Flexible learning of quantum states with generative query neural networks

Yan Zhu,[1] Ya-Dong Wu,[1, *] Ge Bai,[1] Yuexuan Wang,[1, 2] and Giulio Chiribella[1, 3, 4, †]

[1] *QICI Quantum Information and Computation Initiative, Department of Computer Science,*
The University of Hong Kong, Pokfulam Road, Hong Kong
[2] *College of Computer Science and Technology, Zhejiang University, Hangzhou, China*
[3] *Department of Computer Science, Parks Road, Oxford, OX1 3QD, United Kingdom*
[4] *Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada*

Deep neural networks are a powerful tool for characterizing quantum states. In this task, neural networks are typically trained with measurement data gathered from the quantum state to be characterized. But is it possible to train a neural network in a general-purpose way, which makes it applicable to multiple unknown quantum states? Here we show that learning across multiple quantum states and different measurement settings can be achieved by a generative query neural network, a type of neural network originally used in the classical domain for learning 3D scenes from 2D pictures. Our network can be trained offline with classically simulated data, and later be used to characterize unknown quantum states from real experimental data. With little guidance of quantum physics, the network builds its own data-driven representation of quantum states, and then uses it to predict the outcome probabilities of requested quantum measurements on the states of interest. This approach can be applied to state learning scenarios where quantum measurement settings are not informationally complete and predictions must be given in real time, as experimental data become available, as well as to adversarial scenarios where measurement choices and prediction requests are designed to expose learning inaccuracies. The internal representation produced by the network can be used for other tasks beyond state characterization, including clustering of states and prediction of physical properties. The features of our method are illustrated on many-qubit ground states of Ising model and continuous-variable non-Gaussian states.

I. INTRODUCTION

Accurate characterization of quantum systems is crucial for the development, certification, and benchmarking of new quantum technologies [1]. Accordingly, major efforts have been invested into developing suitable techniques for characterizing quantum states, including quantum state tomography [2–6], classical shadow estimation [7, 8], partial state characterization [9, 10] and, more theoretically, quantum state learning [11–13]. Recently, the dramatic development of artificial intelligence inspired new methods of quantum state characterization, in which techniques from the field of machine learning [14] are used to learn descriptions of quantum states from experimental data. In particular, a sequence of works explored the possibility of using neural networks for quantum state characterization [15–26].

When training neural networks for quantum state characterization, most existing works use experimental data generated from a specific quantum state under consideration. As a consequence, the information learnt by the network cannot be directly transferred to other states: for a new quantum state, the training must restart afresh. This limitation affects the efficiency of the network in scenarios where multiple quantum states need to be characterized, and prevents the application of the state characterization method to important tasks, such as quantum state clustering [27], quantum state classification [28], and quantum cross-platform verification [29], where information from multiple quantum states needs to be compared.

In this paper, we develop a flexible neural network capable of learning across multiple quantum states. To achieve this result, we adopt a type of neural network, known as generative query network, which was originally developed for the task of learning 3D scenes from 2D images taken from different viewpoints without human labels [30]. By adapting these techniques to the quantum case, we develop a generative query network for quantum state learning, named GQNQ. The key idea of GQNQ is to let the network work out its own representations for a family of quantum states. While quantum tomographic protocols strive to output a quantum state that bests fits data, GQNQ learns its own internal representation of a quantum state, and uses it to predict the statistics of new measurements that have not been performed yet. The basic structure of GQNQ, together with a comparison with tomographic reconstruction, is illustrated in Fig. 1. GQNQ satisfies the following properties in learning quantum states:

- Generality: there is no constraint on the type of quantum states as well as on the measurements.
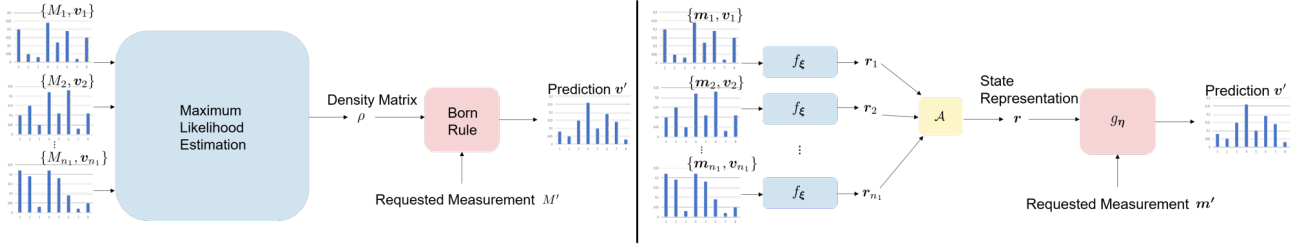
---

* yadongwu@hku.hk
† giulio@cs.hku.hk

Figure 1. Comparison between a tomographic quantum state reconstruction and GQNQ. In quantum state tomography, the experimenter infers a mathematical description of the quantum state, typically in the form of a density matrix, using maximal likelihood estimation or some other technique. Using the estimated density matrix, the experimenter can then predict the outcome probabilities for unperformed quantum measurements. In GQNQ, the task of finding a mathematical representation of the quantum state takes place inside a neural network. First, a representation network learns a concise representation of quantum states from the set of measurement data. Then, a generation network uses this representation to predict outcome probability distribution for any measurement of interest.

- Flexibility: after the model is trained for a certain class of states and measurements, given outcomes of any number of measurements in the class, the model can predict the outcomes of any remaining measurement.

- Real-time prediction: training can be done offline, and state predictions can be generated immediately after the measurement data are given.

- Learner-blindness: information of the quantum state and measurements is kept hidden from the learner.

With these features, GQNQ can be applied in real-time control and calibration of various quantum state preparation devices. GQNQ can also be adapted to online quantum state learning without the need for storing measurement data or postprocessing, and is even applicable in an adversarial scenario, where the measurements are chosen by an adversary. Furthermore, GQNQ can be applied in blind quantum computing, where GQNQ successfully predicts measurement outcomes for requested measurements, but is kept blind to the true quantum state, as GQNQ receives only certain parametrizations of the measurements, instead of their full descriptions.

To the best of our knowledge, no previous approach possesses all these features. For example, the data-driven approach in Ref. [21] fixes measurement settings and the number of measurements in their network, thus violating flexibility, and the approach of maximizing likelihood and entropy [31] can neither give predictions in real time due to its computational complexity nor hide information of density matrix.

We test GQNQ by multi-qubit states, including ground states of Ising models, with local Pauli basis measurement data, as well as continuous-variable quantum states, including Gaussian states, cat states [32], and Gottesman-Kitaev-Preskill (GKP) states [33], with homodyne measurement data. We demonstrate that the state representation, learned by GQNQ from few measurement data, retains the key information about the corresponding quantum state, in the sense that this representation successfully classifies different classes of quantum states and predicts their important physical properties. We also find that GQNQ can mitigate stochastic noises introduced by finite repetitions of measurements.

