Generalized Transfer Matrix States from Artificial Neural Networks

Lorenzo Pastori,[1,2] Raphael Kaubruegger,[2,3,4] and Jan Carl Budich[1]

[1] *Institute of Theoretical Physics, Technische Universität Dresden, 01062 Dresden, Germany*
[2] *Department of Physics, University of Gothenburg, SE 412 96 Gothenburg, Sweden*
[3] *Center for Quantum Physics, Faculty of Mathematics, Computer Science and Physics, University of Innsbruck, A-6020 Innsbruck, Austria*
[4] *Institute for Quantum Optics and Quantum Information of the Austrian Academy of Sciences, A-6020 Innsbruck, Austria*

(Dated: October 31, 2019)

Identifying variational wavefunctions that efficiently parametrize the physically relevant states in the exponentially large Hilbert space is one of the key tasks towards solving the quantum many-body problem. Powerful tools in this context such as tensor network states have recently been complemented by states derived from artificial neural networks (ANNs). Here, we propose and investigate a new family of quantum states, coined generalized transfer matrix states (GTMS), which bridges between the two mentioned approaches in the framework of deep ANNs. In particular, we show by means of a constructive embedding that the class of GTMS contains generic matrix product states while at the same time being capable of capturing more long-ranged quantum correlations that go beyond the area-law entanglement properties of tensor networks. While generic deep ANNs are hard to contract, meaning that the corresponding state amplitude can not be exactly evaluated, the GTMS network is shown to be analytically contractible using transfer matrix methods. With numerical simulations, we demonstrate how the GTMS network learns random matrix product states in a supervised learning scheme, and how augmenting the network by long-ranged couplings leads to the onset of volume-law entanglement scaling. We argue that this capability of capturing long-range quantum correlations makes GTMS a promising candidate for the study of critical and dynamical quantum many-body systems.

I. INTRODUCTION

The quantum many-body problem is one of the outstanding challenges in physics. Besides providing deep theoretical insights, its solution may enable revolutionary technological applications including room-temperature superconductivity and new nano-technology enabled by the understanding of complex macro-molecules. The key issue in this context is the exponential complexity of generic quantum many-body states with the number of constituents. A widely applicable and successful approach towards taming this exploding complexity is to devise families of variational states that efficiently parametrize the physical scenario under investigation.

A paradigmatic example along these lines is provided by tensor network states such as matrix product states (MPS) [1–5], and their higher dimensional generalizations [6–11]. The tensor network ansatz describes exponentially decaying correlations as reflected in the area-law entanglement of the wavefunction, thus successfully capturing ground states of gapped local Hamiltonians [12–14]. Tensor network states have also become an important tool for the study of critical systems [15–17] and dynamical properties [18–20] even though it is clear that these situations exhibit quantum correlations beyond area-law entanglement. The price to pay for encompassing such scenarios is to increase the size of the tensors, i.e. the number of variational parameters, with system size and (exponentially) with time, respectively. Hence, the stronger growth of entanglement limits the amenable system sizes and periods of time-evolution.

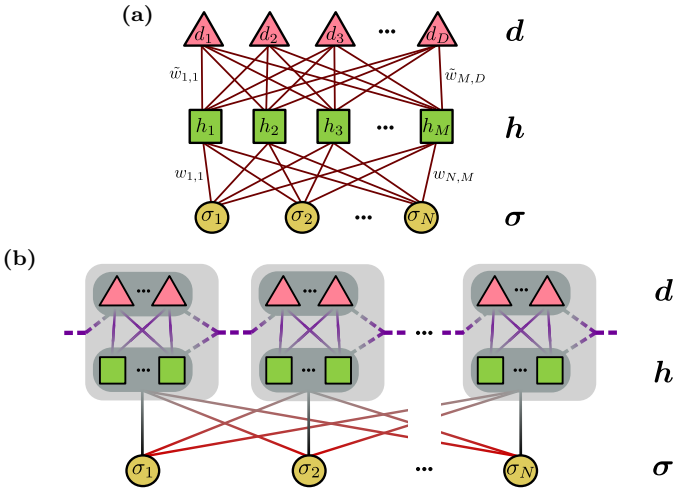

Figure 1: (color online) **(a)**: Three-layer DBM architecture. Yellow circles denote the physical sites $\boldsymbol{\sigma} = \{\sigma_1, ..., \sigma_N\}$, the green boxes the hidden units $\boldsymbol{h} = \{h_1, ..., h_M\}$ and the pink triangles the deep units $\boldsymbol{d} = \{d_1, ..., d_D\}$. Complex weights $w_{i,j}$ and $\tilde{w}_{j,k}$ are associated to the links between physical/hidden and hidden/deep layers respectively. Not drawn in the picture are the bias terms. **(b)** Schematic representation of a GTMS network. Hidden and deep units are grouped in blocks (grey-shaded boxes) corresponding to transfer matrices. Keeping only black and purple links yields a MPS, while the red links make the transfer matrices dependent on all the physical spins. By cutting the purple links in the GTMS network, one retrieves a RBM.

Complementing existing variational approaches [1–

11, 21–26], quantum states derived from artificial neural networks (ANN) [27–31] have recently been introduced and studied. There, the physical degrees of freedom are coupled to a set of auxiliary units (see Fig. 1a), and the wavefunction is obtained summing over all configurations of the auxiliary degrees of freedom (contracting the network), thus retaining the couplings as variational parameters. Analytically contracting one of the simplest ANN architectures known as restricted Boltzmann machine (RBM) [30] already leads to quantum states [31] that exhibit volume-law entanglement [32], thus offering an alternative variational wavefunction for those situations in which short range quantum correlations captured by conventional tensor network methods may not be sufficient [33–36]. Using the more complex ANN class of deep Boltzmann machines (DBM) [30, 37–39] (see Fig. 1a), it has recently been proven that the imaginary time evolution towards the ground state of a generic many-body Hamiltonian can be exactly represented at polynomial network complexity [39]. However, this does not imply an efficient solution of a given many-body problem, since the exact contraction of a DBM in general is a exponentially hard problem. Hence, the explicit form of the wavefunction is in general not accessible even if it can be efficiently represented graphically in the DBM framework.

The purpose of this work is to develop a hybrid approach bridging between tensor network and ANN states. To this end, we introduce and study a class of exactly contractible DBM networks, which we coin generalized transfer matrix state (GTMS) networks (see Fig. 1b). There the wavefunction is analytically evaluated using transfer matrix methods. Quite remarkably, the resulting GTMS are capable of arbitrarily interpolating between MPS and RBM states thus combining key physical properties of these two powerful variational methods. As a limiting case, we obtain conventional MPS (RBM states) from the GTMS architecture by cutting the red (purple) couplings in Fig. 1b. To demonstrate how arbitrary MPS are efficiently parameterized in the proposed framework, we show that GTMS networks can indeed *learn* random MPS by optimizing the coupling parameters in a supervised learning scheme. Furthermore, we argue how the GTMS generalizes and augments the class of MPS by making the tensors non-locally dependent on the physical degrees of freedom. This more complex structure allows the GTMS to capture correlations beyond area-law entanglement. Our analysis is supported by numerical studies on the scaling of the 2^{nd} Rényi entanglement entropy, showing that with the addition of non-local neural couplings in the network (red links in Fig. 1b) the GTMS indeed acquires volume-law entanglement. This increased representational power makes the GTMS a promising candidate for the study of critical and time-dependent systems.

Several previous works have investigated the relationship between ANN sates and tensor network states [40–43], establishing a general correspondence between certain Boltzmann machine (among which short-range RBM) architectures and MPS representations of quantum states. Going beyond these previous insights, our present construction provides a constructive and efficient embedding of MPS and RBM states into the general framework of DBM networks. This allows us to continuously interpolate between MPS and RBM states and generalize both approaches in a physically motivated fashion.

The remainder of the Paper is organized as follows. In the next Section we review the concept of a deep Boltzmann machine state. In Section III we introduce the GTMS, discussing the DBM architectures that can be exactly contracted by means of a transfer matrix approach, and showing how it can interpolate between MPS and RBM. In Section IV we explicitly show that MPS are a special case of GTMS, and we construct the paradigmatic AKLT state from a GTMS network as an example. We also numerically show that the GTMS architecture can be trained to learn generic random MPS. In Section V we provide a numerical analysis of the 2^{nd} Rényi entanglement entropy of the GTMS, and present a concluding discussion in Section VI.

