

# EcoAI 랩세미나

## 양자배터리 팀 논문리뷰

○ 2025.09.05 ○

Presenter: 장현석, 박범도 | 소속: EcoAI Lab | E-mail: [seokchu123@gmail.com](mailto:seokchu123@gmail.com)  
[pbeomdo@gmail.com](mailto:pbeomdo@gmail.com)

팀원: [박범도, 박준성, 장현석]

# Contents

**1-1** Introduction

**1-2** Methodology

**1-3** Experiment and Result

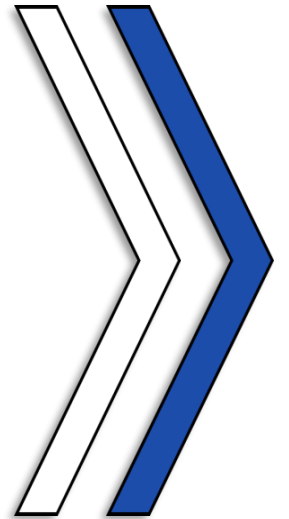
**2** AAI Student Abstract 진행 상황 공유

# 01

EcoAI 랩세미나 - 양자배터리 팀 논문리뷰

## Paper Review (발표: 장현석)

Robust quantum control using reinforcement learning from demonstration  
(Shengyong Li et al.)



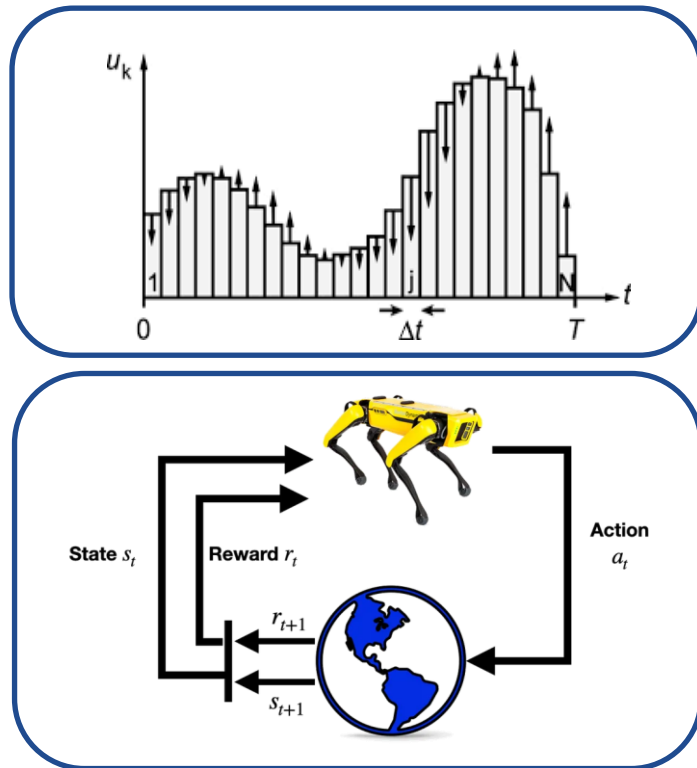
## Motivation

- **Model bias**
  - Model-based 방식 (GRAPE, krotov 등)은 이상적인 시스템 모델 (해밀토니안)을 바탕으로 정밀한 펄스 제어 가능.
  - 물리적 환경에서 외부 요인으로 인한 노이즈 발생 시, 시스템 모델과의 오차 → 편향 (bias) 발생.
- **Sample Inefficiency**
  - Model-free RL은 모델 기반 방식에서 발생하는 편향을 회피할 수 있음.
  - 학습 수렴에 필요한 샘플을 모으는데 시간이 너무 많이 걸림.
  - 초기의 무작위 탐색으로 인한 비효율적인 방식.

