SCIENCE CHINA
Physics, Mechanics & Astronomy

• Article •

May 2022 Vol. 65 No. 5: 250312
https://doi.org/10.1007/s11433-021-1841-2

Experimentally realizing efficient quantum control with reinforcement learning

Ming-Zhong Ai[1,2†], Yongcheng Ding[3,4†], Yue Ban[4,5], José D. Martín-Guerrero[6], Jorge Casanova[4,7], Jin-Ming Cui[1,2*], Yun-Feng Huang[1,2*], Xi Chen[3,4*], 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, China;*
[2] *CAS Center For Excellence in Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei 230026, China;*
[3] *International Center of Quantum Artifcial Intelligence for Science and Technology (QuArtist) and Department of Physics, Shanghai University, Shanghai 200444, China;*
[4] *Department of Physical Chemistry, University of the Basque Country UPV/EHU, Bilbao 48080, Spain;*
[5] *School of Materials Science and Engineering, Shanghai University, Shanghai 200444, China;*
[6] *IDAL, Electronic Engineering Department, University of Valencia, Valencia 46100, Spain;*
[7] *IKERBASQUE, Basque Foundation for Science, Bilbao 48009, Spain*

Received October 19, 2021; accepted December 24, 2021; published online March 10, 2022

We experimentally investigate deep reinforcement learning (DRL) as an artificial intelligence approach to control a quantum system. We verify that DRL explores fast and robust digital quantum controls with operation time analytically hinted by shortcuts to adiabaticity. In particular, the protocol's robustness against both over-rotations and off-resonance errors can still be achieved simultaneously without any priori input. For the thorough comparison, we choose the task as single-qubit flipping, in which various analytical methods are well-developed as the benchmark, ensuring their feasibility in the quantum system as well. Consequently, a gate operation is demonstrated on a trapped $^{171}Yb^{+}$ ion, significantly outperforming analytical pulses in the gate time and energy cost with hybrid robustness, as well as the fidelity after repetitive operations under time-varying stochastic errors. Our experiments reveal a framework of computer-inspired quantum control, which can be extended to other complicated tasks without loss of generality.

quantum control, reinforcement learning, trapped ion, quantum computing, noise robustness

PACS number(s): 03.67.Lx, 32.80.Pj, 07.05.Mh

Citation: M.-Z. Ai, Y. Ding, Y. Ban, J. D. Martín-Guerrero, J. Casanova, J.-M. Cui, Y.-F. Huang, X. Chen, C.-F. Li, and G.-C. Guo, Experimentally realizing efficient quantum control with reinforcement learning, Sci. China-Phys. Mech. Astron. **65**, 250312 (2022), https://doi.org/10.1007/s11433-021-1841-2

1 Introduction

Two-level systems physically realize qubits, which are the basic units of digital quantum computing. In this paradigm, externally controllable parameters should be designed to manipulate the qubits, implementing fast and robust gate operations. Thus, one can construct a universal fault-tolerant quantum computer with physical platforms based on trapped

*Corresponding authors (Jin-Ming Cui, email: jmcui@ustc.edu.cn; Yun-Feng Huang, email: hyf@ustc.edu.cn; Xi Chen, email: xchen@shu.edu.cn)
† These authors contributed equally to this work.

ions and superconducting circuits [1, 2]. In this way, quantum error correction can also be realized physically to reduce the effects of quantum noises and systematic errors. From this perspective, quantum control is bridged to quantum information processing and quantum computing. This connection leads to enormous researches devoted to producing precise quantum control of qubits with driving fields, including adiabatic passages [3, 4], optimized resonant π pulses [5, 6], composite pulses [7-12], pulse-shape engineering [13-15], dynamical decoupling [16], and other optimizations [17-27]. A most straightforward approach to transitionless dynamics obeys the adiabatic theorem by tuning the time-dependent parameter sufficiently slow. However, prolonged operation time destroys the quantum information by induced decoherence, affecting information processing efficiency.

Among these frameworks, the concept of shortcuts to adiabaticity (STA) [28, 29] is proposed, which combines the advantages of both adiabatic passages and resonant pulses. It breaks the adiabatic regime by various techniques, including inverse engineering [30], counter-diabatic driving [31, 32], fast-forward scaling [33, 34], which has been well developed over the past decade. Specifically, inverse engineering leaves enough freedom to further allow other tasks such as, suppressing systematic errors by collaborating with optimal control theory [15, 35-38], dynamical decoupling techniques [39], and machine learning methods [40-42]. However, invariant-based STA requires continuously tunable parameters, limiting the genre of quantum control as analog-only. We consider a more complicated task: designing digital pulses instead of an analog controller with the same output and similar features. In this manner, we would deliver a framework that can be naturally integrated into current quantum computing paradigms based on the application of several digital quantum gates.

Meanwhile, as an artificial intelligence approach, reinforcement learning is a well-known tool for system control [43], and deep learning has been developed for conquering complicated tasks in many areas [44-46], later applied in studying physics [47-52]. Framework of deep learning can be combined with reinforcement learning, searching control pulses for quantum state preparation [53, 54], gate operation [55, 56], and quantum Szilard engine [57]. Since recent researches have employed deep reinforcement learning (DRL) for quantum control [58-63], we are inspired to investigate the connection between DRL and STA. An optimistic expectation is that one can extend STA's concept, introducing DRL as a new technique if it learns the features of STA protocols.

