Thermal states as convex combinations of matrix product states

Mario Berta,[1] Fernando G. S. L. Brandão,[2] Jutho Haegeman,[3] Volkher B. Scholz,[4,3] and Frank Verstraete[3,5]
[1]*Department of Computing, Imperial College London, London SW7 2AZ, United Kingdom*
[2]*Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, California 91125, USA*
[3]*Department of Physics and Astronomy, Ghent University, 9000 Ghent, Belgium*
[4]*Institute for Theoretical Physics, ETH Zurich, 8093 Zürich, Switzerland*
[5]*Department of Physics, University of Vienna, 1090 Vienna, Austria*

(Received 12 January 2018; published 26 December 2018)

We study thermal states of strongly interacting quantum spin chains and prove that those can be represented in terms of convex combinations of matrix product states. Apart from revealing new features of the entanglement structure of Gibbs states, our results provide a theoretical justification for the use of White's algorithm of minimally entangled typical thermal states. Furthermore, we shed new light on time dependent matrix product state algorithms which yield hydrodynamical descriptions of the underlying dynamics.

DOI: 10.1103/PhysRevB.98.235154

I. INTRODUCTION

The theory of entanglement and tensor networks have provided a novel language for describing strongly correlated quantum many body systems. This has led to a deeper understanding of the properties of topological phases of matter [1–4], to novel computational algorithms for simulating such spin systems [5,6] and to rigorous proofs that ground states of gapped one-dimensional quantum spin systems can be represented [7,8] and simulated [9] using matrix product states, as those ground states satisfy the area law for the entanglement theory.

In this paper, we are concerned with representing thermal states of quantum spin systems. It has been proved that such Gibbs states satisfy the area law for the mutual information [10], that matrix product operators [11–13] provide a faithful approximation [14,15] to Gibbs states, and efficient algorithms for finding those operators have been formulated [11,16,17]. Conceptually, those algorithms suffer from a major drawback: no distinction is made between the "classical" and "quantum" correlations. Classical correlations should be dealt with by using Monte Carlo sampling techniques, and one should not waste a large "bond dimension" to those fluctuations. Furthermore, all algorithms dealing with matrix product operators either deal with purifications, which can potentially lead to a huge increase in bond dimension [18], or cannot assure positivity of the matrix product density operator. Those algorithms also become inefficient for low temperatures because, when working with matrix product operator descriptions of pure states, the bond dimension is squared. Those problems can be cured by invoking mixtures of pure matrix product states, and this is the main topic of this paper.

Our main result states that thermal states of one-dimensional local Hamiltonians can be approximately written as such as convex combinations of matrix product states, all of which have a bounded bond dimension. Unlike the case of ground states [7,8], the Hamiltonians do not need to be gapped: we only require a uniform bound on the interaction strength. The proof relies on recent results by one of the authors concerning the Markov structure of Gibbs states [19]. We illustrate this by providing arguments for using matrix product states as subroutines in algorithms dealing with thermal states of quantum spin systems: first, for the minimally entangled typical thermal states (METTS) algorithm of White [20,21] yielding an approximation of Gibbs states using DMRG techniques; and second, for the quantum thermalization algorithm of Leviatan *et al.* revealing hydrodynamical properties of quantum spin chains [22].

Our paper is organized as follows. We first review the definition of matrix product states and state the main result. After sketching the proof (the details can be found in the Appendices), we study the two mentioned numerical algorithms, starting with the static case and following with the dynamical case. We end by discussing further applications of our results.

II. MATRIX PRODUCT STATES

A prominent example of quantum states representable by a tensor network are matrix product states (MPSs) [7,23,24]. They form a submanifold [25] $M_{\mathrm{MPS}}^{D} \subset (\mathbb{C}^d)^{\otimes n}$ of the state space of a one-dimensional quantum lattice system with n sites and a finite d-dimensional local Hilbert space on every site:

$$\left|\psi\left[A_{i_1},\ldots,A_{i_n}\right]\right\rangle = \sum_{i_1,\ldots,i_n} \mathrm{Tr}\left[A_{i_1}\cdots A_{i_n}\right]|i_1\rangle\ldots|i_n\rangle, \quad (1)$$

where here and henceforth we assume periodic boundary conditions,[1] and the A_{i_j}, $j=1,\ldots,n$ are $D\times D$ dimensional matrices. The parameter D is called the bond dimension, and it models the amount of entanglement in the state. Matrix product states with low bond dimension satisfy the area law of entanglement and have been proved to capture the ground

[1]Our main result (Theorem 1) can be extended to open boundary conditions rather trivially, but we choose to work with a periodic boundary condition to keep the notation simple.

 ©2018 American Physical Society

state physics of one-dimensional local gapped Hamiltonians, exactly due to the rapid decay of the entanglement spectrum in these systems [8]. In addition, MPSs allow for an efficient computation of expectation values, independently of the system size. Various algorithms exists which find the best approximate state within the submanifold M^D_{MPS} for the ground state of a local Hamiltonian, either variationally using the density matrix renormalization group (DMRG) [26] or by simulating imaginary time evolution using time-evolving block decimation (TEBD) [27] or the time dependent variational principle (TDVP) [28]. In this paper, we are interested in mixed states which can be written as convex combinations of matrix product states of a fixed bond dimension,

$$\rho[\mu] = \int d\mu(A_{i_1}, \dots) |\psi[A_{i_1}, \dots]\rangle\langle\psi[A_{i_1}, \dots]|. \quad (2)$$

Here $\mu(A_{i_1}, \dots)$ denotes some probability measure on the manifold M^D_{MPS} and is otherwise arbitrary. We note that if we can efficiently sample from this distribution, then we can also efficiently compute expectation values of local observables, as in the case of matrix product states.

III. MAIN RESULT

Sharing many properties with the class of pure matrix product states which model the ground state physics of gapped local Hamiltonians, we expect that convex combinations of matrix product states also possess physical significance. And indeed we find that they approximate thermal states

$$\rho_{H,T} = \frac{1}{Z} \exp(-H/T) \quad (3)$$

of local Hamiltonians $H = \sum_i h_i$, where $Z = \mathrm{Tr}[\exp(-H/T)]$ is the partition function (T denotes temperature). Here local means that each term h_i only acts on a finite number of neighboring sites. In addition, we assume that the interaction terms all satisfy a unique upper bound on their interaction strength, $\|h_i\|_\infty \leqslant C$, where $\|.\|_\infty$ denotes the operator norm.