# 1-1 Introduction

## Proposed Method

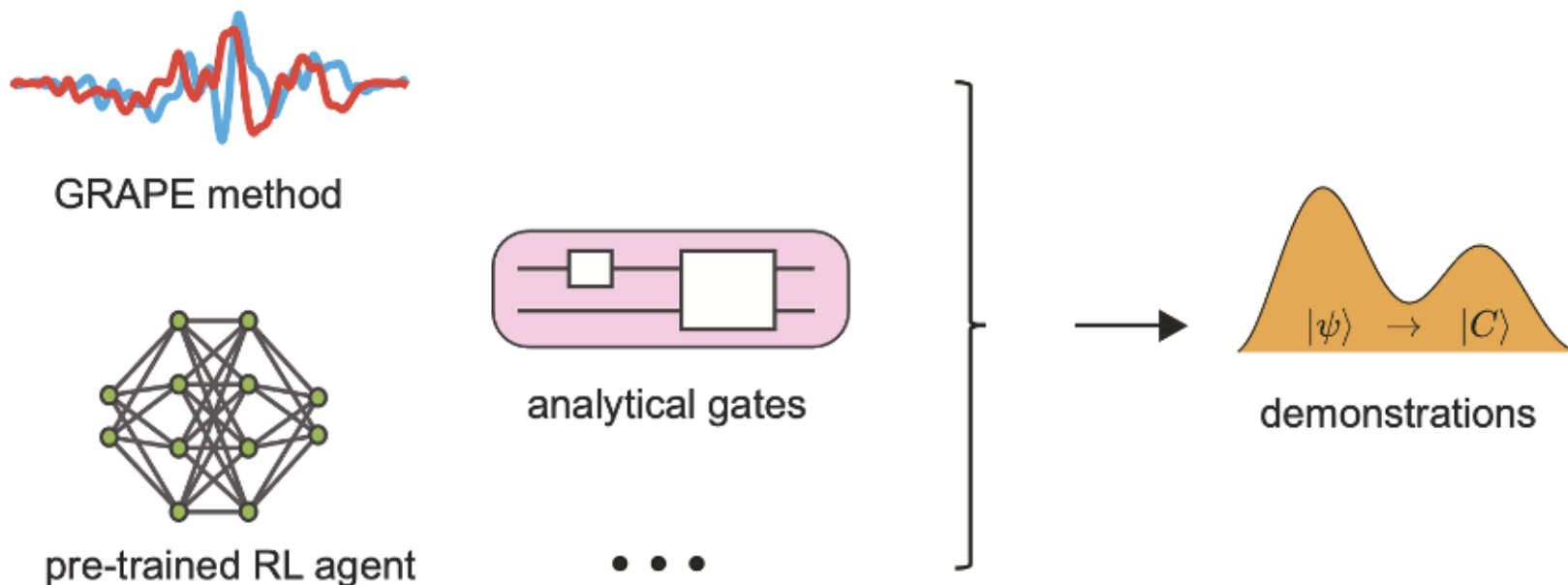
- (model-based) 시스템 모델만으로 연산 결과를 빠르게 출력.
- (model-free) 환경과의 상호작용으로 모델 편향 해소 및 정밀한 연산 수행.



Reinforcement Learning  
from Demonstration

### RLfD (Reinforcement Learning from Demonstrations)

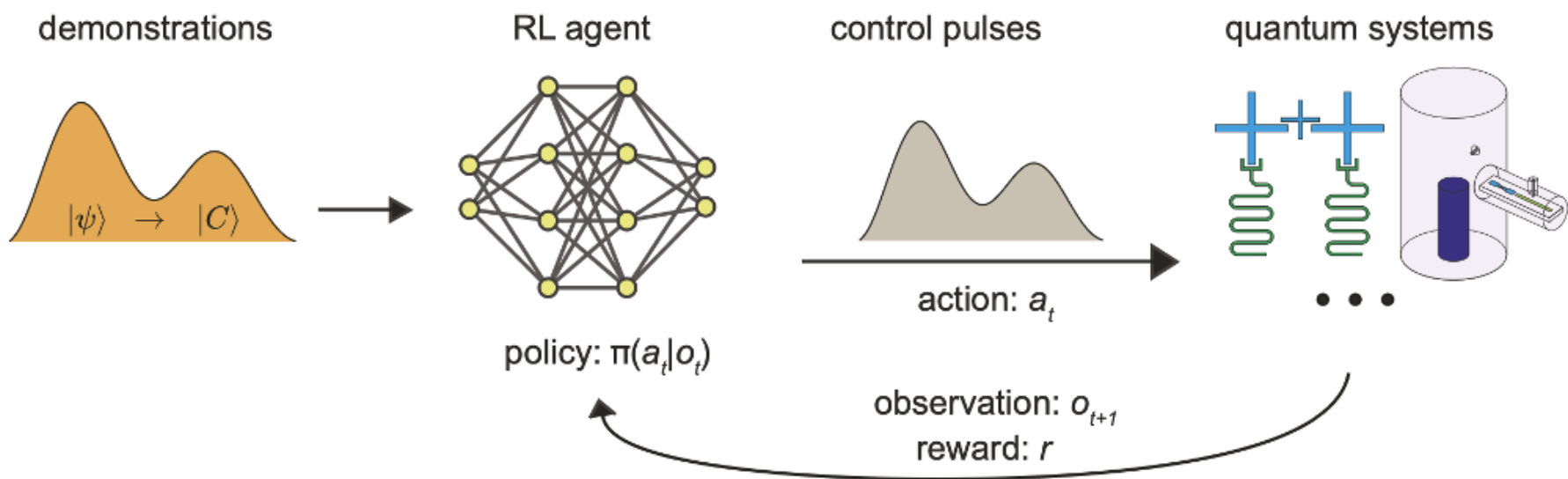
- Demonstrations 생성
  - GRAPE method: 해밀토니안을 기반으로, 최적의 제어 펄스를 빠르게 생성.
  - Pre-trained RL agent: 본 문제보다 더 간단한 저차원 문제 (게이트 문제)에서 미리 학습시킨 에이전트를 활용.
  - Analytical Gates: 물리 이론에 따라 이미 정립된 이상적인 양자 연산 방식을 활용.



