\beta \widetilde{W}]\mathbb{1}$ where recall that for an arbitrary $v \in \mathbb{R}^m$, $W = \sum_i v_i E_i$ and $\widetilde{W}$ are the operators defined in Lemma 27. In Section 5.4, we show that $\widetilde{W}$ is a $\left(1/D, \mathcal{O}(1), \mathcal{O}(1/\beta), c_*\beta^{2D+1}\left(\max_{i\in\Lambda}|v_i|\right)\right)$-quasi-local operator, for $c_* = \mathcal{O}(1)$. Thus, the following operator

$$A^* = \frac{\beta^{-2D-1}}{c_*\max_{i\in\Lambda}|v_i|}(\widetilde{W} - \operatorname{tr}[\rho_\beta\widetilde{W}]\mathbb{1}),$$

is $(\mathcal{O}(1), \mathcal{O}(1), \mathcal{O}(1/\beta), 1)$-quasi-local and satisfies $\operatorname{tr}[A^*\rho_\beta] = 0$. We now apply Theorem 37 to the operator A^* to prove Theorem 21. We need to estimate $\max_i \operatorname{tr}[A^{*2}_{(i)}\eta]$. Consider the following equality obtained from the definition of A^*:

$$\max_{i\in\Lambda}\operatorname{tr}[(A^*_{(i)})^2\eta] = \frac{\beta^{-4D-2}}{c_*^2\left(\max_{i\in\Lambda}|v_i|\right)^2}\left(\max_{i\in\Lambda}\operatorname{tr}[(\widetilde{W}_{(i)})^2\eta]\right).$$

The following lemma is shown in Appendix 5.5.

Lemma 38. *It holds that*

$$\max_{i\in\Lambda}(\operatorname{tr}[(\widetilde{W}_{(i)})^2\eta]) = \frac{\Theta(1)}{(\beta\log(\beta)+1)^{2D+2}}\left(\max_{i\in\Lambda}v_i^2\right).$$

This implies

$$\max_{i\in\Lambda}\operatorname{tr}[(A^*_{(i)})^2\eta] = \frac{\Theta(1)}{\beta^{4D+2}\left(\beta\log(\beta)+1\right)^{2D+2}\left(\max_{i\in\Lambda}|v_i|\right)^2}\left(\max_{i\in\Lambda}v_i^2\right) = \frac{1}{\beta^{\Theta(1)}}.$$

Using this lower bound in Theorem 37, we find

$$\begin{aligned}
\langle(\widetilde{W})^2\rangle - \left(\langle\widetilde{W}\rangle\right)^2 &= c_*^2\beta^{4D+2}\left(\max_{i\in\Lambda}|v_i|\right)^2\left(\langle(A^*)^2\rangle - (\langle A^*\rangle)^2\right) \\
&= \beta^{\Theta(1)}\left(\max_{i\in\Lambda}|v_i|\right)^2\cdot\left(\max_i \mathrm{tr}[(A^*_{(i)})^2\eta]\right)^{\beta^{\Theta(1)}} \\
&= \beta^{\Theta(1)}\cdot\left(\frac{1}{\beta^{\Theta(1)}}\right)^{\beta^{\Theta(1)}}\cdot\left(\max_{i\in\Lambda}|v_i|\right)^2 \\
&\overset{(1)}{=} \beta^{\Theta(1)}\cdot e^{-\beta^{\Theta(1)}}\left(\max_{i\in\Lambda}|v_i|\right)^2 \geq \beta^{\Theta(1)}\cdot\frac{e^{-\beta^{\Theta(1)}}}{m}\left(\sum_i v_i^2\right),
\end{aligned}$$

where we used $\beta^{-\Theta(1)} \geq e^{-\Theta(\beta)}$ in (1). Putting together the bound above with Eq. (48), we find that for every $v \in \mathbb{R}^m$,

$$v^\top\cdot\left(\nabla^2\log Z_\beta(\lambda)\right)\cdot v \geq \beta^2\,\mathrm{Var}[\widetilde{W}] \geq \beta^{\Theta(1)}\cdot\frac{e^{-\beta^{\Theta(1)}}}{m}\sum_{i=1}^m v_i^2.$$

This establishes Theorem 21.

5 Deferred proofs

5.1 Fourier transform of tanh($\beta\omega/2$)/($\beta\omega/2$)

We here derive the Fourier transform of

$$\tilde{f}_\beta(\omega) = \frac{\tanh(\beta\omega/2)}{\beta\omega/2},$$

which is

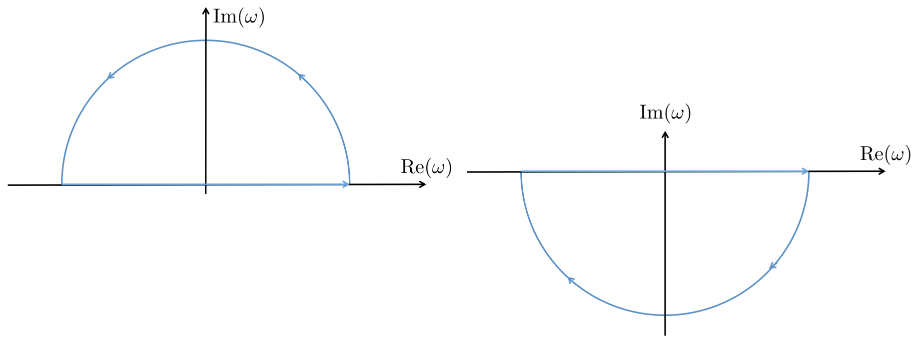
$$f_\beta(t) := \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t}\tilde{f}_\beta(\omega)d\omega.$$

For the calculation of the Fourier transform, we first consider the case of $t > 0$. By defining C^+ as a integral path as in Fig. 3 (a), we obtain

$$\begin{aligned}
\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t}\tilde{f}_\beta(\omega)d\omega &= \frac{1}{2\pi}\int_{C^+} e^{i\omega t}\tilde{f}_\beta(\omega)d\omega \\
&= i\sum_{m=0}^{\infty}\mathrm{Res}_{\omega=i\pi+2im\pi}[e^{i\omega t}\tilde{f}_\beta(\omega)].
\end{aligned} \tag{95}$$

Note that the singular points of $[e^{i\omega t}\tilde{f}_\beta(\omega)]$ are given by $\beta\omega = i\pi(2m+1)$ with m integers. We can calculate the residue as

$$\mathrm{Res}_{\beta\omega=i\pi+2im\pi}[e^{i\omega t}\tilde{f}_\beta(\omega)] = \frac{4e^{-(2m+1)\pi t/\beta}}{\beta\pi}\frac{-i}{2m+1} \tag{96}$$



(a) Integral path C^+ of ω for $t > 0$ (b) Integral path C^- of ω for $t < 0$

Figure 3: Cauchy's integral theorem for the calculation of the Fourier transform.

We thus obtain

$$f_\beta(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t}\tilde{f}_\beta(\omega)d\omega = \frac{4}{\beta\pi}\sum_{m=0}^{\infty}\frac{e^{-(2m+1)\pi t/\beta}}{2m+1}. \tag{97}$$

for $t > 0$.

We can perform the same calculation for $t < 0$. In this case, we define C^- as a integral path as in Fig. 3 (b), and obtain

$$f_\beta(t) = \frac{1}{2\pi}\int_{C^-} e^{i\omega t}\tilde{f}_\beta(\omega)d\omega = -i\sum_{m=0}^{\infty}\mathrm{Res}_{\omega=-i\pi-2im\pi}[e^{i\omega t}\tilde{f}_\beta(\omega)]. \tag{98}$$

By using

$$\mathrm{Res}_{\omega=-i\pi-2im\pi}[e^{i\omega t}\tilde{f}_\beta(\omega)] = \frac{4e^{(2m+1)\pi t/\beta}}{\beta\pi}\frac{i}{2m+1}, \tag{99}$$

we have

$$f_\beta(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t}\tilde{f}_\beta(\omega)d\omega = \frac{4}{\beta\pi}\sum_{m=0}^{\infty}\frac{e^{(2m+1)\pi t/\beta}}{2m+1}. \tag{100}$$

for $t < 0$. By combining the above expressions for $f_\beta(t)$, we arrive at

$$f_\beta(t) = \frac{4}{\beta\pi}\sum_{m=0}^{\infty}\frac{e^{-(2m+1)\pi|t|/\beta}}{2m+1}. \tag{101}$$

The summation is calculated as

$$\sum_{m=0}^{\infty}\frac{e^{-(2m+1)x}}{2m+1} = \int_x^{\infty}\sum_{m=0}^{\infty}e^{-(2m+1)x'}dx' = \int_x^{\infty}\frac{1}{e^{x'}-e^{-x'}}dx' = \frac{1}{2}\log\frac{e^x+1}{e^x-1} \tag{102}$$

for $x > 0$, which yields

$$f_\beta(t) = \frac{2}{\beta\pi}\log\frac{e^{\pi|t|/\beta}+1}{e^{\pi|t|/\beta}-1}. \tag{103}$$

Since

$$\log \frac{e^{\pi|t|/\beta}+1}{e^{\pi|t|/\beta}-1} \leq \frac{2}{e^{\pi|t|/\beta}-1},$$

$f_\beta(t)$ shows an exponential decay in $|t|$.

