Quantum Variational Reinforcement Learning

André Sequeira

24 February 2021

INTERNATIONAL IBERIAN
NANOTECHNOLOGY
LABORATORY

Outline

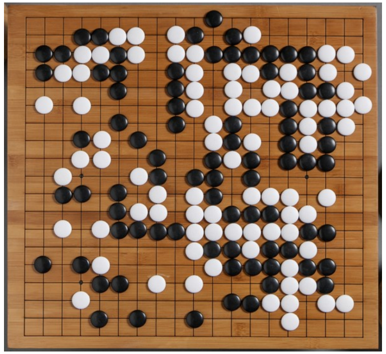
- Motivations
- Variational Quantum Algorithms
- Deep Reinforcement Learning
- Quantum Variational Reinforcement Learning

INTERNATIONAL IBERIAN
NANOTECHNOLOGY
LABORATORY

Motivations

Reinforcement Learning - 1980

Deep Learning + Reinforcement Learning = Deep Reinforcement Learning



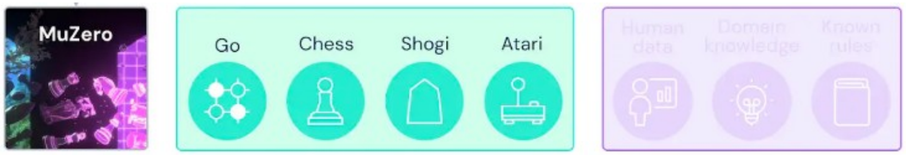
Mastering the game of Go without Human Knowledge, 2017

Dota 2 with Large Scale Deep Reinforcement Learning, 2019

INTERNATIONAL IBERIAN NANOTECHNOLOGY LABORATORY

What about
Hamiltonian Learning ?

Motivations



MuZero learns the rules of the game, allowing it to also master environments with unknown dynamics.
(Dec 2020, Nature)

Learning an accurate model of an environment's dynamics, and then use it to plan

- Self-driving cars – AWS deep racer
- Industry automation – Google DeepMind
- Trading and Finance – IBM

Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model, 2020

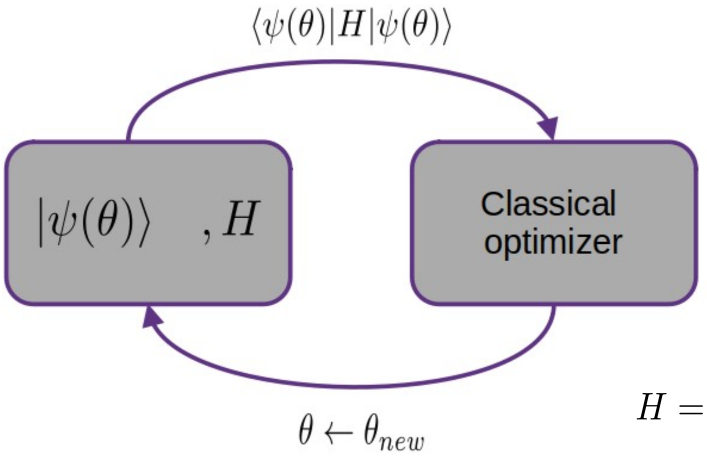
INTERNATIONAL IBERIAN NANOTECHNOLOGY LABORATORY

From Quantum Eigensolvers to Machine Learning models

INTERNATIONAL IBERIAN
NANOTECHNOLOGY
LABORATORY

Variational Quantum Eigensolver

→ Find ground state of Hamiltonian H



Cost-function: Expectation value of Hamiltonian (energy)

Variational principle: $\langle\psi(\theta)|H|\psi(\theta)\rangle \geq E_{min}$

$$\theta^* \leftarrow argmin_\theta \langle\psi(\theta)|H|\psi(\theta)\rangle$$

$$H = \sum_{i\alpha} h^i_\alpha \sigma^i_\alpha + \sum_{ij\alpha\beta} h^{ij}_{\alpha\beta} \sigma^i_\alpha \sigma^j_\beta + \dots$$

↓ *linearity*

$$\langle H\rangle = \sum_{i\alpha} h^i_\alpha \langle\sigma^i_\alpha\rangle + \sum_{ij\alpha\beta} h^{ij}_{\alpha\beta} \langle\sigma^i_\alpha \sigma^j_\beta\rangle + \dots$$

$$|\psi\rangle = U(\theta)|0\rangle$$

$U(\theta)$ → *Ansatz*

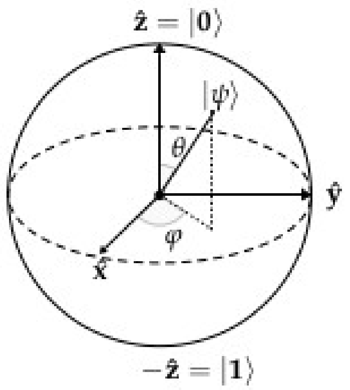
INTERNATIONAL IBERIAN NANOTECHNOLOGY LABORATORY

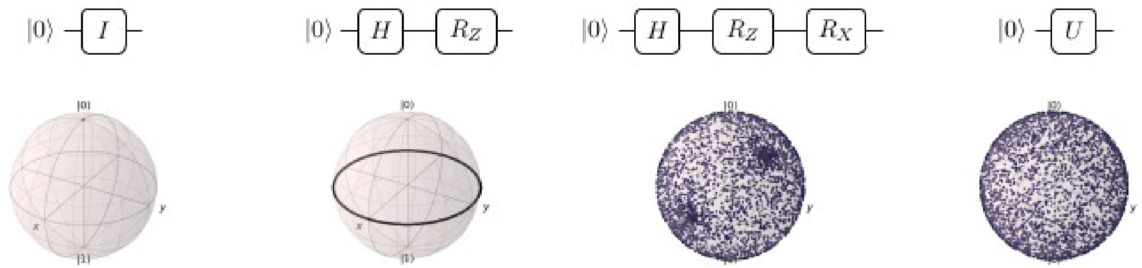
A variational eigenvalue solver on a quantum processor, 2013

