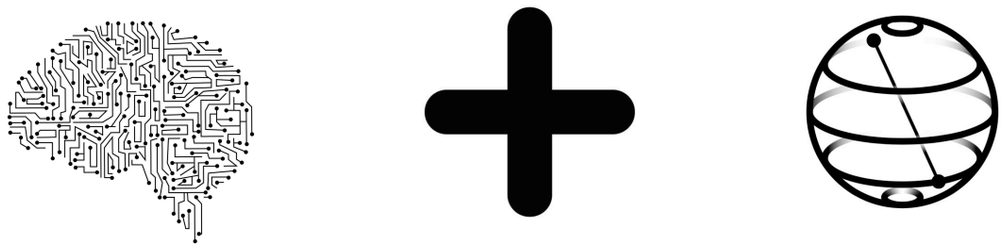
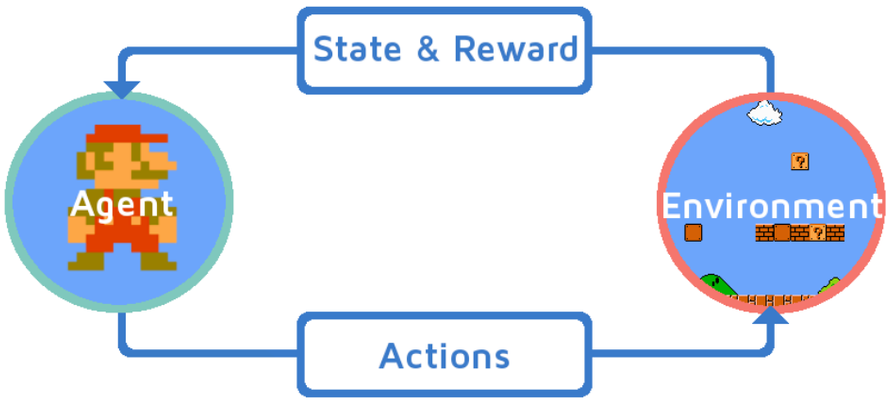
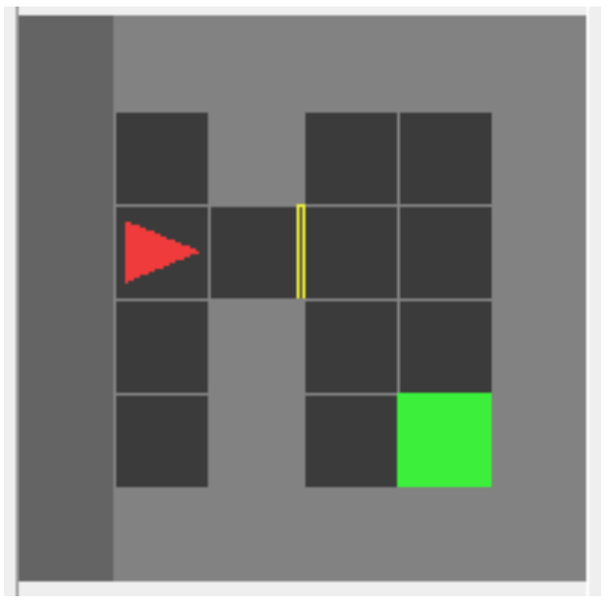
Quantum Reinforcement Learning



Alfredo Ibias, Ana Martín, Santiago Varona,
Pablo Barrio, Pablo Moreno

What's Reinforcement Learning?



