Efficient Description of Many-Body Systems with Matrix Product Density Operators

Jiří Guth Jarkovský[1,2] András Molnár,[1,2,3,4] Norbert Schuch,[1,2,*] and J. Ignacio Cirac[1,2]

[1]*Max-Planck-Institute of Quantum Optics, Hans-Kopfermann-Str. 1, Garching 85748, Germany*

[2]*Munich Center for Quantum Science and Technology, Schellingstr. 4, München 80799, Germany*

[3]*Instituto de Ciencias Matemáticas, Campus Cantoblanco UAM, C/ Nicolás Cabrera, 13-15, Madrid 28049, Spain*

[4]*Dpto. Análisis Matemático y Matemática Aplicada, Universidad Complutense de Madrid, Madrid 28040, Spain*

Matrix product states form a powerful ansatz for the simulation of a wide range of one-dimensional quantum systems that are in a pure state. Their power stems from the fact that they faithfully approximate states with a low amount of entanglement, the "area law." However, in order to accurately capture the physics of realistic systems, one generally needs to apply a mixed-state description. In this work, we establish the mixed-state analog of this characterization. We show that one-dimensional mixed states with a low amount of entanglement, quantified by the entanglement of purification, can be efficiently approximated by matrix product density operators.

I. MOTIVATION

Complex interacting quantum many-body systems cannot be understood without carefully assessing the structure of their quantum correlations, that is, entanglement. This is the key insight behind the density-matrix renormalization group (DMRG) method for the simulation of one-dimensional (1D) systems [1] and it has later allowed to explain DMRG as a variational method over the manifold of matrix product states (MPSs) [2,3]. This understanding has in turn triggered the development of a whole range of methods, such as for the simulation of excitations, time evolution, or scattering. At the same time, MPSs have been generalized to describe mixed states using matrix product density operators (MPDOs), which have proven successful in describing systems in thermal equilibrium, out-of-equilibrium dynamics, and, generally, the simulation of noisy systems in experimental scenarios [4–8].

How can this success be explained from the entanglement structure of the underlying systems? For pure states, this question has been resolved. It is precisely the states with a low degree of entanglement that can be faithfully described by MPSs [9,10]. This is captured by the so-called

entanglement area law, which states for 1D systems that for any contiguous region $\mathcal{R}$, the entanglement $E(\mathcal{R} : \mathcal{R}^c)$ between $\mathcal{R}$ and its complement $\mathcal{R}^c$ is constant or grows at most moderately: $E(\mathcal{R} : \mathcal{R}^c) \leq \text{const.} \times \log |\mathcal{R}|$, where $|\mathcal{R}|$ is the size of $\mathcal{R}$ (as opposed to the $E(\mathcal{R} : \mathcal{R}^c) \approx |\mathcal{R}|$ "volume" scaling expected for a generic state [11]). This gives a simple and tangible entanglement-based criterion for assessing whether an MPS description can be used in a given case and which, in particular, is known to hold, e.g., for ground states and low-lying excited states of reasonable physical systems [12,13].

However, in order to describe realistic physical systems, the effects of noise and coupling to the environment need to be taken into account. This means that the state of those systems will generally be mixed and a description of the system in terms of MPDOs will be required. Unfortunately, however, we do not have a simple and accessible criterion at our disposal that allows us to assess whether the entanglement structure of any specific system of interest, such as in a given experimental setup, makes it susceptible to an MPDO-based description. In fact, the only scenario where such a result is known is the suitability of MPDOs for the description of thermal states [14,15], but even there, it does not provide insights into the relevant structure of the entanglement. At the same time, in all scenarios where the system is out of thermal equilibrium, such as dissipative dynamics, sequential generation of states, relaxation, coupling to nonthermal baths, or any other question of experimental interest, no simple entanglement-based criteria for the applicability of an MPDO-based description are known.

