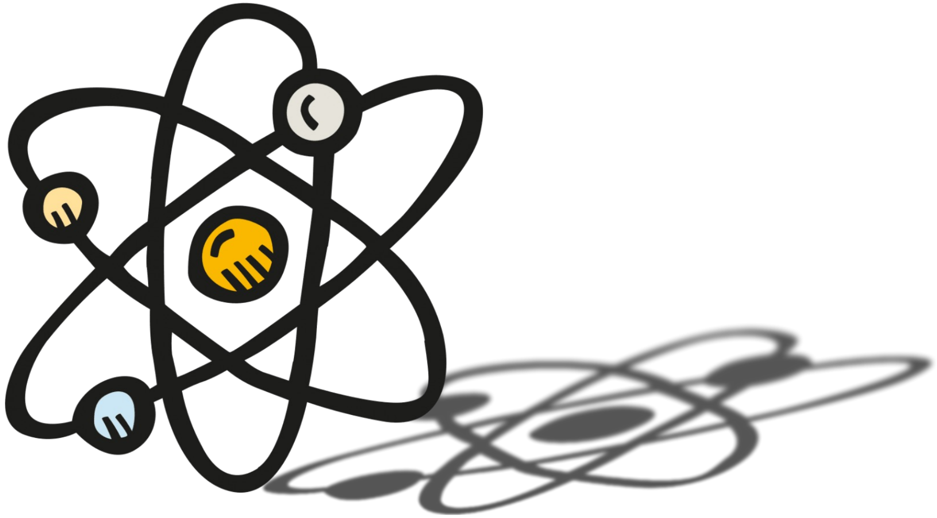
Shadows of Quantum Machine Learning

Sofiene Jerbi, Casper Gyurik, Simon Marshall, Riccardo Molteni, Vedran Dunjko

arXiv:2306.00061



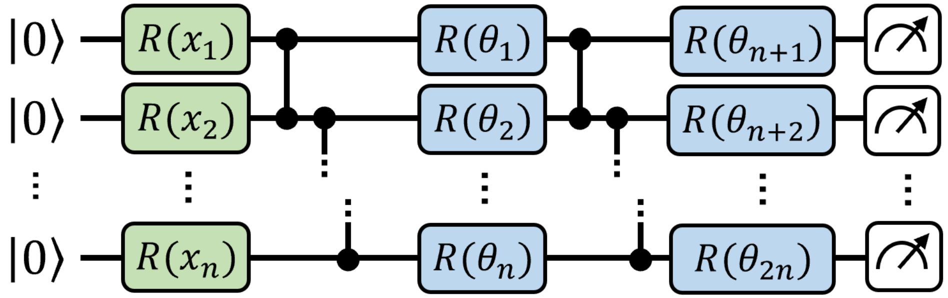
universität innsbruck

Freie Universität Berlin

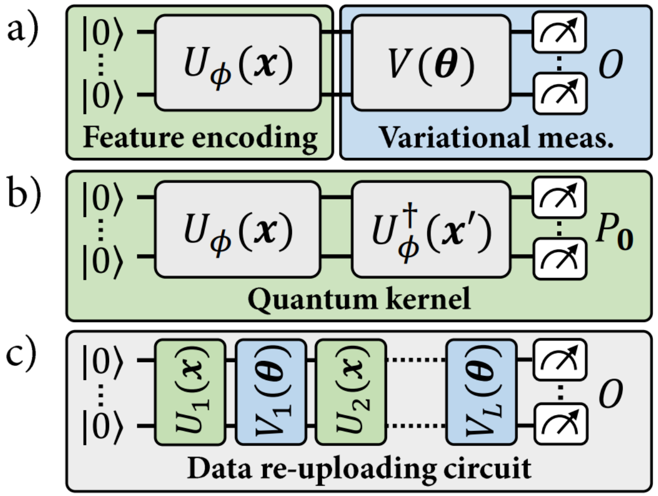
Universiteit Leiden

QTML 2023
22/11/2023

Quantum machine learning models



Various models:



Various names:

Parametrized quantum circuits

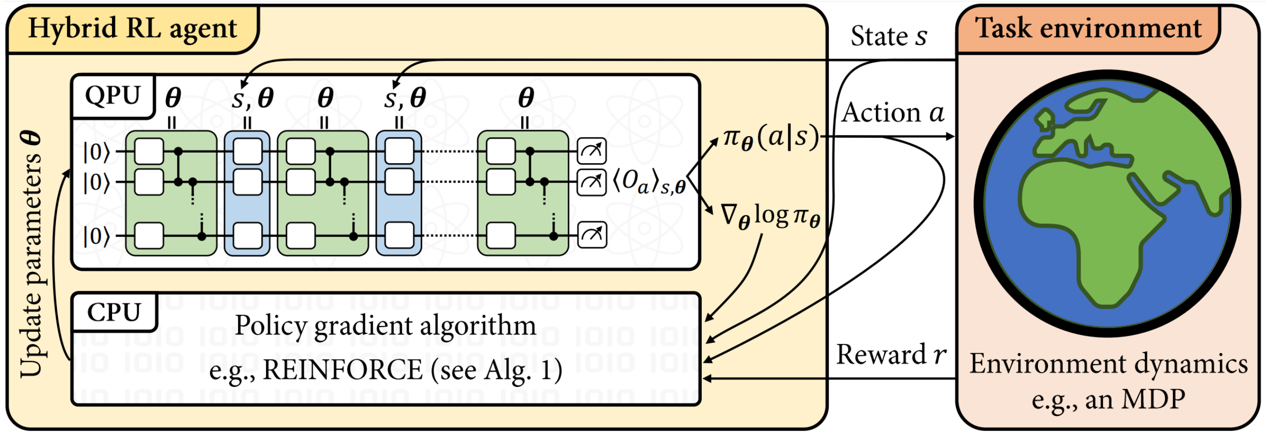
Variational quantum circuits

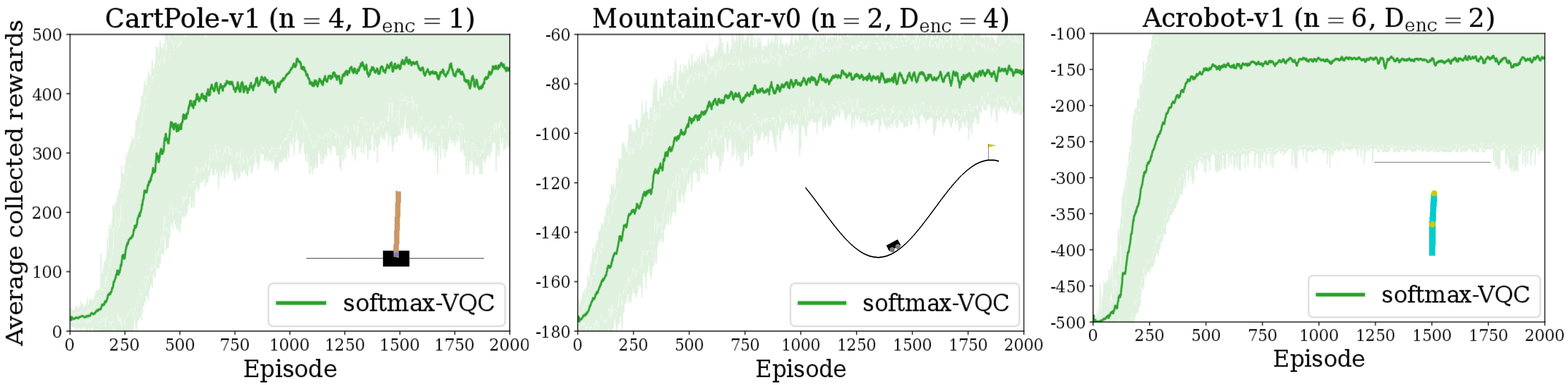
Quantum neural networks

Various applications:

