Probabilistic low-rank factorization accelerates tensor network simulations of critical quantum many-body ground states

Lucas Kohn,1 Ferdinand Tschirsich,1,* Maximilian Keck,2 Martin B. Plenio,3 Dario Tamascelli,3,4 and Simone Montangero1,5,6

1Institute for Complex Quantum Systems and Center for Integrated Quantum Science and Technologies, Universität Ulm, 89069 Ulm, Germany
2NEST, Scuola Normale Superiore and Istituto Nanoscienze CNR, 56126 Pisa, Italy
3Institute for Theoretical Physics, Universität Ulm, 89069 Ulm, Germany
4Dipartimento di Fisica, Università degli Studi di Milano, 20133 Milano, Italy
5Theoretische Physik, Universität des Saarlandes, 66123 Saarbrücken, Germany
6Dipartimento di Fisica e Astronomia, Università degli Studi di Padova, 35131 Padova, Italy

(Received 11 October 2017; published 3 January 2018)

We provide evidence that randomized low-rank factorization is a powerful tool for the determination of the ground-state properties of low-dimensional lattice Hamiltonians through tensor network techniques. In particular, we show that randomized matrix factorization outperforms truncated singular value decomposition based on state-of-the-art deterministic routines in time-evolving block decimation (TEBD)– and density matrix renormalization group (DMRG)–style simulations, even when the system under study gets close to a phase transition: We report linear speedups in the bond or local dimension of up to 24 times in quasi-two-dimensional cylindrical systems.

DOI: 10.1103/PhysRevE.97.013301

I. INTRODUCTION

Tensor network (TN) methods have long proven their power as indispensable tools in simulating quantum and classical many-body systems [1,2]. As first realized by White with the density matrix renormalization group (DMRG) algorithm [3,4], a variational ansatz on the manifold of matrix product states (MPS) [5], TNs provide an efficient parametrization of low-entangled wave functions in quantum many-body state space [6]. While the MPS naturally captures the relevant low-energy spectrum in particular of one-dimensional (1D) gapped Hamiltonians obeying area laws of entanglement [7–10], TNs have been generalized to more complex scenarios: In over two decades of evolution, they have been successfully applied to higher dimensions [11–13], critical phenomena [14–17], finite temperature and closed- and open-system dynamics [18,19], and lattice gauge theories [20–22], just to name a few examples. TNs have also been equipped with structure to encode and exploit symmetries in the model under investigation [23–26].

Truncated singular value decompositions (SVDs) are widely used in TN algorithms to compress states into their respective TN state manifold. Examples include the time-evolving block decimation (TEBD) [27,28], the tensor renormalization group [29], the corner transfer matrix renormalization group (CTMRG) [30], and their application to projected entangled pair states (PEPS) [31–34], but also traditional DMRG which is often formulated in terms of truncated eigenvalue decomposition. In TN numerical practice, SVDs have the additional advantage of providing relevant isometries by orthonormality of the singular vectors and reveal valuable information of the encoded network state, e.g., in the form of entanglement measures based on singular values.

The traditional way to compute a truncated SVD is to first perform the full SVD of a matrix and then discard the smallest singular values. This is reliable and accurate, but also a very costly operation that often dominates computational complexity of TN algorithms. Intuitively, it is also not the most economic protocol: A lot of effort is spent in computing all singular values and vectors, many of which are then discarded. By avoiding the full SVD, a truncated SVD can be obtained more efficiently, especially when the number of retained singular values is small. Well-known methods of this class are simultaneous subspace iteration or Krylov subspace methods like Lanczos or implicitly restarted Arnoldi algorithms [35,36]. Their relevance in large-scale data classification and compression in “big data” applications [37,38], signal processing [39], face recognition [40,41], DNA analysis [42], and other fields is a driving force behind the ongoing development of faster algorithms. A use case in the approximative contraction of unstructured TNs has also been reported [43].

Randomized algorithms outperform prior approaches in both speed and reliability [44]. Specifically, the randomized SVD (RSVD) based on a probabilistic low-rank matrix-factorization algorithm [44] is capable of delivering accurate results with failure probabilities that can be made arbitrarily small, independent of peculiar choices like starting vectors that are common in deterministic methods. RSVD thus promises to significantly accelerate TN methods that spend a considerable amount of resources in truncated SVDs.

Recently, significant speedup due to RSVD has been reported in the TEBD simulation of open-system dynamics [45]. In particular, the authors of Ref. [45] showed that the robust RSVD outperforms deterministic SVD algorithms in
delivering a limited number of largest singular values (and corresponding vectors) while maintaining high accuracy in the simulated dynamics. It is, however, an open question of whether RSVD can be applied with similar success in scenarios beyond the open-system dynamics, since RSVD performance and accuracy are closely tied to the encountered spectra of singular values. This question applies especially to critical systems where the singular values are expected to decay slowly.

In this paper, we demonstrate superior performance of RSVD in the very original application field of TN methods, namely in identifying ground-state properties of low-dimensional quantum lattice Hamiltonians. We confirm significant speedup in different physical scenarios, including situations when the system is critical. Embedded in full-fledged TN simulations, we compare the RSVD against the truncated full SVD from state-of-the-art LAPACK implementations DZGESDD [46], referred to as TSV D in the following. As benchmarks, we use variants of the quantum Ising model in imaginary TEBD time evolution and a DMRG-style ground-state search with the hierarchical binary tree TN (TTN) [16,47–52]. It will become apparent that a simple replacement of state search with the hierarchical binary tree TN (TTN) [16,47–52], referred to as TSV D in the following. As benchmarks, we use variants of the quantum Ising model in imaginary TEBD time evolution and a DMRG-style ground-state search with the hierarchical binary tree TN (TTN) [16,47–52]. It will become apparent that a simple replacement of state search with the hierarchical binary tree TN (TTN) [16,47–52], referred to as TSV D in the following. As benchmarks, we use variants of the quantum Ising model in imaginary TEBD time evolution and a DMRG-style ground-state search with the hierarchical binary tree TN (TTN) [16,47–52]. It will become apparent that a simple replacement of state search with the hierarchical binary tree TN (TTN) [16,47–52]. It will become apparent that a simple replacement of state search with the hierarchical binary tree TN (TTN) [16,47–52]. It will become apparent that a simple replacement of state search with the hierarchical binary tree TN (TTN) [16,47–52].