————

*norbert.schuch@gmail.com

II. RESULT

In this paper, we derive a simple entanglement-based criterion to assess whether a state can be approximated by MPDOs. We show that a suitable family of entanglement area laws—even with a logarithmic correction—implies that the state can be efficiently approximated by an MPDO, that is, an ansatz of the form

$$\sigma_D = \sum A^{[1]}_{i_1 j_1} A^{[2]}_{i_2 j_2} \cdots A^{[N]}_{i_N j_N} |i_1, \ldots, i_N\rangle \langle j_1, \ldots, j_N|. \quad (1)$$

Here, N is the length of the chain, the sum runs over $i_k, j_k = 1, \ldots, d$, with d being the local Hilbert space dimension, and the $A^{[k]}_{i_k j_k}$ are $D \times D$ matrices, except for $A^{[1]}_{i_1 j_1}$ and $A^{[N]}_{i_N j_N}$, which are $1 \times D$ and $D \times 1$ matrices, respectively.

Specifically, given a mixed state ρ on a chain of N d-level spins, we prove that if there exist constants $c > 0$ and $0 < \lambda < 1$ such that the α-Rényi entanglement of purification $E_{p,\alpha}$ (defined below) [16] satisfies

$$E_{p,\alpha}(\mathcal{R} : \mathcal{R}^c) \leq c \log N, \quad \text{for } \alpha = \frac{\lambda}{5 \log_2 N}, \quad (2)$$

then ρ can be efficiently approximated by an MPDO σ_D. As long as the bond dimension D scales polynomially, $D = N^\kappa$ for any $\kappa > 2c/(1 - \lambda)$, the error $\varepsilon := \|\rho - \sigma_D\|_1$ in trace norm goes to zero superpolynomially in N (i.e., faster than any inverse polynomial) [17]. By using the trace norm—which exactly bounds the error in expectation values of arbitrary bounded observables (with largest eigenvalue 1)—we obtain a bound on the error incurred in arbitrary simulations of physical processes. This establishes that MPDOs are precisely the framework needed to faithfully describe mixed states that obey an entanglement area law of the above form.

III. PROOF

A. Proof structure

The proof initially follows the approach in the pure-state case [9]. However, we need to deviate from it halfway through. Rather than sequentially concatenating low-rank approximations across different cuts, we need to take an entirely different route and use a renormalizationlike approach in which we glue cuts in a treelike fashion. The reason is deeply rooted in the different structure of mixed states. We require a good approximation in trace norm, which—unlike the 2-norm—does not induce a scalar product, which in turn is essential to build norm-preserving projections. At the same time, we cannot bound the 2-norm instead. The relative bound $\|\rho\|_1 \leq \sqrt{\mathcal{D}}\|\rho\|_2$, where $\mathcal{D} = d^N$ is the dimension of the total space, is tight (saturated by the maximally mixed state), that is, the trace norm can be exponentially larger in N than the 2-norm, breaking the efficiency of the approximation.
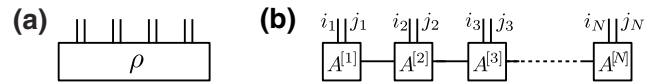

FIG. 1. (a) The tensor notation of density matrix ρ. Each pair of legs denotes the ket and bra index at one site. (b) The tensor network for an MPDO [see Eq. (1)]. The legs denote indices and the connected lines contraction (summation) of indices, corresponding to the matrix products in Eq. (1).

We will use the conventional graphical calculus for MPSs and MPDOs [4–8], where a (mixed) many-body state is denoted as a box with legs (double legs denote ket + bra) [Fig. 1(a)] and an MPDO is expressed as a tensor network, in which tensors are boxes, each leg denotes a tensor index, and connecting legs corresponds to contraction [Fig. 1(b)].

