PHYSICAL REVIEW B **83**, 125104 (2011)

Exploiting translational invariance in matrix product state simulations of spin chains with periodic boundary conditions

B. Pirvu,[1] F. Verstraete,[1] and G. Vidal[2]
[1]*Fakultät für Physik, Universität Wien, Boltzmanngasse 5, A-1090 Wien, Austria*
[2]*School of Mathematics and Physics, The University of Queensland, Brisbane, Queensland 4072, Australia*
(Received 27 May 2010; published 17 March 2011)

We present a matrix product state (MPS) algorithm to approximate ground states of translationally invariant systems with periodic boundary conditions. For a fixed value of the bond dimension D of the MPS, we discuss how to minimize the computational cost to obtain a seemingly optimal MPS approximation to the ground state. In a chain with N sites and correlation length ξ, the computational cost formally scales as $g(D,\xi/N)D^3$, where $g(D,\xi/N)$ is a nontrivial function. For $\xi \ll N$, this scaling reduces to D^3, independent of the system size N, making our method N times faster than previous proposals. We apply the algorithm to obtain MPS approximations for the ground states of the critical quantum Ising and Heisenberg spin-1/2 models as well as for the noncritical Heisenberg spin-1 model. In the critical case, for any chain length N, we find a model-dependent bond dimension $D(N)$ above which the polynomial decay of correlations is faithfully reproduced throughout the entire system.

DOI: 10.1103/PhysRevB.83.125104 PACS number(s): 02.70.−c, 03.67.−a, 05.10.Cc, 75.10.Pq

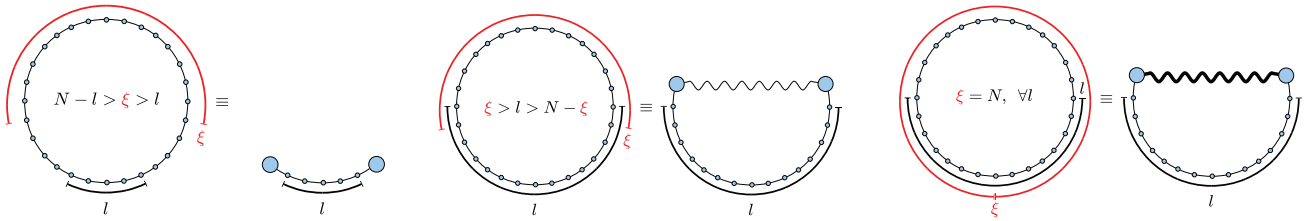
I. INTRODUCTION

Concepts of entanglement for many-body quantum systems have recently proven useful to devise new methods for the numerical simulation of quantum spin chains. It has been shown that the very successful density matrix renormalization group (DMRG) method[1] can be rephrased as a variational method over the class of matrix product states (MPS);[2–5] this realization clarified the relatively poor performance of DMRG for systems with periodic boundary conditions (PBC), as MPS with open boundary conditions (OBC) do not have the right entanglement structure. It was shown in Ref. 4 how this could be cured by using a MPS with PBC. However, due to the cyclic structure of the underlying MPS, the computational cost of the simulation in terms of the MPS bond dimension D grew from $O(D^3)$ to $O(D^5)$. This was subsequently lowered to $O(D^3)$ in Refs. 6 and 7.

An important motivation to study finite chains is that one can compute bulk properties of the system in the thermodynamic limit by extrapolating results obtained for increasingly large chains.[8] In this context, it is relevant whether OBC or PBC are considered. For a finite chain with OBC, local expectation values differ from those in the thermodynamic limit due both to finite-size effects and to boundary effects, and larger chains need to be considered. In contrast, with PBC only finite-size effects are present. This makes the extrapolation to the thermodynamic limit more transparent and smaller systems need to be simulated. Another important advantage of PBC is that only in this case can a finite chain be translationally invariant (TI).[10] This is a crucial feature for the present work, where translation invariance is exploited in order to reduce the computational costs of simulating finite chains.

Pippan, White, and Evertz[7] recently showed how to simulate spin chains with PBC with an MPS algorithm whose computational cost given in terms of D scales as $O(D^3)$. The intuition behind this scaling can be understood if one first considers systems with a correlation length ξ that is much shorter than the system size N. Let us choose a block of sites with size l such that $\xi > l$ [see Fig. 1(a)]. In this case correlations between the left and the right end of the block are mediated only through the sites inside the block. It is clear that the properties of this block are exactly the same as those of a block of equal length embedded in the bulk of a sufficiently large system with OBC. It is then not surprising that the cost for computing observables that are contained within such a block is proportional to D^3, as in the case of OBC. This is basically due to the fact that such calculations involve contracting a tensor network that has, as *uncorrelated* left and right boundary conditions, two *boundary vectors* with D^2 components each.[5] Now imagine we are interested in the description of properties contained in a larger block such that $\xi > l > N - \xi$ [see Fig. 1(b)]. This block is small enough for its ends to have correlations that are mediated via its own sites, yet large enough such that correlations are also mediated via the sites outside the block, since now $N - l < \xi$. If these externally mediated correlations are relatively small, the situation is not very different from the previously described case where $l < N - \xi$. All we have to do is to replace the two uncorrelated *boundary vectors* with a low-rank *boundary matrix* that contains the small amount of correlations. If the rank of the matrix is n, then the cost of this algorithm will be proportional to nD^3.

We emphasize two important aspects of the computational cost of the algorithm in Ref. 7. The first one is that the cost is also proportional to the system size N, due to the usual sweeping procedure that optimizes one site at each instant. We will show below how, in the case of a TI chain, one can basically get rid of this factor. It turns out that for chains where $\xi \ll N$ the cost will not depend on N at all. If $\xi \approx N$, on the other hand, the cost will contain a factor that is smaller than N but is nevertheless an artifact thereof. The reduction of the computational cost is achieved by using a TI MPS, where the N tensors of the MPS are chosen to be identical. For all D, the precision of our results is comparable to that reported in Ref. 7. This indicates that restricting the MPS ansatz to be TI does not lead to a loss of precision, while yielding a substantial reduction of the computational cost. The second important aspect is the multiplicative factor n corresponding to the rank of the boundary matrix that transfers correlations between the ends of a block. In the case where the correlation



(a) Medium ξ, small l: equivalent to OBC environment.

(b) Medium ξ, medium l: equivalent to partially correlated OBC environment.

(c) Large ξ, any l: equivalent to fully correlated OBC environment.

FIG. 1. (Color online) The properties of a block of size l within a PBC system can be equivalent to those of a block with the same size in the bulk of an OBC system. Depending on l and the correlation length ξ, the left and right boundary conditions of the OBC system are more or less correlated.

length ξ is of the order of the system size N [see Fig. 1(c)], this factor may not be small. In a worst-case scenario, where strong correlations between distant sites would force the boundary matrix to be full rank, i.e., $n = D^2$, the approach in Ref. 7 would not be better than the $O(D^5)$ algorithm of Ref. 4. Thus for critical systems where $\xi \approx N$ it is *a priori* unclear what the overall scaling of the computational cost in D will be. However, in Ref. 7 it has been indicated that if D is not too large, the ground-state energy of a critical spin chain obtained using a small constant n is satisfactory, in that its accuracy scales with D in a similar way as it would in an OBC chain of the same size.

Here we shall show how to exploit TI to obtain a faster algorithm that, for instance, does not scale with N when $\xi \ll N$. Nevertheless, except for the case $\xi \ll N$, we still lack a precise characterization of how the cost scales as a function of D and N. We benchmark the present approach by addressing both critical (i.e., $\xi \approx N$) and noncritical (i.e., $\xi \ll N$) chains. An important observation is that in the case of critical systems the finite bond dimension D of the MPS introduces an effective correlation length $\tilde{\xi}_D \approx D^\kappa$ (Refs. 15–18) that depending on D can be much smaller than the actual one. This implies that as N grows, a larger bond dimension $D \approx N^{1/\kappa}$ needs to be considered if correlations between distant sites of the chain with PBC are to be properly captured. Our numerical results are consistent with a complex scenario where the cost of the simulations is dominated by the crossover between finite-N and finite-D corrections, as further discussed in Ref. 19.

The rest of the paper is structured as follows: We start by sketching the main idea of the approach in Sec. II, followed by an in-depth presentation of the algorithm in Sec. III. In Sec. IV we present numerical results for the critical quantum Ising and Heisenberg spin-1/2 models as well as for the noncritical Heisenberg spin-1 model. Finally Sec. V contains some conclusions.

II. OVERVIEW

This work is concerned with the approximation of ground states (GSs) within the variational class of MPS with PBC defined in Ref. 4. Since critical systems are arguably among the most challenging ones from a computational perspective, we will focus our attention on these; noncritical systems will be treated as a simpler special case of these. An important restriction is that we will only consider TI systems, which we will simulate using a TI MPS ansatz, namely, an MPS where the tensors corresponding to different sites are all equal. The resulting variational class is a subclass of the one defined in Ref. 4. The TI MPS with PBC reads

$$|\psi(A_i)\rangle = \sum_{i_1,\ldots,i_N=1}^{d} \mathrm{Tr}\left(A_{i_1} A_{i_2} \cdots A_{i_N}\right) |i_1 i_2 \cdots i_N\rangle \qquad (1)$$

with identical matrices A_i at every site. The basis states are tensor products of the one-site computational basis states, i.e., $|i_1 i_2 \cdots i_N\rangle = |i_1\rangle \otimes |i_2\rangle \otimes \ldots \otimes |i_N\rangle$ where $|i\rangle \in \{|\uparrow\rangle, |\downarrow\rangle\}$ for a spin-1/2 chain. Note that for fixed i each A_i represents a matrix; thus the MPS is completely characterized by the three-dimensional tensor ${A_i}^\alpha{}_\beta =: \mathbf{A}$. The components of $\mathbf{A}$ are then the variational parameters in our ansatz. Furthermore, we would like to point out that we will mostly be interested in Hamiltonians that are real and reflection invariant; these symmetries can be implemented at the level of the MPS by choosing the matrices A_i real and symmetric. This extra constraint does not seem to deteriorate the accuracy of the variational procedure.

Since our ansatz consists of N copies of the same tensor, the energy is not a quadratic expression in the variables defined by the tensors A_i; this implies that we cannot use the sweeping procedure described in Ref. 4 or any other procedure that lowers the energy by minimizing it for one site at a time. While this might seem a reason to be concerned at first, it will actually be the key to reducing computational costs.

The advantages of a TI MPS ansatz (with periodicity 1 or 2) have already been exploited in the context of infinitely long chains.[1,9,11–14] References 1, 9, and 12 used a TI MPS in the context of infinite-system DMRG. In Ref. 11 instead, a (two-site periodic) MPS approximation to ground states was obtained by imaginary-time evolution. References 13 and 14 discussed how to compute ground states with a one-site TI MPS when the imaginary-time evolution operator can be well enough approximated by layers of one-site TI matrix product operators. An attempt to adapt that method to finite chains with PBC yielded results that are not as accurate as one might expect. This is basically due to the fact that the bond dimension truncation method used in Refs. 13 and 14 can be shown to be optimal only for infinitely long chains. We have used a straightforward adaptation of that method for finite chains with PBC and the results are between one and a few orders of magnitude worse than the ones obtained by the gradient

method described in the present work. Finally, we also point out that a TI MPS with PBC was already used in Ref. 6 together with Monte Carlo sampling techniques, with a formal cost $O(ND^3)$. In that case, the use of sampling techniques reduced the cost from $O(D^5)$ to $O(D^3)$, but at the same time enforced the multiplicative factor N, since a TI MPS does not represent a TI state once a given configuration is chosen during the sampling.