In this work, we present an experimental demonstration of a robust and high-precision quantum control task based on the DRL method on a trapped $^{171}Yb^{+}$ ion. DRL agents learn STA-like control as we expected by showing an agreement on features of robustness, with operation time priori suggested by STA. We also verify the protocol's robustness with the existence of both over-rotation and off-resonance errors at the same time. To demonstrate its application in a natural laboratory noise environment, we examine the performance of DRL models under either fixed or time-varying systematic errors. The results show that controls discovered by DRL are robust enough to these errors, requiring much less operation time and energy cost than analytical pulses. We reckon that DRL methods could lead to a promising enhancement in quantum information processing, enabling other complicated tasks, such as closed-loop quantum control, within a similar framework.

2 Theoretical models

Consider the coherent manipulation of a single qubit, whose Hamiltonian reads

$$H = \frac{\hbar}{2}\left[\Omega\sigma_x + \Delta(t)\sigma_z\right], \tag{1}$$

where the Rabi frequency Ω is fixed, while the detuning $\Delta(t)$ is time-varying. To achieve a robust qubit flipping from $|0\rangle$ to $|1\rangle$, a standard π pulse, which corresponds to the Hamiltonian $\frac{\hbar}{2}\Omega\sigma_x$, is convenient and adequate. However, this operation is sensitive to systematic noise and decoherence.

The invariant-based STA suggests that, one can achieve nonadiabatic quantum control of high fidelity and robustness by designed protocols, which satisfy the auxiliary equations derived from Lewis-Riesenfeld (LR) invariant. We parameterize the eigenstates by time-dependent spherical coordinates, which are the colatitude $\theta(t)$ with respect to the z-axis and the longitude $\beta(t)$ with respect to the x-axis. Thus, the LR invariant of a two-level system is constructed by $I(t) = \frac{\hbar}{2}\Omega_0\sum_{\pm}|\psi_{\pm}(t)\rangle\langle\psi_{\pm}(t)|$, where the eigenstates are $|\psi_{+}(t)\rangle = \left(\cos\frac{\theta}{2}e^{-i\frac{\beta}{2}}, \sin\frac{\theta}{2}e^{i\frac{\beta}{2}}\right)^{T}$ and $|\psi_{-}(t)\rangle = \left(\sin\frac{\theta}{2}e^{-i\frac{\beta}{2}}, -\cos\frac{\theta}{2}e^{i\frac{\beta}{2}}\right)^{T}$. The dynamics of the Hamiltonian is governed by time-dependent Schrödinger's equation, whose solution is in superposition of these eigenstates as $|\Psi(t)\rangle = \sum_{\pm}c_{\pm}\exp(i\gamma_{\pm})|\psi_{\pm}(t)\rangle$, with LR phase calculated as:

$$\gamma_{\pm} = \pm\frac{1}{2}\int_0^t\left(\frac{\dot{\theta}\cot\beta}{\sin\theta}\right)dt'. \tag{2}$$

According to the condition for invariant $dI(t)/dt = \partial I(t)/\partial t + (1/i\hbar)[I(t), H(t)] = 0$, we have the auxiliary equations

$$\dot{\theta} = -\Omega\sin\beta, \tag{3}$$

$$\dot{\beta} = -\Omega\cot\theta\cos\beta + \Delta(t), \tag{4}$$

describing the state evolution along the dynamical modes with angular parameters θ and β, which characterize the trajectory on the Bloch sphere. As proposed in ref. [42], the framework can be applied to design robust quantum control, e.g., qubit flipping, against systematic errors with an adequate ansatz of free parameter a, such that

$$\theta(t) = \frac{\Omega T}{a}\left[as - \frac{\pi^2}{2}(1-s)^2 + \frac{\pi^3}{3}(1-s)^3 + \cos(\pi s) + A\right], \tag{5}$$

where $T = -\pi a/[(2 - a - \pi^2/6)\Omega]$, $s = t/T$, and $A = \pi^2/6 - 1$ determined by boundary conditions $\theta(0) = 0$, $\dot{\theta}(0) = \Omega$, $\ddot{\theta}(0) = 0$ and $\theta(T) = \pi$, $\dot{\theta}(T) = \Omega$, $\ddot{\theta}(T) = 0$. Specifically, one can nullify the probability of the first-order transition

$$P = \frac{\hbar^2}{4}\left|\int_0^T \langle\Psi_-(t)|\left(\delta_\Omega\Omega\sigma_x + \delta_\Delta\sigma_z\right)|\Psi_+(t)\rangle\right|^2, \tag{6}$$

which yields the condition for error cancellation

$$\left|\int_0^T \mathrm{d}t e^{\mathrm{i}2\gamma_+(t)}\left(\delta_\Delta \sin\theta - \mathrm{i}2\delta_\Omega\dot{\theta}\sin^2\theta\right)\right| = 0, \tag{7}$$

where systematic errors are characterized by $\Delta(t) \to \Delta(t)+\delta_\Delta$ and $\Omega \to \Omega(1+\delta_\Omega)$, resulting in the configuration $a = 0.604$ and 0.728 for eliminating Δ and Ω-error, respectively. Indeed, smooth detuning pulse $\Delta(t)$ as analog control of single-component is inversely engineered by substituting the ansatz into the following expression:

$$\Delta(t) = -\frac{\ddot{\theta}}{\Omega\sqrt{1-\left(\frac{\dot{\theta}}{\Omega}\right)^2}} + \Omega\cot\sqrt{1-\left(\frac{\dot{\theta}}{\Omega}\right)^2}, \tag{8}$$

which is derived from combining auxiliary equations. The wave-forms of $\Delta(t)$ optimized for different systematic errors in STA (Δ_{STA}) are shown in Figure 1(a) and the maximal tunable detuning $\Delta_{\max}$ for suppressing Δ and Ω errors are 1.5Ω and 1.7Ω, respectively. Concerning our physical realization in trapped ions, the Rabi frequency $\Omega = (2\pi)3.3$ kHz is fixed, where we calculate the corresponding operation time for robust qubit flipping against Δ and Ω-errors as $T_\Delta = 364$ μs and $T_\Omega = 293$ μs.

