# No-Collapse Accurate Quantum Feedback Control via Conditional State Tomography

Sangkha Borah[1,2,3,*] and Bijita Sarma[2,3,†]
[1]Max Planck Institute for the Science of Light, Staudtstraße 2, 91058 Erlangen, Germany
[2]Department of Physics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Staudtstraße 7, 91058 Erlangen, Germany
[3]Okinawa Institute of Science and Technology, Okinawa 904-0495, Japan

(Received 17 January 2023; accepted 30 October 2023; published 21 November 2023)

The effectiveness of measurement-based feedback control protocols is hampered by the presence of measurement noise, which affects the ability to accurately infer the underlying dynamics of a quantum system from noisy continuous measurement records to determine an accurate control strategy. To circumvent such limitations, this Letter explores a real-time stochastic state estimation approach that enables noise-free monitoring of the conditional dynamics including the full density matrix of the quantum system using noisy measurement records within a single quantum trajectory—a method we name as "conditional state tomography." This, in turn, enables the development of precise measurement-based feedback control strategies that lead to effective control of quantum systems by essentially mitigating the constraints imposed by measurement noise and has potential applications in various feedback quantum control scenarios. This approach is particularly useful for reinforcement-learning-(RL) based control, where the RL-agent can be trained with arbitrary conditional averages of observables, and/or the full density matrix as input (observation), to quickly and accurately learn control strategies.

DOI: 10.1103/PhysRevLett.131.210803

Future advancements in quantum technologies will hinge on the ability to effectively manipulate quantum systems by controlling their states through reliable protocols and feedback strategies [1–4]. Broadly speaking, pure control strategies entail using open-loop pulse-based controls for quantum circuits, and such problems have been successfully tackled using conventional optimal control tools like gradient-ascent pulse engineering [5–8]. These methods are fundamentally based on a differentiable model of quantum dynamics that cannot be extended to feedback-based controls [5,9]. For controls employing continuous measurement, nontrivial strategies need to be identified based on conditional dynamics. These measurement-based feedback control (MBFC) techniques are considered pivotal for achieving real-time quantum control in laboratory experiments [10–18]. Reinforcement learning (RL) has recently been proven to be a powerful new ansatz for such control tasks, which, in the quantum domain, was first demonstrated for quantum error correction [19] and optimization of quantum phase transition in 2018 [20]. Following these initial studies, we have recently witnessed its applications in different sets of nonintuitive problems, including applications in quantum control [8,21–24], state transfer [25,26], quantum state preparation and engineering [27–30], and quantum error correction [31]. Very recently, the use of RL controls for real laboratory experiments of a quantum system has become a reality [32,33].

At a fundamental level, the MBFC approaches based on continuous measurements suffer from the limitations of two primary sources. First, such approaches often fail to control the dynamics beyond a specific limit set by the signal-to-noise ratio of the intrinsic and unavoidable measurement-induced noise to the measured quantity. The level of noise increases as $1/\sqrt{\kappa\delta t}$, where $\kappa$ denotes the measurement rate, and $\delta t$ is the measurement time interval, which given the fact that $\delta t$ is related directly to the variance of the noise distribution (in the Wiener noise model) and $\delta t \ll 1$, the actual measured signal can be well hidden in the sea of random noise [24]. This makes it practically impossible for MBFC to find suitable control strategies for the system to achieve the desired dynamics. Second, the continuous measurement process naturally leads to the so-called measurement backaction, which makes the MBFC schemes highly nonintuitive and nontrivial in general [19,24,27,34,35].

In this Letter, we research in this direction and propose an efficient MBFC protocol that can precisely control the dynamics of a quantum system of interest based on noisy, continuous, and real-time measurement data. This is made possible by developing a measurement-based stochastic estimator that can extract the real-time state of the measured system noiselessly and without collapse, thereby controlling the system dynamics in any desired way. Unlike the usual method of state estimation with continuous

measurement using thousands of trajectories from that many copies of the quantum system, this method estimates the conditional state of the system from single trajectory, a method we term as "conditional state tomography." We demonstrate the efficiency of the scheme by applying it to control the dynamics of linear and nonlinear quantum systems where the applied feedback is state based or conditional. We also show the usefulness of the scheme for cases where control laws can be derived based on conditional moments (assuming perfect extraction of the measured signal from the noisy data, which is typically not possible in realistic experiments), which we illustrate with an example of preparing symmetric and antisymmetric entangled states of two qubits. Moreover, our scheme is adaptable for real-time feedback with RL controllers, allowing optimal and efficient training and control.

