Detection of Entangled States Supported by Reinforcement Learning

Jia-Hao Cao,[1] Feng Chen,[1] Qi Liu,[1,*] Tian-Wei Mao,[1] Wen-Xin Xu,[1] Ling-Na Wu,[2,†] and Li You[1,3,4,‡]
[1]State Key Laboratory of Low Dimensional Quantum Physics, Department of Physics, Tsinghua University, Beijing 100084, China
[2]Center for Theoretical Physics and School of Science, Hainan University, Haikou 570228, China
[3]Frontier Science Center for Quantum Information, Beijing, China
[4]Beijing Academy of Quantum Information Sciences, Beijing 100193, China

(Received 6 February 2022; revised 5 June 2023; accepted 14 July 2023; published 14 August 2023)

Discrimination of entangled states is an important element of quantum-enhanced metrology. This typically requires low-noise detection technology. Such a challenge can be circumvented by introducing nonlinear readout process. Traditionally, this is realized by reversing the very dynamics that generates the entangled state, which requires a full control over the system evolution. In this Letter, we present nonlinear readout of highly entangled states by employing reinforcement learning to manipulate the spin-mixing dynamics in a spin-1 atomic condensate. The reinforcement learning found results in driving the system toward an unstable fixed point, whereby the (to be sensed) phase perturbation is amplified by the subsequent spin-mixing dynamics. Working with a condensate of 10 900 ^{87}Rb atoms, we achieve a metrological gain of $6.97^{+1.30}_{-1.38}$ dB beyond the classical precision limit. Our work will open up new possibilities in unlocking the full potential of entanglement caused quantum enhancement in experiments.

DOI: 10.1103/PhysRevLett.131.073201

Introduction.—Entanglement plays a crucial role in quantum metrology [1,2], which aims at beating the standard quantum limit (SQL) or classical limit achievable with uncorrelated particles by using quantum resources. To fully harness the quantum advantages offered by entangled states, it is necessary to discriminate between entangled states with and without parameter perturbations. However, detecting entangled states is often vulnerable to technical noise, compromising the metrological advantage derived from using entanglement. To overcome this challenge, nonlinear readout techniques have been introduced [3–5], based on amplifying the distinction between perturbed and unperturbed entangled states by disentangling the particles [see Fig. 1(a)]. Such a nonlinear process exhibits high sensitivity to perturbations, producing markedly different output states even with minute perturbations, and thus promises quantum-enhanced sensing via signal amplification.

A straightforward implementation of nonlinear readout of entangled states comes from time-reversing the very dynamics that generates the entanglement in the first place [6–12]. Such a procedure, however, requires a precise knowledge of the history of the dynamical process, as well as a full control over the system, which is highly nontrivial for a general many-body quantum system. In this Letter, we demonstrate an easily implementable nonlinear readout process supported by reinforcement learning (RL) [13] that requires no knowledge of entanglement generation history and is realized by modulating only a linear control field.

As an important branch of machine learning, RL targets an optimal strategy for accomplishing a specific task

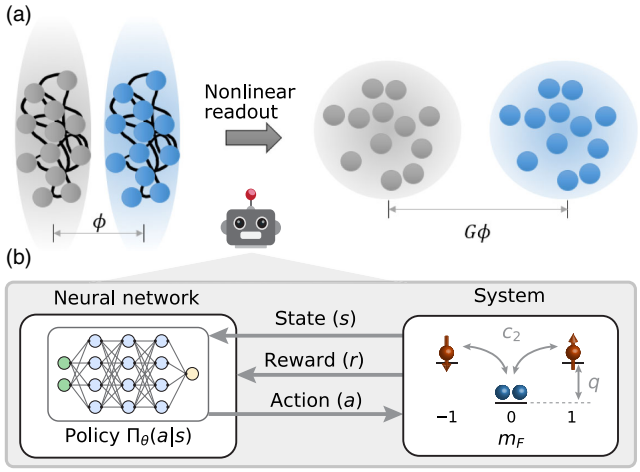

FIG. 1. (a) The precision of measurements is constrained by the quantum fluctuations of probe states, as depicted by the shaded region in the figure. Entangled states with squeezed fluctuation distributions (the fluctuation along the phase-encoding direction is squeezed, as a cost, the fluctuation along the orthogonal direction is enlarged) allow for more precisely estimating parameter ϕ than classical resources (left panel). The benefit of squeezed noise, however, can be easily overshadowed by detection noise, which, combined with quantum fluctuations, establishes the overall noise level. By introducing a nonlinear readout process (right panel), which preferentially amplifies signal over noise, noise-robust detection with high phase sensitivity can be realized. This work exploits RL to guide nonlinear readout of highly entangled states. (b) A typical interaction loop between neural network and system for RL training. In this work, the system is a spin-1 condensate. The QZS q is modulated according to the policy from RL to manipulate the dynamics.

without prior knowledge. It optimizes decision making based on interacting with the system instead of hints or guidances to the solutions. In quantum science and technology, RL has been applied to quantum state engineering [14–21], quantum control [22–26], quantum metrology [27–30], quantum error correction [31–33], and quantum compiling [34–36], etc., with great successes. Here, we employ RL to guide nonlinear readout of highly entangled states for quantum enhanced sensing.

