CHIN. PHYS. LETT. Vol. 38, No. 4 (2021) 040303

Bidirectional Information Flow Quantum State Tomography

Huikang Huang(黄惠康)[1], Haozhen Situ(司徒浩臻)[1*], and Shenggen Zheng(郑盛根)[2*]

[1]*College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China*
[2]*Circuits and Systems Research Center, Peng Cheng Laboratory, Shenzhen 518055, China*

(Received 25 December 2020; accepted 22 February 2021; published online 6 April 2021)

The exact reconstruction of many-body quantum systems is one of the major challenges in modern physics, because it is impractical to overcome the exponential complexity problem brought by high-dimensional quantum many-body systems. Recently, machine learning techniques are well used to promote quantum information research and quantum state tomography has also been developed by neural network generative models. We propose a quantum state tomography method, which is based on a bidirectional gated recurrent unit neural network, to learn and reconstruct both easy quantum states and hard quantum states in this study. We are able to use fewer measurement samples in our method to reconstruct these quantum states and to obtain high fidelity.

DOI: 10.1088/0256-307X/38/4/040303

As a fundamental research topic in quantum information processing, quantum state tomography (QST) has always been a concern. As a data-driven problem, QST aims at obtaining information of a quantum system as much as possible and reconstructing the quantum state density matrix through effective quantum state measurement. Actually, accurate QST is impractical,[1] especially cases in high-dimensional quantum many-body systems. One needs exponential complexity to describe a generic quantum many-body system. Even for small-scale quantum systems to be tomographic, it still requires a lot of resources. Therefore, we want to reconstruct the most accurate quantum system through as less measurement samples as possible. In other words, we want to capture the associations between these limited numbers of quantum state measurement samples and get more information to serve tomography.

After years of investigation, several key technologies have been developed in QST, including, but not limited to, the following methods, compressed sensing tomography,[2,3] permutationally invariant tomography[4,5] and tomographic schemes based on tensor networks.[6] In recent years, quantum machine learning has attract a great deal of attention in quantum computing.[7–9] Meanwhile, machine learning has made great progress in assisting the research of quantum physics problems.[10–15] Meanwhile, QST driven by neural network generative models has also received widespread attention and some relevant research results have appeared. For example, based on the probabilistic undirected graph model, Torlai *et al.*[16] used the restricted Boltzmann machine (RBM) QST method to learn amplitude and phase of quantum state and reconstruct quantum state wavefunction. Based on the powerful autoregressive model recurrent neural networks (RNNs), Carrasquilla *et al.*[17] introduced RNN-QST, which is able to use the informationally complete (IC) positive-operator valued measures (POVMs) samples, to reconstruct quantum states with high classical fidelity. There is also a transformer[18] QST method based on attention mechanism-based generative network, which reconstructs mixed state density matrix of a noisy quantum state.[19] Moreover, there are some other QST methods driven by generative models.[20,21] All of these methods demonstrate that machine learning techniques can effectively deal with specific quantum states.

In this Letter, we propose to use a bidirectional recurrent neural network (BiRNN) generative model based on contextual semantics to perform QST. By slicing quantum state measurement samples as time series information flow, we are able to make full use of the contextual semantics of these messages to perform QST by this bidirectional neural network. At the same time, we propose a network training standard to conduct early stopping, which can help one to find effectively better training model using fewer training samples. These methods enable us to use fewer quantum state measurement samples than RNN[17] and AQT[19] neural network tomography to achieve over 99% classical fidelity on GHZ state. We test our method in dealing with "easy quantum states" and "hard quantum states" that have different sampling difficulties.[22] Finally, we have a brief discussion about why this QST method can effectively process some specific quantum states.