An obvious way to find the TI MPS with minimal energy is a multidimensional minimization procedure that requires only evaluations of the energy function itself, such as the downhill simplex method.[20] When no further information about the function is available, this is indeed the method of choice. It is extremely robust but also extremely slow. However, if there is a feasible way to obtain more elaborate information such as the gradient or the Hessian, there are methods relying on these quantities that are clearly superior as regards the speed of convergence and the required storage space.

In the following we will present an efficient algorithm to calculate the gradient of the energy $\nabla E(\mathbf{a})$ where the argument $\mathbf{a} = \mathrm{vec}(\mathbf{A})$ denotes the vector containing all entries of the MPS tensor $\mathbf{A}$. The result will then be used by a standard numerical library conjugate gradient algorithm to find a minimum of $E(\mathbf{a})$. We must emphasize that this minimum is by no means guaranteed to be the global one, i.e., the optimal ground-state approximation within the subspace defined by our special MPS ansatz. However, our numerical results seem to be slightly more accurate than previous results,[7] while we have obtained a reduction in computational costs. We will illustrate the accuracy of this approach by applying it to two exactly solvable models in order to give exact values for the numerical errors.

The computational cost will turn out to scale as $O(mnD^3) + O(n^2D^3)$ where D is the virtual bond dimension and m and n are some parameters to be specified below. Briefly speaking, the scaling can be understood as follows: First we approximate large powers of the MPS transfer matrix, whose exact definition will be given later in the text, within a reduced subspace of dimension n. Treating each of the n dimensions separately allows us to transform the contraction of a tensor network with PBC [which scales as $O(D^5)$] into n contractions of tensor networks with OBC [each of which scales as $O(D^3)$]. As we will explain in more detail in the next section, the resulting tensor networks will still contain at most one portion represented by say m adjacent transfer matrices that is not directly connected to the already approximated one. If m is large, this second portion can again be approximated within an n-dimensional subspace, thereby yielding the scaling $O(n^2D^3)$. If m is small, we are forced to contract the transfer matrices one after the other, which gives the scaling $O(mnD^3)$.

III. THE ALGORITHM

Let us rearrange the MPS tensor components in a vector $\mathbf{a} = \mathrm{vec}(\mathbf{A})$ which allows us to write the energy as a function over the manifold of free parameters in the MPS:

$$E(\mathbf{a}) = \frac{\langle\psi(\mathbf{a})|\, H\, |\psi(\mathbf{a})\rangle}{\langle\psi(\mathbf{a})|\psi(\mathbf{a})\rangle} \equiv \frac{\langle\psi(\mathbf{A})|\, H\, |\psi(\mathbf{A})\rangle}{\langle\psi(\mathbf{A})|\psi(\mathbf{A})\rangle}. \tag{2}$$

Note that, due to the constraints that the matrices are real and symmetric, the number of vector components in $\mathbf{a}$ has been reduced to $\frac{1}{2}dD(D+1)$. Furthermore we will denote expectation values taken with respect to the MPS defined by the tensor $\mathbf{A}$ as $\langle O\rangle_{\mathbf{A}} := \langle\psi(\mathbf{A})|\, O\, |\psi(\mathbf{A})\rangle$.

Note that (2) can have local extrema as opposed to $E(\Psi) = \langle\Psi|\, H\, |\Psi\rangle$ where $|\Psi\rangle$ is an arbitrary, normalized vector in the exponentially large Hilbert space. The MPS parametrization restricts the full parameter space to a submanifold, thus possibly generating local extrema where all derivatives in this subspace vanish. If one uses as a starting point of the conjugate gradient algorithm a random vector $\mathbf{a}_{\mathrm{rand}}$, the search algorithm will typically get stuck in a local minimum. In order to avoid getting stuck in one of these, we will choose as a starting point a vector $\mathbf{a}_0$ of which we can be sure that it is close to the global minimum. This approach turns out to be very robust and fast. If we are interested in ground states of chains with very large N, the most natural choice for the starting vector is an MPS approximation of the GS of the same model in the thermodynamic limit. Note that this MPS must have exactly the same symmetry properties as our ansatz. It was shown in previous work[14] how to obtain this MPS and we will actually use the tensors computed there as starting points for the present algorithm. It is obvious why the MPS for the GS of the infinite chain is a good choice if one is interested in finite PBC chains with $N \gg \tilde{\xi}_D$, where $\tilde{\xi}_D$ is the correlation length induced by finite D. However, it turns out that this approach also works satisfactorily for moderately large N. Of course, if there already is any PBC solution available, using that one as a starting point may provide a gain in convergence time, especially if the chain lengths are similar.

The gradient $\nabla E(\mathbf{a})$ reads explicitly

$$\begin{aligned}\nabla E(\mathbf{a}) = {} & \frac{1}{\langle\psi(\mathbf{a})|\psi(\mathbf{a})\rangle}\nabla\, \langle\psi(\mathbf{a})|\, H\, |\psi(\mathbf{a})\rangle \\ & - \frac{\langle\psi(\mathbf{a})|\, H\, |\psi(\mathbf{a})\rangle}{\langle\psi(\mathbf{a})|\psi(\mathbf{a})\rangle^2}\nabla\, \langle\psi(\mathbf{a})|\psi(\mathbf{a})\rangle\,. \end{aligned} \tag{3}$$

It turns out that this quantity can be computed efficiently. First, since we assume a translationally invariant Hamiltonian with nearest-neighbor interactions, we have

$$\langle H\rangle_{\mathbf{A}} = \langle H_{N,1}\rangle_{\mathbf{A}} + \sum_{s=1}^{N-1}\langle H_{s,s+1}\rangle_{\mathbf{A}} = N\langle H_{s,s+1}\rangle_{\mathbf{A}}.$$

Hence the first term in (3) is proportional to the gradient of the energy density $\rho_E(\mathbf{a}) = \langle H_{s,s+1}\rangle_{\mathbf{A}}$, $\forall\, s \in [1,N]$ [see Fig. 2(d)]. Second, we can obtain gradients such as the ones occurring in (3) numerically at a given point $\mathbf{a}$ by expanding the differentiated quantity in powers of $\delta\mathbf{a}$ and computing the coefficient of the linear term. Thus the derivative in the first term is obtained via

$$\rho_E(\mathbf{a}+\delta\mathbf{a}) = \rho_E(\mathbf{a}) + \delta\mathbf{a}[\nabla_{\mathbf{a}'}\rho_E(\mathbf{a}')]_{\mathbf{a}'=\mathbf{a}} + O(\delta\mathbf{a}^2) \tag{4}$$

and the one in the second via

$$\begin{aligned}&\langle\psi(\mathbf{a}+\delta\mathbf{a})|\psi(\mathbf{a}+\delta\mathbf{a})\rangle \\ &\quad = \langle\psi(\mathbf{a})|\psi(\mathbf{a})\rangle + \delta\mathbf{a}[\nabla_{\mathbf{a}'}\,\langle\psi(\mathbf{a}')|\psi(\mathbf{a}')\rangle]_{\mathbf{a}'=\mathbf{a}} + O(\delta\mathbf{a}^2).\end{aligned} \tag{5}$$

Let us first consider (4) which can be computed explicitly by taking a sum of completely contracted tensor networks [see Fig. 2(e)]. Let $H_{\mathrm{eff}}(\mathbf{A})$ denote the object that is obtained

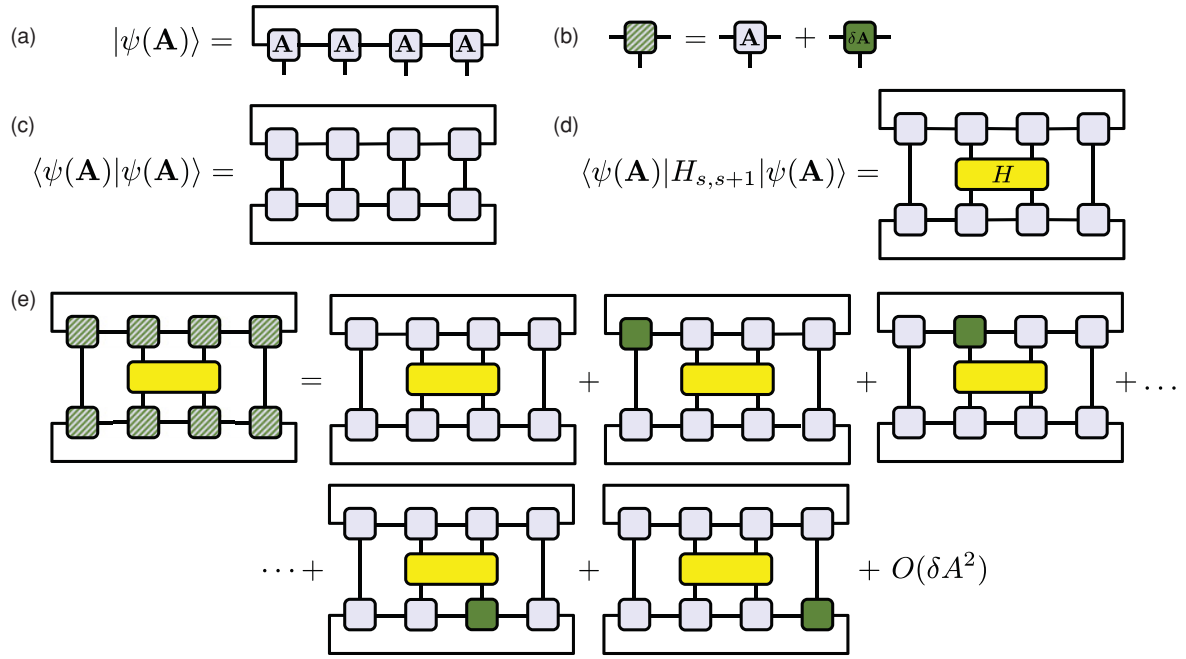


FIG. 2. (Color online) (a) Graphical representation of the TI PBC MPS $|\psi(\mathbf{A})\rangle$ of a TI spin chain with four sites. Note the identical tensors $\mathbf{A}$ at each site. (b) Small perturbation $\delta\mathbf{A}$ is added to the MPS tensor $\mathbf{A}$. (c) Norm of a state $\langle\psi(\mathbf{A})|\psi(\mathbf{A})\rangle$. (d) Expectation value of a two-site operator, e.g., $\langle\psi(\mathbf{A})|H_{s,s+1}|\psi(\mathbf{A})\rangle$. (e) The expectation value is expanded in powers of $\delta\mathbf{A}$.

by removing the tensor $\delta\mathbf{A}$ from each term of $\rho_E(\mathbf{a}+\delta\mathbf{a})$ that is linear in $\delta\mathbf{a}$ (see Fig. 3). This is a tensor with three indices that, reshaped in vector form, yields the desired derivative $\nabla\rho_E(\mathbf{a}) = \mathrm{vec}[H_{\mathrm{eff}}(\mathbf{A}))]$. The computational cost for the *exact* contraction of the tensor networks in $H_{\mathrm{eff}}(\mathbf{A})$ scales as $O(ND^5)$.[4] We will give below a prescription of how this can be improved to $O(mnD^3) + O(n^2D^3)$ by making an ansatz for the approximation of the exact result that depends on two integer parameters m and n. Subsequently we will show how to choose the smallest possible m and n such that no loss in precision occurs and why the scaling reduces to $O(mD^3) + O(nD^3)$ in the case of very long chains.

The other piece that is necessary for the computation of $\nabla E(\mathbf{a})$ is the derivative occurring in the second term of (3); this term can be obtained in a very similar way (see Fig. 3). We will use the notation $N_{\mathrm{eff}}(\mathbf{A})$ for the object defined by $\nabla\langle\psi(\mathbf{a})|\psi(\mathbf{a})\rangle =: \mathrm{vec}[N_{\mathrm{eff}}(\mathbf{A})]$. Due to the simpler structure of the tensor network the computational cost here will scale as $O(nD^3)$ for arbitrary chains and as $O(D^3)$ for very long chains.