Theorem 1. Let H be a local one-dimensional Hamiltonian such that its interaction terms possess a uniform upper bound on their interaction strength. Then, for any temperature T and any $\varepsilon > 0$ there exist a bond dimension D and a probability distribution μ_ε on the manifold M^D_{MPS} of matrix product states with bond dimension D such that the associated convex combination of matrix product states $\rho[\mu]$ is ε-close to the thermal state at temperature T:

$$\|\rho_{H,T} - \rho[\mu_\varepsilon]\|_1 \leqslant \varepsilon. \quad (4)$$

Here, the $\|\cdot\|_1$ denotes the trace-norm. The bond dimension D scales quasi-polynomially in the system size and ε^{-1}, and doubly exponential in the inverse temperature T^{-1}.

Note that such an approximation is trivial for a bond dimension D scaling exponentially in system size, whereas we show quasi-polynomial scaling. We emphasise that a bound on the trace-norm is in particular sufficient to guarantee a bound on the error in observables, but is probably not required. Hence, numerical simulations might observe a faster convergence, especially in local observables. In the following, we present the main ideas which go into the proof of

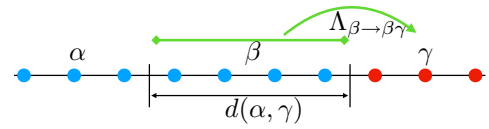


FIG. 1. Markov chain structure of Gibbs states: for every tripartite split of the lattice $\alpha\beta\gamma$, there exists a local quantum channel $\Lambda_{\beta\to\beta\gamma}$ only acting on the region β and approximately recovering the global Gibbs state from the reduced state on the region $\alpha\beta$.

Theorem 1, and refer the reader interested in the exact quantitative bounds and mathematical details to the Appendices.

To start with, we mention that the entanglement structure of Gibbs states with finite correlation length is known to fulfill an area law for the correlation measure quantum mutual information [10],

$$I(A:B)_\rho := H(A)_\rho + H(B)_\rho - H(AB)_\rho, \quad (5)$$

where $H(A)_\rho := -\mathrm{Tr}[\rho_A \log_2 \rho_A]$ denotes the von Neumann entropy. Now, to prove Theorem 1 it would be sufficient to extend that to an area law for another correlation measure: the so-called entanglement of formation [29],

$$E_F(A:B)_\rho := \inf \sum_i p_i H(A)_{\rho^i}, \quad (6)$$

where the infimum is over all pure state decompositions $\rho_{AB} = \sum_i p_i |\rho^i\rangle\langle\rho^i|_{AB}$. However, since the mutual information is neither a convex nor a concave function in the state it generally behaves very differently than the entanglement of formation. In particular, there exist quantum states with $E_F(A:B)_\rho \gg I(A:B)_\rho$ [30] and hence this argument cannot be used to prove the desired statement.

Another way to proceed for $L = \alpha\beta\gamma$ would be to focus on the quantum conditional mutual information, $I(\alpha:\gamma|\beta)_\rho = H(\alpha\beta)_\rho + H(\beta\gamma)_\rho - H(\beta)_\rho - H(\alpha\beta\gamma)_\rho$, where β connects the regions α and β (see Fig. 1). If now $I(\alpha:\gamma|\beta)_\rho$ would be decaying exponentially in the size of β, then recent results in quantum information theory [31–33] show that we could recover the state on $\alpha\beta\gamma$ from the one on just $\alpha\beta$ using a completely positive trace-preserving map acting only on β and having only a few Kraus operators. As already noted in [34,35], this would allow us to repeat that argument and obtain a representation of the global state using a sequence of quasilocal maps, also called a local Markov chain structure.

This in turn will be the entry point of our arguments, as discussed in the following. Unfortunately, while we expect that for many interesting physical systems the quantum conditional mutual information behaves as expected, a general statement for thermal states of local Hamiltonians is not known. However, the local Markov chain structure of Gibbs states of one-dimensional local Hamiltonians still holds approximately [19].

Lemma 2 (Markov chain structure of Gibbs states). Let H be a local one-dimensional Hamiltonian. Then, for any tripartite split $L := \alpha\beta\gamma$ as depicted in Fig. 1, there exists a local quantum channel $\Lambda_{\beta\to\beta\gamma}$ such that

$$\|\rho_{H,T} - (\mathcal{I}_\alpha \otimes \Lambda_{\beta\to\beta\gamma})(\rho_{\alpha\beta})\|_1 \leqslant \exp(-q(T)\sqrt{d(\alpha,\gamma)}), \quad (7)$$

where $d(\alpha,\gamma)$ measures the distance in system size between the regions α and γ, and $q(T)$ denotes some temperature dependent constant.

As shown in [19] the intuition behind the recovery channel in Lemma 2 is as follows. First one uses the quantum belief propagation equations from [36] to obtain the following approximate decomposition of the Gibbs state:

$$\exp(-H_{\alpha\beta\gamma}/T) \approx P_{\beta,t}\big(\exp\big(-H_{\alpha\beta_L}/T\big) \otimes \exp(-H_{\beta_R\gamma}/T)\big)P_{\beta,t}^{\dagger}, \tag{8}$$

where $\beta_L\beta_R$ are regions each composed of half the qubits of β, and $P_{\beta,t}$ is a local operator of size $t := d(\alpha,\gamma)/2$ centered in the cut between β_L and β_R with $\|P_{\beta,t}\|_\infty \leqslant 2^{O(1/T)}$. The approximation error in Eq. (8) is exponentially small in t. From this decomposition one can define the completely positive and trace-nonincreasing map

$$K_{\beta\to\beta\gamma}(\cdot) := P_{\beta,t}\big(\mathrm{Tr}_{\beta_R}\big[P_{\beta,t}^{-1}(\cdot)\big(P_{\beta,t}^{-1}\big)^{\dagger}\big]\otimes\rho_{H_{\beta_R\gamma}}\big)P_{\beta,t}^{\dagger}, \tag{9}$$

and show that $\rho_{\alpha\beta\gamma} \approx (\mathcal{I}_A \otimes K_{\beta\to\beta\gamma})(\rho_{\alpha\beta})$ up to an error $2^{-O(t)}$. This almost achieves our goal; however, the map $K_{B\to BC}$ is not trace preserving. To make it trace preserving one uses a repeat-until-success strategy by probabilistically implementing $K_{\beta\to\beta\gamma}$: if it succeeds the state is recovered and if it fails one traces out the region it was applied to and a shield region just next to it. Now since the Gibbs state has an exponential decay of correlations [37] and since the map is applied with probability $2^{-O(1/T)}$ independently of the system size, one obtains a good approximation of the reduced state of the Gibbs state (on a slightly smaller length). One can then repeat the procedure until it is successful, and this increases the total error to $2^{-O(\sqrt{t})}$ (see [19] for more details).