Method. Our method belongs to unsupervised machine learning. The training samples are produced

Supported by the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2020A1515011204), and the National Natural Science Foundation of China (Grant No. 61602532).
*Corresponding authors. Email: situhaozhen@gmail.com; zhengshg@pcl.ac.cn
© 2021 Chinese Physical Society and IOP Publishing Ltd

based on the IC-POVMs mentioned in Ref. [17]. The IC-POVM operators which we use here are Pauli-4 operators. Each N-qubit quantum state measurement sample is denoted by $\boldsymbol{a}$. We slice each of them into $T = N$ parts as time series information flow, i.e., $\boldsymbol{a} = [a_1, a_2, \ldots, a_N]$, where $a_i \in \{0, 1, 2, 3\}$. The probability distribution corresponding to the Pauli-4 IC-POVM is denoted as $\boldsymbol{P} = \{P(\boldsymbol{a})\}_{\boldsymbol{a}}$ with $P(\boldsymbol{a}) \geq 0$ and $\sum_{\boldsymbol{a}} P(\boldsymbol{a}) = 1$.

Bidirectional Recurrent Neural Networks. Recurrent neural network can process effectively time series data in natural language processing (NLP) problems,[23–25] which is extended to a bidirectional model by Schuster and Paliwal.[26] It can be trained without the limitation of using information just up to a preset future frame. In our experiments, we use a bidirectional gated recurrent unit (BiGRU) neural network, a network model developed by combining the BiRNN network model and a gated recurrent unit.[27] BiRNN has been proved to capture effectively the relationship of contextual semantics in natural language processing. GRU is able to prevent the gradient disappearance and gradient explosion of RNN. Furthermore, GRU uses less network parameters and it is easier to be trained than long short-term memory (LSTM) network.[28]

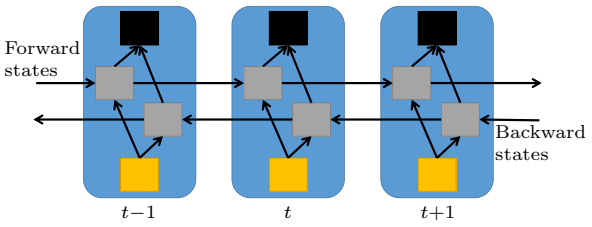

Fig. 1. The framework of general bidirectional recurrent neural networks shown unfolded in time for three time steps. The yellow blocks represent the inputs of quantum state samples and the gray blocks are the hidden states learned from forward and backward information flow of quantum state samples. The black blocks are the quantum state samples of next time step, which are fitted and predicted by the yellow and gray blocks.

In order to process QST by neural network with less quantum state measurement samples, we need to make better use of the limited training samples. According to the general bidirectional model, we divide the training samples into forward information flow and backward information flow. Given a batch of $N_{\rm s}$ quantum state measurement samples $\boldsymbol{E} = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \ldots, \boldsymbol{a}_{N_{\rm s}}\}$, we slice $\boldsymbol{a}_i$ according to time series T, $1 \leq t \leq T$. The training procedure of this unfolded bidirectional network over time can be summarized as follows:

1) Forward flow training: By receiving all input data for one time slice from $t = 1$ to T, the hidden layer outputs $\overrightarrow{h}_t = f(a_t, \overrightarrow{h}_{t-1})$.

2) Backward flow training: By receiving time series data a_i from $t = T$ to 1, the hidden layer outputs $\overleftarrow{h}_t = f(a_t, \overleftarrow{h}_{t+1})$.

3) Consolidate information flow: Outputs of the network training through the joint bidirectional information flow $P(a_{t+1}) = f(\overrightarrow{h}_t, \overleftarrow{h}_t)$.

This procedure is a general bidirectional model framework and the activation function f has different forms in different recurrent units. See Fig. 1 for the framework of general BiRNN for three time steps. The model is trained by minimizing the log-likelihood loss function L over θ, which is as follows:

$$L(\theta) = -\frac{1}{T}\sum_{t=1}^{T} \log[P_\theta(a_t)],$$

where θ is a set of model parameters.

