Practical learning method for multi-scale entangled states

2012 New J. Phys. 14 085004

(http://iopscience.iop.org/1367-2630/14/8/085004)

New Journal of Physics
The open-access journal for physics

Practical learning method for multi-scale entangled states

Olivier Landon-Cardinal and David Poulin
Département de Physique, Université de Sherbrooke, Sherbrooke, Québec, J1K 2R1, Canada
E-mail: olivier.landon-cardinal@usherbrooke.ca
and david.poulin@usherbrooke.ca

New Journal of Physics **14** (2012) 085004 (25pp)
Received 5 April 2012
Published 1 August 2012
Online at http://www.njp.org/
doi:10.1088/1367-2630/14/8/085004

Abstract. We describe two related methods for reconstructing multi-scale entangled states from a small number of efficiently-implementable measurements and fast post-processing. Both methods only require single-particle measurements and the total number of measurements is polynomial in the number of particles. Data post-processing for state reconstruction uses standard tools, namely matrix diagonalization and conjugate gradient method, and scales polynomially with the number of particles. Both methods prevent the build-up of errors from both numerical and experimental imperfections. The first method is conceptually simpler but requires unitary control. The second method circumvents the need for unitary control but requires more measurements and produces an estimated state of lower fidelity.

Contents

1. Introduction

Quantum state tomography [1] is a method to learn a quantum state from measurements performed on many identically prepared systems. This task is crucial not only to assess the degree of control exhibited during the preparation and transformation of quantum states, but also in comparing theoretical predictions to experimental realizations. For instance, numerical methods are used to compute the ground states or thermal states of model quantum systems. Quantum state tomography could be used to check that the experimental state corresponds to the predicted one, thus providing an essential link between theory and experiments. For example, one could in principle use tomography to settle the question [2] of which states correctly describe the quantum Hall fluid at various filling parameters.

In practice, however, the state of n particles is described by a number of parameters that scales exponentially with n. Therefore, tomography requires an exponential number of identically prepared systems on which to perform exponentially many measurements needed to span a basis of observables that completely characterizes the state. Furthermore, solving the inference problem to determine the quantum state that is compatible with all these measurement outcomes requires an exponential amount of classical post-processing. These factors limit tomography to at most a few tens of particles. Thus, the ability to efficiently check the state of experimental systems is a current roadblock to demonstrating quantum control over increasingly large quantum systems.

While this exponential blowup in resources is unavoidable for a generic state due to the exponentially large dimension of Hilbert space, many states encountered in nature have special properties that could be exploited to simplify the task of tomography. Indeed, many states of interest can be described by only a polynomial number of parameters. In fact, the

overwhelming majority of tomographic experiments performed to date [3–9] were used to learn states described with only a few parameters. Such variational states—belonging to a family of states specified with only a few parameters—are omnipresent in many-body physics because they are tailored for numerical calculations and can predict many phenomena observed in nature (Kondo effect, superconductivity, fractional statistics, etc). One example, familiar to both the quantum information and computational many-body communities, is matrix product states (MPS) [10–13] that are at the heart of the density matrix renormalization group (DMRG) numerical method [14, 15], suitable for the description of one-dimensional (1D) quantum systems with finite correlation length [16].

Variational tomography consists of identifying a state within a variational family of states. Since those states are described by a few parameters, variational tomography amounts to learning those parameters and might require fewer resources than required by general tomography.

Recently, we and others have demonstrated [17] that tomography can be performed efficiently on MPS, i.e. such states can be learned from a small number of simple measurements and efficient classical post-processing. Here, we take this result one step further and demonstrate that it is possible to efficiently learn the states associated with the multi-scale entanglement renormalization ansatz (MERA) introduced by Vidal [18], for which efficient numerical algorithms to minimize the energy of local Hamiltonians exist [19]. As opposed to MPS, these MERA states are not restricted to one dimension and can describe systems with algebraic decay of correlations. This last distinction is important because one of the most interesting phenomena in physics, quantum phase transitions, leads to a diverging correlation length and are therefore not suitably described by MPS. In contrast, MERA have been successfully used to study numerous many-body models, such as the critical Ising model in 1D [19–22] and 2D [23], and can also accurately describe systems with topological order [24–26].

In this article, we present two related methods to learn the 1D MERA description of a state using tomographic data obtained from local measurements performed on several copies of the states. Our learning methods for MERA are based on the identification of the unitary gates in the quantum circuit that outputs the MERA state. In that regard, this article is a continuation of our work on MPS and is reminiscent of early methods to numerically optimize MERA tensors [27]. However, going from MPS to MERA is nontrivial because MERA exhibits a spatial arrangement of gates that is more elaborate. Since MERA is a powerful numerical tool, our learning method bridges the gap between numerical simulations and experiments by allowing the direct comparison of numerical predictions to experimental states.

The first method we present requires unitary control of the system and the ability to perform tomography on blocks of a few particles, which can be realized using the correlations between single-particle measurements. Crucially, the size of those blocks does not depend on the total size of the system, making it a *scalable* method. In an experiment, one cannot know beforehand if the state in the lab is a MERA. However, our method contains a built-in certification procedure from which one can assess the proper functioning of the method as the experiments are performed and conclusively determine if the state is well described by the MERA. This method is experimentally challenging since it requires the ability to apply generic quantum gates on a few particles. We therefore present a second method that builds on the first one, but completely circumvents the need for unitary control. Thus, this second MERA learning method can be implemented with existing technologies. While we consider the second method

to be the most interesting from a practical point of view, we chose to present the first approach for pedagogical reasons as it is conceptually simpler.

The rest of the paper is organized as follows. We begin with a more comprehensive exposition of variational tomography in the next section. For pedagogical reasons, we then present the method for MERA learning which uses unitary control in section 3. Section 3.1 explains how to identify the disentanglers. We start by deriving a necessary condition for the existence of suitable disentangler and then turn this criterion into a heuristic objective function that we minimize numerically. In section 3.2, we carefully analyze the buildup of errors in our procedure and show that errors only accumulate linearly with the size of the system. In section 3.3, we present numerical benchmarks of our tomography method. In section 4, we present the simplified method that does not require unitary control. We demonstrate in section 4.1 that it is not necessary to apply the disentanglers to the experimental state since we can simulate the effect of those disentanglers numerically, albeit at the cost of more repeated measurements, as analyzed in section 4.2. We emphasize the practical aspect of the method with an example in section 4.3. Finally, we discuss the error scaling and certification in section 4.4. In section 5, we discuss the relationship between the numerical tractability of a variational family of states and the ability to learn efficiently the variational parameters. Finally, we present in appendix B a tool to contract two different MERA states, which allows for the efficient comparison of a MERA whose parameters have been identified experimentally using our method to a predicted theoretical MERA state.

2. Variational tomography

Consider a variational family of states $\mathcal{C} \equiv \{|\phi(\alpha)\rangle\}_\alpha$, i.e. a family of states described by the possible values of a collection of variational parameters collectively denoted α. To be useful and interesting, this family of states has to be numerically tractable and physically faithful. Numerical tractability requires that states are specified by a few (at most polynomial) parameters, but also that this efficient representation allow for the efficient numerical computation of quantities of interest, such as the energy of the system, correlation functions, or more generally expectation values of local observables. Physical faithfulness means that states of the variational family exhibit the properties of the system of interest.

Given access to multiple copies of the system in a state σ, one is interested in (i) verifying that the experimental state σ belongs to the variational family, i.e. check that σ is close to a state $|\phi(\alpha_0)\rangle \in \mathcal{C}$; and (ii) learn those variational parameters α_0 directly from experimental measurements. In our methods, those two questions will be answered simultaneously. One will perform *variational tomography* to extract the values α_0 and use the same experimental data to *certify* that the experimental state is indeed close to the state $|\phi(\alpha_0)\rangle$.

The main motivation for variational tomography is that, from a logical standpoint, it might be possible to entirely learn the state from a small number of measurements since it is described by a small number of parameters. This is to be contrasted with quantum state tomography whose goal is to learn a generic state in the full Hilbert space, and thus has to identify an exponential number of parameters. However, in variational tomography, it is not necessarily the case that the variational parameters can be accessed directly by measurements; one could imagine that the only way to identify a quantum state inside a variational family would be to

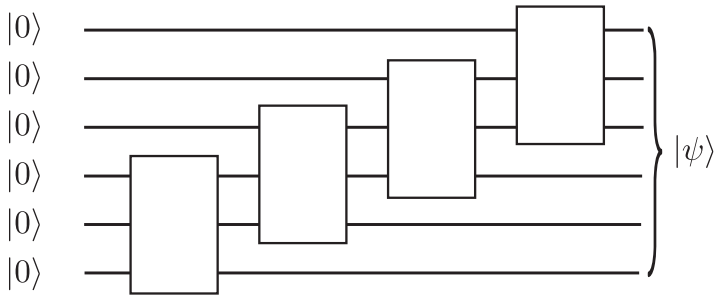

Figure 1. MPS generation from a staircase circuit of local unitary gates.

perform quantum state tomography in the full Hilbert space and then compute the variational description of the state. The relevant questions thus become (i) for which variational families is variational tomography efficient? and (ii) for such classes, what experimental toolbox is required?

A natural variational family of states are those produced by a polynomial-size quantum circuit starting from, e.g. the all $|0\rangle$ state. In that case, the variational parameters α correspond to the description of the circuit namely a description of its gates and their locations. More refined variational families are obtained by further restricting these quantum circuits, e.g. by fixing the layout of gates but leaving the value of each gate arbitrary. In that case, the parameter space of α scales precisely with the number of gates.

The prime example is MPS. These states are the output of a quantum circuit whose gate layout is a staircase pattern shown in figure 1. Furthermore, MPS allow for efficient computation of expectation values of local observables and can reproduce physics of gapped 1D system. In [17], we and others have shown how to identify those gates efficiently. In this work, we turn our attention to MERA states, that are also described by quantum circuits but with a more elaborate gate layout.

