

Critical states preparation in light-matter systems with deep reinforcement learning

Jia-Wen Yu,^{1,2} Yiming Yu ,^{1,2} Ke-Xiong Yan,^{1,2} Jun-Hao Lin,^{1,2} Jie Song,³ Ye-Hong Chen,^{1,2,4,*} and Yan Xia ,^{1,2,†}

¹Department of Physics, [Fuzhou University](#), Fuzhou 350116, China

²Fujian Key Laboratory of Quantum Information and Quantum Optics, [Fuzhou University](#), Fuzhou 350116, China

³Department of Physics, [Harbin Institute of Technology](#), Harbin 150001, China

⁴Quantum Information Physics Theory Research Team, Center for Quantum Computing, [RIKEN](#), Wako-shi, Saitama 351-0198, Japan



(Received 24 September 2025; revised 14 April 2026; accepted 24 June 2026; published 15 July 2026)

The fast and efficient preparation of quantum critical states is a challenging yet crucial task for various quantum technologies. This difficulty is most particularly for systems near a quantum phase transition, where the closure of the energy gap fundamentally limits the timescale of adiabatic processes and thus precludes rapid state preparation. We propose a framework using deep reinforcement learning (DRL) to rapidly prepare quantum critical states, with broad extendibility to light-matter interaction systems. Specifically, a DRL agent optimizes a set of time-dependent control Hamiltonians to drive the system from an initial noncritical state to a target critical state within a finite time and over experimentally accessible parameter ranges. As a concrete application, we focus on the quantum Rabi model. The DRL-optimized time-dependent control Hamiltonian yields a final state with high fidelity (>0.999) to the target critical state. The protocol can be readily extended to other quantum critical systems described by light-matter interaction models, such as quantum Dicke model. This investigation provides a powerful framework for preparing and manipulating quantum critical states.

DOI: [10.1103/34yc-wj82](https://doi.org/10.1103/34yc-wj82)

I. INTRODUCTION

Quantum systems near critical points exhibit a wealth of exotic physical phenomena, where small variations in physical parameters can lead to drastic changes in equilibrium-state properties [1–4]. A prominent class of quantum critical systems is described by light-matter interaction models, such as the quantum Rabi model (QRM) [5,6], the energy gap closes and levels converge to degeneracy as system parameters approach the critical point.

The ground states of these quantum critical systems are central to the study of quantum phase transitions, as they exhibit universal features of quantum criticality [7–9]. These unique properties make them a vital resource for quantum technologies [10–12], particularly in quantum metrology [13–15]. Nevertheless, preparing such critical ground states remains a formidable challenge. Conventional methods, such as adiabatic evolution [16,17], require an infinitely slow passage through the critical point to avoid excitations induced by the closing energy gap, rendering them impractical. Shortcuts to adiabaticity reduce the evolution time but typically rely on experimentally infeasible driving Hamiltonians [18]. Consequently, fast, robust, and experimentally feasible protocols for

generating quantum critical ground states are a crucial open challenge in the field [19,20].

The fast and high-fidelity preparation of a target quantum state can be formulated as an optimization problem [21–28]. Given a system Hamiltonian, the goal is to design time-dependent control protocols that drive an initial state to a target state within timescales shorter than the decoherence time [29–35]. Among existing approaches, gradient-based optimal control methods (e.g., gradient-ascent pulse engineering [36–38]) efficiently identify optimal parameters via gradient evaluation, but rely on explicit knowledge of the system dynamics. However, in strongly or ultrastrongly coupled quantum critical systems, the dynamics are often analytically intractable and gradient evaluation becomes unreliable, which limits the applicability of such methods. In these cases, control strategies based on final measurement outcomes (e.g., fidelity or observables) become essential [38].

Recently, deep reinforcement learning (DRL) has emerged as a powerful tool for solving complex optimization problems in quantum physics, particularly those involving measurement feedback [39–47]. DRL agents learn effective control strategies through trial-and-error interactions with simulated environments, without requiring prior knowledge of the system dynamics [48]. This approach has achieved remarkable success across various quantum tasks, including many-body ground-state preparation [49], high-fidelity quantum gate design [50], and quantum error-correction decoding [51]. Owing to its ability to explore high-dimensional parameter spaces, DRL can identify efficient and nonintuitive control protocols beyond the reach of traditional optimization methods.

In this article, we propose a broadly applicable DRL framework for the rapid preparation of quantum critical states.

*Contact author: yehong.chen@fzu.edu.cn

†Contact author: xia-208@163.com

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](#) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

A DRL agent optimizes time-dependent control Hamiltonians to drive a system from a noncritical ground state to a target critical state within a short time and with high fidelity. As a concrete example, we consider the quantum Rabi model, where the optimized control yields a final state with fidelity exceeding 0.999 using only a single control field. The robustness of the protocol against systematic errors and environmental dissipation is also examined. Numerical simulations show that the fidelity loss remains below 5% on average for $|\beta| \leq 5\%$, while the effect of dissipation stays below 1%, demonstrating strong robustness. This work establishes DRL as a powerful tool for engineering quantum critical states and can be readily extended to other light-matter systems, such as the quantum Dicke model.

II. PHYSICAL MODEL AND METHOD

We consider a class of quantum systems with Hamiltonian

$$H[g(t)] = H_0 + H_{\text{int}}[g(t)], \quad (1)$$

where H_0 is the free Hamiltonian and $H_{\text{int}}[g(t)]$ denotes the interaction term controlled by a tunable parameter $g(t)$ [19]. As $g(t)$ approaches the critical value g_c , strong entanglement develops, making controlled state preparation challenging. Given $g(0)$ and $g(T) = g_c$ over a time interval $[0, T]$, the goal is to design a time-dependent protocol that drives the system from the initial state $|\Phi[g(0)]\rangle$ to the eigenstate $|\Phi[g(T)]\rangle$ of $H[g(T)]$.

To address this, we introduce a sequence of control Hamiltonians, such that the total Hamiltonian reads

$$H_{\text{tot}}(t) = H[g(t)] + \sum_{i=1}^n H_i^c(t), \quad (2)$$

with

$$H_i^c(t) = \Lambda_i \cos(\omega_d t + \phi_i) H_i^c. \quad (3)$$

Here, ω_d is the driving frequency, while Λ_i and ϕ_i denote the amplitudes and phases of the control fields, respectively.

In the first step, the full set of control Hamiltonians is employed to achieve control over all degrees of freedom of the system Hamiltonian, driving the system from $|\Phi(0)\rangle = |\Phi[g(0)]\rangle$ to $|\Phi(T)\rangle = |\Phi[g(T)]\rangle$, while $H[g(t)]$ is tuned accordingly. However, implementing all controls simultaneously is experimentally challenging. To overcome this limitation, we adopt a second step: each control field is applied individually to generate test trajectories $|\Phi_i^{\text{test}}(t)\rangle$, and a trajectory similarity is defined as

$$\Delta_i = \int_0^T \left| \langle \Phi_i^{\text{test}}(t) | \Phi(t) \rangle \right|^2 dt, \quad (4)$$

which quantifies the contribution of the i th control field. Based on Δ_i , the dominant control fields are reoptimized to reduce resources while preserving the target state transfer. In general, determining a sequence of control Hamiltonians $H_i^c(t)$ that drives a complex quantum system to a desired evolution is challenging. The system dynamics are rarely analytically solvable as the $g(t)$ increases. Moreover, in the first step full control analysis, one must explore a high-dimensional parameter space, which leads to a substantial optimization overhead. We therefore employ DRL to design control protocols that

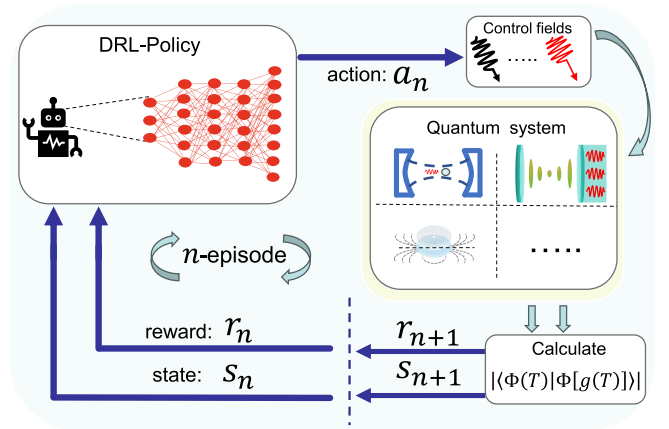


FIG. 1. Schematic of the DRL approach for quantum critical states preparation. At each episode n , the agent observes state s_n , receives a fidelity-based reward r_{n+1} , and selects an action a_n via policy $\pi(A|S)$ to update the control fields, driving the system to s_{n+1} . This iterative process optimizes the control toward the target critical state.

circumvent these analytical limitations. The central idea is that the DRL model (workflow in Fig. 1) emulates the behavior of agents interacting with an environment, learning through rewards to adapt its decisions based on observations [50,52]. To ensure experimental feasibility, we design a physically constrained reward function that incorporates fidelity incentives and penalties on control parameters.