Ansätze

→No consideration of the target domain – Ry , RyRz

→ Domain specific knowledge – Unitary Coupled Cluster Single Double UCCSD



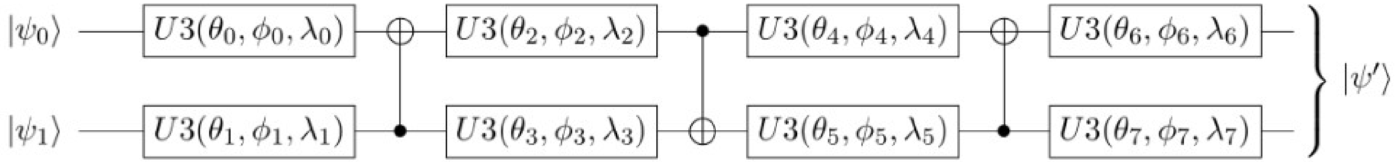


Expressibility and entangling capability of parametrised quantum circuits for hybrid quantum-classical algorithms, 2019

$$U = e^{i\Phi} R_z(\theta) R_y(\phi) R_z(\lambda)$$

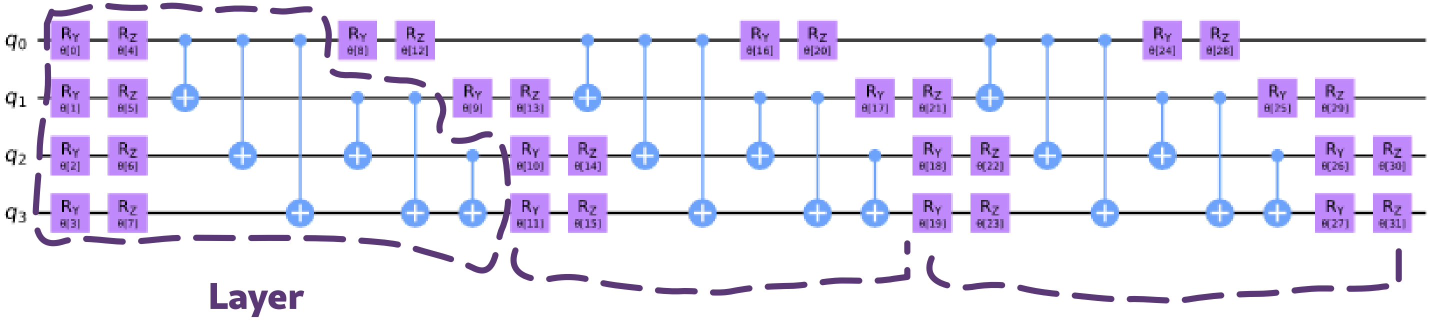
Ansätze

→2 qubit universality: two body interactions, and thus entanglement, must be considered.



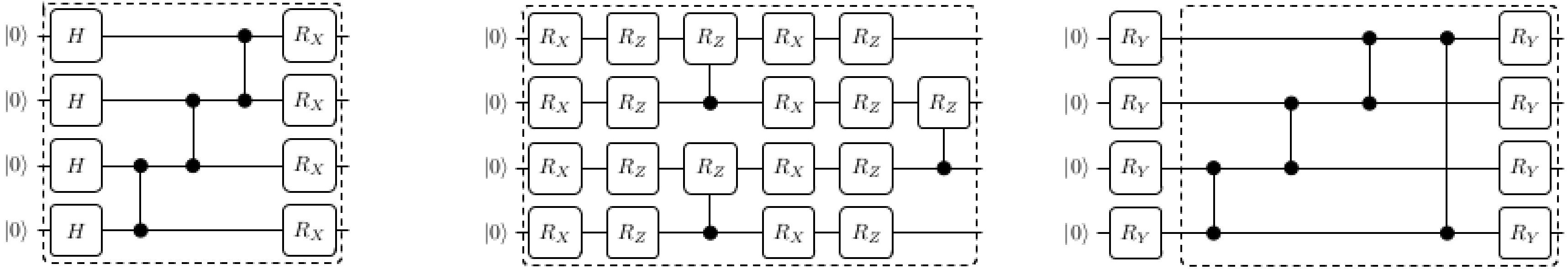
Minimal Universal Two-Qubit CNOT-based Circuits, 2003

→Full entanglement - Highly correlated states



INTERNATIONAL IBERIAN NANOTECHNOLOGY LABORATORY

Ansätze

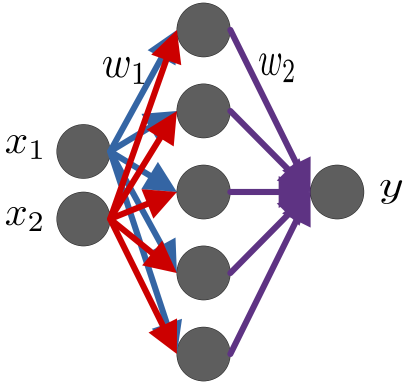


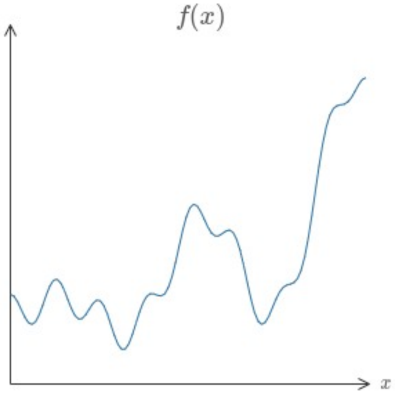
Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms, 2019

- More layers / entanglement may reduce expressibility
- Increase in depth and number of parameters → Optimization more challenging

