

Quantum state estimation based on deep learning

Haowen Xiao(肖皓文) and Zhiguang Han(韩枝光)

Citation: Chin. Phys. B, 2024, 33 (**12**): 120307. DOI: 10.1088/1674-1056/ad78d7
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

In-situ deposited anti-aging TiN capping layer for Nb superconducting quantum circuits

Hao-Ran Tao(陶浩然), Lei Du(杜磊), Liang-Liang Guo(郭亮亮), Yong Chen(陈勇), Hai-Feng Zhang(张海峰), Xiao-Yan Yang(杨小燕), Guo-Liang Xu(徐国良), Chi Zhang(张 驰), Zhi-Long Jia(贾志龙), Peng Duan(段鹏), and Guo-Ping Guo(郭国平)

Chin. Phys. B, 2024, 33 (**9**): 090310. DOI: 10.1088/1674-1056/ad6a3c

Variational data encoding and correlations in quantum-enhanced machine learning

Ming-Hao Wang(王明浩) and Hua Lü(吕桦)

Chin. Phys. B, 2024, 33 (**9**): 090307. DOI: 10.1088/1674-1056/ad5c3b

Quantum block coherence with respect to projective measurements

Pu Wang(王璞), Zhong-Yan Li(李忠艳), and Hui-Xian Meng(孟会贤)

Chin. Phys. B, 2024, 33 (**8**): 080308. DOI: 10.1088/1674-1056/ad50c1

General multi-attack detection for continuous-variable quantum key distribution with local local oscillator

Zhuo Kang(康茁), Wei-Qi Liu(刘维琪), Jin Qi(齐锦), and Chen He(贺晨)

Chin. Phys. B, 2024, 33 (**5**): 050308. DOI: 10.1088/1674-1056/ad2bf2

Fringe removal algorithms for atomic absorption images: A survey

Gaoyi Lei(雷高益), Chencheng Tang(唐陈成), and Yueyang Zhai(翟跃阳)

Chin. Phys. B, 2022, 31 (**5**): 050313. DOI: 10.1088/1674-1056/ac3758

Quantum state estimation based on deep learning

Haowen Xiao(肖皓文) and Zhiguang Han(韩枝光)†

School of Information and Communication Engineering, Hainan University, Haikou 570228, *China*

(Received 14 June 2024; revised manuscript received 28 August 2024; accepted manuscript online 10 September 2024)

We used deep learning techniques to construct various models for reconstructing quantum states from a given set of coincidence measurements. Through simulations, we have demonstrated that our approach generates functionally equivalent reconstructed states for a wide range of pure and mixed input states. Compared with traditional methods, our system offers the advantage of faster speed. Additionally, by training our system with measurement results containing simulated noise sources, the system shows a significant improvement in average fidelity compared with typical reconstruction methods. We also found that constraining the variational manifold to physical states, i.e., positive semi-definite density matrices, greatly enhances the quality of the reconstructed states in the presence of experimental imperfections and noise. Finally, we validated the correctness and superiority of our model by using data generated on IBM Quantum Platform, a real quantum computer.

Keywords: deep learning, quantum state estimation, IBM quantum processor

PACS: 03.67.–a, 03.67.Lx **DOI:** 10.1088/1674-1056/ad78d7

1. Introduction

Quantum state estimation is the process of determining the state of an unknown quantum system through a series of measurements. For multi-qubit quantum states, the number of measurement settings required grows exponentially with the number of qubits.[1] Additionally, determining the physically valid density matrix corresponding to the measurement data itself requires a substantial amount of resources. Quantum state tomography (QST) was the earliest method used for quantum state estimation.[2] Smith implemented four-qubit quantum state estimation from noisy data using the least-squares method.[3] Mastriani proposed an optimal quantum state estimation method based on improved Kalman filtering.[4] Lukens *et al.* introduced an improved Bayesian quantum state estimation method.[5] In 1986, Rumelhart proposed the backpropagation learning algorithm for multilayer artificial neural networks, addressing the training issues of multilayer networks and leading to widespread applications of artificial neural networks in pattern recognition, automatic control, signal processing and other fields.[6] In the field of quantum state estimation, Xu demonstrated QST using neural networks and proved the superiority of using neural networks for quantum state estimation in terms of time complexity.[7] Tao *et al.* achieved QST from local measurement information using fully connected neural networks, providing an effective method for scalable state tomography of large-scale quantum systems.[8] Giacomo successfully implemented QST of entangled states with over a hundred qubits using neural networks and achieved high precision.

In this paper we first use traditional methods to measure the density matrix of quantum states. Then we try different neural networks to establish the mapping relationship between the measurement results of quantum states and the density matrix of quantum states through large-scale data training, in order to estimate quantum states. Different neural networks are designed first. Then, the input–output pairs for training and validation of the neural network are constructed based on the correspondence between the quantum density matrix and the system output measurement values. Subsequently, using the generated samples, supervised training of the designed neural network model is performed using backpropagation to obtain the weight parameters of the network estimator. Finally, any output value of a quantum system is input into the trained network estimator, and while obtaining the network output a normalization transformation is applied to obtain a density matrix that satisfies the properties of being positive semi-definite, having unit trace and being conjugate symmetric. The density matrix estimated by the neural network is then compared with the density matrix constructed by traditional QST to determine the optimal neural network estimation.

2. Standard QST method

Quantum estimation techniques, including those applied in optomechanics, provide a robust framework for enhancing the accuracy of rotational speed measurements.[9] The density matrix contains all the information about a quantum system. Therefore, by obtaining the density matrix of a system its properties can be fully understood. QST is a method for reconstructing the density matrix of a quantum state using statistical information.[10,11] In a 2^n-dimensional Hilbert space, the density matrix of any n-qubit system can be determined by 4^n-1 real parameters. Determining these 4^n-1 real parame-

†Corresponding author. E-mail: hanzhiguang@hainanu.edu.cn

© 2024 Chinese Physical Society and IOP Publishing Ltd

ters involves performing projection measurements on multiple copies of the same system.[12–14] Quantum bits can be realized in many quantum systems. This section focuses on the fundamental principles of standard quantum state reconstruction for qubit systems and its implementation on IBM quantum processors.

2.1. Single-qubit tomography

Single-qubit QST is a method for reconstructing the density matrix of a single-qubit system. The density matrix in such a system is defined by three real parameters for the diagonal elements and one complex parameter for the phase of the off-diagonal elements. By conducting a series of measurements in various bases, statistical data can be gathered to reconstruct the density matrix.

Single-qubit QST is an important technique in the field of quantum information that can be used to verify the state of quantum bits, assess the performance of quantum systems and play a crucial role in applications such as quantum computing and quantum communication.