II. RESULTS

A. Framework

Let us first briefly introduce a problem of quantum state learning in an operational way. For any unknown quantum state ρ, a measurement device can apply any measurement in a set of n d-outcome measurements $\mathcal{M} := \{M_i\}_{i=1}^n$, where each measurement is described by a positive operator-valued measure (POVM) $M_i := \left\{M_i^{(j)}\right\}_{j=1}^d$, satisfying the conditions $M_i^{(j)} \geq 0 \,\forall j \in \{1, \ldots d\}$ and $\sum_{j=1}^d M_i^{(j)} = \mathbb{1}$. A referee randomly chooses n_1 POVMs from $\mathcal{M}$, forming a subset $\mathcal{M}_1 \subset \mathcal{M}$, which may or may not be informationally incomplete. For each POVM measurement $M_i \in \mathcal{M}_1$, the measurement device repeats this measurement (ideally) infinite times, and returns a d-dimensional probability distribution $\boldsymbol{v}_i := \left\{v_i^{(j)}\right\}_{j=1}^d \in \mathbb{R}^d$, where $v_i^{(j)} = \mathrm{tr}\left(\rho M_i^{(j)}\right)$, as well as the parametrization of M_i, denoted by $\boldsymbol{m}_i$. For example, for continuous-variable quantum states, $\boldsymbol{m}_i$ can be the quadrature phase of a homodyne measurement POVM M_i. Intuitively, parametrization of measurements can be considered as dials at the panel of a measurement

device. The referee hands the measurement data, i.e. the set $\{\boldsymbol{m}_i, \boldsymbol{v}_i\}_{i=1}^{n_1}$, to a state learner, who may have no knowledge about quantum physics. The referee then poses to the state learner a query question $\boldsymbol{m}'$, which is the parametrization of a randomly chosen POVM $M' \in \mathcal{M} \setminus \mathcal{M}_1$. By giving parametrization $\boldsymbol{m}'$ instead of the full description of M', the referee provides the learner with the minimum amount of information necessary to correctly predict the outcome statistics for any query measurement. The goal of the learner is to predict the correct outcome probability distribution $\boldsymbol{v}' = \left\{ \operatorname{tr}\left(\rho M'^{(j)} \right) \right\}_{j=1}^{d}$. This setting includes as special cases quantum state reconstruction and quantum statistical query learning [34].

The above state learning problem can be further extended to online learning of quantum states [13], where measurement data are not handed in to a state learner at once, but revealed to the learner one by one. Online state learning can be considered as an adversarial game between the referee and the learner. The referee randomly chooses a sequence of POVM measurements $M_1, M_2, \ldots, M_l$, which is unknown to the learner. At time stage i ($i \in \mathbb{N}^+, i \leq l-1$), the referee applies POVM measurement M_i to sufficient copies of state ρ, and gives the measurement data $\{\boldsymbol{m}_i, \boldsymbol{v}_i\}$ to the state learner. At any time stage t ($t \leq l-1$), when the learner has collected the set of measurement data $\{\boldsymbol{m}_i, \boldsymbol{v}_i\}_{i=1}^{t}$, the referee asks the learner to predict the measurement outcome probabilities for all the remaining POVM measurements in $\mathcal{M} \setminus \{M_i\}_{i=1}^{t}$.

In practical scenarios, the number of repetitions of measurements is finite. Hence, the inputs $\tilde{\boldsymbol{v}}_i$, received by GQNQ does not exactly equal to ideal $\boldsymbol{v}_i$. From these noisy inputs $\{\boldsymbol{m}_i, \tilde{\boldsymbol{v}}_i\}_i$, the aim of GQNQ is still to predict the ideal outcome probability distribution $\{\operatorname{tr}(\rho M'^{(j)})\}_{j=1}^{d}$ for a query POVM M'.

B. Neural Network Structure

To solve the above quantum state learning problem, GQNQ is composed of two main parts: the representation network and the generation network. Consider a given set of n_1 POVM measurements $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_{n_1}$, and its corresponding outcome probabilities $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n_1}$. For each pair of input $\{\boldsymbol{m}_i, \boldsymbol{v}_i\}$, the representation network $f_{\boldsymbol{\xi}}$ with parameter $\boldsymbol{\xi}$ outputs a state representation $\boldsymbol{r}_i := f_{\boldsymbol{\xi}}(\boldsymbol{m}_i, \boldsymbol{v}_i)$. Using the set of representations $\{\boldsymbol{r}_i\}_{i=1}^{n_1}$, an aggregate function outputs a representation $\boldsymbol{r}$ of state ρ. Here, for simplicity, we use the average representation $\boldsymbol{r} := \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{r}_i$. Given the representation $\boldsymbol{r}$ and a random query $\boldsymbol{m}'$, the generation network $g_{\boldsymbol{\eta}}$ with parameter $\boldsymbol{\eta}$ produces $g_{\boldsymbol{\eta}}(\boldsymbol{r}, \boldsymbol{m}') \in \mathbb{R}^d$, which is a prediction of $\boldsymbol{v}'$. We use variational approximations [35] and optimize those parameters $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ by batch gradient descent in the training phase.

During the whole training phase, we do not give any label of quantum states to GQNQ or perform any data preprocessing for the input data. Our representation network and generation network are trained jointly to minimize the average difference between predicted outcome distributions and the real distributions. As the representation network has no knowledge about which POVM M' is queried for prediction, it must produce such a representation $\boldsymbol{r}$ that contains sufficient information of ρ for the purpose of precise predictions of the outcome probabilities $\boldsymbol{v}'$ of any measurement $\boldsymbol{m}' \in \mathcal{M} \setminus \mathcal{M}_1$, instead of only measurements in $\mathcal{M}_1$. By training GQNQ through back-propagation across many different quantum states, and randomizing $\mathcal{M}_1$ of measured POVMs as well as queried POVMs $\mathcal{M} \setminus \mathcal{M}_1$ for each state, our representation network can obtain an accurate and concise state representation $\boldsymbol{r}$. The entire training phase is carried out offline using data from classical simulation. After training is done, the trained neural network can be used for learning quantum states from experimental data.

C. Quantum State Learning: Numerical Experiments

We test GQNQ using the ground state of the Hamiltonian of a one-dimensional Ising model with L sites,

$$H = -\left(\sum_{i=0}^{L-2} J_i \sigma_i^z \sigma_{i+1}^z + g \sum_{j=0}^{L-1} \sigma_j^x \right), \tag{1}$$

