University of Sheffield

Talk @ QT4HEP 2025
CERN, Geneva

Quantum Machine Learning with Physics-Informed and Symmetry-Aware Models

Oleksandr Kyriienko
Professor @ University of Sheffield, UK
Sheffield Quantum Centre

o.kyriienko@sheffield.ac.uk https://kyriienko.github.io/

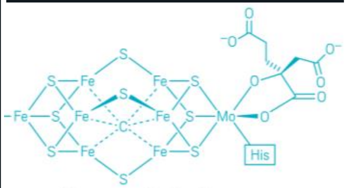
University of Sheffield

QUANTUM APPLICATIONS

Our goal is to **solve outstanding computing problems** and **advance data analysis** with **quantum computing**.

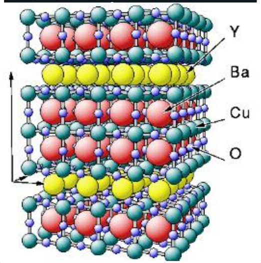
However, not all use cases will be suitable, and only particular areas may benefit from quantum advantage. **Examples:**

Chemistry & Pharma

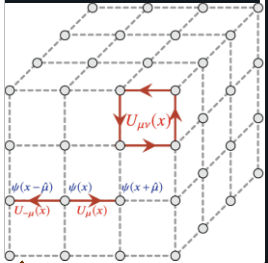


Nitrogenase Fe-Mo cofactor

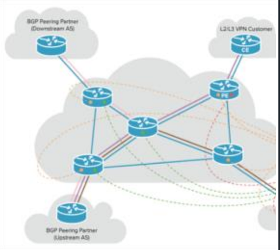
Materials



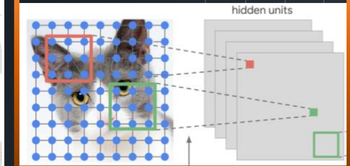
HEP



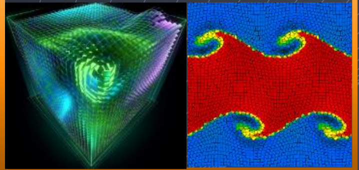
Logistics/finance

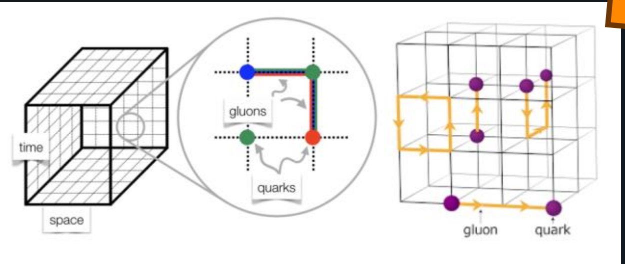


Machine Learning



Scientific computing





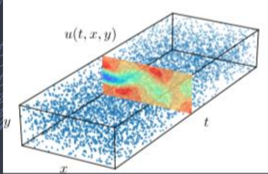
$$\hat{H}_{\text{KS}} = \frac{g^2}{2a}\sum_{\vec{n},i,b}\hat{E}^b_{\vec{n},i}\hat{E}^b_{\vec{n},i} - \frac{1}{2g^2a}\sum_{\vec{n},j,k}\text{Tr}[\hat{U}_{\vec{n},jk} + \hat{U}^\dagger_{\vec{n},jk}]$$

lattice gauge theory (quantum simulation)

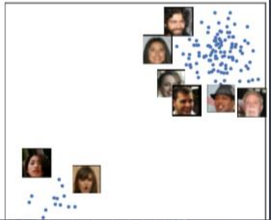
properties of nuclear structure and hadron interactions

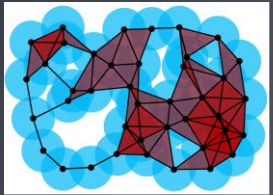
use quantum computers to perform **lattice simulations**, at **scale inaccessible before**

Quantum Scientific Machine Learning (QuaSciML)



- classifying phases
- generative modelling
- solving diff. eqns.
- topological data analysis

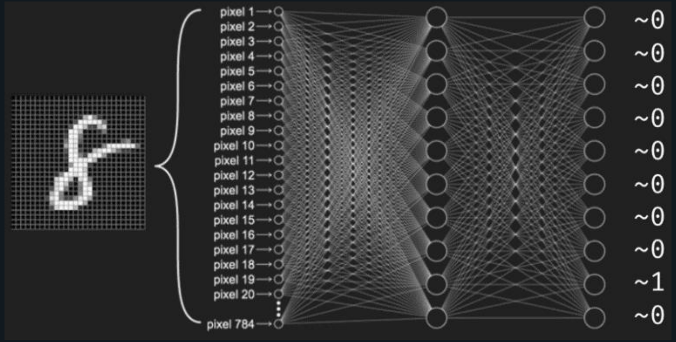




MACHINE LEARNING

Classical machine learning

○ complex models can be designed via vector-matrix multiplication and nonlinear transforms to achieve **universal function approximation**



deep neural network

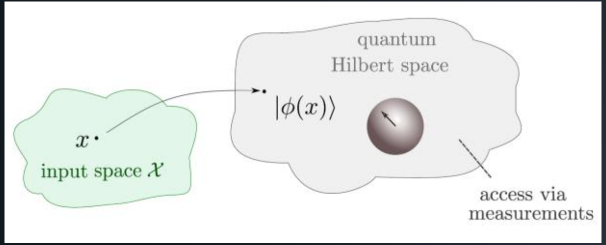
○ nonlinear models can represent complicated correlations and **multivariate** functions

○ the power of learning often comes from **high-dimensional latent space** representation

○ can benefit from kernel-based methods

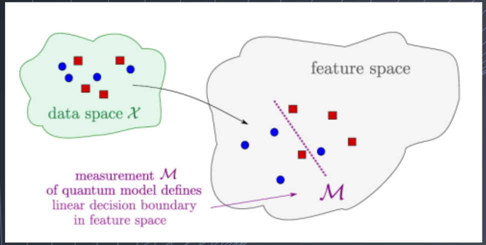
Quantum machine learning

○ quantum models can be designed by mapping data into quantum states and **adaptive** search for a suitable **measurement operator**

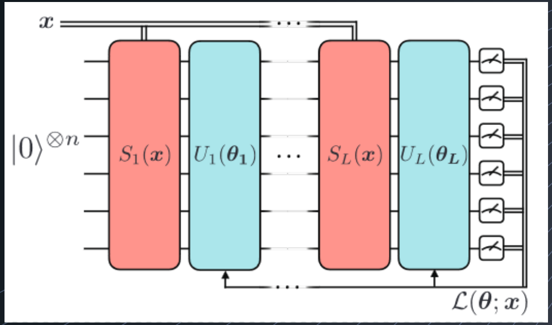


quantum embedding (feature map)

[M. Schuld, arXiv:2101.11020 (2021)]



drawing decision boundary in a feature space



quantum neural network (QNN)

○ use **parametrized unitaries $S_k(x)$** to embed data into quantum states

○ use **variational ansatz $U_k(\theta_k)$** to adjust model and adapt measurement

