

Algorithms for entanglement renormalization

G. Evenbly and G. Vidal

School of Physical Sciences, University of Queensland, Brisbane 4072, Australia

(Received 15 December 2008; revised manuscript received 8 March 2009; published 7 April 2009; corrected 13 April 2009)

We describe an iterative method to optimize the multiscale entanglement renormalization ansatz for the low-energy subspace of local Hamiltonians on a D -dimensional lattice. For translation-invariant systems the cost of this optimization is logarithmic in the linear system size. Specialized algorithms for the treatment of infinite systems are also described. Benchmark simulation results are presented for a variety of one-dimensional systems, namely, Ising, Potts, XX , and Heisenberg models. The potential to compute expected values of local observables, energy gaps, and correlators is investigated.

DOI: [10.1103/PhysRevB.79.144108](https://doi.org/10.1103/PhysRevB.79.144108)

PACS number(s): 05.30.-d, 02.70.-c, 03.67.Mn, 05.50.+q

I. INTRODUCTION

Entanglement renormalization¹ is a numerical technique based on locally reorganizing the Hilbert space of a quantum many-body system with the aim to reduce the amount of entanglement in its wave function. It was introduced to address a major computational obstacle in real-space renormalization-group (RG) methods,²⁻⁴ responsible for limitations in their performance and range of applicability, namely, the proliferation of degrees of freedom that occurs under successive applications of a RG transformation.

Entanglement renormalization is built around the assumption that, as a result of the local character of physical interactions, some of the relevant degrees of freedom in the ground state of a many-body system can be decoupled from the rest by unitarily transforming small regions of space. Accordingly, unitary transformations known as *disentangler*s are applied locally to the system in order to identify and decouple such degrees of freedom, which are then safely removed and therefore do no longer appear in any subsequent coarse-grained description. This prevents the harmful accumulation of degrees of freedom and thus leads to a sustainable real-space RG transformation, able to explore arbitrarily large one-dimensional (1D) and two-dimensional (2D) lattice systems, even at a quantum critical point. It also leads to the *multiscale entanglement renormalization ansatz* (MERA), a variational ansatz for many-body states.⁵

The MERA, based in turn on a class of quantum circuits, is particularly successful at describing ground states at quantum criticality^{1,5-11} or with topological order.^{12,13} From the computational viewpoint, the key property of the MERA is that it can be manipulated efficiently, due to the causal structure of the underlying quantum circuit.⁵ As a result, it is possible to efficiently evaluate the expected value of local observables, and to efficiently optimize its tensors. Thus, well-established simulation techniques for matrix product states, such as energy minimization¹⁴ or simulation of time evolution,¹⁵ can be readily generalized to the MERA.¹⁶⁻¹⁸

In this paper we describe a simple algorithm (and several variations thereof) to compute an approximation of the low-energy subspace of a local Hamiltonian with the MERA, and present benchmark calculations for 1D lattice systems. Our goal is to provide the interested reader with a rather self-contained explanation of the algorithm, with enough infor-

mation to implement it. Sections II and III review and elaborate on the theoretical foundations of the MERA,^{1,5} and establish the notation and nomenclature used in the rest of the paper. Specifically, Sec. II presents the MERA, both from the perspective of quantum circuits and of the renormalization group, and describes several realizations in 1D and 2D lattices. Then Sec. III explains how to compute expected values of local observables and two-point correlators. Central to this discussion is the *past causal cone* of a small block of lattice sites and the *ascending* and *descending* superoperators, which can be used to move local observables and density matrices up and down the causal cone.

Section IV considers how to optimize a single tensor of the MERA during an energy minimization. This optimization involves linearizing a quadratic cost function for the (isometric) tensor, and computing its *environment*. In Sec. V we describe algorithms to minimize the energy of the state or subspace represented by a MERA. A highlight of the algorithms is their computational cost. For an inhomogeneous lattice with N sites, the cost scales as $O(N)$, whereas for translation-invariant systems it drops to just $O(\log N)$. Other variations in the algorithm allow us to address infinite systems, and scale-invariant systems (e.g., quantum critical systems), at a cost independent of N .

Section VI presents benchmark calculations for different 1D quantum lattice models, namely, Ising, three-level Potts, XX , and Heisenberg models. We compute ground-state energies, magnetizations, and two-point correlators throughout the phase diagram, which includes a second-order quantum phase transition. We find that at the critical point of an infinite system, the error in the ground-state energy decays exponentially with the refinement parameter χ , whereas the two-point correlators remain accurate even at distances of millions of lattice sites. We then extract critical exponents from the order parameter and from two-point correlators. Finally, we also compute a MERA that includes the first excited state, from which the energy gap can be obtained and seen to vanish as $1/N$ at criticality.

This paper replaces similar notes on MERA algorithms presented in Ref. 18. For the sake of concreteness, we have not included several of the algorithms of Ref. 18, which nevertheless remain valid proposals. We have also focused the discussion on a ternary MERA for 1D lattices (instead of the binary MERA used in all previous references) because it is somewhat computationally advantageous (e.g., see compu-

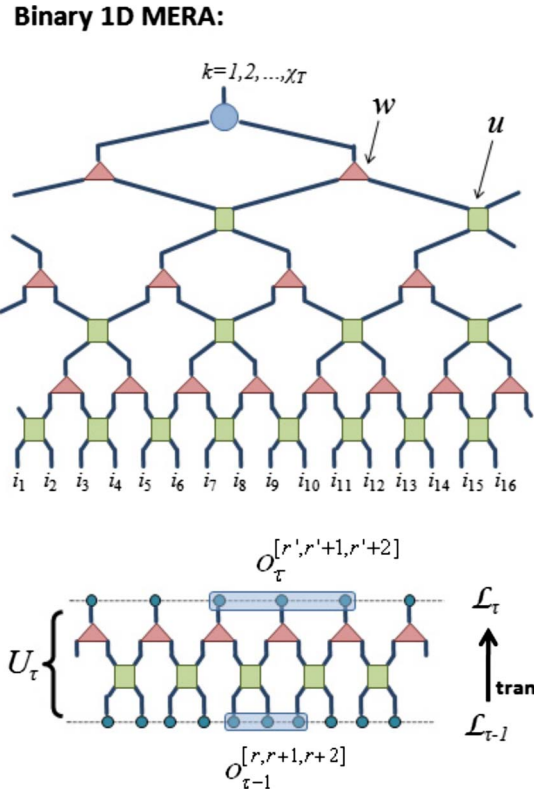
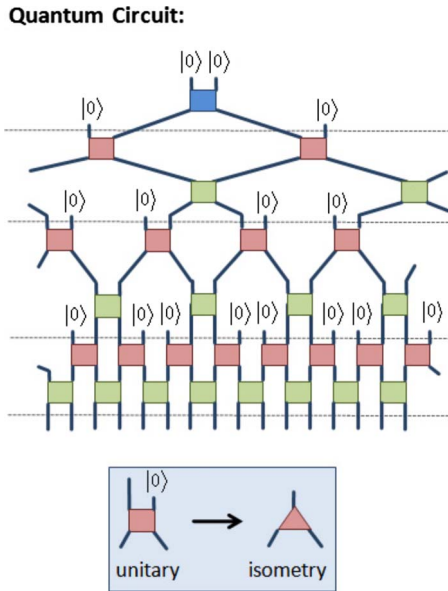


FIG. 1. (Color online) Quantum circuit \mathcal{C} corresponding to a specific realization of the MERA, namely, the binary 1D MERA of Fig. 2. In this particular example, circuit \mathcal{C} is made of gates involving two incoming wires and two outgoing wires, $p=p_{in}=p_{out}=2$. Some of the unitary gates in this circuit have one incoming wire in the fixed state $|0\rangle$ and can be replaced with an isometry w of type (1,2). By making this replacement, we obtain the isometric circuit of Fig. 2.

tation of two-point correlators) and also leads to a much more convenient generalization in two dimensions.

II. MERA

Let \mathcal{L} denote a D -dimensional lattice made of N sites, where each site is described by a Hilbert space \mathbb{V} of finite dimension d , so that $\mathbb{V}_{\mathcal{L}} \cong \mathbb{V}^{\otimes N}$. The MERA is an ansatz used to describe certain pure states $|\Psi\rangle \in \mathbb{V}_{\mathcal{L}}$ of the lattice or, more generally, subspaces $\mathbb{V}_U \subseteq \mathbb{V}_{\mathcal{L}}$.

There are two useful ways of thinking about the MERA that can be used to motivate its specific structure as a tensor network, and also help understand its properties and how the algorithms ultimately work. One way is to regard the MERA as a quantum circuit \mathcal{C} whose output wires correspond to the sites of the lattice \mathcal{L} .⁵ Alternatively, we can think of the MERA as defining a coarse-graining transformation that maps \mathcal{L} into a sequence of increasingly coarser lattices, thus leading to a renormalization-group transformation.¹ Next we briefly review these two complementary interpretations. Then we compare several MERA schemes and discuss how to exploit space symmetries.

A. Quantum circuit

As a quantum circuit \mathcal{C} , the MERA for a pure state $|\Psi\rangle \in \mathbb{V}_{\mathcal{L}}$ is made of N quantum wires, each one described by a Hilbert space \mathbb{V} , and unitary gates u that transform the unentangled state $|0\rangle^{\otimes N}$ into $|\Psi\rangle$ (see Fig. 1).

In a generic case, each unitary gate u in the circuit \mathcal{C} involves some small number p of wires,

FIG. 2. (Color online) (Top) Example of a binary 1D MERA for a lattice \mathcal{L} with $N=16$ sites. It contains two types of isometric tensors, organized in $T=4$ layers. The input (output) wires of a tensor are those that enter it from the top (leave it from the bottom). The *top tensor* is of type (1,2) and the rank χ_T of its upper index determines the dimension of the subspace $\mathbb{V}_U \subseteq \mathbb{V}_{\mathcal{L}}$ represented by the MERA. The *isometries* w are of type (1,2) and are used to replace each block of two sites with a single effective site. Finally, the *disentangler* u are of type (2,2) and are used to disentangle the blocks of sites before coarse-graining. (Bottom) Under the renormalization-group transformation induced by the binary 1D MERA, three-site operators are mapped into three-site operators.

$$u: \mathbb{V}^{\otimes p} \rightarrow \mathbb{V}^{\otimes p}, \quad u^\dagger u = u u^\dagger = I, \quad (1)$$

where I is the identity operator in $\mathbb{V}^{\otimes p}$. For some gates, however, one or several of the input wires are in a fixed state $|0\rangle$. In this case we can replace the unitary gate u with an isometric gate w

$$w: \mathbb{V}_{in} \rightarrow \mathbb{V}_{out}, \quad w^\dagger w = I_{\mathbb{V}_{in}}, \quad (2)$$

where $\mathbb{V}_{in} \cong \mathbb{V}^{\otimes p_{in}}$ is the space of the p_{in} input wires that are not in a fixed state $|0\rangle$ and $\mathbb{V}_{out} \cong \mathbb{V}^{\otimes p_{out}}$ is the space of the $p_{out}=p$ output wires. We refer to w as a (p_{in}, p_{out}) gate or tensor.

Figure 2 shows an example of a MERA for a 1D lattice \mathcal{L} made of $N=16$ sites. Its tensors are of types (1,2) and (2,2). We call the (1,2) tensors isometries w and the (2,2) tensors disentangler u for reasons that will be explained shortly, and refer to Fig. 2 as a binary 1D MERA, since it becomes a binary tree when we remove the disentangler. Most of the previous work for 1D lattices^{1,5-7,16-18} has been done using the binary 1D MERA. However, there are many other pos-

Ternary 1D MERA:

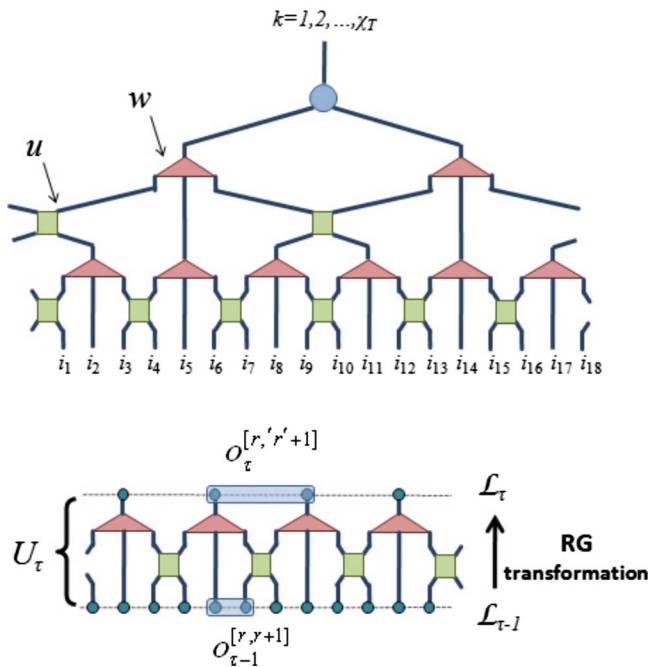


FIG. 3. (Color online) (Top) Example of ternary 1D MERA (rank χ_T , $T=3$) for a lattice of 18 sites. It differs from the binary 1D MERA of Fig. 2 in that the isometries are of type (1,3), so that blocks of three sites are replaced with one effective site. (Bottom) As a result, two-site operators are mapped into two-site operators during the coarse graining.

sible choices. In this paper, for instance, we will mostly use the ternary 1D MERA of Fig. 3, where the isometries w are of type (1,3) and the disentanglers u remain of type (2,2). Figure 4 makes more explicit the meaning of Eq. (2) for these tensors. Notice that describing tensors and their manipulations by means of diagrams is fully equivalent to using equations and often much more clear.

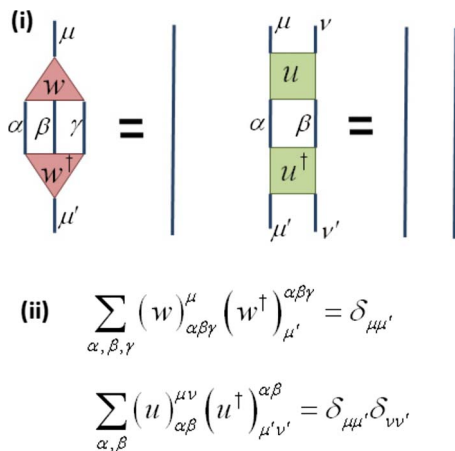


FIG. 4. (Color online) The tensors which comprise a MERA are constrained to be isometric; cf. Eq. (2). The constraints for the isometries w and disentanglers u of the ternary MERA can be equivalently expressed (i) diagrammatically or (ii) with equations. In this paper we will mostly use the diagrammatic notation, which remains simple for complicated tensor networks.

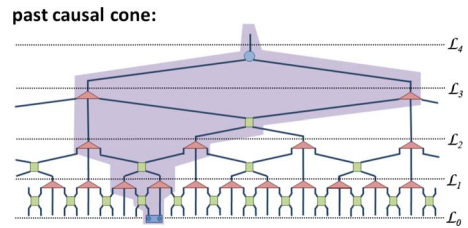


FIG. 5. (Color online) The past causal cone of a group of sites in $\mathcal{L}_0 \equiv \mathcal{L}$ is the subset of wires and gates that can affect the state of those sites. The example shows the causal cone of a pair of nearest-neighbor sites of \mathcal{L}_0 for the ternary 1D MERA. Notice that for each lattice \mathcal{L}_τ , $\tau=0, 1, 2, 3, 4$, the causal cone involves at most two sites. This can be seen to be the case for any pair of contiguous sites of \mathcal{L}_0 . We refer to this property by saying that the causal cones of the MERA have bounded width.

Equation (2) encapsulates a distinctive property of the MERA as a tensor network: each of its tensors is isometric [notice that Eq. (1) is a particular case of Eq. (2)]. A second key feature of the MERA refers to its causal structure. We define the past *causal cone* of an outgoing wire of circuit \mathcal{C} as the set of wires and gates that can affect the state on that wire. A quantum circuit \mathcal{C} leads to a MERA only when the causal cone of an outgoing wire involves just a constant (that is, independent of N) number of wires at any fixed past time (Fig. 5). We refer to this property by saying that the causal cone has a bounded “width.”

The usefulness of the quantum circuit interpretation of the MERA will become clear in Sec. III, in the context of computing expected values for local observables. There, the two defining properties, namely, Eq. (2) and the peculiar structure of the causal cones of \mathcal{C} , will be the key to making such calculations efficient.

