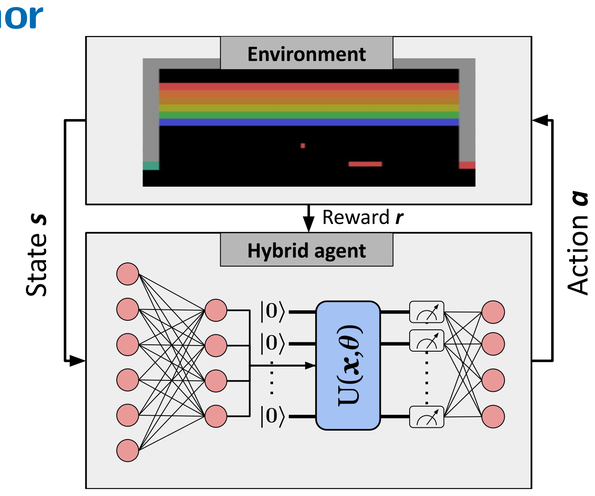
A Hybrid Quantum-Classical Framework for Reinforcement Learning of Atari Games

Master's Thesis Computational Science and Engineering

Carried out at the Faculty of Informatics TU Vienna
Supervised by Univ.Prof. Dr. R. Grosu, Dr. S. Jerbi & Dipl.-Ing. J. Lemmel

Author



Dominik Freinberger MSc

- **Bachelors:** Physics
- **Key Area Mathematics**
  - Machine Learning (R. Mayer, N. Musliu)
  - GPU Architectures and Computing (E. Bartocci)
  - Adv. Regression and Classifictaion (P. Filzmoser)
- **Key Area Informatics**
  - Seminary Computational Mathematics (J. Schöberl)
  - Scientific Computing for FEM (J. Schöberl)
  - Computational Finance (A. Jüngel)
- **Current Occupation:** Researcher in Quantum Machine Learning at RISC Software GmbH
- **Favorite Lectures:** Numerical methods for PDE's, Numerical Simulation and Scientific Computing I & II, Computational Science on Many-Core Architectures, Machine Learning

Introduction

In this work, we introduce a hybrid quantum-classical model based on parameterized quantum circuits (PQCs) for reinforcement learning (RL) in high-dimensional observation spaces, using the Atari 2600 games Pong and Breakout as testbeds. We show that our model solves Pong and achieves a performance comparable to a classical baseline in Breakout. We also present an in-depth analysis of the design choices driving performance gains, highlighting that reward rescaling and output-layer learning-rate adjustments significantly affect learning. We attribute these findings to the distinct Q-function landscapes learned by hybrid and classical models. This research extends previous studies that first demonstrated the learning capabilities of quantum and hybrid models in an approximate Q-learning setting [1, 2, 3] in simple benchmarking environments from the OpenAI Gym.

Quantum Machine Learning

Given an initial n-qubit quantum state $|\psi\rangle := |0\rangle^{\otimes n}$, a PQC applies the following parameter-dependent unitary transformation $U(\boldsymbol{x}, \boldsymbol{\theta})$ to its qubits:

$$U(\boldsymbol{x}, \boldsymbol{\theta}) = \prod_{l=1}^{L} V_l(\boldsymbol{\theta}) U_l(\boldsymbol{x}). \tag{1}$$

The $U_l(\boldsymbol{x})$ encode parts of the feature vector $\boldsymbol{x} \in R^d$ into the quantum state, and the $V_l(\boldsymbol{\theta})$ depend on adjustable parameters $\boldsymbol{\theta} \in \mathbb{R}^k$ that are optimized by classical hardware. The expectation value of a measurement observable on the resulting state defines a deterministic quantum machine learning model:

$$\langle \mathcal{M} \rangle_{x,\theta} = \langle \psi(\boldsymbol{x}, \boldsymbol{\theta}) | \mathcal{M} | \psi(\boldsymbol{x}, \boldsymbol{\theta}) \rangle =: f_\theta(\boldsymbol{x}). \tag{2}$$

We incorporate classical convolutional layers to preprocess high-dimensional input into fewer, informative features for the PQC. We define the hybrid model as:

$$\mathcal{Q}_{hybrid} = L_{w_{out}}\big(f_\theta\big(L_{w_{in}}(\tilde{x})\big)\big). \tag{3}$$

Here, $L_{w_{in}} : \mathbb{R}^{n_{in}} \to \mathbb{R}^{n_q}$ maps raw inputs $\tilde{x}$ to a lower dimension; $f_\theta : \mathbb{R}^{n_q} \to \mathbb{R}^{n_q}$ is the PQC from Eq. 2; and $L_{w_{out}} : \mathbb{R}^{n_q} \to \mathbb{R}^{n_{out}}$ is a classical post-processing layer that yields the final output. The parameters $\boldsymbol{\theta}$, $\boldsymbol{w}_{in}$, and $\boldsymbol{w}_{out}$ are jointly optimized on classical hardware.

