Enhanced qubit readout via reinforcement learning

Aniket Chatterjee [*]
Department of Physics, University of Oxford and
Centre for Quantum Technologies, National University of Singapore

Jonathan Schwinger
Centre for Quantum Technologies, National University of Singapore

Yvonne Y. Gao [†]
Centre for Quantum Technologies, and
Department of Physics, National University of Singapore, Singapore

(Received 12 December 2024; revised 3 March 2025; accepted 24 April 2025; published 22 May 2025)

Measurement is an essential component of robust and practical quantum computation. For superconducting qubits, the measurement process involves the effective manipulation of the joint qubit-resonator dynamics, and it should ideally provide the highest quality for qubit state discrimination with the shortest readout pulse and resonator reset time. Here, we harness model-free reinforcement learning (RL), together with a tailored training environment, to achieve this multifaceted optimization task. Using the IBM quantum device, we demonstrate that the pulse obtained by the RL agent not only successfully achieves state-of-the-art performance, with an assignment error of $(4.6 \pm 0.4) \times 10^{-3}$, but also executes the readout and the subsequent resonator reset almost three times faster than the system's default process. Furthermore, the learned waveforms are robust against realistic parameter drifts and follow an intuitive form, making them readily implementable on existing hardware with little computational overhead. Our results provide an effective readout strategy to boost the performance of superconducting quantum processors and demonstrate the value of RL in providing optimal and practical solutions for complex quantum information processing tasks.

DOI: 10.1103/PhysRevApplied.23.054057

I. INTRODUCTION

Efficient and accurate measurements are critical building blocks for quantum information processing. In circuit quantum electrodynamics (cQED) devices, information from qubits is typically extracted by mapping their states to a resonator, which is strongly coupled to a transmission line. This simple readout mechanism is capable of achieving high-fidelity single-shot state discrimination in typical cQED systems; however, with the ongoing enhancements in both the scale and robustness of current superconducting systems—with gate fidelities now exceeding 99.5% [1] on sizable devices—even small inefficiencies in the readout become a considerable limitation on the overall system performance. As we strive toward practical quantum computation, rapid and effective readout is essential for many critical tasks, such as adaptive quantum error correction protocols [2,3] or dynamic quantum circuits that demand many mid-circuit measurements [4]. Thus, optimizing the readout process to consistently maximize its efficacy continues to drive active and diverse efforts within the cQED community.

Remarkable progress has been made regarding hardware improvements, such as enhancing the coherence properties of qubits [5,6], creating robust quantum-limited parametric amplifiers [7–9], and developing clever integration control capabilities [10–14]. In parallel, many techniques have also been explored to augment the readout process on existing devices by incorporating optimal integration weights [15,16], accelerating resonator reset [17,18], or applying model-based *in situ* optimization of readout parameters [19].

Here, we harness the power of deep reinforcement learning [20–23] (RL) to optimize the full readout pulse holistically and demonstrate its state-of-the-art performance on a standard superconducting system. More specifically,

[*]Contact author: aniket.chatterjee@chch.ox.ac.uk
[†]Contact author: yvonne.gao@nus.edu.sg

we construct a physically motivated RL agent capable of rapidly finding waveforms that significantly outperform a default square pulse and offer noticeable improvements against other optimization strategies, such as the cavity-level excitation and reset (CLEAR) pulse [18] in fidelity and reset time. We demonstrate the performance of these optimized readout pulses by executing them on IBMQ [1] machines. We find that the readout pulses discovered through deep RL achieve overall assignment errors of $(4.6 \pm 0.4) \times 10^{-3}$, on par with the state of the art, while affording a threefold reduction in the time needed to reach maximum fidelity and then reset the resonator back to vacuum compared to the default IBM configurations. Remarkably, in the typical parameter regimes of IBMQ cloud devices, the agent consistently converges on a physically meaningful waveform that can be associated with an analytical form. We dub this the active four-tone readout (A4R), and we show that it can be tuned up through a series of simple calibration protocols that allow direct adaptation to any existing cQED hardware. Finally, we demonstrate that the optimized pulses are robust against system parameter drifts. Our work demonstrates that by leveraging the power of RL, we can effectively enhance the quality and efficiency of transmon qubit readout. Our strategy is, by design, adapted to hardware constraints and easily implementable on generic superconducting devices, making it a powerful tool for quantum information processing with cQED systems. All source code and parameters are provided on GitHub [24] for easy reproducibility and adaptation.

The process to read out the state of a superconducting qubit via a resonator can be described by the coherent quantum Langevin equation [25], with

$$\dot{\alpha}_{g/e}(t) = -\left(\frac{\kappa}{2} \mp i\chi\right)\alpha_{g/e}(t) - iA(t), \tag{1}$$

where $\alpha_{g/e}$ is the coherent complex-valued resonator state corresponding to the qubit in $|g\rangle$ and $|e\rangle$, respectively, $\dot{\alpha}_{g/e}(t)$ is the time-derivative of the resonator state, κ is the energy decay rate, χ is the qubit-resonator dispersive coupling strength, and $A(t)$ is the coherent input field to the resonator.

Naively, a longer and more energetic pulse should lead to the largest phase-space separation between the trajectories corresponding to $|g\rangle$ and $|e\rangle$, allowing the most effective state discrimination; however, the dynamics on real hardware are more subtle. Increasing the pulse duration also increases the likelihood of qubit decay due to its finite energy relaxation time. In addition, increases in the pulse amplitude and, correspondingly, the photon population in the resonator, can also worsen qubit coherence due to ionization [26]. Thus, we must recognize that improving the overall readout performance is a complex, multifaceted optimization process. In previous studies, various strategies have been developed to improve a specific figure of merit of the readout process in isolation. One of the most notable techniques in this framework is CLEAR [18]. This introduces two rectangular segments of fixed duration to the standard square readout pulse to accelerate the resonator ring-up and reset process for devices in the regimes of $\kappa \approx \chi$. In addition, other techniques, such as those investigated in Refs. [17,27,28], have also shown effective improvements via targeted optimization of the ring-up or reset time individually.

In this study, we use RL to develop a generalized qubit readout process for standard cQED devices that can rapidly achieve the best state discrimination without perturbing the qubit state. This translates into an optimization of the microwave waveform that requires the shortest time and the lowest possible number of photons in the resonator to achieve the highest assignment fidelity, defined as

$$\mathcal{F} = 1 - \frac{P(g|e) + P(e|g)}{2}, \tag{2}$$

where $P(g|e)$ $[P(e|g)]$ represents the probability of measuring state $|g\rangle$ $(|e\rangle)$ when state $|e\rangle$ $(|g\rangle)$ was prepared. Concurrently, we also target an accelerated resonator reset, which enables more rapid repetition of measurements with minimal downtime in between.

II. LEVERAGE REINFORCEMENT LEARNING FOR READOUT OPTIMIZATION

To achieve the target readout performance, we employ a model-free deep RL [20,22] framework that seeks out the best waveform that optimizes all relevant figures of merit associated with the readout concurrently and effectively accounts for known system constraints. A conceptual illustration of our implementation is shown in Fig. 1. Compared to other numerical algorithms—such as Nelder-Mead [29] and evolutionary strategies [30], which have been used in classical-quantum optimization problems [31]—deep RL uses neural networks to achieve more efficient optimization for complex and multiparameter quantum processes. Using an effective decision-making agent for a reward-oriented problem, RL has been successfully applied to implement real-time quantum error correction [2], for gate calibration [32], and to discover new decoders [33]. Given the strength of neural networks in learning arbitrary optimal function mappings with little model-based input, a model-free RL optimization framework is particularly well suited to the multipronged readout process.

