

Autonomous Floquet Engineering of Bosonic Codes via Reinforcement Learning

Zheping Wu,¹ Lingzhen Guo,^{2,3} Haobin Shi,¹ and Wei-Wei Zhang^{1,*}

¹*School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China*

²*Center for Joint Quantum Studies and Department of Physics,
School of Science, Tianjin University, Tianjin 300072, China*

³*Tianjin Key Laboratory of Low Dimensional Materials Physics and Preparing Technology,
Department of Physics, Tianjin University, Tianjin 300354, China*

(Dated: August 6, 2026)

SUPPLEMENTARY NOTE 1: NONCOMMUTATIVE FOURIER TRANSFORMATION (NCFT)

In this section, we introduce the decomposition of a target Hamiltonian in the form of a noncommutative Fourier transformation (NcFT). The target Hamiltonian $\hat{H}_{\text{tar}}$ is presented as a sum of plane-wave operators

$$\hat{H}_{\text{tar}} = \frac{\beta}{2\pi} \int \int dk_x dk_p f_T(k_x, k_p) e^{i(k_x \hat{x} + k_p \hat{p})}. \quad (\text{S1.1})$$

The Fourier coefficients $f_T(k_x, k_p)$ are the inverse transformation,

$$f_T(k_x, k_p) = \frac{e^{\frac{\lambda}{4}(k_x^2 + k_p^2)}}{2\pi\beta} \int \int dx dp H_Q^{(T)}(x, p) e^{-i(k_x x + k_p p)}. \quad (\text{S1.2})$$

The $H_Q^{(T)}(x, p)$ can be obtained using

$$H_Q^{(T)}(\alpha, \alpha^*) \equiv \langle \alpha | \hat{H}_F^{(T)} | \alpha \rangle = \sum_{n,m} \chi_{nm} (\alpha^*)^n \alpha^m \quad (\text{S1.3})$$

with the replacement of

$$\begin{cases} x \equiv \langle \alpha | \hat{x} | \alpha \rangle = \sqrt{\frac{\lambda}{2}}(\alpha^* + \alpha) \\ p \equiv \langle \alpha | \hat{p} | \alpha \rangle = i\sqrt{\frac{\lambda}{2}}(\alpha^* - \alpha). \end{cases} \quad (\text{S1.4})$$

Note that from the hermicity of Hamiltonian operator, f_T follows the relationship

$$f_T(k_x, k_p) = f_T^*(-k_x, -k_p). \quad (\text{S1.5})$$

The detailed analytical expression of the NcFT coefficient $f_T(k \cos \tau, k \sin \tau)$ for the preparation of q -leg cat state is given by Ref. [1]. Here, $\tau = \omega t$ is used for the sake of formulaic simplicity.

$$\begin{aligned}
f_T(k \cos \tau, k \sin \tau) &= \frac{e^{\lambda k^2/4}}{2\pi\beta} \int_0^{+\infty} r dr \int_0^{2\pi} d\phi H_{Q\gamma}^{(T)}(r \cos \phi, r \sin \phi) e^{-ikr \cos(\phi-\tau)} \\
&= \frac{e^{\lambda k^2/4}}{|\alpha_0 e^\gamma|^{2q}} \int_0^{+\infty} r dr e^{-r^2/(2\lambda\sigma_\gamma^2)} \left[\frac{r^{2q}}{(2\lambda)^q} J_0(kr) - \frac{(ir\alpha_0 e^{\gamma-i\tau})^q}{(2\lambda)^{q/2}} J_{-q}(kr) \right. \\
&\quad \left. - \frac{(-ir\alpha_0^* e^{\gamma+i\tau})^q}{(2\lambda)^{q/2}} J_q(kr) + |\alpha_0 e^\gamma|^{2q} J_0(kr) \right] \\
&= \frac{\lambda e^{\lambda k^2/4} \sigma_\gamma^{2(q+1)}}{|\alpha_0 e^\gamma|^{2q}} \left[q! {}_1F_1\left(1+q; 1; -\frac{\lambda}{2} \sigma_\gamma^2 k^2\right) \right. \\
&\quad - \left(-ie^{-i\tau} \alpha_0 e^\gamma \sqrt{\frac{\lambda}{2}} \right)^q k^q {}_1F_1\left(1+q; 1+q; -\frac{\lambda}{2} \sigma_\gamma^2 k^2\right) \\
&\quad - \left(-ie^{i\tau} \alpha_0^* e^\gamma \sqrt{\frac{\lambda}{2}} \right)^q k^q {}_1F_1\left(1+q; 1+q; -\frac{\lambda}{2} \sigma_\gamma^2 k^2\right) \\
&\quad \left. + \left| \frac{\alpha_0 e^\gamma}{\sigma_\gamma} \right|^{2q} {}_1F_1\left(1; 1; -\frac{\lambda}{2} \sigma_\gamma^2 k^2\right) \right]. \tag{S1.6}
\end{aligned}$$

18 The detailed analytical expression of the NcFT coefficient $f_T(k \cos \tau, k \sin \tau)$ for the preparation of binomial superpo-
 19 sition state is [2]

$$f_T(k \cos \tau, k \sin \tau) = -\frac{\Delta}{4} \left[f_{00} + 3f_{22} + 3f_{44} + f_{66} + \sqrt{3}(f_{04} + f_{26} + \text{c.c.}) \right]. \tag{S1.7}$$

20 where $f_{nm} \equiv f_{nm}(k \cos \tau, k \sin \tau)$ denotes the NcFT component associated with the operator $|n\rangle\langle m|$.

