

Density matrix reconstruction using non-negative matrix product states

Donghong Han ¹, Chu Guo ^{2,3,*}, and Xiaoting Wang ^{1,†}

¹*Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, 610051, China*

²*Henan Key Laboratory of Quantum Information and Cryptography, Zhengzhou, Henan, 450000, China*

³*Key Laboratory of Low-Dimensional Quantum Structures and Quantum Control of Ministry of Education, Department of Physics and Synergetic Innovation Center for Quantum Effects and Applications, Hunan Normal University, Changsha 410081, China*



(Received 12 May 2022; accepted 4 October 2022; published 21 October 2022)

Quantum state tomography is a key technique for quantum information processing but is challenging due to the exponential growth of its complexity with the system size. In this work we propose an algorithm which iteratively finds the best non-negative matrix product state approximation based on a set of measurement outcomes whose size does not necessarily grow exponentially. Compared to the tomography method based on neural network states, our scheme utilizes a so-called tensor train representation that allows straightforward recovery of the unknown density matrix in the matrix product operator form. As applications, the effectiveness of our algorithm is numerically demonstrated to reconstruct the ground state of the XXZ spin chain under depolarizing noise.

DOI: [10.1103/PhysRevA.106.042435](https://doi.org/10.1103/PhysRevA.106.042435)

I. INTRODUCTION

Characterizing an unknown quantum state is of central importance in developing quantum technologies. Standard quantum state tomography (QST) reconstructs a generic quantum state by performing projective measurements on an informationally complete basis [1,2]. The number of projective measurements required grows exponentially with the system size. In the meanwhile, current quantum technologies have pushed the number of qubits to close to 100 [3–5], and scalable quantum state tomography schemes are in great need. With additional assumptions on the underlying quantum state, more efficient schemes than the standard QST have been proposed, for example, QST for a sparse quantum state by compressive sensing [6–9], QST for quantum states which are permutationally invariant [10,11], and QST for quantum states which can be efficiently represented with a low-depth parametric quantum circuit [12]. In particular, QST methods based on tensor network representation [13,14] as well as neural network ansatz [15–19] are promising to extend QST to a much larger scale, and both approaches have been demonstrated on several tens of qubits based on synthetic data. We also note those efficient QST methods in case only partial information of the unknown quantum state is required [20–22].

For an unknown L -qubit pure state that can be well approximated by a matrix product state (MPS), it is proved that a set of $O(L)$ local n -body reduced density matrix tomography suffices to reconstruct the unknown state, where n is a constant and independent of L if the underlying pure state has bounded entanglement. Thus only $O(\text{poly}(L))$ number of measurements are required [13,14,23]. It has also been shown that a similar approach can be applied to reconstructing an unknown mixed state, with an additional assumption of its

invertibility [24]. However, such methods based on local density matrices are not easy to implement in practice, since (1) exact tomography of a series of local density matrices may already be hard, and (2) we can only reconstruct approximate density matrices by local tomography using a finite number of measurements, and the approximation errors could accumulate and affect the overall tomography performance of the entire state. Another method based on an MPS ansatz is proposed using an unsupervised machine learning algorithm, where only global measurement data on a randomly prepared basis are required [25]. However, such a method only considers the reconstruction of pure states. Neural network state-based algorithms constitute another important class of heuristic approaches for QST with excellent precision and scalability in practice. Specifically, neural network states have been used to model (1) the pure states [15,26] or the density matrices [16] as classical neural networks, and (2) the measured probability distributions [17] as classical neural networks. In the latter approach, it will generally be exponentially hard to further reconstruct the state as a vector or a density matrix from the probability distribution. In the former approach, one could efficiently compute amplitudes based on the trained neural network state, but for other tasks such as computing expectation values one still needs to perform a sampling process based on the trained neural network state, which may not be as convenient or efficient, if the underlying quantum state could be well approximated and written as an MPS. Another possible drawback of the QST methods based on neural network states is, *a priori*, it is not clear which neural network representation is suitable for an unknown quantum state.

Inspired by the fact that MPS has been adapted to represent the multivariate probability distribution function, often referred to as the tensor train representation [27], we propose a QST scheme that combines the advantages of both the tensor network approach and the neural network approach. Specifically, in the first stage, a tensor train representation of the multivariate distribution function is constructed based on

*guochu604b@gmail.com

†xiaoting@uestc.edu.cn

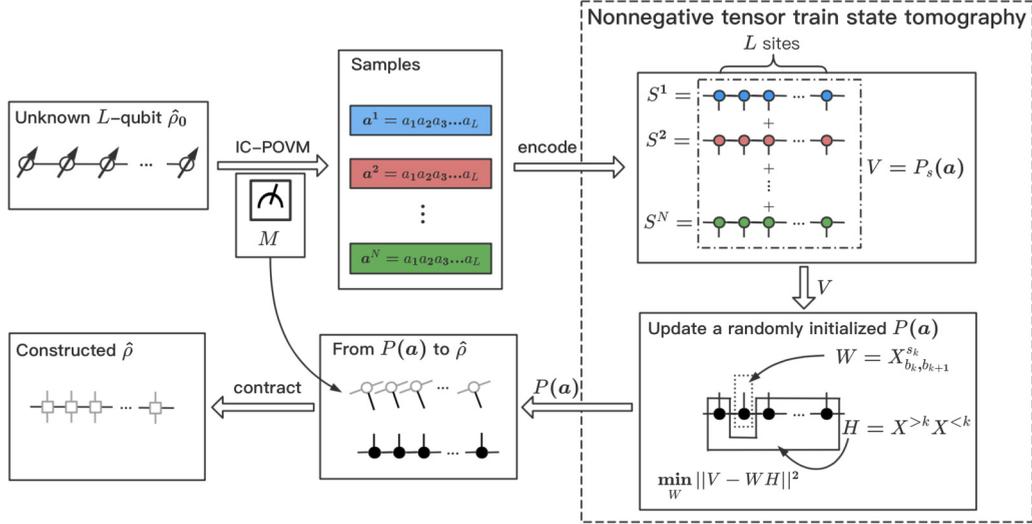


FIG. 1. A flowchart of our non-negative tensor train state tomography algorithm. Given an unknown L -qubit quantum state $\hat{\rho}_0$, an IC-POVM is performed to obtain a sample of bit strings $\{a^j\}$, each of which is then encoded into a *one hot* MPS S^j . Based on these S^j , an optimal non-negative MPS $P(a)$ with a fixed bond dimension is found through optimization, satisfying that it is closest to $P_s(a)$, namely, the superposition of all states S^j , by optimizing each site tensor of the randomly initialized MPS ansatz $P(a)$. Then an MPO $\hat{\rho}$ is reconstructed, with the same bond dimension as $P(a)$, by applying the inverse of the IC-POVM locally on each site of $P(a)$ using Eq. (8). Simulations in this work are based on synthetic data.