5.2 Proof of Claims 33 and 34

Proof of Claim 33. Recall that the goal is to show that for every $X \subseteq \Lambda$ and arbitrary unitaries U_X,

$$\|Q_\gamma^A U_X \sqrt{\rho_\beta}\|_F^2 \leq c_1 e^{\lambda|X|} \delta_\gamma^{\frac{c_2}{c_2+\beta}}, \tag{104}$$

where $Q_\gamma^A = \mathbb{1} - P_\gamma^A$ and P_γ^A was defined in Eq. (69) and $H' := H - \frac{1}{\beta}\log Z_\beta$, which allows us to write $\rho_\beta = e^{-\beta H'}$. To prove this inequality, we start from the following expression:

$$\|Q_\gamma^A U_X \sqrt{\rho_\beta}\|_F^2 = \left\| \sum_{m\in\mathbb{Z}} Q_\gamma^A U_X P_m^{H'} \sqrt{\rho_\beta} \right\|_F^2 = \sum_{m\in\mathbb{Z}} \left\| Q_\gamma^A U_X P_m^{H'} \sqrt{\rho_\beta} \right\|_F^2 \tag{105}$$

with

$$P_m^{H'} := \sum_{j:\mathcal{E}_j \in (m,m+1]} |j\rangle\langle j|, \tag{106}$$

where $|j\rangle$ is the eigenvector of the Hamiltonian H' with $\mathcal{E}_j$ the corresponding eigenvalue. Note that $\sum_{m\in\mathbb{Z}} P_m^{H'} = \mathbb{1}$ and we have $P_m^{H'} = 0$ for $m \notin [-\|H'\|, \|H'\|]$. For some $\Delta \in \mathbb{N}$ which we pick later, we now decompose $\left\| Q_\gamma^A U_X P_m^{H'} \sqrt{\rho_\beta} \right\|_F^2$ as a sum of the following quantities

$$\left\| Q_\gamma^A U_X P_m^{H'} \sqrt{\rho_\beta} \right\|_F^2 = \left\| Q_\gamma^A (P_{>m+\Delta}^{H'} + P_{<m-\Delta}^{H'} + P_{[m-\Delta,m+\Delta]}^{H'}) U_X P_m^{H'} \sqrt{\rho_\beta} \right\|_F^2, \tag{107}$$

where

$$P_{>m+\Delta}^{H'} := \sum_{m'>m+\Delta} P_{m'}^{H'}, \quad P_{<m+\Delta}^{H'} := \sum_{m'<m+\Delta} P_{m'}^{H'}, \quad P_{[m-\Delta,m+\Delta]}^{H'} := \sum_{m-\Delta\leq m'\leq m+\Delta} P_{m'}^{H'}. \tag{108}$$

Summing over all $m \in \mathbb{Z}$ in Eq. (107) and using Eq. (105) followed by the triangle inequality gives us the following inequality

$$\begin{aligned}
&\|Q_\gamma^A U_X \sqrt{\rho_\beta}\|_F^2 \\
&\leq \underbrace{2\sum_{m\in\mathbb{Z}} \left\| Q_\gamma^A P_{[m-\Delta,m+\Delta]}^{H'} U_X P_m^{H'} \sqrt{\rho_\beta} \right\|_F^2}_{:=(1)} + \underbrace{2\sum_{m\in\mathbb{Z}} \left\| Q_\gamma^A (P_{>m+\Delta}^{H'} + P_{<m-\Delta}^{H'}) U_X P_m^{H'} \sqrt{\rho_\beta} \right\|_F^2}_{:=(2)}.
\end{aligned} \tag{109}$$

We first bound (1) in Eq. (109). Note that for every m,

$$\begin{aligned}
\left\| Q_\gamma^A P_{[m-\Delta,m+\Delta]}^{H'} U_X P_m^{H'} \sqrt{\rho_\beta} \right\|_F^2 &\leq \|P_m^{H'} \sqrt{\rho_\beta}\|^2 \cdot \left\| Q_\gamma^A P_{[m-\Delta,m+\Delta]}^{H'} \right\|_F^2 \\
&\leq e^{-\beta m} \left\| Q_\gamma^A P_{[m-\Delta,m+\Delta]}^{H'} \right\|_F^2,
\end{aligned} \tag{110}$$

where the first inequality used Eq. (1). The expression in the last line can be upper bounded as

$$\begin{aligned} e^{-\beta m}\left\|Q_\gamma^A P^{H'}_{[m-\Delta,m+\Delta]}\right\|_F^2 &= e^{-\beta m}\mathrm{tr}\left[Q_\gamma^A P^{H'}_{[m-\Delta,m+\Delta]}\right] \\ &\le e^{-\beta m}e^{\beta(m+\Delta+1)}\mathrm{tr}\left[Q_\gamma^A P^{H'}_{[m-\Delta,m+\Delta]}\rho_\beta P^{H'}_{[m-\Delta,m+\Delta]}\right] \\ &= e^{\beta(\Delta+1)}\left\|Q_\gamma^A P^{H'}_{[m-\Delta,m+\Delta]}\sqrt{\rho_\beta}\right\|_F^2. \end{aligned}$$

where the inequality follows from

$$e^{-\beta(m+\Delta+1)}P^{H'}_{[m-\Delta,m+\Delta]} \preceq P^{H'}_{[m-\Delta,m+\Delta]}\rho_\beta P^{H'}_{[m-\Delta,m+\Delta]}.$$

Thus we conclude, from Equation (110), that

$$\begin{aligned} \left\|Q_\gamma^A P^{H'}_{[m-\Delta,m+\Delta]}U_X P^{H'}_m\sqrt{\rho_\beta}\right\|_F^2 &\le e^{\beta(\Delta+1)}\left\|Q_\gamma^A P^{H'}_{[m-\Delta,m+\Delta]}\sqrt{\rho_\beta}\right\|_F^2 \\ &= e^{\beta(\Delta+1)}\sum_{m'\in[m-\Delta,m+\Delta]}\left\|Q_\gamma^A P^{H'}_{m'}\sqrt{\rho_\beta}\right\|_F^2. \end{aligned}$$

So the first term (1) in Eq. (109) can be bounded by

$$\begin{aligned} \sum_{m\in\mathbb{Z}}\left\|Q_\gamma^A P^{H'}_{[m-\Delta,m+\Delta]}U_X P^{H'}_m\sqrt{\rho_\beta}\right\|_F^2 &\le e^{\beta(\Delta+1)}\sum_{m\in\mathbb{Z}}\sum_{m'\in[m-\Delta,m+\Delta]}\left\|Q_\gamma^A P^{H'}_{m'}\sqrt{\rho_\beta}\right\|_F^2 \\ &\overset{(1)}{=} e^{\beta(\Delta+1)}\cdot(2\Delta+1)\sum_{m'\in\mathbb{Z}}\left\|Q_\gamma^A P^{H'}_{m'}\sqrt{\rho_\beta}\right\|_F^2 \\ &= (2\Delta+1)e^{\beta(\Delta+1)}\left\|Q_\gamma^A\sqrt{\rho_\beta}\right\|_F^2 = (2\Delta+1)e^{\beta(\Delta+1)}\delta_\gamma, \end{aligned} \tag{111}$$

where in (1) we use the fact that each m' appears $(2\Delta+1)$ times in the summation $\sum_{m\in\mathbb{Z}}\sum_{m'\in[m-\Delta,m+\Delta]}$.

We now move on to upper bound (2) in Eq. (109) as follows. We have

$$\left\|Q_\gamma^A(P^{H'}_{>m+\Delta}+P^{H'}_{<m-\Delta})U_X P^{H'}_m\sqrt{\rho_\beta}\right\|_F^2 \le \left\|(P^{H'}_{>m+\Delta}+P^{H'}_{<m-\Delta})U_X P^{H'}_m\right\|^2\cdot\|P^{H'}_m\sqrt{\rho_\beta}\|_F^2, \tag{112}$$

where we use $\|Q_\gamma^A\|\le 1$. Using Lemma 8, we obtain

$$\left\|(P^{H'}_{>m+\Delta}+P^{H'}_{<m-\Delta})U_X P^{H'}_m\right\| \le Ce^{-\lambda(\Delta-|X|)}, \tag{113}$$

where C and λ are universal constants. Plugging Eq. (113) into Eq. (112), we get

$$\left\|Q_\gamma^A(P^{H'}_{>m+\Delta}+P^{H'}_{<m-\Delta})U_X P^{H'}_m\sqrt{\rho_\beta}\right\|_F^2 \le C^2e^{-2\lambda(\Delta-|X|)}\|P^{H'}_m\sqrt{\rho_\beta}\|_F^2. \tag{114}$$

With this, we can bound (2) in Eq. (109) by

$$\sum_{m\in\mathbb{Z}}\left\|Q_\gamma^A(P^{H'}_{>m+\Delta}+P^{H'}_{<m-\Delta})U_X P^{H'}_m\sqrt{\rho_\beta}\right\|_F^2 \le \sum_{m\in\mathbb{Z}}C^2e^{-2\lambda(\Delta-|X|)}\|P^{H'}_m\sqrt{\rho_\beta}\|_F^2 = C^2e^{-2\lambda(\Delta-|X|)}, \tag{115}$$

where the equality used the fact that $\sum_{m\in\mathbb{Z}} \|P_m^{H'}\sqrt{\rho_\beta}\|_F^2 = \mathrm{tr}(\rho_\beta) = 1$. Putting together Eq. (111) and (115) into Eq. (109), we finally obtain the upper bound of

$$\|Q_\gamma^A U_X \sqrt{\rho_\beta}\|_F^2 \le (4\Delta + 2)e^{\beta(\Delta+1)}\delta_\gamma + 2C^2 e^{-2\lambda(\Delta-|X|)}. \tag{116}$$

We let $\Delta = c\beta^{-1}\log(1/\delta_\gamma)$, which gives

$$\|Q_\gamma^A U_X \sqrt{\rho_\beta}\|_F^2 \le c_1 e^{\lambda|X|}\delta_\gamma^{\frac{c_2}{c_2+\beta}}, \tag{117}$$

for some universal constants c_1, c_2. This proves the claim statement. $\square$

We now proceed to prove Claim 34.