II. DEEP BOLTZMANN MACHINE STATES

To set the general stage for our construction, we start with a brief discussion of deep Boltzmann machine (DBM) states [30, 37, 38]. We focus on a three-layer architecture, in which the auxiliary units can be organized in two distinct neural layers which we call hidden and deep layers (see Fig 1a). The input to such a DBM network is the physical layer, i.e. the set of N physical spins (or any quantum number locally associated to the sites of the physical system) σ_i ($i = 1, ..., N$). The hidden and deep layers are chosen as sets of classical Ising spins with values $h_j = \pm 1$, $d_k = \pm 1$ ($j = 1, ..., M$, $k = 1, ..., D$). We denote the physical spin configuration with $\boldsymbol{\sigma} = \{\sigma_1, ..., \sigma_N\}$, the hidden spin configuration with $\boldsymbol{h} = \{h_1, ..., h_M\}$ and the deep spin configuration with $\boldsymbol{d} = \{d_1, ..., d_D\}$. The physical and hidden spins are coupled by a set of (complex) weights $w_{i,j}$ representing the links of the network. Hidden and deep layers are coupled as well by a set of weights $\tilde{w}_{j,k}$. Additionally, the weights c_i, b_j and a_k are referred to as bias terms, and play the role of local fields for the physical σ_i, hidden h_j and deep d_k spins, respectively. All these couplings, collectively denoted by $\boldsymbol{w} = \{c_i, b_j, w_{i,j}, \tilde{w}_{j,k}, a_k\}_{i,j,k}$ play the role of the variational parameters for the DBM state. The connectivity of the network is encoded in a function,

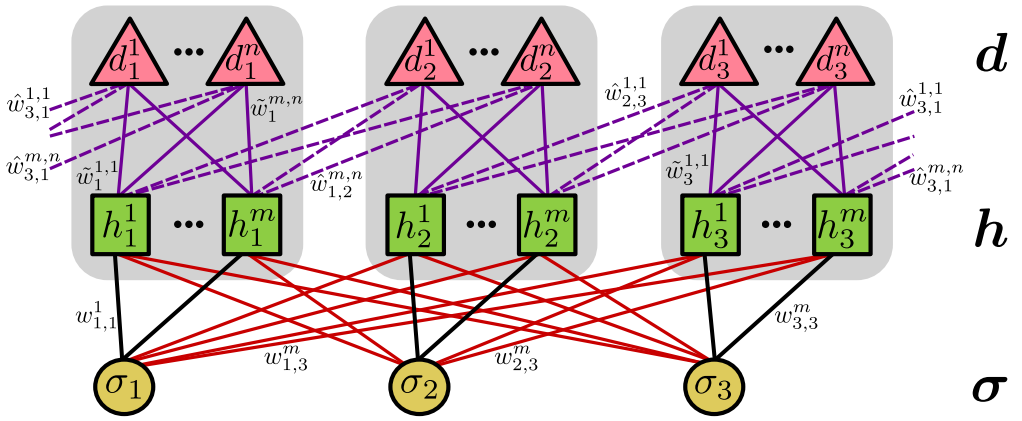

Figure 2: (color online) GTMS network for $N = 3$ physical sites. Each transfer matrix is denoted with a grey box. The red links $w^{\mu}_{i,j\neq i}$ are the RBM weights, which result in the dependence of the transfer matrices on all the physical spins. By erasing the RBM links in the architecture one obtains a MPS with bond dimension 2^n. The purple links are the MPS weights. Bias terms as well as the direct links between sets of deep variables in neighboring blocks are not shown.

called network energy, which in the present case reads as

$$\mathcal{E}_{\rm nw}(\boldsymbol{\sigma},\boldsymbol{h},\boldsymbol{d};\boldsymbol{w}) = -\sum_{i=1}^{N} c_i\sigma_i + \sum_{k=1}^{D} a_k d_k + \sum_{j=1}^{M}\big(b_j + \sum_{i=1}^{N}\sigma_i w_{i,j} + \sum_{k=1}^{D}\tilde{w}_{j,k}d_k\big)h_j \ . \tag{1}$$

The network configurations are then assigned generalized complex Boltzmann weights $\mathrm{e}^{-\mathcal{E}_{\rm nw}(\boldsymbol{\sigma},\boldsymbol{h},\boldsymbol{d};\boldsymbol{w})}$, and the variational wavefunction $\psi_{\boldsymbol{w}}(\boldsymbol{\sigma})$ is obtained after a partial partition sum, i.e. by tracing over the hidden and deep layer configurations:

$$\psi_{\boldsymbol{w}}(\boldsymbol{\sigma}) = \sum_{\{\boldsymbol{d}\}}\sum_{\{\boldsymbol{h}\}} \mathrm{e}^{-\mathcal{E}_{\rm nw}(\boldsymbol{\sigma},\boldsymbol{h},\boldsymbol{d};\boldsymbol{w})} \ . \tag{2}$$

As a simpler limiting case obtained by discarding the deep layer, let us briefly recall the notion of restricted Boltzmann machine (RBM) states [31]. In a RBM network the auxiliary spins consist of only one hidden layer of classical Ising spins $h_j = \pm 1$ and there are no couplings within the set of hidden units. The network energy (see (1)) for a RBM simplifies to $\mathcal{E}_{\rm nw}(\boldsymbol{\sigma},\boldsymbol{h};\boldsymbol{w}) = -\sum_i c_i\sigma_i + \sum_j(b_j + \sum_i \sigma_i w_{i,j})h_j$. To obtain $\psi_{\boldsymbol{w}}(\boldsymbol{\sigma})$, for the RBM we only need to sum over all $\boldsymbol{h}$ configurations, which yields the RBM state $\psi_{\boldsymbol{w}}(\boldsymbol{\sigma}) = \mathrm{e}^{\sum_i c_i\sigma_i}\prod_{j=1}^{M} 2\cosh\left(b_j + \sum_i \sigma_i w_{i,j}\right)$ [31].

The power of going from RBM to DBM networks lies in the universal representational capabilities [30] of the latter, which has been demonstrated in a quantum physics context by showing that a three-layer DBM is capable of exactly representing the (imaginary) time evolution of generic quantum many-body systems [39]. Concretely, Ref. [39] proves that a suitable DBM network of polynomial complexity in system size and imaginary time β with weigths $\boldsymbol{w}_\beta$ can exactly represent the imaginary time evolution of an initial quantum state $|\psi_0\rangle$ with respect to a generic many-body Hamiltonian H, i.e.

$$\psi_{\boldsymbol{w}_\beta}(\boldsymbol{\sigma}) = \langle\boldsymbol{\sigma}|\mathrm{e}^{-\beta H}|\psi_0\rangle \ . \tag{3}$$

However, the major caveat limiting the immediate applicability of this strong result is that it is practically impossible in general to exactly evaluate the wavefunction amplitude $\psi_{\boldsymbol{w}_\beta}(\boldsymbol{\sigma})$ by performing the sum on the right-hand side of Eq. (2). This is in stark contrast to the simpler RBM, where the wavefunction is readily be calculated analytically. In more physical terms, Boltzmann machine state amplitudes $\psi_{\boldsymbol{w}}(\boldsymbol{\sigma})$ resemble an effective action for the physical spins obtained by tracing out a bath of hidden spin variables. Within this analogy, for the RBM architecture the hidden layer amounts to a free spin-system, while for a DBM the auxiliary variables represent an interacting spin system which is hard to solve in general. In the remainder of this article we will identify and study a class of DBM architectures, coined generalized transfer matrix state (GTMS) networks, that can still be exactly contracted, and leads to a unifying generalization of MPS and RBM states.

III. GENERALIZED TRANSFER MATRIX STATES

We now define the central entity of this work, namely the generalized transfer matrix state (GTMS) network, as a particular exactly contractible, deep Boltzmann machine. Our construction is inspired by the aforementioned interpretation of the DBM wavefunction as an effective action obtained by tracing out an interacting spin system representing the set of auxiliary units. This raises

the natural question what kind of wavefunctions are obtained when constraining the couplings so as to make this auxiliary spin-system exactly solvable, which leads us to a substantially larger class of networks than the previously considered RBM architecture (corresponding to a non-interacting auxiliary spin system). Specifically, we will group the auxiliary spins into blocks and limit the connectivity between hidden and deep layers to a nearest-neighbor connectivity between these blocks (see grey boxes in Fig. 1b), while retaining all-to-all connectivity between the physical and the hidden layer. Once this constraint has been implemented, the sum over the auxiliary variables can be evaluated adopting a transfer matrix method, well known from the solution of the 1D Ising model.