Now let us introduce the following convention for denoting incomplete tensor networks where merely one of the MPS tensors is missing: $\langle O\rangle_{\mathbf{A}}^{[\bar{s}]}$ shall henceforth denote the expectation value of the operator O with respect to the TI MPS defined by the tensor $\mathbf{A}$, where one tensor $\mathbf{A}$ has been removed from $|\psi(\mathbf{A})\rangle$ at site s. Following this definition, the first term in the graphical representation of $H_{\mathrm{eff}}(\mathbf{A})$ (see Fig. 3) reads $\langle H_{2,3}\rangle_{\mathbf{A}}^{[\bar{1}]}$. If a tensor has been removed from $\langle\psi(\mathbf{A})|$ at site s, we will denote this by underlining the site index; thus we write $\langle O\rangle_{\mathbf{A}}^{[\underline{s}]}$. Using this convention we can write $H_{\mathrm{eff}}(\mathbf{A})$ as

$$H_{\mathrm{eff}}(\mathbf{A}) = \sum_{s=1}^{N}\left(\langle H_{1,2}\rangle_{\mathbf{A}}^{[\bar{s}]} + \langle H_{1,2}\rangle_{\mathbf{A}}^{[\underline{s}]}\right). \tag{6}$$

For real Hamiltonians and real MPS this reduces of course to

$$H_{\mathrm{eff}}(\mathbf{A}) = 2\sum_{s=1}^{N}\langle H_{1,2}\rangle_{\mathbf{A}}^{[\bar{s}]}. \tag{7}$$

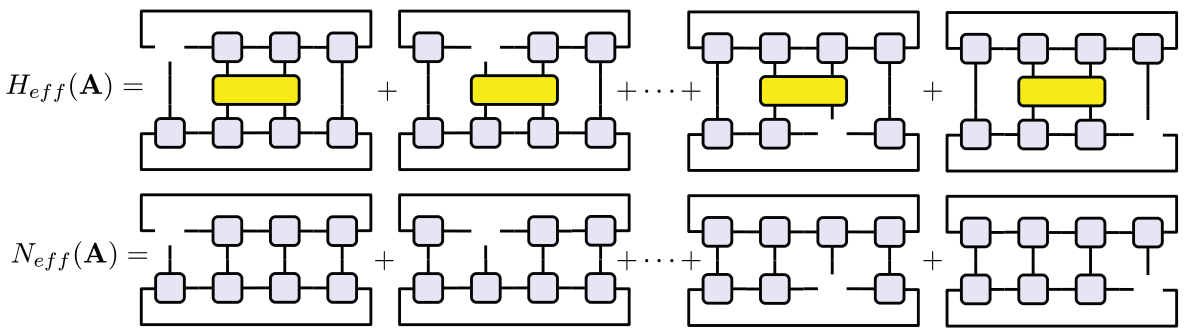


FIG. 3. (Color online) Graphical representation of the tensor $H_{\mathrm{eff}}(\mathbf{A})$ and $N_{\mathrm{eff}}(\mathbf{A})$.

Similar considerations hold for $N_{\mathrm{eff}}(\mathbf{A})$. Thus, using I to denote the identity operator, we can rewrite the gradient of the energy (3) as

$$\nabla E(\mathbf{a}) = \frac{N H_{\mathrm{eff}}(\mathbf{A})}{\langle\psi(\mathbf{A})|\psi(\mathbf{A})\rangle} - \frac{N\rho_E(\mathbf{a})N_{\mathrm{eff}}(\mathbf{A})}{\langle\psi(\mathbf{A})|\psi(\mathbf{A})\rangle^2} = 2N\sum_{s=1}^{N}\left(\frac{\langle H_{1,2}\rangle_{\mathbf{A}}^{[\bar{s}]}}{\langle I\rangle_{\mathbf{A}}} - \frac{\langle H_{1,2}\rangle_{\mathbf{A}}\langle I\rangle_{\mathbf{A}}^{[\bar{s}]}}{\langle I\rangle_{\mathbf{A}}^2}\right) \tag{8}$$

Let us make a brief excursion to sketch how a gradient-based procedure can be employed to find ground states of PBC chains if one is dealing with *complex* Hamiltonians and thereby *complex* MPS. One possibility is to use a gradient-based algorithm that converges to a minimum of the real-valued function $E:\mathbb{C}^n\to\mathbb{R}$ within the complex manifold (n stands here for the number of independent complex parameters in the MPS). It can be shown that in this case one obtains the same expression (8) for the gradient of the energy albeit the individual terms are now complex-valued vectors. However, since standard library routines for gradient-based search cannot minimize over complex manifolds, let us mention the second possibility just for the sake of completeness. Because $\mathbf{a}=\mathbf{x}+i\mathbf{y}$ with $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$, one can treat the energy as an analytic function over a real manifold with twice as many degrees of freedom, i.e., $E:\mathbb{R}^{2n}\to\mathbb{R}$. Similar considerations to the ones leading to (8) yield then for the gradient

$$\begin{aligned}\nabla_{\mathbf{x}}E(\mathbf{x},\mathbf{y}) &= 2N\sum_{s=1}^{N}\left(\frac{\mathrm{Re}\langle H_{1,2}\rangle_{\mathbf{A}}^{[\bar{s}]}}{\langle I\rangle_{\mathbf{A}}} - \frac{\langle H_{1,2}\rangle_{\mathbf{A}}\mathrm{Re}\langle I\rangle_{\mathbf{A}}^{[\bar{s}]}}{\langle I\rangle_{\mathbf{A}}^2}\right),\\ \nabla_{\mathbf{y}}E(\mathbf{x},\mathbf{y}) &= -2N\sum_{s=1}^{N}\left(\frac{\mathrm{Im}\langle H_{1,2}\rangle_{\mathbf{A}}^{[\bar{s}]}}{\langle I\rangle_{\mathbf{A}}} - \frac{\langle H_{1,2}\rangle_{\mathbf{A}}\mathrm{Im}\langle I\rangle_{\mathbf{A}}^{[\bar{s}]}}{\langle I\rangle_{\mathbf{A}}^2}\right).\end{aligned} \tag{9}$$

Returning to real MPS, we still have to show how $H_{\mathrm{eff}}(\mathbf{A})$ and $N_{\mathrm{eff}}(\mathbf{A})$ can be computed efficiently. Since this is somewhat technical we will put the details into the Appendix and give here only a brief overview of the algorithm. First note that for big chains (i.e., large N) tensor networks of the form shown in Fig. 3 will contain big powers of the transfer matrix

$$T=\sum_{i=1}^{d}A_i\otimes A_i. \tag{10}$$

The main idea is to exploit the fact that the eigenvalues of this $D^2\times D^2$ transfer matrix usually decay rapidly enough such that we can approximate big powers of T within a low-dimensional subspace spanned by its dominant eigenvectors. We will denote the dimension of this subspace by n. It is important, however, to realize that also small powers of T occur and that these may eventually require exact contraction. Let us denote the largest power of the transfer matrix that we contract exactly by m. As shown in the Appendix, it turns out that in general it is possible to perform the contractions in such a way that the overall computational cost for $H_{\mathrm{eff}}(\mathbf{A})$ scales like $O(mnD^3)+O(n^2D^3)$ and that for $N_{\mathrm{eff}}(\mathbf{A})$ like $O(nD^3)$. It is easy to see that the scalar expectation values in (3) can be obtained in an analogous yet simpler way. The fact that there are no vacant sites in the corresponding tensor networks enables us to use in that case a contraction method that is similar to the one used for $N_{\mathrm{eff}}(\mathbf{A})$. Thus the computational cost for our algorithm scales as its most expensive part, namely, as $O(mnD^3)+O(n^2D^3)$.

It is also not difficult to check that for very large chains (i.e., either when $N\gg\xi$ for noncritical systems or $N\gg\tilde{\xi}_D$ for critical ones, where $\tilde{\xi}_D$ is the effective correlation length induced by finite D) this scaling can be improved. First note that, as shown in the Appendix, we have in every tensor network at least one portion of the chain expressed as a power of T that we approximate using its dominant eigenvectors. Now, for any bond dimension D there exists an N above which all approximated portions are long enough such that all eigenvalues except the largest one are suppressed by the very large exponent. In this case the overall scaling is $O(mD^3)+O(nD^3)$. Note that in the scaling for the "extremal-s" (Appendix A1) terms we cannot get rid of m because there will always be short portions between the $H_{1,2}$ and the vacant site that must be contracted exactly. Similarly, for the "medium-s" terms (Appendix A2) only the combinations of $\lambda_\alpha^m\lambda_\beta^m$ where both α and β are large will be negligible. Factors like $\lambda_1^m\lambda_\beta^m$ must usually always be taken into account. In any case, the ultimate check of whether our approximations are justified must be done in the simulations, where one must verify if there exists an n beyond which our approximated ground-state energy does not decrease.

We would like to compare our scaling of the computational cost to that of Ref. 7 once again. Note that, expressed in the terms used in this work, the scaling from Ref. 7 is $O(NnD^3)$. On one hand, as previously mentioned, our TI algorithm yields an improvement of one factor N. On the other hand there is an additional factor n that appears in our scaling. This is due to the fact that we compute the gradient of the energy explicitly. It is easy to see that the computational cost for the evaluation of the energy itself is $O(nD^3)$. However, if we were to restrict ourselves to evaluations of the energy only, we would have to use something like a downhill simplex method as the outer function that scans the MPS manifold for the energy minimum. In this case the outer function would call the energy evaluator a huge number of times, thereby yielding an overall cost much higher than the one factor of n that we must pay when computing the gradient.

IV. NUMERICAL RESULTS

We have studied both critical and noncritical nearest-neighbor interaction spin models. The first one is the quantum Ising model for spins 1/2,

$$H_{IS}=-\sum_{i=1}^{N}\sigma_i^z\sigma_{i+1}^z - B\sum_i\sigma_i^x, \tag{11}$$

which we have simulated at its critical point $B=1$. The periodic boundary conditions are implemented as usual by identifying σ_{N+1}^α with σ_1^α. The second one is the antiferromagnetic Heisenberg model

$$H_{HB}=\frac{1}{2}\sum_{i=1}^{N}\left(\sigma_i^x\sigma_{i+1}^x+\sigma_i^y\sigma_{i+1}^y+\sigma_i^z\sigma_{i+1}^z\right). \tag{12}$$

This model is critical for spin-1/2 chains but noncritical for spin-1 chains. We have studied both cases. Note that (12) is not very well suited for description with one-site TI MPS

