SISSA
Scuola
Internazionale
Superiore di
Studi Avanzati

Physics Area – PhD course in
Theory and Numerical Simulation of Condensed Matter

Towards next-generation methods to optimize two-dimensional tensor networks: Algorithmic differentiation and applications to quantum magnets

Candidate:
Juraj Hasik

Advisor:
Dr. Federico Becca

Academic Year 2018-2019

Acknowledgements

First of all, I would like to thank Federico Becca for his supervision over the course of this work.

I want to thank all my friends and colleagues from SISSA, M. Angeli, C. Artiaco, L. Arceci, K. Baumann, M. Berovic, L. Coronel, N. Costa, L. Crippa, L. Ercole, L. Fanfarillo, F. Ferrari, M. Ferri, C. de Franco, C. Genovese, L. Gigli, G. Gori, F. Grandi, F. Grasselli, D. Guerci, L. Isaeva, A. Isidori, S. Karakuzu, D. Karki, L. Kohn, S. Laporte, J. Marcheselli, G. Mbeng, K. Nakano, A. Nava, E. Panizon, S. Pappalardi, M. Paolan, G. Piccitto, N. Qaisrani, M. Rusishvili, A. Sartori, M. Seclí, Y. Shaidu, F. Surace, M. Teruzzi, D. Tisi, A. Urru, M. Wauters. Piece by piece they filled the mosaic of my life in Trieste.

I am also grateful to P. Corboz, P. Czarnik, A. Francuz, O. Gauthé, A. Gendiar, G. Giudici, R. Haghshenas, A. Lauchli, D. Poilblanc, M. Rader, M. Rams, S. Sorella, M. Stoudenmire, and E. Tossatti for a wealth of stimulating discussions.

Finally, I owe my deepest gratitude to Nastasia.

Contents

1

Introduction

> *The fundamental laws necessary for the mathematical treatment of a large part of physics and the whole of chemistry are thus completely known, and the difficulty lies only in the fact that application of these laws leads to equations that are too complex to be solved.*
>
> — Paul A. M. Dirac 'Quantum Mechanics of Many-Electron Systems', Proceedings of the Royal Society (1929), A, 123, 714-733.

The field of condensed matter is concerned with a complete understanding of all forms of ordinary matter, in particular solids and liquids. It seeks to explain their structure, their mechanical and electromagnetic properties both in ambient conditions and when subjected to high or low temperature, pressures or magnetic fields. The discovery of quantum mechanics allowed posing this problem in a unifying picture of many nuclei and electrons all interacting quantum mechanically through Coulomb force. However, the systems of interest, whether a tip of a pencil or a tea in a cup, consist of a monstrous number of these interacting nuclei and electrons. Indeed, just within such tip of a pencil there is roughly 10^{26} carbon atoms forming a solid structure - far more than there are grains of sand in the whole Sahara. And so, as Dirac stated, although we can write down the *ab initio* equations governing such system, the pursuit of their exact solution is futile.

In the face of this seemingly intractable task, we adopt a different viewpoint. For every question, we first reduce the original *ab initio* formulae down to their simplest possible form, which however still captures the essence of the problem and only then we look for their solution. Indeed, this approach has proven to be extremely fruitful. Take the electrons. Already at a seemingly superficial level of analysis by neglecting all the interactions between them and preserving just their fermionic nature in form of Pauli's exclusion principle a surprisingly rich array of phenomena in Nature can be explained. For example, the heat capacity of metals and their electronic density of states or the emission and absorption spectra of light atoms. Sometimes, one substitutes the original system altogether. Rather then thinking in terms of interacting nuclei and electrons take just an ensemble of a few hundred classical point particles interacting through simple Lennard-Jones pair potential. By simulating such a system across a range of temperatures and densities a phase diagram emerges, containing three common phases of matter: Gas, liquid, and solid. In some parts of the diagram, these phases are separated by sharp boundaries and they represent the points where the phase transitions occur.

At the outset, studying such simple artificial systems might seem detached from the complicated reality, instead, the opposite is true. The collective behavior of many

interacting constituents is *universal*. By tuning the two parameters of Lennard-Jones potential, the energy scale and the characteristic distance, one can infer the temperature-density phase diagram of noble gases, water or even more complex molecules. Another instance of such universality is the celebrated Laudau's Fermi liquid theory, which describes the low-temperature behavior of most metals in terms of quasiparticles, representing the collective motion of the underlying electrons. While the electrons might be interacting strongly, these quasiparticles will behave as almost free with their effective mass dictated by the detailed chemical composition of the material. It is this notion of universality, which underpins the study of realistic condensed matter systems by simplified effective models.

In this work, we are concerned with a particular subset of such effective models, the so-called quantum magnets. They describe interacting magnetic moments, spins, arranged on a lattice. Their origin can be traced to a problem of interacting electrons in a periodic potential provided by the inert nuclei arranged on a lattice, which is a generic description of a typical solid with a crystal structure. Suppose, there is just one orbital with a single valence electron per lattice site, which is free to participate in the interaction as the rest is tightly bound to the nuclei and in the chemical bonds forming the crystal. If the strong interaction suppresses the mobility of these electrons, the whole system can be described just by interacting magnetic moments frozen on lattice sites

$$H = \sum_{\langle ij \rangle} J_{1,ij} \mathbf{S}_i \cdot \mathbf{S}_j + \sum_{\langle\langle ij \rangle\rangle} J_{2,ij} \mathbf{S}_i \cdot \mathbf{S}_j + \sum_{\langle ijkl \rangle} Q_{ijkl} (\mathbf{S}_i \cdot \mathbf{S}_j)(\mathbf{S}_k \cdot \mathbf{S}_l) + \dots, \quad (1.1)$$

where $\mathbf{S}_i$ is the spin operator on site i and $\langle ij \rangle$ denotes the nearest neighbors, $\langle\langle ij \rangle\rangle$ the next-nearest neighbors, $\langle ijkl \rangle$ the quartet of sites and so on. The interaction constants $J_{1,ij}$, $J_{2,ij}$, $Q_{ijkl}, \dots$ are given by the detailed chemical composition and the crystal structure of the material.

In this work we will be concerned with paradigmatic examples of such systems in two dimensions. The spin-1/2 isotropic Heisenberg antiferromagnet, defined on a square lattice

$$H = J \sum_{\langle ij \rangle} \mathbf{S}_i \cdot \mathbf{S}_j, \quad (1.2)$$

with $J > 0$ and its frustrated extension, the so-called $J_1 - J_2$ model

$$H = J_1 \sum_{\langle ij \rangle} \mathbf{S}_i \cdot \mathbf{S}_j + J_2 \sum_{\langle\langle ij \rangle\rangle} \mathbf{S}_i \cdot \mathbf{S}_j, \quad (1.3)$$

with $J_1, J_2 > 0$. Among many others, the $J_1 - J_2$ model captures the magnetic properties of cuprates, a large family of strongly-correlated materials. The complete characterization of its phase diagram is still an open question. While the limit $J_2 = 0$ corresponding to Heisenberg antiferromagnet is Néel ordered and has been understood in terms of Landau-Ginzburg theory, the nature of the phase in its highly frustrated regime $J_2/J_1 \approx 0.5$ remains unclear. The tremendous effort of the last thirty years has been concentrated on unraveling its true nature as it might

host a spin liquid, which is an exotic, highly entangled phase of matter with unusual fractional excitations. Its existence was conjectured by P.W. Anderson in 1973 [1].

Due to their strongly-correlated character and a presence of sign problem the frustrated quantum magnets have been primarily studied by variational methods and they represent an extremely challenging problem. Tensor networks are a newcomer among the more established variational methods such as variational Monte Carlo or density-matrix renormalization group. In two dimensions the infinite projected entangled-pair states represent efficient tensor network ansatz for lattice models motivated by the entanglement properties of gapped ground states of local Hamiltonians. The main objective of this work will be to advance the current state-of-the-art in optimization of such tensor networks through algorithmic differentiation, which is a fast and generic method for computing numerically precise gradients of complicated scalar functions of many variables, in our case, the variational energy. We believe, that with these improvements the tensor networks will set the new standards for variational studies of frustrated magnets.

This thesis is organized as follows

- In Chapter 2 we first motivate and define infinite projected entangled-pair states and then formulate the basic tools necessary for their treatment
- In Chapter 3 we introduce the optimization methods for infinite projected entangled-pair states. The imaginary-time evolution and in particular, the gradient optimization using algorithmic differentiation
- In Chapter 4 we investigate complications arising in the imaginary-time optimization across continuous phase transition between non-magnetic and magnetically ordered phases of Heisenberg antiferromagnet on coupled ladders.
- In Chapter 5 we move to gradient optimization with automatic differentiation and study the frustrated $J_1 - J_2$ model.

and finally in Chapter 6 we summarize the outcome of our work and discuss the future directions.

2

Tensor networks in two dimensions

CONTENTS

2.1 Motivation: Entanglement properties of the ground States of lattice models

The conception of the (infinite) matrix product states, (i)MPS, and later the (infinite) projected entangled-pair states, or shortly (i)PEPS, was to a great degree motivated by the distinctive property of the ground states of quantum lattice models which are governed by local interactions. If the system is also gapped then, in the plain terms, the degree of correlation between the different spatial parts of the

system will be strongly limited. This short preliminary section aims to give a more precise description of the above property, the so-called area law, in the terms of the entanglement entropy (for a comprehensive review see Ref. [2]). Afterward, we will argue why the states respecting the area law stand out from the rest of the states in the Hilbert space and how (i)MPS and (i)PEPS take advantage of this fact in the search for the ground states of local Hamiltonians. At the end of this section, we address the violations of the area law in the systems that are still governed by the local Hamiltonians which however are not gapped. The violations are of a direct relevance since the gap generically closes at the point of continuous phase transition or is not present at all in the phases with the spontaneously broken $SU(2)$-spin symmetry. Finally, in this context, we formulate the central problem which will be the focus of this work.

2.1.1 Preliminary: The area law of entanglement entropy

In order to state the area law, we first define the notion of the entanglement entropy, which quantifies the degree of all correlations between two distinct spatial parts of the system. Let us make an arbitrary bipartition of the underlying lattice into a compact region A and its complement B

$$lattice = A \cup B. \tag{2.1}$$

Taking a generic pure state $|\psi\rangle$ of the system, we aim at constructing a reduced density matrix ρ_A of the subsystem A which carries all the information about the expectation values of operators supported on A. The state $|\psi\rangle$ can be expresssed in the Hilbert space given by the tensor product of Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ of the regions A and B

$$|\psi\rangle \in \mathcal{H}_{A\cup B} \text{ where } \mathcal{H}_{A\cup B} = \mathcal{H}_A \otimes \mathcal{H}_B. \tag{2.2}$$

Choosing an orthonormal basis $\{|a\rangle\}$ inside the space $\mathcal{H}_A$ and $\{|b\rangle\}$ in $\mathcal{H}_B$, the generic state $|\psi\rangle$ can be written as

$$|\psi\rangle = \sum_{ab} c_{ab}|a\rangle|b\rangle, \tag{2.3}$$

using a shorthand notation for the tensor product $|a\rangle|b\rangle := |a\rangle \otimes |b\rangle$ [1]. The reduced density matrix ρ_A is then defined to be

$$\rho_A := \mathrm{Tr}_B(|\psi\rangle\langle\psi|), \tag{2.4}$$

where the operation of partial trace $\mathrm{Tr}_B(.)$ maps the operators from the full Hilbert space $\mathcal{H}_{A\cup B}$ into the operators on the subsystem A. In this case, a simple pure state density matrix $|\psi\rangle\langle\psi|$ is mapped into some mixed state reduced density matrix ρ_A

[1] The inner product in $\mathcal{H}_A \otimes \mathcal{H}_B$ is naturally given elementwise as $\langle a'|\langle b'||a\rangle|b\rangle = \langle a'|a\rangle\langle b'|b\rangle = \delta_{a'a}\delta_{bb'}$.

over $\mathcal{H}_A$. Using the selected bases inside the Hilbert spaces of regions A and B we can define the result of the partial trace, component by component, as follows

$$\begin{aligned} \rho_A = \mathrm{Tr}_B(|\psi\rangle\langle\psi|) &= \sum_{a,a'} (\rho_A)_{a,a'} |a'\rangle\langle a| \\ (\rho_A)_{a,a'} &:= \left(\sum_b c^*_{a'b} c_{ab} \right). \end{aligned} \tag{2.5}$$

The reduced density matrix ρ_A contains all the information about the state of the system living in the region A. We can use it to quantify how strongly is the region A correlated with its complement B. The good measure of the "quantity" of all correlations is the entanglement entropy S_A, defined as the Von-Neumann entropy of the reduced density matrix ρ_A

$$S_A := -Tr_A(\rho_A \log \rho_A). \tag{2.6}$$

It is important to state that the S_A is exactly equal to the S_B, and so the definition is consistent [1].

Now we can give a more precise statement of the area law as the relation between the entanglement entropy of the subsystem A with its complement B. Defining the basic terms

$$\partial A := \text{boundary of } A \tag{2.7}$$

$$|A| := \text{volume of } A; \quad |\partial A| := \text{length of } \partial A \tag{2.8}$$

the celebrated area law [3], reads

Area law A state $|\psi\rangle \in \mathcal{H}_{A\cup B}$ obeys the area law if for any bipartition of the system into the region A and its complement B in the thermodynamic limit where both $|A|, |B| \to \infty$ while $|A| \ll |B|$, the entanglement entropy S_A scales at most with the length of the boundary of A: $S_A = O(|\partial A|)$.

In one dimension the area law translates into $S_A^{1D} = const$, since the boundary of a line split into two parts is just a single point. In two dimensions it instead becomes $S_A^{2D} = O(L)$ where $L = |\partial A|$ is the usual length of ∂A, e.g. if A is a disk then L is the length of its perimeter.

2.1.2 Projected entangled-pair states: The 21st-century ansatz for two-dimensional local and gapped models

Let us ask the following question: among all the states in the Hilbert space, are the states obeying the area law exceptional ? The answer is positive. They are, and

[1] this fact can be readily verified by choosing orthonormal bases $\{|a\rangle\}$ in $\mathcal{H}_A$ and $\{|b\rangle\}$ in $\mathcal{H}_B$ such that the coefficient matrix c_{ab} in the Eq. 2.3 is diagonal. Such choice of the bases is dubbed the Schmidt decomposition.

spectacularly so, since a random state in the Hilbert space will have the entanglement entropy scaling with the volume $|A|$ of the subsystem instead as conjectured by Page in 1993 [4] and proved few years later (see Refs. [5, 6]). This result established that the area law states are *rare*, yet by itself it did not make them physically interesting.

In classical systems (e.g., spin models), the whole state is given just by a product of states of the individual spins, the so-called *product state*

$$|\psi\rangle = |\uparrow\rangle_1 |\downarrow\rangle_2 |\uparrow\rangle_3 |\uparrow\rangle_4 \ldots |\rightarrow\rangle_i |\downarrow\rangle_{i+1} \ldots \tag{2.9}$$

For all such product states, the area law holds trivially. The reduced density matrix ρ_A for any A will be simply a pure state density matrix and hence $S_A = 0$, which quantifies the intuitive notion of no correlations between the region A and the rest of the system. Mean-field solutions of spin systems are a prime example of such states [1]. But typically, and especially in one and two dimensions, the ground state of a spin system defined in the spin basis would be a (complicated) superposition of such product states

$$|\psi\rangle = \sum_{s_1 s_2 \ldots} c_{s_1 s_2 \ldots} |s_1 s_2 \ldots\rangle \,, \tag{2.10}$$

where the number of the relevant coefficients c is exponentially large with the system size. Can we make any statement about the entanglement properties in such case ? Or perhaps at least in the specific case of local and gapped Hamiltonians ?

The tremendous effort at the beginning of the 21st century into the study of the entanglement properties of quantum mechanical systems led to a key insight. The ground states of all local and gapped Hamiltonians in one dimension obey the area law. In two dimensions, the ground states of local and gapped Hamiltonians seems to obey the area law. No counterexample is known. The rigorous result for the systems in the dimension lower than two was established by Hasting and Koma in 2006 [7], who in their seminal work connected the gap and locality with the exponential decay of all connected correlations, and eventually the area law.

The physical intuition behind the proof in one dimension and the belief for the area law to hold also in two dimensions stems from the decay of the correlations. If the system is gapped and local, all correlations should decay exponentially with the distance as $exp(-r/\xi)$, where ξ is the correlation length, which is inversely proportional to the gap. Hence, the correlation between the regions A and B is, roughly speaking, coming only from the strip of width ξ around the boundary ∂A. Such consideration naturally leads to the entanglement entropy S_A scaling at most with the length $|\partial A|$ of the boundary.

This property of the ground states was the guiding principle towards finding their efficient parametrization. The states respecting the area law form just an exponentially tiny subspace within the whole Hilbert space as conjectured by Page. Moreover, this subspace must contain the ground state of all local and gapped one

[1] Mean-field in the sense of all spin operators being replaced by their expectation values, which are then determined by minimizing the energy.

dimensional systems. It also contains the ground state of plenty (if not all) local and gapped two dimensional systems.

The vast part of the condensed-matter landscape, including the domain of quantum magnetism, is described by the effective models, which are local. Therefore, this "physical" or "relevant" corner of the Hilbert space is among the first places where one should look for the ground state. This is the fundamental realization that motivates the class of variational states that satisfy the area law by construction hence targeting solely the most relevant part of the Hilbert space. It is at this point where the (i)MPS and their two-dimensional counterpart (i)PEPS enter.

Now we give the specific definitions of these variational states. For the MPS we adopt the definition due to the Schwollock [8], which for a system of N spin-S sites with the open boundary condition reads

$$\begin{aligned} |MPS\rangle &= \sum_{s_1 s_2 \ldots s_N} c_{s_1 s_2 \ldots s_N} |s_1 s_2 \ldots s_N\rangle \,, \\ c_{s_1 s_2 \ldots s_N} &= \mathrm{Tr}_{aux}(M^{s_1} M^{s_2} \ldots M^{s_{N-1}} M^{s_N}) := \sum_{a_1 \ldots a_N} M^{s_1}_{1 a_1} M^{s_2}_{a_1 a_2} \ldots M^{s_{N-1}}_{a_{N-1} a_N} M^{s_N}_{a_N 1}. \end{aligned} \tag{2.11}$$

This MPS is parametrized by the set of N rank-3 tensors $\{M\}$ with the auxiliary bond dimension D (indexed by a's). The tensors M need not to be the same for the different sites $s_1, \ldots, s_N$. Thus, the exponentially large number of coefficients c is now expressed using only $N d_S D^2$ parameters, where d_S is the dimension of the physical spin S

$$\text{dimension of the Hilbert space of spin-}S: \quad d_S = 2S + 1. \tag{2.12}$$

The entanglement entropy under any bipartition is at most $S_A = log\ D$. The MPS provide, in a sense, the complete solution to all local and gapped Hamiltonians in one dimension. To approximate the related ground state for a system of size N such that $| |MPS\rangle - |\psi\rangle | < \epsilon$, the necessary D is at most polynomial in $1/\epsilon$ and independent of N [9, 10].

The PEPS were conceived in 2004 as a direct generalization of MPS to finite two dimensional systems by Verstraete and Cirac [11, 12]. Few years later, together with Jordan [13], they extended PEPS to homogenous, infinite lattices under the name iPEPS. Both PEPS and in particular iPEPS are a very efficient variational wavefunctions, which directly paramatrize the area law corner, since they obey it by construction. We define a PEPS for a spin-S system of size $N = L \times L$ with the open boundary conditions on a square lattice as follows

$$\begin{aligned} |PEPS\rangle &= \sum_{s_1 s_2 \ldots s_N} c_{s_1 s_2 \ldots s_N} |s_1 s_2 \ldots s_N\rangle \,, \\ c_{s_1 s_2 \ldots s_N} &= \mathrm{Tr}_{aux}(a^{s_1} a^{s_2} \ldots a^{s_N}), \end{aligned} \tag{2.13}$$

$$\mathrm{Tr}_{aux}(a^{s_1}a^{s_2}\dots a^{s_N}) := \sum_{\substack{r_1\dots r_N\\ c_1\dots c_N}} (a^{s_1}_{11c_1r_1} \quad a^{s_2}_{1r_1c_2r_2}\dots \quad a^{s_L}_{1r_Lc_L1}$$
$$a^{s_{L+1}}_{c_11c_{L+1}r_{L+1}} \quad a^{s_{L+2}}_{c_2r_{L+1}c_{L+2}r_{L+2}}\dots a^{s_{2L}}_{c_Lr_{2L}c_{2L}1}$$
$$\vdots \quad \vdots \quad \vdots$$
$$a^{s_{N-L+1}}_{c_{N-2L+1}11r_{N-L+1}}\dots \quad a^{s_N}_{c_{N-L}r_N11}). \quad (2.14)$$

This PEPS is parametrized in the terms of N rank-5 tensors $\{a\}$, where each tensor a^s_{uldr} has a site (or spin) index s of dimension d_S and four auxiliary indices of bond dimension D, related to the four directions of the square lattice: up, left, down, and right. Hence, the suggestive arrangement of a's in the above expression. For such a PEPS, the entanglement of a bipartition into two halves is at most $S_A = L\ log\ D$ and again we have used only polynomial number of parameters Nd_SD^4. The iPEPS, to be introduced in detail in Sec. 2.2, exploit the translational symmetry and use only polynomial number of parameters independently of the system size. Therefore, they stand out as an extremely promising candidate for the variational studies of the effective lattice models found in the field of condensed matter.

2.1.3 Gapless systems and the violations of the area law

If the ground state has to belong to the area law corner, besides locality, which is physically well-motivated, the only constraint on the Hamiltonian is the existence of the gap between the ground state and the rest of the spectrum. Yet in this work, the main topics of our interests are

- *Exotic phases of spin systems.* In particular, the spin-liquids and the valence bond solids (VBS) both of which preserve the $SU(2)$-spin symmetry.
- *Phase transitions.* The quantum phase transitions between these exotic phases and the more ordinary broken-symmetry phases captured by the Landau-Ginzburg theory.

Both in the symmetry-broken phases and at the point of the continuous phase transition the gap vanishes. Out of the several consequences, the single most relevant one for (i)MPS or (i)PEPS based description of such phenomena is the violation of the area law. The true ground state does no longer belong into the area law corner. How does the violation impact the accuracy of the ansatz ?

In one dimension, the violation is at most logarithmic

$$S_A = O(log\ L), \quad (2.15)$$

where $L = |A|$ is the length of the subsystem [14]. This is a relatively mild violation and as such (i)MPS still stand out as the state-of-the-art variational method even for one-dimensional local but gapless systems. Highly-accurate description of physics at the criticality is possible due to the two main factors. Firstly, the (i)MPS can be efficiently and reliably optimized with D pushed up to 10^4. Secondly, gathering all the information across the four decades of D, the well understood scaling of the

observables with the bond dimension ultimately leads to the reliable estimates for the thermodynamic limit.

In two dimensions the situation is more delicate. If the ground state has a Fermi surface, e.g. the case of free fermions, the area law will be violated by a multiplicative factor $S_A = O(L\ log\ L)$, as shown by Wolf [14]. A different, less severe yet very important scenario concerns the ground state which spontaneously breaks the continuous symmetry. In such a case, as shown by Metlitsky and Grover [15], the Area law is violated by a subleading logarithmic term

$$S_A \propto \alpha L + (N_G/2)\ log\ L, \tag{2.16}$$

where N_G is the number of Nambu-Goldstone modes in the system. Hence, this violation is common to be found in the realm of quantum magnetism, an ubiquitous Néel phase of anti-ferromagnets being an example. Therefore, the central question of this work which we want to address is

The descriptive power of iPEPS Can iPEPS still give us an insight into the continuous phase transitions and exotic phases of both frustrated and frustration-free quantum magnets, despite the violations of the area law which is the core principle behind the ansatz ?

At this point, contrary to the case of MPS the limits of iPEPS are poorly understood. Moreover, the optimization of iPEPS is extremely challenging and only very small values of bond dimension, between $D = 10$ and $D = 24$ depending on the exact method used, can be considered. In conclusion, we consider the ability of an iPEPS-based variational approach to explain the physics of two-dimensional quantum magnets to be an open question.

2.2 The infinite projected entangled-pair states, their diagrammatic notation, and reduced density matrices

This brief section aims to introduce the basics of iPEPS. First, to define iPEPS, as an appealing extension of the finite PEPS into the thermodynamic limit. Second, to introduce the diagrammatic notation of tensors and their contractions, as it will ease the manipulations and clarify tensor network expressions appearing throughout the rest of this work. Using this notation, we define a tensor network that represents the iPEPS itself. Finally, to build the reduced density matrices of iPEPS which are well defined even in the thermodynamic limit and contain accessible information about the observables. We close this section by a short comment on iPEPS with partially broken translational symmetry.

2.2.1 The iPEPS as a Thermodynamic Limit of the Translationally Invariant PEPS

In order to introduce iPEPS we follow the original exposition due to Jordan [13]. Taking the finite PEPS for spin-S system on a square lattice of size $L \times L$ with the open boundary conditions, as defined in the Eqs. (2.13) and (2.14), we take the limit of $L \to \infty$. Moreover, since we are interested only in the translationally invariant states, we place the same tensor a on every site

$$\begin{aligned} |iPEPS\rangle &= \sum_{s_1 s_2 \ldots} c_{s_1 s_2 \ldots} |s_1 s_2 \ldots\rangle \,, \\ c_{s_1 s_2 \ldots} &= \mathrm{Tr}_{aux}(a^{s_1} a^{s_2} \ldots), \end{aligned} \tag{2.17}$$

$$\mathrm{Tr}_{aux}(a^{s_1} a^{s_2} \ldots) := \sum_{\substack{r_1 r_2 \ldots \\ c_1 c_2 \ldots}} \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \iddots \\ \cdots & a^{s_7}_{c_7 r_9 c_6 r_{10}} & a^{s_8}_{c_8 r_{10} c_1 r_{11}} & a^{s_9}_{c_9 r_{11} c_{10} r_{12}} & \cdots \\ \cdots & a^{s_6}_{c_6 r_8 c_5 r_1} & a^{s_1}_{c_1 r_1 c_2 r_2} & a^{s_2}_{c_{10} r_2 c_{11} r_3} & \cdots \\ \cdots & a^{s_5}_{c_5 r_7 c_4 r_6} & a^{s_4}_{c_2 r_6 c_3 r_5} & a^{s_3}_{c_{11} r_5 c_{12} r_4} & \cdots \\ \iddots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{2.18}$$

where the tensors a within the trace are explictily arranged on a square lattice. In the very same way as in the finite PEPS case, the single rank-5 tensor a^s_{ulrd} has a physical index s of dimension d_S and four auxiliary indices of bond dimension D. These auxiliary indices are again interpreted as indices along the up, left, down, and the right directions of the square lattice. The iPEPS gives the variational approximation for the (infinitely many) coefficients c in terms of a single tensor a with only $d_S D^4$ parameters. The value of any single coefficient $c_{s_1 s_2 \ldots}$ is given by the contraction of infinitely many tensors a arranged on the square lattice as suggestively done in the defining equation[1].

The proper interpretation of this infinite sum is called for. We will address this point and put the iPEPS ansatz on a solid footing by introducing the reduced density matrices of iPEPS. However, before doing that, we first define a convenient diagrammatic notation which will be instrumental in manipulating expressions with (infinitely) many tensors.

2.2.2 Diagrammatic notation and the tensor networks

We define any rank-r tensor as a shape (circle, square, ...) with r legs corresponding to the individual indices of the tensor. A rank-1 tensor is equivalent to a vector, rank-2 tensor is equivalent to a matrix. They will constitute the basic building blocks of more complicated expressions. Let us remark that the notion of a rank of a tensor is not strict in the following sense: any two (or more) indices $i_1, i_2, \ldots, i_N$

[1] we arranged the tensors in a (clockwise) spiral, growing from the central site s^1 outwards. This is just a matter of convenience and does not have any significance for the later analysis

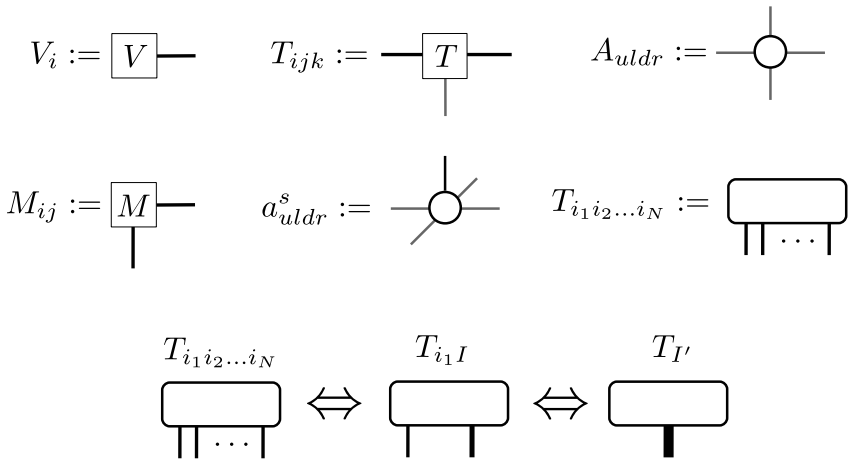


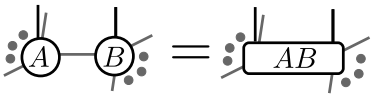
Figure 2.1: The basic diagrammatic definitions. Top row: vector V, rank-3 tensor T, and rank-4 tensor A. Middle row: matrix M, rank-5 tensor a, and rank-N tensor T. The shape is arbitrary. The indices might be assigned to individual legs when necessary. Bottom row: the equivalence between the three different forms of the rank-N tensor T, from left to right, the original rank-N tensor, a certain matrix form with fused indices $I = (i_2 \ldots i_N)$, and finally its vector form with all indices fused into a single index $I' = (i_1 i_2 \ldots i_N)$.

of a tensor can be combined, or fused, to form a single index I

$$fusion(i_1 \ldots i_N \to I) := \begin{cases} I & := i_1 \otimes i_2 \otimes \ldots \otimes i_N \\ dim(I) & = dim(i_1) \times dim(i_2) \times \ldots \times dim(i_N) \end{cases} \tag{2.19}$$

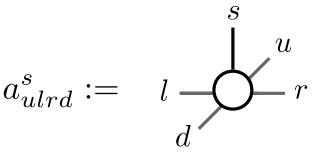
where the dimension of the index I is given by the product of the dimensions of the combined indices. We will exploit this freedom to give an interpretation of the different tensors appearing through the manipulation of the tensor expressions. The corresponding diagrammatic notation of all the basic building blocks is defined in Fig. 2.1. So far, we have defined just an operation of fusion, which changes the rank of the tensor. The essential operation of the tensor calculus is the contraction, which takes two tensors A and B and performs their contraction along the specified pair of indices (which have to share the same dimension). The result can be interpreted as a new tensor, inheriting the uncontracted indices of both A and B.

Tensor contraction The diagrammatic definition



$$\Leftrightarrow \sum_{i_1 i_1'} A^s_{i_1 i_2 \ldots} \delta_{i_1 i_1'} B^{s'}{}_{i_1' i_2' \ldots} = (AB)^{s s'}_{i_2 \ldots i_2' \ldots}. \tag{2.20}$$

All tensor expressions, which we will encounter throughout the rest of this work are composed solely of tensors contracted through their indices. The diagrammatic equivalent is a (possibly infinite) set of shapes connected by lines. Hence the name tensor networks.

$a^s_{ulrd} :=$ 

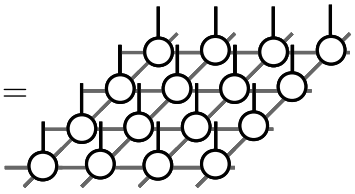
$\mathrm{Tr}_{aux}(a^{s_1}a^{s_2}\ldots) :=$ 

Figure 2.2: The iPEPS tensor network. Left: on-site tensor a with the legs labeled by the corresponding indices. The vertical leg is the spin index (black line) and the in-plane legs are the auxiliary indices (grey lines). Right: tensor network representing the vector of coefficients $c_{s_1 s_2 \ldots}$ of iPEPS. The network is infinite in all in-plane directions, auxiliary indices pointing out of the network are understood to be contracted with the next copies of the tensor a.

Using the introduced diagrams for tensors and their contraction allows us to give the diagrammatic definition of the iPEPS from Eq. (2.17). More specifically, we give the tensor network which represents all the coefficients c as the contraction of an infinite tensor network, defined in the Eq. (2.18) and composed only from copies of the single on-site tensor a.

2.2.3 The reduced density matrices of iPEPS

As mentioned before, the infinite sum that encodes the vector of coefficients c of the iPEPS [Eq. (2.18)] is not a tractable object. Instead, we adopt the following point of view: iPEPS primarily serve as the source of the reduced density matrices. Suppose we are interested in the observables that live in the compact region R. As we will show, we can extract the reduced density matrix ρ_R for the region R in a well-defined manner. Afterwards, all the observables $\mathcal{O}$ in the region R can be readily evaluated as

$$\langle \mathcal{O} \rangle = Tr_R(\rho_R \mathcal{O}). \tag{2.21}$$

To construct ρ_R, first split the system into the region R, and its complement $\bar{R}$. Then express the total wave function accordingly, as in the Eq. (2.2), on the product space of the Hilbert spaces of R and $\bar{R}$

$$|iPEPS\rangle = \sum_{s_1 s_2 \ldots} c_{s_1 s_2 \ldots} |s_1 s_2 \ldots\rangle = \sum_{\substack{s_1 \ldots s_r \\ s_{r+1} \ldots}} c_{s_1 \ldots s_r s_{r+1} \ldots} |s_1 \ldots s_r\rangle |s_{r+1} \ldots\rangle, \tag{2.22}$$

where the indices $s_1 \ldots s_r$ belong into the region R, while the remaining indices $s_{r+1} \ldots$ belong into the complement $\bar{R}$ instead. The reduced density matrix ρ_R is obtained by taking the partial trace over the complement $\bar{R}$. Using the result of the Eq. (2.5) we get

$$\begin{aligned} \rho_R &= \mathrm{Tr}_{\bar{R}}(|iPEPS\rangle \langle iPEPS|) \\ &= \sum_{\substack{s_1 \ldots s_R \\ s'_1 \ldots s'_R}} \left(\sum_{s_{R+1}, \ldots} {c^*}_{s'_1 \ldots s'_R s_{R+1} \ldots} c_{s_1 \ldots s_R s_{R+1} \ldots} \right) |s_1 \ldots s_R\rangle \langle s'_1 \ldots s'_R| \,. \end{aligned} \tag{2.23}$$

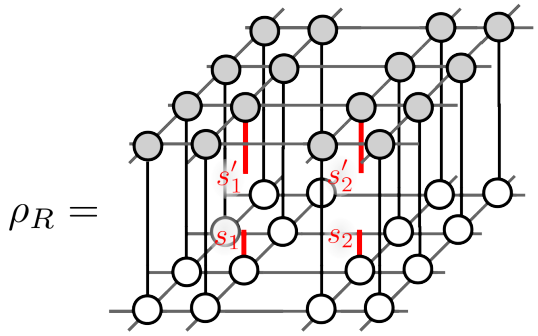


Figure 2.3: The reduced density matrix $(\rho_R)_{s_1 s_2, s_1' s_2'}$ represented as a double-layer tensor network, where the region R contains two spins $s_1 s_2$ (red vertical lines). The network is understood to be infinite, with the auxiliary indices on the edge connecting to next on-site tensors in both layers. All other spin indices are contracted (vertical black lines). All the auxiliary indices are contracted (gray in-plane lines).

The above expression is completely general. Notice, that the matrix elements of ρ_R are given by the contraction of the coefficient vectors c, c^* along all the indices $s_{R+1}, \ldots$ inside the region $\bar{R}$. Now we substitute for the coefficients c, c^* the iPEPS parametrization and without a loss of generality, take the region R that contains just two neighbouring sites s_1 and s_2. The corresponding reduced density matrix, given in components, is

$$(\rho_R)_{s_1 s_2, s_1' s_2'} = \sum_{s_3, \ldots} \mathrm{Tr}_{aux}((a^*)^{s_1'} (a^*)^{s_2'} (a^*)^{s_3} \ldots) \times \mathrm{Tr}_{aux}(a^{s_1} a^{s_2} a^{s_3} \ldots). \qquad (2.24)$$

Let us make use of the diagrammatic tensor network expression, as defined in Fig. 2.2, for the iPEPS coefficients. The above equation is equivalent to two such tensor networks with all their spin indices except on the sites one and two contracted. Hence, we have obtained a very illustrative interpretation of the iPEPS reduced density matrix as a double-layer tensor network, shown in the Fig. 2.3. After normalization, given by $\mathrm{Tr}_R(\rho_R)$, we have a well-defined density matrix for two neighboring sites. The reduced density matrices for any compact region R have a similar form. Let us remark, that the reduced density matrices for finite PEPS are obtained in the same way.

In this work, we are mainly concerned with the effective models of quantum magnetism, which typically contain just nearest- or next-nearest-neighbor interaction. As a consequence, it is enough to consider just a few basic regions. In fact, taking a square region $R = 2 \times 2$, we can already evaluate the energy of the Hamiltonians of the Heisenberg model or the $J_1 - J_2$ model. The second most important regions to consider are thin strips of length L, e.g. $R = L \times 1$, as their reduced density matrices contain the information about all two-point correlation functions.

It is clear now, that the reduced density matrices are the most important objects parametrized by iPEPS. Hence, we have to understand how to construct them, which means to make sense of a contraction of the infinite double-layer tensor network.

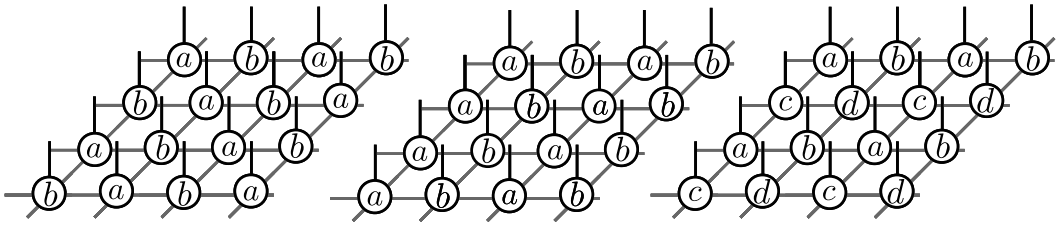


Figure 2.4: Tensor network representation of iPEPS which partially break the symmetries of the square lattice. Left: bipartite iPEPS with two unique tensors $\{a, b\}$ in 2×1 unit cell. Middle: iPEPS targeting the columnar VBS with two unique tensors $\{a, b\}$ in 2×1 unit cell. Right: iPEPS with four unique tensors $\{a, b, c, d\}$ arranged in 2×2 unit cell. Besides the AFM order and the columnar VBS, it can also support a plaquette VBS.

Before we move onto that topic, we complete the introduction of iPEPS by briefly describing their generalization to unit cells with more than one tensor.

2.2.4 The iPEPS with partially broken translational symmetry

When targeting the ground states of Hamiltonians, which have the spatial symmetry of the lattice, an obvious choice is the translationally invariant iPEPS with the same tensor on every site. However, the ground state might spontaneously break the spatial symmetry, at least partially. Such instances can still be straightforwardly captured by iPEPS. To do so, we introduce a unit cell with just a few unique tensors. The whole lattice is then tiled with this unit cell. For example, a bipartite tiling of the square lattice requires two tensors a, b

$$\begin{aligned} |iPEPS(a,b)\rangle &= \sum_{s_1 s_2 \ldots} c_{s_1 s_2 \ldots} |s_1 s_2 \ldots\rangle \\ &= \sum_{s_1 s_2 s_3 s_4 \ldots} \mathrm{Tr}_{aux}(a^{s_1} b^{s_2} a^{s_3} b^{s_4} \ldots) |s_1 s_2 s_3 s_4 \ldots\rangle, \end{aligned} \tag{2.25}$$

which are then arranged appropriately on the lattice. The tensor network representing such iPEPS is shown in Fig. 2.4. Depending on the unit cell and the pattern in which the lattice is tiled a rich set of wave functions with reduced spatial symmetries can be built. Among simple but important examples are iPEPS which can support various VBS orders such as columnar or plaquette VBS (see Fig. 2.4). The number of variational parameters grows only linearly, as the number of unique tensors within the unit cell. The reduced density matrices for these extended iPEPS are defined in the very same way as in the translationally invariant case. They are obtained by the contraction of the corresponding infinite double-layer tensor networks.

2.3 Contracting infinite tensor networks - Part I

We have established that iPEPS ansatz provides a very efficient parametrization of the area law abiding wave functions and their reduced density matrices. However, to obtain these reduced density matrices for generic iPEPS, in principle we have to contract an infinite double-layer network. To make the problem tractable we have to frame it differently. After all, our true aim is to formulate a well defined numerical procedure, which can be implemented on a computer and used to compute the observables of iPEPS, such as the energy or the correlation functions.

Let us go back for a moment to a conceptually simpler and well-defined case of a finite PEPS of size $L \times L$ on a square lattice with the open boundary conditions. Moreover, let this PEPS, which we dub uniform PEPS, be constructed from the copies of a single tensor a. Suppose, that we are interested only in the observables in the bulk of the PEPS. More broadly, in the reduced density matrix $\rho_R(L)$ of a region in the center of the double-layer network of this PEPS. For a finite uniform PEPS of sufficiently large size L , the reduced density matrices in the bulk are indistinguishable from the reduced density matrices of the translationally invariant iPEPS built from the same tensor a. Indeed, for the (i)PEPS with the exponentially decaying correlations there is a length $L \propto 1/\xi$, with ξ being the correlation length, beyond which there are virtually no correlations. Therefore, we define the reduced density matrix ρ_R of iPEPS by the following limit

$$\rho_R := \lim_{L\to\infty} \rho_R(L), \tag{2.26}$$

where the region R is in the centre of the $L \times L$ uniform PEPS.

As such, the reduced density matrices of iPEPS are well-defined and to compute them, we instead look for the best approximation[1] of $\rho_R(L)$ of the uniform PEPS with the progressively increasing size $L \times L$. Once there is no appreciable change between the reduced density matrices above some length L_c, we take such fixed point $\rho_R(L_c)$ to be equivalent to the reduced density matrices of the corresponding iPEPS itself.

In the rest of this section, we will introduce in the necessary detail the principal algorithms that execute the above-described procedure and find the fixed point reduced density matrices. We will present the so-called corner transfer matrix method (CTM), which can handle both simple iPEPS composed from a single unique tensor as well as the more complicated iPEPS with several tensors within the unit cell.

2.3.1 Prologue: Statistical mechanics of classical two-dimensional models and the corner transfer matrix

The first appearance of the corner transfer matrix method can be traced all to way back to the early work of R. J. Baxter on classical statistical mechanics (see Ref. [16]

[1] for the finite PEPS the exact contraction of the finite double-layer network is possible, yet feasible only for very small systems. Such a task is a #P-hard problem and no algorithm with polynomial running time is known [2]

for a more recent exposition). In his pursuit for the solution to the eight-vertex model [2], Baxter introduced CTM as a variational approximation for its partition function [17]. The partition function is given by a contraction of (infinitely) many identical Boltzmann weight factors A_{uldr}, with the indices $\{uldr\}$ representing a classical spin degrees of freedom, living on the links of the square lattice. When the partition function is expressed diagrammatically, we immediately recognize it as an instance of a tensor network, shown in Fig. 2.5(a)-(c)). Hence, in a sense, Baxter was the first one to address the problem of contracting infinite tensor networks, at least in the context of lattice models in physics.

