Reinforcement Learning for Control Variational Quantum Algorithm

Sanjeev Shapkota[1], Yayu Mo[1], Chenxu Liu[2], Yanzhu Chen[3]
, Brian T Kirby [4,5], Thomas A Searles [6], Sanjaya Lohani [1,*]
[1]Southern Methodist University, Dallas, TX, USA
[2]Pacific Northwest National Laboratory, Richland, WA, USA
[3]Florida State University, Tallahassee, FL, USA
[4]Tulane University, New Orleans, LA, USA
[5]DEVCOM Army Research Laboratory, Adelphi, MD, USA
[6]University of Illinois Chicago, Chicago, IL, USA
Email: *slohani@smu.edu

***Abstract*—Variational quantum algorithms (VQAs) are key enablers of practical computations on near-term quantum computing hardware. However, gate-based ansatz suffer from decoherence and control errors, limiting scalability. Pulse-level approaches, such as control variational quantum eigen solver (ctrl-VQE), address these challenges by directly optimizing analog control fields and enabling time-optimal state preparation. Building on this paradigm, we present a reinforcement-learning (RL) framework that autonomously discovers high-fidelity, time-efficient control pulses. Using a hydrogen dimer testbed, our RL agent achieves state-fidelity above 0.995 and ground state energy errors near 0.005 Hartree, demonstrating the promise of AI-driven, hardware-efficient quantum algorithms.**

***Index Terms*—Variational quantum algorithms, VQE, quantum control, reinforcement learning, deep Q-learning**

I. Introduction

Near-term quantum computing devices face significant challenges in implementing deep quantum circuits due to short coherence times, limited connectivity, and gate errors. Variational quantum algorithms (VQAs), such as the Variational Quantum Eigensolver (VQE), mitigate these limitations by optimizing parameterized circuits using hybrid quantum-classical methods for tasks such as molecular ground-state energy estimation [1, 2]. Despite their success in quantum chemistry applications, gate-based ansatz often require deep circuits and costly gate decompositions, increasing resource demands and error rates [1].

Pulse-level approaches offer a promising alternative by directly tuning analog control signals that drive quantum hardware, bypassing explicit gate sequences [3]. Methods such as ctrl-VQE and PANSATZ enable faster, time-optimal state preparation and reduce accumulated errors, achieving energy accuracies comparable to or better than gate-based circuits [4]. Recent experiments have demonstrated the feasibility of ctrl-VQE implementations on superconducting platforms, highlighting their potential for hardware-efficient quantum simulations [5]. Recent work on quantum noise, distributed circuits, and task-specific quantum architectures highlights the importance of robust and adaptive design in quantum systems, motivating our use of reinforcement learning (RL) for pulse-level quantum control [6–9].

At the same time, RL has emerged as a powerful tool for quantum circuit optimization, capable of navigating large design spaces and autonomously discovering efficient strategies [10]. In this work, we integrate RL with ctrl-VQE to develop an AI-driven control framework. Using a hydrogen dimer mapped to a qubit–transmon model, we show that an RL agent can learn time-optimal, high-fidelity pulse sequences that outperform Monte-Carlo samplers in accuracy and runtime. These results demonstrate the promise of combining RL and pulse-level control for scalable, problem-aware quantum algorithms.

II. Ctrl-VQE and device-level Hamiltonians

The ctrl-VQE framework evolves the quantum state under an effective Hamiltonian composed of the device Hamiltonian (H_D) and the control Hamiltonian (H_C) [11]. After a total evolution time T, the resulting state is expressed as:

$$|\psi_I(T;\mathbf{\Omega})\rangle = \mathcal{T}\exp\left(-i\int_0^T dt\ \tilde{H}_{\mathrm{C}}(t)\right)|\psi_I(0)\rangle, \quad (1)$$

where $\tilde{H}_{\mathrm{C}}(t)$ is an effective Hamiltonian, which is equal to $e^{iH_{\mathrm{D}}t}\sum_q \Omega_q(t)\left(e^{i\nu_q t}a_q + e^{-i\nu_q t}a_q^\dagger\right)e^{-iH_{\mathrm{D}}t}$ [5]. Similarly, a_q and $a_q^\dagger$ are device operator rather than problem operator (like Fermionic operators corresponding to the molecular Hamiltonian), $\mathcal{T}$ denotes time ordering, and $\mathbf{\Omega}$ collects all control amplitudes.