With the help of Lemma 2 we are now ready to give a proof of our main result.

Proof of Theorem 1. Our argument is based on three steps:

(i) Using Lemma 2 we construct a matrix product state $|\Psi_{D,\varepsilon}\rangle$ with bond dimension D quasipolynomial in the system size n and ε^{-1}.

(ii) We show that $|\Psi_{D,\varepsilon}\rangle$ is the purification of a convex combination of matrix product states with bond dimension D as in step (i), denoted by $\rho[\mu_\varepsilon]$.

(iii) We show that $\rho[\mu_\varepsilon]$ is close to the Gibbs state $\rho_{H,T}$.

For step (i) we split the lattice into three consecutive regions (as depicted in Fig. 2):

$$L := A_1B_1C_1\ A_2B_2C_2\ \cdots\ A_IB_IC_I \tag{10}$$

of dimension $|A_i| = |B_i| = 2^l$ and $|C_i| = 2^{5l\xi}$, where $\xi \leqslant \exp(c/T)$ denotes the correlation length of the Gibbs state [37]. We then prepare a purification $|\rho^i\rangle_{\bar{A}_iA_i\bar{B}_iB_i}$ of all the reduced states on A_iB_i of the Gibbs state $\rho_{H,T}$ and fill in the missing C_i pieces together with their purifications $\bar{C}_i$ by making use of the Markov chain structure of the Gibbs state (Lemma 2). In more detail, we apply Lemma 2 to the full lattice L with the decomposition

$$\gamma_i := C_i,\ \beta_i := B_iA_{i+1}, \text{and } \alpha_i := L/(\beta_i\gamma_i), \tag{11}$$

leading to quantum channels $\Lambda^i_{B_iA_{i+1}\to B_iA_{i+1}C_i}$ with the approximation property as in Eq. (7). We then apply all the corresponding (minimal) Stinespring dilations of these channels leading to

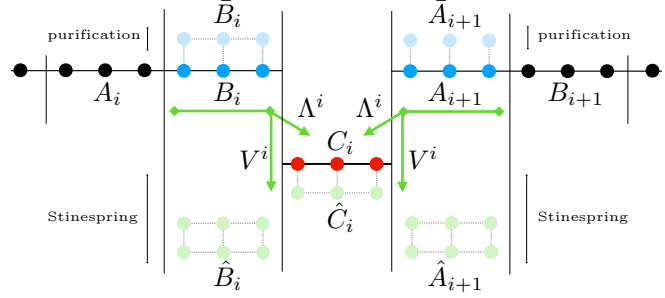


FIG. 2. Proof of our main result (Theorem 1): starting from the purifications of the reduced states on A_iB_i of the Gibbs state (blue region) we create the C_i pieces together with their purifications (red region) by applying the Stinespring dilations V^i of the local Markov channels Λ^i from Lemma 2 (green region). We then get the desired convex combination of matrix product states on the main lattice $A_iB_iC_i$ by tracing out all the purifying registers $\bar{A}_{i+1}\bar{B}_i$ and the Stinespring registers $\hat{B}_i\hat{A}_{i+1}\hat{C}_i$.

$$|\Psi_{D,\varepsilon}\rangle := \bigotimes_{i=1}^{I} V^i_{B_iA_{i+1}\to B_iA_{i+1}C_i\hat{B}_i\hat{A}_{i+1}\hat{C}_i}|\rho^i\rangle_{\bar{A}_iA_i\bar{B}_iB_i}, \tag{12}$$

where $|\hat{B}_i| = |B_i|^2$, $|\hat{A}_{i+1}| = |A_{i+1}|^2$, and $|\hat{C}_i| = |C_i|$. It is straightforward to check that the resulting global pure state $|\Psi_{D,\varepsilon}\rangle$ becomes a matrix product state with bond dimension upper bounded by $D \leqslant 2^{l(8+10\xi)}$ and choosing $l = \log_2^2(n/\varepsilon)$ establishes step (i).

For step (ii) we arrive at $\rho[\mu_\varepsilon]$ from $|\Psi_{D,\varepsilon}\rangle$ by tracing out the purifying registers $\bar{A}_{i+1}\bar{B}_i$ as well as the Stinespring registers $\hat{B}_i\hat{A}_{i+1}\hat{C}_i$. By the monotonicity of the Schmidt rank under stochastic local operations assisted by classical communication (SLOCC) [38] this exactly creates a convex combination of matrix product states with bond dimension upper bounded by D from step (i).

Finally, step (iii) is deduced from the approximate quantum Markov structure as in Eq. (7) together with a telescoping sum argument as well as a nonlockability bound from [39, Lem. 20]. We refer to the Appendices for a derivation of the exact error bounds. ■

IV. GIBBS STATES AND THE METTS ALGORITHM

An algorithm to construct thermal ensembles of MPSs is already available in the literature. It was introduced by White and goes by the name of the minimally entangled typical thermal states (METTS) algorithm. Our theorem provides a theoretical justification for this algorithm. As in the MPO case, we can start from the trivial infinite temperature state and evolve it with imaginary time $\beta/2$ on both sides. However, instead of keeping the exact representation, we can represent the infinite temperature state by uniformly sampling over the basis of product states $|\vec{i}\rangle = |i_1\rangle\cdots|i_n\rangle$, and capture the action of the evolution $\exp(-\beta/2H)$ on $|\vec{i}\rangle$ by promoting it to an

MPS. We thus obtain the representation

$$\frac{1}{Z}e^{-\beta H} = \sum_{\vec{i}} p(\vec{i})|\phi_{T,\vec{i}}\rangle\langle\phi_{T,\vec{i}}|, \tag{13}$$

with $p(\vec{i}) = Z^{-1}\,\langle\vec{i}\,|\exp(-\beta H)|\vec{i}\,\rangle$ being the probability of sampling the METTS $|\phi_{T,\vec{i}}\rangle = p(\vec{i})^{-1/2}\exp(-\beta H/2)|\vec{i}\rangle$. To accurately approximate the probability distribution $p(\vec{i})$, a Markov process was defined using standard time evolution algorithms for MPSs, and, because the steps are small and the entanglement growth is limited, this process satisfies an approximate version of detailed balance and typically leads to a good approximation of the equilibrium state. Note, however, that there is no guarantee that this algorithm converges to the true Gibbs state, as otherwise we would be able to simulate systems that are believed to be intractable (see, e.g., [40]). Our theorem shows that a representation in terms of a mixture of MPSs exists, not that it is easy to find this mixture.

