A class of quantum many-body states that can be efficiently simulated

G. Vidal

1School of Physical Sciences, the University of Queensland, QLD 4072, Australia
(Dated: February 1, 2008)

We introduce the multi-scale entanglement renormalization ansatz (MERA), an efficient representation of certain quantum many-body states on a D-dimensional lattice. Equivalent to a quantum circuit with logarithmic depth and distinctive causal structure, the MERA allows for an exact evaluation of local expectation values. It is also the structure underlying entanglement renormalization, a coarse-graining scheme for quantum systems on a lattice that is focused on preserving entanglement.

PACS numbers:

A better understanding of quantum entanglement has enabled significant progress in the numerical simulation of quantum many-body systems over the last few years. 1, 2, 3, 4. Building on the density matrix renormalization group (DMRG) 2—a well-established technique for one-dimensional (1D) systems on a lattice—, new insight from quantum information science has led, e.g., to efficient algorithms to simulate time-evolution 1 and address 2D systems 2.

A key ingredient of such algorithms is the use of a network of tensors to efficiently represent quantum many-body states. Examples of tensor networks include matrix product states (MPS) 2 for 1D systems, tree tensor networks (TTN) 2 for systems with tree shape, and projected entangled-pair states (PEPS) 2 for 2D systems and beyond, the three structures differing in the way that defines how the tensors are interconnected into a network: the graphs for MPS, TTN and 2D PEPS are, respectively, a chain, a tree and a 2D lattice. Importantly, from these tensor networks the expectation value of local observables can be computed efficiently. But whereas from a MPS and a TTN such calculations are exact, from a PEPS—which has a much wider range of applications—local expectation values can only be obtained efficiently after a number of approximations.

In this manuscript we present the multi-scale entanglement renormalization ansatz (MERA), a structure that efficiently encodes quantum many-body states of D-dimensional lattice systems and from which local expectation values can be computed exactly. A MERA consists of a network of isometric tensors in D+1 dimensions, where the extra dimension can be interpreted in two alternative ways: either as the time of a peculiar class of quantum computations, or as parameterizing different length scales in the system, according to successive applications of a lattice coarse-graining procedure known as entanglement renormalization 2. Focused on preserving entanglement, MERA is a promising candidate to describe emergent quantum phenomena, including quantum phase transitions, quasi-particle excitations and topological order. Here we establish its connection to entanglement renormalization and explore some of its basic properties: efficient contractibility of the network, leading to an efficient evaluation of local expectation values; inherent support of algebraically decaying correlations and of an area law for entanglement; versatility to adapt to both the local and global structure of the system’s underlying lattice; and ability to naturally assimilate symmetries such as invariance under translations or rescaling – which result in substantial gains in computational efficiency.

Let us consider a square lattice L in D spatial dimensions consisting of N sites, where each site s ∈ L is described by a complex vector space V of finite dimension χ. Let \(|\Psi\rangle \in V^\otimes N\) denote a pure state of lattice L and \(\rho^{[s]} = \text{tr}_L(\langle \Psi | \Psi \rangle)|\langle \Psi | \Psi \rangle\rangle\) its reduced density matrix for site s. We study states |\Psi\rangle that can be generated by means of a certain quantum circuit \(C\) of depth \(\Theta \equiv 2 \log_2(N) - 1\), see Fig. 1 for the \(D = 1\) case. Each site in \(L\) corresponds to one outgoing wire of \(C\), and we label both with the same index s.

The causal cone \(C^{[s]}\) of outgoing wire s, defined as the set of gates and wires that may influence state \(\rho^{[s]}\), plays an important role in our discussion. The key feature of circuit \(C\) is that each causal cone \(C^{[s]}\) has bounded width, that is, the number of wires in a time slice \(C^{[s]}_\theta\) of \(C^{[s]}\) is
two properties: (i) the causal cone of any outgoing wire in Fig. (1) involves at most four wires at any time \( \theta \), as highlighted in Fig. 2. More generally, the causal cone of outgoing wires in the \( D \)-dimensional case can be seen to comprise at most \( 3^D-4 \) wires at a given time \( \theta \), see Fig. 3.