where each J_i denotes the coupling strength between nearest neighbour sites i and $i+1$, and g denotes the strength of an external magnetic field. Here each J_i is a Gaussian random number with mean value J and variance $\sigma^2 = 0.1$. We choose the parameter region $J \in [0.5, 1.5]$ and $g = 1$. We consider two different scenarios: In the first scenario, when $L = 6$, we consider all possible six-qubit Pauli basis measurements, while in the second scenario, when $L = 10, 20$ and 50, we only consider Pauli basis measurements on nearest-neighbour qubits. In both scenarios, given measurement data in only a few random Pauli bases, GQNQ predicts the outcome probabilities of all the other possible Pauli basis measurements. The classical fidelity of predictions with respect to the ground truth are plotted in Fig. 2.
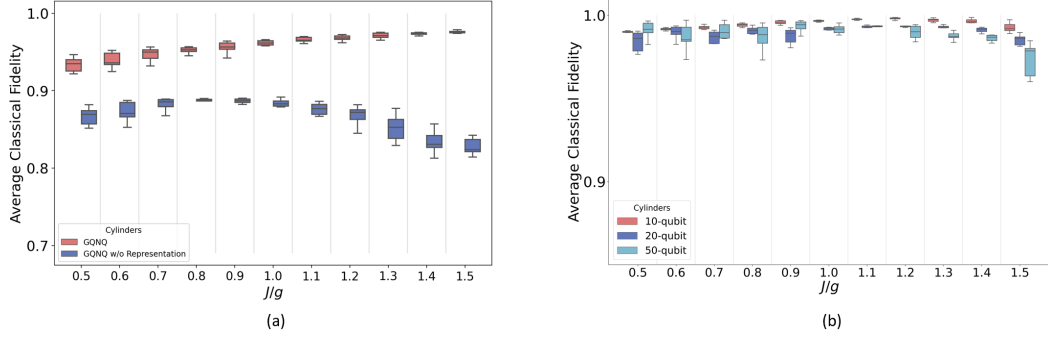
Figure 2. Performance of GQNQ on Ising model ground states as a function of J/g visualized by boxplots [36]. In figure (a), ρ is a six-qubit Ising model ground state. The class of measurements $\mathcal{M}$ is the set of all $3^6 = 729$ six-qubit Pauli-basis measurements. In figure (b), ρ is a ten-, twenty- or fifty-qubit Ising model ground state, and the class of measurements $\mathcal{M}$ is the set of all nearest-neighbour two-qubit Pauli-basis measurements. In all cases, the subset $\mathcal{M}_1$ contains $n_1 = 30$ measurements randomly chosen from $\mathcal{M}$. Given outcome probability distributions for all $\boldsymbol{m} \in \mathcal{M}_1$, the boxes show the classical fidelities of predicted outcome probabilities, averaged over all query measurements in $\mathcal{M}\setminus\mathcal{M}_1$. In the left-hand figure, the red boxes denotes the performance of GQNQ, and in comparison, the blue boxes show the classical fidelities of predicted outcomes of an alternative model, where the same measurement data is fed into the generation network without going through the representation network.

We also test our model for continuous-variable quantum states, encoded in harmonic oscillators. Our examples of continuous-variable states include Gaussian states, cat states, and GKP states, the latter two of which are both non-Gaussian states. Those Gaussian states we study are single-mode squeezed thermal states with thermal variance $V \in [1, 2]$, and squeezing parameter s satisfying $|s| \in [0, 0.5]$, $\arg(s) \in [0, \pi]$. The cat states are superpositions of coherent states with opposite amplitudes $|\alpha, \phi\rangle_{\text{cat}} := \frac{1}{\sqrt{\mathcal{N}}}(|\alpha\rangle + \mathrm{e}^{\mathrm{i}\phi}|-\alpha\rangle)$, where $\mathcal{N} = 2(1 + \mathrm{e}^{-|\alpha|^2}\cos\phi)$, $|\alpha| \in [1, 3]$ and $\phi \in \{0, \frac{\pi}{8}, \ldots, \pi\}$. The GKP states are superpositions of displaced squeezed states

$$|\epsilon, \theta, \phi\rangle_{\text{gkp}} := \mathrm{e}^{-\epsilon\hat{n}}\left(\cos\theta\,|0\rangle_{\text{gkp}} + \mathrm{e}^{\mathrm{i}\phi}\sin\theta\,|1\rangle_{\text{gkp}}\right)$$

where $\epsilon \in [0.05, 0.2]$, $\theta \in [0, 2\pi)$, $\phi \in [0, \pi]$, and $|0\rangle_{\text{gkp}}$ and $|1\rangle_{\text{gkp}}$ are ideal GKP states [33].

The measurements for continuous-variable states are homodyne measurements $\{dx\,|x, \theta\rangle\langle x, \theta|\}_{x\in\mathbb{R}}$, where $|x, \theta\rangle$ is an eigenstate of quadrature operator $\cos\theta\hat{q} + \sin\theta\hat{p}$ with eigenvalue x, and $\hat{q}$ and $\hat{p}$ are quantized position and momentum operators. Given measurement outcome probability distributions with respect to n_1 random quadrature phases, GQNQ predicts the outcome probability distributions at the other $300 - n_1$ quadrature bases, which, together with the previous n_1 quadrature phases, are evenly distributed among $[0, 2\pi)$. We truncate outcome x at ± 6 and project x onto one of 100 bins within $[-6, 6]$ with equal width.

GQNQ can be trained either over the measurement data of one type of the states mentioned above, or the union of the measurement data of more than one type of states without any type label. We test GQNQ using the measurement data of quantum states chosen from any types of states in the training data set. The averaged classical fidelity between the predicted outcome probability distributions and the real distributions are shown in Table I and an example of comparison between predictions and ground truths for a cat state is plotted in Fig. 3. Although the prediction for multiple types of quantum states using one neural network is harder than predicting one particular type of states, it can be seen that GQNQ can predict outcome statistics for different types of states with high classical fidelity, after being trained over the measurement data of multiple types of quantum states without labels.

GQNQ can also be applied to online learning of quantum states. In each round of online learning, newly collected measurement data are fed into the representation network to update the state representation, and the updated representation is fed into the generation network to update the predictions. Hence, the state learner neither needs to store measurement data during data collection, nor to perform time-consuming processing on measurement outcomes. Fig. 4 shows that the prediction error keeps decreasing as more measurement data are collected.

We further use GQNQ to mitigate stochastic errors in measurement data introduced by finite number of measurement samples. This error can be approximately simulated by adding a Gaussian noise at each outcome probability. We train GQNQ using both noiseless and noisy data and test the trained GQNQ on noisy measurement data of continuous-variable quantum states, as shown in Table I.

Table I. Average classical fidelity between predicted outcome statistics and real outcome statistics, averaged over all the test states and random query measurements. The five rows correspond to five different scenarios, where GQNQ is trained and tested over measurement data of five sets of states containing i) Gaussian states alone, ii) cat states alone, iii) the same number of Gaussian states and cat states, iv) GKP states alone, and v) the same number of cat states, GKP states and Gaussian states, respectively. In the first column, the measurement data provided are noiseless. In contrast, in the second and the third columns, the measurement data provided are noisy, where the variances of added Gaussian noise are 0.01 and 0.05, respectively. Although the provided data are noisy, the classical fidelity of predictions are evaluated with respect to the noiseless measurement outcome statistics. The average values of the average classical fidelity over 100 random choices of $\mathcal{M}_1$ are listed in the table. The average fidelitiy of predictions are not given for noisy cases in the scenarios of more than one type of states.