Since an analog quantum control can be derived from the STA framework, it is more challenging to consider the digital quantum control of the Landau-Zener problem. The problem is reformulated to the following expression: how should we manipulate a quantum system for a certain target with a step controller of N intervals within a fixed time? The combinational optimization problem is equivalent to dynamic programming, i.e., a multi-step decision problem whose complexity grows exponentially with step number, allowing an approximation solution by artificial neural networks (ANN) or other universal function approximators; the use of deep ANN architectures with many layers leads to the concept of deep learning, and this, in turn, to DRL. In the framework of DRL, one assumes that there exists an unknown global optimal policy Π for a task, which gives an action $\mathbf{a}(t_i)$ once observing an arbitrary state $\mathbf{s}(t_i)$ at time t_i. The state-action relation $\Pi(\mathbf{s}|\mathbf{a})$ is approximated by an Agent ANN, containing propagation of information between layers and nonlinear activation of neurons, whose parameters are tuned by optimizing algorithms for maximizing the accumulated reward. Details about the implementation of deep reinforcement learning can be found in Supporting Information S1.

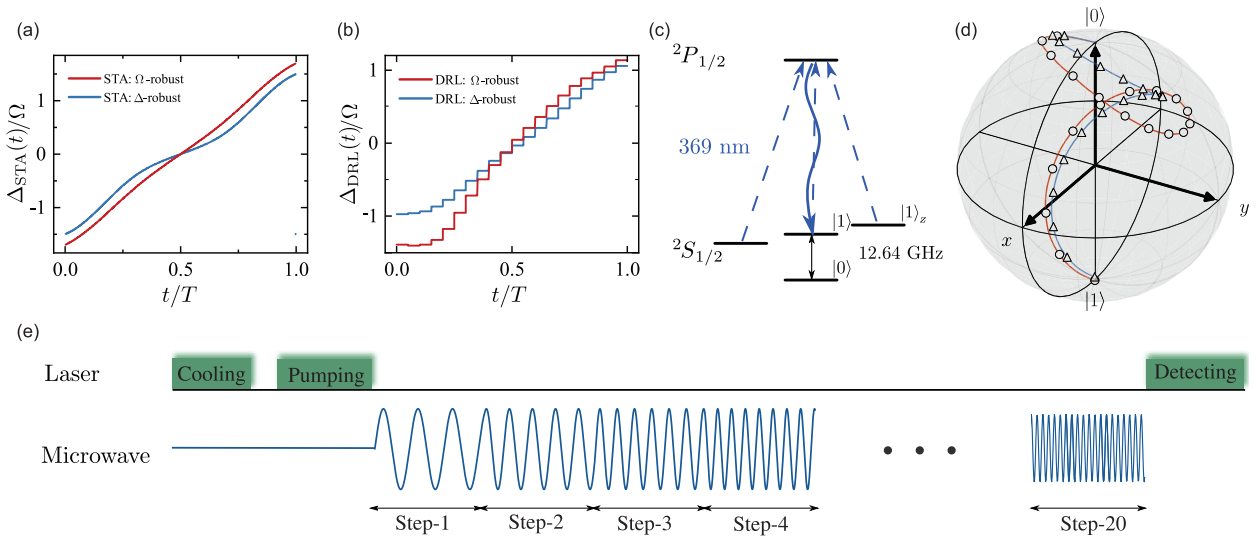

Figure 1 (Color online) Experimental sequences and model wave-forms. (a) Optimized detuning with time under STA method. The time is normalized to [0, 1]. (b) Optimized detuning with time under DRL method. The time is normalized to [0, 1]. (c) Energy level of $^{171}\mathrm{Yb}^+$ ion. (d) Evolution of state in Bloch sphere under the driving of DRL model. Red solid line represents the trajectory optimized for Ω errors while blue solid line represents the trajectory optimized for Δ errors. Hollow circle and hollow triangle represent the state at the end of each driving step. (e) Experimental sequences in DRL model.

In our numerical experiments, the tunable range of detuning $[-\Delta_{\max}, \Delta_{\max}]$ is renormalized into $\tilde{\Delta} \in [0, 1]$ with $\Delta_{\max}$ being the maximal reachable value of $\Delta_{\mathrm{STA}}(t)$, which is the output of ANN as the encoded action at time step t_i: $\tilde{\Delta}(t_i) = [\Delta_{\mathrm{DRL}}(t_i) + \Delta_{\max}]/2\Delta_{\max}$. Information of the two-level system, specifically, the expectation of spin on Z direction $\langle\sigma_z\rangle$, the renormalized detuning $\tilde{\Delta}(t_{i-1})$ that drives the system to the current state, and the system time i/N, are fed to the input layer of the ANN. The quantum dynamics are simulated by Liouville-von Neumann equation, which can be generalized to the Lindblad master equation for taking quantum noises into consideration. While network configuration, hyperparameters, and training details are explained in ref. [42], we introduce the reward functions that we artificially design, which are similar to invariant-based STA that chooses an ansatz for obtaining quantum control. For converging the Agent to robust control of LZ-type, we firstly pre-train the Agent with $r(t_i) = -|\tilde{\Delta}(t_i) - \frac{i-1}{N-1}|$, punishing the deviations from linear growth of detuning, later rewarding a constant if $\langle\sigma_z\rangle > 0.997$ at the final time step for fine-tuning under random systematic errors.

