Variational Monte Carlo with the Multi-Scale Entanglement Renormalization Ansatz

Andrew J. Ferris[1,2] and Guifre Vidal[1,3]

[1]*The University of Queensland, School of Mathematics and Physics, Queensland 4072, Australia*
[2]*Département de Physique, Université de Sherbrooke, Québec, J1K 2R1, Canada*
[3]*Perimeter Institute for Theoretical Physics, Waterloo, Ontario, N2L2Y5, Canada*
(Dated: January 20, 2012)

Tensor network states are powerful variational ansätze for ground states of quantum many-body systems on a lattice. Monte Carlo sampling techniques have been proposed as a strategy to reduce the computational cost of contractions in tensor network approaches. Here we put forward a variational Monte Carlo approach for the multi-scale entanglement renormalization ansatz (MERA), which is a unitary tensor network. Two major adjustments are required compared to previous proposals with non-unitary tensor networks. First, instead of sampling over configurations of the original lattice, made of L sites, we sample over configurations of an effective lattice, which is made of just $\log(L)$ sites. Second, the optimization of unitary tensors must account for their unitary character while being robust to statistical noise, which we accomplish with a modified steepest descent method within the set of unitary tensors. We demonstrate the performance of the variational Monte Carlo MERA approach in the relatively simple context of a finite quantum spin chain at criticality, and discuss future, more challenging applications, including two dimensional systems.

I. INTRODUCTION

Understanding the collective behaviour of quantum many-body systems remains a central topic in modern physics, as well as one of the greatest computational challenges in science. Quantum Monte Carlo sampling techniques are capable of addressing a large class of (unfrustrated) bosonic and spin lattice models, but fail when applied to other models such as frustrated antiferromagnets and interacting fermions due to the so-called sign problem. Variational approaches, on the other hand, are sign-problem free but are typically strongly biased towards specific many-body wavefunctions. An important exception is given by the density matrix renormalization group,[1] a variational approach based on the matrix product state (MPS),[2] which is capable of providing an extremely accurate approximation to the ground state of most one-dimensional lattice models. The success of DMRG is based on the fact that an MPS can reproduce the structure of entanglement common to most ground states of one-dimensional lattice models.

In order to extend the success of DMRG to other contexts, new tensor networks generalizing the MPS have been proposed. For instance, the multi-scale entanglement renormalization ansatz (MERA),[3] with a network of tensors that extends in an additional direction corresponding to length scales, is particularly suited to address quantum critical systems. Most significant has also been the proposal of tensor networks for systems in two and higher dimensions, where the MPS becomes inefficient. Scalable tensor networks include the projected entangled-pair states PEPS[4] (a direct generalization of the MPS to larger dimensions) and higher dimensional versions of the MERA.[5,6] They can be used to address frustrated antiferromagnets and interacting fermions, since they are free of the sign problem experienced by quantum Monte Carlo approaches.

In a tensor network state, the size of the tensors is measured by the bond dimension χ. This bond dimension χ indicates how many variational coefficients are used. Crucially, it also regulates both the cost of the simulation, which scales as $\mathcal{O}(\chi^p)$ for some large power p, and how much ground state entanglement the many-body ansatz can reproduce. In the large χ regime, PEPS and MERA are essentially unbiased methods, but with a huge computational cost that is often unaffordable. More affordable simulations are obtained in the small χ regime, but there these methods are biased in favour of weakly entangled phases (e.g. symmetry-breaking phases) and against strongly entangled phases (e.g. spin liquids and systems with a Fermi surface). Identifying more efficient strategies for tensor network contraction, so that larger values of the bond dimension χ can be used and the bias towards weakly entangled states is suppressed, is therefore a priority in this research area.

Refs. 7, 8 proposed the use of Monte Carlo sampling as a means to decrease computational costs in tensor network algorithms. [We note that there are other variational ansätze, such as so-called correlated product states, entangled plaquette states, and string-bond states, whose contractibility relies on sampling; see the introduction of Ref. 9 for a review]. In a tensor network approach such as MPS, MERA or PEPS, sampling over specific configurations of the lattice allows to reduce the cost of contractions (for single samples) from $\mathcal{O}(\chi^p)$ to $\mathcal{O}(\chi^q)$, where q is significantly smaller than p, typically of the order of $p/2$. Needless to say, sampling introduces statistical errors. However, if less than $O(\chi^{q-p})$ samples are required in order to achieved some pre-stablished accuracy, then overall sampling results in a reduction of computational costs.

The proposal of Refs. 7, 8 is based on computing the overlap of the tensor network state with a product state (representing the sampled configuration). As such, it cannot be directly applied to the MERA, because the overlap of a MERA with a product state cannot be computed efficiently. Luckily, as discussed in Ref. 10, a sampling strategy specific to unitary tensor networks (such as MERA and unitary versions of MPS and tree tensor networks) is not only possible, but it actually
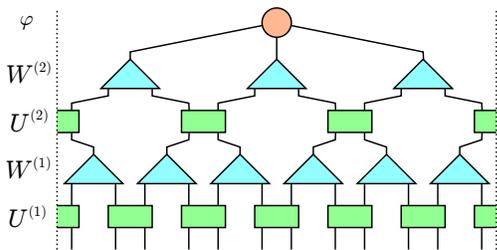
FIG. 1: (Color online) Tensor network diagram for a binary MERA presenting the state $|\Psi\rangle$ of a translation invariant lattice $\mathcal{L}$ made of $L = 12$ sites and with periodic boundary conditions. The tensors on each layer are identical and their labels are displayed to the left. The $U^{(n)}$ tensors (green rectangles) are unitary operators (acting top-to-bottom), $W^{(n)}$ (cyan triangles) are isometric, and φ (red circle) is a normalized 'wavefunction'.

has several advantages. Most notably, sampling takes place over configurations of a reduced, effective lattice; and it is possible to perform perfect sampling, by means of which uncorrelated configurations are drawn directly according to the correct probability.

The main goals of this paper are to propose a variational Monte Carlo scheme for the MERA and to demonstrate its feasibility. We also discuss possible future applications. Let us briefly list some of the highlights of the approach. (i) In a lattice of size L, the sampled configurations correspond to an effective lattice of size $\mathcal{O}(\log(L))$; in this way, the cost of evaluating the expectation value of a local observable scales just as $\mathcal{O}(\log(L))$ and not as $\mathcal{O}(L)$ as in Refs. 7, 8. (ii) We employ the perfect sampling strategy of Ref. 10, thus avoiding the loses of efficiency in the Markov chain Monte Carlo of Refs. 7, 8 due to equilibration and autocorrelation times. (iii) Variational parameters are optimized while explicitly preserving the unitary constraints that the tensors in the MERA are subjected to. This is accomplished by a steepest descent method within the set of unitary tensors, which is much more robust to statistical noise than the singular value decomposition methods employed in MERA algorithms without sampling.[11]

We demonstrate the performance of our approach by computing an approximation to the ground state of a finite Ising chain with transverse magnetic field. For the binary MERA under consideration, sampling lowers the costs of elementary contractions from $\mathcal{O}(\chi^9)$ to $\mathcal{O}(\chi^5)$. We find that the resulting (approximate) ground state energy decreases as the number of samples is increased, thus obtaining a demonstration of principle of the approach. We also notice that the number of samples required to achieve a given accuracy increases as the transverse magnetic field approaches its critical value.

To our knowledge, this is the first instance of sampling-based optimization of a relatively complex tensor network. Previous similar optimizations included that of an MPS,[8] which is a considerably simpler tensor network (with only three-legged tensors), and of tensor networks that under sampling break into smaller, simpler tensor networks (e.g. into MPS, single plaquette states, etc).[7,9,12,13] In more complex tensor networks, such as MERA and PEPS, the optimiza-

tion becomes much harder due to high sensitivity to statistical noise. Thus, for instance, Ref. 14 spells out a full variational Monte Carlo approach for PEPS but uses an alternative method, not based on sampling, in order to optimize the tensors. Indeed, in Ref. 14 sampling is only used to aid in the computation of expectation values. Here, instead, we use sampling both to optimize the MERA and to compute expectation values.