3. Multi-scale entanglement renormalization ansatz (MERA) learning with unitary control

3.1. Identifying the disentanglers

MERA states can be described as the output of a quantum circuit [18] whose structure is represented on figure 2 (as seen with inputs on the top and output at the bottom). For simplicity, we will focus in section 3 on a 1D binary MERA circuit for qubits, but our method generalizes to all 1D MERA states, i.e. particles could have more internal states—thus accounting for a larger MERA refinement parameter χ—and isometries could renormalize several particles to one effective particle. The circuit contains three classes of unitaries. Disentanglers (represented as $\square$) are two-qubit unitary gates; isometries (represented as $\triangle$) are also two-qubit gates but with one input qubit always in the $|0\rangle$ state; the top tensor (represented as $\bigcirc$) is a special case of isometry that takes as input two qubits in the $|00\rangle$ state. Each renormalization layer is made of a row of disentanglers and a row of isometries. Disentanglers remove the short-scale entanglement between adjacent blocks of two qubits while isometries renormalize each pair of qubits to a single qubit. Each renormalization layer performs theses operations on a different length scale. The quantum circuit thus mirrors the renormalization procedure that underlies the MERA.

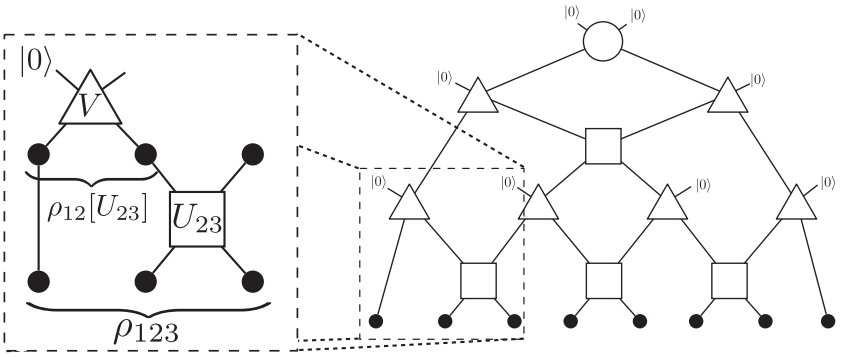

Figure 2. The optimal disentangler $\tilde{U}$ can be computed from the tomographic estimation of the density matrix ρ_{123} on the first three qubits. Once applied, the resulting state $\tilde{\rho}_{12}[\tilde{U}]$ is very close to a rank 2 matrix. Thus, there exist a unitary V that transform $\tilde{\rho}_{12}[\tilde{U}]$ into a state with the first qubit in the state $|0\rangle$.

Learning a MERA state amounts to identifying the various gates in that circuit. By inverting the time direction, we can think of this circuit as transforming a MERA state into the all $|0\rangle$ state, by sequentially disentangling 'ancillary' qubits from the system. The intuitive idea behind our scheme is to proceed by varying the isometries and disentanglers until these ancillary qubits reach the state $|0\rangle$ for each row of isometries. We will exploit this feature to numerically determine each disentangler.

3.1.1. Necessary condition for disentangler. Consider the n qubits at the lowest layer of the MERA. Let ρ_{123} be the reduced density matrix on the first three qubits (see figure 2). If the state is exactly a MERA, there exists a unitary U_{23} acting on qubits 2 and 3 (see left of figure 2) for which

$$\rho_{12}\left[U_{23}\right] = \mathrm{Tr}_3\left[\left(\mathbb{I}_1 \otimes U_{23}\right)\rho_{123}\left(\mathbb{I}_1 \otimes U_{23}^{\dagger}\right)\right] \tag{1}$$

has rank at most 2. Indeed, if the rank of $\rho_{12}[U_{23}]$ was strictly greater than 2, it would be impossible for the isometry V (see left of figure 2) to map the density matrix $\rho_{12}[U_{23}]$ to a state with one of the qubit in the state $|0\rangle$ because the dimension of the space $|0\rangle \otimes \mathbb{C}^2$ would be strictly smaller than the dimension of the support of the density matrix. Hence, we have the necessary criterion

$$\exists \tilde{U}_{23} \quad \rho_{12}[\tilde{U}_{23}] \text{ has rank less or equal than 2.} \tag{2}$$

To find a unitary that fulfills this criterion, it is necessary to know the state ρ_{123}, and this can be achieved by brute-force tomography on these three qubits. Once the original state on the three qubits is known, one has to perform a search over the space of unitaries to find a suitable disentangler. To do this, we will define in section 3.1.2 an objective function to minimize numerically.

Once this optimal unitary operator $\tilde{U}$ has been found numerically, it is necessary to consider how it modifies the quantum state before learning the other elements of the circuit. One obvious way to do so is to apply the unitary transformation to the experimental state and continue the procedure on the transformed state. This amounts to executing the circuit, and should in the end map the experimental state to the all $|0\rangle$ state. For pedagogical reasons, we

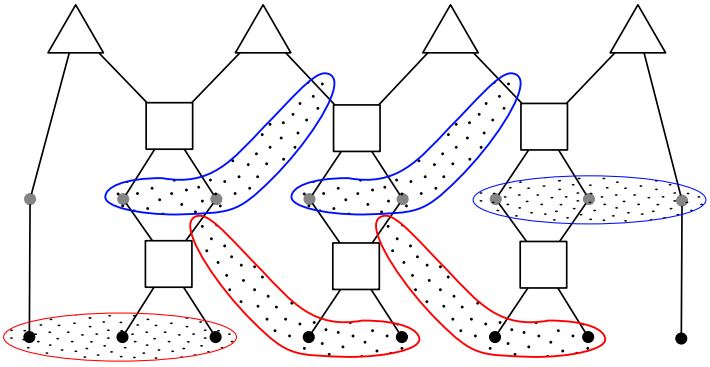

Figure 3. Identification of the disentanglers using two successive sweeps of the chain. Dotted regions cover particles on which brute-force tomography is performed. The first sweep (red dotted regions) finds unitaries starting from the left end of the chain. Those unitaries will be used as initial guesses for the second sweep (blue striped regions) that starts from the right end of the chain.

will first present our scheme assuming that the state is transformed at every step this way. Of course, such unitary control increases the complexity of the scheme and could be out of the reach of current technologies. However, in section 4, we will explain how this unitary transformation can be circumvented at the cost of increasing the number of measurements.

After the optimal disentangler $\tilde{U}$ has been applied to the state, we need to identify the unitary V that rotates the density matrix on the first two qubits such that the first qubit is brought to the $|0\rangle$ state; cf figure 2 left. This does not require any additional tomographic estimate since we already know the descriptions of the state on the first three qubits and the disentangler. We can thus compute the state on the first two qubits $\rho_{12}[\tilde{U}]$ and diagonalize it to obtain the eigenvectors corresponding to its two nonzero eigenvalues. The unitary V is chosen to map those two eigenvectors to the space $|0\rangle \otimes \mathbb{C}^2$, i.e. V rotates the qubits such that the support of the density matrix is mapped to a space where the first qubit is in the $|0\rangle$ state.

All other disentanglers of this layer can be found by recursing the above procedure. Once a disentangler has been identified, it is experimentally applied to the system and brute-force tomography is performed on the next block of three qubits.

Notice that for the last block of a layer of an open boundary MERA, a single unitary is responsible for minimizing the rank of two density matrices. One possible way to handle this is to get a tomographic estimate of the state on the last four qubits and to try to minimize the rank of both reduced matrices. Another way, for which we have opted in our numerical simulations, is to perform multiple sweeps over the layer. For instance, the disentanglers will first be identified from left to right and then the next sweep will be performed from right to left, using the disentanglers found in the first sweep as initial guesses in the space of unitaries (see figure 3).

The number of sweeps can be increased for better accuracy. Additional sweeps requires either to extract the tomographic estimates after each new sweep or to perform tomography on blocks of slightly larger size. Multiple sweeps would also allow to apply our method to MERA states with periodic boundary conditions in 1D and could be useful for 2D-MERA states. While

this would be an interesting continuation of our work, we focus on 1D-MERA for the rest of the paper.

3.1.2. Heuristic objective function. One of the steps in our protocol consists of identifying the unitary $\tilde{U}$ that minimizes the rank of $\rho_{12}[U]$; cf equation (1). There are many distinct ways this can be done and in this section, we present a practical heuristic to accomplish this task. Minimizing the rank of the density matrix $\rho_{12}[U]$ is not a suitable numerical task because, even if the experimental state is an exact MERA, the inferred density matrix will typically have full rank due to machine precision and the imperfect tomographic estimation of ρ_{123}. Thus, we turn the problem of finding $\tilde{U}_{23}$ into an optimization problem by considering the spectral decomposition of $\rho_{12}[U] = \sum_k \lambda_k |\psi_k\rangle\langle\psi_k|$ where the eigenvalues are sorted in decreasing order $\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3 \geqslant \lambda_4$. If $\rho_{12}[U]$ has most of its support on a 2D space, it will have two small eigenvalues that are typically nonzero due to imperfections. We thus consider the objective function

$$f(U, \rho_{123}) = \sum_{k>2} \lambda_k \tag{3}$$

and we perform a minimization over the space of unitaries to determine the optimal unitary $\tilde{U}$. This objective function has a well-defined operational meaning—it is the probability of measuring the disentangled qubit in the state $|1\rangle$ after the isometry V has been applied. We will see in section 3.2 that this property can be used to certify the distance between the experimental and the reconstructed states.

Another way to think about this objective function is to consider the characteristic polynomial $P[X]$ of $\rho_{12}[U]$ which is of the form

$$P[X] = X^4 - X^3 + aX^2 - bX + c, \tag{4}$$

where the coefficients a, b and c are positive since they correspond to the sum of product of the positive eigenvalues of the density matrix. In particular, coefficient b is the sum of all products of three eigenvalues, i.e. $b = \lambda_1\lambda_2\lambda_3 + \lambda_1\lambda_2\lambda_4 + \lambda_1\lambda_3\lambda_4 + \lambda_2\lambda_3\lambda_4$. In order for the rank of the density matrix to be 2, it is sufficient for all 4 products of three eigenvalues to vanish, i.e.

$$\rho_{12}[U] \text{ of rank less than } 2 \iff b = 0. \tag{5}$$

Thus, another suitable objective function is the positive coefficient b, which is a polynomial in the entries of $\rho_{12}[U]$. Indeed, using Bocher formula, coefficient b can be expressed as $6b = 1 - 3\,\mathrm{Tr}A^2 + 2\,\mathrm{Tr}A^3$ where $A = \rho_{12}[U]$. Thus, b is a well-behaved function with respect to the density matrix. Note also that b can be expressed without diagonalizing the density matrix $\rho_{12}[U]$. We will focus on minimizing equation (3) in all subsequent discussions and numerical results.

