

This is the accepted manuscript made available via CHORUS. The article has been published as:

## Plaquette renormalization scheme for tensor network states

Ling Wang, Ying-Jer Kao, and Anders W. Sandvik

Phys. Rev. E **83**, 056703 — Published 6 May 2011

DOI: [10.1103/PhysRevE.83.056703](https://doi.org/10.1103/PhysRevE.83.056703)

# Plaquette Renormalization Scheme for Tensor Network States

Ling Wang,<sup>1,2</sup> Ying-Jer Kao,<sup>3</sup> and Anders W. Sandvik<sup>1,3</sup>

<sup>1</sup>*Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215*

<sup>2</sup>*University of Vienna, Faculty of Physics, Boltzmannngasse 5, 1090 Vienna, Austria*

<sup>3</sup>*Department of Physics, National Taiwan University, Taipei, Taiwan 106*

We present a method for contracting a square-lattice tensor network in two dimensions, based on auxiliary tensors accomplishing successive truncations (renormalization) of 8-index tensors for  $2 \times 2$  plaquettes into 4-index tensors. Since all approximations are done on the wave function (which also can be interpreted in terms of different kind of tensor network), the scheme is variational, and, thus, the tensors can be optimized by minimizing the energy. Test results for the quantum phase transition of the transverse-field Ising model confirm that even the smallest possible tensors (two values for each tensor index at each renormalization level) produce much better results than the simple product (mean-field) state.

PACS numbers: 02.70.Ss, 75.10.Jm, 75.40.Mg, 75.40.Cx

## I. INTRODUCTION

Tensor network states (TNSs) [1–4] are emerging as a promising route toward unbiased modeling of challenging quantum many-body systems, such as frustrated spins. These correlated states are higher-dimensional generalizations of matrix product states (MPSs) [5, 6], which are implicitly produced in density matrix renormalization group (DMRG) calculations [7, 8] and are known to faithfully represent ground states of one-dimensional (1D) hamiltonians with short-range interactions [9]. The matrix size  $m$  has to increase at most polynomially with the system size  $N$ , which underlies the success of the DMRG method in 1D. For 2D and 3D systems, correlations are not reproduced properly by MPSs [3], due to the inherently 1D nature of the local quantum entanglement in these states (although improved schemes [11, 12] can restore 2D or 3D uniformity), and  $m$  then has to grow exponentially with  $N$ . In the TNSs, the matrices are replaced by tensors of rank corresponding to the coordination number of the lattice, e.g., on a 2D square lattice the tensors  $T_{ijkl}^s(\sigma_s)$  for each site  $s$  have four indices, in addition to their physical index (here the  $z$ -component  $\sigma_s$  of a spin), as illustrated in Fig. 1. Contracting over the “bond indices” ( $i, j, k, l$ ) gives the wave function coefficient for given spin state  $\sigma_1, \dots, \sigma_N$ .

While it is believed, based on entanglement entropy considerations [3], that TNSs can represent ground states of short-range 2D and 3D hamiltonians, a serious problem in practice is that contracting the tensors is, in general, a problem which scales exponentially in the system size and the dimension of the tensors. To overcome this challenge, approximate ways to compute the contraction have been proposed [1, 13, 14]. Another approach is to use tree-tensor networks [4], or more sophisticated extensions of these [9], which can be efficiently contracted. Promising results based on TNSs have already been reported for several quantum spin models, but further reduction of the computational complexity, while maintaining the ability of the TNSs to properly account for en-

tanglement, will still be necessary before the most challenging systems can be studied reliably.

In this paper, we introduce a plaquette renormalization scheme for 2D TNSs inspired by a method introduced by Levin and Nave [15] in the context of classical Ising models. They suggested to replace the effective tensors for  $2 \times 2$  plaquettes on the square lattice, which have  $m^8$  elements, by some “renormalized” tensors with  $m_{\text{cut}}^4$  elements, as illustrated in Fig. 2(a). This can be done exactly if all tensor elements are kept, whence  $m_{\text{cut}} = m^2$ , as this corresponds just to a regrouping of the tensor indices. The idea is that the scheme may provide a good approximation even if the tensors are drastically truncated, e.g., with  $m_{\text{cut}} \sim m$ . For a square  $L \times L$  lattice, the new tensors describing  $2 \times 2$  spins should be contracted on a new lattice of length  $L/2$ . If the original  $L$  is a power of 2, this decimation can be continued until there is a single tensor left, which is then contracted with itself (under periodic boundary conditions) to give the wave function. The question is then how to accomplish the tensor truncation in a way which preserves the quantum state in the best way.