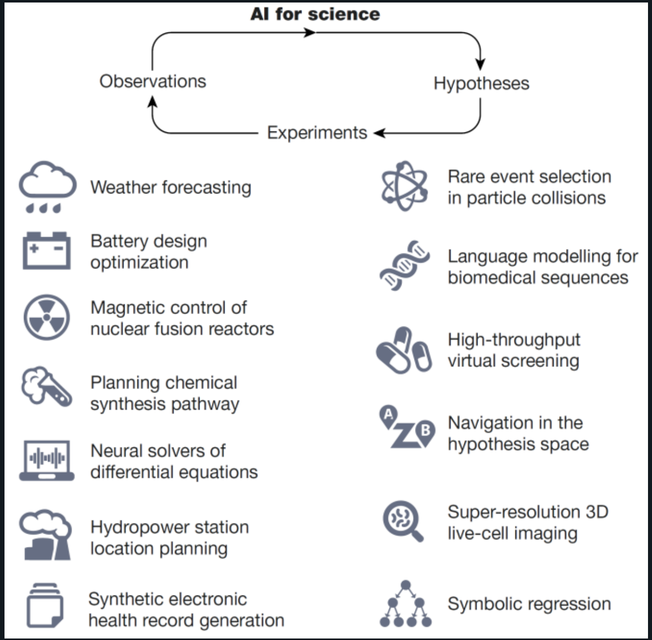
○ minimize the **loss $L(\theta;x)$** depending on the task, benefitting from **exponential scaling** of the latent space

University of Sheffield

PHYSICS-AWARE LEARNING

Scientific **machine learning**

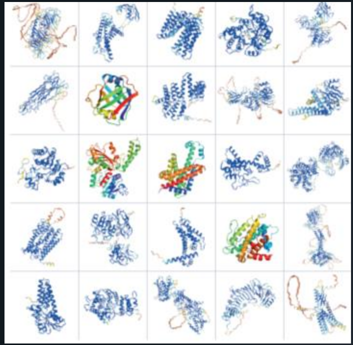
○ use of unconstrained DNNs requires **big data** and **trainable** architectures



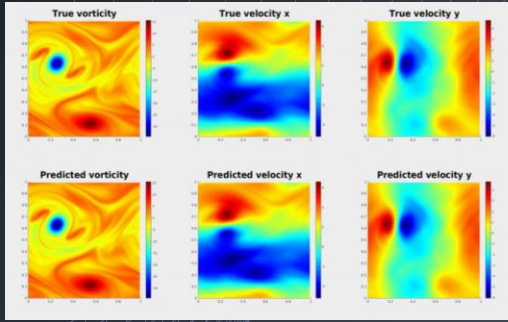
[Wang et al. (DeepMind), Nature 620, 47 (2023)]

Physics-aware **SciML**

○ we can improve model performance (accuracy and generalization) by **embedding physics** and **symmetries** at different levels of the workflow



protein structure prediction with AlphaFold 3
[Abramson et al. (DeepMind), Nature 630, 493 (2024)]



turbulence modelling with PINNs and FNOs
[Li et al. (Caltech), arXiv:2010.08895 (2020)]

problem level

data level

architecture level

loss level

University of Sheffield

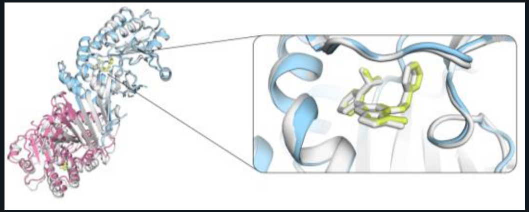
PHYSICS-AWARE LEARNING

Physics-aware **SciML**

○ we can improve model performance (accuracy and generalization) by **embedding physics** and **symmetries** at different *levels* of the workflow [S. Brunton]:

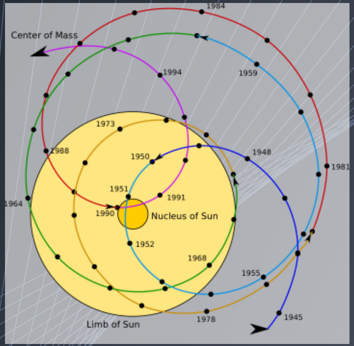
problem level

○ make sure that there is an **underlying physical model** when selecting the problem



data level

○ supply **scientific data** in a suitable format and **augment** it with symmetric configurations

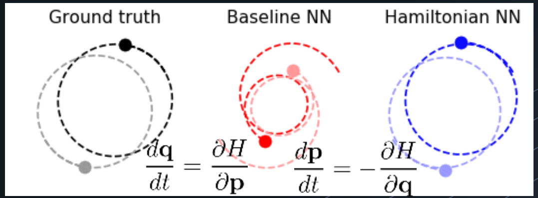


○ choose appropriate **data representation** and coordinate system

architecture level

○ add **inductive bias** to restrict models and include conservation laws

○ embed **symmetries** into the architecture

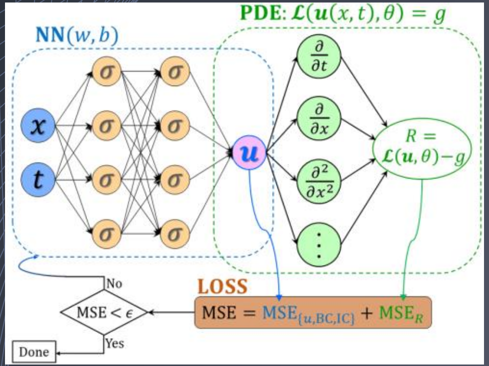


[Greydanus et al., HNNs, NeurIPS (2019)]

loss level

○ combine data-driven modelling with use of **differential equations**

○ add **differential constraints** for training models ➡ **PINNs**



University of Sheffield

PHYSICS-INFORMED: DE-based LOSS

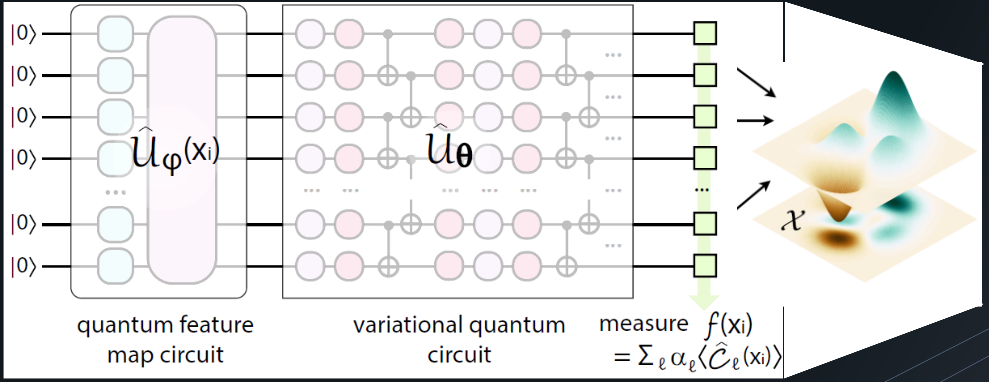
Quantum PINNs

○ quantum models (functions) can be encoded with rotation-based gates and differentiated in the quantum space

[Mitarai et al., Phys. Rev. A 98, 032309 (2018)]

○ we circuit differentiation to develop a strategy for **solving differential equations** with trainable quantum circuits

[OK et al., Phys. Rev. A 8, 103, 052416 (2021)]

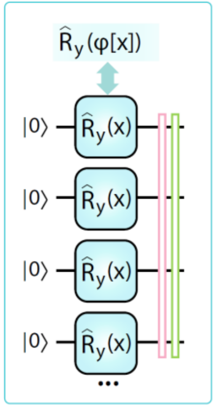


quantum neural network (QNN)

○ one advantage is the use of **automatic differentiation** and **physics-informed** differential constraints (like PINNs)

