

OPEN ACCESS

Minimally entangled typical thermal state algorithms

To cite this article: E M Stoudenmire and Steven R White 2010 *New J. Phys.* **12** 055026

View the [article online](#) for updates and enhancements.

Related content

- [Out-of-equilibrium dynamics with matrix product states](#)
M L Wall and Lincoln D Carr
- [Tensor network techniques for the computation of dynamical observables in one-dimensional quantum spin systems](#)
Alexander Müller-Hermes, J Ignacio Cirac and Mari Carmen Bañuls
- [From density-matrix renormalization group to matrix product states](#)
Ian P McCulloch

Recent citations

- [Quantum Computation of Finite-Temperature Static and Dynamical Properties of Spin Systems Using Quantum Imaginary Time Evolution](#)
Shi-Ning Sun *et al*
- [Performance evaluation of the discrete truncated Wigner approximation for quench dynamics of quantum spin systems with long-range interactions](#)
Masaya Kunimi *et al*
- [Localization of rung pairs in a hard-core Bose-Hubbard ladder](#)
Shang-Shu Li *et al*

Minimally entangled typical thermal state algorithms

E M Stoudenmire¹ and Steven R White²

¹ Department of Physics, University of California, Santa Barbara,
CA 93106-4030, USA

² Department of Physics and Astronomy, University of California, Irvine,
CA 92697, USA

E-mail: e.m.stoudenmire@gmail.com and srwhite@uci.edu

New Journal of Physics **12** (2010) 055026 (32pp)

Received 8 February 2010

Published 28 May 2010

Online at <http://www.njp.org/>

doi:10.1088/1367-2630/12/5/055026

Abstract. We discuss a method based on sampling minimally entangled typical thermal states (METTS) that can simulate finite temperature quantum systems with a computational cost comparable to the ground state density matrix renormalization group (DMRG). Detailed implementation of each step of the method is presented, along with efficient algorithms for working with matrix product states and matrix product operators. Furthermore, we explore how the properties of METTS can reveal characteristic order and excitations of systems and discuss why METTS form an efficient basis for sampling. Finally, we explore the extent to which the average entanglement of a METTS ensemble is minimal.

Contents

1. Introduction	2
2. Producing minimally entangled typical thermal states (METTS) with the pure state method	4
2.1. Imaginary time evolution	6
2.2. Measurement of observables	12
2.3. Collapsing METTS into CPS	14
3. Properties of METTS	20
3.1. Observing thermal fluctuations with METTS	20
3.2. Energy measurements and the efficiency of METTS	22
3.3. In what sense are METTS minimally entangled?	24
Acknowledgments	29
Appendix. Working with MPSs and MPOs	29
References	32

1. Introduction

According to the elementary principles of quantum statistical mechanics, the average of an observable is found by computing

$$\langle A \rangle = \text{Tr}[\rho A] = \frac{1}{\mathcal{Z}} \text{Tr}[e^{-\beta H} A]. \quad (1)$$

While performing such a calculation directly is usually intractable for quantum many body systems, one can still approximate the expectation value of A by strategies based on sampling. For instance, many quantum Monte Carlo methods take advantage of the Trotter decomposition of ρ to express an expectation value as a sum over virtual classical paths, which can then be numerically sampled. In such approaches, the act of randomly choosing a path corresponds to sampling both the thermal and the quantum fluctuations together.

However, it is also possible to sample only the thermal fluctuations, randomly selecting quantum states rather than classical ones. This may be done by expanding the trace in equation (1) in terms of an orthonormal basis $|i\rangle$ such that the expectation value of A takes the form

$$\begin{aligned} \langle A \rangle &= \frac{1}{\mathcal{Z}} \sum_i \langle i | e^{-\beta H/2} A e^{-\beta H/2} | i \rangle \\ &= \frac{1}{\mathcal{Z}} \sum_i P(i) \langle \phi(i) | A | \phi(i) \rangle. \end{aligned} \quad (2)$$

In the first line we have used the cyclic property of the trace and in the second line the set of normalized states $|\phi(i)\rangle$ are defined as

$$|\phi(i)\rangle = P(i)^{-1/2} e^{-\beta H/2} |i\rangle \quad (3)$$

together with the (unnormalized) probability distribution

$$P(i) = \langle i | e^{-\beta H} | i \rangle. \quad (4)$$

One may then estimate $\langle A \rangle$ by sampling the states $|\phi(i)\rangle$ with probability $P(i)/\mathcal{Z}$ and averaging the expectation values $\langle \phi(i) | A | \phi(i) \rangle$ computed at each step.

Although such a procedure could be carried out using any orthonormal basis of states $|i\rangle$, not every choice would lead to an efficient algorithm. In particular, the computational cost of working with a quantum state goes up the more entangled it is, as measured by the von Neumann entanglement entropy across bipartitions of the system.

Therefore a natural choice for the basis $|i\rangle$ is the set of classical product states (CPS). These are states with wavefunctions of the form

$$|i\rangle = |i_1\rangle |i_2\rangle |i_3\rangle \dots |i_N\rangle, \quad (5)$$

where i_j are the states in a local basis that may be chosen arbitrarily for each site j . As a result of their product structure, CPS have an entanglement entropy that is exactly zero.

One therefore might expect that, of all ensembles of states $|\phi(i)\rangle$, those produced from CPS have the least entanglement. Moreover, such $|\phi(i)\rangle$ have all of the properties one intuitively expects of typical states of a thermal system. At high temperatures they are effectively classical, while at lower temperatures they spontaneously break symmetries of the Hamiltonian in systems

that exhibit long-range order. They do not exhibit unphysical non-local correlations, and remain factorized over parts of the system that do not interact. And expectations of observables $\langle \phi(i) | A | \phi(i) \rangle$ lie very close to the thermal average. For such reasons as well as others that we will discuss below, these states have therefore been designated minimally entangled typical thermal states (METTS) [1].

Not only does each METTS provide a good characterization of the thermal ensemble, but there exists a simple algorithm, the pure state algorithm, for sampling many of them efficiently. This is achieved by a random walk through the set of METTS where the next state is constructed from a CPS obtained by measuring, or collapsing, the previous one. This collapsing step automatically ensures that the METTS are sampled with the correct distribution, and it allows one to optimize the sampling process by an appropriate choice of basis. Moreover, collapsing each state may be seen as sampling the quantum fluctuations of the system, a fact that can be used to accelerate the calculation of certain physical observables.

In this paper, we elaborate upon the implementation of the pure state algorithm for sampling METTS and its usage in realistic simulations. First, in section 2, we provide a more detailed explanation of the pure state algorithm and its implementation. We begin with a discussion of various methods for carrying out the imaginary time evolution implicit in equation (3), as it is the most computationally intensive part of the algorithm and the step where the most significant approximations must be made. Then, we turn to the measurement of physical observables. While it is most convenient to calculate quantities like the total energy using matrix product operators (MPOs), we will show how local observables can be calculated more efficiently by taking advantage of the redundancy inherent in the matrix product state (MPS) formalism (for readers unfamiliar with MPOs and MPS, a brief introduction is provided in the appendix). We then conclude the section by discussing an efficient method for collapsing a pure state into a CPS. The basis into which the CPS is measured can be arbitrary, and we will demonstrate how this freedom may be exploited both to improve the sampling autocorrelation time and to help calculate challenging observables.

In section 3, we turn our attention to the properties of the METTS themselves. The properties of individual METTS can provide insight into the order that is present in a system as well as characteristic thermal fluctuations. On the other hand, studying the entire METTS ensemble will show us why it is such an efficient basis for sampling. Finally, we quantify the sense in which METTS are minimally entangled and conjecture that although there exist thermal decompositions with less entanglement than METTS at some temperatures, no decomposition can outperform an optimal METTS ensemble at all temperatures.

Throughout, whenever we find it helpful to show the results of an explicit calculation, we will choose as our model of interest the $S = 1$ bilinear–biquadratic chain with the Hamiltonian

$$H = \cos \theta \sum_i \vec{S}_i \cdot \vec{S}_{i+1} + \sin \theta \sum_i (\vec{S}_i \cdot \vec{S}_{i+1})^2. \quad (6)$$

In particular, we will focus on two special cases of this model, namely $\theta = 0$, the $S = 1$ Heisenberg antiferromagnet and $\theta = \arctan[1/3]$, the Affleck–Kennedy–Lieb–Tasaki (AKLT) model possessing an $S = 1$ valence bond solid ground state [2]. At $T = 0$, both models are in the same phase, the Haldane phase, which is characterized by, among other things, a gap to all excitations [3] and a doubly degenerate entanglement spectrum [4].

2. Producing minimally entangled typical thermal states (METTS) with the pure state method

The pure state method consists of the following simple algorithm for producing a series of METTS such that they are correctly distributed with probability $P(i)/\mathcal{Z}$:

1. Choose a CPS $|i\rangle$.
2. Compute the METTS $|\phi(i)\rangle = e^{-\beta H/2}|i\rangle P(i)^{-1/2}$ and calculate observables of interest.
3. Collapse a new CPS $|i'\rangle$ from $|\phi(i)\rangle$ with probability $p(i \rightarrow i') = |\langle i'|\phi(i)\rangle|^2$ and return to step 2.

We will shortly discuss how to carry out these steps in practice. However, let us first see why producing METTS in this way is guaranteed to give the correct distribution.

Consider the ensemble of all CPS $|i\rangle$ initially distributed with probability $P(i)/\mathcal{Z}$. If we randomly choose a CPS $|i\rangle$ from this ensemble and then follow the above steps, the probability of measuring a particular CPS $|j\rangle$ at the end of step 3 will be

$$\begin{aligned} \sum_i \frac{P(i)}{\mathcal{Z}} p(i \rightarrow j) &= \sum_i \frac{P(i)}{\mathcal{Z}} |\langle j|\phi(i)\rangle|^2 \\ &= \sum_i \frac{P(i)}{\mathcal{Z}} \frac{|\langle j|e^{-\beta H/2}|i\rangle|^2}{P(i)} \\ &= \langle j|e^{-\beta H}|j\rangle/\mathcal{Z} \\ &= P(j)/\mathcal{Z}. \end{aligned} \tag{7}$$

That the resulting ensemble of all such $|j\rangle$ have the same distribution as the original ensemble is a consequence of the detailed balance condition

$$\frac{P(i)}{\mathcal{Z}} p(i \rightarrow j) = \frac{P(j)}{\mathcal{Z}} p(j \rightarrow i). \tag{8}$$

Therefore the ensemble of CPS with distribution $P(i)/\mathcal{Z}$ is a fixed point of this process. Finally then, because each METTS $|\phi(i)\rangle$ is produced deterministically from a CPS $|i\rangle$ in step 2, they will be generated with probability $P(i)/\mathcal{Z}$ as well.

In addition to generating METTS with the correct distribution, the pure state method has other important properties that make it advantageous for performing simulations. First, it may be defined in a completely general way, allowing one to choose a specific implementation based on the problem of interest. Next, it does not rely on a rejection method to perform sampling. That is, every METTS produced (in what is the costliest step of the algorithm) can be used to perform measurements. And the algorithm is readily parallelizable since its sampling method is that of a Markovian random walk.