We here propose a renormalization procedure based on auxiliary tensors introduced at the level of the wave function, as illustrated in Fig. 3, in contrast to recent schemes

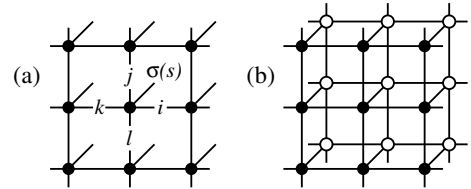


FIG. 1: Tensor networks on the 2D square lattice. (a) For each site  $s$  there is a 4-index tensor  $T_{ijkl}^s(\sigma_s)$ , where  $\sigma_s$  is the physical index (here a spin  $\sigma_s = \pm 1$ ) and  $i, j, k, l \in \{1, \dots, m\}$ . The wave function  $\Psi(\sigma_1, \dots, \sigma_N)$  is the  $N$ -tensor product (contraction) defined as a summation over all shared indices on the lattice bonds. (b) The norm  $\langle \Psi | \Psi \rangle$  is obtained by contracting also over the physical indices.

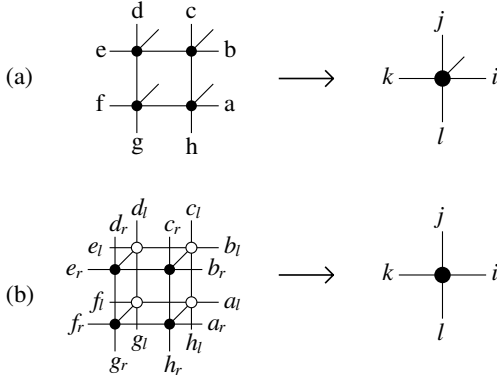


FIG. 2: (a) Truncation (renormalization) of an 8-index plaquette tensor into an effective 4-index tensor. The indices  $a, \dots, h \in \{1, \dots, m\}$  and  $i, j, k, l \in \{1, \dots, m_{\text{cut}}\}$ . The diagonal lines indicate physical indices, which for an  $S = 1/2$  spin system can take two values before the renormalization and 16 values after (and is further multiplied by 16 after each successive renormalization). In (b) the physical indices are traced out first, leading to double tensors. To stay with the same level of truncation as in (a), the indices after the renormalization should then take values  $i, j, k, l \in \{1, \dots, m_{\text{cut}}^2\}$ .

[13, 14] that apply singular-value decompositions on the “double tensor” product obtained when the physical indices have been traced out. Our approach is strictly variational, and can be used also in combination with Monte Carlo sampling of the spins [16]. Here we carry out calculations on the 2D transverse-field Ising model, performing the contraction over the spins exactly.

The outline of the paper is as follows: In Sec. II we first discuss tensor renormalization schemes in general and then provide details of our variant. We test the method by studying the quantum phase transition in the 2D transverse-field Ising model in Sec. III. In Sec. IV we summarize and discuss several aspects of the method and its possible future extensions and applications.

## II. TENSOR RENORMALIZATION

In order to compute physical expectation values based on a TNS, one has to contract the tensors of a bra and ket state over their physical (e.g., spin) indices in addition to the bond indices of the tensors. The full contraction of such a 2D double-tensor network, illustrated in Figs. 1(b,c), is the norm  $\langle \Psi | \Psi \rangle$ . A matrix element  $\langle \Psi | A | \Psi \rangle$  of some operator involving one or several sites can be treated in a similar way. In practice, in most TNS approaches, one would first construct the double tensors

$$D_{abcd}^s = \sum_{\sigma_s = \uparrow, \downarrow} T_{i_2 j_2 k_2 l_2}^{s*}(\sigma_s) T_{i_1 j_1 k_1 l_1}^s(\sigma_s), \quad (1)$$

where the labels  $a, b, c, d$  is a suitable combination of the indices of the bra ( $T^{2*}$ ) and ket ( $T^s$ ) tensors, i.e.,  $a = i_1 + m(i_2 - 1)$ , etc. Similar tensors are constructed

for the sites at which operators act in a local expectation value. The full contraction of the double tensors  $D$ , which have bond dimensions  $m^2$ , is then carried out in some approximate way. A plaquette renormalization in the Levin-Nave scheme with double tensors on the square lattice is depicted in Fig. 2(b). This scheme is exact if the full tensors (with bond dimension  $m^4$ ) are kept, but this is infeasible in practice. The scheme may be a good approximation for some judiciously chosen truncated (renormalized) tensor, but how to find the optimal way to construct it is an open question. Gu, Levin, and Wen implemented a singular value decomposition (SVD) scheme [13] for the double tensors, and a similar method was proposed by Jiang, Weng, and Xiang [14].