Easy and Hard Quantum States. Rocchetto *et al.*[22] proposed to classify quantum states based on the hardness of sampling probability distribution of measurement result in a given basis. The Greenberger–Horne–Zeilinger (GHZ) quantum state has been discussed in many articles, which is a highly non-classical state, specified by $|\Psi_{\rm GHZ}\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes N} + |1\rangle^{\otimes N})$. As a simple quantum pure state, GHZ state is widely used in quantum communication protocols. This kind of states can be sampled easily in a large quantum system and we will learn and reconstruct it in this Letter.

We consider a kind of random pure states as hard states which are generated by normalizing a 2^n dimensional complex vector drawn from unit sphere according to Haar measure.[22] This kind of hard states cannot be prepared easily on any realistic quantum computing device with a polynomial large circuit. It means that quantum measurement samples can only be obtained for few-qubit states. Up to now, the performances of reconstructing the probability distributions,[22] or the quantum state wavefunctions,[16] of these states are not very ideal.

Learning Standards. Quantum fidelity F is a comprehensive measure of quantum state reconstruction. Let $\varrho_{\rm GHZ}$ and $\varrho_{\rm BiGRU}$ be two N-qubit quantum states, the quantum fidelity between them is defined as

$$F(\varrho_{\rm GHZ}, \varrho_{\rm BiGRU}) = {\rm tr}\left(\sqrt{\sqrt{\varrho_{\rm GHZ}}\varrho_{\rm BiGRU}\sqrt{\varrho_{\rm GHZ}}}\right).$$

We say that $\varrho_{\rm BiGRU}$ will be a good representation of $\varrho_{\rm GHZ}$ if the quantum fidelity $F \geq 1 - \epsilon$ for a small $\epsilon > 0$. However, due to exponential large dimension of the representation of density matrices, we can only calculate the quantum fidelity for reconstruction of small quantum systems. For large many-body quantum systems, we use classical fidelity $F_{\rm c}$ to evaluate the reconstruction performance, which is defined as

$$F_{\rm c}(\boldsymbol{P}_{\rm GHZ}, \boldsymbol{P}_{\rm BiGRU}) = \mathbb{E}_{\boldsymbol{a}\sim\boldsymbol{P}_{\rm BiGRU}}\sqrt{\frac{P_{\rm GHZ}(\boldsymbol{a})}{P_{\rm BiGRU}(\boldsymbol{a})}}.$$

F_c is a standard measure of proximity between two distributions. $P_{\mathrm{GHZ}}(\boldsymbol{a})$ and $P_{\mathrm{BiGRU}}(\boldsymbol{a})$ are the exact probability and the probability generated by BiGRU that correspond to the same quantum state measurement sample $\boldsymbol{a}$. Theoretically, $F_c(\boldsymbol{P}_{\mathrm{GHZ}}, \boldsymbol{P}_{\mathrm{BiGRU}}) = 1$ only if $\boldsymbol{P}_{\mathrm{GHZ}} = \boldsymbol{P}_{\mathrm{BiGRU}}$. In general, F_c will serve as an upper bound of F, meaning that a small error in F_c will be amplified in F.[17]

The stop criterion is a difficult problem in training generative models. In this study, we observe that a large number of training epoch which leads to overfitting problem. Even if the loss function continues to decrease, the fidelity may decrease. Therefore, we cannot stop the training according to the loss function itself. Especially for the tomography of an unknown quantum state, how to use the fewest quantum state measurement samples and decide when to stop the network training to obtain the best fidelity will be the key issue. We propose a simple but effective metric for the stop criterion, i.e., degree of loss function's fluctuation: $D = \{d_i\}$ with $d_i = \mathrm{loss}_i - \mathrm{loss}_{i-1}$, which is the difference between the losses of two consecutive epochs. This metric can reflect the training quality of the network and help us to find effectively the best quantum state reconstruction model with the fewest training samples. The detailed process is to select the range with the least fluctuation: $D = [d_j, \cdots, d_{j+n}]$ and choose the trained network with the smallest loss in this range $\mathrm{BiGRU}_{\min(d_j,\cdots,d_{j+n})}$, where n is an appropriate value, e.g., 50.