Most importantly, the computational cost of the pure state algorithm is only modestly greater than that of ground state density matrix renormalization group (DMRG). The typical bond dimension m of the MPS representing each METTS produced by the imaginary time evolution step ranges from $m = 1$ for $\beta = 0$ to $m = m_0$ for very large β , where m_0 is the bond dimension necessary to represent the ground state using DMRG. As shown below, the imaginary time evolution for a system with N sites scales as Nm^3 per time step. Therefore the cost

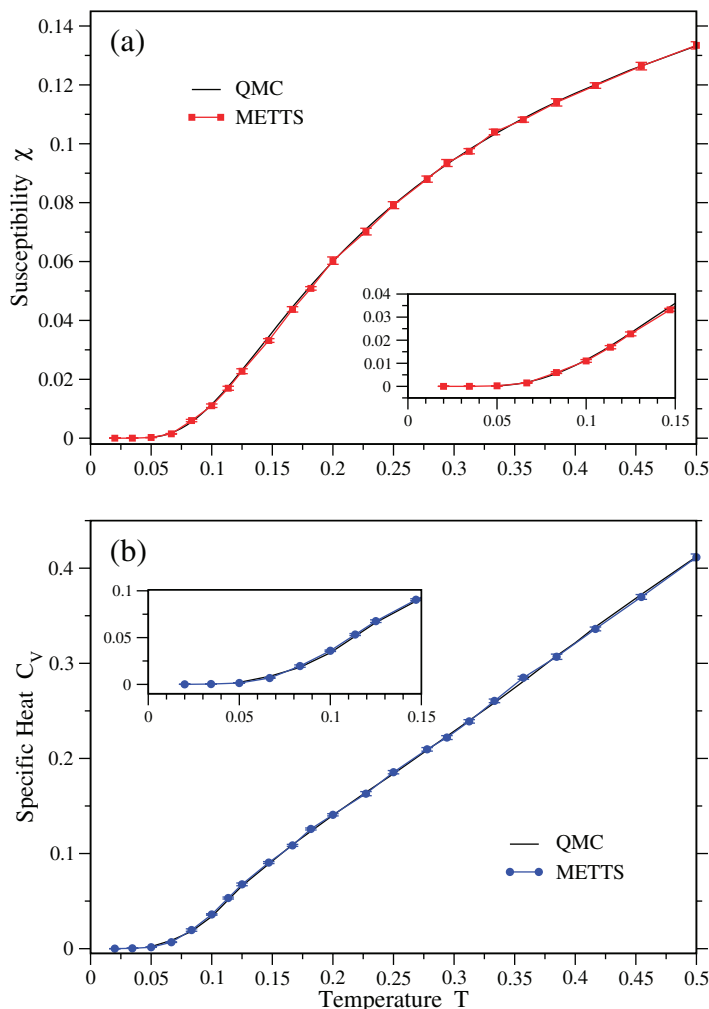


Figure 1. Low-temperature measurements of the (a) susceptibility and (b) specific heat of the $S = 1$ Heisenberg chain with $N = 100$ sites. The quantum Monte Carlo data shown by the solid black lines is treated here as exact.

of producing each METTS scales as $Nm^3\beta$ (ground state DMRG scales as Nm^3). The CPS collapse step is much less costly, scaling only as Nm^2 .

We have therefore been able to reach significantly lower temperatures with the METTS approach than with previous DMRG-based finite temperature methods. For example, Feiguin and White were able to simulate the $S = 1$ Heisenberg chain down to about $T = 0.05$ using the ancilla method, where about $m = 500$ DMRG states had to be kept at the lowest temperatures when using a truncation error cutoff of 1×10^{-10} . This is because their MPSs had to encode both the system and environment spins [5].

In contrast, to reach comparable temperatures using METTS requires only about $m = 60$ states when using the same cutoff. This has allowed us to produce accurate results down to at least $T = 0.02$, as shown in figure 1.

Since the computational cost of METTS is comparable to that of ground state DMRG, we expect that it can treat comparably large ladders and two-dimensional (2D) systems. This promises to be of great value for studying frustrated and fermionic models beyond one dimension—especially models with nontrivial phases at finite temperature.

Having defined the pure state method in a general setting, let us look more closely at each step. We will discuss how to produce and sample METTS in detail, and we will demonstrate that there is great flexibility within the basic algorithm that can allow us to optimize it for our system of interest.

2.1. Imaginary time evolution

Since imaginary time evolution is both the costliest step of the algorithm and the source of most numerical errors, we will begin by discussing some of the best methods for evolving a state in imaginary time and their relative advantages and disadvantages with regard to producing METTS.

In what follows, we will find it convenient to represent wavefunctions as MPSs and, in certain cases, operators as MPOs. A brief review of this formalism may be found in the [appendix](#).

2.1.1. The Trotter–Suzuki method. For a one-dimensional (1D) system, the simplest way to implement time evolution is to use a Trotter–Suzuki decomposition of the time evolution operator, evolving each bond one at a time. When using this method for our 1D benchmark calculations, we chose to do a second-order breakup as in the time-dependent DRMG method of White and Feiguin [6]. Assuming that the Hamiltonian consists of a sum of nearest-neighbor bond Hamiltonians $H = \sum_n H_n$, the operator $e^{-\tau H}$ may be decomposed as

$$e^{-\tau H} \simeq e^{-\tau H_1/2} e^{-\tau H_2/2} \dots e^{-\tau H_2/2} e^{-\tau H_1/2} + \mathcal{O}(\tau^3). \quad (9)$$

We emphasize here that the labeling of the terms in the Hamiltonian is arbitrary. However, for a 1D system with nearest-neighbor interactions it is convenient to let H_n refer to the interactions on the n th bond, making the operator decomposition above perfectly suited for a single DMRG sweep. Because each factor, or ‘gate’, in this breakup is a local operator, it can be computed exactly such that the only error in this treatment of the time evolution operator is the finite time step Trotter error. Then, each gate may be applied directly to the MPS representing the state of the system by using DMRG to expose the two sites around the bond to be time evolved.

While the Trotter–Suzuki method is accurate and convenient, it is directly applicable only to systems with nearest-neighbor interactions along the path taken by the MPS through the sites of the lattice. This makes it hard to apply to 2D systems, including those with nearest-neighbor interactions. Even quasi-1D ladders and 1D chains with longer range interactions cannot be treated by the above method without some modifications.

Various methods have therefore been developed to extend Trotter time evolution to such systems, for example Feiguin and White’s method based on a Runge–Kutta differential equation solver [7]. However, such methods turn out to be too inefficient for our purposes. Hence, we will discuss below some alternative methods that will be more effective for these challenging systems.

2.1.2. Trotter–Suzuki with swap gates. The first method extends the Trotter–Suzuki method by utilizing swap gates, a tool that is familiar in quantum computing. A swap gate $S(i, j)$ exchanges the states on two identical sites i and j . For a spin or boson system, $S(i, j) = \delta_{s_i s'_j} \delta_{s_j s'_i}$, the diagrammatic representation of which is shown in figure 2(a). Swap gates can be used for fermionic

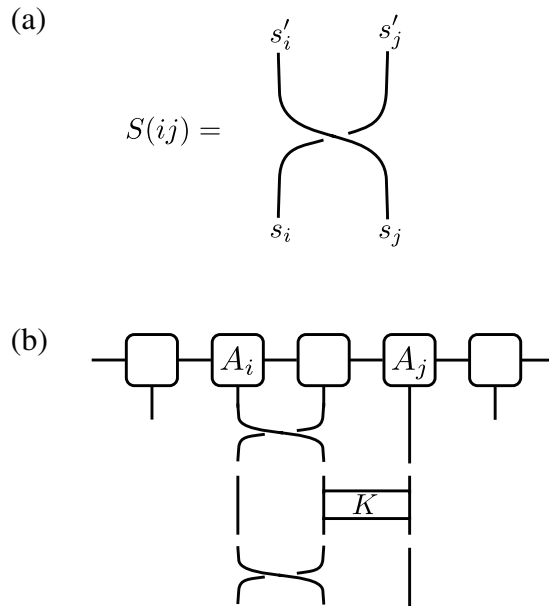


Figure 2. Diagrams illustrating (a) the structure of a swap gate tensor, which is a product of two identity tensors and exchanges the states of two lattice sites, and (b) the application of swap gates to an MPS, allowing a Trotter time evolution gate to act on effective nearest-neighbor sites while actually evolving a system with longer range interactions.

systems as well. In the case of spinless fermions, for instance, $S(ij)$ takes the same form as for bosons, but with an overall minus sign multiplying the case in which both sites are occupied.

Now, a generic Trotter–Suzuki breakup of the time evolution operator for a Hamiltonian containing only one site or two site interaction terms can be written in the form

$$e^{-\tau H} \simeq K_{i_1 j_1} K_{i_2 j_2} K_{i_3 j_3} \cdots \quad (10)$$

where the sites acted on by a given gate K_{ij} are not necessarily nearest-neighbors along the path taken by the MPS through the lattice. However, K_{ij} can be written in terms of a gate that does act on neighboring sites by using the identity

$$K_{ij} = \Theta_{i j-1} K_{j-1 j}^{(ij)} \Theta_{j-1 i}, \quad (11)$$

where the swap operators Θ can be defined as a product of nearest-neighbor swap gates as

$$\begin{aligned} \Theta_{i j-1} &= S_{i i+1} S_{i+1 i+2} \cdots S_{j-2 j-1}, \\ \Theta_{j-1 i} &= S_{j-2 j-1} \cdots S_{i+1 i+2} S_{i i+1} = \Theta_{i j-1}^\dagger. \end{aligned} \quad (12)$$

The operator $K_{j-1 j}^{(ij)}$ is the exponentiation of the same two-body term in the Hamiltonian as K_{ij} , but acts instead on the sites $j-1$ and j .

To make use of the above identity for the purpose of applying the single gate K_{ij} , one first applies the swap gates comprising $\Theta_{j-1 i}$ to the original MPS $|\psi\rangle$, yielding an effective MPS $|\psi'\rangle$. Then, the gate $K_{j-1 j}^{(ij)}$ is applied to $|\psi'\rangle$ as a local operator. Finally, the swap gates composing $\Theta_{i j-1}$ are applied, restoring the original order of the sites.

Finally, when applying many Trotter gates K_{ij} to an MPS, it is useful to observe that because $S(ij)^2 = 1$, the ordering of the Trotter gates in the decomposition can be chosen such

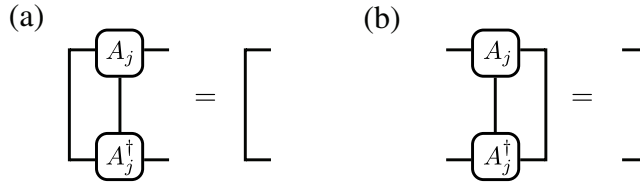


Figure 3. Diagrams illustrating (a) the left orthogonality condition equation (13) and (b) the right orthogonality condition equation (14).

that many of the swap gates cancel. Whenever this is the case, the repeated pair of gates can simply be omitted from the algorithm, leading to a significant speedup.

One way of arranging the Trotter decomposition to take advantage of this cancellation is as follows: taking $j > i$, order the K_{ij} first by i , from left to right, and then for each i , by j from left to right. Then all of the gates are applied in a single sweep that is arranged as a nested loop, where i is iterated over in the outer loop and j in the inner loop. In practice, all of the gates can be produced before applying any of them to the MPS so that cancellations can be found and the corresponding pairs of gates omitted. Even when taking advantage of cancellations, however, this method scales as $L_x L_y^2 m^3$ for an $L_x \times L_y$ ladder instead of scaling proportionally to the number of sites as in the case of a chain.

2.1.3. Imaginary time evolution with MPOs. The remaining methods for time evolution that we will discuss are based on MPOs. But before we turn to the formation of the time evolution MPO itself, let us first consider how to apply it to an MPS.

A concept that will be useful is that of the orthogonality center of an MPS. If we take the center to be at site i , then all of the MPS matrices A^{s_j} such that $j < i$ have been made left orthogonal in the sense that

$$\sum_{\alpha_{j-1} s_j} (A^{s_j})_{\alpha'_j \alpha_{j-1}}^\dagger A_{\alpha_{j-1} \alpha_j}^{s_j} = \delta_{\alpha'_j \alpha_j} \quad (13)$$

and all of the matrices A^{s_j} such that $j > i$ have been made right orthogonal such that

$$\sum_{s_j \alpha_j} A_{\alpha_{j-1} \alpha_j}^{s_j} (A^{s_j})_{\alpha_j \alpha'_{j-1}}^\dagger = \delta_{\alpha_{j-1} \alpha'_{j-1}}. \quad (14)$$

These conditions are illustrated in figure 3.

When an MPS has a well-defined orthogonality center, the set of A matrices to its left and right define an orthonormal basis for each half of the system. The matrix at the orthogonality center then defines the coefficients of the wavefunction in this basis, that is,

$$|\psi\rangle = \sum_{s_i \alpha_{i-1} \alpha_i} A_{\alpha_{i-1} \alpha_i}^{s_i} |\alpha_{i-1}^L\rangle |s_i\rangle |\alpha_i^R\rangle, \quad (15)$$

where the site index $s_i = 1, 2, \dots, d$, while the bond indices $\alpha_j = 1, 2, \dots, m$. Alternatively, when performing two-site operations one can take the orthogonality center to be a bond that can optionally include the sites adjacent to it.

When performing the usual singular value decomposition (SVD) or density matrix diagonalization in DMRG, one takes the tensor $\psi(\alpha_{i-1} s_i s_{i+1} \alpha_{i+1})$ representing the wavefunction and splits it into matrices A^{s_i} and $A^{s_{i+1}}$. These matrices give an optimal representation of the

tensor ψ and because they define expansion coefficients of the wavefunction in an orthonormal basis, the truncation of the wavefunction is also optimal. If there was no orthogonality center, or an SVD was performed away from the center, such an operation would optimally represent a local tensor but not an optimal truncation of the wavefunction.