Proof of Claim 34. Recall that the aim is to prove that for every $X \subseteq \Lambda$ and unitary U_X we have

$$\left\|AQ_\gamma^A U_X \sqrt{\rho_\beta}\right\|_F^2 \le \frac{e^{\Theta(1)|X|}\delta_\gamma^{\Theta(1)/\beta}}{\gamma} + \frac{\Theta(1)|X|^5}{\gamma^5}\langle A^2\rangle$$

We let c_5, λ_1, τ_1 be $\Theta(1)$ constants as defined in Lemma 9 and $c_1, c_2, \lambda = \mathcal{O}(1)$ be constants given by Claim 33. For the proof, we first decompose Q_γ^A as

$$Q_\gamma^A = \sum_{s=1}^{\infty} P_s^A, \quad P_s^A := P_{(s\gamma,(s+1)\gamma]}^A + P_{[-(s+1)\gamma,-s\gamma)}^A, \quad P_0^A := P_{[-\gamma,\gamma]}^A, \tag{118}$$

where $P_{[a,b]}^A$ is defined as $P_{[a,b]}^A := \sum_{a\le\omega\le b}\Pi_\omega$ (where Π_ω is the subspace spanned by the eigenvectors of A with eigenvalue ω). Using this notation, observe that $\|AQ_\gamma^A U_X\sqrt{\rho_\beta}\|_F$ can be bounded by

$$\begin{aligned}\left\|AQ_\gamma^A U_X \sqrt{\rho_\beta}\right\|_F^2 = \sum_{s=1}^{\infty}\left\|AP_s^A U_X\sqrt{\rho_\beta}\right\|_F^2 &\le \sum_{s=1}^{\infty}\left\|AP_s^A\right\|^2 \cdot \left\|P_s^A U_X\sqrt{\rho_\beta}\right\|_F^2 \\ &\le \gamma^2\sum_{s=1}^{\infty}(s+1)^2\left\|P_s^A U_X\sqrt{\rho_\beta}\right\|_F^2,\end{aligned} \tag{119}$$

where we use $\|AP_s^A\| \le \gamma(s+1)$ from the definition (118) of P_s^A. The norm $\left\|P_s^A U_X\sqrt{\rho_\beta}\right\|_F$ is bounded from above by

$$\left\|P_s^A U_X\sqrt{\rho_\beta}\right\|_F = \left\|P_s^A U_X \sum_{s'=0}^{\infty} P_{s'}^A\sqrt{\rho_\beta}\right\|_F \le \sum_{s'=0}^{\infty}\left\|P_s^A U_X P_{s'}^A\sqrt{\rho_\beta}\right\|_F, \tag{120}$$

where we use $\sum_{s'=0}^{\infty} P_{s'}^A = \mathbb{1}$ in the first equation and in the second inequality we use the triangle inequality for the Frobenius norm. Using Lemma 9 we additionally have

$$\|P_s^A U_X P_{s'}^A\| \le c_5|X|e^{-(\lambda_1\gamma|s-s'|/|X|)^{1/\tau_1}} \quad \text{for every } s, s' \ge 0, \tag{121}$$

where c_5, λ_1 are given in Lemma 9. Using this, we have

$$
\begin{aligned}
\left\|P_s^A U_X P_0^A \sqrt{\rho_\beta}\right\|_F &= \left\|P_s^A U_X P_0^A \sqrt{\rho_\beta}\right\|_F^{1/2} \cdot \left\|P_s^A U_X P_0^A \sqrt{\rho_\beta}\right\|_F^{1/2} \\
&\overset{(1)}{\leq} \left(2c_1 e^{\lambda|X|} \delta_\gamma^{\frac{c_2}{c_2+\beta}}\right)^{1/2} \cdot \|P_s^A U_X P_0^A \sqrt{\rho_\beta}\|_F^{1/2} \\
&\overset{(2)}{\leq} \left(2c_1 e^{\lambda|X|} \delta_\gamma^{\frac{c_2}{c_2+\beta}}\right)^{1/2} \cdot \left\|P_s^A U_X P_0^A\right\|^{1/2} \cdot \|\sqrt{\rho_\beta}\|_F^{1/2} \\
&\overset{(3)}{\leq} \left(2c_1 e^{\lambda|X|} \delta_\gamma^{\frac{c_2}{c_2+\beta}}\right)^{1/2} \cdot \left(c_5 |X| e^{-(\lambda_1 \gamma |s|/|X|)^{1/\tau_1}}\right)^{1/2} \cdot 1 \\
&\overset{(4)}{=} (\delta'_\gamma)^{1/2} \cdot \left(c_5 |X| e^{-(\lambda_1 \gamma |s|/|X|)^{1/\tau_1}}\right)^{1/2},
\end{aligned}
\tag{122}
$$

where inequality (1) uses $\left\|P_s^A U_X P_0^A \sqrt{\rho_\beta}\right\|_F \leq 2c_1 \delta_\gamma^{\frac{c_2}{c_2+\beta}}$ [7], inequality (2) uses Eq. (1), inequality (3) uses Eq. (121) and the fact that $\|\sqrt{\rho_\beta}\|_F = \mathrm{tr}(\rho_\beta) = 1$ and equality (4) defines $\delta'_\gamma := 2c_1 e^{\lambda|X|} \delta_\gamma^{\frac{c_2}{c_2+\beta}}$. Using Eq. (122), we obtain the following

$$
\begin{aligned}
\sum_{s'=0}^{\infty} \left\|P_s^A U_X P_{s'}^A \sqrt{\rho_\beta}\right\|_F &\leq \left\|P_s^A U_X P_0^A \sqrt{\rho_\beta}\right\|_F + \sum_{s'=1}^{\infty} \left\|P_s^A U_X P_{s'}^A\right\| \cdot \left\|P_{s'}^A \sqrt{\rho_\beta}\right\|_F \\
&\leq \delta'^{1/2}_\gamma c_5^{1/2} |X|^{1/2} e^{-(\lambda_1 \gamma s/|X|)^{1/\tau_1}/2} \\
&\quad + \sum_{s'=1}^{\infty} c_5 |X| e^{-(\lambda_1 \gamma |s-s'|/|X|)^{1/\tau_1}} \left\|P_{s'}^A \sqrt{\rho_\beta}\right\|_F,
\end{aligned}
\tag{123}
$$

where the first term in the inequality was obtained from Eq. (122) and the second term was obtained from Eq. (121). We now upper-bound the summation in the second term of Eq. (123) by using the Cauchy–Schwarz inequality as follows:

$$
\begin{aligned}
&\sum_{s'=1}^{\infty} e^{-(\lambda_1 \gamma |s-s'|/|X|)^{1/\tau_1}} \left\|P_{s'}^A \sqrt{\rho_\beta}\right\|_F \\
&= \sum_{s'=1}^{\infty} e^{-(\lambda_1 \gamma |s-s'|/|X|)^{1/\tau_1}/2} \cdot \left(e^{-(\lambda_1 \gamma |s-s'|/|X|)^{1/\tau_1}/2} \left\|P_{s'}^A \sqrt{\rho_\beta}\right\|_F\right) \\
&\leq \left(\sum_{s'=1}^{\infty} e^{-(\lambda_1 \gamma |s-s'|/|X|)^{1/\tau_1}}\right)^{1/2} \left(\sum_{s'=1}^{\infty} e^{-(\lambda_1 \gamma |s-s'|/|X|)^{1/\tau_1}} \left\|P_{s'}^A \sqrt{\rho_\beta}\right\|_F^2\right)^{1/2} \\
&\overset{(1)}{\leq} \xi^{1/2} \left(\sum_{s'=1}^{\infty} p_{s'}^A e^{-(\lambda_1 \gamma |s-s'|/|X|)^{1/\tau_1}}\right)^{1/2},
\end{aligned}
\tag{124}
$$

[7] Since $Q_\gamma^A \leq \mathbb{1}$ and $P_0^A = \mathbb{1} - Q_\gamma^A$, we have

$$
\left\|P_s^A U_X P_0^A \sqrt{\rho_\beta}\right\|_F \leq \left\|Q_\gamma^A U_X (\mathbb{1} - Q_\gamma^A) \sqrt{\rho_\beta}\right\|_F \leq \left\|Q_\gamma^A U_X \sqrt{\rho_\beta}\right\|_F + \left\|Q_\gamma^A U_X Q_\gamma^A \sqrt{\rho_\beta}\right\|_F \leq \left\|Q_\gamma^A U_X \sqrt{\rho_\beta}\right\|_F + \delta_\gamma \leq 2c_1 e^{\lambda|X|} \delta_\gamma^{\frac{c_2}{c_2+\beta}},
$$

where the first inequality used $P_s^A \preceq Q_\gamma^A$ $(s \geq 1)$ and the last inequality used $\|Q_\gamma^A U_X \sqrt{\rho_\beta}\|_F \leq c_1 e^{\lambda|X|} \delta_\gamma^{\frac{c_2}{c_2+\beta}}$ from Claim 33.

with $\xi = 2 + \frac{2\tau_1|X|}{\lambda_1\gamma}(2\tau_1)^{\tau_1}$, where $p_{s'} := \left\|P^A_{s'}\sqrt{\rho_\beta}\right\|_F^2$ and we used Fact 1 in inequality (1). Note that $\sum_{s'=1}^{\infty} p_{s'} = \|Q^A_\gamma\sqrt{\rho_\beta}\|_F^2$ because of Eq. (118). We can obtain the following upper bound by combining the equations Eq. (120), (123) and (124):

$$
\begin{aligned}
&\left\|P^A_s U_X\sqrt{\rho_\beta}\right\|_F^2\\
&\leq \Big(\sum_{s'=0}^{\infty}\left\|P^A_s U_X P^A_{s'}\sqrt{\rho_\beta}\right\|_F\Big)^2\\
&\leq \Big({\delta'}_\gamma^{1/2} c_5^{1/2}|X|^{1/2}e^{-(\lambda_1\gamma s/|X|)^{1/\tau_1}/2} + \sum_{s'=1}^{\infty} c_5|X|e^{-(\lambda_1\gamma|s-s'|/|X|)^{1/\tau_1}}\left\|P^A_{s'}\sqrt{\rho_\beta}\right\|_F\Big)^2\\
&\leq \Big({\delta'}_\gamma^{1/2} c_5^{1/2}|X|^{1/2}e^{-(\lambda_1\gamma s/|X|)^{1/\tau_1}/2} + c_5|X|\xi^{1/2}\cdot\left(\sum_{s'=1}^{\infty} p^A_{s'}e^{-(\lambda_1\gamma|s-s'|/|X|)^{1/\tau_1}}\right)^{1/2}\Big)^2\\
&\leq \underbrace{2c_5{\delta'}_\gamma|X|e^{-(\lambda_1\gamma s/|X|)^{1/\tau_1}}}_{:=f_1(s)} + \underbrace{2c_5^2|X|^2\xi\cdot\left(\sum_{s'=1}^{\infty} p^A_{s'}e^{-(\lambda_1\gamma|s-s'|/|X|)^{1/\tau_1}}\right).}_{:=f_2(s)}
\end{aligned}
$$

Recall that the goal of this claim was to upper bound Eq. (119), which we can rewrite now as

$$
\left\|AQ^A_\gamma U_X\sqrt{\rho_\beta}\right\|_F^2 \leq \gamma^2\sum_{s=1}^{\infty}(s+1)^2\left\|P^A_s U_X\sqrt{\rho_\beta}\right\|_F^2 \leq \gamma^2\sum_{s=1}^{\infty}(s+1)^2 f_1(s) + \gamma^2\sum_{s=1}^{\infty}(s+1)^2 f_2(s). \tag{125}
$$