In the learning process, a fixed total evolution time T is discretized into a finite number of steps $k = 1, 2, \dots, K$. At each episode n , the agent observes the current state $s_n \in S$ and chooses an action $a_n \in A$ according to the policy $\pi(A|S)$. The action spaces are defined by the control parameters of the control Hamiltonians $H_i^c(t)$, such as driving frequency, amplitudes, driving time, and phases, and the state s_n is described by the fidelity $F = |\langle \Phi(T) | \Phi[g(T)] \rangle|$. Note that we adopt this form of the fidelity to mitigate the barren plateau problem that may arise during the initial training stage due to low-fidelity values. Other distance measures, such as the Jensen-Shannon divergence and the trace distance, can also be used to characterize the state s_n . However, near the critical point, these measures require significantly larger computational resources. Therefore, the fidelity is chosen here as the training objective. To find the optimal control Hamiltonians $H_i^c(t)$, we define the reward function for the DRL agent as

$$\begin{aligned} R &= r_{\text{fid}} - \zeta_{\text{amp}} P_{\text{amp}} - \zeta_{\text{freq}} P_{\text{freq}} - \zeta_{\text{smooth}} P_{\text{smooth}}, \\ r_{\text{fid}} &= [a + b(F^4)][-\log_{10}(1 - F)], \\ P_{\text{amp}} &= \sum_{i=1}^n \left[\frac{1}{n} \sum_{k=1}^K \exp(\Lambda_{k,i}) \right], \\ P_{\text{freq}} &= \exp(\omega_d), \\ P_{\text{smooth}} &= \sum_{i=1}^n \left[\sum_{k=1}^K (\Lambda_{k+1,i} - \Lambda_{k,i})^2 + 2 \sum_{k=1}^K (\Lambda_{k+2,i} \right. \\ &\quad \left. - 2\Lambda_{k+1,i} + \Lambda_{k,i})^2 \right]. \end{aligned} \quad (5)$$

Here, the term r_{fid} denotes the fidelity reward and constitutes the primary driving component of the reward function R , and a and b are positive hyperparameters (chosen to balance training stability and reward sensitivity). The penalty term P_{amp} suppresses excessive driving amplitudes, P_{freq} biases the agent toward lower driving frequencies, and P_{smooth} reduces abrupt changes in the driving amplitude between adjacent time steps. The coefficients ζ_{amp} , ζ_{freq} , and ζ_{smooth} represent the weights of the corresponding penalty terms. The structure of the reward function enforces strong physical constraints during training. In Appendix C, we compare the above DRL protocol with other methods.

III. APPLICATION-QUANTUM RABI MODEL

To validate our DRL framework, we focus on the quantum Rabi model—a paradigmatic light-matter interaction system exhibiting a superradiant quantum phase transition [5,6]. Its Hamiltonian is

$$H_{\text{Rabi}}[g(t)] = \omega a^\dagger a + \frac{\Omega}{2} \sigma_z + \frac{g(t)\sqrt{\omega\Omega}}{2} (a + a^\dagger) \sigma_x, \quad (6)$$

where $\sigma_{x,z}$ are Pauli matrices for the two-level atom, ω is the cavity field frequency, Ω is the energy splitting of a two-level atom, and $a^\dagger(a)$ is the creation (annihilation) operator. We define the dimensionless parameter $g(t) = 2\lambda(t)/\sqrt{\omega\Omega}$, where $\lambda(t)$ denotes the time-dependent cavity-atom coupling strength. The function $g(t)$ is parametrized as $g(t) = \sum_{j=0}^2 a_j t^j$ [a_j is a dimensionless parameter that determines the shape of the function $g(t)$] and satisfies the boundary conditions $g(0) = g_0$ and $g(T) = g_c$. At the critical point $g_c = 1$ (i.e., $\lambda = \sqrt{\omega\Omega}/2$), the energy gap closes, and the QRM undergoes a superradiant phase transition [5].

Here, we consider five commonly used control Hamiltonians in cavity quantum electrodynamics systems, which provide full control over the degrees of freedom of the QRM [53]:

$$\begin{aligned} H_1^c(t) &= \Lambda_1 \cos(\omega_d t + \phi_1) (a + a^\dagger), \\ H_2^c(t) &= \Lambda_2 \cos(\omega_d t + \phi_2) (a + a^\dagger)^2, \\ H_3^c(t) &= \Lambda_3 \cos(\omega_d t + \phi_3) a^\dagger a, \\ H_4^c(t) &= \Lambda_4 \cos(\omega_d t + \phi_4) \sigma_z, \\ H_5^c(t) &= \Lambda_5 \cos(\omega_d t + \phi_5) \sigma_x. \end{aligned} \quad (7)$$

In each episode, the amplitudes are discretized as sequences $\Lambda_i = \{\Lambda_i(t_1), \dots, \Lambda_i(t_K)\}$. To ensure experimental feasibility, we impose $\Lambda_i(t_1) = \Lambda_i(t_K) = 0$, such that $H_i^c(t)$ vanishes at the initial and final times. At each DRL episode, the evolution operator can be described as $U(T, 0) = \prod_{k=K}^1 U(t_k, t_{k-1})$, where the single-step evolution operator takes the form

$$U(t_k, t_{k-1}) = \mathcal{T} \exp \left[-i \int_{t_{k-1}}^{t_k} \left(H_{\text{Rabi}}[g(\tau)] + \sum_{i=1}^5 H_i^c(\tau) \right) d\tau \right]. \quad (8)$$

Here, $\mathcal{T}$ denotes the time-ordering operator. The final state of the quantum system is given by $|\Phi(T)\rangle = U(T, 0)|\Phi[0]\rangle$. Through repeated training, the agent adjusts the system evolution operator to maximize the fidelity $F(|\Phi(T)\rangle, |\Phi[g(T)]\rangle)$.

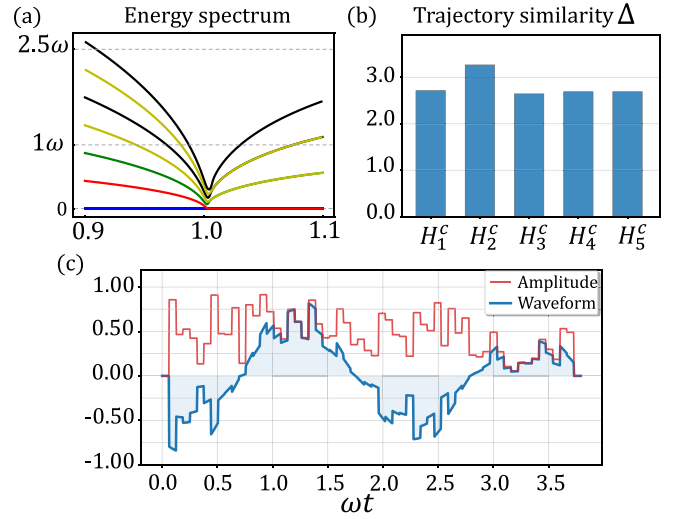


FIG. 2. (a) Energy spectrum ($E_n - E_0$) of the Rabi Hamiltonian $H_{\text{Rabi}}[g(t)]$ across the critical point, calculated with a Hilbert-space truncation dimension of $N = 5000$. (b) Analysis of the trajectory similarity Δ_i , indicating that the control field $(a + a^\dagger)^2$ plays a dominant role in the system evolution. (c) Waveform and amplitude of the control pulses identified by the DRL agent. The control Hamiltonian is given by $H_2^c(t) = f_2(t)(a + a^\dagger)^2$, with the driving field $f_2(t) = A_2 \cos(\omega_d t + \phi_2)$. The parameters used in the simulation are $\Omega = 10^4 \omega$, $g(0) = g_0 = 0.01$, $g(T) = g_c = 1$, and $K = 60$.

The trajectory similarity defined in Eq. (4) is then analyzed to identify the control field with the largest contribution, which is subsequently selected for retraining in order to maximize the fidelity while minimizing resource consumption. Systemic state and reward are obtained by simulating the Schrödinger equation using the QuTiP package [54,55].

In the first step, the DRL-optimized control sequence with the full set of control fields achieves a fidelity of $F(|\Phi(T)\rangle, |\Phi[g(T)]\rangle) = 0.9990$ at $T = 4.1531/\omega$. The trajectory similarity Δ_2 attains the largest value, indicating that the control field $(a + a^\dagger)^2$ plays a dominant role in the evolution. Based on this result, the second step employs only the control field $(a + a^\dagger)^2$ for the evolutionary training. The resulting optimized control sequence achieves the same fidelity, $F(|\Phi(T)\rangle, |\Phi[g(T)]\rangle) = 0.9991$, at a shorter duration of $T = 3.7913/\omega$.

The corresponding optimal parameters $\{\omega_d, \phi_2, \Lambda_2\}$ are shown in Fig. 2. In Appendix B, control sequences for different system parameters Ω/ω and the smoothing of the control waveforms are presented. The populations of the time-evolved state $|\Phi(T)\rangle$ in terms of the instantaneous eigenstates of $H_{\text{Rabi}}[g(t)]$ are presented in Fig. 3. The results show that DRL agent identifies a shortcut distinct from the adiabatic path, enabling the system to evolve to the critical ground state $|\Phi[g(T)]\rangle$ within a markedly short time. This evolution is achieved using a single driving field $(a + a^\dagger)^2$, with the amplitude remaining within experimentally accessible limits and the duration of each single-stage operation t_k well within the range of conventional quantum control times. Our proposal is feasible in circuit QED systems, where a transmission line resonator interacts with a superconducting qubit described by the Jaynes-Cummings model [56,57]. A two-photon drive

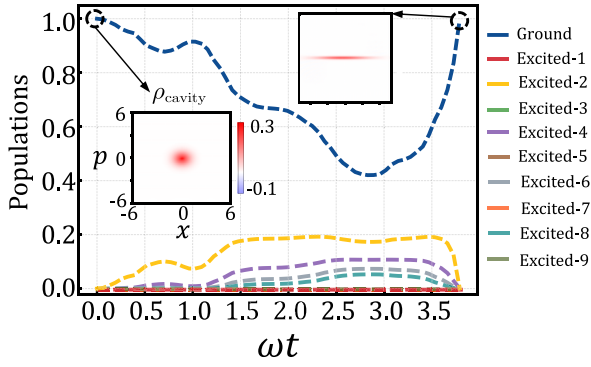


FIG. 3. Time evolution of the populations. At each time t , the evolving state $|\Phi(t)\rangle$ is projected onto the first ten eigenstates $|E_n(t)\rangle$ of the instantaneous Hamiltonian $H_{\text{Rabi}}[g(t)]$, yielding the population distribution $P_n(t) = |\langle E_n(t)|\Phi(t)\rangle|^2$. We also show the Wigner functions of the initial and final states of the cavity subspace, obtained by tracing out the qubit degrees of freedom.

(Josephson parametric amplification process) can be realized by time-dependent modulation of the magnetic flux through a superconducting quantum interference device [58,59]. In Appendix A, a potential experimental implementation scheme based on superconducting circuits is presented.

IV. ANALYSIS OF NOISES

Without loss of generality, the influence of control errors and environmental dissipation on the DRL-control sequence is considered separately.