B. Renormalization-group transformation

Let us now review how the MERA defines a coarse-graining transformation for lattice systems that leads to a real-space renormalization-group scheme, known as entanglement renormalization.¹ We start by grouping the tensors in the MERA into $T \approx \log N$ different layers, where each layer contains a row of isometries w and a row of disentanglers u . We label these layers with an integer $\tau=1, 2, \dots, T$, with $\tau=1$ for the lowest layer and with increasing values of τ as we climb up the tensor network, and denote by U_τ the isometric transformation implemented by all tensors in layer τ , see Figs. 2 and 3. Notice that the incoming wires of each U_τ define the Hilbert space of a lattice \mathcal{L}_τ with a number of sites N_τ that decreases exponentially with τ (specifically, as $N2^{-\tau}$ and $N3^{-\tau}$ for the binary and ternary 1D MERAs). That is, the MERA implicitly defines a sequence of lattices

$$\mathcal{L}_0 \rightarrow \mathcal{L}_1 \rightarrow \dots \rightarrow \mathcal{L}_T, \tag{3}$$

where $\mathcal{L}_0 \equiv \mathcal{L}$ is the original lattice, and where we can think of lattice \mathcal{L}_τ as the result of coarse-graining lattice $\mathcal{L}_{\tau-1}$.

Specifically, as illustrated in Figs. 2 and 3, this coarse-graining transformation is implemented by the operator U_τ^\dagger that maps pure states of the lattice $\mathcal{L}_{\tau-1}$ into pure states of the lattice \mathcal{L}_τ .

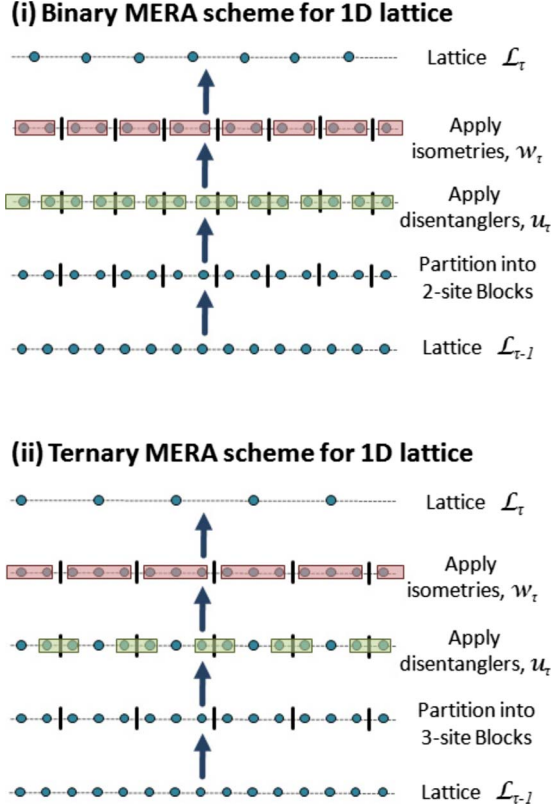


FIG. 6. (Color online) Detailed description of the real-space renormalization-group transformation for 1D lattices induced by (i) the binary 1D MERA and (ii) the ternary 1D MERA.

$$U_{\tau}^{\dagger}: \mathbb{V}_{\mathcal{L}_{\tau-1}} \rightarrow \mathbb{V}_{\mathcal{L}_{\tau}}, \quad (4)$$

and that proceeds in two steps (Fig. 6). Let us partition the lattice $\mathcal{L}_{\tau-1}$ into blocks of neighboring sites. The first step consists of applying the disentanglers u on the boundaries of the blocks, aiming to reduce the amount of short-range entanglement in the system. Once (part of) the short-range entanglement between neighboring blocks has been removed, the isometries w are used in the second step to map each block of sites of lattice $\mathcal{L}_{\tau-1}$ into a single effective site of lattice \mathcal{L}_{τ} .

By composition, we obtain a sequence of increasingly coarse-grained states,

$$|\Psi_0\rangle \rightarrow |\Psi_1\rangle \rightarrow \cdots \rightarrow |\Psi_T\rangle, \quad (5)$$

for the lattices $\{\mathcal{L}_0, \mathcal{L}_1, \dots, \mathcal{L}_T\}$, where $|\Psi_{\tau}\rangle \equiv U_{\tau}^{\dagger}|\Psi_{\tau-1}\rangle$ and $|\Psi_0\rangle \equiv |\Psi\rangle$ is the original state. Overall the MERA corresponds to the transformation $U \equiv U_1 U_2 \cdots U_T$,

$$U: \mathbb{V}_{\mathcal{L}_T} \rightarrow \mathbb{V}_{\mathcal{L}_0}, \quad (6)$$

with $|\Psi_0\rangle = U|\Psi_T\rangle$.

Regarding the MERA from the perspective of the renormalization group is quite instructive. It tells us that this ansatz is likely to describe states with a specific structure of internal correlations, namely, states in which the entanglement is organized in different length scales. Let us briefly explain what we mean by this.

We say that the state $|\Psi\rangle$ contains entanglement at a given length scale λ if by applying a unitary operation (i.e., a disentangler) on a region R of linear size λ , we are able to decouple (i.e., disentangle) some of the local degrees of freedom, that is, if we are able to convert the state $|\Psi\rangle$ into a product state $|\Psi'\rangle \otimes |0\rangle$, where $|0\rangle$ is the state of the local degrees of freedom that have been decoupled and $|\Psi'\rangle$ is the state of the rest of the system. (Here we assumed, of course, that the decoupling is not possible with a unitary operation that acts on a subregion R' of the region R , where the size λ' of R' is smaller than the size of R , $\lambda' < \lambda$.)

What makes the MERA useful is that the entanglement in most ground states of local Hamiltonians seems to decompose into moderate contributions corresponding to different length scales. We can identify two behaviors, depending on whether the system is in a phase characterized by symmetry-breaking order or by topological order (see Ref. 19 and references therein). In systems with symmetry-breaking order, ground-state entanglement spans all length scales λ smaller than the correlation length ξ in the system—and, consequently, at a quantum critical point, where the correlation length ξ diverges, entanglement is present at all length scales.¹ In a system with topological order, instead, the ground state displays some form of (topological) entanglement affecting all length scales even when the correlation length vanishes.^{12,13}

C. Choose your MERA

We have presented the MERA as a tensor network originating in a quantum circuit. Its tensors have incoming and outgoing wires or indices according to a well-defined direction of time in the circuit. Therefore, a MERA can be regarded as a tensor network equipped with a (fictitious) time direction and with two properties:

- (a) Its tensors are isometric [Eq. (2)].
- (b) Past causal cones have bounded width (Fig. 5).

From a computational perspective, these are the only properties that we need to retain. In particular, there is no need to keep the vector space dimension of the quantum wires (equivalently, of the sites in the coarse-grained lattice) constant throughout the tensor network. Accordingly, we will consider a MERA where the vector space dimension of a site of lattice \mathcal{L}_{τ} denoted χ_{τ} may depend on the layer τ . (This dimension could also be different for each individual site of layer τ , but for simplicity we will not consider this case here.) Notice that $\chi_0 = d$ corresponds to the sites of the original lattice \mathcal{L} .

Bond dimension. Often, however, the sites in most layers will have the same vector space dimension (except, for instance, the sites of the original lattice \mathcal{L} , with $\chi_0 = d$, or the single site of the top lattice \mathcal{L}_T , with dimension χ_T). In this case we denote the dominant dimension simply by χ , and we refer to the MERA as having *bond dimension* χ . The computational cost of the algorithms described in subsequent sections is often expressed as a power of the bond dimension χ .

Rank. We refer to the dimension χ_T of the space $\mathbb{V}_{\mathcal{L}_T}$ (corresponding to the single site of the uppermost lattice \mathcal{L}_T)

as the rank of the MERA. For $\chi_T=1$, the MERA represents a pure state $|\Psi\rangle \in \mathbb{V}_{\mathcal{L}}$. More generally, a rank- χ_T MERA encodes a χ_T -dimensional subspace $\mathbb{V}_U \subseteq \mathbb{V}_{\mathcal{L}}$. For instance, given a Hamiltonian H on the lattice \mathcal{L} , we could use a rank- χ_T MERA to describe the ground subspace of H (assuming it had dimension χ_T), or the ground state of H (if it was not degenerate) and its χ_T-1 excitations with lowest energy. The isometric transformation U in Eq. (6) can be used to build a projector $P \equiv UU^\dagger$,

$$P: \mathbb{V}_{\mathcal{L}} \rightarrow \mathbb{V}_{\mathcal{L}}, \quad P^2 = P, \quad \text{tr}(P) = \chi_T, \quad (7)$$

onto the subspace $\mathbb{V}_U \subseteq \mathbb{V}_{\mathcal{L}}$.

Given the above definition of the MERA, many different realizations are possible depending on what kind of isometric tensors are used and how they are interconnected. We have already met two examples for a 1D lattice, based on a binary and ternary underlying tree. Figure 7 shows two schemes for a 2D square lattice. It is natural to ask, given a lattice geometry, what realization of the MERA is the most convenient from a computational point of view. A definitive answer to this question does not seem simple. An important factor, however, is given by the fixed-point size of the support of local observables under successive RG transformations—which corresponds to the width of the past causal cones.

Support of local observables. In each MERA scheme, under successive coarse-graining transformations a local operator eventually becomes supported in a characteristic number of sites. This is the result of two competing effects: disentangles u tend to extend the support of the local observable (by adding new sites at its boundary), whereas the isometries w tend to reduce it (by transforming blocks of sites into single sites). For instance, in the binary 1D MERA, local observables end up supported in three contiguous sites (Fig. 2), whereas in the ternary 1D MERA local observables become supported in two contiguous sites (Fig. 3).

Therefore, an important difference between the binary and ternary 1D schemes is in the natural support of local observables. This can be seen to imply that the cost of a computation scales as a larger power of the bond dimension χ for the binary scheme than for the ternary scheme, namely, as $O(\chi^9)$ compared to $O(\chi^8)$. However, it turns out that the binary scheme is more effective at removing entanglement, and as a result a smaller χ is already sufficient in order to achieve the same degree of accuracy in the computation of, say, a ground-state energy. In the end, we find that for the 1D systems analyzed in Sec. VI, the two effects compensate and the costs required in both schemes in order to achieve the same accuracy are comparable. On the other hand, in the ternary 1D MERA, two-point correlators between selected sites can be computed at a cost $O(\chi^8)$, whereas analogous calculations in the binary 1D MERA are much more expensive. Therefore in any context where the calculation of two-point correlators is important, the ternary 1D MERA is a better choice.

The number of possible realizations of the MERA for 2D lattices is greater than for 1D lattices. For a square lattice, the two schemes of Fig. 7 are obvious generalizations of the above ones for 1D lattices. The first scheme, proposed in Ref. 6 (see also Ref. 8), involves isometries of type (1,4) and the natural support of local observables is a block of 3×3

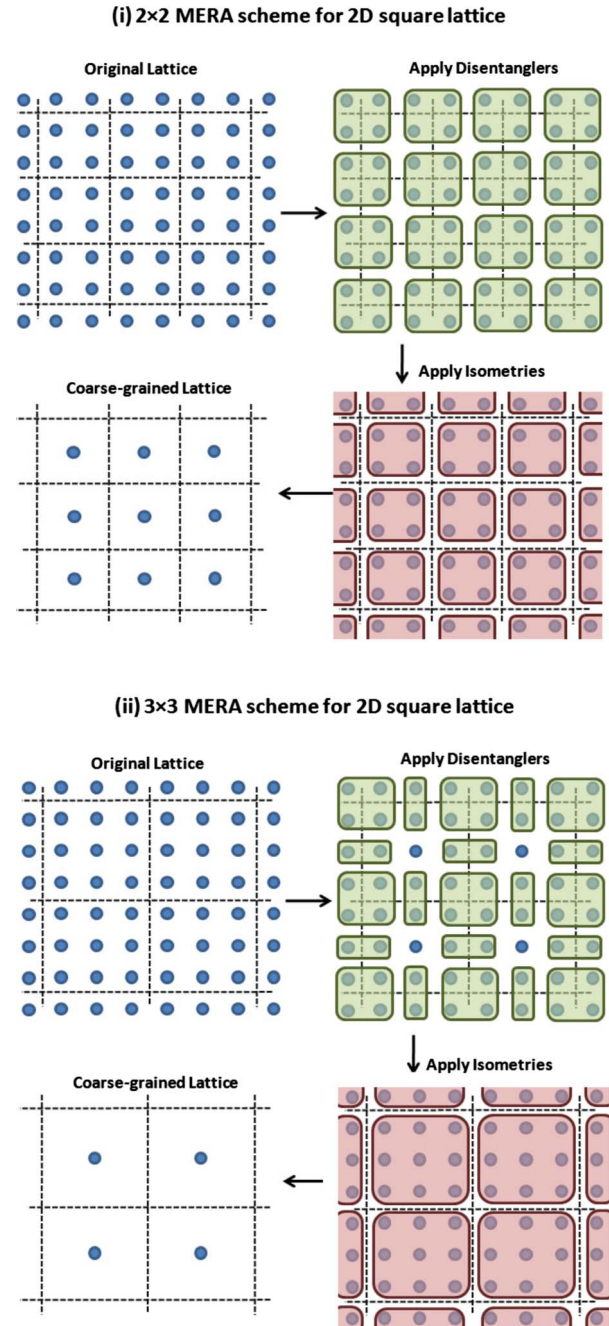


FIG. 7. (Color online) Detailed description of the real-space renormalization-group transformation for a 2D square lattice induced by two possible realizations of the MERA, generalizing the 1D schemes of Fig. 6. In the first case the isometries map a block of 2×2 sites into a single site, which can be seen to imply that the natural size of a local operator, equivalently the causal width of the scheme, is 3×3 sites. In the second case the isometries map a block of 3×3 sites into a single site and the natural size of a local operator is 2×2 . As a result, the computational cost in the second scheme is much smaller than in the first scheme.

sites. The second scheme involves isometries of type (1,9) and local observables end up supported in blocks of 2×2 sites. Here, the much narrower causal cones of the second scheme leads to a much better scaling of the computational

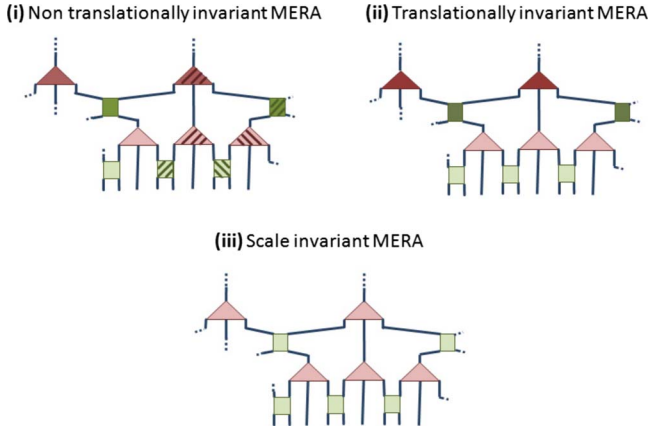


FIG. 8. (Color online) Ternary 1D MERA in the presence of space symmetries. (i) In order to represent an inhomogeneous state or subspace, all disentanglers u and isometries w are different (denoted by different colors). Notice that there are $N/3$ disentanglers (isometries) in the first layer, $N/9$ in the second, and more generally $N/3^\tau$ in layer τ , so that the total number of tensors is $2N\sum_{\tau=1}^{\log N} 1/3^\tau < 2N$. Therefore the total number of parameters required to specify the MERA is proportional to the size N of the lattice \mathcal{L} . (ii) In order to represent a state or subspace that is invariant under translations, we choose all disentanglers and isometries on a given layer of the MERA to be the same. In this case the MERA is completely specified by $O(\log N)$ disentanglers and isometries. (iii) In a scale-invariant MERA, the same disentangler and isometry is in addition used in all layers.

cost with χ , only $O(\chi^{16})$ compared to $O(\chi^{28})$ for the first scheme.

Another remark in relation to possible realizations concerns the type of tensors we use. So far we have insisted in distinguishing between disentanglers u (unitary tensors of type $p \rightarrow p$) and isometries w (isometric tensors of type $1 \rightarrow p'$). We will continue to use this terminology throughout this paper, but we emphasize that a more general form of isometric tensor, e.g., of type (2,4), which both disentangles the system and coarse-grained sites, is possible and is actually used in some realizations.⁹

D. Exploiting symmetries

Symmetries have a direct impact on the efficiency of computations, because they can be used to drastically reduce the number of parameters in the MERA. Important examples are given by space symmetries, such as translation and scale invariance; see Fig. 8.

