Scalable quantum tomography with fidelity estimation

Jun Wang,[1] Zhao-Yu Han,[2] Song-Bo Wang,[1] Zeyang Li,[3] Liang-Zhu Mu,[1,*] Heng Fan,[4,5,†] and Lei Wang[4,5,‡]

[1]*School of Physics, Peking University, Beijing 100871, China*

[2]*Department of Physics, Stanford University, Stanford, California 94305, USA*

[3]*Department of Physics, MIT-Harvard Center for Ultracold Atoms and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

[4]*Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China*

[5]*CAS Central for Excellence in Topological Quantum Computation, University of Chinese Academy of Sciences, Beijing 100190, China*

(Received 8 December 2018; accepted 19 February 2020; published 16 March 2020)

We propose a quantum tomography scheme for pure qudit systems which adopts a certain version of random basis measurements and a generative learning method, along with a built-in fidelity estimation approach to assess the reliability of the tomographic states. We prove the validity of the scheme theoretically, and we perform numerically simulated experiments on several target states that have compact matrix product state representation, demonstrating its efficiency and robustness. We find the number of replicas required by a fixed fidelity criterion grows only linearly as the system size scales up, which saturates a lower bound from information theory. Thus the scheme achieves the highest possible scalability that is crucial for practical quantum state tomography.

DOI: 10.1103/PhysRevA.101.032321

I. INTRODUCTION

With the fast developing techniques of fabricating quantum devices, we can now manipulate a growing number of entangled qudits. Medium sized quantum devices (10–100 qubits) have been implemented in the platforms of superconducting circuits, trapped ions, and ultracold atoms [1–7]. Quantum state tomography (QST), which aims at reconstructing an unknown quantum state from suitable measurements on replicas of the state, is a gold standard for verifying and benchmarking the merits of the implementations. In particular, QST is necessary for proving the completeness of information that could be provided by all practical operations and measurements on a quantum processor.

Early studies of QST focused on mixed states and found that it requires the information provided by projective measurements on a minimal set of $O(d)$ mutually unbiased bases [8–10] or by $O(d^2)$ expectations of positive-operator-valued measures (POVMs) [11–14]. This soon becomes impractical as Hilbert space dimension d grows exponentially with the number of constituents (e.g., particles). For pure states, it was recently proved that in terms of information the adequate number of POVMs can be drastically reduced to $O(d)$ [15–17] and that of measurement bases can be reduced to four [18–20]. However, it is still experimentally intractable to realize these delicately designed nonlocal measurements and to acquire corresponding converged probability distributions, since the size of the sample space d is exponentially large [21].

After a long history of developing its mathematical ground, we are now at a stage to consider the pragmatic aspects of QST. Specifically, the cost of preparing multiple copies of the state and measurements is crucial for large-scale many-body states tomography. There have been several efforts towards scalable QST schemes [22–30], mostly by the mean of exploiting the property of short-range entanglement in a matrix product state (MPS) [31]. Some colleagues [32–42] applied compressed sensing [43,44] from the thriving field of machine learning to this classical problem, introducing randomness to the choice of measurements. We note that there are indeed similarities between QST and unsupervised machine learning tasks such as density estimation [45]. In both tasks, one aims at modeling high-dimensional probability functions from observed data. While QST is more complicated because probability distributions under different bases are inherently related and modeled simultaneously.

Besides preparing replicas for reconstructing the measured state, another practical concern that is drawing attention is fidelity estimation [46–48]. It refers to the task of assessing the proximity between two quantum states. When the two states are the laboratory state and the modeled state, it helps quantifying the quality and progress of the tomography scheme; as for the case of the laboratory state and the ideal aimed state, it decides the merit of experimental preparation for quantum devices.

Our work tries to resolve both problems as concentrating on reconstructing pure states in qudit systems through local projective measurements. Viewing quantum states as integrated generative models and QST as an unsupervised learning task, we adapts its key components: measurement and reconstruction. We design a compressed-sensing-inspired approach to acquire data via single-shot measurements on locally random bases, which avoids full statistics on any specific set of physical observables. We adopt MPS and the associated optimization algorithm proposed by our previous work [49], which is a variation of density matrix

*muliangzhu@pku.edu.cn
†hfan@iphy.ac.cn
‡wanglei@iphy.ac.cn

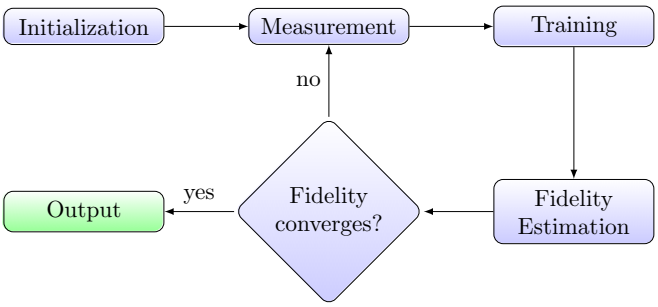


FIG. 1. A flow chart of the QST scheme.

renormalization group (DMRG) [50], as the model and learning approach, respectively. We devise a cost function based on the averaged negative log likelihood with an entanglement entropy penalty, thus our method could also be viewed as a standard maximal likelihood estimation (MLE) approach. All these efforts help the scheme achieve highest possible scalability by saturating an information bound in our computer-simulated experiments for target states that have compact MPS description. Moreover, our scheme allows a fidelity estimation approach that requires neither overhead in measurements nor *a priori* knowledge in the target state.

In this paper we will first present the workflow of the scheme and argue its validity in Sec. II, then demonstrate its efficiency and robustness in Sec. III by computer-simulated experiments.

II. PROCEDURE

The proposed QST scheme includes iterative projective measurements on the target state and training of the model, until the stop criterion concerning the fidelity estimation is met, as shown in the flow chart in Fig. 1. The target and the tomographic states are denoted by their density operators σ, $\tilde{\rho}$, respectively. We emphasize that the theoretical arguments in this section are not confined to pure states, although the following experiments are; the general validity of the scheme does not depend on the property of the target state, though the efficiency may.

A. Measurement

The bases on which one performs projective measurements are required to comprise an informationally complete set of bases in order to avoid candidate states that are indistinguishable based on the measurement outcomes. In practice, one should also consider the complexity of implementing these measurements. Local measurements are often preferred because of the simplicity of single qudit gates. We note that arbitrary local spin measurements could be expediently realized, e.g., in cold atoms experimental platforms [51]. Here we use the spin-S picture to demonstrate our sets of bases on a qudit system, where the qudits have dimension $q = 2S + 1$ and S is a half-integer. The basis denoted by $\mathcal{B}(\{\boldsymbol{n}_1, \ldots, \boldsymbol{n}_N\})$, sometimes $\mathcal{B}(\{\boldsymbol{n}\})$ for short, refers to the eigenbasis corresponding to the product of local spin operators $\boldsymbol{n}_1 \cdot \boldsymbol{s}_1 \otimes \boldsymbol{n}_2 \cdot \boldsymbol{s}_2 \otimes \cdots \otimes \boldsymbol{n}_N \cdot \boldsymbol{s}_N$, where $\boldsymbol{n}_i$ is the unit vector indicating the spin measurement direction on site i, and $\boldsymbol{s}_i = (s_i^x, s_i^y, s_i^z)$ are the single site spin operators.

We propose to perform measurements on *random bases*. This is an analogy of compressed sensing, which recovers a sparse signal by utilizing observations from random perspectives. Each time in demand of a measuring outcome accumulation, one can randomly sample a measuring basis $\mathcal{B}(\{\boldsymbol{n}\})$ from a certain probability density f on $S_2^{\otimes N}$.

In a single projective measurement on basis $\mathcal{B}(\{\boldsymbol{n}\})$, one obtains an outcome $(m_1, m_2, \ldots, m_N)$ and a corresponding product state $|\{\boldsymbol{n}, m\}\rangle \equiv \otimes_{j=1}^N |\boldsymbol{n}_j, m_j\rangle$. Hence with a certain basis sampling distribution, a generic state ρ is naturally a generative model whose sample space $\Omega \equiv \{\otimes_{j=1}^N |\boldsymbol{n}_j, m_j\rangle | \boldsymbol{n}_j \in S_2, m_j \in \{-S, -S+1, \ldots, S\}\}$ comprises all this kind of product states. Note that $|-\boldsymbol{n}_j, -m_j\rangle = |\boldsymbol{n}_j, m_j\rangle$, the probability density is

$$\mathbb{P}[\rho](|\{\boldsymbol{n}, m\}\rangle) = \sum_{c_j=\pm 1} f\left(\{c_j \boldsymbol{n}_j\}_{j=1}^N\right) \langle\{\boldsymbol{n}, m\}|\rho|\{\boldsymbol{n}, m\}\rangle. \quad (1)$$

From this perspective, our scheme is a generative modeling process based on only the outcome product states from the target σ. It is proved in Appendix A that information in the probability density is complete, for the Kullback-Leibler (KL) divergence $D_{\mathrm{KL}}(\mathbb{P}[\sigma]||\mathbb{P}[\tilde{\rho}])$ bounds the matrix norm $||\tilde{\rho} - \sigma||$ as long as f is not ill-shaped, which means $D_{\mathrm{KL}}(\mathbb{P}[\sigma]||\mathbb{P}[\tilde{\rho}])$ specifies the target as its unique minimum.