For a given molecular Hamiltonian H_{mol}, the ctrl-VQE framework computes the expectation value of the Hamiltonian as the cost function. This cost function is iteratively minimized by optimizing the control parameters under the constraint of a maximum allowable evoluation time T [11].

III. Reinforcement-Learning Framework for Control-VQE

In ctrl-VQE, we do not rely on an ideal gate set. Instead, we directly tune the control pulses applied to the hardware [12].

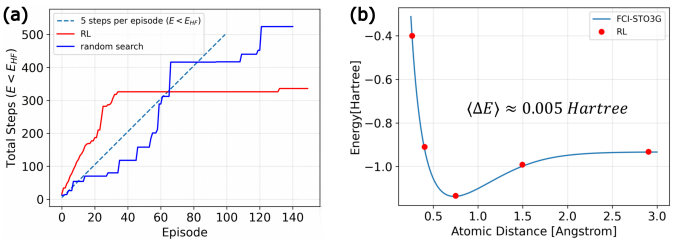

Fig. 1. (a) Learning metric for the RL agent. The plot shows the cumulative number of time steps with $E < E_{\mathrm{HF}}$ versus episode for the RL policy (red) and a random-search baseline (blue). The dashed line is a fixed reference rate of five such steps per episode; the RL curve crosses this line earlier, indicating faster exploration of useful regions. (b) Ground-state energy of the H–H molecule as a function of atomic separation. The solid line shows the reference FCI-STO3G energy curve, while the red markers indicate energies obtained from RL-optimized control pulses. The mean absolute deviation between the two is $\langle \Delta E \rangle \approx 0.005$ Hartree.

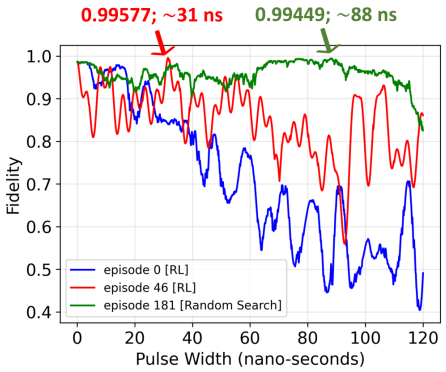

Fig. 2. State fidelity as a function of total pulse width. Curves show an early RL episode (blue), a later RL episode after training (red), and a high-fidelity random-search pulse (green). The trained RL policy reaches high fidelity at a significantly shorter pulse duration than the random-search baseline.

This is helpful because for larger systems it can be better to "abandon gates" and optimize the pulses that the device can actually implement. The goal is to find pulse parameters that prepare a final state whose molecular energy is as small as possible.

For small problems, the action-value function, $Q(s, a)$, can be stored in a table. However, in pulse control, the number of possible states and actions grows quickly (more qubits, more pulse slices, more amplitude choices). This makes a tabular representation unrealistic. That is why we implement the deep Q learning (DQN) setup to find the optimal Q-function [13]. The DQN framework can be approximated as,

$$Q^*(s, a) \approx Q(s, a; \boldsymbol{\theta}), \tag{2}$$

where $\boldsymbol{\theta}$ are the DQN's parameters.

In addition, the replay memory and exploration procedures are integrated directly into the learning pipeline. The reward structure for the learning agent is formulated to advance the ctrl-VQE objective by guiding the search toward the optimal energy landscape [14]. The ctrl-VQE is inherently one of the continuous-control problems. However, in this paper, we adopt a discrete bang-bang approximation and construct a sequence of square pulses. This limits the RL agent to searching over a binary bang-bang controls than the full continuous control space.

