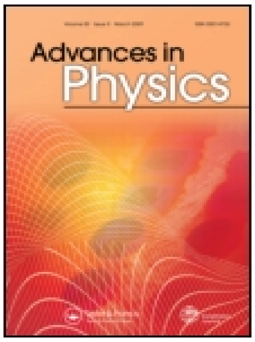

Advances in Physics

ISSN: 0001-8732 (Print) 1460-6976 (Online) Journal homepage: https://www.tandfonline.com/loi/tadp20

Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems

F. Verstraete , V. Murg & J.I. Cirac

To cite this article: F. Verstraete , V. Murg & J.I. Cirac (2008) Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems, Advances in Physics, 57:2, 143-224, DOI: 10.1080/14789940801912366

To link to this article: https://doi.org/10.1080/14789940801912366

Published online: 07 Jul 2008.

Submit your article to this journal

Article views: 5363

View related articles

Citing articles: 695 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tadp20

Advances in Physics
Vol. 57, No. 2, March–April 2008, 143–224

Taylor & Francis
Taylor & Francis Group

Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems

F. Verstraete[a]*, V. Murg[b] and J.I. Cirac[b]

[a]*Faculty of Physics, University of Vienna, Austria*
[b]*Max-Planck-Institut für Quantenoptik, Garching, D-85748, Germany*

(*Received 9 May 2007; final version received 10 January 2008*)

This article reviews recent developments in the theoretical understanding and the numerical implementation of variational renormalization group methods using matrix product states and projected entangled pair states.

Keywords: matrix product states; projected entangled pair states; numerical renormalization group; density matrix renormalization group; entanglement; quantum spin chains; strongly correlated quantum systems

Contents

PAGE
1. Introduction 144
2. Spin systems: general features 146
3. Wilson's NRG method 152
4. MPSs and ground states of spin chains 153
    4.1. Construction and calculus of MPSs 153
        4.1.1. The AKLT model 153
        4.1.2. MPSs 155
        4.1.3. Calculus of MPSs 156
        4.1.4. Generalization of MPSs 160
    4.2. Reformulating NRG methods as variational methods in the class of MPSs 163
    4.3. DMRG 165
        4.3.1. Excitations and spectral functions 168
    4.4. DMRG and PBCs 171
5. Time evolution using MPSs 173
    5.1. Variational formulation of time evolution with MPSs 174
        5.1.1 Time-evolving block-decimation 177
        5.1.2 Exchanging time and space 177
    5.2. Finding ground states by imaginary time evolution 178
        5.2.1. Infinite spin chains 179

---

*Corresponding author. Email: fverstraete@gmail.com

ISSN 0001–8732 print/ISSN 1460–6976 online
© 2008 Taylor & Francis
DOI: 10.1080/14789940801912366
http://www.informaworld.com

6. MPOs 181
    6.1. Finite-temperature systems 182
    6.2. Random quantum spin systems 183
    6.3. Classical partition functions and thermal quantum states 185
    6.4. Density of states 191
7. PEPSs and ground states of two-dimensional quantum spin systems 191
    7.1. Construction and calculus of PEPSs 192
    7.2. Calculus of PEPS 194
    7.3. Variational method with PEPSs 196
    7.4. Time evolution with PEPS 197
        7.4.1. Examples 201
        7.4.2. PEPSs and fermions 204
    7.5. PEPSs on infinite lattices 205
8. Conclusion 206
Acknowledgements 207
References 207
Appendix A. Local reduced density operators of spin systems 211
Appendix B. MPSs represent ground states faithfully 213
Appendix C. Matlab code 216
    C.1. Minimization of the energy 216
    C.2. Time evolution 218
    C.3. Auxiliary functions 219
    C.4. Examples 222

1. Introduction

One of the biggest challenges in physics is developing efficient methods for simulating Hamiltonians: there is a well-developed understanding of the basic laws of quantum mechanics, but the crucial understanding of the collective behaviour of interacting elementary particles and of emerging phenomena relies on our ability to approximate or simulate the corresponding equations. Simulation is particularly relevant in the context of strongly interacting quantum systems, where conventional perturbation theory fails. Popular methods to simulate such systems are the numerical renormalization group (NRG) algorithms by Wilson [1] and White [2,3] (see also [4,5]) and quantum Monte Carlo methods. Both of these have been used with great success, but they also have severe limitations: the NRG and density matrix renormalization group (DMRG) algorithms only work in cases where the original Hamiltonian can be mapped to a local Hamiltonian defined on a one-dimensional chain, while Monte Carlo methods suffer from the so-called sign problem [6] which makes them inappropriate for the description of fermionic and frustrated quantum systems. Very recently, however, new insights coming from the field of quantum information and entanglement theory have shown how those NRG methods can be generalized to higher dimensions, to finite temperature, to random systems, etc. [7–16]. These developments have the potential to revolutionize the way quantum many-body systems can be simulated and understood, which will prove essential in light of the ongoing miniaturization of electronic devices and of fundamental problems such as the study of the role of the Hubbard model in high T_c superconductivity.

We focus on those new developments in this review. In particular, we focus on the simulation of quantum lattice systems of spins and/or fermions. The Hamiltonians arising in this context can be thought of as effective Hamiltonians of more complicated systems that capture the physics for the relevant low-energy degrees of freedom. Studying quantum spin systems has a long and very interesting history, starting with Dirac and Heisenberg in the 1920s, who proposed the so-called Heisenberg model [17,18] as being illustrative of the basic mechanism giving rise to magnetism. The simulation of quantum spin systems and fermionic systems on a lattice, however, has turned out to be extremely complicated. Consider, for example, the Hubbard model in two dimensions which represents the simplest possible model of fermions hopping on a lattice and exhibiting on-site interactions: despite considerable efforts because of the connection with high T_c superconductivity, it is absolutely unclear how the low-energy wavefunctions look in the relevant parameter range. This is the most notable illustration of the biggest bottleneck in condensed matter theory: numerical simulation of effective simple Hamiltonians. The new tools described in this review can be seen as a step towards filling that gap.

This review is organized as follows. In the first, rather technical, Section, we review the basic properties of quantum spin systems and provide intuitive justification for the numerical tools used later. This Section can be skipped by those who are more interested in the variational numerical renormalization methods and less interested in the underlying theory; however, one of the major advantages of the methods is precisely that there is a solid theoretical justification for using them. In Section 2, we review the NRG method as introduced by Wilson, and show how this gives rise to the concept of matrix product states (MPSs). Section 3 discusses the features of MPSs, and goes on to applications on spin chains such as a reformulation of the DMRG and generalizations that allow us to treat excited states. Section 4 is devoted to the issue of real and imaginary time evolution of one-dimensional quantum spin systems, and Section 5 is devoted to the concept of matrix product operators (MPOs) leading to extensions of DMRG to finite-temperature systems, to random spin chains and to the simulation of two-dimensional classical partition functions. In Section 6, we show how all of these ideas can be generalized to the two-dimensional case by introducing the class of projected entangled pair states (PEPSs), and we discuss the applicability and limitations of these methods. Finally, we discuss the convex set of reduced density operators of translational-invariant states in Appendix A, the relation between block entropy and the accuracy of a MPS approximation in Appendix B, and we include some explicit Matlab programs of MPS algorithms in Appendix C.

At this point, we would like to warn the reader that this is not a review of DMRG methods and possible extensions. In fact, there already exist very good reviews on that topic (see, for example [4,5]), where the important progress experienced by DMRG-like methods is discussed extensively. Here we offer a review of the new methods that have been introduced during the last few years, and that have evolved from ideas in the context of quantum information theory. To make the review self-contained and uniform, we focus mainly on the methods introduced by ourselves and collaborators, and also touch upon the approach advocated by Vidal. The main reason is that these approaches can be viewed as (generalized) variational techniques over sets of entangled states, and thus we can present the vast variety of algorithms, whose description is thus-far scattered in several publications, in a unified form. Furthermore, we also discuss some issues which have not been published so far.

2. Spin systems: general features

One of the main characteristics of quantum mechanics is that the underlying Hilbert space is endowed with a tensor product structure: the Hilbert space of two interacting systems is given by the tensor product space of the two individual ones. This structure of the Hilbert space is a direct consequence of the superposition principle, and opens up the possibility of entanglement and new phases of matter. This simple tensor product structure, however, makes it very clear what the main obstacle is in developing a general theory of many-body quantum systems: the size of the Hilbert space grows exponentially in the number of basis constituents, and hence we would, in principle, need exponentially many variables to specify the wavefunction of a N-particle system. However, it is *a priori* not clear whether nature can fully exploit and explore these vast territories of Hilbert space, because another main characteristic of quantum mechanics is that interactions always seem to happen locally and only between a few bodies.

As it turns out, all physical states live on a tiny submanifold of Hilbert space. For example, consider a quantum system of N spin-1/2s on a cubic lattice and all spins polarized in the z-direction, and let us pick a random state in the 2^N-dimensional Hilbert space according to the unitarily invariant Haar measure. Let us furthermore address the problem of finding a lower bound on the time it would take to evolve the polarized state into a state that has an overlap with the random one that is not exponentially small, and this with local interactions. Using tools developed in the context of quantum information theory, one can show that this lower bound scales exponentially in the number of spins [19]: any evolution over a time that scales only polynomially in the number of spins would only allow us to obtain an exponentially small overlap with the chosen random state. If we were, for example, to consider a system of a few hundred spins and assume realistic units for the norm of local Hamiltonian terms, it would already take much longer than the lifetime of the universe to come close to any random point in its Hilbert space. This shows that almost all points in the Hilbert space of a many-body quantum system are unphysical as they are inaccessible: all physical states, that is, all states that can ever be created, live on a tiny submanifold of measure zero.

This concept opens up a very interesting perspective in the context of the description of quantum many-body systems, as there might exist an efficient parametrization of such a submanifold that would provide the natural language for describing those systems. In the context of many-body quantum physics, one is furthermore mainly interested in describing the low-energy sector of local Hamiltonians, and as we discuss later, this puts many extra constraints on the allowed wavefunctions.

As a trivial example of such a submanifold, let us consider a system of $N\to\infty$ spins that all interact with each other via a permutation invariant two-body Hamiltonian such as the Heisenberg interaction [17,18]. Note that there is always a ground state that exhibits the same symmetry as the Hamiltonian. As a consequence of the quantum de-Finnetti theorem [20] one can show that the manifold of all reduced density operators with permutation symmetry is exactly the set of all separable states $\rho = \sum_i p_i \rho_i^{\otimes N}$ when $N\to\infty$. Ground states correspond to the extreme points of this set, which are exactly the product states that have no entanglement. All ground states of permutational-invariant systems therefore lie on the submanifold of separable states which indeed have an efficient representation (only N vectors, one for each individual spin, have to be specified); this statement is equivalent to saying that mean-field theory becomes exact in the thermodynamic limit. This can also be understood in the context of the monogamy property of entanglement [21,22]: a spin has only a finite entanglement susceptibility, and if it has to share the entanglement with infinitely many other spins, then the amount of entanglement between two spins will go to zero. (In the case of a translational- and rotational-invariant system on a regular lattice, the finite versions of the de-Finnetti theorem [23,24] allow one to find lower bounds on the distance of the reduced two-body density operators to the separable operators; a scaling of the form $1/c$ is obtained where c is the coordination number. See also Appendix A.)

In the more general situation when no permutation symmetry is present, only a smaller symmetry group such as the group of translations, the characterization of the relevant manifold is much harder. In contrast to the case of permutation symmetry, where every pure state of N spin s systems can be written as a linear combination of at most $N^{2s}/(2s)!$ Dicke states (Dicke states are well studied in the context of quantum optics, and are obtained by taking the linear superposition of all permutations of a pure separable state in the standard basis: in the case of spin-1/2, you can label them by their expectation value of $\sum_i S_i^z$; see, e.g. [25] for an extensive discussion of their entanglement properties), the size of a complete basis of translational-invariant states is exponentially increasing as $(2s+1)^N/N$, and hence the manifold of interest has exponentially many parameters. However, ground states of local Hamiltonians have many more non-generic properties, most notably the fact that they have extremal local properties such as the minimization of the energy: as the energy is only dependent on the local properties, and the ground state is determined by the condition that its energy is extremal, ground states have extremal local properties and the global properties only emerge to allow for these local properties to be extremal. As an example, let us consider a spin-1/2 antiferromagnetic chain with an associated Hamiltonian

$$\mathcal{H}_{\text{Heis}} = \sum_{\langle i,j \rangle} \vec{S}_i \vec{S}_j$$

where the notation $\langle i,j \rangle$ denotes the sum over nearest neighbours. Each individual term in the Hamiltonian corresponds to an exchange interaction and would be minimized if spins i and j are in the singlet state

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle),$$

however, due to the monogamy or frustration properties of entanglement, a spin-1/2 cannot be in a singlet state with more than one neighbour. As a result, the ground state becomes a complicated superposition of all possible singlet coverings, and as an interesting by-product quasi-long-range order may arise. The important point, however, is that this wavefunction arises from the condition that its local properties are extremal: finding ground states of local translational-invariant two-body Hamiltonians is equivalent to characterizing the convex set of two-body density operators compatible with the fact that they originate from a state with the right symmetry (see Appendix A).

Clearly, we would like to parameterize the manifold of states $\{|\psi_{\text{ex}}\rangle\}$ with extremal local properties. In practice, it is enough to parameterize a manifold of states $\{|\psi_{\text{appr}}\rangle\}$ such that there always exist a large overlap with the exact states $\{|\psi_{\text{ex}}\rangle\}$: for all $|\psi_{\text{ex}}\rangle$, there exists $|\psi_{\text{appr}}\rangle : \||\psi_{\text{ex}}\rangle - |\psi_{\text{appr}}\rangle\| \leq \epsilon$. Let us consider any local Hamiltonian of N spins that exhibits the property that there is a unique ground state $|\psi_{\text{ex}}\rangle$ and that the gap is $\Delta(N)$. Let us furthermore consider the case when $\Delta(N)$ decays no faster than an inverse polynomial in N (this condition is satisfied for all gapped systems and for all known critical translational-invariant systems in one dimension). Then let us assume that there exists a state $|\psi_{\text{appr}}\rangle$ that reproduces well the local properties of all nearest-neighbour reduced density operators: $\|\rho_{\text{appr}} - \rho_{\text{ex}}\| \leq \delta$. Then it follows that the global overlap is bounded by

$$\||\psi_{\text{ex}}\rangle - |\psi_{\text{appr}}\rangle\|^2 \leq \frac{N\delta}{\Delta(N)}.$$

This is remarkable as it shows that it is enough to reproduce the local properties well to guarantee that the global properties are also reproduced accurately: for a constant global accuracy ϵ, it is enough to reproduce the local properties well to an accuracy δ that scales as an inverse polynomial (as opposed to exponential) in the number of spins. This is very relevant in the context of variational simulation methods: if the energy is well reproduced and if the computational effort to obtain a better accuracy in the energy only scales polynomially in the number of spins, then a scalable numerical method can be constructed that reproduces all global properties well (here scalable means essentially a polynomial method).

The central question is thus: is it possible to find an efficient parametrization of a manifold of states whose local properties approximate well all possible local properties? A very interesting new development, inspired a great deal by developments in the field of quantum information and entanglement theory, has shown that this is indeed possible. The main idea is to come up with a class of variational wavefunctions that captures the physics of the low-energy sector of local quantum spin Hamiltonians.

So what other properties do ground states of local quantum Hamiltonians exhibit besides the fact that all global properties follow from their local properties? The key concept to understand their structure is to look at the amount of entanglement present in those states [26]: entanglement is the crucial ingredient that forces quantum systems to behave differently to classical systems, and it is precisely the existence of entanglement that is responsible for such exotic phenomena a quantum phase transitions and topological quantum order [27,28]. It is also the central resource that gives rise to the power of quantum computing [29], and it is known that a lot of entanglement is needed between the different qubits as otherwise the quantum computation can be simulated on a classical computer [7,30]. This is because the amount of entanglement effectively quantifies the relevant number of degrees of freedom that have to be taken into account, and if this is small, then the quantum computation could be efficiently simulated on a classical computer. In the case of ground states of strongly correlated quantum many-body systems, there is also a lot of entanglement (in the case of pure states, the connected correlation functions can be non-zero if an only if there is entanglement), but the key question is obviously to ask how much entanglement is present there: perhaps the

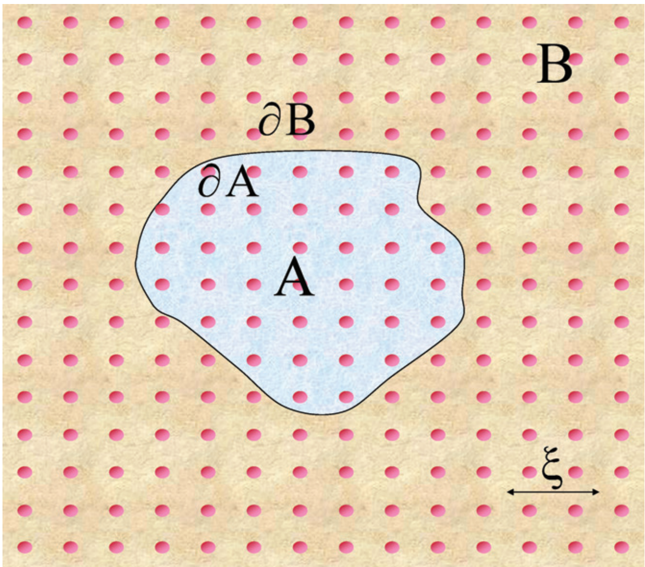

Figure 1. A quantum spin system on a lattice: the lattice is divided into two regions, A and B, with borders ∂A and ∂B, respectively. In the case of ground states of local Hamiltonians, the entanglement entropy between the two regions A and B scales as the area ∂A as opposed to the volume.

amount of entanglement is not too big such that those systems can still be simulated classically.

For example, let us consider a quantum spin system on a n-dimensional infinite lattice, and look at the reduced density operator ρ_L of a block of spins in an $L \times L \times \cdots \times L$ hypercube (Figure 1). The von-Neumann entropy of ρ_L is a coarse-grained measure that quantifies the number of modes in that block that are entangled with the outside [29], and the relevant quantity is to study how this entropy scales with the size of the cube. This question was first studied in the context of black-hole entropy [31–35] and has recently attracted a lot of attention [26,36,37]. Ground states of local Hamiltonians of spins seem to have the property that the entropy is not an extensive property but that the leading term in the entropy only scales as the boundary of the block (hence the name area law):

$$S(\rho_L) \simeq cL^{n-1}. \tag{1}$$

This has a very interesting physical meaning: it shows that most of the entanglement must be concentrated around the boundary, and therefore there is much less entanglement than would be present in a random quantum state (where the entropy would be extensive and scale as L^n). This is very encouraging, as it indicates that the wavefunctions involved exhibit some form of locality, and we might be able to exploit this to come up with efficient local parameterizations of those ground states.

The area law (14) is mildly violated in the case of one-dimensional critical spin systems where the entropy of a block of spins scales as [26,37]

$$S(\rho_L) \simeq \frac{c+\bar{c}}{6}\log L,$$

but even in that case the amount of entanglement is still exponentially smaller then the amount present in a random state. It is at present not clear to what extent such a logarithmic correction will occur in the case of higher dimensional systems: the block-entropy of a critical two-dimensional system of free fermions scales as $L\log L$ [38–41], while critical two-dimensional spin systems were reported where no such logarithmic corrections are present [42], but in any case the amount of entanglement will be much smaller than for a random state. It is interesting to note that this violation of an area law is a pure quantum phenomenon as it occurs solely at zero temperature: in a recent paper [43], it has been shown that the block entropy, as measured by the mutual information (which is the natural measure of correlations for mixed states), obeys an exact area law for all local classical and quantum Hamiltonians. The logarithmic corrections therefore solely arise due to the zero-temperature quantum fluctuations. From the practical point of view, that might indicate that thermal states at low temperature are simpler to simulate than exact ground states.

The existence of an area law for the scaling of entropy is intimately connected to the fact that typical quantum spin systems exhibit a finite correlation length. In fact, Hastings has recently proven that all connected correlation functions between two blocks in a gapped system have to decay exponentially as a function of the distance of the blocks [44]. Let us therefore consider a one-dimensional gapped quantum spin system with correlation length ξ_{corr}. Owing to the finite correlation length, the reduced density operator ρ_{AB} obtained when tracing out a block C of length $l_{AB} \gg \xi_{\text{corr}}$ (see Figure 2) will be equal to

$$\rho_{AB} \simeq \rho_A \otimes \rho_B \tag{2}$$

up to exponentially small corrections. (From a purely mathematical point of view, this is not correct as, surprisingly, there exist states for which all connected correlations functions are negligible while they are very far from being tensor product states [45,46]: examples exist with negligible correlation functions but with the mutual information in the order of the number of qubits in every block. To remedy this, we pointed out in a recent paper that a more sensible way of defining a correlation length is by using the concept

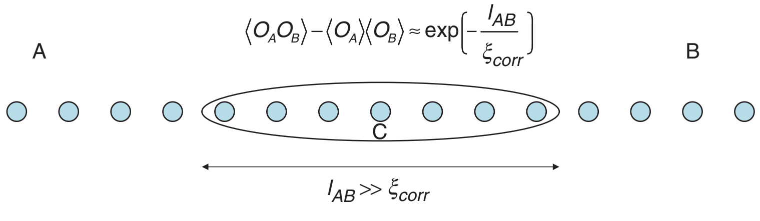
$\langle O_A O_B\rangle - \langle O_A\rangle\langle O_B\rangle \approx \exp\left(-\frac{l_{AB}}{\xi_{corr}}\right)$
A
B
C
$l_{AB} >> \xi_{corr}$

Figure 2. A one-dimensional spin chain with finite correlation length ξ_{corr}; l_{AB} denotes the distance between the block A (left) and B (right). As l_{AB} is much larger than the correlation length ξ_{corr}, the state ρ_{AB} is essentially a product state.

of mutual information, and when there is an exponential decay of mutual information, then the state ρ_{AB} is guaranteed to be close to a tensor product state [43].) The original ground state $|\psi_{ABC}\rangle$ is a purification (a purification of a density operator ρ_X is any pure state $|\psi_{XY}\rangle$ on the bigger Hilbert space X, Y such that its reduced density operator $\mathrm{Tr}_Y|\psi_{XY}\rangle\langle\psi_{XY}| = \rho_X$) of this mixed state, but it is of course also possible to find another purification of the form $|\psi_{AC_l}\rangle\otimes|\psi_{BC_r}\rangle$ (up to exponentially small corrections) with no correlations whatsoever between A and B; here C_l and C_r together span the original block C. It is, however, well known that all possible purifications of a mixed state are equivalent to each other up to local unitaries on the ancillary Hilbert space. This automatically implies that there exists a unitary operation U_C on the block C (see Figure 2) that completely disentangles the left from the right part:

$$I_A \otimes U_C \otimes I_B|\psi_{ABC}\rangle \simeq |\psi_{AC_l}\rangle \otimes |\psi_{BC_r}\rangle.$$

This implies that there exists a tensor $A^i_{\alpha,\beta}$ with indices $1 \leq \alpha,\ \beta,\ i \leq D$ (where D is the dimension of the Hilbert space of C) and states $|\psi^A_\alpha\rangle, |\psi^C_i\rangle, |\psi^B_\beta\rangle$ defined on the Hilbert spaces belonging to A, B, C such that

$$|\psi_{ABC}\rangle \simeq \sum_{\alpha,\beta,i} A^i_{\alpha,\beta}|\psi^A_\alpha\rangle|\psi^C_i\rangle|\psi^B_\beta\rangle.$$

Applying this argument recursively leads to a MPS description of the state (we define those MPSs later) and gives a strong hint that ground states of gapped Hamiltonians are well represented by MPSs. It turns out that this is even true for critical systems [47]; a proof of this is presented in Appendix B.

A remarkable feature of any gapped spin chain is thus that one could imagine dividing the whole chain into segments of size $l \gg \xi_{\mathrm{corr}}$, and then apply a disentangling operation on all blocks in parallel. This would then lead to a product state of many parts, and as such gives a procedure of how such a ground state could be prepared using a quantum circuit with a logical depth that is only dependent on the correlation length (and independent of the number of spins).

In the case of critical systems, we do not expect this disentangling procedure to work as correlations on all length scales appear. However, Vidal showed how this can be remedied by introducing some more advanced disentangling scheme that acts on many different length scales [48]. Basically, the idea is follow up a disentangling step with a coarse-graining step, and do this recursively until there is only one spin left. The procedure for doing this is called the multiscale entanglement renormalization ansatz (MERA), and may lead to alternative methods for simulating quantum spin systems.

It is interesting to note that the situation can again be very different in two dimensions: in that case, gapped quantum spin systems can exhibit topological quantum order, and in [49] it was proven that the depth of any quantum circuit preparing such a topological state has to scale linearly in the size of the system. In other words, it is impossible to device a scheme or pattern by which one could disentangle such states in a parallel way as can be done in one dimension. However, in this two-dimensional case it is also possible to come up with a variational class of states, the so-called PEPSs [12,13], that captures the essential physics for describing those systems. By a recent argument of Hastings [46,50], it can again be proven that basically every ground state of a local Hamiltonian can be well represented by a state within this class.

3. Wilson's NRG method

The exact description of strongly correlated condensed matter systems poses formidable difficulties due to the exponentially large dimension of the associated Hilbert space. However, Wilson was the first to understand that the locality of the interactions between particles or modes enforces ground states to be of a very specific form and could be exploited to simulate them. This insight led to the development of NRG algorithms [1]. The reason for the remarkable accuracy of NRG if applied to quantum impurity problems (e.g. the Kondo and Anderson Hamiltonians [51,52]) can be traced back to the ability to map the related Hamiltonians to momentum space such that they become inhomogeneous one-dimensional quantum lattice Hamiltonians with nearest-neighbour interactions. (The basic reason why a one-dimensional model is obtained even though the original model is concerned with a magnetic impurity in a three-dimensional system is that the Kondo Hamiltonian only affects the *s*-wave part of the wavefunction; the *s*-wave modes are therefore dominant at low energies). NRG is then a recursive method for finding the low-energy spectrum of such Hamiltonians, and yields very accurate results when there is a clear separation of energies, reflected by, for example, an exponential decay of the couplings within the one-dimensional hopping Hamiltonian. The NRG method then recursively diagonalizes the Hamiltonian from large to small energies: at each iteration, a tensor product of the larger energy modes with lower energy modes is made and then projected on a subspace of the lower energy modes of the combined system. Thereafter the Hamiltonian is rescaled. Hence, the basic assumption is that the low-energy modes are affected by their high-energy counterparts, but not vice-versa.

Let us illustrate Wilson's method using the NRG treatment of the single-impurity Anderson model (SIAM). This model can be mapped to a hopping Hamiltonian after a logarithmic discretization of the conduction band [52]:

$$\mathcal{H}=\sum_{n=0}^{N}\xi_n\big(f^\dagger_{n\mu}f_{(n+1)\mu}+\text{h.c.}\big)+\frac{1}{D}\left(\epsilon_d+\frac{U}{2}\right)c^\dagger_{d\mu}c_{d\mu}+\sqrt{\frac{2\Gamma}{\pi D}}\big(f^\dagger_{0\mu}c_{d\mu}+\text{h.c.}\big)+\frac{U}{2D}\big(c^\dagger_{d\mu}c_{d\mu}-1\big)^2, \tag{3}$$

$$\xi_n=\frac{\Gamma^{-n/2}}{2}\frac{(1+\Gamma^{-1})(1-\Gamma^{-n-1})}{\sqrt{(1-\Gamma^{-2n-1})(1-\Gamma^{-2n-3})}}. \tag{4}$$

Here, μ can take the values $\downarrow,\uparrow$, $c_{d\mu}$ denotes the annihilation operator of the impurity and $f_{n\mu}$ of the nth fermion with spin μ, summation over μ has been assumed, and $N\rightarrow\infty$. As the dimension of the associated Hilbert space is 2^{2N}, an exact diagonalization is impossible, and approximations must be made. The hopping terms are decaying exponentially in n, and the basic idea of NRG is to treat the largest n_0 terms first which involves the diagonalization of a 2^{2n_0} matrix. Next we specify a control parameter $D\leq 2^{2n_0}$, retain only the eigenvectors

$$\{|\psi^{n_0}_\alpha\rangle\}_{\alpha=1\ldots D}$$

corresponding to the D lowest eigenvalues, and project the first n_0 terms of the Hamiltonian onto that subspace using the projector

$$P^{[n_0]}=\sum_\alpha|\alpha\rangle\langle\psi^{n_0}_\alpha|$$

yielding the $D \times D$ matrix $\mathcal{H}^{n_0}$. In the first step of the iteration, the extra term involving the coupling between the n_0th and the (n_0+1)th fermions is considered, and an exact diagonalization in the corresponding $4D$-dimensional Hilbert space is performed yielding the eigenvectors. To avoid that the dimensions of the effective Hamiltonian blow up, we project the Hamiltonian onto the new D-dimensional eigenspace corresponding to the lowest eigenvalues

$$P^{[n_0+1]} = \sum_{\alpha} |\alpha\rangle\langle\psi_{\alpha}^{n_0+1}|$$

yielding $\mathcal{H}^{n_0+1}$. Note that $P^{[n_0+1]}$ is a $D \times 4D$ matrix, and for later reference we write its coefficients in tensor form as $P^{i\,[n_0+1]}_{\alpha,\beta}$, $1 \leq \alpha, \beta \leq D, 1 \leq i \leq 4$. Now we iterate this procedure $N - n_0$ times, and NRG is typically said to have converged when $\Gamma\mathcal{H}^{N-1} = \mathcal{H}^{N} + \text{cst}$ up to a unitary transformation. Note that the computational complexity of the NRG procedure scales as ND^3.

Let us now consider the subspace on which we projected the original Hamiltonian more closely. The D states $\{|\psi^{N}_{\alpha_N}\rangle\}$ at the end of the iterations can be written as

$$|\psi^{N}_{\alpha_N}\rangle = \sum_{\alpha_{n_0}\ldots\alpha_{n_{N-1}}} \sum_{i_{n_0} i_{n_0+1}\ldots i_N} P^{i_{n_0}[n_0]}_{\alpha_{n_0}} P^{i_{n_0+1}[n_0+1]}_{\alpha_{n_0}\alpha_{n_0+1}} \ldots P^{i_N[N]}_{\alpha_{N-1}\alpha_N} |i_{n_0}\rangle|i_{n_0+1}\rangle \ldots |i_N\rangle. \tag{5}$$

These states are of exactly the same form as those mentioned in the introduction, and due to the feature that they are defined as a product of matrices, these are called MPSs [53–57] and were originally introduced in the mathematical physics community under the name of finitely correlated states [58,59] (a precursor of this appeared in the context of quantum Markov chains [60]). The energies calculated using NRG are thus energies of an effective Hamiltonian which is the original one projected onto a subspace of MPS. In practice, the NRG method is highly successful for problems where the different terms in the Hamiltonian act on a different scale of energy, and in that case results up to essentially machine precision can be obtained; this can only be true if the class of the MPSs indeed captures all of the physics needed to describe the low-energy physics of these Hamiltonians.

However, by looking at NRG by means of MPSs, it is already clear that it can in principle be formulated as a variational method within the set of MPS. It turns out that this is exactly what White did by introducing the DMRG [2,3], but the way of looking at both methods from that point of view of MPS was only discovered much later [61].

4. MPSs and ground states of spin chains

4.1. *Construction and calculus of MPSs*

4.1.1. *The AKLT model*

The notion of MPSs already appeared naturally in Section 2 describing spin chains with finite correlation length and in the context of NRG. They were first studied in the work of Affleck, Kennedy, Lieb and Tasaki (AKLT) [62], where it was proven that the exact ground state of the spin-1 spin chain with the Hamiltonian

$$\mathcal{H}_{\text{AKLT}} = \sum_{\langle i,j\rangle} \underbrace{\left(\vec{S}_i.\vec{S}_j + \frac{1}{3}(\vec{S}_i.\vec{S}_j)^2 + \frac{2}{3}\right)}_{=P_{ij}}$$

can be parameterized exactly as a MPS. To see this, they observed that the terms P_{ij} are projectors $(P_{ij})^2 = P_{ij}$ onto the five-dimensional spin-2 subspace of two spin-1s, and proceeded by constructing the unique ground state $|\psi_{\text{AKLT}}\rangle$ which is annihilated by all projectors P_{ij} acting on nearest neighbours. This state $|\psi_{\text{AKLT}}\rangle$ can be constructed as follows.

