

Quantum state tomography with tensor train cross approximation

Alexander Lidiak,¹ Casey Jameson,¹ Zhen Qin,² Gongguo Tang,³ Michael B. Wakin,⁴ Zhihui Zhu,² and Zhexuan Gong^{1,*}

¹*Department of Physics, Colorado School of Mines, Golden, Colorado 80401, USA*

²*Department of Electrical & Computer Engineering, University of Denver, Denver, Colorado 80210, USA*

³*Department of Electrical & Computer Engineering, University of Colorado, Boulder, Colorado 80309, USA*

⁴*Department of Electrical Engineering, Colorado School of Mines, Golden, Colorado 80401, USA*

It has been recently shown that a state generated by a one-dimensional noisy quantum computer is well approximated by a matrix product operator with a finite bond dimension independent of the number of qubits. We show that full quantum state tomography can be performed for such a state with a minimal number of measurement settings using a method known as tensor train cross approximation. The method works for reconstructing full rank density matrices and only requires measuring local operators, which are routinely performed in state-of-art experimental quantum platforms. Our method requires exponentially fewer state copies than the best known tomography method for unstructured states and local measurements. The fidelity of our reconstructed state can be further improved via supervised machine learning, without demanding more experimental data. Scalable tomography is achieved if the full state can be reconstructed from local reductions.

With the rapid development of quantum computing and quantum simulation, how to characterize and validate large quantum many-body states generated by experimental quantum devices becomes a major challenge. Among various methods [1–5], quantum state tomography (QST) [6–12] remains the gold standard, as it provides complete information about the experimental state. However, for a generic mixed state in a d -dimensional Hilbert space, QST requires a number of state copies at least proportional to d^2 to guarantee high reconstruction fidelity [10], thus it is generally inefficient for a quantum system of many particles. Fortunately, states generated by physical quantum systems are often structured, requiring much fewer resources in order to gain complete information about them. For example, if the quantum system is well isolated from the environment, then the state of the system is often close to a pure state and thus can be well approximated by a low-rank density matrix. In this case, one simply needs to prepare and measure the state $\sim d$ times [8, 11, 13–18].

To make QST truly scalable, however, the number of state copies needed should only scale polynomially with the number of particles N . This is only possible if the state has a compact representation with only $\text{poly}(N)$ independent parameters. Examples of such states include matrix product states (MPS) [19], matrix product operators (MPO) [20], tensor network states [21], and quantum neural network states [22]. However, having an efficient representation of the quantum state does not imply that an efficient QST method exists. For a pure state represented by an MPS with a finite bond dimension, efficient QST methods have been found and tested for various physical states [7, 23, 24], although a rigorous bound on the number of state copies needed to guarantee high fidelity QST is yet to be found. States that can be efficiently represented by artificial neural networks have also been targeted for QST [22, 25–34], but it is not clear whether efficient QST schemes exist in general for such states.

In this paper, we focus on QST for a mixed state represented by an MPO with a finite bond dimension independent

of the system size. Such an MPO has recently been shown to describe most states generated by a one-dimensional quantum computer with a finite error rate for elementary quantum gates [35]. Given the popularity of one-dimensional quantum computers such as those based on trapped ions [36], efficient QST methods for states generated by such devices are highly desired. However, no guaranteed efficient method to perform QST with bounded error on these states has been developed. We also note that such states can have very high entropy, making QST methods designed for low-rank density matrices generally not applicable. The method currently demonstrating highest efficiency is developed in Ref. [9], which attempts to reconstruct the MPO representing the target state from local reduced density matrices each involving at most R particles. The method requires performing full tomography on the reduced density matrices, which involves a number of state copies exponential in R . If R is small, this method is efficient, but R cannot be determined without knowing the state explicitly or even predicted from the bond dimension, and thus the efficiency of the whole method is not guaranteed.

Here we take an important step towards addressing this problem. Using a method in signal processing and compressed sensing known as tensor train cross approximation [37–40], we show that an N -particle state represented by an MPO with a finite constant bond dimension can be reconstructed by measuring the state in only $O(N)$ different bases. To our best knowledge, the application of cross approximation to QST has not been studied in depth before [41]. Importantly, cross approximation only requires local measurements on individual quantum particles, which are routinely performed in current quantum experiments. For example, on qubit systems, we need only measure one of the three Pauli operators $\sigma^{x,y,z}$ for each qubit. As a result, our method is easily implementable experimentally. For a generic mixed state made of N qubits, one needs to measure all 3^N different combinations of local Pauli operators in order to gain complete information of the state [11]. Therefore, our method requires only a small fraction $O(N)/3^N$ of measurement bases (as well as the number of state copies) compared to such unstructured tomography. We emphasize that our $O(N)$ scaling of the number of measurement bases is optimal since the target MPO state contains

* gong@mines.edu

$O(N)$ independent parameters.

Nevertheless, our method does not guarantee that the number of total state copies is polynomial in N . Due to the statistical errors in quantum measurements, an exponentially large number of copies of states per measurement basis needs to be used to ensure a bounded error (in the Hilbert-Schmidt norm) of the reconstructed state. Without additional assumptions on the target MPO state, we expect this limitation to be fundamental, as it also applies to the best-known unstructured tomography method using local measurements [10]. We develop a supervised machine learning method to alleviate the effects of statistical errors and the requirement of a large number of measurements. Finally, we point out that we can combine our method with the protocol in Ref. [9] to achieve efficient QST if a local reduction exists for the target MPO state. Our method can perform QST on the reduced states much more efficiently than standard methods, thus improving the protocol in Ref. [9] without affecting its scalability condition.

This paper is organized as follows: In Section I, we introduce the method of cross approximation for both matrices and tensors. Section II applies cross approximation to QST for physical target states represented by MPOs. In Section III, we analyze the effects of statistical errors in quantum measurements. We then show that the detrimental effects of such measurement errors can be further reduced via a supervised machine learning method in Section IV. The paper ends with a discussion and outlook section.

I. INTRODUCTION OF CROSS APPROXIMATION

The tensor train cross approximation we use for QST is a generalization of the cross approximation for a matrix, which is also known as *skeleton decomposition* or *CUR decomposition* [42–45]. The main idea of matrix cross approximation is that given a low rank matrix, we can possibly approximate it using a small number of its rows and columns. Formally, following the standard notation for the CUR decomposition, we express the cross approximation for a general $m \times n$ complex valued matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$ as

$$\mathbf{A} \approx \mathbf{C}\mathbf{U}^+\mathbf{R}, \quad (1)$$

where $\mathbf{C} = \mathbf{A}(:, J)$, $\mathbf{R} = \mathbf{A}(I, :)$, and $\mathbf{U} = \mathbf{A}(I, J)$ with I and J respectively denoting some subsets of the indices of \mathbf{A} 's rows and columns. \mathbf{U}^+ denotes the pseudo inverse of the matrix \mathbf{U} (not to be confused with Hermitian conjugate).