Numerical Results. Deep learning framework Pytorch is used in our numerical experiment. We use a three-layer BiGRU in all experiments. The training is performed using backpropagation and the Adam optimizer,[29] with initial learning rate of 0.001. The classical fidelity is calculated by 50000 generative samples drawn from the trained network.

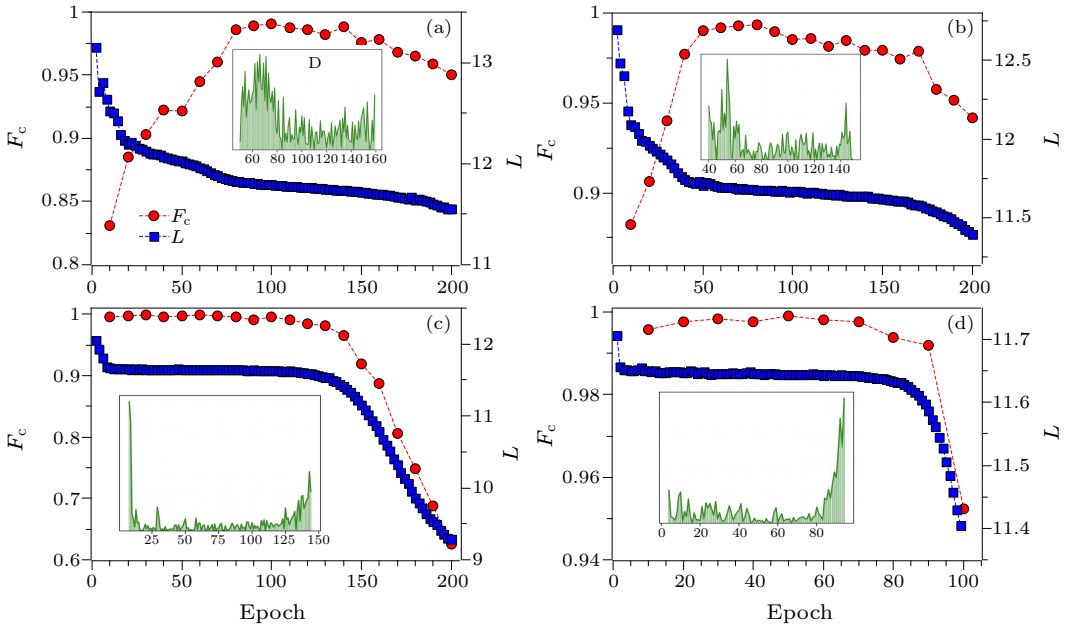

Fig. 2. The BiGRU generative model reconstructs 10-qubit GHZ quantum states. Subplots a, b, c, d show the results that use 500, 1000, 5000, and 30000 training samples, respectively. Red line and blue line are classic fidelity F_c and loss function L. Green line is the degree of fluctuation D. We can observe clearly that the following points: (1) F_c first increases, then drops in the training process. This phenomenon can be explained by the overfitting problem. (2) Throughout the whole training process, when F_c reaches its maximum, the fluctuation D of the loss function L is relatively small, thus D can be used to overcome the overfitting problem. (3) As the number of training samples increases, the fluctuation D is able to judge the performance of the model more accurately.

We now turn our attention to the reconstruction of GHZ state. We show the numerical results of reconstruction of 10-qubit GHZ quantum state using different numbers of training samples in Fig. 2. It is found that no matter how large the number of training samples is, the training model will overfit gradually as the number of training iterations increases. Even if we try to exceed the number of training samples mentioned by Carrasquilla *et al.* in Ref. [17], as shown in Fig. 2(d), overfitting occurs also and results in poor reconstruction fidelity. In other words, how to prevent overfitting is the key point in this neural network QST technology. As far as we know, referencing the degree of loss function's fluctuation can assist us to find better models.