### RLfD (Reinforcement Learning from Demonstrations)

- RLfD 학습 과정

- 초기화: Demonstrations를 바탕으로 RL 에이전트의 초기 정책을 설정.
- 행동: RL 에이전트가 현재 정책에 따라 최적의 제어 펄스를 생성.
- 상호작용: 생성된 제어 펄스를 실제 양자 시스템에 적용하여 상태를 변화시킴.
- 피드백 및 학습: 시스템을 측정하여 얻은 관측 및 보상 정보를 바탕으로, RL 에이전트가 자신의 정책을 더 나은 방향으로 피드백하고 이 과정을 반복.

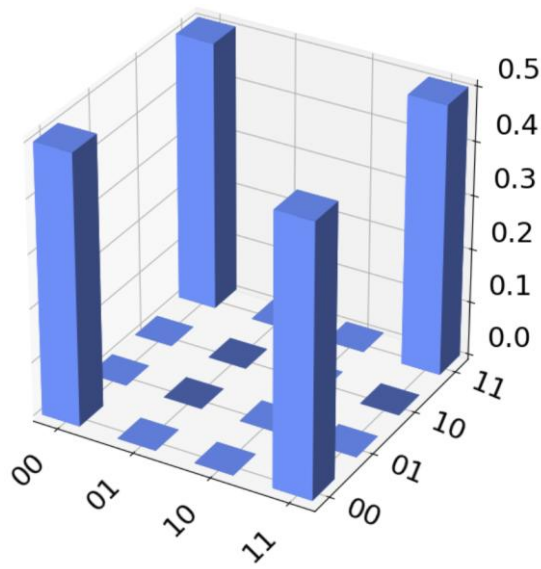


# 1-3 Experiment and Result

## Experiment Setting

### ❖ a. Bell state (2큐비트 얽힘)

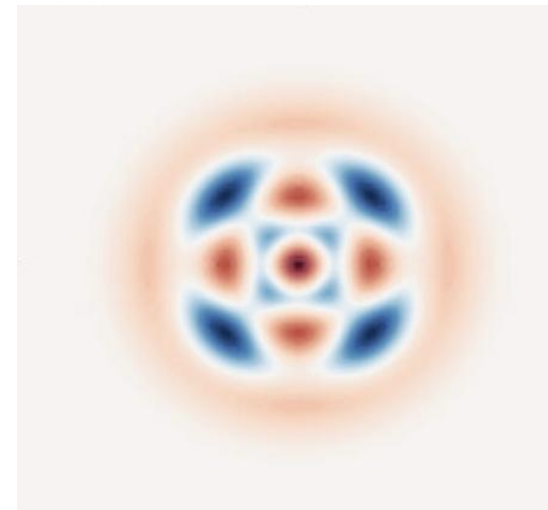
- 기본적인 얽힘 상태
- RLfD의 학습 효율성 및 안정성 검증



a. Bell state

### ❖ b. Binomial state

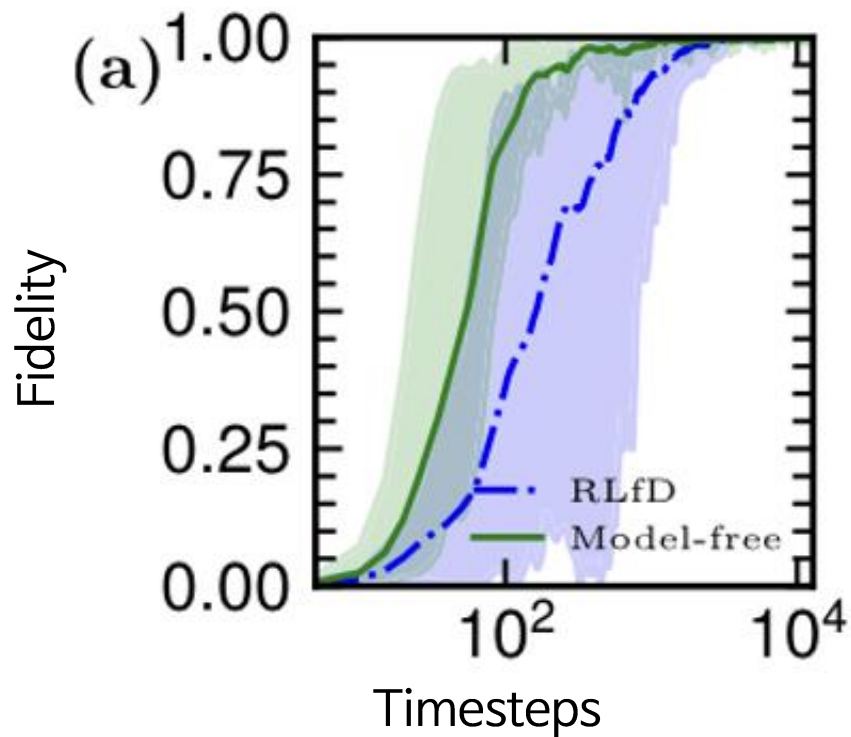
- 공진기 안의 이산적인 조합으로 만들어지는 특수한 양자 상태
