Approximating Gibbs states of local Hamiltonians efficiently with projected entangled pair states

Andras Molnar,[1] Norbert Schuch,[2] Frank Verstraete,[3] and J. Ignacio Cirac[1]
[1]*Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Strasse 1, D-85748 Garching, Germany*
[2]*JARA Institute for Quantum Information, RWTH Aachen University, D-52056 Aachen, Germany*
[3]*Faculty of Physics, University of Vienna, Boltzmanngasse 5, 1090 Vienna, Austria*
and Department of Physics and Astronomy, Ghent University, Ghent, Belgium
(Received 29 September 2014; revised manuscript received 23 December 2014; published 29 January 2015)

We analyze the error of approximating Gibbs states of local quantum spin Hamiltonians on lattices with projected entangled pair states (PEPS) as a function of the bond dimension (D), temperature (β^{-1}), and system size (N). First, we introduce a compression method in which the bond dimension scales as $D = e^{O(\log_2^2(N/\epsilon))}$ if $\beta < O(\log_2 N)$. Second, building on the work of Hastings [M. B. Hastings, Phys. Rev. B **73**, 085115 (2006)], we derive a polynomial scaling relation, $D = (N/\epsilon)^{O(\beta)}$. This implies that the manifold of PEPS forms an efficient representation of Gibbs states of local quantum Hamiltonians. From those bounds it also follows that ground states can be approximated with $D = N^{O(\log_2 N)}$ whenever the density of states only grows polynomially in the system size. All results hold for any spatial dimension of the lattice.

DOI: 10.1103/PhysRevB.91.045138 PACS number(s): 75.10.Jm, 03.65.Ud

I. INTRODUCTION

Problems dealing with quantum many-body systems in lattices appear very often in different branches of physics and chemistry. They typically correspond to discretized versions of first-principles continuum models, such as in high-energy physics, atomic physics, or quantum chemistry, or provide a phenomenological description of a complex system, as in condensed matter physics. They are characterized in terms of a lattice Hamiltonian, H, which describes the motion as well as the interactions among the different constituents. Apart from generating the dynamics via the Schrödinger equation, the Hamiltonian defines the quantum state of the system in thermal equilibrium through the Gibbs density operator,

$$\rho = \frac{e^{-\beta H}}{Z} = \frac{e^{-\beta H}}{\mathrm{tr}[e^{-\beta H}]}, \tag{1}$$

where Z is the partition function and $\beta = 1/\kappa_B T$ is the inverse temperature (we set the Bolzmann constant $\kappa_B = 1$). This operator encodes all the (static) physical properties of our systems. Extracting that information becomes a hard problem, even for systems consisting of very few particles. The reason is that, in order to determine expectation values of observables, we have to express ρ in a basis of the corresponding Hilbert space, and the dimension of the latter grows exponentially with the number of lattice sites, N (i.e., volume) of the lattice. This fact is ultimately related to the tensor product structure inherent in quantum mechanical problems dealing with composite objects, and thus ubiquitous in several branches of science.

There exist different ways around this problem, at least in some specific situations. For instance, one can employ sampling techniques in certain models (not suffering from the sign problem), to accurately determine the physical properties of a system in thermal equilibrium. Alternatively, one can restrict oneself to simple tractable families of states depending on few parameters, which can then be determined by variational techniques. This last approach typically requires a good intuition to select which family will encompass all the physical properties that one has to describe, and can easily lead to either wrong or inaccurate results. Yet another approach is that of quantum simulation, where the Hamiltonian of interest is implemented on a different system on which one has enough control [1].

Strictly speaking, the exponential scaling of the dimension of the Hilbert space with the size of the lattice should not be the ultimate reason for the difficulty of quantum many-body problems, at least for the ones that naturally appear in nature. For instance, if H is the sum of terms acting nontrivially only on at most x lattice sites, then we can characterize all possible Hamiltonians with a number of parameters that scales only polynomially with N. If those terms are local, meaning that the distance between the sites on which term of H acts is bounded by a constant, this scale is even linear in N. Thus, for all those problems, ρ itself only depends on few parameters. One says that the states can only explore a very small "corner" of the Hilbert space [2]. Consequently, it may be possible to utilize this fact to find families of states that describe all possible many-body lattice problems with x-body interactions in thermal equilibrium, and that depend on a number of parameters that only grows polynomially with N. Thus, a central problem in this context is to find and characterize such a family of states. A first and fundamental step would be to solve that problem for local Hamiltonians, on which we will concentrate in the following.

Matrix product states (MPS) [3,4] provide the answer for one-dimensional models at zero temperature for both gapped [5,6] and critical models [7]. Specifically, if Ψ_0 is the ground state of such a Hamiltonian there exists a MPS of bond dimension D, Ψ_{MPS}, such that $\|\Psi_0 - \Psi_{\mathrm{MPS}}\| < \epsilon$ with $D = O(\mathrm{poly}(N/\epsilon))$. Note that, in turn, the number of parameters to characterize the MPS scales polynomially with D. This result is strongly connected to the area law [8,9], which is fulfilled (or only slightly violated) for those models and MPS. In higher dimensions and still at zero temperature, it is conjectured (and proven under certain assumptions [10,11]), that the area law still holds (with logarithmic corrections for certain critical models [12,13]). In that case, one would expect that the projected entangled pair states (PEPS) [14,15], which extend MPS to higher dimensions, would provide us with the efficient description of that corner of the Hilbert

space [2]. Moreover, for any finite temperature (independent of N), an area law has been proven [16] both for Gibbs states (1), as well as for projected entangled pair operators (PEPOs), the extension of PEPS to mixed sates. This also suggests that PEPOs can efficiently describe Gibbs states of local Hamiltonians. From the physics point of view, this is actually the relevant question, as any extended system can only be cooled down to a certain temperature independent of the system size.

Hastings [17] has already derived some remarkable results addressing that question. He has shown that in d spatial dimensions, one can build a PEPO, ρ_{PEPO}, such that $||\rho - \rho_{\mathrm{PEPO}}||_1 < \epsilon$ with bond dimension scaling as

$$D = e^{O(\beta \log_2(N/\epsilon)^d)}. \tag{2}$$

This gives a polynomial scaling for one dimension, and a subexponential (although superpolynomial) one for higher ones. This result also implies a bound for the approximation of the ground state. In fact, if H is gapped and the density of states for a fixed energy only grows as poly(N), then choosing $\beta = O(\log_2 N)$ in (1) we obtain a state that is as close as we want to the ground state [18]. This means that, under those conditions, we can find a PEPS approximation of the ground state with

$$D = e^{O(\log_2(N/\epsilon)^{d+1})}. \tag{3}$$

In the present paper we derive the following results. First, we use a method based on the Trotter expansion to obtain a bound for $\beta \leqslant O(\log_2 N)$ independent of the dimension (although still superpolynomial in N),

$$D = e^{O(\log_2^2(N/\epsilon))}. \tag{4}$$

Under the same condition on the density of states as before, we also obtain that the ground state can be approximated with

$$D = e^{O(\log_2^2(N/\epsilon))}, \tag{5}$$

independent of the dimension. Finally, using Hastings' construction of the PEPO (see also [19]), with the help of an efficient encoding we show that it is possible to have a polynomial scaling for any temperature; i.e.,

$$D = (N/\epsilon)^{O(\beta d \log_2 d)}. \tag{6}$$

Even though Hastings' cluster expansion method gives better results at finite temperature than the first construction, that is more intuitive as it extends the case of commuting Hamiltonians.

The paper is organized as follows. In Sec. II we define the problem we are addressing in this work. Section III derives the bounds (4) and (5) using the technique based on the Trotter expansion. In Sec. IV we use a different encoding of the PEPO based on Hastings' construction to obtain the polynomial bound (6). In all these sections we quote the results and explain how we have proven them. In the appendices we give details of the proofs.

II. PROBLEM

We consider a growing sequence of finite spin systems, S_n, with two-body interactions. To every system, S_n, we assign a graph, $\mathcal{G}_n = (\mathcal{V}_n, \mathcal{E}_n)$, where the vertices $\mathcal{V}_n$ correspond to the individual spins and the edges $\mathcal{E}_n$ to interactions. The Hamiltonian is such that only the connected points interact:

$$H_n = \sum_{e \in \mathcal{E}_n} h_e, \tag{7}$$

where h_e acts nontrivially on spins v and w if $e = (v,w)$. Even though for simplicity we have considered only nearest-neighbor interactions, the results generalize to local more-body interactions. We will assume that the (operator) norm of all the terms in the Hamiltonians is bounded by 1; i.e., $\|h_i\| \leqslant 1$. If the norm of the Hamiltonians were bounded by J instead of 1, this factor could be included in the definition of the temperature.

We assume that all graphs are connected, and that their degree is uniformly bounded. That is, the number of edges starting from a given point is smaller than some constant z. This implies that $2|\mathcal{E}_n|/z < |\mathcal{V}_n| \leqslant |\mathcal{E}_n| + 1$. Thus, we can equally characterize the size of the system by the number of spins or interactions, $N = |\mathcal{V}_n|$ and $|\mathcal{E}_n|$, respectively. For convenience we will denote $|\mathcal{E}_n|$ by K and omit the index n in the following.

