Multiscale Entanglement Renormalisation Ansatz: a review of its formulation and function

Monu Kumar, 2011090, *School of Physical Sciences, NISER.*

ABSTRACT

This study presents a comprehensive overview of MERA, its motivations and origins, its mathematical formulation and its applications. The key concepts and motivations behind MERA are briefly introduced, along with the fundamentals of entanglement in quantum mechanics, numerical renormalisation techniques, and tensor networks. The ansatz is explicitly described, followed by implementation of MERA in simulation of 1D Ising model in the presence of transverse magnetic field. Results obtained with MERA illustrate the method's efficacy and highlight its strengths and limitations. Furthermore, parallels are drawn between MERA's geometric structure and key features of holographic duals; this underscores the method's potential in the study of strongly interacting field theories.

INTRODUCTION

Several quantum many-body systems exhibit complex collective behaviors and exotic phenomena, leading to fundamental questions in physics and potential technological applications. Especially, when extended systems are considered, the high number of interacting degrees of freedom demonstrate strong quantum collective behaviour. Theoretical work in this field is challenging due to these strong interactions and the exponential growth of state space with increasing degrees of freedom. Numerical tools like quantum Monte Carlo and density matrix renormalisation group (DMRG) have been developed to tackle these problems. The sign problem limits the applicability of quantum Monte Carlo in some systems of notable interest, which is otherwise widely successful.

Numerical renormalisation group techniques form another successful class of methods, based on Wilson's theory of real-space renormalisation (physics studied at different length scales). One such method, DMRG, can be used to simulate low-temperature behaviour of strongly interacting systems with feasible computational resources, and requires almost no apriori knowledge of the model studied. DMRG has proven extremely effective for non-critical 1D systems, but it too has met its limits, with higher dimensions and critical points of phase transitions causing problems.

The multiscale entanglement renormalisation ansatz (MERA) is a new method designed for simulating critical points in quantum systems, drawing inspiration from DMRG's mathematical formulation in terms of tensor network states and offering a promising alternative to traditional numerical approaches.

MERA is designed as a technique used for simulating critical points in quantum systems that generalises to higher dimensions. Unlike Monte Carlo methods, MERA does not face the sign problem, making it a viable option in certain cases. MERA is also linked to holography and holographic dualities, connecting quantum gravity theories to gauge quantum field theories. The method offers a concrete realization of the AdS/CFT duality, with potential implications for better understanding these dualities and developing theories of quantum gravity.

ENTANGLEMENT ENTROPY

Quantum entanglement is the phenomenon of a group of particles being generated, interacting, or sharing spatial proximity in such a way that the quantum state of each particle of the group cannot be described independently of the state of the others, including when the particles are separated by a large distance. This can be generalized to two (or more) subsystems (of a

larger encompassing system) which are entangled with each other.

Consider state spaces (subsystems) A and B such they form a larger space $A \otimes B$. Let the corresponding sets of eigenstates be $\{|\psi_i\rangle\} \in A$ and $\{|\phi_i\rangle\} \in B$ respectively. Then,

$$\sum_{i,j} a_{ij} |\psi_i\rangle |\phi_j\rangle \in A \otimes B$$

However, it may be that $\nexists\ |\psi\rangle \in A$ and $|\phi\rangle \in B$ such that,

$$\sum_{i,j} a_{ij} |\psi_i\rangle |\phi_j\rangle = |\psi\rangle \otimes |\phi\rangle$$

Then, the linear combination of states is an entangled state. For example, the qubit state (one of the Bell states):

$$|\phi^+\rangle = \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle)$$

More generally, entanglement captures the correlations between systems that are purely quantum mechanical in nature.

A general density matrix is given by,

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$$

where p_i are the classical probabilities associated to states $|\psi_i\rangle$.

For the total density matrix ρ associated with system $A \otimes B$, we can write,

$$\rho = \rho_A \otimes \rho_B$$

Here,

$$\rho_A = Tr_B(\rho)$$

encodes all information we can have about A, but it is limited by our lack of knowledge of B. Thus, even if ρ is pure, ρ_A may be mixed. The expectation value of an operator O acting on $A \otimes B$ is given by,

$$\langle O \rangle = Tr(\rho O)$$

The von Neuman entropy of the system (quantum mechanical equivalent of Shannon entropy) is defined by,

$$S(\rho) = -Tr(\rho \log(\rho)) = -\sum_{i=1}^{N} p_i \log(p_i) \geq 0$$

where, p_i are the eigenvalues of ρ ($0 \log(0)$ is taken to be zero by limit).

Von Neuman entropy is a measure of mixedness of a density matrix. It is 0 for a pure state, and attains the maximum value $S = \log(N)$ for a maximally entangled state ($p_i = 1/N\ \forall i$). Its strong subadditivity is non trivial and significant. Subadditivity states that for the density matrix ρ,

$$S(\rho) \leq S(\rho_A) + S(\rho_B)$$

Also, Araki-Lieb inequality holds so that,

$$|S(\rho_A) - S(\rho_B)| \leq S(\rho) \leq S(\rho_A) + S(\rho_B)$$

Thus, entropy is a measure of lack of information, which increases when separating two entangled subsystems. Thus, we quantify the *entanglement entropy* of a subsystem as the von Neumann entropy of its reduced density matrix. This definition is based on how the larger system is partitioned, since entanglement between the constituent subsystems will be consequently different.

We further define mutual information between the two subsystems as,

$$I(A, B) = S(\rho_A) + S(\rho_A) - S(\rho)$$

i.e., the information loss between the description as a whole and as sum of parts (the information encoding the correlations between the parts). This is characterized by the constrains correlators of operators defined on the two subsystems. If O_A and O_B are operators acting on A and B respectively, then,

$$\frac{(\langle O_A \otimes O_B \rangle - \langle O_A \rangle_A \langle O_B \rangle_B)^2}{2\langle O_A^2 \rangle_A \cdot \langle O_B^2 \rangle_B} \leq I(A, B)$$

Now, consider a physical system and its state, and the entanglement entropy S of a region of the system, such as a part of spacetime or some part of a lattice. An area law would state that this entanglement entropy scales (asymptotically) at most with area of the boundary of the region, so that in a d-dimensional system for a region of linear size l, entropy scales as $S \propto l^{d-1}$. This is in contrast with randomised quantum states, that have entanglement equally between all the degrees of

freedom and entropy proportional to the number of degrees of freedom. We see that many physical states indeed do not obey area laws, but rather let the entanglement entropy scale with the volume. For example, though the ground state in local, gapped Hamiltonians has been found to satisfy an area law in most cases, the rules are broken at critical points of 1-dimensional systems. There, typically a logarithmic violation of the area law is found (Plenio et al, 2010).

REAL-SPACE RENORMALIZATION

Renormalization is a method to investigate the variance of physics as a function of the scale at which we are observing it. The central idea is that the values of the parameters describing the physics, such as masses and couplings, may change as functions of the energy or, complementarily, length scale at which the theory works. More precisely, we can describe the same system at different scales by a common Hamiltonian, if we make the parameters of the Hamiltonian dependent on the scale. This means that as we move from considering the microscopic interactions to treating larger wholes as the constituents of some effective theory, this process can simply be described as a flow in parameter space, the *renormalization group flow*.

It's observed that the phenomenological parameters like observed masses and couplings of the Lagrangian change at different energies, due to for example screening by virtual particles. Thus, one can fix a scale at which we say we are observing the theory and require observables at this scale to have finite values, disregarding the fact that this might cause singularities at the very high energy end of the theory. This allows us to detach these high energy extremities from the physics of the observable scale, enabling us to do effective predictions at the cost of knowledge of processes at extreme scales.

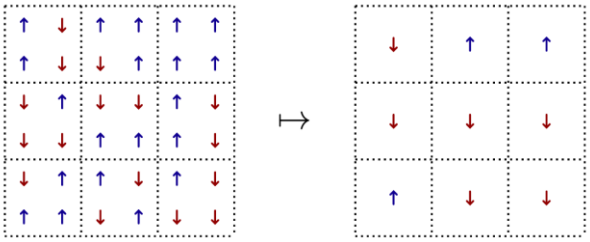

Figure 1: A block-spin transformation in the spirit of Kadanoff. A lattice of spins is approximated by a new lattice with fewer spins. This is done by dividing the original lattice into blocks and describing the state of each block with a single spin, that is taken here to be the mode of the original spins (here down is chosen to win in case of a tie).