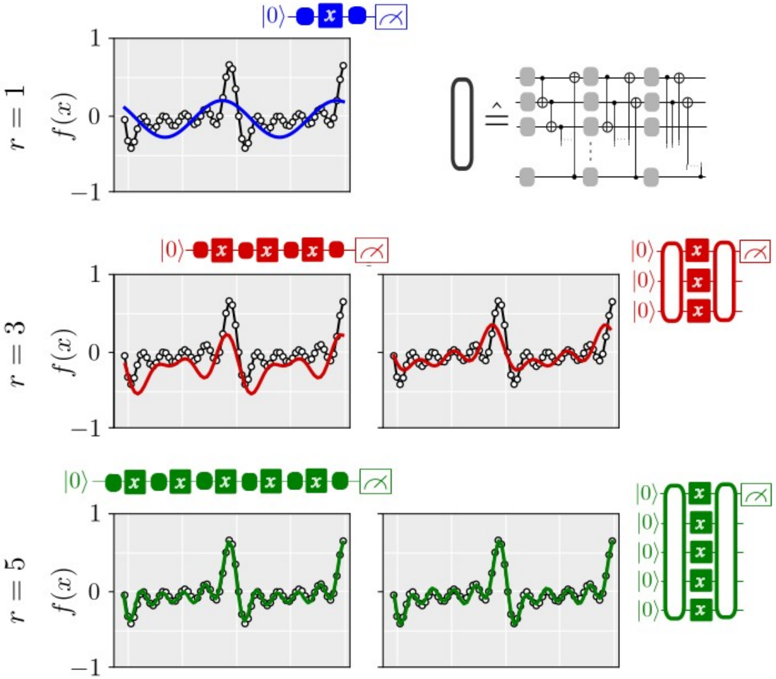
INTERNATIONAL IBERIAN NANOTECHNOLOGY LABORATORY

Universality





A visual proof that neural networks can compute any function, Michael Nielsen



The effect of data encoding on the expressive power of variational quantum machine learning models, 2020

INL INTERNATIONAL IBERIAN NANOTECHNOLOGY LABORATORY

VQC as a ML model

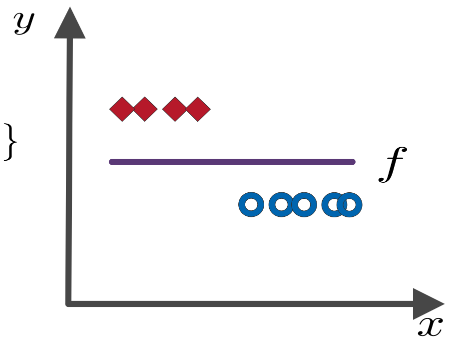
$x \longrightarrow$ $f(x, \theta)$ $\longrightarrow y$ $\longrightarrow$ $f \mapsto VQC$

→ State preparation routine
→ Objective function

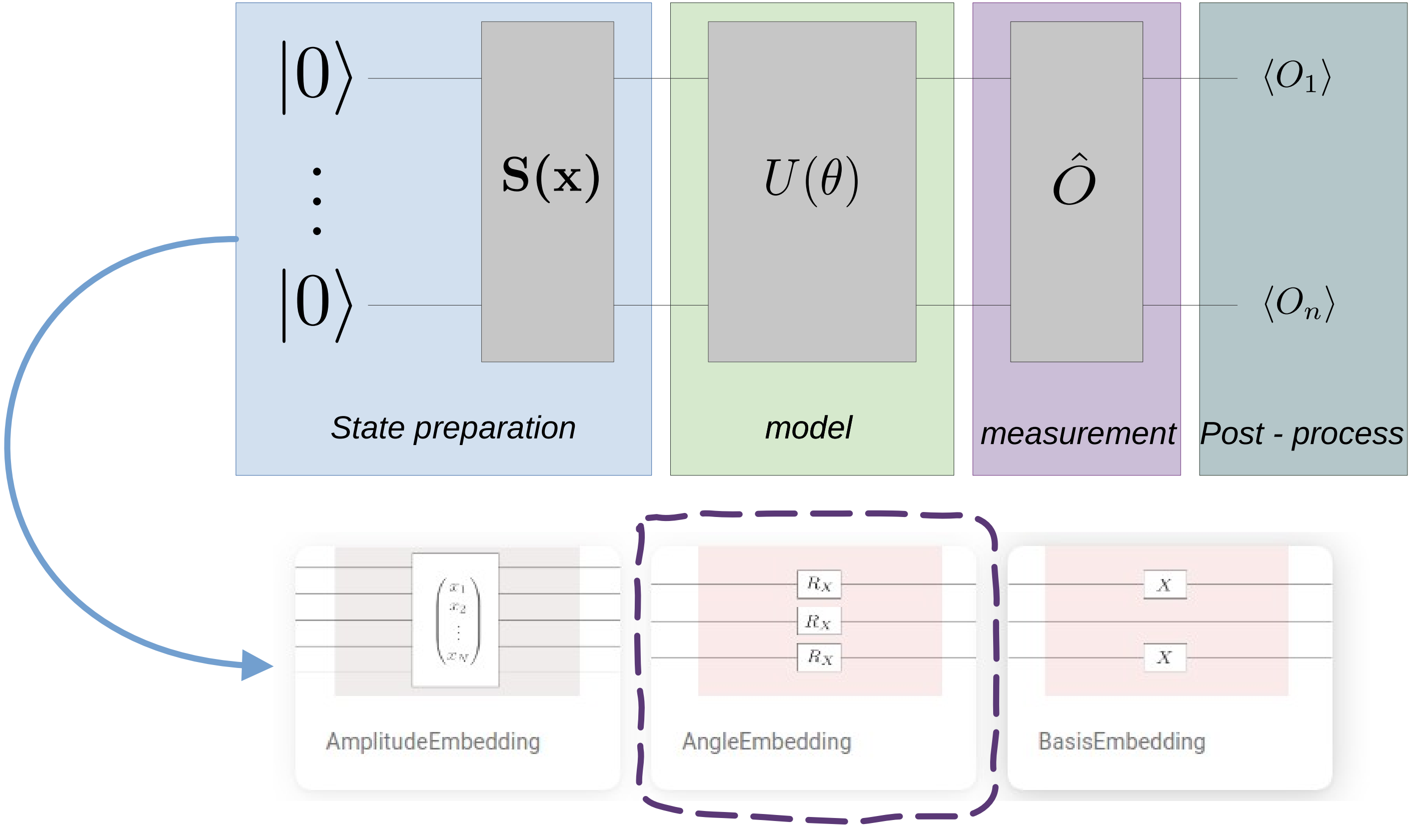
$$D = \{(x_0, y_0), (x_1, y_1), \ldots, (x_{N-1}, y_{N-1})\}$$