We bound each of these terms separately. In order to bound the first term, observe that

$$
\begin{aligned}
\gamma^2\sum_{s=1}^{\infty}(s+1)^2 f_1(s) &= 2\gamma^2 c_5\delta'_\gamma|X|\sum_{s=1}^{\infty}(s+1)^2 e^{-(\lambda_1\gamma s/|X|)^{1/\tau_1}}\\
&\overset{(1)}{\leq} 2\gamma^2 c_5{\delta'}_\gamma|X|\cdot 8\tau_1\cdot\left(\frac{(3\tau_1)^{\tau_1}|X|}{\lambda_1\gamma}\right)^3 \leq \frac{16c_5\delta'_\gamma|X|^4\tau_1(3\tau_1)^{3\tau_1}}{\lambda_1^3\gamma},
\end{aligned}
$$

where inequality (1) uses Fact 1. We now bound the second term in Eq. (125) as follows

$$
\begin{aligned}
\gamma^2\sum_{s=1}^{\infty}(s+1)^2 f_2(s) &= 2\gamma^2 c_5^2|X|^2\xi\sum_{s=1}^{\infty}(s+1)^2\left(\sum_{s'=1}^{\infty} p^A_{s'}e^{-(\lambda_1\gamma|s-s'|/|X|)^{1/\tau_1}}\right).\\
&= 2\gamma^2 c_5^2|X|^2\xi\sum_{s'=1}^{\infty} p^A_{s'}\Big(\sum_{s=1}^{\infty}(s+1)^2 e^{-(\lambda_1\gamma|s-s'|/|X|)^{1/\tau_1}}\Big)\\
&\leq 2\gamma^2 c_5^2|X|^2\xi\sum_{s'=1}^{\infty} p^A_{s'}(2s')^2\Big(\sum_{s=1}^{\infty}(1+|s-s'|)^2 e^{-(\lambda_1\gamma|s-s'|/|X|)^{1/\tau_1}}\Big)\\
&\overset{(1)}{\leq} 2\gamma^2 c_5^2|X|^2\xi\cdot\xi'\sum_{s'=1}^{\infty} p^A_{s'}(2s')^2,
\end{aligned}
\tag{126}
$$

with $\xi' = 2 + 16\left(\frac{(3\tau_1)^{\tau_1}|X|}{\lambda_1\gamma}\right)^3$, where inequality (1) follows from Fact 1. Further upper bound this expression by simplifying the pre-factors, we get

$$\begin{aligned}\gamma^2\sum_{s=1}^{\infty}(s+1)^2 f_2(s) &\le 8\gamma^2 c_5^2|X|^2\xi\xi'\sum_{s'=1}^{\infty}p_{s'}^A(s')^2\\ &= 8c_5^2|X|^2\xi\xi'\sum_{s'=1}^{\infty}(\gamma s')^2 p_{s'}^A\\ &= 8c_5^2|X|^2\xi\xi'\sum_{s'=1}^{\infty}\left\|(\gamma s')P_{s'}^A\sqrt{\rho_\beta}\right\|_F^2\\ &\overset{(2)}{\le} 8c_5^2|X|^2\xi\xi'\sum_{s'=0}^{\infty}\left\|P_{s'}^A A\sqrt{\rho_\beta}\right\|_F^2 = 8c_5^2|X|^2\xi\xi'\langle A^2\rangle,\end{aligned} \tag{127}$$

In inequality (2), we used $(\gamma s')P_{s'}^A \preceq AP_{s'}^A$ from the definition (118) of $P_{s'}^A$. Note that $|X|^2\xi\xi'$ is bounded from above by $\mathcal{O}(|X|^6)$). By combining the above inequalities altogether, we prove Eq. (75). □

5.3 Derivation of the sub-exponential concentration

Recall that the goal in this appendix is to prove the following lemma.

Lemma 39 (Restatement of Lemma 9). *Let A be a $(\tau, a_1, a_2, 1)$-quasi-local operator with $\tau < 1$, as given in Eq. (4). For an arbitrary operator O_X supported on a subset $X \subseteq \Lambda$ with $|X| = k_0$ and $\|O_X\| = 1$, we have*

$$\|P_{\ge x+y}^A O_X P_{\le x}^A\| \le c_5 \cdot k_0 \exp\Big(-(\lambda_1 y/k_0)^{1/\tau_1}\Big), \tag{128}$$

where $\tau_1 := \frac{2}{\tau} - 1$ and c_5 and λ_1 are constants which only depend on a_1 and a_2. In particular, the a_2 dependence of c_5 and λ_1 is given by $c_5 \propto a_2^{2/\tau}$ and $\lambda_1 \propto a_2^{-2/\tau}$ respectively.

Before proving this lemma, let us elaborate upon the method. Recall that

$$P_{\le x}^A = \sum_{\omega\le x}\Pi_\omega, \quad P_{>y}^A = \sum_{\omega>y}\Pi_\omega, \tag{129}$$

where Π_ω is the projector onto the eigenvalue ω eigenspace of A. One way to prove the upper bound in the estimation of the norm (128) is to utilize the technique in Ref. [AKL16] (i.e., Lemma 8). The argument proceeds by considering

$$\begin{aligned}\|P_{\ge x+y}^A O_X P_{\le x}^A\| &= \|P_{\ge x+y}^A e^{-\nu A}e^{\nu A}O_X e^{-\nu A}e^{\nu A}P_{\le x}^A\|\\ &\le \|P_{\ge x+y}^A e^{-\nu A}\|\cdot\|e^{\nu A}O_X e^{-\nu A}\|\cdot\|e^{\nu A}P_{\le x}^A\|\\ &\le e^{-\nu y}\|e^{\nu A}O_X e^{-\nu A}\|,\end{aligned} \tag{130}$$

which reduces the problem to estimation of the norm $\|e^{\nu A}O_X e^{-\nu A}\|$. Additionally, by definition of A in Theorem 37 we have

$$A = \sum_{\ell=1}^{n} g_\ell \bar{A}_\ell, \tag{131}$$

where $\bar{A}_\ell$ is κ-local and g_ℓ is sub-exponentially decaying function for ℓ (as made precise in Eq. (4)), namely $g_\ell \propto \exp(-\mathcal{O}(\ell^\tau))$. In this case, for $\nu = \Theta(1)$ in (130), the norm of the imaginary time evolution can be finitely bounded only in the case $D = 1$ [Kuw16]. That is, the norm $\|e^{\nu A} O_X e^{-\nu A}\|$ diverges to infinity for $\tau < 1$. However, our main contribution in this section is that we are able to prove the lemma statement *without* going through the inequalities in (130) (which in turn used earlier results of [Kuw16, AKL16]). We now give more details.

5.3.1 Proof of Lemma 39

In order to estimate the norm, we need to take a different route from (130). Recall that A is a $(\tau, a_1, a_2, 1)$-quasi local operator. Let I be any interval of the real line and P_I^A be the projector onto the eigenspace of A with eigenvalues in I. Using the operator inequality

$$P^A_{\geq z}(A - \omega \mathbb{1})^m \succeq (z-\omega)^m P^A_{\geq z},$$

we obtain

$$\|(A-\omega\mathbb{1})^m O_X P_I^A\| \geq \|P^A_{\geq z}(A-\omega)^m O_X P_I^A\| \geq (z-\omega)^m \|P^A_{\geq z} O_X P_I^A\|, \tag{132}$$

hence

$$\|P^A_{\geq z} O_X P_I^A\| \leq \frac{\|(A-\omega)^m O_X P_I^A\|}{(z-\omega)^m}. \tag{133}$$

Our strategy to establish Eq. (128) will be to expand

$$\|P^A_{\geq x+y} O_X P^A_{\leq x}\| \leq \sum_{j=0}^{\infty} \|P^A_{\geq x+y} O_X P^A_{I_j}\|, \tag{134}$$

for carefully chosen intervals $I_j := (x - a_1(j+1), x - a_1 j]$ (the term a_1 is as given in the statement of Lemma 39). Towards this, let us fix an arbitrary ω and an interval $I := (\omega - a_1, \omega]$, and prove an upper bound on $\|P^A_{\geq \omega+\theta} O_X P_I^A\|$ (for all θ). We show the following claim.

Claim 40. *Let A be a $(\tau, a_1, a_2, 1)$-quasi local operator as defined in (4). There is a constant c_6 such that*

$$\|P^A_{\geq \omega+\theta} O_X P_I^A\| \leq \frac{1}{\tau} \exp\left[-[\theta/(ec_6 k_0)]^{1/\tau_1} + 1\right]. \tag{135}$$

The claim is proved in subsection 5.3.2. Let us use the claim to establish the lemma. In the inequality (134), we need to estimate $\|P^A_{\geq x+y} O_X P^A_{I_j}\|$ with $I_j := (x-(j+1)a_1, x - ja_1]$. Setting $\omega = x - ja_1$ and $\theta = y + ja_1$ in Claim 40, we have

$$\|P^A_{\geq x+y} O_X P^A_{I_j}\| \leq \frac{1}{\tau} \exp\left\{-\left(\frac{y + a_1 j}{ec_6 k_0}\right)^{1/\tau_1} + 1\right\}. \tag{136}$$

In order to complete the bound on Equation (134), we need to take summation with respect to j. We have

$$\sum_{j=0}^{\infty} \|P^A_{\geq x+y} O_X P^A_{I_j}\| \leq \sum_{j=0}^{\infty} \frac{1}{\tau} \exp\left\{-\left(\frac{y + a_1 j}{ec_6 k_0}\right)^{1/\tau_1} + 1\right\} \leq \frac{1}{\tau} e^{-\frac{1}{2}\left(\frac{y}{ec_6 k_0}\right)^{1/\tau_1}} \left(1 + \frac{ec_6 k_0 \tau_1}{a_1} (2\tau_1)^{1/\tau_1}\right), \tag{137}$$

where in last inequality we used Fact 1 (3) with $c = (ec_6 k_0/a_1)^{-1/\tau_1}$, $p = 1/\tau_1$ and $a = y/a_1$. This gives the form of (128) and completes the proof.