The CTM is a slight twist to the more familiar row-to-row transfer matrix (RTM) approach. Within RTM approach, the partition function $Z = \sum_X exp\{-\beta E(X)\}$ is expressed as a product of the matrices R. Given a configuration of the system X, their elements encode the interaction between the degrees of freedom X_i and X_{i+1} in the consecutive rows of the lattice [see Fig. 2.5(d)]. The matrices R themselves are exponentials of certain Hermitian matrices and hence non-negative. Under the generic assumptions (local interactions, translational symmetry, independence on the boundary condition in the thermodynamic limit) the partition function is then compactly expressed as

$$Z = \mathrm{Tr}(R^L) \quad \Rightarrow \quad \text{diagonalize R} \quad \Rightarrow \quad Z = \sum_n \lambda_n^L. \tag{2.27}$$

Hence, after diagonalizing R, it is clear that the partition function and its dependence on the temperature can be fully elucidated from the spectrum $\{\lambda_0, \lambda_1, \ldots\}$ (ordered in a descending fashion) of the row-to-row transfer matrix. In fact, one of the dominant features which governs the physics of the system in question is the ratio between the leading and the subleading eigenvalue. In the case of the gapped spectrum, $\lambda_1/\lambda_0 < 1$, in the thermodynamic limit both the partition function as well as other thermodynamic quantities given by the logarithmic derivates of Z with respect to some external parameter p are completely dominated by the leading eigenvalue only

$$Z = \lambda_0^L \left[1 + \sum_{n>0} \left(\frac{\lambda_n}{\lambda_0}\right)^L\right]; \quad \partial_p lnZ = \frac{L\lambda_0^L}{Z}\left[\frac{\partial_p \lambda_0}{\lambda_0} + \sum_{n>0}\left(\frac{\lambda_n}{\lambda_0}\right)^L \frac{\partial_p \lambda_n}{\lambda_n}\right], \tag{2.28}$$

$$\Rightarrow \lim_{L\to\infty} \Rightarrow Z = \lambda_0^L; \quad \partial_p lnZ = L\left[\frac{\partial_p \lambda_0}{\lambda_0}\right]. \tag{2.29}$$

Therefore, a good approximation of the leading eigenvalue λ_0 immediately leads to estimates of intensive quantities such as magnetization, obtained by taking its derivative with respect to the applied magnetic field. Computing directly the free energy might not be possible as it depends on the actual magnitude of λ_0. Fortunately the intensive quantities do not share this dependence.

[2] vertex or ice-type models represent a family of models relevant for crystal lattices with hydrogen bonds. Linus Pauling introduced first such model in 1935, to account for residual entropy of water ice.

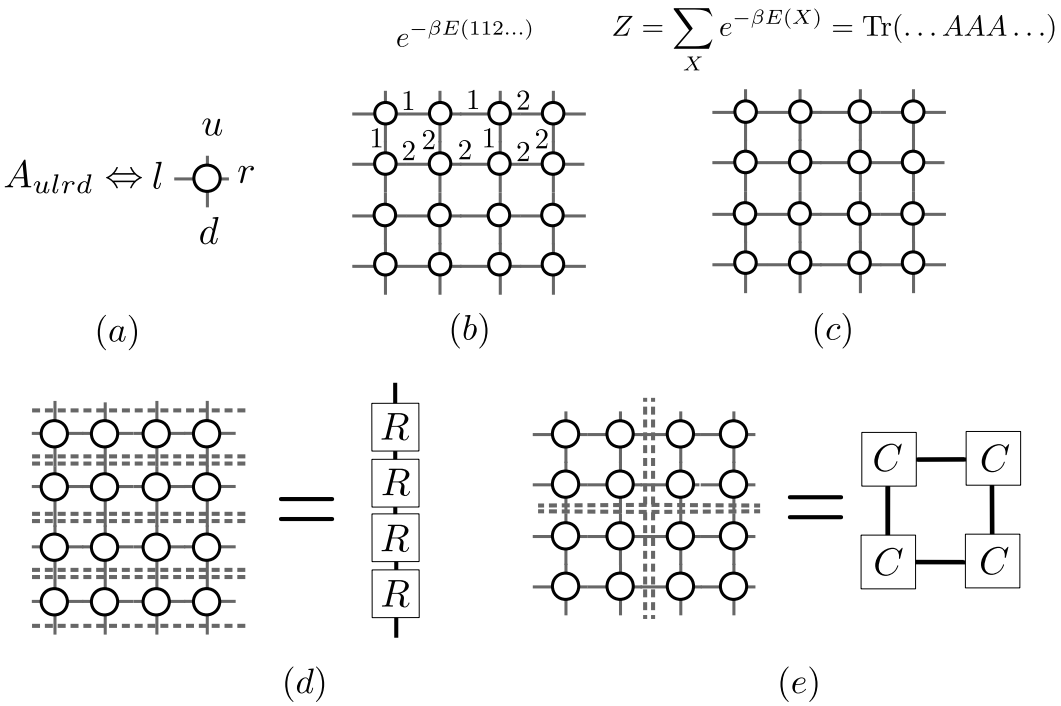


Figure 2.5: Vertex model: (a) Boltzmann weight A on each vertex of a square lattice associated with a particular configuration of four classical spin variables $(uldr)$. A is a real rank-4 tensor which elements are functions of β and any other relevant parameter e.g. interaction strength. (b) Boltzmann weight of a concrete configuration of spins on the square lattice. It is a product of numbers, elements of tensors A obtained by fixing the values of all their indices. (c) Partition function of the vertex model given by the sum of weights for all possible configuration. A result of contracting the tensor network. (d) Partition function expressed as a product of row-to-row transfer matrices R (the trace is omitted). Each matrix R is given by the contraction of one of the rows of the network. (e) Partition function expressed as a product of four corner transfer matrices C. Each matrix C is given by the contraction of one of the four corners of the network.

While the RTM approach splits the partition function into the product of matrices associated with the rows of the system, Baxter in his original work split the partition function into the product of four matrices C, and coined them corner matrices for the obvious reason, as shown in Fig. 2.5(e). To see the contrast, take a system of size $2L \times 2L$ with periodic boundaries, and express the partition function as $Z = \mathrm{Tr}(R^{2L})$ with RTMs. Leaving the size of the system unchanged, assume an open boundary conditions instead. The partition function can now be expressed by the product four matrices $Z = \mathrm{Tr}(C^4)$, each matrix C accounting for all interactions within one of the four $L \times L$ corners of the system. By diagonalizing these matrices, the partition function is given just by their eigenvalues

$$Z = \mathrm{Tr}(C^4) = \sum_i \mu_i^4 \tag{2.30}$$

The role of the spectra of the corner matrices C and the RTMs are identical, they contain all the information about the system. For sufficiently large L, the dependency on the boundary vanishes, and the two approaches coincide. Whereas the expression (2.29) for the partition function by RTMs greatly simplifies as $L \to \infty$, no such obvious simplification happens for the spectra of the corners. However, almost all the elements of these corners are exponentially tiny numbers, made up of sums of products of Boltzmann weights A. Thus, it is reasonable to expect that the spectrum of the corners decays rapidly and so just a first few eigenvalues μ already provide a very good approximation for the partition function and its derivatives.

Baxter found an analytic expression for the eigenvalues of the corners and solved the particular limit of the eight-vertex model. However, for generic models, an analytical solution is not known. Nevertheless, Baxter's idea is at the heart of later developed numerical CTM methods.

The fundamental (and surprising) relation between Baxter's work and the computation of the reduced density matrices of iPEPS is given by the following realization. The double-layer network of iPEPS is formally equivalent to the partition function of a vertex model. By contraction of all spin indices and the fusion of the selected auxiliary indices in the double-layer network of iPEPS one obtains a network identical to the partition function of a vertex model in the thermodynamic limit. This fact is illustrated diagrammatically in Fig. 2.6. The equivalence is only formal since the Boltzmann weights A of any vertex model are non-negative. There is no such restriction on the tensors (or "weights") A for the double-layer network of iPEPS. The same equivalence holds between the double-layer network of uniform PEPS and the partition function of a vertex model of $L \times L$ system.

In the upcoming section, we introduce a numerical method that provides an efficient approximation of the corner matrices for arbitrary "vertex-like" models. Doing so not only allows for the study of many classical two-dimensional models, but more importantly it will also solve the main problem of our interest, which is how to compute the reduced density matrices of iPEPS.

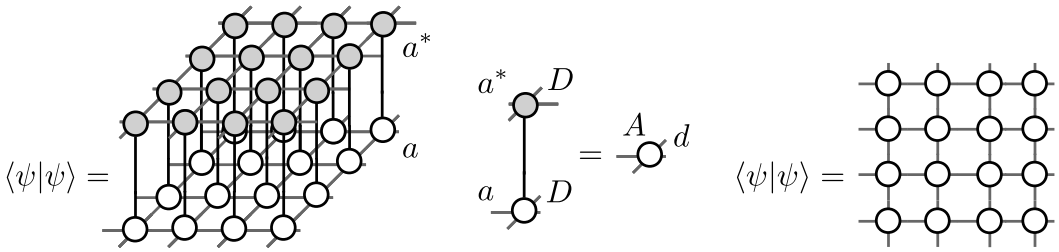


Figure 2.6: Left: contracted double-layer network representing the norm of iPEPS. Contracting first the on-site tensors a, a^* that are on top of each other gives a network equivalent to the partition function of a vertex model. Right: detailed contraction of the on-site tensor a^s_{uldr} with its complex conjugate $a^*_{\bar{u}\bar{l}\bar{d}\bar{r}}$. The result is a new tensor $A_{u'l'd'r'}$ with the auxiliary indices given by the fusion of the paired indices $(u\bar{u}), (l\bar{l}), (d\bar{d})$, and $(r\bar{r})$. The dimension of the fused indices of tensor A is $d = D^2$.

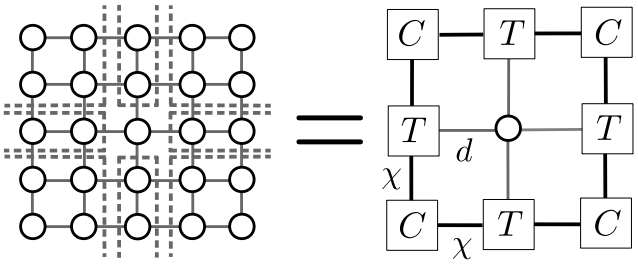


Figure 2.7: Partition function of 5×5 system with an open boundary condition. Its contraction is split and accumulated into eight distinct parts, corner tensors C of dimension $\chi \times \chi$ and half-row/-column tensors T of dimensions $\chi \times \chi \times d$. We call χ the environment bond dimension, or just an environment dimension. For the equality to be exact, the environment dimension must be (at least) $\chi = d^{L/2}$.

2.3.2 Corner transfer matrix renormalization group

In 1992, the DMRG algorithm was invented by White [18] and quickly became the state-of-the-art method for simulation of one-dimensional quantum systems. Shortly after, Nishino and Okunishi in their pioneering work [19, 20] derived a variational approximation for Baxter's corner matrices of the vertex model [1] and solved it using the ideas of DMRG. In doing so, they also provided us with a variational solution to the contraction of double-layer networks of uniform PEPS since these two problems are equivalent.

Their approach starts by considering a small finite system of size $L \times L$ with open boundary conditions. Its partition function is given by the contraction of just a small number of tensors A. As a first step, all the contractions are accumulated into eight distinct parts except the single site in the center of the network. Four corner tensors C, familiar from the Baxter's CTM, and four new tensors T, the so-called half-row/-column tensors. Together, we call these tensors *environment tensors* $E = \{C, T\}$ and give their diagrammatic definition in the Fig. 2.7. For small enough systems, the tensors C and T can be computed explicitly. However, that quickly becomes impossible as the required environment dimension grows exponentially with the system size L. Therefore, for large L, the tensors C and T can be only approximate. The contraction of the partition function is thus approximated by the environment tensors

$$Z(L) = \mathrm{Tr}_{L\times L}(AA\ldots A) = \lim_{\chi \to d^{L/2}} \mathrm{Tr}(CTCTATCTC), \tag{2.31}$$

with the exact result recovered in the limit of exponentially large environment dimension χ. A basic variational principle for the best rank-χ environment tensors $\{C, T\}$ can be formulated by maximizing the approximate partition function

$$\{C, T\} = \underset{rank-\chi}{argmin} |Z(L) - \mathrm{Tr}(CTCTATCTC)|. \tag{2.32}$$

Hence, the next goal is to find the solution of the above variational problem. Before

[1] in their work Nishino and Okunishi were considering the symmetric 16-vertex model on a square lattice, which includes Ising or Potts model as a special case

proceeding, let us mention, that the weights A_{uldr} are symmetric under arbitrary permutation p of the indices. Such tensor are dubbed *C4v symmetric* [1]

$$\text{C4v symmetric tensor:} \quad A_{uldr} = A_{p(uldr)}, \tag{2.33}$$

This permutation symmetry is retained by the environment tensors C and T, hence their elements satisfy the following identities

$$C_{ij} = C_{ji} \quad T_{iju} = T_{jiu}. \tag{2.34}$$

This detail will be important soon.

The central question is how to find these best approximate tensors for a system of (large) size L ? Nishino and Okunishi proposed an elegant solution analogous to the process of DMRG. Suppose we start with the small system, of size 3×3, and hence we can compute the tensors $\{C^{(3)}, T^{(3)}\}$ exactly. Now the key step comes: let us grow the system by a single layer of sites as illustrated in Fig. 2.8. The system becomes effectively of size 5×5 with the outermost layer formed from the environment tensors $\{C^{(3)}, T^{(3)}\}$ [1]. In the spirit of RG, we want to reduce the enlarged network back into its original form with only four corners C, four half-row/-column tensor T and a central site A. However, just contracting the tensors in order to form the new environment tensors $\{C^{(5)}, T^{(5)}\}$ will not do since every such contraction will increase their environment dimension by the factor d

$$\tilde{C}_{(id)(jr)} = \sum_{i'j'ul} C^{(3)}_{i'j'} T^{(3)}_{i'iu} T^{(3)}_{j'jl} A_{uldr}, \tag{2.35}$$

$$\tilde{T}_{(iu)(jd)r} = \sum_{l} T^{(3)}_{ijl} A_{uldr}. \tag{2.36}$$

Thus, in order to complete the RG step, we have to truncate the enlarged environment dimension from χd back to χ. Therefore, the next step is to find the projector P, which truncates the enlarged tensors $\{\tilde{C}, \tilde{T}\}$ in the optimal way.

Clearly, by truncation of the enlarged tensors, some of the information about the system will be lost. The question is how to throw away the least important bits. Consider the matrix $\rho = (C^{(3)})^4$. Due to the symmetries of $\{C, T\}$, given in the Eq. (2.34), ρ is hermitian and it can be interpreted as a reduced density matrix of a particular subsystem. Cut the partition function network from the center to the edge, the ρ is then the reduced density matrix of the degrees of freedom associated with the indices that have been cut in this way (see Fig. 2.9). In DMRG, one seeks the optimal truncation of a Hamiltonian of the enlarged one-dimensional quantum system. By diagonalizing it, the most relevant subspace is identified. Nishino and Okanishi realized, that the enlarged reduced density matrix $\tilde{\rho} = \tilde{C}^4$ can play a similar role. And hence, now we have a lead on how to construct the optimal projectors. Take $\tilde{\rho}$ and diagonalize it

$$\tilde{\rho} = USU^\dagger, \tag{2.37}$$

[1] the permutation of the indices can be interpreted as the rotation of the underlying square lattice. The group of symmetries of the square lattice is a C4v point group.

[1] in contrast with Baxter's approach, the introduction of half-row/-column tensors T is necessary in order to express the enlarged network through the environment tensors as done in Fig. 2.8

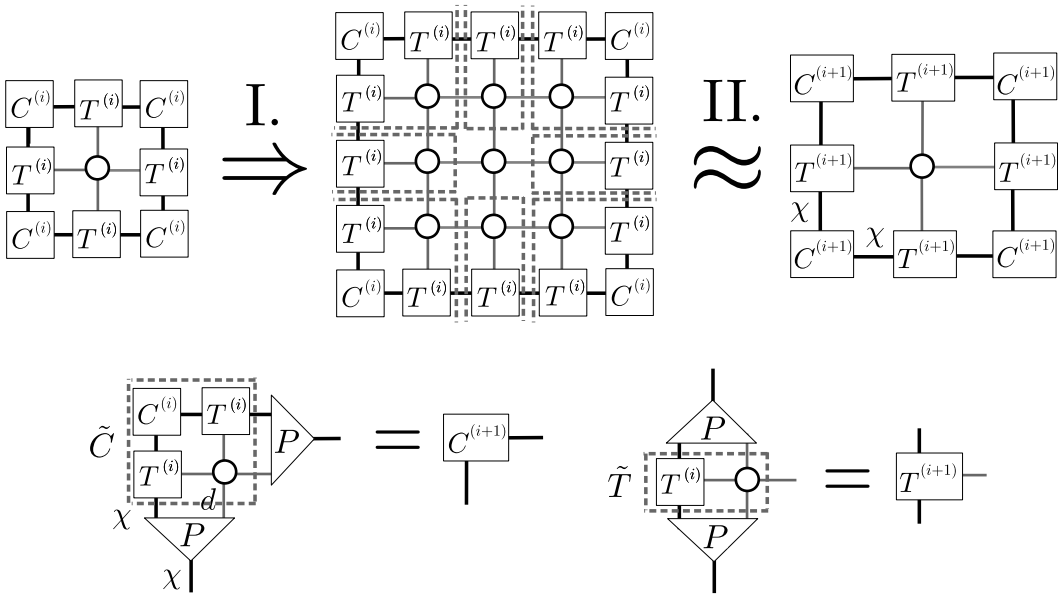


Figure 2.8: Top: A single RG step. First, the partition function network is enlarged by a single layer of sites A and the necessary environment tensors $\{C^{(i)}, T^{(i)}\}$. Second, the enlarged network is approximated by the new network with the updated environment tensors $\{C^{(i+1)}, T^{(i+1)}\}$ of the same dimensions. Tensor $\{C^{(i+1)}, T^{(i+1)}\}$ have "absorbed" the added layer of sites A and the original environment tensors in the eight distinct parts of the network (denoted by dashed lines). Bottom: Equations defining the updated environment tensors $\{C^{(i+1)}, T^{(i+1)}\}$. The projector P, a tensor of dimensions $\chi \times \chi \times d$, truncates the dimension of the enlarged corners $\tilde{C}$ and half-row/-column tensors $\tilde{T}$ in the optimal way (see text).

where we assume that the eigenvalues S are ordered in the descending manner. The most important subspace is then spanned by the eigenvectors with the eigenvalues of the largest magnitude [2]. Hence, the optimal projector is built up from precisely the χ leading eigenvectors

$$P := U_{[:,\chi]}. \tag{2.38}$$

Now we are finally in the position to complete the RG step, by truncating the enlarged tensors $\{\tilde{C}, \tilde{T}\}$ as illustrated in Fig. 2.8 back to their original environment dimension χ. This concludes a single RG step, resulting in a set of new environment tensors $\{C^{(5)}, T^{(5)}\}$, which however effectively describe a larger system of size 5×5. The new environment tensors approximate the larger system in the optimal way, that is minimizing the variational principle of the Eq. (2.45). Each such RG step enlarges the system from the size L to $L + 2$.

The necessary truncation introduces an error, which we estimate now. Consider the enlarged density matrix $\tilde{\rho} = \tilde{C}^4$ (shown in Fig. 2.9). Its trace gives an approximate partition function $Z(L-1)$ of a system of size $(L-1) \times (L-1)$, due to leaving

[2] in the case of iPEPS, the eigenvalues might be negative since weights A are no longer Boltzmann weights as in the case of vertex models.

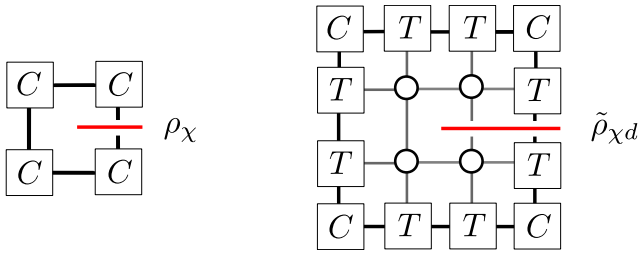


Figure 2.9: Left: reduced density matrix ρ_χ of dimensions $\chi \times \chi$ with respect to the cut from the center of the system to the edge (red line). Right: reduced density matrix $\tilde{\rho}_{\chi d}$ of the enlarged system of dimensions $\chi d \times \chi d$ to the same cut.

out the enlarged half-row/-column tensors $\tilde{T}$

$$Z(L-1) \approx \mathrm{Tr}(\tilde{C}^4) = \sum_{i=1}^{\chi d} S_{ii}^4, \tag{2.39}$$

where the second equality comes from the eigendecompostion of the enlarged corner, Eq. (2.37). The projector truncates the enlarged corner down to the rank-χ. Hence, the sum of the fourth powers of the eigenvalues S of the ignored subspace estimates the error on the partition function due to truncation

$$error(\chi) = \sum_{i=\chi+1}^{\chi d} S_{ii}^4 \tag{2.40}$$

Thus, the sufficiently fast decay of the spectrum of $\tilde{C}$ is crucial for the error to remain small.

With the single RG step done, we can now iterate the procedure and obtain the best rank-χ environment tensors for a $L \times L$ system of any size L. Simply start with the small 3×3 system and the corresponding set of environment tensors $E^{(0)} = \{C^{(0)}, T^{(0)}\}$. Perform a series of $L/2$ RG steps. At each step i, one has the optimal environment $E^{(i)}$ for system of size $(3+2i) \times (3+2i)$. After $L/2$ steps, the tensors $E^{(L)}$ provide the optimal rank-χapproximation of the partition function of the system of size $L \times L$ and by the equivalence also a double-layer network of uniform PEPS. This process constitutes the CTMRG algorithm.

Iterating the RG steps until no appreciable change in the environment tensors is observed results in the so-called fixed point environment tensors. They represent the best rank-χ approximation of the environment tensors for the partition function/double-layer network in the thermodynamic limit. Typically, these fixed point tensors are independent of the initial environment. We summarize the whole CTMRG procedure in the concise algorithmic form together with the details concerning the numerical implementation in Appendix B.1.

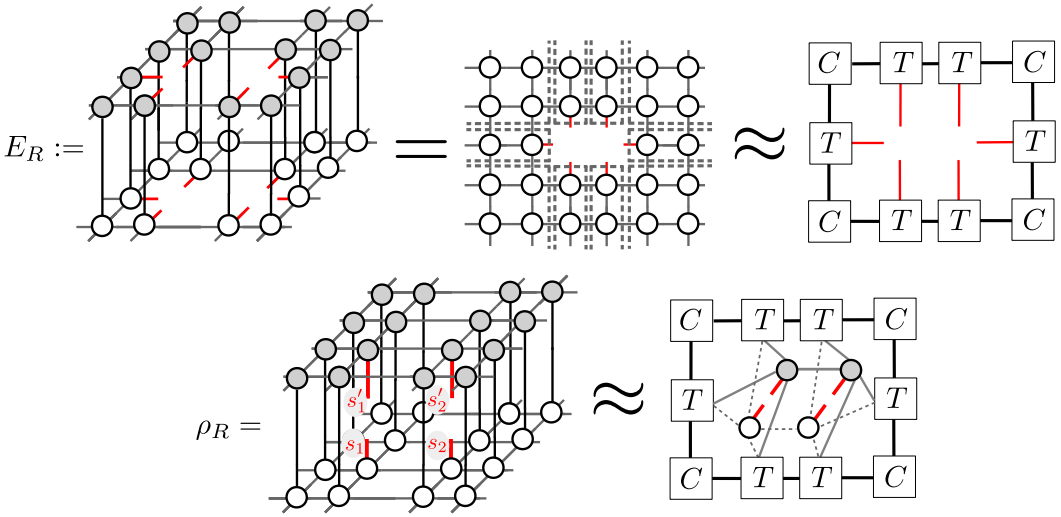


Figure 2.10: Top: The definition of the environment E_R and its rank-χ approximation by the fixed point environment tensors from CTMRG. Example for the region $R = 2 \times 1$. Bottom: The rank-χ approximation of the reduced density matrix of iPEPS. Taking the environment E_R, the indices on the boundary of the region R originating from the tensors T (red in-plane lines), are unfused into their original form (see Fig. 2.6 for their origin). Afterwards, the on-site tensors a, a^* are contracted with the unfused indices completing the construction of the rank-χ approximation of the reduced density matrix.

2.3.3 Reduced density matrices of iPEPS from the Corner Transfer Matrix Renormalization Group

At last, we can connect the results of the CTMRG algorithm with the construction of the reduced density matrices of iPEPS. For a given iPEPS, the CTMRG gives the optimal rank-χ fixed point environment tensors $\{C, T\}$ representing the contraction of the eight distinct parts of the corresponding double-layer network in the thermodynamic limit. Let us isolate any rectangular region R of the double-layer network. Now, we can replace the rest of the double-layer network with the fixed point environment tensors. These tensors, contracted along their environment dimensions form the so called environment E_R of the region R, which we define diagrammatically in the Fig. 2.10. The examples of such environments for the most important regions are

$$E_{1\times1} = \mathrm{Tr}_{env}(CTCTCTCT), \tag{2.41}$$
$$E_{2\times1} = \mathrm{Tr}_{env}(CTTCTCTTCT), \tag{2.42}$$
$$E_{2\times2} = \mathrm{Tr}_{env}(CTTCTTCTTCTT), \tag{2.43}$$
$$E_{L\times1} = \mathrm{Tr}_{env}(CT\ldots TCTCT\ldots TCT), \tag{2.44}$$

where the $\mathrm{Tr}_{env}(.)$ denotes the contraction of all environment indices. From top to bottom: $E_{1\times1}$ is the most compact and holds the information about expectation values of local observables, such as magnetization $m = \sqrt{\sum_{\alpha=x,y,z}\langle S^\alpha\rangle^2}$, $E_{2\times1}$ and

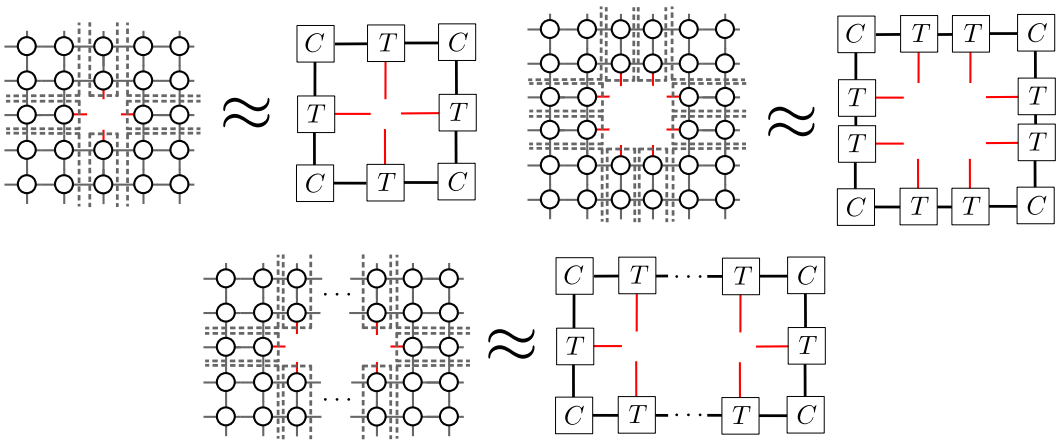


Figure 2.11: Four important examples of the environments for different regions R built from the fixed point tensor $\{C, T\}$ (as with Lego™ bricks). In the clockwise order: $E_{1\times1}$, $E_{2\times1}$, $E_{2\times2}$, and $E_{L\times1}$ for arbitrary length L of the strip. The red lines represent the uncontracted auxiliary indices localized at the boundary ∂R.

$E_{2\times2}$ are minimal environments necessary to evaluate the energy of the nearest and the next-nearest neighbor Hamiltonians respectively, and finally $E_{L\times1}$ is the environment used to compute two-point correlation functions. The illustrative diagrammatic representation of these environments is given in the Fig. 2.11. The final step is to fill these environments E_R with the on-site tensors a, a^* which belong into the enclosed region R, as shown in the Fig. 2.10. And thus we ultimately recover the reduced density matrices ρ_R of iPEPS.

These reduced density matrices of iPEPS we have just constructed are only approximate, more specifically they are the rank-χ approximation $\rho_R(\chi)$ of the true ones. Their accuracy is governed by the environment dimension χ of the environment tensors $\{C, T\}$ used in their construction. Hence, to obtain the definite estimate of observables in the thermodynamic limit for the given iPEPS, the following extrapolation is necessary

$$\langle \mathcal{O} \rangle_{iPEPS} = \langle iPEPS | \, \mathcal{O} \, | iPEPS \rangle = \lim_{\chi \to \infty} \mathrm{Tr}(\rho_R(\chi)\mathcal{O}). \tag{2.45}$$

Therefore, in practice, the CTMRG is performed for increasingly larger environment dimension χ. For each χ, construct the reduced density matrix $\rho_R(\chi)$ and compute the expectation value $\mathcal{O}_\chi = \mathrm{Tr}(\rho_R(\chi)\mathcal{O})$. Continue until the $\mathcal{O}_\chi$ converges to the desired precision. The speed of convergence with χ is directly related to the decay of the spectra of the (enlarged) corner tensors C.

2.4 Example I: spin $S = 2$ AKLT state

We can now demonstrate the validity, performance, and properties of the CTMRG algorithm introduced above. In order to so, we choose certain paradigmatic spin states, which are represented exactly as iPEPS of finite bond dimension D. Since their physical properties were analyzed both analytically and by other numerical approaches, such as Monte Carlo, they provide a vital benchmark for CTMRG.

As the first example, we present the analysis of spin $S = 2$ AKLT state on a square lattice. This valence bond (VB) state is a particular instance of a much broader family of VB states named after Affleck, Kennedy, Lieb, and Tasaki who constructed them in 1988 [21]. In their seminal work, the AKLT quartet showed how to build these $SU(2)$-invariant VB states and their parent Hamiltonians in one dimension, two dimensions or on arbitrary lattice. Also, they proved results about some of their important properties such as the existence of a gap and the exponential decay of all correlations in one dimension. Tensor networks emerge as a natural candidate for a framework to investigate AKLT states due to the compact representation they provide. We construct the iPEPS representation of the $S = 2$ AKLT state and give its parent Hamiltonian in Appendix A.1.

2.4.1 Results from CTMRG: The expectation values of AKLT state

Taking the iPEPS of the spin $S = 2$ AKLT state, we now execute the CTMRG and obtain the fixed point tensors. Afterward, from the environments as the ones given in the Fig. 2.11 we build the reduced density matrices and compute the observables. First, let us remark that the convergence of the CTMRG to the fixed point tensors $\{C, T\}$ is very rapid in spite of the distance between the spectra of the corner matrices used as the convergence measure. It is more stringent than local observable such as energy. The main results are summarized in the Fig. 2.12. Both the energy and the on-site magnetization, computed as $m = \sqrt{\sum_{\alpha=x,y,z}\langle S^\alpha\rangle^2}$, are (nearly) numerically exact zeros. We also compute the dimer-dimer correlation $D_{hh}(r)$ between the two horizontal dimers $D_h(0)$ and $D_h(r\mathbf{x})$ shifted by r sites along the x-direction

$$D_{hh}(r) = \langle D_h(0) D_h((2+r)\mathbf{x})\rangle, \tag{2.46}$$

where the horizontal dimer operator is defined as $D_h(\mathbf{R}) = \mathbf{S}_\mathbf{R} \cdot \mathbf{S}_{\mathbf{R}+\mathbf{x}}$. Interestingly, the dimer-dimer correlation decays exponentially to its finite value $D_{hh}(r \to \infty) = 12.663307$, converged with respect to χ as well. This is a manifestation of the underlying VB structure of the AKLT state.

Partial insight into the performance of CTMRG is given by the spectra of the corner matrices shown in Fig. 2.13. The precision and the speed of convergence is governed directly by the decay of the spectra , see Eq. (2.40). In the case of AKLT, the corner spectra decay very rapidly, with C_{ii} decreasing below 10^{-3} already for $i > 27$. The last remark concerns the degeneracies appearing in the spectra of the corner matrices. For all selected values of χ except $\chi = 36$, the multiplets are

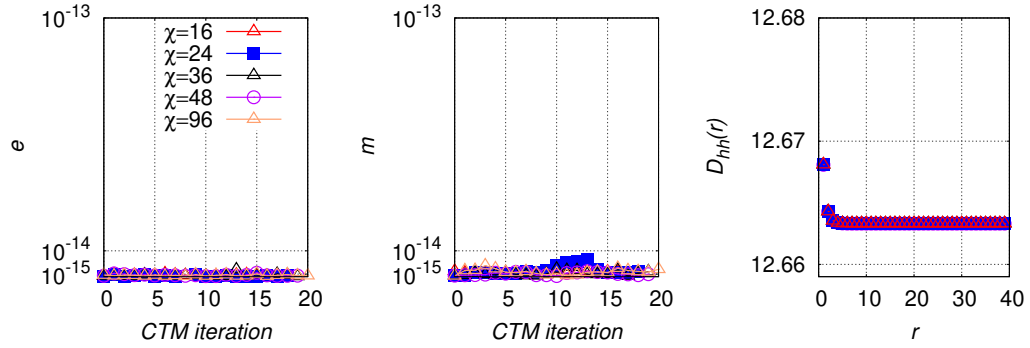


Figure 2.12: CTMRG applied to spin $S = 2$ AKLT state on a square lattice. The environment was considered converged once the distance between the spectra of the corner matrices in two consecutive iterations was smaller than $\epsilon_{CTM} = 10^{-8}$. Left: energy under the parent Hamiltonian computed from the environment E_{2x1} after each iteration. Middle: on-site magnetization computed after each iteration. Right: dimer-dimer correlation of horizontal dimers computed from the environment $E_{L\times 1}$ for increasing length L.

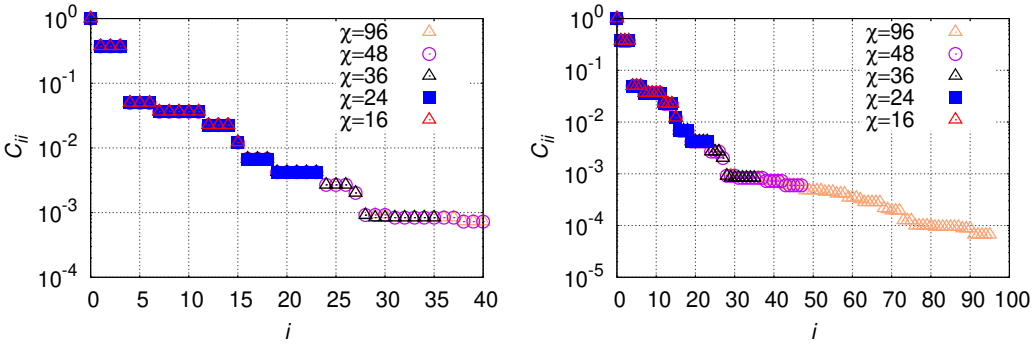


Figure 2.13: Detailed look at the corner spectra of $S = 2$ AKLT state C_{ii} of the converged environment $\{C, T\}$ for an increasing environment dimension χ. Left: First 40 eigenvalues of the corner spectrum. The degeneracies are the reflection of the $SU(2)$ symmetric nature of the AKLT state and they are exact up to the numerical precision, except in the case of $\chi = 36$ (see text). Right: Decay of the spectrum up to the maximal value of $\chi = 96$. The relative importance of the subspace, related to C_{ii}^4, captured by the environment bond dimension over $\chi \gtrsim 75$ is less then 10^{-16} for any observable.

numerically almost exact. They are the direct consequence of the $SU(2)$ invariance of the state. The only exception is the $\chi = 36$ case since this choice of the subspace truncation cuts the multiplet and induces a tiny splitting ($\sim 10^{-5}$).

2.5 Example II: Short range resonating valence bond state

The second example for a demonstration of the CTMRG in action will be the short-range resonating valence bond (RVB) state. Originally put forward by Baskaran, Zou, and Anderson in 1987 as a possible explanation of high-T_c superconductivity in cuprates [22], the RVB states have been investigated in many contexts since then. Taking a lattice in two dimensions, half-filled with electrons, we can imagine pairing all the electrons in VBs (singlets) without imposing any limitation on the distance between the paired electrons. First, consider a single lattice covering, which is a product state of VBs. The RVB states are then constructed from the superposition of a macroscopic number of different lattice coverings.

The short-range RVB (srRVB) is the simplest state from this family, containing only the lattice coverings with the nearest-neighbor valence bonds. The srRVB is then defined as an equal weight superposition of all such coverings. We give the pictorial definition of srRVB on a square lattice in the Fig. 2.14. In the seminal paper introducing the PEPS ansatz, Verstraete and Cirac also gave an elegant representation of srRVB in the terms of C4v symmetric iPEPS with bond dimension $D = 3$ [12]. We give a detailed account of their construction in Appendix A.3.

2.5.1 Results from CTMRG: Observables in the srRVB

The main results for the local observables computed by the CTMRG are shown in the Fig. 2.15. Contrary to the AKLT state, a parent Hamiltonian of the srRVB cannot be written in the terms of nearest- or even next-nearest neighbor spin-spin interaction. Instead, we compute the variational energy of srRVB under the Hamiltonian of antiferromagnetic Heisenberg model. The immediate difference with respect to the AKLT state is the number of iterations necessary to converge energy with the precision 10^{-8}. At least 10^3 iterations were needed for the environment bond dimension as small as $\chi = 108$. And for the largest χ considered, $\chi = 288$, the amount of necessary iterations is nearly 4×10^3. For the state such as srRVB, it is necessary

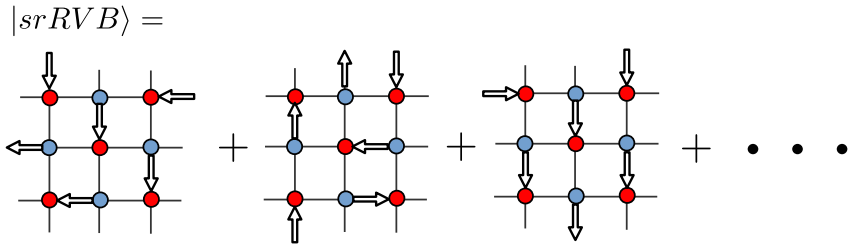


Figure 2.14: Short-range RVB state on a square lattice defined as the equal weight superposition of nearest-neighbor valence bond coverings. The orientation of the singlets is chosen such, that they all point into the sublattice-A sites (red circles).

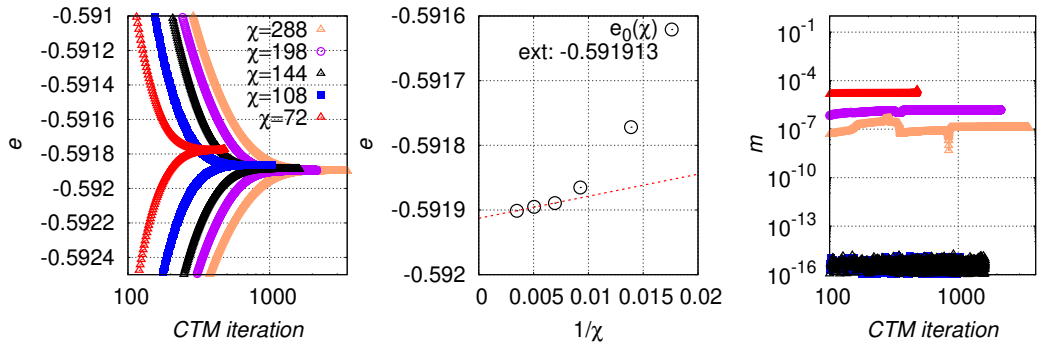


Figure 2.15: Energy and magnetization of srRVB computed from CTMRG. Left: Convergence of the energy computed for the antiferromagnetic Heisenberg model with CTMRG iterations. The energy converges by merging of the two distinct branches. Middle: Extrapolation of energy as a function of $1/\chi$. Right: convergence of the magnetization. Some choices of the environment bond dimension lead to the tiny symmetry breaking.

to perform extrapolation in $1/\chi$ to obtain observables in the thermodynamic limit. The extrapolated value of the energy is $e(\chi \to \infty) = -0.59191(3)$.

While the srRVB is by construction $SU(2)$ invariant, the truncation in CTMRG can induce spurious finite magnetization. Its appearance coincides with the cutting of the multiplet in the corner spectra. The magnitude of this symmetry breaking vanishes with the increasing χ, being only of the order 10^{-7} for the largest environment bond dimension considered. However, if the χ is chosen with the attention to preserving the multiplets in their entirety, such as for $\chi = 108, 144$ the magnetization vanishes exactly.

Next, we evaluate the correlation functions. In particular spin-spin correlations

$$C(r) = \langle \mathbf{S}(0).\mathbf{S}(r\mathbf{x}) \rangle. \tag{2.47}$$

and the connected dimer-dimer correlations defined as

$$D^C_{hh}(r) = \langle D_h(0) D_h((2+r)\mathbf{x}) \rangle - \langle D_h(0) \rangle \langle D_h((2+r)\mathbf{x}) \rangle. \tag{2.48}$$

The results for the correlation functions are summarized in the Fig. 2.16. The spin-spin correlations decay exponentially, except for the cases where the choice of the truncation induces a tiny magnetization leading to the saturation of the spin-spin correlations to a finite value. Connected dimer-dimer correlation functions reveal the critical character of the srRVB. They decay as a power law with an exponent η, only to eventually crossover into the exponential decay regime around the length scale ξ_χ induced by the finite environment bond dimension χ. This behavior is the genuine property of the srRVB since the power-law exponent η grows with the increasing χ. For the largest χ considered the exponent has a value of $\eta(\chi = 288) \approx 1.03$.