The MERA is made of $O(N)$ disentanglers and isometries. In order to describe an inhomogeneous state $|\Psi\rangle \in \mathbb{V}_{\mathcal{L}}$ or subspace $\mathbb{V}_U \subseteq \mathbb{V}_{\mathcal{L}}$, all these tensors are chosen to be different. Therefore, for fixed χ the number of parameters in the MERA scales linearly in N .

However, in the presence of translation invariance, one can use a *translation-invariant* MERA, where we choose all the disentanglers u and isometries w of any given layer τ to be the same, thus reducing the number of parameters to $O(\log N)$ (if there are $T \approx \log N$ layers). We emphasize that a translation-invariant MERA, as just defined, does not neces-

sarily represent a translation-invariant state $|\Psi\rangle \in \mathbb{V}_{\mathcal{L}}$ or subspace $\mathbb{V}_U \subseteq \mathbb{V}_{\mathcal{L}}$. The reason is that different sites of \mathcal{L} are placed in inequivalent positions with respect to the MERA. As a result, often the MERA can only approximately reproduce translation-invariant states or subspaces, although the departure from translation invariance is seen to typically decrease fast with increasing χ . In order to further mitigate inhomogeneities, we often consider an average of local observables or reduced density matrices over all possible sites, as will be discussed in Sec. III.

In systems that are invariant under changes in scale, we will use a *scale-invariant* MERA, where all the disentanglers and isometries can be chosen to be the same and we only need to store a constant number of parameters. The scale-invariant MERA is useful in representing the ground state of some quantum critical systems¹ and the ground subspace of systems with topological order at the infrared limit of the RG flow.^{12,13}

A reduction in parameters (as a function of χ) is also possible in the presence of internal symmetries, such as $U(1)$ (e.g., particle conservation) or $SU(2)$ (e.g., spin isotropy). We defer their analysis to Ref. 21.

For the sake of concreteness, the explanations in the rest of this paper refer to the ternary 1D scheme of Fig. 3. However, analogous considerations also apply to any other realization of the MERA.

III. COMPUTATION OF EXPECTED VALUES OF LOCAL OBSERVABLES AND CORRELATORS

Let $o^{[r,r+1]}$ denote a local observable defined on two contiguous sites $[r, r+1]$ of \mathcal{L} . In this section we explain how to compute the expected value

$$\langle o^{[r,r+1]} \rangle_{\mathbb{V}_U} \equiv \text{tr}(o^{[r,r+1]}P). \quad (8)$$

Here P is a projector [see Eq. (7)] onto the χ_T -dimensional subspace $\mathbb{V}_U \subseteq \mathbb{V}_{\mathcal{L}}$ represented by the MERA. For a rank $\chi_T=1$ MERA, representing a pure state $|\Psi\rangle \in \mathbb{V}_{\mathcal{L}}$, the above expression reduces to

$$\langle o^{[r,r+1]} \rangle_{\Psi} \equiv \langle \Psi | o^{[r,r+1]} | \Psi \rangle. \quad (9)$$

Evaluating Eq. (8) is necessary in order to extract physically relevant information from the MERA, such as, e.g., the energy and magnetization in a spin system. In addition, the manipulations involved are also required as a central part of the optimization algorithms described in Secs. IV and V. The results of this section remain relevant even in cases where no optimization algorithm is required (for instance, when an exact expression of the MERA is known^{12,13}).

As explained below, the expected value of Eq. (8) can be computed in a number of ways:

(a) By repeated use of the *ascending superoperator* \mathcal{A} , the local operator $o^{[r,r+1]}$ is mapped onto a coarse-grained operator o_T on lattice \mathcal{L}_T . Equation (8) can then be evaluated as the trace of the coarse-grained operator o_T , $\text{tr}(o^{[r,r+1]}P) = \text{tr}(o_T)$.

(b) Alternatively, by repeated use of the *descending superoperator* \mathcal{D} , a two-site reduced density matrix $\rho^{[r,r+1]}$ for lattice \mathcal{L} is obtained. Equation (8) can then be evaluated as $\text{tr}(o^{[r,r+1]}P) = \text{tr}(o^{[r,r+1]}\rho^{[r,r+1]})$.

(c) More generally, the ascending and descending superoperators \mathcal{A} and \mathcal{D} can be used to compute an operator $o_{\tau}^{[r',r'+1]}$ and density matrix $\rho_{\tau}^{[r',r'+1]}$ for the coarse-grained lattice \mathcal{L}_{τ} . Equation (8) can then be evaluated as $\text{tr}(o_{\tau}^{[r,r+1]}P) = \text{tr}(o_{\tau}^{[r',r'+1]}\rho_{\tau}^{[r',r'+1]})$.

First we introduce the ascending and descending superoperators \mathcal{A} and \mathcal{D} and explain in detail how to perform the computation of the expected value of Eq. (8). Then we address also the computation of the expected value

$$\langle O \rangle_{\mathcal{V}_U} \equiv \text{tr}(OP), \quad O \equiv \sum_r o^{[r,r+1]}, \quad (10)$$

where O is an operator that decomposes as a sum of local operators in \mathcal{L} , as well as the computation of two-point correlators. Finally, we revisit these tasks in the presence of translation invariance and scale invariance.

The ascending and descending superoperators are an essential part of the MERA formalism that was introduced in Ref. 5 (see, e.g., Fig. 4 of Ref. 5 for an explicit representation of the descending superoperator \mathcal{D}). These superoperators have been also called MERA quantum channel or MERA transfer matrix in Ref. 10.

A. Ascending and descending superoperators

In Sec. II we have seen that the MERA defines a sequence of increasingly coarser lattices $\{\mathcal{L}_0, \mathcal{L}_1, \dots, \mathcal{L}_T\}$. Under the coarse-graining transformation U_{τ}^{\dagger} of Eq. (4), a local operator $o_{\tau-1}^{[r,r+1]}$, supported on two consecutive sites $[r, r+1]$ of lattice $\mathcal{L}_{\tau-1}$, is mapped onto another local operator $o_{\tau}^{[r',r'+1]}$ supported on two consecutive sites $[r', r'+1]$ of lattice \mathcal{L}_{τ} (Fig. 3). This is so because in $U_{\tau}^{\dagger} o_{\tau-1}^{[r,r+1]} U_{\tau}$ most disentanglers and isometries of U_{τ} and U_{τ}^{\dagger} are annihilated in pairs according to Eq. (2). The resulting transformation is implemented by means of the *ascending* superoperator \mathcal{A} described in Fig. 9,

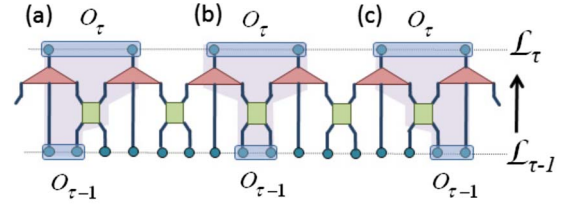
$$o_{\tau}^{[r',r'+1]} = \mathcal{A}(o_{\tau-1}^{[r,r+1]}). \quad (11)$$

In order to keep our notation simple, we do not specify on which lattice or sites the superoperator \mathcal{A} is applied, even though \mathcal{A} actually depends on τ , r , and r' . Instead, when necessary we will simply indicate which of its three structurally different forms (namely, left \mathcal{A}_L , center \mathcal{A}_C , or right \mathcal{A}_R in Fig. 9) is being used.

As above, let $[r, r+1]$ denote two consecutive sites of lattice $\mathcal{L}_{\tau-1}$ and let $[r', r'+1]$ denote two consecutive sites of lattice \mathcal{L}_{τ} that lay inside the past causal cone of $[r, r+1] \in \mathcal{L}_{\tau-1}$. Given a density matrix $\rho_{\tau}^{[r',r'+1]}$ in \mathcal{L}_{τ} , the *descending superoperator* \mathcal{D} of Fig. 10 produces a density matrix $\rho_{\tau-1}^{[r,r+1]}$ in $\mathcal{L}_{\tau-1}$,

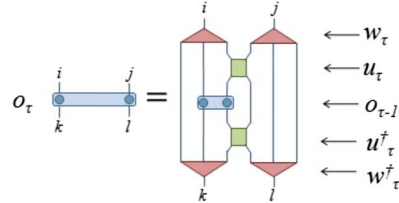
$$\rho_{\tau-1}^{[r,r+1]} = \mathcal{D}(\rho_{\tau}^{[r',r'+1]}). \quad (12)$$

Notice that the descending superoperator \mathcal{D} (which depends on τ , r , and r') is the dual of the ascending superoperator \mathcal{A} , $\mathcal{D} = \mathcal{A}^*$. Indeed, as can be checked in Fig. 11, by construction we have for any $o_{\tau-1}^{[r,r+1]}$ and $\rho_{\tau}^{[r',r'+1]}$,



Ascending Superoperators:

$$(a') o_{\tau} = \mathcal{A}_L(o_{\tau-1})$$



$$(b') o_{\tau} = \mathcal{A}_C(o_{\tau-1})$$

$$(c') o_{\tau} = \mathcal{A}_R(o_{\tau-1})$$

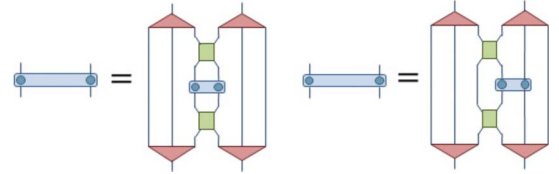


FIG. 9. (Color online) The ascending superoperator \mathcal{A} transforms a local operator $o_{\tau-1}$ of lattice $\mathcal{L}_{\tau-1}$ into a local operator o_{τ} of lattice \mathcal{L}_{τ} (for simplicity we omit the label $[r, r+1]$ that specifies the sites on which $o_{\tau-1}$ and o_{τ} are supported). Depending on the relative position between the support of $o_{\tau-1}$ and the closest disentangler, the operator can be lifted to lattice \mathcal{L}_{τ} in three different ways, indicated in the figure as (a)–(c). Correspondingly, there are three structurally different forms of the ascending superoperator \mathcal{A} , namely, left \mathcal{A}_L , center \mathcal{A}_C , and right \mathcal{A}_R , indicated as (a')–(c'). Notice that the figure completely specifies the tensor network representation of the superoperator, which is written in terms of the relevant disentanglers and isometries (and their Hermitian conjugates). An explicit form for the *average ascending superoperator* $\bar{\mathcal{A}}$ of Eq. (34) is obtained by averaging the above three tensor networks.

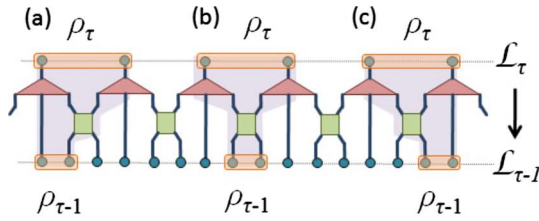
$$\text{tr}[o_{\tau-1}^{[r,r+1]}\mathcal{D}(\rho_{\tau}^{[r',r'+1]})] = \text{tr}[\mathcal{A}(o_{\tau-1}^{[r,r+1]})\rho_{\tau}^{[r',r'+1]}]. \quad (13)$$

Correspondingly, there are also three structurally different forms of the descending superoperators, namely, left \mathcal{D}_L , center \mathcal{D}_C , and right \mathcal{D}_R in Fig. 10.

B. Evaluation of a two-site operator

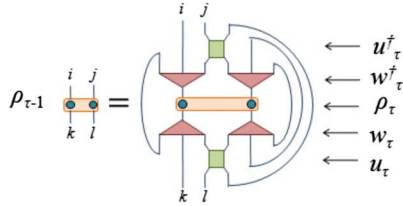
We can now proceed to compute the expected value $\text{tr}(o_{\tau}^{[r,r+1]}P)$ of Eq. (8) from the MERA. This computation corresponds to contracting the tensor network depicted in the upper half of Fig. 12.

In a key first step, the contraction of the tensor network for $\text{tr}(o_{\tau}^{[r,r+1]}P)$ is significantly simplified by the fact that, by virtue of Eq. (2), each isometric tensor outside the past causal cone of sites $[r, r+1] \in \mathcal{L}$ is annihilated by its Hermitian conjugate. As a result, we are left with a new tensor

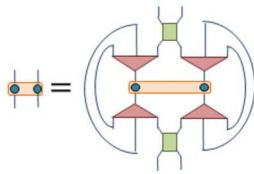


Descending Superoperator:

(a') $\rho_{\tau-1} = \mathcal{D}_L(\rho_\tau)$



(b') $\rho_{\tau-1} = \mathcal{D}_C(\rho_\tau)$



(c') $\rho_{\tau-1} = \mathcal{D}_R(\rho_\tau)$

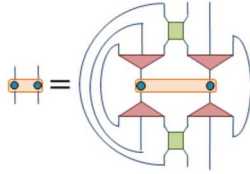


FIG. 10. (Color online) The descending superoperator \mathcal{D} transforms a local density matrix ρ_τ of lattice \mathcal{L}_τ into a local density matrix $\rho_{\tau-1}$ of lattice $\mathcal{L}_{\tau-1}$. Depending on the relative position between the support of $\rho_{\tau-1}$ and the closest disentangler, the density matrix ρ_τ can be lowered to lattice $\mathcal{L}_{\tau-1}$ in three different ways, indicated in the figure as (a)–(c). Correspondingly, there are three structurally different forms of the descending superoperator \mathcal{D} , namely, left \mathcal{D}_L , center \mathcal{D}_C , and right \mathcal{D}_R , indicated as (a')–(c'). An explicit form for the *average descending superoperator* $\overline{\mathcal{D}}$ of Eq. (40) is obtained by averaging the above three tensor networks.

network that contains only (two copies of) the tensors in the causal cone, as represented in the second half of Fig. 12. Because the past causal cones in the MERA have a bounded width, this tensor network can now be contracted with a computational effort that grows with N just as $O(\log N)$. One can proceed in several ways.

Duality:

$$\text{tr}\{\mathcal{A}_\tau(o_{\tau-1})\rho_\tau\} = \text{tr}\{o_{\tau-1}\mathcal{D}_\tau(\rho_\tau)\}$$

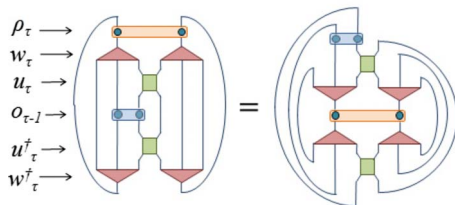


FIG. 11. (Color online) The ascending and descending superoperators, \mathcal{A} and \mathcal{D} , are dual to each other; see Eq. (13). This becomes evident by inspecting the above figure, where the superoperators are explicitly decomposed in terms of disentanglers and isometries.

Expected value of a local observable:

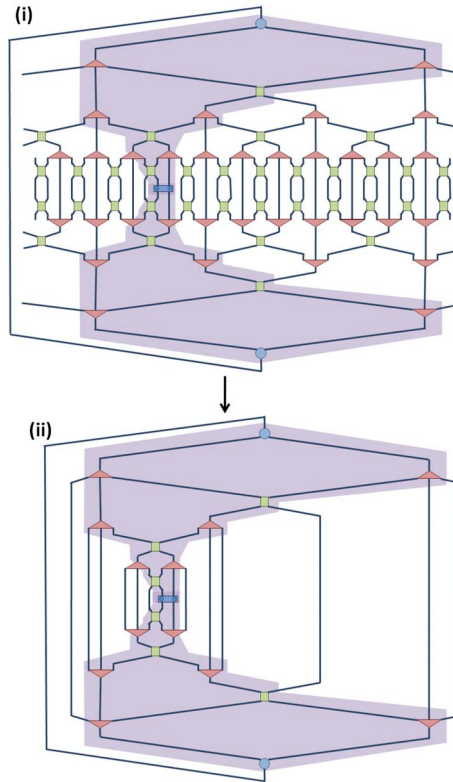


FIG. 12. (Color online) (i) Tensor network corresponding to the expected value $\text{tr}(o^{[r,r+1]}\rho)$ of Eq. (8). The two-site operator $o^{[r,r+1]}$ is represented by a four-legged rectangle in the middle of the tensor network. The shaded region represents the past causal cone of sites $r, r+1 \in \mathcal{L}$. (ii) All isometric tensors that lay outside the past causal cone of sites $r, r+1 \in \mathcal{L}$ annihilate and we are left with a simpler tensor network.

