Quantum Computing for Machine Learning

Examples from High Energy Physics

QUANTUM
TECHNOLOGY
INITIATIVE

Sofia.Vallecorsa@cern.ch

Alpaca Workshop – November 2023

Outline

- Introduction
  - Qubits and algorithms
- The CERN Quantum Technology Initiative
- Quantum Machine Learning
  - Trainability and generalisation
- Example applications at CERN
  - Anomaly Detection
  - Beam Optimisation in linear accelerators
- Summary & Outlook

Quantum potential…

Principles of quantum mechanics enhance computations

Superposition leads to parallelism → **exponential speedup?**

Entanglement → **non linear correlation and classical intractability?**

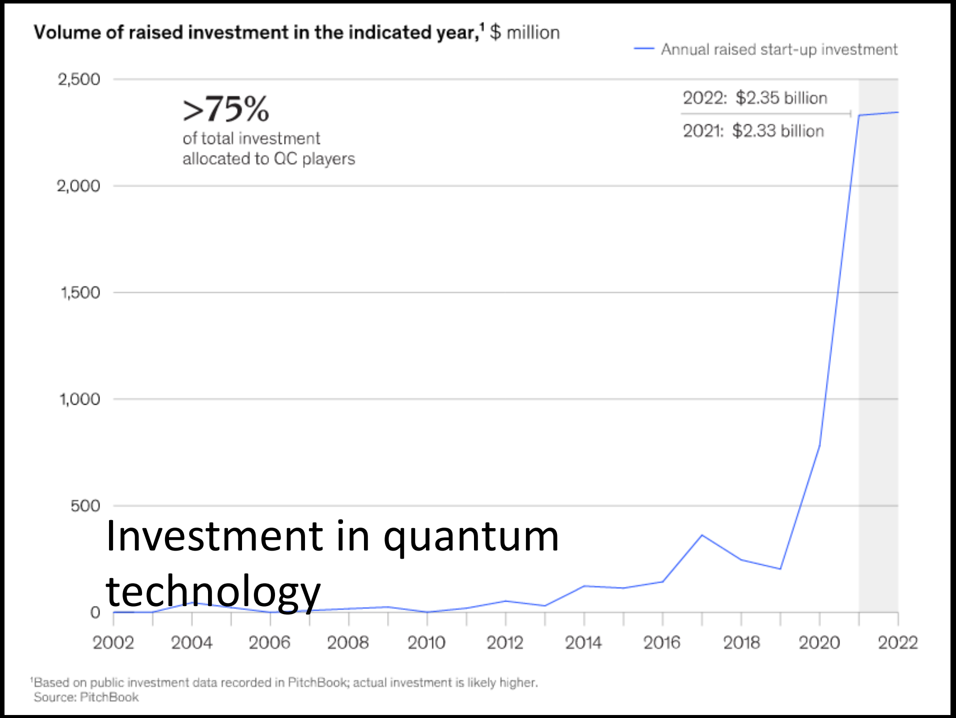
Operations (gates) are unitary transformations → **reversible computing?**

Output is the result of a quantum state measurement according to Born rule → **stochastic computation ?**

No-cloning theorem → **information security**

Quantum state coherence and isolation → **computation stability and errors**

Qubit state collapses → **reproducibility?**



Investment in quantum technology

[1] Based on public investment data recorded in PitchBook; actual investment is likely higher.
Source: PitchBook

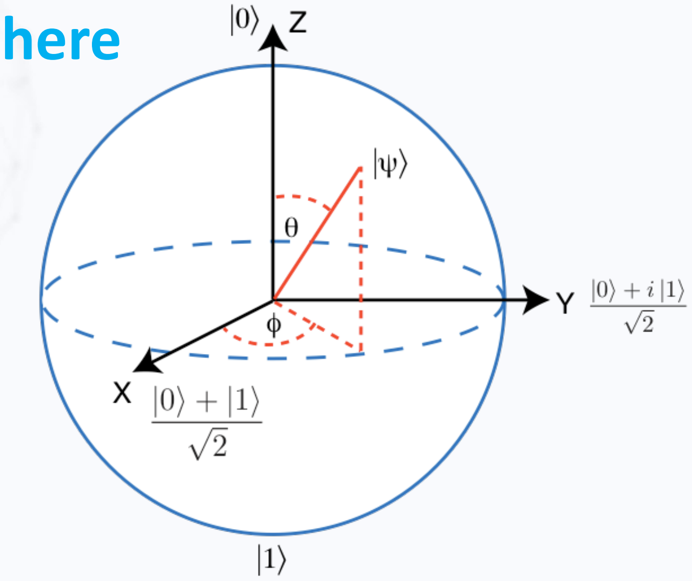
Source: McKinsey 2023
https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/quantum-technology-sees-record-investments-progress-on-talent-gap

Qubits and algorithms

- Basic Unit of Quantum Computation
  - **Classical bits are binary** “0 or 1”
- Quantum Mechanics predicts **superposition states**
- **Dirac notation**