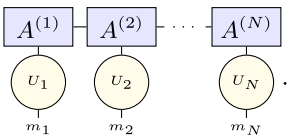
Though some schemes have mentioned random bases measurements [52] for QST, ours is radically different in terms of whether a statistical quantity under a certain basis is necessary. Since our scheme does not deliberately repeat measurements on any single basis, it is possible to avoid the unbearable efforts introduced by complete statistics over an exponentially large number of possible outcomes.

After accumulating a batch of measurement outcomes, we append them into data set $\mathcal{V}$, whose size $|\mathcal{V}|$ is also the number of replicas of the target, and subsequently train the model with the updated $\mathcal{V}$.

B. Training

The model for pure states in our scheme is the wave functions under $\mathcal{B}(\{z_1, z_2, \ldots, z_N\})$ represented with an MPS. Under another basis $\mathcal{B}(\{\boldsymbol{n}\})$ the wave function is straightforwardly obtained by performing local unitary transformations, schematically:

$$\tilde{\Psi}(\{m\}; \{\boldsymbol{n}\}) = \quad (2)$$



where each box denotes a tensor in the MPS, and the circle U_k denotes the single qudit unitary transformation $U(\boldsymbol{n}_k)$ that rotates the direction $\boldsymbol{n}_k$ to $\boldsymbol{e}_z$. For introduction to the graphical notations of tensor networks, we refer to [53,54]. For $\boldsymbol{n}$ in (θ, ϕ) direction in a conventional spherical coordinate, $U(\boldsymbol{n})$ could be realized as $\exp(i\theta s^y/\hbar)\exp(i\phi s^z/\hbar)$.

When the model $\tilde{\rho}$ minimizes the KL divergence $D_{\mathrm{KL}}(\mathbb{P}[\sigma]||\mathbb{P}[\tilde{\rho}])$, it is exactly the target state σ, yet σ itself is what we are reconstructing and currently inaccessible. According to Eq. (1), with both σ, f being fixed, it is equivalent

for the model to minimize

$$\mathcal{L}_N \equiv -\int_\Omega d\mu\, \mathbb{P}[\sigma](|\{\boldsymbol{n}, m\}\rangle) \ln\langle\{\boldsymbol{n}, m\}|\tilde{\rho}|\{\boldsymbol{n}, m\}\rangle, \quad (3)$$

where μ is the natural measure on the sample space Ω. Thus with a finite $\mathcal{V}$, $\mathcal{L}_N$ is estimated as

$$\mathcal{L}_N = \frac{-1}{|\mathcal{V}|}\sum_{|\{\boldsymbol{n},m\}\rangle\in\mathcal{V}} \ln|\tilde{\Psi}(\{m\};\{\boldsymbol{n}\})|^2, \quad (4)$$

whose minimum approaches the target σ as outcomes accumulate in $\mathcal{V}$, so we employ it as the cost function for training the MPS.

$\mathcal{L}_N$ is indeed highly similar with the negative log likelihood (NLL) especially when f is uniform, yet generally $|\tilde{\Psi}(\{m\},\{\boldsymbol{n}\})|^2$ is not the probability density. Moreover, one can optionally add penalty terms in the cost function $\mathcal{L} = \mathcal{L}_N + \lambda\mathcal{P}$ as long as he adiabatically tunes $\lambda \to 0$. For instance, when training the merged kth and $(k+1)$th tensors, take $\mathcal{P}$ as the second order Rényi entropy $-\ln \mathrm{Tr}\,\rho_{\mathrm{R},k}^2$, where $\rho_{\mathrm{R},k}$ is the reduced density matrix. It renders the MPS preference in low entanglement description of the data in training, and thus make the algorithm converge faster. Encoding information from different bases in the cost function, we may not worry that the MPS overfits the probability distribution on certain bases.

To minimize the cost function, we improve the unsupervised learning approach proposed in our previous work [49]. The algorithm is similar with but not exactly 2-site DMRG [50] method and adjusts the parameter allocation as well as the parameters of the MPS to the best description of the measured data. DMRG is known as one of the most successful numerical methods in studying the ground state property of one-dimensional quantum systems [50,55]. The relation between our approach and DMRG has been elaborated in Ref. [49].

The major difference from the previous work is the gradient calculated for tuning, which results from the complex-value form of quantum wave functions and the application of unitary transformations. Details about calculating the gradients of $\mathcal{L}$ are in the Supplemental Material [56].

C. Fidelity estimation

Estimating the proximity between the tomographic state and the target enables one to determine the current quality of tomography process. Quantum fidelity $\mathcal{F} \equiv \sqrt{\mathrm{Tr}[\tilde{\rho}\sigma]}$ and the distance $\mathcal{R} \equiv ||\tilde{\rho} - \sigma||/2$ are constantly used for this quantification, but we cannot use these definitions for fidelity estimation, since we only have access to the measurement outcomes and the training history instead of the target σ itself.

We assume that similar states' tomography have similar converging behavior when all the scheme parameters are fixed. Thus, when our tomographic state is trained to be a sufficiently good approximation, we use it as a virtual target state $\sigma_{\mathrm{vir}} = |\tilde{\Psi}\rangle\langle\tilde{\Psi}|$ to simulate the tomography scheme in a computer: generate measuring outcomes from σ_{vir} and train a new model $\tilde{\rho}_{\mathrm{vir}}$ in the same way. Since we have access to σ_{vir}, this process could be completely monitored and be used to approximate the real process at the same measurements accumulation stage.

For two pure states, $\mathcal{F} = |\langle\psi|\phi\rangle|$ and $\mathcal{R} = \sqrt{\frac{1}{2}(1-\mathcal{F}^2)}$ quantify their proximity. In our experimental tests, it is observed that the distance between the tomographic MPS and the target state $\mathcal{R}_{\mathrm{real}}$ is asymptotically proportional to $|\mathcal{V}|^{-1/2}$, and the distance between two successive tomographic states $\mathcal{R}_{\mathrm{succ}}$ is asymptotically proportional to $|\mathcal{V}|^{-1}$, at large $|\mathcal{V}|$. This is required by the asymptotic normality of unbiased MLE (see Appendix B). This indicates that $\frac{\mathcal{R}_{\mathrm{real}}^2}{\mathcal{R}_{\mathrm{succ}}}$ gradually approaches to a constant C as measurements accumulate. By our assumption that similar states' tomography have close converging processes, we can extract a good approximation of this unknown constant through $\mathcal{R}'_{\mathrm{real}}$ and $\mathcal{R}'_{\mathrm{succ}}$ monitored in the virtual tomography process, so that we can use $\sqrt{C\mathcal{R}_{\mathrm{succ}}}$ to estimate the actual distance $\mathcal{R}_{\mathrm{real}}$ and therefore the fidelity between $\tilde{\rho}$ and the target σ. Moreover, one can extrapolate $|\mathcal{V}|$ that is adequate for the fidelity criterion from the asymptote so that the batch size of measurement outcomes could be properly increased to save training efforts. We notice that similar asymptotic behavior of $\mathcal{R}_{\mathrm{real}}$ has been generally observed in previous QST methods [40,57,58], thus the procedure here may also be adapted for them.

To meet the desired accuracy, we may have to iterate the measurement (Sec. II A) and the training (Sec. II B) steps until the convergence criterion of real fidelity is satisfied. When successively trained MPS' as virtual target states give converging estimate of C, the confidence in the fidelity estimation also increases.

When the scheme stops, one can combine the final estimation of the distance $||\sigma - \tilde{\rho}||$ and the distance from σ to the aimed ideal state σ_{aimed}, to bound the distance $||\sigma - \sigma_{\mathrm{aimed}}|| \leqslant ||\sigma - \tilde{\rho}|| + ||\sigma_{\mathrm{aimed}} - \tilde{\rho}||$. The bound can be directly calculated since both quantities are known.

III. EXPERIMENTAL TESTS

In this section we test our QST scheme on N-qudit systems in computer-simulated experiments. We represent the target states with compact MPS [59] and employ the direct sampling approach [49] to generate independent samples as measurement outcomes.

