Reconstruction of a Photonic Qubit State with Quantum Reinforcement Learning

Shang Yu,[1,2] F. Albarrán-Arriagada,[3,4] J. C. Retamal,[3,4] Yi-Tao Wang,[1,2] Wei Liu,[1,2] Zhi-Jin Ke,[1,2] Yu Meng,[1,2] Zhi-Peng Li,[1,2] Jian-Shun Tang,[1,2,*] E. Solano,[5,6,7] L. Lamata,[5] Chuan-Feng Li,[1,2,†] and Guang-Can Guo[1,2]

[1]*CAS Key Laboratory of Quantum Information, University of Science and Technology of China, Hefei 230026, People's Republic of China*
[2]*CAS Center For Excellence in Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei 230026, People's Republic of China*
[3]*Departamento de Física, Universidad de Santiago de Chile (USACH), Avenida Ecuador 3493, 9170124, Santiago, Chile*
[4]*Center for the Development of Nanoscience and Nanotechnology 9170124, Estación Central, Santiago, Chile*
[5]*Department of Physical Chemistry, University of the Basque Country UPV/EHU, Apartado 644, 48080 Bilbao, Spain*
[6]*IKERBASQUE, Basque Foundation for Science, Maria Diaz de Haro 3, 48013 Bilbao, Spain*
[7]*Department of Physics, Shanghai University, 200444 Shanghai, China*

(Dated: August 29, 2018)

Extracting information from an unknown quantum state is an important task in quantum information. For the most general quantum state tomography, one has to measure the averages of a set of observables for reconstructing the density matrix. However, this method requires enough copies of the unknown state to be provided. In this work, we perform an experiment to reconstruct an unknown photonic quantum state with a limited amount of copies. We employ a quantum reinforcement learning approach to adapt one qubit state, an "agent", to an unknown quantum state, an "environment", by successive single-shot measurements and feedback, in order to achieve maximum overlap. Our experimental learning device, composed of a quantum photonics setup, can adjust the corresponding parameters to rotate the agent system based on the measurement outcomes "0" or "1" on the environment (i.e., reward/punishment signals). Our results show that, when assisted by such a quantum machine learning technique, the fidelities of the deterministic single-photon agent states can achieve over 88% under a proper reward/punishment ratio within 50 iterations. Therefore, only 50 photonic copies of the environment system are consumed. Meanwhile, these results all surpass the fidelities obtained by standard quantum tomography methods with the same number of photons as a resource. This protocol offers a tool for reconstructing an unknown quantum state when only limited copies are provided, and can also be extended to high dimensions, multipartite, and mixed quantum state scenarios.

Introduction.—Human intelligence is one of the most mysterious things which has not been comprehensively understood to date, while attracting much interest both for the exploration of its working mechanism and for potential applications [1, 2]. For the past fifty years, scientists have attempted to produce intelligent machines, which can learn to perform complex tasks, in the realm of the so-called "artificial intelligence (AI)". Although still far from reaching human levels, the drastic development of computer science and semiconductor industry has tremendously improved machine capacities. The most remarkable event in this area recently is the success of AlphaGo [3] in beating the Go world champion. Besides this smart machine playing Go, many other kinds of intelligent devices have been developed to replace traditional ones in various applications [4].

Machine learning [5], as a subtopic within AI realm, has already become a powerful tool for data mining, pattern recognition, among others. Meanwhile, there are many recent works combining ML techniques with quantum information tools [6–27]. These include expressing and witnessing quantum entanglement by artificial neural networks (ANN) [22, 26], analyzing and restructuring a quantum state by restricted Boltzmann machines (RBM) [20, 24, 27], as well as detecting quantum change points, and learning Hamiltonians by Bayesian inference [19, 21, 23, 25]. Besides the above tasks, here, we focus on the topic of quantum reinforcement learning (QRL). In this context, a quantum system, named "agent", can learn how to behave correctly through interactions with a quantum "environment" state, and the "reinforcement" signals—rewards or punishments [18], obtained from interactions between both. The quantum versions of reinforcement learning (RL) have been recently proposed [10, 15, 18, 28–30], being inspired on RL algorithms existing since the beginning of AI.