A detailed exemplary visualization of this GTMS network architecture is shown in Fig. 2. The hidden and deep auxiliary spins are grouped into N_T blocks (the grey shaded areas in Figs. 1b and 2), containing n deep and m hidden spins per block. Within these blocks, the connectivity between hidden and deep variables is all-to-all, but to make the network contractible the couplings between different blocks are limited to nearest neighbors (the purple dashed lines in the Fig. 2). We impose periodic boundary conditions (PBC) to the network, i.e. the last and the first blocks of auxiliary spins are also coupled. In general the number of blocks N_T can be different from the number of physical sites N. Also, arbitrary (i.e. from 2-body to n-body) direct couplings between deep variables in the same block and in neighboring blocks, as well as direct all-to-all couplings between deep and physical layers, can be introduced, still keeping the network contractible (these connections are not shown in Fig. 2 - see the Appendix for more detailed discussion). Finally, we point out that the number m and n of hidden and deep spins per block can in general depend on the block index itself, i.e. $m = m_j$ and $n = n_j$ with $j = 1, ..., N_T$.

Next, we explicitly contract the GTMS network illustrated in Fig. 2, so as to derive an analytical form of the GTMS amplitude. For a straightforward extension to the aforementioned slightly more general connectivity we refer to the Appendix. The network energy of the GTMS network reads as

$$\begin{aligned}\mathcal{E}_{\rm nw}(\boldsymbol{\sigma},\boldsymbol{h},\boldsymbol{d};\boldsymbol{w}) = -\sum_{i=1}^{N} c_i\sigma_i + \sum_{j=1}^{N_T}\Big\{ \sum_{\nu=1}^{n} a_j^\nu d_j^\nu + \\ \sum_{\mu=1}^{m}\Big[b_j^\mu + \sum_{i=1}^{N}\sigma_i w_{i,j}^\mu + \\ \sum_{\nu=1}^{n}\left(\tilde{w}_j^{\mu,\nu} d_j^\nu + \hat{w}_{j,j+1}^{\mu,\nu} d_{j+1}^\nu\right)\Big] h_j^\mu\Big\}\ .\end{aligned} \tag{4}$$

Here the set of weights $\boldsymbol{w}$ contains: c_i, b_j^μ, a_j^ν which are the complex on-site bias weights for σ_i, h_j^μ, d_j^ν respectively (not explicitly shown in Fig. 2), $w_{i,j}^\mu$ which denote the couplings between physical σ_i and hidden h_j^μ (red and purple links between physical and hidden layers in Fig. 2), $\tilde{w}_j^{\mu,\nu}$ that couple h_j^μ and d_j^ν within the same block, and $\hat{w}_{j,j+1}^{\mu,\nu}$ that couple h_j^μ and d_{j+1}^ν in neighboring blocks (the dashed purple links in Fig. 2). We refer to the weights $w_{i,j}^\mu$ with $i \neq j$ as RBM weigths (red links in Fig. 2), while the rest of the links, except for the $w_{i,i}^\mu$ (the black links in Fig. 2) are referred to as MPS weights. This nomenclature is motivated by the fact that if the network is restricted to contain only MPS weights together with the $w_{i,i}^\mu$'s (i.e. if one sets to 0 the RBM weights), the state obtained after its contraction can be recast as an MPS. This will be explained in more detail in Section IV. If we keep instead only the RBM weights with the $w_{i,i}^\mu$'s (that is, we set to 0 the MPS weights) the dependence of the network energy (Eq. (4)) on the deep spins would disappear yielding eventually the network energy of a RBM, and therefore a RBM wavefunction after the contraction of the network.

To perform the sum over hidden and deep variables configurations of Eq. (2) we first trace out the hidden layer. Since the hidden spins are not directly coupled, this sum is easily performed (analogous to the RBM case), and yields:

$$\psi_{\boldsymbol{w}}(\boldsymbol{\sigma}) = \mathrm{e}^{\sum_i c_i\sigma_i}\sum_{\{\boldsymbol{d}\}}\prod_j t_{j,j+1}(\boldsymbol{\sigma},\boldsymbol{d}_j,\boldsymbol{d}_{j+1}) \tag{5}$$

where $\boldsymbol{d}_j$ denotes the deep spin configuration at block j, i.e. $\boldsymbol{d}_j = \{d_j^\nu\}_{\nu=1,...,n}$. The elements $t_{j,j+1}$ of the product read as

$$t_{j,j+1}(\boldsymbol{\sigma},\boldsymbol{d}_j,\boldsymbol{d}_{j+1}) = \mathrm{e}^{\sum_\nu a_j^\nu d_j^\nu}\prod_{\mu=1}^{m} 2\cosh\left(\varphi_j^\mu(\boldsymbol{\sigma},\boldsymbol{d}_j,\boldsymbol{d}_{j+1})\right) \tag{6}$$

with

$$\begin{aligned}\varphi_j^\mu(\boldsymbol{\sigma},\boldsymbol{d}_j,\boldsymbol{d}_{j+1}) = b_j^\mu + \sum_{i=1}^{N}\sigma_i w_{i,j}^\mu + \\ \sum_{\nu=1}^{n}\left(\tilde{w}_j^{\mu,\nu} d_j^\nu + \hat{w}_{j,j+1}^{\mu,\nu} d_{j+1}^\nu\right)\ .\end{aligned} \tag{7}$$

Now, in order to perform the sum over all the configurations of the deep variables, we can interpret the complex numbers $t_{j,j+1}(\boldsymbol{\sigma},\boldsymbol{d}_j,\boldsymbol{d}_{j+1})$ as elements of a *transfer matrix* $T_j(\boldsymbol{\sigma})$ associated to block j. One can uniquely associate a index, running from 1 to 2^n, to a deep spin configuration $\boldsymbol{d}_j$ at block j, by interpreting the n Ising spins d_j^ν in $\boldsymbol{d}_j$ as bits. We can define the elements of the transfer matrix as

$$\Big(T_j(\boldsymbol{\sigma})\Big)_{\boldsymbol{d}_j,\boldsymbol{d}_{j+1}} = t_{j,j+1}(\boldsymbol{\sigma},\boldsymbol{d}_j,\boldsymbol{d}_{j+1})\ . \tag{8}$$

Tracing out the deep layer in Eq. (5) is then equivalent

to taking the product of these transfer matrices, i.e.

$$\psi_{\boldsymbol{w}}(\boldsymbol{\sigma}) = \mathrm{e}^{\sum_i c_i \sigma_i} \,\mathrm{tr}\Big(\prod_{j=1}^{N_T} T_j(\boldsymbol{\sigma})\Big), \tag{9}$$

where the trace comes from the periodic boundary conditions imposed on the blocks of auxiliary spins of the network. We stress that the GTMS network does not require the introduction of PBC for being exactly contractible, and can equally well be used for physical systems with open boundary conditions (see the Appendix for a more detailed discussion).

In case $m = m_j$ and $n = n_j$ then the dimension of T_j would depend on j as well, being equal to $2^{n_j} \times 2^{n_{j+1}}$. The transfer matrices depend in general on the index j as well as on the physical spin configuration $\boldsymbol{\sigma}$ over the entire system, as opposed to the well known case of MPS where each matrix depends locally on the spin quantum number on one physical site. For this reason we call the state of Eq. (9) a generalized transfer matrix state (GTMS), and we will show below that this non-local dependence on the physical quantum numbers allows the GTMS to capture long-range correlations going beyond the area-law typical of MPS.

