Clustering of conditional mutual information for quantum Gibbs states above a threshold temperature

Tomotaka Kuwahara[*]

Mathematical Science Team, RIKEN Center for Advanced Intelligence Project (AIP),
1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan
Department of Mathematics, Faculty of Science and Technology,
Keio University, 3-14-1 Hiyoshi, Kouhoku-ku, Yokohama 223-8522, Japan and
interdisciplinary Theoretical & Mathematical Sciences Program
(iTHEMS) RIKEN 2-1, Hirosawa, Wako, Saitama 351-0198, Japan

Kohtaro Kato[†]

Institute for Quantum Information and Matter,
California Institute of Technology, Pasadena, CA 91125, USA

Fernando G. S. L. Brandão[‡]

Amazon Web Services, AWS Center for Quantum Computing, Pasadena, CA 91125, USA and
Institute for Quantum Information and Matter,
California Institute of Technology, Pasadena, CA 91125, USA

We prove that quantum Gibbs states of spin systems above a certain threshold temperature are approximate quantum Markov networks, meaning that the conditional mutual information decays rapidly with distance. We prove exponential decay (power-law decay) for short-ranged (long-ranged) interacting systems. As consequences, we establish the efficiency of quantum Gibbs sampling algorithms, a strong version of the area law, the quasi-locality of effective Hamiltonians on subsystems, a clustering theorem for mutual information, and a polynomial-time algorithm for classical Gibbs state simulation.

Introduction.— Quantum Gibbs states describes the thermal equilibrium properties of quantum systems. The advent of quantum information science opened up new investigation avenues in the study of Gibbs states, such as the stability of topological quantum memory [1–4], thermalization in isolated quantum systems [5–10], and Hamiltonian complexity [11–14]. Efficient methods to prepare quantum Gibbs states in quantum computers have also found useful in giving quantum speed-ups for problems such as semidefinite programming [15–17] and quantum machine learning [18–22].

Quantum Gibbs state also inherit the locality property of their parent Hamiltonian, which allows for an efficient classical description in many cases. One of the simple characterizations is the exponential decay of bipartite correlation functions which is shown to be true in general one-dimension quantum spin lattices [23] and in higher dimensions above a threshold temperature [24–28]. Another characterization is that at arbitrary finite temperatures, the mutual information between a region and its complement obeys the area law [29]. Quantum Gibbs states also have efficient representations in terms of tensor networks [30, 31].

In classical systems, there are even stronger structural results for Gibbs states. For instance, the Hammersley–Clifford theorem [32] states that classical Gibbs states are equivalent to a class of probability distributions called *Markov networks.* They satisfy the Markov property, that is, a site is independent from all others conditioned on its neighbors. Therefore, for classical Gibbs states all the correlations between two separated vertices are induced by intermediate vertices connecting them.

Although the notion of conditional probability distiburion is missing in quantum systems, we can still generalize Markov networks to quantum systems by using the (quantum) conditional mutual information.

$$\begin{aligned} \mathcal{I}_\rho(A:C|B) \\ := S(\rho^{AB}) + S(\rho^{BC}) - S(\rho^{ABC}) - S(\rho^B), \end{aligned} \quad (1)$$

where ρ^{AB} is the reduced density matrix in the subsystem ($AB = A \cup B$) and $S(\rho^{AB})$ is the von Neumann entropy, namely, $S(\rho^{AB}) := \operatorname{tr}(\rho^{AB} \log \rho^{AB})$ with the natural base. In classical systems, the conditional mutual information become zero if and only if the state is conditionally independent, and therefore it provides a natural measure of conditional independence.

The quantum version of the Hammersley–Clifford theorem has been established for the case where the Hamiltonian is short-range and commuting [33, 34]: any quantum Gibbs state of such Hamiltonian on a triangle-free graph is a Markov network and vise versa. More recently, it has been shown that the Hammersley–Clifford theorem *approximately* holds in one-dimensional lattice [35], in the sense that the conditional mutual information of any Gibbs state decays subexponentially with respect to distance.

In the present work, we establish the approximate Markov property for quantum Gibbs states in spin systems interacting on generic graphs at high temperatures; in other words, we prove the decay of conditional mutual information between two separated subsystems with respect to the size of the conditioning region (using the Manhattan distance on the graph). Herein, we consider not only short-range interactions but also long-range (i.e., power-law decaying) interactions on graphs. We show that above a certain threshold tempera-
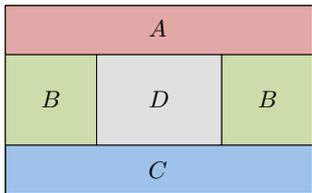
FIG. 1. (color online) Decomposition of the total system into A, B, C, and D. It is possible that a quantum state has no correlation between A and C when looking at only the subsystems A and C but has a strong correlation when looking at them via the subsystem B. This kind of correlation between A and B related to C is measured by conditional mutual information (1). Physically, conditional mutual information characterizes tripartite correlations between A, B and C. A representative example is the topological entanglement entropy [36, 37], which is a special form of the conditional mutual information.

ture, the conditional mutual information decays exponentially (pollynomialy) for short ranged (long ranged) models. Above the temperature threshold, our result strengthens the 1D result from Ref. [35], the area law for mutual information [29] and the standard clustering theorem [27, 28]. Moreover, we rigorously establish a quasipolynomial-time quantum Gibbs sampling algorithm for high temperatures by following the discussion in Ref. [38]. In the computation of thermodynamic quantities (e.g., the partition function), we can give a polynomial-time classical algorithm. Finally, our bound prohibits topological order above a threshold temperature in short-range and long-range interacting systems, as we find a quasi-local constant depth circuit to prepare such states.

Setup.— We consider a quantum system with n spins, where each spin has a d-dimensional Hilbert space. We assume that the spins sit on the vertices of a graph $G = (V, E)$ where V is the total spin set ($|V| = n$). For arbitrary subsystems $A, B \subset V$, we define $d_{A,B}$ as the shortest path length on the graph that connects A and B. If $A \cap B \neq \emptyset$, $d_{A,B} = 0$.

We define the system Hamiltonian H as

$$H = \sum_{|X| \leq k} h_X, \qquad (2)$$

where $\| \cdot \|$ is the operator norm and each of the interaction terms $\{h_X\}$ acts on the spins in $X \subset V$. The Hamiltonian [Eq. (2)] describes generic k-body interacting systems. We characterize the locality of the interactions as follows:

$$\sum_{\substack{X | X \ni v \\ \mathrm{diam}(X) \geq R}} \|h_X\| \leq f(R) \quad \text{with} \quad f(1) \leq 1 \qquad (3)$$

for $\forall v \in V$, where $\mathrm{diam}(X)$ is the diameter of X, namely, $\mathrm{diam}(X) := \max_{\{v_1, v_2\} \in X} d_{v_1, v_2}$ for $X \subset V$. For example, if $k = 2$ and $f(R) = 0$ for $R \geq 2$, the Hamiltonian is described by bipartite nearest-neighbor interactions as $H = \sum_{\{i,j\} \in E} h_{i,j}$. We consider the Gibbs state for the Hamiltonian H at an inverse temperature β as follows:

$$\rho := \frac{1}{Z} e^{-\beta H}, \quad Z := \mathrm{tr}(e^{-\beta H}). \qquad (4)$$

Our purpose is to discuss the Markov property of Gibbs states. Let $V_0 \subseteq V$ be an arbitrary subsystem. Consider a tripartite partitioning of V_0 as $V_0 = ABC$, where we denote $A \cup B$ by AB for simplicity. We notice that the subsystems $\{A, B, C\}$ are *not* necessarily concatenated on the graph (see Fig. 1). If any two nonadjacent subsystems A and C are conditionally independent of the other subsystem B ($= V_0 \setminus AC$), we say that ρ^{V_0} is the *quantum Markov network* on V_0. Mathematically, this implies $\mathcal{I}_\rho(A : C|B) = 0$ for $d_{A,C} > 0$ [33, 39], where $\mathcal{I}_\rho(A : C|B)$ is defined in Eq. (1). It is noteworthy that the Markov property of ρ^{V_0} strongly depends on the selection of the subsystem $V_0 \subseteq V$. To see this point, let us consider a one-dimensional graph. Then, the GHZ state is a Markov network for $\forall V_0 \subset V$, but not globally, namely, $\mathcal{I}_\rho(A : C|B) = 1$ for $ABC = V$. In contrast, the cluster state [40] is globally a Markov network, but not for particular selections of V_0 (e.g., $V_0 = \{2, 4, 6, 8, \ldots, 2\lfloor n/2 \rfloor\}$) [41, 42] (see also [43]). Based on the example of the cluster state, which has a finite correlation length and is described by the matrix product state with bond dimension 2 [44], we cannot ensure the Markov property only by the clustering theorem and the matrix product (or tensor network) representation of the quantum Gibbs state.

The Markov property has a clear operational meaning in terms of a recovery map as follows: If ρ^{V_0} is a Markov network, we can always find a quantum channel $\tau_{B \to BC}$ referred to as the Petz recovery map [39, 45], which recovers ρ^{ABC} from ρ^{AB} ($V_0 = ABC$):

$$\tau_{B \to BC}(\rho^{AB}) = \rho^{ABC}. \qquad (5)$$

The above local reconstruction is not possible for generic quantum states. Note that any quantum Markov network on a tree graph can be constructed from a sequence of n local quantum channels.

In realistic situations, we frequently encounter cases where the density matrix is not given by the exact Markov network but by an approximate Markov network, that is, the conditional mutual information $\mathcal{I}_\rho(A : C|B)$ approaches zero as the distance $d_{A,C}$ increases. In the case where $\mathcal{I}_\rho(A : C|B) = \epsilon$, the celebrated Fawzi–Renner theorem [46] (see also [47]) ensures the existence of the recovery map such that

$$\left\| \tau_{B \to BC}(\rho^{AB}) - \rho^{ABC} \right\|_1^2 \leq \epsilon \log 2, \qquad (6)$$

where $\| \cdot \|_1$ is the trace norm and the form of $\tau_{B \to BC}$ is given by the rotated Petz map (see Ref. [48] for the explicit definition). Based on this theorem, we can still relate the approximate Markov property to the local reconstruction of the state.

The main purpose of this study is to characterize the decay rate of the conditional mutual information $\mathcal{I}_\rho(A : C|B)$ with respect to the distance $d_{A,C}$ on the graph. To concentrate on the physics given by the theorems, we have provided the proofs of our main theorems in the supplementary material [49].

Main result.— We prove the exponential decay of the conditional mutual information above a temperature threshold ($1/\beta_c$), where β_c does not depend on the system size n but only on k in Eq. (2).

Theorem 1. *Let the interaction length r be finite, namely, $f(R) = 0$ for $R > r \in \mathbb{N}$ in Ineq. (3). Then, the condition*

$$\beta < \beta_c := \frac{1}{8e^3 k} \tag{7}$$

implies that the Gibbs state ρ is an approximate Markov network on an arbitrary subset $V_0 \subseteq V$ in the sense that

$$\mathcal{I}_\rho(A : C | B) \le e \min(|\partial A_r|, |\partial C_r|) \frac{(\beta/\beta_c)^{d_{A,C}/r}}{1 - \beta/\beta_c}, \tag{8}$$

where $V_0 = ABC$. We define the surface region of an arbitrary subsystem $L \subseteq V$ as $\partial L_l \subseteq V$ $(l \in \mathbb{N})$:

$$\partial L_l := \{v \in L | d_{v,L^c} \le l\}, \tag{9}$$

where L^c is the complementary set of L (i.e., $L \cup L^c = V$).

We notice that if we select B as an empty set (i.e., $B = \emptyset$), the conditional mutual information reduces to bipartite mutual information:

$$\mathcal{I}_\rho(A : C | \emptyset) = \mathcal{I}_\rho(A : C),$$

where $\mathcal{I}_\rho(A : C) := S(\rho^A) + S(\rho^C) - S(\rho^{AC})$. Therefore, inequality (8) also implies the exponential decay of the mutual information between two separated subsystems. It is an improved version of the standard clustering theorem for the bipartite operator correlation $\mathrm{Cor}_\rho(O_A, O_B) := \mathrm{tr}(\rho O_A O_B) - \mathrm{tr}(\rho O_A)\mathrm{tr}(\rho O_B)$, where O_A and O_B are arbitrary operators with unit norm (i.e., $\|O_A\| = \|O_B\| = 1$) supported on subsystems A and B, respectively. From the relation $[\mathrm{Cor}_\rho(O_A, O_B)]^2 \le 2\mathcal{I}_\rho(A : B)$ [29], the clustering theorem can be derived from the exponential decay of the mutual information. Moreover, it is well known [50, 51] in the context of data hiding that even if the operator correlation is arbitrarily small in a quantum state, the state may still be highly correlated in the mutual information [29].

An important implication of this theorem is related to the quantum sampling of Gibbs states. Based on the Fawzi–Renner theorem (6), an approximate Markov network can be efficiently reconstructed from its reduced density matrix using a quantum computer. According to Ref. [38], the clustering and Markov properties ensure an efficient preparation of quantum Gibbs states on finite-dimensional lattices. By combining our theorem 1 with Theorem 5 in Ref. [38], we obtain the following corollary:

Corollary 2. *Let the graph G be a D-dimensional lattice. Then, under the assumption of Theorem 1, there exists a $(D+1)$-depth circuit of quantum channels $\mathbb{F} = \mathbb{F}_{D+1} \cdots \mathbb{F}_2 \mathbb{F}_1$ such that*

$$\|\mathbb{F}(\psi) - \rho\|_1 = 1/\mathrm{poly}(n), \tag{10}$$

where ψ is an arbitrary quantum state and each quantum channel $\{\mathbb{F}_s\}_{s=1}^{D+1}$ is composed of quasilocal CPTP maps that act on $\mathcal{O}(\log^D n)$ spins.

The number of the elementary gates for the each quasilocal channel $\{\mathbb{F}_s\}_{s=1}^{D+1}$ is of order $\exp[\mathcal{O}(\log^D n)] =$
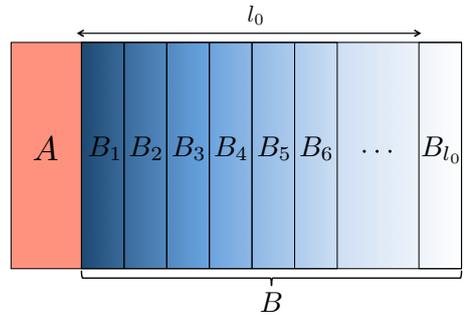

FIG. 2. (color online). Strengthening of the area law resulting from the Markov property. In the figure, we consider a 2D system and decompose it into A and $B = B_1 B_2 \ldots B_{l_0}$ with $d_{A,B_l} = l$ $(1 \le l \le l_0)$. The total mutual information between A and B obeys the area law. The Markov property (8) ensures that the mutual information $\mathcal{I}_\rho(A : B_1 \cdots B_l)$ exponentially converges to $\mathcal{I}_\rho(A : B)$ as l increases. In other words, the information between A ad B is exponentially localized around the boundary.

$n^{\mathcal{O}(\log^{D-1} n)}$ [52, 53]. This also provides the computational time of Gibbs sampling by the quantum computer. This algorithm requires only quasipolynomial computational time, and it is considerably better than a few existing algorithms [54, 55]. The algorithm is worse than the algorithms proposed in Refs. [56] and [57], which require polynomial computational time. However, our method has advantages in the following senses: the method in [56] is applicable only to commuting Hamiltonians and the method in [57] requires twice the number of qubits (i.e., $2n$ qubits) for implementation.

The second implication of the theorem is the strengthening of the area law. The area law for mutual information has been derived at arbitrary temperatures in Ref. [29] in the form of

$$\mathcal{I}_\rho(A : B) \le c\beta |\partial A|, \tag{11}$$

where $AB = V$ and c is an $\mathcal{O}(1)$ constant. The area law implies that $\mathcal{I}_\rho(A : B')$ saturates as $B' \subset B$ grows to B, however Eq. (11) does not provide the saturation rate. Our result implies it saturate exponentially fast, and the mutual information between two subsystems is exponentially localized around the boundary between A and B. To see more details, let us decompose B into l_0 slices, $B_1 B_2 \ldots B_{l_0}$, with $d_{A,B_l} = l$ for $l = 1, 2, \ldots, l_0$ (see Fig. 2). Then, the question is how rapidly the mutual information $\mathcal{I}_\rho(A : B_1 B_2 \cdots B_l)$ saturates to $\mathcal{I}_\rho(A : B)$. From the relation $\mathcal{I}(A : C | B) = \mathcal{I}(A : BC) - \mathcal{I}(A : B)$ and Ineq. (8), we have

$$\mathcal{I}_\rho(A : B_1 \cdots B_l) - \mathcal{I}_\rho(A : B_1 \cdots B_{l-1})$$
$$= \mathcal{I}_\rho(A : B_l | B_1 \cdots B_{l-1}) \sim (\beta/\beta_c)^{l/r}, \tag{12}$$

which shows exponential decay with respect to l.

