16 DMRG: Ground States, Time Evolution, and Spectral Functions

Ulrich Schollwöck
Ludwig-Maximilians-Universität München
Theresienstraße 37, 80333 Munich, Germany

Contents

E. Pavarini, E. Koch, and U. Schollwöck
Emergent Phenomena in Correlated Matter
Modeling and Simulation Vol. 3
Forschungszentrum Jülich, 2013, ISBN 978-3-89336-884-6
http://www.cond-mat.de/events/correl13

1 DMRG: A young adult

On November 9, 1992, Physical Review Letters published a paper entitled "Density Matrix Formulation for Quantum Renormalization Groups" by Steven R. White [1]. After introducing a new algorithm for calculating the low-lying states and their observables of one-dimensional lattice models, the paper presented impressively precise results for both the spin-$\frac{1}{2}$ and spin-1 Heisenberg antiferromagnet. It concludes: "This new formulation appears extremely powerful and versatile, and we believe it will become the leading numerical method for 1D systems; and eventually will become useful for higher dimensions as well." I must admit that I am very surprised that these very confident sentences made it past the referees; but this was just as well: the successes of the *density matrix renormalization group* (DMRG) within the last two decades have, if anything, far exceeded the hopes Steve White might have harboured for his new algorithm at that time.

As all youths do, DMRG underwent puberty, developing a completely new personality: around 2004, when it was 12 years old, the (much older) realization [2, 3] that DMRG is closely linked to a special quantum state class, so-called *matrix product states* (MPS), suddenly spawned a number of algorithmic extensions, which drastically enhanced the reach of DMRG (to name but the first few: [4–10]). These algorithmic extensions could all be expressed in both DMRG and MPS language, so many practitioners at first preferred to stay with the old way of speaking. However, at least in my view it is now abundantly clear that formulating DMRG in the language of MPS is notationally much cleaner and conceptually much more adequate, so I will present the entire lecture in this way. The price to pay is to get used to a notation which will definitely be unfamiliar to a physicist with standard training, as opposed to the DMRG language. But it is worth the effort! For those readers who develop a deeper interest, an overview of DMRG in the old language, with some focus on fields of application, is given by [11], whereas a very technical, but hopefully thorough introduction into the structure and manipulation of MPS can be found in [12]. A more conceptual orientation which presents the story as seen by quantum information theory is found in [13].

2 Matrix product states

Let us consider a quantum system that lives on L lattice sites with d local states $\{\sigma_i\}$ on each site $i \in \{1, 2, \ldots, L\}$. A good example to think of would be interacting spins-$\frac{1}{2}$ where the local states are $|\uparrow\rangle, |\downarrow\rangle$ and $d = 2$. The lattice may at this point be of arbitrary dimension, we just have to give an ordering to the site labels. In view of what is to come, one may of course think of a one-dimensional chain with sites 1 through L.

Pure states are then defined on the L^d-dimensional Hilbert space

$$\mathcal{H} = \otimes_{i=1}^{L} \mathcal{H}_i \qquad \mathcal{H}_i = \{|1_i\rangle, \ldots, |d_i\rangle\} \tag{1}$$

and the most general state reads

$$|\psi\rangle = \sum_{\sigma_1,\ldots,\sigma_L} c^{\sigma_1\ldots\sigma_L}|\sigma_1\ldots\sigma_L\rangle. \tag{2}$$

In the following, we will often abbreviate as $\{\sigma\} = \sigma_1\ldots\sigma_L$. The usual problem of numerical simulations is that the number of state coefficients $c^{\{\sigma\}}$ grows exponentially with system size L. A standard first approximation in order to reduce the exponential number of coefficients, the so-called *mean-field approximation*, consists in factorizing the state coefficients as

$$c^{\sigma_1\ldots\sigma_L} = c^{\sigma_1}\cdot c^{\sigma_2}\cdot\ldots\cdot c^{\sigma_L}. \tag{3}$$

Instead of d^L coefficients, we now have dL coefficients (in the special case that we can assume translational invariance of the state, the number reduces even more drastically to d). The motivation of this ansatz dates back to the molecular field theory of Weiss (1907), where the orientation of small elementary magnets (spin was not known at that time) is calculated by assuming they are exposed to an external magnetic field and an additional effective magnetic field, which self-consistently models the interaction with all other elementary magnets. Successful as mean-field theories have been over the decades (the BCS theory of superconductivity (1957), for example, is structurally a mean-field theory of Cooper pairs), they do not capture the essential feature of quantum physics: *entanglement*.

What makes quantum mechanics fundamentally different from classical physics is the different nature of the states: whereas they are points in phase space in classical physics, they are rays in Hilbert space in quantum mechanics; moreover, the Hilbert space of a many-particle system is given by the tensor product of the single-particle Hilbert spaces. To consider the most simple example, take 2 spins-$\frac{1}{2}$. Then $\mathcal{H}_i = \{|\uparrow_i\rangle, |\downarrow_i\rangle\}$ and $\mathcal{H} = \mathcal{H}_1\otimes\mathcal{H}_2$. The combination of the superposition principle (implied by states living in Hilbert space) and the tensor product structure means that the most general state is

$$|\psi\rangle = c^{\uparrow\uparrow}|\uparrow\uparrow\rangle + c^{\uparrow\downarrow}|\uparrow\downarrow\rangle + c^{\downarrow\uparrow}|\downarrow\uparrow\rangle + c^{\downarrow\downarrow}|\downarrow\downarrow\rangle, \tag{4}$$

subject merely to the normalization condition. It is now very easy to show that not every state of this form factorizes, i.e. $c^{\uparrow\downarrow} = c^{\uparrow}c^{\downarrow}$ and so forth; to see this, just consider the singlet state

$$|\psi\rangle = \frac{1}{\sqrt{2}}|\uparrow\downarrow\rangle - \frac{1}{\sqrt{2}}|\downarrow\uparrow\rangle. \tag{5}$$

States that factorize are called *product states*, whereas all others are called *entangled states*. The importance of entangled states is of course given by the fact that they carry non-local information and superclassical correlations, both of which are essential to quantum physics [14–16]. How can we generalize the product state by Eq. (3) to describe (at least certain) entangled states, while remaining numerically convenient? As general states are sums over products of local states, we are led to think about matrices replacing the scalars of Eq. (3); we are then looking at the sum of product states; they can be entangled. So the most simple generalization would be

$$c^{\sigma_1}\cdot c^{\sigma_2}\cdot\ldots\cdot c^{\sigma_L} \to M^{\sigma_1}\cdot M^{\sigma_2}\cdot\ldots\cdot M^{\sigma_L}, \tag{6}$$

where all M^{σ_i} are (2×2)-matrices, except on sites 1 and L, where they must be (1×2) and (2×1) row and column vectors respectively, such that the matrix product yields a scalar. In fact, the famous Affleck-Kennedy-Lieb-Tasaki (AKLT) model has a ground state that can be cast in exactly this form and contains a wealth of non-trivial physics [17,18]. In fact, this simple ansatz can also be used as a variational ansatz for the ground state of entire classes of Hamiltonians giving quite deep insights into their physics; see, for example, [19–21]. Of course, (2×2)-matrices are only of limited descriptive power given the wealth of quantum states.
We therefore consider the following generalization, which is a generic matrix product state:

$$|\psi\rangle = \sum_{\sigma_1,\ldots,\sigma_L} M^{\sigma_1} M^{\sigma_2} \ldots M^{\sigma_L} |\sigma_1 \sigma_2 \ldots \sigma_L\rangle, \tag{7}$$

where at each site we introduce d matrices M^{σ_i}, which therefore depend on the local state $|\sigma_i\rangle$. The dimensions of the matrices are $(1 \times D_1), (D_1 \times D_2), \ldots, (D_{L-2} \times D_{L-1}), (D_{L-1} \times 1)$, with equal row and column indices of matrices associated to neighboring sites such that the matrix product can be carried out, with the very first and last dimension 1, to yield a scalar.
A given state $|\psi\rangle$ does *not* have a unique decomposition into matrices M^{σ_i}: to see this consider an arbitrary, but invertible matrix X of dimension $(D_i \times D_i)$. Then the matrix product state does not change under the insertion of $XX^{-1} = 1$ between matrices M^{σ_i} and $M^{\sigma_{i+1}}$, which implies a gauge transformation

$$M^{\sigma_i} \to M^{\sigma_i} X \qquad M^{\sigma_{i+1}} \to X^{-1} M^{\sigma_{i+1}}. \tag{8}$$

Later on, we will exploit this gauge degree of freedom to bring MPS into a particularly efficient form for practical use.
Why is this state class so interesting? There are, in my view, five reasons for this.

1. Any quantum state can be represented as an MPS, although the representation may be numerically inefficient. Nevertheless, it is therefore a mathematical structure of general interest.

2. There is a hierarchy of MPS in the sense that states with low entanglement can be represented more efficiently (using smaller matrices) than highly entangled states. So-called area laws reveal that these low-entanglement states are particularly important for low-temperature quantum physics. This makes MPS useful in practice. For the link between MPS and the entanglement of quantum states, I refer to the lecture of Jens Eisert.

3. They emerge naturally in the context of renormalization group schemes, connecting the DMRG framework to more conventional RG schemes.

4. They can be manipulated easily and efficiently: this concerns the application of operators ranging from local creation or annihilation operators through Hamiltonians to time evolution operators, the evaluation of overlaps and expectation values.

5. They can be searched efficiently: which state, given $\hat{H}$, has the lowest energy among a given class of MPS? This is the nucleus of a variational method (none other but DMRG).

To prove the first statement, we need either of two matrix decompositions from linear algebra, the QR or the singular value decomposition (SVD); for an excellent first introduction into these techniques of numerical linear algebra, I recommend [22]. In this presentation, I will use the SVD because it will cover all numerical needs we are going to have; however, in numerical practice, the QR factorisation is much faster and should be used whenever possible instead of the SVD (I will indicate these occasions). For an arbitrary $(m \times n)$-matrix A, we have, with $k = \min(m, n)$, the following decomposition

$$A = USV^\dagger, \tag{9}$$

where the matrices have the following special properties:

- U is $(m \times k)$-dimensional and consists of orthonormal columns, i.e., $U^\dagger U = I$; if $m = k$, then $UU^\dagger = I$ too, and U is unitary.
- S is $(k \times k)$-dimensional and diagonal. The entries on the diagonal are called the *singular values* s_i and are real and non-negative, $s_i \geq 0$. The number $r \leq k$ of strictly positive singular values is equal to the rank of A. All texts and computer codes assume that singular values are sorted in descending order, $s_1 \geq s_2 \geq s_2 \geq \ldots$.
- $V^\dagger$ is $(k \times n)$-dimensional and consists of orthonormal rows, i.e., $V^\dagger V = I$; if $k = n$, then $VV^\dagger = I$ too, and V is unitary.

"Old-fashioned" DMRG makes heavy use of the *eigenvalue decomposition (EVD)* $AU = U\Lambda$ for quadratic (and hermitean) A, with Λ a diagonal matrix with real (but not necessarily non-negative) eigenvalues λ_i on the diagonal and U a unitary matrix whose column vectors $|u_i\rangle$ can be taken to be an orthonormal basis. Here, incidentally, I have introduced a very useful notation in linear algebra, namely that matrices are also read as sets of column vectors, $U = [|u_1\rangle|u_2\rangle \ldots]$, or of row vectors (which is used less frequently). SVD and EVD are however closely connected: if $A = USV^\dagger$, then

$$A^\dagger A = VSU^\dagger USV^\dagger = VS^2V^\dagger \Rightarrow (A^\dagger A)V = VS^2 \tag{10}$$

and similarly

$$AA^\dagger = USV^\dagger VSU^\dagger = US^2U^\dagger \Rightarrow (AA^\dagger)U = US^2. \tag{11}$$

Comparing to the EVD, this means that the singular values squared are the eigenvalues of both $A^\dagger A$ and $AA^\dagger$, and the respective eigenvectors are the columns of U and V respectively. This allows to translate between SVD based procedures in MPS-based algorithms and EVD based procedures in classic DMRG.

Before concluding this tiny excursion into the truly rich properties of the SVD, let me mention that it is also behind a very important decomposition of quantum states, the so-called *Schmidt decomposition*. Its key importance is that it allows for a direct readout of the entanglement properties of a state: consider a bipartition of the "universe" AB (in our case: the L sites) into

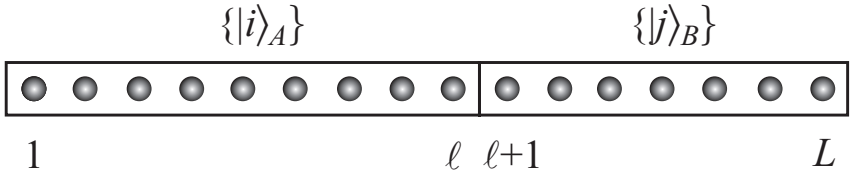

Fig. 1: *Bipartitioning a quantum system in blocks A and B with block-local orthonormal bases.*

subsystems A and B (in our case e.g. sites 1 through ℓ and $\ell+1$ through L); see Fig. 1. Then the most general pure quantum state $|\psi\rangle$ reads

$$|\psi\rangle = \sum_{i=1}^{\dim \mathcal{H}_A} \sum_{j=1}^{\dim \mathcal{H}_B} \psi_{ij} |i\rangle_A |j\rangle_B, \tag{12}$$

where the $\{|i\rangle_A\}$ and $\{|j\rangle_B\}$ form orthonormal bases of subsystems A and B respectively; in practice, these bases will usually be product bases resulting from local orthonormal bases like $\{|\sigma_1 \ldots \sigma_\ell\rangle\}$. Interpreting the ψ_{ij} as the entries of a rectangular matrix Ψ, an SVD of Ψ gives $\Psi = USV^\dagger$ and we can rewrite our state in the *Schmidt decomposed* form

$$|\psi\rangle = \sum_{\alpha=1}^{r} s_\alpha |\alpha\rangle_A |\alpha\rangle_B \tag{13}$$

with s_α the (rank) r non-vanishing singular values of Ψ and

$$|\alpha\rangle_A = \sum_{i=1}^{\dim \mathcal{H}_A} U_{i\alpha} |i\rangle_A \qquad |\alpha\rangle_B = \sum_{j=1}^{\dim \mathcal{H}_B} V^*_{j\alpha} |j\rangle_B \tag{14}$$

It is crucial to note that, due to the properties of U and V, the $\{|\alpha\rangle_A\}$ and $\{|\alpha\rangle_B\}$ form orthonormal sets respectively, because a pitfall of the MPS world is that one encounters state decompositions that look like Eq. (13), but where orthonormality does *not* hold; hence they are not Schmidt decompositions. The physics of the subsystems A and B is encoded by the reduced density operators

$$\hat\rho_A = \mathrm{tr}_B |\psi\rangle\langle\psi| = \sum_{\alpha=1}^{r} s_\alpha^2 |\alpha\rangle_A \ {}_A\langle\alpha| \qquad \hat\rho_B = \mathrm{tr}_A |\psi\rangle\langle\psi| = \sum_{\alpha=1}^{r} s_\alpha^2 |\alpha\rangle_B \ {}_B\langle\alpha| \tag{15}$$

which the Schmidt decomposition allows to be read off in eigenrepresentation. This allows interesting observations like the identity of the non-vanishing eigenvalues of the reduced density operators on A and B, even if A and B are very different: the difference only shows up in the eigenvectors. This, by the way, is a neat illustration of the SVD/EVD link.