V. HYDRODYNAMICS AND CONTINUOUS TIME UPDATES

So far, we have only considered the static case, that is, thermal equilibrium. But now consider a one-dimensional physical system which is first in thermal equilibrium and then, at time zero, subject to a quench such as a local spin flip at the origin, given by a unitary U. The system will eventually converge again into thermal equilibrium, albeit at a different temperature due to the injection of energy at time zero. Our theorem shows that at both ends of the time evolution, the state can be represented as a mixture of MPSs. It is hence natural to assume that the time dependent density matrix can at all times be represented as a mixture of matrix product states, as the "classical" entropy suppresses the quantum correlations. We are hence interested in developing an algorithm for doing time evolution with mixtures of MPSs.

Recently, starting from [28] Leviatan and co-workers [22] suggested an algorithm for time-evolving quantum systems at infinite temperature. Their algorithm commences by sampling matrix product states from a given distribution. In [22], mostly the uniform distribution was considered, but given our thoughts it is of course very plausible to start with the state given by the METTS algorithm, as this is expected to approximate the thermal state before the spin flip. Next, we apply the local spin flip, which is easily implemented on each matrix product state. Indeed, acting with U on an MPS produces just another MPS. Hence acting with the local spin flip on the output of the METTS algorithm gives rise to another convex combination of MPSs,

$$\sum_{\vec{i}} q(\vec{i})\,|\psi_{T,\vec{i}}\rangle\langle\psi_{T,\vec{i}}| \to \sum_{\vec{i}} q(\vec{i})\,U^\dagger|\psi_{T,\vec{i}}\rangle\langle\psi_{T,\vec{i}}|U \tag{14}$$

$$= \sum_{\vec{i}} q(\vec{i})\,|\psi'_{T,\vec{i}}\rangle\langle\psi'_{T,\vec{i}}|, \tag{15}$$

with $|\psi'_{T,\vec{i}}\rangle = U^\dagger|\psi_{T,\vec{i}}\rangle$. The system is then subject to the time evolution under the Hamiltonian H, $|\psi'_{T,\vec{i}}\rangle \to \exp(-itH)|\psi'_{T,\vec{i}}\rangle$. In general, this action will not produce an MPS, and the approximation would be even quite bad [41]. However, the key idea of [22] is to ignore this difficulty and just project the state $\exp(-itH)|\psi'_{T,\vec{i}}\rangle$ back onto the manifold M^D_{MPS} using the method of TDVP applied to MPSs [28,42] such as to preserve all local constants of motion. This leads to an update rule which as a whole takes the initial convex combination of MPSs to a new mixture of MPSs, which gives a completely new interpretation of their algorithm. Their algorithm can immediately be generalized as follows.

Considering an infinitesimal step of the evolution, we can transform the update rule to act directly on the probability distribution μ, giving rise to a Fokker-Planck-like equation. Starting from the pure state evolution $i\frac{\partial|\psi(t)\rangle}{\partial t} = H|\psi(t)\rangle$, we can systematically decompose the right-hand side into terms acting parallel to the state, in the tangent space, in the double tangent space, and so on:

$$H|\psi\rangle = E|\psi\rangle + h^i|\partial_i\psi\rangle + h^{ij}|\partial_i\partial_j\psi\rangle + \cdots, \tag{16}$$

where the first three terms already give rise to an equality in case of a nearest neighbor Hamiltonian. Here, partial derivatives correspond to derivatives with respect to the variational parameters, i.e., every single entry of every single MPS tensor A. Furthermore, we have $E = \langle\psi|H|\psi\rangle$, $h^i = g^{i,\bar{\jmath}}\langle\partial_{\bar{\jmath}}\psi|H - E|\psi\rangle$ where $g^{i,\bar{\jmath}}$ is the inverse of the metric $g_{\bar{\imath},j} = \langle\partial_{\bar{\imath}}\psi|\partial_j\psi\rangle$ of the MPS manifold,

$$h^{ij} = g^{ij,\bar{k}\bar{l}}\big[\langle\partial_{\bar{k}}\partial_{\bar{l}}\psi|H - E|\psi\rangle - \langle\partial_{\bar{k}}\psi|H - E|\psi\rangle\langle\partial_{\bar{l}}\psi|\psi\rangle - \langle\partial_{\bar{l}}\psi|H - E|\psi\rangle\langle\partial_{\bar{k}}\psi|\psi\rangle\big] \tag{17}$$

with $g_{\bar{k}\bar{l},ij} = \langle\partial_{\bar{k}}\partial_{\bar{l}}\psi|\partial_i\partial_j\psi\rangle$, and so on. Partial integration in the convex combination of MPS leads to

$$\begin{aligned}\frac{\partial\mu(\bar{A},A,t)}{\partial t} &= i[\partial_i(h^i(\bar{A},A,t)\mu(\bar{A},A,t)) \\ &\quad - \partial_{\bar{\imath}}(\overline{h^i}(\bar{A},A,t)\mu(\bar{A},A,t))] \\ &\quad - i[\partial_i\partial_j(h^{i,j}(\bar{A},A,t)\mu(\bar{A},A,t)) \\ &\quad - \partial_{\bar{\imath}}\partial_{\bar{\jmath}}(\overline{h^{i,j}}(\bar{A},A,t)\mu(\bar{A},A,t))].\end{aligned} \tag{18}$$

The first-order derivative term gives rise to a drift term of the probability distribution, and corresponds to the prediction of the TDVP (as applied in Ref. [22]). In particular, this term will preserve a pure state. While the second derivative term might be interpreted as a diffusion term at first, one easily observes that the corresponding diffusion matrix is not positive definite. This is to be expected, as we would otherwise capture the exact evolution (for a nearest neighbor Hamiltonian) with our convex combination of MPSs, even when starting from an initial pure state. In that case, the above equation will immediately give rise to negative values of the probability distribution μ and therefore to a sign problem when trying to sample. However, for initial distributions that are sufficiently broad, the above evolution might not destroy positivity right away and could yield an improvement over the pure drift case considered in Ref. [22].