Our system is composed of a ^{87}Rb atomic condensate in the ground hyperfine $F = 1$ manifold $(m_F = 0, \pm 1)$, described by the Hamiltonian ($\hbar = 1$ hereafter)

$$H = \frac{c_2}{2N}[(2a_0^\dagger a_0^\dagger a_1 a_{-1} + \text{H.c.}) + (2N_0 - 1)(N - N_0)] - q(t)N_0$$

under the assumption of the same spatial mode for the three spin components [38]. Here, a_{m_F} ($a_{m_F}^\dagger$) denotes annihilation (creation) operator and $N_{m_F} = a_{m_F}^\dagger a_{m_F}$ counts the number of atoms in m_F component with $N = \sum_{m_F} N_{m_F}$ the total number of atoms. The Hamiltonian can also be written as $H = (c_2/2N)\boldsymbol{L}^2 - q(t)N_0$ in terms of collective spin $\boldsymbol{L} \equiv \sum_{\mu,\nu} a_\mu^\dagger \boldsymbol{F}_{\mu\nu} a_\nu$ with $\boldsymbol{F}_{\mu\nu}(\mu, \nu = -1, 0, +1)$ the spin-1 matrix element. It conserves magnetization $L_z = N_{+1} - N_{-1}$, as well as the total particle number N. The first term in the Hamiltonian describes spin-exchange interaction of ferromagnetic type at strength $c_2 < 0$, whereby atoms in $m_F = 0$ are transferred to $m_F = \pm 1$ in pairs and vice versa. The second term describes an effective quadratic Zeeman shift (QZS) q which can be tuned experimentally [39]. The competition between the above interactions gives rise to intriguing spin-mixing dynamics [38]. In this work, q is modulated according to the policy from RL to manipulate the dynamics.

Reinforcement learning.—In a typical RL task [see Fig. 1(b)], the agent learns from trial and error to achieve a prespecified goal. By inspecting the system through some observables (state), the agent makes decisions (action) according to a certain tactic (policy, a mapping between state and action) to alter the state of the system. At the same time, it collects a feedback (reward) from the system, which measures whether the decision is constructive or not. After many rounds of such interactions, the agent refines the tactic based on the collected information and updates continuously until it gains sufficient experience to arrive at an optimal tactic for achieving the goal (see Supplemental Material [40] for more details).

We first benchmark the RL training in a small system with $N = 50$ atoms. In the quantum metrology application concerned here, our aim is to precisely detect a phase ϕ that is encoded by a rotation operation described by $U_\phi = e^{-i\phi L_x}$ with $L_x = (a_1^\dagger a_0 + a_0^\dagger a_{-1} + \text{H.c.})/\sqrt{2}$. Starting from the polar state with all atoms in $m_F = 0$ component, the probe state is prepared by targeting the balanced spin-1 Dicke state [17,40], which is the ground state of the system at $q = 0$. Instead of direct detection after phase encoding [41], we employ RL to guide a nonlinear readout operation before the measurement, thereby inducing significant distinctions between the final states evolved from without and with encoding rotation U_ϕ. The encoded phase is extracted by measuring the fractional population in $m_F = 0$ component, ρ_0. Inferred from error propagation, the phase sensitivity of ϕ reads

$$\Delta\phi = \Delta\rho_0/|\partial_\phi\langle\rho_0\rangle|, \tag{1}$$

which is determined by the fluctuation of ρ_0 and the slope of its mean with respect to ϕ. The reward of RL training is set to maximize the metrological gain, $-20\log_{10}(\Delta\phi/\Delta\phi_{\text{SQL}})$, over the three-mode SQL $\Delta\phi_{\text{SQL}} = 1/(2\sqrt{N})$.

We present the results of RL training for a global search (without constraints on q) in Fig. 2(a) by blue data, with the q profile for the nonlinear readout process, the corresponding metrological gain, and ρ_0 evolution (in the absence of phase encoding) shown from top to bottom. From the ρ_0 evolution, one can see that the learnt readout operation drives the system toward the initial polar state where all atoms occupy the $m_F = 0$ component ($\rho_0 = 1$). Similar behavior is found when noises due to experimental imperfections are taken into account [40].

The pivotal role played by the polar state in achieving high phase sensitivity is further elaborated in the learning progress of readout operation; see Fig. 2(b). Here, we inspect the metrological performance of the readout processes governed by a collection of $q(t)$ profiles. For each $q(t)$ profile, one can extract the maximal ρ_0 during the readout process and the maximal metrological gain. A summary of them from 2000 trajectories each for three different training epochs (0, 10, 100) is shown in Fig. 2(b). As the learning proceeds, the distribution shrinks toward the upper right corner where achieving enhanced metrological performance and passing through the initial polar state (large maximal ρ_0) are found to be strongly correlated.