- Classification:
  - ✓ Havlíček *et al., Nature* 2019
  - ✓ Schuld *et al., PRL* 2019
- Regression:
  - ✓ Mitarai *et al., PRA* 2018
- Generative modeling:
  - ✓ Liu *et al., PRA* 2018
- Reinforcement learning:
  - ✓ Chen *et al., IEEE* 2020
  - ✓ Jerbi *et al., NeurIPS* 2021

Promising empirical performance

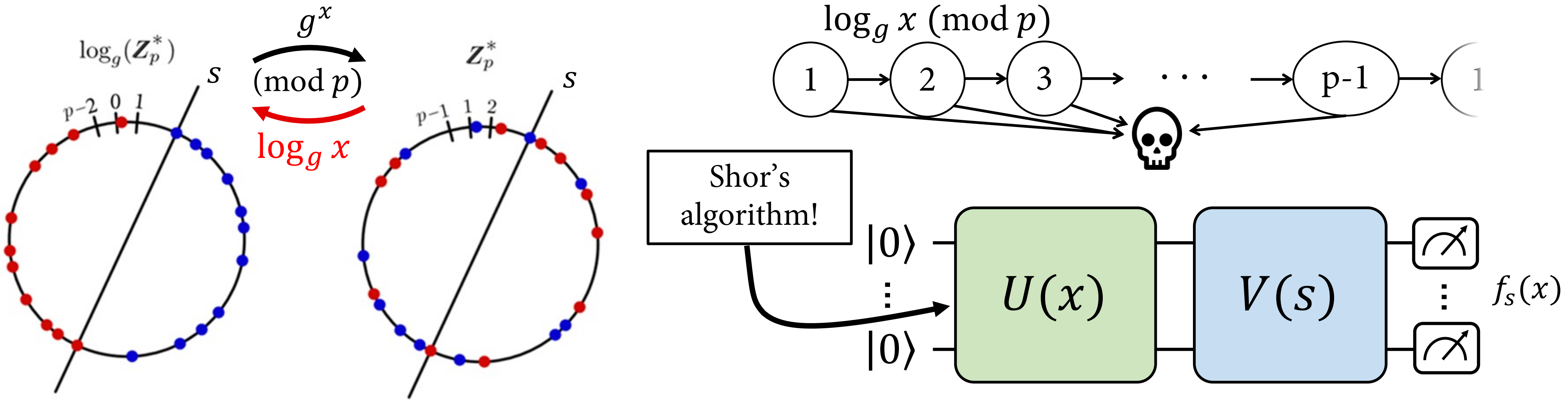




Jerbi, Gyurik, Marshall, Briegel & Dunjko, **Parametrized quantum policies for reinforcement learning**, NeurIPS 2021

Provable quantum advantage

- Learning tasks with a **provable** quantum advantage over any classical learner:



Theorem (informal). There exist learning tasks where:

1. Quantum agents can achieve close-to-optimal performance with high prob.
2. Classical agents cannot achieve a performance (much) better than random, under hardness of $\log_g$

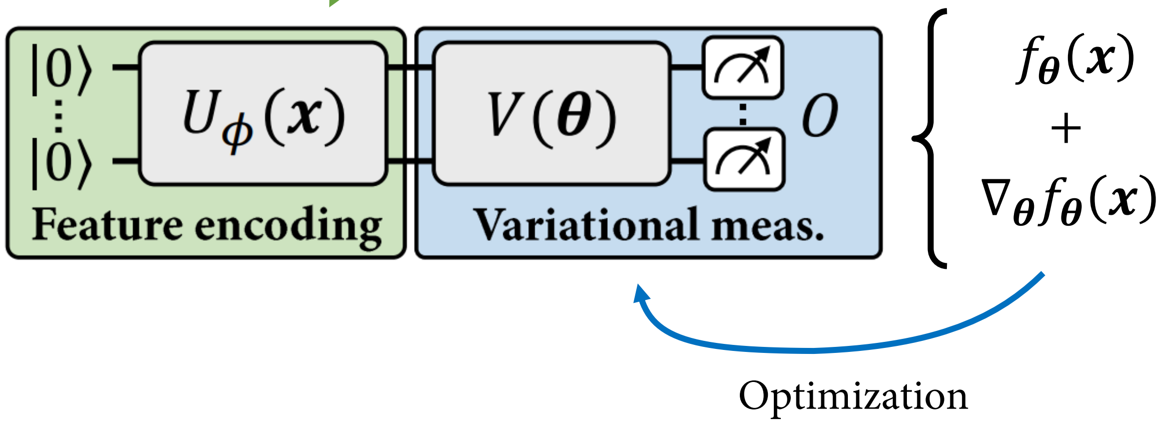
Liu, Arunachalam & Temme, **A rigorous and robust quantum speed-up in supervised machine learning**, Nature Physics (2021)

Making use of QML

Training phase

Given data $\mathcal{D} = \left\{\left(\boldsymbol{x}^{(1)}, g\left(\boldsymbol{x}^{(1)}\right)\right), \ldots, \left(\boldsymbol{x}^{(M)}, g\left(\boldsymbol{x}^{(M)}\right)\right)\right\}$,

fit the model $f_{\boldsymbol{\theta}}$ to the function g



Outputs a trained model $f_{\boldsymbol{\theta}}$

Deployment phase

Evaluate the model $f_{\boldsymbol{\theta}}$ on new data points $\boldsymbol{x}$

Problem: evaluation still needs a quantum computer!

