단계적 강화학습 프레임워크를 통한 강건한 양자 배터리 충전

박준성 [1], 장현석 [1], 박범도 [1], 정훈 [2], 허태욱 [2,] *이상금 [1]
*국립한밭대학교 [1], 한국전자통신연구원 [2]

{js0309, seokchu123, pbeomdo}@gmail.com, {hjeong, htw398}@etri.re.kr,
sangkeum@hanbat.ac.kr

Progressive Reinforcement Learning Framework for Robust Quantum Battery Charging

Junseong Park[1], Hyeonseok Jang[1], Beomdo Park[1], Hoon Jeong[2], Taewook Heo[2]
and *Sangkeum Lee[1]
*Hanbat National University[1], Electronics and Telecommunications Research Institute[2]

요 약

양자 배터리는 결맞음과 얽힘 같은 양자 효과를 활용하여 기존 화학 배터리를 능가할 잠재력을 지닌다. 그러나 환경 노이즈로 인한 결어긋남은 충전 효율을 저하시켜 실질적 구현을 어렵게 만든다. 이러한 한계를 극복하기 위해 Proximal Policy Optimization (PPO) 알고리즘과 커리큘럼 학습 전략을 결합한 강건한 양자 배터리 충전 프레임워크를 제안한다. 단일 큐비트 기반 Jaynes-Cummings (JC) 모델을 적용하고, Lindblad 마스터 방정식을 통해 원자 붕괴, 위상 이탈, 광자 손실이 포함된 개방형 양자 환경을 시뮬레이션한다. 제안된 프레임워크는 노이즈가 없는 환경에서 학습을 시작해 점진적으로 노이즈 강도를 높이며 복잡한 양자 동역학에 안정적으로 적응하도록 설계된다. 시뮬레이션 결과, 강화학습 기반 제어는 단순 정적 제어보다 결맞음을 유지하며 높은 에르고트로피와 에너지 축적 효율을 달성한다. 향후 다중 큐비트 Tavis-Cummings 모델로 확장되어 초흡수와 같이 집단적 양자 효과를 탐구함으로써, 고효율 양자 에너지 저장 기술의 상용화에 기여할 것으로 기대된다.

Ⅰ. 서 론

양자 배터리는 결맞음 및 얽힘과 같은 고유한 양자 효과를 활용하여 기존 화학 배터리 성능을 제공할 잠재력을 지닌다. 그러나 양자 시스템의 내재적 취약성으로 인해 발생하는 환경 노이즈는 결어긋남을 유발하여 그 성능을 심각하게 저해하고, 시스템에서 추출 가능한 유용한 에너지인 에르고트로피를 감소시킨다[1]. 따라서 노이즈 환경에 강건하고 효율적인 충전 프로토콜 개발이 필요하다. 강화학습은 양자 시스템에서 최적 제어 프로토콜을 탐색하는 효과적인 전략으로 활용되고 있다[2]. 노이즈 환경에서는 비볼록한 보상 지형으로 인해 학습이 불안정해지고 에이전트가 전역 최적해 대신 국소 최적해에 수렴하는 한계를 가진다. 본 논문은 이러한 문제를 해결하기 위해 Proximal Policy Optimization (PPO)과 커리큘럼 학습 전략을 결합한 강건한 충전 프레임워크를 제안한다. 커리큘럼 학습은 에이전트가 노이즈가 없는 단순한 환경에서 학습을 시작해 점진적으로 노이즈가 강화된 환경으로 전이하도록 설계되어, 복잡한 양자 동역학에 대한 적응력을 높이고, 국소 최적해에 빠지는 문제를 완화한다. 이를 통해 제안된 프레임워크는 노이즈 환경에서도 높은 에르고트로피를 유지하는 강건하고 효율적인 제어 정책을 학습할 수 있음을 기대한다.