In our scheme, the renormalization is instead accomplished with the aid of auxiliary 3-index tensors  $S_{abc}^n$  in the wave function, which transform and truncate pairs of indices of the plaquette tensors, as shown in Fig. 3. A sequence of plaquette renormalizations,  $n = 1, 2, \dots$ , effectively corresponds to a different kind of tensor network, which is illustrated in the case of the wave function on an  $8 \times 8$  lattice in Fig. 4. An advantage of working with the wave function, instead of the contracted double-tensor network, is that the method remains variational, regardless of how the  $S$ -tensors are chosen. This is not strictly the case when approximations are done when contracting the double tensors.

Here we will apply this scheme to the transverse-field Ising model (which has become the bench-mark of choice for initial tests within TNS approaches);

$$H = -J \sum_{\langle ij \rangle} \sigma_i^z \sigma_j^z - h \sum_i \sigma_i^x, \quad (2)$$

where  $\sigma_i^x$  and  $\sigma_i^z$  are standard Pauli matrices and  $\langle ij \rangle$  denotes nearest-neighbor site pairs on a 2D square  $L \times L$  lattice with periodic boundary conditions. The ground-state wave function of this system is translationally invariant and positive-definite. We can then take the original tensors to be site-independent, i.e., there are just two tensors  $T_{ijkl}(\sigma_s = \pm 1)$  and the tensors  $S^n$  are the same for all plaquettes on a given renormalization level  $n = 1, \dots, \log_2(L) - 1$ . At the last level, four tensors remain (see Fig. 4), which we contract directly. The problem is now to optimize the elements of the  $T$  and  $S$  tensors, to minimize the energy. Before proceeding to calculations, several comments are in order.

It is clear from Fig. 4 that the plaquette renormalization (no matter how it is accomplished) breaks translational symmetry, which, in the optimized state, should gradually be restored with increasing  $m$ . A way to restore the symmetry for finite  $m_{\text{cut}}$  is to sum over all (here four) symmetrically non-equivalent ways of arranging the plaquettes on the lattice at each level [12]. In a similar way, one can also ensure that the wave function is symmetric under other lattice transformations (rotations and reflections) for arbitrary  $T$  and  $S^n$  (as an alternative to enforcing these symmetry in the individual tensors), and spin-inversion symmetry can be implemented in a similar

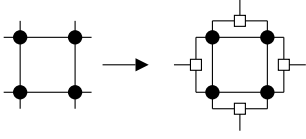


FIG. 3: Renormalization of an 8-index plaquette tensor using auxiliary 3-index tensors. The solid circles denote either the original tensors  $T = T^0$ , which depend on the physical spins (which are not shown here), or tensors  $T^n$  arising after  $n$  renormalization steps have been carried out (and depend on the  $2^{2^n}$  spins within the block). The squares denote 3-index tensors  $S^n$  by which the external indices are decimated by contracting common  $T^{n-1}$  and  $S^n$  indices. The remaining four free indices take values  $1, \dots, m_{\text{cut}}$ .

way. These symmetrization procedures (which also enable studies of states with different quantum-numbers of the symmetry operators) can be carried out if the spins are sampled using Monte Carlo simulations [16], but cannot be easily used with the double-tensor network. Here, in this initial test of the scheme, we will trace out the spins exactly and work with the double tensors. When minimizing the energy, it is then important to calculate the full, translationally averaged energy, not just the site and bond energies on a single plaquette (as is done in the SVD scheme of Gu *et al.* [13]), in order to maintain the variational property of the scheme. No approximations are made when contracting the effective renormalized tensor network, also when using the double tensors. In contrast, the SVD applied to the double-tensor network [13] introduces an approximation due to which the calculated energy becomes non-variational. This can cause problems when optimizing the tensors. Note that in our double-tensor approach, there is a pair of equal  $S$  tensors for each plaquette edge (one from  $\langle \Psi |$  and one from  $|\Psi \rangle$ ) and one cannot combine these into arbitrary tensors, which would make the scheme non-variational.

The renormalization of a plaquette according to Fig. 3 requires 12 internal index summations for each combination of the four external indices of the renormalized tensor, which can be carried out with  $\sim m^8$  operations (assuming  $m_{\text{cut}} = m$ , which we will use here). One example of a sequence of contractions giving this scaling is shown in Fig. 5. When working with double tensors (as we do here, contracting first the spin dependent tensors at the lowest level over the spin indices), the scaling becomes  $m^{16}$ , since all external and internal indices should then take  $m_{\text{cut}}^2 = m^2$  values. Note that the structure with  $S$  tensors renormalizing the bra and ket state (the same for the bra and the ket) has to be preserved at each level, i.e., it is not correct to just compute the double tensor with the lowest-level spin dependent tensors and after that use a single  $S$  tensor on each level to renormalize the double-tensor network. Such a scheme would not be variational.

