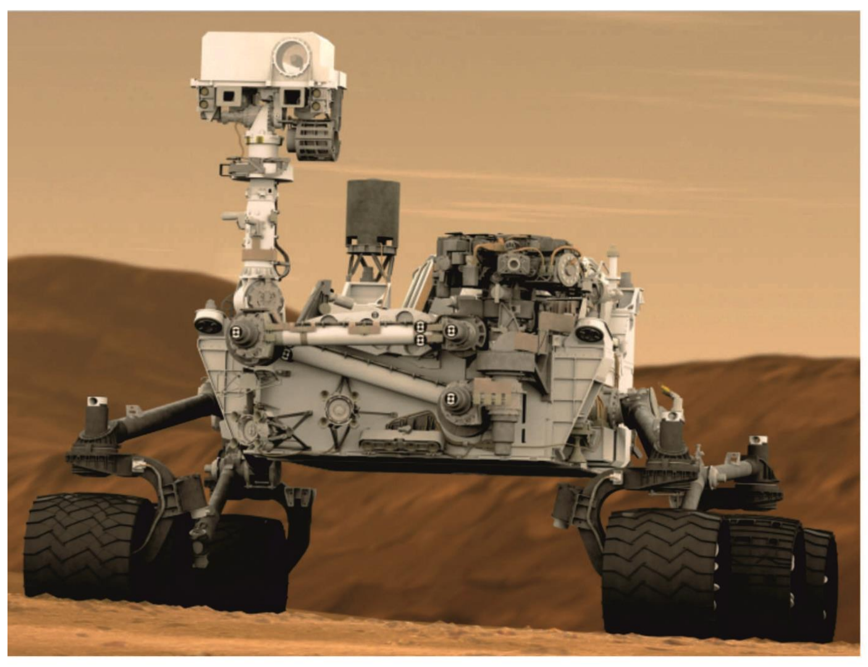
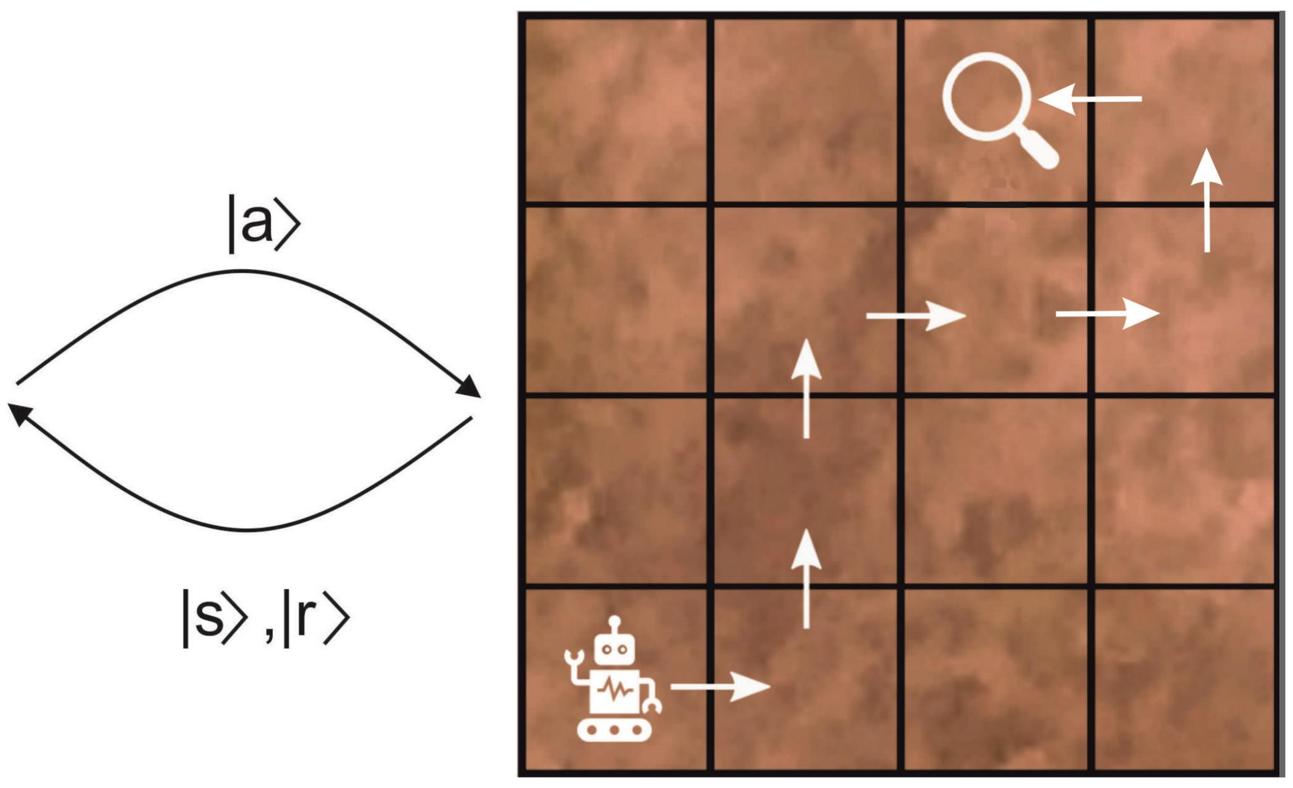
QLearning
Quantum Processors for reinforcement learning