One can understand the observed behavior by inspecting the system from phase space spanned by mean fields [37], ρ_0 and the spinor phase $\theta = \theta_1 + \theta_{-1} - 2\theta_0$, where θ_{m_F} denotes the phase of the m_F component. These two variables are constrained by the single particle mean field energy $\varepsilon = c_2\rho_0(1 - \rho_0)(1 + \cos\theta) - q\rho_0$. The phase space for $|q| < 2|c_2|$ [see top panel of Fig. 2(d)] is divided by a separatrix (thick black line). The polar state sits at the north pole, which is an unstable fixed point of the system for $0 < q < 2|c_2|$ [see the energy surface in the middle panel of Fig. 2(d)], rendering the dynamics starting from it highly susceptible to perturbations. This explains why the readout operation drives the system toward the dynamically unstable polar state.

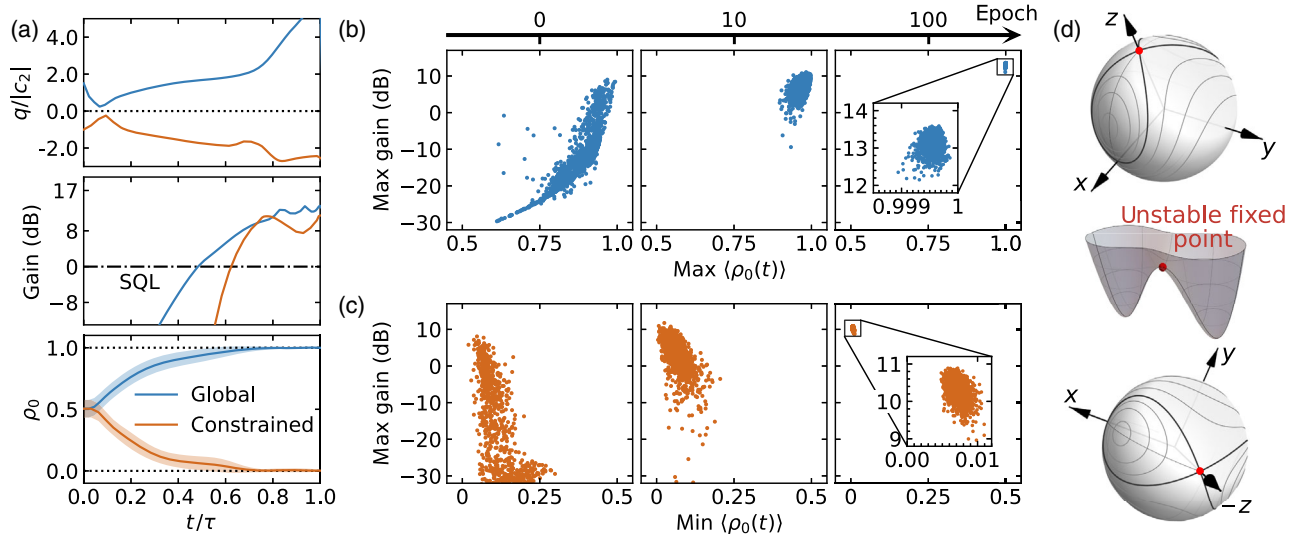

FIG. 2. (a) RL training results for the readout operation in a small system of $N = 50$ particles, with the q profile, the metrological gain, and fractional population ρ_0 evolution (in the absence of phase encoding) shown from top to bottom. The blue (orange) data denotes the results for a global (constrained) search without (with) constraints ($q < 0$) on q. The shaded regions denote the fluctuation of ρ_0, $\Delta\rho_0$. (b) Maximal achievable metrological gain vs maximal accessible ρ_0 during the readout process from 2000 trajectories for each sampled according to the policy at 0, 10, and 100 epoch. As the learning proceeds (from left to right), the readout operation tends to drive the system toward the polar state with a large $\langle\rho_0\rangle$ for getting a large metrological gain. (c) Similar results as (b), but for the case with $q < 0$ constraint. (d) Spin-nematic sphere [37] with energy contours for $q = |c_2|$ (top) and $q = -|c_2|$ (bottom). The three axes are $x = \sqrt{1-z^2}\cos(\theta/2)$, $y = \sqrt{1-z^2}\sin(\theta/2)$, and $z = 2\rho_0 - 1$, with θ being the spinor phase. Middle panel: energy surface. The polar state (marked with a red sphere) sits at the north pole, which is a saddle point for $0 < q < 2|c_2|$. The south pole becomes an unstable fixed point when $-2|c_2| < q < 0$.

To justify our interpretation, we perform another RL training, with q constrained in the negative regime, where the unstable fixed point shifts to the south pole [see bottom panel of Fig. 2(d)]. If our above understanding is correct, the optimal strategy is expected to drive the system toward the south pole with $\rho_0 = 0$. This is indeed what we observe, as shown by the orange data in Figs. 2(a) and 2(c). It is worthy to point out these results suggest that for realizing nonlinear readout it is in fact not necessary to return to the starting point of entanglement generation, which was unanimously believed as a guiding principle in traditional protocols [3–10,12,42]. Such an interesting finding offers an alternative to implementing nonlinear readout.