The derivatives of the energy with respect to all the tensor elements, which we use to minimize the energy, can be evaluated with  $m^{12}$  ( $m^{24}$  with the double ten-

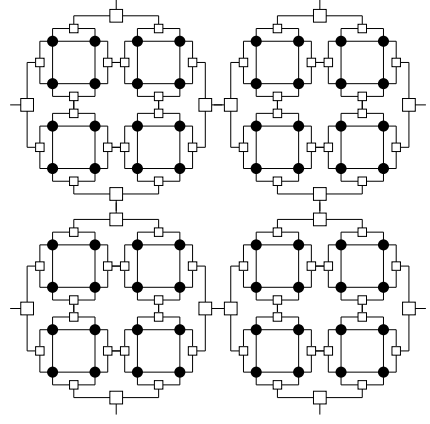


FIG. 4: The effective reduced tensor network for the wave function on an  $8 \times 8$  lattice. Here there are two sets of  $S$  tensors;  $S^1$  and  $S^2$ , denoted by the smaller and larger squares. The solid circles correspond to the original, spin dependent tensor  $T^0$ . At the lowest level (the black circles) there is also a spin index of the tensors which is not indicated here.

sors), using a chain-rule procedure carried out along with the renormalization steps. We need the derivative of  $E$  with respect to the elements of the spin-dependent tensors  $T^0(\sigma)$ , as well as the renormalization tensors  $S^n$  for all levels  $n = 1, \dots, n_{\text{max}}$ . The procedure is very similar for both  $T$  and  $S$  and we write down expressions only for the former case. To simplify the notation we suppress the spin dependence of the tensors  $T^0(\sigma)$  (i.e., there are equations for both  $\sigma = \pm 1$ ). We first note that we can write the derivative needed in terms of the renormalized tensor at any level  $n$  as

$$\frac{\partial E}{\partial T^0_{i_0 j_0 k_0 l_0}} = \sum_{T^n_{i_n j_n k_n l_n}} \frac{\partial E}{\partial T^n_{i_n j_n k_n l_n}} \frac{\partial T^n_{i_n j_n k_n l_n}}{\partial T^0_{i_0 j_0 k_0 l_0}}. \quad (3)$$

At each level we apply the chain rule

$$\frac{\partial T^n_{i_n j_n k_n l_n}}{\partial T^0_{i_0 j_0 k_0 l_0}} = \frac{\partial T^n_{i_n j_n k_n l_n}}{\partial T^{n-1}_{i_{n-1} j_{n-1} k_{n-1} l_{n-1}}} \frac{\partial T^{n-1}_{i_{n-1} j_{n-1} k_{n-1} l_{n-1}}}{\partial T^0_{i_0 j_0 k_0 l_0}}, \quad (4)$$

which leads to the form

$$\frac{\partial E}{\partial T^0_{i_0 j_0 k_0 l_0}} = \sum_{T^n_{i_n j_n k_n l_n}} \dots \sum_{T^2_{i_2 j_2 k_2 l_2}} \sum_{T^1_{i_1 j_1 k_1 l_1}} \frac{\partial E}{\partial T^n_{i_n j_n k_n l_n}} \times \frac{\partial T^n_{i_n j_n k_n l_n}}{\partial T^{n-1}_{i_{n-1} j_{n-1} k_{n-1} l_{n-1}}} \dots \frac{\partial T^2_{i_2 j_2 k_2 l_2}}{\partial T^1_{i_1 j_1 k_1 l_1}} \frac{\partial T^1_{i_1 j_1 k_1 l_1}}{\partial T^0_{i_0 j_0 k_0 l_0}}. \quad (5)$$

All the factors can be easily computed and stored during the renormalization procedure, albeit at the rather high cost  $m^{24}$  when all contractions are carried out exactly (while with Monte Carlo sampling the cost is  $m^{12}$ ).

We update the tensors, using the steepest decent method or a stochastic scheme where only the sign of the derivatives is used [18] (which to a large extent avoids trapping in local minimas). In this process we normalize

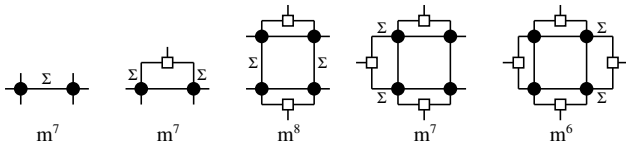


FIG. 5: Steps in the contraction of the renormalized plaquette tensor of Fig. 3. Partial contractions are constructed from left to right. The summations carried out are indicated by  $\Sigma$  and the scaling of each step with  $m$  is shown beneath the diagrams. Each summation and free index contributes a factor  $m$ .