- 양자 오류 보정에 활용 가능
- 복잡한 조건에서도 높은 충실도(Fidelity) 달성



b. binomial state



### Experiment Result



(a) 기존 Model-free RL의 성능과 RLfD 방법의 성능 비교

- RLfD는 0.995 Fidelity 도달에 1340 timesteps 필요
- Model-free RL은 같은 기준에 3060 timesteps 필요

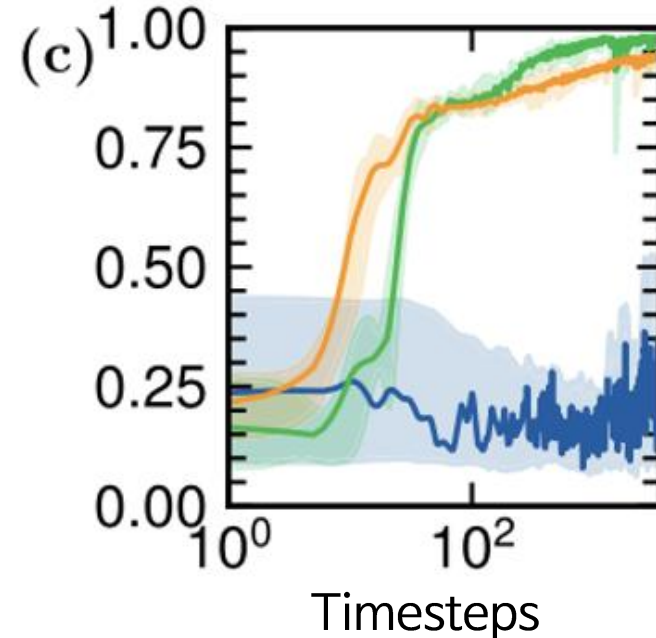
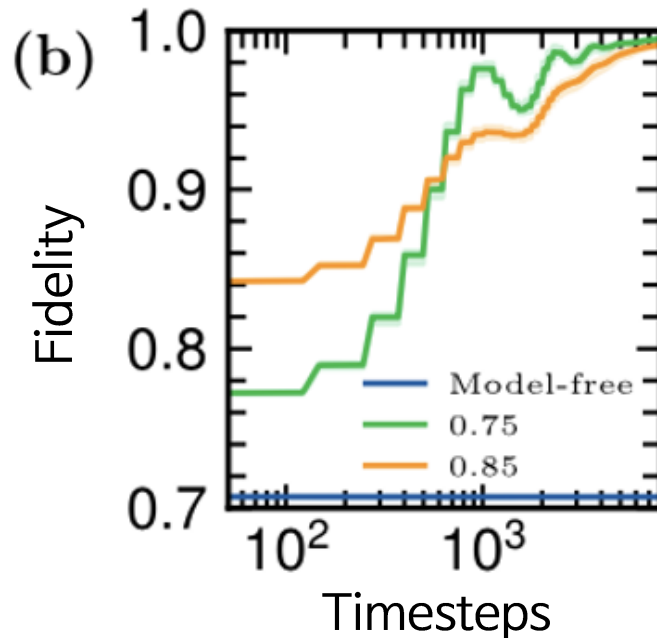
→ RLfD가 Model-free RL보다 훨씬 빠르게 높은 Fidelity에 도달함

# 1-3 Experiment and Result

## Experiment Result

### (b, c) Binomial state 준비 과정

- (b) PPO (Proximal Policy Optimization), (c) SAC (Soft Actor-Critic) 알고리즘을 적용했을 때의 학습 성능 비교.
- 초기 시범 Fidelity (0.75, 0.85)에 관계없이 최종적으로 높은 Fidelity 달성  
→ RLfD 학습이 더 안정적임을 보여줌.