A MERA for \( |\Psi\rangle \) is a tensor network \( \mathcal{M} \) that corresponds to quantum circuit \( \mathcal{C} \) except for some minor geometric changes. Each two-body unitary gate \( u \) of \( \mathcal{C} \) gives rise to a tensor of \( \mathcal{M} \). But incoming wires in state \( |0\rangle \) are eliminated in the tensor network, producing three kinds of tensors: (i) the top tensor \( u(0)|0\rangle \) of \( \mathcal{M} \) has two indices and is normalized to 1,

\[
(t)_{\mu\nu} \equiv (u)_{\alpha\beta}^{\mu\nu}|\alpha,\beta=0, \sum_{\mu\nu}(t^*)_{\mu\nu}(t)_{\mu\nu} = 1; \tag{1}
\]

(ii) tensors in every second row are isometries \( u = |0\rangle \),

\[
(w)_{\mu\nu}^\alpha \equiv (u)_{\alpha\beta}^{\mu\nu}|\beta=0, \sum_{\mu\nu}(w^*)_{\mu\nu}(w)_{\mu\nu} = \delta_{\alpha\alpha'}; \tag{2}
\]

(iii) the rest of tensors in \( \mathcal{M} \) are unitary gates \( u \),

\[
\sum_{\mu\nu}(u^*)_{\alpha\beta}^{\mu\nu}(u)_{\alpha'\beta'}^{\mu\nu} = \delta_{\alpha\alpha'}\delta_{\beta\beta'}, \tag{3}
\]

\[
\sum_{\alpha\beta}(u^*)_{\mu\nu}^{\alpha\beta}(u)_{\alpha\beta'}^{\mu\nu} = \delta_{\mu\mu'}\delta_{\nu\nu'}, \tag{4}
\]

that we call disentanglers for reasons that will become clear later. Notice that the computational space required to store \( \mathcal{M} \) grows as \( O(\chi^4N) \), that is, linearly in \( N \), given that there are \( 2N-1 \) tensors and each tensor depends on at most \( \chi^4 \) parameters.

Thus a MERA is an efficient representation of \( |\Psi\rangle \) consisting of a tensor network \( \mathcal{M} \) in \( D+1 \) dimensions with two properties: (i) tensors are constrained by Eqs. 1, 2, or 3-4; (ii) each open wire \( s \), associated to one site of the underlying lattice \( \mathcal{L} \), has a causal cone \( C^{|s|} \) with bounded width. As a consequence of this peculiar causal structure, the reduced density matrix of a small number of lattice sites can be computed exactly with remarkably small cost. In what follows \( p_1 \) and \( p_2 \) are integers that depend on the spatial dimension \( D \) of \( \mathcal{L} \).

**Lemma 1:** The one-site density matrix \( \rho^{|s|} \) can be computed from a MERA with time \( O(\chi^{p_1} \log N) \).

**Proof:** For each time slice \( C^{|s|}_\theta \) of the causal cone \( C^{|s|} \), we compute its reduced density matrix \( \sigma_\theta \). As shown in Fig. 4 for \( D = 1 \), \( \sigma_{\theta+1} \) can be obtained from \( \sigma_\theta \) with a cost polynomial in \( \chi \) and independent of \( N \). Recall, finally, that \( \rho^{|s|} = \sigma_\theta \) and that \( \Theta = 2 \log_2(N) - 1 \).

**Lemma 2:** The two-site density matrix \( \rho^{|s_1,s_2|} \) can be computed with \( O(\chi^{p_2} \log N) \) time.

**Proof:** Again, the causal cone \( C^{|s_1,s_2|} \) has logarithmic depth and a width independent of \( N \), see Fig. 5, and the reduced density matrix \( \sigma_{\theta+1} \) for time slice \( C^{|s_1,s_2|}_\theta \) can be obtained from \( \sigma_\theta \) for \( C^{|s_1,s_2|}_\theta \) at bounded cost.

The above lemmas can be easily extended to a k-site reduced density matrix \( \rho^{|s_1,...,s_k|} \) (with computational time scaling exponentially in \( k \)) and they imply that the expectation value of local observables, such as two-site correlators

\[
C_2(s_1, s_2) \equiv \langle \Psi | A^{[s_1]} B^{[s_2]} | \Psi \rangle = \text{tr}(\rho^{[s_1,s_2]} A^{[s_1]} B^{[s_2]}), \tag{5}
\]

can be computed efficiently.