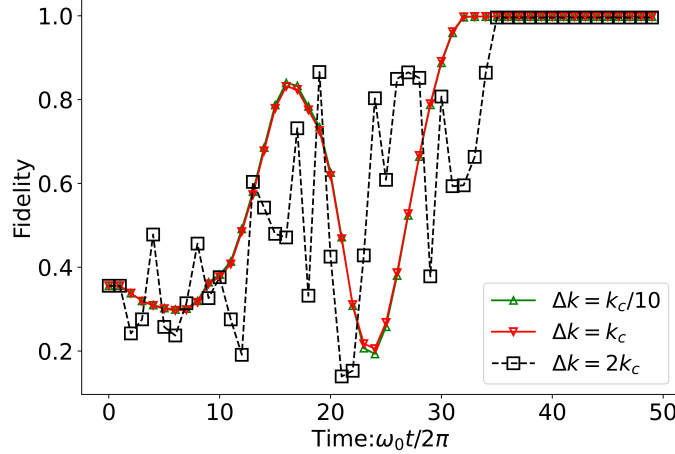


Figure S1. Dynamics with various discretized wavenumbers. Time evolution of the fidelity of the prepared state based on the approximate driving potential superposed by a finite number of cosine functions with discretized wave numbers, $\Delta k = k_c/10$ (green, fidelity=0.9982), $\Delta k = k_c$ (red, fidelity=0.9984) and $\Delta k = 2k_c$ (black, fidelity=0.9961).

21 SUPPLEMENTARY NOTE 2: EXPERIMENTAL REALIZATION OF THE DRIVING FIELD

22 In this section, we introduce the physical implementation of the proposed driving protocol, including oscillator en-
 23 codings in cold atoms and superconducting circuits, finite wave-vector discretization, RWA validity, noise robustness,
 24 the role of $\beta_{\max}$, and comparison with gradient-based optimization.

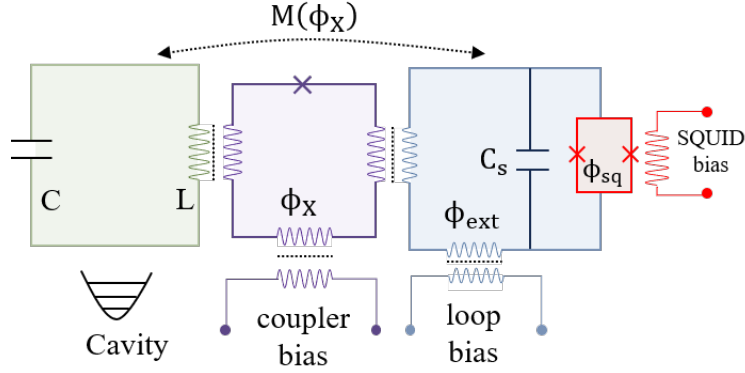


Figure S2. Experimental implementation of a cosine potential in the superconducting circuit platform. A superconducting cavity, represented by the LC circuit on the left, is coupled to a capacitively shunted SQUID circuit on the right through the intermediate coupler. The coupler generates an effective mutual inductance ($M(\Phi_x)$) between the cavity and the SQUID circuit. The fluxes (Φ_x), (Φ_{ext}), and (Φ_{sq}), applied to the coupler, the external SQUID loop, and the SQUID itself, respectively, can be independently controlled through the coupler-bias, loop-bias, and SQUID-bias lines.

A. The physical realization details driving potential

The realization of the designed driving potential in Eq. (4) in the main text can be implemented in the platforms of cold atoms [3–5] or superconducting circuits [1, 6–11].

In the platform of cold atoms, the bosonic mode is the external center-of-mass motion of an atom confined along one spatial direction by an approximately harmonic trap, where the Hamiltonian is given by

$$\hat{H}_{\text{at}}(t) = \frac{\hat{P}^2}{2M} + \frac{1}{2}M\omega_0^2\hat{X}^2 + V_{\text{opt}}(\hat{X}, t). \quad (\text{S2.1})$$

The bosonic code is encoded in the Fock space of the external center-of-mass motion

$$\{|0\rangle_{\text{mot}}, |1\rangle_{\text{mot}}, |2\rangle_{\text{mot}}, \dots\}, \quad (\text{S2.2})$$

rather than in internal hyperfine or electronic states. The cosine components of the driving field, cf. Eq. (4) in the main text, are realized with a pair of laser beams, which produces an optical-dipole potential of the form

$$V_n(X, t) = A_n(t) \cos[k_n X + \phi_n(t)]. \quad (\text{S2.3})$$

For two beams of equal wavelength crossing at an angle θ_n , the projected lattice wave number is $k_n = 2\frac{2\pi}{\lambda_L} \sin(\frac{\theta_n}{2})$, where λ_L is the wavelength of the laser beams. The lattice depth $A_n(t)$ is controlled through the laser intensity and detuning. The wave number k_n is controlled by the laser wavelength and intersection angle. The relative optical phase determines the parameter $\phi_n(t)$. A frequency difference between the two beams generates a time-dependent phase and hence a moving optical lattice.