Bottom-top. In the bottom-top approach, we would start by contracting the indices of $o^{[r,r+1]}$ and the disentanglers and isometries of the first layer ($\tau=1$) of the causal cone; then we would contract the indices of disentanglers and isometries of the second layer ($\tau=2$); and so on (Fig. 13). However, this corresponds to repeatedly applying the ascending superoperator \mathcal{A} on $o_0^{[r,r+1]} \equiv o^{[r,r+1]}$. Therefore this is precisely how we proceed, obtaining a sequence of increasingly coarse-grained operators

Top-tensor contractions:

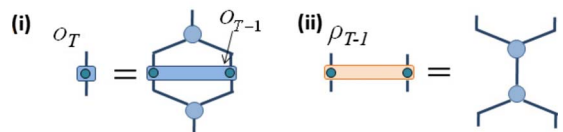


FIG. 13. (Color online) (i) The top tensor transforms a two-site operator o_{T-1} defined on lattice \mathcal{L}_{T-1} into a one-site operator (a $\chi_T \times \chi_T$ matrix) o_T on the top of the MERA. (ii) The two-site density matrix ρ_{T-1} on lattice \mathcal{L}_{T-1} is obtained through contraction of the top tensor with its conjugate. Notice that ρ_{T-1} , as well as all ρ_τ , are normalized to have trace $\text{tr}(\rho_\tau) = \chi_T$.

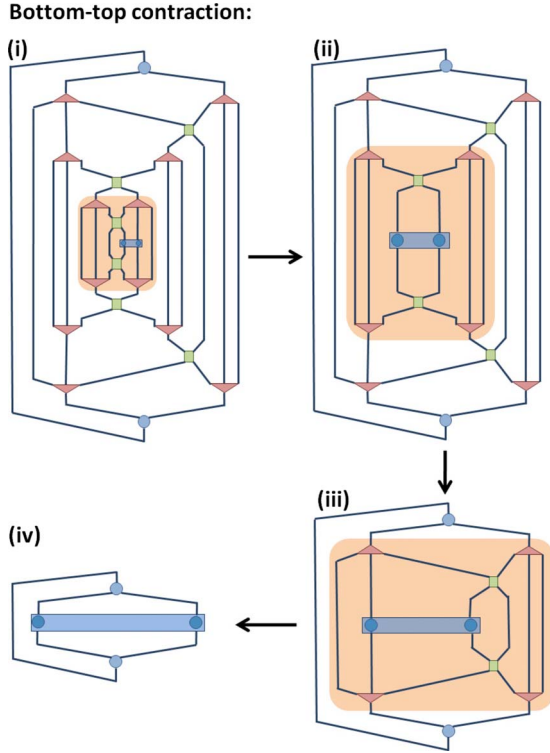


FIG. 14. (Color online) The contraction of the tensor network in the lower half of Fig. 12 using the bottom-top approach corresponds to employing the ascending super operator \mathcal{A} a number of times. In this particular case, we first use (i) \mathcal{A}_R , then (ii) \mathcal{A}_C , and then (iii) \mathcal{A}_L , to bring the tensor network into a simple form whose contraction gives a complex number: the expected value of Eq. (8).

$$o_0^{[r,r+1]} \xrightarrow{\mathcal{A}} o_1^{[r_1,r_1+1]} \xrightarrow{\mathcal{A}} o_2^{[r_2,r_2+1]} \rightarrow \dots \rightarrow o_T \quad (14)$$

supported on lattices $\mathcal{L}_0, \mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_T$, respectively. Here, the $\chi_T \times \chi_T$ matrix o_T at the top of the MERA is obtained according to Fig. 14 and the expected value of Eq. (8) corresponds to its trace,

$$\text{tr}(o^{[r,r+1]}P) = \text{tr}(o_T). \quad (15)$$

Top-bottom. In the top-bottom approach, we would instead start by contracting the indices of the tensors in the top layer ($\tau=T$) of the causal cone; then we would contract the indices of the tensors in the layer right below ($\tau=T-1$); and so on (Fig. 15). However, that corresponds to first computing a density matrix ρ_{T-1} for the two sites of \mathcal{L}_{T-1} according to Fig. 14 and then repeatedly applying the descending superoperator \mathcal{D} . Therefore this is how we proceed, producing a sequence of two-site density matrices

$$\rho_{T-1} \xrightarrow{\mathcal{D}} \dots \rho_2^{[r_2,r_2+1]} \xrightarrow{\mathcal{D}} \rho_1^{[r_1,r_1+1]} \xrightarrow{\mathcal{D}} \rho_0^{[r,r+1]} \quad (16)$$

supported on lattices $\mathcal{L}_{T-1}, \dots, \mathcal{L}_2, \mathcal{L}_1, \mathcal{L}_0$, respectively.²² The last density matrix, $\rho^{[r,r+1]} \equiv \rho_0^{[r,r+1]}$, describes the state of the two sites of \mathcal{L} on which the local operator $o^{[r,r+1]}$ is supported. Therefore we can evaluate the expected value of $o^{[r,r+1]}$,

$$\text{tr}(o^{[r,r+1]}P) = \text{tr}(o^{[r,r+1]}\rho^{[r,r+1]}). \quad (17)$$

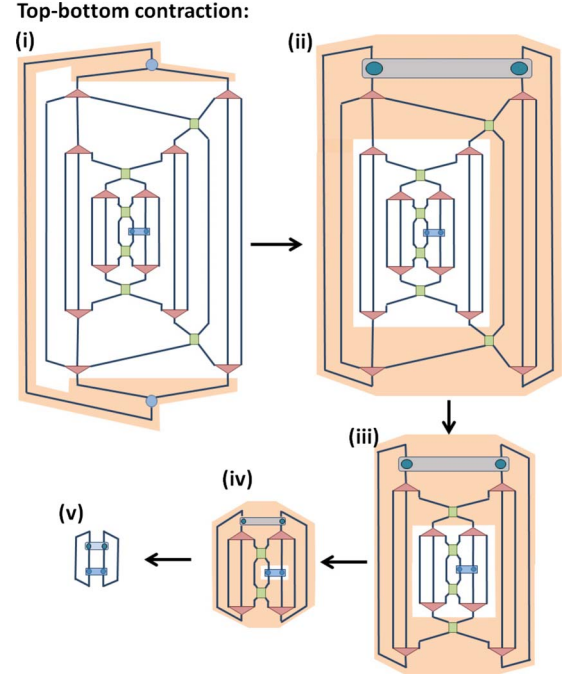


FIG. 15. (Color online) The contraction of the tensor network in the lower half of Fig. 12 using the top-bottom approach corresponds to first implementing (i) a *top tensor* contraction followed by repeated application of the descending super operator \mathcal{D} . Specifically, here we first use (ii) \mathcal{D}_L , then (iii) \mathcal{D}_C , and then (iv) \mathcal{D}_R , in order to compute the appropriate density matrix $\rho^{[r,r+1]}$ for two sites $[r, r+1] \in \mathcal{L}$. With the density matrix $\rho^{[r,r+1]}$ we can finally compute (v) the expectation value $\text{tr}(o^{[r,r+1]}P) = \text{tr}(o^{[r,r+1]}\rho^{[r,r+1]})$.

Middle ground. More generally, one can also evaluate the expected value of Eq. (8) through a mixed strategy where the ascending and descending superoperators are used to compute the operator $o_{\tau}^{[r_{\tau},r_{\tau}+1]}$ and density matrix $\rho_{\tau}^{[r_{\tau},r_{\tau}+1]}$ supported on lattice \mathcal{L}_{τ} which fulfill

$$\text{tr}(o^{[r,r+1]}P) = \text{tr}(o_{\tau}^{[r_{\tau},r_{\tau}+1]}\rho_{\tau}^{[r_{\tau},r_{\tau}+1]}). \quad (18)$$

In all the cases above, one needs to use the ascending/descending superoperators about $T \approx \log N$ times, at a cost $O(\chi^8)$, so that the total computational cost is $O(\chi^8 \log N)$.

C. Evaluation of a sum of two-site operators

In order to compute the expected value

$$\langle O \rangle_{\mathcal{V}_U} \equiv \text{tr}(OP), \quad O \equiv \sum_r o^{[r,r+1]} \quad (19)$$

of an operator O on \mathcal{L} that decomposes as the sum of two-site operators, we can write

$$\text{tr}(OP) = \sum_r \text{tr}(o^{[r,r+1]}P) \quad (20)$$

and individually evaluate each contribution $\text{tr}(o^{[r,r+1]}P)$ by using, e.g., the bottom-top strategy of Sec. III B, with a cost $O(\chi^8 N \log N)$. However, by properly organizing the calculation, the cost of computing $\text{tr}(OP)$ can be reduced to $O(\chi^8 N)$. We next describe how this is achieved. The strategy is

closely related to the computation of expected values in the presence of translation invariance, as discussed later in this section. Again, there are several possible approaches.

Bottom-top. We consider the sequence of operators

$$O_0 \xrightarrow{U_1^\dagger} O_1 \xrightarrow{U_2^\dagger} O_2 \xrightarrow{U_3^\dagger} \cdots O_T, \quad O_0 \equiv O, \quad (21)$$

where the operator O_τ is the sum of $N/3^\tau$ local operators,

$$O_\tau = \sum_{r=1}^{N/3^\tau} o_\tau^{[r,r+1]}. \quad (22)$$

$O_{\tau-1}$ is obtained from O_τ by coarse graining, $O_\tau = U_\tau^\dagger O_{\tau-1} U_\tau$. Each local operator $o_\tau^{[r,r+1]}$ in O_τ is the sum of three local operators from $O_{\tau-1}$ [see (i)–(iii) of Fig. 9], which are lifted to \mathcal{L}_τ by the three different forms of the ascending superoperator, \mathcal{A}_L , \mathcal{A}_C , and \mathcal{A}_R . Since $O_{\tau-1}$ has $N/3^{\tau-1}$ local operators, O_τ is obtained from $O_{\tau-1}$ by using the ascending superoperator \mathcal{A} only $N/3^{\tau-1}$ times. Then, since $\sum_{\tau=0}^T 3^{-\tau} < 2$, this means that the entire sequence of Eq. (21) requires using \mathcal{A} only $O(N)$ times. Once O_T is obtained, the expected value of O follows from

$$\text{tr}(OP) = \text{tr}(O_T). \quad (23)$$

Top-bottom. Here we consider instead the sequence of ensembles of density matrices

$$E_{T-1} \xrightarrow{U_{T-1}} \cdots E_2 \xrightarrow{U_2} E_1 \xrightarrow{U_1} E_0, \quad (24)$$

where E_τ is an ensemble of the $N/3^\tau$ two-site density matrices $\rho_\tau^{[r,r+1]}$ supported on nearest-neighbor sites of \mathcal{L}_τ

$$E_\tau \equiv \{\rho_\tau^{[1,2]}, \rho_\tau^{[2,3]}, \dots, \rho_\tau^{[N/3^\tau, 1]}\}. \quad (25)$$

From each density matrix in the ensemble E_τ we can generate three density matrices in the ensemble $E_{\tau-1}$ by applying the three different forms of the descending superoperator, \mathcal{D}_L , \mathcal{D}_C , and \mathcal{D}_R [see (i)–(iii) of Fig. 10]. All the $N/3^{\tau-1}$ density matrices of ensemble $E_{\tau-1}$ can be obtained from density matrices of E_τ in this way. Since $\sum_{\tau=0}^T 3^{-\tau} < 2$, we see that by using the descending superoperator \mathcal{D} only $O(N)$ times, we are able to compute all the density matrices in the sequence of ensembles of Eq. (24). Once the ensemble $E_0 \equiv E$ has been obtained,

$$E = \{\rho^{[1,2]}, \rho^{[2,3]}, \dots, \rho^{[N,1]}\}, \quad (26)$$

the expected value of O follows from

$$\text{tr}(OP) = \sum_r \text{tr}(o_\tau^{[r,r+1]} \rho_\tau^{[r,r+1]}). \quad (27)$$

Middle ground.—More generally, we could build operator O_τ as well as ensemble E_τ and evaluate the expected value of O from the equality

$$\text{tr}(OP) = \sum_r \text{tr}(o_\tau^{[r,r+1]} \rho_\tau^{[r,r+1]}). \quad (28)$$

Each of the strategies above require the use of the ascending/descending superoperators $O(N)$ times and therefore can indeed be accomplished with cost $O(\chi^8 N)$.

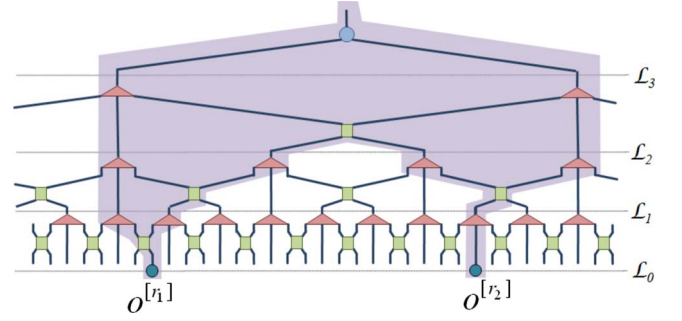


FIG. 16. (Color online) In order to compute a two-point correlator $C_2(r_1, r_2)$, we need to consider the union of the past causal cones of sites r_1 and r_2 . Notice that, in contrast with the case of a single local operator, the joint causal cone of two distant sites typically involves more than two contiguous sites of some lattice \mathcal{L}_τ . This makes the computational cost scale as a power of χ larger than χ^8 .

D. Evaluation of two-point correlators

Let us now consider the computation of a two-point correlator of the form

$$C_2(r_1, r_2) \equiv \langle \Psi | o^{[r_1]} \otimes o^{[r_2]} | \Psi \rangle, \quad (29)$$

where $o^{[r]}$ and $o^{[s]}$ denote two one-site operators applied on two arbitrary sites r and s of \mathcal{L} ; see Fig. 16. Figure 17 shows the tensor network to be contracted. Again, we can use Eq. (2) to remove all disentanglers and isometries that lay outside the joint past causal cone for sites r and s . Then, we can proceed to contract the resulting tensor network, for instance, through a bottom-top or top-bottom approach, with the help of the ascending and descending superoperators (and generalizations thereof). Notice that since at intermediate layers the two legs of the causal cone may contain two sites each one, in general we will need to compute operators or density matrices that span more than just two sites, and the cost of their computation will be larger than $O(\chi^8)$.

However, for specific choices of sites $r, s \in \mathcal{L}$, we are still able to compute $C_2(r, s)$ with overall cost $O(\chi^8 \log N)$, as illustrated in Fig. 18. We emphasize that this was not possible in the binary 1D MERA and is one of the main reasons to work with the ternary 1D MERA. For such choices of sites r and s , each of the two legs of the joint past causal cone contains just one site until, at some layer τ_0 , they fuse into a single two-site leg. We can introduce one-site ascending and descending superoperators $\mathcal{A}^{(1)}$ and $\mathcal{D}^{(1)}$ (Fig. 19), in terms of which we can express, for $\tau \leq \tau_0$, the transformation of a product operator $o_{\tau-1}^{[r]} \otimes o_{\tau-1}^{[s]}$ into a product operator

$$o_\tau^{[r']} \otimes o_\tau^{[s']} = \mathcal{A}^{(1)}(o_{\tau-1}^{[r]}) \otimes \mathcal{A}^{(1)}(o_{\tau-1}^{[s]}), \quad (30)$$

or of a density matrix $\rho_\tau^{[r',s']}$ into a density matrix

$$\rho_{\tau-1}^{[r,s]} = (\mathcal{D}^{(1)} \otimes \mathcal{D}^{(1)})(\rho_\tau^{[r',s]}), \quad (31)$$

where $r, s \in \mathcal{L}_{\tau-1}$ and $r', s' \in \mathcal{L}_\tau$ are sites corresponding to single-site legs of the causal cone. In, say, the bottom-top approach we can compute the correlator of Eq. (29) by using the single-site ascending superoperator $\mathcal{A}^{(1)}$ for layers τ

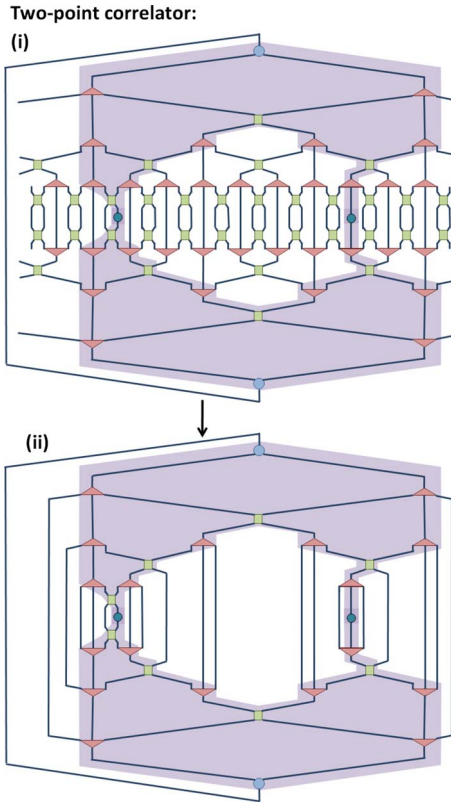


FIG. 17. (Color online) (i) Tensor network to be contracted in order to evaluate a two-point correlator $C_2(r_1, r_2)$. Similar to the case of a local observable in Fig. 3, the tensors outside of the casual cone annihilate in pairs due to their isometric character [Eq. (2)]. (ii) The resulting tensor network is much simpler network. However, for a generic pair of sites $r_1, r_2 \in \mathcal{L}$, the joint past causal cone will contain more than just two sites per layer, resulting in a computational cost that scales with χ as a power larger than χ^8 .