We have defined a MERA for \( |\Psi\rangle \) in terms of a quantum circuit \( \mathcal{C} \) that transforms a product state \( |0\rangle^\otimes N \) into \( |\Psi\rangle \) by means of \( \Theta \) layers of unitary gates. Let \( |\Psi_\tau\rangle \) denote the non-trivial part of the state of circuit \( \mathcal{C} \) at time \( \theta = \Theta - 2\tau \), with \( |\Psi_0\rangle \equiv |\Psi\rangle \). An alternative interpretation of the MERA can be obtained by considering the sequence of states \( \{|\Psi_0\rangle, |\Psi_1\rangle, |\Psi_2\rangle, \ldots\} \), which correspond to undoing the quantum evolution of \( \mathcal{C} \) back in time. Notice that \( |\Psi_{\tau+1}\rangle \) is obtained from \( |\Psi_\tau\rangle \) by applying two layers of tensors in \( \mathcal{M} \). The first layer is made of disentanglers that transform \( |\Psi_\tau\rangle \) into a less entangled state \( |\Psi'_\tau\rangle \). The second layer is made of isometries that...
combine pairs of nearest neighbor wires into single wires, turning the state $|\Psi_0\rangle$ of $N_s$ wires into the state $|\Psi_{s+1}\rangle$ of $N_s/2$ wires, where $N_s = 2^{\Theta - \tau}$. That is, $\cal M$ implements a class of real space coarse-graining transformations known as entanglement renormalization.

More generally, we can use the MERA to transform the sites of lattice $L_0 \equiv L$, as well as operators defined on $L$ such as a Hamiltonian $H_0$, so as to obtain a sequence of increasingly coarse-grained lattices $\{L_0, L_1, L_2, \ldots \}$ and corresponding effective Hamiltonians $\{H_0, H_1, H_2, \ldots \}$. Since an operator defined on site $s \in L$ is mapped into an operator contained in the causal cone $C[s]$, local operators in $L$ remain local when mapped into $L_s$, $1 \leq s \leq \log_2(N) - 1$. Notice that one site in $L_s$ is obtained from a D-dimensional hypercube with $O(2^s)$ sites in $L$. Thus $|\Psi_\tau\rangle$ can be understood as retaining the structure of $|\Psi\rangle$ at length scales greater than $2^\tau / D$.

In Fig. 3, it has been shown that a MERA can encode, in a markedly more efficient way than a MPS, accurate approximations to the ground state of a quantum critical system in a 1D lattice. There correlators $C_2(s_1, s_2)$ decay as a power law with the distance $r$ between sites $s_1$ and $s_2$ and the entanglement entropy $S(\rho_{[B_1]})$ of a block $B_1$ of $l$ adjacent sites scales as $\log l$. Next we give an intuitive justification why MERA supports algebraic decay of correlations in any dimension $D$. We also show that a MERA can carry an amount of block entanglement compatible with logarithmic scaling in 1D and with a boundary law $l^{D-1}$ for $D > 1$.

It follows from the $\lambda$-shaped causal cone $C^{[s_1, s_2]}$ that we can compute $\rho^{[s_1, s_2]}$ from the density matrix $\rho_\tau$ corresponding to a hypercube of (at most) $4^D$ sites of $L_\tau$, where $\tau \approx D \log_2 r$, by means of a sequence of density matrices $\{\rho_1, \ldots, \rho_s, \rho_1, \rho^{[s_1, s_2]}\}$, see Fig. 4. That is, $\rho^{[s_1, s_2]}$ is obtained from $\rho_\tau$ after $O(\log r)$ transformations.

If, as it can be argued in a (scale invariant) critical ground state, each of these transformations reduces correlations by a constant factor $z < 1$, we readily obtain a power law scaling for $C_2(s_1, s_2)$,

$$C_2(s_1, s_2) \approx \tau \log r = r^{-q}, \quad q = \log \frac{1}{z}.$$  \hfill (6)

This property is enabled by the fact that two sites at a distance $r$ in $L$ are connected through a path of length $O(\log r)$ in $\mathcal{M}$, and it indicates that a MERA is particularly suited to describe states with quasi-long-range order, such as critical ground states.