input

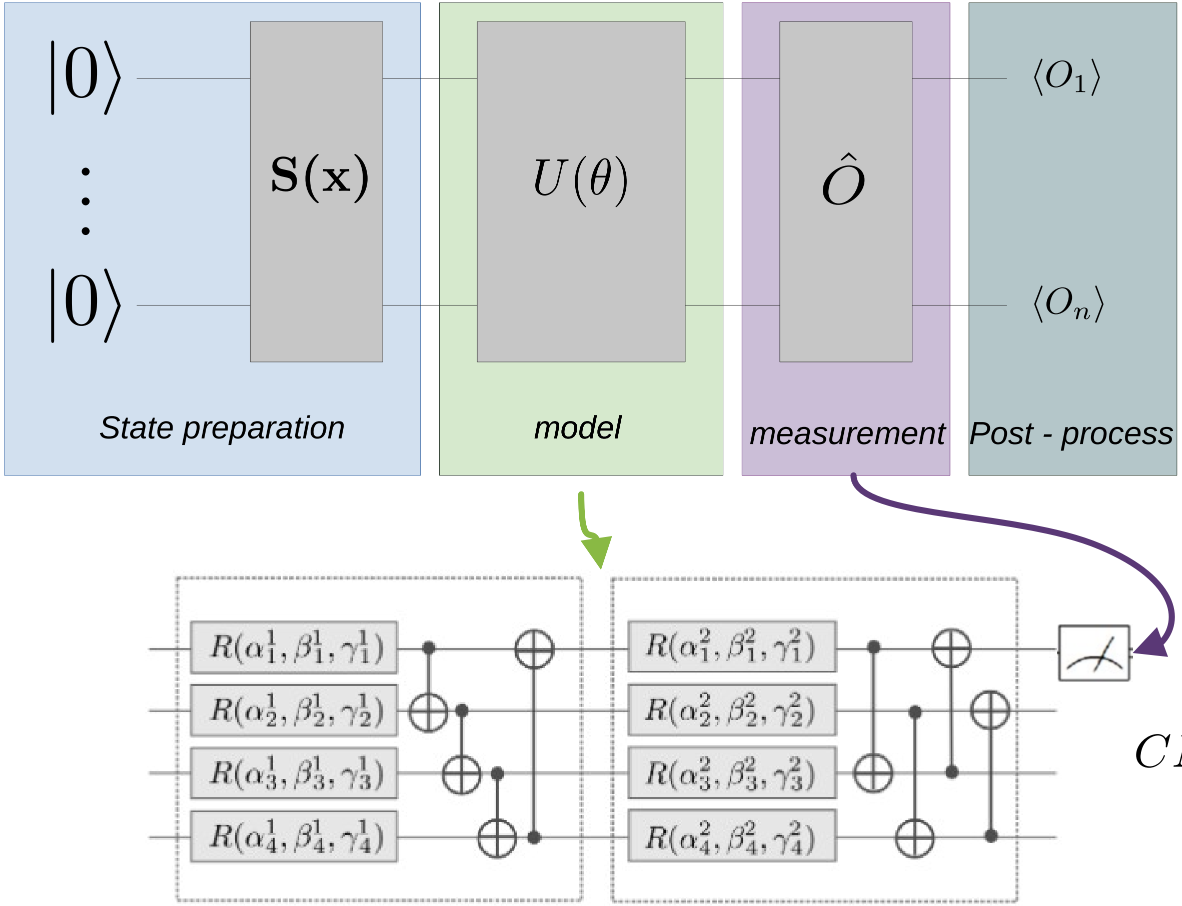
True label



$$L(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^{N-1} (y_i - \hat{y}_i)^2 = \frac{1}{N} \sum_{i=0}^{N-1} (y_i - f(x_i, \theta))^2 \longmapsto \min_{\theta} L(y, \hat{y})$$



$$e^{jx\hat{\sigma}_x} = cos(x)|0\rangle - isin(x)|1\rangle$$



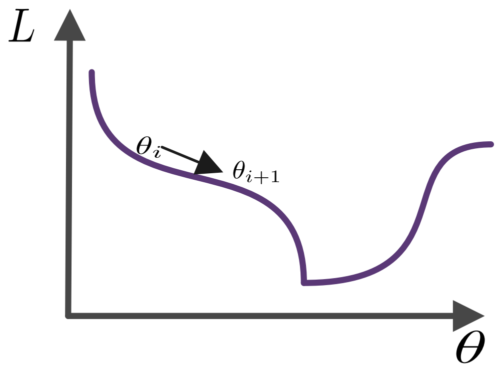
$$CNOT\,(i,(i+l)modN)$$

Strongly Entangling Layers

Circuit-centric quantum classifiers, 2018

$$\langle\hat{\sigma}_z\rangle = \langle\psi|\hat{\sigma}_z|\psi\rangle = \langle 0|S^\dagger U(\theta)^\dagger\hat{\sigma}_z U(\theta)S|0\rangle$$