We emphasize, however, that our results only demonstrate a gain over optimization schemes based on exact contractions (i.e. without sampling) in the low accuracy regime, where only a relatively small number of samples are required. The specific MERA (namely binary MERA for a one-dimensional lattice) and low value of χ ($\chi = 4$) considered here for illustrative purposes implies that the cost per sample is $\chi^9/\chi^5 = \chi^4 = 256$ times smaller than an exact contraction. Recall that the statistical error decreases only as $\sqrt{N}$ with the number N of samples. If more than 256 samples are required in order to obtain a sufficiently accurate approximation of the exact contraction, then the sampling scheme may be overall less efficient than the exact contraction scheme.

The advantage of sampling over exact contraction schemes is expected to be more evident in MERA settings where the cost $\mathcal{O}(\chi^p)$ scales with a larger exponent p, and for larger values of χ. In particular, we envisage that the method described in this paper, possibly with further improvements, will improve the range of applicability of MERA in two and higher dimensions.

The content of this paper is distributed in five more sections. In Sec. II, we discuss methods for sampling with the MERA. In Sec. III we propose an optimization scheme using sampling techniques. In Sec. IV we benchmark the approach with the quantum Ising model. In Sec. V we discuss future applications including extensions to higher dimensions and extracting long-range correlations, before concluding in Sec. VI.

II. LOCAL EXPECTATION VALUES WITH SAMPLING

In this section we explain how to use sampling in order to speed-up the computation of expectation values with the MERA. We present both perfect sampling and partial sampling strategies, building on the proposals of Ref. 10 for generic unitary tensor networks. We also discuss the importance of the choice of local basis in sampling. We start by reviewing some necessary background material on the MERA.

A. MERA, expectation values and causal cones

The MERA[3] is a variational wavefunction for ground states of quantum many-body systems on a lattice. The state $|\Psi\rangle$ of a lattice $\mathcal{L}$ made of L sites is represented by means of a tensor network made of two types of tensors, called *disentanglers* and *isometries*. The tensor network is based on a real-space renormalization group transformation, known as *entanglement renormalization*: disentanglers are used to remove
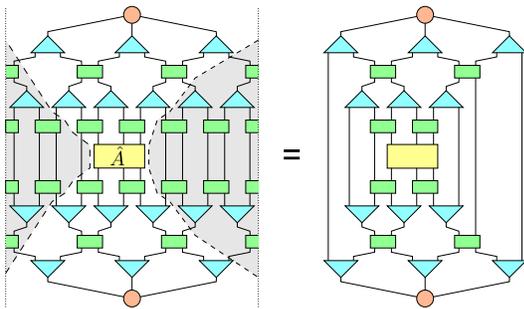
FIG. 2: (Color online) Expectation value $\langle\Psi|\hat{A}|\Psi\rangle = \langle\Psi^{\mathcal{C}}|\hat{A}|\Psi^{\mathcal{C}}\rangle$ of the local operator $\hat{A}$ (yellow rectangle), which acts on (at most) three neighboring sites. The flipped tensors, on the bottom half of the diagram, are the Hermitian conjugates of the respective tensors above. The 'causal cone' is delimited by the dotted line in the left diagram, corresponding to $\langle\Psi|\hat{A}|\Psi\rangle$. The tensors outside the causal cone cancel, significantly simplifying the diagram on the right, corresponding to $\langle\Psi^{\mathcal{C}}|\hat{A}|\Psi^{\mathcal{C}}\rangle$.

FIG. 3: (Color online) States $|\Psi\rangle$ and $|\Psi^{\mathcal{C}}\rangle$ represented by the MERA. (a) Tensor network for the state $|\Psi\rangle$ of the original lattice $\mathcal{L}$. The causal cone is delimited by a discontinuous line. (b) Tensor network for the state $|\Psi^{\mathcal{C}}\rangle$ of the effective lattice $\mathcal{L}^{\mathcal{C}}$. Sites further from the center effectively represent increasingly large length scales in the original lattice.

short-range entanglement from the system, whereas isometries are used to coarse-grain blocks of site into single, effective sites. An example of a MERA on a periodic 1D lattice with $L = 12$ sites is depicted in Fig. 1. This structure is called 'binary' MERA because of the 2-to-1 course-graining transformation in each repeating layer. Ascending upwards in the figure, the disentanglers $U^{(n)}$ remove short-range entanglement in between each course-graining transformation, implemented by isometries $W^{(n)}$ until the remaining Hilbert space is small enough to deal with directly with some wavefunction φ.

The MERA can also be viewed in the reverse — starting from the top of Fig. 1, we descend downwards in a unitary quantum circuit, adding (initially unentangled) sites in each layer. For instance, let us flow downwards in Fig. 1. To a three-site system in state φ, we first add three additional unentangled sites, turning it into a six-site system; and later we add another six unentangled sites, producing the final twelve-site system. This unitary structure can be exploited when calculating the expectation value $\langle\hat{A}\rangle \equiv \langle\Psi|\hat{A}|\Psi\rangle$ of a local operator $\hat{A}$ acting on a few neighboring sites. Specifically, all tensors not 'causally' connected to the few sites supporting $\hat{A}$ cancel, as depicted in Fig. 2. The resulting diagram is significantly simpler and can be interpreted as the expectation value $\langle\Psi^{\mathcal{C}}|\hat{A}|\Psi^{\mathcal{C}}\rangle$ of $\hat{A}$ for a state $|\Psi^{\mathcal{C}}\rangle$ of an effective lattice $\mathcal{L}^{\mathcal{C}}$ made of $O(\log(L))$ sites. We emphasize that, by construction,

$$\langle\Psi|\hat{A}|\Psi\rangle = \langle\Psi^{\mathcal{C}}|\hat{A}|\Psi^{\mathcal{C}}\rangle. \tag{1}$$

Therefore, we can evaluate the expectation value $\langle\Psi|\hat{A}|\Psi\rangle$ by contracting the tensor network corresponding to $\langle\Psi^{\mathcal{C}}|\hat{A}|\Psi^{\mathcal{C}}\rangle$. The numerical cost of performing this contraction grows linearly with the number of sites in $\mathcal{L}^{\mathcal{C}}$, and thus only logarithmically with the number of site L in the original lattice $\mathcal{L}$.

The dimension of the Hilbert space after each coarse-graining transformation is an adjustable parameter, the *bond dimension* χ, which plays a central role in the present dis-
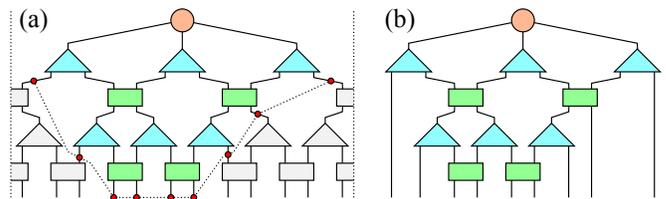
cussion. Increasing the bond dimension χ implies including a larger fraction of the original Hilbert space and leads to greater accuracy, but also requires greater computational resources. Optimization algorithms to approximate ground states, and to evaluate local expectation values and correlators, are present in the literature.[11] The numerical cost of finding an expectation value or performing a single optimization iteration using the binary MERA scales as $\mathcal{O}(\chi^9 \log L)$ for a translation invariant system. For more complex MERA structures, such as those representing 2D lattices, the power of χ for the cost increases dramatically. For instance, the 2D MERA presented in Ref. 5 has a numerical cost of $\mathcal{O}(\chi^{16} \log L)$, which on current computers restricts $\chi < 8$. For many systems, this does not allow for enough entanglement to accurately describe the ground state, limiting the accuracy of the approach.

Here we hope to alleviate this problem by reducing the numerical cost as a function of χ using Monte Carlo techniques. We will find that the cost of a single sample scales as $\mathcal{O}(\chi^5)$ for binary 1D MERA, compared to the $\mathcal{O}(\chi^9)$ cost of the 'exact' contraction.