For systematic errors, the actual control Hamiltonian is written as $H_2^{\text{ac}}(t) = H_2^s(t) + H_2^e$, where H_2^e denotes the control error Hamiltonian. In practice, imperfections may occur in one or more control parameters $\{\omega_d, \phi_2, \Lambda_2\}$. To assess the robustness of the protocol, we examine the influence of deviation in each parameter independently while keeping the others fixed at their optimal values. A perturbed parameter is modeled as $\chi^a = (1 + \beta)\chi$, where $\chi \in \{\omega_d, \phi_2, \Lambda_2\}$ and $\beta \in [-5\%, 5\%]$ characterizes the relative error strength. In Fig. 4(a), the fidelity loss induced by errors is shown. The results show that deviations in the phase ϕ_2 and driving amplitude Λ_2 have little influence on the final evolved state, while the system is more sensitive to deviations in the control frequency ω_d . The reason is that variations in the control frequency ω_d can lead to significant distortions of the entire control waveform, thereby altering the evolution path compared with the error-free case. Nevertheless, our control protocol can still maintain a high preparation fidelity within the typical experimental error range of $[-5\%, 5\%]$ [58,59]. We further account for environmental dissipation during the system evolution. The dynamics is described by a Lindblad master equation [13]

$$\frac{d\rho(t)}{dt} = -i[H_{\text{tot}}(t), \rho(t)] + \sum_{i=1}^3 \left(L_i \rho(t) L_i^\dagger - \frac{1}{2} \{L_i^\dagger L_i, \rho(t)\} \right), \quad (9)$$

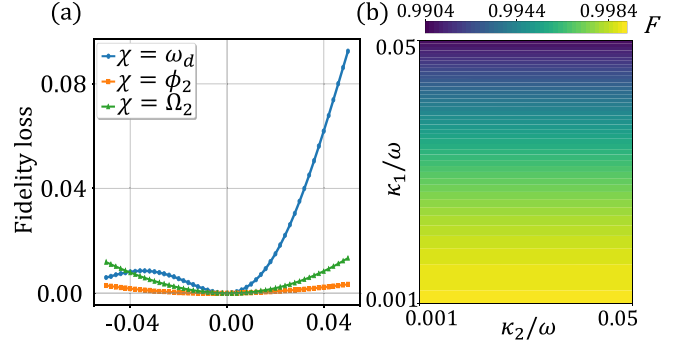


FIG. 4. (a) Analysis of control errors. Different broken lines correspond to errors in different control parameters χ . The results show that the average fidelity loss of the final evolved state due to systematic errors remains below 5%. (b) Analysis of dissipative effects. The dephasing rate of the qubit is fixed at $\kappa_3/\omega = 0.01$. The results indicate that qubit relaxation has a negligible impact on the final evolved state, and even at relatively large dissipation rates, the fidelity between the final state and the target state remains above 0.99.

where $H_{\text{tot}}(t) = H_{\text{Rabi}}[g(t)] + H_2^e(t)$ and $\rho(t)$ is the system density matrix. The Lindblad operators $L_1 = \kappa_1 a$, $L_2 = \kappa_2 \sigma_-$, and $L_3 = \kappa_3 \sigma_z$ correspond to photon loss, qubit relaxation, and qubit dephasing, respectively, with κ_1 , κ_2 , and κ_3 the associated decay rates. The effects of environmental dissipation are summarized in Fig. 4(b). The results indicate that the protocol maintains high fidelity (>0.99) even in the presence of dissipation when $\kappa_1, \kappa_2 \simeq 0.05\omega$. In addition, we apply the DRL framework to train the control sequence directly in an open quantum system. In this case, the system state and reward are obtained by numerically simulating the Lindblad master equation. The highest fidelity between the evolved state and the target state reaches $F = 0.9965$, with $T = 2.6047/\omega$ (the parameters used in the simulation are $\Omega = 100\omega$, $g_0 = 0.01$, $g_c = 1$, $K = 60$, and decay rates $\kappa_1/\omega = \kappa_2/\omega = \kappa_3/\omega = 0.05$). These results demonstrate that the physically constrained reward function and the DRL environment designed in our protocol remain effective even in the presence of dissipation. These features demonstrate the robustness of the designed control scheme and its suitability for realistic experimental implementations.

V. DISCUSSIONS AND CONCLUSIONS

We propose a DRL-based framework for the fast and robust preparation of quantum critical states, where a DRL agent optimizes time-dependent control Hamiltonians $H_i^c(t)$ to drive a system from an initial noncritical ground state $|\Phi[g(0)]\rangle$ to the target critical state $|\Phi[g(T)]\rangle$ ($g(T) = g_c$) within a finite time. As an example, we consider the QRM, which exhibits a second-order mean-field quantum phase transition at $g_c = 1$. The optimized protocol achieves a fidelity of $F(|\Phi(T)\rangle, |\Phi(g_c)\rangle) = 0.9991$ at $T = 3.7913/\omega$. To assess practical feasibility, we analyze the robustness of the control sequence against systematic errors and environmental dissipation. Numerical results show that fidelity loss remains below 5% on average for $|\beta| \leq 5\%$, while dissipation effects stay below 1%, demonstrating strong

robustness. In Appendix D, quantum fisher information (QFI) analysis further confirms the criticality of the prepared state: I_g sharply increases toward divergence at the end of the evolution, reflecting the quantum critical behavior.

Importantly, the proposed framework is readily extendible to other light-matter quantum critical systems, such as the quantum Dicke model. For the quantum Dicke model, the protocol yields a fidelity $F(|\Phi(T)\rangle, |\Phi[g'(T)]\rangle) = 0.9953$ at $T = 3.0279/\omega$. Additional details are provided in Appendix E, highlighting the broad applicability of the framework beyond the QRM. These results demonstrate the effectiveness of DRL for preparing and manipulating quantum critical states and provide a foundation for hybrid or hardware-efficient reward schemes enabling partially model-free experimental training [46].

ACKNOWLEDGMENTS

Y.-H.C. was supported by the National Natural Science Foundation of China under Grants No. 12304390 and No. 12574386, the National Postdoctoral Overseas Talent Recruitment Program of China, the Fujian 100 Talents Program, and the Fujian Minjiang Scholar Program. Y.X. was supported by the National Natural Science Foundation of China under Grant No. 62471143 and the Key Program of National Natural Science Foundation of Fujian Province under Grant No. 2024J02008.

DATA AVAILABILITY

The data used for obtaining the presented numerical results as well as for generating the plots are available on request.

APPENDIX A: PROPOSAL OF AN EXPERIMENTAL REALIZATION

We introduce an architecture composed of a superconducting qubit strongly coupled to a harmonic oscillator. Under the rotating-wave approximation (RWA), the Hamiltonian can be effectively approximated as ($\hbar = 1$)

$$H = \omega_0 a^\dagger a + \frac{\omega_q}{2} \sigma_z + \lambda(\sigma_+ a + \sigma_- a^\dagger), \quad (\text{A1})$$

where $a^\dagger$ (a) denotes the creation (annihilation) operator of the harmonic oscillator with frequency ω_0 and ω_q denotes the transition frequency of the qubit. The Pauli operators are defined as $\sigma_z = |e\rangle\langle e| - |g\rangle\langle g|$, where $|g\rangle$ and $|e\rangle$ are the ground and excited eigenstates of the qubit, respectively. The parameter λ represents the coupling strength between the qubit and the cavity.

Now consider that the qubit is orthogonally driven by two classical fields. The Hamiltonian of the driven system then reads [60]

$$\begin{aligned} H = & \omega_0 a^\dagger a + \frac{\omega_q}{2} \sigma_z + \lambda(\sigma_+ a + \sigma_- a^\dagger) \\ & - \varepsilon_1 [\cos(\omega_1 t) \sigma_x + \sin(\omega_1 t) \sigma_y] \\ & - \varepsilon_2 [\cos(\omega_2 t) \sigma_x + \sin(\omega_2 t) \sigma_y], \end{aligned} \quad (\text{A2})$$

where $\varepsilon_{1,2}$ and $\omega_{1,2}$ denote the amplitudes and frequencies of the two classical driving fields, respectively. The Pauli

operators are defined as $\sigma_x = \sigma_+ + \sigma_- = |e\rangle\langle g| + |g\rangle\langle e|$ and $\sigma_y = i(\sigma_- - \sigma_+) = i(|g\rangle\langle e| - |e\rangle\langle g|)$. Next, by applying the unitary transformation $U = \exp[i\omega_1(a^\dagger a + \frac{\sigma_z}{2})t]$ to Eq. (A2) and invoking the rotating-wave approximation once again, we obtain

$$\begin{aligned} H' = & (\omega_0 - \omega_1) a^\dagger a + \frac{\omega_q - \omega_1}{2} \sigma_z + \lambda(\sigma_+ a + \sigma_- a^\dagger) \\ & - \varepsilon_1(\sigma_+ + \sigma_-) \\ & - \varepsilon_2(e^{i(\omega_2 - \omega_1)t} \sigma_- + e^{-i(\omega_2 - \omega_1)t} \sigma_+). \end{aligned} \quad (\text{A3})$$

After the above operations, we obtain the original first driving term in a time-independent form, $H_1^c = -\varepsilon_1(\sigma_+ + \sigma_-)$. We then transform the Hamiltonian in Eq. (A3) into the interaction picture with respect to H_1^c

$$\begin{aligned} H_I(t) = & e^{iH_1^c t} (H' - H_1^c) e^{-iH_1^c t} \\ = & (\omega_0 - \omega_1) a^\dagger a - \frac{\omega_q - \omega_1}{2} (e^{-2i\varepsilon_1 t} |+\rangle\langle -| + \text{H.c.}) \\ & + \frac{\lambda}{2} [|+\rangle\langle +| - |-\rangle\langle -| + e^{-2i\varepsilon_1 t} |+\rangle\langle -| \\ & - e^{2i\varepsilon_1 t} |-\rangle\langle +|] a + \text{H.c.}] \\ & - \frac{\varepsilon_2}{2} [|+\rangle\langle +| - |-\rangle\langle -| - e^{-2i\varepsilon_1 t} |+\rangle\langle -| \\ & + e^{2i\varepsilon_1 t} |-\rangle\langle +|] e^{i(\omega_2 - \omega_1)t} + \text{H.c.}], \end{aligned} \quad (\text{A4})$$

where $|\pm\rangle = (|g\rangle \pm |e\rangle)/\sqrt{2}$. By tuning the parameters of the external driving fields to satisfy $\omega_1 - \omega_2 = 2\varepsilon_1$ and choosing a relatively strong amplitude for the first driving field ε_1 , a time-independent effective Hamiltonian can be obtained under a second RWA

$$H_{\text{eff}} = (\omega_0 - \omega_1) a^\dagger a + \frac{\varepsilon_2}{2} \sigma_z + \frac{\lambda}{2} (a^\dagger + a) \sigma_x. \quad (\text{A5})$$

Although the coupling strength λ is fixed in Eq. (A5), the effective parameters can still be flexibly engineered by tuning the amplitudes and frequencies of the driving fields. This procedure yields an effective quantum Rabi Hamiltonian, in which the effective qubit frequency and the effective cavity frequency can be tuned independently via the driving frequencies and amplitudes.