Finally, we comment on the spectra of the corner matrices (see Fig. 2.17) which show extremely slow decay compared to the example of the AKLT state. The multiplet structure appears, similarly as in the AKLT case. A careful look at the multiplets reveals that for $\chi = 72, 198$ (and most probably also 288) the truncation cut

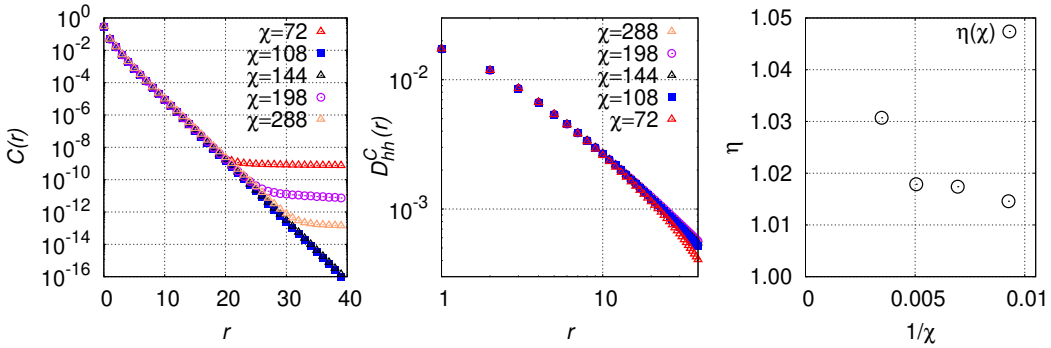


Figure 2.16: Correlation functions of srRVB. Left: spin-spin correlations showing exponential decay. For the cases of $\chi = 72, 198, 288$ the saturation is due to the tiny magnetization induced by the truncation cutting the multiplet of the corner spectra. Middle: connected dimer-dimer correlations with the crossover from algebraic into exponential decay around the length scale ξ_χ. Right: power law exponents for the dimer-dimer correlations extracted by the fit to the Ornstein-Zernike formula $r^{-\eta} exp(-r/\xi_\chi)$.

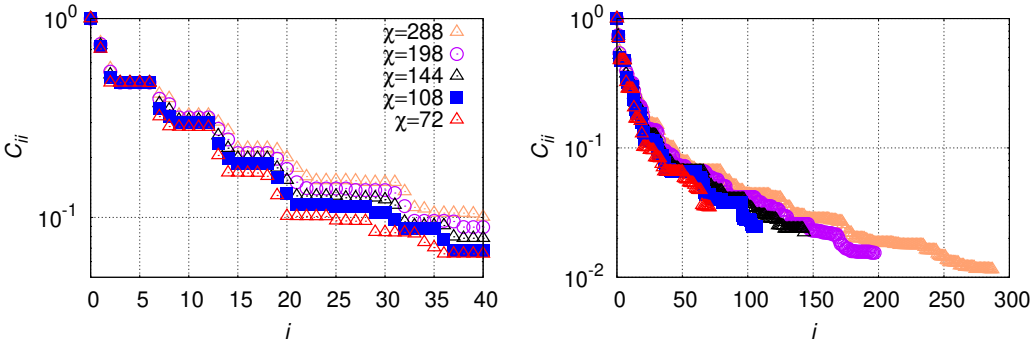


Figure 2.17: Corner spectra C_{ii} of srRVB as a function of the environment bond dimension χ. Left: First 40 eigenvalues of the corner spectrum. For the value of $\chi = 108, 144$ the multiplets are numerically almost exact. Right: The full spectrum decays only algebraically and does not seem to saturate even at small values of i.

the multiplet thus inducing a tiny finite magnetization. The overall slow decay of the corner spectra explains the significantly higher number of CTMRG iterations necessary to converge the observables. Moreover, contrary to the AKLT state which appears to be gapped, all of the individual eigenvalues of the corners grow with the increasing environment bond dimension. This is in stark contrast to the usual behavior of states which satisfy the area law. The srRVB is an example of finite-D iPEPS which belongs into the class of rather exceptional states, displaying critical behavior without violating the area law.

2.6 Contracting infinite tensor networks - Part II

The CTMRG method has been originally developed for symmetric vertex models, where the weights A are invariant under any permutation p of their indices $A_{uldr} = A_{p(uldr)}$. On the other hand, it is not unusual for the ground state to break the spatial symmetry of the Hamiltonian. In the case of quantum magnetism, the clear examples are various VBS ground states, which (partially) break the translational and rotational symmetry of the square lattice. To address such cases, it is necessary to extend the ansatz beyond the single tensor and drop the assumptions on the invariance with respect to permutations of the auxiliary indices. But doing so comes with a price, the CTMRG method must be appropriately generalized as well.

In this section, we will first extend the CTMRG to single-site iPEPS but without any assumptions on the symmetry properties of the on-site tensor a, and consecutively its double-layer version A, under the permutation of its auxiliary indices. Afterward, as the last point, we will consider iPEPS with more tensors within a unit cell (as in the Fig. 2.4) and formulate the final version of the so-called directional CTM algorithm. It will serve us as the principal computational tool for the study of the vast array of iPEPS.

Alternatively, one can use the so-called channel environments [23, 24] which are based on the boundary MPS, inspired by the RTM approach to classical two-dimensional systems.

2.6.1 Directional CTM: Moving beyond the C4v symmetry

The generalization of CTMRG beyond C4v symmetric tensors was given by Orus and Vidal [25] and later perfected by Corboz [26]. The method is known as directional CTM. The basic idea is the same as in the symmetric case. Approximate the contraction of the (infinite) tensor network by a set of finite rank-χ environment tensors. Apply RG procedure to these environment tensors and eventually converge to their fixed point forms. Recover the thermodynamic limit properties in the limit of $\chi \to \infty$. However, since we cannot exploit the symmetry on the auxiliary indices anymore, several complications have to be addressed.

First, we have to consider four different corner tensors $\{C_1, C_2, C_3, C_4, \}$, and four different tensors half-row/-column tensors $\{T_1, T_2, T_3, T_4\}$[1], defined diagrammaticaly in the Fig. 2.18. Growing the system by a single layer of sites, leads to a set of four enlarged corners $\tilde{C}_1, \ldots, \tilde{C}_4$ and four enlarged half-row/-column tensors $\tilde{T}_1, \ldots, \tilde{T}_4$. We would like to identify the most important subspaces of dimension at most χ of these enlarged corners as to truncate them optimally.

In the case of C4v symmetric iPEPS, it was enough to diagonalize the reduced density matrix $\tilde{\rho} = \tilde{C}^4$. Its leading χ eigenvectors provided the optimal projector along the cut (see Fig. 2.9). For generic iPEPS instead, we have to deal with four

[1] contrary to the symmetric case there is no relation between, for example, upper left corner tensor C_1 and lower right C_3. The same is true for any pair of half-row/-column tensors.

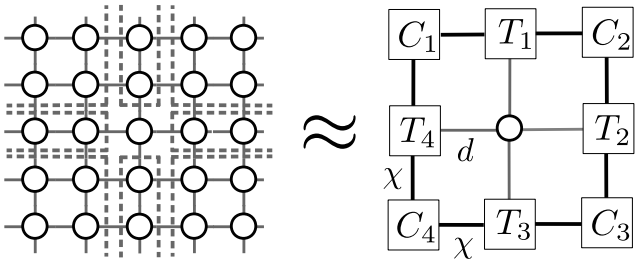


Figure 2.18: Contraction of the tensor network is split and accumulated into eight distinct parts. The contraction of each one of these eight parts is approximated by a finite rank-χ tensor. These tensors are four corner tensors $C_1, \ldots, C_4$ of dimension $\chi \times \chi$ and four half-row/-column tensors $T_1, \ldots, T_4$ of dimensions $\chi \times \chi \times d$. For arbitrary an on-site tensor A, Cs and Ts are assumed to be different with no simple relation between them.

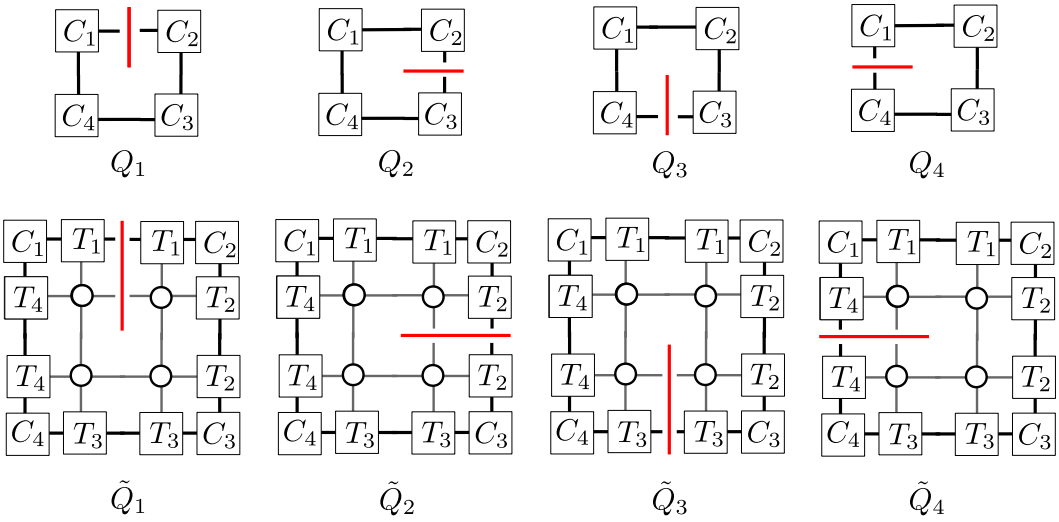


Figure 2.19: Four matrices Q_1, Q_2, Q_3, Q_4 of dimensions $\chi \times \chi$ and their enlarged counterparts $\tilde{Q}_1, \tilde{Q}_2, \tilde{Q}_3, \tilde{Q}_4$ of dimensions $\chi d \times \chi d$ associated to the different cuts of the whole network.

different cuts and hence four different enlarged matrices $\tilde{Q}_1$, ..., $\tilde{Q}_4$, defined diagrammaticaly in the Fig. 2.19. These matrices are not density matrices in a strict sense as they are not symmetric. Nevertheless, they contain the information about the most relevant subspaces along these cuts. However, contrary to the case of symmetric iPEPS, we are facing two conceptual complications. First, since the matrices $\tilde{Q}$ are not symmetric in general they will have different left and right subspaces. Second, the most relevant subspaces for each of the cuts might differ. The last point makes a crucial difference, since for the symmetric iPEPS the projector obtained from the diagonalization of $\tilde{C}$ was optimal for each of the four cuts.

There is no known rigorous approach to find the globally optimal truncation for all the enlarged corners $\tilde{C}_1, \ldots, \tilde{C}_4$ at once. To make the problem tractable, Orus and Vidal introduced a heuristic method, the so-called directional CTM. Instead

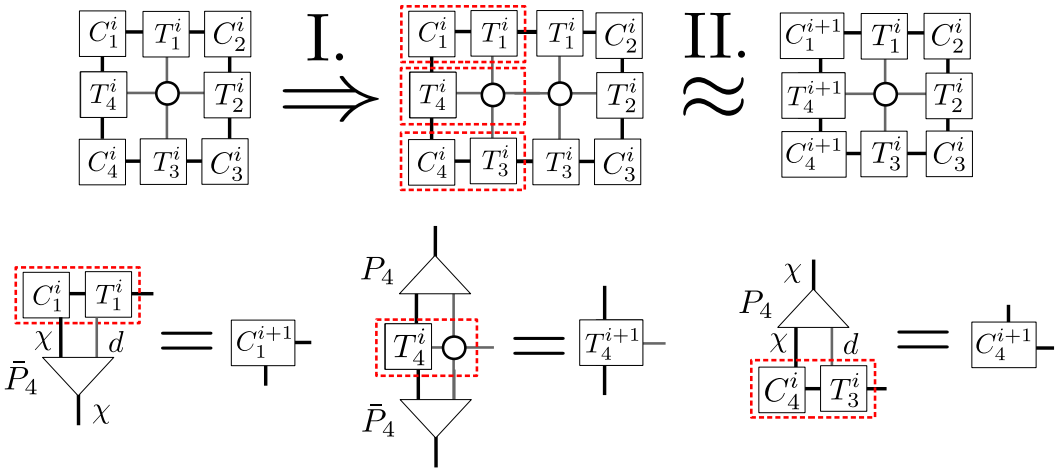


Figure 2.20: Left move of the directional CTM. Top: the network with environment tensors after i-th update is first grown by inserting a column. Enlarged environment tensors, enclosed inside the red dashed rectangles, are identified on left side of the network. Finally, the enlarged tensor are truncated to rank χ. Bottom: definition of the truncated tensors using the projectors $P_4, \bar{P}_4$. The resulting tensors with incremented index are taken to form the new environment.

of growing the system by a single layer at once, proceed by "one lattice direction" $\theta = \{1 \sim \text{up}, 2 \sim \text{right}, 3 \sim \text{down}, 4 \sim \text{left}\}$ at the time. At each RG step grow the system by inserting a single row or column in the chosen direction θ. Afterward, absorb the added tensors into the new enlarged environment tensors. Finally, using a set of projectors $P_\theta, \bar{P}_\theta$ of dimensions $\chi d \times \chi$, related to current direction of the growth, truncate the enlarged tensors back to the original environment dimension χ. The full account of this process, for a single RG step dubbed *move*, along the left direction is illustrated in the Fig. 2.20. The directional CTM continues to iteratively grow and truncate the system alternating the moves for the different directions until the fixed point is reached. While the outline of the procedure is straightforward, the construction of the projectors is crucial for the environment tensors to converge correctly. The caveat lies in the mutual dependence of the truncations along the different cuts. Careless truncation along a single cut might spoil optimality for other cuts, thus preventing the directional CTM from converging either to the correct state or completely.

Currently, the most widely used and empirically well tested recipe for the construction of the projectors is due to Corboz [26] and we adopt it in this work as well. It is based on the so called biorthonormalization developed in the context of the finite-temperature DMRG by Huang [27] where a finite rank approximation to the non-Hermitian transfer matrix is sought. To motivate his approach, lets make a step back and consider an enlarged network split by a horizontal cut as shown in the Fig. 2.21, leaving all the cut indices uncontracted. To execute a left move, we identify the two parts of the split network as R_4 and $\bar{R}_4$. By contracting the indices on the left part of the cut, we recover $R_4\bar{R}_4 = \tilde{Q}_2$, while the contraction of the right

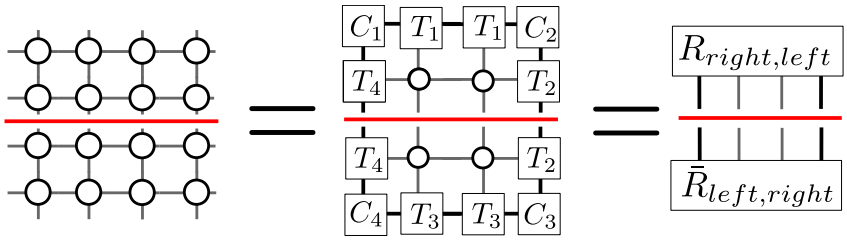


Figure 2.21: Definition of the top and bottom halves $R_4, \bar{R}_4$ of the network with respect to left move ($\theta = 4$). We interpret both R_4 and $\bar{R}_4$ as a matrix of dimensions $\chi d \times \chi d$. In order to move from diagrammatic to the algebraic representation without any ambiguity, we define the matrix R_4 with the indices to the *right* side of the cut as row indices, while the indices on the *left* side of the cut are the column indices. For the matrix $\bar{R}_4$ the role of row and column indices is exactly the opposite. Under this definition, its clear that $R_4\bar{R}_4 = \tilde{Q}_2$ and $\bar{R}_4 R_4 = \tilde{Q}_4$. Bipartitions into $R_\theta, \bar{R}_\theta$ for the other directions θ are defined by keeping the cut (red line) still while rotating the network by a multiple of 90° degrees.

indices gives $\bar{R}_4 R_4 = \tilde{Q}_4$. Now lets ask the following question: what is the best pair of projectors $P_4, \bar{P}_4$, that truncate the dimension along the left cut (one in $\tilde{Q}_4$) from χd down to χ such, that the truncated matrix $\tilde{Q}_2^P = R_4\bar{P}_4 P_4 \bar{R}_4$ is as close to its original $\tilde{Q}_2$ as possible ? The answer is to take

$$\{P_4, \bar{P}_4\} = \underset{rank(R_4\bar{P}_4P_4\bar{R}_4)=\chi}{argmin} |R_4\bar{P}_4P_4\bar{R}_4 - R_4\bar{R}_4|. \tag{2.49}$$

The crucial point here is to disentangle truncations along different cuts. Suppose, that we have found the optimal projectors $P_2, \bar{P}_2$ for the right cut by analyzing the matrix $\tilde{Q}_2$. If the truncation by projectors $P_4, \bar{P}_4$ leads to a $\tilde{Q}_2^P$, which differs substantially from its original $\tilde{Q}_2$, then the projectors $P_2, \bar{P}_2$ found earlier might no longer be the best choice. Unless taken care of, we might never arrive to the environment tensors which are at least reasonably close to the optimal ones by alternating the directions of truncation as outlined in the directional CTM. The solution is found by first taking the SVD of $R_4\bar{R}_4 = \tilde{Q}_2$, as it leads to the best rank χ approximation. Afterwards, we can relate the truncated SVD to projectors $P_4, \bar{P}_4$

$$R_4\bar{R}_4 \overset{SVD}{=} \tilde{U}\tilde{S}\tilde{V}^\dagger \;\rightarrow\; R_4\bar{P}_4P_4\bar{R}_4 = USV^\dagger \;\rightarrow \bar{P}_4P_4 = R_4^{-1}USV^\dagger\bar{R}_4^{-1} \approx \mathbb{1}_{\chi d\times\chi d} \tag{2.50}$$

$$\text{where} \quad U = \tilde{U}_{:,\chi};\; S = \tilde{S}_\chi;\; V = \tilde{V}_{:,\chi}. \tag{2.51}$$

The last step is to express the projectors without the use of the explicit inverses $R_4^{-1}, \bar{R}_4^{-1}$. In order to do so we exploit the following resolution of the identity

$$R_4^{-1}R_4\bar{R}_4\bar{R}_4^{-1} = R_4^{-1}\tilde{U}\tilde{S}\tilde{V}^\dagger\bar{R}_4^{-1} = \bar{R}_4(R_4\bar{R}_4)^{-1}R_4 = \bar{R}_4\tilde{V}\tilde{S}^{-1}\tilde{U}^\dagger R_4. \tag{2.52}$$

Substituting the truncated SVD into the last expression, we define the projectors

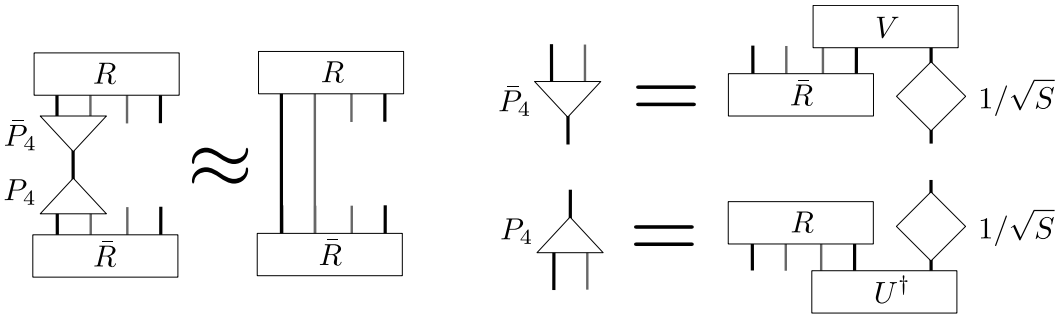


Figure 2.22: Left: The projectors $\bar{P}_4, P_4$ truncate the spaces of the left cut, while minimizing the distance between the projected enlarged matrix $\tilde{Q}_2^P = R_4\bar{P}_4P_4\bar{R}_4$ and its original $\tilde{Q}_4$. Right: projectors $\bar{P}_4, P_4$ are constructed from the result of the truncated SVD of the matrix $\tilde{Q}_2 = R_4\bar{R}_4$. The projectors for other cuts are obtained by the same formula from the appropriate R_θ and $\bar{R}_\theta$ matrices.

$P_4, \bar{P}_4$

$$\bar{R}_4\tilde{V}\tilde{S}^{-1}\tilde{U}^\dagger R_4 \approx (\bar{R}_4 V S^{-1/2})(S^{-1/2}U^\dagger R) = \bar{P}_4 P_4, \tag{2.53}$$

$$\bar{P}_4 = \bar{R}_4 V S^{-1/2}; \quad P_4 = S^{-1/2}U^\dagger R_4. \tag{2.54}$$

This choice of the projectors leads precisely to the condition that the rows of P_4 and the columns of $\bar{P}_4$ form the so called biorthonormal bases $(P_4)_{ik}(\bar{P}_4)_{kj} = \delta_{ij}$. For the illustrative diagrammatic representation of these projectors see Fig. 2.22. Recently, a new construction of the projectors has been proposed [28] that targets directly their final form at the fixed point of CTM, leading to faster convergence for critical iPEPS such as srRVB. Finally, truncating the enlarged tensors with the biorthonormal projectors completes the directional move. By alternating these RG moves, the directional CTM reliably converges to the fixed point. We formulate the complete algorithm and address its details in Appendix B.2.

In conclusion, given a single-site iPEPS, the directional CTM provides the set $E = \{C_1, C_2, C_3, C_4, T_1, T_2, T_3, T_4\}$ of fixed point environment tensors of rank χ that approximate the contraction of the infinite network. Combining them, an environment E_R of an arbitrary rectangular regions R of the iPEPS can be constructed in the same way as for the symmetric iPEPS. We present such environments for the most relevant examples of region R in the Fig. 2.23. Environments along the different directions of the lattice are no longer constrained to be identical.

2.6.2 Directional CTM: Beyond single-site iPEPS

After formulating the CTM for iPEPS without the C4v symmetry, the next and the final necessary step is to extend it to iPEPS with unit cells. The formulation we will present, due to Corboz [26], is in the essence a heuristic, but a well-motivated extension of the directional CTM. The wealth of the empirical evidence of its performance for numerous iPEPS with unit cell sizes ranging from 2×1 up to 12×2 [29] establish it as the state-of-the-art method. We will adhere to the notation introduced by Corboz, as it conveniently translates into the actual implementation.

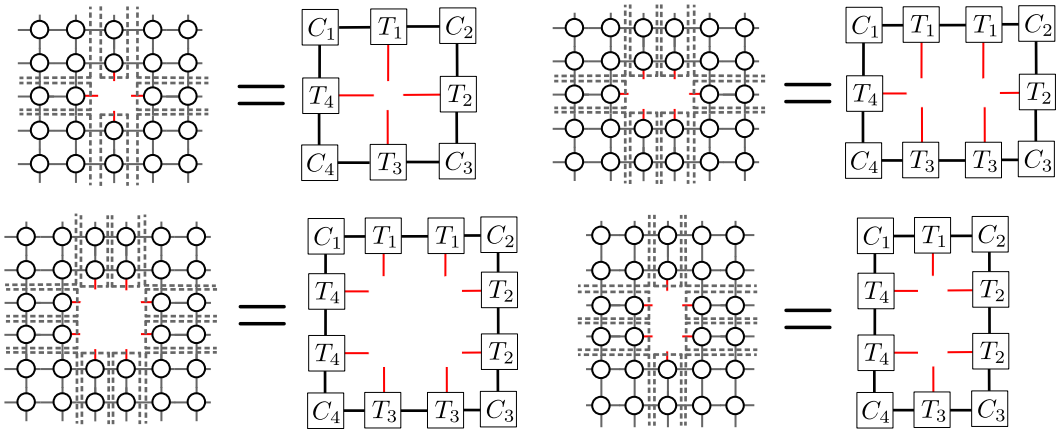


Figure 2.23: Four important examples of the environments for different regions R built from the converged $E = \{C_1, C_2, C_3, C_4, T_1, T_2, T_3, T_4\}$. In the clock-wise order: $E_{1\times1}$, $E_{2\times1}$, $E_{1\times2}$, $E_{2\times2}$. Environments for the correlation functions are obtained by elongating the environment $E_{2\times1}$ ($E_{1\times2}$) along the horizontal (vertical) dimension. The red lines represent the uncontracted auxiliary indices localized at the boundary ∂R.

First, the iPEPS ansatz now assigns a (double-layer) on-site tensor $a^{\mathbf{r}}$ ($A^{\mathbf{r}}$) to each site of the square lattice tilling it with the unit cells. The generalization of the directional CTM starts with making the environment tensors position dependent. For every site $\mathbf{r} = (x, y)$ there is now a set of environment tensors $E^{\mathbf{r}} = \{C_1^{\mathbf{r}}, \ldots, C_4^{\mathbf{r}}, T_1^{\mathbf{r}}, \ldots, T_4^{\mathbf{r}}\}$. Together, they form a complete environment $E = \{E^{\mathbf{r}}\}$. Since the iPEPS is still translationally invariant (at the level of unit cells), the actual number of different environment tensors is given by the number of the unique on-site tensors within a unit cell. For example, for iPEPS with two-site unit cell with the tensors a, b there are just two sets of different environment tensors. Other tensors instrumental in the directional CTM acquire space-dependence in exactly the same manner. Those are specifically the halves of the network $R_\theta^{\mathbf{r}}$ and $\bar{R}_\theta^{\mathbf{r}}$, which are defined with respect to a cut, and the projectors $P_\theta^{\mathbf{r}}$ and $\bar{P}_\theta^{\mathbf{r}}$ that are built from these halves. We give the diagrammatic definition of $R_\theta^{\mathbf{r}}$ and $\bar{R}_\theta^{\mathbf{r}}$ for all four directional moves in the Fig. 2.24.

The directional CTM then proceeds by alternating the up, left, down, and the right moves. But this time, each move composed of inserting the row/column followed by the truncation is done for every non-equivalent lattice site $\mathbf{r}$, that is, not occupied by the identical on-site tensor. For every truncation the projectors $P_\theta^{\mathbf{r}}$ and $\bar{P}_\theta^{\mathbf{r}}$ are constructed according to the Eq. (2.53) from the halves $R_\theta^{\mathbf{r}}$ and $\bar{R}_\theta^{\mathbf{r}}$ of the network. We give the diagrammatic definition of the position-dependent move for the left direction in the Fig 2.25. The complete algorithm and the details concerning its implementation are presented in Appendix B.3.

Once the convergence is reached, the set of fixed point environment tensors E can be used to build rank-χ environments E_R and the subsequent approximation of the reduced density matrices of iPEPS. The environments E_R are constructed

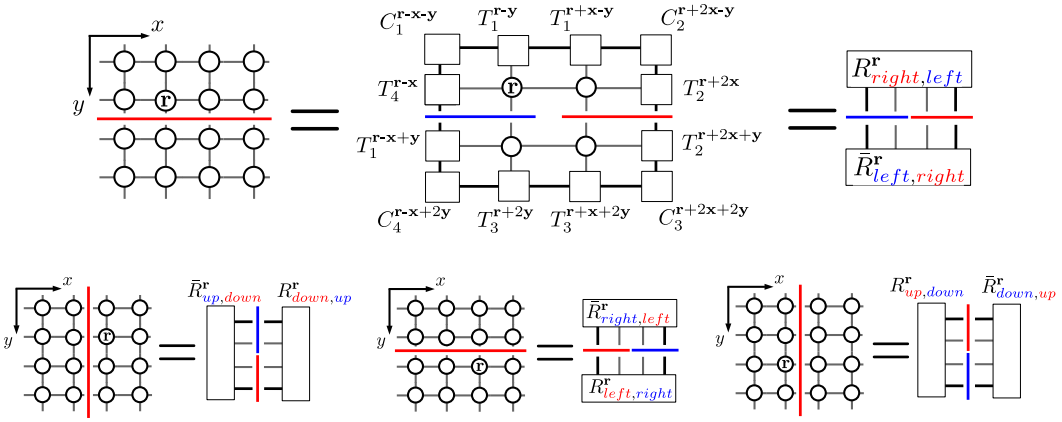


Figure 2.24: Definition of the position dependent halves of the network $R^{\mathbf{r}}_{\theta}$ and $\bar{R}^{\mathbf{r}}_{\theta}$. Top: the definition of the halves $R^{\mathbf{r}}_{4}$ and $\bar{R}^{\mathbf{r}}_{4}$ for the left move with respect to the selected site $\mathbf{r}$. The left (blue) and right (red) part of the horizontal cut is distinguished. The $\mathbf{x}$ and $\mathbf{y}$ are unit vectors in horizontal and vertical directions of the lattice. Bottom: Definition of the halves with respect to the site $\mathbf{r}$ for the remaining three moves, in order up, right, and down.

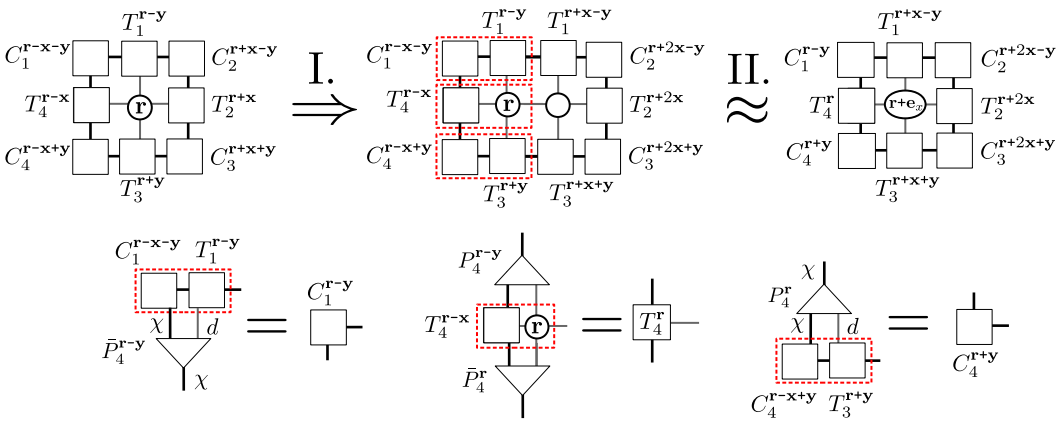


Figure 2.25: Left move of the directional CTM with unit cells. Top: the network with environment tensors around the site $\mathbf{r}$ is grown by a column. Enlarged environment tensors, enclosed inside the red dashed rectangles, are identified on left side of the network. Finally, the enlarged tensor are truncated to rank χ using the projectors $P^{\mathbf{r}}_{4}$, $\bar{P}^{\mathbf{r}}_{4}$, $P^{\mathbf{r-y}}_{4}$, and $\bar{P}^{\mathbf{r-y}}_{4}$. This process is done for all non-equivalent sites $\mathbf{r}$ within the unit cell. Bottom: definition of the truncated environment tensors using the projectors $P^{\mathbf{r}}_{4}$, $\bar{P}^{\mathbf{r}}_{4}$, $P^{\mathbf{r-y}}_{4}$, and $\bar{P}^{\mathbf{r-y}}_{4}$. The resulting environment tensors are shifted by a unit vector $\mathbf{y}$.

analogously to the previous case (see Fig. 2.23), but with the attention to position $\mathbf{r}$ which dictates what environment tensors $C^{\mathbf{r}}_{\theta}$, $T^{\mathbf{r}}_{\theta}$ have to be contracted to create the environment. Allowing for different on-site tensors on different lattice a plethora of new states can be described. This extension of the directional CTM will be the cornerstone of iPEPS approach to several systems in the rest of this work, such as the dimerized Heisenberg model or the $J_1 - J_2$ model.

3

Optimization of iPEPS

CONTENTS

3.1 The overview

The iPEPS tensor networks provide a very efficient parametrization of the ground states of local and gapped Hamiltonians with only a polynomial number of parameters independent of the system size. This is the key property that motivates us to study the quantum magnets through iPEPS-based variational wave functions. However, selecting a variational wave function through physical insight is just the first part of the variational approach. The second, equally important part is the optimization of the variational parameters guided by the energy minimization. The zero-temperature variational principle dictates that the best approximation of the true ground state is achieved by the variational wave function with the parameters tuned to give the lowest possible energy.

Within the domain of tensor network wave functions for local Hamiltonians, the difference between the optimization of (i)MPS in one dimension and (i)PEPS in two dimensions is dramatic. For MPS the question of efficient optimization is to a large degree settled. The state-of-the-art methods, in particular, DMRG and TDVP [30], reliably lead to the best variational states of finite-size systems for MPS with as many as $O(10^8)$ parameters[1]. In contrast, the optimization of iPEPS represents a formidable problem and is an active area of research. The question of what is *the*

[1] the reason why these methods work so well is still not fully understood [2, 9]. In detail, both DMRG and TDVP optimize the system by the so-called alternating least-squares. The hard non-linear problem of optimizing all tensors within the MPS is split into a series of "local" quadratic problems where only a single tensor is optimized at any given time. Yet, there is no guarantee that such an approach leads to the global minimum.

method, leading to the best variational iPEPS for a wide range of local Hamiltonians remains open.

The complexity of optimizing an iPEPS can be traced to three main sources. First, contrary to (i)MPS, the observables cannot be computed exactly and instead an approximation scheme such as directional CTM has to be used. Second, any figure of merit, e.g. energy or the distance between the two iPEPS wave functions, guiding the optimization is a highly non-linear function of the variational parameters. It is notoriously difficult to find the global optimum of such functions. In fact, it is an NP-hard problem. Ultimately, there is a significant redundancy in the iPEPS representation due to the gauge freedom associated with the contracted auxiliary indices. For (i)MPS with open boundaries, a so-called canonical form can be enforced, which fixes the gauge and eliminates the redundancy. In iPEPS, the presence of the loops in the network prevents such gauge-fixing [31].

After setting the scene, we now turn to two main approaches that led to considerable progress. The first one is the imaginary-time evolution, while the second, more recent, is the direct energy minimization based on gradients. We will introduce the imaginary-time evolution in the context of iPEPS and formulate two flavors of the algorithms widely used to optimize iPEPS. Finally, we will turn our attention to more rigorous gradient-based methods. We will introduce the paradigm of differentiable tensor networks, a novel approach to the computation of the gradients. It is based on the well established numerical technique called reverse-mode algorithmic differentiation (AD), also known as the backpropagation in the machine learning community. The AD optimization is readily applicable to all types of iPEPS, with extremely favorable properties, in terms of complexity, robustness and most importantly the resulting variational iPEPS. We consider it the new state-of-the-art for iPEPS optimization.

3.2 Imaginary-time evolution

The imaginary-time evolution is a general method to find the ground state of a many-body system. It is rooted in the observation, that a long time limit $\tau \to \infty$ of imaginary-time evolution $exp(-\tau H)$ under the Hamiltonian H is equivalent to a projector on the ground state. Suppose, that we are concerned with finding the ground state $|0\rangle$, which has the energy E_0. Therefore, the imaginary-time evolved initial state $|\psi_0\rangle$ expanded in the basis of the energy eigenstates reads

$$\begin{aligned} e^{-\tau H}|\psi_0\rangle &= \sum_n |n\rangle e^{-\tau E_n}\langle n|\psi_0\rangle \\ &= |0\rangle e^{-\tau E_0}\langle 0|\psi_0\rangle + \sum_{n>0} |n\rangle e^{-\tau(E_n-E_0)}\langle n|\psi_0\rangle. \end{aligned}$$

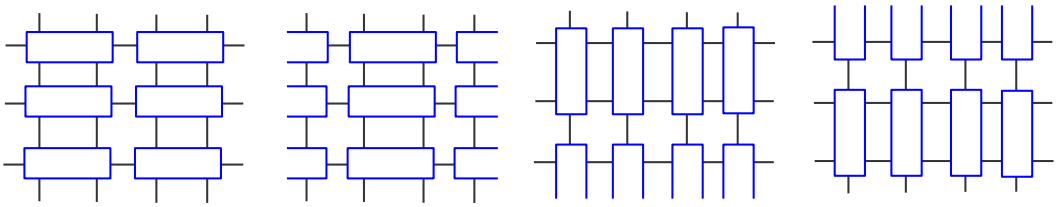


Figure 3.1: A complete set of four partial Hamiltonians H_p for NN Hamiltonian on a square lattice. Each blue-edged rectangle represents a two-site interaction term.

Under the assumption of a finite gap Δ it follows that $E_n - E_0 \geq \Delta$ for all n. Hence, by taking a long time τ limit we recover the ground state

$$\tau \to \infty \ : \ e^{-\tau H}|\psi_0\rangle = |0\rangle\langle 0|\psi_0\rangle + \mathcal{O}(e^{-\tau\Delta}). \tag{3.1}$$

As far as the initial state has a non-zero overlap with the ground state, which is usually the case if we choose $|\psi_0\rangle$ at random, and the gap Δ is finite then the imaginary-time evolution projects out the ground state exponentially fast [1].

3.2.1 Imaginary-time evolution of iPEPS under local Hamiltonians

While the general outline is conceptually simple, the imaginary-time evolution operator

$$U(\tau) := e^{-\tau H}, \tag{3.2}$$

is a complicated object and its size grows exponentially with the system size. However, since we are interested in local Hamiltonians, we can greatly simplify the evolution operator $U(\tau)$ by the means of Suzuki-Trotter decomposition. It is a controlled approximation of the exponential of the sum of non-commuting operators by the product of their exponentials. Take a generic nearest-neighbor (NN) Hamiltonian H. On a regular lattice with coordination z we can split it into a sum of z non-commuting *partial* Hamiltonians H_p

$$H = \sum_{p=1}^{z} H_p, \tag{3.3}$$

$$H_p := \sum_{\langle i,j\rangle(p)} h_{ij}. \tag{3.4}$$

where each H_p represents a mutually commuting subset of NN interaction terms h_{ij}, e.g. spin-spin interaction $\mathbf{S}_i \cdot \mathbf{S}_j$, captured by $\langle i,j\rangle(p)$. In other words, H_p is a certain covering of the lattice by non-overlapping NN interaction terms. See Fig. 3.1 for an example with NN Hamiltonian on a square lattice.

[1] imaginary-time projection is well behaved for finite-size systems, due to the presence of a finite-size gap and hence enjoys the exponential convergence property. The intrinsic gap Δ of the Hamiltonian in the thermodynamic limit is recovered by size-scaling. If the Δ is vanishing, the evolution experiences the so-called critical slowing-down.

Since each individual H_p consists only of mutually commuting terms, taking a small time step ϵ we may approximate the evolution operator $U(\epsilon)$ as

$$U(\epsilon)\,|\psi\rangle = exp(-\epsilon \sum_{p=1}^{z} H_p)\,|\psi\rangle \approx \prod_{p=1}^{z} exp(-\epsilon H_p)\,|\psi\rangle + O(\epsilon^2), \tag{3.5}$$

$$U(2\epsilon)\,|\psi\rangle \approx \prod_{p=1}^{z} exp(-\epsilon H_p) \prod_{p=z}^{1} exp(-\epsilon H_p)\,|\psi\rangle + O(\epsilon^3). \tag{3.6}$$

The first row corresponds to the simple first order Trotter decomposition. The second row instead, gives the second order, or symmetric Trotter decomposition as the sequence of H_p in the second product is reversed. The so-called Trotter error introduced by this approximation can by systematically improved by reducing the time step ϵ. Due to H_p being just a sum of commuting terms, its exponential can be expanded exactly into the product of individual exponentials. Doing so for the first order Trotter decomposition leads to the final form $U_T(\epsilon)$ of the approximate evolution operator

$$U_T(\epsilon)\,|\psi\rangle := \prod_{p=1}^{z} exp(-\epsilon H_p)\,|\psi\rangle = \prod_{p=1}^{z} \prod_{\langle i,j\rangle(p)} u_{ij}\,|\psi\rangle\,, \tag{3.7}$$

where $u_{ij} := exp(-\epsilon h_{ij})$ and they are dubbed Trotter gates. The result for the symmetric decomposition is analogous. Any Trotter gate acts just on two sites and as such their exact form can be readily computed. This is a tractable expression, which can be directly applied in computer simulations.

Now we are in the position to perform an approximate long time evolution $\tau \to \infty$ as to project out the ground state. Since the evolution operator satisfies the following composition property

$$U(\tau_1 + \tau_2) = U(\tau_1)U(\tau_2), \tag{3.8}$$

the long time evolution can be approximated simply by the repeated application of the Trotter decomposed short-time evolution operators

$$U(\tau)|\psi_0\rangle = \prod_{n=1}^{M} U(\epsilon)|\psi_0\rangle \approx \prod_{n=1}^{M} U_T(\epsilon)|\psi_0\rangle + M\mathcal{O}(\epsilon^2), \tag{3.9}$$

where $\epsilon = \tau/M$ is the time step. The total Trotter error scales as $M\mathcal{O}(\epsilon^2) \approx O(\epsilon)$ in the case of first order Trotter decomposition. The second order decomposition improves the error scaling to $O(\epsilon^2)$. The systematic improvement is obtained by increasing the number of discretization steps M. The intricate long-time evolution is thus approximated by a (large) product of Trotter gates

$$U(\tau)|\psi_0\rangle \approx \prod_{n=1}^{M} \left(\prod_{p=1}^{z} \prod_{\langle i,j\rangle(p)} u_{ij} \right) |\psi_0\rangle + \mathcal{O}(\epsilon). \tag{3.10}$$

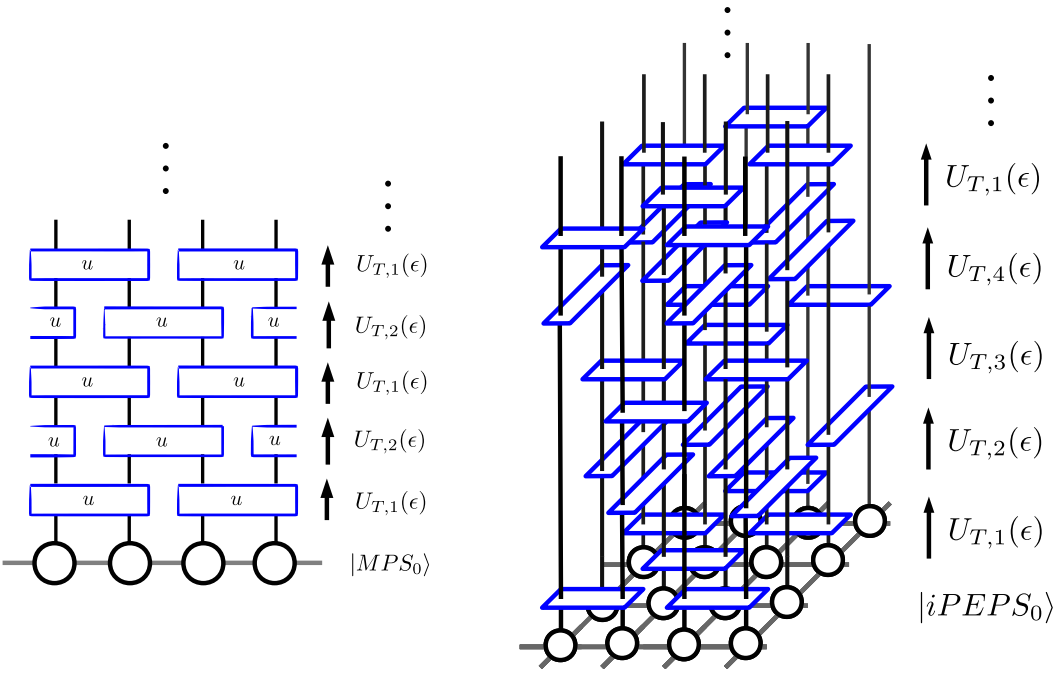


Figure 3.2: Trotter decomposed imaginary-time evolution operator in one and two dimensions. Left: time evolved iMPS $|iMPS_0\rangle$. At step m the partial evolution operator $U_{T,p}$ for $p = 1 + (m \bmod 2)$ is applied. Each $U_{T,p}$ is represented by a layer of non-overlapping two-site Trotter gates u acting on the physical indices (vertical lines). Right: time evolved iPEPS $|iPEPS_0\rangle$. Each horizontal layer of Trotter gates (blue shapes) represents an action of a single partial evolution operator $U_{T,p}$ in similar manner as in the case of iMPS. The Trotter gates act on the physical indices of iPEPS (black vertical lines).

