

AI / Data Engineering Track

Quantum Reinforcement Learning for Joint Control, Communication, and Computing in Stabilized Reusable Space Rocket

Gyu Seon Kim, Jaehyun Chung, Soyi Jung, Soohyun Park, Joongheon Kim

IEEE Transactions on Aerospace and Electronics Systems (2026)

Contents

01 Introduction

02 Related Work

03 Quantum Neural Network

04 Joint Control in Stabilized Reusable Rocket

05 Communication in Stabilized Reusable Rocket

06 Computing in Stabilized Reusable Rocket

07 Performance Evaluation

08 Conclusion

01 Introduction

SpaceX의 Falcon 9와 같은 **재사용 로켓**은 우주 탐사 비용을 매우 크게 절약하기에
우주 산업의 판도를 바꾸는 핵심 기술로 부상함



01 Introduction

재사용 로켓의 임무 성공을 위해서는 다음 세 가지 요소가 필수:

견고한 제어 시스템

착륙 단계에서 정확한 속도 및 자세 제어가 필수적임

원활한 통신 프레임워크

로켓의 실시간 시스템 상태를 지상국과 모니터링해 임무 성능을 최적화함

에너지 효율적인 컴퓨팅 알고리즘

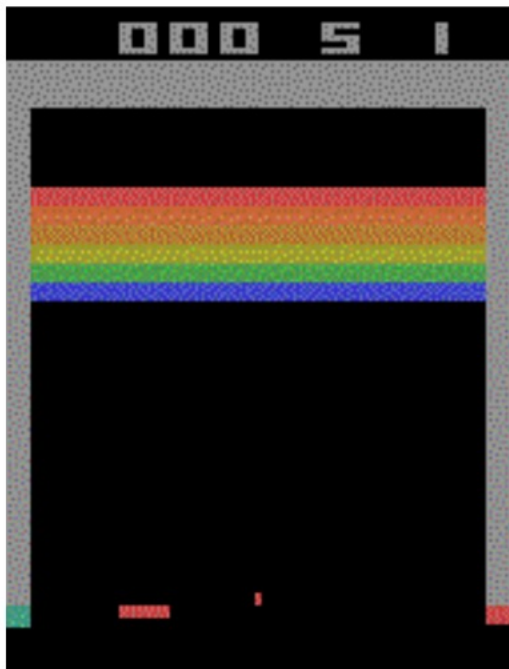
한정된 자원 내에서 필요한 계산을 효율적으로 하고 정보를 제때 처리해야 함

세 가지 요소가 상호 의존적

01 Introduction

동적이고 불확실한 환경? → **강화학습**을 사용하자!

하지만, 로켓은 너무나도 복잡하고 예측 불가능한 환경 → 학습 과정에서 오히려 성능이 나빠지는 경우가 발생할 수 있음



핵심 변수는 적음
물리가 단순하고 노이즈가 거의 없음
행동도 몇 개 안 됨 (좌/우/정지)



실제로는 저차원 문제에 가까움



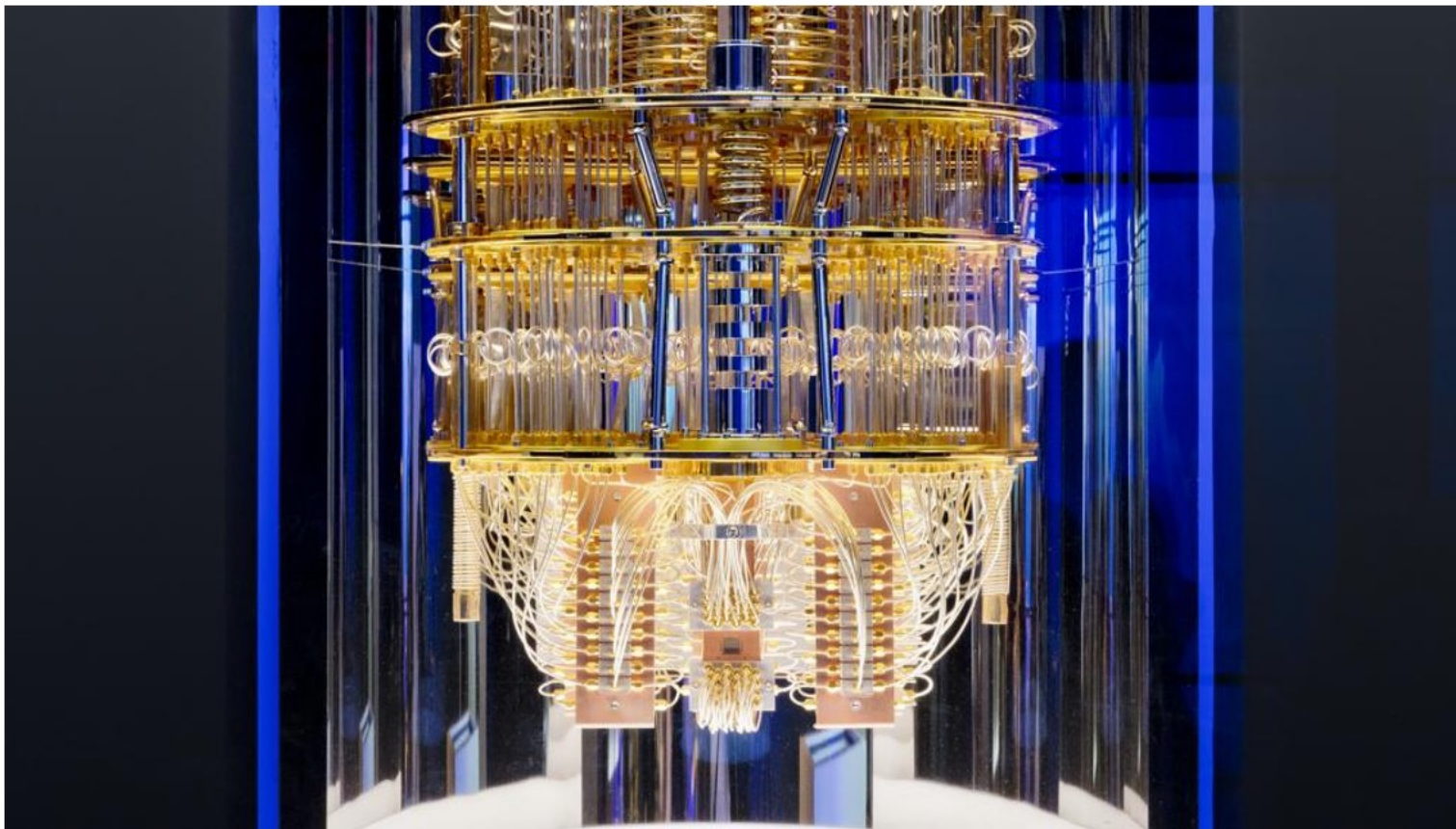
위치, 속도, 자세, 각속도,
연료량, 엔진 상태, 추력 벡터, 바람,
진동, 센서 오차, 지면 상태 등



단순화해도 수십 차원

01 Introduction

일반 컴퓨터로 풀기 어려운 문제 → 양자컴퓨터를 사용한다면?

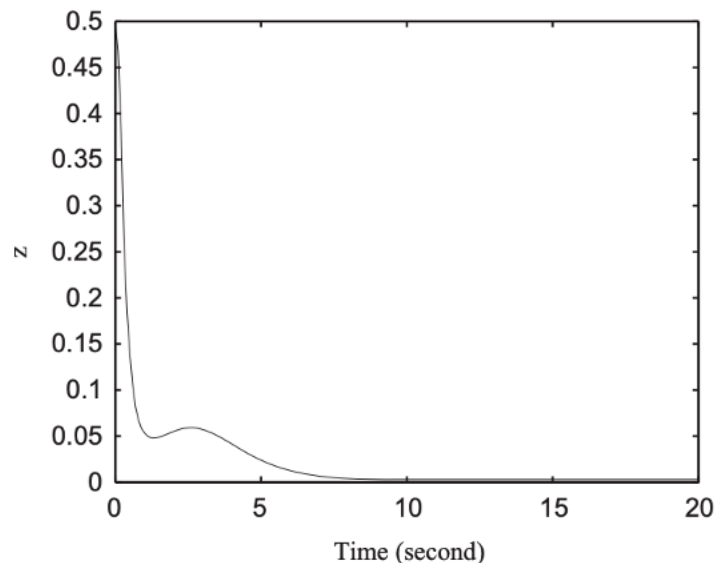


02 Related Work

Robust Control (강인 제어)

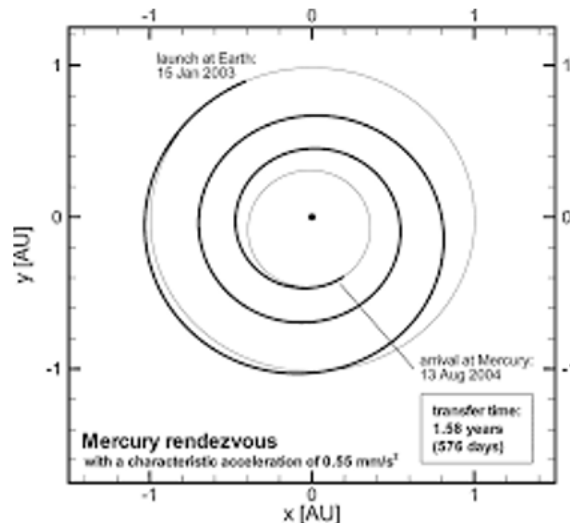
신경망을 활용하여 로켓이나 우주선을 착륙시키는 연구는 과거부터 꾸준히 지속되었음

그러나 실제 착륙 과정은 **매우 복잡**하고, 고려해야 할 **변수가 너무 많음**



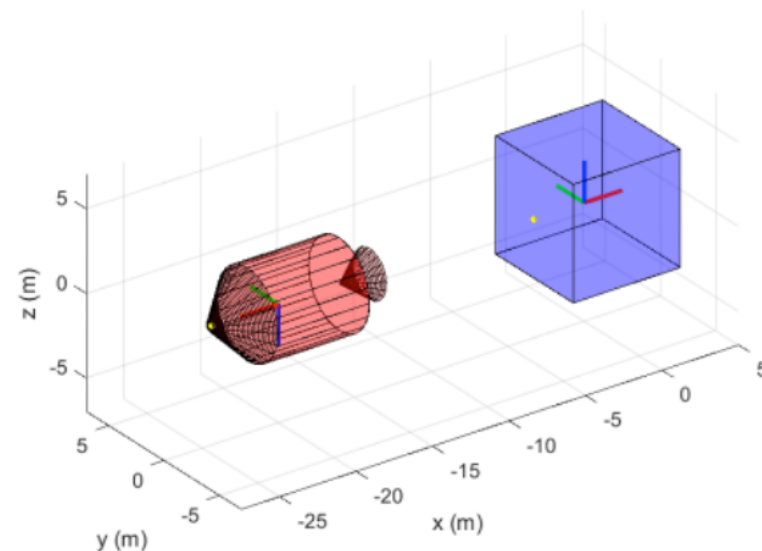
신경 제어를 사용해 우주선의 각운동 안정화한 연구

(Learning from adaptive neural network control of an underactuated rigid spacecraft)



신경 제어 알고리즘을 이용한 우주선 행성 간 경로 설계 최적화 연구

(Optimization of Interplanetary Solar Sailcraft Trajectories Using Evolutionary Neurocontrol)