IV. Numerical Results: Hydrogen Dimer Case Study

We consider a hydrogen dimer (H–H) mapped to an effective two-qubit Hamiltonian using a minimal STO-3G basis, with the FCI ground-state energy as a reference [15]. The drift and control Hamiltonians are simulated to be compatible with a realistic transmon-based superconducting device with several accessible levels per qubit. We adopt the same fixed-frequency transmon setting with always-on coupling as described in [11], and we apply one microwave drive tone to each qubit to drive the interactions.

We begin by examining the exploration metric illustrated in Fig. 1 (a). Points located on or to the left of the dashed reference line (the "red line") denote higher exploration efficiency, and the trained RL agent progressively transitions into this region as training advances. For comparison, the blue line depicts the trajectory followed by an agent employing a Monte Carlo sampling technique.

After training, the RL agent learns pulse schedules, a sequence of pulses, that give energies very close to the exact (Full Configuration Interaction - FCI) value as shown in Fig. 1 (b). Averaging over several runs with different random starting points for the agent and environment, the mean absolute energy error is

$$\langle \Delta E \rangle = \left\langle \left| E(\boldsymbol{\Omega}) - E_{\mathrm{FCI}} \right| \right\rangle \approx 0.005 \text{ Hartree}. \tag{3}$$

The error magnitude is somewhat higher than what has been reported in earlier studies [4, 5, 16], but it remains within an acceptable range. This also suggests that the RL agent may require additional fine-tuning.

The main result is the balance between how accurately (fidelity) and quickly we can prepare the target state. To demonstrate the proof of principle, we use the same device parameters and time discretization for RL and random sampling methods. We then find the state fidelity against the target state by the RL agent against a random-search baseline over pulse parameters.

In a representative run, the RL agent finds a sequence of pulses that prepares a final-state fidelity of $F_{\mathrm{RL}} \approx 0.99577$ using only about $\sim$ 31 ns of the total evolution time. In contrast, the best sequence of pulses discovered by random search reaches $F_{\mathrm{rand}} \approx 0.99449$ but only after about ~ 88 ns as indicated in Fig 2. Note that, a single cross-resonance CNOT gate typically has a duration exceeding 31 ns in fixed-frequency transmon hardware [17].

V. Discussion and Outlook

This work advances pulse-level variational algorithms by showing that RL agent can efficiently discover compact, bang–bang–like control sequences consistent with prior

ctrl-VQE studies. By operating directly at the pulse level, our method unifies quantum optimal control and VQE within a single learning framework, enabling the agent to search for effective control signals while minimizing energy.

The approach might also align with hardware-aware strategies. Pulse-level control, combined with error-mitigation techniques, has demonstrated benefits on real devices, and an RL agent trained at edge could further adapt to device-specific noise and calibration drift. This adaptability positions RL-based control as a complementary tool to improve near-term quantum simulations.

In the future, we plan to scale the method to larger systems, develop reward functions that promote robustness, and benchmark against stronger classical optimizers. In general, integrating AI with control-based variational algorithms may offer a promising route toward resource-efficient, hardware-tailored quantum computation on near-term quantum devices.

Acknowledgments: TAS and SL are supported, in part, by the US Department of Energy, Office of Science. Co-design center for Quantum Advantage (C2QA) under contract number DE-SC0012704. SL also gratefully acknowledges partial support from the Sam Taylor Fellowship Award 2025 and startup funding from SMU.

References

[1] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio *et al.*, "Variational quantum algorithms," *Nature Reviews Physics*, vol. 3, no. 9, pp. 625–644, 2021.

[2] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, "Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz," *Quantum Science and Technology*, vol. 4, no. 1, p. 014008, 2018.

[3] K. N. Smith, G. S. Ravi, T. Alexander, N. T. Bronn, A. R. Carvalho, A. Cervera-Lierta, F. T. Chong, J. M. Chow, M. Cubeddu, A. Hashim *et al.*, "Programming physical quantum systems with pulse-level control," *Frontiers in physics*, vol. 10, p. 900099, 2022.

