Quantum-Assisted Reinforcement Learning for Adaptive Intelligent Systems

Junedul Haque
School of Computer Science and Engineering
Sandip University
Nashik, India

***Abstract*-Quantum computing and reinforcement learning (RL) represent two of the most transformative paradigms in modern computing and artificial intelligence. This review paper surveys the emerging field of quantum-assisted reinforcement learning (QARL), examining how quantum mechanical phenomena such as superposition, entanglement, and quantum parallelism can be leveraged to enhance the efficiency, scalability, and adaptability of RL algorithms. We discuss the theoretical underpinnings of quantum RL, survey landmark algorithms including variational quantum circuits for policy optimization and quantum approximate optimization algorithms (QAOA) applied to decision-making tasks and examine practical implementations on near-term noisy intermediate-scale quantum (NISQ) devices. The review further addresses key challenges including quantum decoherence, barren plateaus in quantum training landscapes, and hardware limitations, while outlining promising future research directions. Our analysis suggests that QARL holds significant potential for achieving quantum advantage in complex adaptive intelligent systems, particularly in combinatorial optimization and high-dimensional state-space problems.**

Keywords: quantum reinforcement learning, variational quantum circuits, NISQ devices, quantum advantage, adaptive intelligent systems, policy optimization

1. INTRODUCTION

The confluence of quantum computing and machine learning has attracted considerable research interest over the past decade, driven by the promise of exponential speedups for specific computational tasks. Reinforcement learning, a branch of machine learning in which an agent learns optimal behavior through interaction with an environment, has achieved remarkable successes in domains ranging from game playing [1] to robotic control [2] and drug discovery [3]. However, classical RL algorithms face fundamental scalability challenges when confronted with high-dimensional state and action spaces, often requiring prohibitive computational resources.

Quantum computing offers a fundamentally different computational paradigm, exploiting quantum mechanical effects to perform calculations that are intractable for classical computers. The demonstration of quantum supremacy by Google in 2019, wherein a quantum processor solved a sampling problem in 200 seconds that would take a classical supercomputer approximately 10,000 years [4], catalyzed renewed interest in practical quantum applications. Within this context, quantum-assisted reinforcement learning has emerged as a promising hybrid framework that combines parameterized quantum circuits with classical optimization routines to enhance RL performance.

This review is structured as follows. Section II provides background on classical RL and quantum computing fundamentals. Section III surveys core QARL algorithms and theoretical results. Section IV examines hardware implementations and NISQ-era considerations. Section V discusses open challenges, and Section VI identifies future research directions before the paper concludes in Section VII.

2. BACKGROUND

A. Classical Reinforcement Learning

Reinforcement learning is formalized through the Markov Decision Process (MDP) framework, defined by the tuple (S, A, P, R, γ), where S denotes the state space, A the action space, $P: S \times A \times S \to [0,1]$ the transition probability function, $R: S \times A \to \mathbb{R}$ the reward function, and $\gamma \in [0,1]$ the discount factor [5]. The agent's goal is to find a policy $\pi: S \to A$ that maximizes the expected cumulative discounted return. Classical RL algorithms are broadly categorized into model-based methods, value-based methods such as Q-learning and deep Q-networks (DQN) [1], and policy gradient methods including REINFORCE [6] and Proximal Policy Optimization (PPO) [7].

Despite their successes, classical RL algorithms face the "curse of dimensionality," wherein the computational and memory requirements scale exponentially with the dimension of the state-action space. Deep RL alleviates some of these issues through function approximation using neural networks, but training can be sample-inefficient and prone to instability [8].

B. Quantum Computing Fundamentals

A quantum computer operates on quantum bits (qubits), which unlike classical bits can exist in a superposition of states $|0\rangle$ and $|1\rangle$, represented as $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ where $|\alpha|^2 + |\beta|^2 = 1$. Quantum entanglement allows qubits to exhibit correlations with no classical analog, enabling information to be encoded exponentially more compactly than with classical bits [9]. Quantum gates manipulate qubit states through unitary transformations, analogous to logic gates in classical computing.

Variational quantum circuits (VQCs), also called parameterized quantum circuits (PQCs), are hybrid quantum-classical structures in which a quantum circuit with tunable parameters is optimized using classical gradient-based methods [10]. VQCs are particularly well-suited to NISQ devices, which have limited qubit counts (50–1000 qubits) and are susceptible to noise and decoherence, but are available today through platforms such as IBM Quantum, Google Sycamore, and IonQ [11].