Effective Hamiltonian on subsystem and classical simulation of Gibbs state.— Theorem 1 is related to the locality of the effective Hamiltonian. We define the effective Hamiltonian of the local reduced density matrix as

$$\tilde{H}_L := -\beta^{-1} \log \mathrm{tr}_{L^c}(e^{-\beta H}). \tag{13}$$

We formally describe $\tilde{H}_L$ as

$$\tilde{H}_L = H_L + \Phi_L, \tag{14}$$

where H_L id composed of the original interacting terms in H on subsystem L, namely, $H_L = \sum_{X \subset L} h_X$, and Φ_L is the effective interaction term. We are interested in the locality of Φ_L. Typically, it is computationally difficult to determine the effective term, even in classical Gibbs states [58]. Our present question is whether the (quasi-)locality of Φ_L can be ensured or not (Fig. 3). In classical Gibbs states or systems with commuting Hamiltonians, Φ_L is exactly localized around the surface region of L (not necessarily localized along the boundary). This point is crucial for the Gibbs states to be the exact Markov network [33, 38]. Additionally, for systems with non-commuting Hamiltonians, the quasi-locality of Φ_L is numerically verified in Ref. [59]. By following the same analysis as the proof of Theorem 1, we can rigorously prove the quasi-locality of Φ_L not only the direction orthogonal to the boundary, but also along the boundary.

Theorem 3. *Under the setup and assumption of Theorem 1, Φ_L is approximated by a localized operator $\Phi_{\partial L_l}$ as follows:*

$$\|\Phi_L - \Phi_{\partial L_l}\| \leq \frac{e}{4\beta} \frac{(\beta/\beta_c)^{l/r}}{1 - \beta/\beta_c} |\partial L_r|, \qquad (15)$$

where $\Phi_{\partial L_l}$ is supported on the region ∂L_l that is separated from the boundary ∂L by a distance of at most l (see Eq. (9) for the definition). In addition, $\Phi_{\partial L_l}$ is composed of local operators that act on at most $(k\lfloor l/r \rfloor)$ spins (see the supplementary materials [49] for an explicit form of $\Phi_{\partial L_l}$). Moreover, computation of Φ_L up to an norm error of $n\epsilon$ is performed with the runtime bounded from above by

$$n(1/\epsilon)^{\mathcal{O}(k \log(dd_G^r))}, \qquad (16)$$

where d_G is the degree of the graph G.

This theorem immediately implies that the classical simulation of the Gibbs states is possible in polynomial time within an error of $1/\text{poly}(n)$. We note that the definition (13) implies $\Phi_L = -\beta^{-1}\log(Z)$ for $L = \emptyset$, i.e. we can calculate the partition function by the same algorithm. We can also calculate the expectation values of local observables or the local entropy by explicitly obtaining expression of $\rho_L = e^{-\beta\tilde{H}_L}$. This is summarized as the following corollary.

Corollary 4. *Thermodynamic properties such as local observables (e.g., energy and magnetization), the partition function $\log(Z)$, and local entropy $-\text{tr}(\rho^L \log \rho^L)$ are classically simulated in polynomial time $\text{poly}(n)$ as long as an error of $1/\text{poly}(n)$ is allowed.*

Long-range interacting systems. — Finally, we extend Theorem 1 from short-range interacting systems to long-range interacting systems. We define the Hamiltonian with the power-law decay interaction by assuming that $f(R)$ in (3) is given by

$$f(R) = R^{-\alpha}, \qquad (17)$$

where $\alpha > 0$. To consider a more general form as $f(R) = gR^{-\alpha}$, we must only scale the inverse temperature from β to β/g. For example, we can consider the
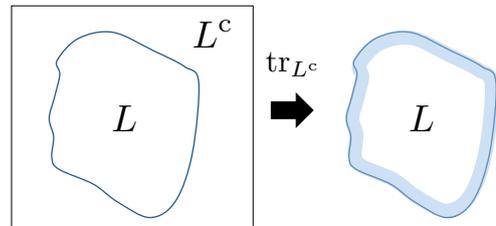

FIG. 3. (color online) Effective Hamiltonian $\tilde{H}_L$ for the reduced density matrix ρ^L. We decompose $\tilde{H}_L$ as $\tilde{H}_L = H_L + \Phi_L$, where H_L is the original Hamiltonian in L and Φ_L the effective term that originates outside L. Theorem 3 implies that Φ_L is exponentially localized around the surface region of L.

following Hamiltonian on a graph with a D-dimensional structure:

$$H = \sum_{i,j \in V} \frac{J}{R_{i,j}^{\alpha+D}} h_{i,j} \quad \text{with} \quad \|h_{i,j}\| = 1, \qquad (18)$$

where $R_{i,j}$ is the distance between spins i and j defined by the graph structure (V, E) and J is determined so that inequality (3) is satisfied. This kind of Hamiltonian is now controllable in realistic experiments and attracts much attention both in experimental [60–64] and theoretical aspects [65–69].

Similar to the case of short-range interacting systems, we prove the decay of the conditional mutual information for long-range interacting systems for $\alpha > 0$.

Theorem 5. *Let A, B, and C be arbitrary subsystems in V ($A, B, C \subset V$). Then, under the assumptions of $\beta < \beta_c/11$ and $d_{A,C} \geq 2\alpha$, the Gibbs state ρ satisfies the approximate Markov property as follows:*

$$\mathcal{I}_\rho(A : C|B) \leq \beta \min(|A|, |C|) \frac{C_\beta}{d_{A,C}^\alpha}, \qquad (19)$$

where $C_\beta := \frac{11e^{1/k}/\beta_c}{1 - 11\beta/\beta_c}$ and β_c was defined in (7).

By selecting $B = \emptyset$, we can also derive the power-law decay of the mutual information between two separated subsystems. To the best of our knowledge, the clustering theorem for the Gibbs state with long-range interaction is limited for classical cases [70–75] and special quantum cases [76, 77]. Our result provides the first general proof of the clustering theorem at finite temperatures in long-range interacting quantum systems. Moreover, we can discuss the saturation rate of the area law, similar to the case of short-range interacting systems. In the setup of Eq. (12), the mutual information $\mathcal{I}_\rho(A : B_1 \cdots B_l)$ approaches $\mathcal{I}_\rho(A : B)$ with an error of $l^{-\alpha}$.

The vanishing of the conditional mutual information implies the absence of the topological order above a temperature threshold. The stability of the topological order at finite temperatures has been extensively investigated in short-range interacting systems [1–4, 78–83]. Even though it is natural to expect that the topological order vanishes at sufficiently high temperatures, there is no general proof that topological entanglement entropy vanishes above a threshold temperature. In long-range interacting systems, the problem

is even more nontrivial, and there are few reports on the stability of the topological order [84]. To discuss the condition of α for the vanishing of $\mathcal{I}_\rho(A : C|B)$ in the thermodynamic limit, let us consider the case with $\min(|A|, |C|) = l_s^D$ and $d_{A,C} = l_s$, where l_s is the length of the total system. Then, inequality (19) implies $\mathcal{I}_\rho(A : C|B) \leq l_s^{D-\alpha}$, and hence, $\alpha > D$ is the sufficient condition for $\mathcal{I}_\rho(A : C|B) \to 0$ in the thermodynamic limit (i.e., $l_s \to \infty$).

Future perspective.— We mention several open problems. The most important problem is the Markov property in low-temperature regimes, where our present analytical technique (i.e., the generalized cluster expansion [49]) breaks down. It is no longer desirable that the Markov property holds for the arbitrary selections of the subregions A, B, and C because the topological order can exist at finite temperatures in four-dimensional systems [1]. Even if we restrict ourselves to the case of $ABC = V$, the problem is still challenging because only few mathematical techniques can access the conditional mutual information at arbitrary temperatures, except for one-dimensional systems [35]. Also, throughout the work, we consider the von Neumann entropy to charac-

terize conditional mutual information. It is interesting to generalize the present results to the Rényi-type conditional mutual information [85]. In this case, we must consider the Rényi entropy, S_ν, for reduced density matrix ρ^L, where $S_\nu(\rho^L) := \log\left(\mathrm{tr}[(\rho^L)^\nu]\right)/(1 - \nu)$ with $0 \leq \nu < \infty$. We expect that there exists a critical $\nu_c \approx \beta_c/\beta$, below which the exponential decay of the Rényi mutual information can be proved [86]

Note added.— On the classical simulations of quantum Gibbs states, we are aware of a related result obtained from a similar approach [87].

[1] Matthew B. Hastings, "Topological Order at Nonzero Temperature," Phys. Rev. Lett. **107**, 210501 (2011).

[2] Olivier Landon-Cardinal and David Poulin, "Local Topological Order Inhibits Thermal Stability in 2D," Phys. Rev. Lett. **110**, 090502 (2013).

[3] Sergey Bravyi and Jeongwan Haah, "Quantum Self-Correction in the 3D Cubic Code Model," Phys. Rev. Lett. **111**, 200501 (2013).

[4] Benjamin J. Brown, Daniel Loss, Jiannis K. Pachos, Chris N. Self, and James R. Wootton, "Quantum memories at finite temperature," Rev. Mod. Phys. **88**, 045005 (2016).

[5] Sandu Popescu, Anthony J Short, and Andreas Winter, "Entanglement and the foundations of statistical mechanics," Nature Physics **2**, 754 (2006).

[6] Markus P. Müller, Emily Adlam, Lluís Masanes, and Nathan Wiebe, "Thermalization and Canonical Typicality in Translation-Invariant Quantum Lattice Systems," Communications in Mathematical Physics **340**, 499–561 (2015).

[7] Fernando GSL Brandao and Marcus Cramer, "Equivalence of statistical mechanical ensembles for non-critical quantum systems," arXiv preprint arXiv:1502.03263 arXiv:1502.03263.

[8] Hal Tasaki, "On the Local Equivalence Between the Canonical and the Microcanonical Ensembles for Quantum Spin Systems," Journal of Statistical Physics **172**, 905–926 (2018).

[9] Tomotaka Kuwahara and Keiji Saito, "Ensemble equivalence and eigenstate thermalization from clustering of correlation," arXiv preprint arXiv:1905.01886 (2019), arXiv:1905.01886.

[10] Tomotaka Kuwahara and Keiji Saito, "Gaussian concentration bound and Ensemble equivalence in generic quantum many-body systems including long-range interaction," arXiv preprint arXiv:1906.10872 (2019), arXiv:1906.10872.

[11] David Poulin and Matthew B. Hastings, "Markov Entropy Decomposition: A Variational Dual for Quantum Belief Propagation," Phys. Rev. Lett. **106**, 080403 (2011).

[12] Tobias J Osborne, "Hamiltonian complexity," Reports on Progress in Physics **75**, 022001 (2012).

[13] Dorit Aharonov, Itai Arad, and Thomas Vidick, "Guest Column: The Quantum PCP Conjecture," SIGACT News **44**, 47–79 (2013).

[14] Sevag Gharibian, Yichen Huang, Zeph Landau, and Seung Woo Shin, "Quantum Hamiltonian Complexity," Foundations and Trends® in Theoretical Computer Science **10**, 159–282 (2015).

[15] F. G. S. L. Brandão and K. M. Svore, "Quantum Speed-Ups for Solving Semidefinite Programs," in *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)* (2017) pp. 415–426.

[16] J. Van Apeldoorn, A. Gilyén, S. Gribling, and R. de Wolf, "Quantum SDP-Solvers: Better Upper and Lower Bounds," in *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)* (2017) pp. 403–414.

[17] Fernando GSL Brandao, Amir Kalev, Tongyang Li, Cedric Yen-Yu Lin, Krysta M Svore, and Xiaodi Wu, "Quantum SDP solvers: Large speed-ups, optimality, and applications to quantum learning," arXiv preprint arXiv:1710.02581 arXiv:1710.02581.

[18] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd, "Quantum machine learning," Nature **549**, 195 (2017).

[19] Mohammad H. Amin, Evgeny Andriyash, Jason Rolfe, Bohdan Kulchytskyy, and Roger Melko, "Quantum Boltzmann Machine," Phys. Rev. X **8**, 021050 (2018).

[20] Mária Kieferová and Nathan Wiebe, "Tomography and generative training with quantum Boltzmann machines," Phys. Rev. A **96**, 062327 (2017).

[21] Hai-Jing Song, Tieling Song, Qi-Kai He, Yang Liu, and D. L. Zhou, "Geometry and symmetry in the quantum Boltzmann machine," Phys. Rev. A **99**, 042307 (2019).

[22] Daniel Crawford, Anna Levit, Navid Ghadermarzy, Jaspreet S Oberoi, and Pooya Ronagh, "Reinforcement learning using quantum Boltzmann machines," arXiv preprint arXiv:1612.05695 (2016), arXiv:1612.05695.

[23] Huzihiro Araki, "Gibbs states of a one dimensional quantum lattice," Communications in Mathematical

Physics **14**, 120–157 (1969).

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

[25] 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**, 223–271 (1995).

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

[27] M. Kliesch, C. Gogolin, M. J. Kastoryano, A. Riera, and J. Eisert, "Locality of Temperature," Phys. Rev. X **4**, 031019 (2014).

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

[29] Michael M. Wolf, Frank Verstraete, Matthew B. Hastings, and J. Ignacio Cirac, "Area Laws in Quantum Systems: Mutual Information and Correlations," Phys. Rev. Lett. **100**, 070502 (2008).

[30] M. B. Hastings, "Solving gapped Hamiltonians locally," Phys. Rev. B **73**, 085115 (2006).

[31] Andras Molnar, Norbert Schuch, Frank Verstraete, and J. Ignacio Cirac, "Approximating Gibbs states of local Hamiltonians efficiently with projected entangled pair states," Phys. Rev. B **91**, 045138 (2015).

[32] John M Hammersley and Peter Clifford, "Markov fields on finite graphs and lattices," Unpublished manuscript **46** (1971).

[33] Winton Brown and David Poulin, "Quantum Markov networks and commuting Hamiltonians," arXiv preprint arXiv:1206.0755 (2012), arXiv:1206.0755.

[34] Farzad Ghafari Jouneghani, Mohammad Babazadeh, Rogayeh Bayramzadeh, and Hossein Movla, "Investigation of Commuting Hamiltonian in Quantum Markov Network," International Journal of Theoretical Physics **53**, 2521–2530 (2014).

[35] Kohtaro Kato and Fernando G. S. L. Brandão, "Quantum Approximate Markov Chains are Thermal," Communications in Mathematical Physics (2019), 10.1007/s00220-019-03485-6.

[36] Alexei Kitaev and John Preskill, "Topological Entanglement Entropy," Phys. Rev. Lett. **96**, 110404 (2006).

[37] Michael Levin and Xiao-Gang Wen, "Detecting Topological Order in a Ground State Wave Function," Phys. Rev. Lett. **96**, 110405 (2006).

[38] Fernando G. S. L. Brandão and Michael J. Kastoryano, "Finite Correlation Length Implies Efficient Preparation of Quantum Thermal States," Communications in Mathematical Physics **365**, 1–16 (2019).

[39] Patrick Hayden, Richard Jozsa, Dénes Petz, and Andreas Winter, "Structure of States Which Satisfy Strong Subadditivity of Quantum Entropy with Equality," Communications in Mathematical Physics **246**, 359–374 (2004).

[40] Robert Raussendorf and Hans J. Briegel, "A One-Way Quantum Computer," Phys. Rev. Lett. **86**, 5188–5191 (2001).

[41] Dominic J. Williamson, Arpit Dua, and Meng Cheng, "Spurious Topological Entanglement Entropy from Subsystem Symmetries," Phys. Rev. Lett. **122**, 140506 (2019).

[42] Liujun Zou and Jeongwan Haah, "Spurious long-range entanglement and replica correlation length," Phys. Rev. B **94**, 075151 (2016).

[43] For the cluster state on a ring with n as an even integer, let ρ^{V_0} be the reduced density matrix with respect to the odd spins, $V_0 = \{1, 3, 5, \cdots, 2s-1\}$ ($s := n/2$).