We also assume that there is a uniformly bounded lattice growth constant. This means that there is a universal constant, γ, such that for any given $e \in \mathcal{E}$ and all $l \in \mathbb{Z}^+$

$$|\{\mathcal{I} \subseteq \mathcal{E} | \mathcal{I} \text{ connected}, e \in \mathcal{I}, |\mathcal{I}| = l\}| \leqslant \gamma^l. \tag{8}$$

That is, the number of connected regions having l edges that include a specific edge, e, grows at most exponentially with l. In particular, this is the case if $\mathcal{G}_n$ is a regular lattice in any spatial dimension [20]. Thus, our treatment includes all those cases.

We consider the Gibbs state corresponding to H given by (1). We will construct a PEPO, $\tilde{\rho}$, of bond dimension D, that is close to that state. In particular, for any $\varepsilon > 0$,

$$\|e^{-\beta H} - \tilde{\rho}\|_1 \leqslant \varepsilon \|e^{-\beta H}\|_1, \tag{9}$$

where $\|x\|_p = [\mathrm{tr}(x^\dagger x)^{p/2}]^{1/p}$ stands for the Schatten p norm ($\|x\| = \|x\|_\infty$ for the operator norm). We will be interested in how D scales with N (or equivalently, with K) and ε.

By a PEPO on a graph $\mathcal{G} = (\mathcal{E}, \mathcal{V})$ we mean that the operator $\tilde{\rho}$ admits the following form:

$$\tilde{\rho} = \sum_{\alpha: \mathcal{E} \to \{1,\ldots,D\}} \bigotimes_{v \in \mathcal{V}} X^v_{\alpha(e_1^v),\ldots,\alpha(e_{z(v)}^v)}. \tag{10}$$

Here, $X^v_{\alpha(e_1^v),\ldots,\alpha(e_{z(v)}^v)}$ are operators acting on the vertex v alone, $z(v)$ is the degree of v, and $e_1^v, \ldots, e_{z(v)}^v$ are the edges going through v. This definition is the straightforward generalization of PEPS [14] for operators [21,22]. One can readily see [2] that this operator can be written as a tensor network on the graph $\mathcal{G}$, where the bond dimension is D.

III. CONSTRUCTION BASED ON A TROTTER EXPANSION AND COMPRESSION

In this section we use a Trotter expansion combined with a compression method to approximate the Gibbs state. The intuition about why this expansion should give rise to a PEPO description is the following (see also [23]). Let us assume that the operators h_e commute with each other. Then, the Gibbs state (1) is proportional to a product of exponentials, each of

them of the form $e^{-\beta h_e}$. One can easily show that each term in that product creates a link in the PEPO [16]. The bond dimension, D_0, is simply the maximum number of singular values of h_e, when decomposed in terms of the vertices it connects, and thus it is independent of K and the temperature. In the general case where the h_e do not commute with each other, we can still perform a Trotter expansion and approximate ρ (up to a constant factor) by $(\tau^\dagger\tau)^M$ where

$$\tau = \prod_{i=1}^{K} e^{-\beta h_i/2M}. \tag{11}$$

The integer M has to be chosen such that the approximation is good; i.e.,

$$\|e^{-\beta H} - (\tau^\dagger\tau)^M\|_1 \leqslant \varepsilon \|e^{-\beta H}\|_1 \tag{12}$$

for some $\varepsilon > 0$. Now, if we use the same argument we see that each time we apply τ, we create a bond between each pair of vertices that are connected in the graph. That is, we multiply the bond dimension by D_0. Thus, naively, the final bond dimension will be D_0^{2M}, and since M has to grow polynomially with K, we get a very bad bound. However, for large M each of the terms in τ is close to the identity operator. Thus, this operator creates very little entanglement and it should be possible to compress the information that is contained in the bond variables for any pair of connected vertices, and therefore to decrease the bond dimension. In fact, in the case of commuting Hamiltonians one can reduce it to D_0, independent of M. This is, in fact, what we do in this section: we first find M such that (12) holds, and then we compress the bond to get a better scaling of the bond dimension with K.

More specifically, we write $e^{-\beta h_i/2M} = \mathbb{1} + (e^{-\beta h_i/2M} - \mathbb{1})$, then, after collecting the K terms of τ and $\tau^\dagger$ into one product of $2K$ terms, we obtain

$$(\tau^\dagger\tau)^M = \prod_{j=1}^{M}\prod_{i=1}^{2K} e^{-\beta \tilde{h}_i/2M} = \prod_{j=1}^{M}\prod_{i=1}^{2K}(1 + x_i), \tag{13}$$

where $\tilde{h}_i$ denotes h_{K+1-i} if $i \leqslant K$, and h_{i-K} otherwise, and $x_i = e^{-\beta \tilde{h}_i/2M} - \mathbb{1}$. After expanding the product, this operator takes the form

$$(\tau^\dagger\tau)^M = \sum_{\lambda \in \mathcal{M}^b_{M,2K}} \prod_{j=1}^{M}\prod_{i=1}^{2K} x_i^{\lambda_{i,j}}. \tag{14}$$

The sum runs over all $M \times 2K$ matrices with entries 0 or 1, denoted by $\mathcal{M}^b_{M,2K}$. From this sum we only keep those terms in which any given x_i appears at most L times in (14). In Sec. III B we show that the resulting operator $\tilde{\rho}$ is a good approximation to $(\tau^\dagger\tau)^M$ if $L \approx \log_2 K$.

In Sec. III C we show then that the resulting operator can be written as a PEPO, in the sense of (10), with bond dimension $M^{O(L)}$. The reason why this operator admits a PEPO form can be understood as follows. First we identify each particular term in the expansion of $(\tau^\dagger\tau)^M$ with the help of indices defined on the edges. This can be done by specifying at every edge, i, the position where x_{K+1-i} and/or x_{K+i} appear out of the M possibilities. Once a term is identified, we proceed with the Schmidt decomposition of that term in order to build the local operators $X^v_{\alpha(e_1^v),\ldots,\alpha(e^v_{z(v)})}$. Let us notice that the latter only depends on the order in which the operators $x_{e_1^v}$, $x_{e_2^v}$, ..., $x_{e^v_{z(v)}}$ appear in the given term, where $e_1^v, \ldots, e^v_{z(v)}$ are the edges starting from point v. This order can be obtained locally from the edges that surround v, which contain information about the x involved in each of them. As a result of that, at every edge we have to specify $\binom{M}{L}^2 \approx M^{2L}$ natural numbers. As $M = \text{poly}(K)$ and $L = O(\log_2 K)$, this gives a bond dimension $K^{O(\log_2 K)}$ for the approximating operator. Therefore, as $N \leqslant 2K/z$, we obtain a bond dimension that scales like $N^{O(\log_2 N)}$.

A. Trotter expansion

We know that $(\tau^\dagger\tau)^M$ [τ as in Eq. (11)] tends to $e^{-\beta H}$ if $M \to \infty$. The question is how big M has to be chosen such that we obtain a good approximation in one-norm. Although the Trotter expansion has been extensively investigated in the literature (see also [24]), we are interested in a bound on the one-norm instead of the operator norm. Here we prove that setting $M = 360\beta^2K^2/\varepsilon$ is enough for an error ε in one-norm.

We present the proof in two steps. First we show that $\|e^{-\beta H} - (\tau^\dagger\tau)^M\|_1$ is small compared to $\|e^{-\beta H}\|_1$ as long as $\|\eta - \tau\|_{2M}$ is small compared to $\|\eta\|_{2M}$ where $\eta = e^{-\beta H/2M}$. Second, we show that $\|\eta - \tau\|_{2M}$ is small compared to $\|e^{-\beta H/2M}\|_{2M}$. The key point is that both $e^{-\beta H}$ and $(\tau^\dagger\tau)$ are close to $(\mathbb{1} - \beta H/M)^M$. We state the first step as a proposition:

Proposition 1. If $\varepsilon < 1/3$ and

$$\|\eta - \tau\|_{2M} \leqslant \frac{\varepsilon}{M}\|\eta\|_{2M}, \tag{15}$$

then

$$\|\eta^{2M} - (\tau^\dagger\tau)^M\|_1 \leqslant 9\varepsilon\|\eta^{2M}\|_1. \tag{16}$$

The proof combines the identity $a^m - b^m = \sum_i a^i(a - b)b^{m-i-1}$ with the Hölder inequality for matrices [25] and it is presented in Appendix A. We state the second statement (that η is close to τ) as a lemma.

Lemma 1. If $M > 36\beta^2K^2/\epsilon$ and $\epsilon < 1$, then

$$\|\eta - \tau\|_{2M} \leqslant \frac{\epsilon}{M}\|\eta\|_{2M}.$$

The main idea is that it is enough to prove the statement for the operator norm, as $\|\eta - \tau\|_{2M}$ is bounded by the Hölder inequality

$$\|\eta - \tau\|_{2M} = \|\eta^{-1}\eta(\eta - \tau)\|_{2M} \leqslant \|\eta^{-1}\|\|\eta\|_{2M}\|\eta - \tau\|,$$

and $\|\eta^{-1}\|$ is not too big as η is close to the identity operator. In order to show that $\|\eta - \tau\|$ is close to zero, by a simple series expansion we obtain that $\|\eta - \mathbb{1} + \beta H/M\|$ is small and so is $\|\tau - \mathbb{1} + \beta H/M\|$. The statement then follows from the triangle inequality. The detailed proof is presented in Appendix B.

Putting together Proposition 1 and Lemma 1, we obtain that the Trotter approximation is ε-close (in one-norm) if the trotter steps are chosen to be $M > 360\beta^2K^2/\varepsilon$.

