# Efficient VQE by Conditional Reinforcement Learning

DongHa Kim, June-Koo Kevin Rhee

School of Electrical Engineering. & KAIST Institute of IT Convergence, KAIST, Daejeon, S. Korea,
Qunova Computing, Daejeon, S.Korea

## Abstract

This study proposes the use of conditional Reinforcement Learning to address the optimization challenges in the Variational Quantum Eigensolver (VQE). While VQE is recognized as a promising algorithm, it faces the issues to re-optimize and redesign the ansatz structure whenever a new Hamiltonian is introduced. This paper aims to overcome this problem by employing conditional-Reinforcement Learning (RL) techniques. The proposed approach seeks to automatically learn the optimal ansatz and optimization strategies for various Hamiltonians, significantly enhancing the applicability and efficiency of VQE.

## Conditional Soft Actor Critic

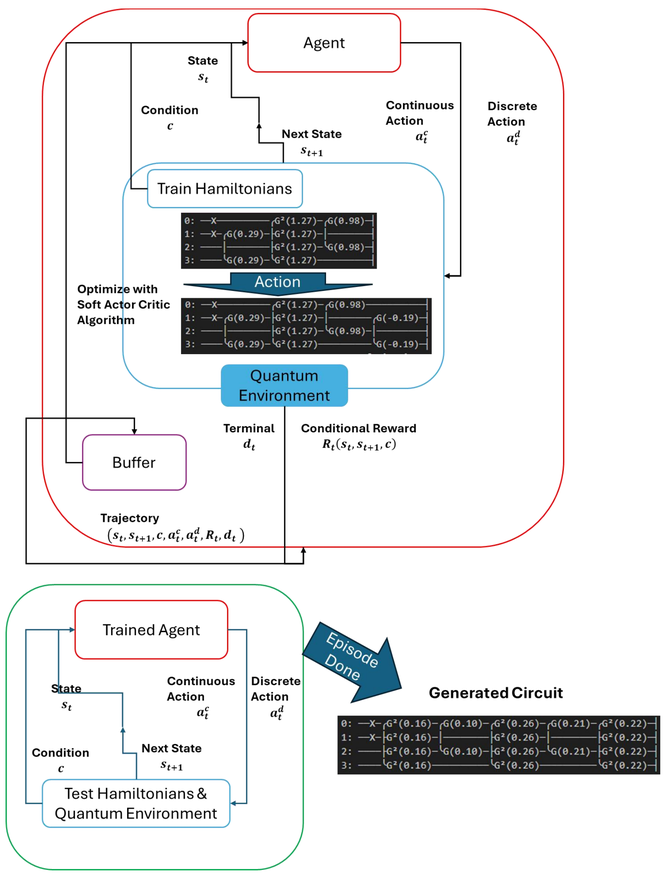
Our goal is to construct circuits from MDP that maximizes the reward for a given condition when it enters the actor. Therefore, one of the inputs to both the policy and Q networks, namely the state, is replaced with state and condition. Moreover, the reward function varies not only based on the state and action but also depending on the condition.

i.e.

$$\pi(\mathrm{s}) \rightarrow \pi(\mathrm{s},\mathrm{c})$$
$$Q(\mathrm{s},\mathrm{a}) \rightarrow Q(\mathrm{s},\mathrm{a},\mathrm{c})$$
$$\mathrm{r}(\mathrm{s},\mathrm{a}) \rightarrow \mathrm{r}(\mathrm{s},\mathrm{a},\mathrm{c})$$

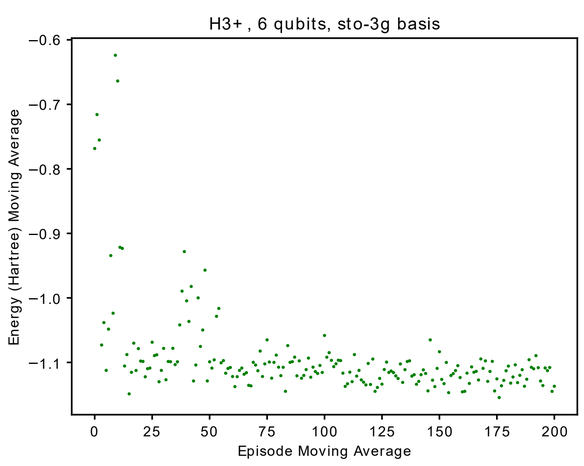
## Algorithm Overview

Here is diagram illustrating the algorithm. Using Conditional Soft Actor-Critic (C-SAC), the agent is trained on the training conditions. After training, the test condition is provided as the condition input to the agent, enabling the actor to generate a circuit.