3.1.3. Numerical minimization over unitary space. Minimization of equation (3) is performed using a conjugate gradient method. We first have to account for the fact that the unitary manifold is not a vector space. To get around this problem, we go to the tangent space by writing any unitary U as the result of a Hamiltonian evolution, i.e., there exists a Hermitian matrix H such that $U = \mathrm{e}^{\mathrm{i}H}$. It is then possible to use the standard conjugate gradient method. Let us sketch the algorithm in more detail.

First, we select a unitary U_0 either at random or from an initial guess (provided for instance by a previous sweep). We will search the unitary space by generating a sequence of unitaries $\{U_k\}$. At the kth step of the minimization, the algorithm is the following.

1. We center the unitary space at point U_{k-1} by defining $\rho_k = (\mathbb{I} \otimes U_{k-1})\rho_{k-1}(\mathbb{I} \otimes U_{k-1})^\dagger$.
2. We compute the gradient $G^{(k)}$ by parametrizing the Hamiltonian H on 2 qubits by its decomposition on the Pauli group $H = \sum_{\mu\nu} h_{\mu\nu}\sigma_\mu \otimes \sigma_\nu$ where $\sigma_\mu \in \{\mathbb{I}, \sigma_x, \sigma_y, \sigma_z\}$ is a Pauli matrix. We successively evaluate the component of the gradient $G^{(k)}$ in the direction (μ, ν) by looking at the effect of the test unitary $U_{\mu,\nu} = \mathbb{I} + i\epsilon\sigma_\mu \otimes \sigma_\nu$ on the objective function, i.e. $G^{(k)}_{\mu,\nu} = \frac{f(U_{\mu,\nu}, \rho_k) - f(\mathbb{I}, \rho_k)}{\epsilon}$ where ϵ is a small number.
3. Instead of following the gradient, which would generally undo some of the minimization performed in the previous steps, we use a conjugate gradient method where the new direction of search $\tilde{G}^{(k)}$ is optimized by taking into account the direction used in the previous step $\tilde{G}^{(k-1)}$ through the Polak–Ribière formula. More precisely, $\tilde{G}^{(k)} = G^{(k)} + \beta\tilde{G}^{(k-1)}$ in which the real parameter β is defined as $\beta = \max\left(0, \frac{G^{(k)}\cdot(\tilde{G}^{(k-1)} - G^{(k)})}{\tilde{G}^{(k-1)}\cdot\tilde{G}^{(k-1)}}\right)$.
4. We perform a line search along the direction $\tilde{G}^{(k)}$ by considering the family of unitaries $\exp(-it\sum_{\mu,\nu}\tilde{G}_{\mu,\nu}\sigma_\mu \otimes \sigma_\nu)$ and optimizing the parameter t to find t_{opt}. We then define

$$U_k = \exp\left(-it_{\text{opt}}\sum_{\mu,\nu}\tilde{G}_{\mu,\nu}\sigma_\mu \otimes \sigma_\nu\right), \tag{6}$$

which ends the kth iteration.

We iterate until the objective function is close enough to zero or that improvement has stopped. Note K the total number of iteration steps. The disentangler returned by the algorithm is $\tilde{U} = U_K U_{K-1} \cdots U_0$.

This method is *heuristic* since the objective functions present no characteristic that would ensure the convergence of the conjugate gradient method. In particular, our search over unitary space depends on the starting point, i.e., the unitary chosen in the first iteration. Indeed, some starting points will lead the heuristic to a local minima where it will get stuck. In order to avoid this phenomenon, we can repeat the overall search by picking at random (according to the unitary Haar measure) different initial points which lead to potentially different minima and keep the smallest of those minima. In any case, this is a minimization problem over a space of *constant* dimension, so the method used to solve it does not affect the scaling with the number of particles n. Ultimately, we can always use a finite mesh over the unitary space and use brute-force search. Nevertheless, we found numerically that this heuristic works well.

A more serious problem is that a choice of unitary that is optimal *locally*, in the sense that it minimizes equation (3), could be sub-optimal *globally* as it might lead to a state for which it is impossible to find a disentangler obeying equation (3) elsewhere in the circuit. This is a phenomenon that is more likely to occur when the minimum is degenerate, i.e., there exists several distinct (modulo gauge) exact disentanglers for the state. However, we have performed numerical experiments on randomly generated MERA states as well as physically motivated states and found that the conjugate gradient performs well (see section 3.3).

3.2. Error analysis

In practice, due to numerical and experimental imperfections, the first disentangled qubit will not be exactly in the $|0\rangle$ state, but merely close to it. This situation arises from the conjunction of three causes: (i) the experimental state of the system is not exactly a MERA, but merely close to one, (ii) the tomographic estimate of the density matrices on blocks of three qubits are slightly inaccurate due to noisy measurements and experimental finite precision, (iii) the numerical minimization did not find the exact minimum.

3.2.1. Preventing error amplification by post-selection. Our error analysis will show that the buildup of errors is linear in the number of disentanglers of the MERA circuit, which is itself linearly proportional to the number of particles in the experimental state. Essentially, the distance between the reconstructed state and the experimental state is the sum of the error made at each elementary step when estimating a disentangler and an isometry. Fortunately, the error made at each elementary step is not amplified by errors made at previous steps. The key to isolate each step from the others is to measure the qubit that should have been disentangled in the computational basis. With high probability the qubit will be found in the $|0\rangle$ state. While the probability of measuring the $|0\rangle$ outcome depends on previous errors, the post-selected state is now free from previous errors. The interest of this post-selection is two-fold. Firstly, it prevents errors in previous steps to contaminate the state and amplify the error made at the current step, thus limiting the error propagation. Secondly, by accumulating statistics on this measurement, we can estimate the probability of outcome $|0\rangle$ and use it to bound the distance of the reconstructed state to the actual state in the lab. Therefore, our procedure comes with a *built-in certification process*. We now describe the error analysis in more detail.

3.2.2. Error at each elementary step. Recall the notation of figure 2. Due to numerical and experimental imperfections, the state on qubits 1, 2 and 3 after applying the disentangler $\tilde{U}_1$ and the isometry V_1 is not exactly in the $|0\rangle \otimes \mathbb{C}^{2(n-1)}$ subspace but contains a small component orthogonal to that space. Thus, it has the form

$$V_1 \tilde{U}_1 |\psi\rangle = \frac{|0\rangle|\eta_1\rangle + |e_1\rangle}{\sqrt{1 + \langle e_1|e_1\rangle}}, \tag{7}$$

where $|\eta_1\rangle$ is the normalized pure state on qubits 2 to n if qubit 1 had been completely disentangled from the chain and $|e_1\rangle$ is some *sub-normalized* vector supported on the subspace $|1\rangle \otimes \mathbb{C}^{2(n-1)}$. The isometry V_1 is chosen to minimize the norm of $|e_1\rangle$, i.e. to minimize $\epsilon_1 \equiv \langle e_1|e_1\rangle$.

Further along the layer, the state after applying k disentanglers and k isometries will be of the form

$$V_k \tilde{U}_k \cdots V_1 \tilde{U}_1 |\psi\rangle = \frac{|0\rangle^{\otimes k}|\eta_k\rangle + |e_k^{\mathrm{ac}}\rangle}{\sqrt{1 + \epsilon_k^{\mathrm{ac}}}}, \tag{8}$$

where the first term $|0\rangle^{\otimes k}|\eta_k\rangle$ is the normalized state had the k qubits in position 1, 3 ... $2k-3$ been completely disentangled from the chain and $|e_k^{\mathrm{ac}}\rangle$ is the accumulated error vector orthogonal to the space where those k qubits are in the $|0\rangle^{\otimes k}$ state, whose square norm is $\epsilon_k^{\mathrm{ac}} \equiv \langle e_k^{\mathrm{ac}}|e_k^{\mathrm{ac}}\rangle$. In order to find the optimal disentangler and isometry, we measure the last disentangled qubit in the computational basis and post-select on the $|0\rangle$ outcome, which occurs

with probability $(1+\epsilon_k^{\mathrm{ac}})^{-1}$. We then perform brute force tomography and identify numerically the disentangler and the isometry that minimizes the norm of the error vector $|e_{k+1}\rangle$ such that

$$V_{k+1}\tilde{U}_{k+1}|\eta_k\rangle = \frac{|0\rangle|\eta_{k+1}\rangle + |e_{k+1}\rangle}{\sqrt{1+\epsilon_{k+1}}}. \tag{9}$$