- Imagine that the three-dimensional Hilbert space of a spin-1 particle is effectively the low-energy subspace of the Hilbert space spanned by two spin-1/2 particles, that is, the three-dimensional Hilbert space is the symmetric subspace of two spin-1/2 particles.
- To ensure that the global state defined on the spin chain has spin zero, let us imagine that each one of the spin-1/2s is in a singlet state with a spin-1/2 of its neighbours (see Figure 3).
- The AKLT state can now be represented by locally projecting the pair of spin-1/2s in the symmetric subspace onto the spin-1 basis $\{|1\rangle, |0\rangle, |-1\rangle\}$:

$$\begin{aligned} P &= |-1\rangle\left(\frac{\langle 00| - \langle 11|}{\sqrt{2}}\right) + |0\rangle\left(\frac{\langle 01| + \langle 10|}{\sqrt{2}}\right) + |1\rangle\left(\frac{\langle 00| + \langle 11|}{\sqrt{2}}\right) \\ &\equiv \sum_{\alpha=x,y,z} |\alpha\rangle\left(\frac{\langle 00| + \langle 11|}{\sqrt{2}}\right)\tau_\alpha \otimes \tau_y, \end{aligned}$$

where $\tau_x = \sigma_x$, $\tau_y = i\sigma_y$, $\tau_z = \sigma_z$ with $\{\sigma_\alpha\}$ the Pauli matrices and where we identified $|-1\rangle = |x\rangle$, $|0\rangle = |z\rangle$, $|1\rangle = |y\rangle$.

Historically, the AKLT state was very important as it shed new light on the conjecture due to Haldane [63,64] that integer spin Heisenberg chains give rise to a gap in the spectrum. That is a feature shared by all generic MPSs: they are always ground states of local gapped quantum Hamiltonians.

Let us first try to rewrite $|\psi_{\text{AKLT}}\rangle$ in the MPS representation. Let us assume that we have an AKLT system of N spins with periodic boundary conditions (PBCs); projecting the wavefunction in the computational basis leads to the following identity:

$$\langle \alpha_1, \alpha_2, \ldots, \alpha_N | \psi_{\text{AKLT}}\rangle = \text{Tr}(\tau_{\alpha_1}.\tau_y.\tau_{\alpha_2}.\tau_y \ldots \tau_{\alpha_N}.\tau_y).$$

The different weights can therefore be calculated as a trace of a product of matrices. The complete AKLT state can therefore be represented as

$$|\psi_{\text{AKLT}}\rangle = \sum_{\alpha_1,\alpha_2,\ldots,\alpha_N} \text{Tr}(\tau_{\alpha_1}.\tau_y.\tau_{\alpha_2}.\tau_y \ldots \tau_{\alpha_N}.\tau_y)|\alpha_1\rangle|\alpha_2\rangle \ldots |\alpha_N\rangle$$

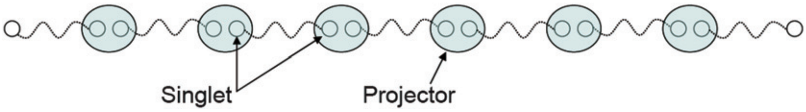

Figure 3. Building up the AKLT state by partial projections on bipartite singlets.

which is of almost exactly the same form as the MPSs introduced in 5. The only difference between them is the occurrence of the matrices τ_y between the different products. This is, however, only a consequence of the fact that we connected the different nodes with singlets, and we could as well have used maximally entangled states of the form

$$|I\rangle = \sum_{i=1}^{2} |ii\rangle$$

and absorbing τ_y into the projector; this way we recover the standard notation for MPSs.

So what did we learn from this AKLT example? Basically, that there is a way of parameterizing the exact ground state of a particular strongly correlated quantum spin chain using a construction involving virtual bipartite entanglement and projections. From the point of view of quantum information theory, this parametrization is very appealing, as it gives an explicit way of constructing a highly entangled multipartite quantum state out of bipartite building blocks. More importantly, it is immediately clear how this picture can be generalized [10,65]: instead of taking spin $D=2$ bonds corresponding to spin-1/2s, we can take much larger D for the virtual spins, and furthermore it is obvious that the projectors can be replaced with any linear map. What is really exciting in doing so is the fact that the states arising from this are translational invariant by construction [58]; this is highly relevant as there does not seem to be another simple way of parameterizing translational-invariant states. Recalling the discussion in Section 1, it is now obvious that this class of translational-invariant MPSs, originally introduced in the literature under the name of finitely correlated states [58], would form a very good ansatz for parameterizing ground states. Indeed, ground states of local Hamiltonians are characterized by extremal local properties that are still compatible with the global translational symmetry, and as the MPSs are built up by projecting the underlying maximally entangled states with extremal local correlations, it is very plausible that the MPSs are perfectly suited for this. In addition to the AKLT model, many variations of that Hamiltonian have been studied with exact MPSs as ground states [66–68]; this is particularly interesting because analytical solutions of spin chains are very rare.

Note that the number of parameters that we have to our disposal in these translational-invariant MPSs scales as dD^2 with d the physical dimension of the spins (indeed, we have to parameterize d matrices A_α of dimension $D \times D$), and hence the natural question now is how does the convex set of all local reduced density operators so obtained compare with the exact convex set of all possible translational-invariant systems. This problem is considered in appendices A and B, and it is found that the number of parameters needed has to scale as a constant or at most polynomially in the number of spins.

4.1.2. *MPSs*

As already explained in the previous Section, the obvious generalization of the AKLT states is obtained by making the dimensions of the virtual spins D larger and considering general linear maps A_α instead of projectors, but we can make a further generalization by making D (the dimension of the virtual subsystems) and those maps site-dependent and writing them as D_i and A^i_α. The most general form of a MPS on N spins of dimension d is then given by

$$|\psi\rangle = \sum_{\alpha_1,\alpha_2,\dots,\alpha_N}^{d} \mathrm{Tr}\big(A^1_{\alpha_1} A^2_{\alpha_2} \dots A^N_{\alpha_N}\big)|\alpha_1\rangle|\alpha_2\rangle \dots |\alpha_N\rangle. \tag{6}$$

The matrices A^i have dimension $D_i \times D_{i+1}$ (here we take the convention that $D_{N+1} = D_1$), and a system with open boundary conditions (OBCs) is obtained by choosing $D_1 = 1$ (i.e. no 'singlet' between the endpoints). A pictorial representation is given in Figure 4. Before continuing, let us remark that every state of N spins has an exact representation as a MPS if we let D grow exponentially in the number of spins; this can easily be shown by making use of the tool of quantum teleportation as shown in [10]. However, the whole point of MPSs is that ground states can typically be represented by MPSs where the dimension D is small and scales at most polynomially in the number of spins; this is the basic reason why renormalization group methods are exponentially more efficient than exact diagonalization.

4.1.3. *Calculus of MPSs*

Next, let us explain the calculus of those MPSs. A crucial aspect of those MPSs is indeed the fact that the expectation value of a large class of observables can be calculated efficiently. More specifically, this holds for all observables that are a tensor product of local observables, such as $\sigma^i_x \otimes \sigma^j_x$. The easiest way to understand this is by considering a so-called tensor network, in which connected bonds correspond to contracted indices and open bonds to uncontracted indices (see Figure 5). More specifically, the tensor network in Figure 5(a) corresponds to the quantity

$$A^1_{\alpha_1}.A^2_{\alpha_2} \dots A^N_{\alpha_N}$$

as the 'virtual' indices are all contracted (corresponding to taking products of matrices) and the physical ones are left open. To calculate expectation values of observables, it is clear that we also have to contract those physical indices, after sandwiching in the observables, and hence expectation values can be represented by a completely contracted network. The relevant thing of course is now to quantify the computational complexity of actually doing those contractions. Naively, one expects that the calculation of the expectation value of a quantum state consisting of, say, N spin-1/2s would involve a number of operations of the order of the size of the Hilbert space and, hence, exponential in N. This strategy would correspond to contracting the tensor network

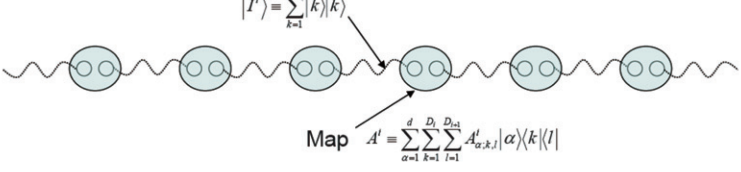
$|I^i\rangle \equiv \sum_{k=1}^{D_i}|k\rangle|k\rangle$
Map $A^i \equiv \sum_{\alpha=1}^{d}\sum_{k=1}^{D_i}\sum_{l=1}^{D_{i+1}} A^i_{\alpha;k,l}|\alpha\rangle\langle k|\langle l|$

Figure 4. A general MPS.

from top to bottom. However, we can be smarter and contract the tensor network from left to right such that we never have to store more than $D_i \times D_i$ variables during the computation (i.e. D_i from the upper row and D_i from then lower row). Furthermore, the contraction of tensors from left to right can be done in a smart way. Let us, for example, assume that all virtual bonds have dimension D and physical bonds dimension d; the leftmost contraction corresponds to contracting the physical bond, leaving a tensor with two indices i, j of dimension D and D. Next we contract one of those with a tensor A_i, and thereafter with $\bar{A}_i$. The total number of multiplications that had to be performed in that procedure is given by d^2D^3, and this computational cost of contracting the whole network is therefore Nd^2D^3 as we have to repeat the previous step N times. If D is not too big (i.e. $D < 1000$), this can still be performed efficiently on present-day computers. (Calculations up to $D = 10{,}000$ have been performed in cases where very high precision was needed, but such cases are only tractable when explicitly making use of quantum numbers [4].) Note that it is also obvious how to calculate the expectation value of any observable that is defined as a tensor product over local observables; this is useful when calculating, for example, the decay of correlations.

In the case of PBCs, the network to contract is very similar (Figure 5(c)) but one has to keep track of more variables (namely those going to the left and to the right), and the total computational cost amounts to Nd^2D^5; the extra factor of D^2 is responsible for the fact that simulations with PBCs are considerably slower than with OBCs [2,3,10], but one has to note that the scaling is still only polynomial and that a D of the order of 100 is still achievable in principle, which is more than enough to obtain a good accuracy in actual calculations.

The rest this subsection is pretty technical and can be skipped for a reader who is not really interested in technical details that are relevant for actually implementing

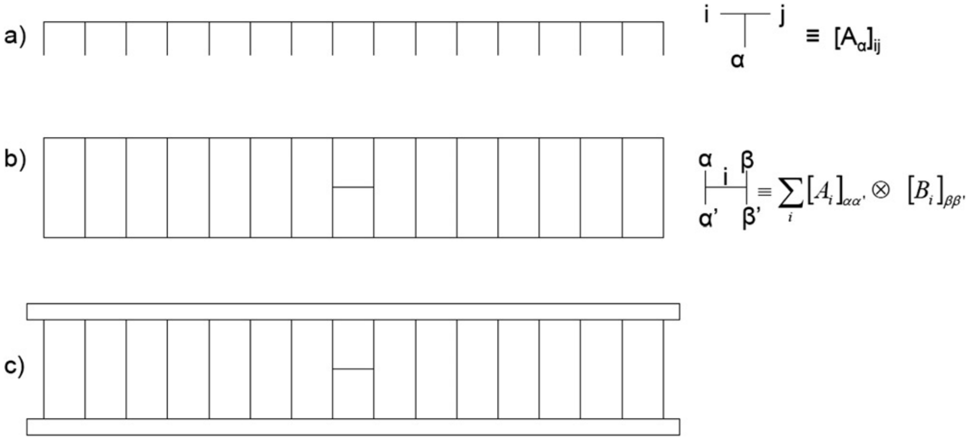
a)
i
j
α
$\equiv [A_\alpha]_{ij}$
b)
α
β
i
α'
β'
$\equiv \sum_i [A_i]_{\alpha\alpha'} \otimes [B_i]_{\beta\beta'}$
c)

Figure 5. (a) An MPS with OBCs as a tensor network. The open bonds correspond to the uncontracted physical indices and the closed bonds to contracted indices arising from taking the products. (b) Calculating the expectation value of an nearest-neighbour operator over a MPS with OBCs by contracting the tensor network completely. (c) An extra bond has to be contracted in the case of PBCs.

NRG methods. Let us first show how to efficiently calculate expectation values of generic Hamiltonians with only terms that act on nearest-neighbour spins:

$$\mathcal{H} = \sum_{k=1}^{N} O^k + \sum_{k=1}^{N-1}\sum_{\gamma} B_\gamma^k \otimes B_\gamma^{k+1}.$$

Instead of calculating the expectation value of each term in the sum separately, one can be a bit smarter and store the information of the result of the contraction from left to right and arrange everything such that the total computational cost is basically still Nd^2D^3, which in practice means a huge gain in computation time. See Appendix C for an actual implementation of this in Matlab.

Another technical tool that is of great practical value is the fact that a MPS does not have a unique representation in terms of tensors A^i but that so-called gauge transformation on the virtual level leaves the physical state invariant. This follows from the fact that a multiplication of two matrices is left invariant when a non-trivial resolution of the identity is inserted between them:

$$\begin{aligned} AB &= AXX^{-1}B = A'B', \\ A' &= AX, \\ B' &= X^{-1}B. \end{aligned}$$

It happens that those gauge degrees of freedom can be exploited to make the forthcoming numerical optimization methods better conditioned. In the particular case of a MPS with OBCs, we would like them to have some orthonormality properties. For example, consider again the NRG of Wilson that was sketched in Section 2. There, one obtained collections of MPSs $\{|\psi_{\alpha_k}^k\rangle\}$ at each step of the recursion such that they were all orthonormal to each other. It is easy to see that a gauge transformation can always be found such that this is fulfilled in the case of MPS with OBCs. Consider therefore a MPS with tensors $A^1, A^2, \ldots, A^N$, and let us look at the two collections of D_k MPS defined on the left and right parts of the qubits, respectively, by opening the kth contracted bonds

$$|\psi_n^k\rangle = \sum_{\alpha_1\alpha_2\ldots\alpha_k} A_{\alpha_1}A_{\alpha_2}\ldots A_{\alpha_k}.e_n|\alpha_1\rangle|\alpha_2\rangle\ldots|\alpha_k\rangle \tag{7}$$

$$|\chi_n^k\rangle = \sum_{\alpha_{k+1}e_n^T\alpha_{k+2}\ldots\alpha_N} A_{\alpha_{k+1}}A_{\alpha_{k+2}}\ldots A_{\alpha_N}|\alpha_{k+1}\rangle|\alpha_{k+2}\rangle\ldots|\alpha_N\rangle \tag{8}$$

where e_n denote the different unit vectors in a D_k-dimensional vector space. Let us furthermore consider the matrices

$$A_{nn'} = \langle\psi_{n'}^k|\psi_n^k\rangle \tag{9}$$

$$B_{nn'} = \langle\chi_{n'}^k|\chi_n^k\rangle \tag{10}$$

and take their respective square roots $A = XX^\dagger$ and $B = YY^\dagger$. It is clear that if we were now to perform the gauge transformation $A_i^k \to A_i^k X^{-1}$ and repeat the procedure described above, we would see that the set $\{|\psi_n^{'k}\rangle\}$ would form an orthonormal set as A' would be equal to the identity. Similarly, we could make the MPS on the right-hand side orthonormal.

It is interesting to note that the Schmidt coefficients obtained by considering the bipartite cut over the kth bond are precisely given by the singular values of the matrix XY^T; indeed, the matrices X^{-1} and Y^{-T} were used to make the left- and right-hand side orthonormal, and the original MPS can of course be written as

$$|\psi\rangle = \sum_{n=1}^{D_k} |\psi_n^k\rangle|\chi_n^k\rangle.$$

Writing out the singular value decomposition (SVD) of $XY^T = U\Sigma V^T$ explicitly, a natural gauge transformation to represent the MPS with OBCs would then be obtained by implementing the gauge-transformation

$$A_i^k \to A_i^k X^{-1} U \quad A_i^{k+1} \to V^T Y^{-T} A_i^{k+1}$$

for all $k = 1\ldots N$ and by writing out the MPS in such a form that the Schmidt coefficients are appearing explicitly:

$$|\psi\rangle = \sum_{\alpha_1\alpha_2\ldots\alpha_N} A^1_{\alpha_1}\Sigma_1 A^2_{\alpha_2}\Sigma_2 \ldots A^N_{\alpha_N}|\alpha_1\rangle\ldots|\alpha_N\rangle.$$

Here, the Σ_i are the diagonal matrices containing the Schmidt coefficients. This parametrization of MPS with OBCs coincides exactly with the definition of the family of states introduced by Vidal in the context of the simulation of real-time evolution of quantum many-body systems [69], and we have just shown that every MPS with OBCs can be written like that.

It would be very appealing to have a similar parametrization in the case of MPS with PBCs. However, it is clear that no notion of singular values or orthonormality can exist in that case as the left-hand side is always connected to the right-hand side. However, in a quantum spin system defined on a ring with a correlation length much shorter than the length of the ring, the boundary effects are not expected to be too pronounced, and an approximate orthonormalization can still be carried out [10]. This is discussed in a later Section. Related to this issue, let us suppose that we have a MPS with PBCs for which we know that there exists a MPS description with all tensors A^i equal to each other (the state is hence obviously translational invariant). However, let us assume that we have a MPS description of that state with all tensors A^i different from each other (i.e. the symmetry has been spoiled by site-dependent gauge transformations), for example, as the result of a variational optimization. Is there a way to recover the symmetric description? For this, we have to find the gauge transformation $\tilde{A}^i_\alpha = X_i A^i_\alpha X^{-1}_{i+1}$, $i = 1\ldots N$, such that

$$\tilde{A}^1_\alpha = \tilde{A}^2_\alpha = \cdots = \tilde{A}^N_\alpha \equiv \tilde{A}_\alpha.$$

Without loss of generality, we can assume that X_1 is equal to the identity (indeed, we still have the freedom of a global gauge transformation). However, then we can find X_2 and X_N as the equations $A^1_\alpha X_2^{-1} = X_N A^N_\alpha$, $\alpha = 1\ldots d$ has at least as many equations as unknowns (in practice, this should be done using a least squares algorithm). This can then be iterated until all gauge transformations X_i have been determined. Alternatively, the new matrices $\tilde{A}_\alpha$ can be found by considering the gauge-invariant normal form of the MPS: in terms of A^i_α, the norm is expressed as $\mathrm{Tr}(E_1 \cdots E_N)$, with $E_i = \sum_{\alpha=1}^d A^i_\alpha \otimes \bar{A}^i_\alpha$. This expression is equal to the norm expressed in terms of $\tilde{A}_\alpha$, namely $\tilde{E}^N$, with $\tilde{E} = \sum_{\alpha=1}^d \tilde{A}_\alpha \otimes \bar{\tilde{A}}_\alpha$.

Therefore, the $\tilde{A}_\alpha$ are obtained by taking the Nth root of $E_1 \cdots E_N$ and performing a Schmidt decomposition of the result.

To summarize, we have shown that MPSs are very appealing from the computational point of view as all correlations functions and the expectation value of local Hamiltonians can be calculated efficiently. Moreover, we have shown how gauge transformations can be used to bring a MPS with OBCs into a normal form.

A relevant observation is that we can also efficiently calculate the overlap between two different MPS. This makes the MPS approach a very nice tool to detect quantum phase transitions, because small changes in the Hamiltonian led to big changes in the ground state around those transition points, and those changes can be detected by calculating the overlap between the different MPS approximations of the respective ground states [70].

A final remark is that it is also simple to calculate expectation values of Hamiltonians with long-range terms. This allows us to easily generalize the methods introduced in later Sections to situations of long-range Hamiltonians.

4.1.4. *Generalization of MPSs*

The valence bond state representation can readily be generalized to trees or higher dimensions. The generalization to higher dimensions is discussed in Section 6, and there it will become clear that the calculation of expectation values is much more involved than in the one-dimensional case and can only be done approximately. In contrast to this, it is easy to see that the generalization of MPSs to tree networks without loops leads to a family of states from whom all expectation values can be calculated efficiently without having to make approximations [71].

Let us illustrate this for the particular case of a MPS on a Cayley tree with coordination number 3 (see Figure 6). To calculate expectation values of such a state $\langle\psi|O|\psi\rangle$, we start contracting the indices from the boundary inwards. As there are no loops, it is easy to see that the number of variables to keep track of does not explode exponentially but remains bounded just like in the case of MPSs. For simplicity, let us assume that all projectors P are equal to each other. At any point of the contraction when going from the outside to the inside, we have two incoming and one outgoing vertices; as the earlier contractions of the incoming vertices did never entangle with each other, their joint state is in a product $\rho_k \otimes \rho_k$. Analogously to the MPS case, the role of the projector P is to apply a completely positive map on this product state: if we define the set of Kraus operators $A^i_{\alpha;\beta\gamma} = P^i_{\alpha\beta\gamma}$, then

$$\rho_{k+1} = \sum_{i=1}^{d} A_i \rho_k \otimes \rho_k A_i^\dagger$$

and yields the input for the next level of contractions. As the map from $\rho_k \rightarrow \rho_{k+1}$ is clearly a non-linear completely positive map, its structure is much richer than in the case of MPSs and associated completely positive linear maps [72]: it happens that such non-linear maps have multiple fixed points with different basins of attractions, and hence the boundary terms can affect the expectation values in the bulk of the system (note that this seems to be a consequence of the fact that we work with a Cayley tree, where the number of vertices increases exponentially as a function of the distance to the centre of the tree).

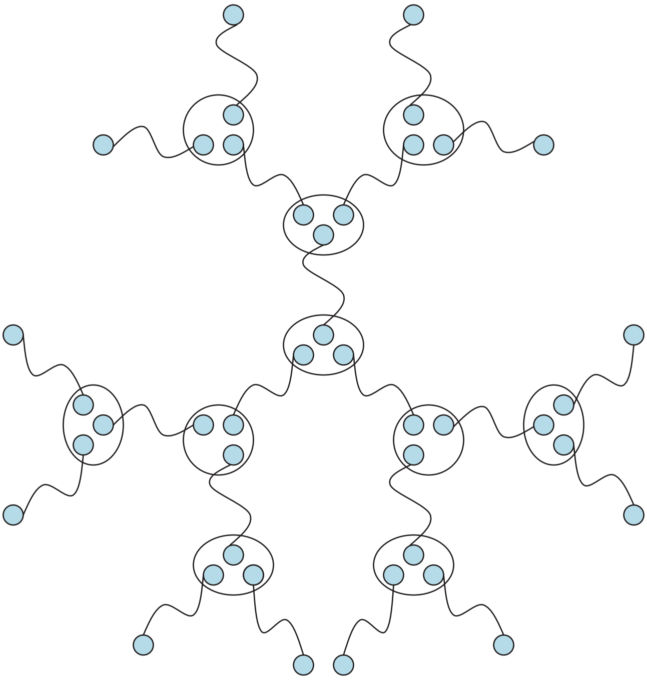

Figure 6. A MPS in the form of a Cayley tree with coordination number 3: just as in the case of spin chains, the big open circles represent projections of three virtual spins to one physical spin.

Anyway, the important point to make is that expectation values of any product observable can be readily calculated efficiently by multiplying matrices with each other: in the case of a tree with coordination number c, the computational cost is D^{c+1} with D the dimension of the bonds (note that this yields D^3 for $c=2$ (i.e. MPS) and D^4 for the Cayley tree with $c=3$). Note also that it is straightforward to obtain a canonical normal form for the tree in the same way as for the MPS by making use of the SVD [73]. Most of the simulation techniques described in the following Sections can therefore also immediately be generalized to this setting of trees.

It can be seen that it is also possible to contract the tensor network exactly and efficiently if there are loops, but not too many of them. A clear demonstration of this is of course the one-dimensional case of MPSs with PBCs: numbering the d-dimensional spins from 1 to $2N$ in the case of a ring with an even number of sites, one can define N new spins by merging spins i and $2N-i+1$ into composite pairs. The corresponding MPS with OBCs has spin dimension d^2 and bond dimension D^2, which can still be contracted efficiently. Similarly, we will be able to contract a generic tree with loops efficiently if it can be mapped onto a different tree without loops by merging collections of spins together; this procedure will still lead to a polynomial computational complexity if and only if the local dimension of the new spins is bounded by a polynomial in the size of the system. This condition is equivalent to the requirement that the maximal amount of spins that has been merged into a new superspin scales at most logarithmic in the size of the system (which guarantees a complexity $poly(D,N)$). It turns out that this notion of transforming an arbitrary graph into a tree by merging vertices together is a well-studied problem in graph theory, and that the tree width of a graph follows from the optimal way of doing this such as minimizing the maximal number of vertices that has to be merged [74] over all possible mappings from a graph to a tree: the tree width is then the maximal number of vertices that are merged in this optimal solution. Although calculating the tree width of a general graph seems to be NP-hard, efficient (i.e. P) approximation algorithms exist that approximate the tree width to within a constant. The connection between tree width and MPSs was pointed out in [75].

One can also formulate a variant of the tree network case, in which the nodes of the tree do not carry any physical spins but only the endpoints of the branches do; in other words, we start from a collection of virtual singlets, order them in the form of a tree, and put a projector on all of the vertices that maps the D^c-dimensional Hilbert space to a one-dimensional Hilbert space. Contraction of such a tensor network can again be done in an efficient way by starting the contraction at the leaves and working oneself inwards. An interesting observation is the fact that every MPS with a given D can be expressed as such a Cayley tree with coordination number $c=3$ and bond dimension D^2 (this is true both for the case of OBCs and PBCs). Note that the opposite situation is different: if we want to represent a generic state on the Cayley tree of bond dimension D as a MPS, the MPS will have a site-dependent bond dimension bounded above by $D^{\log_2(N)} = N^{\log_2(D)}$ (which is still polynomial). This can easily be seen from the construction depicted in Figure 7. Such a description was used to develop a formalism in which one can perform successive renormalization group transformations on quantum states as opposed on the usual level of Hamiltonians [76].

As a further extension, Vidal observed that more general types of tensor networks can be contracted efficiently: if all tensors appearing in the tree (with only physical spins at the bottom) are isometries, then one can additionally put a collection of so-called disentangling unitaries between the different branches. The reason why expectation values of local observables (or more generally observables that act non-trivially only on a constant number of spins) can still be calculated exactly is that in the expression $\langle\psi|O|\psi\rangle$ most of those unitaries that disappear cancel each other out and hence play no role: one only has to keep track of the unitaries within the *lightcone* of the local observable, and it can be shown that the cost of this is still polynomial (albeit with a very large power, especially if one considers the two-dimensional variants of this construction). This approach is called the MERA [48,77], and has a particularly nice interpretation as successive rescaling transformations in the sense of the renormalization group. Due to the presence of the tree network, the Schmidt number in the one-dimensional setting can grow logarithmically with the system size, and the method is therefore very promising for describing critical systems. As opposed to the class of MPSs, however, there does not seem to be a straightforward well-conditioned way of performing a variational optimization over the class of MERA to find the MERA with minimal energy. Several promising approaches have been described in [48,77,78]; the simplest approach is to parameterize the unitaries and isometries as exponentials of anti-Hermitian operators, and implement the steepest descent optimization algorithm on those parameters. (Such optimization techniques over the manifold of unitaries have been studied in great detail in the context of control theory and are called flow equations; they were also used in the context of entanglement theory [79] and first implemented on the level of MERA in [80].) Further extensions are possible, and there is currently an effort in trying to identify the broadest class of quantum states for which all local properties can be calculated efficiently

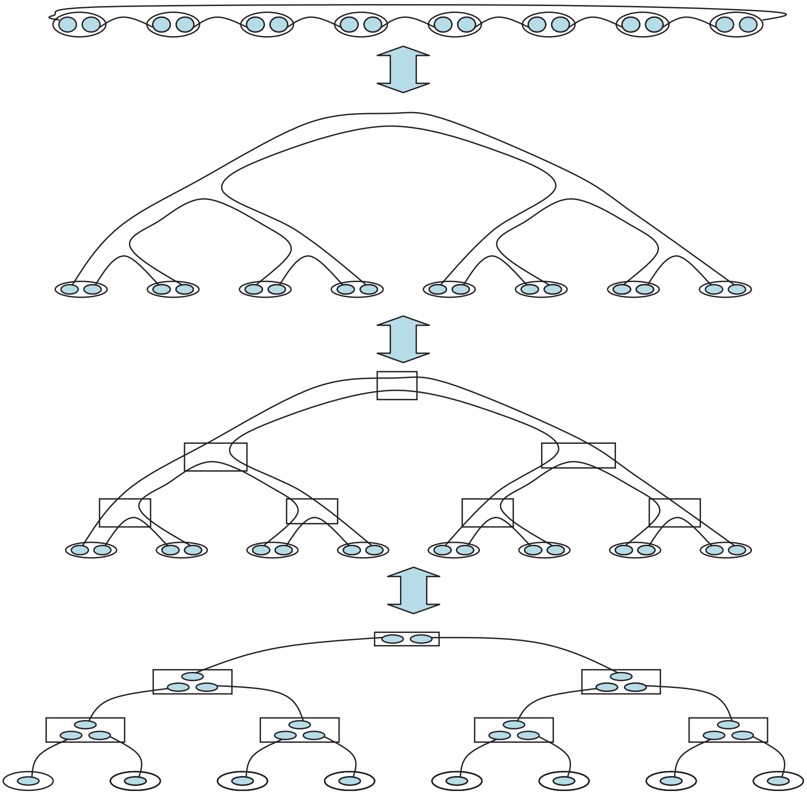

Figure 7. Every MPS can be represented in the form of a tree network in which the vertices are projectors onto a one-dimensional subspace (depicted as square boxes as opposed to ellipses for the case of projectors to physical spins); in the last step, it is understood that the original matrices $A^i_{\alpha\beta}$ are absorbed into the square blocks on the lowest level which are tensors with six D-dimensional indices.

(i.e. polynomial complexity). As an example, one can make use of the concept of MPOs to show that an efficient contraction is still possible when one acts with a quantum circuit of arbitrary depth but consisting only of commuting gates (e.g. $\exp(i\alpha\sigma_z^i \otimes \sigma_z^j)$) between any two spins (even those that are very far apart) of an arbitrary MPS; such states can violate the area law in an extreme sense because they might lead to volume laws, and such states might be relevant in simulations of systems very far from equilibrium.

4.2. *Reformulating NRG methods as variational methods in the class of MPSs*

Let us now look back at the NRG, and reformulate it as a variational method within the class of MPSs. To start with, let us aim to find the *best possible* description of the ground state of, for example, the Heisenberg model within the space of all MPSs of the form (19), using the matrix elements of the matrices $\{A^n\}$ as *variational parameters* to minimize the energy (note that we no longer impose the condition that the A^n are projectors).

Using a Lagrange multiplier to ensure normalization, this leads to the following optimization problem:

$$\min_{|\psi^N\rangle\in\{\mathrm{MPS}_D\}}[\langle\psi^N|\mathcal{H}^N|\psi^N\rangle-\lambda\langle\psi^N|\psi^N\rangle].$$

This cost function is multiquadratic in the dN matrices $\{A_k\}$ with a multiquadratic constraint; indeed, every matrix appears only once in the bra and once in the ket, and so this problem is basically equivalent to a cost function of the form

$$\min_{x^1,x^2,\dots}\sum_{k_1,k_2,\dots l_1,l_2,\dots} x^1_{k_1}\bar{x}^1_{l_1}x^2_{k_2}\bar{x}^2_{l_2}\cdots Q_{k_1,k_2,\dots,l_1,l_2,\dots},$$

where the x^k are vectors with $d\times D_k\times D_{k+1}$ elements and the big tensor Q contains the information of the Hamiltonian. The Lagrange constraint is of a similar form, and it that case $Q_{k_1,k_2,\dots}=\delta_{k_1l2}\delta_{k_2l_2}\dots$ There is a standard way of solving such an optimization problem which is called alternating least squares (ALS). (Multiquadratic optimization problems can in principle be NP-hard to solve [81], and so there is no guarantee that the ALS method will converge to the global optimum. However, in practice this does not seem to occur. See also [82].) This ALS is an iterative method that works as follows: after making an initial guess of the vectors x^k, we keep $x^2,\dots,x^N$ fixed and optimize over x^1. That subproblem is of the form

$$\min_{x^1} x^{1\dagger}H_{\mathrm{eff}}x^1-\lambda x^{1\dagger}N_{\mathrm{eff}}x^1$$

(note that H_{eff} and N_{eff} depend on all of the other x^k) and is exactly solvable as it is a quadratic problem with quadratic constraints. The solution is that we have to choose x equal to the smallest eigenvalue of the generalized eigenvalue problem $H_{\mathrm{eff}}x=\lambda N_{\mathrm{eff}}x$. A crucial point in all of this is that there is an efficient way of calculating H_{eff} and N_{eff} given an MPS and the Hamiltonian: these are obtained by contracting tensor networks of the type discussed in 3.1.3, and an important point in the actual implementation of this is to store the relevant matrices for later use. In the next step, we fix x^1,x^3,x^4,... and repeat the same procedure, until we reach N and repeat the procedure again by going from N to 1 and so on until convergence. Note that at each step of this procedure, the energy is going down and we are hence guaranteed to converge to some value that is hopefully close to the minimal energy of the Hamiltonian [10]. We call this procedure the variational MPS (VMPS) method.