# 1-3 Experiment and Result

## Experiment Result

### SAC vs PPO 성능 분석

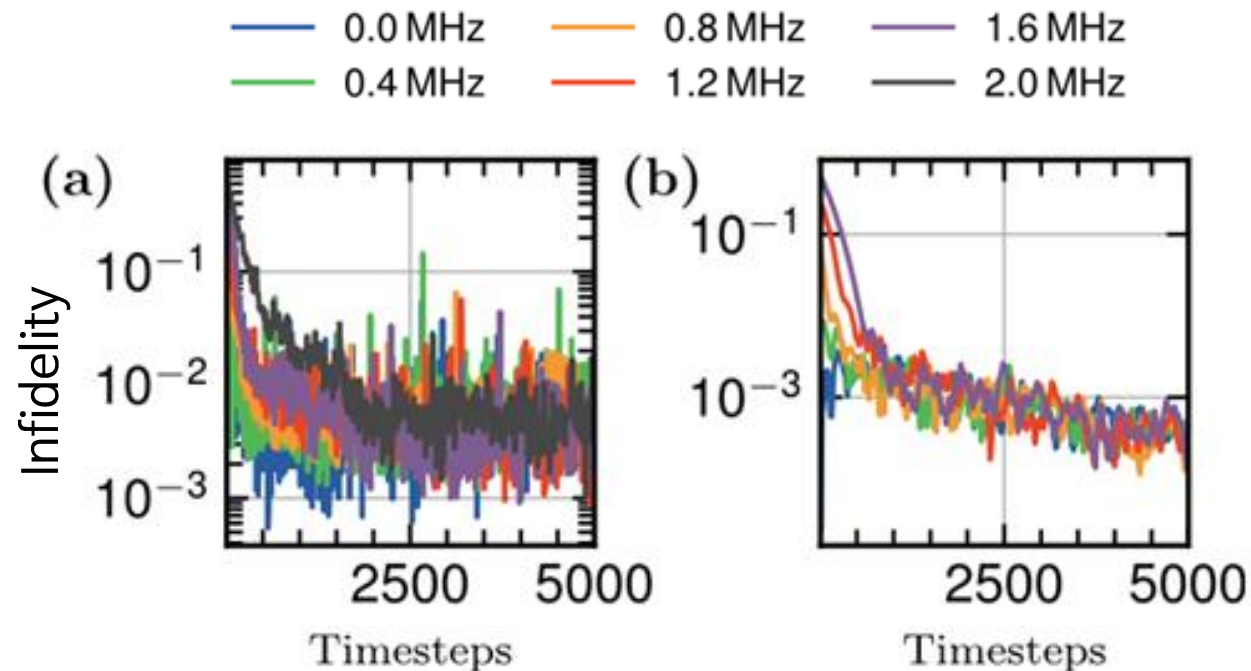
#### (a) SAC

- 학습 속도 빠름, 샘플 효율 높음
- 안정성 낮고 진동 발생 가능

#### (b) PPO

- timesteps 증가 → infidelity 감소
- 학습은 느리지만 안정적
- 최종적으로 더 높은 Fidelity (낮은 Infidelity) 달성

→ RLfD는 복잡한 양자 제어 문제에도 적용 가능  
→ SAC와 PPO는 문제 특성에 따라 선택적으로 활용할 수 있음

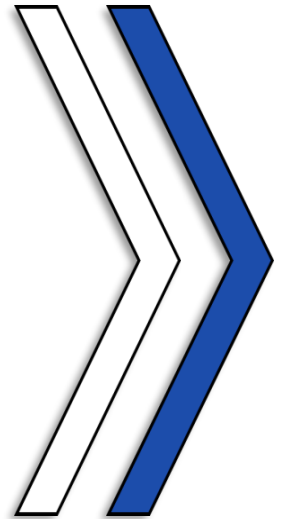


# 02

EcoAI 랩세미나

## AAAI Student Abstract 진행 상황 공유

Robust Charging Framework for Quantum Batteries Using Transfer Reinforcement Learning



