

Autonomous Floquet Engineering of Bosonic Codes via Reinforcement Learning

Corresponding Author: Professor Wei-Wei Zhang

This manuscript has been transferred from another nature journal. This document only contains information relating to versions considered at Communications Physics.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Wu et al. combine Floquet engineering and reinforcement learning techniques to develop control protocols for the high-fidelity preparation of rotational bosonic cat states of a superconducting resonator, for applications in bosonic quantum codes. The authors optimize the time-dependent amplitude and frequency of the driving field, which simultaneously Floquet-engineers a target effective Hamiltonian and prepares one of its eigenstates with high fidelity. The shape of the driving field is linked to the desired effective Hamiltonian through an ansatz proposed in earlier works. The optimization in the presence of photon loss and decoherence is also investigated, revealing a good degree of robustness of the optimal solutions found and potential improvements via noise-informed retraining of the neural network.

The presentation and the main messages are clear, except for some points detailed below. The problem addressed is surely very interesting and the results (fidelities, robustness and, especially, speed up in the state preparation) seem to bring a remarkable improvement as compared to previous works. More broadly, methods for the efficient preparation of eigenstates of high-frequency Floquet Hamiltonians, alongside the realization of the Hamiltonian itself, are a compelling contribution not only for the preparation of bosonic codes, specifically treated here, but more generally for quantum simulation. From this perspective, I believe the paper may be a good fit for Communications Physics. However, there are also some weaker aspects that make me refrain from a fully positive recommendation at this stage:

1) The setup of the physical problem analysed is quite heavily based on the results of Ref. 24. First of all, this makes it hard to follow some details without jumping back and forth to Ref. 24 (is λ after Eq 5 defined in this paper? and τ ?). More fundamentally, the authors put much emphasis on this paper having the merit of integrating Floquet engineering and reinforcement learning. However, in practice the novelty seems to be entirely in the parameter optimization using reinforcement learning, while the Floquet engineering and quantum control aspects are taken directly from previous work. If so, the authors should shift the emphasis in the text accordingly.

2) It is not clear to me how the optimal driving field/potential actually looks like, in terms of the control operators involved: is it really easy to implement experimentally?

3) There is no comparison with other optimal control methods or any comment about this: it is hard for the reader to gauge the upsides of the proposed reinforcement learning approach, as compared to a straightforward application of black-box optimal control routines from widespread libraries (e.g., QuTiP, ...)? Given that the focus of the paper is on the optimization method, this is an important point, in my opinion.

If the authors can convincingly address at least some of these points, I believe the paper would improve significantly and warrant publication in Communications Physics.

Other minor points:

- Comments about scalability are missing: for instance, is it computationally feasible to apply the same method to optimize q-

fold rotational bosonic cat states with larger q ?

- the notations ϕ and φ seem to be used for the same quantity in Eq. (4) and thereafter. Also, why are φ referred to as "frequencies" after Eq. (4)?

- the axis labelling and units are at times inconsistent: e.g., in Fig. 2(c) and its inset, the quantity $\omega(t)$ is indicated both in units of ω_0 and without units.

- there are overall a few typos throughout the text. I just highlight "origianl" in the legend of Fig. 5(a), which may be missed in a spell-check.

Reviewer #2

(Remarks to the Author)

In the manuscript entitled "Autonomous Floquet Engineering of Bosonic Codes via Reinforcement Learning" by Wu et al., the authors use reinforcement learning (RL) to optimize the drive amplitude and frequency of a periodically driven resonator, aiming to accelerate Floquet engineering of bosonic states. They demonstrate rapid, high-fidelity preparation of a q -legged cat state compared with previous Floquet-based protocols, and they study robustness against photon loss and dephasing. The idea is interesting and may be useful as a numerical optimization layer for practical quantum control.

However, in its current form, the authors' claim that the work "establishes a scalable and experimentally feasible route to fault-tolerant bosonic quantum computation" appears overstated, and several key aspects require clarification and/or additional supporting evidence. I suggest the authors consider the following issues.

1. The numerical results are restricted to a single code family (q -legged cat code) and essentially one instance ($q = 4$). This is not sufficient to justify a title and narrative suggesting broad applicability across bosonic code families. As noted in the Introduction, other bosonic encodings such as binomial and GKP codes have qualitatively different structures from q -legged cat states, so it is not straightforward to generalize the conclusions beyond the single demonstrated example.

- The authors should revise the title/abstract to explicitly reflect the demonstrated codewords, or add results for at least one additional, qualitatively distinct code family.

2. The authors employ a driven single-mode harmonic oscillator Hamiltonian. This is a reasonable effective model for single-cavity state control, but the manuscript frames the result in the context of bosonic encoding for quantum information processing. In most experimental implementations of bosonic encoding in circuit-QED systems, a cavity-qubit architecture is widely used, as mentioned in the main text. In that case, neglecting dispersive shifts and self-Kerr terms may fail to capture standard circuit-QED physics. Refs. 22 - 24 discuss concrete JJ/SQUID-based implementations/modulations, whereas the present manuscript treats the driving potential as directly programmable.

- The authors should either narrow the scope/claims to effective single-mode Floquet state engineering or provide a concrete circuit-level mapping (as in Refs. 22 - 24) and discuss when the reduced single-mode description remains valid.

- If a standard cavity-qubit architecture is assumed, the Hamiltonian and discussion should incorporate (or justify neglecting) dispersive coupling and Kerr-type nonlinearities.

3. The manuscript states β/ω_0 is restricted below 0.025 to "satisfy the RWA", however, a quantitative justification for this threshold is missing. The same cap as Ref. [24] may have been chosen for direct comparison; nevertheless, the conclusions may depend on this value and therefore require independent justification.

- The authors should provide a quantitative RWA/Floquet-validity criterion.

- It would be helpful to show how preparation time and reachable fidelity depend on the upper bound β/ω_0 .

4. A central claim is noise-robust, high-fidelity generation. However, in the extremely high-noise case motivating retraining, the retrained policy appears to reach only moderate fidelities around $F \sim 0.8$. This does not support unqualified "high-fidelity" language unless the authors clearly state the operating conditions under which high fidelity is achieved, and what limits fidelity in the strongly lossy regime.

- The authors should mention what limits the fidelity to $F \sim 0.8$ in the lossy environment ($\kappa/\omega_0 \sim 10^{-3}$).

5. In the lossy case, the retrained $\beta(t)$ appears to saturate at the imposed upper bound $\beta/\omega_0 \sim 0.025$ after $t \sim 160 \times 2\pi/\omega_0$. The authors suggest that this may indicate the drive contributes to correcting photon loss errors, but the manuscript currently provides insufficient evidence to support this interpretation.

- The authors should test how the achievable fidelity depends on the cap $\beta_{\max}/\omega_0$ in the $\kappa/\omega_0 \sim 10^{-3}$ case.

6. To isolate what RL contributes beyond generic numerical optimization, comparison to at least one non-RL baseline (e.g., constrained gradient-based optimization under the same parameter constraints) would strengthen the manuscript.

7. The term "autonomous" can mean either (i) autonomous design of control waveforms (via RL), or (ii) autonomous preparation/stabilization as an attractor/steady state under driven-dissipative evolution. Please clarify which meaning is intended.

If the authors can address the above issues well, I can suggest its publication in Communications Physics.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

I appreciate that the authors revised the manuscript in response to my criticism, including additional analyses, and I acknowledge the efforts made by the authors in addressing all points raised. I find most of the responses convincing and the manuscript has certainly improved. Still, I can only recommend publication if the following point is adequately improved:

- I understand that this paper is not aiming at proposing a specific experimental realization in detail, but, as it stands, the discussion about potential implementations is too superficial. For cold atoms, it is not even explained what the oscillator encoding the bosonic code would be. For superconducting circuits, it is not explained in sufficient depth how the Hamiltonian of the proposed resonator+JJs circuit connects to the drive considered.

Other minor points:

- While it's interesting information, I do not fully understand why the authors report the fidelity statistics for the protocols optimised by the GRAPE-like method and assess those protocols based on the average fidelity. I would rather think that one takes the best result over stochastically sampled initial conditions for the control parameters. Or is this to show that the control values optimized for given values of the system parameters do not perform well if the system parameters are slightly perturbed?

- When introducing Fig. 1 on page 2 right column, the wording about the cavity being confined in a dilution refrigerator somewhat suggests that, in applications of the proposed protocol, the control is synthesized based on data from the experimental system. I recommend re-phrasing this to avoid confusion (or comment more about this), as the current RL algorithm requires the full density matrix, hence full tomography, which would be challenging experimentally.

There are still a few typos. For instance:

- an "and" or "or" missing in "in superconducting qubit [...] neutral atoms [19]" on line 54
- a misplaced $\hat{}$ on p and a missing $\hat{}$ on x in Eq. (2)
- $V(x, \omega t)$ in Eq. (4) and in Eq. (3) cannot be the same object. The notation should be adjusted.
- "frequency" on line 165, a $\dagger$ missing in line 166,
- ...

Reviewer #2

(Remarks to the Author)

The authors have responded adequately to my previous comments. The revised manuscript includes useful additional analyses and clarifications, including the discussion of other bosonic encodings, the comparison with a non-RL baseline, and the clarification of the physical model and approximations.

I am satisfied with the authors' response and the current version of the manuscript, and recommend acceptance after minor textual revisions. Before publication, the authors should carefully proofread the manuscript.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have addressed all my remaining questions and comments satisfactorily. In particular, they have now clarified and elaborated on the proposed experimental implementations and explained the rationale behind the statistical characterization of the gradient-based optimisation approach. I thus recommend publication of the manuscript.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

List of changes and replies to the Referees

June 17, 2026

Re: Autonomous Floquet Engineering of Bosonic Codes via Reinforcement Learning
by Zheping Wu, Lingzhen Guo, Haobin Shi, and Wei-Wei Zhang
Manuscript: COMMSPHYS-25-1918-T

Dear Editor and Reviewers,

Thank you very much for the message regarding the above paper and the referees' reports.

