Extending the Hybrid Agent for Reinforcement Learning Beyond Fixed-Length Scenarios

Oliver Sefrin, Sabine Wölk
Institute of Quantum Technologies, German Aerospace Center (DLR), Ulm

Abstract

- Based on amplitude amplification / Grover's algorithm, a hybrid algorithm can speed up Reinforcement Learning (RL) quadratically
- This algorithm requires knowledge of a sufficient number of steps per RL episode (=episode length)
- Here, we extend the algorithm to function without this knowledge, using a probabilistic strategy
- Simulations show a possible advantage towards "harder" RL environments

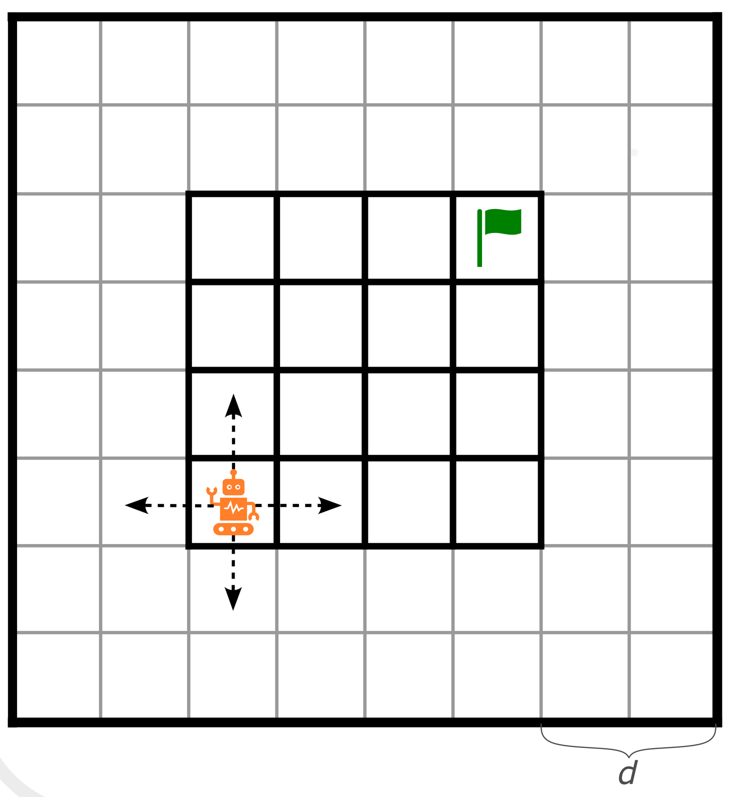
Extended Hybrid Algorithm

- Without knowledge of a sufficient episode length L, we need to vary it
- Main idea: start small, double L probabilistically
- Re-use parameter m of Boyer's algorithm (upper bound for Grover iterations) as "impatience"

Algorithm 1 Probabilistic Hybrid Algorithm

Require: policy $\pi(\vec{a})$
  $L \leftarrow 1,\ m \leftarrow 1,\ \lambda \leftarrow 6/5$ ▷ L: episode length
  rewarded ← false
  while not rewarded **do**
    $r \leftarrow$ random number in $[0, 1]$
    if $r < \frac{\log(m)}{\log(2)\cdot L}$ **then** ▷ prob. to double L: $p_L(m) = \frac{\log(m)}{\log(2)\cdot L}$
      $L \leftarrow 2 \cdot L$
      $m \leftarrow 1$
    $k \leftarrow$ random integer in $[0, m)$
    $|\psi\rangle \leftarrow \sum_{\vec{a}\in\mathcal{A}^{\otimes L}} \sqrt{\pi(\vec{a})}\, |\vec{a}\rangle_A |0\rangle_S |-\rangle_R$ ▷ state preparation
    $|\psi'\rangle \leftarrow G^k |\psi\rangle$ ▷ amplitude amplification
    $\vec{a}' \leftarrow$ **measure** $|\psi'\rangle$
    if $r(\vec{a}') = 1$ **then**
      rewarded ← true
    else
      $m \leftarrow \min(\lambda \cdot m,\ 2^L)$

