

## Improved Thermal Area Law and Quasilinear Time Algorithm for Quantum Gibbs States

Tomotaka Kuwahara<sup>1,2,\*</sup> Álvaro M. Alhambra<sup>3,†</sup> and Anurag Anshu<sup>4,‡</sup><sup>1</sup>Mathematical Science Team, RIKEN Center for Advanced Intelligence Project (AIP),  
1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan<sup>2</sup>Interdisciplinary Theoretical and Mathematical Sciences Program (iTHEMS) RIKEN 2-1,  
Hirosawa, Wako, Saitama 351-0198, Japan<sup>3</sup>Max-Planck-Institut für Quantenoptik, D-85748 Garching, Germany<sup>4</sup>University of California, Berkeley, California 94720, USA (Received 18 August 2020; revised 16 November 2020; accepted 22 December 2020; published 9 March 2021)

One of the most fundamental problems in quantum many-body physics is the characterization of correlations among thermal states. Of particular relevance is the thermal area law, which justifies the tensor network approximations to thermal states with a bond dimension growing polynomially with the system size. In the regime of sufficiently low temperatures, which is crucially important for practical applications, the existing techniques do not yield optimal bounds. Here, we propose a new thermal area law that holds for generic many-body systems on lattices. We improve the temperature dependence from the original  $\mathcal{O}(\beta)$  to  $\mathcal{O}(\beta^{2/3})$  up to a logarithmic factor, thereby suggesting subballistic propagation of entanglement by imaginary-time evolution. This qualitatively differs from the real-time evolution, which usually induces linear growth of entanglement. We also prove analogous bounds for the Rényi entanglement of purification and the entanglement of formation. Our analysis is based on a polynomial approximation to the exponential function which provides a relationship between the imaginary-time evolution and random walks. Moreover, for one-dimensional (1D) systems with  $n$  spins, we prove that the Gibbs state is well approximated by a matrix product operator with a sublinear bond dimension for  $\beta = o[\log(n)]$ . This proof allows us to rigorously establish, for the first time, a quasilinear time classical algorithm for constructing a matrix product state representation of 1D quantum Gibbs states at arbitrary temperatures of  $\beta = o[\log(n)]$ . Our new technical ingredient is a block decomposition of the Gibbs state that bears a resemblance to the decomposition of real-time evolution given by Haah *et al.* [*Proceedings of the 2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, New York, 2018), pp. 350–360].

DOI: [10.1103/PhysRevX.11.011047](https://doi.org/10.1103/PhysRevX.11.011047)Subject Areas: Condensed Matter Physics,  
Quantum Physics,  
Quantum Information**I. INTRODUCTION****A. Background**

One of the most important challenges in quantum many-body physics is to understand their thermal equilibrium properties. Recently, with the advent of large quantum simulators [1–5], the size and controllability of quantum Gibbs states accessible for experiments have dramatically improved. In fact, recent experiments have even succeeded

in implementing imaginary-time evolution [6]. These developments are of considerable interest for quantum computation, because quantum Gibbs states play crucial roles in quantum machine learning [7–14] and quantum algorithms such as semidefinite program solvers [15–17]. Beyond quantum computation, understanding and characterizing quantum Gibbs states is relevant to many open problems in quantum statistical physics and condensed matter physics. Thus, understanding (i) the nature of entanglement structures in quantum Gibbs states and (ii) their simulability via tensor network methods is of great interest.

It is now widely accepted that the area law plays a crucial role [18,19] in the characterization of low-temperature physics of many-body systems. This law states that the entanglement entropy between two subsystems is at most as large as the size of their boundaries. A similar notion also applies to finite-temperature systems. Although a rigorous proof of the area law at the zero temperature appears to be a

\*tomotaka.kuwahara@riken.jp

†alvaro.alhambra@mpq.mpg.de

‡anuraganshu@berkeley.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

notoriously challenging problem [20–27], an analogous area law at finite temperatures has been proved by Wolf *et al.* [28] in a simple and elegant manner. The authors prove the following inequality:

$$I(L:R)_{\rho_\beta} \leq 2\beta \|H_{\partial L}\| \propto \beta |\partial L|, \quad (1)$$

with  $\|\cdots\|$  being the operator norm and  $\partial L$  being the surface region of  $L$ , where  $I(L:R)_{\rho_\beta}$  is the mutual information between the subsets  $L$  and  $R$  [see Eq. (7) below] and  $H_{\partial L}$  denotes the boundary interaction Hamiltonian. The upper bound (1) roughly denotes that the correlations between two complementary regions are concentrated around a distance  $\mathcal{O}(\beta)$  of their boundary.

The thermal area law (1) is optimal at high temperatures [ $\beta \approx \mathcal{O}(1)$ ], because the dependence on  $|\partial L|$  cannot be improved. One may similarly expect that, at low temperatures ( $\beta \gg 1$ ), linear dependence on  $\beta$  should be optimal. This expectation is suggested by the theory of *belief propagation* [29], which indicates that the nonlocal quantum effects can be induced in a length scale of  $\mathcal{O}(\beta)$ . However, there are no definite numerical or theoretical examples that achieve the upper bound (1). Indeed, for specific systems [30–32], we can get much better area-law bounds than (1). This result motivates the possibility of the following improvement of the thermal area law (see Fig. 1):

$$I(L:R)_{\rho_\beta} \lesssim \beta^\gamma |\partial L| \quad (\gamma < 1). \quad (2)$$

Any improvement along these lines is intimately associated with new advances in our understanding of the low-temperature physics. For instance, the widely known relation between area laws and tensor networks suggests that the identification of the minimum  $\gamma_c$  would also lead to optimal representations of Gibbs states. This outcome would result in faster algorithms for computing local expectation values and evaluating the partition functions.

We now turn our attention to the simulability of the quantum Gibbs state. There exists a large number of classical [29,33–50] and quantum [51–59] algorithms to study the properties of the quantum Gibbs states. At high temperatures [ $\beta = \mathcal{O}(1)$ ], the Gibbs states have numerous analytical properties, such as the exponential decay of bipartite correlations [60–65], the large deviation principle [66–68], and the approximate quantum Markov property [69,70]. As a consequence, in this temperature regime, the Gibbs states are proved to be generated by a finite-depth quantum circuit [58,70], and the quantum partition function can be computed in polynomial time [70–73].

Unfortunately, at lower temperatures, computational complexity theory results severely limit the applicability of the algorithms discussed above. Indeed, computing the partition function of Gibbs states in two and higher dimensions is already known to be NP-hard [74,75] (see also Ref. [76]) except for special cases (e.g., ferromagnetic

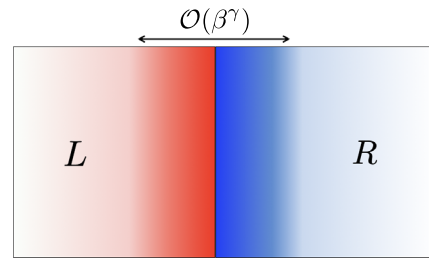


FIG. 1. Schematic depiction of our problem. By decomposing the total system into  $L$  and  $R$ , we consider the mutual information  $I(L:R)_{\rho_\beta}$  between  $L$  and  $R$ . Then, the thermal area law in Ref. [28] gives  $I(L:R)_{\rho_\beta} \lesssim \beta |\partial L|$  ( $\gamma = 1$  in the above picture). We aim to establish a new thermal area law in the form of  $I(L:R)_{\rho_\beta} \lesssim \beta^\gamma |\partial L|$  with  $\gamma < 1$ . In particular, it is a highly nontrivial and fundamental question to identify the best exponent  $\gamma_c$  for which the thermal area law holds in generic many-body systems. Our main result provides the nontrivial upper bound of  $\gamma_c \leq 2/3$ .

spin systems [77,78]). This bottleneck is serious for several practical applications in which the Gibbs states are employed at low temperatures. For example, in the quantum algorithm for semidefinite programming [15], the quantum Gibbs states with  $\beta = \mathcal{O}[\log(n)]$  ( $n$ , system size;  $\beta$ , the inverse temperature) must be sampled. Similar challenges are faced in the imaginary-time evolution, the implementation of which is a central aim of near-term quantum devices [6,79–85]. Thus far, below a threshold temperature where the cluster expansion technique does not work [86], little is known about the universal properties of Gibbs states that may hold independent of the system’s details. This lack provides a strong motivation to identify the optimal thermal area laws.

## B. Description of the main results

For the first main result of the present study, we prove the inequality (2) for  $\gamma = 2/3$ . On the other hand, we also prove the lower bound of  $\gamma_c \geq 1/5$ , using the example constructed in Ref. [87] (see Sec. III B), which means

$$1/5 \leq \gamma_c \leq 2/3.$$

There are two remarks: (i) The result is applicable only to a finite-dimensional lattice, while there may be a counterexample in general graph systems [88], and (ii) in high-dimensional cases, the obtained result is slightly weaker, as is given by  $I(L:R)_{\rho_\beta} \lesssim \beta^\gamma |\partial L| \log(\beta^\gamma |\partial L|)$  (see Theorem 1). To understand why the result is counterintuitive at first sight, let us consider the case of real-time evolution  $e^{iHt}$ . The small-incremental-entangling (SIE) theorem [89–92] predicts the linear increase of the entanglement with respect to time, which translates to the fact that the Schmidt rank of the operator  $e^{iHt}$  grows as  $e^{\mathcal{O}(t)}$ . This result suggests the same linear dependence for the imaginary-time evolution operator

$e^{-\beta H}$ . However, the inequality (2) shows that the scaling of the exponent is sublinear in  $\beta$ . This result means that the entropy growth due to the imaginary-time evolution is more diffusive in nature. We explain this difference in Sec. III A, which can be traced back to a better polynomial approximation to  $e^{-x}$  compared with  $e^{-ix}$  [93]. This polynomial approximation is caused by a random walk interpretation of the Chebyshev basis expansion of  $e^{-x}$  (see Sec. III A), which is not available for  $e^{-ix}$ . This random walk interpretation further suggests that the entropy production in the imaginary-time evolution is diffusive.

The improved area law not only is of fundamental interest but also provides important insights regarding the efficient representation of the quantum Gibbs states. In previous studies [64,94,95], the approximations by matrix product operators (MPOs) and projected entangled pair operators (PEPOs) are investigated through cluster expansion techniques. Furthermore, Ref. [95] explicitly gives the PEPO or MPO construction scheme with the bond dimensions of

$$D = (n/\epsilon)^{\mathcal{O}(\beta)} \quad (\epsilon, \text{approximation error}). \quad (3)$$

If we use the cluster expansion technique, this result is expected to be the best estimation. However, the polynomial-size bond dimension of  $n^{\mathcal{O}(\beta)}$  may still be a significant overestimation. Improvements are strongly motivated by the practical use of tensor network techniques in approximating thermal states [33,41,48], which appears to be much more successful than is guaranteed by the current analytical bounds.

Our second main result focuses on classical algorithms for approximating thermal states in one dimension (1D). By applying our new analyses, we establish a sublinear dependence of the bond dimension of the MPO approximation to the thermal state as

$$D = e^{\tilde{\mathcal{O}}(\beta^{2/3}) + \tilde{\mathcal{O}}[\sqrt{\beta \log(n/\epsilon)}]} \quad (4)$$

with  $\epsilon$  the approximation error, where we write  $\mathcal{O}[n \log(n)]$  as  $\tilde{\mathcal{O}}(n)$  by using the notation  $\tilde{\mathcal{O}}$ . The estimated bond dimension is smaller than any power of  $(n/\epsilon)$  and is well suited for numerical simulations.

Finally, we consider the computational complexity of the construction of the MPO, which approximates 1D quantum Gibbs states. Establishing provably efficient quasilinear algorithms for physical systems is a central target in the field of Hamiltonian complexity [96,97]. The general difficulty lies in that the existence of an efficient MPO description (4) does not necessarily imply an efficient algorithm to find such a description [98,99]. So far, the state-of-the-art algorithm [95] is based on cluster expansion, and MPO construction requires a computation cost which is proportional to  $n \times (n/\epsilon)^{\mathcal{O}(\beta)}$ , where the estimated exponent of  $(n/\epsilon)$  is usually impractically large. However, most classical heuristic algorithms employed in practice

usually require only (quasi)linear computational time with respect to the system size [29,33–50]. We, for the first time, give a quasilinear time algorithm that constructs the approximate MPO, with a run time of

$$n \times e^{\tilde{\mathcal{O}}(\beta) + \tilde{\mathcal{O}}[\sqrt{\beta \log(n/\epsilon)}]},$$

which is quasilinear in  $(n/\epsilon)$  for arbitrary  $\beta = o[\log(n)]$ .

The rest of this paper is organized as follows. In Sec. II, we formulate the precise setting and notations used throughout the paper. In Sec. III, we state the main theorems on the area law and the MPO approximation. In addition, in Sec. III A, we show the relationship between imaginary-time evolution and the random walk, and in Sec. III B, we show the lower bound on the critical  $\gamma_c$ . In Sec. IV, we give the quasilinear algorithm to compute the MPO approximation of the 1D quantum Gibbs states. We also provide a brief explanation regarding why the algorithm works well. In Sec. V, we discuss several physical implications from our analytical techniques. The proofs of the main statements are given in Sec. VI. Finally, in Sec. VII, we summarize the paper, along with a brief discussion. To concentrate on the physics, we provide the more intricate aspects of the proofs in Appendixes.

## II. SETUP AND NOTATION

We consider a quantum system with  $n$  qudits, each of which has a  $\varsigma$ -dimensional Hilbert space. We denote the Hilbert space dimension of a subset  $S \subseteq \Lambda$ , where  $\Lambda$  is a lattice, by  $\mathcal{D}_S$ . For the present discussion, let us restrict ourselves to the case of 1D lattice; we consider higher-dimensional lattices later (see Sec. II A). We define the Hamiltonian  $H$  as follows:

$$H = \sum_{i=1}^n h_{i,i+1}, \quad \|h_{i,i+1}\| + \|h_{i-1,i}\| \leq g, \quad (5)$$

where  $h_{i,i+1}$  contains interactions between  $i$  and  $i+1$  and  $\|\cdot\|$  is the operator norm. By taking the energy units appropriately, we set  $g=1$ . Here, we assume two-body interactions of the Hamiltonian, but the generalization to arbitrary  $k$ -body interactions (i.e.,  $k$ -local Hamiltonian) with  $k = \mathcal{O}(1)$  is straightforward (see Appendix A). For an arbitrary operator  $O$ , we define the Schmidt rank  $\text{SR}(O, i)$  as the minimum integer such that

$$O = \sum_{m=1}^{\text{SR}(O,i)} O_{\leq i,m} \otimes O_{> i,m}, \quad (6)$$

where  $\{O_{\leq i,m}\}$  and  $\{O_{> i,m}\}$  are operators acting on subsets  $\{j\}_{j \leq i}$  and  $\{j\}_{j \geq i+1}$ , respectively. Note that the Schmidt rank  $\text{SR}(h_{i,i+1}, i)$  is always smaller than the local Hilbert dimension  $\varsigma$  [i.e.,  $\text{SR}(h_{i,i+1}, i) \leq \varsigma$ ].

Throughout the paper, we focus on the Gibbs state  $\rho_\beta$  with an inverse temperature  $\beta$ :

$$\rho_\beta = \frac{e^{-\beta H}}{\text{tr}(e^{-\beta H})}.$$

To extend the concept of the entanglement area law from the ground states to finite temperatures, we often utilize the mutual information  $I(L:R)_\rho$  as in Ref. [28]. The mutual information  $I(L:R)_\rho$  reduces to entanglement entropy when  $\rho$  is given by a quantum pure state. For an arbitrary decomposition of the total system  $\Lambda$  into  $\Lambda = L \cup R$ , it is defined as

$$I(L:R)_{\rho_\beta} := S(\rho_\beta^L) + S(\rho_\beta^R) - S(\rho_\beta), \quad (7)$$

where  $S(\cdots)$  is the von Neumann entropy, i.e.,  $S(\rho) := -\text{tr}[\rho \log(\rho)]$ , and  $\rho_\beta^L$  ( $\rho_\beta^R$ ) is the reduced density matrix in subsets  $L$  ( $R$ ). We define the subsets  $L$  and  $R$  as  $L = \{1, 2, \dots, i_0\}$  and  $R = \{i_0 + 1, i_0 + 2, \dots, n\}$ , respectively. Then, the boundary Hamiltonian  $H_{\partial L}$  is given by  $h_{i_0, i_0+1}$ , which gives the previously known thermal area law (1) of

$$I(L:R)_{\rho_\beta} \leq 2\beta \|h_{i_0, i_0+1}\|. \quad (8)$$

For a more detailed characterization of the structure of the quantum Gibbs state, we focus on the MPO representation. We aim to approximate the Gibbs state  $\rho_\beta$  by the following operator:

$$M_D = \sum_{\substack{s_1, s_2, \dots, s_n=1 \\ s'_1, s'_2, \dots, s'_n=1}}^{\xi} \text{tr}(A_1^{[s_1, s'_1]} A_2^{[s_2, s'_2]} \dots A_n^{[s_n, s'_n]}) |s_1, s_2, \dots, s_n\rangle \langle s'_1, s'_2, \dots, s'_n|, \quad (9)$$

where each of the matrices  $\{A_i^{[s_i, s'_i]}\}_{i, s_i, s'_i}$  is described by the  $D \times D$  matrix. We refer to the matrix size  $D$  as the bond dimension. By choosing  $D$  to be sufficiently large as  $D = e^{O(n)}$ , we can describe arbitrary operators in the form of MPO; however, only a relatively small bond dimension is often required in practical applications [e.g.,  $D = o(n)$ ]. To relate this requirement to the mutual information, notice that  $I(L:R)$  can, in general, be bounded by the bond dimension of the purification. For the subclass of MPOs with local purifications [28], this bound cannot exceed  $2 \log D$  for an arbitrary decomposition  $\Lambda = L \sqcup R$ , although no upper bound exists for general MPOs [100,101]. To circumvent this difficulty, we directly give a bound on the bond dimension of a purification that scales as Eq. (4) (see Sec. VI A). The primary problem is to estimate how large the bond dimension needs to be to achieve a certain precision error.

In quantitatively estimating the approximation error, we utilize the Schatten  $p$ -norm, which is defined for arbitrary operator  $O$  as follows:

$$\|O\|_p := [\text{tr}(O^\dagger O)^{p/2}]^{1/p}. \quad (10)$$

Note that  $\|O\|_1$  corresponds to the trace norm and  $\|O\|_\infty$  corresponds to the standard operator norm, which we denote by  $\|O\|$  for simplicity. When  $O$  is a density operator, it is a common practice to consider the trace norm (i.e.,  $p = 1$ ) for the approximation error. However, for estimating approximation errors in the present context, calculations in terms of the general Schatten  $p$ -norm are crucially important. For example, let us consider the situation where we have obtained a good approximation for  $O$  by  $\tilde{O}$  and are interested in approximating  $O^s$  by  $\tilde{O}^s$ . In order to achieve  $\|O^s - \tilde{O}^s\|_1 \ll 1$ , we need to prove  $\|O - \tilde{O}\|_s \ll 1$ ; evidently, approximation solely in terms of the trace norm is not sufficient. This point is clarified in Lemmas 11 and 12 in Appendix B, which are based on the analyses in Ref. [95]. This kind of the technique is crucial in developing a quasilinear time algorithm for the quantum Gibbs states (Sec. IV).

The state-of-the-art results [95] ensure the existence of  $M_D$  such that  $\|\rho_\beta - M_D\|_1 \leq \epsilon$  with the bond dimension as in Eq. (3). The bond dimension  $D$  is roughly related to the mutual information  $I(L:R)_{\rho_\beta}$  as  $I(L:R)_{\rho_\beta} \lesssim \log(D)$ , and, thus, the estimation (3) implies the area-law bound of Eq. (8).

### A. High-dimensional setup

In extending to the high-dimensional systems, we consider a quantum system on a  $d$ -dimensional rectangular lattice with  $d$  the spatial dimension (we note that our analysis can also be applied to other lattices). For simplicity of notation, we consider nearest-neighbor interactions as follows:

$$H = \sum_{\langle i, j \rangle} h_{i, j}, \quad \max_{i \in \Lambda} \sum_j \|h_{i, j}\| \leq g, \quad (11)$$

where  $\langle i, j \rangle$  denotes the pairs of adjacent qudits and  $\|\cdots\|$  is the operator norm. By taking the energy unit appropriately, we set  $g = 1$ .

For convenience, we consider a vertical cut of the total system (see Fig. 1, for example); however, the same argument can be applied to a rectangular cut. For any partition  $\Lambda = L \sqcup R$ , we define an upper bound with the size of the surface region as  $|\partial\Lambda|$ , which is as large as  $O(n^{1-1/d})$ . Note that  $|\partial\Lambda| = 1$  in the 1D lattice.

## III. IMPROVED THERMAL AREA LAW

We first show our main result in the thermal area law. The following theorem holds for arbitrary lattice dimensions.

*Theorem 1.*—For an arbitrary cut  $\Lambda = L \cup R$ , the mutual information  $I(L:R)_{\rho_\beta}$  is upper bounded by



$$I(L:R)_{\rho_\beta} \leq C\beta^{2/3}|\partial\Lambda|[\log^{2/3}(|\partial\Lambda|) + \log(\beta)], \quad (12)$$

where  $C$  is a constant of  $\mathcal{O}(1)$ . In particular, for one-dimensional systems ( $|\partial\Lambda| = 1$ ), we have

$$I(L:R)_{\rho_\beta} \leq C\beta^{2/3} \log(\beta) = \tilde{\mathcal{O}}(\beta^{2/3}). \quad (13)$$

We show the proof in Sec. VI A. The above result has a logarithmic correction of  $\log(|\partial\Lambda|)$  to the area law in high dimensions. Even then, for  $\beta \gtrsim \log^2(|\partial\Lambda|)$ , our result provides a qualitatively better upper bound than the previous one (1). We expect that this correction should be removed using refined analyses on the Schmidt rank for polynomials of the Hamiltonian.

Moreover, for the MPO representation of the 1D quantum Gibbs states, we obtain the following theorem.

*Theorem 2.*—For arbitrary 1D quantum Gibbs state  $\rho_\beta$ , there exists a MPO  $M_D$  as in Eq. (9) such that, for the Schatten  $p$ -norm with  $p = 1$  and  $p = 2$ ,

$$\|\rho_\beta - M_D\|_p \leq \epsilon \|\rho_\beta\|_p \quad (14)$$

with

$$D \leq \exp[\tilde{q}_\epsilon^* \log(\tilde{q}_\epsilon^*)], \quad (15)$$

where  $\tilde{q}_\epsilon^* := C'_0 \max\{\beta^{2/3}, [\beta \log(\beta n/\epsilon)]^{1/2}\}$ .

This proof is shown in Sec. VI B. We believe that the above MPO could also be used to construct a quantum circuit with depth polynomial in the stated bond dimensions. As far as we know, explicit constructions of quantum circuits for generic MPOs [or matrix product states (MPSs)] have been an open problem except for special cases [102–104].

We remark on the generalization to high-dimensional cases. As for the MPO representation, we can improve the  $\beta$  dependence of the bond dimension in high dimensions. However, the MPO representation for high-dimensional systems is not useful, since the bond dimension is inherently subexponentially large with respect to the system size. For an arbitrary bipartition  $\Lambda = L \sqcup R$ , the bond dimension scales as  $e^{\mathcal{O}(\beta^{2/3})|\partial\Lambda|} = e^{\mathcal{O}(\beta^{2/3})n^{1-1/d}}$ . In order to obtain a meaningful representation for a high-dimensional Gibbs state, we need to consider the PEPO [64,94,95]. We expect that the bond dimension of the PEPO might be also sublinear as Eq. (15) in order to achieve a good approximation (14). So far, this remains open and one of the most important future directions (see also Sec. VII).

Now, we discuss the key principles that allow us to improve the original thermal area law (see Appendix C for the details). Our analysis utilizes various recent techniques employed in the proofs of the area law for ground states [23,26,27]. Inspired by these studies, we construct an approximation of the quantum Gibbs state using an appropriate polynomial of low degree [93] and then perform a Schmidt-rank analysis adapted from Ref. [23].

As mentioned in the introduction, the main insight is that the polynomial used by us satisfies the random walk property, which we explain below.

### A. Physical intuition from the random walk behavior

Before the main discussion, let us consider an illustrative example of the random walk behavior in imaginary-time evolution. We here consider a one-particle tight-binding model as

$$H = \sum_{x=-R}^R (|x\rangle\langle x+1| + |x+1\rangle\langle x| - 2|x\rangle\langle x|), \quad (16)$$

where  $|x\rangle$  is the state of the particle on site  $x$ . Then, the real-time Schrödinger equation gives the ballistic propagation of the particle. We consider a time-evolved quantum state  $|0(t)\rangle = e^{-iHt}|0\rangle$ , where the initial state  $|0\rangle$  is the localized state on  $x = 0$ . In Fig. 2(a), we show the fluctuation of the position, which is given by the square root of the variance  $\text{Var}(X) := \langle 0(t)|X^2|0(t)\rangle - [\langle 0(t)|X|0(t)\rangle]^2$ , where  $X = \sum_{x=-R}^R x|x\rangle\langle x|$ . In contrast, the imaginary-time Schrödinger equation is formally equivalent to the random walk differential equation. Hence, the fluctuation for the state  $|0(-i\beta)\rangle = e^{-\beta H}|0\rangle$  grows diffusively with time  $t$  [see Fig. 2(b)]. This result indicates that the imaginary-time evolution may generally induce a diffusive propagation of information in quantum many-body systems. In the following sections, we mathematically justify this intuition.

Suppose  $x$  is fixed to be in a range  $[0, b]$ . As shown in Ref. [93],  $e^{-x}$  can be approximated by a polynomial of degree  $\mathcal{O}(b^{1/2})$ , for a constant error. This approximation is the consequence of a random walk that is concentrated around degree  $\mathcal{O}(b^{1/2})$  after  $b$  steps. Let us introduce  $y \in [-1, 1]$  such that  $x = b(1+y)/2$  and  $e^{-x} = (e^{-(1/2)(1+y)})^b$ . Below, we show that the exponential function  $e^{-b(1+y)/2}$  ( $b \in \mathbb{N}$ ) can be expanded in terms of the Chebyshev polynomials as (see also Fig. 3)

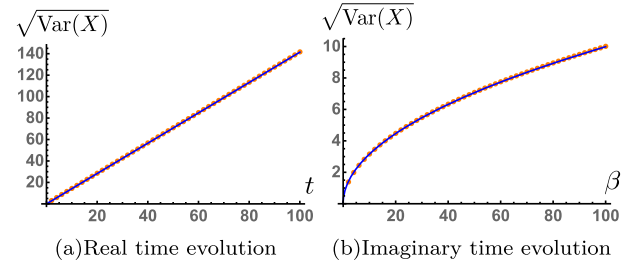


FIG. 2. Comparison between real-time evolution and imaginary time evolution in the tight-binding model (16) with  $R = 500$ . In (a) and (b), we plot the fluctuation of the position  $\sqrt{\text{Var}(X)}$  after the real and imaginary time evolutions, respectively. Here, the state at the initial time is given by  $|0\rangle$ . The fitting functions for (a) and (b) are given by  $\sqrt{2}t$  and  $0.998769\beta^{1/2}$ , respectively. This clearly indicates that the real-time evolution induces a ballistic propagation, whereas the imaginary-time evolution induces a diffusive propagation.

$$\begin{aligned}
 (e^{-(1/2)(1+y)})^b &= \sum_{r_b=-\infty}^{\infty} P(r_b) T_{r_b}(y) \\
 &= \sum_{r_b=-\infty}^{\infty} \left( p(r_b|r_{b-1}) \sum_{r_{b-1}=-\infty}^{\infty} p(r_{b-1}|r_{b-2}) \cdots \sum_{r_2=-\infty}^{\infty} p(r_2|r_1) \sum_{r_1=-\infty}^{\infty} p(r_1|r_0) \right) T_{r_b}(y) \quad (17)
 \end{aligned}$$

with  $r_0 = 0$ , where  $T_r(x)$  is the Chebyshev polynomial and  $p(r|r')$  is a random walk probability from  $r_{b-1}$  to  $r_b$  which is defined below.

For its application to  $e^{-\beta H}$ , we choose  $b = \beta \|H\|$ . Because the Schmidt rank and polynomial degree are closely related [23], we get a diffusive interpretation of the Schmidt rank of  $e^{-\beta H}$ . We thus infer a sublinear  $\beta$  dependence of the mutual information, namely,  $\gamma_c < 1$ . There are two main issues while achieving this value. First, the above polynomial gives an approximation to  $e^{-\beta H}$  only in the operator norm, whereas we are searching for an approximation in a family of norms. Second, even for a constant error approximation in the operator norm, degree  $\sqrt{b} = \sqrt{\beta \|H\|}$  scales with the system size. We solve both the problems using the quantum belief propagation in 1D and a refined version of Suzuki-Trotter decomposition in higher dimensions, which allows us to reduce the problem to a local Hamiltonian  $H_S$ , where  $S$  is a much smaller region. The loss incurred because of the belief propagation and the conversion from the operator norm to other norms leads to our main result of  $\gamma_c \leq 2/3$ .

### 1. Derivation of Eq. (17)

As a first step, we expand

$$e^{-(1/2)(1+y)} = \sum_{j=0}^{\infty} \frac{e^{-1/2}}{2^j j!} \cdot (-y)^j, \quad (18)$$

which is an expectation of  $(-y)^j$  according to the distribution  $q(j) := (e^{-1/2}/2^j j!)$ . Next, we introduce Chebyshev polynomials  $T_r(y)$  (for an integer  $r$ ) and utilize the observation from Ref. [93] that, for  $j > 0$  and integer  $k$ ,

$$\begin{aligned}
 (-y)^j T_r(-y) &= \sum_{r'=-\infty}^{\infty} B_j(r'|r) T_{r'}(-y), \\
 B_j(r'|r) &= 2^{-j} \binom{j}{(j+r'-r)/2}, \quad (19)
 \end{aligned}$$

where we set  $\binom{j}{s+1/2} = 0$  ( $s \in \mathbb{N}$ ) and  $\binom{j}{s} = 0$  for  $s < 0$  and  $s > j$ . Here,  $B_j(r'|r)$  is the binomial distribution which is centered at  $r$  with a variance of  $\sqrt{j}$  (see also Ref. [105]). Now, we have all the tools to set up the random walk over integers. By combining Eqs. (18) and (19), we start with the first random walk step of

$$e^{-(1/2)(1+y)} = \sum_{r_1=-\infty}^{\infty} p(r_1|0) T_{r_1}(y), \quad (20)$$

where the symmetric distribution  $p(r_1|0)$  [with mean 0 and variance  $\mathcal{O}(1)$ ] is defined using  $p(r_1|0) := \sum_{j=0}^{\infty} q(j) B_j(r_1|0)$ . The subsequent steps are obtained by writing

$$\begin{aligned}
 T_{r_1}(y) e^{-(1/2)(1+y)} &= \sum_{j=0}^{\infty} \frac{e^{-1/2}}{2^j j!} \cdot T_{r_1}(y) (-y)^j \\
 &= \sum_{r_2=-\infty}^{\infty} p(r_2|r_1) T_{r_2}(y)
 \end{aligned}$$

with  $p(r_2|r_1) := \sum_{j=0}^{\infty} q(j) B_j(r_2|r_1)$ . One can show that the function  $p(r_2|r_1)$  is symmetric around its mean  $r_1$  and has a variance of 0.5 [see Fig. 3 for the shape of  $p(r|2)$ ]. By repeating the process, we can arrive at Eq. (17). Thus,  $(e^{-(1/2)(1+y)})^b$  is an expectation over  $T_r(y)$ , according to a distribution obtained by performing  $b$  steps of a symmetric random walk with constant variance. It is now clear that the degree is strongly concentrated around  $\mathcal{O}(b^{1/2})$ . This random walk behavior is not available for  $e^{ix}$ , because the distribution  $p(r_2|r_1)$  is not given by a real number. It leads to  $\mathcal{O}(b)$  approximate degree for real-time evolution.

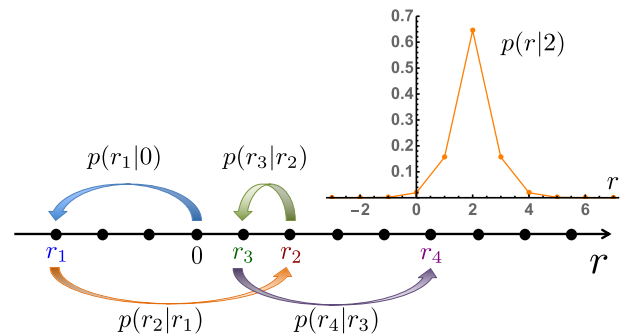


FIG. 3. Schematic picture of the random walk. The exponential function  $e^{-b(1+y)/2}$  is given by the expectation of  $T_{r_b}(x)$  with the probability  $P(r_b)$ , as in Eq. (17). The probability  $P(r_b)$  is generated from the  $b$ -step random walk. In each step, the probability from  $r$  to  $r'$  is given by  $p(r'|r)$ , which is a symmetric function around  $r$ . In the picture, we give the numerical plot of  $p(r|2)$ , where the shape of  $p(r|r')$  does not depend on  $r'$ .

### B. Lower bound on the critical $\gamma_c$

We here show that the exponent  $\gamma$  in Eq. (2) is at least larger than  $1/5$ . According to Ref. [87], there exists a frustration-free local Hamiltonian system with  $n$  qudits ( $\zeta = 3$ ) such that the half-chain entanglement entropy is linear in system size  $n$ , and the spectral gap  $\Delta$  is given as

$$\Delta = \frac{c_\Delta}{n^4 \log n}, \quad (21)$$

where  $c_\Delta$  is a constant of  $\mathcal{O}(1)$ . For this Hamiltonian, let us consider a quantum Gibbs state at the inverse temperature of  $\beta = 2c_\Delta^{-1} \log(\zeta) n^5 \log n$ . Then, the total weight of the excited state is at most as large as  $\zeta^n e^{-\beta\Delta} = e^{-n \log(\zeta)}$ . Therefore, this Gibbs state is exponentially close to the ground state. Using the Fannes inequality [106], the half-chain mutual information in the Gibbs state is

$$I(L:R)_{\rho_\beta} = \Omega(n) = \frac{\Omega(1)}{\log^{1/5} \beta} \beta^{1/5}, \quad (22)$$

which implies that  $I(L:R)_{\rho_\beta}$  should be at least larger than  $\beta^{1/5}$ .