Consider a single-qubit system whose state is fully described by a 2×2 density matrix. We select a set of measurement bases that satisfy the condition of information completeness for performing projection measurements. From the obtained statistical results, we estimate the values of the density matrix. Here, we choose to utilize the Pauli operators $(\sigma_1,\sigma_2,\sigma_3)$ as observable projectors for the measurements.[15,16] The relationship between the density matrix ρ of an n-qubit system and the Stokes parameters S_i is

$$\rho=\sum_{i_1,i_2,\ldots,i_n}^{n} S_{i_1,i_2,\ldots,i_n}\sigma_{i_1}\otimes\sigma_{i_2}\otimes\ldots\sigma_{i_n}, \tag{1}$$

$$S_{i_1,i_2,\ldots,i_n}=\mathrm{Tr}\left\{\rho\sigma_{i_1}\otimes\sigma_{i_2}\otimes\ldots\sigma_{i_{n-1}}\right\}. \tag{2}$$

According to the single-qubit system, we have

$$\begin{aligned}
\mathrm{Tr}\{\hat{\sigma_0}\hat{\rho}\} &= P_{|0\rangle}+P_{|1\rangle}=P_{|H\rangle}+P_{|V\rangle},\\
\mathrm{Tr}\{\hat{\sigma_1}\hat{\rho}\} &= P_{\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)}-P_{\frac{1}{\sqrt{2}}(|0\rangle-|1\rangle)}=P_{|D\rangle}-P_{|A\rangle},\\
\mathrm{Tr}\{\hat{\sigma_2}\hat{\rho}\} &= P_{\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)}+P_{\frac{1}{\sqrt{2}}(|0\rangle-|1\rangle)}=P_{|R\rangle}-P_{|L\rangle},\\
\mathrm{Tr}\{\hat{\sigma_3}\hat{\rho}\} &= P_{|0\rangle}-P_{|1\rangle}=P_{|H\rangle}-P_{|V\rangle}.
\end{aligned} \tag{3}$$

By measuring these probabilities, a density matrix can be constructed.

Measurement on a real quantum computer on IBM Quantum Platform (IBM Q) (defaults to the $\{|0\rangle,|1\rangle\}$ eigenstates of the Pauli operator σ_3. Therefore, it is necessary to apply appropriate quantum gates for transformation: adding a Hadamard gate (H gate) is equivalent to measuring in the eigenstates $\left\{\frac{|0\rangle+|1\rangle}{\sqrt{2}},\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right\}$ of the Pauli operator σ_1; adding $S^\dagger$ rotation gates and H gates is equivalent to measuring in the eigenstates $\left\{\frac{|0\rangle+\mathrm{i}|1\rangle}{\sqrt{2}},\frac{|0\rangle-\mathrm{i}|1\rangle}{\sqrt{2}}\right\}$ of the Pauli operator σ_2; not adding any gate is equivalent to measurement without applying any operator, equivalent to measuring in the eigenstates of σ_3.

The gates applied in multi-qubit system tomography are tensor products of the gates applied in single-qubit system tomography (Fig. 1). For example, on IBM Q default measurements are performed in the eigenstates $\{|00\rangle,|01\rangle,|10\rangle,|11\rangle\}$ of the Pauli operator $\sigma_3\otimes\sigma_3$. Applying the gate $H\otimes H$ to measure $\mathrm{Tr}(\sigma_1\rho)\sigma_1$, H gates are applied to each qubit, followed by measurement. The gate $S^\dagger\otimes H$ is applied to measure $\mathrm{Tr}(\sigma_2\rho)\sigma_2$, where $S^\dagger$ gates and H gates are applied to the first qubit, and an H gate is applied to the second qubit, followed by measurement. The gate $I\otimes H$ is applied to measure $\mathrm{Tr}(\sigma_3\rho)\sigma_3$, where no operation is applied to the first qubit and an H gate is applied to the second qubit, followed by measurement.

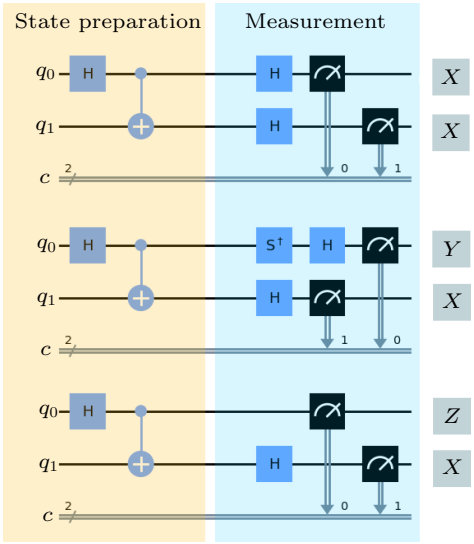


Fig. 1. The tomography process of a two-qubit system on IBM Q, with X, Y and Z representing the Pauli operators σ_1, σ_2 and σ_3.

2.2. Measurement on IBM Q

Since the required probabilities are fixed, we do not need to measure all the Stokes parameters. Instead, we can just measure the following Stokes parameters to obtain all the required probabilities and then calculate the remaining Stokes parameters:

$$\begin{aligned}
S_{1,1} &= \left(P_{|D\rangle}-P_{|A\rangle}\right)\otimes\left(P_{|D\rangle}-P_{|A\rangle}\right)\\
&= P_{|DD\rangle}-P_{|DA\rangle}-P_{|AD\rangle}+P_{|AA\rangle},\\
S_{1,2} &= \left(P_{|D\rangle}-P_{|A\rangle}\right)\otimes\left(P_{|R\rangle}-P_{|L\rangle}\right)\\
&= P_{|DR\rangle}-P_{|DL\rangle}-P_{|AR\rangle}+P_{|AL\rangle},\\
S_{1,3} &= \left(P_{|D\rangle}-P_{|A\rangle}\right)\otimes\left(P_{|H\rangle}-P_{|V\rangle}\right)\\
&= P_{|DH\rangle}-P_{|DV\rangle}-P_{|AH\rangle}+P_{|AV\rangle},\\
S_{2,1} &= \left(P_{|R\rangle}-P_{|L\rangle}\right)\otimes\left(P_{|D\rangle}-P_{|A\rangle}\right)\\
&= P_{|RD\rangle}-P_{|RA\rangle}-P_{|LD\rangle}+P_{|LA\rangle},\\
S_{2,2} &= \left(P_{|R\rangle}-P_{|L\rangle}\right)\otimes\left(P_{|R\rangle}-P_{|L\rangle}\right)
\end{aligned}$$