the quantum measurement data instead of a neural network representation, and then a density matrix renormalization group (DMRG)-like algorithm is used to find the optimal tensor train representation. After that, the tensor train is transformed back into a matrix product operator (MPO) representation of the unknown density matrix. Compared to the existing QST methods based on tensor network states, our scheme directly uses a tensor train representation for the multivariate probability distribution function, instead of the unknown density matrix. Compared to the existing QST methods based on neural network states, our scheme could easily construct the unknown density matrix as an MPO, which usually allows more convenient and efficient evaluations of observables. The flowchart of our algorithm is summarized in Fig. 1. This work is organized as follows: we show the details of our QST scheme in Sec. II and then numerically demonstrate our algorithm for the ground state of the XXZ chain perturbed by depolarizing noise in Sec. III. We conclude in Sec. IV.

II. METHOD

For QST, we use quantum measurements given by informationally complete (IC) positive operator-valued measures (POVMs) [28–30], which describes the most general quantum measurements allowed by quantum theory [31]. We denote the single-qubit IC-POVM as $\{M^s\}$, where each M^s is a positive semidefinite matrix satisfying $\sum_s M^s = \hat{I}$, with \hat{I} the identity matrix. For a single qubit, a minimal IC-POVM can be chosen as $M^s = \frac{1}{2} |\psi^s\rangle \langle \psi^s|$ with

$$|\psi^0\rangle = |0\rangle, \quad (1)$$

$$|\psi^1\rangle = \sqrt{\frac{1}{3}} |0\rangle + \sqrt{\frac{2}{3}} |1\rangle, \quad (2)$$

$$|\psi^2\rangle = \sqrt{\frac{1}{3}} |0\rangle + \sqrt{\frac{2}{3}} e^{i\frac{2\pi}{3}} |1\rangle, \quad (3)$$

$$|\psi^3\rangle = \sqrt{\frac{1}{3}} |0\rangle + \sqrt{\frac{2}{3}} e^{i\frac{4\pi}{3}} |1\rangle, \quad (4)$$

which form the vertices of a regular tetrahedron in the Bloch sphere [32]. The single-qubit IC-POVM $\{M^s\}_{s=0,1,2,3}$ can also be viewed as a three-dimensional tensor in total, written as $M_{\sigma,\sigma'}^s$ with two physical indices σ, σ' of dimension 2 and another index s of dimension 4 corresponding to different measurement outcomes. If we group the two physical indices σ and σ' together, then $M_{\sigma,\sigma'}^s$ is a 4×4 invertible matrix, representing a one-to-one mapping between the single-qubit density matrix and the single-qubit probability distribution.

As in [17], for an L -qubit quantum system, we consider the quantum measurement defined by

$$M^{\otimes a} \equiv M^{a_1} \otimes M^{a_2} \otimes \dots \otimes M^{a_L}, \quad (5)$$

where $\mathbf{a} = (a_1, \dots, a_L)$ represents a string of integers specifying the local projectors, and each integer $a_l \in \{0, 1, 2, 3\}$. The probability distribution $P(\mathbf{a})$ forms an L -variable distribution function in which each local dimension d is equal to 4, satisfying $P(\mathbf{a}) \geq 0$ and $\sum_{\mathbf{a}} P(\mathbf{a}) = 1$. Interestingly, if we assume that the unknown quantum state can be efficiently represented as an MPO,

$$\hat{\rho} = \sum_{b_1, b_2, \dots, b_{L+1}} W_{b_1, b_2}^{\sigma_1, \sigma'_1} W_{b_2, b_3}^{\sigma_2, \sigma'_2} \dots W_{b_L, b_{L+1}}^{\sigma_L, \sigma'_L}, \quad (6)$$

where b_l denotes the auxiliary index, then $P(\mathbf{a})$ can be written as an MPS

$$P(\mathbf{a}) = \sum_{b_1, b_2, \dots, b_{L+1}} X_{b_1, b_2}^{s_1} X_{b_2, b_3}^{s_2} \dots X_{b_L, b_{L+1}}^{s_L}, \quad (7)$$

with each site tensor $X_{b_l, b_{l+1}}^{s_l} = \sum_{\sigma_l, \sigma'_l} W_{b_l, b_{l+1}}^{\sigma_l, \sigma'_l} M_{\sigma_l, \sigma'_l}^{s_l}$. Therefore the bond dimensions of $P(\mathbf{a})$, which are defined as the dimensions of the auxiliary indices $D_l = \dim(b_l)$, are exactly the same as the bond dimensions of $\hat{\rho}$. From Eqs. (6) and (7), we can see that the total number of parameters for both $\hat{\rho}$ and $P(\mathbf{a})$ are bounded by $4LD^2$ if we fix the dimension of all the auxiliary indices to be D , due to the fact that each local-site tensor has at most $4D^2$ parameters. Since each $M_{\sigma_l, \sigma'_l}^{s_l}$ is invertible (the two indices σ_l and σ'_l are assumed to be grouped together), we can obtain each $W_{b_l, b_{l+1}}^{\sigma_l, \sigma'_l}$ from $X_{b_l, b_{l+1}}^{s_l}$ by

$$W_{b_l, b_{l+1}}^{\sigma_l, \sigma'_l} = \sum_{s_l} X_{b_l, b_{l+1}}^{s_l} (M^{-1})_{\sigma_l, \sigma'_l}^{s_l}. \quad (8)$$

Therefore we can efficiently transform back and forth between the density matrix $\hat{\rho}$ and the probability distribution $P(\mathbf{a})$. In this work we will first reconstruct $P(\mathbf{a})$ as an MPS and then we transform it back into a density matrix in MPO form using Eq. (8). One advantage of this approach is that as long as the constructed $P(\mathbf{a})$ is a proper probability distribution, then the density matrix $\hat{\rho}$ from this approach will automatically be Hermitian (which, however, may not be positive if there are not enough measurement data).