An illustration of the matrix cross approximation is shown in Fig. 1. One can mathematically prove that if $\text{rank}(\mathbf{U}) = \text{rank}(\mathbf{A})$, then the cross approximation becomes exact, i.e. $\mathbf{A} = \mathbf{C}\mathbf{U}^{-1}\mathbf{R}$. In this case, one can also perform a singular value decomposition (SVD) of \mathbf{A} to obtain an exact decomposition of \mathbf{A} in a similar form. However, computing the SVD requires full knowledge of the matrix \mathbf{A} , while the cross approximation only requires a small fraction of \mathbf{A} 's rows and columns be known for a low rank matrix. As we will show, this advantage of cross approximation allows us to measure only a small number of observables for a target quantum state represented by a compact MPO.

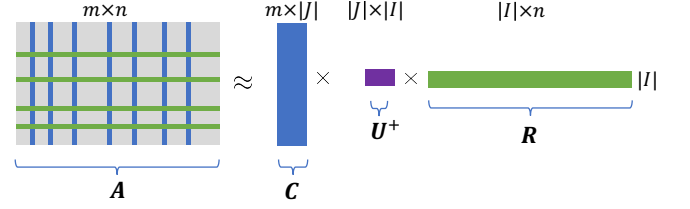


Figure 1. Illustration of matrix cross approximation. An $m \times n$ matrix \mathbf{A} is approximated by a product of three matrices \mathbf{C} , \mathbf{U}^+ , and \mathbf{R} , where \mathbf{C} is formed by selecting $|J|$ columns from \mathbf{A} , \mathbf{R} is formed by selecting $|I|$ rows from \mathbf{A} , and \mathbf{U} is formed by the union of \mathbf{C} and \mathbf{R} .

If the rank of the matrix \mathbf{A} is not known a priori, then cross approximation is in general less accurate and robust than the SVD, which provides the best approximation of \mathbf{A} for a chosen rank. For cross approximation to be optimal, one needs to choose the rows and columns of \mathbf{A} in a way that maximizes the volume (i.e. determinant in modulus) of \mathbf{U} in Eq. (1). The search for the maximum volume submatrix \mathbf{U} is in general computationally expensive, but many efficient methods have been developed to achieve quasi-optimal results [46–51].

We now generalize the matrix cross approximation to tensor train cross approximation. We denote an order- N tensor \mathcal{A} 's elements using $\mathcal{A}(\gamma_1, \gamma_2, \dots, \gamma_N)$, where $\gamma_i = 0, 1, 2, \dots, d_i - 1$ is the index for the i^{th} dimension of the tensor. We focus on the case where \mathcal{A} has a tensor train decomposition of the form:

$$\mathcal{A}(\gamma_1, \dots, \gamma_N) = G_1^{\gamma_1} G_2^{\gamma_2} \dots G_N^{\gamma_N} \quad (2)$$

where each $G_i^{\gamma_i}$ is a matrix of dimension $\chi_{i-1} \times \chi_i$ with $\chi_0 = \chi_N = 1$. We point out that the tensor train in Eq. (2) is identical to an MPS or MPO when the tensor \mathcal{A} represents a wavefunction or density matrix of a quantum many-spin system [19], with $\{\chi_i\}$ often referred to as the *bond dimensions*.

Given an arbitrary tensor \mathcal{A} , one can always find the matrices $\{G_i^{\gamma_i}\}$ in the tensor train decomposition Eq. (2) via successive SVDs [19]. Such successive SVDs further allow one to compress the dimension $\{\chi_i\}$ of the the matrices $\{G_i^{\gamma_i}\}$ [19]. However, the entire tensor \mathcal{A} , which contains exponentially many elements in N , needs to be known in order to perform the SVD, which is impractical for large N .

Similar to the matrix case, we can apply cross approximation instead of SVD for the tensor \mathcal{A} to improve the efficiency of the tensor train decomposition. This consists of the following steps (illustrated in Fig. 2). In the first step, we reshape the tensor \mathcal{A} with dimensions $d_1 \times d_2 \times \dots \times d_N$ into a wide matrix \mathbf{A}_1 of dimension $d_1 \times (d_2 d_3 \dots d_N)$ and perform cross approximation on this matrix by selecting r_1 rows and columns of \mathbf{A}_1 , resulting in $\mathbf{A}_1 \approx \mathbf{C}_1 \mathbf{U}_1^+ \mathbf{R}_1$. One can then show that \mathbf{A}_1 has a rank of at most χ_1 , and thus if $r_1 \geq \chi_1$, the matrix cross approximation of \mathbf{A}_1 can be exact. The matrix $\mathbf{C}_1 \mathbf{U}_1^+$, which has a dimension of $d_1 \times r_1$, creates d_1 row vectors $\{\tilde{G}_1^{\gamma_1}\}$ each of dimension r_1 .

In the second step, we reshape the previously obtained matrix \mathbf{R}_1 into a matrix \mathbf{A}_2 of dimension $(r_1 d_2) \times (d_3 d_4 \dots d_N)$

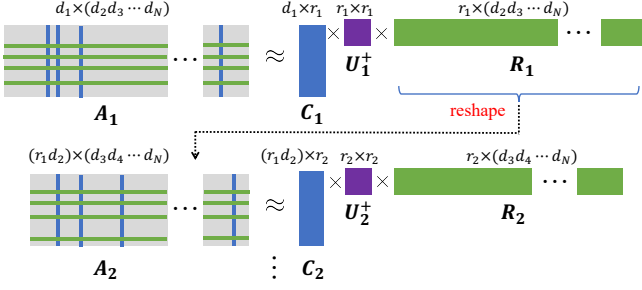


Figure 2. Illustration of tensor train cross approximation. In step one, we perform a matrix cross approximation of \mathbf{A}_1 , which is obtained from reshaping the tensor \mathcal{A} . In step two, we reshape the matrix \mathbf{R}_1 obtained in step 1 into a matrix \mathbf{A}_2 and perform a matrix cross approximation of \mathbf{A}_2 . We iterate this step for a total of $N - 1$ steps.

and perform a cross approximation of $\mathbf{A}_2 \approx \mathbf{C}_2 \mathbf{U}_2^+ \mathbf{R}_2$ by choosing r_2 rows and columns. Again, one can show that the matrix \mathbf{A}_2 has a rank of at most χ_2 , thus the cross approximation of \mathbf{A}_2 can be exact if $r_2 \geq \chi_2$. The matrix $\mathbf{C}_2 \mathbf{U}_2^+$, which has a dimension of $(r_1 d_2) \times r_2$, can be reshaped to create d_2 matrices $\{\tilde{G}_2^{\gamma_2}\}$ each of dimension $r_1 \times r_2$.