B. Monte Carlo sampling with the MERA

Our goal is to compute the expectation value $\langle\hat{A}\rangle \equiv \langle\Psi|\hat{A}|\Psi\rangle$ by contracting the tensor network corresponding to $\langle\Psi^{\mathcal{C}}|\hat{A}|\Psi^{\mathcal{C}}\rangle$, see Fig. 2. The first step is to re-express the tensor network contraction as a summation over indices corresponding to the sites of the effective lattice, as shown in Fig. 4. We then get

$$\langle\hat{A}\rangle = \sum_{\mathbf{r}\in\mathcal{R}} \langle\Psi^{\mathcal{C}}|\mathbf{r}\rangle\langle\mathbf{r}|\hat{A}|\Psi^{\mathcal{C}}\rangle, \tag{2}$$

where $\mathcal{R}$ is an orthonormal basis of product states $|\mathbf{r}\rangle = |r_1\rangle \otimes |r_2\rangle \otimes \ldots$ on the effective lattice $\mathcal{L}^{\mathcal{C}}$.

We will approximately evaluate the sum in Eq. (2) by using Monte Carlo sampling over the states $|\mathbf{r}\rangle \in \mathcal{R}$. Notice that in the effective lattice, the sites away from the support of $\hat{A}$ have undergone one or more course-graining transformations. In other words, sites further from the center represent increasingly larger length-scales. Thus, sampling over sites of the effective lattice corresponds to sampling the system *at different length scales*. This property is reminiscent of global or
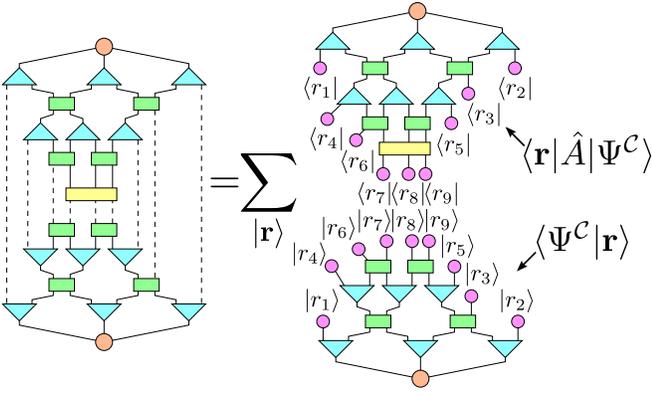
FIG. 4: (Color online) Tensor network for $\langle \Psi^{\mathcal{C}} | \hat{A} | \Psi^{\mathcal{C}} \rangle$. The dashed lines indicate the indices that will be sampled. On the right-hand-side we explicitly write the tensor contractions outside the causal cone as a sum over a complete, orthonormal set of 'wavefunctions' $|\mathbf{r}\rangle = |r_1\rangle \otimes |r_2\rangle \otimes \ldots$ (pink circles). Monte Carlo sampling will be performed over this set. Each term of the sum can be expressed as the product $\langle \Psi^{\mathcal{C}} | \mathbf{r} \rangle \langle \mathbf{r} | \hat{A} | \Psi^{\mathcal{C}} \rangle$.

cluster updates used in existing Monte Carlo methods to solve critical systems.

A naïve scheme for approximating the sum in Eq. (2) would be to choose $|\mathbf{r}\rangle$ at random from $\mathcal{R}$, and evaluate $\langle \Psi^{\mathcal{C}} | \mathbf{r} \rangle \langle \mathbf{r} | \hat{A} | \Psi^{\mathcal{C}} \rangle$ according to the tensor networks in Fig. 4. The cost of obtaining a single sample scales as $\mathcal{O}(\chi^5 \log L)$. However, the statistical variance of a sampling scheme can be substantially reduced by implementing importance sampling — in this case choosing configurations $|\mathbf{r}\rangle$ that are more likely. More precisely, sampling is implemented according to the wavefunction *weight*, $P(\mathbf{r}) \equiv |\langle \Psi^{\mathcal{C}} | \mathbf{r} \rangle|^2$, which can be calculated efficiently as indicated in Fig. 5. We can express Eq. (2) in a form more convenient for importance sampling,

$$\langle \hat{A} \rangle = \sum_{\mathbf{r} \in \mathcal{R}} P(\mathbf{r}) A^{\mathcal{C}}(\mathbf{r}), \qquad (3)$$

where

$$A^{\mathcal{C}}(\mathbf{r}) \equiv \frac{\langle \Psi^{\mathcal{C}} | \mathbf{r} \rangle \langle \mathbf{r} | \hat{A} | \Psi^{\mathcal{C}} \rangle}{\langle \Psi^{\mathcal{C}} | \mathbf{r} \rangle \langle \mathbf{r} | \Psi^{\mathcal{C}} \rangle} = \frac{\langle \mathbf{r} | \hat{A} | \Psi^{\mathcal{C}} \rangle}{\langle \mathbf{r} | \Psi^{\mathcal{C}} \rangle}. \qquad (4)$$

Note that because the MERA is normalized by construction, and therefore the weights sum to one, $\sum_{\mathbf{r} \in \mathcal{R}} P(\mathbf{r}) = 1$. However, during sampling only some subset $\tilde{\mathcal{R}}$ of the configurations are considered. One needs to renormalize the weights accordingly, so that the expectation value $\langle \hat{A} \rangle$ is approximated as

$$\langle \hat{A} \rangle \approx \sum_{\mathbf{r} \in \tilde{\mathcal{R}}} P(\mathbf{r}) A^{\mathcal{C}}(\mathbf{r}) / \sum_{\mathbf{r} \in \tilde{\mathcal{R}}} P(\mathbf{r}). \qquad (5)$$

1. Perfect sampling

In the case of MERA, and indeed any state that can be written as a unitary quantum circuit, a 'perfect' sample can be generated according to the probability distribution $P(\mathbf{r})$ in a single sweep.[10] This makes Markov Chain Monte Carlo unnecessary, simplifying the algorithm and eliminating a source of statistical error (i.e. autocorrelation effects). This is one advantage of this technique over other tensor network sampling methods in the literature.[7–9,12–14]

The sample can be constructed by sampling just one index at a time. Beginning at the top layer of the MERA, and aiming to sample just the first index (left-most in Fig. 4 (b)), we can construct the one-site reduced density matrix ρ_1 by the tensor contraction in Fig. 5 (b). The probability $P(r_1) = \langle r_1 | \rho_1 | r_1 \rangle$ can then be found *for all possible* $|r_1\rangle$. A value of $|r_1\rangle$ is then randomly selected according to any complete basis of our choosing.

After this selection is made, we can then sample the next (top-most) index, according to the conditional weights $P(r_2 \,|\, r_1)$ as calculated by the diagram in Fig. 5 (b) (we refer to Ref. 10 for further details). We continue to sample the state of each site, until we have sampled every site.

Each of the diagrams in Fig. 5 can be calculated with cost $\mathcal{O}(\chi^5)$, while there are $\mathcal{O}(\log L)$ layers to the MERA. A single sample can therefore be generated with cost $\mathcal{O}(\chi^5 \log L)$, compared to $\mathcal{O}(\chi^9 \log L)$ for the exact contraction. So long as the number of samples N is significantly less than χ^4, Monte Carlo sampling will be faster than exact contraction.

In practice, we will perform $N \gg 1$ samples in order to get a good estimate of $\langle \hat{A} \rangle$. As the samples are completely uncorrelated, the variance of $A^{\mathcal{C}}(\mathbf{r})$

$$\mathrm{Var}\left[A^{\mathcal{C}}\right] \equiv \sum_{\mathbf{r}} P(\mathbf{r}) \left(A^{\mathcal{C}}(\mathbf{r}) - \bar{A}^{\mathcal{C}}\right)^2 \qquad (6)$$

can be used to estimate statistical error ΔA,

$$\Delta A \equiv \sqrt{\mathrm{Var}\left[A^{\mathcal{C}}\right]/N}, \qquad (7)$$

while $\mathrm{Var}\left[A^{\mathcal{C}}\right]$ is itself upper bounded by the variance of the operator to be measured,

$$\mathrm{Var}\left[A^{\mathcal{C}}\right] \leq \langle \hat{A}^2 \rangle - \langle \hat{A} \rangle^2. \qquad (8)$$

It is easy to show that the upper bound is saturated when sampling in the diagonal basis of $\hat{A}$. However, the actual value of this variance is dependent on the choice of basis $\{|\mathbf{r}\rangle\}$, as well as the state being sampled. Therefore, it is worth spending some effort in order to minimize this quantity.