If we quantify entanglement by the von Neumann entropy of entanglement, then the entanglement between subsystems A and B given a quantum state $|\psi\rangle$ is obtained by the (conventional) von Neumann entropy of either of the subsystems A or B:

$$S_{A|B}(|\psi\rangle) = -\mathrm{tr}_A \hat\rho_A \ln \hat\rho_A = -\mathrm{tr}_B \hat\rho_B \ln \hat\rho_B = -\sum_{\alpha=1}^{r} s_a^2 \ln s_a^2 \,. \tag{16}$$

We can use this formula and work out the entanglement of the two states already encountered, the general product state and the singlet state of two spins.
In the case of the product state, we consider some bipartitioning into sites 1 through ℓ (subsystem A) and $\ell+1$ through L (subsystem B). We can then write

$$|\psi\rangle = |\alpha\rangle_A |\alpha\rangle_B \quad \text{with} \quad |\alpha\rangle_{A,B} = \sum_{\{\sigma_{A,B}\}} c^{\sigma_{A,B}} |\sigma_{A,B}\rangle, \tag{17}$$

where $\sigma_A \equiv \sigma_1, \ldots, \sigma_\ell$ and similarly σ_B. The state is right away in a Schmidt representation, and the eigenvalue spectrum of the reduced density operators is $(1, 0, 0, \ldots)$; inserting this in the von Neumann entropy formula and using $0 \ln 0 = \lim_{\epsilon \to 0^+} \epsilon \ln \epsilon = 0$, we find that this state is *unentangled*. As the product state is an MPS with matrix dimensions 1, we can state that an MPS is unentangled if and only if it can be represented exactly by an MPS with matrix dimensions 1. (I should add "for a given basis", because entanglement has the unnerving and deeply unphysical property that it is not invariant under global basis transformations: for a simple example, consider the triplet states at $S^z = \pm 1$ which are product states, but which can be rotated by a global basis transformation into the $S^z = 0$ triplet state which is entangled.) The take-home message for us is that, as already mentioned, $D = 1$ MPS are classical (in the sense of unentangled) states and that life gets quantum mechanical (interesting) only from $D = 2$ onwards.
In the case of the singlet state, the reduced density operator obtained by a partial trace is already in diagonal form, $\hat{\rho}_A = \hat{\rho_B} = \mathrm{diag}(\frac{1}{2}, \frac{1}{2})$. Entanglement then is given by $-2 \cdot \frac{1}{2} \ln \frac{1}{2} = \ln 2$. (In statistical physics texts the base of the logarithm is e; in quantum information texts the base of the logarithm is 2 as befits the world of qubits, such that entanglement here would be exactly 1.) The interest of this result is that, using Lagrangian multipliers, one can easily show that this is the maximum entanglement any two-spin state could have and that it is achieved if and only if the reduced density operators are *maximally mixed*, i.e., all eigenvalues are identical. This makes sense as entanglement is a measure of the amount of non-local information in a quantum system: a maximally mixed reduced density operator implies that the amount of local information is minimal. We can generalize the result to a reduced density operator of dimension D: the maximally mixed density operator then has eigenvalues D^{-1}, and the maximal entanglement is given by $-D \cdot D^{-1} \ln D^{-1} = \ln D$.
After this lengthy detour, I am now going to demonstrate that every state can in principle be represented as an MPS. This is done by successively peeling off site after site (say, starting with 1, but the reverse procedure is also possible). Consider $c^{\sigma_1 \sigma_2 \ldots \sigma_L}$, the coefficients of the d^L-dimensional state vector and *reshape* them into a $(d \times d^{L-1})$-dimensional matrix, which is then SV decomposed:

$$c^{\sigma_1 \sigma_2 \ldots \sigma_L} \to \Psi_{\sigma_1, \sigma_2 \ldots \sigma_L} = \sum_{a_1} U_{\sigma_1, a_1} S_{a_1, a_1} V^\dagger_{a_1, \sigma_2 \ldots \sigma_L}. \tag{18}$$

The matrix U is now sliced into d row vectors A^{σ_1}, which we interpret as $(1 \times d)$ matrices (d being the maximum rank possible):

$$U_{\sigma_1, a_1} \to \{A^{\sigma_1}\} \quad \text{with} \quad A^{\sigma_1}_{1, a_1} = U_{\sigma_1, a_1}. \tag{19}$$

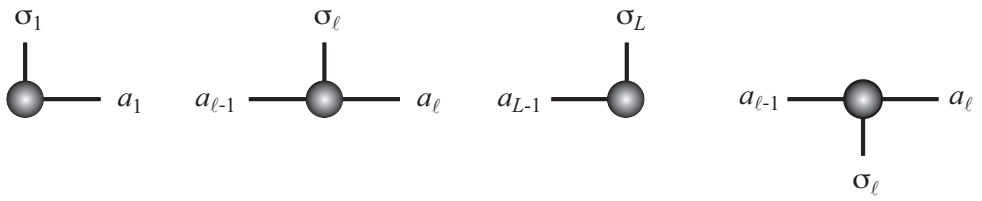

Fig. 2: *Building blocks of an MPS: matrices for first and last sites, as well as bulk sites. Physical indices point vertically, matrix indices horizontally. Complex-conjugation mirrors matrices along the horizontal axis.*

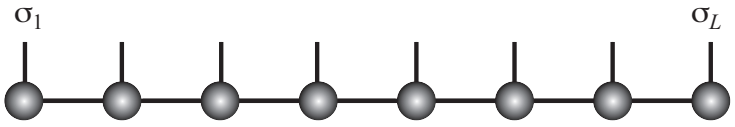

Fig. 3: *Graphical representation of an MPS: all connected lines between building blocks are contracted over.*

If we lump together $c^{a_1\sigma_2\sigma_3\ldots\sigma_L} = S_{a_1,a_1}V^\dagger_{a_1,\sigma_2\ldots\sigma_L}$, we have as a first step

$$c^{\sigma_1\sigma_2\ldots\sigma_L} = \sum_{a_1} A^{\sigma_1}_{1,a_1} c^{a_1\sigma_2\sigma_3\ldots\sigma_L}. \tag{20}$$

In a second step (and all others will be the same), we will again use a sequence of reshaping and a SVD,

$$c^{a_1\sigma_2\sigma_3\ldots\sigma_L} \to \Psi_{a_1\sigma_2,\sigma_3\ldots\sigma_L} = \sum_{a_2} U_{a_1\sigma_2,a_2} S_{a_2,a_2} V^\dagger_{a_2,\sigma_3\ldots\sigma_L}. \tag{21}$$

U is now sliced "horizontally" into d matrices A^{σ_2} of dimension $(d \times d^2)$, where

$$A^{\sigma_2}_{a_1,a_2} = U_{a_1\sigma_2,a_2}, \tag{22}$$

and if S and V are again lumped together to form a new $c^{a_2\sigma_3\sigma_4\ldots\sigma_L}$, the state coefficients read after the second step

$$c^{\sigma_1\sigma_2\ldots\sigma_L} = \sum_{a_1,a_2} A^{\sigma_1}_{1,a_1} A^{\sigma_2}_{a_1,a_2} c^{a_2\sigma_3\sigma_3\ldots\sigma_L}. \tag{23}$$

It is easy to see that upon continuation the coefficients can be represented as a product of matrices as in the definition of an MPS. After half the chain, matrix dimensions which first grow as $(1 \times d)$, $(d \times d^2)$, and so on, will shrink again because in SVD $\min(m, n)$ sets the dimension. The largest matrix dimension is $d^{L/2}$ (L assumed even), hence again exponential in L and it seems nothing has been gained: we still need to approximate, i.e. replace matrices by smaller ones (say, of some maximum dimension $D \sim 1000$) while minimizing the loss of accuracy.
As should be obvious by now, the manipulation of MPS is extremely index-heavy. Fortunately, there is by now a generally accepted graphical representation of MPS (Fig. 2) and their building blocks (up to one degree of freedom, which is nothing compared to the confusing notational wealth surrounding Green's functions or the sign conventions in general relativity). Each matrix is drawn as a dot with two horizontal and one vertical line sticking out; the two horizontal lines

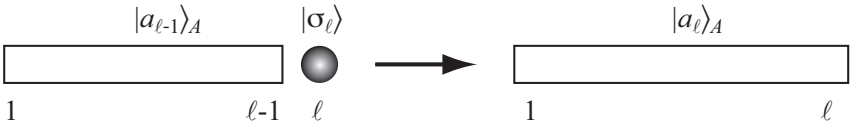

Fig. 4: *Block growth of classic DMRG.*

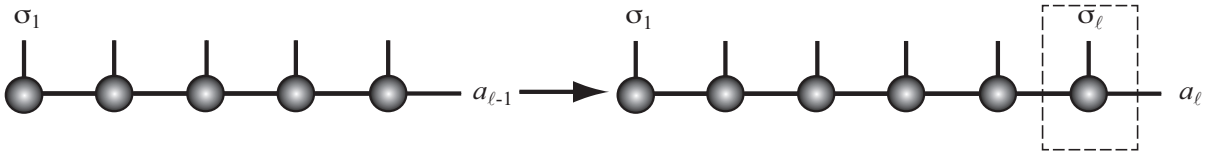

Fig. 5: *Block growth of classic DMRG represented in the MPS framework.*

correspond to the row and column indices and the vertical line to the physical index labeling the matrix. The degree of freedom is whether the vertical line points up or down. I let it point up. If the index only takes the value 1, hence is a dummy index, the corresponding line is dropped (this concerns first and last sites). If one wants to represent the complex-conjugated (not adjoint!) matrix $M^{\sigma *}$, an object we will need for the bra $\langle\psi|$, the direction of the vertical line is inverted, pointing down in my convention. For building an entire state or other more complicated structures, there is a single rule: all connected lines are contracted, i.e., the connected objects are multiplied and the joint index is summed over; for two matrices this is obviously a matrix multiplication. A generic MPS as in Eq. (7) would then look like a comb (Fig. 3).
Some readers of these notes may be familiar with DMRG in the original notation of White, using blocks (of sites) and sites. Let me put this notation into a more general context, that of a generic 1D renormalization scheme: consider a semi-infinite lattice of sites with d local degrees of freedom. If we group the first ℓ sites into a block, a state on the block has d^ℓ coefficients, which is exponentially large. If we decide that we build such blocks *iteratively*, starting from a "block" containing just one site and adding site after site (Fig. 4), then we have to devise a *decimation* scheme of discarding basis states to avoid the exponential growth $d \to d^2 \to d^3 \to \ldots$ and to keep the number of states manageable. If we decide to keep D states for a block description and assume that we have a basis $\{|a_{\ell-1}\rangle\}$ for the block of length $\ell - 1$ and add site ℓ (local states $\{|\sigma_\ell\rangle\}$), then the D states building the (incomplete!) basis of the block of length ℓ will read

$$|a_\ell\rangle = \sum_{a_{\ell-1},\sigma_\ell} \langle a_{\ell-1}, \sigma_\ell | a_\ell\rangle |a_{\ell-1}\rangle|\sigma_\ell\rangle \equiv \sum_{a_{\ell-1},\sigma_\ell} M^{\sigma_\ell}_{a_{\ell-1},a_\ell} |a_{\ell-1}\rangle|\sigma_\ell\rangle, \tag{24}$$

where we have reorganized the expansion coefficients into d matrices M labelled by the local state; the entries are given by $M^{\sigma_\ell}_{a_{\ell-1},a_\ell} = \langle a_{\ell-1}, \sigma_\ell | a_\ell\rangle$ and connect states on the smaller block and the larger block. The matrix entries therefore encode the decimation scheme, which can now be represented as in Fig. 5. The reorganization of the coefficients is of interest because we can iterate the scheme such that

$$|a_\ell\rangle = \sum_{\sigma_1,\ldots,\sigma_\ell} (M^{\sigma_1} M^{\sigma_2} \ldots M^{\sigma_\ell})_{1,a_\ell} |\sigma_1\sigma_2\ldots\sigma_\ell\rangle. \tag{25}$$

Fig. 6: *Graphical representation of left and right normalization conditions (left and right part of the figure). The single lines represent identities.*