강화학습을 도킹 절차를 위한 피드백 제어 규칙 탐색에 사용한 연구

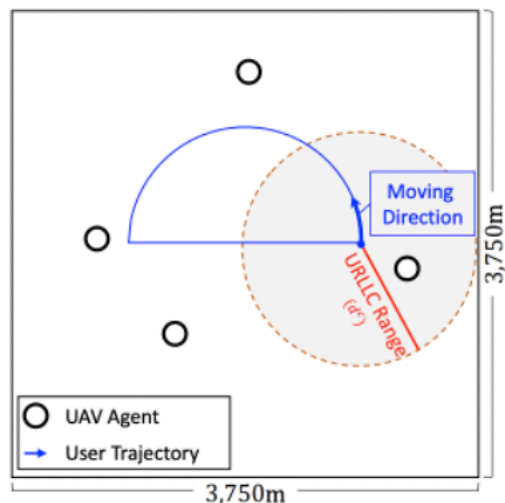
(Autonomous six-degree-of-freedom spacecraft docking maneuvers via reinforcement learning)

02 Related Work

Seamless Communication (원활한 통신)

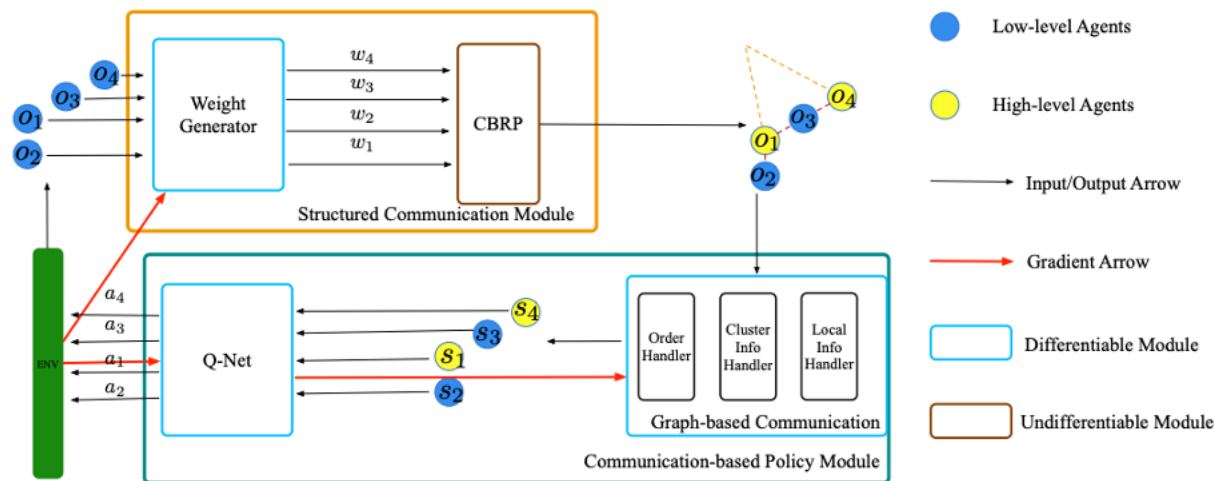
기존에는 다중 에이전트 강화학습 기술을 통한 로켓과 기지국의 통신 방법이 연구됨

그러나 이 방법 또한 로켓이나 기지국이 **많아질수록 최적의 통신 방법을 찾기 어렵다** 는 문제가 존재함



Graph Attention Exchange Network (GAXNet)

(Attention-based Reinforcement Learning for Real-Time UAV Semantic Communication)



Learning Structured Communication (LSC) Framework

(Learning Structured Communication for Multi-agent Reinforcement Learning)

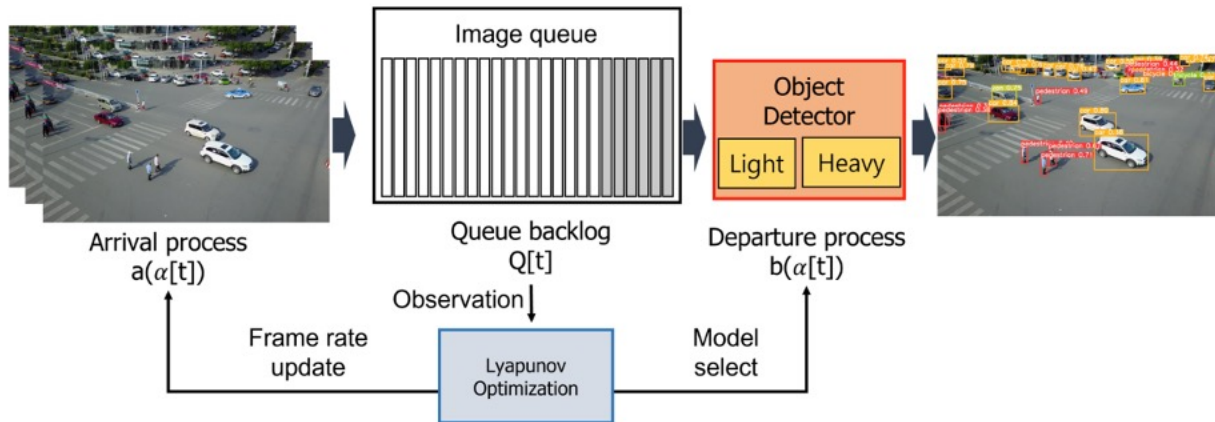
02 Related Work

Image Processing and Low Power Computing (이미지 처리 및 저전력 컴퓨팅)

안전한 착륙을 위해서는 주변 환경을 촬영/분석해 착륙 지점을 정확하게 찾아야 하며,

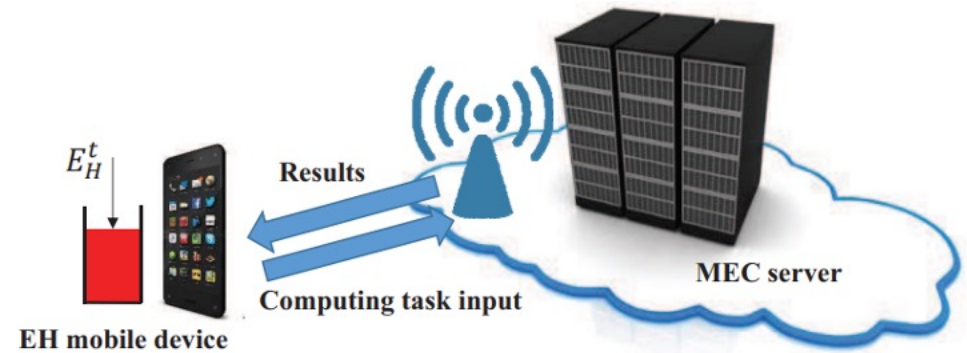
이 과정에서 사진을 처리하고 데이터를 전송할 때의 전력 소모를 최소화하는 것이 중요함

실시간으로 복잡한 환경에서 수많은 이미지를 처리하고 전력을 관리하는 데에는 어려움이 존재함



제한된 연산능력과 배터리 자원 내에서 객체 인식 성능을 최대화하기 위해 자원을 최적으로 관리하는 연구

(Joint frame rate adaptation and object recognition model selection for stabilized unmanned aerial vehicle surveillance)



컴퓨터가 할 작업을 여러 기기에 분산시켜 전력 손실을 줄이는 연구

(Dynamic Computation Offloading for Mobile-Edge Computing with Energy Harvesting Devices)

02 Related Work

Robust Control

너무 복잡하고 고려해야 할 변수가
많은 실제 착륙 과정

Seamless Communication

기지국 수에 따라 기하급수적으로
증가하는 통신 방식 탐색 난이도

Image Processing & Low Power Computing

한정된 자원 내에서 최상의 결과를
도출하는 자원 관리 능력 필요

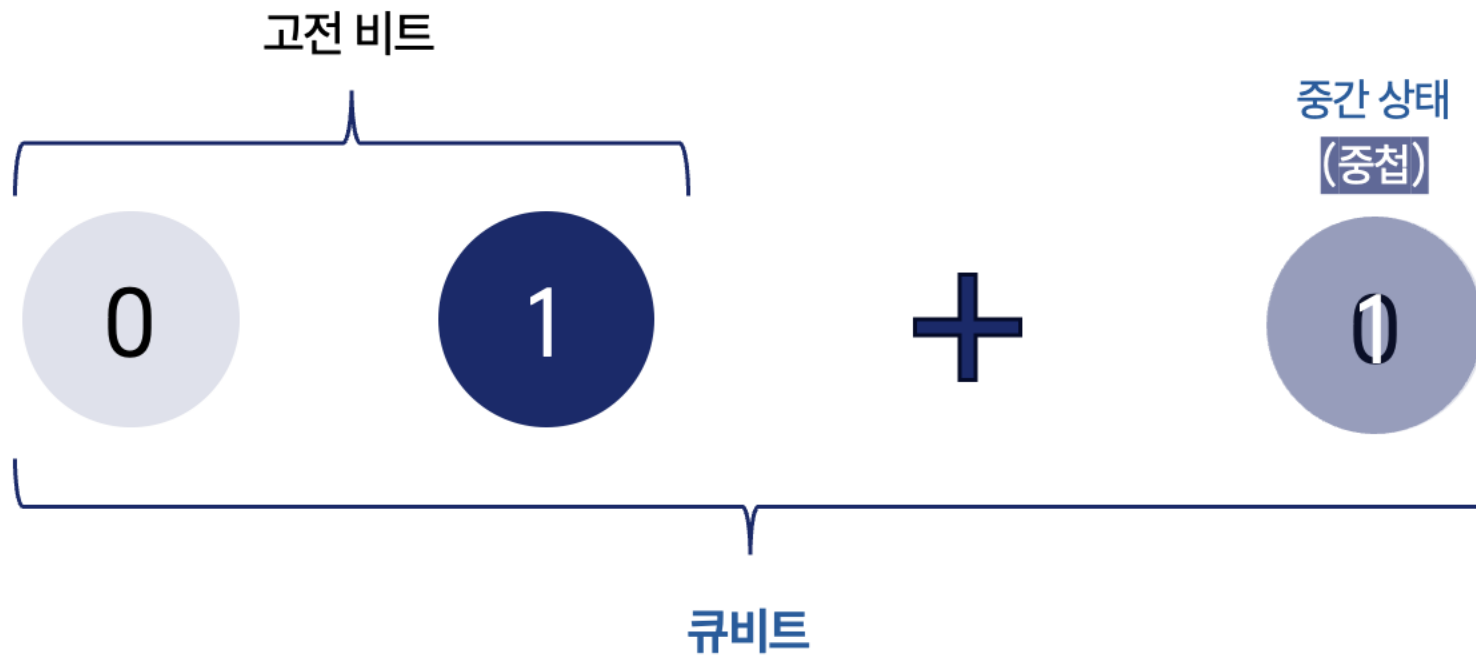


양자 강화학습 (QRL)을 사용하자!