PD Dr. Sabine Wölk
DLR Institute of Quantum Technologies



Credit: NASA/JPL-Caltech



A learning agent learns to solve a given problem by interacting with its environment by performing actions $|a\rangle$ and receiving information $|s\rangle$ about the new environmental state and rewards $|r\rangle$. Its behavior is described by some probability distribution $\Pi(a|s)$ called policy. Learning is achieved by updating the policy Π in order to maximize the long term expected reward.

A quantum enhanced learning agent

For round based problems, we can define a quantum enhanced learning agent in the following way:

1. Given the behavior of a classical agent for a complete round/epoch described by the probability distribution $\Pi(\vec{a})$, prepare the quantum states:

$$|\psi\rangle_A = \sum_{\vec{a}} \sqrt{\Pi(\vec{a})}\,|\vec{a}\rangle_A$$

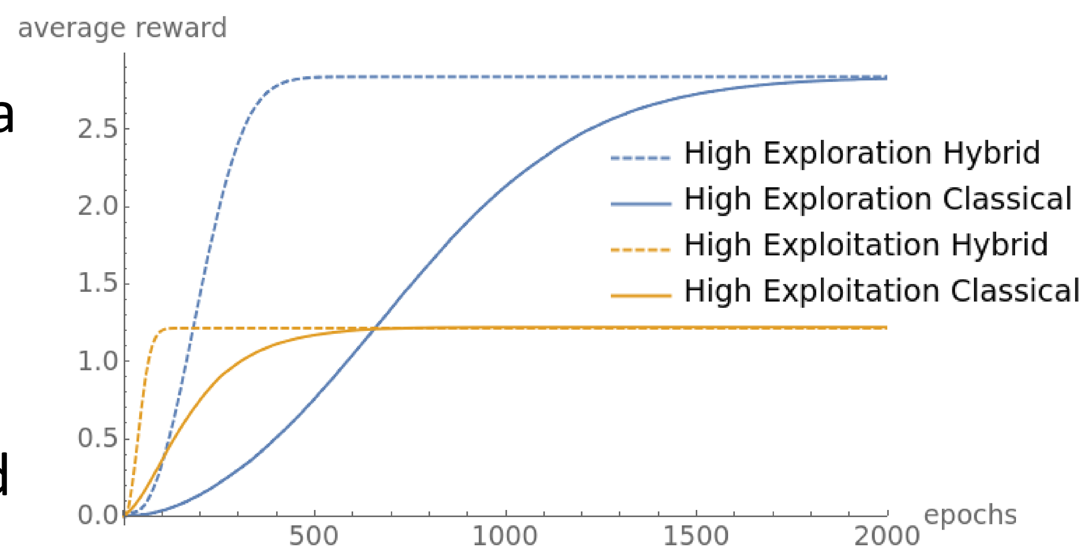
$$|-\rangle_R = (|0\rangle_R - |1\rangle_R)/\sqrt{2}$$

2. Perform several rounds of interactions to perform amplitude amplification leading to

$$|\tilde{\psi}\rangle_A = \sum_{\vec{a}} \sqrt{\tilde{\Pi}(\vec{a})}\,|\vec{a}\rangle_A$$

3. A measurement on $|\tilde{\psi}\rangle_A$ leads with higher probability to a sequence of actions $\vec{a}$ with a high reward r.
4. Perform the sequence of actions $\vec{a}$ (classically) to observe the sequence of states $\vec{s}$ and reward r.
5. Update the classical policy
6. Play another classical or quantum round/epoch