## IV. QUASILINEAR TIME ALGORITHM FOR 1D GIBBS STATE

### A. Main statement

Here, we show that the classical algorithm generating an MPO approximation of the Gibbs state  $\rho_\beta$  is possible with a run time of  $\mathcal{O}(n^{1+o(1)})$  as long as  $\beta = o[\log(n)]$ . We prove the following theorem.

*Theorem 3.*—For arbitrary  $\beta$ , we can efficiently compute a matrix product operator  $M_\beta$  which approximates  $e^{-\beta H}$  in the sense that

$$\|M_\beta - e^{-\beta H}\|_p \leq \epsilon \|e^{-\beta H}\|_p \quad (\epsilon \leq 1), \quad (23)$$

where the bond dimension of  $M_\beta$  is given by  $\exp(Q_\epsilon)$ . Also, the computational time to calculate  $M_\beta$  is  $n\beta \exp(Q_\epsilon)$  with

$$Q_\epsilon := C \max[\beta, \sqrt{\beta \log(n/\epsilon)}] \log[\beta \log(n/\epsilon)], \quad (24)$$

where  $C$  is an  $\mathcal{O}(1)$  constant. When  $\beta \lesssim \log(n/\epsilon)$  and  $\epsilon = 1/\text{poly}(n)$ , the time complexity is given by

$$n \exp\{\tilde{\mathcal{O}}[\sqrt{\beta \log(n)}]\}. \quad (25)$$

We compare the bond dimensions of  $M_{\beta_0}^{(\beta/\beta_0)}$  with that of the theoretical bound in Eq. (15). For  $\beta \lesssim \log(n)$ , both estimations are in the form of  $e^{\tilde{\mathcal{O}}[\sqrt{\beta \log(n)}]}$ , whereas for  $\beta \gg \log(n)$ , the estimation (15) gives a slightly better bound.

### B. MPO for ground space

We also discuss the consequences regarding the calculation of quantum ground states. Let us assume the following condition for the density of states in an energy shell  $(E-1, E]$  for the low-energy regime [107–109]:

$$\mathcal{N}_{E,1} \leq n^{cE} \quad (26)$$

with  $c$  a constant of  $\mathcal{O}(1)$ , where  $\mathcal{N}_{E,1}$  is the number of eigenstates within the energy shell of  $(E-1, E]$ . This condition is typically observed for quantum Hamiltonians which have a spectral gap between the ground state and the first excited state [107]. Under this assumption, the quantum Gibbs state is approximated by the ground state up to an error of  $1/\text{poly}(n)$  for  $\beta = \mathcal{O}[\log(n)]$ ; i.e.,  $\|\rho_\beta - \rho_\infty\|_1 = 1/\text{poly}(n)$ . Then, the computation of the quantum Gibbs state for  $\beta = \mathcal{O}[\log(n)]$  is closely related to the computation of ground states.

By applying  $\beta = \mathcal{O}[\log(n)]$  to Eq. (25), we obtain the time complexity of an almost polynomial form, as  $n^{\mathcal{O}[\log \log(n)]}$ . This result rigorously justifies the empirical success of the imaginary-time-evolving block decimation (TEBD) methods in the computation of the ground states [43,45,47,48]. Our estimation, however, is still slightly worse than the polynomial form (i.e.,  $n^{\mathcal{O}(1)}$ ). In the case of the gapped ground states, the existing algorithms [98,99] already achieve polynomial computational costs without the assumption (26). Any small improvement of Eq. (25) will allow us to obtain a quasilinear time algorithm for the computation of the ground states under the assumption of Eq. (26).

### C. Details of the algorithm and proof of Theorem 3

The algorithm proceeds as follows. Suppose we are at a high temperature  $\beta_0 \leq 1/16$ . First, the 1D Hamiltonian is split into blocks of length  $\ell_0 = \mathcal{O}[\log(n/\epsilon)]$ , as

$$H = \sum_{j=1}^{n_0} H_j, \quad H_j = \sum_{s=(j-1)\ell_0+1}^{j\ell_0} h_{j,j+1}, \quad (27)$$

where  $n_0$  is the number of blocks. We then write  $e^{-\beta H}$  as follows:

$$e^{-\beta H} = \prod_{j=1}^{n_0} e^{\beta_0 H_{1:j-1}} e^{-\beta_0 H_{1:j}} = \prod_{j=1}^{n_0} \Phi_j, \quad (28)$$

where  $H_{1:j} = \sum_{s \leq j} H_s$  and  $H_{1:0} = \hat{0}$ . Here, the operator  $\Phi_j$  is the nonlocal operator on the qudits  $\{1, 2, \dots, j\ell_0\}$ . We first approximate  $\Phi_j$  by the following operator on the local region:

$$\tilde{\Phi}_j = e^{\beta_0 H_{j-1}} e^{-\beta_0 (H_{j-1} + H_j)}, \quad (29)$$

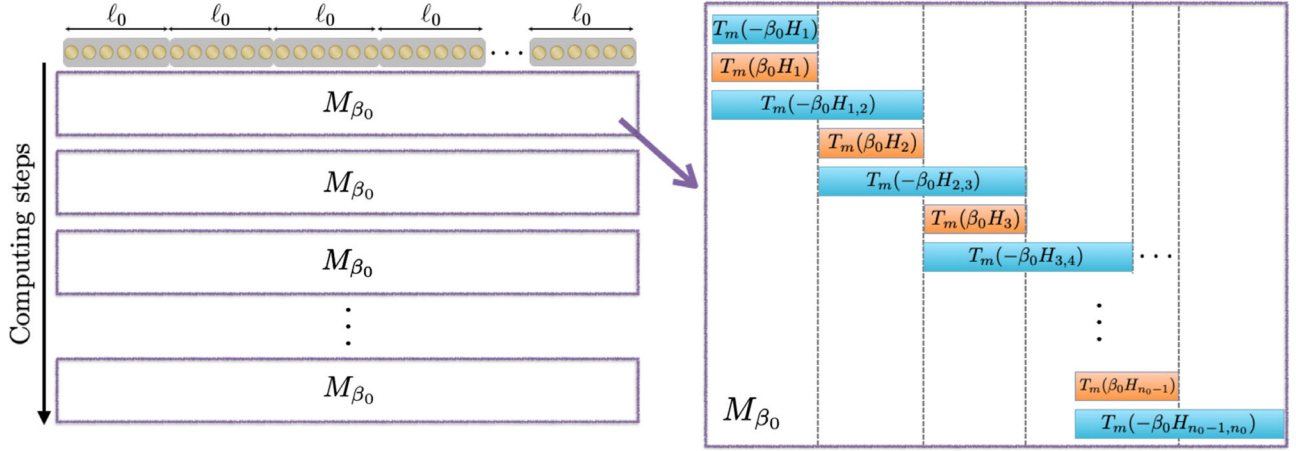


FIG. 4. Our algorithm proceeds by iterated approximations of  $e^{-\beta_0 H}$ , performed  $\beta/\beta_0$  times. In each step, we approximate the Gibbs operator  $e^{-\beta_0 H}$  by the operator  $M_{\beta_0}$ . For this approximation, we establish a decomposition of  $e^{-\beta_0 H}$  as a product of operators shown on the right-hand side. This decomposition uses an imaginary-time version of the Lieb-Robinson bound and the Taylor truncation of the exponential function.

The second approximation is the low-degree polynomial expression of  $\tilde{\Phi}_j$ :

$$\tilde{\Phi}_j^{(m)} = T_m(\beta_0 H_{j-1}) T_m[-\beta_0 (H_{j-1} + H_j)], \quad (30)$$

where  $T_m(x) = \sum_{s=0}^m x^s / m!$  is the truncated Taylor expansion of the order of  $m = \mathcal{O}[\log(n/\epsilon)]$ .

Using the above notation, we can approximate the high-temperature Gibbs state by

$$M_{\beta_0} := \prod_{j=1}^{n_0} \tilde{\Phi}_j^{(m)}. \quad (31)$$

We illustrate this construction in Fig. 4. We notice that our construction resembles the decomposition of the real-time evolution developed in Ref. [110]. Crucially, this approximation is justified using an imaginary-time version of the Lieb-Robinson bound (see Appendix E 1 for the proof).

*Proposition 4.*—For  $\beta \leq 1/16$ , Eq. (31) gives the approximation of the Gibbs state up to an error of

$$\|M_{\beta_0} e^{\beta H} - 1\| \leq \epsilon, \quad (32)$$

where  $M_{\beta_0}$  has the bond dimension of  $e^{\tilde{\mathcal{O}}[\sqrt{\log(n/\epsilon)}]}$ . The sufficient computational time for the construction is given by

$$n \exp\{\tilde{\mathcal{O}}[\sqrt{\log(n/\epsilon)}]\}. \quad (33)$$

We notice that the inequality (32) immediately reduces to

$$\|M_{\beta_0} - e^{-\beta H}\|_p \leq \epsilon \|e^{-\beta H}\|_p \quad (34)$$

for an arbitrary positive  $p$ .

The computational time (33) is qualitatively explained as follows. The operator  $M_{\beta_0}$  is a product of degree- $m$  polynomials  $T_m(x)$ . From Ref. [23], the Schmidt rank of each of  $\{\tilde{\Phi}_j^{(m)}\}_{j=1}^{n_0}$  in Eq. (31) is upper bounded by  $m^{\mathcal{O}(\sqrt{m})} \sim \log(n) \sqrt{\log n}$  along every cut. Because  $\{\tilde{\Phi}_j^{(m)}\}_{j=1}^{n_0}$  are locally defined, for every cut, a constant number of operators in  $\{\tilde{\Phi}_j^{(m)}\}_{j=1}^{n_0}$  contribute to the Schmidt rank. Therefore, the computational time to construct  $M_{\beta_0}$  is at most  $ne^{\tilde{\mathcal{O}}[\sqrt{\log(n/\epsilon)}]}$ .

To extend this result to arbitrary  $\beta$ , we utilize the following upper bound (see Lemma 12 in Appendix B), which slightly extends the analyses in Ref. [95]:

$$\|e^{-2q\beta_0 H} - (M_{\beta_0}^\dagger M_{\beta_0})^q\|_p \leq 3\epsilon_0 q e^{3\epsilon_0 q} \|e^{-\beta H}\|_p \quad (35)$$

for arbitrary positive integers  $q$  and  $p$ , where  $M_{\beta_0}$  satisfies the inequality (34) with  $\epsilon = \epsilon_0$  for arbitrary  $p \in \mathbb{N}$ . We get  $e^{-\beta H} = (e^{-\beta_0 H})^{(\beta/\beta_0)}$  and then multiply the above MPO construction  $\beta/\beta_0$  times, where  $\beta_0$  is appropriately chosen so that  $\beta/\beta_0$  becomes an even integer [i.e.,  $q = \beta/(2\beta_0)$ ]. To make  $3\epsilon_0 q e^{3\epsilon_0 q} \leq \epsilon$  ( $\leq 1$ ), we need to choose  $\epsilon_0 = \epsilon/(6q) = \epsilon\beta_0/(3\beta)$ .

By extending the Schmidt-rank estimation in Ref. [23], we can ensure that the Schmidt rank of  $M_{\beta_0}^{(\beta/\beta_0)}$  is at most as large as  $\exp(Q_\epsilon)$ . In more detail, we can prove the following lemma (see Appendix E 6 for the proof).

*Lemma 5.*—Let  $M_{\beta_0}$  be an approximate operator that has been defined in Eq. (31). Then, for arbitrary  $q \in \mathbb{N}$ , the Schmidt rank of the power of  $M_{\beta_0}$  is upper bounded by



$$\log[\text{SR}(M_{\beta_0}^q)] \leq C' \max(q, \sqrt{mq}) \log(mq) \quad (36)$$

for an arbitrary cut, where  $C'$  is an  $\mathcal{O}(1)$  constant.

Because  $m$  has been chosen as  $m = \mathcal{O}[\log(n/\epsilon)]$ , this upper bound is proportional to  $Q_\epsilon$  for  $q = \mathcal{O}(\beta)$ . Therefore, the quantum Gibbs state  $e^{-\beta H}$  is well approximated by the MPO, with its bond dimensions of  $\exp(Q_\epsilon)$ .

We have already prepared the MPO form of  $M_{\beta_0}$  in Proposition 4. Using the standard results regarding the canonical form of MPOs [111,112], we can efficiently calculate  $M_{\beta_0}^q$  ( $q \lesssim \beta$ ) from  $M_{\beta_0}^{q-1}$  in a computational time of at most  $\text{poly}[\exp(Q_\epsilon)]$ . We notice that, in each of the steps, we can compress the MPO without any truncation error so that the bond dimension of  $M_{\beta_0}^q$  is smaller than the bound in Eq. (36). By recursively constructing  $M_{\beta_0}^q$ , the computation of  $M_{\beta_0}^{(\beta/\beta_0)}$  requires the time steps as many as Eq. (25). We thus prove Theorem 3. ■

Finally, let us compare our method with the imaginary-time-evolving block decimation (TEBD) methods [33,41,48], which proceed by truncation of the Schmidt rank at each imaginary-time Trotter step. A major limitation of these studies is the lack of rigorous justification of the Schmidt-rank truncation, as explained below.

In the TEBD algorithms [48], we start with a matrix product operator  $M_1$  which gives the approximation of  $e^{-\beta_1 H}$  for a certain  $\beta_1$ . We then connect two MPOs as  $M_1^\dagger M_1$ , which is expected to approximate  $e^{-2\beta_1 H}$ . To ensure the precision of approximation, we use Lemma 11 for the MPO  $M_1$ , which necessitates the approximation in terms of general Schatten  $p$ -norm. Now, the main technical difficulty comes from the Schmidt-rank truncation of  $M_1^\dagger M_1$ , which gives MPO  $M_2$  in the next step. After the Schmidt-rank truncation, we connect  $M_2^\dagger M_2$  to approximate the Gibbs state  $e^{-4\beta_1 H}$ . However, to ensure the good approximation from Lemma 11, we have to truncate the Schmidt rank of  $M_1^\dagger M_1$  so that  $M_2$  is close to  $M_1^\dagger M_1$  in terms of the general Schatten  $p$ -norm. The Schmidt-rank truncation based on the singular value decomposition ensures only the approximation in terms of Schatten 2-norm, as in Lemma 1 in Ref. [113]. So far, we have no mathematical tools to perform Schmidt-rank truncation, which guarantees approximation in terms of the general Schatten  $p$ -norm. In summary, even though the quantum Gibbs state can be approximated by an MPO with a small bond dimension, it is highly nontrivial to show whether the truncation of the Schmidt rank retains the good approximation.

We can circumvent this problem by constructing  $e^{-\beta_0 H}$  as a product of local polynomial approximations, which covers the whole chain. Thus, the operator  $M_{\beta_0}$  (or  $M_{\beta_0}^{(\beta/\beta_0)}$ ) has a *finitely bounded Schmidt rank*, and we do not need to approximate it further using the Schmidt-rank truncation.

## V. FURTHER DISCUSSIONS

### A. Rényi entanglement of purification

To characterize the bipartite correlations beyond mutual information, we also consider the Rényi entanglement of purification  $E_{p,\alpha}$  [114], defined as follows.

*Definition 1.*—Let  $\Lambda'$  be a copy of the total system with the Hilbert space  $\mathcal{H}'$ . For an arbitrary quantum state  $\sigma$ , we define  $E_{p,\alpha}(\sigma)$  for the partition  $\Lambda = L \cup R$  as

$$E_{p,\alpha}(\sigma) := \inf_{|\phi\rangle \in \mathcal{H} \otimes \mathcal{H}'} E_\alpha(\phi),$$

$$E_\alpha(\phi) = S_\alpha(\sigma_{L,L'}), \quad (37)$$

where  $|\phi\rangle$  is the purification of  $\sigma$  [i.e.,  $\text{tr}_{\Lambda'}(|\phi\rangle\langle\phi|) = \sigma$ ],  $S_\alpha(\cdot)$  is the Rényi entropy, and  $\sigma_{L,L'} := \text{tr}_{R,R'}(|\phi\rangle\langle\phi|)$ , namely,  $S_\alpha(\sigma_{L,L'}) = (1/\alpha) \log[\text{tr}(\sigma_{L,L'}^\alpha)]$ .

A bound on the Rényi entanglement of purification imposes a stronger restriction to the structure of the quantum state than the mutual information in Eq. (7). For instance, Ref. [115] shows that an upper bound on the entanglement of purification of a 1D system guarantees an efficient approximation by MPOs.

The mutual information  $I(L:R)_{\rho_\beta}$  is related to the Rényi entanglement of purification with  $\alpha = 1$  (see Ref. [116]):

$$I(L:R)_\sigma \leq 2E_{p,1}(\sigma). \quad (38)$$

We also present an upper bound on this quantity as follows (see Sec. VI A for the proof).

*Theorem 6.*—For arbitrary nonzero  $0 < \alpha \leq 1$ , the Rényi entanglement of purification  $E_{p,\alpha}$  is upper bounded as follows:

$$E_{p,\alpha}(\rho) \leq \tilde{C}_0 \max \left[ \beta^{2/3} \log(\beta), \frac{(1-\alpha)\beta}{\alpha} \log\left(\frac{\beta}{\alpha}\right) \right]. \quad (39)$$

The above upper bound implies that, for  $\alpha < 1$ , the entanglement scaling may be linear to  $\beta$  instead of  $\beta^{2/3}$ . This result can be explained as follows. To calculate the Rényi entanglement of purification, we need to obtain the MPO which has an approximation error  $\epsilon$  such that  $D_\epsilon \epsilon^\alpha \lesssim 1$ , where  $D_\epsilon$  is the bond dimension to achieve error  $\epsilon$ . From the MPO with this property, we have  $E_{p,\alpha}(\rho) \lesssim \log(D_\epsilon)$ . Because of  $D_\epsilon \lesssim e^{\sqrt{\beta \log(1/\epsilon)}}$ , the condition  $D_\epsilon \epsilon^\alpha \lesssim 1$  reduces to  $\sqrt{\beta \log(1/\epsilon)} - \alpha \log(1/\epsilon) \lesssim 0$  or  $\log(1/\epsilon) \gtrsim \beta/\alpha^2$ , which gives  $\log(D_\epsilon) \approx \beta/\alpha$ .

Let us compare our result to those of previous studies [64,94,95]. The bond dimension scales as  $D = (1/\epsilon)^{\mathcal{O}(\beta)}$ . By using this estimation, there exists a critical  $\alpha_c$  [ $= 1 - \mathcal{O}(\beta^{-1})$ ] that violates the finite upper bound of  $E_{p,\alpha}(\rho)$  for  $\alpha < \alpha_c$ .

## B. Convex combination of matrix product states

Reference [117] shows that the thermal state can be expressed as a convex combination of MPSs with bond dimension scaling doubly exponentially with  $\beta$ . We can prove the following corollary, which substantially improves their main result (see Sec. VIC for the proof).

*Corollary 7.*—The quantum Gibbs state  $\rho_\beta$  is given by a convex combination of the matrix product states in the following sense:

$$\left\| \rho_\beta - \sum_{i=1}^{\mathcal{D}_\Lambda} p_i |M_i\rangle\langle M_i| \right\|_1 \leq \epsilon, \quad (40)$$

where  $\{|M_i\rangle\}$  are matrix product states with the bond dimension of

$$D = \exp[\tilde{\mathcal{O}}(\tilde{q}_\epsilon^*)], \quad (41)$$

where  $\tilde{q}_\epsilon^*$  has been defined in Theorem 2.

This result can be used to justify the minimally entangled typical thermal states algorithm [33] and the algorithm of Ref. [118] (see also Ref. [117] for more detailed motivations to study the convex combinations of MPSs). Using our bounds, we provide further analytical evidence regarding why these work in practice.

A related quantity in the study of mixed-state entanglement is the *entanglement of formation*. It captures the “average bond dimension” in the convex combination shown in Eq. (40) and can be defined as follows.

*Definition 2.*—Let  $\Lambda'$  be a copy of the total system with the Hilbert space  $\mathcal{H}'$ . For arbitrary quantum state  $\sigma$ , we define  $E_{f,\alpha}(\sigma)$  for the partition  $\Lambda = L \cup R$  as

$$E_{f,\alpha}(\sigma) := \inf \sum_i p_i S_\alpha(\sigma_L^{(i)}), \quad (42)$$

where again  $S_\alpha(\cdot)$  is the Rényi entropy and the minimization is over all pure-state decompositions  $\sigma = \sum_i p_i \sigma^{(i)}$ .

Entanglement of formation is upper bounded by the entanglement of purification [114] (see also Ref. [119]):

$$E_{f,\alpha}(\sigma) \leq E_{p,\alpha}(\sigma), \quad (43)$$

where the equality holds for pure states. When  $\sigma$  is given by the quantum Gibbs state  $\rho_\beta$ , an upper bound follows from Theorem 6:

$$E_{f,\alpha}(\rho_\beta) \leq \tilde{C}_0 \max \left[ \beta^{2/3} \log(\beta), \frac{(1-\alpha)\beta}{\alpha} \log\left(\frac{\beta}{\alpha}\right) \right].$$

## C. Real-time evolution

Our analyses can be partially applied to real-time evolution. In this case, we approximate the unitary time evolution

$e^{-iHt}$  instead of the quantum Gibbs state  $e^{-\beta H}$ . The most essential difference is that the random-walk-like behavior [i.e., Eq. (17)] cannot be justified. Mathematically, the polynomial approximation based on Eq. (17) (see also Lemma 15) is applicable only to imaginary-time evolution. Hence, the MPO approximation of  $e^{-iHt}$  requires the bond dimensions of  $e^{\mathcal{O}(t)}$  instead of  $e^{\mathcal{O}(t^{2/3})}$ . This requirement is expected and consistent with the numerical calculations and the theoretical upper bound [89–92].

Still, our results on the quasilinear time algorithm can be also applied to real-time evolution, where we utilize only the Taylor expansion (30). Let us approximate the unitary time evolution  $e^{-iHt}$  by using a MPO  $M_t$ . For an arbitrary quantum state  $|\psi\rangle$ , we obtain

$$\|(M_t - e^{-iHt})|\psi\rangle\| \leq \|M_t - e^{-iHt}\|_\infty.$$

Recall that the Schatten norm with  $p = \infty$  is equivalent to the operator norm. Hence, by applying Theorem 3 to the case of  $\beta = it$  and  $p = \infty$ , we can obtain the following corollary.

*Corollary 8.*—For arbitrary  $t$ , we can efficiently compute a matrix product operator  $M_t$  that approximates  $e^{-iHt}$  in the sense that

$$\|M_t - e^{-iHt}\| \leq 1/\text{poly}(n), \quad (44)$$

where the bond dimension of  $M_t$  is given by  $\exp\{\tilde{\mathcal{O}}(|t|) + \tilde{\mathcal{O}}[\sqrt{|t| \log(n)}]\}$ . The computational time to calculate  $M_t$  is given by

$$n \exp\{\tilde{\mathcal{O}}(|t|) + \tilde{\mathcal{O}}[\sqrt{|t| \log(n)}]\}. \quad (45)$$

For  $|t| \lesssim \log(n)$ , our result gives a quasilinear computational cost for  $n$ ; thus, it is better than the previous computational cost  $e^{\mathcal{O}(t) + \mathcal{O}[\log(n/\epsilon)]}$ , which is derived from the Lieb-Robinson bound [120,121]. However, for  $|t| \gtrsim \log(n)$ , the computational cost (45) grows exponentially with  $t$  and has the same limitation as the previous methods.

## D. Entanglement rate by imaginary-time evolution

The quantum Gibbs state is regarded as an imaginary-time evolution of the uniformly mixed state, namely,  $\rho_\beta \propto e^{-\beta/2} \rho_{\beta=0} e^{-\beta/2}$ . Thus, the entropy-production rate of the imaginary-time evolution is sublinear with respect to  $\beta$ . Can we extend it to general quantum states instead of the uniformly mixed state? Clearly, when we consider the arbitrary quantum state  $|\psi\rangle$ , the answer is no; that is, the entanglement generation by  $e^{-\beta H}$  for a given cut (e.g.,  $\Lambda = L \sqcup R$ ) is usually unbounded. Even if there are no interactions between  $L$  and  $R$  or  $e^{-\beta H} = e^{-\beta H_L} \otimes e^{-\beta H_R}$ , the entanglement rate can be nonzero if an initial state is arbitrarily chosen. For example, let us consider the initial state  $|\psi\rangle$  as

$$|\psi\rangle = C \sum_i e^{\beta(E_{L,i}+E_{R,i})} |E_{L,i}\rangle \otimes |E_{R,i}\rangle,$$

where  $C$  is a normalization constant and  $|E_{L,i}\rangle$  ( $|E_{R,i}\rangle$ ) is the eigenstate of  $H_L$  ( $H_R$ ) with corresponding eigenvalue  $E_{L,i}$  ( $E_{R,i}$ ). Then, we have

$$\frac{e^{-\beta H_L} \otimes e^{-\beta H_R} |\psi\rangle}{\|e^{-\beta H_L} \otimes e^{-\beta H_R} |\psi\rangle\|} \propto \sum_i |E_{L,i}\rangle \otimes |E_{R,i}\rangle,$$

which is the maximally entangled state. In the above case, the entanglement entropy is significantly increased by the Hamiltonian with no boundary-boundary interactions.

In order to obtain a nontrivial result, we here consider the imaginary-time evolution for a product state  $|P_{L,R}\rangle$  as

$$|P_{L,R}(\beta)\rangle := \frac{e^{-\beta H} |P_{L,R}\rangle}{\|e^{-\beta H} |P_{L,R}\rangle\|}. \quad (46)$$

This setup is feasible in experimental settings [6,82]. When we consider the real-time evolution (i.e.,  $\beta = it$ ), the SIE theorem [91] gives the upper bound for the entanglement rate as  $\mathcal{O}(t)$ . In contrast, no theoretical studies have given an upper bound of the entanglement generation by the imaginary-time evolution. It is an intriguing open problem whether or not the entanglement rate is finitely bounded for large  $\beta$ .

Using our current analyses, we can partially answer this question. To approximate  $|P_{L,R}(\beta)\rangle$ , we use an operator  $O_D$  that satisfies  $\text{SR}(O_D) = D$  for the cut of  $\Lambda = L \sqcup R$  and approximates  $|P_{L,R}(\beta)\rangle$  as  $|P_{L,R}(\beta)\rangle \approx O_D |P_{L,R}\rangle$ . We aim to estimate the approximation error of  $|P_{L,R}(\beta)\rangle$  depending on the Schmidt rank  $D$ . Let us set the ground-state energy of  $H$  equal to zero. Then, from the inequality (49) in Proposition 9 with  $p = \infty$ , there exists  $O_D$  such that

$$\| |P_{L,R}(\beta)\rangle - O_D |P_{L,R}\rangle \| \leq \frac{\epsilon}{\|e^{-\beta H} |P_{L,R}\rangle\|},$$

$$D = e^{\tilde{\mathcal{O}}(\beta^{2/3}) + \tilde{\mathcal{O}}[\sqrt{\beta \log(\beta/\epsilon)}]}. \quad (47)$$

If  $\|e^{-\beta H} |P_{L,R}\rangle\| = \mathcal{O}(1)$ , the entanglement entropy of  $P_{L,R}$  satisfies the same inequality as Eq. (12) and scales as  $\beta^{2/3}$ . However, in general, the quantity  $\|e^{-\beta H} |P_{L,R}\rangle\|$  is exponentially small for  $n$ , and, hence, the value of  $\epsilon$  should be as small as  $e^{\mathcal{O}(n)}$ , which gives the entanglement scaling as  $\sqrt{n\beta}$ . This result is still nontrivial but is rather worse than the expected scaling of  $\beta^{2/3}$ .

To improve the bound, a refined approximation error is required, which is given by the following form of

$$\|O_D e^{\beta H} - 1\| \leq \epsilon \quad (48)$$

instead of the approximation  $\|e^{-\beta H} - O_D\|_p \leq \epsilon \|e^{-\beta H}\|_p$  for the Schatten  $p$ -norm. The approximation of the form of

Eq. (48) can be derived for sufficiently high temperatures (see Proposition 4). If we can extend Proposition 9 in Sec. VI to the form (48), we are able to prove that the entanglement rate by the imaginary-time evolution (46) is upper bounded by  $\tilde{\mathcal{O}}(\beta^{2/3})$ .

Finally, we mention that in various cases the entanglement rate can be smaller than  $\tilde{\mathcal{O}}(\beta^{2/3})$ . In particular, when the ground state is noncritical (or gapped), the imaginary-time evolution for  $e^{-\beta H} |P_{L,R}\rangle$  is expected to rapidly converge to the ground state [41–48]. Indeed, in the case where the Hamiltonian is gapped and defined on a spin chain, there exists a product state which has an  $\mathcal{O}(1)$  overlap with the ground state [22,23,27]. If we choose it as the initial state  $|P_{L,R}\rangle$ , the entanglement entropy for  $e^{-\beta H} |P_{L,R}\rangle / \|e^{-\beta H} |P_{L,R}\rangle\|$  approaches a constant value (i.e., the entanglement entropy of the ground state) exponentially fast with  $\beta$ . Thus, the entanglement rate should be much smaller than  $\mathcal{O}(\beta^{2/3})$ . It is an intriguing question to investigate which class of quantum many-body systems shows a nontrivial entanglement rate for the imaginary-time evolution.

## VI. PROOFS OF THE MAIN THEOREMS

Here we prove Theorems 1 (Theorem 6) and 2 regarding the thermal area law. For simplicity, we focus on one-dimensional systems; however, the essence of the proof is the same in high-dimensional cases (see Appendix D). In Sec. VIC, we also prove Corollary 7, which is based on Theorem 2.

Both Theorems 1 and 2 are based on the following basic approach. We aim to approximate the Gibbs state  $\rho_\beta$  by another operator  $\hat{\rho}_\beta$  which has a smaller Schmidt rank for a given cut  $\Lambda = L \cup R$ . This approximation is formalized in the following proposition, which plays a central role in deriving our main results.

*Proposition 9.*—Let  $\epsilon$  be an arbitrary error such that  $\epsilon \leq e$ . Then, there exists an operator  $\hat{\rho}_\beta$  which approximates  $\rho_\beta$  as follows:

$$\|\rho_\beta - \hat{\rho}_\beta\|_p \leq \epsilon \|\rho_\beta\|_p \quad (49)$$

for arbitrary  $p \in \mathbb{N}$ , and

$$\text{SR}(\hat{\rho}_\beta, i_0) \leq \exp[q_\epsilon^* \log(q_\epsilon^*)] \quad (50)$$

with

$$q_\epsilon^* = C_0 \max\{\beta^{2/3}, [\beta \log(\beta/\epsilon)]^{1/2}\}, \quad (51)$$

where  $C_0$  is a constant of  $\mathcal{O}(1)$ .

The proof is shown in Appendix C. For sufficiently small  $\epsilon$ , this estimation gives a sublinear dependence of the Schmidt rank with respect to  $(1/\epsilon)$ . For example, for  $\epsilon = 1/\text{poly}(n)$ , we have  $\text{SR}(\hat{\rho}_\beta, i_0) \leq n^{\log^{-1/2}(n)}$ , which is

slower than any power of  $n$ . If the Schmidt rank of the approximating operator  $\hat{\rho}_\beta$  exceeds  $e^{\tilde{O}(\beta^{2/3})}$ , the error  $\epsilon$  decays superpolynomially as a function of the Schmidt rank.

### A. Proof of Theorems 1 and 6

Theorems 1 and 6 give upper bounds on the mutual information and the Rényi entanglement of purification, respectively. Because of the inequality (38), Theorem 6 includes Theorem 1, once we consider  $E_{p,1}(\rho)$ . Hence, we need to prove only Theorem 6.

We start from the purification in the form of

$$|\psi\rangle = Z^{-1/2}(e^{-\beta H/2} \otimes \hat{1}) \sum_{j=1}^{D_\Lambda} |j\rangle_\Lambda \otimes |j\rangle_{\Lambda'}, \quad (52)$$

where  $\{|j\rangle\}_{j=1}^{D_\Lambda}$  is an arbitrary orthonormal basis and we denote the partition function  $\text{tr}(e^{-\beta H})$  by  $Z$ . Note that from the above definition  $\text{tr}_{\Lambda'}(|\psi\rangle\langle\psi|) = e^{-\beta H}/Z$ . Then, from the definition of the Rényi entanglement of purification (37), we have

$$E_{p,\alpha}(\rho) \leq E_\alpha(\psi). \quad (53)$$

Next, we estimate an upper bound on  $E_\alpha(\psi)$ .