TREE RENORMALIZATION

Let us consider a d-dimensional lattice L of quantum systems and a Hamiltonian H that is taken to be a sum of nearest-neighbour terms, that dictates the interactions between these degrees of freedom. This could be for example a lattice of spin-$1/2$ fermions, i.e. qubits, with the Hamiltonian of say the Ising model, or something more intricate like a lattice of interacting harmonic oscillators or atoms.

The goal is to calculate the properties of the ground state of the lattice, such as expectation values of local operators and correlators between points, in a manner that is feasible with modern classical computers. Using exact diagonalisation to obtain the ground state with brute force is not feasible for even the simplest systems if the lattice size is anything but extremely small (state space of the system grows exponentially with the number of lattice sites N). Applying perturbation theory around a free field theory is limited to weak couplings between the degrees of freedom.

So, consider a part of the lattice $B \subset L$. If each lattice point $s \in B$ has the state space V_s, then the state space of B is,

$$V_B = \bigotimes_{s \in B} V_s$$

Consider another lattice L' such that each point $s' \in L'$ corresponds to a region of several points B in the original lattice. However, we demand $dim\, \boldsymbol{V}_{s'} < dim\, \boldsymbol{V}_B$, so that the state on L' only retains some of the characteristic features of the

state and is easier to handle computationally. This is similar to Kadanoff's block transformation.

We define a tensor:

$$w: \boldsymbol{V}_{s'} \to \boldsymbol{V}_B$$

such that its dual $w^\dagger: \boldsymbol{V}_B \to \boldsymbol{V}_{s'}$ maps the state of the block to a state of the new lattice point. Furthermore,

$$w^\dagger w = I_{V_{s'}}$$

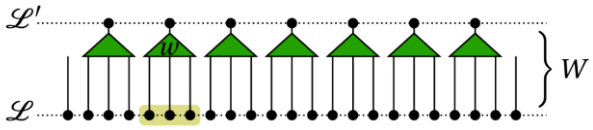

Figure 2: A coarse-graining transformation of a 1D lattice. The black dots mark lattice sites, our choice of the region B is shown in yellow. Each individual w connects a site of the coarser lattice L' to a region in L. Together they form the map W between the two lattices

By dividing the whole lattice L into blocks and constructing an isometry from each of these blocks to L', we create a mapping

$$W: \boldsymbol{V}_{L'} \to \boldsymbol{V}_L; \quad W = \bigotimes_{s' \in L'} w$$

By definition W is also an isometry,

$$W^\dagger W = I_{V_{L'}}$$

For simplicity, we have assumed translational invariance of the system (all the blocks are identical).

Given an operator O, we map it as follows:

$$O \to O' = W^\dagger O W$$

Now, we demand that W preserves ground state expectation values, i.e.,

$$\langle G|O|G\rangle = \langle G'|O'|G'\rangle = \langle G|WW^\dagger OWW^\dagger|G\rangle$$

where, $|G'\rangle = W^\dagger|G\rangle$ and $|G\rangle$ is the ground state of the lattice. Equivalently,

$$WW^\dagger|G\rangle = |G\rangle$$

This is trivially satisfied for a unitary W. However, that is undesirable since it necessitates $\dim \boldsymbol{V}_{L'} = \dim \boldsymbol{V}_L$, which results in no computational advantage. Instead, we require the subspace $V_{L'} \subset V_L$ to be as small as possible, but such that the ground state is preserved.

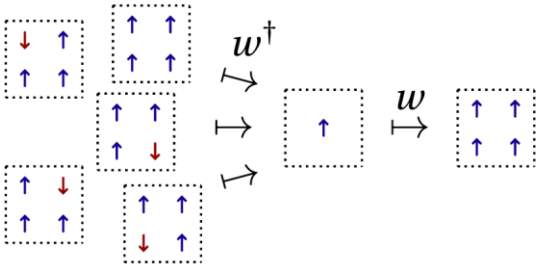

Figure 3: An illustration of how the coarse-graining isometry w would work in Kadanoff's block-spin transformation for blocks that have up as their mode state, and back mapping using $w^\dagger$. Note the non-injectivity of $w^\dagger$ and non-surjectivity of w and how $ww^\dagger$ projects the states on the left to the one state on the right.

Consider the Schmidt decomposition of the ground state:

$$|G\rangle = \sum_{i=1}^{\chi} \sqrt{p_i}\,|\phi_i\rangle|\psi_i\rangle$$

where, $|\phi_i\rangle \in V_B$ for a block and $|\psi_i\rangle$ are the states of the rest of the lattice. Now if we choose w such that

$$ww^\dagger = \sum_{i=1}^{\chi} |\phi_i\rangle\langle\phi_i|$$

then the ground state is preserved since,

$$ww^\dagger|G\rangle = \sum_{i=1}^{\chi} \sqrt{p_i}\,|\phi_i\rangle\langle\phi_i|\phi_j\rangle|\psi_j\rangle = |G\rangle$$

This choice of w is known as White's rule.

The operator $P \equiv ww^\dagger$ is a projection ($P^2 = P$), and it projects to the space spanned by $|\phi_i\rangle$. It leaves the dimensions of V_B that are relevant for the ground state ρ_B intact, but discards the rest of the space. A simple suitable choice of w is

$$w = \sum_{i=1}^{\chi} |\phi_i\rangle|\alpha_i\rangle$$

where $|\alpha_i\rangle$ are a basis for $V_{s'}$. This w trivially fulfills the isometry condition. Also, $\dim(V_{s'}) = \chi \leq \dim(V_B)$, where the inequality is possible because eigenvalues of ρ_B that are zero do not contribute to P and w. Thus, the dimension of the Hilbert space is reduced in the coarse-graining.

The state space of the whole L' is $\bigotimes_{s' \in L} V_{s'}$. It can accommodate the ground state properties of L, but neglects the rest of $\boldsymbol{V}_L$. To disregard contributions from eigenvectors of ρ_B with small eigenvalues in the name of further computational efficiency, we set a cut m such that,

$$1 - \sum_{i=1}^{m} p_i < \epsilon$$

where, it is assumed that the eigenvalues are ordered as $p_1 \geq p_2 \geq \cdots \geq p_\chi$. The upper limit ϵ is a preset truncation error that represents the accuracy of our approximation, and instead we set the approximate rule,

$$ww^\dagger = \sum_{i=1}^{m} |\phi_i\rangle\langle\phi_i|$$

More generally, for a low energy subspace that includes lowest excited states as well, we can make $ww^\dagger$ project to this subspace and otherwise proceed as usual.

Now that we have successfully mapped L to a coarse-grained lattice L', we can continue with this process and coarse-grain further, mapping L' to L'' and that further on to L_3, et cetera. At all stages the ground state is left untouched. At each coarse-graining, a site of the new lattice corresponds to region of sites in the previous one, which again corresponds to a larger region of sites in the one before it. Thus, higher the hierarchy of L_τ, larger a part of the original system is described by a single lattice point, so that the lattices L_τ are approximate descriptions of the system at growing length scales.

The sequence of lattices, states and operators that we get by consecutively coarse-graining is called the *renormalization group flow*.

$$\begin{aligned}
&\mathbb{V}_{\mathscr{L}_0} \xrightarrow{W_0^\dagger} \mathbb{V}_{\mathscr{L}_1} \xrightarrow{W_1^\dagger} \cdots \xrightarrow{W_T^\dagger} \mathbb{V}_{\mathscr{L}_T},\\
&|\mathbb{G}_0\rangle \xrightarrow{W_0^\dagger} |\mathbb{G}_1\rangle \xrightarrow{W_1^\dagger} \cdots \xrightarrow{W_T^\dagger} |\mathbb{G}_T\rangle,\\
&O_0 \xrightarrow{W_0^\dagger} O_1 \xrightarrow{W_1^\dagger} \cdots \xrightarrow{W_T^\dagger} O_T.
\end{aligned}$$

Here, T denotes the level where our coarse-graining flow ends. For an infinite lattice this is when the flow reaches a static state, where subsequent applications of W do not change the system further. In the case of a finite lattice, T is naturally reached when the whole system has shrunk to number of sites smaller than the coarse-graining region. At that point we map the remaining sites with a final top tensor $t : \boldsymbol{V}_{L_T} \to \boldsymbol{V}_{L_{T-1}}$, which is another isometric tensor quite like the w's that have been used, but it coarse-grains the whole remaining lattice L_{T-1} to a single site. We can also just choose to terminate the flow at certain depth, in which case we have several top tensors that are simply the isometries that produce the final, coarsest lattice of the flow.