2. Partial sampling

The statistical noise is caused by Monte Carlo sampling of the tensor contraction, and it stands to reason that sampling *less* indices in the tensor network diagram (see Fig. 4) would reduce the statistical error. Indeed, it is possible to exactly contract the three physical indices at the lowest level (those connected to operator $\hat{A}$) while keeping the computational cost at $\mathcal{O}(\chi^5 \log L)$, as depicted in Fig. 6. That is, we sample over configurations $\mathbf{r}^{\diamond} \in \mathcal{R}^{\diamond}$ of the effective lattice $\mathcal{L}^{\mathcal{C}}$ *minus* the three central sites on which $\hat{A}$ is supported.
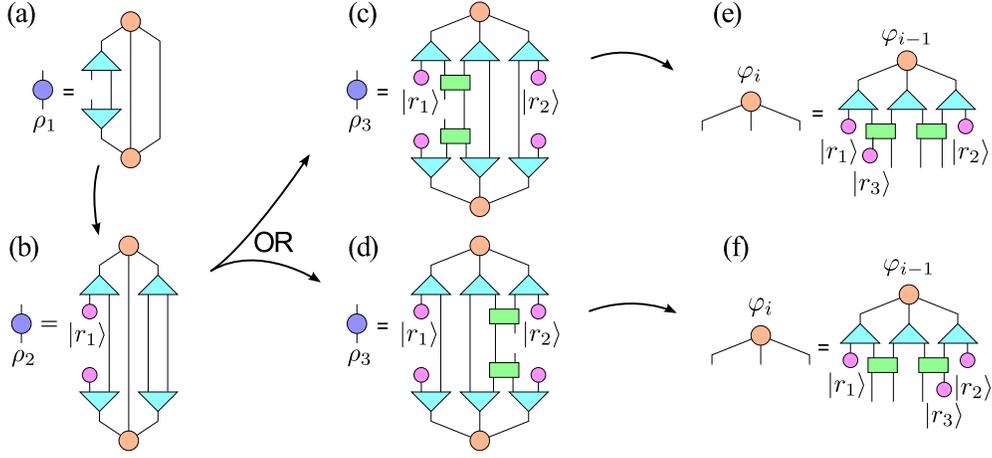
FIG. 5: (Color online) Diagrams required to create a sample, sequentially sampling indices from the top to the bottom of the MERA. First the reduced density matrix, ρ_i (blue circles), of an index is computed. We then randomly generate a state $|r_i\rangle$ according to the diagonal basis of ρ_i, and proceed to calculate ρ_{i+1}. In each 'layer' of the MERA, we sample the outermost indices in (a,b) and *one* of the inner indices in (c,d), depending on the physical location we are sampling. A projected, three-site wavefunction for the following layer is calculated in the appropriate diagram (e) or (f), from which the algorithm begins again with (a), until we have sampled all the indices indicated in Fig. 4 or 6.

This method effectively generates (unnormalized) three-site wave functions $|\varphi_0(\mathbf{r}^\diamond)\rangle \equiv \langle \mathbf{r}^\diamond|\Psi^\mathcal{C}\rangle$ from the reduced density operator (as seen in Fig. 5 (e,f)). Conversely, the reduced density matrix for the three central sites is $\sum_{\mathbf{r}^\diamond} |\varphi_0(\mathbf{r}^\diamond)\rangle\langle\varphi_0(\mathbf{r}^\diamond)|$. Importance sampling is now achieved by selecting $\mathbf{r}^\diamond$ according to the weight

$$P(\mathbf{r}^\diamond) \equiv \langle \Psi^\mathcal{C}|\mathbf{r}^\diamond\rangle\langle \mathbf{r}^\diamond|\Psi^\mathcal{C}\rangle = \langle\varphi_0(\mathbf{r}^\diamond)|\varphi_0(\mathbf{r}^\diamond)\rangle. \quad (9)$$

In turn, perfect sampling of the sites proceeds exactly as before, but it stops before the three central sites, which are not sampled. We define the estimator

$$A^\diamond(\mathbf{r}^\diamond) \equiv \frac{\langle \Psi^\mathcal{C}|\mathbf{r}^\diamond\rangle\hat{A}\langle \mathbf{r}^\diamond|\Psi^\mathcal{C}\rangle}{\langle \Psi^\mathcal{C}|\mathbf{r}^\diamond\rangle\langle \mathbf{r}^\diamond|\Psi^\mathcal{C}\rangle} = \frac{\langle\varphi_0(\mathbf{r}^\diamond)|\hat{A}|\varphi_0(\mathbf{r}^\diamond)\rangle}{\langle\varphi_0(\mathbf{r}^\diamond)|\varphi_0(\mathbf{r}^\diamond)\rangle}, \quad (10)$$

whose expectation value obeys,

$$\langle \hat{A} \rangle = \sum_{\mathbf{r}^\diamond \in \mathcal{R}^\diamond} P(\mathbf{r}^\diamond)A^\diamond(\mathbf{r}^\diamond), \quad (11)$$

in analogy with Eq. (3). Notice that, by construction, the statistical variance $\mathrm{Var}\big[A^\diamond\big]$ is smaller or equal than $\mathrm{Var}\big[A^\mathcal{C}\big]$ in Eq. (8), and therefore the numerical accuracy is increased without affecting the computational cost.

3. Diagonal basis selection

In general, we are free to choose any complete basis $\mathcal{R}$ (or $\mathcal{R}^\diamond$) from which to draw individual samples. A good choice is one that produces a small statistical error, Eqs. (7,8). Intuitively, the goal of importance sampling is to decrease the statistical variance by choosing configurations $\mathbf{r}$ (or $\mathbf{r}^\diamond$) with a large overlap with the state $|\Psi^\mathcal{C}\rangle$. With this in mind, one could aim to maximize the 'average' weight,

$$\overline{P} \equiv \sum_{\mathbf{r}} P(\mathbf{r})^2 \Big/ \sum_{\mathbf{r}} P(\mathbf{r}). \quad (12)$$
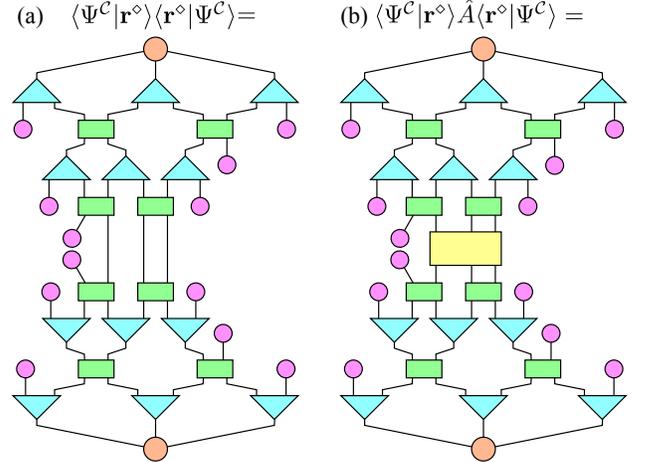

FIG. 6: (Color online) Tensor network diagrams for the quantities (a) $\langle\Psi^\mathcal{C}|\mathbf{r}^\diamond\rangle\langle \mathbf{r}^\diamond|\Psi^\mathcal{C}\rangle = P(\mathbf{r})$ and (b) $\langle\Psi^\mathcal{C}|\mathbf{r}^\diamond\rangle\hat{A}\langle \mathbf{r}^\diamond|\Psi^\mathcal{C}\rangle$ with partial sampling.

It is easy to show that for a given quantum probability distribution, specified by a density matrix, the above quantity is maximized in the diagonal basis of the density matrix. Inspired by this fact, here we choose to sample site i in the basis in which the reduced density matrix ρ_i is diagonal. The $\chi \times \chi$ density matrices ρ_i, calculated in Fig. 5, can be diagonalized with cost $\mathcal{O}(\chi^3)$. Note that the chosen basis will depend on previously sampled sites, and that the resulting sampling basis is still a complete, orthonormal basis of product states.