First results

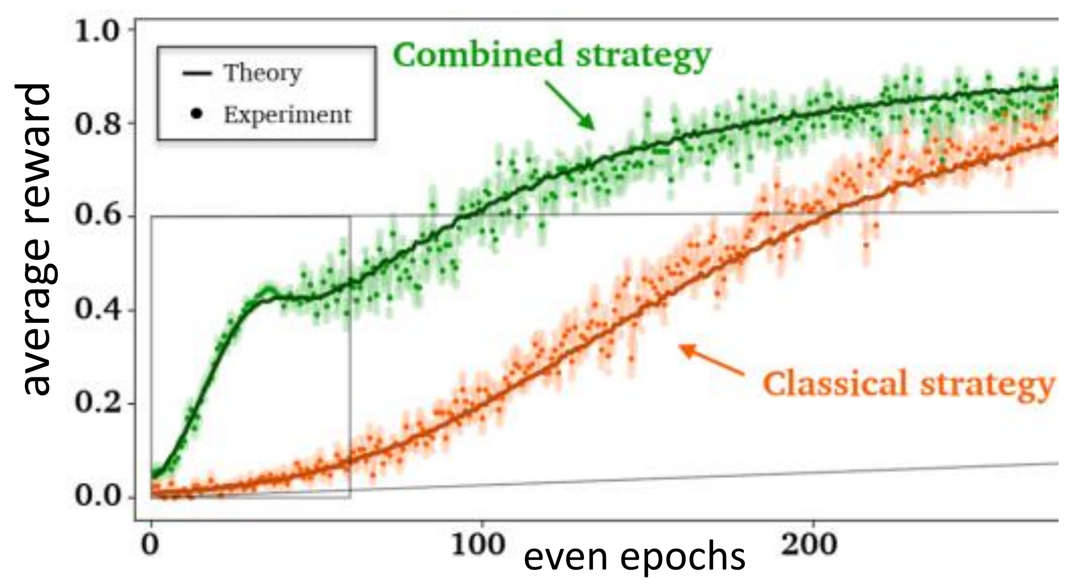


- Many classical reinforcement learning agents can be transformed into hybrid quantum-classical learning agents
- Combining quantum exploration with classical learning updates leads to faster learning in luck-favoring settings
- We proved a quadratic speed-up in the learning time $\langle T\rangle$ for reward-based learning updates

$$\langle T\rangle_Q \leq \alpha\sqrt{\langle T\rangle_C \langle J\rangle}$$

with $\langle J \rangle$ being the average number of rewards necessary to learn

- In a proof-of-principle experiment with photons at Vienna University we demonstrated a speed-up from $\langle T\rangle_C = 270$ epochs to $\langle T\rangle_Q = 100$



Applications

- Routing optimization (minimal time, minimal cost, minimal environmental impact, ….)
- Development of routines for self-driving cars such as parking
- Development of routines for robotics
- Optimization in trading and finance time series problems
- Development of dynamic treatment regimes in healthcare
- Optimization of work flows in engineering
- Development of quantum routines for e.g. state preparation or quantum error correction
- In general, all problems where we need to sequentially combine several actions in order to solve the problem in an optimal way

Industry cooperation

- Development and investigation of use cases
- Investigation and adaption of quantum hardware for quantum enhanced reinforcement learning
- Optimal implementation of multi-qubit quantum gates
- Creation of software tools and packages for customers of quantum enhanced reinforcement learning

Literature

- A. Hamann and S. Wölk, *Performance analysis of a hybrid agent for quantum-accessible reinforcement learning*, New J. Phys. **24**, 229 (2022)
- A. Hamann, V. Dunjko and S. Wölk, *Quantum-accessible reinforcement learning beyond strictly epochal environments*, Quantum Mach. Intell. **3**, 22 (2021)
- V. Saggio et al., *Experimental quantum speed-up in reinforcement learning agents*, Nature **591**, 229 (2021)

Contact

PD Dr. Sabine Wölk
DLR Institute of Quantum Technologies
Theoretical Quantum Physics
sabine.woelk@dlr.de

Gefördert durch:
Bundesministerium für Wirtschaft und Klimaschutz
aufgrund eines Beschlusses des Deutschen Bundestages

DLR
DLR Quantencomputing Initiative