Shadow models

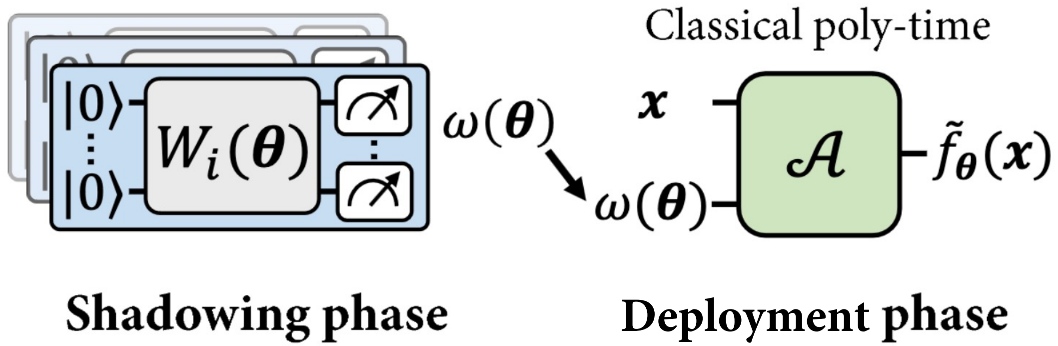
Main question:

Can we design QML models that are **trained** on a **quantum computer**, but,

after data collection (or "shadowing") phase,

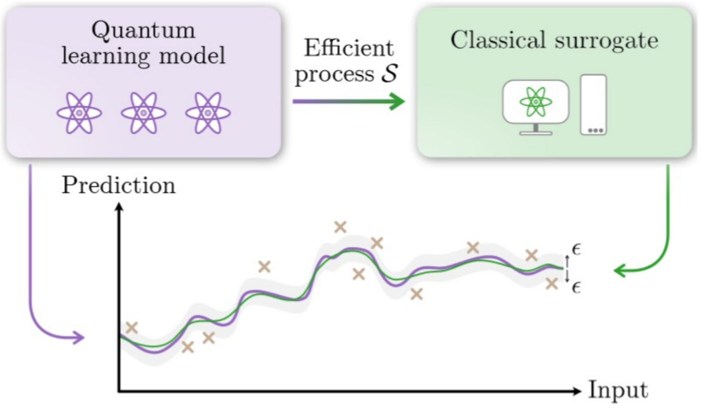
can later be **evaluated** on new data **classically**?

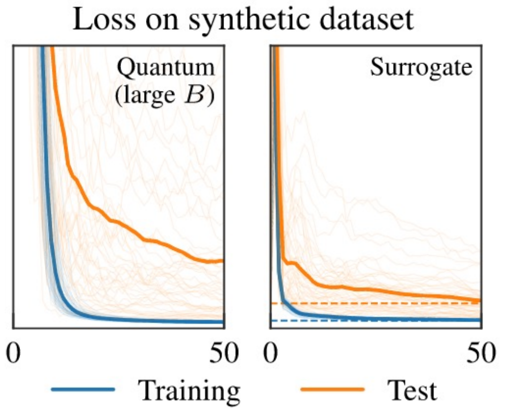
After training phase:



Classical surrogates

An existing proposal:





Based on Fourier representation of quantum models:

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \sum_{\boldsymbol{\omega}\in\Omega} c_{\boldsymbol{\omega}}(\boldsymbol{\theta}) e^{-\mathrm{i}\boldsymbol{\omega}\cdot\boldsymbol{x}}, \qquad \boldsymbol{x} \in \mathbb{R}^d$$

Simply learn the coefficients $c_{\boldsymbol{\omega}}(\boldsymbol{\theta})$!

Max frequency

Sample complexity: $\tilde{O}\left(\frac{{\omega_M}^d \, \|O\|_\infty^2}{\varepsilon^2}\right)$

To guarantee: $\left|\tilde{f}_{\boldsymbol{\theta}}(\boldsymbol{x}) - f_{\boldsymbol{\theta}}(\boldsymbol{x})\right| \leq \varepsilon \quad \forall \boldsymbol{x} \in \mathbb{R}^d$

$\boxed{\text{But also suggests that surrogate trained directly can outperform!}}$

Schreiber, Eisert & Meyer, **Classical surrogates for quantum learning models**. PRL (2023)

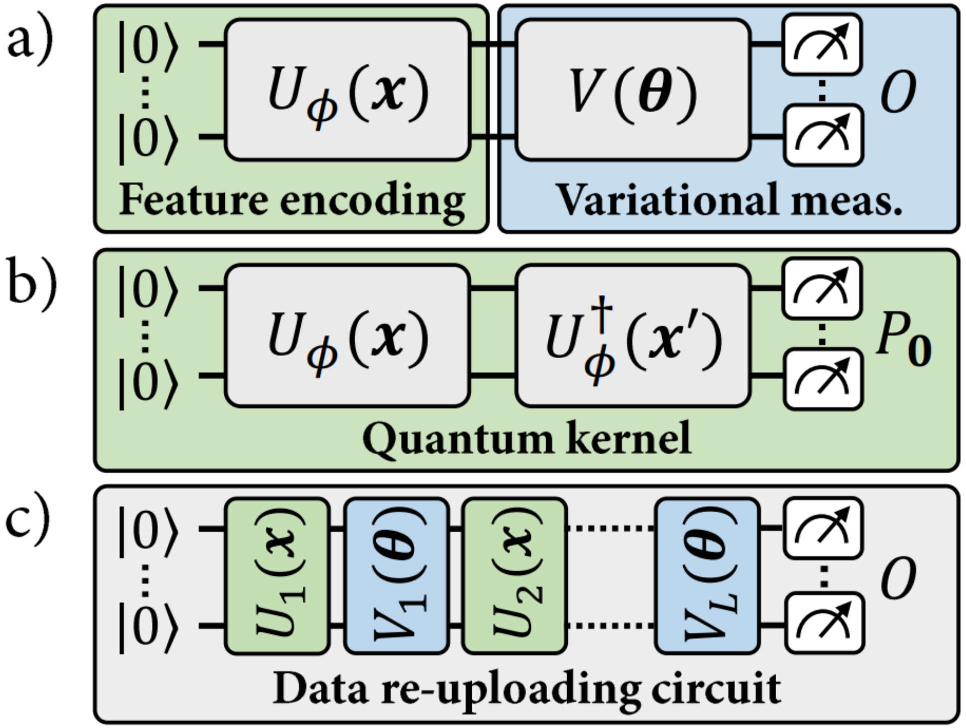
Shadow models

Main questions (revisited):

Q1. Can shadow models achieve a **quantum advantage** over entirely classical (classically trained and classically evaluated) models?

Q2. Do there exist quantum models that **do not admit shadow models**?