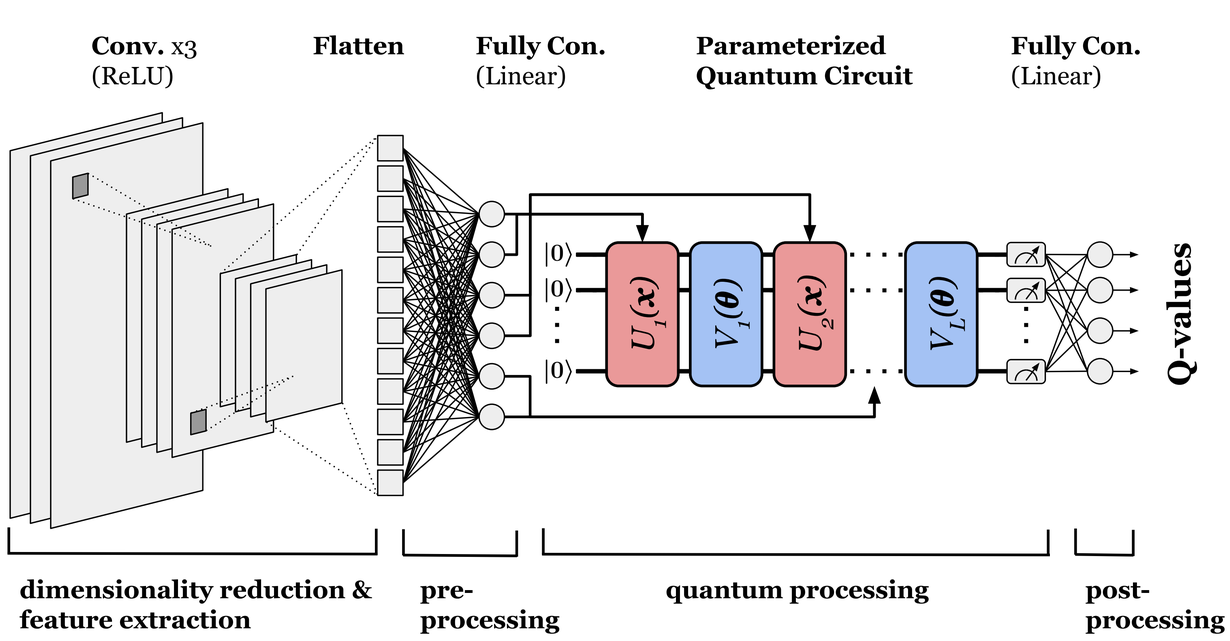
Performance Evaluation

We evaluate the hybrid quantum-classical model's performance within the Q-learning framework, comparing it to a classical reference model. The agent learns an approximation $Q(s, a; \boldsymbol{\theta})$ of the optimal Q-function, which represents expected future rewards for a given state-action pair. From this Q-function, an optimal policy is derived. The loss, minimized via gradient descent, incorporates the temporal difference (TD) error and stabilizes training using a target model $\hat{Q}$:

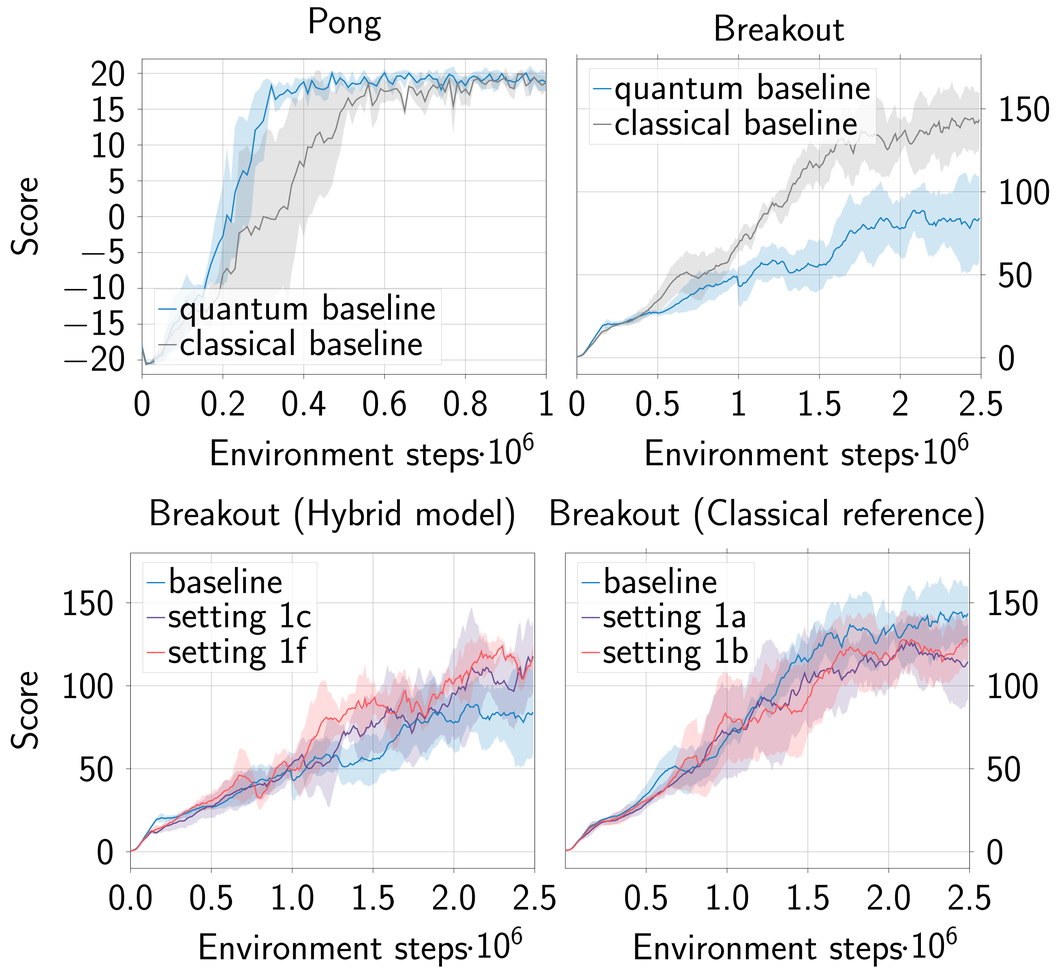
$$\mathcal{L}(\boldsymbol{\theta}) = \Big(r_t + \gamma \max_{a'} \hat{Q}(s_{t+1}, a'; \boldsymbol{\theta}^-) - Q(s_t, a_t; \boldsymbol{\theta})\Big)^2, \tag{4}$$

Experiments were conducted in Pong and Breakout, starting with baseline comparisons between the hybrid and classical models. To explore the influence of certain design choices, we up-scale the environment rewards by factors of 10 (settings 1c and 1a) and 100 (settings 1f and 1b) and adapt the post-processing layer learning rate by corresponding factors of 100 and 1000 for both models. Performance is measured as the total undiscounted reward per episode, averaged over multiple runs to ensure statistical significance. To ensure a fair comparison, the classical reference model incorporates a bottleneck analogous to the hybrid model's pre-processing layer by adding a layer with an equivalent number of neurons.

Hybrid Quantum-Classical Model



Experimental Results



Discussion and Conclusion

The hybrid model solves the game of Pong and demonstrates competitive performance in Breakout compared to the classical model. In Breakout, up-scaling the rewards in the environment and higher learning rates in the hybrid model's post-processing layer significantly enhance its performance (settings 1c and 1f), achieving rewards over 100, while the classical model shows no such benefit (settings 1a and 1b), likely due to differences in the form of the predicted Q-value hypersurfaces. These results contributes to the understanding of near-term quantum learning models and makes an important step towards their deployment in real-world RL scenarios. Future research could explore the robustness of the proposed architecture in noisy simulators and, ultimately, on real quantum devices.

References

[1] Samuel Yen-Chi Chen et al. "Variational Quantum Circuits for Deep Reinforcement Learning". In: *IEEE Access* 8 (2020), pp. 141007–141024. DOI: 10.1109/ACCESS.2020.3010470.

[2] Owen Lockwood and Mei Si. "Reinforcement Learning with Quantum Variational Circuit". In: *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* 16.1 (Oct. 2020), pp. 245–251. DOI: 10.1609/aiide.v16i1.7437.

[3] Andrea Skolik, Sofiene Jerbi, and Vedran Dunjko. "Quantum agents in the Gym: a variational quantum algorithm for deep Q-learning". In: *Quantum* 6 (May 2022). Publisher: Verein zur Förderung des Open Access Publizierens in den Quantenwissenschaften, p. 720. ISSN: 2521-327X. DOI: 10.22331/q-2022-05-24-720.