A. RL implementation

We choose the proximal policy optimization (PPO) [34] algorithm due to its strong convergence properties for computationally inexpensive environments. PPO is an actor-critic method [35], in which a policy neural network (actor) learns optimal actions for a given reward function and a

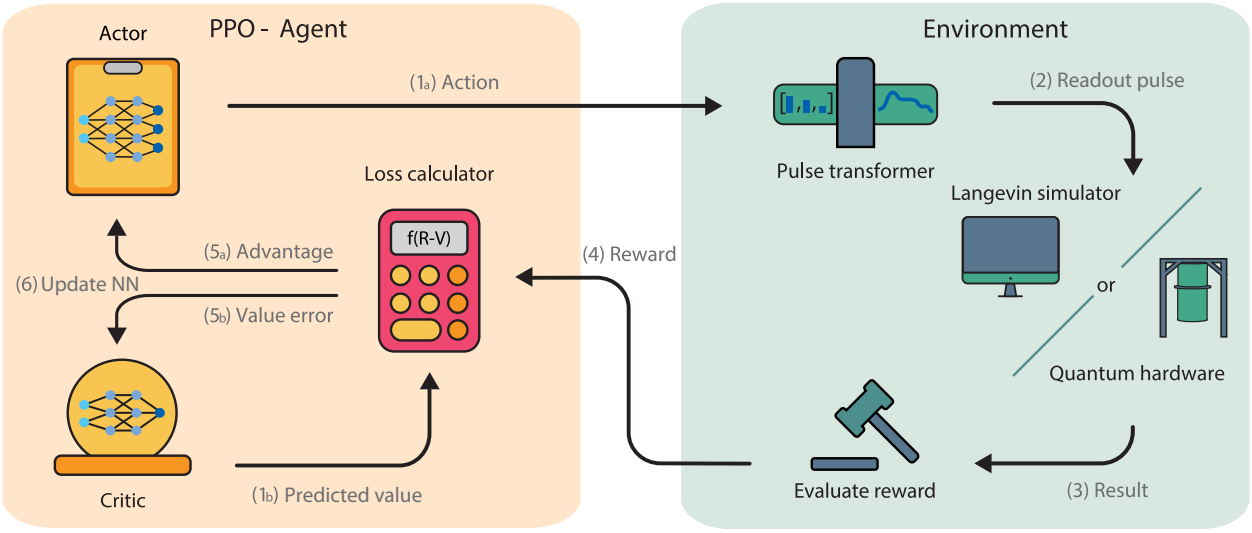

FIG. 1. Depiction of the learning process of a proximal policy optimization (PPO) agent interacting with our environment. Our RL-based readout-optimization framework consists of two main elements: the PPO agent and the environment. In each step of the learning process, the PPO agent samples a batch of actions (1_a) from the policy provided by the actor network. Concurrently, the critic network predicts the value (1_b) of the sampled actions. These actions are processed into the readout pulse (2) through a series of transformations, programmed to reflect the realistic constraints in hardware—i.e., smoothing of the edges to account for limited bandwidth in control electronics—and then tested in a Langevin simulator constructed with real device parameters, emulating the response of a real qubit-resonator system. This simulation's result (3) is then used to calculate the reward (4). The loss calculator then computes the advantage (5_a) and the value error (5_b) by comparing the critic's evaluation against the reward. Finally, the actor and critic neural networks are updated (6) based on the advantage and value error losses, respectively. Our PPO agent iteratively produces increasingly more optimized readout waveforms using this process.

value network (critic) provides a predicted estimate of the reward for the policy. PPO simultaneously ensures good exploration of the action space via the actor and stable learning via the critic. The workflow of the PPO agent is described in Fig. 1. Concretely, for each update of the learning process, the actor outputs the means and standard deviations of a multivariate normal distribution from which an action is sampled. The critic estimates the value of the actions sampled from this action distribution [20] (Fig. 1). During training, the critic network aims to minimize the value error, while the actor aims to produce actions that produce higher rewards than the value estimates of the critic. Training is performed until the critic network accurately estimates the value of the actions and the actor cannot obtain a higher reward than the value estimates, corresponding to finding optimal actions. For the readout problem, this corresponds to estimating the discretized waveform that most effectively maximizes the reward.

The neural-network training can be implemented using either direct measurement data or realistic simulations of the readout dynamics. Here, due to constraints of cloud access, we opted to perform the training via simulations conducted using the quasiclassical Langevin equations [Eq. (1)], which efficiently capture the full dynamics of a resonator coupled to a transmon during a dispersive readout without any costly Hamiltonian simulations. As a result, our method is computationally inexpensive and allows us to obtain all key metrics accurately without simulating high-dimensional Hilbert spaces. With this, we extract the phase-space separation of the resonator state and compute the assignment fidelity, the maximum number of photons, and the residual photon population in the resonator after a reset.

One notable and advantageous feature of PPO is the clipped updates, which corresponds to updating the actor and critic networks with clipped losses, i.e., clipping the loss attributed to the value error and advantage to be between $-\epsilon$ and $+\epsilon$, where ϵ is a small hyperparameter. This prevents overshooting, whereby the networks might fall into some local minima of reward with suboptimal actions, and it also allows for better convergence to reduce the computational overhead incurred by the learning process. Further details about the implementation of the algorithm and the hyperparameters used are presented in Appendix A.

B. Reward function

The crux of our optimization process involves implementing the appropriate reward function, which is crucial for improving learning and optimization efficiency [36], as well as physical viability. All of the relevant optimization objectives and constraints described above are integrated

into our reward function:

$$\begin{aligned} R = &-k_1 \log_{10}(1 - \max(\mathcal{F}(t))) - k_2 \kappa \tau_R \\ &- k_3 \int (A''(t))^2 dt - k_4 (A(0) + A(t_{\min,N})) \\ &- k_5 \mathrm{ReLU}(\max(N(t)) - N_0) \\ &- k_6 \mathrm{heaviside}(t_{\max,\mathcal{F}} - t_{\min,N}), \end{aligned} \tag{3}$$

where k_i are the individual coefficients associated with each reward term; their specific values are detailed in Appendix B.

First, we maximize $\mathcal{F}$, which is computed from the phase-space trajectories [25]. We use a term proportional to the logarithm of the assignment error to steadily increase the reward over a wide range of assignment errors. This motivates the agent to focus on obtaining a maximum in $\mathcal{F}$ before optimizing other auxiliary objectives. Next, we minimize the quantity $\kappa \tau_R$, where τ_R is the total reset time. To ensure practical feasibility on real devices, we account for a conservative bandwidth of 50 MHz on the control system's digital-to-analog converter output and encourage smooth pulses by minimizing $\int \left(A''(t)\right)^2 dt$ [where $A''(t)$ is the second derivative of the waveform] and penalizing the terminal amplitudes $A(0)$ and $A(t_{\min,N})$ at the start and reset point of the readout. Finally, the photon population $N(t)$ is limited to be below that of the default square readout (N_0) on the test device. A final order penalty term ensures that the RL agent always achieves high fidelity and resets the resonator in the correct order.