By introducing parametric pumping into the cavity and operating in the effective regime, a two-photon drive is generated [61]. The resulting effective system Hamiltonian, including the control term, is given by

$$\begin{aligned} H'_{\text{eff}} = & (\omega_0 - \omega_1) a^\dagger a + \frac{\varepsilon_2}{2} \sigma_z + \frac{\lambda}{2} (a^\dagger + a) \sigma_x \\ & + \varepsilon_3(t) \cos(\omega_3 t + \phi) (a + a^\dagger)^2, \end{aligned} \quad (\text{A6})$$

where $\varepsilon_3(t)$ denotes the time-dependent drive amplitude, ω_3 denotes the drive frequency, and ϕ denotes the phase of the parametric pump.

In a superconducting circuit, the Hamiltonian describing the above process is given by [62,63]

$$\begin{aligned} H = & 4E_c(n - n_g)^2 - E_J \cos(\varphi) + \lambda n N + 4E_C N^2 \\ & - N_0 E_J [\Phi(t)] \cos\left(\frac{\phi}{N_0}\right), \end{aligned} \quad (\text{A7})$$

where $E_j = E_{j,\max} |\cos(\pi \Phi / \Phi_0)|$ denotes the tunable Josephson energy, E_c is the charging energy, and λ represents the coupling constant between the transmon qubit and the Josephson-junction array. The effective offset charge of the transmon qubit is denoted by n_g , while n and φ represent the number of Cooper pairs and the superconducting phase difference, respectively. For the Josephson-junction array, E_C and E_J denote the charging energy and the Josephson energy of each junction. The operators N and ϕ represent the number of Cooper pairs and the overall superconducting phase across the array, respectively. The parameter N_0 denotes the number of superconducting quantum interference devices in the array and $\Phi(t)$ is the external magnetic flux used to modulate the parametric oscillator.

The transmon circuit can be treated as a Duffing nonlinear oscillator in the regime $E_J \gg E_C$. When only the lowest two energy levels are retained, the Hamiltonian of the transmon qubit can be written as $H_q = \frac{\omega_q}{2} \sigma_z$. By performing a Taylor expansion of $\cos(\phi/N_0)$ and truncating the expansion at fourth order, with $E_J[\Phi(t)] = E_j + \delta E_j \cos(\omega_p t)$, one obtains

$$\begin{aligned} N_0 E_J[\Phi(t)] \cos\left(\frac{\phi}{N_0}\right) &= N_0 E_j \left(1 - \frac{\phi^2}{2N_0^2} + \frac{\phi^4}{24N_0^4}\right) \\ &+ N_0 \delta E_j \left(1 - \frac{\phi^2}{N_0^2}\right) \cos(\omega_p t). \end{aligned} \quad (\text{A8})$$

Consider the standard quantization procedure for superconducting circuits

$$N = -in_0(a - a^\dagger), \quad \phi = \phi_0(a + a^\dagger), \quad (\text{A9})$$

and neglect the Kerr effect. Under these conditions, the Hamiltonian in Eq. (A6) can be obtained. By supplying the control waveform through an Arbitrary Waveform Generator, the evolution process described in the main text can be implemented.

Therefore, the system required by the protocol presented in the main text, together with the corresponding control Hamiltonian, can be realized.

APPENDIX B: CONTROL FOR DIFFERENT VALUES OF Ω/ω AND WAVEFORM SMOOTHING

As discussed in the main text, when the parameter $g(t)$ approaches the critical point g_c , the energy-level spacing of the quantum Rabi model decreases. Moreover, the ratio between the qubit energy splitting and the cavity frequency, Ω/ω , further influences the energy-level spacing at the critical point. As shown in Fig. 5, increasing Ω/ω leads to an intermittent linear reduction of the energy-level spacing at the critical point. This behavior renders the critical ground state more sensitive to external perturbations and further increases the difficulty of preparing the critical ground state through adiabatic and quasiadiabatic processes. In this case, the time required for adiabatic evolution increases sharply [64].

Next, the procedure for obtaining control sequences that realize the evolution toward the critical ground state within the DRL framework is discussed for different values of Ω/ω . The resulting control sequences are shown in Fig. 6. These results demonstrate that the physical constraint-based reward function adopted in this protocol yields control sequences

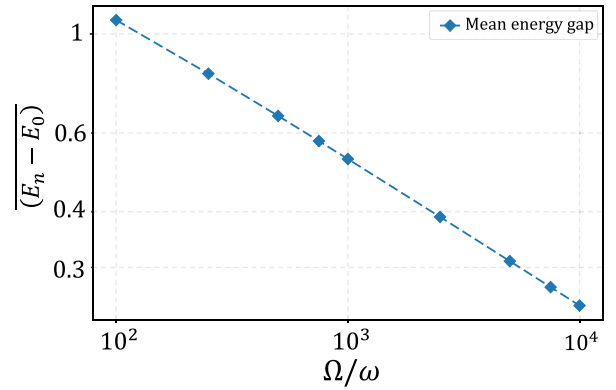


FIG. 5. Energy gap behavior at the critical point g_c for different values of Ω/ω . The energy-level gap between the ground state and the first six ($n = 1, 2, \dots, 7$) excited states is calculated and averaged. The results show that increasing Ω/ω leads to a progressive reduction of the energy gap, which significantly increases the difficulty of accurately preparing the critical ground state. The calculations are performed using a Hilbert-space truncation dimension of $N = 5000$.

capable of preparing the critical ground state with high fidelity across different system parameters Ω/ω and distinct critical points. Although the control sequences shown in Fig. 6 achieve high-fidelity evolution toward the target critical ground states, the presence of abrupt variations in these sequences may lead to practical difficulties in experimental control. To make the control waveforms smoother, Gaussian smoothing is applied to the control amplitudes. This procedure is described as follows [65]:

$$\begin{aligned} G[m] &= \frac{\exp\left(-\frac{m^2}{2\sigma^2}\right)}{\sum_{j=-M}^M \exp\left(-\frac{j^2}{2\sigma^2}\right)}, \\ \tilde{\Lambda}[k] &= \sum_{m=-M}^M G[m] \Lambda[k - m], \end{aligned} \quad (\text{B1})$$

where $M = \lfloor N/2 \rfloor$ is the half-width of the smoothing window of size N and $m \in \mathbb{Z} \cap [-M, M]$ denotes the relative index. Here, σ is the standard deviation of the Gaussian function. Gaussian smoothing acts as a filtering procedure that reduces abrupt variations in the waveform by convolving each discrete control amplitude with a Gaussian function. Figure 7 compares the original control amplitudes and waveforms with the Gaussian-smoothed amplitudes and waveforms, and examines the influence of smoothing on the fidelity of the final evolved state. The results show that significantly smoother waveforms can be obtained while only a small reduction in fidelity is incurred.

APPENDIX C: COMPARISON WITH OTHER METHODS

In this Appendix, we compare our protocol for generating the critical ground state of the quantum Rabi model with methods based on adiabatic evolution and shortcuts to adiabaticity (STA).

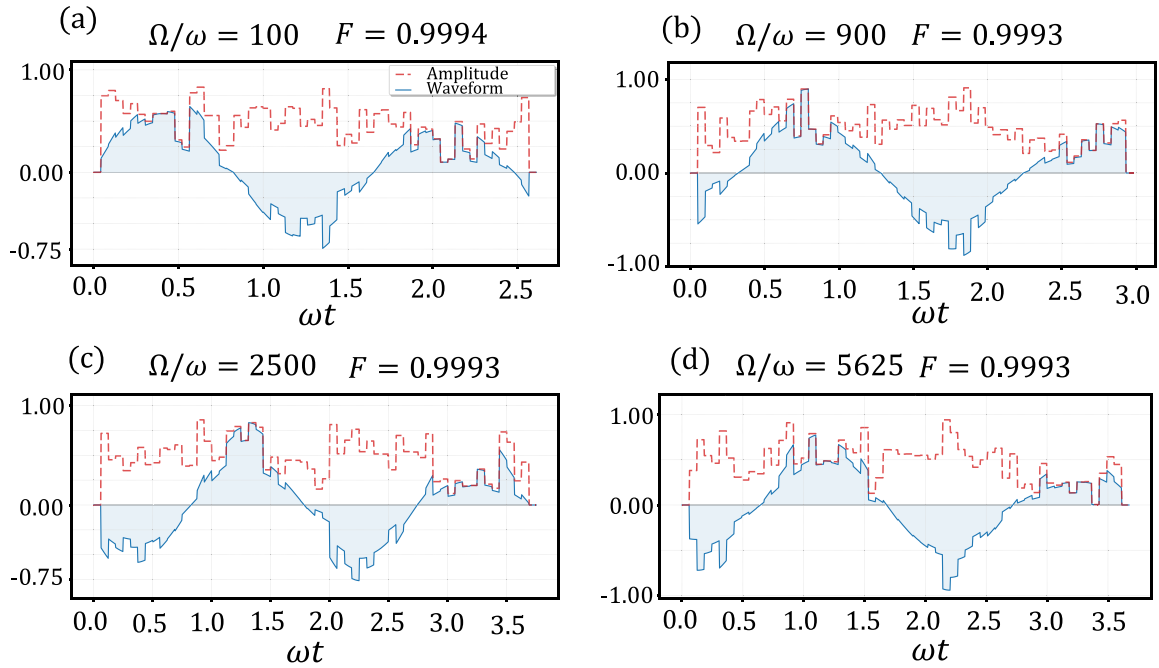


FIG. 6. Panels (a)–(d) show the control sequences obtained using the DRL framework for different values of Ω/ω , which realize the evolution toward the critical ground state with high fidelity.

