



Quantum Approximate Markov Chains are Thermal

Kohtaro Kato^{1,2} , Fernando G. S. L. Brandão²

¹ Department of Physics, Graduate School of Science, The University of Tokyo, Tokyo, Japan

² Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, CA, USA.
E-mail: koh1192035@gmail.com

Received: 16 October 2017 / Accepted: 22 April 2019

Published online: 21 June 2019 – © Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract: We prove that any one-dimensional (1D) quantum state with small quantum conditional mutual information in all certain tripartite splits of the system, which we call a *quantum approximate Markov chain*, can be well-approximated by a Gibbs state of a short-range quantum Hamiltonian. Conversely, we also derive an upper bound on the (quantum) conditional mutual information of Gibbs states of 1D short-range quantum Hamiltonians. We show that the conditional mutual information between two regions A and C conditioned on the middle region B decays exponentially with the square root of the length of B . These two results constitute a variant of the Hammersley–Clifford theorem (which characterizes Markov networks, i.e. probability distributions which have vanishing conditional mutual information, as Gibbs states of classical short-range Hamiltonians) for 1D quantum systems. The result can be seen as a strengthening— for 1D systems—of the mutual information area law for thermal states. It directly implies an efficient preparation of any 1D Gibbs state at finite temperature by a constant-depth quantum circuit.

1. Introduction

A sequence of discrete random variables X_1, \dots, X_n forms a Markov chain if X_{i+1} is uncorrelated from X_1, \dots, X_{i-1} conditioned on the value of X_i . Markov chains are a central concept in probability theory, statistics and beyond. In this paper we consider a combination of two natural generalizations of the concept of a Markov chain.

In the first we only require *approximate* independence from previous random variables, i.e. X_{i+1} should only be *almost* independent from X_1, \dots, X_{i-1} conditioned on X_i . One way to make this notion quantitative is to use the conditional mutual information, defined for every three random variables X, Y, Z drawn from the distribution $p(X, Y, Z)$ as

$$I(X : Z|Y)_p := H(XY)_p + H(YZ)_p - H(XYZ)_p - H(Y)_p,$$

where $H(X)_p := -\sum_{x_i \in X} p_X(x_i) \log p_X(x_i)$ is the Shannon entropy of the marginal distribution on X .¹ In terms of the conditional mutual information, X_1, \dots, X_n is a Markov chain if

$$I(X_1 \dots X_{i-1} : X_{i+1} \dots X_n | X_i)_p = 0 \quad \forall i \in [1, n].$$

The conditional mutual information can also be written as

$$I(X : Z | Y)_p = \mathbb{E}_{Y \sim p(Y)} I(X : Z)_{p_Y}, \quad (1)$$

where $I(X : Z)_{p_Y}$ is the mutual information:

$$I(X : Z)_{p_Y} := H(X)_{p_Y} + H(Z)_{p_Y} - H(XZ)_{p_Y}, \quad (2)$$

with $p_Y(X, Z)$ which is the conditional distribution of X and Z for given $Y = y$. Thus if $I(X : Z | Y)_p$ is small, X and Z are almost uncorrelated conditioned on Y . For $\varepsilon > 0$, we say X_1, \dots, X_n is a ε -approximate Markov chain if

$$I(X_1 \dots X_{i-1} : X_{i+1} \dots X_n | X_i)_p \leq \varepsilon \quad \forall i \in [1, n].$$

In the second, instead of considering random variables, we consider a n -partite quantum state given by a density matrix $\rho_{A_1 \dots A_n} \in \mathcal{D}(\mathcal{H}_{A_1} \otimes \dots \otimes \mathcal{H}_{A_n})$.² The quantum conditional mutual information of a tripartite state ρ_{ABC} is defined as

$$I(A : C | B)_\rho := S(AB)_\rho + S(BC)_\rho - S(ABC)_\rho - S(B)_\rho,$$

where $S(X)_\rho := -\text{Tr}(\rho_X \log \rho_X)$ is the von Neumann entropy of the reduced state on subsystem X . Quantum states satisfying $I(A : C | B)_\rho = 0$ are analogues of Markov chains of three random variables. As in the classical case (I), a multipartite quantum state ρ_{A_1, \dots, A_n} is a quantum Markov chain if

$$I(A_1 \dots A_{i-1} : A_{i+1} \dots A_n | A_i)_\rho = 0 \quad \forall i \in [1, n].$$

In this paper, we are interested in *quantum approximate Markov chains*, a combination of both generalizations. Such concept is already non-trivial for tripartite quantum states ρ_{ABC} . We can say ρ_{ABC} forms a quantum ε -approximate Markov chain if

$$I(A : C | B)_\rho \leq \varepsilon.$$

However there is no quantum analogue of Eq. (1) (see Ref. [1]) and therefore it is unclear if the definition in the above has a nontrivial meaning. A recent result in quantum information theory reveals its meaning [2]. It shows that

$$I(A : C | B)_\rho \geq \min_{\Delta: B \rightarrow BC} -2 \log F(\rho_{ABC}, \Delta_{B \rightarrow BC}(\rho_{AB})), \quad (3)$$

where the minimum is over all completely-positive and trace-preserving (CPTP) map $\Delta_{B \rightarrow BC}$ mapping $\mathcal{D}(\mathcal{H}_B)$ to $\mathcal{D}(\mathcal{H}_B \otimes \mathcal{H}_C)$, and $F(\rho, \sigma) := \text{Tr}((\sigma^{1/2} \rho \sigma^{1/2})^{1/2})$ is the fidelity. Thus if the conditional mutual information is small, A is only correlated to C through B up to a small error, in the sense that C can be approximately recovered given the information contained in B only (see Refs. [3–5]). More generally, we say ρ_{A_1, \dots, A_n} is a quantum ε -approximate Markov chain if

$$I(A_1, \dots, A_{i-1} : A_{i+1}, \dots, A_n | A_i)_\rho \leq \varepsilon \quad \forall i \in [1, n].$$

¹ In the following we always use 2 as the base of log.

² $\mathcal{D}(\mathcal{H}_{A_1} \otimes \dots \otimes \mathcal{H}_{A_n})$ is the set of density matrices over the finite-dimensional Hilbert space $\mathcal{H}_{A_1} \otimes \dots \otimes \mathcal{H}_{A_n}$.

1.1. The Hammersley–Clifford theorem. In this paper we will be interested in finding a structural characterization of quantum approximate Markov chains. Our motivation is a powerful result in statistics called the Hammersley–Clifford Theorem [6]. It states that Markov chains (and more generally Markov networks³), in which all elements of the distribution are non-zero, are equivalent to the set of Gibbs (thermal) states of nearest-neighbor Hamiltonians on a 1D open spin chain⁴:

$$p(X_1 = x_1, \dots, X_n = x_n) = \frac{1}{Z} \exp \left(- \sum_i h_i(x_i, x_{i+1}) \right),$$

for functions $h_i : \mathbb{R}^2 \rightarrow \mathbb{R}$, where

$$Z := \sum_{x_1, \dots, x_n} \exp \left(- \sum_i h_i(x_i, x_{i+1}) \right)$$

is the partition function. Here, the “temperature” is included in the interaction terms.

In Refs. [7, 8], the Hammersley–Clifford theorem was generalized to quantum Markov chains (and Markov networks): A full-rank quantum state $\rho_{A_1 \dots A_n}$ is a quantum Markov chain if, and only if, it can be written as

$$\rho_{A_1 \dots A_n} = \frac{1}{Z} \exp \left(- \sum_i h_{i, i+1} \right),$$

where $Z = \text{Tr}(\exp(-\sum_i h_{i, i+1}))$ and each $h_{i, i+1}$ only acts on subsystems $A_i A_{i+1}$, such that $[h_{i, i+1}, h_{j, j+1}] = 0$ for all i, j . Therefore we have a characterization of full-rank quantum Markov chains as Gibbs states of 1D commuting short-range quantum Hamiltonians.⁵ Conversely, this result also clarifies that correlations in Gibbs states of 1D commuting short-range Hamiltonians are always mediated through interactions between neighboring regions.

The characterization above only involves exact quantum Markov chains and a special set of short-range Hamiltonians. A natural question is whether there is a similar relation between quantum approximate Markov chains and more general quantum Gibbs states. The main result of this paper answers the question in the affirmative: we prove that quantum approximate Markov chains are equivalent to Gibbs states of 1D short-range quantum Hamiltonians, which we also call *local Gibbs states* in short.

Notation: In the following, we consider a quantum spin system Λ on a graph $G = (V, E)$, where $V = \{1, \dots, n\}$ and $E = \{(i, i+1)\}_{i=1}^{n-1}$ for $n \in \mathbb{Z}_{>0}$, i.e., a 1D open spin chain. Sometimes we also consider a closed chain by adding additional edge $(n, 1)$. The Hilbert space of a local subsystem \mathcal{H}_i corresponding to spin i is associated to each $i \in V$, which has finite dimension $d < \infty$. For a subsystem specified by subregion $X \subset V$ (we abuse the notation by using same X to denote both a subsystem and a subregion),

³ A Markov network is a generalization of a Markov chain given by random variables X_1, \dots, X_n defined on the vertices $1, \dots, n \in V$ of a graph $G = (V, E)$, such that X_i is uncorrelated from all other random variables conditioned on the random variables $\{X_j\}_{(i,j) \in E}$ associated to neighboring vertices.

⁴ For Markov networks, in turn, the Hamiltonian is a sum of local functions of variables on all cliques of the graph.

⁵ In Ref. [8] a more general result was shown for quantum Markov networks. In contrast to the classical case, positive (i.e. full-rank) quantum Markov networks are only equivalent to Gibbs states of commuting Hamiltonians with terms on the cliques of the graph if the graph is triangle free.

we denote $\mathcal{H}_X := \bigotimes_{i \in X} \mathcal{H}_i$. The logarithm of $\dim \mathcal{H}_X$ is simply denoted by $|X|$. We denote the operator norm of an operator O by $\|O\|$, and the trace norm of O by $\|O\|_1$. We say that the support of O is X if an operator O on Λ can be written as

$$O = O_X \otimes \mathbf{1}_{X^c},$$

i.e., the tensor product of some operator O_X on region X and the identity operator $\mathbf{1}_{X^c}$ acting on the complement of X (which we denoted by X^c). We will denote the support of an operator O by $\text{supp}(O)$, unless explicitly mentioned.

A short-range Hamiltonian H is a bounded Hermitian operator on \mathcal{H}_Λ which can be decomposed into

$$H = \sum_i h_i,$$

where each $\|h_i\|$ is bounded by a constant and $\text{supp}(h_i)$ only contains spins within graph distance $r < \infty$ from the spin i . We also consider a restricted Hamiltonian H_X on a region X , defined as

$$H_X = \sum_{\text{supp}(h_i) \subset X} h_i,$$

i.e., the sum of interactions acting on spins sitting *inside* of X . The Gibbs state ρ^{H_X} of the Hamiltonian H_X at (inverse) temperature $\beta > 0$ is defined as

$$\rho^{H_X} = \frac{e^{-\beta H_X}}{Z_X}, \quad (4)$$

where $Z_X = \text{Tr}[e^{-\beta H_X}]$. Note that we will omit β in the notations and simply denote ρ^{H_X} and Z_X , while they depend on β . The reduced state of the Gibbs state on a subregion $Y \subset X$ is denoted by $\rho_Y^{H_X}$.

Throughout the paper, we often consider a disjoint tripartition ABC of Λ . We say B shields A from C if A and C are indirectly connected through B . We denote by $d(A, C)$ the graph distance between A and C . $d(A, C) = |B|$ if B shields A from C and is connected.

2. Results

In this section we present the main results of this paper. For two quantum states ρ, σ , we denote their (quantum) relative entropy as

$$S(\rho \parallel \sigma) := \text{Tr}(\rho(\log \rho - \log \sigma))$$

if $\text{supp}(\rho) \subset \text{supp}(\sigma)$, and $S(\rho \parallel \sigma) := +\infty$ otherwise (Here “ $\text{supp}(\rho)$ ” means the subspace spanned by eigenvectors of ρ with nonzero eigenvalues).

2.1. Approximation of quantum approximate Markov states by local Gibbs states. Let us divide 1D spin chain Λ into m connected regions $A_1 A_2 \dots A_m$. We denote the coarse-

grained 1D spin chain $A_1 A_2 \dots A_m$ by A . Our first result is the following theorem (see Sect. 3 for the proof):

Theorem 1. *Let $\rho_{A_1 \dots A_m}$ be a quantum ε -approximate Markov chain on a 1D open chain for $\varepsilon > 0$. Then there exists a short-range Hamiltonian $H = \sum_{i=1}^{m-1} h_{A_i, A_{i+1}}$ with $\text{supp}(h_{A_i, A_{i+1}}) = A_i A_{i+1}$, such that*

$$S\left(\rho \left\| \frac{e^{-H}}{Z}\right.\right) \leq \varepsilon m, \quad (5)$$

where $Z = \text{Tr} e^{-H}$.

From the relation

$$S(\rho \|\sigma) \geq -2 \log F(\rho, \sigma),$$

we find that the state ρ is also close in fidelity to a local Gibbs state. Note that Theorem 1 is not restricted to full-rank states.

It is natural to expect that there exists a similar bound for 1D closed chains. In this paper, we say a state $\rho_{A_1 \dots A_m}$ is a quantum ε -approximate Markov chain on a 1D closed chain if

$$I(A_i : A \setminus A_{i-1} A_i A_{i+1} | A_{i-1} A_{i+1})_\rho \leq \varepsilon, \quad \forall i \in [1, m] \quad (6)$$

in analogy with a quantum Markov network.⁶ Here, we imposed the periodic boundary condition on labels, e.g., $m + 1 \equiv 1$. In this situation, the proof of Theorem 1 does not work straightforwardly. To solve this difficulty, we consider two sufficient conditions as assumptions which show the closeness to local Gibbs states respectively.

The first assumption is that the existence of the finite correlation length in terms of the quantum mutual information, which is obtained by replacing the Shannon entropy by the von Neumann entropy in Eq. (2). The second assumption is called the uniform Markov property [9] requiring that the reduced states of the state are also approximately Markov chains. Note that the second assumption may be derived from the definition of the quantum approximate Markov chains while we have not had any result on it.

Theorem 2. *Let $\rho_{A_1 \dots A_m}$ be a quantum ε -approximate Markov chain on a 1D closed chain for $\varepsilon > 0$.*

(i) *Assume that $\rho_{A_1 \dots A_m}$ also satisfies*

$$I(A_i : A \setminus A_{i-1} A_i A_{i+1})_\rho \leq \varepsilon \quad \forall i \in [1, m].$$

Then there exists a short-range Hamiltonian $H = \sum_i h_{A_{i-1}, A_i, A_{i+1}}$, with $\text{supp}(h_{A_{i-1}, A_i, A_{i+1}}) = A_{i-1} A_i A_{i+1}$, such that

$$S\left(\rho \left\| \frac{e^{-H}}{Z}\right.\right) \leq \varepsilon m.$$

(ii) *Assume that for any $i \in [1, m]$, $\text{Tr}_{A_i}(\rho_{A_1 \dots A_m})$ is a quantum ε -approximate Markov chain for the 1D open chain $A_{i+1} A_{i+2} \dots A_{i-1}$ (we used $m + 1 \equiv 1$). Then there exists a short-range Hamiltonian $H = \sum_i h_{A_{i-1}, A_i, A_{i+1}}$, with $\text{supp}(h_{A_{i-1}, A_i, A_{i+1}}) = A_{i-1} A_i A_{i+1}$, such that*

$$S\left(\rho \left\| \frac{e^{-H}}{Z}\right.\right) \leq \varepsilon m.$$

⁶ Any quantum ε -approximate Markov chain $\rho_{A_1 \dots A_m}$ satisfies $I(A_i : A \setminus A_{i-1} A_i A_{i+1} | A_{i-1} A_{i+1})_\rho \leq 2\varepsilon$ for all $i \in [3, m - 2]$.