From Proposition 9, we can find an approximation  $\hat{\rho}_{\beta/4}$  of  $e^{-\beta H/4}$  such that

$$\|e^{-\beta H/4} - \hat{\rho}_{\beta/4}\|_p \leq \epsilon \|e^{-\beta H/4}\|_p \quad (54)$$

for all  $p$ , where the Schmidt rank of  $\hat{\rho}_{\beta/4}$  is upper bounded by Eq. (50). Define  $|\tilde{\psi}\rangle$  as

$$|\tilde{\psi}\rangle := \tilde{Z}^{-1/2} \hat{\rho}_{\beta/4}^\dagger \hat{\rho}_{\beta/4} \otimes \hat{1} \sum_{j=1}^{D_\Lambda} |j\rangle \otimes |j\rangle, \quad (55)$$

where we define  $\tilde{Z} := \text{tr}(\hat{\rho}_{\beta/4}^\dagger \hat{\rho}_{\beta/4} \hat{\rho}_{\beta/4} \hat{\rho}_{\beta/4}^\dagger)$ . Using the inequality (B9) with  $p = 2$ ,  $O = e^{-\beta H/4}$ , and  $\tilde{O} = \hat{\rho}_{\beta/4}$ , we first obtain for  $\tilde{Z}^{1/2} = \|\hat{\rho}_{\beta/4}^\dagger \hat{\rho}_{\beta/4}\|_2$

$$\begin{aligned} \tilde{Z}^{1/2} &\leq \|\hat{\rho}_{\beta/4}^\dagger \hat{\rho}_{\beta/4} - e^{-\beta H/2}\|_2 + \|e^{-\beta H/2}\|_2 \\ &\leq (3\epsilon + 1) \|e^{-\beta H/2}\|_2 = (3\epsilon + 1) Z^{1/2}, \end{aligned} \quad (56)$$

where we use the triangle inequality in the first inequality. We then obtain the fidelity between  $|\psi\rangle$  and  $|\tilde{\psi}\rangle$  as follows:

$$\begin{aligned} \langle \tilde{\psi} | \psi \rangle &= Z^{-1/2} \tilde{Z}^{-1/2} \|\hat{\rho}_{\beta/4} e^{-\beta H/4}\|_2^2 \\ &\geq \frac{Z^{-1}}{3\epsilon + 1} \|\hat{\rho}_{\beta/4} e^{-\beta H/4}\|_2^2, \end{aligned} \quad (57)$$

where we apply the inequality (56) to  $\tilde{Z}$  in the last inequality. From the triangle inequality, we obtain the upper bound of  $\|\hat{\rho}_{\beta/4} e^{-\beta H/4}\|_2$  in the following form:

$$\begin{aligned} \|\hat{\rho}_{\beta/4} e^{-\beta H/4}\|_2 &\geq \|e^{-\beta H/2}\|_2 - \|(\hat{\rho}_{\beta/4} - e^{-\beta H/4}) e^{-\beta H/4}\|_2 \\ &\geq \|e^{-\beta H}\|_1^{1/2} - \|\hat{\rho}_{\beta/4} - e^{-\beta H/4}\|_4 \cdot \|e^{-\beta H/4}\|_4 \\ &\geq \|e^{-\beta H}\|_1^{1/2} - \epsilon \|e^{-\beta H/4}\|_4^2 = (1 - \epsilon) Z^{1/2}, \end{aligned} \quad (58)$$

where we use  $\|e^{-\beta H/2}\|_2 = \|e^{-\beta H}\|_1^{1/2}$  and the Hölder inequality in the second inequality, in the third inequality we use the inequality (54), and the last equation is derived from  $\|e^{-\beta H/4}\|_4 = \|e^{-\beta H}\|_1^{1/4}$ . By applying inequality (58) to Eq. (57), we obtain the inequality of

$$\langle \tilde{\psi} | \psi \rangle \geq \frac{(1 - \epsilon)^2}{3\epsilon + 1} \geq 1 - 5\epsilon, \quad (59)$$

which implies

$$\| |\psi\rangle - |\tilde{\psi}\rangle \|^2 \leq 2 - 2\langle \tilde{\psi} | \psi \rangle \leq 10\epsilon. \quad (60)$$

In the following, using the above upper bound, we estimate the upper bound of Rényi entanglement entropy for arbitrary  $\alpha > 0$ . We consider the cases of  $\alpha = 1$  and  $\alpha < 1$  separately.

#### 1. Case of $\alpha = 1$

We first consider the case of  $\alpha = 1$ . We define  $|\tilde{\psi}_s\rangle$  as an approximation of  $|\psi\rangle$  which satisfies

$$\| |\psi\rangle - |\tilde{\psi}_s\rangle \|^2 \leq 1/s^2, \quad (61)$$

where we use Eq. (55) for the representation of  $|\tilde{\psi}_s\rangle$ . From Theorem 1, the Schmidt rank of  $|\tilde{\psi}_s\rangle$ , say,  $D_s$ , is upper bounded from above by

$$D_s \leq e^{q_s \log(q_s)} \quad (62)$$

with  $q_s = \tilde{C} \max\{\beta^{2/3}, [\beta \log(s)]^{1/2}\}$ . We define  $\bar{s}$  as an integer such that

$$q_s \begin{cases} = \tilde{C} \beta^{2/3} & \text{for } s \leq \bar{s}, \\ = \tilde{C} [\beta \log(s)]^{1/2} & \text{for } s > \bar{s}, \end{cases}$$

where  $\bar{s}$  is in the order of  $\exp[\mathcal{O}(\beta^{1/3})]$ .

Let us denote the Schmidt decomposition of  $|\psi\rangle$  in Eq. (52) as follows:

$$|\psi\rangle = \sum_{m=1}^{D_\psi} \mu_m |\psi_{L,L',m}\rangle \otimes |\psi_{R,R',m}\rangle, \quad (63)$$



where  $|\psi_{L,L',m}\rangle$  and  $|\psi_{R,R',m}\rangle$  are defined on the Hilbert space of  $L \sqcup L'$  and  $R \sqcup R'$ , respectively. From the above representation, we obtain the Rényi entropy with  $\alpha = 1$  as

$$S_1(|\psi\rangle) = -\sum_{m=1}^{\infty} \mu_m^2 \log(\mu_m^2), \quad (64)$$

which is equal to the standard entanglement entropy.

To estimate  $S_1(|\psi\rangle)$ , we utilize the Eckart-Young theorem. By applying the inequality (B5) to  $|\psi\rangle$  and  $|\tilde{\psi}_s\rangle$ , we obtain the following inequality:

$$\sum_{m>D_s} \mu_m^2 \leq \| |\psi\rangle - |\tilde{\psi}_s\rangle \|^2 \leq 1/s^2, \quad (65)$$

where, in the second inequality, we use the condition (61). To upper bound the Rényi entropy, we first define

$$\Gamma_s^2 := \sum_{m=D_s+1}^{D_{s+1}} \mu_m^2, \quad (66)$$

where we define  $D_0 = 0$ . We then obtain

$$\begin{aligned} S_1(|\psi\rangle) &= -\sum_{m=1}^{D_{\bar{s}}} \mu_m^2 \log(\mu_m^2) - \sum_{s=\bar{s}}^{\infty} \sum_{m=D_s+1}^{D_{s+1}} \mu_m^2 \log(\mu_m^2) \\ &\leq \log(D_{\bar{s}}) - \sum_{s=\bar{s}}^{\infty} \sum_{m=D_s+1}^{D_{s+1}} \Gamma_s^2 \log \frac{\Gamma_s^2}{D_{s+1} - D_s}, \end{aligned} \quad (67)$$

where we use the fact that the uniform distribution maximizes  $\sum_{m=D_s+1}^{D_{s+1}} \mu_m^2 \log(\mu_m^2)$ , i.e.,  $\mu_{D_s+1}^2 = \mu_{D_s+2}^2 = \dots = \mu_{D_{s+1}}^2 = \Gamma_s^2 / (D_{s+1} - D_s)$ . Because of the inequalities (62) and (65), we have  $\Gamma_s^2 \leq 1/s^2$ :

$$\begin{aligned} S_1(|\psi\rangle) &\leq \tilde{C}\beta^{2/3} \log(\tilde{C}\beta^{2/3}) + \sum_{s=\bar{s}}^{\infty} (1/s)^2 \log(3s^2) \\ &\quad + \sum_{s=\bar{s}}^{\infty} \frac{\tilde{C}[\beta \log(s)]^{1/2} \log\{\tilde{C}[\beta \log(s)]^{1/2}\}}{s^2}, \end{aligned}$$

where we apply the inequality  $-x \log x \leq -x \log(x/3) \leq -y \log(y/3)$  for  $0 < x \leq y \leq 1$  to  $-\Gamma_s^2 \log(\Gamma_s^2)$ . Using  $\bar{s} = \exp[\mathcal{O}(\beta^{1/3})]$ , the second and the third terms become less dominant in comparison with the first term when  $\beta$  is large. We thus obtain the main inequality (39) in the theorem for  $\alpha = 1$ .

## 2. Case of $\alpha < 1$

We follow the same analyses as in the case of  $\alpha = 1$ . In this case, we define  $|\tilde{\psi}_s'\rangle$  as an approximation of  $|\psi\rangle$  which satisfies

$$\| |\psi\rangle - |\tilde{\psi}_s'\rangle \|^2 \leq s^{-2/\alpha}, \quad (68)$$

where the Schmidt rank of  $|\tilde{\psi}_s'\rangle$ , say,  $D'_s$ , is upper bounded from above by

$$D'_s \leq e^{q'_s \log(q'_s)} \quad (69)$$

with  $q'_s = \tilde{C} \max\{\beta^{2/3}, [\alpha^{-1}\beta \log(s)]^{1/2}\}$ . We define  $\bar{s}'$  as an integer such that

$$q'_s \begin{cases} = \tilde{C}\beta^{2/3} & \text{for } s \leq \bar{s}', \\ = \tilde{C}[\alpha^{-1}\beta \log(s)]^{1/2} & \text{for } s > \bar{s}', \end{cases}$$

where we have  $\bar{s}' = \exp[\mathcal{O}(\alpha\beta^{1/3})]$ .

Using the Schmidt decomposition as in Eq. (63), the  $\alpha$ -Rényi entropy is given by

$$S_\alpha(|\psi\rangle) = \frac{1}{1-\alpha} \log \left( \sum_{s=0}^{\infty} \sum_{m=D'_s+1}^{D'_{s+1}} \mu_m^{2\alpha} \right). \quad (70)$$

For  $\alpha < 1$ , we obtain the upper bound of

$$\sum_{m=D'_s+1}^{D'_{s+1}} \mu_m^{2\alpha} \leq (D'_{s+1} - D'_s) \left( \frac{\Gamma_s'^2}{D'_{s+1} - D'_s} \right)^\alpha \leq \frac{D'^{1-\alpha}_{s+1}}{s^2},$$

where we adopt the similar notation (66) for  $\Gamma_s'$ , and, to derive  $\Gamma_s'^2 \leq s^{-2/\alpha}$ , we use the condition (68) and the Eckart-Young theorem as in Eq. (65). Therefore, we have the following upper bound for the summation:

$$\begin{aligned} &\sum_{s=0}^{\infty} \sum_{D_s < m \leq D_{s+1}} \mu_m^{2\alpha} \\ &\leq D_{\bar{s}}^{1-\alpha} + \sum_{s \geq \bar{s}} \frac{D_{s+1}^{1-\alpha}}{s^2} \\ &\leq e^{(1-\alpha)\tilde{C}\beta^{2/3} \log(\tilde{C}\beta^{2/3})} \\ &\quad + \sum_{s \geq \bar{s}} \frac{\exp\{(1-\alpha)\tilde{c}_\alpha \log^{1/2}(s) \log[\tilde{c}_\alpha \log^{1/2}(s)]\}}{s^2}, \end{aligned}$$

where we define  $\tilde{c}_\alpha := \tilde{C} \sqrt{\beta/\alpha}$ . For the estimation of the summation for  $\sum_{s \geq \bar{s}}$ , we also use the inequality of

$$\begin{aligned} &\int_1^\infty \frac{\exp\{(1-\alpha)\tilde{c}_\alpha \log^{1/2}(x) \log[\tilde{c}_\alpha \log^{1/2}(x)]\}}{x^2} dx \\ &= \int_0^\infty 2te^{-t^2 + (1-\alpha)\tilde{c}_\alpha t \log(\tilde{c}_\alpha t)} dt \leq e^{\tilde{C}_1(1-\alpha)^2 \tilde{c}_\alpha^2 \log^2(\tilde{c}_\alpha)}, \end{aligned}$$

where  $\tilde{C}_1$  is a constant of  $\mathcal{O}(1)$ . By combining the above inequalities together, we obtain

$$\begin{aligned} & \frac{1}{1-\alpha} \log \left( \sum_{s=0}^{\infty} \sum_{D_s < m \leq D_{s+1}} \mu_m^{2\alpha} \right) \\ & \leq \tilde{C}_0 \max [\beta^{2/3} \log(\beta), (1-\alpha)(\beta/\alpha) \log(\beta/\alpha)]. \end{aligned}$$

This result gives the main inequality (39) in the theorem for  $\alpha < 1$ . This completes the proof. ■

## B. Proof of Theorem 2

Here, we prove Theorem 2, which gives the MPO approximation of the quantum Gibbs state.

We first prove the case of  $p = 2$ . Let  $\hat{\rho}_\beta$  be an approximation of  $\rho_\beta$  such that for a given cut  $\Lambda = L \sqcup R$ . We define the Schmidt rank of  $\hat{\rho}_\beta$  as  $D_{\epsilon_0}$ , which satisfies the inequality (50) with the approximation error  $\epsilon_0$ , namely,

$$\|\rho_\beta - \hat{\rho}_\beta\|_p \leq \epsilon_0 \|\rho_\beta\|_p \quad (71)$$

with

$$\text{SR}(\hat{\rho}_\beta, i_0) := D_{\epsilon_0} \leq \exp[q_\epsilon^* \log(q_\epsilon^*)], \quad (72)$$

where  $q_\epsilon^*$  is defined in Eq. (51). For the cut, we define the Schmidt decomposition of  $\rho_\beta$  as follows:

$$\rho_\beta = \sum_m \mu_m \Phi_{L,m} \otimes \Phi_{R,m} \quad (\mu_m > 0), \quad (73)$$

where  $\{\Phi_{L,m}\}$  ( $\{\Phi_{R,m}\}$ ) are orthonormal operator bases which satisfy

$$\|\Phi_{L,m}\|_2 = 1, \quad \text{tr}(\Phi_{L,m} \Phi_{L,m'}) = 0 \quad (74)$$

for  $m \neq m'$ . Note that, from the above definition, we have

$$\|\rho_\beta\|_2^2 = \sum_m \mu_m^2. \quad (75)$$

By applying the Eckart-Young theorem (B7) to  $\rho_\beta$  and  $\hat{\rho}_\beta$ , we obtain

$$\sum_{m > D_{\epsilon_0}} \mu_m^2 \leq \|\rho_\beta - \hat{\rho}_\beta\|_2^2 \leq \epsilon_0^2 \|\rho_\beta\|_2^2, \quad (76)$$

where we use the inequality (71) with  $p = 2$ . Then, from Lemma 1 in Ref. [113], there exists an MPO  $M_{D_{\epsilon_0}}$  such that

$$\|\rho_\beta - M_{D_{\epsilon_0}}\|_2^2 \leq 2\epsilon_0^2 n \|\rho_\beta\|_2^2. \quad (77)$$

Therefore, by choosing  $\epsilon_0 = [\epsilon/(2n)]^{1/2}$ , we obtain the desired approximation error (14), and the bond dimension  $D_{\epsilon_0}$  satisfies the inequality (15).

Second, we prove the case of  $p = 1$ . For this proof, we consider the purification of the quantum Gibbs state  $\rho_{\beta/2}$  as in Eq. (52), which is denoted by  $|\psi\rangle$ :

$$\begin{aligned} |\psi\rangle &= Z^{-1/2} (e^{-\beta H/2} \otimes \hat{1}) \sum_{j=1}^{D_\Lambda} |j\rangle_\Lambda \otimes |j\rangle_{\Lambda'} \\ &= \sum_{m=1}^{D_\Psi} \nu_m |\psi_{L,L',m}\rangle \otimes |\psi_{R,R',m}\rangle, \end{aligned} \quad (78)$$

where, in the second equation, we use an expression of the Schmidt decomposition similar to that of Eq. (63). If we can obtain an MPS  $|M_{\tilde{D}_\epsilon}\rangle$  such that

$$\| |\psi\rangle - |M_{\tilde{D}_\epsilon}\rangle \| \leq \epsilon, \quad (79)$$

we obtain

$$\| \text{tr}_{\Lambda'} (|\psi\rangle\langle\psi| - |M_{\tilde{D}_\epsilon}\rangle\langle M_{\tilde{D}_\epsilon}|) \|_1 = \|\rho_\beta - M_{\tilde{D}_\epsilon}\|_1 \leq \epsilon, \quad (80)$$

where we define  $M_{\tilde{D}_\epsilon} := \text{tr}_{\Lambda'} (|M_{\tilde{D}_\epsilon}\rangle\langle M_{\tilde{D}_\epsilon}|)$ . Note that  $|M_{\tilde{D}_\epsilon}\rangle\langle M_{\tilde{D}_\epsilon}|$  is given by a MPO with the bond dimension of  $\tilde{D}_\epsilon^2$ .

Our task is now to find an MPS  $|M_{\tilde{D}_\epsilon}\rangle$  which satisfies Eq. (79). For this purpose, we consider the purification of  $\hat{\rho}_{\beta/4}^\dagger \hat{\rho}_{\beta/4}$  as in Eq. (55), which we denote by  $|\tilde{\psi}\rangle$ . Here,  $\hat{\rho}_{\beta/4}$  gives the approximation of  $\rho_{\beta/4}$  as

$$\|\rho_{\beta/4} - \hat{\rho}_{\beta/4}\|_2 \leq \epsilon_1 \|\rho_{\beta/4}\|_2 \quad (81)$$

with  $\text{SR}(\hat{\rho}_{\beta/4}, i_0) = D_{\epsilon_1} \leq \exp[q_{\epsilon_1}^* \log(q_{\epsilon_1}^*)]$  for a given cut  $\Lambda = L \sqcup R$ . The Schmidt rank of  $|\tilde{\psi}\rangle$  along the cut is upper bounded by  $D_{\epsilon_1}^2$ . In contrast, from the inequality (60), we obtain

$$\| |\psi\rangle - |\tilde{\psi}\rangle \| \leq 10\epsilon_1, \quad (82)$$

and, hence, the Eckart-Young theorem gives the same inequality as Eq. (76):

$$\sum_{m > D_{\epsilon_1}^2} \nu_m^2 \leq \| |\psi\rangle - |\tilde{\psi}\rangle \|^2 \leq 10\epsilon_1. \quad (83)$$

Thus, from Lemma 1 in Ref. [113], there exists an MPS  $|M_{D_{\epsilon_1}^2}\rangle$  such that

$$\| |\psi\rangle - |M_{D_{\epsilon_1}^2}\rangle \| \leq \sqrt{20n\epsilon_1}. \quad (84)$$

To obtain the approximation error  $\epsilon$ , we need to choose  $\epsilon_1 = \epsilon/(20n)$ . Therefore, if we choose  $D = D_{\epsilon^2/(400n^2)}$ , there exists an MPO  $M_D$  that satisfies the inequality (14) with  $p = 1$ . Note that the bond dimension  $D_{\epsilon^2/(400n^2)}$  satisfies the inequality (15) by choosing  $C'_0$  appropriately. This completes the proof of Theorem 2. ■

### C. Proof of Corollary 7

Here, we prove that the quantum Gibbs state is well approximated by a convex combination of matrix product states as in Eq. (40):

$$\rho_\beta \approx \sum_{i=1}^{D_\Lambda} p_i |M_i\rangle\langle M_i|. \quad (85)$$

We then show that the approximation error  $\epsilon$  is achieved by taking the bond dimension as in Eq. (41).

The proof is based on Theorem 2. We first consider the MPO approximation of  $e^{-\beta H/2}$  as follows:

$$\|e^{-\beta H/2} - M_{\beta/2}\|_2 \leq \frac{\epsilon}{6} \|e^{-\beta H/2}\|_2, \quad (86)$$

where the bond dimension of  $M_{\beta/2}$  is given by Eq. (15) [or Eq. (41)]. By using Lemma 11 with  $p = 1$ , we get

$$\|e^{-\beta H} - M_{\beta/2} M_{\beta/2}^\dagger\|_1 \leq \frac{\epsilon}{2} \|e^{-\beta H}\|_1. \quad (87)$$

By inserting  $\hat{1} = \sum_{i=1}^{D_\Lambda} |P_i\rangle\langle P_i|$  with  $\{P_i\}_{i=1}^{D_\Lambda}$  the product-state basis, we obtain

$$\left\| \frac{e^{-\beta H}}{\text{tr}(e^{-\beta H})} - \sum_{i=1}^{D_\Lambda} \frac{M_{\beta/2} |P_i\rangle\langle P_i| M_{\beta/2}^\dagger}{\text{tr}(e^{-\beta H})} \right\|_1 \leq \frac{\epsilon}{2}, \quad (88)$$

where we use  $\|e^{-\beta H}\|_1 = \text{tr}(e^{-\beta H})$ .

We now define

$$|M_i\rangle := \frac{M_{\beta/2} |P_i\rangle}{\|M_{\beta/2} |P_i\rangle\|}, \quad p_i := \frac{\|M_{\beta/2} |P_i\rangle\|^2}{\|M_{\beta/2}\|_2^2},$$

$$\sigma_\beta := \sum_{i=1}^{D_\Lambda} p_i |M_i\rangle\langle M_i|, \quad (89)$$

where  $\sigma_\beta$  is the normalized quantum state and satisfies  $\|\sigma_\beta\|_1 = 1$  because of  $\sum_i \|M_{\beta/2} |P_i\rangle\|^2 = \text{tr}(M_{\beta/2} M_{\beta/2}^\dagger) = \|M_{\beta/2}\|_2^2$ . The MPO  $M_{\beta/2}$  has the bond dimension of Eq. (41), and, hence, the quantum state  $M_{\beta/2} |P_i\rangle$  is also given by a matrix product state with Eq. (41). We obtain the norm difference between  $\rho_\beta$  and  $\sigma_\beta$  as

$$\|\rho_\beta - \sigma_\beta\|_1 \leq \left\| \rho_\beta - \frac{\|M_{\beta/2}\|_2^2}{\text{tr}(e^{-\beta H})} \sigma_\beta \right\|_1 + \left\| \sigma_\beta - \frac{\|M_{\beta/2}\|_2^2}{\text{tr}(e^{-\beta H})} \sigma_\beta \right\|_1$$

$$\leq \frac{\epsilon}{2} + \left| 1 - \frac{\|M_{\beta/2}\|_2^2}{\text{tr}(e^{-\beta H})} \right| \cdot \|\sigma_\beta\|_1 \leq \epsilon, \quad (90)$$

where we use Eq. (88) for the first term and for the second term we use  $\|\sigma_\beta\|_1 = 1$  and

$$|\text{tr}(e^{-\beta H}) - \|M_{\beta/2}\|_2^2| = |\text{tr}(e^{-\beta H} - M_{\beta/2} M_{\beta/2}^\dagger)|$$

$$\leq \|e^{-\beta H} - M_{\beta/2} M_{\beta/2}^\dagger\|_1$$

$$\leq \frac{\epsilon}{2} \text{tr}(e^{-\beta H}). \quad (91)$$

We thus prove the inequality (40). This completes the proof. ■

## VII. CONCLUSION

We have shown two main results in this work. The first one is the improved thermal area law that gives a scaling of  $\tilde{\mathcal{O}}(\beta^{2/3})$  over all lattices (Theorem 1). This scaling behavior is qualitatively explained by the fact that the imaginary-time evolution is intrinsically related to the random walk as in Eq. (17). In the 1D case, we also give an MPO representation of the quantum Gibbs state with a sublinear bond dimension with respect to the system size  $n$  (Theorem 2). The second one is a quasilinear time algorithm for preparing an MPO approximation to the 1D thermal state (Theorem 3), which improves upon all the prior rigorous constructions. It also justifies the quasilinear runtime of several heuristic algorithms inspired by the MPO-based techniques. Moreover, our algorithm can be applied to the computation of the ground state under the low-energy-density assumption of Eq. (26). Our first technical insight is the use of polynomial approximations of the exponential function, which are based on Taylor truncation and Chebyshev expansion (17). The second technical contribution is a Trotter-Suzuki-type decomposition of the Gibbs state (see Fig. 4). It would be interesting to see the possibility to further develop our approximation by using the results in Ref. [122].

We leave the following questions to be considered in future work.

- (i) *High-dimensional PEPO representation with sub-linear bond dimension.*—Our analytical approach has improved the bond dimension of the MPO for 1D quantum Gibbs states. Here, the point is to utilize the estimation in Ref. [23] to efficiently encode the polynomial of the Hamiltonian to the MPO representation. We expect that the same improvement should be possible in the PEPO approximation for the high-dimensional Gibbs state. Even though the PEPO representation of the quantum Gibbs state does not imply an efficient simulation by itself [74,75], it is of great importance in the implementation of numerical algorithms employing the PEPO ansatz. The key question is how to encode the polynomial of the Hamiltonian to a PEPO representation with a nontrivial bond dimension. Such a representation will also be useful in the context of area laws for ground states in higher dimensions.
- (ii) *Improving the run time of the algorithm.*—Our algorithm presented in Theorem 3 has a run time

of  $ne^{\tilde{O}(\beta)+\tilde{O}[\sqrt{\beta\log(n)}]}$ . We expect that this time could be improved to  $ne^{\tilde{O}(\beta^{2/3})+\tilde{O}[\sqrt{\beta\log(n)}]}$ , because this result matches the bond dimension of the MPO constructed in Theorem 2. Another challenge is to improve the run time to the subexponential form with respect to  $\log(n)$  for  $\beta = \mathcal{O}[\log(n)]$ . This improvement would lead to quasilinear time algorithms for ground states under the assumption (26). The main difficulty lies in constructing a better polynomial approximation to the quantum Gibbs state than  $M_{\beta_0}^{(\beta/\beta_0)}$  in Eq. (31).

- (iii) *Stronger norm inequality for imaginary-time evolution.*—As discussed in Sec. V D, we observed that an approximation of the form  $\|O_D e^{\beta H} - 1\| \leq \epsilon$  instead of the current one  $\|e^{-\beta H} - O_D\|_p \leq \epsilon \|e^{-\beta H}\|_p$  would lead to an imaginary-time version of the SIE theorem.
- (iv) *Circuit complexity of preparing 1D quantum Gibbs state.*—As discussed after Theorem 2, we believe that our MPO approximation could be used to construct a quantum circuit for preparing the quantum Gibbs state. So far, the best estimation requires  $n^{\mathcal{O}(\beta)}$  to prepare the 1D quantum Gibbs states on the quantum computer [52]. The quantum preparation of the quantum Gibbs state is expected to be easier than the MPO construction on the classical computer. Hence, we conjecture that the sufficient number of the elementary quantum gates should be also quasilinear as in Eq. (25).  
For instance, the adiabatic algorithm presented in Ref. [55] could be used in this context, by establishing the injectivity of the MPO in Eq. (14). As another route, we may be able to employ the techniques in Appendix B in Ref. [17], which implemented the smooth function of a Hamiltonian (see also Sec. V.3 in Ref. [59] for further discussions). By using this method, which relies on polynomial approximations to  $e^{-\beta H}$ , the polynomial presented in Theorem 3 could be efficiently implemented on a quantum computer.
- (v) *Improving the thermal area law to  $\beta^{1/2}|\partial L|$ .*—In this work, we identified the critical  $\gamma_c$  satisfying Eq. (2) as  $1/5 \leq \gamma_c \leq 2/3$ . From the random walk behavior in Sec. III A, we expect that  $\gamma_c$  may be equal to  $1/2$  or even smaller, which would suggest the diffusive propagation of information by the imaginary-time evolution. For the characterization of entanglement structures of quantum many-body systems at finite temperatures, identification of the optimal  $\gamma$  is one of the most fundamental future problems.

## ACKNOWLEDGMENTS

A. A. thanks David Gosset for introducing the excellent survey [93] on polynomial approximations. The authors

also thank John Preskill for the suggestion that an improvement to the thermal area law on graph networks might not hold. The work of T. K. is supported by the RIKEN Center for AIP and JSPS KAKENHI Grant No. 18K13475. Part of the work was done when T. K. was visiting the Perimeter Institute. The work was done when A. A. was affiliated with the Institute for Quantum Computing and the Department of Combinatorics and Optimization, University of Waterloo, and A. M. A. and A. A. were with the Perimeter Institute for Theoretical Physics. A. A. was supported by the Canadian Institute for Advanced Research, through funding provided to the Institute for Quantum Computing by the Government of Canada and the Province of Ontario. This research was supported in part by the Perimeter Institute for Theoretical Physics. Research at the Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Colleges and Universities.

## APPENDIX A: MORE DETAILED SETUP

We here recall the setup. We consider a quantum spin system with  $n$  spins, where each of the spins sits on a vertex of the  $d$ -dimensional graph (or  $d$ -dimensional lattice) with  $\Lambda$  the total spin set, namely,  $|\Lambda| = n$ . We assume that a finite-dimensional Hilbert space ( $\zeta$  dimension) is assigned to each of the spins. For a partial set  $X \subseteq \Lambda$ , we denote the cardinality, that is, the number of vertices contained in  $X$ , by  $|X|$  (e.g.,  $X = \{i_1, i_2, \dots, i_{|X|}\}$ ). We also denote the complementary subset of  $X$  by  $X^c := \Lambda \setminus X$ . We denote the Hilbert space of a subset  $X \subseteq \Lambda$  and its dimension by  $\mathcal{H}_X$  and  $\mathcal{D}_X$ , respectively.

For arbitrary subsets  $X, Y \subseteq \Lambda$ , we define  $d_{X,Y}$  as the shortest path length on the graph that connects  $X$  and  $Y$ ; that is, if  $X \cap Y \neq \emptyset$ ,  $d_{X,Y} = 0$ . When  $X$  is composed of only one element (i.e.,  $X = \{i\}$ ), we denote  $d_{\{i\},Y}$  by  $d_{i,Y}$  for the simplicity. We also define  $\text{diam}(X)$  as follows:

$$\text{diam}(X) := 1 + \max_{i,j \in X} (d_{i,j}). \quad (\text{A1})$$

### 1. One-dimensional $k$ -local Hamiltonian

Let us now define one-dimensional systems, where the Hamiltonian  $H$  is given by the general  $k$ -local operator:

$$H = \sum_{X \subset \Lambda, \text{diam}(X) \leq k} h_X, \quad \max_{i \in \Lambda} \sum_{X: X \ni i} \|h_X\| \leq g, \quad (\text{A2})$$

where  $h_X$  are the interaction terms acting on the subset  $X$ . Here,  $\sum_{X: X \ni i}$  means the summation which picks up all the subsets  $X \subset \Lambda$  such that  $X \ni i$ . In the main text, we consider the Hamiltonian in the form of Eq. (5). By choosing  $k = 2$  and  $g = 1$ , the Hamiltonian (A2) reduces to the form of Eq. (5).



We here define  $\Lambda_{\leq i}$  ( $\Lambda_{> i}$ ) for an arbitrary  $i \in \Lambda$  as the subset  $\{j\}_{j \leq i}$  ( $\{j\}_{j > i}$ ). We denote  $v_i$  by the interaction between  $\Lambda_{\leq i}$  and  $\Lambda_{> i}$ :

$$v_i = \sum_{X: X \cap \Lambda_{\leq i} \neq \emptyset, X \cap \Lambda_{> i} \neq \emptyset} h_X. \quad (\text{A3})$$

For the Hamiltonian (5) in the main text,  $v_i$  is simply given by  $h_{i,i+1}$ . We then define the Schmidt rank  $\text{SR}(v_i, i)$  as  $D_{\text{loc}}$ :

$$\text{SR}(v_i, i) \leq D_{\text{loc}}, \quad (\text{A4})$$

where  $D_{\text{loc}}$  is at most of  $\zeta^{\mathcal{O}(k)}$ .

## 2. High-dimensional $k$ -local Hamiltonian

In considering  $d$ -dimensional systems, we also consider the  $k$ -local operator:

$$H = \sum_{\substack{X \subset \Lambda, |X| \leq k \\ \text{diam}(X) \leq k}} h_X, \quad \max_{i \in \Lambda} \sum_{X: X \ni i} \|h_X\| \leq g. \quad (\text{A5})$$

We slice the total system  $\Lambda$  into  $l_\Lambda$  pieces:

$$\begin{aligned} \Lambda &= \Lambda_1 \sqcup \Lambda_2 \sqcup \cdots \sqcup \Lambda_{l_\Lambda}, \\ |\Lambda_j| &\leq |\partial\Lambda| = \mathcal{O}(n^{d-1/d}), \end{aligned} \quad (\text{A6})$$

where  $l_\Lambda$  is the system length, namely,  $l_\Lambda = \mathcal{O}(n^{1/d})$ , and we define  $|\partial\Lambda|$  as an integer which gives the upper bounds for  $|\Lambda_j|$ .

Similar to the one-dimensional case, we define  $\Lambda_{\leq i}$  ( $\Lambda_{> i}$ ) for an arbitrary  $i \in \Lambda$  as the subset  $\bigsqcup_{j \leq i} \Lambda_j$  ( $\bigsqcup_{j > i} \Lambda_j$ ). We then

define the Schmidt rank  $\text{SR}(O, i)$  in the same way as Eq. (6). We also define  $v_i$  as the interaction between  $\Lambda_{\leq i}$  and  $\Lambda_{> i}$ :

$$v_i = \sum_{X: X \cap \Lambda_{\leq i} \neq \emptyset, X \cap \Lambda_{> i} \neq \emptyset} h_X. \quad (\text{A7})$$

Here, each of the  $\{v_i\}_{i=1}^{l_\Lambda}$  consists of at most of  $\mathcal{O}(|\partial\Lambda|)$  local interaction terms  $h_X$ . We define  $D_{\text{loc}}$  as the upper bound for the Schmidt ranks of  $\{v_i\}$ :

$$\text{SR}(v_i, i) \leq D_{\text{loc}} = \zeta^{\mathcal{O}(k)} |\partial\Lambda|. \quad (\text{A8})$$

## APPENDIX B: BASIC ANALYTICAL TOOLS

### 1. Generalized Hölder inequality for Schatten norm

For a general Schatten  $p$ -norm, we can prove the following generalized Hölder inequality (see Proposition 2.5 in Ref. [123]):

$$\left\| \prod_{j=1}^s O_j \right\|_p \leq \prod_{j=1}^s \|O_j\|_{p_j}, \quad (\text{B1})$$

where  $\sum_{j=1}^s 1/p_j = 1/p$ . From the inequality, we can immediately obtain

$$\|O_1 O_2\|_p \leq \|O_1\|_p \|O_2\|, \quad (\text{B2})$$

where we set  $p_1 = p$  and  $p_2 = \infty$  in Eq. (B1).

### 2. The Eckart-Young theorem

We here show the Eckart-Young theorem [124] without the proof.

*Lemma 10 (the Eckart-Young theorem).*—Let us consider a normalized state  $|\psi\rangle$  and give its Schmidt decomposition as

$$|\psi\rangle = \sum_{m=1}^{D_\psi} \mu_m |\psi_{1,m}\rangle \otimes |\psi_{2,m}\rangle, \quad (\text{B3})$$

where  $\mu_1 \geq \mu_2 \geq \mu_3 \cdots \geq \mu_{D_\psi}$  and  $\{|\psi_{1,m}\rangle\}_{m=1}^{D_\psi}$  and  $\{|\psi_{2,m}\rangle\}_{m=1}^{D_\psi}$  are orthonormal states, respectively. We then consider another quantum state  $|\hat{\psi}\rangle$  with its Schmidt rank  $D$  and define the overlap with the state  $|\psi\rangle$  as  $\| |\psi\rangle - |\hat{\psi}\rangle \|$ . Then, for the Schmidt-rank truncation as

$$|\psi_D\rangle = \sum_{m \leq D} \mu_m |\psi_{1,m}\rangle \otimes |\psi_{2,m}\rangle, \quad (\text{B4})$$

the Eckart-Young theorem gives the following inequality:

$$\| |\psi\rangle - |\psi_D\rangle \|^2 = \sum_{m > D} \mu_m^2 \leq \| |\psi\rangle - |\hat{\psi}\rangle \|^2, \quad (\text{B5})$$

where  $|\hat{\psi}\rangle$  can be unnormalized.