IV. MATRIX PRODUCT STATES FROM GTMS

In this Section we demontrate how, by removing the RBM weights, the GTMS network is able to parametrize generic MPS, with a bond dimension 2^n set by the number n of deep spins per physical site. Defined as a product of tensors with elements $A^{[i]\sigma_i}_{a_{i-1}a_i}$ associated to each physical site i with one physical index σ_i and two auxiliary indices $a_1, a_2 = 1, ..., \chi$ with the bond dimension χ, a generic MPS is of the form [3, 5, 12]:

$$\psi_{\mathrm{MPS}}(\boldsymbol{\sigma}) = \mathrm{tr}\Big(\prod_{i=1}^{N} A^{[i]\sigma_i}\Big) \tag{10}$$

We can immediately see that this form is similar to the one of Eq. (9) with the number of transfer matrices N_T equal to the number of physical sites N, apart form the fact that here the matrices in the product depend only on the quantum number σ_i of the physical site i. To reduce Eq. (9) to the MPS form of Eq. (10) we simply restrict the connectivity of the GTMS network so as to make the T_i depend only on σ_i. We note that in Eq. (9) the dependence of T_j on the entire physical layer enters via the angles φ^μ_j of Eq. (7), where the term $\sum_i \sigma_i w^\mu_{i,j}$ appears. Therefore, if we set to 0 all the couplings $w^\mu_{i,j}$ where $i \neq j$ each φ^μ_j depends only on σ_j. Pictorially this amounts to erasing all the red links in Fig. 2. In physical terms, this corresponds to neglecting the long-range quantum correlations which are mediated by the RBM couplings, keeping only the short-range correlations encoded in the MPS couplings between neighboring transfer matrices. This way, Eq. (9) becomes formally analogous to the MPS in Eq. (10), with bond dimension $\chi = 2^n$ (assuming n constant throughout the system and $N_T = N$):

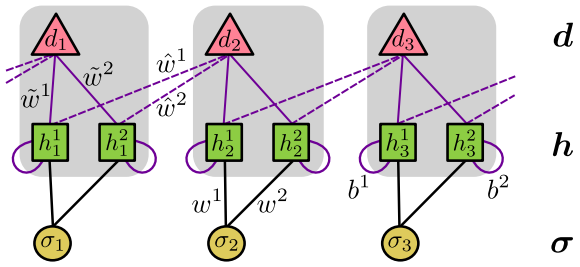

Figure 3: (color online) GTMS network for the construction of the AKLT state, for $N = 3$ physical sites, with the bias terms for hidden units explicitly shown. Since the AKLT state is translation invariant we have that $\forall\, i = 1, ..., N$ $w^\mu_i = w^\mu$, $\tilde{w}^\mu_i = \tilde{w}^\mu$, $\hat{w}^\mu_{i,i+1} = \hat{w}^\mu$, $b^\mu_i = b^\mu$, with $\mu = 1, 2$.

$$\psi_{\boldsymbol{w}}(\boldsymbol{\sigma}) = \mathrm{tr}\Big(\prod_{i=1}^{N} \mathrm{e}^{c_i\sigma_i} T_i(\sigma_i)\Big) \tag{11}$$

where $\mathrm{e}^{c_i\sigma_i} T_i(\sigma_i)$ can be identified with the tensor $A^{[i]\sigma_i}$ in Eq. (10) and the notation $T_i(\sigma_i) \equiv T_i(\boldsymbol{\sigma})$ has been introduced to make the local dependence of the transfer matrices on the physical spins manifest. We note that for being able to parametrize an arbitrary MPS with bond dimension 2^n and spin s physical degrees of freedom, one would in general need $2^{2n}(2s+1)$ complex free parameters per matrix. Simple parameter counting shows that, in order to have enough free parameters, the number m of hidden spins per block should scale with n according to $m \approx \frac{(2s+1)2^{2n}}{2(n+1)}$.

A. AKLT State from GTMS

As an emblematic example we explicitly construct the AKLT state [44, 45] from a GTMS network, shown in Fig. 3. The AKLT state is one of the simplest MPS with bond dimension 2 and tensors independent of position. This suggests that the use of a GTMS architecture with constant $n = 1$, and $m = 2$ hidden variables per site will be sufficient for fully parametrizing the state. The AKLT state is the ground state of a modified spin-1 quantum Heisenberg model [44], hence the physical spin variables which constitute the inputs of our network can take values $\sigma_i = -1, 0, +1$ (while in the remainder of the paper we will use spin 1/2 degrees of freedom). The normalized

AKLT matrices read

$$A^+ = \frac{2}{\sqrt{3}}\begin{pmatrix} 0 & \frac{1}{\sqrt{2}} \\ 0 & 0 \end{pmatrix} \qquad A^- = \frac{2}{\sqrt{3}}\begin{pmatrix} 0 & 0 \\ -\frac{1}{\sqrt{2}} & 0 \end{pmatrix}$$
$$A^0 = \frac{2}{\sqrt{3}}\begin{pmatrix} -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} . \tag{12}$$

For the ANN architecture shown in Fig. 3 we obtain the 2×2 transfer matrices

$$\big(T(\sigma_i)\big)_{d_i,d_{i+1}} = \prod_{\mu=1}^{2} \cosh\left(\varphi^\mu(\sigma_i, d_i, d_{i+1})\right) \tag{13}$$

with the angular arguments

$$\varphi^\mu(\sigma_i, d_i, d_{i+1}) = b^\mu + \sigma_i w^\mu + \tilde{w}^\mu d_i + \hat{w}^\mu d_{i+1} \ , \tag{14}$$

where the independence of several quantities on the physical site index i reflects the translation invariance of the AKLT state. By choosing $w^1 = i\pi/4$, $w^2 = i\pi/2$, $\tilde{w}^1 = -i\pi/4$, $\tilde{w}^2 = i3\pi/4$, $\hat{w}^1 = 0$, $\hat{w}^2 = i3\pi/4$, $b^1 = 0$ and $b^2 = -i\pi/2$, we find $\sqrt{2/3}\,T(\sigma_i) = A^{\sigma_i}$, where the normalization factor $\sqrt{2/3}$ can formally be added to the DBM network by a constant shift to the network energy.

This explicit parameterization demonstrates how the AKLT state is exactly represented by a *short range* DBM network, in which the connectivity is limited to neighboring blocks of auxiliary units. Interestingly, this simple state *cannot* be directly represented by a short-range RBM network. Indeed, as noticed in [41] short-range RBM states correspond to so called entangled-plaquette states, which are products of complex numbers associated to local clusters (plaquettes) of physical sites. Physically, this product structure of commuting local factors makes it impossible for such states to encode the hidden infinitely-ranged string order of the AKLT state [46]. In more practical terms, the string order constrains the AKLT wavefunction to vanish whenever two subsequent A^+ or two subsequent A^- matrices at the physical sites i and j are separated only by A^0 matrices, no matter how large the distance between i and j is. This is encoded in the basic (non-commuting) algebra of these matrices. Clearly, such a constraint cannot be achieved by a product of complex numbers that depend only locally on the physical variables, as in the case of a short-range RBM.

B. Learning Random MPS

Generalizing from the basic example of the AKLT state, we now numerically show that GTMS networks without RBM weights (red links in Figs. 1 and 2) can *learn* a generic MPS. In the context of artificial neural networks, the word *learning* means that the weights of the network are iteratively optimized to find the minimum of a certain cost function [27–29, 31–36]. Assuming translation invariance, we optimize only for one MPS tensor with a network (corresponding to a single grey shaded block in Fig. 2), containing n deep variables per physical site, thus yielding a bond dimension $\chi = 2^n$. The resulting network representation of the MPS tensor elements is visualized in Fig. 4. By fixing σ and the deep spin configurations $\boldsymbol{d}_1 = \{d_1^1, ..., d_1^n\}$ and $\boldsymbol{d}_2 = \{d_2^1, ..., d_2^n\}$ one obtains, after tracing out the hidden layer in the network of Fig. 4, the $(\boldsymbol{d}_1, \boldsymbol{d}_2)$ element of the transfer matrix $T(\sigma_i)$.

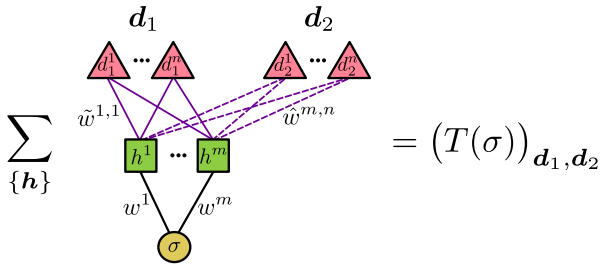

Figure 4: (color online) Block of GTMS network representing the elements $\big(T(\sigma)\big)_{\boldsymbol{d}_1,\boldsymbol{d}_2}$ of a MPS tensor with bond dimension $\chi = 2^n$ and physical index σ. A single matrix element is obtainend by fixing the deep spin configurations $\boldsymbol{d}_1, \boldsymbol{d}_2$ and σ, and then computing the trace over the hidden spin configurations $\boldsymbol{h} = \{h_1, ..., h_m\}$.