The approximate Markov property appears in analysis of gapped ground states of many-body systems (see e.g., Refs. [10, 11]). In these cases, the additional assumptions seem to be satisfied for certain choice of subsystems. Interestingly, there exists a class of states which are locally quantum approximate Markov chains but globally not. A simple example is the n -qubit GHZ state

$$|GHZ_n\rangle = \frac{1}{\sqrt{2}} (|0000\dots 0\rangle + |1111\dots 1\rangle)$$

on a spin chain (either open or closed). When we trace out one qubit from the chain, the reduced state exactly become a quantum Markov chain. However, this state has $I(A : C|B) = 1$ for any tripartition ABC of the whole system where B shields A from C . Therefore, it is not a quantum Markov chain globally. A similar situation arises when considering a ring-like regions in systems with topological order [12]. We show that for these cases, the value of the conditional mutual information for the whole system approximately represents the distance from the set of local Gibbs states. We will discuss an application of this result for analysis of entanglement spectrum in 2D topologically ordered phases in a complementary work [13].

Theorem 3. *Consider a 1D spin chain $X = X_1 X_2 \dots X_m$ with the size $N = |X_1 \dots X_m|$. Let $\rho_{X_1 \dots X_m}$ be a state such that the reduced state obtained by tracing out X_i is a quantum ε -approximate Markov chain for all $i \in [1, m]$. Define the set of Gibbs states of short-range Hamiltonians with interaction strength K as*

$$E_{nn}^K := \left\{ e^{-H} \left| H = \sum_i h_{X_i X_{i+1}} \text{Tr}(e^{-H}) = 1, \|h_{X_i X_{i+1}}\| \leq K \right. \right\}.$$

Note that here we include the normalization factor in the Hamiltonians. Then, for $K = \Theta(N)$ and sufficiently small $\varepsilon > 0$, there exists a constant $c > 0$ such that for any tripartition ABC of the whole system such that B shields A from C , it holds that

$$\min_{\mu \in E^K} S(\rho_X \|\mu) = I(A : C|B)_\rho + \varepsilon(N, \delta) \quad (7)$$

and

$$|\varepsilon(N, \delta)| \leq cN^{\frac{5}{2}} \delta^{\frac{1}{16}},$$

where $\delta = 8\sqrt{\varepsilon} + 2^{-N}$.

Here we used X_i to label subsystems instead of A_i to avoid confusion with $A \subset X$.

2.2. Quantum approximate Markov property of 1D Gibbs states. Our second main result is a kind of converse to Theorems 1 and 2. (see Sect. 4 for the proof):

Theorem 4. *Let $H = \sum_i h_i$ be a short-range 1D Hamiltonian with $\|h_i\| \leq 1$ and $l_0, c, c' > 0$ be universal constants. For an inverse temperature $\beta > 0$ and any partition ABC with $d(A, C) \geq l_0$, there exists a CPTP-map $\Lambda_{B \rightarrow BC} : \mathcal{D}(\mathcal{H}_B) \rightarrow \mathcal{D}(\mathcal{H}_B \otimes \mathcal{H}_C)$ such that*

$$\left\| \rho^{H_{ABC}} - \Lambda_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}}) \right\|_1 \leq e^{-q(\beta)\sqrt{d(A,C)}}, \quad (8)$$

where $q(\beta) = ce^{-c'\beta}$ if the correlation length of $\rho^{H_{ABC}}$ is $\xi = e^{\mathcal{O}(\beta)}$.

The theorem above states that if we choose the region B sufficiently large, the Gibbs state can be approximately recovered from the partial trace over C by performing a recovery map on B . In turn, the statement implies the corresponding conditional mutual information decays similarly:

Corollary 5. *Under the setting of Theorem 4,*

$$I(A : C|B)_{\rho^{H_{ABC}}} \leq 6 \left(d(A, C) + \frac{8(1 + \frac{q(\beta)}{2} \sqrt{d(A, C)})}{q(\beta)^2} \right) e^{-\frac{q(\beta)}{2} \sqrt{d(A, C)}}.$$

Therefore, $I(A : C|B)_{\rho} \leq \mathcal{O}(e^{-\Theta(\sqrt{d(A, C)})})$ and thus any 1D local Gibbs state is a quantum approximate Markov chain with small ε after certain coarse-graining. Conversely, Theorem 1 shows that any quantum approximate Markov chain can be well-approximated by some 1D local Gibbs state. Therefore the combination of the two results can be regarded as a variant of the Hammersley–Clifford theorem for quantum approximate Markov chains. Below we discuss two implications of our results:

2.3. Saturation rate of area law for 1D Gibbs states. A Gibbs state of a short-range Hamiltonian obeys an area law in terms of the mutual information. For instance, for any Gibbs state ρ^H of a short-range Hamiltonian on a lattice, it holds that [14]

$$I(A : A^c)_{\rho^H} \leq \beta J |\partial A|, \quad (9)$$

where $J > 0$ is a constant only depending on the locality of H and the norm of the local interactions. The upper bound is a constant when the system is 1D.

The area law represented by Eq. (9) implies a decay of the conditional mutual information. Consider a 1D spin chain system and let A be a simply connected region. We divide A^c into $\{b_i\}_i$ so that b_1 shields b_2 from A , b_2 shields b_3 from Ab_1 and so on. We set $B_l = b_1 b_2 \dots b_l$. By the monotonicity of the mutual information under the partial trace, we have

$$I(A : B_l)_{\rho^H} \leq I(A : B_{l+1})_{\rho^H} \leq I(A : A^c)_{\rho^H} \leq \beta J |\partial A|.$$

Since the upper bound is independent of l , $I(A : b_{l+1}|B_l)_{\rho^H} = I(A : B_{l+1})_{\rho^H} - I(A : B_l)_{\rho^H}$ eventually vanishes when l grows. Corollary 5 goes one step further and quantifies the rate at which $I(A : B_l)_{\rho^H}$ saturates when l grows. Indeed if each size of b_i and l are sufficiently large, we have

$$I(A : B_{l+1})_{\rho^H} - I(A : B_l)_{\rho^H} = I(A : b_{l+1}|B_l)_{\rho^H} \leq C e^{-c\sqrt{l}},$$

for some constants $C, c > 0$. Therefore, the mutual information of a 1D local Gibbs state saturates the upper bound of the area law at least subexponentially fast in l .

2.4. A short depth representation of 1D Gibbs states. Theorem 4 ensures that there exist local CPTP-maps which approximately recover a 1D Gibbs state from tracing out operations on local regions. In other words, the Gibbs state can be prepared by local operations on reduced states on separated subregions (Fig. 1).

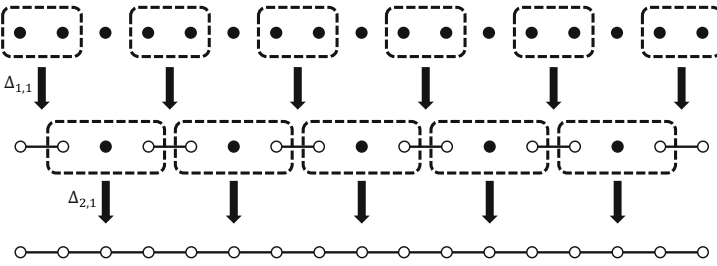


Fig. 1. A schematic picture of the preparation algorithm for 1D Gibbs states. At the first step (upside), we perform CPTP-map $\otimes_i \Delta_{1,i}$ on a product state (black dots). Each $\Delta_{1,i}$ acts on a small set of spins (dotted circle) and produces the reduced state of the target Gibbs state on the set (connected dots). At the second step (downside), we perform another CPTP-map to concatenate these reduced states locally. Due to the approximate Markov property of the Gibbs state, the output state is close to the Gibbs state

Corollary 6. A Gibbs state of any 1D short-range Hamiltonian at constant (inverse) temperature $\beta > 0$ can be well-approximated by a depth-two (mixed) circuit with each gate acting on $O(\log^2(n))$ qubits.

In more detail, there is a CPTP-map (corresponding to the circuit) of the form

$$\Delta = \left(\bigotimes_i \Delta_{2,i} \right) \left(\bigotimes_i \Delta_{1,i} \right),$$

with each local CPTP-map $\Delta_{k,i}$ acting on $O(e^{O(\beta)} \log^2(n/\varepsilon))$ sites, with $\Delta_{k,i}$ and $\Delta_{k,j}$ acting on non-overlapping sites for $i \neq j$, such that

$$\left\| \Delta(\tau) - \frac{e^{-\beta H}}{Z} \right\|_1 \leq \varepsilon,$$

with τ the maximally mixed state.

The proof is given in Sect. 4.4. An earlier result [15] proved that 1D Gibbs states of local Hamiltonians at finite constant temperature can be approximated by a matrix product operator of polynomial bond dimension, which implies they can be constructed efficiently on a quantum computer. However this result does not give that the state can be constructed by a short depth quantum circuit, as Corollary 6 shows. A similar construction methods appear in Refs. [16, 17] under certain assumptions on the approximate Markov property.

3. Gibbs State Representations of 1D States with the Approximate Markov Property

In this section, we prove Theorems 1 and 2 which states that quantum approximate Markov chains can be approximated by 1D local Gibbs states. We also prove Theorem 3. In these proofs, a relationship between Gibbs states and the maximum entropy principle plays an important role. We first review the relationship and then turn to the proofs.

3.1. The maximum entropy state and Gibbs states. The maximum entropy principle, introduced by E. T. Jaynes in classical statistical mechanics [18, 19], is a method to choose an inference under partial information represented by some linear constraints. According to the maximum entropy principle, the most “unbiased” inference is given by the probability distribution with maximum entropy among all distributions satisfying the linear constraints. The solution of this optimization problem is given by a Gibbs distribution, as can be shown by the method of Lagrange multipliers.

This framework has been generalized to quantum systems (see, e.g., Refs. [20, 21]). Especially we are interested in the case where the linear constraints are given by reduced density matrices. Let ρ be a quantum state in $\mathcal{D}(\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n)$. Consider sets of subsystems labeled by X_1, \dots, X_m with $X_i \subset \{1, \dots, n\}$ and let $\mathbf{X} = \{X_1, \dots, X_m\}$. We define the set $R_\rho(\mathbf{X})$ by

$$R_\rho(\mathbf{X}) := \left\{ \sigma \in \mathcal{D}(\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n) \mid \sigma_{X_i} = \rho_{X_i}, (1 \leq \forall i \leq m) \right\}. \quad (10)$$

$R_\rho(\mathbf{X})$ is the set of all states with the same reduced states as ρ for all X_i . Since $R_\rho(\mathbf{X})$ is a closed convex set, there exists a unique state such that

$$\sigma_{\max} := \arg \max_{\sigma \in R_\rho(\mathbf{X})} S(\sigma).$$

We call σ_{\max} *the maximum entropy state in $R_\rho(\mathbf{X})$* . Similar to the classical setting, σ_{\max} is given by a Gibbs state of a Hamiltonian with a specific structure. Let us consider the set of Gibbs states $E(\mathbf{X})$ defined as

$$E(\mathbf{X}) := \left\{ \frac{e^{-H}}{Z} \in \mathcal{D}(\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n) \mid H = \sum_{i=1}^m H_{X_i} \right\}, \quad (11)$$

where $\text{supp}(H_{X_i}) = X_i$. For any $\omega \in E(\mathbf{X})$, it is proven that the Pythagorean theorem

$$S(\rho \parallel \omega) = S(\rho \parallel \sigma_{\max}) + S(\sigma_{\max} \parallel \omega) \quad (12)$$

holds (Corollary 3.7 & Theorem 6.16, [20]). The maximum entropy state σ_{\max} is the unique element of the intersection of $R_\rho(\mathbf{X})$ and $\overline{E(\mathbf{X})}$: the closure of $E(\mathbf{X})$ in the reverse-information topology [20] defined as

$$\overline{E(\mathbf{X})} := \left\{ \sigma \in \mathcal{D}(\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n) \mid \inf_{\omega \in E(\mathbf{X})} S(\sigma \parallel \omega) = 0 \right\}.$$

Therefore, the Pythagorean theorem (12) implies that

$$\inf_{\omega \in E(\mathbf{X})} S(\rho \parallel \omega) = S(\rho \parallel \sigma_{\max}), \quad (13)$$

since $\inf_{\omega \in E(\mathbf{X})} S(\sigma_{\max} \parallel \omega) = 0$. By choosing ω in Eq. (12) as the completely mixed state, which is always contained in $E(\mathbf{X})$, we have

$$S(\rho \parallel \sigma_{\max}) = S(\sigma_{\max}) - S(\rho) \quad (14)$$

and therefore

$$\inf_{\omega \in E(\mathbf{X})} S(\rho \parallel \omega) = S(\sigma_{\max}) - S(\rho). \quad (15)$$

We will use these formulas in the proof of theorems.

3.2. *Proof of Theorem 1.* We now prove:

Theorem 1. *Let $\rho_{A_1 \dots A_m}$ be a quantum ε -approximate Markov chain on a 1D open chain for $\varepsilon > 0$. Then there exists a short-range Hamiltonian $H = \sum_{i=1}^{m-1} h_{A_i, A_{i+1}}$ with $\text{supp}(h_{A_i, A_{i+1}}) = A_i A_{i+1}$, such that*

$$S\left(\rho \left\| \frac{e^{-H}}{Z}\right.\right) \leq \varepsilon m.$$

Proof. Let $\sigma_{A_1 \dots A_m}$ be the maximum entropy state such that

$$\sigma_{A_i A_{i+1}} = \rho_{A_i A_{i+1}} \quad (16)$$

for all $i \in [1, m-1]$. We will show

$$S(\sigma) \leq S(\rho) + \varepsilon m.$$

The result then follows from Eq. (15), since $\sigma_{A_1 \dots A_m}$ is an element of $\overline{E(\mathbf{X})}$ for $\mathbf{X} = \{A_i A_{i+1}\}_{i=1}^{m-1}$ which is a (closure of) set of local Gibbs states.