03 Quantum Neural Network

Qubit (큐비트)

양자컴퓨터의 정보를 처리하는 기본 단위로, 0과 1 외에도 0과 1의 중간 상태를 가질 수 있음



03 Quantum Neural Network

큐비트 기호는 다음과 같으며, 행렬로도 표현할 수 있음

0

$$|0\rangle := \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

1

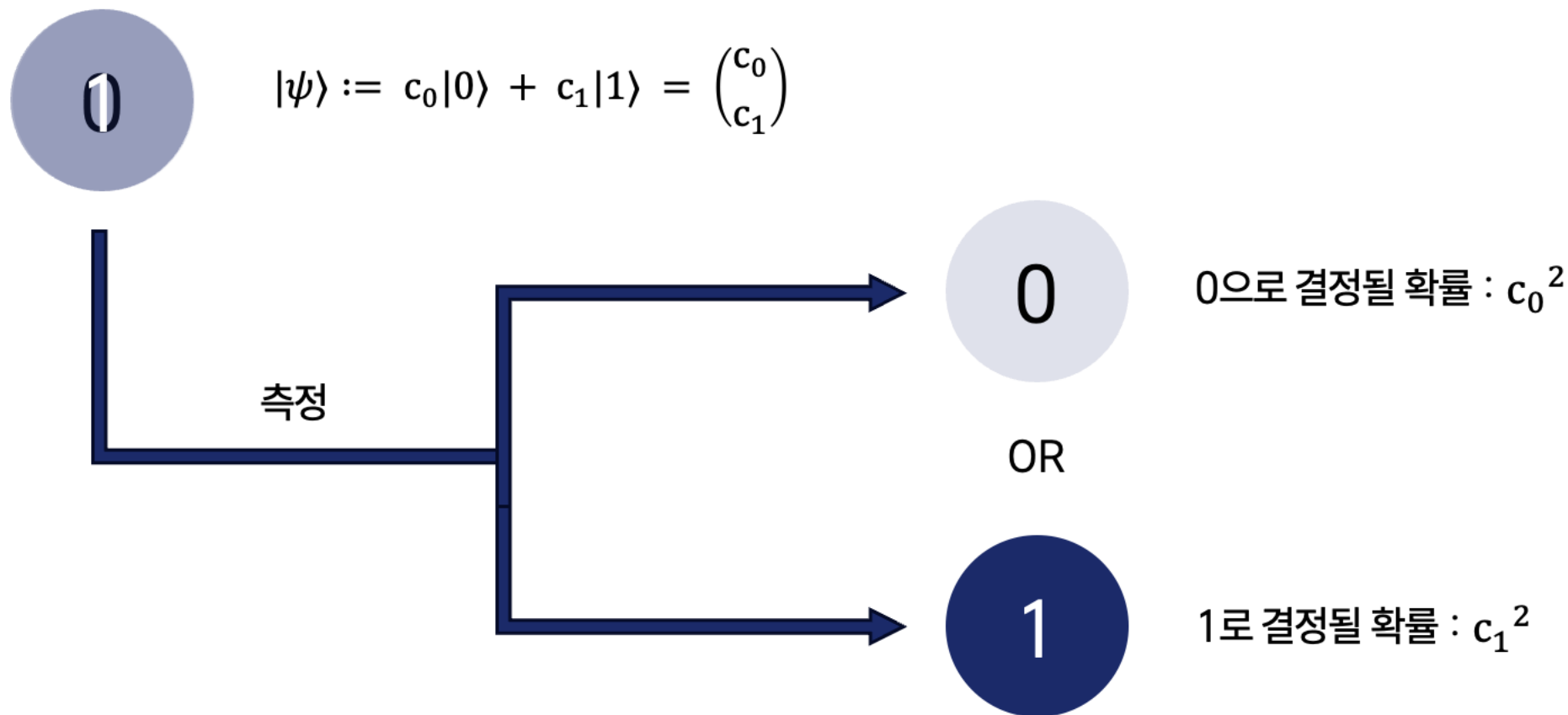
$$|1\rangle := \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

1

$$|\psi\rangle := c_0|0\rangle + c_1|1\rangle = \begin{pmatrix} c_0 \\ c_1 \end{pmatrix}$$

03 Quantum Neural Network

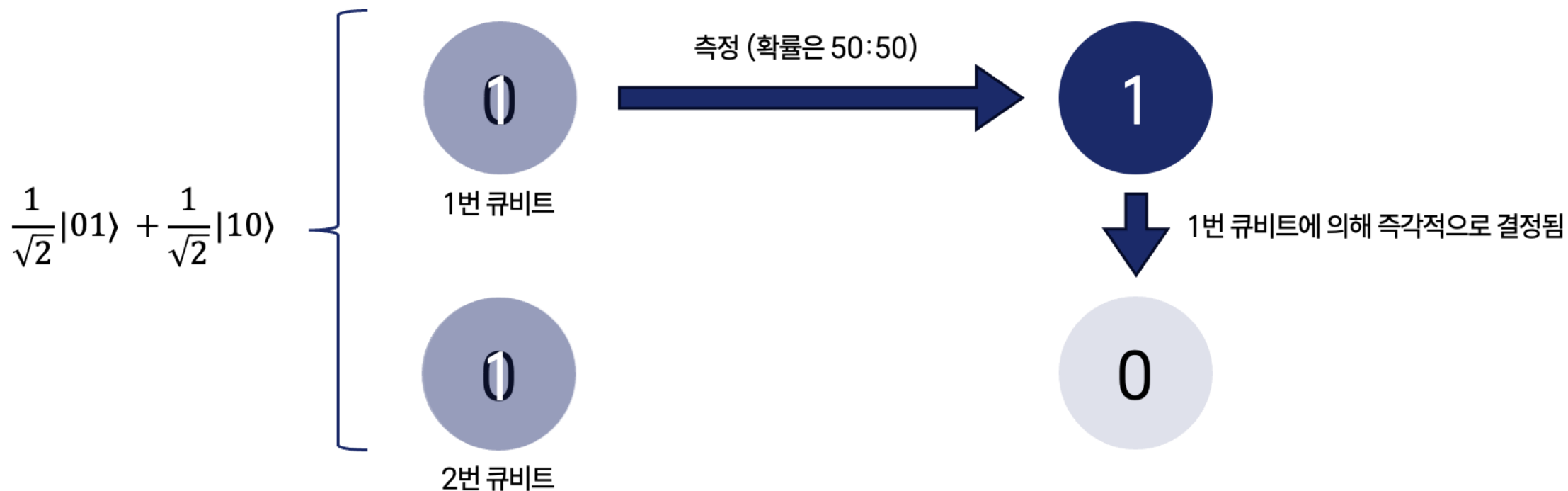
중첩 상태인 큐비트는 최종적으로 **측정** 시 정해진 확률에 따라 0과 1 중 하나의 상태로 결정됨



03 Quantum Neural Network

중첩 외에도, 양자는 얽힘이라는 특성을 지니고 있음

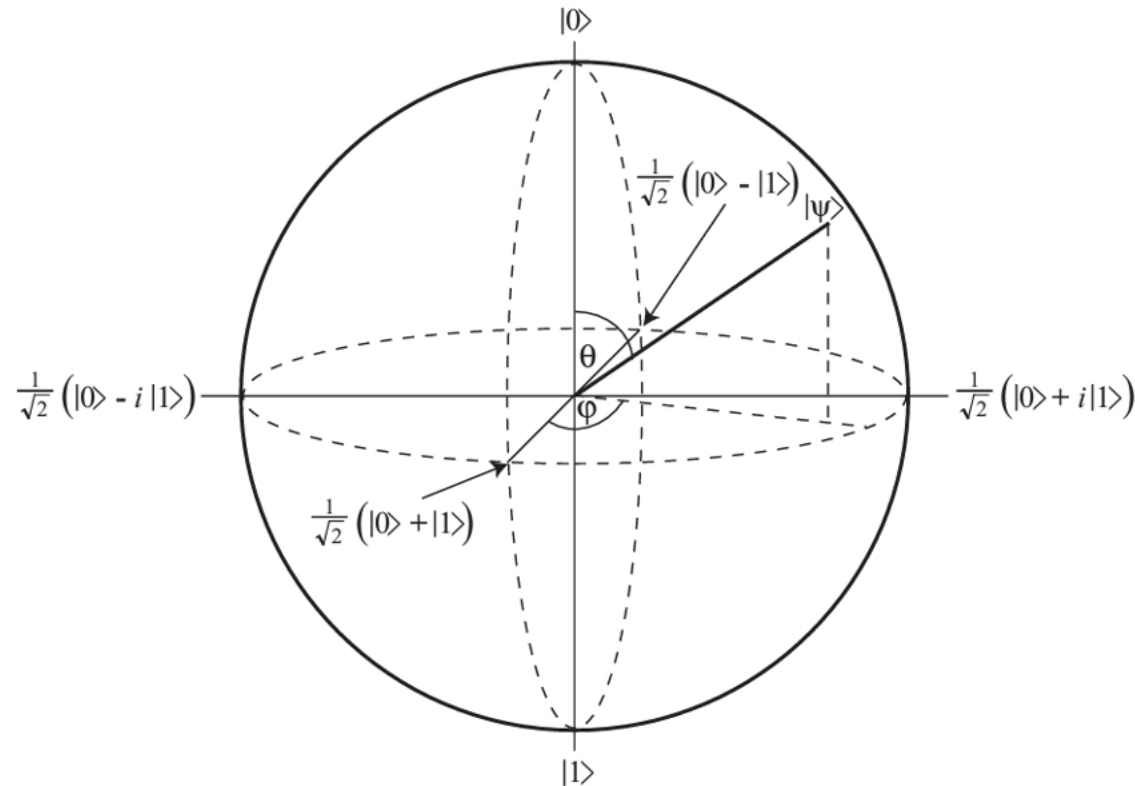
이는 하나의 입자의 상태가 결정되면 다른 입자의 상태도 즉각적으로 결정되는 특성임



03 Quantum Neural Network

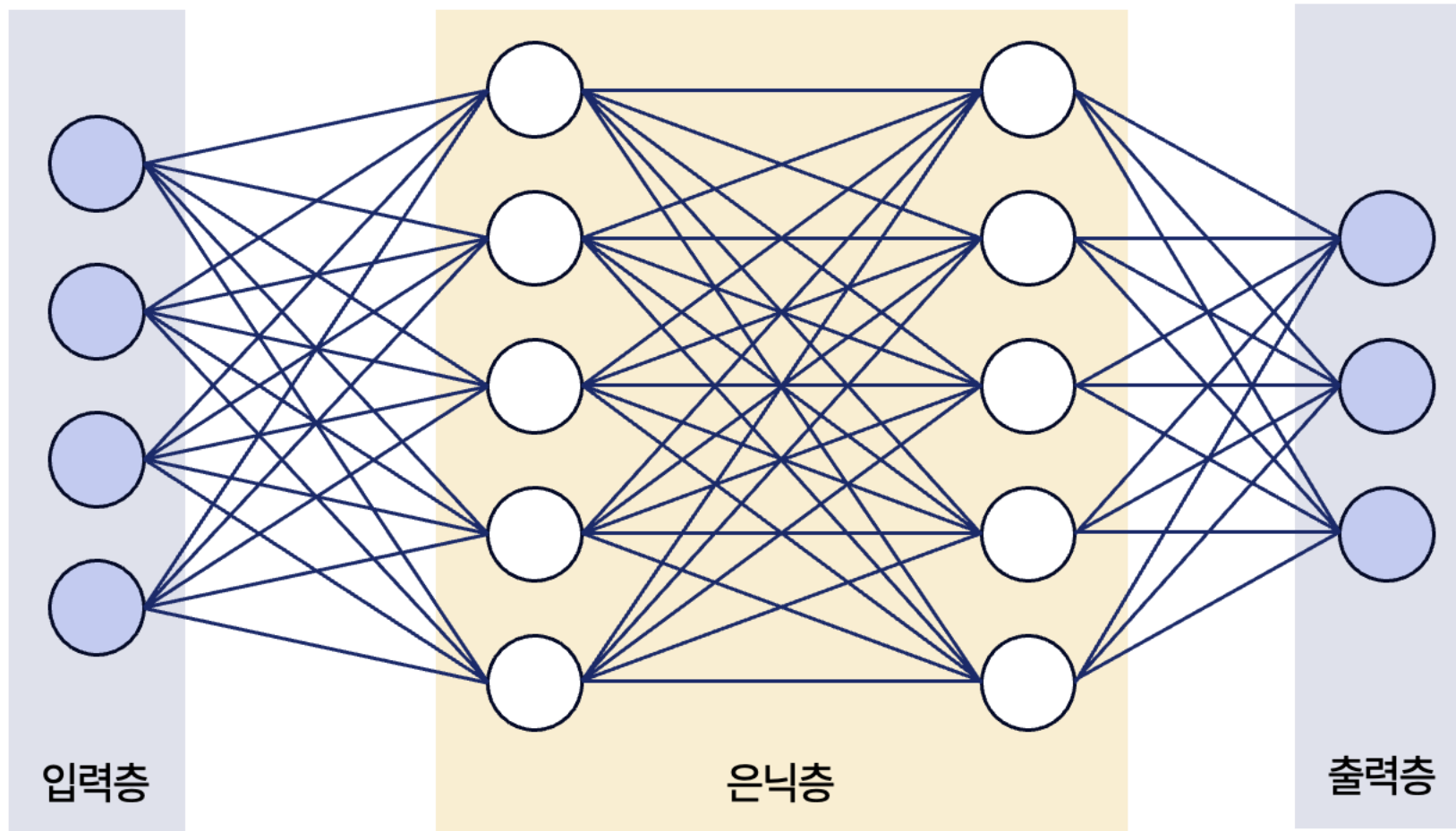
큐비트의 중첩 상태를 기하학적으로 표현하기 위해, 블로흐 구면을 사용함

블로흐 구면은 복잡한 큐비트의 중첩 상태를 직관적으로 이해할 수 있게 도와줌



03 Quantum Neural Network

고전 신경망(NN)의 구조



03 Quantum Neural Network

고전 신경망의 입력층: 숫자 벡터를 그대로 입력

양자 신경망은:

숫자를 그대로 받을 수 없음 → **값들을 회전각처럼 사용**해 큐비트를 특정 양자 상태로 변환



03 Quantum Neural Network

고전 신경망의 은닉층: 입력 특징 조합, 중요 패턴 추출, 비선형 관계 학습

양자 신경망은:

회전과 얽힘을 통해 중요한 특징을 추출

Ex.