3. QUANTUM-ASSISTED REINFORCEMENT LEARNING ALGORITHMS

A. Quantum Q-Learning

One of the earliest approaches to QARL replaces the classical neural network in DQN with a variational quantum circuit to approximate the Q-function. Chen et al. [12] proposed a quantum deep Q-network (QDQN) in which a PQC encodes state information through angle encoding and outputs Q-values via expectation measurements of Pauli observables. Empirical evaluations on OpenAI Gym environments demonstrated competitive performance with classical DQN using significantly fewer trainable parameters, suggesting a potential representational advantage. Subsequent work by Lockwood and Si [13] extended QDQN to more complex environments and analyzed the role of entanglement depth in representation capacity.

The theoretical basis for quantum Q-learning speedups draws from quantum amplitude amplification, a generalization of Grover's algorithm that can yield quadratic speedups in search-based RL subroutines [14]. However, realizing these speedups in practice requires quantum random access memory (QRAM), a technology that remains largely theoretical at meaningful scales.

B. Variational Quantum Policy Gradients

Policy gradient methods have been extended to the quantum domain by parameterizing the policy as a PQC. Jerbi et al. [15] formalized quantum policy gradient (QPG) algorithms and established that quantum policies can represent certain probability distributions exponentially more efficiently than classical neural networks. Their framework employs data re-uploading—a technique in which classical input data is encoded into the quantum circuit at multiple layers—to improve express ability. Schuld et al. [16] provided rigorous analysis of PQC express ability and entangling capability, introducing metrics that guide circuit architecture design for policy representation.

A critical challenge in quantum policy optimization is the barren plateau phenomenon, wherein the gradient of the loss function vanishes exponentially with increasing qubit count for randomly initialized circuits [17]. McClean et al. [17] formally characterized this problem, and subsequent research has proposed mitigation strategies including layer-wise training [18], identity block initialization [19], and problem-inspired ansatz design [20].

C. Quantum Actor-Critic Methods

Actor-critic architectures, which maintain separate function approximators for the policy (actor) and value function (critic), have been adapted to incorporate quantum components. Kwak et al. [21] proposed a hybrid quantum-classical actor-critic framework in which the actor is implemented as a PQC and the critic as a classical neural network. This asymmetric design leverages the potential representational advantages of quantum circuits for policy representation while relying on classical networks for the more stable value estimation task. Their experiments demonstrated faster convergence on continuous control tasks compared to classical Soft Actor-Critic (SAC) baselines.

D. QAOA for Sequential Decision-Making

The Quantum Approximate Optimization Algorithm (QAOA), introduced by Farhi et al. [22], is a variational algorithm designed for combinatorial optimization problems on graphs. Its application to RL is natural in settings where the action selection problem can be mapped to a combinatorial structure. Wauters et al. [23] applied QAOA to reinforcement learning of quantum control protocols, demonstrating that quantum-enhanced exploration strategies accelerate convergence in quantum system control tasks. Similarly, Fitzek et al. [24] showed that QAOA-based action selection in multi-armed bandit problems can achieve better exploration-exploitation trade-offs than classical ε-greedy strategies under specific problem structures.

E. Quantum Model-Based RL

Model-based RL, which explicitly learns a model of the environment dynamics, offers an avenue for quantum speedup through quantum simulation. Dunjko and Briegel [25] provided a theoretical framework demonstrating that quantum agents interacting with classical environments can achieve quadratic speedup in the number of training episodes required to learn optimal policies, provided the agent has access to coherent superpositions of environmental states. This "quantum speedup for learning agents" result is among the most rigorous theoretical contributions to the QARL literature, though its practical realization requires fault-tolerant quantum hardware beyond current capabilities.

4. HARDWARE IMPLEMENTATIONS AND NISQ CONSIDERATIONS

A. Near-Term Device Implementations

Current NISQ devices, with qubit counts ranging from tens to hundreds and gate fidelities of 99–99.9% for two-qubit gates, impose significant constraints on QARL implementations [11]. Typical implementations use 4–16 qubits with circuit depths of 10–50 layers to remain within coherence times. Skolik et al. [26] conducted systematic experiments on IBM Quantum hardware, finding that noise-induced gradient variance increases substantially with circuit depth, underscoring the need for shallow circuit architectures. Mitigation techniques including zero-noise extrapolation and probabilistic error cancellation have been applied to improve the reliability of gradient estimates on real hardware [27].

