

개방 양자계 환경에서의 양자 배터리 충전 프로토콜 최적화를 위한 커리큘럼 강화학습 프레임워크

박준성¹, 장현석¹, 박범도¹, 정훈², 허태욱², *이상금¹
*국립한밭대학교¹, 한국전자통신연구원²

{js03093351, seokchu123, pbeomdo}@gmail.com, {hjeong, htw398}@etri.re.kr,
sangkeum@hanbat.ac.kr

Curriculum Reinforcement Learning Framework for Optimizing Quantum Battery Charging Protocols in Open Quantum Systems

Junseong Park¹, Hyeonseok Jang¹, Beomdo Park¹, Hoon Jeong², Taewook Heo² and
*Sangkeum Lee¹

*Hanbat National University¹, Electronics and Telecommunications Research Institute²

요 약

집단적 다중 큐비트 양자 배터리를 대상으로 개방 양자계 환경에서도 강건한 충전 성능을 달성하기 위한 커리큘럼 기반 강화학습 제어 프레임워크를 제안한다. 다중 큐비트가 단일 캐비티 모드와 집단적으로 결합된 TC (Tavis-Cummings) 모델을 물리적 배경으로 사용하고, 원자 붕괴, 위상 이탈, 광자 손실을 포함한 Lindblad 마스터 방정식을 통해 개방계 동역학을 기술한다. 충전 제어 문제는 연속 상태와 연속 행동을 갖는 강화학습 문제로 정의되며, 보상 함수는 실제로 추출 가능한 에너지를 반영하는 에르고트로피 증가량을 기준으로 설계된다. 정책 학습에는 PPO (Proximal Policy Optimization) 알고리즘을 적용하고, 노이즈가 없는 단일 큐비트 환경에서 시작하여 큐비트 수와 노이즈 강도를 단계적으로 증가시키는 커리큘럼 학습 전략을 도입한다. 시뮬레이션 결과, 제안된 커리큘럼 기반 PPO는 무작위 제어 및 고정 구동 방식 대비 에르고트로피를 빠르게 증가시키며, 노이즈가 존재하는 다중 큐비트 환경에서도 안정적인 충전 성능을 유지한다. 이러한 결과는 커리큘럼 학습이 집단적 양자 배터리 충전 제어에서 효과적인 강화학습 전략이 될 수 있음을 보여준다.

I. 서 론

양자 배터리는 결맞음과 얽힘과 같은 양자역학적 특성을 활용하여 기존 고전적 에너지 저장 장치보다 빠른 충전 속도와 높은 에너지 효율을 달성할 수 있는 차세대 에너지 저장 개념으로 주목받고 있다. 특히 다중 큐비트가 공통의 모드와 집단적으로 결합된 양자 배터리는 초흡수와 같은 집단적 양자 효과를 통해 단일 큐비트 시스템과는 질적으로 다른 충전 성능을 보인다. 실제 양자 배터리는 원자 붕괴, 위상 이탈, 광자 손실과 같은 환경 노이즈에 노출된 개방 양자계로 동작하며, 다중 큐비트 시스템에서는 이러한 노이즈가 집단적 자유도 전체에 비선형적으로 작용하여 충전 과정에서 결맞음 붕괴와 에르고트로피 감소를 유발한다. 강화학습은 복잡한 양자 동역학에서 최적의 제어 프로토콜을 자동으로 탐색할 수 있는 접근법이지만, 집단적 다중 큐비트 개방계 환경에서는 학습이 불안정해지고 정책이 국소 최적해에 수렴하거나 실패하는 문제가 발생한다[1].

본 연구에서는 이러한 문제를 해결하기 위해 집단적 다중 큐비트 양자 배터리를 대상으로 한 커리큘럼 기반 PPO (Proximal Policy Optimization) 충전 제어 프레임워크를 제안하며, 노이즈가 없는 환경에서 시작해 시스템 복잡도를 단계적으로 증가시키는 커리큘럼 학습을 통해 개방 양자계 환경에서도 에르고트로피와 충전 효율을 안정적으로 최대화하는 제어 정책 학습을 목표로 한다[2].

II. 본론

2.1 집단적 다중 큐비트 양자 배터리 모델

다중 큐비트가 단일 캐비티 모드와 집단적으로 결합된 TC (Tavis-Cummings) 모델을 양자 배터리의 물리적 모델로 사용한다. 동일한 주파수를 갖는 N 개의 2 준위 큐비트는 전체를 하나의 집단적 자유도로 표현되며, 집단적 대칭성이 유지되는 상태들만을 고려함으로써 상태 공간을 효과적으로 축소할 수 있다. 시스템 해밀토니안은 캐비티 모드, 집단 스핀, 그리고 시간 의존적 외부 구동 필드를 포함하며, Lindblad 마스터 방정식을 통해 집단 상태 전체에 작용하는 환경 노이즈를 고려한다. TC 모델에서 집단적 동역학은 큐비트 수에 따라 비선형적으로 변화하며, 외부 제어 입력은 개별 큐비트가 아닌 집단 상태 전체에 작용함으로써 단일 큐비트 시스템과는 구별되는 동역학적 특성을 보인다. 이러한 특성으로 인해 다중 큐비트 양자 배터리는 강화학습 관점에서 비선형·고난이도의 제어 문제로 해석된다.