B. Compression

We approximate now $(\tau^\dagger\tau)^M$ by an operator $\tilde{\rho}$ starting from Eq. (14). This expansion can be pictured as follows. We can

think of the resulting operator as a sum:

$$(\tau^\dagger\tau)^M = \sum_{\text{all fillings}} \begin{array}{|c||c|c|c|c|}\hline & x_1 & x_2 & \dots & x_{2K} \\ \hline\hline 1 & X & & & \\ \hline 2 & & & X & \\ \hline \vdots & & X & & \\ \hline M & X & & & X \\ \hline\end{array}, \tag{17}$$

where the table can be understood as follows. We begin to read from the upper-left corner, from left to right, row by row. Whenever we meet an X in the actual cell, we write down the corresponding operator x_i (according to the column), and otherwise the identity operator. The value assigned to a given table is then the product of those operators. We finally have to sum up the resulting operators for all possible fillings of the table.

The approximating operator $\tilde{\rho}$ can be thought of in the same way, just limiting the number of X's in each of the columns.

$$\tilde{\rho} = \sum_{\substack{\text{filling per} \\ \text{column} \leqslant L}} \begin{array}{|c||c|c|c|c|}\hline & x_1 & x_2 & \dots & x_{2K} \\ \hline\hline 1 & X & & & \\ \hline 2 & & & X & \\ \hline \vdots & & X & & \\ \hline M & X & & & X \\ \hline\end{array}. \tag{18}$$

We want to prove that this is a good approximation: $\|(\tau^\dagger\tau)^M - \tilde{\rho}\|_1 \leqslant \varepsilon\|e^{-\beta H}\|_1$ if the maximal number of X's per column, L, is chosen big enough. We will show that $L = O(\log_2 K)$ is enough.

Let us first explain the main idea of the proof. Given a set of columns $\mathcal{I} \subseteq \{1,2,\dots,K\}$, define $S(\mathcal{I})$ to be the sum of all tables containing more than L X's in all columns $i \in \mathcal{I}$, but with no restriction for the columns not belonging to $\mathcal{I}$. Formally, let $\mathcal{Q}(\mathcal{I})$ denote the set of these tables:

$$\mathcal{Q}(\mathcal{I}) = \left\{ \lambda \in \mathcal{M}^b_{M,2K} \mid i \in \mathcal{I} \Rightarrow \sum_j \lambda_{i,j} > L \right\},$$

then $S(\mathcal{I})$ is the sum

$$S(\mathcal{I}) = \sum_{\lambda\in\mathcal{Q}(\mathcal{I})} \prod_{j=1}^{M}\prod_{i=1}^{2K} x_i^{\lambda_{i,j}}. \tag{19}$$

In any column that has no restriction, the sum can be evaluated, giving back $e^{-\beta\tilde{h}_i/2M}$ in every row of that column. By evaluating those sums we arrive at a sum containing only a few terms. In these remaining terms still a large number of X's appear; therefore the norm of each such term is small. Thus the one-norm of $S(\mathcal{I})$ can be bounded. We will express $\tilde{\rho}$ with the help of the sums $S(\mathcal{I})$ in order to be able to bound its norm.

We use this observation in order to upper bound the one-norm of $(\tau^\dagger\tau)^M - \tilde{\rho}$. That difference contains one or more columns where there are more than L appearances of X. We regroup the tables as follows. First, given a set of columns, $\mathcal{I}$, we sum up all tables that have more than L appearances of X in the columns $i \in \mathcal{I}$, albeit at most L in all columns $i \notin \mathcal{I}$. This set of tables is the following set:

$$\mathcal{T}(\mathcal{I}) = \left\{ \lambda \in \mathcal{M}^b_{M,2K} \mid \sum_j \lambda_{i,j} > L \Leftrightarrow i \in \mathcal{I} \right\}.$$

The sum of these tables will be called $R(\mathcal{I})$:

$$R(\mathcal{I}) = \sum_{\lambda\in\mathcal{T}(\mathcal{I})} \prod_{j=1}^{M}\prod_{i=1}^{2K} x_i^{\lambda_{i,j}}. \tag{20}$$

Note that the operator $\tilde{\rho}$ is expressed by $R(\emptyset)$, as $\tilde{\rho}$ is the sum of tables that in each column contain at most L X's.

We can express the sum $S(\mathcal{I})$ with the help of $R(\mathcal{I})$:

$$S(\mathcal{I}) = \sum_{\mathcal{J}\supseteq\mathcal{I}} R(\mathcal{J}), \tag{21}$$

because in any table in $S(\mathcal{I})$, the columns containing more than L X's form a set $\mathcal{J} \supseteq \mathcal{I}$. Note that $(\tau^\dagger\tau)^M = S(\emptyset)$, as $(\tau^\dagger\tau)^M$ contains all tables, with no restriction on the number of X's in any column.

The difference $(\tau^\dagger\tau)^M - \tilde{\rho}$ is then

$$(\tau^\dagger\tau)^M - \tilde{\rho} = S(\emptyset) - R(\emptyset). \tag{22}$$

To bound the norm of this difference, we need to express $R(\emptyset)$ with the help of the $S(\mathcal{I})$'s; that is, we need the inverse relation of Eq. (21). This inverse relation is given by the Möbius inversion formula, which is used, for example, in the context of the Kirkwood-Salzburg equations, for a cluster expansion for the partition function [26,27]. The statement of the Möbius inversion is the following.

Let $\mathcal{A}$ be a finite set, $\mathcal{P}(\mathcal{A})$ the set of all its subsets, and V a vector space. Given a function $f : \mathcal{P}(\mathcal{A}) \to V$, we define the following transformations:

$$\hat{f}(\mathcal{I}) := \sum_{\mathcal{J}:\mathcal{A}\supseteq\mathcal{J}\supseteq\mathcal{I}} f(\mathcal{J}), \tag{23}$$

$$\check{f}(\mathcal{I}) := \sum_{\mathcal{J}:\mathcal{A}\supseteq\mathcal{J}\supseteq\mathcal{I}} (-1)^{|\mathcal{J}\setminus\mathcal{I}|} f(\mathcal{J}). \tag{24}$$

Lemma 2. Möbius inversion:

$$\hat{\check{f}} = \check{\hat{f}} = f.$$

This lemma just expresses that the second transformation is the inverse of the first one. The proof is presented in Appendix C. We will use the lemma by setting $\mathcal{A}$ to be the set of columns, and $f = R$. Thus, comparing the definitions (21) and (23) we deduce that $\hat{f} = S$. Applying the lemma we obtain the desired relation

$$R(\emptyset) = \sum_{\mathcal{I}} (-1)^{|\mathcal{I}|} S(\mathcal{I}),$$

and thus substituting back to Eq. (22)

$$(\tau^\dagger\tau)^M - \tilde{\rho} = S(\emptyset) - \sum_{\mathcal{I}} (-1)^{|\mathcal{I}|} S(\mathcal{I}); \tag{25}$$

therefore

$$(\tau^\dagger\tau)^M - \tilde{\rho} = -\sum_{\mathcal{I}\neq\emptyset} (-1)^{|\mathcal{I}|} S(\mathcal{I}). \tag{26}$$

The one-norm of the difference can be bounded by the triangle inequality:

$$\|(\tau^\dagger\tau)^M - \tilde{\rho}\|_1 \leqslant \sum_{m=1}^{2K} \binom{2K}{m} \max_{\mathcal{I}:|\mathcal{I}|=m} \|S(\mathcal{I})\|_1. \tag{27}$$

We obtained this form by counting the number of subsets $\mathcal{I}$ of the $2K$ columns that have $|\mathcal{I}| = m$. Now, we need to bound the one-norm of $S(\mathcal{I})$. First of all, as noted before, we can sum up over all indices possessing no restriction. That is, over all $\lambda_{i,j}$ with $i \notin \mathcal{I}$. For example, if $2 \notin \mathcal{I}$ then

$$S(\mathcal{I}) = \sum_{\substack{\text{filling} > L \\ \text{for column } i \in \mathcal{I}}} \begin{array}{|c||c|c|c|c|} \hline & x_1 & x_2 & \dots & x_{2K} \\ \hline\hline 1 & X & e^{-\beta h_2} & & \\ \hline 2 & & e^{-\beta h_2} & X & \\ \hline \vdots & & e^{-\beta h_2} & & \\ \hline M & X & e^{-\beta h_2} & & X \\ \hline \end{array}, \tag{28}$$

where we have already summed up for all $\lambda_{2,j}$. Let μ be such a term in $S(\mathcal{I})$ in which each x_i ($i \in \mathcal{I}$) is appearing exactly $k_i > L$ times. Formally,

$$\mu = \prod_{j=1}^{M} \prod_{i=1}^{2K} y_{i,j},$$

with the notation

$$y_{i,j} = \begin{cases} x_i^{\lambda_{i,j}} & \text{if } i \in \mathcal{I}, \\ e^{-\beta \tilde{h}_i} & \text{if } i \notin \mathcal{I}. \end{cases}$$

Here $\lambda \in \mathcal{M}^b_{M,|\mathcal{I}|}$ is restricted to the columns $i \in \mathcal{I}$, and the above defined k_i are given by the expression

$$k_i = \sum_{j=1}^{M} \lambda_{i,j}.$$

The one-norm of this term is bounded by the following lemma.