III. IMPLEMENTATION ON IBMQ DEVICES

We test the efficacy of our RL-based optimization strategy on IBMQ systems via cloud access. We chose two devices, Kyoto and Brisbane, which cover the range of typical system parameters used in superconducting qubit processors.

We obtain relevant system parameters to construct an appropriate learning environment through a series of standard calibration measurements. Specifically, we characterize the dispersive coupling strength χ between the qubit and the readout resonator as well as the decay rate of the resonator κ. For the results reported in this work, we choose two devices: ibm_brisbane Qubit 2 (Brisbane) with $\kappa = 21.4$ MHz and $2\chi/2\pi = 0.31$ MHz, operating in the regime $\kappa/\chi \approx 22$, and ibm_kyoto Qubit 2 (Kyoto) with $\kappa = 10.4$ MHz, $2\chi/2\pi = 0.85$ MHz, and $\kappa/\chi \approx 4$. Both devices have a default measurement time of 1400 ns, consisting of a 720-ns square readout pulse and a 680-ns delay for passive resonator reset. The resulting assignment errors using these default square pulses are $(5.8 \pm 0.9) \times 10^{-3}$ and $(7.0 \pm 1.0) \times 10^{-3}$, respectively.

Testing our RL-enabled readout strategy on these two devices covers the range of typical system parameters used in superconducting qubit readout, with Brisbane representing the fast-readout configuration $\kappa \gg \chi$ made possible through the inclusion of Purcell filters [37], while Kyoto operates at the more familiar $\kappa \sim \chi$ point.

IV. RESULTS AND DISCUSSIONS

A. Learning outcomes

Using the parameters obtained from our calibration measurements on these two devices, we train the PPO agent with batch sizes of 128 steps for 5000 updates. This procedure is implemented on an Nvidia Tesla P100 graphics processing unit with a wall-clock training time of approximately 3 min, which indicates a very efficient learning process and makes it a practical yet powerful strategy for readout optimization. The learning curve for the Brisbane environment is shown in Fig. 2(a), where the normalized sum reward and its dominant contributions due to assignment fidelity, reset amplitude, pulse smoothness, and total reset time are shown as functions of the learning updates. We visualize the intermediate learned policies by plotting the mean action across the batch at the 80th, 300th, and 5000th updates in Fig. 2(b) to gain insights into the agent's progression during this process.

Initially, the agent is completely stochastic and produces random waveforms; however, within the first 80 updates, the agent develops features corresponding to the ring-up, readout, and reset processes. Initially, the agent prioritizes maximizing fidelity, leading to a slightly higher amplitude ring-up in the 80th update waveform and a flat readout segment to reach the appropriate fidelity. From the 80th to 300th update, the agent iteratively improves the reset time, as shown by the higher-amplitude ring-up tone, along with the notable formation of two reset tones for a high-quality reset. From the 300th update onward until the 5000th update, the agent improves the reset amplitude deviation and smooths the waveform till we obtain a final converged waveform at the 5000th update. This is a clear demonstration that the agent approaches the task holistically, balancing the different figures of merit as well as practical constraints of an effective readout throughout the learning process to achieve the optimal solution.

Remarkably, the sequences learned by our RL agent for both Brisbane and Kyoto consistently follow a well-defined structure consisting of four distinct segments. We dub this the active four-tone readout (A4R). A4R follows a simple analytical form, with a single high-amplitude ring-up tone, a constant calibrated segment, and a two-tone reset. A detailed description of the parameters governing A4R is presented in Appendix D. As an illustration, we show the default square pulse, the learned pulse obtained by the RL agent, and the corresponding A4R waveform in Fig. 3(a). In addition, we show that the action of the

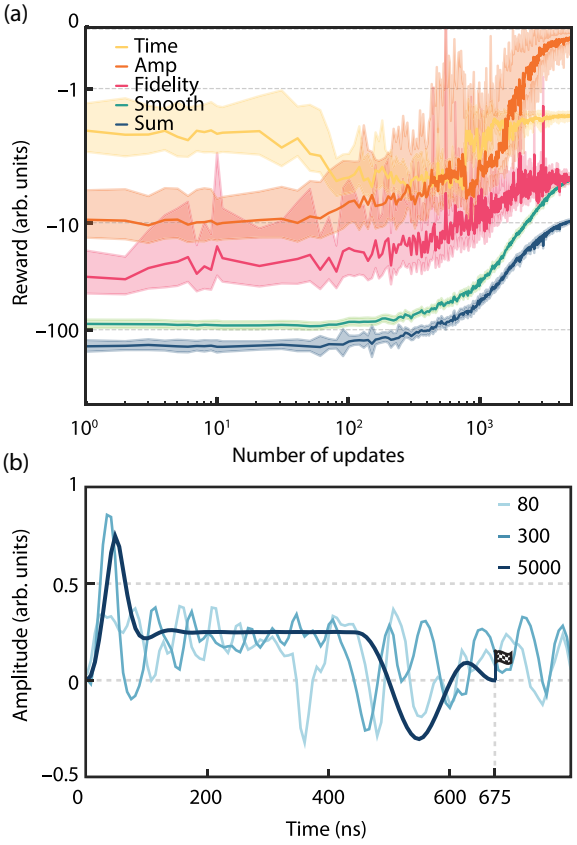

FIG. 2. RL learning curves and resulting waveforms for Brisbane. (a) Learning curves showing the evolution of the total reward, and individual reward contributions over a training run of 5000 updates. The shaded regions show the standard deviations within the batch during training. Initially, fidelity optimization is prioritized while the time reward decreases slightly. Between 1000 and 5000 updates, the fidelity converges to the maximum value allowed by the physical system, while the reset time and auxiliary rewards, such as smoothness and reset amplitudes, are further optimized until convergence. (b) Intermediate waveforms after the action of the pulse transformer at 80, 300, and 5000 updates during training. The initial waveforms are all seeded from randomly generated pulses. Within 300 updates, a distinct high-amplitude ring-up within the limit of the maximum photon population emerges. The agent makes significant improvements between 300 and 5000 updates to smooth the waveform and achieve a fast reset, within 675 ns.

optimized pulses (RL and A4R) follows the intuitive trajectories in phase space [Fig. 3(b)], starting with a sharp increase in the I component, followed by a constant-amplitude drive, which leads to maximum separation in the Q component without necessarily driving the resonator to its steady state. Finally, the active reset is performed to effectively reduce the final photon population to below 0.05, similar to the residual population associated with the default readout on the test device. Although the first reset tone can clear the resonator state from >50 to <1 photon, a short second kickback tone is required to ensure that the small fraction of remaining excitations is rapidly evacuated. This two-tone reset, with segments of variable durations and amplitudes, ensures a faster, more flexible, and more complete reset of the resonator compared to the CLEAR sequence, which requires the two reset tones to be of equal duration [18]. A more detailed comparison with CLEAR is presented in Appendix E. The key practical advantage of the A4R protocol is the generality of its dynamics. By modeling it as a photon-transfer process, we can construct a set of simple calibration measurements, as detailed in Appendix D, to obtain the optimal A4R parameters without further numerical optimization.