VI. CONCLUSIONS

Using recently developed tools from quantum information theory, we proved that thermal states of local, one-dimensional Hamiltonians can be approximately written as a

convex combination of matrix product states. We employed this fact to reconsider two algorithms developed within condensed matter theory, and showed how our result provides a theoretical justification, much like the work of Hastings [8] explains the success of DMRG. We believe that the description of thermal states in term of mixtures will provide the right framework for simulating quantum hydrodynamical effects. The main question left open by our work is if the bound dimension scaling can be improved from quasipolynomial to polynomial in the system size n and ε^{-1}. Note that this would be the case if we could obtain bounds on the conditional mutual information as strong as those argued for in [34,35]. We emphasize that, in contrast to Hasting's result for the MPS representation of ground states [8], our approximation of thermal states in Theorem 1 does not assume the Hamiltonian to be gapped. Hence, correlations can be in general long-ranged and there is the possibility that Theorem 1 can only be improved for, e.g., gapped Hamiltonians. Finally, it would be interesting to study extensions of our result to nonlocal Hamiltonians. The limitation of our current proof strategy is that we crucially employ a result of Araki [37], which states that Gibbs states of one-dimensional local Hamiltonians have an exponential decay of correlations.

ACKNOWLEDGMENTS

We thank Isaac Kim for pointing out an error in an earlier version of this paper and Brian Swingle for discussions. J.H. acknowledges support from the ERC via Grant No. 715861 (ERQUAF). F.V. acknowledges funding from the ERC grant QUTE, and the sfb projects foqus and vicom.

APPENDIX A: PROOF OF THEOREM 1

Let $H = \sum_i h_i$ be a one-dimensional short-range Hamiltonian with $\|h_i\|_\infty \leqslant 1$. For a temperature $T > 0$, let

$$\rho^{H,T} := \frac{\exp(-H/T)}{\mathrm{Tr}[\exp(-H/T)]} \tag{A1}$$

be the corresponding Gibbs state. We will make use of the following restatement of Lemma 2 from the main text, which shows that Gibbs states of one-dimensional local Hamiltonians have an approximate local Markov chain structure [19, Thm. 1].

Lemma 3 (Local Markov chain structure of Gibbs states [19]). *For every tripartite split of the lattice $\alpha\beta\gamma$, there exists a completely positive and trace preserving map $\Lambda_{\beta\to\beta\gamma}$ such that*

$$\left\|\rho^{H,T}_{\alpha\beta\gamma} - (\mathcal{I}_\alpha \otimes \Lambda_{\beta\to\beta\gamma})\left(\rho^{H,T}_{\alpha\beta}\right)\right\|_1 \leqslant \exp(-q(T)\sqrt{d(\alpha,\gamma)}), \tag{A2}$$

for any $d(\alpha,\gamma) \geqslant l_0$ and $q(T) := C\exp(-c/T)$ [where $0 < l_0, C, c < 100$ are universal constants and $d(\alpha,\gamma)$ quantifies the minimal distance in system size between α and γ].

More precisely we will make use of an adapted, purified version of [19, Cor. 4]. This original corollary says that Gibbs states of one-dimensional local Hamiltonians can be well approximated by a depth-2 circuit with each gate acting locally on $O(\log_2^2(n))$ qubits. The following is a more precise restatement of Theorem 1 from the main text that we seek to prove.

Theorem 4. For an n-qubit system with fixed temperature $T > 0$ we can write

$$\left\|\rho^{H,T} - \sum_{j=1}^{J} p_j |\varphi^j\rangle\langle\varphi^j|\right\|_1 \leqslant \varepsilon, \tag{A3}$$

where p_j denotes a probability distribution and all the $|\varphi^j\rangle$ are matrix product states with bond dimension

$$D = 2^{O(\log_2^2(n/\varepsilon))}. \tag{A4}$$

Furthermore, J is exponential in n and the temperature dependence of the bond dimension is $\exp\{\exp[O(1/T)]\}$.

The full argument is shown here for maximal readability, where some steps similar to the proof of [19, Cor. 4] are reproduced.

Proof. The proof is done in multiple steps:

(i) We construct one global matrix product state $|\Psi_{D,\varepsilon}\rangle$ with bond dimension D as in (A4).

(ii) We show that this state $|\Psi_{D,\varepsilon}\rangle$ is the purification of some state $\sum_{j=1}^{J} p_j|\varphi^j\rangle\langle\varphi^j|$ with all the $|\varphi^j\rangle$ matrix product states having bond dimension D as in (A4).

(iii) We show that this state $\sum_{j=1}^{J} p_j|\varphi^j\rangle\langle\varphi^j|$ is ε-close to the Gibbs state $\rho^{H,T}$ in trace distance.

For (i) split the lattice into three consecutive regions $A_1B_1C_1\,A_2B_2C_2\,\cdots A_IB_IC_I$ of dimensions $|A_i| = |B_i| = 2^l$ and $|C_i| = 2^{5l\xi}$, with $\xi \leqslant \exp(c/T)$ the correlation length of the Gibbs state where c is the constant from Lemma 3.[2] Set $\rho^{H,T} =: \rho_{A_1B_1C_1\cdots A_IB_IC_I}$. To construct the global state $|\Psi_{D,\varepsilon}\rangle$ we first prepare a purification of all the marginals $\rho_{A_iB_i}$ of the Gibbs state: $|\zeta^i\rangle_{A_i\bar{A}_iB_i\bar{B}_i}$ with $|\bar{A}_i| = |A_i|$ and $|\bar{B}_i| = |B_i|$. Now the local Markov chain structure of Gibbs states (Lemma 3) implies the existence of completely positive trace preserving maps $\Lambda^i_{B_iA_{i+1}\to B_iA_{i+1}C_i}$ such that[3]

$$\left\|\Lambda^i_{B_iA_{i+1}\to B_iA_{i+1}C_i}\left(\rho_{A_1B_1C_1A_iB_i\cdots A_IB_IC_I}\right) - \rho_{A_1B_1C_1\cdots A_IB_IC_I}\right\|_1 \leqslant \exp(-q(T)\sqrt{\log_2|B_i|}) \tag{A5}$$

$$= \exp(-q(T)\sqrt{l}). \tag{A6}$$

Denote Stinespring isometries of the $\Lambda^i_{B_iA_{i+1}\to B_iA_{i+1}C_i}$ by $V^i_{B_iA_{i+1}\to B_iA_{i+1}C_i\hat{D}_i}$, where we chose the dilation registers as $\hat{D}_i := \hat{A}_i\hat{B}_{i+1}\hat{C}_i$, with $|\hat{A}_i| = |A_i|^2$, $|\hat{B}_{i+1}| = |B_{i+1}|^2$, and

[2] Gibbs states of 1D local Hamiltonians have an exponential decay of correlations [37].

