# Quantum circuit architectures via quantum observable Markov decision process planning

Tomoaki Kimura[1] Kodai Shiba[1,2] Chih-Chieh Chen[2,a)]
Masaru Sogabe[2] Katsuyoshi Sakamoto[1,3] Tomah Sogabe[1,2,3,b)]

***Abstract***: Algorithms for designing quantum circuit architectures are important steps toward practical quantum computing technology. Applying agent-based artificial intelligence methods for quantum circuit design could improve the efficiency of quantum circuits. We propose a quantum observable Markov decision process planning algorithm for quantum circuit design. Our algorithm does not require state tomography, and hence has low readout sample complexity. Numerical simulations for entangled states preparation and energy minimization are demonstrated. The results show that the proposed method can be used to design quantum circuits to prepare the state and to minimize the energy.

***Keywords***: Quantum circuit, Partially observable Markov decision process, planning algorithm

## 1. Introduction

Quantum computers are considered as potential technology to surpass the computational power of classical computers [1,2,3,4]. Variational quantum algorithms [5] have been actively researched. However, the design of a quantum circuit for solving a specific task sometimes requires empirical rules and domain knowledge. Reinforcement Learning (RL) method in Artificial Intelligence (AI) [6,7] has been successful in the areas such as robot control [8] and games [9, 10]. Applying RL to the control of quantum systems has been studied [11, 12, 13, 14, 15]. Most of these studies consider low level control at the hardware (Hamiltonian) level. To perform concrete quantum computation, however, it is also important to control at the circuit level, which is a higher level of abstraction [16]. For simple circuits, it is demonstrated that the closed-loop control can lead to better control performance for trapped-ion quantum processors [17]. State-of-the-art ion trap qubits have coherence time more than 10 minutes [18, 19], which provides enough running time for on-line decision process on a classical computer.

In this report, we consider applying RL to quantum feedback control at the circuit level [20]. The basic RL algorithms solve for Markov Decision Process (MDP), where the current state of the agent can be exactly known from the observation of the environment. But for a quantum system, the Born rule asserts that an observation result is drawn from a probabilistic distribution over the state space. Therefore, it is necessary to formulate the problem as a partially observable problem. Quantum Observable Markov Decision Process (QOMDP) [21, 22, 23, 24] was proposed as a quantum extension of the Partially Observable Markov Decision Process (POMDP) framework for the classical partially observable problems [25, 7], but no specific application of QOMDP was proposed. Our QOMDP planning approach is Bayesian, and does not rely on state tomography [26, 27, 28] or expectation evaluation [29, 30]. Hence it improves the quantum machine sample complexity per time step from $O(\epsilon^{-2}N_{obs})$ (or $O(\epsilon^{-4}(\log N_{obs})^4 \log(2^n))$ with shadow tomography [31]) to $O(1)$ for number of observables $N_{obs}$ and accuracy $\epsilon$. However, our approach still requires exponentially expensive classical planning. We formulate quantum control at the circuit level as a QOMDP reinforcement learning problem to solve for the quantum circuit design problem [20]. The exact QOMDP Bellman equation for value iteration is derived, and a concrete planning algorithm is proposed. In the exact POMDP planning for quantum state, there are three computational intractable parts. Firstly, the size of history set grows exponentially in time. Secondly, the Hilbert space is an uncountable set. Thirdly, the Hilbert space dimension grows exponentially with respect to the circuit width. We introduce the point-based value iteration (PBVI) algorithm from classical POMDP to make the approximating planning tractable and resolve the first and second issues. For the quantum Hilbert space, we perform exact filtering and do not make any approximation. Hence the calculations involving the belief state scale exponentially with respect to the number of qubits. We further consider two types of applications: the problem of state preparation and energy minimization. The proposed algorithm was able to make Bell state and GHZ state for state preparation. The algorithm is able to discover the low energy states for the H2 and H-He+. The experimental results show the applicability of QOMDP to quantum control at the circuit level. Comparing to variational quantum eigen solver (VQE) [32, 33, 34, 5] approach where the variational ansatz has to be chosen empirically, the QOMDP approach allows automatic search over a wide range of possible ansatzes.