B. Performance of optimized waveforms

We execute the resulting optimized waveforms on IBM Kyoto (Q2) and Brisbane (Q2) and verify their performance on IBMQ devices via cloud access. The default readout configurations for these qubits consist of a square pulse and a passive set, with total length 1400 ns and maximum numbers of photons of 26 and 59, respectively. In comparison, the RL agent achieves equal or slightly superior fidelities ($\approx 99.5\%$) using the same number of photons but offers significantly shorter measurement and reset times, with total durations of 675 and 470 ns, respectively. An exemplary phase-space trajectory of the resonator state is shown in Fig. 3(b). This shows that both the RL pulse and A4R follow intuitive dynamics to achieve effective state discrimination [Fig. 3(c)] and realize effective state reset at the end of the process. A summary of all relevant figures of merit for the readout performance is provided in Appendix C.

Moreover, the resulting assignment fidelities also exhibit strong stability against realistic system-parameter drifts [38]. We test this by simulating the readout properties over a range of κ and χ that mimics a $\pm 10\%$ fluctuation on the real hardware. The landscapes of the resulting performance metrics are shown in Figs. 3(d) and 3(e). The assignment fidelity varies by $\leq 1\%$ for κ/χ drifts within $\pm 10\%$, while the reset time varies by <80 ns and is only dependent on changes in κ.

C. Connection to other optimization techniques

These demonstrations show that we have constructed a physically informed RL agent to effectively optimize the readout process of a superconducting qubit with minimal computational overhead. We implemented the learned protocol obtained by the agent on IBMQ devices and showed that they successfully achieve high-fidelity readout with accelerated resonator reset, offering a significant reduction in the total measurement duration.

Importantly, the optimized sequences exhibit well-defined intuitive structures and are robust against hardware parameter drifts, highlighting their practical applicability to real quantum processors. The convergence to the

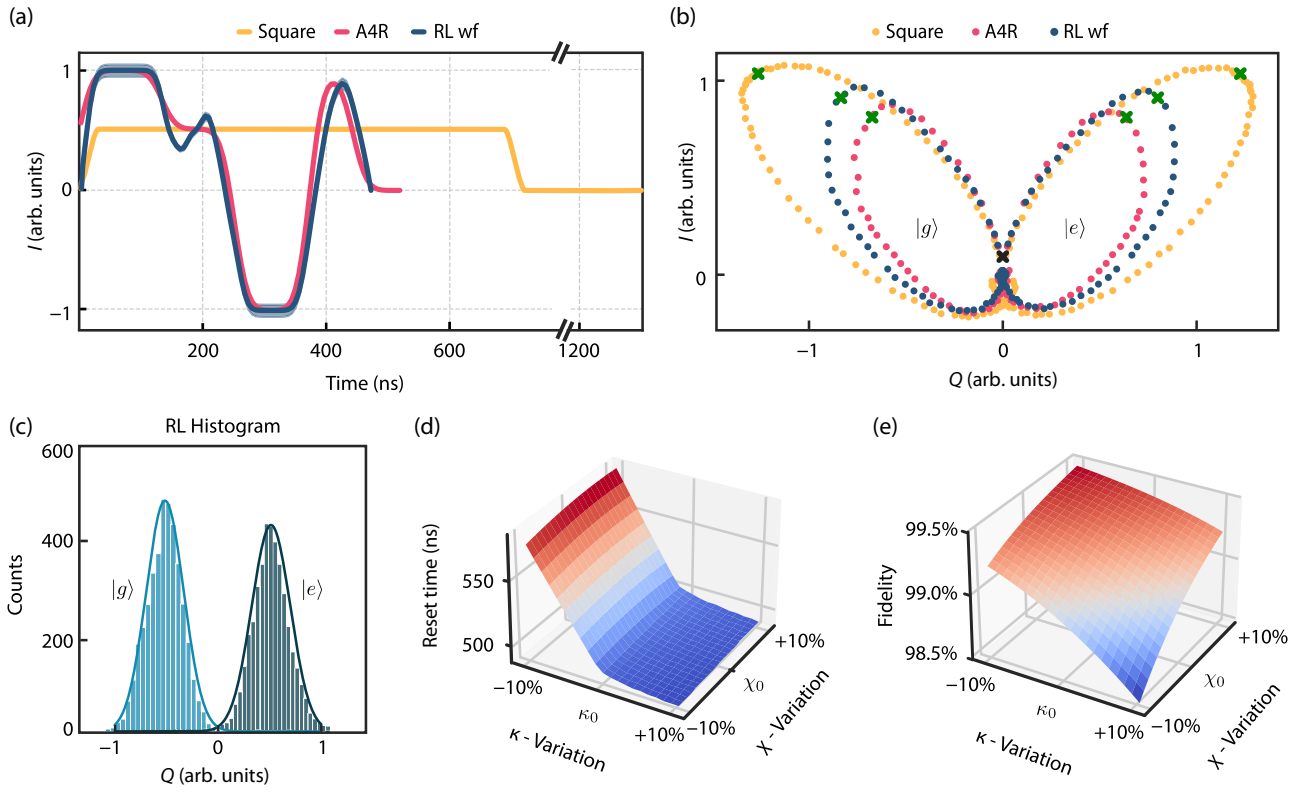

FIG. 3. Performance and stability of optimized waveforms discovered via RL. (a) An exemplary waveform (blue) for IBM Kyoto discovered by our physically informed RL agent. Its smooth amplitude modulation and low bandwidth allow direct implementation on hardware. Both the learned waveform and its generalized analytical counterpart, A4R (red), lead to low assignment errors of $(4.6 \pm 0.4) \times 10^{-3}$ and $(5.7 \pm 0.8) \times 10^{-3}$ and reset the resonator within 470 ns, significantly faster than the 1400 ns needed using the default readout configuration in Kyoto. We use calibrated integration weights around the time of maximum assignment fidelity at 264 ns for A4R and the RL waveform. More details are included in Appendix C. (b) Measured phase-space trajectories of the coherent states in the readout resonator corresponding to the qubit in states $|g\rangle$ and $|e\rangle$, achieved with the default readout pulse (yellow), RL optimized pulse (blue), and A4R (red), respectively. Notably, the RL and A4R trajectories achieve maximum fidelities at much smaller separations than the square pulse, demonstrating that shorter readouts, which do not necessarily reach steady state, can be more optimal. The black cross marks the beginning of the readout and the green crosses indicate the end of the drive to populate the readout resonator and the start of the reset to vacuum. (c) An exemplary readout histogram from 4096 single-shot measurements on IBM Kyoto using the waveform discovered by our RL agent, showing effective state discrimination with $>5\sigma$ separation between the $|g\rangle$ and $|e\rangle$ distributions. Simulated stability landscapes for the (d) reset time and (e) assignment fidelity under κ and χ variations of up to 10%. Both show that these key figures of merit do not fluctuate significantly under realistic parameter drifts.

generalized A4R sequence closely resembles the analytical pulses such as CLEAR [18] and that in Ref. [39]. A more quantitative comparison with CLEAR is provided in Appendix E. This feature indicates that our framework is capable of discovering physically meaningful sequences without any predefined models, making it a versatile tool that can be used to first seek out optimal waveforms when the physical model of the system is not well understood. Once a stable solution is found, as in the case of A4R here, it can be generalized into a few-parameter numerical optimization for easy implementation on real hardware.