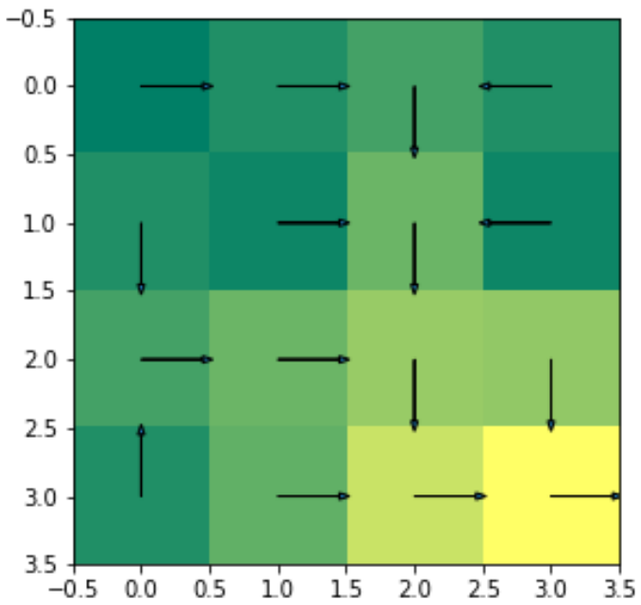


IBM Watson

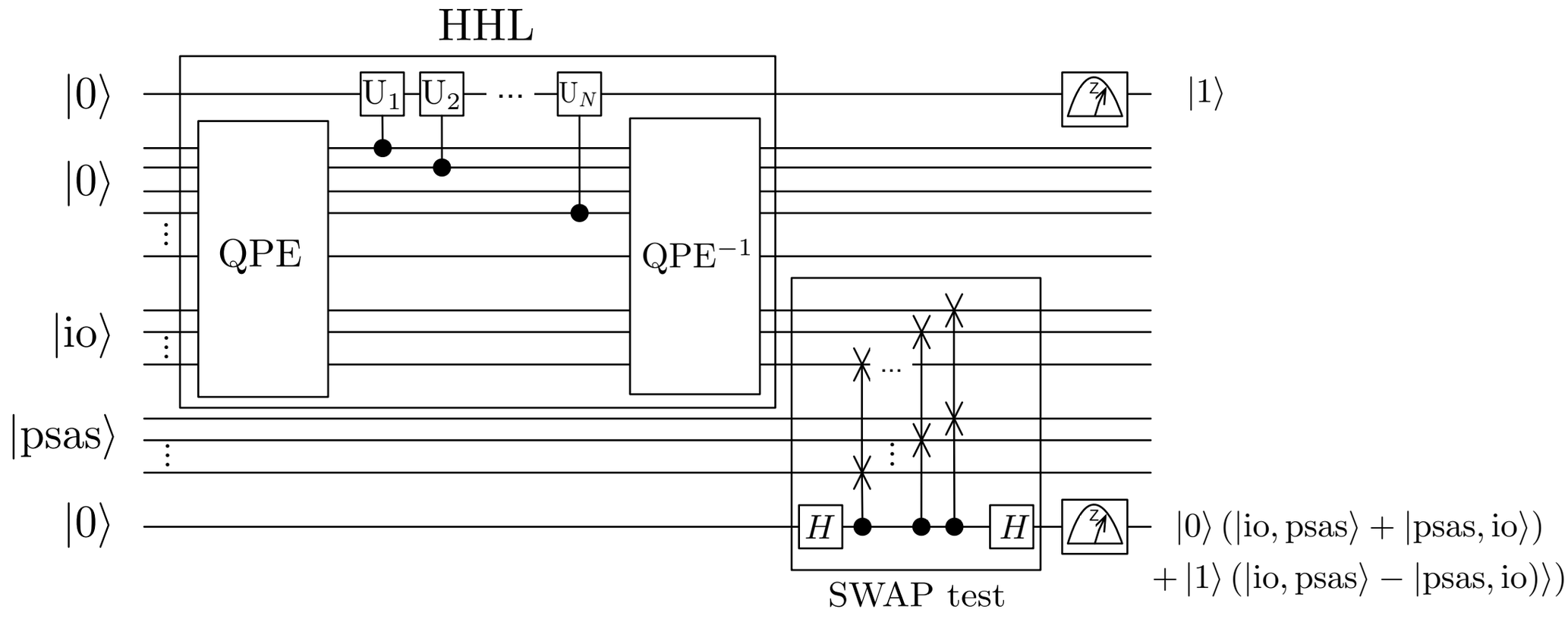
Policy Iteration

How to find the best policy?

Initialize a policy π' arbitrarily
Repeat
$\pi \leftarrow \pi'$
Compute the values using π by
solving the linear equations
$V^{\pi}(s) = E[r|s, \pi(s)] + \gamma \sum_{s' \in S} P(s'|s, \pi(s)) V^{\pi}(s')$
Improve the policy at each state
$\pi'(s) \leftarrow \arg\max_a (E[r|s, a] + \gamma \sum_{s' \in S} P(s'|s, a) V^{\pi}(s'))$
Until $\pi = \pi'$