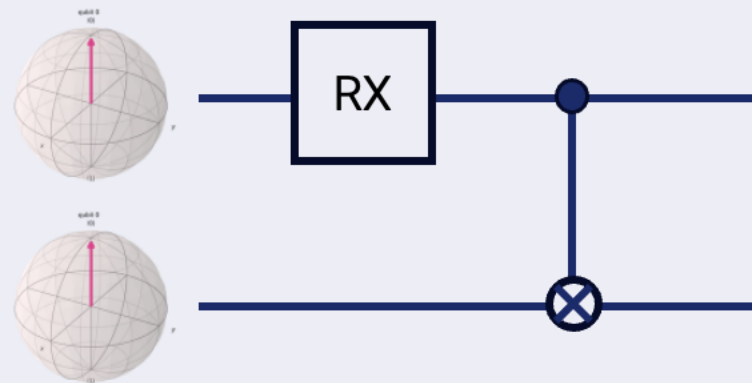
$$z = 0.5x_1 + 0.3x_2 + 0.7x_3$$

계산 후 활성화 함수 적용 / 반복으로 특징 추출

고전 신경망

Ex.

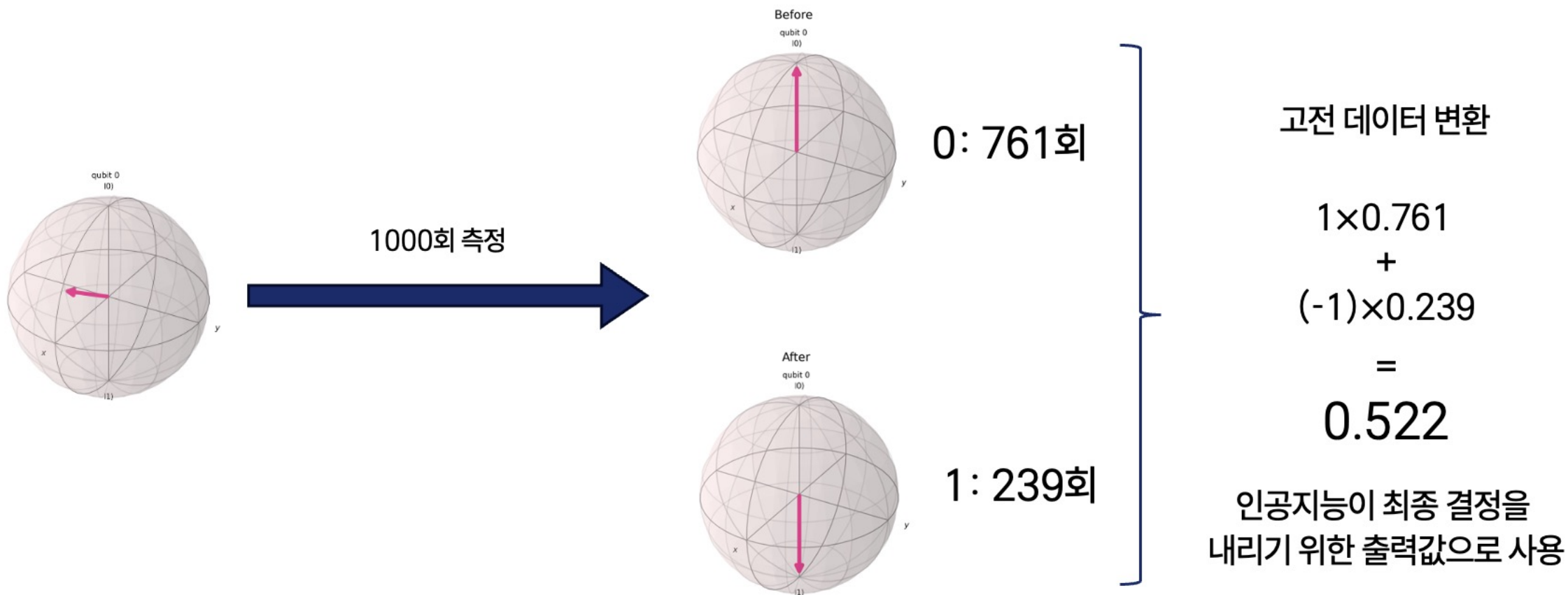
회전 및 얽힘 등을 통해 특징 추출



양자 신경망

03 Quantum Neural Network

양자회로를 거쳐 계산된 양자 상태는 수 차례의 측정을 통해 다시 고전 데이터로 변환됨



04 Joint Control in Stabilized Reusable Rocket

로켓의 착륙 과정을 컴퓨터가 이해할 수 있는 문제로 정의: Markov Decision Process (MDP)

$$\text{MDP} = (S(t), A(t), R(\cdot), P(s'|s, a), \gamma)$$

04 Joint Control in Stabilized Reusable Rocket

$$\text{MDP} = (S(t), A(t), R(\cdot), P(s'|s, a), \gamma)$$

상태 (State)

$$S_L(t) = \underbrace{\bigcup_{t=0}^{T-1} \{\Delta x(t)\}, \bigcup_{t=0}^{T-1} \{\Delta y(t)\}, \bigcup_{t=0}^{T-1} \{\Delta z(t)\}}_{\text{로켓과 착륙 지점 간의 위치 차이}}, \underbrace{\bigcup_{t=0}^{T-1} \{v_x(t)\}, \bigcup_{t=0}^{T-1} \{v_y(t)\}, \bigcup_{t=0}^{T-1} \{v_z(t)\}}_{\text{로켓의 속도}}, \underbrace{\bigcup_{t=0}^{T-1} \{B_x(t)\}, \bigcup_{t=0}^{T-1} \{B_y(t)\}, \bigcup_{t=0}^{T-1} \{B_z(t)\}}_{\text{로켓의 각도}},$$

04 Joint Control in Stabilized Reusable Rocket

$$\text{MDP} = (S(t), A(t), R(\cdot), P(s'|s, a), \gamma)$$

행동 (Action)

$$A_L(t) = [\Pi(t), \Lambda(t)]$$



엔진 추력 벡터의 기울기

엔진 추력 벡터의 방향

04 Joint Control in Stabilized Reusable Rocket

$$\text{MDP} = (S(t), A(t), R(\cdot), P(s'|s, a), \gamma)$$

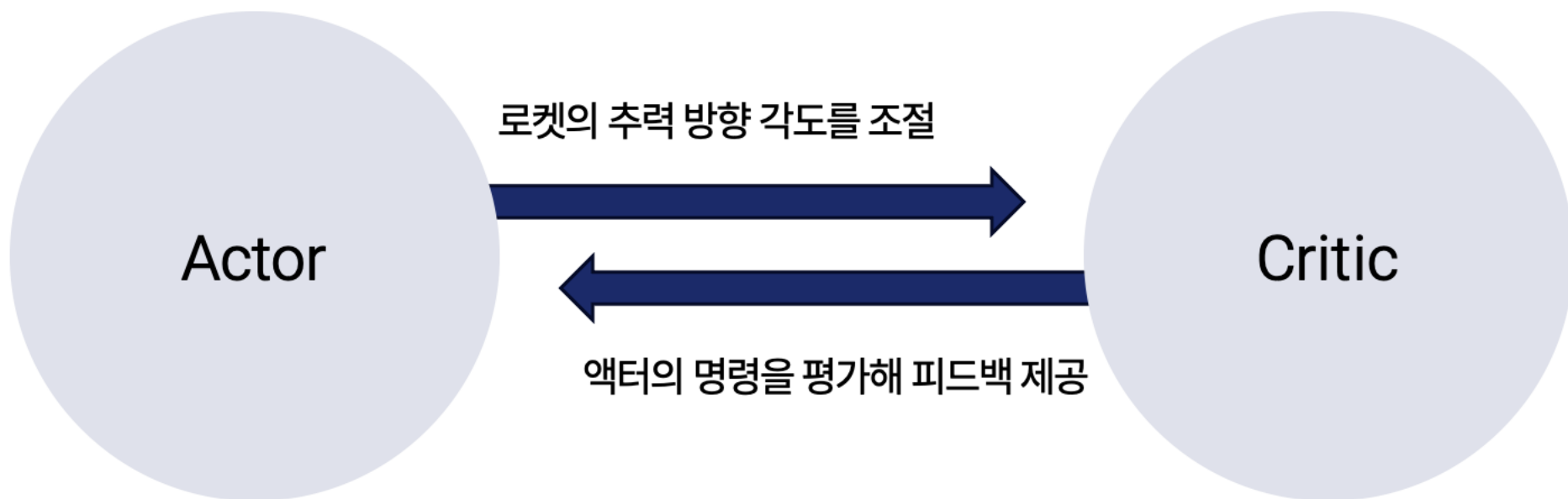
보상 (Reward)

$$R_L(t) = \frac{1}{\underbrace{\sqrt{\Delta x(t)^2 + \Delta y(t)^2 + \Delta z(t)^2}}_{\text{정확한 위치}} + \underbrace{\sqrt{v_x(t)^2 + v_y(t)^2 + v_z(t)^2}}_{\text{느린 속도}} + \underbrace{\sqrt{B_x(t)^2 + B_y(t)^2 + B_z(t)^2}}_{\text{안정적인 각도}}}$$

로켓과 착륙 지점 간의 위치 차이, 속도, 각도가 작아질수록 높은 보상 지급

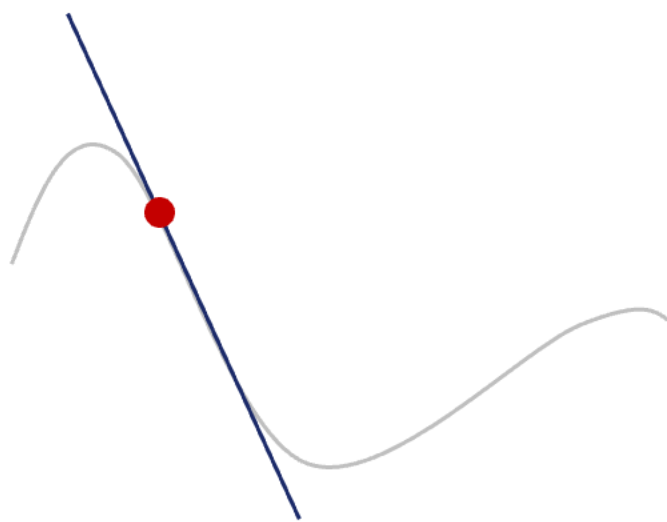
04 Joint Control in Stabilized Reusable Rocket

