Quantum Circuit Design: A Reinforcement Learning Challenge*

Extended Abstract

Philipp Altmann
LMU Munich
philipp.altmann@ifi.lmu.de

Adelina Bärligea
TU Munich

Jonas Stein
LMU Munich

Michael Kölle
LMU Munich

Thomas Gabor
LMU Munich

Thomy Phan
University of Southern California

Claudia Linnhoff-Popien
LMU Munich

ABSTRACT

To assess the prospects of using reinforcement learning (RL) for selecting and parameterizing quantum gates to build viable circuit architectures, we introduce the quantum circuit designer (QCD). By considering quantum control a decision-making problem, we strive to profit from advanced RL exploration mechanisms to overcome the need for granular specification and hand-crafted architectures. To evaluate current state-of-the-art RL algorithms, we define generic objectives that arise from quantum architecture search and circuit optimization. Those evaluation results reveal challenges inherent to learning optimal quantum control.

KEYWORDS

Reinforcement Learning, Quantum Computing, Circuit Optimization, Architecture Search

ACM Reference Format:
Philipp Altmann, Adelina Bärligea, Jonas Stein, Michael Kölle, Thomas Gabor, Thomy Phan, and Claudia Linnhoff-Popien. 2024. Quantum Circuit Design: A Reinforcement Learning Challenge: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

In the current NISQ era, quantum computing (QC) faces limitations in terms of size and precision [15]. Hybrid applications address these constraints to gain early insights and advantages. Hybrid quantum machine learning (QML) involves the use of QC to improve machine learning (ML) and vice versa, using ML to improve QC [4, 6]. This work focuses on the latter aspect, employing reinforcement learning (RL) to enhance the search for viable quantum architectures. Quantum architecture search denotes finding a sequence of unitary operations or gates to execute on a quantum computer and altering its qubits' state to achieve a specific objective [19]. Previous work mainly examined using RL to solve application-specific tasks [14]. To assess the overall capabilities of RL for quantum circuit design and quantum control, we posit a bottom-up approach and formulate a generic gymnasium [18] environment alongside common objectives.

2 QUANTUM CIRCUIT DESIGNER

The *quantum circuit designer* (QCD) is parameterized by the number of available qubits η and the maximum feasible circuit depth δ. To constantly monitor the current circuit, we base the QCD on quantum simulation [3], allowing efficient readout of the state vector. Thus, the **observation** is given by the full complex vector representation of its current state $s = |\Psi\rangle \in \mathbb{C}^{2^\eta}$, similar to [11]. To manipulate the carried quantum state, we use the ϕ-parameterized X-Rotation (1), the unparameterized CNOT operation (2), and the parameterized PhaseShift (3) and Controlled-PhaseShift (4):

$$\boldsymbol{RX}(\phi) = \exp\left(-i\frac{\phi}{2}\boldsymbol{X}\right) \tag{1}$$

$$\boldsymbol{CX} = |0\rangle\langle 0| \otimes \boldsymbol{I} + |1\rangle\langle 1| \otimes \boldsymbol{X} \tag{2}$$

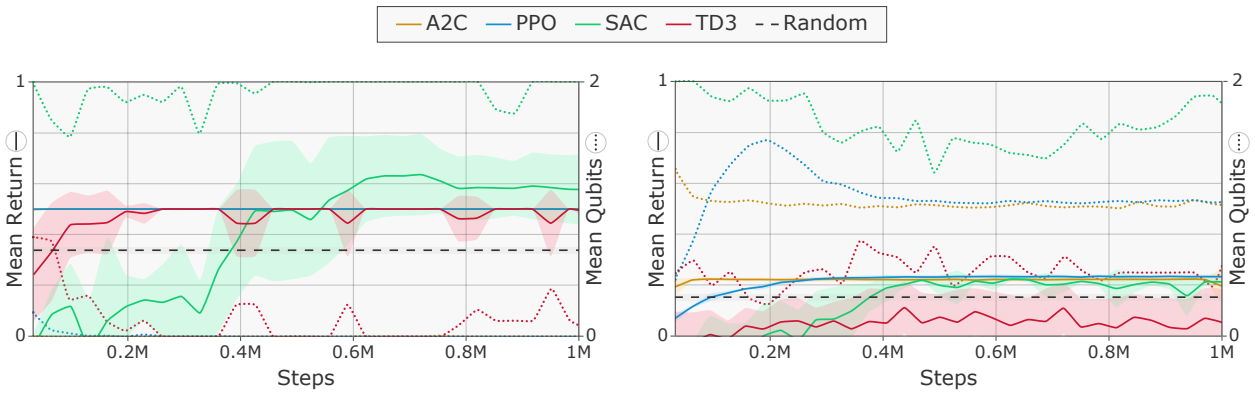
$$\boldsymbol{P}(\phi) = \exp\left(i\frac{\phi}{2}\right) \cdot \exp\left(-i\frac{\phi}{2}\boldsymbol{Z}\right) \tag{3}$$