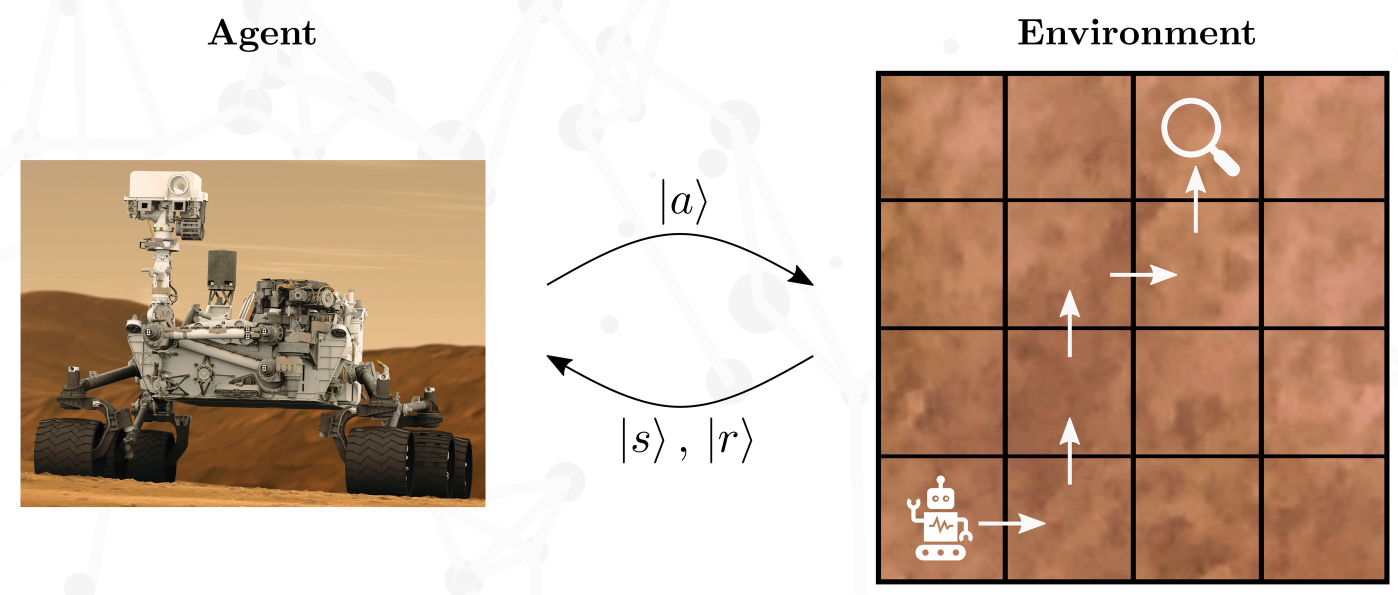
Simulation Details



Assume:

- $\mathcal{A} = \{\text{up, down, left, right}\}$
- untrained agent
  → uniform policy $\pi(a) = \frac{1}{|\mathcal{A}|}\ \forall a \in \mathcal{A}$
- quadratic *base area* (inner square), no inner walls, start & goal in opposite corners
- *outer walls* in a distance d of the base area