Quantum Reinforcement Learning



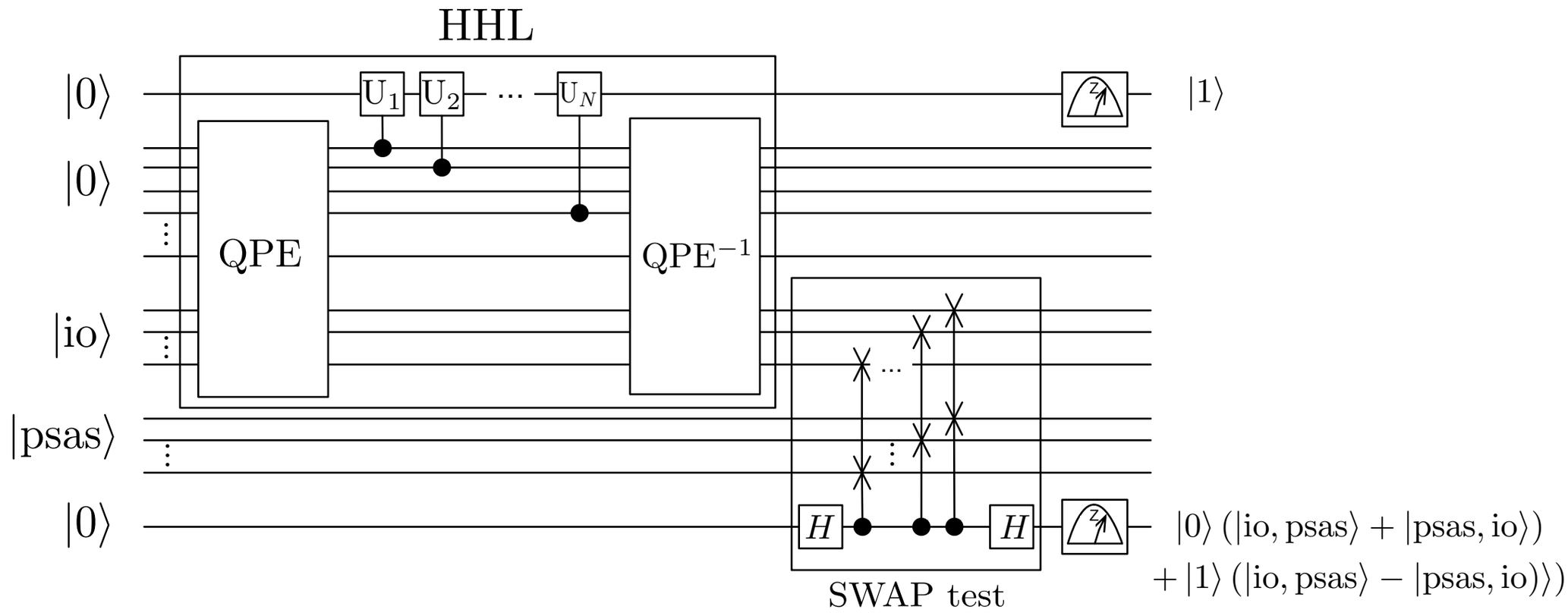
- Depth (original problem proposal): ~$3 \cdot 10^6$ gates. (20 time slices)

Quantum Reinforcement

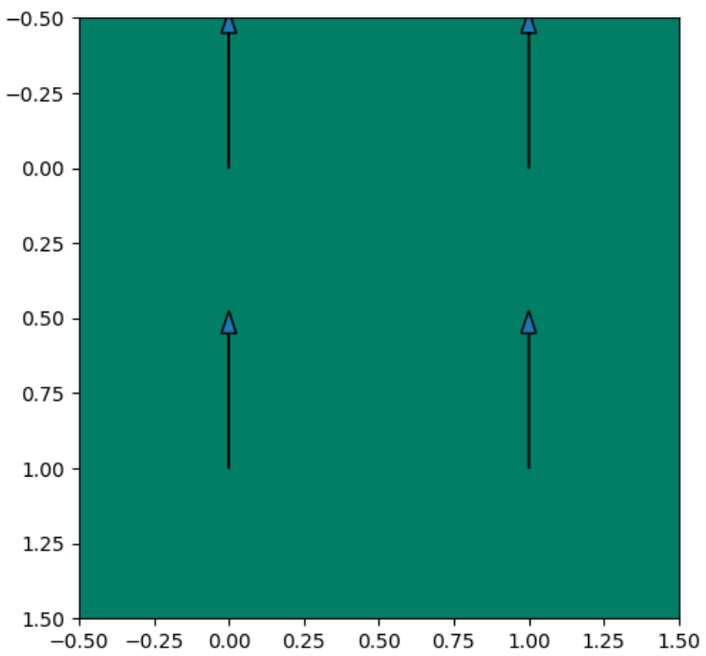
```
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Input matrix is not hermitian. It will be expanded to a hermitian matrix automatically.
Got a 413 code response to /api/Network/ibm-q/Groups/open/Projects/main/jobs: <html>
<head><title>413 Request Entity Too Large</title></head>
<body bgcolor="white">
<center><h1>413 Request Entity Too Large</h1></center>
<hr><center>nginx</center>
</body>
</html>
```