[3] For choosing $\alpha_i = A_1B_1C_1A_{i-1}B_{i-1}C_{i-1}A_iB_{i+1}C_{i+1}\cdots A_IB_IC_I$, $\beta_i = B_iA_{1+1}$, and $\gamma_i = C_i$.

$|\hat{C}_i| = |C_i|$ (cf. Fig. 2 from the main text).[4] We define

$$|\Psi_{D,\varepsilon}\rangle_{A_1\bar{A}_1B_1\bar{B}_1C_1\hat{D}_1\cdots A_I\bar{A}_IB_I\bar{B}_IC_I\hat{D}_I} := \bigotimes_{i=1}^{I} V^i_{B_iA_{i+1}\to B_iA_{i+1}C_i\hat{D}_i}\left(\bigotimes_{i=1}^{I}|\zeta^i\rangle_{A_i\bar{A}_iB_i\bar{B}_i}\right), \quad \text{(A7)}$$

where the last $V^I_{B_I\to B_IC_I\hat{D}_I}$ is only acting on B_I. Since small Stinespring dilations imply small Schmidt rank (Lemma 5) this becomes a matrix product state with bond dimension[5]

$$\max_{i=1,\ldots,I}|A_i||B_i|\times|A_{i+1}||B_i|\times|A_{i+1}|^2|B_i|^2|C_i|^2 \leqslant 2^{l(8+10\xi)}. \quad \text{(A8)}$$

Choosing $l = \log_2^2(n/\varepsilon)$ establishes step (i).

For (ii) we trace out the purifying systems $\bar{A}_1\bar{B}_1\cdots\bar{A}_I\bar{B}_I$ as well as the Stinespring dilation registers $\hat{D}_1\cdots\hat{D}_I$, leading to the state

$$\bigotimes_{i=1}^{I}\Lambda^i_{B_iA_{i+1}\to B_iA_{i+1}C_i}\left(\bigotimes_{i=1}^{I}\rho_{A_iB_i}\right) =: \sum_{j=1}^{J}|\bar{\varphi}^j\rangle\langle\bar{\varphi}^j|_{A_1B_1C_1\cdots A_IB_IC_I}, \quad \text{(A9)}$$

where each $|\bar{\varphi}^j\rangle\langle\bar{\varphi}^j|$ is generated by conditioning on a basis element of the traced out systems. We normalize to

$$|\varphi^j\rangle\langle\varphi^j| := |\bar{\varphi}^j\rangle\langle\bar{\varphi}^j|/p_j \text{ with } p_j := \||\varphi^j\rangle\|_2^2, \quad \text{(A10)}$$

and choose a local basis for the traced out systems $\bar{A}_1\bar{B}_1\hat{D}_1\cdots\bar{A}_I\bar{B}_I\hat{D}_I$ (see Fig. 2 from the main text). We can then use the fact that the Schmidt rank is monotone under local operations and post-selection—so-called SLOCC monotonicity (see, e.g., [38])—and hence conclude that all the $|\varphi^j\rangle$ are again matrix product states with bond dimension D as in (A4). J becomes exponential in n.

For (iii) we need to show that the state in Eq. (A9) is close to the Gibbs state $\rho_{A_1B_1C_1\cdots A_IB_IC_I}$ (the following part is a slightly simplified version of the proof of [19, Cor. 4]). We first note that, by the monotonicity of the trace distance under partial trace, Eq. (A6) implies

$$\left\|\Lambda^i_{B_iA_{i+1}\to B_iA_{i+1}C_i}\left(\rho_{A_1B_1C_1\cdots A_{i-1}B_{i-1}C_{i-1}A_iB_iA_{i+1}B_{i+1}}\right) - \rho_{A_1B_1C_1\cdots A_iB_iC_iA_{i+1}B_{i+1}}\right\|_1 \leqslant \exp(-q(T)\sqrt{l}). \quad \text{(A11)}$$

We estimate

$$\left\|\bigotimes_{i=1}^{I}\Lambda^i_{B_iA_{i+1}\to B_iA_{i+1}C_i}\left(\bigotimes_{i=1}^{I}\rho_{A_iB_i}\right) - \rho_{A_1B_1C_1\cdots A_IB_IC_I}\right\|_1 \leqslant \left\|\Lambda^1_{B_1A_2\to B_1A_2C_1}(\rho_{A_1B_1}\otimes_{A_2B_2}) - \rho_{A_1B_1C_1A_2B_2}\right\|_1 + \left\|\bigotimes_{i=2}^{I}\Lambda^i_{B_iA_{i+1}\to B_iA_{i+1}C_i}\left(\rho_{A_1B_1C_1A_2B_2}\bigotimes_{i=3}^{I}\rho_{A_iB_i}\right) - \rho_{A_1B_1C_1\cdots A_IB_IC_I}\right\|_1, \quad \text{(A12)}$$

where we have used the triangle inequality for the trace distance and the monotonicity of the trace distance under partial trace (plus the subadditivity with respect to tensor products). To bound the first term in Eq. (A12) we use the estimate

$$\begin{aligned}&\left\|\rho_{A_1B_1C_1\cdots A_{i-1}B_{i-1}A_iB_i} - \rho_{A_1B_1C_1\cdots A_{i-1}B_{i-1}}\otimes\rho_{A_iB_i}\right\|_1 \\ &\leqslant (|A_i||B_i|)^2\max_{\|X\|_\infty,\|Y\|_\infty\leqslant 1}\left|\mathrm{Tr}\left[\left(X_{A_1B_1C_1\cdots A_{i-1}B_{i-1}A_iB_i}\otimes Y_{A_iB_i}\right)\left(\rho_{A_1B_1C_1\cdots A_{i-1}B_{i-1}A_iB_i} - \rho_{A_1B_1C_1\cdots A_{i-1}B_{i-1}}\otimes\rho_{A_iB_i}\right)\right]\right| \quad \text{(A13)}\\ &\leqslant 2^{4l}\times 2^{\frac{-\log_2|C_{i-1}|}{\xi}} \quad \text{(A14)}\\ &\leqslant \exp(-l), \quad \text{(A15)}\end{aligned}$$

where we used [39, Lem. 20] for the first inequality and the exponential decay of correlations for the second inequality [37]. In particular, this then implies