In applying an MPO W to an MPS $|\psi_A\rangle$, one destroys its orthogonality. Apart from issues of efficiency, the simplest approach for forming the product $W|\psi_A\rangle$ would be as follows. Define the matrices B as

$$B_{\beta_{i-1}\beta_i}^{s_i} = B_{(\omega_{i-1}\alpha_{i-1})(\omega_i\alpha_i)}^{s_i} = \sum_{s'_i} W_{\omega_{i-1}\omega_i}^{s_i s'_i} A_{\alpha_{i-1}\alpha_i}^{s'_i} \quad (16)$$

for each site i . If the bond dimension of W is k and $|\psi_A\rangle$ is m , this defines a new MPS $|\psi_B\rangle = W|\psi_A\rangle$ of bond dimension mk if we think of the $\beta_i = (\omega_i\alpha_i)$ as combined, or fat, indices.

To guarantee proper truncation in this naive approach, one would first sweep left to right, performing SVDs to make the basis to the left orthogonal, but leaving the bond dimension as mk since the basis to the right is not orthogonal. Once at the right edge, the SVDs in the reverse sweep could truncate the bond dimension to m or to some specified truncation error. However, this procedure scales as Nm^3k^3 and would be highly inefficient if $k \sim 10\text{--}100$.

Verstraete and Cirac recommended a better approach (in a different context) that scales as $Nm^3k + Nm^2k^2$ based on fitting an MPS $|\psi_B\rangle$ to the product $W|\psi_A\rangle$ [8]. Form a random MPS $|\psi_B\rangle$ of bond dimension m and orthogonalize it to have any arbitrary orthogonality center. The optimal two-site wavefunction $\psi(\beta_{i-1}s_i s_{i+1}\beta_{i+1})$ at this center is then found by forming tensors L and R representing the product in the basis for the left and right halves of the system and combining them with the local W and A matrices as in figure 4. This ψ_B minimizes $\| |\psi_B\rangle - W|\psi_A\rangle \|^2$. One may then split ψ_B into matrices B^{s_i} and $B^{s_{i+1}}$ and sweep back and forth through the system, repeating this process until it converges.

Verstraete and Cirac originally advocated a single-site version of this approach but we have found the two-site version better—for example, the value of m can be automatically set to achieve a specified truncation error much more readily, particularly if one is keeping track of conserved quantum numbers. However, even when using the two-site version, this fitting procedure may fail to converge for systems with non-nearest-neighbor interactions in a reasonable number of iterations. The reason seems to be related to the ‘sticking’ problem of DMRG. When two distant sites are strongly entangled, the method has trouble building up this entanglement when improving the state near each site one at a time. On the other hand, the naive method above does not have such a sticking problem.

We have therefore developed an alternative approach, the zip-up algorithm, which has similar scaling to the fitting method but which does not get stuck. We begin by identifying the MPO W with an MPS $|W\rangle$ by combining its primed and unprimed (i.e. bra and ket) site indices. The orthogonality centers of $|\psi_A\rangle$ and $|W\rangle$ are then moved to the first site. This arrangement guarantees that the product of W and $|\psi\rangle$ at the first site has a basis to the right that, while not orthogonal, is not drastically ill-conditioned. By this we mean that while some linear dependence of the basis can exist, no basis state is expected to have a norm that is significantly bigger than one (norms much less than one are less of a problem).

Then, starting from the first site, we form the product MPS $|\psi_B\rangle$ as follows. Define the tensor C_1 as

$$C_{\omega_1\alpha_1}^{s_1} = \sum_{s'_1} W_{\omega_1}^{s_1 s'_1} A_{\alpha_1}^{s'_1}. \quad (17)$$

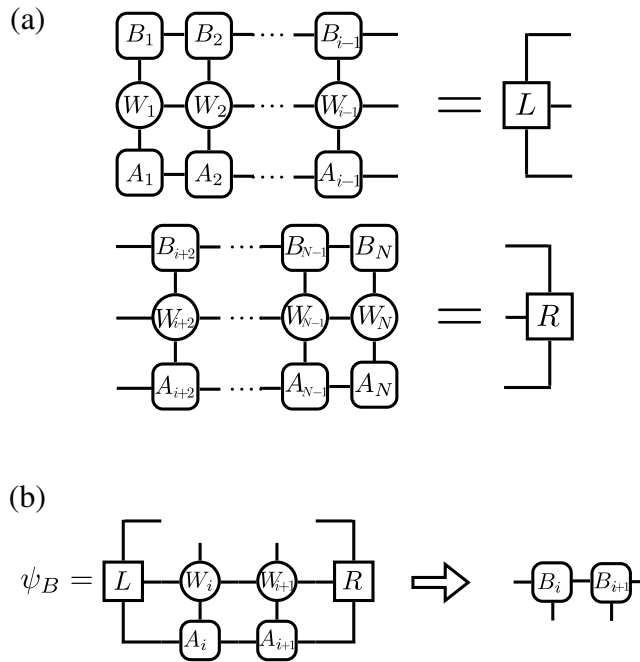


Figure 4. The fitting algorithm for multiplying an MPO times an MPS. Firstly, the tensors (a) L and R are formed by combining $|\psi_A\rangle$, the MPO W acting on it and the resulting state $|\psi_B\rangle$ moving from the edges of the system toward the orthogonality center. Then, ψ_B is defined by the tensor in (b) and split via SVD into matrices B_i and B_{i+1} . Finally, the orthogonality center of $|\psi_B\rangle$ is shifted and the process is repeated until convergence.

Thinking of C_1 as a matrix $C_{s_1(\omega_1\alpha_1)}$, where $(\omega_1\alpha_1)$ is a combined or ‘fat’ index, one may compute its SVD as

$$C_{s_1(\omega_1\alpha_1)} = U_{s_1\beta_1} \Lambda_{\beta_1} V_{\beta_1(\omega_1\alpha_1)}^\dagger. \quad (18)$$

At this step, it is crucial to perform a truncation, keeping only the m largest singular values contained in Λ_1 . This may be done by setting m explicitly or by specifying a cutoff on the truncation error. After truncating, the first matrix of $|\psi_B\rangle$ is defined as $B_{\beta_1}^{s_1} = U_{s_1\beta_1}$ and is therefore left orthogonal by construction.

The remaining B matrices may now be found recursively. To go on to the next step, one computes

$$C_{\beta_1\omega_2\alpha_2}^{s_2} = \sum_{s'_1, \omega_1, \alpha_1, s'_2} U_{\beta_1 s'_1}^\dagger C_{\omega_1\alpha_1}^{s'_1} W_{\omega_1\omega_2}^{s_2 s'_2} A_{\alpha_1\alpha_2}^{s'_2} \quad (19)$$

this time thinking of the result as a matrix $C_{(\beta_1 s_2)(\omega_2 \alpha_2)}$ and computing its SVD to find $U_{(\beta_1 s_2)\beta_2}$. The matrix B^{s_2} may then be set as $B_{\beta_1\beta_2}^{s_2} = U_{(\beta_1 s_2)\beta_2}$ and the process continued by defining $C_3 = C_{\beta_2\omega_3\alpha_3}^{s_3}$ analogously in terms of U_2, C_2, A_3 and W_3 .

For clarity, the construction of the C tensors is depicted graphically in figure 5. Asymptotically, if one assumes that $k < m$, this algorithm scales as Nm^3kd .

We note that when performing the SVDs in the above procedure, it is important to use a truncation error cutoff versus a fixed m during the initial left to right sweep to compensate for the modest loss of efficiency stemming from the fact that each truncation optimizes a local

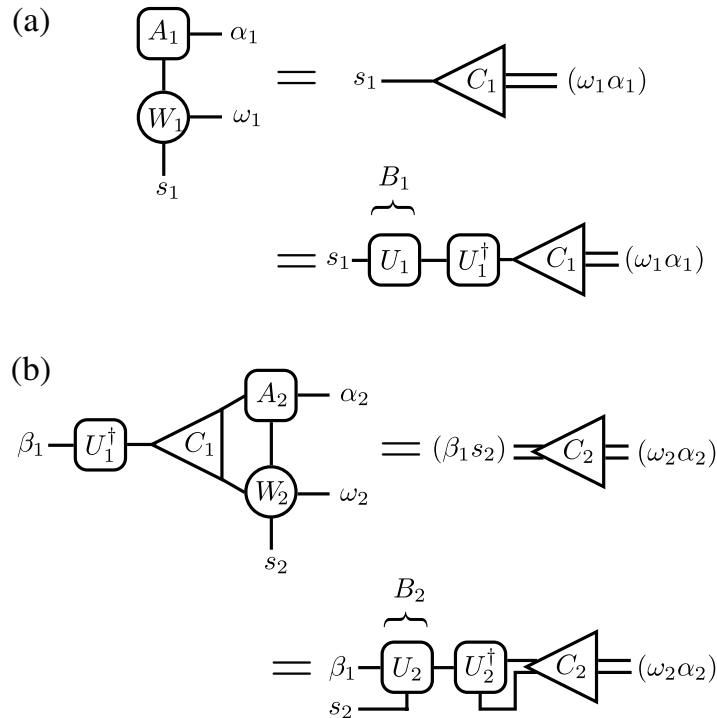


Figure 5. The tensors (a) C_1 and (b) C_2 generated in the first two steps of the zip-up method for multiplying an MPO times an MPS. Indices grouped by parentheses are to be thought of as a single fat index for the purpose of computing an SVD.

tensor, not $|\psi_B\rangle$ itself. Once the computation of $|\psi_B\rangle$ reaches the right boundary, the return SVD is fully efficient since the B matrices are left orthogonal and during this return sweep m can be reduced to an optimum value. For typical 2d clusters, we find that such a procedure leads to an intermediate \tilde{m} that is about twice m .

Finally, it may be useful to combine the above two approaches. One could perform an initial left to right sweep using the zip-up algorithm with a fixed $\tilde{m} \simeq m$ to obtain an initial guess for $|\psi_B\rangle$ and then switch to the fitting algorithm for the return sweep (and possibly additional sweeps) until convergence is reached.

2.1.4. Taylor series construction of the time evolution MPO. Having discussed the application of an MPO to an MPS, let us now turn to the construction of an MPO representation of the time evolution operator $K_\tau = e^{-\tau H}$. The first method we will discuss relies on expanding K_τ in a Taylor series.

To construct the MPO for K_τ using this approach, one begins by constructing an MPO representation of the Hamiltonian H . While it is sometimes possible to work out the MPO for H explicitly, such as in the case of translationally invariant 1D systems, more complex systems can be treated by constructing MPOs for each term in the Hamiltonian and summing them together (see the [appendix](#) for further details on adding and multiplying MPOs).

Having found an MPO representation of H , the Taylor series approach begins with breaking the timestep τ into small fractions $\epsilon = \tau/n$. One may then approximate $K_\epsilon = e^{-\epsilon H}$

in a Taylor series that is truncated at a high order p such that the error is controlled by ϵ^p . Finally, this series may be summed by utilizing MPO multiplication and addition algorithms. In doing so, it is important to arrange the computational steps so as not to repeat any expensive calculations such as computing powers of H . One such arrangement is given by the following algorithm.

Compute the MPO $W_0 = -\epsilon H$. Then compute the MPOs W_j recursively as

$$W_j = -\epsilon H \left(1 + \frac{1}{p-j+1} W_{j-1} \right), \quad (20)$$

where $j = 1, \dots, p$. The timestep operator itself is then given by $K_\epsilon = 1 + W_p$.

Finally, the MPO for K_τ may be found by computing the n th power of K_ϵ using the zip-up method for MPO multiplication described in the [appendix](#). In doing so, it is helpful to choose n to be a power of two so that K_τ may be computed by repeatedly squaring K_ϵ . In this way, fewer MPO multiplications are required.

2.1.5. Trotter–Suzuki construction of the time evolution MPO. An alternative method for constructing the time evolution MPO $K = e^{-\tau H}$ is to make use of the Trotter–Suzuki decomposition of equation (9), where K is approximated as a product of gates K_{ij} . But instead of applying the gates directly to the MPS representing the system, one combines them together to form an MPO for K .