Let us briefly sketch the proof strategy. First, we show that for any bipartition, a bound on the entanglement implies that the target state ρ can be well approximated by a low-rank decomposition across that cut. An area law thus implies that ρ has low-rank approximations across every cut. The crucial step will then be to merge these approximations. To start, we show how to merge two approximations in such a way that (i) we still obtain a good approximation and (ii) the internal structure of the two states is preserved (specifically, existing lower-rank approximations across other cuts), as this allows us to iterate the procedure. In a final step, we then show how to nest this merging procedure in such a way as to obtain a good MPDO approximation of the target state ρ.

B. Entanglement versus approximability for bipartitions

First, we relate entanglement and approximability across a bipartition. We start by defining the α-Rényi entanglement of purification $E_{p,\alpha}(\rho_{AB})$ [16]. For a bipartite state ρ_{AB}, it is given by

$$E_{p,\alpha}(\rho_{AB}) = \min_{|\psi\rangle} E_\alpha(|\psi\rangle), \quad (3)$$

where the minimum is taken over all purifications $|\psi\rangle_{AA'BB'}$ of ρ_{AB}, i.e., $\text{tr}_{A'B'}|\psi\rangle\langle\psi| = \rho_{AB}$, and $E_\alpha(|\psi\rangle) = S_\alpha(\text{tr}_{BB'}|\psi\rangle\langle\psi|)$, with $S_\alpha(\rho) = (1/(1-\alpha)) \log(\text{tr}\rho^\alpha)$, the α-Rényi entanglement entropy quantifying the pure-state entanglement between AA' and BB'. For the remainder of this paper, we restrict to $0 \leq \alpha < 1$.

A key result from the pure-state case [9] is that a small $E_\alpha(|\psi\rangle)$ implies a rapid decay of the Schmidt coefficients and, thus, there exists a low-rank approximation to $|\psi\rangle$. Concretely, for any D_p, there exists a $|\chi_{D_p}\rangle =$

$\sum_{i=1}^{D_p} |\chi_i^L\rangle_{AA'} |\chi_i^R\rangle_{BB'}$ such that

$$\eta := 1 - |\langle \psi | \chi_{D_p} \rangle|^2 \leq \left(\frac{(1-\alpha) \exp[E_{p,\alpha}(\rho_{AB})]}{D_p} \right)^{(1-\alpha)/\alpha}$$

(4)

(and thus D_p scales as an inverse polynomial in the error η). This is equivalent to

$$\| |\psi\rangle\langle\psi| - |\chi_{D_p}\rangle\langle\chi_{D_p}| \|_1 \leq 2\sqrt{1 - |\langle\psi|\chi_{D_p}\rangle|^2} = 2\sqrt{\eta},$$

(5)

where $\| \cdot \|_1$ is the trace norm (i.e., the sum of the singular values) [18]. By tracing $A'B'$ and using the fact that tracing (as a completely positive trace-preserving map) is contractive under the trace norm, we arrive at

$$\delta := \|\rho_{AB} - \sigma_D\|_1 \leq 2\sqrt{\eta},$$

(6)

for some

$$\sigma_D = \sum_{i,j=1}^{D_p} A_{ij} \otimes B_{ij}$$

(7)

with rank $D = D_p^2$ (where $A_{ij} = \text{tr}_{A'} |\chi_i^L\rangle\langle\chi_j^L|$ and $B_{ij} = \text{tr}_{B'} |\chi_i^R\rangle\langle\chi_j^R|$).

Let us now turn toward a spin chain of length N the state ρ of which obeys an area law, that is, there is an $E_{\max}^\alpha$ such that $E_{p,\alpha}(\rho_{AB}) \leq E_{\max}^\alpha$ for any bipartition $A = 1, \ldots, L$, $B = L+1, \ldots, N$. Combining Eqs. (6) and (4), we have that for each cut, there exists a rank $D = D_p^2$ decomposition of the form (7) with trace-norm error

$$\delta \leq 2\sqrt{\eta} \leq 2 \left(\frac{(1-\alpha) \exp[E_{\max}^\alpha]}{D^{1/2}} \right)^{(1-\alpha)/2\alpha}.$$

(8)

What remains to be seen is whether it is possible to merge these different low-rank approximations. However, at this point we can no longer use the purifications to resort to the pure-state result, since the optimal purifications (minimizing $E_{p,\alpha}$) for different cuts need not be related [19]. We thus require a different approach.