In the platform of superconducting circuits, the bosonic mode is provided by the high Q-factor 3D microwave cavity or 1D on-chip resonator. The quantum state is encoded in the photon-number Hilbert space of this resonator. Josephson junctions supply the nonlinearity required to generate the cosine components of the driving field. The specific experimental realization scheme of the cosine potential is shown in Fig. S2. The superconducting cavity on the left is coupled to the SQUID circuit on the right through an intermediate Josephson-junction-based element, referred to as the coupler. The SQUID is shunted by a large capacitance C_s and incorporates an external flux loop. The coupler generates a tunable effective mutual inductance $M(\Phi_x)$ between the cavity and the SQUID circuit [12]. The fluxes Φ_x , Φ_{ext} , and Φ_{sq} , applied to the coupler, the external SQUID loop, and the SQUID itself, respectively, can be controlled independently through the coupler-bias, loop-bias, and SQUID-bias lines. The resulting cavity dynamics are described by the effective Hamiltonian

$$\mathcal{H}(t) = \hbar\omega_0\hat{a}^\dagger\hat{a} + V_{\text{SQUID}}(\hat{\varphi}), \quad (\text{S2.4})$$

where $\omega_0 = \frac{1}{\sqrt{LC}}$ is the bare resonance frequency of the cavity. The large shunt capacitance C_s suppresses the charging energy of the SQUID circuit [13], while its inductive energy is determined by the SQUID potential as described by

$$V_{\text{SQUID}}(\hat{\varphi}) = -E_J \cos\left(\frac{\Phi_{\text{sq}}}{\Phi_0}\pi\right) \cos\left[\frac{M(\Phi_x)}{L}\hat{\varphi} + \frac{\Phi_{\text{ext}}}{\Phi_0}\right]. \quad (\text{S2.5})$$

Here, E_J is the Josephson energy of a single junction, $\Phi_0 = 2\pi\hbar/(2e)$ is the magnetic-flux quantum, and

$$\hat{\varphi} = \sqrt{\frac{\lambda}{2}}(\hat{a}^\dagger + \hat{a}) = \frac{2\pi\hat{\Phi}}{\Phi_0} \quad (\text{S2.6})$$

is the dimensionless cavity-flux operator, with $\lambda = 4e^2\omega_0 L/\hbar$. By adjusting the coupler-bias flux, the effective mutual inductance $M(\Phi_x)$ can be tuned to zero, made arbitrarily large, or even driven to negative values [12]. By tuning either the cavity-SQUID coupling through the mutual inductance or the Josephson energy of the SQUID, the cosine potential correspondence lattice gates can, in principle, be realized, as proposed in Ref. [1]. The gate parameters γ , ρ , and δ are controlled by the applied fluxes Φ_{sq} , Φ_x , and Φ_{ext} , respectively.

B. The finite number of wavevectors

As the driving field in Eq. 4 is represented as a superposition of cosinusoidal potentials which could be approximated by replacing the integral with the sum of a finite number of cosine lattice potentials, as following,

$$V(x, t) \approx \sum_{n=1}^{k_{\text{cut}}/\Delta k} A(n\Delta k, t) \cos[(n\Delta k)x + \phi(k, t)]\Delta k. \quad (\text{S2.7})$$

The number of the required cosine lattice potentials corresponds to the number of Josephson junction (JJ) in the superconducting circuits [1, 14] or the number of laser beams in cold atoms systems [3]. As shown in Fig. S1, our protocol remains highly effective for discretized wavenumber spacings $\Delta k = k_c/10$ (100 JJs), k_c (10 JJs), and $2k_c$ (5 JJs), yielding final fidelities of 0.9982, 0.9984, and 0.9961, respectively. Remarkably, even with the coarsest discretization with five JJs in the superconducting-circuit, still gives a fidelity of 0.9961, showing a good experimental feasibility of our work.

C. The rotating wave approximation (RWA) conditions

Generally, the Floquet Hamiltonian $\hat{H}$ is an arbitrary function of noncommutative operators $\hat{x}$ and $\hat{p}$, and it can not be directly decomposed into a combination of kinetic and potential terms. The Floquet-Magnus expansion [15, 16] is the technique for this situation, while the driving frequency ω_d (and the chosen Floquet frequency Ω) is much larger than the characteristic frequency of the system and $\beta \ll \omega_0$, with

$$\lim_{\omega_0/\beta \rightarrow \infty} \hat{H}_F(\hat{x}, \hat{p}) = \frac{1}{T} \int_0^T dt \hat{H}(t). \quad (\text{S2.8})$$

In experimentally relevant superconducting-circuit implementations, the Floquet drive high-frequency condition is understood relative to the effective low-energy scales of the problem. Up to leading order (RWA), the right-hand side of Eq. (S2.8) coincides with the target Hamiltonian $\hat{H}_F^{(T)}$, where the ensuing solution becomes exact in the high-frequency limit $\omega_0/\beta \rightarrow \infty$.

D. State preparation in dissipative environments

Generally, superconducting circuits interact with a dirtier environment due to charge and flux noise. Considering a typical superconducting aluminium cavity with frequency $\omega_0/2\pi = 5\text{GHz}$, the cavity relaxation time $T_1 = 1/\kappa$ can be longer than $500\mu\text{s}$ [17, 18], which corresponds to $\kappa/\omega_0 \approx 6.4 \times 10^{-8}$. The pure dephasing rate for a cavity is usually much weaker than the relaxation rate where the typical pure dephasing time $T_\phi = 1/\eta$ is longer than 1ms [17, 18]