We appreciate referee 1 for the encouraging assessment that "The problem addressed is surely very interesting, and the results ... bring a remarkable improvement as compared to previous works. ... are a compelling contribution not only for the preparation of bosonic codes, ..., but more generally for quantum simulation. ... may be a good fit for Communications Physics. ... If the authors can convincingly address at least some of these points, I believe the paper would improve significantly and warrant publication in Communications Physics." We also appreciate referee 2 for commenting that "the idea is interesting and may be useful as a numerical optimization layer for practical quantum control.", "If the authors can address the above issues well, I can suggest its publication in Communications Physics."

To address the reviewers' questions and concerns, we have performed additional simulations and revised the manuscript accordingly, adding 10 new figures (Fig 2(d,e), Fig. 5(c), Figs. (6, 7, 9-14)) and an appendix section.

Below, we attach a list of changes and our point-to-point response to the two reviewers' comments. In order to facilitate the reading of the response, we mark the referees' comments in blue, and quote the updated manuscript content with grey backgrounds. All the changes are highlighted in red in our revised version of the manuscript.

Best Regards,

Zheping Wu, Lingzhen Guo, Haobin Shi, and Wei-Wei Zhang

List of major changes

- We have corrected a typographical error in Eqs. (1)–(2) and updated the related text to ensure consistency. This correction does not affect the results or conclusions of the manuscript.
- We add the model’s detailed description (lines 153-177) around Eq. (3) in the Section **Model** .
- We add the formula of driving potential, cf. Eqs. (4)-(5) in the revised manuscript (lines 179-190).
- We add new Figs. 2(d-e) showing the time evolution of driving field in the real space.
- We add the a new Appendix A to explain the Noncommutative Fourier transformation (NcFT).
- We add more sentences (378) and a new Fig. 5(c) to study the photon number compensation in our reinforcement learning approach.
- We add a new paragraph (lines 493-512) at the end of Section **Methods** to discuss the possible experimental implementations of our approach.
- We add a new Appendix B to discuss the details on rotating wave approximation (RWA) and the dispersive dynamics.
- We add a new paragraph (lines 379-406), a new Fig. 6 and a new Appendix C for benchmarking against a conventional gradient-based optimal-control method, to better clarify the advantages of the proposed reinforcement-learning approach relative to standard optimal-control routines.
- We add a new paragraph (lines 407-421) and a new Fig. 7 to study the preparation of a superposition state of binomial codes with our reinforcement learning approach.
- We add a new Appendix D to study the scalability of our approach, considering large q cat state preparation.

Response COMMSPHYS-25-1918-T

We appreciate the referees' insightful comments and suggestions. We have modified our manuscript in accordance with their reports. Below, we address the referees' concerns point-by-point in detail. In order to facilitate the reading of the response, the referees' comments are marked in blue, and our updated manuscript content is quoted with blue backgrounds. All the changes are highlighted in red in the new version of the manuscript.

Response to Referee A

Comment 1

Wu et al. combine Floquet engineering and reinforcement learning techniques to develop control protocols for the high-fidelity preparation of rotational bosonic cat states of a superconducting resonator, for applications in bosonic quantum codes. The authors optimize the time-dependent amplitude and frequency of the driving field, which simultaneously Floquet-engineers a target effective Hamiltonian and prepares one of its eigenstates with high fidelity. The shape of the driving field is linked to the desired effective Hamiltonian through an ansatz proposed in earlier works. The optimization in the presence of photon loss and decoherence is also investigated, revealing a good degree of robustness of the optimal solutions found and potential improvements via noise-informed retraining of the neural network.

The presentation and the main messages are clear, except for some points detailed below. The problem addressed is surely very interesting and the results (fidelities, robustness and, especially, speed up in the state preparation) seem to bring a remarkable improvement as compared to previous works. More broadly, methods for the efficient preparation of eigenstates of high-frequency Floquet Hamiltonians, alongside the realization of the Hamiltonian itself, are a compelling contribution not only for the preparation of bosonic codes, specifically treated here, but more generally for quantum simulation. From this perspective,

I believe the paper may be a good fit for Communications Physics. However, there are also some weaker aspects that make me refrain from a fully positive recommendation at this stage:

Response

We thank this referee for the thorough assessment of our work. We are encouraged by the referee's recognition of the impact, importance, and novelty of our work. We understand the reviewer's reservations regarding certain aspects of the manuscript and have carefully addressed all points raised by the referee. We believe these revisions adequately address the reviewer's comments and strengthen the manuscript.

Comment 2

The setup of the physical problem analysed is quite heavily based on the results of Ref. 24. First of all, this makes it hard to follow some details without jumping back and forth to Ref. 24 (is λ after Eq 5 defined in this paper and τ ?).

Response

We thank this referee for the comments and apologize for the missing definition of symbols.

We have added a more comprehensive brief description of the underlying model and ensured that all symbols in our manuscript are well-defined. The revision is at the three paragraphs starting from line 153 on page 3. The λ is introduced below Eq.(3) of our revised manuscript, as an effective dimensionless Planck constant. And τ is used and defined in our Appendix A as $\tau = \omega t$. The revised content is as below.

The 1st related revision is on page 3, line 153.

In the rotating frame of cavity frequency, the Hamiltonian (3) is transformed to $\hat{H}(t) = \hat{O}(t)\hat{H}(t)\hat{O}(t) + i\lambda\hat{O}(t)\hat{O}^\dagger(t)$ with $\hat{O}(t) \equiv e^{i\hat{a}^\dagger\hat{a}\omega_0 t}$, **where λ is the scaled dimensionless Planck constant ($\hbar$ divided by the characteristic angular momentum of system).**

The 2nd related revision is added on page 10, 587.

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. [22]. Here, $\tau = \omega t$ is used for the sake of formulaic simplicity.

Comment 3

More fundamentally, the authors put much emphasis on this paper having the merit of integrating Floquet engineering and reinforcement learning. However, in practice the novelty seems to be entirely in the parameter optimization using reinforcement learning, while the Floquet engineering and quantum control aspects are taken directly from previous work. If so, the authors should shift the emphasis in the text accordingly.

Response

We thank this referee for the comment. Indeed, the final goal of our work is to optimize the driving field parameter. In our work, we successfully formulated our quantum bosonic code-state preparation task as a reinforcement learning problem. Remarkably, a fusion model of the quantum Floquet engineering and reinforcement learning was successfully designed and implemented, and satisfactory results were achieved.

We would like to clarify that the fusion of two methods is not merely a straightforward combination. It includes the construction of suitable agent state representations, agent action representations, the design of reward and target function design, which together make the problem tractable for reinforcement learning. On top of this formulation, we merge the TD3 algorithm to learn an effective control policy for our task. The reinforcement learning enables adaptive and efficient exploration of high-dimensional control landscapes that are otherwise challenging to optimize with traditional methods. Together, these elements make the control optimization problem amenable to a reinforcement-learning-based solution. Therefore, we would like to clarify that our main contribution is not a direct combination of two existing methods, but rather the development of a reinforcement-learning-based control framework tailored to the existing Floquet-engineering approach.

To address the reviewers' concerns, we revised the manuscript to more clearly describe our contributions to the reinforcement learning-based quantum optimization framework. The revisions are listed as below.

Abstract:

Here, we present a Floquet engineering framework in which reinforcement learning is used to realize the autonomous design of the control waveforms for bosonic code preparation, yielding a general, efficient and noise-resilient strategy.

The revised content has been added on page 3 line 153.

In this work, we introduce a reinforcement learning-assisted quantum-control wave-design strategy, via engineering a temporal periodic driving field (Floquet engineering), to achieve efficient quantum state preparation.

The revised content has been added on page 9 line 546.

Our results highlight the prospects for the deep integration of artificial intelligence and quantum control as a powerful paradigm for the development of fault-tolerant bosonic quantum computation.

Comment 4

It is not clear to me how the optimal driving field/potential actually looks like, in terms of the control operators involved: is it really easy to implement experimentally?

Response

We thank this referee for the suggestions.

We have added two subfigures showing representative optimized driving fields as shown in Fig. 2(c),(d) of the revised manuscript. For the referee's convenience, we reproduce them here in this reply as Fig. S 1. The optimized driving field obtained using our method is shown in Fig. S 1(a). As a comparison, we also added the previously proposed adiabatic-

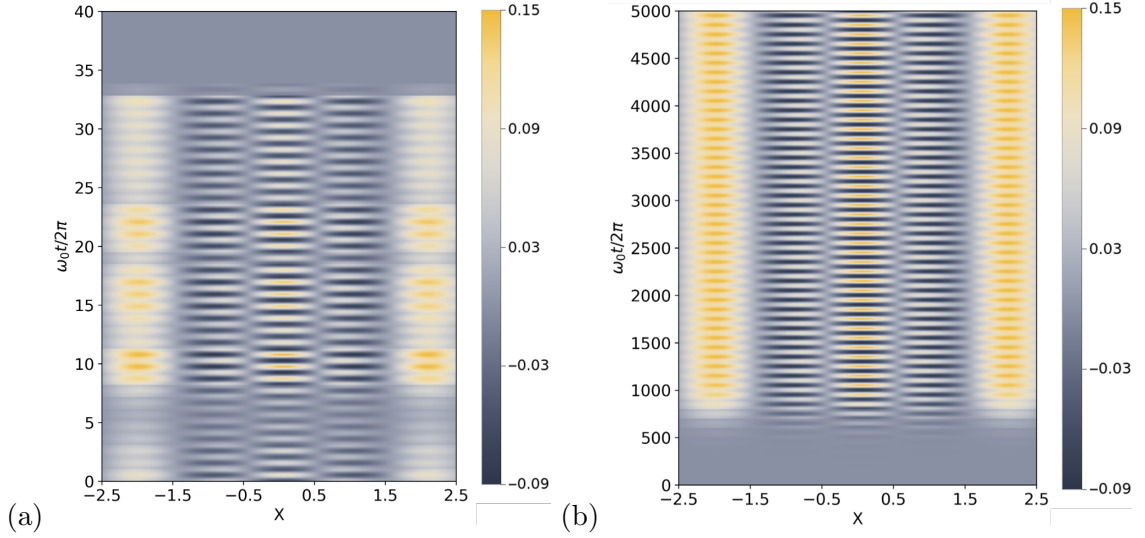


Fig. S 1: driving field with (a) our reinforcement learning approach (b) the original adiabatic ramp in Ref. [1]

ramp protocol driving field in Ref. [1] as shown in Fig. S 1(b). These results demonstrate that the obtained protocols exhibit a similar structured modulation as the previously suggested heuristic Sigmoid function used in adiabatic ramping method. The adiabatic ramp increases the driving strength slowly during roughly the first 700 periods and then keeps it nearly constant. In contrast, our method turns on the driving field at the beginning, the strength of the driving field varies until the system reaches the target state in 34 periods, afterwards the driving is essentially switched off. It demonstrates the much higher efficiency of our method, which significantly reduces the preparation time compared with the conventional adiabatic-ramp approach.