We note that the Eckart-Young theorem can be also applied to operator by regarding it as the vector with  $D_\Lambda^2$  elements. For an operator  $O$ , we can obtain the Schmidt decomposition as

$$O = \sum_{m=1}^{D_O} \mu_m O_{1,m} \otimes O_{2,m}, \quad (\text{B6})$$

where  $\{O_{1,m}\}$  and  $\{O_{2,m}\}$  are operator bases with the property of  $\|O_{1,m}\|_2 = 1$  and  $\text{tr}(O_{1,m} O_{1,m'}) = \delta_{m,m'}$  for  $m \neq m'$ . For an arbitrary operator  $\hat{O}$  with its Schmidt rank  $D$ , we obtain

$$\|O - O_D\|_2^2 = \sum_{m > D} \mu_m^2 \leq \|O - \hat{O}\|_2^2, \quad (\text{B7})$$

where we define  $O_D := \sum_{m \leq D} \mu_m O_{1,m} \otimes O_{2,m}$ . We note that in applying the operator the Eckart-Young theorem is applied only to the Schatten 2-norm. As far as we know, the Eckart-Young theorem cannot be extended to the general Schatten  $p$ -norm.

### 3. Approximation of square operators

In the analyses, we often use the following lemma, which connects the closeness between two operators to that between square of the two operators.

*Lemma 11.*—Let  $O$  and  $\tilde{O}$  be operators which are close to each other in the following sense:

$$\|O - \tilde{O}\|_{2p} \leq \delta \|O\|_{2p} \quad (\delta \leq 1). \quad (\text{B8})$$

Then, the square of the operator  $O$ , which is  $O^\dagger O$ , is close to  $\tilde{O}^\dagger \tilde{O}$  as follows:

$$\|O^\dagger O - \tilde{O}^\dagger \tilde{O}\|_p \leq 3\delta \|O^\dagger O\|_p. \quad (\text{B9})$$

The proof is straightforward by extending the result in Ref. [95], where the positivity of  $O$  is assumed. We show the proof in the following.

#### a. Proof of Lemma 11

Following Ref. [95], we start from

$$\begin{aligned} \|O^\dagger O - \tilde{O}^\dagger \tilde{O}\|_p &= \|O^\dagger(O - \tilde{O}) - (\tilde{O}^\dagger - O^\dagger)\tilde{O}\|_p \\ &\leq \|O^\dagger(O - \tilde{O})\|_p + \|(\tilde{O}^\dagger - O^\dagger)\tilde{O}\|_p, \end{aligned}$$

where the inequality is derived from the triangle inequality. By using the Hölder inequality (B1) with  $p_1 = p_2 = 2p$ , we obtain

$$\|O^\dagger(O - \tilde{O})\|_p \leq \|O^\dagger\|_{2p} \|O - \tilde{O}\|_{2p} \leq \delta \|O\|_{2p}^2,$$

where we use the inequality (B8) and  $\|O^\dagger\|_{2p} = \|O\|_{2p}$ . In the same way, we obtain

$$\|(\tilde{O}^\dagger - O^\dagger)\tilde{O}\|_p \leq \delta \|O\|_{2p} \|\tilde{O}\|_{2p} \leq \delta \|O\|_{2p}^2 (1 + \delta),$$

where the last inequality is derived from  $\|\tilde{O}\|_{2p} = \|\tilde{O} - O + O\|_{2p} \leq \|\tilde{O} - O\|_{2p} + \|O\|_{2p} \leq \|O\|_{2p} (1 + \delta)$ . The definition of the Schatten norm (10) implies

$$\begin{aligned} \|O\|_{2p}^2 &:= [\text{tr}(O^\dagger O)^p]^{1/p} = \{\text{tr}[(O^\dagger O)(O^\dagger O)^\dagger]^{p/2}\}^{1/p} \\ &= \|O^\dagger O\|_p, \end{aligned}$$

where we use Hermiticity of  $O^\dagger O$ . By combining all the above inequalities, we arrive at the inequality of

$$\|O^\dagger O - \tilde{O}^\dagger \tilde{O}\|_p \leq \delta(2 + \delta) \|O^\dagger O\|_p \leq 3\delta \|O^\dagger O\|_p,$$

where we use the condition  $\delta \leq 1$  in the last inequality. This completes the proof of the inequality (B9). ■

### 4. Approximation of $q$ th power of operators

The statement in Lemma 11 is extended to arbitrary powers.

*Lemma 12.*—Let  $O$  and  $\tilde{O}$  be operators which satisfy the inequality

$$\|O - \tilde{O}\|_{2qp} \leq \delta \|O\|_{2qp} \quad (\delta \leq 1). \quad (\text{B10})$$

Then, the  $p$ th power of the operator  $O^\dagger O$  is close to  $(\tilde{O}^\dagger \tilde{O})^p$  as follows:

$$\|(O^\dagger O)^q - (\tilde{O}^\dagger \tilde{O})^q\|_p \leq 3\delta q e^{3\delta q} \|(O^\dagger O)^q\|_p. \quad (\text{B11})$$

The proof is a simple generalization of Proposition 1 in Ref. [95] to arbitrary Schatten  $p$ -norms.

#### a. Proof of Lemma 12

Following Ref. [95], we start from the equation as follows:

$$(O^\dagger O)^q - (\tilde{O}^\dagger \tilde{O})^q = \sum_{s=1}^q (O^\dagger O)^{q-s} (O^\dagger O - \tilde{O}^\dagger \tilde{O}) (\tilde{O}^\dagger \tilde{O})^{s-1}. \quad (\text{B12})$$

We can easily check that the above equation holds for arbitrary  $q$ . By using the triangle inequality for the Schatten norm, we have

$$\begin{aligned} \|(O^\dagger O)^q - (\tilde{O}^\dagger \tilde{O})^q\|_p &\leq \sum_{s=1}^q \|(O^\dagger O)^{q-s} (O^\dagger O - \tilde{O}^\dagger \tilde{O}) (\tilde{O}^\dagger \tilde{O})^{s-1}\|_p. \end{aligned} \quad (\text{B13})$$

Then, our task is to estimate the upper bound of the norm of  $(O^\dagger O)^{q-s} (O^\dagger O - \tilde{O}^\dagger \tilde{O}) (\tilde{O}^\dagger \tilde{O})^{s-1}$ . From the generalized Hölder inequality (B1), we obtain

$$\begin{aligned} &\|(O^\dagger O)^{q-s} (O^\dagger O - \tilde{O}^\dagger \tilde{O}) (\tilde{O}^\dagger \tilde{O})^{s-1}\|_p \\ &\leq \|(O^\dagger O)^{q-s}\|_{pq/(q-s)} \|O^\dagger O - \tilde{O}^\dagger \tilde{O}\|_{pq} \|(\tilde{O}^\dagger \tilde{O})^{s-1}\|_{pq/(s-1)} \\ &\leq \|O^\dagger O\|_{pq}^{q-s} \cdot 3\delta \|O^\dagger O\|_{pq} \cdot \|\tilde{O}^\dagger \tilde{O}\|_{pq}^{s-1}, \end{aligned} \quad (\text{B14})$$

where the equations  $\|(O^\dagger O)^{q-s}\|_{pq/(q-s)} = \|O^\dagger O\|_{pq}^{q-s}$  and  $\|(\tilde{O}^\dagger \tilde{O})^{s-1}\|_{pq/(s-1)} = \|\tilde{O}^\dagger \tilde{O}\|_{pq}^{s-1}$  are straightforwardly derived from the definition (10) and we use the inequality (B9) for  $\|O^\dagger O - \tilde{O}^\dagger \tilde{O}\|_{pq}$ . Furthermore, by using  $\|\tilde{O}^\dagger \tilde{O}\|_{pq} = \|\tilde{O}^\dagger \tilde{O} - O^\dagger O + O^\dagger O\|_{pq} \leq \|\tilde{O}^\dagger \tilde{O} - O^\dagger O\|_{pq} + \|O^\dagger O\|_{pq} \leq (3\delta + 1) \|O^\dagger O\|_{pq}$ , the inequality (B14) reduces to

$$\begin{aligned} &\|(O^\dagger O)^{q-s} (O^\dagger O - \tilde{O}^\dagger \tilde{O}) (\tilde{O}^\dagger \tilde{O})^{s-1}\|_p \\ &\leq 3\delta(3\delta + 1)^q \|O^\dagger O\|_{pq}^q \leq 3\delta e^{3\delta q} \|(O^\dagger O)^q\|_p. \end{aligned} \quad (\text{B15})$$

By applying the inequality (B15) to Eq. (B13), we obtain the main inequality (B11). This completes the proof of the inequality (B11). ■

### 5. Upper bound on the norm for multicommutators

For the norm of multicommutators, we can prove the following lemma (see Lemma 3 in Ref. [125]).

*Lemma 13.*—Let  $\{A_s\}_{s=1}^M$  be  $k_s$ -local operators such that

$$A_s = \sum_{|X| \leq k_s} a_{s,X}, \quad \max_{i \in \Lambda} \sum_{X: X \ni i} \|a_{s,X}\| \leq g_s. \quad (\text{B16})$$

Then, for an arbitrary operator  $O_X$  supported on a subset  $X$ , the norm of the multicommutator is bounded from above by

$$\|\text{ad}_{A_M} \text{ad}_{A_{M-1}} \cdots \text{ad}_{A_1}(O_X)\| \leq \prod_{m=1}^M (2g_m K_m) \|O_X\|, \quad (\text{B17})$$

where  $K_m := |X| + \sum_{s \leq m-1} k_s$ .

For  $g_1 = g_2 = \cdots = g_M = g$  and  $k_1 = k_2 = \cdots = k_M = k$ , we have

$$\begin{aligned} & \|\text{ad}_{A_M} \text{ad}_{A_{M-1}} \cdots \text{ad}_{A_1}(O_X)\| \\ & \leq (2gk)^M \frac{|X|}{k} \left( \frac{|X|}{k} + 1 \right) \cdots \left( \frac{|X|}{k} + M - 1 \right) \|O_X\|. \end{aligned} \quad (\text{B18})$$

## APPENDIX C: FULL PROOF OF PROPOSITION 9

In this section, we show the proof outline of Proposition 9 in Sec. VI which plays key roles in the proofs of the main results (Theorems 1, 2, and 6). We prove it based on several essential Lemmas 14–17. Throughout the proof, while considering the Schmidt rank for a target decomposition  $\Lambda = L \cup R$ , we denote  $\text{SR}(O, i_0)$  by  $\text{SR}(O)$  for simplicity.

### 1. Proof strategy

We here relabel each of the sites such that  $L = \{i\}_{i \leq \ell/2}$  and  $R = \{i\}_{i \geq \ell/2+1}$ , where the length  $\ell$  is a multiple of 4 to

be chosen later. We can arbitrarily extend the system size  $\Lambda \rightarrow \Lambda \sqcup \delta\Lambda$  without changing the Hamiltonian. We have only to add zero operators:

$$H'_{\Lambda \sqcup \delta\Lambda} = H_\Lambda + \hat{O}_{\delta\Lambda}, \quad (\text{C1})$$

where  $\hat{O}_{\delta\Lambda}$  is the zero operator acting on  $\delta\Lambda$ . Note that  $\hat{O}_{\delta\Lambda}$  still satisfies the form of Eq. (A2).

We then decompose the total system into three pieces  $L_0$ ,  $S$ , and  $R_0$  (see Fig. 5), where  $L_0 = \{i\}_{i \leq 0}$ ,  $S = \{i\}_{1 \leq i \leq \ell}$ , and  $R_0 = \{i\}_{i \geq \ell+1}$ . Accordingly, we also decompose the Hamiltonian as follows:

$$\begin{aligned} H &= H_S + H_{L_0} + H_{R_0} + v_0 + v_\ell, & H_S &:= \sum_{X \subset S} h_X, \\ H_{L_0} &:= \sum_{X \subset L_0} h_X, & H_{R_0} &:= \sum_{X \subset R_0} h_X, \end{aligned} \quad (\text{C2})$$

where  $v_0$  and  $v_\ell$  are defined by Eq. (A3). We note that  $H_S$ ,  $H_{L_0}$ , and  $H_{R_0}$  commute with each other. By shifting the energy origin appropriately, we set

$$H_S \geq 0, \quad (\text{C3})$$

where  $\geq$  means that  $H_S$  is positive semidefinite. We divide  $\beta$  into  $2q$  pieces ( $q \in \mathbb{N}$ ) and introduce

$$\rho_0 := e^{-\beta_0 H}, \quad \beta_0 := \beta / (2q).$$

The first step of the proof is the approximation of  $\rho_0$ , which is in the following form:

$$\begin{aligned} \tilde{\rho}_0 &:= \tilde{\Phi}_0^\dagger e^{-\beta_0(H_{L_0} + H_{R_0})} \mathcal{F}_m(\beta_0 H_S) \tilde{\Phi}_0, \\ \tilde{\Phi}_0 &:= \Phi_{L_1} \otimes \Phi_{R_1}, \end{aligned} \quad (\text{C4})$$

where  $\Phi_{L_1}$  and  $\Phi_{R_1}$  are operators supported on  $L_1$  and  $R_1$ , respectively (i.e.,  $L_1 = \{i\}_{i \leq \ell/4}$  and  $R_1 = \{i\}_{i \geq 3\ell/4+1}$ ), and the degree  $m$  polynomial  $\mathcal{F}_m(x)$  approximates the exponential function  $e^{-\beta_0 x}$ . For every  $\delta \leq 1/(3q)$ , we estimate the length  $\ell$  and the degree  $m$  such that

$$\|\rho_0 - \tilde{\rho}_0\|_{2qP} \leq \delta \|\rho_0\|_{2qP}. \quad (\text{C5})$$

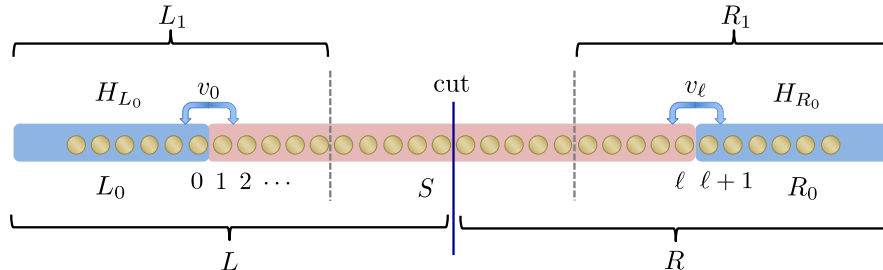


FIG. 5. The decomposition of the system considered in the proof.

Then, by applying inequality (B11), we have

$$\|e^{-\beta H} - \tilde{\rho}_0^{2q}\|_p \leq 3\delta q e^{3\delta q} \|\rho_0^{2q}\|_p \leq \epsilon \|e^{-\beta H}\|_p \quad (\text{C6})$$

with  $\epsilon = 3eq\delta$ , where we use  $\rho_0^{2q} = e^{-\beta H}$  and  $3\delta q e^{3\delta q} \leq 3eq\delta$  from  $\delta \leq 1/(3q)$ . Therefore, by choosing  $\hat{\rho} = \tilde{\rho}_0^{2q}/\text{tr}(e^{-\beta H})$ , we can achieve the bound (49). Note that the condition  $\epsilon \leq e$  in Proposition 9 is due to the equations  $\epsilon = 3eq\delta$  and  $\delta \leq 1/(3q)$ .

The second step is to estimate the upper bound of the Schmidt rank of  $\tilde{\rho}_0^{2q}$ , which is given by

$$[\tilde{\Phi}_0^\dagger e^{-\beta_0(H_{L_0} + H_{R_0})} \mathcal{F}_m(\beta_0 H_S) \tilde{\Phi}_0]^{2q}. \quad (\text{C7})$$

Then, the sufficient Schmidt rank to achieve the inequality (C6) is given by a function of  $q$  [see Eq. (C52) below for more details]. By choosing  $q$  so that the Schmidt rank is minimum, we show that the Schmidt rank is upper bounded by Eq. (50). We thus prove Proposition 9. In the following, we show the details of the above arguments.

## 2. Approximation of $\rho_0$

In the following, we define a parameter  $\nu$  as follows:

$$\nu = \max[\beta_0, \log(6/\delta)]. \quad (\text{C8})$$

In addition, we choose  $q$  such that

$$q^2 \geq \beta. \quad (\text{C9})$$

Let  $H_0 := H_S + H_{L_0} + H_{R_0}$ . We first relate the two operators  $\rho_0 = e^{-\beta_0 H}$  and  $e^{-\beta_0 H_0}$ . We can formally write the following:

$$\rho_0 = \Phi_0 e^{-\beta_0 H_0} \Phi_0^\dagger, \quad (\text{C10})$$

where  $\Phi_0$  is usually a highly nonlocal operator. The first lemma ensures that the  $\Phi$  is approximated by an operator supported on  $L_1 \sqcup R_1$ .

*Lemma 14.*—There exists an operator  $\tilde{\Phi}_0 = \Phi_{L_1} \otimes \Phi_{R_1}$  such that, for

$$\rho'_0 = \tilde{\Phi}_0 e^{-\beta_0 H_0} \tilde{\Phi}_0^\dagger, \quad (\text{C11})$$

we have

$$\|\rho'_0 - \rho_0\|_{p_0} \leq 3e^{-c_1 \ell/\beta_0 + c_2 \beta_0} \|\rho_0\|_{p_0} \quad (\text{C12})$$

for arbitrary  $p_0 \in \mathbb{N}$ , where we assume  $-c_1 \ell/\beta_0 + c_2 \beta_0 \leq 0$  and  $c_1$  and  $c_2$  are  $\mathcal{O}(1)$  constants.

The proof of this lemma is based on the belief propagation [29,69] and the Lieb-Robinson bound [126,127].

## a. Proof of Lemma 14

For the proof, we start from the belief propagation [29], which gives

$$\rho_0 = e^{-\beta_0 H} = \Phi_0 e^{-\beta_0 H_0} \Phi_0^\dagger, \quad (\text{C13})$$

where the operator  $\Phi_0$  is defined as

$$\begin{aligned} \Phi_0 &:= \mathcal{T} e^{\int_0^1 \phi(\tau) d\tau}, \\ \phi(\tau) &:= \frac{-\beta_0 v_0 - \beta_0 v_\ell}{2} \\ &\quad + i\beta_0 \int_{-\infty}^{\infty} g(t)[v_0(t, H_\tau) + v_\ell(t, H_\tau)] dt, \end{aligned} \quad (\text{C14})$$

where  $H_\tau = H_0 + \tau(v_0 + v_\ell)$ ,  $\mathcal{T}$  denotes the ordering operator,  $v_0(t, H_\tau) = e^{itH_\tau} v_0 e^{-itH_\tau}$ ,  $v_\ell(t, H_\tau) = e^{itH_\tau} v_\ell e^{-itH_\tau}$ , and  $g(t)$  is defined as

$$g(t) := \text{sgn}(t) \frac{e^{-2\pi|t|/\beta_0}}{1 - e^{-2\pi|t|/\beta_0}}. \quad (\text{C15})$$

Note that the function  $g(t)$  decays exponentially with  $t$  and, hence, the operator  $\phi(\tau)$  is quasilocal due to the Lieb-Robinson bound [126,127]. We aim to obtain the approximation  $\Phi_0 \approx \Phi_{L_1} \otimes \Phi_{R_1} =: \tilde{\Phi}_0$  and consider the norm difference of

$$\epsilon_\Phi := \|\tilde{\Phi}_0 e^{-\beta_0 H_0} \tilde{\Phi}_0^\dagger - e^{-\beta_0 H}\|_{p_0} \quad (\text{C16})$$

for arbitrary  $p_0 \in \mathbb{N}$ .

In order to quantitatively evaluate the quasilocality of  $\phi(\tau)$ , we first define  $v_0(t, H_\tau, L_1)$  as an approximation of  $v_0(t, H_\tau)$  in the region  $L_1$ :

$$v_0(t, H_\tau, L_1) := \frac{1}{\mathcal{D}_{\Lambda \setminus L_1}} \text{tr}_{\Lambda \setminus L_1}[v_0(t, H_\tau)] \otimes \hat{1}_{\Lambda \setminus L_1}.$$

We define  $v_\ell(t, H_\tau, R_1)$  in the same way. By utilizing the Lieb-Robinson bound [127], we obtain the approximation error of

$$\|v_0(t, H_\tau) - v_0(t, H_\tau, L_1)\| \leq c|t|e^{-c'(\ell/4 - vt)}, \quad (\text{C17})$$

where  $c$ ,  $c'$ , and  $v$  are constants of  $\mathcal{O}(1)$  and we obtain the same upper bound for  $\|v_\ell(t, H_\tau) - v_\ell(t, H_\tau, R_1)\|$ . By using the notations of  $v_0(t, H_\tau, L_1)$  and  $v_\ell(t, H_\tau, R_1)$ , we define  $\tilde{\phi}_{L_1}(\tau)$  and  $\tilde{\phi}_{R_1}(\tau)$  as follows:

$$\begin{aligned} \tilde{\phi}_{L_1}(\tau) &:= \frac{-\beta_0}{2} v_0 + i\beta_0 \int_{-\infty}^{\infty} g(t) v_0(t, H_\tau, L_1) dt, \\ \tilde{\phi}_{R_1}(\tau) &:= \frac{-\beta_0}{2} v_\ell + i\beta_0 \int_{-\infty}^{\infty} g(t) v_\ell(t, H_\tau, R_1) dt. \end{aligned}$$

We notice that  $\tilde{\phi}_{L_1}(\tau)$  and  $\tilde{\phi}_{R_1}(\tau)$  are supported on the subsets  $L_1$  and  $R_1$ , respectively. We then approximate  $\phi(\tau)$  by  $\tilde{\phi}(\tau) = \tilde{\phi}_{L_1}(\tau) + \tilde{\phi}_{R_1}(\tau)$  with an error of



$$\|\phi(\tau) - \tilde{\phi}(\tau)\| \leq c_0 \beta_0^2 e^{-c_1 \ell / \beta_0} \quad (0 \leq \tau \leq \beta_0) \quad (\text{C18})$$

with  $c_0$  and  $c_1$  constants of  $\mathcal{O}(1)$ , where the inequality is derived from the approximation error in Eq. (C17) and the exponential decay of  $g(t)$  as in Eq. (C15).

From the approximation of  $\phi(\tau)$  by  $\tilde{\phi}(\tau)$ , we define  $\tilde{\Phi}_0$  as

$$\tilde{\Phi}_0 := \mathcal{T} e^{-\int_0^1 \tilde{\phi}(\tau) d\tau} = \Phi_{L_1} \otimes \Phi_{R_1}, \quad (\text{C19})$$

where we define  $\Phi_{L_1} := \mathcal{T} e^{-\int_0^1 \tilde{\phi}_{L_1}(\tau) d\tau}$  and  $\Phi_{R_1} := \mathcal{T} e^{-\int_0^1 \tilde{\phi}_{R_1}(\tau) d\tau}$ . By using the inequality (C18), we can obtain the approximation error of  $\tilde{\Phi}_0$  by

$$\begin{aligned} \|1 - \tilde{\Phi}_0 \Phi_0^{-1}\| &\leq c_0 \beta_0^2 e^{-c_1 \ell / \beta_0} e^2 \int_0^1 \|\phi(\tau)\| d\tau \\ &\leq c_0 \beta_0^2 e^{-c_1 \ell / \beta_0 + 2c_1' \beta_0}, \end{aligned} \quad (\text{C20})$$

with  $c_1'$  an  $\mathcal{O}(1)$  constant, where the upper bound  $\|\phi(\tau)\| \leq c_1' \beta_0$  can be derived by following Ref. [69] [see Eq. (42) therein]. By letting  $O_0 := \tilde{\Phi}_0 \Phi_0^{-1}$ , we have, using the triangle inequality,

$$\begin{aligned} \|\Phi_0 e^{-\beta_0 H_0} \Phi_0^\dagger - \tilde{\Phi}_0 e^{-\beta_0 H_0} \tilde{\Phi}_0^\dagger\|_{p_0} &= \|\rho_0 - O_0 \rho_0 O_0^\dagger\|_{p_0} \\ &\leq \|(1 - O_0) \rho_0 O_0^\dagger\|_{p_0} + \|O_0 \rho_0 (1 - O_0^\dagger)\|_{p_0} \\ &\quad + \|(1 - O_0) \rho_0 (1 - O_0^\dagger)\|_{p_0}. \end{aligned} \quad (\text{C21})$$

From the upper bound (C20), the norm of  $1 - O_0$  satisfies the following inequality:

$$\|1 - O_0\| = \|1 - \tilde{\Phi}_0 \Phi_0^{-1}\| \leq e^{-c_1 \ell / \beta_0 + c_2 \beta_0},$$

where we choose  $c_2 = \mathcal{O}(1)$  such that  $c_0 \beta_0^2 e^{2c_1' \beta_0} \leq e^{c_2 \beta_0}$ . Then, the condition  $-c_1 \ell / \beta_0 + c_2 \beta_0 \leq 0$  in the lemma implies  $\|1 - O_0\| \leq 1$ . Therefore, by applying the Hölder inequality (B2) to each of the terms in Eq. (C21), we obtain

$$\begin{aligned} \|\Phi_0 e^{-\beta_0 H_0} \Phi_0^\dagger - \tilde{\Phi}_0 e^{-\beta_0 H_0} \tilde{\Phi}_0^\dagger\|_{p_0} &\leq \|\rho_0\|_{p_0} (\|1 - O_0\|^2 + 2\|1 - O_0\|) \\ &\leq 3e^{-c_1 \ell / \beta_0 + c_2 \beta_0} \|\rho_0\|_{p_0}, \end{aligned} \quad (\text{C22})$$

where, in the second inequality, we get  $\|1 - O_0\|^2 \leq \|1 - O_0\|$  due to  $\|1 - O_0\| \leq 1$ . This completes the proof. ■

The lemma implies that, as the length  $\ell$  becomes large, the approximation error decays exponentially with  $e^{-\mathcal{O}(\ell/\beta_0)}$ . Thus, in order to achieve the inequality

$$\|\rho'_0 - \rho_0\|_{2qp} \leq \frac{\delta}{2} \|\rho_0\|_{2qp}, \quad (\text{C23})$$

we need to choose  $\ell$  as

$$\ell \geq \frac{c_2}{c_1} \beta_0^2 + \frac{\beta_0}{c_1} \log(6/\delta). \quad (\text{C24})$$

By using the parameter  $\nu$  in Eq. (C8), we can write

$$\ell = \tilde{c}_1 \beta_0 \nu = \tilde{c}_1 \nu \beta / (2q), \quad (\text{C25})$$

where  $\tilde{c}_1$  is a constant of  $\mathcal{O}(1)$ .

Second, we approximate  $e^{-\beta_0 H_0}$  by an operator with small Schmidt rank. For this purpose, we use the fact that  $H_S$ ,  $H_{L_0}$ , and  $H_{R_0}$  commute with each other and write  $e^{-\beta_0 H_0} = e^{-\beta_0 (H_{L_0} + H_{R_0})} e^{-\beta_0 H_S}$ . Then, we approximate  $e^{-\beta_0 H_S}$  by a low-degree polynomial of  $H_S$ . The most straightforward approximation is given by the truncation of the Taylor expansion, which gives a good approximation of  $e^{-\beta_0 H_S}$  by taking the polynomial degree as large as  $\|\beta_0 H_S\| + \log(1/\delta_0)$  with  $\delta_0$  the precision error. Unfortunately, we cannot get any improvement of the thermal area law if we utilize the Taylor expansion.

One of the key aspects of our proof is the use of the following lemma from Theorem 4.1 in Ref. [93][], which allows us to achieve the improved thermal area law.

*Lemma 15.*—Let  $\delta_0 \in (0, 1)$ . For any  $m$  satisfying

$$m > c_f \sqrt{\max[\beta_0 \|H_S\|, \log(1/\delta_0)] \log(1/\delta_0)} \quad (\text{C26})$$

[with  $c_f = \mathcal{O}(1)$ ], there exists a polynomial  $\mathcal{F}_m(x)$  with degree  $m$  that satisfies

$$|\mathcal{F}_m(x) - e^{-x}| \leq \delta_0 \quad \text{for } x \in [0, \beta_0 \|H_S\|]. \quad (\text{C27})$$

When  $\beta_0 \|H_S\| \gg \log(1/\delta_0)$ , the above estimation gives a significantly better polynomial degree than that from the Taylor expansion.

We recall that this polynomial approximation is obtained from the Chebyshev polynomial expansion (17) in Sec. III A, which is characterized by the random walk behavior (see Fig. 3).

By using the polynomial  $\mathcal{F}_m(x)$  defined above, we approximate the operator  $\rho'_0$  in Eq. (C11) as

$$\tilde{\rho}_0 := \tilde{\Phi}_0 e^{-\beta_0 (H_{L_0} + H_{R_0})} \mathcal{F}_m(\beta_0 H_S) \tilde{\Phi}_0^\dagger. \quad (\text{C28})$$

Because of Eq. (C3), the spectrum of  $\beta_0 H_S$  is included in the span of  $[0, \beta_0 \|H_S\|]$ , and, hence, the inequality (C27) gives

$$\|\mathcal{F}_m(\beta_0 H_S) - e^{-\beta_0 H_S}\| \leq \delta_0. \quad (\text{C29})$$

We note that the current approximation (C29) is obtained in terms of the operator norm (i.e., Schatten  $\infty$ -norm) instead of the generic Schatten  $p$ -norm. The next problem is to estimate the approximation error  $\|\rho'_0 - \tilde{\rho}_0\|_{p_0}$  for an arbitrary Schatten  $p_0$ -norm. We prove the following lemma.

*Lemma 16.*—Let  $p_0 \in \mathbb{N}$  and  $\delta_0 \in (0, 1)$ . Under the choice of  $\Phi_{L_1} \otimes \Phi_{R_1}$  in Lemma 14,  $\ell$  in Eq. (C24), and  $m, \mathcal{F}_m(x)$  in Lemma 15, we have

$$\|\rho'_0 - \tilde{\rho}_0\|_{p_0} \leq \mathcal{D}_S^{1/p_0} \delta_0 e^{c_3 \beta_0} \|\rho_0\|_{p_0}, \quad (\text{C30})$$

where  $c_3$  is an  $\mathcal{O}(1)$  constant.

### b. Proof of Lemma 16

From the definitions (C11) and (C28) of  $\rho'_0$  and  $\tilde{\rho}_0$ , respectively, we start from the inequality

$$\begin{aligned} \|\rho'_0 - \tilde{\rho}_0\|_{p_0} &= \|\tilde{\Phi}_0^\dagger e^{-\beta_0(H_{L_0} + H_{R_0})} [\mathcal{F}_m(H_S) - e^{-\beta_0 H_S}] \tilde{\Phi}_0\|_{p_0} \\ &\leq \|e^{-\beta_0(H_{L_0} + H_{R_0})} [\mathcal{F}_m(H_S) - e^{-\beta_0 H_S}]\|_{p_0} \cdot \|\tilde{\Phi}_0\|^2, \end{aligned} \quad (\text{C31})$$

where we use Hölder's inequality (B2). From the definition (C19) of  $\Phi_{L_1} \otimes \Phi_{R_1}$ , we obtain

$$\|\tilde{\Phi}_0\|^2 = \|\Phi_{L_1} \otimes \Phi_{R_1}\|^2 \leq e^{2c' \beta_0}. \quad (\text{C32})$$

We next consider

$$\begin{aligned} &\|e^{-\beta_0(H_{L_0} + H_{R_0})} [\mathcal{F}_m(H_S) - e^{-\beta_0 H_S}]\|_{p_0}^{p_0} \\ &= \sum_{s=1}^{\mathcal{D}_{L_0 R_0}} \sum_{s'=1}^{\mathcal{D}_S} e^{-p_0 \beta_0 \tilde{E}_s} |\mathcal{F}_m(\varepsilon_{s'}) - e^{-\beta_0 \varepsilon_{s'}}|^{p_0}, \end{aligned} \quad (\text{C33})$$

where  $\{\tilde{E}_s\}_{s=1}^{\mathcal{D}_{L_0 R_0}}$  and  $\{\varepsilon_{s'}\}_{s'=1}^{\mathcal{D}_S}$  are eigenvalues of  $H_{L_0} + H_{R_0}$  and  $H_S$ , respectively. Note that the Hamiltonians  $H_{L_0}$ ,  $H_{R_0}$ , and  $H_S$  commute with each other and are diagonalizable simultaneously. From the assumption (C3), we have  $\varepsilon_1 = 0$ , and  $\varepsilon_{\mathcal{D}_S} \leq \|H_S\|$ .

From the inequality (C27), which is

$$|\mathcal{F}_m(x) - e^{-x}| \leq \delta_0 \quad \text{for } x \in [0, \beta_0 \|H_S\|], \quad (\text{C34})$$

we have

$$|\mathcal{F}_m(\varepsilon_{s'}) - e^{-\beta_0 \varepsilon_{s'}}|^{p_0} \leq \delta_0^{p_0}. \quad (\text{C35})$$

By applying the above inequality to Eq. (C33), we obtain

$$\begin{aligned} &\|e^{-\beta_0(H_{L_0} + H_{R_0})} [\mathcal{F}_m(H_S) - e^{-\beta_0 H_S}]\|_{p_0}^{p_0} \\ &\leq \sum_{s=1}^{\mathcal{D}_{L_0 R_0}} \sum_{s'=1}^{\mathcal{D}_S} e^{-p_0 \beta_0 \tilde{E}_s} \delta_0^{p_0} \\ &\leq \frac{\mathcal{D}_S \delta_0^{p_0}}{\sum_{s'=1}^{\mathcal{D}_S} e^{-p_0 \beta_0 \varepsilon_{s'}}} \sum_{s=1}^{\mathcal{D}_{L_0 R_0}} \sum_{s'=1}^{\mathcal{D}_S} e^{-p_0 \beta_0 \tilde{E}_s} e^{-p_0 \beta_0 \varepsilon_{s'}} \\ &\leq \mathcal{D}_S \delta_0^{p_0} \|e^{-\beta_0(H_{L_0} + H_{R_0} + H_S)}\|_{p_0}^{p_0}, \end{aligned} \quad (\text{C36})$$

where we use  $\sum_{s'=1}^{\mathcal{D}_S} e^{-p_0 \beta_0 \varepsilon_{s'}} \geq e^{-p_0 \beta_0 \varepsilon_1} = 1$ .