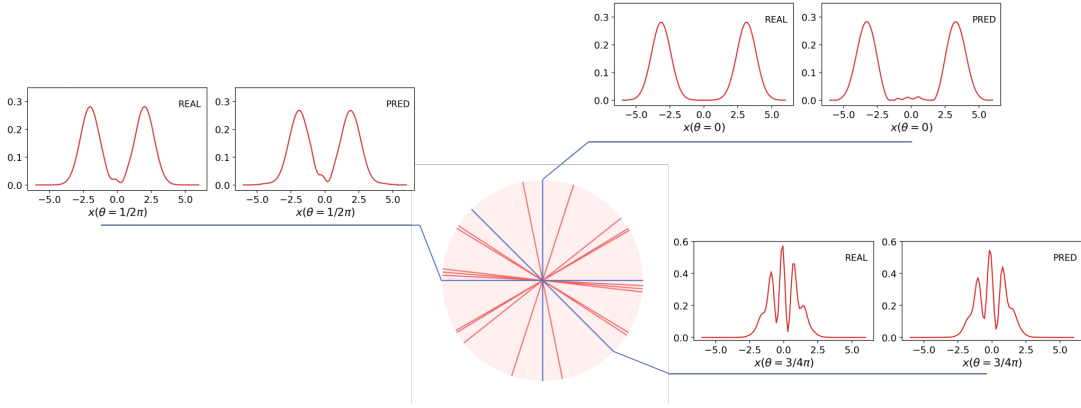|	Noiseless	Var(noise) = 0.01	Var(noise) = 0.05
Gaussian	0.9973	0.9969	0.9874
Cat	0.9827	0.9800	0.9706
Cat+Gaussian	0.9878	–	–
GKP	0.9762	0.9730	0.9586
Cat+Gaussian+GKP	0.9658	–	–

Figure 3. The real outcome probability density (left) and the predicted probability density (right) for cat state $|2.22 + 1.41i, \pi/4\rangle_{\text{cat}}$ at quadrature phases $\theta = 0$, $\theta = \pi/2$ and $\theta = 3\pi/4$, respectively, given the measurement outcome densities at ten random quadrature phases. In the middle circle, ten red lines passing through the center represent those quadrature phases at which measurement outcome statistics are known, and three blue lines passing through the center represent those quadrature phases at which measurement outcome statistics are to be predicted.
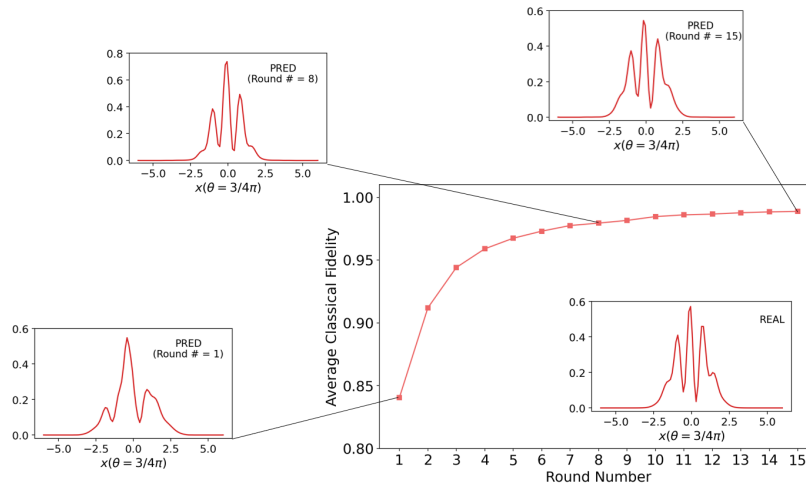

Figure 4. Average classical fidelity between the prediction and the ground truth increases when the number of applied POVMs increases. Each point is the average of classical fidelities of 1215 different cat states over all the possible query POVM measurements. Real outcome statistics and predicted outcome statistics at quadrature phase $\theta = 3/4\pi$ for an example cat state are plotted.
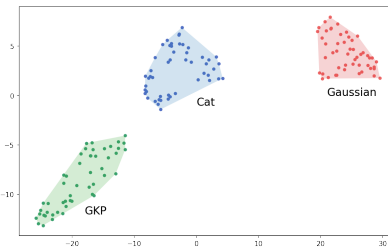
Figure 5. Clustering of state representations. It shows two-dimensional t-SNE embeddings of representations of Gaussian states, cat states and GKP states.

D. Using Representations to Cluster and to Predict Physical Properties

The state representation r constructed in the last subsection contains key information about the associated quantum state ρ, and can be used to perform other downstream tasks beyond the prediction of outcome probability distributions for unmeasured POVMs. To demonstrate this, we keep the representation network after training, and exploit an unsupervised algorithm or train another neural network in a supervised way to perform other tasks using the state representations.

One task is unsupervised clustering of state representations, which is to group state representations into multiple sets such that quantum states belonging to one class fall into one set. We investigate clustering the representations r of different types of continuous-variable states. Those high-dimensional representation vectors, each of which corresponds to one state, are projected onto a two-dimensional plane by the t-distributed stochastic neighbor embedding (t-SNE) [37] and grouped into three sets as shown in Fig. 5. The figure shows a clear separation between the three clusters.

The state representation r can be further used to predict other physical properties in a supervised model by adding additional labels. One example is to predict quantum phases of ground states of Ising models. We train an additional neural network to learn whether a ground state falls into the ordered phase or the disordered phase from its state representation r. The prediction reaches a success rate of 100%, 100% and 99% for ten-qubit, twenty-qubit and fifty-qubit ground states in our test sets, respectively.

III. DISCUSSION

Many works have been done on utilizing machine learning for the purpose of quantum state reconstruction [5, 16, 18, 20–25, 38]. In these works, a neural network is trained to reconstruct a particular experimental state, and hence, the trained network cannot be directly transferred to the reconstruction of a different experimental state. In contrast, our GQNQ model is not restricted to the measurement data of one particular quantum state, nor a certain class of quantum states. We train GQNQ on the measurement data across multiple different classes of states without labels, and the trained GQNQ predicts the outcome statistics of random measurements for different quantum states with high accuracy, implying that GQNQ is capable of learning the features of different classes of states autonomously. As our training phase uses simulated data and no further training is required thereafter, GQNQ is significantly faster than any other data-driven approach to the reconstruction of quantum state in real time from experimental data. Besides, GQNQ does not need the actual description of measurements in terms of POVMs. Instead, it just uses a parametrization, which can be considered as an encrypted description of measurements. This point can be important for the purpose of information security, in a similar way as in blind quantum computing [39]. Here, the use of an encryted description can guarantee that GQNQ does not have all the information of the quantum state when successfully making predictions of its measurement outcomes.

The key idea of GQNQ is that, rather than feeding raw measurement data directly into a generation network, we use a representation network to learn a concise and abstract description of a quantum state from its measurement data before feeding it into a generation network that predicts outcome statistics for other measurements. Then the combination of representation network and generation network is trained jointly such that the learned representation contains all necessary information to accurately predict outcome statistics for any measurement. Our model is, conceptually, closer to "pretty good tomography" [11]. In pretty good tomography, a hypothesis state consistent with expectations of few observables, is a good approximation of the true state if the expectations of almost all other observables, chosen from an ensemble decided in advance, can be successfully predicted using this hypothesis state with high probability. Instead of trying to find such a hypothesis state, in GQNQ, the representation network searches for an abstract representation of the state from the measurement data of very few measurements.

We have tested GQNQ with different multi-qubit states and continuous-variable states. The numerical experiments include ground states of Ising models, which are significant for many-body quantum simulation [40–42], cat states and GKP states, which are both important for realizing Bosonic quantum error correcting codes [33, 43]. Furthermore, we find that our state learning model can mitigate stochastic noises introduced by finite number of measurements, making it suitable for practical application of learning quantum states from noisy data.