○ one of the most **important** QNN building blocks is the **feature map**, which often is chosen as a sequence of rotations

$$|\psi(x)\rangle = \prod_{i=1}^{N} \hat{R}^i_\alpha(\phi_i(x))|\psi_0\rangle$$



product feature map (unitary circuit for data embedding)

$$|f_{\varphi,\theta}(x)\rangle = \hat{\mathcal{U}}_\theta \hat{\mathcal{U}}_\varphi(x)|\emptyset\rangle$$

embedded state

$$f(x) = \langle f_{\varphi,\theta}(x)|\hat{C}|f_{\varphi,\theta}(x)\rangle$$

quantum model
(latent space function)

autodiff via **parameter shift rule**

$$df(x)/dx = \frac{1}{2}\sum_j \Big(\langle f^+_{d\varphi,j,\theta}(x)|\hat{C}|f^+_{d\varphi,j,\theta}(x)\rangle - \langle f^-_{d\varphi,j,\theta}(x)|\hat{C}|f^-_{d\varphi,j,\theta}(x)\rangle \Big)$$

diff. model

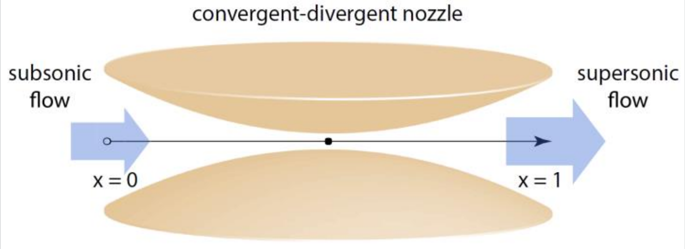
$$\min_\theta(\mathcal{L}_\theta[d_x f, f, x])$$

We can embed circuits and their derivatives into loss function for nonlinear data-driven problems

PHYSICS-INFORMED: DE-based LOSS

As an example in **fluid dynamics** we solved **Navier-Stokes equations** for a quasi-1D nozzle.

[OK, A. Paine, V. Elfving, PRA 8, 103, 052416 (2021)]

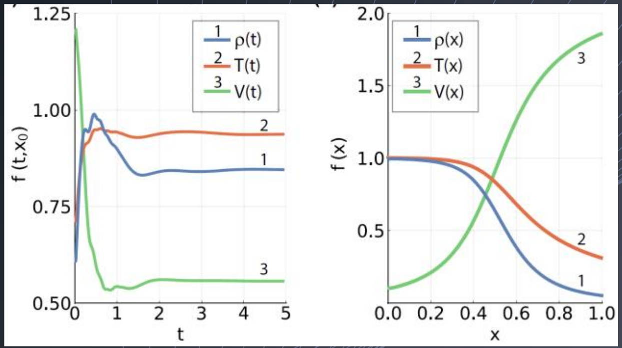


system geometry

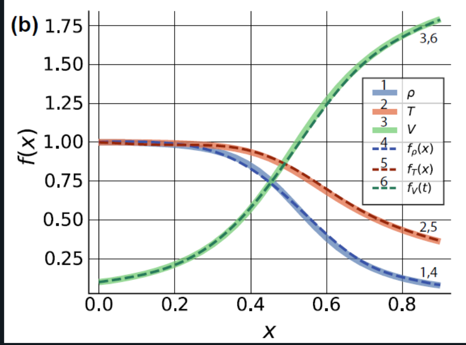
$$\frac{\partial \rho}{\partial t} = -\rho \frac{\partial V}{\partial x} - \rho V \frac{\partial (\log A)}{\partial x} - V \frac{\partial \rho}{\partial x},$$

$$\frac{\partial T}{\partial t} = -V \frac{\partial T}{\partial x} - (\gamma - 1) T \left(\frac{\partial V}{\partial x} + V \frac{\partial (\log A)}{\partial x} \right)$$

$$\frac{\partial V}{\partial t} = -V \frac{\partial V}{\partial x} - \frac{1}{\gamma} \left(\frac{\partial T}{\partial x} + \frac{T}{\rho} \frac{\partial \rho}{\partial x} \right),$$



Navier-Stokes equations and reference solutions*



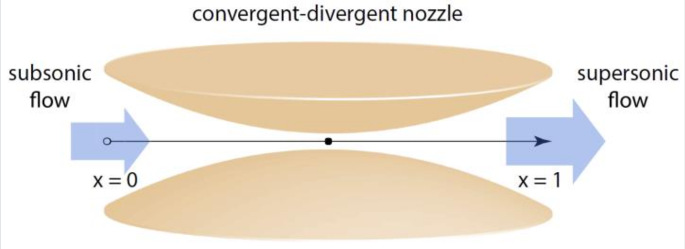
DQC solution

○ we recovered solutions for density, temperature, and velocity by training in domains before and after nozzle

PHYSICS-INFORMED: DE-based LOSS

As an example in **fluid dynamics** we solved **Navier-Stokes equations** for a quasi-1D nozzle.

[OK, A. Paine, V. Elfving, PRA 8, 103, 052416 (2021)]



system geometry

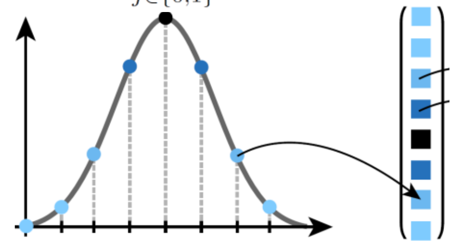
$$\frac{\partial \rho}{\partial t} = -\rho\frac{\partial V}{\partial x} - \rho V\frac{\partial(\log A)}{\partial x} - V\frac{\partial \rho}{\partial x},$$

$$\frac{\partial T}{\partial t} = -V\frac{\partial T}{\partial x} - (\gamma-1)T\left(\frac{\partial V}{\partial x} + V\frac{\partial(\log A)}{\partial x}\right)$$

$$\frac{\partial V}{\partial t} = -V\frac{\partial V}{\partial x} - \frac{1}{\gamma}\left(\frac{\partial T}{\partial x} + \frac{T}{\rho}\frac{\partial \rho}{\partial x}\right),$$

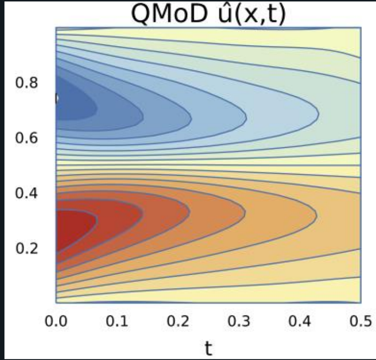
Navier-Stokes equations and reference solutions*

$$|\psi\rangle = \sum_{j\in\{0,1\}^n} g(x_j)|j\rangle$$

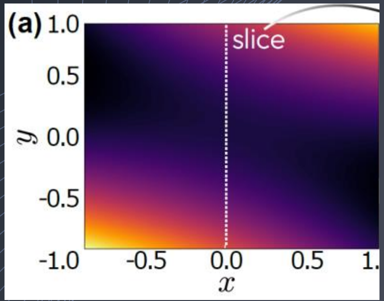


function embedding

$$\frac{\partial u}{\partial t} = \varepsilon\frac{\partial^2 u}{\partial x^2} - \alpha u\frac{\partial u}{\partial x}$$



- we recovered solutions for density, temperature, and velocity by training in domains before and after nozzle
- we can also learn models with **quantum model discovery** and **physics-informed** constraints
- advancing the tools for embedding and mapping models, we discovered solvers that have **parallel grid** evaluation – **exponential** improvement



2D PDE w/ parallel grid evaluation

[N. Heim, A. Gosh, OK, V. Elfving, arXiv:2111.06376]

[A. Paine, V. Elfving, OK, arXiv:2308.01827]

We continue expanding the toolbox for solving differential equations in a data-driven way

PHYSICS-INFORMED: PROBABILISTIC

Going beyond loss constraints, we show that QML can embed **probabilistic models**.

