

ARTICLE OPEN

Local-measurement-based quantum state tomography via neural networks

Tao Xin^{1,2,8}, Sirui Lu^{2,8}, Ningping Cao^{3,4,8}, Galit Anikeeva⁴, Dawei Lu¹, Jun Li^{1*}, Guilu Long^{2,5,6} and Bei Zeng^{3,4,7*}

Quantum state tomography is a daunting challenge of experimental quantum computing, even in moderate system size. One way to boost the efficiency of state tomography is via local measurements on reduced density matrices, but the reconstruction of the full state thereafter is hard. Here, we present a machine-learning method to recover the ground states of k -local Hamiltonians from just the local information, where a fully connected neural network is built to fulfill the task with up to seven qubits. In particular, we test the neural network model with a practical dataset, that in a 4-qubit nuclear magnetic resonance system our method yields global states via the 2-local information with high accuracy. Our work paves the way towards scalable state tomography in large quantum systems.

npj Quantum Information (2019)5:109; <https://doi.org/10.1038/s41534-019-0222-3>

INTRODUCTION

Quantum state tomography (QST) plays a vital role in validating and benchmarking quantum devices,^{1–5} because it can completely capture properties of an arbitrary quantum state. However, QST is not feasible for large systems because of its need for exponential resources. In recent years, there has been extensive research on methods for boosting the efficiency of QST.^{6–12} One of the promising candidates among these methods is QST via reduced density matrices (RDMs);^{13–19} because local measurements are convenient and accurate on many experimental platforms.

QST via RDMs is also a useful tool for characterizing ground states of local Hamiltonians. A many-body Hamiltonian H is k -local if $H = \sum_i H_i^{(k)}$, where each term $H_i^{(k)}$ acts non-trivially on at most k particles. For k -local Hamiltonians, only polynomial number of parameters are needed to characterize the whole system. Moreover, generally, a single eigenstate of such k -local Hamiltonian can encode the information of the system.^{18,20,21} Therefore, for these ground states, one only needs k -local measurements for state tomography. Although local measurements are efficient and even if $|\psi\rangle$ is uniquely determined by its k -local measurements, reconstructing $|\psi\rangle$ from its k -local measurements is computationally hard.²² We remark that this is not due to the problem that $|\psi\rangle$ needs to be described by exponentially many parameters. In fact, in many cases, ground states of k -local Hamiltonians can be effectively represented by tensor product states.^{18,23}

The state reconstruction problem naturally connects to the regression problem in supervised learning. Regression analysis, in general, seeks to discover the relation between inputs and outputs, i.e., to recover the underlying mathematical model. Unsupervised learning techniques have been applied to QST in various cases, such as in refs.^{24,25} In our case, as shown in Fig. 1, by knowing the Hamiltonian H , it is relatively easy to get the ground state $|\psi_H\rangle$ since the ground state is nothing but the eigenvector corresponding to the smallest eigenvalue. And then we could naturally achieve the k -local measurements \mathbf{M} of $|\psi_H\rangle$.

Therefore, the data for tuning our reverse engineering model is accessible, which allows us to realize QST through supervised learning practically. Additionally, artificial neural networks are often noise tolerable,^{26–28} so they are favorable for working with experimental data.

In this work, we propose a local-measurement-based QST by fully connected feedforward neural network, in which every neuron connects to every neuron in the next layer and information only passes forward (i.e., there is no loop in the network). We first build a fully connected feedforward neural network for 4-qubit ground states of fully connected 2-local Hamiltonians. Our trained 4-qubit network not only analyzes the test dataset with high fidelity but also reconstruct 4-qubit nuclear magnetic resonance (NMR) experimental states accurately. We use the 4-qubit case to demonstrate the potential of using neural networks to realize QST via k -local measurements. The versatile framework of neural networks for recovering ground states of k -local Hamiltonians could be extended to more qubits and various interaction structures; we then apply our methods to the ground states of seven-qubit 2-local Hamiltonians with nearest-neighbor couplings. In both cases, neural networks give accurate estimates with high fidelities. We observe our framework yields higher efficiency and better noise tolerance compared with the least-squares tomography (the approximated maximum likelihood estimation (MLE)) when the added noise is $>5\%$.

RESULTS

Theory

The universal approximation theorem²⁹ states that every continuous function on the compact subsets of \mathbb{R}^n can be approximated by a multi-layer feedforward neural network with a finite number of neurons, i.e., computational units. And by observing the relation between k -local Hamiltonian and local measurements of its ground state, as shown in Fig. 1, we are empowered to turn the

¹Shenzhen Institute for Quantum Science and Engineering and Department of Physics, Southern University of Science and Technology, Shenzhen 518055, China. ²State Key Laboratory of Low-Dimensional Quantum Physics and Department of Physics, Tsinghua University, Beijing 100084, China. ³Department of Mathematics & Statistics, University of Guelph, Guelph N1G 2W1 ON, Canada. ⁴Institute for Quantum Computing, University of Waterloo, Waterloo, ON N2L 3G1, Canada. ⁵Tsinghua National Laboratory of Information Science and Technology and The Innovative Center of Quantum Matter, Beijing 100084, China. ⁶Beijing Academy of Quantum Information Sciences, Beijing 100193, China. ⁷Department of Physics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China. ⁸These authors contributed equally: Tao Xin, Sirui Lu, Ningping Cao. *email: lij3@sustech.edu.cn; zengb@ust.hk

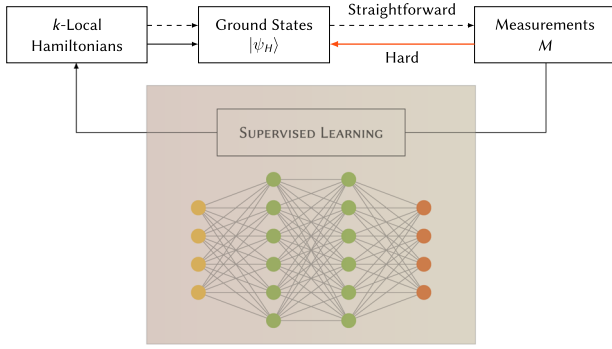


Fig. 1 Procedure of our neural network based local quantum state tomography method. As shown by the dashed arrows, we first construct training and test dataset by generating random k -local Hamiltonians H , calculate their ground states $|\psi_H\rangle$, and obtain local measurement results \mathbf{M} . We then train the neural network with the generated training dataset. After training, as represented by the black arrows, we first obtain the Hamiltonian H through local measurement results \mathbf{M} from the neural network, then recover the ground states from the obtained Hamiltonian. In contrast, the red arrow represents the direction of the normal QST process, which is computationally hard.

tomography problem to a regression problem which fits perfectly into the neural network framework.