Lemma 3. If $M > 72\beta^2 K^2$, then

$$\|\mu\|_1 \leqslant 3\|e^{-\beta H}\|_1 \left(\frac{3\beta}{M}\right)^{k_1+\cdots+k_m}.$$

This bound is the consequence of the fact that the x_i's, whose norm is small, appear exactly $k_1 + k_2 + \cdots + k_m$ times in μ, while the rest of the operators, that is, $e^{-\beta \tilde{h}_i}$, give almost a Trotter approximation of $e^{-\beta H}$. The proof is presented in Appendix D. Note that this bound depends only on the numbers $k_1, k_2, \ldots, k_m$, while μ is determined by the coefficients $\lambda_{i,j}$. Therefore there are more terms μ in which each x_i appears exactly k_i times and thus have the same bound. The number of such terms is given by

$$\binom{M}{k_1}\binom{M}{k_2}\cdots\binom{M}{k_m}, \tag{29}$$

as at each column $i \in \mathcal{I}$ one has to choose k_i rows out of the total number of M rows to place the appearing x_i's. Thus the one-norm of $S(\mathcal{I})$ is bounded by the following sum:

$$\|S(\mathcal{I})\|_1 \leqslant \sum_{k_1>L} \cdots \sum_{k_m>L} 3\|e^{-\beta H}\|_1 \prod_{i=1}^{m} \binom{M}{k_i} \left(\frac{3\beta}{M}\right)^{k_i}, \tag{30}$$

as $S(\mathcal{I})$ is the sum of those terms μ in which each x_i ($i \in \mathcal{I}$) appear $k_i > L$ times in Eq. (28). Therefore

$$\|S(\mathcal{I})\|_1 \leqslant 3\|e^{-\beta H}\|_1 \left[\sum_{k>L} \binom{M}{k} \left(\frac{3\beta}{M}\right)^k\right]^m. \tag{31}$$

The sum in the parentheses can be upper bounded by

$$\sum_{k>L} \binom{M}{k} \left(\frac{3\beta}{M}\right)^k \leqslant e^{3\beta} \left(\frac{3e\beta}{L}\right)^L$$

(see Lemma 5 in Appendix E) and thus

$$\|S(\mathcal{I})\|_1 \leqslant 3\|e^{-\beta H}\|_1 \left[e^{3\beta} \left(\frac{3e\beta}{L}\right)^L\right]^m. \tag{32}$$

Substituting the obtained bound into Eq. (27) the following holds for the error of the compression:

$$\|(\tau^\dagger\tau)^M - \tilde{\rho}\|_1 \leqslant 3\|e^{-\beta H}\|_1 \sum_{m=1}^{2K} \binom{2K}{m} \left[e^{3\beta} \left(\frac{3e\beta}{L}\right)^L\right]^m. \tag{33}$$

Thus, after evaluating the sum, we obtain

$$\|(\tau^\dagger\tau)^M - \tilde{\rho}\|_1 \leqslant 3\|e^{-\beta H}\|_1 \left(\left[1 + e^{3\beta} \left(\frac{3e\beta}{L}\right)^L\right]^{2K} - 1\right). \tag{34}$$

As $(1 + x/K)^K \leqslant e^x \leqslant 1 + 2x$ as long as $x < 1$, this yields the bound

$$\|(\tau^\dagger\tau)^M - \tilde{\rho}\|_1 \leqslant 12\|e^{-\beta H}\|_1 K e^{3\beta} \left(\frac{3e\beta}{L}\right)^L. \tag{35}$$

Therefore, if $\beta \leqslant b \log_2 K$, setting $L = O(\log_2 K/\epsilon)$ implies

$$\|(\tau^\dagger\tau)^M - \tilde{\rho}\|_1 \leqslant \epsilon \|e^{-\beta H}\|_1; \tag{36}$$

thus the error of the compression is bounded by ϵ if $L = O(\log_2 K/\epsilon)$ and $M > 72\beta^2 K^2$.

C. Coding as a PEPO

We show that the resulting operator $\tilde{\rho}$ admits a PEPO form as in Eq. (10):

$$\tilde{\rho} = \sum_{\alpha:\mathcal{E}\to\{1,\ldots,D\}} \bigotimes_{v\in\mathcal{V}} X^v_{\alpha(e^v_1),\ldots,\alpha(e^v_{z(v)})}. \tag{37}$$

First, let us consider the Schmidt decomposition of the operators x_i.

$$x_i = e^{-\beta \tilde{h}_i/2M} - 1 = \sum_{\nu=1}^{s} A^{v,i}_\nu \otimes A^{w,i}_\nu, \tag{38}$$

with s being at most d^2_{spin}, where d_{spin} is the dimension of the Hilbert space describing the individual spins, and the edge corresponding to column i is composed of the two particles v and w. Note that there are two columns associated with a Hamiltonian term h_i, $K + 1 - i$, and $K + i$.

After this decomposition, we can think of $\tilde{\rho}$ as the following sum:

$$\tilde{\rho} = \sum_{\substack{\text{filling per} \\ \text{column} \leqslant L}} \begin{array}{|c||c|c|c|c|} \hline & x_1 & x_2 & \dots & x_{2K} \\ \hline\hline 1 & 3 & 1 & 0 & 0 \\ \hline 2 & 0 & 0 & 3 & 0 \\ \hline \vdots & 0 & 3 & 0 & 4 \\ \hline M & 2 & 0 & s & 0 \\ \hline \end{array}, \tag{39}$$

where the sum runs over all fillings that have at most L cells different from 0 in every column. The table means the following. We begin to read the table from left to right, row-by-row. Whenever we meet a cell in column i containing the number k we write down the operator $A_k^{v,i} \otimes A_k^{w,i}$ as in Eq. (38). Otherwise we write down the identity operator. The value of the table is again the product of these operators.

Every term in the above sum is now a tensor product. The local operator acting on particle v depends only on the columns corresponding to the edges surrounding v. Indeed, operators acting nontrivially on particle v occur only in these columns.

Therefore, the index $\alpha(e)$ at edge e will specify a possible filling of the two columns corresponding to e, and the operator $X^v_{\alpha(e_1^v),\dots,\alpha(e^v_{z(v)})}$ will mean the product of the corresponding Schmidt coefficients.

For a given edge $\alpha(e)$ can take

$$D = \left[\sum_{k\leqslant L} \binom{M}{k} s^k\right]^2 \leqslant L^2 (sM)^{2L} \tag{40}$$

different values, as the positions of the nonzero elements and their values are needed to be specified for the two columns corresponding to edge e.

In Sec. III we have shown that we should set $M > 360\beta^2 K^2/\epsilon$ in order for the Trotter approximation to be ϵ-close to the Gibbs state. In Sec. III B we have seen that one can choose L such that the compressed operator, $\tilde{\rho}$, is ϵ-close to the Trotter expansion. Therefore, by the triangle inequality, for any given ϵ that decreases at most polynomially in the system size, one can approximate the Gibbs state with error ϵ, if the Trotter steps are taken to be poly(K) and the compression, L, to be $O(\log_2 K)$. Thus, our method gives a PEPO approximation with bond dimension $K^{O(\log_2 K)}$. As $2K/z \leqslant N$, this is a PEPO with bond dimension $N^{O(\log_2 N)}$. Writing out explicitly the dependence on the dimension of the lattice we get $N^{O(\log_2 N)+O(\log_2 d)}$. As $d \ll N$ this bound is essentially $D = N^{O(\log_2 N)}$.

In Sec. III B we only have supposed that $\beta \leqslant b \log_2 K$, or equivalently, $\beta \leqslant b \log_2 N$. If H is gapped and the density of states for a fixed energy only grows as poly(N), then by setting $\beta = O(\log_2 N)$, the ground state projector is approximated by the Gibbs state with an error decreasing as poly(N). Therefore, our method also gives an $N^{O(\log_2 N)}$ bond dimensional PEPO approximation of the ground state projector, and thus an $N^{O(\log_2 N)}$ bond dimensional PEPS approximation for the ground state [for any prescribed error ϵ that decreases at most as poly(N)] under the same condition.

IV. poly(N) BOND DIMENSIONAL APPROXIMATION

In this section we show that with the help of the cluster expansion technique [17] we can approximate the thermal state by a PEPO with $N^{O(\beta)}$ bond dimension. For that, we just have to modify theorem 15 in [19] and introduce a more efficient way of encoding the PEPO. That theorem says that for $\beta < \beta^*$ (β^* is a constant) the density operator can be well approximated with the truncated cluster expansion, where only clusters of size at most $O(\log_2 K)$ [equivalently, $O(\log_2 N)$] are included. By a clever choice of the coding of the PEPO, we show that for that temperature one just needs a poly(N) bond dimension, and then, as in [17], we extend the result to lower (but finite) temperatures.

A. Cluster expansion

Before restating theorem 15 in [19] we need to introduce some notation. Let $\mathcal{E}^* = \cup_{k=0}^{\infty} \mathcal{E}^k$, that is, a word w from $\mathcal{E}^*$ denotes a sequence of edges: $w = (w_1 w_2 \cdots w_k)$. Let h_w denote the product of the Hamiltonian terms corresponding to those edges, $h_w = h_{w_1} h_{w_2} \cdots h_{w_k}$, and let supp($w$) be the set of all edges occurring in w.

