Adaptive Quantum Circuit Pruning with Reinforcement Learning

Muhammad Hamza, Muhammad Mustafa Umar Gondel, and Hyundong Shin
Department of Electronics and Information Convergence Engineering, Kyung Hee University, Korea
Email: hshin@khu.ac.kr

***Abstract*—Quantum machine learning is increasingly and readily being used for noisy intermediate-scale quantum (NISQ) networks. With such a fast development, the limitations of NISQ hardware present challenges in scalability and reliability. As the number of used gates increase, decoherence and output errors are increased with it. In this work, we present an adaptive optimization and pruning technique, based on reinforcement learning (RL), particularly the proximal policy optimization (PPO) algorithm, to optimize and prune quantum circuits by reducing gate count while preserving fidelity. Unlike fixed-threshold approaches, our RL agent learns pruning strategies directly from circuit features and performance feedback. After evaluating the model on different 3-qubits circuits, it achieves up to 38% gate reduction with fidelity above 88%, providing reliable results. Our framework generalizes seamlessly to circuit structures not encountered during training.**

I. Introduction

Quantum circuits form the core of all quantum algorithms, ranging from Grover's search to variational algorithms. A recent approach utilizes quantum witness machines to infer properties directly from circuit behaviors, showcasing the rapid growth in quantum computing fields [1], [2]. In the noisy intermediate-scale quantum (NISQ) era, quantum devices are constrained by short coherence times and error-prone gate operations. Minimizing circuit complexity while preserving fidelity is key to reducing errors [3]. Practical systems, including ion-trap architectures, face some additional issues such as motional heating and detuning errors [4]. Reinforcement learning (RL) offers an effective framework for quantum circuit optimization, allowing an agent to iteratively explore configurations and learn strategies that balance fidelity with resource efficiency. Recent work has demonstrated the potential of quantum deep reinforcement learning (QDRL) in complex environments, such as integrating digital twins to optimize and secure 6G networks [5], highlighting the broader applicability of quantum learning methods. Thus circuit pruning can eliminate gates that contribute minimally to the computation, thereby reducing complexity and mitigating decoherence effects.

II. Methodology

RL has been shown to optimize quantum circuits. In particular, ZX-calculus combined with graph neural networks within an RL framework has been used to reduce controlled-NOT (CNOT) gates [6].

Pruning in quantum circuits involves identifying and removing redundant or less significant gates to reduce circuit complexity while ensuring maximum fidelity. By eliminating unnecessary gates, the circuit becomes more efficient, which is crucial for near-term quantum devices with limited coherence times and gate fidelity. In our approach, reinforcement learning is employed to guide the pruning process where an agent learns to select which gates to remove by optimizing a reward function that balances gate reduction and fidelity preservation. Over time, the agent learns pruning strategies that maximize compression while maintaining the circuit's functional accuracy, enabling more efficient and scalable quantum computations. Pruning is modeled as a continuous-action task in a custom Gym environment. Each state encodes circuit-level features such as gate distribution (Pauli-X, Y, Z, Hadamard, R_x, R_y, R_z), depth, gate count, fidelity, and reduction ratio. The action is a threshold $\theta \in [0, \pi]$ which is used to prune rotation gates based on their angles.

The reward encourages fidelity preservation and gate reduction:

$$R_t = 2F_t + 1.5R_{g,t} + b_t - p_t \tag{1}$$

where R_t is the reward, F_t is the fidelity, $R_{g,t}$ is the gate reduction ratio, b_t is the bonus, and p_t is the penalty, all at time step t.

Fidelity is measured as the modulus squared of the inner product of the original and pruned statevectors:

$$F = |\langle \psi_{\text{ideal}} | \psi_{\text{pruned}} \rangle|^2 \tag{2}$$

where $|\psi_{\text{ideal}}\rangle$ and $|\psi_{\text{pruned}}\rangle$ are the statevectors of the original and pruned circuits.

A proximal policy optimization (PPO) agent learns a stochastic policy $\pi_\theta(a_t|s_t)$, mapping states to action probabilities. Training stability is ensured through the clipped surrogate objective, which restricts large policy updates.

$$L^{\text{CLIP}}(\theta) = \mathbb{E}_t\Big[\min\big(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\big)\Big] \tag{3}$$

where $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}$ denotes the ratio of probabilities between the current policy and the previous policy, $\hat{A}_t$ represents the advantage estimate, measuring the relative benefit of an action compared to the expected outcome, and ϵ is a small hyperparameter that sets the clipping threshold.

The advantage $\hat{A}_t$ is computed using Generalized Advantage Estimation (GAE):

$$\hat{A}_t = \delta_t + (\gamma\lambda)\delta_{t+1} + \cdots + (\gamma\lambda)^{T-t+1}\delta_{T-1} \tag{4}$$

with the temporal-difference residual:

$$\delta_t = R_t + \gamma V(s_{t+1}) - V(s_t) \tag{5}$$

where $V(s_t)$ is the value function estimating expected future rewards, γ is the discount factor, and λ is a smoothing parameter. The policy network is updated iteratively using gradient ascent on $L^{\text{CLIP}}(\theta)$ to maximize expected cumulative reward.

This framework allows dynamic adaptation of pruning thresholds, prioritizing fidelity preservation while achieving significant gate reduction.