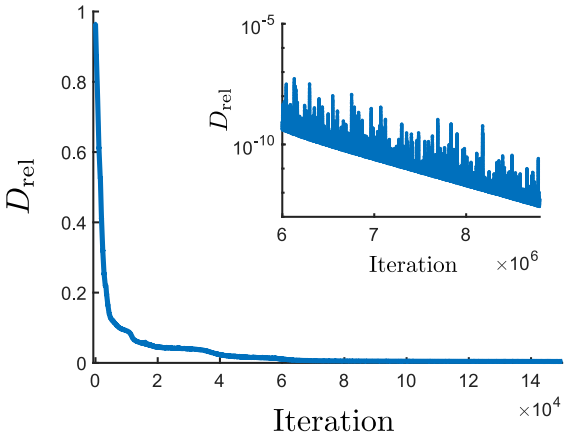

Figure 5: (color online) Convergence plot for the learning of a random MPS tensor with bond dimension $\chi = 8$ and local Hilbert space dimension 2, with a network having $n = 3$, $m = 16$. Here the optimization was done using AdaGrad method. The inset shows a closeup of the cost function at later iterations, plotted in logarithmic scale.

The cost function $D_{\rm rel}$ that we optimize for is defined using the Frobenius norm of the difference between the GTMS transfer matrix T and the random MPS tensor A to be learned:

$$D_{\rm rel} \equiv \|T - A\|^2_{\rm rel} = \frac{\|T - A\|^2}{\|A\|^2}, \tag{15}$$

where the dependence on the variational paramenters $\boldsymbol{w}$ lies in the transfer matrix T (see Eqs. (6-8)).

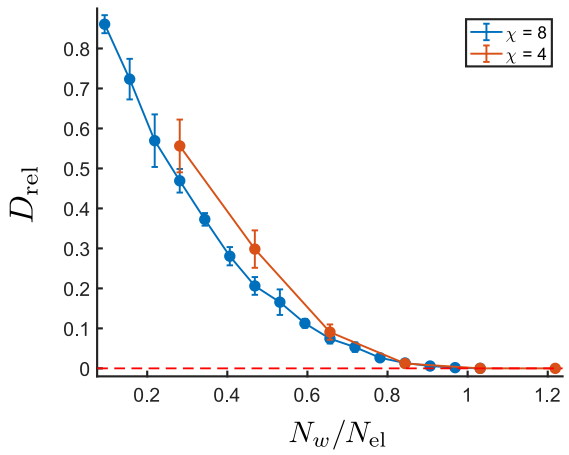

Figure 6: (color online) Attempted MPS compression for bond dimension $\chi = 4$ (orange dots) and $\chi = 8$ (blue dots), performed by removing hidden spins from the network to be optimized. The converged value of $D_{\rm rel}$, averaged over 10 random MPS tensor realizations, is plotted vs. the ratio $N_w/N_{\rm el}$. The dashed red line indicates the position of the 0, for guiding the eye.

We considered the case of spin-1/2 degrees of freedom per physical site, so the total number of elements of the MPS tensor is $N_{\rm el} = 2\chi^2$. The optimization has been performed using stochastic gradient descent methods [47, 48] such as AdaGrad and Adam [49, 50] with $N_w = 1 + 2m + n + 2mn \simeq 2\chi^2$ parameters. An example of a convergence plot for $\chi = 8$ using AdaGrad for optimizing the network with $n = 3$ and $m = 16$ is given in Fig. 5. By using Adam optimizer implemented in the Phyton TensorFlow libraries we were able to learn random MPS tensors up to bond dimension $\chi = 16$ to a final relative accuracy $D_{\rm rel} \sim 10^{-4}$, using $n = 4$ and $m = 52$ on an ordinary desktop computer.

At last we would like to address the question whether it is possible to learn a MPS tensor with a number N_w of variational parameters that is lower than the number of tensor elements $N_{\rm el}$, thus attempting an approximate compression of the MPS. The results of this compression are shown in Fig. 6 relative to a set of 10 realizations of random MPS tensors with bond dimension $\chi = 4$ (orange data) and $\chi = 8$ (blue data). We can see that for $N_w \geq N_{\rm el}$ our network can indeed be optimized to learn the MPS tensors (each data point in Fig. 6 represents the best achieved relative accuracy $\langle D_{\rm rel}\rangle$ averaged over the 10 random MPS tensor realizations). As $N_w < N_{\rm el}$ we find however $\langle D_{\rm rel}\rangle > 0$, meaning that an efficient network compression in general is not viable when learning generic (random) MPS tensors.

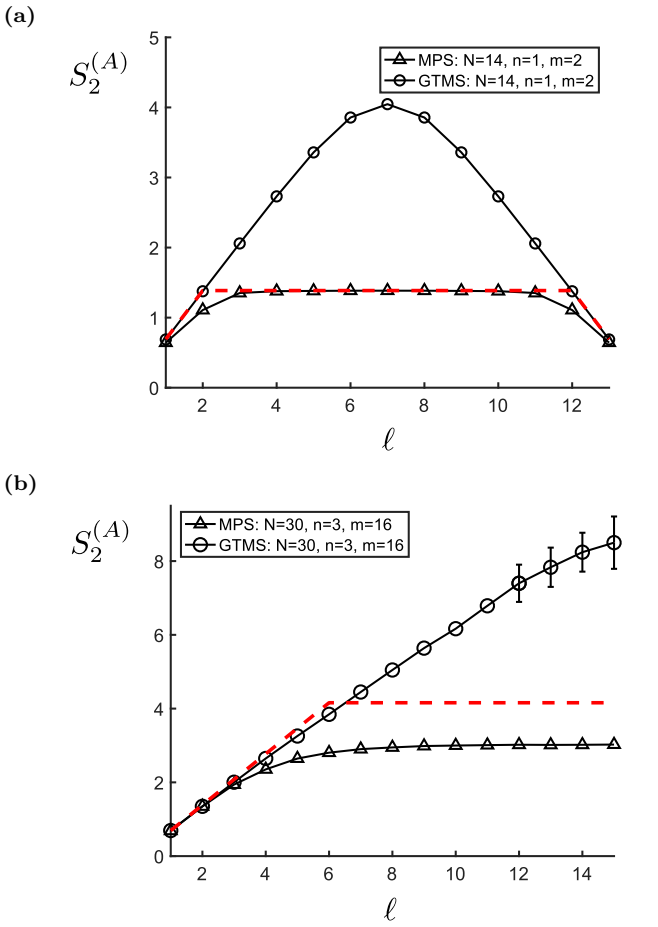

Figure 7: (color online) $S_2^{(A)}$ as function of the length ℓ of subsystem A. Black triangles show $S_2^{(A)}$ for a translation invariant MPS obtained from a GTMS network with randomly chosen weights. Black circles are for GTMS obtained by adding random non-local RBM couplings to the network parametrizing the MPS. The dashed red line shows the upper bound for the entanglement entropy of a MPS with $\chi = 2^n$. Weights are drawn within uniform distribution around 0 with width 0.2 and 8π for real and imaginary parts, respectively. **(a)** Exact results for $n = 1$, $m = 2$ and $N = 14$ sites. **(b)** Monte Carlo results for $n = 3$, $m = 16$ with $N = 30$ sites. The statistical error is smaller than the data symbol, when the errorbar is not shown. For translation invariance, we show $S_2^{(A)}$ only up to half of system size.

V. ENTANGLEMENT ANALYSIS OF GTMS

In this Section, we present a numerical analysis of the entanglement entropy of GTMS for one dimensional systems with PBC and spin 1/2 degrees of freedom per site. We will show that the addition of non-local RBM weights to a GTMS network representing a MPS results in the onset of volume-law entanglement, as opposed to the area-law scaling obtained when keeping the MPS weights only. This is a clear indication of the improved representational power of generalized transfer matrix states.

Specifically, we calculate the second Rényi entropy $S_2^{(A)}$ for a bipartition of a one dimensional spin-1/2 system with PBC in two subsystems A and B, with the total system being in the pure state $|\psi_{\boldsymbol{w}}\rangle$ with GTMS wavefunctions of Eq. (9). The second Rényi entropy of subsystem A is given by

$$S_2^{(A)} = -\log\left(\mathrm{tr}\,\rho_A^2\right) \tag{16}$$

with $\rho_A = \mathrm{tr}_B\,|\psi_{\boldsymbol{w}}\rangle\langle\psi_{\boldsymbol{w}}|$ being the reduced density matrix of subsystem A (the natural logarithm is used). The algorithm introduced in [51] offers a simple and efficient way for calculating $S_2^{(A)}$ with Monte Carlo, which requires Metropolis sampling of two copies of the system, as the trace of ρ_A^2 needs to be evaluated. For the α^{th} Rényi entropy one would need to sample configurations of α copies (see [51] and the Appendix for more details).