$$\begin{aligned}&\left\|\Lambda^1_{B_1A_2\to B_1A_2C_1}\left(\rho_{A_1B_1}\otimes_{A_2B_2}\right) - \rho_{A_1B_1C_1A_2B_2}\right\|_1 \\ &\leqslant \left\|\Lambda^1_{B_1A_2\to B_1A_2C_1}\left(\rho_{A_1B_1A_2B_2}\right) - \rho_{A_1B_1C_1A_2B_2}\right\|_1 + \exp(-l) \quad \text{(A16)}\\ &\leqslant \exp(-q(T)\sqrt{l}) + \exp(-l), \quad \text{(A17)}\end{aligned}$$

where we used Eq. (A11) in the second step together with the monotonicity of the trace distance under partial trace. To estimate the second term in Eq. (A12) we iterate the argument, leading to

$$\begin{aligned}&\left\|\bigotimes_{i=1}^{I}\Lambda^i_{B_iA_{i+1}\to B_iA_{i+1}C_i}\left(\bigotimes_{i=1}^{I}\rho_{A_iB_i}\right) - \rho_{A_1B_1C_1\cdots A_IB_IC_I}\right\|_1 \\ &\leqslant I\times[\exp(-q(T)\sqrt{l}) + \exp(-l)] \quad \text{(A18)}\\ &\leqslant I\times O(\varepsilon/n), \quad \text{(A19)}\end{aligned}$$

since $l = \log_2^2(n/\varepsilon)$ as chosen before. We have $I\leqslant n$ and hence the claim follows. ■

APPENDIX B: MISCELLANEOUS LEMMAS

Lemma 5 (Small Stinespring dilations imply small Schmidt rank). Let $|\varphi\rangle_{\alpha_2\alpha_1\beta_1\beta_2} := |\varphi\rangle_{\alpha_2\alpha_1}\otimes|\varphi\rangle_{\beta_1\beta_2}$ with Schmidt-rank upper bounded by d for any cut through $\alpha_2\alpha_1$ or through $\beta_1\beta_2$. Moreover, let $\Lambda_{\alpha_1\beta_1}$ be a completely positive and trace

[4]The dimension of the Stinespring dilation register is, without loss of generality, as small as input dimension times output dimension of the completely positive trace preserving map in question (see, e.g., [43]).

[5]In more detail, to apply Lemma 5 we need to extend the maps $\Lambda^i_{B_iA_{i+1}\to B_iA_{i+1}C_i}$ to maps $\Lambda^i_{B_iA_{i+1}C_i}$, e.g., by just first throwing the C_i register away and then applying $\Lambda^i_{B_iA_{i+1}\to B_iA_{i+1}C_i}$. Using Lemma 5 with $\alpha_2 = A_1\bar{A}_1B_1\bar{B}_1\cdots A_i\bar{A}_i\bar{B}_i$, $\alpha_1 = B_iC_i$, $\beta_1 = A_{i+1}$, and $\beta_2 = \bar{A}_{i+1}B_{i+1}\bar{B}_{i+1}\cdots A_k\bar{A}_kB_k\bar{B}_k$ gives Eq. (A8).

preserving map with Kraus decomposition $\{K^i_{\alpha_1\beta_1}\}_{i=1}^I$. Then, there exists a Stinespring dilation isometry $V_{\alpha_1\beta_1\to\alpha_1\gamma\beta_1}$ of $\Lambda_{\alpha_1\beta_1}$ such that the Schmidt rank of

$$|\Psi_{D,\varepsilon}\rangle_{\alpha_2\alpha_1\gamma\beta_1\beta_2} := \left(1_{\alpha_2}\otimes V_{\alpha_1\beta_1\to\alpha_1\gamma\beta_1}\otimes 1_{\beta_2}\right)|\varphi\rangle_{\alpha_2\alpha_1\beta_1\beta_2} \tag{B1}$$

in any bipartite cut through $\alpha_2\alpha_1\gamma\beta_1\beta_2$ is upper bounded by $d|\alpha_1||\beta_1|\times I$.

Proof. The total system is given by $\alpha_2\alpha_1\gamma\beta_1\beta_2$ and we treat the different possible bipartite cuts separately:

(i) For any cut through α_2 as well as for the cut $\alpha_2|\alpha_1\gamma\beta_1\beta_2$ the Schmidt rank is upper bounded by $\log_2 d$ since $V_{\alpha_1\beta_1\to\alpha_1\gamma\beta_1}$ does not act on $\alpha_2\beta_2$.

(ii) For the cut $\alpha_2\alpha_1|\gamma\beta_1\beta_2$ we choose

$$V_{\alpha_1\beta_1\to\alpha_1\gamma\beta_1}(\cdot) = \sum_{i=1}^{I} K^i_{\alpha_1\beta_1}(\cdot)\otimes|i\rangle_\gamma. \tag{B2}$$

Taking the operator Schmidt decomposition for each $K^i_{\alpha_1\beta_1}$ in the cut $\alpha_1|\beta_1$ gives

$$K^i_{\alpha_1\beta_1} = \sum_{k_i=1}^{|\alpha_1||\beta_1|} K^{k_i}_{\alpha_1}\otimes K^{k_i}_{\beta_1}, \tag{B3}$$

and we find for the output state

$$|\Psi_{D,\varepsilon}\rangle_{\alpha_2\alpha_1|\gamma\beta_1\beta_2} = \sum_{i=1}^{I}\sum_{k_i=1}^{|\alpha_1||\beta_1|} \underbrace{\left(1_{\alpha_2}\otimes K^{k_i}_{\alpha_1}\right)|\varphi\rangle_{\alpha_2\alpha_1}}_{=:|\varphi^{ik_i}\rangle_{\alpha_2\alpha_1}} \otimes \underbrace{|i\rangle_\gamma\otimes\left(K^{k_i}_{\beta_1}\otimes 1_{\beta_2}\right)|\varphi\rangle_{\beta_1\beta_2}}_{=:|\varphi^{ik_i}\rangle_{\gamma\beta_1\beta_2}}. \tag{B4}$$

(iii) Cuts through γ are already treated by Eq. (B4).