학습 방식은 액터-크리틱(Actor-Critic) 방식을 사용함



04 Joint Control in Stabilized Reusable Rocket

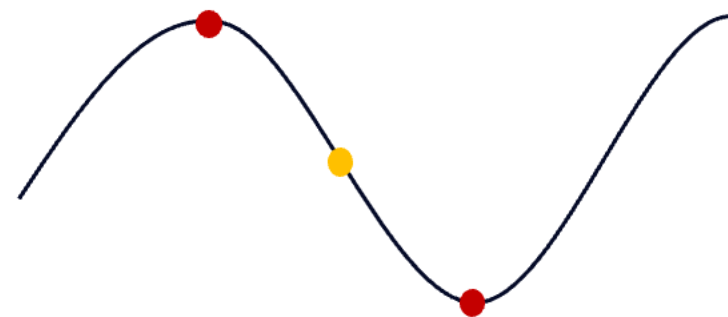
매개변수 시프트 규칙 (Parameter Shift Rule)



고전 신경망:

함수를 미분해 기울기를 계산 (역전파)

양자컴퓨터는 역전파가 불가



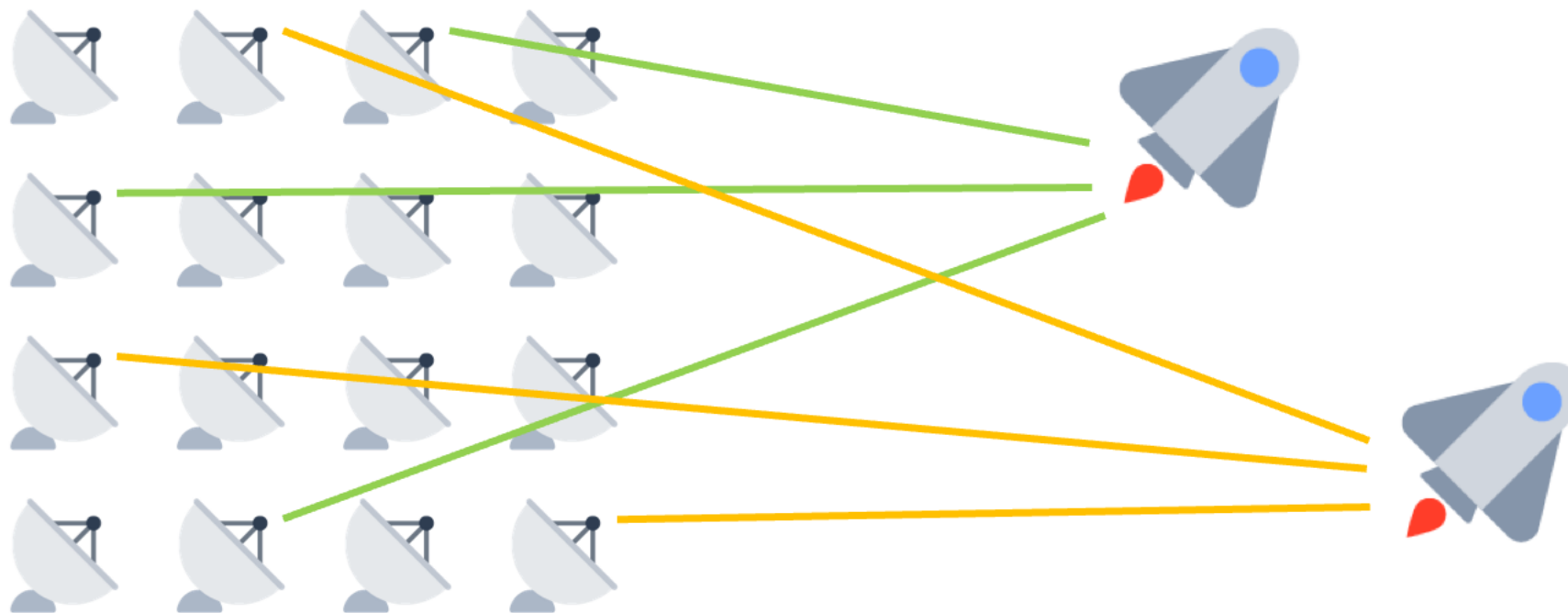
양자 신경망:

파라미터를 $\pm\pi/2$ 만큼 이동시킨 두 회로의 출력 차이로 기울기 계산
(양자 회전게이트의 출력이 삼각함수로 표현되기 때문에 가능)

05 Communication in Stabilized Reusable Rocket

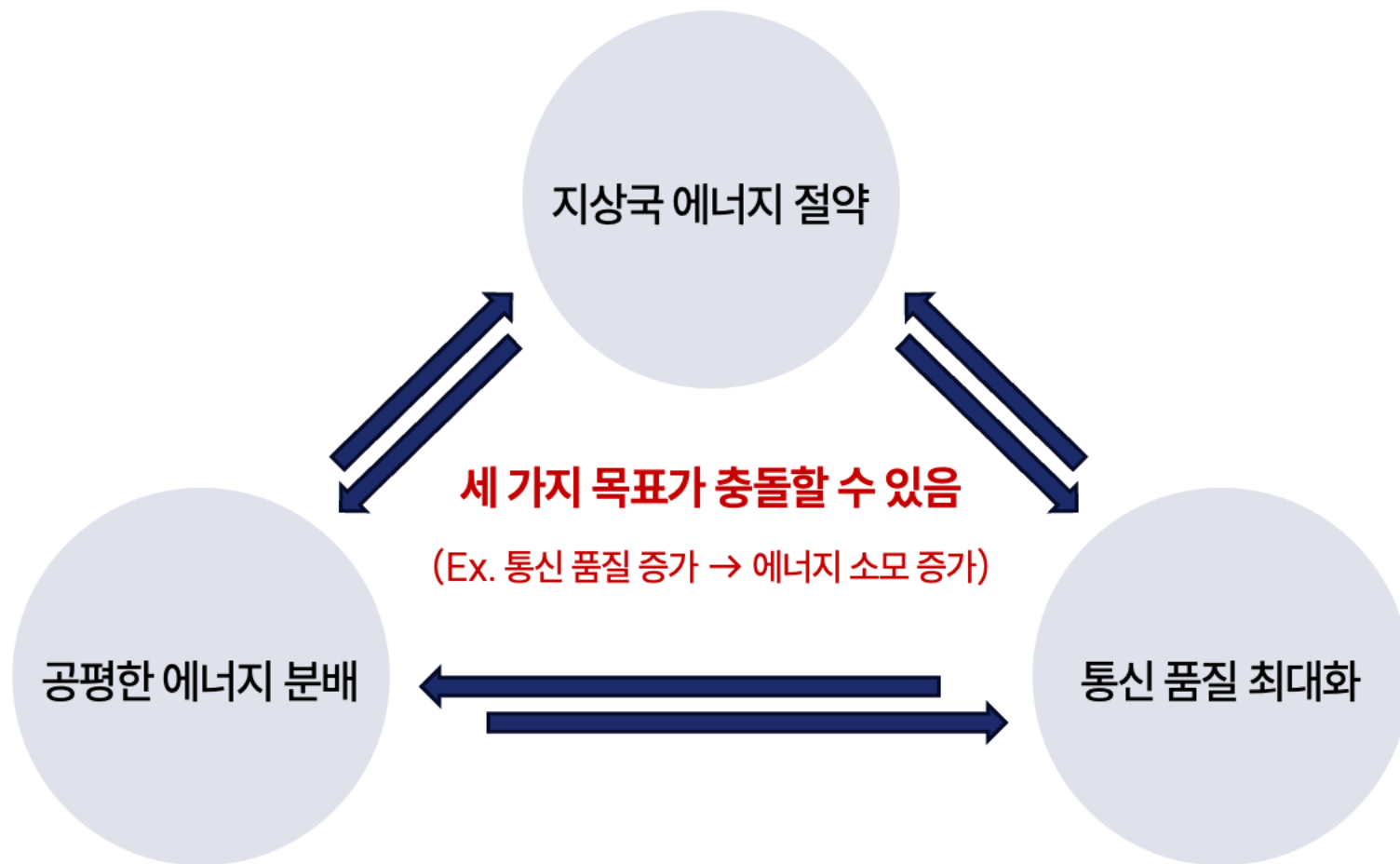
로켓은 재진입/착륙 과정에서 여러 지상국과 효율적으로 통신해야 함

스케줄링: 어떤 지상국에 어떤 로켓을 할당해야 할지 결정



05 Communication in Stabilized Reusable Rocket

스케줄링 목표



05 Communication in Stabilized Reusable Rocket

보상 함수의 구성

보상 함수 = 통신 성능 - 에너지 비용



통신 성능: 통신이 얼마나 잘 되는가?

채널 용량 : 한번에 전송 가능한 데이터의 양

CNR(Carrier-to-Noise Ratio): 신호의 품질 지표



에너지 비용: 지상국이 통신에 에너지를 얼마나 썼는가?

잔여 에너지: 최대 에너지 - 통신에 사용한 에너지 + 협력 항

협력 항: 모든 지상국이 에너지를 골고루 쓰도록 유도

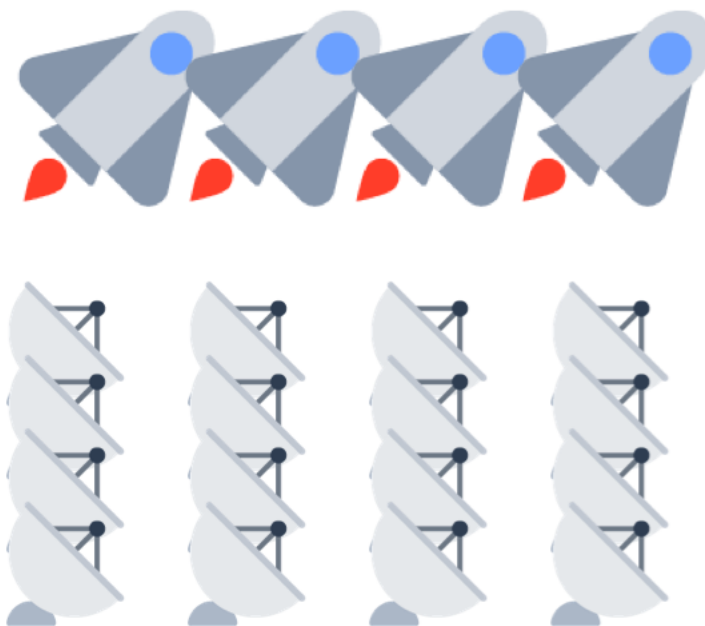
05 Communication in Stabilized Reusable Rocket