And what about excitations? Within the framework discussed up to this point, this can easily be done variationally [83]. Suppose we found the MPS $|\psi_0\rangle$ with lowest energy, the variational problem is then to find the MPS $|\psi_1\rangle$ with the lowest energy that is orthogonal to the $|\psi_0\rangle$. This amounts to taking another Lagrange constraint in the optimization problem:

$$\min_{|\psi_1\rangle\in\{\mathrm{MPS}_D\}}[\langle\psi_1|\mathcal{H}^N|\psi_1\rangle-\lambda\langle\psi_1|\psi_1\rangle-\mu\langle\psi_1|\psi_0\rangle]$$

and this can still be solved iteratively in a very similar way. Basically, at every step we have a subproblem of the form

$$\min_x x^\dagger H_{\mathrm{eff}}x-\lambda x^\dagger N_{\mathrm{eff}}x-\mu y^\dagger x$$

whose solution is given by the solution to the generalized eigenvalue problem $PH_{\mathrm{eff}}Px=\lambda PN_{\mathrm{eff}}Px$ with P the projector on the subspace orthogonal to the vector y.

Doing this iteratively, this will again converge to a state $|\psi_1\rangle$ that is the best possible approximation of the first excited state using a MPS. Of course, a similar procedure can now be used to construct more excited states.

Let us now look back at the NRG method of Wilson. In essence, the goal is to find the low-energy spectrum of the Hamiltonian. NRG can be understood as a different way of performing the optimization discussed above, but the NRG method is suboptimal and is only applicable in the particular case where there is a clear separation of energies (such as for a Kondo impurity or SIAM). Basically, the steps done in the case of NRG are equivalent to the steps one would do using the variational method discussed above during the first sweep from left to right. However, the NRG method is stopped after that first sweep, which effectively means that one never takes into account the influence of the low-energy modes on the larger energy modes. (This feedback may be small in practice, but it is not strictly zero, and its importance increases as the logarithmic discretization is refined by taking $\Lambda \to 1$.) This is clearly suboptimal, and a better result can be obtained by sweeping back and forth a few more times until complete convergence is obtained [61]. The computational complexity of this is no greater than that of the original NRG method. The method so obtained is then basically equivalent to the DMRG introduced by White [2,3], and it is indeed well known that DMRG has a much wider range of applicability than NRG. Note, however, that NRG and DMRG were always considered to be rather different, and it is only by reformulating everything in terms of MPSs that the many similarities become apparent.

Let us now come back to the variational method explained above for doing NRG. The effective Hamiltonian at chain length n, the central object in NRG, is given by $\tilde{\mathcal{H}}^n_{\alpha\beta} = \langle\tilde{\psi}^n_\alpha|\mathcal{H}^n|\tilde{\psi}^n_\beta\rangle$, and can hence also easily be recovered. Let us now illustrate the above by applying them to the SIAM given in equation (9). The following example is taken out of the paper [61]. Since the Hamiltonian couples $\uparrow$ and $\downarrow$ band electrons only via the impurity, it is possible (see also [84]) to "unfold" the semi-infinite Wilson chain into an infinite one, with $\uparrow$ band states to the left of the impurity and $\downarrow$ states to the right, and hopping amplitudes decreasing in both directions as $\Gamma^{-|n|/2}$. Since the left and right end regions of the chain, which describe the model's low-energy properties, are far apart and hence interact only weakly with each other, the effective Hamiltonian for these low energies will be of the form $\mathcal{H}^{\text{eff}}_{\uparrow} \otimes 1_{\downarrow} + 1_{\uparrow} \otimes \mathcal{H}^{\text{eff}}_{\downarrow}$. This is illustrated by the black and blue lines in Figure 8(a). Since $\mathcal{H}^{\text{eff}}_{\uparrow}$ and $\mathcal{H}^{\text{eff}}_{\downarrow}$ can be calculated separately, instead of simultaneously as in typical NRG programs, the dimensions of the effective Hilbert spaces needed in the VMPS approach and NRG approaches to capture the low-energy properties with the same precision are related by $D_{\text{MPS}} \simeq \sqrt{D_{\text{NRG}}}$, implying significant computational gain with VMPS (i.e. a square root speed-up).

Figure 8 compares the energy level flows calculated using NRG and VMPS. They agree remarkably well, even though we used $D_{\text{MPS}} = 32 = \sqrt{D_{\text{NRG}}}$. The accuracy can still be improved significantly by using larger D and smaller Λ (and exploiting symmetries, not done here for VMPS).

4.3. *DMRG*

As discussed in the previous Section, DMRG can effectively be understood as the more advanced version of NRG in the sense that one also sweeps back and forth.

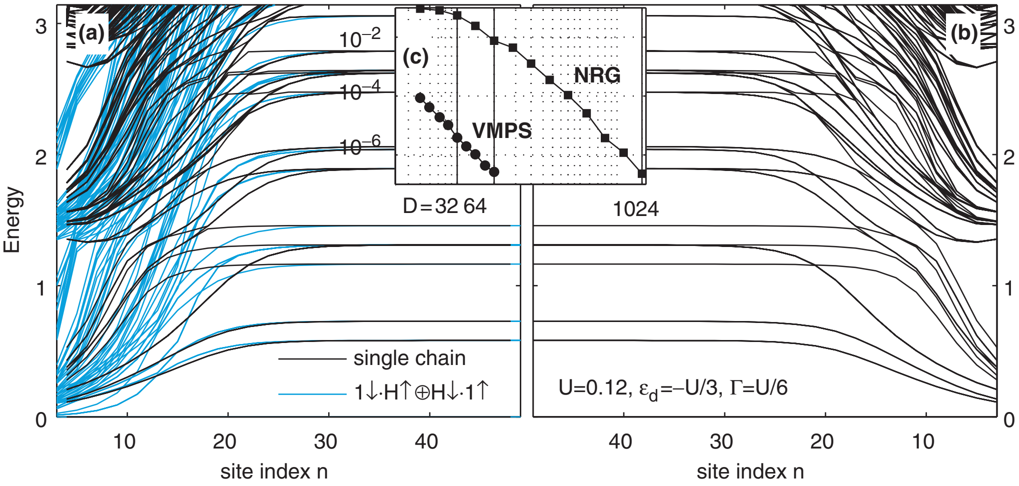

Figure 8. Energy level flow of the SIAM as a function of the site index *n* calculated with (a) VMPS using $D_{\mathrm{MPS}} = 32$ calculated after sweeping and (b) NRG using $D_{\mathrm{NRG}} = 32^2 = 1024$. The blue lines in (a) denote the spectrum of $\mathcal{H}_{\uparrow}^{\mathrm{eff}} \otimes 1_{\downarrow} + 1_{\uparrow} \otimes \mathcal{H}_{\downarrow}^{\mathrm{eff}}$ and demonstrate the decoupling of the ↑ and ↓ spin chains for large *n*. The inset compares the ground state energy as function of *D* for VMPS (circles) and NRG (squares). The subtracted extrapolated ground state energy is $E^{*} \simeq -3.0714$. (Reprinted from [61] with kind permission of the authors.)

Historically, the discovery of the DMRG algorithm had an enormous impact on the field of strongly correlated quantum spin systems as it led to the first algorithm for finding the low-energy spectrum of generic spin chains with local or quasi-local interactions. One of the first checks of the accuracy of the method was the calculation of the ground state energy of the critical spin-1/2 Heisenberg chain, and a dazzling precision was obtained when *D* was taken to be of the order of a few hundred. The DMRG method also allows one to calculate gaps and gave very strong evidence for the existence of the so-called Haldane gap in the case of the spin-1 Heisenberg spin chain.

Still, the VMPS approach discussed in the previous Section and the traditional DMRG differ in crucial details, and when looking back at the way DMRG has been understood and derived, it is not completely obvious to see the parallels with the VMPS approach. We will not explain here how DMRG works, but we refer to the many good review papers on the subject [4,5]. One important difference is that the VMPS approach also works in the case of open as well as PBCs, whereas DMRG uses an ansatz that is specifically tailored for systems with OBCs: the basic reason for this is that DMRG implicitly uses the orthonormalized normal form of MPS discussed in Section 3.1.3, but such a normal form does not exist in the case of PBCs [10]. In practice, simulations of spin chains with PBCs have been done using DMRG, but this is precisely done in the way that was discussed in Section 3.1.4 by folding the MPS with PBCs to one with OBCs by squaring the dimension of the virtual spins $D \rightarrow D^2$; the cost to pay is an algorithm whose computational cost scales as D^6 and, more importantly, for which it is not possible to formulate an ansatz of excited states with a definite momentum (see Section 3.4).

Another difference has to do with the fact that the standard DMRG algorithm is set up in such a way that one performs a variational minimization over two sites together instead of over one site as described in the VMPS. Such as step is then followed by performing a SVD and throwing away the least relevant modes; that step is strictly speaking no longer variational, but there is evidence that the conditioning and the numerical convergence is better when using this method. Another reason why this is done is that the so-called *truncation error* gives some idea about the quality of the energy estimates obtained in a particular DMRG run. However, from the point of view of VMPS, the one-site optimization procedure should perform better for the same computational cost, as the computational cost in the two-site setup goes up with a factor of d^2 and one could better use that time to increase D. This was already observed in the DMRG community [4,85–87], as DMRG can also readily be reformulated with one-site optimizations, but it took until 2005 and a nice paper by White [88] before the numerical conditioning problems were solved (those techniques of course also apply to VMPS) and the one-site DMRG became the standard DMRG method.

However, instead of pointing out the differences, let us discuss a few technical improvements of VMPS that parallel the nice features of DMRG. Following Feynman's credo that you should never underestimate the pleasure of readers to read what they already know, let us repeat again how the whole VMPS procedure works. The goal is to minimize

$$\min_{|\psi^N\rangle \in \{\mathrm{MPS}_D\}} \left[\langle\psi^N|\mathcal{H}^N|\psi^N\rangle - \lambda\langle\psi^N|\psi^N\rangle\right]; \tag{11}$$

which is a multiquadratic optimization problem and can be solved using the ALS method. This leads to the problem of solving the generalized eigenvalue problem $H_{\mathrm{eff}}x = \lambda N_{\mathrm{eff}}x$ which is a standard optimization problem itself that is efficiently solvable if the condition number (i.e. smallest singular value) of N_{eff} is not too small. It happens now that in the case of OBCs, one can always make appropriate gauge transformations (see Section 3.1.3) such as to ensure that N_{eff} is equal to the identity. (Note that in the case of PBCs, no such gauge transformation exists that makes N_{eff} equal to the identity, and hence more heuristic methods have to be used. We discuss this in more detail in the Section dealing with PBCs; that makes the VMPS algorithms with PBCs a bit more tricky to implement, but the basic idea and resulting numerical accuracy is very similar.) Those gauge transformations are always implicitly done in the case of DMRG, and should also be done in the VMPS approach as this obviously leads to the best conditioning (the computational cost for doing so is small; to see this in practice and also to see how easy this VMPS approach is to implement, we refer to Appendix C for the explicit Matlab code for doing so).

Let us now see whether there is any good justification for using this ALS method in the context of minimizing the ground state energy of local spin Hamiltonians, used in both the VMPS and the DMRG approach. Suppose for example that we are optimizing over some site in the VMPS method and the corresponding $D = 128$ while the physical spin dimension is $d = 2$; then we are effectively performing an optimization over $m = \log_d(dD^2) + 1 = 15$ sites, as the number of degrees of freedom we are varying over is equal to the number of degrees of freedom in 15 qubits. Ground states of spin models are typically such that there is a finite correlations length (for critical systems, correlations are also decaying pretty fast), such that 'boundary effects' of more than seven sites away will not play a big role. Colloquially, the bulk convergence to a local optimum cannot occur if the gap in this system of 15 qubits is larger than the effect from that boundary, and this gives a handwaving argument why DMRG almost always converges to the global minimum.

In other words, the success of the MPS approach is related to its inherent capability of patching together solutions of local (e.g. 15 sites) optimization problems, together with the fact that MPS are of course rich enough to approximate ground states of arbitrary ground states that obey an area law (see Appendix B for a proof).

Before proceeding, let us look at a completely different way of performing variational optimizations using MPSs when the Hamiltonian under interest is translational invariant, and let us consider the thermodynamic limit (i.e. infinitely many sites). We know that the ground state will be translational invariant, so we might well use an ansatz that reflects this by choosing all A^i equal to each other. This approach has been originally proposed by Rommer and Östlund [54] and later studied by several other authors (see [86] and references therein). The optimization problem is now reduced to minimizing the expectation value of one term in the Hamiltonian (i.e. energy per site), and this can be calculated as follows: first consider the 'transfer matrix' $E=\sum_\alpha A_\alpha \otimes \bar{A}_\alpha$ and its eigenvalue decomposition $E=\sum_i \lambda_i |r_i\rangle\langle l_i|$ with the λ_i in decreasing order. The energy can now be expressed as

$$E=\frac{1}{\lambda_0^2}\sum_{\alpha_1\alpha_2\alpha_1'\alpha_2'} H_{\alpha_1\alpha_2;\alpha_1'\alpha_2'}\langle l_0|(A_{\alpha_1}\otimes \bar{A}_{\alpha_1'})(A_{\alpha_2}\otimes \bar{A}_{\alpha_2'})|r_0\rangle,$$

where we assumed that the Hamiltonian is only acting on nearest neighbours. This cost function is clearly a very non-linear function of the variables A^i, and standard techniques such as conjugate gradient methods can be used to minimize that expression. However, it happens that this optimization procedure may get stuck in local minima, and the situation only gets worse when increasing D. In comparison with the DMRG or VMPS approach, this method does not seem to work very well for several problems. At first sight, this seems to be strange as in those latter approaches one has many more variables (i.e. one tensor per site). However, this can be understood from the point of view of optimization theory, where it is standard practice to introduce more variables than needed such as to ensure that the problem becomes better behaved. This is precisely what is happening here. However, as we will see later, the idea of using translational-invariant MPSs can be turned into a successful algorithm by combining it with the concept of imaginary time evolution of MPS. It is a bit of a mystery why that approach works better than, for example, conjugate gradient methods, but it is probably related to the inherent robustness of algorithms evolving in imaginary time.

As a last remark, we note that it can be very useful from a computational point of view to exploit symmetries in the system. This can be done both in the DMRG and in the MPS approach, and we refer to [3,89,90] for more details.

4.3.1. *Excitations and spectral functions*

An issue where the usefulness of VMPS really reveals itself is in the context of the study of excitations. In the standard DMRG approach, the idea is basically to store all of the information regarding the ground state and excited states into one big MPS: during the iteration step at, for example, site k, one keeps all tensors $A^1, A^2, \ldots, A^{k-1}, A^{k+1}, \ldots, A^N$ fixed and within the dD^2-dimensional subspace one identifies a number of lowest lying states. One then moves to the next site, and continues until convergence (again, for a more complete understanding of how DMRG works, we refer the interested reader to the many review papers on the subject [4,5]). That procedure is clearly suboptimal and no longer variational, as there is no reason why the same tensors should be used for the ground and excited states. In the worst-case scenario, the tensors A^i are block-diagonal, each block containing the information of a different excited state, and hence the computational cost scales badly with the number of excited states encoded in this manner (i.e. the cost of performing a simulation with k excited states is k^3 times the cost of doing it for the ground state; also, the memory requirements scale as k^2).

Looking back at the VMPS approach described above however, this is a more natural way to deal with excitations at a lower cost, both with respect to memory and computational cost. To repeat again what was explained earlier, one can build up the spectrum in a sequential way: first look for the ground state, after this start over again and find the MPS with minimal energy that is furthermore orthogonal to the ground state, and so further [83]. This procedure does not have any of the drawbacks mentioned above in the case of DMRG, and is fast and reliable, although it requires the whole variational procedure to be run for each required excited state.

If the Hamiltonian under consideration has some symmetries, there might of course be alternative ways of finding excited states. Consider, for example, the spin-1 Heisenberg chain. It is well known that the ground state lives in the sector with total spin 0, and the really interesting quantity in this case is to find the gap between this ground state and the state with minimal energy out of the spin-1 sector. As the total spin commutes with the Hamiltonian, we can add a small magnetic field to the Hamiltonian that plays the role of a chemical potential, and within some parameter range, which can easily be found by doing some numerics, we are guaranteed that the ground state of the complete Hamiltonian will have spin 1. It will, however, still be necessary to implement the procedure mentioned above if more excitation energies are to be found.

Let us continue now and find out whether the formulation of DMRG in terms of MPS also allows one to obtain lower bounds to the energies. More precisely, suppose that the variational method converges to a MPS $|\psi\rangle$ with associated energy $E = \langle\psi|\mathcal{H}|\psi\rangle$. Let us next define the quantity

$$\epsilon = \sqrt{\langle\psi|(\mathcal{H} - E)^2|\psi\rangle}.$$

It is a simple exercise to prove that there exists an exact eigenvalue $E_{\rm ex}$ of $\mathcal{H}$ such that $E \geq E_{\rm ex} \geq E - \epsilon$. We show that, in the case of MPSs, one can calculate the quantity ϵ at essentially the same computational cost as E. This implies that the VMPS approach outlined allows us to obtain both upper and lower bounds to eigenvalues of a given Hamiltonian $\mathcal{H}$; to the best of our knowledge, this is a truly unique feature for a variational method. In typical applications such as the calculation of the ground state energy of Heisenberg antiferromagnets, $\epsilon \simeq 10^{-8}$.

Let us now sketch how ϵ can be calculated efficiently [61]. First of all, we note that expectation values of tensor products of local observables can be calculated by multiplying vectors by matrices:

$$\langle\psi|\hat{O}_1 \otimes \hat{O}_2 \cdots \hat{O}_N|\psi\rangle = A^{[1]}_{\hat{O}_1} A^{[2]}_{\hat{O}_2} \cdots A^{[N-1]}_{\hat{O}_{N-1}} A^{[N]}_{\hat{O}_N}$$

$$\left(A^{[n]}\right)_{\alpha\beta,\alpha'\beta'} = \sum_{ij} P^{i[n]}_{\alpha\alpha'} \bar{P}^{j[n]}_{\beta\beta'} \langle i|\hat{O}_n|j\rangle.$$

Let us now try to evaluate an expression such as $\langle\psi|(\mathcal{H}-E)^2|\psi\rangle$. For simplicity, let us assume that $\mathcal{H}$ represents a spin chain with only nearest neighbour couplings. Naively, one expects that one will have to evaluate of the order of N^2 expectation values of local observables. There is, however, a much more efficient method: going recursively from the first site to the last, one keeps three D^2-dimensional vectors v_0, v_1, v_2 containing terms with respectively $0, 1, 2$ interaction terms (note that $\mathcal{H}^2$ contains at most two interaction terms). At each recursion step, one can easily update v_0 as a function of v_0 and the local terms of the MPS, and equivalently v_1 as a function of v_1 and v_0 plus the local Hamiltonian; v_2 can be updated as a function of the local MPS terms, the local Hamiltonian and v_0, v_1, v_2. Therefore, the computational complexity of calculating ϵ scales as ND^3, just as for the case of evaluating the energy $\langle\psi|\mathcal{H}|\psi\rangle$. Another and perhaps more direct way to see this is do make use of the formalism of MPOs (see Section 5): there we show that every Hamiltonian with only nearest-neighbour interaction has a very simple parametrization as a MPO, and hence also the square of it.

As a side product, this insight allows one to devise efficient algorithms for calculating highly excited eigenstates and eigenenergies of Hamiltonians. Suppose for example that one would like to know the closest eigenstate and eigenenergy to a certain prespecified energy E_{sp}. This could be interesting when one wants to calculate gaps in the spectrum. The variational problem that has to be solved in that case is the following:

$$\min_{|\psi\rangle\in\{\mathrm{MPS}\}} \langle\psi|(\mathcal{H}-E_{\mathrm{sp}})^2|\psi\rangle - \lambda\langle\psi|\psi\rangle$$

This is indeed again a multiquadratic problem that can be solved using the same variational techniques as outlined before, yielding an algorithm with the same computational complexity as the original.

Similar techniques also allow us to calculate Green's functions in a variational way [61]. Green's function have been calculated using NRG methods [91,92], but because of the strong interplay between the different energy levels in that context, the MPS approach should be able to give more precise results. The DMRG techniques that have been developed for calculating Green's functions of spin systems [93–95] are related to our approach but differ in several aspects.

The typical Green's functions of interest are of the form

$$G(\omega+i\eta) = \left\langle\psi\left| f\frac{1}{\mathcal{H}-\omega-i\eta}f^\dagger\right|\psi\right\rangle,$$

where $|\psi\rangle$ represents the ground state and $f^\dagger$ is a creation operator. Here we assume that $|\psi\rangle$ has been calculated using the MPS approach, and we show that this can be done to a very good and controlled precision. The basic idea is now to calculate the Green's function variationally within the set of MPS by finding the unnormalized MPS $|\chi\rangle$, commonly called a correction vector [61], that minimizes the weighted norm

$$\mathcal{N} = \left\| |\chi\rangle - \frac{1}{\mathcal{H}-\omega-i\eta}f^\dagger|\psi\rangle\right\|_{W=(\mathcal{H}-\omega)^2+\eta^2}.$$

Here, we used the notation $\||\xi\rangle\|_W^2 = \langle\xi|W|\xi\rangle$, and the weight $W > 0$ was introduced to make the problem tractable (it should not really affect the precision as all positive

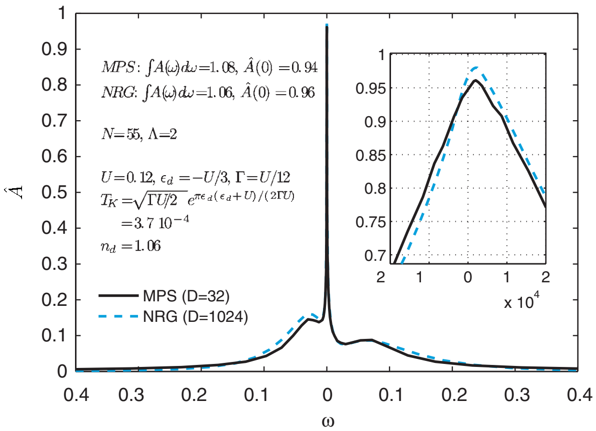

Figure 9. Impurity spectral function for the SIAM, $\hat{\mathcal{A}}_{\uparrow}(\omega) \equiv \mathcal{A}_{\uparrow}(\omega)\pi\Gamma/\sin^2(n_d\pi/2)$, normalized such that $\hat{\mathcal{A}}_{\uparrow}(0)$ should be 1, calculated with VMPS (solid) and NRG (dashed line). (Reprinted from [61] with kind permission of the authors.)

definite norms are essentially equivalent). Writing $|\chi\rangle$ in its real and imaginary part $|\chi\rangle = |\chi_{\rm r}\rangle + i|\chi_{\rm i}\rangle$ and assuming $\mathcal{H}$, $|\psi\rangle$ real, this norm can be written as

$$\mathcal{N} = \langle\chi_{\rm r}|(\mathcal{H}-\omega)^2 + \eta^2|\chi_{\rm r}\rangle - 2\langle\chi_{\rm r}|(\mathcal{H}-\omega)f^{\dagger}|\psi\rangle + \langle\chi_{\rm i}|(\mathcal{H}-\omega)^2 + \eta^2|\chi_{\rm i}\rangle - 2\langle\chi_{\rm i}|f^{\dagger}|\psi\rangle + \langle\psi|\psi\rangle. \tag{12}$$

Minimizing $\mathcal{N}$ clearly involves two independent optimizations over $|\chi_{\rm r}\rangle, |\chi_{\rm i}\rangle$, which we will both parameterize as MPSs. Both of the optimizations involve minimizing the sum of a multiquadratic and a multilinear term; as terms such as $\langle\chi|(\mathcal{H}-\omega)^2|\chi\rangle$ can be calculated efficiently, we can use again the same tricks and keep all but one projectors $\{P^{[i]}\}$ fixed and optimize over the remaining projector. As a multilinear term instead of a quadratic constraint is present, each iteration step can be done efficiently by solving a sparse linear set of equations [61]. Iterating this procedure until convergence, one can evaluate both $|\chi_{\rm r}\rangle, |\chi_{\rm i}\rangle$, and exactly determine the precision of the result by calculating the norm (12); if this norm is too large, one can always increase the control parameter D associated to the MPS $|\chi_{\rm r}\rangle, |\chi_{\rm i}\rangle$. Finally, one can easily evaluate $G(\omega + i\eta)$ as a function of $|\chi\rangle$. The precision of this evaluation can again be bounded. To illustrate this with an example, we again consider the SIAM model (see Figure 9).

4.4. *DMRG and PBCs*

The VMPS approach is applicable to the cases of both OBCs and PBCs. The main difference between the methods is the computational cost of the contraction of the associated tensor network (ND^3 versus ND^5) and the fact that a useful orthonormalization can be used in the case of OBCs. On the other hand, there are several reasons why simulations with PBCs are preferred.

(1) The boundary effects are much smaller than in the case with OBCs. This has been identified as a serious problem for standard DMRG since its conception, as even for very long chains the boundary effects can be seen in the bulk.

(2) Owing to the fact that any non-trivial system with OBCs cannot be translational invariant, it is not possible to study excitations with a definite momentum. Contrasting this to the case of PBCs, it is possible there to construct a variational class of MPSs with a definite momentum, allowing us to, for example, obtain a convenient picture of the energy-momentum relation. The overall price to be paid however is that the computational cost for working with MPSs with PBCs scales as D^5 as opposed to D^3 for OBC.

Note that the traditional DMRG algorithm has also been used to treat systems with PBCs. It is clear, however, from the structure of MPSs that the only way to represent a MPS with PBCs by one with OBCs is to use the virtual bonds to 'teleport' a maximally entangled state from the first to the last site. This leads to an effective D that is equal to the original value squared (this is the only possibly way to accommodate the extra degrees of freedom to take into account this teleportation), and hence the computational cost for the same accuracy would scale as D^6. However, even in doing so, it would be very hard to construct states with a definite momentum.

When implementing the VMPS approach to minimize the energy of a given Hamiltonian with PBCs [10], special precautions have to be taken such that the problem remains well conditioned. In practice, this means that we have to make sure that the matrix N_{eff} in the generalized eigenvalue problem $H_{\text{eff}}x = \lambda N_{\text{eff}}x$ is positive definite and that its smallest eigenvalue is as large as possible. There are several options to achieve this. First of all, we can still play with the gauge conditions which would correspond to a transformation

$$N_{\text{eff}} \to (X \otimes Y \otimes I_d) N_{\text{eff}} \big(X^\dagger \otimes Y^\dagger \otimes I_d\big).$$

Looking back at the case with OBCs, the reason why we could always make $N_{\text{eff}} = I$ is that the left- and right-hand side of the chain factorize such that N_{eff} was always a tensor product to start with. Here, the PBCs enforce the existence of correlations between those sides, hence enforcing N_{eff} to be 'correlated'. However, in practice the number of correlations will be small (as these are finite size effects), and hence N_{eff} will be close to a tensor product. A sensible way of choosing the gauge transformation therefore consists of two steps:

1. find the tensor product $N_1 \otimes N_2 \otimes I_d$ that approximates best N_{eff} (this can easily be achieved by doing a SVD in the space of Hermitian operators);
2. choose the gauge transformations X, Y such that $X N_1 X^\dagger = I = Y N_2 Y^\dagger$.

During the first sweeps, when absolute accuracy is not yet relevant and N_{eff} could still be far from a tensor product, one could, for example, add some identity to N_{eff} such as to make it better conditioned ($N_{\text{eff}} \to N_{\text{eff}} + \epsilon I$). The procedure outlined here is only one of many possible procedures, and it might well be that another choice of gauge conditions leads to better performance, but the procedure described here seems to give good results in practice [10]. An example is provided in Figure 10 where we looked at the Heisenberg spin-1/2 chain with $N = 28$ (we choose this example with a small number of sites as we can still compare this with exact results, and for reference also the results of a DMRG calculation for the same Hamiltonian is provided). As seen in the figure, the scaling of the error for MPSs with PBCs nicely follows the scaling of the error obtained in a DMRG calculation for a Hamiltonian with OBCs.

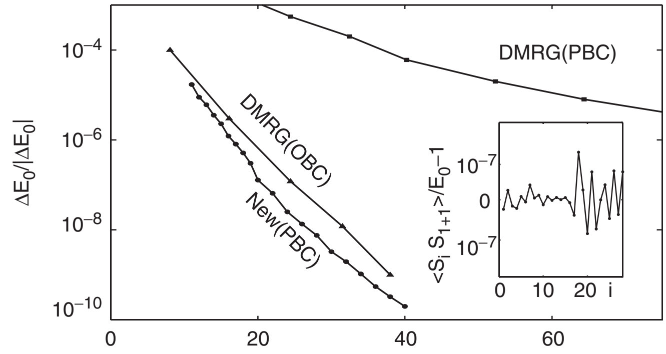

Figure 10. Comparison between DMRG (squares) [3] and the (new) VMPS method (circles) for PBCs, and $N = 28$. For reference the DMRG results [3] for the Heisenberg chain with OBCs (triangles) are also shown. Inset: variation of the local bond strength from the average along the chain, calculated with the new method and $D = 40$. (Reprinted Figure 3 with permission from [10] F. Verstraete *et al.*, Phys. Rev. Lett. 93 (2004), p. 227205. © 2004 by the American Physical Society.)

As already explained in the previous Section, it is easy to look for excited states using an extra Lagrange constraint. However, the VMPS approach can also easily be generalized such as to become a variational method over states with a definite momentum. The basic idea is very simple. Consider the state [83]

$$|\psi_k\rangle = \sum_{n=0}^{N-1} \frac{e^{ikn}}{\sqrt{N}} \hat{T}_n |\chi\rangle, \quad |\chi\rangle = \sum_{\alpha_1,\ldots} \mathrm{Tr}\Big(A^1_{\alpha_1} A^2_{\alpha_2} \ldots A^N_{\alpha_N}\Big) |\alpha_1\rangle |\alpha_2\rangle \ldots |\alpha_N\rangle,$$

where the operator $\hat{T}_k$ is the shift operator implementing a translation over k sites. The state $|\psi_k\rangle$ has the momentum k by construction and is obviously a superposition of N MPSs (note that this implies that the block entropy of the state $|\psi_k\rangle$ can scale as $\log(ND^2)$ as opposed to $\log(D^2)$ for a normal MPS). What is clear, however, is that the whole VMPS procedure outlined above can now be repeated for this extended class of MPSs: the energy is still a multiquadratic function of all tensors A^i involved, and expectation values of local observables can still be calculated efficiently by contracting the corresponding tensor network. The price to pay is an extra factor of N (the number of sites) in all of the calculations, due to the fact that cross terms between the different superpositions enter the optimization. To achieve this rather mild slowdown, a rather involved scheme of bookkeeping has to be maintained, and this is explained in great detail in the paper [83]. As an illustration of the power of this technique, Figure 11 represents the calculation of the energy–momentum relation for two values of the bilinear–biquadratic $S = 1$ spin chain parameterized by the Hamiltonian

$$\mathcal{H} = \sum_i \cos(\theta) \vec{S}_i \vec{S}_{i+1} + \sin(\theta) \big(\vec{S}_i \vec{S}_{i+1}\big)^2.$$

5. Time evolution using MPSs

One of the big advantages of the formulation of renormalization group methods in terms of MPSs is that it also allows us to describe time evolution. This opens up the possibility of simulating the non-equilibrium properties of spin chains, a topic that is currently very

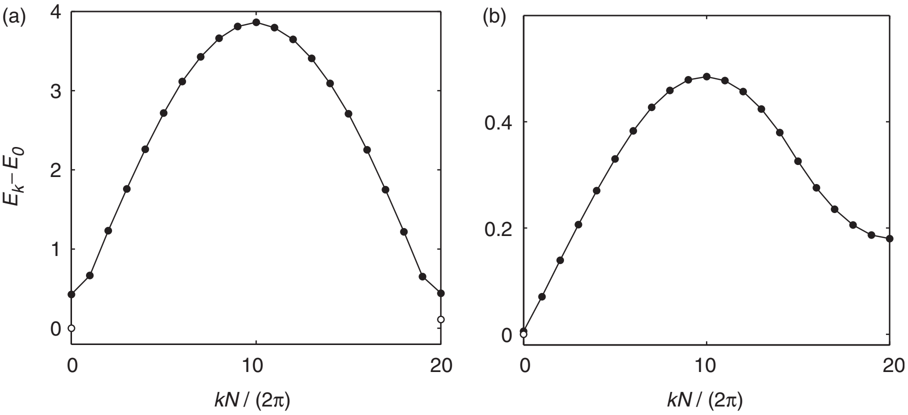

Figure 11. Lowest states of a bilinear–biquadratic $S=1$ chain, $N=40$ sites, $D=10$: (a) $\theta=-\pi/2$, $E_0=-2.7976N$; (b) $\theta=-0.74\pi$, $E_0=-1.4673N$. Empty circles: lowest energy states. Filled circles: first branch of excitations. We estimate the absolute error as $\Delta E_k\approx 10^{-3}$, by comparison with calculations with larger D. (Reprinted Figure 6 with permission from [83] D. Porras *et al.*, Phys. Rev. B 73 (2006), p. 014410. © 2004 by the American Physical Society.)

relevant given the recent breakthroughs in creating strongly correlated quantum spin systems in, for example, optical lattices [96]. Since the development of DMRG, several methods have been proposed [97–99], but the real breakthrough was obtained in a paper of Vidal [69] which was followed up with a variety of related approaches [11,100–104]. In the spirit of this paper, we only review the variational approach, and refer to the review of Schollwöck and White for DMRG-related approaches [105].