Thus the problem reduces to reconstructing an approximate probability distribution $P(\mathbf{a})$ closest to the ideal distribution corresponding to the unknown quantum state, denoted as $P_i(\mathbf{a})$, based on a set of N samples $\mathbf{a}^1, \mathbf{a}^2, \dots, \mathbf{a}^N$ collected from experiment. We assume that among those samples there are only N_s different ones, denoted as $\mathbf{a}^1, \mathbf{a}^2, \dots, \mathbf{a}^{N_s}$, where each distinct sample \mathbf{a}^j has multiplicity n_j , namely, each \mathbf{a}^j appears with an *observed probability* $P_s(\mathbf{a}^j) = \frac{n_j}{N}$, $j = 1, \dots, N_s$. $P_s(\mathbf{a}^j)$ will eventually converge to the exact distribution $P_i(\mathbf{a}^j)$ as N increases. For a limited number of samples, $P_s(\mathbf{a})$ are only approximately equal to the corresponding entries of $P_i(\mathbf{a})$ for each \mathbf{a} . Now the original QST is reduced to the following mathematical problem: given some approximated values, namely, $P_s(\mathbf{a})$ of the nonzero elements for an unknown $P_i(\mathbf{a})$, how can we construct a tensor train approximation with a minimum bond dimension D , denoted as $P(\mathbf{a})$, which is closest to $P_i(\mathbf{a})$?

To this end we note that each sample \mathbf{a}^j can be encoded as a *one hot* MPS S^j as

$$\mathbf{a}^j \mapsto S^j = \sum_{c_1, c_2, \dots, c_{L+1}} A_{j, c_1, c_2}^{s_1} A_{j, c_2, c_3}^{s_2} \dots A_{j, c_L, c_{L+1}}^{s_L}, \quad (9)$$

such that $\dim(c_l) = 1$ and $\dim(s_l) = 4$ for all $1 \leq l \leq L$, and each tensor $A_{j, c_l, c_{l+1}}^{s_l}$ satisfies $A_{j, 0, 0}^{s_l} = 1$ iff s_l equals a_l , and 0 otherwise. For example, for a specific \mathbf{a}^j , if a_l in \mathbf{a}^j is 3, then $A_{j, c_l, c_{l+1}}^{s_l}$ satisfies $A_{j, 0, 0}^0 = 0$, $A_{j, 0, 0}^1 = 0$, $A_{j, 0, 0}^2 = 0$, and $A_{j, 0, 0}^3 = 1$. With Eq. (9) we can rewrite the probability distribution formed by N samples as

$$P_s(\mathbf{a}) = \sum_{j=1}^{N_s} \frac{n_j}{N} S^j, \quad (10)$$

where $P_s(\mathbf{a})$ can be viewed as a superposition of all observed one hot states S^j , weighted by their multiplicities. $P_s(\mathbf{a})$ can be directly taken as the best approximation for $P_i(\mathbf{a})$, that

is, setting $P(\mathbf{a}) = P_s(\mathbf{a})$, $P_s(\mathbf{a})$ can be directly evaluated from Eq. (10) using simple MPS arithmetics. However, given a limited set of samples, the bond dimension of the resulting MPS could be much larger than that of the target distribution. Moreover, this might result in an over-fitting problem, since $P_s(\mathbf{a})$ will be perfect for known samples but will be 0 for unknown samples. For better efficiency and generalizability, one can search for $P(\mathbf{a})$, which is approximately equal to $P_s(\mathbf{a})$, under the condition that the bond dimension is bounded by a fixed value D . This could be done in two approaches [33]: (1) evaluating Eq. (10) exactly and then compressing the resulting MPS using singular value decomposition (SVD), and (2) iteratively searching for the solution to the following optimization problem:

$$\text{Loss}(P(\mathbf{a})) \equiv \sum_{j=1}^{N_s} \|P(\mathbf{a}) - P_s(\mathbf{a}^j)\|^2, \quad (11)$$

with a maximal bond dimension D , where $\|P\|$ denotes the Frobenius norm of the tensor P . We will follow the latter approach which is more precise in practice. One complication here is that if one directly uses the approaches in [33,34], where the MPS ansatz is kept in a canonical form by iteratively using either SVD or QR decomposition, then the solution generally contains negative values, which is undesirable for a probability distribution. To ensure the non-negativity of $P(\mathbf{a})$, one could represent $P(\mathbf{a})$ as a non-negative MPS instead, that is, each site tensor $X_{b_l, b_{l+1}}^{s_l}$ in Eq. (7) is non-negative. Several algorithms have been proposed to approximate a target probability distribution using a non-negative MPS with a fixed bond dimension [35–37]. Here we use a refined approach based on [37], the central idea of which is to use a non-negative matrix decomposition instead of SVD or QR decomposition.

Specifically, we first define the following tensors $X_{b_{k+1}; s_{k+1}, \dots, s_L}^{>k}$ and $X_{s_1, \dots, s_{k-1}; b_k}^{<k}$:

$$X^{>k} = \sum_{b_{k+2}, \dots, b_{L+1}} X_{b_{k+1}, b_{k+2}}^{s_{k+1}} \dots X_{b_L, b_{L+1}}^{s_L}, \quad (12a)$$

$$X^{<k} = \sum_{b_1, \dots, b_{k-1}} X_{b_1, b_2}^{s_1} \dots X_{b_{k-1}, b_k}^{s_{k-1}}, \quad (12b)$$

and $G_{b_{k+1}, b'_{k+1}}^{>k}$ and $G_{b_k, b'_k}^{<k}$:

$$G^{>k} = \sum_{s_{k+1}, \dots, s_L} X_{b_{k+1}; s_{k+1}, \dots, s_L}^{>k} X_{b'_{k+1}; s_{k+1}, \dots, s_L}^{>k}, \quad (13a)$$

$$G^{<k} = \sum_{s_1, \dots, s_{k-1}} X_{s_1, \dots, s_{k-1}; b_k}^{<k} X_{s_1, \dots, s_{k-1}; b'_k}^{<k}. \quad (13b)$$