$$
\begin{aligned}
&= P_{|RR\rangle} - P_{|RL\rangle} - P_{|LR\rangle} + P_{|LL\rangle},\\
S_{2,3} &= \left(P_{|R\rangle} - P_{|L\rangle}\right) \otimes \left(P_{|H\rangle} - P_{|V\rangle}\right)\\
&= P_{|RH\rangle} - P_{|RV\rangle} - P_{|LH\rangle} + P_{|LV\rangle},\\
S_{3,1} &= \left(P_{|H\rangle} - P_{|V\rangle}\right) \otimes \left(P_{|D\rangle} - P_{|A\rangle}\right)\\
&= P_{|HD\rangle} - P_{|HA\rangle} - P_{|VD\rangle} + P_{|VA\rangle},\\
S_{3,2} &= \left(P_{|H\rangle} - P_{|V\rangle}\right) \otimes \left(P_{|R\rangle} - P_{|L\rangle}\right)\\
&= P_{|HR\rangle} - P_{|HL\rangle} - P_{|VR\rangle} + P_{|VL\rangle},\\
S_{3,3} &= \left(P_{|H\rangle} - P_{|V\rangle}\right) \otimes \left(P_{|H\rangle} - P_{|V\rangle}\right)\\
&= P_{|HH\rangle} - P_{|HV\rangle} - P_{|VH\rangle} + P_{|VV\rangle}. \qquad (4)
\end{aligned}
$$

Therefore, we can calculate the following Stokes parameters:

$$
\begin{aligned}
S_{0,1} &= \left(P_{|H\rangle} + P_{|V\rangle}\right) \otimes \left(P_{|D\rangle} - P_{|A\rangle}\right)\\
&= P_{|DD\rangle} - P_{|DA\rangle} - P_{|AD\rangle} + P_{|AA\rangle},\\
S_{1,0} &= \left(P_{|D\rangle} - P_{|A\rangle}\right) \otimes \left(P_{|H\rangle} + P_{|V\rangle}\right)\\
&= P_{|DR\rangle} - P_{|DL\rangle} - P_{|AR\rangle} + P_{|AL\rangle},\\
S_{0,2} &= \left(P_{|H\rangle} + P_{|V\rangle}\right) \otimes \left(P_{|R\rangle} - P_{|L\rangle}\right)\\
&= P_{|DH\rangle} - P_{|DV\rangle} - P_{|AH\rangle} + P_{|AV\rangle},\\
S_{2,0} &= \left(P_{|R\rangle} - P_{|L\rangle}\right) \otimes \left(P_{|H\rangle} + P_{|V\rangle}\right)\\
&= P_{|RD\rangle} - P_{|RA\rangle} - P_{|LD\rangle} + P_{|LA\rangle},\\
S_{0,3} &= \left(P_{|H\rangle} + P_{|V\rangle}\right) \otimes \left(P_{|H\rangle} - P_{|V\rangle}\right)\\
&= P_{|RR\rangle} - P_{|RL\rangle} - P_{|LR\rangle} + P_{|LL\rangle},\\
S_{3,0} &= \left(P_{|H\rangle} - P_{|V\rangle}\right) \otimes \left(P_{|H\rangle} + P_{|V\rangle}\right)\\
&= P_{|RH\rangle} - P_{|RV\rangle} - P_{|LH\rangle} + P_{|LV\rangle}. \qquad (5)
\end{aligned}
$$

According to Eqs. (1) and (2), we obtain the density matrix

$$
\begin{aligned}
\hat{\rho} = \frac{1}{4}\Bigg[&\left(\frac{S_{0,0}}{S_{0,0}}\sigma_0\otimes\sigma_0 + \frac{S_{0,1}}{S_{0,0}}\sigma_0\otimes\sigma_1 + \frac{S_{0,2}}{S_{0,0}}\sigma_0\otimes\sigma_2 + \frac{S_{0,3}}{S_{0,0}}\sigma_0\otimes\sigma_3\right)\\
&+\left(\frac{S_{1,0}}{S_{0,0}}\sigma_1\otimes\sigma_0 + \frac{S_{1,1}}{S_{0,0}}\sigma_1\otimes\sigma_1 + \frac{S_{1,2}}{S_{0,0}}\sigma_1\otimes\sigma_2 + \frac{S_{1,3}}{S_{0,0}}\sigma_1\otimes\sigma_3\right)\\
&+\left(\frac{S_{2,0}}{S_{0,0}}\sigma_2\otimes\sigma_0 + \frac{S_{2,1}}{S_{0,0}}\sigma_2\otimes\sigma_1 + \frac{S_{2,2}}{S_{0,0}}\sigma_2\otimes\sigma_2 + \frac{S_{2,3}}{S_{0,0}}\sigma_2\otimes\sigma_3\right)\\
&+\left(\frac{S_{3,0}}{S_{0,0}}\sigma_3\otimes\sigma_0 + \frac{S_{3,1}}{S_{0,0}}\sigma_3\otimes\sigma_1 + \frac{S_{3,2}}{S_{0,0}}\sigma_3\otimes\sigma_2 + \frac{S_{3,3}}{S_{0,0}}\sigma_3\otimes\sigma_3\right)\Bigg]. \qquad (6)
\end{aligned}
$$

We constructed nine quantum circuits based on IBM Q for observing Bell states. According to Eqs. (4)–(6), we obtained the density matrix (Fig. 2)

$$
\rho = \begin{bmatrix}
0.4854+0.0000\mathrm{i} & -0.0010-0.0381\mathrm{i} & 0.0010-0.0054\mathrm{i} & 0.4082+0.0259\mathrm{i}\\
-0.0010+0.0381\mathrm{i} & 0.0499+0.0000\mathrm{i} & -0.0058+0.0122\mathrm{i} & 0.0029+0.0024\mathrm{i}\\
0.0010+0.0054\mathrm{i} & -0.0085-0.0122\mathrm{i} & 0.0235+0.0000\mathrm{i} & 0.0010-0.0146\mathrm{i}\\
0.4082-0.0259\mathrm{i} & 0.0029-0.0024\mathrm{i} & 0.0010+0.0146\mathrm{i} & 0.4463+0.0000\mathrm{i}.
\end{bmatrix}. \qquad (7)
$$

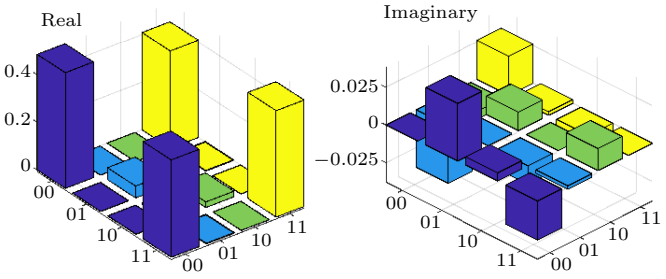


Fig. 2. Density matrix.