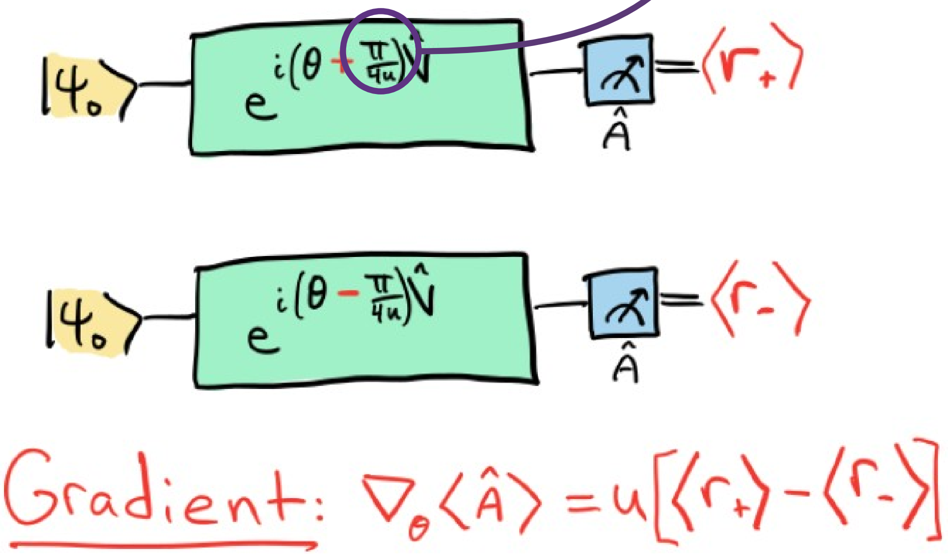
Quantum Gradients



Single qubit gates → $\frac{\pi}{2}$

Gradient-based optimization: $\theta_{i+1} \leftarrow \theta_i - \eta \nabla_\theta L(\theta)$

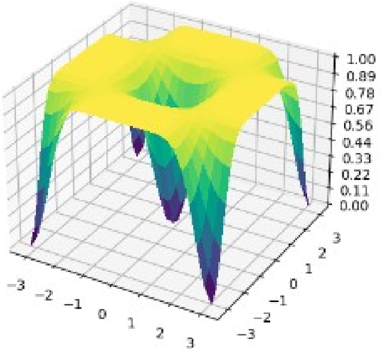
$$\nabla_\theta \langle \hat{A} \rangle = u \left[\langle \hat{A}(\theta + \tfrac{\pi}{4u}) \rangle - \langle \hat{A}(\theta - \tfrac{\pi}{4u}) \rangle \right]$$

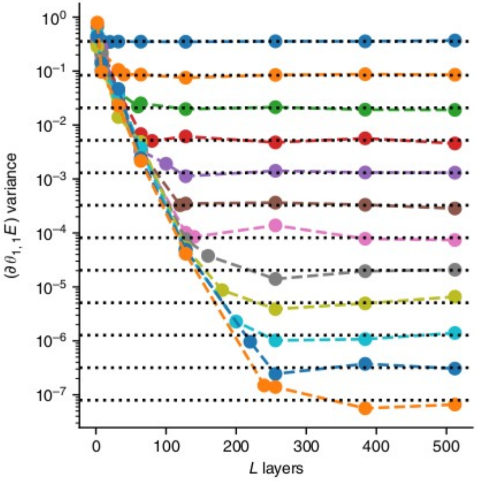


Pennylane.ai

INTERNATIONAL IBERIAN NANOTECHNOLOGY LABORATORY

Loss-function landscape





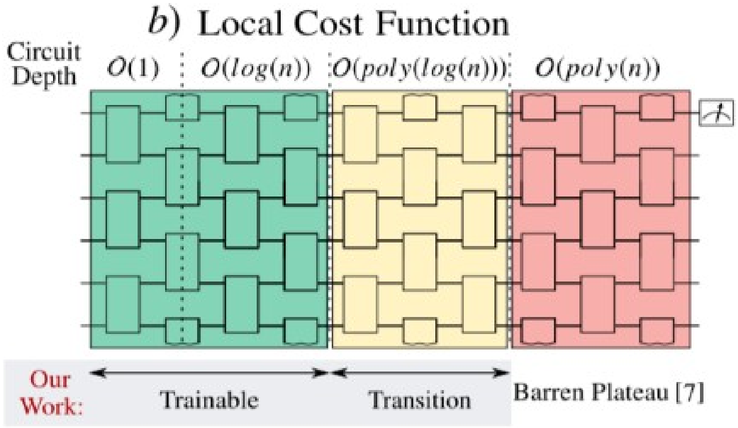
Barren Plateaus

→ Classic deep networks – Gradient vanish exponentially with number of layers

→ **Randomly initialised VQC's** – Gradient vanish exponentially with number of qubits

Towards mitigation:

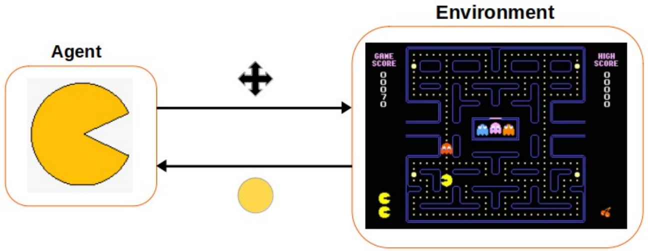
- structured initial guesses – UCCSD
- pre-training segment by segment
- Local cost-functions



Barren plateaus in quantum neural network training landscapes, 2018

Cost-Function-Dependent Barren Plateaus in Shallow Quantum Neural Networks,2020

Deep Reinforcement Learning



RL AGENTS GOAL: find optimal policy π^*

$$G = R_0 + \gamma R_1 + \gamma^2 R_2 + \cdots + \gamma^{h-1} R_{h-1} = \sum_{t=0} \gamma^t R_t$$

Agent	Policy	$\pi : S \mapsto A$
Environment	MDP	$\langle S, A, P, R, \gamma \rangle$

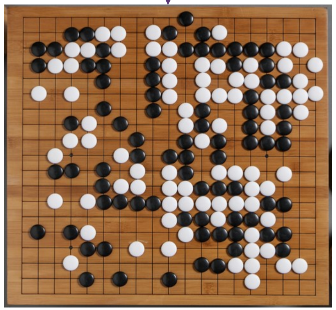
	s_1	s_2	s_3	s_4	s_5
(s,a)	0.3	0	0	0.7	0

(a) Stochastic MDP

	s_1	s_2	s_3	s_4	s_5
(s,a)	0	0	0	1	0

(b) Deterministic MDP

$Q(s,a), \forall s \in S, \forall a \in A$



MDP: Fully observable (MDP) VS Partially Observable (POMDP)