The protocol is shown schematically in Fig. 1. It consists of two operation steps, the estimation stage and the control stage. In the estimation stage, the to be controlled quantum system (shown on the left), with an unknown initial state [given by the density matrix $\rho(0)$] is measured using a (weak) continuous measurement approach. The noisy current streams from the measurement are then used to construct a stochastic estimator (shown on the right), which is a computational model of the measured quantum system, with the same Hamiltonian but with any random initial quantum state $\rho^e(0)$. The estimator can track the dynamics of the measured quantum system in real time after a while, as the conditional state of the estimator converges to that of the physical quantum system. In the control stage of operation, a controller is developed to mediate between the real system and the estimator by applying feedback on the systems based on the conditional dynamics of the latter while continuing to control the systems through the real-time measured data of the physical quantum system.

We first describe the theory behind the measurement-based stochastic estimator and the feedback control method. Suppose the laboratory quantum system (top left in Fig. 1), with Hamiltonian $\hat{H}_0$, is being measured continuously with a weak probe for the measurement operator $\hat{A}$ (suitably scaled to make it dimensionless). Such a continuous measurement process leads to conditional stochastic dynamics of the system density matrix in time $\rho_c(t)$ and is described by the so-called quantum stochastic master equation (SME),

$$\frac{d\rho_c(t)}{dt} = -i[\hat{H}_0, \rho_c(t)] + \kappa\mathcal{D}[\hat{A}]\rho_c(t) + \sqrt{\kappa\eta}\mathcal{H}[\hat{A}]\rho_c(t)d\xi(t). \quad (1)$$

Here, $\kappa$ is the measurement rate (the rate at which information is extracted from the detector), $\eta$ is the measurement efficiency of the detector, and $d\xi(t)$ represents an instantaneous random Wiener noise increment (white noise model with zero mean and variance $\sqrt{dt}$,

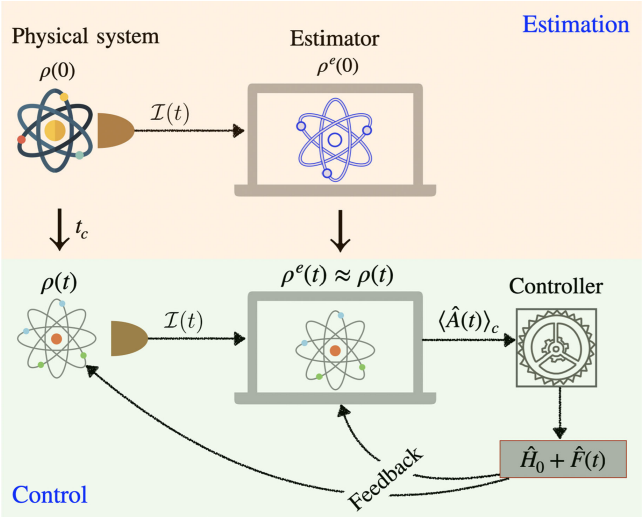


FIG. 1. The schematic of the proposed protocol. Top: the estimation stage. A physical quantum system (left) described by Hamiltonian $\hat{H}_0$ is continuously monitored to probe the observable $\hat{A}$, and the noisy measurement outcomes are fed to the estimator (right)—a *simulator* based on the mathematical model of the real physical system on a classical processor, e.g., a field programmable gate array. The state of the physical system (estimator) at time $t$ is described by $\rho(t)$ [$\rho^e(t)$] which becomes equal at $t \geq t_c$. Bottom: the control stage. A controller is used to apply accurate feedback $\hat{F}(t)$ to both the physical as well as the estimator systems as a function of the estimated noiseless conditional signal $\langle\hat{A}(t)\rangle_c$ obtained through the estimator.

where $dt$ is the time interval between successive measurements). $\mathcal{D}[\hat{A}]$ and $\mathcal{H}[\hat{A}]$ are the superoperators describing, respectively, the backaction and diffusion terms in the SME [1,3], see Supplemental Material [36] for more details. Probing the system with a weakly coupled meter that, in effect, has a broad probability distribution of the quantum state leads to noisy measurement records given by

$$\mathcal{I}(t) = \langle\hat{A}(t)\rangle_c + \frac{1}{\sqrt{4\kappa\eta}}d\xi(t). \quad (2)$$

The first term on the right-hand side of the above equation denotes the conditional mean of the measurement operator (the signal) and the second term represents the contribution of the measurement noise, which depends on $\eta$ and $\kappa$.