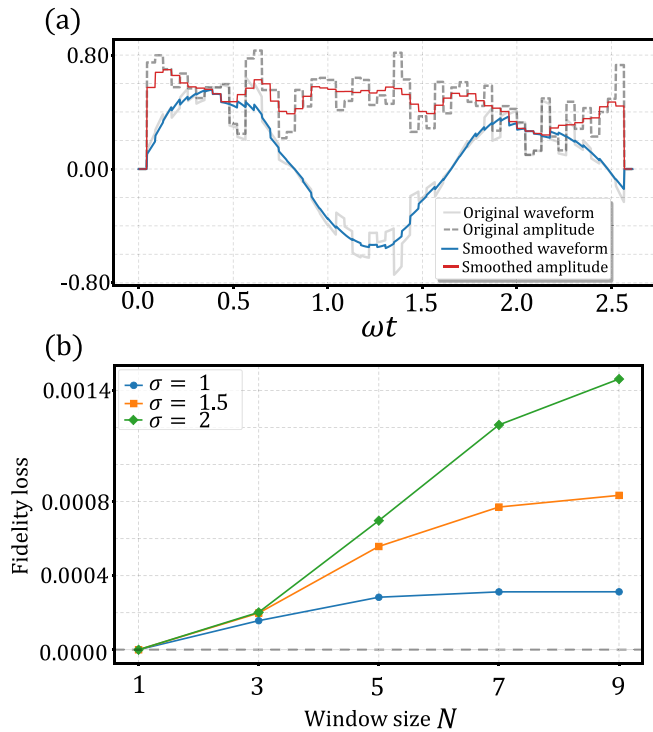


FIG. 7. (a) Effect of Gaussian smoothing on the control waveform. The standard deviation of the Gaussian function is set to $\sigma = 2$, and the smoothing window is chosen as $N = 5$. The control sequence corresponding to $\Omega/\omega = 100$ is selected as a representative example. (b) Fidelity loss induced by Gaussian smoothing in the final evolved state. Even when a large smoothing window of $N = 9$, corresponding to nearly one-sixth of the total number of amplitude segments $K = 60$, is employed with $\sigma = 2$, the resulting fidelity reduction remains very small and does not exceed 0.2%.

1. Adiabatic method

Considering a time-dependent quantum system with Hamiltonian $H_0(t)$, where eigenstates are denoted as $|\phi_n(t)\rangle$

$$H_0(t)|\phi_n(t)\rangle = E_n|\phi_n(t)\rangle. \quad (C1)$$

Setting the initial state $|\Phi_n(0)\rangle = |\phi_n(0)\rangle$, the system at the time t becomes

$$|\Phi_n(t)\rangle = \sum_n c_n(t) e^{-i \int_0^t E_n dt} |\phi_n(t)\rangle, \quad (C2)$$

where $c_n(t)$ is the probability amplitude for the system to be in the state $|\phi_n(t)\rangle$. Substituting $|\Phi_n(t)\rangle$ into Eq. (C1), we can obtain

$$\begin{aligned} \dot{c}_m &= -c_m \langle \phi_m(t) | \dot{\Phi}_m(t) \rangle \\ &= - \sum_{n \neq m} c_n(t) e^{-i \int_0^t (E_n - E_m) dt} \langle \phi_m(t) | \dot{\phi}_n(t) \rangle. \end{aligned} \quad (C3)$$

When the adiabatic approximation condition is satisfied

$$\left| \frac{\langle \phi_m(t) | \dot{H}_0(t) | \phi_n(t) \rangle}{(E_n - E_m)^2} \right| \ll 1, \quad (C4)$$

we can obtain $c_m(t) = 1$, indicating that the system evolves along its initial eigenstate without transition to other eigenstates.

We now analyze the explicit condition for generating the critical ground state using the adiabatic method. For the Hamiltonian $H_{\text{Rabi}}(g)$ in Eq. (6), when the parameter g varies with time slowly enough, i.e.,

$$\frac{dg}{dt} \approx 0, \quad (C5)$$

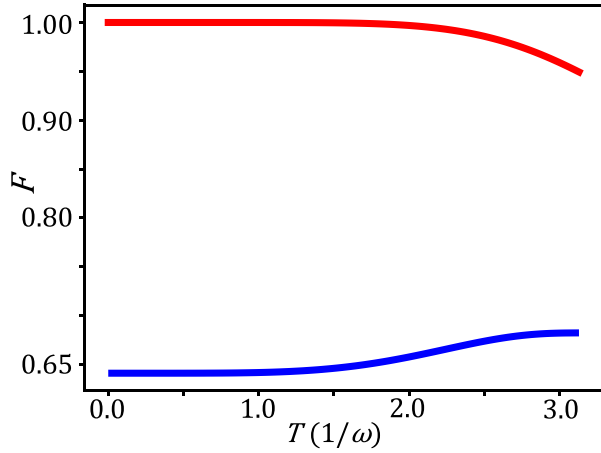


FIG. 8. Time evolution of the state fidelity during a short-time protocol using the adiabatic method. The fidelity of the evolved state $|\Phi(t)\rangle$ with respect to the initial ground state $|\Phi(g_i)\rangle$ (solid red line) and the target critical ground state $|\Phi(g_c)\rangle$ (dashed blue line).

it mean that the adiabatic approximation condition $d(H_{\text{Rabi}}(g))/dt \approx 0$ is satisfied, the system evolves along its initial eigenstate without transition to other eigenstates.

Specially, the instantaneous ground state $|\Phi(g)\rangle$ of $H_{\text{Rabi}}(g)$ is satisfied

$$H_{\text{Rabi}}(g)|\Phi(g)\rangle = E_0(g)|\Phi(g)\rangle, \quad (\text{C6})$$

where $E_0(g)$ is the ground energy. By slowly varying the parameter g ($T \rightarrow \infty$) in the Hamiltonian $H_{\text{Rabi}}(g)$ from an initial value g_i to a final value g_c , the system evolves from the initial state $|\Phi(g_i)\rangle$ to the final state $|\Phi(g_c)\rangle$.

A polynomial ansatz for the tunable parameter g is constructed to satisfy the boundary conditions at the initial ($t = 0, g = g_i$) and final ($t = T, g = g_c$) times. For comparison, we test the adiabatic method in the evolution time $T = 1$ of our protocol, but the result in Fig. 8 demonstrates that this approach is entirely ineffective.

2. STA method

In general, STA often involves introducing a counterdiabatic (CD) term into the original Hamiltonian to enforce the quantum system evolution along the adiabatic path. Here, we will provide a brief introduction to the CD term.

The aforementioned Hamiltonian $H_0(t)$ can be described as

$$H_0(t) = \sum_n |\phi_n(t)\rangle E_n \langle \phi_n(t)|, \quad (\text{C7})$$

where the approximate time-dependent adiabatic solution is

$$\Phi_n(t) = e^{i\zeta(t)} |\phi_n(t)\rangle, \quad (\text{C8})$$

where $\zeta(t)$ are the adiabatic phases with dynamical and geometric parts. Defining now the unitary operator

$$U = \sum_n e^{i\zeta(t)} |\phi_n(t)\rangle \langle \phi_n(0)|, \quad (\text{C9})$$

it must obey

$$i \frac{d}{dt} U = H(t)U, \quad (\text{C10})$$

from which we solve the Hamiltonian $H(t) = i(\partial_t U)U^\dagger$. Consequently, we can obtain the Hamiltonian $H(t)$ to drive the system exactly along the adiabatic channel of $H_0(t)$

$$\begin{aligned} H(t) &= H_0(t) + H_{\text{CD}}(t), \\ H_{\text{CD}}(t) &= i \sum_n [|\dot{\phi}_n(t)\rangle \langle \phi_n(t)| \\ &\quad - \langle \phi_n(t)| \dot{\phi}_n(t)\rangle |\phi_n(t)\rangle \langle \phi_n(t)|]. \end{aligned} \quad (\text{C11})$$

Here, an extra Hamiltonian $H_{\text{CD}}(t)$ to the original Hamiltonian $H_0(t)$ to ensure the system evolves along the adiabatic path without any nonadiabatic transitions over a short period of time. Consequently, we can achieve the adiabatic evolution by constructing STA using the CD term $H_{\text{CD}}(t)$ for a quantum system.

In the limit $\omega/\Omega \rightarrow 0$, the eigenstates of the Hamiltonian $H_{\text{Rabi}}(g)$ are given by

$$|\phi_n(t)\rangle = S(r)|n\rangle |\downarrow\rangle, \quad (\text{C12})$$

where $|n\rangle$ is a Fock state and $S(r) = \exp[(r/2)(a^{\dagger 2} - a^2)]$ is the squeeze operator with the squeezing parameter $r(g) = -\frac{1}{4} \ln(1 - g^2)$ [29,66]

$$H_{\text{CD}}(t) = i \frac{g\dot{g}}{4(1 - g^2)} (a^{\dagger 2} - a^2). \quad (\text{C13})$$

We introduce the CD term $H_{\text{CD}}(t)$ in the original Hamiltonian $H_{\text{Rabi}}(g)$. The complete Hamiltonian of the system becomes

$$\begin{aligned} H(t) &= \omega a^\dagger a + \frac{\Omega}{2} \sigma_z + g\sqrt{\omega\Omega} (a + a^\dagger) \sigma_x \\ &\quad + i \frac{g\dot{g}}{4(1 - g^2)} (a^{\dagger 2} - a^2), \end{aligned} \quad (\text{C14})$$

where we also impose the conditions $\dot{g}(t=0) = \dot{g}(t=T) = 0$ to ensure that $H_{\text{CD}}(t)$ vanishes at the beginning ($t = 0$) and end ($t = T$) of the evolution. Figure 9 shows the evolution of the state fidelity with respect to the initial and target critical ground states. The results are obtained by driving the system with the CD term $H_{\text{CD}}(t)$, whose time-dependent amplitude is plotted in Fig. 10. The STA method achieves a high-fidelity ($F = 0.9980$) final state for the target critical ground state $|\Phi(g_c)\rangle$ within the same evolution time T . However, the implementation of this method is experimentally challenging. The amplitude of the required CD driving term $H_{\text{CD}}(t)$ exhibits a sharp peak at the final time, with an amplitude exceeding the cavity frequency ω . In contrast, our protocol achieves a comparable fidelity using controls that are significantly more feasible to implement experimentally.