$$\boldsymbol{CP}(\phi) = \boldsymbol{I} \otimes |0\rangle\langle 0| + \boldsymbol{P}(\phi) \otimes |1\rangle\langle 1| \tag{4}$$

To provide a balanced action space, we define the 4-dimensional join **action** $\langle o, q, c, \Phi\rangle = a \in \mathcal{A} = \{\Gamma \times \Omega \times \Omega \times \Theta\}$, with the discrete operation choice o, target and control qubits $q, c \in \Omega = [0, \eta - 1]$, and continuous parameterization $\Phi \in [-\pi, \pi]$. To the best of our knowledge, we are the first to consider learning the placement and parameterization of gates in a single closed loop. In contrast, most related work considers using a discrete action space, where circuits must be further optimized post hoc [8, 17]. To reduce the complexity of the operation decision $o \in \Gamma = \{\mathbb{X}, \mathbb{P}, \mathbb{M}, \mathbb{T}\}$, we apply an uncontrolled operation ($\boldsymbol{RX}$ or $\boldsymbol{P}$) iff. $q = c$ and a controlled operation ($\boldsymbol{CX}$ or $\boldsymbol{CP}$) otherwise. Furthermore, the agent can measure a specific qubit ($\mathbb{M}$) or terminate the current episode ($\mathbb{T}$), which is otherwise terminated when all available qubits are measured or the available depth δ is reached. Thus, given a deterministic action selection policy $\pi(a|s)$ and an operation mapping $g : \mathcal{A} \mapsto U$, circuits can be generated as $\Sigma_t = \langle g(a)\rangle_t$ at step $t \leq \eta \cdot \delta \cdot 2 = \sigma$. This budget of available operations per episode σ allows us to define a step cost $C_t = \max\left(0, \frac{3}{2\sigma}\left(t - \frac{\sigma}{3}\right)\right)$ to foster short and compact circuits. In addition to this agnostic cost, we consider the objectives

*This work is a short version of [1] and is part of the Munich Quantum Valley, which is supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus. The code is available at https://github.com/philippaltmann/QCD.

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).



(a) Bell State Preparation | $\eta = 2$, $\delta = 12$ **(b) Random Unitary Composition** | $\eta = 2$, $\delta = 12$

Figure 1: Benchmarking A2C (orange) [13], PPO (blue) [16], SAC (green) [10], and TD3 (red) [5] against a random baseline (dashed line) for the Bell State Preparation and Random Unitary Composition, challenges with regards to the Mean Return (solid) and Mean Qubits utilized (dotted) averaged over eight runs. Shaded areas mark the 95% confidence intervals.

of preparing a specific quantum state (SP) [9] and composing an arbitrary unitary (UC) [20]. Therefore, we define a task-specific **reward** $r_t = \mathcal{R}_t - C_t \in [0, 1]$, with $\mathcal{R}_t^{SP} = |\langle s_t|\Psi\rangle|^2$, given by the fidelity between the final state s_t and the target state Ψ, and $\mathcal{R}_t^{UC} = 1 - \arctan(\|\boldsymbol{U} - \boldsymbol{V}(\Sigma_t)\|)$, given by the Frobenius norm $\|\cdot\|$ between the unitary of the final circuit $\boldsymbol{V}(\Sigma_t)$ and the target unitary $\boldsymbol{U}$. As we are mainly interested in the resulting state post-measurement, the agent receives the reward sparsely upon episode termination. In addition to the *undiscounted return* $G(\tau) = \sum_{t=0}^{T} r_t$ averaged over 100 episodes τ (*Mean Return*), we use the *Mean Qubits* utilized in the generated circuits (bounded by η).