$\leq \tau_0$ and then the two-site ascending superoperator \mathcal{A} for layer $\tau > \tau_0$.

E. Translation invariance

The computation of the expected value $\text{tr}(o^{[r,r+1]}P)$ of a single local operator $o^{[r,r+1]}$ in the case of a translation-invariant MERA can proceed as explained earlier in this section. In the present case one would expect the result to be independent of the sites $[r, r+1] \in \mathcal{L}$ on which the operator is supported, but a finite bond dimension χ typically introduces small space inhomogeneities in the reduced density matrix $\rho^{[r,r+1]}$ and therefore also in $\text{tr}(o^{[r,r+1]}P) = \text{tr}(o^{[r,r+1]}\rho^{[r,r+1]})$.

Given a two-site operator o , an expected value that is independent of $[r, r+1]$ can be obtained by computing an average over sites,

$$\text{tr}(o^{[r,r+1]}P) \rightarrow \frac{1}{N} \sum_r \text{tr}(o^{[r,r+1]}P) \quad (32)$$

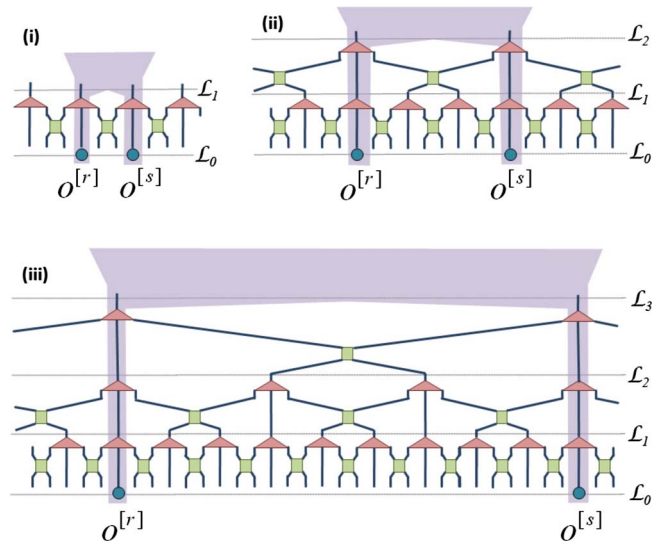


FIG. 18. (Color online) Two-point correlators for specific pairs of sites r, s (at distances of 3^q sites for $q=1, 2, 3, \dots$) can be computed with cost $O(\chi^8 \log N)$. This is due to the fact that the causal cones for each of r, s contains only one site until they meet—(i) at \mathcal{L}_0 , (ii) at \mathcal{L}_1 , or (iii) at \mathcal{L}_2 .

$$= \frac{1}{N} \sum_r \text{tr}(o^{[r,r+1]}\rho^{[r,r+1]}), \quad (33)$$

where the terms $o^{[r,r+1]}$ denote translations of the same operator o . This average can be computed, e.g., by obtaining the N density matrices $\rho^{[r,r+1]}$ individually and then adding them together, with an overall cost $O(\chi^8 N)$. However, with a better organization of the calculation the cost can be reduced to $O(\chi^8 \log N)$.

We first need to introduce average versions of the ascending and descending superoperators. Given a two-site operator

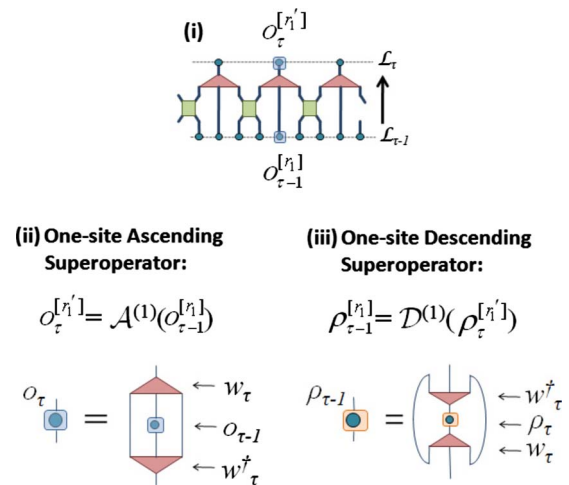


FIG. 19. (Color online) A one-site operator $o_{\tau-1}$ supported on certain sites of $\mathcal{L}_{\tau-1}$ (corresponding to the central wire of an isometry w_{τ}) is mapped onto a single-site operator on \mathcal{L}_{τ} . In this case the (i) ascending and (ii) descending superoperators $\mathcal{A}^{(1)}$ and $\mathcal{D}^{(1)}$ have a very simple form.

$o_{\tau-1}$ in lattice $\mathcal{L}_{\tau-1}$, we can build a two-site operator o_τ by using an average of the three two-site operators resulting from lifting $o_{\tau-1}$ to lattice \mathcal{L}_τ namely, $\mathcal{A}_L(o_{\tau-1})$, $\mathcal{A}_C(o_{\tau-1})$, and $\mathcal{A}_R(o_{\tau-1})$. In terms of the *average ascending superoperator* $\overline{\mathcal{A}}$,

$$\overline{\mathcal{A}} \equiv \frac{1}{3}(\mathcal{A}_L + \mathcal{A}_C + \mathcal{A}_R), \quad (34)$$

this transformation reads

$$o_\tau = \overline{\mathcal{A}}(o_{\tau-1}). \quad (35)$$

Importantly, if we coarse grain the translation-invariant operator

$$\frac{1}{N_{\tau-1}} \sum_r o_{\tau-1}^{[r,r+1]}, \quad N_x \equiv N/3^x, \quad (36)$$

where $N_{\tau-1}$ is the number of sites of $\mathcal{L}_{\tau-1}$ and the terms $o_{\tau-1}^{[r,r+1]}$ denote translations of $o_{\tau-1}$, the resulting operator can be written as

$$\frac{1}{N_\tau} \sum_r o_\tau^{[r,r+1]}, \quad (37)$$

where the terms $o_{\tau-1}^{[r,r+1]}$ denote translations of o_τ and where o_τ and $o_{\tau-1}$ are related through Eq. (35). In other words, the average ascending superoperator $\overline{\mathcal{A}}$ can also be used to characterize the coarse graining, in the translation-invariant case, of operators of the form of Eq. (19).

Let $\bar{\rho}_\tau$ denote the two-site density matrix obtained by averaging over all density matrices $\rho_\tau^{[r,r+1]}$ on different pairs $[r, r+1]$ of two contiguous sites of \mathcal{L}_τ

$$\bar{\rho}_\tau \equiv \frac{1}{N_\tau} \sum_r \rho_\tau^{[r,r+1]}, \quad (38)$$

and similarly for lattice $\mathcal{L}_{\tau-1}$,

$$\bar{\rho}_{\tau-1} \equiv \frac{1}{N_{\tau-1}} \sum_r \rho_{\tau-1}^{[r,r+1]}. \quad (39)$$

Recall that each density matrix $\rho_\tau^{[r,r+1]}$ on lattice \mathcal{L}_τ gives rise to three density matrices in $\mathcal{L}_{\tau-1}$ according to the three versions of the descending superoperator, namely, \mathcal{D}_L , \mathcal{D}_C , and \mathcal{D}_R . It follows that the density matrix $\bar{\rho}_{\tau-1}$ can be obtained from the density matrix $\bar{\rho}_\tau$ by using the *average descending superoperator*,

$$\overline{\mathcal{D}} \equiv \frac{1}{3}(\mathcal{D}_L + \mathcal{D}_C + \mathcal{D}_R), \quad (40)$$

that is,

$$\bar{\rho}_{\tau-1} = \overline{\mathcal{A}}(\bar{\rho}_\tau). \quad (41)$$

We can now proceed to compute the average expected value of Eqs. (32) and (33). This can be accomplished in several alternative ways.

Bottom-top. Given a two-site operator o , we compute a sequence of increasingly coarse-grained operators

$$\overline{\mathcal{A}} \quad \overline{\mathcal{A}} \quad \overline{\mathcal{A}} \\ o_0 \rightarrow o_1 \rightarrow o_2 \rightarrow \cdots \rightarrow o_T, \quad o_0 \equiv o, \quad (42)$$

where o_τ is obtained from $o_{\tau-1}$ by means of the average ascending superoperator $\overline{\mathcal{A}}$. Then we simply have

$$\frac{1}{N} \sum_r \text{tr}(o^{[r,r+1]} P) = \text{tr}(o_\tau). \quad (43)$$

Top-bottom. Alternatively, we can compute the sequence of average density matrices

$$\overline{\mathcal{D}} \quad \overline{\mathcal{D}} \quad \overline{\mathcal{D}} \\ \bar{\rho}_{T-1} \rightarrow \cdots \rightarrow \bar{\rho}_2 \rightarrow \bar{\rho}_1 \rightarrow \bar{\rho}_0, \quad (44)$$

where $\bar{\rho}_{\tau-1}$ is obtained from $\bar{\rho}_\tau$ by means of the average descending superoperator $\overline{\mathcal{D}}$ and where $\bar{\rho} \equiv \bar{\rho}_0$ corresponds to the average density matrix on lattice \mathcal{L} ,

$$\bar{\rho} \equiv \frac{1}{N} \sum_r \rho^{[r,r+1]}, \quad (45)$$

in terms of which we can express the average expected value as

$$\frac{1}{N} \sum_r \text{tr}(o^{[r,r+1]} P) = \text{tr}(o \bar{\rho}). \quad (46)$$

Middle ground. As customary, we can also use both $\overline{\mathcal{A}}$ and $\overline{\mathcal{D}}$ to compute o_τ and $\bar{\rho}_\tau$ and evaluate the average expected value as

$$\frac{1}{N} \sum_r \text{tr}(o^{[r,r+1]} P) = \text{tr}(o_\tau \bar{\rho}_\tau). \quad (47)$$

In all the above strategies the average ascending and descending superoperators $\overline{\mathcal{A}}$ and $\overline{\mathcal{D}}$ are used $O(\log(N))$ times and therefore the computational cost scales as $O(\chi^8 \log N)$.

To summarize, with a translation-invariant MERA we can coarse grain a single two-site operator o (with a transformation that involves averaging over all possible causal cones) or compute the average density matrix $\bar{\rho}$ by using the average ascending or descending superoperators. This leads to a sequence of operators o_τ and density matrices $\bar{\rho}_\tau$

$$\overline{\mathcal{A}} \quad \overline{\mathcal{A}} \quad \overline{\mathcal{A}} \\ o_0 \rightarrow o_1 \rightarrow \cdots \rightarrow o_T, \quad o_0 \equiv o, \quad (48)$$

$$\overline{\mathcal{D}} \quad \overline{\mathcal{D}} \quad \overline{\mathcal{D}} \\ \bar{\rho}_0 \leftarrow \bar{\rho}_1 \leftarrow \cdots \leftarrow \bar{\rho}_T, \quad \bar{\rho}_0 \equiv \bar{\rho}, \quad (49)$$

from which the expected value of o is obtained as $\text{tr}(o \bar{\rho})$, as $\text{tr}(o_T)$ or, more generally, as $\text{tr}(o_\tau \bar{\rho}_\tau)$.

F. Scale invariance

In the case of a translation-invariant MERA that is also scale invariant, the average ascending superoperator $\overline{\mathcal{A}}$ is identical on each layer τ , since it is always made of the same disentangler u and isometry w . We then refer to it as the *scaling superoperator* \mathcal{S} .¹¹ Its dual \mathcal{S}^* corresponds to the descending superoperator $\overline{\mathcal{D}}$.

As derived in Ref. 10, the expected value of a local observable o in the thermodynamic limit can be obtained by analyzing the spectral decomposition of the scaling superoperator \mathcal{S} ,

$$\mathcal{S}(\bullet) = \sum_\alpha \lambda_\alpha \phi_\alpha \text{tr}(\hat{\phi}_\alpha \bullet), \quad \text{tr}(\hat{\phi}_\alpha \phi_\beta) = \delta_{\alpha\beta}. \quad (50)$$

We refer to the eigenoperators ϕ_α of \mathcal{S} ,

$$\mathcal{S}(\phi_\alpha) = \lambda_\alpha \phi_\alpha, \quad (51)$$

as the *scaling operators*. Notice that the operators $\hat{\phi}_\alpha$, which are biorthonormal to the operators ϕ_α , are eigenoperators of \mathcal{S}^* ,

$$\mathcal{S}^*(\hat{\phi}_\alpha) = \lambda_\alpha \hat{\phi}_\alpha. \quad (52)$$

We recall that the scaling operator \mathcal{S} is made of isometric tensors [cf. Eq. (2)] and therefore the identity operator \mathbb{I} is an eigenoperator of \mathcal{S} with eigenvalue 1 (that is, \mathcal{S} is unital),

$$\mathcal{S}(\mathbb{I}) = \mathbb{I}. \quad (53)$$

On the other hand, since the MERA is built as a quantum circuit—and descending through the causal cone corresponds to advancing in the time of a quantum evolution—it is obvious that the descending superoperator \mathcal{D} is a quantum channel, and so are $\overline{\mathcal{D}}$ and \mathcal{S}^* (see also Ref. 10). In particular, \mathcal{S}^* is a contractive superoperator,²³ which means that the eigenvalues λ_α in Eq. (50) are constrained to fulfill $|\lambda_\alpha| \leq 1$. In practical simulations¹¹ one finds that the identity operator \mathbb{I} is the only eigenoperator of \mathcal{S} with eigenvalue of 1, $\lambda_{\mathbb{I}}=1$, and that $|\lambda_\alpha| < 1$ for $\alpha \neq \mathbb{I}$. Let $\hat{\rho}$ denote the corresponding unique fixed point of \mathcal{S}^* ,

$$\mathcal{S}^*(\hat{\rho}) = \hat{\rho}. \quad (54)$$

In an infinite system, the local density matrix of any lattice \mathcal{L}_τ (with finite τ) results from applying \mathcal{S}^* on ρ^T an infinite number of times, and it is therefore equal to the fixed point $\hat{\rho}$.¹⁰ Consequently, Eqs. (48) and (49) are then replaced with

$$\begin{array}{c} \mathcal{S} \quad \mathcal{S} \quad \mathcal{S} \\ o_0 \rightarrow o_1 \rightarrow o_2 \rightarrow o_3 \cdots, \quad o_0 \equiv o, \end{array} \quad (55)$$

$$\begin{array}{c} \mathcal{S}^* \quad \mathcal{S}^* \quad \mathcal{S}^* \\ \hat{\rho} \leftarrow \hat{\rho} \leftarrow \hat{\rho} \leftarrow \hat{\rho} \cdots, \end{array} \quad (56)$$

where in addition, by decomposing o in terms of the scaling operators ϕ_α ,

$$o = \sum_{\alpha} c_{\alpha} \phi_{\alpha}, \quad c_{\alpha} \equiv \text{tr}(\hat{\phi}_{\alpha} o), \quad (57)$$

we can explicitly compute o_τ :

$$o_\tau = (\underbrace{\mathcal{S} \circ \cdots \circ \mathcal{S}}_{\tau \text{ times}})(o) = \sum_{\alpha} c_{\alpha} (\lambda_{\alpha})^{\tau} \phi_{\alpha}. \quad (58)$$

This expression shows that unless $c_1 \neq 0$, the operator o_τ decreases exponentially with τ (recall that $|\lambda_\alpha| < 1$ for $\alpha \neq \mathbb{I}$) and its expected value must vanish. The average expected value of o then reads

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_r \text{tr}(o^{[r,r+1]} P) = \text{tr}(o \hat{\rho}). \quad (59)$$

Two-point correlators for selected positions can also be expressed in a simple way, by considering the *one-site scaling superoperator* $\mathcal{S}^{(1)}$, which is how we refer to the superoperator $\mathcal{A}^{(1)}$ in Fig. 19 in the case of a scale-invariant MERA. Its spectral decomposition,

$$\mathcal{S}^{(1)}(\bullet) = \sum_{\alpha} \mu_{\alpha} \psi_{\alpha} \text{tr}(\hat{\psi}_{\alpha} \bullet), \quad \text{tr}(\hat{\psi}_{\alpha} \psi_{\beta}) = \delta_{\alpha\beta}, \quad (60)$$

provides us with a new set of (one-site) scaling operators ψ_{α} . Given two arbitrary one-site operators o and o' , we can always decompose them in terms of these ψ_{α} [similarly as in Eq. (57)]. Thus we can focus directly on a correlator of the form $\langle \psi_{\alpha}^{[r]} \psi_{\beta}^{[s]} \rangle$. Here r and s are restricted to selected positions as in Fig. 18. Then we have

$$\langle \psi_{\alpha}^{[r]} \psi_{\beta}^{[s]} \rangle = \frac{C_{\alpha\beta}}{|r-s|^{\Delta_{\alpha} + \Delta_{\beta}}}, \quad (61)$$

where $\Delta_{\alpha} \equiv \log_3 \mu_{\alpha}$ is the *scaling dimension* of the scaling operator ψ_{α} , whereas $C_{\alpha\beta}$ is given by

$$C_{\alpha\beta} \equiv \text{tr}[(\psi_{\alpha} \otimes \psi_{\beta}) \hat{\rho}], \quad (62)$$

with $\hat{\rho}$ from Eq. (54).