In our quantum photonics experiments, we explore whether a quantum state adapts onto another unknown quantum state via limited single-shot measurements, where only one photon is measured in each iteration, such that the "reinforcement" signal can be detected. The landmark principle that a single unknown quantum state cannot be cloned [31], prevents one from exactly copying a quantum state in a single-shot. However, here, we are interested in how to obtain the largest amount of information from the unknown state with minimal resources, i.e., with minimal number of identical copies, assisted by a QRL algorithm.

In this work, we employ the quantum reinforcement learning protocol of Ref. [30] to reconstruct an unknown

quantum state with minimal number of photonic copies. We employ 50 iterations, namely, we only consume 50 copies of photons of the environment system, while the fidelity of the agent system reaches over 88% under a proper reward/punishment ratio. Interestingly, we can see the fidelities obtained by our QRL protocol all surpass the one obtained by quantum tomography with the same number of photonic copies. This puts in evidence the advantage of quantum machine learning techniques under a limited-resource scenario. We point out that the primary goal of our protocol is not to achieve larger fidelities than standard tomography with "unlimited" resources, but to achieve better fidelities in a semiautonomous quantum device with reduced resources as compared to usual tomography with equivalent resources.

Quantum reinforcement learning algorithm.—In this section, we review the quantum reinforcement learning algorithm proposed in Ref. [30]. In this work, we mainly focus on the case of reconstructing a qubit state with minimal resources. Our quantum learning algorithm involves three systems: the environment system (E), the register system (R), and the agent system (A), as shown in Fig. 1(a). We assume that the environment system (learning target) can be an arbitrary single-qubit state

$$|E\rangle = \cos(\theta_E/2)|0\rangle + e^{i\phi_E}\sin(\theta_E/2)|1\rangle, \qquad (1)$$

while agent and register qubits are both initialized in $|0\rangle$.

First, we interact the register with the environment system. This step is not necessary if one can measure the environment system directly. Here, we present a more robust method in which the measurements are acted on the register qubit, such that this can be achieved irrespective of whether the environment system is measurable or not. To realize this interaction, we perform a controlled-NOT (CNOT) gate with E as control and R as target, which produces the state

$$\begin{aligned} |\psi\rangle &= U_{E,R}^{\mathrm{CNOT}}|E\rangle|0\rangle_R \\ &= \cos(\frac{\theta_E}{2})|0\rangle_E|0\rangle_R + e^{i\phi_E}\sin(\frac{\theta_E}{2})|1\rangle_E|1\rangle_R. \end{aligned} \qquad (2)$$

Then, we extract the information from the register qubit by measuring it on the basis $\{|0\rangle, |1\rangle\}$. If we detect the signal in the $|0\rangle$ state (the outcome $m = 0$), it indicates that the agent state is similar to the environment state with probability $p = (\cos(\theta_E/2))^2$. On the contrary, if the signal is detected in the $|1\rangle$ state (the outcome $m = 1$), it means the agent state is opposed to the environment state with probability $p = (\sin(\theta_E/2))^2$.

According to the information above, we perform a corresponding action on the agent qubit. When the outcome is 0, we just do nothing, i.e., $U_A = \mathbb{I}$. However, when $m = 1$, we perform a partially-random unitary operation $U_A = \bar{U}(\phi)\bar{U}(\theta)$ on the agent qubit. At ith iteration, $\bar{U}(\phi) = e^{-iS_z^{(i)}\phi^{(i)}}$ and $\bar{U}(\theta) = e^{-iS_x^{(i)}\theta^{(i)}}$ present the phase and amplitude operations along the direction