Every word's support is a set $\mathcal{I} \subseteq \mathcal{E}$. One can break it into connected components: $\mathcal{I} = \cup_i \mathcal{I}_i$ where the $\mathcal{I}_i$'s are connected, and different components do not contain common points. These connected components are also called clusters. Then, let $\mathcal{W}_L \subseteq \mathcal{E}^*$ be the set of all words whose support contains only connected components of size at most L. β^* will denote a constant such that $\gamma e^{(2z-1)\beta^*}(e^{\beta^*} - 1) < 1$ (recall that z is the maximal degree of the graph and γ is the lattice growth constant 8), and

$$\tilde{\rho} = \sum_{w\in\mathcal{W}_L} \frac{(-\beta)^{|w|}}{|w|!} h_w. \tag{41}$$

Theorem 15 in [19] contains the following statement:

Theorem 1. If $\beta \leqslant \beta^*$, then

$$\|e^{-\beta H} - \tilde{\rho}\|_1 \leqslant \|e^{-\beta H}\|_1 \left[\exp\left(K \frac{x^L}{1-x}\right) - 1\right] \tag{42}$$

with $x = \gamma e^{(2z-1)\beta}(e^{\beta} - 1) < 1$.

Similarly to Eqs. (A38)–(A39) in [19] one can show that the operator $\tilde{\rho}$ admits the following form:

$$\tilde{\rho} = \sum_{\substack{\mathcal{I}\in\mathcal{C}_L \\ \mathcal{I}=\uplus\mathcal{I}_i}} \prod_i \check{f}(\mathcal{I}_i), \tag{43}$$

where $\mathcal{C}_L$ means the subsets of edges $\mathcal{I}$ that does not contain a connected component of size bigger than L, and the connected components of $\mathcal{I}$ are $\mathcal{I}_i$'s. The operators $\check{f}(\mathcal{I}_i)$ act locally on $\mathcal{I}_i$ and are defined as

$$\check{f}(\mathcal{I}) = \sum_{\substack{w\in\mathcal{I}^* \\ \text{supp}(w)=\mathcal{I}}} \frac{(-\beta)^{|w|}}{|w|!} h_w. \tag{44}$$

We show in Appendix F that $\check{f}(\mathcal{I})$ is the Möbius transform of $f(\mathcal{I}) = e^{-\beta H(\mathcal{I})}$, with $H(\mathcal{I}) = \sum_{e\in\mathcal{I}} h_e$. This observation makes it easier to show that $\tilde{\rho}$ admits the form (43).

B. Coding

We show in this subsection that the truncated cluster expansion $\tilde{\rho}$ (43) can be written as a PEPO [cf. Eq. (10)]. This operator has a very special form. It is a sum of products of local operators, such that the operator acting on a vertex v only depends on the cluster in which v is contained. Therefore, coding $\tilde{\rho}$ as a PEPO will be carried out in two steps. First, we enumerate all subsets of edges $\mathcal{I} \in \mathcal{C}_L$ with the help of an index $\alpha_1 : \mathcal{E} \to \{1,2,\ldots,B_1\}$. This indexing will be such that for any given vertex $v \in \mathcal{V}$ the surrounding edges encode the information in which cluster v is located. Once the cluster $\mathcal{I}_i \ni v$ is identified, the operator $\check{f}(\mathcal{I}_i)$ is written as a PEPO with the help of an index $\alpha_2 : \mathcal{E} \to \{1,2,\ldots,B_2\}$. The index α used at the description of the PEPO is then the composition of α_1 and α_2 taking $B_1 B_2$ different values.

a. Identifying the clusters. Let the different values of $\alpha_1(e)$ enumerate all clusters containing e and of size at most L. For a given cluster size l, there are at most γ^l clusters containing e [see Eq. (8)]; therefore there are at most $L\gamma^L$ such clusters. As $L = O(\log_2 K)$, this means that α_1 takes at most $B_1 \leqslant \mathrm{poly}(K)$ different values. Let us now examine how this indexing is related to the original goal: to enumerate all subsets of edges $\mathcal{I} \in \mathcal{C}_L$. For any given subset $\mathcal{I} \in \mathcal{C}_L$ one can find the corresponding values $[\alpha_1(e)]_{e\in\mathcal{E}}$. However, given an indexing, α_1, it might not correspond to such a subset of edges. The reason is the following. Given an indexing $[\alpha_1(e)]_{e\in\mathcal{E}}$, each index means a cluster $\mathcal{I}_e$. The subset $\mathcal{I} \in \mathcal{C}_L$ corresponding to this α_1 is $\cup_e \mathcal{I}_e$, if for any two edges e and f either $\mathcal{I}_e = \mathcal{I}_f$, or the two clusters $\mathcal{I}_e$ and $\mathcal{I}_f$ do not have common point. Therefore the indexing does *not* correspond to a subset $\mathcal{I} \in \mathcal{C}_L$ if and only if there are two edges e and f such that $\alpha_1(e)$ and $\alpha_1(f)$ denote two different but overlapping clusters. Let us join e and f with a path of edges going in the union of the two clusters $\mathcal{I}_e$ and $\mathcal{I}_f$. Along that path there is a contradiction locally; otherwise, e and f cannot specify contradictory information (see Fig. 1). Therefore, if an indexing α_1 does not correspond to a subset of edges, then there is a point $v \in \mathcal{V}$ where it can be detected.

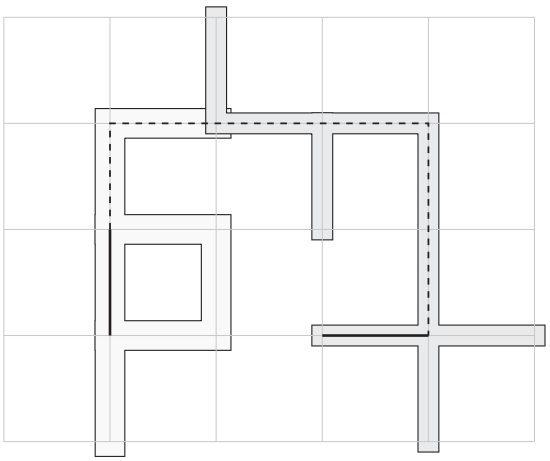

FIG. 1. Two clusters specified by the thick edges. The information contained in those edges becomes contradictory as the clusters overlap. However, the contradiction appears locally somewhere along the dashed line. Thus, our coding will give the 0 operator for this configuration.

b. Coding the local operators. Any operator defined on at most L particles can be written as a PEPO with bond dimension d_{spin}^{2L}, where d_{spin} is the dimension of the Hilbert space of the particles. For example, an expansion in a product basis of the operators supported on L particles can be viewed as a PEPO. As $\check{f}(\mathcal{I}_i)$ is such a local operator with $L = O(\log_2 K)$, this coding requires an index α_2 with $B_2 = \mathrm{poly}(K)$ different values. The local operators used for this construction will be $Y^v_{\alpha_2(e_1^v),\ldots,\alpha_2(e_z^v(v))}(\mathcal{I}_i)$.

With the help of the index $\alpha = (\alpha_1,\alpha_2)$ the operators $X^v_{\alpha(e_1^v),\ldots,\alpha(e_{z(v)}^v)}$ are constructed as follows. If $\alpha_1(e_1^v),\alpha_1(e_2^v),\ldots,\alpha_1(e_{z(v)}^v)$ both specify the same cluster $\mathcal{I}_i$ (or some of them the empty cluster, if compatible with $\mathcal{I}_i$), then let

$$X^v_{\alpha(e_1^v),\ldots,\alpha(e_{z(v)}^v)} = Y^v_{\alpha_2(e_1^v),\ldots,\alpha_2(e_z^v(v))}(\mathcal{I}_i); \tag{45}$$

otherwise, if both of them specify the empty cluster, let $X^v_{\alpha(e_1^v),\ldots,\alpha(e_{z(v)}^v)}$ be $\mathbb{1}$, otherwise let $X^v_{\alpha(e_1^v),\ldots,\alpha(e_{z(v)}^v)}$ be 0. By construction, the contraction of these tensors really gives $\tilde{\rho}$.

As the index used at the coding, $\alpha = (\alpha_1,\alpha_2)$, can only take $B_1 B_2 = \mathrm{poly}(K)$ different values, the above coding is a PEPO with $\mathrm{poly}(K)$ [equivalently $\mathrm{poly}(N)$] bond dimension. Thus, for any $\beta < \beta^*$, we gave an efficient PEPO description of the Gibbs state. Moreover, Theorem 1 holds for $\beta' = \beta/2M$ instead of β if the trace norm is replaced by $\|\cdot\|_{2M}$ without any essential modification. Therefore, by taking M such that $\beta' < \beta^*$, that is, $M = O(\beta)$, this result can be extended to lower (but finite) temperatures as well (see Proposition 1). However, after this step, the approximating operator will be a PEPO exponentiated M times. Therefore the bond dimension required for the PEPO description of the Gibbs state at arbitrary temperature is $N^{O(\beta)}$. Writing out the dependence on the physical dimension one gets $D = N^{O(\beta d \log_2 d)}$ as $\beta^* \propto 1/d$.

V. SUMMARY AND OUTLOOK

We have analyzed the ability of tensor networks to describe thermal (Gibbs) equilibrium states of lattice Hamiltonians with local interactions. First, using a Trotter expansion and a compression method, we have shown that it is possible to approximate that state with a PEPO whose bond dimension scales as $N^{O(\log_2 N)}$, where N is the system size (number of vertices in the lattice). This result is valid for any finite temperature and spatial dimension. It also holds true at zero temperature as long as the Hamiltonian is gapped and the density of states for any energy interval only grows polynomially with the system size. Second, building on Hastings' construction [17], we have shown that it is possible to find a PEPO with a $\mathrm{poly}(N)$ bond dimension at any finite temperature and spatial dimension. Even though the second construction gives better results than the first one at finite temperature, the first construction has the advantage that it is more natural. Indeed, it generalizes the commuting case, in which case one can easily construct an exact PEPO expression for the Gibbs state.