Then, we obtain the von Neumann entropy for ρ^{V_0} as $S(\rho^{V_0}) = (|V_0| - 1)\log 2 = (s-1)\log 2$. On the contrary, for arbitrary subset $V_0' \subset V_0$, the reduced density matrix, $\rho^{V_0'}$, is given by the identity matrix, and hence, $S(\rho_{V_0'}) = |V_0'|\log 2$. Therefore, for an arbitrary decomposition $V_0 = ABC$, we have $\mathcal{I}(A:C|B) = \log 2$ and the Markov property breaks down.

[44] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, "Matrix Product State Representations," Quantum Info. Comput. **7**, 401–430 (2007).

[45] Dénes Petz, "Sufficient subalgebras and the relative entropy of states of a von Neumann algebra," Communications in Mathematical Physics **105**, 123–131 (1986).

[46] Omar Fawzi and Renato Renner, "Quantum Conditional Mutual Information and Approximate Markov Chains," Communications in Mathematical Physics **340**, 575–611 (2015).

[47] Fernando G. S. L. Brandão, Aram W. Harrow, Jonathan Oppenheim, and Sergii Strelchuk, "Quantum Conditional Mutual Information, Reconstructed States, and State Redistribution," Phys. Rev. Lett. **115**, 050501 (2015).

[48] Marius Junge, Renato Renner, David Sutter, Mark M. Wilde, and Andreas Winter, "Universal Recovery Maps and Approximate Sufficiency of Quantum Relative Entropy," Annales Henri Poincaré **19**, 2955–2978 (2018).

[49] Supplementary material.

[50] Patrick Hayden, Debbie Leung, Peter W. Shor, and Andreas Winter, "Randomizing Quantum States: Constructions and Applications," Communications in Mathematical Physics **250**, 371–391 (2004).

[51] M. B. Hastings, "Entropy and entanglement in quantum ground states," Phys. Rev. B **76**, 035114 (2007).

[52] Adriano Barenco, Charles H. Bennett, Richard Cleve, David P. DiVincenzo, Norman Margolus, Peter Shor, Tycho Sleator, John A. Smolin, and Harald Weinfurter, "Elementary gates for quantum computation," Phys. Rev. A **52**, 3457–3467 (1995).

[53] Juha J. Vartiainen, Mikko Möttönen, and Martti M. Salomaa, "Efficient Decomposition of Quantum Gates," Phys. Rev. Lett. **92**, 177902 (2004).

[54] David Poulin and Pawel Wocjan, "Sampling from the Thermal Quantum Gibbs State and Evaluating Partition Functions with a Quantum Computer," Phys. Rev. Lett. **103**, 220502 (2009).

[55] Ersen Bilgin and Sergio Boixo, "Preparing Thermal States of Quantum Systems by Dimension Reduction," Phys. Rev. Lett. **105**, 170405 (2010).

[56] Michael J. Kastoryano and Fernando G. S. L. Brandão, "Quantum Gibbs Samplers: The Commuting Case," Communications in Mathematical Physics **344**, 915–957 (2016).

[57] Yimin Ge, András Molnár, and J. Ignacio Cirac, "Rapid Adiabatic Preparation of Injective Projected Entangled Pair States and Gibbs States," Phys. Rev. Lett. **116**, 080503 (2016).

[58] When Φ_L is given for all $L \subset V$ with $|L| \leq k$, we can compute the average energy $\mathrm{tr}(\rho H)$ in polynomial time. However, the computational complexity of calculating the average energy is known to be NP-hard [88].

[59] Ersen Bilgin and David Poulin, "Coarse-grained belief propagation for simulation of interacting quantum systems at all temperatures," Phys. Rev. B **81**, 054106 (2010).

[60] Bo Yan, Steven A Moses, Bryce Gadway, Jacob P Covey, Kaden RA Hazzard, Ana Maria Rey, Deborah S Jin, and Jun Ye, "Observation of dipolar spin-exchange interactions with lattice-confined polar molecules," Nature **501**, 521 (2013).

[61] Philip Richerme, Zhe-Xuan Gong, Aaron Lee, Crys-

tal Senko, Jacob Smith, Michael Foss-Feig, Spyridon Michalakis, Alexey V Gorshkov, and Christopher Monroe, "Non-local propagation of correlations in quantum systems with long-range interactions," Nature **511**, 198 (2014).

[62] Petar Jurcevic, Ben P Lanyon, Philipp Hauke, Cornelius Hempel, Peter Zoller, Rainer Blatt, and Christian F Roos, "Quasiparticle engineering and entanglement propagation in a quantum many-body system," Nature **511**, 202 (2014).

[63] R. Islam, C. Senko, W. C. Campbell, S. Korenblit, J. Smith, A. Lee, E. E. Edwards, C.-C. J. Wang, J. K. Freericks, and C. Monroe, "Emergence and Frustration of Magnetism with Variable-Range Interactions in a Quantum Simulator," Science **340**, 583–587 (2013).

[64] Jiehang Zhang, Guido Pagano, Paul W Hess, Antonis Kyprianidis, Patrick Becker, Harvey Kaplan, Alexey V Gorshkov, Z-X Gong, and Christopher Monroe, "Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator," Nature **551**, 601 (2017).

[65] Thomas Koffel, M. Lewenstein, and Luca Tagliacozzo, "Entanglement Entropy for the Long-Range Ising Chain in a Transverse Field," Phys. Rev. Lett. **109**, 267203 (2012).

[66] Davide Vodola, Luca Lepori, Elisa Ercolessi, Alexey V. Gorshkov, and Guido Pupillo, "Kitaev Chains with Long-Range Pairing," Phys. Rev. Lett. **113**, 156402 (2014).

[67] Tomotaka Kuwahara, "Exponential bound on information spreading induced by quantum many-body dynamics with long-range interactions," New Journal of Physics **18**, 053034 (2016).

[68] Tomotaka Kuwahara, Itai Arad, Luigi Amico, and Vlatko Vedral, "Local reversibility and entanglement structure of many-body ground states," Quantum Science and Technology **2**, 015005 (2017).

[69] Tomotaka Kuwahara and Keiji Saito, "Area law of non-critical ground states in 1D long-range interacting systems," arXiv preprint arXiv:1908.11547 arXiv:1908.11547.

[70] Robert B. Israel and Chiara R. Nappi, "Exponential clustering for long-range integer-spin systems," Communications in Mathematical Physics **68**, 29–37 (1979).

[71] Camillo Cammarota, "Decay of correlations for infinite range interactions in unbounded spin systems," Communications in Mathematical Physics **85**, 517–528 (1982).

[72] Herbert Spohn and Wilhelm Zwerger, "Decay of the Two-Point Function in One-Dimensional O(N) Spin Models with Long-Range Interactions," Journal of Statistical Physics **94**, 1037–1043 (1999).

[73] Aldo Procacci and Benedetto Scoppola, "On Decay of Correlations for Unbounded Spin Systems with Arbitrary Boundary Conditions," Journal of Statistical Physics **105**, 453–482 (2001).

[74] Alina Kargol, "Decay of correlations in multicomponent

ferromagnets with long-range interaction," Reports on Mathematical Physics **56**, 379 – 386 (2005).

[75] Georg Menz and Robin Nittka, "Decay of Correlations in 1D Lattice Systems of Continuous Spins and Long-Range Interaction," Journal of Statistical Physics **156**, 239–267 (2014).

[76] Alina Kargol, "Decay of correlations in N-vector ferromagnetic quantum models with long-range interactions," Journal of Statistical Mechanics: Theory and Experiment **2014**, P10006 (2014).

[77] Senaida Hernández-Santana, Christian Gogolin, J. Ignacio Cirac, and Antonio Acín, "Correlation Decay in Fermionic Lattice Systems with Power-Law Interactions at Nonzero Temperature," Phys. Rev. Lett. **119**, 110601 (2017).

[78] Claudio Castelnovo and Claudio Chamon, "Entanglement and topological entropy of the toric code at finite temperature," Phys. Rev. B **76**, 184442 (2007).

[79] Claudio Castelnovo and Claudio Chamon, "Topological order in a three-dimensional toric code at finite temperature," Phys. Rev. B **78**, 155120 (2008).

[80] Mehdi Kargarian, "Finite-temperature topological order in two-dimensional topological color codes," Phys. Rev. A **80**, 012321 (2009).

[81] S. Iblisdir, D. Pérez-García, M. Aguado, and J. Pachos, "Scaling law for topologically ordered systems at finite temperature," Phys. Rev. B **79**, 134303 (2009).

[82] Dalimil Mazáč and Alioscia Hamma, "Topological order, entanglement, and quantum memory at finite temperature," Annals of Physics **327**, 2096 – 2111 (2012).

[83] Matthew B. Hastings, Grant H. Watson, and Roger G. Melko, "Self-Correcting Quantum Memories Beyond the Percolation Threshold," Phys. Rev. Lett. **112**, 070501 (2014).

[84] Olivier Landon-Cardinal, Beni Yoshida, David Poulin, and John Preskill, "Perturbative instability of quantum memory based on effective long-range interactions," Phys. Rev. A **91**, 032303 (2015).

[85] Mario Berta, Kaushik P Seshadreesan, and Mark M Wilde, "Rényi generalizations of the conditional quantum mutual information," Journal of Mathematical Physics **56**, 022205 (2015), https://doi.org/10.1063/1.4908102.

[86] According to Theorem 3, the νth power of the reduced density matrix ρ^L is given by $e^{-\nu\beta(H_L+\Phi_L)}/Z^\nu$, where Φ_L is defined in Eq. (14). Thus, the state $[\rho^L]^\nu$ has an effective temperature of $\nu\beta$. Hence, we expect that we can utilize the same analytical techniques for $\nu\beta \lesssim \beta_c$.

[87] Aram Harrow, Saeed Mehraban, and Mehdi Soleimanifar, "Classical algorithms, correlation decay, and complex zeros of partition functions of quantum many-body systems," to appear in arXiv (2019).

[88] F Barahona, "On the computational complexity of Ising spin glass models," Journal of Physics A: Mathematical and General **15**, 3241–3253 (1982).

Appendix A: Proof of Theorem 1

1. Preliminaries

We here recall the setup. We consider a quantum spin system with n spins, where each of the spin sits on a vertex of the graph $G = (V, E)$ with V the total spin set ($|V| = n$). For a partial set L of spins, we denote the cardinality, that is, the number of vertices contained in L, by $|L|$ (e.g. $L = \{i_1, i_2, \ldots, i_{|L|}\}$). We also denote the complementary subset of L by $L^c := V \setminus L$. We denote the local Hilbert space by $\mathcal{H}^v$ ($v \in V$) with $\dim(\mathcal{H}^v) = d$

and the entire Hilbert space is given by $\mathcal{H} := \bigotimes_{v \in V} \mathcal{H}^v$ with $\dim(\mathcal{H}) = d^n$. We also define the local Hilbert space of the subset $L \subset V$ as $\mathcal{H}^L$ and denote the dimension by d_L, namely $d_L := d^{|L|}$. We define $\mathcal{B}(\mathcal{H})$ as the space of bounded linear operators on $\mathcal{H}$.

When we consider a reduced operator on a subsystem L, we denote it as

$$O^L = \text{tr}_{L^c}(O) \otimes \hat{1}_{L^c} \in \mathcal{B}(\mathcal{H}) \tag{A1}$$

by using the superscript index, where $\hat{1}$ is the identity operator and tr_{L^c} is the partial trace operation with respect to the Hilbert space $\mathcal{H}^{L^c}$.

We also define the following set:

$$E^{(x)} := \{X \subset V | \text{diam}(X) = x, \quad |X| \le k\} \tag{A2}$$

with

$$\text{diam}(X) := \max_{v_1, v_2 \in X} d_{v_1, v_2}, \tag{A3}$$

where we defined $d_{A,B}$ as the shortest path length via E which connects A and B ($A \subset V$, $B \subset V$).

In the setup of Theorem 1, we consider the Hamiltonian as

$$H = \sum_{X \in E_r} h_X, \quad \text{with} \quad \sum_{X | X \ni v} \|h_X\| \le 1 \quad \text{for} \quad \forall v \in V \tag{A4}$$

with

$$E_r := E^{(1)} \sqcup E^{(2)} \sqcup \cdots \sqcup E^{(r)} \quad (r \in \mathbb{N}). \tag{A5}$$

Here, the Hamiltonian (A4) describes an arbitrary k-body interacting systems with finite interaction length r.

Throughout the manuscript, we denote the natural logarithm by $\log(\cdot)$ for the simplicity, namely $\log(\cdot) = \log_e(\cdot)$.

a. Cluster notation

We then define several basic terminologies. On the graph (V, E), we call a multiset of subsystems $w = \{X_1, X_2, \ldots, X_{|w|}\}$ ($X_j \in E_r$ for $j = 1, 2, \ldots, |w|$) as "cluster", where $|w|$ is the cardinality of w. Note that each of the elements $\{X_j\}_{j=1}^{|w|}$ satisfies $\text{diam}(X_j) \le r$ from the definition (A5). We denote $\mathcal{C}_{r,m}$ by the set of w with $|w| = m$ and let $V_w \subseteq V$ and $E_w \subseteq E_r$ be the set of different vertices (or spins) and subsystems which are contained in w, respectively. Also, we define connected clusters as follows:

Definition 1. (Connected cluster) *For a cluster $w \in \mathcal{C}_{r,|w|}$, we say that w is a connected cluster if there are no decompositions of $w = w_1 \sqcup w_2$ such that $V_{w_1} \cap V_{w_2} = \emptyset$. We denote by $\mathcal{G}_{r,m}$ the set of the connected clusters with $|w| = m$.*

Definition 2. (Connected cluster to a region, FIG. 4) *Similarly, we say that $w \in \mathcal{C}_{r,|w|}$ is a connected cluster to a subsystem L if there are no decompositions of $w = w_1 \sqcup w_2$ such that $(L \cup V_{w_1}) \cap V_{w_2} = \emptyset$. We denote by $\mathcal{G}_{r,m}^L$ the set of the connected clusters to L with $|w| = m$.*

Definition 3. (Connected cluster with a link between two regions, FIG. 5) *Finally, for a connected cluster $w \in \mathcal{G}_{r,|w|}$, we say that w has links between A and B if there exist a path from A to B in E_w. We denote by $\mathcal{G}_{r,m}^{A,B}$ the set of the connected clusters with $|w| = m$ which have a link A and B.*

b. Basic lemmas for logarithmic operators

Before going to the proof, we prove the following basic lemmas:

Lemma 6. *Let $O \in \mathcal{B}(\mathcal{H})$ be an arbitrary non-negative operator written as*

$$O = \Gamma_{L_1} \otimes \Gamma_{L_2} \otimes \cdots \otimes \Gamma_{L_m}, \tag{A6}$$

where $\{\Gamma_{L_j}\}_{j=1}^L \in \mathcal{B}(\mathcal{H})$ are supported on the subsystems $\{L_j\}_{j=1}^m$, respectively and we assume $L_1 \sqcup L_2 \sqcup \cdots \sqcup L_m = V$. Then, for arbitrary subsystems $A, B, C \subset V$, we have

$$\log O^{AB} + \log O^{BC} - \log O^{ABC} - \log O^B = \sum_{j=1}^m (\log \Gamma_{L_j}^{AB} + \log \Gamma_{L_j}^{BC} - \log \Gamma_{L_j}^{ABC} - \log \Gamma_{L_j}^B). \tag{A7}$$

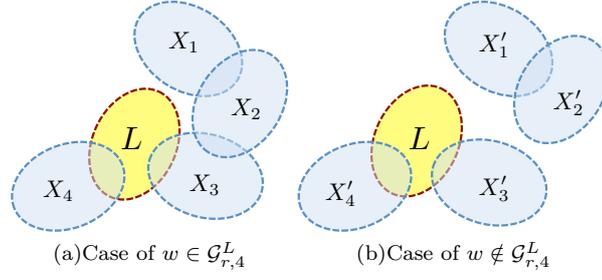*Note that $\{O^{AB}, O^{BC}, O^{ABC}, O^B\}$ are reduced operators as defined in Eq. (A1).*

FIG. 4. Schematic pictures of clusters of $w \in \mathcal{G}_4^L$ and $w \notin \mathcal{G}_4^L$. Each of the elements $\{X_s | X_s \in E_r\}$ is a subset of the total set V (i.e., $X \subset V$). In (a), there there are no decompositions of $w = w_1 \sqcup w_2$ such that $(L \cup V_{w_1}) \cap V_{w_2} = \emptyset$ for $w = \{X_1, X_2, X_3, X_4\}$, whereas in (b) the decomposition $w' = w'_1 \sqcup w'_2$ with $w'_1 = \{X'_2, X'_3\}$ and $w'_2 = \{X'_1, X'_4\}$ satisfies $(L \cup V_{w_1}) \cap V_{w_2} = \emptyset$.
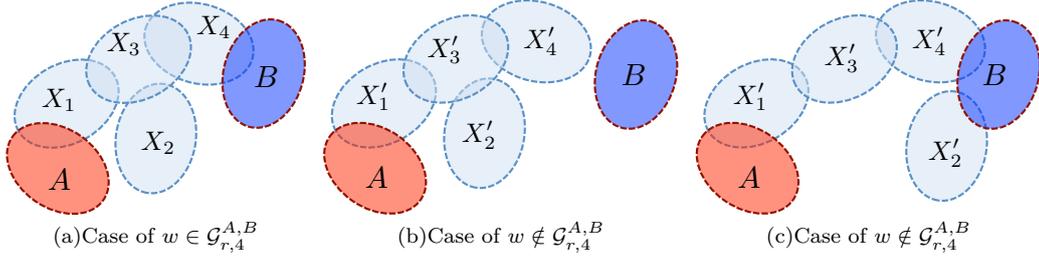

FIG. 5. Schematic pictures of clusters of $w \in \mathcal{G}_{r,4}^{A,B}$ and $w \notin \mathcal{G}_{r,4}^{A,B}$. In (a), subsystems A and B are connected with each other by the cluster w. On the other hand, in (b), the cluster w does not have a link between A and B, and in (c), the cluster has the link but is not connected.