the elements of the  $T^0(\sigma)$  and  $S^n$  tensors such that the largest (in magnitude) elements in each are of order 1 [in the case of  $T^0(\sigma)$  taken as the largest among all elements for  $\sigma = \pm 1$ , since the  $\sigma = \uparrow$  and  $\downarrow$  tensors cannot be rescaled independently of each other). This can lead to problems with elements either too large or too small during the contraction of the full tensor network. At this stage, we therefore perform a separate rescaling in the following way: The individual  $T^0(\sigma)$ -tensors are normalized by contracting them on themselves as if they described a single isolated spin, i.e., first forming the double tensor  $D_{abcd}$  in (1) and then contracting by summing  $D_{aacc}$  over  $a$  (assumed to correspond to the “left” and “right” bond indices) and  $c$  (the “up” and “down” indices). The elements of  $T^0(\sigma)$  are then rescaled so that this single-tensor contraction equals one. A similar rescaling is done for the  $S^n$  tensors at each level, after contracting the renormalized plaquette tensors  $T^n$  with themselves according to the above scheme. We have not encountered any numerical instabilities when this scheme is used.

It should be noted that the optimized  $T$  tensors do not necessarily constitute a good TNS when contracted without the  $S$  tensors, because they are optimized together. On the other hand, it should also be possible to construct special  $S$  tensors that effectively perform something very similar to a SVD (although globally optimized, not locally as in Ref. 13, 14) and then the optimized  $T$  tensors by themselves should also form a good TNS when assembled into a standard 2D tensor network. Here we do not impose any such conditions on the  $S$  tensors.

### III. RESULTS

We now discuss calculations for the hamiltonian (2). For  $L \rightarrow \infty$ , the ground state of this system undergoes a quantum phase transition in the 3D Ising universality class, at a critical field  $h_c/J \approx 3.04$  determined using quantum Monte Carlo simulations [17]. Here we study the model using the smallest possible tensors and truncation,  $m = m_{\text{cut}} = 2$ , using the double-tensor approach for lattices of size  $L = 4, 8, 16, 32$ , and  $64$ . For the  $T^0(\sigma)$  tensors, we enforce symmetry with respect to rotations of the indices, for a total of 6 free parameters each for  $T^0(1)$  and  $T^0(-1)$ . The  $S$  tensors are symmetric in the two indices used in the internal contractions of the pla-

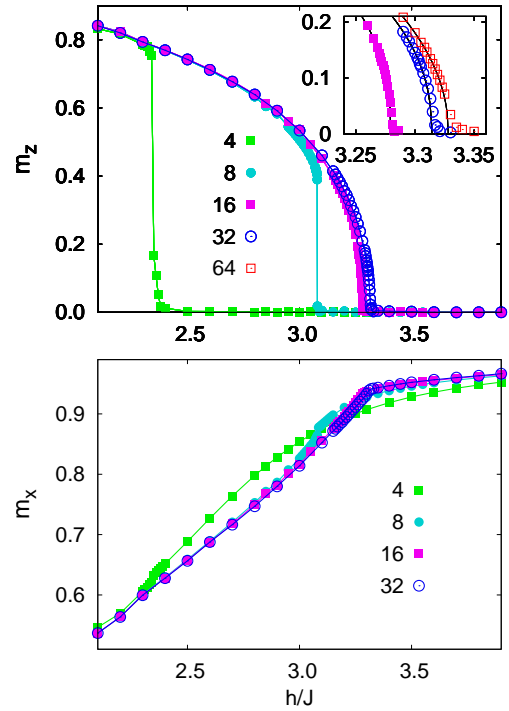


FIG. 6: (Color on-line) Spontaneous magnetization  $m_z$  (upper panel) and field-induced polarization  $m_x$  (lower panel) versus the transverse field for different  $L \times L$  lattices. The inset in the upper panel shows the behavior close to the critical point in greater detail; the solid curves are power-law fits, as discussed in the text and shown on a different scale in Fig. 7.

quettes, and so there are 6 free parameters also for each  $S^n$ .

Starting with random tensor elements, we calculate the energy and its derivatives with respect to all the parameters. We then update the parameters based on the derivatives, normally using a stochastic optimization scheme of the kind discussed in Ref. 18. The plaquette renormalizations and the derivative calculations are highly parallelizable, which we take advantage of by using a massively parallel computer [19]. In general the method performs very well, although occasionally the energy converges to a local minimum, especially close to the critical point.