A. Efficiency

Four typical states with compact MPS expression are invoked: the W, dimer, cluster, and the AKLT ground [53,60] states. W states are decorated with local phases $|W\rangle = \sum_{k=1}^{N} e^{ik\theta}|0_1 0_2 \cdots 1_k \cdots 0_N\rangle$, $\theta = 0.1$ rad. The AKLT ground states are qutrit systems while the others are qubit systems.

As shown in Fig. 2, the number of replicas $|\mathcal{V}|$ sufficing the criterion of $\mathcal{F} = 0.995$ linearly scales with N, and the test on randomly initiated states show the same linear dependence on N in Fig. 3. To better illustrate the scalability, we consider fixed "per-site fidelity" $\mathcal{F}_{\mathrm{ps}} \equiv \mathcal{F}^{1/N}$, which could be viewed as a normalized criterion for systems with different sizes. In this case, only a nearly constant number of replicas are necessary, which is a general corollary of the linear dependence combined with the asymptotic behavior of $\mathcal{R}_{\mathrm{real}}$.

To understand this remarkable relation, we first notice that we need $O(NqD_{\mathrm{max}}^2)$ parameters to describe a MPS, while a

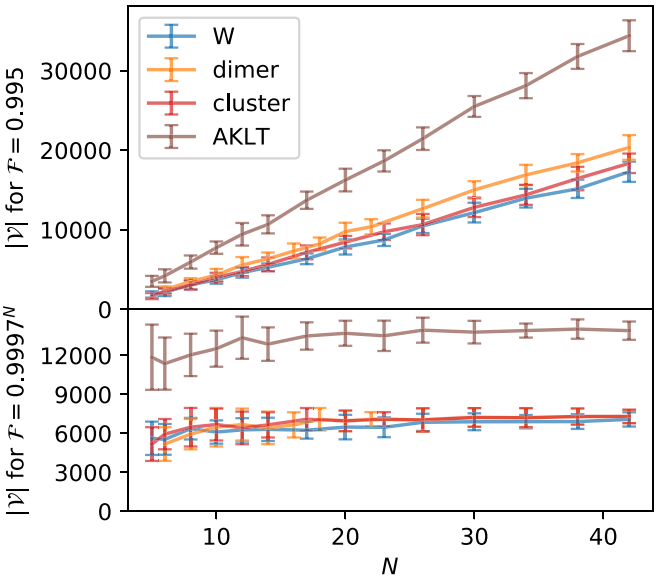


FIG. 2. Efficiency of the scheme tested on the typical states. The number of single-shot outcomes for a fidelity criterion is recorded when the real fidelity get stably higher than that level. The error bars corresponds to the standard deviation of 24 cases with different random seeds.

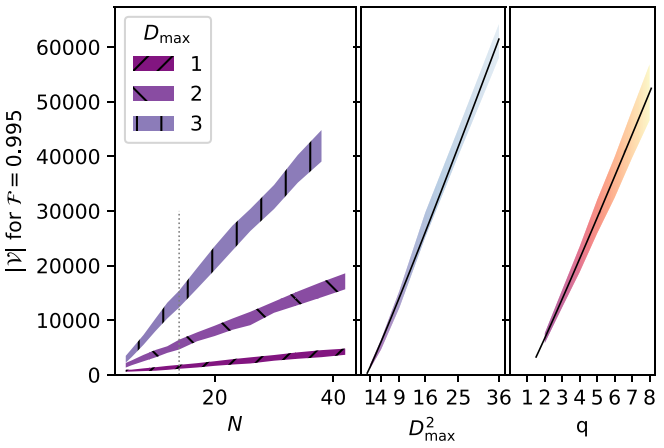


FIG. 3. Efficiency of the scheme tested on randomly initiated states. The standard deviation of the number of replicas $|\mathcal{V}|$ over 72 different random targets are covered in the shades. (Left) Linear dependence of $|\mathcal{V}|$ on size N of $q = 2$ states. (Middle) Quadratic dependence of $|\mathcal{V}|$ on the maximal bond dimension $D_{\max}$ of $N = 14$, $q = 2$ states, and the solid line is the fitting $|\mathcal{V}| = \gamma D^{\beta}_{\max}$ with the resulting parameters being $\beta = 2.1(1)$, $\ln\gamma = 7.2(1)$. (Right) Linear dependence of $|\mathcal{V}|$ on the local Hilbert space dimension $q = 2S + 1$ of $N = 14$, $D = 2$ states, and the solid line is the fitting $|\mathcal{V}| = \lambda S^{\mu}$ with the resulting parameters being $\mu = 1.05(7)$, $\ln\lambda = 9.54(4)$.

single shot projective measurement could provide a sample embodying information that is proportional to N. Hence a constant demand of measurements is necessary to identify a MPS within a normalized error. Then since MPS with finite bond dimensions imposes *sparse* constrains in their Schmidt space, the compressed-sensing-inspired approach could optimize the utility of the information from the measurements on random bases, like the original compressed sensing algorithm in its various applications. To make more concrete interpretation, we prove in Appendix B that the minimal required number of repeated measurement in a tomography process targeting a compact MPS at a fixed fidelity criterion would be lower bounded by $O(N)$. The proof is valid for all maximum likelihood QST schemes with MPS as a model. Our result saturating this bound demonstrates the scheme's ability to optimally exploit the information from each measurement.

To further confirm our argument, we perform experiments on a randomly generated target state with maximal bond dimension $D_{\max}$. As it is shown in Fig. 3, for fixed $\mathcal{F} = 0.995$, we observe quadratic and linear scaling of $|\mathcal{V}|$ when respectively varying $D_{\max}$ and N, and when q is varied, the scaling is linear, which are consistent with our expectations. We thus confirm both the general validity and scalable complexity of our scheme.

The remarkably high efficiency can be compared with previous other schemes proposed for MPS such as the one in Ref. [24], which applied the methods in Refs. [22,23]. For a constant estimation fidelity, they report cubic scaling in N for the total number of measurements, while our scheme scales only linearly in N, thus achieves higher scalability.

B. Fidelity estimation

The asymptotic behavior of $\mathcal{R}_{\text{real}} \propto |\mathcal{V}|^{-1/2}$ and $\mathcal{R}_{\text{succ}} \propto |\mathcal{V}|^{-1}$ has been generally observed in previous tomography schemes [10] and our simulations, which can be explained by the asymptotic normality of MLE methods (see Appendix B). Fitting the asymptotic history into $\mathcal{R} = C|\mathcal{V}|^{\alpha}$ for all cases in the experiments mentioned above, we obtain $\alpha_{\text{real}} = -0.52(7)$, $\alpha_{\text{succ}} = -1.03(4)$, and no dependence on N or $D_{\max}$ has been observed. These results solidly support our claim about the asymptotic behavior.

Our fidelity estimation approach is tested on those typical states and random states mentioned above. We demonstrate the result on a W state of size $N = 30$ in Fig. 4, and more results are available in the Supplemental Material [56]. Aside from confirming the asymptotic behaviors and that $\mathcal{R}^2_{\text{real}}/\mathcal{R}_{\text{succ}}$ approaches to a constant, we note that when $\tilde{\rho}$ is in the vicinity of the target σ and plays the role of a virtual target state, it provides a good estimate of the constant C that the actual target state's $\mathcal{R}^2_{\text{real}}/\mathcal{R}_{\text{succ}}$ converges to. Therefore, with C properly estimated through the virtual process, we are able to assess the real fidelity of our tomographic state to the physical target state.

C. Robustness

Concerning more pragmatic aspects, errors and noises in the implementation of transformations and states are inevitable, and QST is sometimes applied to evaluate the quality of implementation. As it has been shown in the experiments on the typical states and random states, as long as the measurement outcomes obey the probability density, our scheme is acute in reconstruction of the pure target state. Noises in the outcomes can be generated in two stages: in preparation of the replicas and in local transformations for the measurements, and they should be considered as well. Although with the MPS model our experiments are confined to pure target states, we can nevertheless apply the scheme to cases where

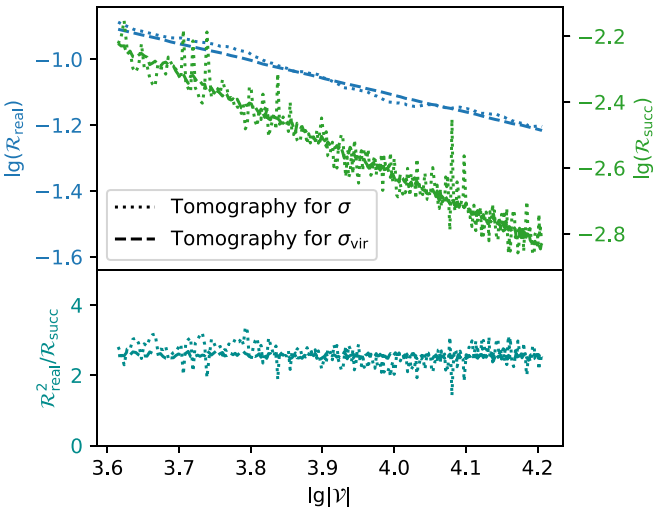


FIG. 4. Fidelity estimation in the tomography for a $N = 30$ W state. A MPS is trained up to $|\mathcal{V}| = 10\,000$ outcomes per our scheme, and $\mathcal{R}_{real}$, $\mathcal{R}_{succ}$, and $\mathcal{R}^2_{real}/\mathcal{R}_{succ}$ in this process are plotted in dotted lines. Then we set it as the virtual target state and simulate our scheme in 24 different random cases. We plot the averaged virtual process in dashed lines.

measurement outcomes are noisy. We mix the target with a portion of quantum noise through the depolarizing channel, namely the measurement is equivalently simulated on

$$\sigma_\epsilon = (1-\epsilon)\sigma + \frac{\epsilon}{q^N} I, \tag{5}$$

where σ is the density operator of the pure target state and I is the identity.