Proof of Lemma 6. We define $A_j := A \cap L_j$, $B_j := B \cap L_j$, and $C_j := C \cap L_j$ for $j = 1, 2, \ldots m$. We notice that $\bigsqcup_{j=1}^m A_j = A$, $\bigsqcup_{j=1}^m B_j = B$ and $\bigsqcup_{j=1}^m C_j = C$ because of $\bigsqcup_{j=1}^m L_j = V$. Then, from the definition (A1), the reduced operator of O with respect to the subsystem B is given by

$$O^B = \mathrm{tr}_{L_1 \backslash B_1} (\Gamma_{L_1}) \otimes \mathrm{tr}_{L_2 \backslash B_2} (\Gamma_{L_2}) \otimes \cdots \otimes \mathrm{tr}_{L_m \backslash B_m} (\Gamma_{L_m}) \otimes \hat{1}_{B^c}$$
$$= \Gamma_{B_1} \otimes \Gamma_{B_2} \otimes \cdots \otimes \Gamma_{B_m} \otimes \hat{1}_{B^c}, \tag{A8}$$

where $\Gamma_{B_j} := \mathrm{tr}_{L_j \backslash B_j} \left(\Gamma_{L_j} \right) \otimes \hat{1}_{B_j^c}$ for $j = 1, 2, \ldots, m$. We define Γ_j^{BC}, Γ_j^{ABC} and Γ_j^B in the same way. We thus obtain

$$\log O^B = \sum_{j=1}^m \log \Gamma_{B_j}. \tag{A9}$$

On the other hand, we have from the definition (A1)

$$\Gamma_{L_j}^B = \mathrm{tr}_{B^c} (\Gamma_{L_j}) \otimes \hat{1}_{B^c} = d^{|B^c| - |L_j \backslash B_j|} \Gamma_{B_j} = d^{n - |B| - |L_j| + |B_j|} \Gamma_{B_j}, \tag{A10}$$

which reduces Eq. (A9) to

$$\log O^B = -(m-1)(n - |B|) \log(d) + \sum_{j=1}^m \log \Gamma_{L_j}^B. \tag{A11}$$

We obtain the similar form to Eq. (A11) for O^{AB}, O^{BC} and O^{ABC}. After a straightforward calculation, we prove the equation (A7). $\square$

Second, we prove the following lemma:

Lemma 7. *For an arbitrary non-negative operator $O \in \mathcal{B}(\mathcal{H})$ which is given by the form of*

$$O = O_L \otimes \hat{1}_{L^c} \tag{A12}$$

with $L \cap C = \emptyset$, we have

$$\log O^{AB} + \log O^{BC} - \log O^{ABC} - \log O^B = 0. \tag{A13}$$

Proof of Lemma 7. From the definition, we obtain

$$O^{ABC} = O^{AB} \otimes \hat{1}_C. \tag{A14}$$

Thus, we obtain $\log O^{BC} = \log(O^B \otimes \hat{1}_C)$ and $\log O^{ABC} = \log O^{AB}$, and hence we immediately obtain Eq. (A13). This completes the proof. $\square$

2. Generalized cluster Expansion

We first parametrize H by using a parameter set $\vec{a} := \{a_X\}_{X \in E_r}$ as

$$H_{\vec{a}} = \sum_{X \in E_r} a_X h_X, \tag{A15}$$

where $H = H_{\vec{1}}$ with $\vec{1} = \{1, 1, \ldots, 1\}$. Note that there are $|E_r|$ parameters in total. By using Eq. (A15), we define a parametrized Gibbs state $\rho_{\vec{a}}$ as

$$\rho_{\vec{a}} := \frac{e^{-\beta H_{\vec{a}}}}{Z_{\vec{a}}}, \tag{A16}$$

where $Z_{\vec{a}} := \mathrm{tr}(e^{-\beta H_{\vec{a}}})$.

In the standard cluster expansion, we consider the Taylor expansion of $e^{-\beta H_{\vec{a}}}$ with respect to the parameters $\vec{a}$. It works well in analyzing a correlation function or tensor network representation, while it is not appropriate to analyze the entropy or effective Hamiltonian of a reduced density matrix. To overcome it, we generalize the standard cluster expansion. We parametrize a target function of interest by $f_{\vec{a}}$ and directly expand it with respect to $\vec{a}$, where $f_{\vec{a}}$ can be chosen not only as a scholar function but also as a operator function. Here, we choose the conditional mutual information as the function $f_{\vec{a}}$. By using $\rho_{\vec{a}}$, we parameterize the conditional mutual information by $\mathcal{I}_{\vec{a}}(A : C|B)$ in the following form:

$$\begin{aligned}
\mathcal{I}_{\vec{a}}(A : C|B) &= -\mathrm{tr}\left[\rho\left(\log \rho_{\vec{a}}^{AB} + \log \rho_{\vec{a}}^{BC} - \log \rho_{\vec{a}}^{ABC} - \log \rho_{\vec{a}}^{B}\right)\right] \\
&= -\mathrm{tr}\left[\rho\left(\log \tilde{\rho}_{\vec{a}}^{AB} + \log \tilde{\rho}_{\vec{a}}^{BC} - \log \tilde{\rho}_{\vec{a}}^{ABC} - \log \tilde{\rho}_{\vec{a}}^{B}\right)\right],
\end{aligned} \tag{A17}$$

where $\rho = \rho_{\vec{1}}$ and we define $\tilde{\rho}_{\vec{a}}$ as

$$\tilde{\rho}_{\vec{a}} := e^{-\beta H_{\vec{a}}} \tag{A18}$$

with

$$\tilde{\rho}_{\vec{a}}^{L} = \left(e^{-\beta H_{\vec{a}}}\right)^{L} = \mathrm{tr}_{L^c}\left(e^{-\beta H_{\vec{a}}}\right) \otimes \hat{1}_{L^c}. \tag{A19}$$

Note that we use the definition (A1) for $\tilde{\rho}_{\vec{a}}^{L}$ $(L \subset V)$

In the following, we define

$$\tilde{H}_{\vec{a}}(A : C|B) := \log \tilde{\rho}_{\vec{a}}^{AB} + \log \tilde{\rho}_{\vec{a}}^{BC} - \log \tilde{\rho}_{\vec{a}}^{ABC} - \log \tilde{\rho}_{\vec{a}}^{B}, \tag{A20}$$

which gives

$$\mathcal{I}_{\vec{a}}(A : C|B) = \mathrm{tr}\left[\rho \tilde{H}_{\vec{a}}(A : C|B)\right] \le \|\tilde{H}_{\vec{a}}(A : C|B)\|. \tag{A21}$$

Then, the Taylor expansion with respect to $\vec{a}$ to the operator $\tilde{H}_{\vec{a}}(A : C|B)$ reads

$$\tilde{H}_{\vec{1}}(A : C|B) = \sum_{m=0}^{\infty} \frac{1}{m!}\left[\left(\sum_{X \in E_r} \frac{\partial}{\partial a_X}\right)^m \tilde{H}_{\vec{a}}(A : C|B)\right]_{\vec{a}=\vec{0}}, \tag{A22}$$

where $\vec{0} = \{0, 0, \ldots, 0\}$. By using the cluster notation, we obtain

$$\sum_{X_1, X_2, \ldots, X_m \in E_r} = \sum_{w \in \mathcal{C}_{r,m}} n_w, \tag{A23}$$

which yields

$$\tilde{H}_{\vec{1}}(A : C|B) = \sum_{m=1}^{\infty} \frac{1}{m!} \sum_{X_1, X_2, \ldots, X_m \in E_r} \prod_{j=1}^{m} \frac{\partial}{\partial a_{X_j}} \tilde{H}_{\vec{a}}(A : C|B)\Big|_{\vec{a}=\vec{0}} = \sum_{m=1}^{\infty} \frac{1}{m!} \sum_{w \in \mathcal{C}_{r,m}} n_w \mathcal{D}_w \tilde{H}_{\vec{a}}(A : C|B)\Big|_{\vec{a}=\vec{0}}, \tag{A24}$$

where $w = \{X_1, X_2, \ldots, X_m\}$ and n_w is the multiplicity that w appears in the summation, and we defined

$$\mathcal{D}_w := \prod_{j=1}^{m} \frac{\partial}{\partial a_{X_j}} \quad \text{with} \quad w = \{X_1, X_2, \ldots, X_m\}. \tag{A25}$$

We notice that the partial derivatives $\frac{\partial}{\partial a_X}$ and $\frac{\partial}{\partial a_{X'}}$ commute with each other because $\log(\tilde{\rho}_{\vec{a}}^{L})$ is a C^{∞}-smooth function with respect to $\vec{a}$ as long as the system size n is finite. The C^{∞}-smoothness of $\log(\tilde{\rho}_{\vec{a}}^{L})$ is proved as follows:

For a finite system size n, the C^∞-smoothness of $e^{-\beta H_{\vec{a}}}$ is ensured, and hence $\tilde{\rho}_{\vec{a}}^L$ is also C^∞-smooth from the definition (A19). Also, we can set

$$\|e^{-\tau\hat{1}}\tilde{\rho}_{\vec{a}}^L\| \le 1. \tag{A26}$$

by choosing a finite energy $\tau < \infty$ appropriately. Notice that $e^{-\tau\hat{1}}\tilde{\rho}_{\vec{a}}^L$ is Hermitian and $e^{-\tau\hat{1}}\tilde{\rho}_{\vec{a}}^L \succeq 0$. This implies the absolute convergence of the following expansion:

$$\log(\tilde{\rho}_{\vec{a}}^L) = \tau\hat{1} + \log(e^{-\tau\hat{1}}\tilde{\rho}_{\vec{a}}^L) = \tau\hat{1} + \log(\hat{1} + e^{-\tau\hat{1}}\tilde{\rho}_{\vec{a}}^L - \hat{1}) = \tau\hat{1} + \sum_{m=1}^{\infty} \frac{(-1)^{m-1}}{m}(e^{-\tau\hat{1}}\tilde{\rho}_{\vec{a}}^L - \hat{1})^m. \tag{A27}$$

Thus, the C^∞-smoothness of $\tilde{\rho}_{\vec{a}}^L$ implies of C^∞-smoothness of $\log(\tilde{\rho}_{\vec{a}}^L)$.

Note that the case of $m = 0$ (i.e., $|w| = 0$) does not contribute to the expansion because of $\tilde{H}_{\vec{0}}(A : C|B) = 0$. In order to calculate the summation of $\sum_{w \in \mathcal{C}_{r,m}}$, we utilize the following proposition:

Proposition 8. *The cluster expansion* (A24) *reduces to the summation of connected clusters which have links between A and C:*

$$\tilde{H}_{\vec{1}}(A : C|B) = \sum_{m=1}^{\infty} \frac{1}{m!} \sum_{w \in \mathcal{G}_{r,m}^{A,C}} n_w \mathcal{D}_w \tilde{H}_{\vec{a}}(A : C|B)\Big|_{\vec{a}=\vec{0}}, \tag{A28}$$

where the definition of $\mathcal{G}_{r,m}^{A,C}$ has been given in Def. 3.

From this proposition, we only need to estimate the contribution of clusters in $\mathcal{G}_{r,m}^{A,C}$ to upper-bound the conditional mutual information $\mathcal{I}_{\vec{1}}(A : C|B) = \mathrm{tr}[\rho\tilde{H}_{\vec{1}}(A : C|B)]$.

a. Proof of Proposition 8

We first introduce the notation $\vec{a}_w$ as a parameter vector such that the elements $\{a_X\}_{X \notin w}$ are vanishing, that is,

$$(\vec{a}_w)_X = 0 \quad \text{for} \quad X \notin w, \tag{A29}$$

where we denote an element of a_X in $\vec{a}$ by $(\vec{a})_X$. We then obtain

$$\mathcal{D}_w \tilde{H}_{\vec{a}}(A : C|B)\Big|_{\vec{a}=\vec{0}} = \mathcal{D}_w \tilde{H}_{\vec{a}_w}(A : C|B)\Big|_{\vec{a}_w=\vec{0}}. \tag{A30}$$

In the following, we aim to prove

$$\mathcal{D}_w \tilde{H}_{\vec{a}_w}(A : C|B)\Big|_{\vec{a}_w=\vec{0}} = 0 \quad \text{for} \quad w \notin \mathcal{G}_{r,|w|}^{A,C}. \tag{A31}$$

We notice that if $w \notin \mathcal{G}_{r,|w|}^{A,C}$ the cluster w satisfies either one of the following two properties (see Figs. 5 (b) and (c)):

$$L_w \cap A = \emptyset \quad \text{or} \quad L_w \cap C = \emptyset \tag{A32}$$

and

$$w \notin \mathcal{G}_{r,|w|}. \tag{A33}$$

In the first case (A32), we can immediately obtain $\tilde{H}_{\vec{a}_w}(A : C|B) = 0$ by choosing $O = e^{-\beta H_{\vec{a}_w}}$ in the lemma 7. In the second case (A33), there exists a decomposition of $w = w_1 \sqcup w_2$ ($|w_1|, |w_2| > 0$) such that $V_{w_1} \cap V_{w_2} = \emptyset$. Hence, we have $e^{-\beta H_{\vec{a}_w}} = e^{-\beta H_{\vec{a}_{w_1}}} \otimes e^{-\beta H_{\vec{a}_{w_2}}}$, and from Lemma 6 we obtain

$$\tilde{H}_{\vec{a}_w}(A : C|B) = \tilde{H}_{\vec{a}_{w_1}}(A : C|B) + \tilde{H}_{\vec{a}_{w_2}}(A : C|B). \tag{A34}$$

Because of $\mathcal{D}_{w_2}\tilde{H}_{\vec{a}_{w_1}}(A : C|B) = \mathcal{D}_{w_1}\tilde{H}_{\vec{a}_{w_2}}(A : C|B) = 0$, we have $\mathcal{D}_w\tilde{H}_{\vec{a}_w}(A : C|B) = 0$. This completes the proof of Proposition 8. $\square$

―――――――――――――――――――――――――――――――――[**End of Proof of Proposition 8**]

3.   Estimation of the expanded terms

In order to estimate the summation (A28) with respect to $\sum_{w \in \mathcal{G}_{r,m}^{A,C}}$, we consider a derivative of

$$\mathcal{D}_w \log \tilde{\rho}_{\vec{a}}^L \Big|_{\vec{a}=\vec{0}} = \mathcal{D}_w \log \tilde{\rho}_{\vec{a}_w}^L \Big|_{\vec{a}_w=\vec{0}} \tag{A35}$$

for an arbitrary subsystem $L \subset V$. We choose the subsets AB, BC, ABC and B as L afterward. We here give an explicit form of the derivative $\mathcal{D}_w \log \tilde{\rho}_{\vec{a}}^L$ in the following proposition 9.