Regarding experimental feasibility, our proposal is able to be implemented in superconducting circuit with Josephson junctions (JJs) or ultra-cold atom in optical lattice with beam lasers. In our method, the driving field 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. The number of the required cosine lattice potentials corresponds to the number of Josephson junction (JJ) in the superconducting circuits or the number of laser beams in cold atoms systems. As shown in Fig. S 2 in this reply, the 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.

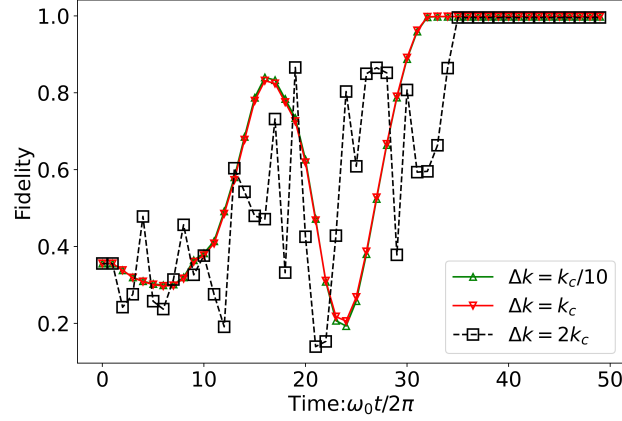


Fig. S 2: Time evolution of the prepared-state fidelity for approximate driving potentials with discretized wavenumber spacings $\Delta k = k_c/10$ (green, final fidelity 0.9982), $\Delta k = k_c$ (red, 0.9984), and $\Delta k = 2k_c$ (black, 0.9961).

Remarkably, even with the coarsest discretization of five JJs in the superconducting-circuit, our approach still achieves a fidelity of 0.9961, showing a good experimental feasibility of our work.

We have incorporated these sections into our revised manuscript and included further discussion of the experimental realization.

First, we have added the content about “the dynamical driving potential” on page 5, line 287.

The revised content has been added on page 5 line 287.

The dynamical driving potential in real space obtained with our reinforcement learning approach is shown in Fig. 2 (d), where the pattern is similar as that obtained with the adiabatic ramp method, except the driving strength varies as shown in Fig. 2 (e).

Second, we have added the experimental realizations discussions on page 9 line 493.

The revised content has been added on page 9 line 493.

Experimental realizations. Our RL approach provides a general constructive method to design the driving potential, up to leading order in the Floquet-Magnus expansion, generating bosonic states in a single-mode cavity. The designed driving field can be implemented in the state-of-the-art experimental platforms. Instead of realizing the driving field in real-space $V(x, t)$, an equivalent route is to directly use Eq. 4 by replacing the integral with sum of a finite number of cosine lattice potentials. In the Appendix B, we demonstrate that, even with 5 such potentials, the Fidelity of the prepared 4-leg cat state can reach 99.61%. For the cold atom platform, the cosine lattice can be realized with laser beams intersecting at various angles [37-39]. In superconducting circuit platform, the cosine driving fields can be directly realized with Josephson junction circuit loops together with standard Gaussian operations [22,40,41].

Third, the more details are added in Appendix B.

Comment 5

There is no comparison with other optimal control methods or any comment about this: it is hard for the reader to gauge the upsides of the proposed reinforcement learning approach, as compared to a straightforward application of black-box optimal control routines from widespread libraries (e.g., QuTiP, ...)? Given that the focus of the paper is on the optimization method, this is an important point, in my opinion.

Response

We thank this referee for the suggestions. We agree that it is essential to benchmark our reinforcement-learning (RL) approach against a more conventional optimal-control baseline. For comparison, we have added a demonstration of a gradient-based optimization method that uses our reinforcement-inspired objective function. The gradient-based optimization requires a prefixed number of evolution steps to define the optimization spaces with the details shown in our Pseudocode of Alg. 1 in this reply.

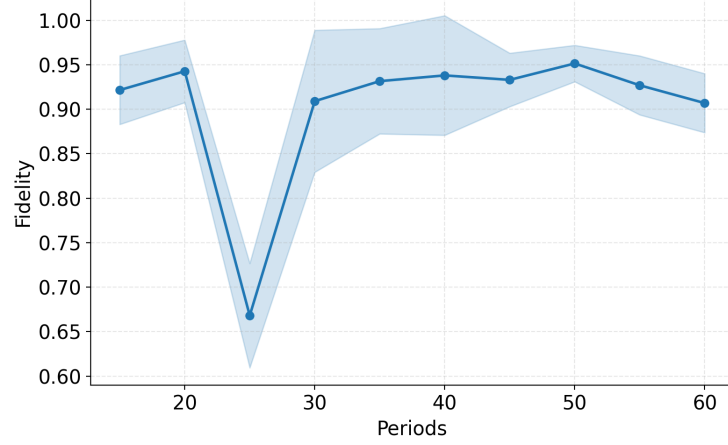


Fig. S 3: Under a noise-free environment, the statistics for the performance of the gradient-based optimization method for the preparation of a bosonic 4-leg state. For each prefixed evolution step T , 20 trials of the simulations are implemented.

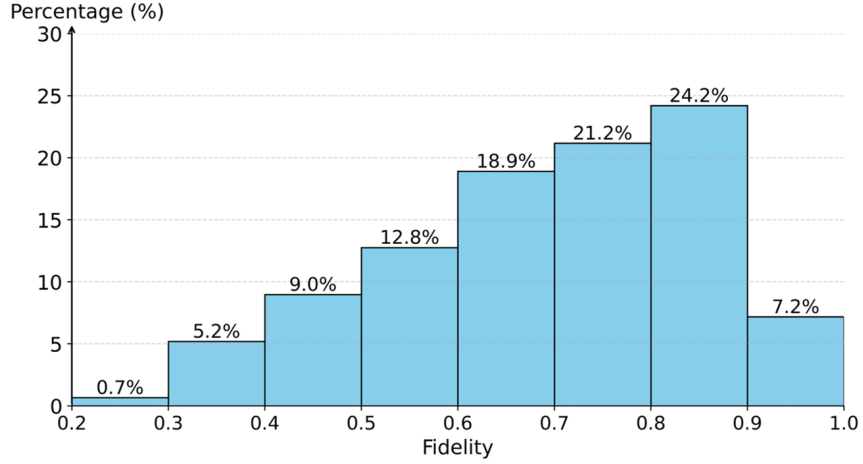


Fig. S 4: The statistics of the prepared state fidelity obtained with 1265 trials of the gradient-based optimization method under a noisy environment with $\kappa = 10^{-7}\omega_0$, $\eta = 0$.

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. Here, 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. S 3 of this reply. 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.

Algorithm 1: Gradient-based optimization method

```

Initialize periods 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  $K_{\text{max}}$  do
    Propagate  $\psi$  under  $x$ 
    Compute fidelity trajectory  $\{F_1, \dots, F_T\}$ 
    Let  $\bar{F}$  and  $F_{\text{min}}$  be the mean and minimum of the last  $M$  fidelities
    Compute loss  $\mathcal{L}(x) = (1 - \bar{F})^2 + (1 - F_{\text{min}})^2$ 
    Update the control parameters using gradient-based optimization
    if converged then
        break
    end
end
Return  $x^*$ 

```

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. S 4 of this reply, where we discover that the gradient-based optimization method is extremely 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.

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,

- Unlike gradient-based optimization methods, RL learns control policies directly from interaction with the environment and reward feedback, without requiring an accurate differentiable model of the system. This makes RL particularly advantageous in experimental settings with noise, or limited system information.
- The RL agent can learn structured policies that guide exploration beyond local minima typically encountered in conventional optimization.
- Our framework is highly reusable for similar tasks, allowing efficient optimization to be implemented quickly and yielding strong performance.
- Compatibility with constrained or experimentally motivated control spaces, where the gradient optimization methods may require significant modifications.

We have added these results on page 6 line *379-406* at the end of the Results section, and we add additional details in Appendix C.

Comment 6

If the authors can convincingly address at least some of these points, I believe the paper would improve significantly and warrant publication in Communications Physics.

Response

We thank the referee for this constructive and encouraging assessment, and the potential recommendation of the publication. We have carefully addressed all of the aforementioned points and believe that the revised manuscript has been significantly improved accordingly.

Comment 7

Other minor points: Comments about scalability are missing: for instance, is it computationally feasible to apply the same method to optimize q -fold rotational bosonic cat states with larger q ?

Response

We thank this referee for the suggestions. In response, we implemented extensive numerical simulations on the study of large q and added a discussion on scalability in Appendix D.

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 policy network parameter 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.

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. S I: Scaling of the Hilbert-space truncation and RL state space for different values of q .

To demonstrate the scalability of our work, we further conducted numerical experiments for $q = \{3, 4, 5, 6\}$. The results are summarized in Table. S I of this reply. These results indicate that for our work, the scaling is computational managable for larger q cases, considering that

the number of model parameters in currently popular large models is at the level of several billions.

We have added this part in the Appendix D of our revised manuscript.

Comment 8

- the notations ϕ and φ seem to be used for the same quantity in Eq. (4) and thereafter. Also, why are φ referred to as “frequencies” after Eq. (4)?