Since the model is a representation of pure state, we could only expect it to approach the pure state with the highest fidelity with respect to σ_ϵ, which is clearly σ. In Fig. 5 we demonstrate the robustness of our scheme over the noises. For real fidelity $\mathcal{F} = 0.995$ between $\tilde{\rho}$ and σ, we obtained the

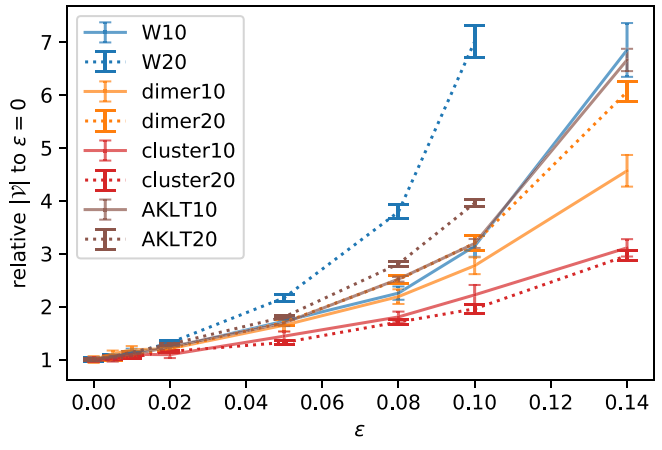


FIG. 5. Robustness over noises in measurement outcomes. ϵ is the noise level, which is simulated as the probability of generating a uniformly random strings instead of obeying the probability distribution of the state in the measurement under a certain basis. The number of replicas necessary for $\mathcal{F} = \sqrt{\mathrm{Tr}[\sigma\tilde{\rho}]} \geqslant 0.995$ are recorded, and their ratio to the $\epsilon = 0$ case are displayed. The error bars indicate the standard deviation over 24 random seeds.

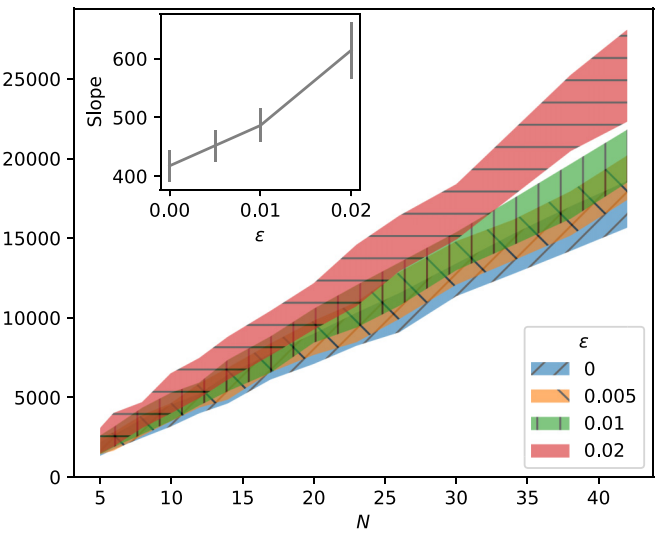


FIG. 6. The robustness of scalability. The number of replicas necessary for the certain criterion $\mathcal{F} = 0.995$ scales linearly with the system size even the outcomes have a certain form of noise. The standard deviation of 72 different random targets are covered in the shades. The inset shows how the slope varies with the noise level.

number of replicas relative to the noiseless case. Within $\epsilon < 0.05$, $|\mathcal{V}|$ sufficing the fidelity criterion does not significantly rise, yet some states are more sensitive than the others. As shown in Fig. 6, the linear scaling with system size is also observed in the noisy cases, which implies that our scheme is scalable even with noise in outcomes, although the slope naturally increases with the noise. The asymptotic behaviors are also noticed not to be significantly affected by noises.

The noises come in two ways as mentioned. Those in preparation of the replicas make the laboratory state mixed, while those in local transformations for measurements are not related to the state itself. These experiments demonstrates the robustness over noises from both ways. Although it leaves out information about how pure the target is in the cases where the laboratory state mixed, applying pure state quantum tomography to physical potentially mixed states is widely accepted in experiments [5,18,24].

IV. CONCLUSION AND OUTLOOK

As the size of implemented quantum devices have increased to medium size, the conventional quantum state tomography approaches are on the verge of intractability. We propose a scalable QST scheme for a qudit system from a generative model perspective, which employs projective measurements under random bases, a learning algorithm with MPS model, and a built-in approach to estimate the fidelity of the tomographic state through the process history. The scheme is featured by: the bases setting in our scheme enjoys experimental accessibility of local transformations and measurements; the algorithm enables adaptive parameter allocation in the model; and the fidelity estimation has no measurement overhead. Most importantly, the scheme does not require statistics on any set of observable quantities, thus it achieves high scalability in simulated experiments on the states with

compact MPS representations: the number of replicas for prescribed fidelity increases linearly and quadratically with the increase in system size and the maximal entanglement spectra size of the target states, respectively. In terms of the required number of replicas, our remarkable scaling result not only stands out among existing schemes, but also saturates the information bound that necessitates $O(N)$ measurements. We justify the efficacy of the fidelity estimation method and also demonstrate the robustness of the scheme in a numerical simulation.

We note that our arguments validating the random measuring bases and the optimization function are applicable to any target quantum states, even for mixed states. We expect that they could pave paths to more practical tomography schemes, by reasonably exploiting other suitable models like multiscale entanglement renormalization ansatz [61,62] and neural networks [63,64], targeting other states with a certain kind of sparse constraint. Moreover, in a sense that the success of QST marks the full characterization of a quantum device, our work suggests that the more operations we could feasibly choose from, the higher efficiency of information contraction from a quantum resource we would achieve.

The data of the simulated experiments mentioned above is available here [65].

ACKNOWLEDGMENTS

We thank X.-Z. Luo, P. Zhang, and especially J. Ouyang for helpful discussions. This work was supported by Ministry of Science and Technology of China (Grants No. 2016YFA0302104 and No. 2016YFA0300600) and National Natural Science Foundation of China (Grants No. 11934018, No. 11774406 and No., 11774398). J.W. is supported by National Training Program of Innovation for Undergraduates. L.W. and H.F. are supported by Chinese Academy of Sciences (XDB28000000).

J.W., Z.-Y.H., and L.W. developed the procedure. S.-B.W. proved the validity for the qubit case, and J.W. generalized it to qudit. Z.-Y.H. proved the information bound. J.W., Z.-Y.H., and Z.L. conducted the efficiency experiments. J.W. conducted the virtual tomography and the robustness experiments. L.W., H.F., and L.-Z.M. supervised this work. All the authors were involved in drafting the manuscript.

J.W. and Z.-Y.H. contributed equally to this work.

APPENDIX A: PROOF OF THE VALIDITY OF RANDOM BASES

We shall prove that the KL divergence

$$D_{\mathrm{KL}}(\mathbb{P}[\sigma]||\mathbb{P}[\tilde{\rho}]) = \int_{\Omega} d\mu \mathbb{P}[\sigma](\{\boldsymbol{n}, m\}) \ln \frac{\mathbb{P}[\sigma](\{\boldsymbol{n}, m\})}{\mathbb{P}[\tilde{\rho}](\{\boldsymbol{n}, m\})} \tag{A1}$$

bounds the Schatten norm $||\sigma - \tilde{\rho}||$.

There are two physical implications. First, because it distinguishes arbitrary two different mixed states, the information it is based on is complete for QST; Second, if we know the target state distribution $\mathbb{P}[\sigma]$, then the KL divergence is valid for quantifying the proximity between $\tilde{\rho}$ and σ, and so is the NLL because it only differs from the KL divergence by a constant depending on σ and f. The distribution $\mathbb{P}[\sigma]$ can be attained by an infinite number of measurements, by the law of large number, and the speed of convergence of the distribution is governed by the central limit theorem.