3. The gradient-ascent pulse engineering method

Another widely used optimal quantum control method, known as gradient-ascent pulse engineering (GRAPE), is introduced to achieve critical ground-state preparation. This method discretizes the continuous time evolution into short time segments and assumes a constant control field within each segment. The chain rule is employed to compute the

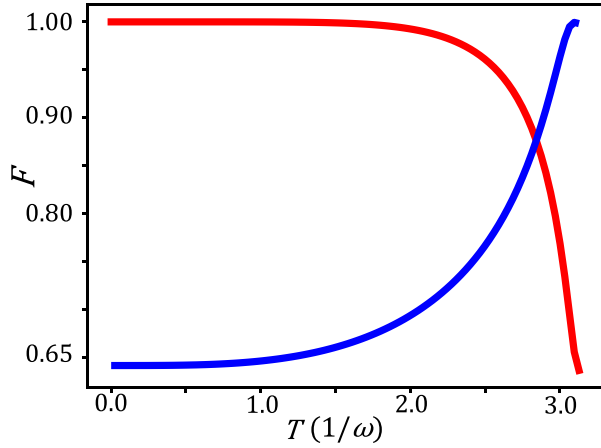


FIG. 9. Time evolution of the state fidelity during a short-time protocol using the STA method. The fidelity of the evolved state $|\Phi(t)\rangle$ with respect to the initial ground state $|\Phi(g_i)\rangle$ (solid red line) and the target critical ground state $|\Phi(g_c)\rangle$ (dashed blue line).

gradient of the objective function, typically the fidelity, with respect to the control amplitude in each time interval. The optimization procedure is iterated until the improvement of the objective function becomes negligible. To satisfy the requirements of the GRAPE method [36], the total Hamiltonian of the system is written as

$$H_{\text{tot}}(t) = H_{\text{Rabi}}(g_c) + \sum_{i=1}^n \Lambda_i(t) H_i^c, \quad (\text{C15})$$

where the system Hamiltonian is chosen as the terminal-time Hamiltonian of the time-varying system, $H_{\text{Rabi}}(g(T) = g_c = 1)$, as defined in the main text, in order to satisfy the GRAPE formalism. The set of control operators H_i^c is identical to that used in the main text. The quantity $\Lambda_i(t)$ denotes the time-discretized control amplitude.

To ensure a consistent benchmark, the number of time segments, the total evolution time, the upper and lower bounds of the control amplitude, and the number of training episodes for the DRL approach in the main text. This training is performed using the open-source `qutip-qtrl` package [67].

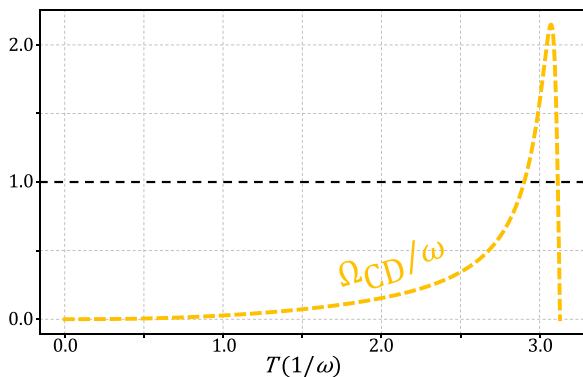


FIG. 10. The time-dependent amplitude of the CD term $H_{\text{CD}}(t)$, where $\Omega_{\text{CD}} = \frac{gg}{4(1-g^2)}$. Here, $g_{\text{STA}} \rightarrow 1$.

Under the above conditions, the control field obtained using the GRAPE method does not achieve high fidelity, with the maximum fidelity reaching 0.5300. The corresponding control waveform is shown in Fig. 11. The results indicate that, for such a time-varying system, evaluating the gradient of the fidelity with respect to the control-field amplitude is challenging. Under the fixed total evolution time and the specified number of time segments, the GRAPE method fails to obtain a control sequence that significantly improves the fidelity using accurate gradient information. In contrast, the DRL approach overcomes the strong dependence on gradient information by employing sampling-based feedback and yields control-field sequences that are not accessible through analytical methods.

APPENDIX D: QUANTUM FISHER INFORMATION

Quantum metrology relies on quantum Cramér-Rao bound that limits the sensitivity of estimating an unknown parameter Θ [66]

$$\text{Var}[\Theta] \geq \frac{1}{\sqrt{I_{\Theta}}}, \quad (\text{D1})$$

where the QFI I_{Θ} for a pure state $|\Phi\rangle$ with respect to a parameter Θ is

$$I_{\Theta} = 4(\langle \partial_{\Theta} \Phi | \partial_{\Theta} \Phi \rangle - |\langle \partial_{\Theta} \Phi | \Phi \rangle|^2), \quad (\text{D2})$$

with $\partial_{\Theta} = \partial/\partial\Theta$. To verify that the prepared state exhibits criticality-specific parameter sensitivity, a hallmark of quantum critical states, the quantum Fisher information with respect to the parameter $g(t)$ is analyzed. In the following discussion, the time dependence of g is suppressed for clarity. The states $|\Phi(t)\rangle = |\Phi(t, g)\rangle$ during evolution are implicitly g dependent, so we approximate the derivative via the finite difference method [68]. Therefore, we set the estimating parameter $\Theta = g$ and calculate the QFI of the states $|\Phi(t, g)\rangle$ at any time

$$\frac{\partial |\Phi(t, g)\rangle}{\partial g} \approx \frac{|\Phi(t, g + \delta)\rangle - |\Phi(t, g)\rangle}{\delta}, \quad (\text{D3})$$

where δ is a small perturbation. The physical meaning of Eq. (D3) is to quantify how much the geometric shape of the quantum state $|\Phi(t, g)\rangle$ (its position in Hilbert space) “shifts” or “distorts” at a given time t due to a minute change in the parameter g . Consequently, we obtain the complete QFI I_g and its time-dependent trajectory by performing this calculation at each time point t .

Figure 12 presents the results of our numerical simulation. The value of curve I_g sharply increases toward divergence at the end of the evolution, implying that the final state has acquired the heightened sensitivity to parameter variations that is characteristic of a critical point.

APPENDIX E: APPLICATION-QUANTUM DICKE MODEL

To demonstrate the generality of our approach, we also investigate another fundamental light-matter interaction system: the Dicke model. This model describes the collective interaction of N identical two-level systems with a single-mode bosonic field via a dipole coupling. The Hamiltonian of the

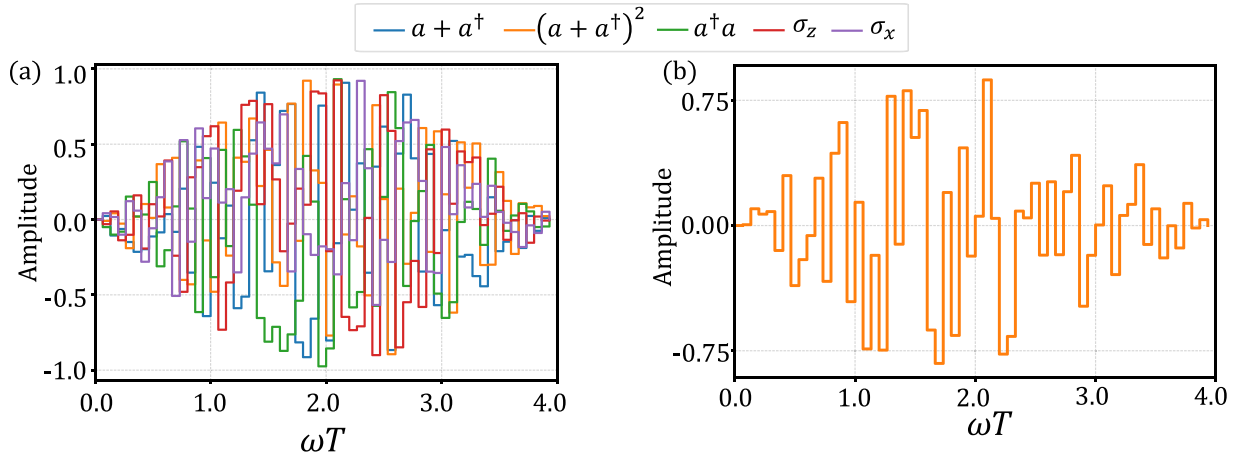


FIG. 11. Control-field waveforms obtained using the GRAPE method. During the optimization, the number of time segments is set to $K = 60$, the total evolution time satisfies $\omega T = 4$, and the upper and lower bounds of the control amplitude obey $|\Lambda_i/\omega| < 1$. The number of training episodes is set to 30 000, consistent with the DRL settings in the main text, and five random seeds are employed. Two control configurations are considered, (a) full control with five independent control fields and (b) single control with one control field. The final results show that the maximum fidelity achieved with five control fields is 0.5116, while the maximum fidelity achieved with a single control field is 0.5300. These results indicate that the GRAPE method has difficulty obtaining effective control sequences for such a rapidly time-varying system.