Response

We apologize for the inconsistency in notation. In the revised version, we have unified the notation and consistently use φ for the phase parameter. We have corrected the wording accordingly throughout the manuscript. The revised version is as below.

$$V(x, \omega t) = \int_0^{+\infty} A(k, \omega t) \cos[kx + \varphi(k, \omega t)] dk, \quad (4)$$

Comment 9

The axis labelling and units are at times inconsistent: e.g., in Fig. 2(c) and its inset, the quantity $\omega(t)$ is indicated both in units of ω_0 and without units.

Response

We thank this referee for pointing out this inconsistency and apologize for it.

In the revised manuscript, we have corrected this issue by consistently expressing $\omega(t)$ in units of $\omega_0 t / (2\pi)$ throughout the figure (including the inset), and we have updated the axis labels accordingly. We have also carefully checked all figures in the manuscript to ensure consistent notation and unit conventions.

Comment 10

There are overall a few typos throughout the text. I just highlight "origianl" in the legend of Fig. 5(a), which may be missed in a spell-check.

Response

We thank this referee for the comment. We have corrected the spelling of "origianl" to "original" in the legend of Fig. 5(a). We have carefully proofread the entire manuscript and corrected several minor typographical errors to improve clarity and presentation.

Response to Referee B

Comment 1

In the manuscript entitled "Autonomous Floquet Engineering of Bosonic Codes via Reinforcement Learning" by Wu et al., the authors use reinforcement learning (RL) to optimize the drive amplitude and frequency of a periodically driven resonator, aiming to accelerate Floquet engineering of bosonic states. They demonstrate rapid, high-fidelity preparation of a q-legged cat state compared with previous Floquet-based protocols, and they study robustness against photon loss and dephasing. The idea is interesting and may be useful as a numerical optimization layer for practical quantum control.

Response

We thank this referee for the thorough assessment of our work and for pointing out the novelty and impact of our work.

Comment 2

However, in its current form, the authors' claim that the work "establishes a scalable and experimentally feasible route to fault-tolerant bosonic quantum computation" appears overstated, and several key aspects require clarification and/or additional supporting evidence. I suggest the authors consider the following issues.

Response

We thank this referee for the comments, and we have rephrased our work to provide a proof-of-principle demonstration of reinforcement learning-based optimization for Floquet engineering of bosonic states, rather than a complete route to fault-tolerant quantum computation. We appreciate the reviewer's suggestion and have carefully addressed the specific points raised below.

Comment 3

The numerical results are restricted to a single code family (q-legged cat code) and essentially one instance ($q = 4$). This is not sufficient to justify a title and narrative suggesting broad applicability across bosonic code families. As noted in the Introduction, other bosonic encodings such as binomial and GKP codes have qualitatively different structures from q-legged cat states, so it is not straightforward to generalize the conclusions beyond the single demonstrated example.

–The authors should revise the title/abstract to explicitly reflect the demonstrated codewords, or add results for at least one additional, qualitatively distinct code family.

Response

We thank this referee for the suggestions. To demonstrate the universality of our approach, we apply our method for the preparation of a superposition state of binomial code [2, 3], where the logical $|\bar{0}_b\rangle$ and $|\bar{1}_b\rangle$ are defined as following,

$$\begin{cases} |\bar{0}_b\rangle \equiv \frac{1}{2}(|0\rangle + \sqrt{3}|4\rangle) \\ |\bar{1}_b\rangle \equiv \frac{1}{2}(\sqrt{3}|2\rangle + |6\rangle), \end{cases} \quad (1)$$

where $|n\rangle$ is the Fock state with n photons. The target superposition state is $|\psi\rangle_{\text{tar}} = \frac{1}{\sqrt{2}}(|\bar{0}_b\rangle + |\bar{1}_b\rangle)$ and the target Hamiltonian is set as $H_{\text{tar}} = -|\psi\rangle_{\text{tar}}\langle\psi|$. As shown in Fig. S 5 of this reply, we see that with around 105 steps of evolution, our reinforcement learning approach prepares the target state while the adiabatic ramp requires 4000 steps [2]. And

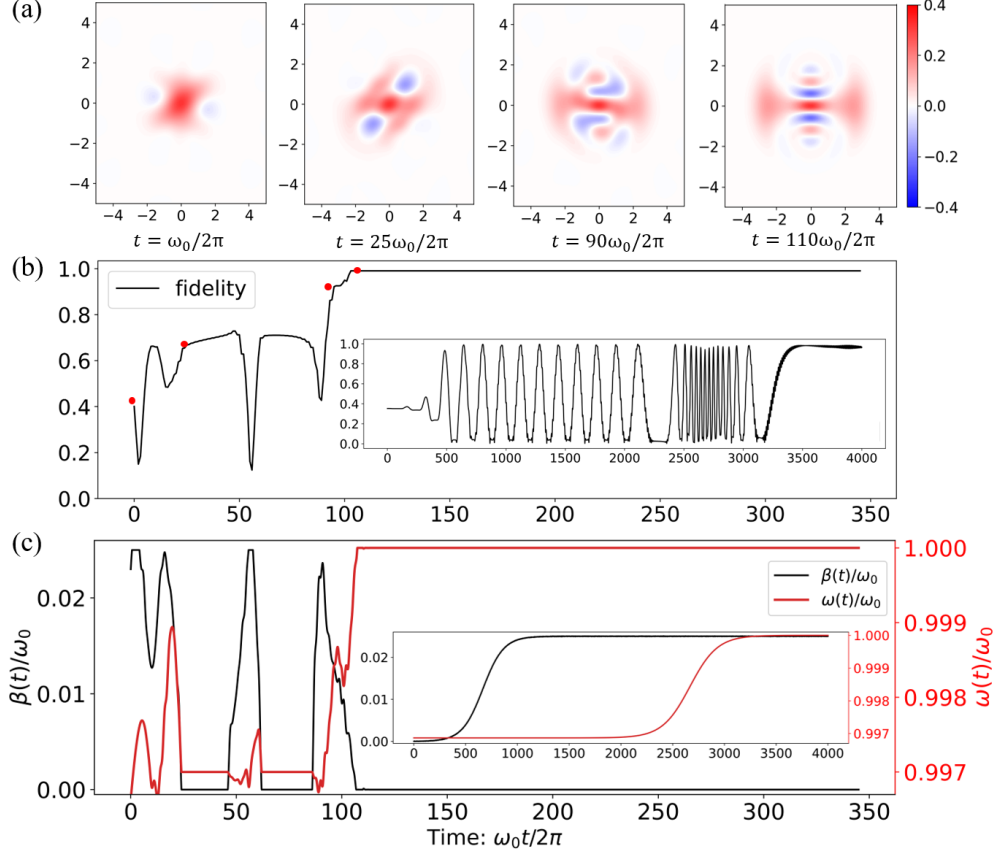


Fig. S 5: Preparation of a superposition state of binomial code $|\psi\rangle_{\text{tar}} = \frac{1}{\sqrt{2}}(|\bar{0}_b\rangle + |\bar{1}_b\rangle)$ with our RL approach. (a) Snapshots of Husimi Q-functions of prepared states at four different time moments marked by the red dots in (b). (b) Stroboscopic time evolution of the prepared state fidelity with the final fidelity remaining at 0.991. Inset: the adiabatic ramp method in Ref. [2] with the final fidelity around 0.969. (c) The learned driving amplitude $\beta(t)$ (black curve) and the driving frequency $\omega(t)$ (red curve) with the preparation time $105 \times 2\pi/\omega_0$. Inset: reproduced driving amplitude (black curve) and driving frequency (red curve) from the adiabatic ramp method with total preparation time $4000 \times 2\pi/\omega_0$ in Ref. [2].

the fidelity of the prepared state reaches 0.991 while with the adiabatic ramp method, the prepared state has a fidelity of 0.969.

Our results show that although different code families require tailored target states and control parameterizations, the underlying optimization framework remains unchanged. The binomial code example thus provides further evidence of the flexibility and generality of the

proposed method.

In addition, we have also conducted extensive simulations for the preparation of cat states with different values of q . Our simulations demonstrate the efficient preparation of these states with the details shown in Table. S II of this reply, which shows that for various q leg cat state, our approach only requires around 30 steps of preparation evolution.

We added these results on page 6, lines 407 to 421, and added Fig.7 in our revised manuscript. We also added more details in Appendix D.

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. S II: Scaling of the Hilbert-space truncation and RL state space for different values of q .

Comment 4

The authors employ a driven single-mode harmonic oscillator Hamiltonian. This is a reasonable effective model for single-cavity state control, but the manuscript frames the result in the context of bosonic encoding for quantum information processing. In most experimental implementations of bosonic encoding in circuit-QED systems, a cavity-qubit architecture is widely used, as mentioned in the main text. In that case, neglecting dispersive shifts and self-Kerr terms may fails to capture standard circuit-QED physics. Refs. 22 - 24 discuss concrete JJ/SQUID-based implementations/modulations, whereas the present manuscript treats the driving potential as directly programmable.

- The authors should either narrow the scope/claims to effective single-mode Floquet state engineering or provide a concrete circuit-level mapping (as in Refs. 22 - 24) and discuss when the reduced single-mode description remains valid.
- If a standard cavity-qubit architecture is assumed, the Hamiltonian and discussion should incorporate (or justify neglecting) dispersive coupling and Kerr-type

nonlinearities.

Response

We thank this referee for the comments and suggestions.

We have narrowed the claims to effective single-mode Floquet state engineering.

Furthermore, as the referee said, the cavity-qubit model ignores the higher-order nonlinear terms of the JJ potential and thus does not utilize the full nonlinearity as a quantum resource. Even worse, the residual nonlinearity causes errors as a destructive channel and could become a dominant source of error for engineering large states. In contrast, we use the JJ/SQUID-cavity scheme which directly utilizes the entire nonlinear JJ cosine potential by harnessing all the high-order nonlinear terms as a valuable quantum resource, which is exactly the advantage of our approach and the key point for the engineering arbitrary quantum Hamiltonian in Ref. [1]

In our revised manuscript, we have adjusted our claims accordingly, as listed below.