We determine the scaling of the second Rényi entropy with the length ℓ of subsystem A, comparing the two cases of a GTMS network parametrizing a conventional MPS, and the augmented GTMS to which nonlocal RBM couplings have been added while keeping the existing couplings unchanged. In Fig. 7, we show exact data on $S_2^{(A)}$ for a system of $N = 14$ sites in the case of a GTMS with $n = 1$, $m = 2$ (panel 7a), and Monte Carlo data from Metropolis sampling for a system of $N = 30$ sites in the case of a GTMS with $n = 3$, $m = 16$ (panel 7b). We observe that the addition of RBM couplings results in a volume-law (i.e. linear in ℓ) scaling of the entanglement, and that $S_2^{(A)}$ exceeds the MPS bound $2\log\chi$ (the dashed red line in Fig. 7) which is set by the bond dimension $\chi = 2^n$. In this sense, the GTMS familiy combines the properties of conventional MPS and RBM states.

VI. CONCLUDING DISCUSSION

In summary, we proposed a deep ANN architecture that is exactly contractible and yields a class of quantum states called GTMS. The GTMS family is shown by means of a constructive mapping to include both generic MPS and RBM states, and allows to continuously interpolate between them. More specifically, GTMS networks are a family of deep Boltzmann machine networks which are exactly and efficiently contractible by means of a transfer matrix method. Our findings are corroborated by numerical data showing that the GTMS network is indeed able of efficiently parametrizing random MPS, where efficiently means that the number of variational weights scales as the number of independent parameters of the MPS. Moreover we show with a numerical analysis of the second Rényi entropy, that GTMS initially parametrizing a MPS (therefore a state with area-law entanglement) can, upon addition of RBM weights, encode long-range correlations with volume-law entanglement. On a general note, representation theorems [30] tell us that the proposed augmentation of an existing network by additional couplings can only improve the capabilities of the network in representing quantum states. In our present construction, the onset of the volume-law scaling provides a concrete intuition for this increased representational power compared to conventional MPS.

The potential of RBM of representing states with volume-law entanglement and encoding up to N-body correlations was already discussed [32], and the general correspondence between RBM states and tensor network states is well known [40–43]. However, an efficient mapping between MPS and RBM states has remained elusive, since it is unclear how the required number of RBM couplings scales with the bond dimension. Here, we have used the higher representational power of DBMs to efficiently and constructively embed MPS into the general framework of ANN states. By efficiently, we mean that the number of DBM couplings needed scales as the number of free parameters in the MPS tensors. In this sense, the GTMS combines key representational properties of MPS or RBM states and has stronger representational power than either of the two alone.

The great advantage of GTMS is that both short-ranged MPS correlations and long-ranged entanglement can be efficiently captured. This makes GTMS a promising *ansatz* for problems where the entanglement growth poses severe limitations to tensor network studies. Such cases include the study of critical systems or of time evolution in quantum many-body systems far from equilibrium, where the MPS *ansatz* would require us to increase the bond dimension with system size and time, respectively [12–20]. However, it is fair to say that the generalization from MPS to GTMS comes at a price: While observables can be efficiently represented directly in the space of MPS [3, 5], for most variational wavefunctions including RBM states and also the proposed GTMS the understanding of the corresponding variational space is far less complete. Therefore, evaluating expectation values of physical observables and optimizing the variational parameters so far requires stochastic methods. Developing powerful optimization methods in the framework of ANN states that draw intuition from tensor network methods such as DMRG, TEBD, and TRG is an interesting direction of future research.

VII. ACKNOWLEDGEMENTS

We acknowledge discussions with E. Bergholtz and H.-H. Tu. LP and JCB acknowledge financial support from the German Research Foundation (DFG) through the Collaborative Research Centre SFB 1143. RK is supported by the Austrian Science Fund SFB FoQuS (FWF Project No. F4016-N23) and the European Research Council (ERC) Synergy Grant UQUAM. The numeri-

cal calculations were performed on resources at the TU Dresden Center for Information Services and High Performance Computing (ZIH), and at the Chalmers Centre for Computational Science and Engineering (C3SE) provided by the Swedish National Infrastructure for Computing (SNIC).

APPENDIX

AI. GENERAL EXACTLY CONTRACTABLE DBM NETWORK

The network shown in Fig. 2 is not the most general exactly contractible architecture. As discussed in Section III, one can introduce arbitrary (from 2-body to n-body) couplings between deep variables in the same block and in neighboring blocks by still keeping the network contractible. In this Appendix we want to elaborate more on this structure, explaining how the presence of an hidden layer is fundamental for having enough variational parameters to parametrize generic MPS.

The network energy for a GTMS where arbitrary couplings between sets of deep variables within the same block and between neighboring blocks have been introduced, is obtained simply by adding a term

$$-\sum_{j=1}^{N_T}\left\{C_j^{(\mathrm{o.s.})}(\boldsymbol{d}_j)+C_j^{(\mathrm{n.n.})}(\boldsymbol{d}_j,\boldsymbol{d}_{j+1})\right\} \tag{17}$$

to the espression of $\mathcal{E}_{\mathrm{nw}}$ in Eq. (4), where $C_j^{(\mathrm{o.s.})}(\boldsymbol{d}_j)$ denotes the sum of all possible direct couplings between the deep spins contained in block j:

$$\begin{aligned} C_j^{(\mathrm{o.s.})}(\boldsymbol{d}_j) = &\sum_{\nu_1\neq\nu_2}\tilde{\omega}_j^{\nu_1,\nu_2}d_j^{\nu_1}d_j^{\nu_2}+ \\ &\sum_{\nu_1\neq\nu_2\neq\nu_3}\tilde{\omega}_j^{\nu_1,\nu_2,\nu_3}d_j^{\nu_1}d_j^{\nu_2}d_j^{\nu_3}+\ldots \end{aligned} \tag{18}$$

and $C_j^{(\mathrm{n.n.})}(\boldsymbol{d}_j,\boldsymbol{d}_{j+1})$ denotes the sum of all possible direct couplings between the deep spins in neighboring blocks j and $j+1$:

$$\begin{aligned} C_j^{(\mathrm{n.n.})}(\boldsymbol{d}_j,\boldsymbol{d}_{j+1}) = &\sum_{\nu_1,\nu_2}\hat{\omega}_{j,j+1}^{(\nu_1)(\nu_2)}d_j^{\nu_1}d_{j+1}^{\nu_2}+ \\ &\sum_{\nu_1,\nu_2\neq\nu_3}\hat{\omega}_{j,j+1}^{(\nu_1)(\nu_2,\nu_3)}d_j^{\nu_1}d_{j+1}^{\nu_2}d_{j+1}^{\nu_3}+ \\ &\sum_{\nu_1\neq\nu_2,\nu_3}\hat{\omega}_{j,j+1}^{(\nu_1,\nu_2)(\nu_3)}d_j^{\nu_1}d_j^{\nu_2}d_{j+1}^{\nu_3}+\ldots\,. \end{aligned} \tag{19}$$

The addition of these couplings does maintain the network exactly contractible. The only modification to the wavefunction amplitude is that the product elements $t_{j,j+1}$ of Eq. (6) for block j are now multiplied by the corresponding factors $\mathrm{e}^{C_j^{(\mathrm{o.s.})}+C_j^{(\mathrm{n.n.})}}$:

$$t_{j,j+1}(\boldsymbol{\sigma},\boldsymbol{d}_j,\boldsymbol{d}_{j+1})\to \mathrm{e}^{C_j^{(\mathrm{o.s.})}+C_j^{(\mathrm{n.n.})}}t_{j,j+1}(\boldsymbol{\sigma},\boldsymbol{d}_j,\boldsymbol{d}_{j+1})\,.$$

One may now ask the question whether, with the addition of such arbitrary links between deep variables, the number of variational parameters is sufficient to parametrize arbitrary MPS *without* the addition of the hidden layer. It turns out that this is not the case. With n deep spin variables per block, the number $\#\omega|_j$ of direct links, and therefore of complex weights $\tilde{\omega}_j^{\{\nu_i\}}$ and $\hat{\omega}_{j,j+1}^{(\{\nu_i\})(\{\nu_k\})}$ in the sum $C_j^{(\mathrm{o.s.})}+C_j^{(\mathrm{n.n.})}$ between blocks j and $j+1$ is

$$\#\omega|_j=\sum_{k=2}^{2n}\binom{2n}{k}-\sum_{k=2}^{n}\binom{n}{k}=2^{2n}-2^n-n-1\,.$$

Therefore, with the addition of the bias terms for the n deep spin variables, the total number of complex weights per transfer matrix in absence of hidden spin variables, would be $2^{2n}-2^n+n-1$, insufficient to parametrize an arbitrary MPS tensor, for which in general $2^{2n}(2s+1)$ would be required ($2s+1$ being the local Hilbert space dimension). If we require the network to be able to parametrize generic MPS, we therefore must add additional hidden spins to be traced out before the sum over the deep variables configurations is computed, in order to have enough variational parameters.