qiskit-terra issue #2518

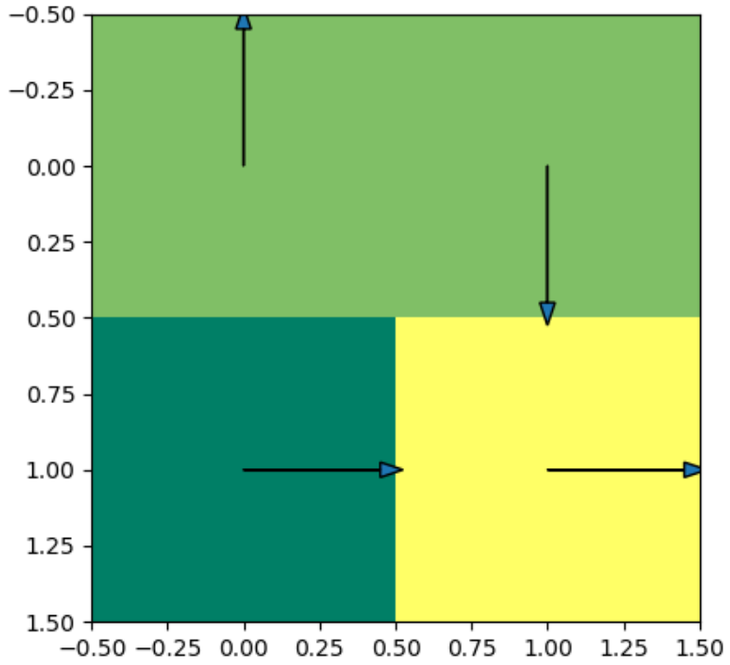
Quantum Reinforcement Learning



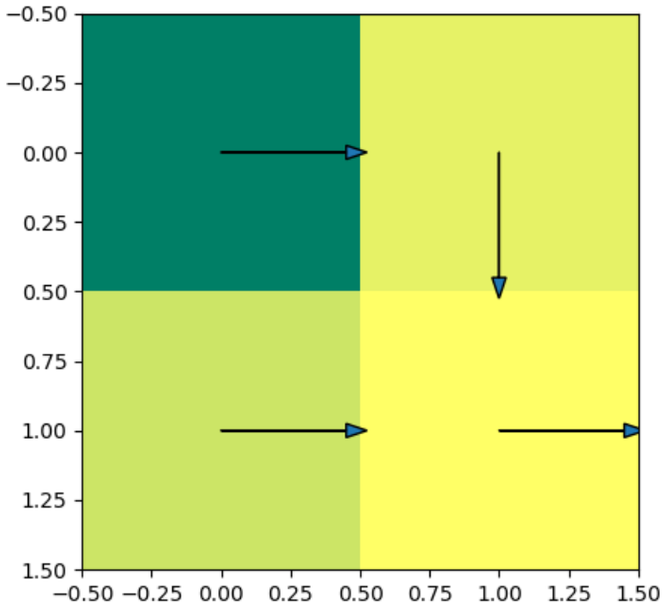
- Depth (original problem proposal): ~$3 \cdot 10^6$ gates gates, 14 qubits (20 time slices)
- Depth feasible proposal: ~158000 gates 10 qubits (20 time slices)



1st iteration



2nd iteration



3rd iteration

Quantum advantage

- Number of iterations of Policy Iteration: $\frac{m(n-m)}{1-\gamma}\log\left(\frac{m^2}{1-\gamma}\right)$
  - For m the number of states,
  - For a_i the number of actions in state i
  - $n = \sum_i a_i$

Complexity per iteration (using classical parallelism wherever possible):

$$O(s\kappa^2\epsilon^{-1}\mathrm{poly}\log(s,\kappa^2,\epsilon^{-1},n))$$

Yinyu Ye, (2011) The Simplex and Policy-Iteration Methods Are Strongly Polynomial for the Markov Decision Problem with a Fixed Discount Rate. Mathematics of Operations Research 36(4):593-603. http://dx.doi.org/10.1287/moor.1110.0516

Conclusions & future work

- We intend to make future **benchmarks** to **prove the quantum advantage.**

- We intend to **publish** the results, either with our algorithm alone or quantising others