Furthermore, the stable structure of the resulting waveform across different system parameters also makes it a versatile tool to augment other hardware-based readout-optimization approaches. For instance, in Ref. [40], the speedup is attained by tailoring the readout and filter mode on the device such that an extremely fast readout (i.e., a vastly enhanced κ) can be performed without degrading the qubit coherence. In another recent work [14], a speedup of the readout process was achieved through physically manipulating the spectral separations of the qubit and readout resonator such that the effective χ was enhanced during the readout. With the ability to achieve a stable and robust readout performance across different κ/χ regimes, our technique can be readily adapted to work in conjunction with these hardware upgrades to further enhance the overall readout performance.

Overall, our work harnesses the intersectional knowledge between model-free reinforcement learning and quantum information science to attain tangible performance enhancement in superconducting readout. Our work also

serves to demonstrate the general efficacy and versatility of RL in complex optimization problems for practical quantum computing applications.

ACKNOWLEDGMENTS

We thank Mr. Arthur Strauss, Dr. Lirande Pira, Mr. Beng Yee Gan, and Dr. Patrick Rebentrost for the insightful discussions and feedback on this work. We acknowledge support from Horizon Quantum Computing in providing access to the IBM cloud devices, and we thank Mr. Kyle Timonthy Chu for his assistance in executing test demonstrations. We acknowledge funding from the National Research Foundation (NRFF12-2020-0063) and the Ministry of Education (MOE-21-0054-P0001), Singapore.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX A: REINFORCEMENT LEARNING IMPLEMENTATION

For the readout-optimization task, we use policy-based reinforcement learning. Broadly speaking, the algorithm tries to find the optimal policy function $\pi_\theta(s|a)$ that chooses the best action a to take at a given state s, parametrized by a neural network with weights θ. In this context, s does not refer to the quantum state but instead represents all current information describing the environment, e.g., the system parameters and final qubit state. Thus, s is simply a constant vector, as the system does not change during the training process and is reset after each step of the optimization.

More concretely, we use the proximal policy optimization (PPO), which is a widely used policy-based algorithm known for its robust learning and convergence capabilities without expansive memory requirements. PPO is an actor-critic method, in which the actor network (policy) outputs the actions to take at a given state, and the critic network tries to predict the value of the actions taken. The actor tries to maximize its advantage $A = R - V$, where R is the reward for a given action and V is the value predicted by the critic. The critic focuses on improving its target estimates $(V - R)^2$. As the actor, through the learning process, increases its likelihood of producing high-advantage actions, the critic improves its estimates of the actor's reward.

Here, we implement the PPO algorithm in JAX [41], based on PUREJAXRL [42]. We opt for shared hidden layers [43] for the actor and critic to reduce the training complexity and for faster learning. In addition, we perform batch rollouts [44] of simulations to speed up training. The hyperparameters were optimized using a grid search, and the results of this search are listed in Table I.

TABLE I. Optimal hyperparameters used for the PPO algorithm. The same set of parameters is used to obtain the waveforms for both Kyoto and Brisbane.

Hyperparameter	Value
Learning rate	0.0003
Hidden layer size	128
Hidden layers	2
No. vectorized environments	128
Activation function	ReLU6
Optimizer	Adam
Update epochs	4
No. minibatches	4
Clip epsilon	0.2
Value clip epsilon	0.2
Entropy coefficient	0
Maximum normed gradient	0.5

APPENDIX B: READOUT ENVIRONMENT

The readout environment takes a 1D real vector of pulse amplitudes as input, which corresponds to the readout waveform sent to the resonator. As the default readout pulse is approximately 720 ns long on IBM Quantum devices, we discretize the RL readout pulse to take 121 real amplitudes, corresponding to a sampling time of 6 ns. Real-valued pulses are sufficient, as we focus on systems with low self-Kerr and operate within $N \ll N_c$, where N_c is the critical photon population [45]. We clip the pulse amplitudes to $\pm\mu$, where μ is the maximum amplitude available on the IBMQ system relative to the default amplitude used for measurement. We also smooth the waveform using a Gaussian convolution to mimic the low-pass filter response of a typical arbitrary waveform generator. Finally, we scale up the waveform by the steady-state amplitude $A_0 = 0.5\sqrt{N_0(\kappa^2 + 4\chi^2)}$, where N_0 is the default measurement photon population. We obtain the values of N_0 and the resonator energy decay rate κ through ac Stark shift measurements [46] on each of the test devices.

The processed waveform is passed to the Langevin simulation, described by Eq. (1). From the simulated coherent state amplitude, we obtain the photon population $N(t) = |\alpha(t)|^2$ and the state separation $S(t) = |\alpha_+(t) - \alpha_-(t)|$. To calculate the resultant time-dependent fidelity, we model the resonator as a Gaussian state and assume we integrate for a short boxcar window that can acquire the instantaneous signal. Hence, the extracted fidelity from the state separation in phase space is given by

$$F_{\mathrm{SNR}}(t) = 0.5\,(1 + \mathrm{erf}\,(\lambda S(t)))\,, \tag{B1}$$

where λ is a scaling constant that captures amplification from a quantum-limited amplifier and assumes constant noise for the distribution. The state-separation fidelity F_{SNR} represents the fidelity of the measurement if we assume perfect state preparation and no qubit evolution during the readout. We account for the infidelity associated with qubit decay and measurement-induced state transitions by adding a qubit decay rate that increases linearly with the resonator photon population [47]. This qubit fidelity term, F_q, is given by

$$F_q(t) = \exp\left(-\gamma_0 t - \gamma_P \int_0^t N(\tau)d\tau\right). \tag{B2}$$

Finally, the overall fidelity is given by

$$\mathcal{F}(t) = 0.5\left(1 + \mathrm{erf}\left(F_0 \times \lambda S(t) \times F_q(t)\right)\right), \tag{B3}$$

where F_0 is the qubit initialization fidelity. Notably, F_q decays with time, while the state separation $S(t)$ increases with time. Thus, it is clear that there is an optimal measurement duration, and this is not necessarily the steady-state duration due to realistic qubit decay. Finally, we evaluate $\max \mathcal{F}(t)$, which is then used in the reward function. To capture the realistic behavior of the hardware, we extract these parameters from the chosen devices by fitting Eq. (B3) to the measured assignment infidelities obtained with various different acquisition times.

To calculate the readout duration, we assume that the measurement ends at the first photon minimum after maximum fidelity is achieved, as enforced by the order penalty in the reward function. The reset time of the readout corresponds to $t_{\mathrm{min},N}$ along with the additional time required to reach the target photon population N_I. We calculate this by assuming standard exponential decay of the photon population. Thus, the total reset time τ_R is

$$\tau_R = t_{\mathrm{min},N} + \frac{m}{\kappa} \times \log\left(\frac{N(t_{\mathrm{min},N})}{N_I}\right), \tag{B4}$$

where m is a T_1 scaling factor added to discourage the agent from conducting a partial reset and then waiting. We find $m = 8$ is sufficient to force the agent to actively reset all systems regardless of κ.

The reward function for the environment is given by Eq. (3), as detailed in the main text. We choose reward coefficients k_i to prioritize fidelity, after which the auxiliary objectives of faster readout and smoothing are included. Hence, we set the fidelity coefficient $k_1 = 10$, the time coefficient $k_2 = 2$, and the smoothness coefficient $k_3 = 1$. The time and smoothness coefficients are determined through grid sweeps of training runs, and provided they are significantly smaller than the fidelity coefficient, we recover the waveforms presented in the main text. We set penalty-term coefficients, such as the amplitude coefficient k_4 and the photon coefficient k_5, to be large values of 100, such that they are enforced throughout the training run.