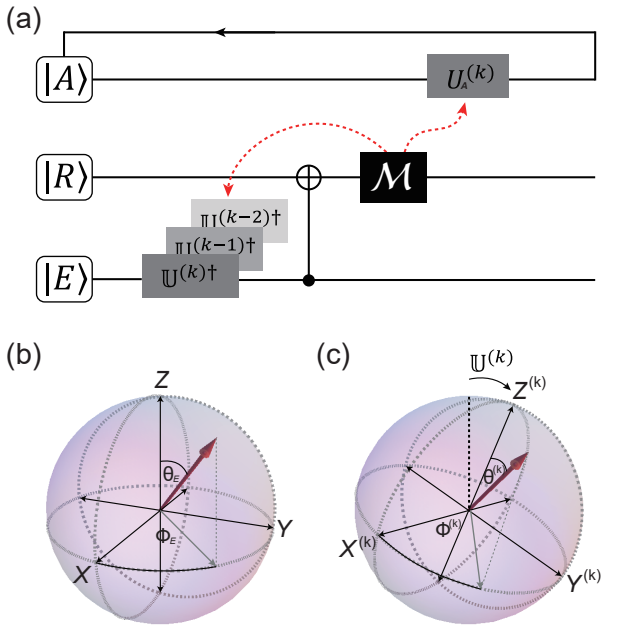

FIG. 1: (a) Quantum circuit diagram for the quantum reinforcement learning protocol. The register qubit first interacts with the environment, and then is detected by a single-shot measurement. The measurement outcomes "0" or "1" which decide the rewards and punishments are input to a computer (not shown in the figure), and control the learning devices (unitary operations) by classical communications, denoted by red dashed lines. (b) Environment state at the initial iteration, shown as the red arrow. (c) Environment state at the kth iteration, in which the reference axes are rotated by previous iterations.

of the spin of the agent. Here, $S_z^{(i)}$ and $S_x^{(i)}$ are the rotated Pauli matrix at ith iteration, and ϕ, θ are the random angles with a range Δ, which can be written as $\phi, \theta \in [-\Delta/2, \Delta/2]$. The angle range is modified by a reward function, which is decided by the outcome of the last step: $\Delta^{(i)} = [(1-m^{(i-1)})\epsilon + m^{(i-1)}1/\epsilon]\Delta^{(i-1)}$. Here, ϵ $(0 < \epsilon < 1)$ controls the reward and punishment ratios, i.e., the value of Δ will be reduced when state $|0\rangle$ is detected, and vice versa. In this step, the partially-random unitary operation becomes $\mathbb{U}^{(i)} = U_A^{(i)}\mathbb{U}^{(i-1)}$ $(\mathbb{U}^{(1)} = \mathbb{I})$. This can be regarded as performing a unitary operation $U_A^{(i)}$ along the reference axes obtained at $i-1$ iteration, shown in Fig. 1 (b) and (c).

The action of the operator $U_A^{(k)}\mathbb{U}^{(k-1)}$ over A is equivalent to the action of the operator $\mathbb{U}^{(k)\dagger}$ over E, which changes the basis of the environment in order to perform the measurement process in the logical basis 0/1.

Experimental setup.—The experimental setup is shown in Fig. 2, which can be recognized as composed of five parts: the generation of photon pairs by periodically poled KTP (PPKTP) crystal through the spontaneous parametric down-conversion (SPDC) process; the environment system that can provide many copies of the

unknown quantum state; the register system which can interact with the environment; and the agent, which can generate deterministic single-photons (polarized at $|H\rangle$) and is equipped with a learning device that can adjust the polarization component and the relative phase of the photon. The PPKTP crystal is pumped by a 404 nm laser with horizontal polarization, and a pair of photons at 808nm can be generated with horizontal and vertical polarization, respectively. A polarizing beam splitter (PBS) thereafter divides these two photons into different paths: one of them is sent to the environment system, and the other one to the register system. In the environment system, the photons are prepared in an unknown state in a black box. Then, a unitary operation which is controlled by the measurement results adjusts the reference axis [32]. The register qubit is prepared in state $|H\rangle$ and interacts with the environment system by a CNOT gate.