In particular, we first construct a deep neural network for 4-qubit ground states of full 2-local Hamiltonians as follows:

$$H = \sum_{i=1}^4 \sum_{1 \leq k \leq 3} \omega_k^{(i)} \sigma_k^{(i)} + \sum_{1 \leq i < j \leq 4} \sum_{1 \leq n, m \leq 3} J_{nm}^{(ij)} \sigma_n^{(i)} \otimes \sigma_m^{(j)}, \quad (1)$$

where $\sigma_k, \sigma_n, \sigma_m \in \Delta$, and $\Delta = \{\sigma_1 = \sigma_x, \sigma_2 = \sigma_y, \sigma_3 = \sigma_z, \sigma_4 = I\}$. We denote the set of Hamiltonian coefficients as $\vec{h} = \{\omega_k^{(i)}, J_{nm}^{(ij)}\}$. The coefficient vector \vec{h} is the vector representation of H according to the basis set $\mathbf{B} = \{\sigma_m \otimes \sigma_n : n + m \neq 8, \sigma_m, \sigma_n \in \Delta\}$. The configuration of the ground states is illustrated in Fig. 2a.

The number of parameters of the local observables of ground states determines the amount of input units of the neural network. Concretely, $\mathbf{M} = \{s_{m,n}^{(ij)} : s_{m,n}^{(ij)} = \text{Tr}(\text{Tr}_{(ij)} \rho \cdot B_{(m,n)}), B_{(m,n)} \in \mathbf{B}, 1 \leq i < j \leq 4, 1 \leq n, m \leq 4\}$, where $\sigma_n, \sigma_m \in \Delta$ and ρ is the density matrix of the ground state. \mathbf{M} is a set of true expectation values $s_{m,n}^{(ij)}$ of the local observables $B_{(m,n)}$ in the ground states ρ . Notice that we are using the true values of expectation values instead of their estimations (which contain statistical fluctuations), since we are theoretically generating all the training and testing data. The input layer has 66 neurons since the cardinality of the set of measurement results is 66. Our network then contains two fully connected hidden layers, in which every neuron in the previous layer is connected to every neuron in the next layer. The number of output units equals to the number of parameters of our 2-local Hamiltonian, which is 66 in our 4-qubit case. More details of our neural network can be found in “Methods” section.

Our training data consist of the 120,000 randomly generated 2-local Hamiltonians as output and the local measurements of their corresponding ground states. The test data include 5000 pairs of Hamiltonians and local measurement results (H_i, \mathbf{M}_i) .

We train the network by a popular optimizer in the machine-learning community called Adam (adaptive moment estimation).^{30,31} For loss function, we choose cosine proximity $\cos(\theta) = (\vec{h}_{\text{pred}} \cdot \vec{h}) / (\|\vec{h}_{\text{pred}}\| \cdot \|\vec{h}\|)$, where \vec{h}_{pred} is the estimate of the neural network and \vec{h} is the desired output. Generally speaking, the role of loss functions in supervised learning is to efficiently measure the distance between the true value and the estimated outcome. (In our case, it is the distance between \vec{h} and \vec{h}_{pred}). And the training procedure seeks to

minimize this distance. We find the cosine proximity function fits our scenario better than the more commonly chosen loss functions, such as mean square error or mean absolute error.³² The reason can be understood as follows. Because the parameter vector \vec{h} is a representation of the corresponding Hamiltonian in the Hilbert space expanded by the local operators \mathbf{B} , the angle θ between the two vectors \vec{h} and \vec{h}_{pred} is a “directional distance measure” between two corresponding Hamiltonians.²⁰ Notice that the Hamiltonian corresponding to the parameter \vec{h} has the same eigenvectors as those of the Hamiltonian of $c \cdot \vec{h}$, where $c \in \mathbb{R}$ is a constant. In other words, we only care about the “directional distance”. Instead of forcing every single element close to its true value (as mean squared error or mean absolute error does), the cosine loss function tends to train the angle θ towards zero, which is more adapted to our situation.

As illustrated in Fig. 1, after getting estimated Hamiltonian from the neural network, we calculate the ground state ρ_{nn} of the estimated Hamiltonian and take the result as the estimate of the ground state that we attempt to recover. We remark that our estimated Hamiltonian is not necessarily exactly the same as the original Hamiltonian; Even if that happens, our numeric results suggest their ground states are still close.

There are two different fidelity functions that we can use to measure the distance between the randomly generated states ρ_{rd} and our neural network estimated states ρ_{nn} , namely:

$$f(\rho_1, \rho_2) \equiv \text{Tr} \sqrt{\sqrt{\rho_1} \rho_2 \sqrt{\rho_1}}, \quad (2)$$

$$C(\rho_1, \rho_2) \equiv \frac{\text{Tr}(\rho_1 \rho_2)}{\sqrt{\text{Tr}(\rho_1^2)} \cdot \sqrt{\text{Tr}(\rho_2^2)}}. \quad (3)$$

The fidelity measure f defined in Eq. (2) is standard,³³ which requires the matrix ρ_1 and ρ_2 to be positive semi-definite. Considering that the density matrix obtained directly from the raw data of a state tomography experiment may possibly not be positive definite, we usually adopt the definition of C for processing the raw data in NMR.³⁴ In this work, there are not negative matrices after constraining the raw density matrices to be positive semi-definite. The values of the fidelities are calculated by f if there is no additional explanation in the following.

After supervised learning on the training data, our neural network is capable of estimating the 4-qubit output of the test set with high performance. The fidelity averaged over the whole test set is 98.7%. The maximum, minimum, standard deviations of fidelities for the test set are shown in Table 1. Figure 2c illustrates the fidelities between 100 random states ρ_{rd} and our neural network estimates ρ_{nn} .

Our framework generalizes directly to more qubits and different interaction patterns. We apply our framework to recover 7-qubit ground states of 2-local Hamiltonians with nearest-neighbor interaction. The configuration of our 7-qubit states is shown in Fig. 2b. The Hamiltonian of this 7-qubit case is

$$H = \sum_{i=1}^7 \sum_{1 \leq k \leq 3} \omega_k^{(i)} \sigma_k^{(i)} + \sum_{i=1}^6 \sum_{1 \leq n, m \leq 3} J_{nm}^{(i)} \sigma_n^{(i)} \otimes \sigma_m^{(i+1)}, \quad (4)$$

where $\sigma_k, \sigma_n, \sigma_m \in \Delta$, $\omega_k^{(i)}$ and $J_{nm}^{(i)}$ are coefficients. We trained a similar neural network with 250,000 pairs of randomly generated Hamiltonians and 2-local measurements of the corresponding ground states. For the 5000 randomly generated test sets, the network estimates have an average fidelity of 97.9%. More statistical performance are shown in Table 1 and fidelity results of 100 random generated states are shown in Fig. 2d.

Due to the variance inherent to this method, it is natural to ask how to determine whether a neural network estimate ρ_{nn} is acceptable without knowing the true state ρ_{rd} . This problem can be easily solved by calculating the measurement estimate \mathbf{M}_{pred} , i.e., using the estimate ρ_{nn} to measure the set of local operators \mathbf{B} .