[OK, A. Paine, V. Elfving, Phys. Rev. Res. 6, 033291 (2024)]
[S. Kasture, OK, V. Elfving, Phys. Rev. A 108, 042406 (2023)]
[A. Paine, V. Elfving, OK, Adv. Q. Tech. 6, 033291 (2024)]

We developed mappings between continuous and discrete models for solving **Fokker-Planck PDEs** for relevant **SDEs**.

$$dX_t = -\nu(X_t - \mu)dt + \sigma dW_t$$

SDE for Ornstein-Uhlenbeck process

$$\nu p(x,t_s) + \nu(x-\mu)\frac{d}{dx}p(x,t_s) + \frac{\sigma^2}{2}\frac{d^2}{dx^2}p(x,t_s) = 0$$

Fokker-Planck equation at stationarity

derivatives?

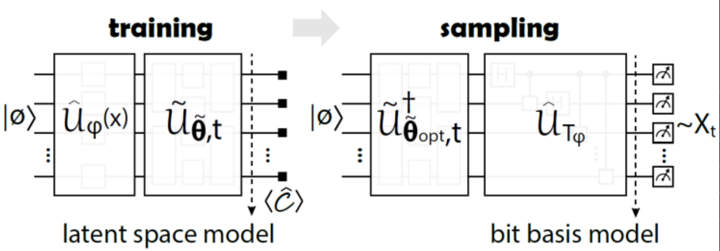
$$p_\theta(x) = |\langle x|\psi_\theta\rangle|^2$$

QCBM (quantum circuit Born machine)

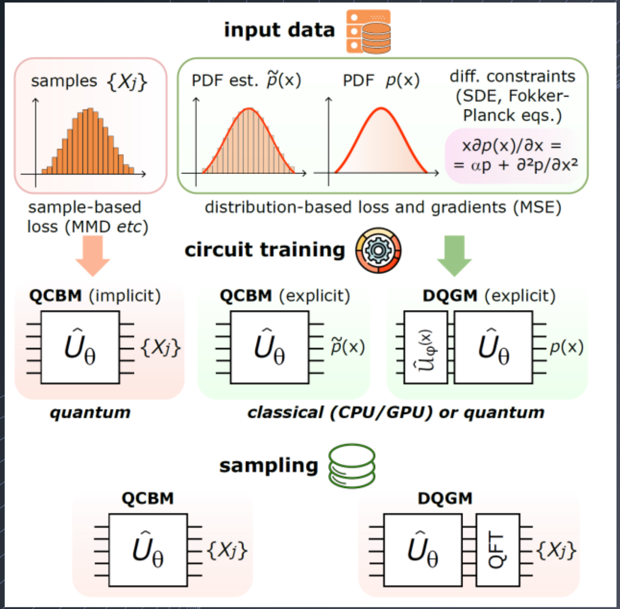
DQGM (differentiable quantum generative model)

$$\widetilde{p}_{\tilde{\theta},t}(x) = \mathrm{tr}\{\hat{\rho}_{\tilde{x}}\widetilde{U}^\dagger_{\tilde{\theta},t}\hat{\rho}_0\widetilde{U}_{\tilde{\theta},t}\}$$

$$\widetilde{p}_{\tilde{\theta},t}(x) = \mathrm{tr}\{|x\rangle\langle x|\hat{\mathcal{W}}^\dagger_{\tilde{\theta},t}\hat{\rho}_0\hat{\mathcal{W}}_{\tilde{\theta},t}\}$$



training and sampling stages if DQGM



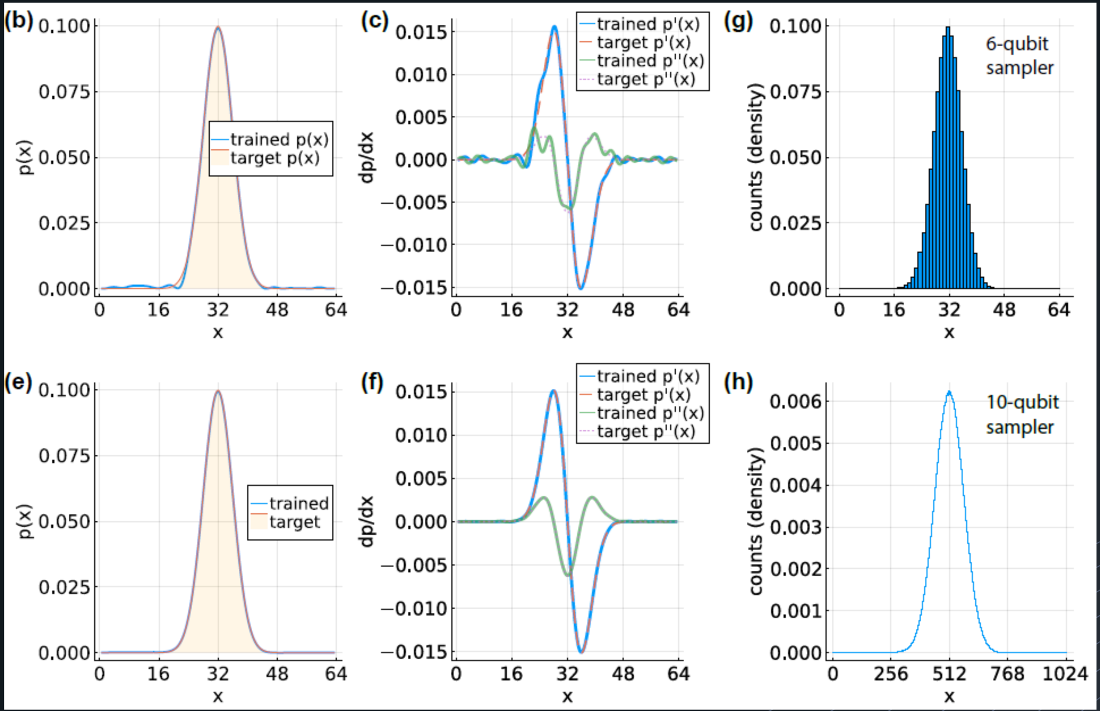
implicit and explicit models

University of Sheffield

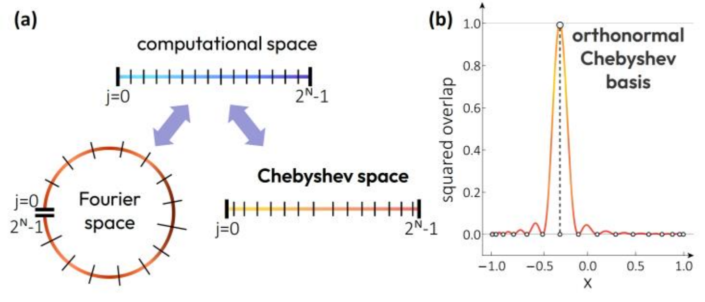
PHYSICS-INFORMED: BASIS SELECTION

FPE-based loss improves **generalization** and enables sampling with **extended** registers.