There are some straightforward implications of the results derived here. First, even though we have concentrated on PEPOs, it is trivial to express our results in terms of (pure) PEPS. At finite temperature, we can just consider the PEPO corresponding to half the temperature, and apply it to locally maximally entangled states in order to obtain a purification in

terms of a PEPS with a polynomially growing bond dimension [21]. At zero temperature, we can simply apply the constructed PEPO to a random product state in order to show that there exists a PEPS with $D = N^{O(\log_2 N)}$. Second, for translationally invariant problems in regular lattices, our construction may break translational invariance (as we select some order of the bonds). But it is always possible [4] to make a PEPO (or PEPS) translationally invariant with an increase of the bond dimension by just a factor of N. Third, even though we have considered Hamiltonians interacting along the edges in the graph, our construction can be easily extended to the case in which the local Hamiltonians act on plaquettes. The idea is that at the Trotter decomposition we have made no assumption on the support of the individual Hamiltonian terms, whereas at the coding procedure, we still need to keep information contained in a constant number of columns: in an edge $e = (v, w)$, we can keep the information contained in the columns corresponding to Hamiltonian terms that act nontrivially on either v or w. In such a coding the same piece of information is specified in more than one edge, but their consistency can be checked locally, at the vertices. The cluster expansion technique can be applied with no essential modification as the number of terms acting on the boundary of a cluster can still be upper bounded by a constant times the size of the cluster, and the number of clusters containing l terms is still bounded by γ^l, where γ is a lattice growth constant [20]. Finally, our construction can also be straightforwardly extended to fermions with the result that we just have to use fermionic PEPS [28].

ACKNOWLEDGMENTS

We thank M. C. Bañuls, T. Vidick, Z. Landau, and U. Vazirani for discussions. This work has been partially supported by the EU project SIQS, and the DFG project NIM. We also thank the Benasque Center for Science, the Perimeter Institute (Waterloo), and the Simons Center for the Theory of Computing (Berkeley), where part of the work was carried out, for their hospitality. N.S. acknowledges support from the Alexander von Humboldt foundation and the EU project QALGO. J.I.C. acknowledges support from Adolph C. and Mary Sprague Miller Institute for Basic Research in Science, University of California Berkeley.

APPENDIX A: PROOF OF PROPOSITION 1

Here we present the proof of Proposition 1. The proof consists of two steps. First, by the positivity of η, we show that if $\varepsilon < 1/3$ and

$$\|\eta - \tau\|_{2M} \leqslant \frac{\varepsilon}{M}\|\eta\|_{2M}, \tag{A1}$$

then

$$\|\eta^2 - \tau^\dagger\tau\|_M \leqslant 3\varepsilon\|\eta^2\|_M. \tag{A2}$$

Using the identity $a^2 - b^2 = a(a-b) + (a-b)b$ and the triangle inequality we obtain

$$\|\eta^2 - \tau^\dagger\tau\|_M = \|\eta(\eta-\tau)\|_M + \|(\eta-\tau^\dagger)\tau\|_M; \tag{A3}$$

thus using the Hölder inequality and that $\|X\|_{2M} = \|X^\dagger\|_{2M}$, we conclude that

$$\|\eta^2 - \tau\tau^\dagger\|_M \leqslant (\|\eta\|_{2M} + \|\tau\|_{2M})\|\eta - \tau\|_{2M}. \tag{A4}$$

$\|\eta - \tau\|_{2M}$ is bounded by the assumptions of the statement, and so is $\|\tau\|_{2M}$ by the triangle inequality. Therefore

$$\|\eta^2 - \tau\tau^\dagger\|_M \leqslant \left(2 + \frac{\varepsilon}{M}\right)\frac{\varepsilon}{M}\|\eta\|_{2M}^2. \tag{A5}$$

η is positive; thus $\|\eta\|_{2M}^2 = \|\eta^2\|_M$. If $\varepsilon < 1$, then

$$\|\eta^2 - \tau\tau^\dagger\|_M \leqslant 3\frac{\varepsilon}{M}\|\eta^2\|_M. \tag{A6}$$

This completes the proof of the first step.

Second, we prove that if Eq. (A6) holds, then

$$\|\eta^{2M} - (\tau^\dagger\tau)^M\|_1 \leqslant 9\varepsilon\|\eta^{2M}\|_1.$$

The proof is basically the same as that of the first step. Using the identity $a^m - b^m = \sum_i a^i(a-b)b^{m-i-1}$ and the triangle inequality we obtain

$$\|\eta^{2M} - (\tau^\dagger\tau)^M\|_1 \leqslant \sum_{i=0}^{M-1}\|\eta^{2i}(\eta^2 - \tau^\dagger\tau)(\tau^\dagger\tau)^{M-i-1}\|_1. \tag{A7}$$

Hence by Hölder's inequality the difference is upper bounded by

$$\sum_{i=0}^{M-1}\|\eta^{2i}\|_{\frac{M}{i}}\|\eta^2 - (\tau^\dagger\tau)\|_M\|(\tau^\dagger\tau)^{M-i-1}\|_{\frac{M}{(M-i-1)}}. \tag{A8}$$

For X positive semidefinite, and any real number r, $X^r \geqslant 0$, and thus by the definition of the Schatten norms

$$\|X^r\|_{M/r} = \|X\|_M^r. \tag{A9}$$

Applying (A9) to η and $\tau^\dagger\tau$ in (A8), the inequality takes the following form:

$$\|\eta^{2M} - (\tau^\dagger\tau)^M\|_1 \leqslant \sum_{i=0}^{M-1}\|\eta^2\|_M^i\|\eta^2 - \tau^\dagger\tau\|_M\|\tau^\dagger\tau\|_M^{M-i-1}. \tag{A10}$$

$\|\eta^2 - \tau^\dagger\tau\|_M$ is bounded by Eq. (A6). Hence, by the triangle inequality, $\|\tau^\dagger\tau\|_M$ is bounded as well,

$$\|\tau^\dagger\tau\|_M \leqslant \|\eta^2\|_M + \|\eta^2 - \tau^\dagger\tau\|_M \leqslant \left(1 + \frac{3\varepsilon}{M}\right)\|\eta^2\|_M.$$

As $1 + 3\varepsilon/M > 1$, we can upper bound the sum by taking $(1 + 3\varepsilon/M)^M$ as common factor in every term,

$$\|\eta^{2M} - (\tau^\dagger\tau)^M\|_1 \leqslant M\frac{\varepsilon}{M}\left(1 + \frac{3\varepsilon}{M}\right)^M\|\eta^2\|_M^M. \tag{A11}$$

Since $(1 + 3\varepsilon/M)^M < e^{3\varepsilon} < 3$, if $\epsilon \leqslant 1/3$, the statement of the proposition follows:

$$\|\rho^{2M} - (\tau^\dagger\tau)^M\|_1 \leqslant 9\varepsilon\|\eta^2\|_M^M. \tag{A12}$$

APPENDIX B: PROOF OF LEMMA 1

Here we present the proof of Lemma 1, and derive how big the number of Trotter steps should be chosen for a good approximation of the Gibbs state. The proof relies on the fact that if M is big enough, then both η and τ are close to $\mathbb{1} - \beta H/2M$.

By the use of Hölder's inequality we obtain

$$\|\eta-\tau\|_{2M}=\|\eta^{-1}\eta(\eta-\tau)\|_{2M}\leqslant\|\eta\|_{2M}\|\eta^{-1}\|\|\eta-\tau\|.$$

The norm of η^{-1} can be upper bounded by a constant if $M>\beta K/2$:

$$\|\eta^{-1}\|\leqslant e^{\beta K/2M}\leqslant 3. \tag{B1}$$

The norm of $\eta-\tau$ will be bounded with the help of the triangle inequality, by adding and subtracting $\mathbb{1}-\beta H/2M$:

$$\|\eta-\tau\|\leqslant\left\|\eta-\mathbb{1}+\frac{\beta H}{2M}\right\|+\left\|\tau-\mathbb{1}+\frac{\beta H}{2M}\right\|. \tag{B2}$$

We will use the following bound on the Taylor expansion of the exponential function to upper bound these expressions.

Lemma 4. The following two bounds hold:

$$\|e^A-\mathrm{Id}\|\leqslant\|A\|e^{\|A\|},$$

$$\|e^A-\mathrm{Id}-A\|\leqslant\frac{\|A\|^2}{2}e^{\|A\|}.$$

Proof. $e^A=\sum_n\frac{A^n}{n!}$; thus

$$e^A-\sum_{n=0}^{k}\frac{A^n}{n!}=\sum_{n=k+1}^{\infty}\frac{A^n}{n!}. \tag{B3}$$

Therefore the norm of the difference can be upper bounded by the triangle inequality

$$\left\|e^A-\sum_{n=0}^{k}\frac{A^n}{n!}\right\|\leqslant\sum_{n=k+1}^{\infty}\frac{\|A\|^n}{n!}\leqslant\frac{\|A\|^{k+1}}{(k+1)!}\sum_{n=0}^{\infty}\frac{\|A\|^n}{n!}, \tag{B4}$$

since $(n+k+1)!\geqslant n!(k+1)!$. Summing up we have the desired inequality

$$\left\|e^A-\sum_{n=0}^{k}\frac{A^n}{n!}\right\|\leqslant\frac{\|A\|^{k+1}}{k+1!}e^{\|A\|}. \tag{B5}$$

The statements correspond to the particular cases $k=0,1$. ■

Due to the previous lemma, we can bound the first part of the right-hand side of Eq. (B2):

$$\left\|e^{-\frac{\beta H}{2M}}-1+\frac{\beta H}{2M}\right\|\leqslant\frac{\beta^2K^2}{8M^2}e^{\beta K/2M}\leqslant\frac{\varepsilon}{2M}. \tag{B6}$$

If $M\geqslant\beta^2K^2/\epsilon$ and $M\geqslant\frac{\beta K}{2}$, because then $e^{\beta K/2M}\leqslant 3$ and $3/8\leqslant 1/2$.