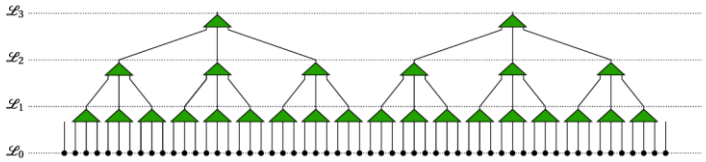

Figure 4: An example of the renormalisation group flow created by tree renormalisation. At each level three neighbouring sites are grouped together and mapped with an isometry to a new lattice point at a coarser lattice

For finite systems, one needs to consider the boundary conditions. Usually periodic boundary conditions are adopted, because they make both numerics and analytical treatment simple, though explicit boundaries are also possible (Evenbly et al, 2010).

At every level of the flow we apply either White's rule or the approximate rule, which ensures that

$$\langle G_0|O_0|G_0\rangle = \langle G_\tau|O_\tau|G_\tau\rangle \;\; \forall\, \tau \in [0,1,\ldots,T]$$

Then, White's rule dictates the Hilbert space dimension χ_τ at every level, which depends on the spectrum of the reduced density matrices on the previous level $\tau - 1$. Now, the computational cost of any observable scales as a growing function of χ_τ.

For on all levels of the renormalisation group flow, χ is related to the amount of entanglement between B and the rest of the lattice. The entanglement entropy is given by,

$$S(\rho_B) = -\sum_{i=1}^{\chi} p_i \log(p_i)$$

For maximum entropy, $(p_i = 1/\chi)$, we get,

$$S(\rho_B) = -\sum_{i=1}^{\chi} \chi^{-1} \log(\chi^{-1}) = \log \chi$$

Therefore, for a fixed S, $\chi \geq e^S$. Further limits for the scaling of χ can be obtained from area laws of entanglement entropy. For a 1D system the boundary of our region *B* stays constant even as the region grows under the renormalisation group flow. If the correlation length is finite (lattice sites very far away from each other are not entangled), we get a boundary law and the entanglement entropy will saturate to a constant maximum value.

Instead, at a critical point of a second-order quantum phase transition, the correlation length diverges and there is a logarithmic violation of the area law:

$$S \propto \log l$$

where, *l* is the linear size of the region.

Further if the system is *d* -dimensional with $d > 1$ we typically get an area law both on and off criticality,

$$S \propto l^{d-1}$$

where, *l* could be the length of the side of a (hyper)cube that we choose to be our *B*.

Looking at a coarse-graining block B_τ, its linear size grows exponentially in τ, as a larger and larger region is described by a single lattice point. For example, for a square lattice with a coarse-graining region of 3^d sites, $l_\tau = 3^\tau$. If we then assume the best-case scenario $\chi \approx e^S$, we get the following results:

$$\chi_\tau \leq \chi_{max} \sim e^{S_{max}} \qquad d = 1 \; non-critical$$

$$\chi_\tau \sim e^{\tau} \qquad d = 1 \; critical$$

$$\chi_\tau \sim e^{e^{\tau}} \qquad d \geq 2$$

Thus, the coarse graining scheming only finds applicability in limited cases (mostly 1D non-critical systems), beyond which it loses feasibility due to extremely rapid growth of χ for larger systems. This can be attributed to the fact that some of the short-range degrees of freedom, which present themselves in entanglement between neighbouring blocks, are not integrated out by coarse- graining. They persist and even accumulate under the renormalisation group flow, forcing the Hilbert space dimension χ to grow with τ and causing calculations to quickly become prohibitively expensive computationally. The situation is worse in higher dimensions because the boundaries between blocks are larger, as evident in double exponential growth.

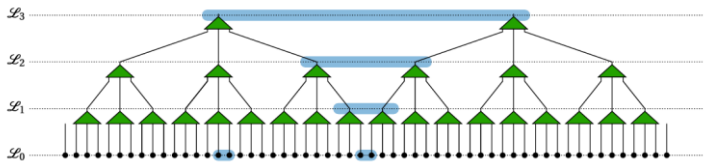

Figure 5: The unphysical accumulation of short-scale entanglement in a renormalisation group flow. Local entanglement between neighbouring sites (shown in blue) is treated differently depending on its location.

ENTANGLEMENT RENORMALIZATION

Entanglement renormalisation, proposed by Guifre Vidal (Vidal, 2006), is built to solve the problem encountered with the tree renormalisation scheme. It aims to remove short-range entanglement between blocks before applying the coarse-graining by introducing *disentanglers*, denoted by u. They are unitary transformations that operate on the sites at the borders of blocks to reduce local entanglement between them. To the 1D tree renormalisation that was illustrated earlier, u is added as a map between the two sites on either side of the border of two blocks, and two new sites.

$$u: V_{s_1'} \otimes V_{s_1'} \rightarrow V_{s_1} \otimes V_{s_2}$$

$$uu^\dagger = u^\dagger u = 1$$

These two new sites will then be fed to the coarse-graining isometries. In higher dimensions the principle remains the same.

As with the isometries w, we can define the disentangling operator for the whole lattice as

$$U = \otimes u$$

where the tensor product is over all the pairs of sites at the borders between blocks. This is illustrated in figure 6.

As unitary transformations, the disentanglers preserve information including the dimension of the state space of the sites on which they operate. This is unlike what happens with the isometries, which pack three sites into one, retaining only the ground state properties and projecting states to the

low energy subspace. The disentangling transformation is akin to reorganising or adjusting the border between blocks, so that the degrees of freedom that are entangled would be on the same side of the border, minimising cross-block entanglement that would cause problems as it accumulates.

Consider four spin-$1/2$ particles in a chain, $r_1\ s_1\ s_2\ r_2$, and an isometry w that is about to coarse-grain the two middle ones. Let the whole state of the system be,

$$|\psi\rangle = \frac{1}{2}\left(|0\rangle_{r_1}|1\rangle_{s_1} + |1\rangle_{r_1}|0\rangle_{s_1}\right)\left(|0\rangle_{r_2}|1\rangle_{s_2} + |1\rangle_{r_2}|0\rangle_{s_2}\right)$$

$$= \frac{1}{2}(|0101\rangle + |0110\rangle + |1001\rangle + |1010\rangle)$$

This is a worst-case scenario, where the *r* and *s* spins are pair-wise in maximally entangled Bell states. This is seen in the reduced density matrix

$$\rho_{s_1 s_2} = Tr_{r_1 r_2}(\rho) = Tr_{r_1 r_2}(|\psi\rangle\langle\psi|)$$

$$= \frac{1}{4}(|00\rangle\langle 00| + |11\rangle\langle 11| + |10\rangle\langle 10| + |01\rangle\langle 01|)$$

which is a maximally mixed state. Coarse-graining this would either cause great errors in the truncation or force us to retain the whole state space. Now consider adding disentanglers u that operate on the pairs $r_1 s_1$ and $s_2 r_2$. We write u as

$$u = \sum_{i,j,a,b=0}^{1} u_{ij}^{ab}\ |a\rangle_r |b\rangle_s \langle i|_r \langle j|_s$$

with

$$u_{01}^{00} = u_{10}^{00} = u_{01}^{01} = -u_{10}^{01} = \frac{1}{\sqrt{2}}$$

$$u_{00}^{10} = u_{11}^{11} = 1$$

$$u_{ij}^{ab} = 0 \qquad otherwise$$

Such a u maps:

$$\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \rightarrow |00\rangle$$

Thus, after the action of both the disentanglers on $|\psi\rangle$, the state of four spins is $|0000\rangle$. The reduced density matrix is given by,

$$\rho_{s_1 s_2} = |00\rangle\langle 00|$$

Now, coarse-graining has no issues. Only one of the eigenvalues is non-zero, and so the use of disentanglers has reduced the dimension of the effective site state space from 4 to 1.