We make use of the above observation to simplify the training task for larger systems. The calculations of phase sensitivity [Eq. (1)] involve time-consuming computations of various output states under a span of encoded phases. To simplify the numerics, we replace the training task for large systems by targeting the initial polar state. To further speed up training and avoid sparse reward, we adopt transfer learning [40], with experience gained from trained neural networks for smaller sized systems in the absence of atom loss applied as initial values to larger or/and realistic experimental systems with loss. This benefits from the remarkable generalization ability of the RL policy [17]. For treating large systems including atom loss, truncated Wigner approximation method [40,42–45,55–59] is used in our numerical simulations.

Experimental implementation.—The RL protocols are implemented experimentally in a spinor condensate of about 10 900 ^{87}Rb atoms at $c_2 = -2\pi \times 2.6(2)$ Hz. The net QZS $q = q_B + q_{\mathrm{MW}}$ includes a magnetic field contribution q_B, and a microwave dressing field (detuned from the $|F = 1, m_F = 0\rangle$ to $|F = 2, m_F = 0\rangle$ clock transition) contribution q_{MW} [60]. The bias magnetic field is set at 0.537 G corresponding to $q_B \sim 8|c_2|$ as a trade-off between minimizing the influence of rf noise and maintaining the stability of QZS [40]. The experiment starts from a polar state BEC at $q \sim 17|c_2|$, followed by quenching q to $1.5|c_2|$ through tuning the microwave power before the state generation operation. Subsequently, q is ramped to prepare the probe states [40], which serve as the input states for the nonlinear readout operations. Here, three different probe states are studied, prepared by targeting the ground states of the system at $q_{\mathrm{tg}} = 0, 0.5|c_2|$, and $|c_2|$, respectively. The metrological potential of a probe state $|\psi_p\rangle$ is limited by the quantum Cramér-Rao bound [2] to, $(\Delta\phi)^2 \geqslant 1/F_Q(|\psi_p\rangle, L_x)$, with the quantum Fisher information $F_Q(|\psi_p\rangle, L_x) = 4(\Delta L_x)^2 = 4\langle L_x^2\rangle = 2\langle \boldsymbol{L}^2\rangle$ in the $L_z = 0$ subspace for a pure state. In our system, the ground state possesses larger collective spin length $\langle \boldsymbol{L}^2\rangle$ than excited states, and $\langle \boldsymbol{L}^2\rangle$ increases when $|q|$ decreases [see Fig. 3(a)], with the balanced spin-1 Dicke state at $q = 0$ offering the highest quantum Fisher information, and thus the highest theoretical phase sensitivity. On the other hand, preparing the ground