We have found that this approach can radically increase the average value of the weight, Eq. (12) with the effect becoming stronger for larger systems and values of χ. More importantly, we find that the statistical variance in the observables is decreased (see Sec. IV A and Fig. 8).

This technique to select the sampling basis is not specific

to unitary tensor networks nor perfect sampling methods, and could thus be of benefit to other variational quantum Monte Carlo algorithms.

4. Operators that decompose as sum of local terms

Finally, we may wish to compute the expectation value of an operator that is the sum of local terms, such as a Hamiltonian made of nearest-neighbor interactions:

$$\hat{H} = \sum_i \hat{H}_i. \tag{13}$$

In this case we sample each local term $\hat{H}_i$ as indicated previously, noticing that the causal cone of each $\hat{H}_i$ depends on the location of the sites of lattice $\mathcal{L}$ where the local operator is supported. One can either choose to (uniformly) sample the position i in the lattice, or systematically sweep through all the positions i. A complete sweep, where each site i is visited once, costs $\mathcal{O}(\chi^5 L \log L)$.

III. ENERGY MINIMIZATION WITH SAMPLING

In order to find an approximation to the ground state, we need to minimize the energy of the MERA. The direction of steepest *ascent* is given by the complex derivative with respect to the conjugate[16] of each element of each tensor. Inserting Eq. (4) (or Eq. (10)) into Eq. (5) and differentiating gives

$$\frac{\partial \langle \hat{H} \rangle}{\partial U^{(n)*}} = \sum_i \frac{\partial \langle \hat{H}_i \rangle}{\partial U^{(n)*}}, \tag{14}$$

$$\frac{\partial \langle \hat{H}_i \rangle}{\partial U^{(n)*}} = \sum_{\mathbf{r}} P(\mathbf{r}) \left[\frac{1}{P(\mathbf{r})} \frac{\partial \langle \Psi^{\mathcal{C}} | \mathbf{r} \rangle \langle \mathbf{r} | \hat{H}_i | \Psi^{\mathcal{C}} \rangle}{\partial U^{(n)*}} \right.$$
$$\left. - \frac{\langle \hat{H}_i \rangle}{P(\mathbf{r})} \frac{\partial \langle \Psi^{\mathcal{C}} | \mathbf{r} \rangle \langle \mathbf{r} | \Psi^{\mathcal{C}} \rangle}{\partial U^{(n)*}} \right], \tag{15}$$

where the derivative with respect to a tensor is element-wise, and similar expressions hold for $W^{(n)}$ and φ. The derivatives on the right-hand-side of Eq. (15) can be found by using the usual rules for calculating derivatives in the diagrams, see Figs. 6 and 7.[19] In practice $\langle \hat{H}_i \rangle$ and $\frac{\partial \langle \hat{H}_i \rangle}{\partial U^{(n)*}}$ will be estimated simultaneously by sampling.

Multiple approaches are possible for updating the MERA to minimize the energy and find a good approximation to the ground state. One approach often used iteratively optimizes each tensor in the MERA according to the following algorithm[11].

$$usv = -\frac{\partial \langle \hat{H} \rangle}{\partial U^{(n)*}}$$
$$U^{(n)} \rightarrow uv \tag{16}$$

In the above u and v are unitary, while s is diagonal and positive, thus representing the singular value decomposition

(SVD) of the derivative. This algorithm finds the unitary tensor $U^{(n)}$ that minimizes the trace value of its product with the above, called environment— with the requirement that $\hat{H}$ is negative-semidefinite. Similar steps apply to $W^{(n)}$ and φ, where the SVD ensures that they remain isometric or normalized, respectively.

Unfortunately, the above scheme is extremely sensitive to the statistical noise inherent to Monte Carlo sampling, and results in very poor optimization. Ideally, we would prefer a method in which the statistical noise is able to average out over many iterations.

The most obvious scheme satisfying this requirement is straightforward steepest descent. Again, one must ensure that the tensors obey the unitary/isometric constraints characteristic of the MERA, so one can utilize the SVD to find the unitary tensor closest (with respect to the L_2–norm) to the usual downhill update. With this method, the ith step is given by

$$usv = U^{(n)} - \mu_t \frac{\partial \langle \hat{H} \rangle}{\partial U^{(n)*}},$$
$$U^{(n)} \rightarrow uv, \tag{17}$$

where μ_t is a number modulating the size of change at the step t.

In this paper we avoid using the SVD entirely by explicitly remaining in the unitary subspace, along the lines of Ref. 17. We define the tangent vector $G_U^{(n)}$ as the derivative projected onto the tangent space of all unitaries located about $U^{(n)}$,

$$G_U^{(n)} = \frac{\partial \langle \hat{H} \rangle}{\partial U^{(n)*}} - U^{(n)} \frac{\partial \langle \hat{H} \rangle}{\partial U^{(n)*}}^{\dagger} U^{(n)}. \tag{18}$$

The matrix $U^{(n)} - \mu_t G_U^{(n)}$ is within $\mathcal{O}(\mu_t^2)$ of a unitary matrix. Noting that $U^{(n)\dagger} G_U^{(n)}$ is anti-Hermitian, then the update

$$U^{(n)} \rightarrow U^{(n)} \exp\left[-\mu_t \left(U^{(n)\dagger} \frac{\partial \langle \hat{H} \rangle}{\partial U^{(n)*}} - \frac{\partial \langle \hat{H} \rangle}{\partial U^{(n)*}}^{\dagger} U^{(n)} \right) \right] \tag{19}$$

both travels in the direction of the tangent vector while $U^{(n)}$ remains precisely unitary. The same approach works for $n \times m$ isometric matrices, taking care that $n \leq m$, with computational cost scaling similarly to the SVD approach as $\mathcal{O}(n^2 m)$ (see Appendix).

The performance of the algorithm is highly dependent on the behaviour of μ_t, as well as the number of Monte Carlo samples, N_t, taken in each step. Simple schemes will keep μ_t and N_t constant, which is the approach we take here. On the other hand, one may choose to increase N_t with t so that harmful noise is reduced when approaching the optimal solution; or to decrease μ_t with t for much the same reason; or a combination of both.
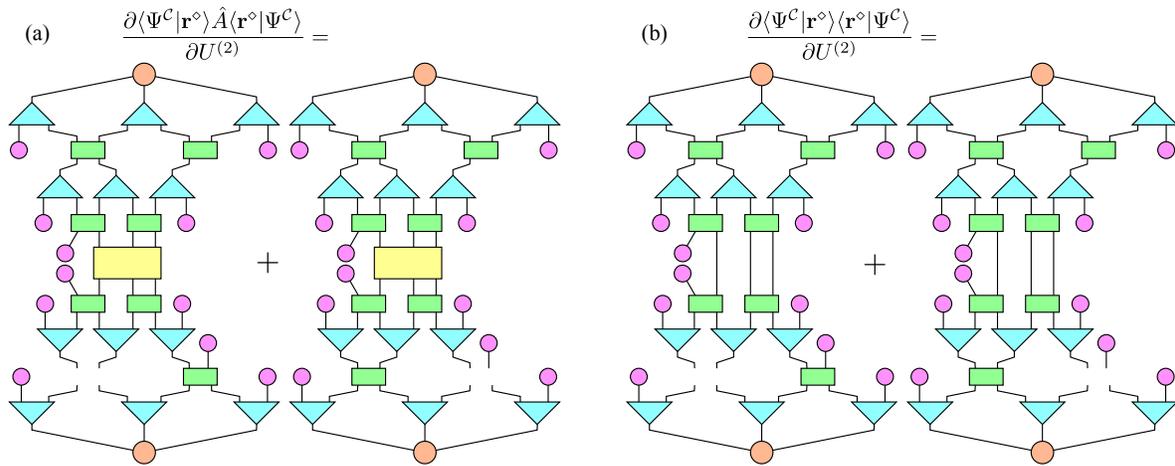
FIG. 7: (Color online) Examples of tensor network diagrams to find the derivatives of $\langle\Psi^{\mathcal{C}}|\mathbf{r}^{\diamond}\rangle\hat{A}\langle\mathbf{r}^{\diamond}|\Psi^{\mathcal{C}}\rangle$ and $\langle\Psi^{\mathcal{C}}|\mathbf{r}^{\diamond}\rangle\langle\mathbf{r}^{\diamond}|\Psi^{\mathcal{C}}\rangle$ with respect to the tensor $U^{(2)\dagger}$ (i.e. the transpose of those in Eq. (15)). The derivative with respect to a particular tensor is given by it's *environment*, or the contraction of all the other tensors in the diagram. The chain rule produces multiple terms where a tensor appears more than once.