5.1. *Variational formulation of time evolution with MPSs*

Mathematically, the problem is to evolve an initial MPS in real time by updating the tensors in the MPS description under Hamiltonian evolution. In practice, this could be used to investigate how excitations travel through the spin chain or to obtain spectral information about a Hamiltonian of interest. In the spirit of the VMPS approach, we would like to do this in a variational way: given a Hamiltonian and an initial MPS, evolve that state within the manifold of MPSs in such a way that the error in approximating the exact evolution is minimized at every infinitesimal step [11].

We are particularly interested in the case where the Hamiltonian, which can be time-dependent, is a sum of local terms of the form

$$\mathcal{H}(t)=\sum_{\langle ij\rangle} f_{ij}(t)\hat{O}_i\otimes\hat{O}_j$$

and where $\langle ij\rangle$ means that the sum has to be taken over all pairs of nearest neighbours (note that the situation with long-range interactions can be treated in a very similar way). There are several tools to discretize the corresponding evolution. This is not completely trivial because generically the different terms in the Hamiltonian do not commute. A standard tool is to use the Trotter decomposition [106,107]

$$e^{A+B}=\lim_{n\to\infty}\left(e^{\frac{A}{n}}e^{\frac{B}{n}}\right)^n.$$

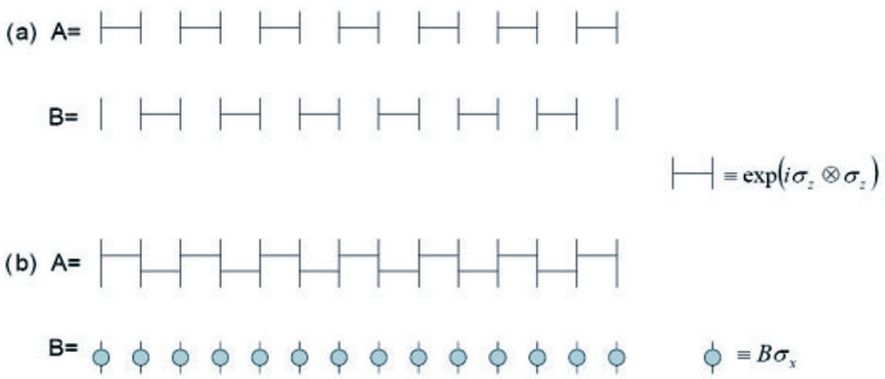

Figure 12. Spin network representation of the different operators obtained in the case of a Trotter expansion of the Ising Hamiltonian in transverse field with (a) nearest-neighbour decomposition and (b) Ising versus magnetic field decomposition.

Suppose, for example, that the Hamiltonian can be split into two parts A and B such that all terms within A and within B are commuting: $\mathcal{H} = A + B$; $A = \sum_i A_i$; $B = \sum_i B_i$; $[A_i, A_j]_- = 0 = [B_i, B_j]_-$. This ensures that one can efficiently represent e^{iA} as a product of terms. The evolution can then be approximated by evolving first under the operator $e^{i\delta t A}$, then under $e^{i\delta t B}$, again under $e^{i\delta t A}$ and so further. The time step can be chosen such as to ensure that the error made due to this discretization is smaller than a prespecified error, and there is extensive literature of how to improve on this by using, for example, the higher order Trotter decompositions [108,109]. In the case of nearest-neighbour Hamiltonians, a convenient choice for A is to take all terms that couple the even sites with the odd sites to the right of it and for B the sites to the left of it. In that case, e^{iA} and e^{iB} are tensor products of nearest-neighbour two-body operators; see Figure 12(a) for a pictorial representation of this in terms of spin networks. However, it is important to note that different choices of A and B might work better in practice. Consider, for example, the Ising Hamiltonian in an transversal field and its decomposition into two different terms A and B (see also Figure 12(b)):

$$\mathcal{H}_{\text{Ising}} = \underbrace{\sum_k \sigma_z^k \otimes \sigma_z^{k+1}}_{A} + H \underbrace{\sum_k \sigma_x^k}_{B}. \tag{13}$$

Obviously, a similar kind of decomposition is possible in the case of the Heisenberg model, but there three terms are needed instead. The advantage of evolution with such a decomposition is that it does not break translational invariance as in the even–odd case.

Let us next investigate how to treat this time evolution in a variational way within the class of MPSs. A sensible cost function to minimize is given by

$$\| A|\psi(k)\rangle - |\psi(k+1)\rangle \|_2^2, \tag{14}$$

where $|\psi(k)\rangle$ is the initial MPS, $|\psi(k+1)\rangle$ is the MPS to be found, and A is the operator arising out of the Trotter expansion. (We remark that as alternative cost function we could have chosen $|\langle\psi(k+1)|A|\psi(k)\rangle|^2/\langle\psi(k+1)|\psi(k+1)\rangle$ which has the aim of finding the normalized MPS $|\psi(k+1)\rangle$ that has maximal overlap with the evolved version of the original MPS. It is, however, an easy exercise to find out that this optimization problem leads to exactly the same optimal MPS, up to a normalization factor.) First of all, it is important to note that this cost function can indeed be evaluated efficiently when $|\psi(k)\rangle$ and $|\psi(k+1)\rangle$ are MPSs and A is of the form discussed above. In principle, this cost function can be made equal to zero by increasing the bond dimension D by a factor of at most d^2 (this is indeed the largest possible Schmidt number of an operator acting on two sites). However, the whole point of using MPSs is that MPSs with low bond dimension are able to capture the physics needed to describe the low-energy sector of the Hilbert space. So if we stay within that sector, the hope is that the bond dimension will not have to be multiplied by a constant factor at each step (which would lead to an exponential computational cost as a function of time), but will hopefully saturate. This is certainly the case if we evolve using, for example, imaginary time evolution of a constant local Hamiltonian, as we know that in that case the ground state is indeed well represented with a MPS with not too large a bond dimension. It is, however, important to keep in mind that an exponential explosion is, in principle, possible for other kinds of evolution: the worst-case computational complexity for time evolution using MPSs is exponential as a function of time.

Taking this into account, a justified way of dealing with time evolution is to prespecify an error ϵ that can be tolerated, and then look for the minimal D for which there exists a MPS $|\psi(k+1)\rangle$ that yields an error smaller than ϵ. Looking back at the cost function (14), it looks pretty familiar how to minimize it: the cost function has only quadratic and linear terms, and we will hence be able to minimize it using a method very similar to the ALS method discussed in the previous Section. More specifically, the cost function has only multiquadratic and multilinear terms in the variables of the MPS, and we will solve this in a recursive way where at each recursion step an optimization of the form

$$x^{\dagger}A_{\text{eff}}x - 2x^{\dagger}y_{\text{eff}}$$

has to be solved. The solution to this problem is the simple solution of the linear set of equations

$$A_{\text{eff}}x = y_{\text{eff}},$$

and this hence leads to a very efficient way of minimizing the cost function for time evolution in a recursive way: sweeping back and forth, we solve the above optimization subproblem at each step, and we are guaranteed that the total cost function goes down at every step and hence will converge. After convergence, we can check how large the error has been for the particular value of D that we choose, and if this error is too large, we can increase D and repeat the optimization. Note that the only requirement for the complete Trotter step evolution to be successful is that the tensor network when sandwiching the operators A and B between two MPSs can be contracted efficiently. It is clear that this method both works in real and imaginary time evolution, and both for time-independent and time-dependent Hamiltonians. As a note, and this will be crucial in the applications of this variational time evolution of MPSs to higher-dimensional spin systems, we also want to point out that we did not use the fact that the operator A was close to the identity, that is, this method is applicable in a more general way than only for small Trotter steps.

Note also that this method is applicable to systems with both OBCs and PBCs, and also with long-range interactions.

5.1.1. *Time-evolving block-decimation*

There has recently been a lot of attention on studying time evolution using MPSs, but instead of using the optimal variational way described above, the vast majority of the works has been using the so-called time-evolving block-decimation (TEBD) procedure introduced by Vidal [69]. The reason for this is that this method can readily be implemented with existing DMRG code [100,101]. It explicitly uses the Schmidt normal form described in Section 3.1.3, and is hence more suitable for MPSs with OBCs. Also, one has to use the Trotter expansion with even–odd sites decomposition. It can also be understood as a variational method, but in the more restricted sense in which we consider the following setup: given a MPS and one (possibly non-unitary) gate acting on two nearest neighbours, find the new MPS with given dimension D that approximates this optimally. This can be done using the SVD: using the normal form for MPSs with OBCs, we know that the original MPS is of the form

$$|\psi\rangle = \sum_{n=1}^{D} |\psi_n^{\mathrm{L}}\rangle |\chi_n^{\mathrm{R}}\rangle$$

with $\{|\psi_n^{\mathrm{L}}\}$ and $\{|\psi_n^{\mathrm{R}}\}$ orthogonal sets. The gate which acts on the nearest neighbours locally increased the dimension of the bond with a factor of at most d^2, and we would like to reduce the dimension of that bond to D again. One can orthonormalize everything again and obtain the Schmidt decomposition over that particular bond, and then the reduction can trivially be done by discarding (i.e. projecting out) the smallest Schmidt coefficients. (This is indeed the virtue of the SVD: if we want to approximate a given matrix with one of lower rank in an optimal way, so as to minimize the Frobenius norm of the difference, then the solution is given by taking the SVD and putting the smallest singular values equal to zero.) In the next step, evolution between the next nearest neighbours is considered, and so on. Note that the variational method discussed above could deal with all those gates at once; nevertheless, the computational cost of both methods is very similar, and in practice both methods seem to achieve a similar accuracy if the time steps are small. The first, however, can be improved by choosing longer time steps and higher Trotter orders [103].

It is also possible to devise more sophisticated methods that somehow make use of the fact that superpositions of MPSs can still be dealt with in an efficient way. For a nice review that compares all such methods and those described before, see [103]. For nice examples of the power of real-time evolution, we refer the reader to the review article [105].

5.1.2. *Exchanging time and space*

In the previous exposition on how to treat time evolution in MPS systems, the space and time directions were treated on distinct footings: the goal was to find good approximations to the states $|\psi(t)\rangle$. However, in the end we are mainly interested in calculating expectation values of some specific operators $\langle\psi(t)|O|\psi(t)\rangle$. Using a Trotter decomposition as before, the exact calculation of this quantity involves the contraction of a tensor network as

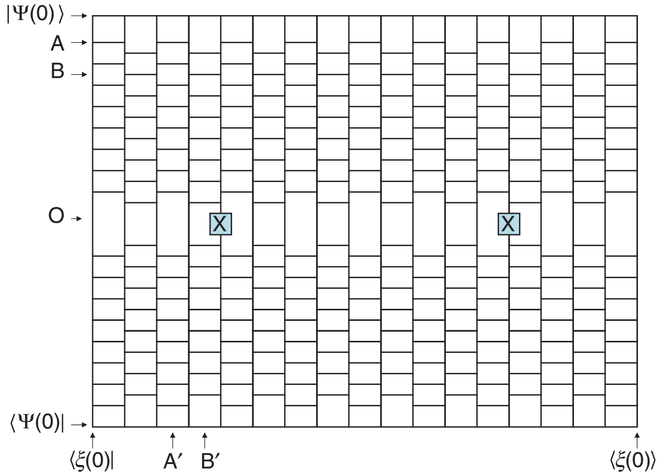

Figure 13. When contracting the spin network for time evolution, one can also exchange time and space directions. In this particular case, this depicts a network for calculating the X–X correlation function for the state $|\psi(0)\rangle$ evolved with eight Trotter steps parameterized by the operators A, B. This is equivalent to evolving the state $|\xi\rangle$ with different operators A', B'. The operators A', B' are obtained trivially by exchanging the physical and virtual dimensions of A, B.

depicted in Figure 13. The time-evolution algorithms described above are particular ways of approximating the contraction of this tensor network. However, it is clear that within this picture, space and time directions can easily be exchanged, leading to an alternative way of calculating expectation values. This is particularly interesting when we want to perform, for example, imaginary time evolution for a finite spin chain but where we want the time steps go to infinity: exchanging space and time, this is equivalent to evolving an infinite MPS over a finite period of time, and we will describe how to do this in the next Section. Note also that this picture of reinterpreting the problem of time evolution in terms of the contraction of spin networks turns out to be the central tool needed to perform calculations with PEPSs, the extensions of MPSs to higher dimensions.

5.2. *Finding ground states by imaginary time evolution*

The tools discussed in the previous Section apply both to real and imaginary time evolution. This provides a completely different way of finding ground states of local quantum spin Hamiltonians: if we start with an arbitrary state $|\psi_0\rangle$, evolution in imaginary time leads to

$$|\psi(t)\rangle = e^{-Ht}|\psi_0\rangle = \sum_{k=1}^{n} e^{-\lambda_k t}|\chi_k\rangle\langle\chi_k|\psi_0\rangle$$

with $H = \sum_{k=1}^{n} \lambda_k |\chi_k\rangle\langle\chi_k|$ the eigenvalue decomposition of H. Hence, as long as the ground state is not degenerate and for times longer than the inverse gap, the state $|\psi(t)\rangle$ will converge exponentially fast to the ground state, and the speed of convergence is exactly quantified by the gap. This is indeed something that is a recurring theme: the smaller the gap, the slower all variational methods seem to converge. This is of course not unexpected because it is difficult to discriminate between excited states with small energy and the ground state. However, it is interesting to note that the closer a system is to criticality (i.e. gap goes to zero), the larger D has to be for a MPS to approximate the ground state for a fixed error (see Appendix B, which shows that in critical systems this scaling is polynomial in the number of spins in the worst case scenario). A very interesting question would be to relate the density of states above the gap to the decay of the Schmidt coefficients that one obtains by dividing the ground state into two pieces.

In practice, finding a ground state using imaginary time evolution is pretty reliable. Of course, the time steps have to be adjusted such that they become smaller and smaller, but one of the great features of imaginary time evolution is its inherent robustness: it does not really matter if one makes mistakes in the beginning, as there is anyway an exponential convergence to the ground state afterwards. This is in contrast to the evolution in real time.

5.2.1. *Infinite spin chains*

Another advantage of the imaginary time evolution approach is that one can easily treat systems with PBCs or treat the thermodynamic limit (number of sites $\to\infty$) by making use of the inherent translational invariance of MPSs with all tensors A^i equal to each other.

To illustrate how this can be done, let us consider the specific case treated in [72] and assume that we want to model the ground state of the Ising Hamiltonian in a transverse field defined on a spin chain. As discussed in Section 4.1 following equation (13), one can choose the decomposition in the Trotter step in such a way that both the operators A and B

$$A = \sum_k \sigma_z^k \otimes \sigma_z^{k+1}, \quad B = h \sum_k \sigma_x^k,$$

are completely translational invariant, and hence we can stay within the manifold of translational-invariant states to describe its evolution. The next step is to see what $\exp(\delta t A)$ and $\exp(\delta t B)$ look like. The latter is simple, as it is just equal to

$$\exp(\delta t B) = \otimes_k \exp(\delta t h \sigma_x).$$

The former expression is obviously a product of commuting operators, and a simple exercise allows one to see that there is a simple matrix product description for it:

$$\begin{aligned}
\exp(\delta t A) &= \sum_{i_1 i_2 \ldots} \mathrm{Tr}(C_{i_1} C_{i_2} \ldots) X_{i_1} \otimes X_{i_2} \otimes \ldots, \\
C_0 &= \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}, \quad C_1 = \begin{pmatrix} 0 & \sqrt{\sinh(\delta t)\cosh(\delta t)}, \\ \sqrt{\sinh(\delta t)\cosh(\delta t)} & 0 \end{pmatrix}, \\
X_0 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.
\end{aligned} \tag{15}$$

This can justifiably be called a MPO (see the next Section). The associated bond dimension is two, and this means that when acting on a MPS of dimension D with this MPO, the exact representation of the new MPS will at most be $2D$. In this particular example of the Ising Hamiltonian, imaginary time evolution would now amount to act subsequently with

$\exp(\delta tA)$ and $\exp(\delta tB)$ on an initial state. Combined together, it happens that their product $Q = \exp(\delta tB/2)\exp(\delta tA)\exp(\delta tB/2)$ is again exactly of the form (15) with unchanged tensor C_i and X_i but where we have to replace $X_0 = I$ with $X_0 = \exp(\delta t\sigma_x)$. So Q is a very simple MPO with bond dimension two, and the goal is to evolve a translational-invariant MPS using this MPO. This can be done in a very simple way: given a MPS with tensor A_i of dimension D, the action of Q is such that we have to replace

$$A_i \rightarrow \sum_{k,l} A_k \otimes C_l \langle i|X_l|k\rangle.$$

The new MPS corresponding to this A_i has bond dimension $2D$, and this has to be reduced as otherwise its size would increase exponentially. The optimal choice is again simple if we consider a system with OBCs: we consider its normal form (Section 3.1.3), and cut it in the appropriate way without losing translational invariance. (More specifically, a possible procedure is as follows [72]: calculate the leading left $|v_l\rangle$ and right eigenvector $|v_r\rangle$ of $\sum_k C_k \otimes \bar{C}_k$; note that this can be done in a sparse way. Reshape those eigenvectors in the square matrices X_l and X_r, and as those matrices are the fixed points of the CP-maps $\sum_k C_k \cdot C_k^\dagger$ and $\sum_k C_k^\dagger \cdot C_k$, X_l and X_r are guaranteed to be positive. Next take the SVD of $\sqrt{X_r}\sqrt{X_l} = U\Sigma V^\dagger$, and reduce the number of columns of U and V to D and discard the D lowest singular values in Σ; U and V hence become $2D \times D$ matrices and Σ a $D \times D$ matrix. Define $G_r = \sqrt{X_r^{-1}}U\sqrt{\Sigma}$ and $G_l = \sqrt{\Sigma}V^\dagger\sqrt{X_l^{-1}}$, and make the transformation $A_i \rightarrow G_l \cdot A_i G_r$ such that it corresponds to a MPS with D-dimensional bonds instead of $2D$.)

The cutting step is not so easy to justify rigorously, but seems to work very well in practice: what we do is calculate the Schmidt normal form with respect to the $2D$-dimensional bonds, and then cut all bonds together. The tricky thing is that cutting the bond somewhere changes the Schmidt coefficients somewhere else, but the point is that these changes are only of the order of the Schmidt coefficients that are cut and those are very small anyway. Amazingly, this procedure works very well, and even a small bond dimension of $D = 32$ already reproduces the ground state energy density E of the critical Ising model ($h = 1$) up to a precision better than $E - 4/\pi < 10^{-7}$. Much better conditioning can also be obtained [72] by working with Hermitian matrices $\{A^i\}$.

Clearly, this procedure can be repeated for any translational-invariant Hamiltonian such as the Heisenberg model (note that the bond dimension will be larger there), and the big advantage is that it allows us to treat infinite systems. The finite system with PBCs can be dealt with in a similar way, although the cutting procedure is more involved there.

A variant of this procedure can be obtained by using the even–odd Trotter decomposition. In that case, exact translational invariance is broken, but it is still perfectly translational invariant with period 2:

$$|\psi\rangle = \sum_{i_1 i_2 \ldots} \ldots A_{i_1} B_{i_2} A_{i_3} B_{i_4} \ldots |i_1\rangle|i_2\rangle|i_3\rangle|i_4\rangle.$$

This type of imaginary time evolution has been studied in detail by Vidal [9], and convergence seems to be fast. The updating and cutting works in a very similar way as discussed in the example above. Note that from the point of view of variational states, it could be advantageous to work with such an $ABAB\ldots$ scheme when an antiferromagnetic ordering is expected.

6. MPOs

Instead of restricting our attention to pure MPSs, we can readily generalize the approach and deal with MPOs. In its most general case, a MPO is defined as [8,11]

$$\hat{O} = \sum_{i_1 i_2 \ldots} \mathrm{Tr}\Big(A^1_{i_1} A^2_{i_2} \ldots\Big) \sigma_{i_1} \otimes \sigma_{i_2} \otimes \ldots \tag{16}$$

with σ_i a complete single particle basis (e.g. the Pauli matrices for a qubit). We have already encountered an example of such a MPO in the previous Section, where the evolution following a Trotter step was expressed by a MPO; that MPO essentially played the role of a transfer matrix during the evolution. As we will show later, any transfer matrix arising in the context of the classical partition function will have an exact representation in terms of such a MPO. This is the reason why all of the renormalization methods described above can also be used in the context of two-dimensional classical spin systems. As a side remark, it is also true that any translational-invariant spin Hamiltonian with nearest-neighbour interactions has an exact MPO representation. For this, let us consider the spin-1/2 case. We first need the fact that there always exists a basis such that the Hamiltonian is of the form

$$\mathcal{H} = \sum_{\alpha,i} \lambda_\alpha \sigma^i_\alpha \otimes \sigma^{i+1}_\alpha + \sum_j \hat{O}^j$$

where $\hat{O}$ can be any one-qubit operator. It is now a small exercise to prove that

$$\mathcal{H} = \sum_{i_1 i_2 \ldots} \big(v_l A_{i_1} A_{i_2} \ldots v_r^T\big) X_{i_1} \otimes X_{i_2} \otimes \ldots,$$

$$X_0 = I \qquad X_1 = \sigma_x \qquad X_2 = \sigma_y \qquad X_3 = \sigma_z \qquad X_4 = \hat{O},$$

$$v_l = |0\rangle \qquad v_r = |4\rangle,$$

$$A_0 = |0\rangle\langle 0| + |4\rangle\langle 4|, \quad A_1 = |0\rangle\langle 1| + |1\rangle\langle 4| \qquad A_2 = |0\rangle\langle 2| + |2\rangle\langle 4|,$$

$$A_3 = |0\rangle\langle 3| + |3\rangle\langle 4|, \quad A_4 = |0\rangle\langle 4|.$$

Actually, one can easily prove that $D = 5$ is optimal because this is the operator Schmidt number of the Hamiltonian when splitting it into two pieces.

MPOs have been shown to be very useful in obtaining spectral information about a given Hamiltonian. Examples are the parametrization of Gibbs states (Section 5.1), the simulation of random quantum spin systems (Section 5.2), the calculation of classical partition functions (Section 5.3), and the determination of the density of states (Section 5.4).

If a MPO is a positive operator and has trace one, then it becomes a matrix product density operator (MPDO). A simple systematic way of constructing a MPDO [11] is by making use of the fact that every mixed state can be seen as a part of a bigger pure system, namely its purification. If we model this purification as a MPS, then the MPDO is obtained by tracing over the purifying degrees of freedom. This picture is especially useful if the purification is such that to every original spin, a locally accompanying purifying spin is present, and that the pure state is such that these pairs of spins correspond to one bigger site:

$$\begin{aligned}
|\psi\rangle &= \sum_{i_1 i_2 \dots} \mathrm{Tr}\Big(A^1_{i_1 j_1} A^2_{i_2 j_2} \dots\Big)|i_1\rangle|j_1\rangle \otimes |i_2\rangle|j_2\rangle \otimes \dots, \\
\rho &= \mathrm{Tr}_{j_1 j_2 \dots}(|\psi\rangle\langle\psi|), \\
&= \sum_{i_1 i_2 \dots i'_1 i'_2 \dots} \mathrm{Tr}\left(\underbrace{\sum_{j_1 j'_1} A^1_{i_1 j_1} \otimes \bar{A}^1_{i'_1 j'_1}}_{E^1_{i_1 i'_1}} \underbrace{\sum_{j_2 j'_2} A^2_{i_2 j_2} \otimes \bar{A}^2_{i'_2 j'_2}}_{E^2_{i_2 i'_2}} \dots \right) |i_1\rangle\langle i'_1| \otimes |i_2\rangle\langle i'_2| \otimes \dots, \\
&= \sum_{i_1 i_2 \dots i'_1 i'_2 \dots} \mathrm{Tr}\Big(E^1_{i_1 i'_1} E^2_{i_2 i'_2} \dots\Big)|i_1\rangle\langle i'_1| \otimes |i_2\rangle\langle i'_2| \otimes \dots. \qquad (17)
\end{aligned}$$

This representation of MPDOs in terms of purifications will turn out to be very useful to describe Gibbs states in thermal equilibrium. Note that not all MPO can arise from locally tracing out an ancilla: the tensors E^i arising in the above description are very special in the sense that they correspond to completely positive maps and hence live in a convex positive cone.

6.1. *Finite-temperature systems*

MPOs are very convenient method for approximating Gibbs states of local Hamiltonians. More specifically, we would like to approximate the operator

$$e^{-\beta\mathcal{H}} = e^{-(\beta/2)\mathcal{H}} I e^{-(\beta/2)\mathcal{H}},$$

where I is the identity operator. In the spirit of last few Sections, this amounts to evolving the maximally mixed state in imaginary time for a period $\beta/2$ (see [8,11]). However, instead of implementing this evolution on the MPO directly, a square root speed up in computational complexity can be gained (note that such a square root speed-up is guaranteed when the temperature is very low as there as the number of classical correlations is zero for pure states) by considering the evolution on its *purification* (see the previous Section). More specifically, the initial state I can be seen as a collection of half of the maximally entangled pairs. In the notation of equation (17), the initial state is a MPS with $D = 1$ and where $A_{ij} = \delta_{ij}$. We can now use the variational algorithm discussed above to evolve this initial state over a time $\beta/2$, and once we have that state it can be used, for example, to calculate the free energy and any correlation function.

Note that without time evolution, it would be very hard to perform a variational calculation over all possible MPOs to approximate the Gibbs state, as that state is variationally characterized by the condition that it minimizes the free energy $\mathrm{Tr}(\rho\mathcal{H}) - TS(\rho)$; the problem here is the calculation of the von Neumann entropy $S(\rho)$, which we do not know how to calculate for MPDOs in general. However, as a by-product of the imaginary time approach, we can readily calculate the entropy of the Gibbs state ρ

$$S(\rho) = \log \mathrm{Tr}\big(e^{-\beta H}\big) + \beta \frac{\mathrm{Tr}\big(He^{-\beta H}\big)}{\mathrm{Tr}(e^{-\beta H})},$$

as all of the quantities on the right-hand side of this equation can efficiently be calculated and approximated using the MPS/MPO representation obtained via evolution in imaginary time.

Let us now see whether it is possible to calculate non-equilibrium properties of such MPOs by evolving them. The most general type of evolution allowed by quantum mechanics is described by the Lindblad equation, which takes into account noise and can therefore increase the entropy. In general, this means that the system is interacting with an extra auxiliary system, and this poses problems for treating the evolution via the purification described above: the dimension of the ancilla has to grow continuously. The straightforward remedy is to work with the full MPO picture, and evolution can again be treated in a straightforward variational way [8,11].

6.2. *Random quantum spin systems*

Another non-trivial application of MPOs and more specifically of the corresponding purifications is the fact that those allow us to simulate random quantum spin systems in a highly efficient way. By making effective use of the entanglement present between system and an appropriately chosen ancilla, it happens that one can simulate exponentially many realizations of the random system in one run; this is quantum parallelism in the strongest possible sense. To explain this, we closely follow the exposition given in the paper [15].

Let us consider a quantum system with Hilbert space $\mathcal{H}$ that evolves accordingly to a Hamiltonian $H(r_1, \ldots, r_n)$ where $r_1, \ldots, r_n$ are random variables that take values within a finite discrete set, $r_\ell \in \Gamma_\ell = \{\lambda_1^\ell, \ldots, \lambda_{m_\ell}^\ell\}$, with a probability distribution given by $p(r_1, \ldots, r_n)$. In order to simulate exactly the dynamics of such a system one would need to perform $\prod_{\ell=1}^n m_\ell$ simulations, one per each possible realization of the set of random variables $\mathbf{r} = (r_1, \ldots, r_n)$. For each realization the system evolves to a different state $|\psi_\mathbf{r}(t)\rangle = e^{-iH(\mathbf{r})t/\hbar}|\psi_0\rangle$, where $|\psi_0\rangle$ is the initial state. Given this set of evolved states and a physical observable $\hat{O}$, one is typically interested in the average of the expectation values of that observable in the different evolved states, that is, in quantities of the form:

$$\langle\langle \hat{O}(t)\rangle\rangle := \sum_\mathbf{r} p(\mathbf{r})\langle\psi_\mathbf{r}(t)|\hat{O}|\psi_\mathbf{r}(t)\rangle. \tag{18}$$

In the following we describe an algorithm that allows us to simulate in parallel all possible time evolutions of the random system described above. We consider an auxiliary system with Hilbert space $\mathcal{H}_a$ and a Hamiltonian acting on $\mathcal{H} \otimes \mathcal{H}_a$ of the form $\tilde{H} = H(\hat{R}_1, \ldots, \hat{R}_n)$, where $\hat{R}_1, \ldots, \hat{R}_n$ are operators that act in $\mathcal{H}_a$, commute with each other and have spectra $\Gamma_1, \ldots, \Gamma_n$. Note that we have replaced the set of random variables $\mathbf{r}$ by a set of quantum operators $\hat{\mathbf{R}}$ with the same spectra. The algorithm works as follows.

(1) *Initialization*. Let us prepare the auxiliary system in an initial superposition state of the form:

$$|\psi_a\rangle = \sum_\mathbf{r} \sqrt{p(\mathbf{r})}\,|\mathbf{r}\rangle, \tag{19}$$

where the states $|\mathbf{r}\rangle$ are simultaneous eigenstates of the set of operators $\hat{\mathbf{R}}$, with $\hat{R}_\ell|\mathbf{r}\rangle = r_\ell|\mathbf{r}\rangle$. Each state $|\mathbf{r}\rangle$ is therefore in one-to-one correspondence with one realization of the set of random variables $\mathbf{r}$, its weight in the superposition state (19) being equal to the probability with which the corresponding realization occurs for the random system.

(2) *Evolution*. We evolve the initial state of the composite system $|\psi_0\rangle \otimes |\psi_a\rangle$ under the Hamiltonian $\widetilde{H}$. The evolved state is

$$|\Psi(t)\rangle = \sum_{\mathbf{r}} \sqrt{p(\mathbf{r})}\, |\psi_{\mathbf{r}}(t)\rangle \otimes |\mathbf{r}\rangle. \tag{20}$$

This superposition state contains the complete set of evolved states we are interested in

(3) *Read-out*. In order to obtain the quantities (18) we just need to measure the observable $\hat{O} \otimes 1$,

$$\langle \Psi(t)|\hat{O} \otimes 1|\Psi(t)\rangle = \langle\langle \hat{O}(t)\rangle\rangle. \tag{21}$$

The algorithm above allows us, in particular, to obtain the averaged properties of a random system over the collection of all possible *ground states*. Let us assume that the interaction between the system and the ancilla is introduced adiabatically, so that the Hamiltonian is now $\widetilde{H}(t) = H(\beta(t)\hat{\mathbf{R}})$, where $\beta(t)$ is a slowly varying function of time with $\beta(0) = 0$, $\beta(T) = 1$, T being the time duration of the evolution. If the system is prepared in the ground state of the Hamiltonian $H(\mathbf{0})$, the algorithm above will simulate in parallel all possible adiabatic paths, so that the composite superposition state (20) will contain all possible ground states of the random system [15].

In addition, the scheme above can be easily extended for the computation of other moments of the distribution of physical observables (higher than (18)), which are sometimes important in the understanding of quantum random systems (QRS) [15]. For example, quantities such as $\langle\langle \hat{O}^2\rangle - \langle \hat{O}\rangle^2\rangle$, can be computed by using an additional copy of the system.