[4] D. Meirom and S. H. Frankel, "Pansatz: Pulse-based ansatz for variational quantum algorithms," *Frontiers in Quantum Science and Technology*, vol. 2, p. 1273581, 2023.

[5] A. Asthana, C. Liu, O. R. Meitei, S. E. Economou, E. Barnes, and N. J. Mayhall, "Leakage reduces device coherence demands for pulse-level molecular simulations," *Phys. Rev. Appl.*, vol. 19, p. 064071, Jun 2023. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevApplied.19.064071

[6] Y. Mo, M. A. Thornton, and S. Lohani, "Analyzing quantum noise in image encodings: A circuit-based hybrid integration approach," in *2025 IEEE International Conference on Quantum Computing and Engineering (QCE)*, vol. 02, 2025, pp. 276–279.

[7] S. Lohani, S. Regmi, J. M. Lukens, R. T. Glasser, T. A. Searles, and B. T. Kirby, "Dimension-adaptive machine learning-based quantum state reconstruction," *Quantum Machine Intelligence*, vol. 5, no. 1, p. 1, 2023.

[8] Y. Mo, S. Lohani, and M. Thornton, "Distorted edge feature extraction using quantum convolutional structure," in *2025 IEEE 18th Dallas Circuits and Systems Conference (DCAS)*, 2025, pp. 1–6.

[9] Y. Mo, M. A. Thornton, and S. Lohani, "Potential of distributed circuits for mitigating quantum image noise," in *2025 IEEE International Conference on Quantum Computing and Engineering (QCE)*, vol. 02, 2025, pp. 608–609.

[10] V. Seetohul, H. Jahankhani, S. Kendzierskyj, and I. S. Will Arachchige, "Quantum reinforcement learning: Advancing ai agents through quantum computing," in *Space Law Principles and Sustainable Measures*. Springer, 2024, pp. 55–73.

[11] O. R. Meitei, B. T. Gard, G. S. Barron, D. P. Pappas, S. E. Economou, E. Barnes, and N. J. Mayhall, "Gate-free state preparation for fast variational quantum eigensolver simulations," *npj Quantum Information*, vol. 7, no. 1, p. 155, 2021.

[12] K. M. Sherbert, H. Amer, S. E. Economou, E. Barnes, and N. J. Mayhall, "Parametrization and optimizability of pulse-level variational quantum eigensolvers," *Physical Review Applied*, vol. 23, no. 2, p. 024036, 2025.

[13] B. Schimkowitsch, "Control of pulsed lasers with high repetition rate using reinforcement learning," Ph.D. dissertation, Technische Universität Wien, 2024.

[14] K. Karuppasamy, V. Puram, S. Johnson, and J. P. Thomas, "A comprehensive review of quantum circuit optimization: Current trends and future directions," *Quantum Reports*, vol. 7, no. 1, p. 2, 2025.

[15] C.-L. Hong, T. Tsai, J.-P. Chou, P.-J. Chen, P.-K. Tsai, Y.-C. Chen, E.-J. Kuo, D. Srolovitz, A. Hu, Y.-C. Cheng *et al.*, "Accurate and efficient quantum computations of molecular properties using daubechies wavelet molecular orbitals: A benchmark study against experimental data," *PRX Quantum*, vol. 3, no. 2, p. 020360, 2022.

[16] D. J. Egger, C. Capecci, B. Pokharel, P. K. Barkoutsos, L. E. Fischer, L. Guidoni, and I. Tavernelli, "Pulse variational quantum eigensolver on cross-resonance-based hardware," *Physical Review Research*, vol. 5, no. 3, p. 033159, 2023.

[17] S. Sheldon, E. Magesan, J. M. Chow, and J. M. Gambetta, "Procedure for systematically tuning up cross-talk in the cross-resonance gate," *Phys. Rev. A*, vol. 93, p. 060302, Jun 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.93.060302