We iterate the above step for a total of $N - 1$ steps. This allows us to obtain a set of matrices $\{\tilde{G}_i^{\gamma_i}\}$ for $i = 1, 2, \dots, N$ and $\gamma_i = 1, 2, \dots, d_i - 1$, where in the last step the matrix \mathbf{R}_{N-1} is used to obtain $\{\tilde{G}_N^{\gamma_N}\}$. As we mentioned, if we choose $r_i \geq \chi_i$ in the matrix cross approximation in step i , we can exactly recover the original tensor as $\mathcal{A}(\gamma_1, \dots, \gamma_N) = \tilde{G}_1^{\gamma_1} \tilde{G}_2^{\gamma_2} \dots \tilde{G}_N^{\gamma_N}$. Most importantly, this process is highly efficient, as in each step we only need to find the small matrices \mathbf{C}_i of dimension $(r_{i-1} d_i) \times r_i$ and \mathbf{U}_i of dimension $r_i \times r_i$, and they can be constructed directly from the elements of the tensor \mathcal{A} . Assuming that $d = \max_i d_i$ and $r = \max_i r_i$, then we only need to know approximately Ndr^2 elements of \mathcal{A} to fully reconstruct it. And in the case of $r_i = \chi_i$, this reconstruction method is optimal because \mathcal{A} contains the same number of independent parameters.

In reality, \mathcal{A} may not have an exact tensor train decomposition as in Eq. (2), or we may not know the bond dimension $\{\chi_i\}$ of the exact tensor train decomposition. In this case, we can choose r_i in the tensor train cross approximation heuristically, starting from a small initial guess. If $r_i < \chi_i$, then the cross approximation is no longer exact but we can use maximum volume principle based algorithms, of which two have been developed, to find the optimal entries of the tensor \mathcal{A} used to perform the tensor train decomposition. The first algorithm is introduced in Ref. [38] and known as the “DMRG-cross” algorithm, since it is similar to the density matrix renormalization group (DMRG) method used in variationally finding the ground state of a 1D quantum many-body system [19]. Similar to DMRG, it can choose r_i adaptively between 1 and a preset maximum value based on a chosen local error threshold. The second is introduced in Ref. [39] and is known as the “greedy-cross” algorithm, which uses the method of greedy approximation [52] and chooses an aggressively small r_i . We will mainly use the DMRG-cross algorithm for QST as it performs more reliably in our calculations.

II. CROSS APPROXIMATION BASED QST

We now apply the above-mentioned tensor train cross approximation method to quantum state tomography. For simplicity, we focus on a quantum state of N qubits, but it is straightforward to generalize our method to a quantum system of N qudits. A general mixed state of N qubits can be described by a density operator ρ in the N -qubit Pauli operator basis.

$$\rho = \sum_{\gamma_1, \dots, \gamma_N} \mathcal{A}(\gamma_1, \gamma_2, \dots, \gamma_N) (\sigma_1^{\gamma_1} \sigma_2^{\gamma_2} \dots \sigma_N^{\gamma_N}) \quad (3)$$

where each $\gamma_i = 0, 1, 2, 3$ and $\sigma_i^{0,1,2,3}$ denotes respectively the identity operator and Pauli operator $\sigma_i^{x,y,z}$ acting non-trivially only on qubit i . As a result, the full state is described by the order- N tensor \mathcal{A} we discussed in Section I with $d = d_1 = d_2 \dots = d_N = 4$. Importantly, each element of \mathcal{A} can be measured experimentally since

$$\mathcal{A}(\gamma_1, \dots, \gamma_N) = \langle \sigma_1^{\gamma_1} \sigma_2^{\gamma_2} \dots \sigma_N^{\gamma_N} \rangle / 2^N. \quad (4)$$

For state-of-art quantum computers and quantum simulators, the expectation value of $\sigma_1^{\gamma_1} \sigma_2^{\gamma_2} \dots \sigma_N^{\gamma_N}$ can be measured by locally measuring each qubit (spin-1/2) in the X , Y , or Z direction in an arbitrary order. Note that if $\gamma_i = 0$, one can measure the qubit i in any direction (or not measure it) and simply replace the operator σ_i^0 by unity in evaluating the expectation value.

If \mathcal{A} for our target state can be represented or well approximated by a tensor train (or MPO) defined in Eq. (2) with a maximum bond dimension $\chi \equiv \max_i \chi_i$, then we can perform QST by reconstructing \mathcal{A} (and hence the full state) using tensor-train cross approximation. The main advantage of this QST method is that only about $4N\chi^2$ elements of \mathcal{A} need to be measured, which is also the number of different measurement bases required. As we mentioned in Section I, this is the minimal number of measurement bases required to gain full information of the target state.

Another advantage of our QST protocol is that we obtain an efficient representation of the target state in an MPO form (consisting of the matrices $\{\tilde{G}_i^{\gamma_i}\}$). The full density matrix of the target state is never reconstructed or stored, explicitly. However, the reconstructed MPO can be used to compute the expectation values of most physically interesting observables efficiently [53]. In particular, these observables include N -body correlation functions, such as $\sigma_1^{\gamma_1} \sigma_2^{\gamma_2} \dots \sigma_N^{\gamma_N}$ with no $\gamma_i = 0$. The expectation value of such global observable is hard to obtain using shadow tomography techniques [1, 4] and full QST is usually needed.

We now benchmark the performance of cross approximation based QST using two different classes of physical target states represented by MPOs with a given bond dimension. First, we consider thermal states of a 1D quantum Ising model, described by the Hamiltonian

$$H = \sum_{i=1}^{N-1} \sigma_i^z \sigma_{i+1}^z + g \sum_{i=1}^N \sigma_i^x. \quad (5)$$

We will set $g = 1$, which makes the ground state of H at its quantum critical point in the thermodynamic limit. Such ground state requires the largest bond dimension among all values of g for it to be approximated by an MPS or MPO. The thermal state of H is defined by $\rho_T = \frac{e^{-H/T}}{\text{Tr}(e^{-H/T})}$, and it can be well approximated by an MPO with a small bond dimension for tens or even hundreds of spins [53]. We will set $T = 0.2$ (corresponding to a low temperature state close to the ground state) and $T = 2$ (corresponding to a high temperature state) in our following calculations. We generate ρ_T in an MPO form for up to $N = 40$ qubits using the Open Source Matrix Product State (OSMPS) software package [54, 55], which uses an imaginary time evolution of an MPO ansatz to approximate ρ_T with a maximum bond dimension of 32.