The paper is organized as follows: First, we explain the use of truncated SVD as a tool of information compression in typical TN scenarios in Sec. II. We continue with a short review of the RSVD method and how it can help achieve faster compression in Sec. II B. We then introduce our benchmark models in Sec. III and present a detailed performance analysis by switching from TSV D to RSVD in Sec. IV. Section V concludes the paper with a discussion of the results and with practical tips for the implementation and identification of situations that may benefit from RSVD.

II. LOW-RANK FACTORIZATION

The maximal bond dimension $\chi$ of a TN is a fundamental parameter: It can be linked to the amount of quantum entanglement that can be hosted in the network state [6]. At the same time, $\chi$ determines the computational complexity of algorithms performed on the network. Typical operations include the computation of expectation values, propagation in real or imaginary time, and renormalization steps updating the network description in iterative algorithms. All these operations can result in the growth of index dimensions beyond the maximally allowed bond dimension. A compression step is then achieved by means of a truncated SVD.

A. Truncated SVD

Let $A$ be a real- or complex-valued $m$-by-$n$ matrix with $m \geq n$. In our case, $A$ usually represents the contraction of two tensors, and it can also be given in the form of a matrix product $X'Y$. The compression step then provides a rank-$\chi$ factorization $X'Y$ which is a good approximation $A \approx X'Y$, but also limits $X$ to an $m$-by-$\chi$ matrix and $Y$ to a $\chi$-by-$n$ matrix. A standard solution is to compute the rank-$\chi$ truncated SVD as follows:

\[
\text{Algorithm RSVD}
\]

| Input | $m$-by-$n$ matrix $A$, integers $\chi, \ell, q$ |
|-------|------------------------------------------|
| Output | approximate rank-$\chi$ truncated SVD of $A$ |

1. Generate an $n$-by-$\ell$ Gaussian matrix $\Omega$
2. Compute $Y := (AA^\dagger)^qA\Omega$
3. Store in $Q$ the orthonormalized columns of $Y$
4. Compute the rank-$\chi$ truncated SVD of $B := Q^\dagger A$

In detail, the algorithm begins by drawing a random test matrix $\Omega$ from a standard Gaussian distribution in step 1. Note that other choices may work as well and that the quality of random numbers is not of crucial importance. Step 2 then produces an $m$-by-$\ell$ sample $Y$ of the range of $A$, by multiplying the columns of the test matrix with $(AA^\dagger)^qA$.
This process emphasizes the most relevant singular vectors, associated with large singular values \( \sigma \), by a factor of \( \sigma^{2q+1} \). In order to maintain numerical stability when these factors range over several orders of magnitude, step 2 is carried out as a power iteration with subsequent QR factorizations to keep the sample orthonormal (see Ref. [44], Algorithm 4.4). Step 3 then provides a basis of the sampled, relevant contributions to the range of A in the orthonormal columns of the \( m \)-by-\( \ell \) matrix \( Q \). A rank-\( \ell \) approximation of A is now available by projection into that subspace: \( A_{\ell} := QQ^*A \). Such an explicit construction is, however, not required. Instead, step 4 invokes a rank-\( \chi \) TSVD factorization \( \hat{U}\Sigma V^\dagger \) of the typically much smaller \( \ell \)-by-\( n \) matrix \( B := Q^*A \). Due to \( \chi \) is much smaller than \( \ell \), the \( \chi \) largest singular values of \( A \) are approximated in \( \Sigma \). If required, approximate associated left and right singular vectors of \( A \) are given by \( \hat{Q}U \) and \( \hat{V} \), respectively. Both are exact isometries.

The average RSVD compression error \( \varepsilon_{\text{RSVD}} := \mathbb{E}(\|A - \hat{Q}U\hat{V}\|^2) \) depends on the spectrum of singular values and can be made arbitrarily close to the minimal truncation error \( \delta_{\text{trunc}} \) in either Frobenius or spectral norm: Following the analysis in Ref. [57], a minimal oversampling of \( \ell \geq \chi + 2 \) already guarantees

\[
\varepsilon_{\text{RSVD}} \leq \sqrt{\delta_{\text{trunc}}^2 + C^2(n, \ell) \delta_{\text{trunc}}^2 (\sigma_{\ell-1}/\sigma_{\chi})^{4q}}, \tag{1}
\]

where \( C^2(n, \ell) = (4e/3)^2(\sqrt{n - \ell + 2} + \sqrt{\chi + 7})^2 \delta \) (\( e \) being Euler’s number) is of complexity \( O(n\ell) \), i.e., bounded by a bilinear growth in \( n \) and \( \ell \) from above, with \( \ell \leq n \). Note that \( \delta_{\text{trunc}} \) depends on the selected norm, unlike the additional terms introduced by the randomized approach. While highest accuracy is expected for quickly decreasing singular values, a striking feature of RSVD is that already small powers \( q > 0 \) drive those contributions, which add to \( \delta_{\text{trunc}} \) in Eq. (1), to zero exponentially fast, even in cases of slowly decaying singular values. Furthermore, sufficient oversampling in \( \ell \) makes the probability of a substantial deviation from the average error bound arbitrarily small [57].

Throughout our benchmarks, we make the conservative choice \( \ell = 2\chi \), which is suitable to keep the RSVD error within a small factor of \( \delta_{\text{trunc}} \) even for \( q = 0 \) [44]. In this configuration, RSVD promises an asymptotic speedup over TSVD in the order of the compression ratio

\[
T_T/T_R \sim n/\chi, \tag{2}
\]

where \( T_T \) and \( T_R \) are the times required by TSVD and RSVD to decompose the same input matrix \( A \). The proportionality is due to the lower RSVD computational complexity, which is dominated by the matrix-matrix products of \( O(nm\ell(q + 1)) \) in sampling \( Y \).