By strong subadditivity, we find

$$\begin{aligned} S(A_1 \dots A_m)_\sigma &\leq S(A_1 A_2)_\sigma - S(A_2)_\sigma + S(A_2 \dots A_m)_\sigma \\ &\leq S(A_1 A_2)_\sigma - S(A_2)_\sigma + S(A_2 A_3)_\sigma - S(A_3)_\sigma + S(A_3 \dots A_m)_\sigma \\ &\vdots \\ &\leq \sum_{i=1}^{m-2} [S(A_i A_{i+1})_\sigma - S(A_{i+1})_\sigma] + S(A_{m-1} A_m)_\sigma \\ &= \sum_{i=1}^{m-2} [S(A_i A_{i+1})_\rho - S(A_{i+1})_\rho] + S(A_{m-1} A_m)_\rho. \end{aligned} \quad (17)$$

The last equality follows from Eq. (16). Since $\rho_{A_1 \dots A_m}$ is a quantum ε -approximate Markov chain,

$$I(A_i : A_{i+2} \dots A_m | A_{i+1})_\rho \leq \varepsilon \quad \forall i \in [1, m-2],$$

which can be rewritten as

$$S(A_i A_{i+1})_\rho - S(A_{i+1})_\rho + S(A_{i+1} \dots A_m)_\rho \leq S(A_i A_{i+1} \dots A_m)_\rho + \varepsilon,$$

i.e., the strong subadditivity is saturated up to error ε . Therefore we obtain

$$\begin{aligned} &\sum_{i=1}^{m-2} [S(A_i A_{i+1})_\rho - S(A_{i+1})_\rho] + S(A_{m-1} A_m)_\rho \\ &\leq \sum_{i=1}^{m-2} S(A_i | A_{i+1})_\rho + S(A_{m-2} A_{m-1} A_m) + \varepsilon \\ &\leq \sum_{i=1}^{m-3} S(A_i | A_{i+1})_\rho + S(A_{m-3} \dots A_m) + 2\varepsilon \\ &\vdots \\ &\leq S(A_1 | A_2)_\rho + S(A_2 \dots A_m)_\rho + (m-2)\varepsilon \\ &\leq S(A_1 \dots A_m)_\rho + (m-1)\varepsilon, \end{aligned} \quad (18)$$

where $S(A|B)_\rho := S(AB)_\rho - S(B)_\rho$ is the conditional entropy. Combining Eqs. (17) and (18) we have

$$S(\rho_A \| \sigma_A) = S(\sigma_A) - S(\rho_A) \leq (m - 1)\varepsilon.$$

Since $\sigma_A \in \overline{E(\mathbf{X})}$, there exists a Gibbs state

$$\omega_A = \frac{1}{Z} \exp \left(- \sum_{i=1}^{m-1} h_{A_i A_{i+1}} \right) \in E(\mathbf{X})$$

which satisfies

$$S(\sigma_A \| \omega_A) \leq \varepsilon.$$

Using the Pythagorean theorem, we obtain

$$S(\rho_A \| \omega_A) = S(\rho_A \| \sigma_A) + S(\sigma_A \| \omega_A) \leq m\varepsilon,$$

which completes the proof. \square

3.3. Proof of Theorem 2. Let us first restate the theorem:

Theorem 2. *Let $\rho_{A_1 \dots A_m}$ be a quantum ε -approximate Markov chain on a 1D closed chain for $\varepsilon > 0$.*

(i) *Assume that $\rho_{A_1 \dots A_m}$ also satisfies*

$$I(A_i : A \setminus A_{i-1} A_i A_{i+1})_\rho \leq \varepsilon \quad \forall i \in [1, m].$$

Then there exists a short-range Hamiltonian $H = \sum_i h_{A_{i-1}, A_i, A_{i+1}}$, with $\text{supp}(h_{A_{i-1}, A_i, A_{i+1}}) = A_{i-1} A_i A_{i+1}$, such that

$$S \left(\rho \left\| \frac{e^{-H}}{Z} \right. \right) \leq \varepsilon m.$$

(ii) *Assume that for any $i \in [1, m]$, $\text{Tr}_{A_i}(\rho_{A_1 \dots A_m})$ is a quantum ε -approximate Markov chain for the 1D open chain $A_{i+1} A_{i+2} \dots A_{i-1}$ (we used $m + 1 \equiv 1$). Then there exists a short-range Hamiltonian $H = \sum_i h_{A_{i-1}, A_i, A_{i+1}}$, with $\text{supp}(h_{A_{i-1}, A_i, A_{i+1}}) = A_{i-1} A_i A_{i+1}$, such that*

$$S \left(\rho \left\| \frac{e^{-H}}{Z} \right. \right) \leq \varepsilon m.$$

Proof. Theorem 2(i) can be shown in the following way. Let $\sigma_{A_1 \dots A_m}$ be the maximum entropy state such that

$$\sigma_{A_{i-1} A_i A_{i+1}} = \rho_{A_{i-1} A_i A_{i+1}}$$

for all $i \in [1, m]$ (with the periodic boundary condition). As in the previous section, we recursively use the strong subadditivity and obtain

$$\begin{aligned}
S(A_1 \dots A_m)_\sigma &\leq S(A_1 A_2 A_3)_\sigma - S(A_1 A_3)_\sigma + S(A_1 A_3 A_4 \dots A_m)_\sigma \\
&\leq S(A_2 | A_1 A_3)_\sigma + S(A_3 A_4 A_5)_\sigma - S(A_3 A_5)_\sigma + S(A_1 A_3 A_5 \dots A_m)_\sigma \\
&\quad \vdots \\
&\leq \sum_{i=1}^{\lfloor \frac{m}{2} \rfloor} S(A_{2i} | A_{2i-1} A_{2i+1})_\sigma + S(A_1 A_3 A_5 \dots A_{2\lceil \frac{m}{2} \rceil - 1})_\sigma \\
&\leq \sum_{i=1}^{\lfloor \frac{m}{2} \rfloor} S(A_{2i} | A_{2i-1} A_{2i+1})_\sigma + \sum_{i=1}^{\lceil \frac{m}{2} \rceil} S(A_{2i-1})_\sigma \\
&= \sum_{i=1}^{\lfloor \frac{m}{2} \rfloor} S(A_{2i} | A_{2i-1} A_{2i+1})_\rho + \sum_{i=1}^{\lceil \frac{m}{2} \rceil} S(A_{2i-1})_\rho,
\end{aligned}$$

where we used subadditivity $S(A_i A_j) \leq S(A_i) + S(A_j)$ in the second inequality. From the additional assumption on the mutual information,

$$0 \leq S(A_1)_\sigma + S(A_3)_\sigma - S(A_1 A_3)_\sigma \leq I(A_1 : A_3 A_4 \dots A_{m-1})_\sigma \leq \varepsilon.$$

By using this type of inequalities, we obtain that

$$\begin{aligned}
\sum_{i=1}^{\lceil \frac{m}{2} \rceil} S(A_{2i-1})_\rho &\leq S(A_1 A_3)_\rho + \sum_{i=3}^{\lceil \frac{m}{2} \rceil} S(A_{2i-1})_\rho + \varepsilon \\
&\quad \vdots \\
&\leq S(A_1 A_3 A_5 \dots A_{2\lceil \frac{m}{2} \rceil - 1})_\rho + \left(\left\lceil \frac{m}{2} \right\rceil - 1 \right) \varepsilon.
\end{aligned}$$

Since $\rho_{A_1 \dots A_n}$ is a quantum ε -approximate Markov chain on a closed chain, one can further show that

$$\begin{aligned}
S(A_1 \dots A_m)_\sigma &\leq \sum_{i=1}^{\lfloor \frac{m}{2} \rfloor} S(A_{2i} | A_{2i-1} A_{2i+1})_\rho + S(A_1 A_3 A_5 \dots A_{2\lceil \frac{m}{2} \rceil - 1})_\rho + \left(\left\lceil \frac{m}{2} \right\rceil - 1 \right) \varepsilon \\
&\leq S(A_1 A_2 \dots A_m)_\rho + (m-1)\varepsilon.
\end{aligned}$$

The rest part of the proof is the same as the proof of Theorem 1.

To prove Theorem 2(ii), we use the strong subadditivity to the maximum entropy state $\sigma_{A_1 \dots A_m}$ to obtain:

$$\begin{aligned}
S(A_1 \dots A_m)_\sigma &\leq S(A_m A_1 A_2)_\sigma - S(A_m A_2)_\sigma + S(A_2 A_3 \dots A_m)_\sigma \\
&\leq S(A_1 | A_m A_2)_\sigma + S(A_2 A_3)_\sigma - S(A_3)_\sigma + S(A_3 \dots A_m)_\sigma \\
&\quad \vdots \\
&\leq S(A_1 | A_m A_2)_\sigma + \sum_{i=2}^{m-2} S(A_i | A_{i+1})_\sigma + S(A_{m-1} A_m)_\sigma \\
&\leq S(A_1 | A_m A_2)_\rho + \sum_{i=2}^{m-2} S(A_i | A_{i+1})_\rho + S(A_{m-1} A_m)_\rho.
\end{aligned}$$

Recall that $\rho_{A_2 \dots A_m}$ is a quantum ε -approximate Markov chain by assumption. Therefore, we have

$$\begin{aligned}
S(A_1 | A_m A_2)_\rho + \sum_{i=2}^{m-2} S(A_i | A_{i+1})_\rho + S(A_{m-1} A_m)_\rho \\
\leq S(A_1 | A_m A_2)_\rho + S(A_2 \dots A_m)_\rho + (m-2)\varepsilon \\
\leq S(A_1 \dots A_m)_\rho + (m-1)\varepsilon,
\end{aligned}$$

from which we complete the proof in the same way as in the proof of Theorem 1. \square

3.4. *Proof of Theorem 3.* Finally, we prove Theorem 3:

Theorem 3. *Consider a 1D spin chain $X = X_1 X_2 \dots X_m$ with the size $N = |X_1 \dots X_m|$. Let $\rho_{X_1 \dots X_m}$ be a state such that the reduced state obtained by tracing out X_i is a quantum ε -approximate Markov chain for all $i \in [1, m]$. Define the set of Gibbs states of short-range Hamiltonians with interaction strength K as*

$$E_{nm}^K := \left\{ \frac{1}{Z} e^{-H} \mid H = \sum_i h_{X_i X_{i+1}}, \|h_{X_i X_{i+1}}\| \leq K \right\}.$$

Then, for $K = \Theta(N)$ and sufficiently small $\varepsilon > 0$, there exists a constant $c > 0$ such that for any tripartition ABC of the whole system such that B shields A from C , it holds that

$$\min_{\mu \in E^K} S(\rho_X \parallel \mu) = I(A : C | B)_\rho + \varepsilon(N, \delta)$$

and

$$|\varepsilon(N, \delta)| \leq c N^{\frac{5}{2}} \delta^{\frac{1}{16}},$$

where $\delta = 8\sqrt{\varepsilon} + 2^{-N}$.

The strategy of the proof is as follows. We first construct a global state $\tilde{\rho}'_{ABC}$ on ABC from a reduced state of ρ_{ABC} by using recovery maps. We then introduce $\tilde{\rho}_{ABC}$, a modification of $\tilde{\rho}'_{ABC}$ and define a Gibbs state of a nearest-neighbor Hamiltonian $\tilde{\pi}_{ABC}$ from its reduced states. Finally we show that this Gibbs state is almost the closest state in the set of Gibbs states of bounded nearest-neighbor Hamiltonians.

Proof. Without loss of generality, we consider a system defined on a 1D closed spin chain, e.g., $X_1 \equiv X_{m+1}$, and assume that $|X_1| = \max_i |X_i|$. Let $A \equiv X_1$, $B_1 \equiv X_m X_2$, $B_2 \equiv X_{m-1} X_3$ and $C \equiv X_4 X_5 \dots X_{m-2}$. In the following, we use both notations $X_1 \dots X_m$ and ABC interchangeably. Note that if we choose another tripartition $A'B'C'$ satisfying the condition instead of ABC , the chain rule of the conditional mutual information:

$$I(A : CD|B)_\rho = I(A : C|B)_\rho + I(A : D|BC)_\rho \quad (19)$$

and the assumption imply that

$$|I(A : C|B)_\rho - I(A' : C'|B')_\rho| \leq \varepsilon.$$

From the results in Ref. [2], there exist CPTP maps called (approximate) recovery maps such that

$$\|\rho_{AB_1 B_2} - \Lambda_{B_1 \rightarrow AB_1}(\rho_B)\|_1 \leq 2\sqrt{\varepsilon}, \quad (20)$$

$$\|\rho_{B_1 B_2 C} - \Lambda_{B_2 \rightarrow B_2 C}(\rho_B)\|_1 \leq 2\sqrt{\varepsilon}, \quad (21)$$

where we omitted the identity maps for simplicity. By using these maps, we define a global state

$$\tilde{\rho}'_{ABC} := \Lambda_{B_2 \rightarrow B_2 C} \circ \Lambda_{B_1 \rightarrow AB_1}(\rho_B).$$

Since the CPTP-maps recover reduced states with good accuracy, ρ'_{ABC} has almost same bipartite marginals on, e.g., AB .

$$\begin{aligned} \|\rho_{AB} - \tilde{\rho}'_{AB}\|_1 &\leq \|\rho_{AB} - \Lambda_{B_1 \rightarrow AB_1}(\rho_B)\|_1 + \|\Lambda_{B_1 \rightarrow AB_1}(\rho_B) - \tilde{\rho}'_{AB}\|_1 \\ &\leq \|\rho_{AB} - \Lambda_{B_1 \rightarrow AB_1}(\rho_B)\|_1 + \|\Lambda_{B_1 \rightarrow AB_1}(\rho_{BC}) - \tilde{\rho}'_{ABC}\|_1 \\ &\leq 2\sqrt{\varepsilon} + \|\Lambda_{B_1 \rightarrow AB_1}(\rho_{BC}) - \Lambda_{B_1 \rightarrow AB_1} \circ \Lambda_{B_2 \rightarrow B_2 C}(\rho_B)\|_1 \\ &\leq 4\sqrt{\varepsilon}. \end{aligned} \quad (22)$$

The first inequality follows from the triangle inequality of the trace norm, and the second inequality follows from the monotonicity of the trace-norm under Tr_C . The same bound holds for marginals on BC as well. Equations (20) and (22) imply that $\tilde{\rho}'_{ABC}$ is approximately recoverable state:

$$\|\tilde{\rho}'_{ABC} - \Lambda_{B_2 \rightarrow B_2 C}(\tilde{\rho}'_{AB})\|_1 \leq \|\tilde{\rho}'_{ABC} - \Lambda_{B_2 \rightarrow B_2 C}(\rho_{AB})\|_1 + \|\rho_{AB} - \tilde{\rho}'_{AB}\|_1 \leq 6\sqrt{\varepsilon}. \quad (23)$$

For two quantum states ρ_{AB}, σ_{AB} satisfying $\|\rho_{AB} - \sigma_{AB}\|_1 < \delta \leq 1$, the (Alicki-)Fannes inequality [22]

$$\begin{aligned} |S(A|B)_\rho - S(A|B)_\sigma| &\leq 4\delta|A| + 2h_2(\delta) \\ &\leq 6\sqrt{\delta}|A| \end{aligned}$$

holds with the binary entropy $h_2(\delta) = H(p = \{\delta, 1 - \delta\})$. This inequality and Eq. (23) yields

$$\begin{aligned} I(A : C|B)_{\tilde{\rho}'} &= S(A|B)_{\tilde{\rho}'} - S(A|BC)_{\tilde{\rho}'} \\ &\leq S(A|BC)_{\tilde{\rho}''} - S(A|BC)_{\tilde{\rho}'} \\ &\leq 24\sqrt{\varepsilon}|A| + 2h_2(6\sqrt{\varepsilon}) \\ &\leq 6\sqrt{6}|A|\varepsilon^{\frac{1}{4}}, \end{aligned}$$

where we denote $\tilde{\rho}'_{ABC} := \Lambda_{B_2 \rightarrow B_2 C}(\tilde{\rho}'_{AB})$ in the second line which follows from the data processing inequality for $\Lambda_{B_2 \rightarrow B_2 C}$. Therefore, $\tilde{\rho}'_{ABC}$ is a quantum approximate Markov chain for small $\varepsilon > 0$.