## 2. Method

QOMDP [21] is defined by $\mathrm{Q} = \{\mathcal{S}, \mathcal{O}, A, R, \gamma, |s_0\rangle\}$. $\mathcal{S}$ is the Hilbert space of the system. $\mathcal{O} = \{o^1, \cdots, o^{|\mathcal{O}|}\}$ is the set of observations, where $|X|$ denotes the cardinality of a set $X$. $A =$

1 Engineering Department, The University of Electro-Communications, 182-8585 Tokyo, Japan
2 Grid Inc., 107-0061 Tokyo, Japan
3 i-PERC, The University of Electro-Communications, 182-8585 Tokyo, Japan
a) chen.chih.chieh@gridsolar.jp
b) sogabe@uec.ac.jp

$\left\{A^{a^1}, \cdots, A^{a^{|A|}}\right\}$ is the set of transition operators, and each operator $A^a = \left\{A^a_{o^1}, \cdots, A^a_{o^{|\mathcal{O}|}}\right\}$ has $|\mathcal{O}|$ Kraus matrices. The conditional probability of getting the observation $o$ when executing the action $a$ in the state $|s\rangle$ is

$$\Pr(o||s\rangle, a) = \langle s|A_o^{a\dagger}A_o^a|s\rangle.$$

The state transition is defined by

$$|s'\rangle(|s\rangle, a, o) \leftarrow \frac{A_o^a|s\rangle}{\sqrt{\langle s|A_o^{a\dagger}A_o^a|s\rangle}}.$$

$R = \left\{R_{a^1}, \cdots, R_{a^{|A|}}\right\}$ is the set of operators for rewards. The reward of executing action $a$ in state $|s\rangle$ is calculated by

$$r(|s\rangle, a) = \langle s|R_a|s\rangle.$$

$\gamma$ is the discount rate. $|s_0\rangle$ is the initial state. The structure of our QOMDP algorithm [20] is presented in Fig. 1. The agent selects an action according to the policy and executes the action for the environment. The operation $A_o^a$ corresponding to the action $a$ performed in the environment is executed, and the observation $o$ is fed back to the agent. The agent also receives a reward. The above action-observation-reward sequence is for a single time step. This is repeated until the end of the episode. The pseudocode is presented in Fig. 2. We simply present the algorithm here. The detail derivation and analysis for the algorithm could be found in [20].

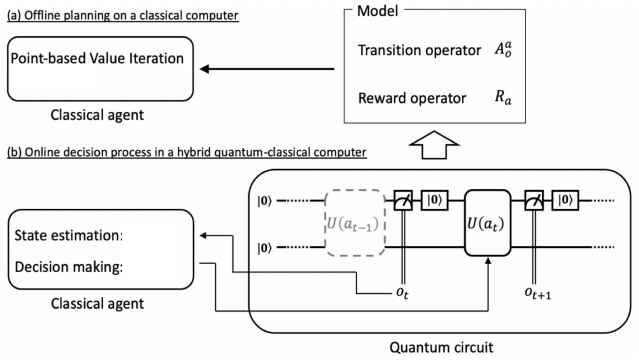


Fig. 1: QOMDP setting

**Algorithm 1**

Define Q = $\{\mathcal{S}, \mathcal{O}, A, R, \gamma, |s_0\rangle\}$
Define horizon H
Define maximum iteration I
def expand($\mathcal{S}^{\sim}$, Q)
  $\mathcal{S}^{\sim\prime} \leftarrow \{\}$
  for $|s\rangle$ in $\mathcal{S}^{\sim}$
    for $a$ in $\mathcal{A}$
      Sample observation $o$ according to probability $\Pr(o||s\rangle, a) = \langle s|A_o^{a\dagger}A_o^a|s\rangle$
      $|s'_a\rangle \leftarrow \frac{A_o^a|s\rangle}{\sqrt{\langle s|A_o^{a\dagger}A_o^a|s\rangle}}$
      Calculate minimum distance $d_a$ between $|s'_a\rangle$ and $\mathcal{S}^{\sim}$
    $a_{\max} = \operatorname{argmax}_a d_a$
    $\mathcal{S}^{\sim\prime} \leftarrow \mathcal{S}^{\sim\prime} \cup \{|s'_{a_{\max}}\rangle\}$
return $\mathcal{S}^{\sim} \cup \mathcal{S}^{\sim\prime}$