The second part of the right-hand side of Eq. (B2) can be written as

$$\tau-\mathbb{1}+\frac{\beta H}{2M}=\prod_i[\mathbb{1}+x_i]-\mathbb{1}+\frac{\beta H}{2M}, \tag{B7}$$

where x_i is as in Eq. (14). Let us expand the product. The zeroth term cancels out, whereas the first-order term is

$$\sum_i x_i+\frac{\beta H}{2M}.$$

The norm of the sum of the kth-order terms can be upper bounded by

$$\binom{K}{k}\left(\frac{\beta}{2M}e^{\frac{\beta}{2M}}\right)^k\leqslant\left(\frac{3\beta K}{2M}\right)^k$$

if $M\geqslant\beta/2$, because there are $\binom{K}{k}$ kth-order terms, and the norm of $\|x_i\|$ can be bounded by Lemma 4:

$$\|x_i\|\leqslant\frac{\beta}{2M}e^{\frac{\beta}{2M}}.$$

Therefore, after expanding the product in Eq. (B7), we obtain that

$$\left\|\tau-1+\frac{\beta H}{2M}\right\|\leqslant\left\|\sum_i x_i+\frac{\beta h_i}{2M}\right\|+\sum_{k=2}^{\infty}\left(\frac{3K\beta}{2M}\right)^k. \tag{B8}$$

The first term can be again bounded by Lemma 4 as $x_i+\beta h_i/2M=e^{-\beta h_i/2M}-\mathbb{1}+\beta h_i/2M$:

$$\left\|\sum_i x_i-\frac{\beta h_i}{2M}\right\|\leqslant K\frac{\beta^2}{4M^2}e^{\frac{\beta}{2M}}\leqslant\frac{K\beta^2}{M^2}, \tag{B9}$$

since if $M>\beta/2$, then $e^{\frac{\beta}{2M}}<4$. The second term can be upper bounded by

$$\sum_{k=2}^{\infty}\left(\frac{3K\beta}{2M}\right)^k=\left(\frac{3K\beta}{2M}\right)^2\frac{1}{1-\frac{3K\beta}{2M}}\leqslant\frac{5K^2\beta^2}{M^2}, \tag{B10}$$

since if $M>3K\beta$, then $\frac{1}{1-\frac{3K\beta}{2M}}\leqslant 2$, and $9/2\leqslant 5$. Finally, $K>1$ and thus the sum of the bounds obtained in Eqs. (B9) and (B10) can be upper bounded by

$$\left\|\tau-1+\frac{\beta H}{2M}\right\|\leqslant\frac{6K^2\beta^2}{M^2}\leqslant\frac{\varepsilon}{2M}, \tag{B11}$$

if $M>12K^2\beta^2\frac{1}{\varepsilon}$.

Putting together the two bounds in Eqs. (B6) and (B11), we obtain that

$$\|\eta-\tau\|\leqslant\frac{\epsilon}{M} \tag{B12}$$

if $M>12K^2\beta^2/\varepsilon$. Therefore, the statement follows: if $M>36K^2\beta^2/\varepsilon$, then

$$\|\eta-\tau\|_{2M}\leqslant\frac{\epsilon}{M}\|\eta\|_{2M}. \tag{B13}$$

APPENDIX C: PROOF OF THE MÖBIUS INVERSION

Here we prove Lemma 2. The first part of the statement is that $\check{\hat{f}}=f$. Let us define $f'(I)$ as

$$f'(\mathcal{I})=\sum_{\mathcal{J}\supseteq\mathcal{I}}(-1)^{|\mathcal{J}\setminus\mathcal{I}|}\hat{f}(\mathcal{J}). \tag{C1}$$

Then, the statement is that $f'(\mathcal{I})=f(\mathcal{I})$. Let us express $\hat{f}$ with the help of f as in Eq. (23):

$$f'(\mathcal{I})=\sum_{\mathcal{J}\supseteq\mathcal{I}}(-1)^{|\mathcal{J}\setminus\mathcal{I}|}\sum_{\mathcal{K}\supseteq\mathcal{J}}f(\mathcal{K}). \tag{C2}$$

By changing the order of the sums we obtain

$$f'(\mathcal{I})=\sum_{\mathcal{K}\supseteq\mathcal{I}}f(\mathcal{K})\sum_{\mathcal{J}:\mathcal{K}\supseteq\mathcal{J}\supseteq\mathcal{I}}(-1)^{|\mathcal{J}\setminus\mathcal{I}|}. \tag{C3}$$

We evaluate now the second sum. Suppose first $\mathcal{K} \neq \mathcal{I}$; then

$$\sum_{\mathcal{J}:\mathcal{K}\supseteq\mathcal{J}\supseteq\mathcal{I}} (-1)^{|\mathcal{J}\setminus\mathcal{I}|} = \sum_{\mathcal{J}'\subseteq\mathcal{K}\setminus\mathcal{I}} (-1)^{|\mathcal{J}'|} = (1-1)^{|\mathcal{K}\setminus\mathcal{I}|} = 0.$$

Otherwise, if $\mathcal{K} = \mathcal{I}$, then the sum is 1. Substituting this back in the expression of $f'(\mathcal{I})$, we get

$$f'(\mathcal{I}) = \sum_{\mathcal{K}\supseteq\mathcal{I}} f(\mathcal{K}) \sum_{\mathcal{J}:\mathcal{K}\supseteq\mathcal{J}\supseteq\mathcal{I}} (-1)^{|\mathcal{J}\setminus\mathcal{I}|} = f(\mathcal{I}). \tag{C4}$$

This proves the first part of the statement. The second part, $\hat{\check{f}} = f$, works similarly. Let us define now f'' as follows:

$$f''(\mathcal{I}) = \sum_{\mathcal{J}\subseteq\mathcal{I}} \check{f}(\mathcal{J}). \tag{C5}$$

Thus, we have to prove that $f'' = f$. Substituting back the expression for $\check{f}$ [as in Eq. (24)] in this equation, we obtain

$$f''(\mathcal{I}) = \sum_{\mathcal{J}\subseteq\mathcal{I}} \sum_{\mathcal{K}\subseteq\mathcal{J}} (-1)^{|\mathcal{J}\setminus\mathcal{K}|} f(\mathcal{K}). \tag{C6}$$

By changing the order of the two sums we obtain

$$f''(\mathcal{I}) = \sum_{\mathcal{K}\subseteq\mathcal{I}} f(\mathcal{K}) \sum_{\mathcal{J}:\mathcal{K}\subseteq\mathcal{J}\subseteq\mathcal{I}} (-1)^{|\mathcal{J}\setminus\mathcal{K}|}. \tag{C7}$$

The second sum is again $\delta_{\mathcal{K},\mathcal{I}}$, and thus

$$f''(\mathcal{I}) = f(\mathcal{I}). \tag{C8}$$

APPENDIX D: PROOF OF LEMMA 3

Here we present the proof of Lemma 3. In μ two types of terms occur. First, if i refers to a column that has been summed up ($i \notin \mathcal{I}$), then in every row of that column the term $e^{-\beta\tilde{h}_i}$ appears. Second, if $i \in \mathcal{I}$, then the sum on that column has not been evaluated; therefore the corresponding term in row j is $x_i^{\lambda_{i,j}}$. We now separate these terms:

$$\mu = \prod_{j=1}^{M} \prod_{i\in\mathcal{I}} y_i^{\lambda_{i,j}} \prod_{i\in\{1..2K\}} e^{-\beta\tilde{h}_i/2M}, \tag{D1}$$

where we have introduced

$$y_i = \prod_{j<i} e^{-\beta\tilde{h}_j/2M} x_i \prod_{j<i} e^{\beta\tilde{h}_j/2M} e^{\beta\tilde{h}_i/2M}. \tag{D2}$$

The norm of y_i can be bounded by the norm of x_i as follows:

$$\|y_i\| \leqslant \|x_i\| e^{\beta K/M} \tag{D3}$$

because $\|e^{-\beta\tilde{h}_j/2M}\| \leqslant 1$ and $\|e^{\beta\tilde{h}_i/2M}\| \leqslant e^{\beta/2M}$ and there are at most $2K$ such terms in y_i. Thus, by applying Lemma 4 to x_i, we obtain

$$\|y_i\| \leqslant \|x_i\| e^{\beta K/M} \leqslant \frac{\beta}{2M} e^{\beta/2M} e^{\beta K/M} \leqslant \frac{3\beta}{2M}, \tag{D4}$$

since $e^{\beta(2K+1)/2M} \leqslant 3$ if $M > 2\beta K > \beta(K+1/2)$. We now apply Hölder's inequality to Eq. (D1) in order to bound $\|\mu\|_1$:

$$\|\mu\|_1 \leqslant \prod_i \|y_i\|^{\sum_j \lambda_{i,j}} \left\| \prod_{i\in\{1..2K\}} e^{-\beta\tilde{h}_i/2M} \right\|_M^M.$$

The last expression on the right-hand side is $(\tau^\dagger\tau)$ from the Trotter expansion formula. By the use of another Hölder's inequality

$$\|\tau^\dagger\tau\|_M \leqslant \|\tau\|_{2M}^2.$$

Using the triangle inequality and Lemma 1 with the choice $\epsilon = 1/2$, we obtain that

$$\|\tau\|_{2M} \leqslant \left(1 + \frac{1}{2M}\right) \|e^{-\beta H/2M}\|_{2M} \tag{D5}$$

if $M > 72\beta^2K^2$; therefore

$$\|\mu\|_1 \leqslant \prod_i \|y_i\|^{k_i} \left(1 + \frac{1}{2M}\right)^{2M} \|e^{-\beta H/2M}\|_{2M}^{2M}. \tag{D6}$$

Using the bound (D4) on $\|y_i\|$, and the fact that $(1 + 1/2M)^{2M} < e < 3$, we obtain the statement of the lemma,

$$\|\mu\|_1 \leqslant 3\left(\frac{3\beta}{M}\right)^{k_1+\cdots+k_n} \|e^{-\beta H}\|_1. \tag{D7}$$

APPENDIX E: LEMMA ON THE SUM OF BINOMIAL COEFFICIENTS

We need the following lemma to upper bound a sum of binomial coefficients in Eq. (31):

Lemma 5. $\sum_{k>L} \binom{M}{k} x^k \leqslant e^{Mx} (\frac{eMx}{L})^L$.