With Eqs. (12a) and (12b) we can rewrite $P(\mathbf{a})$ as

$$P(\mathbf{a}) = \sum_{b_k, b_{k+1}} X_{b_k, b_{k+1}}^{s_k} X_{\dots, s_{k-1}; b_k}^{<k} X_{b_{k+1}; s_{k+1}, \dots}^{>k}, \quad (14)$$

for each $1 \leq k \leq L$. Substituting Eq. (14) into Eq. (11), the loss function becomes $\|V - WH\|^2$, with $V = P_s(\mathbf{a})$, $W = X_{b_k, b_{k+1}}^{s_k}$, and $H = X_{\dots, s_{k-1}; b_k}^{<k} X_{b_{k+1}; s_{k+1}, \dots}^{>k}$. Thus the goal is to find the best non-negative factorization of V . One of the most well-known approaches for solving this problem is the following

updating rule:

$$W \leftarrow W \circ \frac{[VH^t]}{[WHH^t]}, \quad (15a)$$

$$H \leftarrow H \circ \frac{[W^tV]}{[W^tWH]}, \quad (15b)$$

where \circ means Hadamard (element-wise) product, and $\frac{[A]}{[B]}$ denotes the element-wise division of the matrices A and B [38]. With Eqs. (15a) and (15b) the loss function is guaranteed to decrease monotonically. From Eq. (15a), one can update the tensor $X_{b_k, b_{k+1}}^{s_k}$ as

$$X_{b_k, b_{k+1}}^{s_k} \leftarrow X_{b_k, b_{k+1}}^{s_k} \circ \frac{[\sum_{s_l \neq k} P_s(\mathbf{a}) X^{<k} X^{>k}]}{[\sum_{b'_k, b'_{k+1}} X_{b'_k, b'_{k+1}}^{s_k} G_{b_{k+1}, b'_{k+1}}^{>k} G_{b_k, b'_k}^{<k}]}. \quad (16)$$

The denominator on the right-hand side of Eq. (16) can be efficiently evaluated without computing the summation in Eq. (10), for which we define two tensors $\tilde{G}_{j, b_{k+1}, c_{k+1}}^{>k}$ and $\tilde{G}_{j, b_k, c_k}^{<k}$:

$$\tilde{G}_j^{>k} = \sum_{s_{k+1}, \dots, s_L} X_{b_{k+1}, s_{k+1}, \dots, s_L}^{>k} A_{j, c_{k+1}, s_{k+1}, \dots, s_L}^{>k}, \quad (17a)$$

$$\tilde{G}_j^{<k} = \sum_{s_1, \dots, s_{k-1}} X_{s_1, \dots, s_{k-1}, b_k}^{<k} A_{j, s_1, \dots, s_{k-1}, c_k}^{<k}, \quad (17b)$$

where $A^{>k}$ and $A^{<k}$ are defined similarly as $X^{>k}$ and $X^{<k}$ in Eqs. (12a) and (12b). Then we have

$$\begin{aligned} & \sum_{s_l \neq k} P_s(\mathbf{a}) X^{<k} X^{>k} \\ &= \sum_j \frac{n_j}{N} \sum_{b'_k, b'_{k+1}} A_{j, c_k, c_{k+1}}^{s_k} \tilde{G}_{j, b_{k+1}, c_{k+1}}^{>k} \tilde{G}_{j, b_k, c_k}^{<k}. \end{aligned} \quad (18)$$

The complete algorithm to find the optimal $P(\mathbf{a})$ that minimizes Loss($P(\mathbf{a})$) in Eq. (11) is summarized in Algorithm 1. Once $P(\mathbf{a})$ is found, the best MPO $\hat{\rho}$ can be reconstructed by applying the inverse of the IC-POVM locally on each site of $P(\mathbf{a})$, as illustrated in Fig. 1.

III. RESULTS

We demonstrate our algorithm by reconstructing the density matrix corresponding to the ground state of the XXZ chain subjecting to depolarizing noise. The Hamiltonian of the XXZ chain can be written as

$$\hat{H} = \sum_{l=1}^{L-1} J(\hat{\sigma}_l^x \hat{\sigma}_{l+1}^x + \hat{\sigma}_l^y \hat{\sigma}_{l+1}^y + \gamma \hat{\sigma}_l^z \hat{\sigma}_{l+1}^z) + h \sum_{l=1}^L \hat{\sigma}_l^z, \quad (19)$$

where L is the number of the spins, J is the tunneling strength which we fix to 1, h is the magnetization strength, and γ is the interaction strength. We choose $h = 1$ to break the degeneracy of the ground state due to the spin-flip symmetry. The depolarizing noise is described by the CPTP map:

$$\hat{\rho} \rightarrow \mathcal{E}(\hat{\rho}) = \frac{p\hat{I}}{d} + (1-p)\hat{\rho}, \quad (20)$$

with $d = 2^L$ the dimension of the Hilbert space, $\hat{\rho}$ the density matrix corresponding to the exact ground state, and p

Algorithm 1. Non-negative tensor train state tomography.