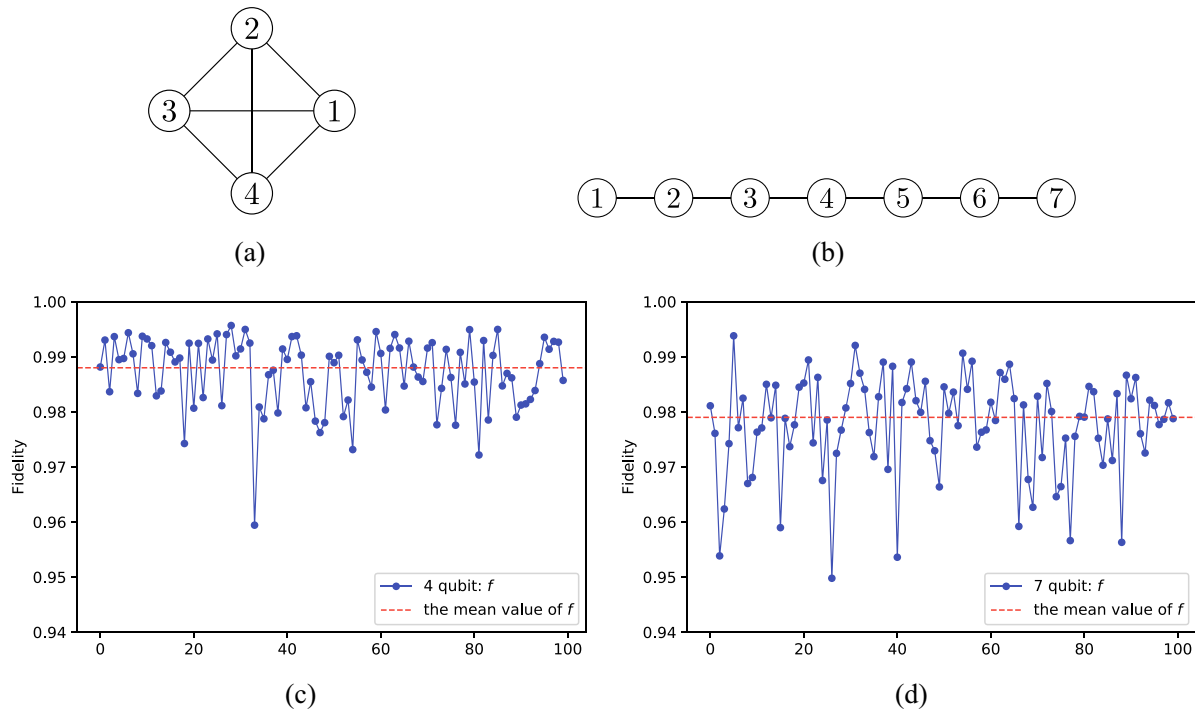


Fig. 2 Theoretical results for 4 qubits and 7 qubits. **a** The configuration of our 4-qubit states. Each dot presents a qubit, and every qubit interacts with each other. **b** The configuration of our 7-qubit states: only nearest qubits have interactions. **c** The fidelities of 100 random 4-qubit states ρ_{rd} and our neural network estimates ρ_{nn} . Notice that the x -coordinate does not have physical meaning, we randomly pick 100 states and label them from 0 to 99. It is the same for the 7-qubit case. The average fidelity (Eq. (2)) is 98.7%. **d** The fidelities of 100 random 7-qubit states ρ_{rd} and our neural network estimates ρ_{nn} . The average fidelity (Eq. (2)) of the whole test data set is 97.9%.

Table 1. The statistical performance of our neural networks for 4-qubit and 7-qubit cases.

| | Max (%) | Min (%) | Standard deviation | Average fidelity (%) |
|---------|---------|---------|--------------------|----------------------|
| 4-qubit | 99.8 | 91.4 | 5.93e−3 | 98.7 |
| 7-qubit | 99.6 | 85.2 | 10.4e−3 | 97.9 |

By setting an acceptable error bound and comparing \mathbf{M}_{pred} with the true measurements \mathbf{M} , one can decide whether to accept ρ_{nn} or not. Please see the “Methods” section for details.

Our neural-network-based framework is also significantly faster than the approximated MLE method. Once the network is trained sufficiently well, it can be used to deal with thousands of datasets without much effort on a regular computer. Calculating ρ_{nn} from \overline{h}_{pred} , which is essentially the computation of the eigenvector corresponding to the smallest eigenvalue, is the only part that may take some time. Detailed discussions could be found in the “Methods” section.

Experiment

So far, our theoretical model is noise-free. To demonstrate that our trained machine-learning model is resilient to experimental noises, we experimentally prepare the ground states of the random Hamiltonians and then try to reconstruct the final quantum states from 2-local measurements using a four-qubit NMR platform.^{35–38} The four-qubit sample is ^{13}C -labeled *trans*-crotonic acid dissolved in d_6 -acetone, where C_1 – C_4 are encoded as the four work qubits, and the rest spin-half-nuclei are decoupled throughout all experiments. Figure 3 describes the parameters and structure of this molecule. Under the weak-coupling approximation, the

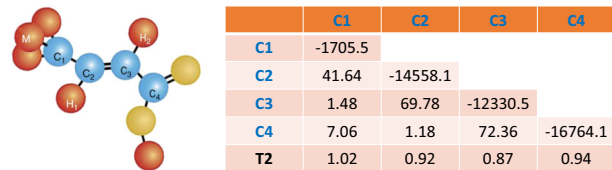


Fig. 3 The molecular structure and Hamiltonian parameters of the ^{13}C -labeled *trans*-crotonic acid. The atoms C_1 – C_4 are used as the four qubits in the experiment, and the atoms M , H_1 and H_2 are decoupled throughout the experiment. In the table, the chemical shifts with respect to the Larmor frequency and J -coupling constants (in Hz) are listed by the diagonal and off-diagonal numbers, respectively. The relaxation timescales T_2 (in seconds) are shown at the bottom.

Hamiltonian of the system writes

$$\mathcal{H}_{\text{int}} = \sum_{j=1}^4 \pi(\nu_j - \nu_0) \sigma_z^j + \sum_{j < k=1}^4 \frac{\pi}{2} J_{jk} \sigma_z^j \sigma_z^k, \quad (5)$$

where ν_j are the chemical shifts, J_{jk} are the J -coupling strengths, and ν_0 is the reference frequency of ^{13}C channel in the NMR platform. All experiments were carried out on a Bruker AVANCE 400 MHz spectrometer at room temperature. We briefly describe our three experimental steps here and leave the details in the “Methods” section: (i) *Initialization*: The pseudo-pure state^{39–41} for being the input of quantum computation $|0000\rangle$ is prepared. (More details are provided in the “Methods” section). (ii) *Evolution*: Starting from the state $|0000\rangle$, we create the ground state of the random two-body Hamiltonian by applying the optimized shaped pulses. (iii) *Measurement*: In NMR experiments, the expectation values of all 2-qubit Pauli products can be measured by the ensemble measurement. From them, we can directly obtain all 2-local measurements, and perform four-qubit QST to estimate the

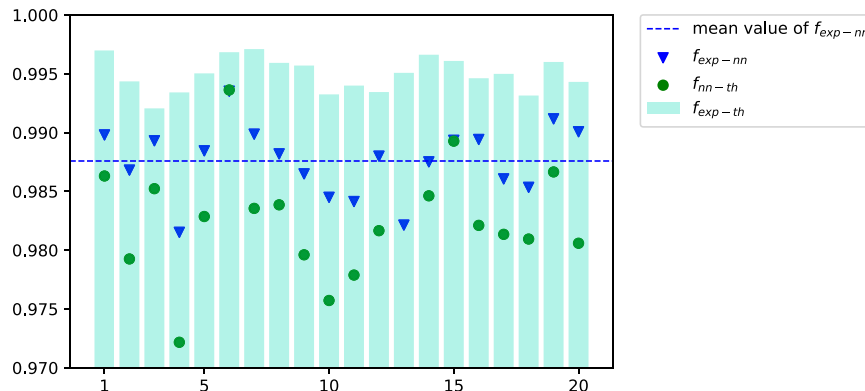


Fig. 4 The predication results with experimental data. Here we list three different fidelities for 20 experimental instances. The horizontal axis is the dummy label of the 20 experimental states. The cyan bars, $f_{\text{exp-th}}$, are the fidelities between the theoretical states ρ_{th} and the experimental states ρ_{exp} . The blue triangles, $f_{\text{exp-nn}}$, are fidelities between our neural network estimates ρ_{nn} and the experimental states ρ_{exp} with the average fidelity over 98.8%. And the green dots, $f_{\text{nn-th}}$, are the fidelities between our neural network estimates and the theoretical states.

quality of our implementations, which is accomplished by the least-squares tomography from the experimental data. More details about the least-squares tomography can be found in “Methods” section.

