Multi-Stage Reinforcement Learning for Robust Charging of Quantum Batteries (Student Abstract)

Beomdo Park[1], Hyeonseok Jang[1], Junseong Park[1], Minu Baek[1], Gihun Gil[1], Minsung Jung[1], Woohyeon Kwon[1], Harin Jang[1], Yeojin Jang[1], Hoon Jeong[2], Taewook Heo[2], Sangkeum Lee[1]*

[1]Hanbat National University, South Korea
[2]Electronics and Telecommunications Research Institute (ETRI), South Korea
{pbeomdo, seokchu123, js03093351, bmw5779, minegihun, jmss1101, mfireon0520, hrjang713, jyeoj251}@gmail.com,
{hjeong, htw398}@etri.re.kr, sangkeum@hanbat.ac.kr*

Abstract

Quantum batteries have emerged as a next-generation energy storage solution, leveraging quantum phenomena such as superabsorption to overcome the limitations of conventional energy technologies. However, noise arising from interactions with the external environment degrades the charging efficiency and stability of the battery by disrupting the system's quantum coherence. To address this challenge, this study proposes a robust charging framework for a single-qubit quantum battery based on the Jaynes-Cummings (JC) model. The proposed framework combines the Proximal Policy Optimization (PPO) algorithm with a multi-stage reinforcement learning structure. The agent first learns fundamental control principles in a noise-free, ideal environment and subsequently performs robust learning in progressively noisier and more complex settings. Simulation results demonstrate that the trained agent navigates a stable charging trajectory on the Bloch sphere, thereby achieving high ergotropy even in the presence of noise. These findings suggest that multi-stage reinforcement learning is an effective solution for control problems in noisy quantum systems and provides a theoretical foundation for designing charging protocols for multi-qubit systems.

Introduction

Quantum batteries promise to surpass classical batteries by utilizing quantum effects like coherence and entanglement, but their performance is severely hampered by environmental noise causing decoherence (Quach, Cerullo, and Virgili 2023). This leads to energy loss and unstable charging, reducing the usable energy, or ergotropy (Alicki and Fannes 2013).

Reinforcement learning (RL) has emerged as a powerful tool for discovering optimal control protocols in quantum systems. (Ferraro et al. 2018). However, training an RL agent directly in a noisy environment is inefficient and often leads to suboptimal policies due to the non-convex reward landscape. To overcome this, we propose a robust charging framework combining Proximal Policy Optimization (PPO) with a multi-stage training framework (Schulman et al. 2017). This multi-stage approach guides the agent's exploration from a simple to a complex problem space. By first mastering control in a noise-free environment, the agent develops a foundational policy that is then fine-tuned in progressively noisier settings. This staged approach allows the agent to develop a robust policy that avoids local optima and efficiently adapts to complex system dynamics. Our simulations on a single-qubit JC model validate this framework, demonstrating stable and efficient charging even under high-noise conditions. Recent studies show quantum(-inspired) RL is effective for continuous-state energy control (Nengroo et al. 2025)

*Corresponding author.

Methodology

Quantum System Model

Our system is a single-qubit quantum battery modeled by the Jaynes-Cummings (JC) Hamiltonian, which includes terms for the cavity (charger), the qubit (battery), and their interaction. Specifically, we model three primary noise channels: atomic decay (γ_e), dephasing (γ_p), and photon loss (κ). The agent's objective is to discover a time-dependent control pulse that maximizes the final ergotropy—the energy extractable via unitary transformations—while actively mitigating these detrimental effects.

PPO for Multi-Stage Quantum Control

To manage a quantum environment characterized by progressively increasing noise levels, we employ the PPO algorithm, which is an actor-critic method particularly well-suited for continuous control tasks.

- **State (S_t):** A vector of physical observables derived from the density matrix $\rho(t)$, including energy, ergotropy, photon count, and spin expectation values ($\langle J_x \rangle, \langle J_y \rangle, \langle J_z \rangle$).
- **Action (A_t):** The real and imaginary amplitudes of the external control pulse applied at each timestep.
- **Reward (R_t):** A shaped reward function defined as a weighted sum of the incremental increase in ergotropy and the final ergotropy achieved, encouraging high efficiency charging.

Category	Hyperparameter	Value
Physical System	Photon Number (N_{photons})	8
	Qubit/Cavity Freq. (ω_q, ω_c)	1.0
	Coupling Strength (g)	0.1
RL Environment	Total Charging Time (T)	20.0
	Number of Time Steps	100
	Max Control Amplitude ($A_{\max}$)	0.3
PPO Algorithm	Learning Rate (η)	Linear Decay
	Steps (per update) / Batch Size	2048 / 128
	Discount (γ)/GAE Lambda (λ)	1.0 / 0.95
	Epochs / Clip Range	10 / 0.1

Table 1. Key hyperparameters for the RL model.

PPO's signature clipped objective function prevents excessive policy updates, which is critical for maintaining stability when adapting a policy to a new, more complex noise environment. A linearly decaying learning rate is also employed to ensure smooth convergence during fine-tuning at each stage.