The sums over the matrix row and column indices was simply absorbed into a compact matrix multiplication notation, and structurally the basis states look like MPS. As we have specified no decimation procedure, we can draw several conclusions: (i) DMRG is a method that grows blocks using decimation yields states in the MPS format (with subtle modifications that can be ignored here); (ii) any RG scheme in 1D that can be characterized by an iterative growth and decimation scheme yields states in the MPS format; this holds in particular for Wilson's Numerical Renormalization Group (NRG) [23,24] which is the method of choice for the famous Kondo problem. In fact, as a historical remark, the failure of Wilson's NRG for general strongly correlated problems in 1D was what motivated White's work in 1992; the underlying MPS structure finally allowed to understand the connection between NRG and DMRG [25].
I already mentioned in the beginning that there is a gauge degree of freedom in MPS. Both in the general decomposition and in the block growth procedure we have unwittingly chosen a gauge which will turn out to be extremely useful in practice. In the block growth procedure both $\{|a_{\ell-1}\rangle\}$ and $\{|a_\ell\rangle\}$ form orthonormal sets, respectively. Therefore

$$\begin{aligned}\delta_{a'_\ell,a_\ell} &= \langle a'_\ell|a_\ell\rangle = \sum_{a'_{\ell-1}\sigma'_\ell a_{\ell-1}\sigma_\ell} M^{\sigma'_\ell *}_{a'_{\ell-1},a'_\ell} M^{\sigma_\ell}_{a_{\ell-1},a'_\ell} \langle a'_{\ell-1}\sigma'_\ell|a_{\ell-1}\sigma_\ell\rangle \\ &= \sum_{a_{\ell-1}\sigma_\ell} M^{\sigma_\ell *}_{a_{\ell-1},a'_\ell} M^{\sigma_\ell}_{a_{\ell-1},a'_\ell} = \sum_{\sigma_\ell} (M^{\sigma_\ell\dagger} M^{\sigma_\ell})_{a'_\ell,a_\ell}\end{aligned}$$

or

$$I = \sum_{\sigma_\ell} M^{\sigma_\ell\dagger} M^{\sigma_\ell} \equiv \sum_{\sigma_\ell} A^{\sigma_\ell\dagger} A^{\sigma_\ell}. \tag{26}$$

Matrices that obey this relationship are called *left-normalized* and will be denominated by A; exactly the same property follows from the general state decomposition from the column orthonormality $U^\dagger U = I$. If one builds blocks from the right, adding sites at the left end, or if one carries out an SVD on a general state starting from the right, i.e. at site L, one obtains similarly *right-normalized* matrices B with

$$I = \sum_{\sigma_\ell} B^{\sigma_\ell} B^{\sigma_\ell\dagger}. \tag{27}$$

I will refer to MPS that consist entirely of A-matrices as left-canonical, MPS that consist entirely of B-matrices I will call *right-canonical*. A third class, which is the most important one in numerical practice, is called *mixed-canonical* and has structure $AAAAAMBBBBBBBBBB$, with one matrix without special normalization property sandwiched in between. We will discuss further below how to convert between all three representations. Graphically, the normalization conditions are represented as in Fig. 6.

3 Matrix product operators

Given that quantum mechanics works by applying operators (unitary operators for time evolution, projection operators in measurements, ...) to states, it is quite surprising that the systematic representation of operators in a generalization of the MPS scheme did not really start until quite recently (see e.g. [26]). The most general operator on our L sites reads

$$\hat{O} = \sum_{\{\sigma\}} \sum_{\{\sigma'\}} c^{\sigma_1 \ldots \sigma_L, \sigma'_1 \ldots \sigma'_L} |\sigma_1 \ldots \sigma_L\rangle\langle\sigma'_1 \ldots \sigma'_L|, \tag{28}$$

where the primed variables label the ingoing state the operator acts on and the unprimed variables the outgoing state. If we reshuffle the indices to group states on the same site, we have

$$c^{\sigma_1 \ldots \sigma_L, \sigma'_1 \ldots \sigma'_L} \rightarrow c^{\sigma_1 \sigma'_1 \sigma_2 \sigma'_2 \ldots \sigma_L \sigma'_L} \tag{29}$$

and a "mean-field approximation" to the operator would read

$$c^{\sigma_1 \sigma'_1 \sigma_2 \sigma'_2 \ldots \sigma_L \sigma'_L} \rightarrow c^{\sigma_1 \sigma'_1} \cdot c^{\sigma_2 \sigma'_2} \cdot \ldots \cdot c^{\sigma_L \sigma'_L}. \tag{30}$$

While there is no physical reason why this should be in any ways a good approximation, in fact this is an *exact* representation of many operators: consider the operator $\hat{S}^z_i$ acting on a spin on site i. As this operator is in fact defined only on $\mathcal{H}_i$, a pedantic notation would be

$$\hat{S}^z_i \rightarrow \hat{I}_1 \otimes \hat{I}_2 \otimes \ldots \otimes \hat{S}^z_i \otimes \ldots \otimes \hat{I}_L \tag{31}$$

and the coefficients would read

$$c^{\sigma_1 \sigma'_1 \sigma_2 \sigma'_2 \ldots \sigma_L \sigma'_L} = \delta_{\sigma_1, \sigma'_1} \cdot \delta_{\sigma_2, \sigma'_2} \cdot \ldots \cdot (\hat{S}^z)_{\sigma_i, \sigma'_i} \cdot \ldots \cdot \delta_{\sigma_L, \sigma'_L}. \tag{32}$$

So, if we introduce *matrix product operators* (MPO) as a straightforward generalization of the MPS notation as

$$\hat{O} = \sum_{\{\sigma\}} \sum_{\{\sigma'\}} M^{\sigma_1 \sigma'_1} M^{\sigma_2 \sigma'_2} \ldots M^{\sigma_L \sigma'_L} |\sigma_1 \ldots \sigma_L\rangle\langle\sigma'_1 \ldots \sigma'_L| \tag{33}$$

with the usual rules for the matrix dimensions, the above operator would be encoded simply by scalar (i.e., $D = 1$) matrices, and this would also hold for operators as used for n-point correlators such as $\hat{S}^+_i \hat{S}^-_j$. Any more complicated operator can also be turned into an MPO: reconsider the construction of an MPS for an arbitrary state, and group indices $(\sigma_1 \sigma'_1)(\sigma_2 \sigma'_2) \ldots$. Of course, the question arises whether this neat result will generalize in the sense that more complicated operators like a Hamiltonian $\hat{H}$ still find an *exact* representation with *small* D. As opposed to MPS of interest which will usually involve an approximation, we will find this to be true for many Hamiltonians of interest.

What happens if we apply an MPO $\hat{O}$ to an MPS $|\psi\rangle$? Assuming that the MPS is formed by matrices M^{σ_i} and the MPO by matrices $N^{\sigma_i \sigma'_i}$ it is a simple calculation to show that the new state $\hat{O}|\psi\rangle$ is again an MPS, with matrices $\tilde{M}^{\sigma_i}$ that have entries

$$\tilde{M}^{\sigma_i}_{(ab),(a'b')} = \sum_{\sigma'_i} N^{\sigma_i \sigma'_i}_{aa'} M^{\sigma'_i}_{bb'}. \tag{34}$$

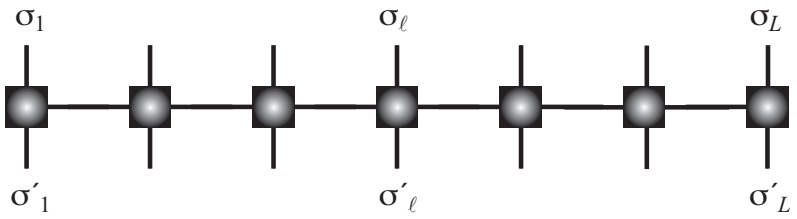

Fig. 7: *Graphical representation of an MPO: the vertical lines sticking out to the bottom represent ingoing physical states, the lines sticking out to the top outgoing physical states.*

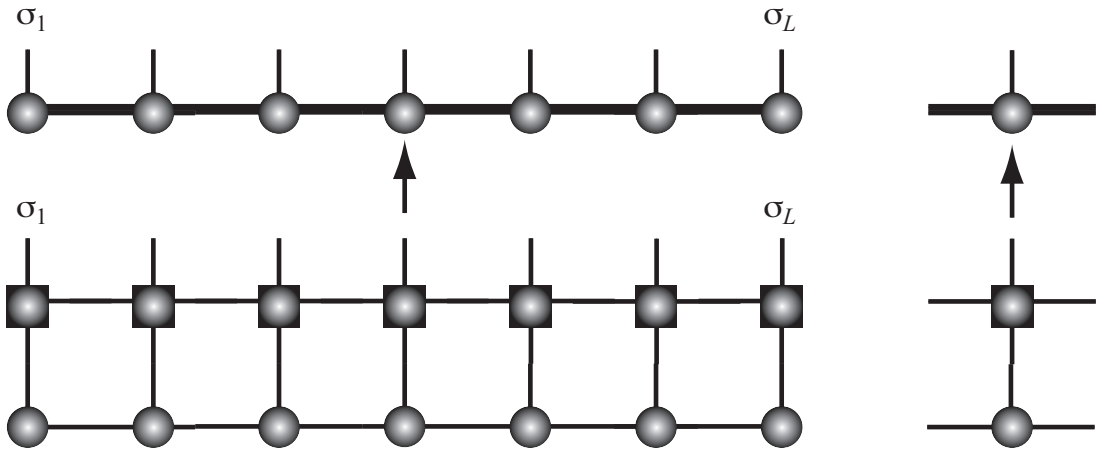

Fig. 8: *Graphical representation of the application of an MPO to an MPS: all connected lines between building blocks are contracted over. A new MPS results, with matrix dimensions being the product of the original matrix dimensions.*

Numerically, this operation can be implemented very efficiently; the important observation is that the dimensions of the new matrices are given by the *product* of the old matrix dimensions. This is potentially disastrous, as at least MPS dimensions can become very large for some desired quality of approximation. But as most Hamiltonians in MPO form have dimensions less than 10, the new MPS may still be barely manageable; nevertheless this observation indicates that we need a *compression procedure* such that an MPS with undesirably large (unmanageable) matrix dimensions can be approximated optimally by an MPS with smaller matrix dimensions at some loss of accuracy.

Graphically, an MPO (Fig. 7) is represented in analogy to an MPS, with *two* vertical legs sticking out, corresponding to the "ingoing" (line pointing down) and "outgoing" state (line pointing up). The MPO times MPS gives MPS rule finds a simple graphical representation as in Fig. 8. Lines sticking out in the same direction can be unified into a single one, with the rule that the dimension of the unified line is the product of the original dimensions.

4 Normalization and compression

As we have seen, the dimensions of MPS will usually grow as we proceed with calculations, making them potentially useless if we do not counter this growth. How can we compress an MPS with minimal loss of information? Let us assume (don't worry that the assumption looks a bit artificial) that we have a quantum state in the following MPS representation,

$$|\psi\rangle = \sum_{\{\sigma\}} A^{\sigma_1} A^{\sigma_2} \dots A^{\sigma_\ell} M^{\sigma_{\ell+1}} B^{\sigma_{\ell+2}} \dots B^{\sigma_L} |\sigma_1 \dots \sigma_L\rangle \tag{35}$$

and the row and column dimensions of $M^{\sigma_{\ell+1}}$ (and the adjacent column dimension of A^{σ_ℓ} and row dimension of $B^{\sigma_{\ell+2}}$) are too big and we want to reduce them with minimum of loss of accuracy in the state description. To achieve this, we stack the d matrices $M^{\sigma_{\ell+1}}$ columnwise into a single matrix, i.e.

$$M_{a_\ell, \sigma_{\ell+1} a_{\ell+1}} = M^{\sigma_{\ell+1}}_{a_\ell, a_{\ell+1}} \tag{36}$$

and carry out an SVD on the new matrix as $M = USV^\dagger$. If we absorb U it into $A^{\sigma_\ell} \leftarrow A^{\sigma_\ell} U$, this retains the left-normalization due to $U^\dagger U = I$ and corresponds to a basis transformation of the (reduced) orthonormal block basis for block A formed from sites 1 through ℓ,

$$|a_\ell\rangle_A := \sum_{\sigma_1, \dots, \sigma_\ell} (A^{\sigma_1} \dots A^{\sigma_\ell})_{1, a_\ell} |\sigma_1 \dots \sigma_\ell\rangle. \tag{37}$$

Similarly, we have a new orthonormal block basis on the block formed by sites $\ell + 1$ through L by slicing as

$$B^{\sigma_{\ell+1}}_{a_\ell, a_{\ell+1}} = V^\dagger_{a_\ell, \sigma_{\ell+1} a_{\ell+1}}, \tag{38}$$

where $B^{\sigma_{\ell+1}}$ are right-normalized and

$$|a_\ell\rangle_B := \sum_{\sigma_{\ell+1}, \dots, \sigma_L} (B^{\sigma_{\ell+1}} \dots B^{\sigma_L})_{a_\ell, 1} |\sigma_{\ell+1} \dots \sigma_L\rangle \tag{39}$$

form an orthonormal set. Identifying $s_{a_\ell} = S_{a_\ell, a_\ell}$, we therefore have a Schmidt decomposition of $|\psi\rangle$ as

$$|\psi\rangle = \sum_{a_\ell} s_{a_\ell} |a_\ell\rangle_A |a_\ell\rangle_B. \tag{40}$$

Hence (and this is why the mixed-canonical representation is so important, ensuring orthonormal bases on A and B) a mixed-state representation can be turned into a Schmidt decomposition and vice versa. The Schmidt decomposition, on the other hand, allows to read off the correct compression strategy: the s_a^2 corresponding to the statistical weights in the reduced density operators, the optimal truncation is given by *retaining those pairs of Schmidt states that have the maximum Schmidt coefficients*. If we can afford matrix dimensions D, and assume ordering by descending singular values, we simply cut down the column dimension of A^{σ_ℓ} to D and similarly the row dimension of $B^{\sigma_{\ell+1}}$, and both dimensions of S.

The problem with this procedure is that it only works at the seam between the left-normalized and right-normalized parts of the state, but in general we have to truncate everywhere. Our

prodecure immediately indicates how to remedy this: if (after truncation) we multiply S to the left, $M^{\sigma_\ell} \leftarrow A^{\sigma_\ell} S$, the matrices on site ℓ will lose their normalization property and the state will have the form

$$|\psi\rangle = \sum_{\{\sigma\}} A^{\sigma_1} A^{\sigma_2} \dots A^{\sigma_{\ell-1}} M^{\sigma_\ell} B^{\sigma_{\ell+1}} \dots B^{\sigma_L} |\sigma_1 \dots \sigma_L\rangle. \tag{41}$$

Compared to our initial state, the seam between left- and right-normalized states has shifted by one to the left, and we can continue our truncation procedure now one site to the left. This, of course, would also have been the case if we had not truncated at all.