C. Merging two approximations

To start, consider a bipartite state $\rho \equiv \rho_{AB}$ (obtained by blocking sites), a truncated approximation

$$\sigma_1 = \sum_{i=1}^D A_i \otimes B_i, \quad \|\rho - \sigma_1\| \leq \delta_1$$

(9)

across some cut, and another approximation σ_2,

$$\|\rho - \sigma_2\| \leq \delta_2,$$

(10)

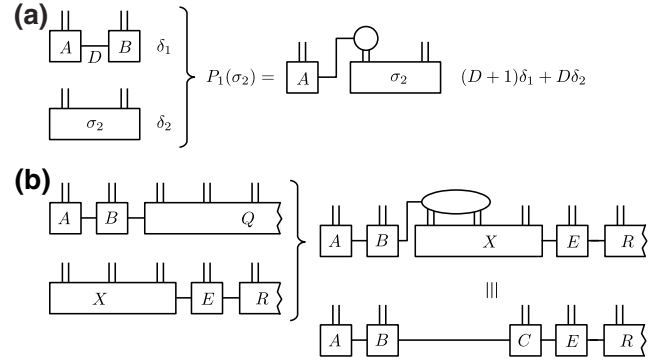obtained, e.g., by truncating across a different cut. Let us now try to connect those two approximations. To

FIG. 2. (a) The merging of two approximations with errors δ_1 and δ_2 yields an approximation with error $(D+1)\delta_1 + D\delta_2$. Here, the circular tensor denotes $\sum_i c_{ik}\hat{A}_i'$ [cf. Eq. (12)]. (b) Merging preserves existing cuts (i.e., the local MPDO structure); here, we apply (a) to the first two sites versus the rest.

this end, consider a (not necessarily orthogonal) projection P_1 onto span$\{A_i\}$, $P_1(A_i) = A_i$, which can be written as $P_1(X) = \sum_{i=1}^D \hat{A}_i \text{tr}[(\hat{A}_i')^\dagger X]$ for some basis $\hat{A}_i = \sum c_{ik}A_k$ of span$\{A_i\}$ and some (dual) matrices $\hat{A}_i'$ satisfying $\text{tr}[(\hat{A}_i')^\dagger \hat{A}_j] = \delta_{ij}$. P_1 can be naturally embedded into the full space as

$$P_1(X) = \sum_{i=1}^D \hat{A}_i \otimes \text{tr}_A[(\hat{A}_i' \otimes \mathbb{1}_B)^\dagger X].$$

(11)

Now consider

$$P_1(\sigma_2) = \sum_{i,k=1}^D c_{ik}A_k \otimes \text{tr}_A[(\hat{A}_i' \otimes \mathbb{1}_B)^\dagger \sigma_2]$$

(12)

[see Fig. 2(a)]. First, it also has rank D across the cut; second, the left part is spanned by A_k and thus inherits the structure of the left part of σ_1; and, third, the right part is obtained from σ_2 by tracing its left part with $(A_i')^\dagger$ and thus inherits the structure of the right part of σ_2. In particular, if σ_1 and σ_2 have parts on the left and right, respectively, which are already in matrix product form; both of these are inherited by $P_1(\sigma_2)$ [see Fig. 2(b)]. We can then iterate this scheme, starting from truncations at individual cuts, to obtain an MPDO approximation.