The improved scaling of the RSVD algorithm is complemented by its conceptual simplicity, which directly translates to a fast, stable, and easily parallelizable implementation in terms of highly optimized linear algebra routines as provided by level-3 BLAS and LAPACK [46]. Various RSVD implementations are available, for instance in MATLAB [58], in R [59], and via C libraries such as RSVDPACK [60] or RRSVD-RACKAGE [45], which our benchmarks are based on.

### III. BENCHMARKS

We benchmark RSVD against TSVD performance, when employed in state-of-the-art TN algorithms. Our focus lies on closed-system ground states, and we compare both run time and precision of the relevant physical quantities in the outcomes.

We first outline the TN algorithms that drive our benchmark simulations and the role played by compression. Afterward, we report model Hamiltonians and parameters. We close this section with a brief account on the numerical implementation.

#### A. TN algorithms

We employed TEBD imaginary time evolution on MPSs and DMRG-style variational ground-state search in the TTN [51] with double-tensor optimization. Both algorithms are well-established techniques in ground-state search of quantum lattice Hamiltonians. They iteratively approximate those ground states in TNs of a selected maximal bond dimension \( \chi \), defined over \( d \)-dimensional “physical” tensor indices that correspond to lattice sites [see Figs. 1(a) and 1(d)]. Specifically, both algorithms perform local update steps on adjacent tensors, which require a truncated SVD to reconstruct bond indices. Note that it is the absence of loops (network cycles) in MPS and TTN geometries that makes truncated SVD an optimal protocol here, as it maintains maximum quantum fidelity between the states before and after the compression of a single bond [3,4,61].

The two methods, however, rely on different local update steps:

In the TEBD algorithm, designed for time evolution with nearest-neighbor interactions, the update step consists of an application of a (real or imaginary) time-evolution exponential
on two adjacent lattice sites. In standard TEBD, the exponential
takes the form of a single four-index tensor or “block” (B) \(u_{NN}\).
It can also be given by a sum of Kronecker products (P) of
single-site operators \(\sum_{k=1}^{K} u_{rk} \otimes u_{ik}\), which can be more time
and memory efficient for \(K < d\). Both strategies pose different
compression problems [Figs. 1(b) and 1(c)]: The block
update contracts directly into a square matrix of dimension
\(\chi d\), while in the product update we obtain a \((\chi K)\)-by-(\(\chi d\))
matrix instead. In both cases, the resulting matrix \(A\) must
be compressed into a rank-\(\chi\) factorization with compression ratios \(d\) and min(\(K,d\)), respectively. Consequently, we expect
RSVD to significantly speed up TEBD simulations on lattices
with larger local dimensions \(d\): In terms of computational
complexity, TEBD with typical bond dimension \(\chi \geq d\) is
dominated by the TSVD compression step of \(O(\chi^3d^3)\) in block
(B) and \(O(\chi^3K^2d)\) in product (P) updates for \(K \leq d\). The
asymptotic RSVD speedup can reduce this scaling to \(O(\chi^3d^2)\)
(for B), as demonstrated in Ref. [45], and to \(O(\chi^3Kd)\) (for
P), respectively—which are typical costs exhibited by other
operations within TEBD as well.

In the TTN setting, instead, the update step directly replaces
two adjacent tensors with a matrix \(A\) associated to the lowest
eigenvector of an effective Hamiltonian. The matrix \(A\) is at
most a \(\chi^2\)-by-\(\chi^2\) square matrix. On some lower levels of the
tree geometry, smaller dimensions can be encountered, with \(d^2\)
at the physical indices on the bottom [Fig. 1(e)]. The majority
of run time, however, is spent on the large update matrices, and
these require a compression by a ratio \(\chi\). A massive speedup
of the compression step, in the order of the bond dimension,
can thus be expected from employing RSVD instead of TSVD.

A feature of all simulations is that we explicitly target the
symmetry-invariant ground states under certain global Abelian
symmetries of the Hamiltonian. These grant us an inner block
structure in all tensors, which enhances efficiency and precision
of the simulation [23,24]. In the compression problem, we therefore
encounter strictly block-diagonal matrices \(A\), encoded in \(N\) nontrivial blocks. The dimensions of these blocks
correspond to degeneracies of symmetry sectors, and add up
to the respective full dimensions of \(A\). In all benchmarked
situations, \(N\) equals the (small) number of global symmetry
sectors, and the optimal TN ground-state approximations display
more or less evenly sized block dimensions. Since matrix
factorizations can be done blockwise, all the actual matrix
dimensions passed to the truncated SVD algorithm are thus
roughly those of \(A\) divided by \(N\). But as the truncation
rank \(\chi_t \approx \chi/N\) per block is similarly reduced, no change in
the compression ratio and hence in the asymptotic speedup
occurs.

Note that the truncation rank per block is usually not known
\textit{a priori}, as it depends on the number of large singular
values \(\sigma_j > \sigma_f\) therein. This information is only directly
available with TSVD, where all singular values of all blocks are
computed. RSVD, on the other hand, delivers just the requested
number of singular values for each block, and some estimate of
the appropriate truncation \(\chi_t' \approx \chi_t\) must be made beforehand.
After RSVDs are then performed in all sectors, we postselect
the \(\chi\) largest singular values and obtain the new optimal
block dimensions \(\chi_t\). In our TEBD simulations, we choose
a blockwise truncation rank \(\chi_t' = \chi/N + c\) with a small constant
c that allows for some variation in sector sizes (typically less
than 5%). For TTN, we instead make the simplest maximal
choice \(\chi_t' = \chi\), which reduces the achievable speedup by a
(small) factor \(N\) but does not require any estimates.