Apart from the addition of disentanglers the procedure of entanglement renormalisation is the same as tree renormalisation. Like the isometries, the u tensors may or may not be the same on different layers of the renormalisation group flow and at different points on the lattices.

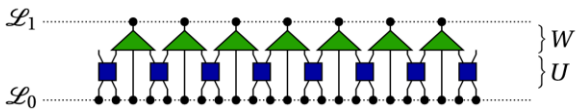

Figure 6: The coarse-graining transformation of entanglement renormalisation. The green isometries are now accompanied by the blue disentanglers u.

Due to U, the cross-block entanglement is now reduced, leading to more efficient coarse-graining, leaving us with a smaller χ. For typical local Hamiltonians χ_τ need not grow at all as τ grows. For scale invariant states, χ is found to be constant and for states with finite correlation length it approaches 1. This allows for arbitrarily many coarse-grainings and fulfills the natural criterion that scale invariant states produce a respectively symmetric renormalisation group flow. In addition, the U_τ's on different layers deal with entanglement over growing distances and thus the description of the state $|\mathbf{G}\rangle$ which is now encoded in the *u* and *w* tensors and the state $|\mathbf{G}_T\rangle$ has entanglement neatly organised according to length scale.

Vidal considers the 1-dimensional quantum Ising model with a transverse magnetic field on an infinite lattice

$$H = \sum_i \sigma_i^x \sigma_{i+1}^x + h \sum_i \sigma_i^z$$

He first obtains an approximation of the ground state using matrix product states and then applies entanglement renormalisation on this state and observes the effect the use of disentanglers has. This is illustrated in Figures 7 & 8 (from Vidal, 2006).

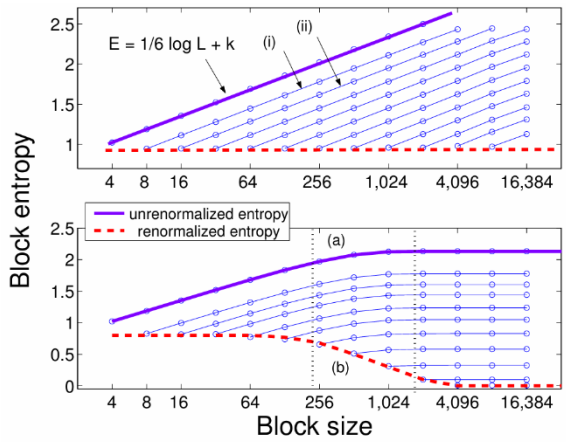

Figure 7: Scaling of the entanglement entropy in 1-dimensional quantum Ising model with transverse magnetic field, using entanglement and tree renormalisation. ***Top:*** *The critical case h = 1.* ***Bottom:*** *The noncritical case h = 1.001. The purple lines correspond to tree renormalisation, the red ones to entanglement renormalisation. The blue lines (i), (ii), etc. come from using disentanglers only on the first layer, first two layers, etc*

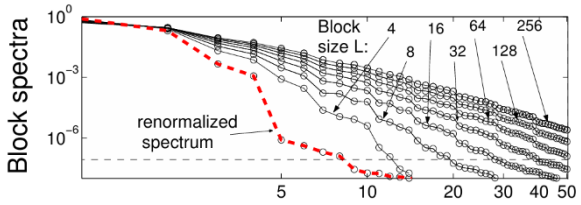

Figure 8: Spectrum of the density matrix of a coarse-graining block for various block sizes at the critical point. Without disentanglers, as the block size L grows, the number m of eigenvalues required to achieve the desired precision grows roughly exponentially in $\tau = \log_2 L$*. As disentanglers are added, the spectrum settles to a much steeper curve, which results in a lower m that is independent of the number of coarse-grainings done*

Figure 7 shows the ground state entanglement between a coarse-graining block and the rest of the lattice, as the block grows under successive renormalisations, compared with and without disentanglers. Without disentanglers the amount of entanglement grows with the block size, as short-range entanglement accumulates. This growth is linear in the critical case $h = 1$, but when disentanglers are added the growth stops and entanglement is left at a constant level. Off-criticality and without disentanglers the growth saturates to a maximum at the correlation length, whereas with disentanglers it decreases to zero.

Figure 8 shows the spectrum of the reduced density matrix of a block of spins, i.e. the distribution of eigenvalues of different magnitude, at the critical point. Without disentanglers the tail of the distribution grows with the block size, making computations more taxing. Adding disentanglers cuts the tail and makes it independent of the number of coarse-grainings applied, allowing an arbitrary number of iterations.

TENSOR NETWORK STATES

PENROSE NOTATION

The notation is closely analogous to the abstract index notation of tensors, where the contractions of different tensors, corresponding to composition of functions or equivalently taking traces of exterior products, is denoted by using the same index for the tensors that are contracted. For example, to denote $\Gamma : V^\dagger \times V \to \mathbf{C}$ contracted over the second argument with the first argument of $\Sigma : V^\dagger \times V^\dagger \to \mathbf{C}$, we write $\Gamma^a{}_b \, \Sigma^{bc}$. The appearance of the notation is exactly the same as with the Einstein summation convention, but in abstract index notation indices do not have numeric values and the symbols such as $\Gamma^a{}_b$ do not denote any numerical factors in some given basis, but the abstract tensor itself as a multilinear function.

In the graphical Penrose notation, a rank (n, m) tensor is represented by some shape that has n open-ended wires or 'legs' coming from it that point down and m wires pointing up.

A product of tensors is represented by grouping the symbols for the tensors together, and commutativity of the tensor symbols in the abstract index notation enables freedom of order. A contraction over indices is denoted by joining together the open ends of wires, where one must be pointing down and the other up, so that the types of the arguments of the tensors match. This is the same thing as giving the indices the same symbol in the abstract index notation.

Two upper and two lower indices cannot be contracted, as that would mean feeding a dual vector to a function that takes vectors or the other way around. However, we often have a preferred isomorphism between V and the dual space $V^\dagger$, so that each vector has a corresponding dual vector. All these are denoted by bending an open wire so that it points in the opposite direction. The abstract

index notation equivalent is raising and lowering indices.

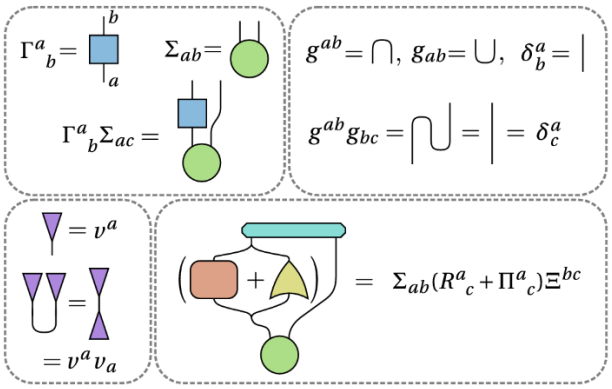

Figure 9: Some made up examples of tensor expressions drawn in the Penrose graphical tensor notation. ***Top left:*** *A simple example of two tensors and their contraction. Note that the blue tensor Γ that is of rank (1,1) is in fact nothing but a matrix.* ***Top right:*** *In relativity the metric gab and its inverse g ab are used to raise and lower indices. In the graphical notation this corresponds to bending upwards pointing wires down and vice versa. This is done by the cup and cap presented. The contraction of the metric with its inverse is just the identity, the Kronecker delta, denoted by a straight line.* ***Bottom left:*** *A vector v^a is denoted here by a triangle, and it has just one outgoing wire. It is then contracted with itself using the cup of the metric. A new notation is also introduced here: Sometimes the dual of a tensor that has all the indices lowered or raised (here we only have one index though) is denoted by simply mirroring the shape so that it is upside down.* ***Bottom right:*** *A somewhat more complicated tensor contraction just for illustration. Note that the final expression has no outgoing wires just as the index notation has no free indices. Thus, the end result of the contraction is a scalar.*

TENSOR NETWORK STATES

Since the state space of a system is a vector space we can consider its elements, the quantum states, and the bounded linear operators operating on them as tensors of varying types. Alternatively, one can do tensor calculus using Dirac's bra-ket notation by denoting vectors, i.e. type (1,0) tensors, by $|kets\rangle$ and dual vectors, i.e. type (0,1) tensors, by $\langle bras|$. Thus, a rank (2,3) tensor Γ can be written as

$$\Gamma = \Gamma^{\mathrm{ab}}_{cde} = |\phi_a\rangle|\phi_b\rangle\langle\phi_c|\langle\phi_d|\langle\phi_e|$$

This allows one to handle many-body states as $(N, 0)$ rank tensors, where there are many-body states as living in the tensor product of N single particle state spaces.