The estimator is a model quantum system with the same Hamiltonian $\widehat{H_0}$, as depicted in Fig. 1 (top right), which is initialized in any arbitrary quantum state $\rho^e(0)$, and is driven by the noisy measurement current of the real laboratory quantum system, $\mathcal{I}(t)$ [Eq. (2)]. The dynamics of the estimator is described by the modified SME [1,35,44],

$$d\rho_c^e(t) = -i[\hat{H}_0, \rho_c^e(t)]dt + \kappa\mathcal{D}[\hat{A}]\rho_c^e(t)dt + 2\kappa\eta[\mathcal{I}(t) - \langle A(t)\rangle_c^e]\mathcal{H}[\hat{A}]\rho_c^e(t)dt, \quad (3)$$

where $\rho_c^e(t)$ denotes the conditional density matrix of the estimator independent of the real system, and $\langle\hat{A}(t)\rangle_c^e = \mathrm{Tr}[\rho^e\hat{A}]$ is the conditional mean calculated for the estimator at time $t$. In essence, the estimator dynamics is driven by the noisy real-time measurement currents from the meter and the conditional means of the estimator itself. It can be shown that the overlap between the states $\rho(t)$ and $\rho^e(t)$ following Eqs. (1) and (3) monotonically increases until it reaches unity: $\delta\mathrm{Tr}[\rho\rho^e](t) \sim \mathrm{Tr}[\sqrt{\rho}(\hat{A}+\langle\hat{A}\rangle)\rho^e(\hat{A}+\langle\hat{A}\rangle)\times\sqrt{\rho}]\delta t$. Thus, provided the estimator gets a sufficient amount of measurement data, the convergence of its dynamic state to that of the physical quantum system, i.e., $\rho_e(t) \sim \rho(t)$ can always be guaranteed, except for the cases where $[\hat{H}_0, \hat{A}] = 0$. In case of the latter, the observable $\hat{A}$ is a constant of motion, and the continuous measurement of it does not provide any information about the state of the system, causing $\rho^e(t)$ to remain in one of the eigenstates of $\hat{H}_0$. It is possible to ensure convergence regardless of the values of $\eta$ and $\kappa$, although the time it takes to reach convergence, $t_f$, will be longer if $\eta$ and $\kappa$ are lower (see the Supplemental Material [36] for an example of a qubit to illustrate the protocol). Once this estimation stage is complete, the second stage of the MBFC scheme, namely the control stage, is initiated, as shown in Fig. 1 (bottom).

We first apply the scheme for dynamic feedback cooling of a linear quantum harmonic oscillator and demonstrate how it becomes possible to employ accurate state-based feedback control to achieve this. The Hamiltonian of the linear quantum harmonic oscillator is given by $\hat{H}_0 = \hat{p}^2/2m + m\omega^2\hat{x}^2/2$, where $\hat{x}$ and $\hat{p}$ are the position and momentum operators, respectively, $m$ is the mass of the oscillator, and $\omega$ denotes the frequency of oscillation. Consider making a measurement of the position operator such that $\hat{A} = \hat{x}$. In Fig. 2(a), the instantaneous fidelity between the states of the real system and the estimator $\mathcal{F}(t)$ is shown during the estimation stage of the control protocol. As shown in terms of the monotonically improved fidelity, the estimator starts mimicking the dynamics of the measured quantum system; also shown in the inset of Fig. 2(a), where the evolution of the conditional means of $\hat{x}$ for the measured system and estimator are compared. After the estimation stage is completed, which is typically smaller than $\kappa^{-1}$, the control stage is initialized. We now use a state-based control strategy given by $\hat{H}(t) = \hat{H}_0 - \langle\hat{x}(t)\rangle_c\hat{p}$, where $\langle\hat{x}(t)\rangle_c$ denotes the conditional mean of $\hat{x}$ at time $t$. Such a feedback represents a damping control scheme, where the controller applies feedback based on the conditional mean of the position operator to effectively reduce the momentum as it approaches $\langle\hat{x}(t)\rangle_c \to 0$. The feedback is applied to both the measured system and the estimator based on the noise-free conditional mean of the position extracted by the estimator. The results are shown in Fig. 2(b), where it is found that the proposed control protocol leads to fast and accurate dynamic cooling of the quantum harmonic oscillator. The inset of Fig. 2(b) shows how the control protocol could keep the quantum state at a dynamical minimum to any length of time, which is crucial.