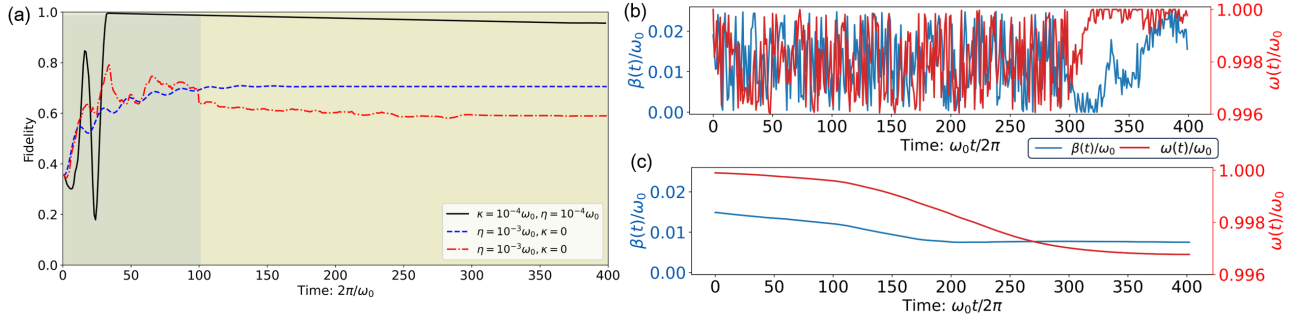


Figure S3. (a) Stroboscopic fidelity of the prepared state as a function of time under various noisy environments, considering the dephasing noise. The black solid line corresponds to $\kappa = 10^{-4}\omega_0$ and $\eta = 10^{-4}\omega_0$. The red dashed line shows the fidelity for $\eta = 10^{-3}\omega_0$ and $\kappa = 0$ obtained with the original policy, while the blue dash-dotted line shows the fidelity obtained with a retrained policy under the same noise condition. (b) Time dependent control of $\beta(t)$ (blue line) and $\omega(t)$ (red line) obtained with the original policy. (c) Time-dependent control of $\beta(t)$ (blue line) and $\omega(t)$ (red line) obtained with the retrained policy.

corresponding to $\eta/\omega_0 < 3.2 \times 10^{-8}$. The dynamics of the prepared state in a noisy environment are described by the Lindblad master equation,

$$\frac{d}{dt}\rho(t) = -\frac{i}{\lambda}[\hat{H}(t), \rho(t)] + \kappa\left(\hat{a}\rho\hat{a}^\dagger - \frac{1}{2}\{\hat{a}^\dagger\hat{a}, \rho\}\right) + \eta\left(\hat{a}^\dagger\hat{a}\rho\hat{a}^\dagger\hat{a} - \frac{1}{2}\{(\hat{a}^\dagger\hat{a})^2, \rho\}\right). \quad (\text{S2.9})$$

As shown in Fig. ??, while our RL model is trained within a low noise environment ($\kappa = 10^{-7}\omega_0, \eta = 10^{-7}\omega_0$), the trained neural network is able to prepare target states in various high noise environments, such as with noise $\kappa = 10^{-4}\omega_0, \eta = 10^{-4}\omega_0$.

While increasing the noise level further to a high photon loss environment with $\kappa = 10^{-3}\omega_0, \eta = 0\omega_0$, our trained model is not able to prepare the target state with high fidelity with a fidelity around 0.3. But by retraining our RL model, the retrained neural network can improve the fidelity of the prepared state to around 0.8, as shown in Fig. 4. We investigate the photon number dynamics in the photon loss environment. As shown in Fig. 5, the photon number is being stabilized around 2 with our retrained neural networks, while without retraining, the photon number drops to around 1. Meanwhile, the retrained neural networks find the driving field is always fluctuating around 0.025 to compensate for the environmental photon loss.

As shown in Fig. S3, while increasing the noise level further to a high dephasing environment with $\eta = 10^{-3}\omega_0, \kappa = 0\omega_0$, our trained model is not able to prepare the target state with high fidelity with a fidelity around 0.6. By retraining the model, the Fidelity can be improved to around 0.7. We conjecture that to further improve the fidelity, we need more complicated formula components for the driving field, which will be our future work.

E. The role of $\beta_{\max}$ in our reinforcement learning approach

In this section, we investigate the effect of $\beta_{\max}$ in our reinforcement learning approach for the preparation of bosonic codes. The setting of $\beta/\omega_0 \ll 1$ ensures the weak nonlinearity condition, and the system evolution in the rotating frame is slow, which allows the approximation of the Floquet Hamiltonian with the leading order of the Floquet-Magnus expansion in Eq. S2.8 to be valid.

In our reinforcement learning (RL) approach, we train the neural networks to find an optimized control sequence for the driving field. The value of $\beta_{\max}/\omega_0$ directly sets the optimization space for our RL algorithm. In Fig. S4, we observe that while $\beta_{\max} \leq 0.025$ the value of $\beta_{\max}/\omega_0$ closely related with the state preparation time. Specifically, with $\beta_{\max}/\omega_0 = 0.015$, the state preparation time is 40 steps, while the state preparation time with $\beta_{\max}/\omega_0 = 0.025$ is 34 steps. We conjecture that within the validity range of RWA, the larger the value of $\beta_{\max}$ is, the shorter the optimized state preparation time is. In the early stage of our training process of a total of 300 episodes, β/ω_0 explores the whole range $[0, 0.1]$, as shown by the case of the 1st training episode in Fig. S5 (a). As training progresses, our RL approach learns to concentrate the exploration into a narrower interval. As shown in Fig. S5 (b), in the 156th training episode, β/ω_0 mainly explores within $[0, 0.025]$. Using the final policy obtained after the training of a total of 300 episodes, we test the preparation of the 4-leg cat state with $\beta_{\max}/\omega_0 = 0.1$, and the RL agent accomplishes