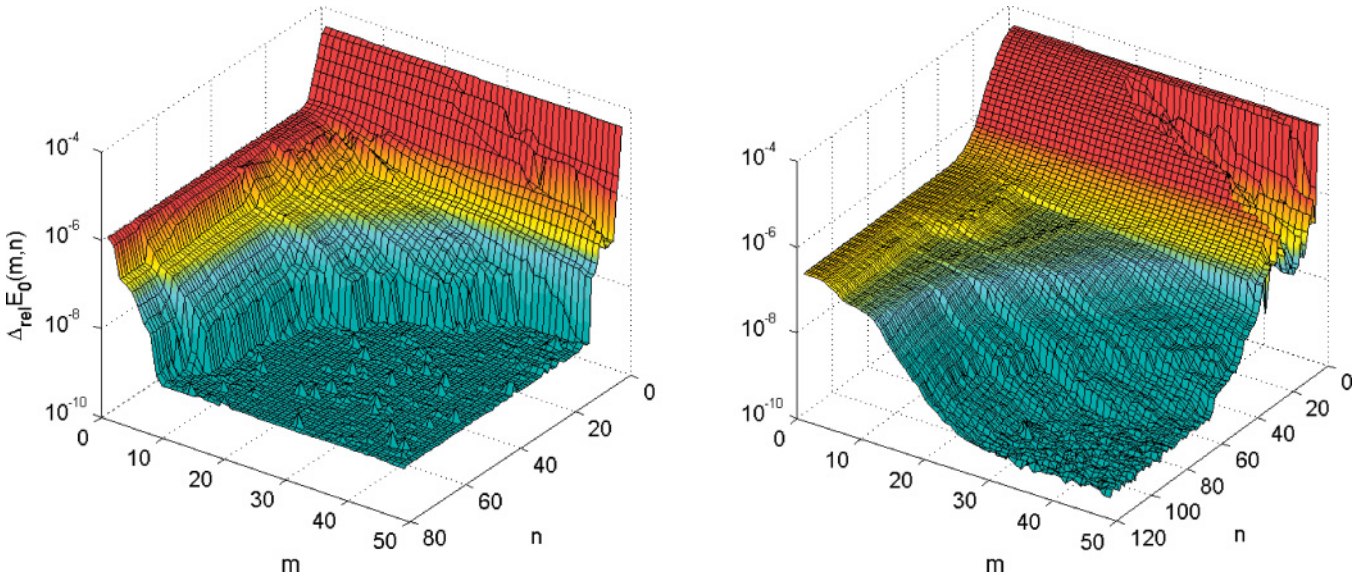


FIG. 4. (Color online) Critical quantum Ising chain with $N = 100$: Relative precision of the MPS ground-state energy as compared to the analytical result as a function of the parameters (m,n) for $D = 16$ (left) and 32 (right).

due to its antiferromagnetic character. In order to cure this problem we apply in the case of the spin-1/2 chain a unitary transformation consisting of Pauli σ^y matrices on each second site.[21] This leaves the spectrum unchanged and after we have found the 1-site TI MPS for the ground state, we can recover the one for the unchanged Hamiltonian by a new application of a unitary transformation. The resulting MPS is then of course two-site TI. The rotated Heisenberg Hamiltonian reads

$$H_{HB} = \frac{1}{2}\sum_{i=1}^{N}\left(-\sigma_i^x\sigma_{i+1}^x + \sigma_i^y\sigma_{i+1}^y - \sigma_i^z\sigma_{i+1}^z\right). \quad (13)$$

A. Critical systems

Let us illustrate the strategy for the scan of the parameter space spanned by $\{m,n\}$ on the basis of results obtained for small critical chains of 100 and 400 sites. Figures 4 and 5 show the relative precision $\Delta_{\text{rel}}E_0(m,n) = [E_0^{\text{exact}} - E_0^{\text{MPS}}(m,n)]/E_0^{\text{exact}}$ of the MPS ground-state energy compared to the exact solution as a function of the algorithm parameters m and n for the quantum Ising and Heisenberg chains, respectively. The first observation is that there exist $m_{\max}$ and $n_{\max}$ such that for all $m \geqslant m_{\max}$, $n \geqslant n_{\max}$ the precision does not improve any more. In the featured plots the plateau $\mathcal{P}$ with minimal energy is reached within the plot range. The

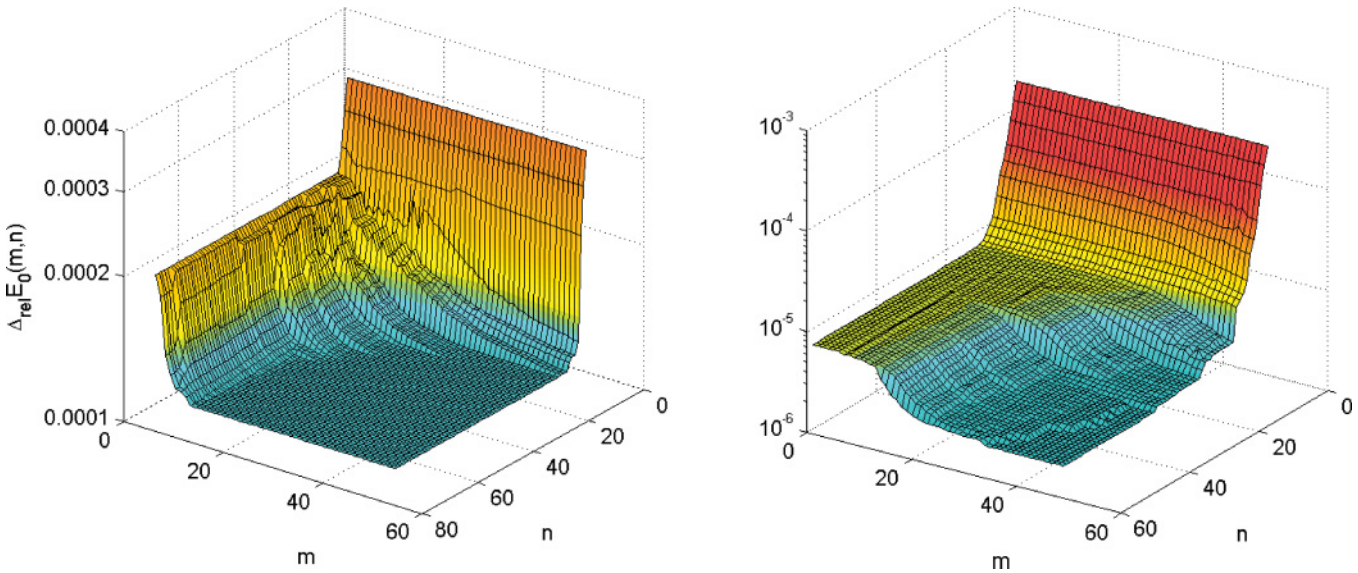


FIG. 5. (Color online) Critical Heisenberg chain with $N = 100$: Relative precision of the MPS ground-state energy as compared to the analytical result as a function of the parameters (m,n) for $D = 16$ (left) and 32 (right).

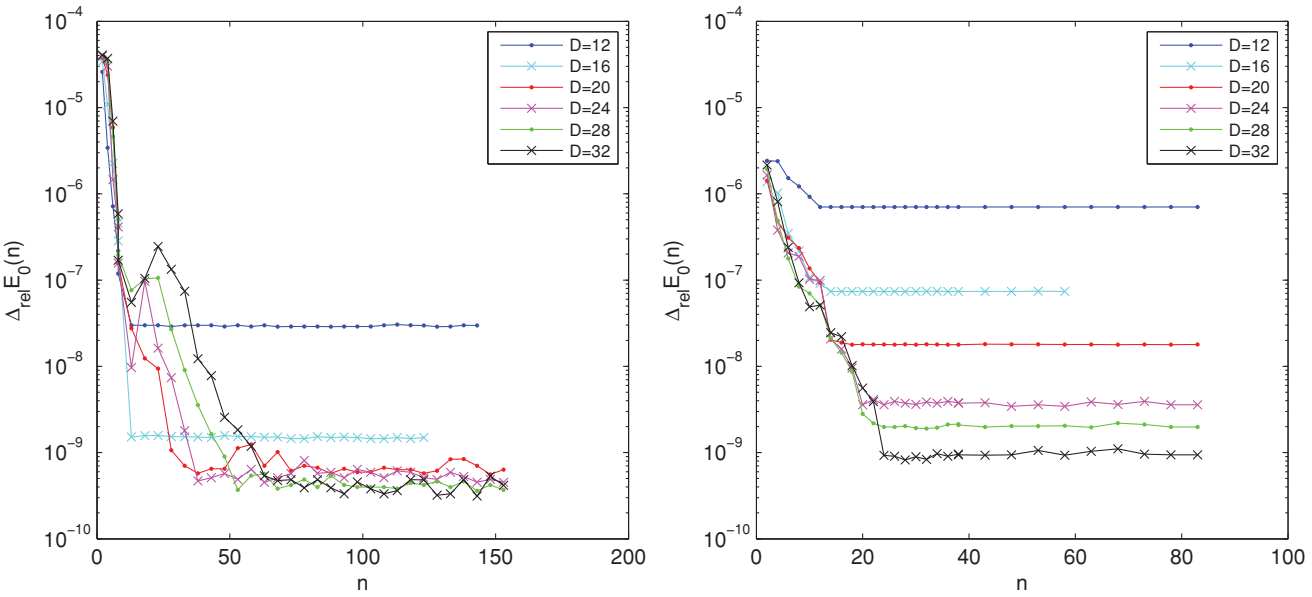


FIG. 6. (Color online) Critical quantum Ising chain with $N = 100$ (left) and 400 (right): Relative precision of the MPS ground-state energy as a function of the parameter n for different bond dimensions D. The scan has been performed along the line $m = 5n$ up to the maximal value of m and then along the line with constant $m = (N - 2)/2$.

optimal point $\{m_{\text{opt}},n_{\text{opt}}\}$ is then the point of $\mathcal{P}$ that minimizes the scaling of the computational cost $O(mnD^3) + O(n^2D^3)$, i.e., $\{m_{\text{opt}},n_{\text{opt}}\} = \min|_{\{m,n\}\in\mathcal{P}}(mn + n^2)$. Clearly, the optimal parameters m_{opt} and n_{opt} will be different for different models and different values of the chain length N and the MPS bond dimension D.

The plots reveal a further detail: if we are not very pedantic about the optimal $\{m,n\}$ pair, it is not necessary to scan the entire plane, which is computationally very expensive. If we are willing to settle for any pair $\{m,n\}$ that yields maximal precision, we can scan along any line $n = km$ and we can be sure that at some point we will hit $\mathcal{P}$. This pair is quasioptimal in the sense that we have found the optimal n for the corresponding m and vice versa. This is due to the fact that for any point of $\mathcal{P}$, especially for its boundary, walking along lines with increasing m or n does not take us out of $\mathcal{P}$. As one can see in Figs. 4 and 5, $\mathcal{P}$ is roughly symmetric in m and n, so a sensible line to scan along would be given by $n = m$. In practice it might be better to choose $k < 1$ since there are parts of the algorithm with the scaling $O(nD^3)$ multiplied by a big constant factor. In our simulations we have used $k = 1/5$. As we have explained in the Appendix, our algorithm allows us to increase m only up to $(N - 2)/2$. If until then the results obtained along $n = m$ have not converged yet, we must continue the scan along the line given by the constant maximal m toward larger n.

The relative precision of the MPS ground-state energy for such line scans is plotted in Fig. 6. We notice that with increasing D the maximally reachable precision gets better in concordance with what one would expect. The fact that m_{opt} and n_{opt} increase with D is also intuitive. What is a bit surprising is that for small n the results obtained for small bond dimensions are either similar to or even better than the ones obtained for higher bond dimensions. This means that if one is not willing to go to larger values of n, there is no point in increasing D.

Another interesting point is that for fixed D, as we increase N, the plateau $\mathcal{P}$ is reached sooner and sooner (i.e., for smaller

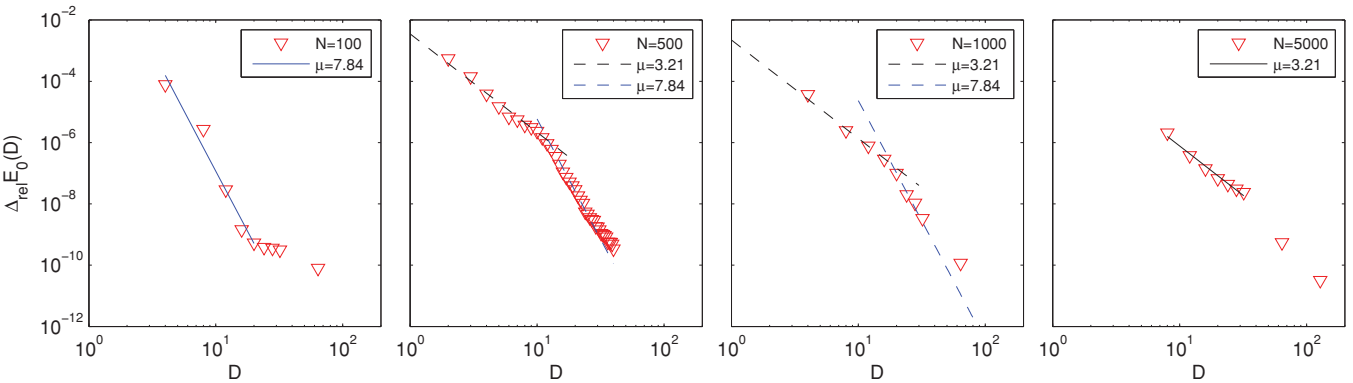


FIG. 7. (Color online) Critical quantum Ising model: relative precision of the MPS ground-state energy for different N as a function of D.