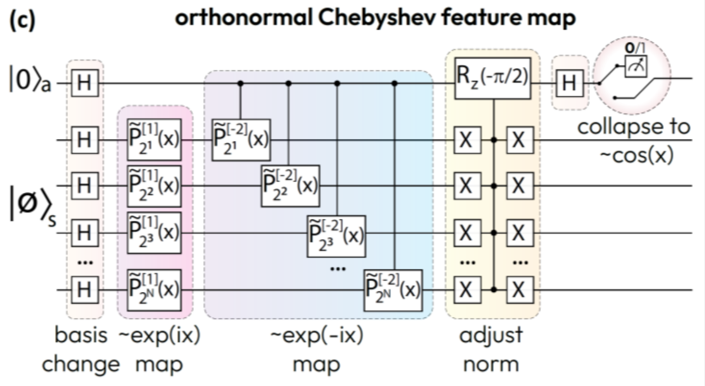
To modify DQGM bases we also developed **quantum Chebyshev toolbox**.
[C. Williams, A. Paine, H-Y. Wu, V. Elfving, OK, arXiv:2306.17026 (2023)]



training probabilistic models with differential constraints



mapping to Chebyshev space

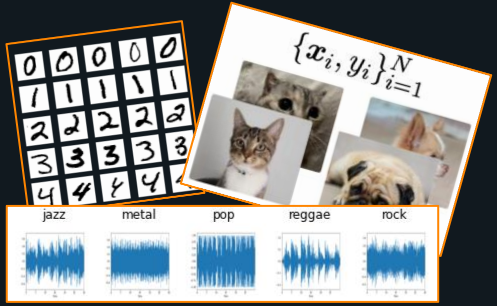


quantum Chebyshev feature map/transform

University of Sheffield

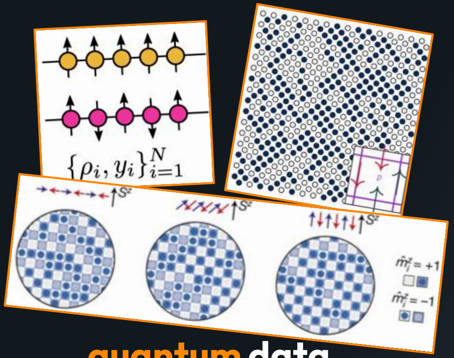
CLASSICAL vs QUANTUM DATA

Recent works have highlighted that QML can be good for analyzing **quantum datasets** – collections of quantum states.



classical data

vs



quantum data

[I. Cong, Choi, Lukin (Harvard), Nature Phys. (2019)]

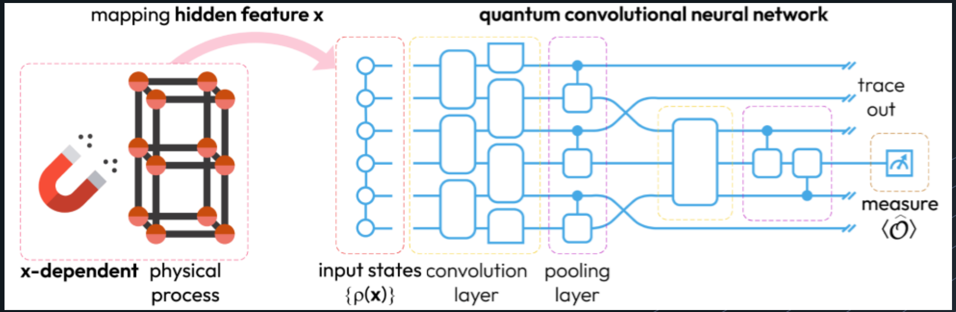
[Huang et al. (Google Q.AI), Science 376, 1182 (2022)]

○ **ground state** can be classified by **QCNNs** (quantum convolutional neural networks) for distinguishing different phases

○ appealed to entanglement but not the basis

○ apparent **gap in understanding** QML models for classical and quantum datasets

We proposed to formalize the use of quantum data and analyze how it introduces an inductive bias



[C. Umeano, ..., OK, Adv Q. Technol. 2400325 (2024)]

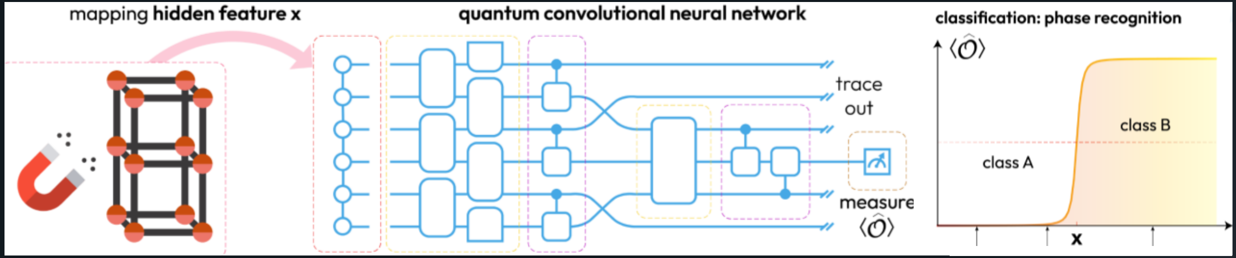
○ QCNNs show that we can build models that depend on **hidden features** (parameters of a system Hamiltonian etc)

○ in this case **ground state preparation** becomes the quantum **feature map** making models **physics-aware**

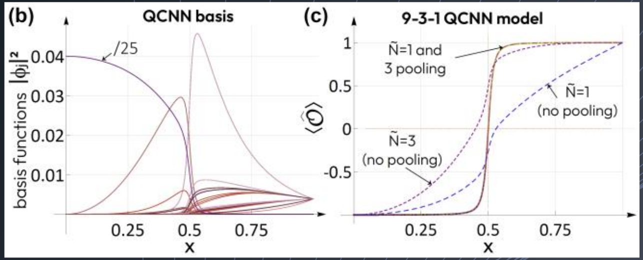
University of Sheffield

PHYSICS-AWARE: QUANTUM DATA

We consider an operator for the **ground state** preparation as a **function of x** and plot typical basis functions showing criticality.



QCNN for ground state classification



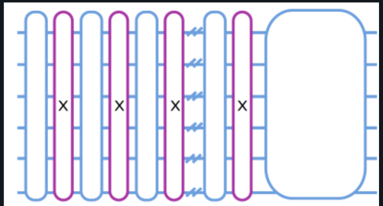
associated basis and model

[C. Umeano, …, OK, Adv Q. Technol. 2400325 (2024)]

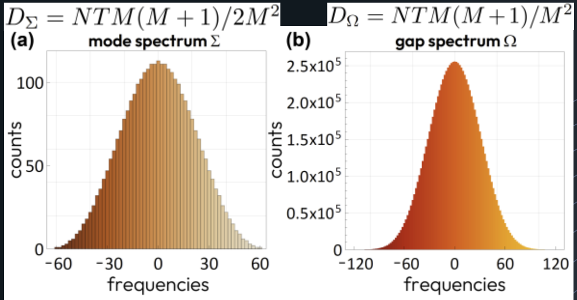
Formally GSP can be implemented as a **sequence of unitaries** and allows to analyze the **spectrum** of the embedding.

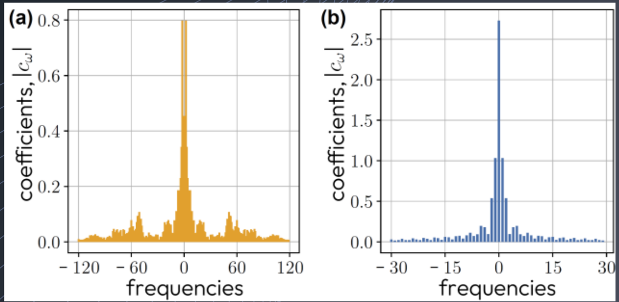
[C. Umeano & OK, arXiv:2404.07174 (2024)]

$$|\psi_G(x)\rangle = \hat{U}_\varphi(x)|\psi_0\rangle$$



ground state preparation
≈
Trotterized adiabatic evolution





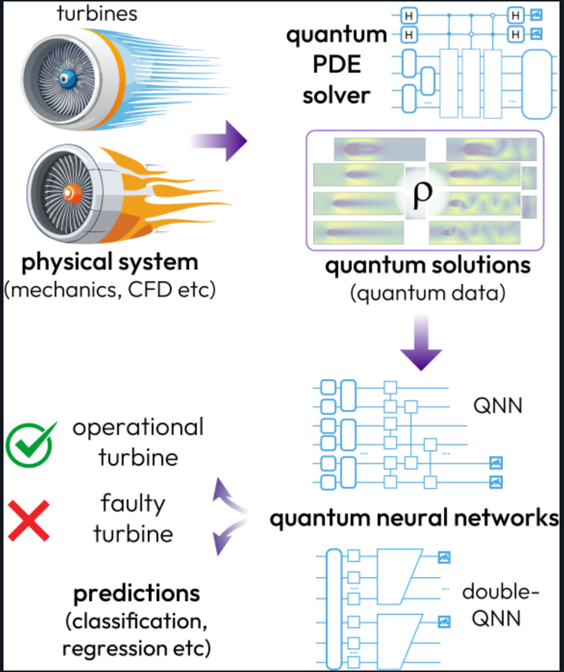
ultra-high degeneracy spectrum with **inductive bias** from quantum data

University of Sheffield

PHYSICS-AWARE: QUANTUM DATA

While **quantum data** may sound **exotic**, we point out that it is ubiquitous if we look at the **solutions of quantum diff. equations**.