## ● 주제: 강화학습을 이용한 개방계 양자 시스템의 충전 최적화

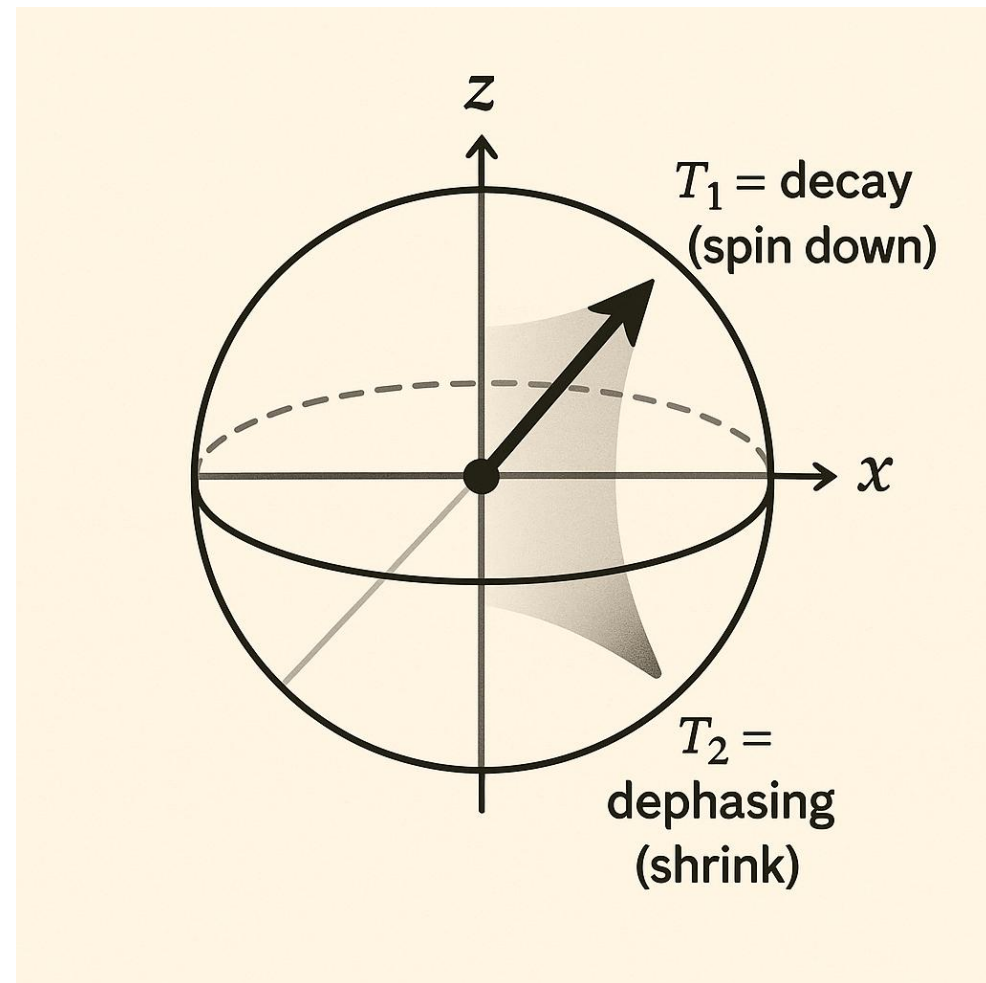
- 개방계 양자배터리 충전시 **노이즈의 문제 존재** - Decoherence(결잃음)
- $T_1$ : Decay(바닥 상태로 되돌아가려는 에너지 감소-z축 이동)
- $T_2$ : Dephasing (에너지는 잃지 않으면서 위상 정보 손실-xy평면 이동)

## ● 문제점

- 노이즈는 양자배터리의 충전 효율성, 안정성 저해.
- 사용 가능 에너지인 Ergotropy를 감소시킴.

## ● 목표

→ 노이즈가 존재하는 개방형 양자 시스템 환경에서 단일 큐비트 **양자배터리의 Ergotropy를 최대화하는 최적의 제어 프로토콜**을 어떻게 찾을 것인가?



디코히런스 ( $T_1$ : 에너지 손실) ( $T_2$ : 위상 손실)

### ● 기존 접근법의 한계

- 양자 게이트 등 Static pulse 제어 방식은 노이즈에 취약함
- 강화학습을 노이즈 환경에 직접 적용하는 것은 비효율적, 복잡한 보상 공간 때문에 발생하는 Local Optima 문제

### ● 연구 목표

- 다중 스테이지를 통한 에이전트의 정책 전파를 활용
- 노이즈 강건성을 가지는 제어 프로토콜 최적화를 제안

## Multi-Stage Reinforcement Learning for Robust Charging of Quantum Batteries

Anonymous submission

### Abstract

Quantum batteries have emerged as a promising 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 transfer learning-based 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 successfully navigates a stable charging trajectory on the Bloch sphere, thereby achieving high ergotropy even in the presence of noise. These findings suggest that transfer learning-based reinforcement learning is an effective solution for control problems in noisy quantum systems and provides a theoretical foundation for designing charging protocols for future multi-qubit systems.