We also demonstrate that the learned state representations retain essential features of the corresponding quantum states, by showing that these representations can be used for other downstream tasks, for example, clustering of quantum states and prediction of physical properties, apart from generating outcome statistics. Noticeably, the dimensions of the representations are set to 32 for the ground states of Ising model, and 16 for continuous-variable quantum states, both of which are much smaller than the degrees of freedom in reconstructed density matrices [21] and restricted Boltzmann machines [20].

In contrast to unsupervised learning approaches to the reconstruction of quantum states [16–18], our learning method requires sufficient training data coming from either classical simulations or quantum experiments. Although all the data in our examples come from classical simulations, a experimenter, who possesses a quantum device, can utilize our network to learn experimental quantum states from real measurement data. One interesting open question is whether one can reduce the size of training data while keeping the generality and flexibility of the neural network.

IV. METHODS

Our neural networks are implemented by the pytorch [44] framework and trained on four NVIDIA GeForce GTX 1080 Ti GPUs. We exploit the Adam optimizer [45] and batch gradient descent [46] technique to perform the training. The training time is less than two hours for each task discussed in this paper. The ground states of the Hamiltonian of one-dimensional Ising models utilized in our numerical experiments are solved by the exact method for the scenario of $L = 6$ and are approximately solved by density-matrix renormalization group (DMRG) [47] for the scenario of $L = 10, 20$ and 50. The data of continuous-variable quantum states are generated by simulation tools provided in Strawberry Fields [48].

We present the whole training procedure by pseudocode in Algorithm 1 following the notations introduced in Sec. I.

Algorithm 1: Training of generative query network for quantum state learning.

Data: number of states in training set N, state measurement results $\{\{\boldsymbol{m}_i, \boldsymbol{v}_i^k\}_{i=1}^n\}_{k=1}^N$, maximum number of known
POVM measurement results for each state a, maximum number of epochs E, learning rate δ, batch size B.

Initialize parameters $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ randomly, $e = 0$;

while $e < E$ **do**

  $\mathcal{L} = 0$;

  for $k = 1$ **to** N **do**

    Generate a random integer number n_1 from $[1, a]$;

    Randomly select n_1 pairs of $\{\boldsymbol{m}_i, \boldsymbol{v}_i\}$ from $\{\boldsymbol{m}_i, \boldsymbol{v}_i^k\}_{i=1}^n$ and denote them as $\{\boldsymbol{m}_{i_j}, \boldsymbol{v}_{i_j}^k\}_{j=1}^{n_1}$, where $\{i_j\}_{j=1}^n$ is a
permutation of $\{1, \ldots, n\}$;

    Input each of $\{\boldsymbol{m}_{i_j}, \boldsymbol{v}_{i_j}^k\}_{j=1}^{n_1}$ into the representation network $f_{\boldsymbol{\xi}}$ to obtain the representations $\{\boldsymbol{r}_{i_j}\}_{j=1}^{n_1}$ as
$\boldsymbol{r}_{i_j} = f_{\boldsymbol{\xi}}(\boldsymbol{m}_{i_j}, \boldsymbol{v}_{i_j}^k)$;

    Calculate the state representation by an aggregate function $\mathcal{A}$ as $\boldsymbol{r} = \mathcal{A}(\{\boldsymbol{r}_{i_j}\}_{j=1}^{n_1})$;

    Input $\boldsymbol{r}$ and the remaining $\{\boldsymbol{m}_{i_j}\}_{j=n_1+1}^n$ into the generation network $g_{\boldsymbol{\eta}}$ to obtain the predictions $\{\boldsymbol{v}_{i_j}'^k\}_{j=n_1+1}^n$ of
measurement outcome distributions as $\boldsymbol{v}_{i_j}'^k = g_{\boldsymbol{\eta}}(\boldsymbol{r}, \boldsymbol{m}_{i_j})$;

    Calculate the loss l by comparing $\{\boldsymbol{v}_{i_j}'^k\}_{j=n_1+1}^n$ with $\{\boldsymbol{v}_{i_j}^k\}_{j=n_1+1}^n$ and update $\mathcal{L}$ as $\mathcal{L} = \mathcal{L} + l$;

    if $k \bmod B = 0$ **then**

      Calculate $\nabla_{\boldsymbol{\xi}} \mathcal{L}$ and $\nabla_{\boldsymbol{\eta}} \mathcal{L}$;

      Update $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ as $\boldsymbol{\xi} = \boldsymbol{\xi} - \delta \nabla_{\boldsymbol{\xi}} \mathcal{L}$, $\boldsymbol{\eta} = \boldsymbol{\eta} - \delta \nabla_{\boldsymbol{\eta}} \mathcal{L}$;

      $\mathcal{L} = 0$;

  $e = e + 1$;

V.   ACKNOWLEDGEMENT

VI.   APPENDIX

A.   Implementation details of GQNQ

1.   Structure of GQNG

As shown in Fig. 6, our proposed Generative Query Network for quantum state learning (GQNG) is mainly composed of a representation network $f_{\boldsymbol{\xi}}$, an aggregate function $\mathcal{A}$ and a generation network $g_{\boldsymbol{\eta}}$.
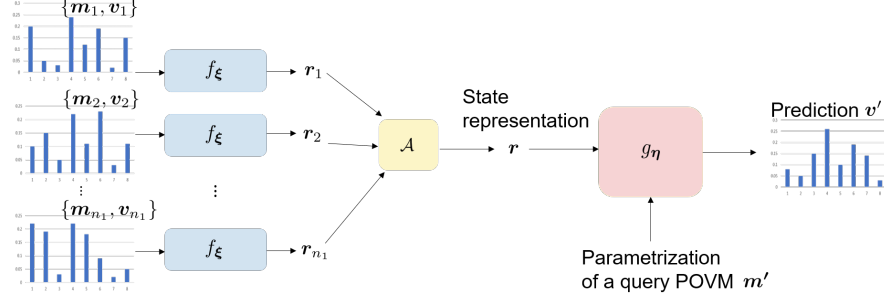

Figure 6. Structure of GQNQ.

The representation network $f_{\boldsymbol{\xi}}$ consists of multiple dense layers [49], also called full-connected layers and we depict its structure in Fig. 7. $\boldsymbol{\xi}$ contains trainable parameters of all layers. The input of the representation network is a pair of $(\boldsymbol{m}_i, \boldsymbol{v}_i)$, where $\boldsymbol{m}_i$ is parameterization of a POVM measurement and $\boldsymbol{v}_i$ is its corresponding measurement outcome probabilities. The output $\boldsymbol{r}_i$ can be regarded as an abstract representation of $\{\boldsymbol{m}_i, \boldsymbol{v}_i\}$. Here, for simplicity, we just use the average function $\boldsymbol{r} := \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{r}_i$ as the aggregate function but we believe other more sophisticated architecture such as recurrent neural network [49] may achieve better performance.
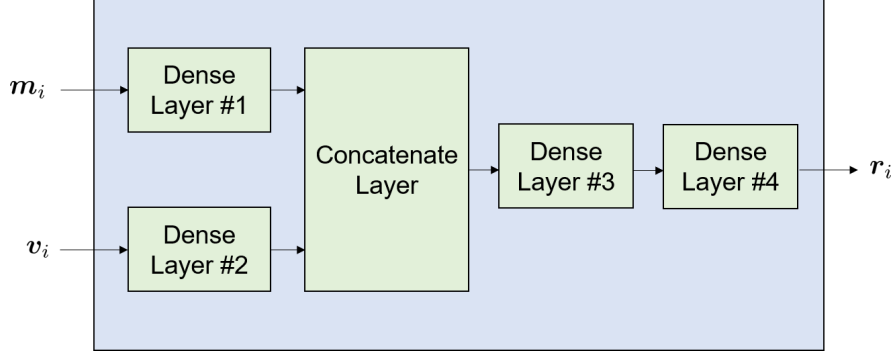

Figure 7. Structure of the representation network.