We next consider the upper bound of  $\|e^{-\beta_0 H_0}\|_{p_0}$  in terms of  $\|e^{-\beta_0 H}\|_{p_0}$ . Recall that  $H_0 = H_{L_0} + H_{R_0} + H_S$  and, hence,  $e^{-\beta_0 H_0} = e^{-\beta_0(H - v_0 - v_\ell)}$ . By using the Golden-Thompson inequality, we have

$$\begin{aligned} \text{tr}(e^{-p_0 \beta_0 H_0}) &\leq \text{tr}(e^{-p_0 \beta_0 H} \cdot e^{-p_0 \beta_0 (v_0 + v_\ell)}) \\ &\leq e^{p_0 \beta_0 \|v_0 + v_\ell\|} \text{tr}(e^{-p_0 \beta_0 H}) \\ &\leq e^{2gk p_0 \beta_0} \|e^{-\beta_0 H}\|_{p_0}^{p_0}, \end{aligned} \quad (\text{C37})$$

where we use  $\|v_0 + v_\ell\| \leq 2gk$  from the condition in Eq. (A2). Note that  $\text{tr}(e^{-p_0 \beta_0 H_0}) = \|e^{-\beta_0 H_0}\|_{p_0}^{p_0}$ . By combining the inequalities (C36) and (C37), we arrive at the inequality

$$\begin{aligned} &\|e^{-\beta_0(H_{L_0} + H_{R_0})} [\mathcal{F}_m(H_S) - e^{-\beta_0 H_S}]\|_{p_0}^{p_0} \\ &\leq \mathcal{D}_S \delta_0^{p_0} e^{2gk p_0 \beta_0} \|e^{-\beta_0 H}\|_{p_0}^{p_0}. \end{aligned} \quad (\text{C38})$$

By applying the inequalities (C32) and (C38) to Eq. (C31), we obtain the main inequality (C30) with  $c_3 = 2c' + 2gk$ . This completes the proof.  $\blacksquare$

Let us substitute  $p_0 = 2qp$  in Lemma 16 and choose  $\delta_0$  that satisfies

$$\mathcal{D}_S^{1/(2qp)} \delta_0 e^{c_3 \beta_0} = \frac{\delta}{2}.$$

This choice ensures that  $\|\rho'_0 - \tilde{\rho}_0\|_{2qp} \leq (\delta/2) \|\rho_0\|_{2qp}$ , and we conclude that

$$\|\rho_0 - \tilde{\rho}_0\|_{2qp} \leq \|\rho_0 - \rho'_0\|_{2qp} + \|\rho'_0 - \tilde{\rho}_0\|_{2qp} \leq \delta \|\rho_0\|_{2qp},$$

where we use the inequality (C23). Therefore, the choice of  $\tilde{\rho}_0$  as in Eq. (C28) achieves the inequality (C5).

Let us simplify the expression for all the parameters appearing so far. We consider

$$\begin{aligned} \delta_0 &= \frac{\delta}{2} e^{-c_3 \beta_0} \mathcal{D}_S^{-1/(2qp)} = \frac{\delta}{2} e^{-c_3 \beta_0 - c_\varepsilon \ell / (2qp)} \\ &\rightarrow \log(1/\delta_0) = \log(2/\delta) + c_3 \beta_0 + \frac{c_\varepsilon \tilde{c}_1 \nu \beta_0}{2qp}, \end{aligned} \quad (\text{C39})$$

where we define  $\mathcal{D}_S = \zeta^\ell = e^{c_\zeta \ell}$  and use the expression of  $\ell$  in Eq. (C25). From the assumption (C9), we have  $\beta_0 = (\beta/2q) \leq q/2$ . This result yields

$$\begin{aligned} \log(2/\delta) + c_3\beta_0 &\leq \log(1/\delta_0) \leq \log(2/\delta) + c_3\beta_0 + \frac{c_\zeta \tilde{c}_1 \nu}{4p} \\ &\leq \log(2/\delta) + c_3\beta_0 + \frac{c_\zeta \tilde{c}_1 \nu}{4}, \end{aligned}$$

where we use  $p \geq 1$ . Using the definition (C8) of  $\nu$ , we can thus write

$$\log(1/\delta_0) = \tilde{c}_2 \nu \quad (\text{C40})$$

for some constant  $\tilde{c}_2$ . From the choice of  $\ell$  in Eq. (C24), we have

$$\max[\beta_0 \|H_S\|, \log(1/\delta_0)] = \mathcal{O}(\beta_0 \ell), \quad (\text{C41})$$

where we use  $\|H_S\| \leq g\ell$  from the condition in Eq. (A2). Hence, we obtain the following simpler form of  $m$ :

$$\begin{aligned} m &= \lceil c_f \sqrt{\max[\beta_0 \|H_S\|, \log(1/\delta_0)] \log(1/\delta_0)} \rceil \\ &= \tilde{c}'_2 \sqrt{\nu \beta_0 \ell}. \end{aligned} \quad (\text{C42})$$

### 3. Schmidt-rank analysis

The remaining task is to estimate the Schmidt rank of the operator  $\tilde{\rho}_0^{2q}$  which is given by Eq. (C7). For this purpose, we consider the following more general problem for the simplicity of notation. We also utilize the lemma in the subsequent sections. Let us define a decomposition of the total system into  $\tilde{L}$ ,  $S$ , and  $\tilde{R}$  (see Fig. 6). We then aim to estimate the Schmidt rank of an operator of the form

$$\hat{\mathcal{G}}_{m,M} = [\Phi_1 \mathcal{G}_m(H_S) \Phi_2]^M, \quad (\text{C43})$$

where  $\mathcal{G}_m(x)$  is an arbitrary degree  $m$  polynomial, the operators  $\Phi_1$  and  $\Phi_2$  are supported on  $\tilde{L}$  and  $\tilde{R}$ , respectively, and  $H_S$  is a local Hamiltonian on the subset  $S$  ( $|S| = \ell$ ). The Schmidt-rank estimation for an arbitrary polynomial of  $H$  is given in Ref. [23]. However, in the present case, the additional operators  $\Phi_1$  and  $\Phi_2$  prohibit the direct application of that results to Eq. (C43). In the following lemma, we can obtain the

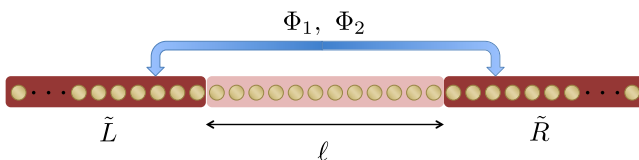


FIG. 6. The decomposition of the system in the Schmidt-rank analysis.

modified version of the Schmidt-rank estimation in Ref. [23]. For the generalization to high-dimensional systems in Appendix D, we consider the high-dimensional Hamiltonian (A5).

*Lemma 17.*—For an arbitrary operator in the form of Eq. (C43), the Schmidt rank across the bipartition of the system to the left and right at the point  $i \in S$  is upper bounded by

$$\text{SR}(\hat{\mathcal{G}}_{m,M}, i) \leq \min_{\tilde{\ell}: \tilde{\ell} \leq \ell} [\zeta^{\tilde{\ell} |\partial\Lambda|} (10mMD_{\text{loc}})^{2M+2\tilde{\ell}+(2kmM/\tilde{\ell})}], \quad (\text{C44})$$

where  $\partial\Lambda$  and  $D_{\text{loc}}$  are defined in Eqs. (A6) and (A8), respectively. If we consider a one-dimensional Hamiltonian with two-body interactions ( $k = 2$ ), we have  $|\partial\Lambda| = 1$  and  $D_{\text{loc}} \leq \zeta$ .

We can further extend Lemma 17 to the following operator:

$$\hat{\mathcal{G}}_{m,M}^{(p)} = [\Phi_1 \mathcal{G}_m(H_{S_1}) \mathcal{G}_m(H_{S_2}) \cdots \mathcal{G}_m(H_{S_p}) \Phi_2]^M,$$

where  $S_j \subseteq S$  ( $j = 1, 2, \dots, p$ ) with  $|S_j| = \ell$ . We then obtain the following corollary.

*Corollary 18.*—For an arbitrary operator in the form of Eq. (C43), the Schmidt rank across the bipartition of the system to the left and right at the point  $i \in S$  is upper bounded by

$$\text{SR}(\hat{\mathcal{G}}_{m,M}^{(p)}, i) \leq \min_{\tilde{\ell}: \tilde{\ell} \leq \ell} [\zeta^{\tilde{\ell} |\partial\Lambda|} (10mMD_{\text{loc}})^{2pM+2p\tilde{\ell}+(2pkmM/\tilde{\ell})}], \quad (\text{C45})$$

where  $\partial\Lambda$  and  $D_{\text{loc}}$  are defined in Eqs. (A6) and (A8), respectively.

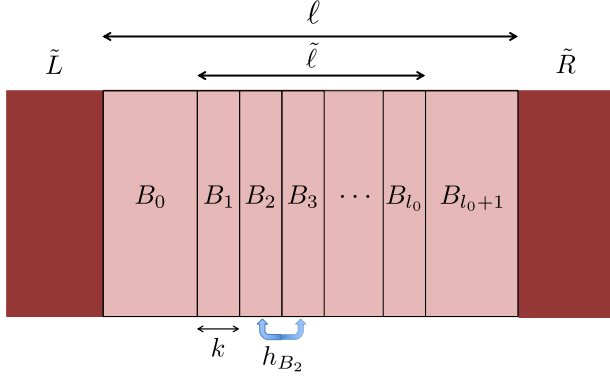
*Proof of Corollary 18.*—The proof is the same as that of Lemma 17. The difference is that the inequality (C47) is replaced by

$$\begin{aligned} \text{SR}(\hat{\mathcal{G}}_{m,M}^{(p)}, i) &\leq \min_{\tilde{\ell}: \tilde{\ell} \leq \ell} \left[ \zeta^{\tilde{\ell} |\partial\Lambda|} \prod_{j=1}^p m \binom{M}{\tilde{\ell}}^2 \max_{s \in [l_0]} \text{SR}_s(\hat{\mathcal{G}}_{m,M}^{(j), \leq (mM/l_0), s}) \right], \end{aligned}$$

where  $\hat{\mathcal{G}}_{m,M}^{(j), \leq (mM/l_0), s}$  for the Hamiltonian  $H_{S_j}$  is defined in the same way as  $\hat{\mathcal{G}}_{m,M}^{\leq (mM/l_0), s}$  in Eq. (C47) for the Hamiltonian  $H_S$ . We then obtain the same inequality as Eq. (C50) and prove the inequality (C45). This completes the proof. ■

#### a. Proof of Lemma 17

We apply an analysis similar to that in Ref. [26], which modifies the proof in Ref. [23] for the Schmidt-rank estimation.


 FIG. 7. The decomposition of the subset  $S$  into blocks.

First, we decompose  $S$  into  $(l_0 + 2)$  blocks  $\{B_s\}_{s=0}^{l_0+1}$  with  $|B_s| = k$  ( $s = 1, 2, \dots, l_0$ ) and  $l_0 = \tilde{\ell}/k$  (see Fig. 7).

$$\text{SR}(\hat{\mathcal{G}}_{m,M}, i) \leq \min_{\tilde{\ell}: \tilde{\ell} \leq \ell} \left[ \binom{mM}{\tilde{\ell}}^2 \zeta^{\tilde{\ell}|\partial\Lambda} \max_{s \in [l_0]} \text{SR}_s(\hat{\mathcal{G}}_{m,M}^{\leq (mM/l_0), s}) \right], \quad l_0 = \tilde{\ell}/k, \quad (\text{C47})$$

where  $\text{SR}_s(\dots)$  is the Schmidt rank across the bipartition between  $B_s$  and  $B_{s+1}$ . Also, the operator  $\hat{\mathcal{G}}_{m,M}^{\leq (mM/l_0), s}$  is derived from  $\hat{\mathcal{G}}_{m,M}$  by considering only those terms in which  $h_{B_s}$  occurs at most  $(mM/l_0)$  times. Let  $H_S = P + h_{B_s} + Q$ , where  $P$  is to the “left” of  $h_{B_s}$  and  $Q$  is to the “right” of  $h_{B_s}$  and expand the powers  $H_S$ . From  $[P, Q] = 0$ , any particular power  $(H_S)^T$  is a linear combination of the following terms:

$$(P^{p_1} Q^{q_1}) h_{B_s} (P^{p_2} Q^{q_2}) h_{B_s} \dots (P^{p_{T'}} Q^{q_{T'}}) h_{B_s} (P^{p_{T'}} Q^{q_{T'}})$$

with  $\sum_{i=1}^{T'} (p_i + q_i) \leq T$  and  $T' \leq T$ . This combination allows us to expand  $\hat{\mathcal{G}}_{m,M}$  as a linear combination of the following terms:

$$\begin{aligned} & \Phi_1 (P^{p_{1,1}} Q^{q_{1,1}}) h_{B_s} (P^{p_{1,2}} Q^{q_{1,2}}) h_{B_s} \dots (P^{p_{1,T_1-1}} Q^{q_{1,T_1-1}}) h_{B_s} (P^{p_{1,T_1}} Q^{q_{1,T_1}}) \\ & \Phi_2 \Phi_1 (P^{p_{2,1}} Q^{q_{2,1}}) h_{B_s} (P^{p_{2,2}} Q^{q_{2,2}}) h_{B_s} \dots (P^{p_{2,T_2-1}} Q^{q_{2,T_2-1}}) h_{B_s} (P^{p_{2,T_2}} Q^{q_{2,T_2}}) \\ & \dots \\ & \Phi_2 \Phi_1 (P^{p_{M,1}} Q^{q_{M,1}}) h_{B_s} (P^{p_{M,2}} Q^{q_{M,2}}) h_{B_s} \dots (P^{p_{M,T_M-1}} Q^{q_{M,T_M-1}}) h_{B_s} (P^{p_{M,T_M}} Q^{q_{M,T_M}}) \Phi_2. \end{aligned} \quad (\text{C48})$$

Above, the positive integers  $T_i$  and the powers  $p_{i,k}, q_{i,k} \geq 0$  are such that

$$\sum_{i=1}^M \sum_{k=1}^{T_i} (p_{i,k} + q_{i,k}) \leq mM, \quad (\text{C49})$$

since the total degree is  $mM$ . But, recall that we are interested in  $\hat{\mathcal{G}}_{m,M}^{\leq (mM/l_0), s}$  where  $h_{B_s}$  occur at most  $(mM/l_0)$  times, which enforces the following constraint:

$$\sum_{i=1}^M (T_i - 1) \leq \frac{mM}{l_0} \Rightarrow \sum_{i=1}^M T_i \leq \frac{mM}{l_0} + M.$$

Here,  $\tilde{\ell}$  is a control parameter such that  $\tilde{\ell} \leq \ell$ . We then decompose the Hamiltonian  $H_S$  as

$$H_S = h_{B_0} + h_{B_{l_0+1}} + \sum_{s=1}^{l_0} h_{B_s}, \quad (\text{C46})$$

where  $h_{B_s}$  is comprised of the internal interactions in  $B_s$  and block-block interactions between  $B_s$  and  $B_{s+1}$ . Note that the interaction length is at most  $k$ , and, hence, only adjacent blocks can interact with each other. Also, from the inequality(A8), the Schmidt rank of  $h_{B_s}$  is upper bounded by  $D_{\text{loc}} = \zeta^{O(k)} |\partial\Lambda|$ .

We expand  $\mathcal{G}_m(H_S) = \sum_{j=0}^m a_j (H_S)^j$  by using the decomposition (C46). Using the polynomial interpolation argument in Ref. [23], it holds that (see Lemmas 5.2 and 5.3 in Ref. [26])

The number of the combinations of positive integers  $\{T_1, T_2, \dots, T_M\}$  satisfying  $\sum_{i=1}^M T_i = r$  is smaller than  $r$  multicomination from a set of  $M$  elements and, hence, is upper bounded by

$$\binom{\binom{M}{r}}{r} = \binom{M+r-1}{r} \leq 2^{M+r-1}.$$

Then, the combinations of positive integers  $\{T_1, T_2, \dots, T_M\}$  satisfying  $\sum_{i=1}^M T_i \leq (mM/l_0) + M$  is smaller than

$$\sum_{r=0}^{(mM/l_0)+M} 2^{M+r-1} \leq 2^{2M+(mM/l_0)}.$$



When a tuple  $\{T_1, T_2, \dots, T_M\}$  is given, the number of the nonzero integers in  $\{p_{i,k}, q_{i,k}\}_{i \in [M], k \in [T_i]}$  which appears in Eq. (C48) is equal to

$$\sum_{i=1}^M 2T_i \leq 2 \left( \frac{mM}{l_0} + M \right).$$

Therefore, for a fixed  $\{T_1, T_2, \dots, T_M\}$ , the number of the combinations of positive integers  $\{p_{i,k}, q_{i,k}\}_{i \in [M], k \in [T_i]}$  satisfying Eq. (C49) is upper bounded by  $(mM)$  multicomination from a set of  $(\sum_{i=1}^M 2T_i)$  elements:

$$\begin{aligned} \left( \binom{\sum_{i=1}^M 2T_i}{mM} \right) &\leq \left( \binom{\frac{2mM}{l_0} + 2M}{mM} \right) \\ &= \binom{\frac{2mM}{l_0} + 2M + mM - 1}{mM} \\ &\leq (5mM)^{(2mM/l_0) + 2M}. \end{aligned}$$

For each of the nonzero integers in  $\{p_{i,k}, q_{i,k}\}_{i \in [M], k \in [T_i]}$ , the Schmidt rank of the expression in Eq. (C48), across the cut between  $B_s$  and  $B_{s+1}$ , is at most  $D_{\text{loc}}^{mM/l_0}$ . It is because only  $h_{B_s}$  increases the Schmidt rank across the cut between  $B_s$  and  $B_{s+1}$  and the number of  $h_{B_s}$  appearing in Eq. (C48) is smaller than  $(mM/l_0)$  from the definition of  $\hat{\mathcal{G}}_{m,M}^{\leq (mM/l_0), s}$ . Therefore, we finally arrive at the inequality of

$$\begin{aligned} \text{SR}_s(\hat{\mathcal{G}}_{m,M}^{\leq (mM/l_0), s}) &\leq 2^{2M + (mM/l_0)} (5mM)^{(2mM/l_0) + 2M} D_{\text{loc}}^{mM/l_0} \\ &\leq (10mMD_{\text{loc}})^{2M + (2mM/l_0)} \\ &= (10mMD_{\text{loc}})^{2M + (2kmM/\tilde{\ell})}, \end{aligned}$$

where we use  $l_0 = \tilde{\ell}/k$  in the last equation. By applying the above inequality to Eq. (C47), we obtain

$$\begin{aligned} m \binom{M^2}{\tilde{\ell}} \varsigma^{\tilde{\ell}|\partial\Lambda|} \max_{s \in [l_0]} \text{SR}_s(\hat{\mathcal{G}}_{m,M}^{\leq (mM/l_0), s}) \\ \leq m \binom{M^2}{\tilde{\ell}} \varsigma^{\tilde{\ell}|\partial\Lambda|} (10mMD_{\text{loc}})^{2M + (2kmM/\tilde{\ell})} \\ \leq \varsigma^{\tilde{\ell}|\partial\Lambda|} (10mMD_{\text{loc}})^{2M + 2\tilde{\ell} + (2kmM/\tilde{\ell})}, \end{aligned} \quad (\text{C50})$$

where we use  $(\frac{mM}{\tilde{\ell}})^2 \leq (mM)^{2\tilde{\ell}}$ . This completes the proof.  $\blacksquare$

In considering one-dimensional systems in Lemma 17 with  $\tilde{\ell} = \ell$ , we have

$$\begin{aligned} \text{SR}(\hat{\mathcal{G}}_{m,M}) &\leq \varsigma^\ell (10mM\varsigma^k)^{2M + 2\ell + (2kmM/\ell)} \\ &\leq (10mM\varsigma^k)^{2M + 3\ell + (2kmM/\ell)} \end{aligned} \quad (\text{C51})$$

because of  $|\partial\Lambda| = 1$  and  $D_{\text{loc}} \leq \varsigma^k$  [see also the inequality (A4)]. By applying the above inequality to  $\tilde{\rho}_0$  in Eq. (C28)

with  $M = 2q$  [Eq. (C7)],  $m = \tilde{c}_2' \sqrt{\nu\beta_0\tilde{\ell}}$  [Eq. (C42)], and  $\ell = \tilde{c}_1\nu\beta_0 = \tilde{c}_1\nu\beta/(2q)$  [Eq. (C25)], we obtain

$$\begin{aligned} \text{SR}(\tilde{\rho}_0^{2q}) &\leq (20mq\varsigma^k)^{4q + 3\tilde{c}_1\nu\beta/(2q) + 4\tilde{c}_2'kq\sqrt{\nu\beta_0/\tilde{\ell}}} \\ &= (20mq\varsigma^k)^{(4 + 4k\tilde{c}_2'/\sqrt{\tilde{c}_1})q + (3/2)\tilde{c}_1\nu/q}, \end{aligned} \quad (\text{C52})$$

Now, we specify the choice of  $q$  by solving for

$$q^2 = \beta\nu = \beta \max \left( \log(2/\delta), \frac{\beta}{2q} \right), \quad (\text{C53})$$

where we use the definition of  $\nu$  in Eq. (C8). This choice gives the result of

$$q \propto \max \{ \beta^{2/3}, [\beta \log(2/\delta)]^{1/2} \}, \quad (\text{C54})$$

where we choose  $q$  appropriately so that the condition (C9) may be satisfied (i.e.,  $\beta \leq q^2$ ). From  $\delta = \epsilon/(3eq) = \mathcal{O}(\epsilon/\beta)$ , by applying the notation of  $q_\epsilon^*$  in Eq. (51) to Eq. (C52), we finally obtain

$$\text{SR}(\tilde{\rho}_0^{2q}) \leq e^{q_\epsilon^* \log(q_\epsilon^*)}. \quad (\text{C55})$$

This completes the proof of Proposition 9.  $\blacksquare$

## APPENDIX D: PROOF OF THEOREM 1 IN HIGH-DIMENSIONAL CASES

We here prove the improved thermal area law for high-dimensional Hamiltonians (A5).

### 1. Restatement

For the convenience of the reader, we restate the statement in the form of the following theorem.

*Theorem 19.*—Let us consider a  $d$ -dimensional lattice and a vertical cut of the total system:  $\Lambda = L \sqcup R$  with  $L = \Lambda_1 \sqcup \Lambda_2 \sqcup \dots \sqcup \Lambda_i$  and  $R = \Lambda_{i+1} \sqcup \Lambda_{i+2} \sqcup \dots \sqcup \Lambda_\Lambda$ , where we use the notation in Eq. (A6). Then, we obtain the improved area law for the mutual information as follows:

$$I(L:R)_{\rho_\beta} \leq C |\partial\Lambda| \beta^{2/3} \log^{2/3}(\beta |\partial\Lambda|), \quad (\text{D1})$$

where  $C$  is a constant which depends on  $k, g, \varsigma$ , and  $d$ .

*Remark.*—The above upper bound is qualitatively better than the established thermal area law of  $I(L:R)_{\rho_\beta} \lesssim \beta |\partial\Lambda|$  for  $\beta \gtrsim \log^2(|\partial\Lambda|)$ . For simplicity, we here consider a vertical cut of the total system, but the generalization to a rectangular cut is straightforward.

We notice that the logarithmic correction originates from the superexponential dependence of  $m$  in Lemma 17. If we can improve the  $m$  independence in Lemma 17,

$$(10mMD_{\text{loc}})^{2M + 2\tilde{\ell} + (2kmM/\tilde{\ell})} \rightarrow (\text{const})^{2M + 2\tilde{\ell} + (2kmM/\tilde{\ell})},$$

we can prove the improved area law in the form of  $I(L:R)_{\rho_\beta} \leq C|\partial\Lambda|\beta^{2/3}$ .

## 2. High-level overview

We, in the following, restrict ourselves to the inverse temperature such that

$$\beta \geq \log^2(|\partial\Lambda|), \quad (\text{D2})$$

since the regime of  $\beta < \log^2(|\partial\Lambda|)$  in Eq. (D1) is already covered by the previous thermal area law [28].

The proof strategy is very close to that in the one-dimensional case. We here relabel each of the sites such that  $L = \{\Lambda_i\}_{i \leq \ell/2}$  and  $R = \{\Lambda_i\}_{i \geq \ell/2+1}$  [see Eq. (A6) for the definition of  $\Lambda_i$ ], where the length  $\ell$  is an integer which is a multiple of 4 to be chosen later. We then decompose the total system into three pieces  $L_0$ ,  $S$ , and  $R_0$  (see Fig. 5), where  $L_0 = \{\Lambda_i\}_{i \leq 0}$ ,  $S = \{\Lambda_i\}_{1 \leq i \leq \ell}$ , and  $R_0 = \{\Lambda_i\}_{i \geq \ell+1}$ . Accordingly, we also decompose the Hamiltonian as follows:

$$\begin{aligned} H &= H_S + H_{L_0} + H_{R_0} + v_0 + v_\ell, & H_S &:= \sum_{X:XC_S} h_X, \\ H_{L_0} &:= \sum_{X:XC_{L_0}} h_X, & H_{R_0} &:= \sum_{X:XC_{R_0}} h_X, \end{aligned} \quad (\text{D3})$$

where  $v_i$  is defined in Eq. (A7). We note that  $H_S$ ,  $H_{L_0}$ , and  $H_{R_0}$  commute with each other. As in the one-dimensional case, by shifting the energy origin appropriately, we set

$$H_S \succeq 0, \quad (\text{D4})$$

where  $\succeq$  means that  $H_S$  is positive semidefinite. We divide  $\beta$  into  $2q$  pieces ( $q \in \mathbb{N}$ ) and introduce

$$\rho_0 := e^{-\beta_0 H}, \quad \beta_0 := \beta/(2q). \quad (\text{D5})$$

The first difference from the one-dimensional case is that we cannot derive Lemma 14 as in the case of 1D, since we cannot utilize the belief propagation technique [29] in high-dimensional systems. In high-dimensional cases, the operator  $\phi(\tau)$  in Eq. (C14) has the norm of  $\mathcal{O}(\beta_0|\partial\Lambda|)$ , while in the one-dimensional case, it has the norm of  $\mathcal{O}(\beta_0)$ . This fact reduces the approximation error in Eq. (C12) to  $e^{-\mathcal{O}(\ell/\beta_0) + \mathcal{O}(\beta_0|\partial\Lambda|)} \|\rho_0\|_{\rho_0}$  in high-dimensional systems. Hence, we need to choose  $\ell = \mathcal{O}(\beta_0^2|\partial\Lambda|)$  to ensure a good approximation error, but this choice is too large to be utilized in the derivation of the improved thermal area law.

In order to overcome this difficulty, we choose  $q = \mathcal{O}(\beta)$  such that

$$\beta_0 := \beta/(2q) \leq \frac{1}{32gk}. \quad (\text{D6})$$

As shown in Lemma 20 below, this condition allows us to construct the operator  $\tilde{\rho}_0$  as in Eq. (C5) (i.e.,  $\|\rho_0 - \tilde{\rho}_0\|_{2qp} \leq \delta\|\rho_0\|_{2qp}$ ) in the following form:

$$\begin{aligned} \tilde{\rho}_0 &:= \tilde{\Phi}_0 e^{-\beta_0(H_{L_0} + H_{R_0})} \mathcal{F}_m(\beta_0 H_S), \\ \tilde{\Phi}_0 &:= \Phi_{L_1} \otimes \Phi_{R_1}, \end{aligned} \quad (\text{D7})$$

where  $\Phi_{L_1}$  and  $\Phi_{R_1}$  are operators supported on  $L_1$  and  $R_1$ , respectively (i.e.,  $L_1 = \{\Lambda_i\}_{i \leq \ell/4}$  and  $R_1 = \{\Lambda_i\}_{i \geq 3\ell/4+1}$ ), and the degree  $m$  polynomial  $\mathcal{F}_m(x)$  approximates the exponential function  $e^{-\beta_0 x}$ . As in the inequality (C6), this operator gives the approximation

$$\|e^{-\beta H} - \tilde{\rho}_0^{2q}\|_p \leq \epsilon \|e^{-\beta H}\|_p \quad \text{with} \quad \epsilon := 3eq\delta. \quad (\text{D8})$$

The mutual information is roughly determined by the upper bound of the Schmidt rank of  $\tilde{\rho}_0^{2q}$ , which is given by

$$[\tilde{\Phi}_0 e^{-\beta_0(H_{L_0} + H_{R_0})} \mathcal{F}_m(\beta_0 H_S)]^{2q}. \quad (\text{D9})$$

The Schmidt rank of the above operator is  $(mq)^{\mathcal{O}(q) + \mathcal{O}(\ell|\partial\Lambda|) + \mathcal{O}(mq/\ell)}$  from Lemma 17. In the one-dimensional case, for  $q = \mathcal{O}(\beta)$ , this estimation gives the Schmidt rank of  $e^{\mathcal{O}(\beta)}$  and spoils the improved thermal area law. However, in high-dimensional systems, the contribution of  $e^{\mathcal{O}(\beta)}$  is much smaller than  $e^{\beta^{2/3}|\partial\Lambda|}$  as long as  $\beta \leq |\partial\Lambda|^3$ . Therefore, it is still possible to derive an improved area law from the approximation by Eq. (D9). This point is the second difference between the 1D case and high-dimensional cases.

In the proof of the area law, we roughly choose (see Appendix D4 for more details)

$$m \approx |\partial\Lambda|\sqrt{\ell}, \quad q \approx \beta, \quad \ell \approx \beta^{2/3}, \quad (\text{D10})$$

which gives the Schmidt rank of the operator (D9) as  $(mq)^{\mathcal{O}(q) + \mathcal{O}(\ell|\partial\Lambda|) + \mathcal{O}(mq/\ell)} \approx \exp[\beta^{2/3}|\partial\Lambda| \log(\beta|\partial\Lambda|)]$ . We thus obtain the inequality (D1).

In the following, we show how the basic lemmas in the one-dimensional case are extended to the high-dimensional cases.

## 3. Approximation of $\rho_0$ of Eq. (D7)

We relate the two operators  $\rho_0 = e^{-\beta_0 H}$  and  $e^{-\beta_0 H_0}$ . We can formally write

$$\rho_0 = \Phi_0 e^{-\beta_0 H_0}, \quad (\text{D11})$$

where  $\Phi_0 = e^{-\beta_0 H} e^{\beta_0 H_0}$  is usually a highly nonlocal operator. The lemma below ensures that the  $\Phi_0$  is approximated by an operator supported on  $L_1 \sqcup R_1$ .

*Lemma 20.*—Let  $\tilde{H}$  and  $\tilde{H}_0$  be Hamiltonians as follows:

$$\begin{aligned}\tilde{H} &= H_{L_1} + H_{R_1}, \\ \tilde{H}_0 &= H_{L_1} + H_{R_1} - v_0 - v_\ell.\end{aligned}\quad (\text{D12})$$

We then define

$$\begin{aligned}\tilde{\Phi}_0 &:= \tilde{\Phi}_{L_1} \otimes \tilde{\Phi}_{R_1} = e^{-\beta_0 \tilde{H}} e^{\beta_0 \tilde{H}_0}, \\ \tilde{\Phi}_{L_1} &:= e^{-\beta_0 H_{L_1}} e^{\beta_0 (H_{L_1} - v_0)}, \\ \tilde{\Phi}_{R_1} &:= e^{-\beta_0 H_{R_1}} e^{\beta_0 (H_{R_1} - v_\ell)}.\end{aligned}\quad (\text{D13})$$

Then, for  $\beta_0 \leq 1/(32gk)$  and  $\ell \geq 2k \log(|\partial\Lambda|)$ , the approximated operator

$$\rho'_0 = \tilde{\Phi}_0 \rho_0 = \tilde{\Phi}_0 e^{-\beta_0 H_0} \quad (\text{D14})$$

satisfies

$$\|\rho'_0 - \rho_0\|_{p_0} \leq 2|\partial\Lambda| e^{-\ell/(2k)} \|\rho_0\|_{p_0} \quad (\text{D15})$$

for arbitrary  $p_0 \in \mathbb{N}$ .

As mentioned in the previous subsection, this decomposition has an advantage over the belief propagation method used in Lemma 14. By using the decomposition

(D14), we can achieve the upper bound of  $|\partial\Lambda| e^{-\mathcal{O}(\ell)}$  instead of  $e^{-\mathcal{O}(\ell) + \mathcal{O}(|\partial\Lambda|)}$ .

### a. Proof of Lemma 20

We first define  $V$  as  $V := H - H_0 = \tilde{H} - \tilde{H}_0$ , namely,

$$V = v_0 + v_\ell. \quad (\text{D16})$$

We also define  $V' := H_0 - \tilde{H}_0$ .

We here aim to prove

$$\|e^{-\beta_0 \tilde{H}} e^{\beta_0 \tilde{H}_0} e^{-\beta_0 H_0} e^{\beta_0 H} - 1\| \leq 2|\partial\Lambda| e^{-\ell/(2k)}. \quad (\text{D17})$$

When we obtain the above upper bound, we arrive at the main inequality as follows:

$$\begin{aligned}\|\tilde{\Phi}_0 \rho'_0 - \rho_0\| &= \|e^{-\beta_0 \tilde{H}} e^{\beta_0 \tilde{H}_0} e^{-\beta_0 H_0} - e^{-\beta_0 H}\|_p \\ &= \|e^{-\beta_0 \tilde{H}} e^{\beta_0 \tilde{H}_0} e^{-\beta_0 H_0} e^{\beta_0 H} e^{-\beta_0 H} - e^{-\beta_0 H}\|_p \\ &\leq \|e^{-\beta_0 \tilde{H}} e^{\beta_0 \tilde{H}_0} e^{-\beta_0 H_0} e^{\beta_0 H} - 1\| \cdot \|e^{-\beta_0 H}\|_p \\ &\leq 2|\partial\Lambda| e^{-\ell/(2k)} \|e^{-\beta_0 H}\|_p,\end{aligned}\quad (\text{D18})$$

where the first inequality comes from Hölder's inequality (B2).

