

Distributed quantum architecture search using multi-agent reinforcement learning

M.Sergeev¹, D.Rabinovich¹, G.Paradezhenko¹ and V.Palyulin¹

¹Skolkovo Institute of Science and Technology, 30 Bolshoy Boulevard, building 1, Moscow, 121205, Russia
e-mail: Mikhail.Sergeev@skoltech.ru

Quantum architecture search (QAS) automates the design of parameterized quantum circuits for variational algorithms. While reinforcement learning (RL) is a promising QAS method, recent single-agent implementations suffer from poor scalability due to the growing action space and computational cost [2]. To overcome this, we introduce MARL-QAS, a novel multi-agent RL algorithm for QAS utilizing multi-agent reinforcement learning (MARL). In MARL-QAS, multiple agents act independently on partitioned blocks of the quantum circuit, significantly reducing the action space per agent and enabling scalable, distributed training.

We benchmark MARL-QAS on the MaxCut combinatorial problem, comparing it with the standard QAOA approach as a baseline [3]. The environment involves circuit composition and parameter optimization. The available gate set is

$$\mathcal{G} = \{R_j^x(\theta), R_j^y(\theta), \text{CNOT}_{j,j-1}, \text{CNOT}_{j,j+1}\}, \quad (1)$$

where $R_j^\alpha(\theta)$ are single-qubit rotations, and CNOT gates connect adjacent qubits with periodic boundary conditions.

The agents receive a shared reward $r = 2\eta - \rho t$, where $\rho > 0$ is a depth penalty, t is the step number, and η is the average approximation ratio over N_p problem Hamiltonians H_p :

$$\eta = \frac{1}{N_p} \sum_{p=1}^{N_p} \frac{\lambda_{\max}(H_p) - E_p}{\lambda_{\max}(H_p) - \lambda_0(H_p)}. \quad (2)$$

Here, λ_0 and $\lambda_{\max}$ are the ground and highest excited state energies, and E_p is the energy of the constructed ansatz. The algorithm thus aims to minimize the averaged energy.

As shown in Fig. 1, DQMARL converges as effectively as single-agent counterparts but enables distributed execution. This eliminates the computational overhead of simulating the entire system on real hardware. The approach can also be extended to meta-RL scenarios for various quantum machine learning tasks.

Fig. 2 compares MARL-QAS and QAOA circuits. For 3-regular graphs, QAOA requires $3np$ CNOT gates for depth p . While MARL-QAS uses slightly more variational parameters (e.g., 13 vs 10 for $n = 12$), it requires 2–3 times fewer CNOT gates, making it highly favorable for NISQ devices.

Next, we benchmark the proposed MARL-QAS algorithm on the problem of ground state energy search

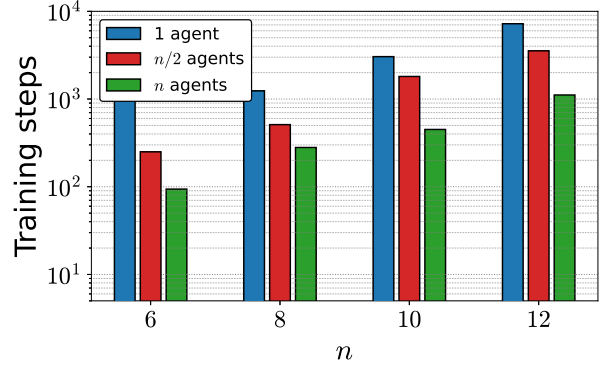


Figure 1: Training steps to reach $\eta \geq 0.95$ for 3-regular MaxCut instances of size n , averaged over multiple runs for single-agent and MARL-QAS ($m = n/2$ and n agents).

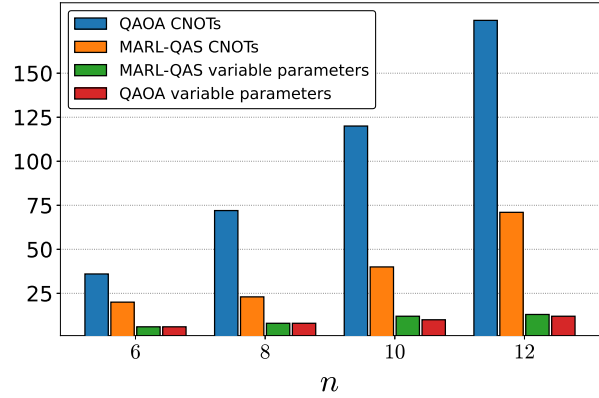


Figure 2: Number of CNOT gates and variational parameters for MARL-QAS vs QAOA for 3-regular MaxCut graphs of size n . QAOA depth p is chosen to reach $\eta = 0.95$.