B. Quantum Simulation Platforms

Beyond superconducting qubit systems, trapped-ion platforms such as those developed by IonQ and Honeywell offer higher gate fidelities and all-to-all qubit connectivity, which can reduce circuit depth requirements. Photonic quantum systems, while not yet competitive in qubit count, offer unique advantages for certain continuous-variable RL formulations [28]. Neutral atom arrays, demonstrated by Lukin et al. [29] at scales exceeding 200 qubits, are emerging as a promising platform for near-term QARL experiments due to their long coherence times and reconfigurability.

C. Hybrid Classical-Quantum Architectures

Given current hardware limitations, the predominant implementation paradigm is hybrid quantum-classical, wherein a classical computer handles data preprocessing, loss computation, and parameter updates, while quantum hardware evaluates parameterized circuits. Frameworks such as PennyLane [30], Qiskit Machine Learning [31], and TensorFlow Quantum [32] provide software infrastructure for developing and benchmarking QARL algorithms on both simulators and real quantum hardware, significantly lowering the barrier to experimental research.

5. OPEN CHALLENGES

A. Barren Plateaus and Trainability

The barren plateau problem represents perhaps the most significant theoretical obstacle to practical QARL. As demonstrated by McClean et al. [17] and extended by Cerezo et al. [33], the variance of cost function gradients decreases exponentially in the number of qubits for hardware-efficient ansätze, making training infeasible at scale. While problem-inspired circuit designs, correlating ansatz structure to the problem symmetry, can partially mitigate this, a general solution remains elusive. Recent work by Larocca et al. [34] has provided a more complete characterization of barren plateaus in terms of the Lie algebras generated by the circuit's gate set, offering new insights for ansatz design.

B. Quantum Decoherence and Noise

Quantum decoherence, the loss of quantum information due to interaction with the environment, fundamentally limits the depth and reliability of quantum circuits on current hardware. For QARL applications, decoherence introduces systematic biases into gradient estimates that can destabilize training. Quantum error correction, which can in principle eliminate these effects, requires fault-tolerant hardware with physical qubit overhead of 100–1000 physical qubits per logical qubit [9], far exceeding current capabilities. Error mitigation techniques provide a pragmatic intermediate solution at the cost of additional circuit evaluations.

C. Classical Simulation Overhead

A subtle but important challenge is that many proposed QARL algorithms require exponential classical computational resources to simulate, making it difficult to rigorously benchmark quantum approaches against classical baselines at scale. Abbas et al. [35] highlighted this simulation barrier, noting that claims of quantum advantage in machine learning must be carefully scrutinized with respect to the full computational cost including data loading, circuit compilation, and measurement overhead.

D. Encoding and Readout Bottlenecks

Encoding classical data into quantum states (state preparation) and extracting useful information through measurements (readout) can incur overheads that negate quantum advantages. Efficient amplitude encoding requires exponentially many gates in the worst case, while measurement statistics require many circuit repetitions (shots) for reliable gradient estimates. These encoding and readout bottlenecks are critical practical constraints that must be addressed for QARL to achieve practical quantum advantage [36].

6. FUTURE RESEARCH DIRECTIONS

Several promising directions emerge from this review. First, the development of problem-structured ansätze that exploit the symmetries of specific RL environments represents a high-priority research direction, as such designs can circumvent barren plateaus while maintaining expressibility [20]. Second, the integration of QARL with quantum communication networks opens the possibility of quantum multi-agent reinforcement learning, wherein agents share quantum states to enable fundamentally new coordination protocols [37].

Third, the application of QARL to quantum control problems—optimizing protocols for state preparation, error correction, and quantum sensing—represents a natural domain where the quantum agent and environment share the same physical substrate [23]. Fourth, advances in fault-tolerant quantum computing will eventually enable access to quantum algorithms with provable exponential speedups, such as quantum linear system solvers [38] applied to value function computation. Fifth, rigorous theoretical analysis of the sample complexity of QARL algorithms, analogous to classical RL theory, is needed to establish when and why quantum policies provide advantages [15].