QML models are linear models



$f_{\boldsymbol{\theta}}(\boldsymbol{x})$

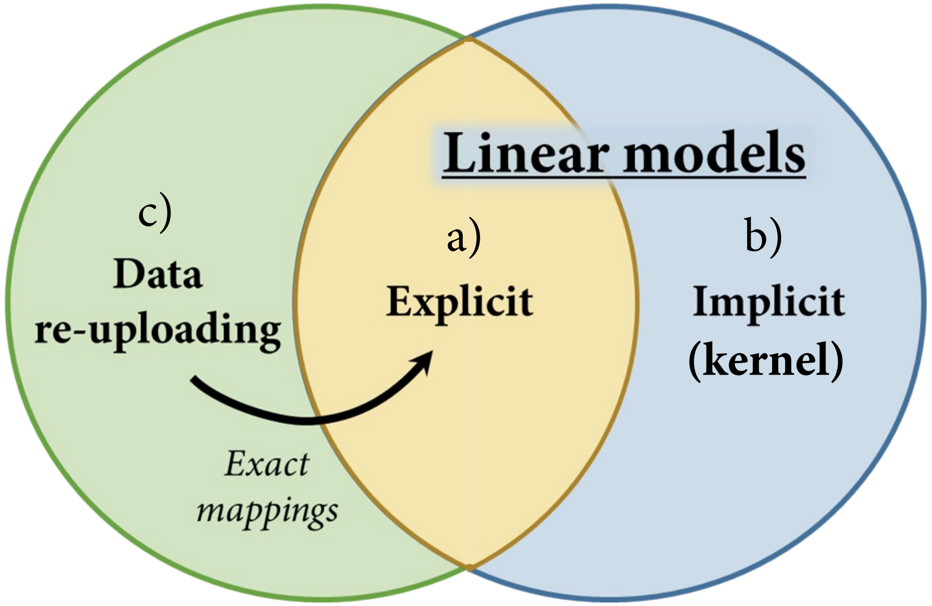
$=$

$\mathrm{Tr}[\rho(\boldsymbol{x})O(\boldsymbol{\theta})]$

$\mathrm{Tr}[\rho(\boldsymbol{x})O_{\boldsymbol{\alpha}}]$

$\sum_m \alpha_m \rho(\boldsymbol{x}^{(m)})$

$\mathrm{Tr}[\rho(\boldsymbol{x},\boldsymbol{\theta})O(\boldsymbol{\theta})]$

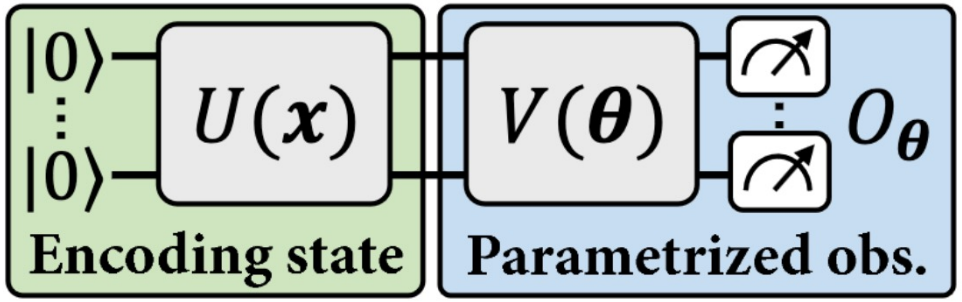


Jerbi, …, Briegel & Dunjko, **Quantum machine learning beyond kernel methods**. Nature Communications (2023)

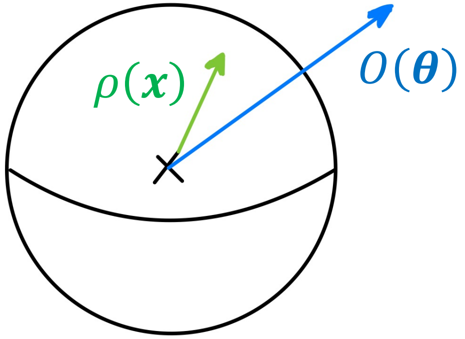
From explicit model to flipped model

Explicit model

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \mathrm{Tr}[\rho(\boldsymbol{x})O(\boldsymbol{\theta})]$$

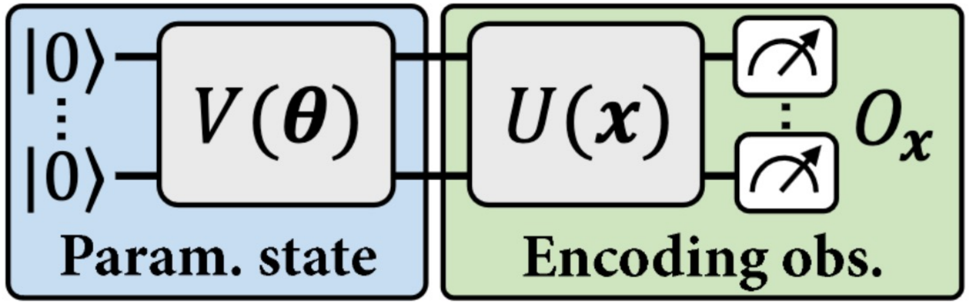


$\rho(\boldsymbol{x}) = |\psi(\boldsymbol{x})\rangle\langle\psi(\boldsymbol{x})|$ $\quad O(\boldsymbol{\theta}) = V^\dagger(\boldsymbol{\theta})OV(\boldsymbol{\theta})$

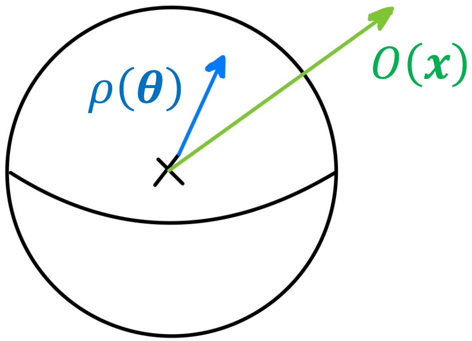


Flipped model

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \mathrm{Tr}[\rho(\boldsymbol{\theta})O(\boldsymbol{x})]$$

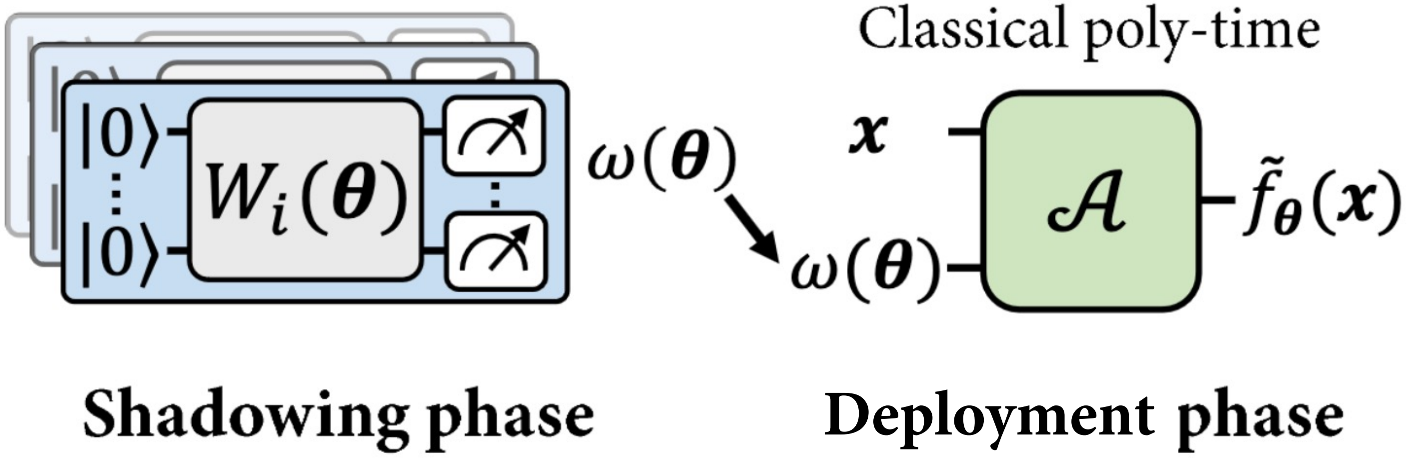


$\rho(\boldsymbol{\theta}) = |\psi(\boldsymbol{\theta})\rangle\langle\psi(\boldsymbol{\theta})|$ $\quad O(\boldsymbol{x}) = U^\dagger(\boldsymbol{x})O_{\boldsymbol{x}}U(\boldsymbol{x})$