Ⅱ. 본론

2.1 양자 배터리 시스템의 물리 모델링

단일 큐비트 기반 Jaynes-Cummings (JC) 모델을 사용하여 양자 배터리 시스템을 구성한다. 노이즈가 존재하는 개방형 환경에서의 시간 진화는 Lindblad 마스터 방정식을 통해 시뮬레이션 되며, 이를 통해 강화학습 에이전트가 제어하는 유니터리 동역학(unitary dynamics)과 환경 노이즈로 인해 발생하는 비유니터리(non-unitary) 과정을 동시에 반영할 수 있다. 본 연구에서는 원자 붕괴, 위상 이탈, 광자 손실의 세 가지 주요 노이즈 채널을 고려한다. 에이전트의 목표는 이러한 노이즈 효과를 최소화하면서, 유니터리 변환을 통해 시스템으로부터 추출 가능한 유용한 에너지인 에르고트로피를 최대화하는 시간 의존 제어 펄스를 학습하는 것이다. 강화학습 환경과 알고리즘의 구성은 JC 모델의 물리 파라미터, 학습 환경 설정 및 PPO 알고리즘의 주요 하이퍼파라미터로 정의되며, 그 세부 내용은 표 1 과 같다. 이러한 구성은 시스템의 양자 동역학을 정밀하게 반영하면서도 학습 안정성을 확보하여 양자 배터리 충전 과정의 효율적 제어를 가능하게 한다.

2.2 강화학습 기반 프레임워크 설계

점진적으로 노이즈 강도가 증가하는 환경에서 효율적인 제어 정책을 학습하기 위해, 연속 제어 문제에 적합한 Actor-Critic 구조의 강화학습 알고리즘인 PPO 를 적용한다. PPO 의 클리핑(clipping) 기법은 급격한 정책 업데이트를 억제하여 학습의 안정성을 유지하며, 정책이 더 복잡한 노이즈 환경으로 전이될 때 안정적인 성능 유지에 핵심적인 역할을 한다. 또한 선형적으로 감소하는 학습률을 도입하여 각 단계의 미세조정(fine-tuning) 과정에서 안정적인 수렴을 유도한다. 이러한 커리큘럼 학습 절차는 노이즈가 없는 환경에서 학습된 초기 정책을 기반으로, 점차 노이즈가 강화된 조건으로 학습 환경을 전이시키는 방식으로 구성된다. 이를 통해 에이전트는 복잡한 양자 동역학에 점진적으로 적응하면서 국소 최적해에 빠지는 문제를 효과적으로 완화할 수 있다[3].

구분	하이퍼파라미터	값
물리 시스템	Photon Number (N_{photons})	8
	Qubit/Cavity Freq. (ω_q, ω_c)	1.0
	Coupling Strength (g)	0.1
강화학습 환경	Total Charging Time (T)	20.0
	Number of Time Steps	100
	Max Control Amplitude ($A_{\max}$)	0.3
PPO 알고리즘	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