Input: The set of samples from POVM measurement;
Output: Near-optimal non-negative MPS form of $P(\mathbf{a})$;

- 1: Encode each \mathbf{a}^j into A^j according to Eq. (9);
- 2: Randomly initialize $P(\mathbf{a})$ as in [39];
- 3: $\tilde{G}_{j, b_1, c_1}^{<1} = 1, \tilde{G}_{j, b_{L+1}, c_{L+1}}^{>L} = 1$
- 4: $G_{b_1, b'_1}^{<1} = 1, G_{b_{L+1}, b'_{L+1}}^{>L} = 1$
- 5: **for** $k = 1, 2, \dots, L-1$ **do**
- 6: $G_{b_{k+1}, b'_{k+1}}^{<k+1} = \sum_{s_k, b_k, b'_k} G_{b_k, b'_k}^{<k} X_{b_k, b_{k+1}}^{s_k} X_{b'_k, b'_{k+1}}^{s_k}$;
- 7: **for** $j = 1, 2, \dots, N_s$ **do**
- 8: $\tilde{G}_{j, b_{k+1}, c_{k+1}}^{<k+1} = \sum_{s_k, b_k, c_k} \tilde{G}_{j, b_k, c_k}^{<k} X_{b_k, b_{k+1}}^{s_k} A_{j, c_k, c_{k+1}}^{s_k}$;
- 9: **end for**
- 10: **end for**
- 11: **for** $k = L-1, L-2, \dots, 1$ **do**
- 12: $G_{b_{k+1}, b'_{k+1}}^{>k} = \sum_{s_{k+1}, b_{k+2}, b'_{k+2}} G_{b_{k+2}, b'_{k+2}}^{>k+1} X_{b_{k+1}, b_{k+2}}^{s_{k+1}} X_{b'_{k+1}, b'_{k+2}}^{s_{k+1}}$;
- 13: **for** $j = 1, 2, \dots, N_s$ **do**
- 14: $\tilde{G}_{j, b_{k+1}, c_{k+1}}^{>k} = \sum_{s_{k+1}, b_{k+2}, c_{k+2}} \tilde{G}_{j, b_{k+2}, c_{k+2}}^{>k+1} X_{b_{k+1}, b_{k+2}}^{s_{k+1}} A_{j, c_{k+1}, c_{k+2}}^{s_{k+1}}$;
- 15: **end for**
- 16: **end for**
- 17: **while true do**
- 18: **for** $k = 1, 2, \dots, L-1$ **do**
- 19: update $X_{b_k, b_{k+1}}^{s_k}$ using Eq. (16);
- 20: $G_{b_{k+1}, b'_{k+1}}^{<k+1} = \sum_{s_k, b_k, b'_k} G_{b_k, b'_k}^{<k} X_{b_k, b_{k+1}}^{s_k} X_{b'_k, b'_{k+1}}^{s_k}$;
- 21: **for** $j = 1, 2, \dots, N_s$ **do**
- 22: $\tilde{G}_{j, b_{k+1}, c_{k+1}}^{<k+1} = \sum_{s_k, b_k, c_k} \tilde{G}_{j, b_k, c_k}^{<k} X_{b_k, b_{k+1}}^{s_k} A_{j, c_k, c_{k+1}}^{s_k}$;
- 23: **end for**
- 24: **end for**
- 25: **for** $k = L-1, L-2, \dots, 1$ **do**
- 26: update $X_{b_{k+1}, b_{k+2}}^{s_{k+1}}$ using Eq. (16);
- 27: $G_{b_{k+1}, b'_{k+1}}^{>k} = \sum_{s_{k+1}, b_{k+2}, b'_{k+2}} G_{b_{k+2}, b'_{k+2}}^{>k+1} X_{b_{k+1}, b_{k+2}}^{s_{k+1}} X_{b'_{k+1}, b'_{k+2}}^{s_{k+1}}$;
- 28: **for** $j = 1, 2, \dots, N_s$ **do**
- 29: $\tilde{G}_{j, b_{k+1}, c_{k+1}}^{>k} = \sum_{s_{k+1}, b_{k+2}, c_{k+2}} \tilde{G}_{j, b_{k+2}, c_{k+2}}^{>k+1} X_{b_{k+1}, b_{k+2}}^{s_{k+1}} A_{j, c_{k+1}, c_{k+2}}^{s_{k+1}}$;
- 30: **end for**
- 31: **end for**
- 32: **if stopping criterion is met, then**
- 33: break;
- 34: **end if**
- 35: **end while**
- 36: **return** $P(\mathbf{a})$

the strength of the noise. We note that for $p = 0$, namely, for pure states, there already exists an efficient MPS-based tomography algorithm which directly uses MPS as the ansatz to represent an unknown pure state [13].

Similar to [17], we use both the quantum fidelity and the classical fidelity to measure the learning accuracy. Specifically, the quantum fidelity is defined as

$$\mathcal{F}_q = \text{tr}^2(\sqrt{\sqrt{\hat{\rho}_1} \hat{\rho}_2 \sqrt{\hat{\rho}_1}}), \quad (21)$$

for two density matrices $\hat{\rho}_1$ and $\hat{\rho}_2$, and the classical fidelity is defined as

$$\mathcal{F}_c = \mathbb{E}_{\mathbf{a} \sim P_i}[\sqrt{P(\mathbf{a})/P_i(\mathbf{a})}], \quad (22)$$

where $P(\mathbf{a})$ is the reconstructed probability distribution, and $P_i(\mathbf{a})$ is the ideal probability distribution. We also define the quantum and the classical infidelities as $\mathcal{I}_q = 1 - \mathcal{F}_q$ and $\mathcal{I}_c = 1 - \mathcal{F}_c$, respectively. In our numerical simulations we