We now consider the entropy $S(\rho_{[B_1]})$ for the density matrix $\rho^{[B_1]}$ of a hypercube $B_1$ made of $l^D$ sites. The causal cone of $B_1$ shrinks exponentially fast with $\tau$ and we see that, once more, we can compute $\rho_{[B_1]}$ from the density matrix $\eta_\tau$ for a hypercube made of (at most) $4^D$ sites of lattice $L_\tau$, for $\tau \approx D \log_2 l$, through a sequence of density matrices $\{\eta_1, \ldots, \eta_\tau, \rho^{[B_1]}\}$, see Fig. 5. Density matrix $\eta_\tau$ is obtained from $\eta_{\tau+1}$ by: (i) applying a layer of isometries and a layer of disentanglers that do not change the entropy of $\eta_{\tau+1}$; (ii) tracing out $n_\tau = O(2(\tau-\tau)(D-1)/D)$ boundary sites. Tracing out one boundary site increases the entropy by at most $\log_2 \chi$ bits, the total increase in entropy $\Delta S_\tau$ being at most $n_\tau \log_2 (\chi)$ bits. Thus the entropy of $\rho^{[B_1]}$ fulfills

$$S(\rho^{[B_1]}) - S(\eta_\tau) = \sum_{\tau=1}^{\bar{\tau}} \Delta S_\tau \leq \log_2(\chi) \sum_{\tau=1}^{\bar{\tau}} n_\tau,$$  \hfill (7)

where $S(\eta_\tau)$ is at most $4^D \log_2 \chi$.

In a 1D MERA the number $n_\tau$ of sites that are traced out is bounded by a constant $c$ [i.e., a hypercube has only two boundary sites] for any $\tau$, $1 \leq \tau \leq \bar{\tau} = O(\log l)$, and

$$S(\rho_{[1]}) - S(\eta_\tau) \leq \log_2(\chi) \bar{\tau} = O(\log l).$$  \hfill (8)

As numerically confirmed in [viEr,evER], in a MERA for critical 1D ground states $\Delta S_\tau$ is independent of $\tau$, leading
to the logarithmic scaling $S(\rho^{[B_l]}) \approx \log(l)$, whereas for a non-critical ground state $\Delta S_l$ vanishes for $\tau \gg \log_2 l$, where $\xi$ is the correlation length, so that $S(\rho^{[B_l]})$ saturates for $l \gg \xi$.

In $D > 1$, instead, the $n_\tau$ decays exponentially with $\tau$, the upper bound for the $S(\rho^{[B_l]})$ being dominated by the contribution from small $\tau$,

$$S(\rho^{[B_l]}) - S(\eta_\tau) \leq \log_2(\chi) \sum_{\tau=1}^{\tilde{\tau}} 2^{(\tilde{\tau} - \tau) \frac{D}{D-1}} \approx \log_2(\chi) \cdot 2^{D \log_2(l) \frac{D}{D-1}} = \log_2(\chi) l^{D-1}.$$ (9)

That is, a MERA in $D > 1$ supports block entanglement that scales at most according to a boundary law $S^{[B_l]} \approx l^{D-1}$.

For the sake of concreteness, we have analyzed the case where $\mathcal{L}$ is a square lattice. The structure of the MERA, however, can be adapted to a more generic lattice, with arbitrary local, geometric and topological properties, while preserving its distinctive causal structure. For instance, in $D = 2$ dimensions a specific MERA can be built to represent states of a triangular lattice; or of a lattice with random vacancies or linear or bulk defects; or to account for a variety of boundary conditions (e.g., plane, cylinder, sphere or torus). In addition, the number of levels $\chi$ can vary throughout $\mathcal{M}$. Adjusting the MERA to the specifics of a problem often leads to computational gains.

In particular, the symmetries of $|\Psi\rangle$ can be assimilated into the MERA. An internal symmetry, such as $SU(2)$ invariance, results in a series of constraints for the tensor in $\mathcal{M}$, which then depend on less parameters. More substantial gains are obtained when $|\Psi\rangle$ is invariant under translations [that is, cyclic shifts by one lattice site in a system with periodic boundary conditions], since all the tensors in a layer of $\mathcal{M}$ can be chosen to be the same and the MERA depends only on $\chi^4 \log N$ parameters. But the most dramatic savings occur for states that are invariant under entanglement renormalization transformations, even in an infinite lattice. Here all tensors in $\mathcal{M}$ are the same and the MERA depends just on $O(\chi^4)$ parameters. Scale invariant critical ground states can be shown to belong to this class.