The algorithm described above reduces the simulation of a quantum random system to the simulation of an equivalent non-random interacting problem. This exact mapping allows us to integrate the simulation of randomness in quantum systems within the framework of numerical methods that are able to efficiently simulate the corresponding interacting problem. As an illustrative example we consider the case of a one-dimensional spin $s = 1/2$ system with random local magnetic field. The Hamiltonian of the system is

$$H(b_1, \ldots, b_N) = H_0 + B \sum_{\ell=1}^{N} b_\ell S_\ell^z, \tag{22}$$

where H_0 is a short-range interaction Hamiltonian, $\mathbf{b} = (b_1, \ldots, b_N)$ is a set of classical random variables that take values $\{1/2, -1/2\}$ with probability distribution $p(\mathbf{b})$. Following the algorithm above the 2^N simulations required for the exact simulation of the dynamics (or the ground-state properties) of this random problem can be simulated in parallel as follows. We consider an auxiliary one-dimensional spin $\sigma = 1/2$ system. We prepare this ancilla in the initial state $|\psi_a\rangle = \sum_{\mathbf{b}} \alpha_{\mathbf{b}} |\mathbf{b}\rangle$, where the states $|\mathbf{b}\rangle$ have all z components of the N spins well defined, $\widehat{\sigma_\ell^z}|\mathbf{b}\rangle = b_\ell|\mathbf{b}\rangle$, and $\alpha_{\mathbf{b}} = \sqrt{p(\mathbf{b})}$. The entangled properties of the state of the ancilla reflect the classical correlations among the random variables. For example, for a uniform distribution of the random field, $p(\mathbf{b}) = 1/2^N$, the state of the ancilla is just

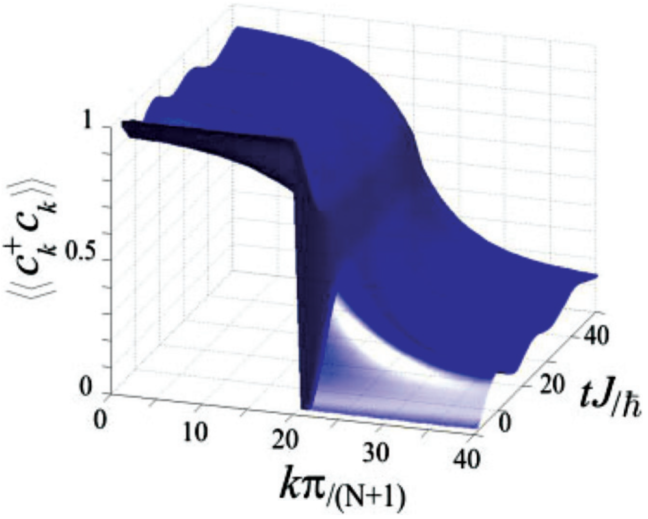

Figure 14. Numerical simulation of the time evolution of a random field XY spin chain with $N = 40$. We show the correlation function $\langle\langle c_k^\dagger c_k \rangle\rangle$ as a function of time and momentum k. Here $c_k \propto \sum_\ell sin(k\ell)\widetilde{c}_\ell$, $k = \pi/(N+1), \ldots, \pi N/(N+1)$ and $\widetilde{c}_\ell = \prod_{\ell<\ell'} S^z_{\ell'}(S^x_\ell + iS^y_\ell)$ are the fermionic operators given by the Jordan–Wigner transformation. The evolution Hamiltonian is $H(b_1, \ldots, b_N) = H_0 + B\sum_{\ell=1}^N b_\ell S^z_\ell$ with H_0 being the XY Hamiltonian with $B_0 = 0$, $B/J = -4$ and $p(\mathbf{b}) = 1/2^N$. The initial state $|\psi_0\rangle$ is the ground state of H_0. As the system evolves in time the initially sharp Fermi sea disappears.

a product state, $|\psi_a\rangle \propto (|\uparrow\rangle + |\downarrow\rangle)^{\otimes N}$. We evolve the system and the ancilla under the interaction Hamiltonian

$$\widetilde{H} = H(\widehat{\sigma}_1^z, \ldots, \widehat{\sigma}_n^z) = H_0 + \beta \sum_\ell \widehat{\sigma}_\ell^z \widehat{S}_\ell^z. \quad (23)$$

Here, $\beta = B$ if we want to simulate dynamics under the Hamiltonian (22), and β is a slowly varying function of time with $\beta(0) = 0$ and $\beta(T) = B$ for the simulation of the ground state properties. We have then reduced the simulation of the random problem to that of the time evolution of two coupled spin-1/2 chains with the Hamiltonian (23). This problem is equivalent to a one-dimensional lattice problem of N sites with physical dimension $d = 2 \times 2$, which can be easily incorporated into the framework of the variational numerical methods.

6.3. *Classical partition functions and thermal quantum states*

In this Section, we show how the concept of MPOs can be used to calculate the free energy of a classical two-dimensional spin system. This also leads to an alternative way of treating thermal states of one-dimensional quantum spin systems, as those can be mapped to classical two-dimensional spin systems by the standard classical-quantum mapping. (In the context of MPSs and especially PEPSs, there also exist a different mapping between classical and quantum spin models in the same dimension [42]. There, the thermal classical fluctuations map onto quantum ground state fluctuations, and this leads to a lot of insight into the nature of quantum spin systems.) Historically, Nishino was the first to pioneer the use of DMRG techniques in the context of calculating partition functions of classical spin systems [110]. By making use of the Suzuki–Trotter decomposition, his method has then subsequently been used to calculate the free energy of translational-invariant one-dimensional quantum systems [111–113], but the main restriction of those methods is that it cannot be applied in situations in which the number of particles is finite and/or the system is not homogeneous; furthermore, one has to explicitly use a system with PBCs in the quantum case, a task that is not well suited for standard DMRG. The VMPS-approach gives an easy solution to those problems.

Our method relies on reexpressing the partition and correlation functions as a contraction of a collection of 4-index tensors, which are arranged according to a two-dimensional configuration [14]. We perform this task for both two-dimensional classical and one-dimensional quantum systems.

Let us consider first the partition function of an inhomogeneous classical two-dimensional n-level spin system on an $L_1 \times L_2$ lattice. For simplicity we concentrate on a square lattice and nearest-neighbour interactions, although our method can be easily extended to other short-range situations. We have

$$Z = \sum_{x^{11},\ldots,x^{L_1L_2}} \exp\left[-\beta H(x^{11},\ldots,x^{L_1L_2})\right],$$

where

$$H(x^{11},\ldots) = \sum_{ij}\left[H^{ij}_{\downarrow}(x^{ij},x^{i+1,j}) + H^{ij}_{\rightarrow}(x^{ij},x^{i,j+1})\right]$$

is the Hamiltonian, $x^{ij} = 1,\ldots,n$ and β is the inverse temperature. The SVD allows us to write

$$\exp\left[-\beta H^{ij}_{q}(x,y)\right] = \sum_{\alpha=1}^{n} f^{ij}_{q\alpha}(x) g^{ij}_{q\alpha}(y),$$

with $q \in \{\downarrow,\rightarrow\}$. Defining the tensors

$$X^{ij}_{lrud} = \sum_{x=1}^{n} f^{ij}_{\downarrow d}(x) g^{i-1,j}_{\downarrow u}(x) f^{ij}_{\rightarrow r}(x) g^{i,j-1}_{\rightarrow l}(x),$$

the partition function can now be calculated by contracting all 4-index tensors X^{ij} arranged on a square lattice in such a way that, for example, the indices l, r, u, d of X^{ij} are contracted with the indices r, l, d, u of the respective tensors $X^{i,j-1}, X^{i,j+1}, X^{i-1,j}, X^{i+1,j}$. In order to determine the expectation value of a general operator of the form $O(\{x^{ij}\}) = Z\prod_{ij} O^{ij}(x^{ij})$, one just has to replace each tensor X^{ij} by

$$X^{ij}_{lrud}(O^{ij}) = \sum_{x=1}^{n} O^{ij}(x) f^{ij}_{\downarrow d}(x) g^{i-1,j}_{\downarrow u}(x) f^{ij}_{\rightarrow r}(x) g^{i,j-1}_{\rightarrow l}.$$

The quantum case if very similar. We consider the partition function of an inhomogeneous one-dimensional quantum system composed of L n-level systems,

$$Z = \operatorname{tr}\exp(-\beta H).$$

It is always possible to write the Hamiltonian H as a sum $H=\sum_k H_k$ with each part consisting of a sum of commuting terms. Let us, for simplicity, assume that $H=H_1+H_2$ and that only local and two-body nearest-neighbour interactions occur, that is, $H_k=\sum_i O_k^{i,i+1}$ and $\left[O_k^{i,i+1}, O_k^{j,j+1}\right]=0$, with $i,j=1,\ldots,L$. The more general case can be treated in a similar way. Let us now consider a decomposition

$$\exp\left(-\frac{\beta}{M}O_k^{i,i+1}\right)=\sum_{\alpha=1}^{\kappa}\hat{S}_{k\alpha}^{i}\otimes\hat{T}_{k\alpha}^{i+1}. \tag{24}$$

The SVD guarantees the existence of such an expression with $\kappa\le n^2$. As we will see later, a smart choice of $H=\sum_k H_k$ can typically decrease κ drastically. Making use of the Suzuki–Trotter formula (note that, in practice, it will be desirable to use the higher order versions of the Trotter decomposition)

$$Z=\mathrm{Tr}\left(\prod_k \exp\left(-\frac{\beta}{M}H_k\right)\right)^M+\mathrm{O}\left[\frac{1}{M}\right]$$

it can be readily seen that the partition function can again be calculated by contracting a collection of 4-index tensors X^{ij} defined as

$$X^{ij}_{(ll')(rr')ud}\equiv\left[\hat{T}^j_{1l}\hat{S}^j_{1r}\hat{T}^j_{2l'}\hat{S}^j_{2r'}\right]_{[ud]},$$

where the indices (l,l') and (r,r') are combined to yield a single index that may assume values ranging from 1 to κ^2. Note that now the tensors X^{ij} and $X^{i'j}$ coincide, and that the indices u of the first and d of the last row have to be contracted with each other as well, which corresponds to a classical spin system with PBCs in the vertical direction. A general expectation value of an operator of the form $O=ZO^1\otimes\cdots\otimes O^N$ can also be reexpressed as a contraction of tensors with the same structure: it is merely required to replace each tensor X^{1j} in the first row by

$$X^{1j}_{(ll')(rr')ud}(O^j)=\left[O^j\hat{T}^j_{1l}\hat{S}^j_{1r}\hat{T}^j_{2l'}\hat{S}^j_{2r'}\right]_{[ud]}.$$

Let us next move on to explaining how the tensor contraction can be performed. We use the techniques that were originally developed in the context of PEPSs in order to contract the tensors X^{ij} introduced above in a controlled way. The main idea is to express the objects resulting from the contraction of tensors along the first and last column in the two-dimensional configuration as MPSs and those obtained along the columns $2,3,\ldots,L-1$ as MPOs. More precisely, we define

$$\langle\mathbb{X}^1|:=\sum_{r_1\ldots r_M=1}^{m}\mathrm{tr}\left(\mathbb{X}^{11}_{r_1}\ldots\mathbb{X}^{M1}_{r_M}\right)\langle r_1\ldots r_M|$$

$$|\mathbb{X}^{\mathrm{L}}\rangle:=\sum_{l_1\ldots l_M=1}^{m}\mathrm{tr}\left(\mathbb{X}^{1L}_{l_1}\ldots\mathbb{X}^{ML}_{l_M}\right)|l_1\ldots l_M\rangle$$

$$\mathbb{X}^j:=\sum_{l_1,r_1,\ldots=1}^{m}\mathrm{tr}\left(\mathbb{X}^{1j}_{l_1r_1}\ldots\mathbb{X}^{Mj}_{l_Mr_M}\right)|l_1\ldots\rangle\langle r_1\ldots|,$$

where $m=n$ for two-dimensional classical systems and $m=\kappa^2$ for on-dimensional quantum systems. These MPSs and MPOs are associated with a chain of M m-dimensional systems and their virtual dimension amounts to $D=n$. Note that for two-dimensional classical systems the first and last matrices under the trace in the MPS and MPO reduce to vectors. The partition function (and similarly other correlation functions) reads $Z=\langle\mathbb{X}^1|\mathbb{X}^2\cdots\mathbb{X}^{L-1}|\mathbb{X}^L\rangle$. Evaluating this expression iteratively by calculating step by step $\langle\mathbb{X}^j|:=\langle\mathbb{X}^{j-1}|\mathbb{X}^j$ for $j=2,\ldots,L-1$ fails because the virtual dimension of the MPS $\langle\mathbb{X}^j|$ increases exponentially with j. A way to circumvent this problem is to replace in each iterative step the MPS $\langle\mathbb{X}^j|$ by a MPS $\langle\tilde{\mathbb{X}}^j|$ with a reduced virtual dimension $\tilde{D}$ that approximates the state $\langle\mathbb{X}^j|$ best in the sense that the norm $\delta K:=\|\langle\mathbb{X}^j|-\langle\tilde{\mathbb{X}}^j|\|$ is minimized. Owing to the fact that this cost function is multiquadratic in the variables of the MPS, this minimization can be carried out very efficiently; the exponential increase of the virtual dimension can hence be prevented and the iterative evaluation of Z becomes tractable, such that an approximation to the partition function can be obtained from $Z\simeq\langle\tilde{\mathbb{X}}^{L-1}|\mathbb{X}^L\rangle$. The accuracy of this approximation depends only on the choice of the reduced dimension $\tilde{D}$ and the approximation becomes exact for $\tilde{D}\geq D^{\mathrm{L}}$. As the norm δK can be calculated at each step, $\tilde{D}$ can be increased dynamically if the obtained accuracy is not large enough. In the worst-case scenario, such as in the NP-complete Ising spin glasses [114], $\tilde{D}$ will probably have to grow exponentially in L for a fixed precision of the partition function. However, in less pathological cases, it seems that $\tilde{D}$ only has to grow polynomially in L; indeed, the success of the methods developed by Nishino [110] in the translational-invariant case indicate that even a constant $\tilde{D}$ will produce very reliable results.

We illustrate this with an example of bosons in optical lattices. A system of trapped bosonic particles in a one-dimensional optical lattice of L sites is described by the Bose–Hubbard Hamiltonian [115]

$$H=-J\sum_{i=1}^{L-1}(a_i^\dagger a_{i+1}+h.c.)+\frac{U}{2}\sum_{i=1}^{\mathrm{L}}\hat{n}_i(\hat{n}_i-1)+\sum_{i=1}^{\mathrm{L}}V_i\hat{n}_i,$$

where $a_i^\dagger$ and a_i are the creation and annihilation operators on site i and $\hat{n}_i=a_i^\dagger a_i$ is the number operator. This Hamiltonian describes the interplay between the kinetic energy due to the next-neighbour hopping with amplitude J and the repulsive on-site interaction U of the particles. The last term in the Hamiltonian models the harmonic confinement of magnitude $V_i=V_0(i-i_0)^2$. The variation of the ratio U/J drives a phase-transition between the Mott-insulating and the superfluid phase, characterized by localized and delocalized particles, respectively [116]. Experimentally, the variation of U/J can be realized by tuning the depth of the optical lattice [115,117]. On the other hand, one typically measures the momentum distribution directly by letting the atomic gas expand and then measuring the density distribution. Thus, here we are mainly interested in the (quasi-)momentum distribution

$$n_k=\frac{1}{L}\sum_{r,s=1}^{L}\langle a_r^\dagger a_s\rangle e^{i2\pi k(r-s)/L}.$$

Our goal is now to use our numerical method to study the finite-temperature properties of this system for different ratios U/J. We thereby assume that the system is in a thermal state corresponding to a grand canonical ensemble with chemical potential μ, such that the partition function is obtained as $Z = \operatorname{tr} e^{-\beta(H-\mu\hat{N})}$. Here, $\hat{N} = \sum_{i=1}^{\mathrm{L}} \hat{n}_i$ represents the total number of particles. For the numerical study, we assume a maximal particle number q per lattice site, such that we can project the Hamiltonian H on the subspace spanned by Fock states with particle numbers per site ranging from 0 to q. The projected Hamiltonian $\tilde{H}$ then describes a chain of L spins, with each spin acting on a Hilbert-space of dimension $n = q + 1$. A Trotter decomposition that turned out to be advantageous for this case is

$$e^{-\beta(\tilde{H}-\mu\hat{N})} = (\hat{V}^{\dagger}\hat{V})^{M} + \mathrm{O}\left[\frac{1}{M^2}\right], \tag{25}$$

with $\tilde{H} = H_R + H_S + H_T$, $H_R = -(J/2)\sum_{i=1}^{L-1} R^{(i)}R^{(i+1)}$, $H_S = -(J/2)\sum_{i=1}^{L-1} S^{(i)}S^{(i+1)}$, $H_T = \sum_{i=1}^{\mathrm{L}} T^{(i)}$, $R^{(i)} = \tilde{a}_i^{\dagger} + \tilde{a}_i$, $S^{(i)} = -i(\tilde{a}_i^{\dagger} - \tilde{a}_i)$, $T^{(i)} = (1/2)\tilde{n}_i(\tilde{n}_i - 1) + V_i\tilde{n}_i$ and $\hat{V} = e^{-(\beta/2M)H_R}e^{-(\beta/2M)H_S}e^{-(\beta/2M)(H_T-\mu\hat{N})}$. Thereby, $\tilde{a}_i^{\dagger}, \tilde{a}_i$ and $\tilde{n}_i$ denote the projections of the creation, the annihilation and the number operators $a_i^{\dagger}$, a_i and n_i on the q-particle subspace. The decomposition (24) of all two-particle operators in expression (25) then straightforwardly leads to a set of 4-index tensors X_{lrud}^{ij}, with indices l and r ranging from 1 to $(q+1)^3$ and indices u and d ranging from 1 to $q + 1$. Note that the typical second-order Trotter decomposition with $H = H_{\mathrm{even}} + H_{\mathrm{odd}}$ would make the indices l and r range from 1 to $(q+1)^6$.

Let us start out by considering the limit $U/J \to \infty$ in which double occupation of single lattice sites is prevented and the particles in the lattice form a Tonks–Girardeau gas [118]. In this limit, the Bose–Hubbard Hamiltonian maps to the Hamiltonian of the exactly solvable (inhomogeneous) XX model, which allows us to benchmark our algorithm. The comparison of our numerical results with the exact results can be garnered from Figure 15. Here, the density and the (quasi-)momentum distribution are considered for the special case $\beta J = 1$, $L = 40$, $N = 21$ and $V_0/J = 0.034$. The numerical results shown have been obtained for a Trotter number of $M = 10$ and two different reduced virtual dimensions $\tilde{D} = 2$ and $\tilde{D} = 8$. The norm δK was of the order of 10^{-4} for $\tilde{D} = 2$ and 10^{-6} for $\tilde{D} = 8$. (We note that we have stopped our iterative algorithm at the point where the variation of δK was less than 10^{-8}.) From the insets, it can be garnered that the error of the numerical calculations is already very small for $\tilde{D} = 2$ (of the order of 10^{-3}) and decreases significantly for $\tilde{D} = 8$. This error can be decreased further by increasing the Trotter number M.

As the ratio U/J becomes finite, the system becomes physically more interesting, but lacks an exact mathematical solution. In order to judge the reliability of our numerical solutions in this case, we check the convergence with respect to the free parameters of our algorithm (q, $\tilde{D}$ and M). As an illustration, the convergence with respect to the parameter q is shown in Figure 16. In this figure, the density and the (quasi-)momentum distribution are plotted for $q = 2, 3$ and 4. We thereby assume that $\beta J = 1$, $L = 40$ and $N = 21$ and consider interaction strengths $U/J = 4$ and 8. The harmonic potential V_0 is chosen in such a way so as to describe Rb atoms in a harmonic trap of a frequency in Hertz (along the lines of [118]). We note that we have taken into account that changes of the ratio U/J are obtained from changes in both the on-site interaction U and the hopping amplitude J due to variations in the depth of the optical lattice. The numerical calculations have been performed with $M = 10$ and $\tilde{D} = q + 1$. From the figure it can be garnered that convergence with respect to q is achieved for $q \geq 3$.

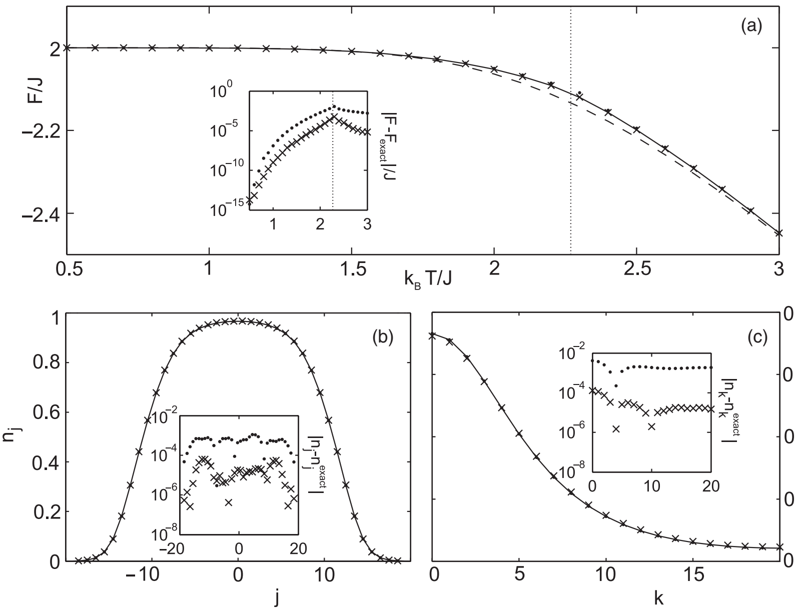

Figure 15. Density and (quasi-)momentum distribution in the Tonks–Girardeau gas limit, plotted for $\beta J = 1$, $L = 40$, $N = 21$ and $V_0/J = 0.034$. The dots (crosses) represent the numerical results for $\tilde{D} = 2$ ($\tilde{D} = 8$) and the solid line illustrates the exact results. The error of the numerical results can be garnered from the insets. (Reprinted with permission from [14] V. Murg *et al.*, Phys. Rev. Lett. 95 (2005), p. 057206. © 2005 by the American Physical Society.)

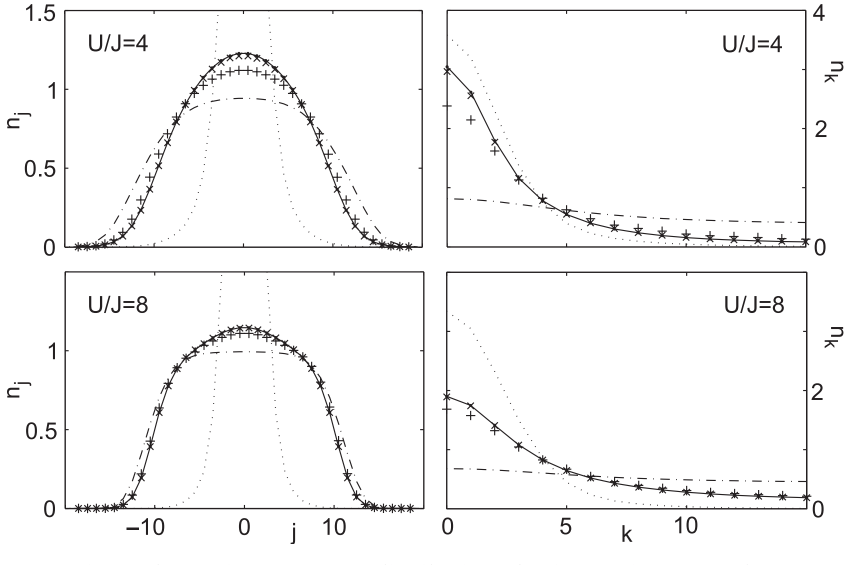

Figure 16. Density and (quasi-)momentum distributions for interaction strengths $U/J = 4$ and 8. Here, $\beta J = 1$, $L = 40$, $N = 21$ and $M = 10$. Numerical results were obtained for $q = 2$ (plus signs), $q = 3$ (crosses) and $q = 4$ (solid line). For comparison, the distributions for $U/J = 0$ (dotted lines) and $U/J \to \infty$ (dash-dotted lines) are also included. (Reprinted with permission from [14] V. Murg *et al.*, Phys. Rev. Lett. 95 (2005), p. 057206. © 2005 by the American Physical Society.)

6.4. *Density of states*

Information about the density of states can be obtained by studying the quantity

$$f(t) = \mathrm{Tr}\left(e^{-iHt}\right)$$

as a function of time. Indeed, assume that we know the function $f(t)$ exactly from $t = -\infty \to +\infty$. Its Fourier transform

$$F(\omega) = \int_{t=-\infty}^{\infty} e^{i\omega t}\mathrm{Tr}\, e^{-iHt} = \sum_{k=0}^{\infty} \delta(\omega - E_k)$$

with $\{E_k\}$ the spectrum of the Hamiltonian H. Hence, the Fourier transform of $f(t)$ gives all of the information about the spectrum. Of course, in practice we will not be able to determine $f(t)$ for all times, and we will only be able to approximate it within some time window. The effect on its Fourier transform is that it becomes convolved with a sinc-function of some width inversely proportional to the time window. (We can also make use of more advanced signal processing tricks to obtain a better conditioning. We refer the interested reader to [119] for more details.) This means that the resolution of such simulations will depend on the time frames in which we can calculate $f(t)$. Furthermore, only a discrete number of points will be available, such that we have to perform a discrete Fourier transform instead, but this is still sufficient to obtain the spectral information we are interested in.

As first discussed by Osborne, the calculation of $f(t)$ can efficiently be implemented with the techniques discussed before [119]. Basically, it is the real-time equivalent of calculating $e^{-\beta H}$ and, as has already been discussed, there are several options for doing this. First of all, we can evolve a collection of maximally entangled states with $I \otimes e^{-iHt}$ using small Trotter steps, and then calculate its overlap with the original maximally entangled states (note again that several options exist for how to choose the Trotter decomposition). Indeed, it happens that

$$\langle I | A \otimes I | I \rangle = \mathrm{Tr}\, A$$

with $|I\rangle = \sum_k |k\rangle|k\rangle$ a maximally entangled state.

Alternatively, we can consider the Suzuki–Trotter decomposition for an infinitesimal step, write down the complete tensor network corresponding to $f(t)$, and contract it using the techniques discussed in the previous Section 5.3. Note that in this particular case, there is again the possibility of contracting the corresponding tensor network in two directions, namely in the time or space direction. Experience tell us which way is more reliable.

In any case, it is really exciting to see how many new tools and possibilities can be explored by reformulating it in the MPS language.

7. PEPSs and ground states of two-dimensional quantum spin systems

In this Section, we present a natural generalization of the one-dimensional MPS to two and higher dimensions and build simulation techniques based on those states which effectively extend DMRG to higher dimensions. We call those states PEPSs [12,13], since they can be understood in terms of pairs of maximally entangled states of some auxiliary systems that are locally projected in some low-dimensional subspaces. This class of states includes the generalizations of the two-dimensional AKLT states known as tensor product states [120–126] which have been used for two-dimensional problems but is much broader since every state can be represented as a PEPS (as long as the dimension of the entangled pairs is large enough). We also develop an efficient algorithm to calculate correlation functions of these PEPSs, and which allows us to extend the one-dimensional algorithms to higher dimensions. This leads to many interesting applications, such as scalable variational methods for finding ground or thermal states of spin systems in higher dimensions as well as simulating their time evolution. For the sake of simplicity, we restrict our attention to a square lattice in two dimensions. The generalization to higher dimensions and other geometries is straightforward.

We want to emphasize, however, that the PEPS method is not yet as well established as the one-dimensional MPS or DMRG methods; this is mainly due to its higher complexity, but also to a large extent due to the fact that these PEPS methods are relatively unexplored which means that there is a lot of room for improvement and exciting research.

7.1. *Construction and calculus of PEPSs*

There have been various attempts at using the ideas developed in the context of the NRG and DMRG to simulate two-dimensional quantum spin systems. However, in hindsight it is clear why those methods were never very successful: they can be reformulated as variational methods within the class of one-dimensional MPSs, and the structure of those MPS is certainly not well suited to describing the ground states of two-dimensional quantum spin systems. This can immediately be understood when reconsidering the area law discussed in Section 2 (see Figure 17): if we look at the number of degrees of freedom needed to describe the relevant modes in a block of spins, this has to scale as the boundary of the block, and hence this increases exponentially with the size of that boundary. This means that it is impossible to use a NRG or DMRG approach, where the number of degrees of freedom is bounded to D. (Clearly, the VMPS/DMRG methods can reveal very valuable information in the case of quasi-two-dimensional systems such as ladders with a few rungs; see e.g. [127] for a nice illustration.)

However, it is straightforward to generalize the MPS picture to higher dimensions: the main reason of the success of the MPS approach is that it allows a good representation of local properties that are compatible with, for example, the translational symmetry in the system. These strong local correlations are obtained by sharing maximally entangled states between neighbours, and the longer range correlations are basically mediated by the intermediate particles. This is, of course, a very physical picture, as the Hamiltonian does not force any long-range correlations to exist *a priori*, and those only come into existence because of frustration effects. This generalization to higher dimensions can therefore be obtained by distributing virtual maximally entangled states between all neighbouring sites [126], and as such a generalization of the AKLT picture is obtained.

More specifically, each physical system at site i is represented by four auxiliary systems a_i, b_i, c_i, and d_i of dimension D (except at the borders of the lattices). Each of those systems is in a maximally entangled state

$$|I\rangle = \sum_{i=1}^{D} |ii\rangle$$

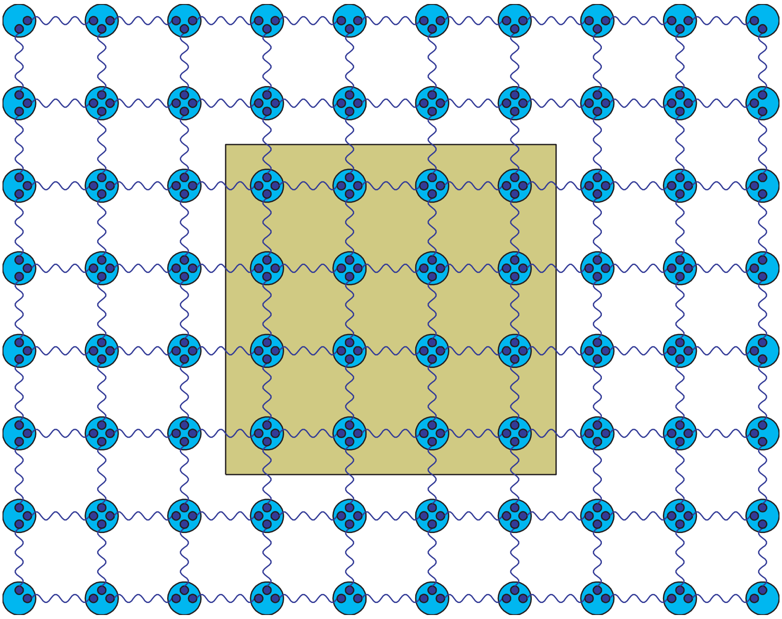

Figure 17. Representation of a quantum spin system in two dimensions using the PEPS representation. If we calculate the entropy of a block of spins, then this quantity will obey an area law and scale as the length of the boundary between the block and the rest. PEPS states are constructed such as to have this property built in.

with one of its neighbours, as shown in Figure 17. The PEPS $|\Psi\rangle$ is then obtained by applying to each site one operator Q_i that maps the four auxiliary systems onto one physical system of dimension d. This leads to a state with coefficients that are contractions of tensors according to a certain scheme. Each of the tensors is related to one operator Q_i according to

$$[A_i]^k_{lrud} = \langle k|Q_i|l, r, u, d\rangle$$

and thus associated with one lattice site i. All tensors possess one physical index k of dimension d and four virtual indices l, r, u and d of dimension D. The scheme according to which these tensors are contracted mimics the underlying lattice structure: the four virtual indices of the tensors are related to the left, right, upper and lower bond emanating from the corresponding lattice site. The coefficients of the PEPS are then formed by joining the tensors in such a way that all virtual indices related to same bonds are contracted. This is illustrated in Figure 18 for the special case of a 4 × 4 square lattice. Assuming this contraction of tensors is performed by the function $\mathcal{F}(\cdot)$, the resulting PEPS can be written as

$$|\Psi\rangle = \sum_{k_1,\ldots,k_M=1}^{d} \mathcal{F}\big([A_1]^{k_1},\ldots,[A_M]^{k_M}\big)|k_1,\ldots,k_M\rangle.$$

This construction can be generalized to any lattice shape and dimension and one can show that any state can be written as a PEPS if we allow the bond dimension to become

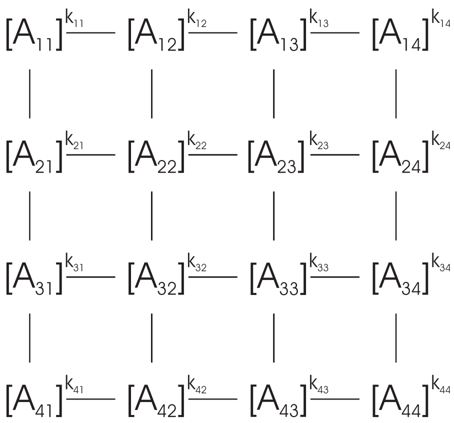
[A11]^k11 [A12]^k12 [A13]^k13 [A14]^k14
[A21]^k21 [A22]^k22 [A23]^k23 [A24]^k24
[A31]^k31 [A32]^k32 [A33]^k33 [A34]^k34
[A41]^k41 [A42]^k42 [A43]^k43 [A44]^k44

Figure 18. Structure of the coefficient related to the state $|k_{11},\dots,k_{44}\rangle$ in the PEPS $|\Psi_A\rangle$. The bonds represent the indices of the tensors $[A_i]^k$ that are contracted.

very large. In this way, we also resolve the problem of the entropy of blocks mentioned above, since now this entropy is proportional to the bonds that connect such a block with the rest, and therefore to the area of the block. Note also that, in analogy to the MPS [58], the PEPSs are guaranteed to be ground states of local Hamiltonians.