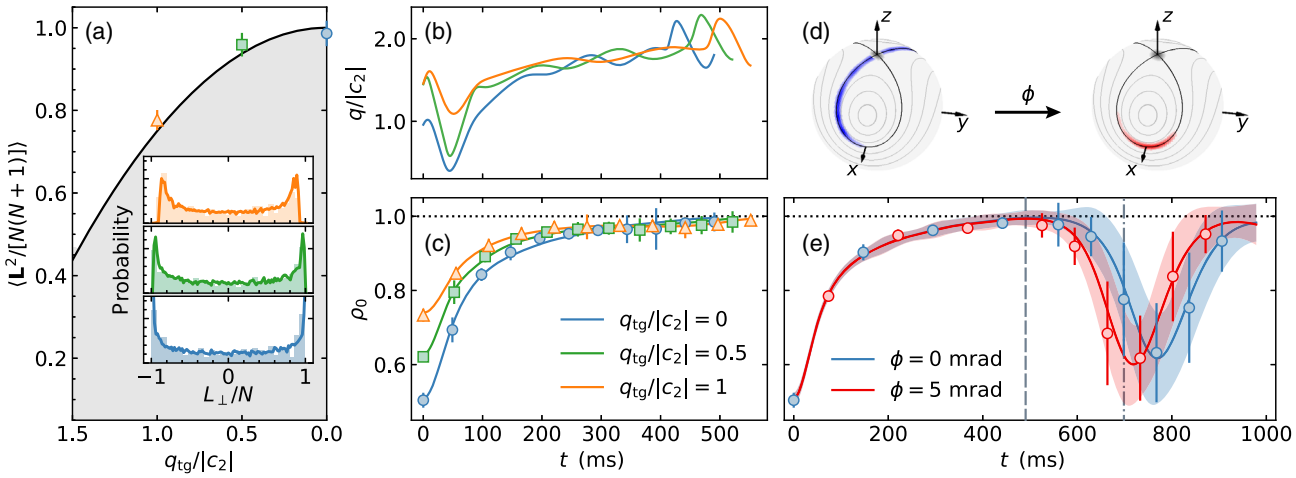

FIG. 3. (a) The collective spin length of the spinor BEC. The black solid line (gray shaded region) denotes theoretical results of the ground state (excited states). The markers denote results of the correspondingly prepared target probe states at $q_{tg}/|c_2| = 0$ (blue), 0.5 (green), and 1 (orange) from 500 continuous experimental runs on measuring transverse spins [40], whose distributions are shown in the insets. The solid lines in the insets come from numerical simulations. (b) The RL-learnt q profile for the nonlinear readout of the three probe states. (c) The evolution of ρ_0 during the readout process. (d) Illustration of states at the end of the RL-learnt readout operation [$t = 491$ ms, marked by vertical dashed line in (e)] (gray data) and at the detection time [$t = 699$ ms, marked by vertical dot-dashed line in (e)] (colored data) on the spin-nematic sphere. The left (right) panel denotes the case when no (a) phase is encoded into the probe state. (e) The evolution of ρ_0 when no phase (blue data) or a phase (red data) is encoded to the balanced spin-1 Dicke state prepared at $t = 0$. The QZS is $q = |c_2|$. For (c) and (e), markers are data from 100 experimental runs. The solid lines (shaded regions) denote numerically computed mean (uncertainty).

state by ramping q from $q \gg |c_2|$ to smaller q_{tg} requires longer time [40], and thus suffers more from atom loss. Therefore, it is nontrivial to predict which state provides the highest phase sensitivity in experiment.

Phase encoding is enacted by a well-calibrated Rabi rotation of the state using rf field that SU(2)-symmetrically couples all three m_F components. To detect the phase in a noise-robust way, we perform nonlinear readout operation guided by RL. Namely, q is ramped according to the RL profile, as shown in Fig. 3(b). The evolution of ρ_0 during the readout process, measured by absorption imaging after spatial Stern-Gerlach separation [40], is shown in Fig. 3(c). The experimental data (markers) shows an excellent agreement with theory (solid lines). Approaching $\rho_0 = 1$ at the end of the readout process signals the success of the RL protocol.

To illustrate the sensitivity of the system to phase perturbation, we compare the dynamics of the system when a finite phase (red data) or no phase (blue data) is encoded to the probe state (balanced spin-1 Dicke state). The results are shown in Figs. 3(d) and 3(e). Before the system approaches the polar state ($t < 491$ ms), one can hardly distinguish these two cases, while a substantial difference is observed after the system passes through the polar state ($t > 491$ ms). Such a sensitive dependence of ρ_0 on phase perturbation exemplifies the discrimination of the encoded states with slightly different phases.

We show the metrological performance of the three probe states in Fig. 4. The mean value of ρ_0 and its fluctuation $\Delta\rho_0$ (measured at 208 ms after the system

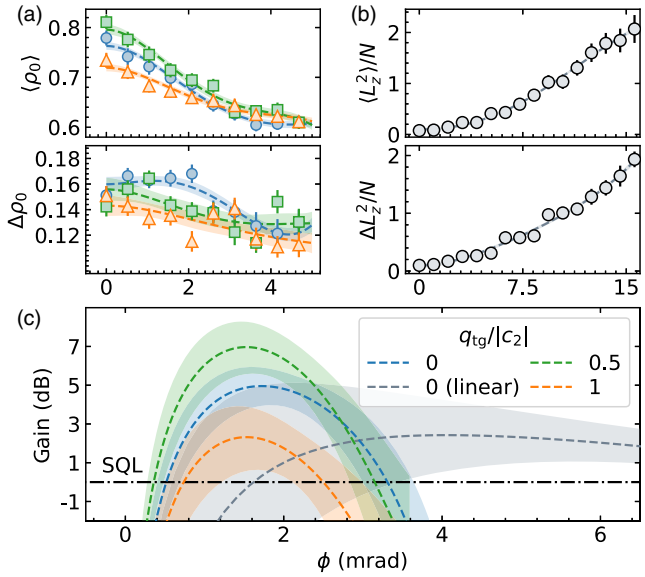

FIG. 4. (a) The averages (upper panel) and standard deviations (lower panel) of fractional spin population in $m_F = 0$, ρ_0, in the vicinity of phase $\phi = 0$. Each data point comes from 100 continuous experimental runs. (b) The averages (upper panel) and standard deviations (lower panel) of L_z^2, for the balanced spin-1 Dicke state without nonlinear readout operation. Each data point comes from 50 continuous experimental runs. For (a) and (b), dashed lines are fitted results. (c) The corresponding metrological gains over three-mode SQL (dot-dashed line). The dashed lines are obtained based on error propagation by using the fitted results in (a) and (b). The shaded region indicates the fitting uncertainty.

approaches the polar state [61]) for small encoded phases are shown in Fig. 4(a), with circles denoting experimental data and dashed lines the fitted curves. Phase sensitivity is calculated based on error propagation [Eq. (1)] by using the fitted data [dashed lines in Fig. 4(a)] and shown as colored data in Fig. 4(c). The optimal metrological gains for the three states with $q_{\mathrm{tg}}/|c_2| = 0$, 0.5, and 1 are found to be $\xi^2 = -20\log_{10}\{\Delta\phi/[1/(2\sqrt{N})]\} \simeq 4.96^{+0.96}_{-1.02}$, $6.97^{+1.30}_{-1.38}$, and $2.32^{+1.56}_{-1.73}$ dB, respectively, beyond three-mode SQL [41] of $1/(2\sqrt{N})$ with the atom number of the probe state $N \simeq 10\,400$. The probe state at $q_{\mathrm{tg}}/|c_2| = 0.5$ therefore outperforms due to the aforementioned tradeoff between theoretical metrological gain and atom loss.