Now, explicitly writing an MPO representation for a gate K_{ij} can be challenging, as it is not a simple product of single site operators and will not generally act on sites adjacent to each other in the MPS. However, one solution is to use the swap gate technique described above. One first works out the set of swap gates and local $K_{j-1,j}^{(ij)}$ gate operators that form K and removes any redundant gates. Then the MPO for K may be computed by creating an identity MPO and applying the gates directly to it, taking care to truncate the MPO bond dimension after each gate is applied. Or, if even greater accuracy is required, one may follow the steps above to construct an MPO for $K_\epsilon = e^{-\epsilon H}$ with $\epsilon = \tau/2^n$ and then use the zip-up method to square it n times and obtain K .

An important advantage of this Trotter–Suzuki method over the Taylor series approach is that errors are easier to identify and control. In particular, we have found that the parameters controlling the error in the Taylor series method must be reduced whenever the number of lattice sites N is increased in order to maintain the same accuracy. This can be traced to the fact that the Hamiltonian generally possesses at least one eigenvalue that scales as N , while other eigenvalues scale as N^p with $p < 1$.

In contrast, the Trotter–Suzuki approach to computing K does not suffer from this difficulty since one works with gates acting only on pairs of sites, and such gates may be constructed exactly. The only sources of error that remain, then, are the finite timestep Trotter error and the truncation error in the construction of the MPO.

2.2. Measurement of observables

After producing a METTS, one wants to measure a number of bulk quantities, such as the energy, as well as local observables, such as the magnetization on each site. For bulk observables that consist of local terms summed over the entire lattice, it is usually convenient to use an MPO representation. We must therefore find an efficient way to compute expectation values of such MPOs.

On the other hand, it is more efficient to work with local observables explicitly, and to compute their expectation values only in terms of the orthogonality center of the MPS. As we will see later, this second approach is particularly important as it is the key to performing an efficient CPS collapse in the last step of the pure state method.

2.2.1. Computing the expectation value of an MPO. Operators formed by summing over local terms, most notably the Hamiltonian, can generally be written as an MPO with relatively small bond dimension k . Since we will need to compute the expectation value of these MPOs with respect to MPS that have a relatively large bond dimension m , let us see how to do this efficiently.

Consider computing the matrix element $\langle \psi_A | \hat{W} | \psi_B \rangle$, where \hat{W} is an MPO of bond dimension k , and $|\psi_A\rangle$ and $|\psi_B\rangle$ are MPS of bond dimension m . Of course, this reduces to an expectation value if $|\psi_A\rangle = |\psi_B\rangle$.

One starts the computation at the beginning of the chain, defining L_1 as

$$L_{\alpha_1 \omega_1 \beta_1} = \sum_{s_1 t_1} (A^{s_1})_{\alpha_1}^\dagger W_{\omega_1}^{s_1 t_1} B_{\beta_1}^{t_1}. \quad (21)$$

Then L_2 can be produced from L_1 by computing

$$L_{\alpha_2 \omega_2 \beta_2} = \sum_{t_2 \beta_1} \left[\sum_{s_2 \omega_1} \left[\sum_{\alpha_1} L_{\alpha_1 \omega_1 \beta_1} (A^{s_2})_{\alpha_1 \alpha_2}^\dagger \right] W_{\omega_1 \omega_2}^{s_2 t_2} \right] B_{\beta_1 \beta_2}^{t_2}. \quad (22)$$

Here the order of operations in performing the sum is crucial to ensure good scaling—in particular one should not sum over both the α_1 and β_1 indices simultaneously.

One then proceeds to define L_3 in terms of L_2 in a similar manner until the end of the chain is reached. In this way, one computes the matrix element exactly as there is no truncation step. If the steps of the algorithm are followed in the order given above and we assume that $k < m$, the computation scales as Nm^3kd .

2.2.2. Computing the expectation value of a local operator. While any operator may be applied to a state by representing it as an MPO, it is often more efficient to compute the action of *local* operators on an MPS explicitly. The basic idea is to shift the orthogonality center of the MPS (defined in section 2.1.3 above) such that it lies within the set of sites on which the local operator acts. This way any MPS matrix on which the operator acts trivially is excluded from the computation.

For example, consider the computing of the expectation value of a single site operator \hat{O}_i acting on site i . If we take the orthogonality center to be the site i , then because the matrices A^{s_j} for all $j < i$ obey the left orthogonality condition equation (13) and the matrices A^{s_j} for all $j > i$ obey the right orthogonality condition equation (14), the expectation value of \hat{O}_i reduces to

$$\langle \psi_A | \hat{O}_i | \psi_A \rangle = \sum_{\alpha_{i-1} \alpha_i s'_i s_i} (A^{s'_i})_{\alpha_{i-1} \alpha_i}^\dagger \langle s'_i | \hat{O}_i | s_i \rangle A_{\alpha_{i-1} \alpha_i}^{s_i}. \quad (23)$$

Or, if we are interested in applying the operator to the state to obtain $\hat{O}_i | \psi_A \rangle$, this can be done by making the replacement

$$A_{\alpha_{i-1} \alpha_i}^{s_i} \rightarrow \sum_{s'_i} \langle s_i | \hat{O}_i | s'_i \rangle A_{\alpha_{i-1} \alpha_i}^{s'_i} \quad (24)$$

while keeping all other A matrices fixed.

Expectation values of multi-site local operators such as $\langle \psi_A | \hat{O}_i \hat{O}_{i+1} | \psi_A \rangle$ may be calculated in a similar way. Assuming that the orthogonality center is at site i or $i + 1$, one now includes both the matrices A^{s_i} and $A^{s_{i+1}}$. However, in contrast to the single-site computation, it is important in such multi-site expectation values to break up the tensor contractions such that the scaling of any one operation never exceeds $m^3 d$.

Finally, while each local expectation value may be computed with a scaling no worse than $m^3 d$, one usually wants to perform such a calculation at each site in the system. Since this necessarily involved shifting of the orthogonality center, the overall scaling will then be $N m^3 d^2$ as an SVD must be computed at each step.

2.2.3. Estimation of errors. While the methods discussed above give an efficient way to calculate observables for a single METTS, one must also consider the properties of the METTS ensemble when estimating averages and computing error bars. This is because correlations exist both between sequential METTS and between related observables within a METTS.

As we will discuss in the next section, the autocorrelation time of the METTS pure state method can be made quite short (less than five steps or so) by properly choosing the basis into which the CPS are collapsed. However, when estimating the error in a series of expectation values, it is still best to remove the effects of correlations by using a binning procedure as follows.

After calculating expectation values of an observable A , one collects sequential values into bins whose size is larger than the autocorrelation time. The average value within each bin is then computed. Each bin average may now be taken to be statistically independent and the error estimated as the standard error of these bin averages.

However, for quantities such as the specific heat or magnetic susceptibility, given by $C_v = \frac{\beta^2}{N} [\langle H^2 \rangle - \langle H \rangle^2]$ and $\chi = \frac{\beta}{N} [\langle m^2 \rangle - \langle m \rangle^2]$, computing the error as the sum of the errors in the first and second moments greatly overestimates the true error even when using binning. This can be traced to the fact that, from one step of the pure state algorithm to the next, the first and second moments of a bulk observable A are strongly correlated. That is, if $\langle A \rangle$ is relatively large for a particular METTS, so is $\langle A^2 \rangle$.

A simple way to deal with this difficulty is therefore to use a resampling method, such as the bootstrap method but with correlations taken into account. In other words, from the entire set of A expectation values, one randomly samples a new set of the same size as the original but with replacement, such that a value may appear more than once. One then calculates $\langle A \rangle$ for this resample. But then, instead of repeating this procedure independently for $\langle A^2 \rangle$, one resamples instead the *same* subset of METTS. So, if the expectation value of A calculated from a particular METTS is included in the resample n times, the corresponding expectation value of A^2 is included n times as well.

The bootstrap method then consists of repeating the above process a large number of times, each time obtaining an estimate for the second cumulant $\langle A^2 \rangle - \langle A \rangle^2$. If the process is repeated enough times, the resulting distribution of second cumulant estimates should be approximately Gaussian. Finally, the error in the second cumulant may be estimated as the standard deviation of this Gaussian distribution.

2.3. Collapsing METTS into CPS

After generating a METTS by imaginary time evolution and calculating observables of interest, the METTS must be collapsed into a CPS so that a new METTS can be generated with the

correct probability. This can be done in any basis, a fact that one can exploit to reduce the autocorrelation time of the pure state algorithm and to accelerate the measurement of certain observables.

It is also important, for reasons of efficiency, to carry out this collapse for each site one at a time. We will therefore discuss how to collapse a METTS in practice and show that the resulting algorithm scales as Nm^2d^2 .

2.3.1. CPS collapse algorithm. Since collapsing METTS is simplest for the case of a spin-1/2 model, let us begin by considering this case before discussing the general method.

Say one wishes to collapse the spin of a pure state $|\Psi\rangle$ at site i along the \hat{n} -axis. Define the states $|+\rangle$ and $|-\rangle$ at site i such that $\hat{n} \cdot \vec{S}_i|\pm\rangle = \pm\frac{1}{2}|\pm\rangle$. Then the probability of finding the spin at site i to be in the $|+\rangle$ state is $p_+ = \langle\Psi|\hat{n} \cdot \vec{S}_i|\Psi\rangle + \frac{1}{2}$ and to be in the $|-\rangle$ state is $p_- = 1 - p_+$.

After determining the new state for site i using a random number generator, the actual collapse is enacted by the projection

$$|\Psi\rangle \rightarrow \begin{cases} p_+^{-1/2} |+\rangle\langle+|\Psi\rangle & \text{prob } p_+ \\ p_-^{-1/2} |-\rangle\langle-|\Psi\rangle & \text{prob } p_- \end{cases} \quad (25)$$

Finally, once this procedure is carried out for every site in the system, one obtains a CPS $|\Psi'\rangle$ with probability $|\langle\Psi'|\Psi\rangle|^2$.

Now consider the more general case of a lattice system where the local Hilbert space at each site i can be expanded in an orthonormal basis $|m\rangle_i$ with $m = 1, \dots, d_i$. We again emphasize that this basis can be chosen arbitrarily for each site i and for each step of the algorithm.

Define $P_i(m) = |m\rangle_i\langle m|_i$ as the projection operator into the state $|m\rangle_i$. Then the probability of finding site i to be in the state $|m\rangle_i$ is just $p_m = \langle\Psi|P_i(m)|\Psi\rangle$.

Having selected a particular state $|\tilde{m}\rangle_i$ according to this probability, one completes the collapse by making the replacement

$$|\Psi\rangle \rightarrow p_{\tilde{m}}^{-1/2} P_i(\tilde{m})|\Psi\rangle. \quad (26)$$

Before discussing the detailed implementation of this method for MPS, let us conclude with some explicit examples of projection operators. From the above discussion, we can immediately identify the projectors for the $S = 1/2$ basis states $|+\rangle$ and $|-\rangle$ as

$$\begin{aligned} P_i(+)&= \frac{1}{2} + \hat{n} \cdot \vec{S}_i, \\ P_i(-)&= \frac{1}{2} - \hat{n} \cdot \vec{S}_i. \end{aligned} \quad (27)$$

Or, if we consider an $S = 1$ system and take $|m\rangle_i$ with $m = \bar{1}, 0, 1$ to be the eigenstates of $\hat{n} \cdot \vec{S}$, the $P_i(m)$ may be written explicitly as

$$\begin{aligned} P_i(1) &= \frac{1}{2}(\hat{n} \cdot \vec{S}_i) + \frac{1}{2}(\hat{n} \cdot \vec{S}_i)^2, \\ P_i(0) &= 1 - (\hat{n} \cdot \vec{S}_i)^2, \\ P_i(\bar{1}) &= -\frac{1}{2}(\hat{n} \cdot \vec{S}_i) + \frac{1}{2}(\hat{n} \cdot \vec{S}_i)^2. \end{aligned} \quad (28)$$

2.3.2. Collapsing an MPS into a CPS. Having defined the CPS collapse algorithm in a general setting, let us see how to apply it to a state $|\psi_A\rangle$ represented as an MPS. First, assume that the orthogonality center of the MPS has been chosen to be the first site. In other words, assume that the matrices A^{s_j} such that $j = 2, 3, \dots, N$ have been made right orthogonal as in equation (14).