Now, we consider that that our state is composed of a certain network (graph) of relatively simple tensors contracted with each other in various ways.

In Penrose's graphical notation these tensors and their contractions can be represented as a network.

A simple example is what are called matrix product states or MPS. Consider an N -body quantum state $|\psi\rangle \in \boldsymbol{V} = \underbrace{\bigotimes^{N} V_i}_{i=1}$. MPS is then the ansatz that $|\psi\rangle$ decomposes into a product of N rank three tensors as in figure 10. Each tensor has a free index or an open wire that represents the state vector of that site, plus two indices that are contracted with its nearest neighbours (as in a matrix product, hence the name). The contracted indices are also called *bonds* between the tensors.

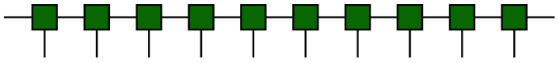

Figure 10: A matrix product state, MPS. The free legs are the indices of the state while the bonds form matrix products between the tensors at each site. The MPS resembles the structure of the lattice itself with nearest neighbour bonds. The legs sticking out to the sides can be taken to mean periodic boundaries or an infinite lattice.

The same MPS state written out the usual way would be

$$|MPS\rangle = M^{\alpha_1\beta_1}_{\beta_2} M^{\alpha_2\beta_2}_{\beta_3} \dots M^{\alpha_N\beta_N}_{\beta_1} |\phi_{a_1}\rangle|\phi_{a_2}\rangle \dots |\phi_{a_N}\rangle$$

Here we have set periodic boundaries. Each M is one of the green cubes in figure 10, the αs are the downwards pointing open legs and the βs, which are implicitly summed over, are the bonds. We say each of the contracted indices can obtain χ different values, which we call the *bond dimension* of the tensor network. This would correspond to the dimension of the matrix in an MPS. For a free index the number of values it can obtain is the dimension of the local Hilbert space of that site. However, in terms of tensor network states, χ has no relation to the spectrum of some operator, or any other similar criterion. Fixing χ (which could be different for each bond, generally assumed to be the same) is part of the structure of the tensor network.

Now, χ can be tuned to control the efficiency and accuracy of the ansatz. For $\chi = 1$ there is no contraction and the connecting legs vanish, leaving us with N rank 1 tensors multiplied with each other: a product state. With growing χ we can expand the reach of our ansatz, meaning our tensor

network can represent a larger part of the original Hilbert space. If we make χ exponential in N the MPS class of states covers the whole Hilbert space. This is illustrated in figure 11.

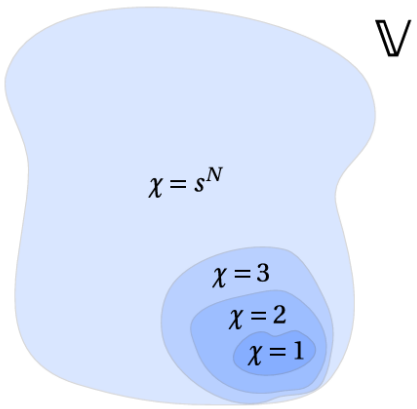

Figure 11: A symbolic illustration of the reach of a tensor network ansatz. (s here is the local Hilbert space dimension of a single degree of freedom). The key is in making the structure of the tensor network such that even at low χ the physically relevant parts of the state space are within the reach of the ansatz

Increasing χ makes all computations more resource intensive, reducing the usefulness of the ansatz. So, one needs to balance to do numerics on the tensor network and making sure it can represent the states one wants it to represent. Usually, the better the structure of our tensor network corresponds to the structure of the physical state (e.g. in terms of correlations between sites), the lower we can keep our χ, though increasing the number of tensors and bonds in the network does increase the computational cost.

For matrix product states, with low χ they naturally support correlations between sites of the chain that decay exponentially with distance, and also naturally lead to saturation of entanglement entropy at large scales. Both of these properties are essentially due to nearest neighbour connections. Thus, MPS can efficiently represent states that have these features. Importantly, the structure of MPS is also such that observables for the state can be calculated with a cost of the order $N\chi^p$, i.e. only linear in N. This ensures that we can efficiently not only gain a representation of a state, but also extract information from it.

Tree renormalisation leads to a tree-like structure of coarse-graining isometries. If we treat those isometries as type $(3,1)$ tensors then as a whole the tree is a type $(N,1)$ tensor and is called the tree tensor network (TTN). When using TTN as a tensor network ansatz the tensor elements are free variables one may vary to get the tensor network to represent different states. This enables us to numerically search for the ground state by keeping the structure of the network and the symmetry properties of the tensors fixed and varying the tensor elements, without any reference anymore to the roles of the tensors as disentanglers or coarse-grainers or requirements such as White's rule.

It is to be noted that for a TTN to be a physical N -body state it would need to be of type $(N\,,0)$, not $(N,1)$. The one dual index comes from the top tensor of TTN, which needs to be contracted with a type $(1,0)$ vector to obtain a state.

For tensor network for entanglement renormalisation, we now add disentanglers to TTN. This leads to the sought-after *multiscale entanglement renormalisation ansatz (MERA)*. As with TTN, we consider MERA to be a variational ansatz for quantum states, with the structure of its network following from entanglement renormalisation. Many other kinds of tensor networks could also be devised.

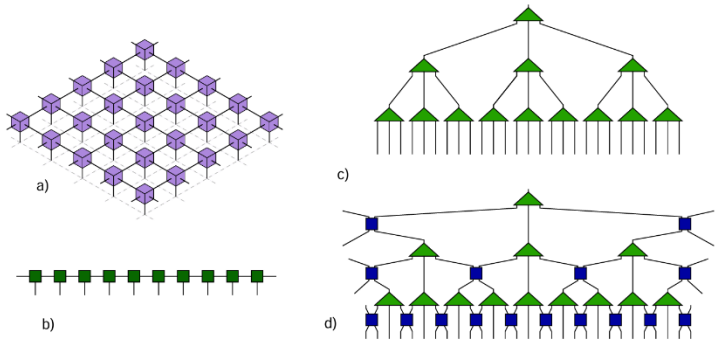

Figure 12: All the different tensor networks mentioned in the text. a) PEPS, b) MPS. c) (ternary) TTN, d) MERA

MERA: THE ANSATZ

The tensor network obtained from applying entanglement renormalization to a given lattice state is the MERA ansatz. Given the lattice and corresponding Hamiltonian, we consider that the ground state can be represented as a MERA of fixed bond dimension χ. (For a finite lattice of N sites, the tensor network has bounded depth, since entanglement renormalisation will eventually

coarse-grain the whole lattice to a few tensors that cannot be coarse-grained further).

At the end of the renormalization group flow, the last few sites are mapped by the *top tensor t* of the MERA, as illustrated in figure 13. The range of the one outgoing dual index of *t* is defined as the rank of the MERA (χ_T). It is the dimension of the local Hilbert space of the one effective site at the top.

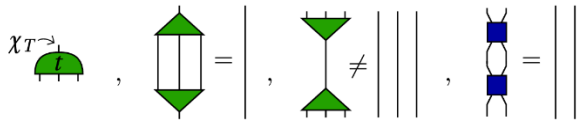

Figure 13: The tensors of MERA in the graphical tensor notation. On the left a top tensor. In the middle the isometry condition $w^\dagger w = I$ and the note that $ww^\dagger \neq I$. On the right the unitarity condition $uu^\dagger = u^\dagger u = I$

The rank of a MERA is also the dimension of the subspace $\boldsymbol{V}_{MERA}(\subset \boldsymbol{V}_L)$ that it describes (the tensor elements are fixed). For $\chi_T = 1$, the corresponding MERA represents a single pure state, usually the ground state. The top tensor has no dual indices at all and the MERA is a tensor network of type $(N, 0)$. If $\chi_T > 1$, the MERA can describe the χ_T degenerate ground states, or the ground state and $\chi_T - 1$ lowest excited states. In this case the type of the MERA is $(N, 1)$.