This allows us to define two strategies: in order to bring *any* MPS into form $AAAA\dots$ or $BBBB\dots$, we start either from site 1 (for A) or site L (for B) and work our way through the chain by a sequence of SVDs *without* truncations (in such a case, faster QR decompositions do as well). In the previous paragraphs, we have just seen how a step to the left generates a B. The fact that we had A-matrices to the left there only mattered for the truncation. Bringing any MPS into left-canonical or right-canonical form I refer to as *normalization* (indeed, as one can see easily, in the very last step of the procedure, a scalar survives, which is nothing but the norm of the state, so one can use it as a "conventional" normalization procedure $\langle\psi|\psi\rangle \stackrel{!}{=} 1$ as well). Partially right-normalizing a left-normalized MPS generates a mixed-normalized MPS.

As a second strategy for compressing an MPS to some acceptable matrix dimension, we take a state that is e.g. in form $AAAAA\dots$ (which can be achieved by the first strategy) and move through all mixed-canonical representations from the right, truncating along the way. Here, the QR factorization, which does not give access to the singular values, is not useful.

At this point it is now easy to see for which types of states MPS can yield good approximations: the quality of truncations depends on how quickly the singular values s_a of the Schmidt decompositions (or the statistical weights of the reduced density operators) decay with a: if they decay rapidly, the truncated statistical weight is negligible, and in practice truncated weigths for ground states are often of the order 10^{-10} or even less if we keep matrix dimensions $D \sim 1000$. Usually, we do not know the spectrum of the reduced density operators; as entanglement is derived from this spectrum, it contains similar (but much less) information: a state with a rapidly decaying spectrum has lower entanglement than a state with slowly decaying spectrum. Hence, the feasability of the representation of a state by an MPS rests on its entanglement properties, MPS being a low-entanglement representation. For more details I refer to Jens Eisert's lecture in this volume.

Looking more closely, one realizes that this compression scheme cannot be absolutely optimal: while it is optimal at each step, the compression at some site depends on the outcome of earlier compressions on other sites, but not on the compression that will happen later. Hence there is an informational asymmetry, which usually is not so important because the amount of compression at each site is small, as just mentioned. In cases where this is a problem, there exists an alternative variational technique which is strictly optimal, see [12].

5 Time-evolution: tDMRG, TEBD, tMPS

Traditionally, the exposition of DMRG starts with explaining the ground state algorithm, DMRG proper, which for a given Hamiltonian $\hat{H}$ looks for its ground state within, as it turns out, the space of MPS. This reflects the historical course of events: time-dependent DMRG [4–6, 8] (with the variants of TEBD and tMPS, but this is all very much the same) was invented 12 years after ground state DMRG. We will reverse this sequence, as time-dependent DMRG is much easier to explain and also implement *and* can be used, albeit in a quite inefficient way, to find the ground state of a given Hamiltonian.

We restrict our attention to time-independent Hamiltonians $\hat{H}$; this captures a large number of the problems encountered in practice. As all more important time-evolution schemes currently in use consider small ("infinitesimal") time steps, time-dependent Hamiltonians can be modeled by a sequence of Hamiltonians that change after each small time step.

Assume we have an initial state $|\psi(0)\rangle$ in MPS form; such a state can be constructed by hand (in simple cases like a Néel state, which is just a $D = 1$ MPS) or is obtained by some other MPS calculation, e.g. as the ground state of some (other) Hamiltonian (otherwise there would be no non-trivial dynamics) – this is the typical setup in ultra cold atom experiments where non-equilibrium dynamics is generated by Hamiltonian quenches, i.e. abrupt changes in Hamiltonian parameters. In the case of coherent evolution, the state at time t is given by

$$|\psi(t)\rangle = \mathrm{e}^{-\mathrm{i}\hat{H}t}|\psi(0)\rangle. \tag{42}$$

If we manage to give $\mathrm{e}^{-\mathrm{i}\hat{H}t}$ an MPO representation, the problem would be solved within the MPS framework, as applying an MPO to an MPS yields a new MPS. There are several issues with this idea: (i) no one knows how to do this exactly in an efficient form on a classical computer, let alone a piece of paper; (ii) usually we are interested in the entire evolution, i.e. the state for an entire sequence of times; (iii) if the dimension of the resulting MPO is too large, the resulting MPS will not be numerically manageable. While problems (i) and (ii) can be resolved at some cost in accuracy, we will see (iii) to be a fundamental issue.

One approach, that is also historically the first, is to *Trotterize* time evolution. This is a well-known analytical scheme first used in quantum field theory, and then first applied in numerical physics by Suzuki [27] to quantum Monte Carlo schemes. Assume we want to calculate $|\psi(t)\rangle$ for times in $[0, T]$. We split interval length T into $N \to \infty$ time steps $\tau \to 0$ with $N\tau = T$. In numerical practice, τ will of course be solidly finite, e.g. $\tau = 0.01$ when a typical time scale of the problem is 1. Let us also assume that our Hamiltonian consists only of at most nearest-neighbor terms, like $\mathbf{S}_i \cdot \mathbf{S}_{i+1}$ in the Heisenberg model or the hopping term $\sum_\sigma (\hat{c}^\dagger_{i\sigma}\hat{c}_{i+1\sigma} + \mathrm{h.c.})$ in the Hubbard model. Then we can split the Hamiltonian in nearest-neighbour terms, $\hat{H} = \sum_{i=1}^{L-1} \hat{h}_i$, where in the case of the Heisenberg model $\hat{h}_i = \mathbf{S}_i \cdot \mathbf{S}_{i+1}$. On-site terms like the Hubbard interaction are "distributed" across the two $\hat{h}$ that share a site and counted only half, e.g. $U\hat{n}_{i\uparrow}\hat{n}_{i\downarrow}$ enters as $(U/2)\hat{n}_{i\uparrow}\hat{n}_{i\downarrow}$ in both $\hat{h}_{i-1}$ and $\hat{h}_i$. Watch out for first and last sites!

Then we can rewrite the evolution operator as

$$\mathrm{e}^{-\mathrm{i}\hat{H}T} = \prod_{i=1}^{N} \mathrm{e}^{-\mathrm{i}\hat{H}\tau} = \prod_{k=1}^{N} \mathrm{e}^{-\mathrm{i}\sum_{i=1}^{L-1}\hat{h}_i\tau} \stackrel{!}{=} \prod_{k=1}^{N}\prod_{i=1}^{L-1} \mathrm{e}^{-\mathrm{i}\hat{h}_i\tau}. \tag{43}$$

Not only do we now have access to all times $\tau, 2\tau, \ldots$, we now also only have to calculate the (infinitesimal) time-evolution on *two sites*, $\mathrm{e}^{-\mathrm{i}\hat{h}_i\tau}$. This is a $(d^2 \times d^2)$ matrix obtained by diagonalizing the $(d^2 \times d^2)$ matrix representation H_i of $\hat{h}_i$ and exponentiating it,

$$H_i U = U\Lambda \;\Rightarrow\; H_i = U\Lambda U^\dagger \;\Rightarrow\; \mathrm{e}^{-\mathrm{i}H_i\tau} = U\mathrm{e}^{-\mathrm{i}\Lambda\tau}U^\dagger = U\cdot\mathrm{diag}(\mathrm{e}^{-\mathrm{i}\lambda_1\tau}, \mathrm{e}^{-\mathrm{i}\lambda_2\tau}, \ldots)\cdot U^\dagger, \tag{44}$$

which is easily implemented using standard diagonalization and matrix multiplication routines. So the fundamental building block can be calculated, but there is a catch, indicated by the exclamation mark on the last identity in (43). This factorization does *not* work, due to Glauber's formula,

$$\mathrm{e}^{\hat{A}+\hat{B}} = \mathrm{e}^{\hat{A}}\mathrm{e}^{\hat{B}}\mathrm{e}^{\frac{1}{2}[\hat{A},\hat{B}]}, \tag{45}$$

which only holds under some restrictions on $\hat{A}$ and $\hat{B}$ (there may be further terms), but the decisive point is that if two operators do not commute, the exponential of their sum will *not* factorize in general. As our operators $\hat{A}$ and $\hat{B}$ scale as τ, however, in our case the commutator will scale with τ^2. Hence, in the limit $\tau \to 0$, factorization becomes exact; for finite τ, the error will scale as τ^2 and can therefore be excellently extrapolated to the exact $\tau \to 0$ limit. This decomposition of the evolution operator is called the *first-order Trotter decomposition*, because after $N = T/\tau$ time steps, the accumulated error will be of order τ. Higher-order decompositions are available (fourth-order decompositions being the most popular, with an error of order τ^4; for details, I refer to the literature), but for our purpose of explaining the method first-order serves perfectly well.

As bond evolution operators commute if they do not share a site, it is customary to split the Hamiltonians into odd and even bonds as

$$\hat{H} = \hat{H}_{\mathrm{odd}} + \hat{H}_{\mathrm{even}}; \qquad \hat{H}_{\mathrm{odd}} = \sum_i \hat{h}_{2i-1}, \quad \hat{H}_{\mathrm{even}} = \sum_i \hat{h}_{2i} \tag{46}$$

and arrange time evolution as

$$\mathrm{e}^{-\mathrm{i}\hat{H}T} = \mathrm{e}^{-\mathrm{i}\hat{H}_{\mathrm{even}}\tau}\mathrm{e}^{-\mathrm{i}\hat{H}_{\mathrm{odd}}\tau}; \qquad \mathrm{e}^{-\mathrm{i}\hat{H}_{\mathrm{even}}\tau} = \prod_i \mathrm{e}^{-\mathrm{i}\hat{h}_{2i}\tau}, \quad \mathrm{e}^{-\mathrm{i}\hat{H}_{\mathrm{odd}}\tau} = \prod_i \mathrm{e}^{-\mathrm{i}\hat{h}_{2i-1}\tau}\,. \tag{47}$$

Assuming we know the two-site time-evolution MPOs, we can represent the Trotter-decomposed MPO for the global time evolution as in Fig. 9.

The two-site time-evolution MPO can easily be derived by our SVD decomposition procedure developed for MPS. It is easy to calculate $U^{\sigma_1\sigma_2,\sigma_1'\sigma_2'} = \langle\sigma_1\sigma_2|\mathrm{e}^{-\mathrm{i}\hat{h}_1\tau}|\sigma_1'\sigma_2'\rangle$. Then we proceed by reshuffling of indices and SVD as

$$U^{\sigma_1\sigma_2,\sigma_1'\sigma_2'} = \overline{U}_{\sigma_1\sigma_1',\sigma_2\sigma_2'} \stackrel{SVD}{=} \sum_b W_{\sigma_1\sigma_1',b} S_{b,b} W_{b,\sigma_2\sigma_2'} = \sum_b M_{1,b}^{\sigma_1\sigma_1'} M_{b,1}^{\sigma_2\sigma_2'}$$

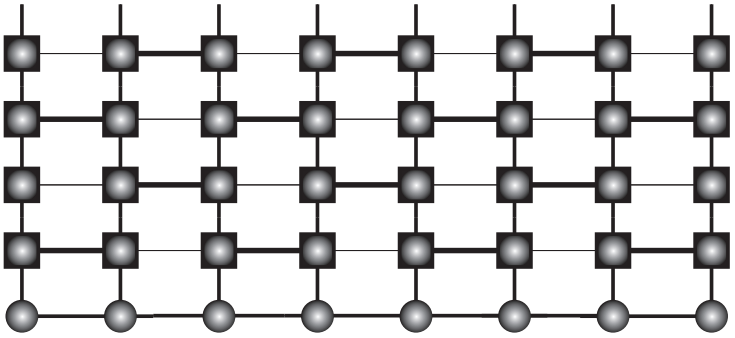

Fig. 9: *MPO representation of the Trotterized time-evolution operator* $\mathrm{e}^{-\mathrm{i}\hat{H}T}$*; every second line corresponds to an odd-bond or even-bond infinitesimal time-evolution; at the bottom, there is the state that is evolved in time contracted into the MPO.*

which is just an MPO where I have absorbed two factors of $\sqrt{S_{b,b}}$ into the left and right matrices. As the original matrix is of dimension $(d^2 \times d^2)$, the MPO dimension is $D_W = d^2$; more precisely, the matrices for odd bond evolutions have dimensions $(1 \times d^2), (d^2 \times 1), (1 \times d^2), \ldots$ and vice versa for even bonds. After one Trotter step, MPS matrix dimensions will therefore have grown everywhere by a factor d^2. This means that after each time step, we will have to compress the MPS, and the time evolution algorithm takes a very simple form:

1. Apply infinitesimal Trotter time step in MPO form to MPS to obtain $|\psi(t)\rangle \to |\psi(t+\tau)\rangle$
2. Compress $|\psi(t+\tau)\rangle$
3. Continue with the next time step

In principle, it looks as if this procedure could be continued forever. The Trotter decomposition error, which is $O(\tau^2)$ per time step will accumulate as $t/\tau O(\tau^2)$, hence grow linearly in time. It can be made arbitrarily small by sending $\tau \to \infty$ or, more efficiently, using higher-order Trotter decompositions where the error will scale as $O(\tau^3)$ or even $O(\tau^5)$ in the most frequently used approaches. The problem rests in the compression; compression as such is not a problem if we can describe the state $|\psi(t+\tau)\rangle$ with the same accuracy as $|\psi(t)\rangle$ by an MPS of given dimension D. This is however not the case: building on the Lieb-Robinson theorem [28], it has been observed that entanglement in a time-evolving quantum state will grow up to linearly in time, $S(t) \leq S(0) + \nu t$, with some constant ν, and that this linear bound is met in "global quenches" where the non-equilibrium is generated by a sudden change of the Hamiltonian operator. As we have seen, a reduced density operator of dimension D (hence an MPS of dimension D) can encode at most entanglement $S = \ln D$; hence we have to increase matrix sizes in the MPS at worst *exponentially* as time evolves to maintain the same accuracy. Time-dependent simulations therefore hit an exponential wall after some time. No one has found a solution for that problem yet. While observing thermalization is therefore excluded, in many situations of interest, ν is small enough that we can see the physics of interest; as a few examples, consider spin-charge separation [29], the propagation of bosonic correlations in ultracold atom gases [30], the

relaxation of a density wave in strongly interacting ultracold atoms [31, 32], but there are many more.