**Algorithm 2**

```
Define Q = {S, O, A, R, γ, |s_0⟩}
Define horizon H
Define maximum iteration I
Define minimum number of point set N
def point_based_update(S~, η, Q)
    η' ← { }
    η^{a,o} = {A_o^{a†} Υ A_o^a : Υ ∈ η}
    for |s⟩ in S~
        Υ^{a,|s⟩} = R_a + γ Σ_o argmax_{Υ∈η^{a,o}} ⟨s|Υ|s⟩
        backup(|s⟩) = argmax_{Υ∈{η_a^{|s⟩}}_{a∈A}} ⟨s|Υ|s⟩
        if not backup(|s⟩) in η'
            η' ← η' ∪ {backup(|s⟩)}
return η'
def plan(Q)
    S~ ← {|s_0⟩}
    While |S~| < N do
        S~ ← expand(S~, Q)
    Initialize η
    for iteration = 0, 1, ···, I - 1 do
        If iteration > 0
            S~ ← expand(S~, Q)
        for horizon = 0, 1, ···, H - 1 do
            η ← point_based_update(S~, η, Q)
return η
```

Fig. 2: Pseudocode for QOMDP-PBVI planning algorithm

## 3. Experiments

### 3.1 State preparation

For the state preparation task, the goal is to find a quantum circuit which produces a target state. The circuit is depicted in Fig. 3. Some examples of the planning results are demonstrated in Fig. 4. The reward matrix is given by

$$\boldsymbol{R_a} = \sum_o A_o^{a\dagger}|\boldsymbol{s}_{\mathbf{Target}}\rangle\langle\boldsymbol{s}_{\mathbf{Target}}|A_o^a,$$

and hence the reward is the state fidelity.

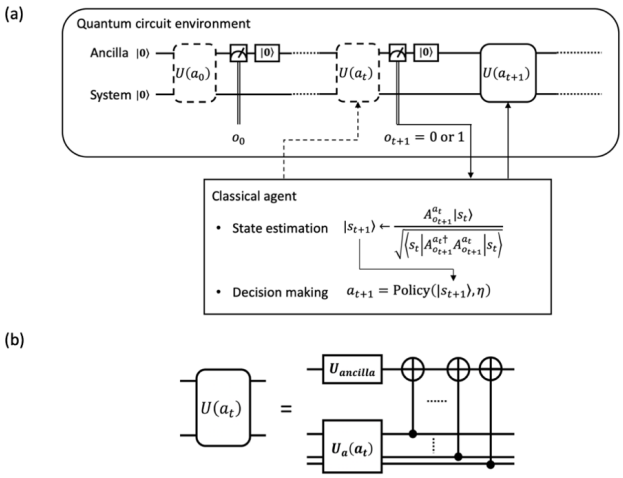


Fig. 3: State preparation circuits

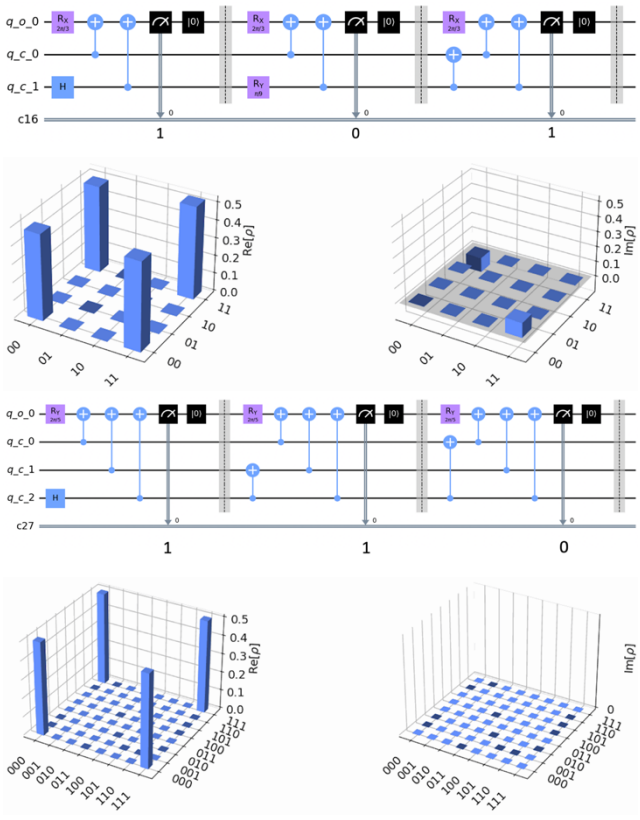


Fig. 4: QOMDP planning result for Bell-GHZ state. The circuits and the output density matrices are shown.