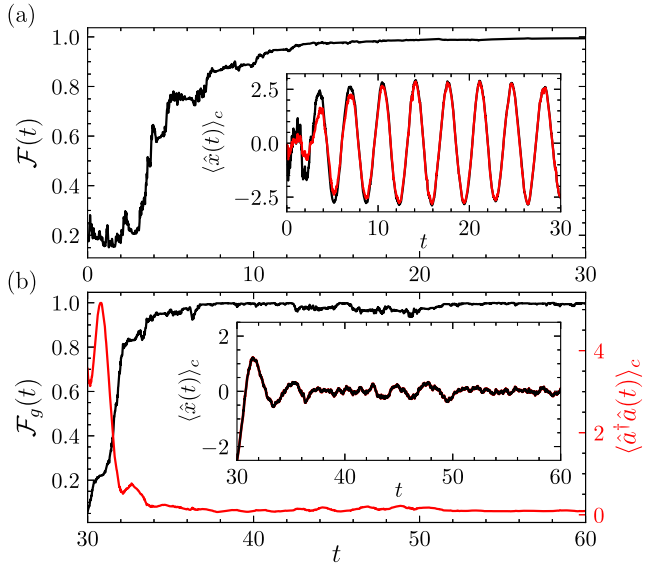


FIG. 2. Control of a linear quantum harmonic oscillator using the protocol. (a) In the estimation phase, the fidelity $\mathcal{F}(t)$ between the physical system and the estimator steadily converges. The inset displays the conditional means of the observable $\hat{x}$. (b) Subsequently, a state-based controller is applied, swiftly guiding the particle's motion around the center $\langle\hat{x}(t)\rangle_c = 0$. The instantaneous fidelity $\mathcal{F}_g(t)$ (depicted in black) quantifies the closeness between the physical system or estimator state and the target state, i.e., the ground state of the oscillator. The conditional mean population $\langle\hat{a}^\dagger\hat{a}(t)\rangle_c$ in the oscillator is shown in red.

Next, we consider a nonlinear quartic potential with the unperturbed Hamiltonian given by $\hat{H}_0 = \hat{p}^2/2m + \lambda\hat{x}^4$, where we have chosen $m = 1/\pi$ and $\lambda = \pi/25$ with proper dimensions. We apply artificial control viz. RL [37–39], to devise proper feedback strategies in this case. It is noteworthy that with the designed stochastic estimator, it is now possible to apply the full density matrix as well as the means and moments of the operators for choosing any accurate feedback scheme. Therefore, the scheme allows using accurate conditional means of observables as the input $s_t$ (observation) to the RL agent; for example, here we use $s_t = \{\langle\hat{x}\rangle, \langle\hat{p}\rangle\}$. Another advantage of the estimator control is that state fidelities are now realizable, which are usually pervasive in real experimental measurements. Therefore, given that we have access to the fidelity $\mathcal{F}(t)$ of the estimator, it can be used as a simple and efficient reward function that needs to be maximized by the RL agent in the training process. The agent is first trained with a given initial state, which, due to the generalizability of the trained model, permits use for controlling the system started with other (random) initial states. The learning curve as the mean fidelity $\overline{\mathcal{F}}(N)$ over each training episode

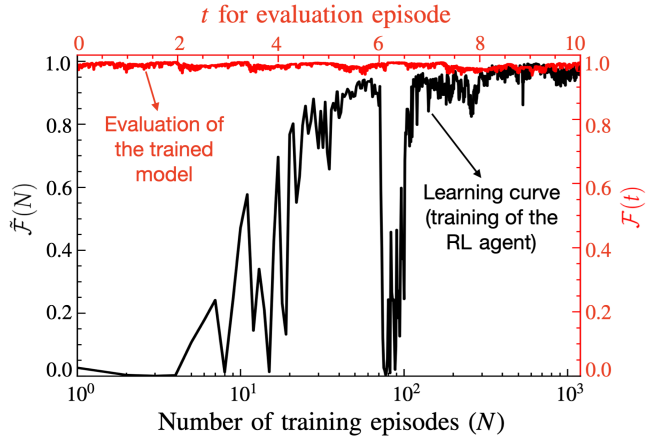


FIG. 3. The protocol is applied to control a particle's motion in a nonlinear quartic potential to cool it to its dynamic ground state using RL-based control. The training process is shown in a black colored line as the average fidelity over each episode $N$ with respect to the target state (ground state) $\bar{\mathcal{F}}(N)$, which is maximized through training. Note that the sudden drop at $N \sim 100$ is due to the exploration of the RL agent. The performance of the trained agent is shown in the red line.