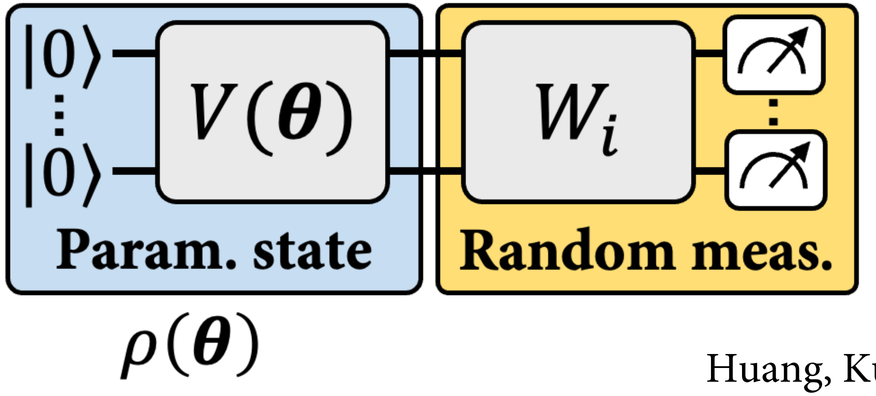


Shadows of flipped models

After training phase:



E.g., for flipped models:



Construct a classical shadow $\hat{\rho}(\boldsymbol{\theta}) = \omega(\boldsymbol{\theta})$

That allows to estimate

$\tilde{f}_{\boldsymbol{\theta}}(\boldsymbol{x}) \approx \mathrm{Tr}[\rho(\boldsymbol{\theta})O(\boldsymbol{x})]$

for a certain family $\{O(\boldsymbol{x})\}_{\boldsymbol{x}}$

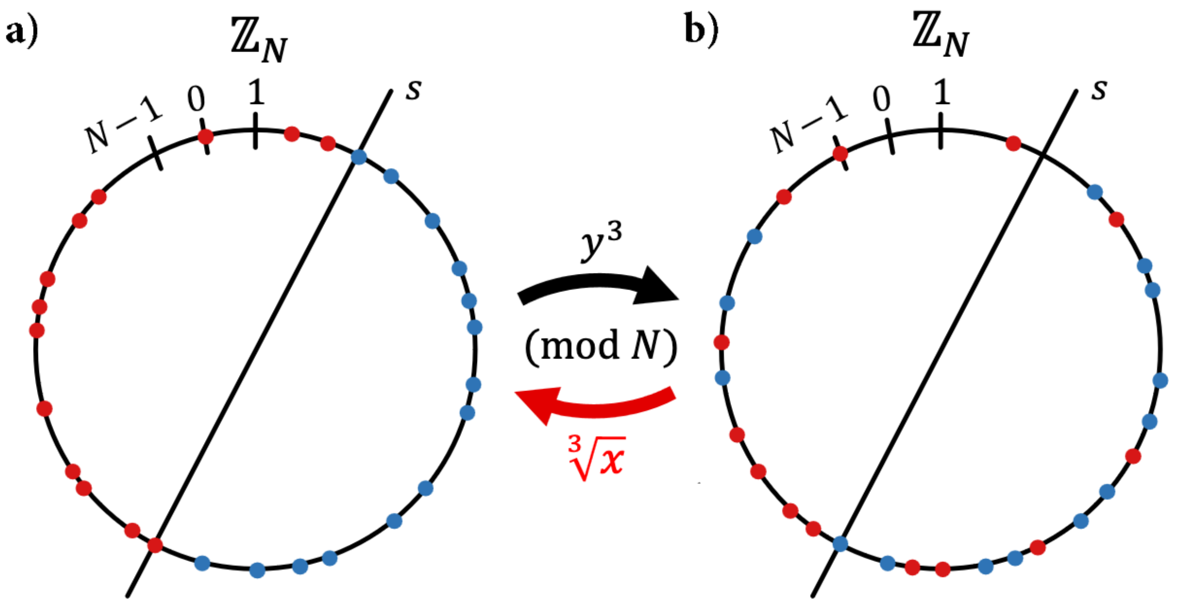
Huang, Kueng & Preskill, **Predicting many properties of quantum states from very few measurements**, Nature Physics (2020)

Q1: Quantum advantage with shadow models

- Cannot construct a shadow model for the discrete-log learning task

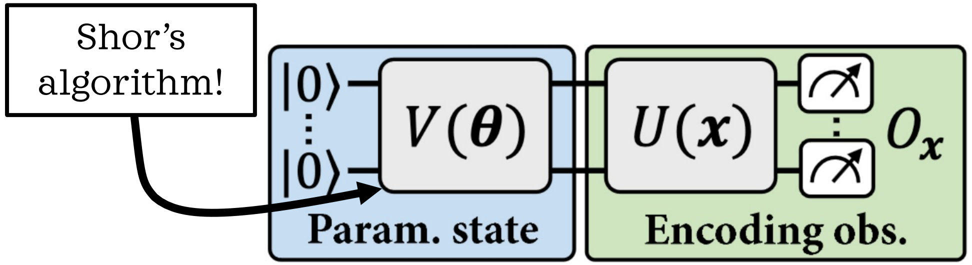
(Flipping leads to non-classically-evaluatable $O(\boldsymbol{x})$)

- Turn instead to another learning task:



Cube root has a **trap door**:
$\exists d$ such that $x^d \bmod N = \sqrt[3]{x} \bmod N$

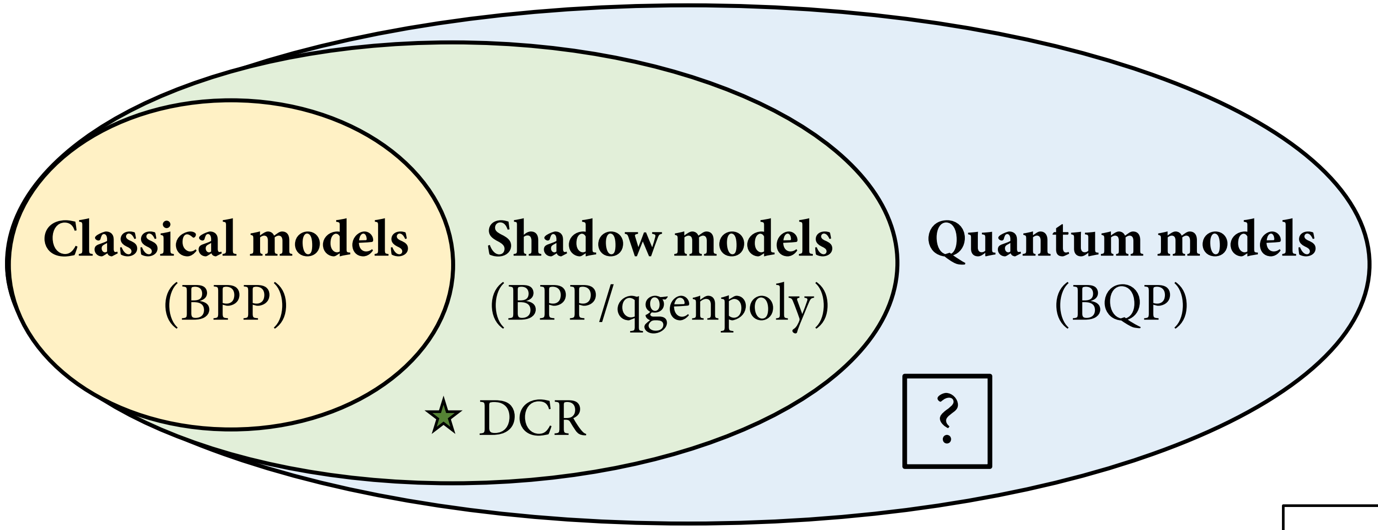
A shadowfiable flipped model solves the task!