Next, define a full-rank modification of $\tilde{\rho}'_{ABC}$, that is,

$$\tilde{\rho}_{ABC} := \left(1 - \frac{1}{2^{N-1}}\right) \tilde{\rho}'_{ABC} + \frac{1}{2^{N-1}} \tau_{ABC}, \quad (24)$$

where τ_{ABC} is the completely mixed state on ABC (recall that N represents the logarithm of the total dimension). Since by definition

$$\|\tilde{\rho}_{ABC} - \tilde{\rho}'_{ABC}\|_1 \leq 2^{-N}, \quad (25)$$

the Fannes inequality implies that

$$I(A : C|B)_{\tilde{\rho}} \leq I(A : C|B)_{\tilde{\rho}'} + \frac{12|A|}{\sqrt{2^N}} \quad (26)$$

$$\leq 6\sqrt{6}|A|\varepsilon^{\frac{1}{4}} + \frac{12|A|}{\sqrt{2^N}}. \quad (27)$$

Therefore $\tilde{\rho}_{ABC}$ is still an approximate Markov chain for large N .

Then, we construct a Gibbs state $\tilde{\pi}_X$

$$\tilde{\pi}_X := \frac{1}{Z} e^{-H_X^{\tilde{\rho}}} \quad (28)$$

from the reduced states of $\tilde{\rho}_{ABC}$, where Z is the normalizer and

$$H_X^{\tilde{\rho}} := - \sum_{i=1}^m (\ln \tilde{\rho}_{X_i X_{i+1}} - \ln \tilde{\rho}_{X_i}).$$

$\tilde{\pi}_X$ is an element of E_{nn}^K with $K = \Theta(N)$. In the following, we show that $\tilde{\pi}$ is close to $\tilde{\rho}$.

By definition of $H_X^{\tilde{\rho}}$, it holds that

$$S(\tilde{\rho}_X \| e^{-H_X^{\tilde{\rho}}}) = \sum_{i=1}^m S(X_i | X_{i+1})_{\tilde{\rho}} - S(X_1 \dots X_m)_{\tilde{\rho}}. \quad (29)$$

Note that $\sum_{i=1}^m S(X_i | X_{i+1})_{\tilde{\rho}} = \sum_{i=1}^m S(X_{i+1} | X_i)_{\tilde{\rho}}$ by $X_{m+1} = X_1$. By an iterative calculation, we have

$$\begin{aligned} \sum_{i=1}^m S(X_i | X_{i+1})_{\tilde{\rho}} &= S(X_1 | X_2)_{\tilde{\rho}} + S(X_2 | X_3)_{\tilde{\rho}} + S(X_3 | X_4)_{\tilde{\rho}} + \sum_{i=4}^m S(X_i | X_{i+1})_{\tilde{\rho}} \\ &= S(X_1 | X_2)_{\tilde{\rho}} + I(X_2 : X_4 | X_3)_{\tilde{\rho}} + S(X_2 X_3 | X_4)_{\tilde{\rho}} + \sum_{i=4}^m S(X_i | X_{i+1})_{\tilde{\rho}} \\ &\vdots \\ &= S(X_1 | X_2)_{\tilde{\rho}} + S(X_2 \dots X_{m-1} | X_m)_{\tilde{\rho}} + S(X_m | X_1)_{\tilde{\rho}} \end{aligned}$$

$$\begin{aligned}
& + \sum_{i=3}^{m-1} I(X_2 \dots X_{i-1} : X_{i+1} | X_i)_{\tilde{\rho}} \\
& = S(X_2 \dots X_{m-1} | X_m)_{\tilde{\rho}} - S(X_2)_{\tilde{\rho}} + S(X_m X_1 X_2)_{\tilde{\rho}} \\
& \quad + I(X_m : X_2 | X_1)_{\tilde{\rho}} + \sum_{i=3}^{m-1} I(X_2 \dots X_{i-1} : X_{i+1} | X_i)_{\tilde{\rho}}.
\end{aligned}$$

By using the subadditivity $S(X_2 X_m) \leq S(X_2) + S(X_m)$, we obtain that

$$\begin{aligned}
(29) & \leq I(X_1 : X_3 \dots X_{m-1} | X_2 X_m)_{\tilde{\rho}} + I(X_m : X_2 | X_1)_{\tilde{\rho}} \\
& \quad + \sum_{i=3}^{m-1} I(X_2 \dots X_{i-1} : X_{i+1} | X_i)_{\tilde{\rho}}.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
S\left(\tilde{\rho}_X \left\| e^{-H_X^{\tilde{\rho}}}\right.\right) & \leq I(X_m : X_2 | X_1)_{\tilde{\rho}} + \sum_{i=2}^{m-2} I(X_2 \dots X_i : X_{i+2} | X_{i+1})_{\tilde{\rho}} \\
& \quad + I(A : B_2 | B_1)_{\tilde{\rho}} + I(A : C | B_1 B_2)_{\tilde{\rho}}, \tag{30}
\end{aligned}$$

where we used the chain rule (19). The first three terms only depend on marginals of $\tilde{\rho}_{ABC}$ on AB or BC . Since $\tilde{\rho}_{ABC}$ is close to $\tilde{\rho}'_{ABC}$ (25), $\tilde{\rho}_{ABC}$ also has marginals close to ρ_{ABC} on these regions:

$$\|\tilde{\rho}_{AB} - \rho_{AB}\|_1, \|\tilde{\rho}_{BC} - \rho_{BC}\|_1 \leq 8\sqrt{\varepsilon} + 2^{-N} \equiv \delta.$$

Thus, as in Eq. (26), the Fannes inequality implies that

$$\begin{aligned}
& I(X_m : X_2 | X_1)_{\tilde{\rho}} + \sum_{i=2}^{m-2} I(X_2 \dots X_i : X_{i+2} | X_{i+1})_{\tilde{\rho}} \\
& \leq I(X_m : X_2 | X_1)_{\rho} + \sum_{i=2}^{m-2} I(X_2 \dots X_i : X_{i+2} | X_{i+1})_{\rho} \\
& \quad + 12 \left(|X_2| + \sum_{i=2}^{m-2} |X_{i+2}| \right) \sqrt{\delta} \\
& \leq (m-1)\varepsilon + 12(m-2)|X_1| \sqrt{\delta}.
\end{aligned}$$

Combining with Eq. (27), we obtain

$$\begin{aligned}
S\left(\tilde{\rho}_X \left\| e^{-H_X^{\tilde{\rho}}}\right.\right) & \leq (m-1)\varepsilon + 12(m-2)|X_1| \sqrt{\delta} + 6\sqrt{6}|X_1| \varepsilon^{\frac{1}{4}} + \frac{12|X_1|}{\sqrt{2^N}} \\
& \leq \mathcal{O}\left(N\sqrt{\delta}\right),
\end{aligned}$$

where we used $m|X_1| = \Theta(N)$ and the asymptotic notation $\mathcal{O}(f(N, \varepsilon))$ as $N \rightarrow \infty$ and $\varepsilon \rightarrow 0$ (and therefore $\delta \rightarrow 0$).

Let us estimate the partition function $Z = \text{Tr} e^{-H_X^{\tilde{\rho}}}$. By Pinsker inequality, the above bound implies

$$\begin{aligned} \left| \text{Tr} \left(e^{-H_X^{\tilde{\rho}}} \right) - 1 \right| &\leq \left\| \tilde{\rho}_X - e^{-H_X^{\tilde{\rho}}} \right\|_1 \\ &\leq \sqrt{2S \left(\tilde{\rho}_X \left\| e^{-H_X^{\tilde{\rho}}} \right. \right)} \\ &\leq \mathcal{O} \left(N^{\frac{1}{2}} \delta^{\frac{1}{4}} \right) \end{aligned} \quad (31)$$

where we used the inequality $\| \|A\|_1 - \|B\|_1 \leq \|A - B\|_1$. For given N , we assume δ is sufficiently small so that Eq. (31) is smaller than 1. We then obtain

$$|\log Z| \leq \mathcal{O} \left(N^{\frac{1}{2}} \delta^{\frac{1}{4}} \right).$$

Thus, the difference between $\tilde{\rho}_X$ and $\tilde{\pi}_X$ is bounded as

$$\begin{aligned} \|\tilde{\rho}_X - \tilde{\pi}_X\|_1 &\leq \sqrt{2S(\tilde{\rho}_X \|\tilde{\pi}_X)} \\ &= \sqrt{2 \left(S(\tilde{\rho}_X \| e^{-H_X^{\tilde{\rho}}}) - \log Z \right)} \\ &\leq \mathcal{O} \left(N^{\frac{1}{2}} \delta^{\frac{1}{8}} \right). \end{aligned} \quad (32)$$

Again, by the Fannes inequality, the conditional mutual information of $\tilde{\pi}_X$ is bounded as

$$I(A : C|B)_{\tilde{\pi}} \leq \mathcal{O} \left(|X_1| N^{\frac{1}{4}} \delta^{\frac{1}{16}} \right).$$

The marginal of $\tilde{\pi}$ on AB satisfy

$$\begin{aligned} \|\rho_{AB} - \tilde{\pi}_{AB}\|_1 &\leq \|\rho_{AB} - \tilde{\rho}'_{AB}\|_1 + \|\tilde{\rho}'_{AB} - \tilde{\rho}_{AB}\|_1 + \|\tilde{\rho}_{AB} - \tilde{\pi}_{AB}\|_1 \\ &\leq \mathcal{O} \left(N^{\frac{1}{2}} \delta^{\frac{1}{8}} \right). \end{aligned} \quad (33)$$

Here, we used Eqs. (22), (25) and (32) in the second inequality. In the same way, we also obtain

$$\|\rho_{BC} - \tilde{\pi}_{BC}\|_1 \leq \mathcal{O}(N^{\frac{1}{2}} \delta^{\frac{1}{8}}). \quad (34)$$

Finally, we show that $I(A : C|B)_\rho$ approximates the distance between ρ and the set of Gibbs states E_{nn}^K in terms of the relative entropy. By combining Eqs. (33) and (34) with the Fannes inequality, we have that

$$\begin{aligned} &\left| I(A : C|B)_\rho - \left(S(ABC)_{\tilde{\pi}} - S(ABC)_\rho \right) \right| \\ &\leq \left| S(A|B)_\rho + S(BC)_\rho - S(ABC)_{\tilde{\pi}} \right| \\ &\leq \left| S(A|B)_\rho - S(A|B)_{\tilde{\pi}} \right| + \left| S(BC)_\rho - S(BC)_{\tilde{\pi}} \right| + I(A : C|B)_{\tilde{\pi}} \\ &\leq \mathcal{O} \left(N^{\frac{3}{4}} \delta^{\frac{1}{16}} \right). \end{aligned} \quad (35)$$

By definition, it holds that

$$\min_{\mu \in E_{nn}^K} S(\rho_{ABC} \|\mu_{ABC}) = S(ABC)_\rho - \min_{\mu \in E_{nn}^K} \text{Tr}(\rho \log \mu).$$

Here, $\log \mu \propto H_{AB} + H_{BC}$ for some bounded Hermitian operators H_{AB} and H_{BC} satisfying $\|H_{AB}\| + \|H_{BC}\| \leq \mathcal{O}(mK)$. Equations (33) and (34) yield that

$$\begin{aligned} |\mathrm{Tr}(\rho_Y O_Y) - \mathrm{Tr}(\tilde{\pi}_Y O_Y)| &\leq \|O\| \|\rho_Y - \tilde{\pi}_Y\| \\ &\leq \mathcal{O}\left(\|O\| N^{\frac{1}{2}} \delta^{\frac{1}{8}}\right) \end{aligned} \quad (36)$$

for $Y = AB, BC$. Therefore, we can approximate $\mathrm{Tr}(\rho \log \mu)$ by $\mathrm{Tr}(\tilde{\pi} \log \mu)$, and we have

$$\begin{aligned} \min_{\mu \in E_{mn}^K} S(\rho_{ABC} \|\mu_{ABC}) &= S(ABC)_\rho - \min_{\mu \in E_{mn}^K} \mathrm{Tr}(\tilde{\pi} \log \mu) + \mathcal{O}\left(mK N^{\frac{1}{2}} \delta^{\frac{1}{8}}\right) \\ &= S(ABC)_{\tilde{\pi}} - \min_{\mu \in E_{mn}^K} \mathrm{Tr}(\tilde{\pi} \log \mu) + I(A : C|B)_\rho \\ &\quad + \mathcal{O}\left(N^{\frac{5}{4}} \delta^{\frac{1}{16}}\right) + \mathcal{O}\left(N^{\frac{5}{2}} \delta^{\frac{1}{8}}\right) \\ &= \min_{\mu \in \mathcal{E}_{mn}^K} S(\tilde{\pi}_{ABC} \|\mu_{ABC}) + I(A : C|B)_\rho + \mathcal{O}\left(N^{\frac{5}{2}} \delta^{\frac{1}{16}}\right) \\ &= I(A : C|B)_\rho + \mathcal{O}\left(N^{\frac{5}{2}} \delta^{\frac{1}{16}}\right), \end{aligned}$$

where we used $mK = \Theta(N^2)$ and Eq. (35) in the secondline. The third inequality follows from Eq. (35) and the last line follows from $\tilde{\pi} \in E_{mn}^K$. \square

4. 1D Quantum Gibbs states are Quantum Approximate Markov Chain

In this section, we provide a proof of Theorem 4, Corollaries 5 and 6. A key point of the proof is that if a short-range Hamiltonian changes locally, the corresponding Gibbs state also changes quasi-locally. To obtain operators representing changes of the Gibbs state, we employ quantum belief propagation equations which have been studied in Refs. [23,24]. We first introduce these technical tools, and then show the proofs. We discuss another approach to prove Theorem 4 based on the results of Ref. [25] in Appendix A.