There has recently been a lot of progress in justifying this PEPS picture; Hastings has shown [46] that indeed every ground state of a local quantum spin Hamiltonian has an efficient representation in terms of a PEPS, that is, one whose bond dimension D scales subexponentially with the number of spins under interest. Also, he has shown that all thermal states have an efficient representation in terms of MPOs. This is great news, as it basically shows that we have identified the relevant manifold describing the low-energy physics of quantum spin systems. This can lead to many applications in theoretical condensed matter physics, as questions regarding the possibility of some exotic phase of matter can now be answered by looking at the set of PEPSs, hence skipping the bottleneck of simulation of ground states.

The family of PEPSs also seems to be very relevant in the field of quantum information theory. For example, all quantum error-correcting codes such as Kitaev's toric code [128] exhibiting topological quantum order have a very simple and exact description in terms of PEPSs [42]. Furthermore, the PEPS picture has been used to show the equivalence between different models of quantum computation [126]; more specifically, the so-called cluster states [129] have a simple interpretation in terms of PEPSs, and this picture demystifies the inner workings of the one-way quantum computer.

7.2. *Calculus of PEPS*

We now show how to determine expectation values of operators in the state $|\Psi\rangle$. We consider a general operator $O=\prod_i O_i$ and define the $D^2\times D^2\times D^2\times D^2$–tensors

$$\left[E_j^{O_j}\right]_{(ll')(rr')}^{(uu')(dd')}=\sum_{k,k'=1}^{d}\langle k|O_j|k'\rangle\left[A_j^*\right]_{lrud}^{k'}\left[A_j\right]_{l'r'u'd'}^{k}.$$

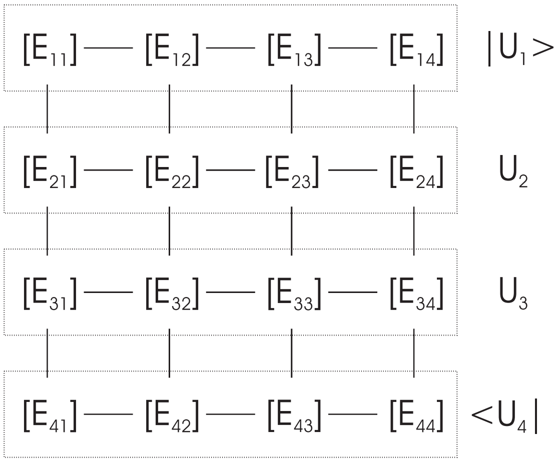
[E11] — [E12] — [E13] — [E14]
|U1>
[E21] — [E22] — [E23] — [E24]
U2
[E31] — [E32] — [E33] — [E34]
U3
[E41] — [E42] — [E43] — [E44]
<U4|

Figure 19. Structure of the contractions in $\langle\Psi_A|\Psi_A\rangle$. In this scheme, the first and last rows can be interpreted as MPSs $|U_1\rangle$ and $\langle U_4|$ and the rows in between as MPOs U_2 and U_3. The contraction of all tensors is then equal to $\langle U_4|U_3U_2|U_1\rangle$.

In this definition, the symbols (ll'), (rr'), (uu') and (dd') indicate composite indices. We may interpret the four indices of this tensor as being related to the four bonds emanating from site j in the lattice. Then, $\langle\Psi|O|\Psi\rangle$ is formed by joining all tensors $E_j^{O_j}$ in such a way that all indices related to same bonds are contracted, as in the case of the coefficients of PEPSs. These contractions have a rectangular structure, as depicted in Figure 19. In terms of the function $\mathcal{F}(\cdot)$, the expectation value reads

$$\langle\Psi|O|\Psi\rangle = \mathcal{F}\left(E_1^{O_1},\ldots,E_N^{O_N}\right).$$

The contraction of all tensors $E_j^{O_j}$ according to this scheme requires a number of steps that scales exponentially with N, and makes calculations intractable as the system grows larger. Owing to this, an approximate method has to be used to calculate expectation values.

The approximate method suggested in [12] is based on MPSs and MPOs. The main idea is to interpret the first and last row in the contraction scheme as MPSs and the rows in between as MPOs. The horizontal indices thereby form the virtual indices and the vertical indices are the physical indices. Thus, the MPSs and MPOs have both virtual dimension and physical dimension equal to D^2. Explicitly written, the MPSs read as

$$|U_1\rangle = \sum_{\tilde{d}_1,\ldots,\tilde{d}_L=1}^{D^2} \mathrm{tr}\left(\left[E_{11}^{O_{11}}\right]^{1\tilde{d}_1}\cdots\left[E_{1L}^{O_{1L}}\right]^{1\tilde{d}_L}\right)|\tilde{d}_1,\ldots,\tilde{d}_L\rangle$$

$$\langle U_L| = \sum_{\tilde{u}_1,\ldots,\tilde{u}_L=1}^{D^2} \mathrm{tr}\left(\left[E_{L1}^{O_{L1}}\right]^{\tilde{u}_1 1}\cdots\left[E_{LL}^{O_{LL}}\right]^{\tilde{u}_L 1}\right)\langle\tilde{u}_1,\ldots,\tilde{u}_L|$$

and the MPO at row r is

$$U_r = \sum_{\substack{\tilde{u}_1,\ldots,\tilde{u}_L=1\\ \tilde{d}_1,\ldots,\tilde{d}_L=1}}^{D^2} \mathrm{tr}\left(\left[E_{r1}^{O_{r1}}\right]^{\tilde{u}_1\tilde{d}_1}\cdots\left[E_{rL}^{O_{rL}}\right]^{\tilde{u}_L\tilde{d}_L}\right)|\tilde{u}_1,\ldots,\tilde{u}_L\rangle\langle\tilde{d}_1,\ldots,\tilde{d}_L|.$$

In terms of these MPSs and MPOs, the expectation value is a product of MPOs and MPSs:

$$\langle\Psi|O|\Psi\rangle = \langle U_L|U_{L-1}\cdots U_2|U_1\rangle.$$

The evaluation of this expression is, of course, intractable. With each multiplication of a MPO with a MPS, the virtual dimension increases by a factor of D^2. Thus, after L multiplications, the virtual dimension is D^{2L}, which is exponential in the number of rows. The expression, however, reminds of the time–evolution of a MPS. There, each multiplication with a MPO corresponds to one evolution step. The problem of the exponential increase in the virtual dimension is circumvented by restricting the evolution to the subspace of MPSs with a certain virtual dimension $\tilde{D}$. This means that after each evolution step the resulting MPS is approximated by the 'nearest' MPS with virtual dimension $\tilde{D}$. This approximation can be done efficiently, as shown in [11]. In this way, $\langle\Psi|O|\Psi\rangle$ can also be calculated efficiently: first, the MPS $|U_2\rangle$ is formed by multiplying the MPS $|U_1\rangle$ with MPO U_2. The MPS $|U_2\rangle$ is then approximated by $|\tilde{U}_2\rangle$ with virtual dimension $\tilde{D}$. In this fashion the procedure is continued until $|\tilde{U}_{L-1}\rangle$ is obtained. The expectation value $\langle\Psi|O|\Psi\rangle$ is then simply

$$\langle\Psi|O|\Psi\rangle = \langle U_L|\tilde{U}_{L-1}\rangle.$$

Interestingly enough, this method to calculate expectation values can be adopted to develop very efficient algorithms to determine the ground states of two-dimensional Hamiltonians and the time evolution of PEPSs by extending DMRG and the time evolution schemes to two dimensions.

7.3. *Variational method with PEPSs*

Let us start with an algorithm to determine the ground state of a Hamiltonian with short-range interactions on a square $L \times L$ lattice. The goal is then to determine the PEPS $|\Psi\rangle$ with a given dimension D which minimizes the energy:

$$\langle H\rangle = \frac{\langle\Psi|H|\Psi\rangle}{\langle\Psi|\Psi\rangle}. \tag{26}$$

Following [10], the idea is to iteratively optimize the tensors A_i one by one while fixing all of the other tensors until convergence is reached. The crucial observation is the fact that the exact energy of $|\Psi\rangle$ (and also its normalization) is a quadratic function of the components of the tensor A_i associated with *one* lattice site i. Owing to this, the optimal parameters A_i can simply be found by solving a generalized eigenvalue problem.

The challenge that remains is to calculate the matrix pair for which the generalized eigenvalues and eigenvectors shall be obtained. In principle, this is done by contracting all indices in the expressions $\langle\Psi|H|\Psi\rangle$ and $\langle\Psi|\Psi\rangle$ except those connecting to A_i. By interpreting the tensor A_i as a dD^4-dimensional vector $\boldsymbol{A}_i$, these expressions can be written as

$$\langle\Psi|H|\Psi\rangle = \boldsymbol{A}_i^\dagger \mathcal{H}_i \boldsymbol{A}_i, \tag{27}$$

$$\langle\Psi|\Psi\rangle = \boldsymbol{A}_i^\dagger \mathcal{N}_i \boldsymbol{A}_i. \tag{28}$$

Thus, the minimum of the energy is obtained by the generalized eigenvector $\boldsymbol{A}_i$ of the matrix pair $(\mathcal{H}_i, \mathcal{N}_i)$ to the minimal eigenvalue μ:

$$\mathcal{H}_i \boldsymbol{A}_i = \mu \mathcal{N}_i \boldsymbol{A}_i.$$

It turns out that the matrix pair $(\mathcal{H}_i, \mathcal{N}_i)$ can be efficiently evaluated by the method developed for the calculation of expectation values: $\mathcal{N}_i$ relies on the contraction of all but one tensors E_j^I (with I denoting the identity) according to the same rectangular scheme as before. The one tensor that has to be omitted is E_i^I, the tensor related to site i. Assuming this contraction is performed by the function $\mathcal{G}_i(\cdot)$, $\mathcal{N}_i$ can be written as

$$\left[\mathcal{N}_i\right]_{lrud\ k'}^{k\ \ \ \ l'r'u'd'} = \mathcal{G}_i\left(E_1^I, \ldots, E_N^I\right)_{lrud}^{l'r'u'd'} \delta_{k'}^{k}.$$

If we join the indices $(klrud)$ and $(k'l'r'u'd')$, we obtain the $dD^4 \times dD^4$ matrix that fulfills equation (28). To evaluate $\mathcal{G}_i(\cdot)$ efficiently, we proceed in the same way as before by interpreting the rows in the contraction structure as MPSs and MPOs. First, we join all rows that lie above site i by multiplying the topmost MPS $|U_1\rangle$ with subjacent MPO and reducing the dimension after each multiplication to $\tilde{D}$. Then, we join all rows lying below i by multiplying $\langle U_L|$ with the adjacent MPOs and reducing the dimension as well. We end up with two MPSs of virtual dimension $\tilde{D}$, which we can contract efficiently with all but one of the tensors E_j^I lying in the row of site i.

The effective Hamiltonian $\mathcal{H}_i$ can be determined in an analogous way, but here the procedure has to be repeated for every term in the Hamiltonian (i.e. of the order of $2N$ times in the case of nearest-neighbour interactions). Assuming a single term in the Hamiltonian has the tensor-product structure $H^s \equiv \prod_i h_i^s$, the effective Hamiltonian $\mathcal{H}_i^s$ corresponding to this term is obtained as

$$\left[\mathcal{H}_i^s\right]_{lrud\ k'}^{k\ \ \ \ l'r'u'd'} = \mathcal{G}_i\left(E_1^{h_1^s}, \ldots, E_N^{h_N^s}\right)_{lrud}^{l'r'u'd'} \left[h_i^s\right]_{k'}^{k}.$$

The complete effective Hamiltonian $\mathcal{H}_i$ that fulfills equation (27) is then produced as

$$\mathcal{H}_i = \sum_s \mathcal{H}_i^s.$$

Thus, both of the matrices $\mathcal{N}_i$ and $\mathcal{H}_i$ are directly related to the expressions $\mathcal{G}_i\left(E_1^I, \ldots, E_N^I\right)$ and $\mathcal{G}_i\left(E_1^{h_1^s}, \ldots, E_N^{h_N^s}\right)$. These expressions, however, can be evaluated efficiently using the approximate method introduced before for the calculation of expectation values. Therefore, the optimal A_i can be determined, and one can proceed with the following site, iterating the procedure until convergence.

7.4. *Time evolution with PEPS*

Let us next move to describe how a time evolution can be simulated on a PEPS. We assume that the Hamiltonian only couples nearest neighbours, although more general settings can be considered. The principle of simulating a time-evolution step is as follows: first, a PEPS $|\Psi_A^0\rangle$ with physical dimension $d=2$ and virtual dimension D is chosen as a starting state. This state is evolved by the time-evolution operator $U = e^{iH\delta t}$ (we assume $\hbar = 1$) to yield another PEPS $|\Psi_B\rangle$ with a virtual dimension D_B increased by a factor η:

$$|\Psi_B\rangle = U|\Psi_A^0\rangle.$$

The virtual dimension of this state is then reduced to D by calculating a new PEPS $|\Psi_A\rangle$ with virtual dimension D that has minimal distance to $|\Psi_B\rangle$. This new PEPS is the starting state for the next time-evolution step. The crucial point in simulating a time evolution with PEPS is thus the development of an efficient algorithm for reducing the virtual dimension of a PEPS.

Before formulating this algorithm, let us recite how to express the product $U|\Psi_A^0\rangle$ in terms of a PEPS. This is done by means of a Trotter approximation: first, the interaction terms in H are classified as *horizontal* and *vertical* according to their orientation and *even* and *odd* depending on whether the interaction is between even–odd or odd–even rows (or columns). The Hamiltonian can then be decomposed into a *horizontal–even*, a *horizontal–odd*, a *vertical–even* and a *vertical–odd* part:

$$H = H_{he} + H_{ho} + H_{ve} + H_{vo}.$$

The single-particle operators of the Hamiltonian can simply be incorporated into one of the four parts (note that different Trotter decompositions are again possible, for example, grouping all Pauli operators of the same kind into three different groups as we discussed earlier, and in some cases this leads to a clear computational advantage). Using the Trotter approximation, the time-evolution operator U can be written as a product of four evolution operators:

$$U = e^{-iH\delta t} \approx e^{-iH_{he}\delta t} e^{-iH_{ho}\delta t} e^{-iH_{ve}\delta t} e^{-iH_{vo}\delta t}. \tag{29}$$

Since each of the four parts of the Hamiltonian consist of a sum of commuting terms, each evolution operator is equal to a product of two-particle operators w_{ij} acting on neighbouring sites i and j. These two-particle operators have a Schmidt decomposition consisting of, say, η terms:

$$w_{ij} = \sum_{\rho=1}^{\eta} u_i^\rho \otimes v_j^\rho.$$

One such two-particle operator w_{ij} applied to the PEPS $|\Psi_A^0\rangle$ modifies the tensors A_i^0 and A_j^0 associated with sites i and j as follows: assuming that the sites i and j are horizontal neighbours, A_i^0 has to be replaced by

$$[B_i]^k_{l(r\rho)ud} = \sum_{k'=1}^{d} [u_i^\rho]^k_{k'} [A_i^0]^{k'}_{lrud}$$

and A_j^0 becomes

$$[B_j]^k_{(l\rho)rud} = \sum_{k'=1}^{d} [v_j^\rho]^k_{k'} [A_j^0]^{k'}_{lrud}.$$

These new tensors have a joint index related to the bond between sites i and j. This joint index is composed of the original index of dimension D and the index ρ of dimension η that enumerates the terms in the Schmidt decomposition. Thus, the effect of the two-particle operator w_{ij} is to increase the virtual dimension of the bond between sites i and j by a factor of η. Consequently, $e^{-iH_{he}\delta t}$ and $e^{-iH_{ho}\delta t}$ increase the dimension of every second horizontal bond by a factor of η; $e^{-iH_{ve}\delta t}$ and $e^{-iH_{vo}\delta t}$ do the same for every second vertical bond.

By applying all four evolution operators consecutively, we have found an approximate form of the time-evolution operator U that, when applied to a PEPS $|\Psi_A^0\rangle$, yields another PEPS $|\Psi_B\rangle$ with a virtual dimension multiplied by a constant factor η.

The aim of the approximate algorithm is now to optimize the tensors A_i related to a PEPS $|\Psi_A\rangle$ with virtual dimension D, such that the distance between $|\Psi_A\rangle$ and $|\Psi_B\rangle$ tends to a minimum. The function to be minimized is thus

$$K(A_1, \ldots, A_M) = \left\| |\Psi_A\rangle - |\Psi_B\rangle \right\|^2.$$

This function is non-convex with respect to all parameters $\{A_1, \ldots, A_M\}$. However, due to the special structure of PEPSs, it is quadratic in the parameters A_i associated with *one* lattice site i. Owing to this, the optimal parameters A_i can simply be found by solving a system of linear equations. The concept of the algorithm is to perform this one-site optimization site-by-site until convergence is reached.

The coefficient matrix and the inhomogeneity of the linear equations system can be calculated efficiently using the method developed for the calculation of expectation values. In principle, they are obtained by contracting all indices in the expressions for the scalar products $\langle\Psi_A|\Psi_A\rangle$ and $\langle\Psi_A|\Psi_B\rangle$ except those connecting to A_i. By interpreting the tensor A_i as a dD^4-dimensional vector $\boldsymbol{A}_i$, these scalar–products can be written as

$$\langle\Psi_A|\Psi_A\rangle = \boldsymbol{A}_i^\dagger \mathcal{N}_i \boldsymbol{A}_i, \tag{30}$$

$$\langle\Psi_A|\Psi_B\rangle = \boldsymbol{A}_i^\dagger \mathcal{W}_i. \tag{31}$$

Since

$$K = \langle\Psi_B|\Psi_B\rangle + \langle\Psi_A|\Psi_A\rangle - 2\mathrm{Re}\langle\Psi_A|\Psi_B\rangle,$$

the minimum is attained as

$$\mathcal{N}_i \boldsymbol{A}_i = \mathcal{W}_i.$$

The efficient calculation of $\mathcal{N}_i$ has already been described in the previous Section. The scalar product $\langle\Psi_A|\Psi_B\rangle$ and the inhomogeneity $\mathcal{W}_i$ are calculated in an efficient way following the same ideas. First, the $DD_B \times DD_B \times DD_B \times DD_B$ tensors

$$\left[F_j\right]_{(ll')(rr')}^{(uu')(dd')} = \sum_{k=1}^{d} \left[A_j^*\right]_{lrud}^k \left[B_j\right]_{l'r'u'd'}^k$$

are defined. The scalar product $\langle\Psi_A|\Psi_B\rangle$ is then obtained by contracting all tensors F_j according to the previous scheme, which is performed by the function $\mathcal{F}(\cdot)$:

$$\langle\Psi_A|\Psi_B\rangle = \mathcal{F}(F_1, \ldots, F_M)$$

The inhomogeneity $\mathcal{W}_i$ relies on the contraction of all but one of the tensors F_j, namely the function $\mathcal{G}_i(\cdot)$, in the sense that

$$\left[\mathcal{W}_i\right]_{lrud}^k = \sum_{l'r'u'd'=1}^{D} \mathcal{G}_i(F_1, \ldots, F_M)_{lrud}^{l'r'u'd'} \left[B_i\right]_{l'r'u'd'}^k.$$

Joining all indices ($klrud$) in the resulting tensor leads to the vector of length dD^4 that fulfills equation (31). Thus, both the scalar product $\langle\Psi_A|\Psi_B\rangle$ and the inhomogeneity $\mathcal{W}_i$ are directly related to the expressions $\mathcal{F}(F_1,\dots,F_M)$ and $\mathcal{G}_i(F_1,\dots,F_M)$. These expressions, however, can be evaluated efficiently using the approximate method from before.

Even though the principle of simulating a time-evolution step has now been recited, the implementation in this form is numerically expensive. The following notes describe how to make the simulation more efficient.

(1) *Partitioning of the evolution.* The number of required numerical operations decreases significantly as one time-evolution step is partitioned into four substeps: first the state $|\Psi_A^0\rangle$ is evolved by $e^{-iH_{vo}\delta t}$ only and the dimension of the increased bonds is reduced back to D. Next, evolutions according to $e^{-iH_{ve}\delta t}$, $e^{-iH_{ho}\delta t}$ and $e^{-iH_{he}\delta t}$ follow. Even though the partitioning increases the number of evolution steps by a factor of four, the number of multiplications in one evolution step decreases by a factor of η^3.

(2) *Optimization of the contraction order.* Most critical for the efficiency of the numerical simulation is the order in which the contractions are performed. We have optimized the order in such a way that the scaling of the number of multiplications with the virtual dimension D is minimal. For this, we assume that the dimension $\tilde{D}$ that tunes the accuracy of the approximate calculation of $\mathcal{N}_i$ and $\mathcal{W}_i$ is proportional to D^2, that is, $\tilde{D} = \kappa D^2$. The number of required multiplications is then of the order of $\kappa^2 D^{10} L^2$ and the required memory scales as $d\eta\kappa^2 D^8$. (The scaling D^{10} is obtained when a sparse matrix algorithm is used at all steps in the algorithm. In particular, we have to use an iterative sparse method for solving the linear set of equations in the approximation step.)

(3) *Optimization of the starting state.* The number of sweeps required to reach convergence depends on the choice of the starting state for the optimization. The idea for finding a good starting state is to reduce the bonds with increased virtual dimension ηD by means of a Schmidt decomposition. This is done as follows: assuming the bond is between the horizontal neighbouring sites i and j, the contraction of the tensors associated with these sites, B_i and B_j, along the bond i–j forms the tensor

$$[\mathcal{M}_{ij}]^{k\ \ k'}_{lud\ r'u'd'} = \sum_{\rho=1}^{D\eta} [B_i]^{k}_{l\rho ud}[B_j]^{k'}_{\rho r'u'd'}.$$

By joining the indices ($klud$) and ($k'r'u'd'$), this tensor can be interpreted as a $dD^3 \times dD^3$-matrix. The Schmidt–decomposition of this matrix is

$$\mathcal{M}_{ij} = \sum_{\rho=1}^{dD^3} c_\rho \mathcal{A}_i^\rho \otimes \mathcal{A}_j^\rho$$

with the Schmidt coefficients c_ρ ($c_\rho \geq 0$) and corresponding matrices $\mathcal{A}_i^\rho$ and $\mathcal{A}_j^\rho$. We can relate these matrices to a new pair of tensors A_i^0 and A_j^0 associated with sites i and j:

$$[A_i^0]^k_{l\rho ud} = \sqrt{c_\rho}[\mathcal{A}_i^\rho]^k_{lud},$$

$$[A_j^0]^k_{\rho rud} = \sqrt{c_\rho}[\mathcal{A}_j^\rho]^k_{rud}.$$

The virtual dimension of these new tensors related to the bond between sites i and j is equal to the number of terms in the Schmidt decomposition. Since these terms are weighted with the Schmidt coefficients c_ρ, it is justified to keep only the D terms with coefficients of largest magnitude. Then, the contraction of the tensors A_i^0 and A_j^0 along the bond i–j with dimension D yields a good approximation to the true value $\mathcal{M}_{ij}$:

$$[\mathcal{M}_{ij}]^{k\ \ \ k'}_{lud\ r'u'd'} \approx \sum_{\rho=1}^{D} [A_i^0]^k_{l\rho ud}[A_j^0]^{k'}_{\rho r'u'd'}.$$

When we apply this method to all bonds with increased dimension we obtain the starting state for the optimization.

7.4.1. *Examples*

Let us now illustrate the variational methods with some examples. Models to which the PEPS algorithms have already been applied to include the Heisenberg antiferromagnet [12], the Shastry–Sutherland model [130] and the system of hard-core bosons in a two-dimensional optical lattice [13]. In the following, we recite the results for the latter system, which include calculations of ground-state properties and studies of the time evolution after sudden changes in the parameters.

The system of bosons in a two-dimensional optical lattice is characterized by the Bose–Hubbard Hamiltonian

$$H = -J\sum_{\langle i,j\rangle}\left(a_i^\dagger a_j + \text{h.c.}\right) + \frac{U}{2}\sum_i \hat{n}_i(\hat{n}_i - 1) + \sum_i V_i\hat{n}_i,$$

where $a_i^\dagger$ and a_i are the creation and annihilation operators on site i and $\hat{n}_i = a_i^\dagger a_i$ is the number operator. This Hamiltonian describes the interplay between the kinetic energy due to the next-neighbour hopping with amplitude J and the repulsive on-site interaction U of the particles. The last term in the Hamiltonian models the harmonic confinement of magnitude $V_i = V_0(i - i_0)^2$. Since the total number of particles $\hat{N} = \sum_i \hat{n}_i$ is a symmetry of the Hamiltonian, the ground state will have a fixed number of particles. This number can be chosen by appending the term $-\mu\hat{N}$ to the Hamiltonian and tuning the chemical potential μ. In the limit of hard-core interaction, $U/J \to \infty$, two particles are prevented from occupying a single site. This limit is especially interesting in one dimension where the particles form the so-called Tonks–Girardeau gas [118,131]. The particles in this gas are strongly correlated, which leads to algebraically decaying correlation functions. In two dimensions, the model was studied in detail in [132]. In the hard-core limit, the Bose–Hubbard model is equivalent to a spin system with XX interactions described by the Hamiltonian

$$H = -\frac{J}{2}\sum_{\langle i,j\rangle}\left(\sigma_x^{(i)}\sigma_x^{(j)} + \sigma_y^{(i)}\sigma_y^{(j)}\right) + \frac{1}{2}\sum_i \left(V_i - \mu\right)\sigma_z^{(i)}.$$

Here, σ_x^i, σ_y^i and $\sigma_z^{(i)}$ denote the Pauli operators acting on site i. This Hamiltonian has a structure that can be simulated with the PEPS algorithm: it describes L^2 physical systems of dimension $d = 2$ on a $L \times L$ square lattice.

In Figure 20, the energy in the case of a 4×4 lattice is plotted as the system undergoes an imaginary time evolution. Thereby, a time step $\delta t = -i0.03$ is assumed and the

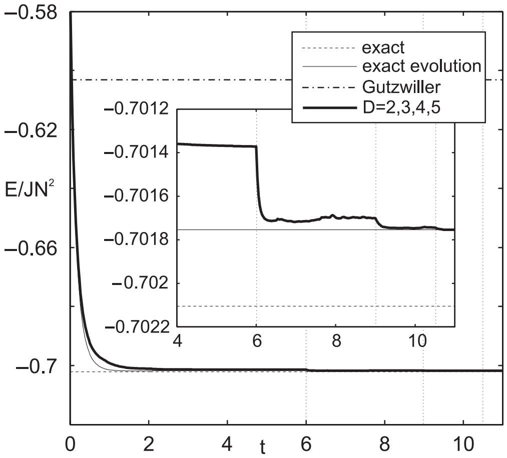

Figure 20. Energy as a function of time for the imaginary time evolution of the system of hard-core bosons on a 4×4 lattice. The evolutions are performed sequentially with PEPSs of virtual dimension $D = 2$, 3, 4 and 5. The times at which D is increased are indicated by vertical lines. For comparison, the exact ground-state energy, the exact imaginary time evolution and the energy of the optimal Gutzwiller ansatz are included. (Reprinted with permission from [13] V. Murg *et al.*, Phys. Rev. A 75 (2007), p. 033605. © 2007 by the American Physical Society.)

magnitude of the harmonic confinement (in units of the tunnelling constant) is chosen as $V_0/J = 36$. In addition, the chemical potential is tuned to $\mu/J = 3.4$ such that the ground state has particle number $\langle \hat{N} \rangle = 4$. With this configuration, the imaginary time evolution is performed both exactly and variationally with PEPSs. As a starting state a product state is used that represents a Mott-like distribution with four particles arranged in the centre of the trap and no particles elsewhere. The variational calculation is performed with $D = 2$ first until convergence is reached; then, evolutions with $D = 3$, 4 and 5 follow. At the end, a state is obtained that is very close to the state obtained by exact evolution. The difference in energy is $|E_{D=5} - E_{\text{exact}}| \simeq 6.4614 \times 10^{-5} J$. For comparison, the exact ground-state energy obtained by an eigenvalue calculation and the energy of the optimal Gutzwiller ansatz are also included in Figure 20. The difference between the exact result and the results of the imaginary time evolution is due to the Trotter error and is of the order of $O(\delta t^2)$. The energy of the optimal Gutzwiller ansatz is well separated from the exact ground-state energy and the results of the imaginary time evolution.

In Figure 21, the energy as a function of time is plotted for the imaginary time evolution on an 11×11 lattice. Again, a time step $\delta t = -i0.03$ is assumed for the evolution. The other parameters are set as follows: the ratio between harmonic confinement and the tunnelling constant is chosen as $V_0/J = 100$ and the chemical potential is tuned to $\mu/J = 3.8$ such that the total number of particles $\langle \hat{N} \rangle$ is 14. The starting state for the imaginary time evolution is, similar to before, a Mott-like distribution with 14 particles arranged in the centre of the trap. This state is evolved within the subset of PEPS with $D = 2$, 3, 4 and 5. As can be garnered from the plot, this evolution shows a definite convergence. In addition, the energy of the final PEPS lies well below the energy of the optimal Gutzwiller ansatz.

An application of the time-evolution algorithm with PEPSs is found in the study of dynamic properties of hard-core bosons on a lattice of size 11×11. Here, the responses of this system to sudden changes in the parameters are investigated and the numerical results

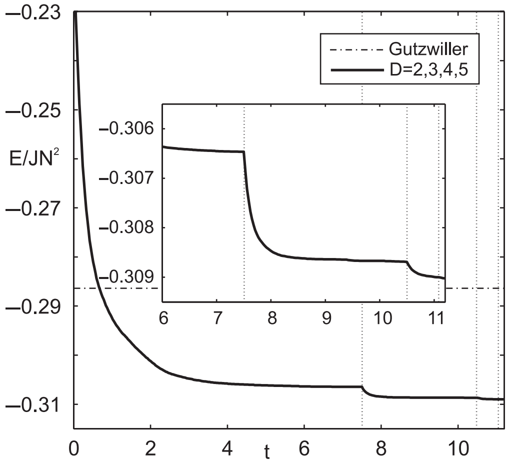

Figure 21. Energy as a function of time for the imaginary time evolution of the system of hard-core bosons on a 11×11 lattice. The evolutions are performed sequentially with PEPSs of virtual dimension $D = 2$, 3, 4 and 5. The times at which D is increased are indicated by vertical lines. For comparison, the energy of the optimal Gutzwiller ansatz is included. (Reprinted with permission from [13] V. Murg *et al.*, Phys. Rev. A 75 (2007), p. 033605. © 2007 by the American Physical Society.)

are compared with the results obtained by a Gutzwiller ansatz. An interesting property that is observed is the fraction of particles that are condensed. For interacting and finite systems, this property is measured best by the condensate density ρ which is defined as the largest eigenvalue of the correlation matrix $\langle a_i^\dagger a_j \rangle$.

In Figure 22, the evolution of a Mott distribution with 14 particles arranged in the centre of the trap is studied. It is assumed that $V_0/J = 100$, $\mu/J = 3.8$ and $\delta t = 0.03$. To ensure that the results are accurate, the following procedure was used for simulating the time evolution: first, the simulation has been performed using PEPSs with $D = 2$ and 3 until the overlap between these two states fell below a certain value. Then, the simulation has been continued using PEPS with $D = 3$ and 4 as long as the overlap between these two states was close to 1. The results of this calculation can be garnered from Figure 22. What can be observed is that there is a definite increase in the condensate fraction. The Gutzwiller ansatz is in contrast to this result since it predicts that the condensate density will remain constant. The inset in Figure 22 shows the overlap of the $D = 2$ with the $D = 3$ PEPSs and the $D = 3$ with the $D = 4$ PEPSs.