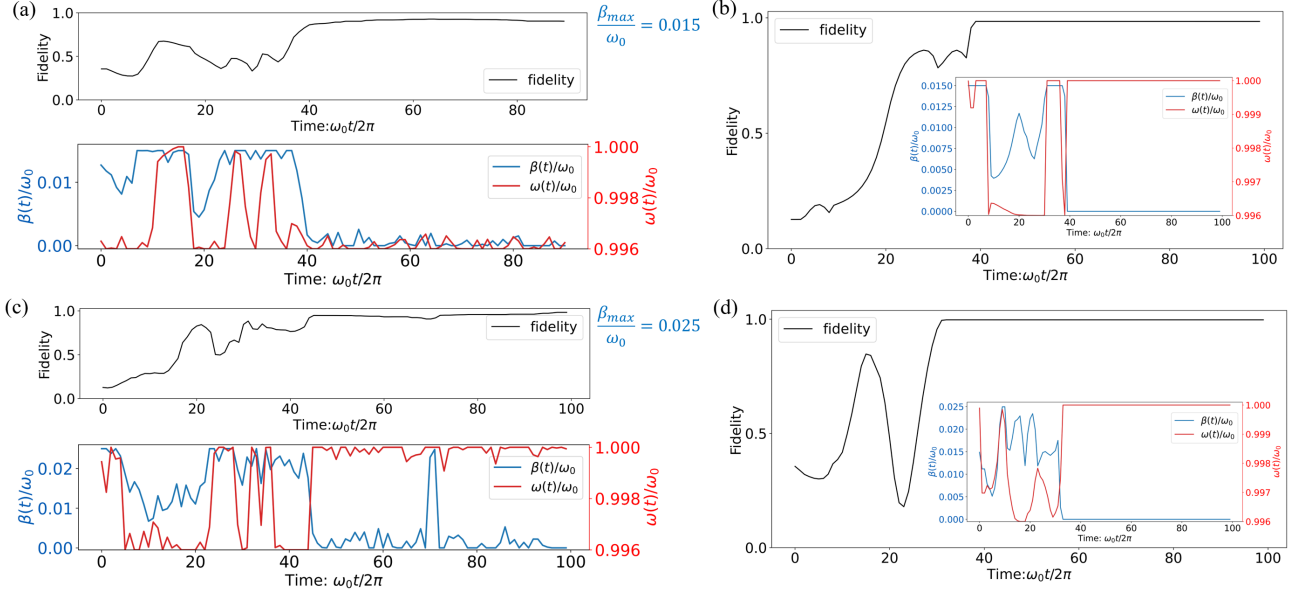


Figure S4. The training (a, c) and testing (b, d) trajectories for (a, b) $\beta_{\max}/\omega_0 = 0.015$ and (c, d) $\beta_{\max}/\omega_0 = 0.025$. The black, blue, and red curves represent the fidelity, $\beta(t)$, and $\omega(t)$ respectively.

the task in only 20 steps. The maximum value of $\beta(t)/\omega_0$ in this testing process is slightly above 0.025, as shown in Fig. S5 (c) (the 300th episode). With $\beta_{\max}/\omega_0 = 0.1$, the achieved state fidelity is 0.991, which is lower than the value of 0.998 obtained with $\beta_{\max}/\omega_0 = 0.025$. We conjecture that the lower fidelity in the $\beta_{\max}/\omega_0 = 0.1$ case arises from the breakdown of the rotating-wave approximation (RWA).

We further investigate the effect of $\beta_{\max}$ on the fidelity of the prepared state achieved with our retrained RL model under a large noise environment. We discover that while $\beta_{\max} < 0.025\omega_0$, the fidelity is slightly lower than that obtained with $\beta_{\max} = 0.025\omega_0$, while $\beta_{\max} > 0.025\omega_0$, the fidelity is around the same value as that obtained with $\beta_{\max} = 0.025\omega_0$, as shown in Fig. S6. These numerical studies also support our choice of $\beta_{\max} = 0.025\omega_0$ as a good set.

Algorithm 1: Gradient-based optimization method

```

Initialize period length T
Initialize control parameters  $x = \{\beta_1, \dots, \beta_T, \omega_1, \dots, \omega_T\}$ 
Initialize  $H_0$ ,  $V(t)$ ,  $\psi_{\text{init}}$ , and  $\psi_{\text{tar}}$ 
Set control bounds and tail window length  $M = 20$ 
for  $k = 1$  to  $K_{\max}$  do
    Propagate  $\psi$  under  $x$ 
    Compute fidelity trajectory  $\{F_1, \dots, F_T\}$ 
    Let  $\bar{F}$  and  $F_{\min}$  be the mean and minimum of the last  $M$  fidelities
    Compute loss  $\mathcal{L}(x) = (1 - \bar{F})^2 + (1 - F_{\min})^2$ 
    Update the control parameters using gradient-based optimization
    if converged then
        break
    end
end
Return  $x^*$ 

```

SUPPLEMENTARY NOTE 3: GRADIENT-BASED OPTIMIZATION METHOD

In this section, we present a gradient-based optimization method for the same task, which is a variant of the Quantum Gradient Ascent Pulse Engineering (GRAPE) method [19], and uses our reinforcement-inspired reward. The gradient-