For comparison, we also directly detect the phase-encoded spin-1 Dicke state [40,41], i.e., without nonlinear readout process. In this case, the phase can be extracted by measuring L_z^2, whose mean values and fluctuations for small phases are shown in Fig. 4(b). The inferred phase sensitivity [gray line in Fig. 4(c)] gives $\xi^2 \simeq 2.42^{+1.78}_{-1.84}$ dB beyond three-mode SQL. Hence, an enhancement of 2.5 dB is observed by introducing the nonlinear readout operation. The main constraint for further improvement on the achievable phase sensitivity is atom loss, which is an extra price to pay for performing the nonlinear readout operation.

In conclusion, we employ RL to guide nonlinear readout of entangled states based on manipulation of the spin-mixing dynamics of a spin-1 condensate by modulating the QZS. Remarkably, the optimal strategy is found to drive the system toward an unstable fixed point, where small perturbations grow in time. Our results suggest that there are various approaches to realizing nonlinear readout besides the paradigmatic one of returning to the starting point of entanglement generation. Our method can be generalized to other systems; see Ref. [40] for a discussion of applying our method to a spin-1/2 system.

We thank Dr. M. K. Tey and Dr. J. Yu for helpful discussions. This work is supported by the National Natural Science Foundation of China (NSFC) (Grants No. U1930201, No. 11654001, and No. 92265205), and by the National Key R&D Program of China (Grant No. 2018YFA0306504) and the Innovation Program for Quantum Science and Technology (2021ZD0302100).

J.-H. C. and F. C. contributed equally to this work.

*Present address: Laboratoire Kastler Brossel, Collège de France, CNRS, ENS-PSL University, Sorbonne Université, Paris, France.
†Corresponding author.
lingna.wu@hainanu.edu.cn
‡Corresponding author.
lyou@tsinghua.edu.cn

[1] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum-enhanced measurements: Beating the standard quantum limit, Science **306**, 1330 (2004).
[2] L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied, and P. Treutlein, Quantum metrology with nonclassical states of atomic ensembles, Rev. Mod. Phys. **90**, 035005 (2018).
[3] E. Davis, G. Bentsen, and M. Schleier-Smith, Approaching the Heisenberg Limit without Single-Particle Detection, Phys. Rev. Lett. **116**, 053601 (2016).
[4] F. Fröwis, P. Sekatski, and W. Dür, Detecting Large Quantum Fisher Information with Finite Measurement Precision, Phys. Rev. Lett. **116**, 090801 (2016).
[5] T. Macrì, A. Smerzi, and L. Pezzè, Loschmidt echo for quantum metrology, Phys. Rev. A **94**, 010102(R) (2016).
[6] O. Hosten, R. Krishnakumar, N. J. Engelsen, and M. A. Kasevich, Quantum phase magnification, Science **352**, 1552 (2016).
[7] S. C. Burd, R. Srinivas, J. J. Bollinger, A. C. Wilson, D. J. Wineland, D. Leibfried, D. H. Slichter, and D. T. C. Allcock, Quantum amplification of mechanical oscillator motion, Science **364**, 1163 (2019).
[8] K. A. Gilmore, M. Affolter, R. J. Lewis-Swan, D. Barberena, E. Jordan, A. M. Rey, and J. J. Bollinger, Quantum-enhanced sensing of displacements and electric fields with two-dimensional trapped-ion crystals, Science **373**, 673 (2021).
[9] S. Colombo, E. Pedrozo-Peñafiel, A. F. Adiyatullin, Z. Li, E. Mendez, C. Shu, and V. Vuletić, Time-reversal-based quantum metrology with many-body entangled states, Nat. Phys. **18**, 925 (2022).
[10] D. Linnemann, H. Strobel, W. Muessel, J. Schulz, R. J. Lewis-Swan, K. V. Kheruntsyan, and M. K. Oberthaler, Quantum-Enhanced Sensing Based on Time Reversal of Nonlinear Dynamics, Phys. Rev. Lett. **117**, 013001 (2016).
[11] F. Hudelist, J. Kong, C. Liu, J. Jing, Z. Y. Ou, and W. Zhang, Quantum metrology with parametric amplifier-based photon correlation interferometers, Nat. Commun. **5**, 3049 (2014).
[12] Y. Liu, J. Li, L. Cui, N. Huo, S. M. Assad, X. Li, and Z. Y. Ou, Loss-tolerant quantum dense metrology with SU(1,1) interferometer, Opt. Exp. **26**, 27705 (2018).
[13] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, MA, 2018).
[14] 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).
[15] M. Dalgaard, F. Motzoi, J. J. Sørensen, and J. Sherson, Global optimization of quantum dynamics with alphazero deep exploration, npj Quantum Inf. **6**, 6 (2020).
[16] M. M. Wauters, E. Panizon, G. B. Mbeng, and G. E. Santoro, Reinforcement-learning-assisted quantum optimization, Phys. Rev. Res. **2**, 033446 (2020).
[17] S.-F. Guo, F. Chen, Q. Liu, M. Xue, J.-J. Chen, J.-H. Cao, T.-W. Mao, M. K. Tey, and L. You, Faster State Preparation Across Quantum Phase Transition Assisted by Reinforcement Learning, Phys. Rev. Lett. **126**, 060401 (2021).
[18] J. Yao, L. Lin, and M. Bukov, Reinforcement Learning for Many-Body Ground-State Preparation Inspired by Counterdiabatic Driving, Phys. Rev. X **11**, 031070 (2021).