Proposition 9. *Let us take $m-1$ copies of the partial Hilbert space $\mathcal{H}^{L^c}$ and distinguish them by $\{\mathcal{H}_j^{L^c}\}_{j=1}^m$. Then, we define the extended Hilbert space as $\mathcal{H}^L \otimes \mathcal{H}_{1:m}^{L^c}$ with*

$$\mathcal{H}_{1:m}^{L^c} := \mathcal{H}_1^{L^c} \otimes \mathcal{H}_2^{L^c} \otimes \cdots \otimes \mathcal{H}_m^{L^c}. \tag{A36}$$

Then, for an arbitrary operator $O \in \mathcal{H}$, we extend the domain of definition and denote $O_{\tilde{\mathcal{H}}_s} \in \mathcal{B}(\mathcal{H}^L \otimes \mathcal{H}_{1:m}^{L^c})$ by the operator which acts only on the space $\mathcal{H}^L \otimes \mathcal{H}_s^{L^c}$. Now, for an arbitrary cluster $w = \{X_1, X_2, \ldots, X_m\}$, we have

$$\mathcal{D}_w \log \tilde{\rho}_{\vec{a}}^L \Big|_{\vec{a}=\vec{0}} = \frac{(-\beta)^m}{m! d_{L^c}^m} \mathcal{P}_m \mathrm{tr}_{L_{1:m}^c} \left(\tilde{h}_{X_1}^{(0)} \tilde{h}_{X_2}^{(1)} \cdots \tilde{h}_{X_m}^{(m-1)} \right), \tag{A37}$$

where $\mathrm{tr}_{L_{1:m}^c}$ denotes the partial trace with respect to the Hilbert space $\mathcal{H}_{1:m}^{L^c}$ and we define

$$\tilde{O}^{(0)} := O_{\tilde{\mathcal{H}}_1}, \quad \tilde{O}^{(s)} := O_{\tilde{\mathcal{H}}_1} + O_{\tilde{\mathcal{H}}_2} + \cdots + O_{\tilde{\mathcal{H}}_s} - s O_{\tilde{\mathcal{H}}_{s+1}} \tag{A38}$$

for $s = 1, 2, \ldots, m$. Note that $\mathcal{P}_m$ is the symmetrization operator as

$$\mathcal{P}_m \tilde{h}_{X_1}^{(0)} \tilde{h}_{X_2}^{(1)} \cdots \tilde{h}_{X_m}^{(m-1)} = \sum_\sigma \tilde{h}_{X_{\sigma_1}}^{(0)} \tilde{h}_{X_{\sigma_2}}^{(1)} \cdots \tilde{h}_{X_{\sigma_m}}^{(m-1)}, \tag{A39}$$

where $\sum_\sigma$ denotes the summation of $m!$ terms which come from all the permutations.

a.   Proof of Proposition 9

For the proof, we consider the Taylor expansion with respect to β:

$$\log \tilde{\rho}_{\vec{a}}^L = \sum_{m=0}^\infty \frac{\beta^m}{m!} \frac{\partial^m}{\partial \beta^m} \log \tilde{\rho}_{\vec{a}}^L \Big|_{\beta=0}. \tag{A40}$$

Next, because of

$$\frac{\partial^m}{\partial \beta^m} \log(d_{L^c}) = 0 \quad \text{for} \quad m \geq 1, \tag{A41}$$

we have

$$\frac{\partial^m}{\partial \beta^m} \log \tilde{\rho}_{\vec{a}}^L \Big|_{\beta=0} = \frac{\partial^m}{\partial \beta^m} \log \left[\mathrm{tr}_{L^c} (e^{-\beta H_{\vec{a}}} / d_{L^c}) \right] \Big|_{\beta=0} \tag{A42}$$

for $m \geq 1$.

We aim to prove the following lemma which gives the explicit form of the derivatives with respect to β:

Lemma 10. *The derivatives of $\log \tilde{\rho}_{\vec{a}}^L$ with respect to β can be written as*

$$\frac{\partial^m}{\partial \beta^m} \log \left[\mathrm{tr}_{L^c} (e^{-\beta H_{\vec{a}}} / d_{L^c}) \right] \Big|_{\beta=0} = \frac{(-1)^m}{d_{L^c}^m} \mathrm{tr}_{L_{1:m}^c} \left(\tilde{H}_{\vec{a}}^{(0)} \tilde{H}_{\vec{a}}^{(1)} \cdots \tilde{H}_{\vec{a}}^{(m-1)} \right), \tag{A43}$$

where the definitions of $\tilde{H}_{\vec{a}}^{(s)}$ ($s = 0, 1, 2, \ldots, m-1$) and $\mathcal{H}_{1:m}^{L^c}$ have been given in Eqs. (A38) and Eq. (A36), respectively. We give the proof of the lemma afterward.

By assuming the above lemma, we can prove Eq. (A37) as follows. In considering $\mathcal{D}_w \log \tilde{\rho}_{\vec{a}}^L \big|_{\vec{a}=\vec{0}}$ with $|w| = m$, only the mth order terms of β in the expansion (A40) contribute to the derivative. Hence, we have

$$\mathcal{D}_w \log \tilde{\rho}_{\vec{a}}^L \Big|_{\vec{a}=\vec{0}} = \frac{\beta^m}{m!} \mathcal{D}_w \left(\frac{\partial^m}{\partial \beta^m} \log \left[\mathrm{tr}_{L^c} (e^{-\beta H_{\vec{a}}} / d_{L^c}) \right] \right) \Big|_{\beta=0}. \tag{A44}$$

By combining Eqs. (A35), (A43) and (A44), we have

$$\mathcal{D}_w \log \tilde{\rho}_{\vec{a}}^L \big|_{\vec{a}=0} = \frac{(-\beta)^m}{m!} \frac{1}{d_{L^c}^m} \mathcal{D}_w \mathrm{tr}_{L_{1:m}^c} \left(\tilde{H}_{\vec{a}}^{(0)} \tilde{H}_{\vec{a}}^{(1)} \cdots \tilde{H}_{\vec{a}}^{(m-1)} \right)$$

$$= \frac{(-\beta)^m}{m!} \frac{1}{d_{L^c}^m} \mathcal{P}_m \mathrm{tr}_{L_{1:m}^c} \left(\tilde{h}_{X_1}^{(0)} \tilde{h}_{X_2}^{(1)} \cdots \tilde{h}_{X_m}^{(m-1)} \right). \tag{A45}$$

We therefore obtain Eq. (A37) in Proposition 9. This completes the proof. □

[**Proof of Lemma 10**] In order to prove Eq. (A43), we first expand $\log \left[\mathrm{tr}_{L^c} (e^{-\beta H_{\vec{a}}}/d_{L^c}) \right]$ as follows:

$$\log \left[\frac{\mathrm{tr}_{L^c}(e^{-\beta H_{\vec{a}}})}{d_{L^c}} \right] = \log \left[\hat{1} + \sum_{m=1}^{\infty} \frac{(-\beta)^m}{m!} \frac{\mathrm{tr}_{L^c}(H_{\vec{a}}^m)}{d_{L^c}} \right] = \sum_{q=1}^{\infty} \frac{(-1)^{q-1}}{q} \left(\sum_{m=1}^{\infty} \frac{(-\beta)^m}{m!} \frac{\mathrm{tr}_{L^c}(H_{\vec{a}}^m)}{d_{L^c}} \right)^q, \tag{A46}$$

where in the first equation we use the fact that 0th term of the expansion gives $\mathrm{tr}_{L^c}(\hat{1}/d_{L^c}) = \hat{1}$. We then pick up the terms of β^m. Because of

$$\left(\sum_{m=1}^{\infty} \frac{(-\beta)^m}{m!} \frac{\mathrm{tr}_{L^c}(H_{\vec{a}}^m)}{d_{L^c}} \right)^q$$

$$= \sum_{m=q}^{\infty} \sum_{\substack{m_1+m_2+\cdots+m_q=m \\ m_1 \geq 1, m_2 \geq 1, \ldots, m_q \geq 1}} \frac{(-\beta)^{m_1+m_2+\cdots+m_q}}{m_1! m_2! \cdots m_q!} \frac{\mathrm{tr}_{L^c}(H_{\vec{a}}^{m_1}) \mathrm{tr}_{L^c}(H_{\vec{a}}^{m_2}) \cdots \mathrm{tr}_{L^c}(H_{\vec{a}}^{m_q})}{d_{L^c}^q}, \tag{A47}$$

the mth-order term in Eq. (A46) is given by

$$\beta^m \sum_{q=1}^{m} \frac{(-1)^{q-1}}{q} \sum_{\substack{m_1+m_2+\cdots+m_q=m \\ m_1 \geq 1, m_2 \geq 1, \ldots, m_q \geq 1}} \frac{(-1)^m}{m_1! m_2! \cdots m_q!} \frac{\mathrm{tr}_{L^c}(H_{\vec{a}}^{m_1}) \mathrm{tr}_{L^c}(H_{\vec{a}}^{m_2}) \cdots \mathrm{tr}_{L^c}(H_{\vec{a}}^{m_q})}{d_{L^c}^q}. \tag{A48}$$

We thus obtain

$$\frac{\partial^m}{\partial \beta^m} \log \left[\mathrm{tr}_{L^c}(e^{-\beta H_{\vec{a}}}/d_{L^c}) \right] \Big|_{\beta=0}$$

$$= \sum_{q=1}^{m} \frac{(-1)^{q-1}}{q} \sum_{\substack{m_1+m_2+\cdots+m_q=m \\ m_1 \geq 1, m_2 \geq 1, \ldots, m_q \geq 1}} \frac{m!(-1)^m}{m_1! m_2! \cdots m_q!} \frac{\mathcal{P}_q \mathrm{tr}_{L^c}(H_{\vec{a}}^{m_1}) \mathrm{tr}_{L^c}(H_{\vec{a}}^{m_2}) \cdots \mathrm{tr}_{L^c}(H_{\vec{a}}^{m_q})}{q! d_{L^c}^q}, \tag{A49}$$

where $\mathcal{P}_q$ is the symmetrization operator with respect to $\{m_1, m_2, \ldots, m_q\}$. In the same manner, we can formally expand

$$\frac{(-1)^m}{d_{L^c}^m} \mathrm{tr}_{L_{1:m}^c} \left(\tilde{H}_{\vec{a}}^{(0)} \tilde{H}_{\vec{a}}^{(1)} \cdots \tilde{H}_{\vec{a}}^{(m-1)} \right)$$

$$= \sum_{q=1}^{m} \sum_{\substack{m_1+m_2+\cdots+m_q=m \\ m_1 \geq 1, m_2 \geq 1, \ldots, m_q \geq 1}} \mathcal{C}_{m_1, m_2, \ldots, m_q}^{(q)} \mathcal{P}_q \mathrm{tr}_{L^c}(H_{\vec{a}}^{m_1}) \mathrm{tr}_{L^c}(H_{\vec{a}}^{m_2}) \cdots \mathrm{tr}_{L^c}(H_{\vec{a}}^{m_q}). \tag{A50}$$

For the proof of Lemma 10, we need to check whether each of the coefficients of $\mathcal{P}_q \mathrm{tr}_{L^c}(H_{\vec{a}}^{m_1}) \mathrm{tr}_{L^c}(H_{\vec{a}}^{m_2}) \cdots \mathrm{tr}_{L^c}(H_{\vec{a}}^{m_q})$ for all the pairs of $\{m_1, m_2, \ldots, m_q\}$ is equal between Eqs. (A49) and (A50). Instead of directly writing down the explicit form of $\mathcal{C}_{m_1, m_2, \ldots, m_q}^{(q)}$, we will take the following step. First, we prove

$$\frac{\partial^m}{\partial \beta^m} \log \left[\mathrm{tr}_{L^c}(e^{-\beta H_{\vec{a}}}/d_{L^c}) \right] \Big|_{\beta=0} = \frac{(-1)^m}{d_{L^c}^m} \mathrm{tr}_{L_{1:m}^c} \left(\tilde{H}_{\vec{a}}^{(0)} \tilde{H}_{\vec{a}}^{(1)} \cdots \tilde{H}_{\vec{a}}^{(m-1)} \right) \tag{A51}$$

in the case of $L^c = V$. The proof of Eq. (A51) implies that the coefficients of $\mathcal{P}_q \mathrm{tr}_{L^c}(H_{\vec{a}}^{m_1}) \mathrm{tr}_{L^c}(H_{\vec{a}}^{m_2}) \cdots \mathrm{tr}_{L^c}(H_{\vec{a}}^{m_q})$ are equal between Eqs. (A49) and (A50) for $L^c = V$. Then, because the coefficients $\mathcal{C}_{m_1, m_2, \ldots, m_q}^{(q)}$ do not depend on the form of L^c, the proof in the case of $L^c = V$ also results in the proof in the other cases (i.e., $L^c \neq V$). Therefore, in the following, we aim to give the proof of Eq. (A51) for $L^c = V$.

For $L^c = V$, we have

$$\frac{\partial}{\partial \beta} \log \left[\frac{\mathrm{tr}_V(e^{-\beta H_{\vec{a}}})}{d_V} \right] = -\mathrm{tr}(H_{\vec{a}} \rho_{\vec{a}}), \tag{A52}$$
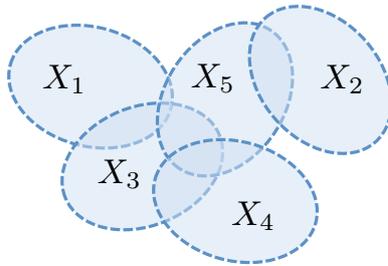
FIG. 6. $N_{X_s|w}$ is defined by a number of subsystems in w that have overlap with X_s. When $w = \{X_1, X_2, X_3, X_4, X_5\}$ is given as above, we have $N_{X_1|w} = 2$, $N_{X_2|w} = 1$, $N_{X_3|w} = 2$, $N_{X_4|w} = 2$ and $N_{X_5|w} = 4$.

and hence our task is to calculate

$$\frac{\partial^m}{\partial \beta^m} \log \left[\frac{\mathrm{tr}_V (e^{-\beta H_{\vec{a}}})}{d_V} \right] = -\mathrm{tr}_V \left(H_{\vec{a}} \frac{\partial^{m-1}}{\partial \beta^{m-1}} \rho_{\vec{a}} \right). \tag{A53}$$

By using Lemma 2 in Ref. [10], we have

$$\frac{\partial^{m-1}}{\partial \beta^{m-1}} \mathrm{tr} \left(H_{\vec{a}} \rho_{\vec{a}} \right) \Big|_{\beta=0} = \frac{(-1)^{m-1}}{d_V^m} \mathrm{tr}_{V_{1:m}^c} \left(\tilde{H}_{\vec{a}}^{(0)} \tilde{H}_{\vec{a}}^{(1)} \cdots \tilde{H}_{\vec{a}}^{(m-1)} \right), \tag{A54}$$

where in the inequality (B.3) in [10], we choose as $m_1 = 0$, $m_2 = m - 1$ and $\omega_X = H_{\vec{a}}$. We thus obtain the equation (A51). This completes the proof of Lemma 10. $\square$

―――――――――――――――――――――――――――[**End of Proof of Proposition 9**]

We then aim to obtain an upper bound of $\left\| \mathrm{tr}_{L_{1:m}^c} \left(\tilde{h}_{X_1}^{(0)} \tilde{h}_{X_2}^{(1)} \cdots \tilde{h}_{X_m}^{(m-1)} \right) \right\|$. For the purpose, we utilize the following proposition.

Proposition 11. *Let $\{O_s\}_{s=0}^m$ be operators supported on a subset $w := \{X_s\}_{s=0}^m$, respectively. When they satisfy $\mathrm{tr}_{L^c}(O_s) = 0$ for $s = 0, 1, 2 \ldots, m$, we obtain*

$$\frac{1}{d_{L^c}^m} \left\| \mathrm{tr}_{L_{1:m}^c} \left(\tilde{O}_0^{(0)} \tilde{O}_1^{(1)} \tilde{O}_2^{(2)} \cdots \tilde{O}_{m-1}^{(m-1)} \right) \right\| \le \|O_0\| \prod_{s=1}^m 2 N_{X_s|w_L} \|O_s\|, \tag{A55}$$

where we define $\tilde{O}_s^{(s)}$ as in Eq. (A38). $N_{X_s|w}$ is a number of subsets in w that have overlap with X_s (Fig. 6):

$$N_{X_s|w} = \#\{X \in w | X \ne X_s, X \cap X_s \ne \emptyset\}. \tag{A56}$$

The proof is the same as that of Proposition 3 in Ref. [10], which proves Ineq. (A55) for $L^c = V$.

