ACM SAC ‘26 – Quantum Software Engineering Practices (QSE) Track

Variational Quantum Rainbow Deep Q-Network for Optimizing Resource Allocation Problem

Hung Nguyen[1], Thinh Nguyen[2], Hung Cao[1]

[1]Analytics Everywhere Lab, University of New Brunswick, Canada
[2]University of Science and Technology of Hanoi, Vietnam

Motivation

Real-World Resource Allocation in Complex Operations

Many real-world systems must **assign multiple tasks to different roles simultaneously:**

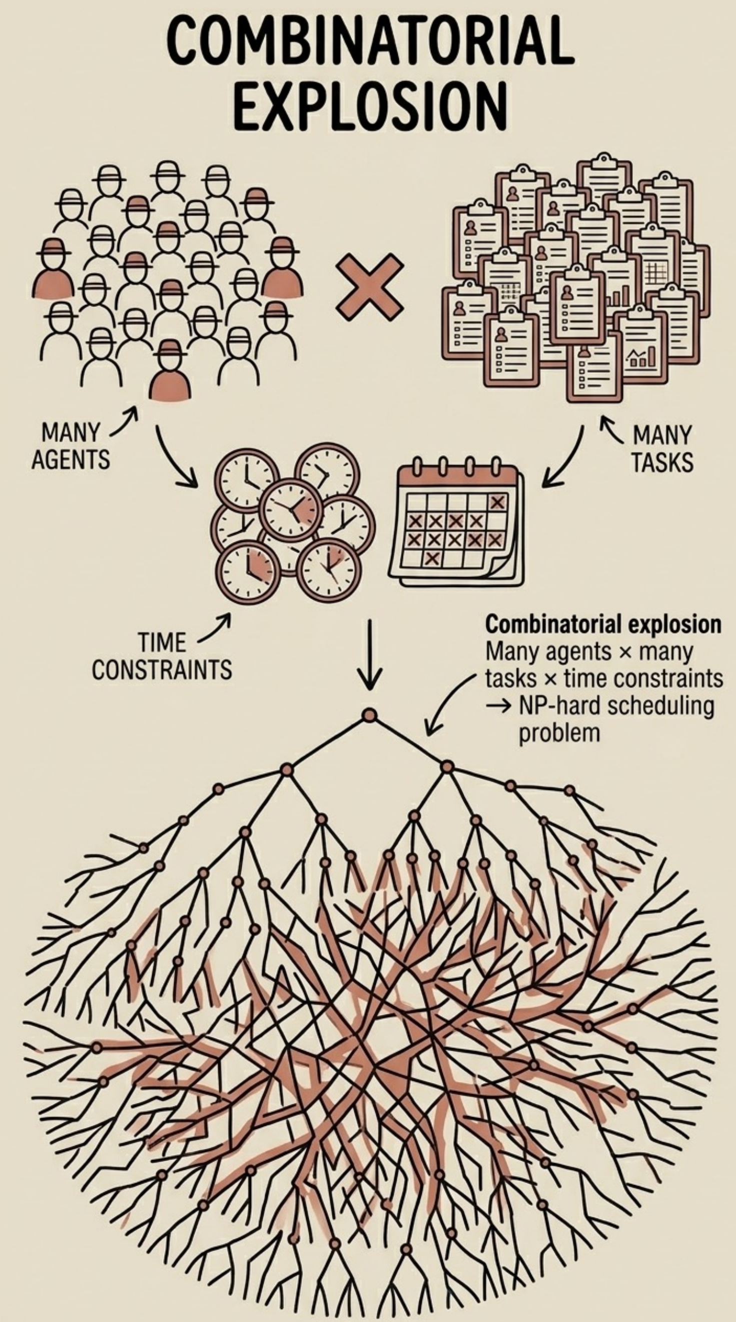
- Healthcare emergency response
- Wildfire disaster management
- Police and public safety operations

Art: Generated by Flow

Combinatorial Explosion

Many agents × many tasks × time constraints

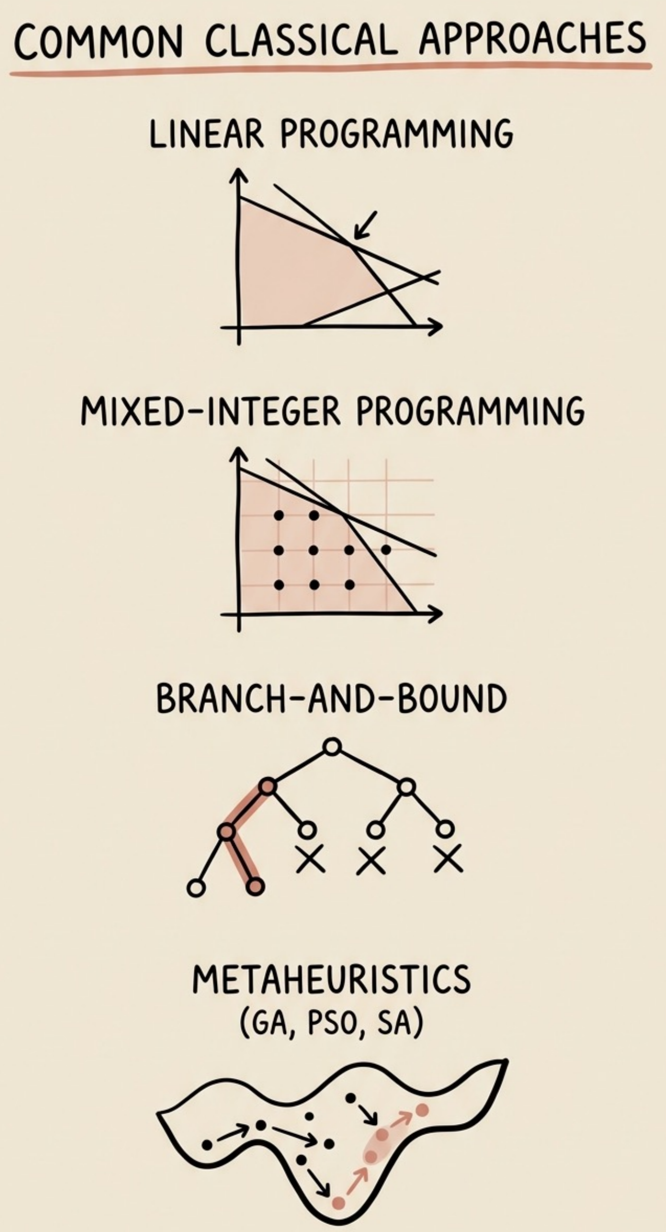
→ NP-hard scheduling problem



Art: Generated by Flow

Traditional Optimization Approaches

- Linear Programming
- Mixed-Integer Programming
- Branch-and-Bound
- Metaheuristics (Genetic Algorithm, Particle Swarm Optimization, Simulated Annealing)