What is the approximation error of the merged truncation $P_1(\sigma_2)$? Using $P_1(\sigma_1) = \sigma_1$ from Eq. (9), we have

$$\|P_1(\sigma_2) - \rho\|_1 \leq \|P_1(\sigma_2) - P_1(\sigma_1)\|_1 + \|P_1(\sigma_1) - \rho\|_1$$

$$\leq \|P_1(\sigma_2 - \sigma_1)\|_1 + \|\sigma_1 - \rho\|_1$$

$$\leq \|P_1(X)\|_1 + \delta_1,$$

(13)

with $X := \sigma_2 - \sigma_1$, $\|X\|_1 \le \delta_1 + \delta_2$. Starting from Eq. (11), we find

$$
\begin{aligned}
\|P_1(X)\|_1 &\le \sum \|\hat{A}_i \otimes \mathrm{tr}_A[(\hat{A}'_i \otimes \mathbb{1})^\dagger X]\|_1 \\
&= \sum \|\hat{A}_i\|_1 \, \|\mathrm{tr}_A[(\hat{A}'_i \otimes \mathbb{1})^\dagger X]\|_1 \\
&\overset{(*)}{\le} \sum \|\hat{A}_i\|_1 \, \|(\hat{A}'_i \otimes \mathbb{1})^\dagger X]\|_1 \\
&\le \sum \|\hat{A}_i\|_1 \, \|\hat{A}'_i \otimes \mathbb{1}\|_\infty \, \|X\|_1 \\
&\le \sum_{i=1}^{D} \|\hat{A}_i\|_1 \, \|\hat{A}'_i\|_\infty \, \|X\|_1,
\end{aligned} \tag{14}
$$

where $(*)$ uses the contractivity of the partial trace and all sums run over $i = 1, \ldots, D$.

To keep $\|P_1(X)\|_1$ small, we thus ideally want to choose $\hat{A}_i$ and $\hat{A}'_i$ such that $\|\hat{A}_i\|_1 = \|\hat{A}'_i\|_\infty = 1$ (this is optimal as $\delta_{ij} = \mathrm{tr}[(\hat{A}'_i)^\dagger \hat{A}_j] \le \|\hat{A}_i\|_1 \|\hat{A}'_i\|_\infty$). It turns out that such $\{\hat{A}_i\}$, $\{\hat{A}'_i\}$ indeed exist, a standard result in functional analysis [20]. To start, note that the space $\mathcal{A} = \mathrm{span}\{A_i\}$ with norm $\|\cdot\|_1$ is a normed space. In any such space, there exists an *Auerbach basis*: a basis $\{\hat{A}_i\}$ of $\mathcal{A}$, together with a set of linear functionals $\hat{a}'_j : \mathcal{A} \to \mathbb{C}$ such that $\hat{a}'_j(\hat{A}_i) = \delta_{ij}$ and $\|\hat{A}_i\|_1 = \|\hat{a}'_j\| = 1$. Using the Hahn-Banach theorem, we can now extend the bounded functionals $\hat{a}'_j$ defined on $\mathcal{A}$ to bounded linear functionals $\hat{a}'_{j,\mathrm{ext}} : \mathcal{M}_n(\mathbb{C}) \to \mathbb{C}$ defined on the full matrix space $\mathcal{M}_n(\mathbb{C})$ (where n is the size of the A_i), such that $\|\hat{a}'_{j,\mathrm{ext}}\| = 1$. Clearly, all these $a'_{j,\mathrm{ext}}$ can be uniquely expressed as $a'_{j,\mathrm{ext}}(Y) = \mathrm{tr}[(\hat{A}'_j Y]$, where $\|\hat{A}'_j\|_\infty = \|\hat{a}'_{j,\mathrm{ext}}\| = 1$. We thus obtain a set of matrices $\{\hat{A}_i\}_i$ and $\{\hat{A}'_j\}_j$, $i,j = 1, \ldots, D$, such that $\mathrm{span}\{\hat{A}_i\} = \mathcal{A}$, $\|\hat{A}_i\|_1 = \|\hat{A}'_j\|_\infty = 1$, and $\mathrm{tr}[(\hat{A}'_j)^\dagger A_i] = \delta_{ij}$, as desired.