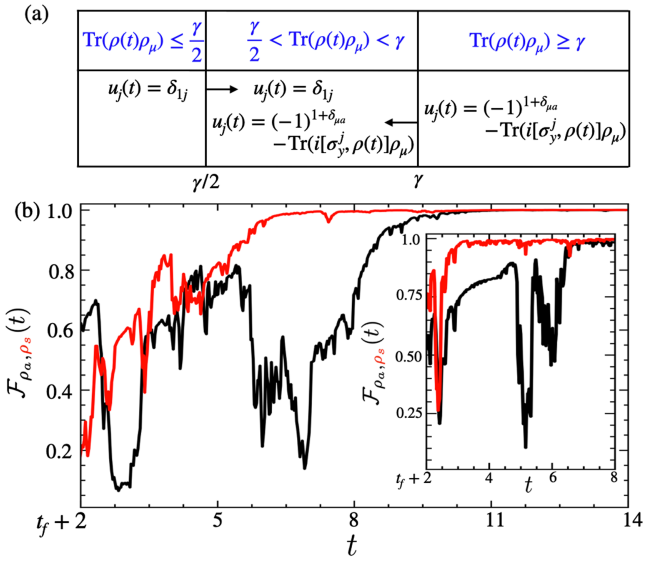


FIG. 4. Demonstration of the proposed MBFC protocol for the preparation of symmetric, $\rho_s$, and antisymmetric, $\rho_a$, entangled states between two qubits as an example for when it is possible to derive control laws based on conditional moments within stochastic dynamics. Control laws $u_1$ and $u_2$ are selected depending on the conditional value of $\rho(t)\rho_\mu$, where $\mu \in \{s, a\}$ (symmetric and antisymmetric) are in the three regimes, conveniently demonstrated in (a), and the arrows represent the direction of the entrance boundary of $\rho(t)$ to the middle section. $\gamma$ is the damping parameter, the measurement rate $\kappa$ is assumed to be 0.1, and the efficiency $\eta = 0.5$ for this simulation. After the estimation stage (not shown), these control laws are applied on conditional mean data (density matrices to compute instantaneous fidelity), which leads to convergence to the target states ($\rho_a$: black and $\rho_s$: red), shown in (b). In the absence of such laws, RL can be used—the performance is shown in the inset of figure (b) with similar color settings.

$N$ is shown in black in Fig. 3. Using conditional means for training the RL agent makes learning quicker and more accurate. The evaluated episodic fidelity variation $\mathcal{F}(t)$ is shown in the red colored line in Fig. 3 in the biaxial plot's second scale, demonstrating accurate feedback control by the trained RL model.

Besides, it is often possible to derive control laws for systems undergoing continuous measurement based on the conditional means of observables (without the noise component). Although such control laws would not have much value in realistic situations due to the unavailability of accurate noiseless signal, we now show in the following that in such context too, our proposed scheme would be useful. To illustrate it, we consider the preparation of symmetric ($\rho_s$) and antisymmetric ($\rho_a$) entangled states of two qubits, where the states are given by $\rho_{(s/a)} = \frac{1}{2}(\psi_{\uparrow\downarrow} \pm \psi_{\downarrow\uparrow})(\psi_{\uparrow\downarrow} \pm \psi_{\downarrow\uparrow})^*$. Here, $\psi_{\uparrow\downarrow} = (\uparrow) \otimes (\downarrow)$ and $\psi_{\downarrow\uparrow} = (\downarrow) \otimes (\uparrow)$ are the tensor product states of the individual qubit states in the ground and excited states. The quantum filtering equation under feedback with control variables $u_1(t)$ and $u_2(t)$ is given by

$$\begin{aligned} d\rho(t) = &-iu_1(t)[\sigma_y^{(1)}, \rho(t)]dt - iu_2(t)[\sigma_y^{(2)}, \rho(t)]dt \\ &-\frac{1}{2}[F_z, [F_z, \rho(t)]]dt + \sqrt{\eta}\{F_z\rho(t) + \rho(t)F_z \\ &- 2\mathrm{Tr}[F_z\rho(t)]\rho(t)\}dW_t, \end{aligned} \tag{4}$$