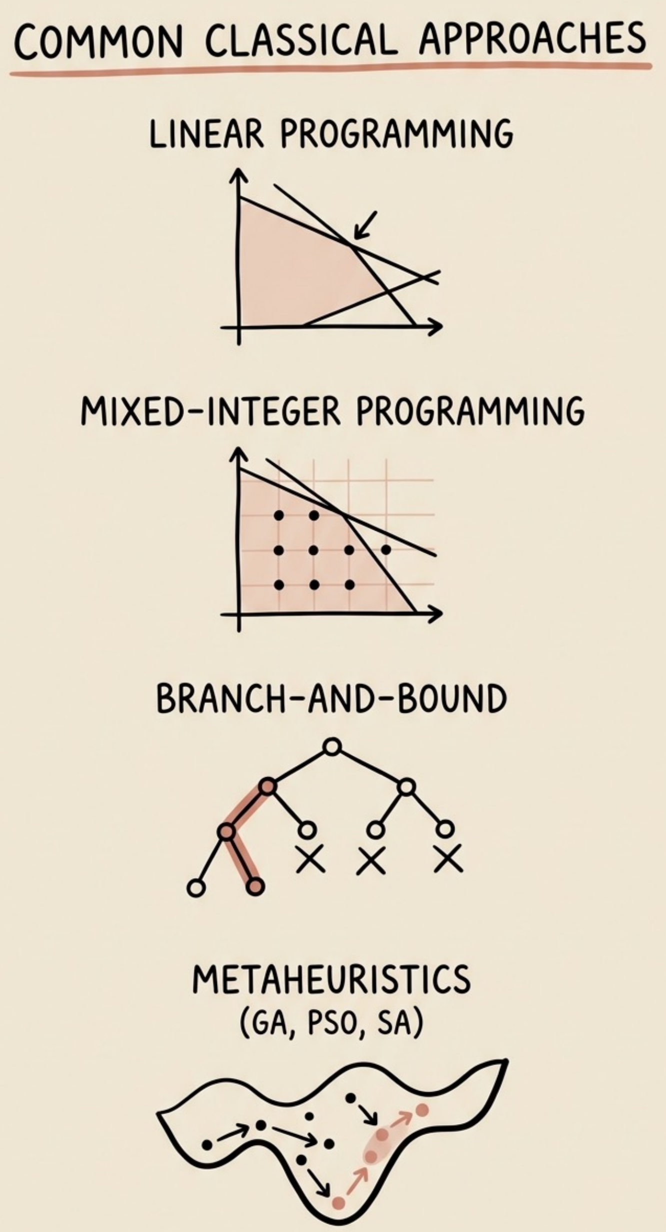


Art: Generated by Flow

Traditional Optimization Approaches

Limitations:

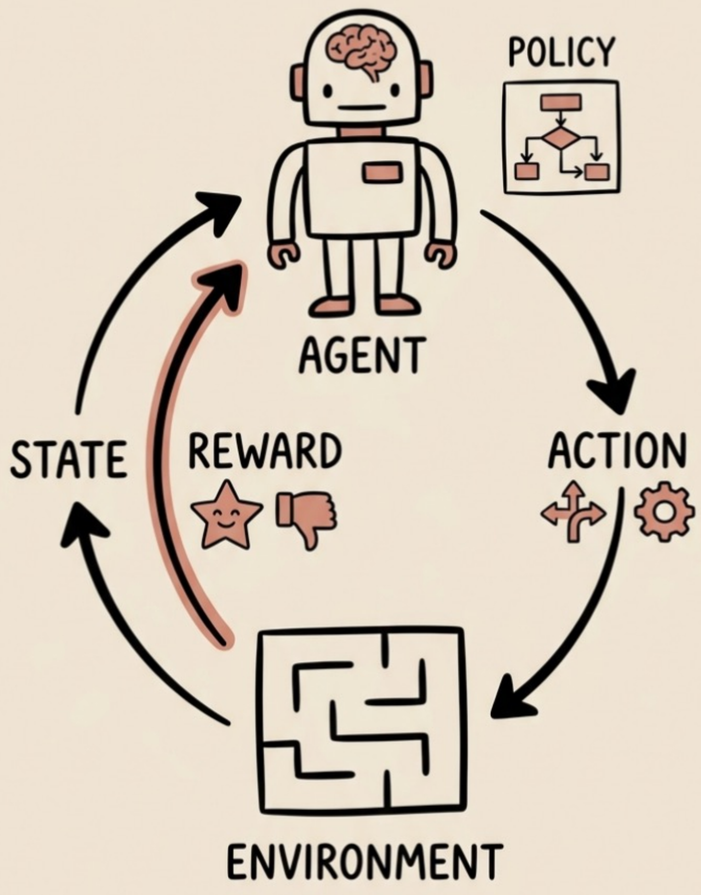
- Poor scalability
- High computational cost
- Static environment modelling assumptions
- Requires hand-crafted constraints and parameter tuning



Art: Generated by Flow

Why Reinforcement Learning

- Learning good decisions through interaction
- Agent learns a scheduling policy
- Used in many complex decision problems