IV. BENCHMARK: CRITICAL QUANTUM ISING CHAIN

In this section we demonstrate the above techniques with the well-known transverse-field quantum Ising model,

$$\hat{H} = -\sum_i \hat{\sigma}_i^z \hat{\sigma}_{i+1}^z + h\hat{\sigma}_i^x. \tag{20}$$

Such a Hamiltonian can be expressed as a sum of nearest-neighbour terms $\hat{H}_i$, such that

$$\hat{H}_i = -\hat{\sigma}_i^z \hat{\sigma}_{i+1}^z - h/2\big(\hat{\sigma}_i^x + \hat{\sigma}_{i+1}^x\big). \tag{21}$$

We will pay particular attention to the region around the critical point at $h = 1$, which is the most demanding computationally. For concreteness, we use a three-layer binary MERA with periodic boundary conditions, resulting in a lattice of 24 sites in the bottom of the MERA structure. However, each of these sites corresponds to a block of 3 physical spins, making a total of 72 spins. We choose this blocking so that for $\chi \leq 2^3 = 8$ the bond dimension only ever decreases when ascending through the MERA.

A. Expectation values

We now analyze the effectiveness of extracting expectation values from the MERA using Monte Carlo sampling. For perfect sampling techniques, the accuracy can be easily extracted from the variance using Eq. (7). The scaling of the error in the energy, $\Delta E \propto N^{-1/2}$, is shown explicitly for the critical ($h = 1$) system in Fig. 8 (a). The variance of the energy estimator E as a function of h is shown in Fig. 8 (b). Here we have used MERA wavefunctions previously optimized using standard techniques[11] *without* Monte Carlo sampling, that is, sampling is only employed to extract the expectation values.

Notice that the variance is maximal near the critical point at $h = 1$. As the Monte Carlo code effectively samples wavefunctions from the reduced three-site (i.e. nine-spin) density

operator, one would expect the energy variance to increase with the amount of entanglement in the system. For reference we have included the entropy of the three-site density matrix $\rho^{(3)}$ in Fig. 8 (c). This entropy mostly[20] corresponds to the entanglement entropy of three sites with the remainder of the system. We see a strong correlation between the amount of entanglement and the size of the variance of the energy estimator E.

Let us emphasize that Fig. 8 (b) shows that our scheme performs significantly better than directly sampling $\rho^{(3)}$ in either the x or z spin basis. The measured variances are very similar to a diagonal sampling of $\rho^{(3)}$ (i.e. in its diagonal basis). This indicates that the sampling scheme is performing as intended in Sec. II B 3.

In general, an accurate representation of wavefunctions with greater amounts of entanglement will require greater bond dimension χ. These results suggest that the statistical variance generated by this scheme will also increase with the entanglement. To achieve a certain precision in the expectation value of local observables, the number of required samples grows with this variance, and thus with the amount of entanglement and with the minimum suitable value of χ. Therefore, although a *single* sample has cost $\mathcal{O}(\chi^5)$, the *total* cost to obtain a certain precision may have some additional dependence on χ. Nevertheless, no additional dependence was clearly manifest in our simulations at fixed h.

B. Optimization

Finally, we combine Monte Carlo sampling with our unitary-subspace steepest descent algorithm to obtain optimized wavefunctions. In Fig. 9 we plot the energy of the MERA during the optimization process at $h = 1$ and $\chi = 4$, where the simulation progresses through a range of different number of sweeps N per optimization step. In all cases the step size is fixed at $\mu_t = 0.1$. We observe that increasing N
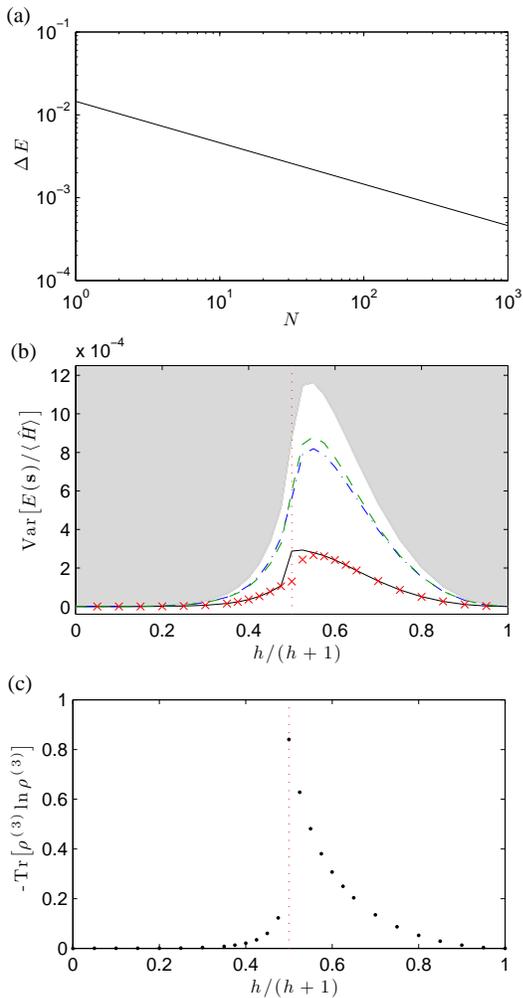
FIG. 8: (Color online) (a) The statistical error of the energy per site at $h = 1$ as a function of the number of samples N follows the classic $N^{-1/2}$ scaling. (b) Variance of the energy estimator (normalized to the expectation value $\langle \hat{H} \rangle$) for optimized MERAs with bond-dimension $\chi = 8$ for various values of h (red crosses). Estimates of the statistical uncertainty are smaller than the symbols. The grey area is eliminated by inequality Eq. (8), bounded by the variance of $\hat{H}$. For comparison, we include the variances expected from several hypothetical samplings of $\rho^{(3)}$. Variances from sampling the spins in the z (or x) basis is indicated by a blue, dash-dot (or green, dashed) line. Our numerical results (red crosses) show remarkable similarity with sampling in the diagonal basis of $\rho^{(3)}$ (black, solid line). (c) The entanglement entropy of $\rho^{(3)}$. In (b) and (c), the critical point at $h = 1$ is indicated with a red dotted vertical line.

improves the quality of the optimized wavefunction and for large values of N the simulation tends to converge towards the same energy obtained with exact contractions, as expected.