Proof. First, as $\binom{M}{k} \leqslant M^k/k!$, we have

$$\sum_{k>L} \binom{M}{k} x^k \leqslant \sum_{k\geqslant L} \frac{1}{k!} (Mx)^k. \tag{E1}$$

We then use $(L+n)! \geqslant n!L!$ and sum up over $n = k - L$.

$$\sum_{k\geqslant L} \binom{M}{k} x^k \leqslant \frac{1}{L!} (Mx)^L e^{Mx}. \tag{E2}$$

Finally, by Stirling's formula, we have the desired result:

$$\sum_{k>L} \binom{M}{k} x^k \leqslant e^{Mx} \left(\frac{eMx}{L}\right)^L. \tag{E3}$$

■

APPENDIX F: ON THE CLUSTER EXPANSION

In this appendix we show how to use the Möbius inversion to reproduce the cluster expansion. In particular, we show that

$$g(\mathcal{I}) = \sum_{\substack{w\in\mathcal{I}^* \\ \mathrm{supp}(w)=\mathcal{I}}} \frac{(-\beta')^{|w|}}{|w|!} h_w \tag{F1}$$

is the (inverse) Möbius transform [29] of

$$f(\mathcal{J}) = e^{-\beta' H(\mathcal{J})}. \tag{F2}$$

Let us consider the Möbius transform of g:

$$\hat{g}(\mathcal{I}) = \sum_{\mathcal{J}\subseteq\mathcal{I}} g(\mathcal{J}) = \sum_{\mathcal{J}\subseteq\mathcal{I}} \sum_{\substack{w\in\mathcal{J}^* \\ \mathrm{supp}(w)=\mathcal{J}}} \frac{(-\beta')^{|w|}}{|w|!} h_w. \tag{F3}$$

This means that in $\hat{g}$ we have to sum up for all words in $\mathcal{I}^*$. Indeed, in the sum every word is counted exactly once as we sum up all possible supports. This implies that

$$\hat{g}(\mathcal{I}) = e^{-\beta' H(\mathcal{I})} = f(\mathcal{I}), \tag{F4}$$

and therefore by the Möbius inversion formula $g = \check{f}$.

We now show that obtaining the form Eq. (43) of $\tilde{\rho}$ is much easier with these tools. The proof follows from the multiplicativity of $\check{f}$: if $\mathcal{I}$ and $\mathcal{J}$ are nonoverlapping clusters, then $\check{f}(\mathcal{I} \cup \mathcal{J}) = \check{f}(\mathcal{I})\check{f}(\mathcal{J})$. Indeed,

$$\check{f}(\mathcal{I} \cup \mathcal{J}) = \sum_{\mathcal{K} \subseteq \mathcal{I} \cup \mathcal{J}} (-1)^{|\mathcal{I} \cup \mathcal{J} \setminus \mathcal{K}|} e^{-\beta H(\mathcal{K})}, \tag{F5}$$

where we have used the multiplicativity of the exponential. $\mathcal{K}$ can be broken into two parts: $\mathcal{K}_\mathcal{I} = \mathcal{K} \cap \mathcal{I}$ and $\mathcal{K}_\mathcal{J} = \mathcal{K} \cap \mathcal{J}$. Then both the -1 factor and the exponential factorizes as follows:

$$\check{f}(\mathcal{I} \cup \mathcal{J}) = \sum_{\substack{\mathcal{K}_\mathcal{I} \subseteq \mathcal{I} \\ \mathcal{K}_\mathcal{J} \subseteq \mathcal{J}}} (-1)^{|\mathcal{I} \setminus \mathcal{K}_\mathcal{I}|} (-1)^{|\mathcal{J} \setminus \mathcal{K}_\mathcal{J}|} e^{-\beta H(\mathcal{K}_\mathcal{I})} e^{-\beta H(\mathcal{K}_\mathcal{J})} \tag{F6}$$

and this sum is nothing but $\check{f}(\mathcal{I})\check{f}(\mathcal{J})$. This implies that the Gibbs state admits the following form:

$$\tilde{\rho} = \sum_{\mathcal{I} \subseteq \mathcal{E}} \prod_{\substack{i : \mathcal{I}_i \text{ are the} \\ \text{clusters in } \mathcal{I}}} \check{f}(\mathcal{I}_i), \tag{F7}$$

and thus the approximation $\tilde{\rho}$ is nothing but

$$\tilde{\rho} = \sum_{\substack{\mathcal{I} \in \mathcal{C}_L \\ \mathcal{I} = \uplus \mathcal{I}_i}} \prod_i \check{f}(\mathcal{I}_i) \tag{F8}$$

as in Eq. (43).

[1] J. I. Cirac and P. Zoller, Goals and opportunities in quantum simulation, Nat. Phys. **8**, 264 (2012).

[2] J. I. Cirac and F. Verstraete, Renormalization and tensor product states in spin chains and lattices, J. Phys. A: Math. Theor. **42**, 504004 (2009).

[3] M. Fannes, B. Nachtergaele, and R. F. Werner, Finitely correlated states on quantum spin chains, Commun. Math. Phys. **144**, 443 (1992).

[4] D. Pérez-García, F. Verstraete, M. M. Wolf, and J. I. Cirac, Matrix product state representations, Quantum Inf. Comput. **7**, 401 (2007).

[5] M. B. Hastings, An area law for one dimensional quantum systems, J. Stat. Mech: Theory Exp. (2007) P08024.

[6] Z. Landau, U. Vazirani, and T. Vidick, A polynomial-time algorithm for the ground state of 1D gapped local Hamiltonians, arXiv:1307.5143.

[7] F. Verstraete and J. I. Cirac, Matrix product states represent ground states faithfully, Phys. Rev. B **73**, 094423 (2006).

[8] M. Srednicki, Entropy and area, Phys. Rev. Lett. **71**, 666 (1993).

[9] J. Eisert, M. Cramer, and M. B. Plenio, Area laws for the entanglement entropy—A review, Rev. Mod. Phys. **82**, 277 (2010).

[10] L. Masanes, Area law for the entropy of low-energy states, Phys. Rev. A **80**, 052104 (2009).

[11] E. Hamza, S. Michalakis, B. Nachtergaele, and R. Sims, Approximating the ground state of gapped quantum spin systems, J. Math. Phys. **50**, 095213 (2009).

[12] M. M. Wolf, Violation of the entropic area law for fermions, Phys. Rev. Lett. **96**, 010404 (2006).

[13] D. Gioev and I. Klich, Entanglement entropy of fermions in any dimension and the Widom conjecture, Phys. Rev. Lett. **96**, 100503 (2006).

[14] F. Verstraete and J. I. Cirac, Renormalization algorithms for Quantum-Many Body Systems in two and higher dimensions, arXiv:cond-mat/0407066.

[15] F. Verstraete and J. I. Cirac, Valence-bond states for quantum computation, Phys. Rev. A **70**, 060302 (2004).

[16] 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).

[17] M. B. Hastings, Solving gapped hamiltonians locally, Phys. Rev. B **73**, 085115 (2006).

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

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

[20] D. Klarner, Cell growth problems, Can. J. Math. **19**, 851 (1967).

[21] 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).

[22] 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).

[23] M. B. Hastings, Locality in quantum systems, arXiv:1008.5137.

[24] D. W. Berry, G. Ahokas, R. Cleve, and B. C. Sanders, Efficient quantum algorithms for simulating sparse hamiltonians, Commun. Math. Phys. **270**, 359 (2006).

[25] R. Bhatia, *Matrix Analysis* (Springer, New York, 1997).

[26] R. Kotecký and D. Preiss, Cluster expansion for abstract polymer models, Commun. Math. Phys. **103**, 491 (1986).

[27] R. B. Griffiths, Rigorous results and theorems, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic Press, New York, 1980), pp. 7–109.

[28] C. V. Kraus, N. Schuch, F. Verstraete, and J. I. Cirac, Fermionic projected entangled pair states, Phys. Rev. A **81**, 052338 (2010).

[29] Note that the definition of Möbius inversion is slightly different in this context: we use a sum over $\mathcal{J} \subseteq \mathcal{I}$ instead of a sum over $\mathcal{J} \supseteq \mathcal{I}$. The inverse is defined likewise.