In order to testify our method, we reconstruct GHZ states with system sizes ranging from $N = 10$ to 50 qubits. The necessary number of training samples and the classical fidelities are shown in Fig. 3. Compared with the state-of-the-art RNN-QST[17] and AQT[19] methods, our BiGRU-QST method uses almost the fewest number of measurement samples to achieve over

99% fidelity. We can also see a linear growth of the number of training samples with respect to the number of qubits.

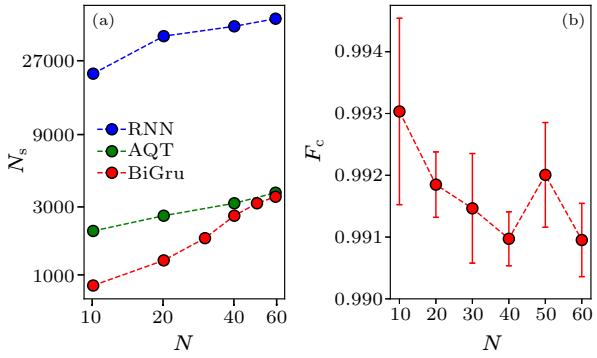

Fig. 3. The performance of reconstructing GHZ quantum states by BiGRU. Based on the fluctuation D to execute early stopping, we test BiGRU to reconstruct GHZ quantum states with 10, 20, 30, 40, 50, 60 qubits. For each GHZ state, we obtain a network model and then sample five groups of generative samples to calculate five F_c values. Each group contains 50000 generative samples. We require all five F_c values to achieve over 99%. (a) The log-log plot shows the necessary number of training samples (N_s = 900, 1300, 1800, 2500, 3000, 3300), which is performed by BiGRU-QST. Compared with RNN-QST and AQT, our method needs the fewest training samples to reach over 99% F_c. (b) The plot shows the average F_c over 5 calculations which is reconstructed by BiGRU-QST. The error bar stands for standard deviation.

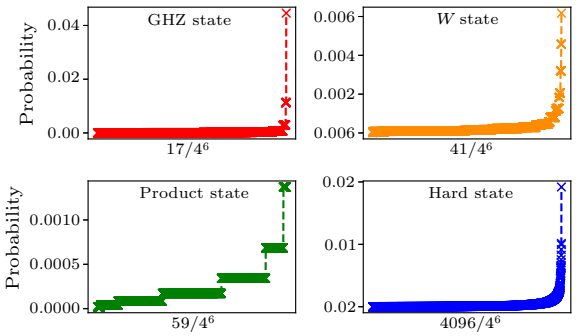

Fig. 4. The distribution of measurement results of 6-qubit GHZ, W, product and hard states. The probability values shown have been sorted. There are 17, 41, 59 and 4096 different probability values for these states, respectively. More probability values will make the state more difficult to be reconstructed.

When we use BiGRU to perform QST on the hard state, we find a limitation of this method. In our experiments, we prepare synthetic datasets mimicking experimental measurements of the 6-qubit exact distributions of quantum states $\boldsymbol{P}_{\rm GHZ}$, $\boldsymbol{P}_{\rm Hard}$, $\boldsymbol{P}_W$ and $\boldsymbol{P}_{\rm product}$. The W state is written as $|\Psi_W\rangle = \frac{1}{\sqrt{N}}(|100\cdots\rangle + |010\cdots\rangle + \cdots + |\cdots001\rangle)$ and the product state is a fully polarized state $|\Psi_{\rm Product}\rangle = H^{\otimes N}|0\rangle^{\otimes N} = \frac{1}{\sqrt{2^N}}\sum_{i=0}^{2^N-1}|i\rangle$, where $|i\rangle$ is the computational basis. The probability distribution of measurement results of 6-qubit quantum system under the pauli-4 measurement operators has $4^6 = 4096$ probability values. The probability distributions after sorting for these four states are shown in Fig. 4. Due to the relatively simple structure of W state, GHZ state and product state, these 4^6 probability values have only 41, 17 and 59 different values. In other words, these few probability values will occupy most of the quantum information and directly affect tomographic fidelity. In numerical results, we investigate F_c as a function of N_s, the reconstructed results are shown in Fig. 5. We find that F_c of product state, W state and GHZ state reach quickly 99% fidelity with fewer measurement samples and F_c of the hard state increases much more slowly because the probability values of its measurement distribution are all different. In short word, we can get an approximate rule for the QST method driven by these time series generative models, that is, the number of different probability values will seriously affect reconstruction results. The efficiency of QST will be better when the number of different measurement probability values is smaller.