where $dW_t$ is the Winner noise increment at time $t$. $\sigma_g^i$, $g \in \{x, y, z\}$ and $i = \{1, 2\}$ are tensored Pauli operators for qubit $i$, and $F_z = \sigma_z^1 + \sigma_z^2$ [40]. The control laws dictate nonintuitive choices of the control parameters $u_1(t)$ and $u_2(t)$ provided the real-time conditional fidelity between the current and the target states $\rho_s$ and $\rho_a$ could be accurately extracted via conditional tomography of the quantum states, which is often a difficult task if not impossible. These are discussed in Supplemental Material [36] and conveniently represented in Fig. 4(a). Using these control laws with the MBFC scheme makes it possible to evaluate the controls $u_1(t)$ and $u_2(t)$ in real time, which leads to a guaranteed preparation of the states $\rho_a$ and $\rho_s$, shown in black and red lines, respectively, in Fig. 4(b). It becomes also possible to use RL for control similar to the case shown for a quartic oscillator above, in which case one can use the full density matrix for training along with conditional means, and the performance is shown in the inset of the figure. Compared to the control laws, the RL controller can help the system reach its target state in a shorter time scale.

Finally, we will mention possible shortcomings of the proposed scheme. First, the protocol leans toward a model-based approach, aiming to maximize controlled output accuracy based on a highly precise physical model, and

therefore one should care about potential model bias. To remove model bias, one can integrate model learning techniques such as Hamiltonian learning beforehand [45]. It is also possible to use machine learning techniques such as Bayesian estimation [46] and RL [47,48] for estimating model parameters. Second, when dealing with real-time feedback control problems, it is likely to have potential delays between the measurement and feedback operations. The estimator, being a simulator in a classical processor, needs finite time for simulation that can add to this delay event, especially for large systems. While the estimation stage of the protocol can be streamlined by completing it in a single pass by providing all previous measurement results at once to the estimator; for the control stage, it would be advantageous to provide the estimator and the controller with frequent measurement results, to discover finer controllability, tailored to the system's complexity. In such cases, RL-based methods can be especially effective [32,33].

In conclusion, even when employing sophisticated noise filtering techniques such as linear quadratic regulator, linear quadratic Gaussian, and Kalman filters in standard MBFC experiments, extracting the exact signal from the noisy measurement results remains a formidable task [18]. Consequently, conventional feedback strategies fall short of achieving accurate control. The proposed protocol circumvents this by estimating accurate conditional state tomography, thereby enabling precise quantum feedback control within the realm of continuous measurement. Furthermore, this protocol integrates seamlessly with RL-based control methods, enabling efficient training and control.

The authors thank Gerard Milburn, Jason Twamley, and Michael Kewming for useful discussions.