Applying this disentangler and isometry to the whole state of the chain, one gets

$$V_{k+1}\tilde{U}_{k+1}\cdots V_1\tilde{U}_1|\psi\rangle = \frac{|0\rangle^{\otimes k+1}|\eta_{k+1}\rangle + |e_{k+1}^{\mathrm{ac}}\rangle}{\sqrt{1+\epsilon_{k+1}^{\mathrm{ac}}}}, \tag{10}$$

where the accumulated error vector at step $k+1$ is

$$|e_{k+1}^{\mathrm{ac}}\rangle = |e_{k+1}\rangle + \sqrt{1+\epsilon_{k+1}}V_{k+1}\tilde{U}_{k+1}|e_k^{\mathrm{ac}}\rangle \tag{11}$$

and the square of its norm $\epsilon_{k+1}^{\mathrm{ac}} \equiv \langle e_{k+1}^{\mathrm{ac}}|e_{k+1}^{\mathrm{ac}}\rangle$ obeys the recurrence relation

$$1+\epsilon_{k+1}^{\mathrm{ac}} = (1+\epsilon_{k+1})\left(1+\epsilon_k^{ac}\right) \tag{12}$$

since the elementary error vector $|e_{k+1}\rangle$, for which all previous ancillary particles have been disentangled, is orthogonal to the vector $V_{k+1}\tilde{U}_{k+1}|e_k^{\mathrm{ac}}\rangle$. Thus,

$$1+\epsilon_{k+1}^{\mathrm{ac}} = \prod_{i=1}^{k+1}(1+\epsilon_i)\,. \tag{13}$$

3.2.3. Global error. After the choice of m disentanglers and m isometries, the reconstructed state is $|\phi\rangle = V_m^\dagger\tilde{U}_m^\dagger\cdots V_1^\dagger\tilde{U}_1^\dagger|0\rangle^{\otimes m+1}|\eta_m\rangle$. Its distance to the actual experimental state $|\psi\rangle$ can be stated in terms of the (in) fidelity as

$$\begin{aligned} 1-|\langle\phi|\psi\rangle|^2 &= 1-1/\left(1+\epsilon_m^{\mathrm{ac}}\right) \\ &= 1-1\Big/\prod_{i=1}^{m}1+\epsilon_i. \end{aligned} \tag{14}$$

Practically, one is interested in guaranteeing that the reconstructed state is close to the experimental state, up to global error E, i.e. to guarantee that $1-|\langle\phi|\psi\rangle|^2 \leqslant E$. Suppose that all error vectors are bounded, i.e. that for all step i, we have $\epsilon_i \leqslant \varepsilon$. Inverting equation (14), it suffices that

$$\varepsilon \leqslant (1-E)^{-1/m} - 1 \simeq E/m \tag{15}$$

in the limit where the tolerable global error E is small. Thus, we see that *errors accumulate linearly* and that a precision inversely proportional to the number of disentanglers is sufficient to ensure a constant global error. Furthermore, statistics on the post-selection performed at each step allows to estimate each ϵ_k^{ac}—and therefore ϵ_m^{ac} through equation (13)—that gives direct access to the distance between the reconstructed and experimental states.

Finally, from these estimates of ϵ_i, one can identify particular steps of the procedure that have gone wrong. This information can be used to turn the scheme into an *adaptive* one. Suppose the error is particularly large for a given step. This might be due to an important amount of entanglement concentrated in one region of space, e.g. near a defect, which can be accounted for by increasing the MERA refinement parameter χ locally, i.e. by using disentanglers acting on a larger number of qubits. In practice, χ could be increased until the error is below some target threshold.

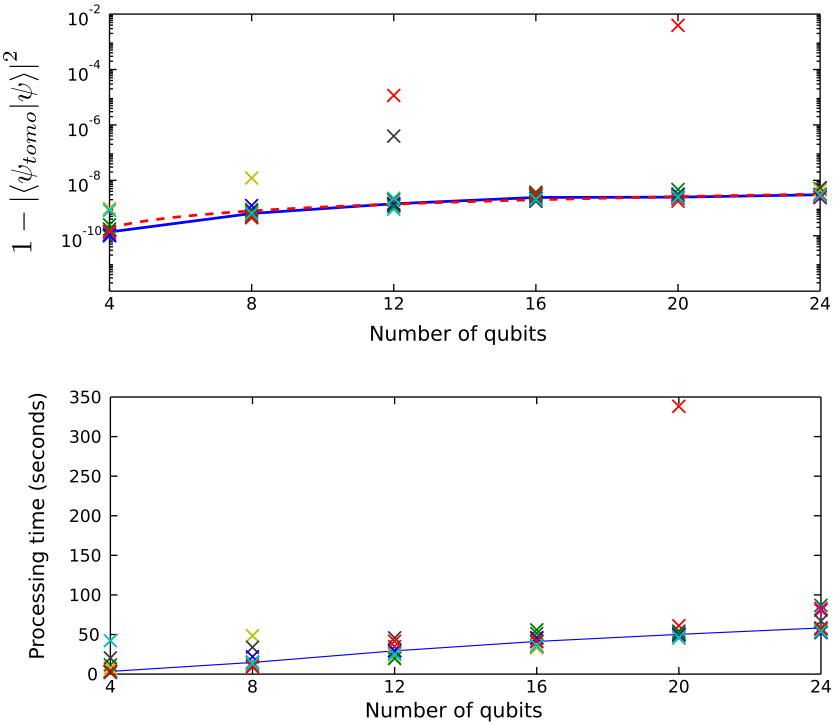

Figure 4. (top) Infidelity to the 'experimental state', i.e. $1 - |\langle\psi_{\text{tomo}}|\psi\rangle|^2$ where $|\psi\rangle$ is a random MERA on n qubits and $|\psi_{\text{tomo}}\rangle$ is the state reconstructed from the MERA tomography method using three sweeps. (bottom) Processing time (on a standard laptop) to perform MERA tomography using three sweeps. Both figures exhibit 10 runs for each number of qubits $n \in \{8, 12, 16, 20, 24\}$. In both figures, each $\times$ represents results for one random MERA. The full lines represent median for each number of qubits. The dashed line on the top figure is the linear approximation to the median. Notice that the numerical minimization can fail to converge as illustrated by the atypical data points. For instance, for one of the 20-qubit MERA, the processing time was one order of magnitude longer than the average and the infidelity six orders of magnitude larger than average.

3.3. Numerical performance

3.3.1. Benchmarking results. We have performed numerical simulations to benchmark the performances of the conjugate gradient method in our setting. We have generated random MERA states—by picking each unitary gate in the circuit from the unitary group Haar measure—, simulated the experiment on those states, and use our algorithm to infer the initial MERA state. We did not introduce noise in measurements to simulate experimental errors since the error analysis indicates how those errors would build up.

As mentioned before, there is no guarantee that our minimization procedure converges to the true minimum, resulting in small imperfections in the reconstructed state. Figure 4 (top) shows the distance between the reconstructed state and the actual state. As indicated by the dashed line, these results are in good agreement with a linear scaling of the error, where the

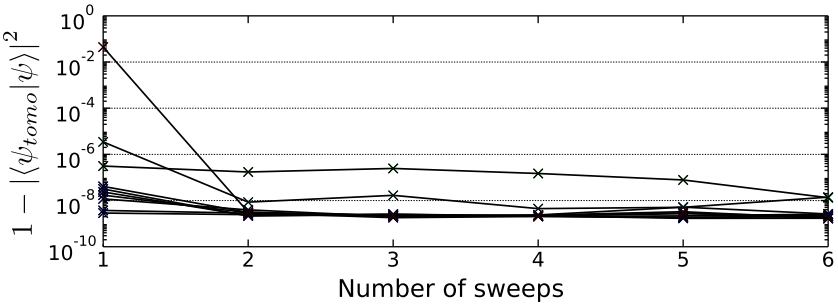

Figure 5. Infidelity to a 20-qubit state using a reconstructed method with a variable number of sweeps. Each line corresponds to a different random MERA.

source of errors is due to finite machine precision and approximate minimization of the objective function.

The inference algorithm's complexity is dominated by the conjugate gradient descents, and therefore scales linearly with the number of disentanglers or the number of particles in the system. Figure 4 (bottom) shows the actual run time of the inference algorithm for different randomly chosen MERA states and of various sizes. Once again, we see a good agreement with a linear dependence with the system size. Systems of up to 24 qubits can easily be handled in a few minutes of computation and requires 1197 different measurement settings for each sweep of the 24 qubit system. This is to be contrasted with the 656 100 experiments needed to reconstruct the state of 8 qubits in [3] and the post-processing of the data that took approximately a week [28]. Additional sweeps improve the convergence as showed on figure 5.

We also tested our method on a physical model, namely the 1D Ising model with transverse field at the critical point. The results obtained where coherent with what is expected from the approximation of the true ground state with a MERA with refinement parameter $\chi = 2$.

3.3.2. Possible improvements. Note the presence of isolated points on the graphs of figure 4 that achieve a lower fidelity and required a longer processing time. These cases appear because the heuristic fails to find a global minimum. It appears that in some cases, a unitary transformation U_{23} meeting criterion equation (3) is not sufficient to guarantee that it will be possible to find subsequent disentanglers obeying equation (3). Put another way, *locally* minimizing the objective function might not lead to a *global* optimum. Indeed, consider the following example. Let $|\psi\rangle$ be a MERA state whose first qubit is disentangled from the rest of the chain, i.e. $|\psi\rangle = |0\rangle|\phi\rangle$. The rank of the density matrix on the first two qubits is at most 2 and that remains true after *any* unitary is applied on qubits 2 and 3. Thus, any choice of disentangler minimizes the objective function equation (3)—we say that the minimum is degenerate. In this case, even the identity, i.e. applying no disentangler at all, achieves the minimum. However, suppose the state $|\phi\rangle$ on qubits 2 to n is highly entangled and that removing part of this entanglement between qubits 2 and 3 was crucial to be able to reconstruct its MERA description. In this case, applying the identity on qubits 2 and 3, even if locally optimal, was not globally optimal. Hence, minimizing the objective function equation (3) seems to be necessary but not sufficient to successively identify all disentanglers.

Although our numerical simulations suggest that this situation is rather atypical and can be suppressed with additional sweeps (see figure 3), it is possible to overcome this problem

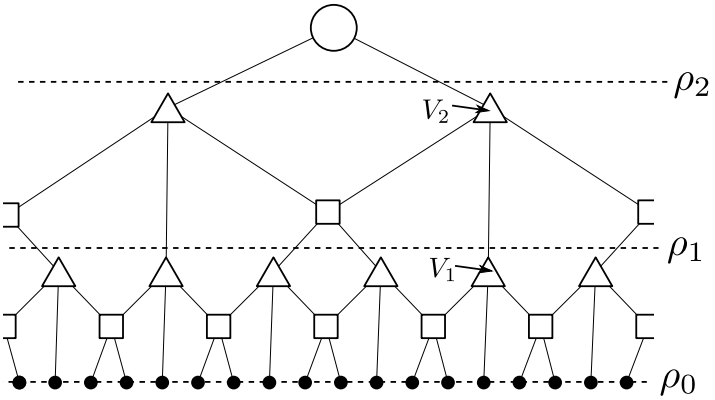

Figure 6. MERA as a renormalization procedure that creates a sequence of states $\{\rho_\tau\}_\tau$.

by imposing additional constraint on the disentangler. For instance, one can demand that the second qubit be in a state as pure as possible, effectively minimizing the entanglement between the last qubit of one block and the first qubit of the next block. This corresponds to the following modified objective function

$$f\left(\tilde{\rho}_{12}\left[U\right]\right) = \sum_{k>2} \lambda_k + \epsilon\lambda_2 \tag{16}$$

i.e., we add a small perturbation that will only take action when the two smallest eigenvalues of $\tilde{\rho}_{12}[U_{23}]$ are very small and will further constrain the search. This slight modification solved the problematic situation we considered, and there exist many other heuristics to improve the method. This problem, and its heuristic solutions, are similar to those encountered when using the MERA to estimate the ground states of a Hamiltonian numerically.