For evaluating the DRL-inspired robust quantum control, we perform two numerical experiments as follows: (i) We set the operation time as $T_\Delta = 364$ and $T_\Omega = 293$ μs, being split uniformly by 20 pulses as the only hint from STA. The digital wave-forms output from our DRL model optimized for different systematic errors are shown in Figure 1(b) (specific values can be found in Supporting Information S4). We emphasize that the STA framework clarifies the upper bound of robustness in the Landau-Zener problems, which could be employed for benchmarking the capability of the Agent, as an artificial intelligence approach to digital quantum control with the alike feature. (ii) The operation time is arbitrarily set to be $T = 300$ μs for checking if the Agent can explore desired protocols against hybrid systematic errors without any field knowledge of STA. We clarify that DRL is more general for this task since invariant-based STA no longer eliminates the hybrid errors perfectly but on certain proportion of δ_Δ and δ_Ω instead. All wave-forms used in real experiments are from these two numerical experiment models.

3 Results and discussion

Our experiments are performed on a $^{171}\mathrm{Yb}^+$ ion trapped in a harmonic Paul trap, with the simplified structure being described in detail in Appendix. As shown in Figure 1(c), the two-level system (TLS) is encoded in the $^2\mathrm{S}_{1/2}$ ground state of the ion, with $|0\rangle = \left|^2\mathrm{S}_{1/2}, F = 0, m_F = 0\right\rangle$ and $|1\rangle = \left|^2\mathrm{S}_{1/2}, F = 1, m_F = 0\right\rangle$. The difference of energy level $|0\rangle$ and $|1\rangle$ is about $\omega_{01} = 12.6428$ GHz. The microwaves used to drive the TLS are generated through a mixing method. More specifically, a signal generator (Agilent E8257D) generates a microwave around 12.4428 GHz and an arbitrary waveform generator (AWG) generates a 200 MHz microwave signal. These two microwave is then mixed with each other through a mixer (see Appendix). After a high pass filter (HPF), this signal will be amplified to about 10 W and then transmitted to the ion with a microwave horn [64]. Our trap device is shielded with a 1.5-mm thick single layer Mu-metal [65], making the final coherence time about 200 ms for $|0\rangle \leftrightarrow |1\rangle$ transition, which is characterized by Ramsey experiments.

In each cycle, the experiment takes the following process: after 1 ms Doppler cooling, the state of the ion is initialized to $|0\rangle$ state through 20 μs optical pumping with 99.5% fidelity. The wave-form output from DRL model is transformed into driving microwave through modulating the detuning, which is shown in Figure 1(e). Then the driving microwave is transmitted to the ion to drive the TLS. Finally, a NA (numerical aperture) = 0.4 objective is used for state-dependent fluorescence detection to determine the probability in state $|1\rangle$ (the qubit state preparation and measurement can be found in Appendix). In all of our experiments, we set the Rabi frequency to $\Omega = (2\pi)$ 3.3 kHz. That is to say, the corresponding 2π time is about 300 μs.

To verify the robustness of the DRL control method against static errors, we compare the performance of STA, DRL, standard π pulse and BB1 (BroadBand 1) [66] or CORPSE methods [67] (Compensate for Off-Resonance with a Pulse SEquence, BB1 and CORPSE are described in detail in Supporting Information S2) in a single-qubit X gate task under different over-rotation (δ_Ω) errors or off-resonance (δ_Δ) errors, respectively. The DRL models are pre-trained according to the preliminary information provided by STA methods optimized in δ_Ω and δ_Δ errors. The state evolution under DRL driving in Bloch sphere is shown in Figure 1(d). The comparison results are shown in Figure 2(a) and (b), and the more refined information presented on a log-scale is shown in Figure 2(c) and (d). The solid lines represent numerical results while the distinct points correspond to the experiments in all Figure 2. The DRL method performs as well as STA in most cases, in addition to the case that δ_Ω and δ_Δ errors are too large. Meanwhile, they are all more robust than the π pulse method and a little worse than the BB1 method under static over-rotation errors. As for detuning errors, DRL and STA methods perform better than π pulse methods and perform as well as CORPSE method. It is worth noting that the operation time of BB1 and CORPSE are 750 and 650 μs, respectively, which are more than twice that of DRL method. Since the microwave power is directly proportional to the Rabi frequency Ω of the system. The energy cost depends only on Ω and operation time, this means BB1 and CORPSE methods

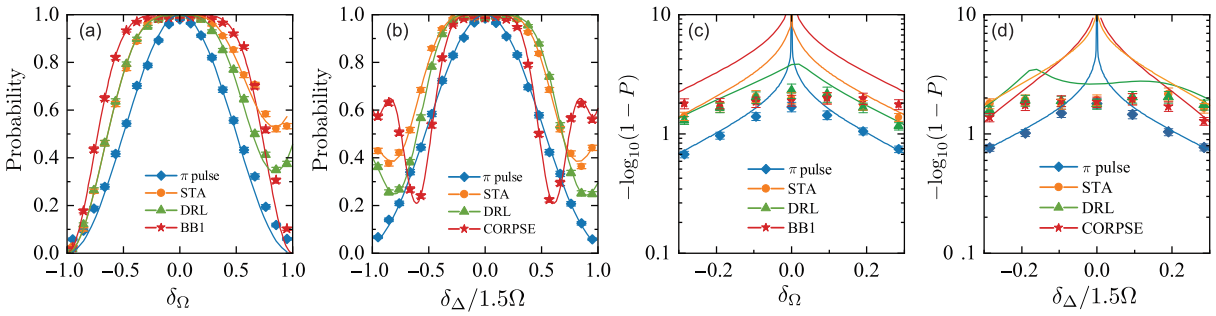

Figure 2 (Color online) Noise robustness comparison of π pulse, STA, DRL and BB1 or CORPSE methods in single-qubit X gate task under different over-rotation or off-resonance errors, respectively. (a), (b) The performance of four control methods under different δ_Ω and δ_Δ errors. The DRL method is as robust as STA method in most cases and they are all more robust than π pulse in both kinds of errors. (c), (d) The detailed comparison presented on a log-scale of results. For each point, the experiment was repeated 1000 times. The error bars correspond to quantum projection noise. The solid lines represent the result of numerical simulation.

will consume twice as much energy as DRL method for the same task.