Let me conclude this section by briefly discussing the calculation of ground states using *imaginary* time evolution. Quite generally, starting from a random state $|\psi\rangle = \sum_n c_n |n\rangle$, with eigenstates $\hat{H}|n\rangle = E_n|n\rangle$, $E_0 \leq E_1 \leq E_2 \leq \ldots$,

$$\begin{aligned} \lim_{\beta\to\infty} \mathrm{e}^{-\beta\hat{H}}|\psi\rangle &= \lim_{\beta\to\infty} \sum_n \mathrm{e}^{-\beta E_n} c_n |n\rangle = \lim_{\beta\to\infty} \mathrm{e}^{-\beta E_0}(c_0|0\rangle + \sum_{n>0} \mathrm{e}^{-\beta(E_n-E_0)} c_n |n\rangle \\ &= \lim_{\beta\to\infty} \mathrm{e}^{-\beta E_0} c_0 |0\rangle, \end{aligned}$$

where for simplicity I have also assumed that the ground state is non-degenerate. We see that if the random starting state has some overlap with the ground state, this contribution will survive longest in the $\beta \to \infty$ limit. Of course, except for the unlikely case that $E_0 = 0$, this surviving contribution will either diverge or decay exponentially. But as we have to truncate after each infinitesimal (imaginary) time step anyways, we can use this occasion to normalize the state at every step, yielding the ground state $|0\rangle$ in the large-β limit.

Numerically, imaginary time evolution is more benign than real time evolution, as any errors made numerically are exponentially suppressed by further applications of $\mathrm{e}^{-\tau\hat{H}}$, whereas it is a hallmark of a unitary (real time) evolution that errors are with us to stay and will only be compounded with further errors. However, the procedure is slow compared to a direct variational ground state search in MPS space, which is classic DMRG, and will be discussed below.

6 Overlaps and expectation values

Calculating time-evolving states as such is of course of little interest; in the end, we want to calculate how much a state is changing from the original state (i.e. the overlap $\langle\psi(t)|\psi(0)\rangle$) or, even more frequently, how some observable evolves in time, e.g. $\langle S_i^z(t)\rangle = \langle\psi(t)|\hat{S}_i^z|\psi(t)\rangle$. Let us first focus on the overlap, because the latter calculation will be a simple generalization. Mathematically, representing $|\psi\rangle$ by matrices M and $|\phi\rangle$ by matrices $\tilde{M}$, we have

$$\langle\phi|\psi\rangle = \sum_{\{\sigma\}} \sum_{\{\sigma'\}} \langle\{\sigma'\}|\tilde{M}^{\sigma'_1 *} \ldots \tilde{M}^{\sigma'_L *} M^{\sigma_1} \ldots M^{\sigma_L}|\{\sigma\}\rangle = \sum_{\{\sigma\}} \tilde{M}^{\sigma_1 *} \ldots \tilde{M}^{\sigma_L *} M^{\sigma_1} \ldots M^{\sigma_L}. \tag{48}$$

Graphically, this can be represented as in Fig. 10. In which order are we to carry out the contractions? At first sight, this may seem a detail, but it is not: imagine we contract all the horizontal lines first. Each contraction is a matrix-matrix multiplication that costs $O(D^3)$ assuming matrix size D. Hence there are $O(2LD^3)$ operations (we ignore edge effects and other details) for *each* of the d^L product basis configurations, and the overall operation count is $O(2LD^3d^L)$ which is *exponentially large* in system size even if we truncate MPS matrices. This would conclude the story, but there is an exponential speedup: contractions are carried out moving through the network from the left to the right, adding one matrix after another into the contracted part. The operational count is now $O((2L-1)D^3d)$, which is linear in L and only weakly polynomial in

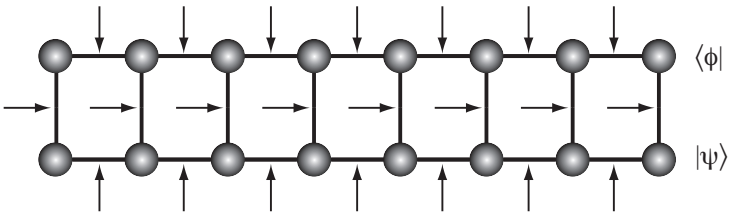

Fig. 10: *Contraction scheme for the overlap of two MPS. All arrows point to contraction points.*

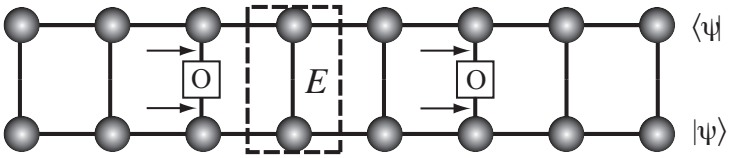

Fig. 11: *Contraction scheme for a two-point correlator, indicating the transfer operator E.*

D. This can best be seen by rearranging the overlap equation as

$$\begin{aligned}
\langle\phi|\psi\rangle &= \sum_{\{\sigma\}} \tilde{M}^{\sigma_1 *} \dots \tilde{M}^{\sigma_L *} M^{\sigma_1} \dots M^{\sigma_L} \\
&= \sum_{\{\sigma\}} \tilde{M}^{\sigma_L \dagger} \dots \tilde{M}^{\sigma_1 \dagger} M^{\sigma_1} \dots M^{\sigma_L} \\
&= \sum_{\sigma_L} \tilde{M}^{\sigma_L \dagger} \left(\dots \left(\sum_{\sigma_2} \tilde{M}^{\sigma_2 \dagger} \left(\sum_{\sigma_1} \tilde{M}^{\sigma_1 \dagger} M^{\sigma_1} \right) M^{\sigma_2} \right) \dots \right) M^{\sigma_L} .
\end{aligned}$$

If one works out the contractions from inside out, the first bracket costs $O(dD^3)$, ignoring that the first and last matrices are in reality vectors. The result is again a matrix, so the next bracket is essentially a product of three matrices (carried out as two matrix-matrix multiplications) and the cost is $O(2dD^3)$. Overall, we obtain the contraction count announced. This is the optimal scheme, and the contraction is exact; I am mentioning this because in the higher-dimensional generalizations of MPS contractions are not exact unless exponential complexity is accepted (which is of course impossible) and finding the optimal contraction scheme is NP hard (i.e. impossible to determine in practice).

Evaluating expectation values of operators is now very easy: assume we want to evaluate the expectation value of $\hat{O}$ acting on site i, we just have to replace $\sum_{\sigma_i}$ by $\sum_{\sigma_i,\sigma_i'} O_{\sigma_i,\sigma_i'}$, a double sum over the matrix elements of $\hat{O}$ in the local basis, information which is trivially available. So the computational cost is hardly growing at all, and graphically, one would replace the vertical line at site i by the operator. If we consider the object

$$E^{(a_{\ell-1}a'_{\ell-1}),(a_\ell,a'_\ell)} := \sum_{\sigma_\ell} A^{\sigma_\ell *}_{a_{\ell-1},a_\ell} A^{\sigma_\ell}_{a'_{\ell-1},a'_\ell}, \tag{49}$$

which is shown in Fig. 11, one can work out analytically or see graphically that it acts as a transfer matrix in the calculation of correlators, determining the decay of correlators by its eigenvalues. As is typical for transfer matrix calculations, all correlators are either long-ranged

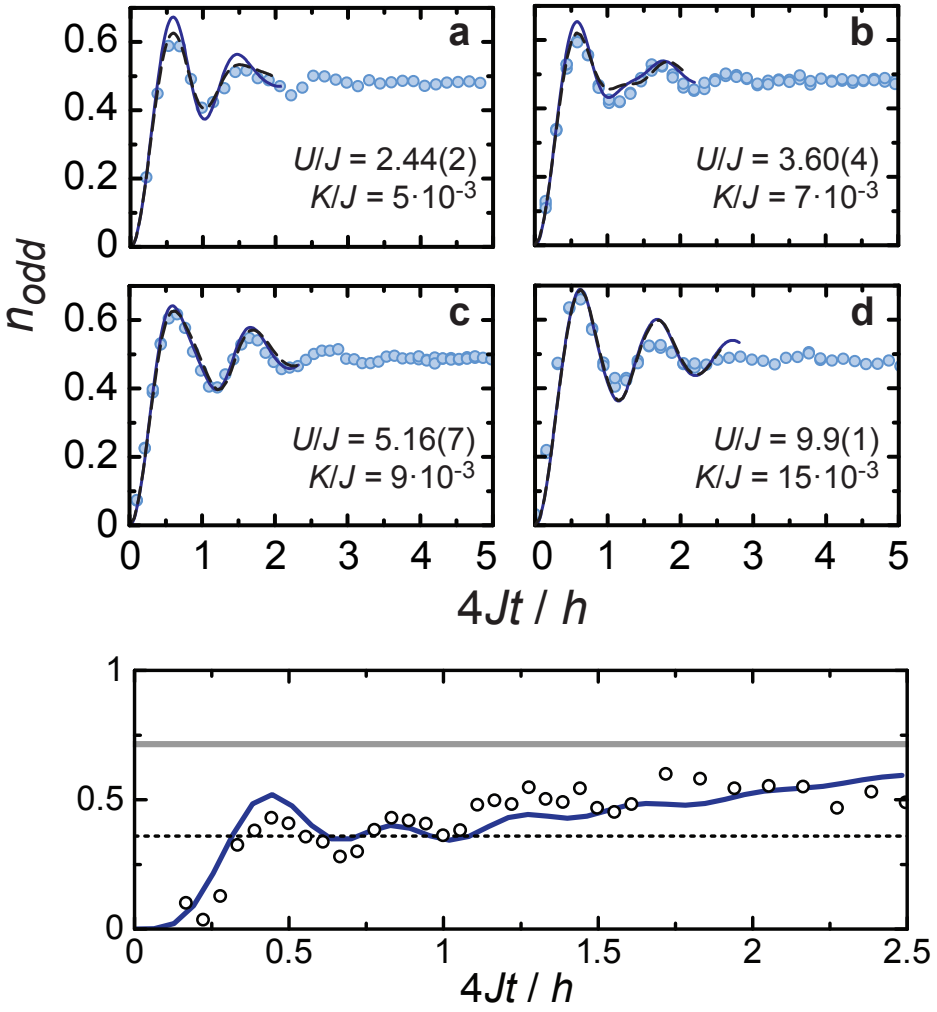

Fig. 12: *Comparison between tDMRG simulations (lines) and experiment (points). In both cases, the starting wave function is a density wave in an optical lattice with one boson on even and no bosons on odd sites, that evolves according to a Bose-Hubbard-Hamiltonian with hopping amplitude J and on-site repulsion U. The upper panel shows how the density relaxes from 0 to 0.5 for various interaction strengths. For $U/J = 9.9$ one-dimensionality is partially lost, leading to disagreement. The lower panel shows $4\mathrm{Re}\langle b_i^\dagger b_{i+1}\rangle$ as a function of time: nearest-neighbour quantum correlations are built up over time by the relaxation of the density wave. Taken from [32].*

or decay as a superposition of exponentials. Critical power-law decays are not possible for MPS. Therefore, for a quantum state that has critical correlations, the MPS will approximate the power law decay by a superposition of exponentials, but eventually switch to an exponential decay on very long length scales, when only the slowest exponential decay in the superposition survives. Increasing D, the MPS will model the critical decay on increasing length scales.

As an example of the accuracy achieved nowadays both by experiment and theory for non-trivial strongly interacting systems, using time-dependent DMRG, consider the comparisons shown in Fig. 12 between tDMRG and a state-of-the-art quantum simulator using ultracold atoms in an optical lattice [32].