지상국과 로켓의 수가 많아지게 되면 경우의 수가 너무 많아져 기존 AI로 해결하는 것이 어려워지게 됨



경우의 수가 적음, 단순함

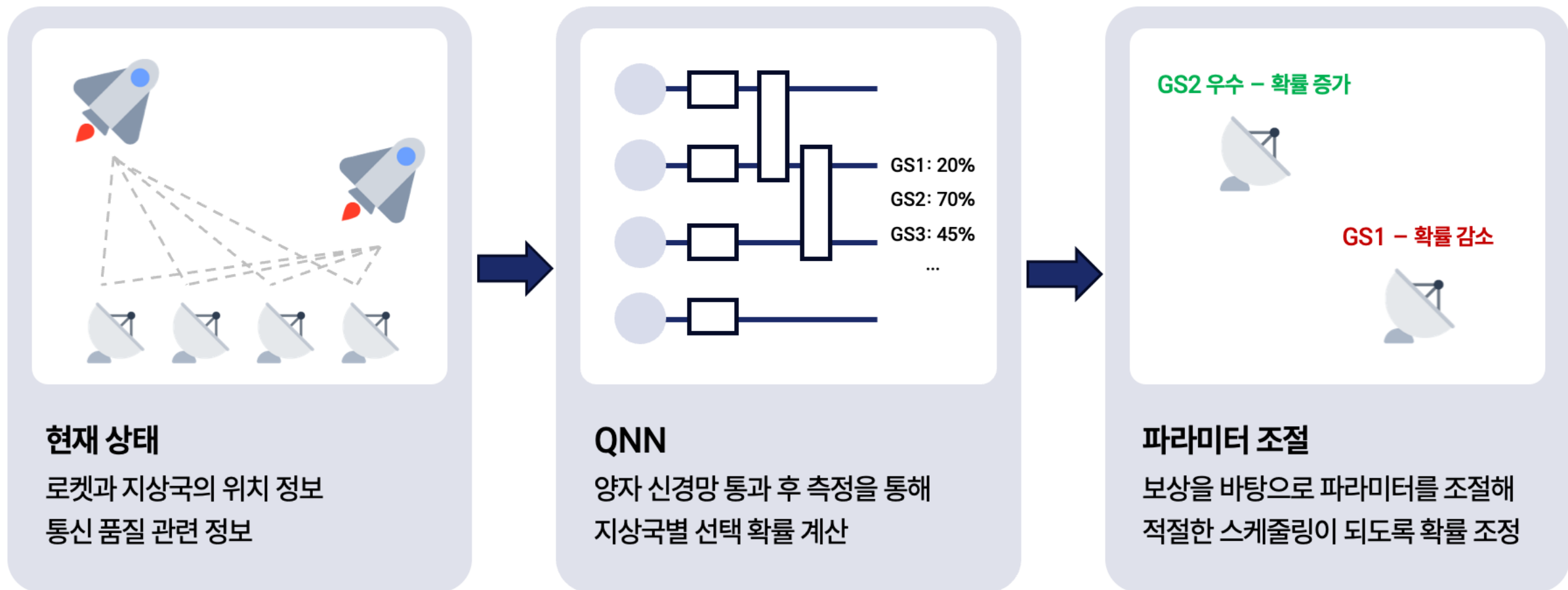
그러나 로켓과 지상국이 많아진다면



경우의 수가 너무 많음, 차원의 저주 발생

양자 멀티에이전트 강화학습을 사용

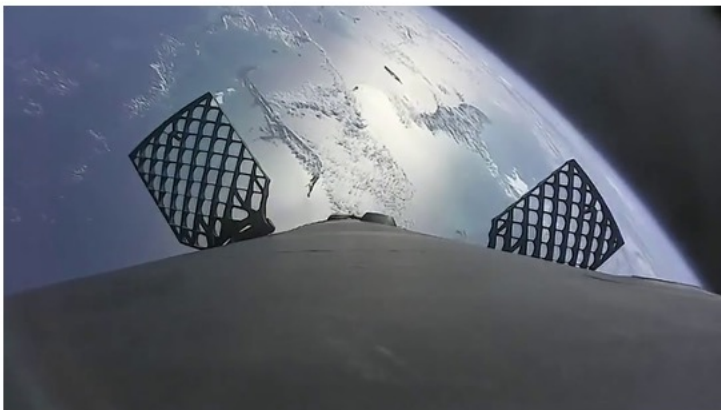
05 Communication in Stabilized Reusable Rocket



06 Computing in Stabilized Reusable Rocket

로켓 착륙 시에 이미지를 촬영해 전송해야 함

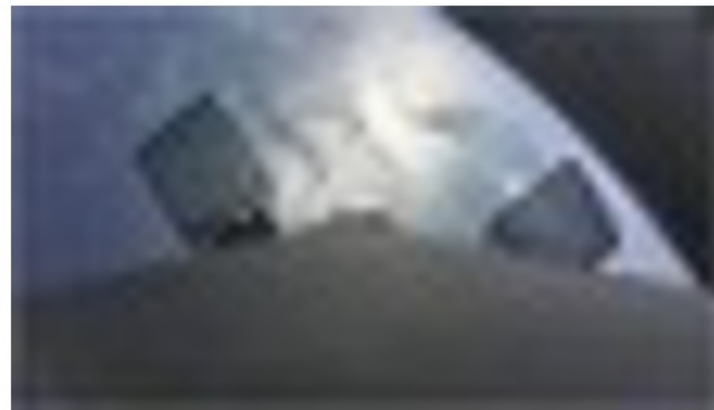
고화질과 저화질의 장단점이 명확하기 때문에, 최적의 균형을 찾는 방법에 대해 연구가 필요함



고화질 이미지

사고 원인 분석, 제어 정확도 향상

큐 오버플로 발생 가능성



저화질 이미지

데이터 전송량 감소, 지연 감소

제어 정확도 감소

목표: 고화질 사진을 전송하면서 큐 오버플로가 발생하지 않도록 제어

06 Computing in Stabilized Reusable Rocket

큐 백로그가 거의 비어 있다면



Queue



고해상도 이미지를 전송해 정확도 향상

큐 백로그가 가득 차 데이터가 밀린다면



Queue

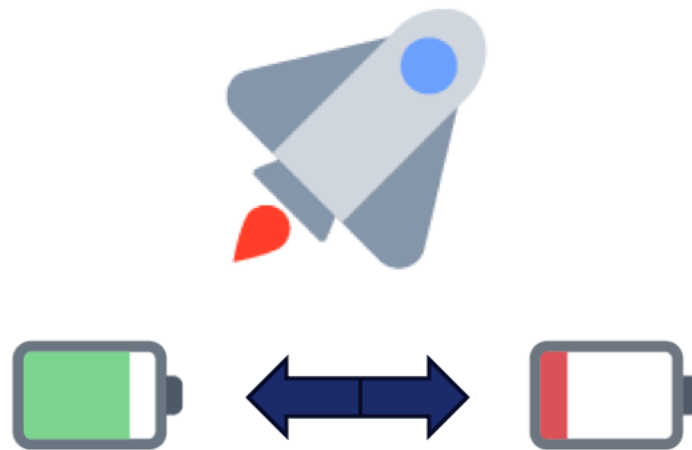


저해상도 이미지를 전송해 데이터 처리 속도 향상

이 과정을 수학 공식으로 만들어서 AI가 최적의 결정을 내리도록 학습함

06 Computing in Stabilized Reusable Rocket

로켓이 자체적으로 가지는 에너지 자원은 매우 제한적
따라서 통신 과정에서 효율적으로 에너지를 사용해야 함

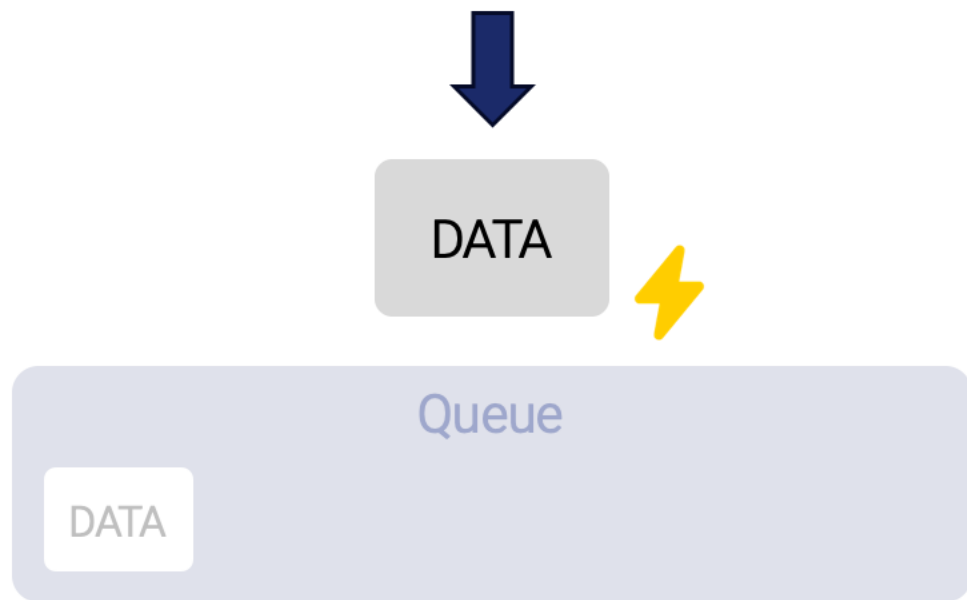


항상 최대 전력을 소비할 필요가 없음

06 Computing in Stabilized Reusable Rocket

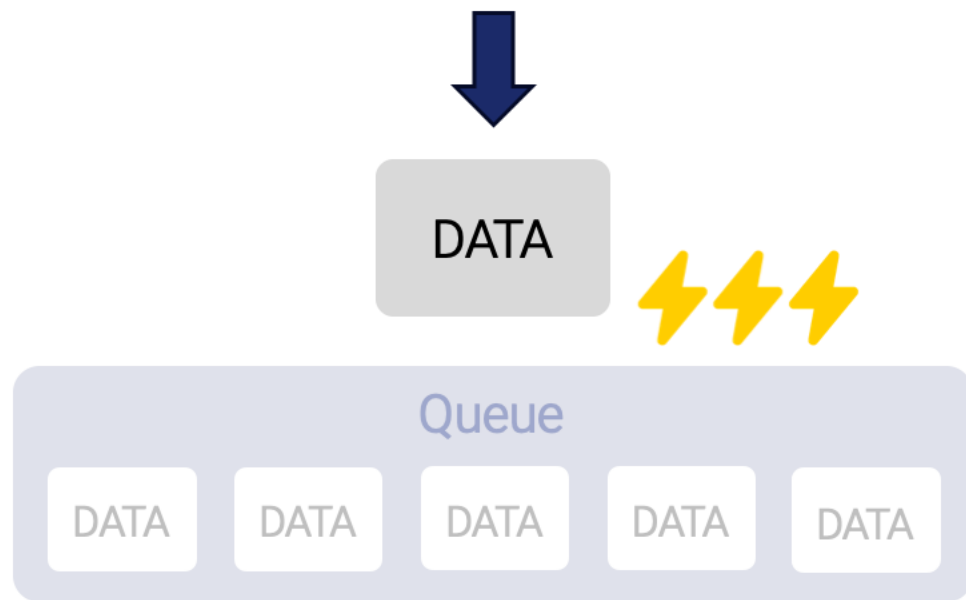
이 또한 이미지 처리와 마찬가지로 큐 백로그를 확인해 전력 소모를 조절

큐 백로그가 거의 비어 있다면



전력을 낮춰 에너지 절약

큐 백로그가 가득 차 데이터가 밀린다면



전력을 높여 데이터 처리 속도 향상

07 Performance Evaluation



로켓 제어 시나리오

비교 대상: 기존 강화학습 알고리즘
(DDPG, Actor-Critic, DQN)



통신 시나리오

비교 대상: 기존 다중 에이전트 강화학습 알고리즘
(Cooperative MARL, Independent Q-Learning,
Random Scheduler)



컴퓨팅 시나리오

비교 대상: 극단적인 케이스
(이미지 - 항상 고화질 / 항상 저화질
저전력 컴퓨팅 - 항상 최대 전력 / 항상 최소 전력)