### 3.2 Energy minimization

For the energy minimization task, the goal is to find a quantum circuit which produces a minimum energy state for a given Hamiltonian H. The reward matrix is given by

$$\boldsymbol{R_a} = -\sum_o A_o^{a\dagger} H A_o^a \ ,$$

and hence the reward is the energy expectation. The result for H2 molecule energy minimization is shown in Fig. 5.

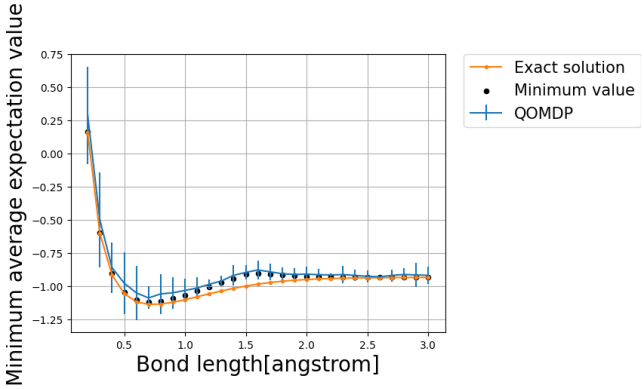


Fig. 5: Energy minimization results for H2 molecule. Energy unit is Hartree. Error bar denotes one standard deviation over 100 executions.

## 4. Conclusion

In this report, we propose a point-based value iteration QOMDP planning algorithm. State preparation and energy minimization applications are demonstrated. Numerical simulations provide proof-of-concept examples of the applicability of the proposed algorithm. One future direction is to look for suitable method to reduce the computational cost for the classical planning part of the algorithm.

## Reference

[1]John Preskill. 2018. Quantum computing in the NISQ era and beyond. Quantum 2 (2018), 79. https://doi.org/10.22331/q- 2018- 08- 06- 79

[2]Arute, F., Arya, K., Babbush, R. *et al.* Quantum supremacy using a programmable superconducting processor. *Nature* **574,** 505–510 (2019). https://doi.org/10.1038/s41586-019-1666-5

[3]Lov K. Grover. 1996. A fast quantum mechanical algorithm for database search. In Proceedings of the twenty-eighth annual ACM symposium on Theory of Computing (STOC '96). Association for Computing Machinery, New York, NY, USA, 212–219. DOI:https://doi.org/10.1145/237814.237866

[4]P. W. Shor, "Algorithms for quantum computation: discrete logarithms and factoring," Proceedings 35th Annual Symposium on Foundations of Computer Science, 1994, pp. 124-134, doi: 10.1109/SFCS.1994.365700

[5]Cerezo, M., Arrasmith, A., Babbush, R. et al. Variational quantum algorithms. Nat Rev Phys 3, 625–644 (2021). https://doi.org/10.1038/s42254-021-00348-9

[6]R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, A Bradford Book, Cambridge, MA, USA, 2018.

[7]S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, Pearson Education, London, UK, 4th edition, 2021

[8]Kormushev, P.; Calinon, S.; Caldwell, D.G. Reinforcement Learning in Robotics: Applications and Real-World Challenges. Robotics 2013, 2, 122-148. https://doi.org/10.3390/robotics2030122

[9]V. Mnih, K. Kavukcuoglu, D. Silver et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[10]D. Silver, J. Schrittwieser, K. Simonyan et al., "Mastering the game of Go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354–359, 2017.

[11]Sangkha Borah, Bijita Sarma, Michael Kewming, Gerard J. Milburn, and Jason Twamley, Phys. Rev. Lett. **127**, 190403 – Published 2 November 2021