In experiments, we created the ground states of 20 random Hamiltonians of the form in Eq. (1) and performed 4-qubit QST for them after the state preparations. It is worth emphasizing that the experimental raw density matrices obtained from ensemble measurements on NMR are usually negative. First, we further performed the least-squares QST from the raw density matrices in experiments as ρ_{exp} , and estimated that the fidelities between the experimental states ρ_{exp} and the target ground state ρ_{th} are over 99.2%. It is noted that the purpose of reconstructing the states ρ_{exp} is to use them to compare with the results estimated by our neural network. We collected the expectation values of all 2-qubit Pauli product operators, such as $\langle \sigma_x \otimes I \otimes I \otimes I \rangle$ and $\langle \sigma_x \otimes \sigma_y \otimes I \otimes I \rangle$, which were directly obtained by measuring the expectation values of these Pauli strings in NMR. Then we fed them into our neural-network-based framework to reconstruct the 4-qubit states, obtaining an average fidelity of 98.8% between ρ_{exp} and ρ_{nn} , where ρ_{nn} is the neural network estimated state. Figure 4 shows the fidelity details of these density matrices. The results indicate that the original 4-qubit state can be efficiently reconstructed by our trained neural network using only 2-local measurements, instead of the traditional full QST.

DISCUSSION

As a famous double-edged sword in experimental quantum computing, QST captures full information of quantum states on the one hand, while on the other hand, its implementation consumes a tremendous amount of resources. Unlike traditional QST that requires exponentially many experiments with the growth of system size, the recent approach by measuring RDMs and reconstructing the full state thereafter opens up a new avenue to efficiently realize experimental QST. However, there is still an obstacle in this approach, that it is in general computationally hard to construct the full quantum state from its local information.

This is a typical problem empowered by machine learning. In this work, we apply the neural network model to solve this problem and demonstrate the feasibility of our method with up to seven qubits in the simulation. It should be noticed that 7-qubit QST in experiments is already a significant challenge in many platforms—the largest QST to date is of 10 qubits in superconducting circuits, where the theoretical state is a GHZ state with rather simple mathematical form.⁶¹ We further demonstrate that

our method works well in a 4-qubit NMR experiment, thus validating its usefulness in practice. We anticipate this method to be a powerful tool in future QST tasks of many qubits due to its accuracy and convenience. Comparing with the MLE, our method has acceptable fidelities, better noise tolerance and also has a significant advantage in terms of speed.

Our framework can be extended in several ways. First, we can consider excited states. As stated in the “Results” section, the Hamiltonian recovered by our neural network is not necessarily the original Hamiltonian, but their ground states are fairly close. We preliminarily examined the eigenstates of estimated Hamiltonians. Although the ground states have considerable overlap, the excited states are not close to each other. It means, in this reverse engineering problem, ground states are numerically more stable than excited states. To recover excited states using our method, one may need to use more sophisticated neural networks, such as convolutional neural network⁶² (CNN) or residual neural network⁶³ (ResNet). Second, although we have not included noise in the training and test data, our network estimates the experimental 4-qubit fully connected 2-local states with high fidelities. This indicates our method has certain error tolerant ability. For future study, one can add different noise to the training and test data. Third, one can also study how to incorporate the current method into the existing quantum tomography methods, such as compressive sensing techniques.^{9,64,65}

METHODS

Machine learning

In this subsection, we discuss our training/test dataset generation procedure, the structure, and hyperparameters of our neural network, and the required number of training data during training. And we also provide a criterion for determining whether the neural network estimate is acceptable without knowing the true states.

The training and test data sets are formed by random k -local Hamiltonians and k -local measurements of corresponding ground states. For our 4-qubit case, the 2-local Hamiltonians are defined in Eq. (1). The parameter vector \vec{h} of random Hamiltonians are uniformly drawn from random normal distributions without uniform mean values and standard deviations. It is realized by applying function `np.random.normal` in Python. Similarly, for the 7-qubit case, Hamiltonian is defined in Eq. (4), and the corresponding parameter vector \vec{h} is generated by the same method. As the dashed lines in Fig. 1 shows, after getting random Hamiltonians H , we calculate the ground states $|\psi_H\rangle$ (the eigenvector corresponds to the smallest eigenvalue of H) and then get the 2-local measurements \mathbf{M} .

In this work, we use a fully connected feedforward neural network, which is famous as the first and most simple type of neural network.⁴² By fully connected, it means every neuron is connected to every other neuron

Table 2. Average fidelities on the test set by using different numbers of training data and epochs.

| 4-qubit (66-300-300-66) | | | | 7-qubit (75-150-300-300-150-75) | | | |
|-------------------------|----------------|----------------|----------------|---------------------------------|----------------|----------------|----------------|
| Training data | Epoch: 100 (%) | Epoch: 300 (%) | Epoch: 600 (%) | Training data | Epoch: 100 (%) | Epoch: 300 (%) | Epoch: 600 (%) |
| 500 | 69.1 | 56.2 | 56.9 | 1000 | 22.2 | 31.2 | 31.3 |
| 1000 | 78.1 | 66.8 | 58.7 | 10,000 | 93.8 | 91.5 | 88.7 |
| 5000 | 94.7 | 94.1 | 93.6 | 50,000 | 96.5 | 96.6 | 96.9 |
| 10,000 | 96.6 | 96.5 | 96.5 | 100,000 | 96.9 | 97.3 | 97.5 |
| 50,000 | 98.0 | 98.4 | 98.1 | 200,000 | 97.3 | 97.7 | 97.8 |
| 120,000 | 98.5 | 98.6 | 98.6 | 250,000 | 97.6 | 97.8 | 97.8 |

The batch size is 512, and the size of the test dataset is 5000. As the amount of training data increases, we find the average fidelity of estimated states and the true test states goes up, and the neural network reaches a certain performance after we fed sufficient training data. We also observe more training data requires more training epochs

Table 3. Average fidelities on the test set by using different batch sizes.

| 4-qubit (training data: 120,000) | | | |
|----------------------------------|----------------|----------------|----------------|
| Batch size | Epoch: 300 (%) | Epoch: 600 (%) | Epoch: 900 (%) |
| 512 | 98.6 | 98.6 | 98.7 |
| 1028 | 98.7 | 98.7 | 98.7 |
| 2048 | 98.5 | 98.7 | 98.7 |
| 7-qubit (training data: 250,000) | | | |
| Batch size | Epoch: 300 (%) | Epoch: 600 (%) | Epoch: 900 (%) |
| 512 | 97.8 | 97.8 | 97.9 |
| 1028 | 97.6 | 97.8 | 97.8 |