4.1. Perturbative analysis of Gibbs states. Let us consider an one-parameter family of a Hamiltonian H on a spin lattice with a perturbation operator V

$$H(s) = H + sV, \quad (37)$$

where $s \in [0, 1]$. The change of the corresponding Gibbs state due to a small change of s can be computed through a quantum belief propagation equation [23,24]:

$$\frac{d}{ds} e^{-\beta H(s)} = -\frac{\beta}{2} \left\{ e^{-\beta H(s)}, \Phi_\beta^{H(s)}(V) \right\}, \quad (38)$$

where the operator $\Phi_\beta^{H(s)}(V)$ is given by [24]

$$\Phi_\beta^{H(s)}(V)_{ij} := V_{ij} \tilde{f}_\beta(E_i(s) - E_j(s)) \quad (39)$$

in the energy eigenbasis of $H(s) = \sum_i E_i(s)|i\rangle\langle i|$,⁷ with $\tilde{f}_\beta(\omega) = \frac{\tanh(\beta\omega/2)}{\beta\omega/2}$. Using the Fourier transform $f_\beta(t) = \frac{1}{2\pi} \int d\omega \tilde{f}_\beta(\omega)e^{i\omega t}$, $\Phi_\beta^{H(s)}(V)$ can be written as the integral form:

$$\Phi_\beta^{H(s)}(V) = \int_{-\infty}^{\infty} dt f_\beta(t) e^{-iH(s)t} V e^{iH(s)t}. \quad (40)$$

Taking the formal integration of Eq. (38), we obtain

$$e^{-\beta H(1)} = O e^{-\beta H(0)} O^\dagger, \quad (41)$$

where the operator O is defined as

$$\begin{aligned} O &= \mathcal{T} \exp \left[-\frac{\beta}{2} \int_0^1 ds' \Phi_\beta^{H(s')} (V) \right] \\ &= \sum_{n=0}^{\infty} \left(-\frac{\beta}{2} \right)^n \int_0^1 ds'_1 \int_0^{s'_1} ds'_2 \dots \int_0^{s'_{n-1}} ds'_n \Phi_\beta^{H(s'_n)} (V) \dots \Phi_\beta^{H(s'_1)} (V), \end{aligned}$$

with \mathcal{T} the time-ordering operation. Since we have $dt f_\beta(t) = \frac{dt}{\beta} f_1(\frac{t}{\beta})$, it holds that

$$\begin{aligned} \left\| \Phi_\beta^{H(s)}(V) \right\| &= \left\| \int_{-\infty}^{\infty} dt' f_1(t') e^{-i\beta H(s)t'} V e^{i\beta H(s)t'} \right\| \\ &\leq \|V\| \left| \int_{-\infty}^{\infty} dt' f_1(t') \right| \\ &= \|V\|. \end{aligned} \quad (42)$$

The integral in the last equality can be calculated through the series expansion:

$$\frac{\tanh(x)}{x} = \sum_{k=0}^{\infty} \frac{2}{x^2 + (k + \frac{1}{2})^2 \pi^2}. \quad (43)$$

The upper bound of $\|\Phi_\beta^{H(s)}(V)\|$ implies the upper bound of $\|O\|$ that is given by

$$\|O\| \leq e^{\frac{\beta}{2}\|V\|}. \quad (44)$$

When the Hamiltonian defined on a many-body system is short-ranged, time-evolutions of a local operator is restricted by using the Lieb-Robinson bound [26]. Suppose that H is a Hamiltonian obeying the Lieb-Robinson bound, and O_A and O_B are observables supported on local regions A and B , respectively. Then, the Lieb-Robinson bound for these operators is formulated as

$$\left\| \left[O_A, e^{-iHt} O_B e^{iHt} \right] \right\| \leq c \|O_A\| \|O_B\| \min(|A|, |B|) e^{c'(vt - d(A,B))}, \quad (45)$$

where $c, v \geq 0$, $c' > 0$ are constants. Assume that $H(0)$ is a short-range Hamiltonian and $\text{supp}(V)$ is a simply-connected local region. Then $H(s)$ obeys the Lieb-Robinson bound for all $s \in [0, 1]$. Since $f_\beta(t)$ decays fast in $|t|$, the Lieb-Robinson bound implies that the effective support of $\Phi_\beta^{H(s)}(V)$ can be restricted to a local region \mathcal{V}_l , which

⁷ Each $|i\rangle$ depends on s as well as the eigenvalues $E_i(s)$.

contains all sites within distance l from $\text{supp}(V)$. More precisely, there exist positive constants c' and v , which is determined by $H(s)$, such that [24]

$$\left\| \Phi_{\beta}^{H(s)}(V) - \text{Tr}_{\mathcal{V}_l^c} \left[\Phi_{\beta}^{H(s)}(V) \right] \otimes \frac{1}{d_{\mathcal{V}_l^c}} \mathbb{1}_{\mathcal{V}_l^c} \right\| \leq c' \|V\| e^{-\frac{c'l}{1+c'v\beta/\pi}}, \quad (46)$$

where $d_{\mathcal{V}_l^c} := \dim \mathcal{H}_{\mathcal{V}_l^c}$. We also define the integral of the restricted operator in Eq. (46) as

$$O_{\mathcal{V}_l} := \mathcal{T} \exp \left(-\frac{\beta}{2} \int_0^1 ds' \text{Tr}_{\mathcal{V}_l^c} \left[\Phi_{\beta}^{H(s')} (V) \right] \otimes \frac{1}{d_{\mathcal{V}_l^c}} \mathbb{1}_{\mathcal{V}_l^c} \right). \quad (47)$$

This operator is also localized on \mathcal{V}_l and approximates O with good accuracy. Let us choose c' and v so that Eq. (46) holds for all $s \in [0, 1]$. Then, we obtain that

$$\|O - O_{\mathcal{V}_l}\| \leq \frac{c'\beta\|V\|}{2} e^{\frac{(1+c')\beta\|V\|}{2}} e^{-\frac{c'l}{1+c'v\beta/\pi}}. \quad (48)$$

To see this, consider some parametrized operators $Q(s)$ and $\tilde{Q}(s)$ satisfying $\|Q(s)\|, \|\tilde{Q}(s)\| \leq C$ and $\|Q(s) - \tilde{Q}(s)\| \leq \Delta$ for all $s \in [0, 1]$. From the simple calculation, we obtain

$$\begin{aligned} Q(s_n)Q(s_{n-1}) \cdots Q(s_1) &= \tilde{Q}(s_n)\tilde{Q}(s_{n-1}) \cdots \tilde{Q}(s_1) \\ &\quad + \sum_{j=1}^n Q(s_n) \cdots Q(s_{j+1}) \Delta_j \tilde{Q}(s_{j-1}) \cdots \tilde{Q}(s_1) \end{aligned}$$

where $\Delta_j = Q(s_j) - \tilde{Q}(s_j)$ and $\|\Delta_j\| \leq \Delta$. Therefore, we obtain that

$$\|Q(s_n)Q(s_{n-1}) \cdots Q(s_1) - \tilde{Q}(s_n)\tilde{Q}(s_{n-1}) \cdots \tilde{Q}(s_1)\| \leq nC^{n-1}\Delta. \quad (49)$$

In our case, $Q(s) = \Phi_{\beta}^{H(s)}(V)$ and $\tilde{Q}(s) = \text{Tr}_{\mathcal{V}_l^c} \left[\Phi_{\beta}^{H(s)}(V) \right] \otimes \frac{1}{d_{\mathcal{V}_l^c}} \mathbb{1}_{\mathcal{V}_l^c}$. We can choose $\Delta = c' e^{-\frac{c'l}{1+c'v\beta/\pi}} \|V\|$. From Eqs. (42) and (46), their norms can be bounded as

$$\begin{aligned} \|Q(s)\|, \|\tilde{Q}(s)\| &\leq \|Q(s)\| + \|\tilde{Q}(s) - Q(s)\| \\ &\leq \left(1 + c' e^{-\frac{c'l}{1+c'v\beta/\pi}} \right) \|V\| \\ &\leq (1 + c') \|V\|. \end{aligned}$$

Therefore, Eq. (49) holds for $C = (1 + c') \|V\|$. By inserting Eq. (49) to the definition of O , we obtain Eq. (48).

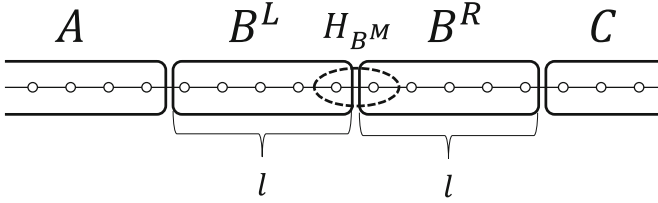


Fig. 2. A schematic picture of the definition of H_{B^M} . We divide B into two halves B^L and B^R . In the case of a nearest-neighbor Hamiltonian, H_{B^M} is the interaction term acting on both B^L and B^R (the dotted circle)

4.2. *Proof of Theorem 4.* For the convenience of the reader, we restate Theorem 4 below.

Theorem 4. *Let $H = \sum_i h_i$ be a short-range 1D Hamiltonian with $\|h_i\| \leq 1$ and $l_0, C, c > 0$ be universal constants. For an inverse temperature $\beta > 0$ and any partition ABC with $d(A, C) \geq l_0$, there exists a CPTP-map $\Lambda_{B \rightarrow BC} : \mathcal{D}(\mathcal{H}_B) \rightarrow \mathcal{D}(\mathcal{H}_B \otimes \mathcal{H}_C)$ such that*

$$\left\| \rho^{H_{ABC}} - \Lambda_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}}) \right\|_1 \leq e^{-q(\beta)\sqrt{d(A,C)}},$$

where $q(\beta) = ce^{-c'\beta}$ if the correlation length of $\rho^{H_{ABC}}$ is $\xi = e^{(C)\beta}$.

The proof of Theorem 4 consists of three steps. In the first step, we show that there exists a CP-map which recovers the Gibbs state from the reduced state with exponentially good accuracy (Lemma 7). In the second, we normalize the CP-map to make it trace non-increasing i.e. we show the existence of a quantum operation which succeed to recover the Gibbs state with some probability (Lemma 9). We also show that the success probability is a constant of system size. Finally, we construct a CPTP-map from the probabilistic operation by employing a repeat-until-success strategy in the third step.

Let us begin with the following lemma.

Lemma 7. *For any 1D Gibbs state $\rho^{H_{ABC}}$ of a short-range Hamiltonian on a system with a partition ABC such that $l := d(A, C)/2 > r$, there exists a CP map $\kappa_{B \rightarrow BC}$, non-negative constants c' and v such that*

$$\left\| \rho^{H_{ABC}} - \kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}}) \right\|_1 \leq C_1(\beta)e^{-q_1(\beta)l}, \quad (50)$$

where $C_1(\beta)$ is a non-negative constant and $q_1(\beta) = \frac{c'}{1+c'v\beta/\pi}$, as defined in Eq. (46).

Proof. Let us consider a short-range Hamiltonian $H = \sum_i h_i$ with the range r . Without loss of generality, we introduce a tripartition ABC of a 1D system so that each subsystem is simply connected and $d(A, C)$ is chosen to be $2l$ for some integer $l > r$. We then split region B into the left half B^L , which touches A , and the right half B^R which touches C (Fig. 2). We denote the sum of the all interactions h_j acting on both B^L and B^R by $H_{B^M} = \sum_{j:i \in \text{supp}(h_j)} h_j$. By assumption, $\|H_{B^M}\| \leq J$ for some constant $J \geq 0$.

Remark 8. When B consists of a fixed number of simply connected regions, each connected component neighboring both A and C is divided into two halves as in the same way. Then, H_{B^M} is the sum of all interaction terms acting on both such divided regions.

We apply the technical tools discussed in the previous section to a parametrized Hamiltonian $H_{ABC}(s) = H_{A^L B^L} + H_{B^R C} + sH_{B^M}$. Here, H_{B^M} corresponds to the perturbation operator V in the previous section. Then, we introduce an operator O_{ABC} defined as

$$O_{ABC} := \mathcal{T}\text{Exp} \left[-\frac{\beta}{2} \int_0^1 ds' \Phi_{\beta}^{H_{ABC}(s')} (H_{BM}) \right].$$

Equations (41) implies that

$$e^{-\beta H_{ABC}} = O_{ABC} e^{-\beta(H_{ABL}+H_{BRC})} O_{ABC}^{\dagger}.$$

We also introduce the inverse operator \tilde{O}_{ABC} , that is given by

$$\tilde{O}_{ABC} := \tilde{\mathcal{T}}\text{Exp} \left[\frac{\beta}{2} \int_0^1 ds' \Phi_{\beta}^{H_{ABC}(s')} (H_{BM}) \right],$$

where $\tilde{\mathcal{T}}$ denotes the inverse time-ordering operator. O_{ABC} and \tilde{O}_{ABC} satisfies the following relation (see e.g., [27])

$$O_{ABC} \tilde{O}_{ABC} = \tilde{O}_{ABC} O_{ABC} = \mathbf{1}_{ABC}.$$

From Eq. (44), the operator norms of O_{ABC} and \tilde{O}_{ABC} can be bounded as

$$\|O_{ABC}\|, \|\tilde{O}_{ABC}\| \leq e^{\frac{\beta}{2} \|H_{BM}\|} = e^{\frac{\beta}{2} J}.$$

Importantly, the upper bound is independent of size of A , B and C . From the definitions of O_{ABC} and its inverse, it is not difficult to see that

$$\rho^{H_{ABC}} = O_{ABC} \left[\text{Tr}_{BRC} \left(\tilde{O}_{ABC} \rho^{H_{ABC}} \tilde{O}_{ABC}^{\dagger} \right) \otimes \rho_{BRC}^{H_{BRC}} \right] O_{ABC}^{\dagger}. \quad (51)$$

From Eq. (48), we know that there exist operators O_B and \tilde{O}_B whose supports are restricted on B . For simplicity, let us denote

$$K(\beta) = \frac{c' \beta J}{2} e^{\frac{(1+c')\beta J}{2}}.$$

Then, O_B and \tilde{O}_B satisfy

$$\|O_{ABC} - O_B\|, \|\tilde{O}_{ABC} - \tilde{O}_B\| \leq K(\beta) e^{-q_1(\beta)l}, \quad (52)$$

where $q_1(\beta) = \frac{c'}{1+c'v\beta/\pi}$ for non-negative constants c' and v which are chosen as in Eq. (48). Then, the operator norm of the local operators O_B and \tilde{O}_B can be bounded by

$$\|O_B\|, \|\tilde{O}_B\| \leq \|O_{ABC}\| + \|O_{ABC} - O_B\| \leq e^{\frac{\beta}{2} J} + K(\beta), \quad (53)$$

which is independent of the size of B . Let $\tilde{O}_{B|B}$ be the non-trivial part of \tilde{O}_B acting on B , i.e.,

$$\tilde{O}_B = \tilde{O}_{B|B} \otimes \mathbf{1}_{AC}.$$

By using this notation, we define a CP-map $\kappa_{B \rightarrow BC}$ by replacing $O_{ABC}(\tilde{O}_{ABC})$ by local operators $O_B(\tilde{O}_B)$ and removing partial trace over C in Eq. (51), i.e.,

$$\kappa_{B \rightarrow BC}(\sigma_B) \equiv O_B \left[\text{Tr}_{B^R} \left(\tilde{O}_{B|B} \sigma_B \tilde{O}_{B|B}^{\dagger} \right) \otimes \rho_{B^R C}^{H_{B^R C}} \right] O_B^{\dagger}. \quad (54)$$