The generation network $g_{\boldsymbol{\eta}}$ is special because its structure is different in the training and test phase. Here $\boldsymbol{\eta}$ contains all of trainable parameters in the generation network. In the test phase, the generation network consists of two dense layers and one Long short-term memory (LSTM) cell [50], and we depict its structure in Fig. 8. The input of this generation network is the state representation $\boldsymbol{r}$ and the parameterization $\boldsymbol{m}'$ of a query POVM measurement and the output $\mathcal{N}'$ is a distribution of the prediction $\boldsymbol{v}'$ of measurement outcome probabilities corresponding to $\boldsymbol{m}'$. $\boldsymbol{h}_0$, $\boldsymbol{c}_0$ and $\boldsymbol{u}_0$ are some internal parameters, all of which are initialized as zero tensors. As we can see, the generation network execute the fist dense layer and the LSTM cell for L times while $\boldsymbol{m}'$ and $\boldsymbol{r}$ are injected to the network for each time. It is worth mentioning that $\boldsymbol{z}_i$ $(i \in \mathbb{N}, i < L)$ can be viewed as a hidden variable obeys a prior Gaussian distribution $\mathcal{N}_i$ generated by the first dense layer from $\boldsymbol{h}_i$. In Fig. 9, we depict the generation network exploited in the training. Here another extra input $\boldsymbol{v}'_{real}$ is available because we know the real outcome probabilities in the training phase. Furthermore, we utilize another LSTM cell and another dense layer to generate a posterior distribution of the hidden variable $\boldsymbol{z}_i$ rather than sampling $\boldsymbol{z}_i$ from a prior distribution. The advantage of such design is that we can make good use of the information of $\boldsymbol{v}'_{real}$ to obtain better $\boldsymbol{z}_i$ during the generation. $\boldsymbol{h}_0^1$, $\boldsymbol{c}_0^1$, $\boldsymbol{h}_0^2$, $\boldsymbol{c}_0^2$ and $\boldsymbol{u}_0$ are some internal parameters, all of which are initialized as zero tensors.
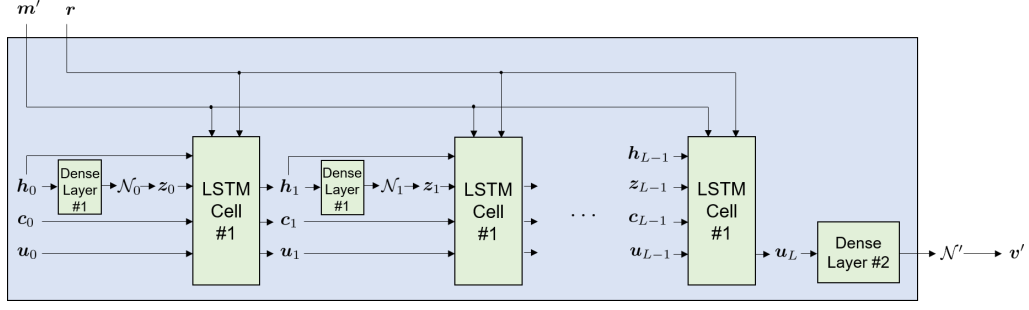
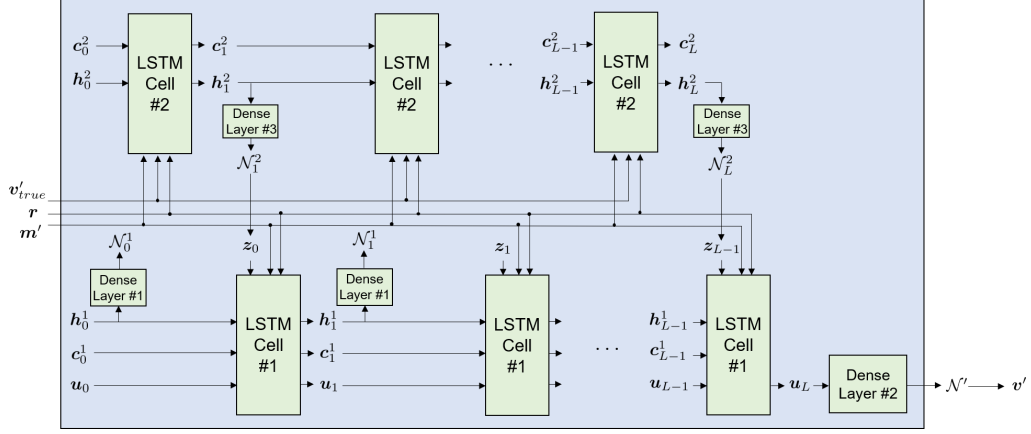Figure 8. Structure of the generation network in the test.

Figure 9. Structure of the generation network in the training.

2. Training of GQNQ

We define the loss function $\mathcal{L}$ of the training in Eqn. 2.

$$\mathcal{L}(\boldsymbol{\xi}, \boldsymbol{\eta}) = \mathbb{E}[-\ln \mathcal{N}'(\boldsymbol{v}'_{real}) + \sum_{j=0}^{L-1} \mathrm{KL}(\mathcal{N}_j^1, \mathcal{N}_{j+1}^2)], \tag{2}$$

where KL represents KL divergence [51] of two Gaussian distributions. The first term of this loss function can be interpreted as the reconstruction loss, which can guide the model to acquire more accurate predictions. The second term is a regularization term utilized for seeking a better prior distribution of the hidden variable $\boldsymbol{z}_i$ in the generation process, which is constructive to improve accuracy of the predictions further.

We adopt batch gradient descent [46] and the Adam optimizer [45] to minimize this loss function in the training. The batch size is set to 10 or 20 in all of our experiments and the learning rate decreases gradually with the increase of the number of training epochs.

B. Additional Experiments

In addition to the numerical experiments described in the main body of the paper, we also perform other experiments and we will present more results in this section.