In order to apply Proposition (11) to $\mathrm{tr}_{L_{1:m}^c} \left(\tilde{h}_{X_1}^{(0)} \tilde{h}_{X_2}^{(1)} \cdots \tilde{h}_{X_m}^{(m-1)} \right)$, the condition $\mathrm{tr}_{L^c}(h_X) = 0$ is necessary, whereas it is not generally satisfied. Thus, instead of considering h_X, we consider $\mathfrak{h}_X$ which is defined as follows:

$$\mathfrak{h}_X := h_X - \frac{h_X^L}{d_{L^c}} \quad \text{for} \quad X \in E_r, \tag{A57}$$

where $\mathfrak{h}_X$ satisfies $\mathrm{tr}_{L^c}(\mathfrak{h}_X) = \mathrm{tr}_{L^c}(h_X) - h_X^L \mathrm{tr}_{L^c}(\hat{1})/d_{L^c} = h_X^L - h_X^L = 0$ from the definition (A1). By using the notation of $\mathfrak{h}_X$, we obtain

$$\mathrm{tr}_{L_{1:m}^c} \left(\tilde{h}_{X_1}^{(0)} \tilde{h}_{X_2}^{(1)} \cdots \tilde{h}_{X_m}^{(m-1)} \right) = \mathrm{tr}_{L_{1:m}^c} \left(\tilde{\mathfrak{h}}_{X_1}^{(0)} \tilde{\mathfrak{h}}_{X_2}^{(1)} \cdots \tilde{\mathfrak{h}}_{X_m}^{(m-1)} \right) + \frac{h_{X_1}^L}{d_{L^c}} \otimes \mathrm{tr}_{L_{1:m}^c} \left(\tilde{\mathfrak{h}}_{X_2}^{(1)} \tilde{\mathfrak{h}}_{X_3}^{(2)} \cdots \tilde{\mathfrak{h}}_{X_m}^{(m-1)} \right), \tag{A58}$$

where we use $\tilde{\mathfrak{h}}_X^{(s)} = \tilde{h}_X^{(s)}$ for $s \ge 1$ which comes from the definition (A38), and apply Eq. (A57) to $\tilde{h}_{X_1}^{(0)}$. We then prove $\mathrm{tr}_{L_{1:m}^c} \left(\tilde{\mathfrak{h}}_{X_2}^{(1)} \tilde{\mathfrak{h}}_{X_3}^{(2)} \cdots \tilde{\mathfrak{h}}_{X_m}^{(m-1)} \right) = 0$. By using the definition (A38) for $\tilde{\mathfrak{h}}_{X_2}^{(1)}$, we have

$$\mathrm{tr}_{L_{1:m}^c} \left(\tilde{\mathfrak{h}}_{X_2}^{(1)} \tilde{\mathfrak{h}}_{X_3}^{(2)} \cdots \tilde{\mathfrak{h}}_{X_m}^{(m-1)} \right) = \mathrm{tr}_{L_{1:m}^c} \left[\left(\tilde{\mathfrak{h}}_{X_2, \tilde{\mathcal{H}}_1} - \tilde{\mathfrak{h}}_{X_2, \tilde{\mathcal{H}}_2} \right) \tilde{\mathfrak{h}}_{X_3}^{(2)} \cdots \tilde{\mathfrak{h}}_{X_m}^{(m-1)} \right]. \tag{A59}$$

Because the operator $\tilde{\mathfrak{h}}_X^{(s)}$ ($s \ge 2$) is invariant under the swapping between the Hilbert spaces $H_1^{L^c}$ and $H_2^{L^c}$ (i.e., $\tilde{\mathcal{H}}_1 \leftrightarrow \tilde{\mathcal{H}}_2$), we have

$$\mathrm{tr}_{L_{1:m}^c} \left(\tilde{\mathfrak{h}}_{X_2, \tilde{\mathcal{H}}_1} \tilde{\mathfrak{h}}_{X_3}^{(2)} \cdots \tilde{\mathfrak{h}}_{X_m}^{(m-1)} \right) = \mathrm{tr}_{L_{1:m}^c} \left(\tilde{\mathfrak{h}}_{X_2, \tilde{\mathcal{H}}_2} \tilde{\mathfrak{h}}_{X_3}^{(2)} \cdots \tilde{\mathfrak{h}}_{X_m}^{(m-1)} \right). \tag{A60}$$

Therefore, the term (A59) vanishes and Eq. (A58) reduces to

$$\mathrm{tr}_{L^{\mathrm{c}}_{1:m}} \left(\tilde{h}^{(0)}_{X_1} \tilde{h}^{(1)}_{X_2} \cdots \tilde{h}^{(m-1)}_{X_m} \right) = \mathrm{tr}_{L^{\mathrm{c}}_{1:m}} \left(\tilde{\mathfrak{h}}^{(0)}_{X_1} \tilde{\mathfrak{h}}^{(1)}_{X_2} \cdots \tilde{\mathfrak{h}}^{(m-1)}_{X_m} \right). \tag{A61}$$

By using Proposition 11, we obtain an upper bound of $\mathrm{tr}_{L^{\mathrm{c}}_{1:m}} \left(\tilde{h}^{(0)}_{X_1} \tilde{h}^{(1)}_{X_2} \cdots \tilde{h}^{(m-1)}_{X_m} \right)$ as follows:

$$\frac{1}{d^m_{L^{\mathrm{c}}}} \left\| \mathrm{tr}_{L^{\mathrm{c}}_{1:m}} \left(\tilde{h}^{(0)}_{X_1} \tilde{h}^{(1)}_{X_2} \cdots \tilde{h}^{(m-1)}_{X_m} \right) \right\| = \frac{1}{d^m_{L^{\mathrm{c}}}} \left\| \mathrm{tr}_{L^{\mathrm{c}}_{1:m}} \left(\tilde{\mathfrak{h}}^{(0)}_{X_1} \tilde{\mathfrak{h}}^{(1)}_{X_2} \cdots \tilde{\mathfrak{h}}^{(m-1)}_{X_m} \right) \right\|$$

$$\leq \| \mathfrak{h}_{X_1} \| \prod_{s=2}^{m} 2 N_{X_s|w} \| \mathfrak{h}_{X_s} \| \leq \frac{1}{2} \prod_{s=1}^{m} 4 N_{X_s|w} \| h_{X_s} \|, \tag{A62}$$

where we use $\| \mathfrak{h}_X \| \leq 2 \| h_X \|$ which comes from the definition (A57). By combining the inequality (A62) with Eq. (A37), we obtain an upper bound of

$$\left\| \mathcal{D}_w \log \tilde{\rho}^{L}_{\vec{a}} \big|_{\vec{a}=\vec{0}} \right\| \leq \frac{1}{2} \prod_{s=1}^{m} 4 \beta N_{X_s|w} \| h_{X_s} \|. \tag{A63}$$

By applying the inequality (A63) to the cases $L = AB$, $L = BC$, $L = ABC$ and $L = B$, we obtain the following inequality:

$$\left\| \mathcal{D}_w \tilde{H}_{\vec{a}}(A : C|B) \big|_{\vec{a}=\vec{0}} \right\| \leq 2(4\beta)^m \prod_{s=1}^{m} N_{X_s|w} \| h_{X_s} \|, \tag{A64}$$

where $\tilde{H}_{\vec{a}}(A : C|B)$ has been defined in Eq. (A20). Then, the final task is to upper-bound the summation with respect to $\sum_{w \in \mathcal{G}^{A,C}_{r,m}}$ in Eq. (A28):

$$\left\| \tilde{H}_{\vec{1}}(A : C|B) \right\| \leq \sum_{m=1}^{\infty} \frac{1}{m!} \sum_{w \in \mathcal{G}^{A,C}_{r,m}} n_w \left\| \mathcal{D}_w \tilde{H}_{\vec{a}}(A : C|B) \big|_{\vec{a}=\vec{0}} \right\|$$

$$\leq \sum_{m=1}^{\infty} \frac{2(4\beta)^m}{m!} \sum_{w \in \mathcal{G}^{A,C}_{r,m}} n_w \prod_{s=1}^{m} N_{X_s|w} \| h_{X_s} \|, \tag{A65}$$

where we use the proposition 8 in the first inequality.

For the estimation of the summation, we first focus on the fact that any cluster in $w \in \mathcal{G}^{A,C}_{r,m}$ must have overlaps with the surface regions of A and C, say ∂A_r and ∂C_r ($r \in \mathbb{N}$):

$$\partial A_r := \{ v \in A | d_{v,A^{\mathrm{c}}} \leq r \}, \quad \partial C_r := \{ v \in C | d_{v,C^{\mathrm{c}}} \leq r \}. \tag{A66}$$

Second, because $d_{A,C}$ is the minimum path length on the graph (V, E) to connect the subsystems A and C, the condition $w \in \mathcal{G}^{A,C}_{r,m}$ implies $|w| \geq d_{A,C}/r$ as the necessary condition. From these two fact, we will replace the summation $\sum_{w \in \mathcal{G}^{A,C}_{r,m}}$ with $\sum_{v \in \partial A_r} \sum_{m \geq d_{A,C}/r} \sum_{w \in \mathcal{G}^v_{r,m}}$ by taking all the clusters with the sizes $|w| \geq d_{A,C}/r$ which have overlap with A into account:

$$\sum_{m=1}^{\infty} \frac{2(4\beta)^m}{m!} \sum_{w \in \mathcal{G}^{A,C}_{r,m}} n_w \prod_{s=1}^{m} N_{X_s|w} \| h_{X_s} \| \leq \sum_{v \in \partial A_r} \sum_{m \geq d_{A,C}/r} 2(4\beta)^m \sum_{w \in \mathcal{G}^v_{r,m}} \frac{n_w}{m!} \prod_{s=1}^{m} N_{X_s|w} \| h_{X_s} \|, \tag{A67}$$

where the same inequality holds for the replacement of $\sum_{v \in \partial A_r}$ by $\sum_{v \in \partial C_r}$.

In order to estimate the summation of $\sum_{w \in \mathcal{G}^v_{r,m}}$, we utilize the following proposition which has been given in Ref. [10]:

Proposition 12 (Proposition 4 in Ref. [10]). *Let* $\{ o_X \}_{X \in E_\infty}$ *be arbitrary operators such that*

$$\sum_{X | X \ni v} \| o_X \| \leq g \quad for \quad \forall v \in V, \tag{A68}$$

where E_∞ *is defined by Eq.* (A5) *and it gives the set of all the subsystems* $X \subset V$ *with* $|X| \leq k$. *Then, for an arbitrary subset* L, *we obtain*

$$\sum_{w \in \mathcal{G}^L_m} \frac{n_w}{m!} \prod_{s=1}^{m} N_{X_s|w_L} \| o_{X_s} \| \leq \frac{1}{2} e^{|L|/k} (2 e^3 g k)^m, \tag{A69}$$

where w_L *is defined as* $w_L := \{ L, X_1, X_2, \ldots, X_{|w|} \}$ *for* $w = \{ X_1, X_2, \ldots, X_{|w|} \}$.

By applying Proposition 12 to the inequality (A67), we have

$$\sum_{w \in \mathcal{G}_{r,m}^v} \frac{n_w}{m!} \prod_{s=1}^m N_{X_s|w} \|h_{X_s}\| \le \frac{1}{2} e^{1/k} (2e^3 k)^m, \tag{A70}$$

where we use $N_{X_s|w_L} \le N_{X_s|w}$ in (A69) and the condition (A4) gives $g = 1$. Therefore, the inequality (A67) reduces to

$$\sum_{m=0}^\infty \frac{1}{m!} \sum_{w \in \mathcal{G}_{r,m}^{A,C}} n_w \left\| \mathcal{D}_w \tilde{H}_{\vec{a}}(A:C|B) \right\|_{\vec{a}=\vec{0}} \le \sum_{v \in \partial A_r} \sum_{m \ge d_{A,C}/r} e^{1/k} (8e^3 k\beta)^m$$

$$\le e|\partial A_r| \frac{(8e^3 k\beta)^{d_{A,C}/r}}{1 - 8e^3 k\beta}, \tag{A71}$$

where we use $k \ge 1$. We notice that the same inequality holds for the replacement of $|\partial A_r|$ by $|\partial C_r|$. By combining the inequalities (A21), (A65) and (A71), we prove Theorem 1. $\square$

Appendix B: Quasi-Locality of effective Hamiltonian on a subsystem: Proof of Theorem 3

We here consider the effective Hamiltonian on a subsystem L, which we define as

$$\tilde{H}_L := -\beta^{-1} \log \tilde{\rho}^L, \tag{B1}$$

where $\tilde{\rho}^L$ is defined in Eq. (A19). We prove the following theorem which refines the Theorem 3:

Theorem 13. *The effective Hamiltonian $\tilde{H}_L$ is given by a quasi-local operator*

$$\tilde{H}_L = H_L + \sum_{m=1}^\infty \sum_{w \in \mathcal{G}_{r,m}^{L,L^c}} n_w h_{L_w} - \frac{\hat{1}}{\beta} \log Z_{L^c} \tag{B2}$$

with

$$H_L := \sum_{X \subset L} h_X, \quad Z_{L^c} := \frac{1}{d_L} \mathrm{tr}(e^{-\beta H_{L^c}} \otimes \hat{1}_L) \tag{B3}$$

for $L \subset V$, where each of $\{h_{L_w}\}_{w \in \mathcal{G}_{r,m}^{L,L^c}}$ is supported on the subsystem $L_w := L \cap V_w$ (see Def. (B18)) and $\mathcal{G}_{r,m}^{L,L^c}$ is defined as a cluster subset defined in Def. 3. The effective interaction terms $\{h_{L_w}\}_{w \in \mathcal{G}_{r,m}^{L,L^c}}$ is exponentially localized around the boundary:

$$\sum_{m > m_0} \sum_{w \in \mathcal{G}_{r,m}^{L,L^c}} n_w \|h_{L_w}\| \le \frac{e}{4\beta} \frac{(\beta/\beta_c)^{m_0+1}}{1 - \beta/\beta_c} |\partial L_r| \tag{B4}$$

for an arbitrary m_0.

From Eq. (B2), the effective interaction term Φ_L is given by

$$\Phi_L = \sum_{m=1}^\infty \sum_{w \in \mathcal{G}_{r,m}^{L,L^c}} n_w h_{L_w} - \frac{\hat{1}}{\beta} \log Z_{L^c}. \tag{B5}$$

Because of $\mathrm{diam}(V_w) \le mr$, the subsystem $L \cap V_w$ ($w \in \mathcal{G}_{r,m}^{L,L^c}$) is separated from the boundary ∂L at most by a distance mr, namely $L \cap V_w \subseteq \partial L_{mr}$, where ∂L_l has been defined in Eq. (9) as follows:

$$\partial L_l := \{v \in L | d_{v,L^c} \le l\}. \tag{B6}$$

Hence, by defining $\Phi_{\partial L_l}$ as

$$\Phi_{\partial L_l} = \sum_{m \le \lfloor l/r \rfloor} \sum_{w \in \mathcal{G}_{r,m}^{L,L^c}} n_w h_{L_w} - \frac{\hat{1}}{\beta} \log Z_{L^c}, \tag{B7}$$

we have

$$\|\Phi_L - \Phi_{\partial L_l}\| \le \frac{e}{4\beta} \frac{(\beta/\beta_c)^{l/r}}{1 - \beta/\beta_c} |\partial L_r|. \tag{B8}$$

This gives the proof of Theorem 3.

1. Proof of Theorem 13

In order to apply the generalized cluster expansion, we first parametrize $\tilde{H}_L$ as

$$\tilde{H}_{L,\vec{a}} := -\beta^{-1} \log \tilde{\rho}_{\vec{a}}^L. \tag{B9}$$

As in Eq. (A24), the generalized cluster expansion for $\tilde{H}_{L,\vec{a}}$ reads

$$\tilde{H}_{L,\vec{1}} = -\frac{1}{\beta} \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{w \in \mathcal{C}_{r,m}} n_w \mathcal{D}_w \tilde{H}_{L,\vec{a}} \Big|_{\vec{a}=\vec{0}}. \tag{B10}$$

We can now prove the following proposition:

Proposition 14. *The summation with respect to the clusters $\sum_{w \in \mathcal{C}_{r,m}}$ reduces to the following form:*

$$\tilde{H}_{L,\vec{1}} = H_L - \frac{1}{\beta} \log Z_{L^c} + \sum_{m=1}^{\infty} \frac{1}{m!} \sum_{w \in \mathcal{G}_{r,m}^{L,L^c}} n_w \mathcal{D}_w \tilde{H}_{L,\vec{a}} \Big|_{\vec{a}=\vec{0}}, \tag{B11}$$

where $H_L := \sum_{X \subset L} h_X$ and $Z_{L_c} := d_L^{-1} \mathrm{tr}(e^{-\beta H_{L^c}})$.

a. Proof of Proposition 14

For the proof, we first prove

$$\mathcal{D}_w \log(\tilde{\rho}_{\vec{a}_w}^L) = 0 \quad \text{for} \quad w \notin \mathcal{G}_{r,|w|}. \tag{B12}$$

The proof is given as follows. Due to the existence of decomposition $w = w_1 \sqcup w_2$ such that $V_{w_1} \cap V_{w_2} = \emptyset$, we have $e^{-\beta H_{\vec{a}_w}} = e^{-\beta H_{\vec{a}_{w_1}}} \otimes e^{-\beta H_{\vec{a}_{w_2}}}$ and hence,

$$\log(\tilde{\rho}_{\vec{a}_w}^L) = \log(\tilde{\rho}_{\vec{a}_{w_1}}^L) + \log(\tilde{\rho}_{\vec{a}_{w_2}}^L) - \log d_{L^c}. \tag{B13}$$

Because $\mathcal{D}_{w_2} \log(\tilde{\rho}_{\vec{a}_{w_1}}^L) = \mathcal{D}_{w_1} \log(\tilde{\rho}_{\vec{a}_{w_2}}^L) = 0$, we obtain Eq. (B12).