Finally, we make a few comments about the accuracy of the algorithm. One indicator for the accuracy is the distance between the time-evolved state and the state with reduced virtual dimension. For the time evolution of the Mott distribution that was discussed before, this quantity is plotted in Figure 23. We find that the distance is typically of the order of 10^{-3} for $D = 2$ and of the order of 10^{-4} for $D = 3$ and 4. Another quantity that is monitored is the total number of particles $\langle \hat{N} \rangle$. Since this quantity is supposed to be conserved during the whole evolution, its fluctuations indicate the reliability of the algorithm. From the inset in Figure 23, the fluctuations of the particle number in case of the time evolution of the Mott distribution can be garnered. We find that these fluctuations are at most of the order of 10^{-5}.

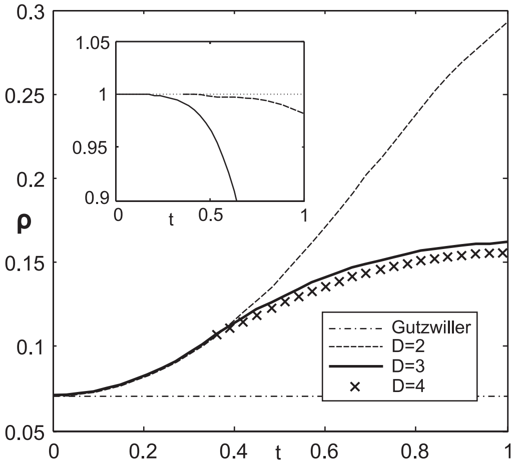

Figure 22. Time evolution of the condensate density starting from a Mott distribution with 14 particles arranged in the centre of the trap. The magnitude of the trapping potential is $V_0/J = 100$. For the evolution, the Gutzwiller ansatz and PEPSs with $D = 2$, 3 and 4 are used. The inset shows the overlap between the $D = 2$ and 3 PEPSs (solid line) and the $D = 3$ and 4 PEPSs (dashed line). (Reprinted with permission from [13] V. Murg *et al.*, Phys. Rev. A 75 (2007), p. 033605. © 2007 by the American Physical Society.)

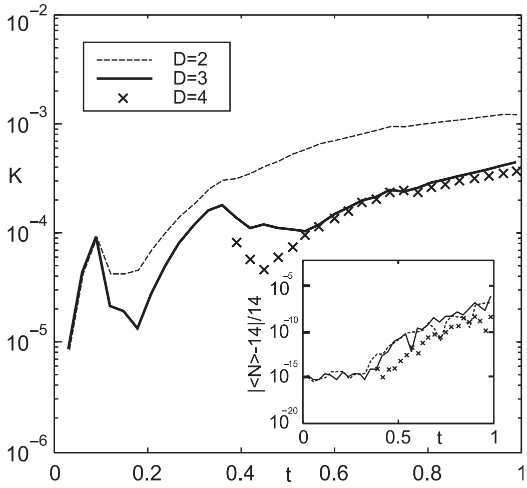

Figure 23. Distance K between the time-evolved state and the state with reduced virtual dimension. The virtual dimensions $D = 2$, 3 and 4 are included. The distance is plotted for the evolution of a Mott distribution with $N = 14$, as explained in Figure 22. From the inset, the deviation of the particle number from the value 14 can be garnered. (Reprinted with permission from [13] V. Murg *et al.*, Phys. Rev. A 75 (2007), p. 033605. © 2007 by the American Physical Society.)

7.4.2. *PEPSs and fermions*

The critical reader should by now have complained that we are only talking about spin systems and not about fermionic systems. Indeed, one of the long-term goals of the numerical approaches discussed here is to be able to simulate, for example, the

Fermi–Hubbard model in the relevant parameter regime, as the big advantage of MPS/PEPS methods is the non-existence of the so-called sign problem that is plaguing Monte Carlo methods.

The methods that we discussed in one dimension are perfectly applicable to fermionic systems, as the amazing Jordan–Wigner transformation allows us to map a local Hamiltonian of fermions to a local Hamiltonian of spins, and we know that the MPS techniques work provably well on the latter. The big problem, however, is that the Jordan–Wigner transformation only works in one dimension: if we use it on a two-dimensional lattice, a local Hamiltonian of fermions is mapped to a highly non-local Hamiltonian of spins. PEPSs, on the other hand, are devised to be such that they have extremal local properties; if the Hamiltonian contains a lot of strong non-local terms, we cannot expect a PEPS to exhibit the corresponding extremal long-range correlations. The natural question to ask is therefore whether there exists a generalization of the Jordan–Wigner transformation to higher dimensions. This was indeed shown to be possible in [133]: given any local Hamiltonian in terms of fermionic operators such as the Hubbard model in two or three dimensions, then there exists a local spin-3/2 Hamiltonian whose low-energy sector corresponds exactly to the original fermionic Hamiltonian. The conclusion is that the PEPS methods are equally applicable to quantum spin systems as to fermionic systems.

Another and more efficient approach is to make use of quantum numbers. In general, it is difficult to keep track of quantum numbers on a two-dimensional lattice. An exception, however, is given by the parity of the occupation number between two sites: by making the PEPS tensors block diagonal, one can invoke the even or odd occupation number between any sites and as such eliminate the fact that spurious effective long-range interactions arise in fermionic lattice systems [72]. This is very relevant as it leads to a huge speed-up of the algorithms for fermionic lattice systems.

7.5. *PEPSs on infinite lattices*

In parallel with the one-dimensional MPSs, we can also explicitly make use of the translational symmetry in the class of PEPSs to simulate low-energy properties of infinite lattices. A naive procedure is to impose complete translational symmetry and perform a line search (such as conjugate gradient) within the parameter space of the tensor; note that this can be done as expectation values of the corresponding PEPSs can be calculated using the appropriate one-dimensional infinite algorithms discussed above. However, the cost function is highly non-linear and the number of parameters grows very fast with D, such that those brute-force methods will typically lead to local minima. A smarter approach is to use imaginary time evolution, but keeping the translational invariance. We have already described how to do this in the one-dimensional case, where a particular kind of Trotter expansion led to a translational-invariant MPO, and we can repeat this in two dimensions leading to a translational-invariant PEPS operator. The only non-trivial part is the question of how to reduce the bond dimension after one time evolution step; but here we can again draw inspiration from the one-dimensional case, and ask the question which projectors can we put on the bonds such as to maximize the overlap of the *projected* state with the evolved one.

A more straightforward approach is to assume an $\ldots ABAB\ldots$ symmetry where each tensor A has all of its neighbouring tensors B and vice versa (of course, this is only possible

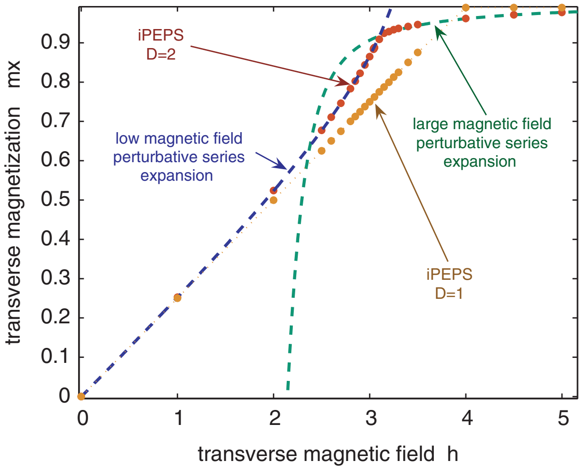

Figure 24. Magnetization $m_x(h)$ in the ground state $|\Psi_h\rangle$ of the two-dimensional quantum Ising model with transverse magnetic field. A trivial iPEPS (inner dimension $D = 1$) produces a magnetization that grows essentially linearly with the magnetic field until it saturates. Instead, the simplest non-trivial iPEPS (inner dimension $D = 2$) produces a magnetization curve that overlaps with the results of series expansions for both small and large magnetic fields. ($D = 3$ leads to results that could hardly be distinguished from those for $D = 2$ in this Figure.) Note that around $h \approx 3.1$ the derivative of the magnetization $m_x(h)$ changes suddenly. (Reprinted with permission from [134] with kind permission of the authors.)

for bipartite lattices; a different choice can be made for different types of lattice). Again, imaginary time evolution can be used to find the ground state, and this was first studied in [134] where the term iPEPS (infinite PEPS) was coined. The idea is as follows: take the even–odd–horizontal–vertical Trotter decomposition as discussed previously, and next evolve with just one operator acting on two nearest-neighbour sites. Effectively, this increases the bond dimension between those two sites (A and B). The *environment* of those spins can be readily calculated using the infinite one-dimensional translational-invariant methods discussed above, and then the variational problem becomes the problem of finding new A' and B' that approximate the state with higher bond optimally. Again, this can be done using the ALS method. Subsequently, we replace all tensors A and B with A' and B', and continue until convergence. The last step, replacing all tensors with the optimal local tensors, is only justified if the time step in the imaginary time evolution is very small, but in practice, this seems to work very well, as illustrated in Figure 24.

8. Conclusion

Recent progress in quantum information theory has provided scientists with new mathematical tools to describe and analyze many-body quantum systems. These new tools have given rise to novel descriptions of the quantum states that appear in nature, which are very efficient both in terms of the number of variables used to parameterize states and the number of operations to determine expectation values of typical observables. In a sense, they allow us to describe the 'corner of Hilbert space' where relevant states are located with an effort that only scales polynomially with the number of particles, as opposed to the exponential scaling resulting with other descriptions. These results have automatically led to a diverse set of new powerful algorithms to simulate quantum systems. Those algorithms allow us to describe ground states, thermal equilibrium, low excitations, dynamics, random systems, etc., of many-body quantum systems, and thus to attack new kinds of problems obtaining very precise results. Moreover, the methods work in one or more spatial dimensions. In the first case, the success of some of those methods is directly related to the extraordinary performance of DMRG. In higher dimensions, they also give rise to a better understanding of several many-body systems for which a description has not been possible with the existing techniques.

This paper has reviewed these new methods in a unified form. We have introduced MPSs and their extensions to higher dimensions, PEPSs, and shown how one can build powerful algorithms that find the best descriptions of states within that family of states. The algorithms are relatively simple to implement, although they require some tricks that have also been reported in this paper. We have also given simple Matlab codes in Appendix C to illustrate how one can program some of the methods. Thus, we believe that the present paper may be very useful both to scientists interested in implementing these new algorithms to describe any kind of many-body quantum system, as well as those interested in creating new algorithms for some specific purpose. We have also provided a large amount of appealing evidence of the fact that most interesting states in nature are well described by MPSs and PEPSs. This, in fact, indicates that our algorithms as well as future extensions can provide us with unique tools to explore the fascinating physics of quantum many-body systems.

Acknowledgements

We thank J. Eisert, M. Fannes, J. García-Ripoll, M. Hastings, A. Kitaev, J. Latorre, M. Martín-Delgado, B. Nachtergaele, G. Ortiz, R. Orus, T. Osborne, B. Paredes, D. Pérez-García, M. Plenio, M. Popp, D. Porras, J. Preskill, E. Rico, U. Schollwöck, N. Schuch, C. Schön, E. Solano, G. Vidal, J. von Delft, A. Weichselbaum, R. Werner, S. White, A. Winter, M. Wolf and M. Zwolak for valuable discussions that stimulated the developments of the ideas discussed in this paper.

References

[1] K.G. Wilson, Rev. Mod. Phys. 47 (1975), p. 773.
[2] S.R. White, Phys. Rev. Lett. 69 (1992), p. 2863.
[3] ——, Phys. Rev. B 48 (1992), p. 10345.
[4] U. Schollwöck, Rev. Mod. Phys. 77 (2005), p. 259.
[5] I. Peschel, X. Wang, M. Kaulke and K. Hallberg, *Density-Matrix Renormalization*, Springer, Berlin, 1998.
[6] M. Troyer and U. Wiese, Phys. Rev. Lett. 94 (2005), p. 170201.
[7] G. Vidal, Phys. Rev. Lett. 91 (2003), p. 147902.
[8] M. Zwolak and G. Vidal, Phys. Rev. Lett. 93 (2004), p. 207205.
[9] G. Vidal, Phys. Rev. Lett. 98 (2007), p. 070201.
[10] F. Verstraete, D. Porras and J.I. Cirac, Phys. Rev. Lett. 93 (2004), p. 227205.
[11] F. Verstraete, J.J. García-Ripoll and J.I. Cirac, Phys. Rev. Lett. 93 (2004), p. 207204.
[12] F. Verstraete and J.I. Cirac, Preprint cond-mat/0407066v1 (2004).
[13] V. Murg, F. Verstraete and J.I. Cirac, Phys. Rev. A 75 (2007), p. 033605.

[14] V. Murg, F. Verstraete and J.I. Cirac, Phys. Rev. Lett. 95 (2005), p. 057206.
[15] B. Paredes, F. Verstraete and J.I. Cirac, Phys. Rev. Lett. 95 (2005), p. 140501.
[16] S. Anders, M.B. Plenio, W. Dür, F. Verstraete and H.J. Briegel, Phys. Rev. Lett. 97 (2006), p. 107206.
[17] W. Heisenberg, Z. Phys. 38 (1926), p. 441.
[18] P. Dirac, Proc. R. Soc. Lond. A 112 (1926), p. 661.
[19] F. Verstraete and A. Winter, in preparation.
[20] E. Størmer, J. Funct. Anal. 3 (1969), p. 48.
[21] V. Coffman, J. Kundu and W.K. Wootters, Phys. Rev. A 61 (2000), p. 052306.
[22] T.J. Osborne and F. Verstraete, Phys. Rev. Lett. 96 (2006), p. 220503.
[23] R. Koenig and R. Renner, J. Math. Phys. 46 (2005), p. 122108.
[24] M. Christandl, R. Koenig, G. Mitchison and R. Renner, Comm. Math. Phys. 273(2) (2007), pp. 473–498.
[25] J.K. Stockton, J.M. Geremia, A.C. Doherty and H. Mabuchi, Phys. Rev. A 67 (2003), p. 022112.
[26] G. Vidal, J.I. Latorre, E. Rico and A. Kitaev, Phys. Rev. Lett. 90 (2003), p. 227902.
[27] X.-G. Wen, *Quantum Field Theory of Many-body Systems*, Oxford University Press, Oxford, 2004.
[28] M. Levin and X.G. Wen, Phys. Rev. Lett. 96 (2006), p. 110405.
[29] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, 2000.
[30] R. Jozsa and N. Linden, Preprint quant-ph/0201143v2 (2002).
[31] L. Bombelli, R.K. Koul, J.H. Lee and R.D. Sorkin, Phys. Rev. D 34 (1986), p. 373.
[32] M. Srednicki, Phys. Rev. Lett. 71 (1993), p. 666.
[33] G′t. Hooft, Nucl. Phys. B 256 (1985), p. 727.
[34] C.G. Callan and F. Wilczek, Phys. Lett. B 333 (1994), p. 55.
[35] T.M. Fiola, J. Preskill, A. Strominger and S.P. Trivedi, Phys. Rev. D 50 (1994), p. 3987.
[36] M.B. Plenio, J. Eisert, J. Dreissig and M. Cramer, Phys. Rev. Lett. 94 (2005), p. 060503.
[37] P. Calabrese and J. Cardy, J. Stat. Mech. 0406 (2004), p. P002.
[38] M.M. Wolf, Phys. Rev. Lett. 96 (2006), p. 010404.
[39] D. Gioev and I. Klich, Phys. Rev. Lett. 96 (2006), p. 100503.
[40] T. Barthel, M.-C. Chung and U. Schollwöck, Phys. Rev. A 74 (2006), p. 022329.
[41] M. Cramer, J. Eisert and M.B. Plenio., Phys. Rev. Lett. 98 (2007), p. 220603.
[42] F. Verstraete, M.M. Wolf, D. Perez-Garcia and J.I. Cirac, Phys. Rev. Lett. 96 (2006), p. 220601.
[43] M.M. Wolf, F. Verstraete, M.B. Hastings and J.I. Cirac, Phys. Rev. Lett. 100 (2008), p. 070502.
[44] M.B. Hastings, Phys. Rev. Lett. 93 (2004), p. 140402.
[45] P. Hayden, D. Leung, P.W. Shor and A. Winter, Commun. Math. Phys. 250 (2004), pp. 371–391.
[46] M.B. Hastings, Phys. Rev. B 76 (2007), p. 035114.
[47] F. Verstraete and J.I. Cirac, Phys. Rev. B 73 (2006), p. 094423.
[48] G. Vidal, Phys. Rev. Lett. 99 (2007), p. 220405.
[49] S. Bravyi, M.B. Hastings and F. Verstraete, Phys. Rev. Lett. 97 (2006), p. 050401.
[50] M.B. Hastings, Phys. Rev. B 73 (2006), p. 085115.
[51] J. Kondo, Prog. Theor. Phys. 32 (1964), p. 37.
[52] H.R. Krishna-murthy, J.W. Wilkins and K.G. Wilson, Phys. Rev. B 21 (1980), p. 1044.
[53] S. Östlund and S. Rommer, Phys. Rev. Lett. 75 (1995), p. 3537.
[54] S. Rommer and S. Östlund, Phys. Rev. B 55 (1997), p. 2164.
[55] D. Perez-Garcia, F. Verstraete, M.M. Wolf and J.I. Cirac, Quantum Inf. Comput. 7 (2007), p. 401.
[56] A. Klümper, A. Schadschneider and J. Zittartz, Europhys. Lett. 24 (1993), p. 293.
[57] C. Lange, A. Klümper and J. Zittartz, Z. Phys. B 96 (1994), p. 267.

[58] M. Fannes, B. Nachtergaele and R.F. Werner, Comm. Math. Phys. 144 (1992), p. 443.
[59] ——, Europhys. Lett. 10 (1989), p. 633.
[60] L. Accardi, Phys. Rep. 77 (1981), p. 169.
[61] F. Verstraete, A. Weichselbaum, U. Schollwöck, J.I. Cirac and J. von Delft, Preprint cond-mat/0504305v2 (2005).
[62] I. Affleck, T. Kennedy, E.H. Lieb and H. Tasaki, Commun. Math. Phys. 115 (1988), p. 477.
[63] F.D. Haldane, Phys. Rev. Lett. 50 (1983), p. 1153.
[64] ——, Phys. Lett. A 93 (1983), p. 464.
[65] F. Verstraete, M.A. Martin-Delgado and J.I. Cirac, Phys. Rev. Lett. 92 (2004), p. 087201.
[66] A.K. Kolezhuk, R. Roth and U. Schollwöck, Phys. Rev. Lett. 77 (1996), p. 5142.
[67] A.K. Kolezhuk and H.J. Mikeska, Phys. Rev. Lett. 90 (1998), p. 2709.
[68] A.K. Kolezhuk, H.J. Mikeska, K. Maisinger and U. Schollwöck, Phys. Rev. B 59 (1999), p. 13565.
[69] G. Vidal, Phys. Rev. Lett. 93 (2004), p. 040502.
[70] M. Cozzini, R. Ionicioiu and P. Zanardi, Phys. Rev. B 76 (2007), p. 104420.
[71] M. Fannes, B. Nachtergaele and R.F. Werner, J. Stat. Phys. 66 (1992), p. 939.
[72] F. Verstraete et al., in preparation.
[73] Y.Y. Shi, L.M. Duan and G. Vidal, Phys. Rev. A 74 (2006), p. 022320.
[74] N. Robertson and P.D. Seymour, J. Algorithms 7 (1986), p. 309.
[75] Y.Y. Shi and I. Markov, Preprint quant-ph/0511069 (2005).
[76] F. Verstraete, J.I. Cirac, J.I. Latorre, E. Rico and M.M. Wolf, Phys. Rev. Lett. 94 (2005), p. 140601.
[77] G. Vidal, Preprint quant-ph/0610099v1 (2006).
[78] ——, Preprint 0707.1454 (2007).
[79] K. Audenaert, F. Verstraete and B. De Moor, Phys. Rev. A 64 (2001), p. 052304.
[80] C.M. Dawson, J. Eisert and T.J. Osborne, Phys. Rev. Lett. 100 (2008), p. 130501.
[81] A. Ben-Tal and A. Nemirovski, Math. Operat. Res. 23 (1998), pp. 769–805.
[82] J. Eisert, Phys. Rev. Lett. 97 (2006), p. 260501.
[83] D. Porras, F. Verstraete and J.I. Cirac, Phys. Rev. B 73 (2006), p. 014410.
[84] C. Raas, G.S. Uhrig and F.B. Anders, Phys. Rev. B 69 (2004), p. 041102(R).
[85] H. Takasaki, T. Hikihara and T. Nishino, J. Phys. Soc. Jpn 68 (1999), p. 1537.
[86] J. Dukelsky, M.A. Martin-Delgado, T. Nishino and G. Sierra, Europhys. Lett. 43 (1998), p. 457.
[87] S.R. White and D.J. Scalapino, Phys. Rev. Lett. 91 (2003), p. 136403.
[88] S.R. White, Phys. Rev. B 72 (2005), p. 180403.
[89] I. McCulloch and M. Gulacsi, Europhys. Lett. 57 (2002), p. 852.
[90] S. Singh, H.Q. Zhou and G. Vidal, Preprint cond-mat/0701427 (2007).
[91] T.A. Costi and C. Kieffer, Phys. Rev. Lett. 76 (1996), p. 1683.
[92] T.A. Costi, Phys. Rev. Lett. 80 (1998), p. 1038.
[93] K.A. Hallberg, Phys. Rev. B 52 (1995), p. R9827.
[94] T.D. Kühner and S.R. White, Phys. Rev. B 60 (1999), p. 335.
[95] E. Jeckelmann, Phys. Rev. B 66 (2002), p. 045114.
[96] J.I. Cirac and P. Zoller, Phys. Today 57 (2004), p. 38.
[97] M.A. Cazalilla and J.B. Marston, Phys. Rev. Lett. 88 (2002), p. 256403.
[98] H.G. Luo, T. Xiang and X.Q. Wang, Phys. Rev. Lett. 91 (2003), p. 049701.
[99] M.A. Cazalilla and J.B. Marston, Phys. Rev. Lett. 91 (2003), p. 049702.
[100] S.R. White and A.E. Feiguin, Phys. Rev. Lett. 93 (2004), p. 076401.
[101] A.J. Daley, C. Kollath, U. Schollwöck and G. Vidal, J. Stat. Mech. Theor. Exp. 4 (2004), p. P04005.
[102] A.E. Feiguin and S.R. White, Phys. Rev. B 72 (2005), p. 220401.
[103] J.J. García-Ripoll, New J. Phys. 8 (2006), p. 305.

[104] D. Gobert, C. Kollath, U. Schollwöck and G. Schütz, Phys. Rev. E 71 (2005), p. 036102.

[105] U. Schollwöck and S.R. White, *Effective models for low-dimensional strongly correlated systems*, in G.G. Batrouni and D. Poilbanc, eds., AIP, Melville, New York, 2006, pp. 155.

[106] M. Suzuki, Phys. Lett. A 146 (1990), p. 319.

[107] ——, J. Math. Phys. 32 (1991), p. 400.

[108] A.T. Sornborger and E.D. Stewart, Phys. Rev. A 60 (1999), p. 156.

[109] I.P. Omelyan, I.M. Mryglod and R. Folk, Comput. Phys. Commun. 146 (2002), p. 188.

[110] T. Nishino, J. Phys. Soc. Jpn. 64 (1995), p. 3598.

[111] R.J. Bursill, T. Xiang and G.A. Gehring, J. Phys. Condensed Matter 8 (1996), p. L583.

[112] N. Shibata, J. Phys. Soc. Jpn. 66 (1997), p. 2221.

[113] X. Wang and T. Xiang, Phys. Rev. B 56 (1997), p. 5061.

[114] F. Barahona, J. Phys. A: Math. Gen. 15 (1982), p. 3241.

[115] D. Jaksch, C. Bruder, J.I. Cirac, C.W. Gardiner and P. Zoller, Phys. Rev. Lett. 81 (1998), p. 3108.

[116] M.P. Fisher, P.B. Weichmann, G. Grinstein and D.S. Fisher, Phys. Rev. B 40 (1989), p. 546.

[117] H.P. Büchler, G. Blatter and W. Zwerger, Phys. Rev. Lett. 90 (2003), p. 130401.

[118] B. Paredes, A. Widera, V. Murg, O. Mandel, S. Fölling, I. Cirac, G.V. Shlyapnikov, T.W. Hänsch and I. Bloch, Nature 429 (2004), p. 277.

[119] T.J. Osborne, Preprint cond-mat/0605194v1 (2006).

[120] Y. Hieida, K. Okunishi and Y. Akutsu, New J. Phys. 1 (1999), p. 7.

[121] K. Okunishi and T. Nishino, Prog. Theor. Phys. 103 (2000), p. 541.

[122] T. Nishino, K. Okunishi, Y. Hieida, N. Maeshima and Y. Akutsu, Nucl. Phys. B 575 (2004), p. 504.

[123] Y. Nishio, N. Maeshima, A. Gendiar and T. Nishino, Preprint cond-mat/0401115v1 (2004).

[124] H. Niggeman, A. Klümper and J. Zittartz, Z. Phys. B 104 (1997), p. 103.

[125] M.A. Martin-Delgado, M. Roncaglia and G. Sierra, Phys. Rev. B 64 (2001), p. 075117.

[126] F. Verstraete and I.J. Cirac, Phys. Rev. A 70 (2004), p. 060302.

[127] S.R. White and A.L. Chernyshev, Phys. Rev. Lett. 99 (2007), p. 127004.

[128] A. Kitaev, Ann. Phys. 303 (2003), p. 2.

[129] R. Raussendorf and H.J. Briegel, Phys. Rev. Lett. 86 (2001), p. 5188.

[130] A. Isacsson and O.F. Syljuasen, Phys. Rev. E 74 (2006), p. 026701.

[131] M. Girardeau, J. Math. Phys. 1 (1960), p. 516.

[132] T. Kennedy, E.H. Lieb and B.S. Shastry, Phys. Rev. Lett. 61 (1988), p. 2582.

[133] F. Verstraete and J.I. Cirac, J. Stat. Mech. 0509 (2005), p. P012.

[134] J. Jordan, R. Orus, G. Vidal, F. Verstraete and J.I. Cirac, Preprint cond-mat/0703788v3 (2007).

[135] M.M. Wolf, F. Verstraete and J.I. Cirac, Phys. Rev. Lett. 92 (2004), p. 087903.

[136] C.N. Yang and C.P. Yang, Phys. Rev. 150 (1966), pp. 321–339.

[137] O.F. Syljuasen, Phys. Rev. B 322 (2004), p. 25.

[138] H.Q. Lin, J.S. Flynn and D.D. Betts, Phys. Rev. B 64 (2001), p. 214411.

[139] M.R. Dowling, A.C. Doherty and S.D. Bartlett, Phys. Rev. A 70 (2004), p. 062113.

[140] I. Peschel, M. Kaulke and O. Legeza, Ann. Phys. (Leipzig) 8 (1999), p. 153.

[141] B.Q. Jin and V.E. Korepin, J. Stat. Phys. 116 (2004), p. 79.

[142] M.B. Hastings, *An area law for one dimensional quantum systems*, J. Stat. Phys. 8 (2007), p. P08024.

[143] N. Schuch, M.M. Wolf, F. Verstraete and J.I. Cirac, Phys. Rev. Lett. 100 (2008), p. 030504.

[144] N. Schuch, J.I. Cirac and F. Verstraete, Preprint 0802.3351v1.

[145] S. Trebst, U. Schollwöck, M. Troyer and P. Zoller, Phys. Rev. Lett. 96 (2006), p. 250402.

Appendix A: Local reduced density operators of spin systems

Due to the variational nature of ground states, there always exists a ground state with the same symmetries as the associated Hamiltonian. If the Hamiltonian has translational symmetry and consists of two-body nearest-neighbour interactions, then it is clear that the energy of a state with the right symmetry is completely determined by its reduced density operator of two neighbouring spins. The reduced density operators arising from these (eventually mixed) states with a given symmetry form a convex set, and the energy for a given Hamiltonian will be minimized for a state whose reduced density operator is an extreme point in this set. More specifically, the equation $\mathrm{Tr}(H\rho) = E$ determines a hyperplane in the space of reduced density operators of states with a given symmetry, and the energy will be extremal when the hyperplane is tangent to the convex set ($E = E_{\mathrm{extr}}$). The problem of finding the ground-state energy of nearest-neighbour translational-invariant Hamiltonians is therefore equivalent to the determination of the convex set of two-body reduced density operators arising from states with the right symmetry. Strictly speaking, these two problems are dual to each other. In the case of quadratic Hamiltonians involving continuous variables, the determination of this convex set was solved for fairly general settings in [135] by means of Gaussian states. The determination of this convex set in the case of spin systems, however, turns out to be much more challenging.

Let us illustrate this with a simple example

$$\mathcal{H} = -\sum_{\langle i,j\rangle} S_i^x S_j^x + S_i^y S_j^y + \Delta S_i^z S_j^z$$

on a lattice of arbitrary geometry and dimension. (We assume that the graph corresponding to the lattice is edge transitive, meaning that any vertex can be mapped to any other vertex by application of the symmetry group of the graph.) Due to the symmetries, the reduced density operator of two nearest neighbours can be parameterized by only two parameters (this can easily be proven by invoking local twirling operations which leave the Hamiltonian invariant):

$$\rho = \frac{1}{4}\big(I \otimes I + x(\sigma_x \otimes \sigma_x + \sigma_y \otimes \sigma_y) + z\sigma_z \otimes \sigma_z\big).$$

Positivity of ρ enforces $-1 \leq z \leq 1 - 2|x|$, and the state is separable if and only if $1 + z \leq 2|x|$. In the case of an infinite one-dimensional spin chain, the ground-state energy $E(\Delta)$ has been calculated exactly [136], and this determines the tangent hyperplanes

$$2x + z\Delta + E(\Delta) = 0$$

whose envelope makes up the extreme points of the convex set of reduced density operators of translationally invariant one-dimensional states: the boundary of this convex set is parameterized by

$$\begin{aligned} z &= -\partial E(\Delta)/\partial\Delta \\ x &= -(E(\Delta) + \partial E(\Delta)/\partial\Delta)/2, \end{aligned}$$

which we plotted in Figure A1. We also plot the boundary for the two-dimensional square lattice. These two-dimensional data were obtained by numerical methods [12,137,138]); of course, this convex set is contained in the previous set, as all of the semidefinite constraints defining the set corresponding to one dimension are strictly included in the set of constraints for the two-dimensional case. Finally, we plot the set of separable states, which contains the reduced density operators of the allowed states for a lattice with infinite coordination number. The boundary of this separable set is given by the inner diamond; this immediately implies that the difference between the exact energy and that obtained by mean field theory will be maximized whenever the hyperplane that forms the boundary of the first set will be parallel to this line. This happens when $\Delta = -1$ (independent of the dimension), which corresponds to the antiferromagnetic case, and this proves that the 'entanglement gap' [139] in the *XXZ* plane is maximized by the antiferromagnetic ground state for any dimension and geometry. Similarly, it proves that the ground state is separable whenever $\Delta \geq 1$ and $\Delta = -\infty$. Note also that in the two-dimensional case, part of the boundary of the convex set consists of a plane parameterized by $2x + z + E(1) = 0$. This indicates a degeneracy of the ground state around the antiferromagnetic

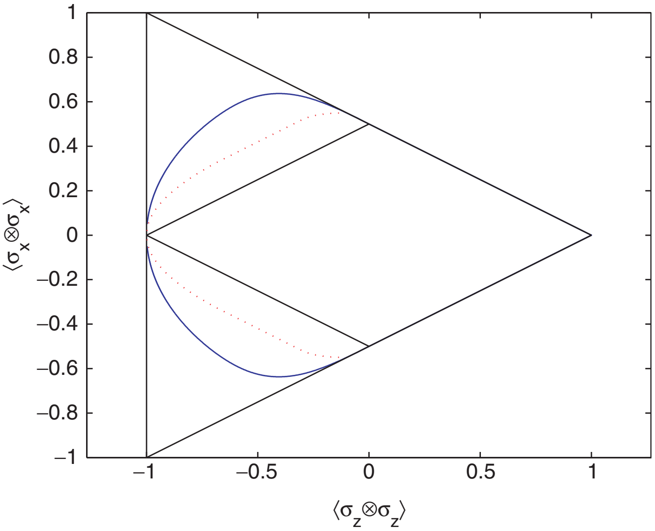

Figure A1. Convex sets of the possible reduced density operators of translational-invariant spin-1/2 states in the *XX–ZZ* plane. The big triangle represents all positive density operators; the inner parallelogram represents the separable states; the union of the separable cone and the convex hull of the full curved line is the complete convex set in the case of a one-dimensional geometry, and the dashed lines represent extreme points in the two-dimensional case of a square lattice. The singlet corresponds to the point with coordinates (−1, −1). (Reprinted with permission from [47] F. Verstraete *et al.*, Phys. Rev. B 73 (2006), p. 094423. © 2006 by the American Physical Society.)