We conclude with a few pointers to future work. On the one hand, most techniques to simulate quantum systems with a MPS can be generalized to a MERA. This include algorithms to compute the ground and thermal states, and to simulate time evolution. For instance, in order to update $\mathcal{M}$ after the action of a unitary gate $U^{[s_1,s_2]}$ on sites $s_1$ and $s_2$, only the tensors in the causal cone $\mathcal{C}^{[s_1,s_2]}$ need to be modified, with a cost logarithmic in $N$.

On the other hand, the potential of a MERA is not restricted to the representation of individual states. Notice that by feeding the incoming wires, labeled by $r$, of quantum circuit $\mathcal{C}$ with state $\otimes_{r=1}^{N} |\phi^r_0\rangle$ instead of $\otimes_{r=1}^{N} |\phi^r\rangle$, we can generate an infinite family of entangled states $|\Psi_{\{\phi_r\}}\rangle$, all represented by a MERA that only differs in the isometric tensors and such that, given the unitarity of $\mathcal{C}$, fulfill

$$\langle \Psi_{\{\phi_r\}} | \Psi_{\{\phi'_r\}} \rangle = \prod_r \langle \phi^r_r | \phi^r_r \rangle.$$ (10)

This can be used to encode, in just one (generalized) MERA, not only the ground state of a lattice system but also its quasi-particle excitations. This is also used, in systems with topological quantum order, to store topological information in the top tensor of a MERA.

The author acknowledges support from the Australian Research Council through a Federation Fellowship.

[1] G. Vidal, Phys. Rev. Lett. 91, 147902 (2003), quant-ph/0301063.
[2] G. Vidal, Phys. Rev. Lett. 93, 040502 (2004), quant-ph/0311080.
[3] S. R. White and A. E. Feiguin, Phys. Rev. Lett. 93, 076401 (2004).
[4] A. J. Daley et al., J. Stat. Mech.: Theor. Exp. (2004) P04005.
[5] F. Verstraete and J. I. Cirac, cond-mat/0407066.
[6] F. Verstraete, M. M. Wolf, D. Perez-Garcia, J. I. Cirac, quant-ph/0601075.
[7] Y.Y. Shi, L.M. Duan and G. Vidal, Phys. Rev. A 74, 022320 (2006), quant-ph/0511070.
[8] For see instance, F. Verstraete, J. J. Garcia-Ripoll, J. I. Cirac, Phys. Rev. Lett. 93, 207204 (2004), cond-mat/0406426.
[9] M. Zwolak and G. Vidal, Phys. Rev. Lett. 93, 207205 (2004), cond-mat/0406440.
[10] B. Paredes, F. Verstraete, J. I. Cirac, cond-mat/0505288.
[11] F. Verstraete, D. Porras, J. I. Cirac Phys. Rev. Lett. 97, 227205 (2004); D. Porras, F. Verstraete, J. I. Cirac, cond-mat/0504717.
[12] T. Osborne, quant-ph/0601019.
[13] Ibid, quant-ph/0603137.
[14] Ibid, cond-mat/0605194.
[15] S. R. White, Phys. Rev. Lett. 69, 2863 (1992), Phys. Rev. B 48, 10345 (1993).
[16] U. Schollwoeck, Rev. Mod. Phys. 77, 259 (2005), cond-mat/0409292.
[17] M. Fannes, B. Nachtergaele and R. F. Werner, Comm. Math. Phys. 144, 3 (1992), pp. 443-490.
[18] S. Ostlund and S. Rommer, Phys. Rev. Lett. 75, 19 (1995), pp. 3537.
[19] G. Vidal, cond-mat/0512165.
[20] In an MPS, correlations between left and right halves of a chain are accounted for through $\chi_{MPS}$ terms in the Schmidt decomposition. A MERA can account for about $2\chi \log_2 N$ terms in the Schmidt decomposition. Thus, it takes an MPS with $\chi_{MPS} \approx 2\chi \log_2 N$ to represent a state stored with a MERA of $\chi$-level wires.
[21] G. Vidal, J.I. Latorre, E. Rico, A. Kitaev, Phys. Rev. Lett. 90 (2003) 227902, quant-ph/0211074.
[22] P. Calabrese and J. Cardy, J.Stat.Mech. (2004) P002, hep-th/0405152.
[23] A. R. Its, B.-Q. Jin and V. E. Korepin, J. Phys. A: Math. Gen. vol 38, pages 2975-2990, 2005, cond-mat/0409027.
[24] G. Evenbly et al., in preparation.
[25] S. Singh et al., in preparation.
[12] M. Aguado et al., *in preparation*. 