Art: Generated by Flow

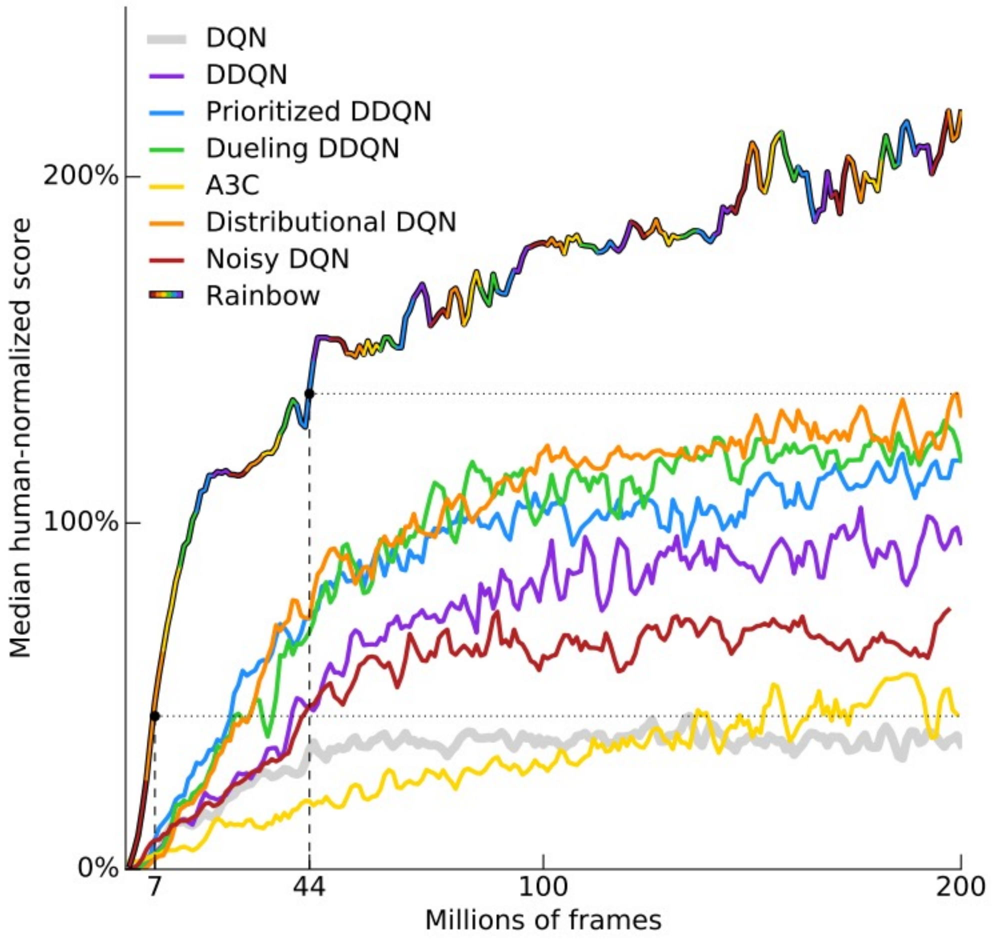
Rainbow DQN

Analytics Everywhere Lab AT THE UNIVERSITY OF NEW BRUNSWICK

One of the strongest DQN variants

Combines multiple RL improvements:

- Double Q-learning
- Prioritized replay
- Distributional RL
- Noisy exploration



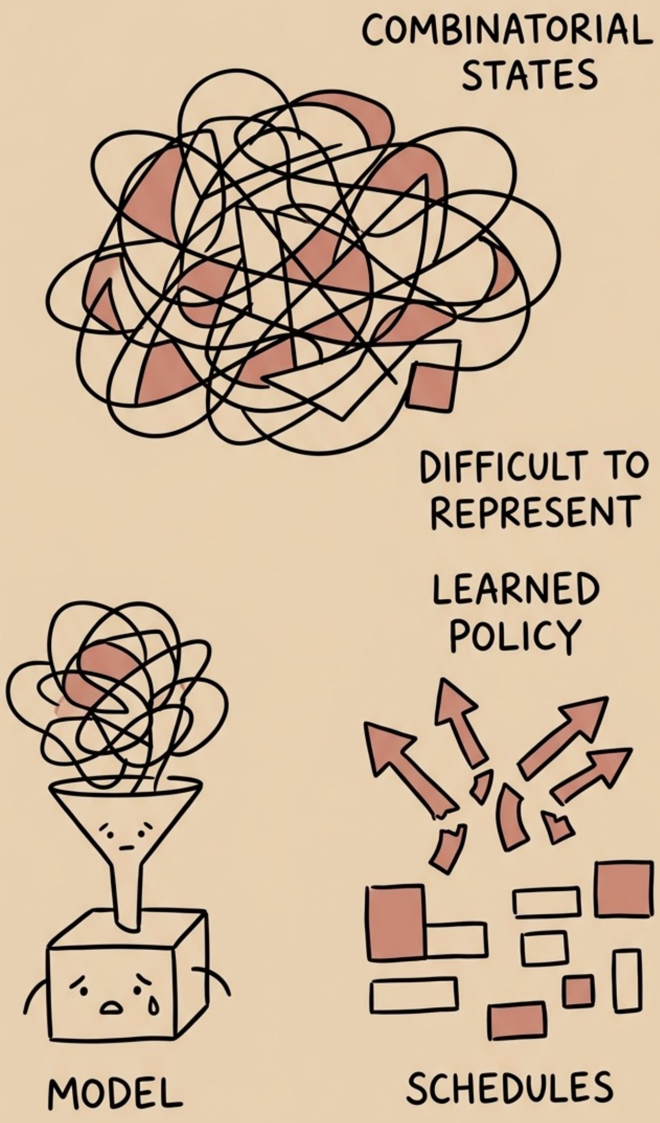
Representation Limit of Classical Deep RL Models

Deep RL models rely on neural networks to approximate value functions.

Key issue: **representation capacity**.

- Complex combinatorial states are difficult to represent.

If the model cannot represent these relationships well, the learned policy may not find optimal schedules.



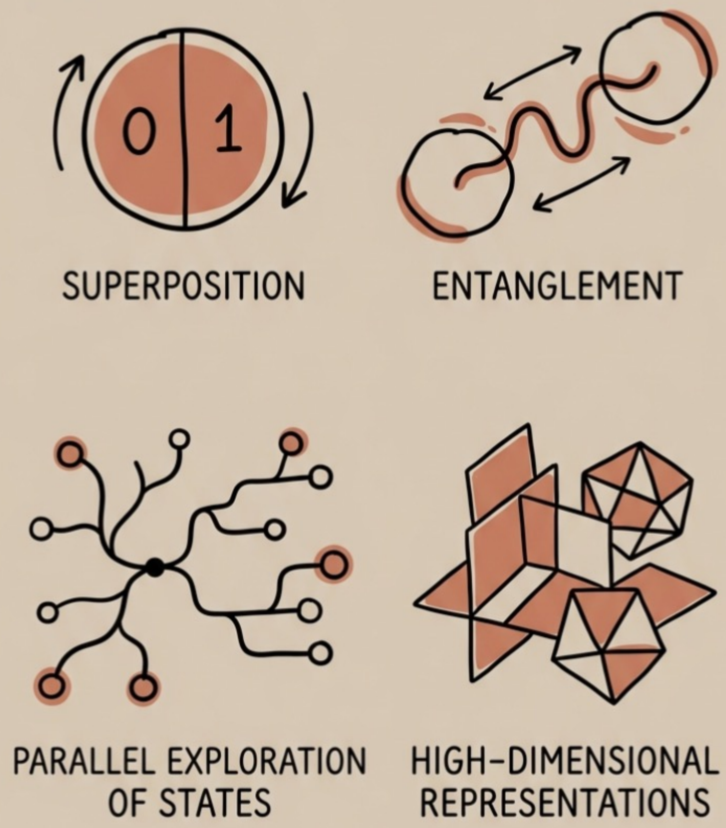
Art: Generated by Flow

Why Quantum RL?