The size of test dataset is 5000. The optimal batch size for the 4-qubit case is 1024 and for the 7-qubit case is 512

in the next layer. Feedforward or acyclic, as the word indicated, means information only passes forward; the network has no cycle. Our machine-learning process is implemented using Keras,⁴³ which is a high-level deep learning library running on top of the popular machine-learning framework: Tensorflow.⁴⁴

As mentioned in the “Results” section, the true value of local measurements have been used as input to our neural network. The input is $\mathbf{M} = \{s_{m,n}^{(ij)} : s_{m,n}^{(ij)} = \text{Tr}(\text{Tr}_{(ij)} \rho \cdot B_{(m,n)}), B_{(m,n)} \in \mathbf{B}, 1 \leq i < j \leq 4, 1 \leq n, m \leq 4\}$. For the 4-qubit case, it is easy to see that \mathbf{M} has $3 \times 4 = 12$ single body terms and $C_4^2 \times 9 = 54$ 2-body terms. By arranging these 66 elements in \mathbf{M} into a row, we set it as the input of our neural network.

The output set to be the vector representation of the Hamiltonian \vec{h} , which also has 66 entries. For the 7-qubit 2-local case, where 2-body terms only appear on nearest qubits, the network takes 2-local measurements as input, and the number of neurons in the input layer is $7 \times 3 + 6 \times 3 \times 3 = 75$. The number of neurons in the output layer is also 75.

The physical aspect of our problem fixes the input and output layers. The principle for choosing the number of hidden layers is efficiency. While training networks, inspired by Occam’s Razor principle, we choose fewer layers and neurons when increasing them do not significantly increase the performance but increases the required training epochs. In our 4-qubit case, two hidden layers of 300 neurons have been inserted between the input layer and the output layer. In the 7-qubit case, we use four fully connected hidden layers with the following number of hidden neurons: 150-300-300-150. The activate function for each layer is rectified linear unit (ReLU),⁴⁵ which is a widely used non-linear activation function. We also choose the optimizer having the best performance in our problem over almost all the built-in optimizers in Tensorflow: AdamOptimizer (adaptive moment estimation).³⁰ The learning rate is set to be 0.001.

The whole training dataset has been split into two parts, 80% used for training, and 20% used for validation after each epoch. A new data set of

5000 data was used as the test set after training. The initial batch size was chosen as 512. As the amount of training data increases, the average fidelity of estimated states and the true test states goes up. The neural network reaches a certain performance after we fed sufficient training data. More training data requires more training epochs; however, replete epochs ebb the neural network performance due to over-fitting. Table 2 shows the average fidelities of using different training data and epochs. The first round of training locks down the optimal amount of training data, then we change the batch size and find the optimal epoch. We report the results for the second round training in Table 3. For the 4-qubit case, appropriate increase in the batch size can benefit the stability of training process, thus improves the performance of the neural network. Though, by choosing the batch size as 512 and 2048, the network can also reach the same performance with larger epochs, we chose the batch size as 1028, since more epochs require more training time. After the same attempting for the 7-qubit case, we find 512 a promising batch size.

The time cost for preparing the network involves two parts—generating training and testing data, and training the networks. Most of the time spending on data generating is to solve the ground states (eigenvector corresponding to the smallest eigenvalue) of randomly generated Hamiltonians. It takes roughly 5 min (2.2 h) to generate the whole data set for 4-qubit (7-qubit) by implementing `eigs` in MATLAB. With sufficient data in hand, the network-training procedure takes about 12 min (49 min) for 4-qubit (7-qubit).

As reported in the “Results” section, the fidelities of neural network estimates have a slight variation, for example, the fidelity of the 4-qubit case is range from 91.4% to 99.8%. One who uses this framework might wonder how precise the neural network outcome is compared to the true state. In contrast of the scenario when we are testing our framework theoretically, we do not have the true state in hand. Now it is natural to ask that how to determine whether the estimate is precise enough. Provisionally, we could solve this question in a straightforward way.

Based on ρ_{nn} , we compute $\mathbf{M}_{\text{pred}} = \{s_{m,n}^{(ij)} : s_{m,n}^{(ij)} = \text{Tr}(\text{Tr}_{(ij)} \rho_{nn} \cdot B_{(m,n)}), B_{(m,n)} \in \mathbf{B}, 1 \leq i < j \leq 4, 1 \leq n, m \leq 4\}$ and compare with the original \mathbf{M} . Root-mean-square-error (RMSE) between two variables \vec{x} and \vec{y} , defined as $\text{rmse}(\vec{x}, \vec{y}) = \sqrt{\frac{1}{d} \sum_{i=1}^d (x_i - y_i)^2}$, is a frequently used quantity to measure the closeness of \vec{x} and \vec{y} . In reality, how bad an error is also depends on the magnitude of the true value. That means, with the same RMSE, the larger the magnitude of the true value is the better the accuracy. A measure referring to the real value reveals more about how precise between an estimation and a expected outcome. We, therefore, define a quantity called relative RMSE, namely $\text{rmse}(\vec{x}, \vec{y}) = \sqrt{\frac{1}{d} \sum_{i=1}^d (x_i - y_i)^2} / \|\vec{y}\| = \text{rmse}(\vec{x}, \vec{y}) / \|\vec{y}\|$, where y is the true value and $\|\vec{y}\|$ is its l^2 -norm. The relative RMSE between \mathbf{M}_{pred} and \mathbf{M} is $\text{rmse}(\mathbf{M}_{\text{pred}}, \mathbf{M}) / \|\mathbf{M}\|$. By bounding the relative RMSE $< 0.2\%$, 4692 out of 5000 (93.8%) estimations of our 4-qubit network are acceptable, and the probability of these estimations having fidelities higher than 97% is 99.8%.

Comparison with the approximated MLE

The standard MLE^{46–49} is usually adopted to reconstruct a legal and full quantum state whose local information is closest to the measured results. It technically maximizes the likelihood of the estimate by the given data. When we make the Gaussian distribution and assume that all

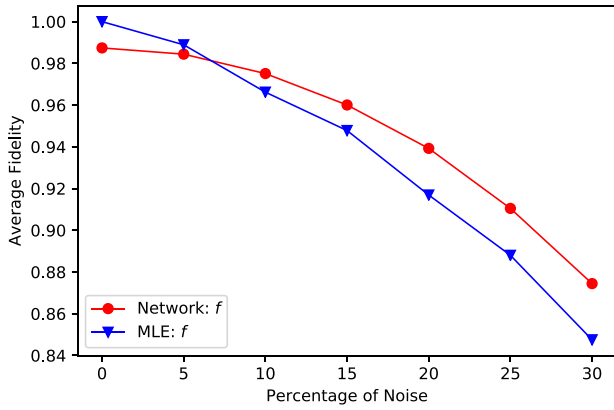


Fig. 5 Performances of the approximated MLE and our neural network framework under noisy inputs. The dots in the figure are the average fidelities f between the true state ρ and the noisy estimates ρ_{noise} corresponding to various percentage of noise. The red circles are estimated by the neural network, and the blue triangles are from the approximated MLE.

measurements have the same standard deviation, the MLE is approximately the least-squares tomography which minimizes the distance between the searched results and the measurement outcomes.⁵⁰ In this section, we make a comparison of efficiency, accuracy, and noise tolerance between the approximated MLE and our method.