We use DIFFRAX [48] and JAX for the ordinary differential equation simulation to implement the readout environment. These tools allow us to leverage just-in-time compilation and automatic vectorization to execute fast parallel training runs and hyperparameter-optimization sweeps.

APPENDIX C: SUMMARY OF SYSTEM PARAMETERS AND READOUT PERFORMANCE

A comparison of the readout performance on both IBM Kyoto and Brisbane is presented in Table II, and a summary of the readout pulse parameters is provided in Table III. We identify the time of maximum fidelity using repeated fidelity measurements at various acquisition times for each readout. For the infidelities of the RL waveforms and A4R, we integrate the signal for a window centered around the time of maximum fidelity with calibrated integration weights provided by the IBM devices. For the default readout, we report the infidelity as integrated with the custom weights provided by IBM, which achieves the same fidelity within shot noise as integrating

TABLE II. Comparison of readout performances between the default measurement protocol, the RL-optimized waveform, and a calibrated A4R waveform on Kyoto and Brisbane. Bold numbers correspond to the best achieved value in the category. On both systems, the RL and A4R waveforms show similar improvements of a two- to threefold speedup over the default pulse with similar or even lower assignment errors. This is attributed to the high amplitude ring-up causing the resonator to achieve minimum error more rapidly. Furthermore, the reset for both devices reduces the resonator photon population to <0.05, comparable with the default configurations used on these test devices.

Readout configuration	Duration (ns)	$1 - \mathcal{F}_{\max}$	$t_{\mathcal{F}_{\max}}$ (ns)	$N_{\max}$	$N_{\min}$
Kyoto default + delay	1400	$(5.8 \pm 0.9) \times 10^{-3}$	560	26.8	<0.05
RL	**470**	$(\mathbf{4.6 \pm 0.4}) \times 10^{-3}$	**264**	**26.2**	(0.04 ± 0.03)
A4R	490	$(5.7 \pm 0.8) \times 10^{-3}$	**264**	26.5	(0.02 ± 0.02)
Brisbane default + delay	1400	$(\mathbf{7.0 \pm 1.0}) \times 10^{-3}$	733	**58.6**	<0.05
RL	**675**	$(\mathbf{7.0 \pm 1.0}) \times 10^{-3}$	**542**	59.1	(0.05 ± 0.03)
A4R	685	$(7.2 \pm 1.0) \times 10^{-3}$	**542**	59.2	(0.03 ± 0.03)

TABLE III. Parameters for the two qubits used in our implementation of the optimized readout pulses.

Parameters	Brisbane Q2	Kyoto Q2
qubit frequency	4.610 GHz	4.733 GHz
resonator frequency	7.229 GHz	7.225 GHz
χ to readout	0.15 MHz	0.42 MHz
readout κ	21.4 MHz	10.4 MHz
qubit T_1	291 μs	344 μs
qubit T_2	222 μs	281 μs

for a window around the time of maximum fidelity. We are not able to further optimize the specific form of these integration weights due to restrictions on IBM devices.

APPENDIX D: GENERALIZED ACTIVE FOUR-TONE READOUT (A4R)

The A4R pulse is designed to achieve high fidelity and rapid resonator reset in the regime of $\kappa/\chi \gg 1$. It consists of four sequential segments: ring-up, steady-state readout, photon depletion, and kickback (see Fig. 4). Each segment i is characterized by an amplitude A_i and a duration τ_i, resulting in a total of eight parameters that can be extracted efficiently from measurement outcomes or a sample-efficient optimizer such as Nelder-Mead [29].

In this regime, a single ring-up segment is sufficient, as the phase induced by the dispersive interaction is small. Thus, one calibrated ring-up tone effectively drives the resonator to reach near steady state in phase space, after which the second tone ensures a high-fidelity state discrimination. The amplitude during the steady-state readout segment, A_2, is fixed at a target steady-state amplitude, while A_0 and the remaining seven parameters can be calibrated using a simple three-step measurement routine.

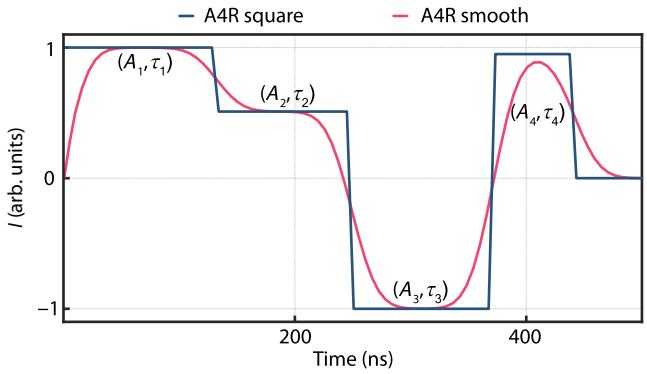

FIG. 4. A4R waveform, which consists of four segments: ring up, steady-state readout, reset, and kickback. Each segment is characterized by two variables: duration τ_i and amplitude A_i. The blue line depicts the square pulse form of A4R, while the red line shows the smoothed form after bandwidth constraints are applied.

We set the ring-up amplitude $A_1 = \mu A_0$ and solve the coherent Langevin equation for a constant-amplitude tone to estimate the ring-up duration τ_1 as

$$\tau_1 = \frac{2}{\kappa} \ln\left(\frac{\mu}{\mu - 1}\right), \tag{D1}$$

where first-order approximations in χt are used to eliminate oscillating terms. As the ring-up should be as quick as possible, we use this estimate of τ_1 to find the largest A_1 within the bandwidth constraint. We sweep the estimated τ_1 value and perform photon-number measurements to get the final value of τ_1. Then, τ_2 is calibrated to maximize fidelity by measuring the instantaneous fidelity while varying this duration.

For resonator reset, two tones are required to account for the impact of κ and χ on the trajectory. To achieve fast depletion of photons in the resonator, we select $A_3 = -\mu A_0$. The duration, τ_3, can then be calculated as

$$\tau_3 = \frac{2}{\kappa} \ln\left(\frac{\mu + 1}{\mu}\right), \tag{D2}$$

which is independent of χ. Any inaccuracies in the reset are corrected by the kickback tone with, the parameters A_4 and τ_4 determined directly via the residual photon population measurements. For the Brisbane setup, the first reset tone removes the bulk of the energy, bringing its photon number from 58.8 to 0.6. The final kickback further reduces the population to <0.05 photons, comparable to that achieved in the default readout via passive decay of the resonator.

APPENDIX E: COMPARISON WITH CLEAR

In this study, we have demonstrated that A4R provides a significant improvement in the readout efficiency for the test devices in the $\kappa \gg \chi$ regime. Here, we provide an analysis of its performance against CLEAR (Ref. [18]), which enacts an active ring-up and reset alongside a standard steady-state square pulse. Focusing on the $\kappa \approx \chi$ regime, CLEAR applies the ansatz of playing two segments of equal duration with variable amplitudes before and after a square pulse. The four optimal amplitudes associated with the ring-up and reset are analytically determined in the linear regime and, more generally, can be obtained numerically with the appropriate target photon populations in mind. For instance, in Ref. [18], the authors used a target of N_0 after ring-up and ≈ 0.1 photons after the reset. Overall, they found a speedup of 50% in resonator reset compared to using a passive process following a square readout on the specific test device.