7. CONCLUSION

This review has surveyed the rapidly evolving field of quantum-assisted reinforcement learning, encompassing theoretical foundations, algorithmic developments, hardware implementations, and open challenges. QARL represents a compelling synthesis of two transformative computational paradigms, with the potential to unlock new capabilities for adaptive intelligent systems in domains where classical RL faces fundamental scalability limitations.

The field has progressed significantly since its inception, with rigorous theoretical results establishing conditions for quantum speedups, and experimental demonstrations on NISQ hardware providing proof-of-concept validation. However, substantial challenges remain, particularly around trainability, noise, and the classical simulation barrier. The path to practical quantum advantage in RL will require coordinated advances in quantum hardware, error mitigation, algorithm design, and theoretical understanding.

As quantum hardware continues to mature and the theoretical foundations of QARL deepen, we anticipate that quantum-assisted approaches will become an important component of the AI toolkit, particularly for applications in quantum chemistry, materials design, logistics, and finance, where the native quantum structure of the problem aligns naturally with quantum computational primitives. The coming decade promises to be a pivotal period for determining the practical scope and limits of quantum advantage in reinforcement learning.

REFERENCES

[1] V. Mnih et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, no. 7540, pp. 529–533, Feb. 2015. doi: 10.1038/nature14236.

[2] T. P. Lillicrap et al., "Continuous control with deep reinforcement learning," in Proc. 4th Int. Conf. Learn. Represent. (ICLR), San Juan, Puerto Rico, 2016.

[3] A. Zhavoronkov et al., "Deep learning enables rapid identification of potent DDR1 kinase inhibitors," Nature Biotechnology, vol. 37, no. 9, pp. 1038–1040, Sep. 2019. doi: 10.1038/s41587-019-0224-x.

[4] F. Arute et al., "Quantum supremacy using a programmable superconducting processor," Nature, vol. 574, no. 7779, pp. 505–510, Oct. 2019. doi: 10.1038/s41586-019-1666-5.

[5] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[6] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," Machine Learning, vol. 8, no. 3–4, pp. 229–256, May 1992. doi: 10.1007/BF00992696.

[7] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv:1707.06347, Jul. 2017.

[8] M. Fortunato et al., "Noisy networks for exploration," in Proc. 6th Int. Conf. Learn. Represent. (ICLR), Vancouver, BC, Canada, 2018.

[9] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information, 10th anniversary ed. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[10] M. Cerezo et al., "Variational quantum algorithms," Nature Reviews Physics, vol. 3, no. 9, pp. 625–644, Aug. 2021. doi: 10.1038/s42254-021-00348-9.

[11] J. Preskill, "Quantum computing in the NISQ era and beyond," Quantum, vol. 2, p. 79, Aug. 2018. doi: 10.22331/q-2018-08-06-79.

[12] S. Y.-C. Chen, C.-H. H. Yang, J. Qi, P.-Y. Chen, X. Ma, and H.-S. Goan, "Variational quantum circuits for deep reinforcement learning," IEEE Access, vol. 8, pp. 141007–141024, 2020. doi: 10.1109/ACCESS.2020.3010470.

[13] O. Lockwood and M. Si, "Reinforcement learning with quantum variational circuits," in Proc. AAAI Conf. Artif. Intell. Interactive Digital Entertain., Oct. 2020.

[14] G. Brassard, P. Høyer, M. Mosca, and A. Tapp, "Quantum amplitude amplification and estimation," in Quantum Computation and Quantum Information, AMS Contemporary Mathematics, vol. 305, Providence, RI, USA: AMS, 2002, pp. 53–74.

[15] S. Jerbi, L. J. Fiderer, H. Poulsen Nautrup, J. M. Kübler, H. J. Briegel, and V. Dunjko, "Quantum machine learning beyond kernel methods," Nature Communications, vol. 14, no. 1, p. 517, Jan. 2023. doi: 10.1038/s41467-023-36159-y.

[16] M. Schuld, R. Sweke, and J. J. Meyer, "Effect of data encoding on the expressive power of variational quantum-machine-learning models," Physical Review A, vol. 103, no. 3, p. 032430, Mar. 2021. doi: 10.1103/PhysRevA.103.032430.

[17] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes," Nature Communications, vol. 9, no. 1, p. 4812, Nov. 2018. doi: 10.1038/s41467-018-07090-4.

[18] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, "An initialization strategy for addressing barren plateaus in parametrized quantum circuits," Quantum, vol. 3, p. 214, Dec. 2019. doi: 10.22331/q-2019-12-09-214.

