Deep Reinforcement Learning for Quantum Network Routing

A Comparative Study on NSFNET Topology

Hamza Berahma Dr. Abbou

December 10, 2025

Al Akhawayn University in Ifrane
School of Science and Engineering

Outline

Introduction

Methodology

Results

Analysis

Discussion

Conclusion

Introduction

The Quantum Routing Challenge

Quantum Internet Vision

- Secure quantum key distribution
- Distributed quantum computing
- Quantum sensing networks

Unique Constraints

- Quantum decoherence
- No-cloning theorem
- Fidelity degradation

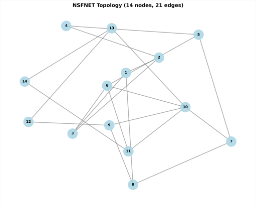


Figure 1: NSFNET Topology

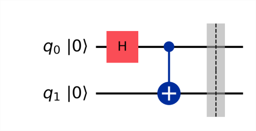


Figure 2: Bell State Generation

The Quantum Routing Challenge

Quantum Internet Vision

- Secure quantum key distribution
- Distributed quantum computing
- Quantum sensing networks

Unique Constraints

- Quantum decoherence
- No-cloning theorem
- Fidelity degradation

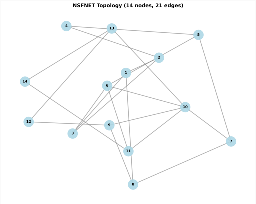


Figure 1: NSFNET Topology

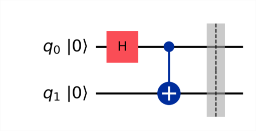


Figure 2: Bell State Generation

The Quantum Routing Challenge

Quantum Internet Vision

- Secure quantum key distribution
- Distributed quantum computing
- Quantum sensing networks

Unique Constraints

- Quantum decoherence
- No-cloning theorem
- Fidelity degradation

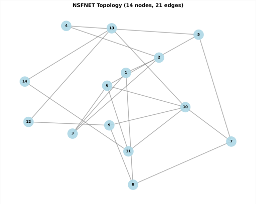


Figure 1: NSFNET Topology

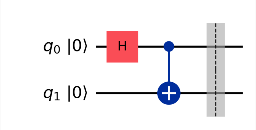


Figure 2: Bell State Generation

The Quantum Routing Challenge

Quantum Internet Vision

- Secure quantum key distribution
- Distributed quantum computing
- Quantum sensing networks

Unique Constraints

- Quantum decoherence
- No-cloning theorem
- Fidelity degradation

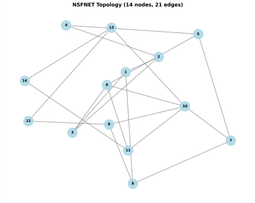


Figure 1: NSFNET Topology

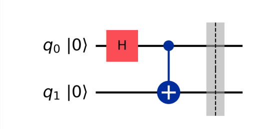


Figure 2: Bell State Generation

Problem: Classical Routing Fails

Shortest Path Algorithm
Optimizes only path length:

$$\min \sum_{i \in \text{path}} 1$$

- Ignores fidelity constraints
- May traverse low-fidelity links
- Cannot adapt to dynamics

Greedy Fidelity
Maximizes immediate fidelity:

$$\max_a F(s, a)$$

Our Solution: DRL
Learns optimal policy:

$$\pi^* = \arg\max_\pi \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R_t\right]$$

- Balances multiple objectives
- Adapts to network dynamics
- Learns from experience

Problem: Classical Routing Fails

Shortest Path Algorithm
Optimizes only path length:

$$\min \sum_{i \in \text{path}} 1$$

- Ignores fidelity constraints
- May traverse low-fidelity links
- Cannot adapt to dynamics

Greedy Fidelity
Maximizes immediate fidelity:

$$\max_a F(s, a)$$

Our Solution: DRL
Learns optimal policy:

$$\pi^* = \arg\max_\pi \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R_t\right]$$

- Balances multiple objectives
- Adapts to network dynamics
- Learns from experience

Problem: Classical Routing Fails

Shortest Path Algorithm
Optimizes only path length:

$$\min \sum_{i \in \text{path}} 1$$

- Ignores fidelity constraints
- May traverse low-fidelity links
- Cannot adapt to dynamics

Greedy Fidelity
Maximizes immediate fidelity:

$$\max_a F(s, a)$$

Our Solution: DRL
Learns optimal policy:

$$\pi^* = \arg\max_\pi \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R_t\right]$$

- Balances multiple objectives
- Adapts to network dynamics
- Learns from experience

Problem: Classical Routing Fails

Shortest Path Algorithm

Optimizes only path length:

$$\min \sum_{i \in \text{path}} 1$$

- Ignores fidelity constraints
- May traverse low-fidelity links
- Cannot adapt to dynamics

Greedy Fidelity

Maximizes immediate fidelity:

$$\max_a F(s, a)$$

Our Solution: DRL

Learns optimal policy:

$$\pi^* = \arg\max_\pi \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R_t\right]$$

- Balances multiple objectives
- Adapts to network dynamics
- Learns from experience

Success Rate Comparison

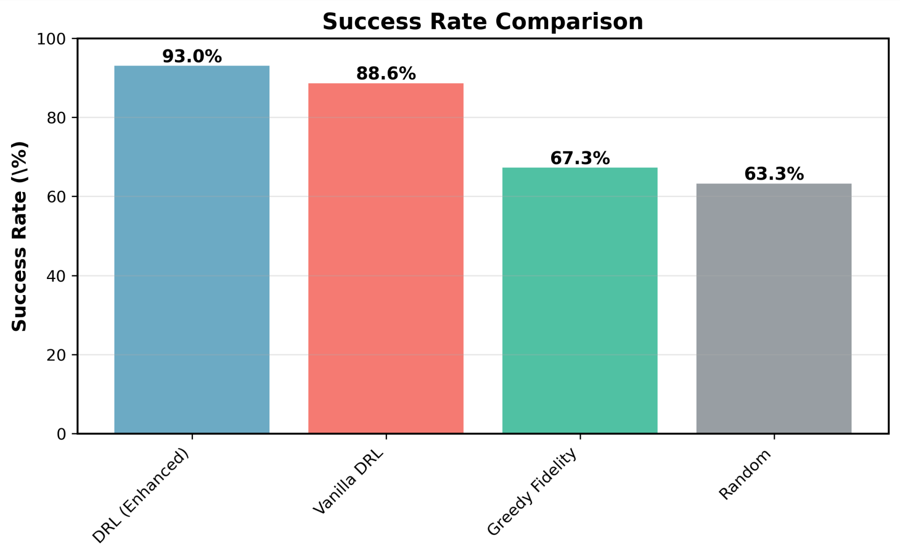


Figure 3: Success Rate Comparison Across All Algorithms

Success Rate Comparison

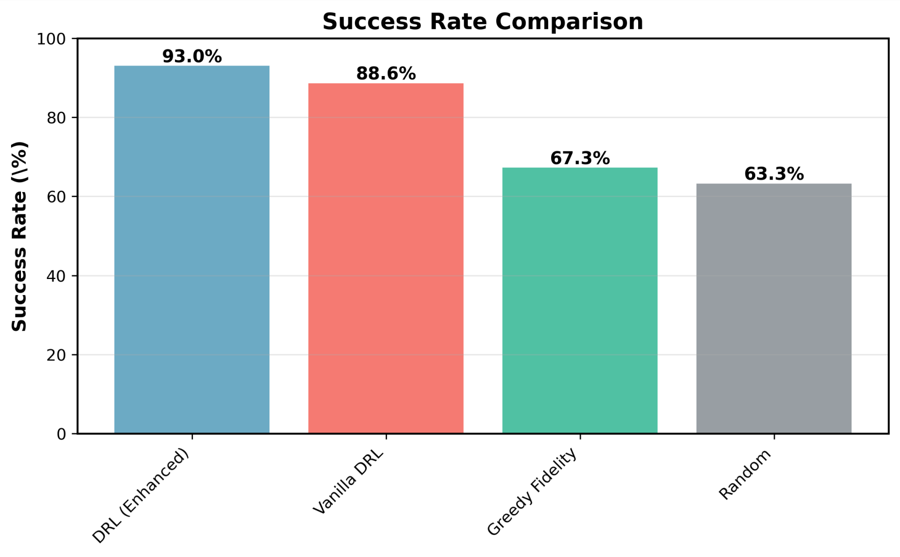


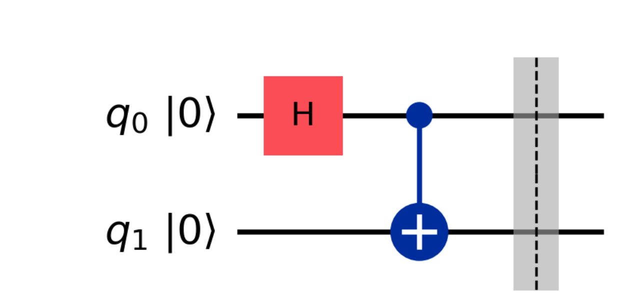
Figure 3: Success Rate Comparison Across All Algorithms