3. Using a neural network to estimate the quantum state

The complexity of traditional QST methods increases exponentially with the number of qubits, and due to noise effects their accuracy often remains limited. On the other hand, some neural network models can be used to accurately and efficiently approximate the nonlinear Schrödinger equation under consideration and predict the dynamic behavior of soliton solutions.[17] Here, we use neural networks to estimate the density matrix of quantum states. Due to their excellent generalization and learning abilities they can efficiently be utilized for quantum state estimation.

3.1. Building neural networks

As the number of qubits increases, noise interference and the time required for traditional measurement methods rapidly increase, making the standard methods impractical. At this point, employing neural networks to estimate quantum states is clearly advantageous as it effectively reduces noise interference and accelerates measurement time. This section focuses on quantum states of two qubits, which result in a total of 16 possible measurement outcomes. These outcomes form a vector set used as input for the neural network, representing the measurement vectors. Therefore, the designed neural network has 16 input nodes. The final output of the network is an estimated density matrix, with a size of 4×4. Considering that density matrices can be complex, they need to be separated

into real and imaginary parts, resulting in two matrices. These two matrices are then flattened column-wise into vectors and concatenated together as the output. Clearly, the output layer has 32 nodes, with the first 16 representing the real part of the density matrix and the latter 16 representing the imaginary part. The loss function adopts the mean squared error (MSE) between the estimated matrix and the true matrix. Subsequently, we will verify the feasibility of the neural network in estimating quantum states with three or more qubits. For the estimation of a three-qubit quantum state, the network architecture is basically the same, with only modifications to the input and output layers needed.

3.2. Training dataset construction

In this section we first utilize supervised learning with multilayer perceptron (MLP) and convolutional neural networks (CNNs) to estimate quantum states. As this is supervised learning, we need to construct training and testing samples to train the model. Additionally, we manually introduce noise into the training set to enhance the model's generalization ability.

Here, we use measurement operators constructed from Pauli matrices to measure the density matrix, where all measurement outcomes form vectors as input samples. For an n-qubit quantum system, the measurement operators, constructed from Pauli matrices, are randomly selected from a set of n matrices.[18] Then, the normalized tensor product of these matrices is calculated to give

$$W_i = \frac{\sigma_1 \otimes \sigma_2 \otimes \cdots \sigma_n}{\sqrt{d}}. \tag{8}$$

For W_i, the measurement outcomes y_i satisfy

$$y_i = \mathrm{tr}\left(W_i^{\dagger} \rho_k\right) = \mathrm{vec}\left(W_i\right)^{\dagger} \mathrm{vec}\left(\rho_k\right). \tag{9}$$

This formula illustrates that given a measurement algorithm W_i, the measurement value of the density matrix ρ_k is y_i. Then we construct the required input as a one-dimensional vector

$$b_k = [y_1, y_2, \ldots, y_{d\times d}] \tag{10}$$

and add a Gaussian noise sequence e_k

$$b_k = \left(\mathrm{vec}\left(W_1\right), \mathrm{vec}\left(W_2\right), \ldots, \mathrm{vec}\left(W_n\right)\right)^{\dagger} \mathrm{vec}\left(\rho_k\right) + e_k. \tag{11}$$

The density matrix ρ[19] corresponding to its measurements is

$$\rho = \frac{\varphi_r^{\dagger} \varphi_r}{\mathrm{tr}\left(\varphi_r^{\dagger} \varphi_r\right)}, \tag{12}$$

where φ_r is a $d \times r$ Wishart matrix, with elements following a normal distribution with mean 0 and standard deviation 1. When the rank $r = 1$, it represents a pure state density matrix. It can be verified that the density matrix generated using this method, which is used as an output sample for the neural network, satisfies the properties of being a Hermitian matrix, positive semi-definite and having unit trace.

During the training of the neural network, since it is supervised learning, we first need to generate a known density matrix. Then, the input to the neural network is a series of observations of this generated density matrix and the output is the estimated density matrix. So then we can obtain a random density matrix ρ by randomly generating a Wishart matrix that satisfies the above conditions.

3.3. Loss function and fidelity

The neural network needs to determine a loss function for training and a method to assess its effectiveness. Here, we adopt mean squared error (MSE) as the loss function and fidelity as the metric to assess the accuracy of the neural network

$$\mathrm{MSE}\left(\rho, \hat{\rho}\right) = \frac{\sum_{i=1}^{2^n} \sum_{j=1}^{2^n} \left|\rho_{i,j} - \hat{\rho}_{i,j}\right|^2}{4^n}, \tag{13}$$

where ρ is the generated true density matrix, $\hat{\rho}$ is the estimated value of the density matrix during the network training process and n is the number of qubits. We train the neural network by minimizing the MSE value. Since the output of the neural network is an approximation of the true density matrix, it may not necessarily satisfy the three properties required by the density matrix: Hermiticity, semi-definiteness and unit trace[20]

$$\rho = \rho^{\dagger}, \quad \rho \geq 0, \quad \mathrm{tr}\left(\rho\right) = 1. \tag{14}$$

We use ρ_r as the density matrix output by the neural network. To ensure that the quantum pure state satisfies the three properties, the output ρ of the neural network is normalized

$$\hat{\rho} = \frac{\rho_r^{\dagger} \rho_r}{\mathrm{tr}\left(\rho_r^{\dagger} \rho_r\right)}. \tag{15}$$

We use fidelity as a metric to evaluate the effectiveness of the experimental results, with $\hat{\rho}$ representing the estimated density matrix and ρ the true density matrix

$$F\left(\rho, \hat{\rho}\right) = \mathrm{tr}\left(\sqrt{\sqrt{\rho}\hat{\rho}\sqrt{\rho}}\right). \tag{16}$$

The fidelity ranges from 0 to 1.[21] When the difference between σ and ρ is smaller the value of $F\left(\rho, \hat{\sigma}\right)$ approaches 1, and when the two matrices are exactly the same the fidelity is 1.

3.4. Using three different neural networks to estimate quantum state

3.4.1. Multilayer perceptron

We employ mini-batch learning with a batch size of 128. The optimizer used is Adam and the activation function is

ReLU; weight decay is applied to prevent overfitting. The network architecture consists of two hidden layers with 1024 nodes each (Fig. 3). We want to generate a training dataset containing 10000 samples.