5.3.2 Proof of Claim 40

From Equation (133), it suffices to upper bound $\|(A-\omega)^m O_X P_I^A\|$. Abbreviate $\tilde{A} := A - \omega \mathbb{1}$. Introduce the multi-commutator

$$\mathrm{ad}_{\tilde{A}}^s(O_X) := \underbrace{[\tilde{A}, \dots [\tilde{A}, [\tilde{A}, O_X]] \dots]}_{s \text{ times}}.$$

Consider the following identity,

$$\tilde{A}^m O_X P_I^A = \sum_{s=0}^{m} \binom{m}{s} \mathrm{ad}_{\tilde{A}}^s(O_X) \tilde{A}^{m-s} P_I^A. \tag{138}$$

This shows that

$$\|\tilde{A}^m O_X P_I^A\| \le \sum_{s=0}^{m} \binom{m}{s} \|\mathrm{ad}_{\tilde{A}}^s(O_X)\| \cdot \|\tilde{A}^{m-s} P_I^A\| \le \sum_{s=0}^{m} \binom{m}{s} a_1^{m-s} \|\mathrm{ad}_{\tilde{A}}^s(O_X)\|, \tag{139}$$

where we use $\|\tilde{A}^{m-s} P_I^A\| = \|(A-\omega)^{m-s} P_I^A\| \le a_1^{m-s}$ for $I = (\omega - a_1, \omega]$. The remaining task is to estimate the upper bound of $\|\mathrm{ad}_{\tilde{A}}^s(O_X)\| = \|\mathrm{ad}_{A}^s(O_X)\|$. This is done in the following claim.

Claim 41. *Let A be a $(\tau, a_1, a_2, 1)$-quasi local operator as defined in (4). Then, for an arbitrary operator O_X which is supported on a subset X ($|X| = k_0$), the norm of the multi-commutator $\mathrm{ad}_A^s(O_X)$ is bounded from above by*

$$\|\mathrm{ad}_A^s(O_X)\| \le \frac{(2a_1)^s (2k_0)^s e^s}{\tau} \cdot \left(\frac{2}{a_2 \tau}\right)^{\frac{2s}{\tau}} \cdot (s^{\tau_1})^s \quad \text{for} \quad s \le m, \tag{140}$$

where the constants a_1 and a_2 have been defined in Eq. (4).

By applying the inequality (140) to (139), we obtain

$$\begin{aligned}
\|\tilde{A}^m O_X P_I^A\| &\le \sum_{s=0}^{m} \binom{m}{s} a_1^{m-s} \frac{(2a_1)^s (2k_0)^s e^s}{\tau} \cdot \left(\frac{2}{a_2 \tau}\right)^{\frac{2s}{\tau}} \cdot (s^{\tau_1})^s \\
&\le \sum_{s=0}^{m} \binom{m}{s} (2a_1)^m \frac{(2ek_0)^m}{\tau} \cdot \left(\frac{2}{a_2 \tau}\right)^{\frac{2m}{\tau}} \cdot (m^{\tau_1})^m \\
&= (4a_1)^m \frac{(2ek_0)^m}{\tau} \cdot \left(\frac{2}{a_2 \tau}\right)^{\frac{2m}{\tau}} \cdot (m^{\tau_1})^m = \frac{1}{\tau} \left[8ea_1 k_0 [2/(a_2\tau)]^{2/\tau} m^{\tau_1}\right]^m.
\end{aligned}$$

Therefore, setting $z = \omega + \theta$ in the inequality (133), we obtain

$$\|P_{\ge \omega+\theta}^A O_X P_I^A\| \le \frac{\|\tilde{A}^m O_X P_I^A\|}{\theta^m} \le \frac{1}{\tau} \left[8ea_1 k_0 [2/(a_2\tau)]^{2/\tau} \frac{m^{\tau_1}}{\theta}\right]^m \tag{141}$$

$$\le \frac{1}{\tau} \left(\frac{c_6 k_0 m^{\tau_1}}{\theta}\right)^m, \tag{142}$$

where $c_6 := 8ea_1[2/(a_2\tau)]^{2/\tau}$. Let us choose $m = \tilde{m}$ with $\tilde{m}$ the minimum integer such that

$$\frac{c_6 k_0 \tilde{m}^{\tau_1}}{\theta} \le 1/e. \tag{143}$$

The above condition is satisfied by $\tilde{m}^{\tau_1} \le \theta/(ec_6k_0)$, which implies

$$\tilde{m} = \left\lfloor [\theta/(ec_6k_0)]^{1/\tau_1} \right\rfloor, \tag{144}$$

where $\lfloor \cdot \rfloor$ is the floor function. From this choice, the claim concludes.

5.3.3 Proof of Claim 41

As stated in Claim 41, we let A be a $(\tau, a_1, a_2, 1)$-quasi local operator as defined in (4). Recall that we need to show, for an arbitrary operator O_X which is supported on k_0 sites, the norm of the multi-commutator $\mathrm{ad}_A^s(O_X)$ is bounded by

$$\|\,\mathrm{ad}_A^s(O_X)\| \le \frac{(2a_1)^s(2k_0)^s e^s}{\tau} \cdot \left(\frac{2}{a_2\tau}\right)^{\frac{2s}{\tau}} \cdot (s^{\tau_1})^s \quad \text{for} \quad s \le m.$$

We start from the following expansion:

$$\mathrm{ad}_A^s(O_X) = \sum_{k_1,k_2,\dots,k_s} g_{k_1}g_{k_2}\cdots g_{k_s}[\bar{A}_{k_s}, [\bar{A}_{k_{s-1}}, \cdots [\bar{A}_{k_1}, O_X]\cdots].$$

By using Lemma 3 in Ref. [KMS16] and setting $\zeta = 1$ (see Definition 3) we obtain

$$\|[[\bar{A}_{k_s}, [\bar{A}_{k_{s-1}}, \cdots [\bar{A}_{k_1}, O_X]\cdots]\| \le 2^s k_0(k_0+k_1)(k_0+k_1+k_2)\cdots(k_0+k_1+k_2+\cdots+k_{s-1}), \tag{145}$$

where we used that A is a $(\tau, a_1, a_2, 1)$-quasi local operator. Recall that we set $\|O_X\| = 1$ and $|X| = k_0$. The norm of $\mathrm{ad}_A^s(O_X)$ is bounded from above by

$$\begin{aligned}
&\|\,\mathrm{ad}_A^s(O_X)\| \\
&\le \sum_{k_1,k_2,\dots,k_s=1}^{\infty} 2^s g_{k_1}g_{k_2}\cdots g_{k_s} k_0(k_0+k_1)(k_0+k_1+k_2)\cdots(k_0+k_1+k_2+\cdots+k_{s-1}) \\
&= \sum_{K\ge s} \sum_{\substack{k_1+k_2+\dots+k_s=K \\ k_1\ge 1, k_2\ge 1,\dots,k_s\ge 1}} 2^s g_{k_1}g_{k_2}\cdots g_{k_s} k_0(k_0+k_1)(k_0+k_1+k_2)\cdots(k_0+k_1+k_2+\cdots+k_{s-1}),
\end{aligned} \tag{146}$$

where the summation over K starts from s because each of $\{k_j\}_{j=1}^s$ is larger than 1. Now, using the inequality $\log[g_k/a_1] \le -a_2k^\tau$ with $\tau \le 1$, we have $\sum_{j=1}^s \log(g_{k_j}/a_1) \le \log(g_{k_1+k_2+\cdots+k_s}/a_1)$. This follows from $\sum_{j=1}^s k_j^\tau \ge (k_1+k_2+\cdots+k_s)^\tau$. Thus, using $k_1+k_2+\cdots+k_s = K$, the summand in the inequality (146) is upper-bounded by

$$g_{k_1}g_{k_2}\cdots g_{k_s}k_0(k_0+k_1)(k_0+k_1+k_2)\cdots(k_0+k_1+k_2+\cdots+k_{s-1}) \le a_1^s(g_K/a_1)k_0(k_0+K)^{s-1}, \tag{147}$$

where we use the inequality $k_1 + k_2 + \cdots + k_j \le K$ for $j = 1, 2, \ldots, s-1$. By combining the two inequalities (146) and (147), we obtain

$$
\begin{aligned}
\|\operatorname{ad}_A^s(O_X)\| &\le \sum_{K\ge s} \sum_{\substack{k_1+k_2+\ldots+k_s=K \\ k_1\ge 1, k_2\ge 1,\ldots,k_s\ge 1}} (2a_1)^s (g_K/a_1) k_0 (k_0+K)^{s-1} \\
&\overset{(1)}{\le} \sum_{K\ge s} \left(\!\!\binom{s}{K-s}\!\!\right) (2a_1)^s (g_K/a_1) k_0 (k_0+K)^{s-1} \\
&= \sum_{K\ge s} \binom{K-1}{s-1} (2a_1)^s (g_K/a_1) k_0 (k_0+K)^{s-1} \\
&\overset{(2)}{\le} (2a_1)^s \frac{(2k_0)^s}{2} \sum_{K\ge s} \frac{e^s K^s}{s^s} (g_K/a_1)(K)^{s-1} \\
&\overset{(3)}{\le} \frac{(2a_1)^s (2k_0)^s e^s}{s^s} \cdot \frac{1}{2} \sum_{K\ge s} K^{2s-1} e^{-a_2 K^\tau} \le \frac{(2a_1)^s (2k_0)^s e^s}{s^s} \cdot \frac{1}{2} \sum_{K\ge 0} K^{2s-1} e^{-a_2 K^\tau} \\
&\overset{(4)}{\le} \frac{(2a_1)^s (2k_0)^s e^s}{s^s \tau} \cdot \left(\frac{2s}{a_2 \tau}\right)^{\frac{2s}{\tau}} = \frac{(2a_1)^s (2k_0)^s e^s}{\tau} \cdot \left(\frac{2}{a_2\tau}\right)^{\frac{2s}{\tau}} \cdot \left(s^{\frac{2}{\tau}-1}\right)^s,
\end{aligned}
$$

where in (1), $\left(\!\binom{}{}\!\right)$ denotes the multi-combination, namely $\left(\!\binom{n}{m}\!\right) = \binom{n+m-1}{n-1}$, in (2) we upper bound $\binom{K-1}{s-1} \le \frac{e^s K^s}{s^s}$, $k_0 + K \le 2k_0 K$, in (3) we use the sub-exponential form of g_K in Eq. (4) and in (4) we use Fact 1. Since $\tau_1 = \frac{2}{\tau} - 1$, this proves the statement.