One begins the collapse by first computing the expectation values of the projectors $P_1(m)$. As discussed in section 2.2, the existence of a well-defined orthogonality center allows us to compute these expectation values in terms of the matrix A^{s_1} alone. Explicitly, one computes

$$\begin{aligned} \langle \psi_A | P_1(m) | \psi_A \rangle &= \sum_{\alpha_1 s'_1} (A^{s'_1})_{\alpha_1}^\dagger \langle s'_1 | P_1(m) | s_1 \rangle A_{\alpha_1}^{s_1} \\ &= \sum_{\alpha_1} \left[\sum_{s'_1} (A^{s'_1})_{\alpha_1}^\dagger \langle s'_1 | m \rangle \right] \left[\sum_{s_1} \langle m | s_1 \rangle A_{\alpha_1}^{s_1} \right]. \end{aligned} \quad (29)$$

Once the new state of site 1 is chosen to be $|s_1\rangle = |m_1\rangle$, say, with probability $p_{m_1} = \langle \psi_A | P_1(m_1) | \psi_A \rangle$, the state could be collapsed by naively applying the projector $P_1(m_1)$ to it. However, while the action of a generic single-site operator would be followed by an SVD to orthogonalize site 1 and turn site 2 into the center site, the fact that the CPS collapse operators are projectors allows for an even more efficient algorithm.

Because the repeated action of single-site projectors on any state produces a product state and a product state can be represented as an MPS with bond dimension 1, if the above step of applying the projector $P_1(m_1)$ to site 1 and left orthogonalizing A^{s_1} was performed, the new bond index α'_1 connecting A^{s_1} to A^{s_2} could be truncated such that it only takes on one value. Anticipating this result, one can therefore directly replace A^{s_1} with the 1×1 matrix $A^{s_1} = \langle s_1 | m_1 \rangle$. Finally then, to ensure that $P_1(m_1) | \psi_A \rangle$ still describes the same state of the remaining sites not acted on by the projector, one must replace A^{s_2} according to

$$A_{\alpha_1 \alpha_2}^{s_2} \rightarrow A_{m_1 \alpha_2}^{s_2} = p_{m_1}^{-1/2} \sum_{s_1 \alpha_1} \langle m_1 | s_1 \rangle A_{\alpha_1}^{s_1} A_{\alpha_1 \alpha_2}^{s_2}. \quad (30)$$

That is, the label m_1 plays the role of a new bond index that runs over just one value. In fact, because it plays a trivial role, this label can be dropped with the result that the new state of the system is a product of a single-site wavefunction for site 1 and an $N - 1$ site MPS for the remaining sites. Having made this identification, the second site now lies on the boundary of an MPS and can be collapsed in exactly the same way as the first.

Although applying a local operator typically scales as $m^3 d^2$ per site, the considerations above show that the CPS collapse step of the pure state algorithm can be optimized even further. The result is an algorithm that scales as $m^2 d^2$ where the costliest step is the formation of the new A matrix in equation (30).

2.3.3. Choosing the CPS basis for spin models. When collapsing a METTS into an CPS, one has the freedom to choose the CPS basis arbitrarily for each METTS and at each site. We would therefore like to exploit this freedom first of all to ensure ergodicity and also to minimize correlations among the METTS.

One simple strategy for improving ergodicity is to pick a random quantization axis \hat{n} for each site at every step, and collapse the spins into the basis of $\hat{n} \cdot \hat{S}$ eigenstates. Then if the Hamiltonian conserves a quantity such as the total \hat{z} magnetization S_{tot}^z , such random measurements will generate CPS that explore many different S_{tot}^z sectors. Moreover, random projections do not explicitly break rotational symmetry in the sense of favoring a particular axis from the outset. We have found that in practice random projections work quite well and give a short autocorrelation time as shown in figure 6.

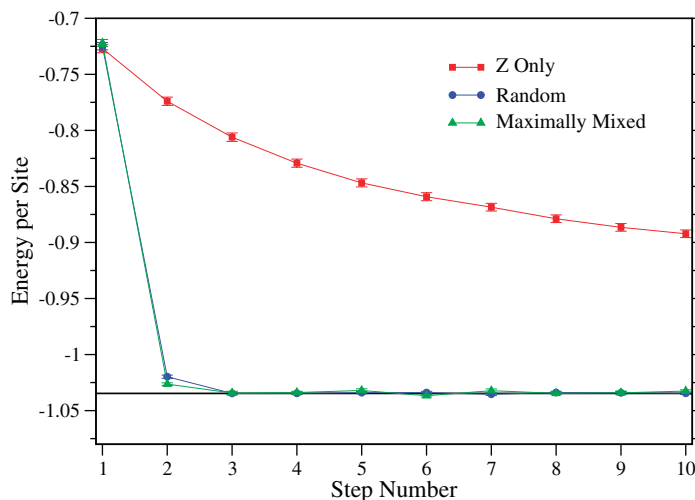


Figure 6. The energy of the $L = 50$, $S = 1$ Heisenberg chain at $\beta = 1.0$ found by choosing a completely random CPS (like those produced in a $\beta = 0$ simulation) and measuring the energy of an METTS produced after an given number of steps of the pure state method. The symbols show the ensemble average of many such runs each starting from a different CPS (the line is the quantum Monte Carlo result, treated here as exact). Depending upon the basis chosen for the CPS collapse, the number of steps needed for the effects of the initial CPS choice to be removed varies widely. Measuring only along the z -axis produces the strongest autocorrelation effects, while choosing random axes or mutually mixed bases gives much better performance and an autocorrelation time of less than 4–5 steps.

In contrast, one could instead collapse every site along the same axis, say \hat{z} . This has a number of advantages, including being simpler to code and making measurements of diagonal operators easier to perform, as we will discuss below. One particularly nice feature of using a \hat{z} -only basis is that, since all of the collapsed spins lie along the same axis, the resulting CPS can be easily visualized and plotted.

For example, in figures 7 and 8, we show the CPS obtained after collapsing subsequent METTS in simulations of the Heisenberg and AKLT models and mark points at which the spins fail to follow the diluted Néel pattern that underlies the string order in the Haldane phase. It is interesting to see the number of defects decrease as the temperature is lowered, and to observe that for the AKLT case, they disappear entirely as $T \rightarrow 0$.

A serious drawback to the \hat{z} -only approach, however, is that it leads to problems with ergodicity. In practice, using the \hat{z} basis only tends to lengthen the autocorrelation time, reducing the efficiency of the algorithm for calculating thermal averages. For example, one can clearly see strong correlations in the positions of the defects plotted in figures 7(a) and 8(a). Correlations in energy measurements can persist over many steps as shown in figure 6.

Since we would like to retain the advantages of the \hat{z} -only basis but cannot afford to sacrifice ergodicity, one solution is to collapse METTS along the \hat{z} -axis only on even numbered steps. If we then collapse along random axes after odd steps, we can maintain a short autocorrelation time.

This compromise suggests an even better approach, however. Instead of switching to random axis CPS collapses, we can use a basis that is maximally mixed relative to the S^z



Figure 7. CPS produced by performing \hat{z} -only measurements of METTS for the $S = 1$ Heisenberg antiferromagnet with $N = 100$. Shown are the middle 60 sites for (a) $\beta = 1.0$, (b) $\beta = 5.0$ and (c) $\beta = 20.0$. The vertical red lines mark defects in the diluted Néel order of the Haldane phase.

eigenstates. For example, in an $S = 1/2$ system, one would perform odd step collapses along the \hat{x} -axis. As S^x eigenstates are in an equal superposition of S^z eigenstates, the autocorrelation time is expected to be minimal since any memory of the previous CPS will be completely erased.

Likewise, for $S = 1$ systems, one can perform odd step collapses in the maximally mixed basis given by

$$\begin{aligned}
 |\mu_1\rangle &= \frac{1}{\sqrt{3}}(e^{i2\pi/3}|1\rangle - |0\rangle - e^{-i2\pi/3}|\bar{1}\rangle), \\
 |\mu_2\rangle &= \frac{1}{\sqrt{3}}(|1\rangle - |0\rangle - |\bar{1}\rangle), \\
 |\mu_3\rangle &= \frac{1}{\sqrt{3}}(e^{-i2\pi/3}|1\rangle - |0\rangle - e^{i2\pi/3}|\bar{1}\rangle).
 \end{aligned}
 \tag{31}$$



Figure 8. CPS produced by performing \hat{z} -only measurements of METTS for the $S = 1$ AKLT antiferromagnet with $N = 100$. The temperatures and symbols are identical to those in figure 7.

As shown in figure 6, using this basis on odd steps reduces correlation effects even more effectively than the random axis approach, for which the autocorrelation time is already quite short.

2.3.4. Diagonal measurements with product states. While the cost of computing expectation values of quantities such as the energy or the magnetization is quite low for each METTS produced, other measurements can be very costly.

For example, we may want to measure real space spin correlations $C(\Delta) = \sum_i \langle \vec{S}_i \cdot \vec{S}_{i+\Delta} \rangle$. One approach would be to create an MPO that encodes the sum over all pairs of spin operators separated by Δ . Or, we can work with MPOs for the individual spin operators and perform the sum above for each METTS. In practice however, we have found that both approaches are very costly because of both the time needed to construct the MPOs and to compute the expectation values.

But assuming we are studying a spin rotationally invariant system, only the $\langle S_i^z S_j^z \rangle$ correlations are needed as the others are equal by symmetry. Since $S_i^z S_j^z$ is diagonal in the product basis of S^z eigenstates, we can take advantage of this fact to perform correlator measurements using the CPS collapsed from each METTS instead of using the METTS themselves. As long as the operator we wish to calculate is diagonal in the basis of the collapsed CPS, we will still obtain the correct thermal average.

So, if one performs \hat{z} -only collapses as discussed in the previous section, the operator content of measuring a diagonal observable such as $C^z(\Delta) = \sum_i \langle S_i^z S_{i+\Delta}^z \rangle$ can be ignored entirely. That is, if we collapse a METTS into an CPS $|\Psi\rangle$ in the \hat{z} basis such that

$$|\Psi\rangle = |s_1\rangle|s_2\rangle|s_3\rangle \dots |s_N\rangle, \quad (32)$$

then we immediately have $C^z(\Delta) = \sum_i s_i s_{i+\Delta}$.

One can make even better use of the resources expended in producing each METTS by re-collapsing it multiple times and measuring each resulting CPS. This procedure also produces the correct thermal average for diagonal observables and is much more efficient for large β than collapsing each METTS only once. By using the pure state algorithm in this way, one explicitly samples not only the thermal fluctuations, but the quantum fluctuations as well.

Finally, since the above method relies on performing CPS collapses along the \hat{z} -axis, it is susceptible to the same ergodicity issues mentioned earlier. However, ergodicity can be restored by alternating between the \hat{z} -axis collapses on even steps and random or maximally mixed collapses on odd steps. For the $S = 1/2$ case, since the maximally mixed basis is just the \hat{x} basis, we can perform measurements of quantities like $C(\Delta)$ for rotationally invariant models on *every* step by treating the \hat{x} basis as an effective \hat{z}' basis.

3. Properties of METTS

While the pure state algorithm is effective for computing thermal averages, it also generates an entire wavefunction at each step that may be thought of as a snapshot of the system. Every METTS it produces can be used to simultaneously calculate many different observables and look for characteristic fluctuations or short range order.

Studying the properties of METTS can also help us to understand why relatively few METTS are needed to accurately estimate observables. By identifying the METTS ensemble that maximizes this sampling efficiency for a simple model, we will discover that while the entanglement entropy of such an ensemble is not always optimal, it may be minimal in a more restricted sense.

3.1. Observing thermal fluctuations with METTS

Figures 9 and 10 each show the properties of a single METTS produced for the spin-one Heisenberg and AKLT chains, respectively. The temperature is taken to be $T = 0.1$ in the units of equation (6).

Throughout both chains, antiferromagnetic spin correlations are clearly visible but thermal fluctuations are also apparent. For example, the AKLT chain in figure 10 shows longitudinal fluctuations in the magnitude of $\langle \vec{S} \rangle$ around site 40. In the inset, one can see that the local bond energy is also correspondingly higher at this point in the chain. This may be understood as a fluctuation away from the spin liquid ground state in which $\langle \vec{S} \rangle = 0$.