In deriving Eq. (61) we have used that, by construction, $|r-s|=3^q$ for some $q=1,2,3,\dots$. Coarse graining $\psi_{\alpha}^{[r]} \psi_{\beta}^{[s]}$ a number q of times produces a multiplicative factor $(\mu_{\alpha} \mu_{\beta})^q$ and the residual two-site operator $\psi_{\alpha}^{[0]} \psi_{\beta}^{[1]}$, whose expected value gives $C_{\alpha\beta}$. On the other hand, by noting that $\mu^q = \mu^{\log_3 |r-s|} = |r-s|^{\log_3 \mu} = |r-s|^{\log_3 \Delta}$, we arrive at $(\mu_{\alpha} \mu_{\beta})^q = |r-s|^{\Delta_{\alpha} + \Delta_{\beta}}$, which explains the denominator in Eq. (61).

We note that the polynomial decay of correlators in the scale-invariant MERA was established in Ref. 5. Their connection with the eigenvalues of the scaling superoperator [Eq. (50)] was formalized in Ref. 10. Its closed expression in Eq. (61), including the coefficients $C_{\alpha\beta}$, was derived in Ref. 11, where also three-point correlators were considered. (Reference 11 also unveiled a connection between the scale-invariant MERA and conformal field theory. The coefficients $C_{\alpha\beta}$ of two-point correlators and the analogous for three-point correlators are the key to identifying the operator algebra of primary fields and their towers of descendant fields.)

In conclusion, from the scale-invariant MERA we can characterize the expected value of local observables and two-point correlators, as expressed in Eqs. (59), (61), and (62). All critical exponents of the theory can be extracted from the scaling dimensions Δ_{α} . The required manipulations include computing $\bar{\rho}$ from \mathcal{S} (using sparse diagonalization techniques) and diagonalizing $\mathcal{S}^{(1)}$, all of which can be accomplished with the ternary 1D MERA with cost $O(\chi^8)$.

IV. OPTIMIZATION OF A DISENTANGLER OR ISOMETRY

In preparation for the algorithms to be described in Sec. V, here we explain how to optimize a single tensor of the MERA. Let H be a Hamiltonian made of nearest-neighbor two-site interactions $h^{[r,r+1]}$,

$$H = \sum_r h^{[r,r+1]}. \quad (63)$$

For purposes of the optimization below, we choose each term $h^{[r,r+1]}$ so that it has no positive eigenvalues, $h^{[r,r+1]} \leq 0$. (This can be achieved with the simple replacement $h^{[r,r+1]} \rightarrow h^{[r,r+1]} - \lambda_{\max} I$, where λ_{\max} is the largest eigenvalue of $h^{[r,r+1]}$.)

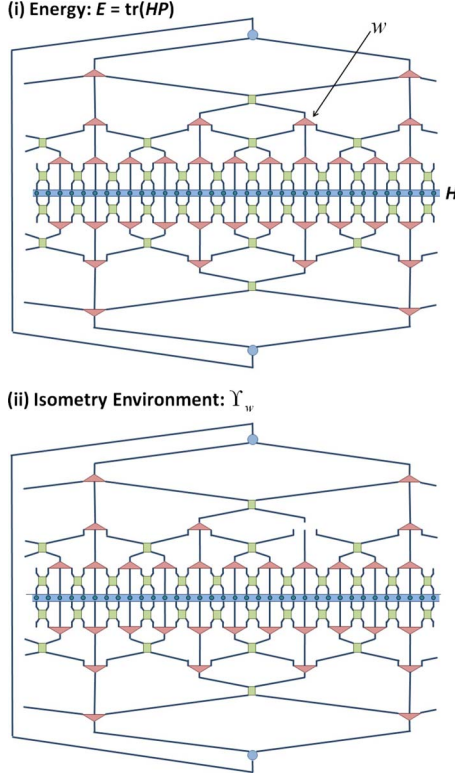


FIG. 20. (Color online) (i) The energy of a MERA, defined as $E \equiv \text{tr}(HP)$, is represented explicitly in terms of a tensor network. The removal of an isometry w from this network gives (ii) the environment Y_w for w (and similarly for disentanglers u). By construction we have that $E = \text{tr}(wY_w)$.

Our goal for the time being will be to minimize the energy [Fig. 20(i)],

$$E \equiv \text{tr}(HP), \tag{64}$$

where P is a projector onto the χ_T -dimensional subspace $V_U \in V_{\mathcal{L}}$ [see Eq. (7)] by modifying only one of the tensors of the MERA. The optimization of a disentangler u is very similar to that of an isometry w , and we can focus on describing the latter in more detail.

Suppose then that given a MERA, we want to optimize an isometry w while keeping the rest of the tensors fixed. The cost function E is quadratic in w (more specifically, it depends bilinearly on w and w^\dagger),

$$E(w) = \text{tr} \left(\sum_s w M_s w^\dagger N_s \right) + c_1, \tag{65}$$

where M_s and N_s are two sets of matrices and c_1 is a constant [that originates in all the Hamiltonian terms of Eq. (63) outside the future causal cone of w]. Unfortunately there is no known algorithm to solve a quadratic problem subject to the additional isometric constraint of Eq. (2). One can, however, attempt several approximate strategies (see Ref. 18 for some possibilities). Here we describe an iterative approach based on linearizing the cost function $E(w)$.

Enviroments of an isometry:

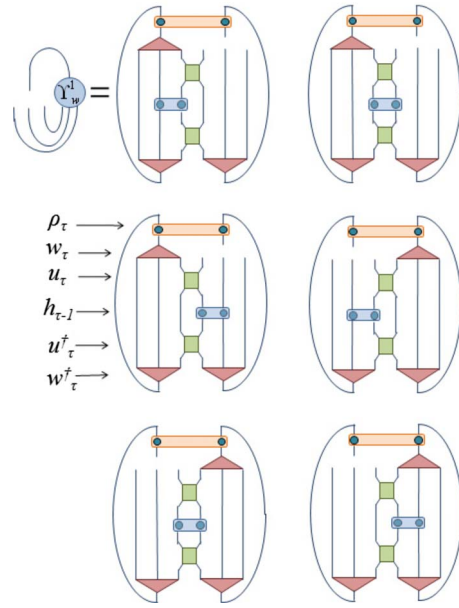


FIG. 21. (Color online) Tensor network corresponding to the 6 different contributions Y_w^i to the environment $Y_w = \sum_{i=1}^6 Y_w^i$ of the isometry w . Notice that at each iteration of L1-L4 we need to recompute each Y_w^i since it depends on the updated w^\dagger . Nevertheless, the Hamiltonian term and density matrix that appears in Y_u remain the same throughout the optimization and only need to be computed once.

In this approach, we temporarily regard w and w^\dagger as independent tensors, and optimize w while keeping w^\dagger fixed. The cost function reads, up to the irrelevant constant, simply

$$E^*(w) \equiv \text{tr}(wY_w), \quad Y_w \equiv \sum_s M_s w^\dagger N_s, \tag{66}$$

where we call the matrix Y_w the *environment* of the isometry w and we treat it as if it was independent of w . $E^*(w)$ is then minimized by the choice $w = -WV^\dagger$, where V and W are the unitary transformations in the singular value decomposition of the environment, $Y_w = VSW^\dagger$,

$$\min_w E^*(w) = \min_w \text{tr}(wVSW^\dagger) = -\text{tr}(S) = -\sum_\alpha s_\alpha \tag{67}$$

(here $s_\alpha \geq 0$ are the singular values of Y_w).

Accordingly, given an initial isometry w , the optimization is performed by iterating the following four steps q_{one} times:

L1. Compute the environment Y_w with the newest version of w^\dagger (as explained below; see also Fig. 21).

L2. Compute the singular value decomposition $Y_w = VSW^\dagger$.

L3. Compute the new isometry $w' = -WV^\dagger$.

L4. Replace w^\dagger with w'^\dagger .

The environment Y_w of an isometry w (at layer τ) can be decomposed as the sum of six contributions Y_w^i ($i=1, \dots, 6$), each one expressed as a tensor network that involves neighboring isometric tensors of the same layer τ (disentanglers and isometries) as well as one Hamiltonian

Environments of a disentangler:

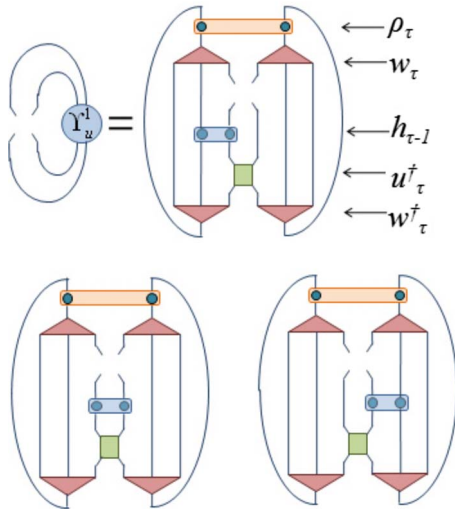


FIG. 22. (Color online) Tensor networks corresponding to the three different contributions Y_u^i to the environment $Y_u = \sum_{i=1}^3 Y_u^i$ of the disentangler u . Notice that at each iteration of L1–L4 we need to recompute each Y_u^i since it depends on the updated u^\dagger . Nevertheless, the Hamiltonian term and density matrix that appears in Y_u remain the same throughout the optimization and only need to be computed once.

term $h_{\tau-1}^{[r,r+1]}$ and one density matrix $\rho_\tau^{[r',r'+1]}$; see Fig. 21. The two-site Hamiltonian term $h_{\tau-1}^{[r,r+1]}$ collects contributions from all the Hamiltonian terms in Eq. (63) included in the future causal cone of the sites $r, r+1$ of $\mathcal{L}_{\tau-1}$ and is computed with the help of the ascending superoperator \mathcal{A} . Similarly, the two-site density matrix $\rho_\tau^{[r',r'+1]}$ is computed with the help of the descending superoperator \mathcal{D} . The computation of $h_{\tau-1}^{[r,r+1]}$ and $\rho_\tau^{[r',r'+1]}$, which only needs to be performed once during the optimization of w , has a cost $O(\chi^8 \log N)$.

On the other hand, once we have $h_{\tau-1}^{[r,r+1]}$ and $\rho_\tau^{[r',r'+1]}$, computing Y_w has a cost $O(\chi^8)$ and needs to be repeated at each iteration of steps L1–L4, with a total cost $O(\chi^8 q_{\text{one}})$. In actual MERA simulations we find that the cost function E_w typically drops very close to the eventual minimum already after a small number of iterations q_{one} on the order of 10.

The optimization of a disentangler u is achieved analogously, but in this case the environment Y_u decomposes into three contributions Y_u^i ($i=1, 2, 3$); see Fig. 22. The required Hamiltonian terms and density matrices can be computed at a cost $O(\chi^8 \log N)$, while the optimization of u following steps L1–L4 has a cost $O(\chi^8 q_{\text{one}})$.

V. OPTIMIZATION OF THE MERA

In this section we explain a simple algorithm to optimize the MERA so that it minimizes the energy of a local Hamiltonian of form (63). We first describe the algorithm for a generic system, and then discuss a number of specialized variations. These are directed to exploit translation invariance and scale invariance and to simulate systems where there is a finite range of correlations.

A. Algorithm

The basic idea of the algorithm is to attempt to minimize the cost function of Eq. (64) by sequentially optimizing individual tensors of the MERA, where each tensor is optimized as explained in Sec. IV.

By choosing to sweep the MERA in an organized way, we are able to update all its $O(N)$ tensors once with cost $O(\chi^8 N)$. Here we describe a bottom-top approach where the MERA is updated layer by layer, starting with the bottom layer $\tau=1$ and progressing upward all the way to the top layer (top-bottom and combined approaches are also possible).

Given a starting MERA and the Hamiltonian of Eq. (63), a bottom-top sweep is organized as follows:

A1. Compute all two-site density matrices $\rho_\tau^{[r,r+1]}$ for all layers τ and sites $r \in \mathcal{L}_\tau$. Starting from the lowest layer and for growing values of $\tau=1, 2, \dots, T-1$, repeat the following two steps.

A2. Update all disentanglers u and isometries w of layer τ .

A3. Compute all two-site Hamiltonian terms $h_\tau^{[r,r+1]}$ for layer τ .

Then, finally,

A4. Update the top tensor of the MERA.

In step A1, we compute all nearest-neighbor reduced density matrices $\rho_\tau^{[r,r+1]}$ for all the lattices \mathcal{L}_τ so that they can be used in step A2. We first compute the density matrix for the two sites of \mathcal{L}_{T-1} as explained in Fig. 14. Then we use the descending superoperator \mathcal{D} to compute the six possible nearest-neighbor two-site density matrices of \mathcal{L}_{T-2} . More generally, given all the relevant density matrices of \mathcal{L}_τ we use \mathcal{D} to obtain all the relevant density matrices of $\mathcal{L}_{\tau-1}$. In this way, the number of operations is proportional to the number of computed density matrices, namely, $O(N)$, and the total cost is $O(\chi^8 N)$.

Step A2 sweeps into a sequence of single-tensor optimizations that sweeps a given layer τ of the MERA. Each individual optimization in that layer is performed as explained in Sec. IV. Note that in order to optimize, say, an isometry w , we build its environment Y_w by using (i) the density matrices computed in step A1; (ii) the Hamiltonian terms that were either given at the start for $\tau=1$ or have been computed in A3 for $\tau>1$; and (iii) the neighboring disentanglers and isometries within the layer τ . We can proceed, for instance, by updating all disentanglers of the layer from left to right, and then update all the isometries. This can be repeated a number q_{lay} of times until the cost function does not change significantly.

In step A3, the new disentanglers and isometries of layer τ are used to build the ascending superoperator \mathcal{A} , which we then apply to the Hamiltonian terms of layer $\tau-1$ to compute the Hamiltonian terms for layer τ . As explained after Eq. (22), each Hamiltonian term in layer τ is built from three contributions from layer $\tau-1$. In step A4, the optimized top tensor corresponds to the χ_T eigenvectors with smaller energy eigenvalues of the Hamiltonian of the two-site lattice \mathcal{L}_{T-1} , obtained by exact diagonalization.

The overall optimization of the MERA consists of iterating steps A1–A4 until some pre-established degree of convergence in the energy E is achieved. Suppose this occurs after q_{iter} iterations. Then the cost of the optimization scales

as $O(\chi^8 N q_{\text{one}} q_{\text{lay}} q_{\text{iter}})$. We observe that it is often convenient to keep q_{one} and q_{lay} relatively small (say, between 1 and 5), since it is not worth spending much effort optimizing a single tensor or layer that will have to be optimized again later on with a modified cost function.