With a personal computer,⁵¹ every single 4-qubit state takes about 1 min to compute for the approximated MLE. The estimating procedure of our method analyzed 5000 data in 2 min (about 0.024 s per data set) using the same computer. For the 7-qubit case, the approximated MLE requires about 168 min to converge for each single data point. Remarkably, our method can process 5000 data sets within <6 min (about 0.070 s per data set). This suggests that our method is substantially faster than the approximated MLE. We can reasonably expect that when the system size gets even larger, our computation time advantage will become more impressive.

In the 4-qubit cases, the approximated MLE can yield estimates with an average fidelity of 99.9%. In the 7-qubit cases, it can still achieve an average fidelity of 99.9%. Therefore in terms of accuracy, the approximated MLE slightly outperforms our method.

We also analyze the noise tolerance of the two methods by adding noise to the input measurements. The set of unbiased noise \vec{n} was generated according to the normal distribution with mean value 0 and standard deviation 1. The percentile noise vector $a\vec{n}$ is formed by multiplying the factor $a \in \{5\%, 10\%, 15\%, 20\%, 25\%, 30\%\}$ to the unbiased noise \vec{n} . By adding $a\vec{n}$ to the true measurements \mathbf{M} , we formed the noisy input $\mathbf{M} + a\vec{n}$. Suppose the approximated MLE or our neural network estimates the noisy output ρ_{noise} . We calculate the fidelities of the estimate ρ_{noise} with the true state ρ for 100 pairs of 4-qubit data. As depicted in Fig. 5, our method has better noise tolerance than the approximated MLE with the pure state constraint when noise >5% is added to the measurements of a pure state.

NMR states preparation

Our experimental procedure consists of three steps: initialization, evolution, and measurement. In this subsection, we discuss these three steps in details.

- (i) *Initialization*: The computational basis state $|0\rangle^{\otimes n}$ is usually chosen as the input state for quantum computation. Most of the quantum systems do not start from such an input state, so a proper initialization processing is necessary before applying quantum circuits. In NMR, the sample initially stays in the Boltzmann distribution at room temperature,

$$\rho_{\text{thermal}} = \mathcal{I}/16 + \epsilon(\sigma_z^1 + \sigma_z^2 + \sigma_z^3 + \sigma_z^4),$$

where \mathcal{I} is the 16×16 identity matrix and $\epsilon \approx 10^{-5}$ is the polarization. We cannot directly use it as the input state for

quantum computation, because such a thermal state is a highly mixed state.^{39,52} We instead create a so-called pseudo-pure state (PPS) from this thermal state by using the spatial averaging technique,^{39–41} which consists of applying local unitary rotations and using z-gradient fields to destroy the unwanted coherence. The form of the 4-qubit PPS can be written as

$$\rho_{0000} = (1 - \epsilon')\mathcal{I}/16 + \epsilon'|0000\rangle\langle 0000|.$$

Here, although the PPS ρ_{0000} is also a highly mixed state, the identity part \mathcal{I} neither changes under any unitary operations nor contributes to observable NMR signal. It means that we can focus on the deviated part $|0000\rangle\langle 0000|$ and consider $|0000\rangle\langle 0000|$ as the initial state of our quantum system. Finally, 4-qubit QST was performed to evaluate the quality of our PPS. We found that the fidelity between the perfect pure state $|0000\rangle$ and the experimentally measured PPS is about 98.7% by the definition of C in Eq. (3), where the raw PPS density matrix obtained directly from the experiment is negative. This sets a solid ground for the subsequent experiments.

- (ii) *Evolution*: In this step, we prepared the ground states of the given Hamiltonians using optimized pulses. The form of the considered Hamiltonian is chosen as Eq. (1). Here, the parameters $\omega_k^{(i)}$ and $J_{nm}^{(j)}$ mean the chemical shift and the J -coupling strength, respectively. In experiments, we create the ground states of different Hamiltonians by randomly changing the parameter set $(\omega_k^{(i)}, J_{nm}^{(j)})$. For the given Hamiltonian, the gradient ascent pulse engineering (GRAPE) algorithm^{53–56} is adapted to optimize a radio-frequency (RF) pulse to realize the dynamical evolution from the initial state $|0000\rangle$ to the target ground state. The GRAPE pulses are designed to be robust to the static field distributions and RF inhomogeneity, and the simulated fidelity is over 0.99 for each dynamical evolution.
- (iii) *Measurement*: In principle, we only need to measure the 2-local measurements to determine the original 4-qubit Hamiltonian through our trained network. Experimentally, we performed 4-qubit QST, which naturally includes the 2-local measurements after preparing these states,^{57–59} to evaluate the performance of our implementations. Hence, we can estimate the quality of the experimental implementations by computing the fidelity between the target ground state $\rho_{\text{th}} = |\psi_{\text{th}}\rangle\langle\psi_{\text{th}}|$ and the experimentally reconstructed density matrix ρ_{exp} .⁶⁰ By reconstructing states ρ_{nn} merely based on the experimental 2-local measurements, the performance of our trained neural network can be evaluated by comparing the experimental states ρ_{exp} with the states ρ_{nn} .

Finally, we attempt to evaluate the confidence of the expected results by analyzing the potential error sources in experiments. The infidelity of the experimental density matrix is mainly caused by some unavoidable factors in experiments, including decoherence effects, imperfections of the PPS preparation, and imprecision of the optimized pulses. From a theoretical perspective, we numerically simulate the influence of the optimized pulses and the decoherence effect of our qubits. Then we compare the fidelity computed in this manner with the ideal case to evaluate the quality of the final density matrix. As a numerical result, about 0.2% infidelity was created on average and the 1.2% error is related to the infidelity of the initial state preparation. Additionally, other errors can also contribute to the infidelity, such as imperfections in the readout pulses and spectral fitting.

The approximated MLE

We briefly describe the approximated MLE we used in the numerical simulation. The standard MLE^{48,49} is also a method to produce satisfactory results in recovering the full states from the experimental measurements. In general, the standard MLE can be divided into three steps.

- (i) Parameterize a density matrix in a legal way. Here, we describe a pure density matrix by

$$\rho(\vec{x}) = V(\vec{x})V^\dagger(\vec{x})/\text{Tr}(V(\vec{x})V^\dagger(\vec{x})).$$

V is a 2^N -dimensional vector with the parameters \vec{x} and the number of qubits N . $\rho(\vec{x})$ is a normalized and non-negative definite Hermitian density matrix under such a parameterization.