[19] T. L. Patti, K. Najafi, X. Gao, and S. F. Yelin, "Entanglement devised barren plateau mitigation," Physical Review Research, vol. 3, no. 3, p. 033090, Aug. 2021. doi: 10.1103/PhysRevResearch.3.033090.

[20] M. Larocca, P. Czarnik, K. Sharma, G. Muraleedharan, P. J. Coles, and M. Cerezo, "Diagnosing barren plateaus with tools from quantum optimal control," Quantum, vol. 6, p. 824, Sep. 2022. doi: 10.22331/q-2022-09-29-824.

[21] Y. Kwak, W. J. Yun, S. Jung, and J. Kim, "Introduction to quantum reinforcement learning: theory and PennyLane-based implementation," in Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC), Jeju, South Korea, 2021, pp. 416–420. doi: 10.1109/ICTC52510.2021.9620885.

[22] E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm," arXiv:1411.4028, Nov. 2014.

[23] M. M. Wauters, E. Panizon, G. B. Mbeng, and G. E. Santoro, "Reinforcement-learning-assisted quantum optimization," Physical Review Research, vol. 2, no. 3, p. 033083, Jul. 2020. doi: 10.1103/PhysRevResearch.2.033083.

[24] D. Fitzek, M. Jonsson, A. Sinayskiy, I. Sinayskiy, and F. Petruccione, "Qreate: The quantum reinforcement learning tookit for gate sequences," arXiv:2003.07348, Mar. 2020.

[25] V. Dunjko and H. J. Briegel, "Machine learning and artificial intelligence in the quantum domain: a review of recent progress," Reports on Progress in Physics, vol. 81, no. 7, p. 074001, Jun. 2018. doi: 10.1088/1361-6633/aab406.

[26] A. Skolik, J. R. McClean, M. Mohseni, P. van der Smagt, and M. Leib, "Layerwise learning for quantum neural networks," Quantum Machine Intelligence, vol. 3, no. 5, 2021. doi: 10.1007/s42484-020-00036-4.

[27] K. Temme, S. Bravyi, and J. M. Gambetta, "Error mitigation for short-depth quantum circuits," Physical Review Letters, vol. 119, no. 18, p. 180509, Nov. 2017. doi: 10.1103/PhysRevLett.119.180509.

[28] N. Killoran et al., "Continuous-variable quantum neural networks," Physical Review Research, vol. 1, no. 3, p. 033063, Nov. 2019. doi: 10.1103/PhysRevResearch.1.033063.

[29] H. Bernien et al., "Probing many-body dynamics on a 51-atom quantum simulator," Nature, vol. 551, no. 7682, pp. 579–584, Nov. 2017. doi: 10.1038/nature24622.

[30] V. Bergholm et al., "PennyLane: Automatic differentiation of hybrid quantum-classical computations," arXiv:1811.04968, 2018.

[31] A. Javadi-Abhari et al., "Quantum computing with Qiskit," arXiv:2405.08810, May 2024.

[32] M. Broughton et al., "TensorFlow Quantum: A software framework for quantum machine learning," arXiv:2003.02989, Mar. 2020.

[33] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, "Cost function dependent barren plateaus in shallow parametrized quantum circuits," Nature Communications, vol. 12, no. 1, p. 1791, Mar. 2021. doi: 10.1038/s41467-021-21728-w.

[34] M. Larocca et al., "A review of barren plateaus in variational quantum computing," arXiv:2405.00781, May 2024.

[35] A. Abbas et al., "On quantum backpropagation, information reuse, and cheating," in Advances in Neural Information Processing Systems (NeurIPS), vol. 36, 2023.

[36] E. Tang, "A quantum-inspired classical algorithm for recommendation systems," in Proc. 51st Annu. ACM Symp. Theory Comput. (STOC), Phoenix, AZ, USA, Jun. 2019, pp. 217–228. doi: 10.1145/3313276.3316310.

[37] N. Kölle et al., "Quantum multi-agent reinforcement learning via variational quantum circuits," arXiv:2408.05885, Aug. 2024.

[38] A. W. Harrow, A. Hassidim, and S. Lloyd, "Quantum algorithm for linear systems of equations," Physical Review Letters, vol. 103, no. 15, p. 150502, Oct. 2009. doi: 10.1103/PhysRevLett.103.150502.