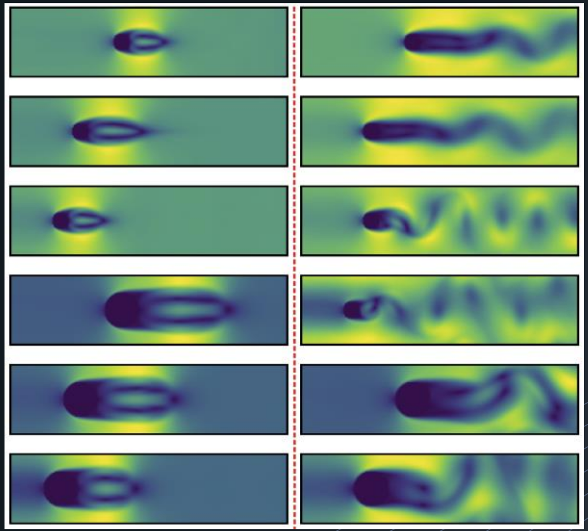
[C. Williams, S. Scali, A. Gentile, D. Berger, OK,arXiv:2411.14259 (2024)]



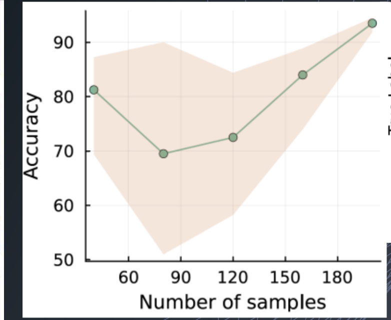
QuaSciML workflow for predictions

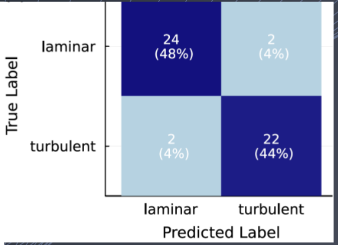
$$\frac{\partial u}{\partial t} + (u \cdot \nabla) u = -\frac{1}{\sigma}\nabla p + \nu \nabla^2 u$$

Navier-Stokes eqns. with turbulent and laminar flow



2 flow classes (quantum data)





generalization and confusion matrix

- customized QML architectures for analyzing correlations can lead to high performance
- choice of basis (quantum representation) is very important for successful readout

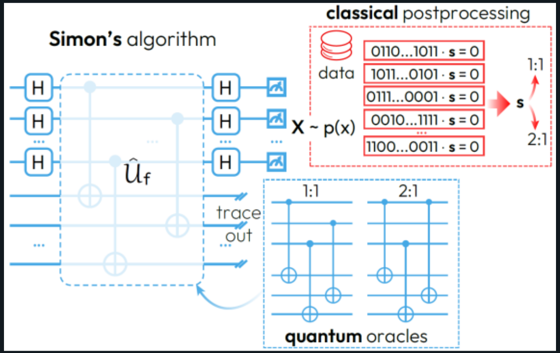
I believe physics-aware QML will be a vital part of feature extraction for PDE solvers.

University of Sheffield

SYMMETRY-AWARE: SIMON's PROBLEM

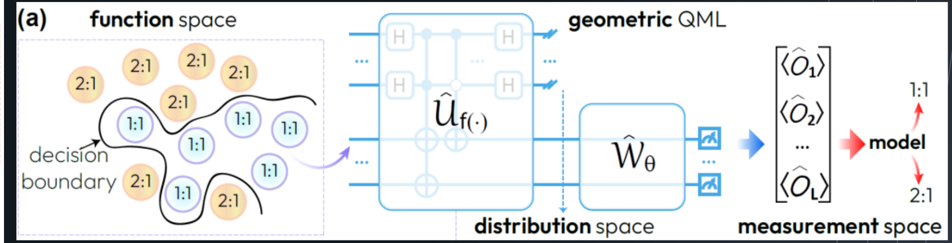
Can we have a **provable advantage** from quantum machine learning? Yes, if we shown the **QML** allows learning known protocols for **hard** problems in **BQP** complexity class.

[C. Umeano, V. Elfving, OK, arXiv:2402.03871 (2024)]

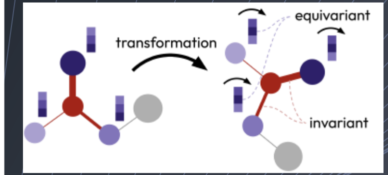


$$f(x) = f(y) = f(x \oplus s)$$

Simon's problem: learn to separate 1:1 or 2:1 functions



geometric quantum machine learning (GQML):
embed **symmetries** of you data into a model



invariance and **equivariance**

[JJ Meyer et al., PRX Quantum 4, 010328 (2023)]

---

1) **invariant initial state** $\hat{U}_\sigma|\psi_0\rangle = |\psi_0\rangle$

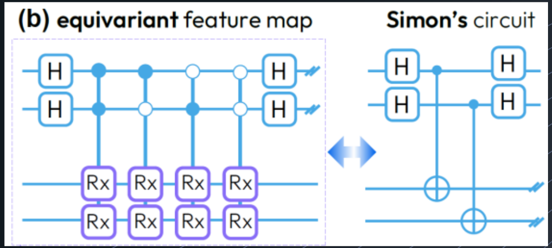
2) **equivariant embedding** $\hat{U}(V_\sigma[x]) = \hat{U}_\sigma \hat{U}(x) \hat{U}_\sigma^\dagger$

3) **equivariant ansatz** $[\hat{W}(\theta), \hat{U}_\sigma] = 0$

4) **invariant measurement** $\hat{U}_\sigma \hat{O} \hat{U}_\sigma^\dagger = \hat{O}$

In case of Boolean functions as data the **symmetries** correspond to **permutations** and **bitflips** of arguments – function type remains the same for any x.

$$g(V_{bp}[f_m]) = g(f_k)$$



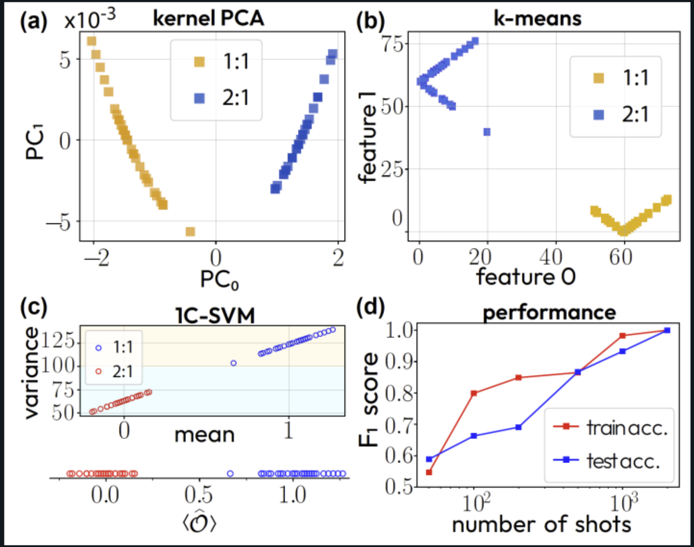
equivariant feature map

Once we introduce equivariant loading of Boolean functions, we can learn the Simon's algorithm that is known to be in BQP^A complexity class.