The method can produce solutions breaking spin-inversion symmetry, in which case we do not obtain the true ground state for a given system size  $L$ , which can have no broken symmetries for finite  $L$ . This is to be expected, in analogy with how mean-field theory (corresponding to the  $m = 1$  simple product state) produces symmetry-broken states for  $h < h_c$ . Thus we can simply compute the magnetization  $m_z = \langle \sigma_s^z \rangle$  (averaged over all sites  $s$ ) and study its behavior for increasing  $L$ . Note that for fixed  $L$ , spin-inversion symmetry should gradually be restored with increasing  $m$ , as the optimized state should approach the true finite- $L$  symmetric ground state. However, for any fixed  $m$ , we expect the symmetry to be broken when  $L \rightarrow \infty$  for  $h$  below some  $m$ -dependent  $h_c$ . Here we only consider  $m = 2$ . Results for  $m_z$  and the

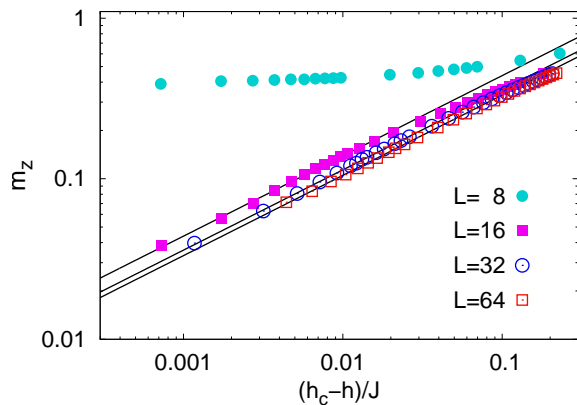


FIG. 7: (Color on-line) Critical scaling of the magnetization (the order parameter). The behavior for the larger system sizes is consistent with mean-field behavior (exponent  $\beta = 1/2$ ), as shown with the straight lines.

field-induced  $m_x = \langle \sigma_s^x \rangle$  are shown versus the field  $h/J$  in Fig. 6. The transition point between the magnetic and paramagnetic states moves toward higher fields with increasing  $L$ , converging to  $h_c/J \approx 3.33$  for the largest sizes. This is much closer to the unbiased quantum Monte Carlo result [17]  $h_c/J \approx 3.04$  than the mean-field ( $m = 1$ ) value  $h_c/J = 4$ .

While other renormalization schemes produce transition points even closer to the correct result for  $m = 2$  [13], the results are not directly comparable because here we use  $m_{\text{cut}} = 2$  throughout the renormalization, while much larger  $m_{\text{cut}}$  as used in [13].

There is some rounding of the magnetization curve close to the transition for  $h > h_c(L)$ , which becomes less pronounced with increasing  $L$ . One the other hand, the magnetization curve for  $h < h_c(L)$  becomes less sharp with increasing  $L$ . For  $L = 4$  and  $8$  the transitions look almost first-order, with apparent jumps of the order parameter which diminish with increasing  $L$ . This is very similar to the behavior found for matrix-product state with small matrices in Ref. 21, where it was shown that there is a crossing of two energy minimas in the space of matrix elements for finite systems. As the system size grows, these minimas move close to each other and for infinite size the transition is continuous. If this mechanism is at play also here, then there should be no finite size rounding associated with the jumps. The rounding that we do observe for  $m^z$  in Fig. 6 can in principle be due to incomplete optimization of the tensors, which we cannot completely rule out. The results are, however, reproducible, and also the rounding effects are smaller for the larger systems (where one would expect the optimization to be more difficult). It is therefore possible that the energy landscape of the plaquette-renormalized TNS is different from that of the MPSs, and the jump is not completely discontinuous.

Analyzing the critical behavior by fitting a power-law,  $(h_c - h)^\beta$ , to the data for  $L \geq 16$  (where we do not ob-

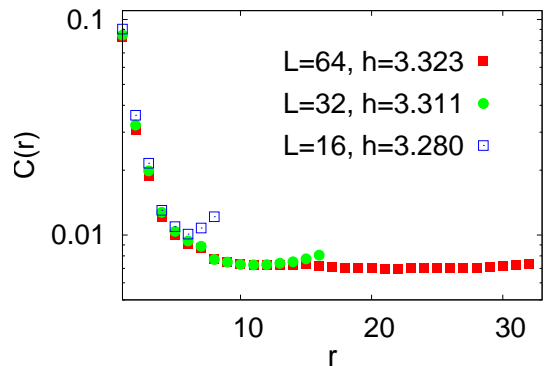


FIG. 8: (Color on-line) Spin-spin correlation versus separation  $r$  at fields  $h$  close to  $h_c$  such that the long-distance correlations are approximately the same for system sizes  $L = 16, 32$ , and  $64$ .

serve any jumps), we find that the behavior close to the transition can be described by a mean-field exponent,  $\beta = 1/2$ , as shown in Fig. 7. For  $L = 8$  the behavior is different, with a very small exponent (related to the “quasi-discontinuities” discussed above). In Ref. 21 it was argued the critical behavior for TNSs should, in general, be expected to be of mean-field form for systems of infinite size and finite tensor (or matrix, in one dimension) size  $m$ . The true critical behavior (exponent) should emerge with increasing  $m$  (for large system sizes) in a window that tends toward  $h_c$  as the system size is increased.