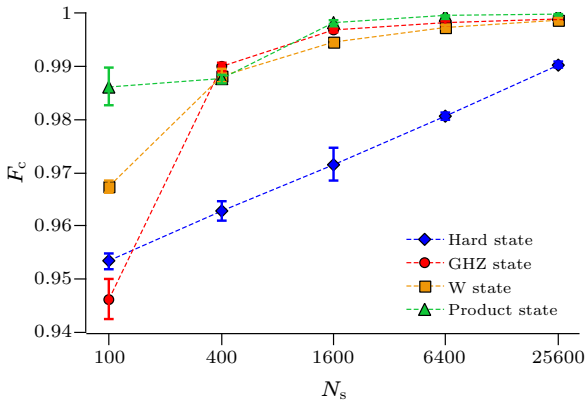

Fig. 5. Using the BiGRU generative model to reconstruct 6-qubit Product state, hard state, GHZ state and W state with different numbers of training samples. For product state, GHZ and W states, we use five groups of training samples to train five models and sample 50000 generative samples from each models to calculate five F_c. For the hard state, we train five models using five groups of training samples produced by five different random pure states. The error bar stands for standard deviation. We find that the standard deviation will be smaller as the number of training samples increases.

In summary, the purpose of quantum state tomography is to reconstruct as complete quantum state as possible through limited quantum state measurement samples. Actually, it is impractical to perform QST in large quantum systems due to exponential "curse of dimensionality" inherent to description of quantum states. Therefore, how to alleviate these scaling issues is what we are pursuing. The QST driven by neural networks helps us process specific quantum states to achieve high fidelity reconstructions with only a small amount of samples. In our work, we propose a BiGRU neural network as a generative model, which can effectively capture contextual associations in the field of

natural language processing. In terms of QST by BiGRU, we encode quantum state measurement samples into bidirectional information flow, and use BiGRU to capture the internal correlation of quantum information, and rely on early stopping skill to achieve better results compared with other neural network QST within the scope of our knowledge.

We show through numerical experiments that these time series generative models are capable of representing, especially product state, GHZ state and W state, whose probability distributions have very few different values. BiGRU can use very few measurement samples to realize QST with high fidelity, over 99%. On the other hand, the reconstruction of hard states with many different measurement probability values needs more samples. Although performing hard state QST has no significant reduction of measurement samples, this method can be effectively used in QST, because the fidelity increases as the number of samples increases, which means this method is credible.

We have given a brief discussion and experimental results to show that these kinds of easy quantum states with few probability values can have a good QST performance by neural network method. In order to obtain more general conclusions, more experimental verification and theoretical explanation are needed. In order to improve the versatility of this method in quantum state tomography technology, as further works, we will investigate QST and find out relationship between the number of different probability values and the necessary number of measurement samples. Furthermore, we will also study characteristics of the probability distribution which may have a direct impact on neural network based tomography.

The authors are thankful to the anonymous referees and editors for their comments and suggestions that greatly help to improve the quality of the manuscript. We thank Dr. Zhiming He for discussion to improve the presentations of some figures.