$$\rho(\boldsymbol{\theta}) = |d, s\rangle\langle d, s|$$

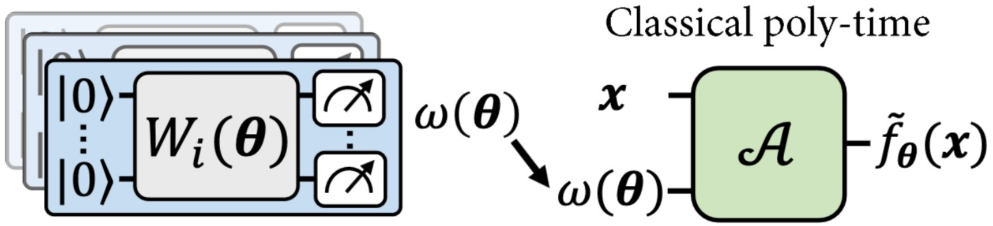
$$O(\boldsymbol{x}) = \sum_{d,s} (x^d \bmod N)|d, s\rangle\langle d, s|$$

A view from complexity theory



BPP/qgenpoly:

($\subset$ BPP/poly = P/poly)



Generation of quantum advice

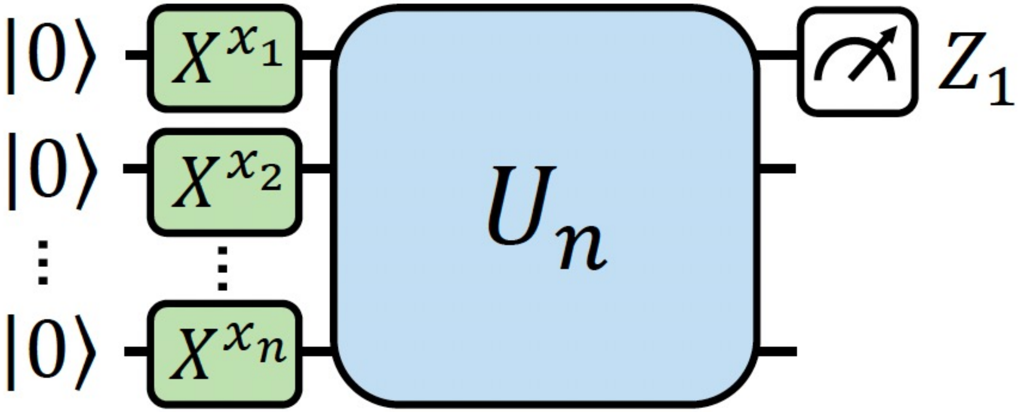
BPP computation aided by advice

Side remark:
A classical surrogate that does not use QC is in BPP.

[1] Schreiber *et al.*, PRL 2023
[2] Landman *et al.*, ICLR 2023
[3] Rudolph *et al.*, 2308.09109
[4] Sweke *et al.*, 2309.11647

Q2: Limitations of shadow models

Easy to construct a universal model for BQP:

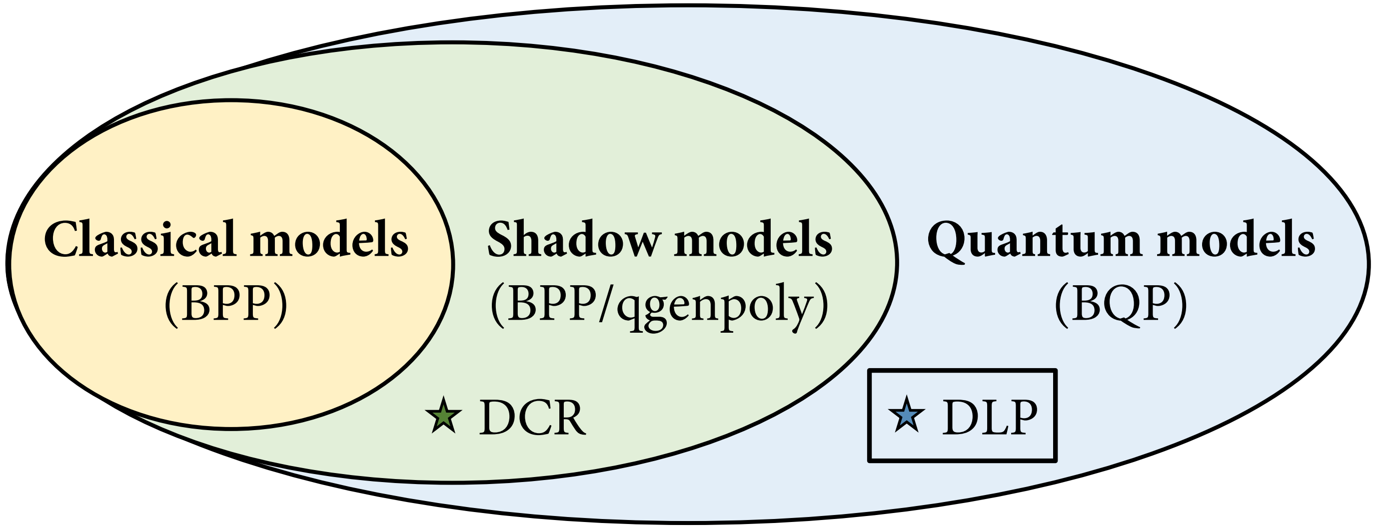


If this BQP-complete model is in BPP/qgenpoly $\implies$ BQP $\subseteq$ P/poly Very unlikely

Similarly,

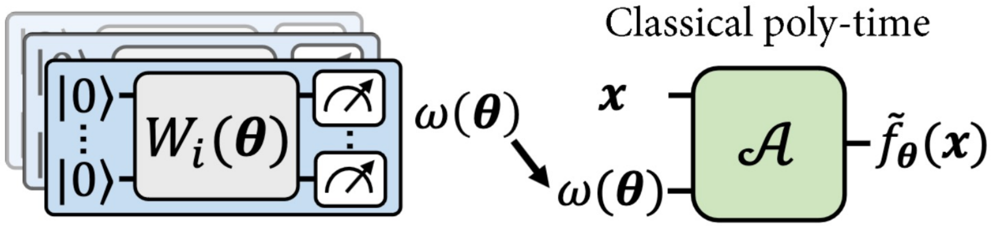
If any model for DLP is in BPP/qgenpoly $\implies$ DLP $\subseteq$ P/poly Very unlikely