2.2 강화학습 환경 및 PPO 설정

충전 제어 문제는 연속 상태와 연속 행동을 갖는 강화학습 문제로 정의된다[3]. 강화학습 에이전트는 시스템의 현재 상태를 관측하고, 연속적인 제어 입력을 출력함으로써 충전 과정을 제어한다. 에이전트의 관측 공간은 배터리 에너지, 에르고트로피, 평균 광자 수, 집단 스핀 성분, 그리고 정규화된 시간 정보를 포함하며, 행동 공간은 캐비티에 인가되는 구동 필드의 실수 및 허수

성분으로 정의된다. 보상 함수는 에르고트로피 증가량으로 설계하여 에이전트는 단순한 에너지 축적이 아닌 실제로 추출 가능한 에너지를 극대화하는 충전 전략을 학습하도록 유도된다. 정책 학습에는 연속 제어 문제에서 안정적인 성능을 보이는 PPO 알고리즘을 사용하며, 주요 하이퍼파라미터 설정은 표 1 과 같으며, 이러한 설정은 비선형적인 집단 동역학과 환경 노이즈가 존재하는 조건에서도 안정적인 학습 수렴을 가능하게 한다.

Parameter	Value
Clipping parameter (ϵ)	0.2
Discount factor (γ)	0.99
GAE parameter (λ)	0.95
Entropy coefficient (c_e)	0.01
Learning rate (η)	Stepwise decay

표 1. PPO 주요 하이퍼파라미터

2.3 커리큘럼 학습 전략

학습의 수렴성과 강건성을 확보하기 위해 커리큘럼 학습 전략을 도입한다. 커리큘럼은 노이즈가 없는 단일 큐비트 환경에서 시작하여, 큐비트 수와 환경 노이즈 강도가 점진적으로 증가하는 6 단계로 구성된다. 각 단계의 세부 설정은 표 2 에 정리되어 있으며, 이를 통해 에이전트는 단순한 동역학에서부터 복잡한 다중 큐비트 개방계 환경까지 안정적으로 적응하는 제어 정책을 학습할 수 있다. 노이즈 강도는 학습 초기에 불안정한 정책 붕괴를 방지하면서도, 후반 단계에서 강건성을 유도할 수 있도록 점진적으로 증가하도록 설정한다.

Stage	Qubits	Noise level	Episodes	Time steps
1	1	0.000	50	25
2	2	0.000	60	35
3	2	0.003	70	40
4	2	0.006	80	45
5	3	0.008	100	50
6	3	0.012	120	55

표 2. 커리큘럼 학습 단계 설정

III. 실험 및 결과

커리큘럼 기반 PPO 의 성능을 검증하기 위해, 서로 다른 큐비트 수와 노이즈 강도를 갖는 커리큘럼 환경에서 학습을 수행한다. 그림 1 은 충전 과정에서의 에르고트로피 시간 변화를 나타내며, 무작위 제어 및 고정 구동 방식과의 비교 결과를 함께 제시한다. 커리큘럼 기반 PPO 는 초기 단계에서 에르고트로피를 빠르게 증가시키고, 노이즈가 존재하는 환경에서도 비교 방식 대비 높은 에르고트로피를 안정적으로 유지함을 확인할 수 있다. 충전 후반부에서 에르고트로피가 일정 수준으로 수렴하는 현상이 관찰되는데, 이는 보상 함수를 에르고트로피 증가량으로 정의함에 따라, 큐비트 수 증가로 동역학이 복잡해지는 영역에서 무리한 제어를 회피하고 안정적인 충전 상태를 유지하는 정책이 학습되기 때문이다. 이러한 결과는 커리큘럼 기반 PPO 가 다중 큐비트 양자계 환경에서도 물리적으로 의미 있는 충전 제어를 효과적으로 수행할 수 있음을 보여준다.

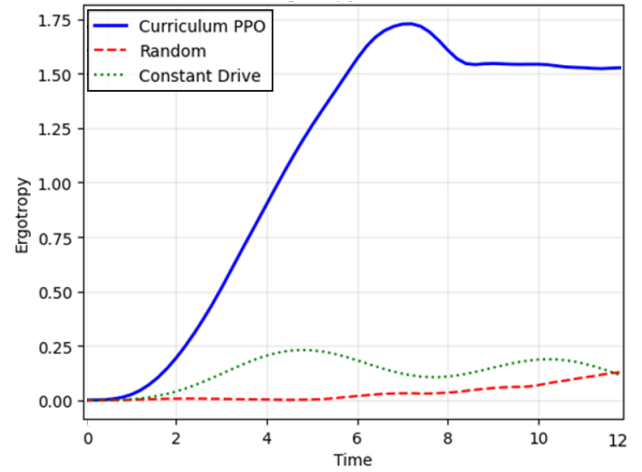


그림 1. 강화학습 기반 제어와 기존 제어 방식의 에르고트로피 변화

IV. 결론

본 연구에서는 집단적 다중 큐비트 양자 배터리를 대상으로, 개방 양자계 환경에서도 강건하게 동작하는 커리큘럼 기반 PPO 충전 제어 프레임워크를 제안하고 그 유효성을 수치 시뮬레이션을 통해 검증하였다. 제안된 방법은 노이즈와 시스템 차원이 증가하는 환경에서도 에르고트로피를 안정적으로 유지하며, 비커리큘럼 강화학습 및 정적 제어 방식 대비 우수한 성능을 보인다. 이러한 결과는 커리큘럼 학습이 집단적 양자 시스템 제어에서 중요한 역할을 할 수 있음을 시사한다. 향후 연구에서는 본 프레임워크를 더 큰 규모의 다중 큐비트 시스템으로 확장하여 초고속충과 같은 집단적 양자 효과와의 연계를 분석하고, 실제 양자 하드웨어 적용 가능성까지 검토할 예정이다.

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] 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
- [2] Ma, H., Dong, D., Ding, S. X., & Chen, C. (2021). Curriculum-based deep reinforcement learning for quantum control. *IEEE Transactions on Neural Networks and Learning Systems*, 34(11), 9000– 9013.
- [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.