4. MERA learning without unitary control

For pedagogical reasons, we presented our learning method in a way that required disentanglers and isometries to be physically applied to the experimental state in order to unravel the circuit. In this section, we will show how to circumvent unitary control at the price of slightly more elaborate numerical processing and consuming more copies of the state. The main idea is to numerically simulate how measurements performed on the original, unaltered experimental system would be transformed if the unraveling circuit had been applied.

4.1. Simulating measurements on renormalized state

A MERA is an ansatz that corresponds to a renormalization procedure. Each renormalization step maps a state to another one on fewer particles and schematically corresponds to a layer of the MERA circuit. Applying the first layer and removing the ancillary particles that have been (approximately) disentangled maps the experimental state ρ_0 on n particles to a state ρ_1 on fewer particles (see figure 6). Recursively, this procedure constructs a sequence of states $\{\rho_\tau\}_\tau$.

To get from $\rho_{\tau-1}$ to ρ_τ, one can either perform this mapping *physically* by experimentally applying the gates corresponding to the MERA layer, or one can *compute* the function mapping

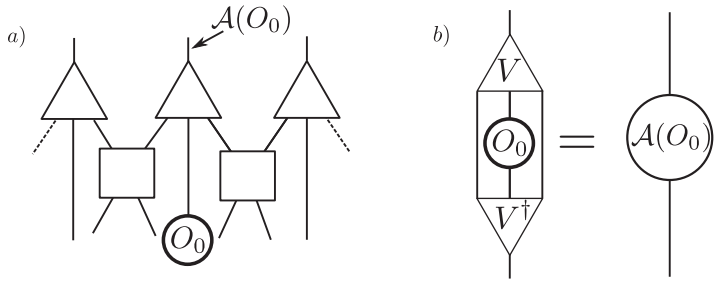

Figure 7. Ascending superoperator and renormalized observables for a ternary MERA. (a) Ternary MERA with one site observable O_0 that is transformed into a renormalized observable $\mathcal{A}(O_0)$ on the renormalized state. (b) Tensor contraction corresponding to the ascending superoperator $\mathcal{A}$.

$\rho_{\tau-1}$ to ρ_τ from the description of the gates. As in [19], define an ascending superoperator $\mathcal{A}$ that maps an operator $O_{\tau-1}$ acting on layer $\tau-1$ to an operator O_τ acting on the next layer τ

$$O_\tau = \mathcal{A}_\tau(O_{\tau-1}) \tag{17}$$

such that

$$\mathrm{Tr}[\rho_\tau \mathcal{A}_\tau(O_{\tau-1})] = \mathrm{Tr}[\rho_{\tau-1} O_{\tau-1}]. \tag{18}$$

This recursively carries over to the experimental state ρ_0

$$\mathrm{Tr}[\rho_\tau \mathcal{A}_\tau \circ \cdots \circ \mathcal{A}_1(O_0)] = \mathrm{Tr}[\rho_0 O_0]. \tag{19}$$

Thus, in order to extract information from a density matrix ρ_τ, one can measure the expectation value of several observables O_0^i on the density matrix ρ_0. Measuring those observables will effectively amount to measuring the observables $O_\tau^i \equiv \mathcal{A}_\tau \circ \cdots \circ \mathcal{A}_1(O_0^i)$ on the density matrix ρ_τ.

The ascending superoperator can be computed from the knowledge of the disentanglers and isometries. Its exact form depends on the physical support of the observable. For instance, for ternary MERA, we can restrict to ascending superoperator that only depends on the isometries of the MERA [22] (see figure 7). This is a simple example where an experimental observable on one particle is mapped to observable on one renormalized particle. More generally, observables on many sites will be ascended to observables on fewer sites. Any choice of observables is valid as long as the renormalized observables $\{O_\tau^i\}_i$ span the support of the reduced density matrix ρ_τ.

4.2. Overhead in the number of measurements

This procedure leads to an overhead in the total number of measurements because renormalized observables are less efficient at extracting information. Suppose (for clarity) that we measure Pauli observables $\{O_0^i\}_i$ on the experimental state. These observables are orthonormal for the Hilbert–Schmidt inner product and thus maximize information extraction. However, the renormalized observables $O_1^i \equiv \mathcal{A}_1(O_0^i)$ need not be orthonormal. Consider their Gram matrix $G_{ij} = \mathrm{Tr}[O_1^i (O_1^j)^\dagger]$ which can be diagonalized by a unitary matrix Z. Its normalized eigenvectors $R_1^i = \frac{1}{\sqrt{\lambda_i}} \sum_j Z_{ij} O_1^j$ are orthonormal observables but cannot be directly measured because they do not correspond to simple observables on the experimental state, but instead

to linear combination of them. Thus, to reconstruct the density matrix $\rho_1 = \sum_i r_1^i R_1^i$, the expectation values $r_1^i = \mathrm{Tr}\rho_1 R_1^i$ have to be computed by taking a linear combination of the expectation values $o_0^j \equiv \mathrm{Tr}\rho_0 O_0^j$ on the experimental state

$$r_1^i = \frac{1}{\sqrt{\lambda_i}} \sum_j Z_{ij} \mathrm{Tr}\rho_1 O_1^j = \frac{1}{\sqrt{\lambda_i}} \sum_j Z_{ij} o_0^j. \tag{20}$$

Due to limited number of repeated measurements, estimation of each o_0^i will present a variance $\mathbb{V}(o_0^i)$. Suppose that measurements are repeated enough times to ensure that all variances are below a precision threshold, i.e. $\mathbb{V}(o_0^i) \leqslant \epsilon$. Since r_1^i is a linear combination of those measurements, it will have a variance $\mathbb{V}(r_1^i) = \frac{1}{\lambda_i} \sum_j |Z_{ij}|^2 \mathbb{V}(o_0^j) \leqslant \frac{\epsilon}{\lambda_i} \sum_j |Z_{ij}|^2$. Therefore, in order to ensure a precision ϵ on the estimate of r_1^i, this imprecision needs to be compensated by multiplying the number of repeated measurements by the *conditioning factor* $\lambda_i^{-1} \sum_j |Z_{ij}|^2$.

When scaling operators on τ layers, the conditioning factors for each layer will multiply (in the worst case) but we expect the conditioning for each layer to be a constant independent of system size. Thus, the total number of measurements will remain *polynomial* in the number of particles since there is only a *logarithmic* number of renormalization layers.

We can make this argument rigorous for critical systems that exhibit scale-invariance, precisely the physical systems for which MERA was introduced. Due to scale-invariance, the ascending operator $\mathcal{A}_\tau$ will not depend on the index of the layer and we refer to it as the scaling superoperator $\mathcal{S}$ [22]. Its diagonalization yields eigenvectors ϕ_α called scaling operators associated to eigenvalues μ_α. In [22], it was shown that those eigenvalues are related to the scaling dimensions Δ_α of the underlying conformal field theory (CFT) by $\Delta_\alpha = \log_3 \mu_\alpha$ where the basis of the log depends on the MERA type (here we consider a ternary MERA for clarity). Scaling operators ϕ_α can be used as observables to extract information about states in higher level of the MERA. Indeed, one can simulate a measurement of $\mathcal{S}^\tau(\phi_\alpha)$ on ρ_τ by measuring the observable ϕ_α on ρ_0. We can analyze the increase in the number of measurements by distinguishing two sources of imprecision. Firstly, to reconstruct ρ_τ one has to use normalized operator $\phi_\alpha^{[\tau]} = 3^{\tau\Delta_\alpha} \mathcal{S}^\tau(\phi_\alpha)$ whose increased statistical fluctuations have to be compensated by performing additional measurements. Secondly, diagonalizing the Gram matrix of the $\phi_\alpha^{[\tau]}$ will introduce another conditioning factor. However, this Gram matrix is independent of the layer since $G_{\alpha\beta}^{[\tau]} = \mathrm{Tr}[\phi_\alpha^{[\tau]} \phi_\beta^{[\tau]}] = \mathrm{Tr}[\phi_\alpha \phi_\beta]$. Thus, the conditioning factor for layer τ will be the product of a factor exponential in the number of layers and a constant factor coming from the orthonormalization. Overall, this amounts to a conditioning factor that scales polynomially with system size.

4.3. Example

We have performed a numerical simulation to approximate the ground state of the critical Ising model $H = -\sum_{\langle i,j \rangle} \sigma_x^i \otimes \sigma_x^j - \sum_k \sigma_z^k$ based on the MERA structure illustrated on figure 6. Due to periodic boundary conditions of the MERA and the translational invariance of the Hamiltonian, all disentanglers and isometries on a given layer are identical. In addition, because the model is critical and hence (nearly) scale invariant, the gates are the same at each level of the MERA, except the first one, where scale invariance has not quite settled in.