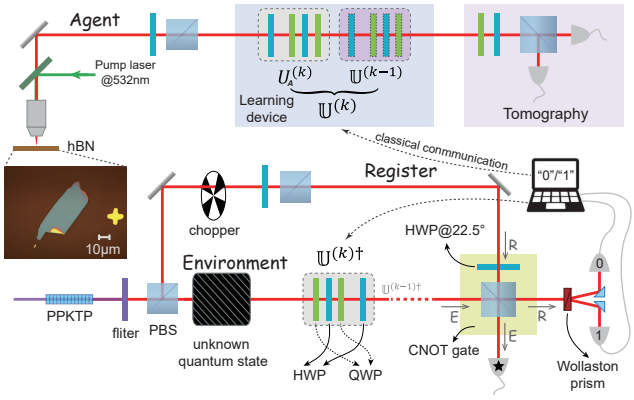

FIG. 2: Experimental setup. The photon pairs are generated from PPKTP by SPDC process, and work as environment system and register system, respectively. On the environment route, photons are randomly prepared and go through a unitary gate which is determined by the last iteration. On the register route, the chopper works as a switch for creating a pseudo-on-demand single photon source assisted by postselection. Then, the photons are prepared at $|H\rangle$ and interact with the environment by a CNOT gate. After the interaction, we measure the register qubit and obtain the outcome "0" or "1" (If we find the first event when chopper is switched on is the coincidence of "0" detector and "★" detector, we deem that the outcome at this iteration is "0"). According to the outcome in the current iteration, we perform the action on the agent system as well as the environment system.

Consider the kth iteration, first we prepare the photons in the environment system at an unknown state, as shown in Eq. (1) in a black box. This time the unitary operation $\mathbb{U}^{k-1}$ is settled by $(k-1)$th iteration. Then, the register qubit, which is prepared at $|H\rangle$, interacts with the environment system by a CNOT gate [33, 34] (more details about the CNOT gate are shown in the Appendix A). After this, we extract the information by measuring the register qubit at $\{|H\rangle, |V\rangle\}$ basis via a Wollaston prism, and we can obtain the outcomes $m = 0$ or $m = 1$. According to the measurement outcomes, we calculate the unitary operation $U_A^{(k)}$, including θ and ϕ, on the classical computer and drive the learning device. Meanwhile, $\mathbb{U}^{(k)\dagger}$ is also operated for adjusting its reference axis and prepared for the next iteration. At the agent system, a deterministic single-photon source is fabricated based on the N_BV_N color-center defect in hexagonal boron nitride (hBN) flakes. More details about the single photon source are shown in the Appendix B. A half-wave plate (HWP) and PBS initialize the single photons at $|H\rangle$ and a learning device, which is controlled by the computer according to the "reinforcement" signal 0/1 equipped after that. After each iteration, we can use the tomography setup at the agent route to check the fidelity of the state at current step. It is worth to note that during each iteration, only one photon is consumed. In our experiment, we use a chopper (working as a switch) and post-selection method to realize a pseudo-on-demand single photon source [25]. That is, when a new iteration begins, we drive the chopper to switch on the register route, and select the first effective coincidence event to be the result [25].

Results of the adaptive learning.—For simplicity, we first try to reconstruct a quantum state without imaginary part, e.g., $|E_1\rangle = \frac{1}{\sqrt{2}}(|H\rangle + |V\rangle)$, and the experimental results of the quantum reinforcement learning progress are shown in Fig. 3. We can observe that a relative larger reward/punishment ratio ϵ will cause a slower convergence speed. When $\epsilon = 0.80$, the fidelity converges at around $k = 36$. However, when $\epsilon = 0.50$, the fidelity converges at around $k = 7$, which is much earlier than the former case. In this situation, the agent can learn the information of the environment state successfully within 50 iterations. A typical example is shown in Fig. 3, while the fidelity of the agent system can reach over 93% after 20 iterations for all ϵ. We perform 20 identical learning experiments in total for this state and the average fidelity at the last iteration can reach 0.955 ± 0.047, 0.947 ± 0.045, 0.931 ± 0.074 at $\epsilon = 0.80, 0.65, 0.50$, respectively [35]. Furthermore, it is worth to note that in our QRL protocol, each iteration just consumes one photonic copy of the environment. Therefore, the 50 iterations can be realized by 50 copies of the photon. If we use the same numbers of photons (i.e., the same resources) to perform quantum tomography on the environment system, the average fidelity can only reach 0.654 ± 0.100, which is obtained by 20 individual experiments (shown as the black line and the gray region—the range of error). This result indeed exhibits the advantages of our applied QML method.