5.4 Quasi-locality of $\widetilde{W}$

We here aim to obtain (τ, a_1, a_2, ζ)-quasi-locality of the operator $\widetilde{W}$, where $\{\tau, a_1, a_2, \zeta\}$ defined in Definition 3. This is used in Section 4.8. In particular, we will show that

$$
(\tau, a_1, a_2, \zeta) = \left(1/D, \mathcal{O}(1), \mathcal{O}(1/\beta), \mathcal{O}(\beta^{2D+1})\left(\max_{j\in\Lambda} v_j\right)\right)
$$

suffices to prove the quasi-locality of $\widetilde{W}$. Recall the definition of $\widetilde{W}$:

$$
\widetilde{W} = \int_{-\infty}^{\infty} f_\beta(t)\, e^{-iHt}\, W\, e^{iHt} dt,
$$

where

$$
f_\beta(t) = \frac{2}{\beta\pi} \log \frac{e^{\pi|t|/\beta}+1}{e^{\pi|t|/\beta}-1}
$$

and

$$
W = \sum_{i\in\Lambda} v_i E_i.
$$

We write

$$
\widetilde{W} = \sum_i v_i \int_{-\infty}^{\infty} f_\beta(t)\, e^{-iHt}\, E_i\, e^{iHt} dt.
$$

Abbreviate

$$\tilde{E}_i(t) := e^{-iHt}\, E_i\, e^{iHt}$$

and recall that $\tilde{E}_i = \int_{\infty}^{\infty} f_\beta(t)\tilde{E}_i(t)$. Moreover, (with some abuse of notation) let $B(r,i) \subseteq \Lambda$ denote the ball of radius r such that: the centre of $B(r,i)$ coincides with the the center of the smallest ball containing E_i. We assume that r ranges in the set $\{m_i, m_i+1, \ldots, n_i\}$, where m_i is the radius of the smallest ball containing E_i and n_i is the number such that $B(n_i, i) = \Lambda$. Define

$$\tilde{E}_i^r(t) := \mathrm{tr}_{B(r,i)^c}[\tilde{E}_i(t)] \otimes \frac{\mathbb{1}_{B(r,i)^c}}{\mathrm{tr}[\mathbb{1}_{B(r,i)^c}]}, \quad \tilde{E}_i^0(t) = 0,$$

i.e., $\tilde{E}_i^r(t)$ traces out all the qudits in $\tilde{E}_i(t)$ that are at outside the $B(r,i)$-ball around $\tilde{E}_i^r$. From [BHV06], we have

$$\|\tilde{E}_i(t) - \tilde{E}_i^r(t)\| \le \|E_i\| \min\left\{1, c_3 r^{D-1} e^{-c_4(r-m_i-v_{\mathrm{LR}}|t|)}\right\}$$

which in particular implies

$$\|\tilde{E}_i^r(t) - \tilde{E}_i^{r-1}(t)\| \le 2\min\left\{1, c_3 e^{c_4 m_i} r^{D-1} e^{-c_4(r-v_{\mathrm{LR}}|t|)}\right\},$$

where we use $\|E_i\| = 1$, v_{LR} is the Lieb-Robinson velocity (as defined in Fact 4) and c_3, c_4 are constants. We note that the $2\min\{1,\cdot\}$ is derived from the trivial upper bound $\|\tilde{E}_i^r(t) - \tilde{E}_i^{r+1}(t)\| \le 2$. This allows us to write the following quasi-local expression:

$$\tilde{E}_i(t) = \sum_{r=m_i}^{n_i} \left(\tilde{E}_i^r(t) - \tilde{E}_i^{r-1}(t)\right).$$

Using this, we can now write the quasi-local representation of $\tilde{E}_i$ as follows.

$$\int_{-\infty}^{\infty} f_\beta(t)\tilde{E}_i(t)dt = \int_{-\infty}^{\infty} f_\beta(t) \sum_{r=m_i}^{n_i} \left(\tilde{E}_i^r(t) - \tilde{E}_i^{r-1}(t)\right).$$

To see that it is quasi-local, observe that the term with radius r has norm

$$\begin{aligned}
&\int_{-\infty}^{\infty} f_\beta(t)\left\|\tilde{E}_i^r(t) - \tilde{E}_i^{r-1}(t)\right\| \\
&\le 2c_3 e^{c_4 m_i} r^{D-1} \cdot \left(e^{-c_4 r}\int_{|t|\le r/(2v_{\mathrm{LR}})} e^{c_4 v_{\mathrm{LR}}|t|} f_\beta(t)dt + \int_{|t|>r/(2v_{\mathrm{LR}})} f_\beta(t)dt\right) \\
&\le 2c_3 e^{c_4 m_i} r^{D-1} \cdot \left(e^{-c_4 r/2}\int_{-\infty}^{\infty} f_\beta(t)dt + \int_{|t|>r/(2v_{\mathrm{LR}})} \frac{2\beta}{\pi|t|} e^{-\pi|t|/\beta}dt\right) \\
&\le 2c_3 e^{c_4 m_i} r^{D-1} \left(e^{-c_4 r/2} + \frac{4\beta v_{\mathrm{LR}}}{\pi r}\frac{e^{-\pi r/(2\beta v_{\mathrm{LR}})}}{\pi/\beta}\right) \\
&= 2c_3 e^{c_4 m_i} r^{D-1} \left(e^{-c_4 r/2} + \frac{4\beta^2 v_{\mathrm{LR}}}{\pi^2} r^{-1} e^{-\pi r/(2\beta v_{\mathrm{LR}})}\right) \le c_5 r^{D-1} e^{-c_4' r},
\end{aligned}$$

with $c_4' = \min(c_4/2, \pi/(2\beta v_{\rm LR}))$, where $c_5 = \mathcal{O}(\beta^2)$ is a constant which does not depend on r, and we use $\int_{-\infty}^{\infty} f_\beta(t) = \tilde{f}_\beta(0) = 1$ and $f_\beta(t) \le 2\beta/(\pi|t|)e^{-\pi|t|/\beta}$. Note that $c_4' = \mathcal{O}(1/\beta)$. Define

$$a_{B(r,i)} := e^{c_4' r/2} \int_{-\infty}^{\infty} f_\beta(t) \left(\tilde{E}_i^r(t) - \tilde{E}_i^{r-1}(t)\right).$$

Here, the operator $a_{B(r,i)}$ is supported on the subset $B(r,i)$. Then, from $|B(r,i)| = \mathcal{O}(r^D)$, the quasi-local representation of $\widetilde{W}$ is given as

$$\widetilde{W} = \sum_{i\in\Lambda} v_i \sum_{r=m_i}^{n_i} e^{-c_4' r/2} a_{B(r,i)} = \sum_{i\in\Lambda}\sum_{r=m_i}^{n_i} e^{-\mathcal{O}(|B(r,i)|^{\frac{1}{D}})} v_i a_{B(r,i)},$$

with $e^{-\mathcal{O}(|B(r,i)|^{\frac{1}{D}})}$ decaying sub-exponentially with rate $\tau = 1/D$, for all $i \in \Lambda$. We also obtain the parameter ζ in Eq. (4), which can be calculated for a fixed r. For an arbitrary r, we have

$$\begin{aligned}\sum_{r,j:B(r,j)\ni i} v_j \|a_{B(r,j)}\| \le c_5 r^{D-1} \sum_{j:B(r,j)\ni i} v_j e^{-c_4' r/2} &\le \left(\max_{j\in\Lambda} v_j\right) c_5 c_B r^{2D-1} e^{-c_4' r/2} \\ &\le \Theta(\beta^{2D+1}) \left(\max_{j\in\Lambda} v_j\right),\end{aligned}$$

where we define c_B such that $|B(r,j)| \le c_B r^D$ and we used $c_5 = \mathcal{O}(\beta^2)$. This completes the representation and shows that $\widetilde{W}$ is a $\left(1/D, \mathcal{O}(1), \mathcal{O}(1/\beta), \mathcal{O}(\beta^{2D+1})\left(\max_{j\in\Lambda} v_j\right)\right)$-quasi-local.

5.5 Proof of Lemma 38

Recall that the goal in this section is to prove that for $\widetilde{W}$ defined in Lemma 27 we have

$$\max_{i\in\Lambda} \mathrm{tr}[(\widetilde{W}_{(i)})^2\eta] = \frac{\Theta(1)}{(\beta\log(\beta)+1)^{2D+2}} \left(\max_{i\in\Lambda} v_i^2\right),$$

where η is the maximally mixed state. In this direction, we will now prove that

$$\max_{i\in\Lambda} \|\widetilde{W}_{(i)}\sqrt{\eta}\|_F \ge \frac{c_7}{(\beta\log(\beta)+1)^{D+1}} \max_{i\in\Lambda}(|v_i|), \tag{148}$$

for a constant $c_7 = \mathcal{O}(1)$. For convenience, let us define $\mathrm{argmax}_{i\in\Lambda}|v_i| = i_+$, or equivalently $|v_{i_+}| = \max_{i\in\Lambda}|v_i|$. In the following, we prove the inequality (148) for $\|\widetilde{W}_{(i+)}\sqrt{\eta}\|_F$ instead of $\max_{i\in\Lambda}\|\widetilde{W}_{(i)}\sqrt{\eta}\|_F$. By using the inequality $\max_{i\in\Lambda}\|\widetilde{W}_{(i)}\|_F \ge \|\widetilde{W}_{(i_+)}\|_F$, we obtain the main statement. We denote the ball region $B(r,i_+)$ by B_r for the simplicity, where r is fixed later. Let us consider $\widetilde{W}[B_r]$ which is defined as follows:

$$\widetilde{W}[B_r] := \int_{-\infty}^{\infty} f_\beta(t) e^{-iHt} W[B_r] e^{iHt} dt, \quad W[B_r] := \sum_{i\in B_r} v_i E_i. \tag{149}$$

Since $\widetilde{W}[B_r]$ is obtained from $W[B_r]$ in an equivalent manner as $\widetilde{W}$ is obtained from W, the following claim follows along the same lines as Theorem 29. We skip the very similar proof.