To compare the A4R readout with CLEAR, we reproduce the optimization protocol outlined in Ref. [18] to determine CLEAR pulse parameters for the Kyoto

device using a Nelder-Mead [29] optimizer implemented from SCIPY [49]. The resulting optimized parameters for CLEAR show a noticeably faster reset (≈ 460 ns) compared to the default (≈ 680 ns); however, no discernible speedup is found for the ring-up stage in this $\kappa \gg \chi$ regime when a constraint is imposed on the maximum number of photons in the optimization. Overall, for Kyoto, CLEAR requires ≈ 1180 ns for a full readout cycle. While this offers an improvement over the default (1400 ns), it significantly underperforms compared to A4R, which completes the full readout process in 470 ns under the same optimization constraints.

Overall, alongside a threefold speedup over the simple square readout with passive decay, A4R demonstrates a 2.5-fold speedup over the CLEAR waveform. While A4R and CLEAR are quite reminiscent of each other structurally, several improvements in A4R—such as a single-segment ring-up, optimized readout durations, and flexible reset segment durations for the $\kappa/\chi \gg 1$ regime—allow for significantly improved readout speeds while preserving a robust assignment fidelity.

To assess the performance comparison more extensively, we also simulated the action of CLEAR and A4R over a wide range of κ/χ values. In Table IV, we compare both the resulting readout fidelities and the durations of the pulses discovered by the RL agent and the CLEAR pulses derived analytically for the same system parameters. In these simulations, we fix the dispersive coupling strength χ to match that of Kyoto Qubit 2, and we vary the κ/χ ratio between 0.5, 2, 5, and 10. We set the other relevant system parameters, such as the photon population of the readout N_0 and the decay rates γ_0 and γ_P, such that a square readout pulse achieves a fidelity of 0.995. In this setting, as there is no reference square readout, we reduce the readout segment of the waveform until the time of maximum fidelity, similar to that in A4R. We note that this in itself represents a significant speedup compared to the CLEAR readout, which previously required a longer duration near steady state. The code for reproducing these results is provided on GitHub [24].

This comparison shows that even in regimes where κ is large, as pursued in some hardware optimization strategies, the optimized sequence produced by RL still offers a clear acceleration in the total readout and reset process compared to the CLEAR pulse. As such, our framework serves as a valuable and versatile tool that can be applied to augment the performance of superconducting qubit readout across different device parameters or hardware configurations.

TABLE IV. Simulated readout fidelity and total readout duration (measurement + reset) for different system parameters using CLEAR and the waveforms discovered by RL.

Device parameter	RL/A4R	CLEAR
$\kappa/\chi = 0.5$	99.4% ; 967 ns	99.4% ; 1100 ns
$\kappa/\chi = 2$	99.5% ; 502 ns	99.5% ; 757 ns
$\kappa/\chi = 5$	99.4% ; 265 ns	99.5% ; 457 ns
$\kappa/\chi = 10$	99.6% ; 164 ns	99.7% ; 276 ns

[1] IBM Quantum, https://quantum.ibm.com/ (2021).

[2] V. V. Sivak, A. Eickbusch, B. Royer, S. Singh, I. Tsioutsios, S. Ganjam, A. Miano, B. L. Brock, A. Z. Ding, and L. Frunzio *et al.*, Real-time quantum error correction beyond break-even, Nature **616**, 50 (2023).

[3] N. Sundaresan, T. J. Yoder, Y. Kim, M. Li, E. H. Chen, G. Harper, T. Thorbeck, A. W. Cross, A. D. Córcoles, and M. Takita, Demonstrating multi-round subsystem quantum error correction using matching and maximum likelihood decoders, Nat. Commun. **14**, 2852 (2023).

[4] A. D. Córcoles, M. Takita, K. Inoue, S. Lekuch, Z. K. Minev, J. M. Chow, and J. M. Gambetta, Exploiting dynamic quantum circuits in a quantum algorithm with superconducting qubits, Phys. Rev. Lett. **127**, 100501 (2021).

[5] A. P. M. Place, L. V. H. Rodgers, P. Mundada, B. M. Smitham, M. Fitzpatrick, Z. Leng, A. Premkumar, J. Bryon, A. Vrajitoarea, and S. Sussman *et al.*, New material platform for superconducting transmon qubits with coherence times exceeding 0.3 milliseconds, Nat. Commun. **12**, 1779 (2021).

[6] C. Wang, X. Li, H. Xu, Z. Li, J. Wang, Z. Yang, Z. Mi, X. Liang, T. Su, and C. Yang *et al.*, Towards practical quantum computers: Transmon qubit with a lifetime approaching 0.5 milliseconds, Npj Quantum Inf. **8**, 3 (2022).

[7] J. Aumentado, Superconducting parametric amplifiers: The state of the art in Josephson parametric amplifiers, IEEE Microw. Mag. **21**, 45 (2020).

[8] R. Vijay, D. H. Slichter, and I. Siddiqi, Observation of quantum jumps in a superconducting artificial atom, Phys. Rev. Lett. **106**, 110502 (2011).

[9] B. Ho Eom, P. K. Day, H. G. LeDuc, J. Zmuidzinas, and A wideband, Low-noise superconducting amplifier with high dynamic range, Nat. Phys. **8**, 623 (2012).

[10] M. D. Reed, B. R. Johnson, A. A. Houck, L. DiCarlo, J. M. Chow, D. I. Schuster, L. Frunzio, and R. J. Schoelkopf, Fast reset and suppressing spontaneous emission of a superconducting qubit, Appl. Phys. Lett. **96**, 203110 (2010).

[11] T. Walter, P. Kurpiers, S. Gasparinetti, P. Magnard, A. Potočnik, Y. Salathé, M. Pechal, M. Mondal, M. Oppliger, C. Eichler, and A. Wallraff, Rapid high-fidelity single-shot dispersive readout of superconducting qubits, Phys. Rev. Appl. **7**, 054020 (2017).

[12] Y. Sunada, S. Kono, J. Ilves, S. Tamate, T. Sugiyama, Y. Tabuchi, and Y. Nakamura, Fast readout and reset of a superconducting qubit coupled to a resonator with an intrinsic Purcell filter, Phys. Rev. Appl. **17**, 044016 (2022).

[13] R. Dassonneville, T. Ramos, V. Milchakov, L. Planat, E. Dumur, F. Foroughi, J. Puertas, S. Leger, K. Bharadwaj, and J. Delaforce *et al.*, Fast high-fidelity quantum nondemolition qubit readout via a nonperturbative cross-Kerr coupling, Phys. Rev. X **10**, 011045 (2020).

[14] F. Swiadek, R. Shillito, P. Magnard, A. Remm, C. Hellings, N. Lacroix, Q. Ficheux, D. C. Zanuz, G. J. Norris, and A. Blais *et al.*, Enhancing dispersive readout of superconducting qubits through dynamic control of the dispersive shift: Experiment and theory, PRX Quantum **5**, 040326 (2024).

[15] J. Gambetta, W. A. Braff, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, Protocols for optimal readout of qubits using a continuous quantum nondemolition measurement, Phys. Rev. A **76**, 012325 (2007).