This is a general result, and as far as the Hamiltonian is local the Trotter decomposition is always available. Although for less localized Trotter gates, e.g. acting on three sites, the minimal number of partial Hamiltonians H_p can quickly grow.

The Trotter decomposed imaginary-time evolution operator of Eq. (3.10) can be readily applied to iPEPS. Interpreting the single two-site Trotter gate as a tensor in the spin basis

$$u_{ij} = \sum_{s_i s_j, s'_i s'_j} (u)_{s_i s_j, s'_i s'_j} \, |s_i s_j\rangle \, \langle s'_i s'_j| \,, \tag{3.11}$$

we can express the whole evolution as a tensor network. We give the diagrammatic interpretation of both time-evolved iMPS and iPEPS in Fig. 3.2, The network representing the time-evolved iPEPS extends into a third dimension, which plays the role of the imaginary time. Each discretization step is represented by a set of layers of Trotter gates, depending on the number of partial Hamiltonians H_p. Each layer represents the action of a single partial evolution operator

$$U_{T,p}(\epsilon) := exp(-\epsilon H_p). \tag{3.12}$$

As we stated before, The Trotter error is reduced by increasing the number of

steps M which in effect makes the network taller. The complete contraction of this network gives the ground state, which in principle can be described by iPEPS. However, directly contracting a network of iPEPS time-evolved by M steps results in iPEPS with bond dimension $D \propto O(k^M)$, where k is the internal dimension of the Trotter gate[1]. This exponential growth of entanglement with time cannot be sustained, therefore it is necessary to use a different strategy to extract the ground state from the time-evolved network.

3.2.2 Contracting time-evolved iPEPS

To make the contraction of time-evolved iPEPS tractable, we want to preserve the efficient representation at any point in the evolution. The basic premise is for the ground state to be well described by iPEPS with a (relatively) small bond dimension D. Therefore, the key is to proceed layer by layer while maintaining the optimal fixed D approximation of the so far contracted network. At each step m contracting a single layer of gates, which make up the partial evolution operator $U^{(m)}_{T,p}(\epsilon)$, with the current state $|\psi_m\rangle$ produces a new iPEPS. However, its bond dimension is increased on the bonds where the Trotter gates act. The elementary task at every step is thus to find an iPEPS of bond dimension D that minimizes the distance from the iPEPS time-evolved by the partial evolution operator $U_{T,p}(\epsilon)$. Formally, taking iPEPS $|\psi_m\rangle$ at step m, the optimal approximation of the time-evolved state in a class of iPEPS with bond dimension D is given by

$$|\psi_{m+1}\rangle \;=\; \underset{|\phi\rangle \in iPEPS(D)}{\operatorname{argmin}} \left| |\phi\rangle - U^{(m)}_{T,p}(\epsilon)|\psi_m\rangle \right|. \tag{3.13}$$

The distance $\mathcal{F}$ to be minimized, induced by the usual inner product on the physical Hilbert space, is composed of four terms

$$\mathcal{F}(\phi) := \| \, |\phi\rangle - |U\psi_m\rangle \, \|^2 = \langle\phi|\phi\rangle - \langle U\psi_m|\phi\rangle - \langle\phi|U\psi_m\rangle + const, \tag{3.14}$$

where $|U\psi_m\rangle = U^{(m)}_{T,p}(\epsilon)|\psi_m\rangle$ is a shorthand notation for the time-evolved iPEPS.

This minimization problem is quite complicated. The individual terms making up the distance $\mathcal{F}$ are equivalent to the contracted double-layer tensor networks, and as such they must be evaluated approximately, e.g. by the means of CTM. Moreover, the distance $\mathcal{F}$ is a highly non-linear function of the on-site tensors, which define the iPEPS $|\phi\rangle$. In effect, one has to resort to iterative methods in an attempt for a solution. But generically within every iteration, the evaluation of the distance requires CTM for both the norm $\langle\phi|\phi\rangle$ and the overlaps $\langle U\psi_m|\phi\rangle$ or $\langle\phi|U\psi_m\rangle$, which makes the computational expense[1] for just a single iteration already too high.

---

[1] internal dimension of the Trotter gate $(u)_{s_i s_j, s'_i s'_j}$ is given by the rank of the matrix $u_{(s_i s'_i)(s_j s'_j)}$.

[1] the complexity of CTM is $O(\chi^3 d^3)$, with $d = D^2$ being the dimension of the double-layer tensors. The time-evolved iPEPS has a bond dimension increased by the factor k, the internal dimension of the Trotter gate, which for the operator such as spin-spin interaction $\mathbf{S}_i \cdot \mathbf{S}_j$ is $k = 4$. Therefore, the contraction of the overlap term $\langle U\psi_m|\phi\rangle$ by CTM is roughly 4^3 times slower than for the norm $\langle\phi|\phi\rangle$.

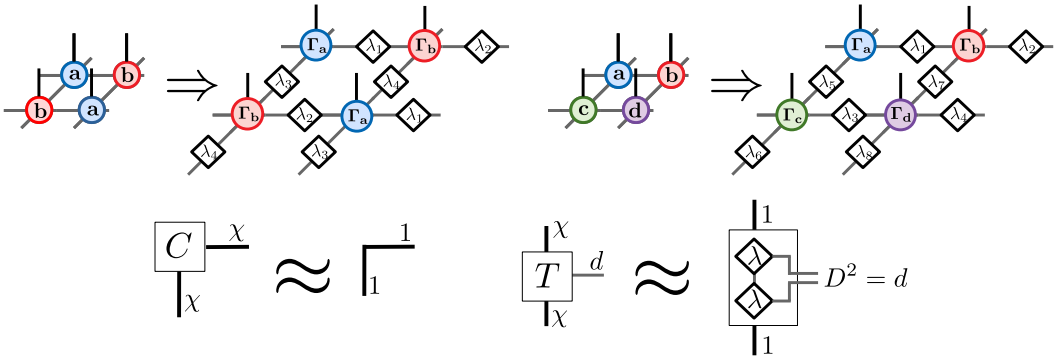


Figure 3.3: iPEPS ansatz for Simple Update Top left: two-site bipartite iPEPS extended by the weights $\lambda_1, \ldots, \lambda_4$ for SU optimization. There are four non-equivalent links between on-site tensors each requiring a diagonal matrix of weights λ. Top right: four-site iPEPS with 2×2 unit cell extended by eight weights $\lambda_1, \ldots, \lambda_8$. Bottom: Simple update ansatz for environment tensors. The environment bond dimension is taken to be one, which makes corner tensor C trivial. The half-row/-column tensors T are built just from the product of weights λ, one from the bottom and one from the top layer of the double-layer network. The weights used to build T depend on its position in the network.

To move forward, the intractable minimization problem of Eq. (3.13)) is instead replaced by a simplified alternative based on the heuristic reasoning. Its solution should provide a good approximation of the imaginary-time evolved iPEPS for sufficiently small time step. We present two such recipes, the so-called Simple and Full update, extensively used to optimize iPEPS tensor networks.

3.2.3 Simple update

The Simple Update (SU) for two-dimensional nearest-neighbor Hamiltonians was introduced by Jiang [32] and later extended to next-nearest neighbour interaction [33] and beyond [34]. In its essence, it is a straightforward generalization of the imaginary-time evolution of iMPS by the so-called time-evolving block decimation (TEBD) [35]. In one dimension, the equivalent minimization of Eq. (3.13) for time-evolved iMPS can be solved efficiently and exactly by imposing the so-called canonical form. The environment of any finite segment of iMPS, composed just from the left and right boundary vectors of dimension D^2, is encoded by the diagonal D-dimensional matrices λ (dubbed weights). These weights are part of the iMPS and allow for fast and exact computation of both the norm and the overlaps of iMPS (for details see [35]). The SU optimization starts by augmenting the iPEPS ansatz with the weights λ on every auxiliary bond. We show examples for two types of iPEPS: two-site bipartite iPEPS and four-site iPEPS in the Fig. 3.3. For other cases, the weights are added analogously. To compute observables by the CTM we can recover the usual iPEPS with no weights by simply absorbing them back into on-site tensors.

Now, we address the two main approximations done by SU in order to simplify

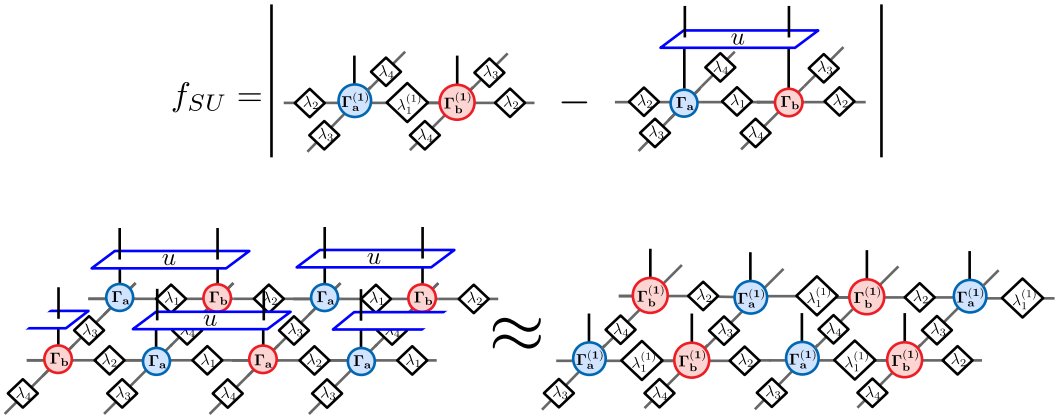


Figure 3.4: Top: The distance f_{SU} minimized during a single step of SU optimization. It is minimized with respect to tensors $\Gamma_a^{(1)}$, $\Gamma_b^{(1)}$ and the weight λ_1 on the link where the gate u acts. Bottom: The SU is performed on two-site bipartite iPEPS, optimized for nearest-neighbor Hamiltonian. At step m, the iPEPS $|\psi_m\rangle$ time-evolved by a single partial evolution operator $U_{T,p}(\epsilon)$ is approximated as a new iPEPS $|\psi_{m+1}\rangle$, formed from the tensors that minimize the distance f_{SU}.

the minimization problem of Eq. (3.13). The first one is done at the level of the partial evolution operator $U_{T,p}(\epsilon)$, which is by itself a product of Trotter gates. For a sufficiently small time step, the $U_{T,p}(\epsilon)$ can be expanded up to first order as

$$U_{T,p}(\epsilon) = \prod_{\langle i,j\rangle(p)} u_{ij}(\epsilon) \approx \sum_{\langle i,j\rangle(p)} u_{ij}(\epsilon) + O(\epsilon^2). \tag{3.15}$$

For translationally invariant Hamiltonians the pairs $\langle i,j\rangle(p)$ define a certain periodic arrangement of the identical Trotter gates, as shown in the Fig. 3.2. At the level of unit cells, the iPEPS is also periodic and hence it is enough to minimize the distance between iPEPS evolved by a single Trotter gate

$$|\psi_{m+1}\rangle = \underset{|\phi\rangle\in iPEPS(D)}{\operatorname{argmin}} \mathcal{F}_{loc}(\phi) := \underset{|\phi\rangle\in iPEPS(D)}{\operatorname{argmin}} \left|\,||\phi\rangle - u(\epsilon)|\psi_m\rangle|\,\right|. \tag{3.16}$$

The on-site tensors found by solving this "local" problem define a new iPEPS which is optimal up to first order in the time step. The second approximation is essential but drastic. The environments of the double-layer networks defining the norm and overlap, that make up the distance $\mathcal{F}_{loc}$ in the above minimization, are approximated just by the weights λ. We give the diagrammatic definition of the equivalent environment tensors C and T in the Fig. 3.3. Such an approximate environment is effectively rank-1 and completely neglects all the correlations between different spatial regions of the system. In fact, it can be shown that with such an environment the SU optimizes the system as though it lives on a Bethe lattice [36].

With the two key approximations in place, the original minimization problem of the Eq. (3.13) is replaced by a significantly easier one, given in the terms of distance f_{SU} which is defined diagrammaticaly in the Fig. 3.4. Inside the distance

f_{SU}, only the small local networks play a role. They consist of the on-site tensors upon which the single gate u acts and the weights connected to these sites. The distance f_{SU} is minimized using SVD since it finds the best rank-D approximation of the time-evolved local network . We give a detailed account of the SU process for both NN and NNN interaction, including the important details relevant for numerical implementation in Appendix B.4. The tensors found by doing so are then replaced in every unit cell of the iPEPS. The SU optimization proceeds by contracting the whole three-dimensional network of time-evolved iPEPS layer by layer alternating the placement of the corresponding gate within the unit cell as per the applied partial evolution operator $U_{T,p}$. After a sufficient number of steps, the evolution reaches a fixed point iPEPS that represents the $\tau \to \infty$ limit of imaginary-time evolution under SU. Once the weights are absorbed into the on-site tensors resulting in a regular iPEPS, the observables can be extracted from the wave function with the CTM. Although very crude, the SU is fast. Taking an iPEPS with bond dimension D its complexity for two-site gate u is just $O(D^3)$, and for a three-site gate, necessary for the NNN interaction, it becomes $O(D^6)$.

Finally, let us comment on the single parameter that controls the SU optimization, which is the time step ϵ of the Trotter decomposition. Even though in principle the Trotter error can be made arbitrarily small by reducing the time step, this fact is not reflected in the results of SU. The truncation error introduced at each SU step by keeping the bond dimension D fixed spoils this feature. As a rule of thumb, smaller time steps improve the final optimized iPEPS but it is not guaranteed to always be the case.

3.2.4 Full update and fast full update

The SU optimization described above is fast, but at rather steep price. Spatial correlations, which are present in the wave function, are basically ignored during the time evolution when solving the local problem of the Eq. (3.16). Hence, the fixed-D iPEPS successively produced by SU are far from being optimal. Instead, they can be very poor approximations of the wave function time-evolved by the partial evolution operator $U_{T,p}$. These shortcomings of SU are even more pronounced for highly-correlated ground states such as spontaneous symmetry-broken phases of which there is plenty to be found in quantum magnets, a prime example being the ground state of Heisenberg antiferromagnet.

Full update (FU) is a more refined approach to imaginary-time evolution of iPEPS, introduced by Jordan et al. [13]. The true minimization problem of the Eq. (3.13) is still beyond reach. The aim of FU is to avoid the severe approximation of the environments that appear in the local problem given in the Eq. (3.16). The distance $\mathcal{F}_{loc}$, between the iPEPS $|\psi_m\rangle$ time-evolved by a single gate u and its fixed-D approximation $|\phi\rangle$ reads

$$\mathcal{F}_{loc}(\phi) = \langle\phi|\phi\rangle - \langle u\psi_m|\phi\rangle - \langle\phi|u\psi_m\rangle + const. \tag{3.17}$$

Since both the norm and overlaps are still highly non-linear functions of the on-site tensors of $|\phi\rangle$ a simplification, although less extreme, is done: the iPEPS $|\phi\rangle$ is taken

to be identical to the initial iPEPS $|\psi_m\rangle$ except the on-site tensors on which the gate u acts. Initially, only the NN Hamiltonians were considered, but later the FU was extended also to systems with NNN interactions requiring three-site gates [37, 38]. The resulting approximate distances, denoted f_{FU}, take the form

$$f_{FU}(\tilde{a},\tilde{b}) = \langle\psi_m(\tilde{a},\tilde{b})|\psi_m(\tilde{a},\tilde{b})\rangle - \langle u\psi_m|\psi_m(\tilde{a},\tilde{b})\rangle - \langle\psi_m(\tilde{a},\tilde{b})|u\psi_m\rangle\,, \tag{3.18}$$

$$f_{FU}(\tilde{a},\tilde{b},\tilde{c}) = \langle\psi_m(\tilde{a},\tilde{b},\tilde{c})|\psi_m(\tilde{a},\tilde{b},\tilde{c})\rangle - \langle u\psi_m|\psi_m(\tilde{a},\tilde{b},\tilde{c})\rangle - \langle\psi_m(\tilde{a},\tilde{b},\tilde{c})|u\psi_m\rangle\,, \tag{3.19}$$

where in the first row the on-site tensors $\{\tilde{a},\tilde{b}\}$ affected by the two-site gate u play the role of variables. The second row is analogous, except the L-shaped gate u is assumed to act on three sites $\{\tilde{a},\tilde{b},\tilde{c}\}$. We illustrate these two important cases of distance f_{FU} covering the NN and NNN Hamiltonians in the Fig. 3.5. Considering the diagrammatic form of f_{FU}, it is natural to interpret each of these minimizations as an "impurity" problem embedded in the (constant) environment of the double-layer iPEPS $|\psi_m\rangle$ that surrounds the affected sites. Once found, the optimal tensors $\{\tilde{a},\tilde{b}\}$ ($\{\tilde{a},\tilde{b},\tilde{c}\}$), are replaced in every unit cell defining a new approximate time-evolved iPEPS $|\psi_{m+1}\rangle$. Effectively, the solution of the local impurity problem is promoted to a global one. The details of solving the above impurity problem by the means of Alternating least-squares are given in Appendix B.5.

This concludes the approximate solution of the original minimization of Eq. (3.13) by FU. Within every FU step the time-evolved iPEPS $|u\psi_m\rangle$ with the increased bond dimension is approximated by the iPEPS $|\psi_{m+1}\rangle$ of a fixed D. Afterwards, the last step before moving on to the next layer of gates is to compute the new fixed point environment tensors E_{m+1} by the CTM. In principle, the CTM should be iterated until convergence to do so. However, that is a very costly process. For sufficiently small time steps the change in the on-site tensors is assumed to be small. Hence only a few iterations of CTM starting from the previous environment E_m should already provide a good enough environment E_{m+1} of the updated iPEPS. Indeed, for time step $\epsilon \to 0$ the on-site tensors remain unchanged and there is no need to recompute the environment[1]. This variation of the Full update, known as Fast full update (FFU) [39, 40] provides a dramatic speed-up by performing only a single CTM step (or just a few) every time the on-site tensors are updated.

With the updated environment, the FU (or FFU) proceeds to contract the next layer in the network of imaginary-time evolution. Eventually, after contracting a sufficient number of layers this approximate evolution reaches the fixed point. Since within FU the environment of the iPEPS is always available, the convergence can be established by directly observing the variational energy. The final iPEPS provides the best variational approximation of the true ground state by the FU optimization. Similarly as in the case of SU, decreasing the time step ϵ generally improves the variational energy but eventually, the truncation error introduced at each step

[1] strictly speaking, this is not true since the FU for vanishing time step leaves the on-site tensors unchanged up to a gauge. Performing at least a step of CTM maintains the compatibility of the environment tensors with the new gauge.

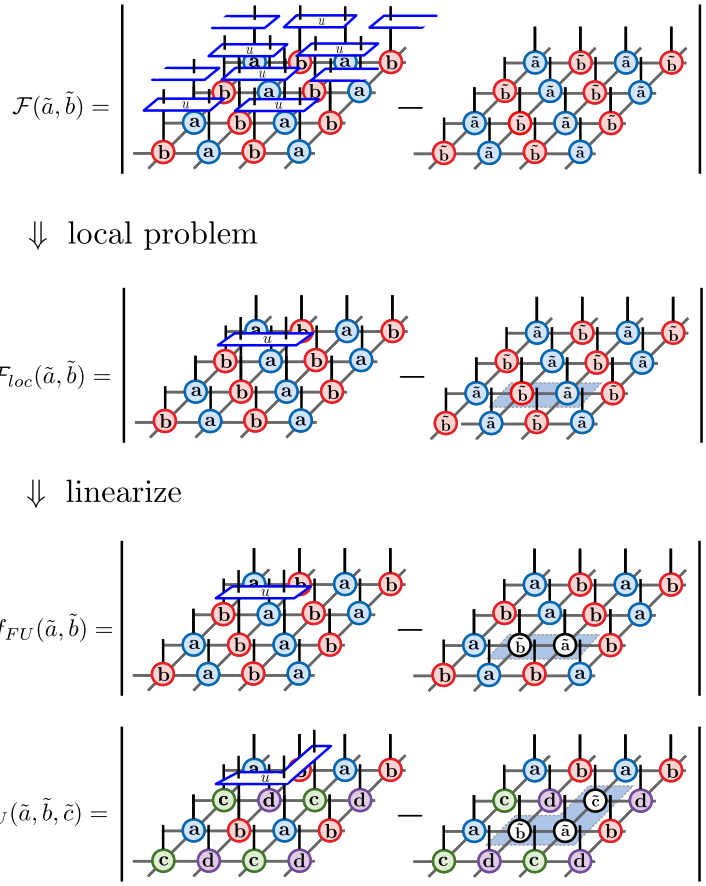


Figure 3.5: Derivation of the approximate distance f_{FU} minimized by the FU for two-site bipartite iPEPS optimized under the action of translationally invariant NN Hamiltonian. Starting from the top, the true distance $\mathcal{F}$ between iPEPS time evolved by $U_{T,p}$ [see Eq. (3.13)] is approximated by the local problem $\mathcal{F}_{loc}$ involving just a single Trotter gate u. Afterward, the problem is linearized by allowing only two tensors directly modified by the gate to vary. The distance between such iPEPS $|\psi_m(\tilde{a},\tilde{b})\rangle$ and the original $|\psi_m\rangle$ evolved by a single two-site gate u is denoted f_{FU}. Bottom: Four-site iPEPS with 2×2 unit cell is optimized by FU with respect to the NNN Hamiltonian. Its Trotter decomposition is given in a terms of L-shaped three-site gates u. The f_{FU} is a distance between the iPEPS $|\psi_m\rangle$ evolved by a single three-site gate u and its modified version now with three different tensors $|\psi_m(\tilde{a},\tilde{b},\tilde{c})\rangle$.

catches up and prevents further improvements. Overall, the FU leads to considerably better variational iPEPS than SU at the expense of complexity, as every (F)FU step invokes at least a single iteration of the costly CTM procedure.

3.3 Gradient based methods

The imaginary-time evolution performed either by SU or the more refined (F)FU ultimately does not provide a good enough variational iPEPS in many challenging cases. The severe and uncontrolled approximations taken within those methods in order to obtain numerically feasible algorithms put a hard limit on their precision. While this might not prevent either of them to describe sufficiently well the gapped and weakly correlated phases, it certainly becomes a problem when addressing the continuous phase transitions.

The direct energy minimization takes a more rigorous approach, treating the optimization of iPEPS as a type of non-linear optimization problem

$$|\Psi\rangle = \underset{|\phi\rangle \in iPEPS(D)}{argmin} \frac{\langle\phi|H|\phi\rangle}{\langle\phi|\phi\rangle}, \tag{3.20}$$

to be solved within a class of iPEPS with a bond dimension D. There are two principal levels of difficulty with such optimization problem. The first and the more ambitious one, is to find the iPEPS $|\Psi\rangle$ which realizes the global minimum of the energy. For a highly non-linear function of many parameters, it is an archetypal case of NP-hard problem. Instead, we will focus on the (relatively) simpler task, which is to find the best local minimum starting from a guess $|\psi_0\rangle$. The standard way to proceed in such a case is to minimize the function using the gradients[1]. Treating all N on-site tensors $\mathbf{a} = \{a, b, \ldots\}$ which define the iPEPS $|\phi(\mathbf{a})\rangle$ as a single point in a large Nd_SD^4-dimensional parameter space, we are interested in computing the gradient of the variational energy

$$\mathbf{g}_0 := \nabla E(\mathbf{a}_0) = \nabla \frac{\langle\phi(\mathbf{a}_0)|H|\phi(\mathbf{a}_0)\rangle}{\langle\phi(\mathbf{a}_0)|\phi(\mathbf{a}_0)\rangle}, \tag{3.21}$$

at arbitrary point $\mathbf{a}_0$. The gradient $\mathbf{g}_0$ completely describes the energy in the immediate vicinity of the point $\mathbf{a}_0$. More precisely, adjusting the tensors in the direction $d\mathbf{a}$ by a small step $\beta \to 0$ the resulting change in the energy is

$$E(\mathbf{a}_0 + \beta d\mathbf{a}) = E(\mathbf{a}_0) + \beta \mathbf{g}_0 . d\mathbf{a} + O(\beta^2). \tag{3.22}$$

Therefore, once the gradient has been computed the tensors can be adjusted accordingly as to decrease the energy. The most straightforward scheme to do so is the so-called *steepest descent* optimization

$$\begin{aligned} \mathbf{a}_{i+1} &= \mathbf{a}_i - \beta_i \mathbf{g}_i, \\ E(\mathbf{a}_{i+1}) &= E(\mathbf{a}_i) - \beta_i |\mathbf{g}_i|^2 + O(\beta_i^2), \end{aligned} \tag{3.23}$$

where at each step i, the tensors are adjusted in the direction opposite to the

[1] in general, the evaluation of the energy is costly as it requires the converged CTM. Thus the alternative gradient-free optimization methods which rely on the fast evaluation of the objective function are ruled out.

gradient by a step β_i. The sufficiently small step guarantees the decrease in the energy. Eventually, the steepest descent arrives at the local minimum of the energy. The corresponding iPEPS $|\Psi\rangle$ is a candidate for the best variational state with bond dimension D. There are more refined ways to reach the local minimum in terms of speed of convergence, all of which rely on the gradients, and we will return to them later.

The fundamental ingredient of the outlined optimization is the gradient $\mathbf{g}$. Its computation for an arbitrary iPEPS $|\phi(\mathbf{a})\rangle$ is the only difficult and conceptually new step. The first breakthrough in this direction came in 2016, by Corboz [41] and Vanderstraeten et al. [24]. Taking a translationally invariant NN Hamiltonian and a one-site iPEPS $|\phi(a)\rangle$, fix all but a single tensor a. We are interested in the gradient of the energy with respect to a^*[2]. The key observation shows, that it can be expressed as an infinite sum of the contributions from all two-site terms, where each individual contribution is naturally interpreted as a tensor network

$$
\begin{aligned}
H &= \sum_{\langle ij \rangle} h_{ij}, \\
\mathbf{g} &= \nabla \langle \phi(a) | H | \phi(a) \rangle \\
&= \sum_{\langle ij \rangle} \nabla \mathrm{Tr}_{aux}(\dots (a^*)^{s_i'} (a^*)^{s_j'} \dots) h_{s_i' s_j', s_i s_j} \mathrm{Tr}_{aux}(\dots a^{s_i} a^{s_j} \dots) \\
&= \sum_{\langle 1j \rangle} \mathrm{Tr}_{aux}(\cancel{(a^*)^{s_1'}} (a^*)^{s_j'} \dots) h_{s_1' s_j', s_1 s_j} \mathrm{Tr}_{aux}(a^{s_1} a^{s_j} \dots) + \\
&\quad \sum_{1 \notin \langle ij \rangle} \mathrm{Tr}_{aux}(\cancel{(a^*)^{s_1'}} \dots (a^*)^{s_i'} (a^*)^{s_j'} \dots) h_{s_i' s_j', s_i s_j} \mathrm{Tr}_{aux}(\dots a^{s_i} a^{s_j} \dots).
\end{aligned}
\tag{3.24}
$$

The final expression has a straightforward diagrammatic interpretation. Each term represents a contracted double layer network including a single two-site operator h but with the fixed tensor a^* missing[3]. Together, they account for the contributions to the gradient from all relative positions of the term h with respect to the fixed tensor a, as illustrated in the Fig. 3.6. In the Refs. [24, 41], an efficient resummation schemes for the infinite sum that makes up the gradient have been proposed. Both approaches are based on the methods to contract the double-layer networks, either by extending the CTM [41] or alternatively using the so called channel environments [24], both with the complexity $O(\chi^3 D^6)$. This progress made the gradient-based optimization of iPEPS a reality, setting the new standard for optimal iPEPS for prototypical models of quantum magnets as spin-1/2 antiferromagnetic Heisenberg model on a square lattice.

Internally, the resummation schemes are tightly connected to the form of the Hamiltonian. In particular, to the form and the extent of the individual interaction terms. Moving on from NN Hamiltonians to NNN Hamiltonians and beyond,

[2] or a. For real iPEPS the distinction is irrelevant. For complex tensor a, the energy is still a real function of a and the gradient is taken only with respect to either real or imaginary part of a.

[3] the operator h used in such computation of the gradient is modified by subtracting its expectation value $h \to h - \langle h \rangle$. This eliminates the (infinite) contribution from the disconnected part of the sum in the Eq. 3.24, see Ref. [24] for details.

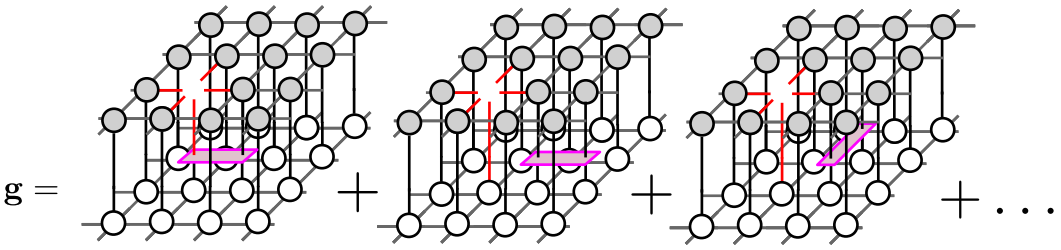


Figure 3.6: Gradient for NN Hamiltonian is given by the (infinite) sum of contributions from all two-site interaction terms h. The red legs represent the unpaired indices of the missing tensor a^* and the violet shapes represent the two-site interaction operator h.

requires a non-trivial extension and also leads to an increase in computational complexity. The pursuit for a more versatile methods to evaluate gradients is thus still on. The most recent development comes from borrowing the state-of-the-art component of the leading machine learning frameworks known as backpropagation or more generally automatic differentiation.

3.3.1 Primer: Automatic Differentiation

The automatic or algorithmic differentiation (AD) represents a set of techniques for highly efficient evaluation of derivatives of complicated functions. The basic premise is that any function, no matter how complex, is ultimately just a composition of (many) elementary functions such as additions, multiplications, or transcendental functions like exponential. Treated as a computer program, possibly with a control flow[1], the AD evaluates the gradients by "differentiating" the program itself by automatic application of the chain rule. For a more comprehensive introduction see the Refs. [42, 43].

Let us analyze in detail the evaluation of the gradient for a scalar function E of N variables

$$E : \mathbb{R}^N \longrightarrow \mathbb{R} \tag{3.25}$$

where E is defined as a composition of several vector functions F^1, F^2, F^3, and final vector to scalar function F^4

$$\begin{aligned} &E : \mathbb{R}^N \xrightarrow{F^1} \mathbb{R}^{M_2} \xrightarrow{F^2} \mathbb{R}^{M_3} \xrightarrow{F^3} \mathbb{R}^{M_4} \xrightarrow{F^4} \mathbb{R} \\ &F^4(F^3(F^2(F^1(\mathbf{x})))) = F^4(F^3(F^2(\mathbf{v}^2))) = F^4(F^3(\mathbf{v}^3)) = F^4(\mathbf{v}^4) =: E \end{aligned} \tag{3.26}$$

Although relatively simple, this model captures all the essentials of the numerical implementation of the variational wave function (or neural network), which given a set of parameters $\mathbf{x}_0$ evaluates the variational energy. In order to optimize the parameters, we are interested in the gradient $\mathbf{g}_0$ of the function E at the point $\mathbf{x}_0$. The conceptually simplest option is to evaluate the gradient using the finite

[1] control flow such as branching statements (`if`) or loops (`for`,...) does not present any problem for AD as far as it is uniquely resolved for any input. This is not the case for symbolic differentiation.

differences

$$(\mathbf{g}_0)_i \approx \frac{E(\mathbf{x}_0 + h\mathbf{e}_i) - E(\mathbf{x}_0)}{h}, \tag{3.27}$$

where $\mathbf{e}_i$ is unit vector in the i-th direction in the parameter space and h is a small step size. There are two main issues. The first is practical, as the finite precision arithmetic introduces numerical errors. The second issue of a major consequence is the complexity. Since the difference has to be evaluated for all N directions, the overall cost becomes $O(N) \times O(E)$, where $O(E)$ is the complexity of the single evaluation of E.

Suppose, that the derivatives of the intermediate functions $F^1, \ldots, F^4$, can be readily computed. This opens another path to the gradient through the direct application of the chain rule. For a given function F^n, its Jacobian reads

$$J^n := \left(\frac{\partial F^n}{\partial \mathbf{v}^n}\right) = \begin{pmatrix} \frac{\partial F_1^n}{\partial (\mathbf{v}^n)_1} & \cdots & \frac{\partial F_1^n}{\partial (\mathbf{v}^n)_{M_n}} \\ \vdots & \vdots & \vdots \\ \frac{\partial F_{M_{n+1}}^n}{\partial (\mathbf{v}^n)_1} & \cdots & \frac{\partial F_{M_{n+1}}^n}{\partial (\mathbf{v}^n)_{M_n}}, \end{pmatrix} \tag{3.28}$$

where $(\mathbf{v}^n)_i$ is the i-th component of a vector $\mathbf{v}^n \in \mathbb{R}^{M_n}$. Evaluating the Jacobian at the given point $\mathbf{v}^n = \mathbf{v}_0^n$ gives simply a $M_{n+1} \times M_n$ matrix of numbers

$$J^n(\mathbf{v}_0^n) = \left(\frac{\partial F^n}{\partial \mathbf{v}^n}\right)\Bigg|_{\mathbf{v}^n = \mathbf{v}_0^n}. \tag{3.29}$$

Taking a point $\mathbf{x}_0$ and a given direction $\mathbf{e}_i$ we can compute both the function E and the corresponding component of the gradient $(\mathbf{g}_0)_i$ as follows

$$\begin{aligned} \mathbf{x}_0 \equiv \mathbf{v}_0^1 \quad &\longrightarrow \mathbf{v}_0^2 = F^1(\mathbf{v}_0^1) \quad &&\longrightarrow \mathbf{v}_0^3 = F^2(\mathbf{v}_0^2) \quad &&\longrightarrow \mathbf{v}_0^4 = F^3(\mathbf{v}_0^3) \\ &\longrightarrow E = F^4(\mathbf{v}_0^4), && && \end{aligned} \tag{3.30}$$

$$\begin{aligned} \mathbf{e}_i \equiv \mathbf{g}_{0,i}^1 \quad &\longrightarrow \mathbf{g}_{0,i}^2 = J^1(\mathbf{v}_0^1) \cdot \mathbf{g}_{0,i}^1 \quad &&\longrightarrow \mathbf{g}_{0,i}^3 = J^2(\mathbf{v}_0^2) \cdot \mathbf{g}_{0,i}^2 \quad &&\longrightarrow \mathbf{g}_{0,i}^4 = J^3(\mathbf{v}_0^3) \cdot \mathbf{g}_{0,i}^3 \\ &\longrightarrow (\mathbf{g}_0)_i = J^4(\mathbf{v}_0^4) \cdot \mathbf{g}_{0,i}^4. && && \end{aligned} \tag{3.31}$$

The i-th element of the gradient $\mathbf{g}_0$ has been evaluated by performing the three matrix-vector products and the final dot product[1] starting from the right

$$(\mathbf{g}_0)_i = J^4(\mathbf{v}_0^4) \cdot (J^3(\mathbf{v}_0^3) \cdot (J^2(\mathbf{v}_0^2) \cdot (J^1(\mathbf{x}_0) \cdot \mathbf{e}_i))). \tag{3.32}$$

This is the so called *forward mode* of AD, as the initial perturbation $\mathbf{e}_i$ is propagated forward through the computation of the function E. The Jacobians J^n, being large-dimensional objects, are never constructed explicitly, it is enough to perform matrix-vector multiplications by computing the individual elements on the fly. But contrary to the finite differences, the components of the gradient are evaluated with full numerical precision. However, in order to obtain the complete gradient, we still need

[1] for $n = 4$, the Jacobian J^4 is in fact a row vector, as the F^4 is a scalar function.

to repeat this procedure N times for each direction in the paramater space. The complexity scaling with the dimensions of the parameter space is too prohibitive.

Now comes the conceptual leap, in what is called the *reverse mode* AD. Suppose, that the function E has been evaluated at the point $\mathbf{x}_0$ and all the intermediate variables $\mathbf{v}_0^4, \ldots, \mathbf{v}_0^2$ are still available. Therefore, the elements of the Jacobians $J^4, \ldots, J^1$ can be computed as well. This allows us to evaluate the product of the Jacobians in the reverse order now

$$1 \cdot J^4(\mathbf{v}_0^4) = \bar{\mathbf{v}}_0^4 \rightarrow \quad \bar{\mathbf{v}}_0^4 \cdot J^3(\mathbf{v}_0^3) = \bar{\mathbf{v}}_0^3 \rightarrow \quad \bar{\mathbf{v}}_0^3 \cdot J^2(\mathbf{v}_0^2) = \bar{\mathbf{v}}_0^2 \rightarrow \quad \bar{\mathbf{v}}_0^2 \cdot J^1(\mathbf{x}_0) = \bar{\mathbf{x}}_0, \tag{3.33}$$

where the row vectors $\bar{\mathbf{v}}_0^4, \ldots, \bar{\mathbf{x}}_0$ are dubbed *adjoints*. The first adjoint in the sequence[1] $\bar{\mathbf{v}}_0^4$, is given just by the elements of the last Jacobian $J^4(\mathbf{v}_0^4)$. Every step is a vector-matrix multiplication, where the adjoint vector multiplies the Jacobian from the left. The final result $\bar{\mathbf{x}}_0$ is actually the gradient itself, which can be simply verified

$$\bar{\mathbf{x}}_0 \cdot \mathbf{e}_i = (((J^4(\mathbf{v}_0^4) \cdot J^3(\mathbf{v}_0^3)) \cdot J^2(\mathbf{v}_0^2)) \cdot J^1(\mathbf{x}_0)) \cdot \mathbf{e}_i = (\mathbf{g}_0)_i. \tag{3.34}$$

And so by performing a single computation of Eq. (3.33), we have obtained the complete gradient $\mathbf{g}_0$ at point $\mathbf{x}_0$. Analogously to the forward mode, the Jacobians do not have to be explicitly constructed. Only the vector-matrix multiplication functions $\bar{F}^n$, dubbed *adjoint functions*, are needed. They are defined as

$$\begin{aligned} &\bar{F}^n : \mathbb{R}^{M_n} \times \mathbb{R}^{M_{n+1}} \xrightarrow{\bar{F}^n} \mathbb{R}^{M_n} \\ &\bar{F}^n(\mathbf{v}^n, \bar{\mathbf{v}}^{n+1}) := \bar{\mathbf{v}}^{n+1} \cdot J^n(\mathbf{v}^n) = \bar{\mathbf{v}}^n. \end{aligned} \tag{3.35}$$

With the help of the adjoint functions we can compactly express the reverse algorithm of the "program" (3.26), which evaluates the gradient

$$\bar{F}^1(\mathbf{x}_0, \bar{F}^2(\mathbf{v}_0^2, \bar{F}^3(\mathbf{v}_0^3, \bar{F}^4(\mathbf{v}_0^4, 1)))) = \bar{\mathbf{x}}_0. \tag{3.36}$$

Not only is the gradient computed with full numerical precision, but this time the crucial difference is in the complexity. It is constant with respect to the number of parameters N. In fact, it can be proven, that the total number of operations needed to compute the gradient of a scalar function via reverse mode AD is at most $6 \times O(E)$ [43]. Let us stress that fast execution of the reverse mode AD relies on the existence of efficient adjoints $\bar{F}$, especially for less elementary functions F such as SVD or diagonalization.

Although we demonstrated reverse AD on the simple example of the Eq. (3.26), the mechanism is easily extended to arbitrary number of intermediate functions F. Moreover, the functions F themselves can be a direct product of several "smaller"

[1] technically, the first adjoint in the sequence is a (co)directional vector $\bar{\mathbf{e}}$ in the space of outputs, but for a scalar function there is just a single direction in the space of outputs so $\bar{\mathbf{e}} = 1$.

functions, for example

$$
\begin{aligned}
&F : \mathbb{R}^N \xrightarrow{F} \mathbb{R}^M \\
&F(\mathbf{x}) := \begin{pmatrix} f(\mathbf{x}_{[1:k']}) \\ g(\mathbf{x}_{[k'+1:l']}) \\ id(\mathbf{x}_{[l'+1:N]}) \end{pmatrix} = \begin{pmatrix} \mathbf{v}_{[1:k]} \\ \mathbf{v}_{[k+1:l]} \\ \mathbf{v}_{[l+1:M]} \end{pmatrix} = \mathbf{v},
\end{aligned}
\tag{3.37}
$$

where f, g are some vector functions and id is an identity $\mathbf{x}_{[l'+1:N]} = \mathbf{v}_{[l+1:M]}$ [the dimensions are equal: $N-(l'+1) = M-(l+1)$]. The Jacobian for such F simplifies into a block-diagonal form and the formula of the Eq. (3.36) for reverse mode computation of the gradient is still valid without any change.

In summary, the reverse mode AD enables fast and numerically precise computation of the gradient for any scalar function E implemented as a program that composes the elementary functions F for which the adjoints $\bar{F}$ are known. First, the function E is evaluated keeping all the intermediate variables. In the second stage, the adjoints are composed in the reverse order, as in the Eq. (3.36), resulting in the gradient $\mathbf{g}$. The price to pay, is the necessity of preserving all the intermediate variables. The reverse mode AD is the integral part of the leading machine learning frameworks such as PyTorch [44] or TensorFlow [45], with backpropagation being just a special case of more general reverse AD. In essence, they provide two basic tools. First, a rich set of tensor algebra operations F together with their adjoints $\bar{F}$. Second, the AD engine which takes a user-supplied program for a function E and evaluates both its value and the gradient automatically by composing in the reversed order the adjoints of the operations F from which the E is built up. As the Liao et al. realized in their recent work [46], there is no fundamental obstacle in applying the machinery of AD to tensor network algorithms.

3.3.2 Differentiating iPEPS algorithms with AD

The properties of the reverse mode AD are very appealing. To employ this technique, we have to first understand whether the process of the computation of the energy for a given iPEPS falls into the category of functions as the studied example $E(\mathbf{x})$ in the Eq. (3.26). The second step is to identify the smallest building blocks of the process and check if their adjoints exist. If these two conditions are met, the gradient of the energy of iPEPS can be evaluated by the reverse AD. We will restrict ourselves to the simple case of one-site C4v symmetric iPEPS and the C4v symmetric NN Hamiltonian $H = \sum_{\langle ij \rangle} h_{ij}$ such as Heisenberg model, and show that it is indeed the case. The following analysis can be straightforwardly extended to directional CTM with unit cells, making it differentiable by AD as well and we will return to this matter later in this section.