A view from complexity theory



BPP/qgenpoly:

($\subset$ BPP/poly = P/poly)



Generation of quantum advice

BPP computation aided by advice

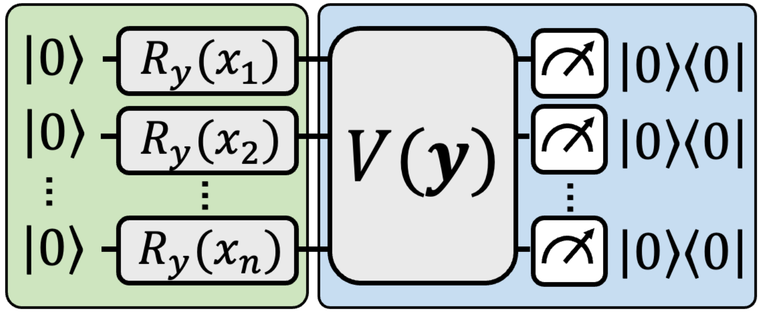
Non Fourier shadow models?

- Bonus question: do there exist models that are not efficiently Fourier shadowfiable?

Black-box queries of $f_{\boldsymbol{\theta}}(\boldsymbol{x})$ (up to add. error) to learn the coefficients $c_{\boldsymbol{\omega}}(\boldsymbol{\theta})$

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \sum_{\boldsymbol{\omega}\in\Omega} c_{\boldsymbol{\omega}}(\boldsymbol{\theta}) e^{-\mathrm{i}\boldsymbol{\omega}\cdot\boldsymbol{x}}, \qquad \boldsymbol{x} \in \mathbb{R}^n$$

Take:



$$f_y(\boldsymbol{x}) = \mathrm{Tr}[\rho(\boldsymbol{x})O(\boldsymbol{y})], \qquad \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{y} \in \{0,1\}^n$$

$$\rho(\boldsymbol{x}) = \otimes_{i=1}^{n} R_y(x_i)|0\rangle\langle 0|R_y^\dagger(x_i)$$

$O(\boldsymbol{y}) = |\boldsymbol{y}\rangle\langle\boldsymbol{y}|$ or, e.g., $O(\boldsymbol{y}) = U_{DLP}|\boldsymbol{y}\rangle\langle\boldsymbol{y}|U_{DLP}^\dagger$

To guarantee: $\left|\tilde{f}_{\boldsymbol{\theta}}(\boldsymbol{x}) - f_{\boldsymbol{\theta}}(\boldsymbol{x})\right| \leq 1/4 \quad \forall \boldsymbol{x} \in \mathbb{R}^n$

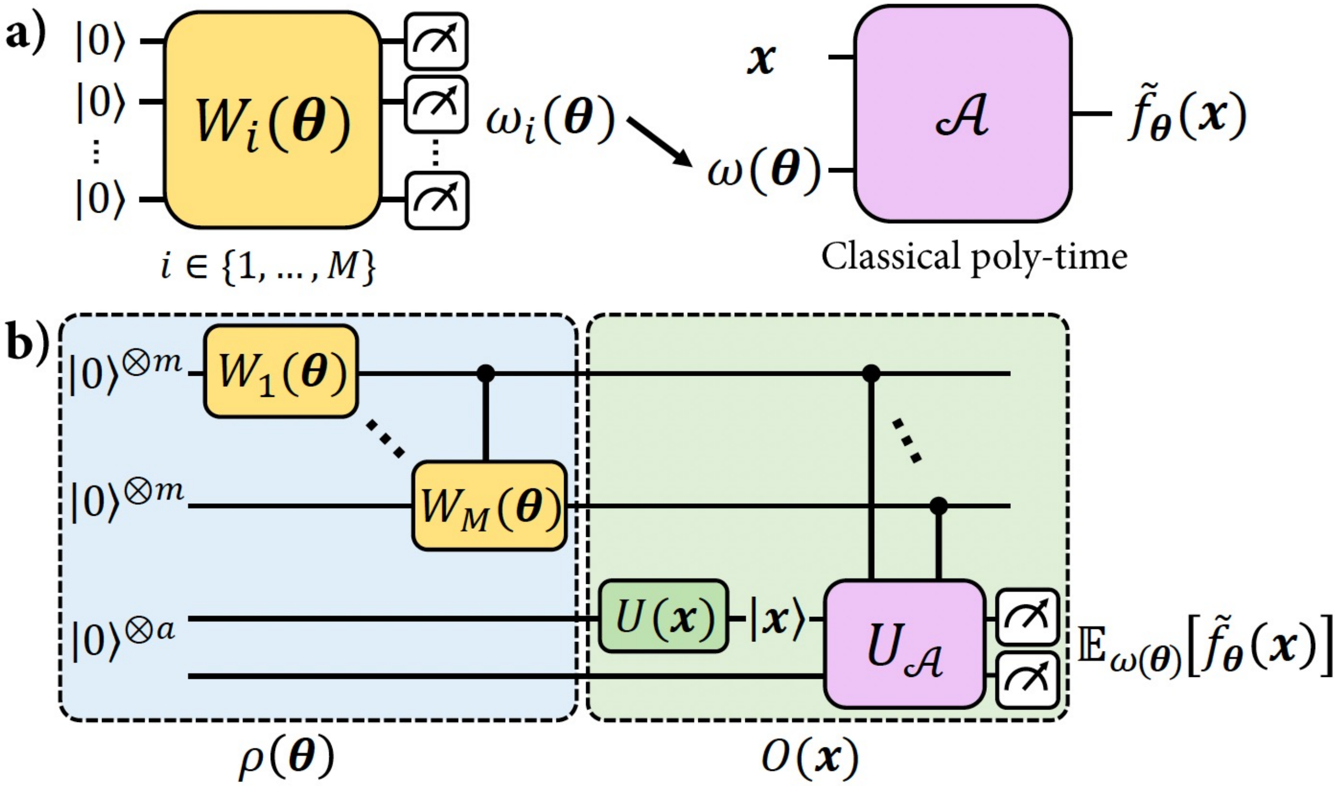
$$\boxed{\text{Sample complexity: } \Omega(2^n) \quad \text{v.s.} \quad \tilde{O}\left(\frac{{\omega_M}^n \|O\|_\infty^2}{\varepsilon^2}\right)}$$

Easy to see for $\boldsymbol{x} \in \left\{0, \frac{\pi}{2}\right\}^n$, it's a Grover oracle.

But trivially shadowfiable when flipped!