We then consider the cases of $V_w \subseteq L$ and $V_w \subseteq L^c$ in Eq. (B10). In the case of $V_w \subseteq L$, the definition (A19) gives

$$\log(\tilde{\rho}_{L,\vec{a}_w}) = -\beta H_{\vec{a}_w} + \log d_{L^c}. \tag{B14}$$

Therefore, we have $\mathcal{D}_w \log(\tilde{\rho}_{\vec{a}_w}^L)$ vanishes for $m \geq 2$, and

$$-\frac{1}{\beta} \sum_{m=1}^{\infty} \frac{1}{m!} \sum_{w \in \mathcal{G}_{r,m}, V_w \subseteq L} n_w \mathcal{D}_w \tilde{H}_{L,\vec{a}} \Big|_{\vec{a}=\vec{0}} = \sum_{X \subset L} h_X = H_L. \tag{B15}$$

On the other hand, in the case of $V_w \subseteq L^c$, $\log(\tilde{\rho}_{\vec{a}_w}^L)$ becomes a constant operator (i.e., $\log(\tilde{\rho}_{\vec{a}_w}^L) \propto \hat{1}$). Hence, we obtain

$$-\frac{1}{\beta} \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{w \in \mathcal{G}_{r,m}, V_w \subseteq L^c} n_w \mathcal{D}_w \tilde{H}_{L,\vec{a}} \Big|_{\vec{a}=\vec{0}} = -\frac{1}{\beta} \log[\mathrm{tr}_{L^c}(e^{-\beta H_{L^c}})] = -\frac{\log Z_{L_c}}{\beta} \hat{1}. \tag{B16}$$

Thus, the summation (B10) reduces to

$$\tilde{H}_{L,\vec{1}} = -\frac{1}{\beta} \sum_{m=1}^{\infty} \frac{1}{m!} \sum_{w \in \mathcal{G}_{r,m}, V_w \subseteq L} n_w \mathcal{D}_w \tilde{H}_{L,\vec{a}} \Big|_{\vec{a}=\vec{0}} - \frac{1}{\beta} \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{w \in \mathcal{G}_{r,m}, V_w \subseteq L^c} n_w \mathcal{D}_w \tilde{H}_{L,\vec{a}} \Big|_{\vec{a}=\vec{0}}$$

$$- \frac{1}{\beta} \sum_{m=1}^{\infty} \frac{1}{m!} \sum_{w \in \mathcal{G}_{r,m}^{L,L^c}} n_w \mathcal{D}_w \tilde{H}_{L,\vec{a}} \Big|_{\vec{a}=\vec{0}}$$

$$= H_L - \frac{1}{\beta} \log Z_{L^c} - \frac{1}{\beta} \sum_{m=1}^{\infty} \frac{1}{m!} \sum_{w \in \mathcal{G}_{r,m}^{L,L^c}} n_w \mathcal{D}_w \tilde{H}_{L,\vec{a}} \Big|_{\vec{a}=\vec{0}}. \tag{B17}$$

This completes the proof. $\square$

We now define h_{L_w} as

$$h_{L_w} := \frac{-\beta^{-1}}{m!} \mathcal{D}_w \tilde{H}_{L,\vec{a}}\Big|_{\vec{a}=\vec{0}}, \tag{B18}$$

where $w \in \mathcal{G}_{r,m}^{L,L^c}$. Note that the operator h_{L_w} is supported on the subsystem $L_w = L \cap V_w$. Then, the effective Hamiltonian $\tilde{H}_{L,\vec{1}}$ is formally written by

$$\tilde{H}_{L,\vec{1}} = H_L - \frac{1}{\beta} \log Z_{L^c} + \sum_{m=1}^{\infty} \sum_{w \in \mathcal{G}_{r,m}^{L,L^c}} n_w h_{L_w}. \tag{B19}$$

By using the proposition 9 with the inequalities (A62) and (A70), we have

$$\sum_{w \in \mathcal{G}_{r,m}^v} n_w \|h_{L_w}\| \le \frac{\beta^{-1}}{m!} \sum_{w \in \mathcal{G}_{r,m}^v} \frac{n_w}{2} \prod_{s=1}^m 4\beta N_{X_s|w} \|h_{X_s}\|$$

$$\le (4\beta)^{m-1} e^{1/k} (2e^3 k)^m \le \frac{e}{4\beta} (\beta/\beta_c)^m, \tag{B20}$$

where we use $e^{1/k} \le e$ due to $k \ge 1$. By using the above inequality, the contribution of mth order terms in the expansion (B11) is bounded from above by

$$\sum_{w \in \mathcal{G}_{r,m}^{L,L^c}} n_w \|h_{L_w}\| \le \sum_{v \in \partial L_r} \sum_{w \in \mathcal{G}_{r,m}^v} n_w \|h_{L_w}\| \le \frac{e}{4\beta} (\beta/\beta_c)^m |\partial L_r|, \tag{B21}$$

where ∂L_r has been defined in Eq. (B6).

$$\sum_{m>m_0}^{\infty} \sum_{w \in \mathcal{G}_{r,m}^{L,L^c}} n_w \|h_{L_w}\| \le \frac{e|\partial L_r|}{4\beta} \sum_{m=m_0+1}^{\infty} (\beta/\beta_c)^m = \frac{e|\partial L_r|}{4\beta} \frac{(\beta/\beta_c)^{m_0+1}}{1 - \beta/\beta_c}. \tag{B22}$$

This completes the proof of Theorem 13. □

2. Computational cost of cluster summation

We here show the computational cost to estimate the effective Hamiltonian $\tilde{H}_L$. For this aim, we start from a slightly weaker expression than Eq. (B11) as follows

$$\Phi_L = \tilde{H}_{L,\vec{1}} - H_L = -\frac{1}{\beta} \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{w \in \mathcal{G}_{r,m}, V_w \subseteq L^c} n_w \mathcal{D}_w \tilde{H}_{L,\vec{a}}\Big|_{\vec{a}=\vec{0}} - \frac{1}{\beta} \sum_{m=1}^{\infty} \frac{1}{m!} \sum_{w \in \mathcal{G}_{r,m}^{L,L^c}} n_w \mathcal{D}_w \tilde{H}_{L,\vec{a}}\Big|_{\vec{a}=\vec{0}}, \tag{B23}$$

where we use the second and third terms in the first equation of (B17). Our task is to estimate the computational cost of $n_w \mathcal{D}_w \tilde{H}_{L,\vec{a}}\big|_{\vec{a}=\vec{0}}$ and the number of clusters in $\{w \in \mathcal{G}_{r,m}, V_w \subseteq L^c\}$ and $w \in \mathcal{G}_{r,m}^{L,L^c}$.

First, we consider $n_w \mathcal{D}_w \tilde{H}_{L,\vec{a}}\big|_{\vec{a}=\vec{0}}$. As defined in Eq. (A24), n_w is immediately calculated, and hence we need to estimate the computational cost to calculate the multiderivative

$$\mathcal{D}_w \tilde{H}_{L,\vec{a}_w}\Big|_{\vec{a}_w=\vec{0}} = \prod_{j=1}^m \frac{\partial}{\partial a_{X_j}} \tilde{H}_{L,\vec{a}_w}\Big|_{\vec{a}_w=\vec{0}} \tag{B24}$$

with $w = \{X_s\}_{s=1}^m$ by using numerical differentiation. The operator $\tilde{H}_{L,\vec{a}_w}$ is given by

$$\tilde{H}_{L,\vec{a}_w} = -\beta^{-1} \log \tilde{\rho}_{\vec{a}_w}^L = -\beta^{-1} \text{tr}_{L^c} \left(e^{-\beta H_{\vec{a}_w}} \right) \otimes \hat{1}_{L^c}, \tag{B25}$$

where we use the definition (A19). Note that $H_{\vec{a}_w}$ is supported on $V_w \subset V$. Hence, the computational cost to calculate $\tilde{H}_{L,\vec{a}_w}$ is at most of $d^{\mathcal{O}(|V_w|)}$. In order to perform the differentiation, we need to calculate $2^{|w|}$ values of $\tilde{H}_{L,\vec{a}_w}$ for $a_{X_s} = \pm\Delta$ ($\Delta \to +0$) for $s = 1, 2, \ldots, |w|$. Thus, for the numerical differentiation we need the computational cost of $2^{|w|} \cdot d^{\mathcal{O}(|V_w|)} = d^{\mathcal{O}(mk)}$ with $|w| = m$, where we use $|V_w| \le |w|k$.

We then need to sum up the contributions from all the clusters in $\{w \in \mathcal{G}_{r,m}, V_w \subseteq L^c\}$ and $w \in \mathcal{G}_{r,m}^{L,L^c}$. For the purpose, we first prove the following theorem on the number of clusters:

Proposition 15. *The total number of different clusters in $\mathcal{G}_{r,m}^{L^c}$ is bounded as follows:*

$$\#\left\{w\in\mathcal{C}_{r,m}\big|w\in\mathcal{G}_{r,m},V_w\subseteq L^c\quad\text{or}\quad w\in\mathcal{G}_{r,m}^{L,L^c}\right\}\le|L^c|\left(3\cdot2^kd_G^{rk}\right)^m. \tag{B26}$$

This roughly gives the total number by $|L^c|d_G^{\mathcal{O}(rkm)}$,

In total, the computation of the m-th order in the expansion (B23) is performed with the runtime bounded from above by

$$d^{\mathcal{O}(mk)}\cdot|L^c|d_G^{\mathcal{O}(rkm)}\le n(d\cdot d_G^r)^{mk}. \tag{B27}$$

Also, the convergence of the expansion (B23) is estimated as in (B21) and (B22)

$$\sum_{w\in\mathcal{G}_{r,m},V_w\subseteq L^c}\left\|n_w\mathcal{D}_w\tilde{H}_{L,\vec{a}}\Big|_{\vec{a}=\vec{0}}\right\|-\sum_{m=1}^\infty\frac{1}{m!}\sum_{w\in\mathcal{G}_{r,m}^{L,L^c}}\left\|n_w\mathcal{D}_w\tilde{H}_{L,\vec{a}}\Big|_{\vec{a}=\vec{0}}\right\|$$
$$\le\sum_{v\in L^c}\sum_{w\in\mathcal{G}_{r,m}^v}n_w\|h_{L_w}\|\le\frac{e}{4\beta}(\beta/\beta_c)^m|L^c|\le\frac{e}{4\beta}(\beta/\beta_c)^mn, \tag{B28}$$

which yields

$$\sum_{m>m_0}^\infty\sum_{w\in\mathcal{G}_{r,m},V_w\subseteq L^c}\left\|n_w\mathcal{D}_w\tilde{H}_{L,\vec{a}}\Big|_{\vec{a}=\vec{0}}\right\|-\sum_{m>m_0}^\infty\sum_{m=1}^\infty\frac{1}{m!}\sum_{w\in\mathcal{G}_{r,m}^{L,L^c}}\left\|n_w\mathcal{D}_w\tilde{H}_{L,\vec{a}}\Big|_{\vec{a}=\vec{0}}\right\|\le\frac{en}{4\beta}\frac{(\beta/\beta_c)^{m_0+1}}{1-\beta/\beta_c}. \tag{B29}$$

Therefore, we need to choose $m=\mathcal{O}(\log(1/\epsilon))$ to calculate Φ_L up to an error $n\epsilon$ as long as $\beta<\beta_c$. Hence, the computational cost is estimated as

$$n(d\cdot d_G^r)^{k\mathcal{O}(\log(1/\epsilon))}=n(1/\epsilon)^{\mathcal{O}(k\log(dd_G^r))}. \tag{B30}$$

This completes the derivation of the computational cost (16) for computing Φ_L. $\square$

a. Proof of Proposition 15

We here prove Proposition 15 which gives an upper bound of the number of cluster connecting to a subset L^c. For the purpose, we estimate the number of clusters in $\mathcal{G}_{r,m}^v$, which gives an upper bound of

$$\#\left\{w\in\mathcal{C}_{r,m}\big|w\in\mathcal{G}_{r,m},V_w\subseteq L^c\quad\text{or}\quad w\in\mathcal{G}_{r,m}^{L,L^c}\right\}\le\sum_{v\in L^c}\#\left\{w\big|w\in\mathcal{G}_{r,m}^v\right\}. \tag{B31}$$

First, we count the number of clusters $w=\{X_s\}_{s=1}^q$ which satisfy $X_s\cap Y\ne\emptyset$ for $\forall X_s$ $(s=1,2,\ldots,q)$, where Y is an arbitrary subset in V. The number is bounded from above by

$$\#\{w\in\mathcal{C}_{r,q}|X_s\cap Y\ne\emptyset,\ s=1,2,\ldots,q\}\le\sum_{\{v_1,v_2,\ldots,v_q\}\subseteq Y}\prod_{s=1}^q\deg(v_s), \tag{B32}$$

where we define $\deg(v)$ as $\deg(v):=\#\{X\in E_r|X\ni v\}$. By using the graph degree d_G, we can upper-bound $\deg(v)$ by

$$\deg(v)=\#\{X\in E_r|X\ni v\}\le\binom{d_G^r}{k}\le d_G^{rk}, \tag{B33}$$

where d_G^r is the upper bound of the number of vertices $\{v'\}_{v'\in V}$ such that $d_{v,v'}\le r$. Also, note that $X\in E_r$ implies $|X|\le k$ from the definitions (A4) and (A5). The summation with respect to $\{v_1,v_2,\ldots,v_q\}$ is equal to the q_1-multicombination from a set of $|L|$ vertices, which is equal to

$$\sum_{\{v_1,v_2,\ldots,v_q\}\subseteq Y}=\left(\!\!\binom{|Y|}{q}\!\!\right)=\binom{q+|Y|-1}{q}\le2^{q+|Y|-1}. \tag{B34}$$

By combining the inequalities (B33) and (B34) with (B32), we obtain

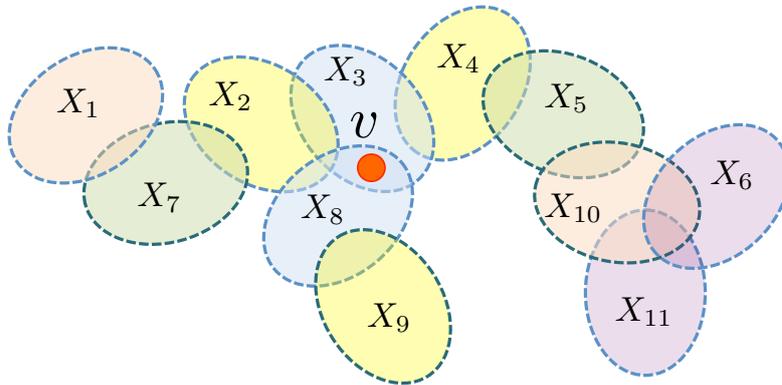$$\#\{w\in\mathcal{C}_{r,q}|X_s\cap Y\ne\emptyset,\ s=1,2,\ldots,q\}\le2^{|Y|-1}(2d_G^{rk})^q. \tag{B35}$$

FIG. 7. Decomposition of w in $\mathcal{G}_{r,m}^v$ as in Eq. (B36). In the picture, we have $w_0 = \{X_3, X_8\}$, $w_1 = \{X_2, X_4, X_9\}$, $w_2 = \{X_5, X_7\}$, $w_3 = \{X_1, X_{10}\}$, $w_4 = \{X_6, X_{11}\}$.