Our second class of target states are random locally purified tensor network (LPTN) states [56]. These states are represented by random MPOs that are guaranteed to be physical. The density operator ρ_{LPTN} of such a random LPTN state is represented by

$$\langle s_1 \cdots s_N | \rho_{\text{LPTN}} | s'_1 \cdots s'_N \rangle = \mathbf{M}_1^{s_1, s'_1} \cdots \mathbf{M}_N^{s_N, s'_N} \quad (6)$$

where $|s_1 \cdots s_N\rangle$ with $s_i = 0, 1$ denotes the computational basis state for N qubits. To make sure that the above density operator ρ is physical, it must be semi-positive definite. This can be guaranteed if each matrix $\mathbf{M}_i^{s_i, s'_i}$ takes the following form [53]:

$$\mathbf{M}_i^{s_i, s'_i} = \sum_{a_i=1}^{K_i} \mathbf{A}_i^{s_i, a_i} \otimes (\mathbf{A}_i^{s'_i, a_i})^* \quad (7)$$

where $*$ denotes complex conjugate and K_i is an arbitrary positive integer. For simplicity, we assume each matrix $\mathbf{A}_i^{s_i, a_i}$ is of dimension $\kappa \times \kappa$, except that $\mathbf{A}_1^{s_1, a_1}$ and $\mathbf{A}_N^{s_N, a_N}$ are of dimension $1 \times \kappa$ and $\kappa \times 1$ respectively. To make the state ρ_{LPTN} sufficiently random, we set each matrix element of $\mathbf{A}_i^{s_i, a_i}$ to be a random complex number with both its real and imaginary parts drawn uniformly from $[-1, 1]$. In addition, we set $K_i = 10$ to make sure the state is sufficiently mixed ($K_i = 1$ will generate a pure state).

It is straightforward to see that ρ_{LPTN} is an MPO with bond dimension $\chi = \kappa^2$ and is always Hermitian. However, to make ρ_{LPTN} physical, it still needs to be normalized. This is done by calculating

$$\text{Tr}(\rho) = (\mathbf{M}_1^{0,0} + \mathbf{M}_1^{1,1}) \cdots (\mathbf{M}_N^{0,0} + \mathbf{M}_N^{1,1}) \quad (8)$$

and dividing each matrix $\mathbf{M}_i^{s_i, s'_i}$ matrix above by $[\text{Tr}(\rho)]^{1/N}$.

Finally, we need to convert the MPO $\{\mathbf{M}_i^{s_i, s'_i}\}$ representing ρ_{LPTN} from the computational basis to the Pauli operator basis to comply with our measurement scheme. This can be done using the following linear transformation that preserves the bond dimension of the MPO:

$$G_i^0 = \frac{M_i^{0,0} + M_i^{1,1}}{2} \quad G_i^1 = \frac{M_i^{0,1} + M_i^{1,0}}{2} \quad (9)$$

$$G_i^2 = i \frac{M_i^{0,1} - M_i^{1,0}}{2} \quad G_i^3 = \frac{M_i^{0,0} - M_i^{1,1}}{2}. \quad (10)$$

To quantify the performance of tensor train cross approximation, we also need to define a distance measure between the reconstructed state and the original target state. Commonly used distance measures for quantum states include trace distance and fidelity [57], but both of these measures cannot be computed efficiently for a large number of qubits even if the states have a compact MPO representation. Here we instead use a normalized Frobenius norm (squared) difference as a distance measure between two density matrices ρ_1 and ρ_2 , defined as [9]

$$D(\rho_1, \rho_2) \equiv \frac{\|\rho_1 - \rho_2\|_2^2}{\|\rho_1\|_2^2} = \frac{\text{Tr}(\rho_1^\dagger \rho_1 + \rho_2^\dagger \rho_2 - \rho_1^\dagger \rho_2 - \rho_2^\dagger \rho_1)}{\text{Tr}(\rho_1^\dagger \rho_1)} \quad (11)$$

where \dagger denotes Hermitian conjugate and $\|A\|_2 = \sqrt{\text{Tr}(A^\dagger A)}$ denotes the Frobenius norm of any matrix A . Importantly, $D(\rho_1, \rho_2)$ can be computed efficiently if both ρ_1 and ρ_2 are represented by compact MPOs. For example, if

$$\rho_1 = \sum_{\gamma_1, \dots, \gamma_N} (G_1^{\gamma_1} G_2^{\gamma_2} \cdots G_N^{\gamma_N}) \sigma_1^{\gamma_1} \sigma_2^{\gamma_2} \cdots \sigma_N^{\gamma_N} \quad (12)$$

$$\rho_2 = \sum_{\gamma_1, \dots, \gamma_N} (\tilde{G}_1^{\gamma_1} \tilde{G}_2^{\gamma_2} \cdots \tilde{G}_N^{\gamma_N}) \sigma_1^{\gamma_1} \sigma_2^{\gamma_2} \cdots \sigma_N^{\gamma_N}, \quad (13)$$

then terms such as $\text{Tr}(\rho_1^\dagger \rho_2)$ in Eq. (11) can be computed efficiently using contractions of the two MPOs:

$$\frac{\text{Tr}(\rho_1^\dagger \rho_2)}{2^N} = \left(\sum_{\gamma_1} (G_1^{\gamma_1})^* \otimes \tilde{G}_1^{\gamma_1} \right) \cdots \left(\sum_{\gamma_N} (G_N^{\gamma_N})^* \otimes \tilde{G}_N^{\gamma_N} \right). \quad (14)$$

The distance measure D can only be calculated for a numerical benchmark experiment. For an actual quantum experiment, we do not have full knowledge of the state (or its MPO representation) and therefore cannot calculate D . However, we can guess if our QST protocol succeeded in well approximating the experimental target using a sampled version of D , denoted by D_s below, to quantify the quality of the QST:

$$D_s(\rho_1, \rho_2) = \frac{\sum'_{\gamma_1 \cdots \gamma_N} |\langle \sigma_1^{\gamma_1} \cdots \sigma_N^{\gamma_N} \rangle_{\rho_1} - \langle \sigma_1^{\gamma_1} \cdots \sigma_N^{\gamma_N} \rangle_{\rho_2}|^2}{\sum'_{\gamma_1 \cdots \gamma_N} |\langle \sigma_1^{\gamma_1} \cdots \sigma_N^{\gamma_N} \rangle_{\rho_1}|^2} \quad (15)$$

where $\sum'_{\gamma_1 \cdots \gamma_N}$ denotes the sum over only the indices $\{\gamma_1, \dots, \gamma_N\}$ used by the cross approximation (and the corresponding local bases in which measurements are performed). Since we only need $O(N)$ measurement bases, D_s can be computed easily from experimental measurement data.

The difference between the target state and the cross approximation reconstructed state comes primarily from two sources: (1) Underestimation of the bond dimension of the target MPO state leads the cross approximation to be inexact. (2) Statistical error in quantum measurements leads the elements of the tensor \mathcal{A} (and therefore the inputs to the cross approximation) to be inexact. For the rest of this section, we will focus on the first error source by assuming zero statistical error in the measurement. This is of course an impractical assumption as it will require an infinite number of state copies

per measurement basis. We will remove this assumption and focus on the effects of the statistic error in Section III.