By inserting these $\{\hat{A}_i\}$, $\{\hat{A}'_i\}$ in Eq. (14), we arrive at $\|P_1(X)\|_1 \le D \|X\|_1$, which together with Eq. (13) and $\|X\|_1 \le \delta_1 + \delta_2$ yields

$$
\|P_1(\sigma_2) - \rho\|_1 \le (D+1)\delta_1 + D\delta_2. \tag{15}
$$

That is, we have merged the two approximations σ_1 and σ_2, with a new error as above; if both $\delta_1, \delta_2 \le \delta$, the new error is at most $(2D+1)\delta$.

D. Merging all approximations

At this point, we can start to concatenate truncations using Eq. (15). However, we cannot do this sequentially as one would do for the 2-norm (where one can choose P_1 as the orthogonal projection for which $\|P_1(X)\| \le \|X\|$; note that this yields an alternative proof for the result of Ref. [9]). The prefactor $(2D+1)$ would grow exponentially with the number of steps, rendering the bound useless. To overcome this issue, we choose a renormalizationlike
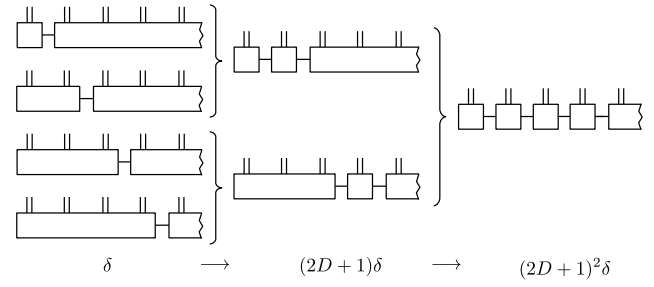

FIG. 3. The merging of cuts in a treelike fashion. In each step, the number of cuts is doubled and the error grows by a factor of $(2D+1)$.

procedure, where we concatenate the cuts in a treelike fashion, as shown in Fig. 3, using Eq. (15) in each step. One can readily check that each step doubles the number of cuts and multiplies the error with $(2D+1)$; if the number of cuts is not a power of 2, we can start some branches of the tree later or we can pad the spin chain with trivial (uncorrelated) spins. For a chain of length N, this scheme thus requires $K = \lceil \log_2(N-1) \rceil \le (\log_2 N) + 1$ steps and thus incurs a total error of

$$
\varepsilon = (2D+1)^K \delta \le (2D+1)^{\log_2 N + 1} \delta. \tag{16}
$$

We are now at the point at which we can combine our results. Combination of Eqs. (8) and (16) yields

$$
\varepsilon \le 2(2D+1)^{\log_2 N + 1} \left(\frac{(1-\alpha)e^{E_{\max}^\alpha}}{D^{1/2}} \right)^{(1-\alpha)/2\alpha}. \tag{17}
$$

If we now—following Eq. (2)—choose

$$
\alpha = \frac{\lambda}{5 \log_2 N}, \quad e^{E_{\max}^\alpha} = N^c, \quad \text{and} \quad D = N^\kappa, \tag{18}
$$

with $0 < \lambda < 1$, $\kappa > 2c/(1-\lambda)$, we have [using $1 - \alpha \le 1$, $(1-\alpha)/2\alpha \ge 2\log_2 N/\lambda$, $2D+1 < 3D$, and $N^c/D^{1/2} \le 1$]

$$
\varepsilon \le 2(3D)^{\log_2 N + 1} \left(\frac{N^c}{D^{1/2}} \right)^{2\log_2 N/\lambda} \tag{19}
$$

$$
\le 6D \left(\frac{3N^{2c/\lambda}}{D^{1/\lambda - 1}} \right)^{\log_2 N} = 6N^\kappa \left(\frac{3}{N^\Delta} \right)^{\log_2 N} \to 0, \tag{20}
$$

with $\Delta = (1/\lambda)(\kappa(1-\lambda) - 2c) > 0$, which thus goes to zero superpolynomially as $N \to \infty$. This completes the proof of our result.