We then consider the following decomposition of $w \in \mathcal{G}_{r,m}^v$ (see Fig. 7):

$$w = w_0 \sqcup w_1 \sqcup w_2 \sqcup \cdots \sqcup w_l, \quad 0 \le l \le m-1, \tag{B36}$$

where $w_j \subset w_L$ satisfy $d(w_j, v) = j$ for $j = 0, 1, 2, \ldots, l$. Here, we define $d(w_j, w_0)$ as the shortest path length in the cluster $w_0 \sqcup w_1 \sqcup \cdots \sqcup w_{j-1}$ which connects from w_j to v. We also define $q_j := |w_j|$ with $q_j \ge 1$. We notice that all the clusters $w \in \mathcal{G}_{r,m}^v$ can be decomposed into the from of (B36).

For fixed $\{q_0, q_1, \ldots, q_l\}$, the number of clusters $\{w_1, w_2, \ldots, w_l\}$ defined as in Eq. (B36) is bounded by

$$\#\left\{w \in \mathcal{C}_{r,q_0} | X_{0,s} \cap v \ne \emptyset, \ s = 1, 2, \ldots, q_0\right\} \prod_{j=1}^l \max_{w_{j-1} \in \mathcal{C}_{r,q_{j-1}}} \left(\#\left\{w \in \mathcal{C}_{r,q_j} | X_{s_j} \cap V_{w_{j-1}} \ne \emptyset, \ s_j = 1, 2, \ldots, q_j\right\}\right)$$

$$\le (2d_G^{rk})^{q_0} \prod_{j=1}^l \left[2^{kq_{j-1}-1}(2d_G^{rk})^{q_j}\right] \le 2^{-l}\left(2^{k+1}d_G^{rk}\right)^m, \tag{B37}$$

where we denote $w_j = \{X_{s_j}\}_{s_j=1}^{q_j}$; note that $\sum_{j=0}^l q_j = m$. Then, by taking the summation with respect to $\{q_0, q_1, \ldots, q_l\}$ and l, we finally obtain the upper bound of $\#\{w | w \in \mathcal{G}_{r,m}^v\}$ as follows:

$$\#\{w | w \in \mathcal{G}_{r,m}^v\} \le \sum_{l=0}^{m-1} \sum_{\substack{q_0+q_1+\cdots+q_l=m \\ q_0 \ge 1, q_1 \ge 1, \ldots, q_l \ge 1}} 2^{-l}\left(2^{k+1}d_G^{rk}\right)^m$$

$$= \sum_{l=0}^{m-1} \left(\binom{l+1}{m-l-1}\right) 2^{-l}\left(2^{k+1}d_G^{rk}\right)^m$$

$$= \sum_{l=0}^{m-1} \binom{m-1}{l} 2^{-l}\left(2^{k+1}d_G^{rk}\right)^m \le \left(3 \cdot 2^k d_G^{rk}\right)^m, \tag{B38}$$

where the summation with respect to $\{q_0, q_1, \ldots, q_l\}$ ($q_0 \ge 1, q_1 \ge 1, \ldots, q_l \ge 1$) is equal to the $(m-l-1)$-multicombination from a set of $l+1$ elements:

$$\sum_{\substack{q_0+q_1+\cdots+q_l=m \\ q_0 \ge 1, q_1 \ge 1, \ldots, q_l \ge 1}} = \left(\binom{l+1}{m-l-1}\right) = \binom{m-1}{l}. \tag{B39}$$

By applying the above upper bound to the inequality (B31), we obtain the main inequality (B26). This completes the proof. $\square$

_____[**End of Proof of Proposition 15**]

Appendix C: Proof of Theorem 5

We here show the proof of Theorem 5 which upper bounds the conditional mutual information in long-range interacting systems. We rewrite the Hamiltonian with the power-law decay interaction by using the notations (A2)

and (A5):

$$H = \sum_{X \in E_\infty} h_X = \sum_{l=1}^{\infty} \sum_{X \in E^{(l)}} h_X. \tag{C1}$$

We here define $\tilde{g}_l$ as

$$\tilde{g}_l := \max_{v \in V} \sum_{X \in E^{(l)} | X \ni v} \|h_X\|. \tag{C2}$$

Then, the assumption (17) implies

$$\sum_{l \geq R} \sum_{X \in E^{(l)} | X \ni v} \|h_X\| \leq \sum_{l \geq R} \tilde{g}_l \leq R^{-\alpha}. \tag{C3}$$

We again show the statement that we would like to prove:

Theorem 16. *Let A, B and C be arbitrary subsystems in V $(A, B, C \subset V)$. Then, under the assumption that the inverse temperature satisfies*

$$\beta < \beta_c/11 = \frac{1}{88e^3 k}, \tag{C4}$$

the Gibbs state ρ satisfies the approximate Markov property as follows:

$$\mathcal{I}_\rho(A : C|B) \leq \beta \min(|A|, |C|) \frac{11 e^{1/k}/\beta_c}{1 - 11\beta/\beta_c} d_{A,C}^{-\alpha}, \tag{C5}$$

where we assume that $d_{A,C} \geq 2\alpha$.

1. Details of the proof

We start from Eq. (A24). By parametrizing the Hamiltonian as

$$H_{\vec{a}} = \sum_{X \in E_\infty} a_X h_X = \sum_{l=1}^{\infty} \sum_{X \in E^{(l)}} a_X h_X, \tag{C6}$$

we have

$$\begin{aligned}
\tilde{H}_{\vec{1}}(A : C|B) &= \sum_{m=1}^{\infty} \frac{1}{m!} \sum_{X_1, X_2, \ldots, X_m \in E_\infty} \prod_{j=1}^{m} \frac{\partial}{\partial a_{X_j}} \log(\tilde{\rho}_{\vec{a}}^L)\Big|_{\vec{a}=\vec{0}} \\
&= \sum_{m=1}^{\infty} \frac{1}{m!} \sum_{l_1, l_2, \ldots, l_m = 1}^{\infty} \sum_{X_1 \in E^{(l_1)}, X_2 \in E^{(l_2)}, \ldots, X_m \in E^{(l_m)}} \prod_{j=1}^{m} \frac{\partial}{\partial a_{X_j}} \log(\tilde{\rho}_{\vec{a}}^L)\Big|_{\vec{a}=\vec{0}} \\
&= \sum_{m=1}^{\infty} \frac{1}{m!} \sum_{l_0 = m}^{\infty} \sum_{w \in \mathcal{C}_m(l_0)} n_w \mathcal{D}_w \tilde{H}_{\vec{a}}(A : C|B)\Big|_{\vec{a}=\vec{0}},
\end{aligned} \tag{C7}$$

where we define $\mathcal{C}_m(l_0) \subset \mathcal{C}_{\infty,m}$ as

$$\mathcal{C}_m(l_0) = \left\{ w = \{X_1, X_2, \ldots, X_m\} \in \mathcal{C}_{\infty,m} \middle| X_j \in E^{(l_j)}, \ j = 1, 2, \ldots, m \quad \text{s.t.} \quad \sum_{j=1}^{m} l_j = l_0 \right\}. \tag{C8}$$

See Eq. (A2) and Sec. A 1 a for the definitions of $\mathcal{C}_{\infty,m}$ and $E^{(l)}$.

Next, from Eq. (C7), we can derive a similar statement to the proposition 8:

$$\begin{aligned}
\tilde{H}_{\vec{1}}(A : C|B) &= \sum_{m=1}^{\infty} \frac{1}{m!} \sum_{l_0 = m}^{\infty} \sum_{w \in \mathcal{C}_m(l_0)} n_w \mathcal{D}_w \tilde{H}_{\vec{a}}(A : C|B)\Big|_{\vec{a}=\vec{0}} \\
&= \sum_{m=1}^{\infty} \frac{1}{m!} \sum_{l_0 \geq d_{A,C}} \sum_{w \in \mathcal{G}_m^{A,C}(l_0)} n_w \mathcal{D}_w \tilde{H}_{\vec{a}}(A : C|B)\Big|_{\vec{a}=\vec{0}},
\end{aligned} \tag{C9}$$

where we define $\mathcal{G}_m^{A,C}(l_0) \subset \mathcal{G}_{\infty,m}^{A,C}$ as

$$\mathcal{G}_m^{A,C}(l_0) = \left\{ w = \{X_1, X_2, \ldots, X_m\} \in \mathcal{G}_{\infty,m}^{A,C} \,\middle|\, X_j \in E^{(l_j)}, \; j = 1, 2, \ldots, m \quad \text{s.t.} \quad \sum_{j=1}^m l_j = l_0 \right\}. \tag{C10}$$

Notice that we have $w \notin \mathcal{G}_m^{A,C}(l_0)$ if $l_0 < d_{A,C}$ from the above definition.

By following the same discussions in the derivation of Ineq. (A67), we obtain

$$\|\tilde{H}_{\bar{1}}(A:C|B)\| \leq \sum_{v \in A} \sum_{m=1}^\infty \sum_{l_0 \geq d_{A,C}} 2(4\beta)^m \sum_{w \in \mathcal{G}_{r,m}^v(l_0)} \frac{n_w}{m!} \prod_{s=1}^m N_{X_s|w} \|h_{X_s}\|, \tag{C11}$$

where in this case, the summation of $v \in \partial A_r$ is replaced by $v \in A$ due to $\partial A_\infty = A$ (see Eq. (A66)). Then, by using the inequality (A70), obtain

$$\sum_{w \in \mathcal{G}_{r,m}^v(l_0)} \frac{n_w}{m!} \prod_{s=1}^m N_{X_s|w} \|h_{X_s}\| \leq \frac{e^{1/k}(2e^3 k)^m}{2} \sum_{l_1+l_2+\ldots+l_m=l_0} \prod_{j=1}^m \tilde{g}_{l_j}, \tag{C12}$$

where we defined $\tilde{g}_l$ in Eq. (C2). By combining the inequalities (C11) and (C12), we obtain

$$\|\tilde{H}_{\bar{1}}(A:C|B)\| \leq \sum_{v \in A} \sum_{m=1}^\infty \sum_{l_1+l_2+\cdots+l_m \geq d_{A,C}} e^{1/k}(8e^3 k\beta)^m \prod_{j=1}^m \tilde{g}_{l_j}. \tag{C13}$$

We can prove the following inequality (see Sec. C 1 a for the proof):

$$\sum_{l_1+l_2+\cdots+l_m \geq l_0} \prod_{j=1}^m \tilde{g}_{l_j} \leq 11^m l_0^{-\alpha} \tag{C14}$$

for arbitrary $l_0 \geq 2\alpha$. By using the above inequality, we obtain

$$\sum_{m=1}^\infty \sum_{l_1+l_2+\cdots+l_m \geq d_{A,C}} e^{1/k}(8e^3 k\beta)^m \prod_{j=1}^m \tilde{g}_{l_j} \leq d_{A,C}^{-\alpha} \sum_{m=1}^\infty e^{1/k}(11\beta/\beta_c)^m \leq \frac{11e^{1/k}\beta/\beta_c}{1 - 11\beta/\beta_c} d_{A,C}^{-\alpha}. \tag{C15}$$

By combining the inequalities (C13) and (C15), we finally obtain

$$\|\tilde{H}_{\bar{1}}(A:C|B)\| \leq \beta|A| \frac{11e^{1/k}/\beta_c}{1 - 11\beta/\beta_c} d_{A,C}^{-\alpha}. \tag{C16}$$

In the same way, we can derive the inequality such that $|A|$ is replaced by $|C|$ in (C16). By combining the above inequality with (A21), we prove Theorem 5. $\square$

a. Proof of the inequality (C14)

For the proof, we start from the following form:

$$\sum_{l_1+l_2+\cdots+l_m \geq l_0} \prod_{j=1}^m \tilde{g}_{l_j} \leq \eta_m l_0^{-\alpha}. \tag{C17}$$

We, in the following, construct a recurrence relation to determine η_m. First, Eq. (C3) immediately implies

$$\sum_{l_1+l_2+\cdots+l_m \geq l_0} \prod_{j=1}^m \tilde{g}_{l_j} \leq \prod_{j=1}^m \sum_{l_j=1}^\infty \tilde{g}_{l_j} \leq 1. \tag{C18}$$

Based on the inequalities (C17) and (C18), we consider the case of $m+1$ as

$$\sum_{l_1+l_2+\cdots+l_{m+1} \geq l_0} \prod_{j=1}^{m+1} \tilde{g}_{l_j} \leq \sum_{l_{m+1}=1}^\infty \tilde{g}_{l_{m+1}} \sum_{l_1+l_2+\cdots+l_m \geq l_0-l_{m+1}} \prod_{j=1}^m \tilde{g}_{l_j}$$

$$\leq \eta_m \sum_{l_{m+1}=1}^\infty \tilde{g}_{l_{m+1}} \max\left[(l_0 - l_{m+1})^{-\alpha}, 1\right]$$

$$\leq \eta_m \sum_{l=1}^{l_0-1} \tilde{g}_l (l_0 - l)^{-\alpha} + \eta_m \sum_{l \geq l_0} \tilde{g}_l \leq \eta_m \sum_{l=1}^{l_0-1} \tilde{g}_l (l_0 - l)^{-\alpha} + \eta_m l_0^{-\alpha}, \tag{C19}$$

where the last inequality comes from the inequality (C3) with $R = l_0$. In order to upper-bound the first term, we decompose the summation as follows:

$$\sum_{l=1}^{l_0-1} \tilde{g}_l (l_0 - l)^{-\alpha} = \left(\sum_{l \in [1,l_1)} + \sum_{l \in [l_1,l_2)} + \sum_{l \in [l_2,l_3)} + \sum_{l \in [l_3,l_0)} \right) \tilde{g}_l (l_0 - l)^{-\alpha}, \tag{C20}$$

for $\alpha > 2$, where $l_1 = \lceil l_0/\alpha \rceil$, $l_2 = \lceil l_0/2 \rceil$, $l_3 = \lceil l_0 - l_0/\alpha \rceil$. For $\alpha \leq 2$, we decompose as

$$\sum_{l=1}^{l_0-1} \tilde{g}_l (l_0 - l)^{-\alpha} = \left(\sum_{l \in [1,l_2)} + \sum_{l \in [l_2,l_0)} \right) \tilde{g}_l (l_0 - l)^{-\alpha}. \tag{C21}$$

Next, for arbitrary choice of $[x,y]$ $(1 \leq x \leq y \leq l_0 - 1)$, we have

$$\sum_{l \in [x,y)} \tilde{g}_l (l_0 - l)^{-\alpha} \leq (l_0 - y + 1)^{-\alpha} \sum_{l \in [x,y)} \tilde{g}_l \leq (l_0 - y + 1)^{-\alpha} \sum_{l \geq x} \tilde{g}_l \leq (l_0 - y + 1)^{-\alpha} x^{-\alpha}, \tag{C22}$$

which reduces the inequality (C20) to

$$\begin{aligned}
\sum_{l=1}^{l_0-1} \tilde{g}_l (l_0 - l)^{-\alpha} \leq & (l_0 - \lceil l_0/\alpha \rceil + 1)^{-\alpha} + (l_0 - \lceil l_0/2 \rceil + 1)^{-\alpha} \lceil l_0/\alpha \rceil^{-\alpha} \\
& + (l_0 - \lceil l_0 - l_0/\alpha \rceil + 1)^{-\alpha} \lceil l_0/2 \rceil^{-\alpha} + \lceil l_0 - l_0/\alpha \rceil^{-\alpha} \\
\leq & 2(l_0 - l_0/\alpha)^{-\alpha} + 2(l_0/2)^{-\alpha} (l_0/\alpha)^{-\alpha} \\
= & 2l_0^{-\alpha} \left[\frac{1}{(1 - 1/\alpha)^\alpha} + \left(\frac{2\alpha}{l_0} \right)^\alpha \right] \leq 10 l_0^{-\alpha}
\end{aligned} \tag{C23}$$

for $\alpha > 2$, where we use $1/(1 - 1/x)^x \leq 4$ for $x \geq 2$ and $l_0 \geq 2\alpha$ from the condition of the theorem. For $\alpha \leq 2$, we also obtain

$$\sum_{l=1}^{l_0-1} \tilde{g}_l (l_0 - l)^{-\alpha} \leq 2(l_0/2)^{-\alpha} \leq 8 l_0^\alpha \tag{C24}$$

from the decomposition (C21), where we use $2^\alpha \leq 4$ for $\alpha \leq 2$.

By applying the inequalities (C23) and (C24) to the inequality (C19), we obtain

$$\sum_{l_1+l_2+\cdots+l_{m+1} \geq l_0} \prod_{j=1}^{m+1} \tilde{g}_{l_j} \leq 11 \eta_m l_0^{-\alpha}, \tag{C25}$$

which gives rise to

$$\eta_{m+1} \leq 11 \eta_m. \tag{C26}$$

This yields the inequality (C14). This completes the proof. $\square$