B. Translation-invariant systems

When the Hamiltonian H is invariant under translations, we can use a translation-invariant MERA.²⁰ In this case, each layer τ is characterized by a disentangler u_τ and an isometry w_τ . In addition, on each lattice \mathcal{L}_τ we have one two-site Hamiltonian h_τ and one average density matrix $\bar{\rho}_\tau$. Then a bottom-top sweep of the MERA breaks into steps A1–A4 for the inhomogeneous case above, but with the following simplifications.

In step A1, we compute $\bar{\rho}_{\tau-1}$ from $\bar{\rho}_\tau$ using the average descending superoperator \bar{D} of Eq. (40),

$$\bar{\rho}_\tau \rightarrow \bar{\rho}_{\tau-1} = \bar{D}(\bar{\rho}_\tau). \quad (68)$$

Then the whole sequence $\{\bar{\rho}_{T-1}, \dots, \bar{\rho}_1, \bar{\rho}_0\}$, with $T \approx \log N$, is computed with cost $O(\chi^8 \log N)$.

In step A2, the minimization of the energy E by optimizing, say, the isometry w_τ is no longer a quadratic problem (since a larger power of w_τ appears now in the cost function). Nevertheless, we still linearize the cost function E and optimize w_τ according to steps L1–L4 of Sec. IV. Namely, we build the environment Y_w (which now contains copies of w_τ and w_τ^\dagger , all of them treated as frozen), compute its singular value decomposition to build the optimal w'_τ , and then replace w_τ and w_τ^\dagger with w'_τ and w'^{\dagger}_τ in the tensor network for Y_w before starting the next iteration.

In step A3, the new Hamiltonian term h_τ is obtained from $h_{\tau-1}$ using the average ascending superoperator \bar{A} ,

$$h_{\tau-1} \rightarrow h_\tau = \bar{A}(h_{\tau-1}). \quad (69)$$

Step A4 proceeds as in the inhomogeneous case. The overall cost of optimizing the MERA is in this case $O(\chi^8 \log(N) q_{\text{one}} q_{\text{lay}} q_{\text{iter}})$.

C. Scale-invariant systems

Given the Hamiltonian H for an infinite lattice at a quantum critical point, where we expect the system to be invariant under rescaling, we can use a scale-invariant MERA to represent its ground state.^{1,5–7,10,11} (A scale-invariant MERA is also relevant in the context of topological order in the infrared limit of the RG flow,^{12,13} both for finite and infinite systems. We will not consider such systems here.)

Let us assume that all disentanglers and isometries are copies of a unique pair (u, w) . Then, as explained in Ref. 11, the optimization algorithm can be specialized to take advantage of scale invariance as follows.

In step A1, we apply sparse diagonalization techniques to compute the fixed-point density matrix $\hat{\rho}$ from the superoperator \mathcal{S}^* . This amounts to applying \mathcal{S}^* a number of times and therefore can be accomplished with cost $O(\chi^8)$.

In step A2, the environment for, e.g., the isometry w , Y_w , is computed as a weighted sum of environments for different

layers $\tau=1, 2, \dots$. In a translation-invariant MERA the environment for layer τ is a function $f(u_\tau, w_\tau, \rho_\tau, h_{\tau-1})$ of the pair (u_τ, w_τ) , the density matrix ρ_τ and the Hamiltonian term $h_{\tau-1}$ (specifically, f is the sum of the diagrams in Fig. 21). A scale-invariant MERA corresponds to the replacements

$$(u_\tau, w_\tau) \rightarrow (u, w), \quad \rho_\tau \rightarrow \hat{\rho}, \quad (70)$$

so that only $h_{\tau-1}$ retains dependence on τ . We then choose the average environment

$$Y_w \equiv \sum_{\tau=1}^{\infty} \frac{1}{3^\tau} f(u, w, \hat{\rho}, h_{\tau-1}), \quad (71)$$

where the weight $1/3^\tau$ reflects the fact that for each isometry at layer τ there are three isometries at layer $\tau-1$. Using linearity of the f in its fourth argument we arrive at

$$Y_w = f(u, w, \hat{\rho}, \bar{h}), \quad \bar{h} \equiv \sum_{\tau=1}^{\infty} \frac{1}{3^\tau} h_{\tau-1}. \quad (72)$$

In practice, only a few terms of the expansion of \bar{h} (say, $\tau=1, 2, 3, 4$) seem to be necessary. Given Y_w , the optimization proceeds as usual with a singular value decomposition.

Steps A3 and A4 are not necessary. That is, the algorithm used to minimize the expected value of H consists simply in iterating the following two steps:

ScInv1. Given a pair (u, w) , compute a pair $(\hat{\rho}, \bar{h})$.

ScInv2. Given the pairs (u, w) and $(\hat{\rho}, \bar{h})$, update the pair (u, w) .

In practical simulations it is convenient to include a few (say, one or two) transitional layers at the bottom of the MERA, each one characterized by a different pair (u_τ, w_τ) . In this way the bond dimension χ of the MERA can be made independent of the dimension d of the sites of \mathcal{L} . These transitional layers are optimized using the algorithm for translation-invariant systems.

D. Finite range of correlations

A third variation in the basic algorithm consists of setting the number T' of layers in the MERA to a value smaller than its usual one, $T \approx \log_3 N$ (in such a way that the number $N_{T'}$ of sites on the top lattice $\mathcal{L}_{T'}$ may still be quite large) and to consider a state $|\Psi\rangle$ of the lattice \mathcal{L} such that after T' coarse-graining transformations it has become a product state,

$$|\Psi_0\rangle \rightarrow |\Psi_1\rangle \rightarrow \dots \rightarrow |\Psi_{T'}\rangle, \quad (73)$$

where $|\Psi_{T'}\rangle = |0\rangle^{\otimes N_{T'}}$. For instance, Fig. 23 shows a MERA for $N=36$ and $T'=2$. The four top tensors of this MERA are of type (0,3), where the lack of upper index indicates that the top lattice $\mathcal{L}_{T'}$ is in a product state of its $N_{T'}=4$ sites.

We refer to this ansatz as the *finite-range* MERA, since it is such that correlations in $|\Psi\rangle$ are restricted to a finite range ζ , roughly $\zeta \approx 3^{T'}$ sites, in the sense that regions separated by more than ζ sites display no correlations. This is due to the fact that the past causal cones of distance regions of \mathcal{L} have zero intersection, see Fig. 23.

Given a ground state $|\Psi\rangle$ with a finite correlation length ξ , the finite-range MERA with $\zeta \approx \xi$ turns out to be a better

Finite range MERA:

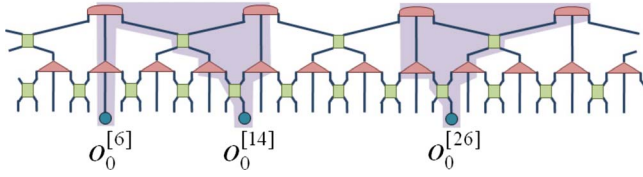


FIG. 23. (Color online) A *finite-correlation-range* MERA with $T'=2$ layers is used to represent a state of $N=36$ sites. Since it lacks the uppermost layers, only sites within a finite distance or range $\zeta \approx 3^{T'}$ of each other may be correlated. More precisely, only pairs of sites (r_1, r_2) whose past casual cones intersect can be correlated, as it is the case with the pair of sites (6,14) but not with (14,26), for which we have $\langle o^{[14]} o^{[26]} \rangle = \langle o^{[14]} \rangle \langle o^{[26]} \rangle$.

option to represent $|\Psi\rangle$ than the standard MERA with $T \approx \log_3 N$ layers, in that it offers a more compact description and the cost of the simulations is also lower since there are less tensors to be optimized. The algorithm is adapted in a straightforward way. The only significant difference is that the top isometries, being of type (0,3), do not require any density matrix in their optimization (their environment is only a function of neighboring disentanglers and isometries, and of Hamiltonian terms).

A clear advantage of the finite-range MERA is in a translation-invariant system, where the cost of a simulation with range $\zeta=3^{T'}$ is $O(\chi^8 \log_3 \zeta)$, that is, independent of N . This allows us to take the limit of an infinite system. We find that given a translation-invariant Hamiltonian $H = \sum_{r=1}^N h^{[r,r+1]}$, where $h^{[r,r+1]}$ is the same for all $r \in \mathcal{L}$, the optimization of a finite-range MERA will lead to the same collection of optimal disentanglers and isometries $\{(u_1, w_1), (u_2, w_2), \dots, (u_{T'}, w_{T'})\}$, for different lattice sizes N, N', N'', \dots larger than ζ . This is due to the existence of disconnected causal cones, which imply that the cost functions for the optimization are not sensitive to the total system size provided it is larger than ζ . As a result, $\{(u_1, w_1), (u_2, w_2), \dots, (u_{T'}, w_{T'})\}$ can be used to define not just one but a whole collection of states $|\Psi(N)\rangle, |\Psi(N')\rangle, |\Psi(N'')\rangle, \dots$, for lattices of different sizes N, N', N'', \dots , such that they all have the same two-site density matrix ρ and therefore also the same expected value of the energy per link,

$$\langle \Psi(N) | h | \Psi(N) \rangle = \langle \Psi(N') | h | \Psi(N') \rangle = \dots \quad (74)$$

In particular, we can use the finite-range MERA algorithm to obtain an upper bound for the ground-state energy of an infinite system, even though only T' pairs (u_τ, w_τ) are optimized.

VI. BENCHMARK CALCULATIONS FOR 1D SYSTEMS

In order to benchmark the algorithms of Sec. V, we have analyzed zero-temperature, low-energy properties of a number of 1D quantum spin systems. Specifically, we have considered the Ising model,²⁴ the three-state Potts model,²⁵ the XX model,²⁶ and the Heisenberg models,²⁷ with Hamiltonians

$$H_{\text{Ising}} = \sum_r (\lambda \sigma_z^{[r]} + \sigma_x^{[r]} \sigma_x^{[r+1]}), \quad (75)$$

$$H_{\text{Potts}} = \sum_r \left(\lambda M_z^{[r]} + \sum_{a=1,2} M_{x,a}^{[r]} M_{x,3-a}^{[r+1]} \right), \quad (76)$$

$$H_{\text{XX}} = \sum_r (\sigma_x^{[r]} \sigma_x^{[r+1]} + \sigma_y^{[r]} \sigma_y^{[r+1]}), \quad (77)$$

$$H_{\text{Heisenberg}} = \sum_r (\sigma_x^{[r]} \sigma_x^{[r+1]} + \sigma_y^{[r]} \sigma_y^{[r+1]} + \sigma_z^{[r]} \sigma_z^{[r+1]}), \quad (78)$$

where σ_x , σ_y , and σ_z are the spin-1/2 Pauli matrices and $M_{x,1}$, $M_{x,2}$, and M_z are the matrices

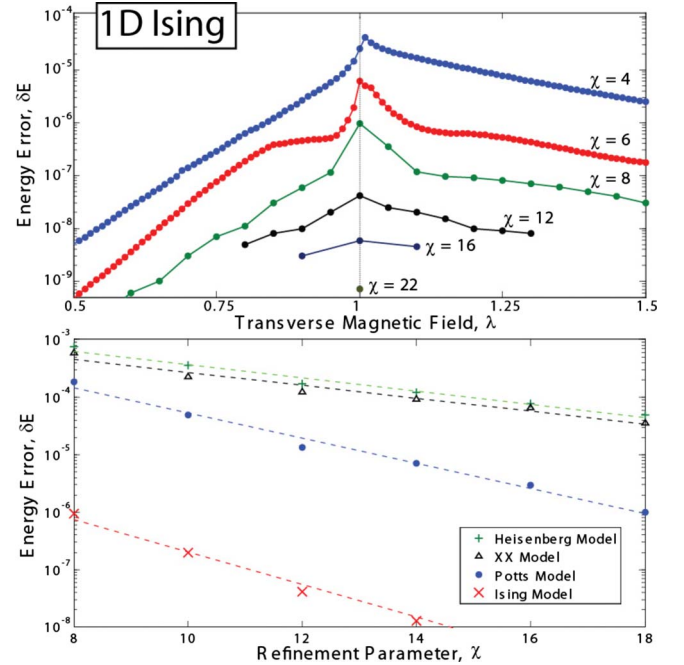


FIG. 24. (Color online) (Top) The energy error of the MERA approximations to the ground state of the infinite Ising model, as compared against exact analytic values, is plotted both for different transverse magnetic field strengths and different values of the MERA refinement parameter χ . The finite-correlation-range algorithm (with at most $T=5$ levels) was used for noncritical ground states, while the scale-invariant MERA was used for simulations at the critical point. It is seen that representing the ground state is most computationally demanding at the critical point, although even at criticality the MERA approximates the ground state to between five digits of accuracy ($\chi=4$) and ten digits of accuracy ($\chi=22$). (Bottom) Scale-invariant MERA are used to compute the ground states of infinite, critical, 1D spin chains of Eqs. (75)–(78) for several values of χ . In all instances one observes a roughly exponential convergence in energy over a wide range of values for χ as indicated by trend lines (dashed line). Energy errors for Ising, XX, and Heisenberg models are taken relative to the analytic values for ground energy, while energy errors presented for the Potts model are taken relative to the energy of a $\chi=22$ simulation.

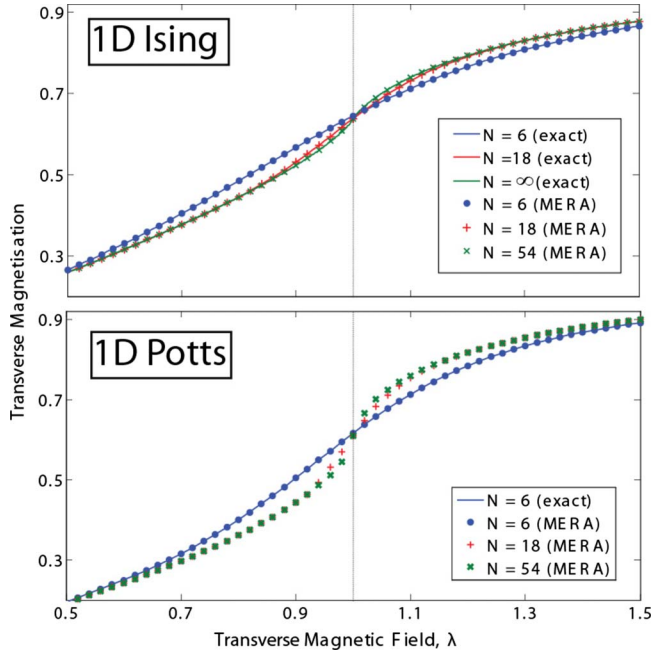


FIG. 25. (Color online) The transverse magnetization $\langle \sigma_z \rangle$ for Ising and $\frac{1}{2}\langle M_z \rangle$ for Potts, is plotted for translation-invariant chains of several sizes N . (Top) For the Ising model, the magnetization given from $\chi=8$ MERA matches those from exact diagonalization for small system sizes ($N=6, 18$), while the magnetization from the $N=54$ MERA approximates that from the thermodynamic limit (known analytically). (Bottom) Equivalent magnetizations for the Potts model, here computed with a $\chi=12$ MERA. Simulations with larger N systems show little change from the $N=54$ data, again indicating that $N=54$ is already close to the thermodynamic limit.

$$M_z \equiv \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad (79)$$

$$M_{x,1} \equiv \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad M_{x,2} \equiv \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \quad (80)$$

We assume periodic boundary conditions in all instances and use a translation-invariant MERA to represent an approximation to the ground state and, in some models, also the first excited state. For Ising and Potts models the parameter λ is the strength of the transverse magnetic field applied along the z axis, with $\lambda_c=1$ corresponding to a quantum phase transition. Both the XX model and Heisenberg model are quantum critical as written.