Given two arbitrary density functions $\mathbb{P}, \mathbb{Q}$, the total variation distance between their probability measures P, Q, defined as $\delta(P, Q) = \sup\{|P(A) - Q(A)||A \text{ is a measurable event}\}$, is bounded by the KL divergence according to Pinsker's inequality [66] $\sqrt{2}\delta(P, Q) \leqslant \sqrt{D_{\mathrm{KL}}(\mathbb{P}||\mathbb{Q})}$. Taking the event $A = \{x \in \text{sample space}|\mathbb{P}(x) \geqslant \mathbb{Q}(x)\}$, one further bounds the L^1 distance $\int d\mu\, |\mathbb{P} - \mathbb{Q}| \leqslant 2\delta(P, Q)$. In our case,

$$\begin{aligned}\sqrt{2D_{\mathrm{KL}}(\mathbb{P}[\sigma]||\mathbb{P}[\tilde{\rho}])} &\geqslant \int d\mu\, |\mathbb{P}[\sigma] - \mathbb{P}[\tilde{\rho}]| \\ &= \int d\mu\, F(\{\boldsymbol{n}\})|\langle\{\boldsymbol{n}, m\}|\sigma|\{\boldsymbol{n}, m\}\rangle \\ &\quad - \langle\{\boldsymbol{n}, m\}|\tilde{\rho}|\{\boldsymbol{n}, m\}\rangle|, \end{aligned} \tag{A2}$$

where we denote $F(\{\boldsymbol{n}\}) = \sum_{c_j=\pm} f(\{c_j \boldsymbol{n}_j\}_{j=1}^N)$.

As $\langle\{\boldsymbol{n}, m\}|\rho|\{\boldsymbol{n}, m\}\rangle \geqslant 0$, $\sum_{\{m\}}\langle\{\boldsymbol{n}, m\}|\rho|\{\boldsymbol{n}, m\}\rangle = 1$, we have $\langle\{\boldsymbol{n}, m\}|\rho|\{\boldsymbol{n}, m\}\rangle \in [0, 1]$, thus the L^1 distance bounds something similar to the L^2 distance

$$\begin{aligned}\int d\mu\, |\mathbb{P}[\sigma] - \mathbb{P}[\tilde{\rho}]| \times 2 \geqslant \int d\mu\, F(\{\boldsymbol{n}\})(\langle\{\boldsymbol{n}, m\}|\sigma|\{\boldsymbol{n}, m\}\rangle \\ - \langle\{\boldsymbol{n}, m\}|\tilde{\rho}|\{\boldsymbol{n}, m\}\rangle)^2. \end{aligned} \tag{A3}$$

In order to prove that it further bounds the usual matrix distances, we introduce an operation for the linear space of q^N by q^N Hermitian matrices ($q = 2S + 1$):

$$(\alpha, \beta) \equiv \int_{\Omega} d\mu\, F(\{\boldsymbol{n}\})\langle\{\boldsymbol{n}, m\}|\alpha|\{\boldsymbol{n}, m\}\rangle\langle\{\boldsymbol{n}, m\}|\beta|\{\boldsymbol{n}, m\}\rangle, \tag{A4}$$

with which $\sqrt{(\alpha, \alpha)}$ directly relates to the right-hand side of Eq. (A3). Its bi-linearity being obvious, in order to justify it as an inner product we prove its positivity as follow:

As long as $F(\{\boldsymbol{n}\}) \geqslant 0$ is not ill-shaped, $0 = (\alpha, \alpha)$ requires the integrand to be 0 almost everywhere in $\Omega = (S_2 \times \{S, S-1, \ldots, -S\})^{\otimes N}$, i.e.,

$$\begin{aligned}&\langle\{\boldsymbol{n}, m\}|\alpha|\{\boldsymbol{n}, m\}\rangle = 0, \quad \forall m \in \{S, S-1, \ldots, -S\}, \\ &\text{a.e.} \boldsymbol{n}_j \in S_2. \end{aligned} \tag{A5}$$

In order to show the conditions in Eq. (A5) restrict α to be 0, let us look into $N = 1$ case first.

1. $N = 1$ case

Denoting $|m\rangle = |z, m\rangle$ for short, with $\boldsymbol{n}$ in (θ, ϕ) direction ($\theta \in [0, \pi]$, $\phi \in [0, 2\pi)$) we have $|\boldsymbol{n}, m\rangle = e^{-i\phi s^z} e^{-i\theta s^y}|m\rangle = \sum_{m'=-S}^{S} e^{-im'\phi}|m'\rangle d^{(S)}_{m'm}(\theta)$, and according to Wigner's formula

$$\begin{aligned} d^{(S)}_{m'm}(\theta) \equiv \langle m'|e^{-i\theta s^y}|m\rangle &= \sum_{k\in\mathbb{Z}} (-1)^{k+m'-m} W(S, m', m, k) \\ &\times \left(\cos\frac{\theta}{2}\right)^{2S-(2k+m'-m)} \left(\sin\frac{\theta}{2}\right)^{2k+m'-m}, \end{aligned} \tag{A6}$$

$$\begin{aligned} &W(S, m', m, k) \\ &\equiv \frac{\sqrt{(S+m')!(S-m')!(S+m)!(S-m)!}}{(S+m-k)!k!(S-k-m')!(k-m+m')!}, \end{aligned} \tag{A7}$$

where the summation over k takes whatever makes the arguments in the factorials in the denominator are non-negative. Thus $0 = \langle \boldsymbol{n}, m|\alpha|\boldsymbol{n}, m\rangle$ gives

$$0 = \sum_{m_1,m_2} e^{i(m_2-m_1)\phi} d_{mm_2}(-\theta)\alpha_{m_2m_1} d_{m_1m}(\theta), \text{ where } \alpha_{m'm} \equiv \langle m'|\alpha|m\rangle. \tag{A8}$$

Since $\{e^{im\phi}|m \in \mathbb{Z}\}$ is a linear independent set of $\phi \in [0, 2\pi)$ functions, in Eq. (A8) the coefficient of $e^{i\mu\phi}$ is zero

$$0 = \sum_{m_1} d_{m,m_1+\mu}(-\theta)\alpha_{m_1+\mu,m_1} d_{m_1,m}(\theta), \tag{A9}$$

where m_1 runs in the range that $m_1, m_1 + \mu \in \{2S, 2S-1, \ldots, -2S\}$ in the sum.

Plugging Eq. (A6) into Eq. (A9), $\forall \mu \in \{2S, 2S-1, \ldots, -2S\}$, $\forall m \in \{S, S-1, \ldots, -S\}$,

$$0 = \sum_{m_1} \alpha_{m_1+\mu,m_1} \sum_{k_1,k_2} (-1)^{k_1+k_2+m-m_1} W(S, m, m_1+\mu, k_1) W(S, m_1, m, k_2) \left(\cos\frac{\theta}{2}\right)^{4S-2(k_1+k_2)+\mu} \left(\sin\frac{\theta}{2}\right)^{2(k_1+k_2)-\mu}, \tag{A10}$$

k_1, k_2 considered in the sum are confined so that in the denominators of

$$W(S, m, m_1+\mu, k_1) = \frac{\sqrt{(S+m)!(S-m)!(S+m_1+\mu)!(S-m_1-\mu)!}}{(S+m_1+\mu-k_1)!k_1!(S-k_1-m)!(k_1+m-m_1-\mu)!}, \tag{A11}$$

$$W(S, m_1, m, k_2) = \frac{\sqrt{(S+m_1)!(S-m_1)!(S+m)!(S-m)!}}{(S+m-k_2)!k_2!(S-k_2-m_1)!(k_2+m_1-m)!} \tag{A12}$$

there are no negative arguments of the factorials. Thus $2(k_1+k_2)-\mu \geqslant 0$ and $4S-2(k_1+k_2)+\mu \geqslant 0$. Meanwhile, note that $\{\cos^{N-n} x \sin^n x|n = 0, 1, \ldots, N\}$ is a linear independent set of $x \in [0, \pi/2]$ functions. Hence $\forall \mu \in \{2S, 2S-1, \ldots, -2S\}$ and $\forall k = k_1 + k_2 \in \{0, 1, \ldots, 2S\} \cap \{\mu, \mu+1, \ldots, \mu+2S\}$, the coefficient of $(\cos\frac{\theta}{2})^{4S-2k+\mu}(\sin\frac{\theta}{2})^{2k-\mu}$ in Eq. (A10) is zero

$$0 = \sum_{m_1} \alpha_{m_1+\mu,m_1} \sum_{k_1} (-1)^{k+m-m_1} W(S, m, m_1+\mu, k_1) W(S, m_1, m, k-k_1). \tag{A13}$$

Consider a subset of requirements where $m = S$, the factors $k_1!(-k_1)!$ and $(S-k_2-m_1)!(k_2+m_1-S)!$ in the denominators *confines* $k_1 = 0$ and $m_1 = S-k$ in the summation in Eq. (A13), respectively. Thus we confirm

$$0 = \alpha_{S-k+\mu,S-k}(2S)![(2S-k+\mu)!(k-\mu)!(2S-k)!k!]^{-1/2}, \tag{A14}$$

which suffices $\alpha = 0$ as k, μ runs over whatever they are allowed.