*sangkha.borah@mpl.mpg.de
†bijita.sarma@fau.de

[1] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control* (Cambridge University Press, Cambridge, England, 2009).
[2] J. Zhang, Y.-x. Liu, R.-B. Wu, K. Jacobs, and F. Nori, Quantum feedback: Theory, experiments, and applications, Phys. Rep. **679**, 1 (2017).
[3] K. Jacobs, *Quantum Measurement Theory and Its Applications* (Cambridge University Press, Cambridge, England, 2014).
[4] A. C. Doherty, S. Habib, K. Jacobs, H. Mabuchi, and S. M. Tan, Quantum feedback control and classical control theory, Phys. Rev. A **62**, 012105 (2000).
[5] P. de Fouquieres, S. G. Schirmer, S. J. Glaser, and I. Kuprov, Second order gradient ascent pulse engineering, J. Magn. Reson. **212**, 412 (2011).
[6] O. V. Morzhin and A. N. Pechen, Krotov method for optimal control of closed quantum systems, Russ. Math. Surv. **74**, 851 (2019).
[7] C. P. Koch, U. Boscain, T. Calarco, G. Dirr, S. Filipp, S. J. Glaser, R. Kosloff, S. Montangero, T. Schulte-Herbrüggen, D. Sugny, and F. K. Wilhelm, Quantum optimal control in quantum technologies. Strategic report on current status, visions and goals for research in Europe, Eur. Phys. J. Quantum Technol. **9**, 19 (2022).
[8] B. Sarma, S. Borah, A. Kani, and J. Twamley, Accelerated motional cooling with deep reinforcement learning, Phys. Rev. Res. **4**, L042038 (2022).
[9] T. Propson, B. E. Jackson, J. Koch, Z. Manchester, and D. I. Schuster, Robust quantum optimal control with trajectory optimization, Phys. Rev. Appl. **17**, 014036 (2022).
[10] L. S. Martin, W. P. Livingston, S. Hacohen-Gourgy, H. M. Wiseman, and I. Siddiqi, Implementation of a canonical phase measurement with quantum feedback, Nat. Phys. **16**, 1046 (2020).
[11] S. Kuang, G. Li, Y. Liu, X. Sun, and S. Cong, Rapid feedback stabilization of quantum systems with application to preparation of multiqubit entangled states, IEEE Trans. Syst. Man Cybern. **52**, 1 (2021).
[12] M. Rossi, D. Mason, J. Chen, Y. Tsaturyan, and A. Schliesser, Measurement-based quantum control of mechanical motion, Nature (London) **563**, 53 (2018).
[13] R. Vijay, C. Macklin, D. H. Slichter, S. J. Weber, K. W. Murch, R. Naik, A. N. Korotkov, and I. Siddiqi, Stabilizing Rabi oscillations in a superconducting qubit using quantum feedback, Nature (London) **490**, 77 (2012).
[14] F. Tebbenjohanns, M. L. Mattana, M. Rossi, M. Frimmer, and L. Novotny, Quantum control of a nanoparticle optically levitated in cryogenic free space, Nature (London) **595**, 378 (2021).
[15] D. J. Wilson, V. Sudhir, N. Piro, R. Schilling, A. Ghadimi, and T. J. Kippenberg, Measurement-based control of a mechanical oscillator at its thermal decoherence rate, Nature (London) **524**, 325 (2015).
[16] W. P. Livingston, M. S. Blok, E. Flurin, J. Dressel, A. N. Jordan, and I. Siddiqi, Experimental demonstration of continuous quantum error correction, Nat. Commun. **13**, 1 (2022).
[17] L. Magrini, P. Rosenzweig, C. Bach, A. Deutschmann-Olek, S. G. Hofer, S. Hong, N. Kiesel, A. Kugi, and M. Aspelmeyer, Real-time optimal quantum control of mechanical motion at room temperature, Nature (London) **595**, 373 (2021).
[18] R. Jiménez-Martínez, J. Kołodyński, C. Troullinou, V. G. Lucivero, J. Kong, and M. W. Mitchell, Signal tracking beyond the time resolution of an atomic sensor by Kalman filtering, Phys. Rev. Lett. **120**, 040503 (2018).
[19] T. Fösel, P. Tighineanu, T. Weiss, and F. Marquardt, Reinforcement learning with neural networks for quantum feedback, Phys. Rev. X **8**, 031084 (2018).
[20] M. Bukov, A. G. R. Day, D. Sels, P. Weinberg, A. Polkovnikov, and P. Mehta, Reinforcement learning in different phases of quantum control, Phys. Rev. X **8**, 031086 (2018).
[21] Z. T. Wang, Y. Ashida, and M. Ueda, Deep reinforcement learning control of quantum cartpoles, Phys. Rev. Lett. **125**, 100401 (2020).
[22] M. Y. Niu, S. Boixo, V. N. Smelyanskiy, and H. Neven, Universal quantum control through deep reinforcement learning, npj Quantum Inf. **5**, 1 (2019).