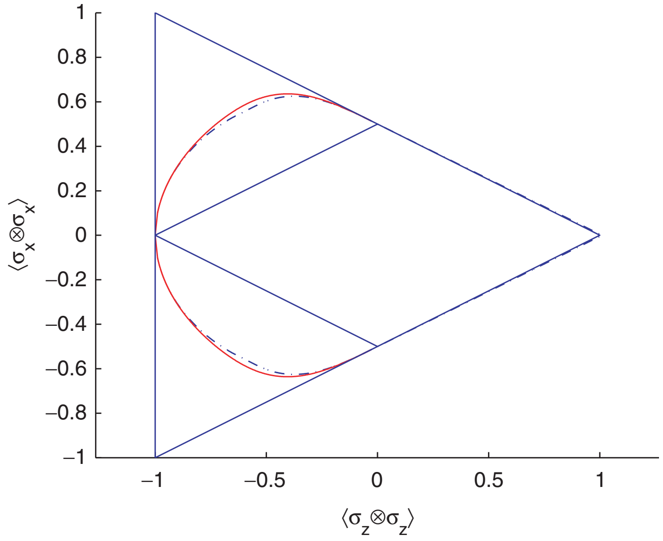

Figure A2. Convex sets in the *XXZ* plane: the inner diamond borders the set of separable states (see Figure 1); dash-dotted lines represent extreme points of the convex set produced by MPSs of $D = 2$. (Reprinted with permission from [47] F. Verstraete *et al.*, Phys. Rev. B 73 (2006), p. 094423. © 2006 by the American Physical Society.)

point, and indicates that a phase transition is occurring at that point (more specifically between an Ising and a Berezinskii–Kosterlitz–Thouless phase).

As a good illustration of the actual accuracy obtained with MPSs, we calculated the convex set obtained with MPSs in the thermodynamic limit for the XXZ chain with $D=2$, where D is the dimension of the matrices in the MPS (see Figure A2). It is almost unbelievable how good the exact convex set can be approximated. Note that typical DMRG calculations have $D\sim 200$, and that the accuracy grows superpolynomial in D. Note also that the $D=1$ case corresponds to mean-field theory, whose corresponding convex set coincides with the set of separable states.

The same argument involving the notion of a correlation length applies in higher dimensions and indicates that PEPSs represent ground states of gapped local Hamiltonians well. Note, however, that the convex set in the two-dimensional case is much closer to the separable set than in the one-dimensional case; this gives a hint that PEPSs of smaller dimension will suffice to give the same accuracy as in the one-dimensional case. In the next Section we quantitatively bound how well a translationally invariant state can be represented in terms of a MPS, and analyze the corresponding implications for the description of ground states of one-dimensional spin chains.

Appendix B: MPSs represent ground states faithfully

We derive an upper bound to the error made by approximating a general ground state of a one-dimensional quantum spin system by a MPS. As we show below, this has very important implications in the performance of the renormalization algorithms in describing ground states of one-dimensional spin chains.

Lemma B1: *There exists a MPS* $|\psi_D\rangle$ *of dimension D such that*

$$\| |\psi\rangle - |\psi_D\rangle \|^2 \leq 2\sum_{\alpha=1}^{N-1} \epsilon_\alpha(D),$$

where $\epsilon_\alpha(D) = \sum_{i=D+1}^{N_\alpha} \mu^{[\alpha]i}$.

***Proof*:** We can always write ψ as a MPS of dimension $D=2^{N/2}$ and fulfilling

$$\sum_i A^{[m]i} A^{[m]i\dagger} = I, \quad \sum_i A^{[m]i\dagger} \Lambda^{[m+1]} A^{[m]i} = \Lambda^{[m]}.$$

Let us now consider the D-dimensional MPS $|\psi_D\rangle$ which is defined by the $D \times D$ matrices $[A^{[\alpha]i}]_{(1\ldots D,1\ldots D)}$ (i.e. the upper-left block of $A^{[\alpha]i}$). The goal is now to bound $\langle\psi|\psi_D\rangle$. The gauge conditions were chosen such as to make the task simple:

$$\langle\psi_D|\psi\rangle = \mathrm{Tr}[\$_2(\cdots \$_{N-2}(\$_{N-1}(\Lambda^{N-1}P)P)P\cdots)P]; \tag{B1}$$

here $P=\sum_{k=1}^{D} |k\rangle\langle k|$ and $\$_m(X) = \sum_i A^{[m]i\dagger} X A^{[m]i}$ represents a trace-preserving completely positive (TPCP) map parameterized by the Kraus operators $A^{[m]i}$. Let us now recursively define

$$Y^{[k]} = \$_k\big(Y^{[k+1]}P\big), \quad Y^{[N-1]} = \Lambda^{[N-1]}P;$$

observe that $\Lambda^{[k]} = \$_k(\Lambda^{[k+1]})$. We want a bound on $\mathrm{Tr}|\Lambda^{[1]} - Y_1|$, as equation (B1) is equal to $\mathrm{Tr}(Y^{[2]})$. The crucial property we need is that TPCP maps are contractive with relation to the trace-norm (this can directly be proven by considering the Neumark representation of a TPCP map as a unitary in a larger space): $\mathrm{Tr}|\$(X)| \leq \mathrm{Tr}|X|$. It follows that

$$\begin{aligned}\mathrm{Tr}|\Lambda^{[k]} - Y^{[k]}| &= \mathrm{Tr}|\$_k\big(\Lambda^{[k+1]} - Y^{[k+1]}P\big)| \\ &\leq \mathrm{Tr}|\Lambda^{[k+1]} - Y^{[k+1]}P| \\ &\leq \mathrm{Tr}|\Lambda^{[k+1]} - Y^{[k+1]}| + \mathrm{Tr}|\Lambda^{[k+1]}(I-P)|.\end{aligned}$$

Note that the last term in the sum is given exactly by $\sum_{\alpha=D+1}^{2^{N/2}} \lambda^{[k+1]\alpha}$. The theorem now follows immediately by recursion and by observing that $\langle\psi_D|\psi_D\rangle \leq 1$ by similar arguments. □

The implications of this result are very strong: it shows that for systems for which the $\epsilon_\alpha(D)$ decay fast in D, there exist MPSs with *small D* which will not only reproduce well the local correlations (such as energy) but also all of the non-local properties (such as correlation length). The following lemma now relates the derived bound to the Renyi entropies of the reduced density operators, through which one can make the connection to the ground states of one-dimensional Hamiltonians. The Renyi entropies of ρ are defined as

$$S^\alpha(\rho) = \frac{1}{1-\alpha}\log(\mathrm{Tr}\,\rho^\alpha),$$

and we consider $0 < \alpha < 1$. We denote as before $\epsilon(D) = \sum_{i=D+1}^{\infty} \lambda_i$ with λ_i the non-increasingly ordered eigenvalues of ρ. Then we have the following.

Lemma B2: *Given a density operator* ρ. *If* $0 < \alpha < 1$, *then*

$$\log(\epsilon(D)) \leq \frac{1-\alpha}{\alpha}\left(S^\alpha(\rho) - \log\frac{D}{1-\alpha}\right).$$

Proof: Let us first characterize the probability distribution that has maximal possible weight in its tail (i.e. $p = \sum_{i=D+1}^{\infty} p_i$) for a given Renyi entropy. Introducing a free parameter $0 < h \leq (1-p)/D$, such a probability distribution must be of the form

$$p_1 = 1 - p - (D-1)h$$

$$h = p_2 = p_3 = \cdots p_{D+p/h}$$

$$p_{D+p/h+1}, \cdots p_\infty = 0$$

because this distribution majorizes all other distributions with given p, D, p_D (Renyi-entropies are Schur-convex functions). For a given p, D, h, it holds that

$$\begin{aligned}\sum_i p_i^\alpha &= (1-p-(D-1)h)^\alpha + (D-1+p/h)h^\alpha \\ &\geq Dh^\alpha + ph^{\alpha-1}.\end{aligned}$$

Minimizing this expression with relation to h, we obtain

$$\sum_i p_i^\alpha \geq (D^{1-\alpha}p^\alpha)/((1-\alpha)^{1-\alpha}\alpha^\alpha).$$

Denoting $S^\alpha(p, D)$ the minimal possible entropy for given p, D, we obtain

$$S^\alpha(p, D) \geq \frac{1}{1-\alpha}\log\left(\frac{D^{1-\alpha}p^\alpha}{(1-\alpha)^{1-\alpha}\alpha^\alpha}\right)$$

and, hence,

$$p \leq \exp\left(\frac{1-\alpha}{\alpha}\left(S^\alpha(p, D) - \log\frac{D}{1-\alpha}\right)\right).$$

The proof now follows by replacing $S^\alpha(p, D)$ by $S^\alpha(\rho)$. □

This lemma is very interesting in the light of the fact that in the case of critical systems, arguably the hardest to simulate (for non-critical systems, the renormalization group flow is expected to increase the Renyi entropies in the UV direction; the corresponding fixed point corresponds to a critical system whose entropy thus upper bounds that of the non-critical system), the Renyi-entropy of a contiguous block of L spins scales as [26,37,140,141]

$$S^{\alpha}(\rho_L) \simeq \frac{c+\bar{c}}{12}\left(1+\frac{1}{\alpha}\right)\log(L) \tag{B2}$$

for all $\alpha > 0$; here c is the central charge. The fact that the eigenvalues of ρ_L decay fast has previously been identified as a indication of the validity of the DMRG approach [140]. The truncation error [2,4], which has been used in the DMRG community as a check for convergence, is essentially given by $\epsilon(D) - \epsilon(2D)$ and therefore indeed gives a good idea of the error in a simulation.

Let us investigate how the computational effort to simulate such critical systems scales as a function of the length $N = 2L$ of the chain. Let us therefore consider the Hamiltonian associated to a critical system, but restrict it to $2L$ sites. The entropy of a half chain (we consider the ground state $|\psi_{\rm ex}\rangle$ of the finite system) will typically scale as in equation (B2) but with an extra term that scales as $1/N$. Suppose that we want to enforce that $\| |\psi_{\rm ex}\rangle - |\psi_D\rangle \|^2 \leq \epsilon_0/L$ with ϵ_0 independent of L. (We choose the $1/L$ dependence so as to ensure that the absolute error in extensive observables does not grow.) Denote the minimal D needed to obtain this precision for a chain of length $2L$ by D_L. Following Lemma B1 and the fact that the entropy of all possible contiguous blocks reaches its maximum in the middle of the chain (hence, $p \leq \epsilon_0/L^2$ is certainly sufficient), Lemmas B.1 and B.2 combined yield

$$D_L \leq cst\left(\frac{L^2}{(1-\alpha)\epsilon_0}\right)^{\alpha/(1-\alpha)} L^{((c+\bar{c})/12)((1+\alpha)/\alpha)}.$$

This shows that D only has to scale polynomially in L to keep the accuracy ϵ_0/L fixed; in other words, there exists an efficient scalable representation for ground states of critical systems (and, hence, also of non-critical systems) in terms of MPSs. Such a strong result could not have been anticipated from just performing simulations. Furthermore, Hastings has proven that ground states of gapped systems always obey a strict area law [142]. This implies that the ground state of any gapped spin chain is indeed well approximated by a MPS. For a more detailed description of the relations between area laws and approximability, we refer the interested reader to [143].

Now what about the complexity of finding this optimal MPS? It has been observed that DMRG converges exponentially fast to the ground state with a relaxation time proportional to the inverse of the gap Δ of the system [4]. For translational-invariant critical systems, this gap seems to close only polynomially. As we have proven that D also only have to scale polynomially, the complexity of deciding whether the ground-state energy of one-dimensional quantum systems is below a certain value is certainly in NP (as it can be checked efficiently if Merlin gives the MPS description). The problem would even be in P if the following conditions are met: (1) the α-entropy of blocks in the exact ground state grow at most logarithmically with the size of the block for some $\alpha < 1$; (2) the gap of the system scales at most polynomially with the system size; (3) given a gap that obeys condition (2), there exists an efficient DMRG-like algorithm that converges to the global minimum. As the VMPS approach [10] is essentially an ALS method of solving a non-convex problem which is at worst NP-hard [81,82], there is *a priori* no guarantee that it will converge to the global optimum, although the occurrence of local minima seems to be unlikely [4]. Surprisingly, one can indeed construct a family of Hamiltonians with nearest-neighbour interactions on a line for which the ground state is an exact MPS with polynomial D, but for which it is an NP-complete problem to find it [144]. As the corresponding Hamiltonian has a gap that closes polynomially in the size of the system, the only hope of VMPS/DMRG methods being in P is in the case of gapped systems. (Let us specify an alternative method which should in principle not become trapped in local minima in the case of gapped systems. As in the adiabatic theorem, we can construct a time-dependent Hamiltonian $H(t)$ with $H(0)$ trivial and $H(1)$ the Hamiltonian to simulate; if we discretize this evolution in a number of steps that grows polynomially in the inverse gap, the adiabatic theorem guarantees that we will end up in the ground state of $H(1)$ if we can follow the ground state of $H(t)$ closely (see, e.g. [145] for a nice application in the context of quantum simulators). The idea is to make D of $|\psi_D(t)\rangle$ large enough such as to follow the ground state $|\psi(t)\rangle$ close enough in such a way that the optimization is always convex around the global optimum within the domain $\| |\chi\rangle - |\psi(t)\rangle \| \leq \epsilon$. As we are simulating this classically, we could even perform imaginary time evolution over a longer time during each step.)

Appendix C: Matlab code

Finally, we would like to give an idea of how to program the variational methods explained in the previous Sections. We present two functions, one for the calculation of ground and first excited states and one for the reduction of the virtual dimension of MPSs. We demonstrate these functions by means of the antiferromagnetic Heisenberg chain.

C.1. *Minimization of the energy*

The function `minimizeE` optimizes the parameters of a MPS in such a way that the expectation value with respect to a given Hamiltonian tends to a minimum. The function expects this Hamiltonian to be defined in a $M \times N$ cell `hset`, where N denotes the number of sites and M the number of terms in the Hamiltonian. Assuming that the Hamiltonian is of the form

$$H = \sum_{m=1}^{M} h_m^{(1)} \otimes \cdots \otimes h_m^{(N)},$$

the element `hset`$\{m, j\}$ is equal to $h_m^{(j)}$. Further arguments are the virtual dimension of the resulting MPS, D, and the expected accuracy of the energy, `precision`.

Output arguments are the optimized energy E and corresponding MPS `mps`. The MPS is stored as a $1 \times N$ cell, each entry corresponding to one matrix.

Optionally, a MPS `mpsB` can be specified as an argument to which the resulting state shall be orthogonal. This is especially useful for calculating the first excited state.

```
function [E, mps] = minimizeE(hset, D, precision, mpsB)

[M, N] = size(hset);
d = size(hset{1, 1}, 1);
mps = createrandommps(N, D, d);
mps = prepare(mps);

% storage-initialization
Hstorage = initHstorage(mps, hset, d);
if ~isempty(mpsB), Cstorage = initCstorage(mps, [], mpsB, N); end
P = [];

% optimization sweeps
while 1
  Evalues = [];

  % ********* cycle 1: j → j + 1 (from 1 to N − 1) *********
  for j = 1:(N − 1)
  % projector-calculation
  if ~isempty(mpsB)
    B = mpsB{j};
    Cleft = Cstorage{j};
    Cright = Cstorage{j + 1};
    P = calcprojector_onesite(B, Cleft, Cright);
  end

  % optimization
  Hleft = Hstorage(:, j);
  Hright = Hstorage(:, j + 1);
  hsetj = hset(:, j);
  [A, E] = minimizeE_onesite(hsetj, Hleft, Hright, P);
  [A, U] = prepare_onesite(A, ’lr’);
  mps{j} = A;
  Evalues = [Evalues, E];

  % storage-update
  for m = 1:M
    h = reshape(hset{m, j}, [1, 1, d, d]);
    Hstorage{m, j + 1} = updateCleft(Hleft{m}, A, h, A);
  end
  if ~isempty(mpsB)
    Cstorage{j + 1} = updateCleft(Cleft, A, [], B);
  end
end

% ***************** cycle 2: j → j − 1 (from N to 2) *****************
for j = N:(-1):2
  % projector-calculation
  if ~isempty(mpsB)
    B = mpsB{j};
    Cleft = Cstorage{j};
    Cright = Cstorage{j + 1};
    P = calcprojector_onesite(B, Cleft, Cright);
  end

  % minimization
  Hleft = Hstorage(:, j);
  Hright = Hstorage(:, j + 1);
  hsetj = hset(:, j);
  [A, E] = minimizeE_onesite(hsetj, Hleft, Hright, P);
  [A, U] = prepare_onesite(A, ’rl’);
  mps{j} = A;
  Evalues = [Evalues, E];

  % storage-update
    for m = 1:M
      h = reshape(hset{m, j}, [1, 1, d, d]);
      Hstorage{m, j} = updateCright(Hright{m}, A, h, A);
    end

    if ~isempty(mpsB)
      Cstorage{j} = updateCright(Cright, A, [], B);
    end

  end

  if (std(Evalues)/abs(mean(Evalues)) < precision)
    mps{1} = contracttensors(mps{1}, 3, 2, U, 2, 1);
    mps{1} = permute(mps{1}, [1, 3, 2]);
    break;
  end

end

% ******************* one-site optimization *********************

function [A, E] = minimizeE_onesite(hsetj, Hleft, Hright, P)
```

```
DAl=size(Hleft{1},1);
DAr=size(Hright{1},1);
d=size(hsetj{1},1);

% calculation of Heff
M=size(hsetj,1);

Heff=0;
for m=1:M
  Heffm=contracttensors(Hleft{m},3,2,Hright{m},3,2);
  Heffm=contracttensors(Heffm,5,5,hsetj{m},3,3);
  Heffm=permute(Heffm,[1,3,5,2,4,6]);
  Heffm=reshape(Heffm,[DAl*DAr*d,DAl*DAr*d]);
  Heff=Heff+Heffm;
end

% projection on orthogonal subspace
if ~isempty(P),Heff=P'*Heff*P; end

% optimization
options.disp=0;
[A,E]=eigs(Heff,1,'sr',options);
if ~isempty(P),A=P*A; end
A=reshape(A,[DAl,DAr,d]);


function [P]=calcprojector_onesite(B,Cleft,Cright)

y=contracttensors(Cleft,3,3,B,3,1);
y=contracttensors(y,4,[2,3],Cright,3,[2,3]);
y=permute(y,[1,3,2]);
y=reshape(y,[prod(size(y)),1]);
Q=orth([y,eye(size(y,1))]);
P=Q(:,2:end);
```

C.2. *Time evolution*

The function `reduceD` forms the basis for the simulation of a time evolution. It multiplies a given MPS with a given MPO and reduces the virtual dimension of the resulting state, that is, it searches a MPS with reduced virtual dimension and minimal distance to the original state. The MPS and the MPO are specified in the arguments `mpsA` and `mpoX`. As before, they are represented by a cell with entries identifying the matrices. The reduced virtual dimension is specified in the argument `DB`. The argument `precision` defines the convergence condition: if fluctuations in the distance are less than `precision`, the optimization is assumed to be finished.

The output argument is the optimized MPS `mpsB` with virtual dimension `DB`.

```
function mpsB=reduceD(mpsA,mpoX,DB,precision)

N=length(mpsA);
d=size(mpsA{1},3);
mpsB=createrandommps(N,DB,d);
mpsB=prepare(mpsB);
% initialization of the storage
Cstorage=initCstorage(mpsB,mpoX,mpsA,N);

% optimization sweeps
while 1
  Kvalues=[];
  % ***************** cycle 1: j → j + 1 (from 1 to N − 1) ****************
  for j = 1:(N − 1)
    % optimization
    Cleft = Cstorage{j};
    Cright = Cstorage{j + 1};
    A = mpsA{j}; X = mpoX{j};
    [B, K] = reduceD2_onesite(A, X, Cleft, Cright);
    [B, U] = prepare_onesite(B, ’lr’);
    mpsB{j} = B;
    Kvalues = [Kvalues, K];
    % storage-update
    Cstorage{j + 1} = updateCleft(Cleft, B, X, A);
  end

% ***************** cycle 2: j → j − 1 (from N to 2) *****************
for j = N:(-1):2
  % optimization
  Cleft = Cstorage{j};
  Cright = Cstorage{j + 1};
  A = mpsA{j}; X = mpoX{j};
  [B, K] = reduceD2_onesite(A, X, Cleft, Cright);
  [B, U] = prepare_onesite(B, ’rl’);
  mpsB{j} = B;
  Kvalues = [Kvalues, K];
  % storage-update
    Cstorage{j} = updateCright(Cright, B, X, A);
  end

  if std(Kvalues)/abs(mean(Kvalues)) lt;precision
    mpsB{1} = contracttensors(mpsB{1}, 3, 2, U, 2, 1);
    mpsB{1} = permute(mpsB{1}, [1, 3, 2]);
    break;
  end

end

% ****************** one-site optimization ********************

function [B, K] = reduceD2_onesite(A, X, Cleft, Cright)

Cleft = contracttensors(Cleft, 3, 3, A, 3, 1);
Cleft = contracttensors(Cleft, 4, [2, 4], X, 4, [1, 4]);

B = contracttensors(Cleft, 4, [3, 2], Cright, 3, [2, 3]);
B = permute(B, [1, 3, 2]);

b = reshape(B, [prod(size(B)), 1]);
K = -b’ * b;
```

C.3. *Auxiliary functions*

The previous two functions depend on several auxiliary functions that are printed in this Section.

- Gauge transformation that prepares the MPS `mpsB` in such a form that N_{eff} is equal to the identity for the first spin (see Section 3.3):

  ```
  function [mps] = prepare(mps)
  ```

```
N=length(mps);

for i=N:-1:2
  [mps{i},U]=prepare_onesite(mps{i},'rl');
  mps{i-1}=contracttensors(mps{i-1},3,2,U,2,1);
  mps{i-1}=permute(mps{i-1},[1,3,2]);
end

function [B,U,DB]=prepare_onesite(A,direction)

[D1,D2,d]=size(A);
switch direction
case 'lr'
  A=permute(A,[3,1,2]); A=reshape(A,[d*D1,D2]);
  [B,S,U]=svd2(A); DB=size(S,1);
  B=reshape(B,[d,D1,DB]); B=permute(B,[2,3,1]);
  U=S*U;
case 'rl'
  A=permute(A,[1,3,2]); A=reshape(A,[D1,d*D2]);
  [U,S,B]=svd2(A); DB=size(S,1);
  B=reshape(B,[DB,d,D2]); B=permute(B,[1,3,2]);
  U=U*S;
end
```

- Initialization of storages:

```
function [Hstorage]=initHstorage(mps,hset,d)
[M,N]=size(hset);
Hstorage=cell(M,N+1);
for m=1:M, Hstorage{m,1}=1; Hstorage{m,N+1}=1; end
for j=N:-1:2
  for m=1:M
    h=reshape(hset{m,j},[1,1,d,d]);
    Hstorage{m,j}=updateCright(Hstorage
      {m,j+1},mps{j},h,mps{j});
  end

end

function [Cstorage]=initCstorage(mpsB,mpoX,mpsA,N)

Cstorage=cell(1,N+1);
Cstorage{1}=1;
Cstorage{N+1}=1;
for i=N:-1:2
  if isempty(mpoX),X=[]; else X=mpoX{i}; end
  Cstorage{i}=updateCright(Cstorage
   {i+1},mpsB{i},X,mpsA{i});
end

function [Cleft]=updateCleft(Cleft,B,X,A)

if isempty(X),X=reshape(eye(size(B,3)),[1,1,2,2]); end

Cleft=contracttensors(A,3,1,Cleft,3,3);
Cleft=contracttensors(X,4,[1,4],Cleft,4,[4,2]);
Cleft=contracttensors(conj(B),3,[1,3],Cleft,4,[4,2]);

function [Cright]=updateCright(Cright,B,X,A)

if isempty(X),X=reshape(eye(size(B,3)),[1,1,2,2]); end
```

```
Cright=contracttensors(A,3,2,Cright,3,3);
Cright=contracttensors(X,4,[2,4],Cright,4,[4,2]);
Cright=contracttensors(conj(B),3,[2,3],Cright,4,[4,2]);
```

- Creation of a random MPS:

```
function [mps]=createrandommps(N,D,d)
mps=cell(1,N);
mps{1}=randn(1,D,d)/sqrt(D);
mps{N}=randn(D,1,d)/sqrt(D);
for i=2:(N-1)
  mps{i}=randn(D,D,d)/sqrt(D);
end
```

- Expectation value of the MPS mps with respect to the operator defined in `hset`:

```
function [e,n]=expectationvalue(mps,hset)

[M,N]=size(hset);
d=size(mps{1},3);

% expectation value
e=0;
for m=1:M
    em=1;
    for j=N:-1:1
      h=hset{m,j};
      h=reshape(h,[1,1,d,d]);
      em=updateCright(em,mps{j},h,mps{j});
    end
    e=e+em;
end

% norm
n=1;
X=eye(d); X=reshape(X,[1,1,d,d]);
for j=N:-1:1
  n=updateCright(n,mps{j},X,mps{j});
end

e=e/n;
```

- Contraction of index `indX` of tensor `X` with index `indY` of tensor `Y` (`X` and `Y` have a number of indices corresponding to `numindX` and `numindY`, respectively):

```
function [X,numindX]=contracttensors
     (X,numindX,indX,Y,numindY,indY)

Xsize=ones(1,numindX); Xsize(1:length(size(X)))=size(X);
Ysize=ones(1,numindY); Ysize(1:length(size(Y)))=size(Y);

indXl=1:numindX; indXl(indX)=[];
indYr=1:numindY; indYr(indY)=[];

sizeXl=Xsize(indXl);
sizeX=Xsize(indX);
sizeYr=Ysize(indYr);
sizeY=Ysize(indY);

if prod(sizeX)~=prod(sizeY)
  error('indX and indY are not of same dimension.');
```

```
end

    if isempty(indYr)
        if isempty(indXl)
            X=permute(X, [indX]);
            X=reshape(X, [1, prod(sizeX)]);

            Y=permute(Y, [indY]);
            Y=reshape(Y, [prod(sizeY), 1]);
            X=X*Y;
            Xsize=1;

    return;

else
    X=permute(X, [indXl, indX]);
    X=reshape(X, [prod(sizeXl), prod(sizeX)]);
    Y=permute(Y, [indY]);
    Y=reshape(Y, [prod(sizeY), 1]);

    X=X*Y;
    Xsize=Xsize(indXl);

X=reshape(X, [Xsize, 1]);

return
end

end

X=permute(X, [indXl, indX]);
X=reshape(X, [prod(sizeXl), prod(sizeX)]);

Y=permute(Y, [indY, indYr]);
Y=reshape(Y, [prod(sizeY), prod(sizeYr)]);

X=X*Y;
Xsize=[Xsize(indXl), Ysize(indYr)];
numindX=length(Xsize);
X=reshape(X, [Xsize, 1]);
```

- Economical SVD:

```
function [U, S, V]=svd2(T)

[m, n]=size(T);
if m gt;=n, [U, S, V]=svd(T, 0); else [V, S, U]=svd(T', 0); end
V=V';
```

C.4. *Examples*

As a first example, we show how to calculate the ground state and the first excited state of the antiferromagnetic Heisenberg chain using the method `minimizeE` from before. In this example, the chain length N is assumed to be 10 and the virtual dimension `D` is set to 5.

```
N=10;
D=5;
precision=1e-5;

% Heisenberg Hamiltonian
M=3*(N-1);
hset=cell(M, N);
```

```
sx = [0, 1;1, 0]; sy = [0, -1i;1i, 0]; sz = [1, 0;0, -1]; id = eye(2);
for m = 1:M, for j = 1:N, hset{m, j} = id; end; end
for j = 1:(N - 1)
   hset{3 * (j - 1) + 1, j} = sx; hset{3 * (j - 1) + 1, j + 1} = sx;
   hset{3 * (j - 1) + 2, j} = sy; hset{3 * (j - 1) + 2, j + 1} = sy;
   hset{3 * (j - 1) + 3, j} = sz; hset{3 * (j - 1) + 3, j + 1} = sz;
end

% ground state energy
randn('state', 0)
[E0, mps0] = minimizeE(hset, D, precision, []);
fprintf('E0 = %g\n', E0);

% first excited state
[E1, mps1] = minimizeE(hset, D, precision, mps0);
fprintf('E1 = %g\n', E1);
```

As a second example, we focus on the real time evolution with respect to the Heisenberg antiferromagnet. The starting state is a product state with all spins pointing in the z-direction except for the central spin which is flipped. We evolve the state with the method `reduceD` and calculate at each step the magnetization m_z of the central spin. As before, `N` is equal to 10 and `D` is set to 5.

```
N = 10;
D = 5;
precision = 1e-5;
dt = 0.03;
jflipped = 5;
% magnetization in z-direction
oset = cell(1, N);

sx = [0, 1;1, 0]; sy = [0, -1i;1i, 0]; sz = [1, 0;0, -1]; id = eye(2);
for j = 1:N, oset{1, j} = id; end;
oset{1, jflipped} = sz;

% time evolution operator
h = kron(sx, sx) + kron(sy, sy) + kron(sz, sz);
w = expm(-1i * dt * h);
w = reshape(w, [2, 2, 2, 2]); w = permute(w, [1, 3, 2, 4]);
w = reshape(w, [4, 4]);
[U, S, V] = svd2(w); eta = size(S, 1);
U = U * sqrt(S); V = sqrt(S) * V;
U = reshape(U, [2, 2, eta]); U = permute(U, [4, 3, 2, 1]);
V = reshape(V, [eta, 2, 2]); V = permute(V, [1, 4, 3, 2]);
I = reshape(id, [1, 1, 2, 2]);
mpo_even = cell(1, N);
mpo_odd = cell(1, N);
for j = 1:N, mpo_even{j} = I; mpo_odd{j} = I; end
for j = 1:2:(N - 1), mpo_odd{j} = U; mpo_odd{j + 1} = V; end
for j = 2:2:(N - 1), mpo_even{j} = U; mpo_even{j + 1} = V; end

% starting state (one spin flipped)
mps0 = cell(1, N);
for j = 1:N
   if j == jflipped, state = [0; 1]; else state = [1; 0]; end
   mps0{j} = reshape(state, [1, 1, 2]);
end

% time evolution
mps = mps0;
```

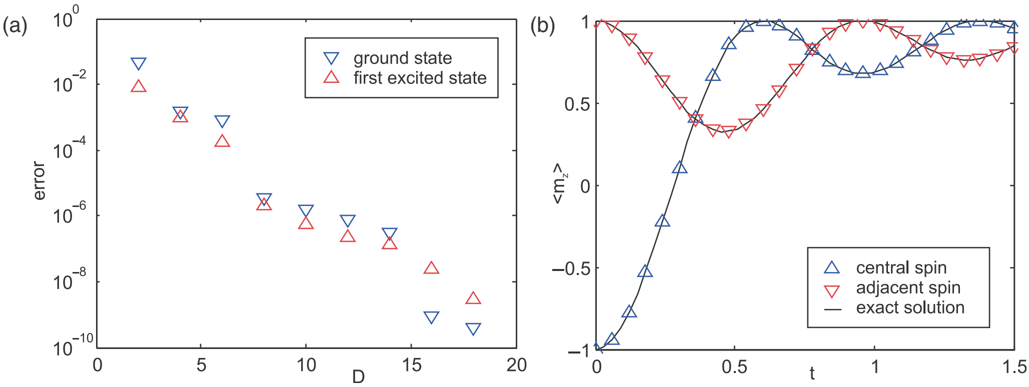

Figure C1. (a) Error of the variational method as a function of the virtual dimension D for $N = 10$ spins. The blue triangles represent the ground state, the red triangles the first excited state. (b) Time evolution as described in the second example. The magnetization of the central spin is represented by blue triangles and the magnetization of the spin adjacent to the central spin by red triangles. For comparison, exact results are also included (black lines). See online version for colour.

```
mzvalues = [];
for step = 1:50
    fprintf(’Step %2d: ’, step);
    [mps, K] = reduceD(mps, mpo_even, D, precision);
    [mps, K] = reduceD(mps, mpo_odd, D, precision);
    mz = expectationvalue(mps, oset);
    mzvalues = [mzvalues, mz];
    fprintf(’mz = %g\n’, mz);
end
```

A comparison of the results produced by these examples to exact calculations is shown in Figure C1. It can be seen that already for moderate values of D the precision is very good.