Let us denote

$$X_1 = \text{Tr}_{B^R C} \left(\tilde{O}_{ABC} \rho^{H_{ABC}} \tilde{O}_{ABC}^\dagger \right) \otimes \rho_{B^R C}^{H_{B^R C}},$$

and

$$X_2 = \text{Tr}_{B^R} \left(\tilde{O}_{B|B} \rho_{AB}^{H_{ABC}} \tilde{O}_{B|B}^\dagger \right) \otimes \rho_{B^R C}^{H_{B^R C}}.$$

We have

$$\begin{aligned} \|X_1 - X_2\|_1 &= \left\| \text{Tr}_{B^R C} \left(\tilde{O}_{ABC} \rho^{H_{ABC}} \tilde{O}_{ABC}^\dagger \right) \otimes \rho_{B^R C}^{H_{B^R C}} \right. \\ &\quad \left. - \text{Tr}_{B^R} \left(\tilde{O}_{B|B} \rho_{AB}^{H_{ABC}} \tilde{O}_{B|B}^\dagger \right) \otimes \rho_{B^R C}^{H_{B^R C}} \right\|_1 \\ &= \left\| \text{Tr}_{B^R C} \left(\tilde{O}_{ABC} \rho^{H_{ABC}} \tilde{O}_{ABC}^\dagger \right) - \text{Tr}_{B^R} \left(\tilde{O}_{B|B} \rho_{AB}^{H_{ABC}} \tilde{O}_{B|B}^\dagger \right) \right\|_1 \\ &= \left\| \text{Tr}_{B^R C} \left(\tilde{O}_{ABC} \rho^{H_{ABC}} \tilde{O}_{ABC}^\dagger \right) - \text{Tr}_{B^R C} \left(\tilde{O}_B \rho^{H_{ABC}} \tilde{O}_B^\dagger \right) \right\|_1 \\ &\leq \left\| \tilde{O}_{ABC} \rho^{H_{ABC}} \tilde{O}_{ABC}^\dagger - \tilde{O}_B \rho^{H_{ABC}} \tilde{O}_B^\dagger \right\|_1 \\ &\leq \left\| (\tilde{O}_{ABC} - \tilde{O}_B) \rho^{H_{ABC}} \tilde{O}_{ABC}^\dagger \right\|_1 + \left\| \tilde{O}_B \rho^{H_{ABC}} (\tilde{O}_{ABC}^\dagger - \tilde{O}_B^\dagger) \right\|_1. \end{aligned} \quad (55)$$

We used the monotonicity of the trace-norm in the last inequality. To address the calculation, we use the following special case of the Hölder's inequality:

$$\|AB\|_1 \leq \|A\|_1 \|B\|. \quad (56)$$

It implies that

$$\begin{aligned} (55) &\leq \left\| (\tilde{O}_{ABC} - \tilde{O}_B) \right\| \left\| \rho^{H_{ABC}} \tilde{O}_{ABC}^\dagger \right\|_1 + \left\| \tilde{O}_B \rho^{H_{ABC}} \right\|_1 \left\| (\tilde{O}_{ABC}^\dagger - \tilde{O}_B^\dagger) \right\| \\ &\leq \left\| (\tilde{O}_{ABC} - \tilde{O}_B) \right\| \left\| \tilde{O}_{ABC}^\dagger \right\| + \left\| \tilde{O}_B \right\| \left\| (\tilde{O}_{ABC}^\dagger - \tilde{O}_B^\dagger) \right\| \\ &\leq 2K(\beta) \left(e^{\frac{\beta J}{2}} + K(\beta) \right) e^{-q_1(\beta)l}. \end{aligned}$$

The first and second lines follow from Eq. (56) and $\|\rho^{H_{ABC}}\|_1 = 1$. In the last line, we used Eqs. (52) and (53).

By using the above bound, we bound the difference between the original Gibbs state $\rho^{H_{ABC}}$ and $\kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})$ as

$$\begin{aligned} &\left\| \rho^{H_{ABC}} - \kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}}) \right\|_1 \\ &= \left\| O_{ABC} X_1 O_{ABC}^\dagger - O_B X_2 O_B^\dagger \right\|_1 \\ &\leq \left\| O_{ABC} X_1 O_{ABC}^\dagger - O_B X_1 O_B^\dagger \right\|_1 + \left\| O_B (X_1 - X_2) O_B^\dagger \right\|_1 \\ &\leq \left\| (O_{ABC} - O_B) X_1 O_{ABC}^\dagger \right\|_1 + \left\| O_B X_1 (O_{ABC}^\dagger - O_B^\dagger) \right\|_1 + \|(X_1 - X_2)\|_1 \left\| O_B^\dagger \right\|^2 \\ &\leq \|O_{ABC} - O_B\| \|\tilde{O}_{ABC}\|^2 \|O_{ABC}\| + \|O_{ABC}^\dagger - O_B^\dagger\| \|\tilde{O}_{ABC}\|^2 \|O_B\| \\ &\quad + \|(X_1 - X_2)\|_1 \left\| O_B^\dagger \right\|^2 \end{aligned}$$

$$\begin{aligned} &\leq K(\beta) \left(e^{\frac{3\beta J}{2}} + e^{\beta J} \left(e^{\frac{\beta J}{2}} + K(\beta) \right) + 2 \left(e^{\frac{\beta J}{2}} + K(\beta) \right)^3 \right) e^{-q_1(\beta)l} \\ &\leq 4K(\beta) \left(e^{\frac{\beta J}{2}} + K(\beta) \right)^3 e^{-q_1(\beta)l}. \end{aligned}$$

Here we used the fact that $\|X_1\|_1 \leq \|\rho^{H_{ABC}}\|_1 \|\tilde{O}_{ABC}\|^2 = \|\tilde{O}_{ABC}\|^2$ in the fourth line.

Choosing $C_1(\beta) = 4K(\beta) \left(e^{\frac{\beta J}{2}} + K(\beta) \right)^3$ completes the proof. \square

Unfortunately, the map $\kappa_{B \rightarrow BC}$ is not a trace non-increasing map in general. Thus, we normalize $\kappa_{B \rightarrow BC}$ to obtain a CP and trace non-increasing map, which corresponds to a probabilistic process.

Lemma 9. *Under the setting of Lemma 7, there exists a CP and trace non-increasing map $\tilde{\Lambda}_{B \rightarrow BC}$ for any $l \geq l_0(\beta) \equiv \left\lceil \frac{\log C_1(\beta)+1}{q_1(\beta)} \right\rceil = \mathcal{O}(\beta^2)$ such that*

$$\left\| \rho^{H_{ABC}} - \frac{\tilde{\Lambda}_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})}{\text{Tr}[\tilde{\Lambda}_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})]} \right\|_1 \leq C_2(\beta) e^{-q_1(\beta)l}, \quad (57)$$

where $C_2(\beta) = \frac{2C_1(\beta)}{(1-e^{-1})}$. Moreover, $p = \text{Tr}[\tilde{\Lambda}_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})]$ is a strictly positive constant which is independent of the size of subsystems A , B and C .

Proof. We denote the maximum eigenvalue of $O_B^\dagger O_B$ ($\tilde{O}_B^\dagger \tilde{O}_B$) by $\lambda_{\max}^{O_B}$ ($\lambda_{\max}^{\tilde{O}_B}$). From Eq. (53) and inequality $\|A^\dagger A\| \leq \|A\|^2$, these eigenvalues are bounded as

$$\lambda_{\max}^{O_B}, \lambda_{\max}^{\tilde{O}_B} \leq \left(e^{\frac{\beta J}{2}} + K(\beta) \right)^2. \quad (58)$$

Define $\lambda_{\max} := \lambda_{\max}^{O_B} \lambda_{\max}^{\tilde{O}_B}$. Then, we define the normalized map $\tilde{\Lambda}_{B \rightarrow BC}$ as

$$\tilde{\Lambda}_{B \rightarrow BC}(\sigma_B) := \frac{1}{\lambda_{\max}} \kappa_{B \rightarrow BC}(\sigma_B).$$

By definition, $\tilde{\Lambda}_{B \rightarrow BC}$ is CP and trace non-increasing. After the normalization, the output state for the input $\rho_{AB}^{H_{ABC}}$ is

$$\frac{\tilde{\Lambda}_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})}{\text{Tr}[\tilde{\Lambda}_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})]}.$$

Let us introduce $l_0(\beta) = \frac{\ln C_1(\beta)+1}{q_1(\beta)}$. For any $l \geq l_0(\beta)$, $C_1(\beta) e^{-q_1(\beta)l} \leq e^{-1} < 1$.

For such l , the probability p for the input $\rho_{AB}^{H_{ABC}}$ is then estimated as

$$\begin{aligned} p &= \text{Tr}[\tilde{\Lambda}_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})] \\ &= \frac{1}{\lambda_{\max}} \left\| \kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}}) \right\|_1 \\ &\geq \frac{1}{\lambda_{\max}} \left| \|\rho^{H_{ABC}}\|_1 - \left\| \rho^{H_{ABC}} - \kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}}) \right\|_1 \right| \\ &\geq \frac{1}{\lambda_{\max}} \left(1 - C_1(\beta) e^{-q_1(\beta)l} \right) \end{aligned}$$

$$\begin{aligned}
&\geq \frac{1}{\lambda_{\max}} \left(1 - \frac{1}{e}\right) \\
&\geq \frac{1 - e^{-1}}{\left(e^{\frac{\beta J}{2}} + K(\beta)\right)^4} \\
&> 0,
\end{aligned} \tag{59}$$

where we used Eq. (58) in the line before the last.

The approximation error of the output is then estimated as

$$\begin{aligned}
&\left\| \rho^{H_{ABC}} - \frac{\tilde{\Lambda}_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})}{\text{Tr}[\tilde{\Lambda}_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})]} \right\|_1 \\
&= \left\| \rho^{H_{ABC}} - \frac{\kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})}{\|\kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})\|_1} \right\|_1 \\
&\leq \left| 1 - \frac{1}{\|\kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})\|_1} \right| \|\rho^{H_{ABC}}\|_1 \\
&\quad + \frac{1}{\|\kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})\|_1} \left\| \rho^{H_{ABC}} - \kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}}) \right\|_1 \\
&\leq \frac{C_1(\beta)e^{-q_1(\beta)l}}{1 - e^{-1}} + \frac{1}{1 - e^{-1}} C_1(\beta)e^{-q_1(\beta)l} \\
&\leq \frac{2C_1(\beta)}{1 - e^{-1}} e^{-q_1(\beta)l}.
\end{aligned}$$

In the third line, we used the fact

$$\left| \|\kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})\|_1 - 1 \right| \leq C_1(\beta)e^{-q_1(\beta)l},$$

which follows from

$$1 - C_1(\beta)e^{-q_1(\beta)l} \leq \|\kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})\|_1 \leq 1 + C_1(\beta)e^{-q_1(\beta)l}.$$

Thus, we conclude that Lemma 9 holds by choosing $C_2(\beta) = \frac{2C_1(\beta)}{1 - e^{-1}}$. \square

We are now in position to prove Theorem 1. Without loss of generality, let us assume that $d(A, C) = |B| = 3l^2 - l$ for $l \in \mathbb{N}$. We divide B into $B = B_l \bar{B}_{l-1} B_{l-1} \dots \bar{B}_1 B_1$ as shown in Fig. 3, where for each i , $|B_i| = 2l$ and $|\bar{B}_i| = l$. From Lemma 9, there exists a CP and trace non-increasing map $\tilde{\Lambda}_{B_i \rightarrow B_i \bar{B}_{i-1} \dots B_1 C}$ for each i which approximately recovers $\rho^{H_{ABC}}$ from the reduced state on $AB_l \dots \bar{B}_i B_i$ (here, we choose B_i as “ B ” and $B_{i-1} \dots B_1 C$ as “ C ” in the lemma). We then introduce a CP and trace non-increasing map $\tilde{E}_{B_i \rightarrow B_i \bar{B}_{i-1} \dots B_1 C}$ for each $\tilde{\Lambda}_{B_i \rightarrow B_i \bar{B}_{i-1} \dots B_1 C}$ so that $\{\tilde{\Lambda}_{B_i \rightarrow B_i \bar{B}_{i-1} \dots B_1 C}, \tilde{E}_{B_i \rightarrow B_i \bar{B}_{i-1} \dots B_1 C}\}$ forms a quantum instrument.⁸ We simply denote $\tilde{\Lambda}_{B_i \rightarrow B_i \bar{B}_{i-1} \dots B_1 C}$ by $\tilde{\Lambda}_i$ and its complementary map $\tilde{E}_{B_i \rightarrow B_i \bar{B}_{i-1} \dots B_1 C}$ by \tilde{E}_i . We also denote the tracing operation over

⁸ For instance, $\tilde{E}_{B \rightarrow BC}$ can be chosen to be $\tilde{E}_{B \rightarrow BC}(\sigma_B) = \sqrt{\mathbf{I} - \overline{M_B}} \sigma_B \sqrt{\mathbf{I} - \overline{M_B}} \otimes |0\rangle\langle 0|_C$, where $M_B := \sum_i K_i^\dagger K_i \leq \mathbf{I}_B$ with $\{K_i\}$ the Kraus operators of $\tilde{\Lambda}_{B \rightarrow BC}$.

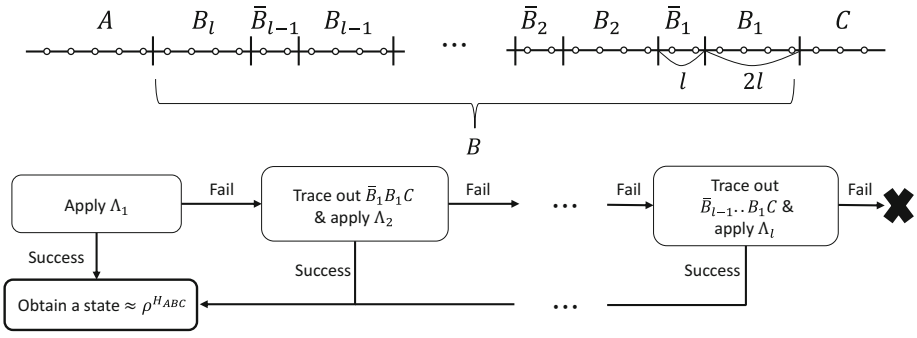


Fig. 3. A schematic picture of the repeat-until-success method. We introduced buffer systems $\{\bar{B}_i\}$ to suppress the effect of failure. The “failure” output at the end corresponds to the CP-map $\bar{E}_l \circ \text{Tr}_{l-1} \bar{E}_{l-1} \cdots \circ \text{Tr}_1 \bar{E}_1$

$\bar{B}_i B_i \dots B_1 C$ by Tr_i . We define a CPTP-map $\Lambda_{B \rightarrow BC}$ as

$$\begin{aligned} \Lambda_{B \rightarrow BC}(\sigma_B) &= \tilde{\Lambda}_1(\sigma_B) + \left(\tilde{\Lambda}_2 + \cdots \left(\tilde{\Lambda}_{l-1} + \left(\tilde{\Lambda}_l + \tilde{E}_l \right) \text{Tr}_{l-1} \tilde{E}_{l-1} \right) \cdots \text{Tr}_2 \tilde{E}_2 \right) \text{Tr}_1 \tilde{E}_1(\sigma_B) \end{aligned} \quad (60)$$

based on the repeat-until-success method (Fig. 3).