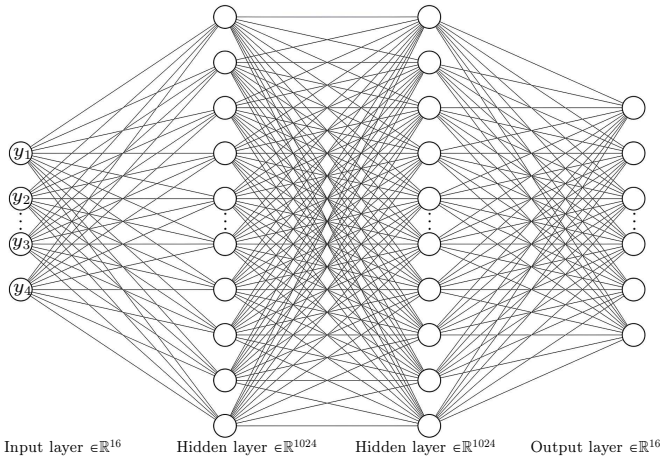


Fig. 3. Multilayer perceptron.

According to Eq. (12), we let the rank $r = 1$ so we can obtain the superposition state density matrix as output samples. Then, based on Eq. (11), we obtain the input samples. A total of 15000 samples were generated, with 10000 used as the training set samples and 5000 as the test set samples. The batch size was set to 256, and the number of training epochs was set to 10. The relationship between the defined loss function and the number of training epochs is depicted in Fig. 4.

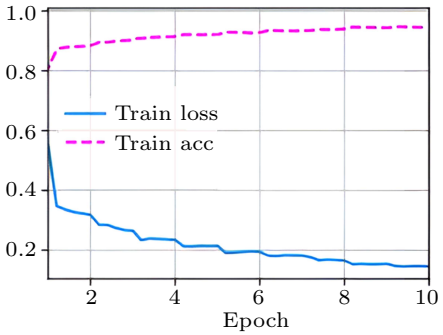


Fig. 4. The training process for a multilayer perceptron.

The final training loss function reaches the order of 10^{-5}, with a training accuracy of 98.6%. This validates the density matrix generated from the real observations of the input for the two-qubit system.

3.4.2. Convolutional neural networks

CNNs possess several advantages. The convolutional operations of a CNN only involve local regions of input and weights, rather than fully connecting to the entire input. This reduces the number of parameters in the network, alleviating the risk of overfitting, and enables effective handling of high-dimensional data. Furthermore, CNNs can automatically learn suitable feature representations from raw data for the task at hand.[22–24] Through the backpropagation algorithm, the network adjusts weights to minimize the loss function, thereby enabling the network to efficiently extract task-relevant features. To convert a one-dimensional vector into an image, the input vector

$$b_k = [y_1, y_2, \ldots, y_{d\times d}]. \tag{17}$$

For a two-qubit system, where $d = 4$, each set of d values should be listed separately on one line

$$\widetilde{b_k} = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 \\ y_5 & y_6 & y_7 & y_8 \\ y_9 & y_{10} & y_{11} & y_{12} \\ y_{13} & y_{14} & y_{15} & y_{16} \end{bmatrix}. \tag{18}$$

As each value in every measurement set represents a probability within the range $[0, 1]$, we need to optimize this input for better performance of the CNN. Here, we apply the following transformation to b to obtain a new matrix $\widetilde{b_k}'$ that better suits the requirements of the CNN

$$\widetilde{b_k}' = 255\widetilde{b_k}. \tag{19}$$

3.4.3. AlexNet

We input a 4×4 image with eight channels and expand it into a 55 image with 96 channels using a max-pooling layer. Then, we transform it into a 27×27 image with 256 channels using a max-pooling layer, and transform this into a 13×13 image with 384 channels using a convolutional layer twice. Next, this is transformed into a 13×13 image with 256 channels using a convolutional layer. Finally, a global pooling layer is used to flatten it into a one-dimensional vector, then it is passed through a fully connected neural network with two hidden layers, each containing 2048 nodes. Dropout is applied with a probability of 0.5 to avoid overfitting during this process, and a 16-parameter output vector is obtained (Fig. 5).

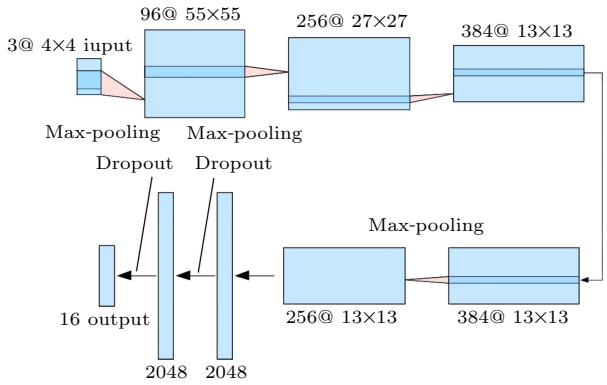


Fig. 5. AlexNet.

With a batch size of 128 and using the Adam optimizer with the ReLU activation function, we conducted experiments on the case where the number of samples for eigenstate selection is 10000. However, the batch size is adjusted to 256. We set the number of epochs to 10. The relationship between the mean squared error loss (MSELoss) and the number of training epochs is depicted in Fig. 6.

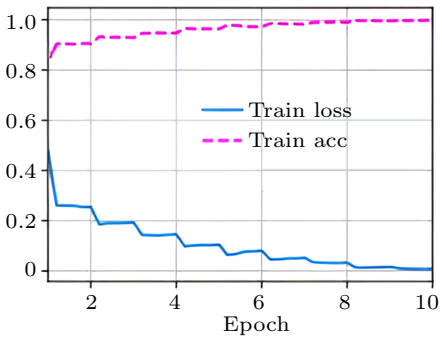


Fig. 6. The training process of AlexNet.

After training is complete, we use the trained estimator to validate the estimation of the density matrices in the test set. For the density matrix to be estimated, diag(1,0,0,0), we utilize the trained state estimator for estimation. Then, we transform the results of the estimator using Eq. (9), retaining five decimal places. The estimated result is as follows: the fidelity is calculated to be 0.999999976 and the RMSE is 0.000002695. Additionally, the estimated result satisfies the three properties of a density matrix. It can be observed that in estimating the eigenstates, the wide neural network estimator designed in this paper achieves an accuracy of over 99% fidelity estimation precision, with an RMSE accuracy reaching 10^{-4}.