[12]V. V. Sivak, A. Eickbusch, H. Liu, B. Royer, I. Tsioutsios, M. H. Devoret, Model-Free Quantum Control with Reinforcement Learning. arXiv:2104.14539 [quant-ph]
[13]Niu, M.Y., Boixo, S., Smelyanskiy, V.N. et al. Universal quantum control through deep reinforcement learning. npj Quantum Inf 5, 33 (2019). https://doi.org/10.1038/s41534-019-0141-3
[14]Nurdin, H.I. and Yamamoto, N., 2017. Linear Dynamical Quantum Systems. In Analysis, Synthesis, and Control. Springer
[15]He, RH., Wang, R., Nie, SS. et al. Deep reinforcement learning for universal quantum state preparation via dynamic pulse control. EPJ Quantum Technol. 8, 29 (2021). https://doi.org/10.1140/epjqt/s40507-021-00119-6
[16]Alfred Aho and Jeffrey Ullman. 2022. Abstractions, their algorithms, and their compilers. Commun. ACM 65, 2 (February 2022), 76–91. DOI:https://doi.org/10.1145/3490685
[17]Negnevitsky, V., Marinelli, M., Mehta, K.K. et al. Repeated multi-qubit readout and feedback with a mixed-species trapped-ion register. Nature 563, 527–531 (2018). https://doi.org/10.1038/s41586-018-0668-z
[18]Wang, Y., Um, M., Zhang, J. *et al.* Single-qubit quantum memory exceeding ten-minute coherence time. *Nature Photon* **11,** 646–650 (2017). https://doi.org/10.1038/s41566-017-0007-1
[19]Wang, P., Luan, CY., Qiao, M. *et al.* Single ion qubit with estimated coherence time exceeding one hour. *Nat Commun* **12,** 233 (2021). https://doi.org/10.1038/s41467-020-20330-w
[20]Tomoaki Kimura, Kodai Shiba, Chih-Chieh Chen, Masaru Sogabe, Katsuyoshi Sakamoto, Tomah Sogabe, submitted to Journal of Physics Communications.
[21]Jennifer Barry, Daniel T. Barry, and Scott Aaronson, Phys. Rev. A **90**, 032311 – Published 9 September 2014
[22]Cidre, Guillermo Andres. Planning in a quantum system. Carnegie Mellon University Pittsburgh, PA, 2016.
[23]S. G. Ying, M. S. Ying. Reachability analysis of quantum Markov decision processes. Information and Computation, vol. 263, pp. 31–51, 2018. DOI: 10.1016/j.ic.2018.09.001.
[24]Ying, MS., Feng, Y. & Ying, SG. Optimal Policies for Quantum Markov Decision Processes. Int. J. Autom. Comput. 18, 410–421 (2021). https://doi.org/10.1007/s11633-021-1278-z
[25]Christos H. Papadimitriou, John N. Tsitsiklis, (1987) The Complexity of Markov Decision Processes. Mathematics of Operations Research 12(3):441-450. https://doi.org/10.1287/moor.12.3.441
[26]Michael A. Nielsen and Isaac L. Chuang. 2011. Quantum computation and quantum information (10th Anniversary Edition). Cambridge University Press, New York.
[27]Esther Ye, Samuel Yen-Chi Chen. Quantum Architecture Search via Continual Reinforcement Learning. arXiv:2112.05779 [quant-ph]
[28]En-Jui Kuo, Yao-Lung L. Fang, Samuel Yen-Chi Chen. Quantum Architecture Search via Deep Reinforcement Learning. arXiv:2104.07715 [quant-ph]
[29]Shi-Xin Zhang, Chang-Yu Hsieh, Shengyu Zhang, Hong Yao, Differentiable Quantum Architecture Search, arXiv:2010.08561 [quant-ph]
[30]Mateusz Ostaszewski, Lea M. Trenkwalder, Wojciech Masarczyk, Eleanor Scerri, Vedran Dunjko, REINFORCEMENT LEARNING FOR OPTIMIZATION OF VARIATIONAL QUANTUM CIRCUIT ARCHITECTURES, arXiv:2103.16089 [quant-ph]
[31]Scott Aaronson. 2018. Shadow tomography of quantum states. In Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing (STOC 2018). Association for Computing Machinery, New York, NY, USA, 325–338. DOI:https://doi.org/10.1145/3188745.3188802
[32]Peruzzo, A., McClean, J., Shadbolt, P. et al. A variational eigenvalue solver on a photonic quantum processor. Nat Commun 5, 4213 (2014). https://doi.org/10.1038/ncomms5213
[33]McClean, J. R., Romero, J., Babbush, R., and Aspuru-Guzik, A., “The theory of variational hybrid quantum-classical algorithms”, New Journal of Physics, vol. 18, no. 2, 2016. doi:10.1088/1367-2630/18/2/023023.
[34]Kandala, A., Mezzacapo, A., Temme, K. *et al.* Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549,** 242–246 (2017). https://doi.org/10.1038/nature23879

**Acknowledgments** The authors gratefully acknowledge the funding from Information-technology Promotion Agency (IPA) under MITOU TARGET program 2021. We thank Naoki Yamamoto for valuable discussions.