When we input $\rho_{AB}^{H_{ABC}}$ to $\Lambda_{B \rightarrow BC}$, the output of each map $\tilde{\Lambda}_i$ corresponds to the success of the recovery process at the i th step (with probability p_i) and \tilde{E}_i corresponds to the failure of the recovery process (with probability $1 - p_i$). If it fails, we trace out both the recovered system and the “buffer” system \bar{B}_i , and then, the effect of the failure can be almost neglected. Thus, we can repeat the probabilistic process to obtain the success outcome. The effect of the failure is estimated by the following lemma, which utilizes the exponential decay of correlation of 1D Gibbs states [25].

Lemma 10. *Under the setting of Lemma 7, there exists a constant $\xi \geq 0$ such that*

$$(1 - p_i) \left\| \text{Tr}_i \left(\rho^{H_{ABC}} \right) - \frac{\text{Tr}_i \tilde{E}_i \left(\rho_{AB_l \dots B_i}^{H_{ABC}} \right)}{1 - p_i} \right\|_1 \leq e^{-\frac{1}{\xi}}.$$

Proof. Define a correlation function $\text{Cor}(X : Y)_\rho$ of regions X and Y by

$$\text{Cor}(X : Y)_\rho = \max_{\|M\|, \|N\| \leq 1} |\text{Tr}[(M \otimes N)(\rho_{XY} - \rho_X \otimes \rho_Y)]|.$$

Consider some CP and trace-decreasing map $(\mathbf{1}_X \otimes \mathcal{E}_{Y \rightarrow Z})(\rho_{XY}) = \sum_i E_i \rho_{XY} E_i^\dagger$. By lemma 9 of Ref. [28], it holds that

$$\text{Cor}(X : Y)_\rho \geq \text{Tr}[M_Y \rho_{XY}] \| \rho_X - \sigma_X \|_1,$$

where $M_Y = \sum_i E_i^\dagger E_i$ and

$$\sigma_X = \frac{1}{\text{Tr}[M_Y \rho_{XY}]} \text{Tr}_Y [M_Y \rho_{XY}] = \text{Tr}_Y \frac{(\text{id}_X \otimes \mathcal{E}_{Y \rightarrow Z})(\rho_{XY})}{\text{Tr}(\text{id}_X \otimes \mathcal{E}_{Y \rightarrow Z})(\rho_{XY})}.$$

It has been shown that any 1D Gibbs states with a short-range Hamiltonian have exponentially decaying $\text{Cor}(X : Y)_\rho$ [25], i.e., there exist constants $c, \xi \geq 0$ such that

$$\text{Cor}(X : Y)_\rho \leq ce^{-d(X:Y)/\xi}. \quad (61)$$

Choosing $X = AB_l \dots B_{i+1}$, $Y = B_i$, $Z = B_i \dots B_1 C$ and $\mathcal{E}_{Y \rightarrow Z} = \tilde{E}_i$ prove Lemma 10. \square

Without loss of generality, let us assume

$$p_i = \text{Tr}[\tilde{\Lambda}_i(\rho_{AB_l \dots B_i}^{H_{ABC}})] = p > 0$$

for all i .⁹ Lemma 10 allows an iterative calculation. First we have

$$\left\| \Lambda_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}}) - \tilde{\Lambda}_1(\rho_{AB}^{H_{ABC}}) + (1-p)\tilde{\Lambda}_2(\rho_{AB_l \dots B_2}^{H_{ABC}}) \right. \\ \left. + (1-p) \left(\tilde{\Lambda}_3 + \left(\dots \left(\tilde{\Lambda}_l + \tilde{E}_l \right) \text{Tr}_{l-1} \tilde{E}_{l-1} \right) \dots \right) \text{Tr}_2 \tilde{E}_2(\rho_{AB_l \dots B_2}^{H_{ABC}}) \right\|_1 \leq e^{-l/\xi}.$$

Here, we used Lemma 10 for $\text{Tr}_1 \tilde{E}_1(\rho_{AB}^{H_{ABC}})$ and $\rho_{AB_l \dots B_2}^{H_{ABC}}$. Then we can obtain

$$\left\| \Lambda_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}}) - \tilde{\Lambda}_1(\rho_{AB}^{H_{ABC}}) + (1-p)\tilde{\Lambda}_2(\rho_{AB_l \dots B_2}^{H_{ABC}}) \right. \\ \left. + (1-p)^2 \tilde{\Lambda}_3(\rho_{AB_l \dots B_3}^{H_{ABC}}) + (1-p)^2 \left(\tilde{\Lambda}_4 + \dots \text{Tr}_4 \tilde{E}_4 \right) \text{Tr}_3 \tilde{E}_3(\rho_{AB_l \dots B_3}^{H_{ABC}}) \right\|_1 \leq 2e^{-l/\xi}.$$

We can proceed in a similar way, where at each i th step, we replace $\text{Tr}_i \tilde{E}_i(\rho_{AB}^{H_{ABC}})$ by $\text{Tr}_i(\rho_{AB}^{H_{ABC}})$ by using the triangle inequality. After iterating $l-1$ steps, we obtain

$$\left\| \Lambda_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}}) - \sum_{i=1}^l (1-p)^{i-1} \tilde{\Lambda}_i(\rho_{AB}^{H_{ABC}}) + (1-p)^{l-1} \tilde{E}_l(\rho_{AB_l}^{H_{ABC}}) \right\|_1 \\ \leq (l-1)e^{-l/\xi}. \quad (62)$$

Since $\left(\sum_{i=1}^l p(1-p)^{i-1} \right) + (1-p)^l = 1$, it follows

$$\rho^{H_{ABC}} = \sum_{i=1}^l p(1-p)^{i-1} \rho^{H_{ABC}} + (1-p)^l \rho^{H_{ABC}}$$

and thus

$$\left\| \rho^{H_{ABC}} - \sum_{i=1}^l (1-p)^{i-1} \tilde{\Lambda}_i(\rho_{AB}^{H_{ABC}}) + (1-p)^{l-1} \tilde{E}_l(\rho_{AB_l}^{H_{ABC}}) \right\|_1 \\ \leq \sum_{i=1}^l p(1-p)^{i-1} \left\| \rho^{H_{ABC}} - \frac{1}{p} \tilde{\Lambda}_i(\rho_{AB_l \dots B_i}^{H_{ABC}}) \right\|_1$$

⁹ Otherwise, we just pick the smallest p among $p_i, i = 1, \dots, l$ and redefine $\tilde{\Lambda}_i$ to be $\frac{p}{p_i} \tilde{\Lambda}_i$ and \tilde{E}_i to be $\tilde{E}_i + (1 - \frac{p}{p_i}) \tilde{\Lambda}_i$.

$$\begin{aligned}
& + (1-p)^l \left\| \rho^{H_{ABC}} - \frac{1}{1-p} \tilde{E}_l(\rho_{AB_i}^{H_{ABC}}) \right\|_1 \\
& \leq \{1 - (1-p)^l\} C_2(\beta) e^{-q_1(\beta)l} + 2(1-p)^l.
\end{aligned} \tag{63}$$

Therefore, by combining Eqs. (62) and (63), we conclude

$$\begin{aligned}
\| \rho^{H_{ABC}} - \Lambda_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}}) \|_1 & \leq \{1 - (1-p)^l\} C_2(\beta) e^{-q_1(\beta)l} \\
& \quad + 2e^{-|\ln(1-p)l|} + (l-1)e^{-l/\xi} \\
& \leq C_2(\beta) e^{-q_1(\beta)l} + 2e^{-|\ln(1-p)l|} + le^{-l/\xi}.
\end{aligned} \tag{64}$$

Here, the probability p can be bounded as in Eq. (59), and thus we have

$$|\ln(1-p)| \geq \left| \ln \left(1 - \frac{(1-e^{-1})e^{-2\beta J}}{\left(1 + e^{-\frac{\beta J}{2}} K(\beta)\right)^4} \right) \right| \geq \frac{(1-e^{-1})e^{-2\beta J}}{\left(1 + e^{-\frac{\beta J}{2}} K(\beta)\right)^4} = e^{-\Theta(\beta)},$$

where the last inequality follows from $\log(1-x) \leq -x$ for any $x \in [0, 1]$. If $\xi = e^{\mathcal{O}(\beta)}$, Eq. (64) can be bounded by

$$2C_2(\beta)le^{-q'(\beta)l} = e^{-q'(\beta)l + \ln(2C_2(\beta)l)}, \tag{65}$$

where $q'(\beta) = \Omega(e^{-\Theta(\beta)})$. Since $d(A, C) = 3l^2 - l$, Eq. (65) is $\Theta(e^{-\Theta(\sqrt{d(A,C)}})$. Therefore, for sufficiently large l , there exists a constant $q(\beta) = \Omega(e^{-\Theta(\beta)})$ such that $e^{-q'(\beta)l + \ln(2C_2(\beta)l)} \leq e^{-q(\beta)\sqrt{d(A,C)}}$.

4.3. Proof of Corollary 5.

Proof. Let us first consider a 1D open spin chain with a tripartition ABC so that a simply connected region B shields A from C . Then, $d(A, C) = |B|$. Without loss of generality, we assume $|A| \leq |B| \leq |C|$. Divide C into $C = C_1 \cup C_2 \cup \dots \cup C_m$, where m is the maximum number such that $|C_i| = |B|$ for $1 \leq i < m$ and each C_i shields C_{i-1} from C_{i+1} (here, $C_0 \equiv B$).

Theorem 4 and the Fannes inequality imply

$$\begin{aligned}
I(A : C_i | BC_1 \dots C_{i-1})_{\rho^{H_{ABC}}} & = S(A | BC_1 \dots C_{i-1})_{\rho^{H_{ABC}}} - S(A | BC_1 \dots C_i)_{\rho^{H_{ABC}}} \\
& \leq S(A | BC_1 \dots C_i)_{\Lambda_{B \dots C_{i-1} \rightarrow BC}(\rho_{AB \dots C_{i-1}}^{H_{ABC}})} \\
& \quad - S(A | BC_1 \dots C_i)_{\rho^{H_{ABC}}} \\
& \leq 6|B|e^{-\frac{q(\beta)}{2}\sqrt{i|B|}}
\end{aligned}$$

with a constant $q(\beta) \geq 0$ for any $i \in [1, m]$ and sufficiently large $|B|$. By the chain rule

$$I(A : C | B) = I(A : C_1 | B) + I(A : C_2 | BC_1) + \dots + I(A : C_m | BC_1 \dots C_{m-1}),$$

we have

$$I(A : C | B)_{\rho^{H_{ABC}}} \leq 6 \sum_{i=1}^m |B| e^{-\frac{q(\beta)}{2}\sqrt{i|B|}}$$

$$\begin{aligned}
&\leq 6 \left(|B| e^{-\frac{q(\beta)}{2} \sqrt{|B|}} + |B| \sum_{i=1}^{m-1} e^{-\frac{q(\beta)}{2} \sqrt{(i+1)|B|}} \right) \\
&\leq 6 \left(|B| e^{-\frac{q(\beta)}{2} \sqrt{|B|}} + \int_1^\infty e^{-\frac{q(\beta)}{2} \sqrt{x|B|}} dx \right) \\
&= 6 \left(|B| + \frac{8(1 + \frac{q(\beta)}{2} \sqrt{|B|})}{q(\beta)^2} \right) e^{-\frac{q(\beta)}{2} \sqrt{|B|}}.
\end{aligned}$$

Again, the upper bound is $e^{-\Theta(\sqrt{d(A,C)})}$. The same strategy works for a more complicated tripartition ABC of both 1D open chains and closed chains. \square

4.4. Proof of Corollary 6.

Proof. Let us consider 1D open spin chain without loss of generality. We first divide the whole chain into consecutive regions $A_1 B_1 C_1 A_2 B_2 C_2 \dots A_k B_k C_k$, where we choose $|A_i| = |B_j| = l \geq l_0$ and $|C_i| = 5\xi(\ln d)l$ for all i , where l_0 is the constant given in Theorem 4, the correlation length ξ is given in Eq. (61) and d is a constant bounding the dimension of the Hilbert space of one spin from above. Let us consider region $(A_1 B_1 C_1 \dots C_{i-1} A_i B_i C_i) (B_i A_{i+1} C_i)$, where $B_i A_{i+1}$ shields $A_1 B_1 \dots B_{i+1}$ from C_i . From Theorem 4, there exists a CPTP-map $\delta_i : \mathcal{D}(\mathcal{H}_{B_i A_{i+1}}) \rightarrow \mathcal{D}(\mathcal{H}_{B_i C_i A_{i+1}})$ such that

$$\left\| \Delta_i \left(\rho_{A_1 B_1 \dots A_i B_i A_{i+1} B_{i+1} C_{i+1}}^H \right) - \rho_{A_1 B_1 \dots A_i B_i C_i A_{i+1} B_{i+1} C_{i+1}}^H \right\|_1 \leq C e^{-q(\beta)\sqrt{l}}. \quad (66)$$

Since the Gibbs state has exponentially decaying correlations, after tracing out C_i , the two remained connected components are almost uncorrelated. By using Lemma 20 of Ref. [28], it follows that

$$\begin{aligned}
&\left\| \rho_{A_1 B_1 C_1 \dots B_{i-1} A_i B_i} - \rho_{A_1 B_1 C_1 \dots B_{i-1}} \otimes \rho_{A_i B_i} \right\|_1 \\
&\leq (\dim \mathcal{H}_{A_i B_i})^2 \text{Cor}(A_1 B_1 C_1 \dots B_{i-1} : A_i B_i)_\rho^H \\
&\leq d^{4l} e^{-5\xi(\ln d)l/\xi} \\
&= e^{-(\ln d)l}.
\end{aligned} \quad (67)$$