Fig. 8 shows the spin-spin correlation function  $C(r_{ij}) = \langle \sigma_i^z \sigma_j^z \rangle$  averaged over all equidistant spins close to  $h_c$ . We have here chosen the particular values of  $h$  for each system size in such a way that the curves for different system sizes coincide approximately and that the long-distance value is relatively small (i.e., a weakly ordered system). The behavior is clearly different from a simple  $m = 1$  product state (mean-field theory), which only gives a constant  $C(r) = \langle \sigma_i^z \rangle^2$  for  $r > 0$ . The correlations instead decay over a distance of several lattice spacings. This shows that the scheme can account for non-trivial quantum fluctuations. Characteristic boundary enhancements of the correlations at  $r \approx L/2$  are also seen. It is not possible, however, to observe truly critical correlations with the small,  $m = 2$ , tensors that we have used here. This is related to the asymptotic mean-field behavior seen in Fig. 7.

We finally confirm that the calculations satisfy the variational bound of the energy. Unbiased results for comparison was obtained using quantum Monte Carlo calculations [22] for the same lattice sizes. Fig. 9 shows the energy versus the field for  $L = 32$ , as well as the relative error for several system sizes. The results of the tensor-network calculations are always higher. As expected, the error is peaked close to the phase transition. It also grows slightly with the system size, which is also not unexpected. For  $L = 32$ , some of the data points (the one at the peak as well as those for the largest fields) are

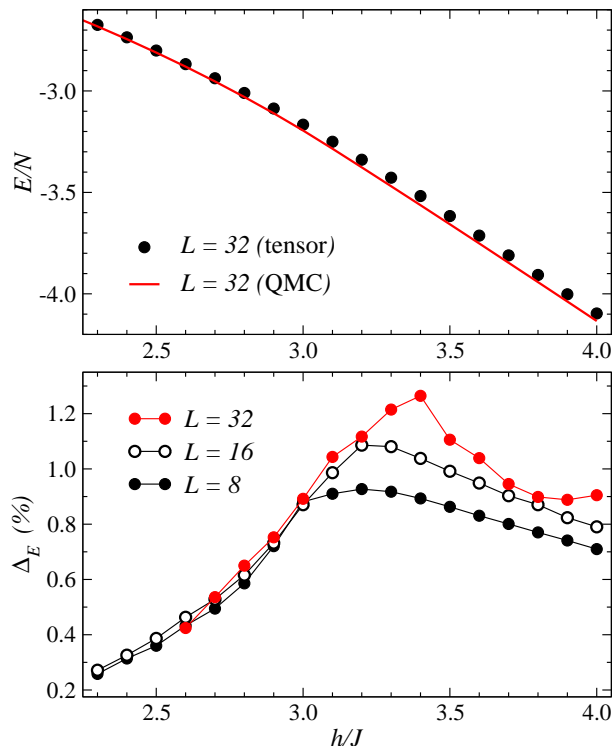


FIG. 9: (Color on-line) Upper panel: Ground state energy per spin obtained with the tensor-network ansatz for an  $L = 32$  lattice compared with unbiased quantum Monte Carlo calculations for the same systems. Lower panel: The relative energy error in the the tensor-network calculation (i.e., the deviation from the QMC results) for system sizes  $L = 8, 16$ , and  $32$ .

likely affected by incomplete optimization of the tensors.

#### IV. SUMMARY AND DISCUSSION

In summary, we have presented a scheme using auxiliary tensors to renormalize plaquette tensors in a 2D tensor network. The approach can also be regarded as a different tensor network, which can be contracted efficiently. Figs. 3 and 4 summarize the approach pictorially. Using the example of the transverse-field Ising model, we have shown that the scheme produces results far better than mean-field theory, even with the smallest possible non-trivial tensors and truncation ( $m = 2$ ). The scaling to the thermodynamic limit is well-behaved. Based on these results, we expect a fast convergence to the ex-

act ground state with increasing  $m$ . Increasing  $m$  to 3 is already quite challenging within the double-tensor approach that we have employed here, since the scaling is  $\propto m^{24}$  (when computing all energy derivatives). However, in variational Monte Carlo simulations (sampling the spins and optimizing the tensors based on stochastic estimates of the derivatives [16]), the scaling is  $m^{12}$ , and it should then be easier to study larger  $m$ . Using some optimization method not requiring derivatives (e.g., the methods discussed in [21]) formally brings the effort down to  $m^8$  and  $m^{16}$  and when using Monte Carlo sampling and exact spin tracing, respectively.