In order to derive the inequality (D17), we define  $\mathcal{G}(\tau)$  as

$$\mathcal{G}(\tau) := e^{-\tau \tilde{H}} e^{\tau \tilde{H}_0} e^{-\tau H_0} e^{\tau H}. \quad (\text{D19})$$

We then obtain

$$\begin{aligned}\frac{d}{d\tau} \mathcal{G}(\tau) &= e^{-\tau \tilde{H}} (\tilde{H}_0 - \tilde{H}) e^{\tau \tilde{H}} \mathcal{G}(\tau) + e^{-\tau \tilde{H}} e^{\tau \tilde{H}_0} e^{-\tau H_0} (H - H_0) e^{\tau H_0} e^{-\tau \tilde{H}_0} e^{\tau \tilde{H}} \mathcal{G}(\tau) \\ &= -e^{-\tau \tilde{H}} (V - e^{\tau \tilde{H}_0} e^{-\tau H_0} V e^{\tau H_0} e^{-\tau \tilde{H}_0}) e^{\tau \tilde{H}} \mathcal{G}(\tau),\end{aligned}\quad (\text{D20})$$

where we use the definition of  $H - H_0 = \tilde{H} - \tilde{H}_0 = V$ . The solution of the above differential equation is given by

$$\mathcal{G}(\beta_0) = \mathcal{T} \exp \left( - \int_0^{\beta_0} e^{-\tau \tilde{H}} (V - e^{\tau \tilde{H}_0} e^{-\tau H_0} V e^{\tau H_0} e^{-\tau \tilde{H}_0}) e^{\tau \tilde{H}} d\tau \right), \quad (\text{D21})$$

where  $\mathcal{T}$  is the ordering operator. From the above equation, we obtain the upper bound of  $\|e^{-\beta_0 \tilde{H}} e^{\beta_0 \tilde{H}_0} e^{-\beta_0 H_0} e^{\beta_0 H} - 1\| = \|\mathcal{G}(\beta_0) - 1\|$  as

$$\|\mathcal{G}(\beta_0) - 1\| \leq \exp \left( \int_0^{\beta_0} \|e^{-\tau \tilde{H}} (V - e^{\tau \tilde{H}_0} e^{-\tau H_0} V e^{\tau H_0} e^{-\tau \tilde{H}_0}) e^{\tau \tilde{H}}\| d\tau \right) - 1. \quad (\text{D22})$$

We can prove the following upper bound (see below for the proof):

$$\|e^{-\tau \tilde{H}} (V - e^{\tau \tilde{H}_0} e^{-\tau H_0} V e^{\tau H_0} e^{-\tau \tilde{H}_0}) e^{\tau \tilde{H}}\| \leq \frac{gk|\partial\Lambda|}{2} 8^{-s_\ell}, \quad s_\ell := \frac{\ell}{4k} - 2. \quad (\text{D23})$$

For  $\ell \geq 2k \log(|\partial\Lambda|)$ , we have

$$\frac{gk\beta|\partial\Lambda|}{2}8^{-s_\ell} \leq |\partial\Lambda|8^{-\ell/(4k)} \leq |\partial\Lambda|e^{-\ell/(2k)} \leq 1, \quad (\text{D24})$$

where we use  $\beta_0 \leq 1/(32gk)$  and  $e^2 < 8$ . We thus use the inequality (D23) to reduce the inequality (D22) to

$$\|\mathcal{G}(\beta_0) - 1\| \leq \exp\left(\frac{gk\beta_0|\partial\Lambda|}{2}8^{-s_\ell}\right) - 1 \leq gk\beta_0|\partial\Lambda|8^{-s_\ell} \leq 2|\partial\Lambda|8^{-\ell/(4k)} \leq 2|\partial\Lambda|e^{-\ell/(2k)}, \quad (\text{D25})$$

where we use  $e^x - 1 \leq 2x$  for  $0 \leq x \leq 1$  and  $gk\beta_0 \leq 1/32$ . This completes the proof.  $\blacksquare$

### b. Proof of the inequality (D23)

We start from the following equation:

$$\begin{aligned} V - e^{\tau\tilde{H}_0}e^{-\tau H_0}Ve^{\tau H_0}e^{-\tau\tilde{H}_0} &= -\int_0^\tau \frac{d}{dx}(e^{x\tilde{H}_0}e^{-xH_0}Ve^{xH_0}e^{-x\tilde{H}_0})dx \\ &= -\int_0^\tau (e^{x\tilde{H}_0}e^{-xH_0}[e^{xH_0}\tilde{H}_0e^{-xH_0} - H_0, V]e^{xH_0}e^{-x\tilde{H}_0})dx \\ &= \int_0^\tau (e^{x\tilde{H}_0}e^{-xH_0}[e^{xH_0}V'e^{-xH_0}, V]e^{xH_0}e^{-x\tilde{H}_0})dx, \end{aligned} \quad (\text{D26})$$

where we use the definition  $V' := H_0 - \tilde{H}_0$ . This result yields

$$\begin{aligned} \|e^{-\tau\tilde{H}}(V - e^{\tau\tilde{H}_0}e^{-\tau H_0}Ve^{\tau H_0}e^{-\tau\tilde{H}_0})e^{\tau\tilde{H}}\| &\leq \int_0^\tau \|e^{-\tau\tilde{H}}e^{x\tilde{H}_0}e^{-xH_0}[e^{xH_0}V'e^{-xH_0}, V]e^{xH_0}e^{-x\tilde{H}_0}e^{\tau\tilde{H}}\|dx \\ &= \int_0^\tau \|e^{-\tau\tilde{H}}e^{x\tilde{H}_0}[V', e^{-xH_0}Ve^{xH_0}]e^{-x\tilde{H}_0}e^{\tau\tilde{H}}\|dx. \end{aligned} \quad (\text{D27})$$

The commutator  $[V', e^{-xH_0}Ve^{xH_0}]$  is decomposed by the Baker-Campbell-Hausdorff expansion:

$$[V', e^{-xH_0}Ve^{xH_0}] = \sum_{s=0}^{\infty} \frac{(-x)^s}{s!} \text{ad}_{V'} \text{ad}_{H_0}^s(V). \quad (\text{D28})$$

Because the supports of  $V$  and  $V'$  are separated at least by a distance of  $\ell/4 - 2k$ , we have

$$\text{ad}_{V'} \text{ad}_{H_0}^s(V) = 0 \quad \text{for } s \leq \frac{\ell/4 - 2k}{k} =: s_\ell. \quad (\text{D29})$$

Furthermore, we have

$$e^{-\tau\tilde{H}}e^{x\tilde{H}_0}[V', e^{-xH_0}Ve^{xH_0}]e^{-x\tilde{H}_0}e^{\tau\tilde{H}} = \sum_{m_2=0}^{\infty} \frac{(-\tau)^{m_2}}{m_2!} \sum_{m_1=0}^{\infty} \frac{x^{m_1}}{m_1!} \sum_{s=s_\ell+1}^{\infty} \frac{(-x)^s}{s!} \text{ad}_{\tilde{H}}^{m_2} \text{ad}_{H_0}^{m_1} \text{ad}_{V'} \text{ad}_{H_0}^s(V). \quad (\text{D30})$$

From Lemma 13 or inequality (B18), the norm of the multicommutator is upper bounded by

$$\|\text{ad}_{\tilde{H}}^{m_2} \text{ad}_{H_0}^{m_1} \text{ad}_{V'} \text{ad}_{H_0}^s(h_X)\| \leq (2gk)^{m_1+m_2+s+1} (m_1 + m_2 + s + 1)! \|h_X\|, \quad (\text{D31})$$

where  $h_X$  is supported on  $X$  such that  $|X| \leq k$ . Then, because of the definition of  $v_i$  in Eq. (A7), we have

$$\begin{aligned} \|\text{ad}_{\tilde{H}}^{m_2} \text{ad}_{H_0}^{m_1} \text{ad}_{V'} \text{ad}_{H_0}^s(v_i)\| &\leq (2gk)^{m_1+m_2+s+1} (m_1 + m_2 + s + 1)! \sum_{X: X \cap \Lambda_{\leq i} \neq \emptyset, X \cap \Lambda_{> i} \neq \emptyset} \|h_X\| \\ &\leq (2gk)^{m_1+m_2+s+1} (m_1 + m_2 + s + 1)! \sum_{j \in \Lambda_i \cup \Lambda_{i-1} \cup \dots \cup \Lambda_{i-k+1}} \sum_{X: X \ni j} \|h_X\| \\ &\leq gk|\partial\Lambda| (2gk)^{m_1+m_2+s+1} (m_1 + m_2 + s + 1)!, \end{aligned} \quad (\text{D32})$$



where we use  $|\Lambda_i| + |\Lambda_{i-1}| + \dots + |\Lambda_{i-k+1}| \leq k|\partial\Lambda|$ . Because of  $V = v_0 + v_\ell$ , we obtain

$$\|\text{ad}_{\tilde{H}}^{m_2} \text{ad}_{\tilde{H}_0}^{m_1} \text{ad}_{V'} \text{ad}_{H_0}^s(V)\| \leq |\partial\Lambda| (2gk)^{m_1+m_2+s+2} (m_1+m_2+s+1)!. \quad (\text{D33})$$

By using the inequality

$$\frac{(m_1+m_2+s+1)!}{m_1!m_2!s!} = \frac{(m_1+1)!}{m_1!1!} \frac{(m_1+m_2+1)!}{(m_1+1)!m_2!} \frac{(m_1+m_2+s+1)!}{(m_1+m_2+1)!m_2!} \leq 8^{m_1+1} 4^{m_2} 2^s, \quad (\text{D34})$$

we have

$$\begin{aligned} & \sum_{m_2=0}^{\infty} \frac{\tau^{m_2}}{m_2!} \sum_{m_1=0}^{\infty} \frac{x^{m_1}}{m_1!} \sum_{s=s_\ell+1}^{\infty} \frac{x^s}{s!} \|\text{ad}_{\tilde{H}}^{m_2} \text{ad}_{\tilde{H}_0}^{m_1} \text{ad}_{V'} \text{ad}_{H_0}^s(V)\| \\ & \leq \sum_{m_2=0}^{\infty} \tau^{m_2} \sum_{m_1=0}^{\infty} x^{m_1} \sum_{s=s_\ell+1}^{\infty} x^s |\partial\Lambda| (2gk)^{m_1+m_2+s+2} \frac{(m_1+m_2+s+1)!}{m_1!m_2!s!} \\ & \leq \frac{32g^2k^2|\partial\Lambda|}{1-16gkx} \frac{1}{1-8gk\tau} \frac{(4gkx)^{s_\ell+1}}{1-4gkx}. \end{aligned} \quad (\text{D35})$$

From  $\max(\tau, x) \leq \beta_0 \leq 1/(32gk)$ , the above inequality reduces to

$$\sum_{m_2=0}^{\infty} \frac{\tau^{m_2}}{m_2!} \sum_{m_1=0}^{\infty} \frac{x^{m_1}}{m_1!} \sum_{s=s_\ell+1}^{\infty} \frac{x^s}{s!} \|\text{ad}_{\tilde{H}}^{m_2} \text{ad}_{\tilde{H}_0}^{m_1} \text{ad}_{V'} \text{ad}_{H_0}^s(V)\| \leq 13g^2k^2|\partial\Lambda|8^{-s_\ell}. \quad (\text{D36})$$

By combining the inequalities (D30) and (D36), we obtain

$$\|e^{-\tau\tilde{H}} e^{x\tilde{H}_0} [V', e^{-xH_0} V e^{xH_0}] e^{-x\tilde{H}_0} e^{\tau\tilde{H}}\| \leq 13g^2k^2|\partial\Lambda|8^{-s_\ell}. \quad (\text{D37})$$

From the inequality (D27), we thus prove the inequality of

$$\|e^{-\tau\tilde{H}} (V - e^{\tau\tilde{H}_0} e^{-\tau H_0} V e^{\tau H_0} e^{-\tau\tilde{H}_0}) e^{\tau\tilde{H}}\| \leq 13g^2k^2\tau|\partial\Lambda|8^{-s_\ell} \leq \frac{gk|\partial\Lambda|}{2} 8^{-s_\ell}, \quad (\text{D38})$$

where we use  $\tau \leq \beta_0 \leq 1/(32gk)$  in the second inequality. This completes the proof. ■

The lemma implies that, as the length  $\ell$  becomes large, the approximation error decays exponentially with  $e^{-\mathcal{O}(\ell)}$ . Thus, in order to achieve the inequality

$$\|\rho'_0 - \rho_0\|_{2qp} \leq \frac{\delta}{2} \|\rho_0\|_{2qp}, \quad (\text{D39})$$

we need to choose  $\ell$  as

$$\ell \geq 2k \log(4|\partial\Lambda|/\delta). \quad (\text{D40})$$

We approximate  $e^{-\beta_0 H_0}$  by an operator with small Schmidt rank. For this purpose, we use the fact that  $H_S$ ,  $H_{L_0}$ , and  $H_{R_0}$  commute with each other and write  $e^{-\beta_0 H_0} = e^{-\beta_0(H_{L_0}+H_{R_0})} e^{-\beta_0 H_S}$ . Then, we approximate  $e^{-\beta_0 H_S}$  by using the polynomial of  $H_S$  in Lemma 15. By using the polynomial  $\mathcal{F}_m(x)$  defined there, we approximate the operator  $\rho'_0$  in Eq. (D14) by

$$\tilde{\rho}_0 := \tilde{\Phi}_0 e^{-\beta_0(H_{L_0}+H_{R_0})} \mathcal{F}_m(\beta_0 H_S). \quad (\text{D41})$$

Because of Eq. (C3), the spectrum of  $\beta_0 H_S$  is included in the span of  $[0, \beta_0 \|H_S\|]$ , and, hence, the inequality (C27) gives

$$\|\mathcal{F}_m(\beta_0 H_S) - e^{-\beta_0 H_S}\| \leq \delta_0 \quad (\text{D42})$$

by choosing  $m$  appropriately following Lemma 15. The next problem is to estimate the approximation error  $\|\rho'_0 - \tilde{\rho}_0\|_{p_0}$  for an arbitrary Schatten  $p_0$ -norm. We prove the following lemma, which is similar to Lemma 16.

*Lemma 21.*—Let  $p_0 \in \mathbb{N}$  and  $\delta_0 \in (0, 1)$ . Under the choice of  $\Phi_{L_1} \otimes \Phi_{R_1}$  in Lemma 14,  $\ell$  in Eq. (D40), and  $m, \mathcal{F}_m(x)$  in Lemma 15, we have

$$\|\rho'_0 - \tilde{\rho}_0\|_{p_0} \leq \mathcal{D}_S^{1/p_0} \delta_0 e^{|\partial\Lambda|/7} \|\rho_0\|_{p_0} \quad (\text{D43})$$

for  $\beta_0 \leq 1/(32gk)$ .

### c. Proof of Lemma 21

From the definitions (D14) and (D41) of  $\rho'_0$  and  $\tilde{\rho}_0$ , respectively, we start from the inequality of

$$\begin{aligned} \|\rho'_0 - \tilde{\rho}_0\|_{p_0} &= \|\tilde{\Phi}_0 e^{-\beta_0(H_{L_0} + H_{R_0})} [\mathcal{F}_m(H_S) - e^{-\beta_0 H_S}]\|_{p_0} \\ &\leq \|e^{-\beta_0(H_{L_0} + H_{R_0})} [\mathcal{F}_m(H_S) - e^{-\beta_0 H_S}]\|_{p_0} \cdot \|\tilde{\Phi}_0\|, \end{aligned} \quad (\text{D44})$$

where we use Hölder's inequality (B2) in the second line. From the definition (D13) of  $\tilde{\Phi}_0$ , we obtain the following upper bound (see below for the proof):

$$\|\tilde{\Phi}_0\| \leq e^{|\partial\Lambda|/15}. \quad (\text{D45})$$

Next, we obtain the same inequality as Eq. (C36), which gives the upper bound of

$$\begin{aligned} \|e^{-\beta_0(H_{L_0} + H_{R_0})} [\mathcal{F}_m(H_S) - e^{-\beta_0 H_S}]\|_{p_0}^{p_0} \\ \leq \mathcal{D}_S \delta_0^{p_0} \|e^{-\beta_0(H_{L_0} + H_{R_0} + H_S)}\|_{p_0}^{p_0}. \end{aligned} \quad (\text{D46})$$

In order to estimate the upper bound of  $\|e^{-\beta_0(H_{L_0} + H_{R_0} + H_S)}\|_{p_0}$  in terms of  $\|e^{-\beta_0 H}\|_{p_0}$ , we use the Golden-Thompson inequality to derive

$$\begin{aligned} \text{tr}(e^{-p_0 \beta_0 H_0}) &\leq \text{tr}(e^{-p_0 \beta_0 (v_0 + v_\epsilon)} \cdot e^{-p_0 \beta_0 H}) \\ &\leq e^{2p_0 \beta_0 gk |\partial\Lambda|} \|e^{-\beta_0 H}\|_{p_0}^{p_0} \\ &\leq e^{p_0 |\partial\Lambda|/16} \|e^{-\beta_0 H}\|_{p_0}^{p_0}, \end{aligned} \quad (\text{D47})$$

where we use  $\text{tr}(e^{-p_0 \beta_0 H_0}) = \|e^{-\beta_0 H_0}\|_{p_0}^{p_0}$ ,  $\beta_0 \leq 1/(32gk)$ , and derive the upper bound of  $\|v_i\|$  from the definition (A7):

$$\begin{aligned} \|v_i\| &\leq \sum_{X: X \cap \Lambda_{\leq i} \neq \emptyset, X \cap \Lambda_{> i} \neq \emptyset} \|h_X\| \\ &\leq \sum_{j \in \Lambda_i \sqcup \Lambda_{i-1} \sqcup \dots \sqcup \Lambda_{i-k+1}} \sum_{X: X \ni j} \|h_X\| \leq gk |\partial\Lambda|. \end{aligned}$$

By combining the inequalities (D46) and (D47), we arrive at the inequality of

$$\begin{aligned} \|e^{-\beta_0(H_{L_0} + H_{R_0})} [\mathcal{F}_m(H_S) - e^{-\beta_0 H_S}]\|_{p_0}^{p_0} \\ \leq \mathcal{D}_S \delta_0^{p_0} e^{p_0 |\partial\Lambda|/16} \|e^{-\beta_0 H}\|_{p_0}^{p_0}. \end{aligned} \quad (\text{D48})$$

By applying the inequalities (D45) and (D48) to Eq. (D44), we obtain the main inequality (C30). This completes the proof.  $\blacksquare$

### d. Proof of the inequality (D45)

By using Eq. (D13), we have

$$\|\tilde{\Phi}_0\| \leq \|e^{-\beta_0 H_{L_1}} e^{\beta_0 (H_{L_1} - v_0)}\| \cdot \|e^{-\beta_0 H_{R_1}} e^{\beta_0 (H_{R_1} - v_\epsilon)}\|.$$

We here consider

$$e^{-\beta_0 H_{L_1}} e^{\beta_0 (H_{L_1} - v_0)} = \mathcal{T} \exp \left( - \int_0^{\beta_0} e^{-x H_{L_1}} v_0 e^{x H_{L_1}} dx \right),$$

which gives rise to the inequality of

$$\|e^{-\beta_0 H_{L_1}} e^{\beta_0 (H_{L_1} - v_0)}\| \leq \exp \left( \int_0^{\beta_0} \|e^{-x H_{L_1}} v_0 e^{x H_{L_1}}\| dx \right). \quad (\text{D49})$$

We thus aim to derive the upper bound of  $\|e^{-x H_{L_1}} v_0 e^{x H_{L_1}}\|$ .

By using the Baker-Campbell-Hausdorff expansion, we have

$$\|e^{-x H_{L_1}} v_0 e^{x H_{L_1}}\| \leq \sum_{m=0}^{\infty} \frac{x^m}{m!} \|\text{ad}_{H_{L_1}}^m(v_0)\|. \quad (\text{D50})$$

By using Lemma 13 or the inequality (B18), the norm of  $\|\text{ad}_{H_{L_1}}^m(v_0)\|$  is upper bounded as follows:

$$\|\text{ad}_{H_{L_1}}^m(v_0)\| \leq gk |\partial\Lambda| (2gk)^m m!, \quad (\text{D51})$$

where we use an analysis similar to Eq. (D32). Hence, we calculate the upper bound of  $\|e^{-x H_{L_1}} v_0 e^{x H_{L_1}}\|$  as

$$\begin{aligned} \|e^{-x H_{L_1}} v_0 e^{x H_{L_1}}\| &\leq \sum_{m=0}^{\infty} \frac{x^m}{m!} \cdot gk |\partial\Lambda| (2gk)^m m! \\ &= \frac{gk |\partial\Lambda|}{1 - 2gkx} \leq \frac{16gk}{15} |\partial\Lambda|, \end{aligned} \quad (\text{D52})$$

where we use  $x \leq \beta_0 \leq 1/(32gk)$ . By applying this inequality to Eq. (D49), we have

$$\|e^{-\beta_0 H_{L_1}} e^{\beta_0 (H_{L_1} - v_0)}\| \leq e^{(16gk\beta_0/15)|\partial\Lambda|} \leq e^{(1/30)|\partial\Lambda|}. \quad (\text{D53})$$

We obtain the same inequality for  $\|e^{-\beta_0 H_{R_1}} e^{\beta_0 (H_{R_1} - v_\epsilon)}\|$ . This completes the proof.  $\blacksquare$

Let us substitute  $p_0 = 2qp$  in Lemma 21 and choose  $\delta_0$  such that it satisfies

$$\mathcal{D}_S^{1/(2qp)} \delta_0 e^{|\partial\Lambda|/7} \leq \frac{\delta}{2}. \quad (\text{D54})$$

This choice ensures that  $\|\rho'_0 - \tilde{\rho}_0\|_{2qp} \leq (\delta/2) \|\rho_0\|_{2qp}$ , and we conclude that

$$\begin{aligned} \|\rho_0 - \tilde{\rho}_0\|_{2qp} &\leq \|\rho_0 - \rho'_0\|_{2qp} + \|\rho'_0 - \tilde{\rho}_0\|_{2qp} \\ &\leq \delta \|\rho_0\|_{2qp}, \end{aligned} \quad (\text{D55})$$

where we use the inequality (D39).

Let us simplify the expression for all the parameters appearing so far. We first consider

$$\mathcal{D}_S = \varsigma^{|S|} \leq e^{\ell|\partial\Lambda|\log(\varsigma)}, \quad (\text{D56})$$

and, hence, from Eq. (D54) with  $q = \beta/(2\beta_0)$  and  $p \geq 1$ , we can choose  $\delta_0$  as

$$\begin{aligned} \log(1/\delta_0) &= \log(2/\delta) + |\partial\Lambda| \left( \frac{1}{7} + \beta_0 \log(\varsigma) \frac{\ell}{\beta} \right) \\ &= |\partial\Lambda| \left( \frac{1}{7} + \beta_0 \log(\varsigma) \frac{\ell}{\beta} + \frac{\log(2/\delta)}{|\partial\Lambda|} \right) \\ &= \nu' |\partial\Lambda| \end{aligned} \quad (\text{D57})$$

with

$$\nu' := \frac{1}{7} + \beta_0 \log(\varsigma) \frac{\ell}{\beta} + \frac{\log(2/\delta)}{|\partial\Lambda|} \leq \text{const} \times \ell. \quad (\text{D58})$$

Also, the norm of the Hamiltonian  $\beta_0 H_S$  is bounded from above by

$$\beta_0 \|H_S\| \leq \beta_0 g |S| \leq \beta_0 g \ell |\partial\Lambda|, \quad (\text{D59})$$

where we use the definition (A6) of  $|\partial\Lambda|$ . Because of the above upper bound, we have

$$\begin{aligned} \max[\beta_0 \|H_S\|, \log(1/\delta_0)] &\leq |\partial\Lambda| \max[\beta_0 g \ell, \nu'] \\ &= \mathcal{O}(\ell |\partial\Lambda|). \end{aligned} \quad (\text{D60})$$

Hence, from the inequality (C26) in Lemma 15, we obtain the following form of  $m$  to achieve the inequality (D42):

$$\begin{aligned} m &= \lceil c_f \sqrt{\max[\beta_0 \|H_S\|, \log(1/\delta_0)] \log(1/\delta_0)} \rceil \\ &= \tilde{c} |\partial\Lambda| \sqrt{\nu' \ell}, \end{aligned} \quad (\text{D61})$$

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

Finally, we apply Lemma 17 to  $\tilde{\rho}_0^{2q}$ . We have  $D_{\text{loc}} \leq \varsigma^k |\partial\Lambda|$ , and, hence,

$$\text{SR}(\hat{\mathcal{G}}_{m,M}) \leq \min_{\tilde{\ell}: \tilde{\ell} \leq \ell} [\varsigma^{\tilde{\ell}|\partial\Lambda} (10mM\varsigma^k |\partial\Lambda|)^{2M+2\tilde{\ell}+(2kmM/\tilde{\ell})}]. \quad (\text{D62})$$

Under the choice of

$$\begin{aligned} \hat{\mathcal{G}}_{m,M} &= \tilde{\rho}_0^{2q}, \quad M = 2q = (\beta/\beta_0), \quad m = \tilde{c} |\partial\Lambda| \sqrt{\nu' \ell}, \\ \ell &\geq 2k \log(4|\partial\Lambda|/\delta) \geq \tilde{c}' \log(|\partial\Lambda|/\delta), \end{aligned} \quad (\text{D63})$$

we reduce the upper bound of Eq. (D62) to

$$\text{SR}(\tilde{\rho}_0^{2q}) \leq \min_{\tilde{\ell}: \tilde{\ell} \leq \ell} \left[ \varsigma^{\tilde{\ell}|\partial\Lambda} \left( \frac{10\tilde{c}\beta\varsigma^k |\partial\Lambda|^2 \sqrt{\nu' \ell}}{\beta_0} \right)^{2\beta/\beta_0 + 2\tilde{\ell} + 2\tilde{c}k(\beta/\beta_0) |\partial\Lambda| \tilde{\ell}^{-1} \sqrt{\nu' \ell}} \right]. \quad (\text{D64})$$

#### 4. Choice of polynomial degree $m$ and region length $\ell$

We here consider how to choose the parameters  $m$  and  $\ell$ . We assume  $|R| \geq |L|$  ( $\geq |\partial\Lambda|$ ) and choose  $\delta$  as  $\delta = 1/|L|^2$ , and the condition for  $\ell$  in Eq. (D63) reads

$$\ell \geq 2k \log(4|\partial\Lambda|/\delta) \geq \tilde{c}_1 \log(|L|), \quad (\text{D65})$$

where  $\tilde{c}_1$  is a constant which depends only on  $g$ ,  $k$ , and  $d$ . Then, under the condition of  $\beta \geq \log^2(|\partial\Lambda|) \propto \log^2(|L|)$ , we can choose  $\ell$  such that

$$\ell \leq \tilde{c}_1' \beta. \quad (\text{D66})$$

We then obtain the upper bound of  $\nu'$  in Eq. (D57) as

$$\nu' = \frac{1}{7} + \tilde{c}_1' \beta_0 \log(\varsigma) + \frac{\log(2|L|^2)}{|\partial\Lambda|} \leq \tilde{c}_2, \quad (\text{D67})$$

where  $\tilde{c}_2$  is a constant which depends only on  $g$ ,  $k$ ,  $d$ , and  $\varsigma$ .

We here denote

$$\frac{10\tilde{c}\beta\varsigma^k |\partial\Lambda|^2 \sqrt{\nu' \ell}}{\beta_0} \leq e^{\tilde{c}_3 \log(\beta|\partial\Lambda|)} \quad (\text{D68})$$

with  $\tilde{c}_3$  an  $\mathcal{O}(1)$  constant.

Then, the upper bound (D64) is simplified as

$$\begin{aligned} \text{SR}(\tilde{\rho}_0^{2q}) &\leq \min_{\tilde{\ell}: \tilde{\ell} \leq \ell} [e^{2\tilde{c}_3(1/\beta_0 + \tilde{c}_1')\beta \log(\beta|\partial\Lambda|)} \cdot e^{\tilde{\ell} \log(\varsigma) |\partial\Lambda| + 2\tilde{c} \sqrt{\tilde{c}_2} k(\beta/\beta_0) \tilde{c}_3 \log(\beta|\partial\Lambda|) \tilde{\ell}^{-1} \ell^{1/2} |\partial\Lambda|}] \\ &= e^{\tilde{c}_4 \beta \log(\beta|\partial\Lambda|)} \min_{\tilde{\ell}: \tilde{\ell} \leq \ell} [e^{\tilde{\ell} \log(\varsigma) |\partial\Lambda| + \tilde{c}_5 \beta \log(\beta|\partial\Lambda|) \tilde{\ell}^{-1} \ell^{1/2} |\partial\Lambda|}], \end{aligned} \quad (\text{D69})$$

where we define  $\tilde{c}_4 := 2\tilde{c}_3(1/\beta_0 + \tilde{c}'_1)$  and  $\tilde{c}_5 := 2\tilde{c}\sqrt{\tilde{c}_2}\tilde{c}_3k/\beta_0$ .

In the above upper bound, we would like to choose

$$\tilde{\ell} = \left\lceil \left( \frac{\tilde{c}_5}{\log(\zeta)} \beta \log(\beta|\partial\Lambda|) \right)^{1/2} \ell^{1/4} \right\rceil. \quad (\text{D70})$$

In order that the choice above is consistent with  $\tilde{\ell} \leq \ell$ , the length  $\ell$  should satisfy

$$\ell \geq \left( \frac{\tilde{c}_5}{\log(\zeta)} \beta \log(\beta|\partial\Lambda|) \right)^{2/3}. \quad (\text{D71})$$

We note that this choice of  $\ell$  exists under the constraints of Eqs. (D65) and (D66) because of  $\beta \geq \log^2(|\partial\Lambda|)$ . By applying the above choice of  $\tilde{\ell}$  with Eq. (D71) to the upper bound (D69), we finally arrive at the inequality

$$\text{SR}(\tilde{\rho}_0^{2q}) \leq \exp[\tilde{c}_4\beta \log(\beta|\partial\Lambda|) + \tilde{c}_6|\partial\Lambda|\beta^{2/3} \log^{2/3}(\beta|\partial\Lambda|)]. \quad (\text{D72})$$

The inequality  $\beta \log(\beta|\partial\Lambda|) \geq |\partial\Lambda|\beta^{2/3} \log^{2/3}(\beta|\partial\Lambda|)$  holds for  $\beta \gtrsim |\partial\Lambda|^3$ . However, when  $\beta = \mathcal{O}(|\partial\Lambda|^3)$ , the upper bound gives  $e^{|\partial\Lambda|^3}$  and is worse than the trivial upper bound  $e^{\mathcal{O}(n)}$  because of  $|\partial\Lambda| = \mathcal{O}(n^{(d-1)/d})$ . We thus conclude that the second term in Eq. (D72) is more dominant than the first term.

We have chosen  $\delta = 1/|L|^2$ , and, hence, the inequality (D8) for  $p = 1$  ensures

$$\|e^{-\beta H} - \tilde{\rho}_0^{2q}\|_1 \leq \frac{3e\beta}{2\beta_0|L|^2}, \quad (\text{D73})$$

where we set  $\|e^{-\beta H}\|_1 = 1$ . Then, by using the Alicki-Fannes inequality [106,128], the main inequality (D1) is obtained:

$$\begin{aligned} I(L:R)_{\rho_\beta} &\leq I(L:R)_{\tilde{\rho}_0^{2q}} + \mathcal{O}(\beta/|L|) \\ &\leq 2\log[\text{SR}(\tilde{\rho}_0^{2q})] + \mathcal{O}(\beta/|L|) \\ &\leq C|\partial\Lambda|\beta^{2/3} \log^{2/3}(\beta|\partial\Lambda|), \end{aligned} \quad (\text{D74})$$

where the inequality  $I(L:R)_{\tilde{\rho}_0^{2q}} \leq 2\log[\text{SR}(\tilde{\rho}_0^{2q})]$  is derived from the purification of  $\tilde{\rho}_0^{2q}$  as

$$|\psi\rangle = (\tilde{\rho}_0^q \otimes \hat{1}) \sum_{j=1}^{D_\Lambda} |j\rangle_\Lambda \otimes |j\rangle_{\Lambda'}, \quad (\text{D75})$$

where  $\{|j\rangle\}_{j=1}^{D_\Lambda}$  is an arbitrary orthonormal basis (see also Sec. VI A). The mutual information  $I(L:R)_{\tilde{\rho}_0^{2q}}$  is smaller than 2 times the entanglement entropy for  $|\psi\rangle$  [see the inequality (38) in the main text], which is trivially smaller than  $2\log[\text{SR}(\tilde{\rho}_0^q)]$ . This completes the proof. ■

## APPENDIX E: PROOFS OF PROPOSITION 4 AND LEMMA 5

### 1. Proposition 4 for general $k$ -local Hamiltonian (A2)

We here prove the following statement about high temperatures which plays a crucial role in obtaining the quasilinear time algorithm.

*Proposition 22.*—For  $\beta \leq 1/(8gk)$ , we can construct a matrix product representation  $M_\beta$  of  $\rho_\beta$  up to an error

$$\|M_\beta - e^{-\beta H}\|_p \leq \epsilon \|e^{-\beta H}\|_p \quad (\text{E1})$$

for an arbitrary positive  $p$ , where  $M_\beta$  has the bond dimension of  $e^{\tilde{\mathcal{O}}[\sqrt{\log(n/\epsilon)}]}$ . The sufficient computational time for this construction is given by

$$ne^{\tilde{\mathcal{O}}[\sqrt{\log(n/\epsilon)}]}. \quad (\text{E2})$$

We notice that the computational cost does not depend on  $p$ .

We here consider general  $k$ -local Hamiltonians. In the main text, the Hamiltonian (5) is considered. By choosing

$$k = 2, \quad g = 1 \quad (\text{E3})$$

in Proposition 22, we can obtain Proposition 4. Here, the equation  $g = 1$  is derived from the condition  $\max_{i \in [n]} (\|h_{i-1,i}\| + \|h_{i,i+1}\|) \leq g = 1$  in Eq. (5).

### 2. Proof strategy

We aim to give an explicit algorithm to obtain the MPO approximation of  $e^{-\beta H}$  for  $\beta \leq 1/(8gk)$ . We decompose the total system into small blocks  $\{B_s\}_{s=1}^{n_0}$  with length  $\ell_0$  (i.e.,  $|B_s| = \ell_0$ ), which gives  $\|\Lambda\| = n_0\ell_0$  (see Fig. 8). In fact, we may not be able to find an integer  $n_0$  satisfying  $n = n_0\ell_0$ , but we can arbitrary extend the system size  $\Lambda \rightarrow \Lambda \sqcup \delta\Lambda$  without changing the Hamiltonian. We have only to add the zero operators, and the form of Eq. (A2) is still retained as follows:

$$H = \sum_{X \subset \Lambda \sqcup \delta\Lambda, |X| \leq k} h_X, \quad \sup_{i \in \Lambda \sqcup \delta\Lambda} \sum_{X: X \ni i} \|h_X\| \leq g,$$

where  $h_X = \hat{0}_X$  if  $X \cap \delta\Lambda \neq \emptyset$ .