Since all the q by q Hermitian matrix constitute a q^2-dimension $\mathbb{R}$ space, denoted by $\mathcal{H}$, it is implied that $\{|\boldsymbol{n}, m\rangle\langle \boldsymbol{n}, m||\boldsymbol{n} \in S_2, m \in \{S, S-1, \ldots, -S\}\}$ includes a basis of $\mathcal{H}$.

2. Back to general N case

The space of all q^N by q^N Hermitian matrices is q^{2N} dimensional, and it is also $\mathcal{H}^{\otimes N}$, thus $\{\otimes_{j=1}^N |\boldsymbol{n}_j, m_j\rangle\langle \boldsymbol{n}_j, m_j||\boldsymbol{n}_j \in S_2, m_j \in \{S, S-1, \ldots, -S\}\}$ includes a basis of $\mathcal{H}^{\otimes N}$. Hence Eq. (A5) suffices $\alpha = 0$.

Thus we confirm that $(\cdot, \cdot)$ is an inner product in the q^N by q^N Hermitian matrices space. Since all inner products in the same space are equivalent to each other, $(\alpha, \alpha) = C||\alpha||^2$. In summary,

$$\sqrt{2D_{\mathrm{KL}}(\mathbb{P}[\sigma]||\mathbb{P}[\rho])} \geqslant 2\delta(P[\rho], P[\sigma]) \geqslant \int d\mu|\mathbb{P}[\rho] - \mathbb{P}[\sigma]| \geqslant \frac{1}{2}\int d\mu F(\{\boldsymbol{n}\})(\langle\{\boldsymbol{n}, m\}|\sigma - \tilde{\rho}|\{\boldsymbol{n}, m\}\rangle)^2 = \frac{C}{2}||\sigma - \tilde{\rho}||^2.$$

APPENDIX B: THE INFORMATION BOUND AND ASYMPTOTIC NORMALITY

1. Scaling of number of replicas with system size

Quantum tomography can in general be regarded as a kind of parameter estimation, in the way that (1) the measurement outcomes are independent identical distributed random variables and (2) the model wave function $\tilde{\Psi}$ is an estimator of the target wave function ψ. The target state wave function under any basis is a q^N-dimensional vector depending on much fewer parameters θ^0 in its modeled description, e.g., MPS representation, and in accordance $\tilde{\Psi}$ is a q^N-dimensional estimator with adjustable parameters. To be exactly a parameter estimation problem, however, the model should be fixed, which in our scheme is achieved after the dynamical allocation of parameters. For simplicity, we denote each outcome as random variable X, the whole bunch of real and imaginary parts of the tensor entries as parameters θ, and the parameters corresponding to the laboratory state ψ as θ^0. Applying the Cramér-Rao bound to the covariance matrix of $\tilde{\Psi}$,

$$\mathrm{Cov}_{\theta^0}(\tilde{\Psi}_i, \tilde{\Psi}_j) \geqslant \frac{1}{|\mathcal{V}|}\sum_{\alpha,\beta}\left(\frac{\partial\psi_i}{\partial\theta_\alpha}[I(\theta)^{-1}]_{\alpha\beta}\frac{\partial\psi_j^*}{\partial\theta_\beta}\right)\Big|_{\theta^0}, \tag{B1}$$

where I is the Fisher information matrix, whose entries are defined as

$$I(\theta)_{\alpha\beta} = \mathbb{E}_{\theta^0}\left[\frac{\partial}{\partial\theta_\alpha}\ln\mathbb{P}(X|\theta)\frac{\partial}{\partial\theta_\beta}\ln\mathbb{P}(X|\theta)\right]. \tag{B2}$$

In this expression, $\mathbb{P}$ is the probability density function for possible outcomes X that is compounded by the choice of measurement basis $\mathcal{B}$ and the corresponding measuring result; the parameters are the ones in the MPS description of the modeled state, while the expectation $\mathbb{E}_\theta$ is taken with the target state parameters; $\mathcal{V}$ is the set of measurement outcomes, whose size is the number of samples for parameter estimation or equivalently the number of measured replicas of the quantum state. The inequality means as matrices, the difference of the two sides is semipositive definite. It is also noteworthy that as a result of asymptotic normality of MLE, in (B1) the equality asymptotically take holds [67].

The distance between the model and the target state can be characterized by the trace of the covariance matrix $\sum_j \mathrm{Var}\tilde{\Psi}_j$. Since inequality (B1) is about positive definiteness, taking trace keeps the inequality, thus we have

$$\mathcal{R}^2_{\mathrm{real}}|\mathcal{V}| \geqslant \mathrm{Tr}\,(KI^{-1})|_{\theta^0} \tag{B3}$$

in which

$$K_{\alpha\beta} = \sum_j \frac{\partial\psi_j}{\partial\theta_\alpha}\frac{\partial\psi_j^*}{\partial\theta_\beta}. \tag{B4}$$

We can easily verify that both K and I are semipositive definite. Inequality (B3) is a lower bound of the average distance between the tomographic state and the target state and become the exact asymptotic behavior as $|\mathcal{V}| \to \infty$.

To further see how the lower bound scales with N, $D_{\max}$, and q, we apply inequalities $\mathrm{Tr}\,AB \geqslant \lambda_{\min}(A)\,\mathrm{Tr}\,B$ [68] and $\mathrm{Tr}(A^{-1}) \geqslant \dim(A)^2/\mathrm{Tr}(A)$ for positive-definite matrices A, B, obtaining

$$\mathrm{Tr}(KI^{-1}) \geqslant \lambda_{\min}(K)\,\mathrm{Tr}(I^{-1}) \geqslant \lambda_{\min}(K)\frac{\Theta^2}{\mathrm{Tr}\,I}, \tag{B5}$$

where Θ is the total number of parameters, $\Theta \approx 2qD^2_{\max}N$ as N grows. We notice that when θ_α and θ_β reside on far separated sites in the MPS, the corresponding matrix element of K would be exponentially small due to the multiplication of transfer matrices. So the minimal eigenvalue of K should not scale with the system size, but only depend on the intrinsic property of the MPS representation of the target state.

On the other hand, $\mathrm{Tr}\,I$ is related to wave functions under different bases by

$$\mathrm{Tr}\,I = \sum_\alpha \int dX \frac{1}{\mathbb{P}(X|\theta)}[\partial_{\theta_\alpha}\mathbb{P}(X|\theta)]^2 = \int d\mu(\mathcal{B})Q(\mathcal{B}), \tag{B6}$$

$$Q(\mathcal{B}) \equiv \sum_\alpha \sum_{|m\rangle\in\mathcal{B}} (\partial_{\theta_\alpha}|\langle m|\psi\rangle|^2)^2/|\langle m|\psi\rangle|^2. \tag{B7}$$

μ represents the probability measure of choosing basis, $d\mu[\mathcal{B}(\{\boldsymbol{n}\})] = d\boldsymbol{n}_1 d\boldsymbol{n}_2\cdots\boldsymbol{n}_N f(\{\boldsymbol{n}\})$. Now let us look into the wave function under basis $\mathcal{B}$, $\varphi_m \equiv \langle m|\psi\rangle$, which is transformed from ψ_m by local unitary transformations. Separating the amplitude and phase of the wave function $\varphi = \sqrt{p}\exp(i\phi)$, we have

$$\begin{aligned}\sum_{m,\alpha} 4|\partial_{\theta_\alpha}\varphi_m|^2 &= \sum_{m,\alpha} 4\left|e^{i\phi_m}\partial_{\theta_\alpha}\sqrt{p_m} + ie^{i\phi_m}\sqrt{p_m}\partial_{\theta_\alpha}\phi_m\right|^2 \\ &\geqslant \sum_{m,\alpha}\frac{(\partial_{\theta_\alpha}p_m)^2}{p_m} = Q(\mathcal{B}).\end{aligned} \tag{B8}$$

For simplicity the notation specifying the basis is omitted for φ, p, and ϕ, although they are basis dependent. $\sum_m|\partial_{\theta_\alpha}\varphi_m|^2$, however, is independent of basis and is less than unity because of the normalization and canonical conditions of the MPS. Therefore,

$$\mathrm{Tr}\,I \leqslant \int d\mu(\mathcal{B})\sum_\alpha 1 \sim \Theta, \tag{B9}$$

and consequently $\mathcal{R}^2_{\mathrm{real}}|\mathcal{V}| \geqslant \lambda_{\min}(K)\Theta$. When N grows, this lower bound scales as $\lambda NqD^2_{\max}$, where λ is a constant independent of N. We note that our above analysis is basis independent, so our result applies to any choice of measurement bases.