Let us recall that for such iPEPS, defined in the terms of a single tensor a, the passage from a to the energy $E(a)$ is composed of two main steps. First, the CTMRG algorithm converges to the rank-χ fixed point environment tensors $\{C, T\}$. Afterwards, in the second step, the reduced density matrix $\rho_{2\times 1}$ is built from the environment tensors $\{C, T\}$ (see the Fig. 2.10) and the expectation value $\langle h \rangle =$

$\mathrm{Tr}(\rho_{2\times1}h)$ is computed. Finally, the on-site energy is given by $E(a) = 2\langle h\rangle$. The above described steps can be formally expressed as the composition of four functions

$$E : \mathbb{R}^{d_s D^4} \xrightarrow{F^{init}} \mathbb{R}^M \xrightarrow{F^{CTMRG}} \mathbb{R}^M \xrightarrow{F^{RDM}} \mathbb{R}^{4\times4} \xrightarrow{F^e} \mathbb{R}. \tag{3.38}$$

First, F^{init} produces the double-layer tensor A and the initial environment tensors $\{C^0, T^0\}$, living in a large space $\mathbb{R}^M$. Next, the F^{CTMRG} performs the whole CTMRG algorithm. In fact, from its detailed description in the Appendix B.1, we can split F^{CTMRG} into N iterations of RG step F^{step}. The N-th step gives the converged environment tensors[1] and the reduced density matrix $\rho_{2\times1}$ is then built in F^{RDM}. Finally, the function F^e evaluates the two-site operator and returns the variational energy. The whole process is thus a sequence

$$E : \mathbb{R}^{d_s D^4} \xrightarrow{F^{init}} \mathbb{R}^M \xrightarrow{F^{step}} \mathbb{R}^M \xrightarrow{F^{step}} \dots \xrightarrow{F^{step}} \mathbb{R}^M \xrightarrow{F^{RDM}} \mathbb{R}^{4\times4} \xrightarrow{F^e} \mathbb{R}, \tag{3.39}$$

where the individual functions, specifying only their arguments and outputs, are

$$\begin{aligned} &F^{init}(a) = (A, C^0, T^0),\\ &F^{step}(A, C^i, T^i) = (A, C^{i+1}, T^{i+1}),\\ &F^{RDM}(a, C^N, T^N) = \rho_{2\times1},\\ &F^e(\rho_{2\times1}) = E. \end{aligned} \tag{3.40}$$

We have expressed the variational energy $E(a)$ as a composition of four different functions, in the form similar to the example from the Eq. (3.26). Still, at this level of granularity, their adjoints (or Jacobians) are not easily deduced. Therefore, we will break down the functions F into their basic constituents.

In general, the tensor network algorithms are ultimately composed of just a few elementary functions. In particular tensor contractions, which consist of reshapes, permutations, and finally matrix multiplication. Out of those, only the matrix multiplication needs to be differentiated as the rest merely reshuffles the elements. In the case of CTMRG, the only other required operation is the diagonalization. With this in mind, we expose the whole scalar function $E(a)$ of the Eq. (3.39) down to such detail in the form of the directed acyclic graph (DAG) shown in the Fig. 3.7. This DAG provides an equivalent definition of the scalar function $E(a)$. The only elementary functions making up the whole DAG are the matrix multiplication and the diagonalization[2]. For both of them, the adjoint functions are known [46, 47]. Therefore, the gradient of the variational energy $E(a)$ can be readily computed by the reverse AD. First, the $E(a)$ is evaluated and all the intermediate variables represented by the vertices of the DAG in the Fig 3.7 are stored. In the second phase, the adjoint functions are evaluated while traversing the DAG in the opposite direction.

[1] the control flow statements in CTMRG as the `while` loop and the `if` convergence condition do not present a problem for the reverse AD. For a given iPEPS and the ϵ_{CTMRG} convergence criterion, there will be just a fixed number N of RG steps executed.

[2] the truncation as well as the normalization of C, T which we have omitted in the DAG are just a special case of matrix multiplication.

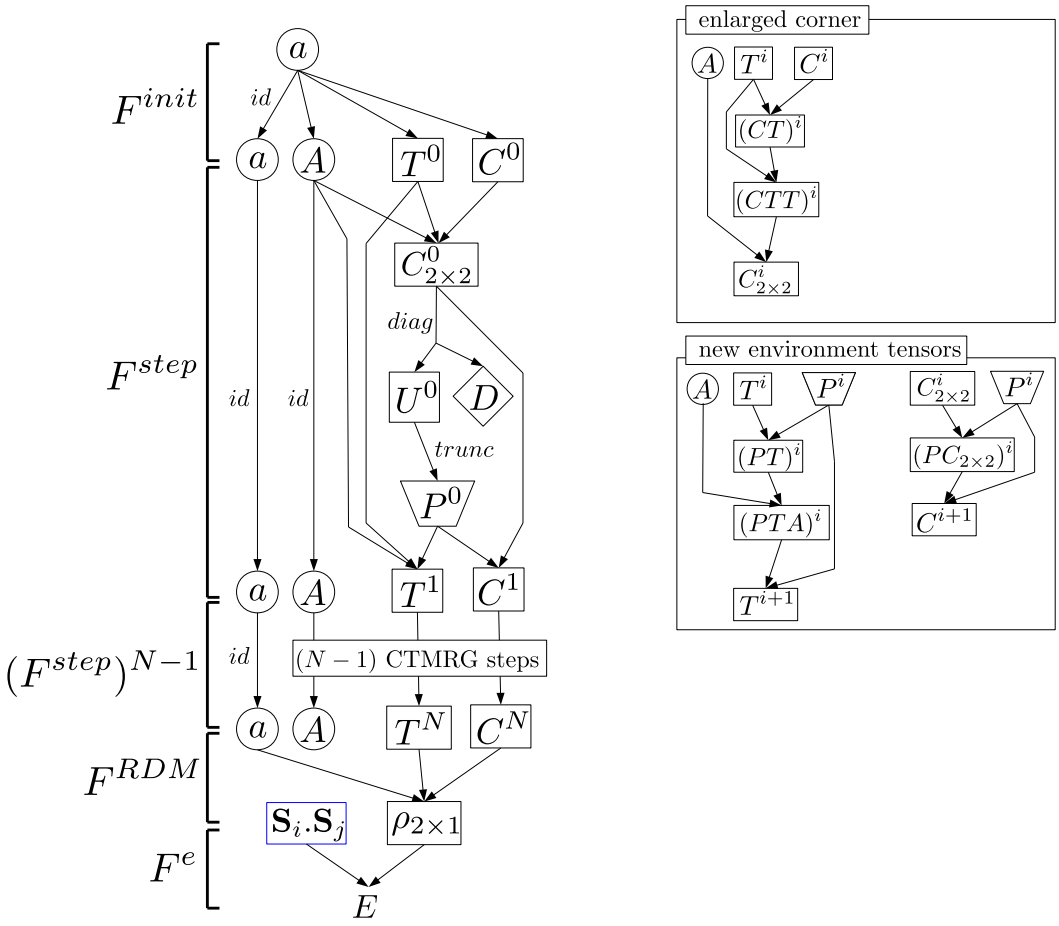


Figure 3.7: Complete DAG representation of the function $E(a)$, which evaluates the variational energy of one-site C4v symmetric iPEPS through CTMRG. The shapes hold the variables and the arrows represent the functions. Every arrow, unless labeled, represents tensor contraction of one or more tensors. Exceptions are the *id* arrows which are just identities, the *trunc* arrow is the truncation, and the *diag* arrow is the symmetric diagonalization returning the eigenvalues and the eigenvectors D, U. The tables on the right describe in detail the sequence of the tensor contractions needed to construct the enlarged corner and the new environment tensors. The reduced density matrix $\rho_{2\times1}$ is constructed analogously, just by a series of tensor contractions from tensors a, C^N, and T^N.

If the same variable v enters several functions F as an argument, its adjoint $\bar{v}$ is a sum of contributions from the adjoint version of all those functions

$$\bar{v} = \sum_{\alpha} \bar{F}^{\alpha}(v, \bar{o}), \tag{3.41}$$

where $\bar{o}$ stands for the adjoints of the outputs. This is the case, for example, of the double-layer tensor A which enters the computation of both the enlarged corner $C_{2\times2}$ and the T tensor in each step of the CTMRG.

Let us conclude this section by a few remarks. First, the reverse AD can be directly applied also to iPEPS with unit cells. Besides more complicated DAG with many more intermediate variables, the only difference is the use of the SVD to construct the projectors used by the directional CTM and the adjoint function of SVD is available [47]. Second, although we have not touched this subject the reverse AD with the necessary adjoints can be extended to the complex-valued tensors [48]. The most recent development in this direction is the adjoint function of the complex SVD [49], thus covering the main iPEPS algorithms completely. Finally, the most important remark concerns the real limits of reverse AD. They are set by the requirement of storing all the intermediate variables during the evaluation of the variational energy. While for the one site C4v iPEPS, the memory requirements are not drastic, for the iPEPS with large unit cells they can grow quickly out of proportions. The need for storing all the intermediate variables can be eliminated by the use of so-called checkpointing [46]. The intermediate variables within the checkpointed section of the DAG are not stored. Only the inputs and the outputs of that section are stored while the rest is recomputed on the fly during the reverse traversal of the DAG. For example, one can checkpoint the whole RG step F^{step}, thus saving just the environment tensors $\{C, T\}$ and avoiding to store large intermediate variables such as the enlarged corner. This increases the number of operations at most by a factor of two. The AD engine as implemented within TensorFlow or PyTorch (and many other implementations) works directly with the DAG representation of the functions to be differentiated. The user-supplied program is turned into DAG either before or during the execution of the variational energy function. Therefore, the checkpointing is readily available.

4

Multiplets in coupled spin ladders

CONTENTS

4.1 Introduction

The iPEPS has been widely used in the recent years to assess the properties of two-dimensional quantum spin systems. Focusing on the spin $S = 1/2$ and $S = 1$ quantum magnets we list just a few of the main results achieved so far. The phase diagram of Shastry-Sutherland model, which provides an effective description of the orthogonal-dimer AFM $SrCu_2(BO_3)_2$ was among the first of such challenging systems addressed by iPEPS [37]. Later, the study of this model in the magnetic field [26] led to an insight explaining the formation of the magnetization plateaus. Large scale iPEPS investigation of spin-1/2 Heisenberg antiferromagnet on kagome lattice [50] provided another piece of evidence supporting the existence of a gapless spin-liquid phase in this system. Recently, complete phase diagrams of spin-1 bilinear-biquadratic Heisenberg models on the square and the triangular lattices were explored [51, 52]. Also the frustrated $J_1 - J_2$ model on a square lattice for both spin $S = 1/2$ and $S = 1$ was addressed by iPEPS [38, 53]. All of these results were obtained by the imaginary-time evolution with SU and FU, still the method presents critical issues when dealing with the unrestricted optimization of tensors.

One of the discrepancies in the context of the imaginary-time optimized iPEPS is the overestimated magnetization of the standard spin-1/2 Heisenberg antiferromagnet on a square lattice, even after the extrapolation to the limit of large bond dimension. The iPEPS primarily targets gapped states with finite correlation length, which is not the case for the symmetry-broken ground state of this model that is gapless with diverging correlation length. Still, the spontaneous symmetry-breaking is a common phenomenon in the quantum magnets and thus it is desirable for iPEPS to provide highly precise description of such phases as well. The recently

developed finite correlation length scaling [54, 55] was used to obtain the accurate results for the Heisenberg antiferromagnet or quantum Ising models from the finite-D iPEPS. However, both mentioned studies relied on the iPEPS optimized by the direct energy minimization.

In this chapter we first benchmark our implementation of imaginary-time optimization of iPEPS for spin-1/2 Heisenberg antiferromagnet with three-site FU instead of the more common two-site FU. Moreover, we show that the iPEPS obtained this way can by analyzed by the finite correlation length scaling leading to a very precise estimates for both the energy and the magnetization in the thermodynamic limit. In principle, this scaling technique, as developed in the Refs. [54, 55], can be performed as far as the system is in the magnetically ordered phase. This raises a tempting prospect of utilizing finite correlation length scaling to improve the iPEPS analysis of continuous transitions between $SU(2)$-symmetric and symmetry-broken phases of quantum magnets. However, there are certain unexpected obstacles that arise in SU and FU optimization when modeling the continuous transition from the ordered to non-magnetic phases. In the second part of this chapter we will connect them to the emergence of a sharp multiplet structure in the "virtual" indices of tensors. In this case, a generic choice of the bond dimension D is not compatible with the multiplets and leads to a symmetry breaking (e.g., generating a finite magnetic order). In addition, varying the initial guess, different final states may be reached, with considerably large deviations in the magnetization value.

This chapter is partly based on the work published as:
[56] J. Hasik and F. Becca, "Optimization of infinite projected entangled pair states: The role of multiplets and their breaking ", Phys. Rev. B **100**, 054429 (2019)

4.2 Benchmark: Isotropic Heisenberg antiferromagnet

The spin-1/2 Heisenberg antiferromagnet on a square lattice is a paradigmatic model of a quantum magnetism and despite its deceptively simple form it exemplifies rich physical behaviour, which can be explained by the theory of spontaneous symmetry breaking. Its Hamiltonian reads

$$H = \sum_{\langle ij \rangle} \mathbf{S}_i \cdot \mathbf{S}_j, \tag{4.1}$$

where $\langle ij \rangle$ denotes the pairs of nearest-neighbour sites. It has a global $SU(2)$ spin symmetry. The pioneering work giving the strong evidence for the symmetry broken ground state was done by P. W. Anderson in 1950's [57] who performed the perturbative spin wave expansion around the classical Néel ground state, estimating the staggered magnetization at $m \approx 0.303$. The finite staggered magnetization is equivalent to the non-vanishing correlation function

$$m^2 = \lim_{r \to \infty} \langle \mathbf{S}(0).\mathbf{S}(r) e^{-i\mathbf{r}.(\pi,\pi)} \rangle \tag{4.2}$$

in the thermodynamic limit. The ground state is thus coined to be long-range ordered. Moreover, the decay of the connected part of the correlation function

$$C_S^C(r) = \langle \mathbf{S}(0).\mathbf{S}(r)e^{-i\mathbf{r}.(\pi,\pi)}\rangle - \langle \mathbf{S}(0)\rangle\langle \mathbf{S}(r)\rangle, \tag{4.3}$$

is algebraic. Spontaneously breaking the $SU(2)$ symmetry gives rise to gapless Goldstone modes. They manifest as spin waves with a transverse polarisation to the vector of the spontaneous magnetization $\mathbf{m} = (\langle S^z\rangle, \langle S^x\rangle, \langle S^y\rangle)$ and are responsible for the algebraic decay of the $C_S^C(r)$. Inspite of the gapless spectrum and the algebraic decay of correlations, the ground state still obeys the Area Law albeit with a subleading logarithmic correction. This is a generic feature of the ground states with spontaneously broken symmetries (see Sec. 2.1.3) and it was verified both by quantum Monte Carlo and the DMRG simulations of antiferromagnetic Heisenberg model on a square lattice [58].

We benchmark the complete implementation of the directional CTM for unit cells and both SU and (F)FU optimization on this system.

4.2.1 The iPEPS ansatz and the computational details

We parametrize the ground state by an iPEPS $|\Psi(a,b,c,d)\rangle$ with a 2×2 unit cell containing four different on-site tensors a, b, c, and d, with auxiliary bond dimension D, see Fig. 4.1(a). Within SU, eight additional weights $\{\lambda_1, \ldots, \lambda_8\}$ are added for each non-equivalent bond between the on-site tensors Γ_a, Γ_b, Γ_c and Γ_d, leading to a state denoted as $|\Phi(\Gamma_a, \Gamma_b, \Gamma_c, \Gamma_d, \lambda_1, \ldots, \lambda_8)\rangle$, see Fig. 4.1(b). For the purpose of computing the environment, we absorb the weights into the tensors, e.g., $a = \Gamma_a\sqrt{\lambda_1}\sqrt{\lambda_2}\sqrt{\lambda_5}\sqrt{\lambda_6}$, thus recovering the original form $|\Psi(a,b,c,d)\rangle$. To evaluate observables for a given state, we employ the directional CTM algorithm extended to unit cells (see Sec. 2.6.2) to construct the environments relative to each site in the 2×2 unit cell. Afterwards, relevant reduced density matrices $\{\rho\}$ can be obtained by combining the environments with on-site tensors. We employ two types of truncated SVD algorithms to speed up the construction of the projectors. The randomized SVD [59] and SVD based on the Arnoldi iteration [60], reducing the computational cost of CTM from standard $O(\chi^3 D^6)$ down to $O(\chi^3 D^4)$.

The Hamiltonian in question, defined in the Eq. (4.1), is the sum of two-site terms $\mathcal{H} = \sum_{R,R'} h_{R,R'}$. For a sufficiently small imaginary time τ, we simplify the evolution operator $U = \exp(-\tau\mathcal{H})$ by the symmetric Trotter-Suzuki decomposition (Eq. (3.6)) into a product of either 2-site or 3-site gates, $u_{R,R'} = \exp(-\tau h_{R,R'})$ and $u_{R,R',R''} = \exp[-\tau(h_{R,R'} + h_{R',R''})]$, respectively. Within the 2×2 unit cell, the three sites R, R', and R'' are arranged to form all possible L-shaped patterns covering the two-dimensional lattice. Even though the Hamiltonian contains just nearest-neighbour interactions, the 3-site gate leads to better optimizations with respect to the case with two sites.

We perform optimizations of the iPEPS using two-site and three-site Trotter gates, both with the SU and the FU. To decrease the computational costs of both SU and FU, we use the scheme with reduced tensors [33, 38]. The in-depth description of these updates is given in the Appendices B.4 and B.5 respectively. Moreover, in

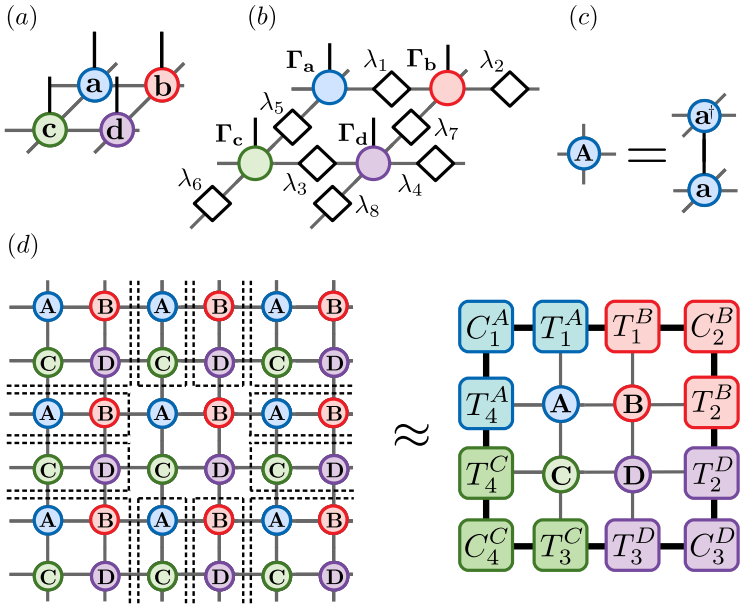


Figure 4.1: (a) Regular iPEPS *ansatz* for 2×2 unit-cell. (b) iPEPS *ansatz* for SU within the same unit cell. (c) Tensors in capitals denote the double-layer version of on-site tensors. For example A is given by the contraction of a and its complex conjugate $a^\dagger$ through the physical index. In this case, all tensors within the ansatz are real. (d) Part of an infinite double-layer tensor network corresponding to the norm of an iPEPS given by ansatz (c). The rank-χ environment of the central 2×2 region is assembled from four sets of environment tensors $E^A = \{C_1^A, \ldots, C_4^A, T_1^A, \ldots, T_4^A\}, E^B, E^C$, and E^D with the environement bond dimension χ. The environments for other regions are constructed analogously.

most of the FU simulations, we do not recompute the environment from scratch after updating the tensors, instead we use the FFU scheme [39], taking only a single iteration of CTM per applied Trotter gate. For the evaluation of the observables the CTM is always interated until convergence. In general, we find that for any fixed time step τ the energy generated by FU optimization reaches a minimum and then starts increasing. To have a well defined convergence criterion for FU we use an adaptive τ. Should the energy increase after the FU iteration, we go back to the previous state and halve the time step, i.e., $\tau \to \tau/2$. The FU optimization is terminated once the time step becomes smaller than 10^{-6}. All the computations have been performed with `pi-peps` [61], a library for running iPEPS simulations built on top of ITensor [62].

4.2.2 Energy, magnetization, and finite correlation length scaling

We perform extensive three-site gate FU optimizations of iPEPS with the bond dimension up to $D = 6$. The best variational iPEPS was obtained for the bond dimension $D = 6$ with the energy $e = -0.669395$ and the average on-site magnetization $m = 0.317930$. In this case, the environment bond dimension for the final

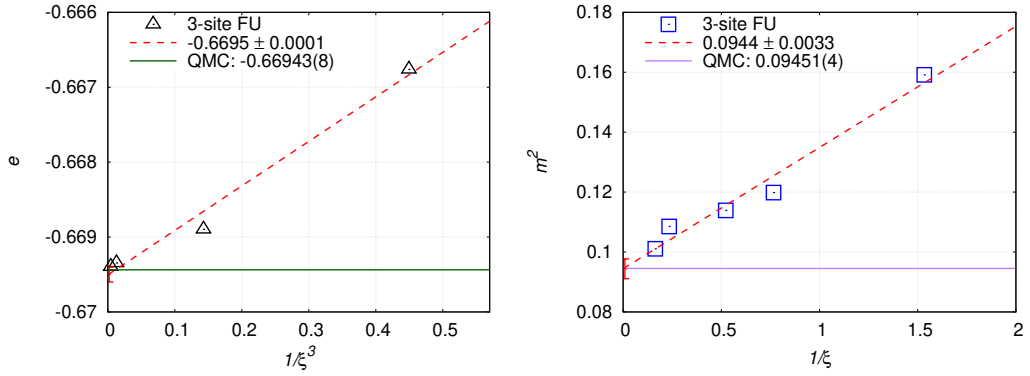


Figure 4.2: Energy and magnetization of the isotropic Heisenberg antiferromagnet from three-site FU with bond dimension up to $D = 6$. Left: the energies from the optimal iPEPS with the bond dimensions $D = \{3, 4, 5, 6\}$ (black triangles) plotted against the expected dependence $1/\xi^3$. The corresponding bond dimensions of the points are in the same order as the correlation length ξ. The linear extrapolation (red dashed line) gives the energy compatible with the QMC estimate. Right: the magnetization squared m^2 for all considered bond dimensions (blue squares) plotted against the $1/\xi$. The extrapolated value (red dashed line) is again in a good agreement with the QMC estimate.

steps of FU was increased up to the value of $\chi = 288$. Altough the iPEPS ansatz does not enforce the C4v symmetry, it is recovered by the optimization. The relative difference between the smallest and the largest value of the NN spin-spin interaction term among the eight non-equivalent bonds in the ansatz is just 0.5%. Both the use of the three-site gates as well as the large χ within (F)FU contributes to the improvement of the optimal iPEPS with respect to the two-site gate results. The best finite-D iPEPS from two-site gate FFU reported in the Ref. [39] with $D = 7$ and $\chi = 70$ has the energy $e = -0.669309$ and the magnetization $m = 0.33490$.

In comparison, currently the most precise results for the order parameter of antiferromagnetic Heisenberg model are based on the quantum Monte Carlo (QMC) for increasingly large lattices coupled with the finite-size scaling [63, 64] leading to the value of $m_{QMC} = 0.30743(1)$. Hence, the $D = 6$ iPEPS still has a relative error of about 3% on the magnetization. In attempt to extrapolate the finite-D observables into the thermodynamic limit and reconcile the results with the QMC we conduct the finite correlation length scaling as in the Ref. [54]. The main idea is to transfer the established finite-size scaling analysis into the context of iPEPS, with the correlation length of iPEPS dictating the finite-size corrections. First, the correlation lengths ξ are extracted from the optimal iPEPS through the analysis of the double-layer transfer matrices as explained in the Appendix C. Afterwards, we plot the energy and the magnetization against their expected dependencies on the correlation length, which are motivated by finite-size scaling for Heisenberg model. The results are summarized in the Fig. 4.2. The extrapolated value of the energy $e_0 = -0.6695 \pm 0.0001$ is in agreement with the reference QMC value $e_{QMC} = -0.66943(8)$ [64]. The improvement is even more pronounced for the magnetization

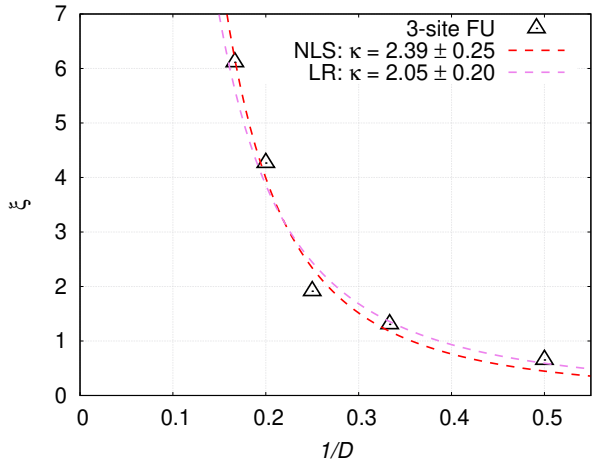


Figure 4.3: Correlation length of the isotropic Heisenberg antiferromagnet from three-site FU with bond dimension up to $D = 6$. The correlation length ξ has a diverging tendency with the increasing bond dimension. The linear regression (LR) and the non-linear least squares fit (NLS) to the power law $\xi(D) = a/D^{-\kappa}$ predict significantly different exponent κ (see text).

$m_0 = 0.307 \pm 0.005$. Hence, we can conclude that the FCLS improves the estimates of the observables in the thermodynamic limit also for the FU optimized iPEPS. However, the error bars on these estimates are relatively large compared to the analysis based on the gradient-optimized iPEPS of the Refs. [54, 55].

The optimized finite-D iPEPS have a finite correlation length. This is in contrast to the diverging correlation length expected for the ground state of Heisenberg antiferromagnet, or spontaneously symmetry-broken phases in general. It is assumed that this is a generic property of the optimized unrestricted iPEPS[1] and it is independent of the optimization method [54, 55]. In fact, the scaling of the correlation length with the bond dimension $\xi(D) \propto 1/D^{-\kappa}$ shown in the Fig. 4.3 gives $\kappa \approx 2$ similar to the results obtained from the gradient-optimization [54]. With this remark we conclude the benchmark of the imaginary-time optimization of iPEPS with unit cells.

4.3 Case study: Heisenberg antiferromagnet on coupled ladders

In order to illustrate the issues facing both the SU and FU optimization, we focus on a particular case of the dimerized $S = 1/2$ Heisenberg antiferromagnet on an

[1] one can construct specific iPEPS using symmetric tensors, which have diverging correlation length [65]. The srRVB of the Sec. 2.5 is an example.

array of coupled two-leg ladders:

$$\mathcal{H} = J \sum_R \mathbf{S}_R \cdot \mathbf{S}_{R+x} + \sum_R J_R \mathbf{S}_R \cdot \mathbf{S}_{R+y}, \tag{4.4}$$

where $\mathbf{S}_R = (S^x_R, S^y_R, S^z_R)$ is the $S = 1/2$ operator on the site $R = (x, y)$ of a square lattice and $J_R = J$ or $J_R = \alpha J$, depending on the parity of y. For $\alpha = 1$ we have an isotropic Heisenberg antiferromagnet on the square lattice and when $\alpha = 0$ the system decouples into an array of two-leg ladders. In the former case, the ground state has Néel antiferromagnetic order and gapless excitations (spin waves); instead, in the latter case, no long-range magnetic order is present and the spectrum is fully gapped. Therefore, a quantum phase transition exists at a finite value of the inter-ladder coupling α [66], as detected by using quantum Monte Carlo methods at zero temperature [67, 68]. In particular, the precise location of the quantum phase transition has been determined with high accuracy, i.e., $\alpha_c = 0.31407(5)$, also suggesting that the critical properties are described by the same universality class as that of the classical three-dimensional Heisenberg model [67].

The phase diagram as well as the nature of the phases is well understood, hence providing a reference for the optimized iPEPS. In particular, the ability of iPEPS to describe correctly non-magnetic ground states with strong local entanglement, notably the existence of nearest-neighbor singlets. This goes beyond the case of the trivial paramagnetic phase that appears in the quantum Ising model [69], which is adiabatically connected to a product state over each lattice site. Indeed, the presence of a local entanglement induces a non-trivial structure in the virtual space, which is easily broken by a blind opimization, thus leading to some symmetry-breaking mechanism, e.g., the generation of magnetic order in the ground-state wave function. We observe that below the critical inter-ladder coupling the vanishing magnetization is recovered only for selected values of bond dimension D, while a blind optimization with a generic D gives rise to a spurious finite magnetization down to the limit of decoupled ladders. This fact has important effects when analyzing a quantum phase transition between magnetically ordered and disordered phases, possibly obscuring its nature and preventing the use of FCLS.

Our calculations show that the paramagnetic phase that is stable for $\alpha < \alpha_c$ is built by tensors having a particular structure that does not fit with a generic value of D. As a consequence, the optimization performed within SU or FU schemes generally leads to a symmetry-broken state with a small but finite magnetization. The correct vanishing magnetization is obtained only for a few selected values of D, making it difficult to perform a scaling for $D \to \infty$. Moreover, for generic D, especially in the paramagnetic region, the effective energy landscape appears very rough, featuring many nearly degenerate states with substantially different magnetizations. Our results strongly suggest that, within iPEPS (or PEPS), it is extremely important to make use of symmetries in the tensors, as suggested in Ref. [70] and developped in Ref. [65].

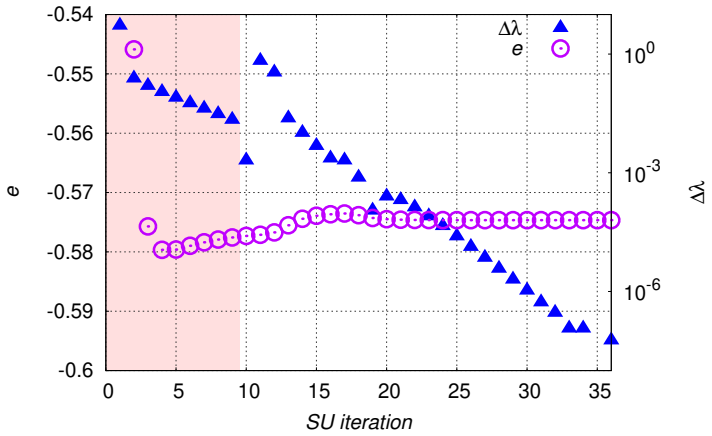


Figure 4.4: Typical SU simulation with $D = 4$ for $\alpha = 0.2$. The initial tensors reproduce a valence-bond solid with singlets on the strong rungs of the ladders; along the optimization 3-site gates are used with time step $\tau = 0.05$. After each SU iteration (corresponding to the application of 32 gates given by symmetric Trotter-Suzuki decomposition) the energy is computed (purple circles - left axis) for the resulting state using CTM with $\chi = 96$. The convergence of SU is tracked by the distance $\Delta\lambda$ of Eq. (4.5) (blue triangles - right axis) of weights $\{\lambda\}$ between consecutive SU iterations. The pink area corresponds to states with vanishing magnetization.

4.3.1 Issues with optimization and the rough landscape

Let us discuss the results of the optimization technique for both the paramagnetic and magnetically ordered phase of the Heisenberg model on coupled two-leg ladders of Eq. (4.4). The details of the ansatz, identical to the one for the Heisenberg model benchmark, as well as the optimization technique have been described in the Sec. 4.2.1. First of all, it is important to emphasize that, within both the SU and FU techniques, the energy can have a non-monotonic behavior along the optimization procedure. Indeed, the minimization problems of Fig. B.2 and 3.5 do not necessarily imply that the energy will decrease at every step of the evolution. In general, after a relatively short transient in which the energy is rapidly decreasing, a minimum is reached and then a slow but inescapable upturn is obtained, no matter how small the imaginary-time discretization is. This is due to the fact that the optimization by SU or FU does not coincide with a true energy minimization [24, 41]. As an example of this behavior, we report in Fig. 4.4 an optimization performed within SU for $D = 4$. Here, we consider $\alpha = 0.2$, initializing the tensors in order to have a valence-bond solid, in which singlets are formed along the strong rungs of the ladders. Tracing the convergence within SU is often done by observing the change in the weights:

$$\Delta\lambda = \sqrt{\sum_{i=1}^{8} \left[\lambda_i^{(m+1)} - \lambda_i^{(m)}\right]^2}, \tag{4.5}$$

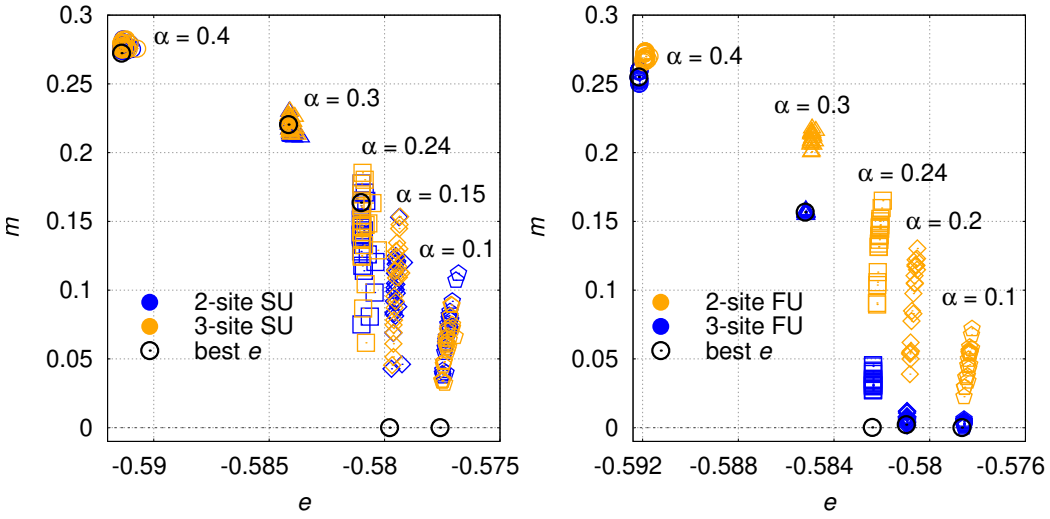


Figure 4.5: Different optimizations by imaginary-time evolution for $D = 4$ with 2-site and 3-site gates for both SU (left) and FU (right) for a range of values α across the transition. The environment bond dimension used in CTM is $\chi = 96$. Black circles denote the best energy states for $D = 4$ obtained throughout all the simulations within SU and FU respectively (see text).

between two subsequent iterations $(m + 1)$ and (m), where the weights are always normalized such that the leading weight $\lambda_1 = 1$. However, while $\Delta\lambda$ eventually decreases down to very small values, signaling a converged SU simulation, the energy (computed with full environment by CTM) shows a non-monotonic behavior with a clear upturn after a few iterations. In this case, a fixed $\tau = 0.05$ is used, in order to emphasize the existence of a minimum in the energy; by using an adaptive time step, as described at the end of Sec. 4.2.1, it would be possible to avoid the rise of the energy, which is otherwise inevitable. Most remarkably, even though the exact ground state has a vanishing magnetization m and the initial state has $m = 0$, a few steps after the minimum, the magnetization becomes finite, spoiling the correct feature of the true ground-state wave function. Hence, in the spirit of the variational principle, we take the lowest-energy state as the "converged" one, for which all the other physical properties (i.e., correlation functions) are computed. At the outset, computing the energy at every iteration of SU seems to betray its purpose, as a crude but fast way to explore the phase diagram; however, if only states given by converged $\Delta\lambda$ are analyzed, the result gives a completely wrong picture with a finite magnetization down to the limit of decoupled ladders.

Now, we would like to stress that both the SU and FU schemes do not always lead to a unique "converged" state, i.e., different starting points may lead to different resulting states. In general, this is not a surprising behavior for nonlinear optimization, a case of both schemes. Yet for the model of Eq. (4.4), whereas the final energy varies in a relatively small range, other quantities might show con-

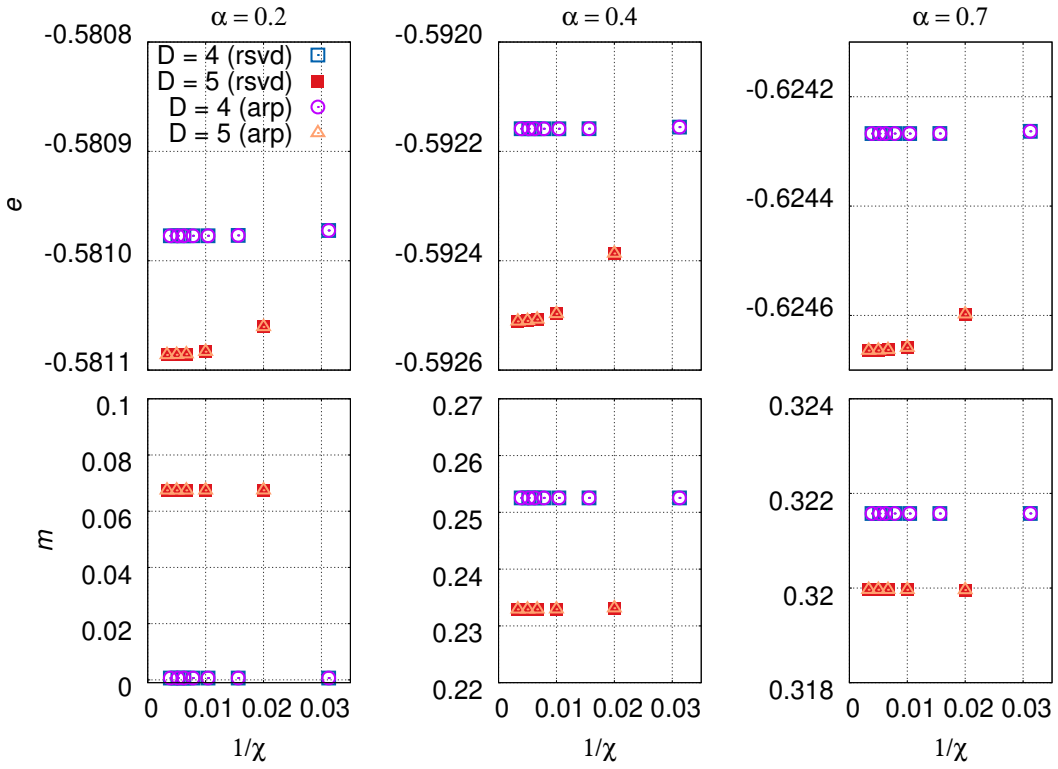


Figure 4.6: Scaling of the energy (top panels) and the magnetization (bottom panels) with the environment bond dimension χ as computed from the best variational states for a set of inter-ladder couplings α. Results from CTM using randomized SVD (rsvd) or Arnoldi iteration (arp) are found to be in a perfect agreement. Magnetization is converged for small values of χ across all the cases. Instead, complete convergence of the energy requires higher values of χ especially close to criticality. Nevertheless, even for $\alpha = 0.4$ (and $D = 5$), a very good accuracy can be achieved for a modest value of $\chi = 100$, with the difference between converged value of energy at $\chi = 300$ and $\chi = 100$ being less than 1.4×10^{-5}.

siderably stronger variation [1]. In addition, we find that imaginary-time evolutions performed with two- or three-site gates may give distinct results, especially within FU. First of all, we briefly discuss the comparison between optimizations done with these two sets of gates for $D = 4$, see Fig. 4.5. Within SU, the difference between two- and three-site gates is small and there is no notable advantage in using three-site gates to perform imaginary-time evolution. Instead, within the FU scheme there is considerable profit in the optimization using three-site gates. Two aspects must be emphasized. The first one is that the distribution of the magnetization is much wider in the paramagnetic phase than in the antiferromagnetic one for both the SU and FU approaches. Indeed, within the magnetically ordered phase, all the final en-

[1] In general, for a variational state with accuracy η on the energy per site, the accuracy on a generic correlation function is $\sqrt{\eta}$. Here, fluctuations in the magnetization are much larger than this estimation.

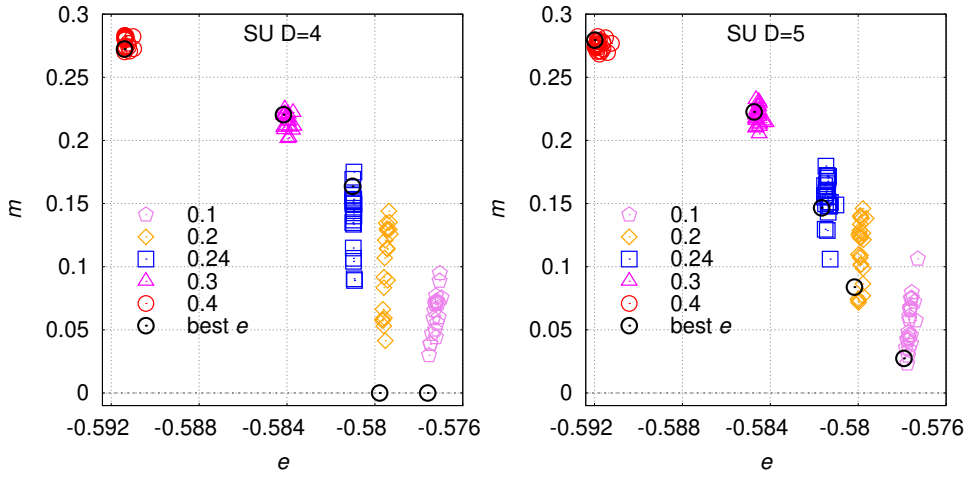


Figure 4.7: Results for the energy and magnetization as obtained within the SU when starting from different (randomized) initial states for different values of the inter-ladder coupling α. The value of the auxiliary bond dimension is $D = 4$ (left) and $D = 5$ (right). Black circles denote the best energy states within SU obtained throughout all the simulations for $D = 4$ and $D = 5$.

ergies and magnetizations are distributed in a very narrow region; most importantly, the fluctuations of m are small with respect to its actual value. By contrast, within the paramagnetic region, it is possible to stabilize states with huge variations in m, still having tiny energy differences (e.g., of the order of $0.0005J$). We would like to emphasize that the presence of large fluctuations in the magnetization persists far away from the critical point, inside the paramagnetic region. This aspect is associated with the nature of the tensor structure of the wave function and is not related to the presence of a quantum phase transition. The second aspect, which is by far much more relevant, is that a generic optimization that starts from random initial tensors does not give the correct vanishing magnetization within the paramagnetic phase. This is particularly true within the SU, while the FU scheme highly improves the quality of the results. Still, paramagnetic states are obtained by requiring both a carefully selected initial state, e.g., valence-bond solids, and a particular value of the auxiliary bond dimension, for example $D = 4$. Let us note that it is possible to induce a finite magnetization by breaking SU(2) symmetry in approximate environment, as observed in Ref. [71]. Its true vanishing value is then recovered only in the limit of $\chi \to \infty$. However, this is not our case, as we show in Fig. 4.6. Indeed, the magnetization is well converged for the values of environment dimension χ used. For the rest of the analysis, we will use three-site gates since, in general, they give better energies with respect to the case with two sites. We now discuss the most important issue of this work, namely, the fact that a paramagnetic state with zero magnetization can be obtained only for selected values of the bond dimension D, i.e., the ones that do not break the multiplet structure of the tensors. In Fig. 4.7, we show the outcomes of several SU optimizations for different inter-ladder couplings α. The cases with $D = 4$ and 5 are presented (the cases with $D = 2$ and 3 give

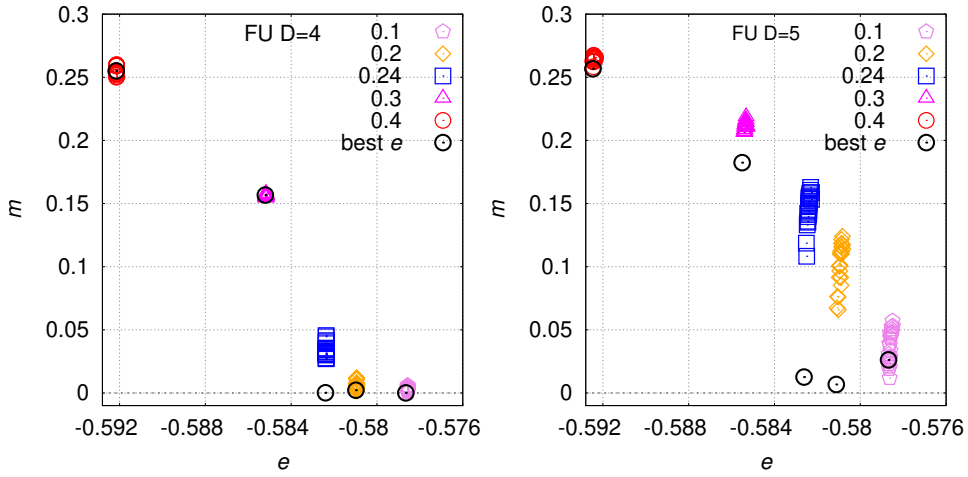


Figure 4.8: The same as in Fig. 4.7 but obtained within FU. Here the value of χ that determines the dimension of environment tensors is 96 for $D = 4$ and 100 for $D = 5$. Black circles denote the best energy states within FU obtained throughout all the simulations for $D = 4$ and $D = 5$ iPEPS respectively (see text).

completely unphysical results, with large values of m down to $\alpha = 0$, and therefore will not be discussed here). The results are qualitatively similar when considering the FU technique (see Fig. 4.8).