The top tensor is reached at the level $\tau = T \approx \log N$. Inverting this gives a logarithm, the base of which depends on the MERA scheme: 2 for binary, 3 for ternary, etc. This slow growth of the depth of the tensor network as a function of *N* is important for computations to scale efficiently. The logarithmic depth also means that, as usually with tree-like structures, there are only $O(N)$ tensors in the whole network. Thus, the memory needed to store a MERA is $O(\chi^{\#legs} \cdot N)$ (the number of elements in each tensor is $\chi^{\#legs}$).

QUANTUM CIRCUIT & CAUSAL STRUCTURE

Since MERA involves the action of related isometries and unitary operators and models the system accordingly, its function can be modelled as an equivalent quantum circuit built out of unitaries. One such use of unitary operator (tensor) is as shown in figure 14.

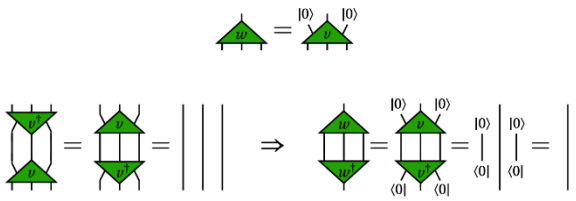

Figure 14: The isometry w as a unitary tensor v that is contracted with two constant vectors

As a quantum circuit a MERA should be read top-down. Depth of the tensor network is an additional dimension associated with the renormalisation group flow. Now, it functions as the time of the quantum computation, with the positive direction running down the tensor networks towards the original lattice L_0(in the renormalisation picture).

A quantum computation simply modifies the state of a qubit register, and does not destroy or create degrees of freedom like a coarse-graining does. So, a quantum circuit MERA *M* takes a lattice state as its input and outputs another lattice state. Consider the default input state to be,

$$|Input\rangle = |\underbrace{0 \ldots 0}_{N}\rangle$$

Then the output is the state at the initial stage in entanglement renormalization. Now we need N inputs, which necessitates adding more dual legs to isometries w and top tensors t so as to make M of type (N, N). These are then contracted with constant $|0\rangle$ vectors, which allows us to decompose the isometries in terms of unitaries (v) as follows:

$$w = w_\sigma^{\alpha\beta\gamma} |\phi_\alpha\rangle \left|\phi_\beta\right\rangle \left|\phi_\gamma\right\rangle \langle\phi_\sigma|$$

$$= v_{\sigma\rho\tau}^{\alpha\beta\gamma} |\phi_\alpha\rangle \left|\phi_\beta\right\rangle \left|\phi_\gamma\right\rangle \langle\phi_\sigma| \langle\phi_\rho|0\rangle \langle\phi_\tau|0\rangle$$

such that,

$$vv^\dagger = v^\dagger v = I$$

$$w^\dagger w = \langle 0|\langle 0|v^\dagger v|0\rangle|0\rangle = I \cdot \langle 0|0\rangle^2 = I$$

Thus, MERA top-down becomes a quantum circuit. At each level the coarse-grainers feed to the computation new degrees of freedom that were previously in the $|0\rangle$ state. Then, what were previously disentanglers are now entanglers, that entangle the state of these new degrees of freedom with the rest of the lattice. Thus, the circuit starts from a product state and first entangles sites quite far away from each other, then proceeds to

entangle intermediate sites in between and continues until it reaches neighbouring sites. By the time it gets there, the circuit has created a state that can have entanglement at all length scales. This state is the result of the algorithm, and the state which the MERA as a tensor network represents.

Now, we can analyse the "time evolution" of MERA as a quantum circuit. Tracing back the computation that led to the state of the site $s \in L_0$, an output wire of the circuit, it is observed that the number of tensors and wires that affects s is limited at all times. This set of tensors is defined as the (past) causal cone of s, denoted C_s. At each coarse-graining step, the disentanglers broaden the causal cone C_s, because the disentangling process mixes neighbouring sites. Countering this, each $w^\dagger$ then contracts the cone by mapping several of the neighbouring tensors to a product of $|0\rangle's$ and a single site state. The exact way these two effects compete and play out depends on the type of MERA we have, but the crucial result is the same: The width of the causal cone C_s, given by the number of effective lattice sites it includes at each level L_τ, is bounded at all times τ to a number independent of N. A causal cone is illustrated in figure 15.

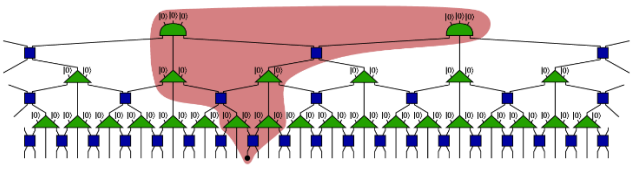

Figure 15: The causal cone of a single site in a 1-dimensional ternary MERA. Counting effective lattice sites inside the causal cone at each level, we see that their number is limited to 2.

This strict causal structure of MERA is essential for its computational efficiency, and is one of its advantages. the width of the causal cone for a MERA is bounded from above at all depths and for all MERAs in all dimensions and shapes. This is crucial because when calculating observables at certain sites, we only need to consider tensors inside the causal cone of those sites.

CALCULATING OBSERVABLES

Consider a MERA for the given state, with density matrix $\rho^{[s',s'+1]}$ on two neighbouring effective sites of the layer L_τ in the tensor network. Consider a local operator $o^{[r,r+1]}$ defined on two neighbouring effective sites of the tensor network, but on another level L_ν. If $\rho^{[s',s'+1]}$ spans the width of the past causal cone of $o^{[r,r+1]}$, this is all we need to consider to calculate the expectation value $\langle o \rangle_\rho$, because we know the tensors of the MERA. However, to carry out the calculation we need a contraction by either raising o to level τ or lowering ρ to level ν.

The ascending superoperator A is defined by the way it operates on local operators such as o. It simply maps $o^{[r,r+1]}$ to the corresponding two-site operator $o^{[r',r'+1]}$ defined on $L_{\nu+1}$:

$$o^{[r',r'+1]} = A o^{[r,r+1]}$$

The explicit form of A depends on the type of MERA (the dimensions of the lattice, the types of disentanglers and isometries used and also on how the sites r and $r + 1$ are positioned on the lattice relative to the disentanglers).

The corresponding dual $D = A^\dagger$ lowers density matrices (or in principle any other operators) down the MERA. From previously defined $\rho^{[s',s'+1]}$, D gives the two-site density matrix $\rho^{[s,s+1]}$ on $L_{\tau-1}$, assuming s and $s + 1$ are in the causal future of s' and $s' + 1$:

$$\rho^{[s,s+1]} = D\rho^{[s',s'+1]}$$

Thus, knowing $\rho^{[r',r'+1]}$, we can write,

$$Tr\left[o^{[r,r+1]} D(\rho^{[r',r'+1]})\right] = Tr\left[A\left(o^{[r,r+1]}\right)\rho^{[r',r'+1]}\right]$$

It is crucial that local operators are mapped to local operators by A. This is due to the limited width of causal cones. If anything that happens at two neighbouring sites can only depend on a limited number of sites in the causal past, then any operators lifted up the MERA must also stay inside this past. It ensures renormalising a system with local interactions produces a coarse-grained system that still has only local interactions.

Now, consider a given two-site operator o_0 at L_0 and a MERA. We denote the whole of MERA with all its tensors by M, $M \equiv U_0 W_0 U_1 \dots W_{T-1} t$. Now to know $\langle o_0 \rangle_{\boldsymbol{v}_{MERA}}$ we need to evaluate,

$$Tr\left[M^{\dagger} o_0 M\right] = Tr[o_0 P]$$

If M is a pure state ($(N, 0)$ type), then the above equation reduces to $\langle M|o_0|M\rangle$.