Consider the isometry V_1 that maps three physical qubits to one renormalized particle, as illustrated on figure 7(b). Given V_1, we can compute how the one-site Pauli observables $\{O_0^i\} = \{\mathbb{I}, \sigma_x, \sigma_y, \sigma_z\}$ are transformed as one-site observables O_1^i on the renormalized particle

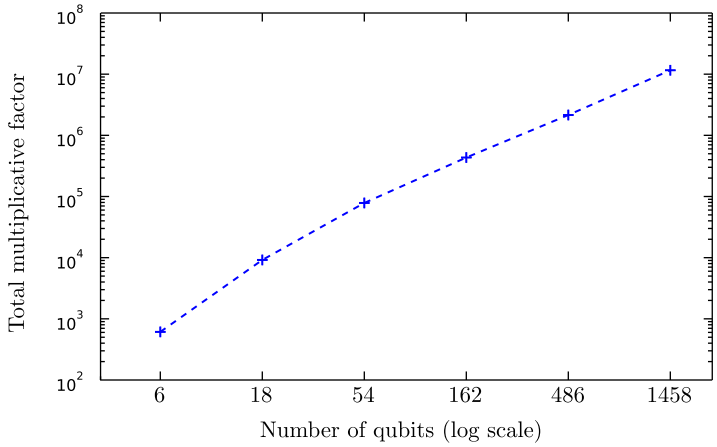

Figure 8. Total multiplicative factor to learn a ternary MERA corresponding to the ground state of the 1D Ising model with critical transverse field. The total number of measurements is the product of this factor with the number of measurements $N(\epsilon) \approx 1/\epsilon^2$ necessary to estimate the expectation value of a Pauli observable to accuracy ϵ on the physical particle. Notice that the curve approaches a straight line, which indicates that the total number of measurements scales *polynomially* with the number of particles.

(see figure 7). This linear transformation can be represented by a matrix M such that $O_1^i = \sum_j M_{ij} O_0^j$. In our particular example,

$$M = \begin{pmatrix} 1 & 0.062 & 0 & 0.460 \\ 0 & 0.779 & 0 & -0.213 \\ 0 & 0 & 0.401 & 0 \\ 0 & 0.330 & 0 & 0.294 \end{pmatrix}.$$

Notice that σ_y is mapped on itself so it is a scaling operator, i.e. the measurement of the observable σ_y on the physical particle is sufficient to simulate the measurement of σ_y on the renormalized particle since $\mathrm{Tr}\rho_1\sigma_y = (0.401)^{-1}\mathrm{Tr}\rho_0\sigma_y$. To estimate the expectation value of σ_y on the renormalized particle to some accuracy ϵ thus requires $(0.401)^{-2}N(\epsilon)$ measurements, where $N(\epsilon) \approx 1/\epsilon^2$ is the number of measurements required to estimate σ_y to accuracy ϵ on the physical particle.

Since M is invertible, one can similarly deduce the expectation values of Pauli observables on the renormalized state by taking linear combination of expectation values of observables on the experimental state ρ_0. The fact that the image by V_1 of one-site observables on ρ_0 spans the space of one-site observables on ρ_1 is quite favorable. In general, one could have to consider physical observables on a few sites to generate all renormalized observables. Furthermore, the choice of observables on the physical state could be optimized to minimize the amplification of repeated measurements.

Proceeding similarly for higher renormalization layers, the total number of measurement will be a product of the standard statistical overhead $N(\epsilon) \approx \epsilon^{-2}$ with a total multiplicative factor that results from the conditioning factors for each layers. Figure 8 shows this total multiplicative factor for the critical Ising model with the MERA structure shown at figure 6. The total number of qubits is $n = 2 \times 3^k$ for $k \leqslant 5$. Recall that quantum state tomography of the

8 qubit W state required more than 656 100 measurements in [3]. Using the same elementary precision, i.e. $N(\epsilon) = 100$, our method allows to learn a 18-qubit system with a comparable number of measurements (917 656 measurements).

4.4. Error propagation and certification without unitary control

In the scheme with unitary control, certification is made possible thanks to experimental measurements that directly project the experimental state on a particular MERA state, by making sure that the ancillary qubits are disentangled after each layer of renormalization. Keeping records on the success rate of these measurements thus yields information about the overlap. Not only does this allow for direct fidelity estimate, but the resulting post-selection also limits error propagation; cf section 3.2.1.

Using a scheme similar to the one explained in section 4.2, one could simulate the physical measurement of the disentangled particles on the level of the experimental state. While this is possible, it requires in principle an exponential number of measurements since a single-qubit observable at a high level of the MERA will typically have support on its entire causal cone. That statement can seem contradictory to our claim of section 4.2 that density matrices on high layers of the MERA can be estimated by using only a polynomial number of measurements. The apparent contradiction arises from the fact that efficient estimation is based on the assumption that the state is well-approximated by a MERA. Obviously, one cannot rely on that property to certify that the state is well approximated by a MERA.

Despite this difficulty, we will now argue that it is possible to certify the state without unitary control, using the 18-qubit ternary MERA of figure 6 as a concrete example. As a first step, we will study how errors propagate during the MERA learning process. Because a small change in the state can in principle alter the MERA structure substantially, it is difficult to rigorously bound this error propagation. Thus, this discussion, as well as appendix A, should be seen as plausible arguments for the scaling of errors rather than a rigorous proof.

We note $|\psi_0\rangle$ the experimental state and $|\psi_1\rangle$ the state after the first layer of renormalization. Note U_1 the unitary that gives rise to the isometry V_1, i.e.

$$V_1 = (|0\rangle\langle 0| \otimes \mathbb{I} \otimes |0\rangle\langle 0|)\, U_1. \tag{21}$$

Note $\mathbb{U}_1$ the unitary that corresponds to the first layer of renormalization. Without post-selection, the state after that layer reads

$$|\psi_1\rangle = \mathbb{U}_1|\psi_0\rangle = \frac{|0\rangle^{\otimes 12}|\eta_1\rangle + |e_1\rangle}{\sqrt{1+\epsilon_1}}, \tag{22}$$

where $|e_1\rangle$ is a sub-normalized error vector whose square norm is $\epsilon_1 \equiv \langle e_1|e_1\rangle$.

Our scheme to learn the MERA presented in section 4.2 uses the identity $\mathrm{Tr}[O_1^i|\eta_1\rangle\langle\eta_1|] = \mathrm{Tr}[O_0^i|\psi_0\rangle\langle\psi_0|]$—valid when $\epsilon_1 = 0$ or when post-selecting on the all $|0\rangle$ state of ancillary qubits—to estimate the expected value of O_1^i on $|\eta_1\rangle$ through measurements of O_0^i on the physical state $|\psi_0\rangle$. Repeating with various O_1^i enabled us to reconstruct the density matrices σ_1^i that are marginals of $|\eta_1\rangle$ on small blocks of particles. However, in the presence of errors and without post-selection, this identity does not hold. Using equations (21) and (22), we obtain

$$\left|\mathrm{Tr}\left[O_1^i|\eta_1\rangle\langle\eta_1|\right] - \mathrm{Tr}\left[O_0^i|\psi_0\rangle\langle\psi_0|\right]\right| \leqslant 2\epsilon_1. \tag{23}$$

Thus the reconstruction algorithm will estimate the reduced density matrices $\sigma_1^{[i]}$ up to some error $E_1 \in O(\epsilon_1)$. The precise relation between E_1 and ϵ_1 depends on the method

(linear inversion, maximum-likelihood, etc) used to reconstruct the density matrix from the measurements. The disentangling algorithm will then take those error-prone contaminated reduced density matrices $\sigma_1^{[i]}$ as input to identify the disentanglers and isometries of the second renormalization layer. As a consequence, even in the absence of any other source of errors, these disentanglers and isometries will typically be chosen sub-optimally. Thus, we see that errors accumulate along the renormalization flow.

The accumulated error E_2^{ac} after the second layer of renormalization will contain a component inherited from the accumulated error E_1^{ac} of the previous layer and an intrinsic error ϵ_2. As before, the intrinsic error can be caused by a numerically sub-optimal choice of disentangler or simply because the state is not exactly a MERA. The distinction between this accumulated error E_k^{ac} and the accumulated error in the presence of post-selection, denoted ϵ_k^{ac} in section 3.2, stems from the fact that the algorithm used to find the disentangler operates on a contaminated state. The analysis presented in appendix A shows that if there is an accumulated error E_k^{ac} after k layers of renormalization, the error E_{k+1}^{ac} is bounded by

$$1 + E_{k+1}^{\mathrm{ac}} \leqslant \left(1 + E_k^{\mathrm{ac}}\right)^3 (1 + \epsilon_{k+1}) \tag{24}$$

which results in an error that grows exponentially with the number of layers. The number of layers being logarithmic in the number of particles, the error thus scales polynomially with system size. The fidelity between the experimental state and the reconstructed state is given by the final accumulated error

$$|\langle\psi_0|\psi_{\mathrm{tomo}}\rangle|^2 = \left(1 + E_{\mathrm{top}}^{\mathrm{ac}}\right)^{-1}. \tag{25}$$

For a ternary MERA with $n = 2 \times 3^k$ particles, we can use equation (24) and (25) to bound the fidelity as a function of the intrinsic error made at each layer of the MERA

$$|\langle\psi_0|\psi_{\mathrm{tomo}}\rangle|^2 \geqslant (1+\epsilon_1)^{-n/3} (1+\epsilon_2)^{-n/3^2} \cdots \left(1 + \epsilon_{\mathrm{top}}\right)^{-1}. \tag{26}$$

Given this relation between fidelity and intrinsic errors, we now turn to the problem of certification.