Then we examine the DRL model under δ_Δ and δ_Ω hybrid errors. It is worthwhile to mention that we set operation time and tunable range of detuning without any knowledge from STA when pre-training the DRL model locally. The performance of π pulse and DRL method under hybrid errors is shown in Figure 3, in which the probabilities are represented in a logarithmic scale to better distinguish the difference between these two methods. The DRL method is more likely to accumulate errors than π pulse due to the multi pulses driving operation; besides, we just stop our training once $\langle\sigma_z\rangle > 0.997$, which can be further improved theoretically. As we can expect, the π pulse method performs a little better than DRL in the case of almost no errors. Nevertheless, with the increase of hybrid errors, the DRL method's performance is much better than π pulse in most cases, which is essential in precise quantum manipulation.

Besides, we also examine the DRL model at the Zeeman energy level of the ion with $|1\rangle_z = \left|^2S_{1/2}, F=1, m_F=1\right\rangle$. The Zeeman energy level is first-order sensitive to the disturbance of the magnetic field, which could induce realistic laboratory noise into the TLS. The corresponding coherence time of $|0\rangle \leftrightarrow |1\rangle_z$ transition is about 0.35 ms. We compare the performance of π pulse, DRL and CinBB (CORPSE in BB1, described in detail in Supporting Information S2) [66, 68] methods in the single-qubit X gate under only the magnetic field noise with different numbers of π flips. As shown in Figure 4(a), the final probability decreases rapidly with the number of π flips in π pulse method owing to inevitable decoherence. However, the DRL method is more robust with the increase of the number of π flips than π pulse and CinBB method. For the propose of examining DRL model under time-varying errors [12], we introduce white noise both in over-rotation and off-resonance errors through mixing and microwave modulation methods in our system (see Supporting Information S3 for details of time-varying systematic errors) [69]. The experimental results, which are shown in Figure 4(b), indicate that DRL model is more robust than π pulse and CinBB methods under time-varying

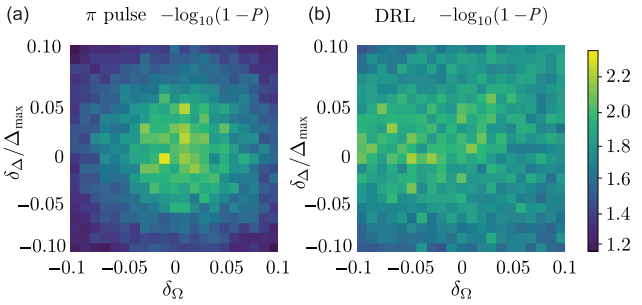

Figure 3 (Color online) The performance of π pulse and DRL model under hybrid errors. The π pulse method (a) performs a little better than DRL method (b) in the case of almost no errors, because the pulses of DRL are more complex that it is easy to accumulate operation errors. However, with the increase of hybrid errors, the performance of DRL model is much better than π pulse control.

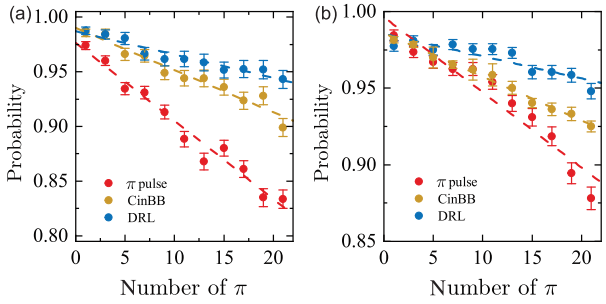

Figure 4 (Color online) Noise-resilient feature of DRL method under systematic and time-varying errors. (a) The performance of π pulse, CinBB and DRL methods under systematic errors with different number of π flips. (b) The performance of π pulse, CinBB and DRL methods under time-varying errors with different number of π flips. For each point, the experiment was repeated 1000 times. The error bars correspond to quantum projection noise. The dashed lines represent the linear fitting of the data.

errors, which is of great significance in noisy quantum information processing experiments. Besides, the operation time of CinBB is 1250 μs, which is almost four times as long as DRL method.

4 Conclusion

In summary, we experimentally demonstrate a fast and robust quantum control task based on DRL, realizing better robustness and lower energy cost in qubit flipping as a simple task. Compared with STA, BB1 (over-rotation error), CORPSE (off-resonance error), and CinBB (both errors) under various noise and errors, the results verify the advantage of the DRL method on driving the system under laboratory environments, providing sufficient and reliable information for precise quantum control and quantum information processing. Besides, this framework allows further fine-tuning for better fidelity, allowing one to study the trade-off between fidelity and robustness under various environment. The framework of this work can also be fitted to other tasks, e.g., designing robust two-qubit gates, controlling many-body physics models without much effort. Another extension could be directly training the DRL agent in interactive quantum systems for other complicated tasks, e.g., closed-loop quantum control with continuous measurement and weak value feedback, which shares the same framework but is challenging for both analytical design by theory and experimental implementation in quantum optical systems.