The equation is evaluated by repeated contractions using $vv^{\dagger} = v^{\dagger}v = I$ and $ww^{\dagger} = I$. The process is illustrated in figure 16. The tensors and their duals cancel each other out everywhere but in the causal cone, where o_0 stands in the way.

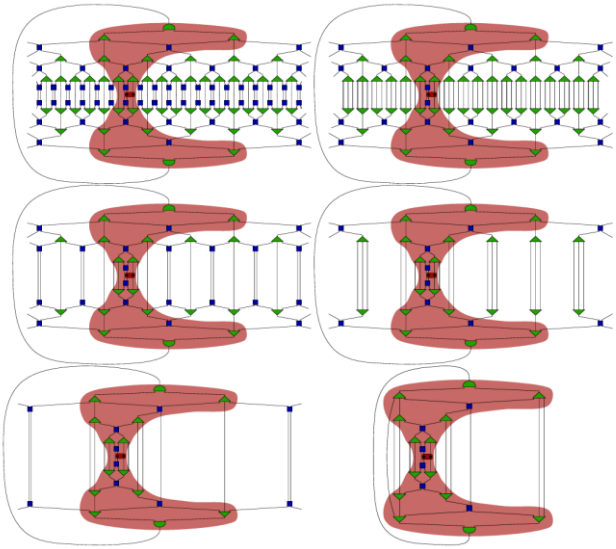

Figure 16: In contracting the whole tensor network, the dual tensors cancel out in many cases, forming simple identity operators. The series of figures here presents how this advances step by step, until there are no more trivial contractions left. The causal cone, marked in light red, emerges as the result of this.

We can start out by taking the tensors next to o_0 and contract them first, and then proceed radially outwards from there until we reach the top tensor and obtain our desired scalar that is the expectation value. This amounts to using A repeatedly on o to raise it all the way up to the level T, and is shown in figure 17.

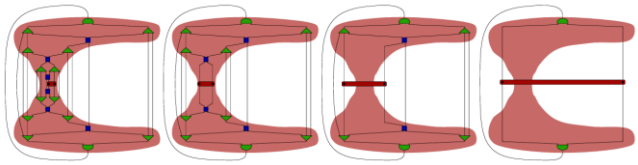

Figure 17: The contraction of a two-site operator (red) with a MERA proceeding from the inside out, i.e. using the ascending superoperator. In the end we have a simple trace over four tensors which then evaluates to a scalar that is the desired expectation value.

Equivalently, we can start by contracting everything near the top tensor first and proceed from there towards o_0. This means first extracting the density matrix $\rho_{T-1} = tt^{\dagger}$ from the top tensor and then repeatedly applying D to get the density matrix relevant for o_0. We can also do something in between the two, using both A and D, and make the operator and the density matrix meet on some level τ, where we can then evaluate $\langle o\rangle = Tr[o_\tau \rho_\tau]$.

Because the depth of the tensor network is log N and the causal cone is bounded, order is a matter of preference, and one ends up contracting $O(\log N)$ layers of tensors, meaning $O(\log N)$ applications of a superoperator. This brings the computational cost of evaluating a local operator to O(cost of contracting log N).

The cost-of-contraction there is the computational cost (time complexity) of contracting the tensors in a superoperator. This depends on the number of legs we need to contract, the types of tensors they are attached to and the bond dimension χ. In general, a contraction of a single index between two tensors has the complexity of $O\,(\chi^{\#legs})$, where $\#legs$ means the total number of legs of both the tensors, including (but not double counting) the leg that is contracted. In index notation this would be the number of different indices required to write the contraction.

OPTIMIZATION

Consider a Hamiltonian with nearest neighbour interactions given by,

$$H = \sum_r h^{[r,r+1]}$$

To find the ground state we naturally want to minimise the energy.

The expectation value of the Hamiltonian for our MERA state that we want to minimise is,

$$E = Tr[HP] = Tr\left[M^{\dagger} H M\right] = \sum_r Tr[M^{\dagger} h^{[r,r+1]} M]$$

where, $P = M^{\dagger}M$ is the projection operator. We can rewrite,

$$E = \sum_i Tr\left[w M_i w^{\dagger} N_i\right] + E_0$$

where M_i and N_i are parts of the tensor network between w and its dual, as we take the trace in E. The constant E_0 is the contribution to energy coming from outside the future and past causal cones of w and $w^\dagger$, and is irrelevant here.

We see that E is quadratic in w. Now, one needs to solve such an optimisation problem.

For simplicity, assume that $w^\dagger$ is independent of w. That makes E linear in $\bar{w}$: $E = Tr[wY]$, where $Y = \sum_i M_i w^\dagger N_i$ in the environment of w. The sum is made out of as many parts as there are $h^{[r,r+1]}$ in the future causal cone of w. This is shown in figure 18.

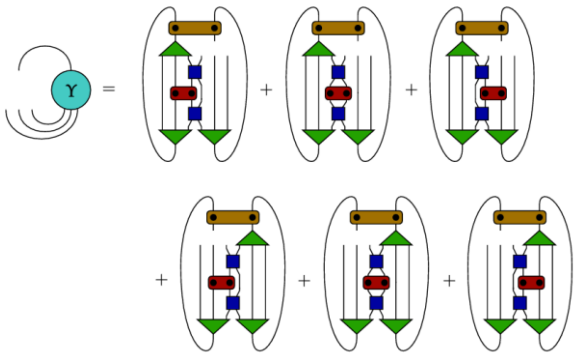

Figure 18: The environment Y of an isometry w in ternary MERA. The density matrix (brown) is calculated by contracting everything from the top tensor down to w, in essence applying D repeatedly. Similarly the terms of the Hamiltonian (red) are results of operating with A on the original terms $h^{[r,r+1]}$. The dual w† changes for every iteration, but the density matrix and the Hamiltonian do not, which makes subsequent calculations of the environment lighter after the first iteration.

Now, we decompose the environment as its singular value decomposition $Y = USV^\dagger$, where U and V are unitary and S is square and diagonal, with non-negative elements. Now, E be minimised by choosing $w = -VJU^\dagger$, where J is a rectangular diagonal matrix with 1s on the main diagonal.

Thus, we obtain,

$$E = Tr\left[-VU^\dagger USV^\dagger\right] = -Tr[S] = -\sum_i s_i$$

where, s_i are the singular values. The time complexity of the singular value decomposition is below $O(\chi^8)$ and so subleading.

This is the basic procedure, which we then iterate as follows:

1. Compute the environment Y of w.
2. Compute the singular value decomposition of Y.
3. Compute $w' = -VJU^\dagger$.
4. Set $w^\dagger = w'^\dagger$.
5. Go to 1.

In step 1 at each iteration only $w^\dagger$ changes. The other parts of the tensor network can thus be computed only once for the whole iteration process. This means that (for ternary MERA) the computational cost is dominated by either $O(\chi^8 \log N)$ coming from contracting most of the tensor network for step 1, or $O(\chi^8 q_{\text{tensor}})$ coming from the iteration. Here q_{tensor} is the number of iterations we do.

Now, we simply loop over the whole MERA in an organised manner. As an initial starting point, a MERA with random tensor elements works fine. We then begin by computing all the necessary local density matrices of the MERA.

We first compute the density matrix of L_{T-1} from the top tensor, then from that we get the density matrices of L_{T-2} and so forth. Thanks to using the previous layer to obtain the next one, this can all be done in $O(\chi^8 N)$ time, i.e. only a single superoperator contraction is needed per tensor.

Then using these and the given Hamiltonian, we optimise the lowest layer of tensors as described before. We may for example start with the disentanglers, moving from left to right, and then move on to the isometries. This is repeated q_{layer} times until the change in energy is small enough.

Next, we use the newly optimised tensors to build the ascending superoperator that we need to get the Hamiltonian terms on the next lattice. Then, using this newly obtained $H_1 = A(H)$ and the previously computed density matrices (optimising a layer below them does not affect them) we then go over the next layer of tensors just as we did for the first one.

Repeating this layer by layer we eventually reach the top tensor. That we then finally set to correspond to the χ_T lowest energy eigenvectors of

the Hamiltonian of L_{T-1}. This can be done with an exact diagonalisation algorithm with negligible cost. We can then start over, calculating the new density matrices from the new top tensor and repeat the whole process. This whole process is repeated q_{MERA} times until a preset degree of convergence is reached for E, and in the end we have our final MERA.