- (ii) Construct a likelihood function to be maximized. The measurements calculated from the parameterized density matrix $\rho(\vec{x})$ are $\text{Tr}(\text{Tr}_{(i,j)}\rho(\vec{x}) \cdot B_{(m,n)})$ with $B_{(m,n)} \in \mathbf{B}$, $1 \leq i < j \leq 4$ and $1 \leq n, m \leq 4$, and the total probability of $\rho(\vec{x})$ yielding the results closely to the

true measurements \mathbf{M} can be written as

$$P(\vec{x}) = \frac{1}{\mathcal{N}} \prod_{i,j,m,n} \exp \left[-\frac{\{\text{Tr}(\text{Tr}_{(i,j)} \rho(\vec{x}) \cdot B_{(m,n)}) - s_{m,n}^{(i,j)}\}^2}{2(\chi_{m,n}^{(i,j)})^2} \right],$$

where $\chi_{m,n}^{(i,j)}$ is the standard deviation of each measurement $s_{m,n}^{(i,j)}$ and \mathcal{N} is the normalization (Gaussian model). $P(\vec{x})$ is the likelihood function we need to maximize. If we assume the standard deviation is the same, the standard MLE is approximately least-squares tomography.⁵⁰ It is equivalent to minimize the following function:

$$\mathcal{F}(\vec{x}) = \sum_{i,j,m,n} \left[\text{Tr}(\text{Tr}_{(i,j)} \rho(\vec{x}) \cdot B_{(m,n)}) - s_{m,n}^{(i,j)} \right]^2.$$

Here, we ignore some constants which do not influence the optimization, e.g., the normalization factor \mathcal{N} . $\mathcal{F}(\vec{x})$ is the cost function that we minimize with the least-squares tomography.

- (iii) Minimize the cost function using some techniques. We use the function `lsqnonlin` of MATLAB with an initial guess and a default setting. It takes a while to optimize a sum of squares like $\mathcal{F}(\vec{x})$. Finally, quantum state $\rho(\vec{x})$ can be recovered from the parameters \vec{x} when the optimization is finished.

DATA AVAILABILITY

The experimental data and the source code that support the findings of this study are available from the corresponding author on reasonable request.

Received: 13 February 2019; Accepted: 10 September 2019;

Published online: 29 November 2019

REFERENCES

- D'Ariano, G. M., De Laurentis, M., Paris, M. G., Porzio, A. & Solimeno, S. Quantum tomography as a tool for the characterization of optical devices. *J. Opt. B* **4**, S127 (2002).
- Häffner, H. et al. Scalable multiparticle entanglement of trapped ions. *Nature* **438**, 643 (2005).
- Leibfried, D. et al. Creation of a six-atom 'schrödinger cat' state. *Nature* **438**, 639 (2005).
- Lvovsky, A. I. & Raymer, M. G. Continuous-variable optical quantum-state tomography. *Rev. Mod. Phys.* **81**, 299 (2009).
- Baur, M. et al. Benchmarking a quantum teleportation protocol in superconducting circuits using tomography and an entanglement witness. *Phys. Rev. Lett.* **108**, 040502 (2012).
- Klimov, A., Munoz, C., Fernández, A. & Saavedra, C. Optimal quantum-state reconstruction for cold trapped ions. *Phys. Rev. A* **77**, 060303 (2008).
- Hou, Z. et al. Full reconstruction of a 14-qubit state within four hours. *New J. Phys.* **18**, 083036 (2016).
- Cramer, M. et al. Efficient quantum state tomography. *Nat. Commun.* **1**, 149 (2010).
- Gross, D., Liu, Y.-K., Flammia, S. T., Becker, S. & Eisert, J. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.* **105**, 150401 (2010).
- Tóth, G. et al. Permutationally invariant quantum tomography. *Phys. Rev. Lett.* **105**, 250403 (2010).
- Li, J. et al. Optimal design of measurement settings for quantum-state-tomography experiments. *Phys. Rev. A* **96**, 032307 (2017).
- Lanyon, B. et al. Efficient tomography of a quantum many-body system. *Nat. Phys.* **13**, 1158 (2017).
- Baldwin, C. H., Deutsch, I. H. & Kalev, A. Strictly-complete measurements for bounded-rank quantum-state tomography. *Phys. Rev. A* **93**(5), 052105 (2016). <https://journals.aps.org/pr/abstract/10.1103/PhysRevA.93.052105>.
- Linden, N., Popescu, S. & Wootters, W. Almost every pure state of three qubits is completely determined by its two-particle reduced density matrices. *Phys. Rev. Lett.* **89**, 207901 (2002).
- Linden, N. & Wootters, W. The parts determine the whole in a generic pure quantum state. *Phys. Rev. Lett.* **89**, 277906 (2002).
- Diósi, L. Three-party pure quantum states are determined by two two-party reduced states. *Phys. Rev. A* **70**, 010302 (2004).
- Chen, J., Ji, Z., Ruskai, M. B., Zeng, B. & Zhou, D.-L. Comment on some results of erdahl and the convex structure of reduced density matrices. *J. Math. Phys.* **53**, 072203 (2012).
- Chen, J., Ji, Z., Zeng, B. & Zhou, D. From ground states to local hamiltonians. *Phys. Rev. A* **86**, 022339 (2012).
- Chen, J. et al. Uniqueness of quantum states compatible with given measurement results. *Phys. Rev. A* **88**, 012109 (2013).
- Qi, X.-L. & Ranard, D. Determining a local Hamiltonian from a single eigenstate. *Quant.* **3**, 159 (2019). <https://quantum-journal.org/papers/q-2019-07-08-159/>.
- Hou, S.-Y. et al. Determining system hamiltonian from eigenstate measurements without correlation functions. Preprint at arXiv:1903.06569 (2019).
- Qi, B. et al. Quantum state tomography via linear regression estimation. *Sci. Rep.* **3**, 3496 (2013).
- Zeng, B., Chen, X., Zhou, D.-L. & Wen, X.-G. Quantum information meets quantum matter—from quantum entanglement to topological phase in many-body systems. Preprint at arXiv:1508.02595 (2015). <https://www.springer.com/gp/book/9781493990825>.
- Kieferová, M. & Wiebe, N. Tomography and generative training with quantum boltzmann machines. *Phys. Rev. A* **96**, 062327 (2017).
- Torlai, G. et al. Neural-network quantum state tomography. *Nat. Phys.* **14**, 447 (2018).
- Chandra, P. & Singh, Y. Fault tolerance of feedforward artificial neural networks—a framework of study. In *Proc. International Joint Conference on Neural Networks*, Vol. 1, 489–494 (IEEE, 2003). <https://ieeexplore.ieee.org/document/1223395>.
- Singh, Y. & Chauhan, A.S. Neural networks in data mining. *J. Theor. Appl. Inf.* **5**, 37–42 (2009).
- Basheer, I. A. & Hajmeer, M. Artificial neural networks: fundamentals, computing, design, and application. *J. Microbiol. Methods* **43**, 3–31 (2000).
- Le Roux, N. & Bengio, Y. Representational power of restricted Boltzmann machines and deep belief networks. *Neural Comput.* **20**, 1631–1649 (2008).
- Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. *Proc. 3rd International Conference on Learning Representations (ICLR, 2015)*. <https://dblp.uni-trier.de/db/conf/iclr/iclr2015.html>.
- Reddi, S. J., Kale, S., and Kumar, S. On the convergence of Adam and beyond. *Proc. 6th International Conference on Learning Representations (ICLR, 2018)*.
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436 (2015).
- Nielsen, M. A. & Chuang, I. *Quantum Computation and Quantum Information* (Cambridge University Press, 2002).
- Fortunato, E. M. et al. Design of strongly modulating pulses to implement precise effective hamiltonians for quantum information processing. *J. Chem. Phys.* **116**, 7599–7606 (2002).
- Xin, T. et al. Nuclear magnetic resonance for quantum computing: techniques and recent achievements. *Chinese Phys. B* **27**, 020308 (2018).
- Vandersypen, L. M. & Chuang, I. L. NMR techniques for quantum control and computation. *Rev. Mod. Phys.* **76**, 1037 (2005).
- Jones, J. A., Vedral, V., Ekert, A. & Castagnoli, G. Geometric quantum computation using nuclear magnetic resonance. *Nature* **403**, 869 (2000).
- Xin, T. et al. Nmrcloudq: a quantum cloud experience on a nuclear magnetic resonance quantum computer. *Sci. Bull.* **63**, 17–23 (2018).
- Cory, D. G., Fahmy, A. F. & Havel, T. F. Ensemble quantum computing by NMR spectroscopy. *Proc. Natl Acad. Sci. USA* **94**, 1634–1639 (1997).
- Fahmy, A. F. & Havel, T. F. Nuclear magnetic resonance spectroscopy: an experimentally accessible paradigm for quantum computing. In *Quantum Computation and Quantum Information Theory: Reprint Volume with Introductory Notes for ISI TMR Network School, Physica D: Nonlinear Phenomena*, Vol. 120, Issues 1–2, 82–101 (1998). [https://doi.org/10.1016/S0167-2789\(98\)00046-3](https://doi.org/10.1016/S0167-2789(98)00046-3).
- Knill, E., Chuang, I. & Laflamme, R. Effective pure states for bulk quantum computation. *Phys. Rev. A* **57**, 3348 (1998).
- Schmidhuber, J. Deep learning in neural networks: an overview. *Neural Net.* **61**, 85–117 (2015).
- Chollet, F. et al. Keras. <https://keras.io> (2015).
- Abadi, M. et al. Tensorflow: a system for large-scale machine learning. In *Proc. of the 12th USENIX conference on Operating Systems Design and Implementation*, Vol. 16, 265–283 (2016). <https://dl.acm.org/citation.cfm?id=3026899>.
- Nair, V. & Hinton, G. E. Rectified linear units improve restricted boltzmann machines. In *ICML'10 Proc. of the 27th International Conference on International Conference on Machine Learning*, (eds Fürnkranz, J. & Joachims, T.) 807–814 (2010). <https://dl.acm.org/citation.cfm?id=3104425>.
- James, D. F., Kwiat, P. G., Munro, W. J. & White, A. G. On the measurement of qubits. In: *Asymptotic Theory of Quantum Statistical Inference: Selected Papers* (ed. Hayashi, M. (Japan Science and Technology Agency & University of Tokyo)) 509–538 (World Scientific, 2005).
- Hradil, Z. & Řeháček, J. Efficiency of maximum-likelihood reconstruction of quantum states. *Fortschr. Phys.* **49**, 1083–1088 (2001).
- Řeháček, J., Hradil, Z., Knill, E. & Lvovsky, A. Diluted maximum-likelihood algorithm for quantum tomography. *Phys. Rev. A* **75**, 042108 (2007).
- Paris, M. & Rehacek, J. *Quantum State Estimation*, Vol. 649 (Springer Science & Business Media, 2004).