The revised content has been added on page 2 line 133.

As shown in Fig. 1, we sketch the preparation process of the bosonic state that is realized in the single-mode cavity bosonic qubit with a circuit chip confined in a Dilution refrigerator.

The revised content has been added on page 3 line 207.

As shown in Fig. 1, the single superconducting cavity evolves from the initial vacuum state under the driven Hamiltonian given by Eq. (3).

The revised content has been added on page 9 line 497.

line 497 ... generating bosonic states in a single-mode cavity.

Comment 5

The manuscript states β/ω_0 is restricted below 0.025 to “satisfy the RWA”, however, a quantitative justification for this threshold is missing. The same

cap as Ref. [24] may have been chosen for direct comparison; nevertheless, the conclusions may depend on this value and therefore require independent justification.

- The authors should provide a quantitative RWA/Floquet–validity criterion.
- It would be helpful to show how preparation time and reachable fidelity depend on the upper bound β/ω_0 .

Response

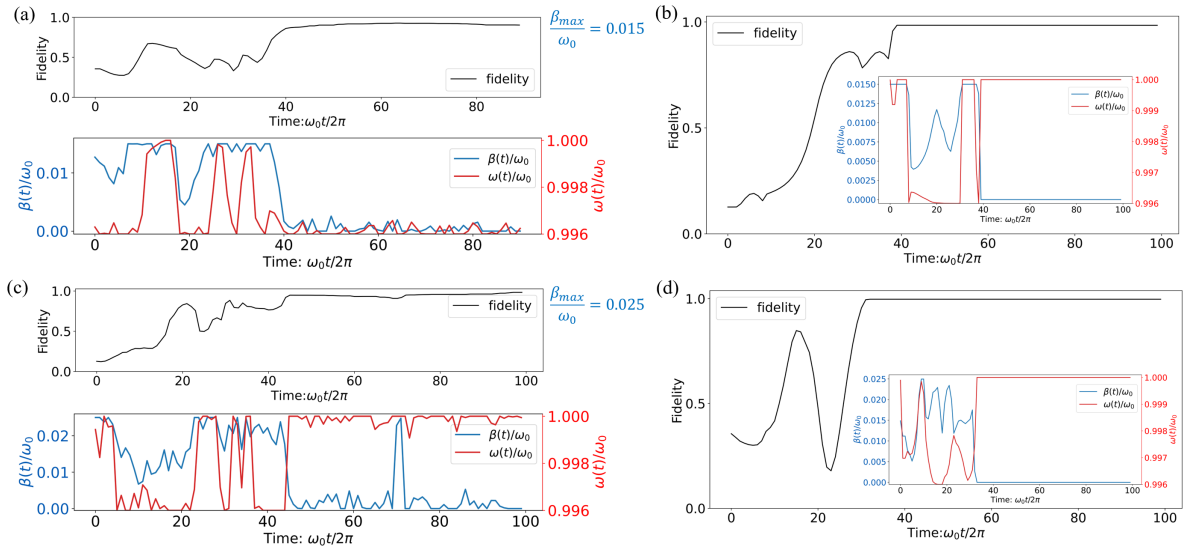


Fig. S 6: The training (a, c) and testing (b, d) trajectories for (a, b) $\beta_{\max}/\omega_0 = 0.015$ and (c, d) $\beta_{\max}/\omega_0 = 0.025$. The black, blue, and red curves represent the fidelity, $\beta(t)$, and $\omega(t)$ respectively.

We thank this referee for this comment.

In the revised manuscript, we have added a discussion clarifying the validity regime of the RWA/Floquet description. In particular, we now relate the condition $\beta/\omega_0 \ll 1$ to the suppression of fast-oscillating counter-rotating terms, and provide a quantitative estimate of their contribution in our parameter regime.

We have included additional numerical results with $\beta/\omega_0 = 0.015, 0.025, 0.1$ respectively as shown in Fig. S 6 and Fig. S 7 of this reply. Our results indicate that with $\beta_{\max} < 0.025$, the smaller the $\beta_{\max}$ is, the longer time our reinforcement learning approach needed for the state preparations. Our simulations show that, with $\beta_{\max} = 0.015$, the required preparation

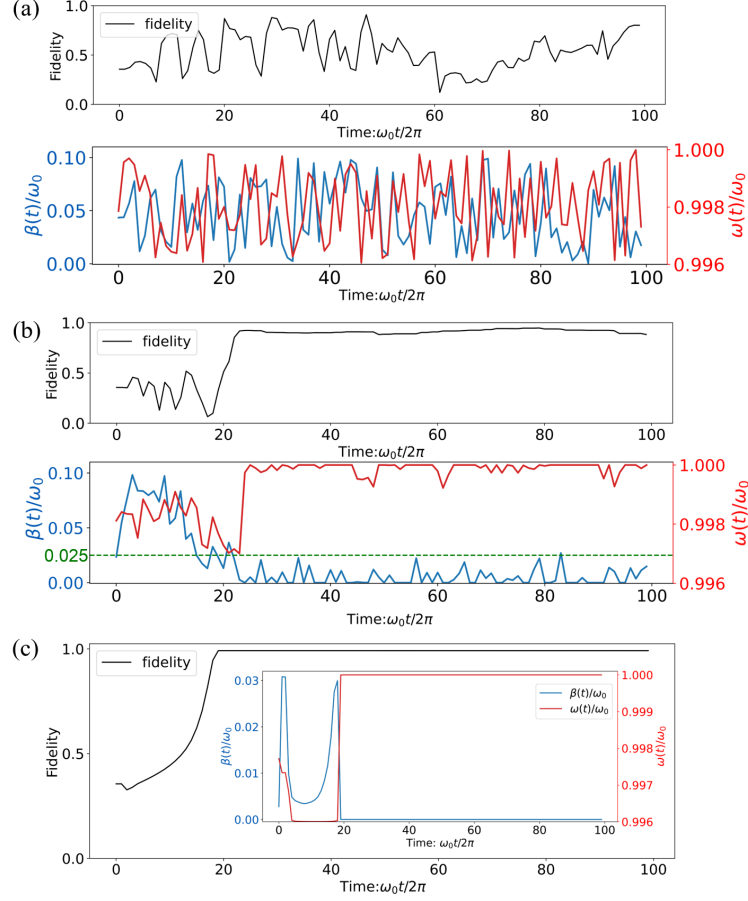


Fig. S 7: (a) The evolution of prepared state fidelity and the corresponding training trajectory from the 1st episode, (b) The evolution of prepared state fidelity and the corresponding training trajectory from the 156th episode, and (c) The evolution of prepared state fidelity and the corresponding 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.

time is 40 steps, compared with a 34 steps required with $\beta_{\max} = 0.025$.

While with a large value of $\beta_{\max}/\omega_0 = 0.1$, which may affect the validity of RWA, our RL approach finds the optimized $\beta_{\max}/\omega_0$ fluctuation range automatically converges to the value of $\beta_{\max}/\omega_0 = 0.025$.

In the early stage of our training process of a total of 300 episodes, β/ω_0 explores the whole range of $[0, 0.1]$ as shown by the case of the 1st training episode in Fig. S 7(a) of this reply. As training progresses, our RL approach learns to concentrate the exploration into a narrower

interval. As shown in Fig. 7 (b) of this reply, 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 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. 7 (c) (the 300th episode) of this reply. 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).

Our simulations indicate that the value of $\beta/\omega_0 = 0.025$ is a reasonably good choice for our case. We have added these results to our revised manuscript.

We have revised the manuscript in page 5 line 315 accordingly to incorporate these results and discussion. The content is shown in the Appendix B.4.

The revised content has been added on page 5 line 315.

We would like to point out that setting the maximum value of $\beta/\omega_0 = 0.025$ is approximately an optimal choice according to our numerical simulations. For setting the maximum value of β/ω_0 smaller than 0.025, it needs longer preparation time. While for setting the maximum value of β/ω_0 larger than 0.025, the RL optimized control sequence of β/ω_0 eventually fluctuates in the range between 0 and 0.025 (see more results and discuss in Appendix B.4).

Comment 6

A central claim is noise-robust, high-fidelity generation. However, in the extremely high-noise case motivating retraining, the retrained policy appears to reach only moderate fidelities around $F \sim 0.8$. This does not support unqualified “high-fidelity” language unless the authors clearly state the operating conditions under which high fidelity is achieved, and what limits fidelity in the strongly lossy regime.

- The authors should mention what limits the fidelity to $F \sim 0.8$ in the lossy environment ($\kappa/\omega_0 \sim 10^{-3}$).

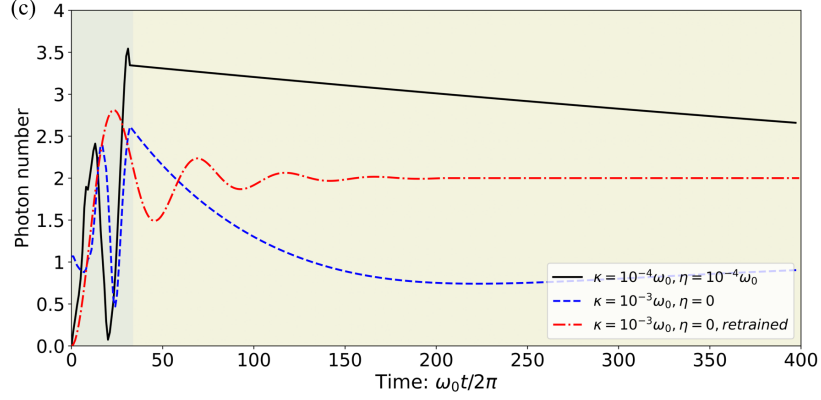


Fig. S 8: The photon number evolution during the state preparation dynamics under various photon loss noise environments with our original trained network and the retrained network.

Response

We thank this referee for the comments. We have revised the content as below.