All shadowfiable models are shadowfiable flipped models

- Bonus answer: Flipped models are shadow-universal
- Proof:

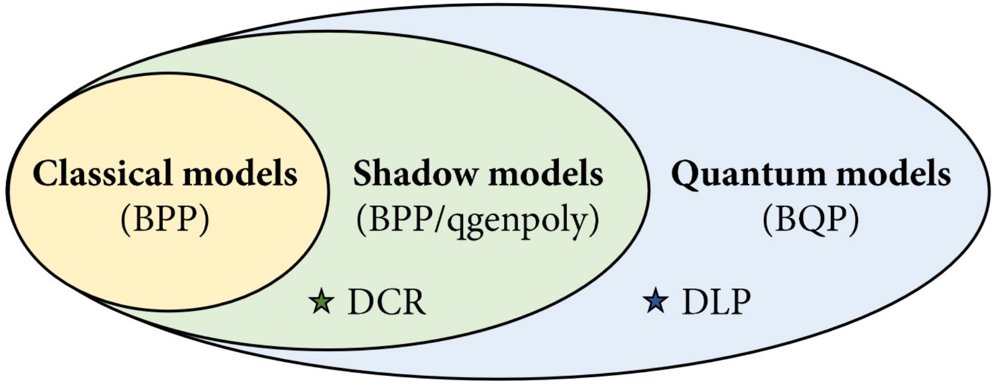


Summary

Main questions:

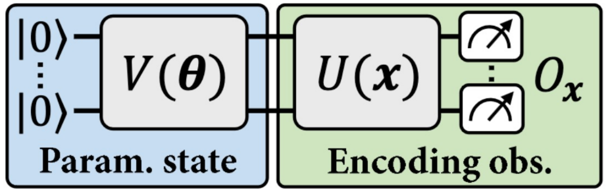
Q1. Can shadow models achieve a **quantum advantage** over entirely classical models?

Q2. Do there exist quantum models that **do not admit shadow models**?

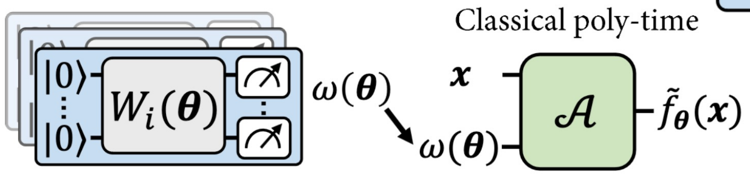


Outlook

- How do (shadowfiable) flipped models do in practice?

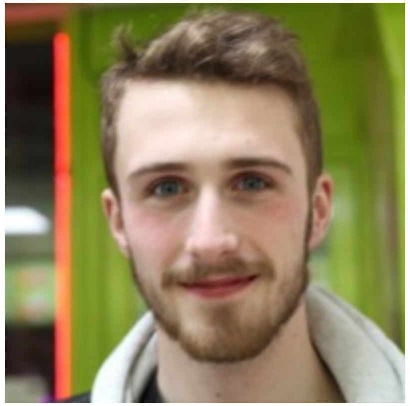


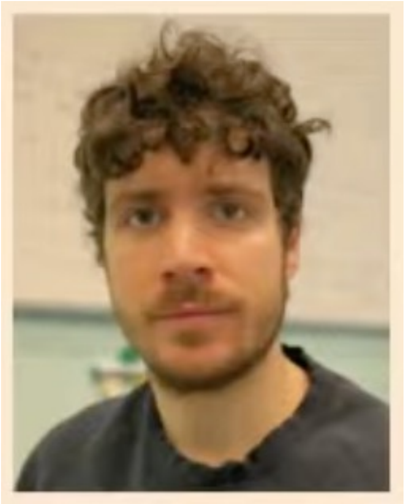
- New designs for shadow models?



Special thanks


Casper Gyurik


Simon Marshall


Riccardo Molteni


Vedran Dunjko

Freie Universität Berlin

Universiteit Leiden

⟨aQa⟩

universität innsbruck

Generalization performance

- Learning performance:

$$\underbrace{\hat{\mathcal{L}}(f_{\boldsymbol{\theta}})}_{\substack{\textit{Training loss} \\ \max_{m\in\{1,\ldots,M\}} |f_{\boldsymbol{\theta}}(\boldsymbol{x}^{(m)}) - g(\boldsymbol{x}^{(m)})|}} \qquad v.s. \qquad \underbrace{\mathcal{L}_\gamma(f_{\boldsymbol{\theta}})}_{\substack{\textit{Expected loss} \\ \Pr_{\boldsymbol{x}\in\mathcal{X}}[|f_{\boldsymbol{\theta}}(\boldsymbol{x}) - g(\boldsymbol{x})| > \gamma]}}$$

- For a flipped model: $f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \mathrm{Tr}[\rho(\boldsymbol{\theta})O(\boldsymbol{x})]$

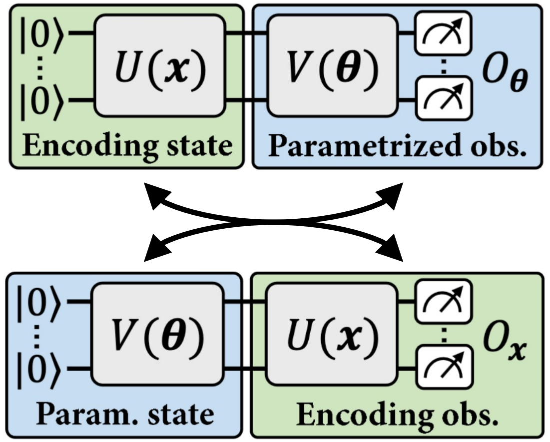
  For $\hat{\mathcal{L}}(f_{\boldsymbol{\theta}}) = \eta$,

if the training set size $\mathrm{M} \geq \tilde{O}\left(\frac{n\|O\|_\infty}{\varepsilon(\gamma-\eta)^2}\right)$

then $\mathcal{L}_\gamma(f_{\boldsymbol{\theta}}) \leq \varepsilon$

Aaronson, **The Learnability of Quantum States**. Proceedings of Royal Society A (2007)

Flipping bounds



- **Lower bound:**

There exist explicit models $\mathrm{Tr}[\rho(\boldsymbol{x})O(\boldsymbol{\theta})]$, with $\|O\|_1 = d$ and $n = \mathcal{O}(\log d)$ qubits,

such that, for any flipped model $\mathrm{Tr}[\rho'(\boldsymbol{\theta})O'(\boldsymbol{x})]$, if

$$|\mathrm{Tr}[\rho(\boldsymbol{x})O(\boldsymbol{\theta})] - \mathrm{Tr}[\rho'(\boldsymbol{\theta})O'(\boldsymbol{x})]| \leq \varepsilon, \forall \boldsymbol{x}, \boldsymbol{\theta}$$

$$\text{then } m \, \|O'\|_\infty^2 \geq \Omega\left(\mathrm{d}^2 \left(\frac{1}{2} - \varepsilon\right)^2\right)$$

- **Upper bound:**

We give an exact procedure ($\varepsilon = 0$) that uses $m = n + 1$ qubits and guarantees $\|O\|_\infty = \|O\|_1$.

Based on a renormalization of $O(\boldsymbol{\theta})$ and importance sampling.