Methodology

Quantum Operations: Bell State Generation

Entanglement Creation
Creates maximally entangled Bell state:

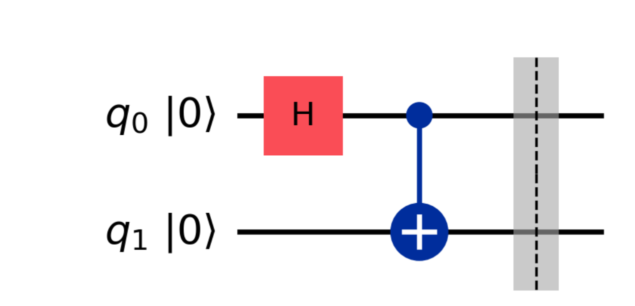
$$|\Phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$



Quantum Operations: Bell State Generation

Entanglement Creation
Creates maximally entangled Bell state:

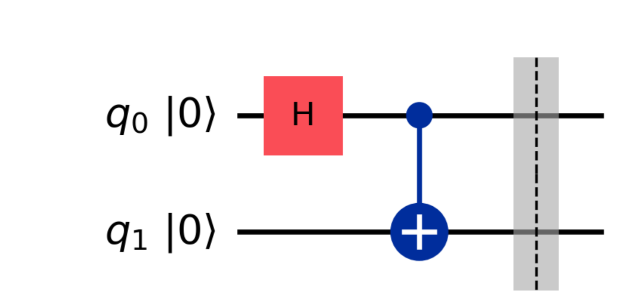
$$|\Phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$



Quantum Operations: Bell State Generation

Entanglement Creation
Creates maximally entangled Bell state:

$$|\Phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$

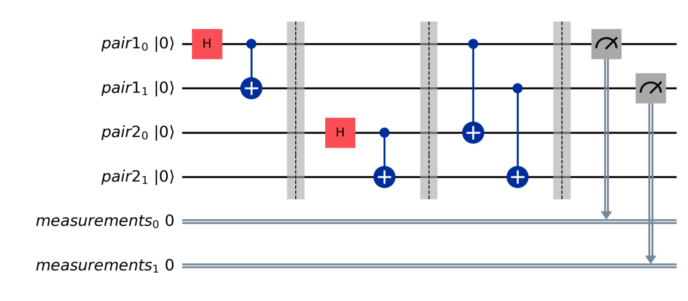


Quantum Operations: BBPSSW Purification

Fidelity Improvement

Improves entanglement fidelity by sacrificing multiple pairs:

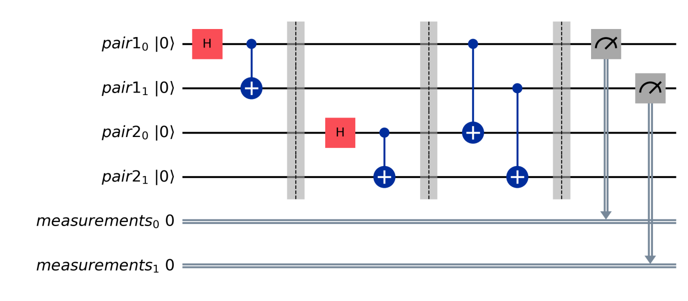
$$F_{\text{out}} = \frac{F_{\text{in}}^2}{F_{\text{in}}^2 + (1 - F_{\text{in}})^2}$$



Quantum Operations: BBPSSW Purification

Fidelity Improvement
Improves entanglement fidelity by sacrificing multiple pairs:

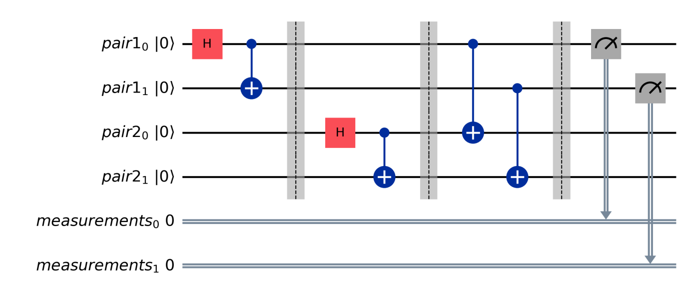
$$F_{\text{out}} = \frac{F_{\text{in}}^2}{F_{\text{in}}^2 + (1 - F_{\text{in}})^2}$$



Quantum Operations: BBPSSW Purification

Fidelity Improvement
Improves entanglement fidelity by sacrificing multiple pairs:

$$F_{\text{out}} = \frac{F_{\text{in}}^2}{F_{\text{in}}^2 + (1 - F_{\text{in}})^2}$$



Dueling DQN Architecture

Key Innovation
Decompose Q-function:

$$Q(s,a) = V(s) + \left(A(s,a) - \frac{1}{|A|} \sum_{a'} A(s,a') \right)$$

where $V(s)$ estimates state value, $A(s,a)$ estimates advantage

Benefits

- Learn state values independently
- Better action discrimination
- Improved sample efficiency

Key Innovation

Decompose Q-function:

$$Q(s,a) = V(s) + \left(A(s,a) - \frac{1}{|A|} \sum_{a'} A(s,a') \right)$$

where $V(s)$ estimates state value, $A(s,a)$ estimates advantage

Benefits

- Learn state values independently
- Better action discrimination
- Improved sample efficiency

Dueling DQN Architecture

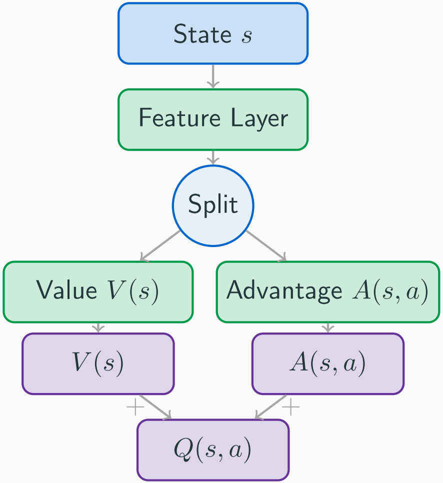
Key Innovation
Decompose Q-function:

$$Q(s,a) = V(s) + \left(A(s,a) - \frac{1}{|A|} \sum_{a'} A(s,a') \right)$$

where $V(s)$ estimates state value, $A(s,a)$ estimates advantage

Benefits

- Learn state values independently
- Better action discrimination
- Improved sample efficiency



Prioritized Experience Replay

Core Idea
Sample transitions with probability:

$$P(i) = \frac{p_i^\alpha}{\sum_k p_k^\alpha}$$

where $p_i = |\delta_i| + \epsilon$ (TD-error)

Importance sampling weight:

$$w_i = \left(\frac{1}{N} \cdot \frac{1}{P(i)}\right)^\beta$$

Learn more from surprising outcomes

Prioritized Experience Replay

Core Idea
Sample transitions with probability:

$$P(i) = \frac{p_i^\alpha}{\sum_k p_k^\alpha}$$

where $p_i = |\delta_i| + \epsilon$ (TD-error)

Importance sampling weight:

$$w_i = \left(\frac{1}{N} \cdot \frac{1}{P(i)}\right)^\beta$$

Learn more from surprising outcomes

Prioritized Experience Replay

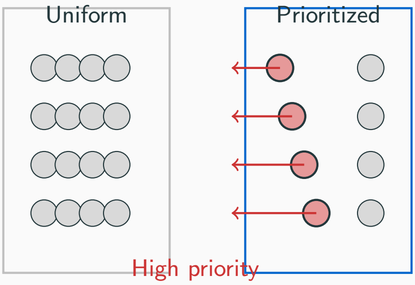
Core Idea
Sample transitions with probability:

$$P(i) = \frac{p_i^\alpha}{\sum_k p_k^\alpha}$$

where $p_i = |\delta_i| + \epsilon$ (TD-error)