4.3.2 Appearance of multiplets in the virtual space

The most remarkable aspect is that the exact result $m = 0$ can be obtained only for a few selected values of the bond dimension, i.e., $D = 4$ and $D = 8$, while for all the other values of D the best energy states break the spin SU(2) symmetry and develop finite magnetization. As a consequence, a smooth extrapolation of the magnetization with increasing bond dimension D is not always possible, while the energy usually has a very regular behavior (see Fig. 4.9). In order to highlight this feature, we compute the spectrum of the singular values of the matrix that is obtained by contracting the index connecting two neighboring tensors and glueing together all the remaining ones of each tensor, thus creating a $2D^3 \times 2D^3$ matrix (where the factor 2 comes from the physical index). This spectrum is particularly simple within the antiferromagnetic phase, where all singular values are not degenerate (see Fig. 4.9). In this case, a given choice of D never spoils the structure of the spectrum, and no appreciable differences are seen in any correlation function. By contrast, within the paramagnetic phase a very peculiar multiplet structure appears, which is preserved only for selected values of D. Indeed, the spectrum shows degeneracies that depend upon the bond: starting from the largest values, we have $1, 3, 3, 1, \ldots$ (when contracting along the weak bonds with αJ and the strong horizontal bonds along the ladder with J) and $2, 2, 4, \ldots$ (when contracting along the strong vertical bonds of the ladder with J). Therefore, it is clear that only particular values of D can accommodate these multiplet structures (e.g., $D = 4$ and

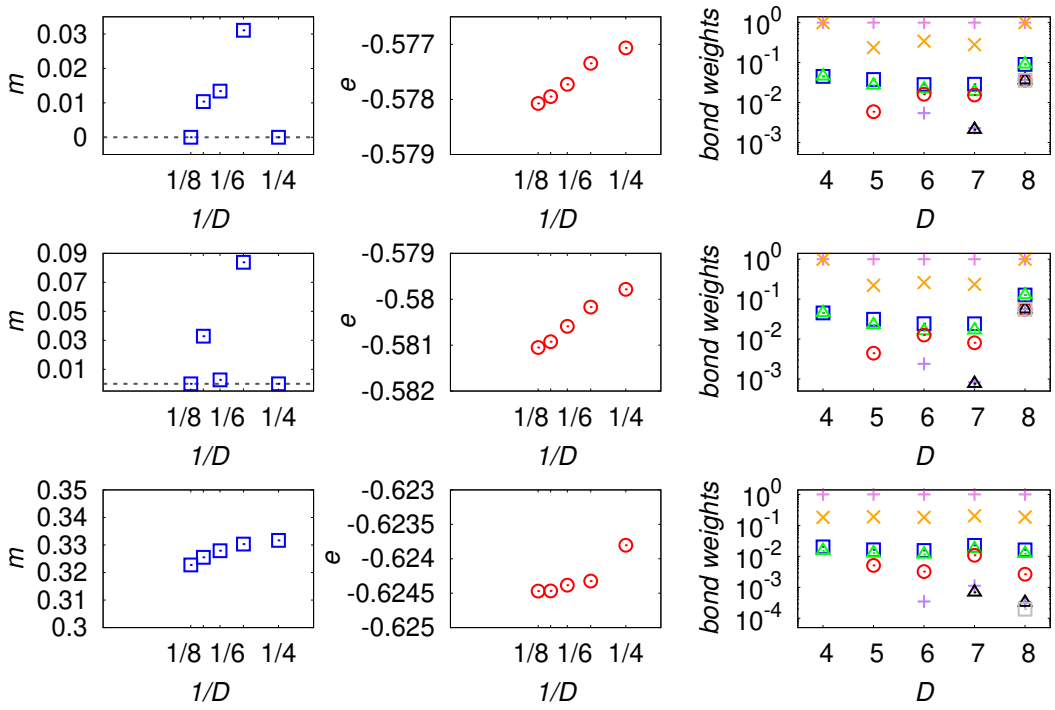


Figure 4.9: Magnetization, energy, and values of the weights along rungs of the ladders (see text) obtained by SU in the paramagnetic phase with $\alpha = 0.05$ (upper panel) and $\alpha = 0.2$ (middle panel), and in the magnetically ordered phase with $\alpha = 0.7$ (lower panel).

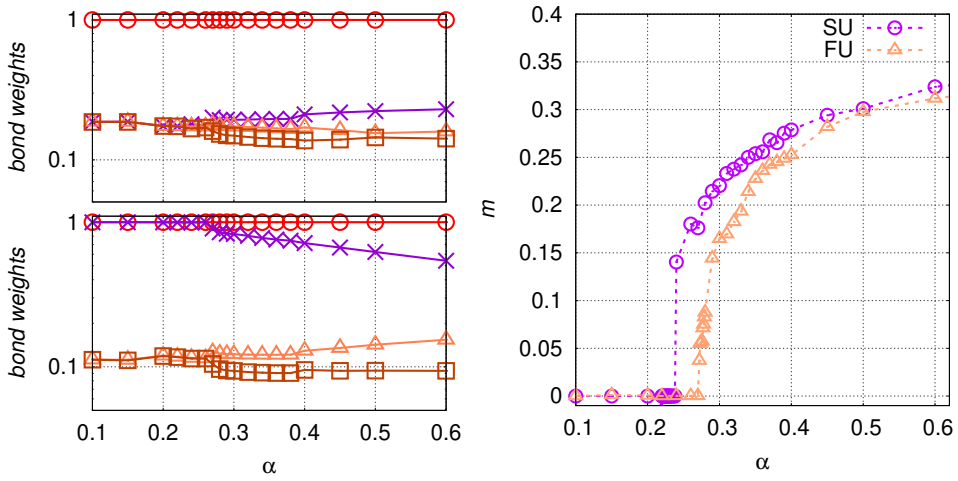


Figure 4.10: (left) Spectra on selected bonds of $D = 4$ optimized by FU across the transition. In the paramagnetic phase, the orizontal bonds running along ladders show degeneracies 1,3 . . . , while strong vertical bonds (bottom) show degeneracies 2, 2 . . . . (right) Phase diagram of the model of Eq. (4.4) obtained by using both SU and FU with bond dimension $D = 4$.

$D = 8$). In all the other cases, multiplets are broken, which leads to a small residual magnetization. Although m can be made relatively small, a faithful extrapolation for $D \to \infty$ is not possible, if not limited to the values of D that give the correct $m = 0$ result. This outcome poses serious problems whenever we want to describe a paramagnetic (e.g., spin-liquid) phase with a complicated (and not *a priori* known) multiplet structure. Indeed, it is clear that in this case a blind optimization will very likely lead to a state with a small but finite magnetization, masking the existence of a truly quantum paramagnet or spin-liquid phase with vanishing magnetization.

As a consequence of the previous results, the magnetization curve by varying the interladder coupling α is reasonable only for $D = 4$ (and $D = 8$, not shown), being finite and smooth (vanishing) for large (small) values of α (see Fig. 4.10). Still, for this (small) value of the bond dimension the transition point is underestimated within SU (i.e., $\alpha \approx 0.24$); in addition, a relatively large jump of the magnetization is observed, in contrast to the exact behavior where a continuous transition takes places. By emplyoing FU, the critical point shifts towards the correct location (i.e., $\alpha \approx 0.27$), and also the jump disappears. Notice that at the quantum critical point the multiplet structure of the tensor is broken and the ground state develops a finite magnetization. For other choices of the bond dimension the results are clearly non-physical: for $D = 2$ and 3 a completely smooth curve may be obtained, with $m > 0$ down to $\alpha = 0$. Instead, for $D = 5$, 6, and 7 it is remarkably hard to work out a smooth curve, and most importantly, finite values of m are still obtained in the paramagnetic regime. This irregular behavior makes it very difficult (if not impossible) to perform an extrapolation for $D \to \infty$ or finite correlation length scaling as we did in the isotropic Heisenberg antiferromagnet ($\alpha = 1$) benchmark. These issues become more severe especially close to criticality.

4.3.3 Conclusions

In this case study, we have highlighted a few relevant issues that appear within the iPEPS optimization. First of all, the widely used SU and FU techniques are very sensitive to the initial state when applied in a phase with no broken continuous symmetry, giving final states that may have considerably different physical properties (e.g., magnetization), while having very close energies. In the example considered here, the spin-1/2 Heisenberg antiferromagnet on coupled two-leg ladders, this situation is particularly evident, since large fluctuations in the magnetization are present within the paramagnetic phase (especially within SU, but also within FU). The second and the most important aspect, which has not been realized in the past, is the strong dependence of the results on the bond dimension D. This feature is intimately related to the presence of multiplets in the tensors of symmetric states. In the studied case, the paramagnetic phase is adiabatically connected to a valence-bond solid (VBS) with singlets along the rungs of the ladders. The FU and SU reach the ground state by progressively adding correlations on top of the initial VBS through iterative application of Trotter gates, which themselves have a multiplet structure (the $\mathbf{S}_i \cdot \mathbf{S}_j$ operator decomposes into a singlet part and a triplet part). Without truncation, the multiplets would remain imprinted in the tensors making up the state. However, the multiplet structure is preserved by truncation only for specific values of D (e.g., $D = 4$ and 8), giving rise to well-behaved simulations and physically correct variational states. Instead, whenever the value of D does not fit the multiplet structure, some breaking mechanism appears, e.g., leading to a finite magnetization and a rough energy landscape. By a similar reasoning, we expect multiplets to play a role also in the VBS phase of the $J - Q$ model [72]. In particularly simple models, such as the one that has been considered here, it is not hard to find out the exact degeneracy of multiplets and obtain reasonable results, possibly even with a scaling analysis with D. In more complicated cases (e.g., the frustrated $J_1 - J_2$ Heisenberg model on a square or triangular lattice), it might not be easy to work out the degeneracy, possibly leading to spurious results, with finite magnetization. In this respect, it is particularly important to either use $SU(2)$-symmetric tensors [65, 70] or at least impose the multiplets by using abelian symmetries [73, 74] and compare with unconstrained optimization, in order to understand the actual physical properties of highly entangled ground states.

5

AD-optimized iPEPS for frustrated magnets

CONTENTS

5.1 $\mathbf{J_1 - J_2}$ model - Prototype of frustrated quantum magnet

The conception of RVB theory by Anderson in 1973 [1] initiated a hunt for a physical realization of such spin liquid states. More than a decade later, together with Baskaran, they conjectured spin liquid state to be the origin of high-T_C superconductivity in cuprates [22]. Within the domain of quantum magnets a promising path towards the spin liquid phase is the frustration. The Heisenberg antiferromagnet on a square lattice is ordered, but the inclusion of antiferromagnetic next-nearest-neighbor interaction can supress the magnetic order and instead lead to a non-magnetic phase. The paradigmatic model of frustrated magnetism, the so-called $J_1 - J_2$ model on a square lattice exemplifies this behaviour. Its Hamiltonian reads

$$\mathcal{H} = J_1 \sum_{\langle ij \rangle} \mathbf{S}_i \cdot \mathbf{S}_j + J_2 \sum_{\langle\langle ij \rangle\rangle} \mathbf{S}_i \cdot \mathbf{S}_j, \tag{5.1}$$

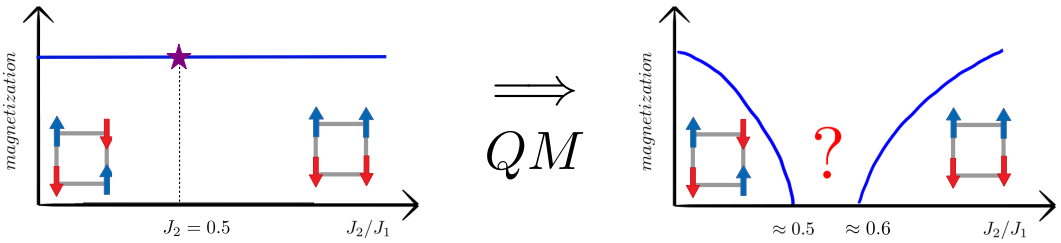


Figure 5.1: Sketch of a phase diagram of classical and quantum frustrated spin-1/2 $J_1 - J_2$ model. Left: Classical phase diagram contains two phases, Néel antiferromagnet and stripe AF separated by a point $J_2 = 0.5$. Right: Quantum fluctuations lead to an emergence of a non-magnetic intermediate phase between two ordered phases.

where $\langle ij \rangle$ denotes nearest-neighbor pairs of sites and $\langle\langle ij \rangle\rangle$ next-nearest-neighbor pairs (along diagonals) respectively. For both couplings $J_1, J_2 > 0$ the model is frustrated and from now on we set J_1 to unity.

At a classical level, the phase diagram of frustrated spin-1/2 $J_1 - J_2$ model consists of two ordered phases separated by point at $J_2 = 0.5$. The Néel phase for $J_2 < 0.5$ and a antiferromagnetic stripe phase for $J_2 > 0.5$, illustrated in Fig. 5.1. The appearance of stripe phase can be understood in the limit $J_1 = 0$, where the system decouples into two interpenetrating square lattices each hosting a Néel phase on its own.

First investigation of quantum spin-S $J_1 - J_2$ model was carried out in 1988 by Chandra and Doucot in terms of linear spin-wave theory [75]. Their results showed an instability of magnetic order (both Néel and stripe) against quantum fluctuations. The effect is most pronounced for the smallest spin $S = 1/2$, while the classical picture is correctly recovered in the limit of $S \to \infty$, where the instability region shrinks to a point at $J_2 = 0.5$. Returning to spin $S = 1/2$, the early perturbative results [76, 77], based on series expansion around dimerized state, predicted a continuous transition into non-magnetic phase around the point $J_2^{c_1} \approx 0.38$, after which an $S = 1$ gap opens. Eventually, an instability towards ordered stripe phase develops and the system undergoes a first order transition at $J_2^{c_2} \approx 0.62$ into stripe phase.

The present numerical results agree on the existence of paramagnetic phase, however its true nature remains unsettled. The exact diagonalization study of lattices up to $N = 40$ sites [78] concluded a gapped non-magnetic region $0.35 \lesssim J_2 \lesssim 0.66$ albeit with substantial finite-size effects. The results from DMRG on long cylinders $2L \times L$ of circumference up to $L = 14$ differ in the detailed description of the non-magnetic phase. Computations by Jiang, Yao, and Balents [79] point to a gapped $\mathbb{Z}_2$ spin liquid in the whole region $0.41 \lesssim J_2 \lesssim 0.62$. In the DMRG study by Gong et al. [80], using larger bond dimension, instead two distinct phases were identified within the non-magnetic region. First, the Néel order vanishes at $J_2^{c1} \approx 0.44$ giving rise to a putative gapless spin liquid phase. Second, a plaquette VBS emerges for $0.5 \lesssim J_2 \lesssim 0.61$. The most recent DMRG study by Wang and Sandvik [81] focused on the level crossing analysis on cylinders with circumference up to $L \leq 10$. Their results point to a gapless spin liquid in the region $0.46 \lesssim J_2 \lesssim 0.52$ and a VBS in the rest of the non-magnetic region extending up to $J_2^{c_2} \approx 0.62$.

The second widely used numerical method to treat this system is variational Monte Carlo (VMC), based on the Gutzwiller projected BCS wave functions with Jastrow factor. Studying $L \times L$ lattices of size up to $L = 18$, Hu et al. [82] report a gapless spin liquid in the region $0.48 \lesssim J_2 \lesssim 0.6$. Similar work by Morita, Kaneko, and Imada [83], in resemblance to some of the DMRG results, finds an evidence for a gapped VBS in the region $0.5 \lesssim J_2 \lesssim 0.6$. Recently, a study by Ferrari and Becca [84] of lattices up to size $L = 22$ strengthened the case for a gapless spin liquid appearing around $J_2^{c_1} \approx 0.48$.

Finally, this brings us to the youngest variational method to be applied to $J_1 - J_2$ model, namely tensor networks and (i)PEPS in particular. Large-scale study by Liu et. al [85], using finite PEPS with $D \leq 8$ on lattices of size up to $L \leq 14$

optimized by stochastic gradient, points to a gapless spin liquid in the whole region $0.42 \lesssim J_2 \lesssim 0.6$. Using the $U(1)$-symmetric iPEPS and FU optimization for $D \leq 9$, Haghshenas and Sheng [38] instead report a columnar VBS in the region $0.53 \lesssim J_2 \lesssim 0.61$, consistent with some of the DMRG and VMC results. An alternative approach, based on the family of RVB wave functions, was recently proposed by Poilblanc, Mambrini, and Capponi [71, 86]. They construct extended RVB states as C4v symmetric iPEPS with $D \leq 7$ from a small set of $SU(2)$-symmetric tensors, which are then optimized by gradients obtained from finite differences. Focusing the optimization on a single point $J_2 = 0.55$, the optimal iPEPS provide very good variational energy and hint at spin liquid ground state with algebraic dimer-dimer correlations, similar to the example of short-range RVB of Sec. 2.5.

Our goal is to perform high-precision unrestricted iPEPS simulations of $J_1 - J_2$ model utilizing reverse mode AD. Doing so will provide context for results from both $U(1)$-symmetric and extended RVB iPEPS, allowing us to compare extrapolations to the limit of $D \to \infty$. Also, such computation will serve as an important benchmark for capabilities of gradient optimization with AD. In the rest of this chapter we will briefly revisit couppled ladders to highlight the benefits of gradient optimization. Afterward, we will first present the phase diagram of $J_1 - J_2$ model for $0 \leq J_2 \leq 0.55$ from AD-optimized C4v symmetric iPEPS with $D \leq 6$. Second, focusing on the point $J_2 = 0.55$, we optimize also $D = 7$ iPEPS and perform extrapolation $D \to \infty$. The extrapolated variational energy is below both the best available DMRG and VMC estimates. Finally, we compare single-site, 2×1, and 2×2 unit cell iPEPS and observe an appearance of VBS for $D = 4$. All simulations have been performed with `tn-torch` [87], an AD-enabled iPEPS library based on PyTorch [44], which has been developed in the course of this work.

5.2 Lesson from the couppled ladders

The optimization of coupled ladders [see Eq. (4.4)] with FU highlighted several of its issues. Here, we return to the same system and optimize the 2×2 iPEPS with four on-site tensors $|\Psi(\mathbf{a})\rangle = |\Psi(a, b, c, d)\rangle$ with gradients computed by AD. At each step, we first perform the directional CTM and then compute the energy per site using reduced density matrices of eight non-equivalent bonds. The CTM together with the computation of energy consitutes the scalar function to be differentiated and intermediate tensors produced during its execution are stored. The CTM is considered converged once the difference of energies between two consecutive CTM iterations goes below $\epsilon_{CTM} = 10^{-8}$. For iPEPS with $D = 5$ and close to criticality at $\alpha = 0.3$ it takes typically forty CTM iterations. Second, the gradient $\mathbf{g}_i$ with respect to all on-site tensors $\mathbf{a}_i$ is evaluated by AD. Finally, the tensors are updated to $\mathbf{a}_{i+1} = \mathbf{a}_i + \beta \mathbf{p}_i$ by L-BFGS method [88], where $\mathbf{p}_i$ is a direction and β is the step size. At each step i, the L-BFGS approximates the Hessian B_i using the past

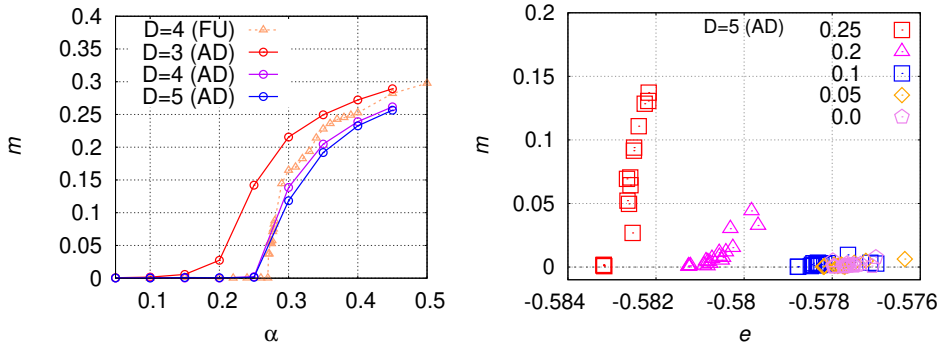


Figure 5.2: Coupled ladders model [see Eq. (4.4)] revisited with AD optimization. Left: Phase diagram obtained by AD for $(D, \chi_{opt}) = (3, 72), (4, 96)$, and $(5, 96)$. At $D = 4$ AD optimization improves the energies by $2 - 3 \times 10^{-4}$ with respect to FU. For $D = 3$ the iPEPS experiences smooth crossover from paramagnet into ordered phase and for $D \geq 4$ the magnetization curves become compatible with continuous phase transition. Moreover, for $D = 5$ small values of $m \approx 10^{-4} - 10^{-3}$ are obtained even in the paramagnetic phase. Right: Local optima from twenty instances of AD optimization intialized by a VBS with a small noise, similar to Fig. 4.8. For $\alpha < 0.3$ AD finds the optimal (leftmost) iPEPS with small magnetizations.

gradients and then solves the first order optimality condition $B_i \mathbf{p}_i = -\mathbf{g}_i$ to find the optimal direction $\mathbf{p}_i$.

For each value of α we prepare twenty initial VBS with noise and optimize them. After two hundred AD steps, the optimal iPEPS candidates for each α are taken to produce the phase diagram, shown in Fig. 5.2. Most importantly, AD gives consistent results for increasing bond dimension. At $D = 3$, we obtain magnetization curve that resembles a smooth crossover. For both $D = 4$ and $D = 5$ we get a curve in agreement with continous transition[1]. The results for $D = 4$ obtained by AD and FU are comparable, but the crucial improvement of AD with respect to FU is the small magnetization for $D = 5$ iPEPS in the whole paramagnetic phase. A detailed look at $D = 5$ reveals that AD also experiences rough energy landscape in the paramagnetic phase, shown in Fig. 5.2. However, in contrast to FU, AD consistently finds minima with lower energy and small magnetization. Hence, we can draw a conlusion that for AD imposing the correct multiplets is not a necessary condition for obtaining iPEPS with small magnetization. On the other hand we conjecture that doing so might still tame the roughness of the energy landscape improving the convergence properties. This is connected with another issue of a more practical character. The overall process of FU simulations, even at $D = 4$, is very tedious. Smoothing out the magnetization curve requires repeated initialization of FU from states at nearby values of α. Gradient optimization on the other hand can already provide optimal iPEPS which form a consistent and smooth magnetization curves with a simpler protocol as described above.

[1] more points between $0.2 < \alpha < 0.3$ are required to distinguish between weak crossover or sharp feature as exhibited by the FU data.

In summary, the gradient based optimization can overcome the issues present in the unrestricted FU. Imposing the multiplets might still be beneficial, and in coupled ladders, it is simple to infer them based on the product state of valence bonds which is adiabatically connected to the $\alpha < \alpha_c$ paramagnetic phase. The same cannot be said for the $J_1 - J_2$ model. Both the Néel and the stripe phase are symmetry-broken phases with no such structure. Moreover, the intermediate phase has no known origin from which it can be adiabatically evolved. For gapless phases, such process is not even possible. Therefore, gradient optimization positions itself as the optimal if not the only tool to study the $J_1 - J_2$ model with unrestricted iPEPS ansatz.

5.3 $\mathbf{J_1 - J_2}$ model with single-site symmetric iPEPS

To study the phase diagram of frustrated $J_1 - J_2$ model we first choose single-site C4v symmetric iPEPS, enforcing the symmetries of the square lattice in accordance with the putative spin liquid phase. The Néel order is also supported by this ansatz through the unitary transformation U acting on the spin index of every on-site tensor belonging to one chosen sublattice of bipartite square lattice (see Eq. A.13 for details). We optimize this iPEPS with gradient optimization using AD up to bond dimension $D = 6$. At each step of the optimization, first the symmetric CTMRG is performed and then the energy is evaluated as illustrated in Fig. 3.7. The only adjustment is the reduced density matrix of 2×2 region used to evaluate the energy since it accommodates both the NN and NNN terms making up the $J_1 - J_2$

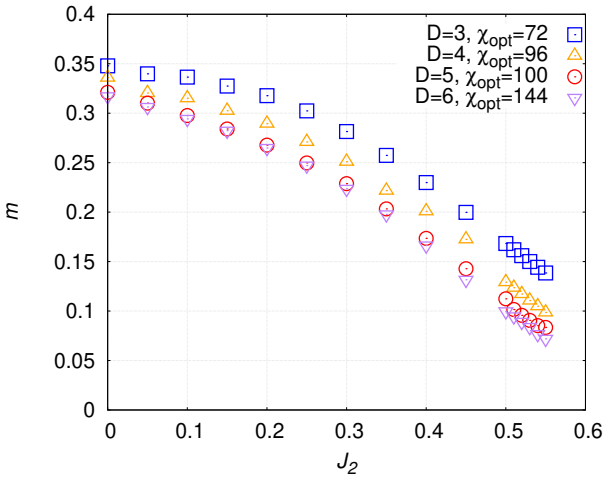


Figure 5.3: The phase diagram of $J_1 - J_2$ model with single-site C4v symmetric iPEPS. Magnetization curves obtained from AD-optimized iPEPS are smooth. In all cases, the optimal iPEPS show finite magnetization down to $J_2 = 0.55$.

Hamiltonian. All intermediate tensors produced in course of this process are stored. Second, the gradient is evaluated by AD and then the on-site tensor is updated according to L-BFGS described in the previous section. Finally, the updated tensor is symmetrized. At each AD step the initial $\{C, T\}$ environment tensors are initialized from the on-site tensor itself. The convergence threshold on the difference between energies in two consecutive CTMRG steps is set to $\epsilon_{CTM} = 10^{-8}$. The necessary computational effort is significantly smaller than in the case of coupled ladders due to the relative simplicity of single-site symmetric CTMRG.

For each value of J_2, a set of simulations is performed starting from a random C4v symmetric tensors and a modest value of environment bond dimension χ_{opt}. Typically, we perform less then hundred AD steps in this initial stage. Afterward, the best candidates are selected for subsequent optimization with increasing value of χ_{opt}. For example, in the final stage of optimization with $D = 6$ iPEPS at $J_2 = 0.55$, we performed five hundred gradient steps at $\chi_{opt} = 144$ converging the energy with a precision $< 10^{-5}$. Once the optimization phase is finished, the observables are recovered by performing χ-scaling typically up to $\chi = 2\chi_{opt}$.

We present the resulting phase diagram of J_1-J_2 model up to $J_2 = 0.55$ obtained by single-site C4v symmetric iPEPS in Fig. 5.3. First of all, the AD optimization proves to be robust and leads to a consistent picture of decreasing magnetization with increasing bond dimension. Overall, all optimal iPEPS show finite magnetization decreasing monotonically with the frustration. Only $D = 5$ data for $J_2 > 0.5$ show a slight change in the curvature of $m(J_2)$, while no such behaviour is observed for $D = 6$ iPEPS. A set of $D = 7$ optimizations might shed light on the nature of this $D = 5$ anomaly, whether it is a particular finite-D deviation or it is a systematic feature of odd D iPEPS. If so, it might be a hint of a sharp feature developing in the $D \to \infty$ limit.

Now, we analyze in more detail two particular points, $J_2 = 0.5$ and $J_2 = 0.55$, where we performed additional optimizations with $D = 7$ iPEPS up to $\chi_{opt} = 147$. First, we briefly comment on the data for $J_2 = 0.5$, shown in Fig. 5.4. The $J_2 = 0.5$ point is closer to the criticality, where the Néel order disappears according to DMRG and VMC computations. However, the finite-D scaling points to a ground state with small yet finite magnetization $m(D \to \infty) \approx 0.04$, at odds with DMRG and VMC. It is instructive to compare our results with the $U(1)$-symmetric FU iPEPS study. The extrapolated magnetization is consistent with the data from the $U(1)$ ansatz. Still, with respect to FU the AD optimization improves the energy by a considerable factor. For $D = 7$, the AD-optimized iPEPS has an energy $e(J_2 = 0.5, D = 7) = -0.4964$, which puts it well below the FU-optimized $D = 7$ state with energy $e_{FU}(J_2 = 0.5, D = 7) \approx -0.4959$. We perform extrapolations in $1/D$ of the energy from the $D > 3$ data to $D \to \infty$, both linear and power law. The resulting estimates are $e^{(lin)}(J_2 = 0.5, D \to \infty) = -0.4986(2)$ and $e^{(pwr)}(J_2 = 0.5, D \to \infty) = -0.4971(2)$ for linear and power-law fit respectively. Linear extrapolation gives surprisingly low energy compared to DMRG and VMC. On the other hand, the power law fit is compatible with VMC extrapolated energy, suggesting that optimizations for additional values of D are necessary to reach a definitive conclusion on the asymptotic behaviour of energy for large D.

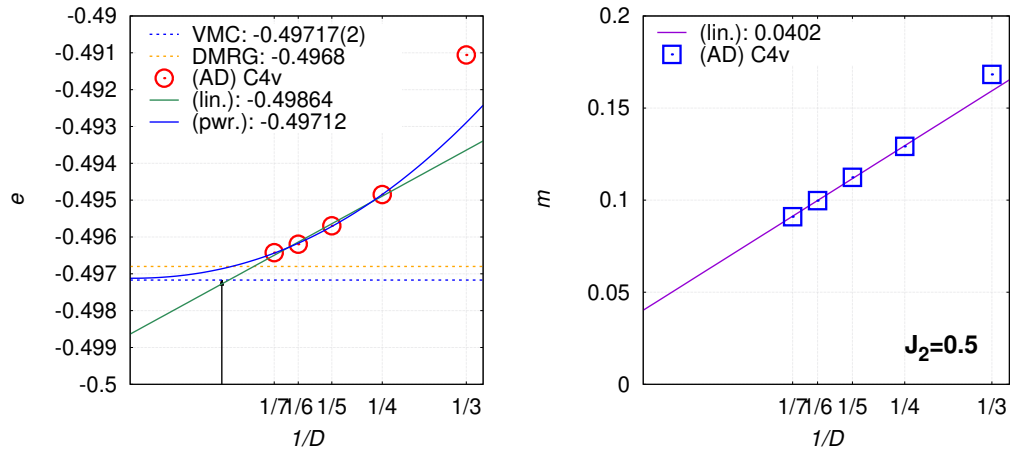


Figure 5.4: The energy and magnetization of C4v symmetric iPEPS at $J_2 = 0.5$. Left: Energy as a function of $1/D$ with both linear (green) and power law (blue) extrapolations of the energy to $D \to \infty$ limit. The VMC [82] and DMRG [80] estimates of energy in the thermodynamic limit are included for comparison. Right: Finite-D scaling of magnetization gives a value compatible with $U(1)$-symmetric iPEPS study [38].

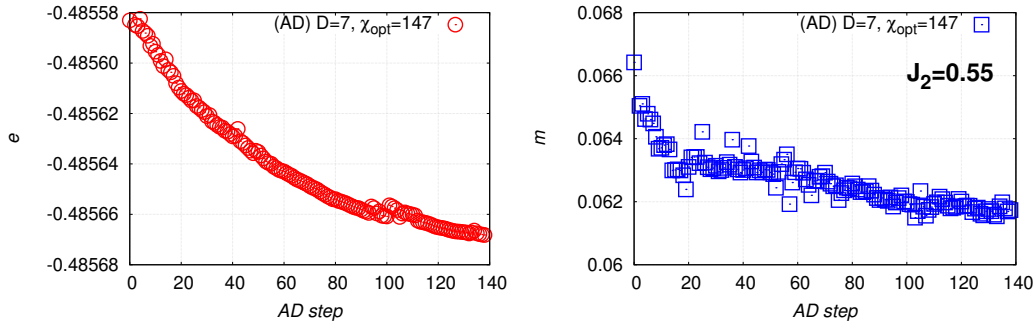


Figure 5.5: Final stage of AD optimization of $D = 7$ C4v symmetric iPEPS at $J_2 = 0.55$. Left: Energy as a function of AD steps. Typically, the convergence of CTMRG below ϵ_{CTM} threshold requires $35 - 40$ RG steps. Occasionally, the energy increases after an AD step (see text). Right: Convergence of magnetization throughout the optimization.

We turn to the highly frustrated $J_2 = 0.55$ point. The final stage of optimizations of $D = 7$ iPEPS, requiring more than a hundred AD steps, is shown in the Fig. 5.5. Eventually, we converge the energy to a precision $< 2 \times 10^{-5}$. Sometimes, AD step results in the increase of the energy due to the overestimated step size β since we opt to not use any strong line-search method[1]. The complete results

[1] to minimize the function along the direction $\mathbf{p}$ given by L-BFGS the braket method is often adopted. The minimum of such one-dimensional problem is reliably pinpointed by iteratively shrinking an interval. However, such line-search methods typically involve (many) function evaluations, which in the case of iPEPS means very costly CTM iterations.

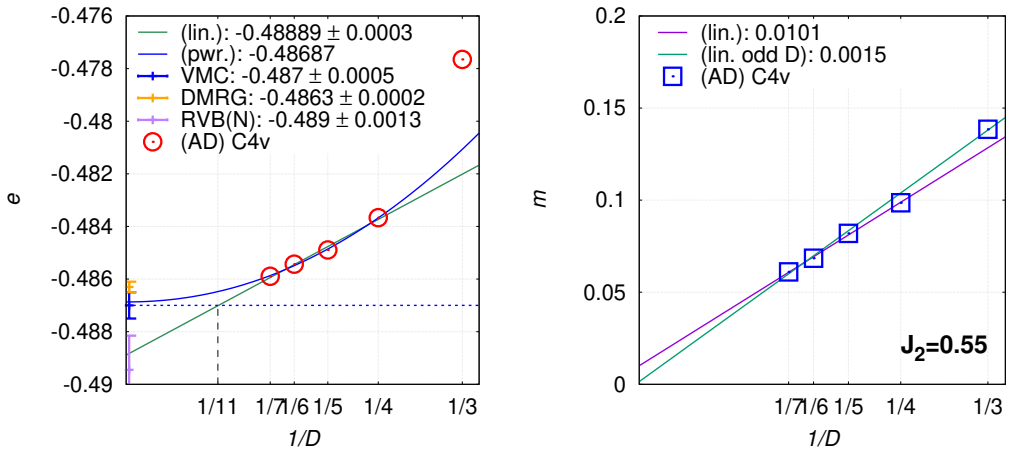


Figure 5.6: Finite-D scaling of the energy and magnetization from AD-optimized C4v symmetric iPEPS at $J_2 = 0.55$. Left: Linear extrapolation (green) of energy is below both DMRG [80] and VMC [82] estimates and it is compatible with the extrapolation from N-colored RVB iPEPS [86]. Power law extrapolation (blue) is compatible with the VMC estimate. Right: Linear extrapolation with all points gives a very small, yet finite magnetization. Using just an odd D iPEPS, which can host coloured RVB states, the extrapolated value of magnetization is an order of magnitude smaller.

for $J_2 = 0.55$ together with finite-D scaling are presented in Fig. 5.6. The linear extrapolation of energy gives $e^{(lin)}(J_2 = 0.55, D \to \infty) = -0.4889(3)$ while the power-law extrapolation instead leads to $e^{(pwr)}(J_2 = 0.55, D \to \infty) = -0.4869(5)$. Both extrapolations are below the DMRG estimate for the energy in the thermodynamic limit. Together with VMC, the present results provide evidence that DMRG is probably overestimating the ground state energy. The most likely reason is that the DMRG simulations end up trapped in a local minimum at this highly frustrated point. An alternative scenario is for the energy of iPEPS as a function of $1/D$ to rapidly change curvature. Reaching the bond dimension $D = 11$ might be sufficient to resolve this question, since for this value of D the expected linear dependence will cross not only DMRG but also the VMC energy estimate. Another reason for optimizing our ansatz up to $D = 11$ comes from the $SU(2)$-symmetric extended RVBs. Current numerical data [86] for these extended RVBs suggest that they should become energetically favorable beyond $D > 11$. Also the magnetization seems to vanish for large bond dimensions as indicated by the extrapolation from odd D data. If it is indeed the case, the non-magnetic C4v symmetric iPEPS is at odds with the tenative VBS phase emerging for $J_2 \gtrsim 0.52$ as seen by some of the DMRG and VMC studies.

We close this section by presenting the χ-scaling of the observables at $J_2 = 0.55$, shown in Fig. 5.7, for optimal iPEPS with $D = 5, 6$ and 7. The overall behaviour of energy and magnetization with χ, characterized by modest corrections, suggests that all the iPEPS considered have finite correlation length, which is expected to be a generic outcome of unrestricted iPEPS optimization [54].

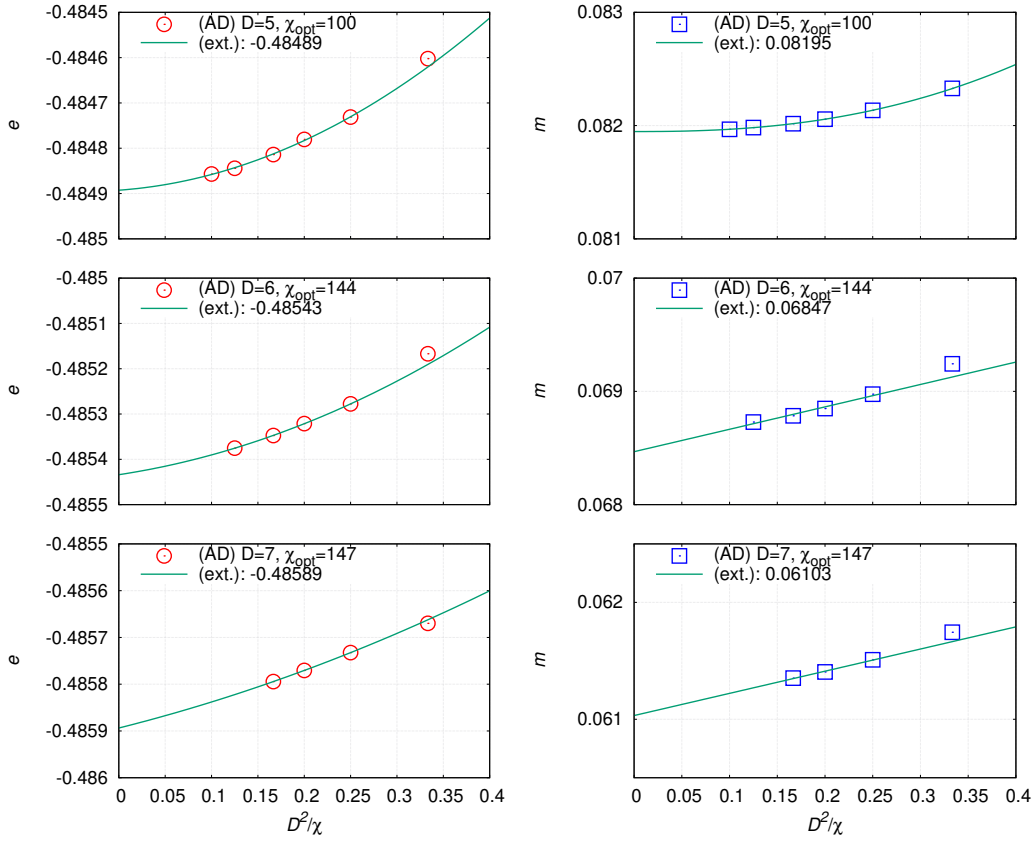


Figure 5.7: Scaling of energy and magnetization with the environment dimension χ at $J_2 = 0.55$. From top to bottom: The case of optimal iPEPS $(D, \chi_{opt}) = (5, 100)$, $(6, 144)$, and $(7, 147)$.

5.4 The 2×1 and 2×2 unit cells

Finally, we give a glimpse at the simulation of $J_1 - J_2$ model beyond C4v symmetric iPEPS. For this model, AD optimization of iPEPS with 2×2 unit cell requires only a slight modification with respect to coupled ladders. To compute the energy we now have to evaluate contribution also from the next-nearest-neighbor terms. Therefore, we use four reduced density matrices $\rho_{2\times2}$, one for each non-equivalent 2×2 region, instead of eight reduced density matrices for nearest-neighbor bonds as in the coupled ladders. In the case of 2×1 unit cell, there are just two such non-equivalent 2×2 regions.