For 6-qubit quantum states, we also consider other two classes of quantum states aside from ground states of the Ising model we introduced in the main body of this paper. One is 6-qubit Greenberger–Horne–Zeilinger (GHZ) states with local rotation unitaries defined as $|\psi\rangle = U_A \otimes U_B \otimes U_C |GHZ\rangle$ and the other is 6-qubit W states local rotation unitaries defined as $|\psi\rangle = U_A \otimes U_B \otimes U_C |W\rangle$, where U_A, U_B and U_C are single-qubit rotations. we consider all possible six-qubit Pauli basis measurements here and trained and tested over measurement data of three sets of states containing i) GHZ states alone, ii) the same number of GHZ states and W states, iii) the same number of GHZ states,

W states and ground states of 6-qubit Ising model. We limit the range of the rotation angle to $[0, \frac{\pi}{5}]$ for GHZ states and W states. The average classical fidelity of predictions with respect to the ground truth are plotted in Fig. 10.
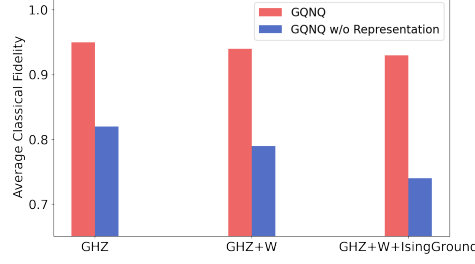

Figure 10. Classical fidelity between predicted outcome statistics and genuine outcome statistics, averaged over all the test states and query Pauli bases, given measurement statistics at 10 to 30 random Pauli bases, for different classes of training and testing states. The three different cases are when GQNQ and GQNQ w/o Representation are trained and tested only over measurement data of locally rotated GHZ states, across equal-sized measurement dataset of locally rotated GHZ states and locally rotated W states, and across equal-sized measurement dataset of locally rotated GHZ states, locally rotated W states, and the ground states of Ising model. The deviation introduced by random selection of POVMs in all cases is negligible with the largest deviation during 100 repetitions of numerical experiments less than 0.2%.

As we mentioned in the main body of the paper, the state representation r can be utilized to cluster different type of states. We do the same thing for three classes of states discussed here and plot the results in Fig. 11.
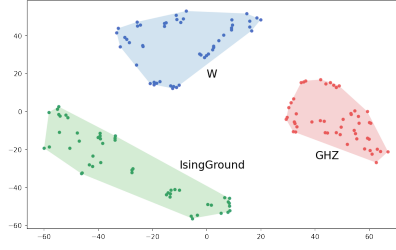

Figure 11. Clustering of state representations. Fig shows two-dimensional t-SNE embeddings of the representations of GHZ states, W states and ground states of Ising model, which are clearly separated and are grouped, respectively.

For the numerical experiments for the learning of continuous-variable quantum states, we provide another example of comparison between predictions and ground truths for a GKP state in Fig. 12 here.
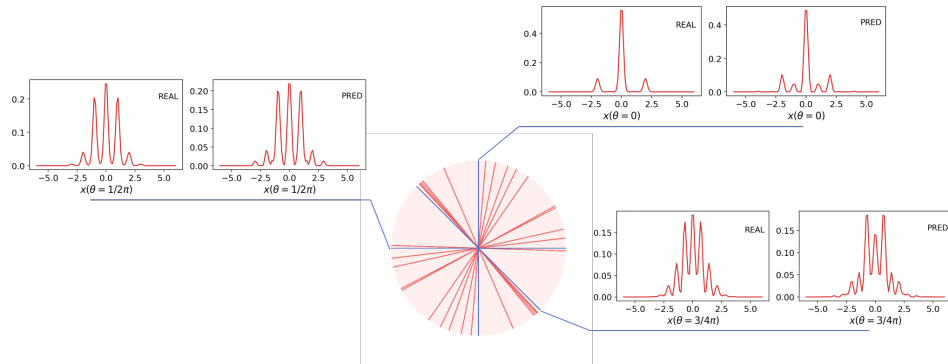

Figure 12. The real outcome probability density (left) and the predicted probability density (right) for gkp state $|0.15, \pi/24, 7\pi/8\rangle_{\mathrm{gkp}}$ at quadrature phases $\theta = 0$, $\theta = \pi/2$ and $\theta = 3\pi/4$, respectively, given the measurement outcome densities at ten random quadrature phases. In the middle circle, fifteen red lines passing through the center represent those quadrature phases at which measurement outcome statistics are known, and three blue lines passing through the center represent those quadrature phases at which measurement outcome statistics are to be predicted.

[1] Jens Eisert, Dominik Hangleiter, Nathan Walk, Ingo Roth, Damian Markham, Rhea Parekh, Ulysse Chabaud, and Elham Kashefi, "Quantum certification and benchmarking," Nat. Rev. Phys. **2**, 382–390 (2020).

[2] G. Tóth, W. Wieczorek, D. Gross, R. Krischek, C. Schwemmer, and H. Weinfurter, "Permutationally invariant quantum tomography," Phys. Rev. Lett. **105**, 250403 (2010).

[3] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert, "Quantum state tomography via compressed sensing," Phys. Rev. Lett. **105**, 150401 (2010).

[4] Marcus Cramer, Martin B Plenio, Steven T Flammia, Rolando Somma, David Gross, Stephen D Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu, "Efficient quantum state tomography," Nat. Commun. **1**, 1–7 (2010).

[5] BP Lanyon, C Maier, Milan Holzäpfel, Tillmann Baumgratz, C Hempel, P Jurcevic, Ish Dhand, AS Buyskikh, AJ Daley, Marcus Cramer, *et al.*, "Efficient tomography of a quantum many-body system," Nat. Phys. **13**, 1158–1162 (2017).

[6] Jordan Cotler and Frank Wilczek, "Quantum overlapping tomography," Phys. Rev. Lett. **124**, 100401 (2020).

[7] Hsin-Yuan Huang, Richard Kueng, and John Preskill, "Predicting many properties of a quantum system from very few measurements," Nat. Phys. **16**, 1050–1057 (2020).

[8] Hsin-Yuan Huang, Richard Kueng, Giacomo Torlai, Victor V Albert, and John Preskill, "Provably efficient machine learning for quantum many-body problems," arXiv:2106.12627 (2021).

[9] Steven T. Flammia and Yi-Kai Liu, "Direct fidelity estimation from few pauli measurements," Phys. Rev. Lett. **106**, 230501 (2011).

[10] Marcus P. da Silva, Olivier Landon-Cardinal, and David Poulin, "Practical characterization of quantum devices without tomography," Phys. Rev. Lett. **107**, 210404 (2011).

[11] Scott Aaronson, "The learnability of quantum states," Proc. R. Soc. A **463**, 3089–3114 (2007).

[12] Scott Aaronson, "Shadow tomography of quantum states," SIAM J. Comput. **49**, STOC18–368 (2019).

[13] Scott Aaronson, Xinyi Chen, Elad Hazan, Satyen Kale, and Ashwin Nayak, "Online learning of quantum states," J. Stat. Mech.: Theory Exp. **2019**, 124019 (2019).

[14] Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová, "Machine learning and the physical sciences," Rev. Mod. Phys. **91**, 045002 (2019).

[15] Qian Xu and Shuqi Xu, "Neural network state estimation for full quantum state tomography," arXiv preprint arXiv:1811.06654 (2018).

[16] Giacomo Torlai and Roger G. Melko, "Latent space purification via neural density operators," Phys. Rev. Lett. **120**, 240503 (2018).

[17] Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo, "Neural-network quantum state tomography," Nat. Phys. **14**, 447–450 (2018).

[18] Juan Carrasquilla, Giacomo Torlai, Roger G Melko, and Leandro Aolita, "Reconstructing quantum states with generative models," Nat. Mach. Intell. **1**, 155–161 (2019).