In order to prove the robustness of our QRL protocol, we select other two quantum states for which an imaginary part is contained. Figure 4 (a,b) show the learning process for the environment state $|E_2\rangle = \frac{1}{\sqrt{2}}(|H\rangle + e^{i\pi/4}|V\rangle)$ and $|E_3\rangle = 0.948|H\rangle + e^{0.890i}0.317|V\rangle$,

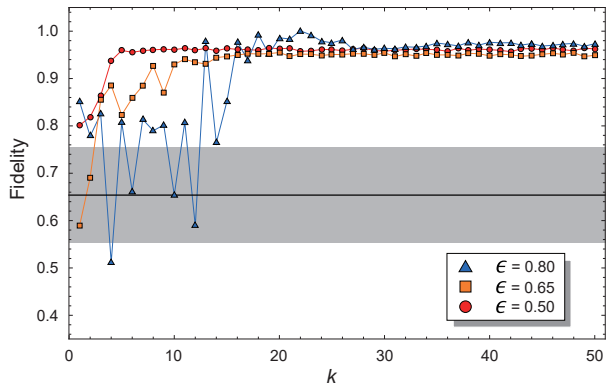

FIG. 3: Experimental results for the quantum reinforcement learning process with environment state $|E_1\rangle$. The triangle (blue), square (orange), and circle (red) represent the fidelity of the agent state in each iteration with different reward/punishment ratios ϵ. The black solid line and the gray region are the average values of the fidelity of the environment state obtained by quantum tomography with 50 photons, and associated errors obtained in 20 individual experiments, respectively.

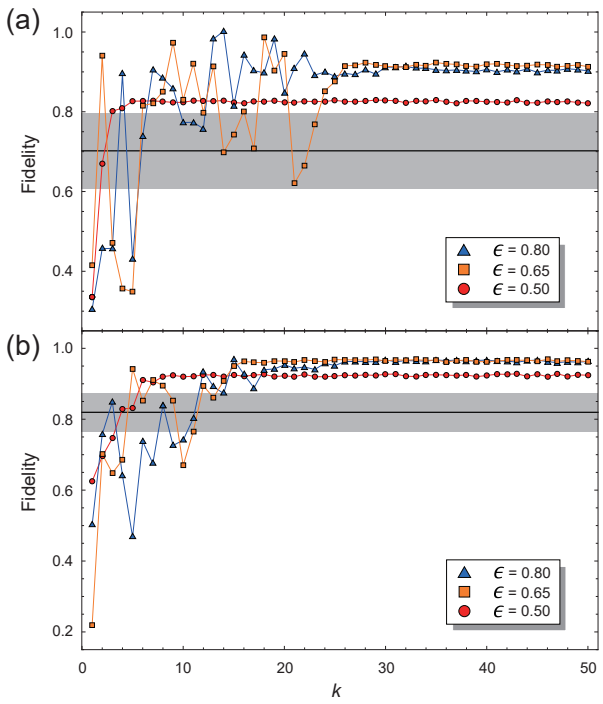

FIG. 4: (a) The fidelity of the agent state in each step for learning environment $|E_2\rangle$. (b) The fidelity of the agent state in each step for learning environment $|E_3\rangle$. The symbols are similar to those in Fig. 3.

respectively. For these more general environment systems, the QRL protocol can also successfully adapt the agent state to an ideal fidelity. The average fidelities obtained from 20 learning experiments are 0.886 ± 0.033, 0.882 ± 0.035, 0.860 ± 0.060 at $\epsilon = 0.80, 0.65, 0.50$, respectively for $|E_2\rangle$; and 0.933 ± 0.044, 0.911 ± 0.052, 0.902 ± 0.046 at $\epsilon = 0.80, 0.65, 0.50$, respectively for $|E_3\rangle$. Besides, in these two cases, we find a relatively higher ϵ (i.e., $\epsilon = 0.80, 0.65$) will benefit the agent to learn more environment information (i.e., a higher fidelity can be obtained). Still, for these cases, the QRL-assisted results all surpass the canonical quantum tomography method with the same resources (the average values of fidelities obtained by quantum tomography are 0.702 ± 0.095 and 0.820 ± 0.054 for $|E_2\rangle$ and $|E_3\rangle$, respectively).