7 Finite-temperature simulations

So far, we have been considering pure-state calculations only; quantum mechanics generally deals with mixed states that are represented by (reduced) density operators $\hat{\rho}$, which are Hermitean, non-negative, and normalized (trace 1). In the early days of MPS-formulations of DMRG, it seemed advantageous to find a MPS-type *state* representation of $\hat{\rho}$ in order to "recycle" existing methods. This is indeed possible, using the concept of purification, first used in this context by [8], which is the Schmidt decomposition read backwards: given a density operator living on physical space P expressed in eigen-representation

$$\hat{\rho}_P = \sum_n \rho_n |n\rangle_P \ {}_P\langle n|, \tag{50}$$

we can interpret it as the reduced density operator on P for a state $|\psi\rangle_{PQ}$ living on a larger space PQ, where Q must be at least as large as P (in practice, simply a copy of P: instead of a spin chain, we simulate a spin ladder, etc.), and where the Schmidt decomposed form of $|\psi\rangle_{PQ}$ is

$$|\psi\rangle_{PQ} = \sum_n \sqrt{\rho_n} |n\rangle_P |n\rangle_Q \,. \tag{51}$$

The choice of the $|n\rangle_Q$ living on the auxiliary space Q is free as long as they form an orthonormal set; this gives a gauge degree of freedom to $|\psi\rangle_{PQ}$, the purification of $\hat{\rho}_P$

$$\hat{\rho}_P = \mathrm{tr}_Q |\psi\rangle_{PQ} \ {}_{PQ}\langle\psi| \,. \tag{52}$$

Provided we know the purification, all further calculations can be done in a pure state framework on PQ. Expectation values are (normalization of $\hat{\rho}$ and $|\psi\rangle$ imply each other)

$$\langle \hat{O}_P \rangle_{\hat{\rho}_P} = \mathrm{tr}_P \hat{O}_P \hat{\rho}_P = \mathrm{tr}_P \hat{O}_P \mathrm{tr}_Q |\psi\rangle_{PQ\,PQ}\langle\psi| = \mathrm{tr}_{PQ} \hat{O}_P |\psi\rangle_{PQ\,PQ}\langle\psi| = {}_{PQ}\langle\psi|\hat{O}_P|\psi\rangle_{PQ}, \tag{53}$$

where we have used the cyclicity of the trace. Time evolution becomes

$$\hat{\rho}_P(t) = \mathrm{e}^{-\mathrm{i}\hat{H}t}\, \hat{\rho}_P\, \mathrm{e}^{+\mathrm{i}\hat{H}t} = \mathrm{e}^{-\mathrm{i}\hat{H}t}\, \mathrm{tr}_Q |\psi\rangle_{PQ\,PQ}\langle\psi| \mathrm{e}^{+\mathrm{i}\hat{H}t} = \mathrm{tr}_Q |\psi(t)\rangle_{PQ\,PQ}\langle\psi(t)| \tag{54}$$

with $|\psi(t)\rangle_{PQ} = \mathrm{e}^{-\mathrm{i}\hat{H}t}|\psi\rangle_{PQ}$. Hence we simply have to carry out a standard time-evolution of a pure state. The problem is, of course, that usually we do not know the eigenrepresentation of $\hat{\rho}_P$. But for the most important application, the thermal state $\mathrm{e}^{-\beta\hat{H}}$, it can be calculated easily: $\mathrm{e}^{-\beta\hat{H}} = \mathrm{e}^{-\beta\hat{H}/2} \cdot \hat{I}_P \cdot \mathrm{e}^{-\beta\hat{H}/2} = \mathrm{tr}_Q \mathrm{e}^{-\beta\hat{H}/2} |\rho_0\rangle_{PQ\,PQ}\langle\rho_0| \mathrm{e}^{-\beta\hat{H}/2}$, where $|\rho_0\rangle_{PQ}$ is the purification of $\hat{\rho}_P(\beta = 0) = \hat{I}_P$ (up to an irrelevant normalization). The infinite temperature ($\beta = 0$) reduced density operator factorizes between individual sites, and so does the purified state between pairs of sites in P and Q associated with each other. But the purification of the totally mixed state $\hat{\rho}_P(\beta = 0)$ on a site is nothing but a maximally entangled state between a site in P and the corresponding site in Q, for example the singlet state if the physical site contains a spin-$\frac{1}{2}$. Now this state can be encoded easily as an MPS on one pair of sites, and as the MPS factorizes between pairs of sites, the matrices have dimension 1 and can be written down by

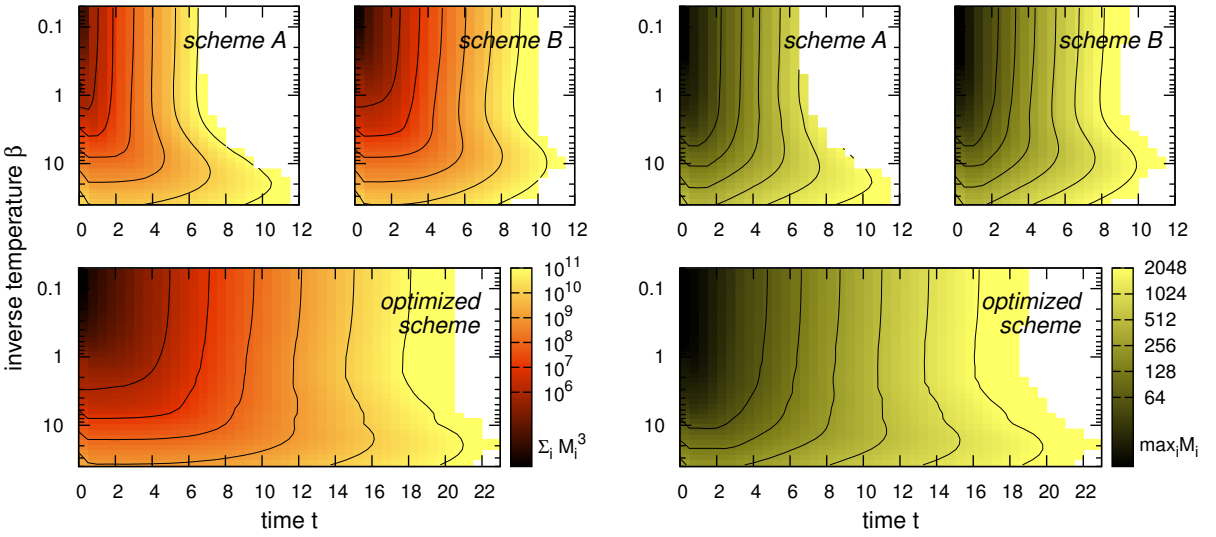

Fig. 13: *Reachable times for the isotropic Heisenberg model with $\hat{B}^\dagger = \hat{A} = \hat{S}^\pm$ at the chain center: Schemes A, B are the conventional scheme [8] and the improved scheme [34] compared to the new scheme of [36]. The left panels show computational cost (proportional to the sum of the third power of the matrix dimensions), the right panels the maximal matrix dimension. Similar greyscales correspond to similar usage of resources. Adapted from [36].*

hand. $|\rho_\beta\rangle_{PQ}$, a purification of $\hat{\rho}_P(\beta)$, is then obtained by an imaginary time evolution up to $\beta/2$. For time-dependent results obtained in this way and enhanced by a prediction technique, consider [33]. However, it was observed that times reached are relatively short.

Let us consider this in more detail. If we subject a thermal state to a real-time evolution using the same Hamiltonian, it remains unchanged as $[\hat{\rho}_P(\beta), \hat{H}] = 0$. In the purification approach, however, resource usage grows, because we are time-evolving the purified state which is not an eigenstate of $\hat{H}$, and the above reasoning for entanglement growth applies. We are therefore moving through increasing costly purifications of the same density operator. As there is a gauge freedom in the auxiliary state space, it might be possible to counteract this resource growth by a suitable transformation of auxiliary states. Indeed, it was found that evolving the auxiliary system backward in time, using $\mathrm{e}^{+\mathrm{i}\hat{H}t}$, resources will not grow [34], and this increase in efficiency also pays off for time-evolutions once a local operator has been applied. Substantial further improvement [35, 36] (and a better understanding of the procedure of [34]) is possible by exploiting the isomorphism between local bounded operators $\hat{B} : \mathcal{H} \mapsto \mathcal{H}$ and states $|\psi\rangle$ on $\mathcal{H} \otimes \mathcal{H}$. Our purified (MPS) state therefore has an (MPO) operator interpretation on $\mathcal{H}$, and we can discard the notion of purification entirely. In fact, in my view, this is the formulation to use in the future, the old purification approach having its usefulness because it allows recycling of tDMRG, whereas in the new approach, MPO manipulation routines are needed. The currently best approach found there is given as

$$\langle \hat{B}(2t)\hat{A}\rangle_\beta = Z(\beta)^{-1}\mathrm{tr}\left([\mathrm{e}^{\mathrm{i}\hat{H}t}\mathrm{e}^{-\beta\hat{H}/2}\hat{B}\mathrm{e}^{-\mathrm{i}\hat{H}t}][\mathrm{e}^{-\mathrm{i}\hat{H}t}\hat{A}\mathrm{e}^{-\beta\hat{H}/2}\mathrm{e}^{\mathrm{i}\hat{H}t}]\right), \tag{55}$$

where the objects in $[,]$ are treated as MPOs that are constructed by Trotterization and compression just as in normal time-evolution. An auxiliary space Q is not needed at all.

8 Ground states with MPS: DMRG

We have already seen that an imaginary time evolution is capable of yielding a ground state. The classic technique, which is more efficient, is by a variational minimisation within MPS space, and is what is traditionally referred to as DMRG, which is a variational minimisation technique in MPS space and hence sometimes, but rarely, also referred to as VMPS. For a given allowed maximum MPS matrix dimension D, the best MPS approximation to the ground state energy (and the ground state) is given by

$$\min \frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle} \quad \Leftrightarrow \quad \min \Big(\langle\psi|\hat{H}|\psi\rangle - \lambda\langle\psi|\psi\rangle \Big) , \tag{56}$$

where we have introduced a Lagrangian multiplier λ to enforce normalization $\langle\psi|\psi\rangle = 1$ that will give the variational approximation to the ground state energy E_0 from above (because we are in a restricted search space). If we increase D, search space is enlarged (as it contains all MPS of smaller dimension), and the approximate energy will decrease monotonically, allowing for an extrapolation in $D \to \infty$, which is the exact limit.
Assuming we know the MPO representation of $\hat{H}$, which I postpone for a moment, we have to minimise the network represented in Fig. 14. This is a multilinear problem, as states $|\psi\rangle$ are products of the unknown M^σ, for which no immediately efficient strategy is known. A time-proven approach to the problem is the *alternating least square (ALS) method*, which runs as follows:

1. Start with a guess for the MPS extremizing energy, $\{M^{\sigma_i}\}$.

2. Pick a site i, $1 \leq i \leq L$, and consider all matrices M^{σ_j}, $j \neq i$, fixed and retain only M^{σ_i} as variables. Eq. (56) is then *quadratic* in M^{σ_i}, and minimization becomes a linear problem, leading to new M^{σ_i} minimizing energy within the "framework" provided by the other matrices.

3. Now pick another site as i, repeating step 2, until all sites have been visited often enough that λ does not decrease anymore. The resulting MPS is the variationally best approximation to the ground state, as is λ to the ground state energy.

Let us elaborate on these steps, because an efficient implementation strongly rests on details here. The starting guess can be random or be constructed by an iterative growth of a chain of length $2 \to 4 \to \ldots \to L$, where at each step the chain is grown by inserting 2 sites at

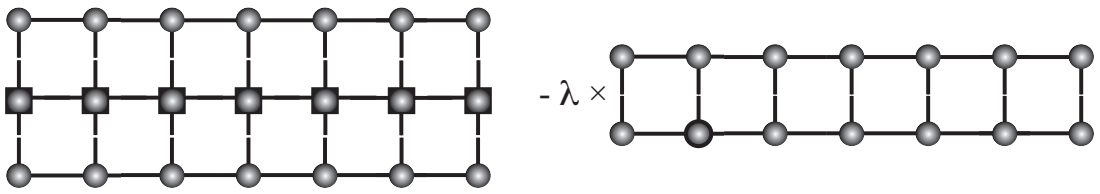

Fig. 14: *Network to be contracted to extremize* $\langle\psi|\hat{H}|\psi\rangle - \lambda\langle\psi|\psi\rangle$.

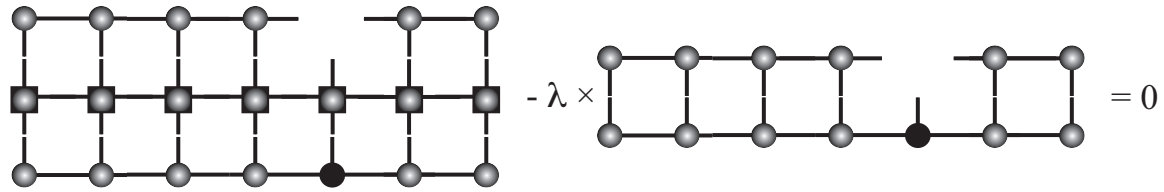

Fig. 15: *Network to be contracted to extremize $\langle\psi|\hat{H}|\psi\rangle - \lambda\langle\psi|\psi\rangle$. The unknown (variable) matrix M^{σ_i} is shown in black.*

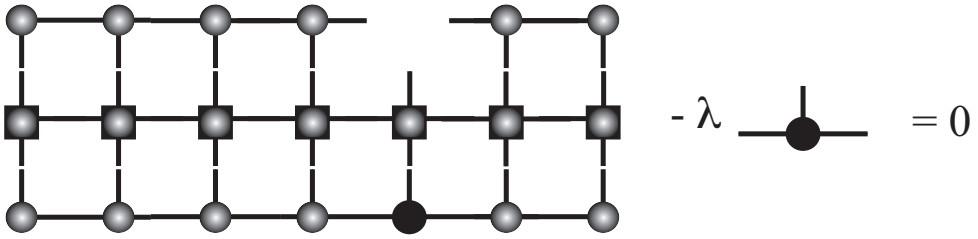

Fig. 16: *Eigenvalue problem to be solved for extremizing $\langle\psi|\hat{H}|\psi\rangle - \lambda\langle\psi|\psi\rangle$ provided the state $|\psi\rangle$ is in adequate mixed canonical form. Again, the unknown (variable) matrix M^{σ_i} is shown in black.*

the center, and determining the two sets of M^σ for these 2 sites such that energy is minimized while keeping all previously found matrices fixed. This relates to the block plus site growth strategy exposed earlier, and is what traditionally is referred to as "infinite system DMRG". Historically, this was even considered the core of the method, whereas it is rather the warm-up to the "real" algorithm! This statement has to be qualified to the extent that for translationally invariant systems the "warm-up" can be turned into a ground state search algorithm in its own right under the name of iDMRG [12, 37].
For carrying out the minimization step, we take the derivative

$$\frac{\partial}{\partial M^{\sigma_i *}} \Big(\langle\psi|\hat{H}|\psi\rangle - \lambda\langle\psi|\psi\rangle \Big) \stackrel{!}{=} 0. \tag{57}$$

Graphically, this corresponds to Fig. 15, where the matrix with respect to which we take the derivative is simply removed, because it contributed linearly to the original network. If we now contract the two remaining networks, we have three free legs each and three legs connecting to the unknown M^{σ_i}. We can rewrite this as

$$\sum_{\sigma_i' a_{i-1}' a_i'} H_{\sigma_i a_{i-1} a_i, \sigma_i' a_{i-1}' a_i'} M_{\sigma_i' a_{i-1}' a_i'} = \sum_{\sigma_i' a_{i-1}' a_i'} N_{a_{i-1} a_i, a_{i-1}' a_i'} \delta_{\sigma_i, \sigma_i'} M_{\sigma_i' a_{i-1}' a_i'} \tag{58}$$

$$\equiv \sum_{\sigma_i' a_{i-1}' a_i'} N_{\sigma_i a_{i-1} a_i, \sigma_i' a_{i-1}' a_i'} M_{\sigma_i' a_{i-1}' a_i'}, \tag{59}$$

where the primed variables run over the legs connecting to M^{σ_i} and the unprimed variables run over the free legs. H and N are the contracted networks, and for convenience I have brought all indices down. The resulting equation has the form of a generalized eigenvalue problem,

$$H\mathbf{m} = \lambda N \mathbf{m}, \tag{60}$$

where the matrices are $(dD^2 \times dD^2)$ dimensional and the vector $\mathbf{m}$ (from the reshaped matrices M^{σ_i}) is dD^2-dimensional. As we are looking for the ground state, we have to find the lowest eigenvalue λ. With $D \sim 1000$ in practice, this cannot be found by full diagonalization techniques, but we can use *large sparse eigensolvers* such as provided by the Lanczos method (for a first introduction, see [22]).