In the revised manuscript, we have clarified the operating conditions of high fidelities are achieved, and we have moderated the corresponding statements to avoid overgeneralization. In particular, we now explicitly distinguish between the weak-noise regime and the reasonable strong-noise, where fidelities close to unity can be obtained, and the strong-loss regime (e.g., $\kappa/\omega_0 \sim 10^{-3}$), where the achievable fidelity is fundamentally limited.

The reduction of fidelity to values around $F \sim 0.8$ in the lossy environment can be primarily attributed to the photon loss rate. In this regime, photon loss events during the preparation process irreversibly degrade the prepared state, and this effect can only be partially mitigated by our current control optimization. As shown in Fig. S 8 of this reply, retraining our reinforcement learning can stabilize the photon number at a higher level compared with the one without retraining in the large noise environment of $\kappa/\omega_0 = 10^{-3}$. It would require an additional complicated higher-order Floquet-Magnus expansions of driving field components [4], which will be our future work.

We have added a discussion in the revised manuscript on page 6 line 361 to clarify these points and to more precisely characterize the achievable performance across different noise

regimes. Here, for the referee's convenience, we reproduce them as below.

Abstract

...while maintaining high-fidelity state generation even under reasonably strong dissipative and dephasing noise.

We revise the content on page 6 line 344

...obtain an improved fidelity value around 0.8.

The revised content has been added on page 6 line 362.

In Fig. 5 (c), we investigate the dynamics of average photon numbers during the state preparation under the large noise environments with parameters $\kappa = 10^{-3}\omega_0, \eta = 0$, using the original trained and retrained reinforcement learning networks respectively. We clearly see that under large noise environment, our retrained neural networks is able to compensate the photon loss dynamically to a much higher level, but is not able to fully remove its effect to restore the photon number to the expected level as the low noise case such as $\kappa = 10^{-7}\omega_0, \eta = 0$. We also witness that the photon numbers during the evolution dynamics in Fig. 5 (c) has the similar same trend as the state fidelity in Fig. 4. Under the large noise environments, it will require other complicated control process that is beyond the scope of this work.

Comment 7

In the lossy case, the retrained $\beta(t)$ appears to saturate at the imposed upper bound $\beta/\omega_0 \sim 0.025$ after $t \sim 160 \times 2\pi/\omega_0$. The authors suggest that this may indicate the drive contributes to correcting photon loss errors, but the manuscript currently provides insufficient evidence to support this interpretation.

– The authors should test how the achievable fidelity depends on the cap β_{max}/ω_0 in the $\kappa/\omega_0 \sim 10^{-3}$ case.

Response

We thank this referee for the comments and suggestions. We have discussed the part of

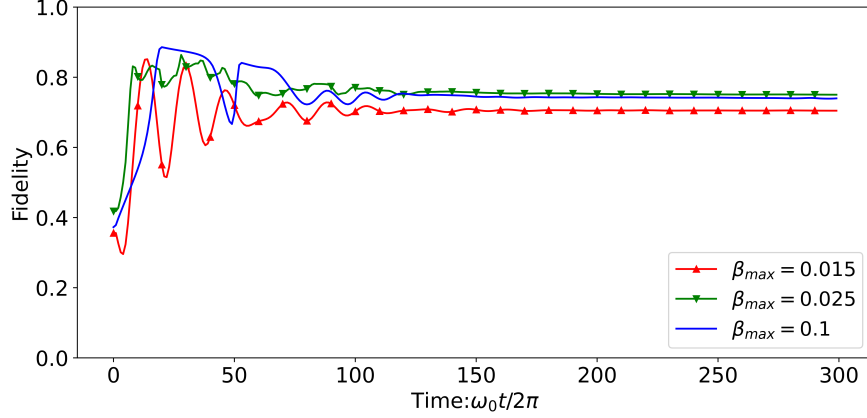


Fig. S 9: The fidelity evolution of our prepared state using a retrained RL model with various $\beta_{\max}$ under the noisy environment of $\kappa = 10^{-3}\omega_0, \eta = 0$.

“drive’s contribution on the photon loss” in previous comment 5 of this referee. We will answer the effect of $\beta_{\max}$ on the achievable fidelity in large noise environment.

We further investigate the effect of $\beta_{\max}$ on the fidelity of the prepared state achieved with our retrained RL model in a high-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. S 9 of this reply. These numerical studies also support our choice of $\beta_{\max} = 0.025\omega_0$ as a good set.

We have revised the manuscript and add the details in Appendix B.4 to clarify our statement.

Comment 8

To isolate what RL contributes beyond generic numerical optimization, comparison to at least one non-RL baseline (e.g., constrained gradient-based optimization under the same parameter constraints) would strengthen the manuscript.

Response

We thank this referee for the suggestions. We agree that it is essential to benchmark our reinforcement-learning (RL) approach against a more conventional optimal-control baseline.

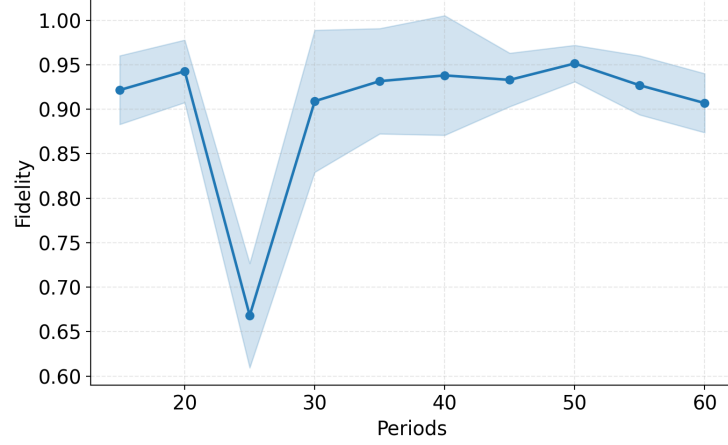


Fig. S 10: Under a noise-free environment, the statistics for the performance of the gradient-based optimization method for the preparation of a bosonic 4-leg state. For each prefixed evolution step T , 20 trials of the simulations are implemented.

For comparison, we have added a demonstration of a gradient-based optimization method that uses our reinforcement-inspired objective function. The gradient-based optimization requires a prefixed number of evolution steps to define the optimization spaces with the details shown in our Pseudocode of Alg. 2 in this reply.

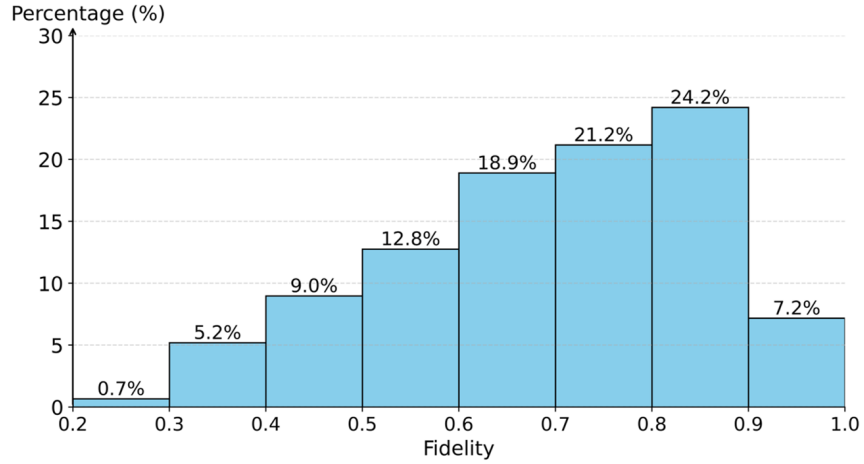


Fig. S 11: The statistics of the prepared state fidelity obtained with 1265 trials of the gradient-based optimization method under a noisy environment with $\kappa = 10^{-7}\omega_0, \eta = 0$.

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. Here,

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. S 10 of this reply. 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.

Algorithm 2: Gradient-based optimization method

```

Initialize periods 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  $K_{\text{max}}$  do
    Propagate  $\psi$  under  $x$ 
    Compute fidelity trajectory  $\{F_1, \dots, F_T\}$ 
    Let  $\bar{F}$  and  $F_{\text{min}}$  be the mean and minimum of the last  $M$  fidelities
    Compute loss  $\mathcal{L}(x) = (1 - \bar{F})^2 + (1 - F_{\text{min}})^2$ 
    Update the control parameters using gradient-based optimization
    if converged then
        break
    end
end
Return  $x^*$ 

```

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. S 11 of this reply, where we discover that the gradient-based optimization method is extremely 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.

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,

- Unlike gradient-based optimization methods, RL learns control policies directly from interaction with the environment and reward feedback, without requiring an accurate differentiable model of the system. This makes RL particularly advantageous in experimental settings with noise, or limited system information.
- The RL agent can learn structured policies that guide exploration beyond local minima typically encountered in conventional optimization.
- Our framework is highly reusable for similar tasks, allowing efficient optimization to be implemented quickly and yielding strong performance.
- Compatibility with constrained or experimentally motivated control spaces, where the gradient optimization methods may require significant modifications.

We have added these results on page 6 line 379-406 at the end of the Results section, and we add additional details in Appendix C.

Comment 9

The term “autonomous” can mean either (i) autonomous design of control waveforms (via RL), or (ii) autonomous preparation/stabilization as an attractor/steady state under driven-dissipative evolution. Please clarify which meaning is intended.

Response

We thank the referee for pointing out this ambiguity in the use of the term “autonomous.”

In our work, “autonomous” is intended to refer to the automated design of control waveforms via reinforcement learning, rather than autonomous state preparation or stabilization in the sense of a driven-dissipative steady-state attractor.

We agree that the current wording may lead to confusion between these two interpretations. In the revised manuscript, we have clarified this point explicitly and adjusted the wording where appropriate to avoid ambiguity. The contents are shown as below.

Abstract, line 13 page 1

we present a Floquet engineering framework in which reinforcement learning is used to realize the autonomous design of the control waveforms for bosonic code preparation, yielding a general, efficient, and noise-resilient strategy.