Here we optimized the tensors variationally, which requires the energy averaged over all non-equivalent sites and bonds. The computational effort then scales with the system size as  $L^2 \log(L)$ . If we optimize only a local energy, which is not a variational estimate of the total energy but can produce good results in the SVD approach [13] (and may work well also in our scheme for some particular classes of  $S$  tensors), the scaling is  $\log(L)$ , as in SVD based schemes. It may also be possible to use imaginary-time evolution (ground-state projection), as is often done in other TNS approaches [20].

Application to other quantum spin systems (and even fermions) is in principle straight-forward, although the convergence with  $m$  can of course be expected to be model dependent.

A technically appealing feature of our scheme is that the plaquette renormalization procedure can be implemented very efficiently on GPUs (graphics processing units, the use of which is emerging as an important trend in high-performance computing [23]). The speed-up relative to a standard CPU can be very significant for large tensors. For  $m = 3$ , we have achieved an efficiency boost of  $\approx 25$  for the plaquette contraction of the double tensor in a more computationally challenging model, the  $J_1 - J_2$  Heisenberg model.[24]. It should be noted that for the current case with  $m = 2$ , the tensor size is too small to obtain any performance gain from GPU as the amount of computations needs to be done can not hide the memory latency. We plan to use this approach in future model studies with larger  $m$ .

We would like to thank Zheng-Cheng Gu, Frank Verstraete, and Xiao-Gang Wen for useful discussions. This work is supported by NSF Grant No. DMR-0803510 (AWS) and by NSC Grant No. 97-2628-M-002-011-MY3 and NTU 99R0066-69 (YJK). YJK. and AWS would also like to thank the NCTS of Taiwan for travel support.

- 
- [1] F. Verstraete and J. I. Cirac, arXiv:cond-mat/0407066.
  - [2] H. Takasaki, T. Hikihara, and T. Nishino, J. Phys. Soc. Jpn. **68**, 1537 (1999); T. Nishino *et al.*, Nucl. Phys. B **575**, 504 (2000).
  - [3] F. Verstraete, M. M. Wolf, D. Perez-Garcia, and J. I.

- Cirac, Phys. Rev. Lett. **96**, 220601 (2006).
- [4] Y.-Y. Shi, L.-M. Duan, and G. Vidal, Phys. Rev. A **74**, 022320 (2006).
- [5] S. Östlund and S. Rommer, Phys. Rev. Lett. **75**, 3537 (1995).

- [6] F. Verstraete, D. Porras, J. I. Cirac, Phys. Rev. Lett. **93**, 227205 (2004).
- [7] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).
- [8] U. Schollwöck, Rev. Mod. Phys. **77**, 259 (2005).
- [9] G. Vidal, Phys. Rev. Lett. **99**, 220405 (2007); Phys. Rev. Lett. **101**, 110501 (2008); G. Evenbly and G. Vidal, arXiv:0811.0879.
- [10] G. Vidal, Phys. Rev. Lett. **93**, 040502 (2004).
- [11] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, Phys. Rev. Lett **100**, 040501 (2008).
- [12] A. W. Sandvik, Phys. Rev. Lett. **101**, 140603 (2008).
- [13] Z.-C. Gu, M. Levin, and X.-G. Wen, Phys. Rev. B **78**, 205116 (2008).
- [14] H. C. Jiang, Z. Y. Weng, and T. Xiang, Phys. Rev. Lett. **101**, 090603 (2008).
- [15] M. Levin and C. P. Nave, Phys. Rev. Lett. **99**, 120601 (2007).
- [16] A. W. Sandvik and G. Vidal, Phys. Rev. Lett. **99**, 220602 (2007).
- [17] H. Rieger and N. Kawashima, Eur. Phys. J. B **9**, 233 (1999).
- [18] J. Lou and A. W. Sandvik, Phys. Rev. B **76**, 104432 (2007).
- [19] The IBM BlueGene L at the Center for Computational Science at Boston University.
- [20] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, Phys. Rev. Lett **98**, 140506 (2007).
- [21] C. Liu, L. Wang, A. W. Sandvik, Y.-C. Su, and Y.-J. Kao, Phys. Rev. B **82**, 060410 (2010).
- [22] A. W. Sandvik, Phys. Rev. E **68**, 056701 (2003).
- [23] See <http://gpgpu.org> for articles and discussion of general-purpose computing with GPUs.
- [24] J. F. Yua, H.-C. Hsiao and Y.-J. Kao, Computers in Fluids (in press).