SYMMETRY-AWARE: SIMON's PROBLEM

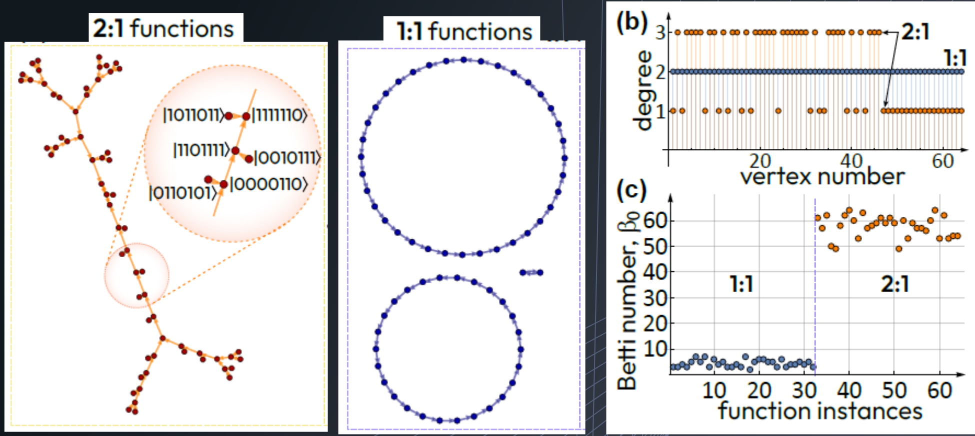
University of Sheffield

We observe that the GQML-based workflow fully represents Simon's algorithm, apart from the classical post-processing.



[C. Umeano, V. Elfving, OK, arXiv:2402.03871 (2024)]

Visualization: what do detect as features of the function-based dataset?



unsupervised detection from sampling

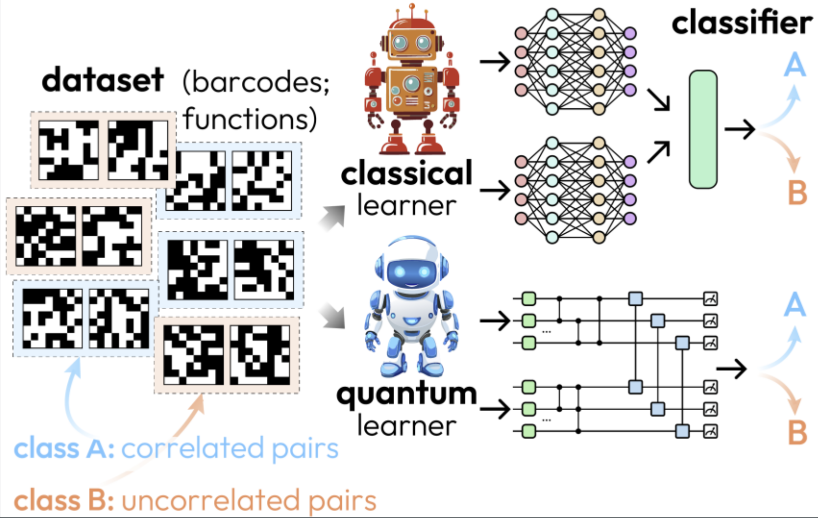
computational directed graph visualization of 1:1 or 2:1 functions and clear separation based on topological properties

We observe that quantum machine learning is capable of learning hard protocols where advantage comes from superior sampling complexity, complex input (probability distribution), and post-processing. I expect more examples to emerge.

SYMMETRY-AWARE: BARCODES

How do we find other examples where **symmetry-aware QML** is used to discover **exponentially-improved** algorithms?

[C. Umeano, S. Scali, OK, arXiv:2409.01496 (2024)]



workflow of quantum barcode classification task with learners trying to spot a similarity between samples

Input data correspond to samples from distributions of 2 types

$x_{|y=0} \sim P_{\mathrm{A}}(x)$ $x_{|y=1} \sim P_{\mathrm{B}}(x)$

correlated vs uncorrelated pairs

Quantum embedding: graph states (or phase or REW states)

$$\left(\sum_{j=1}^{N}(-1)^{\mathbf{x}_1^{[j]}}|j\rangle\right)\otimes\left(\sum_{j=1}^{N}(-1)^{\mathbf{x}_2^{[j]}}|j\rangle\right)/N =: |\phi_{\mathbf{x}_1}\rangle\otimes|\phi_{\mathbf{x}_2}\rangle$$

Symmetries: global correlations (circuit that remaps states)

$\sigma \in S_\Phi$ $\sigma : x_k \mapsto \sigma(x_k) \cong \bar{x}_k$

Quantum symmetry representations (conserved operators)

$\prod_{i=1}^{n} \mathrm{SWAP}_{i,i+n}$ $\hat{U}_\Phi = Y^{\otimes n} \otimes Y^{\otimes n}$

register exchange **bitstring remapping**

We test the ability of classical neural network architectures and QML-based approaches to learn the distribution labels with exponential improvements in generalization.

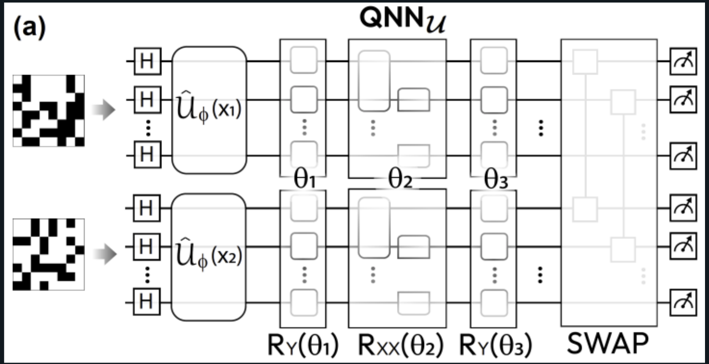
SYMMETRY-AWARE: BARCODES

Following the generic rules of symmetry-aware processing, we developed and followed 2 different **GQML workflows**: variational **basis adaption** and **measurement selection**.

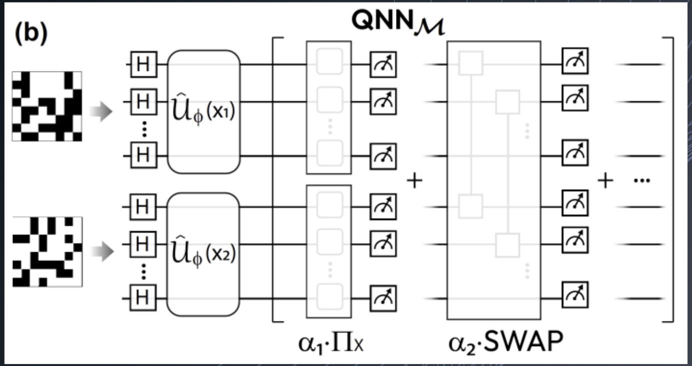
[C. Umeano, S. Scali, OK, arXiv:2409.01496 (2024)]

Pool of symmetric operators: $\left\{\sum_i \hat{Y}_i, \sum_i \hat{X}_i\hat{X}_{i+1}, \sum_i \hat{Y}_i\hat{Y}_{i+1}, \sum_i \hat{Z}_i\hat{Z}_{i+1}\right\}$ $\left\{\hat{X}^{\otimes 2n}, \hat{Z}^{\otimes 2n}, \prod_{i=1}^{n} \mathrm{SWAP}_{i,i+n}, \hat{H}^{\otimes 2n}, \dots\right\}$

local generators **global generators**



basis adaptation circuit for spotting correlations



measurement selection circuit for spotting correlations

Hypotheses:

$$h_\theta(x_m) = \langle\psi_0|\hat{\mathcal{U}}_\phi^\dagger(x_m)\hat{W}^\dagger(\theta)\hat{O}\hat{W}(\theta)\hat{\mathcal{U}}_\phi(x_m)|\psi_0\rangle$$