3 EVALUATION

To evaluate the performance of current RL approaches for QCD, we use the two-qubit *bell state* reflecting basic entanglement and the unitary of a random operation [12], central to versatile simulation capabilities [7]. The two-qubit bell state can be created by combining CNOT with the Hadamard operator, which, however, is precluded from our gate set and therefore needs to be reconstructed itself, e.g., via $H = \boldsymbol{P}(\pi/2)\boldsymbol{RX}(\pi/2)\boldsymbol{P}(\pi/2)$. This dependency constitutes the first central challenge QC yields for RL: Complex circuit architectures are often built upon hierarchical building blocks. This inherent structure justifies multi-level optimization approaches resembling similar hierarchies. However, prospects of the proposed closed-loop approach include exploring unconventional approaches that may provide out-of-the-box solutions to complex multi-level challenges. Indications of these prospects can be observed for the bell state preparation results shown in Fig. 1a, where SAC reaches mean returns above 60%. In contrast, the other approaches converge to a mean return of 0.5, with only TD3 showing some deviation. Looking at the utilized qubits reveals primal convergence to empty circuits, most likely caused by the multi-objective nature of the QCD. Consequently, RL exploits a local optimum introduced by the step cost, which, to be overcome, would require a performance decrease caused by using additional operations. Only SAC utilized both available qubits, presumably due to a more advanced exploration mechanism for continuous control spaces. This exploration challenge is less prominent with the composition of random unitaries, as shown in Fig. 1b. While SAC again shows the highest utilization of the available qubits, A2C and PPO at least explore single-qubit circuits. Except for TD3, all approaches exceed the random baseline with a final mean return of around 20%. As expected, random unitary composition poses a significantly harder objective when compared to bell state preparation, as the exact resemblance of the operator is required. Consequently, another challenge can be identified: In addition to the discrete choice of operation, target, and control qubits, precise control over the parameterization is required.

4 CONCLUSION

We introduced the *Quantum Circuit Designer*, a unified framework for benchmarking RL for low-level quantum control. To assess state-of-the-art RL algorithms, we consider two concrete objectives; preparing an entangled state and composing a random unitary operation, combined with the ubiquitous objective of producing compact circuits within the specified bound for the number of qubits η and circuit depth δ. Consequently, we identified the central challenges current RL algorithms face. Those mainly concern exploring a multi-modal reward landscape in combination with a complex non-uniform high-dimensional action space. Overcoming those challenges and using RL to design viable quantum circuits requires algorithms that support joint action spaces that combine discrete and continuous actions. Furthermore, smoother reward metrics, including intermediate rewards, are needed to ease the exploration of the intricate action space for quantum control. Also, an attention mechanism or partial observability might be helpful to focus on relevant parts of the state space and further decrease the exploration challenge [2]. Future work should also collect further concrete and relevant states and unitaries. Also, the overall objective might be extended to account for mitigating errors. Overall, however, we believe that the *quantum circuit designer* provides a profound base for future research on applying RL to quantum circuit design.

REFERENCES

[1] Philipp Altmann, Adelina Bärligea, Jonas Stein, Michael Kölle, Thomas Gabor, Thomy Phan, and Claudia Linnhoff-Popien. 2023. Challenges for Reinforcement Learning in Quantum Computing. arXiv:2312.11337

[2] Philipp Altmann, Fabian Ritz, Leonard Feuchtinger, Jonas Nüßlein, Claudia Linnhoff-Popien, and Thomy Phan. 2023. CROP: Towards Distributional-Shift Robust Reinforcement Learning Using Compact Reshaped Observation Processing. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, Edith Elkind (Ed.). International Joint Conferences on Artificial Intelligence Organization, Macao, SAR, 3414–3422. https://doi.org/10.24963/ijcai.2023/380 Main Track.

[3] Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, Shahnawaz Ahmed, Vishnu Ajith, M Sohaib Alam, Guillermo Alonso-Linaje, B AkashNarayanan, Ali Asadi, et al. 2022. PennyLane: Automatic differentiation of hybrid quantum-classical computations. arXiv:1811.04968

[4] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermanni Heimonen, Jakob S. Kottmann, Tim Menke, Wai-Keong Mok, Sukin Sim, Leong-Chuan Kwek, and Alán Aspuru-Guzik. 2022. Noisy intermediate-scale quantum algorithms. *Rev. Mod. Phys.* 94 (Feb 2022), 015004. Issue 1. https://doi.org/10.1103/RevModPhys.94.015004

[5] Stephen Dankwa and Wenfeng Zheng. 2020. Twin-Delayed DDPG: A Deep Reinforcement Learning Technique to Model a Continuous Movement of an Intelligent Robot Agent. In *Proceedings of the 3rd International Conference on Vision, Image and Signal Processing* (Vancouver, BC, Canada) *(ICVISP 2019)*. Association for Computing Machinery, New York, NY, USA, Article 66, 5 pages. https://doi.org/10.1145/3387168.3387199