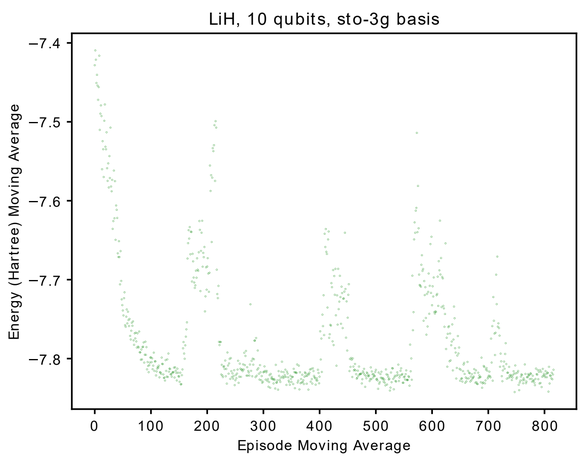



## Results (Training)

We applied our algorithm to two molecular structures: $\mathrm{H}_3^+$ (equilateral triangle) and $\mathrm{LiH}$.
The hamiltonians for these molecules used in the algorithm, were obtained using the Jordan-Wigner transformation and the STO-3G basis set with 2 active electrons.
The bond length data utilized ranged from 1.5 to 5 atomic units for $\mathrm{H}_3^+$ and 2.5 to 6 atomic units for $\mathrm{LiH}$.
Also, we constrained maximal number of quantum gate as 10, 20 for $\mathrm{H}_3^+$, $\mathrm{LiH}$ molecule.



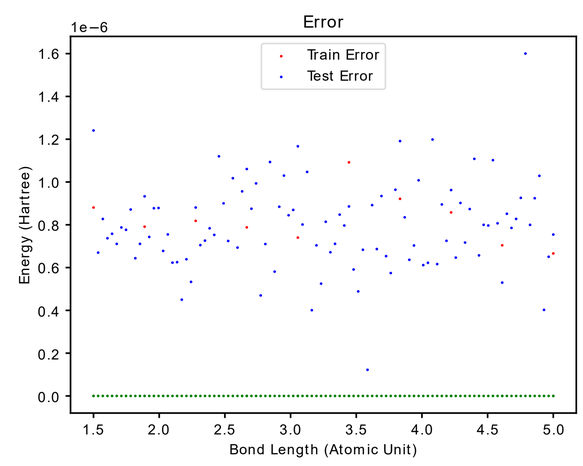

H3+ training graph (moving episode average 50)



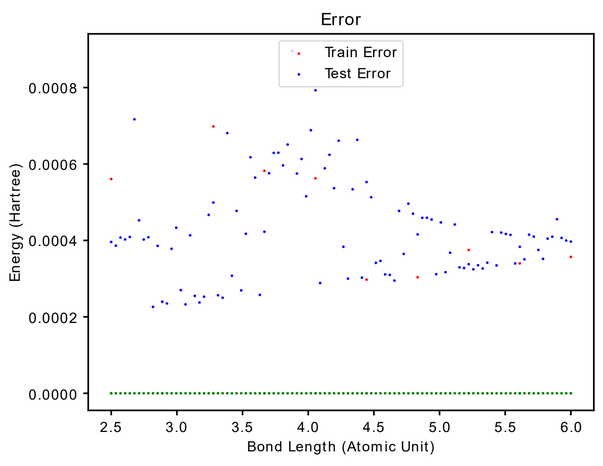

LiH training result (moving episode average 25)