Quantum computing offers:

- Superposition
- Entanglement
- Parallel exploration of states
- High-dimensional representations

Quantum circuits can act as feature extractors in RL systems



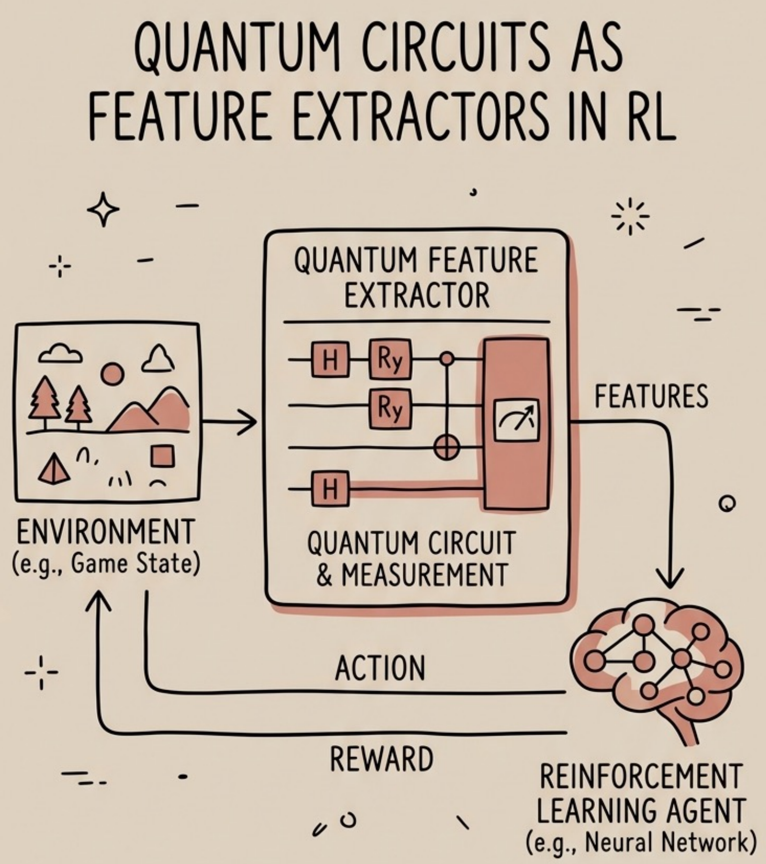
Art: Generated by Flow

Our Proposed Solution

Variational Quantum Rainbow Deep Q-Network (VQR-DQN)

Combine RL with **quantum-enhanced feature extraction:**

- Richer representation of complex decision spaces
- Improved modeling of correlations in combinatorial problems
- Potential for better policies



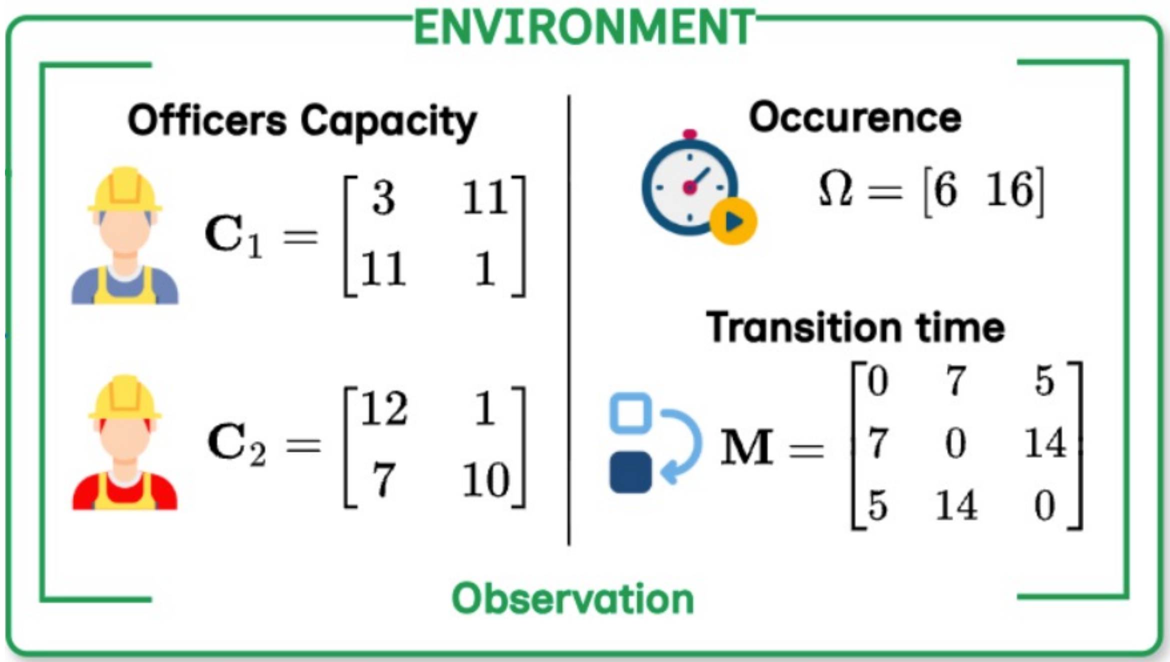
Art: Generated by Flow

Human Resource Allocation Problem (HRAP)

Formulated as an Markov Decision Process (MDP)

Environment simulates:

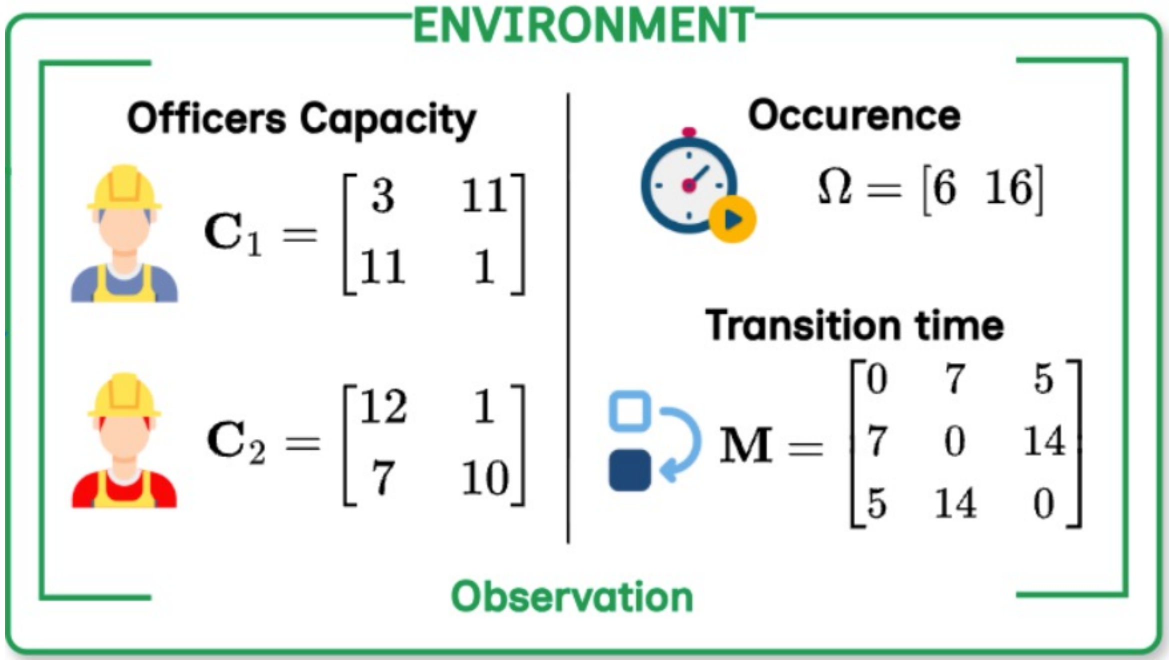
- Officers O
- Events E
- Tasks T



State Representation

State includes:

- Officer capability C_o
- Event occurrence time Ω
- Transition (Travel) time M



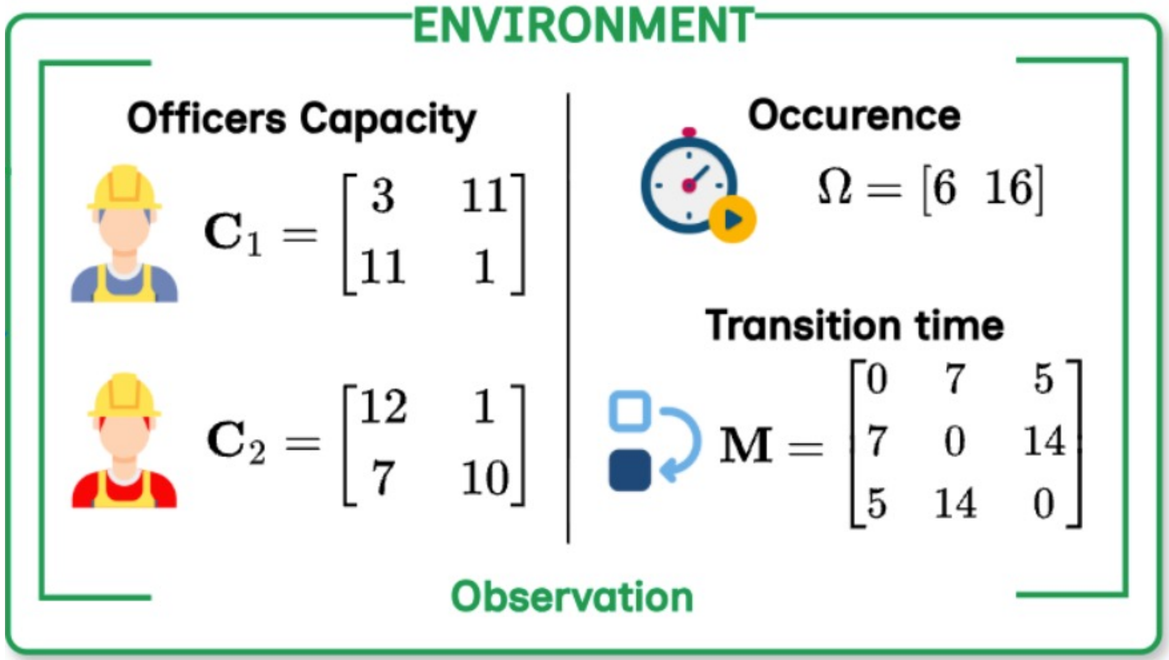
Action Space

Action:

For each task t in event e,
assign officer → task

Action space:

$O^{E \times T}$ grows rapidly with problem size



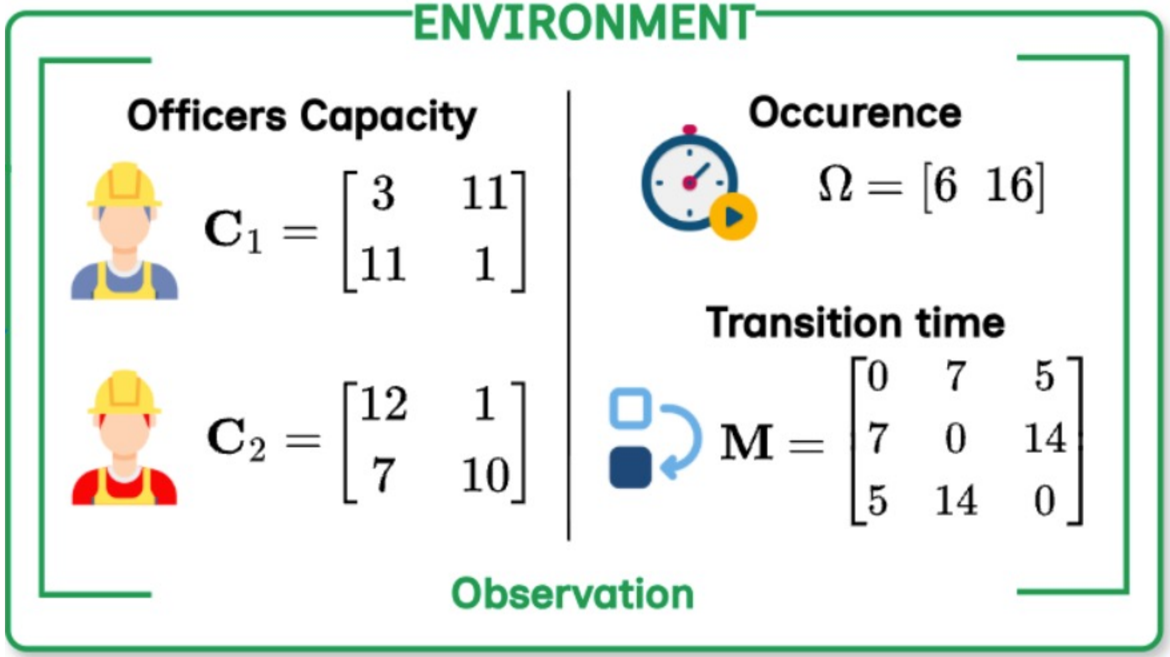
Reward Function

Reward = negative of the maximum completion time across all events

$$r_t = -\frac{\max_e \left(\sum_t C_{o,e,t} + \sum_{\text{transitions}} M_{e_1,e_2}\right)}{\Psi}$$