MSE-based training $\frac{1}{M}\sum_{m=1}^{M}(ah_\theta(x_m)+b-y_m)^2$

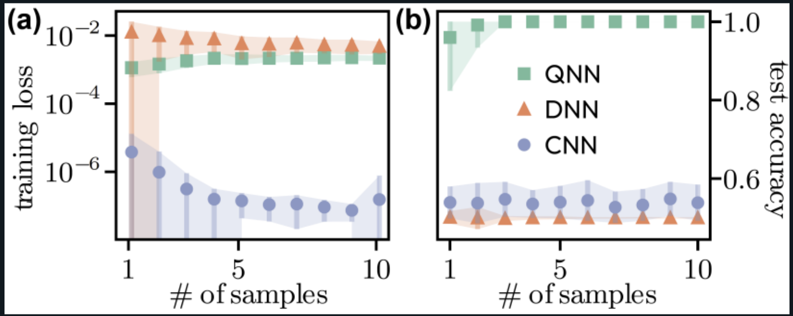
$$h(x_m) = \boldsymbol{\alpha}\cdot\phi(x_m) \leftarrow [\langle\hat{O}_1\rangle_m, \langle\hat{O}_2\rangle_m, \dots, \langle\hat{O}_K\rangle_m]$$

$$\text{LASSO}: \quad \min_{\alpha} \frac{1}{2M}\sum_{m=1}^{M}\left(\boldsymbol{\alpha}\cdot\phi(x_m) - y_m\right)^2 + \lambda\|\boldsymbol{\alpha}\|_{l_1}$$

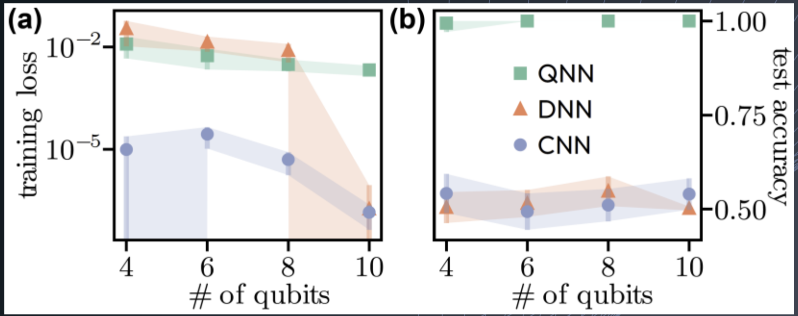
SYMMETRY-AWARE: BARCODES

We proceed to test QML-based workflows on created datasets of correlated/uncorrelated pairs, and benchmark them against classical deep neural networks (**DNN**) and convolutional neural networks (**CNN**) as a function of: **1)** training set; **2)** system size.

Basis adaptation **QML_U** does not perform well, but measurement selection **QML_M** shows excellent generalization.



QML_M-based offers 100% test accuracy when trained on few samples, while DNN/CNN do not generalize



excellent accuracy and generalization is observed at increased system size with 10^{308} possible states

---

To explain this performance we shall look at the problem of **forrelation** by Aaronson & Ambainis [arXiv:1411.5729 (2014) + Raz & Tal] that shows the maximal separation between BQP and PH.

$O(1)$ **vs** $\tilde{\Omega}(\sqrt{N})$ quantum and classical query complexity for distinguishing random and Fourier-related pairs

$$\tilde{\mathbf{x}}_2 \approx \hat{H}^{\otimes n}\tilde{\mathbf{x}}_1$$

correlated pairs

$$F = |\langle\phi_{\mathbf{x}_1}|\hat{H}^{\otimes n}|\phi_{\mathbf{x}_2}\rangle|^2 = |\langle 0|\hat{H}^{\otimes n}\hat{U}^{\dagger}_{\phi}(\mathbf{x}_1)\hat{H}^{\otimes n}\hat{U}_{\phi}(\mathbf{x}_2)\hat{H}^{\otimes n}|0\rangle|^2$$

overlap-based forrelation test

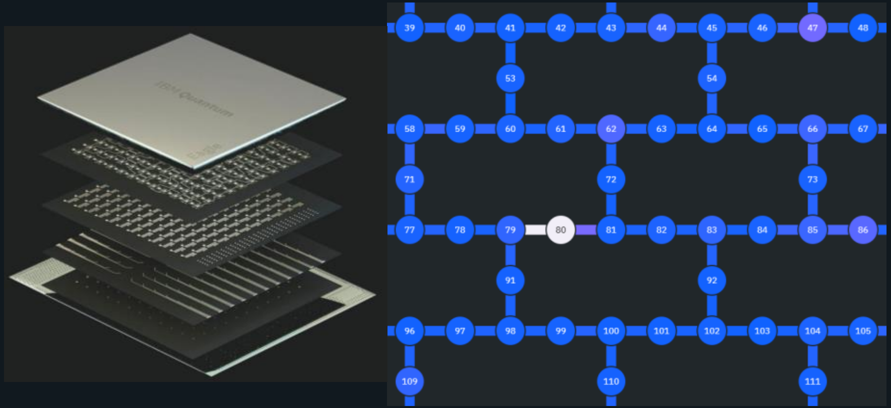
$$\hat{O}(\alpha^*) = \hat{H}^{\otimes 2n}\cdot\left(\prod_{i=1}^{n}\mathrm{SWAP}_{i,i+n}\right)$$

optimal hypothesis

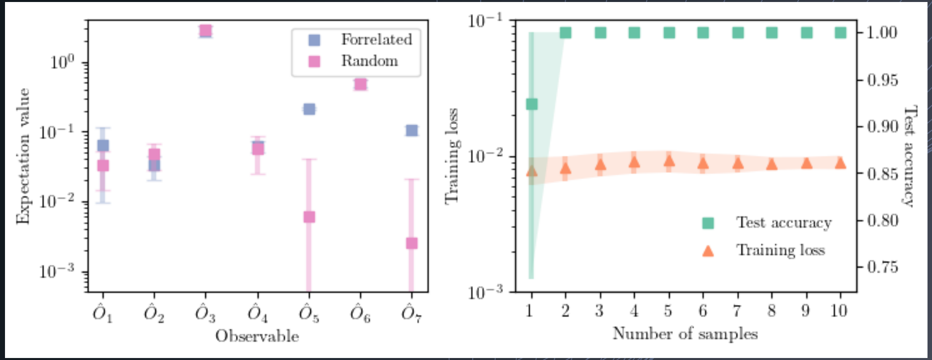
SYMMETRY-AWARE: BARCODES

All good, but will this actually work in practice? Can we learn a sharp decision boundary on **real hardware**?

[C. Umeano, S. Scali, OK, arXiv:2409.01496 (2024)]



IBM Eagle QPU (Kyiv) with 127 qubits: tested full QML_M approach on 40 qubits



measured observables for 2 classes and test accuracy from hardware predictions

- ○ GQML-based solution remains robust even in the presence of noise
- ○ similar approaches can be applied for cases of similarity testing where correlations are important

By targeting the barcode classification problem motivated by forrelation we have shown that QML can use symmetry-aware workflow to find optimal decision boundary with just few samples, and implemented on 40-qubit superconducting QPU.

University of Sheffield

“Success of quantum scientific machine learning largely depends on designing models that are physics-aware and obey relevant symmetries by construction, offering superior generalization. Inductive bias and powerful embedding can bring QuaSciML to next level.