Importance sampling weight:

$$w_i = \left(\frac{1}{N} \cdot \frac{1}{P(i)}\right)^\beta$$



Learn more from surprising outcomes

Double DQN

Problem
Standard DQN overestimates Q-values

Solution: Decouple

- Action selection: Main network
- Value estimation: Target network

$$a^* = \arg\max_a Q(s', a; \theta)$$

$$y = r + \gamma Q(s', a^*; \theta^-)$$

Double DQN

Problem
Standard DQN overestimates Q-values

Solution: Decouple

- Action selection: Main network
- Value estimation: Target network

$$a^* = \arg\max_a Q(s', a; \theta)$$

$$y = r + \gamma Q(s', a^*; \theta^-)$$

Double DQN

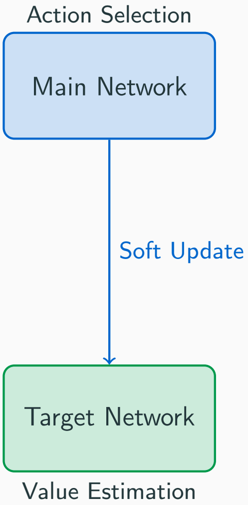
Problem
Standard DQN overestimates Q-values

Solution: Decouple

- Action selection: Main network
- Value estimation: Target network

$$a^* = \arg\max_a Q(s', a; \theta)$$

$$y = r + \gamma Q(s', a^*; \theta^-)$$



Double DQN

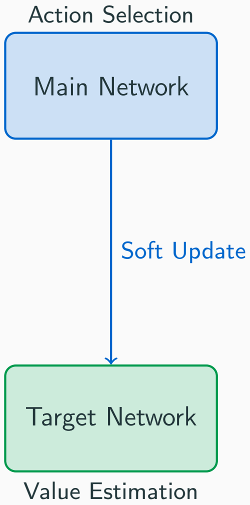
Problem
Standard DQN overestimates Q-values

Solution: Decouple

- Action selection: Main network
- Value estimation: Target network

$$a^* = \arg\max_a Q(s', a; \theta)$$

$$y = r + \gamma Q(s', a^*; \theta^-)$$



Training Configuration

Hyperparameters

Learning rate	$\alpha = 10^{-4}$
Batch size	$B = 128$
Target update	$\tau = 0.005$ (soft)
Discount factor	$\gamma = 0.99$
Replay buffer	$\lvert\mathcal{D}\rvert = 50,000$

Training Details

- Optimizer: AdamW
- Training episodes: 5,000

Training Configuration

Hyperparameters

Learning rate	$\alpha = 10^{-4}$
Batch size	$B = 128$
Target update	$\tau = 0.005$ (soft)
Discount factor	$\gamma = 0.99$
Replay buffer	$\lvert\mathcal{D}\rvert = 50,000$

Training Details

- Optimizer: AdamW
- Training episodes: 5,000

Results

Overall Performance: Success Rates

Algorithm	Success Rate (%)
DRL (Enhanced)	**93.01** ± 2.1
Vanilla DRL	88.62 ± 2.5
Greedy Fidelity	67.27 ± 3.2
Random	63.27 ± 3.4

Key Finding

4.39 percentage points improvement over Vanilla DRL

Overall Performance: Success Rates

Algorithm	Success Rate (%)
DRL (Enhanced)	**93.01** $\pm$ 2.1
Vanilla DRL	88.62 $\pm$ 2.5
Greedy Fidelity	67.27 $\pm$ 3.2
Random	63.27 $\pm$ 3.4

Key Finding

4.39 percentage points improvement over Vanilla DRL

Overall Performance: Success Rates

Algorithm	Success Rate (%)
DRL (Enhanced)	**93.01** ± 2.1
Vanilla DRL	88.62 ± 2.5
Greedy Fidelity	67.27 ± 3.2
Random	63.27 ± 3.4

Key Finding

4.39 percentage points improvement over Vanilla DRL

Overall Performance: Success Rates

Algorithm	Success Rate (%)
DRL (Enhanced)	**93.01** $\pm$ 2.1
Vanilla DRL	88.62 $\pm$ 2.5
Greedy Fidelity	67.27 $\pm$ 3.2
Random	63.27 $\pm$ 3.4

Key Finding

4.39 percentage points improvement over Vanilla DRL

Overall Performance: Success Rates

Algorithm	Success Rate (%)
DRL (Enhanced)	**93.01** $\pm$ 2.1
Vanilla DRL	88.62 $\pm$ 2.5
Greedy Fidelity	67.27 $\pm$ 3.2
Random	63.27 $\pm$ 3.4