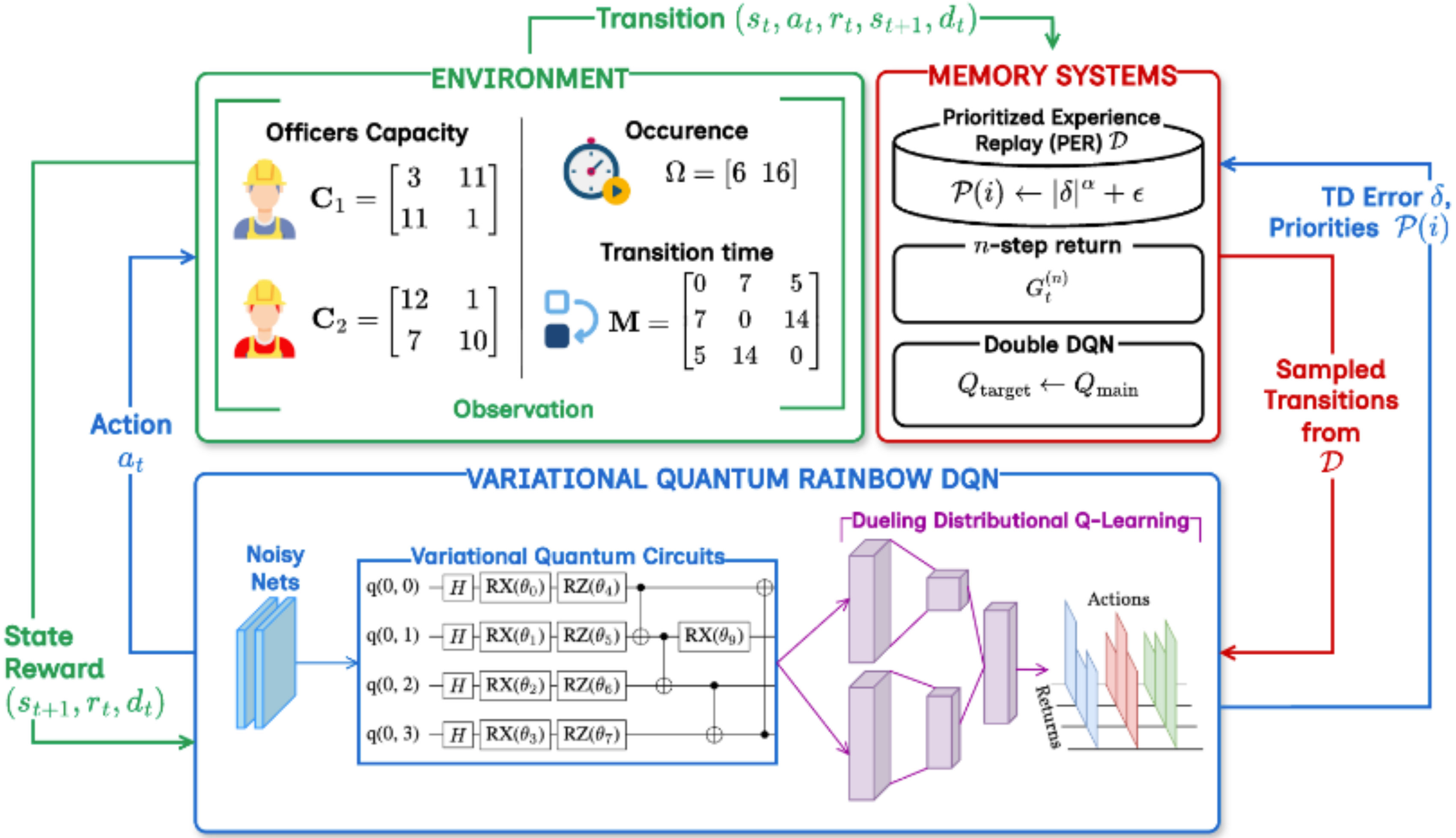
Goal:

- Minimize maximum completion time across all events.
- Prevents one event from becoming a delay bottleneck



VQR-DQN Architecture

Classical layers ⟶ Variational quantum circuit ⟶ RL decision layer

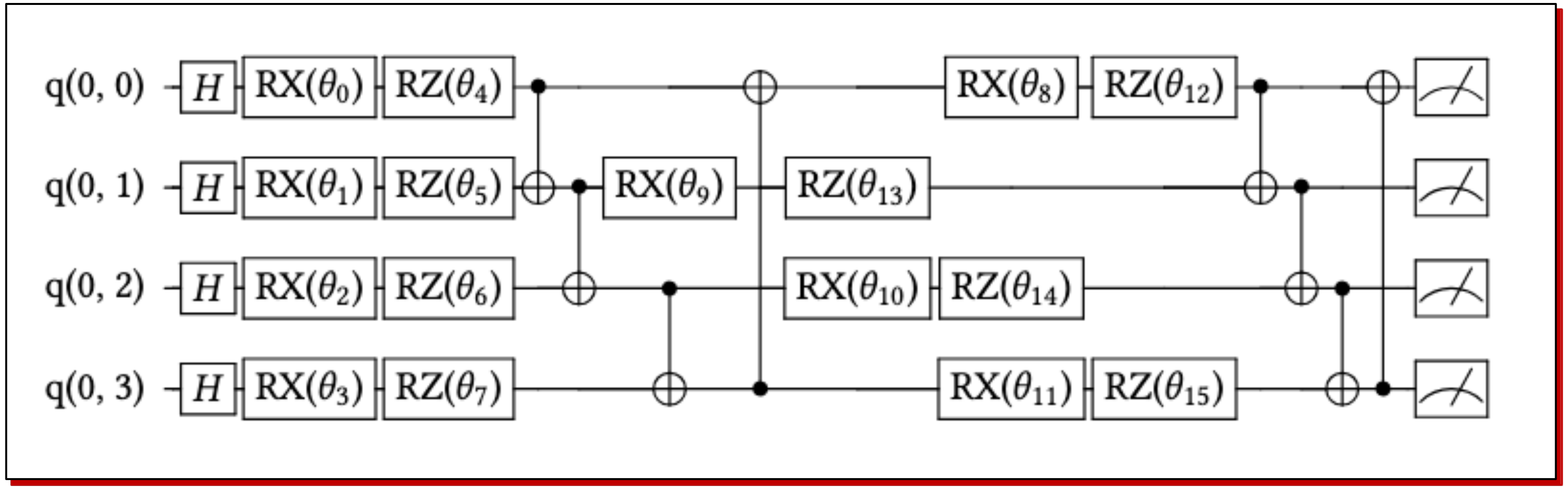


Variational Quantum Circuit

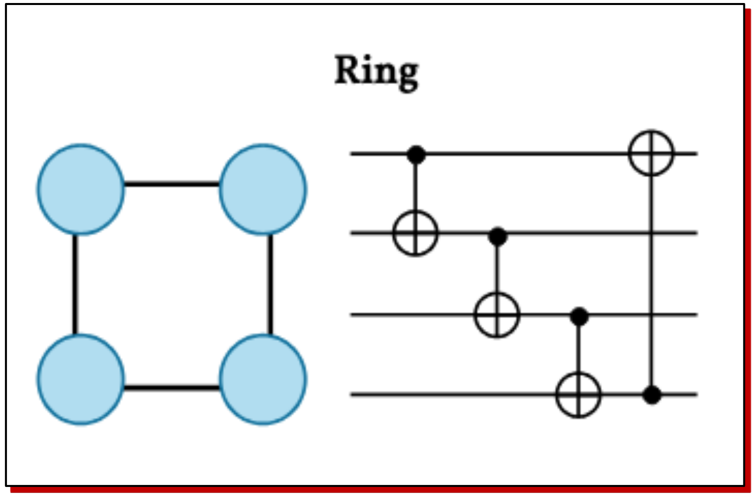
- Variational quantum circuits used as feature extractors
- CNOT gates arranged in ring topology

Ring topology enables:

- Global entanglement
- Better feature correlations
- Improved expressibility



Ansatz with the Ring topology with 4 input qubits and 2 layers



Ring topology

Training Procedure

1. Initialize: Main & target network, PER

2. Agent Interaction: State ⟶ choose action ⟶ receive reward & next state ⟶ store experience

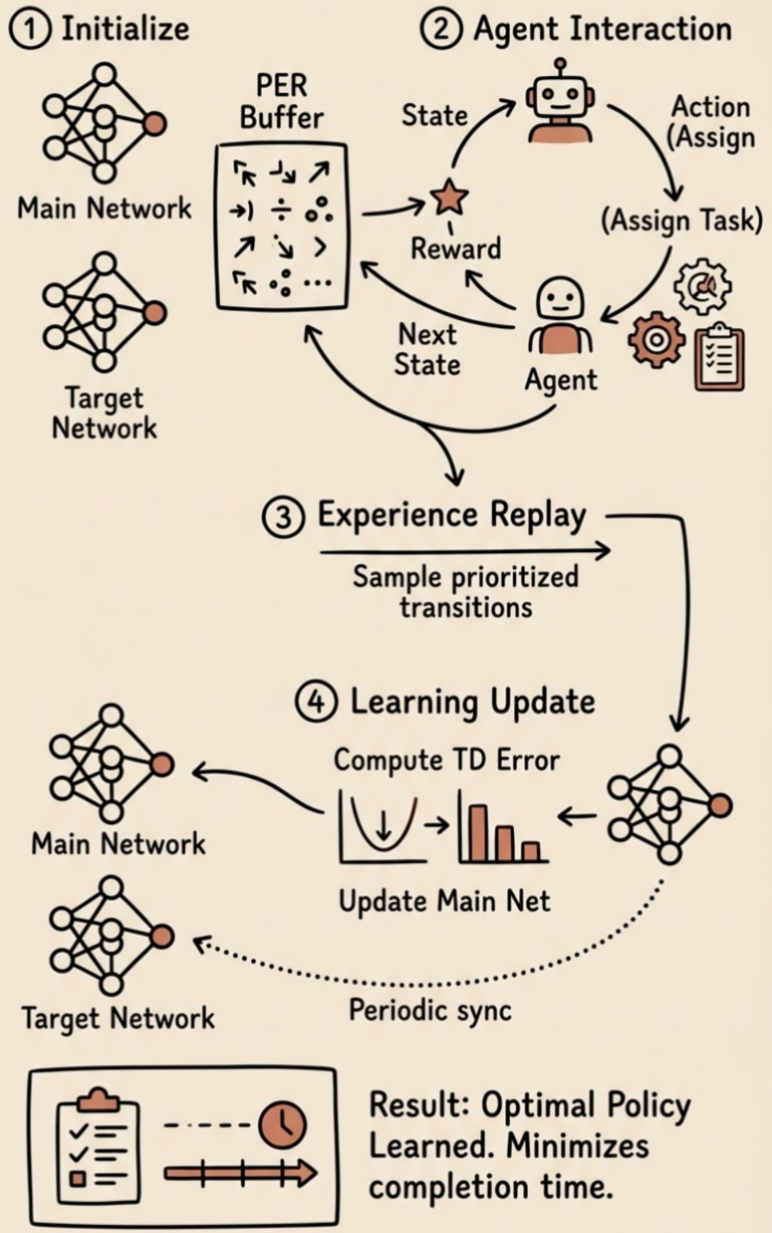
3. Experience Replay: Sample PER past transitions

4. Learning Update: Compute *n*-step return, TD error; Update network parameters

5. Target Network Update: Periodically synchronize

Result: Learns a policy that efficiently allocates roles to tasks and minimizes completion time.

Art: Generated by Flow



Experimental Setup

Simulated with TensorFlow Quantum on IonQ Aria-1

4 benchmark configurations with different officers, events, and travel times

- Training: 50,000 training episodes
- Evaluation: 200 test episodes

Methods:

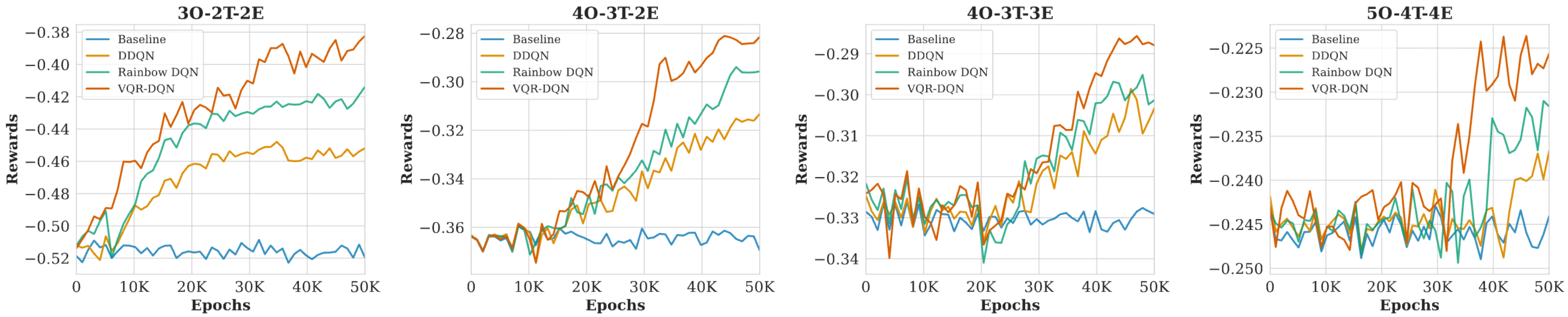
- Baseline: random assignment
- Classical RL: Double DQN (DDQN), Rainbow DQN

Results

Performance improvements:

- 26.8% better than baseline
- 4.9–13.4% better than classical RL

Config.	$\|\mathcal{A}\|$	Baseline	DDQN	Rainbow DQN	VQR-DQN (Ours)
3O-2T-2E	3^4	-0.5225	-0.4539 (▲ 13.1%)	-0.4189 (▲ 19.8%)	**-0.3823 (▲ 26.8%)**
4O-3T-2E	4^6	-0.3689	-0.3132 (▲ 15.1%)	-0.2957 (▲ 19.8%)	**-0.2815 (▲ 23.7%)**
4O-3T-3E	4^9	-0.3316	-0.3032 (▲ 8.6%)	-0.3012 (▲ 9.2%)	**-0.2872 (▲ 13.4%)**
5O-4T-4E	5^{16}	-0.2488	-0.2366 (▲ 4.9%)	-0.2309 (▲ 7.2%)	**-0.2236 (▲ 10.1%)**