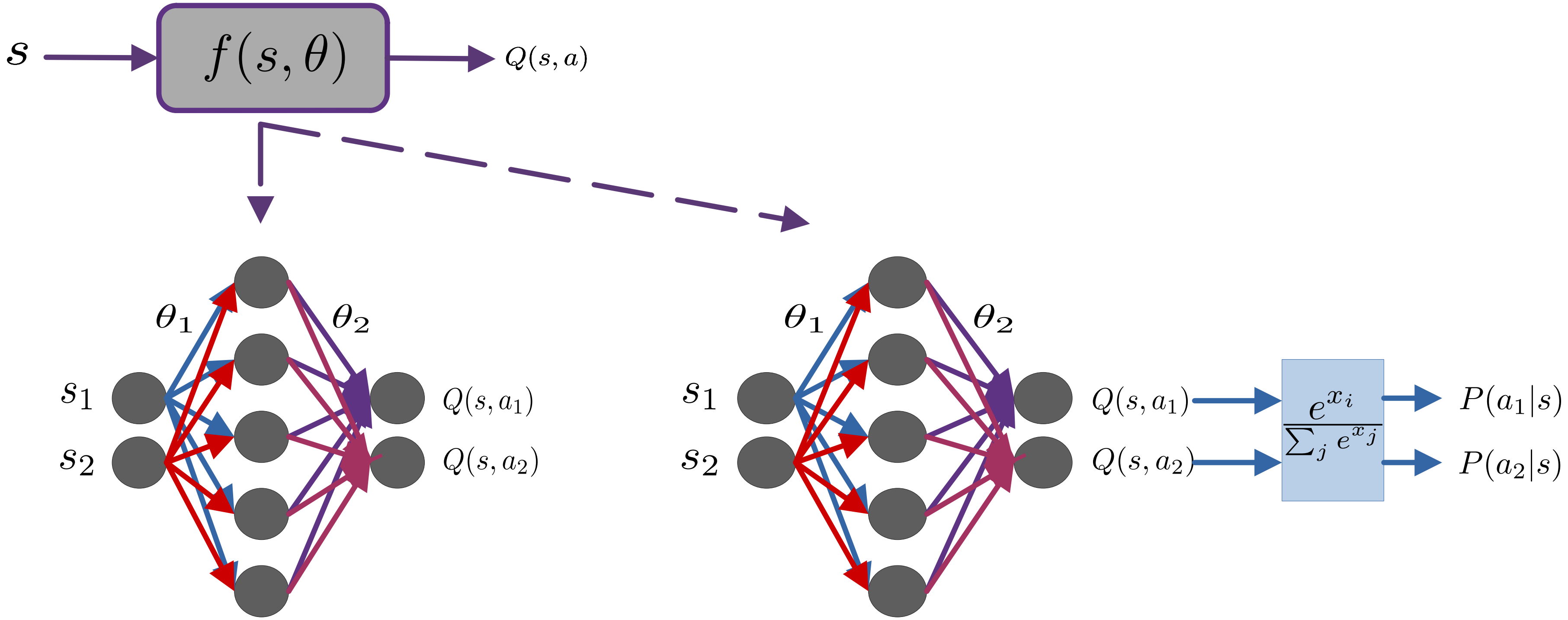
Solving the MDP for the optimal policy π^*:

Model-based RL:
The agent knows the dynamics of the environment (P,R)
Dynamic Programming - Value Iteration / Policy Iteration

Model-Free RL:
Unknown dynamics - Resort to Sampling techniques
Exploration-Exploitation dillema
MC Learning , TD-Learning (SARSA, Q-Learning)

Introduction to Reinforcement Learning, Sutton and Barto, 2018

INTERNATIONAL IBERIAN NANOTECHNOLOGY LABORATORY

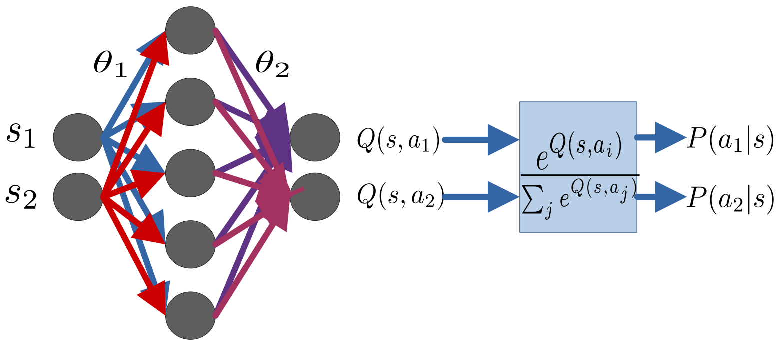


ϵ - greedy

Policy Network

$$\pi(a|s) = \begin{cases} argmax_{a\in A} Q(s,a) \quad with \quad probability \quad 1-\epsilon \\ random(A) \quad with \quad probability \quad \epsilon \end{cases}$$

Q-Network



$$\tau = \{(s_0, a_0, R_0), (s_1, a_1, R_1), \ldots, (s_{T-1}, a_{T-1}, R_{T-1})\}$$

$$G(\tau) = \sum_{t=0}^{T} \gamma^t R_t$$

Gradient-based optimisation

$$\theta_{i+1} \leftarrow \theta_i - \eta \nabla_\theta L(\theta)$$

Vanilla Policy Gradient

$$\nabla_\theta L(\theta) = \mathbb{E}_{\tau \sim \pi_\theta}\left[\sum_{t=0}^{T} G(\tau) \nabla_\theta log \pi_\theta(a_t|s_t)\right]$$