This work was supported by the National Natural Science Foundation of China (Grant Nos. 11874343, 61327901, 11774335, and 11734015), An-hui Initiative in Quantum Information Technologies (Grant Nos. AHY020100, and AHY070000), and Key Research Program of Frontier Sciences, Chinese Academy of Sciences (Grant No. QYZDYSSW-SLH003). The theoretical part of the work was also partially supported by the National Natural Science Foundation of China (Grant No. 12075145), STCSM (Grant No. 2019SHZDZX01-ZX04), Program for Eastern Scholar, QMiCS (Grant No. 820505) and OpenSuperQ (Grant No. 820363) of the EU Flagship on Quantum Technologies, Spanish Government PGC2018-095113-B-I00 (MCIU/AEI/FEDER, UE), Basque Government IT986-16, EU FET Open Grant Quromorphic (Grant No. 828826) as well as EPIQUS (Grant No. 899368). Xi Chen acknowledges the Ramon y Cajal Program (Grant No. RYC-2017-22482). Jorge Casanova acknowledges the Ramon y Cajal Program (Grant No. RYC2018-025197-I) and the EUR2020-112117 Project of the Spanish MICINN, as well as the support from the UPV/EHU through the grant EHUrOPE.

Supporting Information

The supporting information is available online at phys.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

1 K. R. Brown, J. Kim, and C. Monroe, npj Quantum Inf. **2**, 16034 (2016).
2 J. M. Gambetta, J. M. Chow, and M. Steffen, npj Quantum Inf. **3**, 2 (2017).
3 P. Král, I. Thanopulos, and M. Shapiro, Rev. Mod. Phys. **79**, 53 (2007).
4 J. M. Martinis, and M. R. Geller, Phys. Rev. A **90**, 022307 (2014), arXiv: 1402.5467.
5 S. V. Remizov, D. S. Shapiro, and A. N. Rubtsov, Phys. Rev. A **92**, 053814 (2015), arXiv: 1508.00821.
6 D. Stefanatos, and E. Paspalakis, Phys. Rev. A **100**, 012111 (2019), arXiv: 1906.11493.
7 K. R. Brown, A. W. Harrow, and I. L. Chuang, Phys. Rev. A **70**, 052318 (2004), arXiv: quant-ph/0407022.
8 B. T. Torosov, S. Guérin, and N. V. Vitanov, Phys. Rev. Lett. **106**, 233001 (2011).
9 X. Rong, J. Geng, F. Shi, Y. Liu, K. Xu, W. Ma, F. Kong, Z. Jiang, Y. Wu, and J. Du, Nat. Commun. **6**, 8748 (2015).
10 K. R. Brown, A. W. Harrow, and I. L. Chuang, Phys. Rev. A **70**, 052318 (2004), arXiv: quant-ph/0407022.
11 E. Mount, C. Kabytayev, S. Crain, R. Harper, S.-Y. Baek, G. Vrijsen, S. T. Flammia, K. R. Brown, P. Maunz, and J. Kim, Phys. Rev. A **92**, 060301 (2015).
12 C. Kabytayev, T. J. Green, K. Khodjasteh, M. J. Biercuk, L. Viola, and K. R. Brown, Phys. Rev. A **90**, 012316 (2014), arXiv: 1402.5174.
13 M. Steffen, and R. H. Koch, Phys. Rev. A **75**, 062326 (2007).
14 E. Barnes, and S. Das Sarma, Phys. Rev. Lett. **109**, 060401 (2012), arXiv: 1206.0297.
15 D. Daems, A. Ruschhaupt, D. Sugny, and S. Guérin, Phys. Rev. Lett. **111**, 050404 (2013), arXiv: 1304.4016.
16 L. Viola, E. Knill, and S. Lloyd, Phys. Rev. Lett. **82**, 2417 (1999), arXiv: quant-ph/9809071.
17 S. J. Glaser, U. Boscain, T. Calarco, C. P. Koch, W. Köckenberger, R. Kosloff, I. Kuprov, B. Luy, S. Schirmer, T. Schulte-Herbrüggen, D. Sugny, and F. K. Wilhelm, Eur. Phys. J. D **69**, 279 (2015).
18 T. Caneva, M. Murphy, T. Calarco, R. Fazio, S. Montangero, V. Giovannetti, and G. E. Santoro, Phys. Rev. Lett. **103**, 240501 (2009), arXiv: 0902.4193.
19 S. Guérin, V. Hakobyan, and H. R. Jauslin, Phys. Rev. A **84**, 013423 (2011).
20 G. C. Hegerfeldt, Phys. Rev. Lett. **111**, 260501 (2013), arXiv: 1305.6403.
21 A. Garon, S. J. Glaser, and D. Sugny, Phys. Rev. A **88**, 043422 (2013), arXiv: 1310.5226.
22 L. Van Damme, Q. Ansel, S. J. Glaser, and D. Sugny, Phys. Rev. A **95**, 063403 (2017), arXiv: 1704.07653.
23 C. Arenz, B. Russell, D. Burgarth, and H. Rabitz, New J. Phys. **19**, 103015 (2017).
24 H. Ball, M. Biercuk, A. Carvalho, J. Chen, M. R. Hush, L. A. de Castro, L. Li, P. J. Liebermann, H. Slatyer, C. Edmunds, V. Frey, C. Hempel, and A. Milne, Quantum Sci. Technol. **6**, 044011 (2021).
25 A. R. R. Carvalho, H. Ball, M. J. Biercuk, M. R. Hush, and F. Thomsen, Phys. Rev. Appl. **15**, 064054 (2021), arXiv: 2010.08057.
26 J. T. Merrill, and K. R. Brown, *Progress in Compensating Pulse Sequences for Quantum Computation* (John Wiley & Sons, Ltd., Hoboken, 2014), pp. 241-294.
27 G. Dridi, K. Liu, and S. Guérin, Phys. Rev. Lett. **125**, 250403 (2020).
28 D. Guéry-Odelin, A. Ruschhaupt, A. Kiely, E. Torrontegui, S. Martínez-Garaot, and J. G. Muga, Rev. Mod. Phys. **91**, 045001 (2019), arXiv: 1904.08448.
29 E. Torrontegui, S. Ibánez, S. Martínez-Garaot, M. Modugno, A. del Campo, D. Guéry-Odelin, A. Ruschhaupt, X. Chen, and J. G. Muga, *Advances in Atomic, Molecular, and Optical Physics* (Elsevier, Amsterdam, 2013), vol. 62, pp. 117-169.
30 X. Chen, A. Ruschhaupt, S. Schmidt, A. del Campo, D. Guéry-Odelin, and J. G. Muga, Phys. Rev. Lett. **104**, 063002 (2010), arXiv: 0910.0709.