for the Schwinger model Hamiltonian. The latter is a prominent example of a model with a strongly entangled ground state. The problem Hamiltonian reads

$$H = w \sum_{j=1}^{n-1} [\sigma_j^+ \sigma_{j+1}^- + \text{H.c.}] + \frac{m_0}{2} \sum_{j=1}^n (-1)^j \sigma_j^z + \bar{g} \sum_{j=1}^n L_j^z, \quad (3)$$

where $\sigma_j^\pm = \frac{1}{2}(\sigma_j^x \pm i\sigma_j^y)$, $L_j = \epsilon_0 - \frac{1}{2} \sum_{\ell=1}^j (\sigma_\ell^z + (-1)^\ell)$, w is the spin flip-flop coupling, m_0 is the bare mass, $\bar{g}$ is the electric field coupling, and ϵ_0 is the background electric field. In our simulations, in Eq. (3) we set $w = 1$ as the energy unit, $m_0 = 1$, $\bar{g} = 1$ and $\epsilon_0 = 0$.

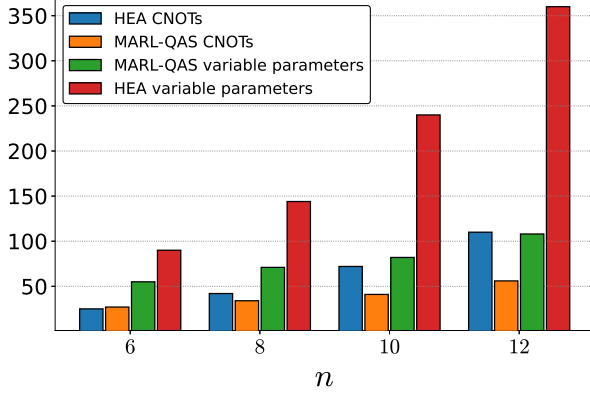


Figure 3: Number of two-qubit CNOT gates and number of variational parameters for circuits designed by the MARL-QAS algorithm for minimizing the Schwinger Hamiltonian (3) in comparison to the standard HEA (4) for different system sizes n . For each n , the minimal depth of HEA L was chosen as the smallest depth, sufficient to reach the threshold of $\eta = 0.95$ in (2).

A standard approach for minimizing (3) is to apply a well-known expressive ansatz, for instance, the hardware-efficient ansatz (HEA). This ansatz consists of L layers: each comprises single-qubit rotations $R^\alpha(\theta)$, $\alpha = x, y, z$ applied to each qubit, followed by CNOT gates arranged in the ladder pattern,

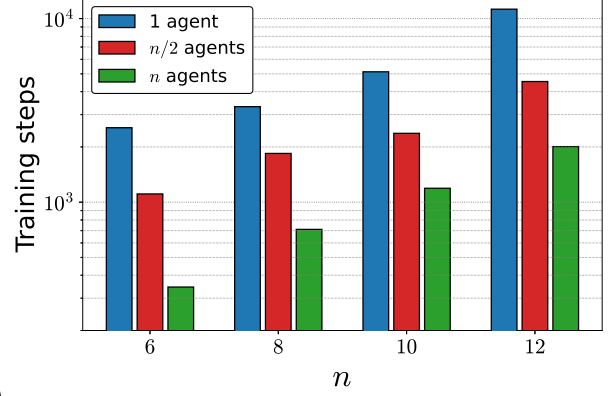
$$U(\theta) = \prod_{k=1}^L \left[\prod_{j=1}^{n-1} \text{CNOT}_{j,j+1} \times \prod_{j=1}^n R_j^z(\theta_{k,j,3}) R_j^y(\theta_{k,j,2}) R_j^x(\theta_{k,j,1}) \right]. \quad (4)$$

Additionally, we compare the properties of the circuits designed by MARL-QAS to HEA (4) for each system size n , graphically represented at Fig. 3. Their performance in terms of the approximation ratio (2) is almost the same. However, the MARL-QAS circuits require about half as much two-qubit gates as in HEA.

Similarly to the MaxCut problem, in Fig. 4 we demonstrate the effect of multiple RL agents in the RL-based QAS on the number of training steps for searching a problem-specific PQC for the Schwinger model (3). As one can see the results are qualitatively the same as for the combinatorial optimization shown in Fig. 1.

[1] Martyniuk D., et al., IEEE Int. Conf. on Quantum Computing and Eng. (QCE), 1, 1695 (2024).

[2] He, Z., et al., The European Physical Journal Plus, 137(4), 491 (2022).



a)

Figure 4: Number of training steps required to converge to the optimal circuit composition in the RL-based QAS with the approximation ratio $\eta \geq 0.95$ for minimizing the Schwinger Hamiltonian (3) for different system sizes n . The results are averaged over multiple runs using the single-agent and MARL-QAS algorithms for $m = n/2$ and n agents.

[3] Farhi E., et al. arXiv preprint arXiv:1411.4028 (2014).