(iv) For any cut $\alpha_1 = \alpha_1^a\alpha_1^b$ we take the Schmidt decomposition

$$|\varphi\rangle_{\alpha_2\alpha_1} = \sum_{j=1}^{d}|\varphi^j\rangle_{\alpha_2\alpha_1^a}\otimes|\varphi^j\rangle_{\alpha_1^b}, \tag{B5}$$

as well as the following operator Schmidt decomposition of the Kraus operators in Eq. (B2):

$$K^i_{\alpha_1\beta_1} = \sum_{k_i=1}^{|\alpha_1||\beta_1|} K^{k_i}_{\alpha_1^a}\otimes K^{k_i}_{\alpha_1^b\beta_1}. \tag{B6}$$

Hence, we find for the output state

$$|\Psi_{D,\varepsilon}\rangle_{\alpha_2\alpha_1^a|\alpha_1^b\gamma\beta_1\beta_2} = \sum_{i=1}^{I}\sum_{k_i=1}^{|\alpha_1||\beta_1|}\sum_{j=1}^{d} \underbrace{\left(1_{\alpha_2}\otimes K^{k_i}_{\alpha_1^a}\right)|\varphi^j\rangle_{\alpha_2\alpha_1^a}}_{=:|\varphi^{jik_i}\rangle_{\alpha_2\alpha_1^a}} \otimes \underbrace{\left(K^{k_i}_{\alpha_1^b\beta_1}\otimes 1_{\gamma\beta_2}\right)\left(|\varphi^j\rangle_{\alpha_1^b}\otimes|i\rangle_\gamma\otimes|\varphi\rangle_{\beta_1\beta_2}\right)}_{=:|\varphi^{jik_i}\rangle_{\alpha_1^b\gamma\beta_1\beta_2}}. \tag{B7}$$

(v) Cuts through β_2, the cut $\alpha_2\alpha_1\gamma\beta_1|\beta_2$, as well as cuts of the form $\beta_1 = \beta_1^a\beta_2^b$ follow by symmetry.

We conclude the claim since we have treated all cuts. ■

[1] D. Pérez-García, M. M. Wolf, M. Sanz, F. Verstraete, and J. I. Cirac, Phys. Rev. Lett. **100**, 167202 (2008).
[2] F. Pollmann, A. M. Turner, E. Berg, and M. Oshikawa, Phys. Rev. B **81**, 064439 (2010).
[3] X. Chen, Z.-C. Gu, and X.-G. Wen, Phys. Rev. B **83**, 035107 (2011).
[4] N. Schuch, D. Pérez-García, and I. Cirac, Phys. Rev. B **84**, 165139 (2011).
[5] F. Verstraete, V. Murg, and J. Cirac, Adv. Phys. **57**, 143 (2008).
[6] U. Schollwöck, Ann. Phys. **326**, 96 (2011).
[7] F. Verstraete and J. I. Cirac, Phys. Rev. B **73**, 094423 (2006).
[8] M. B. Hastings, J. Stat. Mech.: Theory Exp. (2007) P08024.
[9] Z. Landau, U. Vazirani, and T. Vidick, Nat. Phys. **11**, 566 (2015).
[10] M. M. Wolf, F. Verstraete, M. B. Hastings, and J. I. Cirac, Phys. Rev. Lett. **100**, 070502 (2008).
[11] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac, Phys. Rev. Lett. **93**, 207204 (2004).
[12] M. Zwolak and G. Vidal, Phys. Rev. Lett. **93**, 207205 (2004).
[13] B. Pirvu, V. Murg, J. I. Cirac, and F. Verstraete, New J. Phys. **12**, 025012 (2010).
[14] M. B. Hastings, Phys. Rev. B **73**, 085115 (2006).
[15] A. Molnar, N. Schuch, F. Verstraete, and J. I. Cirac, Phys. Rev. B **91**, 045138 (2015).
[16] C. Karrasch, J. H. Bardarson, and J. E. Moore, Phys. Rev. Lett. **108**, 227206 (2012).
[17] T. Barthel, New J. Phys. **15**, 073010 (2013).
[18] G. De las Cuevas, N. Schuch, D. Pérez-García, and J. I. Cirac, New J. Phys. **15**, 123021 (2013).
[19] K. Kato and F. Brandao, arXiv:1609.06636.
[20] S. R. White, Phys. Rev. Lett. **102**, 190601 (2009).
[21] E. M. Stoudenmire and S. R. White, New J. Phys. **12**, 055026 (2010).
[22] E. Leviatan, F. Pollmann, J. H. Bardarson, and E. Altman, arXiv:1702.08894.
[23] M. Fannes, B. Nachtergaele, and R. F. Werner, Commun. Math. Phys. **144**, 443 (1992).
[24] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, Quantum Inf. Comput. **7**, 401 (2007).
[25] J. Haegeman, M. Mariën, T. J. Osborne, and F. Verstraete, J. Math. Phys. **55**, 021902 (2014).
[26] U. Schollwöck, Rev. Mod. Phys. **77**, 259 (2005).
[27] G. Vidal, Phys. Rev. Lett. **93**, 040502 (2004).
[28] J. Haegeman, J. I. Cirac, T. J. Osborne, I. Pižorn, H. Verschelde, and F. Verstraete, Phys. Rev. Lett. **107**, 070601 (2011).
[29] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).

[30] P. Hayden, D. W. Leung, and A. Winter, Commun. Math. Phys. **265**, 95 (2006).
[31] O. Fawzi and R. Renner, Commun. Math. Phys. **340**, 575 (2015).
[32] M. Junge, R. Renner, D. Sutter, M. M. Wilde, and A. Winter, Ann. Henri Poincaré **19**, 2955 (2018).
[33] D. Sutter, M. Berta, and M. Tomamichel, Commun. Math. Phys. **352**, 37 (2017).
[34] B. Swingle and J. McGreevy, Phys. Rev. B **94**, 155125 (2016).
[35] F. Brandao and M. J. Kastoryano, arXiv:1609.07877.
[36] M. B. Hastings, Phys. Rev. B **76**, 201102(R) (2007).
[37] H. Araki, Commun. Math. Phys. **14**, 120 (1969).
[38] G. Vidal, Phys. Rev. Lett. **83**, 1046 (1999).
[39] F. Brandao and M. Horodecki, Commun. Math. Phys. **333**, 761 (2015).
[40] D. Aharonov, D. Gottesman, S. Irani, and J. Kempe, Commun. Math. Phys. **287**, 41 (2009).
[41] T. Prosen and M. Znidaric, Phys. Rev. E **75**, 015202 (2007).
[42] J. Haegeman, C. Lubich, I. Oseledets, B. Vandereycken, and F. Verstraete, Phys. Rev. B **94**, 165116 (2016).
[43] D. Kretschmann, D. Schlingemann, and R. F. Werner, IEEE Trans. Inf. Theory **54**, 1708 (2008).