Conclusions.—In summary, we demonstrate the quantum reinforcement learning protocol that can help a quantum agent to adapt to an unknown environment qubit state. The environment system can provide a limited number of photonic copies, and the register qubit will interact with it by a CNOT gate. Then, we detect the register photons on $\{|H\rangle, |V\rangle\}$ basis by single-shot measurement, and obtain the "reinforcement" signal to decide reward or punishment for the next iteration. Based on the calculation results, we perform the unitary operations on both agent and environment systems. We can observe that, within 50 iterations, the agent can adapt to the three test environment systems with high fidelities, which all surpass the quantum tomography results with same resources. Especially, in some cases, a proper reward/punishment ratio ϵ will give a better outcome, such as in the test states $|E_1\rangle$ and $|E_2\rangle$, for which a larger ϵ produces better results, at the expense of converging later. This is the well-known balance between exploration and exploitation in RL, namely, a larger ϵ produces larger exploration, with which the final fidelity can be larger, but also more fluctuations, such that the final fidelity is achieved later than with a smaller ϵ. This is a standard characteristic present in reinforcement learning. This QRL protocol can also be extended to high dimensions, multipartite, and mixed quantum state situations, which will enable more applications in the quantum information and communication techniques.

Appendix A: Realization of CNOT gate.—Given that in our experiments the control qubit state is arbitrary $(\alpha|H\rangle+\beta|V\rangle)$, and the target qubit state is just $|H\rangle$, the CNOT gate is easy to be realized by only one PBS and one HWP. The target qubit state is transformed to be $(|H\rangle + |V\rangle)/\sqrt{2}$ after the HWP @22.5°. Thus, the whole system can be written as $(\alpha|HH\rangle + \alpha|HV\rangle + \beta|VH\rangle + \beta|VV\rangle)/\sqrt{2}$. We use the post-selection approach here to select the coincidence event of the two outputs, such that the state $(\alpha|HH\rangle + \beta|VV\rangle)/\sqrt{2}$ is obtained, with success probability 1/2 [33, 34].

Appendix B: Brief introduction of the hBN material.—The room-temperature single-photon source (SPS) is fabricated based on the hexagonal boron nitride (hBN) flakes performed with the nitrogen-ion irradiation and the high-temperature anneal treatments. The stable single N_BV_N defects in hBN can be gener-

ated and work as SPSs in the existing environment [36]. The SPS used in the experiment is excited and collected by the home-made confocal microscope with a NA=0.9 objective. The single-photon purity of the SPS is $g^{(2)}(0) = 0.045 \pm 0.045 \ll 0.5$, which indicates the remarkable quantum-emission property.

ACKNOWLEDGMENTS

This work is supported by the National Key Research and Development Program of China (No. 2017YFA0304100), the National Natural Science Foundation of China (Grants Nos. 61327901, 11674304, 61490711, 11774334, 11774335, 11474267, 11325419, 11574291, 91321313 and 11821404), the Key Research Program of Frontier Sciences of the Chinese Academy of Sciences (Grant No. QYZDY-SSW-SLH003), the Youth Innovation Promotion Association of Chinese Academy of Sciences (Grants No. 2017492), the Foundation for Scientific Instrument and Equipment Development of Chinese Academy of Sciences (No. YJKYYQ20170032), Anhui Initiative in Quantum Information Technologies (AHY020100, AHY060300), the Fundamental Research Funds for the Central Universities (No. WK2470000026), Centro Basal FB0807, Ramón y Cajal Grant RYC-2012-11391, MINECO/FEDER FIS2015-69983-P and Basque Government IT986-16.