To certify the fidelity of the reconstructed state, one therefore needs to estimate the intrinsic error ϵ_k made at each layer. This error is the probability that the disentangled particles are not in the $|0\rangle$ state. With unitary control, estimation of ϵ_k could be performed by experimentally projecting all those ancillary qubits in the $|0\rangle$ state and accumulating statistics. Without unitary control, estimation of ϵ_k has to be performed *locally* by estimating the projection orthogonal to $|0\rangle\langle 0| \otimes \mathbb{I} \otimes |0\rangle\langle 0|$ for each isometry. The expectation value of this projector, the *leakage error* f_k^i, is precisely the objective function used in our numerical procedure, c.f. Equation (3). In order to estimate the error ϵ_k for the layer k, we use the union bound, which results in an estimate that scales with the number of isometries in the layer. Supposing that for each isometry the leakage error is smaller than ϵ, we get the upper bound

$$\epsilon_k \leqslant \sum_i f_k^i \leqslant n\epsilon/3^k. \tag{27}$$

Combining equation (26) and (8), we get

$$|\langle\psi_0|\psi_{\mathrm{tomo}}\rangle|^2 \geqslant \prod_{k=1}^{\log_3 n} \left(1 + \frac{n}{3^k}\epsilon\right)^{-n/3^k}. \tag{28}$$

Assuming that ϵ i smaller than $1/n^2$, the right hand side of (28) reduces to $1-\epsilon\sum_{k=1}^{\log_3 n}\left(\frac{n}{3^k}\right)^2 \sim 1-\frac{n^2}{8}\epsilon$. Thus, for states that are at most $1/n^2$ away in fidelity from a MERA state, we can certify the distance between the reconstructed and the experimental states. This certificate is rather loose due to the use of the union bound and assuming that error accumulate in the worst possible way. A more accurate estimate of the fidelity can be performed using the Monte Carlo technique of [29] but requires a number of measurements that will scale exponentially with system size.

5. Discussion

In this paper, we have presented a tomography method that allows to efficiently learn the MERA description of a state by patching together tomography experiments on a few particles and using fast numerical processing. The method is heuristic but works very well in numerical simulations. A complete analytic understanding of how to find an optimal disentangler at each step would be desirable, but may well be intractable. With regards to experimental use, the method should be thought of as a proof of principle and is flexible enough to be adapted to accommodate many experimental constraints.

One issue of fundamental interest raised by our work is the relationship between the numerical tractability of a variational family of states and the ability to learn efficiently the variational parameters. In order to be interesting, variational family of states must not only be described by a small number of parameters, but also allow for the efficient numerical computation of quantities of interest, such as the energy of the system, correlation functions, or more generally expectation values of local observables. On its own, an efficient representation is of limited computational usefulness. For instance, the Gibbs state or ground state of a local Hamiltonian is described by a few parameters—a temperature and a local Hamiltonian—but does not allow to extract physical quantities of interest efficiently. Another example is the variational family of projected entangled pair states or PEPS [30], the generalization of MPS to system in more than one dimension. While PEPS have been instrumental in better understanding of quantum many-body systems, they are in general intractable numerically [31].

Is there a relation between numerical tractability and efficient tomography? The method presented in [17] to learn a MPS from local measurements made explicit use of the energy minimization algorithm for MPS; namely DMRG [14, 15]. This example suggests that numerical tractability could imply that learning the variational parameters is possible. In that regard, MERA are intriguing states because they live at the frontier of tractability. Indeed, in more than 1 dimension, MERA states are a subclass of PEPS [32] with a bond dimension independent of system size [33]. While the computation of expectation values of local observables is believed to be intractable for PEPS, it is efficient for MERA. In one dimension, MERA can be seen as MPS if one allows the bond dimension to grow *polynomially* with the size of the system (while MPS are usually required to have a *constant* bond dimension). Thus, while MPS manipulations typically have a computational cost linear in the number of particles, 1D-MERA manipulations have a computational cost which is super-linear (but yet polynomial).

Beyond MPS and MERA, one could consider states obtained from a quantum circuit where the positions of the gates are known and try to identify those gates. An interesting question is then to characterize what topology of circuits makes it possible to learn gates efficiently. This could lead to formal methods for the testing and verification of quantum hardware.

Acknowledgments

This work was partially funded by the Natural Sciences and Engineering Research Council of Canada (NSERC). OLC acknowledges the support of NSERC through a Vanier scholarship. DP acknowledges financial support by the Lockheed Martin Corporation. We thank Marcus da Silva and Andy Ferris for stimulating discussions and Andy Ferris for sharing numerical data on scale-invariant MERAs.

Appendix A. Error accumulation without post-selection

The modified scheme that circumvents the need for unitary control modifies the error propagation. Namely, the scaling of the overall error increases since the error at each step will depend on previous errors, that will accumulate and amplify subsequent errors. In this section, we present a heuristic argument to understand how this accumulation affects the overall precision of the scheme.

Because there is no post-selection, the algorithm used to find the disentanglers and isometries will not operate on the state $|\eta_k\rangle$ (see equation (8)) but instead will operate on the contaminated state

$$(1-\varepsilon)\,|\eta_k\rangle\langle\eta_k| + \varepsilon |E_k^{\mathrm{ac}}\rangle\langle E_k^{\mathrm{ac}}|, \tag{A.1}$$

where $\varepsilon \equiv \frac{E_k^{\mathrm{ac}}}{1+E_k^{\mathrm{ac}}}$ and $|E_k^{\mathrm{ac}}\rangle$ is a subnormalized error vector resulting from the accumulation of all previous errors, and whose square norm is $E_k^{\mathrm{ac}} \equiv \langle E_k^{\mathrm{ac}}|E_k^{\mathrm{ac}}\rangle$. Thus, the numerical minimization returns a unitary that is typically not the optimal disentangler for $|\eta_k\rangle$.

In the degenerate case—when the objective function equation (3) has many distinct minima (modulo gauge)—this might change the disentangling unitary drastically, either because the objective function is flat or because the solution jumps from one local minima to another. In the latter case, the errors are causing the algorithm to explore different local minima, which is actually an exploration that is desirable to find the global minimum. Degenerate minima correspond to hard instances of the problem, and it is conceivable that in these cases the state can neither be learned nor certified. In the non-degenerate case however, we can heuristically bound the accumulation of errors. We proceed in three steps. Firstly, we analyze how the modification of the input state will affect the disentangling unitary returned by the algorithm. Secondly, we evaluate how this imperfect disentangler impacts the error propagation. Thirdly, we bound the error at step $k+1$ in terms of the error at step k. This technical result is used in section 4.4.

A.1. Disentangling unitary without post-selection

Let us denote by $\tilde{U} = \mathrm{e}^{\mathrm{i}\tilde{H}}$ the unitary that is returned by our algorithm in the presence of post-selection, i.e. the unitary that minimizes the objective function equation (3) for the post-selected state ρ. If we don't post-select on the ancillary particles being disentangled, this minimization is not performed on the perfect state $|\eta_k\rangle$ but rather on the contaminated state given by equation (A.1). We want to know how much $\tilde{U} = \arg\min_U f(U, \rho)$ changes when ρ goes from $|\eta_k\rangle$ to the state of equation (A.1).

Using the chain rule, we formally write $\frac{\partial \tilde{U}}{\partial \rho} = \frac{\partial \tilde{U}}{\partial f}\frac{\partial f}{\partial \rho}$. The first term, $\frac{\partial \tilde{U}}{\partial f}$, quantifies how much $\tilde{U}$ changes when the objective function changes for a given ρ. In the non-degenerate

case, we expect this term to be bounded in norm by a Lipschitz constant η. The second term, $\frac{\partial f}{\partial \rho}$, evaluates how the objective function changes when the state changes. Recalling that the objective function is a sum of eigenvalues and using non-degenerate perturbation theory, this term is going to be proportional to ε defined by equation (A.1). Thus, instead of $\tilde{U} = \mathrm{e}^{\mathrm{i}\tilde{H}}$, the minimization algorithm returns $\mathrm{e}^{\mathrm{i}(\tilde{H}+\varepsilon\eta A)} \approx W\tilde{U}$ where the anomalous unitary

$$W = \mathrm{e}^{\mathrm{i}\varepsilon\eta A} \tag{A.2}$$

quantifies the perturbation to the perfect disentangler due to the presence of error. Note that A is a Hermitian operator of norm of order 1.

A.2. Impact of the imperfect disentangler on error propagation

At step $k+1$, the anomalous unitary acts on the state of equation (10)

$$W_{k+1}\frac{|0\rangle^{\otimes k+1}|\eta_{k+1}\rangle + |e^{\mathrm{ac}}_{k+1}\rangle}{\sqrt{1+\epsilon^{\mathrm{ac}}_{k+1}}}, \tag{A.3}$$

where $\epsilon^{\mathrm{ac}}_{k+1} \equiv \langle e^{\mathrm{ac}}_{k+1}|e^{\mathrm{ac}}_{k+1}\rangle$ is the error resulting from all previous steps but assuming the disentangler used at step $k+1$ is the exact one. Up to an overall phase, we can rewrite the state (A.3) as

$$\frac{|0\rangle^{\otimes k+1}|\eta_{k+1}\rangle + |E^{\mathrm{ac}}_{k+1}\rangle}{\sqrt{1+E^{\mathrm{ac}}_{k+1}}}, \tag{A.4}$$

where the error vector $|E^{\mathrm{ac}}_{k+1}\rangle$ (whose square norm is E^{ac}_{k+1}) now takes into account the imperfect disentangler.

Comparing equation (A.3) and (A.4), we see that error E^{ac}_{k+1} relates to the error $\epsilon^{\mathrm{ac}}_{k+1}$ through

$$1 + E^{\mathrm{ac}}_{k+1} = 1 + \epsilon^{\mathrm{ac}}_{k+1}/\beta^2, \tag{A.5}$$

where $\beta = \left|\langle\eta_{k+1}|\langle 0|^{\otimes k+1} W |0\rangle^{\otimes k+1}|\eta_{k+1}\rangle\right|$.

A.3. Error propagation without post-selection

Using $W = \mathrm{e}^{\mathrm{i}\varepsilon\eta A}$, calculations show that

$$\beta^2 = 1 - \varepsilon^2\eta^2(\langle A^2\rangle - \langle A\rangle^2) = 1 - \varepsilon^2\eta^2\Delta^2 \tag{A.6}$$

plus $O(\varepsilon^4\eta^4)$ terms, where the variance Δ^2 of A with respect to state $|0\rangle^{\otimes k+1}|\eta_{k+1}\rangle$ appears. Recalling that $\varepsilon = \frac{E^{\mathrm{ac}}_k}{1+E^{\mathrm{ac}}_k}$, we can bound β^2 by

$$\beta^2 = \frac{(1+E^{\mathrm{ac}}_k)^2 - (E^{\mathrm{ac}}_k)^2\eta^2\Delta^2}{(1+E^{\mathrm{ac}}_k)^2} \geqslant \frac{1}{(1+E^{\mathrm{ac}}_k)^2}, \tag{A.7}$$

for any E^{ac}_k if $\eta^2\Delta^2 \leqslant 1$ or for small E^{ac}_k otherwise.