We then define  $H_{1:j}$  as

$$H_{1:j} = \sum_{X \subset B_{\leq j}} h_X, \quad B_{\leq j} := B_1 \sqcup B_2 \sqcup \dots \sqcup B_j. \quad (\text{E4})$$

By using this notation, we define operators  $\Phi_j$  and  $\Phi_{1:j}$  as follows:

$$\Phi_j := e^{\beta H_{1:j-1}} e^{-\beta H_{1:j}}, \quad \Phi_{1:j} = \Phi_1 \Phi_2 \dots \Phi_j, \quad (\text{E5})$$



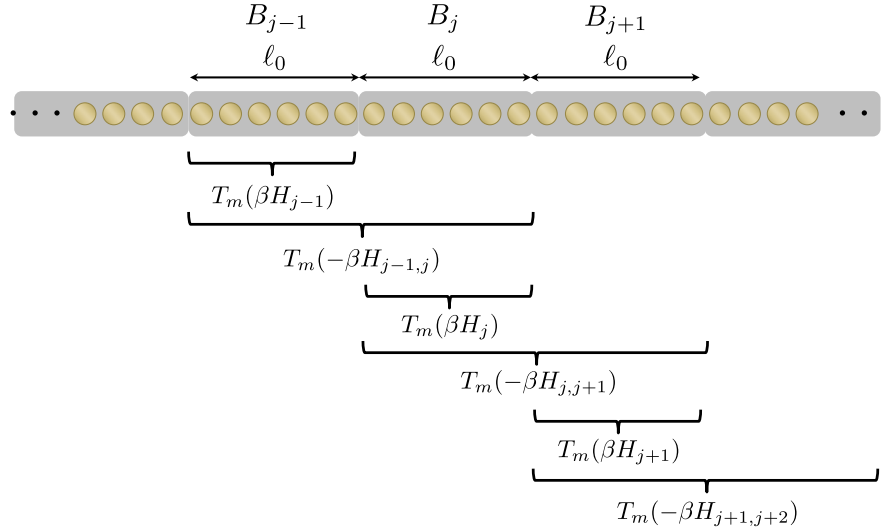


FIG. 8. Basic strategy for the approximation of  $e^{-\beta H}$  [ $\beta \leq 1/(8gk)$ ]. We decompose the total systems into blocks. We then take the two-step approximation: (i)  $e^{-\beta H} \approx \tilde{\Phi}_{1:n_0}$  and (ii)  $\tilde{\Phi}_{1:n_0} \approx \tilde{\Phi}_{1:n_0}^{(m)}$ , which yield  $e^{-\beta H} \approx \tilde{\Phi}_{1:n_0}^{(m)}$ . Here, the approximated quantum Gibbs state  $\tilde{\Phi}_{1:n_0}^{(m)}$  is constructed from the polynomials  $T_m(\beta H_{j-1})$  and  $T_m(-\beta H_{j-1,j})$  as in Eq. (E8).

where we define  $H_{1:0} = \hat{0}$ . Note that each of  $\{\Phi_j\}_{j=1}^{n_0}$  may be highly nonlocal. By using  $\{\Phi_j\}_{j=1}^{n_0}$ , we have

$$e^{-\beta H} = \Phi_{1:n_0}. \quad (\text{E6})$$

We, in the following, derive efficient approximations for  $\{\Phi_j\}_{j=1}^{n_0}$ . For this purpose, we define  $\tilde{\Phi}_j$  and  $\tilde{\Phi}_{1:j}$  as follows:

$$\tilde{\Phi}_j := e^{\beta H_{j-1}} e^{-\beta H_{j-1,j}}, \quad \tilde{\Phi}_{1:j} = \tilde{\Phi}_1 \tilde{\Phi}_2 \cdots \tilde{\Phi}_j,$$

$$H_j := \sum_{X:XC B_j} h_X + \sum_{X:Xn B_j \neq \emptyset, Xn B_{j+1} \neq \emptyset} h_X,$$

$$H_{j,j+1} := H_j + H_{j+1}. \quad (\text{E7})$$

Here,  $H_j$  is comprised of the internal interaction in the block  $B_j$  and the block-block interactions between  $B_j$  and  $B_{j+1}$ . We first approximate  $e^{-\beta H}$  by  $\tilde{\Phi}_{1:n_0}$ . Then, we approximate  $\tilde{\Phi}_{1:n_0}$  by using polynomial approximations as

$$\begin{aligned} \tilde{\Phi}_j^{(m)} &:= T_m(\beta H_{j-1}) T_m(-\beta H_{j-1,j}), \\ \tilde{\Phi}_{1:j}^{(m)} &= \tilde{\Phi}_1^{(m)} \tilde{\Phi}_2^{(m)} \cdots \tilde{\Phi}_j^{(m)}, \end{aligned} \quad (\text{E8})$$

where  $T_m(x) = \sum_{s=0}^m x^s / s!$  is the truncated Taylor expansion. In the following, we estimate the parameters  $\ell_0$  and  $m$  to achieve the precision of

$$\|\tilde{\Phi}_{1:n_0}^{(m)} e^{\beta H} - 1\| \leq \epsilon. \quad (\text{E9})$$

The above upper bound yields, for an arbitrary Schatten  $p$ -norm,

$$\begin{aligned} \|\tilde{\Phi}_{1:n_0}^{(m)} - e^{-\beta H}\|_p &\leq \|(\tilde{\Phi}_{1:n_0}^{(m)} e^{\beta H} - 1) e^{-\beta H}\|_p \\ &\leq \|\tilde{\Phi}_{1:n_0}^{(m)} e^{\beta H} - 1\| \cdot \|e^{-\beta H}\|_p \\ &\leq \epsilon \|e^{-\beta H}\|_p, \end{aligned} \quad (\text{E10})$$

where we use the Hölder inequality (B2) in the second step.

In Appendix E3, we prove that the inequality (E9) is achieved by choosing  $\ell_0$  and  $m$  as

$$\ell_0 = c_0 k \log(6n/\epsilon) \quad \text{and} \quad m = c_1 \log(6n/\epsilon), \quad (\text{E11})$$

where  $c_0$  and  $c_1$  are constants of  $\mathcal{O}(1)$ . Under the choice above, we estimate the Schmidt rank  $\tilde{\Phi}_{1:n_0}^{(m)}$  across an arbitrary cut.

Here,  $\tilde{\Phi}_{1:n_0}^{(m)}$  is given by

$$\tilde{\Phi}_{1:n_0}^{(m)} = T_m(-\beta H_1) T_m(\beta H_1) T_m(-\beta H_{1,2}) T_m(\beta H_2) T_m(-\beta H_{2,3}) \cdots T_m(\beta H_{n_0-1}) T_m(-\beta H_{n_0-1,n_0}). \quad (\text{E12})$$

Let us consider a cut between  $\Lambda_{\leq i}$  and  $\Lambda_{> i}$  for a fixed  $i \in \Lambda$ . Then, at most five polynomials contribute to the Schmidt rank of  $\text{SR}(\tilde{\Phi}_{1:n_0}^{(m)}, i)$  (see Fig. 8), where we denote them as  $T_m(-\beta H_{j-1,j})$ ,  $T_m(\beta H_j)$ ,  $T_m(-\beta H_{j,j+1})$ ,  $T_m(\beta H_{j+1})$ , and  $T_m(-\beta H_{j+1,j+2})$  ( $j \in [n_0]$ ). We thus obtain

$$\begin{aligned} \log [\text{SR}(\tilde{\Phi}_{1:n_0}^{(m)}, i)] &\leq \log \{\text{SR}[T_m(-\beta H_{j-1,j}), i]\} + \log \{\text{SR}[T_m(\beta H_j), i]\} + \log \{\text{SR}[T_m(-\beta H_{j,j+1}), i]\} \\ &\quad + \log \{\text{SR}[T_m(\beta H_{j+1}), i]\} + \log \{\text{SR}[T_m(-\beta H_{j+1,j+2}), i]\}. \end{aligned} \quad (\text{E13})$$

By using Lemma 17 with  $\Phi_1 = \Phi_2 = 1$  and  $M = 1$ , we obtain from Eq. (E11)

$$\begin{aligned} \log \{\text{SR}[T_m(\beta H_j), i]\} &\leq C \max(m/\ell_0, \sqrt{m}) \log(\zeta m) \\ &= \tilde{\mathcal{O}}[\sqrt{\log(n/\epsilon)}], \end{aligned} \quad (\text{E14})$$

where  $C$  is a constant of  $\mathcal{O}(1)$  which depends on  $k$ . Therefore, for an arbitrary cut,  $\log [\text{SR}(\tilde{\Phi}_{1:n_0}^{(m)}, i)]$  is bounded from above by  $\tilde{\mathcal{O}}[\sqrt{\log(n/\epsilon)}]$ . This result ensures that the operator  $\tilde{\Phi}_{1:n_0}^{(m)}$  is expressed by a matrix product operator with a bond dimension of  $e^{\tilde{\mathcal{O}}[\sqrt{\log(n/\epsilon)}]}$ . Because the operator  $\tilde{\Phi}_{1:n_0}^{(m)}$  satisfies the approximation error of Eq. (E10), we prove the first part of the statement in Proposition 22.

In order to prove the second part of the statement, we consider the computational cost to construct the MPO of  $\tilde{\Phi}_{1:n_0}^{(m)}$ . We first note that each of the polynomials  $T_m(\beta H_j)$  and  $T_m(\beta H_{j,j+1})$  is described by a local MPO with bond dimension  $D = e^{\tilde{\mathcal{O}}[\sqrt{\log(n/\epsilon)}]}$ . In the computations of  $H_j^q$  and  $H_{j,j+1}^q$  ( $q \leq m$ ), we can utilize the compression of the MPO which is based on the singular value decomposition. Next, recall that we can express arbitrary local Hamiltonians by the MPO with a constant bond dimension [112]. Using this result, we recursively construct the power of the Hamiltonian  $H_j^q$  from  $H_j^{q-1}$ . At each stage of this recursion, we ensure that the bond dimension is smaller than  $D$ , by compressing the MPO using the singular value decomposition. By representing the MPO in the canonical form [112], this compression can be performed efficiently with a computational cost of  $\text{poly}(D)$  (since the Schmidt coefficient beyond the rank  $D$  is exactly equal to zero, and the error in this compression is equal to zero). These procedures allow us to construct the local MPO of  $T_m(\beta H_j)$  and  $T_m(\beta H_{j,j+1})$  with a run time of  $\text{poly}(D) = e^{\tilde{\mathcal{O}}[\sqrt{\log(n/\epsilon)}]}$ .

The remaining task is to connect all the local MPOs of  $T_m(\beta H_j)$  and  $T_m(\beta H_{j,j+1})$  to construct the operator  $\tilde{\Phi}_{1:n_0}^{(m)}$  in Eq. (E12). From the inequality (E13), the bond dimension is at most  $D^5$ , and, hence, the iterative multiplications of the functions  $T_m(\beta H_j)$  and  $T_m(\beta H_{j,j+1})$  require  $\text{poly}(D)$  computational time, which results in the total computational time of  $n \times \text{poly}(D) = ne^{\tilde{\mathcal{O}}[\sqrt{\log(n/\epsilon)}]}$ . This completes the proof of Proposition 22. ■

### 3. Proof of the choice (E11)

We prove that the choice of Eq. (E11) achieves the approximation error (E9). In order to estimate the lhs in Eq. (E9), we recursively estimate

$$\epsilon_j := \|\tilde{\Phi}_{1:j}^{(m)} \Phi_{1:j}^{-1} - 1\|, \quad (\text{E15})$$

where we set  $\tilde{\Phi}_{1:0} = \Phi_{1:0} = 1$ . Because of  $\Phi_{1:n_0} = e^{-\beta H}$  as in Eq. (E6), we have  $\epsilon_{n_0} = \|\tilde{\Phi}_{1:n_0}^{(m)} e^{\beta H} - 1\|$ . By using  $\epsilon_j$ , we can calculate the upper bound of  $\epsilon_{j+1}$ . From  $\Phi_{1:j} = e^{-\beta H_{1:j}}$ , we have

$$\begin{aligned} \tilde{\Phi}_{1:j+1}^{(m)} \Phi_{1:j+1}^{-1} &= \tilde{\Phi}_{1:j}^{(m)} \Phi_{1:j}^{-1} (e^{-\beta H_{1:j}} \tilde{\Phi}_{j+1}^{(m)} \Phi_{j+1}^{-1} e^{\beta H_{1:j}}) \\ &= \tilde{\Phi}_{1:j}^{(m)} \Phi_{1:j}^{-1} \tilde{\Psi}_j, \end{aligned} \quad (\text{E16})$$

where  $\tilde{\Psi}_j := e^{-\beta H_{1:j}} \tilde{\Phi}_{j+1}^{(m)} \Phi_{j+1}^{-1} e^{\beta H_{1:j}}$ . We then obtain

$$\begin{aligned} \tilde{\Phi}_{1:j+1}^{(m)} \Phi_{1:j+1}^{-1} - 1 &= (\tilde{\Phi}_{1:j}^{(m)} \Phi_{1:j}^{-1} - 1)(\tilde{\Psi}_j - 1) + (\tilde{\Psi}_j - 1) + (\tilde{\Phi}_{1:j}^{(m)} \Phi_{1:j}^{-1} - 1), \end{aligned}$$

and, hence,

$$\epsilon_{j+1} \leq \epsilon_j \delta_j + \epsilon_j + \delta_j, \quad (\text{E17})$$

where  $\delta_j := \|\tilde{\Psi}_j - 1\|$ . When we obtain  $\delta_j \leq \bar{\delta}$ , we have  $\epsilon_{j+1} \leq (1 + \bar{\delta})\epsilon_j + \bar{\delta}$ , which yields  $\epsilon_{n_0} \leq (1 + \bar{\delta})^{n_0} - 1$ . We here use  $\epsilon_0 = 0$ . For  $\bar{\delta} \leq 1/n_0$ , we have

$$\epsilon_{n_0} \leq 2n_0 \bar{\delta}. \quad (\text{E18})$$

Therefore, the problem reduces to the estimation of  $\delta_j$ .

The operator  $\Psi_j$  includes the imaginary-time evolution by  $e^{-\beta H_{1:j}}$ , but the high-temperature assumption of  $\beta < 1/(8gk)$  allows us to prove  $\delta_j \ll 1$ . In order to calculate the upper bound of  $\|\tilde{\Psi}_j - 1\|$ , we define

$$\begin{aligned} \Psi_j^{(m)} &:= e^{-\beta H_{1:j}} \tilde{\Phi}_{j+1}^{(m)} \Phi_{j+1}^{-1} e^{\beta H_{1:j}}, \\ \Psi_j &:= e^{-\beta H_{1:j}} \tilde{\Phi}_{j+1} \Phi_{j+1}^{-1} e^{\beta H_{1:j}}, \\ \delta_{j,1} &= \|\Psi_j^{(m)} - 1\|, \quad \delta_{j,2} = \|\Psi_j - 1\|. \end{aligned} \quad (\text{E19})$$

The above definition implies  $\tilde{\Psi}_j - 1 = \Psi_j^{(m)} \Psi_j - 1$ , and, hence,

$$\begin{aligned} \|\tilde{\Psi}_j - 1\| &\leq \|(\Psi_j^{(m)} - 1)(\Psi_j - 1) + (\Psi_j - 1) + (\Psi_j^{(m)} - 1)\| \\ &\leq \delta_{j,1}\delta_{j,2} + \delta_{j,2} + \delta_{j,1}. \end{aligned} \quad (\text{E20})$$

Indeed, we prove the following lemmas.

*Lemma 23.*—Under the assumption of  $\beta \leq 1/(8gk)$ , we obtain the upper bound of

$$\delta_{j,1} \leq (4/3)^{(2\ell_0/k)+1} \|\tilde{\Phi}_{j+1}^{(m)} \tilde{\Phi}_{j+1}^{-1} - 1\|. \quad (\text{E21})$$

Here,  $m$  is a control parameter and can be chosen appropriately.

*Lemma 24.*—Under the assumption of  $\beta \leq 1/(8gk)$ , we obtain the upper bound of

$$\delta_{j,2} \leq 10g\ell_0 2^{-\ell_0/k} e^{10g\beta\ell_0/3} \leq 10g\ell_0 2^{-\ell_0/(3k)}, \quad (\text{E22})$$

where the second inequality is derived from  $e^{10gk\beta/3} \leq e^{5/12} < 2^{2/3}$ .

Based on the above lemma, we choose the block size  $\ell_0$  as

$$\ell_0 = c_0 k \log(1/\tilde{\epsilon}), \quad (\text{E23})$$

where  $c_0$  is a constant such that  $\delta_{j,2} \leq 10g\ell_0 2^{-\ell_0/(3k)} \leq \tilde{\epsilon}$  and we fix  $\tilde{\epsilon} (< 1)$  afterward. Also, in order to upper bound  $\delta_{j,1}$  in Eq. (E21), we need to estimate the norm of

$$\begin{aligned} &\tilde{\Phi}_{j+1}^{(m)} \tilde{\Phi}_{j+1}^{-1} - 1 \\ &= T_m(\beta H_{j-1}) T_m(-\beta H_{j-1,j}) e^{\beta H_{j-1,j}} e^{-\beta H_{j-1}} - 1. \end{aligned}$$

We then obtain

$$\begin{aligned} \|\tilde{\Phi}_{j+1}^{(m)} \tilde{\Phi}_{j+1}^{-1} - 1\| &\leq \|T_m(\beta H_{j-1}) [T_m(-\beta H_{j-1,j}) e^{\beta H_{j-1,j}} - 1] e^{-\beta H_{j-1}} + T_m(\beta H_{j-1}) e^{-\beta H_{j-1}} - 1\| \\ &\leq \|T_m(\beta H_{j-1})\| \cdot \|e^{-\beta H_{j-1}}\| \cdot \|T_m(-\beta H_{j-1,j}) e^{\beta H_{j-1,j}} - 1\| + \|T_m(\beta H_{j-1}) e^{-\beta H_{j-1}} - 1\|. \end{aligned}$$

Because of  $\|H_{j-1}\| \leq g\ell_0$  and  $\|H_{j-1,j}\| \leq 2g\ell_0$ , we have  $\|T_m(\beta H_{j-1})\| \leq e^{\mathcal{O}(\beta g\ell_0)} = e^{\mathcal{O}(\ell_0/k)}$  and  $\|e^{-\beta H_{j-1}}\| \leq e^{\mathcal{O}(\ell_0/k)}$ .

In order to achieve  $\|\tilde{\Phi}_{j+1}^{(m)} \tilde{\Phi}_{j+1}^{-1} - 1\| \leq \tilde{\epsilon} (4/3)^{-(2\ell_0/k)-1}$  (or  $\delta_{j,1} \leq \tilde{\epsilon}$ ), we need to choose  $m$  such that

$$\|T_m(-\beta H_{j-1,j}) e^{\beta H_{j-1,j}} - 1\| \leq \tilde{\epsilon} e^{-\mathcal{O}(\ell_0/k)}. \quad (\text{E24})$$

From  $\|\beta H_{j-1,j}\| \lesssim \beta g\ell_0 = \mathcal{O}(\ell_0/k)$  for  $\beta \leq 1/(8gk)$ , the above inequality is satisfied by choosing  $m = \mathcal{O}(\ell_0/k) + \mathcal{O}[\log(1/\tilde{\epsilon})]$ . The choice of Eq. (E23) implies

$$m = c_1 \log(1/\tilde{\epsilon}), \quad (\text{E25})$$

where  $c_1$  is a constant of  $\mathcal{O}(1)$ .

Under the above choices of  $\ell_0$  and  $m$ , we obtain  $\delta_{j,1} \leq \tilde{\epsilon}$  and  $\delta_{j,2} \leq \tilde{\epsilon}$ , and, hence, from the inequality (E20), we have

$$\|\tilde{\Psi}_j - 1\| \leq 3\tilde{\epsilon}. \quad (\text{E26})$$

We thus obtain  $\bar{\delta} = 3\tilde{\epsilon}$ , which reduces the inequality (E18) to

$$\epsilon_{n_0} \leq 6\tilde{\epsilon} n_0 \leq 6n\tilde{\epsilon}. \quad (\text{E27})$$

By choosing  $\tilde{\epsilon} = \epsilon/(6n)$ , we can obtain the desired precision (E9) between  $\tilde{\Phi}_{1:n_0}^{(m)}$  and  $e^{-\beta H}$ . This completes the proof. ■

#### 4. Proof of Lemma 23

We here consider an arbitrary operator  $O_S$  supported on  $S$  and derive the upper bound of

$$e^{-\beta H_{1:j}} O_S e^{\beta H_{1:j}} = \sum_{m=0}^{\infty} \frac{(-\beta)^m}{m!} \text{ad}_{H_{1:j}}^m(O_S). \quad (\text{E28})$$

By using Lemma 13 or the inequality (B18), we can derive

$$\|\text{ad}_{H_{1:j}}^m(O_S)\| \leq (2gk)^m \|O_S\| \prod_{s=1}^m [\|S\|/k + (s-1)], \quad (\text{E29})$$

where we use the condition that  $H_{1:j}$  and  $H_j$  are  $k$ -local operators as in Eq. (A2). We then obtain

$$\begin{aligned} \|e^{-\beta H_{1:j}} O_S e^{\beta H_{1:j}}\| &\leq \|O_S\| \sum_{m=0}^{\infty} \frac{(2gk\beta)^m}{m!} \prod_{s=1}^m [\|S\|/k + s - 1] \\ &= \|O_S\| (1 - 2gk\beta)^{-\|S\|/k}, \end{aligned} \quad (\text{E30})$$

where we use the equation of  $(1-x)^{-y} = \sum_{m=0}^{\infty} x^m/m! \prod_{s=1}^m (y+s-1)$ .

We then choose  $O_S$  as  $\tilde{\Phi}_{j+1}^{(m)} \tilde{\Phi}_{j+1}^{-1} - 1$ , which yields

$$e^{-\beta H_{1:j}} O_S e^{\beta H_{1:j}} = \Psi_j^{(m)} - 1. \quad (\text{E31})$$

From the definitions (E7) and (E8), we have

$$\tilde{\Phi}_{j+1}^{(m)} \tilde{\Phi}_{j+1}^{-1} = T_m(\beta H_j) T_m(-\beta H_{j,j+1}) e^{\beta H_{j,j+1}} e^{\beta H_j},$$

and, hence, the support of this operator satisfies

$$|\text{Supp}(\tilde{\Phi}_{j+1}^{(m)} \tilde{\Phi}_{j+1}^{-1})| \leq 2\ell_0 + k. \quad (\text{E32})$$

Therefore, by using the inequality (E30) with  $|S| = 2\ell_0 + k$ , we have

$$\|\Psi_j^{(m)} - 1\| \leq (4/3)^{(2\ell_0/k)+1} \|\tilde{\Phi}_{j+1}^{(m)} \tilde{\Phi}_{j+1}^{-1} - 1\|, \quad (\text{E33})$$

where we use  $1 - 2gk\beta \geq 3/4$  because of  $\beta \leq 1/(8gk)$ . This completes the proof of Lemma 23.  $\blacksquare$

### 5. Proof of Lemma 24

We here estimate the norm of

$$\begin{aligned} \Psi_j - 1 &= e^{-\beta H_{1:j}} \tilde{\Phi}_{j+1} \Phi_{j+1}^{-1} e^{\beta H_{1:j}} - 1 \\ &= (e^{-\beta H_{1:j}} \tilde{\Phi}_{j+1} e^{\beta H_{1:j}}) (e^{-\beta H_{1:j}} \Phi_{j+1}^{-1} e^{\beta H_{1:j}}) - 1. \end{aligned}$$

For the estimation, we are going to simplify the operators  $e^{-\beta H_{1:j}} \tilde{\Phi}_{j+1} e^{\beta H_{1:j}}$  and  $e^{-\beta H_{1:j}} \Phi_{j+1}^{-1} e^{\beta H_{1:j}}$ .

We first consider  $e^{-\beta H_{1:j}} \tilde{\Phi}_{j+1} e^{\beta H_{1:j}}$  and start from the equation of

$$e^{\beta H_{1:j+1}} = (\mathcal{T} e^{-\int_0^\beta e^{\tau H_{1:j}} H_{j+1} e^{-\tau H_{1:j}} d\tau}) e^{\beta H_{1:j}}, \quad (\text{E34})$$

where  $\mathcal{T}$  is the ordering operator. Then, from  $\Phi_{j+1}^{-1} = e^{\beta H_{1:j+1}} e^{-\beta H_{1:j}}$ , the above equation reduces  $e^{-\beta H_{1:j}} \Phi_{j+1}^{-1} e^{\beta H_{1:j}}$  to the following form:

$$\begin{aligned} e^{-\beta H_{1:j}} \Phi_{j+1}^{-1} e^{\beta H_{1:j}} &= \mathcal{T} e^{-\int_0^\beta H_{j+1}^{(\tau)} d\tau}, \\ H_{j+1}^{(\tau)} &:= e^{-(\beta-\tau)H_{1:j}} H_{j+1} e^{(\beta-\tau)H_{1:j}}. \end{aligned} \quad (\text{E35})$$

In a similar way, we can represent  $e^{-\beta H_{j:j+1}}$  as

$$e^{-\beta H_{j:j+1}} = (\mathcal{T} e^{-\int_0^\beta e^{-\tau H_j} H_{j+1} e^{\tau H_j} d\tau}) e^{-\beta H_j}, \quad (\text{E36})$$

and, hence, we have from  $\tilde{\Phi}_{j+1} := e^{\beta H_j} e^{-\beta H_{j:j+1}}$

$$\begin{aligned} e^{-\beta H_{1:j}} \tilde{\Phi}_{j+1} e^{\beta H_{1:j}} &= e^{-\beta H_{1:j}} (\mathcal{T} e^{-\int_0^\beta e^{(\beta-\tau)H_j} H_{j+1} e^{-(\beta-\tau)H_j} d\tau}) e^{\beta H_{1:j}} \\ &= \mathcal{T} e^{-\int_0^\beta e^{-\beta H_{1:j}} e^{(\beta-\tau)H_j} H_{j+1} e^{-(\beta-\tau)H_j} e^{\beta H_{1:j}} d\tau} \\ &= \mathcal{T} e^{-\int_0^\beta \tilde{H}_{j+1}^{(\tau)} d\tau}, \end{aligned} \quad (\text{E37})$$

where we define  $\tilde{H}_{j+1}^{(\tau)}$  as

$$\tilde{H}_{j+1}^{(\tau)} := e^{-\beta H_{1:j}} e^{(\beta-\tau)H_j} H_{j+1} e^{-(\beta-\tau)H_j} e^{\beta H_{1:j}}. \quad (\text{E38})$$

We now prove the following claim.

*Claim 25.*—Let  $\{A_j\}_{j=1}^N$  and  $\{B_j\}_{j=1}^N$  be arbitrary operators. We also define  $\Phi_{A,j} := e^{A_1} e^{A_2} \dots e^{A_j}$  and

$\Phi_{B,j} := e^{B_1} \dots e^{B_2} e^{B_1}$ . We then obtain the following upper bound as

$$\|\Phi_{A,N} \Phi_{B,N} - 1\| \leq \bar{\Phi} \sum_{s=1}^N \|e^{A_s} e^{B_s} - 1\|, \quad (\text{E39})$$

where  $\bar{\Phi} := \exp[\sum_{s=1}^N (\|A_s\| + \|B_s\|)]$ .

*Proof of Claim 25.*—By using the triangle inequality, we first obtain

$$\begin{aligned} \|\Phi_{A,N} \Phi_{B,N} - 1\| &\leq \|\Phi_{A,N-1} \Phi_{B,N-1} - 1\| \\ &\quad + \|\Phi_{A,N-1} (e^{A_N} e^{B_N} - 1) \Phi_{B,N-1}\| \\ &\leq \|\Phi_{A,N-1} \Phi_{B,N-1} - 1\| + \bar{\Phi} \|e^{A_N} e^{B_N} - 1\|, \end{aligned} \quad (\text{E40})$$

where we use  $\|\Phi_{A,N-1}\| \cdot \|\Phi_{B,N-1}\| \leq \bar{\Phi}$ . By iteratively applying the above inequality to  $\|\Phi_{A,s} \Phi_{B,s} - 1\|$ , we arrive at the main inequality (E39).  $\blacksquare$

By using the Trotter decomposition in the expressions of Eqs. (E35) and (E37), we can assign as  $\Phi_{A,N} \rightarrow e^{-\beta H_{1:j}} \tilde{\Phi}_{j+1} e^{\beta H_{1:j}}$  and  $\Phi_{B,N} \rightarrow e^{-\beta H_{1:j}} \Phi_{j+1}^{-1} e^{\beta H_{1:j}}$  in the limit of  $N \rightarrow \infty$ . Then, from Lemma 25, we obtain

$$\|\Psi_j - 1\| \leq \bar{\Phi}_\beta \int_0^\beta \|H_{j+1}^{(\tau)} - \tilde{H}_{j+1}^{(\beta-\tau)}\| d\tau, \quad (\text{E41})$$

where we define  $\bar{\Phi}_\beta$  as

$$\bar{\Phi}_\beta := \exp\left(\int_0^\beta \|H_{j+1}^{(\tau)}\| + \|\tilde{H}_{j+1}^{(\tau)}\| d\tau\right). \quad (\text{E42})$$

To complete the proof, we need to show the following claim.

*Claim 26.*—Under the assumption of  $\beta < 1/(8gk)$ , the following upper bounds hold:

$$\|H_{j+1}^{(\tau)} - \tilde{H}_{j+1}^{(\beta-\tau)}\| \leq 10g\ell_0 2^{-\ell_0/k} \quad (\text{E43})$$

and

$$\bar{\Phi}_\beta \leq e^{10g\beta\ell_0/3}. \quad (\text{E44})$$

By applying the above claim to Eq. (E41), we prove Lemma 24.  $\blacksquare$

#### a. Proof of Claim 26

We first estimate the norm of  $H_{j+1}^{(\tau)} - \tilde{H}_{j+1}^{(\tau)}$ . For this purpose, we first note that the  $H_{j-1}$  is supported on the subset  $B_{j-1} \sqcup \{j\ell_0 + 1, j\ell_0 + 2, \dots, j\ell_0 + k - 1\}$ , namely,

$$\text{Supp}(H_{j-1}) \subset B_j \sqcup \{j\ell_0 + 1, j\ell_0 + 2, \dots, j\ell_0 + k\},$$



where  $\text{Supp}(\cdots)$  denotes the support of the operator. On the other hand, because  $H_j$  includes at most  $k$ -body interactions, the support of  $\text{ad}_{H_j}^q(H_{j+1})$  is given by

$$\text{Supp}[\text{ad}_{H_j}^q(H_{j+1})] \subset \{(j+1)\ell_0 - qk, (j+1)\ell_0 - qk + 1, \dots, (j+1)\ell_0\} \sqcup B_{j+1} \sqcup B_{j+2}. \quad (\text{E45})$$

Therefore, we have

$$[H_{j-1}, \text{ad}_{H_j}^m(H_{j+1})] = 0 \quad \text{if } k + mk \leq \ell_0. \quad (\text{E46})$$

This result implies

$$\text{ad}_{H_j}^m(H_{j+1}) = \text{ad}_{H_{1:j}}^m(H_{j+1}) \quad \text{for } m \leq \ell_0/k - 1. \quad (\text{E47})$$

Hence, from the definition (E38) of  $\tilde{H}_{j+1}^{(\beta-\tau)}$ , we have

$$\begin{aligned} \tilde{H}_{j+1}^{(\beta-\tau)} &= \sum_{m=0}^{\infty} \sum_{m_1+m_2=m} \frac{(-\beta)^{m_1}}{m_1!} \frac{(\beta-\tau)^{m_2}}{m_2!} \text{ad}_{H_{1:j}}^{m_1} \text{ad}_{H_j}^{m_2}(H_{j+1}) \\ &= \sum_{m \leq \ell_0/k-1} \frac{(-\tau)^m}{m!} \text{ad}_{H_{1:j}}^m(H_{j+1}) + \sum_{m > \ell_0/k-1} \sum_{m_1+m_2=m} \frac{(-\beta)^{m_1}}{m_1!} \frac{(\beta-\tau)^{m_2}}{m_2!} \text{ad}_{H_{1:j}}^{m_1} \text{ad}_{H_j}^{m_2}(H_{j+1}). \end{aligned} \quad (\text{E48})$$

Therefore, we have the upper bound of

$$\|H_{j+1}^{(\tau)} - \tilde{H}_{j+1}^{(\beta-\tau)}\| \leq \sum_{m > \ell_0/k-1} \left( \frac{\tau^m}{m!} \text{ad}_{H_{1:j}}^m(H_{j+1}) + \sum_{m_1+m_2=m} \frac{\beta^{m_1}}{m_1!} \frac{|\beta-\tau|^{m_2}}{m_2!} \text{ad}_{H_{1:j}}^{m_1} \text{ad}_{H_j}^{m_2}(H_{j+1}) \right). \quad (\text{E49})$$

The remaining task is to estimate the summations. By applying the inequality (E29) with  $O_S = h_X$ , we obtain

$$\text{ad}_{H_{1:j}}^{m_1} \text{ad}_{H_j}^{m_2}(h_X) \leq (2gk)^{m_1+m_2} (m_1 + m_2)! \|h_X\|, \quad (\text{E50})$$

where  $h_X$  is an interaction operator in  $H_{j+1}$ . From this inequality with  $m_2 = 0$  and the definition of  $H_{j+1}$ , we have

$$\sum_{m > \ell_0/k-1} \frac{\tau^m}{m!} \|\text{ad}_{H_{1:j}}^m(h_X)\| \leq \|h_X\| \sum_{m > \ell_0/k-1} (2gk\tau)^m \leq \|h_X\| \frac{(2gk\beta)^{\ell_0/k-1}}{1 - 2gk\beta}, \quad (\text{E51})$$

where we use  $\tau \leq \beta$ . From the definition of  $H_{j+1}$  in Eq. (E7), we have

$$\sum_{m > \ell_0/k-1} \frac{\tau^m}{m!} \|\text{ad}_{H_{1:j}}^m(H_{j+1})\| \leq \frac{(2gk\beta)^{\ell_0/k-1}}{1 - 2gk\beta} \sum_{X: X \cap B_j \neq \emptyset} \|h_X\| \leq \frac{g\ell_0(2gk\beta)^{\ell_0/k-1}}{1 - 2gk\beta}, \quad (\text{E52})$$

where we use the  $\sum_{X: X \cap B_j \neq \emptyset} \|h_X\| \leq \sum_{i \in B_j} \sum_{X: X \ni i} \|h_X\| \leq g|B_j|$  with the condition in Eq. (A2).