[16] C. A. Ryan, B. R. Johnson, J. M. Gambetta, J. M. Chow, M. P. da Silva, O. E. Dial, and T. A. Ohki, Tomography via correlation of noisy measurement records, Phys. Rev. A **91**, 022118 (2015).

[17] C. C. Bultink, M. A. Rol, T. E. O'Brien, X. Fu, B. C. S. Dikken, C. Dickel, R. F. L. Vermeulen, J. C. de Sterke, A. Bruno, R. N. Schouten, and L. DiCarlo, Active resonator reset in the nonlinear dispersive regime of circuit QED, Phys. Rev. Appl. **6**, 034008 (2016).

[18] D. T. McClure, H. Paik, L. S. Bishop, M. Steffen, J. M. Chow, and J. M. Gambetta, Rapid driven reset of a qubit readout resonator, Phys. Rev. Appl. **5**, 011001 (2016).

[19] A. Bengtsson, A. Opremcak, M. Khezri, D. Sank, A. Bourassa, K. J. Satzinger, S. Hong, C. Erickson, B. J. Lester, and K. C. Miao *et al.*, Model-based optimization of superconducting qubit readout, Phys. Rev. Lett. **132**, 100603 (2024).

[20] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (The MIT Press, Cambridge, MA, USA, 2018), 2nd ed.

[21] S. Kakade, A natural policy gradient, in *Proceedings of the 15th International Conference on Neural Information Processing Systems: Natural and Synthetic*, edited by T. Dietterich, S. Becker, and Z. Ghahramani (MIT Press, Cambridge, MA, USA, 2001), pp. 1531–1538.

[22] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, An introduction to deep reinforcement learning, Foundations and Trends® in Machine Learning **11**, 219 (2018).

[23] J. Schulman, S. Levine, P. Moritz, M. I. Jordan, and P. Abbeel, Trust region policy optimization, arXiv:1502.05477.

[24] AnikenC, https://github.com/AnikenC/RL-meets-Qubit-Readout.git GitHub - anikenc/rl-meets-qubit-readout: Repository for demonstration of enhanced qubit readout via reinforcement learning (2024).

[25] A. Blais, A. L. Grimsmo, S. M. Girvin, and A. Wallraff, Circuit quantum electrodynamics, Rev. Mod. Phys. **93**, 025005 (2021).

[26] M. F. Dumas, B. Groleau-Paré, A. McDonald, M. H. Muñoz Arias, C. Lledó, B. D'Anjou, and A. Blais, Measurement-induced transmon ionization, Phys. Rev. X **14**, 041023 (2024).

[27] C. Hoffer, Superconducting qubit readout pulse optimization using deep reinforcement learning (2021).

[28] B. Lienhard, Machine learning assisted superconducting qubit readout (2021).

[29] J. A. Nelder and R. Mead, A simplex method for function minimization, Comput. J. **7**, 308 (1965).

[30] J. Kennedy and R. Eberhart, Particle swarm optimization, in *Proceedings of ICNN'95 - International Conference on Neural Networks* (IEEE, 1995), Vol. 4, pp. 1942–1948.

[31] X. Cheng, X.-J. Lu, Y.-N. Liu, and S. Kuang, Comparison of differential evolution, particle swarm optimization, quantum-behaved particle swarm optimization, and quantum evolutionary algorithm for preparation of quantum states, Chin. Phys. B **32**, 020202 (2023).

[32] Y. Baum, M. Amico, S. Howell, M. Hush, M. Liuzzi, P. Mundada, T. Merkh, A. R. R. Carvalho, and M. J. Biercuk, Experimental deep reinforcement learning for error-robust gate-set design on a superconducting quantum computer, PRX Quantum **2**, 040324 (2021).

[33] J. Olle, R. Zen, M. Puviani, and F. Marquardt, Simultaneous discovery of quantum error correction codes and encoders with a noise-aware reinforcement learning agent, Npj Quantum Inf. **10**, 126 (2024).

[34] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, Proximal policy optimization algorithms, arXiv:1707.06347.

[35] V. Konda and J. Tsitsiklis, Actor-Critic Algorithms, in *Advances in Neural Information Processing Systems*, edited by S. Solla, T. Leen, and K. Müller (MIT Press, Cambridge, MA, USA, 1999), Vol. 12, https://proceedings.neurips.cc/paper_files/paper/1999/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

[36] H. Sowerby, Z. Zhou, and M. L. Littman, Designing rewards for fast learning, arXiv:2205.15400.

[37] E. Jeffrey, D. Sank, J. Y. Mutus, T. C. White, J. Kelly, R. Barends, Y. Chen, Z. Chen, B. Chiaro, and A. Dunsworth *et al.*, Fast accurate state measurement with superconducting qubits, Phys. Rev. Lett. **112**, 190504 (2014).

[38] T. Proctor, M. Revelle, E. Nielsen, K. Rudinger, D. Lobser, P. Maunz, R. Blume-Kohout, and K. Young, Detecting and tracking drift in quantum information processors, Nat. Commun. **11**, 5396 (2020).

[39] S. Hazra, W. Dai, T. Connolly, P. D. Kurilovich, Z. Wang, L. Frunzio, and M. H. Devoret, Benchmarking the readout of a superconducting qubit for repeated measurements, Phys. Rev. Lett. **134**, 100601 (2025).

[40] P. A. Spring, L. Milanovic, Y. Sunada, S. Wang, A. F. van Loo, S. Tamate, and Y. Nakamura, Fast multiplexed superconducting qubit readout with intrinsic Purcell filtering, arXiv:2409.04967 [quant-ph].

[41] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, http://github.com/google/jax JAX: composable transformations of Python+NumPy programs (2018).

[42] C. Lu, J. Kuba, A. Letcher, L. Metz, C. Schroeder de Witt, and J. Foerster, Discovered policy optimisation, Adv. Neural Inf. Process. Syst. **35**, 16455 (2022).

[43] S. Huang, R. F. J. Dossa, A. Raffin, A. Kanervisto, and W. Wang, in *ICLR Blog Track* (2022), https://iclr-blog-track.github.io/2022/03/25/ppo-implementation-details/.

[44] A. Bou, M. Bettini, S. Dittert, V. Kumar, S. Sodhani, X. Yang, G. De Fabritiis, and V. Moens, TorchRL: A data-driven decision-making library for PyTorch, arXiv:2306.00577.

[45] J. B. Curtis, I. Boettcher, J. T. Young, M. F. Maghrebi, H. Carmichael, A. V. Gorshkov, and M. Foss-Feig, Critical

theory for the breakdown of photon blockade, Phys. Rev. Res. **3**, 023062 (2021).

[46] D. I. Schuster, A. Wallraff, A. Blais, L. Frunzio, R. S. Huang, J. Majer, S. M. Girvin, and R. J. Schoelkopf, ac Stark shift and dephasing of a superconducting qubit strongly coupled to a cavity field, Phys. Rev. Lett. **94** (2005).

[47] J. Gambetta, A. Blais, D. I. Schuster, A. Wallraff, L. Frunzio, J. Majer, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Qubit-photon interactions in a cavity: Measurement-induced dephasing and number splitting, Phys. Rev. A **74**, 042318 (2006).

[48] P. Kidger, *On Neural Differential Equations*, Ph.D. thesis, school University of Oxford, 2021.

[49] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, and J. Bright *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in Python, Nat. Methods **17**, 261 (2020).