The recurrence relation for errors given by equation (12) in the case with post-selection now relates the accumulated error $\epsilon^{\mathrm{ac}}_{k+1}$ to the previous accumulated error E^{ac}_k

$$1 + \epsilon^{\mathrm{ac}}_{k+1} = (1+\epsilon_{k+1})(1+E^{\mathrm{ac}}_k). \tag{A.8}$$

Combining equation (A.8) and the bound on β given by equation (A.7), equation (A.5) becomes

$$1 + E^{\mathrm{ac}}_{k+1} \leqslant \left(1+E^{\mathrm{ac}}_k\right)^3 (1+\epsilon_{k+1}), \tag{A.9}$$

which indicates how errors proliferate when post-selection is not possible.

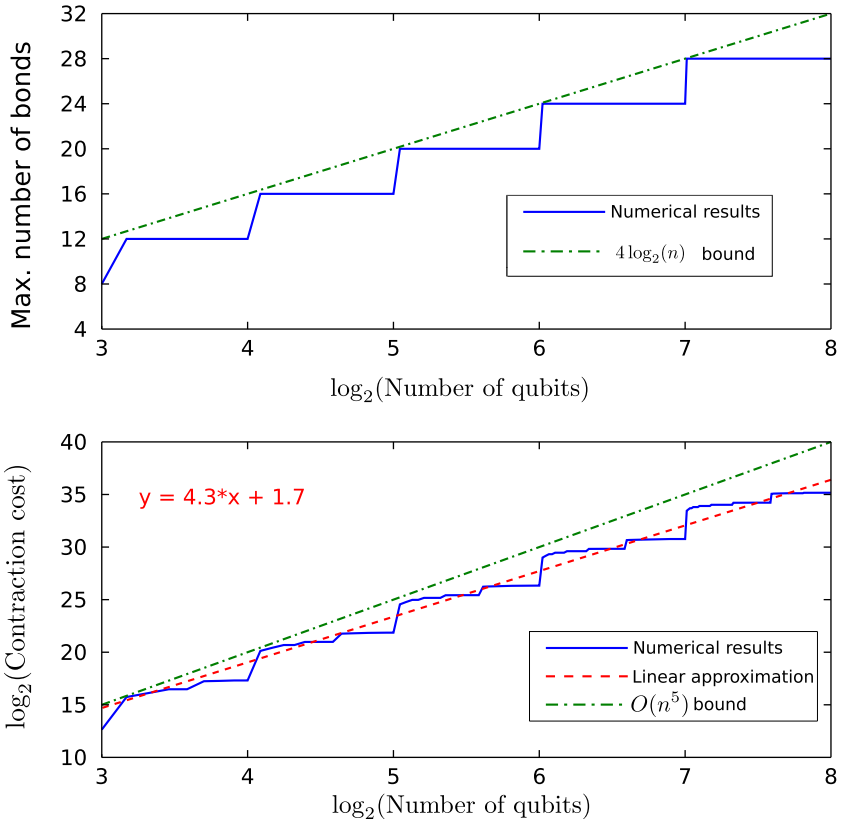

Figure B.1. (top) Maximum number of bonds during the contraction procedure as a function of the logarithm of the number of qubits n. Numerical results (solid blue line) are consistent with the expected bound of $4\log_2 n$. (bottom) Contraction cost C as a function of the number n of qubits on a log scale. Numerical results (solid blue line) are consistent with the $O(n^5)$ bound (dot dashed green line) but linear approximation (dashed red line) indicate that the cost scales like a smaller power of n, namely $C \simeq n^{4.3}$.

Appendix B. Comparing a reconstructed MERA to a predicted MERA

In this section, we describe a polynomial algorithm to contract two MERA states, thus allowing to compute their fidelity. This algorithm is of practical interest for comparing a MERA whose parameters have been identified experimentally using our method to a predicted MERA state—found by numerical optimization for instance. Notice that contracting two different MERA states also allows to compute expectation values of tensor product of local observables $\bigotimes_i A_i$ since it suffices to contract the original state $|\psi\rangle$ and the modified state $|\phi\rangle = \bigotimes_i A_i|\psi\rangle$, which is also a MERA state.

Defining a method to contract two MERA states is equivalent to giving a prescription on how to sequentially contract the tensor network resulting to joining two MERA states. Recall that contracting two tensors $(M)_{i_a j_b}$ and $(N)_{k_b \ell_c}$ to obtain $T_{i_a \ell_c} = \sum_{j_b} M_{i_a j_b} N_{j_b \ell_c}$ has a computational cost of $a \times b \times c$ where a is the number of values that the index i_a can take b and c are defined in the same way with respect to j_b and ℓ_c. In a tensor network, every tensor is usually represented with a number of bonds that each represent an index that has the same

maximal number of possible values. For a MERA, this maximal bond dimension is usually denoted by χ.

The main idea to contract efficiently two MERA states is essentially to turn them into two MPS before contracting them. We look at the MERA circuit as having $n/2$ columns of gates vertically and $\log_\chi n - 1$ renormalization layers horizontally. The sequence of contraction is to sequentially contract every tensor in the leftmost column to create a tensor with a large number of bonds that will then contract with every tensor in the next column. The maximal number of bonds that this leftmost tensor will have throughout the contraction of the network is given by the maximal number of bonds that are opened when taking a vertical cut in the tensor network. For a single MERA, cutting through each of the $\log_\chi n - 1$ layer opens up two bonds, one for the rightmost incoming edge of the isometry and one for the outgoing edge of the isometry. Thus, for the contraction of two MERAs, the maximum number of bonds for a vertical cut is bounded by $\max \# = 2 \times 2 \times \log_\chi n = 4 \log_\chi n$, which is verified numerically (see top of figure B.1). Since at every contraction step, the leftmost tensor with a large number of bonds contract with another tensor that has at most two bonds in addition to the ones being contracted, the maximum cost of one contraction is $\chi^{\max \#} \chi^2 = \chi^2 n^4$. Finally, there are $O(n)$ disentanglers and isometries to contract so the total cost of contracting the network is bounded by $O(n^5)$. Actual numerical simulations show that this bound is probably not tight (see bottom of figure B.1).

References

[1] Vogel K and Risken H 1989 *Phys. Rev.* A **40** 2847
[2] Das Sarma S, Gervais G and Zhou X 2010 *Phys. Rev.* B **82** 115330
[3] Häffner H *et al* 2005 *Nature* **438** 643
[4] Barreiro J T, Schindler P, Gühne O, Monz T, Chwalla M, Roos C F, Hennrich M and Blatt R 2010 *Nature Phys.* **6** 943–6
[5] DiCarlo L *et al* 2009 *Nature* **460** 240
[6] DiCarlo L, Reed M D, Sun L, Johnson B R, Chow J M, Gambetta J M, Frunzio L, Girvin S M, Devoret M H and Schoelkopf R J 2010 *Nature* **467** 574–8
[7] Filipp S *et al* 2009 *Phys. Rev. Lett.* **102** 1
[8] Mikami H, Li Y, Fukuoka K and Kobayashi T 2005 *Phys. Rev. Lett.* **95** 150404
[9] Resch K, Walther P and Zeilinger A 2005 *Phys. Rev. Lett.* **94** 070402
[10] Affleck I, Kennedy T, Lieb E H and Tasaki H 1987 *Phys. Rev. Lett.* **59** 799
[11] Fannes M, Nachtergaele B and Werner R F 1992 *Commun. Math. Phys.* **144** 443
[12] Vidal G 2003 *Phys. Rev. Lett.* **91** 147902
[13] Vidal G 2004 *Phys. Rev. Lett.* **93** 040502
[14] White S R 1992 *Phys. Rev. Lett.* **69** 2863
[15] Schollwöck U 2005 *Rev. Mod. Phys.* **77** 259
[16] Verstraete F and Cirac J 2006 *Phys. Rev.* B **73** 094423
[17] Cramer M, Plenio M, Flammia S, Somma R, Gross D, Bartlett S, Landon-Cardinal O, Poulin D and Liu Y 2010 *Nature Commun.* **1** 149
[18] Vidal G 2008 *Phys. Rev. Lett.* **101** 110501
[19] Evenbly G and Vidal G 2009 *Phys. Rev.* B **79** 144108
[20] Evenbly G, Pfeifer R N C, Picó V, Iblisdir S, Tagliacozzo L, McCulloch I P and Vidal G 2010 *Phys. Rev.* B **82** 161107
[21] Montangero S, Rizzi M, Giovannetti V and Fazio R 2009 *Phys. Rev.* B **80** 113103
[22] Pfeifer R N C, Evenbly G and Vidal G 2009 *Phys. Rev.* A **79** 040301
[23] Cincio L, Dziarmaga J and Rams M 2008 *Phys. Rev. Lett.* **100** 2

[24] Aguado M and Vidal G 2008 *Phys. Rev. Lett.* **100** 070404
[25] König R and Bilgin E 2010 *Phys. Rev.* B **82** 125118
[26] Pfeifer R, Corboz P, Buerschaper O, Aguado M, Troyer M and Vidal G 2010 *Phys. Rev.* B **82** 115126
[27] Vidal G 2008 arXiv:0707.1454v2
[28] Blume-Kohout R 2010 *New J. Phys.* **12** 043034
[29] da Silva M P, Landon-Cardinal O and Poulin D 2011 *Phys. Rev. Lett.* **107** 210404
[30] Verstraete F, Murg V and Cirac J 2008 *Adv. Phys.* **57** 143
[31] Schuch N, Wolf M M, Verstraete F and Cirac J I 2007 *Phys. Rev. Lett.* **98** 140506
[32] Verstraete F and Cirac J 2004 arXiv:cond-mat/0407066
[33] Barthel T, Kliesch M and Eisert J 2010 *Phys. Rev. Lett.* **105** 010502