### B. Models

We simulated the quantum Ising model with ferromagnetic
interaction in a tunable transverse field \(h\), on two different
lattices: First, a 1D spin-\(S\) chain of length \(L\) with Hamiltonian
\[
H_{\text{chain}} = -\frac{1}{S^2} \sum_{j} X_j X_{j+1} + \frac{h}{S} \sum_{j} Z_j ,
\]
where \(X\) and \(Z\) are local spin operators (we set \(h = 1\)) and
subscripts denote application sites. In general, in a computa-
tional spin-\(Z\) eigenbasis (\(|m\rangle\)) of local dimension \(d = 2S + 1\)
with integer or half-integer magnetic quantum numbers \(m \in\{-S,-S+1,\ldots,S\}\), we have
\[
\langle m'|Z|m\rangle = m \delta_{m,m'}, \quad \text{(4a)}
\]
\[
\langle m'|X|m\rangle = \sqrt{(S + 1)(m + m' - 1) - mm'} \times (\delta_{m,m+1} + \delta_{m+1,m})/2 . \quad \text{(4b)}
\]

For \(S = 1/2\), \(X\) and \(Z\) reduce to standard Pauli matrices
and the model is exactly solvable with quantum critical point
at \(|h| = h_c = 1\). As \(S \to \infty\), the transition point shifts with
\(h_c \to 2\) [62]. The TEBD is performed for \(S > 1/2\) in open
boundary conditions with values of \(h\) in various distances to
the critical points, which we estimated from finite-size scaling
techniques [63]. In our TTN benchmark, we focus exclusively
on \(S = 1/2\) in periodic boundary conditions.

The second benchmark is the simulation of a spin-1/2
two-dimensional (2D) square-lattice Ising model in cylindrical
boundary conditions of length \(L\) and circumference (or width)
\(W\). With respective site subscripts \(i\) and \(j\), the Hamiltonian
reads
\[
H_{\text{cyl}} = -\sum_{i,j} X_{i,j} X_{i,j+1} - \sum_{i,j} X_{i,j} X_{i+1,j} + h \sum_{i,j} Z_{i,j} .
\]

By summation over \(i\), we map this Hamiltonian onto an open
chain of length \(L\) with local dimension \(d = 2W\). For reasonably
small values \(W\), the ground state can be approximated in a MPS
and its critical behavior can be studied with DMRG [64]. We
performed imaginary TEBD at various values of \(h\), including
points in proximity of the critical field at around \(h_c \approx 3.044\),
as reported with high precision in Monte Carlo and TN studies
on the square lattice [65,66].

As is well known, in the thermodynamic limit, the one-
and two-dimensional Ising models of Eqs. (3) and (5) exhibit
spontaneous ferromagnetic order for \(|h| < h_c\), which breaks
down in the paramagnetic phase for \(|h| > h_c\). Both phases are
gapped; however, at \(|h| = h_c\), the systems become critical and
gapless.

In the case of 1D lattices, we know that the ground states
of a short-ranged, gapped system obey area laws for the
entanglement entropy, while this is not true for a critical,
gapless system [7–10]. Since squares of the singular values
in loop-free TN compression steps correspond to reduced density
eigenvalues of lattice bipartitions, singular values are directly
linked to bipartite entanglement measures such as the von
Neumann entropy, and thus the error analysis Eq. (1) of RSV
d is linked to the physical properties of the ground state. For
this reason, we perform our benchmarks at various values of
h, including values in close proximity to h_c. We expect the
latter to pose the most demanding situation for RSV
due to a potentially slow decay of tail singular values [67], which make
greater amounts of computational resources necessary (via
parameters q, c) to avoid larger errors in Eq. (1). As a comment,
we remark that the benchmarked MPS and TTN simulations
are best suited for noncritical systems due to finite bond dimen-
sions χ that limit correlations and entanglement. However, the
selected finite lattice sizes admit simulations at and around
h = h_c, as is typical in extrapolating critical properties via
finite-size scaling techniques [63,68]. Furthermore, TTN have
capabilities beyond MPS in encoding quantum critical ground
states [16].

Both Ising models in Eqs. (3) and (5) exhibit a global
parity symmetry because their Hamiltonians commute with
\( \otimes_{j=1}^{L} P_j \), being defined locally by \( \langle m | P | m \rangle = (-1)^{m+S} \delta_{m,m'} \). Local basis states transform as \( P | m^{\pm} \rangle = \pm | m^{\pm} \rangle \) and fall either in the even "+" or odd sector "−" of dimensions \( d_+ \), \( d_- \) \( \approx d/2 \) respectively. Rotations in the cylindrical boundary
conditions provide an additional Abelian \( Z_W \) cyclic symmetry
for (5). As a consequence, even and odd sectors further
decompose into \( W \) different angular momentum sectors. As
mentioned in Sec. III A, we encode these symmetries explicitly,
which allows us to restrict the TN state representation to
the ground-state global invariant sector \( s = 0 \), that is, the
even-parity and rotationally invariant subspace.

C. Implementation

Here we report the detailed implementation of a fair run
time and precision comparison between TSVD and RSV
d, and discuss technical details of the benchmarks.

We performed complete runs of our TEBD and TTN
benchmark algorithms by iterating double-tensor updates until
the energy expectation value of the TN state stagnates within
some threshold \( \delta E \). Each run was repeated for different field
h, maximal bond dimension \( \chi \), lattice width \( L \), and a selected spin \( S \) or width \( W \), either with TSVD or RSV
d in the compression steps.

For the precision comparison, we extracted expectations of
energy and magnetization order, correlation and entanglement
properties, and singular values from the produced final states.
The magnetization order \( M \) was measured from nonlocal
correlations,

\[
M = \sqrt{\sum_{k \neq k'} \langle X_k X_{k'} \rangle / \mathcal{N}},
\]

where \( k \) goes over all lattice sites and \( \mathcal{N} \) counts the number of
expectations summed over. The estimate for the correlation
length \( \xi \) was computed from expectations values of \( X_{(k)} \equiv X_{k} \)
in the chain and \( X_{(d)} \equiv X_{(d)} \) in the cylinder as follows:

\[
\xi = \sqrt{\frac{\sum_{r=0}^{R} (r-1)^2 C_r}{2 \sum_{r=0}^{\infty} C_r}}.
\]

Here, \( C_r \) denotes the bulk average over MPS sites \( j \) of
\( \langle X_{(i,j)} X_{(i,j+r)} \rangle \). The additional site index \( i \) appears only in the
two-dimensional model and is averaged over as well to extract
only the horizontal correlation length subject to compression
through the MPS bond links. Note that \( \xi \) tends to underestimate
the actual correlation length and saturates below \( L/\sqrt{\delta} \) if it be-
comes large compared to the system size. Furthermore, profiles of
the von Neumann entropy \( S_N(f) = -\sum_k \lambda_k^2 \log(\lambda_k^2) \) have been obtained from the compressed singular values at MPS
bonds \( j = 1, \ldots, L - 1 \).