표 1. 강화학습 모델의 주요 하이퍼파라미터

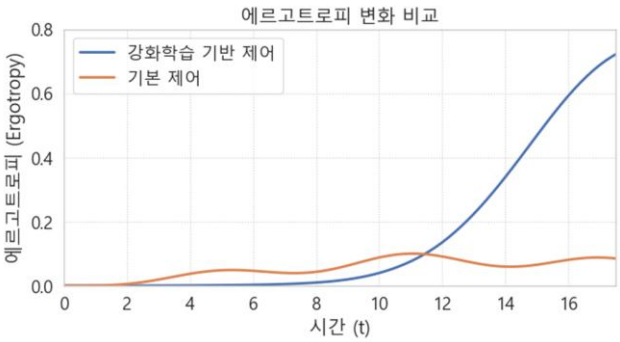

그림 1. 강화학습 기반 제어와 기본 제어의 에르고트로피 변화 비교

2.3. 시뮬레이션 및 결과 분석

제안된 프레임워크는 단일 큐비트 양자 배터리의 충전 과정에 적용하며, 최종 목표 환경은 원자 붕괴율 0.05, 위상 이탈률 0.025, 광자 손실률 0.05 의 높은 노이즈 조건으로 설정한다. 훈련된 에이전트는 이론적 최대값에 근접한 에르고트로피를 달성하며, 단순한 정적 펄스 제어보다 우수한 성능을 보인다. 그림 1 은 시간에 따른 에르고트로피 변화를 비교한 결과를 나타낸다. 기본 제어는 위상 이탈 및 감쇠 효과로 인해 일시적인 상승 이후 성능이 불안정해지는 반면, 강화학습 기반 제어는 커리큘럼 학습을 통해 환경 노이즈에 적응함으로써 결맞음을 유지하고 에너지 충전 효율을 향상시킨다. 큐비트는 JC 모델에서 스핀 입자로 표현되며, 블로흐 구면 상에서 남극과 북극은 각각 바닥상태와 들뜬상태에 해당한다. 이때 큐비트의 상태 벡터는 두 가지 운동을 동시에 수행한다. 첫째, 고유 해밀토니안에 의해 Z 축을 중심으로 끊임없이 회전하는 라모어 세차(Lamor precession)가 발생하며, 둘째, 외부 제어 펄스와의 상호작용을 통해 라비 진동(Rabi oscillation)이 유도되어 에너지가 흡수된다. 그림 2 는 큐비트의 동역학을 블로흐 구면 상에 시각화한 결과이다. 강화학습 에이전트가 생성한 제어 펄스는 큐비트의 라모어 세차와 라비 진동을 최적으로 제어하여, 양자 상태가 나선형 궤적을 따라 노이즈의 영향 속에서도 최종 완충 상태로 안정적으로 유도되는 과정을 보여준다.

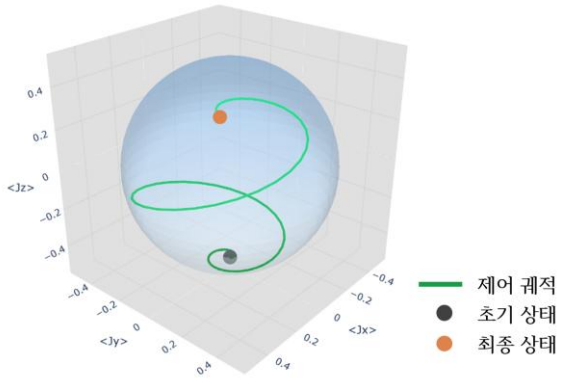

그림 2. 블로흐 구면 상에서 큐비트의 상태 궤적

Ⅲ. 결론

본 연구에서는 PPO 알고리즘과 커리큘럼 학습 전략을 결합하여 노이즈 환경에서도 강건하게 동작하는 양자 배터리 충전 프레임워크를 제안한다. 커리큘럼 학습을 통해 에이전트는 노이즈가 없는 환경에서 학습을 시작하여 점진적으로 노이즈가 강화된 조건으로 전이하며 복잡한 양자 동역학에 안정적으로 적응할 수 있다. 시뮬레이션 결과, 학습된 제어 펄스는 노이즈 환경에서도 결맞음을 유지하며 에너지 축적 효율을 향상시켜 제안된 프레임워크의 유효성을 입증한다. 향후 본 프레임워크를 다중 큐비트 Tavis-Cummings 모델로 확장하여 초흡수와 같은 집단적 양자 효과를 탐구하여, 고효율 양자 에너지 저장 기술의 구현과 양자 배터리 상용화에 기여하고자 한다.

ACKNOWLEDGMENT

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 traditional lithium-ion batteries)

참 고 문 헌

[1] Jiang, C., Pan, Y., Wu, Z.-G., Gao, Q., & Dong, D. (2022). Robust optimization for quantum reinforcement learning control using partial observations. *Physical Review A, 105*, 062443.

[2] Erdman, P. A., Andolina, G. M., Giovannetti, V., & Noé, F. (2024). Reinforcement learning optimization of the charging of a Dicke quantum battery. *Physical Review Letters, 133*(24), 243602.

[3] Nengroo, S. H., Har, D., Jeong, H., Heo, T., & Lee, S. (2025). Continuous variable quantum reinforcement learning for HVAC control and power management in residential building. Energy and AI, 21, 100541.