[19] T. Haug, W.-K. Mok, J.-B. You, W. Zhang, C. E. Png, and L.-C. Kwek, Classifying global state preparation via deep reinforcement learning, Mach. Learn. **2**, 01LT02 (2020).
[20] 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).
[21] R. Porotti, A. Essig, B. Huard, and F. Marquardt, Deep reinforcement learning for quantum state preparation with weak nonlinear measurements, Quantum **6**, 747 (2022).
[22] Z. An and D. L. Zhou, Deep reinforcement learning for quantum gate control, Europhys. Lett. **126**, 60002 (2019).
[23] M. Y. Niu, S. Boixo, V. N. Smelyanskiy, and H. Neven, Universal quantum control through deep reinforcement learning, npj Quantum Inf. **5**, 33 (2019).
[24] Z. T. Wang, Y. Ashida, and M. Ueda, Deep Reinforcement Learning Control of Quantum Cartpoles, Phys. Rev. Lett. **125**, 100401 (2020).
[25] J. Lin, Z. Y. Lai, and X. Li, Quantum adiabatic algorithm design using reinforcement learning, Phys. Rev. A **101**, 052327 (2020).
[26] V. Saggio, B. E. Asenbeck, A. Hamann, T. Strömberg, P. Schiansky, V. Dunjko, N. Friis, N. C. Harris, M. Hochberg, D. Englund, S. Wölk, H. J. Briegel, and P. Walther, Experimental quantum speed-up in reinforcement learning agents, Nature (London) **591**, 229 (2021).
[27] 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**, 82 (2019).
[28] J. Schuff, L. J. Fiderer, and D. Braun, Improving the dynamics of quantum sensors with reinforcement learning, New J. Phys. **22**, 035001 (2020).
[29] L.-Y. Chih and M. Holland, Reinforcement-learning-based matter-wave interferometer in a shaken optical lattice, Phys. Rev. Res. **3**, 033279 (2021).
[30] Y. Qiu, M. Zhuang, J. Huang, and C. Lee, Efficient and robust entanglement generation with deep reinforcement learning for quantum metrology, New J. Phys. **24**, 083011 (2022).
[31] T. Fösel, P. Tighineanu, T. Weiss, and F. Marquardt, Reinforcement Learning with Neural Networks for Quantum Feedback, Phys. Rev. X **8**, 031084 (2018).
[32] 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).
[33] P. Andreasson, J. Johansson, S. Liljestrand, and M. Granath, Quantum error correction for the toric code using deep reinforcement learning, Quantum **3**, 183 (2019).
[34] Y.-H. Zhang, P.-L. Zheng, Y. Zhang, and D.-L. Deng, Topological Quantum Compiling with Reinforcement Learning, Phys. Rev. Lett. **125**, 170501 (2020).
[35] L. Moro, M. G. A. Paris, M. Restelli, and E. Prati, Quantum compiling by deep reinforcement learning, Commun. Phys. **4**, 178 (2021).
[36] Z. He, L. Li, S. Zheng, Y. Li, and H. Situ, Variational quantum compiling with double q-learning, New J. Phys. **23**, 033002 (2021).
[37] T. M. Hoang, C. S. Gerving, B. J. Land, M. Anquez, C. D. Hamley, and M. S. Chapman, Dynamic Stabilization of a Quantum Many-Body Spin System, Phys. Rev. Lett. **111**, 090403 (2013).
[38] C. K. Law, H. Pu, and N. P. Bigelow, Quantum Spins Mixing in Spinor Bose-Einstein Condensates, Phys. Rev. Lett. **81**, 5257 (1998).
[39] F. Gerbier, A. Widera, S. Fölling, O. Mandel, and I. Bloch, Resonant control of spin dynamics in ultracold quantum gases by microwave dressing, Phys. Rev. A **73**, 041602(R) (2006).
[40] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.131.073201, which includes Refs. [17,41–54] for (i) a detailed description of reinforcement learning task in Sec. I and transfer learning in Sec. II; (ii) simulation of dissipative systems in Sec. III; (iii) experimental methods in Sec. IV; (iv) more data in Sec. V; (v) a brief introduction to spin-1 Dicke state in Sec. VI; (vi) a comparison of our protocol with time-reversal protocol in Sec. VII; and (vii) a study on nonlinear readout in a spin-1/2 system in Sec. VIII, which also includes a comparison of RL with traditional optimization methods.
[41] Y.-Q. Zou, L.-N. Wu, Q. Liu, X.-Y. Luo, S.-F. Guo, J.-H. Cao, M. K. Tey, and L. You, Beating the classical precision limit with spin-1 dicke states of more than 10,000 atoms, Proc. Natl. Acad. Sci. U.S.A. **115**, 6381 (2018).
[42] Q. Liu, L.-N. Wu, J.-H. Cao, T.-W. Mao, X.-W. Li, S.-F. Guo, M. K. Tey, and L. You, Nonlinear interferometry beyond classical limit enabled by cyclic dynamics, Nat. Phys. **18**, 167 (2022).
[43] M. J. Steel, M. K. Olsen, L. I. Plimak, P. D. Drummond, S. M. Tan, M. J. Collett, D. F. Walls, and R. Graham, Dynamical quantum noise in trapped Bose-Einstein condensates, Phys. Rev. A **58**, 4824 (1998).
[44] A. Sinatra, C. Lobo, and Y. Castin, The truncated Wigner method for Bose-condensed gases: Limits of validity and applications, J. Phys. B **35**, 3599 (2002).
[45] C. D. Hamley, C. S. Gerving, T. M. Hoang, E. M. Bookjans, and M. S. Chapman, Spin-nematic squeezed vacuum in a quantum gas, Nat. Phys. **8**, 305 (2012).
[46] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, Proximal policy optimization algorithms, arXiv:1707.06347.
[47] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, arXiv:1512.03385.
[48] J. Achiam, Spinning up documentation (2020).
[49] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, OpenAI Gym, arXiv:1606.01540.
[50] X.-Y. Luo, Y.-Q. Zou, L.-N. Wu, Q. Liu, M.-F. Han, M. K. Tey, and L. You, Deterministic entanglement generation from driving through quantum phase transitions, Science **355**, 620 (2017).
[51] T. M. Hoang, H. M. Bharath, M. J. Boguslawski, M. Anquez, B. A. Robbins, and M. S. Chapman, Adiabatic quenches and characterization of amplitude excitations in a continuous quantum phase transition, Proc. Natl. Acad. Sci. U.S.A. **113**, 9475 (2016).
[52] C. Klempt, O. Topic, G. Gebreyesus, M. Scherer, T. Henninger, P. Hyllus, W. Ertmer, L. Santos, and J. J.