Hybrid Agent for Reinforcement Learning

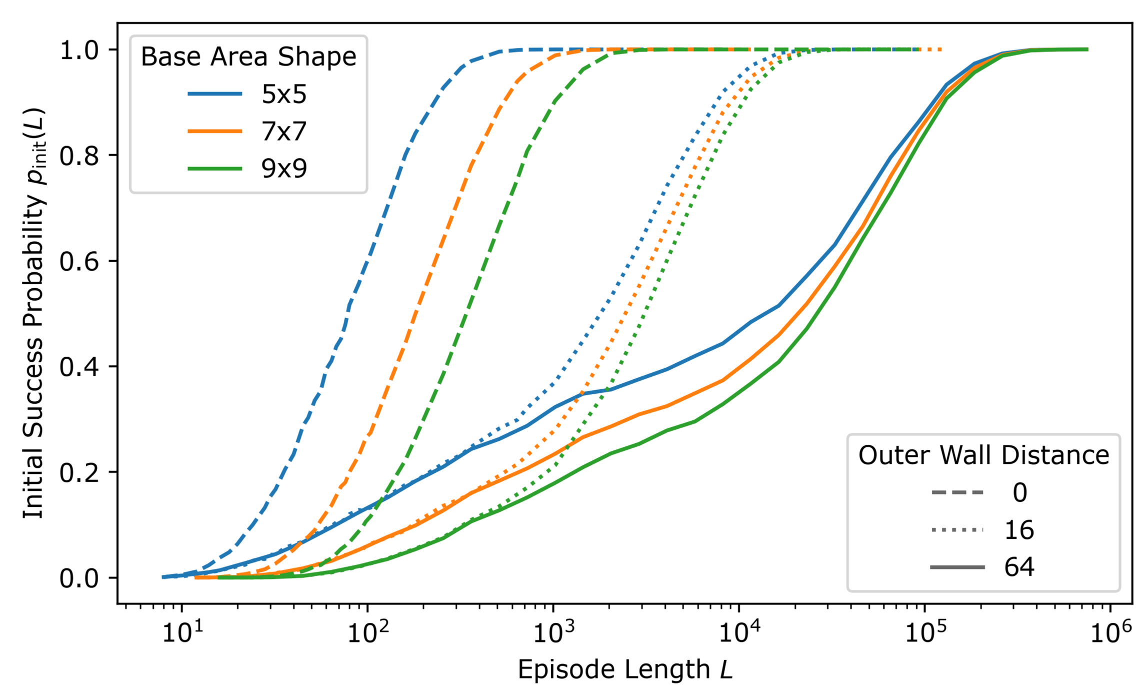


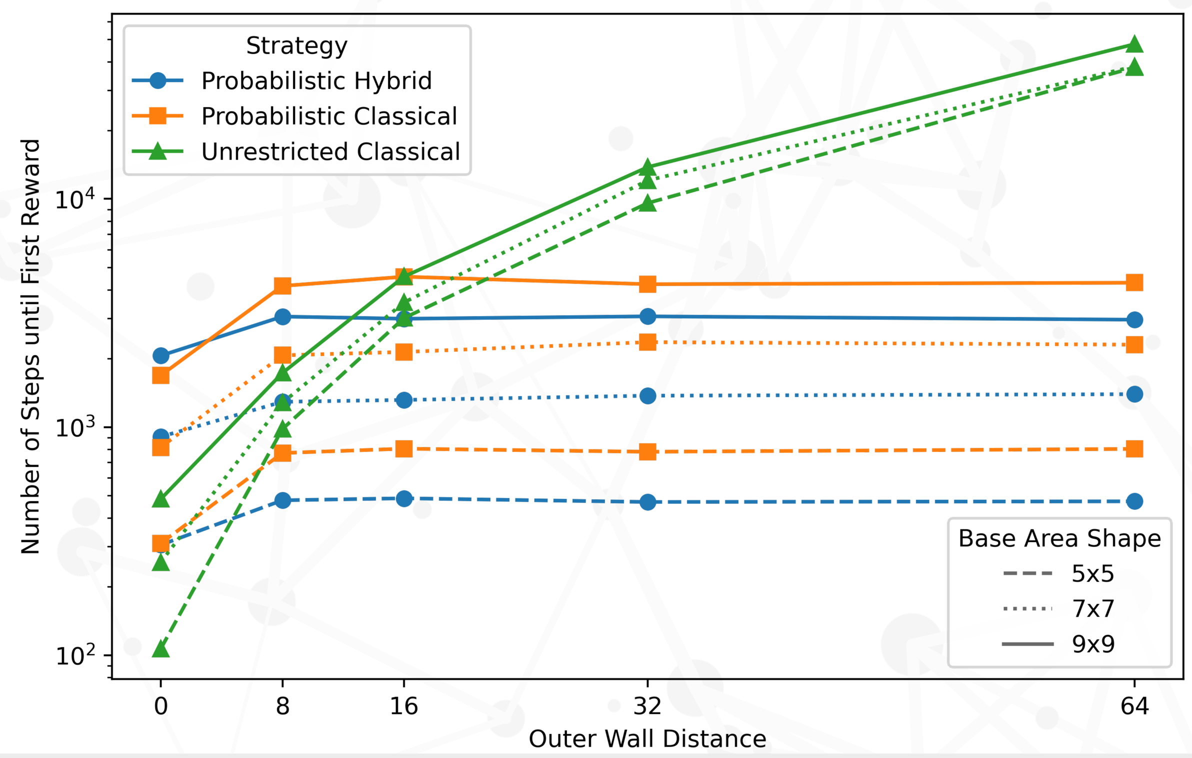
- Agent and Environment interact by exchanging quantum states
- Environment's response U_{env} can be used to create an effective phase kickback oracle and thus a Grover operator G (for a certain class of environments) [1]
- Alternate between:
  quantum round:
  - perform amplitude amplification (AA) using Boyer's AA algorithm for an unkown number of solutions [2]
  - measure an action sequence

  classical round:
  - test the measured action sequence
  - update the policy according to the RL algorithm

⇒ Theoretically proven [3] and experimentally verified [4] quadratic speedup in terms of sample complexity

Results





Conclusion

- The probabilistic strategy provides a valuable addition in RL scenarios with little or no information about the problem layout, especially for finding the very first reward
- No further hyperparameters are introduced, requiring no extra tuning
- In environments with large state spaces and slowly increasing success probabilties (for increasing episode length), the hybrid agent outperforms classical agents (in terms of the total number of steps until a reward is found)

[1] Dunjko, V., Taylor, J. M., & Briegel, H. J. (2016). Quantum-enhanced machine learning. Physical review letters, 117(13), 130501.
[2] Boyer, M., Brassard, G., Høyer, P., & Tapp, A. (1998). Tight bounds on quantum searching. Fortschritte der Physik: Progress of Physics, 46(4-5), 493-505.
[3] Hamann, A., & Wölk, S. (2022). Performance analysis of a hybrid agent for quantum-accessible reinforcement learning. New Journal of Physics, 24(3), 033044.
[4] Saggio, V., Asenbeck, B. E., Hamann, A., Strömberg, T., Schiansky, P., Dunjko, V., ... & Walther, P. (2021). Experimental quantum speed-up in reinforcement learning agents. Nature, 591(7849), 229-233.

DLR Deutsches Zentrum für Luft- und Raumfahrt
Institut für Quantentechnologien
QCI DLR Quantum Computing Initiative
oliver.sefrin@dlr.de