We also profiled the individual run times spent in the
truncated SVDs of compression steps, \( T_T \) and \( T_R \), and the
time spent in all remaining parts of the algorithm, \( \bar{T}_T \) and \( \bar{T}_R \),
for TSVD and RSVD runs, respectively. All these run times
have been divided by the number of iterations performed in the
simulation. However, we have found no substantial differences
in the number of update steps performed with TSVD and
RSVD, as reported in Sec. IV. We therefore obtain the average
speedup in compression due to RSVD from

\[
\tau := f \times \frac{T_T}{T_R},
\]

where \( f := \frac{\bar{T}_R}{\bar{T}_T} \) is the ratio of run times spent outside
compression. Since our benchmarks have been performed on
shared cluster nodes, we introduced the factor \( f \) to equalize
the effect of the computational environment on the bare
compression times. Thus, simulation runs that were slowed
down by other computations on a cluster node can be fairly
compared to faster executed simulation runs.

The complete simulation protocol for TEBD was as fol-
 lows: Starting from a product state with randomized tensors
of bond dimension one, the algorithm is run in imaginary
time with some sufficiently large initial time step \( dt \) in the
local imaginary time evolution exponential. After a few first
iterations out of typically many hundred, the bond dimension
saturates the allowed maximum, and we can safely assume \( \chi \)
to be the typical compression rank. The simulation stops when
convergence of the energy is detected as follows: Throughout
the simulation, the change of the expectation value of the
energy is monitored in regular intervals. Whenever this change
drops below the targeted precision threshold \( \delta E \), the simulation
time step \( dt \) is subsequently reduced by a constant factor.
Convergence is declared when the total energy decrement
between two time-step reductions falls below \( \delta E \), too. With
smaller choices of \( \delta E \), better approximations of the final MPS
to the actual ground state of the system can be expected within
the bond dimension \( \chi \), but at the cost of increased number of
iterations and run time.

The TTN ground-state search employs randomized initial
states remaining at maximal bond dimension throughout the
entire simulation. The same initial states were used in com-
parative TSVD and RSVD runs. The algorithm then performs
sequences of double-tensor updates on adjacent tensors, un-
til the difference in energy expectation between subsequent
sweeps falls below machine precision.

All simulations were carried out in double precision
arithmetic with complex numbers, except for the imaginary
TEBD on the spin-1/2 chain which we benchmarked in
a TN representation with real elements, a common choice
to enhance efficiency under time-reversal invariance. Linear
algebra computations (BLAS, LAPACK) where performed with
the INTEL MATH KERNEL LIBRARY (MKL) in versions 11.x. Our fully truncated TSVD implementation is based on the LAPACK D/ZGESDD divide-and-conquer algorithm. For RSVD, we employed the fixed-rank implementation from the RRSVD package [45] with parameters \( q = 4, \ell = 2 \chi \) (see Sec. II B) for any targeted truncation rank \( \chi \). This implementation employs LAPACK D/ZGESVDD for the final factorization in step 4 of the RSVD algorithm. All simulations were executed with single-threaded compression step on 16-way Intel Xeon E5–2670 (2.6-GHZ) compute nodes.

IV. RESULTS

We first report the speedups obtained from upgrading compression steps from TSVD to RSVD. We then present evidence that no loss of precision occurs due to RSVD. Finally, we present selected ground-state properties and spectra of singular values that we encountered in our benchmarks.

All the following speedups have been obtained from independent simulations according to Eq. (8) with an estimated uncertainty of at most \( \Delta \tau \approx 10 \% \). Equal numbers of RSVD and TSVD compression steps were performed in all TTN simulations. Some imaginary TEBD runs converged in less iterations with either RSVD or TSVD, but those fluctuations were negligible compared to \( \Delta \tau \).

Speedups up to \( \tau \approx 24 \) have been reached in TEBD simulations of increasing local dimensions, as shown in Fig. 2 for the one- and two-dimensional Ising models of Eqs. (3) and (5). We observe that RSVD outperforms TSVD for \( d > 10 \), with speedups directly proportional to \( d \) as predicted by the asymptotic cost analysis in Sec. III A. These speedups remain stable under different algorithm parameters, such as changes in convergence criteria [orange crosses in Fig. 2(b)]. We also found no significant dependency on the transverse field \( h \). Thus, all speedups are geometric means over five (2D) and ten (1D) different values of \( h \) in various distances from (including close proximity to) the critical point, and each speedup falls within the error bars.

In all cases, however, the speedup tends to increase with the bond dimension, as shown in Fig. 3 for selected one- and two-dimensional TEBD simulations. The latter suggests some saturation at high bond dimension. Again, all speedups shown are geometric means over at least ten simulations at transverse fields \( h \) in various distances from \( h_c \), which had no significant impact on the speedup, as can be seen from the error bars that always enclose minimal and maximal speedup.

Complementary to our TEBD results, the TTN benchmarks demonstrate massive RSVD speedups already for spin 1/2, when bond dimensions are scaled up: For instance, at \( \chi = 60 \) we found \( \tau \approx 6 \), while \( \chi = 100 \) already provided us with \( \tau \approx 11 \), both on a lattice of length \( L = 64 \).

Next, we assess the accuracy of the final states delivered by our TSVD and RSVD benchmarks. To this end, we compare the simulation errors in energy expectation value \( E \) and nonlocal magnetization order parameter \( M \) of Eq. (6) for various simulation parameters such as \( h, \chi \), and precision target

FIG. 2. Speedup in compression step due to RSVD in TEBD simulations of increasing local dimensions. (a) 2D Ising model \( (L = 30) \) as a function of the width \( W \). (b) 1D Ising model \( (L = 100) \) as a function of local spin \( S \). Dark and light blue points represent data at bond dimensions \( \chi = 100 \) and \( \chi = 150 \), respectively, both from block update and fixed convergence criteria \( (1D: \delta E = 10^{-13}, 2D: \delta E = 10^{-8} \text{ except } W = 8 \text{ was stopped before convergence}) \). Error bars indicate a 10% error estimate in speedups. Orange crosses show linear fits for \( d > 10 \) with slopes \( \approx 0.10,0.13 \) (2D) and 0.18,0.21 (1D) for \( \chi = 100,150 \) respectively.