We have performed numerical experiments for synthetic target states, those being the aforementioned thermal states of Eq. (5) and the random LPTN states in Eq. (6). For both types of target states, we can efficiently calculate the expectation values of $\sigma_1^{\gamma_1} \sigma_2^{\gamma_2} \cdots \sigma_N^{\gamma_N}$ [53] needed for the tensor train cross approximation. We then apply the DMRG-cross algorithm to reconstruct the target state for up to $N = 40$ qubits. Without fine tuning, we set the maximum bond dimension used in DMRG-cross to 10, and the local truncation error to 10^{-3} .

As expected, we find that the cross approximation works very well (with $D < 10^{-6}$) when the DMRG-cross algorithm does not underestimate the bond dimension of the target state. This is the case for the thermal states of Eq. (5) with $T = 2$ (see Fig. 3a), which are well approximated by MPOs with bond dimensions not exceeding 7 for $N \leq 40$, and DMRG-cross adaptively chooses bond dimensions between 6 and 8 for these states. Ideally, we expect D to be zero (or at machine error level) if we do not underestimate the bond dimension, but numerical instabilities associated with the pseudo-inverse could account for the nonzero D observed [40].

For most other target states, DMRG-cross underestimates the bond dimensions. But remarkably, the cross approximation still works well in such scenarios, achieving $D < 10^{-2}$ everywhere in Fig. 3. This means that the tensor train cross approximation can actually be used as a compression technique for MPOs. Unlike most other MPO compression techniques, the cross approximation does not have any restrictions on the locality of MPOs [20], and is particularly useful for compressing quantum states represented by MPOs. Specifically, the thermal states with $T = 0.2$ are represented by MPOs of increasing bond dimension in N , up to 32. And we clearly see that D increases with N due to an increasingly underestimated bond dimension. This is also the case for the random LPTN states, whose bond dimensions are either 16 or 36. Naively, one would expect the errors due to the underestimated bond dimensions would proliferate exponentially in N due to the iterative process in tensor train cross approximation (see Fig. 2). However, we only see a polynomial increase of the distance measure D as N increases. This observation is consistent with recent mathematical results on the error bounds of tensor train cross approximation showing that the reconstruction error in general grows only polynomially in N [39, 40]. For all target states we studied, the sampled distance D_s is generally well below the value of D .

Fig. 4 shows the number of measurement bases N_b required by the cross approximation based QST protocol. We obtain N_b in our numerical experiments by counting the number of times a distinct Pauli expectation value $\langle \sigma_1^{\gamma_1} \sigma_2^{\gamma_2} \cdots \sigma_N^{\gamma_N} \rangle$ is used in the DMRG-cross algorithm. As we expect, N_b indeed scales linearly in N (and quadratically in the bond dimension used by DMRG-cross). In contrast, unstructured tomography based on local measurements requires 3^N measurement bases, which far exceeds N_b for all $N > 8$.

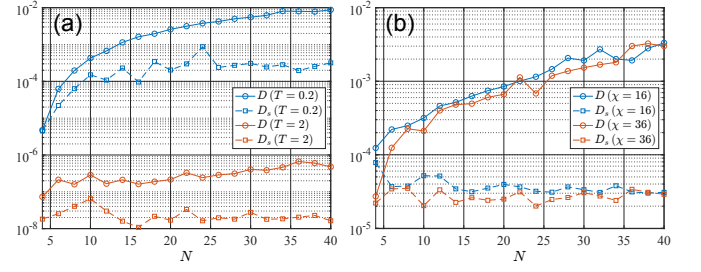


Figure 3. Distance measures (D and D_s) between the target states and the reconstructed states using tensor train cross approximation without measurement errors as a function of the number of qubits N . The target states are thermal states of a quantum Ising model with $T = 0.2$ or 2 (a) and random LPTN states with $\chi = 16$ or 36 (b).

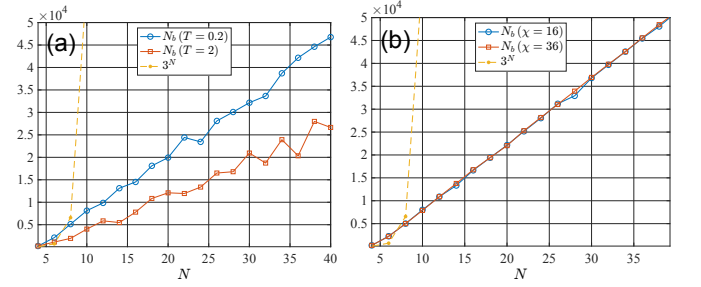


Figure 4. The number of measurement bases N_b required by the tensor train cross approximation as a function of N . We also plot the number of measurement bases required by unstructured tomography that is equal to 3^N . The target states are the same as those in Fig. 3.

III. EFFECTS OF STATISTICAL ERRORS

The benchmark results in Section II assumes that the elements of the tensor \mathcal{A} are measured exactly. In practice, however, we can only measure $\langle \sigma_1^{\gamma_1} \sigma_2^{\gamma_2} \cdots \sigma_N^{\gamma_N} \rangle$ up a standard deviation of $\delta \leq 1/\sqrt{M}$, where M is the number of repeated quantum measurements or identical state copies. Note that throughout this work we assume that the measurements are performed on individual copies of the N -qubit state and do not consider collective measurements [10] that involve measuring multiple state copies simultaneously with possible entangling operations. Collective measurements may require fewer state copies to achieve the same precision in QST [58], but they are very challenging to realize experimentally.

Let us first estimate the precision δ needed in measuring $\langle \sigma_1^{\gamma_1} \sigma_2^{\gamma_2} \cdots \sigma_N^{\gamma_N} \rangle$. The basic idea is that δ should be much smaller than the typical magnitude of $\langle \sigma_1^{\gamma_1} \sigma_2^{\gamma_2} \cdots \sigma_N^{\gamma_N} \rangle$ to ensure a high signal-to-noise ratio. The root mean square value of $\langle \sigma_1^{\gamma_1} \sigma_2^{\gamma_2} \cdots \sigma_N^{\gamma_N} \rangle$ over all values of $\gamma_1, \dots, \gamma_N$ in a target state ρ can be estimated using:

$$\langle \sigma_1^{\gamma_1} \cdots \sigma_N^{\gamma_N} \rangle_{\text{rms}} = \sqrt{\frac{1}{4^N} \sum_{\gamma_1, \dots, \gamma_N} |\langle \sigma_1^{\gamma_1} \cdots \sigma_N^{\gamma_N} \rangle_\rho|^2} = \sqrt{\frac{\text{Tr}(\rho^2)}{2^N}}. \quad (16)$$