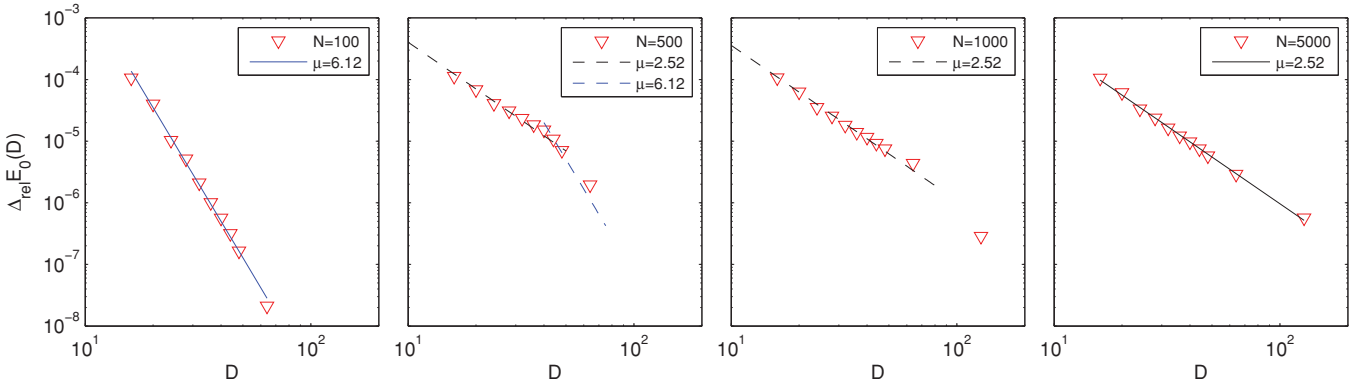


FIG. 8. (Color online) Critical Heisenberg model: relative precision of the MPS ground-state energy for different N as a function of D.

values of n and implicitly of m). This behavior is due to the fact that with increasing N the weight that we lose in our contracted tensor network by choosing $n < D^2$ becomes negligible at smaller n.

B. Observables: Energy and correlation functions

As the computational cost of our algorithm actually decreases if we increase the number of sites N while keeping D constant, we can investigate PBC chains of arbitrary size. Nevertheless, if we want to maintain the same level of precision for very long chains, we must also increase D while going up in N. Figures 7 and 8 show the relative precision of the ground-state energy for the critical quantum Ising and Heisenberg models, respectively, as a function of the MPS bond dimension D. We can see that generally the relative error is decreasing as a polynomial of D, i.e., $\Delta_{\text{rel}} E_0(D) \propto D^{-\mu}$. We have fitted straight lines through the reliable[22] data of the $N = 100$ and 5000 plots and have obtained for the exponent μ the values 7.84 and 3.21 (6.12 and 2.52) for the critical quantum Ising (Heisenberg) model. In the central plots (i.e., $N = 500$ and 1000) one can distinguish between two regions where the relative precision is decaying polynomially with the exponents obtained from the outer plots (i.e., $N = 100$ and 5000). We have emphasized this by drawing dashed lines through the data points in the central plots. Note that the dashed lines are not fitted, they have merely the same slope as the full lines in the outer plots. This behavior can be best understood if one looks at correlation functions.

Let us first consider the critical quantum Ising model. In Fig. 9 we have plotted the ZZ and the XX correlation functions, i.e.,

$$\begin{aligned}\Gamma^{ZZ}(\Delta r) &= \langle \sigma_r^z \sigma_{r+\Delta r}^z \rangle - \langle \sigma_r^z \rangle \langle \sigma_{r+\Delta r}^z \rangle, \\ \Gamma^{XX}(\Delta r) &= \langle \sigma_r^x \sigma_{r+\Delta r}^x \rangle - \langle \sigma_r^x \rangle \langle \sigma_{r+\Delta r}^x \rangle,\end{aligned} \tag{14}$$

in the MPS ground state of a chain with $N = 500$ sites. The solid line represents the exact solution obtained by applying the program of Ref. 23 to the quantum Ising model with PBC. One can clearly see that with increasing D the MPS correlations

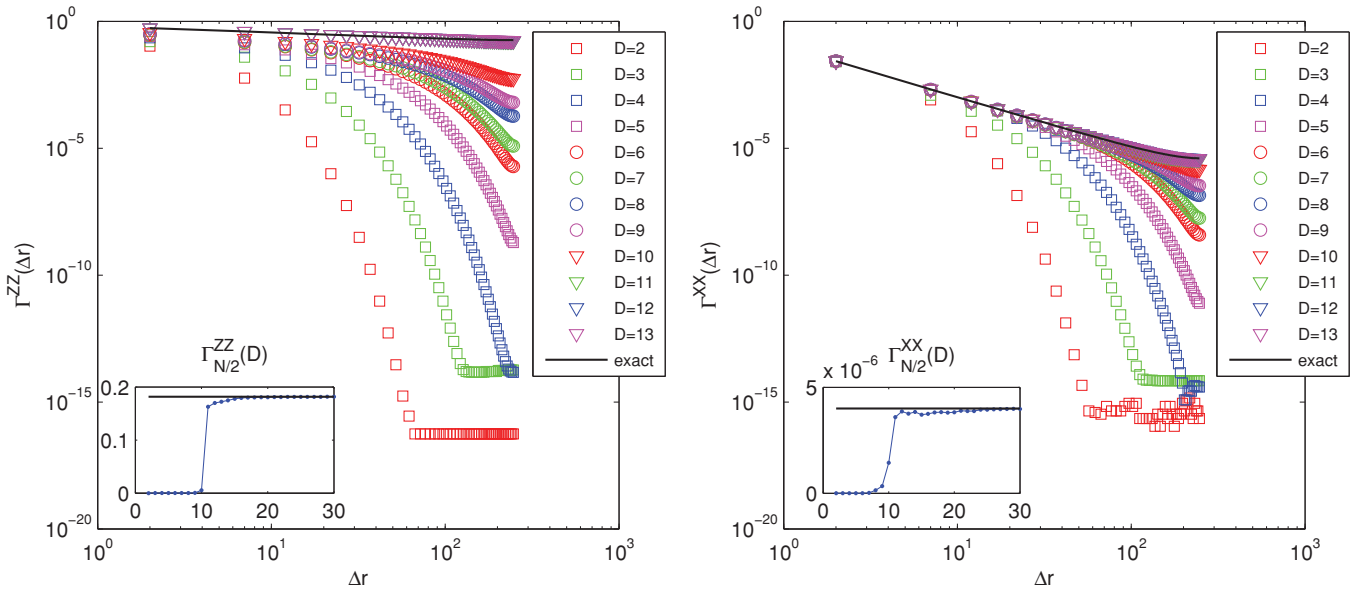


FIG. 9. (Color online) Correlation functions for a critical quantum Ising chain with $N = 500$. Left: order parameter correlator $\Gamma^{ZZ}(\Delta r)$ and as inset the half-chain correlator as a function of D. Right: correlator $\Gamma^{XX}(\Delta r)$ and as inset the half-chain correlator as a function of D.

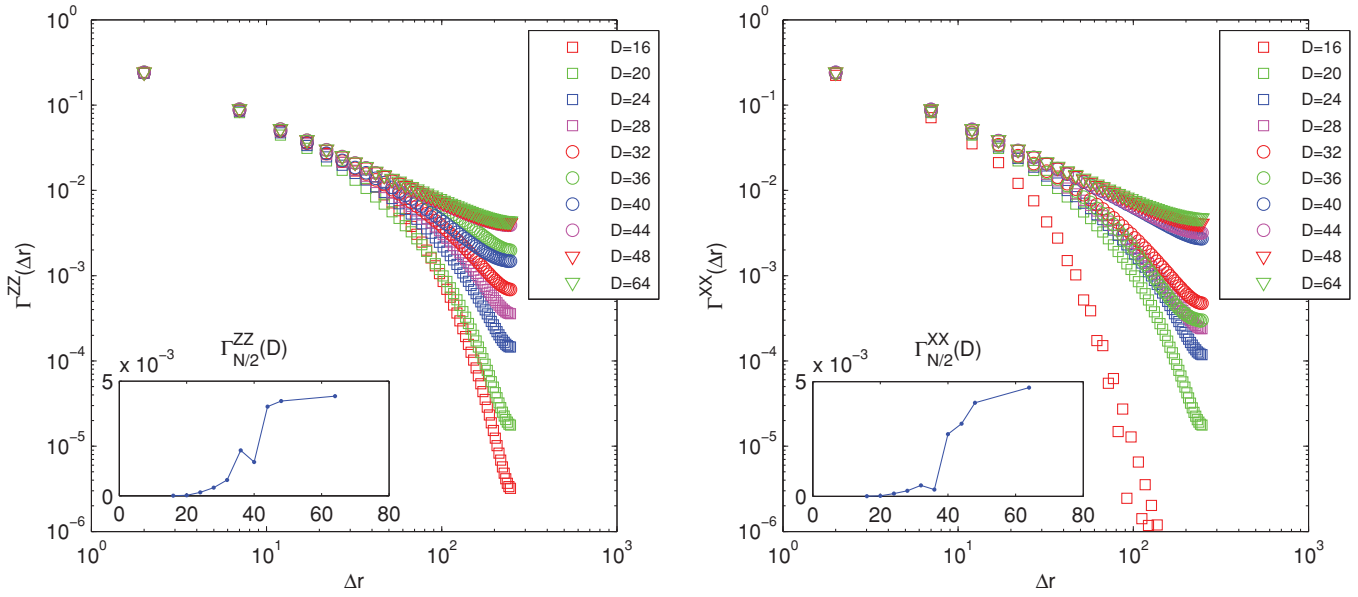


FIG. 10. (Color online) Absolute values of the correlation functions for a critical Heisenberg chain with $N = 500$. Left: correlator $\Gamma^{ZZ}(\Delta r)$ and as inset the half-chain correlator as a function of D. Right: correlator $\Gamma^{XX}(\Delta r)$ and as inset the half-chain correlator as a function of D.

become more and more accurate, just as one would expect. Note that we have only plotted the correlation functions for separations $\Delta r \leqslant N/2$. This is because due to the periodic boundary conditions $\Gamma(\Delta r)$ is symmetric around $\Delta r = N/2$, i.e., $\Gamma(N/2 - i) = \Gamma(N/2 + i)$ for integer $i < N/2$. Strictly speaking this holds only for even N. In the case of odd N we have the slightly different relation $\Gamma((N - i)/2) = \Gamma((N + i)/2)$, $\forall\, i \in \{1,3,5,\ldots,N-2\}$. We would like to point out that while the exact $\Gamma(\Delta r)$ is linear for small Δr in the log-log plot, thus implying polynomial decay of correlations in that regime, it flattens out toward $\Delta r \approx N/2$. This behavior is consistent with the physical requirement that the correlation function is smooth at $\Delta r = N/2$. The insets show the value of the half-chain correlators $\Gamma_{N/2}(D) := \Gamma(\Delta r = N/2, D)$ as a function of D. One can clearly see a jump in $\Gamma_{N/2}(D)$ at some D'. This means that in this model, if one wants to obtain good approximations for long-range correlations in the ground state, one must use MPS with bond dimension $D \geqslant D'(N)$. Note that the jump in the inset of Fig. 9 occurs roughly in the same region as the change of the slope in the second plot of Fig. 7. This allows us to understand why in Fig. 7 the slope for large D is steeper than the one for small D: if D is not large enough such that correlations are faithfully reproduced throughout the entire chain, this represents a further source of error besides the inherent error of MPS with nonexponential bond dimension (i.e., $D \ll d^{N/2}$).