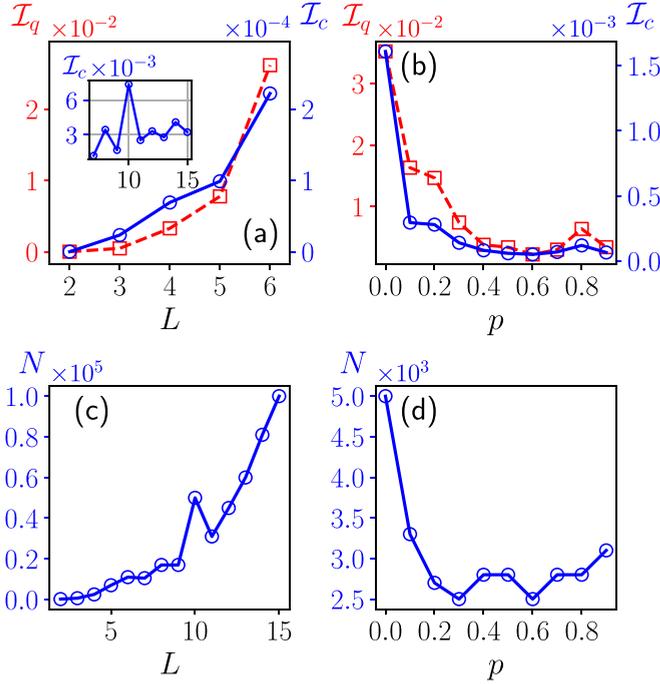


FIG. 2. (a) \mathcal{I}_q (red dashed line with squares) and \mathcal{I}_c (blue solid line with circles) as functions of system size L for fixed $p = 0.6$. The inset shows \mathcal{I}_c as a function of system size for larger system size (\mathcal{I}_q for $L > 6$ is not shown since it is too expensive to compute). (b) \mathcal{I}_q and \mathcal{I}_c as functions of depolarized noise strength p for fixed $L = 4$. (c) The minimum number of required training data N as a function of L such that $\mathcal{I}_c \leq 1\%$, with $p = 0.6$. (d) The minimum number of required training data N as a function of p such that $\mathcal{I}_c \leq 1\%$, with $L = 4$. The other parameters used in these simulations are $\gamma = 2$ and $D = 10$.

have generated two independent synthetic datasets for each parameter setting we have considered, each with 30 000 000 samples. One dataset is used for training (we may only use part of them as the training data) and the other is used for testing. For the quantum fidelity, we directly compute \mathcal{F}_q between the reconstructed density matrix and the target density matrix. For the classical fidelity, we use the test dataset to evaluate Eq. (22).

We first study the final reconstruction quality as a function of the system size L and the noise strength p . We show \mathcal{I}_q and \mathcal{I}_c as functions of L in Fig. 2(a) and of p in Fig. 2(b), respectively. We can see that the final fidelities (both the quantum and the classical) decrease as L increases and increase as p increases. We can also see that it is much easier for a near-perfect reconstruction of the probability distribution than the reconstruction of the underlying quantum state, as in the numerical simulation \mathcal{I}_c is always at least one order of magnitude smaller than \mathcal{I}_q . In Figs. 2(c) and 2(d) we show the minimum number N of required training data for $\mathcal{I}_c \leq 1\%$ as a function of L and p , respectively. Since it is numerically too demanding to compute the quantum fidelity between two quantum matrices with more than 10 qubits, we only consider \mathcal{I}_c in these two panels. We can see that N increases as L increases and decreases as p increases.

Then we fix $L = 6$ and study the convergences of \mathcal{I}_q and \mathcal{I}_c as functions of the number of sweeps and for different values

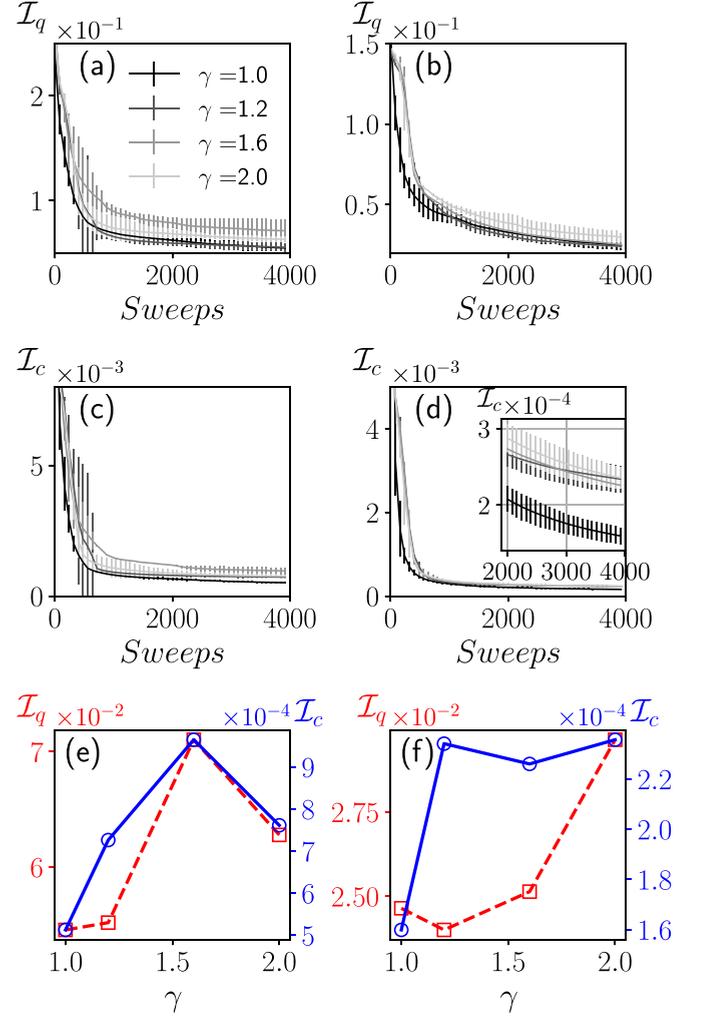


FIG. 3. (a), (c) \mathcal{I}_q (a) and \mathcal{I}_c (c) as functions of the number of sweeps for different γ with $p = 0.4$. (b), (d) \mathcal{I}_q (b) and \mathcal{I}_c (d) as functions of the number of sweeps for different γ with $p = 0.6$. The inset in (d) shows the tail of the convergence for \mathcal{I}_c . In (a), (b), (c), (d) the gray lines from darker to lighter correspond to $\gamma = 1, 1.2, 1.6, 2$, respectively. We have also chosen the five best results according to their loss values out of 100 trials and plotted the mean values of them. (The standard deviations are shown as error bars.) (e) The left and right axis show the final \mathcal{I}_q and \mathcal{I}_c as functions of γ with $p = 0.4$. (f) The left and right axis show the final \mathcal{I}_q and \mathcal{I}_c as functions of γ with $p = 0.6$. Here we have chosen $L = 6$. We have also used $N = 3 \times 10^7$ and $D = 10$ in these simulations.

of γ , which is shown in Fig. 3. In Figs. 3(a) and 3(c) we show \mathcal{I}_q and \mathcal{I}_c as functions of the number of sweeps with $p = 0.4$ for different values of γ , while in Figs. 3(b) and 3(d) we show \mathcal{I}_q and \mathcal{I}_c as functions of the number of sweeps with $p = 0.6$ for different values of γ . We can see that for all the different noise strengths, \mathcal{I}_c converges in about 1000 sweeps and \mathcal{I}_q does not fully converge after 4000 sweeps. The final values of \mathcal{I}_q and \mathcal{I}_c after 4000 sweeps are also shown in Fig. 3(e) for $p = 0.4$ and in Fig. 3(f) for $p = 0.6$. We can see that in both cases \mathcal{I}_c is about two orders of magnitude smaller than the corresponding \mathcal{I}_q and that \mathcal{I}_q and \mathcal{I}_c for $p = 0.6$ is smaller than those for $p = 0.4$.