[23] X.-M. Zhang, Z. Wei, R. Asad, X.-C. Yang, and X. Wang, When does reinforcement learning stand out in quantum control? A comparative study on state preparation, npj Quantum Inf. **5**, 1 (2019).
[24] S. Borah, B. Sarma, M. Kewming, G. J. Milburn, and J. Twamley, Measurement-based feedback quantum control with deep reinforcement learning for a double-well nonlinear potential, Phys. Rev. Lett. **127**, 190403 (2021).
[25] R. Porotti, D. Tamascelli, M. Restelli, and E. Prati, Coherent transport of quantum states by deep reinforcement learning, Commun. Phys. **2** (2019).
[26] I. Paparelle, L. Moro, and E. Prati, Digitally stimulated Raman passage by deep reinforcement learning, Phys. Lett. A **384**, 126266 (2020).
[27] R. Porotti, A. Essig, B. Huard, and F. Marquardt, Deep reinforcement learning for quantum state preparation with weak nonlinear measurements, Quantum **6**, 747 (2022).
[28] X.-M. Zhang, Z. Wei, R. Asad, X.-C. Yang, and X. Wang, When does reinforcement learning stand out in quantum control? A comparative study on state preparation, npj Quantum Inf. **5**, 85 (2019).
[29] J. Mackeprang, D. B. Rao Dasari, and J. Wrachtrup, A reinforcement learning approach for quantum state engineering, Quantum Mach. Intell. **2**, 1 (2020).
[30] T. Haug, W.-K. Mok, J.-B. You, W. Zhang, C. Eng Png, and L.-C. Kwek, Classifying global state preparation via deep reinforcement learning, Mach. Learn. Sci. Technol. **2**, 01LT02 (2020).
[31] H. P. Nautrup, N. Delfosse, V. Dunjko, H. J. Briegel, and N. Friis, Optimizing quantum error correction codes with reinforcement learning, Quantum **3**, 215 (2019).
[32] K. Reuer, J. Landgraf, T. Fösel, J. O'Sullivan, L. Beltrán, A. Akin, G. J. Norris, A. Remm, M. Kerschbaum, J.-C. Besse, F. Marquardt, A. Wallraff, and C. Eichler, Realizing a deep reinforcement learning agent discovering real-time feedback control strategies for a quantum system, Nat. Commun. **14**, 7138 (2023).
[33] V. V. Sivak, A. Eickbusch, B. Royer, S. Singh, I. Tsioutsios, S. Ganjam, A. Miano, B. L. Brock, A. Z. Ding, L. Frunzio, S. M. Girvin, R. J. Schoelkopf, and M. H. Devoret, Real-time quantum error correction beyond break-even, Nature (London) **616**, 50 (2023).
[34] S. Hacohen-Gourgy and L. S. Martin, Continuous measurements for control of superconducting quantum circuits, Adv. Phys. **5**, 1813626 (2020).
[35] J. Zhang, Y.-x. Liu, R.-B. Wu, K. Jacobs, and F. Nori, Quantum feedback: Theory, experiments, and applications, Phys. Rep. **679**, 1 (2017).
[36] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.131.210803 for a brief theory of continuous quantum measurement, continuous feedback control, and reinforcement learning-based closed-loop control, as well as technical details of the implementation and additional results, which contains Refs. [1,3,24,27,32,37–43].
[37] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. (The MIT Press, Cambridge, MA, 2018).
[38] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, Proximal policy optimization algorithms, arXiv:1707.06347.
[39] J. Schulman, S. Levine, P. Moritz, M. I. Jordan, and P. Abbeel, Trust region policy optimization, arXiv:1502.05477.
[40] M. Mirrahimi and R. Van Handel, Stabilizing feedback controls for quantum systems, SIAM J. Control Optim. **46**, 445 (2007).
[41] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning* (MIT Press, Cambridge, MA, 2016), Vol. 1.
[42] A. Essig, Q. Ficheux, T. Peronnin, N. Cottet, R. Lescanne, A. Sarlette, P. Rouchon, Z. Leghtas, and B. Huard, Multiplexed photon number measurement, Phys. Rev. X **11**, 031045 (2021).
[43] V. Gebhart, R. Santagati, A. A. Gentile, E. M. Gauger, D. Craig, N. Ares, L. Banchi, F. Marquardt, L. Pezzè, and C. Bonato, Learning quantum systems, Nat. Rev. Phys. **5**, 141 (2023).
[44] L. Diósi, T. Konrad, A. Scherer, and J. Audretsch, Coupled Ito equations of continuous quantum state measurement and estimation, J. Phys. A **39**, L575 (2006).
[45] V. Gebhart, R. Santagati, A. A. Gentile, E. M. Gauger, D. Craig, N. Ares, L. Banchi, F. Marquardt, L. Pezzè, and C. Bonato, Learning quantum systems, Nat. Rev. Phys. **5**, 141 (2023).
[46] S. Nolan, A. Smerzi, and L. Pezzè, A machine learning approach to Bayesian parameter estimation, npj Quantum Inf. **7**, 1 (2021).
[47] T. Xiao, J. Fan, and G. Zeng, Parameter estimation in quantum sensing based on deep reinforcement learning, npj Quantum Inf. **8**, 1 (2022).
[48] H. Xu, J. Li, L. Liu, Y. Wang, H. Yuan, and X. Wang, Generalizable control for quantum parameter estimation through reinforcement learning, npj Quantum Inf. **5**, 1 (2019).