The absolute values of the correlation functions for the critical Heisenberg chain with $N = 500$ sites can be found in Fig. 10. We have taken the absolute value since due to the antiferromagnetic nature of the Heisenberg model the ground-state correlation function is changing its sign from site to site. Note that these plots only contain the simulation data since we do not have analytical expressions for the long-range correlations. Qualitatively Fig. 10 shows the same behavior as Fig. 9. Quantitatively we can see that the correlation functions converge at much larger D than in the case of the critical quantum Ising model, which is exactly what we would expect. The half-chain correlators $\Gamma_{N/2}(D)$ exhibit a more or less continuous transition to the region where correlations are faithfully reproduced.

We would like to make an interesting final remark regarding the error in the correlation functions as a function of Δr. In the left part of Fig. 11 we have plotted $\Gamma^{ZZ}_{\mathrm{MPS}}(\Delta r) - \Gamma^{ZZ}_{\mathrm{exact}}(\Delta r)$ for different D in the regime where the half-chain correlators have well converged (i.e., $D > 25$). The surprising thing is that the error does not grow monotonically as a function of Δr as one would expect, but that it rather oscillates around zero. Nevertheless the amplitude of the oscillations is growing monotonically with Δr. The right part of Fig. 11 reveals that, similarly to the relative error of the ground-state energy, the relative error of the half-chain correlators $\Delta_{\mathrm{rel}}\Gamma_{N/2}(D)$ obeys power-law decay as a function of D in the large-D regime.

Our numerical analysis thus indicates that for each N there is a minimum value of $D = D'(N)$ such that correlations throughout the entire chain are properly captured. As investigated in Ref. 19, for critical systems this minimum value of $D'(N)$ is seen to be given by a small power of N that depends on the universality class of the model. This dependence will allow us in Ref. 19 to characterize the cost of the algorithm presented in this work as a power of N. For the moment we will settle for a scaling of the overall computational cost of $O(g(D,\xi/N)D^3)$ where $g(D,\xi/N)$ will be seen to become trivial only for noncritical systems.

C. Noncritical systems

We have seen that for critical systems it is quite involved to predict the computational cost of MPS algorithms that find the optimal approximation of the ground state within the manifold defined by MPS with fixed bond dimension D. This turns out to be much easier for noncritical systems where the correlation length ξ is much smaller than the chain length N. We have

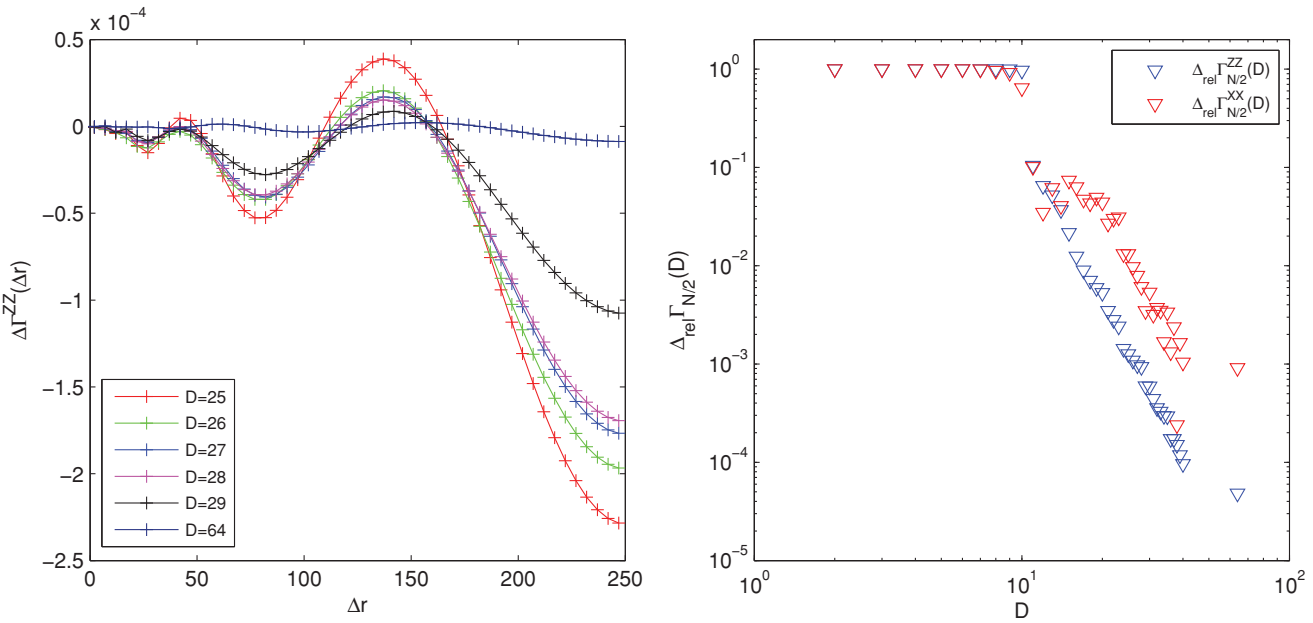


FIG. 11. (Color online) Relative precision of correlation functions in the MPS ground state of the quantum Ising model with $N = 500$. Left: error of the order parameter correlation function $\Gamma^{ZZ}(\Delta r)$ for several different D in the high-precision regime. Right: relative precision of the half-chain correlators $\Gamma^{ZZ}_{N/2}$ and $\Gamma^{XX}_{N/2}$ as a function of the MPS bond dimension D.

studied the spin-1 Heisenberg chain as the prototype of a noncritical quantum spin chain in order to be able to compare our results with the ones presented in Ref. 7. As pictured in Fig. 12, for $N = 100$ and D that is not too big, $n = 4$ is sufficient in order to obtain the optimal MPS approximation to the ground state. This is in agreement with the predictions of Ref. 7. However for D as big as 100, we would have to choose $n = 7$ if we are not willing to lose any precision. This indicates a dependence of n on D which is much weaker than in the case of critical systems. Since due to finite computer memory we cannot increase D arbitrarily, it is safe to say that for systems where $\xi \ll N$, n is given by a small constant. This is exactly what happens for a spin-1 Heisenberg chain with 100 sites since as shown in Ref. 24 the correlation length is roughly $\xi \approx 6$ such that $\xi \ll N$. It is obvious from Fig. 12 that m can be chosen arbitrarily so we can fix it to $m = 1$. Thus in this case the cost of our algorithm scales as $O(D^3)$ which is indeed less by a factor N than the cost from Ref. 7. Nevertheless we must emphasize that for systems where the condition $\xi \ll N$ is not fulfilled anymore, the picture of a small constant n breaks down and the characterization of the computational cost becomes nontrivial.

In Fig. 13 we have plotted the relative energy precision and the correlation functions as functions of D. Note that

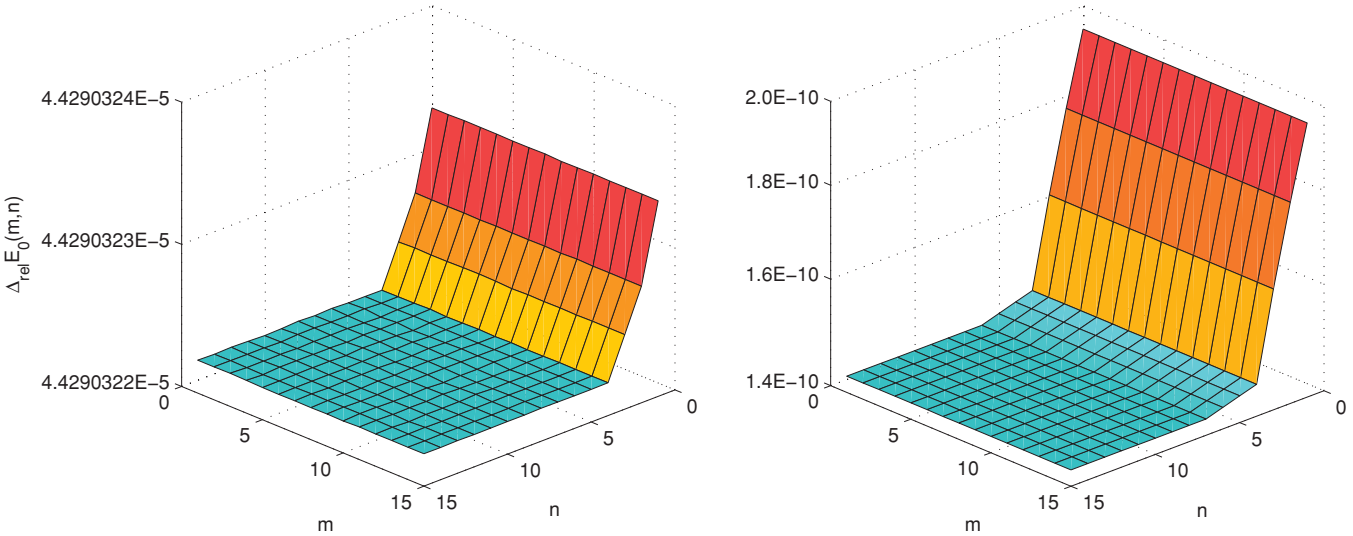


FIG. 12. (Color online) Spin-1 Heisenberg chain with $N = 100$: relative precision of the MPS ground-state energy as compared to the best numerical approximation as a function of the parameters (m,n) for $D = 16$ (left) and 100 (right).

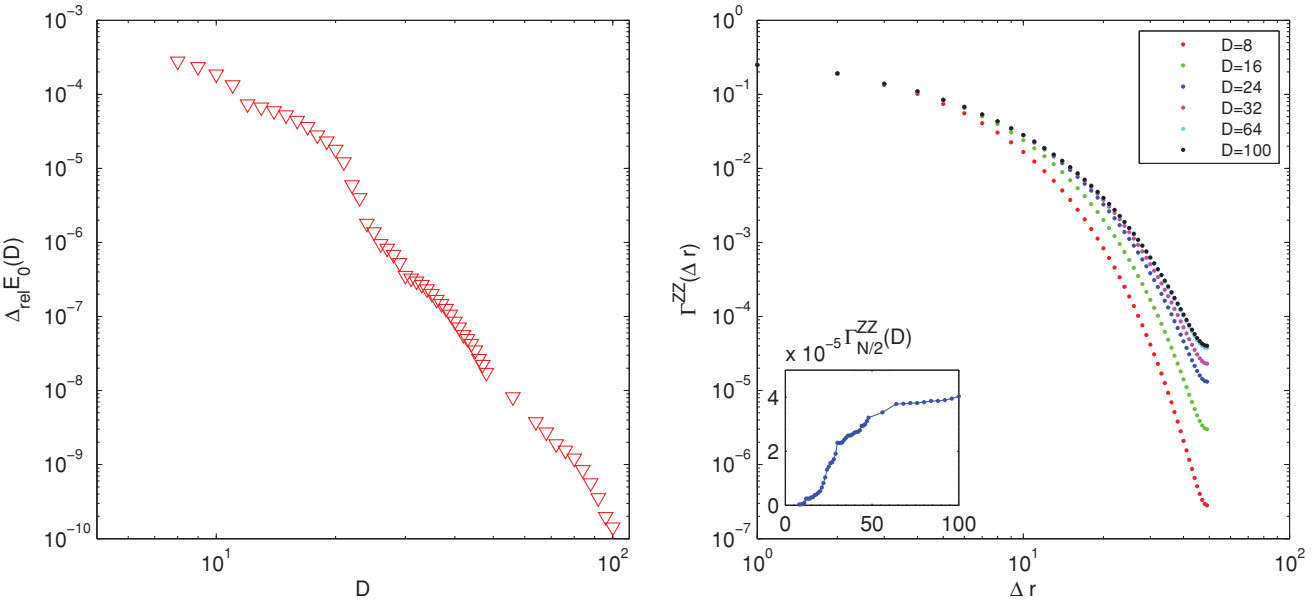


FIG. 13. (Color online) Spin-1 Heisenberg chain with $N = 100$. Left: relative precision of the MPS ground-state energy as a function of D. Right: absolute values of the correlation functions $\Gamma^{ZZ}(\Delta r)$ and as inset the half-chain correlator as a function of D.

for the "exact" ground-state energy density we have used $E_0 = -1.401\,484\,039$ which is the value obtained by an extrapolation of our own finite-D results to infinite D. We have done this since the ground-state energy that we obtain for $D = 100$ is smaller than any other value we have found in the literature, and in particular slightly smaller than the one used as the "exact" ground-state energy in Ref. 7.