31 S. Deffner, C. Jarzynski, and A. del Campo, Phys. Rev. X **4**, 021013 (2014).
32 S. An, D. Lv, A. Del Campo, and K. Kim, Nat. Commun. **7**, 12999 (2016), arXiv: 1601.05551.
33 S. Masuda, and K. Nakamura, Proc. Math. Phys. Eng. Sci. **466**, 1135 (2010).
34 S. Masuda, Phys. Rev. A **86**, 063624 (2012), arXiv: 1208.5650.
35 A. Ruschhaupt, X. Chen, D. Alonso, and J. Muga, New J. Phys. **14**, 093040 (2012).
36 X. J. Lu, X. Chen, A. Ruschhaupt, D. Alonso, S. Guérin, and J. G. Muga, Phys. Rev. A **88**, 033406 (2013), arXiv: 1305.6127.
37 S. Masuda, and S. A. Rice, J. Chem. Phys. **142**, 244303 (2015).
38 H. Wu, L. Xing, Y. Cai, L. Liu, E. He, B. Li, and X. Tian, Appl. Sci. **10**, 1508 (2020).
39 C. Munuera-Javaloy, Y. Ban, X. Chen, and J. Casanova, Phys. Rev. Appl. **14**, 054054 (2020), arXiv: 2007.15394.
40 E. Zahedinejad, J. Ghosh, and B. C. Sanders, Phys. Rev. Appl. **6**, 054005 (2016), arXiv: 1511.08862.
41 B. J. Liu, X. K. Song, Z. Y. Xue, X. Wang, and M. H. Yung, Phys. Rev. Lett. **123**, 100501 (2019), arXiv: 1806.07904.
42 Y. Ding, Y. Ban, J. D. Martín-Guerrero, E. Solano, J. Casanova, and X. Chen, Phys. Rev. A **103**, L040401 (2021).
43 R. S. Sutton, and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. (The MIT Press, Cambridge, 2018).
44 V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, Nature **518**, 529 (2015).
45 V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, arXiv: 1312.5602.
46 D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan, and D. Hassabis, arXiv: 1712.01815.
47 X. Y. Ma, H. Y. Lyu, K. R. Hao, Y. M. Zhao, X. Qian, Q. B. Yan, and G. Su, Sci. Bull. **66**, 233 (2021), arXiv: 2004.08527.
48 A. Nagy, and V. Savona, Phys. Rev. Lett. **122**, 250501 (2019), arXiv: 1902.09483.
49 M. J. Hartmann, and G. Carleo, Phys. Rev. Lett. **122**, 250502 (2019), arXiv: 1902.05131.
50 F. Vicentini, A. Biella, N. Regnault, and C. Ciuti, Phys. Rev. Lett. **122**, 250503 (2019), arXiv: 1902.10104.
51 N. Yoshioka, and R. Hamazaki, Phys. Rev. B **99**, 214306 (2019), arXiv: 1902.07006.
52 R. Iten, T. Metger, H. Wilming, L. Del Rio, and R. Renner, Phys. Rev. Lett. **124**, 010508 (2020), arXiv: 1807.10300.
53 B. M. Henson, D. K. Shin, K. F. Thomas, J. A. Ross, M. R. Hush, S. S. Hodgman, and A. G. Truscott, Proc. Natl. Acad. Sci. USA **115**, 13216 (2018), arXiv: 1809.03124.
54 X.-M. Zhang, Z. Wei, R. Asad, X.-C. Yang, and X. Wang, npj Quantum Inf. **5**, 85 (2019).
55 Z. An, and D. L. Zhou, Eur. Phys. Lett. **126**, 60002 (2019).
56 X. Yang, J. Li, and X. Peng, Sci. Bull. **64**, 1402 (2019).
57 V. B. Sørdal, and J. Bergli, Phys. Rev. A **100**, 042314 (2019), arXiv: 1904.04712.
58 M. Bukov, A. G. Day, D. Sels, P. Weinberg, A. Polkovnikov, and P. Mehta, Phys. Rev. X **8**, 031086 (2018).
59 R. Porotti, D. Tamascelli, M. Restelli, and E. Prati, Commun. Phys. **2**, 61 (2019).
60 M. Y. Niu, S. Boixo, V. N. Smelyanskiy, and H. Neven, npj Quantum Inf. **5**, 33 (2019).
61 X. M. Zhang, Z. W. Cui, X. Wang, and M. H. Yung, Phys. Rev. A **97**, 052333 (2018), arXiv: 1802.09248.
62 R. B. Wu, H. Ding, D. Dong, and X. Wang, Phys. Rev. A **99**, 042327 (2019), arXiv: 1811.01884.
63 Z. T. Wang, Y. Ashida, and M. Ueda, Phys. Rev. Lett. **125**, 100401 (2020), arXiv: 1910.09200.
64 J. M. Cui, M. Z. Ai, R. He, Z. H. Qian, X. K. Qin, Y. F. Huang, Z. W. Zhou, C. F. Li, T. Tu, and G. C. Guo, Sci. Bull. **64**, 1757 (2019).
65 A. Farolfi, D. Trypogeorgos, G. Colzi, E. Fava, G. Lamporesi, and G. Ferrari, Rev. Sci. Instr. **90**, 115114 (2019), arXiv: 1907.06457.
66 C. A. Ryan, J. S. Hodges, and D. G. Cory, Phys. Rev. Lett. **105**, 200402 (2010), arXiv: 1008.2197.
67 S. Wimperis, J. Magne. Res. Ser. A **109**, 221 (1994).
68 A. M. Souza, G. A. Alvarez, and D. Suter, Phys. Rev. Lett. **106**, 240501 (2011), arXiv: 1103.4563.
69 M. Z. Ai, J. M. Cui, R. He, Z. H. Qian, X. X. Gao, Y. F. Huang, C. F. Li, and G. C. Guo, Phys. Rev. A **103**, 012608 (2021), arXiv: 2008.00885.