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 (Zhao, Wang, and et al. 2023). Our simulations on a single-qubit Jaynes-Cummings (JC) model validate this framework, demonstrating stable and efficient charging even in high-noise conditions.

### 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. The system's evolution in a noisy, open-world environment is simulated using the Lindblad master equation. This formalism correctly captures both the coherent dynamics controlled by the RL agent and the incoherent noise processes that cause decoherence. Specifically, we model three primary noise channels: atomic decay ( $\gamma_c$ ), dephasing ( $\gamma_p$ ), and photon loss ( $\kappa$ ). The agent's objective is to discover a

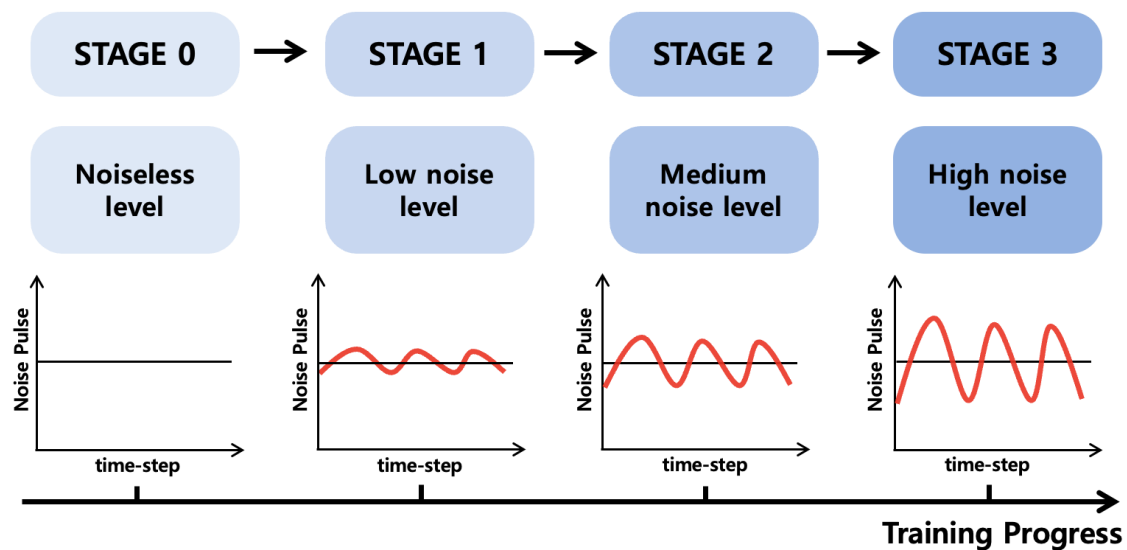
작성 중인 논문 초안

### ● 다단계 강화 학습

- 에이전트가 이상적인 환경 (Noise=0)에서 기초 제어 원리를 먼저 학습한 후
- 학습된 정책을 노이즈가 점차 강해지는 환경으로 전이시켜 적응 및 고도화

### ● 기반 모델:

- 시스템 모델: Jaynes-Cummings (JC) Hamiltonian
- 강화학습: PPO (상태: 물리 관측량, 행동: 제어 펄스, 보상: Ergotropy)



다단계 강화 학습 예시

### ● 결과1: 충전 펄스

#### • (a) 에너지 및 어고트로피 진화

노이즈 환경에서도 높은 어고트로피(사용 가능 에너지)를 달성하였음. 총 에너지와 어고트로피 간의 차이는 노이즈로 인한 결맞음 붕괴를 의미.