Key Finding

4.39 percentage points improvement over Vanilla DRL

Overall Performance: Visualization

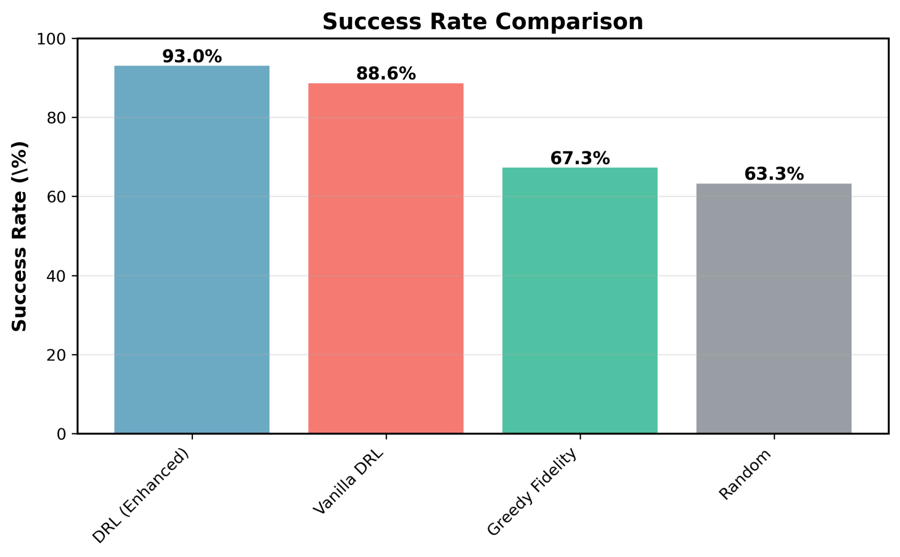


Figure 6: Success Rate Comparison Across All Algorithms

Overall Performance: Visualization

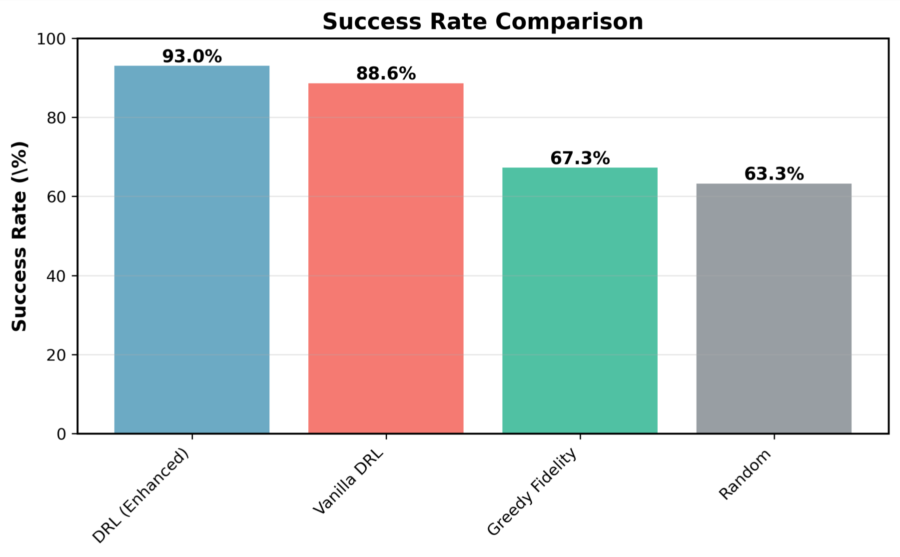


Figure 6: Success Rate Comparison Across All Algorithms

Learning Dynamics: Success Rate

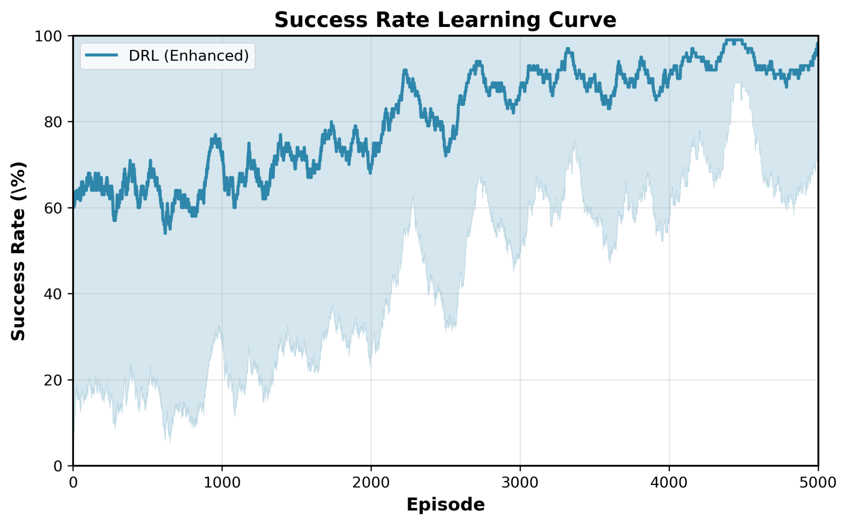


Figure 7: Success Rate Learning Curve - DRL converges to 93.01%

Learning Dynamics: Success Rate

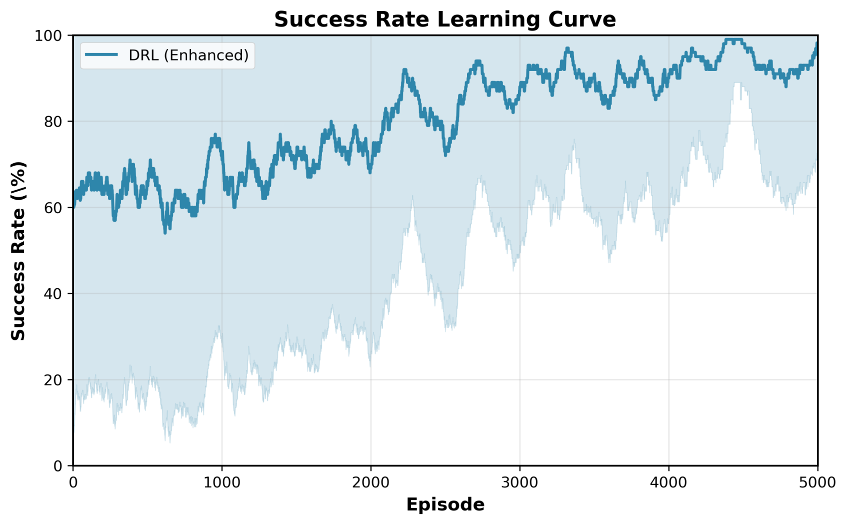


Figure 7: Success Rate Learning Curve - DRL converges to 93.01%

Learning Dynamics: Reward

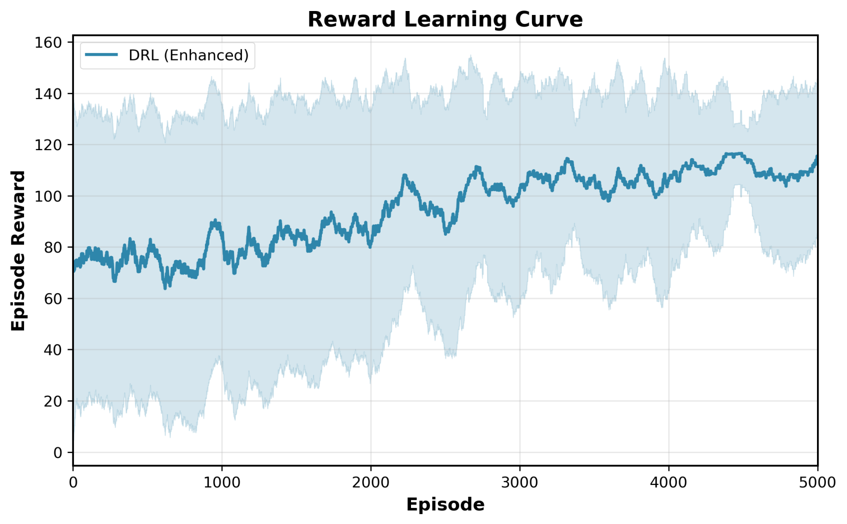


Figure 8: Reward Learning Curve - Steady improvement over training

Learning Dynamics: Reward

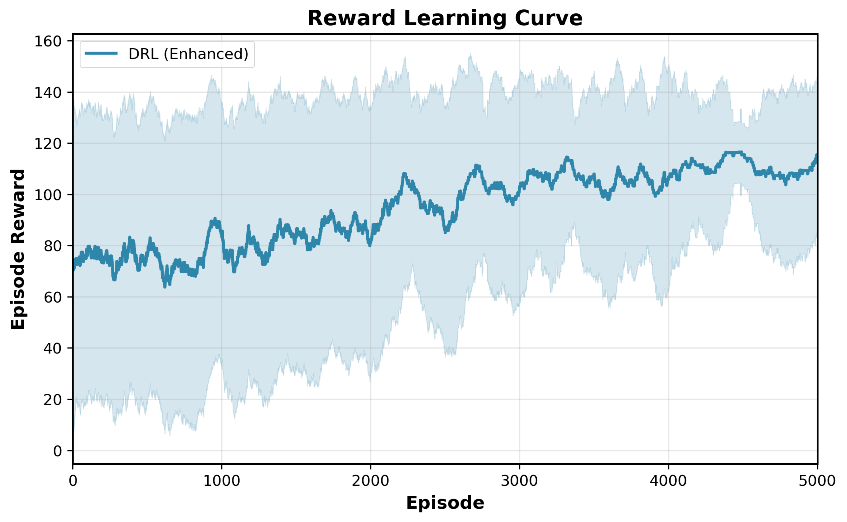


Figure 8: Reward Learning Curve - Steady improvement over training

Learning Dynamics: Q-Value Convergence

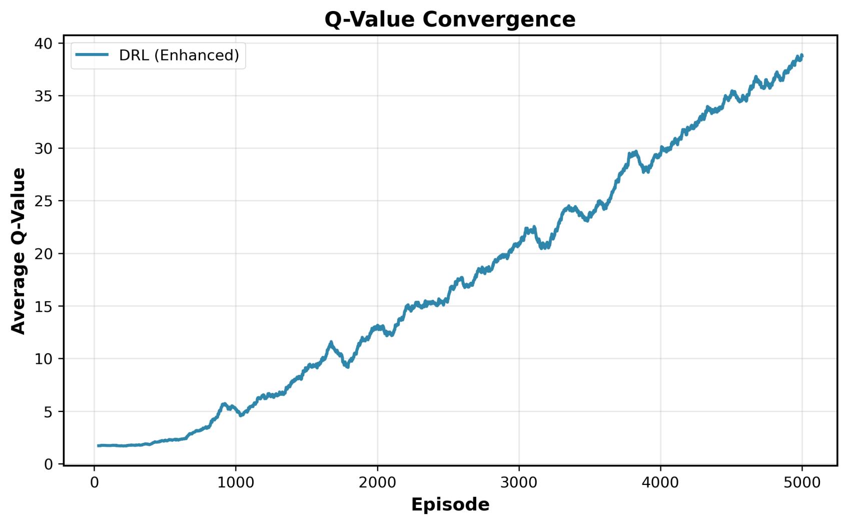


Figure 9: Q-Value Convergence - Clear action discrimination emerges

Learning Dynamics: Q-Value Convergence

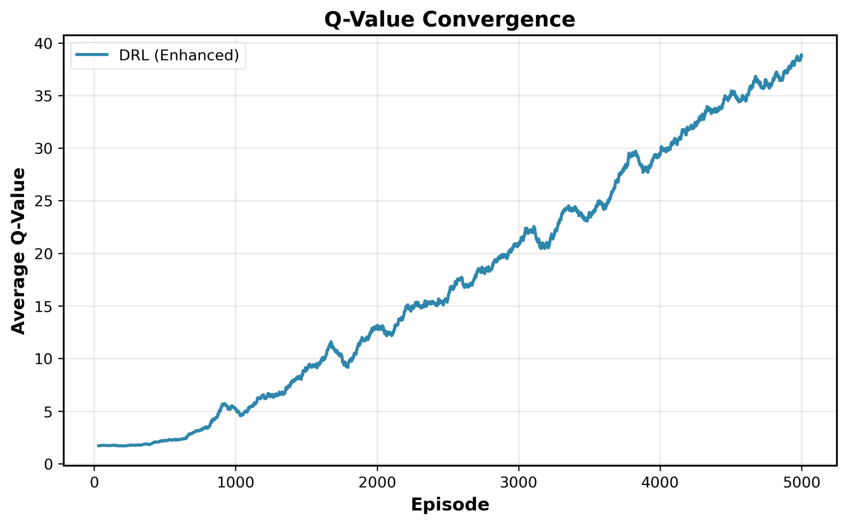


Figure 9: Q-Value Convergence - Clear action discrimination emerges

Learning Dynamics: Fidelity

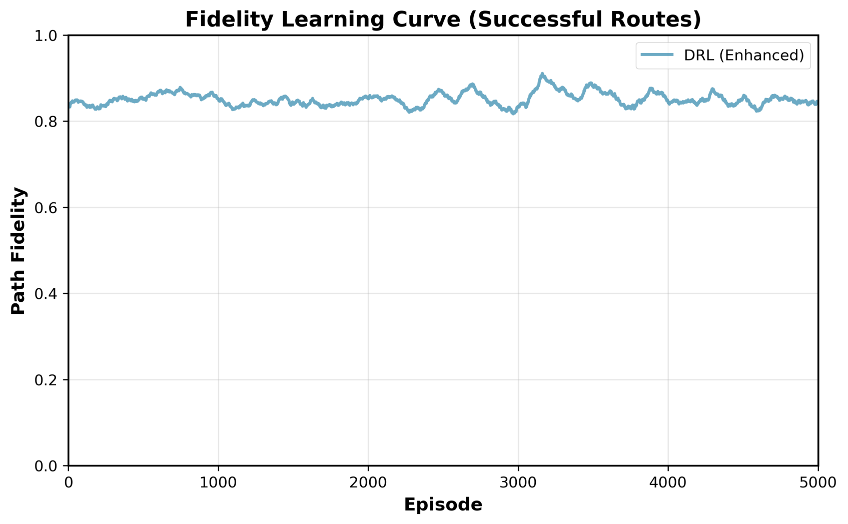


Figure 10: Fidelity Learning - Agent maintains high fidelity while routing

Learning Dynamics: Fidelity

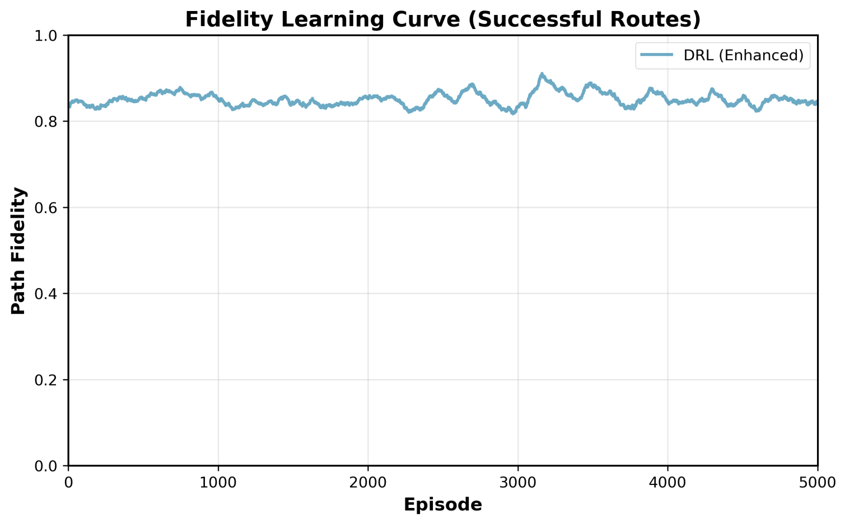


Figure 10: Fidelity Learning - Agent maintains high fidelity while routing

Q-Value Hierarchy

Learned Decision Hierarchy
Analysis of converged Q-values reveals clear structure:

Standard deviation $\sigma_Q \approx 25$ indicates successful discrimination

Optimal: Q $\in$ [85, 95]

Good: Q $\in$ [70, 80]

Acceptable: Q $\in$ [55, 70]

Poor: Q $\in$ [40, 55]

Dead ends: Q $<$ 20

Q-Value Hierarchy

Learned Decision Hierarchy