The computational cost of this whole process scales as $O\left(\chi^8 \cdot N \cdot q_{tensor} \cdot q_{layer} \cdot q_{MERA}\right)$. In practice, it has been found (Evenbly & Gidal, 2009) that it is best to keep q_{tensor} and q_{layer} quite small, say 4 or 5.

IMPLEMENTATION

We considered a crcitical quantum 1D Ising model with 256 sites, in the presence of transverse magnetic field ($J = h = 1/2$). The corresponding errors in energy (compared to the exact values) of ground state for the system (with periodic boundary conditions) as a function of iterations in varMERA is illustrated in figure 19.

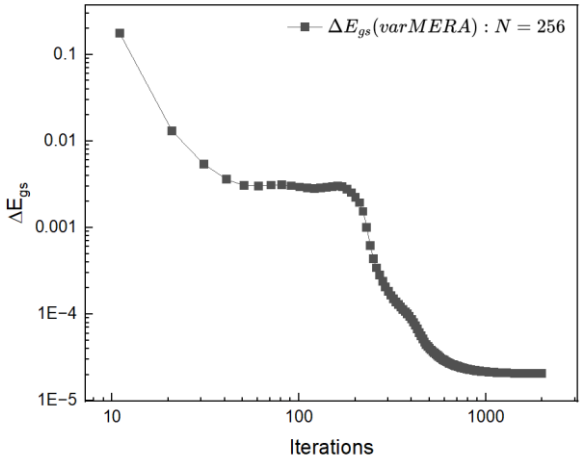

Figure 19: ΔE_{gs} vs iterations for critical quantum Ising model, with N =256, in the presence of transverse magnetic field. The error decreases rapidly with increasing iterations, with decent tolerance value ($\sim 10^{-5}$) attained at small (~ 1000) number of iterations

This simulation was done for a fixed value of bond dimension and magnetic strength. However, it was observed that increase in magnetic field or bond dimension χ lead to better performance (reduced error for given iterations), though no significant change was observed in the convergence rates (w.r.t iterations).

Furthermore, the corresponding conformal scaling dimension was calculated (using MERA) and compared with the analytic values, as shown in figure 20.

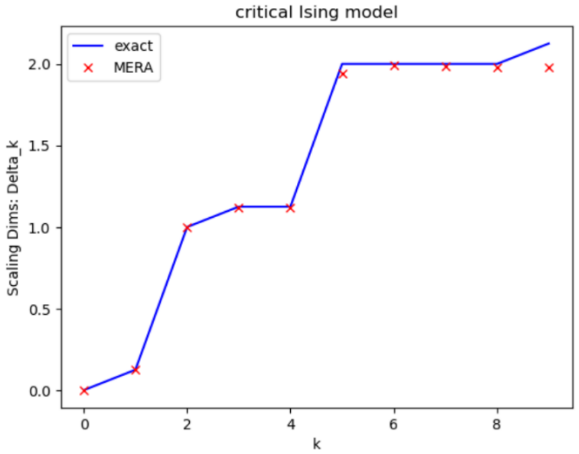

Figure 20: Relative error in scaling dimensions for critical quantum Ising model, with N =256, in the presence of transverse magnetic field.

We obtained excellent agreement with literature (Vidal 2006, 2009) in both results.

APPLICATIONS & EXTENSIONS

TYPES OF MERA

Finite Range MERA: A scale invariant MERA is naturally infinite both in width and depth. A finite MERA has depth log N, because at the log N it coarse-grains to a single top tensor. However, we can of course choose to terminate our MERA before reaching the top. This allows us to then consider infinite translation invariant systems.

Let us say we terminate the MERA at depth T and consider states that factorise at this level. The MERA then ends in a product of several top tensors of type (3,0) (we assume $\chi_T = 1$). Such a MERA necessarily has a maximum range of correlations ζ of the order of $\zeta \approx 3^T$, because the causal cones of the top tensors have finite width at L_0 and so sites further than ζ apart can not be in causal contact. Hence we call such a tensor network a *finite range MERA*.

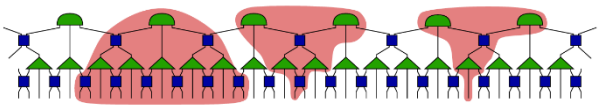

Figure 21: Part of a ternary finite range MERA. The future causal cone of a top tensor, shown on the left, has a fixed width at the bottom layer. This width is the maximum distance for correlations between sites, as sites further than this apart do not have a common past. To the right, the sites are too far apart and their causal cones do not intersect

For a finite system in a state with correlation length of the order of ζ a finite range MERA is a computationally more efficient option than the full MERA up to log N. It captures all the relevant physics because beyond $\tau = \log \zeta$ the state factorises, but is less costly to optimise and use. Thus, a finite range MERA is especially advantageous for a translation invariant system

2D MERA: Conceptually the generalisation to several dimensions poses no problems. The lattice is broken into small pieces, disentanglers are applied on the borders, which now form $d - 1$ dimensional surfaces, and isometries are applied on the disentangled regions. This approach is also computationally feasible, albeit more complicated than simulations for chains.

The extension involves introduction of more disentanglers to accomadate the faster accumulation of entanglement (much larger state space), though balancing the effect of bond dimensions becomes an issue. However, even with the much steeper exponent of χ, benchmark simulations are encouraging.

Branching MERA: This generalization draws on the the strengths of MERA: bound causal structure and special geometry. Its causal structure, though different from that of MERA, also allows for efficient contractibility, but covers a wider class of states, that do not necessarily obey a boundary law for entanglement.

A branching MERA state can exhibit bulk entanglement that scales as the volume l^d of the system, contrasted with the l^{d-1} scaling of the ordinary MERA, for the case $d > 1$. For a one-dimensional system branching MERA exhibits $S_l \propto l$ whereas the ordinary MERA gives logarithmic scaling.

Moreover, for different holographic trees with a variable number of branches, the causal structures are somewhere between those of ordinary MERA and the full width of 2^{T-t} branches. Thus, we get something that lies between the boundary and the bulk laws. This way we can tune our tensor network to naturally support such scalings as $S_l \propto l^{d-1} \log l$ or in the one dimensional case $S_l \propto \log^2 l$.

This makes it a viable choice for studying some systems for which the ordinary MERA is not suitable, such as Fermi liquids and spin-Bose metals etc. The normal MERA arises as a special case of the branching one.

CONCLUSIONS

We have extensively discussed the multiscale entanglement renormalisation ansatz and entanglement renormalisation. The formulation and functioning of the ansatz was discussed in detail, followed by a successful implementation of the method in determining the ground state energy of 1D quantum Ising model with a transverse magnetic field on finite lattice (256 sites). The usability of the method was further discussed in various use cases viz-a-viz its advantages and limitations. Also, possible extensions were briefly commented upon.

REFERENCES

- G. Vidal, *Entanglement renormalization*, arXiv:cond-mat/0512165v2 [cond-mat.str-el] (5 Dec 2006)
- A. Milsted, G. Vidal, *Geometric interpretation of the multi-scale entanglement renormalization ansatz*, arXiv:1812.00529v1 [hep-th] (3 Dec 2018)
- R. Pfeifer, G. Evenbly, G. Vidal, *Entanglement renormalization, scale invariance, and quantum criticality*, arXiv:0810.0580v2 [cond-mat.str-el] 10 Apr 2009
- D. Pomarico, *Multiscale Entanglement Renormalization Ansatz: Causality and Error Correction. Dynamics* 2023, 3, 622–635.https://doi.org/10.3390/dynamics3030033
- J. Eisert, M. Cramer, M.B. Plenio, *Area laws for the entanglement entropy- a review*. Rev. Mod.Phys. 82, 277 (2010). arXiv: 0808.3773 (28th Aug. 2008).

- G. Evenbly et al. *Boundary quantumcritical phenomena with entanglement renormalization*. Phys. Rev. B 82, 161107(R) (2010). arXiv: 0912.1642 (9th Dec. 2009)
- G. Evenbly & G. Vidal, *Algorithms for entanglement renormalization*. Phys. Rev. B 79, 144108 (2009). arXiv: 0707.1454 (10th July 2007)