Each Δ_i acts on different sets of spins and therefore does not overlap. Then we have

$$\begin{aligned}
&\left\| \Delta_1 \otimes \dots \otimes \Delta_k (\rho_{A_1 B_1} \otimes \dots \otimes \rho_{A_k B_k}) - \rho_{A_1 B_1 C_1 \dots A_k B_k C_k} \right\|_1 \\
&\leq \left\| \Delta_1 \otimes \dots \otimes \Delta_k (\rho_{A_1 B_1} \otimes \dots \otimes \rho_{A_k B_k}) - \Delta_2 \right. \\
&\quad \left. \otimes \dots \otimes \Delta_k (\rho_{A_1 B_1 C_1 A_2 B_2} \otimes \rho_{A_3 B_3} \otimes \dots \otimes \rho_{A_k B_k}) \right\|_1 \\
&\quad + \left\| \Delta_2 \otimes \dots \otimes \Delta_k (\rho_{A_1 B_1 C_1 A_2 B_2} \otimes \rho_{A_3 B_3} \otimes \dots \otimes \rho_{A_k B_k}) - \rho_{A_1 B_1 C_1 \dots A_k B_k C_k} \right\|_1 \\
&\leq \left\| \Delta_1 (\rho_{A_1 B_1} \otimes \rho_{A_2 B_2}) - \rho_{A_1 B_1 C_1 A_2 B_2} \right\|_1 \\
&\quad + \left\| \Delta_2 \otimes \dots \otimes \Delta_k (\rho_{A_1 B_1 C_1 A_2 B_2} \otimes \rho_{A_3 B_3} \otimes \dots \otimes \rho_{A_k B_k}) - \rho_{A_1 B_1 C_1 \dots A_k B_k C_k} \right\|_1 \\
&\leq \left\| \Delta_1 (\rho_{A_1 B_1 A_2 B_2}) - \rho_{A_1 B_1 C_1 A_2 B_2} \right\|_1 + e^{-(\ln d)l} \\
&\quad + \left\| \Delta_2 \otimes \dots \otimes \Delta_k (\rho_{A_1 B_1 C_1 A_2 B_2} \otimes \rho_{A_3 B_3} \otimes \dots \otimes \rho_{A_k B_k}) - \rho_{A_1 B_1 C_1 \dots A_k B_k C_k} \right\|_1 \\
&\leq 2e^{-q(\beta)\sqrt{l}} + \left\| \Delta_2 \otimes \dots \otimes \Delta_k (\rho_{A_1 B_1 C_1 A_2 B_2} \otimes \dots \otimes \rho_{A_k B_k}) - \rho_{A_1 B_1 C_1 \dots A_k B_k C_k} \right\|_1.
\end{aligned}$$

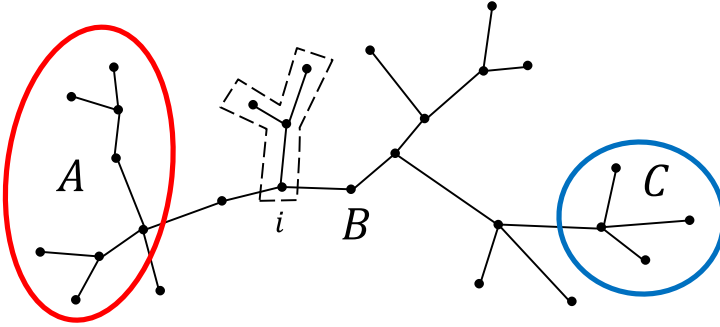


Fig. 4. An example of the region with a partition ABC for a tree graph. Here, B is the set of all spins outside of the circles. In the coarse-grain procedure, the descendants of i (the spins in the dashed region) can be regarded as one large system

The first inequality follows from the triangle inequality, the second from the monotonicity of the trace norm under quantum operations, the third from Eq. (67), and the fourth from Eq. (66) and $e^{-(\ln d)l} \leq e^{-q(\beta)\sqrt{l}}$ for large l .

Iterating the argument above, we find

$$\left\| \Delta_1 \otimes \cdots \otimes \Delta_k (\rho_{A_1 B_1} \otimes \cdots \otimes \rho_{A_k B_k}) - \rho_{A_1 B_1 C_1 \dots A_k B_k C_k} \right\|_1 \leq 2k e^{-q(\beta)\sqrt{l}}.$$

Since $k \leq n$, choosing $l = \mathcal{O}(\log^2(n/\varepsilon))$ gives an error bounded by ε . We denote a CPTP-map which construct $\rho_{A_i B_i}$ by $\Delta_{1,i}$ and relabel Δ_i in the above by $\Delta_{2,i}$. The CPTP-map $\bigotimes_i \Delta_{1,i}$ creates product state of the form of $\rho_{A_1 B_1} \otimes \rho_{A_2 B_2} \otimes \cdots \otimes \rho_{A_k B_k}$, and then $\bigotimes_i \Delta_{2,i}$ approximately creates the target state from this product state. \square

5. Extension to More General Graphs and a Conjecture for Higher Dimension

Our proof for 1D spin chains can be generalized to more general graphs with appropriate partitions. For instance, let us consider a tree graph $G = (E, V)$ with a partition ABC as depicted in Fig. 4. Since G is a tree, there is a unique path connecting A and C . Then, all spins in B are classified as (i) spins belonging to the path (ii) descendants of spins on the path (iii) the rest spins which are separated from the path. We can obtain a coarse grained 1D chain by regarding each spin on the path and its descendants as one system, and removing all spins in (iii). Therefore, we can apply the proof in the previous section to this situation as well. Note that the norm of an interaction term connecting spins on the path is irrelevant to the size of the coarse grained spins.

An important point of the above argument is that the success probability of the recovery map in lemma 9 is bounded by a constant of $d(A, C)$. In general cases, we can consider a partition ABC which cannot be reduced to 1D systems, such as depicted in Fig. 5. Remember that the success probability p is in the order $\Omega(e^{\beta \|H_{B^L}\|})$. In the case of Fig. 5, H_{B^L} is the sum of all interactions along the perimeter of B^L , which is proportional to l . When considering the repeat-until-success method, the success probability decays too rapidly, and therefore our strategy does not work.

The quantum Hammersley–Clifford theorem holds for more general class of Markov networks such as regular D -dimensional lattices. In this case, a partition ABC of the system is chosen so that B shields A from C (Fig. 5). Due to this observation, we expect

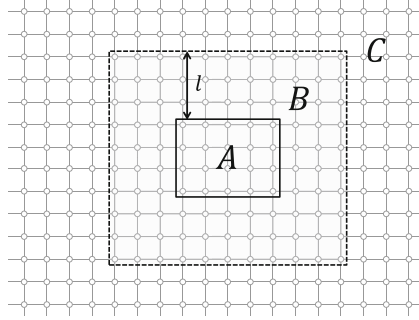


Fig. 5. An example of a 2D lattice with a partition ABC . We expect that the conditional mutual information $I(A : C|B)_\rho$ for any Gibbs state decays fast with respect to $d(A, C)$

that Gibbs states of short-ranged Hamiltonians obey fast decay of the conditional mutual information in D -dimensional systems as well. We have the following conjecture:

Conjecture 1. *Let ρ be a Gibbs state of a short-ranged Hamiltonian defined on a D -dimensional spin lattice. Then, there exist constants $C, c > 0$ such that for every three regions A, B, C with B shielding A from C ,*

$$I(A : C|B)_\rho \leq C e^{-cd(A,C)}.$$

Note that when $D = 1$, this conjecture gives an improved bound. It turns out that the area law for mutual information implies a weak version of the conjecture, as discussed in Sect. 2.3. Consider a Gibbs state in the infinite volume limit (as a KMS state). Let A be a region of the lattice and B_l be a ring around A of width l . Then, because of the area law [14], for every $\varepsilon > 0$ there is an integer l s.t.

$$I(A : B_{l+1}) - I(A : B_l) \leq \varepsilon,$$

which can be written as

$$I(A : C|B_l) \leq \varepsilon,$$

with $C := B_{l+1} \setminus B_l$.

We can also ask whether Theorem 1 can be extended to higher dimensions, i.e. is any state on a D -dim lattice with small $I(A : C|B)$ for B shielding A from C close to thermal? As displayed in Theorem 2, we may need additional conditions for general graphs. Although we do not know any counter-example, we also could not find any partial result whether the additional conditions in those theorems are necessary.

Acknowledgements. Part of this work was done when both of us were working in the QuArC group of Microsoft Research. KK thanks Advanced Leading Graduate Course for Photon Science (ALPS) and JSPS KAKENHI Grant Number JP16J05374 for financial support. We thank Matthew Hastings and Michael Kastoryano for useful discussions.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix A: Another approach to prove Theorem 4

In Sect. 4, we employ a perturbative method to obtain a local operator O determined by the Hamiltonian and a local operator V such that

$$e^{-\beta(H+V)} \approx Oe^{-\beta H} O^\dagger.$$

The existence of such operator plays a central role in the proof of Theorem 4. In this appendix, we introduce another approach to obtain similar operators, which is based on the previous work by Araki [25]. The main difference between these approaches is the origins of locality of the operator O . In the perturbative approach, the locality is originated in Lieb-Robinson bounds, which restrict real time evolutions of operators. Instead, in Araki’s approach, locality of O is originated in a restriction on imaginary-time evolutions of V .

Let us consider a 1D spin chain $\Lambda = [-n, n]$, a short-range Hamiltonian H on Λ , and a local operator V . We denote the maximum strength of H by J . A simple algebra show the following relation holds.

$$\begin{aligned} e^{-\beta(H+V)} &= e^{-\frac{\beta}{2}(H+V)} e^{\frac{\beta}{2}H} e^{-\beta H} e^{\frac{\beta}{2}H} e^{-\frac{\beta}{2}(H+V)} \\ &\equiv E_r(V; H) e^{-\beta H} E_r(V; H)^\dagger, \end{aligned}$$

where we denote $e^{-\frac{\beta}{2}(H+V)} e^{\frac{\beta}{2}H}$ by $E_r(V; H)$. By denoting $V(\beta) = e^{-\beta H} V e^{\beta H}$, $E_r(V; H)$ has another form written as [25]

$$E_r(V; H) = \sum_{n=0}^{\infty} (-1)^n \int_0^{\frac{\beta}{2}} d\beta_1 \int_0^{\beta_1} d\beta_2 \cdots \int_0^{\beta_{n-1}} d\beta_n V(\beta_n) \cdots V(\beta_1).$$

Actually, $E_r(V; H)$ can be approximated by a local operator.

Lemma 11. [25] *The following statements hold for any region $X \subset [-n, n]$ and any bounded operator V with $\text{supp}(V) = [a, b] \subset [-n, n]$.*

(i) *There exists a constant $C \geq 0$ depending on β, J and $\|V\|$ such that*

$$\|E_r(V; H_X)\| \leq C$$

(ii) *There exist constants $C, q \geq 0$ depending on β, J and $\|V\|$ such that*

$$\|E_r(V; H_X) - E_r(V; H_{X \cap [a-l, b+l]})\| \leq C \frac{q^{1+\lfloor \frac{l}{2} \rfloor}}{(1 + \lfloor \frac{l}{2} \rfloor)!}.$$

Since $\log x! \approx x \log x - x$, the denominator grows faster than the numerator with respect to l , and thus, the accuracy of the above approximation is exponentially good with respect to l . Note that similar properties hold for the inverse of $E_r(V; H_X)$, $E_l(V; H_X) \equiv e^{-\frac{\beta}{2}H_X} e^{\frac{\beta}{2}(H_X+V)}$. Therefore, by choosing $V = H_{B_M}$, $E_r(V; H)$ and its local approximation play the same role as O_{ABC} and O_B in Sect. 4.2.

References

1. Ibinson, B., Linden, N., Winter, A.: Robustness of quantum markov chains. *Commun. Math. Phys.* **277**(2), 289–304 (2008)
2. Fawzi, O., Renner, R.: Quantum conditional mutual information and approximate markov chains. *Commun. Math. Phys.* **340**(2), 575–611 (2015)
3. Devetak, I., Yard, J.: Exact cost of redistributing multipartite quantum states. *Phys. Rev. Lett.* **100**, 230501 (2008)
4. Berta, M., Brandao, F.G.S.L., Majenz, C., Wilde, M.M.: Deconstruction and conditional erasure of quantum correlations (2016). arXiv e-prints [arXiv:1609.06994](https://arxiv.org/abs/1609.06994)
5. Sharma, K., Wakakuwa E., Wilde, M.M.: Conditional quantum one-time pad (2017). arXiv e-prints [arXiv:1703.02903](https://arxiv.org/abs/1703.02903)
6. Hammersley, J.M., Clifford, P.E.: Markov field on finite graphs and lattices (1971). <http://www.statslab.cam.ac.uk/grg/books/hammfest/hamm-cliff.pdf>
7. Leifer, M.S., Poulin, D.: Quantum graphical models and belief propagation. *Ann. Phys.* **323**(8), 1899–1946 (2008)
8. Brown, W., Poulin D.: Quantum Markov networks and commuting Hamiltonians (2012). arXiv e-prints [arXiv:1206.0755](https://arxiv.org/abs/1206.0755)
9. Brandao, F.G.S.L., Kastoryano, M.J.: Finite correlation length implies efficient preparation of quantum thermal states (2016). arXiv e-prints [arXiv:1609.07877](https://arxiv.org/abs/1609.07877)
10. Kim, I.H.: Determining the structure of the real-space entanglement spectrum from approximate conditional independence. *Phys. Rev. B* **87**, 155120 (2013)
11. Flammia, S.T., Haah, J., Kastoryano, M.J., Kim, I.H.: Limits on the storage of quantum information in a volume of space. *Quantum* **1**, 4 (2017)
12. Levin, M., Wen, X.-G.: Detecting topological order in a ground state wave function. *Phys. Rev. Lett.* **96**, 110405 (2006)
13. Kato, K., Brandao, F.G.S.L.: Locality of edge states and entanglement spectrum from strong subadditivity (2018). arXiv preprints [arXiv:1804.05457](https://arxiv.org/abs/1804.05457)
14. Wolf, M.M., Verstraete, F., Hastings, M.B., Cirac, J.I.: Area laws in quantum systems: mutual information and correlations. *Phys. Rev. Lett.* **100**, 070502 (2008)
15. Hastings, M.B.: Solving gapped Hamiltonians locally. *Phys. Rev. B* **73**(8), 085115 (2006)
16. Swingle, B., McGreevy, J.: Mixed s -sourcery: building many-body states using bubbles of nothing. *Phys. Rev. B* **94**(15), 155125 (2016)
17. Brandao, F.G.S.L.: Finite correlation length implies efficient preparation of quantum thermal states (2016). arXiv preprints [arXiv:1609.07877](https://arxiv.org/abs/1609.07877)
18. Jaynes, E.T.: Information theory and statistical mechanics. *Phys. Rev.* **106**, 620–630 (1957)
19. Jaynes, E.T.: Information theory and statistical mechanics. II. *Phys. Rev.* **108**, 171–190 (1957)
20. Weis, S.: Information topologies on non-commutative state spaces. *J. Conv. Anal.* **21**(2), 339–399 (2014)
21. Weis, S.: The MaxEnt extension of a quantum Gibbs family, convex geometry and geodesics. In: AIP Conference Proceedings, vol. 1641 (2015)
22. Alicki, R., Fannes, M.: Continuity of quantum conditional information. *J. Phys. A: Math. Gen.* **37**, L55 (2004)
23. Hastings, M.B.: Quantum belief propagation: an algorithm for thermal quantum systems. *Phys. Rev. B* **76**, 201102 (2007)
24. Kim, I.H.: Perturbative analysis of topological entanglement entropy from conditional independence. *Phys. Rev. B* **86**, 245116 (2012)
25. Araki, H.: Gibbs states of a one dimensional quantum lattice. *Commun. Math. Phys.* **14**(2), 120–157 (1969)
26. Lieb, E.H., Robinson, D.W.: The finite group velocity of quantum spin systems. *Commun. Math. Phys.* **28**(3), 251–257 (1972)
27. Araki, H.: Expansional in banach algebras. *Ann. Sci. Ecole Norm. S.* **6**(1), 67–84 (1973)
28. Brandao, F.G.S.L., Horodecki, M.: Exponential decay of correlations implies area law. *Commun. Math. Phys.* **333**(2), 761–798 (2015)

Communicated by M. M. Wolf