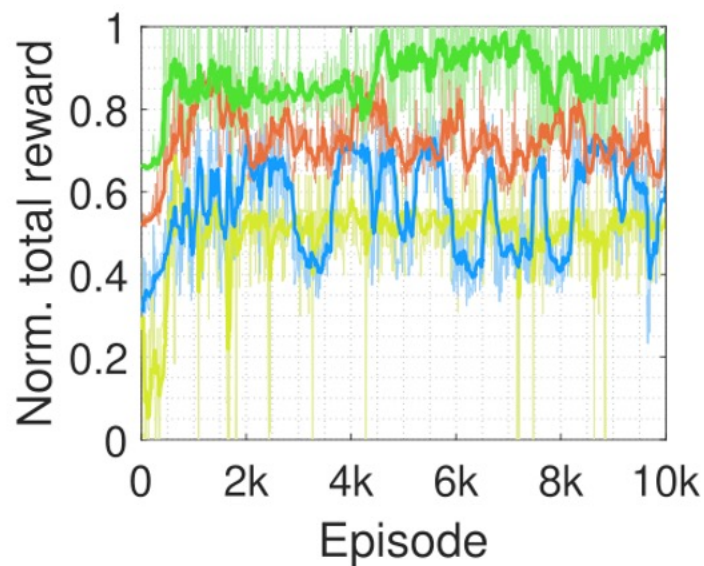


실험은 TorchQuantum (양자컴퓨터 시뮬레이션 프로그램)으로 진행됨
SpaceX – Falcon 9의 파라미터를 반영

07 Performance Evaluation

Robust Control 평가 기준 (1)

누적 보상의 수렴: AI가 학습하면서 얻는 점수가 꾸준히 높아지는지



다른 강화학습 알고리즘보다 높은 점수를 받음

Proposed

DDPG

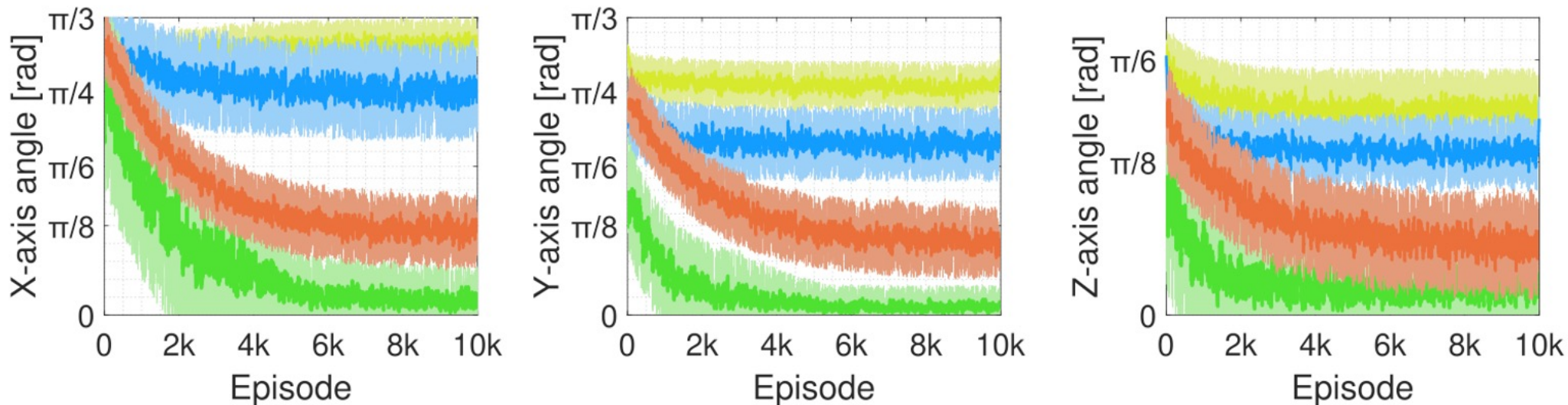
Actor-Critic

DQN

07 Performance Evaluation

Robust Control 평가 기준 (2)

자세 각도: 로켓이 착륙 과정에서 흔들리지 않고 똑바로 착륙하는지



모든 각도 축이 0에 가까워짐 → 착륙 과정에서 흔들림 없이 안정적으로 착지

Proposed

DDPG

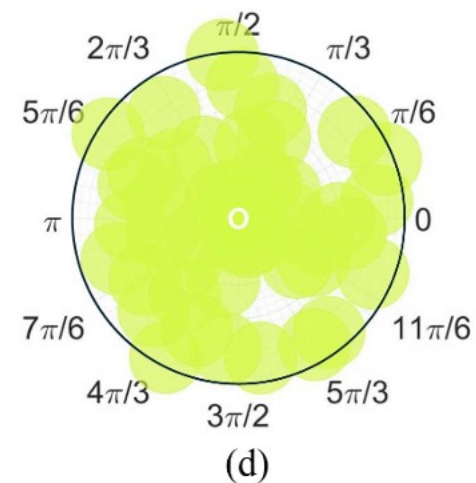
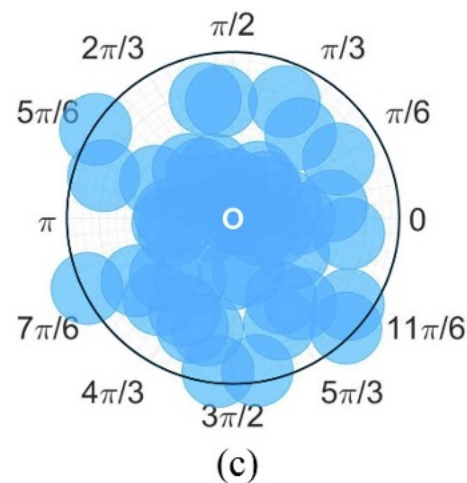
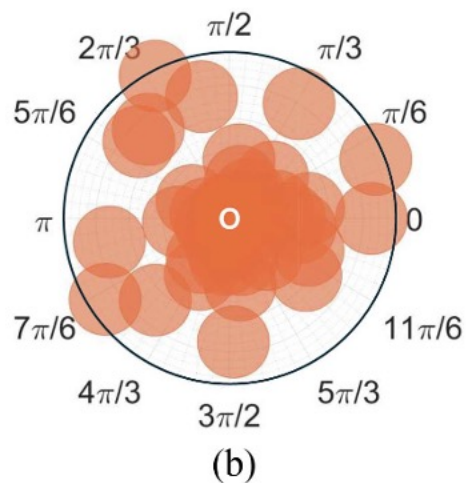
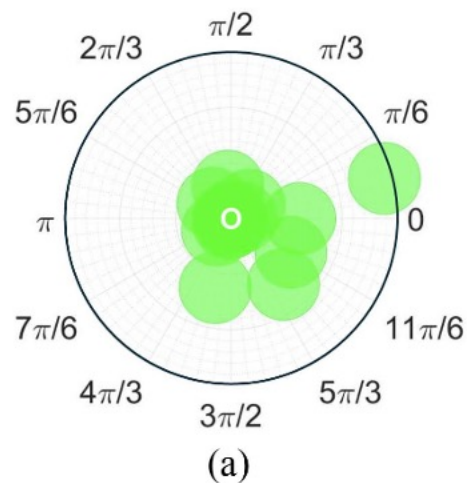
Actor-Critic

DQN

07 Performance Evaluation

Robust Control 평가 기준 (3)

위치 편차: 착륙 지점에서 얼마나 벗어나지 않고 정확하게 착륙하는지



제안한 알고리즘이 목표 지점에 가장 정확하게 착지함

— Proposed

— DDPG

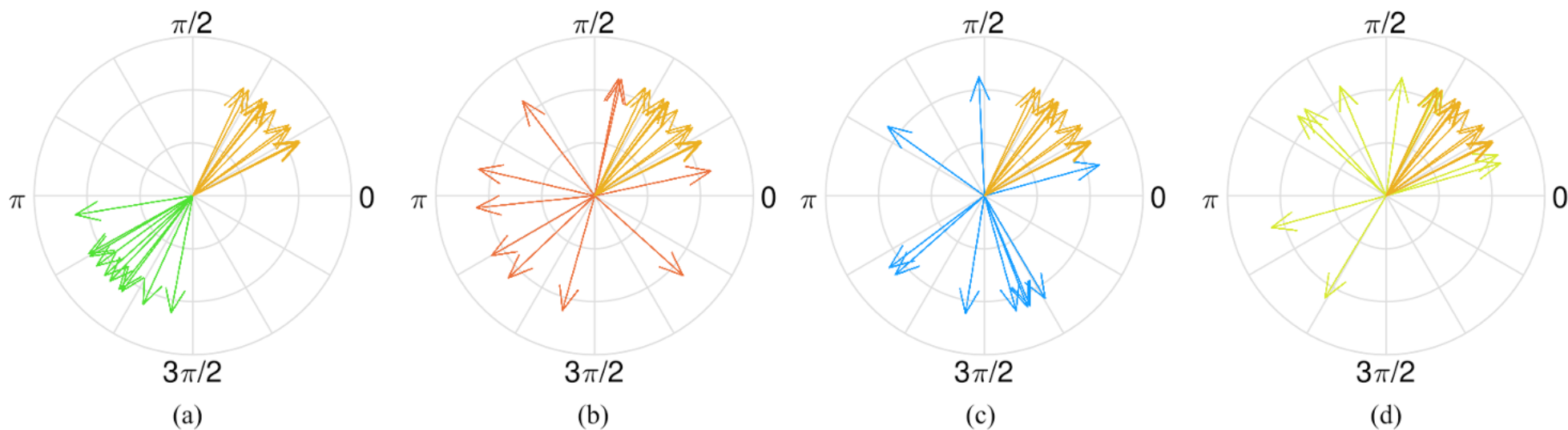
— Actor-Critic

— DQN

07 Performance Evaluation

Robust Control 평가 기준 (4)

추력 벡터 방향: 바람이 불 때 로켓이 바람에 맞춰 엔진의 힘을 올바른 방향으로 조절하는지



바람과 반대 방향으로 정확하게 엔진을 분사함

Proposed

DDPG

Actor-Critic

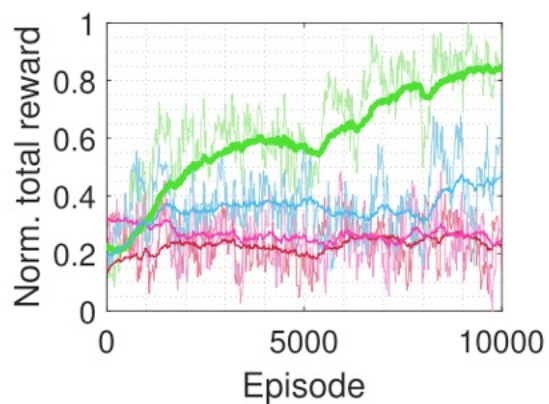
DQN

Wind direction

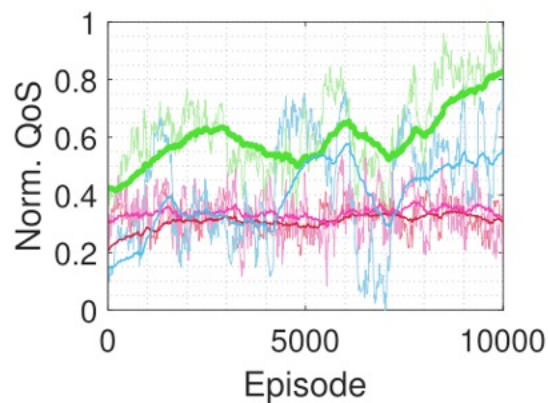
07 Performance Evaluation

Seamless Communication 평가 기준

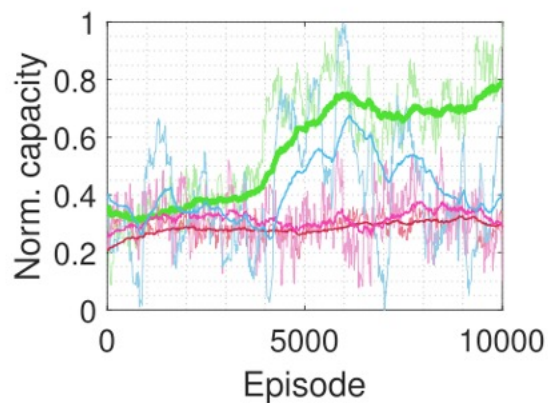
보상 수렴, QoS (서비스 품질), 채널 용량, 잔여 에너지



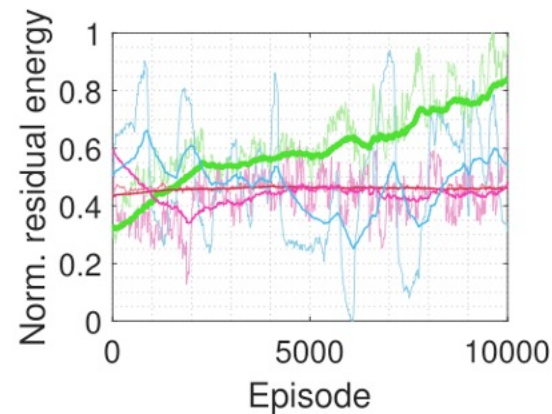
(a)