IV. CONCLUSIONS

In summary, in this work we establish when MPDOs can efficiently describe quantum many-body systems that are in a mixed state, which is key to an accurate description of

realistic physical systems. We derive the conditions that a state ρ has to fulfill such that it can be approximated by an MPDO with a polynomial bond dimension. In particular, we show that, for a sequence of density operators ρ on a spin chain of length N, an entanglement area law implies an efficient approximability of ρ by MPDOs.

To be specific, we consider a family of area-law bounds for the Rényi entanglement of purification that limit the quantum correlations to growing at most logarithmically with the system size N; in this setting, we find that there exist MPDO approximations to ρ with a bond dimension that grows polynomially in the system size N and for which the approximation error decreases faster than any inverse polynomial in N. This shows that MPDOs provide a faithful approximation to density operators that satisfy an area law and are thus well suited for the numerical simulation as well as the analytical study of such systems.

Finally, our work gives rise to a diverse range of follow-up questions. Can a similar result be established for other entanglement measures and what would be their relation to our finding? Can our result, or a variant thereof, be combined with a suitable area law for thermal states (such as in Ref. [21]), to obtain an alternative approximability proof for Gibbs states via an area law, similar to the pure-state case? Is there a way to avoid the treelike merging of cuts in favor of, e.g., a sequential merging scheme or is there a deeper physical reason behind the necessity of this choice? And, finally, can a similar connection between area laws and approximability also be established in higher dimensions?

[1] S. R. White, Density Matrix Formulation for Quantum Renormalization Groups, Phys. Rev. Lett. **69**, 2863 (1992).

[2] S. Östlund and S. Rommer, Thermodynamic Limit of Density Matrix Renormalization, Phys. Rev. Lett. **75**, 3537 (1995).

[3] F. Verstraete, D. Porras, and J. I. Cirac, DMRG and Periodic Boundary Conditions: A Quantum Information Perspective, Phys. Rev. Lett. **93**, 227205 (2004).

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

[5] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, Ann. Phys. **326**, 96 (2011).

[6] N. Schuch, Condensed Matter Applications of Entanglement Theory, arXiv:1306.5551.

[7] R. Orus, A practical introduction to tensor networks: Matrix product states and projected entangled pair states, Ann. Phys. **349**, 117 (2014).

[8] J. C. Bridgeman and C. T. Chubb, Hand-waving and interpretive dance: An introductory course on tensor networks, J. Phys. A: Math. Theor. **50**, 223001 (2017).

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

[10] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, Entropy Scaling and Simulability by Matrix Product States, Phys. Rev. Lett. **100**, 30504 (2008).

[11] P. Hayden, D. Leung, P. W. Shor, and A. Winter, Aspects of generic entanglement, Commun. Math. Phys. **250**, 371 (2004).

[12] M. Hastings, An area law for one dimensional quantum systems, J. Stat. Mech. P08024 (2007).

[13] I. Arad, Z. Landau, U. Vazirani, and T. Vidick, Rigorous RG algorithms and area laws for low energy eigenstates in 1D, Commun. Math. Phys. **356**, 65 (2017).

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

[15] A. Molnar, N. Schuch, F. Verstraete, and J. I. Cirac, Approximating Gibbs states of local Hamiltonians efficiently with PEPS, Phys. Rev. B **91**, 045138 (2015).

[16] B. M. Terhal, M. Horodecki, D. W. Leung, and D. P. DiVincenzo, The entanglement of purification, J. Math. Phys. **43**, 4286 (2002).

[17] Note that due to the monotonicity of the α-Rényi entropy in α, we can also choose any smaller α in Eq. (2).

[18] This is readily checked by working in the two-dimensional space spanned by the two vectors; note that $|\chi_D\rangle$ need not be normalized.

[19] Different purifications are related by a unitary on the purifying system $A'B'$—which, however, mixes A' and B' and thus changes the entanglement properties.

[20] D. Werner, *Funktionalanalysis* (Springer, Heidelberg, 1997).

[21] 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).