Generalized eigenvalue problems can be potentially dangerous, if the condition number of matrix N becomes large. But if we assume that all matrices in $|\psi\rangle$ to the left of M^{σ_i} are left-normalized, $AAAA\ldots$, and all matrices to the right of it are right-normalized, $\ldots BBBB$, the normalization conditions imply that the entire network on the right collapses and only M^{σ_i} remains. Then $N = I$, and a simple eigenvalue problem $H\mathbf{m} = \lambda\mathbf{m}$ remains. In order to achieve this convenient situation, we simply take our starting state, bring it by (partial) normalization into form $AAAAAAMBBBBB$ (wherever we want to start) and replace random choices of locations i by systematic "sweeping" through the chain from right to left to right to left and so on (or inversely), carrying out one suitable normalization step at each iteration moving left or right to keep the seam between the left- and right-normalized matrices moving along with i. This sweeping procedure is exactly the finite-system DMRG invented by White in 1992 and the concurrent normalization procedure is nothing but White's prediction algorithm [38].

A few more remarks are in order (for details, see [12]): contracting the left network (building H) might in principle be very costly, but in practice one keeps three separate parts, namely the part of H to the left of M^{σ_i}, the one on top (the part of the MPO), and the part of H to the right of M^{σ_i}. If we sweep left and right, these parts can be recycled or updated iteratively from previous steps, drastically reducing the numerical cost. At the same time, original DMRG considered two sites for optimization at once, i.e. pairs $M^{\sigma_i}M^{\sigma_{i+1}}$. Looking more closely at how this is done, one realizes that this slightly breaks the variational nature of the method and is a bit more costly numerically (by a factor of d); on the other hand, it is less plagued by the problem of our "single-site" method that it may sometimes get stuck in a non-global minimum (being non-variational can take you out of such dead ends). But this issue can be resolved elegantly by adding suitable "noise" to the procedure [12, 39].

How can we control accuracy? Extrapolating in $D \to \infty$ is fine in principle (energies go down monotonically, due to the variational nature of the procedure, other operators also converge monotonically as a rule of thumb), but if MPS make sense it is because the Schmidt coefficients decay rapidly, usually exponentially fast. But then we expect some kind of exponential convergence in D which however is not so clean as to allow an easy extrapolation. It is easier to monitor, as a function of D, the variation of the energy,

$$\langle\psi|\hat{H}^2|\psi\rangle - (\langle\psi|\hat{H}|\psi\rangle)^2, \tag{61}$$

which can be evaluated without further approximation as the Hamiltonian MPO is exact, and to extrapolate quantities in the variation, which is 0 for $D \to \infty$ ($|\psi\rangle$ will be an exact eigenstate).

9 Constructing the MPO representation of a Hamiltonian

So far we have dodged the issue of constructing the MPO corresponding to a Hamiltonian. In fact, for short-ranged Hamiltonians this is quite easy, and as the construction is best understood by looking at an example, I will consider the Heisenberg model

$$\hat{H} = J \sum_{i=1}^{L-1} \left(\frac{1}{2}(\hat{S}_i^+ \hat{S}_{i+1}^- + \hat{S}_i^- \hat{S}_{i+1}^+) + \hat{S}_i^z \hat{S}_{i+1}^z \right) + h \sum_{i=1}^{L} \hat{S}_i^z. \tag{62}$$

This consists of operator strings of the type $\hat{I} \otimes \hat{I} \otimes \hat{S}^+ \otimes \hat{S}^- \otimes \hat{I} \otimes \hat{I} \ldots$. To simplify the notation, I will introduce *operator valued* matrices in MPOs, namely

$$\hat{M}^{[i]} = \sum_{\sigma_i, \sigma_i'} M^{\sigma_i, \sigma_i'} |\sigma_i\rangle\langle\sigma_i'|. \tag{63}$$

Then the Hamiltonian will take the form $\hat{H} = \hat{M}^{[1]} \hat{M}^{[2]} \ldots \hat{M}^{[L]}$. Let us imagine the construction of the Hamiltonian as the action of an automaton which has internal states (not to be confused with quantum states of our system). It starts from the right end of the chain in some internal state and moves through it to the left end. The action of the automaton is shown in Fig. 17. The automaton starts in internal state 1, acts, and once it has passed site 1, it should be in some final internal state and have produced exactly all terms that contribute to the Hamiltonian. This can be achieved as follows: we associate internal state 1 with "no non-trivial operator to the right". Being in state 1, the automaton has five options at a site: adding another $\hat{I}$ to the operator string (staying in 1), adding a $\hat{S^+}$ term (moving to state 2), a $\hat{S^z}$ term (moving to state 3), or a $\hat{S^-}$ term (moving to state 4). In any of those cases, the automaton now must add the term that completes the interaction at the next site, this is a $(J/2)\hat{S^-}$ term, a $J\hat{S^z}$ term, or a $(J/2)\hat{S^+}$ term. In any case, it moves into state 5, which corresponds to "completed interaction to the right". At the same time, it can also move directly from state 1 to 5 by introducing a $h\hat{S^z}$ field term.

The action of the automaton can now be represented in matrix form where the row and column indices correspond to the outgoing and ingoing internal state of the automaton at each step. Taking into account that at site 1 the automaton must end in state 5 and that at site L it must start in state 1, the matrices read

$$\hat{M}^{[i]} = \begin{bmatrix} \hat{I} & 0 & 0 & 0 & 0 \\ \hat{S}^+ & 0 & 0 & 0 & 0 \\ \hat{S}^z & 0 & 0 & 0 & 0 \\ \hat{S}^- & 0 & 0 & 0 & 0 \\ h\hat{S}^z & (J/2)\hat{S}^- & J^z\hat{S}^z & (J/2)\hat{S}^+ & \hat{I} \end{bmatrix} \tag{64}$$

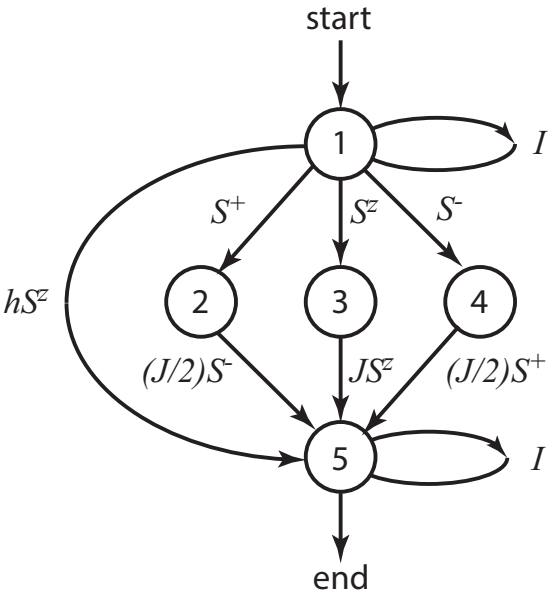

Fig. 17: *States of the automaton constructing the Hamiltonian MPO for the Heisenberg model.*

and on the first and last sites

$$\hat{M}^{[1]} = \begin{bmatrix} h\hat{S}^z & (J/2)\hat{S}^- & J^z\hat{S}^z & (J/2)\hat{S}^+ & \hat{I} \end{bmatrix} \qquad \hat{M}^{[L]} = \begin{bmatrix} \hat{I} \\ \hat{S}^+ \\ \hat{S}^z \\ \hat{S}^- \\ h\hat{S}^z \end{bmatrix} . \tag{65}$$

We can identify $\tilde{D} = 5$ as the dimension of the Hamilton MPO which is given by the above matrices. Similarly, other Hamiltonians can be constructed; for longer-ranged interactions, we have to introduce further internal states such that the automaton can keep track of when to complete an interaction. For a more elaborate discussion, see [12] and references therein.

10 Dynamical DMRG

Dynamical DMRG is the denomination for methods for the calculation of frequency-dependent Green's functions and spectral functions, hence the name which is not to be confused with *time-dependent* DMRG for the calculation of out-of-equilibrium dynamics.
The fundamental object under study are objects such as

$$S_\eta(\omega) := \left\langle 0 \left| \hat{O}^\dagger \frac{1}{E_0 + \omega + \mathrm{i}\eta - \hat{H}} \hat{O} \right| 0 \right\rangle , \tag{66}$$

where $\eta = 0^+$ (numerically, it will be a small, but finite positive number) and $|0\rangle$ is the ground state of $\hat{H}$. This is nothing but the Fourier transform (with a numerical convergence factor $\mathrm{e}^{-\eta t}$) of the $T = 0$ Green's function

$$G_O(t) = \langle 0|\hat{O}^\dagger(t)\hat{O}(0)|0\rangle \qquad (t > 0) \tag{67}$$

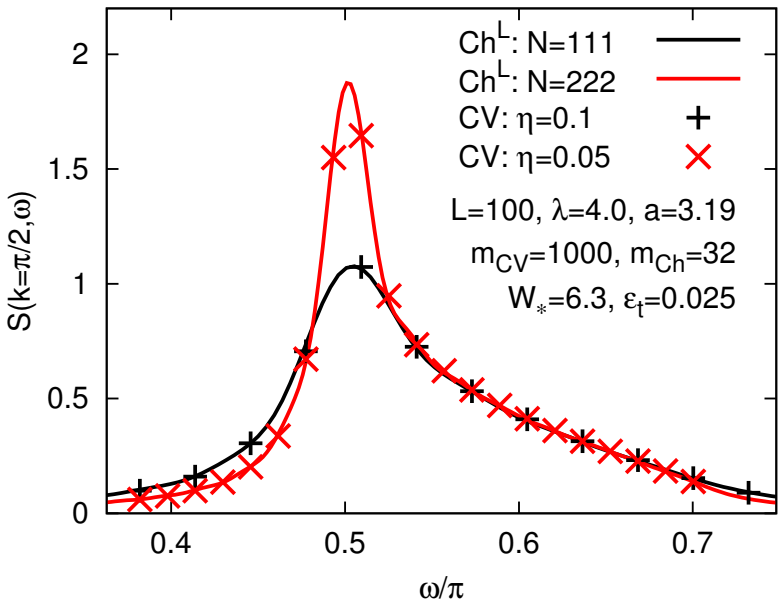

Fig. 18: *Comparison between correction vector (CV) and Chebyshev vector (Ch) results for the $k = \frac{\pi}{2}$ spectral function of the isotropic Heisenberg antiferromagnet. Agreement is excellent, but the Chebyshev results are obtained orders of magnitude faster. From [45].*

where the Heisenberg picture is assumed. For available $|0\rangle$, various techniques have been proposed which are not restricted to MPS, but can be efficiently implemented there.

The fastest (but often least precise) technique is the *continuous fraction* or Lanczos approach. This approach was pioneered in [40], first used in the classic DMRG context by [41], but can be made much more precise if expressed in MPS language [42]. Starting from $|q_1\rangle = \hat{O}|0\rangle/\|\hat{O}|0\rangle\|$, it generates a sequence of orthonormal Krylov vectors (also known as Lanczos vectors) $|q_m\rangle$ as

$$\beta_m|q_{m+1}\rangle = \hat{H}|q_m\rangle - \alpha_m|q_m\rangle - \beta_{m-1}|q_{m-1}\rangle\,, \tag{68}$$

where the α_m and β_m (calculated by normalizing $|q_{m+1}\rangle$) are the diagonal and off-diagonal elements of a tridiagonal matrix representation of $\hat{H}$: $\alpha_m = \langle q_m|\hat{H}|q_m\rangle$ and $\beta_m = \langle q_{m+1}|\hat{H}|q_m\rangle$. Then it can be shown by matrix inversion that

$$S_\eta(\omega) = \cfrac{\langle 0|\hat{O}^\dagger\hat{O}|0\rangle}{E + \mathrm{i}\eta - \alpha_1 - \cfrac{\beta_1^2}{E + \mathrm{i}\eta - \alpha_2 - \cfrac{\beta_2^2}{E + \mathrm{i}\eta - \alpha_3 - \ldots}}}. \tag{69}$$

The attractive feature of this procedure is that the generation of Krylov vectors is usually available for free as the necessary algorithm already is part of the Lanczos large sparse matrix diagonalization routine needed in the ground state search.

The current gold standard, but also slowest, approach is the *correction vector* approach [43,44]. One introduces the so-called correction vector

$$|c\rangle = \frac{1}{E_0 + \omega + \mathrm{i}\eta - \hat{H}}\hat{O}\,|0\rangle, \tag{70}$$

such that the Green's function reads $G_\eta(\omega) = \langle 0|\hat{O}^\dagger|c\rangle$. $|c\rangle$ is determined by solving the large sparse equation system

$$(E_0 + \omega + \mathrm{i}\eta - \hat{H})|c\rangle = \hat{O}|0\rangle, \tag{71}$$

for which reliable, but quite costly techniques such as GMRES are available [22]. Note that it is not a good idea to "square" the system as originally proposed in order to make it hermitian, because the new condition number is now the square of the old one, drastically slowing down convergence. Taking the limit $\eta \to 0$ also implies strongly increasing D of the correction vector MPS to maintain the desired accuracy, for reasons not fully understood at the moment.