III. Results

The RL agent is trained on 2,356 randomly generated three-qubit circuits, each with 25–30 gates sampled from a universal set of single-qubit rotations (X, Y, Z, H, R_x, R_y, R_z) and two-qubit CNOT gates. Rotation parameters were drawn uniformly from $[-\pi, \pi]$. Circuits are regenerated each episode to ensure diversity, prevent overfitting, and promote robust, generalizable policy learning. Training spanned 300,000 time steps with a reward function promoting high fidelity and gate reduction. Bonus rewards were given when fidelity exceeded 90% and reduction surpassed 10%, while penalties discouraged reductions below 4%, ensuring meaningful pruning.

The trained RL agent was evaluated on three unseen quantum circuits to assess generalization capabilities. Circuit 1 contains 20 gates, Circuit 2 has 24 gates, and Circuit 3 has 21 gates, each composed of single-qubit rotations (X, Y, Z, H, R_x, R_y, R_z) and two-qubit CNOT gates. The agent achieved gate reductions of 5%, 12.5%, and 38.10% for Circuits 1–3, respectively, with corresponding fidelities of 98.51%, 99.02%, and 88.28%, illustrating a trade-off between compression and accuracy. These results indicate that the RL agent effectively balances gate reduction with functional preservation. Bonus rewards encouraged meaningful pruning, while penalties prevented trivial solutions where fidelity could be maximized without actual gate removal.

IV. Conclusion

An RL-based approach employing a PPO agent is proposed for adaptive quantum circuit pruning, with the goal of removing redundant gates while maintaining high fidelity. The agent determines which gates to prune by exploring the circuit environment and adjusting its actions to maximize a reward that considers both gate minimization and fidelity maintenance. Experimental results on a variety of quantum circuits demonstrate that the method effectively accomplishes the intended task. These findings highlight the potential of reinforcement learning for automated quantum circuit optimization. Future work will focus on deploying the method on IBM Quantum (IBMQ) hardware, scaling to larger and more complex circuits, and integrating with CNOT gate reduction strategies to further enhance pruning efficiency and overall circuit performance.

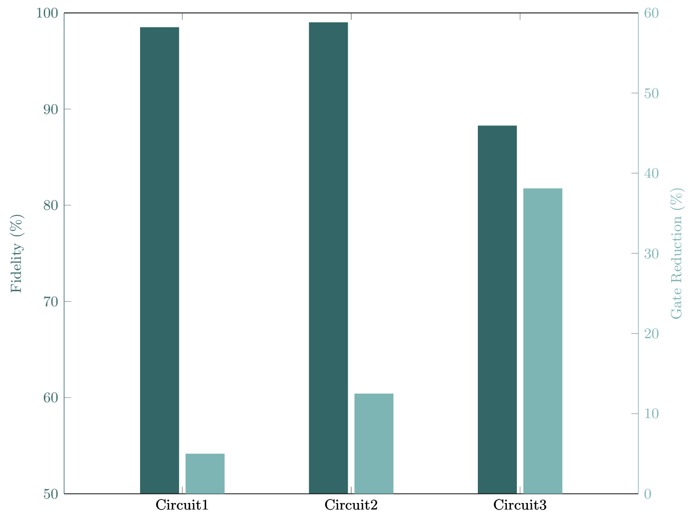

Figure 1. The plot presents the pruning results for three quantum circuits, showing fidelity and gate reduction percentages. As gate reduction increases, fidelity decreases, reflecting how pruning intensity varies with circuit complexity and highlighting the trade-off between reduction and fidelity.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) under RS-2025-00556064, by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2025-RS-2021-II212046) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation), and by a grant from Kyung Hee University in 2023 (KHU-20233663).

References

[1] U. Khalid, J. ur Rehman, H. Jung, T. Q. Duong, O. A. Dobre, and H. Shin, "Quantum property learning for NISQ networks: Universal quantum witness machines," *IEEE Trans. Commun.*, vol. 73, no. 4, pp. 2207–2221, Apr. 2025.

[2] U. Khalid, J. ur Rehman, S. N. Paing, H. Jung, T. Q. Duong, and H. Shin, "Quantum network engineering in the NISQ age: Principles, missions, and challenges," *IEEE Netw.*, vol. 38, no. 1, pp. 112–123, Jan. 2024.

[3] U. I. Paracha, S. M. A. Rizvi, M. M. U. Gondel, W. Park, and H. Shin, "Adaptive Quantum Readout Error Mitigation with Transfer Learning," in *Proc. International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju Island, Korea, Republic of, Oct. 2024.

[4] Y. Nam, J.-S. Chen, N. C. Pisenti, K. Wright, C. Delaney, D. Maslov, K. R. Brown, S. Allen, J. M. Amini, J. Apisdorf, K. M. Beck, A. Blinov, V. Chaplin, M. Chmielewski, C. Collins, S. Debnath, K. M. Hudek, A. M. Ducore, M. Keesan, S. M. Kreikemeier, J. Mizrahi, P. Solomon, M. Williams, J. D. Wong-Campos, D. Moehring, C. Monroe, and J. Kim, "Ground-state energy estimation of the water molecule on a trapped ion quantum computer," *NPJ Quantum Inf.*, vol. 6, no. 1, p. 33, Apr. 2020.

[5] B. Narottama, A. U. Haq, J. A. Ansere, N. Simmons, B. Canberk, S. L. Cotton, H. Shin, and T. Q. Duong, "Quantum deep reinforcement learning for digital twin-enabled 6G networks and semantic communications: Considerations for adoption and security," *IEEE Trans. Netw. Sci. Eng.*, Sep. 2025.

[6] A. Mattick, M. Periyasamy, C. Ufrecht, A. Y. Dubey, C. Mutschler, A. Plinge, and D. D. Scherer, "Optimizing quantum circuits via ZX diagrams using reinforcement learning and graph neural networks," 2025, arXiv:2504.03429.