Therefore, for the statistical errors to be small, we require

$$\delta = \epsilon \sqrt{\frac{\text{Tr}(\rho^2)}{2^N}} \text{ and } M \geq \frac{1}{\delta^2} = \frac{2^N}{\epsilon^2 \text{Tr}(\rho^2)} \quad (17)$$

where ϵ is a small number that quantifies the average relative error in measuring $\langle \sigma_1^{\gamma_1} \sigma_2^{\gamma_2} \cdots \sigma_N^{\gamma_N} \rangle$. Since $2^{-N} \leq \text{Tr}(\rho^2) \leq 1$, the number of state copies per measurement basis scales as $O(2^N)$ to $O(4^N)$, which makes the protocol not scalable. However, we emphasize that this appears to be a fundamental limitation also shared by other QST methods. For example, for an unstructured state with a full rank density operator, the best known protocol for QST using the same local Pauli measurements as ours requires a total number of state copies given by $O(6^N/(\epsilon^2 \text{Tr}(\rho^2)))$ [11] for $D = \epsilon^2$ between the target state and the reconstructed state. Thus the number of state copies per basis in such a protocol is the same as our Eq. (17). Our cross approximation based protocol still has the major advantage of using only $O(N)$ instead of 3^N measurement bases, thus requiring a much smaller number of total state copies.

We now show that the DMRG-cross algorithm can indeed tolerate a small amount of measurement error quantified by a small relative error threshold ϵ . Since the number of measurements per basis scales exponentially in N , for practical reasons we limit our study to $N \leq 12$ and the maximum bond dimension in DMRG-cross to 6. We choose the target states to be either the thermal states of Eq. (5) with $T = 0.2$ or the random LPTN states with bond dimension $\chi = 16$, which have been studied in Section II. To simulate the statistical errors on the quantum measurements, we add Gaussian random noise with a standard deviation given by δ defined in Eq. (17) to each value of $\langle \sigma_1^{\gamma_1} \sigma_2^{\gamma_2} \cdots \sigma_N^{\gamma_N} \rangle$ used in the cross approximation, with $\epsilon = 0.01$. Due to such randomness in each numerical experiment, we average D and N_b over 80 repeated experiments. The results are shown in Fig. 5, where we clearly see that the distance measure D is larger than that in Fig. 3 without measurement errors. However, the increase of D with N is again slow. This is consistent with our recent work [40] showing that the reconstruction error for tensor train approximation scales at most polynomially in N for a finite relative error ϵ in the measurements.

IV. IMPROVEMENT USING MACHINE LEARNING

As we mentioned in Section II, each expectation value $\langle \sigma_1^{\gamma_1} \sigma_2^{\gamma_2} \cdots \sigma_N^{\gamma_N} \rangle$ used in the tensor train cross approximation is obtained by measuring $\sigma_i^{\gamma_i}$ in the target state for all qubits. This means each state copy actually provides N bits of information, as each qubit will yield a measurement outcome of either 1 or -1 upon the measurement of $\sigma^{x,y,z}$. However, we only use one bit of information in evaluating $\langle \sigma_1^{\gamma_1} \sigma_2^{\gamma_2} \cdots \sigma_N^{\gamma_N} \rangle$ from each state copy. This means the experimental measurement data contains far more information than what we used in the cross approximation. The full information of the experimental data can instead be captured by the expectation values of all measured operators $\sigma_1^{\gamma_1} \cdots \sigma_N^{\gamma_N}$ with no $\gamma_i = 0$ plus

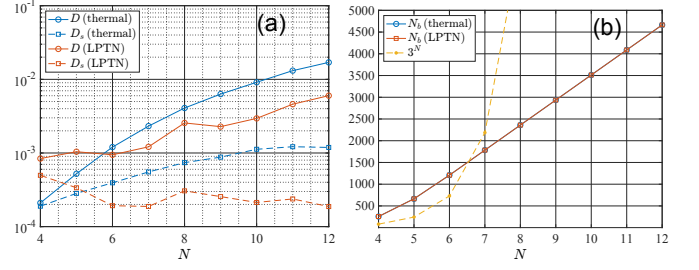


Figure 5. (a) Distance measures (D and D_s) between the target states and the reconstructed states using tensor train cross approximation with measurement errors quantified by $\epsilon = 0.01$ (see main text). The target states are either thermal states of a quantum Ising model with $T = 0.2$ or random LPTN states with $\chi = 16$. (b) The number of measurement bases N_b needed for the cross approximation to reconstruct both target states, compared again with 3^N .

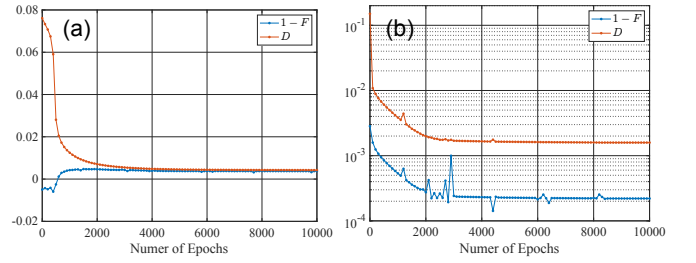


Figure 6. Distance measure D and infidelity $1-F$ between the target states and the reconstructed states obtained using tensor train cross approximation and further trained by supervised machine learning. Statistical errors are included by simulating $M = 10^6$ repeated measurements per basis. The target state is either a thermal state of a quantum Ising model with $T = 1$ (a) or a random LPTN state with $\chi = 16$ (b), both with $N = 8$ qubits. The horizontal axes represent the training process.

those with any one or more γ_i replaced by zero. Alternatively, we can use the full counting statistics of $\sigma_1^{\gamma_1}, \dots, \sigma_N^{\gamma_N}$ to capture the full information of measurement data, which is much more efficient for a large N .

In this section, we show that such extra information contained in the experimental data but unused by the cross approximation can be harnessed to improve the quality of the QST via a supervised machine learning procedure. Our idea is simple: We can use the MPO $\{\tilde{G}_i^{\gamma_i}\}$ obtained using the tensor train cross approximation as a good initial guess, and use the extra entries of \mathcal{A} calculated from the same experimental data to refine $\{\tilde{G}_i^{\gamma_i}\}$ via supervised learning. The loss function during the training is defined as

$$L = \sum_{\gamma_1, \dots, \gamma_N} |\langle \sigma_1^{\gamma_1} \cdots \sigma_N^{\gamma_N} \rangle_{\rho_1} - \langle \sigma_1^{\gamma_1} \cdots \sigma_N^{\gamma_N} \rangle_{\rho_2}|^2 \quad (18)$$

where $\sum_{\gamma_1, \dots, \gamma_N}$ denotes the summation over all indices $\gamma_1, \dots, \gamma_N$ of which $\langle \sigma_1^{\gamma_1} \sigma_2^{\gamma_2} \cdots \sigma_N^{\gamma_N} \rangle$ can be computed from the experimental data, each subject to statistical errors quantified in Section III. ρ_1 here denotes the experimental state being measured and ρ_2 denotes the state represented by the

MPO being trained. Importantly, L can be calculated just based on the experimental data, without knowing the full target state. We then perform a stochastic gradient descent on the loss function L over the parameters in the MPO being trained to minimize L using an adaptive moment estimation (Adam) method [59].