The result for $D = 4$ case, the smallest bond dimension where the columnar VBS develops is shown in Fig. 5.8. The complementary data from optimal iPEPS for both 2×1 and 2×2 unit cells indicate a sharp transition around $J_2^{c_1} \approx 0.52$ into columnar VBS. Although the 2×2 iPEPS can host both the Néel ordered phase and the columnar VBS, the transition from former to latter still appears as first order

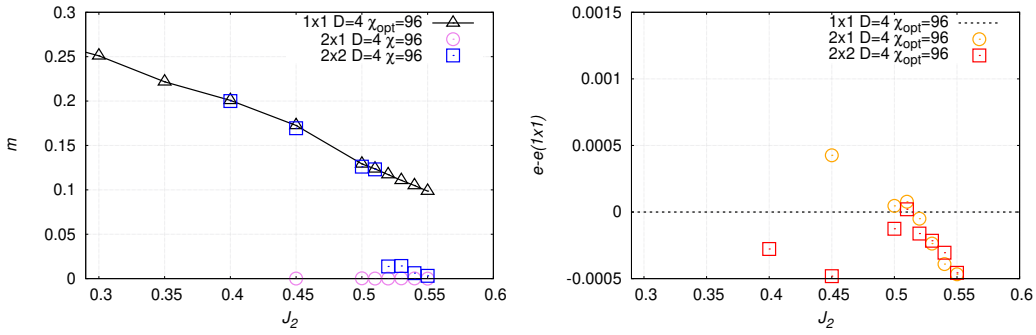


Figure 5.8: AD-optimized $D = 4$ iPEPS with 2×1 and 2×2 unit cell. Left: Magnetization of iPEPS with different unit cells. Around $J_2 \approx 0.52$ a columnar VBS order develops and the magnetization drops to vanishing values. Right: Comparison of energies with respect to C4v symmetric iPEPS. In accordance with the magnetization data, the 2×1 unit cell, which cannot support Néel order, becomes energetically favourable for $J_2 \gtrsim 0.52$ and competitive with 2×2 iPEPS also hosting a columnar VBS.

within the $D = 4$ manifold. As before, we are interested in comparison of our results with $U(1)$-symmetric FU simulations [38] since so far it is the only other iPEPS study addressing this model with general optimization beyond single-site unit cell. The FU-optimized iPEPS also point to a transition into columnar VBS and estimate its location as $J_2^{c_1} \approx 0.53$ by extrapolating the data in D. Still, at $J_2 = 0.53$ and $D = 4$ the FU optimization finds finite magnetization $m(J_2 = 0.53, D = 4) \approx 0.12$. Instead, the AD optimization locates energetically more optimal iPEPS, which hosts columnar VBS with vanishing magnetization.

We will address the consequences of these results in the final conclusions as they also indicate the future directions in which the iPEPS investigation of $J_1 - J_2$ model should proceed.

6

Conclusions and future perspectives

The AD optimization developed in the course of this work provides a powerful and flexible way to optimize iPEPS with large unit cells for complicated short-range Hamiltonians. One of the main outcomes of this work comes in the form of high-performance open-source library `tn-torch` [87], which implements production-ready differentiable CTM algorithms. The simulations can be run on both CPUs and GPUs out of the box thanks to PyTorch, serving as a fundation on top of which the iPEPS algorithms are built. The practical simplicity of AD-optimized iPEPS and the state-of-the-art performance opens up the way to more wide-spread adoption of two-dimensional tensor network techniques, which are otherwise complex and challenging to implement and extend. This simplicity becomes apparent, when passing from nearest-neighbour to next-nearest-neighbour Hamiltonians requires just an introduction of 2×2 reduced density matrix in the final step of the variational energy computation. As a consequence, the scope of systems that can be addressed by AD optimization extends beyond the next-nearest-neighbour interaction and they can be implemented with a minimal effort. There is no need for modification of the CTM algorithms. Hence, the AD optimization brings iPEPS one step closer to become the variational method of choice in two dimensions, similar to what DMRG has become in one dimension.

In the beginning of this work, we set out to answer the question, whether iPEPS can describe quantum magnets in the vicinity of continous transitions and probe their highly entangled phases such as spin liquids. The results from AD optimization of both frustration-free coupled ladders and frustrated $J_1 - J_2$ model make us optimistic. The variational energies obtained from iPEPS in highly frustrated regime at $J_2 = 0.55$ set the new reference. There are two key questions that have to be investigated. First question concerns the behaviour of apparent sharp transition from Néel phase to columnar VBS with increasing D since among the features on which both the VMC and the DMRG agree is the continuous nature of Néel to paramagnet transition. Second question is the ultimate faith of the finite VBS order in the limit of $D \to \infty$. Whether it is a genuine property of the ground state or but a finite-D effect and the columnar VBS eventually diminishes restoring the C4v symmetry.

The answer to both questions crucially depends on our ability to optimize iPEPS with increasingly large bond dimensions. This calls for further algorithmic development. In the short term horizon, the extension of AD optimization to iPEPS with abelian symmetries will decrease both the complexity and memory requirements of AD algorithm. Second appealing and novel direction is the development of adjoints for partial SVD and diagonalization algorithms, which would in turn decrease com-

plexity of projector construction in CTM down to $O(\chi^3 D^4)$ and allow us to easily reach bond dimensions beyond $D = 11$.

In summary, the pursuit of solution to Dirac's original question of solving the many-body problem of condensed matter is still in progress. The presented two-dimensional iPEPS method, together with future advances, offers a promising answer to models arising from the effective lattice description of strongly-correlated materials.

List of publications

This work led to the following publications and open-source computer libraries for conducting high-performance tensor network simulations in two dimensions

- J. Hasik and A. Sartori, *pi-peps: A C++ library built on top of ITensor for running iPEPS simulations of two dimensional spin systems*, `https://github.com/jurajHasik/pi-peps`.
- J. Hasik and F. Becca, "Optimization of infinite projected entangled pair states: The role of multiplets and their breaking", Phys. Rev. B **100**, 054429 (2019) `10.1103/PhysRevB.100.054429`.
- J. Hasik and G. B. Mbeng, *tn-torch: A differentiable tensor network library for two-dimensional lattice models*, to be released soon.

The following publications, done over the course of this PhD, have not been included in this thesis

- J. Hašík, E. Tosatti, and R. Martoňák, "Quantum and classical ripples in graphene" Phys. Rev. B **97**, 140301(R) (2018)

A

Exact $SU(2)$-invariant iPEPS

In this Appendix, we present the construction of iPEPS representation of two paradigmatic $SU(2)$-invariant states on a square lattice. The spin $S = 2$ AKLT state and the short-range RVB state.

A.1 The iPEPS representation of Spin $S = 2$ AKLT state

In this section, we derive the iPEPS with a bond dimension $D = 2$ that represents the spin $S = 2$ AKLT state. In Sec. 2.2, we introduced iPEPS from the top-down perspective, that is, as a specific parametrization of the wave function coefficients motivated by the Area law. Whereas now, we will build the exact iPEPS representation from the bottoms-up. Starting with a specific, but simple state $|AUX\rangle$ living in the auxiliary space and then projecting it into physical Hilbert space of spin $S = 2$ degrees of freedom (DOF). This projection will be defined in terms of the familiar on-site tensors a. Their action on the $|AUX\rangle$ state will induce the iPEPS representation we are looking for.

As the first step, we associate four auxiliary spins $S = 1/2$ to each site of the square lattice. Next, a simple product state is formed by these auxiliary spins. They are paired in VBs $|VB\rangle_{ij} := |\uparrow_i\downarrow_j\rangle - |\downarrow_i\uparrow_j\rangle$ across the nearest-neighbor links of the lattice

$$|AUX\rangle = \prod_{i\in A} \Big(|VB\rangle_{i,i+x} |VB\rangle_{i,i-x} |VB\rangle_{i,i+y} |VB\rangle_{i,i-y} \Big), \tag{A.1}$$

where the product runs only over the sublattice-A sites of a square lattice and the orientation of all singlets is fixed[1] as to point into the sublattice-A sites (see Fig. A.1). Second, we define a projector a acting on the four auxiliary spins associated with each site. It projects their collective state into the local physical Hilbert space of spin $S = 2$, which is also the highest weight irreducible representation (irrep) formed by their tensors product

$$a: \quad \otimes_4 \mathcal{H}(S = 1/2) \to \mathcal{H}(S = 2). \tag{A.2}$$

Choosing the spin S^z eigenstate basis both in the auxiliary and the local physical

[1] it is not possible to define orientation for singlets on any two-dimensional lattice. It must be a bipartite lattice. A triangular or kagome lattice is a counter-example.

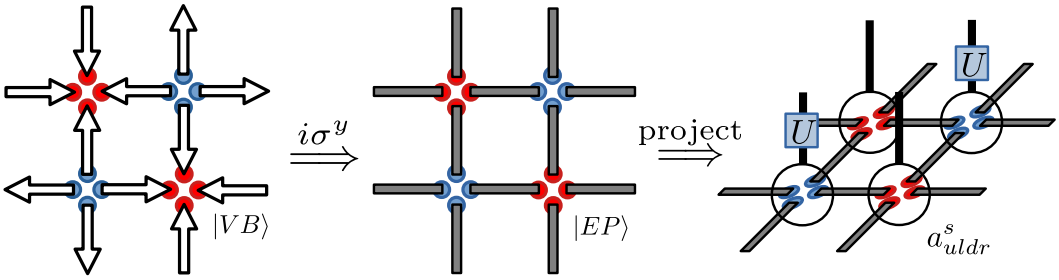


Figure A.1: Construction of $S = 2$ AKLT state on a square lattice. Left: The product state of oriented VBs (arrows) denoted as $|AUX\rangle$ (see text). The orientation is fixed such that all VBs point into the sublattice-A sites (red circles). Middle: Apply rotation operators $i\sigma^y$ on the auxiliary spins of all sublattice-B sites (blue circles) transforming the valence bonds into the $|EP\rangle$ states which do not have an orientation (grey bars). Right: Application of the projectors a. On the sublattice-B sites, the permutation operator U (blue squares) is applied on the physical index (black vertical line). The result is a single-site C4v iPEPS representation of the AKLT state.

space the projector a takes familiar form of rank-5 on-site tensor of iPEPS

$$a = \sum_{suldr} a^s_{uldr} \, |s\rangle \, \langle uldr| \, , \tag{A.3}$$

$$a^s_{uldr} = CGC(2, s|1/2, u; 1/2, l; 1/2, d; 1/2, r), \tag{A.4}$$

where CGC gives the Clebsh-Gordan coefficients of the decomposition of tensor product of four auxiliary spins into spin $S = 2$ irrep. The index s runs over S^z eigenstates of the spin $S = 2$, while the indices $\{u, l, d, r\}$ run over the eigenstates of auxiliary spins associated with the four lattice directions up, left, down, and right. Tensor a is also C4v symmetric since it is invariant with respect to any permutation p of the auxiliary indices

$$a^s_{uldr} = a^s_{p(uldr)}. \tag{A.5}$$

The projectors a are applied to $|AUX\rangle$ on every site of a square lattice. The outcome of the projection is the AKLT state

$$|AKLT(S = 2)\rangle = \left(\prod_i a^i \right) |AUX\rangle \, , \tag{A.6}$$

where the product runs over all sites i of the square lattice. To represent this state as iPEPS we have to express its coefficients in the form of Eq. (2.18) as the contraction of a tensor network.

Let us make the following observation. Take a pair of the projectors a acting on two sites i and j. Suppose there is a single entangled pair $|EP\rangle_{r_i l_j} := |{\uparrow_{r_i}}{\uparrow_{l_j}}\rangle + |{\downarrow_{r_i}}{\downarrow_{l_j}}\rangle$

connecting two auxiliary spins, one from each site. The projection then gives

$$\begin{aligned} a^i a^j \left|EP\right\rangle_{r_i l_j} &= \sum_{r_i l_j r_i' l_j' \ldots} a^{s_i}_{u_i l_i d_i r_i'} \left|s_i\right\rangle \left\langle u_i l_i d_i r_i'\right| a^{s_j}_{u_j l_j' d_j r_j} \left|s_j\right\rangle \left\langle u_j l_j' d_j r_j\right| \delta_{r_i l_j} \left|r_i\right\rangle \left|l_j\right\rangle \\ &= \sum_{\ldots} \left(\sum_x a^{s_i}_{u_i l_i d_i x} a^{s_j}_{u_j x d_j r_j} \right) \left|s_i\right\rangle \left|s_j\right\rangle \left\langle u_i l_i d_i\right| \left\langle u_j r_j d_j\right| \\ &= \sum_{\substack{s_i u_i l_i d_i \\ s_j u_j l_j d_j}} \left(a^i a^j\right)^{s_i s_j}_{u_i l_i d_i u_j r_j d_j} \left|s_i\right\rangle \left|s_j\right\rangle \left\langle u_i l_i d_i\right| \left\langle u_j r_j d_j\right| . \end{aligned} \tag{A.7}$$

In the last line we recognize the expression of Eq. (2.20), a result of the contraction of two tensors by a pair of indices. The indices that are contracted are effectively specified by the entangled pair $|EP\rangle$. Therefore, we have observed that projecting an EP is equivalent to a contraction of the projectors by the paired indices.

With this fact in mind, we define a slightly different state $|AUX'\rangle$ in the auxiliary space. It is given as a product state of EPs, instead of VBs, connecting the auxiliary spins across the nearest-neighbor links of the lattice

$$\left|AUX'\right\rangle = \prod_{i\in A} \Big(\left|EP\right\rangle_{i,i+x} \left|EP\right\rangle_{i,i-x} \left|EP\right\rangle_{i,i+y} \left|EP\right\rangle_{i,i-y} \Big) , \tag{A.8}$$

where the product again runs only over the sublattice-A sites. In contrast to the original $|AUX\rangle$ state, the EPs do not have an orientation (see Fig. A.1). Applying the projectors on every site of $|AUX'\rangle$ state is equivalent to their contraction, which can be seen by the repeated application of Eq. (A.7). Hence, such a state is an iPEPS since its coefficients are given by a contracted tensor network formed from the on-site tensors a. This is the origin of the name Projected entangled-pair state, or PEPS in short. Therefore, if we can transform the VBs forming the $|AUX\rangle$ state into EPs and then apply the projectors, the result would be an iPEPS representation of the AKLT state.

Observe that applying a rotation $i\sigma^y$ on the second index of a VB gives EP

$$(i\sigma^y) \left|\uparrow\right\rangle = -\left|\downarrow\right\rangle ; \ (i\sigma^y) \left|\downarrow\right\rangle = \left|\uparrow\right\rangle , \tag{A.9}$$

$$(i\sigma^y)_j \left|VB\right\rangle_{ij} = (i\sigma^y)_j (\left|\uparrow_i\downarrow_j\right\rangle - \left|\downarrow_i\uparrow_j\right\rangle) = \left|\uparrow_i\uparrow_j\right\rangle + \left|\downarrow_i\downarrow_j\right\rangle = \left|EP\right\rangle_{ij} . \tag{A.10}$$

Hence, using an identity $\mathbb{1} = (-i\sigma^y)(i\sigma^y)$, we act with the rotation $i\sigma^y$ on every VB in the $|AUX\rangle$ state, such that it always rotates the auxiliary spin on the B-sublattice

$$\begin{aligned} \mathbb{1} \left|AUX\right\rangle = \prod_{i\in A} \Big(& (-i\sigma^y)_{i+x} \left|EP\right\rangle_{i,i+x} (-i\sigma^y)_{i-x} \left|EP\right\rangle_{i,i-x} \\ & (-i\sigma^y)_{i+y} \left|EP\right\rangle_{i,i+y} (-i\sigma^y)_{i-y} \left|EP\right\rangle_{i,i-y} \Big) . \end{aligned} \tag{A.11}$$

The last but important detail is the realization that the joint action of four such rotation operators $i\sigma^y$ transforms the projector a on every B-sublattice site in a

simple way. Their action is equivalent to a unitary U acting on the physical spin $S = 2$ instead

$$(i\sigma^y)_{uu'}(i\sigma^y)_{ll'}(i\sigma^y)_{dd'}(i\sigma^y)_{rr'}a^s_{u'l'd'r'} = U_{ss'}a^{s'}_{uldr}, \tag{A.12}$$

$$U_{i,dim(S)+1-i} := (-)^i. \tag{A.13}$$

Therefore, making the use of the unitary U, we can finally give the iPEPS representation of the AKLT spin $S = 2$ state[1] on a square lattice

$$|AKLT(S=2)\rangle = \sum_{\{s_i\}} \left(\prod_{i\in B} U^i \right) \mathrm{Tr}_{aux} \left(\prod_i a^{s_i} \right) |\{s_i\}\rangle \,. \tag{A.14}$$

The state is defined using only a single C4v symmetric on-site tensor a with the bond dimension $D = 2$, which is the dimension of space of auxiliary spin $S = 1/2$. We also illustrate the main steps of the construction in Fig. A.1, notice the unitaries U acting on every B-sublattice site. Finally, we make in important remark: the presence of the unitaries is irrelevant for the CTMRG. They cancel each other site-by-site when the double-layer network is formed.

A.2 The Parent Hamiltonian of spin $S = 2$ AKLT state

The parent Hamiltonian of AKLT is physical since it is composed just from the local interactions. Each term is a projector from a pair of nearest-neighbor $S = 2$ spins into $S = 4$ irrep

$$H_{AKLT(S=2)} = \sum_{\langle ij\rangle} P_{ij}(S = 4|S = 2; S = 2), \tag{A.15}$$

$$P_{ij} = N(\mathbf{S}^2_{ij} - 0)(\mathbf{S}^2_{ij} - 1 \times 2)(\mathbf{S}^2_{ij} - 2 \times 3)(\mathbf{S}^2_{ij} - 3 \times 4). \tag{A.16}$$

Defining $\mathbf{S}_{ij} = \mathbf{S}_i + \mathbf{S}_j$, the P_{ij} in the second equation annihilates all two spin $S = 2$ states that do not overlap with the spin $S = 4$ irrep[1]. The normalization coefficient is fixed such that $P^2_{ij} = P_{ij}$, a necessary condition for P_{ij} being a projector. Expanding the $\mathbf{S}_{ij}$, we can write the Hamiltonian in the terms of more familiar spin-spin interaction

$$H_{AKLT(S=2)} = \sum_{\langle ij\rangle} \frac{1}{14} \left(\mathbf{S}_i.\mathbf{S}_j + \frac{7}{10}(\mathbf{S}_i.\mathbf{S}_j)^2 + \frac{7}{45}(\mathbf{S}_i.\mathbf{S}_j)^3 + \frac{1}{90}(\mathbf{S}_i.\mathbf{S}_j)^4 \right). \tag{A.17}$$

[1] analogous procedure leads to other states from the AKLT family in one dimension or on the bipartite lattices in two and higher dimensions. The physical spin is given by the maximal irrep of the tensor product of z auxilliary spin $S = 1/2$ associated to each site, where z is the lattice coordination.

[1] $SU(2)$ irreps of spin S satisfy the relation for the total spin $\mathbf{S}^2 = S(S + 1)$.

From this form it is clear that the Hamiltonian is $SU(2)$-invariant. It annihilates the AKLT state since between any two nearest-neighbors there is always a singlet at the virtual level and as a corollary the remaining auxiliary spins cannot conspire to form a state from $S = 4$ irrep. Moreover, the Hamiltonian is a sum of positive semi-definite operators, projectors P_{ij}. As such, it is also positive semi-definite and hence its lowest eigenvalue is zero, which puts the AKLT state into its ground state manifold.

A.3 The iPEPS representation of short-range RVB

The main motive behind the construction is analogous to the case of AKLT state, given in Sec. A.1. First, we define the auxiliary space. Each site of the square lattice is decorated with four auxiliary degrees of freedom. Each of them lives in the Hilbert space of dimension three, composed from the auxiliary spins $S = 1/2$ and spin $S = 0$ as the direct sum $\mathcal{H}^{aux} := \mathcal{H}(S = 0) \oplus \mathcal{H}(S = 1/2)$. The projector a acts on the virtual space built up from the four auxiliary spins attached to each site, which hosts a physical spin $S = 1/2$. Specifying the basis in the $\mathcal{H}^{aux}$ as $\{|0\rangle, |\uparrow\rangle, |\downarrow\rangle\}$, the projector a is defined to be

$$a : \quad \otimes_4 \mathcal{H}^{aux} \;\rightarrow\; \mathcal{H}(S = 1/2), \tag{A.18}$$

$$\begin{aligned} a = {} & |\uparrow\rangle (\langle\uparrow 000| + \langle 0\uparrow 00| + \langle 00\uparrow 0| + \langle 000\uparrow|) \\ & + |\downarrow\rangle (\langle\downarrow 000| + \langle 0\downarrow 00| + \langle 00\downarrow 0| + \langle 000\downarrow|). \end{aligned} \tag{A.19}$$

To understand its action, we define an entangled pair $|EP_{1+2}\rangle_{ij}$ between two spaces $\mathcal{H}_i^{aux}$ and $\mathcal{H}_j^{aux}$

$$|EP_{1+2}\rangle_{ij} = |0\rangle_i |0\rangle_j + |\uparrow\rangle_i |\uparrow\rangle_j + |\downarrow\rangle_i |\downarrow\rangle_j . \tag{A.20}$$

The action of a pair of projectors a on the $|EP_{1+2}\rangle$ state linking two auxiliary spaces is equivalent to the contraction of the corresponding tensors, see Eq. (A.7). The result can be seen as being composed from two distinct terms which are interpreted as follows: First, an $|EP\rangle = |\uparrow\uparrow\rangle + |\downarrow\downarrow\rangle$ formed from the physical spins with all other auxiliary spins set to state $\langle 0|$, and second, the pair of projectors restricted to the reduced auxiliary spaces $\otimes_3 \mathcal{H}_i^{aux}$ and $\otimes_3 \mathcal{H}_j^{aux}$

$$\begin{aligned} a^i a^j |EP_{2+1}\rangle_{ij} = {} & (|\uparrow\rangle_i |\uparrow\rangle_j + |\downarrow\rangle |\downarrow\rangle_j) \langle 000|_i \langle 000|_j + \\ & + [|\uparrow\rangle (\langle\uparrow 00| + \langle 0\uparrow 0| + \langle 00\uparrow|) + |\downarrow\rangle (\langle\downarrow 00| + \langle 0\downarrow 0| + \langle 00\downarrow|)]_i \\ & \otimes [|\uparrow\rangle (\langle\uparrow 00| + \langle 0\uparrow 0| + \langle 00\uparrow|) + |\downarrow\rangle (\langle\downarrow 00| + \langle 0\downarrow 0| + \langle 00\downarrow|)]_j. \end{aligned} \tag{A.21}$$

The diagrammatic representation, shown in Fig. A.2, leads to a simple interpretation of the above equation. Either an $|EP\rangle$ state between the physical spins is created and no other pair can be formed with them, or no pair is formed between the sites. As a corollary, there can be at most single $|EP\rangle$ state per physical site. To built an

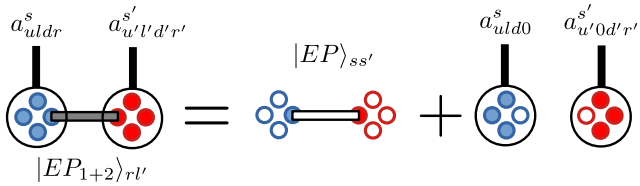


Figure A.2: Two projectors act on the $|EP_{1+2}\rangle$ in the auxiliary space (grey bar). The result is the $|EP\rangle$ between the physical spins (white bar) and two projectors with linked indices fixed to the value zero ($|0\rangle$).

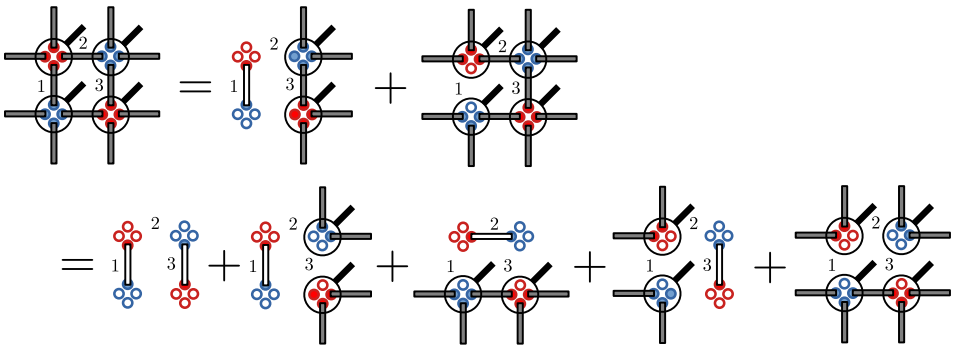


Figure A.3: Successive application of the projectors on the $|AUX\rangle$ state (see text), starting from the bond 1 and continuing with the bonds 2 and 3. The result is a superposition of all possible $|EP\rangle$ coverings over the bonds $1, 2$ and 3.

iPEPS representation, we define $|AUX\rangle$ as a product state of $|EP_{1+2}\rangle$ states linking all nearest-neighbor pairs $\langle ij\rangle$ of auxiliary spaces on a square lattice

$$|AUX\rangle = \prod_{\langle ij\rangle} |EP_{1+2}\rangle_{ij} \,. \tag{A.22}$$

Then, we apply projectors a on every site. Their action can be understood through the inductive use of the relation given in Fig. A.2. This process produces a superposition of all possible EP coverings on a square lattice with equal weights (see Fig. A.3). Afterward, to turn EPs into VBs, it is enough to introduce the rotations $i\sigma^y$ on the auxiliary spaces of every sublattice-B site as in the case of AKLT. They act only on the $\mathcal{H}(S = 1/2)$ part of the Hilbert space $\mathcal{H}^{aux}$. Their joint action is again expressed through the unitary U (see Eq. A.12) acting on the physical spins of the sublattice-B sites. This concludes the construction of C4v symmetric iPEPS representation

$$|srRVB\rangle = \sum_{\{s_i\}} \left(\prod_{i\in B} U^i \right) \mathrm{Tr}_{aux} \left(\prod_i a^{s_i} \right) |\{s_i\}\rangle \tag{A.23}$$

of the srRVB state with bond dimension $D = 3$.

B

The core iPEPS algorithms and the details of their numerical implementation

This appendix lists the essential iPEPS algorithms used throughout this work in a form directly amenable to numerical implementation. We also discuss their details, which are crucial for the stability and the robustness of the implementation. If possible, we comment on the ways certain parts of these algorithms can be accelerated by utilizing the cutting-edge methods in the domain of numerical linear algebra.

B.1 CTMRG for single-site C4v symmetric iPEPS

We give the complete CTMRG algorithm in the table Alg. 1. There are two important details concerning the numerical implementation. First, it is numerically favorable to diagonalize just a single $\tilde{C}$ instead of $\tilde{\rho} = \tilde{C}^4$. The resulting projector is identical. Second, the environment tensors have to be normalized after every RG step since their overall magnitude grows, for example, by the norm or by the largest element. Ultimately, there are some points which need to be addressed to complete

Algorithm 1 Compute rank-χ fixed point environment tensors $\{C, T\}$ for iPEPS with the symmetric double-layer tensor A with auxiliary dimension D^2. Complexity: $\mathcal{O}(\chi^3 D^6)$

Require: A, initial $\{C,\ T\}$
  while $\{C,\ T\}$ not converged **do**
    $\tilde{C} \leftarrow CTTA$
    $U, S \leftarrow$ Diagonalization$(\tilde{C})$
    $P \leftarrow$ truncate U to χ columns
    $C \leftarrow P^\dagger \tilde{C} P$
    $T \leftarrow P^\dagger T A P$
    $C,\ T \leftarrow$ normalize $C,\ T$
  end while
  return $\{C,\ T\}$

the introduction of the CTMRG and its application to contract the double-layer network of iPEPS.

- *Initial* $\{C,\ T\}$. The choice of the initial tensors is in principle arbitrary. Common

choices are random tensors or one can use the on-site tensor a of the iPEPS itself. For example, by the contractions $C_{(u\bar{u})(l\bar{l})} = a_{suldr}a^*_{s\bar{u}\bar{l}dr}$ and $T_{(u\bar{u})(l\bar{l})(d\bar{d})} = a_{suldr}a^*_{s\bar{u}\bar{l}\bar{d}r}$ (the remaining index pairs are then fused into individual indices, each with the dimension $d = D^2$). It is possible that for a certain iPEPS the fixed point environment tensors as well as the observables might depend on the initial $\{C, T\}$. In such cases, a more detailed analysis has to be carried out by varying the initial tensors.

- *Convergence criterion.* Whether or not $\{C,\ T\}$ are converged can be assessed by several figures of merit. For example, a reasonable choice could be (but is not limited to) one of the following quantities: Difference between the spectra of C in two consecutive iterations, difference between the one-site density matrices in two consecutive iterations, or difference between particular observables such as energy.

- *Diagonalization of $\tilde{C}$.* To obtain the projector P, we need only the leading χ eigenvectors of $\tilde{C}$. This allows exploiting approximate diagonalization methods such as the Lanczos algorithm and compute just the leading χ eigenvectors. In doing so, the overall complexity of the algorithm can be reduced to $\mathcal{O}(\chi^3 D^4)$ which is a significant improvement. For example, taking an iPEPS with bond dimension $D = 5$ we can expect a speed-up close to a factor ≈ 25.

B.2 Directional CTM for single-site iPEPS

The complete single-site directional CTM algorithm is listed in the table Alg. 2. The algorithm assumes an existence of a function $getR\bar{R}(\theta, E)$. Given a direction θ of the directional move, $getR\bar{R}(\theta, E)$ contracts the two parts of the enlarged network built from the current environment tensors $E = \{C_1, \ldots, C_4, T_1, \ldots, T_4\}$ into the matrices R_θ and $\bar{R}_\theta$ (see Fig. 2.21). As can be seen from the comparison with the CTMRG for C4v symmetric iPEPS (see Sec. B.1), the final algorithm is more complicated. Each move involves an SVD of a matrix $R_\theta\bar{R}_\theta$ of size $\chi D^2 \times \chi D^2$, which sets the overall complexity to $O(\chi^3 D^6)$. Although, the complexity scales with χ and D as in the symmetric case, the prefactor is significantly higher. Contrary to the symmetric case, it now takes four directional moves to effectively grow the system by the equivalent of a single layer of sites, which makes the directional CTM slower by at least a factor of four. Moreover, within each move, the matrices R_θ and $\bar{R}_\theta$ have to be constructed and then multiplied to form $R_\theta\bar{R}_\theta$. Despite being a highly parallelizable task, composed just from several matrix multiplications, all of these operations scale as $\chi^3 D^6$ adding up to the overall prefactor.

Let us comment on some of the technical details of the single-site directional CTM. The initial choice of the tensors E and the convergence criteria raised for the symmetric CTMRG pertain to the directional CTM as well. Besides those, there are several new important considerations:

Algorithm 2 Compute rank-χ fixed point environment $E = \{C_1, C_2, C_3, C_4, T_1, T_2, T_3, T_4\}$ for iPEPS with the double-layer tensor A with auxiliary dimension D^2. Complexity: $\mathcal{O}(\chi^3 D^6)$.

Require: A, $getR\bar{R}(\theta, E)$, initial E
  $m^+ \leftarrow map\{1 \to 2, 2 \to 3, 3 \to 4, 4 \to 1\}$
  $m^- \leftarrow map\{1 \to 4, 2 \to 1, 3 \to 2, 4 \to 3\}$
  while E not converged **do**
    for $\theta = 1$ to 4 **do**
      $R, \bar{R} \leftarrow getR\bar{R}(\theta, E)$
      $\tilde{U}, \tilde{S}, \tilde{V} \leftarrow \mathrm{SVD}(R\bar{R})$
      $U \leftarrow \tilde{U}_{:,\chi}$, $S \leftarrow \tilde{S}_{\chi}$, $V \leftarrow \tilde{V}_{:,\chi}$
      $P \leftarrow RU^{\dagger}S^{-1/2}$
      $\bar{P} \leftarrow \bar{R}VS^{-1/2}$
      $C_{m^+(\theta)} \leftarrow C_{m^+(\theta)}T_{m^+(\theta)}\bar{P}$
      $T_\theta \leftarrow PT_\theta A\bar{P}$
      $C_\theta \leftarrow C_\theta T_{m^-(\theta)}\bar{P}$
      $C_{m^+(\theta)}, T_\theta, C_\theta \leftarrow$ normalize $C_{m^+(\theta)}, T_\theta, C_\theta$
    end for
  end while
  return E

- *Projectors.* The biorthonormal projectors $P_\theta, \bar{P}_\theta$ are constructed with respect to two opposite cuts of a system split in half. For example, for the "left" move ($\theta = 4$) the pair of cuts $\tilde{Q}_2$ and $\tilde{Q}_4$, or for the "up" move ($\theta = 1$) the pair of cuts $\tilde{Q}_1$ and $\tilde{Q}_3$. For an iPEPS with only a slight breaking of the rotational symmetry this choice is balanced in the sense of both halves (matrices) $R_\theta, \bar{R}_\theta$ having roughly the same magnitude. Ideally, when truncating the indices along the selected cut, we would like to maintain biorthonormality to the best possible degree with respect to all other cuts.

  Last important remark on the projector construction is related to the inverse of the square root of the singular values. This step requires an introduction of some cutoff scale ϵ_{SVD} ensuring the stability of the algorithm. Typically, we set $\epsilon_{SVD} \approx 10^{-8}$, square root of machine precision.

- *SVD.* Since the matrices $\{\tilde{Q}_1, \ldots, \tilde{Q}_4\}$ are no longer symmetric, to compute their low-rank approximations SVD has to be used. Analogously to the symmetric CTM, only the leading χ singular pairs are used in the construction of the projectors which allows exploiting truncated SVD methods such as randomized SVD [59], Arnoldi method [60] or Lanczos bidiagonalization [89]. Employing one of these algorithms instead of the full rank SVD reduces the complexity by a substantial factor of D^2 down to $O(\chi^3 D^4)$.

- *Sequence of directional moves.* The sequence in which the directional moves are executed is arbitrary. Moreover, it might be desirable to perform a different num-

ber of horizontal than vertical directional moves within the CTM. If the system is highly anisotropic, having significantly different correlation lengths along vertical and horizontal directions, such a choice might speed up the convergence to the fixed point.

B.3 Directional CTM for iPEPS with unit cells

The preliminary point, before the CTM algorithm is formulated, is the implementation of iPEPS with a unit cell containing more than one on-site tensor. At the level of the unit cell, we specify a map $\mathbf{a} = \{(0,0) \to a_0, (1,0) \to a_1, \ldots\}$ from the coordinates within the unit cell to a unique on-site tensor. To describe how the unit cells are laid out on the square lattice, we define a tiling function $a^{\mathbf{r}}$. It assigns an on-site tensor to each vertex $\mathbf{r} = (x, y)$ by mapping it back to the unit cell as (x', y') and then picking the appropriate tensor

$$a : \mathbf{r} \to \mathbf{a}, \tag{B.1}$$

$$a^{\mathbf{r}} = \mathbf{a}(x'(\mathbf{r}), y'(\mathbf{r})). \tag{B.2}$$

Such a definition is very flexible and allows for iPEPS with tilings going beyond the simple periodic boundary conditions. We give three different examples of the map $\mathbf{a}$, which lead to the various tilings of the square lattice introduced earlier in the Fig. 2.4. They are 2×1 and 2×2 unit cells with PBC and a bipartite iPEPS built with 2×1 unit cell.

$\mathbf{2 \times 1}$	**bipartite**	$\mathbf{2 \times 2}$
$\mathbf{a} = \{(0,0) \to a, (1,0) \to b\}$	$\mathbf{a} = \{(0,0) \to a, (1,0) \to b\}$	$\mathbf{a} = \{(0,0) \to a, (1,0) \to b,$ $(0,1) \to c, (1,1) \to d\}$
$x' = x \bmod 2$ $y' = 0$	$x' = x \bmod 2$ $y' = \lvert y \rvert$	$x' = x \bmod 2$ $y' = y \bmod 2$
$a^{\mathbf{r}} = \mathbf{a}(x', y')$	$a^{\mathbf{r}} = \mathbf{a}((x' + y') \bmod 2, 0)$	$a^{\mathbf{r}} = \mathbf{a}(x', y')$

It is assumed that the modulo operation takes the sign of the argument $\text{sign}(x \bmod y) = \text{sign}(x)$. These three examples are the natural candidates for the ground states of nearest- and next-nearest neighbor spin models. In summary, the whole iPEPS ansatz is specified by a unit cell map $\mathbf{a}$ together with a tiling function $a^{\mathbf{r}}$ as $|iPEPS\rangle \Leftrightarrow (\mathbf{a}, a^{\mathbf{r}})$.

The directional CTM algorithm assumes an existence of two auxiliary functions. First, function $A^{\mathbf{r}}$ simply returns a double-layer tensor given a site $\mathbf{r}$. The second function, $getR\bar{R}(\mathbf{r}, \theta, E)$, is familiar from the previous section but now with an additional positional dependence. Given the site $\mathbf{r}$ and the direction θ of the directional move, it contracts the two parts of the enlarged network, built from the current

environment tensors $\{E^{\mathbf{r}}\}$, into the matrices $R_\theta^{\mathbf{r}}$ and $\bar{R}_\theta^{\mathbf{r}}$ (see Fig. 2.24). The complete algorithm is listed in table Alg. 3. Within every directional move, the inner loop runs over all non-equivalent sites $\mathbf{r}$ given in the unit cell map $\mathbf{a}$. First, all the projectors are constructed and accumulated in two maps P and $\bar{P}$, indexed by the position $\mathbf{r}$. Afterward, the absorption and the truncation is performed.

Algorithm 3 Compute rank-χ fixed point environment $E = \{E^{\mathbf{r}}\}$ for iPEPS specified as $(\mathbf{a}, a^{\mathbf{r}})$ with N unique tensors of bond dimension D within the unit cell. Complexity: $\mathcal{O}(N\chi^3 d^3)$.

Require: $A^{\mathbf{r}}$, $getR\bar{R}(\mathbf{r}, \theta, E)$, initial E
  $m^+ \leftarrow map\{1 \to 2, 2 \to 3, 3 \to 4, 4 \to 1\}$
  $m^- \leftarrow map\{1 \to 4, 2 \to 1, 3 \to 2, 4 \to 3\}$
  $\delta\mathbf{r} \leftarrow map\{1 \to \mathbf{x}, 2 \to \mathbf{y}, 3 \to -\mathbf{x}, 4 \to -\mathbf{y}\}$
  $\Delta \leftarrow map\{1 \to \mathbf{y}, 2 \to -\mathbf{x}, 3 \to -\mathbf{y}, 4 \to \mathbf{x}\}$
  while $E^{\mathbf{r}}$ not converged **do**
    for $\theta = 1$ to 4 **do**
      for $\mathbf{r}$ in $\mathbf{a}$ **do**
        $R, \bar{R} \leftarrow getR\bar{R}(\mathbf{r}, \theta, E)$
        $\tilde{U}, \tilde{S}, \tilde{V} \leftarrow \mathrm{SVD}(R\bar{R})$
        $U \leftarrow \tilde{U}_{:,\chi}$, $S \leftarrow \tilde{S}_\chi$, $V \leftarrow \tilde{V}_{:,\chi}$
        $P(\mathbf{r}) \leftarrow RU^\dagger S^{-1/2}$
        $\bar{P}(\mathbf{r}) \leftarrow \bar{R}VS^{-1/2}$
      end for
      for $\mathbf{r}$ in $\mathbf{a}$ **do**
        $C_{m^+(\theta)}^{\mathbf{r}+\delta\mathbf{r}(\theta)} \leftarrow C_{m^+(\theta)}^{\mathbf{r}+\delta\mathbf{r}(\theta)+\Delta(\theta)} T_{m^+(\theta)}^{\mathbf{r}+\delta\mathbf{r}(\theta)} \bar{P}(\mathbf{r}+\delta\mathbf{r}(\theta))$
        $T_\theta^{\mathbf{r}} \leftarrow P(\mathbf{r}+\delta\mathbf{r}(\theta)) T_\theta^{\mathbf{r}+\Delta(\theta)} A^{\mathbf{r}} \bar{P}(\mathbf{r})$
        $C_\theta^{\mathbf{r}-\delta\mathbf{r}(\theta)} \leftarrow C_\theta^{\mathbf{r}-\delta\mathbf{r}(\theta)+\Delta(\theta)} T_{m^-(\theta)}^{\mathbf{r}-\delta\mathbf{r}(\theta)} \bar{P}(\mathbf{r})$
        $C_{m^+(\theta)}^{\mathbf{r}+\delta\mathbf{r}(\theta)}, T_\theta^{\mathbf{r}}, C_\theta^{\mathbf{r}-\delta\mathbf{r}(\theta)} \leftarrow \mathrm{normalize}\ C_{m^+(\theta)}^{\mathbf{r}+\delta\mathbf{r}(\theta)}, T_\theta^{\mathbf{r}}, C_\theta^{\mathbf{r}-\delta\mathbf{r}(\theta)}$
      end for
    end for
  end while
  return E

All the remarks raised for the previous CTM algorithms are relevant for the CTM extended to unit cells. The complexity scales linearly with the size of the unit cell, which is quite favorable. Still, due to the certain level of independence, there is a room for speeding up its execution

- *Inner loop over non-equivalent sites.* In general, all pairs of projectors needed during the execution of a single directional move are constructed independently in the first inner loop over $\mathbf{r}$. Afterward, in the second inner loop, the truncation of the enlarged environment tensors is done, again independently for different $\mathbf{r}$. Therefore, both inner loops can be parallelized, offering a significant speedup for large unit cells.

B.4 Simple Update: two-site and three-site gates

First, we give a detailed account on the solution of single SU step for two-site bipartite iPEPS evolved under the translationally invariant NN Hamiltonian. It amounts to minimizing the distance f_{SU}, given in Fig. 3.4. Within each step, only a single weight out of four is updated, depending on the bond where the gate u acts. Important preliminary step is to pass from on-site tensors to reduced tensors before doing the SVD. Doing so decreases the complexity of the SU to $O(D^3)$ from $O(D^9)$ and makes the two-site SU extremely fast. The whole process is described in Fig. B.1. Since the last step leading to the new updated on-site tensors requires the inverse of weights, care should be taken, e.g. by imposing a cutoff on the inverted weights.

The complete SU *iteration* consists of SU steps applied to all bonds within the unit cell. For the case considered here, that is, two-site bipartite iPEPS evolved by two-site gates it could be either four SU updates per iteration or eight updates

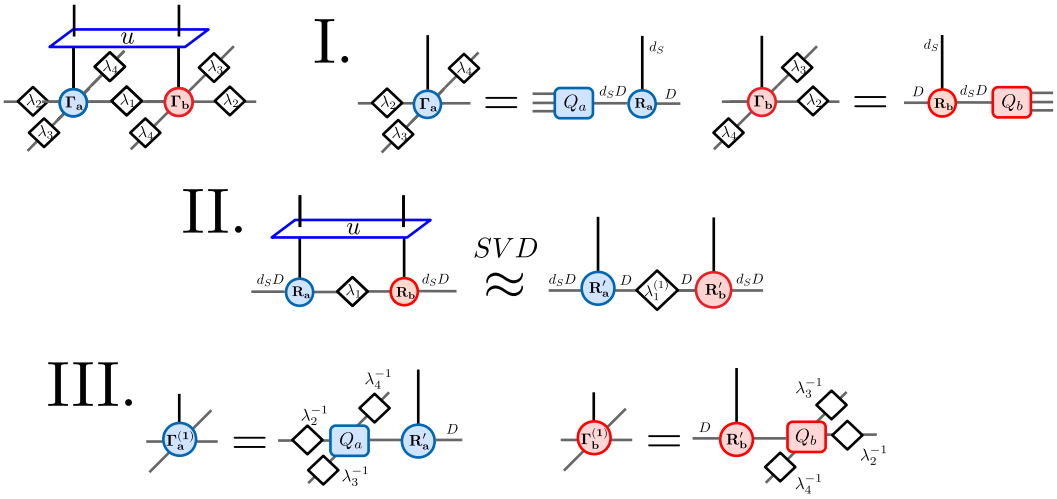


Figure B.1: Single SU step in detail. First, the on-site tensors Γ_a, Γ_b are contracted with the weights and the reduced tensors R_a, R_b are formed from their QR decomposition. Second, the reduced tensors are contracted with the weight λ_1 and the gate u. The resulting tensor, interpreted as a matrix, is decomposed by the SVD and truncated to rank D. The first D left and right singular vectors are taken to be the new reduced tensors R'_a and R'_b respectively. The leading D singular values are normalized such that the dominant singular value is set to unity. They form the new weight $\lambda_1^{(1)}$. The third and the last step is to reconstruct the new on-site tensors $\Gamma_a^{(1)}$ and $\Gamma_b^{(1)}$ by contracting the new reduced tensors with their original "Q" parts and the inverses of the weights. This concludes the single SU step, ending up with the updated $\Gamma_a^{(1)}$, $\Gamma_b^{(1)}$, and $\lambda_1^{(1)}$.