Figure 24 shows the accuracy obtained for ground-state energies of the above models in the limit of an infinite chain, as a function of the refinement parameter χ . Simulations were performed with either the finite-correlation-range algorithm (for the noncritical Ising) or the scale-invariant algorithm (for critical systems). In all cases one observes roughly exponential convergence to the exact energy with increasing χ . For any fixed value of χ , the MERA consistently yields more accuracy for some models than for others. For the Ising

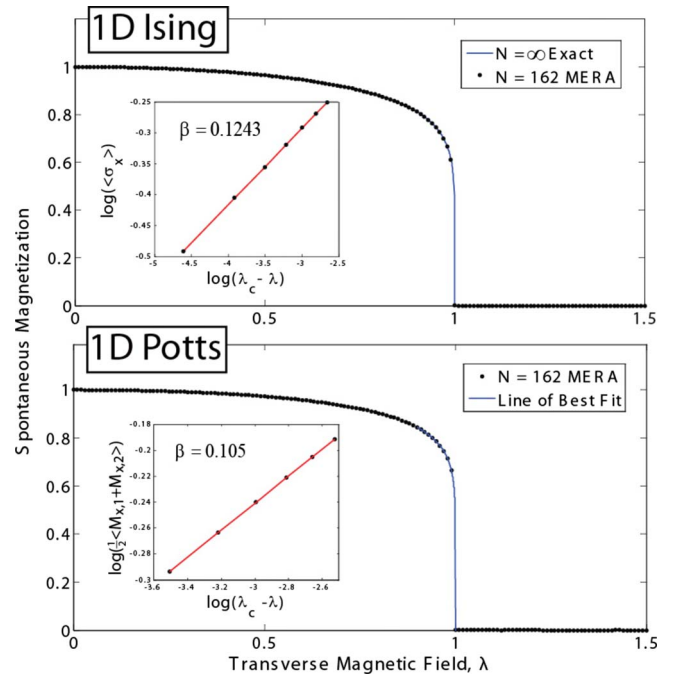


FIG. 26. (Color online) (Top) Spontaneous magnetization $\langle \sigma_x \rangle$ computed with a $\chi=8$ MERA for a periodic Ising system of $N=162$ sites. The results closely approximate the analytic values of magnetization known for the thermodynamic limit. A fit of the data near the critical point yields a critical exponent $\beta_{\text{MERA}}=0.1243$, with the exact exponent known as $\beta_{\text{ex}}=1/8$. (Bottom) An equivalent phase portrait of the Potts model, here with spontaneous magnetization $\frac{1}{2}\langle M_{x,1} + M_{x,2} \rangle$, is computed with a $\chi=14$ MERA and is plotted with a fit of the data near the critical point. The fit yields a critical exponent $\beta_{\text{MERA}}=0.105$ with the exact exponent known to be $\beta_{\text{ex}}=1/9$.

model, the cheapest simulation considered ($\chi=4$) produced five digits of accuracy, while the most computationally expensive simulation ($\chi=22$) produced ten digits of accuracy.²⁸ The time taken for the MERA to converge, running on a 3 GHz dual-core desktop PC with 8 Gbyte of random access memory (RAM), is approximately a few minutes, hours, days, or weeks for $\chi=4, 8, 16$, or 22 , respectively. We stress that these simulations were performed on single desktop computers. A parallel implementation of the code running on a computer cluster might bring significantly larger values of χ within computational reach.

Figure 25 demonstrates the ability of the MERA to reproduce finite-size effects. It shows the transverse magnetization as a function of the transverse magnetic field for several system sizes. The results smoothly interpolate between those for small system sizes and those for an infinite chain, and match the available exact solutions. On the other hand, the MERA can also be used to explore the phase diagram of a system. Figure 26 shows the spontaneous magnetization, which is the system's order parameter, for a 1D chain of $N=162$ sites for both Ising and Potts models, where N has been chosen large enough that the results under consideration do not change significantly with the system size (thermodynamic limit). A fit for the critical exponent of the Ising model gives $\beta_{\text{MERA}}=0.1243$, while the fit for the Potts model pro-

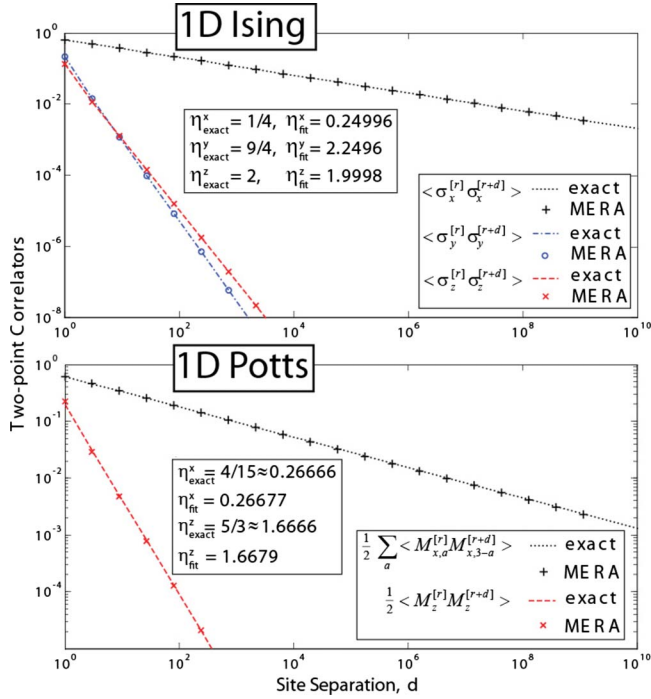


FIG. 27. (Color online) Two-point correlators for infinite 1D Ising and Potts chains at criticality ($\lambda=1$), as computed with $\chi=22$ scale-invariant MERA. Correlators for the Ising model are compared against analytic solutions (Ref. 24), while those for the Potts model are plotted against the polynomial decay predicted from CFT (Ref. 29). A scale-invariant MERA produces polynomial decay of correlators at all length scales. A fit of the form $\langle \sigma_x^{[r]} \sigma_x^{[r+d]} \rangle \propto d^{-\eta}$ for Ising correlators generated by the MERA gives the decay exponent $\eta^x=0.24996$, close to the known analytic value of $1/4$ and similarly for the fits on other correlators. Indeed the MERA here reproduces exact $\langle \sigma_x^{[r]} \sigma_x^{[r+d]} \rangle$ correlators for the Ising model at a distance up to $d=10^9$ sites within 0.6% accuracy. Critical exponents for the Potts are also reproduced very accurately.

duces $\beta_{\text{MERA}}=0.105$. These values are within less than 1% and 6% of the exact exponents $\beta=1/8$ and $\beta=1/9$ for the Ising and Potts models, respectively. Obtaining an accurate value for this critical exponent through a fit of the data near the critical point is difficult due to the steepness of the curve near the critical point. Through an alternative method involving the scaling superoperator \mathcal{S} (Sec. III), more accurate critical exponents were obtained in Ref. 11.

The previous results refer to local observables. Let us now consider correlators. A scale-invariant MERA, useful for the representation of critical systems, gives polynomial correlators at all length scales, as shown in Fig. 27 for the critical Ising and Potts models. Note that Fig. 27 displays the correlators that are most convenient to compute (as per Fig. 18). These occur at distances $d=3^q$ for $q=1,2,3,\dots$ and are evaluated with cost $O(\chi^8)$. Evaluation of arbitrary correlators is possible (see Fig. 16) but its cost is several orders of χ more expensive. The precision with which correlators are obtained is remarkable. A $\chi=22$ MERA for the Ising model gives $\langle \sigma_x^{[r]} \sigma_x^{[r+d]} \rangle$ correlators, at distances up to $d=10^9$ sites, accurate to within 0.6% of exact correlators. Critical exponents η are obtained through a fit of the form $C(r,r+d)$

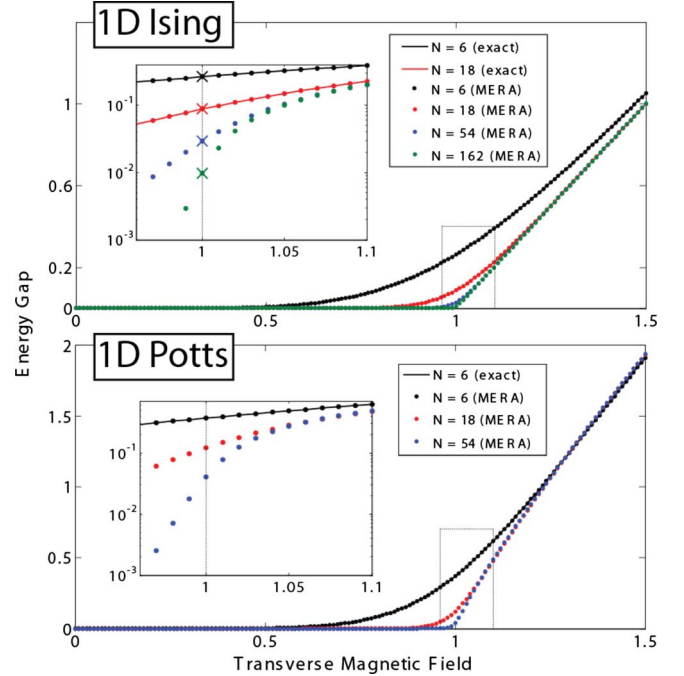


FIG. 28. (Color online) (Top) A $\chi=8$ MERA is used to compute the energy gap ΔE (the energy difference between the ground and first excited states) of the Ising chains as a function of the transverse magnetic field. The gap computed with MERA for $N=6, 18$ sites is in good agreement with that computed through exact diagonalization of the system. Inset: Crosses show analytic values of energy gaps at the critical point for $N=\{6, 18, 54, 162\}$. Even for the largest system considered, $N=162$, the gap computed with MERA, $\Delta E_{\text{MERA}}=9.67 \times 10^{-3}$, compares well with the exact value $\Delta E_{\text{ex}}=9.69 \times 10^{-3}$. (Bottom) Equivalent data for the Potts model where simulations have been performed with a $\chi=14$ MERA to account for the increased computational difficulty of this model.

$\propto d^{-\eta}$, with C as correlators of x, y , or z magnetization. For the Ising model, the exponents for x, y , and z magnetizations are obtained with less than 0.02% error each. For the Potts model exponents are obtained with less than 0.04% and 0.08% errors for x and z magnetizations, respectively.

Finally, we also demonstrate the ability of MERA to investigate low-energy excited states by computing the energy gap in the Ising and Potts models. Figure 28 shows that the gap grows linearly with the magnetic field λ and independent of N in the disordered phase $\lambda \gg \lambda_c$, while at criticality it closes as $1/N$. Even a relatively cheap $\chi=8$ MERA reproduces the known critical energy gaps to within 0.2% for systems as large as $N=162$ sites. The expected value of arbitrary local Hamiltonians (besides the energy) can also be easily evaluated for the excited state.

VII. CONCLUSIONS

After reviewing the conceptual foundations of the MERA (Refs. 1 and 5) (Secs. II and III), in this paper we have provided a rather self-contained description of an algorithm to explore low-energy properties of lattice models (Secs. IV and V), and benchmark calculations addressing 1D quantum spin chains (Sec. VI).

Many of the features of the MERA algorithm highlighted by the present results can also be observed by investigating systems of free fermions⁶ and free bosons⁷ in $D=1, 2$ dimensions. These include (i) the ability to consider arbitrarily large systems, (ii) the ability to compute the low-energy subspace of a Hamiltonian, (iii) the ability to disentangle non-critical systems completely, (iv) the ability to find a scale-invariant representation of critical systems, and finally (v) the reproduction of accurate polynomial correlators for critical systems. However, the algorithms of Refs. 6 and 7 exploit the formalism of Gaussian states that is characteristic of free fermions and bosons and cannot be easily generalized to interacting systems. Instead, the algorithms discussed in this paper can be used to address arbitrary lattice models with local Hamiltonians.

An alternative method to optimize the MERA is with a time-evolution algorithm as described in Ref. 17. The time-evolution algorithm has a clear advantage: it can be used both to compute the ground state of a local Hamiltonian (by simulating an evolution in imaginary time) and to study lattice dynamics (by simulating an evolution in real time). We find, however, that the algorithms described in the present paper are a better choice when it comes to computing ground states. On the one hand, the time-evolution algorithm has a

time step δt that needs to be sequentially reduced in order to diminish the error in the Suzuki-Trotter decomposition of the (Euclidean) time-evolution operator. In the present algorithm, convergence is faster and there is no need to fine tune a time step δt . In addition, the present algorithm allows one to compute not only the ground state but also low-energy excited states. It is unclear how to use the time-evolution algorithm to achieve the same.

The benchmark calculations presented in this paper refer to 1D systems. For such systems, however, DMRG (Ref. 3) already offers an extraordinarily successful approach. The strength of entanglement renormalization and the MERA relies on the fact that the present algorithms can also address large 2D lattices, as discussed in Refs. 8 and 9.

ACKNOWLEDGMENTS

The authors thank Frank Verstraete for key suggestions that lead to some of the optimization methods described in Sec. IV, and thank R. N. C. Pfeifer for useful discussions and comments. Support from the Australian Research Council (APA, Grants No. FF0668731 and No. DP0878830) is acknowledged.

¹G. Vidal, Phys. Rev. Lett. **99**, 220405 (2007).

²K. G. Wilson, Rev. Mod. Phys. **47**, 773 (1975).

³S. R. White, Phys. Rev. Lett. **69**, 2863 (1992); Phys. Rev. B **48**, 10345 (1993); U. Schollwöck, Rev. Mod. Phys. **77**, 259 (2005).

⁴C. J. Morningstar and M. Weinstein, Phys. Rev. Lett. **73**, 1873 (1994); Phys. Rev. D **54**, 4131 (1996).

⁵G. Vidal, Phys. Rev. Lett. **101**, 110501 (2008).

⁶G. Evenbly and G. Vidal, arXiv:0710.0692 (unpublished).

⁷G. Evenbly and G. Vidal, arXiv:0801.2449 (unpublished).

⁸L. Cincio, J. Dziarmaga, and M. M. Rams, Phys. Rev. Lett. **100**, 240603 (2008).

⁹G. Evenbly and G. Vidal, arXiv:0811.0879 (unpublished).

¹⁰V. Giovannetti, S. Montangero, and R. Fazio, Phys. Rev. Lett. **101**, 180503 (2008).

¹¹R. N. C. Pfeifer, G. Evenbly, and G. Vidal, Phys. Rev. A **79**, 040301 (2009).

¹²M. Aguado and G. Vidal, Phys. Rev. Lett. **100**, 070404 (2008).

¹³R. Koenig, B. W. Reichardt, and G. Vidal, arXiv:0806.4583 (unpublished).

¹⁴F. Verstraete, D. Porras, and J. I. Cirac, Phys. Rev. Lett. **93**, 227205 (2004); D. Porras, F. Verstraete, and J. I. Cirac, Phys. Rev. B **73**, 014410 (2006).

¹⁵G. Vidal, Phys. Rev. Lett. **91**, 147902 (2003); **93**, 040502 (2004).

¹⁶C. M. Dawson, J. Eisert, and T. J. Osborne, Phys. Rev. Lett. **100**, 130501 (2008).

¹⁷M. Rizzi, S. Montangero, and G. Vidal, Phys. Rev. A **77**, 052328

(2008).

¹⁸G. Evenbly and G. Vidal, arXiv:0707.1454 (unpublished).

¹⁹M. A. Levin and X.-G. Wen, Phys. Rev. B **71**, 045110 (2005).

²⁰A translation-invariant MERA, characterized by one disentangler and one isometry at each layer, need not represent a translation-invariant state $|\Psi\rangle$, hence the need to consider the average density matrix $\bar{\rho}$.

²¹S. Singh *et al.* (unpublished).

²²Each operator ρ_τ in Eq. (16) is both Hermitian ($\rho_\tau^\dagger = \rho_\tau$) and non-negative ($\langle\phi|\rho_\tau|\phi\rangle \geq 0, \forall \phi$) but its trace is $\text{tr}(\rho_\tau) = \chi_T$. For simplicity, we call ρ_τ a density matrix for any $\chi_T \geq 1$, even though it is only a proper density matrix for $\chi_T = 1$.

²³O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics I* (Springer, New York, 1979).

²⁴P. Pfeuty, Ann. Phys. (N.Y.) **57**, 79 (1970); T. W. Burkhardt and I. Guim, J. Phys. A **18**, L33 (1985).

²⁵J. Sólyom and P. Pfeuty, Phys. Rev. B **24**, 218 (1981).

²⁶E. Lieb, T. Schultz, and D. Mattis, Ann. Phys. (N.Y.) **16**, 407 (1961).

²⁷R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic, New York, 1982).

²⁸We note that the $\chi=22$ simulation for the Ising model gives less accurate results than the trend line observed for smaller χ in Fig. 24 would suggest. It is possible that numerical errors (such as errors in the sparse eigenvalue decomposition used) may have been significant at this level of accuracy.

²⁹P. Di Francesco, P. Mathieu, and D. Senechal, *Conformal Field Theory* (Springer, New York, 1997).