In a similar way, we calculate

$$\sum_{m > \ell_0/k-1} \sum_{m_1+m_2=m} \frac{\beta^{m_1}}{m_1!} \frac{|\beta-\tau|^{m_2}}{m_2!} \|\text{ad}_{H_{1:j}}^{m_1} \text{ad}_{H_j}^{m_2}(H_{j+1})\| \leq g\ell_0 \sum_{m > \ell_0/k-1} (2gk\beta)^m \sum_{m_1+m_2=m} \frac{(m_1 + m_2)!}{m_1! m_2!} = \frac{g\ell_0(4gk\beta)^{\ell_0/k-1}}{1 - 4gk\beta}, \quad (\text{E53})$$

where we use  $\sum_{m_1+m_2=m} [(m_1 + m_2)! / m_1! m_2!] = 2^m$ . By applying the inequalities (E51) and (E53) to Eq. (E49), we obtain

$$\|H_{j+1}^{(\tau)} - \tilde{H}_{j+1}^{(\beta-\tau)}\| \leq \frac{g\ell_0(2gk\beta)^{\ell_0/k-1}}{1 - 2gk\beta} + \frac{g\ell_0(4gk\beta)^{\ell_0/k-1}}{1 - 4gk\beta}. \quad (\text{E54})$$

Therefore, by using the assumption  $\beta \leq 1/(8gk)$ , we prove the inequality (E43).

The above analyses can also be utilized to estimate the norms of  $\|H_{j+1}^{(\tau)}\|$  and  $\|\tilde{H}_{j+1}^{(\tau)}\|$ . From the inequality (E51), we first obtain

$$\|H_{j+1}^{(\tau)}\| \leq \sum_{m=0}^{\infty} \frac{\tau^m}{m!} \|\text{ad}_{H_{1,j}}^m(H_{j+1})\| \leq \frac{g\ell_0}{1-2gk\beta}. \quad (\text{E55})$$

From the inequality (E53), we can also derive

$$\|\tilde{H}_{j+1}^{(\tau)}\| \leq \sum_{m=0}^{\infty} \sum_{m_1+m_2=m} \frac{\beta^{m_1}}{m_1!} \frac{|\beta-\tau|^{m_2}}{m_2!} \|\text{ad}_{H_{1,j}}^{m_1} \text{ad}_{H_j}^{m_2}(H_{j+1})\| \leq \frac{g\ell_0}{1-4gk\beta}. \quad (\text{E56})$$

By applying the above two inequalities to Eq. (E42) under the assumption  $\beta \leq 1/(8gk)$ , we prove the inequality (E44). This completes the proof of Claim 26.  $\blacksquare$

## 6. Proof of Lemma 5

We here prove Lemma 5 in the main text, which gives the upper bound of the Schmidt rank of  $M_\beta^q$  with  $M_\beta$  equal to  $\tilde{\Phi}_{1:n_0}^{(m)}$  in Eq. (E12):

$$M_\beta = T_m(-\beta H_1) T_m(\beta H_1) T_m(-\beta H_{1,2}) T_m(\beta H_2) T_m(-\beta H_{2,3}) \cdots T_m(\beta H_{n_0-1}) T_m(-\beta H_{n_0-1,n_0}), \quad (\text{E57})$$

where  $m$  and  $\ell_0$  are chosen as in Eq. (E11). Our purpose is to prove that, for arbitrary  $q \in \mathbb{N}$ , the Schmidt rank of the  $q$ th power of  $M_\beta$  is upper bounded by

$$\log[\text{SR}(M_\beta^q)] \leq C \max(q, \sqrt{mq}) \log(mq). \quad (\text{E58})$$

As shown in the inequality (E13), for an arbitrary cut, at most five polynomials contribute to the Schmidt rank. We denote them as  $T_m(-\beta H_{j-1,j})$ ,  $T_m(\beta H_j)$ ,  $T_m(-\beta H_{j,j+1})$ ,  $T_m(\beta H_{j+1})$ , and  $T_m(-\beta H_{j+1,j+2})$  ( $j \in [n_0]$ ). We then denote  $M_\beta$  by

$$M_\beta = \Phi_1 T_m(-\beta H_{j-1,j}) T_m(\beta H_j) T_m(-\beta H_{j,j+1}) T_m(\beta H_{j+1}) T_m(-\beta H_{j+1,j+2}) \Phi_2, \quad (\text{E59})$$

where

$$\begin{aligned} \Phi_1 &= T_m(-\beta H_1) T_m(\beta H_1) T_m(-\beta H_{1,2}) \cdots T_m(-\beta H_{j-2,j-1}) T_m(\beta H_{j-1}), \\ \Phi_2 &= T_m(\beta H_{j+2}) T_m(-\beta H_{j+2,j+3}) \cdots T_m(\beta H_{n_0-1}) T_m(-\beta H_{n_0-1,n_0}). \end{aligned} \quad (\text{E60})$$

Note that the Hamiltonians  $H_j$  and  $H_{j,j+1}$  are defined on the subsets  $B_j$  and  $B_j \sqcup B_{j+1}$ , respectively (see Fig. 8). We then apply Corollary 18 to  $M_\beta^q$  with  $p = 5$  and  $\ell = 2\ell_0$ . The inequality (C45) gives

$$\text{SR}(M_\beta^q) \leq \min_{\tilde{\ell}: \tilde{\ell} \leq 2\ell_0} [\zeta^{\tilde{\ell}} (10mq\zeta^k)^{10q+10\tilde{\ell}+(10kmq/\tilde{\ell})}] \leq \min_{\tilde{\ell}: \tilde{\ell} \leq 2\ell_0} [(10mq\zeta^{2k})^{10q+10\tilde{\ell}+(10kmq/\tilde{\ell})}]. \quad (\text{E61})$$

We now choose  $\tilde{\ell}$  as

$$\tilde{\ell} = \sqrt{kmq} = \sqrt{\frac{c_1}{c_0} \ell_0 q}, \quad (\text{E62})$$

where the second equation comes from the choice of Eq. (E11). Because of the constraint  $\tilde{\ell} \leq 2\ell_0$ , the exponent  $q$  should satisfy

$$q \leq \frac{4c_0}{c_1} \ell_0 = \frac{4c_0^2 k}{c_1} \log(6n/\epsilon). \quad (\text{E63})$$

Under this condition, we can choose  $\tilde{\ell}$  as in Eq. (E62), and hence, we obtain

$$\begin{aligned} \log[\text{SR}(M_\beta^q)] &\leq C' \log(mq) [q + \sqrt{mq}] \\ &\leq C \sqrt{mq} \log(mq), \end{aligned} \quad (\text{E64})$$

with  $C'$  and  $C$  constants of  $\mathcal{O}(1)$ , where we use  $q \lesssim n$  because of  $q \leq (2c_0/c_1)\ell_0$  and  $\ell_0 \propto m$  from Eq. (E11).

On the other hand, for  $q > (4c_0/c_1)\ell_0$ , we cannot choose  $\ell_0$  as in Eq. (E62). We here choose  $\tilde{\ell} = 2\ell_0$  and obtain

$$10q + 10\tilde{\ell} + \frac{10kmq}{\tilde{\ell}} = 10q + 20\ell_0 + \frac{5kmq}{\ell_0} \leq \left(10 + \frac{5c_1}{c_0} + \frac{5c_1}{c_0}\right)q, \quad (\text{E65})$$

where we use  $\ell_0 < c_1q/(4c_0)$  and  $m/\ell_0 = c_1/(c_0k)$  from Eq. (E11). We thus obtain

$$\log[\text{SR}(M_\beta^q)] \leq Cq \log(mq). \quad (\text{E66})$$

By combining the inequalities (E64) and (E66), we obtain the main inequality (E58). This completes the proof. ■

- 
- [1] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner *et al.*, *Probing Many-Body Dynamics on a 51-Atom Quantum Simulator*, *Nature (London)* **551**, 579 (2017).
- [2] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z.-X. Gong, and C. Monroe, *Observation of a Many-Body Dynamical Phase Transition with a 53-Qubit Quantum Simulator*, *Nature (London)* **551**, 601 (2017).
- [3] A. D. King, J. Carrasquilla, J. Raymond, I. Ozfidan, E. Andriyash, A. Berkley, M. Reis, T. Lanting, R. Harris, F. Altomare *et al.*, *Observation of Topological Phenomena in a Programmable Lattice of 1,800 Qubits*, *Nature (London)* **560**, 456 (2018).
- [4] C. Hempel, C. Maier, J. Romero, J. McClean, T. Monz, H. Shen, P. Jurcevic, B. P. Lanyon, P. Love, R. Babbush, A. Aspuru-Guzik, R. Blatt, and C. F. Roos, *Quantum Chemistry Calculations on a Trapped-Ion Quantum Simulator*, *Phys. Rev. X* **8**, 031022 (2018).
- [5] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell *et al.*, *Quantum Supremacy Using a Programmable Superconducting Processor*, *Nature (London)* **574**, 505 (2019).
- [6] M. Motta, C. Sun, A. T. K. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. S. L. Brandão, and G. K.-L. Chan, *Determining Eigenstates and Thermal States on a Quantum Computer Using Quantum Imaginary Time Evolution*, *Nat. Phys.* **16**, 205 (2020).
- [7] M. H. Amin, E. Andriyash, J. Rolfe, B. Kulchitsky, and R. Melko, *Quantum Boltzmann Machine*, *Phys. Rev. X* **8**, 021050 (2018).
- [8] M. Kieferová and N. Wiebe, *Tomography and Generative Training with Quantum Boltzmann Machines*, *Phys. Rev. A* **96**, 062327 (2017).
- [9] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, *Quantum Machine Learning*, *Nature (London)* **549**, 195 (2017).
- [10] D. Crawford, A. Levit, N. Ghadermarzy, J. S. Oberoi, and P. Ronagh, *Reinforcement Learning Using Quantum Boltzmann Machines*, *Quantum Inf. Comput.* **18**, 51 (2018), <https://dl.acm.org/doi/10.5555/3370185.3370188>.
- [11] E. Bairey, I. Arad, and N. H. Lindner, *Learning a Local Hamiltonian from Local Measurements*, *Phys. Rev. Lett.* **122**, 020504 (2019).
- [12] G. Torlai and R. G. Melko, *Machine-Learning Quantum States in the NISQ Era*, *Annu. Rev. Condens. Matter Phys.* **11**, 325 (2020).
- [13] N.-H. Chia, A. Gilyén, T. Li, H.-H. Lin, E. Tang, and C. Wang, *Sampling-Based Sublinear Low-Rank Matrix Arithmetic Framework for Dequantizing Quantum Machine Learning*, in *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020* (Association for Computing Machinery, New York, 2020), pp. 387–400.
- [14] A. Anshu, S. Arunachalam, T. Kuwahara, and M. Soleimanifar, *Sample-Efficient Learning of Quantum Many-Body Systems*, [arXiv:2004.07266](https://arxiv.org/abs/2004.07266).
- [15] F. G. S. L. Brandão and K. M. Svore, *Quantum Speed-Ups for Solving Semidefinite Programs*, in *Proceedings of the 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, New York, 2017), pp. 415–426.
- [16] F. G. S. L. Brandão, A. Kalev, T. Li, C. Y.-Yu. Lin, K. M. Svore, and X. Wu, *Quantum SDP Solvers: Large Speed-Ups, Optimality, and Applications to Quantum Learning*, in *Proceedings of the 46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)* (Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Patras, Greece, 2019), pp. 27:1–27:14.
- [17] J. van Apeldoorn, A. Gilyén, S. Gribling, and R. de Wolf, *Quantum SDP-Solvers: Better Upper and Lower Bounds*, *Quantum* **4**, 230 (2020).
- [18] J. Eisert, M. Cramer, and M. B. Plenio, *Colloquium: Area Laws for the Entanglement Entropy*, *Rev. Mod. Phys.* **82**, 277 (2010).
- [19] G. Vidal, J. I. Latorre, E. Rico, and A. Kitaev, *Entanglement in Quantum Critical Phenomena*, *Phys. Rev. Lett.* **90**, 227902 (2003).
- [20] M. B. Hastings, *An Area Law for One-Dimensional Quantum Systems*, *J. Stat. Mech.* (2007) P08024.
- [21] D. Aharonov, I. Arad, U. Vazirani, and Z. Landau, *The Detectability Lemma and Its Applications to Quantum Hamiltonian Complexity*, *New J. Phys.* **13**, 113043 (2011).
- [22] I. Arad, Z. Landau, and U. Vazirani, *Improved One-Dimensional Area Law for Frustration-Free Systems*, *Phys. Rev. B* **85**, 195145 (2012).
- [23] I. Arad, A. Kitaev, Z. Landau, and U. Vazirani, *An Area Law and Sub-exponential Algorithm for 1D Systems*, [arXiv:1301.1162](https://arxiv.org/abs/1301.1162).
- [24] F. G. S. L. Brandão and M. Horodecki, *An Area Law for Entanglement from Exponential Decay of Correlations*, *Nat. Phys.* **9**, 721 (2013).
- [25] N. Abrahamsen, *A Polynomial-Time Algorithm for Ground States of Spin Trees*, [arXiv:1907.04862](https://arxiv.org/abs/1907.04862).
- [26] A. Anshu, I. Arad, and D. Gosset, *Entanglement Sub-volume Law for 2D Frustration-Free Spin Systems*, in *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, Chicago, STOC 2020* (Association for Computing Machinery, New York, 2020), pp. 868–874.

- [27] T. Kuwahara and K. Saito, *Area Law of Noncritical Ground States in 1D Long-Range Interacting Systems*, *Nat. Commun.* **11**, 4478 (2020).
- [28] M. M. Wolf, F. Verstraete, M. B. Hastings, and J. I. Cirac, *Area Laws in Quantum Systems: Mutual Information and Correlations*, *Phys. Rev. Lett.* **100**, 070502 (2008).
- [29] M. B. Hastings, *Quantum Belief Propagation: An Algorithm for Thermal Quantum Systems*, *Phys. Rev. B* **76**, 201102(R) (2007).
- [30] M. Žnidarič, T. Prosen, and I. Pižorn, *Complexity of Thermal States in Quantum Spin Chains*, *Phys. Rev. A* **78**, 022103 (2008).
- [31] H. Bernigau, M. J. Kastoryano, and J. Eisert, *Mutual Information Area Laws for Thermal Free Fermions*, *J. Stat. Mech.* (2015) P02008.
- [32] T. Barthel, *One-Dimensional Quantum Systems at Finite Temperatures Can Be Simulated Efficiently on Classical Computers*, [arXiv:1708.09349](https://arxiv.org/abs/1708.09349).
- [33] S. R. White, *Minimally Entangled Typical Quantum States at Finite Temperature*, *Phys. Rev. Lett.* **102**, 190601 (2009).
- [34] E. M. Stoudenmire and S. R. White, *Minimally Entangled Typical Thermal State Algorithms*, *New J. Phys.* **12**, 055026 (2010).
- [35] M. Binder and T. Barthel, *Minimally Entangled Typical Thermal States versus Matrix Product Purifications for the Simulation of Equilibrium States and Time Evolution*, *Phys. Rev. B* **92**, 125119 (2015).
- [36] M. Rigol, T. Bryant, and R. R. P. Singh, *Numerical Linked-Cluster Approach to Quantum Lattice Models*, *Phys. Rev. Lett.* **97**, 187202 (2006).
- [37] B. Tang, E. Khatami, and M. Rigol, *A Short Introduction to Numerical Linked-Cluster Expansions*, *Comput. Phys. Commun.* **184**, 557 (2013).
- [38] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, *Quantum Monte Carlo Simulations of Solids*, *Rev. Mod. Phys.* **73**, 33 (2001).
- [39] A. W. Sandvik and J. Kurkijärvi, *Quantum Monte Carlo Simulation Method for Spin Systems*, *Phys. Rev. B* **43**, 5950 (1991).
- [40] U. Schollwöck, *The Density-Matrix Renormalization Group*, *Rev. Mod. Phys.* **77**, 259 (2005).
- [41] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac, *Matrix Product Density Operators: Simulation of Finite-Temperature and Dissipative Systems*, *Phys. Rev. Lett.* **93**, 207204 (2004).
- [42] M. Zwolak and G. Vidal, *Mixed-State Dynamics in One-Dimensional Quantum Lattice Systems: A Time-Dependent Superoperator Renormalization Algorithm*, *Phys. Rev. Lett.* **93**, 207205 (2004).
- [43] R. Orús and G. Vidal, *Infinite Time-Evolving Block Decimation Algorithm beyond Unitary Evolution*, *Phys. Rev. B* **78**, 155117 (2008).
- [44] W. Li, S.-J. Ran, S.-S. Gong, Y. Zhao, B. Xi, F. Ye, and G. Su, *Linearized Tensor Renormalization Group Algorithm for the Calculation of Thermodynamic Properties of Quantum Lattice Models*, *Phys. Rev. Lett.* **106**, 127202 (2011).
- [45] J. Haegeman, J. I. Cirac, T. J. Osborne, I. Pižorn, H. Verschelde, and F. Verstraete, *Time-Dependent Variational Principle for Quantum Lattices*, *Phys. Rev. Lett.* **107**, 070601 (2011).
- [46] P. Czarnik and J. Dziarmaga, *Variational Approach to Projected Entangled Pair States at Finite Temperature*, *Phys. Rev. B* **92**, 035152 (2015).
- [47] J. Haegeman, C. Lubich, I. Oseledets, B. Vandereycken, and F. Verstraete, *Unifying Time Evolution and Optimization with Matrix Product States*, *Phys. Rev. B* **94**, 165116 (2016).
- [48] B.-B. Chen, L. Chen, Z. Chen, W. Li, and A. Weichselbaum, *Exponential Thermal Tensor Network Approach for Quantum Lattice Models*, *Phys. Rev. X* **8**, 031082 (2018).
- [49] D. Poulin and E. Bilgin, *Belief Propagation Algorithm for Computing Correlation Functions in Finite-Temperature Quantum Many-Body Systems on Loopy Graphs*, *Phys. Rev. A* **77**, 052318 (2008).
- [50] E. Bilgin and D. Poulin, *Coarse-Grained Belief Propagation for Simulation of Interacting Quantum Systems at All Temperatures*, *Phys. Rev. B* **81**, 054106 (2010).
- [51] D. Poulin and P. Wocjan, *Sampling from the Thermal Quantum Gibbs State and Evaluating Partition Functions with a Quantum Computer*, *Phys. Rev. Lett.* **103**, 220502 (2009).
- [52] E. Bilgin and S. Boixo, *Preparing Thermal States of Quantum Systems by Dimension Reduction*, *Phys. Rev. Lett.* **105**, 170405 (2010).
- [53] K. Temme, T. J. Osborne, K. G. Vollbrecht, D. Poulin, and F. Verstraete, *Quantum Metropolis Sampling*, *Nature (London)* **471**, 87 (2011).
- [54] M.-H. Yung and A. Aspuru-Guzik, *A Quantum–Quantum Metropolis Algorithm*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 754 (2012).
- [55] Y. Ge, A. Molnár, and J. I. Cirac, *Rapid Adiabatic Preparation of Injective Projected Entangled Pair States and Gibbs States*, *Phys. Rev. Lett.* **116**, 080503 (2016).
- [56] M. J. Kastoryano and F. G. S. L. Brandão, *Quantum Gibbs Samplers: The Commuting Case*, *Commun. Math. Phys.* **344**, 915 (2016).
- [57] A. N. Chowdhury and R. D. Somma, *Quantum Algorithms for Gibbs Sampling and Hitting-Time Estimation*, *Quantum Inf. Comput.* **17**, 41 (2017), <https://dl.acm.org/doi/10.5555/3179483.3179486>.
- [58] F. G. S. L. Brandão and M. J. Kastoryano, *Finite Correlation Length Implies Efficient Preparation of Quantum Thermal States*, *Commun. Math. Phys.* **365**, 1 (2019).
- [59] A. Gilyén, Y. Su, .H. Low, and N. Wiebe, *Quantum Singular Value Transformation and beyond: Exponential Improvements for Quantum Matrix Arithmetics*, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, Phoenix, AZ, STOC 2019* (Association for Computing Machinery, New York, 2019), pp. 193–204.
- [60] H. Araki, *Gibbs States of a One Dimensional Quantum Lattice*, *Commun. Math. Phys.* **14**, 120 (1969).
- [61] L. Gross, *Decay of Correlations in Classical Lattice Models at High Temperature*, *Commun. Math. Phys.* **68**, 9 (1979).
- [62] Y. M. Park and H. J. Yoo, *Uniqueness and Clustering Properties of Gibbs States for Classical and Quantum Unbounded Spin Systems*, *J. Stat. Phys.* **80**, 223 (1995).



- [63] D. Ueltschi, *Cluster Expansions and Correlation Functions*, *Moscow Math. J.* **4**, 511 (2004), <https://doi.org/10.17323/1609-4514-2004-4-2-511-522>.
- [64] M. Kliesch, C. Gogolin, M. J. Kastoryano, A. Riera, and J. Eisert, *Locality of Temperature*, *Phys. Rev. X* **4**, 031019 (2014).
- [65] J. Fröhlich and D. Ueltschi, *Some Properties of Correlations of Quantum Lattice Systems in Thermal Equilibrium*, *J. Math. Phys. (N.Y.)* **56**, 053302 (2015).
- [66] M. Lenci and L. Rey-Bellet, *Large Deviations in Quantum Lattice Systems: One-Phase Region*, *J. Stat. Phys.* **119**, 715 (2005).
- [67] K. Netočný and F. Redig, *Large Deviations for Quantum Spin Systems*, *J. Stat. Phys.* **117**, 521 (2004).
- [68] T. Kuwahara and K. Saito, *Gaussian Concentration Bound and Ensemble Equivalence in Generic Quantum Many-Body Systems including Long-Range Interactions*, *Ann. Phys. (Amsterdam)* **421**, 168278 (2020).
- [69] K. Kato and F. G. S. L. Brandão, *Quantum Approximate Markov Chains are Thermal*, *Commun. Math. Phys.* **370**, 117 (2019).
- [70] T. Kuwahara, K. Kato, and F. G. S. L. Brandão, *Clustering of Conditional Mutual Information for Quantum Gibbs States above a Threshold Temperature*, *Phys. Rev. Lett.* **124**, 220601 (2020).
- [71] A. W. Harrow, S. Mehraban, and M. Soleimanifar, *Classical Algorithms, Correlation Decay, and Complex Zeros of Partition Functions of Quantum Many-Body Systems*, in *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020* (Association for Computing Machinery, New York, 2020), pp. 378–386.
- [72] E. Crosson and S. Slezak, *Classical Simulation of High Temperature Quantum Ising Models*, [arXiv:2002.02232](https://arxiv.org/abs/2002.02232).
- [73] R. L. Mann and T. Helmuth, *Efficient Algorithms for Approximating Quantum Partition Functions*, [arXiv:2004.11568](https://arxiv.org/abs/2004.11568).
- [74] F. Barahona, *On the Computational Complexity of Ising Spin Glass Models*, *J. Phys. A* **15**, 3241 (1982).
- [75] L. A. Goldberg and M. Jerrum, *A Complexity Classification of Spin Systems with an External Field*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13161 (2015).
- [76] More in detail, we mean the following statement. Let  $Z$  be a partition function as  $Z := \text{tr}(e^{-\beta H})$  with  $H$  the system Hamiltonian. Then, there exists a constant  $\zeta > 0$  such that approximating  $\log(Z)$  up to an error  $\zeta n$  ( $n$ , system size) is the NP-hard problem [75].
- [77] M. Jerrum and A. Sinclair, *Polynomial-Time Approximation Algorithms for the Ising Model*, *SIAM J. Comput.* **22**, 1087 (1993).
- [78] S. Bravyi and D. Gosset, *Polynomial-Time Classical Simulation of Quantum Ferromagnets*, *Phys. Rev. Lett.* **119**, 100503 (2017).
- [79] H. Lamm and S. Lawrence, *Simulation of Nonequilibrium Dynamics on a Quantum Computer*, *Phys. Rev. Lett.* **121**, 170501 (2018).
- [80] M. J. S. Beach, R. G. Melko, T. Grover, and T. H. Hsieh, *Making Trotters Sprint: A Variational Imaginary Time Ansatz for Quantum Many-Body Systems*, *Phys. Rev. B* **100**, 094434 (2019).
- [81] X. Yuan, S. Endo, Q. Zhao, Y. Li, and S. C. Benjamin, *Theory of Variational Quantum Simulation*, *Quantum* **3**, 191 (2019).
- [82] S. McArdle, T. Jones, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, *Variational Ansatz-Based Quantum Simulation of Imaginary Time Evolution*, *npj Quantum Inf.* **5**, 75 (2019).
- [83] K. Yeter-Aydeniz, R. C. Pooser, and G. Siopsis, *Practical Quantum Computation of Chemical and Nuclear Energy Levels Using Quantum Imaginary Time Evolution and Lanczos Algorithms*, *npj Quantum Inf.* **6**, 63 (2020).
- [84] P. J. Love, *Cooling with Imaginary Time*, *Nat. Phys.* **16**, 130 (2020).
- [85] H. Nishi, T. Kosugi, and Yu.-i. Matsushita, *Implementation of Quantum Imaginary-Time Evolution Method on NISQ Devices: Nonlocal Approximation*, [arXiv:2005.12715](https://arxiv.org/abs/2005.12715).
- [86] Above this threshold temperature, one can ensure that there are no quantum or classical phase transitions.
- [87] D. Gottesman and M. B. Hastings, *Entanglement versus Gap for One-Dimensional Spin Systems*, *New J. Phys.* **12**, 025002 (2010).
- [88] D. Aharonov, A. W. Harrow, Z. Landau, D. Nagaj, M. Szegedy, and U. Vazirani, *Local Tests of Global Entanglement and a Counterexample to the Generalized Area Law*, in *Proceedings of the 2014 IEEE 55th Annual Symposium on Foundations of Computer Science* (IEEE, New York, 2014), pp. 246–255.
- [89] S. Bravyi, *Upper Bounds on Entangling Rates of Bipartite Hamiltonians*, *Phys. Rev. A* **76**, 052319 (2007).
- [90] J. Eisert and T. J. Osborne, *General Entanglement Scaling Laws from Time Evolution*, *Phys. Rev. Lett.* **97**, 150404 (2006).
- [91] K. Van Acoleyen, M. Mariën, and F. Verstraete, *Entanglement Rates and Area Laws*, *Phys. Rev. Lett.* **111**, 170501 (2013).
- [92] M. Mariën, K. M. R. Audenaert, K. Van Acoleyen, and F. Verstraete, *Entanglement Rates and the Stability of the Area Law for the Entanglement Entropy*, *Commun. Math. Phys.* **346**, 35 (2016).
- [93] S. Sachdeva and N. K. Vishnoi, *Faster Algorithms via Approximation Theory*, *Found. Trends Theor. Comput. Sci.* **9**, 125 (2014).
- [94] M. B. Hastings, *Solving Gapped Hamiltonians Locally*, *Phys. Rev. B* **73**, 085115 (2006).
- [95] A. Molnar, N. Schuch, F. Verstraete, and J. I. Cirac, *Approximating Gibbs States of Local Hamiltonians Efficiently with Projected Entangled Pair States*, *Phys. Rev. B* **91**, 045138 (2015).
- [96] T. J. Osborne, *Hamiltonian Complexity*, *Rep. Prog. Phys.* **75**, 022001 (2012).
- [97] S. Gharibian, Y. Huang, Z. Landau, S. W. Shin *et al.*, *Quantum Hamiltonian Complexity*, *Found. Trends Theor. Comput. Sci.* **10**, 159 (2015).
- [98] Z. Landau, U. Vazirani, and T. Vidick, *A Polynomial Time Algorithm for the Ground State of One-Dimensional Gapped Local Hamiltonians*, *Nat. Phys.* **11**, 566 (2015).
- [99] I. Arad, Z. Landau, U. Vazirani, and T. Vidick, *Rigorous RG Algorithms and Area Laws for Low Energy Eigenstates in 1D*, *Commun. Math. Phys.* **356**, 65 (2017).



- [100] G. De las Cuevas, N. Schuch, D. Pérez-García, and J. I. Cirac, *Purifications of Multipartite States: Limitations and Constructive Methods*, *New J. Phys.* **15**, 123021 (2013).
- [101] G. De las Cuevas, T. S. Cubitt, J. I. Cirac, M. M. Wolf, and D. Pérez-García, *Fundamental Limitations in the Purifications of Tensor Networks*, *J. Math. Phys. (N.Y.)* **57**, 071902 (2016).
- [102] C. Schön, E. Solano, F. Verstraete, J. I. Cirac, and M. M. Wolf, *Sequential Generation of Entangled Multiqubit States*, *Phys. Rev. Lett.* **95**, 110503 (2005).
- [103] J.-G. Liu, Y.-H. Zhang, Y. Wan, and L. Wang, *Variational Quantum Eigensolver with Fewer Qubits*, *Phys. Rev. Research* **1**, 023025 (2019).
- [104] S.-J. Ran, *Encoding of Matrix Product States into Quantum Circuits of One- and Two-Qubit Gates*, *Phys. Rev. A* **101**, 032310 (2020).
- [105] We start from the basic formula for the Chebyshev polynomials:  $xT_r(x) = [T_{r+1}(x) + T_{r-1}(x)]/2$ . Let  $Y$  be a random variable taking values 1 or  $-1$  with the probability  $1/2$ . We then obtain  $xT_r(x) = \mathbb{E}_{Y_1}[T_{r+Y_1}(x)] = [T_{r+1}(x) + T_{r-1}(x)]/2$ . In the same way, we can obtain  $x^2T_r(x) = \mathbb{E}_{Y_1}[xT_{r+Y_1}(x)] = \mathbb{E}_{Y_1, Y_2}[T_{r+Y_1+Y_2}(x)]$ . By repeating the process, we obtain  $x^jT_r(x) = \mathbb{E}_{Y_1, \dots, Y_j}[T_{r+D_j}(x)]$  with  $D_j = Y_1 + Y_2 + \dots + Y_j$ . The probability distribution of  $D_j$  obeys the binomial distribution as  $2^{-j} \binom{j}{(j+D_j)/2}$ , which gives Eq. (19).
- [106] M. Fannes, *A Continuity Property of the Entropy Density for Spin Lattice Systems*, *Commun. Math. Phys.* **31**, 291 (1973).
- [107] M. B. Hastings, *Entropy and Entanglement in Quantum Ground States*, *Phys. Rev. B* **76**, 035114 (2007).
- [108] L. Masanes, *Area Law for the Entropy of Low-Energy States*, *Phys. Rev. A* **80**, 052104 (2009).
- [109] T. Kuwahara and K. Saito, *Eigenstate Thermalization from the Clustering Property of Correlation*, *Phys. Rev. Lett.* **124**, 200604 (2020).
- [110] J. Haah, M. Hastings, R. Kothari, and G. H. Low, *Quantum Algorithm for Simulating Real Time Evolution of Lattice Hamiltonians*, in *Proceedings of the 2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, New York, 2018), pp. 350–360.
- [111] G. Vidal, *Efficient Simulation of One-Dimensional Quantum Many-Body Systems*, *Phys. Rev. Lett.* **93**, 040502 (2004).
- [112] U. Schollwöck, *The Density-Matrix Renormalization Group in the Age of Matrix Product States*, *Ann. Phys. (Amsterdam)* **326**, 96 (2011).
- [113] F. Verstraete and J. I. Cirac, *Matrix Product States Represent Ground States Faithfully*, *Phys. Rev. B* **73**, 094423 (2006).
- [114] B. M. Terhal, M. Horodecki, D. W. Leung, and D. P. DiVincenzo, *The Entanglement of Purification*, *J. Math. Phys. (N.Y.)* **43**, 4286 (2002).
- [115] J. G. Jarkovský, A. Molnár, N. Schuch, and J. I. Cirac, *Efficient Description of Many-Body Systems with Matrix Product Density Operators*, *Phys. Rev. X Quantum* **1**, 010304 (2020).
- [116] S. Bagchi and A. K. Pati, *Monogamy, Polygamy, and Other Properties of Entanglement of Purification*, *Phys. Rev. A* **91**, 042323 (2015).
- [117] M. Berta, F. G. S. L. Brandão, J. Haegeman, V. B. Scholz, and F. Verstraete, *Thermal States as Convex Combinations of Matrix Product States*, *Phys. Rev. B* **98**, 235154 (2018).
- [118] E. Leviatan, F. Pollmann, J. H. Bardarson, D. A. Huse, and E. Altman, *Quantum Thermalization Dynamics with Matrix-Product States*, [arXiv:1702.08894](https://arxiv.org/abs/1702.08894).
- [119] At the inequality (9) in Ref. [114], the upper bound of  $E_{f,\alpha}(\sigma) \leq E_{p,\alpha}(\sigma)$  is obtained for the case of  $\alpha = 1$ . However, the proof therein can be easily extended to generic  $\alpha$ .
- [120] T. J. Osborne, *Efficient Approximation of the Dynamics of One-Dimensional Quantum Spin Systems*, *Phys. Rev. Lett.* **97**, 157202 (2006).
- [121] M. B. Hastings, *Observations outside the Light Cone: Algorithms for Nonequilibrium and Thermal States*, *Phys. Rev. B* **77**, 144302 (2008).
- [122] A. M. Childs, Y. Su, M. C. Tran, N. Wiebe, and S. Zhu, *Theory of Trotter Error with Commutator Scaling*, *Phys. Rev. X* **11**, 011020 (2021)..
- [123] D. Sutter, *Approximate Quantum Markov Chains*, in *Approximate Quantum Markov Chains*, Springer Briefs in Mathematical Physics Vol. 28 (Springer, Cham, 2018), pp. 75–100.
- [124] C. Eckart and G. Young, *The Approximation of One Matrix by Another of Lower Rank*, *Psychometrika* **1**, 211 (1936).
- [125] T. Kuwahara, T. Mori, and K. Saito, *Floquet-Magnus Theory and Generic Transient Dynamics in Periodically Driven Many-Body Quantum Systems*, *Ann. Phys. (Amsterdam)* **367**, 96 (2016).
- [126] E. H. Lieb and D. W. Robinson, *The Finite Group Velocity of Quantum Spin Systems*, *Commun. Math. Phys.* **28**, 251 (1972).
- [127] S. Bravyi, M. B. Hastings, and F. Verstraete, *Lieb-Robinson Bounds and the Generation of Correlations and Topological Quantum Order*, *Phys. Rev. Lett.* **97**, 050401 (2006).
- [128] R. Alicki and M. Fannes, *Continuity of Quantum Conditional Information*, *J. Phys. A* **37**, L55 (2004).