Dicke model is

$$H_{\text{Dicke}}[g'(t)] = \omega' a^\dagger a + \Omega' S_z + \frac{g'(t)\sqrt{\omega'\Omega'}}{\sqrt{N}}(a + a^\dagger)S_x, \quad (\text{E1})$$

where ω' is the frequency of single-mode bosonic field described by the bosonic operators $a^{(\dagger)}$, Ω' is the transition frequency, and $S_i = \sum_{j=1}^N \sigma_i$, $i = x, y, z$ are the collective spin operators. Similarly, following the same definition in the main text, we introduce a dimensionless parameter $g'(t) = 2\lambda'(t)/\sqrt{\omega'\Omega'}$, where $\lambda'(t)$ is the coupling strength.

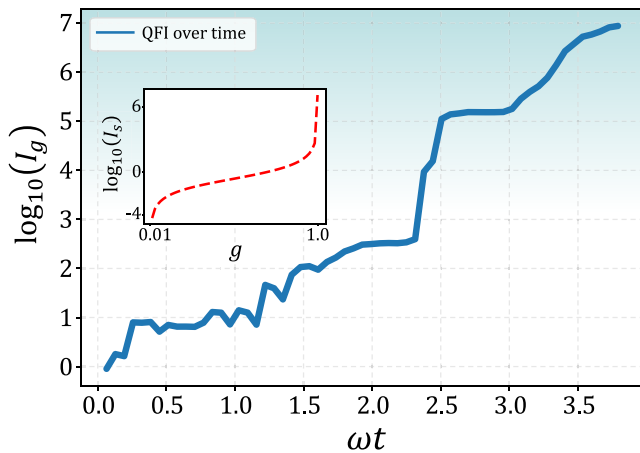


FIG. 12. Time evolution of I_g calculated for the evolving state $|\Phi(t)\rangle$ under the DRL-optimized control protocol, together with $I_s = g^2/2(1 - g^2)^2$ in the limit $\omega/\Omega \rightarrow 0$, shown as the standard upper bound. The results demonstrate that as the evolution approaches the final time ($t \rightarrow T$), the final state acquires extreme sensitivity to the parameter g and I_g approaches I_s . Here, $\delta = 10^{-5}$ and $g_s = 0.9999$.

In the thermodynamic limit $N \rightarrow \infty$, the system undergoes a normal-to-superradiant quantum phase transition at the critical point $g'_c = 1$. In this Appendix, a larger value of Ω'/ω' is chosen to partially emulate the thermodynamic limit that

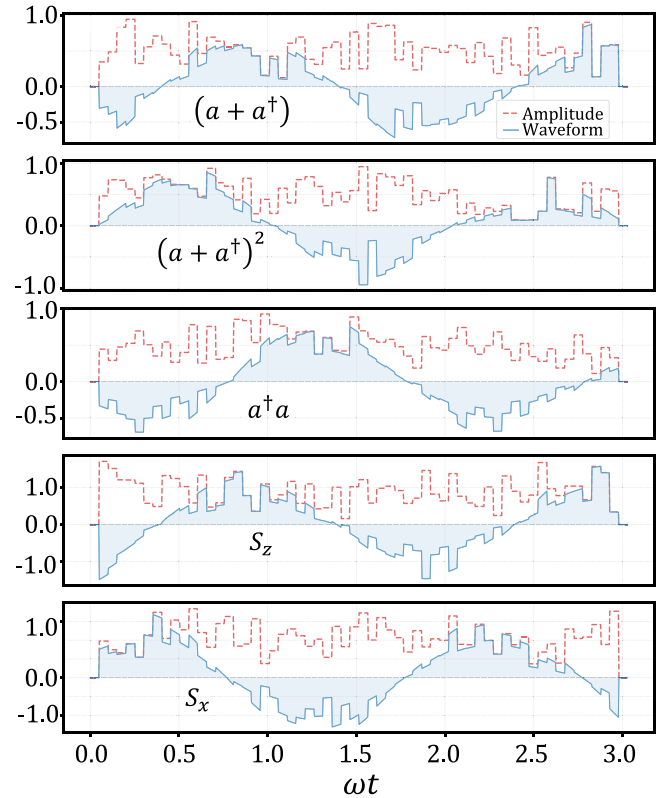


FIG. 13. Waveforms of the control pulses discovered by the DRL agent. The parameters used in the simulation are $\Omega' = 2500\omega'$, $N = 4$, $g'(0) = 0.01$, $g'(T) = g'_c = 1$, and $K = 60$.

would otherwise be achieved by increasing the number of qubits N .

To drive the system from the initial ground state $|\Phi[g'(0)]\rangle$ of $H_{\text{Dicke}}[g'(0)]$ to the target ground state $|\Phi[g'(T)]\rangle$ of $H_{\text{Dicke}}[g'(T) = g'_c]$, we employ the following five control Hamiltonians:

$$\begin{aligned} H_1^c(t) &= \Lambda'_1 \cos(\omega'_d t + \phi'_1)(a + a^\dagger), \\ H_2^c(t) &= \Lambda'_2 \cos(\omega'_d t + \phi'_2)(a + a^\dagger)^2, \\ H_3^c(t) &= \Lambda'_3 \cos(\omega'_d t + \phi'_3)a^\dagger a, \\ H_4^c(t) &= \Lambda'_4 \cos(\omega'_d t + \phi'_4)S_z, \\ H_5^c(t) &= \Lambda'_5 \cos(\omega'_d t + \phi'_5)S_x. \end{aligned} \quad (\text{E2})$$

We select a set of five control Hamiltonians commonly used in cavity quantum electrodynamics systems. The first three

terms, $H_1^c(t)$, $H_2^c(t)$, and $H_3^c(t)$, are identical to the control Hamiltonians used for the QRM in the main text. The remaining two terms, $H_4^c(t)$ and $H_5^c(t)$, represent global X and Z control fields, respectively, applied to all qubits. The control field constraints and all other DRL settings are identical to the specifications in the main text.

Now, we present the specific settings for our numerical experiment and subsequently showcase our findings. The optimal parameters ($N = 4$) found by the DRL framework deliver a fidelity $F(|\Phi(T)\rangle, |\Phi[g'(T)]\rangle) = 0.9953$ at $T = 3.0279/\omega'$. Similarly, the corresponding optimal parameters for the control Hamiltonians $\{\omega'_d, \phi'_i, \Lambda'_i\}$ are depicted in Fig. 13. This high fidelity confirms the protocol's applicability to other critical quantum systems. An optimal control scheme is thus obtained for preparing the critical quantum state within the given parameter constraints.

-
- [1] M. O. Scully and M. S. Zubairy, *Quantum Optics* (Cambridge University Press, Cambridge, 1997).
 - [2] Y.-H. Chen, Y. Qiu, A. Miranowicz, N. Lambert, W. Qin, R. Stassi, Y. Xia, S.-B. Zheng, and F. Nori, Sudden change of the photon output field marks phase transitions in the quantum Rabi model, *Commun. Phys.* **7**, 5 (2024).
 - [3] D. Rossini and E. Vicari, Coherent and dissipative dynamics at quantum phase transitions, *Phys. Rep.* **936**, 1 (2021).
 - [4] S. Sachdev, *Quantum Phase Transitions*, 2nd ed. (Cambridge University Press, Cambridge, 2011).
 - [5] M.-J. Hwang, R. Puebla, and M. B. Plenio, Quantum phase transition and universal dynamics in the Rabi model, *Phys. Rev. Lett.* **115**, 180404 (2015).
 - [6] M.-J. Hwang, P. Rabl, and M. B. Plenio, Dissipative phase transition in the open quantum Rabi model, *Phys. Rev. A* **97**, 013825 (2018).
 - [7] W. H. Zurek, U. Dorner, and P. Zoller, Dynamics of a quantum phase transition, *Phys. Rev. Lett.* **95**, 105701 (2005).
 - [8] P. Zanardi and N. Paunković, Ground state overlap and quantum phase transitions, *Phys. Rev. E* **74**, 031123 (2006).
 - [9] A. Carollo, D. Valenti, and B. Spagnolo, Geometry of quantum phase transitions, *Phys. Rep.* **838**, 1 (2020).
 - [10] A. Osterloh, L. Amico, G. Falci, and R. Fazio, Scaling of entanglement close to a quantum phase transition, *Nature (London)* **416**, 608 (2002).
 - [11] G. Vidal, J. I. Latorre, E. Rico, and A. Kitaev, Entanglement in quantum critical phenomena, *Phys. Rev. Lett.* **90**, 227902 (2003).
 - [12] V. E. Korepin, Universality of entropy scaling in one dimensional gapless models, *Phys. Rev. Lett.* **92**, 096402 (2004).
 - [13] L. Garbe, M. Bina, A. Keller, M. G. A. Paris, and S. Felicetti, Critical quantum metrology with a finite-component quantum phase transition, *Phys. Rev. Lett.* **124**, 120504 (2020).
 - [14] I. Frérot and T. Roscilde, Quantum critical metrology, *Phys. Rev. Lett.* **121**, 020402 (2018).
 - [15] R. Salvia, M. Mehboudi, and M. Perarnau-Llobet, Critical quantum metrology assisted by real-time feedback control, *Phys. Rev. Lett.* **130**, 240803 (2023).
 - [16] J. A. Vaitkus and A. D. Greentree, Digital three-state adiabatic passage, *Phys. Rev. A* **87**, 063820 (2013).
 - [17] R. Menchon-Enrich, J. Mompert, and V. Ahufinger, Spatial adiabatic passage processes in sonic crystals with linear defects, *Phys. Rev. B* **89**, 094304 (2014).
 - [18] Y.-H. Chen, Y. Xia, Q.-C. Wu, B.-H. Huang, and J. Song, Method for constructing shortcuts to adiabaticity by a substitute of counterdiabatic driving terms, *Phys. Rev. A* **93**, 052109 (2016).
 - [19] L. Innocenti, G. De Chiara, M. Paternostro, and R. Puebla, Ultrafast critical ground state preparation via bang-bang protocols, *New J. Phys.* **22**, 093050 (2020).
 - [20] S. Bernier and K. Agarwal, Spatiotemporal quenches for efficient critical ground state preparation in the two-dimensional transverse field Ising model, *Phys. Rev. B* **111**, 054311 (2025).
 - [21] Y. Yamamoto and H. A. Haus, Preparation, measurement and information capacity of optical quantum states, *Rev. Mod. Phys.* **58**, 1001 (1986).
 - [22] R. van Handel, J. K. Stockton, and H. Mabuchi, Modelling and feedback control design for quantum state preparation, *J. Opt. B: Quantum Semiclass. Opt.* **7**, S179 (2005).
 - [23] K.-X. Hu, B.-H. Huang, S.-W. Xu, Z.-C. Shi, Y. Xia, and Y.-H. Chen, High-fidelity control of cat-state qubits by polychromatic pulses, *Phys. Rev. Appl.* **25**, 024083 (2026).
 - [24] D.-S. Li, Y.-H. Kang, Z.-C. Shi, Y. Xiao, Y.-H. Chen, and Y. Xia, Efficient and flexible preparation of photonic noon states in a superconducting system, *Phys. Rev. A* **113**, 012619 (2026).
 - [25] G.-G. Gu, D.-S. Li, Y.-H. Chen, B.-H. Huang, and Y. Xia, Preparation of high-fidelity entangled cat states with composite pulses, *Adv. Quantum Technol.* **8**, 2400518 (2025).
 - [26] D. Girolami, How difficult is it to prepare a quantum state? *Phys. Rev. Lett.* **122**, 010505 (2019).
 - [27] F. Yang, Y.-C. Liu, and L. You, Manipulating photonic quantum states with long-range interactions, *Phys. Rev. A* **99**, 063803 (2019).
 - [28] M. Xue, X. Li, W. Ye, J.-J. Chen, Z.-F. Xu, and L. You, Controlling atomic spin mixing via multiphoton transitions in a cavity, *Phys. Rev. A* **106**, 033708 (2022).
 - [29] J.-W. Yu, K.-X. Yan, Y. Qiu, J. Song, Y.-H. Chen, and Y. Xia, Efficient generation of arbitrary photon-number-squeezed light via shortcuts to adiabaticity, *Opt. Express* **33**, 26356 (2025).