FIG. 3. Dependency of RSVD speedup on the bond dimension in TEBD simulations. (a) 2D Ising model, speedup at width \( W = 6, L = 30, \text{ and } \delta E = 10^{-8} \). Blue and gray: block (B) and product (P) updates, respectively. (b) 1D Ising model, speedup at spin \( S = 5, L = 100 \text{ and } \delta E = 10^{-13} \). Both panels share the same \( \tau \) axis.
The errors are computed from differences $\Delta E = (E - E_{\text{best}})/E_{\text{best}}$ and $\Delta M = |M - M_{\text{best}}|/M_{\text{best}}$ to high-precision data $E_{\text{best}}$ and $M_{\text{best}}$, respectively. In TEBD simulations, $E_{\text{best}}$ and $M_{\text{best}}$ have been extrapolated from bond dimensions and precisions up to $\chi = 150$, $\delta E = 10^{-14}$ using TSVD, with uncertainty smaller than all observed differences $\Delta E$ and $\Delta M$ (typically one or more orders of magnitude). We found that both TSVD and RSVF produce comparable simulation errors in all benchmarks, as exemplified in Fig. 4 for TEBD simulations of the one-dimensional Ising model for $L = 100$ and $S = 5$. We found similar results for up to $L = 400$ in various precision targets and bond dimensions $\chi \leq 100$ in both par- and ferromagnetic phases as well as close to the critical point. In two-dimensional TEBD simulations at $W = 6$, $L = 30$ and in the TTN benchmarks, TSVD and RSVF results even matched within computational precision.

The range of physical properties covered by our benchmarks is demonstrated in Fig. 5, where the upper panels [Figs. 5(a) and 5(b)] show the magnetization $M$ and the estimate for the correlation length $\xi/L$ [see Eq. (7)] in the final TEBD simulation states. These results, taken from TSVD runs of 1D and 2D Ising models for some of the benchmarked transverse fields $h$, display values of magnetic order and correlation lengths spanning the entire spectrum of possible outcomes. Furthermore, the von Neumann entropies $S_N(j)$ on the MPS bonds confirm area laws in both ordered and unordered phases as well as typical corrections near the 1D critical point, which are well described by a fit to $S_N(j) = a + c/6\log(L/\pi \sin(\pi j/L))$ with some constants $a, c$ [69]. The corresponding singular values are detailed in the bottom panels [Figs. 5(c) and 5(d)]. Within the bond dimensions $\chi_s$ of individual symmetry sectors, they are well fitted by power-law decays $\sigma_k \approx (C_1k + C_2)^{-\gamma}$ with fit constants $C_1$, $C_2$, and decay exponents $\gamma$ ranging from $-2$ to $-11$.

This decay of singular values, which relates physical properties to RSVF performance (as discussed in Sec. III) is further analyzed in Fig. 6 where we present complete spectra of singular values from the local compression problems $A_i$, including the truncated tail of small singular values, for a central bond and critical transverse field. In both TEBD [Fig. 6(a)] and TTN [Fig. 6(b)] simulations, the spectrum of singular values $\lambda_k$ can be separated into two parts: For $k \leq \chi_s$, the spectrum appears to undergo only minor changes throughout the algorithm run time and is well described by the actual decay in the final (ground) state (see Fig. 5 for TEBD) over the majority of the run time. For $k > \chi_s$, on the other hand, we observe a tail spectrum that does not necessarily follow the characteristics expected from the actual ground state (i.e., $\chi \to \infty$). Namely, it changes significantly over the run time and exhibits the fastest decay in the final iteration(s) of the algorithm: In the case of TEBD, the tail can be seen to

\[\delta E = 10^{-11}\] (light shaded) and $10^{-13}$. All errors have been obtained from extrapolated ground state values for $S = 5, L = 100$ in the ferro- ($h = 1.0$) and paramagnetic ($h = 2.0$) phases as well as close to the critical point.

\[\log_{10}(\Delta M)\text{ Transverse field } h\]

\[\log_{10}(\Delta E)\text{ Transverse field } h\]
be bounded by a rapid polynomial decay, well separated from the retained singular values as it finally becomes proportional to a very small evolution time step $dt$. In TTN, compression starts from a rather flat tail spectrum that quickly approaches an exponential decay. This demonstrates that the compression problem within the TN approximation becomes increasingly well conditioned for RSVD, even close to the phase transition, as the algorithm converges closer to the ground state. This allows RSVD to deliver higher precision [cf. Eq. (1), due to oversampling] with higher reliability right in the final stages of the algorithms when most needed.

V. DISCUSSION AND OUTLOOK

We provided evidence for substantially accelerated compression of tensor networks in all benchmarked algorithms by simply replacing the full truncated TSVD with the RSVD algorithm. In particular, RSVD outperformed TSVD with the expected asymptotic speedup, that is proportional to the compression ratio, when not more than 10% of singular values were retained.

Remarkably enough, we attained those speedups without loss of precision in the simulated ground states: With RSVD we reproduced local expectation values such as the energy, as well as long-range correlation and entanglement properties, with differences to TSVD simulations far smaller than the inherent ansatz errors due to a finite bond dimension or number of iterations performed. By benchmarking with encoded Abelian symmetries, we confirmed the RSVD speedup in reduced bond and local dimensions per sector. Even though small matrix sizes can reduce speedups, RSVD becomes increasingly useful with the typically large bond dimensions that are required for ground-state approximation. All results, moreover, hold up independently from the various physical scenarios, i.e., off and at quantum critical points over a wide range of correlation lengths and respective spectra of the singular values. The iterative nature shared by many TN algorithms has been observed to work in favor of RSVD in that the truncated tail singular values decayed quickly in the relevant final iterations, even close to phase transitions.