[19] Raban Iten, Tony Metger, Henrik Wilming, Lídia del Rio, and Renato Renner, "Discovering physical concepts with neural networks," Phys. Rev. Lett. **124**, 010508 (2020).

[20] Egor S Tiunov, VV Tiunova, Alexander E Ulanov, AI Lvovsky, and Aleksey K Fedorov, "Experimental quantum homodyne tomography via machine learning," Optica **7**, 448–454 (2020).

[21] Shahnawaz Ahmed, Carlos Sánchez Muñoz, Franco Nori, and Anton Frisk Kockum, "Quantum state tomography with conditional generative adversarial networks," Phys. Rev. Lett. **127**, 140502 (2021).

[22] Shahnawaz Ahmed, Carlos Sánchez Muñoz, Franco Nori, and Anton Frisk Kockum, "Classification and reconstruction of optical quantum states with deep neural networks," Phys. Rev. Res. **3**, 033278 (2021).

[23] Andrea Rocchetto, Edward Grant, Sergii Strelchuk, Giuseppe Carleo, and Simone Severini, "Learning hard quantum distributions with variational autoencoders," NPJ Quantum Inf. **4**, 28 (2018).

[24] Yihui Quek, Stanislav Fort, and Hui Khoon Ng, "Adaptive quantum state tomography with neural networks," NPJ Quantum Inf. **7**, 105 (2021).

[25] Adriano Macarone Palmieri, Egor Kovlakov, Federico Bianchi, Dmitry Yudin, Stanislav Straupe, Jacob D Biamonte, and Sergei Kulik, "Experimental neural network enhanced quantum tomography," NPJ Quantum Inf. **6**, 20 (2020).

[26] Alistair W. R. Smith, Johnnie Gray, and M. S. Kim, "Efficient quantum state sample tomography with basis-dependent neural networks," PRX Quantum **2**, 020348 (2021).

[27] Gael Sentís, Alex Monràs, Ramon Muñoz Tapia, John Calsamiglia, and Emilio Bagan, "Unsupervised classification of quantum data," Phys. Rev. X **9**, 041029 (2019).

[28] Jun Gao, Lu-Feng Qiao, Zhi-Qiang Jiao, Yue-Chi Ma, Cheng-Qiu Hu, Ruo-Jing Ren, Ai-Lin Yang, Hao Tang, Man-Hong Yung, and Xian-Min Jin, "Experimental machine learning of quantum states," Phys. Rev. Lett. **120**, 240501 (2018).

[29] Andreas Elben, Benoît Vermersch, Rick van Bijnen, Christian Kokail, Tiff Brydges, Christine Maier, Manoj K. Joshi, Rainer Blatt, Christian F. Roos, and Peter Zoller, "Cross-platform verification of intermediate scale quantum devices," Phys. Rev. Lett. **124**, 010504 (2020).

[30] SM Ali Eslami, Danilo Jimenez Rezende, Frederic Besse, Fabio Viola, Ari S Morcos, Marta Garnelo, Avraham Ruderman, Andrei A Rusu, Ivo Danihelka, Karol Gregor, *et al.*, "Neural scene representation and rendering," Science **360**, 1204–1210 (2018).

[31] Yong Siah Teo, Huangjun Zhu, Berthold-Georg Englert, Jaroslav Řeháček, and Zden ĕk Hradil, "Quantum-state reconstruction by maximizing likelihood and entropy," Phys. Rev. Lett. **107**, 020404 (2011).

[32] Vladimír Bužek and Peter L Knight, "I: Quantum interference, superposition states of light, and nonclassical effects," in *Progress in optics*, Vol. 34 (Elsevier, 1995) pp. 1–158.

[33] Daniel Gottesman, Alexei Kitaev, and John Preskill, "Encoding a qubit in an oscillator," Phys. Rev. A **64**, 012310 (2001).

[34] Srinivasan Arunachalam, Alex B Grilo, and Henry Yuen, "Quantum statistical query learning," arXiv:2002.08240 (2020).

[35] David M Blei, Alp Kucukelbir, and Jon D McAuliffe, "Variational inference: A review for statisticians," J. Am. Stat. Assoc. **112**, 859–877 (2017).

[36] David F Williamson, Robert A Parker, and Juliette S Kendrick, "The box plot: a simple visual method to interpret data," Annals of internal medicine **110**, 916–921 (1989).

[37] Laurens Van der Maaten and Geoffrey Hinton, "Visualizing data using t-sne." Journal of machine learning research **9** (2008).

[38] Tao Xin, Sirui Lu, Ningping Cao, Galit Anikeeva, Dawei Lu, Jun Li, Guilu Long, and Bei Zeng, "Local-measurement-based quantum state tomography via neural networks," NPJ Quantum Inf. **5**, 109 (2019).

[39] Stefanie Barz, Elham Kashefi, Anne Broadbent, Joseph F Fitzsimons, Anton Zeilinger, and Philip Walther, "Demonstration of blind quantum computing," science **335**, 303–308 (2012).

[40] Axel Friedenauer, Hector Schmitz, Jan Tibor Glueckert, Diego Porras, and Tobias Schätz, "Simulating a quantum magnet with trapped ions," Nat. Phys. **4**, 757–761 (2008).

[41] Kihwan Kim, M-S Chang, Simcha Korenblit, Rajibul Islam, Emily E Edwards, James K Freericks, G-D Lin, L-M Duan, and Christopher Monroe, "Quantum simulation of frustrated ising spins with trapped ions," Nature **465**, 590–593 (2010).

[42] R Islam, EE Edwards, K Kim, S Korenblit, C Noh, H Carmichael, G-D Lin, L-M Duan, C-C Joseph Wang, JK Freericks, *et al.*, "Onset of a quantum phase transition with a trapped ion quantum simulator," Nat. Commun. **2**, 1–6 (2011).

[43] Victor V. Albert, Kyungjoo Noh, Kasper Duivenvoorden, Dylan J. Young, R. T. Brierley, Philip Reinhold, Christophe Vuillot, Linshu Li, Chao Shen, S. M. Girvin, Barbara M. Terhal, and Liang Jiang, "Performance and structure of single-mode bosonic codes," Phys. Rev. A **97**, 032346 (2018).

[44] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, *et al.*, "Pytorch: An imperative style, high-performance deep learning library," Advances in neural information processing systems **32**, 8026–8037 (2019).

[45] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).

[46] Sebastian Ruder, "An overview of gradient descent optimization algorithms," arXiv preprint arXiv:1609.04747 (2016).

[47] Ulrich Schollwöck, "The density-matrix renormalization group," Reviews of modern physics **77**, 259 (2005).

[48] Nathan Killoran, Josh Izaac, Nicolás Quesada, Ville Bergholm, Matthew Amy, and Christian Weedbrook, "Strawberry fields: A software platform for photonic quantum computing," Quantum **3**, 129 (2019).

[49] Charu C Aggarwal *et al.*, "Neural networks and deep learning," Springer **10**, 978–3 (2018).

[50] Sepp Hochreiter and Jürgen Schmidhuber, "Long short-term memory," Neural computation **9**, 1735–1780 (1997).

[51] Solomon Kullback, *Information theory and statistics* (Courier Corporation, 1997).