[6] Vedran Dunjko, Jacob M. Taylor, and Hans J. Briegel. 2016. Quantum-Enhanced Machine Learning. *Phys. Rev. Lett.* 117 (Sep 2016), 130501. Issue 13. https://doi.org/10.1103/PhysRevLett.117.130501

[7] Richard P. Feynman. 1982. Simulating physics with computers. *International Journal of Theoretical Physics* 21, 6 (01 Jun 1982), 467–488. https://doi.org/10.1007/BF02650179

[8] Thomas Fösel, Murphy Yuezhen Niu, Florian Marquardt, and Li Li. 2021. Quantum circuit optimization with deep reinforcement learning. arXiv:2103.07585

[9] Thomas Gabor, Maximilian Zorn, and Claudia Linnhoff-Popien. 2022. The applicability of reinforcement learning for the automatic generation of state preparation circuits. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion* (Boston, Massachusetts) *(GECCO '22)*. Association for Computing Machinery, New York, NY, USA, 2196–2204. https://doi.org/10.1145/3520304.3534039

[10] Tuomas Haarnoja, Aurick Zhou, Kristian Hartikainen, George Tucker, Sehoon Ha, Jie Tan, Vikash Kumar, Henry Zhu, Abhishek Gupta, Pieter Abbeel, and Sergey Levine. 2019. Soft Actor-Critic Algorithms and Applications. arXiv:1812.05905

[11] Michael Kölle., Tom Schubert., Philipp Altmann., Maximilian Zorn., Jonas Stein., and Claudia Linnhoff-Popien. 2024. A Reinforcement Learning Environment for Directed Quantum Circuit Synthesis. In *Proceedings of the 16th International Conference on Agents and Artificial Intelligence (ICAART 2024) - Volume 1*. INSTICC, SciTePress, Rome, Italy, 83–94.

[12] Francesco Mezzadri. 2007. How to generate random matrices from the classical compact groups. arXiv:math-ph/0609050

[13] Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. 2016. Asynchronous Methods for Deep Reinforcement Learning. In *Proceedings of The 33rd International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 48)*, Maria Florina Balcan and Kilian Q. Weinberger (Eds.). PMLR, New York, New York, USA, 1928–1937. https://proceedings.mlr.press/v48/mniha16.html

[14] Mateusz Ostaszewski, Lea M. Trenkwalder, Wojciech Masarczyk, Eleanor Scerri, and Vedran Dunjko. 2021. Reinforcement learning for optimization of variational quantum circuit architectures. In *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (Eds.), Vol. 34. Curran Associates, Inc., 18182–18194. https://proceedings.neurips.cc/paper_files/paper/2021/file/9724412729185d53a2e3e7f889d9f057-Paper.pdf

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

[16] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal Policy Optimization Algorithms. arXiv:1707.06347

[17] Tomah Sogabe, Tomoaki Kimura, Chih-Chieh Chen, Kodai Shiba, Nobuhiro Kasahara, Masaru Sogabe, and Katsuyoshi Sakamoto. 2022. Model-Free Deep Recurrent Q-Network Reinforcement Learning for Quantum Circuit Architectures Design. *Quantum Reports* 4, 4 (2022), 380–389. https://doi.org/10.3390/quantum4040027

[18] Mark Towers, Jordan K. Terry, Ariel Kwiatkowski, John U. Balis, Gianluca de Cola, Tristan Deleu, Manuel Goulão, Andreas Kallinteris, Arjun KG, Markus Krimmel, Rodrigo Perez-Vicente, Andrea Pierré, Sander Schulhoff, Jun Jet Tai, Andrew Tan Jin Shen, and Omar G. Younis. 2023. Gymnasium. https://doi.org/10.5281/zenodo.8127026

[19] Shi-Xin Zhang, Chang-Yu Hsieh, Shengyu Zhang, and Hong Yao. 2022. Differentiable quantum architecture search. *Quantum Science and Technology* 7, 4 (aug 2022), 045023. https://doi.org/10.1088/2058-9565/ac87cd

[20] Yuan-Hang Zhang, Pei-Lin Zheng, Yi Zhang, and Dong-Ling Deng. 2020. Topological Quantum Compiling with Reinforcement Learning. *Phys. Rev. Lett.* 125 (Oct 2020), 170501. Issue 17. https://doi.org/10.1103/PhysRevLett.125.170501