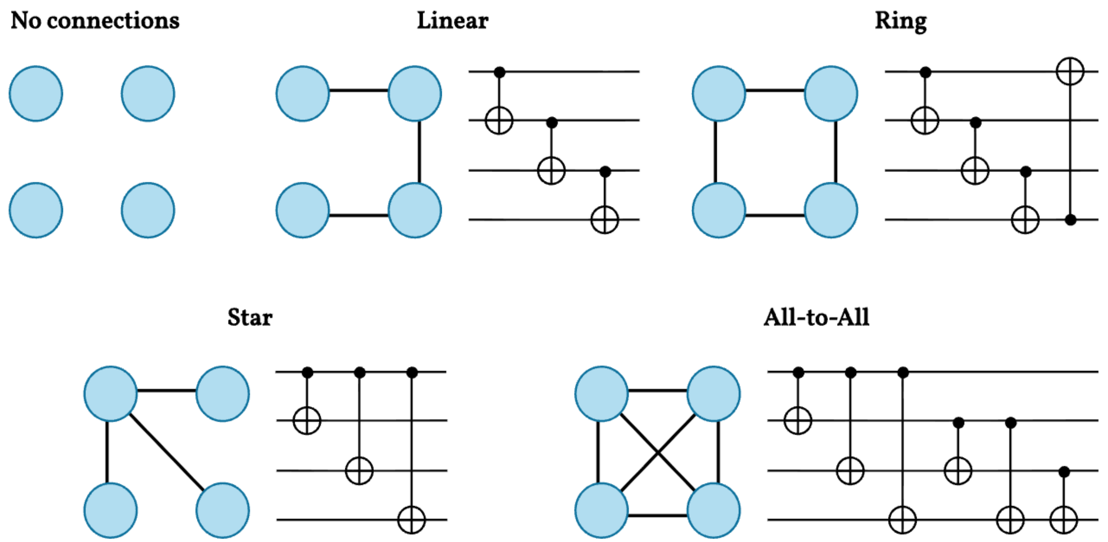


(a) $|\mathcal{A}| = 3^4$ **(b)** $|\mathcal{A}| = 4^6$ **(c)** $|\mathcal{A}| = 4^9$ **(d)** $|\mathcal{A}| = 5^{16}$

Impact of VQC Topologies

Ring topology achieved the best performance:

- Better entanglement distribution
- Higher expressibility
- Stronger ability to capture complex feature correlations



Algorithm	Rewards
Baseline	-0.5225
VQR-DQN + Linear	-0.4249 (▲ 18.7%)
VQR-DQN + Star	-0.4514 (▲ 13.6%)
VQR-DQN + Ring	**-0.3823 (▲ 26.8%)**
VQR-DQN + All-to-All	-0.4103 (▲ 21.5%)

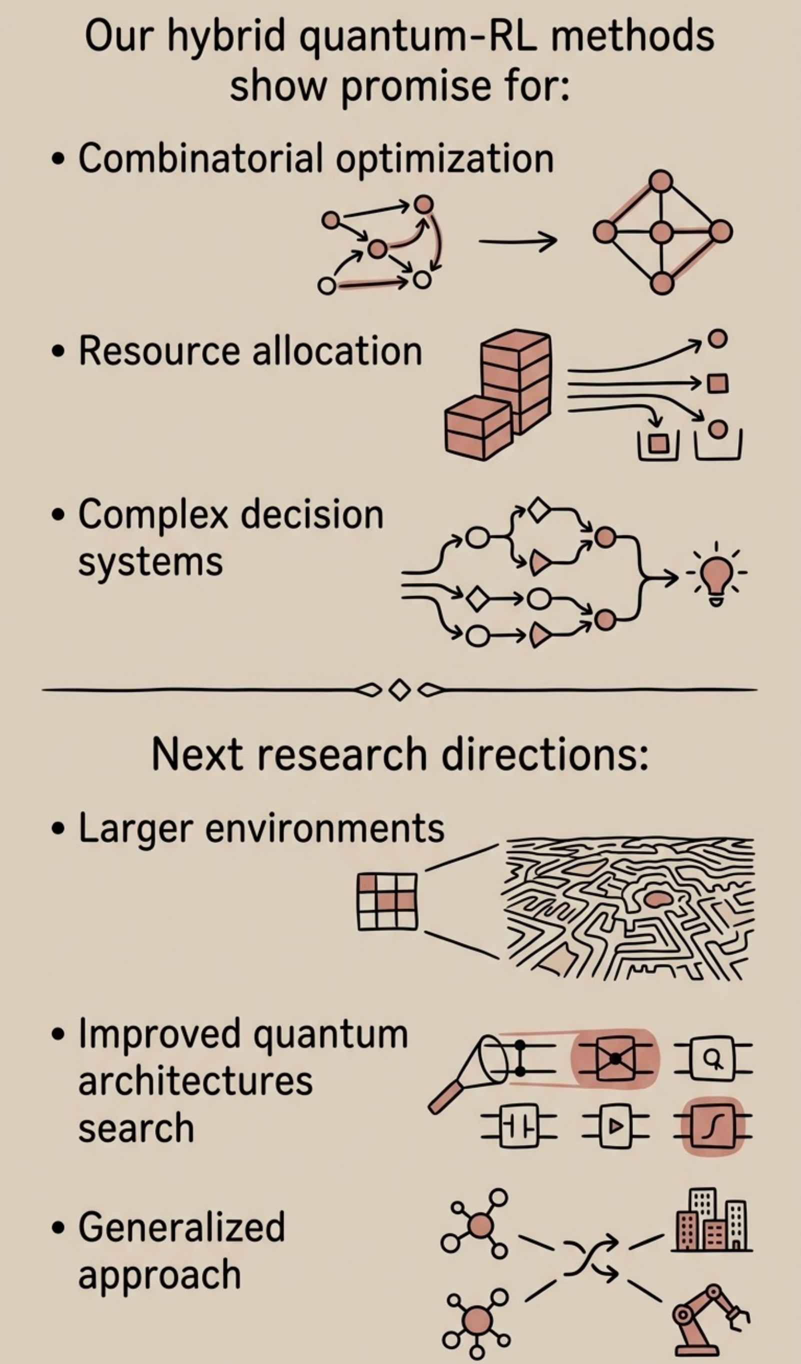
Conclusion & Future Work

Our hybrid quantum-RL methods show promise for:

- Combinatorial optimization
- Resource allocation
- Complex decision systems

Next research directions:

- Larger environments
- Improved quantum architectures search
- Generalized approach



Art: Generated by Flow

Thank you!

Analytics Everywhere Lab AT THE UNIVERSITY OF NEW BRUNSWICK

Our mission

We're recruiting MSc. and PhD students!

If you are interested, contact hung.cao@unb.ca

Hung Cao, PhD

Assistant Professor, Lab Director
Analytics Everywhere Lab
University Of New Brunswick, Canada
hung.cao@unb.ca

Affiliated Faculty

Francis Palma, PhD

Monica Wachowicz, PhD

Trevor Hanson, PhD

Rene Richard, MSc

Students

Hung Nguyen

Asfia Kawnine

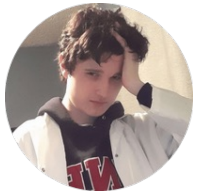
Atah Nuh Mih

Alireza Rahimi

Pavi P

…

Krishno Dey



Connor McLenaghan

Ishan Randeniya

Simran Dadhich

Bohdan Savchuk

…