The contents at line 85 page 2.

we present a Floquet engineering paradigm wherein reinforcement learning is leveraged to automate the synthesis of control waveforms for bosonic code preparation.

The contents at line 130 page 2.

We propose an efficient reinforcement-learning framework that enables the autonomous design of the control waveforms for the rapid and high-fidelity preparation of bosonic states.

The contents at line 153 page 3.

In this work, we introduce a reinforcement learning–assisted quantum-control wave-design strategy, via engineering a temporal periodic driving field (Floquet engineering), to achieve efficient quantum state preparation.

The contents at line 241 page 4.

Here, we present an efficient, robust bosonic code preparation scheme empowered by the automated design of control waveforms via reinforcement learning and Floquet engineering.

The contents at line 531 page 9.

In this work, we have introduced a reinforcement machine learning–assisted Floquet engineering scheme for the autonomous design of the control sequence for bosonic code state preparation.

Comment 10

If the authors can address the above issues well, I can suggest its publication in Communications Physics.

Response

We thank this referee for the encouragement. We have investigated and addressed all the issues pointed out by the referees, which have substantially improved the quality of this manuscript.

-
- [1] L. Guo and V. Peano, Phys. Rev. Lett. **132**, 023602 (2024), URL <https://link.aps.org/doi/10.1103/PhysRevLett.132.023602>.
 - [2] L. Guo, T. Huang, and L. Du, Communications Physics **8**, 414 (2025).
 - [3] S. L. Braunstein and P. van Loock, Rev. Mod. Phys. **77**, 513 (2005), URL <https://link.aps.org/doi/10.1103/RevModPhys.77.513>.
 - [4] Y. Xu and L. Guo, Reports on Progress in Physics **88**, 037602 (2025).

List of changes and replies to the Referees

August 6, 2026

Re: Autonomous Floquet Engineering of Bosonic Codes via Reinforcement Learning
by Zheping Wu, Lingzhen Guo, Haobin Shi, and Wei-Wei Zhang
Manuscript: COMMSPHYS-25-1918-T

Dear Editor and Reviewers,

Thank you very much for the message regarding the above paper and the referees' reports.

We appreciate Referee A for the encouraging assessment "I find most of the responses convincing and the manuscript has certainly improved." We also appreciate Referee B for the recommendation of publication.

According to the reviewers' questions and concerns, we have revised the manuscript, and added the detailed experimental realizations (see the beginning of Supplementary Note 2).

Below, we attach a list of changes and our point-to-point response to the two reviewers' comments. In order to facilitate the reading of the response, we mark the referees' comments in blue, and quote the updated manuscript content with grey backgrounds. All the changes are highlighted in red in our revised version of the manuscript.

In addition, we have adjusted the section structure according to the editor's suggestions.

Best Regards,

Zheping Wu, Lingzhen Guo, Haobin Shi, and Wei-Wei Zhang

List of major changes

- In response to the editor's request, we have shortened the Abstract to fewer than 160 words.
- In response to the editor's request, we have added titles of all Figures and Tables.
- We have added 'and' on line 52.
- In response to the editor's request, we have separated the final paragraph of the Introduction on line 83.
- In response to the editor's request, we have removed the 'Model' section heading, incorporated its three subsections into the Results section, and added the corresponding subsection titles at lines 93, 130, 243, and 378 in the revised manuscript.
- We have added a description of the model's training data collection process through tomography at the beginning of Section **Reinforcement learning for the preparation of bosonic states** (lines 138-145).
- We have corrected a typographical error in the formula of Hamiltonian in the rotating frame of cavity frequency (line 173-174). This correction does not affect the results or conclusions of the manuscript.
- We have added an explanation for the statistics of prepared states fidelity with the gradient-based optimization method (line 385-392, 405-409), and added a new figure showing the best performance of the gradient-based optimization method in Supplementary Note 3.
- We have added a concise description of the experimental realizations in Section **Experimental realizations** (line 536-569), and added the detailed description of the experimental realizations at the beginning of Supplementary Note 2.A.

Response COMMSPHYS-25-1918-A

We appreciate the referees' insightful comments and suggestions. We have modified our manuscript in accordance with their reports. Below, we address the referees' concerns point-by-point in detail. In order to facilitate the reading of the response, the referees' comments are marked in blue, and our updated manuscript content is quoted with gray backgrounds. All the changes are highlighted in red in the new version of the manuscript.

Response to Referee A

Comment 1

I appreciate that the authors revised the manuscript in response to my criticism, including additional analyses, and I acknowledge the efforts made by the authors in addressing all points raised. I find most of the responses convincing and the manuscript has certainly improved. Still, I can only recommend publication if the following point is adequately improved:

Response

We are grateful to the reviewer for their encouraging feedback. Acknowledging their reservations about the manuscript, we have carefully addressed all points raised the referee. We believe these revisions have substantially improved the overall quality of the work.

Comment 2

- I understand that this papers is not aiming at proposing a specific experimental realization in detail, but, as it stands, the discussion about potential implementations is too superficial. For cold atoms, it is not even explained what the oscillator encoding the bosonic code would be. For superconducting circuits, it is not explained in sufficient depth how the Hamiltonian of the proposed resonator+JJs circuit connects to the drive considered.

Response

We thank the referee for this constructive comment and agree that the experimental discussion in the original manuscript was too brief. In the revised main text, we have added a concise explanation of how to implement the driving potential with superconducting circuits (*The revision is marked in red starting from line 536 on page 9.*). For the convenience of referee, we copy added content here.

The key is to realize the cosine component in the driving potential, cf. Eq. (4). In principle, such cosine potential can be directly implemented in the platform of cold atoms with optical lattice potentials [1–3]. Here, for practical scalable quantum control, we outline the general scheme how to implement the driving potential in the platform of superconducting circuits based on the Josephson junctions (JJs). First, a high Q-factor superconducting microwave cavity (3D cavity or 1D on-chip resonator) is coupled to the SQUID circuit, which provides a nonlinear cosine potential, via a mediated JJ-based circuit referred to as the “coupler” that induces a tunable mutual inductance between the cavity and the SQUID circuit [4]. We present more details in the Supplementary Note 2. The total Hamiltonian of the JJ-based superconducting cavity is described by $\hat{\mathcal{H}}(t) = \hbar\omega_0\hat{a}^\dagger\hat{a} + V_{\text{SQUID}}(\hat{\varphi})$ with ω_0 the bare frequency of the cavity and V_{SQUID} the inductive energy of the SQUID circuit given by [4–6]

$$V_{\text{SQUID}}(\hat{\varphi}) = -E_J \cos\left(\frac{\Phi_{\text{sq}}}{\Phi_0}\pi\right) \cos\left[\frac{M(\Phi_x)}{L}\hat{\varphi} + \frac{\Phi_{\text{ext}}}{\Phi_0}\right]. \quad (1)$$

Here, E_J is the single JJ energy, $\Phi_0 = 2\pi\hbar/2e$ is the magnetic flux quantum and $\hat{\varphi} = \sqrt{\frac{\lambda}{2}}(\hat{a}^\dagger + \hat{a}) = 2\pi\hat{\Phi}/\Phi_0$ is the cavity flux operator with dimensionless Planck constant $\lambda = 4e^2\omega_0 L/\hbar$ (L is the inductance of the cavity). The parameters Φ_x , Φ_{ext} , and Φ_{sq} represent the tunable net fluxes applied to the coupler, the external loop of the SQUID, and the SQUID itself, respectively. Compared Eq. (1) to cosine component in Eq. (4), the effective Josephson energy $-E_J \cos\left(\frac{\Phi_{\text{sq}}}{\Phi_0}\pi\right)$ controls the amplitude, the mutual inductance parameter $M(\Phi_x)/L$ determines the effective wave number, and the external flux bias Φ_{ext}/Φ_0 controls the temporal phase. In principle, one can implement the desired driving potential $V(x, t)$ by connect sufficient SQUID circuits to the superconducting cavity.

In addition, we have added contents in Supplementary Note 2. A providing a detailed discussion of how to experimentally implement the needed V_{SQUID} with the platforms of cold atoms in optical lattices and superconducting circuits based on Josephson junctions. We believe these revisions make the connection between our theoretical protocol and experimentally accessible controls substantially clearer.

Comment 3

-While it's interesting information, I do not fully understand why the authors report the fidelity statistics for the protocols optimized by the GRAPE-like method and assess those protocols based on the average fidelity. I would rather think that one takes the best result over stochastically sampled initial conditions for the control parameters. Or is this to show that the control values optimized for given values of the system parameters do not perform well if the system parameters are slightly perturbed?

Response

We thank this referee for the comments and apologize for the unclear descriptions.

The reported statistics are not intended to test robustness against perturbations of the physical system parameters. In all optimization runs, the system parameters are kept fixed; only the initial values of the control parameters are independently randomized and then iteratively optimized by a gradient-based procedure to maximize the target-state fidelity. We use statistical measures because the gradient-based optimization is strongly dependent on its initial control sequence and can converge to different local optima. Consequently, reporting only the best result would demonstrate that a high-fidelity protocol can be found after sufficiently many restarts, but it would not characterize how reliably the method finds such a protocol or how sensitive its success is to the initialization.

The average fidelity, together with the fidelity distribution and success probability, therefore quantifies the typical performance and convergence reliability of the optimization procedure under randomly sampled initial controls. We have clarified this motivation in the revised manuscript and, for completeness, also report the best achieved fidelity in our Supplementary

Note 3 with a new figure as shown in Fig. S 1 here in this reply. As shown in the figure, the gradient-based optimization method reaches a maximum fidelity of 0.9901 at the evolution time marked by the red dot. However, this high fidelity is attained only transiently, and the subsequent pronounced fluctuations prevent the fidelity from stabilizing at a consistently high level.

We have added a clear explanation in our revised manuscript. The revised content is as below.

The revision is on page 6, line 385-392.

As the performance of the gradient-based optimization method heavily relies on its initial control sequence and can converge to different local optima, we implement the algorithm multiple times and study the statistics of its state preparation fidelity performance, which will characterize how reliably the method finds such a protocol or how sensitive its success is to the initialization.