To benchmark this supervised machine learning method, we set $N = 8$ and use either a target state that is a thermal state of Eq. (5) with $T = 1$ or a random LPTN state with bond dimension $\chi = 16$. We use $M = 10^6$ repeated measurements per basis and a total number of roughly 1000 measurement bases as required by the cross approximation. For both target states, the tensor train cross approximation gives us a reconstructed MPO with $D \approx 0.1$, which is reasonably good but not ideal due to a finite statistic error. As shown in Fig. 6, supervised machine learning is able to refine the MPO obtained using cross approximation significantly. For the thermal state, the distance measure D drops by a factor of about 20, while for the random LPTN state D drops by a factor of almost 100. Since the system size is small, here we can also compute the infidelity $1 - F$ between the target state ρ_1 and the reconstructed state ρ_2 (obtained from their MPO representations), with the fidelity defined as $F \equiv \text{Tr} \sqrt{\sqrt{\rho_1} \rho_2 \sqrt{\rho_1}}$. For the thermal state, the infidelity is negative initially, showing that the reconstructed density matrix is not strictly semi-positive definite. But the supervised machine learning is able to correct this and produces a positive density matrix with a fidelity of $F \approx 99.6\%$ at the end of the training. For the random LPTN state, we were able to improve the fidelity from 99.7% initially to 99.98%.

V. DISCUSSION AND OUTLOOK

The supervised machine learning introduced in the previous section can noticeably reduce the number of repeated measurements M per basis while maintaining the same level of reconstruction error. However, we do not expect it to overcome the exponential scaling of M in the system size N . To make the QST fully scalable, we can combine the tensor train cross approximation method with the MPO concatenation method in Refs. [9, 60]. First, we note that if an N -qubit mixed state can be represented by an MPO with a maximum bond dimension χ , then any R -qubit reduced state can also be represented by an MPO with a bond dimension of at most χ . As a result, if the MPO for the full state satisfies the invertibility condition in Ref. [9] such that it can be reconstructed via the reduced states of R consecutive qubits, we can use the tensor train cross approximation to perform QST on $O(N)$ of such R consecutive qubits, as long as R is not too large. This is a significant improvement over tomography methods that do

not take advantage of the structure of the reduced states and hence require exponentially more (in N) measurements. Such improvement can be seen for R as small as 7 based on the examples we studied (see Fig. 5b).

We emphasize that to our best knowledge, no provably efficient QST protocol has been found for generic physical states with a compact MPO representation. Our work highlights an important roadblock towards finding such efficient protocol. Although a compact MPO state has only $O(N)$ independent parameters, the state in general contains an exponentially large (in N) number of nonzero parameters and thus each parameter has an exponentially small typical value. Measuring even $O(N)$ such parameters to a small relative error would thus require an exponentially large number of measurements. This fundamental limitation seems only avoidable if the full state can be obtained by local reductions such that one just need to perform QST on small subsystems instead. However, the existence of such local reductions cannot be guaranteed [9, 60]. A very useful future direction is to derive upper bounds on the size R in the local reductions, provided that the target state is represented by an MPO with a finite bond dimension. This may be possible especially if we can tolerate a small amount of error in the reconstruction of the target state.

On the other hand, one can try QST protocols where no single parameter of the density matrix is measured to a good precision. For example, Ref. [23] shows that one can perform just one measurement per basis and choose a large number of random local bases to reconstruct an MPS target state very efficiently. This is an orthogonal approach from ours that requires a large number of measurements per basis but a minimal number of bases. However, how to generalize this approach to mixed states represented by MPOs is another interesting open question. In addition, instead of choosing the measurement bases randomly in such an approach, one can adaptively choose new bases based on the previous measurement outcomes. Such adaptive measurement protocols have been shown to perform better in QST [10, 33] than those using independent measurement bases. In fact, the maximum volume principle based tensor cross approximation used in this work is already such an adaptive protocol. Whether adaptive measurement protocols can eventually lead to efficient QST for generic MPO target states remains to be seen.

ACKNOWLEDGMENTS

We thank the HPC center at Colorado School of Mines for providing computational resources needed in carrying out this work. We acknowledge funding support from NSF Grants No. CCF-1839232, PHY-2112893, CCF-2106834 and CCF-2106881, as well as the W. M. Keck Foundation.

-
- [1] S. Aaronson, *SIAM Journal on Computing* **49**, STOC18 (2019).
 - [2] G. Struchalin, Y. A. Zagorovskii, E. Kovlakov, S. Straupe, and S. Kulik, *PRX Quantum* **2**, 010307 (2021).

- [3] A. W. Smith, J. Gray, and M. Kim, *PRX Quantum* **2**, 020348 (2021).
- [4] H.-Y. Huang, R. Kueng, and J. Preskill, *Nature Physics* **16**, 1050