Analysis of converged Q-values reveals clear structure:

Standard deviation $\sigma_Q \approx 25$ indicates successful discrimination

Optimal: Q $\in$ [85, 95]

Good: Q $\in$ [70, 80]

Acceptable: Q $\in$ [55, 70]

Poor: Q $\in$ [40, 55]

Dead ends: Q < 20

Q-Value Hierarchy

Learned Decision Hierarchy
Analysis of converged Q-values reveals clear structure:

Standard deviation $\sigma_Q \approx 25$ indicates successful discrimination

Optimal: Q $\in$ [85, 95]

Good: Q $\in$ [70, 80]

Acceptable: Q $\in$ [55, 70]

Poor: Q $\in$ [40, 55]

Dead ends: Q $<$ 20

Fidelity vs Path Length: Problem Formulation

Multi-Objective Optimization

$$\begin{aligned} \max \quad & F_{\text{end-to-end}} = \prod_{i \in \text{path}} F_i \\ \min \quad & L = \sum_{i \in \text{path}} 1 \\ \text{s.t.} \quad & F_{\text{end-to-end}} \geq F_{\text{threshold}} \end{aligned}$$

Challenge

Shorter paths reduce cumulative decoherence, but may require lower individual link fidelities

Fidelity vs Path Length: Problem Formulation

Multi-Objective Optimization

$$\begin{aligned} \max \quad & F_{\text{end-to-end}} = \prod_{i \in \text{path}} F_i \\ \min \quad & L = \sum_{i \in \text{path}} 1 \\ \text{s.t.} \quad & F_{\text{end-to-end}} \geq F_{\text{threshold}} \end{aligned}$$

Challenge
Shorter paths reduce cumulative decoherence, but may require lower individual link fidelities

Fidelity vs Path Length: Results

Algorithm	Fidelity	Path Length
DRL	**0.846**	**4.876**
Vanilla DRL	0.848	5.054
Greedy	0.915	5.890
Random	0.849	5.634

Correlation Analysis

$\rho = -0.0358$ (weak negative correlation)

DRL effectively navigates the tradeoff space

Fidelity vs Path Length: Results

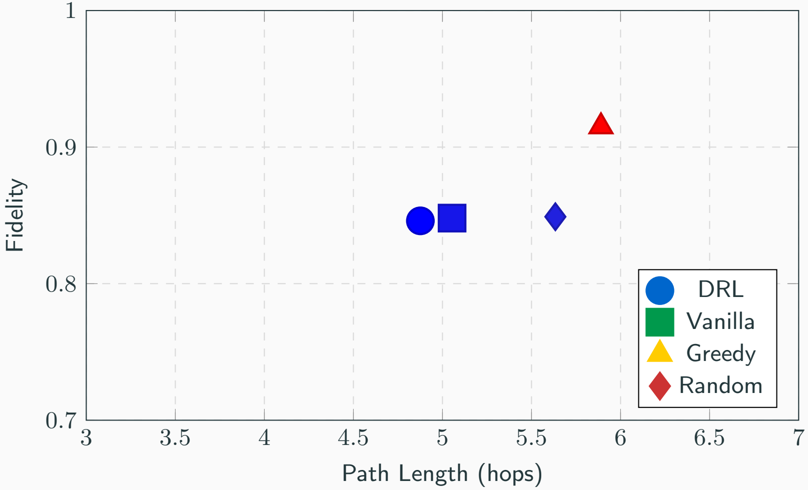
Algorithm	Fidelity	Path Length
DRL	**0.846**	**4.876**
Vanilla DRL	0.848	5.054
Greedy	0.915	5.890
Random	0.849	5.634