#### • (b) RL 에이전트가 학습한 제어 펄스

실수부 (파란선): 배터리에 에너지 주입하여 라모어 세차 유도. Z축으로 에너지를 밀어 올리는 힘

허수부 (주황선): T2의 dephasing 과정을 상쇄하기 위한 정교한 위상 제어를 담당. xy평면 안으로 붕괴하는 것을 억제

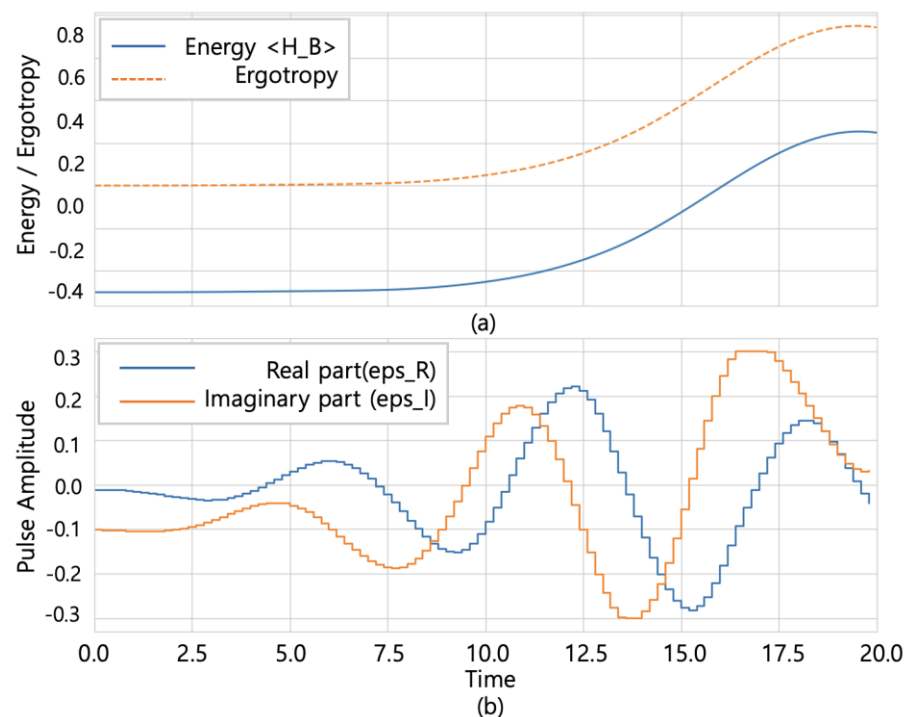


Fig 1. Charging dynamics



### ● 결과2: 블로흐 구면 상의 궤적

- 큐비트상태가방전 ( $|0\rangle$ )에서충전( $|1\rangle$ )방향으로이동
- 나선형 경로는라모어세차에 의한능동적인 위상 제어의 증거다.
- End 지점이 구 표면이 아닌 내부에 위치해 노이즈로 인한 순수도(Purity) 손실을 의미.

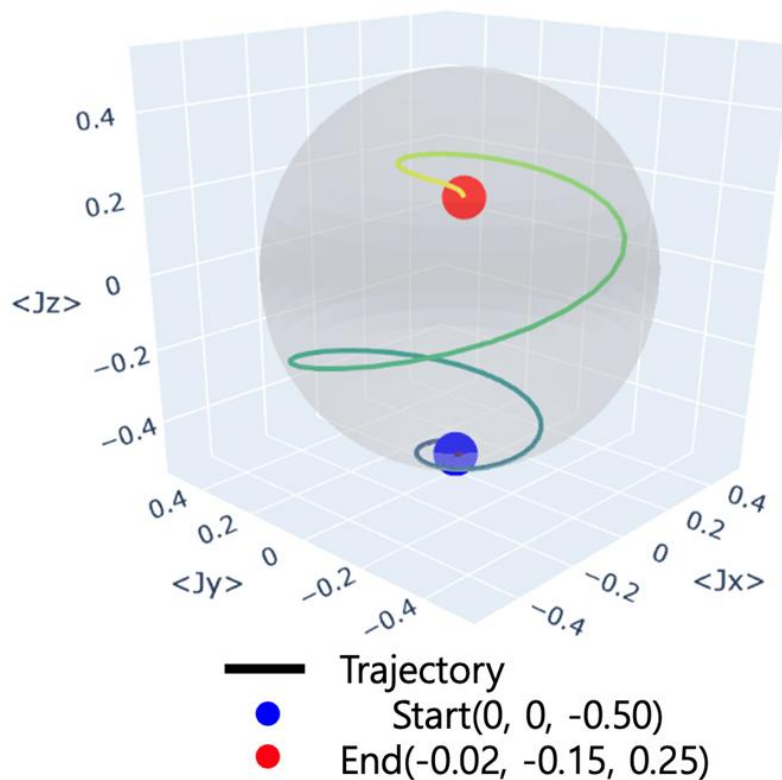


Fig 2. State trajectory of the qubit on the Bloch sphere.

# 감사합니다

---

○ 2025.09.05 ○

Presenter: 박범도 | 소속: EcoAI Lab | E-mail: [pbeomdo@gmail.com](mailto:pbeomdo@gmail.com)

팀원: [박범도, 박준성, 장현석]