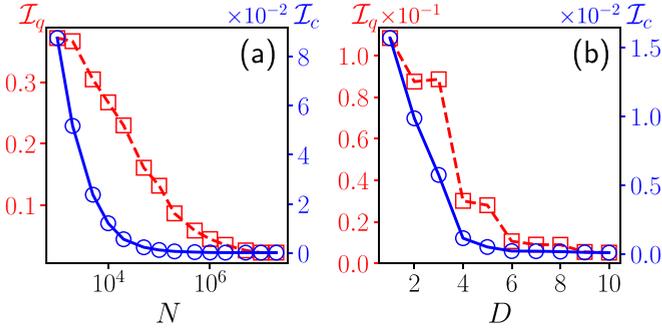


FIG. 4. (a) \mathcal{I}_q (red dashed line with squares) and \mathcal{I}_c (blue solid line with circles) as functions of the number of training data N , with $D = 10$. (b) \mathcal{I}_q (red dashed line with squares) and \mathcal{I}_c (blue solid line with circles) as functions of the bond dimension D , with $N = 3 \times 10^7$. The other parameters used are $L = 6$, $p = 0.6$, and $\gamma = 2$.

In Fig. 3 we have used a large number of training data ($N = 3 \times 10^7$) as well as a large bond dimension $D = 10$ to ensure that we could reach both high quantum and classical fidelities for different values of γ . This may not be necessary in all cases [for example, we can see from Fig. 2(c) that one could already reach classical fidelities higher than 99% for up to 15 qubits with $N \approx 10^5$ only]. In Fig. 4 we plot the dependence of the fidelities on the number of training data and on the bond dimension D used in our reconstruction algorithm, where we have focused on $L = 6$, $p = 0.6$, and $\gamma = 2$. From Fig. 4(a) we can see that with $N = 10^6$, \mathcal{F}_q already reaches 95.5%. We can also see that the classical fidelity converges much faster than the quantum fidelity, for example, with $N < 10^4$ one could already get $\mathcal{F}_c > 90\%$, while for $\mathcal{F}_q > 90\%$ one needs $N > 10^5$. From Fig. 4(b) we can see that both \mathcal{F}_q and \mathcal{F}_c have reached larger than 96% for $D = 4$ and that \mathcal{F}_c also converges faster than \mathcal{F}_q against D .

Due to the variational feature of our algorithm similar to DMRG, it could be trapped in local minima [also because the initial MPS $P(\mathbf{a})$ is randomly initialized] [40]. Therefore, in our numerical results, the same reconstruction algorithm is run for many trials with random initialization of $P(\mathbf{a})$, and the one with the lowest loss value is chosen as the final result. Ideally one hopes to directly choose the trials with the highest fidelity. However, in real experimentation the target state is unknown and it is not possible to compute the fidelities. As a result, it is important that the trials with lower loss values correspond to those with higher fidelities. Such correspondence between loss values and fidelities is shown in Fig. 5, where we have repeated the reconstruction algorithm 100 times. We can see that the loss value indeed has the desired correspondence with the fidelity.

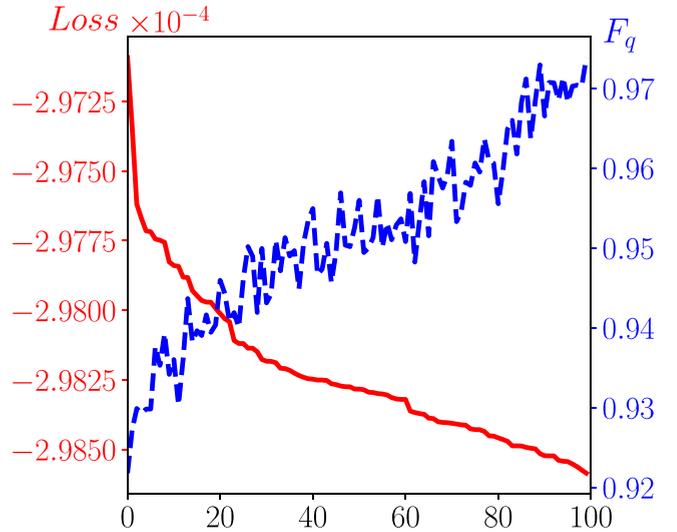


FIG. 5. The x axis denotes different labels of the 100 numerical experiments, labeled 1 to 100, sorted by their final loss values from large to small. In our simulation, we have used $P^2(\mathbf{a}) - 2P(\mathbf{a})P_s(\mathbf{a})$ as the loss value, which simply shifts the original loss value in Eq. (11) by a constant. Here the results are taken from the reconstruction of $P(\mathbf{a})$ for $L = 6$, $\gamma = 2.0$, and $p = 0.6$.

IV. CONCLUSION

We have presented an algorithm based on the non-negative matrix product state for quantum state tomography. Given a number of experimental measurement outcomes, our algorithm iteratively finds the optimal non-negative MPS representation which best approximates the probability distribution of these outcomes. Applying simple local transformations, the reconstructed non-negative MPS can be converted into a density matrix for the unknown quantum state. This is in comparison with the QST methods based on neural network states, for which one generally cannot directly write down the quantum state but only has indirect access to it via sampling from the trained neural networks. As applications, the effectiveness of our algorithm is demonstrated to reconstruct the ground state of the XXZ chain with depolarizing noise.

ACKNOWLEDGMENTS

C.G. acknowledges support from the National Natural Science Foundation of China under Grant No. 11805279. D.H. and X.W. gratefully acknowledge a grant from the National Key R&D Program of China, Grant No. 2018YFA0306703.