Beyond this, accuracy could be improved by increasing χ. It should be noted that we have observed that the steepest descent method (with either exact contractions or sampling) will not always produce wavefunctions of the same quality as the SVD method as it may be more susceptible to local minima or extreme stiffness. However, accuracy can still be systemat-
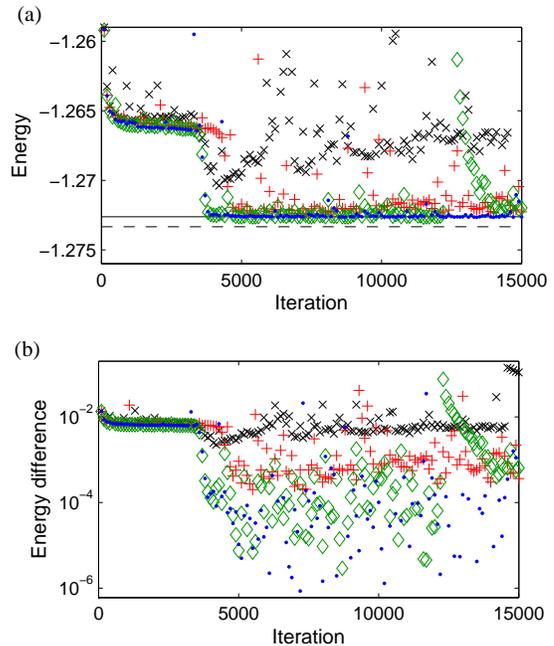

FIG. 9: (Color online) (a) Energy of a wavefunction with $\chi = 4$ during optimization. Each iteration is update using N sweeps, where N is $1, 2, 4, 8$ in the black crosses, red pluses, green diamonds and blue points respectively. Every 100 iterations, we calculate the exact energy corresponding to the current wavefunction, which is plotted here. The solid horizontal line indicates the energy of an optimized $\chi = 4$ MERA using exact contractions and steepest descent (which remains $\approx 7 \times 10^{-4}$ above the the true ground state energy, indicated by the dashed line). The simulation converges for large N, but χ may need to increase for greater accuracy. (b) Difference to the above solid line plotted on a logarithmic scale. The difference reduces with increasing N, though statistical fluctuations are still evident.

ically improved by increasing χ.

In the previous section we noted that the statistical uncertainty peaked around the critical point at $h = 1$, where the entanglement is maximal, and one might expect the optimizations to be most difficult around this point. Plotted in Fig. 10 is the difference in energy between our wavefunctions and the exact, analytic solution for a range of h. We observe that the error decreases away from the critical point, and that there is a clear relationship between the quality of the wavefunction and the number of samples per step, N.

There are several possible limiting factors in variational Monte Carlo optimizations of MERA wavefunctions. One must balance the cost of increasing N, χ and the total number of iterations, to produce results of the desired accuracy. On top of this, the ansatz presents a complicated optimization landscape and one must be careful not to be stuck in local minima.

There is much scope to improve on the above optimization scheme by using more sophisticated approaches. Most obviously, the step-size μ_t and number of samples N_t performed in each iteration could be adjusted as the simulation progresses. For example by choosing the step size to decrease
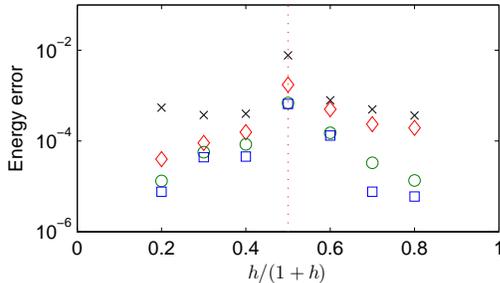
FIG. 10: (Color online) The difference between the final optimized energy and the true ground state for a variety of field strengths h. The markers indicate the number of Monte Carlo sweeps taken between updates, $N = 1, 4, 16, 64$ for the black cross, red diamond, green circle and blue square, respectively. There is a clear trend for improved ground state energy as N increases, and away from the critical point at $h = 1$ (vertical red dotted line).

as $\mu_t \propto 1/t^\alpha$, with N fixed and $0 < \alpha \leq 1$ we are guaranteed convergence to some local minimum[18]. Equivalently, the noise could be reduced at each step by increasing $N \propto t^\beta$, or some combination of both.

In Ref. 8 it was found that using just the sign of the derivative, as well as properties resulting from translational invariance, was sufficient for optimizing a periodic MPS with Monte Carlo sampling. Other approaches existing in the literature may result is significant gains, though it should be noted that approaches requiring the second derivative or Hessian matrix would increase the order of the numerical cost as a function of χ, and would have to be made robust to statistical noise.

V. FUTURE APPLICATIONS

There are several situations where it is most natural to use Monte Carlo sampling to speed-up MERA algorithms. Let us briefly review them.

A. Different MERA structures

In this work we have considered in detail the binary MERA for a 1D lattice with translation invariance. However, even in a translationally invariant, 1D lattice, one has freedom to choose between a large variety of entanglement renormalization schemes, leading to MERA structures with different configurations of isometries and disentanglers.

The ternary MERA, shown in Fig. 11 (a), has a narrower causal cone, with a width of just two sites, and the traditional algorithms for optimizing it have a cost that scales as $\mathcal{O}(\chi^8)$. As a result, the ternary MERA is sometimes favored over the binary MERA. Note that this does not necessarily translate into an improved accuracy in expectation values — the ternary MERA is in general less accurate than binary MERA for the same value of χ.
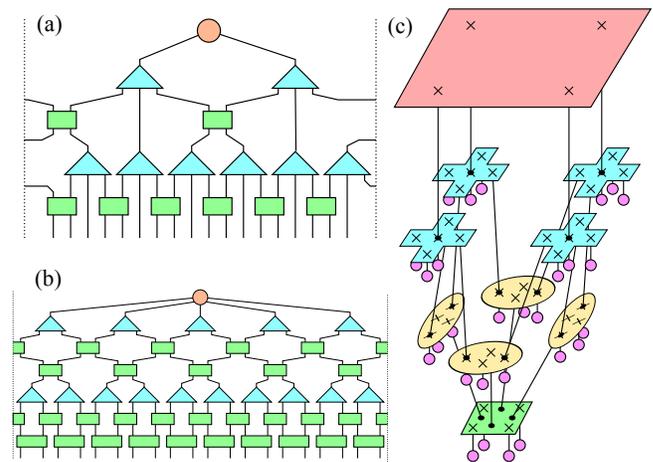

FIG. 11: (Color online) (a) The ternary MERA structure has isometries transforming 3 sites into 1. (b) This two-layer binary MERA structure has added disentanglers which may help represent long range entanglement. (c) Sampling scheme for the lowest layer of the 3×3-to-1 MERA described in Ref. 5.

It is interesting to note that, in the ternary MERA, the perfect sampling algorithm presented here again has cost $\mathcal{O}(\chi^5)$, while Markov chain Monte Carlo, as well as expectation value and environment estimation, is possible with cost $\mathcal{O}(\chi^4)$. It is unclear whether after including autocorrelation effects the Markov chain method performs better, similar, or worse overall compared to the perfect sampling algorithm.

Another possible 1D MERA includes two layers of disentanglers to account for entanglement over larger distances, as depicted in Fig. 11 (b). This MERA has a causal cone that is five sites wide, and traditional algorithms would have numerical cost $\mathcal{O}(\chi^{12})$ and require memory $\mathcal{O}(\chi^{10})$ — limiting χ to rather small values. However, a sampling technique will only require $\mathcal{O}(\chi^7)$ time per sample and $\mathcal{O}(\chi^5)$ memory overall — a huge saving. Note that the power roughly halves when we change from an exact contraction which effectively rescales density matrices from higher layers to lower, to a Monte Carlo scheme which samples wavefunctions from this distribution.

The scaling of computational cost in χ in 2D lattices is even more challenging, mostly because the width of the causal cone (or number of indices included in a horizontal section of the causal cone) is much larger. Once again, sampling wavefunctions will require roughly square-root the number of operations (and memory) needed to calculate the exact reduced-density matrix. For instance, in the 3×3-to-1 MERA presented in Ref. 5, the cost of an exact contraction scales as $\mathcal{O}(\chi^{16} \log L)$, while with Monte Carlo sampling it is possible with just $\mathcal{O}(\chi^8 \log L)$ operations per sample (depicted in Fig. 11 (c)). Memory might be a limiting factor in 2D MERA algorithms, while the temporary memory required for this algorithm is less than that to store the disentanglers and isometries.
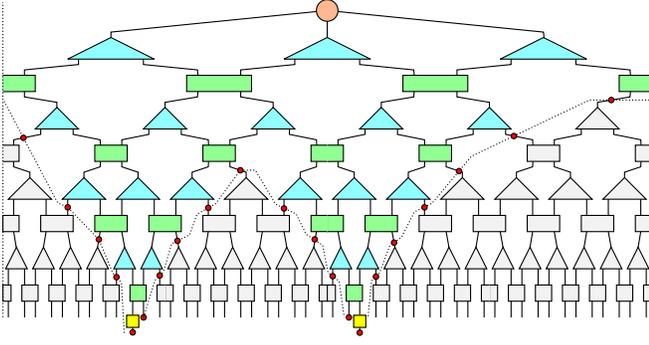
FIG. 12: (Color online) Causal cone of two operators separated by 17 sites in a 4 layer binary MERA. Monte Carlo sampling can be performed on the red points, with a cost $\mathcal{O}(\chi^8)$. An exact contraction of the expectation value costs $\mathcal{O}(\chi^{12})$.