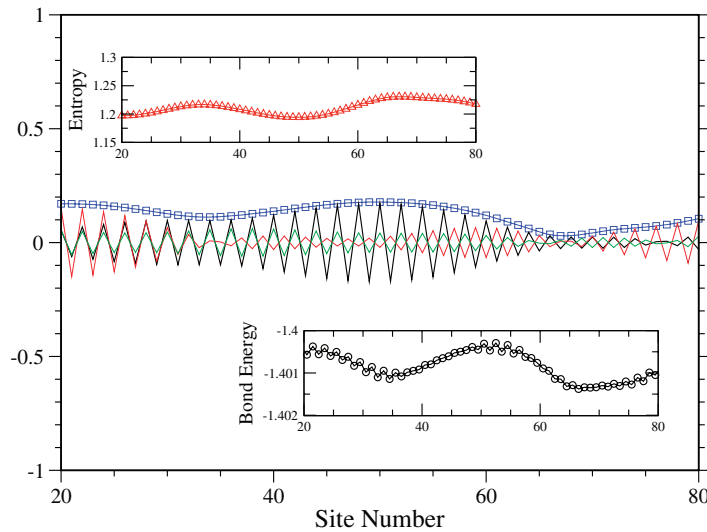


Figure 9. Properties of a METTS produced for the 100 site $S = 1$ Heisenberg chain at $T = 0.1$, central 80 sites. In the main plot, the solid lines (red, green and black) show the three components of $\langle \vec{S} \rangle$, while the (blue) boxes show $|\langle \vec{S} \rangle|$. The entanglement entropy on each bond is shown in the top inset, while the expectation value of each Hamiltonian bond term is shown in the bottom inset.

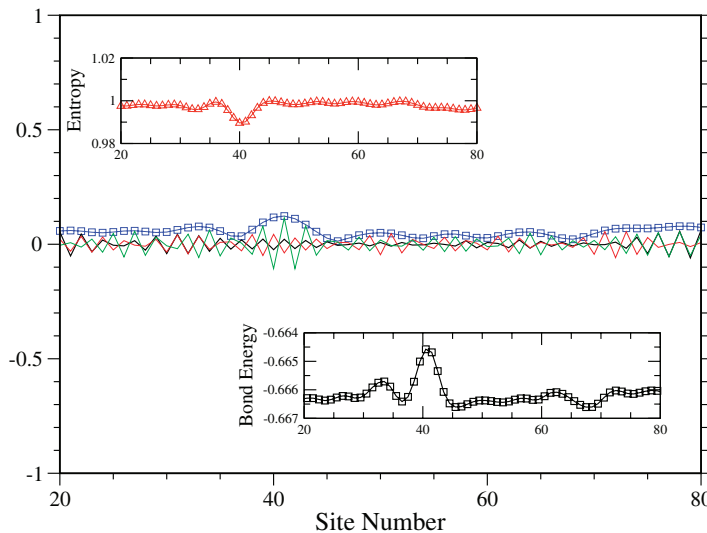


Figure 10. Properties of a METTS produced for the 100 site $S = 1$ AKLT chain, central 80 sites. The temperature and symbols are the same as those in figure 9.

A second type of thermal fluctuation that is visible is a twist in the staggered magnetization, such as the one near site 30 of the Heisenberg chain of figure 9. Near this site, the spin length is also shortened and the local bond energy is relatively low. After the twist, a second region of correlated spins persists for a few correlation lengths until another twist occurs near site 70.

It is interesting to observe that since each METTS is a pure state, the von Neumann entanglement entropy for bipartitions of the system about each bond is well defined, and we normalize it here such that a spin-1/2 singlet has an entropy of 1. As figures 9 and 10 show,

in the regions where the spins are more classical and the bond energy is relatively higher, the entanglement entropy is lower, signifying weaker quantum fluctuations.

3.2. Energy measurements and the efficiency of METTS

Let us now turn our attention from the properties of individual METTS to ensembles of METTS. An important concept will be that of a *thermal decomposition*, by which we mean a set of pure states $|\phi(i)\rangle$ and weights $P(i)$ such that

$$\sum_i P(i) |\phi(i)\rangle \langle \phi(i)| = e^{-\beta H}. \quad (33)$$

In fact, all density matrices can be decomposed in this way and the $|\phi(i)\rangle$ need not be orthogonal, although they should be normalized. From equations (3) and (4), one sees that every METTS ensemble corresponds to such a decomposition.

Implicit in equation (33) is a division of thermal effects into classical fluctuations quantified by the weights $P(i)$, and quantum fluctuations present within the states $|\phi(i)\rangle$, although this distinction is somewhat arbitrary as there exist many different thermal decompositions. But because we are interested in sampling the $|\phi(i)\rangle$ rather than calculating them all, our sampling process will be especially efficient when most of the thermal effects can be captured within the states themselves, rather than in their distribution.

To make this intuition more precise, let us consider the energy fluctuations within a given thermal decomposition. Using the notation $\langle A \rangle_i = \langle \phi(i) | A | \phi(i) \rangle$, we may write the specific heat as

$$C_v = \frac{\beta^2}{N} \left[\sum_i \frac{P(i)}{\mathcal{Z}} \langle H^2 \rangle_i - \left(\sum_i \frac{P(i)}{\mathcal{Z}} \langle H \rangle_i \right)^2 \right]. \quad (34)$$

However, the fact that we are working with an ensemble of pure states also allows us to define a ‘quantum specific heat’

$$C_v^{(q)} = \frac{\beta^2}{N} \sum_i \frac{P(i)}{\mathcal{Z}} \left[\langle H^2 \rangle_i - \langle H \rangle_i^2 \right] \quad (35)$$

that quantifies how much each state $|\phi(i)\rangle$ contributes to the full specific heat on average.

Now it turns out that, up to a constant factor, the difference between the quantum specific heat and the total specific heat is nothing but the variance of the expectation value of the energy within the ensemble, that is,

$$\begin{aligned} C_v - C_v^{(q)} &= \frac{\beta^2}{N} \sigma^2(H) \\ &= \frac{\beta^2}{N} \left[\sum_i \frac{P(i)}{\mathcal{Z}} \langle H \rangle_i^2 - \left(\sum_i \frac{P(i)}{\mathcal{Z}} \langle H \rangle_i \right)^2 \right]. \end{aligned} \quad (36)$$

So, the smaller this difference, the fewer the number of states that must be sampled to estimate the energy to a given accuracy.

Now turning to a particular choice of a thermal decomposition, let us first consider the exact energy eigenstates of the system, taking $|\phi(i)\rangle = |\epsilon_i\rangle$ and $P(i) = e^{-\beta \epsilon_i}$. Since H is diagonal in

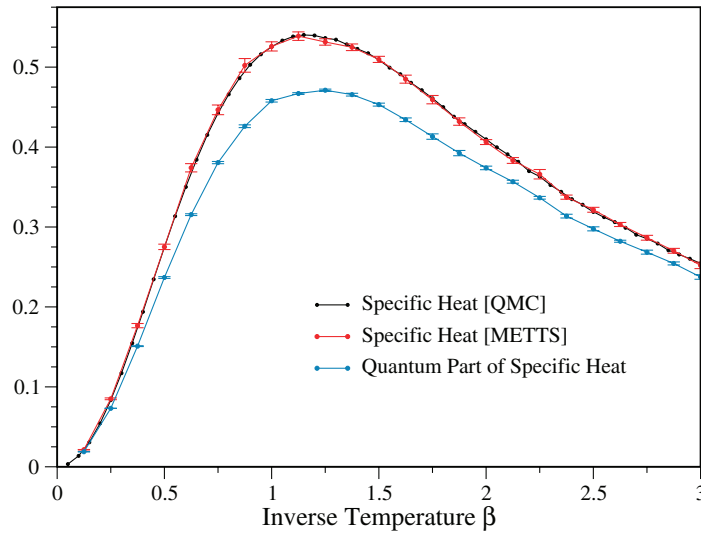


Figure 11. The specific heat of the $S = 1$ Heisenberg chain with $N = 100$ sites calculated using METTS with each CPS collapse performed along random axes. Also shown is the quantum specific heat, that is, the average estimate of the specific heat as calculated from a single METTS.

this basis, it follows that $\langle H^2 \rangle_i = \langle H \rangle_i^2$. Therefore the quantum specific heat is precisely zero and, as a result, the variance in the energy is maximal. So even if one could sample the exact energy eigenstates of the system, they would make an exceedingly poor basis for the purpose of estimating the average energy.

Next, consider a decomposition of the thermal ensemble in terms of METTS. From the calculation of both the specific heat and the quantum specific heat shown in figure 11 for the $S = 1$ Heisenberg chain, we can see that, even near the peak where the difference between C_v and $C_v^{(q)}$ is greatest, the quantum specific heat is only smaller than the total by about a factor of 0.8. This indicates that the METTS decomposition we sampled is quite close to being optimal. Physically, we may attribute this efficiency to the fact that, for a given temperature T , each METTS has a significant overlap with all of the energy eigenstates $|\epsilon_i\rangle$ such that $\epsilon_i \lesssim T$.

Finally then, let us see what would happen if we use a decomposition of the thermal ensemble in which this overlap with the energy eigenstates is maximal. Consider the overcomplete, unnormalized basis of states

$$|\Theta\rangle = \sum_i e^{i\theta_i} |\epsilon_i\rangle \quad (37)$$

each parameterized by the set $\Theta = \{\theta_i\}$. Then the normalized states given by

$$|\phi(\Theta)\rangle = \frac{1}{\sqrt{Z}} e^{-\beta H/2} |\Theta\rangle = \frac{1}{\sqrt{Z}} \sum_i e^{i\theta_i} e^{-\beta \epsilon_i/2} |\epsilon_i\rangle \quad (38)$$

form a decomposition of a thermal ensemble with uniform weights $P(\Theta) = \text{const}$. If we now imagine sampling even just one such state, the resulting estimate of the energy would be exact,

since

$$\begin{aligned} \langle \phi(\Theta) | H | \phi(\Theta) \rangle &= \frac{1}{\mathcal{Z}} \sum_{i,j} e^{i(\theta_j - \theta_i)} e^{-\beta(\epsilon_j + \epsilon_i)/2} \langle \epsilon_j | H | \epsilon_i \rangle \\ &= \frac{1}{\mathcal{Z}} \sum_i e^{-\beta \epsilon_i} \epsilon_i \\ &= \langle H \rangle. \end{aligned} \quad (39)$$

In a similar manner, one may obtain the exact specific heat from just one of these states as well, which implies that $C_v^{(q)} = C_v$ for this decomposition. Formally, then, this particular basis is the most optimal one for sampling the average energy of the system. However, constructing even one state $|\Theta\rangle$ amounts to diagonalizing the entire Hamiltonian, which would render a sampling approach irrelevant.

In the end, then, the METTS decomposition achieves both a remarkably high sampling efficiency in practice, as quantified by $C_v - C_v^q$, while also being one of the least costly decompositions to produce.

3.3. In what sense are METTS minimally entangled?

Consider a system consisting of just two spin-1/2 moments, A and B in a pure state $|\psi\rangle$. We call $|\psi\rangle$ entangled if it cannot be factorized, that is, if it is not a CPS.

To compute the entanglement entropy $S[\psi]$, one first forms the density matrix ρ_A describing spin A alone by tracing out the states of B

$$\rho_A = \text{Tr}_B |\psi\rangle\langle\psi|. \quad (40)$$

One may define ρ_B likewise. In terms of these reduced density matrices, $S[\psi]$ is then

$$S[\psi] = -\text{Tr}[\rho_A \log_2(\rho_A)] = -\text{Tr}[\rho_B \log_2(\rho_B)]. \quad (41)$$

At finite temperature, however, A and B are not in a pure state $|\psi\rangle$, but in the mixed state described by the density matrix $\rho = \frac{1}{\mathcal{Z}} e^{-\beta H}$. While for a mixed state one can no longer define an entanglement entropy in the same manner, one could choose a particular decomposition of ρ as in equation (33) and compute its average entanglement entropy

$$S_\phi[\rho] = \sum_i \frac{P(i)}{\mathcal{Z}} S[\phi(i)]. \quad (42)$$

However, $S_\phi[\rho]$ turns out to depend upon which decomposition we choose.

A more meaningful measure of mixed state entanglement is the *entanglement of formation* $E[\rho]$ defined as the minimum value of $S_\phi[\rho]$ among all possible decompositions [9]

$$E[\rho] = \min_\phi S_\phi[\rho]. \quad (43)$$

While such a minimization would be difficult to compute in general, an exact expression has been found by Wootters for the case of a two-qubit system [10]. So, for this simple system at least, we may unambiguously check whether the average entanglement entropy of the METTS decomposition saturates the lower bound set by $E[\rho]$. In other words, we may explicitly check if METTS are minimally entangled.

Before we make this comparison, however, let us review Wootters' method for computing the entanglement of formation. First define the concurrence of a pure state $|\psi\rangle$ of two-qubits as

$$C[|\psi\rangle] = |\langle\psi|\tilde{\psi}\rangle|, \quad (44)$$

where $|\tilde{\psi}\rangle$ is the time reversal conjugate of $|\psi\rangle$. Concurrence is an alternative measure of entanglement and may be used to compute the standard von Neumann entanglement entropy equation (41) through the relation

$$S[|\psi\rangle] = f\left(\frac{1 + \sqrt{1 - (C[|\psi\rangle])^2}}{2}\right), \quad (45)$$

where $f(x) = -x\log_2(x) - (1-x)\log_2(1-x)$.