References

[1] Häffner H, Hänsel W, Roos C F, Benhelm J, Chek-al K D, Chwalla M, Körber T, Rapol U D, Riebe M, Schmidt P O, Becher C, Gühne O, Dür W and Blatt R 2005 *Nature* **438** 643
[2] Gross D, Liu Y K, Flammia S T, Becker S and Eisert J 2010 *Phys. Rev. Lett.* **105** 150401
[3] Yin Q, Xiang G Y, Li C F and Guo G C 2018 *Chin. Phys. Lett.* **35** 070302
[4] Tóth G, Wieczorek W, Gross D, Krischek R, Schwemmer C and Weinfurter H 2010 *Phys. Rev. Lett.* **105** 250403
[5] Moroder T, Hyllus P, Tóth G, Schwemmer C, Niggebaum A, Gaile S, Gühne O and Weinfurter H 2012 *New J. Phys.* **14** 105001
[6] Baumgratz T, Gross D, Cramer M and Plenio M B 2013 *Phys. Rev. Lett.* **111** 020401
[7] Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N and Lloyd S 2017 *Nature* **549** 195
[8] Situ H, He Z, Wang Y, Li L and Zheng S 2020 *Inf. Sci.* **538** 193
[9] He Z, Li L, Zheng S, Li Y and Situ H 2021 *New J. Phys.* **23** 033002
[10] Carleo G and Troyer M 2017 *Science* **355** 602
[11] Hartmann M J and Carleo G 2019 *Phys. Rev. Lett.* **122** 250502
[12] Deng D L, Li X and Das S S 2017 *Phys. Rev. X* **7** 021021
[13] Cai Z and Liu J 2018 *Phys. Rev. B* **97** 035116
[14] Fournier R, Wang L, Yazyev O V and Wu Q 2020 *Phys. Rev. Lett.* **124** 056401
[15] Yao T S, Tang C Y, Yang M, Zhu K J, Yan D Y, Yi C J, Feng Z L, Lei H C, Li C H, Wang L, Wang L, Shi Y G, Sun Y J and Ding H 2019 *Chin. Phys. Lett.* **36** 068101
[16] Torlai G, Mazzola G, Carrasquilla J, Troyer M, Melko R and Carleo G 2018 *Nat. Phys.* **14** 447
[17] Carrasquilla J, Torlai G, Melko R G and Aolita L 2019 *Nat. Mach. Intell.* **1** 200
[18] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I 2017 arXiv:1706.03762v5 [cs.CL]
[19] Cha P, Ginsparg P, Wu F, Carrasquilla J, McMahon P L and Kim E A 2020 arXiv:2006.12469v1 [quant-ph]
[20] Ahmed S, Sanchez M C, Nori F and Frisk K A 2020 arXiv:2008.03240v1 [quant-ph]
[21] Luchnikov I A, Ryzhov A, Stas P J, Filippov S N and Ouerdane H 2019 *Entropy* **21** 1091
[22] Rocchetto A, Grant E, Strelchuk S, Carleo G and Severini S 2018 *npj Quantum Inf.* **4** 28
[23] Sutskever I, Vinyals O and Le Q V 2014 *Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS'14)* (Cambridge, MA: MIT Press) vol 2 p 3104
[24] Wu Y, Schuster M, Chen Z, Le Q, Macherey W, Krikun M, Cao Y, Gao Q, Macherey K, Klingner J, Shah A, Johnson M, Liu X, Kaiser U, Gouws S, Kato Y, Kudo T, Kazawa H and Dean J 2016 arXiv:1609.08144v2 [cs.CL]
[25] Chiu C, Sainath T N, Wu Y, Prabhavalkar R, Nguyen P, Chen Z, Kannan A, Weiss R J, Rao K, Gonina E, Jaitly N, Li B, Chorowski J and Bacchiani M 2018 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (15–20 April 2018, Calgary, AB, Canada) p 4774
[26] Schuster M and Paliwal K K 1997 *IEEE Trans. Signal Process.* **45** 2673
[27] Cho K, Van Merrienboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H and Bengio Y 2014 arXiv:1406.1078v3 [cs.CL]
[28] Hochreiter S and Schmidhuber J 1997 *Neural Comput.* **9** 1735
[29] Kingma D and Ba J 2014 arXiv:1412.6980v9 [cs.LG]