We expect the presented results to be robust and reproducible in a wide range of tensor network applications. For instance, our choice of RSVD parameters $(q, \ell)$ has been extremely conservative, as confirmed by the small differences to TSVD in the outcomes, and can be fine-tuned for much higher efficiency: Namely, by reducing $q$, RSVD might outperform TSVD for compression ratios as moderate as five or less. With RSVD, precision and efficiency of the compression can further be balanced dynamically, which promises significant reduction of run time in the earlier algorithm stages, as is already standard practice, for instance, in the eigensolver optimization steps in DMRG. In this regard, it may prove specifically useful that RSVD can also deliver a fixed error (instead of fixed rank) approximation: Parameters such as $\chi$, $\ell$, and possibly $q$ are then dynamically adjusted to deliver a compression within a given error bound [44,57]. Such dynamics might also provide an alternative route to fix the compressed sector sizes $\chi_i$ in the presence of symmetric TN, even though good estimates (for instance based on previous iterations) plus added oversampling work well as demonstrated. Moreover, ongoing development of the RSVD method itself may lead to further optimizations, such as modified power iteration schemes for faster convergence [70] or single-view algorithms [71].

With the benchmarked ground-state simulations, it is clear that RSVD is indeed not limited to open system real-time dynamics with TEBD [45], and we foresee a broad impact on DMRG and imaginary or real time evolution codes that operate on ground states, including short-time quenches [72] out of equilibrium via TEBD or the time-dependent variational principle [73]. This in turn could open new possibilities, for instance, in the numerical study of the Kibble-Zurek mechanism [74,75]. More generally, RSVD has great potential in all TN algorithms that make extensive use of truncated SVD with high compression ratios. This includes the various double-tensor update strategies that are regularly employed in DMRG and time evolution codes when Abelian or non-Abelian symmetries are encoded, and to avoid metastabilities that hinder convergence [76,77]. Another particularly promising scenario is the efficient approximation of environments in higher dimensional lattices, especially with PEPS in infinite 2D lattices. Prominent examples are iPEPS [78] by means of infinite TEBD [79,80], coarse-graining renormalization approaches [34,81,82], and corner transfer matrix methods based on CTMRG [33,83] with potentially large compression ratios when bond dimensions are pushed up. Conceivable use cases for RSVD include, for example, lattice models with large local dimensions, lattices in higher dimensions, infinite lattices, and applications of TNs in quantum chemistry.

Note added. S. Morita et al. [84] has recently reported a complexity reduction of the TRG algorithm achieved by means of RSVD.

ACKNOWLEDGMENTS

We thank M. Gerster for discussions and sharing his TTN code. The authors gratefully acknowledge support from the Carl-Zeiss-Stiftung via Nachwuchsförderprogramm, the
state of Baden-Württemberg through bwHPC, the Italian HPC facility CINECA through the TEDDI project, and the German Research Foundation (DFG) through the TWITTER grants. S.M. gratefully acknowledges the support of the DFG via a Heisenberg fellowship. This work was supported by the ERC Synergy grant BioQ.