- (2020).
- [5] C. Harney, S. Pirandola, A. Ferraro, and M. Paternostro, *New Journal of Physics* **22**, 045001 (2020).
 - [6] G. Tóth, W. Wieczorek, D. Gross, R. Krischek, C. Schwemmer, and H. Weinfurter, *Phys. Rev. Lett.* **105**, 250403 (2010).
 - [7] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, *Nature communications* **1**, 1 (2010).
 - [8] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, *New Journal of Physics* **14**, 095022 (2012).
 - [9] T. Baumgratz, D. Gross, M. Cramer, and M. B. Plenio, *Phys. Rev. Lett.* **111**, 020401 (2013).
 - [10] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, *IEEE Transactions on Information Theory* **63**, 5628 (2017).
 - [11] M. Guță, J. Kahn, R. Kueng, and J. A. Tropp, *Journal of Physics A: Mathematical and Theoretical* **53**, 204001 (2020).
 - [12] D. S. França, F. G. L. Brandão, and R. Kueng, in *16th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2021)* (2021) pp. 7:1–7:13.
 - [13] Y.-K. Liu, *Advances in Neural Information Processing Systems* **24** (2011).
 - [14] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, *Phys. Rev. Lett.* **105**, 150401 (2010).
 - [15] A. Kyrillidis, A. Kalev, D. Park, S. Bhojanapalli, C. Caramanis, and S. Sanghavi, *npj Quantum Information* **4**, 1 (2018).
 - [16] C. H. Baldwin, I. H. Deutsch, and A. Kalev, *Phys. Rev. A* **93**, 052105 (2016).
 - [17] J. L. Kim, G. Kollias, A. Kalev, K. X. Wei, and A. Kyrillidis, *arXiv preprint arXiv:2104.07006* (2021).
 - [18] D. Gross, *IEEE Transactions on Information Theory* **57**, 1548 (2011).
 - [19] U. Schollwöck, *Annals of Physics* **326**, 96 (2011), january 2011 Special Issue.
 - [20] J. c. v. Guth Jarkovský, A. Molnár, N. Schuch, and J. I. Cirac, *PRX Quantum* **1**, 010304 (2020).
 - [21] R. Orús, *Annals of Physics* **349**, 117 (2014).
 - [22] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, *Nature Physics* **14**, 447 (2018).
 - [23] J. Wang, Z.-Y. Han, S.-B. Wang, Z. Li, L.-Z. Mu, H. Fan, and L. Wang, *Phys. Rev. A* **101**, 032321 (2020).
 - [24] B. Lanyon, C. Maier, M. Holzäpfel, T. Baumgratz, C. Hempel, P. Jurcevic, I. Dhand, A. Buyskikh, A. Daley, M. Cramer, *et al.*, *Nature Physics* **13**, 1158 (2017).
 - [25] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, *Nature Machine Intelligence* **1**, 155 (2019).
 - [26] T. Schmale, M. Reh, and M. Gärtner, *arXiv preprint arXiv:2109.13776* (2021).
 - [27] A. Rocchetto, E. Grant, S. Strelchuk, G. Carleo, and S. Severini, *npj Quantum Information* **4**, 1 (2018).
 - [28] G. Torlai, B. Timar, E. P. L. van Nieuwenburg, H. Levine, A. Omran, A. Keesling, H. Bernien, M. Greiner, V. Vuletić, M. D. Lukin, R. G. Melko, and M. Endres, *Phys. Rev. Lett.* **123**, 230504 (2019).
 - [29] T. Xin, S. Lu, N. Cao, G. Anikeeva, D. Lu, J. Li, G. Long, and B. Zeng, *npj Quantum Information* **5**, 1 (2019).
 - [30] A. M. Palmieri, E. Kovlakov, F. Bianchi, D. Yudin, S. Straupe, J. D. Biamonte, and S. Kulik, *npj Quantum Information* **6**, 1 (2020).
 - [31] G. Torlai and R. G. Melko, *Phys. Rev. Lett.* **120**, 240503 (2018).
 - [32] S. Ahmed, C. Sánchez Muñoz, F. Nori, and A. F. Kockum, *Phys. Rev. Lett.* **127**, 140502 (2021).
 - [33] H. Lange, M. Kebrič, M. Buser, U. Schollwöck, F. Grusdt, and A. Bohrdt, *arXiv preprint arXiv:2203.15719* (2022).
 - [34] Y. Quek, S. Fort, and H. K. Ng, *npj Quantum Information* **7**, 1 (2021).
 - [35] K. Noh, L. Jiang, and B. Fefferman, *Quantum* **4**, 318 (2020).
 - [36] C. Monroe, W. C. Campbell, L.-M. Duan, Z.-X. Gong, A. V. Gorshkov, P. W. Hess, R. Islam, K. Kim, N. M. Linke, G. Pagano, P. Richerme, C. Senko, and N. Y. Yao, *Rev. Mod. Phys.* **93**, 025001 (2021).
 - [37] I. Oseledets and E. Tyrtshnikov, *Linear Algebra and its Applications* **432**, 70 (2010).
 - [38] D. Savostyanov and I. Oseledets, in *The 2011 International Workshop on Multidimensional (nD) Systems* (2011) pp. 1–8.
 - [39] D. V. Savostyanov, *Linear Algebra and its Applications* **458**, 217 (2014).
 - [40] Z. Qin, A. Lidiak, Z. Gong, G. Tang, M. Wakin, and Z. Zhu, *arXiv:2207.0432* (2022).
 - [41] M. Holzäpfel, M. Cramer, N. Datta, and M. B. Plenio, *Journal of Mathematical Physics* **59**, 042201 (2018).
 - [42] C. Boutsidis and D. P. Woodruff, *SIAM Journal on Computing* **46**, 543 (2017).
 - [43] N. Mitrovic, M. T. Asif, U. Rasheed, J. Dauwels, and P. Jaillet, in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)* (IEEE, 2013) pp. 1475–1480.
 - [44] A. Aldroubi, K. Hamm, A. B. Koku, and A. Sekmen, *Frontiers in Applied Mathematics and Statistics*, 65 (2019).
 - [45] M. Xu, R. Jin, and Z.-H. Zhou, in *International Conference on Machine Learning* (PMLR, 2015) pp. 1412–1421.
 - [46] A. Civril and M. Magdon-Ismael, *Theoretical Computer Science* **410**, 4801 (2009).
 - [47] A. Deshpande and L. Rademacher, in *2010 IEEE 51st annual symposium on foundations of computer science* (IEEE, 2010) pp. 329–338.
 - [48] A. Deshpande, L. Rademacher, S. S. Vempala, and G. Wang, *Theory of Computing* **2**, 225 (2006).
 - [49] A. Cortinovis and D. Kressner, *SIAM Journal on Matrix Analysis and Applications* **41**, 1651 (2020).
 - [50] S. A. Goreinov and E. E. Tyrtshnikov, *Contemporary Mathematics* **280**, 47 (2001).
 - [51] Y. Chen, S. Bhojanapalli, S. Sanghavi, and R. Ward, *The Journal of Machine Learning Research* **16**, 2999 (2015).
 - [52] V. N. Temlyakov, *Acta Numerica* **17**, 235–409 (2008).
 - [53] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac, *Phys. Rev. Lett.* **93**, 207204 (2004).
 - [54] M. J. D. Jäschke, M. L. Wall and L. D. Carr, Matrix product state open source code., <http://sourceforge.net/projects/openmps/>.
 - [55] D. Jäschke, M. L. Wall, and L. D. Carr, *Computer Physics Communications* **225**, 59 (2018).
 - [56] A. H. Werner, D. Jäschke, P. Silvi, M. Kliesch, T. Calarco, J. Eisert, and S. Montangero, *Phys. Rev. Lett.* **116**, 237201 (2016).
 - [57] M. Nielsen and I. Chuang, *Quantum computation and quantum information*, 10th ed. (Cambridge University Press, Cambridge, 2010).
 - [58] R. O’Donnell and J. Wright, in *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing* (ACM, 2016).
 - [59] D. P. Kingma and J. Ba, *arXiv preprint arXiv:1412.6980* (2014).
 - [60] T. Baumgratz, A. Nüßeler, M. Cramer, and M. B. Plenio, *New Journal of Physics* **15**, 125004 (2013).