Arlt, Parametric Amplification of Vacuum Fluctuations in a Spinor Condensate, Phys. Rev. Lett. **104**, 195303 (2010).

[53] M. J. Holland and K. Burnett, Interferometric Detection of Optical Phase Shifts at the Heisenberg Limit, Phys. Rev. Lett. **71**, 1355 (1993).

[54] J. Johansson, P. Nation, and F. Nori, Qutip 2: A python framework for the dynamics of open quantum systems, Comput. Phys. Commun. **184**, 1234 (2013).

[55] A. A. Norrie, R. J. Ballagh, and C. W. Gardiner, Quantum turbulence and correlations in Bose-Einstein condensate collisions, Phys. Rev. A **73**, 043617 (2006).

[56] B. Opanchuk, M. Egorov, S. Hoffmann, A. I. Sidorov, and P. D. Drummond, Quantum noise in three-dimensional BEC interferometry, Europhys. Lett. **97**, 50003 (2012).

[57] P. D. Drummond and B. Opanchuk, Truncated Wigner dynamics and conservation laws, Phys. Rev. A **96**, 043616 (2017).

[58] A. Johnson, S. S. Szigeti, M. Schemmer, and I. Bouchoule, Long-lived nonthermal states realized by atom losses in one-dimensional quasicondensates, Phys. Rev. A **96**, 013623 (2017).

[59] C. Gerving, T. Hoang, B. Land, M. Anquez, C. Hamley, and M. Chapman, Non-equilibrium dynamics of an unstable quantum pendulum explored in a spin-1 Bose–Einstein condensate, Nat. Commun. **3**, 1169 (2012).

[60] L. Zhao, J. Jiang, T. Tang, M. Webb, and Y. Liu, Dynamics in spinor condensates tuned by a microwave dressing field, Phys. Rev. A **89**, 023608 (2014).

[61] From the training results of small systems where noises are taken into account, we learn that after the system approaches the polar state, an additional period of spin-mixing dynamics is needed to submerge the effect of noise by signal amplification. The duration of this extra evolution is optimized in experiments to give the highest metrological gain.