Experiments and Results

Key hyperparameters for the simulation are detailed in Table 1. The final target environment featured significant noise, with rates for atomic decay (γ_e), dephasing (γ_p), and photon loss (κ) set to 0.05, 0.025, and 0.05, respectively.

As shown in Figure 1, the agent achieves a final ergotropy for the given noise level, outperforming simple, static pulse strategies. The learned control pulse exhibits a complex form. The real component primarily drives energy transfer, while the imaginary component actively modulates the qubit's phase to counteract decoherence, showcasing an advanced coherent control strategy.

In the JC model, a qubit is described as a spin particle,

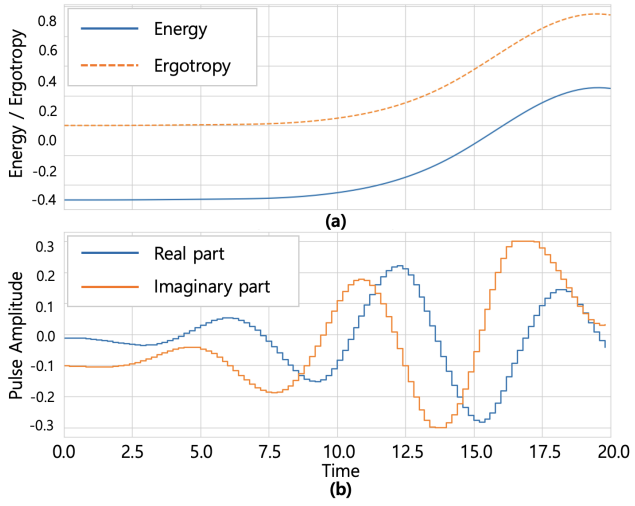

Figure 1: Charging dynamics of the trained agent. (a) Evolution of total energy and ergotropy. (b) Real (Re) and imaginary (Im) parts of the learned control pulse.

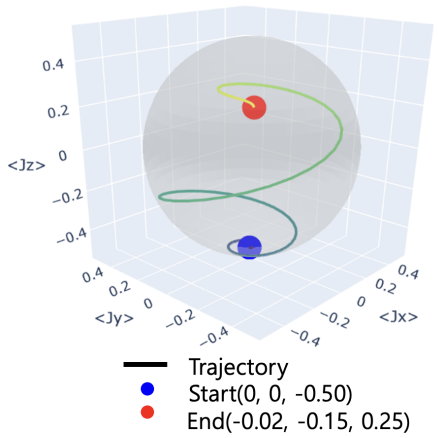

Figure 2: State trajectory of the qubit on the Bloch sphere. The path evolves from the blue point (start) to the red point (end), achieving 75% of the maximum ergotropy.

where the south and north poles of the Bloch sphere correspond to the ground state ($|g\rangle$, discharged) and the excited state ($|e\rangle$, fully charged), respectively. The state vector of the qubit simultaneously undergoes two distinct dynamics. First, governed by the qubit(battery)'s intrinsic Hamiltonian, the state exhibits continuous Larmor precession around the z-axis. Second, an external control pulse, mediated by the interaction with a cavity, induces an energy-charging process akin to a Rabi-like oscillation, driving the qubit state from the south pole towards the north pole. Figure 2 visualizes these composite dynamics, illustrating the quantum state's transition to the charged state along a spiral trajectory, which results from the superposition of its precession around the z-axis and the concurrent energy absorption.

Conclusion and Future Work

We have presented a robust charging framework for quantum batteries using a multi-stage PPO, which overcomes the inefficiencies of direct training in noisy environments. Furthermore, the policy learned in our simulated Markovian environment can serve as a powerful starting point for fine-tuning in more realistic non-Markovian environments characteristic of actual quantum hardware, accelerating the discovery of practical control protocols.

Acknowledgments

This work was partly supported by Korea Evaluation Institute of Industrial Technology(KEIT) grant funded by the Korea government(MOTIE) (No.RS-2025-04752989, Quantum battery core technology for ultra-fast charging 100x faster than a traditional lithium-ion batteries)

References

Alicki, R.; and Fannes, M. 2013. Entanglement boost for extractable work from ensembles of quantum batteries. *Physical Review E*, 87(4): 042123.

Ferraro, D.; Campisi, M.; Andolina, G. M.; Pellegrini, V.; and Polini, M. 2018. High-Power Collective Charging of a Solid-State Quantum Battery. *Physical Review Letters*, 120(11): 117702.

Nengroo, S. H.; Har, D.; Jeong, H.; Heo, T.; and Lee, S. 2025. Continuous variable quantum reinforcement learning for HVAC control and power management in residential building. *Energy and AI*, 21: 100541.

Quach, J. Q.; Cerullo, G.; and Virgili, T. 2023. Quantum batteries: The future of energy storage? *Joule*, 7(10): 2195–2200.

Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347*.