The above scaling relation is the main result of this Appendix. As a special application, we see for a prescribed distance criterion $\mathcal{R}$, the lower bound of the required number of replicas is $O(N)$.

2. Behaviors of distances over number of replicas

Revisiting inequality (B3) and recalling the asymptotic equality, we see the asymptotic behavior of the distance for a certain laboratory state $\mathcal{R}_{\mathrm{real}} \propto |\mathcal{V}|^{-1/2}$, because $\mathrm{Tr}(KI^{-1})|_{\theta^0}$ totally depends on the laboratory state.

To handle the successive distance, denote the estimate at the stage $|\mathcal{V}| = V$ as $\tilde{\theta}^{(V)}$, which optimizes the loss function $\mathcal{L}^{(V)}(\theta) = V^{-1}\sum_{i=1}^V \ln\mathbb{P}(X_i|\theta)$. Recall that X_i are independent random variables obeying the identical distribution density $\mathbb{P}(X|\theta^0)$. As a new measurement is carried out, $\tilde{\theta}^{(V)}$ might no longer be the optimal estimate, but we can nevertheless relate $\tilde{\theta}^{(V+1)}$ to $\tilde{\theta}^{(V)}$ by

$$\frac{-1}{V+1}\partial_{\theta_\alpha}\ln\mathbb{P}(X_{V+1}|\tilde{\theta}^{(V)}) = \partial_{\theta_\alpha}\mathcal{L}^{(V+1)}(\tilde{\theta}^{(V+1)}) - \partial_{\theta_\alpha}\mathcal{L}^{(V+1)}(\tilde{\theta}^{(V)}) = -\sum_\beta G^{(V+1)}_{\alpha\beta}(\hat{\theta})\left(\tilde{\theta}^{(V+1)}_\beta - \tilde{\theta}^{(V)}_\beta\right), \tag{B10}$$

$$G^{(V)}_{\alpha\beta}(\theta) \equiv -\partial_{\theta_\alpha}\partial_{\theta_\beta}\mathcal{L}^{(V)}(\theta), \tag{B11}$$

where the fact $\partial_\theta \mathcal{L}^{(V)}(\tilde{\theta}^{(V)}) = 0$ and Lagrangian mean theorem are used, and thus

$$(V+1)\big(\tilde{\theta}_\alpha^{(V+1)} - \tilde{\theta}_\alpha^{(V)}\big) = \sum_\beta [G^{(V+1)}(\hat{\theta})]^{-1}_{\alpha\beta} \partial_{\theta_\beta} \ln \mathbb{P}(X_{V+1}|\tilde{\theta}^{(V)}). \tag{B12}$$

As V grows, $G^{(V)}_{\alpha\beta}(\theta) \to \mathbb{E}_{\theta^0}[-\partial_{\theta_\alpha}\partial_{\theta_\beta} \ln \mathbb{P}(X|\theta)] = I(\theta)_{\alpha\beta}$, and $\hat{\theta}$ is confined between $\tilde{\theta}^{(V)}$, $\tilde{\theta}^{(V+1)}$ to approach θ^0, therefore

$$(V+1)^2 \mathbb{E}_{\theta^0}\big(\big(\tilde{\theta}_\alpha^{(V+1)} - \tilde{\theta}_\alpha^{(V)}\big)\big(\tilde{\theta}_\beta^{(V+1)} - \tilde{\theta}_\beta^{(V)}\big)\big) \to \sum_{\gamma,\lambda} I(\hat{\theta})^{-1}_{\alpha\gamma} I(\tilde{\theta}^{(V)})_{\gamma\lambda} I(\hat{\theta})^{-1}_{\lambda\beta} \to I(\theta^0)^{-1}_{\alpha\beta}. \tag{B13}$$

The successive distance between the model states is thus

$$V^2 \mathcal{R}^2_{\text{succ}} = V^2 \sum_{\alpha,\beta,j} \frac{\partial \psi_j}{\partial \theta_\alpha} \mathbb{E}_{\theta^0}\big(\big(\tilde{\theta}_\alpha^{(V+1)} - \tilde{\theta}_\alpha^{(V)}\big)\big(\tilde{\theta}_\beta^{(V+1)} - \tilde{\theta}_\beta^{(V)}\big)\big) \frac{\partial \psi_j}{\partial \theta_\beta} \to \text{Tr}(KI^{-1})\Big|_{\theta^0}, \tag{B14}$$

which also supports that the ratio $\mathcal{R}^2_{\text{real}}/\mathcal{R}_{\text{succ}}$ goes to a constant $\sqrt{\text{Tr}(KI^{-1})|_{\theta^0}}$ depending on the target state.

[1] H. Häffner, W. Hänsel, C. F. Roos, J. Benhelm, D. Chek-al kar, M. Chwalla, T. Körber, U. D. Rapol, M. Riebe, P. O. Schmidt, C. Becher, O. Gühne, W. Dür, and R. Blatt, Scalable multiparticle entanglement of trapped ions, Nature (London) **438**, 643 (2005).

[2] J. Kelly, R. Barends, A. G. Fowler, A. Megrant, E. Jeffrey, T. C. White, D. Sank, J. Y. Mutus, B. Campbell, Yu Chen, Z. Chen, B. Chiaro, A. Dunsworth, I.-C. Hoi, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, A. Vainsencher, J. Wenner, A. N. Cleland, and J. M. Martinis, State preservation by repetitive error detection in a superconducting quantum circuit, Nature (London) **519**, 66 (2015).

[3] T. Monz, D. Nigg, E. A. Martinez, M. F. Brandl, P. Schindler, R. Rines, S. X. Wang, I. L. Chuang, and R. Blatt, Realization of a scalable Shor algorithm, Science **351**, 1068 (2016).

[4] C. Song, K. Xu, W. Liu, C.-p. Yang, S.-B. Zheng, H. Deng, Q. Xie, K. Huang, Q. Guo, L. Zhang, P. Zhang, D. Xu, D. Zheng, X. Zhu, H. Wang, Y.-A. Chen, C.-Y. Lu, S. Han, and J.-W. Pan, 10-Qubit Entanglement and Parallel Logic Operations with a Superconducting Circuit, Phys. Rev. Lett. **119**, 180511 (2017).

[5] K. Xu, J.-J. Chen, Y. Zeng, Y.-R. Zhang, C. Song, W. Liu, Q. Guo, P. Zhang, D. Xu, H. Deng, K. Huang, H. Wang, X. Zhu, D. Zheng, and H. Fan, Emulating Many-Body Localization with A Superconducting Quantum Processor, Phys. Rev. Lett. **120**, 050507 (2018).

[6] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, V. Vuletić, and M. D. Lukin, Probing many-body dynamics on a 51-atom quantum simulator, Nature (London) **551**, 579 (2017).

[7] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z.-X. Gong, and C. Monroe, Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator, Nature (London) **551**, 601 (2017).

[8] S. Bandyopadhyay, P. O. Boykin, V. Roychowdhury, and F. Vatan, A new proof for the existence of mutually unbiased bases, Algorithmica **34**, 512 (2002).

[9] T. Durt, B.-G. Englert, I. Bengtsson, and K. Życzkowski, On mutually unbiased bases, Int. J. Quantum. Inform. **08**, 535 (2010).

[10] R. B. A. Adamson and A. M. Steinberg, Improving Quantum State Estimation with Mutually Unbiased Bases, Phys. Rev. Lett. **105**, 030406 (2010).

[11] P. Busch, Informationally complete sets of physical quantities, Int. J. Theor. Phys. **30**, 1217 (1991).

[12] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, Symmetric informationally complete quantum measurements, J. Math. Phys. **45**, 2171 (2004).

[13] S. T. Flammia, On SIC-POVMs in prime dimensions, J. Phys. A: Math. Gen. **39**, 13483 (2006).

[14] D. Appleby, H. Dang, and C. Fuchs, Symmetric informationally-complete quantum states as analogues to orthonormal bases and minimum-uncertainty states, Entropy **16**, 1484 (2014).

[15] S. T. Flammia, A. Silberfarb, and C. M. Caves, Minimal informationally complete measurements for pure states, Found. Phys. **35**, 1985 (2005).

[16] J. Finkelstein, Pure-state informationally complete and "really" complete measurements, Phys. Rev. A **70**, 052107 (2004).