AII. OPEN BOUNDARY CONDITIONS

Here we briefly mention how the GTMS can be contracted without the use of PBC, thereby allowing for the parametrization of MPS with open boundary conditions and their generalization to a GTMS that interpolates between them and RBM states.

Considering the network studied in Section III, described by the network energy of Eq. (4), we set open boundary conditions on it simply by erasing (i.e. setting to 0) the $\hat{w}$ couplings extending from tensor $j=N_T$ to $j=1$. By applying the transfer matrix method to the contraction of a GTMS network with open boundaries we find an exact espression for the wavefunction, which reads

$$\psi_{\boldsymbol{w}}(\boldsymbol{\sigma})=\mathrm{e}^{\sum_i c_i\sigma_i}\bar{V}_1\Big(\prod_{j=2}^{N_T-1}T_j\Big)V_{N_T}\,. \tag{20}$$

The transfer matrices T_j in the bulk of this open boundary GTMS are the same as the ones defined in Eqs. (6-8). The right-boundary tensor ($j=N_T$) is a $2^n\times 1$ column vector with elements

$$\big(V_{N_T}\big)_{\boldsymbol{d}_{N_T}}=\mathrm{e}^{\sum_\nu a_{N_T}^\nu d_{N_T}^\nu}\prod_{\mu=1}^{m}2\cosh\big(\tilde{\varphi}_{N_T}^\mu(\boldsymbol{\sigma},\boldsymbol{d}_{N_T})\big) \tag{21}$$

where

$$\tilde{\varphi}_{N_T}^\mu(\boldsymbol{\sigma},\boldsymbol{d}_{N_T})=b_{N_T}^\mu+\sum_{i=1}^{N}\sigma_i w_{i,N_T}^\mu+\sum_{\nu=1}^{n}\tilde{w}_{N_T}^{\mu,\nu}d_{N_T}^\nu\,. \tag{22}$$

The left-boundary tensor ($j = 1$) is given by

$$\bar{V}_1 = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} T_1 \ . \tag{23}$$

that is a 1×2^n row vector, where T_1 is the transfer matrix matrix with elements $t_{1,2}(\boldsymbol{\sigma}, \boldsymbol{d}_1, \boldsymbol{d}_2)$.

AIII. EFFICIENT CALCULATION OF SECOND RÉNYI ENTROPY

Here we review the algorithm for the computation of the second Rényi entropy applicable to Monte Carlo calculations, introduced in [51]. Consider a system S in a quantum state $|\psi\rangle$, and a bipartition of it into two subsystems A and B. The Rényi entropy of A subsystem is given by

$$S_2^{(A)} = -\log\left(\mathrm{tr}\,\rho_A^2\right) \tag{24}$$

with $\rho_A = \mathrm{tr}_B\,|\psi\rangle\langle\psi|$ being the reduced density matrix of subsystem A. One can re-express $S_2^{(A)}$ in a form which is convenient for Monte Carlo calculation by considering an identical copy S' of the system S with the same bipartition into subsystems A' and B', and defining a Swap_A operator acting on the tensor product of the Hilbert spaces of the two copies, which swaps the configurations in A and A'. More concretely, let $|\boldsymbol{\sigma}\rangle$ be a state in the coordinate basis of S (a spin configuration) which can be written as $|\boldsymbol{\sigma}\rangle = |\boldsymbol{\sigma}_A\boldsymbol{\sigma}_B\rangle$, where $\boldsymbol{\sigma}_A$ and $\boldsymbol{\sigma}_B$ are configurations in A and B respectively. Similarly, $|\boldsymbol{\sigma}'\rangle = |\boldsymbol{\sigma}'_{A'}\boldsymbol{\sigma}'_{B'}\rangle$ in S'. The swap operator acts on the tensor product of the two copies as $\mathrm{Swap}_A\big(|\boldsymbol{\sigma}\rangle \otimes |\boldsymbol{\sigma}'\rangle\big) = |\boldsymbol{\sigma}'_{A'}\boldsymbol{\sigma}_B\rangle \otimes |\boldsymbol{\sigma}_A\boldsymbol{\sigma}'_{B'}\rangle$. Using this definition it is easy to show that the second Rényi entropy can be rewritten as

$$S_2^{(A)} = -\log\left(\mathrm{tr}\,\rho_A^2\right) = -\log\langle \mathrm{Swap}_A\rangle \tag{25}$$

where the expectation value $\langle \mathrm{Swap}_A\rangle$ is taken over the product state $|\psi\rangle \otimes |\psi\rangle$ of the two copies. This expectation value reads

$$\langle \mathrm{Swap}_A\rangle = \sum_{\{\boldsymbol{\sigma}\},\{\boldsymbol{\sigma}'\}} P(\boldsymbol{\sigma})P(\boldsymbol{\sigma}')\frac{\langle \boldsymbol{\sigma}'_{A'}\boldsymbol{\sigma}_B|\psi\rangle\langle\boldsymbol{\sigma}_A\boldsymbol{\sigma}'_{B'}|\psi\rangle}{\langle \boldsymbol{\sigma}_A\boldsymbol{\sigma}_B|\psi\rangle\langle\boldsymbol{\sigma}'_{A'}\boldsymbol{\sigma}'_{B'}|\psi\rangle} \tag{26}$$

where, as before, $|\boldsymbol{\sigma}\rangle = |\boldsymbol{\sigma}_A\boldsymbol{\sigma}_B\rangle$ in S, $|\boldsymbol{\sigma}'\rangle = |\boldsymbol{\sigma}'_{A'}\boldsymbol{\sigma}'_{B'}\rangle$ in S', and $P(\boldsymbol{\sigma}) = \frac{|\langle\boldsymbol{\sigma}|\psi\rangle|^2}{\langle\psi|\psi\rangle}$ is the probability density for the $\boldsymbol{\sigma}$ configuration. The double sum in Eq. (26) can be replaced by sum over two sets of Monte Carlo samples of $P(\boldsymbol{\sigma})$ allowing for an efficient calculation of $S_2^{(A)}$.

AIV. TRANSLATION INVARIANT GTMS

We discuss here how to implement translation invariance in a GTMS network. We start our discussion from the case of a translation invariant MPS parametrized by a GTMS network. For a translation invariant MPS, the individual tensors $A^{[i]\sigma_i}_{a_{i-1}a_i}$ at each lattice site i are the independent of the site index i, namely $A^{[i]\sigma_i}_{a_{i-1}a_i} = A^{\sigma_i}_{a_{i-1}a_i}$. This suggests that a GTMS network parametrizing a translation invariant MPS must as well have the weights (the black and purple links in Fig. 2) independent of the site index, that is $w^{\mu}_{i,i} = w^{\mu}$, $\tilde{w}^{\mu,\nu}_i = \tilde{w}^{\mu,\nu}$, $\hat{w}^{\mu,\nu}_{i,i+1} = \hat{w}^{\mu,\nu}$, $c_i = c$, $b^{\mu}_i = b^{\mu}$ and $a^{\nu}_i = a^{\nu}$ (and $w^{\mu}_{i,j\neq i} = 0$). Therefore, for a translation invariant MPS it is sufficient to calculate the matrices $A^{\sigma_i}_{a_{i-1}a_i}$ for the different values of σ_i once.

On the top of the MPS weights, we can then add non-zero RBM weights $w^{\mu}_{i,j\neq i}$ in such a way that translation invariance is preserved. For this it is sufficient to set the weights $w^{\mu}_{i,j\neq i}$ as dependent only on the distance $|i-j|$ between the the physical site i and the position of the tensor j connected by the link. This means $w^{\mu}_{i,i+d} = w^{\mu}_d$, where one has to apply PBC by setting $i + N = i$ if N is the number of physical sites of the system. However, since now the transfer matrices $T_j(\boldsymbol{\sigma})$ depend in general on the spin configuration on the whole system, one still needs to calculate all of the T_j transfer matrices for each spin configuration.