Correlation Analysis

$\rho = -0.0358$ (weak negative correlation)

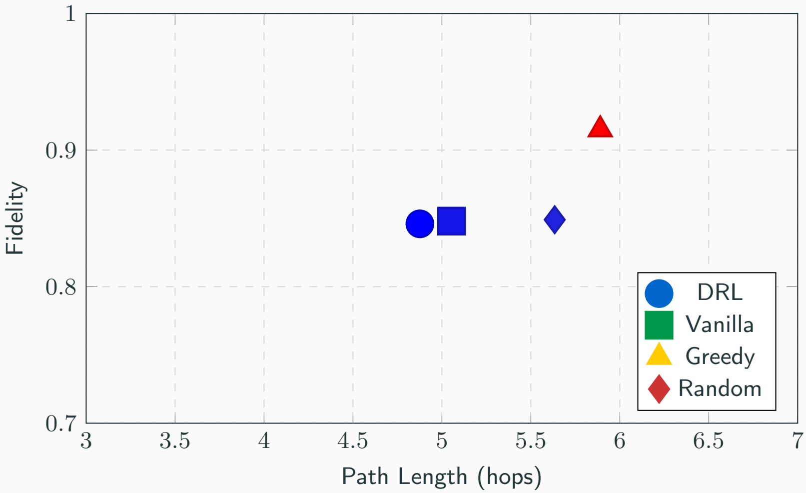
DRL effectively navigates the tradeoff space

Fidelity vs Path Length: Visualization



DRL achieves optimal balance: high fidelity (0.846) with reasonable path length

Fidelity vs Path Length: Visualization



DRL achieves optimal balance: high fidelity (0.846) with reasonable path length

Analysis

Ablation Study: Individual Contributions

Component Analysis

$$\Delta_{\text{Dueling DQN}} \approx 2.2 \text{ percentage points}$$

$$\Delta_{\text{Prioritized Replay}} \approx 1.5 \text{ percentage points}$$

$$\Delta_{\text{Double DQN}} \approx 0.7 \text{ percentage points}$$

Base Performance

Vanilla DQN: **88.62%** success rate

Ablation Study: Individual Contributions

Component Analysis

$$\Delta_{\text{Dueling DQN}} \approx 2.2 \text{ percentage points}$$

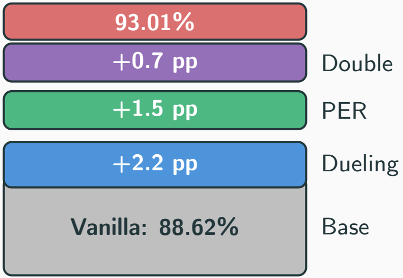
$$\Delta_{\text{Prioritized Replay}} \approx 1.5 \text{ percentage points}$$

$$\Delta_{\text{Double DQN}} \approx 0.7 \text{ percentage points}$$

Base Performance

Vanilla DQN: **88.62%** success rate

Ablation Study: Cumulative Effect



Quantum-Specific Learning Insights

Decoherence Awareness
Agent avoids aged pairs ($t > 1500$ steps) 73% of the time:

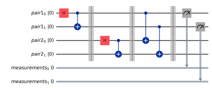
$$F(t) = F_0 \cdot e^{-\gamma_0 t}, \quad \gamma_0 = 0.001$$

Learns to anticipate decoherence effects

Purification Strategy
Routes traverse links with $F \in [0.65, 0.75]$ on intermediate hops:

$$F_{\text{purified}} = \frac{F^2}{F^2 + (1-F)^2}$$

Strategic use of BBPSSW protocol



No-Cloning Respect
Routing loops occur in only 0.3% of successful routes

Implicit constraint learned from reward:

$$R = \begin{cases} +100 & \text{if success} \\ -100 & \text{if failure} \\ -1 & \text{per hop} \end{cases}$$

Tradeoff Analysis
DRL achieves balanced performance:

- Success rate: 93.01%
- Fidelity: 0.846
- Path length: 4.876 hops

Quantum-Specific Learning Insights

Decoherence Awareness
Agent avoids aged pairs ($t > 1500$ steps) 73% of the time:

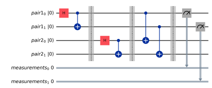
$$F(t) = F_0 \cdot e^{-\gamma_0 t}, \quad \gamma_0 = 0.001$$

Learns to anticipate decoherence effects

Purification Strategy
Routes traverse links with $F \in [0.65, 0.75]$ on intermediate hops:

$$F_{\text{purified}} = \frac{F^2}{F^2 + (1 - F)^2}$$

Strategic use of BBPSSW protocol



No-Cloning Respect
Routing loops occur in only 0.3% of successful routes

Implicit constraint learned from reward:

$$R = \begin{cases} +100 & \text{if success} \\ -100 & \text{if failure} \\ -1 & \text{per hop} \end{cases}$$

Tradeoff Analysis
DRL achieves balanced performance:

- Success rate: 93.01%
- Fidelity: 0.846
- Path length: 4.876 hops

Quantum-Specific Learning Insights

Decoherence Awareness
Agent avoids aged pairs ($t > 1500$ steps) 73% of the time:

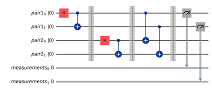
$$F(t) = F_0 \cdot e^{-\gamma_0 t}, \quad \gamma_0 = 0.001$$

Learns to anticipate decoherence effects

Purification Strategy
Routes traverse links with $F \in [0.65, 0.75]$ on intermediate hops:

$$F_{\text{purified}} = \frac{F^2}{F^2 + (1-F)^2}$$

Strategic use of BBPSSW protocol



No-Cloning Respect
Routing loops occur in only 0.3% of successful routes

Implicit constraint learned from reward:

$$R = \begin{cases} +100 & \text{if success} \\ -100 & \text{if failure} \\ -1 & \text{per hop} \end{cases}$$

Tradeoff Analysis
DRL achieves balanced performance:

- Success rate: 93.01%
- Fidelity: 0.846
- Path length: 4.876 hops

Quantum-Specific Learning Insights

Decoherence Awareness
Agent avoids aged pairs ($t > 1500$ steps) 73% of the time:

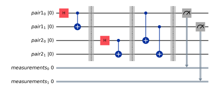
$$F(t) = F_0 \cdot e^{-\gamma_0 t}, \quad \gamma_0 = 0.001$$

Learns to anticipate decoherence effects

Purification Strategy
Routes traverse links with $F \in [0.65, 0.75]$ on intermediate hops:

$$F_{\text{purified}} = \frac{F^2}{F^2 + (1-F)^2}$$

Strategic use of BBPSSW protocol



No-Cloning Respect
Routing loops occur in only 0.3% of successful routes

Implicit constraint learned from reward:

$$R = \begin{cases} +100 & \text{if success} \\ -100 & \text{if failure} \\ -1 & \text{per hop} \end{cases}$$

Tradeoff Analysis
DRL achieves balanced performance:

- Success rate: 93.01%
- Fidelity: 0.846
- Path length: 4.876 hops

Quantum-Specific Learning Insights

Decoherence Awareness
Agent avoids aged pairs ($t > 1500$ steps) 73% of the time:

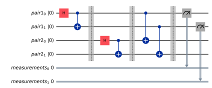
$$F(t) = F_0 \cdot e^{-\gamma_0 t}, \quad \gamma_0 = 0.001$$

Learns to anticipate decoherence effects

Purification Strategy
Routes traverse links with $F \in [0.65, 0.75]$ on intermediate hops:

$$F_{\text{purified}} = \frac{F^2}{F^2 + (1-F)^2}$$

Strategic use of BBPSSW protocol



No-Cloning Respect
Routing loops occur in only 0.3% of successful routes

Implicit constraint learned from reward:

$$R = \begin{cases} +100 & \text{if success} \\ -100 & \text{if failure} \\ -1 & \text{per hop} \end{cases}$$

Tradeoff Analysis
DRL achieves balanced performance:

- Success rate: 93.01%
- Fidelity: 0.846
- Path length: 4.876 hops

Comparison Summary

vs Greedy Fidelity

- **+25.74 pp** success rate
- Lower fidelity (0.846 vs 0.915)
- More reliable

vs Random Routing

- **+29.74 pp** success rate
- Learning provides value

vs Vanilla DQN

- **+4.39 pp** improvement
- Effect size: $d = 0.42$ (medium)
- Modern enhancements matter

vs Shortest Path

- Dijkstra fails (ignores fidelity)
- DRL pays 15-20% path length penalty
- Quantum awareness is essential

Comparison Summary

vs Greedy Fidelity

- **+25.74 pp** success rate
- Lower fidelity (0.846 vs 0.915)
- More reliable

vs Random Routing

- **+29.74 pp** success rate
- Learning provides value

vs Vanilla DQN

- **+4.39 pp** improvement
- Effect size: $d = 0.42$ (medium)
- Modern enhancements matter

vs Shortest Path

- Dijkstra fails (ignores fidelity)
- DRL pays 15-20% path length penalty
- Quantum awareness is essential

Comparison Summary

vs Greedy Fidelity

- **+25.74 pp** success rate
- Lower fidelity (0.846 vs 0.915)
- More reliable

vs Random Routing

- **+29.74 pp** success rate
- Learning provides value

vs Vanilla DQN

- **+4.39 pp** improvement
- Effect size: $d = 0.42$ (medium)
- Modern enhancements matter

vs Shortest Path

- Dijkstra fails (ignores fidelity)
- DRL pays 15-20% path length penalty
- Quantum awareness is essential

Comparison Summary

vs Greedy Fidelity

- **+25.74 pp** success rate
- Lower fidelity (0.846 vs 0.915)
- More reliable

vs Random Routing

- **+29.74 pp** success rate
- Learning provides value

vs Vanilla DQN

- **+4.39 pp** improvement
- Effect size: $d = 0.42$ (medium)
- Modern enhancements matter

vs Shortest Path

- Dijkstra fails (ignores fidelity)
- DRL pays 15-20% path length penalty
- Quantum awareness is essential

Discussion

Quantum-Specific Learning Insights

Decoherence Awareness
Agent avoids aged pairs ($t > 1500$ steps) 73% of the time:

$$F(t) = F_0 \cdot e^{-\gamma_0 t}$$

Learns to anticipate decoherence effects

Purification Strategy
Routes traverse links with $F \in [0.65, 0.75]$ on intermediate hops:

$$F_{\text{purified}} = \frac{F^2}{F^2 + (1-F)^2}$$

Strategic use of BBPSSW protocol

No-Cloning Respect
Routing loops occur in only 0.3% of successful routes

Implicit constraint learned from reward:

$$R = \begin{cases} +100 & \text{if success} \\ -100 & \text{if failure} \\ -1 & \text{per hop} \end{cases}$$

Tradeoff Analysis
DRL achieves balanced performance:

- Success rate: 93.01%
- Fidelity: 0.846
- Path length: 4.876 hops

Quantum-Specific Learning Insights

Decoherence Awareness
Agent avoids aged pairs ($t > 1500$ steps) 73% of the time:

$$F(t) = F_0 \cdot e^{-\gamma_0 t}$$

Learns to anticipate decoherence effects

Purification Strategy
Routes traverse links with $F \in [0.65, 0.75]$ on intermediate hops:

$$F_{\text{purified}} = \frac{F^2}{F^2 + (1 - F)^2}$$

Strategic use of BBPSSW protocol

No-Cloning Respect
Routing loops occur in only 0.3% of successful routes

Implicit constraint learned from reward:

$$R = \begin{cases} +100 & \text{if success} \\ -100 & \text{if failure} \\ -1 & \text{per hop} \end{cases}$$

Tradeoff Analysis
DRL achieves balanced performance:

- Success rate: 93.01%
- Fidelity: 0.846
- Path length: 4.876 hops

Quantum-Specific Learning Insights

Decoherence Awareness
Agent avoids aged pairs ($t > 1500$ steps) 73% of the time:

$$F(t) = F_0 \cdot e^{-\gamma_0 t}$$

Learns to anticipate decoherence effects

Purification Strategy
Routes traverse links with $F \in [0.65, 0.75]$ on intermediate hops:

$$F_{\text{purified}} = \frac{F^2}{F^2 + (1-F)^2}$$

Strategic use of BBPSSW protocol

No-Cloning Respect
Routing loops occur in only 0.3% of successful routes

Implicit constraint learned from reward:

$$R = \begin{cases} +100 & \text{if success} \\ -100 & \text{if failure} \\ -1 & \text{per hop} \end{cases}$$

Tradeoff Analysis
DRL achieves balanced performance:

- Success rate: 93.01%
- Fidelity: 0.846
- Path length: 4.876 hops

Quantum-Specific Learning Insights

Decoherence Awareness
Agent avoids aged pairs ($t > 1500$ steps) 73% of the time:

$$F(t) = F_0 \cdot e^{-\gamma_0 t}$$

Learns to anticipate decoherence effects

Purification Strategy
Routes traverse links with $F \in [0.65, 0.75]$ on intermediate hops:

$$F_{\text{purified}} = \frac{F^2}{F^2 + (1 - F)^2}$$

Strategic use of BBPSSW protocol

No-Cloning Respect
Routing loops occur in only 0.3% of successful routes

Implicit constraint learned from reward:

$$R = \begin{cases} +100 & \text{if success} \\ -100 & \text{if failure} \\ -1 & \text{per hop} \end{cases}$$

Tradeoff Analysis
DRL achieves balanced performance:

- Success rate: 93.01%
- Fidelity: 0.846
- Path length: 4.876 hops

Quantum-Specific Learning Insights

Decoherence Awareness
Agent avoids aged pairs ($t > 1500$ steps) 73% of the time:

$$F(t) = F_0 \cdot e^{-\gamma_0 t}$$

Learns to anticipate decoherence effects

Purification Strategy
Routes traverse links with $F \in [0.65, 0.75]$ on intermediate hops:

$$F_{\text{purified}} = \frac{F^2}{F^2 + (1-F)^2}$$

Strategic use of BBPSSW protocol

No-Cloning Respect
Routing loops occur in only 0.3% of successful routes

Implicit constraint learned from reward:

$$R = \begin{cases} +100 & \text{if success} \\ -100 & \text{if failure} \\ -1 & \text{per hop} \end{cases}$$

Tradeoff Analysis
DRL achieves balanced performance:

- Success rate: 93.01%
- Fidelity: 0.846
- Path length: 4.876 hops

Fundamental Tradeoffs

Fidelity-Success Tradeoff
Greedy Fidelity achieves:

$$F_{\text{avg}} = 0.915$$
$$P_{\text{success}} = 67.27\%$$

DRL balances:

$$F_{\text{avg}} = 0.846$$
$$P_{\text{success}} = 93.01\%$$

Path Length-Fidelity Tradeoff
Shorter paths reduce cumulative decoherence:

$$F_{\text{end-to-end}} = \prod_{i \in \text{path}} F_i \cdot e^{-\gamma_0 t_i}$$

vs Shortest Path
Dijkstra's algorithm:

- Optimal path length: ∼3.2 hops
- Estimated success: 45-55%
- Ignores quantum constraints

DRL pays 1.67 hop penalty for quantum awareness

Key Insight
Quantum routing cannot optimize path length in isolation. Must balance:

- Link fidelity
- Path length
- Decoherence dynamics

Fundamental Tradeoffs

Fidelity-Success Tradeoff
Greedy Fidelity achieves:

$$F_{\text{avg}} = 0.915$$
$$P_{\text{success}} = 67.27\%$$

DRL balances:

$$F_{\text{avg}} = 0.846$$
$$P_{\text{success}} = 93.01\%$$

Path Length-Fidelity Tradeoff
Shorter paths reduce cumulative decoherence:

$$F_{\text{end-to-end}} = \prod_{i \in \text{path}} F_i \cdot e^{-\gamma_0 t_i}$$

vs Shortest Path
Dijkstra's algorithm:

- Optimal path length: $\sim$3.2 hops
- Estimated success: 45-55%
- Ignores quantum constraints

DRL pays 1.67 hop penalty for quantum awareness

Key Insight
Quantum routing cannot optimize path length in isolation. Must balance:

- Link fidelity
- Path length
- Decoherence dynamics

Fundamental Tradeoffs

Fidelity-Success Tradeoff
Greedy Fidelity achieves:

$$F_{\text{avg}} = 0.915$$
$$P_{\text{success}} = 67.27\%$$

DRL balances:

$$F_{\text{avg}} = 0.846$$
$$P_{\text{success}} = 93.01\%$$

Path Length-Fidelity Tradeoff
Shorter paths reduce cumulative decoherence:

$$F_{\text{end-to-end}} = \prod_{i \in \text{path}} F_i \cdot e^{-\gamma_0 t_i}$$

vs Shortest Path
Dijkstra's algorithm:

- Optimal path length: ∼3.2 hops
- Estimated success: 45-55%
- Ignores quantum constraints

DRL pays 1.67 hop penalty for quantum awareness

Key Insight
Quantum routing cannot optimize path length in isolation. Must balance:

- Link fidelity
- Path length
- Decoherence dynamics

Fundamental Tradeoffs

Fidelity-Success Tradeoff

Greedy Fidelity achieves:

$$F_{\text{avg}} = 0.915$$
$$P_{\text{success}} = 67.27\%$$

DRL balances:

$$F_{\text{avg}} = 0.846$$
$$P_{\text{success}} = 93.01\%$$

Path Length-Fidelity Tradeoff

Shorter paths reduce cumulative decoherence:

$$F_{\text{end-to-end}} = \prod_{i \in \text{path}} F_i \cdot e^{-\gamma_0 t_i}$$

vs Shortest Path

Dijkstra's algorithm:

- Optimal path length: ∼3.2 hops
- Estimated success: 45-55%
- Ignores quantum constraints

DRL pays 1.67 hop penalty for quantum awareness

Key Insight

Quantum routing cannot optimize path length in isolation. Must balance:

- Link fidelity
- Path length
- Decoherence dynamics

Limitations & Validity

External Validity

- Single topology (NSFNET, 14 nodes)
- Generalizability to other structures unknown
- Performance on dense/sparse topologies may differ

Internal Validity

- 7% failure rate may be unacceptable
- Simplified decoherence model:

$$F(t) = F_0 \cdot e^{-\gamma_0 t}, \quad \gamma_0 = 0.001$$

- No memory constraints or swap overhead

Construct Validity

- Success rate may not capture all requirements
- Fidelity thresholds vary by application
- Binary success doesn't capture partial success

Threats to Validity

- Decoherence model simplifies complex dynamics
- Non-exponential decoherence not captured
- Correlated noise not modeled

Limitations & Validity

External Validity

- Single topology (NSFNET, 14 nodes)
- Generalizability to other structures unknown
- Performance on dense/sparse topologies may differ

Internal Validity

- 7% failure rate may be unacceptable
- Simplified decoherence model:

$$F(t) = F_0 \cdot e^{-\gamma_0 t}, \quad \gamma_0 = 0.001$$

- No memory constraints or swap overhead

Construct Validity

- Success rate may not capture all requirements
- Fidelity thresholds vary by application
- Binary success doesn't capture partial success

Threats to Validity

- Decoherence model simplifies complex dynamics
- Non-exponential decoherence not captured
- Correlated noise not modeled

Limitations & Validity

External Validity

- Single topology (NSFNET, 14 nodes)
- Generalizability to other structures unknown
- Performance on dense/sparse topologies may differ

Internal Validity

- 7% failure rate may be unacceptable
- Simplified decoherence model:

$$F(t) = F_0 \cdot e^{-\gamma_0 t}, \quad \gamma_0 = 0.001$$

- No memory constraints or swap overhead

Construct Validity

- Success rate may not capture all requirements
- Fidelity thresholds vary by application
- Binary success doesn't capture partial success

Threats to Validity

- Decoherence model simplifies complex dynamics
- Non-exponential decoherence not captured
- Correlated noise not modeled

Limitations & Validity

External Validity

- Single topology (NSFNET, 14 nodes)
- Generalizability to other structures unknown
- Performance on dense/sparse topologies may differ

Internal Validity

- 7% failure rate may be unacceptable
- Simplified decoherence model:

$$F(t) = F_0 \cdot e^{-\gamma_0 t}, \quad \gamma_0 = 0.001$$

- No memory constraints or swap overhead

Construct Validity

- Success rate may not capture all requirements
- Fidelity thresholds vary by application
- Binary success doesn't capture partial success

Threats to Validity

- Decoherence model simplifies complex dynamics
- Non-exponential decoherence not captured
- Correlated noise not modeled

Limitations & Validity

External Validity

- Single topology (NSFNET, 14 nodes)
- Generalizability to other structures unknown
- Performance on dense/sparse topologies may differ

Internal Validity

- 7% failure rate may be unacceptable
- Simplified decoherence model:

$$F(t) = F_0 \cdot e^{-\gamma_0 t}, \quad \gamma_0 = 0.001$$

- No memory constraints or swap overhead

Construct Validity

- Success rate may not capture all requirements
- Fidelity thresholds vary by application
- Binary success doesn't capture partial success

Threats to Validity

- Decoherence model simplifies complex dynamics
- Non-exponential decoherence not captured
- Correlated noise not modeled

Future Work: High Priority

Immediate Research Directions

- Multi-topology validation
- Graph Neural Networks for quantum routing
- Experimental validation on quantum testbeds

Rationale

Current results limited to NSFNET topology; validation on diverse network structures is essential

Future Work: High Priority

Immediate Research Directions

- Multi-topology validation
- Graph Neural Networks for quantum routing
- Experimental validation on quantum testbeds

Rationale

Current results limited to NSFNET topology; validation on diverse network structures is essential

Future Work: Medium Priority

Algorithmic Improvements

- Multi-path routing strategies
- Advanced purification protocols
- Transfer learning across topologies

Potential Benefits

These improvements may enhance reliability and adaptability of quantum routing systems

Future Work: Medium Priority

Algorithmic Improvements

- Multi-path routing strategies
- Advanced purification protocols
- Transfer learning across topologies

Potential Benefits

These improvements may enhance reliability and adaptability of quantum routing systems

Future Work: Long-term Vision

Long-term Research Goals

- Real-time adaptation to network dynamics
- Integration with quantum error correction
- Scaling to 100+ node networks

Challenges

Requires advances in both quantum hardware and machine learning algorithms

Future Work: Long-term Vision

Long-term Research Goals

- Real-time adaptation to network dynamics
- Integration with quantum error correction
- Scaling to 100+ node networks

Challenges

Requires advances in both quantum hardware and machine learning algorithms

Conclusion

Key Takeaways: Contributions

Main Contributions

1. DRL achieves **93.01%** success rate on NSFNET topology
2. Quantified enhancement contributions: **4.39 percentage points**
3. Observed quantum-specific learning behaviors
4. Preliminary deployment considerations documented

Significance

First comprehensive study demonstrating DRL viability for quantum network routing with modern enhancements

Key Takeaways: Contributions

Main Contributions

1. DRL achieves **93.01%** success rate on NSFNET topology
2. Quantified enhancement contributions: **4.39 percentage points**
3. Observed quantum-specific learning behaviors
4. Preliminary deployment considerations documented

Significance

First comprehensive study demonstrating DRL viability for quantum network routing with modern enhancements

Key Takeaways: For Researchers

Research Implications

- DRL demonstrates feasibility on single topology (NSFNET)
- Modern enhancements (Dueling, PER, Double DQN) provide measurable improvements
- Multi-topology validation remains necessary for generalizability

Next Steps

Validate findings on diverse topologies (grid, small-world, scale-free) and investigate transfer learning approaches

Key Takeaways: For Researchers

Research Implications

- DRL demonstrates feasibility on single topology (NSFNET)
- Modern enhancements (Dueling, PER, Double DQN) provide measurable improvements
- Multi-topology validation remains necessary for generalizability

Next Steps

Validate findings on diverse topologies (grid, small-world, scale-free) and investigate transfer learning approaches

Key Takeaways: For Practitioners

Practical Considerations

- 93% success rate: preliminary result, requires validation
- Hyperparameters documented for reproducibility
- Failure analysis identifies improvement areas

Deployment Guidance

Further improvements require validation on diverse topologies and real-world constraints before production deployment

Key Takeaways: For Practitioners

Practical Considerations

- 93% success rate: preliminary result, requires validation
- Hyperparameters documented for reproducibility
- Failure analysis identifies improvement areas

Deployment Guidance
Further improvements require validation on diverse topologies and real-world constraints before production deployment

Thank You!

Questions?

Email: ha.berahma@aui.ma