Same score independent of the action

REINFORCE

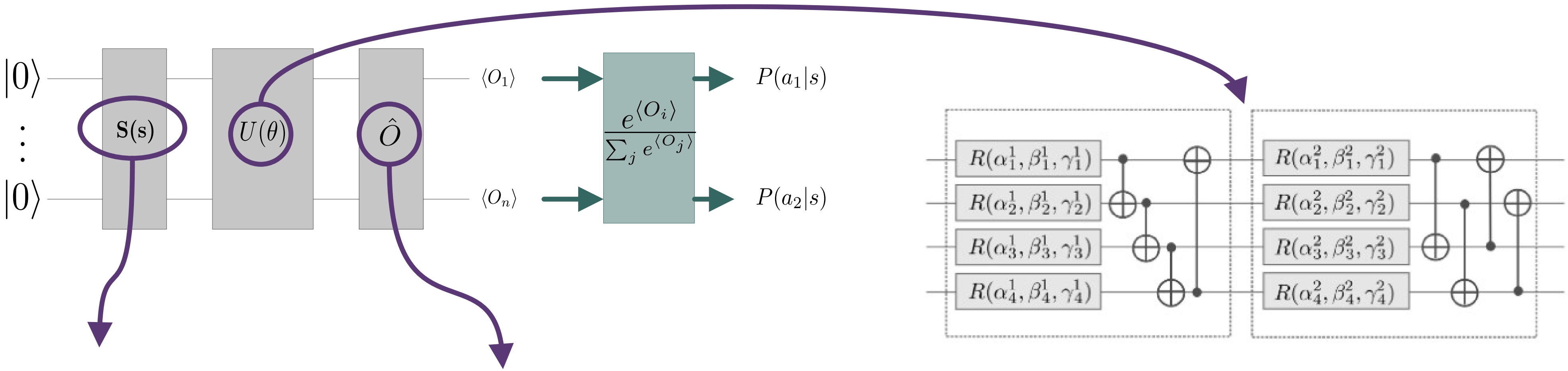
$$\nabla_\theta L(\theta) = \mathbb{E}_{\tau \sim \pi_\theta}\left[\sum_{t=0}^{T} G_t(\tau) \nabla_\theta log \pi_\theta(a_t|s_t)\right]$$

Rank different actions

Quantum Variational Reinforcement Learning

Comparison with State of the art

- First quantum hybrid Policy Network algorithm
- Application to continuous state-space problems



Problem dependent

Computational basis measurements

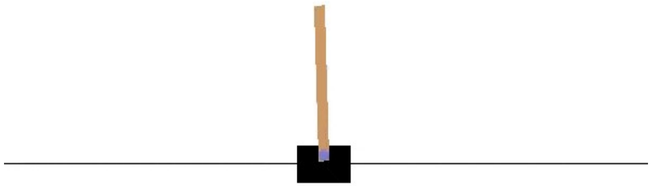
$$\hat{O} = \hat{\sigma}_z$$

Strongly Entangling Layers

$$\nabla_\theta L(\theta) = \mathbb{E}_{\tau\sim\pi_\theta}\left[\sum_{t=0}^{T} G_t(\tau) log\pi_\theta(a_t|s_t)\right]$$

$$\frac{e^{\langle O_i \rangle}}{\sum_j e^{\langle O_j \rangle}}$$

OpenAI Cartpole environment

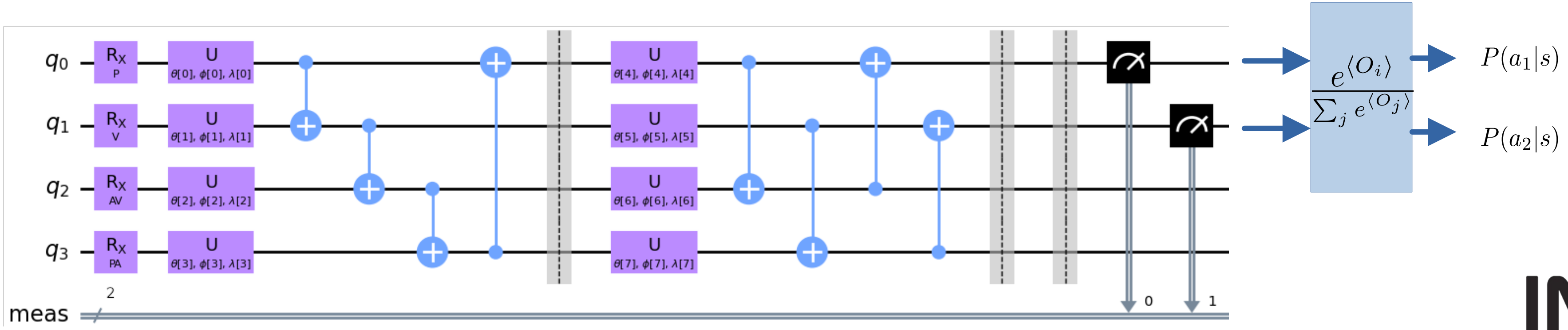


State-Space:
- Position - $[-4.8, 4.8]$
- Velocity - $[-\infty, \infty]$
- Pole angle $[-0.418rad, 0.418rad]$
- Pole angular velocity - $[-\infty, \infty]$

Each time-step receive +1 reward

Goal: score +195 reward for 100 consecutive episodes

Action-space:
Move left – 0
Move right – 1



It does not succeed !

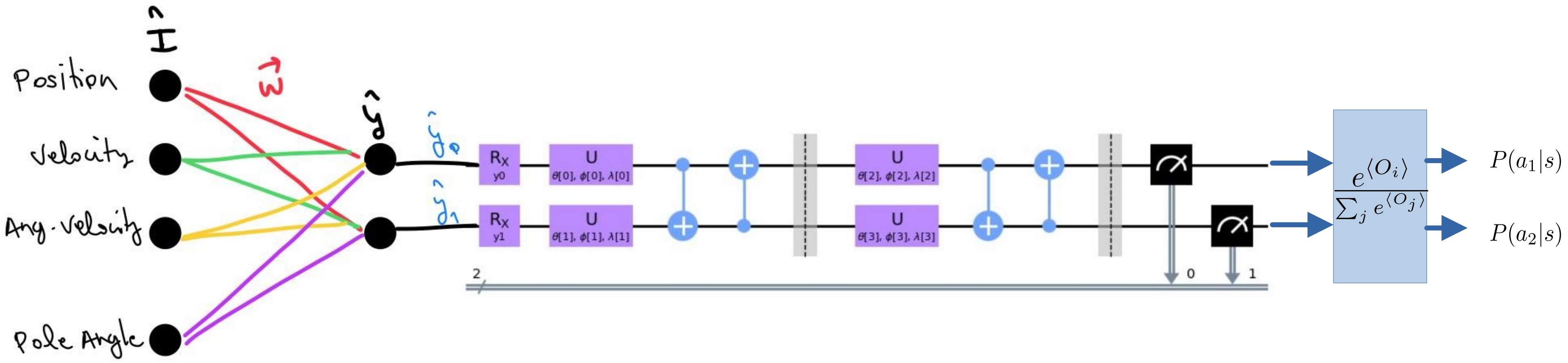
INTERNATIONAL IBERIAN NANOTECHNOLOGY LABORATORY