* Electronic address: tjs@ustc.edu.cn
† Electronic address: cfli@ustc.edu.cn

[1] T. Chouard and L. Venema, Nature **521**, 435 (2015).
[2] J. Stajic, R. Stone, G. Chin, and B. Wilbe, Science **349**, 248 (2015).
[3] D. Silver *et al.*, Nature **529**, 484 (2016).
[4] S. Russell and P. Norvig, Artificial Intelligence: A modern approach (Prentice hall, 1995).
[5] P. Mehta, M. Bukov, C.-H. Wang, A. G. R. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, Preprint at https://arxiv.org/abs/1803.08823 (2018).
[6] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Nature **549**, 195 (2017).
[7] V. Dunjko and H. Briegel, Rep. Prog. Phys. **81**, 074001 (2018).
[8] M. Schuld, I. Sinayskiy, and F. Petruccione, Contemp. Phys. **56**, 172 (2015).
[9] M. Sasaki, A. Carlini, and R. Jozsa, Phys. Rev. A **64**, 022317 (2001).
[10] D. Dong, C. Chen, H. Li, and T. J. Tarn, IEEE Trans. Syst. Man Cybern. B Cybern. **38**, 1207 (2008).
[11] A. W. Harrow, A. Hassidim, and S. Lloyd, Phys. Rev. Lett. **103**, 150502 (2009).
[12] A. Bisio, G. Chiribella, G. M. D'Ariano, S. Facchini, and P. Perinotti, Phys. Rev. A **81**, 032324 (2010).
[13] N. Wiebe, D. Braun, and S. Lloyd, Phys. Rev. Lett. **109**, 050505 (2012).
[14] S. Lloyd, M. Mohseni, and P. Rebentrost, Nat. Phys. **10**, 631-633 (2014).
[15] G. D. Paparo, V. Dunjko, A. Makmal, M. A. Martin-Delgado, and H. J. Briegel, Phys. Rev. X **4**, 031002 (2014).
[16] U. Alvarez-Rodriguez, L. Lamata, P. Escandell-Montero, J. D. Martín-Guerrero, and E. Solano, Scientific Reports **7**, 13645 (2017).
[17] M. Faccin, P. Migda, T. H. Johnson, V. Bergholm, and J. D. Biamonte, Phys. Rev. X **4**, 014012 (2014).
[18] V. Dunjko, J. M. Taylor, and H. J. Briegel, Phys. Rev. Lett. **117**, 130501 (2016).
[19] G. Sentís, E. Bagan, J. Calsamiglia, G. Chiribella, and R. Muñoz-Tapia, Phys. Rev. Lett. **117**, 150502 (2016).
[20] G. Carleo and M. Troyer, Science **355**, 602 (2017).
[21] J. Wang *et al.*, Nat. Phys. **13**, 551 (2017).
[22] D.-L. Deng, X. Li, and S. D. Sarma, Phys. Rev. X **7**, 021021 (2017).
[23] G. Sentís, J. Calsamiglia, and R. Muñoz-Tapia, Phys. Rev. Lett. **119**, 140506 (2017).
[24] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Nat. Phys. **14**, 447 (2018).
[25] S. Yu *et al.*, Preprint at http://arXiv.org/abs/1801.07508 (2018).
[26] J. Gao *et al.*, Phys. Rev. Lett. **120**, 240501 (2018).
[27] G. Torlai and R. G. Melko, Phys. Rev. Lett. **120**, 240503 (2018).
[28] L. Lamata, Scientific Reports **7**, 1609 (2017).
[29] F. A. Cárdenas-López, L. Lamata, J. C. Retamal, and E. Solano, PLOS ONE 13(7): e0200455 (2018).
[30] F. Albarrán-Arriagada, J. C. Retamal, E. Solano, and L. Lamata, Preprint at http://arXiv.org/abs/1803.05340 (2018).
[31] W. K. Wootters and W. H. Zurek, Nature **299**, 802 (1982).
[32] S. Yu, Y.-T. Wang, Z.-J. Ke, W. Liu, W.-H. Zhang, G. Chen, J.-S. Tang, C.-F. Li, G.-C. Guo, Phys. Rev. A **96**, 062324 (2017).
[33] H. He, J. Wu, and X. Zhu, *Advanced Computer Architecture*, 157 (Springer, 2016).
[34] C.-Y. Lu, D. E. Browne, T. Yang, and J.-W. Pan, Phys. Rev. Lett. **99**, 250504 (2007).
[35] Other nineteen experimental data sets are not shown in the picture.
[36] T. T. Tran, K. Bray, M. J. Ford, M. Toth, and I. Aharonovich, Nat. Nanotech. **11**, 37 (2016).