- [30] Y.-H. Chen, W. Qin, X. Wang, A. Miranowicz, and F. Nori, Shortcuts to adiabaticity for the quantum Rabi model: Efficient generation of giant entangled cat states via parametric amplification, *Phys. Rev. Lett.* **126**, 023602 (2021).
- [31] M. Mamaev, L. C. G. Govia, and A. A. Clerk, Dissipative stabilization of entangled cat states using a driven Bose-Hubbard dimer, *Quantum* **2**, 58 (2018).
- [32] Y. Zeng, W. Qin, Y.-H. Chen, C. Gneiting, and F. Nori, Neural-network-based design of approximate Gottesman-Kitaev-Preskill code, *Phys. Rev. Lett.* **134**, 060601 (2025).
- [33] X. Xu, S. C. Benjamin, and X. Yuan, Variational circuit compiler for quantum error correction, *Phys. Rev. Appl.* **15**, 034068 (2021).
- [34] F. A. Cárdenas-López, J. C. Retamal, and X. Chen, Shortcuts to adiabaticity in superconducting circuits for fast multi-partite state generation, *Commun. Phys.* **6**, 167 (2023).
- [35] D. Guéry-Odelin, A. Ruschhaupt, A. Kiely, E. Torrontegui, S. Martínez-Garaot, and J. G. Muga, Shortcuts to adiabaticity: Concepts, methods, and applications, *Rev. Mod. Phys.* **91**, 045001 (2019).
- [36] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, Optimal control of coupled spin dynamics: Design of NMR pulse sequences by gradient ascent algorithms, *J. Magn. Reson.* **172**, 296 (2005).
- [37] S. Machnes, U. Sander, S. J. Glaser, P. de Fouquières, A. Gruslys, S. Schirmer, and T. Schulte-Herbrüggen, Comparing, optimizing, and benchmarking quantum-control algorithms in a unifying programming framework, *Phys. Rev. A* **84**, 022305 (2011).
- [38] R. Porotti, V. Peano, and F. Marquardt, Gradient-ascent pulse engineering with feedback, *PRX Quantum* **4**, 030305 (2023).
- [39] Y. Zeng, Z.-Y. Zhou, E. Rinaldi, C. Gneiting, and F. Nori, Approximate autonomous quantum error correction with reinforcement learning, *Phys. Rev. Lett.* **131**, 050601 (2023).
- [40] 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).
- [41] H. Xu, T. Xiao, J. Huang, M. He, J. Fan, and G. Zeng, Toward Heisenberg limit without critical slowing down via quantum reinforcement learning, *Phys. Rev. Lett.* **134**, 120803 (2025).
- [42] A. Melkani, C. Gneiting, and F. Nori, Eigenstate extraction with neural-network tomography, *Phys. Rev. A* **102**, 022412 (2020).
- [43] F. Chen, J.-J. Chen, L.-N. Wu, Y.-C. Liu, and L. You, Extreme spin squeezing from deep reinforcement learning, *Phys. Rev. A* **100**, 041801(R) (2019).
- [44] J.-H. Cao, F. Chen, Q. Liu, T.-W. Mao, W.-X. Xu, L.-N. Wu, and L. You, Detection of entangled states supported by reinforcement learning, *Phys. Rev. Lett.* **131**, 073201 (2023).
- [45] Y. Zhu, T. Xiao, G. Zeng, G. Chiribella, and Y.-D. Wu, Controlling unknown quantum states via data-driven state representations, *npj Quantum Inf.* (2026), doi:10.1038/s41534-026-01269-0.
- [46] V. V. Sivak, A. Eickbusch, H. Liu, B. Royer, I. Tsioutsios, and M. H. Devoret, Model-free quantum control with reinforcement learning, *Phys. Rev. X* **12**, 011059 (2022).
- [47] Y. Yu, Y. Qiu, X. Zhao, Y.-H. Chen, and Y. Xia, Pulse parameterization and seeded optimization for fast longitudinal qubit readout, *Phys. Rev. Appl.* (2026), doi:10.1103/1dg9-b4vw.
- [48] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, Deep reinforcement learning: A brief survey, *IEEE Signal Process. Mag.* **34**, 26 (2017).
- [49] 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).
- [50] T.-N. Xu, Y. Ding, J. D. Martín-Guerrero, and X. Chen, Robust two-qubit gate with reinforcement learning and dropout, *Phys. Rev. A* **110**, 032614 (2024).
- [51] V. Sivak, M. Newman, and P. Klimov, Optimization of decoder priors for accurate quantum error correction, *Phys. Rev. Lett.* **133**, 150603 (2024).
- [52] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, Proximal policy optimization algorithms, *arXiv:1707.06347*.
- [53] H. Walther, B. T. H. Varcoe, B.-G. Englert, and T. Becker, Cavity quantum electrodynamics, *Rep. Prog. Phys.* **69**, 1325 (2006).
- [54] J. Johansson, P. Nation, and F. Nori, QuTiP: An open-source Python framework for the dynamics of open quantum systems, *Comput. Phys. Commun.* **183**, 1760 (2012).
- [55] 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).
- [56] T. Niemczyk, F. Deppe, H. Huebl, E. P. Menzel, F. Hocke, M. J. Schwarz, J. J. García-Ripoll, D. Zueco, T. Hümmer, E. Solano, A. Marx, and R. Gross, Circuit quantum electrodynamics in the ultrastrong-coupling regime, *Nat. Phys.* **6**, 772 (2010).
- [57] X. Gu, A. F. Kockum, A. Miranowicz, Y.-x. Liu, and F. Nori, Microwave photonics with superconducting quantum circuits, *Phys. Rep.* **718-719**, 1 (2017).
- [58] M.-A. Lemonde, N. Didier, and A. A. Clerk, Enhanced nonlinear interactions in quantum optomechanics via mechanical amplification, *Nat. Commun.* **7**, 11338 (2016).
- [59] P. D. Nation, J. R. Johansson, M. P. Blencowe, and F. Nori, *Colloquium: Stimulating uncertainty: Amplifying the quantum vacuum with superconducting circuits*, *Rev. Mod. Phys.* **84**, 1 (2012).
- [60] D. Ballester, G. Romero, J. J. García-Ripoll, F. Deppe, and E. Solano, Quantum simulation of the ultrastrong-coupling dynamics in circuit quantum electrodynamics, *Phys. Rev. X* **2**, 021007 (2012).
- [61] M. Mirrahimi, Z. Leghtas, V. V. Albert, S. Touzard, R. J. Schoelkopf, L. Jiang, and M. H. Devoret, Dynamically protected cat-qubits: A new paradigm for universal quantum computation, *New J. Phys.* **16**, 045014 (2014).
- [62] J.-H. Lin, K.-X. Yan, J. Song, Y.-H. Chen, and Y. Xia, Approaching the critical point through an exact ground-state evolution, *Phys. Rev. Res.* **8**, 013275 (2026).
- [63] Y.-H. Chen, R. Stassi, W. Qin, A. Miranowicz, and F. Nori, Fault-tolerant multiqubit geometric entangling gates using photonic cat-state qubits, *Phys. Rev. Appl.* **18**, 024076 (2022).
- [64] M. M. Rams, P. Sierant, O. Dutta, P. Horodecki, and J. Zakrzewski, At the limits of criticality-based quantum metrology: Apparent super-Heisenberg scaling revisited, *Phys. Rev. X* **8**, 021022 (2018).

- [65] T. Lindeberg, Discrete approximations of Gaussian smoothing and Gaussian derivatives, [J. Math. Imaging Vision 66, 759 \(2024\)](#).
- [66] K. Gietka, F. Metz, T. Keller, and J. Li, Adiabatic critical quantum metrology cannot reach the Heisenberg limit even when shortcuts to adiabaticity are applied, [Quantum 5, 489 \(2021\)](#).
- [67] N. Lambert, E. Giguère, P. Menczel, B. Li, P. Hopf, G. Suárez, M. Galí, J. Lishman, R. Gadhvi, R. Agarwal, A. Galicia, N. Shammah, P. Nation, J. R. Johansson, S. Ahmed, S. Cross, A. Pitchford, and F. Nori, QuTiP 5: The quantum toolbox in Python, [Phys. Rep. 1153, 1 \(2026\)](#).
- [68] P.-B. Zhou, *Numerical Analysis of Electromagnetic Fields* (Springer, Berlin, 1993), pp. 63–94.