(b)



(c)



(d)

모든 평가 기준에서 가장 좋은 성능을 보임

(높은 보상, 우수한 통신 품질, 높은 채널 용량, 지상국 에너지 절감)

Proposed

Comp-1

Comp-2

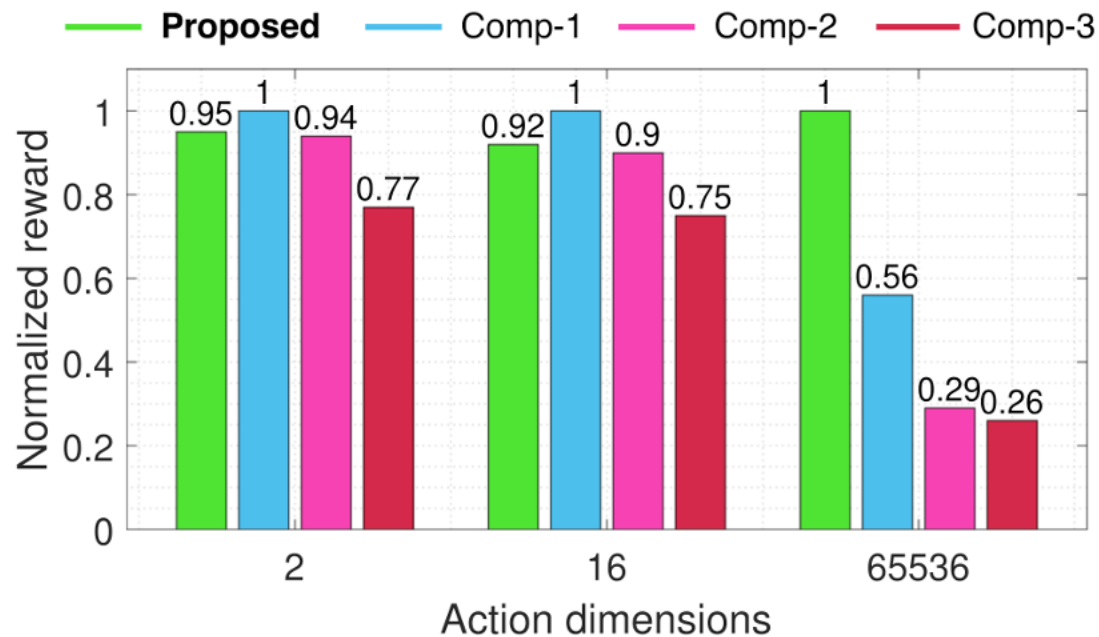
Comp-3

07 Performance Evaluation

차원이 커질수록 압도적으로 유리해지는 알고리즘

기존 알고리즘은 행동의 가짓수(차원)가 많아질수록 학습에 어려움을 겪음

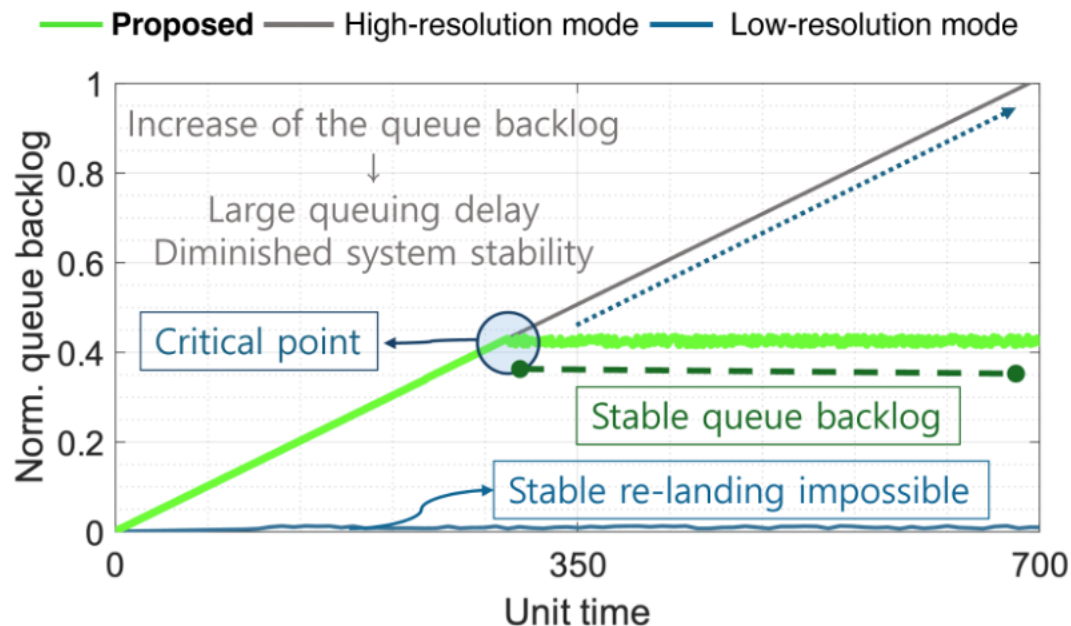
그러나 제시하는 알고리즘은 양자컴퓨터의 성질로 인해 차원이 많아져도 일관된 성능을 보임



07 Performance Evaluation

Energy-Efficient Computing 평가: 이미지 처리

고화질 이미지를 전송하다가, 통신이 지연되는 시점(=큐 백로그 누적)부터 저화질 이미지를 전송

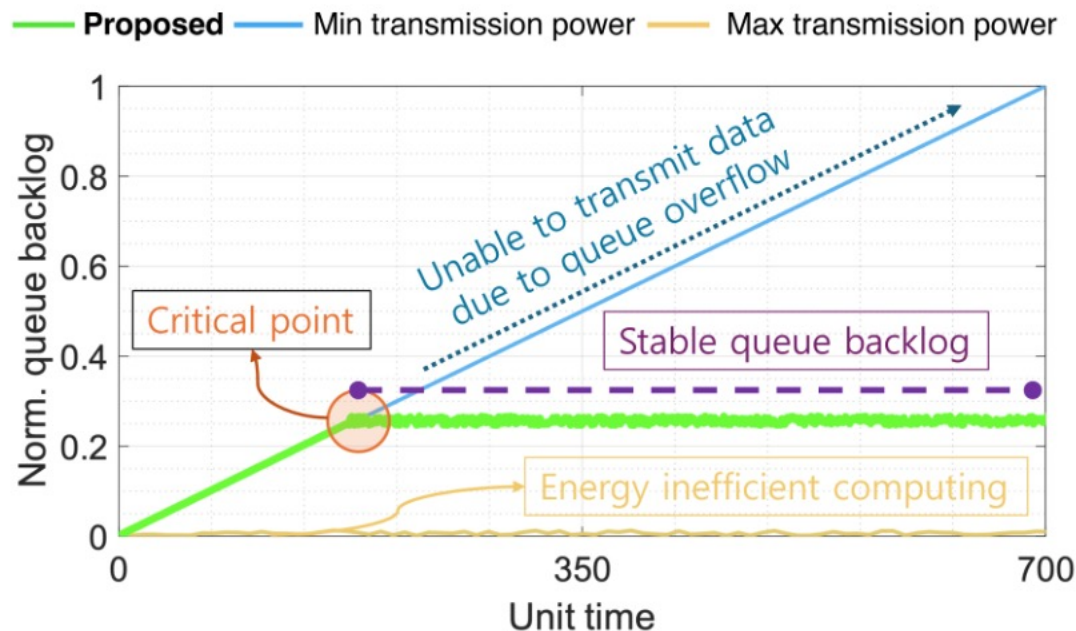


통신 지연 (고화질 이미지의 문제점) & 불안정한 착지 (저화질 이미지의 문제점) 동시 해결

07 Performance Evaluation

Energy-Efficient Computing 평가: 전력 처리

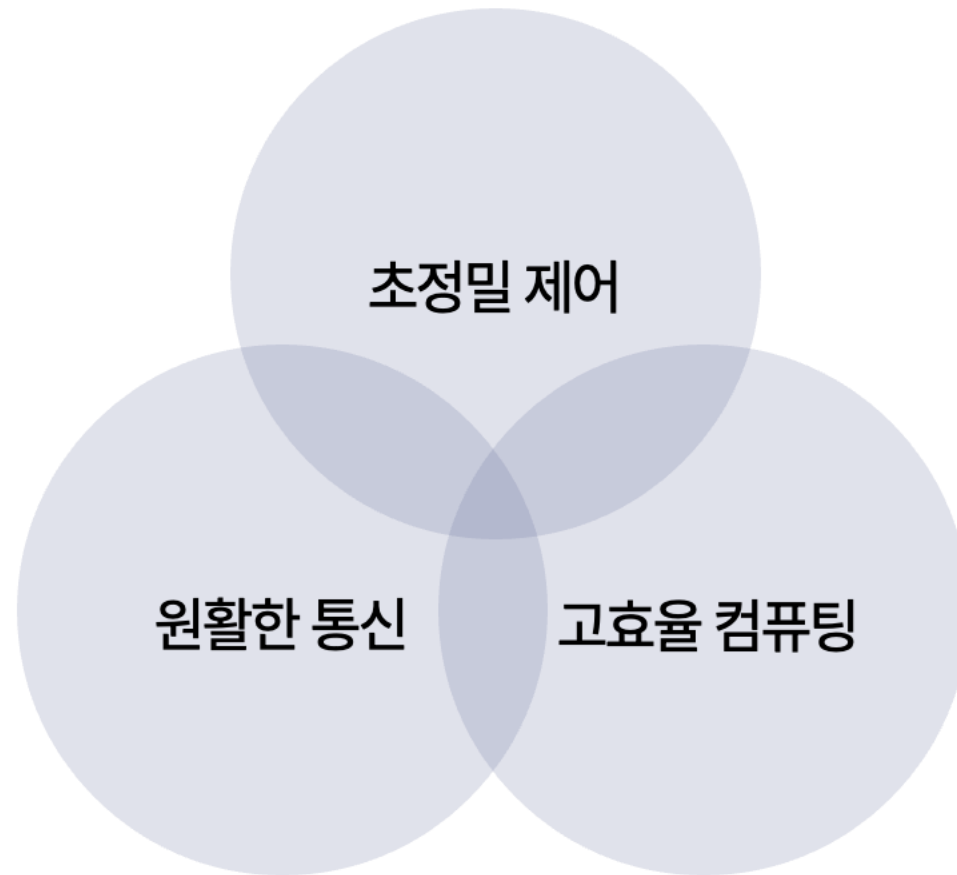
최소 전력으로 데이터를 전송하다가, 통신이 지연되는 시점(=큐 백로그 누적)부터 최대 전력으로 데이터를 전송



통신 지연 (최소 전력의 문제점) & 에너지 고갈 (최대 전력의 문제점) 동시 해결

08 Conclusion

본 기술은 로켓 착륙을 넘어, 우주선 조종, 우주 공간에서의 대규모 IoT 네트워크에도 활용될 가능성이 존재



AI / Data Engineering Track

Thank you

송실대학교 전기공학부 학술 소모임 NOVA
발표자 : 이상윤 (wsy04018@naver.com)