50. Acharya, A., Kypraios, T. and Guță, M. A comparative study of estimation methods in quantum tomography. *J. Phys. A: Math Theor.* **52**(23), 234001, (2019). <https://iopscience.iop.org/article/10.1088/1751-8121/ab1958>.
51. MacBook Pro, Processor: 2.3 GHz Intel Core i5, Memory: 8 GB.
52. Gershenfeld, N. A. & Chuang, I. L. Bulk spin-resonance quantum computation. *Science* **275**, 350–356 (1997).
53. Boulant, N., Edmonds, K., Yang, J., Pravia, M. & Cory, D. Experimental demonstration of an entanglement swapping operation and improved control in NMR quantum-information processing. *Phys. Rev. A* **68**, 032305 (2003).
54. Khaneja, N., Reiss, T., Kehlet, C., Schulte-Herbrüggen, T. & Glaser, S. J. Optimal control of coupled spin dynamics: design of nmr pulse sequences by gradient ascent algorithms. *J. Magn. Reson.* **172**, 296–305 (2005).
55. Ryan, C., Negrevergne, C., Laforest, M., Knill, E. & Lafamme, R. Liquid-state nuclear magnetic resonance as a testbed for developing quantum control methods. *Phys. Rev. A* **78**, 012328 (2008).
56. Lu, D. et al. Enhancing quantum control by bootstrapping a quantum processor of 12 qubits. *npj Quantum Inf.* **3**, 45 (2017).
57. Leskowitz, G. M. & Mueller, L. J. State interrogation in nuclear magnetic resonance quantum-information processing. *Phys. Rev. A* **69**, 052302 (2004).
58. Lee, J.-S. The quantum state tomography on an NMR system. *Phys. Lett. A* **305**, 349–353 (2002).
59. Li, J. et al. Optimal design of measurement settings for quantum-state-tomography experiments. *Phys. Rev. A* **96**, 032307 (2017).
60. Altepeter, J. B., Jeffrey, E. R. & Kwiat, P. G. Photonic state tomography. *Adv. Atom. Mol. Opt. Phys.* **52**, 105–159 (2005).
61. Song, C. et al. 10-qubit entanglement and parallel logic operations with a superconducting circuit. *Phys. Rev. Lett.* **119**, 180511 (2017).
62. Krizhevsky, A., Sutskever, I. & Hinton, G.E. Advances in Neural Information Processing Systems 25. In *Advances in Neural Information Processing Systems* (eds Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q.) 1097–1105 (Curran Associates, Inc., 2012). <https://dl.acm.org/citation.cfm?id=3065386>.
63. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 770–778 (2016).
64. Flammia, S. T., Gross, D., Liu, Y.-K. & Eisert, J. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New J. Phys.* **14**, 095022 (2012).
65. Riofrío, C. et al. Experimental quantum compressed sensing for a seven-qubit system. *Nat. Commun.* **8**, 15305 (2017).

ACKNOWLEDGEMENTS

We thank Yi Shen for helpful discussions. G.L. is grateful to the following funding sources: the National Natural Science Foundation of China (11175094); National Basic Research Program of China (2015CB921002). T.X., D.L., and J.L. are supported by the National Natural Science Foundation of China (Grants nos. 11605153, 11605005, 11875159, U1801661, 11905099, and 11975117), Science, Technology and Innovation

Commission of Shenzhen Municipality (Grants nos. ZDSYS20170303165926217 and JCYJ20170412152620376), Guangdong Innovative and Entrepreneurial Research Team Program (Grant no. 2016ZT06D348). T.X. is also supported by Guangdong Basic and Applied Basic Research Foundation (No. 2019A1515011383). N.C. and B.Z. acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC), Canadian Institute for Advanced Research (CIFAR) and Chinese Ministry of Education (20173080024).

AUTHOR CONTRIBUTIONS

B.Z. conceived the idea of this paper. S.L. and N.C. wrote and implemented the computer code and simulations. T.X. accomplished the NMR experiments. B.Z. supervised the project. T.X., S.L., and N.C. wrote the manuscript with feedback from all authors.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to J.L. or B.Z.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019