Claim 42. *It holds that*

$$\|\widetilde{W}[B_r]\|_F^2 \geq \frac{\mathcal{D}_\Lambda}{c_5[\beta\log(r)+1]^2}\sum_{i\in B_r} v_i^2,$$

where c_5 is a constant of $\mathcal{O}(1)$.

Since the new operator $\widetilde{W}[B_r]$ well approximates the property of $\widetilde{W}$ around the site i_+, as long as r is sufficiently large, we expect that $\widetilde{W}[B_r]_{(i_+)}$ and $\widetilde{W}_{(i_+)}$ are close to each other. The claim below makes this intuition rigorous:

Claim 43. *It holds that*

$$\|\widetilde{W}_{(i_+)} - \widetilde{W}[B_r]_{(i_+)}\| \leq c_1|v_{i_+}|\beta^D e^{-c_2 r/\beta}, \tag{150}$$

where c_1, c_2 are constants of $\mathcal{O}(1)$.

This claim implies that the contribution of all the terms in $\widetilde{W}_{(i_+)}$ which are not included in the B_r ball around i_+ decays exponentially with r. Hence,

$$\begin{aligned}\|\widetilde{W}_{(i_+)}\|_F = \|\widetilde{W}_{(i_+)} - \widetilde{W}[B_r]_{(i_+)} + \widetilde{W}[B_r]_{(i_+)}\|_F &\geq \|\widetilde{W}[B_r]_{(i_+)}\|_F - \|\widetilde{W}_{(i_+)} - \widetilde{W}[B_r]_{(i_+)}\|_F \\ &\geq \|\widetilde{W}[B_r]_{(i_+)}\|_F - \sqrt{\mathcal{D}_\Lambda}\|\widetilde{W}_{(i_+)} - \widetilde{W}[B_r]_{(i_+)}\| \\ &\geq \|\widetilde{W}[B_r]_{(i_+)}\|_F - \sqrt{\mathcal{D}_\Lambda}c_1|v_{i_+}|\beta^D e^{-c_2 r/\beta},\end{aligned} \tag{151}$$

where we use $\|\widetilde{W}_{(i_+)} - \widetilde{W}[B_r]_{(i_+)}\|_F \leq \sqrt{\mathcal{D}_\Lambda}\|\widetilde{W}_{(i_+)} - \widetilde{W}[B_r]_{(i_+)}\|$ in the second inequality. Second, we consider the approximation of $\widetilde{W}[B_r]$ by $\widetilde{W}[B_r, B_{r'}]$ which are supported on $B_{r'}$:

$$\widetilde{W}[B_r, B_{r'}] := \mathrm{tr}_{B_{r'}^{\mathrm{c}}}(\widetilde{W}[B_r]) \otimes \frac{\mathbb{1}_{B_{r'}^{\mathrm{c}}}}{d^{|B_{r'}^{\mathrm{c}}|}}. \tag{152}$$

Because of the quasi-locality of $\widetilde{W}$, we expect $\widetilde{W}[B_r, B_{r'}] \approx \widetilde{W}[B_r]$ for $r' \gg r$. This is shown in the following lemma:

Claim 44. *The norm difference between $\widetilde{W}[B_r, B_{r'}]$ and $\widetilde{W}[B_r]$ is upper-bounded as*

$$\|\widetilde{W}[B_r] - \widetilde{W}[B_r, B_{r'}]\| \leq c_3|v_{i_+}|r^D\beta e^{-c_4|r'-r|/\beta} \tag{153}$$

and

$$\|\widetilde{W}[B_r]_{(i_+)} - \widetilde{W}[B_r, B_{r'}]_{(i_+)}\| \leq 2c_3|v_{i_+}|r^D\beta e^{-c_4|r'-r|/\beta}, \tag{154}$$

where c_3, c_4 are constants of $\mathcal{O}(1)$.

The claim reduces the inequality (151) to

$$\begin{aligned}\|\widetilde{W}_{(i_+)}\|_F &\geq \|\widetilde{W}[B_r]_{(i_+)}\|_F - \sqrt{\mathcal{D}_\Lambda}c_1|v_{i_+}|\beta^D e^{-c_2 r/\beta} \\ &\geq \|\widetilde{W}[B_r, B_{r'}]_{(i_+)}\|_F - \sqrt{\mathcal{D}_\Lambda}\|\widetilde{W}[B_r]_{(i_+)} - \widetilde{W}[B_r, B_{r'}]_{(i_+)}\| - \sqrt{\mathcal{D}_\Lambda}c_1|v_{i_+}|\beta^D e^{-c_2 r/\beta} \\ &\geq \|\widetilde{W}[B_r, B_{r'}]_{(i_+)}\|_F - \sqrt{\mathcal{D}_\Lambda}c_1|v_{i_+}|\beta^D e^{-c_2 r/\beta} - 2\sqrt{\mathcal{D}_\Lambda}c_3|v_{i_+}|r^D\beta e^{-c_4(r'-r)/\beta}.\end{aligned} \tag{155}$$

Next, we relate the norm of $\widetilde{W}[B_r, B_{r'}]_{(i_+)}$ to that of $\widetilde{W}[B_r, B_{r'}]$ using Claim 10. By recalling that $\widetilde{W}[B_r, B_{r'}]_{(i_+)}$ is supported on $B_{r'}$, this gives

$$\|\widetilde{W}[B_r, B_{r'}]_{(i_+)}\|_F \geq \frac{1}{|B_{r'}|}\|\widetilde{W}[B_r, B_{r'}]\|_F, \tag{156}$$

which reduces the inequality (155) to

$$\begin{aligned}\|\widetilde{W}_{(i_+)}\|_F \geq & \frac{1}{|B_{r'}|}\|\widetilde{W}[B_r, B_{r'}]\|_F - \sqrt{\mathcal{D}_\Lambda}c_1|v_{i_+}|\beta^D e^{-c_2 r/\beta} - 2\sqrt{\mathcal{D}_\Lambda}c_3|v_{i_+}|r^D\beta e^{-c_4(r'-r)/\beta} \\ \geq & \frac{1}{|B_{r'}|}\|\widetilde{W}[B_r]\|_F - \sqrt{\mathcal{D}_\Lambda}c_1|v_{i_+}|\beta^D e^{-c_2 r/\beta} - (2+1/|B_{r'}|)\sqrt{\mathcal{D}_\Lambda}c_3|v_{i_+}|r^D\beta e^{-c_4(r'-r)/\beta},\end{aligned} \tag{157}$$

where in the second inequality we apply Claim 44 to $\|\widetilde{W}[B_r, B_{r'}]\|_F$. Finally, we use the lower bound given in Claim 42 and the inequality $\sum_{i\in B_r} v_i^2 \geq v_{i_+}^2$ (since $i_+ \in B_r$) to obtain

$$\|\widetilde{W}[B_r]\|_F^2 \geq \frac{\mathcal{D}_\Lambda}{c_5[\beta\log(r)+1]^2}v_{i_+}^2.$$

This reduces the inequality (157) to the following:

$$\frac{\|\widetilde{W}_{(i_+)}\|_F}{\sqrt{\mathcal{D}_\Lambda}} \geq \frac{|v_{i_+}|}{c_8\,(r')^D\,\sqrt{c_5}[\beta\log(r)+1]} - c_1|v_{i_+}|\beta^D e^{-c_2 r/\beta} - 3c_3|v_{i_+}|r^D\beta e^{-c_4(r'-r)/\beta},$$

where we used $|B_{r'}| \leq c_8\,(r')^D$, for some constant c_8. By choosing $r' = 2r$ and $r = \Theta(1)\cdot D\beta\log(\beta)+1$, we have

$$\frac{\|\widetilde{W}_{(i_+)}\|_F}{\sqrt{\mathcal{D}_\Lambda}} = \|\widetilde{W}_{(i_+)}\sqrt{\eta}\|_F \geq \frac{c_7|v_{i_+}|}{(\beta\log(\beta)+1)^{D+1}}, \tag{158}$$

for some constant c_7. This completes the proof. □

5.5.1 Proof of Claims 43, 44

Proof of Claim 43. Recall that the goal is to prove

$$\|\widetilde{W}_{(i_+)} - \widetilde{W}[B_r]_{(i_+)}\| \leq c_1|v_{i_+}|\beta^D e^{-c_2 r/\beta}$$

for constants c_1, c_2. We start from the integral representation of $\widetilde{W}_{(i_+)}$:

$$\widetilde{W}_{(i_+)} = \widetilde{W} - \int d\mu(U_{i_+})U_{i_+}^\dagger \widetilde{W} U_{i_+}, \tag{159}$$

where $\mu(U_{i_+})$ is the Haar measure for unitary operator U_{i_+} which acts on the i_+th site. This yields

$$\widetilde{W}_{(i_+)} - \widetilde{W}[B_r]_{(i_+)} = \widetilde{W}[B_r^c] - \int d\mu(U_{i_+})U_{i_+}^\dagger \widetilde{W}[B_r^c]U_{i_+}. \tag{160}$$

We thus obtain

$$\begin{aligned}\|\widetilde{W}_{(i_+)} - \widetilde{W}[B_r]_{(i_+)}\| &\leq \sup_{U_{i_+}} \|[U_{i_+}, \widetilde{W}[B_r^{\mathrm{c}}]]\| \\ &\leq \int_{-\infty}^{\infty} f_\beta(t) \sum_{j\in B_r^{\mathrm{c}}} |v_j| \sup_{U_{(i)}} \|[U_{i_+}, e^{-iHt}E_j e^{iHt}]\| dt \\ &\leq f \cdot |v_{i_+}| \sum_{j\in B_r^{\mathrm{c}}} \int_{-\infty}^{\infty} f_\beta(t) \min(e^{-c(\mathrm{dist}(i_+,j)-m_j-v_{\mathrm{LR}}t)}, 1) dt, \end{aligned} \tag{161}$$

where we use $|v_j| \leq |v_{i_+}|$ and the Lieb-Robinson bound (Fact 4) for the last inequality, and m_j is the radius of the support for E_j (see also Sec. 5.4). Because the function $f_\beta(t)$ decays as $e^{-\mathcal{O}(t/\beta)}$ and $\mathrm{dist}(i_+, j) \geq r$ for $j \in B_r^{\mathrm{c}}$, we have

$$|v_{i_+}| \sum_{j\in B_r^{\mathrm{c}}} f e^{cm_j} \int_{-\infty}^{\infty} f_\beta(t) \min(e^{-c(\mathrm{dist}(i_+,j)-v_{\mathrm{LR}}t)}, 1) dt \leq c_1 |v_{i_+}| \beta^D e^{-c_2 r/\beta}. \tag{162}$$