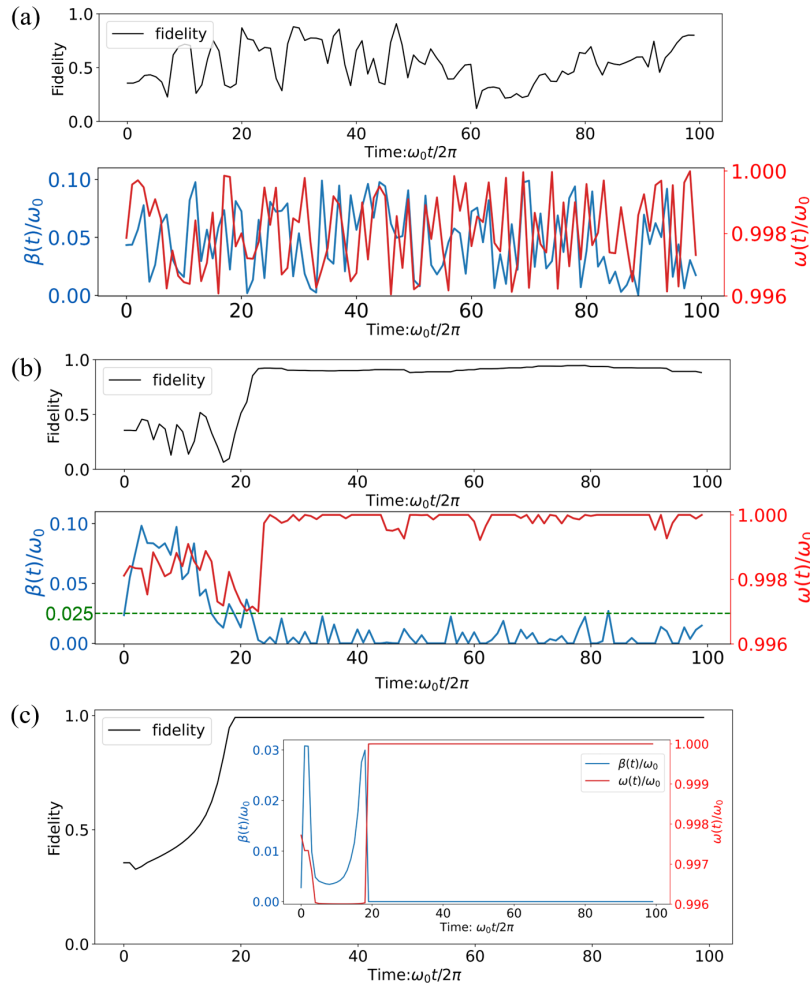


Figure S5. (a) The evolution of prepared state fidelity and the corresponding training trajectory from the 1st episode, (b) The evolution of prepared state fidelity and the corresponding training trajectory from the 156th episode, and (c) The evolution of prepared state fidelity and the corresponding testing trajectory from the 300th (final) episode, with $\beta_{\max}/\omega_0 = 0.1$, achieved using our reinforcement learning approach. The black, blue, and red curves represent the fidelity, $\beta(t)$, and $\omega(t)$, respectively.

based optimization requires a prefixed number of evolution steps to define the optimization spaces. Specifically, the gradient-based optimization method computes the gradient of the fidelity with respect to each control parameter efficiently: $\frac{\partial \mathcal{F}}{\partial \beta(t)}$ and $\frac{\partial \mathcal{F}}{\partial \omega(t)}$, and then updates the controls iteratively: $\beta(t+1) \leftarrow \beta(t) + \epsilon \frac{\partial \mathcal{F}}{\partial \beta(t)}$, $\omega(t+1) \leftarrow \omega(t) + \epsilon \frac{\partial \mathcal{F}}{\partial \omega(t)}$, with ϵ as the learning rate. For the bosonic state preparation task, we design a RL inspired loss function which is $\mathcal{L}(x) = (1 - \bar{F})^2 + (1 - F_{\min})^2$ with $\bar{F}$ and $F_{\min}$ as the mean and minimum of the fidelity values in the last M steps. The corresponding pseudocode is shown in Alg. 1. As the performance of gradient-based optimization relies heavily on the initial parameters, it is not guaranteed that the prepared state has high fidelity to the target state.

We thoroughly study its performance under noise-free and noisy environments. While considering a noise-free environment, we apply the gradient-based optimization method for the preparation of a bosonic 4-leg state with various settings of the prefixed number of evolution steps, as shown in Fig. S7. We can observe that for various settings of T , the achieved fidelity fluctuates. With a $T = 50$, the mean of the obtained fidelity is highest, and the corresponding variance is smallest. Therefore, we adopt $T = 50$ in the following experiments. Under the noise-free setting with $T = 50$, we perform 200 independent runs with different random seeds, obtaining a mean achieved fidelity of 0.9541.

While considering a noisy environment, we apply the gradient-based optimization method with $T = 50$ for the bosonic 4-leg cat state preparation with a noise level of $\kappa = 10^{-7}\omega_0$, $\eta = 0$, as shown in Fig. 6, where we discover that the gradient-based optimization method is unstable and fragile to the environmental noise with the achieved fidelity varies from 0.2 to 1.0, and only 7.2% are above 0.9. The large distribution variance indicates that the gradient-based

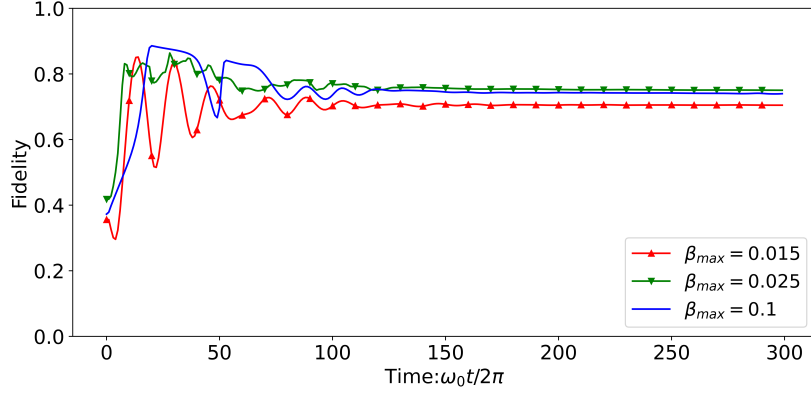


Figure S6. The effect of $\beta_{\max}$ on the prepared state fidelity under large noisy environment with $\kappa = 10^{-3}\omega_0, \eta = 0$.