[17] T. Heinosaari, L. Mazzarella, and M. M. Wolf, Quantum tomography under prior information, Commun. Math. Phys. **318**, 355 (2013).

[18] D. Goyeneche, G. Cañas, S. Etcheverry, E. S. Gómez, G. B. Xavier, G. Lima, and A. Delgado, Five Measurement Bases Determine Pure Quantum States on Any Dimension, Phys. Rev. Lett. **115**, 090401 (2015).

[19] C. Carmeli, T. Heinosaari, J. Schultz, and A. Toigo, How many orthonormal bases are needed to distinguish all pure quantum states? Eur. Phys. J. D **69**, 179 (2015).

[20] C. Carmeli, T. Heinosaari, M. Kech, J. Schultz, and A. Toigo, Stable pure state quantum tomography from five orthonormal bases, Europhys. Lett. **115**, 30001 (2016).

[21] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, Sample-optimal tomography of quantum states, IEEE Trans. Inf. Theory **63**, 5628 (2017).

[22] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, Efficient quantum state tomography, Nat. Commun. **1**, 149 (2010).

[23] T. Baumgratz, A. Nüßeler, M. Cramer, and M. B. Plenio, A scalable maximum likelihood method for quantum state tomography, New J. Phys. **15**, 125004 (2013).

[24] 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, Efficient tomography of a quantum many-body system, Nat. Phys. **13**, 1158 (2017).

[25] B. Qi, Z. Hou, Y. Wang, D. Dong, H.-S. Zhong, L. Li, G.-Y. Xiang, H. M. Wiseman, C.-F. Li, and G.-C. Guo, Adaptive quantum state tomography via linear regression estimation: Theory and two-qubit experiment, npj Quant. Info. **3**, 19 (2017).

[26] C. Granade, C. Ferrie, and S. T. Flammia, Practical adaptive quantum tomography, New J. Phys. **19**, 113017 (2017).

[27] A. Steffens, M. Friesdorf, T. Langen, B. Rauer, T. Schweigler, R. Hübener, J. Schmiedmayer, C. A. Riofrío, and J. Eisert, Towards experimental quantum-field tomography with ultracold atoms, Nat. Commun. **6**, 7663 (2015).

[28] S. T. Flammia, D. Gross, S. D. Bartlett, and R. Somma, Heralded polynomial-time quantum state tomography, arXiv:1002.3839.

[29] A. Steffens, C. A. Riofrío, R. Hübener, and J. Eisert, Quantum field tomography, New J. Phys. **16**, 123010 (2014).

[30] G. Tóth, W. Wieczorek, D. Gross, R. Krischek, C. Schwemmer, and H. Weinfurter, Permutationally Invariant Quantum Tomography, Phys. Rev. Lett. **105**, 250403 (2010).

[31] D. Pérez-García, F. Verstraete, M. M. Wolf, and J. I. Cirac, Matrix product state representations, Quant. Inform. Comput. **7**, 401 (2007).

[32] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Quantum State Tomography via Compressed Sensing, Phys. Rev. Lett. **105**, 150401 (2010).

[33] A. Kalev, R. L. Kosut, and I. H. Deutsch, Quantum tomography protocols with positivity are compressed sensing protocols, npj Quant. Info. **1**, 15018 (2015).

[34] S. Aaronson, The learnability of quantum states, Proc. R. Soc. London Ser. A **463**, 3089 (2007).

[35] A. Rocchetto, Stabiliser states are efficiently PAC-learnable, Quant. Inform. Comput. **18**, 541 (2018).

[36] A. Rocchetto, S. Aaronson, S. Severini, G. Carvacho, D. Poderini, I. Agresti, M. Bentivegna, and F. Sciarrino, Experimental learning of quantum states, Quantum. Inform. Meas. (QIM) V: Quantum Tech. F5A.26 (2019), doi: 10.1364/QIM.2019.F5A.26.

[37] A. Rocchetto, E. Grant, S. Strelchuk, G. Carleo, and S. Severini, Learning hard quantum distributions with variational autoencoders, npj Quant. Info. **4**, 28 (2018).

[38] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, Quantum tomography via compressed sensing: Error bounds, sample complexity and efficient estimators, New J. Phys. **14**, 095022 (2012).

[39] F. Huszár and N. M. T. Houlsby, Adaptive Bayesian quantum tomography, Phys. Rev. A **85**, 052120 (2012).

[40] C. Ferrie, Self-Guided Quantum Tomography, Phys. Rev. Lett. **113**, 190404 (2014).

[41] A. Kyrillidis, A. Kalev, D. Park, S. Bhojanapalli, C. Caramanis, and S. Sanghavi, Provable compressed sensing quantum state tomography via non-convex methods, npj Quant. Info. **4**, 36 (2018).

[42] D. Ahn, Y. S. Teo, H. Jeong, F. Bouchard, F. Hufnagel, E. Karimi, D. Koutný, J. Řeháček, Z. Hradil, G. Leuchs, and L. L. Sánchez-Soto, Adaptive Compressive Tomography with no *A Priori* Information, Phys. Rev. Lett. **122**, 100404 (2019).

[43] D. L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory **52**, 1289 (2006).

[44] E. J. Candes and T. Tao, Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies? IEEE Trans. Inf. Theory **52**, 5406 (2006).

[45] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016).

[46] S. T. Flammia and Y.-K. Liu, Direct Fidelity Estimation from Few Pauli Measurements, Phys. Rev. Lett. **106**, 230501 (2011).

[47] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Practical Characterization of Quantum Devices without Tomography, Phys. Rev. Lett. **107**, 210404 (2011).

[48] O. Gühne, C.-Y. Lu, W.-B. Gao, and J.-W. Pan, Toolbox for entanglement detection and fidelity estimation, Phys. Rev. A **76**, 030305(R) (2007).

[49] Z.-Y. Han, J. Wang, H. Fan, L. Wang, and P. Zhang, Unsupervised Generative Modeling Using Matrix Product States, Phys. Rev. X **8**, 031012 (2018).

[50] S. R. White, Density Matrix Formulation for Quantum Renormalization Groups, Phys. Rev. Lett. **69**, 2863 (1992).

[51] A. Keesling, A. Omran, H. Levine, H. Bernien, H. Pichler, S. Choi, R. Samajdar, S. Schwartz, P. Silvi, S. Sachdev, P. Zoller, M. Endres, M. Greiner, V. Vuletić, and M. D. Lukin, Scalable multiparticle entanglement of trapped ions, Nature (London) **568**, 207 (2019).

[52] C. H. Baldwin, I. H. Deutsch, and A. Kalev, Strictly-complete measurements for bounded-rank quantum-state tomography, Phys. Rev. A **93**, 052105 (2016).

[53] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, Ann. Phys. **326**, 96 (2011).

[54] R. Orús, A practical introduction to tensor networks: Matrix product states and projected entangled pair states, Ann. Phys. **349**, 117 (2014).

[55] S. R. White, Density-matrix algorithms for quantum renormalization groups, Phys. Rev. B **48**, 10345 (1993).

[56] See Supplemental Materials at http://link.aps.org/supplemental/10.1103/PhysRevA.101.032321 for detailed calculation of gradients and extra examples showing the asymptotic behavior.

[57] R. D. Gill and S. Massar, State estimation for large ensembles, Phys. Rev. A **61**, 042312 (2000).

[58] L. Pereira, L. Zambrano, J. Cortés-Vega, S. Niklitschek, and A. Delgado, Adaptive quantum tomography in high dimensions, Phys. Rev. A **98**, 012339 (2018).

[59] J. Eisert, Entanglement and tensor network states, Model. Simul. **3**, 520 (2013).

[60] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki, Rigorous Results on Valence-Bond Ground States in Antiferromagnets, Phys. Rev. Lett. **59**, 799 (1987).

[61] G. Vidal, Class of Quantum Many-Body States that Can be Efficiently Simulated, Phys. Rev. Lett. **101**, 110501 (2008).

[62] O. Landon-Cardinal and D. Poulin, Practical learning method for multi-scale entangled states, New J. Phys. **14**, 085004 (2012).

[63] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).

[64] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Neural-network quantum state tomography, Nat. Phys. **14**, 447 (2018).

[65] J. Wang, Z.-Y. Han, and Z.-Y. Li, Data for "Quantum Tomography with Born Machine", https://github.com/congzlwag/BornMachineTomo (2019).

[66] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems* (Cambridge University Press, Cambridge, 2011).

[67] J. A. Rice, *Mathematical Statistics and Data Analysis* (Cengage Learning, Boston, MA, 2006).

[68] Y. Fang, K. A. Loparo, and X. Feng, Inequalities for the trace of matrix product, IEEE Trans. Autom. Control **39**, 2489 (1994).