This completes the proof. □

Proof of Claim 44. Recall that we wanted to show

$$\|\widetilde{W}[B_r] - \widetilde{W}[B_r, B_{r'}]\| \leq c_3 |v_{i_+}| r^D \beta e^{-c_4|r'-r|/\beta}.$$

In order to prove this, we also utilize the integral representation of $\widetilde{W}[B_r, B_{r'}]$:

$$\widetilde{W}[B_r, B_{r'}] := \int d\mu(U_{B_{r'}^{\mathrm{c}}}) U_{B_{r'}^{\mathrm{c}}}^\dagger \widetilde{W}[B_r] U_{B_{r'}^{\mathrm{c}}}, \tag{163}$$

which yields an upper bound of $\|\widetilde{W}[B_r] - \widetilde{W}[B_r, B_{r'}]\|$ as

$$\|\widetilde{W}[B_r] - \widetilde{W}[B_r, B_{r'}]\| \leq \int d\mu(U_{B_{r'}^{\mathrm{c}}}) \|[\widetilde{W}[B_r], U_{B_{r'}^{\mathrm{c}}}]\|. \tag{164}$$

From the definition (149) of $\widetilde{W}[B_r]$ and the Lieb-Robinson bound (Fact 4), we obtain

$$\begin{aligned}\int d\mu(U_{B_{r'}^{\mathrm{c}}}) \|[\widetilde{W}[B_r], U_{B_{r'}^{\mathrm{c}}}]\| &\leq \int d\mu(U_{B_{r'}^{\mathrm{c}}}) \int_{-\infty}^{\infty} f_\beta(t) \sum_{j\in B_r} |v_j| \cdot \|[e^{-iHt}E_j e^{iHt}, U_{B_{r'}^{\mathrm{c}}}]\| \\ &\leq f \max_{j\in B_r}(\mathrm{Supp}(E_j)) |v_{i_+}| \int_{-\infty}^{\infty} f_\beta(t) \sum_{j\in B_r} \min(e^{-c(r'-r-m_j-v_{\mathrm{LR}}t)}, 1) dt \\ &\leq c_3' |v_{i_+}| \cdot |B_r| \cdot \beta e^{-c_4|r'-r|/\beta}, \end{aligned} \tag{165}$$

with c_3' a constant of $\mathcal{O}(1)$, where $\mathrm{Supp}(E_j) \propto m_j^D = \mathcal{O}(1)$ is the support of E_j. Since $|B_r| \propto r^D$, we obtain the main inequality (153). Now, since

$$\widetilde{W}[B_r]_{(i_+)} = \widetilde{W}[B_r] - \int d\mu(U_{i_+}) U_{i_+}^\dagger \widetilde{W}[B_r] U_{i_+}$$

and

$$\widetilde{W}[B_r, B_{r'}]_{(i_+)} = \widetilde{W}[B_r, B_{r'}] - \int d\mu(U_{i_+}) U_{i_+}^\dagger \widetilde{W}[B_r, B_{r'}] U_{i_+},$$

we obtain the second inequality (154) due to

$$\|\widetilde{W}[B_r]_{(i_+)} - \widetilde{W}[B_r, B_{r'}]_{(i_+)}\| \leq 2\|\widetilde{W}[B_r] - \widetilde{W}[B_r, B_{r'}]\|.$$

This completes the proof. □

References

[Aar18a] Scott Aaronson. Shadow tomography of quantum states. In *STOC'18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 325–338. ACM, New York, 2018.

[AAR+18b] Mohammad H Amin, Evgeny Andriyash, Jason Rolfe, Bohdan Kulchytskyy, and Roger Melko. Quantum Boltzmann machine. *Physical Review X*, 8(2):021050, 2018.

[AKL16] Itai Arad, Tomotaka Kuwahara, and Zeph Landau. Connecting global and local energy distributions in quantum spin models on a lattice. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(3):033301, March 2016.

[AR19] Scott Aaronson and Guy N. Rothblum. Gentle measurement of quantum states and differential privacy. In *STOC'19—Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 322–333. ACM, New York, 2019.

[Ara69] Huzihiro Araki. *Gibbs states of a one dimensional quantum lattice*. *Communications in Mathematical Physics*, 14(2):120–157, Jun 1969.

[BAL19] Eyal Bairey, Itai Arad, and Netanel H Lindner. Learning a local hamiltonian from local measurements. *Physical Review Letters*, 122(2):020504, 2019.

[BHV06] S. Bravyi, M. B. Hastings, and F. Verstraete. Lieb-robinson bounds and the generation of correlations and topological quantum order. *Phys. Rev. Lett.*, 97:050401, Jul 2006.

[BKL+19] F. G. S. L. Brandão, A. Kalev, T. Li, C. Y-Y. Lin, K. M. Svore, and X. Wu. Quantum SDP Solvers: Large Speed-Ups, Optimality, and Applications to Quantum Learning. In *46th International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 132, pages 27:1–27:14, 2019.

[BMBO20] Xavier Bonet-Monroig, Ryan Babbush, and Thomas E. O'Brien. Nearly optimal measurement scheduling for partial tomography of quantum states. *Phys. Rev. X*, 10:031064, Sep 2020.

[CW20] Jordan Cotler and Frank Wilczek. Quantum overlapping tomography. *Physical Review Letters*, 124(10):100401, 2020.

[EHF19] Tim J Evans, Robin Harper, and Steven T Flammia. Scalable bayesian Hamiltonian learning. *arXiv:1912.07636*, 2019.

[FU15] Jürg Fröhlich and Daniel Ueltschi. *Some properties of correlations of quantum lattice systems in thermal equilibrium*. *Journal of Mathematical Physics*, 56(5):053302, 2015.

[Gro79] Leonard Gross. *Decay of correlations in classical lattice models at high temperature*. *Communications in Mathematical Physics*, 68(1):9–27, Feb 1979.

[Has07] Matthew B Hastings. Quantum belief propagation: An algorithm for thermal quantum systems. *Physical Review B*, 76(20):201102, 2007.

[HC71] J. M. Hammersley and P. E. Clifford. Markov field on finite graphs and lattices. 1971.

[HKK08] Masahito Hayashi, Akinori Kawachi, and Hirotada Kobayashi. Quantum measurements for hidden subgroup problems with optimal sample complexity. *Quantum Information and Computation*, 8:0345–0358, 2008.

[HW12] A. W. Harrow and A. Winter. How many copies are needed for state discrimination? *IEEE Transactions on Information Theory*, 58(1):1–2, 2012.

[Jay57] Edwin T Jaynes. Information theory and statistical mechanics. ii. *Physical Review*, 108(2):171, 1957.

[KB19] Kohtaro Kato and Brandão, Fernando G. S. L. Quantum approximate Markov chains are thermal. *Comm. Math. Phys.*, 370(1):117–149, 2019.

[KGK+14] M. Kliesch, C. Gogolin, M. J. Kastoryano, A. Riera, and J. Eisert. *Locality of Temperature*. *Phys. Rev. X*, 4:031019, Jul 2014.

[Kim17] Isaac H Kim. Markovian matrix product density operators: Efficient computation of global entropy. *arXiv:1709.07828*, 2017.

[KKBa20] Tomotaka Kuwahara, Kohtaro Kato, and Fernando G. S. L. Brandão. Clustering of conditional mutual information for quantum gibbs states above a threshold temperature. *Phys. Rev. Lett.*, 124:220601, Jun 2020.

[KMS16] Tomotaka Kuwahara, Takashi Mori, and Keiji Saito. Floquet-magnus theory and generic transient dynamics in periodically driven many-body quantum systems. *Annals of Physics*, 367:96 – 124, 2016.

[Kuw16] Tomotaka Kuwahara. Asymptotic behavior of macroscopic observables in generic spin systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(5):053103, 2016.

[LP08] Matthew S Leifer and David Poulin. Quantum graphical models and belief propagation. *Annals of Physics*, 323(8):1899–1946, 2008.

[LR72] Elliott H. Lieb and Derek W. Robinson. The finite group velocity of quantum spin systems. *Communications in Mathematical Physics*, 28(3):251–257, 1972.

[Mon15] Andrea Montanari. Computational implications of reducing data to sufficient statistics. *Electron. J. Stat.*, 9(2):2370–2390, 2015.

[NS09] Bruno Nachtergaele and Robert Sims. Locality estimates for quantum spin systems. In Vladas Sidoravičius, editor, *New Trends in Mathematical Physics*, pages 591–614, Dordrecht, 2009. Springer Netherlands.

[PY95] Yong Moon Park and Hyun Jae Yoo. *Uniqueness and clustering properties of Gibbs states for classical and quantum unbounded spin systems*. *Journal of Statistical Physics*, 80(1):223–271, Jul 1995.

[QR19] Xiao-Liang Qi and Daniel Ranard. Determining a local Hamiltonian from a single eigenstate. *Quantum*, 3:159, 2019.

[Uel04] Daniel Ueltschi. *Cluster expansions and correlation functions*. *Moscow Mathematical Journal*, 4(2):511–522, 2004.

[VMLC16] Marc Vuffray, Sidhant Misra, Andrey Lokhov, and Michael Chertkov. Interaction screening: Efficient and sample-optimal learning of ising models. In *Advances in Neural Information Processing Systems*, pages 2595–2603, 2016.

[VMN+19] Guillaume Verdon, Jacob Marks, Sasha Nanda, Stefan Leichenauer, and Jack Hidary. Quantum Hamiltonian-based models and the variational quantum thermalizer algorithm. *arXiv:1910.02071*, 2019.

[WGFC14a] Nathan Wiebe, Christopher Granade, Christopher Ferrie, and David Cory. Quantum hamiltonian learning using imperfect quantum resources. *Physical Review A*, 89(4):042314, 2014.

[WGFC14b] Nathan Wiebe, Christopher Granade, Christopher Ferrie, and David G Cory. Hamiltonian learning and certification using quantum resources. *Physical Review Letters*, 112(19):190501, 2014.