More recently, a new technique was proposed which expands the Green's function in terms of *Chebyshev polynomials* [45]; I refer to the literature as it is not yet a widely used standard method; but from Fig. 18 it should be clear that it is a serious contender for the correction vector method, because it achieves similar accuracy orders of magnitude faster. – It remains, however, to be seen whether this advantageous scenario continues to hold for more complex systems.

Concluding this section, let me mention that both at $T = 0$ and $T > 0$ (where the techniques just discussed do not apply) frequency-dependent can be successfully obtained by a combination of time-dependent DMRG for the calculation of two-time correlators and subsequent Fourier transformation (for $T = 0$, see [6], for $T > 0$ see [33]). Frequency resolution is limited by the finite range in time of tDMRG which necessitates some exponential damping of the real-time data.

11 Outlook: DMRG in two dimensions

MPS and DMRG are obviously best suited to the study of one-dimensional quantum systems. However, there has been enormous interest in the notoriously elusive physics of two-dimensional quantum systems for decades, ranging from frustrated magnets to high-T_c superconductors, with a recent surge of interest in the physics of topological quantum spin liquids. Therefore it is not surprising that there have been numerous attempts to apply DMRG (MPS) to

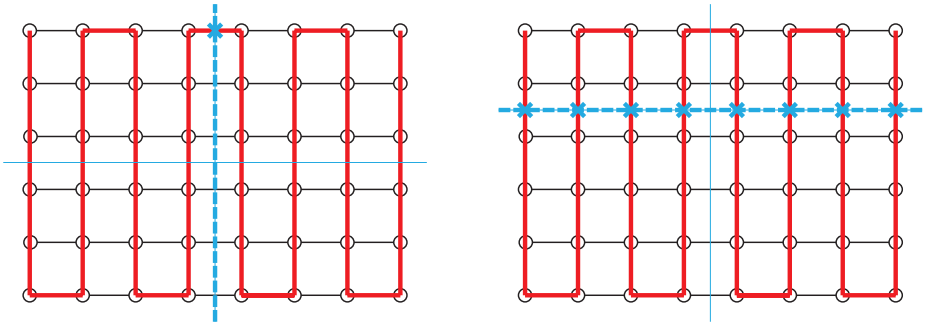

Fig. 19: *Two-dimensional DMRG setup: The 2D lattice is explored by a one-dimensional snake, for which one formulates an MPS. Horizontal interactions become long-ranged, but the real problem is that encoding the entanglement of vertically separated bipartitions of the system becomes exponentially expensive.*

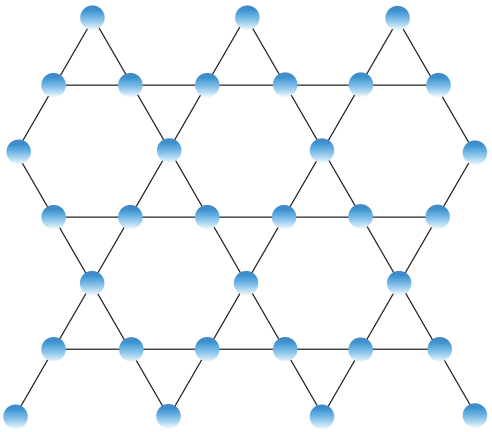

Fig. 20: *The kagome lattice, whose name derives from the structure of Japanese bamboo baskets, but can be found in natural magnetic substances.*

two-dimensional quantum systems, seriously starting with [38] and mainly focusing on Heisenberg and t-J-models in two dimensions.

The fundamental idea is to map (see Fig. 19) the two-dimensional lattice to a one-dimensional snake winding through the system, for which an MPS can be formulated. Technically, the price to pay is that short-ranged interactions perpendicular to the snake direction can become long-ranged (consider horizontal interactions in Fig. 19). In the beginning, this was considered to be the main reason for the comparatively disappointing results obtained (the reasoning need not concern us here). But in reality, the reason is given by the fact that entanglement for ground states of short-ranged Hamiltonians scales linearly with system size (area law [46]), i.e. $S \propto L$. As we have seen, a reduced density operator of dimension D can at most encode entanglement $S = \ln D$. Now the reduced density operators generated by an MPS of dimension D have at most dimension D themselves; hence we need MPS of (at least, because the distribution of the weights enters also) dimension 2^D to encode entanglement S. If we consider the vertical cut through the system, it is crossed by the snake only once, and the matrices therefore have to be *exponentially large* in the vertical size of the lattice, which strongly limits achievable system sizes. On the upside, horizontal cuts are crossed by the snake L times, and we have D^L coefficients if we think about the bottom and top parts of the system being represented by single matrices, and we can encode entanglement $S = L \ln D$, matching the area law. The size restriction therefore only applies in one spatial direction, such that simulations have focused on long stripe systems $L_x \gg L_y$.

As tensorial generalizations of MPS such as PEPS [9], MERA [47] or iPEPS [48] do not suffer from exponential growth of resources in two dimensions, they were at first considered vastly superior to the essentially unsuitable DMRG approach in two dimensions. However, it was realized that they suffer not only from scaling polynomially much worse in tensor dimension (up to D^{16} as opposed to D^3 for MPS operations), which would always be preferrable to exponential growth in resources for sufficiently large system sizes, but also from quite severe issues with normalizability and the conditioning of the arising linear algebra problems as well as further approximations involved in the contractions. Therefore, progress in that direction has not been

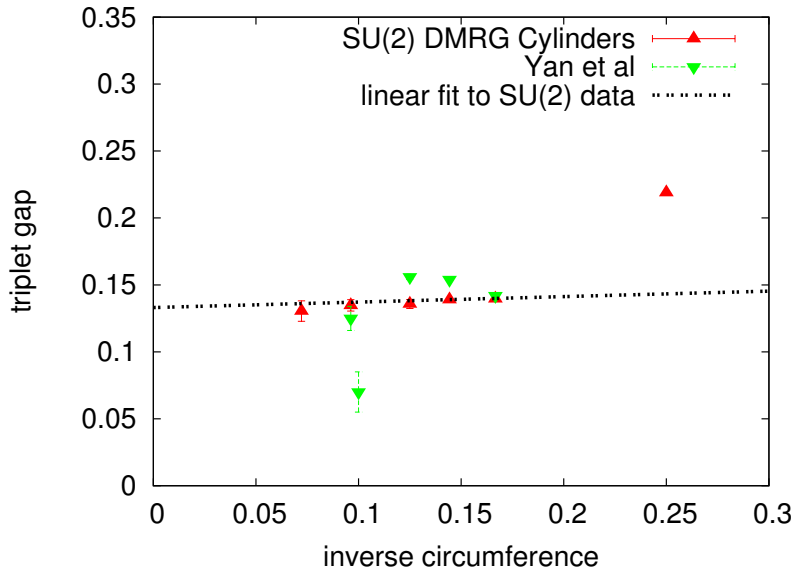

Fig. 21: *Size of the (triplet) gap of the Heisenberg antiferromagnet on the kagome lattice versus inverse cylinder width obtained in two large-scale DMRG simulations. The more advanced SU(2) symmetric simulations allow for the first time for a reliable extrapolation to the thermodynamic limit. Reproduced from [50].*

quite as spectacular as originally hoped for, which has left at least for now quite a bit of space for brute-force, but numerically well controlled DMRG applications, which have progressed both due to more powerful computers and progress in highly efficient implementations of the algorithm.

To illustrate the state of the art, let me mention the recent large-scale numerical studies of the isotropic Heisenberg antiferromagnet on the kagome lattice illustrated in Fig. 20, which have stirred up a lot of interest [49, 50]. The nature of the ground state has been discussed since the late eighties, without any conclusive answer, but numerous competing proposals so far. The most likely candidate for the ground state (in the view of many, the definite solution) could be identified recently as a quantum spin liquid by [49], more precisely, using $D \sim 17000$ and exploiting non-Abelian symmetries of the Hamiltonian in the largest DMRG study done so far, as a Z_2 topological quantum spin liquid [50]; lattices studied were on cylinders of a width of up to 18 lattice spacings and a length of up to 70 lattice spacings. For an example, consider Fig. 21, where the first reliable extrapolation of the (triplet) gap of the model is shown.

References

[1] S.R. White, Phys. Rev. Lett. **69**, 2863 (1992)

[2] S. Östlund and S. Rommer, Phys. Rev. Lett. **75**, 3537 (1995)

[3] J. Dukelsky, M.A. Martín-Delgado, T. Nishino and G. Sierra, Europhys. Lett. **43**, 457 (1998)

[4] G. Vidal, Phys. Rev. Lett. **93**, 040502 (2004)

[5] A.J. Daley, C. Kollath, U. Schollwöck and G. Vidal, J. Stat. Mech.: Theor. Exp. P04005 (2004)

[6] S.R. White and A.E. Feiguin, Phys. Rev. Lett. **93**, 076401 (2004)

[7] F. Verstraete, D. Porras and J.I. Cirac, Phys. Rev. Lett. **93**, 227205 (2004)

[8] F. Verstraete, J.J. Garcia-Ripoll and J.I. Cirac, Phys. Rev. Lett. **93**, 207204 (2004)

[9] F. Verstraete and J.I. Cirac, cond-mat/0407066 (2004)

[10] M. Zwolak and G. Vidal, Phys. Rev. Lett. **93**, 207205 (2004)

[11] U. Schollwöck, Rev. Mod. Phys. **77**, 259 (2005)

[12] U. Schollwöck, Ann. Phys. **326**, 96 (2011)

[13] F. Verstraete, V. Murg and J.I. Cirac, Adv. Phys. **57**, 143 (2008)

[14] A. Einstein, B Podolsky and N. Rosen, Phys. Rev. **47**, 777 (1935)

[15] J. Bell, Physics **1**, 195 (1964)

[16] M.A. Nielsen and I.L. Chuang: *Quantum computation and quantum information* (Cambridge University Press, 2000)

[17] I. Affleck, T. Kennedy, E.H. Lieb and H. Tasaki, Phys. Rev. Lett. **59**, 799 (1987)

[18] I. Affleck, T. Kennedy, E.H. Lieb and H. Tasaki, Comm. Math. Phys. **115**, 477 (1988)

[19] A. Klümper, A. Schadschneider and J. Zittartz, Europhys. Lett. **24**, 293 (1993)

[20] A. Kolezhuk, R. Roth and U. Schollwöck, Phys. Rev. Lett. **77**, 5142 (1996)

[21] A. Kolezhuk and H.J. Mikeska, Phys. Rev. Lett. **80**, 2709 (1998)

[22] L.N. Trefethen and D. Bau III: *Numerical linear algebra* (SIAM, Philadelphia, 1997)

[23] K.G. Wilson, Rev. Mod. Phys. **47**, 773 (1975)

[24] R. Bulla, T.A. Costi and T. Pruschke, Rev. Mod. Phys. **80**, 395 (2008)

[25] A. Weichselbaum, F. Verstraete, U. Schollwöck, J.I. Cirac and J. von Delft, Phys. Rev. B **80**, 165117 (2009)

[26] I.P. McCulloch, J. Stat. Mech.: Theor. Exp. P10014 (2007)

[27] M. Suzuki, Prog. Theor. Phys. **56**, 1454 (1976)

[28] E.H. Lieb and D. W. Robinson, Commun. Math. Phys. **28**, 251 (1972)

[29] C. Kollath, U. Schollwöck and W. Zwerger, Phys. Rev. Lett. **95**, 250402 (2006)

[30] M. Cheneau, P. Barmettler, D. Poletti, M. Endres, P. Schauß, T. Fukuhara, C. Gross, I. Bloch, C. Kollath and S. Kuhr, Nature **481**, 484 (2012)

[31] M. Cramer, A. Flesch, I.P. McCulloch, U. Schollwöck and J. Eisert, Phys. Rev. Lett. **101**, 063001 (2008)

[32] S. Trotzky, Y.-A. Chen, A. Flesch, I.P. McCulloch, U. Schollwöck, J. Eisert and I. Bloch, Nature Physics **8**, 325 (2012)

[33] T. Barthel, U. Schollwöck and S.R. White, Phys. Rev. B **79**, 245101 (2009)

[34] C. Karrasch, J.H. Bardarson and J.E. Moore, Phys. Rev. Lett. **108**, 227206 (2012)

[35] T. Barthel, U. Schollwöck and S. Sachdev, arXiv:1212.3570 (2012)

[36] T. Barthel, arXiv:1301.2246 (2013)

[37] I.P. McCulloch, arXiv:0804.2509 (2008)

[38] S.R. White, Phys. Rev. Lett. **77**, 3633 (1996)

[39] S.R. White, Phys. Rev. B **72**, 180403 (2005)

[40] E.R. Gagliano and C. A. Balseiro, Phys. Rev. Lett. **59**, 2999 (1987)

[41] K. Hallberg, Phys. Rev. B **52**, 9827 (1995)

[42] P.E. Dargel, A. Wöllert, A. Honecker, I.P. McCulloch, U. Schollwöck and T. Pruschke, Phys. Rev. B **85**, 205119 (2012)

[43] T.D. Kühner and S.R. White, Phys. Rev. B **60**, 335 (1999)

[44] E. Jeckelmann, Phys. Rev. B **66**, 045114 (2002)

[45] A. Holzner, A. Weichselbaum, I.P. McCulloch, U. Schollwöck and J. von Delft, Phys. Rev. B **83**, 195115 (2011)

[46] J. Eisert, M. Cramer and M.B. Plenio, Rev. Mod. Phys. **82**, 277 (2010)

[47] G. Vidal, Phys. Rev. Lett. **99**, 220405 (2007)

[48] J. Jordan, R. Orús, G. Vidal, F. Verstraete and J.I. Cirac, Phys. Rev. Lett. **101**, 250602 (2008)

[49] S. Yan, D. A. Huse and S.R. White, Science **332**, 1173 (2011)

[50] S. Depenbrock, I.P. McCulloch and U. Schollwöck, Phys. Rev. Lett. **109**, 067201 (2012)