Square one

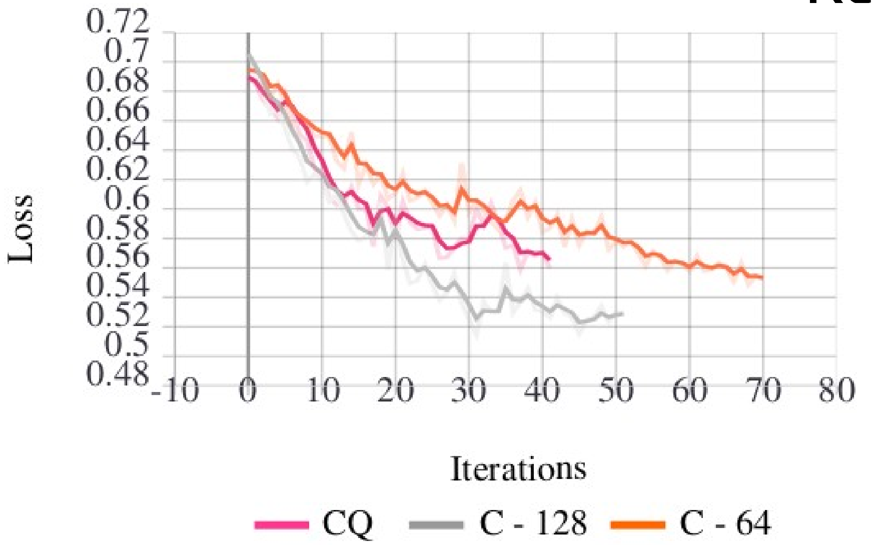
- Change entanglement pattern – Linear, ring…
- Change arbitrary rotation – Ry
- Correlate pole angle/velocity position/velocity ...
- Measure all qubits
- Learning rate / optimizer
- Amplitude Encoding

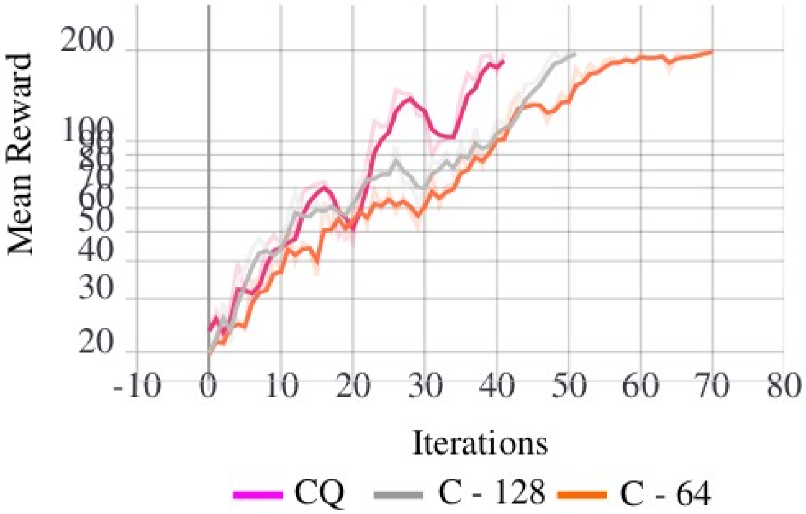
Hybrid approach

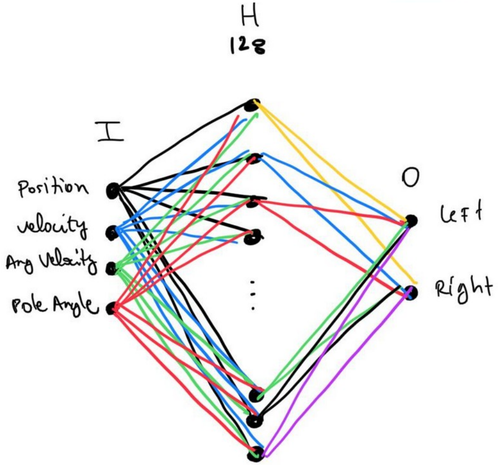
Dimensionality reduction by classic NN



Results







- On average, hybrid nn behaves similarly to the usual 128 hidden-layer classic network

Measuring advantage

INTERNATIONAL IBERIAN NANOTECHNOLOGY LABORATORY

of parameters trained

Classic:

I, input: size 4
H, hidden – layer: 128 neurons
O, output: size 2

$$|C| = I * H + H * O = 4 * 128 + 128 + 2 = 768$$

Hybrid:

classic:

Ic, input: size 4
Oc, output: size 2

quantum:

I, input: 2
L, layers: 3
P, parameters per layer: 3

$$|H| = Ic * Oc + I * L * P = 26$$

$$\frac{|C|}{|H|} \approx 30$$

Future work & open questions

- Apply Hybrid scheme to different problems

- Full quantum approach ?

- Measuring capacity of quantum models ?

INTERNATIONAL IBERIAN NANOTECHNOLOGY LABORATORY

Quantum Variational Reinforcement Learning

André Sequeira

24 February 2021

INTERNATIONAL IBERIAN
NANOTECHNOLOGY
LABORATORY