3.4.4. ResNet

ResNet demonstrates strong performance and interpretability in estimating quantum states, effectively addressing gradient issues, handling complex quantum features and thereby improving the accuracy and stability of quantum state estimation. The introduction of residual connections helps alleviate the gradient vanishing and exploding problems commonly encountered in deep neural networks. This makes it easier for quantum state estimation models to converge during training and obtain stable results. Additionally, its architectural design enables the network to easily construct deep structures exceeding a thousand layers. Such depth structures better capture the complex features of quantum systems, enhancing the model's representational power and ability to learn more complex and abstract feature representations.[25–27] For quantum state estimation, this means that the network can better capture non-locality and entanglement features of quantum systems, thereby improving estimation accuracy. The residual connections of ResNet make the network easier to interpret, and by analyzing residual blocks (Fig. 7), one can gain a clearer understanding of the feature extraction and transformation processes the network undergoes when learning quantum state estimation tasks.

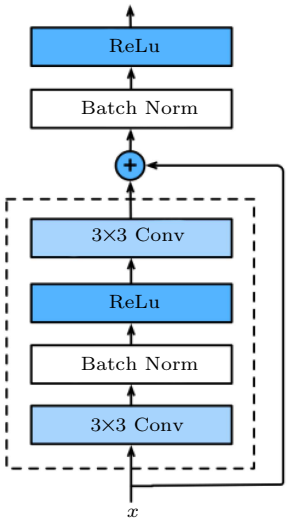


Fig. 7. Residual blocks.

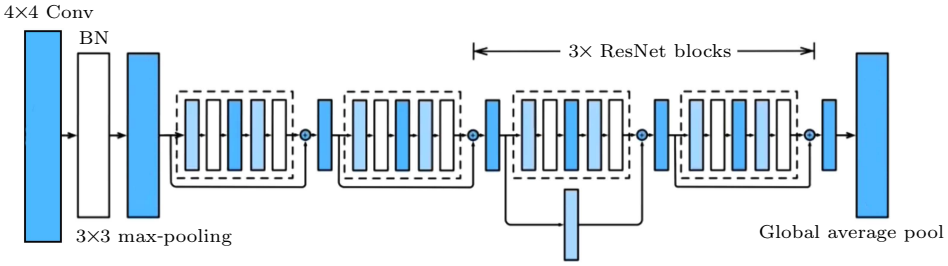


Fig. 8. ResNet.

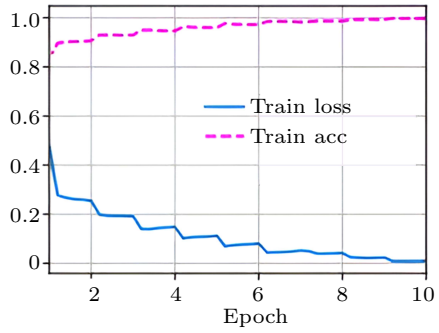


Fig. 9. The training process of ResNet.

To adopt residual blocks batch normalization is used after addition. To transform the observation one-dimensional vector into a 4×4 image and input it into a residual convolutional network the following steps are used: first, the network parameters are optimized using batch normalization. Then, it is passed through a 3×3 max-pooling layer and after that through four residual blocks. Finally, the result is output through a global pooling layer; we use the Adam optimizer and ReLU activation function (Fig. 8).

For the experiment with a sample size of 10000 for eigen-

state selection a batch size of 256 and 10 epochs are used. The relationship between the MSELoss function and the number of training epochs is shown in Fig. 9.

4. Comparison and summary

After determining the density matrix ρ for the quantum state, we use the standard Jacobi method to find its eigenvalues. For the MLP, the eigenvalues of the reconstructed density matrix and fidelity are 0.0013, 0.0024, 0.0258, 0.9643 and $F = 0.9933$, and satisfy $\text{trace}(\rho_{\text{MLP}}) = 1$, $\rho^{\dagger}_{\text{MLP}} = \rho_{\text{MLP}}$. For AlexNet, the eigenvalues and fidelity of the reconstructed density matrix are 0.0000, 0.0020, 0.0248, 0.9732 and $F = 0.9953$, and satisfy $\text{trace}(\rho_{\text{AlexNet}}) = 1$, $\rho^{\dagger}_{\text{AlexNet}} = \rho_{\text{AlexNet}}$. For ResNet, the eigenvalues and fidelity of the reconstructed density matrix are 0.0000, 0.0000, 0.0267, 0.9733 and $F = 0.9962$, and satisfies $\text{trace}(\rho_{\text{ResNet}}) = 1$, $\rho^{\dagger}_{\text{ResNet}} = \rho_{\text{ResNet}}$. All eigenvalues are positive semi-definite. Thus, deep learning methods produce positive semi-definite density matrices. It is important to investigate the quality and effectiveness of the reconstructed density matrix in a noisy environment. We add noise to the data v_j for each measurement of the two-qubit polynomial state using beta, Gaussian and uniform distributions, and add five noise coefficients: 0.02, 0.04, 0.06, 0.08 and 0.1.

We conducted 100 runs for each noise coefficient on each of the three networks. The mean fidelity of the reconstructed density matrix was calculated. Our analysis, as depicted in Figs. 10–12, reveals that ResNet yields the highest-fidelity density matrix, while the lowest fidelity is observed with MLP. The least favorable fidelity achieved by MLP is approximately 0.98 in a Gaussian distribution noise setting with a coefficient of 0.1, around 0.976 for a beta distribution and approximately 0.963 for a uniform distribution. Therefore, all networks exhibit superior performance in a Gaussian distribution noise environment. On one hand the benefits of quantum information stem from the quantum non-locality inherent in entangled states, known for applications such as quantum teleportation and dense coding. Conversely, quantum entanglement offers novel computational avenues, thus rendering its creation and assessment pivotal. Our study employed both network and conventional QST techniques to reconstruct the density matrices of entangled states with two and three qubits, respectively. We assessed the efficacy of these methods in constructing density matrices of entangled states.

Figures 13 and 14 depict the reconstruction of the density matrix of the two-qubit entangled Bell state $|\psi\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$ using standard QST and MLP-based deep learning. QST was further conducted on IBM Q hardware (IBM Q Manila), yielding 36 outcomes from 1000 measurements. The eigenvalues of the reconstructed density matrix via QST were found to be -0.0129, 0.0402, 0.0747 and 0.8970, with a fidelity of 0.9903 and a trace equal to 1, satisfying Hermiticity but not positive semi-definiteness. Subsequently, we employed deep learning for density matrix estimation, utilizing the same 36 measurements as input data for the neural network. Following state estimation using the neural network, all eigenvalues became positive.