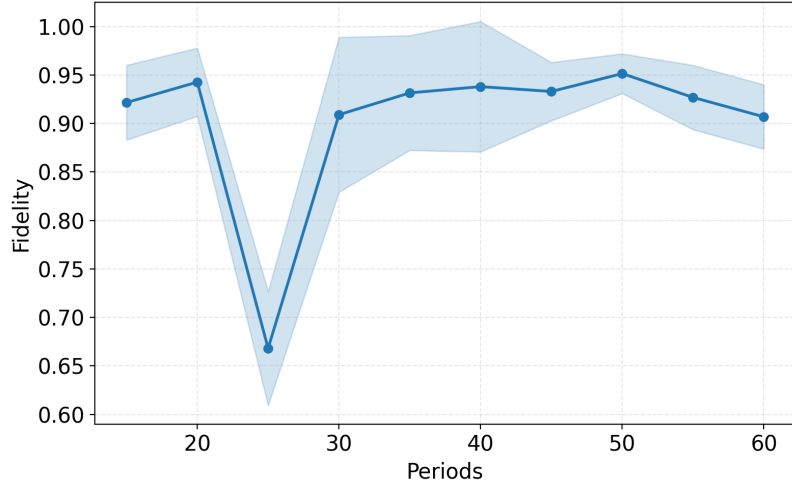


Figure S7. Under a noise-free environment, the statistics for the performance of the gradient-based optimization method for the preparation of a bosonic 4-leg state. For each prefixed evolution step T , 20 trials of the simulations are implemented.

optimization method is quite sensitive to environmental noise and cannot reliably converge to a high-quality control solution for this task. The best fidelity of 0.9901 found with our simulations is shown in Fig. S8, where we can see that the fidelity is not well stabilized.

With the above studies, we would like to clarify the advantage of our reinforcement learning method. Compared to the gradient-based optimization method, the RL agent learns a sequential decision policy that is naturally suited to exploring such a complicated constrained control space, while keeping the driving fields within the prescribed admissible range throughout the evolution. The primary motivation for using reinforcement learning is that the sequential control problem is exactly the type of scenario where RL is good at, but the standard gradient-based or black-box optimization methods become less effective. In particular, we would like to summarize the following advantages of the reinforcement learning method compared with the gradient-based optimization method as follows,

1. Gradient-free optimization: RL does not rely on explicit gradient information, which is beneficial in settings with non-differentiable cost functions or noisy evaluations.
2. Efficient exploration of high-dimensional control landscapes: The agent can learn structured policies that guide exploration beyond local minima typically encountered in conventional optimization.
3. Adaptability and reusability: Once trained, the RL agent can generalize to related tasks or parameter regimes without restarting the optimization from scratch.
4. Compatibility with constrained or experimentally motivated control spaces, where standard methods may require

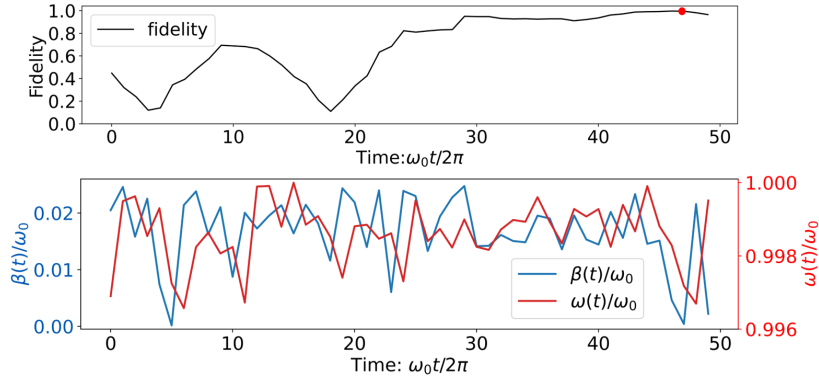


Figure S8. The best performance case with a gradient-based optimization method under a noisy environment with $\kappa = 10^{-7}\omega_0$ and $\eta = 0$ among our simulations shown in Fig.6. The highest fidelity achieved is 0.9901 (marked with the red dot). The black curve represents fidelity, and the blue and red curves represent $\beta(t)$ and $\omega(t)$.

significant modifications.

SUPPLEMENTARY NOTE 4: SCALABILITY CONSIDERING THE PREPARATION OF BOSONIC CAT CODES WITH LARGE q

In this section, we discuss the scalability of the proposed method with respect to the preparation of q -fold rotational bosonic cat states. The scalability of our approach is affected by two main factors: the size of the environment state and the size of the agent neural networks, characterized by the number of parameters. In our simulations, we apply a photon number truncation N_p for the representation of a bosonic state. The dimension of the system density matrix is $N_p \times N_p$, and the corresponding RL state space is $\mathbb{R}^{2N_p^2+1}$, where we use density matrix representation and the extra number 1 represents the index for RL update timestamps. A larger N_p allows the state information to be represented more accurately, but it also increases the computational cost significantly. On the other hand, if N_p is too small, the state information cannot be sufficiently captured, which may degrade the control performance.