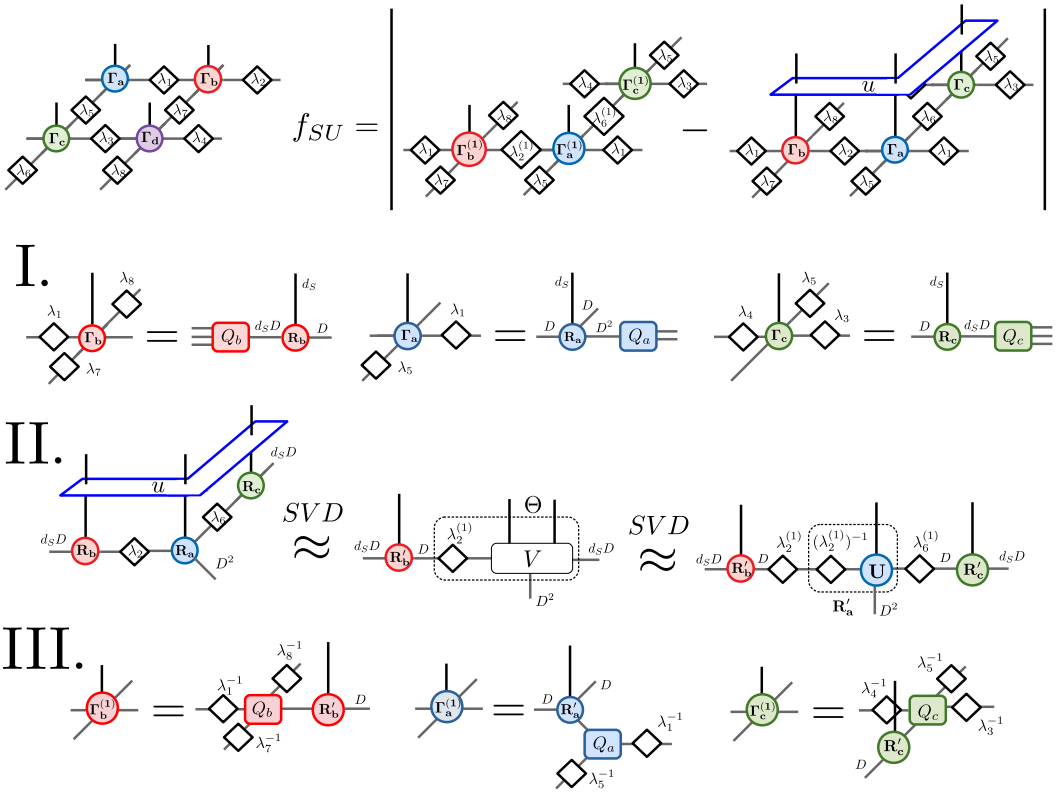


Figure B.2: Single SU step with three-site gate u. Top left: iPEPS with 2×2 unit cell hosting four unique on-site tensors. Top right: distance f_{SU} to be minimized by the SU. Steps (I.-III.): First, the on-site tensors Γ_a, Γ_b, and Γ_c are contracted with the weights and the reduced tensors R_a, R_b, and R_c are formed by their QR decomposition. Second, the reduced tensors are contracted with the weights λ_2, λ_6 and the gate u. The resulting tensor, interpreted as a matrix, is decomposed by the SVD and truncated to rank D. The first D left and right singular vectors are taken to form the new reduced tensor R'_b and a tensor V respectively. The leading D singular values are normalized such that the dominant singular value is set to unity and they form the new weight $\lambda_2^{(1)}$. Tensors V and $\lambda_2^{(1)}$ are contracted into a matrix Θ. Second truncated SVD, this time of Θ, gives tensor U and the new reduced tensor R'_c assembled from the first D left and right singular vectors. The first D singular values form the new weight $\lambda_6^{(1)}$ after normalization. The last new reduced tensor R'_a is obtained from the tensor U and $(\lambda_2^{(1)})^{-1}$, coming from the restored weight. Finally, the updated on-site tensors $\Gamma_a^{(1)}$, $\Gamma_b^{(1)}$, and $\Gamma_c^{(1)}$ are built by contracting the new reduced tensors with their original "Q" parts and the inverses of the weights.

depending on which Trotter decomposition[1] is used [see Eq. (3.5)]. Finally, the optimization is done by repeating the SU iterations until convergence. The difference

[1] in principle the order in which the individual gates are applied might influence the optimization. However, there are no known examples where different order leads to non-negligible differences.

between the weights in the consecutive iterations is often used to establish if the imaginary-time evolution by SU converged.

Importantly, SU was extended to next-nearest neighbor interaction by Corboz [33]. The gates u, which make up the Trotter-decomposed time evolution operator of such Hamiltonians, are acting on three sites. Nevertheless, the overall procedure of the SU step is similar to the two-site gate case, with the crucial difference being the recovery of the updated reduced tensors by the series of SVDs. Taking as an example a 2×2 iPEPS with four tensor in the unit cell, we describe a complete step of SU for three-site gate in Fig. B.2. The complexity of the update is $O(D^6)$, dominated by the construction of the reduced tensors. Each such SU step updates three on-site tensors and two of the eight distinct weights $\{\lambda_1, \ldots, \lambda_8\}$. The complete iteration of SU now consists of sixteen or thirty-two individual SU steps depending on the Trotter decomposition. The number sixteen is given by all the possible ways the gate u can act within the 2×2 unit cell of the iPEPS.

B.5 Full Update: three-site gate

In the main text, the outline of the FU optimization was described. Here, we focus on the concrete step which is the minimization of the approximate distance f_{FU} from Eq. (3.19). We will demonstrate the solution in the case of three-site Trotter gate u applied to iPEPS with 2×2 unit cell containing four tensors. An example relevant for the systems with NNN interactions such as the $J_1 - J_2$ model. The simplification to two-site FU is straightforward.

The impurity problem defined by the f_{FU} can be solved efficiently since the environment of the 2×2 region where the gate u acts is constant. We take this L-shaped gate to act on the three on-site tensors $\{a, b, c\}$, with the tensor a being in between the other two (see Fig. **??**). For other fifteen relative positions of the gate u within the unit cell the procedure is analogous. The f_{FU} is minimized only with respect to the tensors $\{\tilde{a}, \tilde{b}, \tilde{c}\}$, which are supposed to replace the three tensors affected by the gate u. Although there is no unique way to find them the different recipes in the literature often use the Alternating-least squares (ALS) approach. Here, we follow the proposal by Haghshenas [38]. Alternatively one can use a different procedure based on the four isometries between the tensors $\{a, b, c\}$ [37].

The tensors $\{\tilde{a}, \tilde{b}, \tilde{c}\}$ are initialized by the current on-site tensors $\{a, b, c\}$, which is a sound choice since for the small time step ϵ the gate u is very close to identity. Similar to the SU, it is favorable to pass to the reduced tensors by the QR decomposition

$$\begin{aligned} \text{Initialization:} \quad & a = Q_a r_a; \quad b = Q_b r_b; \quad c = Q_c r_c, \\ & \tilde{a} = Q_a \tilde{r}_a; \quad \tilde{r}_a = r_a, \\ & \tilde{b} = Q_b \tilde{r}_b; \quad \tilde{r}_b = r_b, \\ & \tilde{c} = Q_c \tilde{r}_c; \quad \tilde{r}_c = r_c. \end{aligned} \tag{B.3}$$

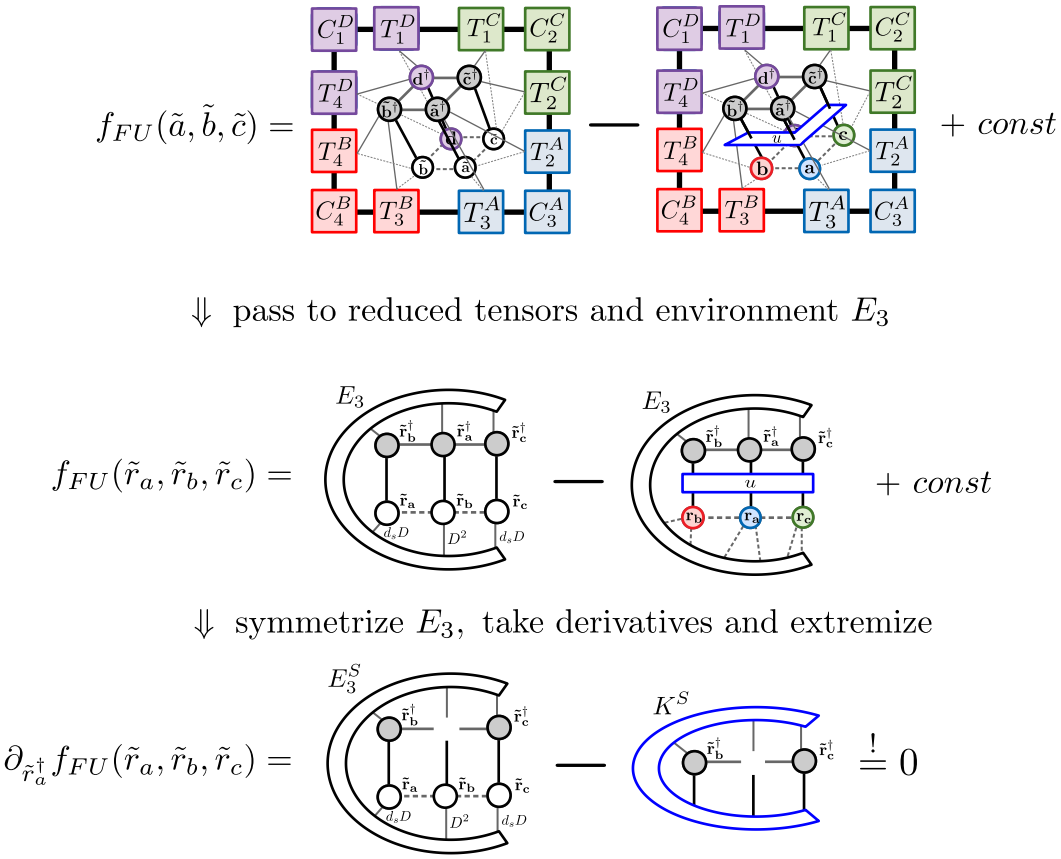


Figure B.3: Minimization of the distance f_{FU} within a single FU step of the imaginary-time evolution of 2×2 iPEPS under the action of NNN Hamiltonian. Starting from the top, the double-layer networks of the norm and the overlap in Eq. (3.19) are approximated using the rank-χ environment defined for iPEPS with the tensors $\{a, b, c, d\}$. We consider only the terms with the conjugates $\{\tilde{r}_a^\dagger, \tilde{r}_b^\dagger, \tilde{r}_c^\dagger\}$. All but the reduced tensors and the gate u are contracted into effective environment E_3 (see text). An instance of the linear system produced by the ALS in terms of constant tensors: The symmetrized environment E_3^S, tensor K^S containing the gate u, and the fixed reduced tensors. This linear system is solved for $\tilde{r}_a^\dagger$.

All the orthonormal bases $\{Q_a, Q_b, Q_c\}$ are taken do be constant and only the reduced tensors $\{\tilde{r}_a, \tilde{r}_b, \tilde{r}_c\}$ will be optimized. The orthonormal bases, the fourth on-site tensor d, $d^\dagger$, and the environment $E_{2\times2}$ are contracted together into a single tensor E_3

$$E_3 := \mathrm{Tr}(E_{2\times2} d^\dagger Q_a^\dagger Q_b^\dagger Q_c^\dagger d Q_a Q_b Q_c), \tag{B.4}$$

which defines the effective environment for the reduced tensors, see Fig. B.3. Both the norm and the overlap figuring in the f_{FU} can be expressed through the tensor E_3, the gate u, and the reduced tensors. In fact, the E_3 can be interpreted as a reduced density matrix and as such it has to be symmetric and non-negative. However, since the CTM provides only approximate environment tensors, the above properties might not be exactly satisfied. Therefore, the E_3 is symmetrized and only

its non-negative part is considered

$$UDU^\dagger = diagonalize\left(\frac{1}{2}(E_3 + E_3^\dagger)\right), \tag{B.5}$$

$$E_3^S := UD^+U^\dagger, \tag{B.6}$$

where D^+ contains only non-negative eigenvalues. Not doing so might lead to instabilities later in the ALS. Using the amended environment E_3^S, the gate u and the reduced tensors $\{r_a, r_b, r_c\}$ in the overlap are contracted into a single tensor K^S (see Fig. B.3). The distance f_{FU} can now be given in a convenient form, ready for the ALS

$$f_{FU}(\tilde{r}_a, \tilde{r}_b, \tilde{r}_c) = \tilde{r}_a^\dagger \tilde{r}_b^\dagger \tilde{r}_c^\dagger E_3^S \tilde{r}_a \tilde{r}_b \tilde{r}_c - \tilde{r}_a^\dagger \tilde{r}_b^\dagger \tilde{r}_c^\dagger K^S - K^S \tilde{r}_a \tilde{r}_b \tilde{r}_c + const. \tag{B.7}$$

With the preliminary part done, the ALS process starts. It minimizes the distance one reduced tensor at the time. At each iteration a derivative of f_{FU} with respect to a single reduced tensor, say $\tilde{r}_a^\dagger$, out of $\{\tilde{r}_a^\dagger, \tilde{r}_b^\dagger, \tilde{r}_c^\dagger\}$ is taken while the other reduced tensors are fixed

$$\partial_{\tilde{r}_a^\dagger} f_{FU} = \tilde{r}_b^\dagger \tilde{r}_c^\dagger E_3^S \tilde{r}_a \tilde{r}_b \tilde{r}_c - \tilde{r}_b^\dagger \tilde{r}_c^\dagger K^S = M\tilde{r}_a - Y \stackrel{!}{=} 0, \tag{B.8}$$

which defines a linear system in terms of the coefficient matrix M and the right hand side Y. The solution of this system gives the new reduced tensor $\tilde{r}_a^\dagger$ that decreases the distance. The tensor should be normalized, for example, by rescaling it with the inverse magnitude of its largest element. This process is repeated alternating the reduced tensors until the f_{FU} converges. Typically, the convergence is established by the difference of distances f_{FU} between ALS iterations reaching a threshold ϵ_f. At last, using the converged reduced tensors, the new on-site tensors $\{\tilde{a}, \tilde{b}, \tilde{c}\}$ are recovered and the minimization of the distance f_{FU} is done.

The linear systems of Eq. (B.8) tend to be very badly conditioned[1]. While for two-site gates a certain form of gauge fixing can improve the conditioning [39, 90], it cannot be easily extended to three-site gates. A robust solution is provided by the pseudoinverse with a cutoff ϵ_{ALS}. The complexity of the ALS then becomes $O(D^{12})$, dominated by the SVD that computes the pseudoinverse. Other iterative methods, such as conjugate gradient, might be used as well. However, the caution must be taken due to the bad conditioning of the matrix M.

[1] negative eigenvalues of the M, coming from the unsymmetrized environment E_3 make the system ill-defined. Taking the symmetrized environment E_3^S still leaves the directions associated with the vanishing eigenvalues free. Hence, the M often does not define a unique solution.

C

Analysis of the Transfer matrices

Static correlation functions are important probes of the quantum state. In this appendix we describe a specific form taken by all the two-point correlation functions of iPEPS. Taking a state $|\psi\rangle$, the two-point correlation function between operators O_1 and O_2 inserted at points $\mathbf{x}$ and $\mathbf{y}$ respectively reads

$$f(\mathbf{x},\mathbf{y})_{O_1O_2} := \langle O_1(\mathbf{x})O_2(\mathbf{y})\rangle = \langle\psi|\, O_1(\mathbf{x})O_2(\mathbf{y})\,|\psi\rangle\,. \tag{C.1}$$

The long distance behavior of the correlation function and its connected part contains information about the phase of the system. The connected part of the correlation function is defined as

$$f^C(\mathbf{x},\mathbf{y})_{O_1O_2} := \langle O_1(\mathbf{x})O_2(\mathbf{y})\rangle - \langle O_1(\mathbf{x})\rangle\langle O_2(\mathbf{y})\rangle. \tag{C.2}$$

In the rest of the appendix, we will restrict the discussion to single-site C4v symmetric iPEPS, hence the correlation functions will depend only on the distance $r = |\mathbf{x}-\mathbf{y}|$.

C.1 Two-point correlation functions of iPEPS

The algebraic form of $f_{O_1O_2}(r)$ is constrained to a be a sum of exponential decays. This fact becomes apparent by considering a transfer matrix $\mathbb{T}$, which naturally appears in the diagram representing the generic two-point correlation function, shown in Fig. C.1. Defining a transfer matrix $\mathbb{T}$ as a contraction of all tensors on the rung of the network, we can express the correlation function as

$$f(r)_{O_1O_2} = \langle L(O_1)|\,\mathbb{T}^{r-1}\,|R(O_2)\rangle\,, \tag{C.3}$$

where the vectors $\langle L(O_1)|\,, |R(O_2)\rangle$ denote the left and the right boundaries of the network in Fig. C.1 and they also contain the operators O_1, O_2. In the rest, we will use just a short-hand notation $\langle L|\,, |R\rangle$. The transfer matrix $\mathbb{T}$, not necessarily symmetric, can be diagonalized as

$$\mathbb{T} = \sum_i |l_i\rangle\,\lambda_i\,\langle r_i|\,, \tag{C.4}$$

with $|l_i\rangle$ and $\langle r_i|$ being left and right eigenvectors respectively, satisfying the orthonormality relation $\langle l_i|r_j\rangle = \delta_{ij}$. Assuming that the leading and the subleading

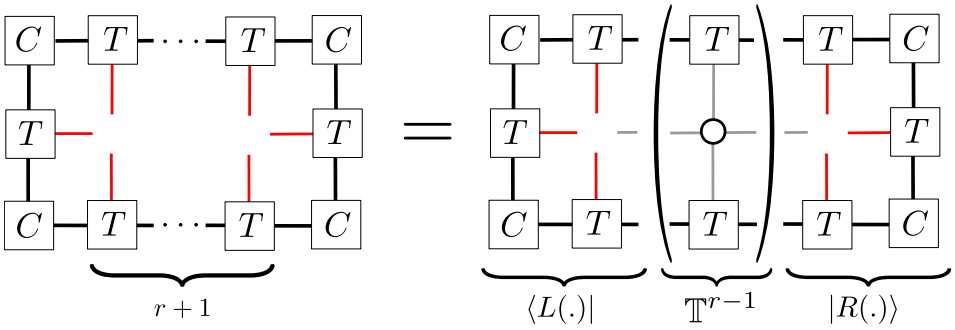


Figure C.1: The environment of two sites separated by a distance $r-1$. Inserting the on-site tensors with operators O_1 and O_2 respectively and contracting the network gives the value of the their correlation function $f(r)_{O_1O_2}$. The network can be seen as one-dimensional ladder with $r+1$ rungs, composed of the left and right boundary vectors $\langle L(.)|$, $|R(.)\rangle$, and $r-1$ transfer matrices $\mathbb{T}$.

eigenvalues are not degenerate, we can rescale the eigenvalues such that λ_0 is unity. Using the eigendecomposition of the transfer matrix the correlation function simplifies to

$$f(r)_{O_1O_2} = \sum_i \lambda_i^{r-1} \langle L|l_i\rangle \langle r_i|R\rangle = \langle L|l_0\rangle \langle r_0|R\rangle + \sum_{i>0} \lambda_i^{r-1} \langle L|l_i\rangle \langle r_i|R\rangle . \quad \text{(C.5)}$$

Leading eigenvectors $|l_0\rangle$ and $\langle r_0|$ of the transfer matrix are equivalent to the left and right boundary $\langle L(\mathbb{1})|, |R(\mathbb{1})\rangle$ with the identity operator inserted. Therefore, the first term in the above expression is simply the disconnected part of the correlation function. Hence, the general form of the connected correlation function is the sum of terms decaying exponentially fast with the distance

$$f^C(r)_{O_1O_2} = \sum_{i>0} \lambda_i^{r-1} \langle L|l_i\rangle \langle r_i|R\rangle . \quad \text{(C.6)}$$

The above form of the connected correlation function can be used to directly extract the correlation length(s) of the iPEPS. By equating the $f^C(r)$ with an exponential decay $exp(-r/\xi)$ and taking a large distance limit, we recover ξ as

$$r \gg 1 : \quad exp(-r/\xi) = |\lambda_1|^r \;\Rightarrow\; \xi = -1/log|\lambda_1|. \quad \text{(C.7)}$$

The correlation length ξ obtained this way is the largest correlation length in the system. For some operators O_1 (O_2) the so-called form factor $\langle L(O_1)|l_1\rangle$ ($\langle r_1|R(O_2)\rangle$) might vanish. In such case, the eigenvalue λ_i for the first non-vanishing set of form-factors sets the correlation length $\xi_i = -1/log|\lambda_i|$ that governs the decay of the corresponding $f(r)^C$. A good example is given by the srRVB state, analyzed in Sec. 2.5, where the dimer-dimer correlations decay substantially slower than the spin-spin correlations.

The environment tensors $\{T\}$, that make up the transfer matrix $\mathbb{T}$, are only approximate. The eigenvalues $\{\lambda\}$ thus depend on the environment bond dimension

χ. This fact has two consequences. First, the finite value of χ induces an effective length scale ξ_χ beyond which all the correlation functions decay exponentially. In effect, even the exact iPEPS with power-law correlations, such as srRVB, will appear to have a finite correlation length. In such cases, the functional form of the connected correlation function can be modeled with the Ornstein-Zernike formula

$$f^C(r) \propto r^{-\eta} exp(-r/\xi_\chi). \tag{C.8}$$

Second, similarly to other observables, to estimate the value of the correlation length(s) in the thermodynamic limit one has to perform their scaling with χ. However, the correlation length tends to converge very slowly with χ and the CTM becomes costly since it scales as $O(\chi^3)$ (for a fixed bond dimension D of iPEPS). Recently, an improved extrapolation method of the correlation length with χ was developed [91]. It is based on the scaling of the gaps between the eigenvalues $\{\lambda\}$ with the environment dimension χ.

Finally, we remark that the above discussion can be extended to iPEPS with unit cells. One has both the horizontal and the vertical transfer matrices $\{\mathbb{T}\}$ given by a product of all rungs spanning the unit cell along those directions.

Bibliography

[1] P. Anderson, "Resonating valence bonds: A new kind of insulator?", Materials Research Bulletin **8**, 153 (1973) https://doi.org/10.1016/0025-5408(73)90167-0.

[2] J. Eisert, M. Cramer, and M. B. Plenio, "Colloquium: Area laws for the entanglement entropy", Rev. Mod. Phys. **82**, 277 (2010) 10.1103/RevModPhys.82.277.

[3] M. Srednicki, "Entropy and area", Phys. Rev. Lett. **71**, 666 (1993) 10.1103/PhysRevLett.71.666.

[4] D. N. Page, "Average entropy of a subsystem", Phys. Rev. Lett. **71**, 1291 (1993) 10.1103/PhysRevLett.71.1291.

[5] S. K. Foong and S. Kanno, "Proof of Page's conjecture on the average entropy of a subsystem", Phys. Rev. Lett. **72**, 1148 (1994) 10.1103/PhysRevLett.72.1148.

[6] S. Sen, "Average Entropy of a Quantum Subsystem", Phys. Rev. Lett. **77**, 1 (1996) 10.1103/PhysRevLett.77.1.

[7] M. B. Hastings and T. Koma, "Spectral Gap and Exponential Decay of Correlations", Communications in Mathematical Physics **265**, 781 (2006) 10.1007/s00220-006-0030-4.

[8] U. Schollwöck, "The density-matrix renormalization group in the age of matrix product states", Annals of Physics **326**, January 2011 Special Issue, 96 (2011) https://doi.org/10.1016/j.aop.2010.09.012.

[9] F. Verstraete and J. I. Cirac, "Matrix product states represent ground states faithfully", Phys. Rev. B **73**, 094423 (2006) 10.1103/PhysRevB.73.094423.

[10] M. B. Hastings, "An area law for one-dimensional quantum systems", Journal of Statistical Mechanics: Theory and Experiment **2007**, P08024 (2007) 10.1088/1742-5468/2007/08/p08024.

[11] F. Verstraete and J. I. Cirac, "Renormalization algorithms for Quantum-Many Body Systems in two and higher dimensions", arXiv:cond-mat/0407066 (2004).

[12] F. Verstraete, M. M. Wolf, D. Perez-Garcia, and J. I. Cirac, "Criticality, the Area Law, and the Computational Power of Projected Entangled Pair States", Phys. Rev. Lett. **96**, 220601 (2006) 10.1103/PhysRevLett.96.220601.

[13] J. Jordan, R. Orús, G. Vidal, F. Verstraete, and J. I. Cirac, "Classical Simulation of Infinite-Size Quantum Lattice Systems in Two Spatial Dimensions", Phys. Rev. Lett. **101**, 250602 (2008) 10.1103/PhysRevLett.101.250602.

[14] M. M. Wolf, "Violation of the Entropic Area Law for Fermions", Phys. Rev. Lett. **96**, 010404 (2006) 10.1103/PhysRevLett.96.010404.

[15] M. A. Metlitski and T. Grover, "Entanglement Entropy of Systems with Spontaneously Broken Continuous Symmetry", arXiv:1112.5166 (2011).

[16] R. J. Baxter, "Corner transfer matrices in statistical mechanics", arXiv:cond-mat/0611167 (2006) 10.1088/1751-8113/40/42/S05.

[17] R. J. Baxter, "Corner transfer matrices of the eight-vertex model. II. The Ising model case", Journal of Statistical Physics **17**, 1 (1977) 10.1007/BF01089373.

[18] S. R. White, "Density matrix formulation for quantum renormalization groups", Phys. Rev. Lett. **69**, 2863 (1992) 10.1103/PhysRevLett.69.2863.

[19] T. Nishino and K. Okunishi, "Corner Transfer Matrix Renormalization Group Method", Journal of the Physical Society of Japan **65**, 891 (1996) 10.1143/JPSJ.65.891.

[20] T. Nishino and K. Okunishi, "A Density Matrix Algorithm for 3D Classical Models", Journal of the Physical Society of Japan **67**, 3066 (1998) 10.1143/JPSJ.67.3066.

[21] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki, "Valence bond ground states in isotropic quantum antiferromagnets", Comm. Math. Phys. **115**, 477 (1988).

[22] G. Baskaran, Z. Zou, and P. Anderson, "The resonating valence bond state and high-Tc superconductivity — A mean field theory", Solid State Communications **63**, 973 (1987) https://doi.org/10.1016/0038-1098(87)90642-9.

[23] L. Vanderstraeten, M. Mariën, F. Verstraete, and J. Haegeman, "Excitations and the tangent space of projected entangled-pair states", Phys. Rev. B **92**, 201111 (2015) 10.1103/PhysRevB.92.201111.

[24] L. Vanderstraeten, J. Haegeman, P. Corboz, and F. Verstraete, "Gradient methods for variational optimization of projected entangled-pair states", Phys. Rev. B **94**, 155123 (2016) 10.1103/PhysRevB.94.155123.

[25] R. Orús and G. Vidal, "Simulation of two-dimensional quantum systems on an infinite lattice revisited: Corner transfer matrix for tensor contraction", Phys. Rev. B **80**, 094403 (2009) 10.1103/PhysRevB.80.094403.

[26] P. Corboz, T. M. Rice, and M. Troyer, "Competing States in the t-J Model: Uniform d-Wave State versus Stripe State (Supplemental Material)", Phys. Rev. Lett. **113**, 046402 (2014) 10.1103/PhysRevLett.113.046402.

[27] Y.-K. Huang, "Biorthonormal transfer-matrix renormalization-group method for non-Hermitian matrices", Phys. Rev. E **83**, 036702 (2011) 10.1103/PhysRevE.83.036702.

[28] M. T. Fishman, L. Vanderstraeten, V. Zauner-Stauber, J. Haegeman, and F. Verstraete, "Faster methods for contracting infinite two-dimensional tensor networks", Phys. Rev. B **98**, 235148 (2018) 10.1103/PhysRevB.98.235148.

[29] B. Ponsioen, S. S. Chung, and P. Corboz, "Period 4 stripe in the extended two-dimensional Hubbard model", arXiv:1907.01909 (2019).

[30] J. Haegeman, J. I. Cirac, T. J. Osborne, I. Pi žorn, H. Verschelde, and F. Verstraete, "Time-Dependent Variational Principle for Quantum Lattices", Phys. Rev. Lett. **107**, 070601 (2011) 10.1103/PhysRevLett.107.070601.

[31] R. Orús, "A practical introduction to tensor networks: Matrix product states and projected entangled pair states", Annals of Physics **349**, 117 (2014) https://doi.org/10.1016/j.aop.2014.06.013.

[32] H. C. Jiang, Z. Y. Weng, and T. Xiang, "Accurate Determination of Tensor Network State of Quantum Lattice Models in Two Dimensions", Phys. Rev. Lett. **101**, 090603 (2008) 10.1103/PhysRevLett.101.090603.

[33] P. Corboz, J. Jordan, and G. Vidal, "Simulation of fermionic lattice models in two dimensions with projected entangled-pair states: Next-nearest neighbor Hamiltonians", Phys. Rev. B **82**, 245119 (2010) 10.1103/PhysRevB.82.245119.

[34] H. Kalis, D. Klagges, R. Orús, and K. P. Schmidt, "Fate of the cluster state on the square lattice in a magnetic field", Phys. Rev. A **86**, 022317 (2012) 10.1103/PhysRevA.86.022317.

[35] G. Vidal, "Classical Simulation of Infinite-Size Quantum Lattice Systems in One Spatial Dimension", Phys. Rev. Lett. **98**, 070201 (2007) 10.1103/PhysRevLett.98.070201.

[36] W. Li, J. von Delft, and T. Xiang, "Efficient simulation of infinite tree tensor network states on the Bethe lattice", Phys. Rev. B **86**, 195137 (2012) 10.1103/PhysRevB.86.195137.

[37] P. Corboz and F. Mila, "Tensor network study of the Shastry-Sutherland model in zero magnetic field", Phys. Rev. B **87**, 115144 (2013) 10.1103/PhysRevB.87.115144.

[38] R. Haghshenas and D. N. Sheng, "$U(1)$-symmetric infinite projected entangled-pair states study of the spin-1/2 square J_1-J_2 Heisenberg model", Phys. Rev. B **97**, 174408 (2018) 10.1103/PhysRevB.97.174408.

[39] H. N. Phien, J. A. Bengua, H. D. Tuan, P. Corboz, and R. Orús, "Infinite projected entangled pair states algorithm improved: Fast full update and gauge fixing", Phys. Rev. B **92**, 035142 (2015) 10.1103/PhysRevB.92.035142.

[40] H. N. Phien, I. P. McCulloch, and G. Vidal, "Fast convergence of imaginary time evolution tensor network algorithms by recycling the environment", Phys. Rev. B **91**, 115137 (2015) 10.1103/PhysRevB.91.115137.

[41] P. Corboz, "Variational optimization with infinite projected entangled-pair states", Phys. Rev. B **94**, 035133 (2016) 10.1103/PhysRevB.94.035133.

[42] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, "Automatic Differentiation in Machine Learning: a Survey", Journal of Machine Learning Research **18**, 1 (2018).

[43] A. Griewank and A. Walther, *Evaluating Derivatives*, Second (Society for Industrial and Applied Mathematics, 2008), 10.1137/1.9780898717761.

[44] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic Differentiation in PyTorch", in NIPS Autodiff Workshop (2017).

[45] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Y. Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, Software available from tensorflow.org, 2015.

[46] H.-J. Liao, J.-G. Liu, L. Wang, and T. Xiang, "Differentiable Programming Tensor Networks", Phys. Rev. X **9**, 031041 (2019) 10.1103/PhysRevX.9.031041.

[47] M. Giles, "An extended collection of matrix derivative results for forward and reverse mode algorithmic differentiation", Tech. Rep. (2008).

[48] C. Hubig, "Use and implementation of autodifferentiation in tensor network methods with complex scalars", arXiv:1907.13422 (2019).

[49] Z.-Q. Wan and S.-X. Zhang, *Automatic Differentiation for Complex Valued SVD*, 2019.

[50] H. J. Liao, Z. Y. Xie, J. Chen, Z. Y. Liu, H. D. Xie, R. Z. Huang, B. Normand, and T. Xiang, "Gapless Spin-Liquid Ground State in the $S = 1/2$ Kagome Antiferromagnet", Phys. Rev. Lett. **118**, 137202 (2017) 10.1103/PhysRevLett.118.137202.

[51] I. Niesen and P. Corboz, "A tensor network study of the complete ground state phase diagram of the spin-1 bilinear-biquadratic Heisenberg model on the square lattice", SciPost Phys. **3**, 030 (2017) 10.21468/SciPostPhys.3.4.030.

[52] I. Niesen and P. Corboz, "Ground-state study of the spin-1 bilinear-biquadratic Heisenberg model on the triangular lattice using tensor networks", Phys. Rev. B **97**, 245146 (2018) 10.1103/PhysRevB.97.245146.

[53] R. Haghshenas, W.-W. Lan, S.-S. Gong, and D. N. Sheng, "Quantum phase diagram of spin-1 J_1-J_2 Heisenberg model on the square lattice: An infinite projected entangled-pair state and density matrix renormalization group study", Phys. Rev. B **97**, 184436 (2018) 10.1103/PhysRevB.97.184436.

[54] M. Rader and A. M. Läuchli, "Finite Correlation Length Scaling in Lorentz-Invariant Gapless iPEPS Wave Functions", Phys. Rev. X **8**, 031030 (2018) 10.1103/PhysRevX.8.031030.

[55] P. Corboz, P. Czarnik, G. Kapteijns, and L. Tagliacozzo, "Finite Correlation Length Scaling with Infinite Projected Entangled-Pair States", Phys. Rev. X **8**, 031031 (2018) 10.1103/PhysRevX.8.031031.

[56] J. Hasik and F. Becca, "Optimization of infinite projected entangled pair states: The role of multiplets and their breaking", Phys. Rev. B **100**, 054429 (2019) 10.1103/PhysRevB.100.054429.

[57] P. W. Anderson, "An Approximate Quantum Theory of the Antiferromagnetic Ground State", Phys. Rev. **86**, 694 (1952) 10.1103/PhysRev.86.694.

[58] H. F. Song, N. Laflorencie, S. Rachel, and K. Le Hur, "Entanglement entropy of the two-dimensional Heisenberg antiferromagnet", Phys. Rev. B **83**, 224410 (2011) 10.1103/PhysRevB.83.224410.

[59] S. Voronin and P.-G. Martinsson, "RSVDPACK: An implementation of randomized algorithms for computing the singular value, interpolative, and CUR decompositions of matrices on multi-core and GPU architectures", arXiv:1502.05366 (2015).

[60] *The ARPACK-NG project*, http://github.com/opencollab/arpack-ng.

[61] J. Hasik and A. Sartori, *pi-peps: A C++ library built on top of ITensor for running iPEPS simulations of two dimensional spin systems*, https://github.com/jurajHasik/pi-peps.

[62] E. Stoudenmire, *ITensor*, https://github.com/ITensor/ITensor.

[63] U. Gerber, C. P. Hofmann, F.-J. Jiang, M. Nyfeler, and U.-J. Wiese, "The constraint effective potential of the staggered magnetization in an antiferromagnet", Journal of Statistical Mechanics: Theory and Experiment **2009**, P03021 (2009) 10.1088/1742-5468/2009/03/p03021.

[64] A. W. Sandvik and H. G. Evertz, "Loop updates for variational and projector quantum Monte Carlo simulations in the valence-bond basis", Phys. Rev. B **82**, 024407 (2010) 10.1103/PhysRevB.82.024407.

[65] M. Mambrini, R. Orús, and D. Poilblanc, "Systematic construction of spin liquids on the square lattice from tensor networks with SU(2) symmetry", Phys. Rev. B **94**, 205124 (2016) 10.1103/PhysRevB.94.205124.

[66] S. Sachdev, "Quantum Criticality: Competing Ground States in Low Dimensions", Science **288**, 475 (2000) 10.1126/science.288.5465.475.

[67] M. Matsumoto, C. Yasuda, S. Todo, and H. Takayama, "Ground-state phase diagram of quantum Heisenberg antiferromagnets on the anisotropic dimerized square lattice", Phys. Rev. B **65**, 014407 (2001) 10.1103/PhysRevB.65.014407.

[68] L. Capriotti and F. Becca, "Quantum phase transition in coupled spin ladders", Phys. Rev. B **65**, 092406 (2002) 10.1103/PhysRevB.65.092406.

[69] S. Sachdev, *Quantum Phase Transitions*, 2nd ed. (Cambridge University Press, 2011), 10.1017/CBO9780511973765.

[70] S. Jiang and Y. Ran, "Symmetric tensor networks and practical simulation algorithms to sharply identify classes of quantum phases distinguishable by short-range physics", Phys. Rev. B **92**, 104414 (2015) 10.1103/PhysRevB.92.104414.

[71] D. Poilblanc and M. Mambrini, "Quantum critical phase with infinite projected entangled paired states", Phys. Rev. B **96**, 014414 (2017) 10.1103/PhysRevB.96.014414.

[72] A. W. Sandvik, "Evidence for Deconfined Quantum Criticality in a Two-Dimensional Heisenberg Model with Four-Spin Interactions", Phys. Rev. Lett. **98**, 227202 (2007) 10.1103/PhysRevLett.98.227202.

[73] B. Bauer, P. Corboz, R. Orús, and M. Troyer, "Implementing global Abelian symmetries in projected entangled-pair state algorithms", Phys. Rev. B **83**, 125106 (2011) 10.1103/PhysRevB.83.125106.

[74] S. Singh, R. N. C. Pfeifer, and G. Vidal, "Tensor network states and algorithms in the presence of a global U(1) symmetry", Phys. Rev. B **83**, 115125 (2011) 10.1103/PhysRevB.83.115125.

[75] P. Chandra and B. Doucot, "Possible spin-liquid state at large S for the frustrated square Heisenberg lattice", Phys. Rev. B **38**, 9335 (1988) 10.1103/PhysRevB.38.9335.

[76] M. P. Gelfand, R. R. P. Singh, and D. A. Huse, "Zero-temperature ordering in two-dimensional frustrated quantum Heisenberg antiferromagnets", Phys. Rev. B **40**, 10801 (1989) 10.1103/PhysRevB.40.10801.

[77] V. N. Kotov, J. Oitmaa, O. P. Sushkov, and Z. Weihong, "Low-energy singlet and triplet excitations in the spin-liquid phase of the two-dimensional J_1-J_2 model", Phys. Rev. B **60**, 14613 (1999) 10.1103/PhysRevB.60.14613.

[78] J. Richter and J. Schulenburg, "The spin-1/2 J1–J2 Heisenberg antiferromagnet on the square lattice:Exact diagonalization for N=40 spins", The European Physical Journal B **73**, 117 (2010) 10.1140/epjb/e2009-00400-4.

[79] H.-C. Jiang, H. Yao, and L. Balents, "Spin liquid ground state of the spin-$\frac{1}{2}$ square J_1-J_2 Heisenberg model", Phys. Rev. B **86**, 024424 (2012) 10.1103/PhysRevB.86.024424.

[80] S.-S. Gong, W. Zhu, D. N. Sheng, O. I. Motrunich, and M. P. A. Fisher, "Plaquette Ordered Phase and Quantum Phase Diagram in the Spin-$\frac{1}{2}$ J_1-J_2 Square Heisenberg Model", Phys. Rev. Lett. **113**, 027201 (2014) 10.1103/PhysRevLett.113.027201.

[81] L. Wang and A. W. Sandvik, "Critical Level Crossings and Gapless Spin Liquid in the Square-Lattice Spin-1/2 $J_1 - J_2$ Heisenberg Antiferromagnet", Phys. Rev. Lett. **121**, 107202 (2018) 10.1103/PhysRevLett.121.107202.

[82] W.-J. Hu, F. Becca, A. Parola, and S. Sorella, "Direct evidence for a gapless Z_2 spin liquid by frustrating Néel antiferromagnetism", Phys. Rev. B **88**, 060402 (2013) 10.1103/PhysRevB.88.060402.

[83] S. Morita, R. Kaneko, and M. Imada, "Quantum Spin Liquid in Spin 1/2 J1–J2 Heisenberg Model on Square Lattice: Many-Variable Variational Monte Carlo Study Combined with Quantum-Number Projections", Journal of the Physical Society of Japan **84**, 024720 (2015) 10.7566/JPSJ.84.024720.

[84] F. Ferrari and F. Becca, "Spectral signatures of fractionalization in the frustrated Heisenberg model on the square lattice", Phys. Rev. B **98**, 100405 (2018) 10.1103/PhysRevB.98.100405.

[85] W.-Y. Liu, S. Dong, C. Wang, Y. Han, H. An, G.-C. Guo, and L. He, "Gapless spin liquid ground state of the spin-$\frac{1}{2}$ $J_1 - J_2$ Heisenberg model on square lattices", Phys. Rev. B **98**, 241109 (2018) 10.1103/PhysRevB.98.241109.

[86] D. Poilblanc, M. Mambrini, and S. Capponi, "Critical colored-RVB states in the frustrated quantum Heisenberg model on the square lattice", arXiv:1907.03678 (2019).

[87] J. Hasik and G. B. Mbeng, *tn-torch: A differentiable tensor network library for two-dimensional lattice models*, to be released soon.

[88] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization", Mathematical Programming **45**, 503 (1989) 10.1007/BF01589116.

[89] R. M. Larsen, *PROPACK - Software for large and sparse SVD calculations*, http://sun.stanford.edu/~rmunk/PROPACK.

[90] M. Lubasch, J. I. Cirac, and M.-C. Bañuls, "Unifying projected entangled pair state contractions", New Journal of Physics **16**, 033014 (2014) 10.1088/1367-2630/16/3/033014.

[91] M. M. Rams, P. Czarnik, and L. Cincio, "Precise Extrapolation of the Correlation Function Asymptotics in Uniform Tensor Network States with Application to the Bose-Hubbard and XXZ Models", Phys. Rev. X **8**, 041033 (2018) 10.1103/PhysRevX.8.041033.