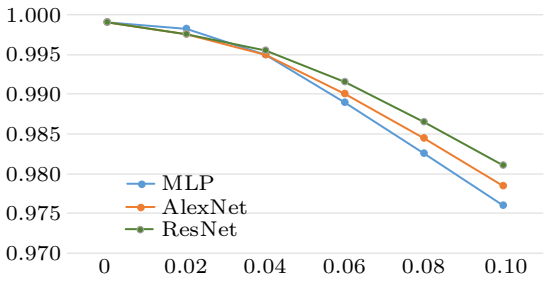


Fig. 10. Performance of MLP, AlexNet and ResNet in Gaussian noise distributions.

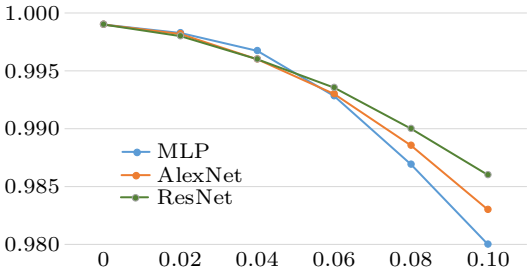


Fig. 11. Performance of MLP, AlexNet and ResNet in beta noise distributions.

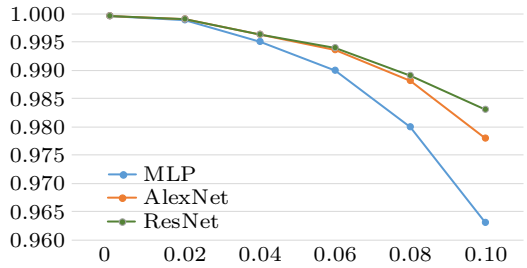


Fig. 12. Performance of MLP, AlexNet and ResNet in uniform noise distributions.

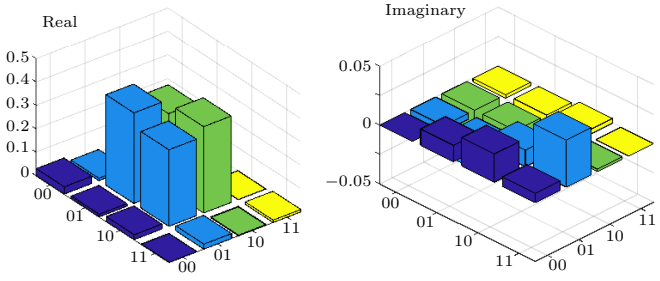


Fig. 13. Reconstruction of the density matrix based on tomography.

The entangled W state of three qubits is represented as $|\psi\rangle = \frac{1}{\sqrt{3}}(|001\rangle + |010\rangle + |100\rangle)$. Studies show that the W state remains quite stable in real physical systems. Unlike other entangled states, the W state can maintain entanglement even with particle loss or decoherence. In contrast, states like the Greenberger–Horne–Zeilinger state lose all entanglement if one particle is lost.[28] This stability means that the W state

is widely used in quantum information processing, forming the basis for quantum memories, quantum secure direct communication and universal quantum cloning machines.

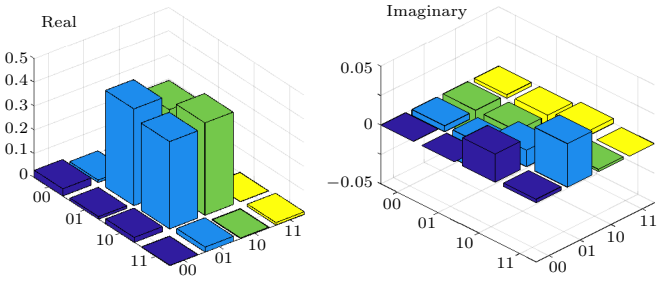


Fig. 14. Reconstruction of the density matrix based on deep learning.

Using IBM Q, we created a three-qubit W state with an initial state of $|000\rangle$ by applying a series of quantum gates, including R_Y rotation gates, controlled-Z gates and controlled-NOT gates, as shown in Fig. 15.

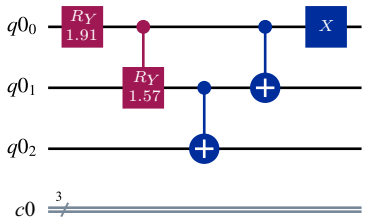


Fig. 15. Measurement of the three-qubit quantum state.

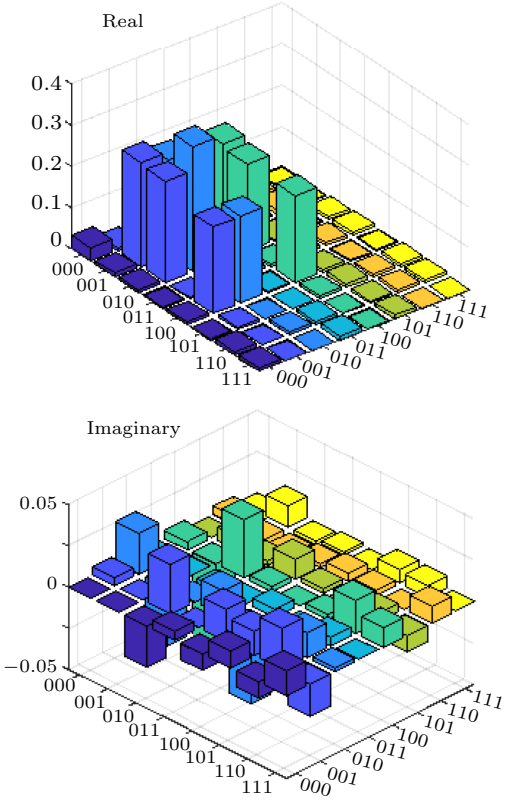


Fig. 16. The density matrix obtained through residual neural networks in deep learning methods is shown in Fig. 17.

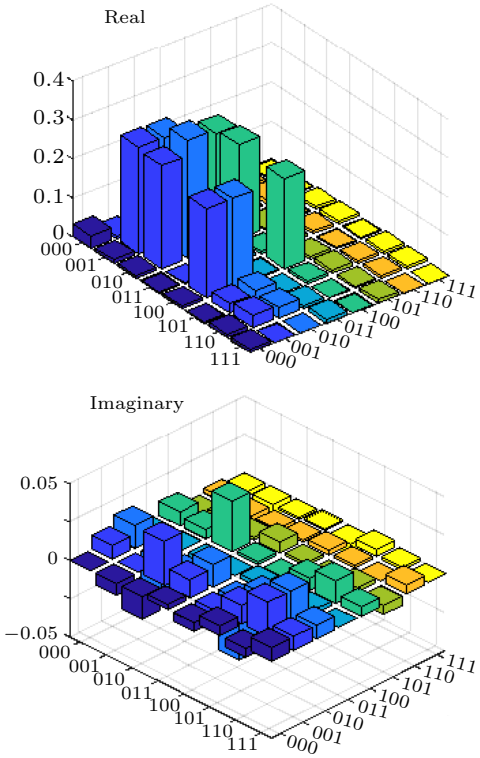


Fig. 17. Density matrix obtained through residual neural networks in deep learning methods.