From the RL agent's perspective, to learn an effective control strategy, the size of the required RL neural network parameters increases as the RL state space increases. Therefore, the second dominating issue for computational overhead during training is the size of the neural network parameters, which is generally determined by empirical engineering experience and experiments.

To demonstrate the scalability of our work, we further conducted experiments for $q = \{3, 4, 5, 6\}$. The results are summarized in Table. S1. These results indicate that for our work, the scaling is computational manageable for larger q cases, considering that the number of model parameters in currently popular large models is at the level of several billions.

q	N_p	RL State space	RL network Parameters	Obtained state fidelity	Total steps
3	4	$\mathbb{R}^{33}$	75k	0.998	30
4	6	$\mathbb{R}^{73}$	85k	0.998	34
5	7	$\mathbb{R}^{99}$	91k	0.996	30
6	8	$\mathbb{R}^{129}$	99k	0.996	28

Table S1. Scaling of the Hilbert-space truncation and RL state space for different values of q .

Supplementary References

- [1] L. Guo, T. Huang, and L. Du, Engineering bosonic codes with quantum lattice gates, *Communications Physics* **8**, 414 (2025).
- [2] L. Guo, T. Huang, and L. Du, Engineering bosonic codes

* Corresponding author: wei-wei.zhang@nwpu.edu.cn

- with quantum lattice gates, *Communications Physics* **8**, 414 (2025).
- [3] L. Guo, V. Peano, and F. Marquardt, Phase space crystal vibrations: Chiral edge states with preserved time-reversal symmetry, *Phys. Rev. B* **105**, 094301 (2022).
- [4] H. Moritz, T. Stöferle, M. Köhl, and T. Esslinger, Exciting collective oscillations in a trapped 1d gas, *Phys. Rev. Lett.* **91**, 250402 (2003).
- [5] Z. Hadzibabic, S. Stock, B. Battelier, V. Bretin, and J. Dalibard, Interference of an array of independent bose-einstein condensates, *Phys. Rev. Lett.* **93**, 180403 (2004).
- [6] F. Chen, A. Sirois, R. Simmonds, and A. Rimberg, Introduction of a dc bias into a high-q superconducting microwave cavity, *Applied Physics Letters* **98** (2011).
- [7] F. Chen, J. Li, A. Armour, E. Brahim, J. Stettenheim, A. Sirois, R. Simmonds, M. Blencowe, and A. Rimberg, Realization of a single-cooper-pair josephson laser, *Physical Review B-Condensed Matter and Materials Physics* (2014).
- [8] A. Armour, M. Blencowe, E. Brahim, and A. Rimberg, Universal quantum fluctuations of a cavity mode driven by a josephson junction, *Physical review letters* **111**, 247001 (2013).
- [9] V. Gramich, B. Kubala, S. Rohrer, and J. Ankerhold, From coulomb-blockade to nonlinear quantum dynamics in a superconducting circuit with a resonator, *Physical Review Letters* **111**, 247002 (2013).
- [10] A. Krasnok, P. Dhakal, A. Fedorov, P. Frigola, M. Kelly, and S. Kutsaev, Superconducting microwave cavities and qubits for quantum information systems, *Applied Physics Reviews* **11** (2024).
- [11] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, *et al.*, Quantum supremacy using a programmable superconducting processor, *nature* **574**, 505 (2019).
- [12] M. S. Allman, F. Altomare, J. D. Whittaker, K. Cicak, D. Li, A. Sirois, J. Strong, J. D. Teufel, and R. W. Simmonds, rf-squid-mediated coherent tunable coupling between a superconducting phase qubit and a lumped-element resonator, *Physical review letters* **104**, 177004 (2010).
- [13] J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Charge-insensitive qubit design derived from the cooper pair box, *Physical Review A—Atomic, Molecular, and Optical Physics* **76**, 042319 (2007).
- [14] X. C. Kolesnikow, R. W. Bomantara, A. C. Doherty, and A. L. Grimsmo, Gottesman-kitaev-preskill state preparation using periodic driving, *Phys. Rev. Lett.* **132**, 130605 (2024).
- [15] S. Blanes, F. Casas, J.-A. Oteo, and J. Ros, The magnus expansion and some of its applications, *Physics reports* **470**, 151 (2009).
- [16] F. Casas, J. Oteo, and J. Ros, Floquet theory: exponential perturbative treatment, *Journal of Physics A: Mathematical and General* **34**, 3379 (2001).
- [17] A. Eickbusch, V. Sivak, A. Z. Ding, S. S. Elder, S. R. Jha, J. Venkatraman, B. Royer, S. M. Girvin, R. J. Schoelkopf, and M. H. Devoret, Fast universal control of an oscillator with weak dispersive coupling to a qubit, *Nature Physics* **18**, 1464 (2022).
- [18] V. V. Sivak, A. Eickbusch, B. Royer, S. Singh, I. Tsioutsios, S. Ganjam, A. Miano, B. L. Brock, A. Z. Ding, L. Frunzio, *et al.*, Real-time quantum error correction beyond break-even, *Nature* **616**, 50 (2023).
- [19] R. Porotti, V. Peano, and F. Marquardt, Gradient-ascent pulse engineering with feedback, *PRX Quantum* **4**, 030305 (2023).