The correlation functions plotted in Fig. 13 show nontrivial behavior around $\Delta r \approx N/2$ where they clearly deviate from exponential decay. The half-chain correlator plotted in the inset seems to converge as a function of D but we do not have compelling evidence for that.

V. CONCLUSIONS AND OUTLOOK

We have demonstrated the performance of a gradient-based algorithm for the simulation of TI spin chains with PBC for both critical and noncritical systems. For critical systems where the correlation length is of the order of the system size, the overall scaling of the computational cost is $O(mnD^3) + O(n^2D^3)$ and we have given an analysis of the parameter space $\{m,n\}$ with a prescription of how to obtain a quasioptimal pair $\{m_{\text{opt}},n_{\text{opt}}\}$. In the special case of a critical system that is simulated by MPS with comparatively small D, such that $\tilde{\xi}_D \ll N$ holds for the induced correlation length, the overall scaling is given by $O(mD^3) + O(nD^3)$. For noncritical systems with a correlation length that is much smaller than the system size, increasing D barely affects the parameters m and n and we can write for the overall scaling $O(D^3)$. In the last two cases the cost is one factor N less than that of the algorithm presented in Ref. 7. However, for critical systems in the large-D regime, the cost of Ref. 7 is improved merely by a factor N/n due to the appearence of n^2 in the scaling of our algorithm.

The different types of scaling of the computational cost are directly related to the entanglement entropy of the studied ground states. For critical systems, the exact dependence of m and n on the universality class is an open question and will be treated in future work.

With a TI MPS approximation of the ground state of a system with PBC at hand, it is possible to develop efficient MPS algorithms for the approximation of excited states. The preliminary results we have obtained using the MPS computed in this work as the backbone for an ansatz for momentum eigenstates are very promising.

In higher dimensions, one can use a gradient-based approach to obtain tensor network approximations of the ground state too. However, a straightforward generalization of the present algorithm to higher-dimensional systems with PBC, e.g., in the context of projected entangled pair states, is not obvious due to the fact that the dimension of the transfer matrix grows exponentially with the system size in that case.

ACKNOWLEDGMENTS

We thank V. Murg, E. Rico, and L. Tagliacozzo for valuable discussions. This work was supported by the FWF doctoral program Complex Quantum Systems (Grant No. W1210) the FWF SFB project FoQuS, the ERC grant QUERG, and the ARC Grants No. FF0668731 and No. DP0878830.

APPENDIX A: COMPUTATION OF $H_{\text{eff}}(\mathbf{A})$

Let us introduce a shorthand notation for the building blocks of $H_{\text{eff}}(\mathbf{A})$ that will allow us to express it in a very compact way. From the graphical representation (see Fig. 14) it should be obvious what the objects H^{AA}_{AA}, H^{A}_{AA}, H^{A}_{AA}, $T = T^{A}_{A}$, and T_A mean; note that T denotes the MPS transfer matrix that has been repeatedly mentioned in the main text. For the sake of completeness we also give the definition of the tensor H^{AA}_{AA} explicitly in terms of its components:

$$\left(H^{AA}_{AA}\right)^{\alpha\ \ \gamma}_{\alpha'\ \ \gamma'} = A_i{}^{\alpha}{}_{\beta} A_j{}^{\beta}{}_{\gamma} (H_{s,s+1})^{ij}{}_{i'j'} A^{i'}{}_{\alpha'}{}^{\beta'} A^{j'}{}_{\beta'}{}^{\gamma'}. \quad \text{(A1)}$$

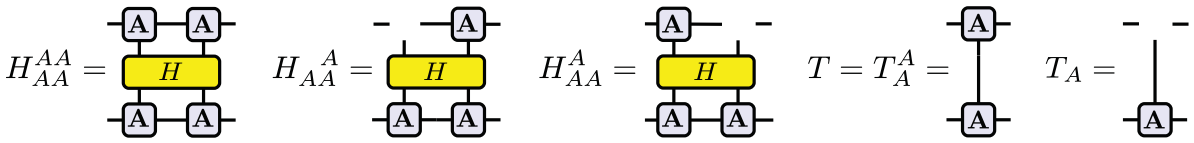


FIG. 14. (Color online) Graphical representation of H_{AA}^{AA}, $H_{AA}{}^{A}$, H_{AA}^{A}, $T = T_A^A$, and T_A.

Here we have used greek letters to label the virtual bonds, latin ones for the physical bonds, and the Einstein summation convention to denote contracted indices. If one combines the left-hand side indices α and α' into one big index and does the same for the right-hand side indices γ and γ', it is clear that H_{AA}^{AA} represents a $D^2 \times D^2$ matrix. The other objects defined in Fig. 14 have similar explicit definitions. $H_{\text{eff}}(\mathbf{A})$ now reads

$$H_{\text{eff}}(\mathbf{A}) = 2\text{Tr}^*\Bigg[H_{AA}{}^{A} T^{N-2} + H_{AA}^{A} T^{N-2} + \sum_{s=0}^{N-3} H_{AA}^{AA} T^s T_A T^{N-3-s} \Bigg], \tag{A2}$$

where $\text{Tr}^*[\cdots]$ indicates that the trace is taken only with respect to the matrix multiplication of the "outer" indices of the "big" $D^2 \times D^2$ matrices. These big matrices may have internal open indices that survive the $\text{Tr}^*[\cdots]$ operation and make sure that $H_{\text{eff}}(\mathbf{A})$ is left with its tensor structure such that it can be later reexpressed as a vector.

The computation of (A2) is the bottleneck of our method. If we were to compute it by straightforward matrix multiplication, even using the sparseness, the computational cost would scale as $O(ND^5)$. In order to improve this scaling, the crucial point is to realize that for large N most terms in (A2) will contain high powers of T, which means that they can be very well approximated within the subspace spanned by the dominant eigenvectors[25] of T. This can be easily seen if we write such factors in their eigenbasis

$$\begin{aligned} T^s &= \sum_{\alpha=1}^{D^2} \lambda_\alpha^s \, |\lambda_\alpha\rangle \langle\lambda_\alpha| \\ &= \lambda_1^s \left[|\lambda_1\rangle \langle\lambda_1| + \sum_{\alpha=2}^{D^2} \left(\frac{\lambda_\alpha}{\lambda_1}\right)^s |\lambda_\alpha\rangle \langle\lambda_\alpha| \right], \end{aligned} \tag{A3}$$

where $|\lambda_1| \geqslant |\lambda_2| \cdots \geqslant |\lambda_{D^2}|$. Obtaining the eigenbasis does not spoil the overall computational cost since, due to the sparse structure of T, one can obtain its n dominant eigenvectors with $O(nD^3)$ operations. Obviously the subspace corresponding to the small-magnitude eigenvalues is suppressed exponentially with increasing s and thus it can be neglected for powers s that are large enough (e.g., for $s = 20$ and $|\frac{\lambda_\alpha}{\lambda_1}| \approx 0.1$, $|\frac{\lambda_\alpha}{\lambda_1}|^s \approx 10^{-20} < 10^{-16}$ which is the machine precision of double-precision floating-point numbers). In these cases it is perfectly fine to restrict ourselves to the subspace spanned by say n dominant eigenvectors, with the parameter n yet to be determined. In fact, we will perform the entire computation a few times, starting with a rather small n and increasing it until the result does not improve any more. When this happens, we know that we have found the optimal n beyond which, when all other parameters are fixed, the precision does not get any better. Thus we will approximate large powers of the transfer matrix as

$$T^s \approx \sum_{\alpha=1}^{n} \lambda_\alpha^s \, |\lambda_\alpha\rangle \langle\lambda_\alpha| . \tag{A4}$$

At this point we must remark that this approximation only works if the moduli of the transfer matrix eigenvalues $|\lambda_\alpha|$ are not concentrated around a certain point (i.e., T is not approximately proportional to unity). In that case, any increment of n will improve the precision and we will end up with very bad overall scaling. In the extremal case of optimal $n = D^2$ the overall scaling becomes $O(D^7)$. For models where this behavior occurs the algorithm presented here may be worse than contracting the tensor networks explicitly, where the scaling is $O(ND^5)$. In these cases the chain length N ultimately decides which method is preferable. Fortunately for the models treated by us, this undesirable behavior does not occur and we end up with relatively small n beyond which the precision does not improve any more.

Let us now return to (A2). There are two different types of terms which must be treated differently. The first and the second terms under the trace can be considered as "easy." They are approximated by

$$\langle H_{1,2}\rangle_{\mathbf{A}}^{[\overline{1}]} = \text{Tr}^*\big[H_{AA}{}^{A} T^{N-2}\big] \approx \sum_{\alpha=1}^{n} \langle\lambda_\alpha| \, H_{AA}{}^{A} \, |\lambda_\alpha\rangle \, \lambda_\alpha^{N-2}, \tag{A5}$$

which is computed within $O(nD^3)$ operations. This is because each contraction $\langle\lambda_\alpha| \, H_{AA}{}^{A} \, |\lambda_\alpha\rangle$ can be performed with cost $O(D^3)$ and this has to be done n times.