[1] U. Schollwöck, Ann. Phys. 326, 96 (2011).
[2] R. Orús, Ann. Phys. 349, 117 (2014).
[3] S. R. White, Phys. Rev. Lett. 69, 2863 (1992).
[4] S. R. White, Phys. Rev. B 48, 10345 (1993).
[5] S. Rommer and S. Östlund, Phys. Rev. B 55, 2164 (1997).
[6] J. Eisert, Model. Simul. 3, 520 (2013).
[7] K. Audenaert, J. Eisert, M. B. Plenio, and R. F. Werner, Phys. Rev. A 66, 042327 (2002).
[8] M. B. Plenio, J. Eisert, J. Dreißig, and M. Cramer, Phys. Rev. Lett. 94, 060503 (2005).
[9] J. Eisert, M. Cramer, and M. B. Plenio, Rev. Mod. Phys. 82, 277 (2010).
[10] M. M. Wolf, Phys. Rev. Lett. 96, 010404 (2006).
[11] F. Verstraete, M. M. Wolf, D. Perez-Garcia, and J. I. Cirac, Phys. Rev. Lett. 96, 226001 (2006).
[12] L. Tagliacozzo, G. Evenbly, and G. Vidal, Phys. Rev. B 80, 235127 (2009).
[13] M. Gerster, M. Rizzi, P. Silvi, M. Dalmonte, and S. Montangero, Phys. Rev. B 96, 195123 (2017).
[14] G. Vidal, Phys. Rev. Lett. 99, 220405 (2007).
[15] G. Vidal, Phys. Rev. Lett. 101, 110501 (2008).
[16] P. Silvi, V. Giovannetti, S. Montangero, M. Rizzi, J. I. Cirac, and R. Fazio, Phys. Rev. A 81, 062335 (2010).
[17] R. Orús, T.-C. Wei, O. Buerschaper, and A. García-Saez, Phys. Rev. Lett. 113, 257202 (2014).
[18] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac, Phys. Rev. Lett. 93, 207204 (2004).
[19] A. H. Werner, D. Jascke, P. Silvi, M. Kliesch, T. Calarco, J. Eisert, and S. Montangero, Phys. Rev. Lett. 116, 237201 (2016).
[20] P. Silvi, E. Rico, T. Calarco, and S. Montangero, New J. Phys. 16, 103015 (2014).
[21] L. Tagliacozzo, A. Celi, and M. Lewenstein, Phys. Rev. X 4, 041024 (2014).
[22] T. Pichler, M. Dalmonte, E. Rico, P. Zoller, and S. Montangero, Phys. Rev. X 6, 011023 (2016).
[23] S. Singh, R. N. C. Pfeifer, and G. Vidal, Phys. Rev. A 82, 050301 (2010).
[24] S. Singh, R. N. C. Pfeifer, and G. Vidal, Phys. Rev. B 83, 115125 (2011).
[25] S. Singh and G. Vidal, Phys. Rev. B 86, 195114 (2012).
[26] A. Weihselbaum, Ann. Phys. 327, 2972 (2012).
[27] G. Vidal, Phys. Rev. Lett. 91, 147902 (2003).
[28] G. Vidal, Phys. Rev. Lett. 93, 040502 (2004).
[29] M. Levin and C. P. Nave, Phys. Rev. Lett. 99, 120601 (2007).
[30] T. Nishino and K. Okunishi, J. Phys. Soc. Jpn. 65, 891 (1996).
[31] F. Verstraete and J. I. Cirac, arXiv:cond-mat/0407066.
[32] V. Murg, F. Verstraete, and J. I. Cirac, Phys. Rev. A 75, 033605 (2007).
[33] R. Orús, Phys. Rev. B 85, 205117 (2012).
[34] H.-H. Zhao, Z.-Y. Xie, T. Xiang, and M. Imada, Phys. Rev. B 93, 125115 (2016).
[35] G. H. Golub and C. F. Van Loan, Matrix Computations (Johns Hopkins University Press, Baltimore, MA, 2012), Vol. 3.
[36] J. W. Demmel, Applied Numerical Linear Algebra (Society for Industrial and Applied Mathematics, Philadelphia, 1997).
[37] N. B. Erichson, K. Manohar, S. L. Brunton, and J. N. Kutz, arXiv:1703.09074.
[38] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. (Springer, New York, 2009).
[39] A. Cichocki, D. P. Mandic, A. H. Phan, C. F. Caiafa, G. Zhou, Q. Zhao, and L. D. Lathauwer, IEEE Signal Process. Mag. 32, 145 (2015).
[40] V. Vijayakumari, World J. Comp. Appl. Technol., 1, 41 (2013).
[41] A. G. Pai, S. L. Fernandes, K. Nayak, Nagesha, K. K. Accamma, K. Sushmitha, and K. Kumari, in Second International Conference on Electronics and Communication Systems (ICECS) (IEEE, New York, 2015), pp. 577–582.
[42] M. E. Wall, A. Rechtsteiner, and L. M. Rocha, in A Practical Approach to Microarray Data Analysis, edited by D. P. Berrar, W. Dubitzky, and M. Grañenz (Springer, Berlin, 2003), pp. 91–109.
[43] A. S. Jermyn, arXiv:1709.03080.
[44] N. Haliko, P.-G. Martinsson, and J. A. Tropp, SIAM Rev. 53, 217 (2011).
[45] D. Tamascelli, R. Rosenbach, and M. B. Plenio, Phys. Rev. E 91, 063306 (2015).
[46] E. Anderson, Z. Bai, C. Bischof, L. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammond, A. McKenney, and D. Sorensen, LAPACK Users’ Guide, 3rd ed. (Society for Industrial and Applied Mathematics, Philadelphia, 1999).
[47] Y.-Y. Shi, L.-M. Duan, and G. Vidal, Phys. Rev. A 74, 022320 (2006).
[48] W. Hackbusch and S. Kühn, J. Fourier Anal. Appl. 15, 706 (2009).
[49] V. Murg, F. Verstraete, O. Legeza, and R. M. Noack, Phys. Rev. B 82, 205105 (2010).
[50] N. Nakatani and G. K.-L. Chan, J. Chem. Phys. 138, 134113 (2013).
[51] M. Gerster, P. Silvi, M. Rizzi, R. Fazio, T. Calarco, and S. Montangero, Phys. Rev. B 90, 125154 (2014).
[52] M. Gerster, M. Rizzi, F. Tschiirsich, P. Silvi, R. Fazio, and S. Montangero, New J. Phys. 18, 015015 (2016).
[53] C. Eckart and G. Young, Psychometrika 1, 211 (1936).
[54] L. Mirsky, Quart. J. Math. 11, 50 (1960).
[55] P.-G. Martinsson, V. Rokhlin, and M. Tygert, Appl. Comput. Harmonic Anal. 30, 47 (2011).
[56] V. Rokhlin, A. Szlam, and M. Tygert, SIAM J. Matrix Anal. Appl. 31, 1100 (2010).
[57] M. Gu, SIAM J. Sci. Comput. 37, A1139 (2015).
[58] A. Szlam, Y. Kluger, and M. Tygert, arXiv:1412.3510.
[59] N. B. Erichson, S. Voronin, S. L. Brunton, and J. N. Kutz, arXiv:1608.02148.
[60] S. Voronin and P.-G. Martinsson, arXiv:1502.05366.
[61] P. Silvi, F. Tschirsich, M. Gerster, J. Jünemann, D. Jaschke, M. Rizzi, and S. Montangero, arXiv:1710.03733.
[62] K. A. Penson and M. Kolb, Phys. Rev. B 30, 1470 (1984).
[63] M. E. Fisher and M. N. Barber, Phys. Rev. Lett. 28, 1516 (1972).
[64] M. S. L. du Croo de Jongh and J. M. J. van Leeuwen, Phys. Rev. B 57, 8494 (1998).
[65] H. W. J. Blöte and Y. Deng, Phys. Rev. E 66, 066110 (2002).
[66] M. Rizzi, S. Montangero, P. Silvi, V. Giovannetti, and R. Fazio, New J. Phys. 12, 075018 (2010).
[67] P. Calabrese and A. Lefèvre, Phys. Rev. A 78, 032329 (2008).
[68] J. Cardy, Finite-Size Scaling (Elsevier, Amsterdam, 2012), Vol. 2.
[69] P. Calabrese and J. Cardy, J. Phys. A 42, 504005 (2009).
[70] C. Musco and C. Musco, in Advances in Neural Information Processing Systems 28, edited by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett (Curran Associates, Inc., New York, 2016), pp. 1396–1404.
[71] J. A. Tropp, A. Yurtsever, M. Udell, and V. Cevher, arXiv:1609.00048.
[72] J. Eisert and T. J. Osborne, Phys. Rev. Lett. 97, 150404 (2006).
[73] J. Haegeman, J. I. Cirac, T. J. Osborne, I. Pižorn, H. Verschelde, and F. Verstraete, Phys. Rev. Lett. 107, 070601 (2011).
[74] T. W. B. Kibble, J. Phys. A 9, 1387 (1976).
[75] W. H. Zurek, Nature (London) 317, 505 (1985).
[76] S. R. White, Phys. Rev. B 72, 180403 (2005).