Now for a general mixed state of two-qubits, it may be shown that there exists at least one optimal decomposition of ρ which saturates the minimum in equation (43) and which consists of pure states all having the *same* concurrence and hence the same entanglement entropy. This minimal concurrence value is given by

$$C[\rho] = \max\{0, \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4\}, \quad (46)$$

where λ_i are the square roots of the eigenvalues of the matrix $\rho(\sigma_y \otimes \sigma_y)\rho^*(\sigma_y \otimes \sigma_y)$. It therefore follows that the entanglement of formation of a two-qubit system is given by

$$E[\rho] = f\left(\frac{1 + \sqrt{1 - (C[\rho])^2}}{2}\right). \quad (47)$$

Equipped with this explicit formula for the entanglement of formation, let us now consider a specific system, namely the two-site $S = 1/2$ Heisenberg model with the Hamiltonian

$$H = J \vec{S}_A \cdot \vec{S}_B \quad (48)$$

and take the antiferromagnetic $J > 0$ case.

As $\beta \rightarrow \infty$, the unique ground state $|\psi_0\rangle$ of H is the singlet formed out of the spins A and B , which has an entanglement entropy $S[|\psi_0\rangle] = 1$. So, the entanglement of formation must also be $E[\rho] = 1$ at zero temperature. In the opposite limit as β tends to zero, $E[\rho]$ must tend to zero. As shown in figure 12, it actually reaches zero at a finite temperature of $\beta J = \ln(3)$. Thus for any smaller value of βJ , the density matrix is separable and may be decomposed entirely in terms of CPS [11].

Now let us compute the average entanglement of a METTS decomposition of this system. To do so, we must first choose the underlying CPS $|i\rangle$ to be used in producing our METTS. One possible choice is to take the set $|i\rangle$ to be the complete basis of \hat{z} eigenstates. Because this yields a decomposition consisting of only four METTS we can calculate them exactly. The resulting average entanglement entropy of this 'ZZ METTS' ensemble is plotted in figure 12.

Note that we could have also used the pure state algorithm to sample this decomposition by collapsing each METTS into the \hat{z} basis on every site and then the \hat{x} basis on every site on alternate steps. The METTS produced after either type of collapse can then be used to measure the entanglement since the system is rotationally invariant. Collapsing into the \hat{z} basis alone, on the other hand, will not sample all four METTS since the Hamiltonian conserves total S^z .

We may also choose CPS bases other than the S^z eigenstates and thereby find METTS decompositions with even lower average entanglement. Consider the XZ METTS derived from

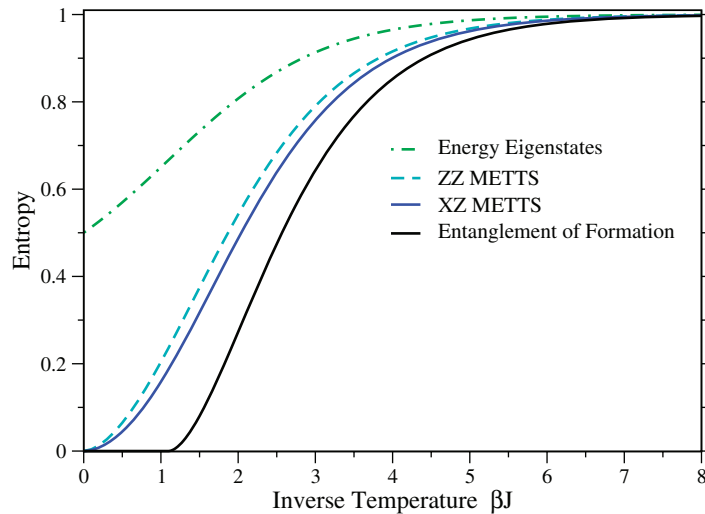


Figure 12. The average entanglement of the ZZ and XZ METTS density matrix decompositions for the two-site $S = 1/2$ Heisenberg antiferromagnet. The XZ METTS form the optimal METTS decomposition of the system, but their average entropy exceeds the entanglement of formation $E[\rho]$. Also shown is the average entanglement of the energy eigenstate decomposition.

the orthonormal CPS basis $\{|\uparrow\rightarrow\rangle, |\uparrow\leftarrow\rangle, |\downarrow\rightarrow\rangle, |\downarrow\leftarrow\rangle\}$, where up and down arrows denote S^z eigenstates and left and right arrows denote S^x eigenstates. As shown in figure 12, the average entanglement of these METTS is uniformly lower than the ZZ METTS.

What is more, the variance of both energy and entropy is exactly zero for the XZ METTS ensemble. This follows from the fact that any XZ METTS can be transformed into any other by global spin rotations that commute with H . Therefore we can conclude that the XZ METTS are the optimal METTS decomposition for this system, at least up to a global spin rotation. Note that one may also sample the XZ METTS with the pure state method by collapsing them into the \hat{x} and \hat{z} bases on alternating steps *and* on alternating sites.

Regardless of the basis we choose, however, one can see from figure 12 that the average entanglement of each METTS decomposition tends to zero smoothly as $\beta \rightarrow 0$ and always exceeds $E[\rho]$. Therefore, because even the optimal METTS decomposition is greater than $E[\rho]$ for any finite temperature, we can conclude that METTS are *not* minimally entangled.

However, we may ask if the optimal METTS decomposition is instead minimally entangled in a more limited sense. For this purpose it will be helpful to look at a special class of thermal density matrix decompositions. Beginning with a decomposition at some particular temperature β_0

$$\rho_0 = \frac{e^{-\beta_0 H}}{\mathcal{Z}_0} = \sum_i p_i |\psi_i\rangle\langle\psi_i|, \quad (49)$$

one can generate a decomposition for ρ at a different temperature β by making use of the relation

$$\begin{aligned} \rho &= \frac{e^{-\beta H}}{\mathcal{Z}} = \frac{\mathcal{Z}_0}{\mathcal{Z}} e^{-(\beta-\beta_0)H/2} \rho_0 e^{-(\beta-\beta_0)H/2} \\ &= \sum_i \frac{p_i \mathcal{Z}_0}{\mathcal{Z}} Q(i) |\psi_i(\beta)\rangle\langle\psi_i(\beta)|, \end{aligned} \quad (50)$$

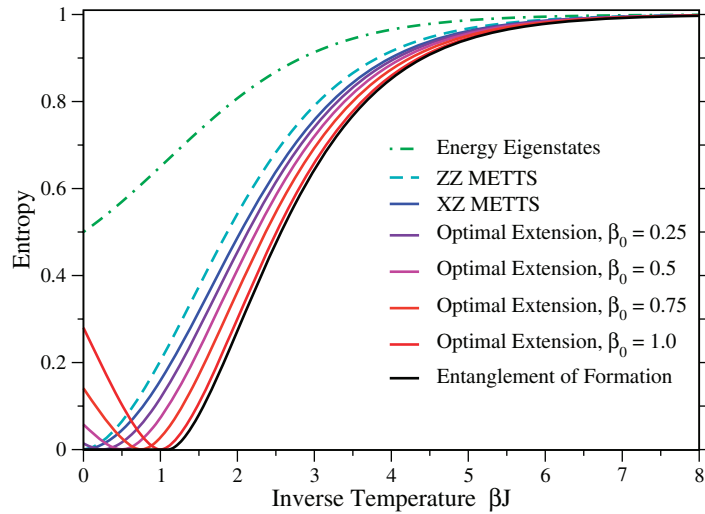


Figure 13. The average entanglement entropy of extended decompositions of the two-site $S = 1/2$ Heisenberg antiferromagnet. Each optimal extension reaches the minimum allowed entanglement at some temperature β_0 , but still exceeds the entanglement of the optimal XZ METTS at high enough temperatures. Also shown are the entanglements of the ZZ METTS decomposition and the energy eigenstate decomposition.

where the pure states defining the new decomposition are given by

$$\begin{aligned} |\psi_i(\beta)\rangle &= e^{-(\beta-\beta_0)H/2} |\psi_i\rangle / Q(i)^{1/2}, \\ Q(i) &= \langle \psi_i | e^{-(\beta-\beta_0)H} | \psi_i \rangle. \end{aligned} \quad (51)$$

We may consider this new decomposition the *extension* of the decomposition of equation (49), or an extended decomposition for short. Note that the entanglement of an extended decomposition is always an analytic function of β . For our two-qubit system then, this implies that no extended decomposition can have an entanglement equal to $E[\rho]$ at all temperatures.

In fact, we would like to conjecture that no extended decomposition can even have an entanglement below the optimal METTS decomposition at all temperatures. To see why this may be the case, let us attempt to construct a counterexample for the two-qubit system. At any temperature β_0 , we can construct an optimal decomposition, following Wootters, such that the entropy of each state $|\psi_i\rangle$ is equal to the entanglement of formation. Then, because the extensions of these decompositions will have the least possible entanglement at β_0 , we expect that they will also have low entanglement at other temperatures.

Now, there can be multiple optimal decompositions for each starting temperature β_0 . However, we have found that there is always one whose extension is less entangled than the others and has zero entropy variance for all temperatures, not just β_0 . As shown in figure 13, while such optimal extensions have less entanglement than the optimal METTS at low temperatures, they eventually exceed the optimal METTS as the temperature is increased. Moreover, the less entangled these extensions become at high temperature, the more they resemble the optimal METTS ensemble itself.

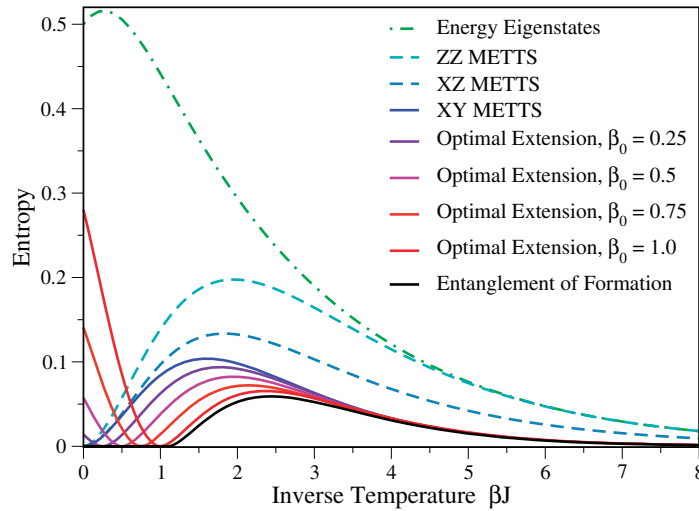


Figure 14. The average entanglement entropy of extended decompositions of the two-site $S = 1/2$ Heisenberg antiferromagnet with an applied field $h = 1.5$ J. In the presence of the field, the optimal METTS decomposition is given by the XY METTS. Also shown are the entanglement of the XZ and ZZ METTS decompositions and the energy eigenstate decomposition.

To check that this behavior is general, we can also study the two-qubit system in a field with the Hamiltonian

$$H = J \vec{S}_A \cdot \vec{S}_B - h (S_A^z + S_B^z). \quad (52)$$

Here we focus on the case $J > 0$ and $h \geq 0$. For small fields, the temperature dependence of the entanglement of formation $E[\rho]$ remains similar to the zero field case. Then at $h = J$, the system undergoes a first-order quantum phase transition such that the ground state is no longer a maximally entangled singlet, but the fully polarized CPS with zero entanglement. Once h is greater than J , the entanglement of formation no longer behaves as a monotonic function of temperature. Regardless of the value of h , however, $E[\rho]$ is strictly zero for $\beta J \leq \ln(3)$, indicating that ρ becomes separable above this temperature.

In figure 14, we have taken a particular value of the field $h > J$ and plotted the entropies of various optimal extensions. As before, each extension either exceeds the optimal METTS entanglement over a significant temperature range or else approaches the optimal METTS decomposition ever more closely as this range is reduced. Also, it is interesting to note that in the presence of a field it is the XY METTS (generated from S^x and S^y eigenstates) that now form the optimal METTS decomposition, at least up to a rotation about the \hat{z} -axis. This is because the XZ METTS are no longer required to be physically identical by the symmetries of the Hamiltonian and will exhibit a finite entropy variance.