The revision is on page 6, line 405-409.

Although the gradient-based optimization method achieves a maximum fidelity of 0.9901 in our simulations, approximately 50% of the optimization runs yield fidelities below 0.7, indicating substantial variability and a lack of reliable convergence to high-fidelity solutions; see more details in Supplementary Note 3.

Comment 4

- When introducing Fig. 1 on page 2 right column, the wording about the cavity being confined in a dilution refrigerator somewhat suggests that, in applications of the proposed protocol, the control is synthesized based on data from the experimental system. I recommend re-phrasing this to avoid confusion (or comment more about this), as the current RL algorithm requires the full density matrix, hence full tomography, which would be challenging experimentally.

Response

We thank this referee for the comments. We have added an extra sentence to make this

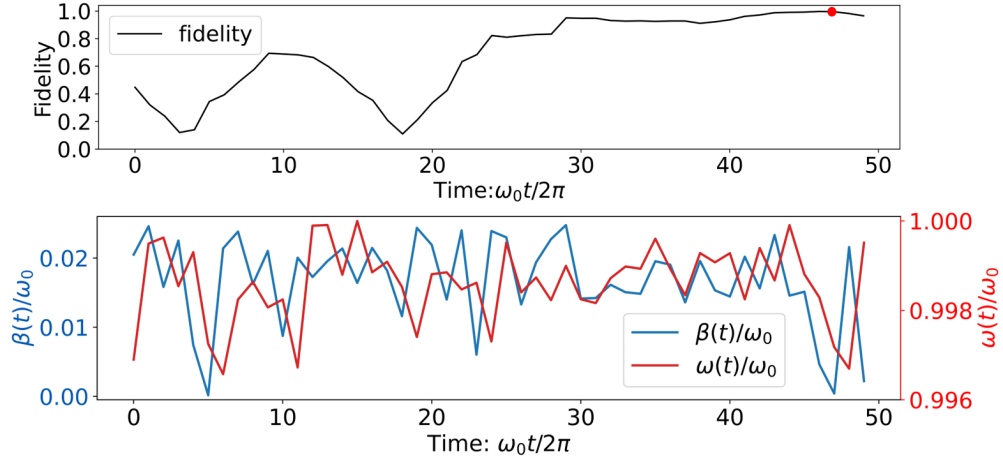


Fig. S 1: The best performance case with a gradient-based optimization method under a noisy environment with $\kappa = 10^{-7}\omega_0$ and $\eta = 0$ among our simulations shown in Fig. 6 of our main text. The highest fidelity achieved is 0.9901 (marked with the red dot). The black curve represents fidelity, and the blue and red curves represent $\beta(t)$ and $\omega(t)$.

clear. The revised content is as below.

The revision is on page 2, line 138-145.

Experimentally, the training data are acquired through tomography, which involves repeated preparation of the same bosonic state followed by controlled analysis operations and measurements. Once trained, the neural-network control policy can be deployed directly on the physical system, without requiring tomography or further optimization during inference.

Comment 5

There are still a few typos. For instance:

- - an "and" or "or" missing in "in superconducting qubit [...] neutral atoms [19]" on line 54
- - a misplaced $\hat{}$ on p and a missing $\hat{}$ on x in Eq. (2)
- - $V(x, \omega t)$ in Eq. (4) and in Eq. (3) cannot be the same object. The notation should be adjusted.

- - "frequency" on line 165, a *\dagger* missing in line 166,
- - ...

Response

We thank this referee for the thorough assessment of our manuscript, and we are sorry for the typos. We have corrected the typos and carefully checked our revisions to ensure their quality.

As for the symbol V , the quantities ' $V(x, \omega t)$ ' appearing in Eqs. (3) and (4) represent the same external driving potential, but they are written at different levels of description.

In Eq. (3),

$$\hat{\mathcal{H}}(t) = \frac{\omega_0}{2} (\hat{p}^2 + \hat{x}^2) + \beta V(\hat{x}, \omega t),$$

$V(\hat{x}, \omega t)$ denotes the time-dependent nonlinear driving potential acting on the cavity. Since the spatial variable is replaced by the position operator $\hat{x}$, $V(\hat{x}, \omega t)$ is an operator-valued function that enters the system Hamiltonian.

Equation (4),

$$V(x, \omega t) = \int_0^\infty A(k, \omega t) \cos[kx + \varphi(k, \omega t)] dk,$$

provides an explicit representation of the same driving potential as a continuous superposition of sinusoidal components with wave number k , amplitude $A(k, \omega t)$, and phase $\varphi(k, \omega t)$.

Therefore, the potential in Eq. (3) is obtained from Eq. (4) by promoting the classical coordinate x to the operator $\hat{x}$:

$$V(\hat{x}, \omega t) = \int_0^\infty A(k, \omega t) \cos[k\hat{x} + \varphi(k, \omega t)] dk.$$

In brief, Eq. (3) specifies how V enters the Hamiltonian, whereas Eq. (4) specifies how V is constructed. The parameter β is an overall driving-strength scale, while $A(k, \omega t)$ determines the relative weight of each spatial frequency component.

Response to Referee B

Comment 1

The authors have responded adequately to my previous comments. The revised manuscript includes useful additional analyses and clarifications, including the discussion of other bosonic encodings, the comparison with a non-RL baseline, and the clarification of the physical model and approximations.

I am satisfied with the authors' response and the current version of the manuscript, and recommend acceptance after minor textual revisions. Before publication, the authors should carefully proofread the manuscript.

Response

We thank this referee for the recommendation of publication.

-
- [1] L. Guo, V. Peano, and F. Marquardt, Phys. Rev. B **105**, 094301 (2022), URL <https://link.aps.org/doi/10.1103/PhysRevB.105.094301>.
 - [2] H. Moritz, T. Stöferle, M. Köhl, and T. Esslinger, Phys. Rev. Lett. **91**, 250402 (2003), URL <https://link.aps.org/doi/10.1103/PhysRevLett.91.250402>.
 - [3] Z. Hadzibabic, S. Stock, B. Battelier, V. Bretin, and J. Dalibard, Phys. Rev. Lett. **93**, 180403 (2004), URL <https://link.aps.org/doi/10.1103/PhysRevLett.93.180403>.
 - [4] L. Guo, T. Huang, and L. Du, Communications Physics **8**, 414 (2025).
 - [5] F. Chen, A. Sirois, R. Simmonds, and A. Rimberg, Applied Physics Letters **98** (2011).
 - [6] F. Chen, J. Li, A. Armour, E. Brahim, J. Stettenheim, A. Sirois, R. Simmonds, M. Blencowe, and A. Rimberg, Physical Review B-Condensed Matter and Materials Physics (2014).

Replies to the Editor

September 2, 2026

Re: Autonomous Floquet Engineering of Bosonic Codes via Reinforcement Learning
by Zheping Wu, Lingzhen Guo, Haobin Shi, and Wei-Wei Zhang
Manuscript: COMMSPHYS-25-1918-T

Dear Editor,

Thank you very much for the message regarding the above paper and the editors' reports.

We also appreciate editors for the recommendation of publication.

According to the editors' questions and concerns, we have revised the manuscript.

Below, we attach our point-to-point response to the editors' comments. In order to facilitate the reading of the response, we mark the editors' comments in blue.

Best Regards,

Zheping Wu, Lingzhen Guo, Haobin Shi, and Wei-Wei Zhang

Response COMMSPHYS-25-1918-A

We appreciate the editors' insightful comments and suggestions. We have modified our manuscript in accordance with their reports. Below, we address the editors' concerns point-by-point in detail. In order to facilitate the reading of the response, the editors' comments are marked in blue.

Response to editor

Comment 1

Fig 1: Please expand the caption to explain the individual elements shown in the figures and the symbols used. More explicitly state the role of reinforcement learning within the autonomous Floquet engineering framework. The caption should also clarify the flow of information between the quantum system and the RL agent.

Response

We thank the editors for this helpful suggestion. We have expanded the caption of Fig. 1 to clarify the individual elements shown in the schematic, define the symbols used, and explain the role of reinforcement learning within the autonomous Floquet-engineering framework. In addition, we have also corrected the label ρ_0 in the figure to ρ_{target} to make its meaning explicit. These revisions have been incorporated into the caption in Fig. 1.

Fig 2: Please make the caption more self-contained by explicitly stating the key result. For example, in (b), why the fidelity corresponding to the RL-designed protocol is being compared to fidelity of the prepared state using the adiabatic ramp method in [24]. In addition, define all colours (colorbars in a,d,e), symbols, and insets and explain the significance of panels (d) and (e).

Response

We thank the editors for this helpful suggestion. We have revised the caption of Fig. 2 to make it self-contained and to explicitly state the key results of each panel. In particular, we now explain that panel (b) compares the fidelity achieved by the RL-designed protocol with that obtained using the Floquet adiabatic-ramp method of Ref. [24] because both methods prepare the same target bosonic-code state, thereby providing a direct comparison of their preparation times. This comparison shows that the RL-driven autonomous Floquet-engineering protocol reaches near-unit fidelity on a substantially shorter timescale. We have also defined the colors, symbols, colorbars, and insets throughout the figure, including the red dots marking the snapshots in panel (a), the black and red curves representing the driving amplitude and frequency in panel (c), and the color scales used in panels (a), (d), and (e). Furthermore, the revised caption explains that panels (d) and (e) visualize the driving potentials generated by the RL-based and adiabatic-ramp protocols, respectively, and that their comparison highlights the rapidly varying, nonadiabatic control sequence discovered by the RL agent, in contrast to the slowly varying adiabatic protocol. We have also explicitly noted that panels (d) and (e) use different horizontal time scales.

Fig 3: In the caption, please explicitly indicate what the different colored bars signify.

Response

We thank the editors for pointing this out. We have revised the caption to explicitly state the meaning of the different bar colors. The revised caption now clarifies that the color of each bar encodes the fidelity of the prepared state, with blue corresponding to lower fidelity and red corresponding to higher fidelity. The bar height and the numerical value shown above each bar represent the same fidelity quantity.