Appendix

A1 Trapped ion platform

The type of platform used in our experiments is needle trap. As shown in Figure a1, the needle trap consists of 6 needles. Two opposite needles are connected to radio frequency (RF) potential to trap the ion and the others are connected to direct current (DC) potential to fine tuning the position of the ion. The size of the needle trap depends mainly on the distance between the two needles tips near the trap center, which is set to 180 μm in our experiment. The trap is installed in an ultrahigh vacuum below 10^{-11} torr, and a helical resonator provides the RF signal with frequency 24 MHz and amplitude of 180 V to the trap. Ion fluorescence is collected by an objective lens with 0.4 numerical aperture, and detected by a photo-multiplier tube (PMT). The total fluorescence detection efficiency is about 2%.

A2 Microwave waveform used in deep reinforcement learning method

We generate required waveform of the microwave field through setting the waveform of arbitrary waveform generation (AWG) for modulation. The carrier microwave $B_c(t) = A_1 \sin(\omega_c t)$, where A_1 is amplitude and $f_c = \omega_c/2\pi = 12.4$ GHz is the frequency. The waveform generated by

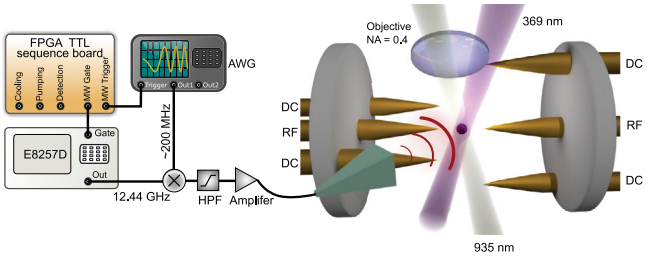

Figure a1 (Color online) Experimental setup. A single $^{171}Yb^{+}$ ion is trapped in the center of the needle trap. Two 369 and 935 nm lasers are used to cooling the ion and 369 nm laser is also used to detect the state of ion. The microwave used to drive the ion is generated through mixing method. The whole experimental sequences are controlled by a TTL sequences board based on field programmable gate array (FPGA).

AWG for modulation is $I(t) = A_2\sin(\phi(t))$. After mixing, the microwave field will be $B(t) = \frac{A_1A_2}{2}(\sin(\omega_c t+\phi(t))+\sin(\omega_c t-\phi(t)))$, where the phase function $\phi(t)$ can be expressed in a piece-wise function for the microwave composed of 20 steps in our DRL experiments. With the qubit resonance frequency $f_0 = \omega_0/2\pi = 12.6$ GHz, we filter out the low frequency components of the microwave through a high pass filter. In our experiments, we only adjust $\Delta(t)$ with discrete steps by changing phase $\phi(t)$ as follows:

$$\phi(t) = \begin{cases} (\omega_0-\omega_c)t + \Delta_1 t, & (0, t_1), \\ (\omega_0-\omega_c)t + \Delta_2 t + \phi_1, & (0, t_2-t_1), \\ (\omega_0-\omega_c)t + \Delta_3 t + \phi_2, & (0, t_3-t_2), \\ \dots & \\ (\omega_0-\omega_c)t + \Delta_{20} t + \phi_{19}, & (0, t_{20}-t_{19}), \end{cases} \tag{a1}$$

where $\Delta_n (n \in [1, 20])$ is the step-wise detuning and $\phi_1 = (\omega_0-\omega_c)t_1 + \Delta_1 t_1$, $\phi_2 = (\omega_0-\omega_c)(t_2-t_1) + \Delta_2(t_2-t_1) + \phi_1$ and so on.

A3 Qubit state preparation and measurement

In ion trap experiments, the state preparation and measurement cannot be perfect and there will always be some limitations. We characterize these errors as follows. The ion is prepared in $|0\rangle$ state through optical pumping and ideally, no photon should be detected as the ion is in dark state. However, due to the dark counts of the photon detector as well as photons scattered from the environment, we will collect some photons sometimes. Then, we apply a π pulse to flip the $|0\rangle$ state to $|1\rangle$ state and detect the fluorescence. Because the collection efficiency problem, no photon will be collected sometimes. The threshold is selected as 2 in our experiments. When the photon number is > 2, the qubit is identified as bright state and the probability of being mistaken as dark state is ϵ_D. By contrary, the probability of being mistaken as bright state when photon number is ≤ 2 is ϵ_B. The total error can be taken as $\epsilon = (\epsilon_B + \epsilon_D)/2$, which is 0.5% in our experiments.