- [1] A. G. White, D. F. V. James, P. H. Eberhard, and P. G. Kwiat, *Phys. Rev. Lett.* **83**, 3103 (1999).
- [2] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, *Phys. Rev. A* **64**, 052312 (2001).
- [3] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell *et al.*, *Nature (London)* **574**, 505 (2019).

- [4] Y. Wu, W.-S. Bao, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan *et al.*, *Phys. Rev. Lett.* **127**, 180501 (2021).
- [5] Q. Zhu, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan, M. Gong *et al.*, *Sci. Bull.* **67**, 240 (2022).
- [6] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, *Phys. Rev. Lett.* **105**, 150401 (2010).

- [7] W.-T. Liu, T. Zhang, J.-Y. Liu, P.-X. Chen, and J.-M. Yuan, *Phys. Rev. Lett.* **108**, 170403 (2012).
- [8] A. Smith, C. A. Ríofrío, B. E. Anderson, H. Sosa-Martinez, I. H. Deutsch, and P. S. Jessen, *Phys. Rev. A* **87**, 030102(R) (2013).
- [9] C. A. Ríofrío, D. Gross, S. T. Flammia, T. Monz, D. Nigg, R. Blatt, and J. Eisert, *Nat. Commun.* **8**, 15305 (2017).
- [10] G. Tóth, W. Wieczorek, D. Gross, R. Kischek, C. Schwemmer, and H. Weinfurter, *Phys. Rev. Lett.* **105**, 250403 (2010).
- [11] T. Moroder, P. Hyllus, G. Tóth, C. Schwemmer, A. Niggebaum, S. Gaile, O. Gühne, and H. Weinfurter, *New J. Phys.* **14**, 105001 (2012).
- [12] Y. Liu, D. Wang, S. Xue, A. Huang, X. Fu, X. Qiang, P. Xu, H.-L. Huang, M. Deng, C. Guo, X. Yang, and J. Wu, *Phys. Rev. A* **101**, 052316 (2020).
- [13] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, *Nat. Commun.* **1**, 149 (2010).
- [14] B. P. Lanyon, C. Maier, M. Holzäpfel, T. Baumgratz, C. Hempel, P. Jurcevic, I. Dhand, A. S. Buyskikh, A. J. Daley, M. Cramer, M. B. Plenio, R. Blatt, and C. F. Roos, *Nat. Phys.* **13**, 1158 (2017).
- [15] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, *Nat. Phys.* **14**, 447 (2018).
- [16] G. Torlai and R. G. Melko, *Phys. Rev. Lett.* **120**, 240503 (2018).
- [17] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, *Nat. Mach. Intell.* **1**, 155 (2019).
- [18] S. Ahmed, C. Sánchez Muñoz, F. Nori, and A. F. Kockum, *Phys. Rev. Lett.* **127**, 140502 (2021).
- [19] A. W. R. Smith, J. Gray, and M. S. Kim, *PRX Quantum* **2**, 020348 (2021).
- [20] H.-Y. Huang, R. Kueng, and J. Preskill, *Nat. Phys.* **16**, 1050 (2020).
- [21] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, *PRX Quantum* **2**, 030348 (2021).
- [22] R. Gupta, R. Xia, R. D. Levine, and S. Kais, *PRX Quantum* **2**, 010318 (2021).
- [23] T. Xin, D. Lu, J. Klassen, N. Yu, Z. Ji, J. Chen, X. Ma, G. Long, B. Zeng, and R. Laflamme, *Phys. Rev. Lett.* **118**, 020401 (2017).
- [24] T. Baumgratz, D. Gross, M. Cramer, and M. B. Plenio, *Phys. Rev. Lett.* **111**, 020401 (2013).
- [25] J. Wang, Z.-Y. Han, S.-B. Wang, Z. Li, L.-Z. Mu, H. Fan, and L. Wang, *Phys. Rev. A* **101**, 032321 (2020).
- [26] A. Rocchetto, E. Grant, S. Strelchuk, G. Carleo, and S. Severini, *npj Quantum Inf.* **4**, 28 (2018).
- [27] I. Oseledets and E. Tyrtshnikov, *Linear Algebra Appl.* **432**, 70 (2010).
- [28] G. García-Pérez, M. A. C. Rossi, B. Sokolov, F. Tacchino, P. K. Barkoutsos, G. Mazzola, I. Tavernelli, and S. Maniscalco, *PRX Quantum* **2**, 040342 (2021).
- [29] S. T. Flammia, *J. Phys. A: Math. Gen.* **39**, 13483 (2006).
- [30] H. Sosa-Martinez, N. K. Lysne, C. H. Baldwin, A. Kalev, I. H. Deutsch, and P. S. Jessen, *Phys. Rev. Lett.* **119**, 150401 (2017).
- [31] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2000).
- [32] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, *J. Math. Phys.* **45**, 2171 (2004).
- [33] U. Schollwöck, *Ann. Phys.* **326**, 96 (2011).
- [34] X. Shi, Y. Shang, and C. Guo, *Phys. Rev. A* **105**, 052424 (2022).
- [35] N. Lee, A.-H. Phan, F. Cong, and A. Cichocki, in *Neural Information Processing*, Vol. 9949, edited by A. Hirose, S. Ozawa, K. Doya, K. Ikeda, M. Lee, and D. Liu (Springer International Publishing, Cham, Switzerland, 2016), pp. 87–95.
- [36] E. M. Shcherbakova, *Lobachevskii J. Math.* **40**, 1863 (2019).
- [37] E. Shcherbakova and E. Tyrtshnikov, in *Large-Scale Scientific Computing*, edited by I. Lirkov and S. Margenov (Springer International Publishing, Cham, Switzerland, 2020), pp. 156–164.
- [38] D. D. Lee and H. S. Seung, *Nature (London)* **401**, 788 (1999).
- [39] S. Holtz, T. Rohwedder, and R. Schneider, *SIAM J. Sci. Comput.* **34**, A683 (2012).
- [40] Z. Liu, L.-W. Yu, L.-M. Duan, and D.-L. Deng, *arXiv:2108.08312*.