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

[2] I. P. McCulloch, *Infinite size density matrix renormalization group*, revisited, arXiv:0804.2509 (2008).

[3] U. Schollwöck, *The Density-Matrix Renormalization Group in the Age of Matrix Product States*, Ann. Phys. **326**, 96 (2011).

[4] S. Östlund, S. Rommer, *Thermodynamic Limit of Density Matrix Renormalization*, Phys. Rev. Lett. **75** 19, 3537 (1995).

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

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

[7] L. Tagliacozzo, G. Evenbly, G. Vidal, *Simulation of two-dimensional quantum systems using a tree tensor network that exploits the entropic area law*, Phys. Rev. B **80**, 235127 (2009).

[8] V. Murg, F. Verstraete, Ö. Legeza, R. M. Noack, *Simulating strongly correlated quantum systems with tree tensor networks*, Phys. Rev. B **82**, 205105 (2010).

[9] M. Gerster, M. Rizzi, P. Silvi, M. Dalmonte, S. Montangero, *Fractional quantum Hall effect in the interacting Hofstadter model via tensor networks*, Phys. Rev. B **96** 195123 (2017).

[10] G. Vidal, *Class of Quantum Many-Body States that Can Be Efficiently Simulated*, Phys. Rev. Lett **101**, 110501 (2008).

[11] F. Verstraete, V. Murg, J.I. Cirac, *Matrix Product States, Projected Entangled Pair States, and Variational Renor-*

malization Group Methods for Quantum Spin Systems, Advances in Physics, **57** (2), 143-224 (2008).

[12] J. Eisert, M. Cramer, M. B. Plenio, *Colloquium: Area Laws for the Entanglement Entropy*, Rev. Mod. Phys. **82**, 277 (2010).

[13] M. B. Hastings, *An Area Law for One-Dimensional Quantum Systems*, J. Stat. Mech. (2007) P08024.

[14] L. Amico, R. Fazio, A. Osterloh, V. Vedral, *Entanglement in Many-Body Systems*, Rev. Mod. Phys. **80**, 517 (2008).

[15] G. Vidal, J. I. Latorre, E. Rico, A. Kitaev, *Entanglement in Quantum Critical Phenomena*, Phys. Rev. Lett **90**, 227902 (2003).

[16] U. Schollwöck, V. Meden, W. Metzner, K. Schönhammer, *DMRG Studies of Impurities in Luttinger Liquids*, Progress of Theoretical Physics Supplement **145** (1), 312?319 (2002).

[17] N. Laflorencie, E. S. Sorensen, M.-S. Chang, I. Affleck, *Boundary Effects in the Critical Scaling of Entanglement Entropy in 1D Systems*, Phys. Rev. Lett. **96**, 100603 (2006).

[18] G. Vidal, *Efficient Simulation of One-Dimensional Quantum Many-Body Systems*, Phys. Rev. Lett **93** (4), 040502 (2004).

[19] S. R. White, A. E. Feiguin, *Real-Time Evolution Using the Density Matrix Renormalization Group*, Phys. Rev. Lett **93** (7), 076401 (2004).

[20] M. P. Zaletel, R. S. K. Mong, C. Karrasch, J. E. Moore, F. Pollmann, *Time-evolving a matrix product state with long-ranged interactions*, Phys. Rev. B **91**, 165112 (2015).

[21] R. Jastrow, *Many-Body Problem with Strong Forces*, Phys. Rev. **98** (5), 1479-1484 (1955).

[22] M. C. Gutzwiller, *Correlation of Electrons in a Narrow s Band*, Phys. Rev. **137** (6A), A1726-A1735 (1965).

[23] W. L. McMillan, *Ground State of Liquid* He^4, Phys. Rev. **138** (2A), A442-A451 (1965).

[24] D. Ceperley, G. V. Chester, M. H. Kalos, *Monte Carlo simulation of a many-fermion study*, Phys. Rev. B **16** (7), 3081-3099 (1977).

[25] S. Sorella, *Wave function optimization in the variational Monte Carlo method*, Phys. Rev. B **71**, 241103(R) (2005).

[26] M. Casula, C. Attaccalite, S. Sorella, *Correlated geminal wave function for molecules: An efficient resonating valence bond approach*, J. Chem. Phys. **121**, 7110 (2004).

[27] Y. LeCun, Y. Bengio, G. Hinton, *Deep Learning*, Nature **521**, 436-444 (2015).

[28] J. Carrasquilla, R. G. Melko, *Machine Learning Phases of Matter*, Nature Physics **13**, 431-434 (2017).

[29] L. Wang, *Discovering Phase Transitions with Unsupervised Learning*, Phys. Rev. B **94**, 195105 (2016).

[30] N. Le Roux, Y. Bengio, *Representational Power of Restricted Boltzmann Machines and Deep Belief Networks*, Neural Comput. **20**, 1631-1649 (2008).

[31] G. Carleo, M. Troyer, *Solving the quantum many-body problem with artificial neural networks*, Science **355**, 602 (2017).

[32] D.-L. Deng, X. Li, S. Das Sarma, *Quantum Entanglement in Neural Network States*, Phys. Rev. X **7**, 021021 (2017).

[33] D.-L. Deng, X. Li, S. Das Sarma, *Machine learning topological states*, Phys. Rev. B **96**, 195145 (2017).

[34] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, *Restricted Boltzmann machine learning for solving strongly correlated quantum systems*, Phys. Rev. B **96**, 205152 (2017).

[35] H. Saito and M. Kato, *Machine Learning Technique to Find Quantum Many-Body Ground States of Bosons on a Lattice*, J. Phys. Soc. Jpn. **87**, 014001 (2018).

[36] R. Kaubruegger, L. Pastori, J. C. Budich, *Chiral Topological Phases from Artificial Neural Networks*, Phys. Rev. B **97**, 195136 (2018).

[37] R. Salakhutdinov, G. R. Hinton, *Deep Boltzmann Machines*, PMLR **5**, 448-455 (2009).

[38] X. Gao, L.-M. Duan, *Efficient representation of quantum many-body states with deep neural networks*, Nature Communications **8**, 662 (2017).

[39] G. Carleo, Y. Nomura, M. Imada, *Constructing exact representations of quantum many-body systems with deep neural networks*, arXiv:1802.09558 (2018).

[40] Y. Huang, J. E. Moore, *Neural network representation of tensor network and chiral states*, arXiv:1701.06246 (2017).

[41] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, *Neural Networks Quantum States, String-Bond States and chiral topological states*, Phys. Rev. X **8**, 011006 (2018).

[42] S. R. Clark, *Unifying Neural-network Quantum States and Correlator Product States via Tensor Networks*, J. Phys. A: Math. Theor. **51**, 135301 (2018).

[43] J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang, *Equivalence of restricted Boltzmann machines and tensor network states*, Phys. Rev. B **97**, 085104 (2018).

[44] I. Affleck, T. Kennedy, E. H. Lieb, H. Tasaki, *Rigorous results on valence-bond ground states in antiferromagnets*, Phys. Rev. Lett. **59** (7), 799-802 (1987).

[45] I. Affleck, T. Kennedy, E. H. Lieb, H. Tasaki, *Valence bond ground states in isotropic quantum antiferromagnets*, Commun. Math. Phys. **115**, 477 (1988).

[46] M. den Nijs, K. Rommelse, *Preroughening transitions in crystal surfaces and valence-bond phases in quantum spin chains*, Phys. Rev. B **40**, 4709 (1989).

[47] L. Bottou, *Large-Scale Machine Learning with Stochastic Gradient Descent*, Proceedings of the 19th International Conference on Computational Statistics, 177-187 (2010).

[48] L. Bottou, F. E. Curtis, J. Nocedal, *Optimization Methods for Large-Scale Machine Learning*, Siam Reviews, **60** (2), 223-311 (2018).

[49] J. Duchi, E. Hazan, Y. Singer, *Adaptive subgradient methods for online learning and stochastic optimization*, Journal of Machine Learning Research **12**, 2121 (2011).

[50] D. P. Kingma, J. L. Ba, *Adam: A method for stochastic optimization*, arXiv:1412.6980 (2014).

[51] M. B. Hastings, I. González, A. B. Kallin, R. G. Melko, *Measuring Renyi Entanglement Entropy in Quantum Monte Carlo Simulations*, Phys. Rev. Lett. **104**, 157201 (2010).