## Results (Testing)

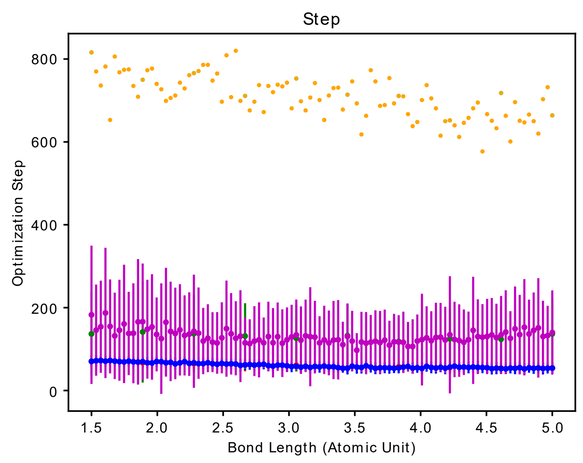
For the molecule $\mathrm{H}_3^+$ and $\mathrm{LiH}$ molecules, gradient descent was applied to a circuit with an ansatz structure generated by a discrete actor, using rotation angles generated by a continuous actor as the initial parameters for optimization.
The results demonstrate that the number of steps required by our algorithm is significantly lower compared to other methods for new hamiltonian condition.



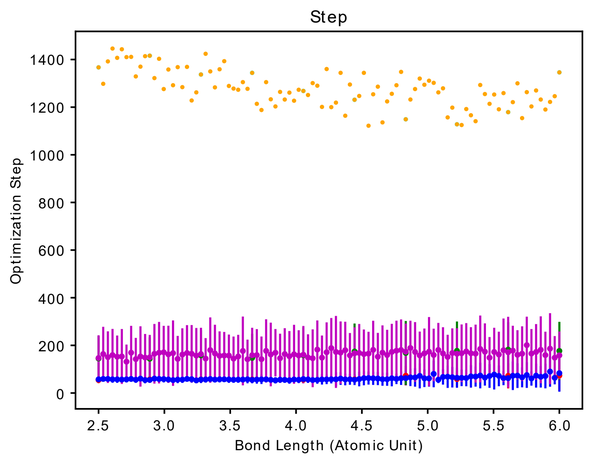

H3+ best error gap generated by agent + gradient descent



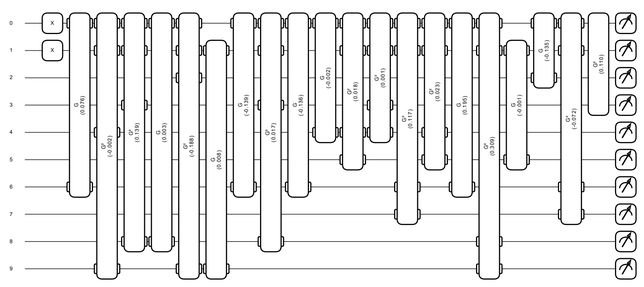

LiH best error gap generated by agent + gradient descent




H3+ average optimization step
(yellow : ADAPT-VQE, purple : random ansatz, blue : ours)




LiH average optimization step
(yellow : ADAPT-VQE, purple : random ansatz, blue : ours)



One of generated circuit with agent trained for LiH and condition at distance 3.7727 atomic unit

## Discussion

In this study, we developed an algorithm optimized for reducing quantum resource requirements when faced with new Hamiltonians.
First, we implemented reinforcement learning to generate quantum circuits via a Markov Decision Process (MDP).
Second, we introduced conditional reinforcement learning, enabling the agent to adapt to new condition values and generate corresponding circuits effectively.
Third, we employed the Soft Actor-Critic (SAC) algorithm in a conditional-continuous-discrete reinforcement learning framework to ensure that the results vary smoothly with changes in conditions.
Figures demonstrate the efficacy of our algorithm in requiring fewer quantum resources compared to traditional methods, optimization starting from random ansatz and random parameter initialization, and the ADAPT-VQE algorithm.
However, it is important to note that our algorithm demands a substantial number of quantum measurements during training, and the resulting error gap was larger than that observed with the ADAPT-VQE approach with same number of gate. This discrepancy can be attributed to the nature of reinforcement learning, which assumes that the reward function without a gradient, making precise convergence more challenging. This limitation highlights an area for further improvement in the accuracy and efficiency of our reinforcement learning-based algorithm.

## Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2021R1A2C2013790)