Going forward, it will be interesting to explore the extent to which the observations made here hold for larger systems and more complex models. It would be especially useful to know if an optimal METTS decomposition with ideal sampling efficiency always exists and whether it can be efficiently computed.

Although the average entanglement entropy of a METTS decomposition is not an entanglement measure in a strict sense (for instance, it is not zero when ρ is separable) [12],

it is conceivable that its scaling behavior could reveal nontrivial information about a system. Whether or not one must find an optimal METTS decomposition to access such information remains to be seen. But because the METTS pure state algorithm is efficient enough to simulate moderately large 2D systems, the entanglement properties of METTS could turn out to be a useful tool for characterizing exotic phases or critical points.

Acknowledgments

EMS thanks Leon Balents for making this work possible and for providing support through NSF grant no. DMR-0804564. EMS also acknowledges helpful discussions with Thomas Barthel and Ryan Mishmash and thanks the 2009 ICTS Condensed Matter Programme for hospitality. SRW acknowledges support from NSF grant no. DMR-0907500. The quantum Monte Carlo calculations appearing in this work were performed using the ALPS looper application [13]. Thanks are due to Microsoft Station Q for providing the computing resources for these calculations.

Appendix. Working with MPSs and MPOs

An MPS is a scheme for representing a wavefunction, either exactly or approximately, by writing its amplitudes in some particular basis as a product of matrices. A general MPS may be written as

$$|\psi_A\rangle = \sum_{\{s\}} A^{s_1} A^{s_2} \dots A^{s_N} |s_1\rangle |s_2\rangle \dots |s_N\rangle \quad (\text{A.1})$$

where the site labels s_i take on values $s_i = 1, 2, \dots, d$ labeling the local basis.

For an adjacent pair of matrices $A_{\mu\nu}^{s_i} A_{\nu\lambda}^{s_{i+1}}$, if the index ν summed between the matrices takes on values $1, 2, \dots, m$ one says that the bond between matrices i and $i+1$ has bond dimension m . Although m can vary from bond to bond, in the following discussion we assume that it has a fixed value for a given MPS. While in the presence of periodic boundary conditions one must trace over the above product of matrices in order to reduce it to a scalar amplitude, in what follows we work exclusively with open boundary conditions such that A^{s_1} and A^{s_N} are just $1 \times m$ and $m \times 1$ matrices, respectively.

An MPO is defined similarly to an MPS except that the decomposition is over a basis of local operators. A general MPO may therefore be written as

$$\hat{W} = \sum_{\{t,s\}} W^{t_1 s_1} \dots W^{t_N s_N} |t_1\rangle \dots |t_N\rangle \langle s_1| \dots \langle s_N| \quad (\text{A.2})$$

where for open boundary conditions the dimensions of the first and the last matrix are again $1 \times m$ and $m \times 1$.

A.1. Creating MPOs

The simplest MPOs to create are those describing products of local operators, that is, $\hat{W} = \Pi_j \hat{O}_j$. In particular, one can represent a single site operator \hat{O}_i this way by taking $\hat{O}_j = 1$ for all $j \neq i$. If the operator \hat{W} is such a site-factorized product, its MPO representation has a simple product structure as well. Each $W^{t_i s_i}$ is actually just a 1×1 matrix defined by $W^{t_i s_i} = \langle t_i | \hat{O}_i | s_i \rangle$.

Another simple class of MPOs is those describing translationally invariant sums of operators in one dimension in which case the matrices W can be taken to be identical (except at the boundaries) and of lower triangular form. For example, the MPO of the operator $S_{\text{tot}}^z = \sum_{i=1}^N S_i^z$ may be represented in terms of the bulk matrices

$$W^{t_i s_i} = \begin{bmatrix} \delta^{t_i s_i} & 0 \\ (S_i^z)^{t_i s_i} & \delta^{t_i s_i} \end{bmatrix}$$

together with the boundary matrices $W^{t_1 s_1} = [0 \ \delta^{t_1 s_1}]$ and $W^{t_N s_N} = [\delta^{t_N s_N} \ 0]^T$. Translationally invariant 1D Hamiltonians can likewise be represented by translationally invariant lower triangular MPOs of bond dimension 3 and higher [14].

However, there may not always be such a simple, explicit prescription for operators of interest, in particular Hamiltonians of 2D systems. For use in such cases, we present below general algorithms for the addition and multiplication of arbitrary MPOs.

Equipped with these algorithms, the MPO representation of any Hamiltonian that is a sum of local terms may be found by first obtaining the MPOs for each term as described above and then summing them together. The MPO multiplication algorithm is then useful for forming other important operators such as H^2 for specific heat measurements or $e^{-\tau H}$ for imaginary time evolution.

A.2. Adding MPS and MPOs

First consider adding two MPS $|\psi_A\rangle$ and $|\psi_B\rangle$ with bond dimensions m_A and m_B . If $|\psi_C\rangle = |\psi_A\rangle + |\psi_B\rangle$, then

$$C^{s_i} = A^{s_i} \oplus B^{s_i}, \quad (\text{A.3})$$

which in block form is

$$C^{s_i} = \begin{bmatrix} A^{s_i} & 0 \\ 0 & B^{s_i} \end{bmatrix}. \quad (\text{A.4})$$

More formally, one may write this definition of C^{s_i} as

$$C_{\chi_{i-1}\chi_i}^{s_i} = \sum_{\alpha_{i-1}, \alpha_i, \beta_{i-1}, \beta_i} \delta_{\chi_{i-1}\alpha_{i-1}} A_{\alpha_{i-1}\alpha_i}^{s_i} \delta_{\alpha_i\chi_i} + \delta_{\chi_{i-1}(\beta_{i-1}+m_A)} B_{\beta_{i-1}\beta_i}^{s_i} \delta_{(\beta_i+m_A)\chi_i}, \quad (\text{A.5})$$

where the χ indices run over values $1, 2, \dots, (m_A + m_B)$.

After constructing the C matrices as above, the addition has been carried out exactly. However, one usually wants to work with an MPS for $|\psi_C\rangle$ that has a bond dimension comparable to m_A or m_B , not $m_A + m_B$. One therefore usually truncates the MPS representation for $|\psi_C\rangle$ as follows.

Treating $C_1 = C_{\chi_1}^{s_1}$ as a matrix $C_{s_1\chi_1}$, compute its SVD

$$C_{s_1\chi_1} = U_{s_1\chi'_1} \Lambda_{\chi'_1} V_{\chi'_1\chi_1}^\dagger. \quad (\text{A.6})$$

During this step, only m_C singular values are kept such that $\chi'_1 = 1, 2, \dots, m_C$. Finally, replace C_1 with $C_{\chi'_1}^{s_1} = U_{s_1\chi'_1}$ and replace C_2 with $C_{\chi'_1\chi_2}^{s_2} = \Lambda_{\chi'_1} V_{\chi'_1\chi_1}^\dagger C_{\chi_1\chi_2}^{s_2}$.

The next bond in the truncated MPS for $|\psi_C\rangle$ is now found by treating the new C_2 as a matrix $C_{(s_2\chi'_1)\chi_2}$ and computing its SVD. The matrix C_2 is then set to U_2 and C_3 is multiplied by $\Lambda_2 V_2^\dagger$ just as C_2 was in the first step. This truncation process may then be likewise repeated for every remaining C matrix.

The addition of MPOs proceeds in complete analogy to the addition of MPS. To compute the MPO \hat{Z} such that $\hat{Z} = \hat{W} + \hat{X}$, one defines the matrices Z as

$$Z^{s_i t_i} = W^{s_i t_i} \oplus X^{s_i t_i}, \quad (\text{A.7})$$

where the only difference from the MPS case is that there are now d^2 matrices to be added instead of d . Thinking of the pair of indices s_i and t_i as a fat index $(s_i t_i)$, one may likewise truncate the matrices Z by performing sequential SVDs as above.

The MPS addition algorithm above has a computational cost of Nm^3d^2 for $m_A = m_B = m_C = m$. Similarly, the addition of MPOs scales as Nk^3d^4 , where k is the MPO bond dimension.

A.3. Multiplying MPOs

Consider the task of multiplying two MPOs \hat{W} and \hat{X} with the goal of producing a product MPO $\hat{Z} = \hat{W}\hat{X}$. Let these MPOs be written as products of W , X and Z matrices with bond dimension k .

This multiplication may be carried out one site at a time in a process analogous to the zip-up algorithm discussed above. Therefore, the first step is to map each MPO to an MPS by combining its t and s (or primed and unprimed) indices at each site into a fat index. These MPS should then be transformed such that their orthogonality center is at the first site and then mapped back into MPOs. In doing so, one ensures that in the truncation step below, the basis of states surrounding the local tensor is not overly ill-conditioned. However, because this basis will not be orthonormal, it is important to use a truncation error cutoff instead of a fixed k to determine the final MPO bond dimension.

Now, to carry out the multiplication beginning with the first site, define the tensor R_1 as

$$R_{\omega_1 \xi_1}^{s_1 t_1} = \sum_{u_1} W_{\omega_1}^{t_1 u_1} X_{\xi_1}^{u_1 s_1}. \quad (\text{A.8})$$

Thinking of R_1 as a matrix $R_{(s_1 t_1)(\omega_1 \xi_1)}$, compute its SVD to obtain

$$R_{(s_1 t_1)(\omega_1 \xi_1)} = U_{(s_1 t_1)\zeta_1} \Lambda_{\zeta_1} V_{\zeta_1(\omega_1 \xi_1)}^\dagger, \quad (\text{A.9})$$

where the singular values are truncated such that only the k largest are kept. The first matrix of \hat{Z} may now be defined as $Z_{\zeta_1}^{s_1 t_1} = U_{(s_1 t_1)\zeta_1}$.

The remaining Z matrices may now be found by proceeding recursively. Define R_2 in terms of R_1 as

$$R_{\zeta_1 \omega_2 \xi_2}^{s_2 t_2} = \sum_{s_1 t_1} U_{\zeta_1(s_1 t_1)}^\dagger \left[\sum_{\xi_1 u_2} \left[\sum_{\omega_1} R_{\omega_1 \xi_1}^{s_1 t_1} W_{\omega_1 \omega_2}^{t_2 u_2} \right] X_{\xi_1 \xi_2}^{u_2 s_2} \right]. \quad (\text{A.10})$$

Again, R_2 may be thought of as a matrix and its SVD computed as

$$R_{(s_2 t_2 \zeta_1)(\omega_2 \xi_2)} = U_{(s_2 t_2 \zeta_1)\zeta_2} \Lambda_{\zeta_2} V_{\zeta_2(\omega_2 \xi_2)}^\dagger. \quad (\text{A.11})$$

The second matrix of \hat{Z} may now be defined as $Z_{\zeta_1 \zeta_2}^{s_2 t_2} = U_{(s_2 t_2 \zeta_1)\zeta_2}$ and the algorithm continued by defining R_3 analogously to R_2 above in terms of U_2 , R_2 , W_3 and X_3 .

If the contractions in equation (A.10) are carried out in the order indicated, this algorithm for multiplying MPOs scales as Nk^4d^3 with the most expensive step being the construction of the R tensors.

References

- [1] White S R 2009 *Phys. Rev. Lett.* **102** 190601
- [2] Affleck I, Kennedy T, Lieb E H and Tasaki H 1987 *Phys. Rev. Lett.* **59** 799
- [3] Haldane F D M 1983 *Phys. Rev. Lett.* **50** 1153
- [4] Pollmann F, Turner A M, Berg E and Oshikawa M 2010 *Phys. Rev. B* **81** 064439
- [5] Feiguin A E and White S R 2005 *Phys. Rev. B* **72** 220401
- [6] White S R and Feiguin A E 2004 *Phys. Rev. Lett.* **93** 076401
- [7] Feiguin A E and White S R 2005 *Phys. Rev. B* **72** 020404
- [8] Verstraete F and Cirac J 2004 arXiv:cond-mat/0407066
- [9] Bennett C H, DiVincenzo D P, Smolin J and Wootters W K 1996 *Phys. Rev. A* **54** 3824
- [10] Wootters W K 1998 *Phys. Rev. Lett.* **80** 2245
- [11] Arnesen M C, Bose S and Vedral V 2001 *Phys. Rev. Lett.* **87** 017901
- [12] Vedral V, Plenio M B, Rippin M A and Knight P L 1997 *Phys. Rev. Lett.* **78** 2275
- [13] Albuquerque A F *et al* 2007 *J. Magn. Magn. Mater.* **310** 1187 (see also <http://alps.comp-phys.org/>)
- [14] McCulloch I P 2007 *J. Stat. Mech.* **2007** P10014