The computationally more expensive terms are the ones under the sum over s, where two different powers of T are involved. We will call these terms "hard." They are approximated by

$$\begin{aligned} \langle H_{1,2}\rangle_{\mathbf{A}}^{[\overline{3+s}]} &= \text{Tr}^*\big[H_{AA}^{AA} T^s T_A T^{N-3-s}\big] \\ &\approx \sum_{\alpha,\beta=1}^{n} \langle\lambda_\beta| \, H_{AA}^{AA} \, |\lambda_\alpha\rangle \, \langle\lambda_\alpha| \, T_A \, |\lambda_\beta\rangle \, \lambda_\alpha^s \lambda_\beta^{N-3-s}. \end{aligned} \tag{A6}$$

Here we must remark two things: (i) It is not necessary to let the second index β run over the same range as α. It would be possible to choose as an upper bound a further parameter n' and also vary this one until the precision did not improve anymore. However, since expression (A6) obviously has left-right symmetry, it is sensible to assume that the optimal result would yield $n = n'$. Even if this were not the case, due to the fact that we scan along n, convergence would be reached only for some $n_{\text{optimal}} \geqslant \sup\{n,n'\}$, so we would find the lowest achievable energy anyway. (ii) For very small or very large s either the left or the right transfer matrix segments in (A6) cannot be well approximated by a small number of eigenvalues n since the lower λ_α are not sufficiently suppressed by the

small exponent. In the worst case we would have to take all D^2 eigenvalues into account, which dramatically increases the computational cost. In order to solve this issue, we will compute these terms by exact contraction of segments of length m, which introduces this further parameter into our algorithm. This will be explained in more detail below. For the moment let us note that, depending on the magnitude of s, we can further separate the sum in (A2) over the hard terms into

$$\sum_{s=0}^{N-3} \equiv \sum_{s=0}^{m-1} + \sum_{s=m}^{N-3-m} + \sum_{s=N-2-m}^{N-3}. \tag{A7}$$

We call the terms over which the second sum is taken "medium-s" terms and will treat them differently from the "extremal-s" terms that appear in the first and third sums. Thus $H_{\text{eff}}(\mathbf{A})$ can be divided into

$$H_{\text{eff}}(\mathbf{A}) = 2\big(H_{\text{eff}}^{\text{easy}}(\mathbf{A}) + H_{\text{eff}}^{\text{hard,extr}}(\mathbf{A}) + H_{\text{eff}}^{\text{hard,med}}(\mathbf{A})\big). \tag{A8}$$

A. Computation of extremal-s terms

In this section we treat the terms with small and large s. The first thing to remark is that for large N, if T^s cannot be well approximated within some low-dimensional subspace because s is too small, it is very likely that for T^{N-3-s} the approximation will work because $N-3-s \gg s$. The same observation holds in the other direction if s is too large. Second, depending on the MPS bond dimension D and the amount of entanglement present in the MPS (i.e., depending on the model one is treating), there is a certain m above which T^s with $s \geqslant m$ can be faithfully approximated within the $(n < D^2)$-dimensional subspace spanned by n dominant eigenvectors. As we know nothing about m *a priori*, we introduce it as a further parameter into our algorithm. We will scan m within its range $[1,1/2(N-2)]$ and in the end we will obtain some optimal pair (m,n). The reason why m does not go all the way up to $N-3$ is that, in order for our algorithm to scale effectively as D^3, we must employ the dominant eigenvector approximation on the other half of the chain. Without it we would get the undesirable scaling $O(ND^5)$. The contraction (see Fig. 15) we must perform for each term with small s thus reads

$$\begin{aligned}\langle H_{1,2}\rangle_{\mathbf{A}}^{[\overline{3+s}]} &= \mathrm{Tr}^*\big[H_{AA}^{AA} T^s T_A T^{N-3-s}\big] \\ &\approx \sum_{\alpha=1}^{n} \langle\lambda_\alpha|\, H_{AA}^{AA} T^s T_A\, |\lambda_\alpha\rangle\, \lambda_\alpha^{N-3-s}, \quad \forall\, s < m,\end{aligned} \tag{A9}$$

and can be done with computational cost $O(nD^3)$ using a sparse matrix contraction scheme. As we have to repeat this procedure m times, the total cost scales as $O(mnD^3)$.

The large-s terms (i.e., when $N-3-m < s \leqslant N-3$) can be easily obtained by making use of the left-right symmetry of the tensor network around the point with $s = (N-2)/2$. The sum over all these s turns out to be related to the sum over the small-s terms by taking the transpose with respect to the open virtual bond indices at the empty site where T_A sits. Thus the computational cost remains unchanged, $O(mnD^3)$.

B. Computation of medium-s terms

For terms where s is neither too small nor too large, both powers of the transfer matrix (i.e., T^s and T^{N-3-s}) can be well approximated within the subspace spanned by n dominant eigenvectors. The good news is that in this case the sum over s can be performed analytically in contrast to the extremal-s case where we had to compute each of the m terms separately. However, there is also bad news, namely, that we now have an additional sum over the eigenvalue index stemming from the approximation of T^{N-3-s}. Explicitly, the sum over all medium-s terms reads

$$\begin{aligned} H_{\text{eff}}^{\text{hard,med}}(\mathbf{A}) &= \mathrm{Tr}^*\Bigg[\sum_{s=m}^{N-3-m} H_{AA}^{AA} T^s T_A T^{N-3-s}\Bigg] \\ &\approx \mathrm{Tr}^*\Bigg[\sum_{s=0}^{N-3-2m} \sum_{\alpha,\beta=1}^{n} H_{AA}^{AA} T^m\, |\lambda_\alpha\rangle \\ &\qquad \times \lambda_\alpha^s\, \langle\lambda_\alpha|\, T_A\, |\lambda_\beta\rangle\, \lambda_\beta^{N-3-2m-s}\, \langle\lambda_\beta|\, T^m\Bigg] \\ &= \sum_{\alpha,\beta=1}^{n} \langle\lambda_\beta|\, H_{AA}^{AA}\, |\lambda_\alpha\rangle\, \langle\lambda_\alpha|\, T_A\, |\lambda_\beta\rangle\, \lambda_\alpha^m \\ &\qquad \times \lambda_\beta^m \frac{\lambda_\beta^{N-2-2m} - \lambda_\alpha^{N-2-2m}}{\lambda_\beta - \lambda_\alpha}. \end{aligned} \tag{A10}$$

In the first step we have shifted the summation variable s and have written the matrices T in their eigenbasis. To move from the second to the third line we have used the cyclic property of the trace to write the entire expression as a sum over products of scalars (actually the factor containing T_A is only a scalar with respect to our specially defined trace since it contains internal free indices). Furthermore we have performed the s sum straightforwardly.

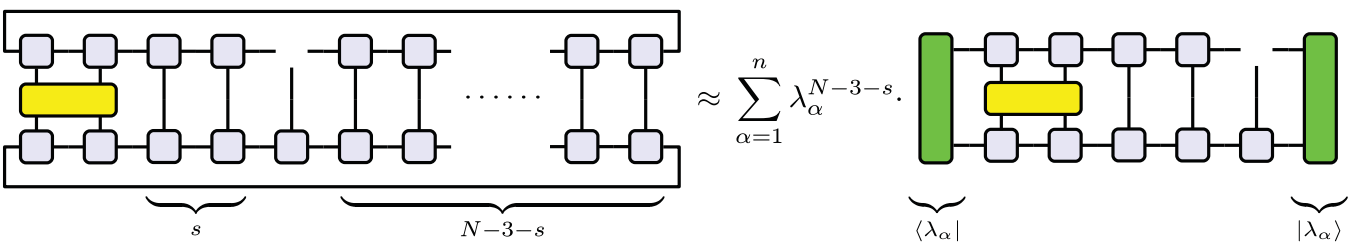


FIG. 15. (Color online) Graphical representation of a term with small s and its approximation within the subspace spanned by n dominant eigenvectors of T.

The computational cost scales here as $O(n^2D^3)$. This is because we have two sums going from 1 to n over terms that are contracted within $O(D^3)$ operations.

APPENDIX B: COMPUTATION OF $N_{\text{eff}}(\mathbf{A})$

Our prescription for the computation of $N_{\text{eff}}(\mathbf{A})$ is also based on the observation that big powers of the transfer matrix T can be very well approximated within the subspace spanned by the dominant eigenvectors. However, here things are much easier than for $H_{\text{eff}}(\mathbf{A})$. This is because the translational invariance is not broken by the two-site Hamiltonian (see Fig. 3) and we can write

$$N_{\text{eff}}(\mathbf{A}) = 2N\langle I\rangle_{\mathbf{A}}^{[\bar{1}]}. \tag{B1}$$

Similarly to $\langle H_{1,2}\rangle_{\mathbf{A}}^{[\bar{1}]}$ in (A5), $\langle I\rangle_{\mathbf{A}}^{[\bar{1}]}$ is approximated by

$$\langle I\rangle_{\mathbf{A}}^{[\bar{1}]} = \mathrm{Tr}^*[T_A T^{N-1}] \approx \sum_{\alpha=1}^{n} \langle\lambda_\alpha|\, T_A\, |\lambda_\alpha\rangle\, \lambda_\alpha^{N-1}, \tag{B2}$$

which is computed within $O(nD^3)$ operations.

[1] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).
[2] S. Rommer and S. Östlund, Phys. Rev. B **55**, 2164 (1997).
[3] G. Vidal, Phys. Rev. Lett. **93**, 040502 (2004).
[4] F. Verstraete, D. Porras, and J. I. Cirac, Phys. Rev. Lett. **93**, 227205 (2004).
[5] F. Verstraete, V. Murg, and J. I. Cirac, Adv. Phys. **57**, 143 (2008).
[6] A. W. Sandvik and G. Vidal, Phys. Rev. Lett. **99**, 220602 (2007).
[7] P. Pippan, S. R. White, and H. G. Evertz, Phys. Rev. B **81**, 081103 (2010).
[8] J. Cardy, *Scaling and Renormalization in Statistical Physics* (Cambridge University Press, Cambridge, 1996).
[9] T. Nishino and K. Okunishi, J. Phys. Soc. Jpn. **64**, 4084 (1995); Y. Hieida, K. Okunishi, and Y. Akutsu, Phys. Lett. A **233**, 464 (1997); K. Okunishi, Y. Hieida, and Y. Akutsu, Phys. Rev. E **59**, R6227 (1999); K. Ueda, T. Nishino, K. Okunishi, Y. Hieida, R. Derian, and A. Gendiar, J. Phys. Soc. Jpn. **75**, 014003 (2006).
[10] TI can also be exploited for MPS simulations with OBC, but this requires addressing an infinite system (Refs. 1, 9, and 11–14). Notice that since the system size N is infinite from the start, there cannot be finite-size or boundary corrections to the bulk properties of the system. However, in this case numerical results are contaminated by effects due to the finite bound dimension D of the MPS. Interestingly, one can apply "finite-D" scaling techniques to extract accurate estimates of bulk properties (Refs. 15–18).
[11] G. Vidal, Phys. Rev. Lett. **98**, 070201 (2007).
[12] I. P. McCulloch, e-print arXiv:0804.2509.
[13] R. Orús and G. Vidal, Phys. Rev. B **78**, 155117 (2008).
[14] B. Pirvu, V. Murg, J. I. Cirac, and F. Verstraete, New J. Phys. **12**, 025012 (2010).
[15] T. Nishino, K. Okunishi, and M. Kikuchi, Phys. Lett. A **213**, 69 (1996).
[16] M. Andersson, M. Boman, and S. Östlund, Phys. Rev. B **59**, 10493 (1999).
[17] L. Tagliacozzo, T. R. de Oliveira, S. Iblisdir, and J. I. Latorre, Phys. Rev. B **78**, 024410 (2008).
[18] F. Pollmann, S. Mukerjee, A. M. Turner, and J. E. Moore, Phys. Rev. Lett. **102**, 255701 (2009).
[19] B. Pirvu *et al.* (unpublished).
[20] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, Cambridge, 2007).
[21] For the spin-1 chain we must apply the operator $M = \exp(i\pi\sigma^y)$ on every second site in order to obtain the same effect.
[22] If D is too large for a given chain length N, the optimal parameter n can get close to its maximal value, i.e., $n \approx D^2$. In these cases the line scan described in Sec. IV A converges at moderate n only due to finite machine precision. However, the precision of the MPS that is obtained in this way is not the one that is theoretically maximally achievable with an MPS of bond dimension D. We emphasize that with infinite machine precision the line scan would converge only close to $n = D^2$ and also the large D points in Figs. 7 and 8 would lie roughly on the line corresponding to polynomial decay.
[23] E. Lieb, T. Schultz, and D. Mattis, Ann. Phys. (NY) **16**, 407 (1961).
[24] S. R. White and D. A. Huse, Phys. Rev. B **48**, 3844 (1993).
[25] Normally one denotes the eigenvector corresponding to the eigenvalue with the largest magnitude as the *dominant eigenvector*. Accordingly, the obvious meaning of the plural (i.e., *dominant eigenvectors*) would be to denote the eigenvectors of a degenerate dominant eigenvalue. However, we rather use the term *dominant eigenvectors* in order to refer to a set of eigenvectors whose corresponding eigenvalues have the largest magnitude among all eigenvalues.