B. Long-range correlations

Another challenge with MERA calculations is the numerical cost of long-range correlations. Take for instance the two-site operator $\langle \hat{A}_i \hat{B}_j \rangle$, for arbitrary sites i and j. The cost required to contract the corresponding tensor network within the binary MERA scheme can scale as much as $\mathcal{O}(\chi^{12})$ — significantly more than the $\mathcal{O}(\chi^9)$ cost for neighbor and next-nearest-neighbor correlations.

However, an estimate of the correlator can be obtained using Monte Carlo sampling at a reduced cost (per sample). In Fig. 12, we depict the causal cone structure of two single-site operators separated by $i - j = 17$ sites in a binary MERA. The cost of Monte Carlo sampling for $\langle \hat{A}_i \hat{B}_j \rangle$ is just $\mathcal{O}(\chi^8)$.

This technique can be extended to 2D lattice systems, where calculating long-range correlations exactly quickly becomes infeasible, even for modest values of χ. Note again that memory constraints are particularly challenging for 2D MERA calculations and Monte Carlo sampling can alleviate this burden.

VI. CONCLUSION AND OUTLOOK

We have outlined and tested a scheme for Monte Carlo sampling with the MERA. Uncorrelated samples can be efficiently generated directly from the wavefunction overlap probability distribution, without needing to resort to Markov chain Monte Carlo methods. From this, expectation values can be extracted and we have demonstrated techniques to reduce the statistical error. We have also presented and demonstrated an algorithm to optimize MERA wavefunctions using sampled energy derivatives.

The numerical results presented here were not intended to be state-of-the-art solutions of the 1D quantum Ising model, but rather to demonstrate feasibility and motivate subsequent applications to 2D systems. In general, Monte Carlo sampling becomes more advantageous for systems with large numbers of degrees of freedom, and we expect 2D MERA to be no exception. Because the reduction in cost in two (and higher) dimensions is so significant, Monte Carlo techniques are a very attractive way to achieve reasonable values of χ with current computers.

Obvious improvements to the code include utilizing symmetries and parallelization to supercomputers, which is straightfoward with our perfect sampling algorithm. Further research into optimization strategies may lead to other improvements (e.g. by reducing the number of iterations or the tendency to find local minima). Reweighting techniques[7] may make the optimization more efficient when approaching the ground state.

Appendix

In this appendix we explain how to compute, with cost $O(n^2 m)$, the $n \times m$ isometric matrix

$$F = A \exp(A^\dagger B - B^\dagger A), \qquad (22)$$

where A is an $n \times m$ isometric matrix, B is a general $m \times n$ matrix, and $n \leq m$. The naïve approach would be to evaluate the $m \times m$ matrix $A^\dagger B - B^\dagger A$ and compute its exponential, with cost $\mathcal{O}(m^3)$, before multiplying by A. However, noting that the exponent does not have full rank (the rank is at most $2n$), we can hope to find a faster method.

Taking the Taylor expansion

$$\begin{aligned} F &= A + B - AB^\dagger A + \frac{1}{2!}\big(BA^\dagger B - BB^\dagger A \\ &\quad - AB^\dagger B + AB^\dagger AB^\dagger A\big) + \dots, \end{aligned} \qquad (23)$$

we observe that the result can be achieved with a series of multiplications between $n \times n$ matrices $C = AB^\dagger$, $C^\dagger$ and $D = BB^\dagger$, post-multiplied by either A or B, requiring total cost $\mathcal{O}(n^2 m)$.

The following algorithm calculates the Taylor expansion to order p.

$C \leftarrow AB^\dagger; D \leftarrow BB^\dagger$
$W \leftarrow I; X \leftarrow 0; Y \leftarrow I; Z \leftarrow 0$
for $i = 1 \rightarrow p$ **do**
    $T \leftarrow -(WC^\dagger + XD)/i$
    $X \leftarrow (W + XC)/i$
    $W \leftarrow T$
    $Y \leftarrow Y + W$
    $Z \leftarrow Z + X$
end for
$F \leftarrow YA + ZB$

In the binary MERA, where isometries are $\chi \times \chi^2$ matrices (i.e. $n = \chi$, $m = \chi^2$), the cost of this algorithm scales as

$\mathcal{O}(\chi^4)$, compared with the cost $\mathcal{O}(\chi^6)$ of the naïve approach. This algorithm becomes particularly important for a tree tensor network and for the MERA in two dimensions, where the naïve approach becomes more expensive, in powers of χ, than

a sampling (thus becoming the bottle neck of an optimization based on sampling), whereas the above algorithm remains competitive.

1. S. R. White, Phys. Rev. Lett **69**, 2863 (1992).
2. S. Östlund and S. Rommer, Phys. Rev. Lett. **75**, 3537 (1995).
3. G. Vidal, Phys. Rev. Lett. **99**, 220405 (2007); G. Vidal, Phys. Rev. Lett. **101**, 110501 (2008).
4. F. Verstraete and J. I. Cirac, eprint arXiv:cond-mat/0407066 (2004).
5. G. Evenbly and G. Vidal, Phys. Rev. Lett. **102**, 180406 (2009).
6. L. Cincio, J. Dziarmaga, and M. M. Rams, Phys. Rev. Lett. **100**, 240603 (2008).
7. N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, Phys. Rev. Lett. **100**, 040501 (2008).
8. A. W. Sandvik and G. Vidal, Phys. Rev. Lett. **99**, 220602 (2007).
9. H. J. Changlani, J. M. Kinder, C. J. Umrigar, and G. K.-L. Chan, (preprint) arXiv:0907.4646 (2009).
10. A. J. Ferris and G. Vidal, In preperation (2011).
11. G. Evenbly and G. Vidal, Phys. Rev. B **79**, 144108 (2009).
12. F. Mezzacapo, N. Schuch, M. Boninsegni, and J. I. Cirac, New J. Phys. **11**, 083026 (2009).
13. K. H. Marti, B. Bauer, M. Reiher, M. Troyer, and F. Verstraete, New J. Phys. **12**, 103008 (2010).
14. H. C. Jiang, Z. Y. Weng, and T. Xiang, Phys. Rev. Lett. **101**, 090603 (2008).
15. R. N. C. Pfeifer, G. Evenbly, and G. Vidal, Phys. Rev. A **79**, 040301 (2009).
16. A. Hjorungnes and D. Gesbert, Signal Processing, IEEE Transactions on **55**, 2740 (2007), ISSN 1053-587X.
17. T. Abrudan, J. Eriksson, and V. Koivunen, Signal Processing, IEEE Transactions on **56**, 1134 (2008), ISSN 1053-587X.
18. A. Harju, B. Barbiellini, S. Siljamäki, R. M. Nieminen, and G. Ortiz, Phys. Rev. Lett. **79**, 1173 (1997), URL http://link.aps.org/doi/10.1103/PhysRevLett.79.1173.
19. Notice that the second term in the brackets of Eq. (15) arises from the change in the normalization of the wave function. Although we are dealing with unitary/isometric tensors which ensure $\langle\Psi^{\mathcal{C}}|\Psi^{\mathcal{C}}\rangle = 1$, small changes in arbitrary directions may break the unitarity and modify the norm. Interestingly, even when projecting into the unitary tangent space (see below) where this term averages to zero, its inclusion is important to reduce the sampling error — sometimes by several orders of magnitude.
20. There may be very small contributions to the entropy resulting from the anisotropy of the MERA. Also note that H and $\rho^{(3)}$ share a $\mathbb{Z}_2$ symmetry which is broken by the MERA wavefunction for $h < 1$, and the *exact* ground state should have $S = \ln 2$ entanglement entropy as $h \to 0$.