Standard quantum tomography and deep learning methods were then used to reconstruct the density matrix of the W state. The experimentally reconstructed density matrix, described in coordinates from $|000\rangle$ to $|111\rangle$, is similar to previous superposition and Bell states; quantum tomography of the W state was also conducted on quantum hardware (IBM Q Manila), with a measurement count set to 1800, yielding 512 measurement outcomes. After obtaining estimated values for 63 parameters $r_{i_1,i_2,i_3} = 1$ (normalized due to the density matrix with $r_{0,0,0} = 1$), they were substituted into Eq. (6) to derive the density matrix ρ. The final density matrix obtained featured negative eigenvalues of 0.0280, 0.0280, 0.0166, 0.0166, 0.0041, 0.0041, 0.0159, 0.0159, 0.0440, 0.0440, 0.0594, 0.0594, 0.0887, 0.0887, 0.8325 and 0.8325. The density matrix obtained through standard QST is shown in Fig. 16.

The fidelity $F = 0.9743$ of the density matrix obtained through classical QST and ResNet corresponds to eigenvalues and fidelities of 0.0000, 0.0000, 0.0042, 0.0126, 0.0231, 0.0509, 0.0668, 0.8420 and $F = 0.9837$. We find that using a network always results in a semi-positive definite density matrix and the fidelity of the density matrix is better than using the standard QST method. In addition to traditional methods, incorporating feedback control and quantum error correction can significantly enhance multi-parameter estimation in quantum systems.[29] Finally, for large quantum systems such as a 10-qubit system the dimensionality of the input vector reaches the order of 10^6. Consequently, using a MLP for estimation is no longer effective. When using CNNs for estimation, ResNet

is employed. The input is a 1024×1024 matrix, and we modify the output to be a two-channel 1024×1024 matrix. One channel represents the real part of the density matrix and the other represents the imaginary part of the density matrix. Normally ResNet can easily handle 1024-pixel images, so training this network to estimate 10-qubit quantum states is clearly feasible. Therefore, this method can effectively handle large quantum systems. However, it is worth noting that the size of the quantum state matrix grows exponentially. For quantum states with more than 14 qubits, the size of the input and output matrices exceeds 10000×10000. At this point, using convolutional networks may also become impractical due to computational complexity.

5. Conclusion

In this paper we successfully employed neural networks based on deep learning to estimate the reconstruction of separable and entangled states of two and three qubits. Subsequently, we compared this approach with the standard quantum state tomography (QST) method. The comparison indicates that neural networks estimate the quantum state as a whole and incorporate prior knowledge of the density matrix from the outset. This incorporation ensures the positive semi-definiteness and normalization of the matrix, thus significantly reducing statistical errors. Additionally, the three models employed in the neural network include multilayer perceptron (MLP), AlexNet and residual networks. These models have limited adjustable parameters, simple adaptation processes and are independent of gradient information. It is noteworthy that MLP demonstrates the highest convergence rate, while the residual neural network produces the highest-quality reconstructed density matrix. Furthermore, all these algorithms outperform the standard QST method in fidelity. Therefore, neural network estimation of quantum states has been proven to be more effective in the presence of statistical noise and limited measurement quantities.

References

[1] D'Alessandro D 2003 *J. Phys. A: Math. Gen.* **36** 9721

[2] Fano U 1957 *Rev. Mod. Phys.* **29** 74

[3] Smith A, Riofrio C A, Anderson B E, Sosa-Martinez H, Deutsch I H and Jessen P S 2013 *Phys. Rev. A* **87** 030102(R)

[4] Mario M 2018 *Journal of Applied Mathematics and Physics* **6** 1363

[5] Lukens J M, Law K J H, Jasra A and Lougovski P 2020 *New J. Phys.* **22** 063038

[6] Rumelhart D E, Hinton G E and Williams R J 1986 *Nature* **323** 533

[7] Xu Q and Xu S Q 2018 *arXiv c-prints* arXiv:1811.06654

[8] Xin T, Lu S R, Cao N P, Galit Anikeeva, Lu D W, Li J, Long G L and Zeng B 2019 *Npj Quantum Inform.* **5** S127

[9] Li H and Cheng J 2023 *Chin. Phys. B* **32** 100602

[10] Chen J, Dawkins H, Ji Z, Johnston N, Kribs D, Shultz F and Zeng B 2013 *Phys. Rev. A* **88** 012109

[11] Toninelli E, Ndagano B, Valles A, Sephton B, Nape I, Ambrosio A, Capasso F, Padgett M J and Forbes A 2019 *Adv. Opt. Photon.* **11** 67

[12] Bayraktar O, Swillo M, Canalias C and Bjork G 2016 *Phys. Rev. A* **94** 020105

[13] Thew R T, Nemoto K, White A G and Munro W J 2002 *Phys. Rev. A* **66** 012303

[14] James D F, Kwiat P G, Munro W J and White A G 20010 *Phys. Rev. A* **64** 052312

[15] Schmied R 2016 *J. Mod. Opt.* **63** 1744

[16] Garcia-Martin D and Sierra G 2018 *JAMP* **06** 1760

[17] Liu X M, Zhang Z Y and Liu W J 2023 *Chin. Phys. Lett.* **40** 070501

[18] Nielsen M A and Chuang I L 2010 *Quantum Computation and Quantum Information* (Cambridge University Press)

[19] Zyczkowski K, Penson K A, Nechita I and Collins B 2011 *J. Math. Phys.* **52** 062201

[20] Bravyi S and Kitaev A 2002 *Annals of Physics* **298** 210

[21] Jozsa R 1993 *Journal of Modern Optics*

[22] LeCun Y, Bengio Y and Hinton G 2015 *Nature* **521** 436

[23] Goodfellow J I, Bengio Y and Courville A 2016 *Genetic Programming and Evolvable Machines* **19** 305

[24] Simonyan K and Zisserman A 2014 arXiv:1409.1556

[25] He K M, Zhang X Y, Ren S Q and Sun J 2015 arXiv:1512.03385

[26] Huang G, Liu Z, L. Van Der Maaten and Weinberger K Q 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 2261–2269

[27] Szegedy C, Liu W, Jia Y Q, *et al.* 2015 *015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 1–9

[28] Dur W, Vidal G and Cirac J I 2000 *Phys. Rev. A* **62** 062314

[29] Hong H Y, Lu X J and Kuang S 2023 *Chin. Phys. B* **32** 040603