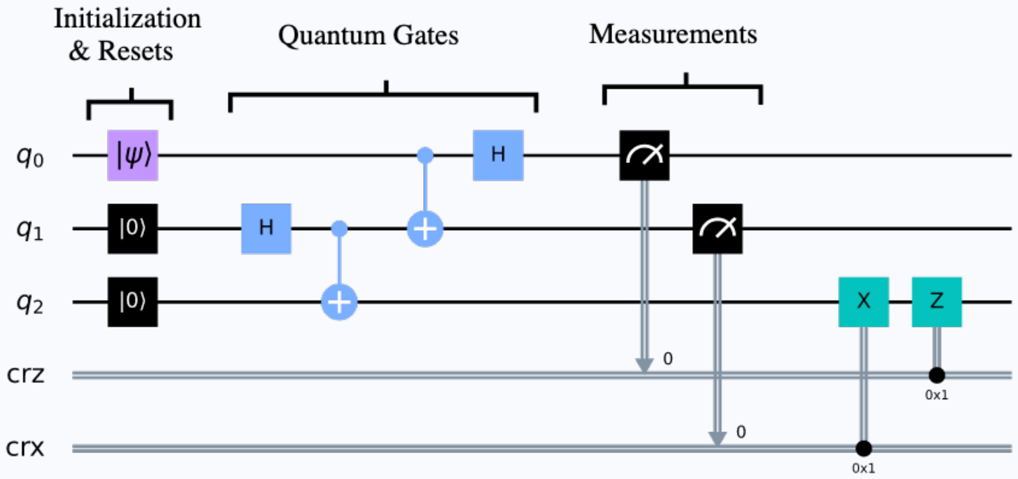
$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$$

Bloch Sphere



$$|\psi\rangle = \cos\frac{\theta}{2}\,|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}\,|1\rangle$$

- Operations are **unitary** matrices
  - Input and output states have the **same dimension**
  - Some classical gates (or, and, nand, xor...) **cannot be implemented directly**
  - Can **simulate** any classical computation with small overhead

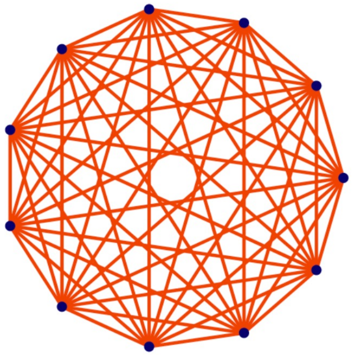


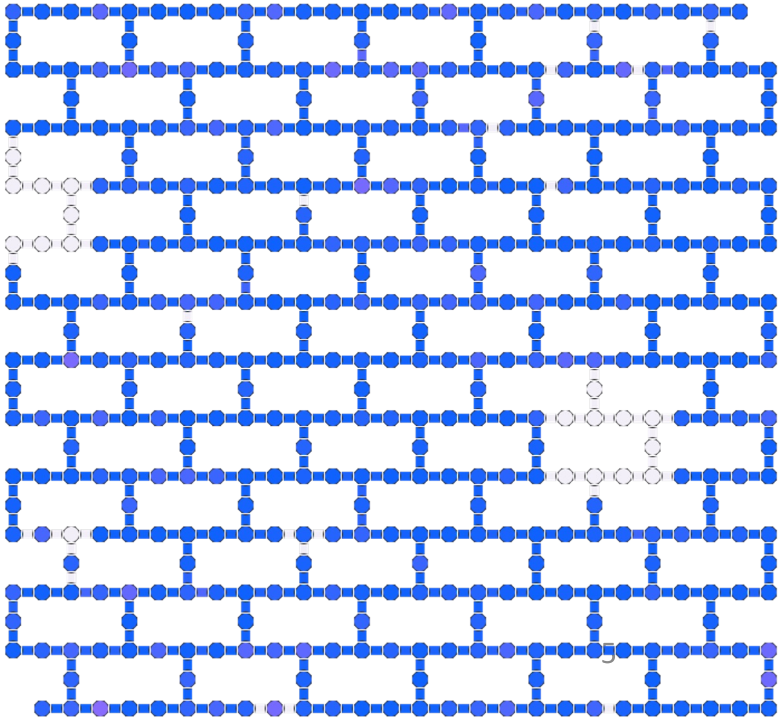
Interest in **multi level** representations: qutrits..

Noisy Intermediate-Scale Quantum devices

- Limitations in terms of **stability** and **connectivity**
    - **Circuit optimisation**
- **De-coherence**, measurement errors or gate level errors (**noise**)
    - Specific **error mitigation techniques**
    - Prefer algorithms **robust against noise**
- **Problem size**
- Initially integrated in **hybrid quantum-classical infrastructure (HPC + QC)**
    - **Quantum Processing Units** as new “hardware accelerators”

Trapped ion technology: *ionQ*
with all-to-all connectivity





Superconducting qubits:
IBM Seattle

- How do we define advantage?
  - Speed-up and complexity
  - Sample efficiency
  - Representational power
  - Energy efficiency???

- Evaluate performance on realistic use cases

2019: Google

nature

QUANTUM SUPREMACY

Classical supercomputer outperformed by quantum chip for the first time

https://www.nature.com/articles/s41586-019-1666-5

The CERN Quantum Technology Initiative was launched in 2020

Voir en *français*

CERN meets quantum technology

The CERN Quantum Technology Initiative will explore the potential of devices harnessing perplexing quantum phenomena such as entanglement to enrich and expand its challenging research programme

30 SEPTEMBER, 2020 | By Matthew Chalmers

The AEgIS 1T antimatter trap stack. CERN's AEgIS experiment is able to explore the multi-particle entangled nature of photons from positronium annihilation, and is one of several examples of existing CERN research with relevance to quantum technologies. (Image: CERN)

Understanding the impact of quantum technologies in HEP

Quantum simulation and HEP theory applications

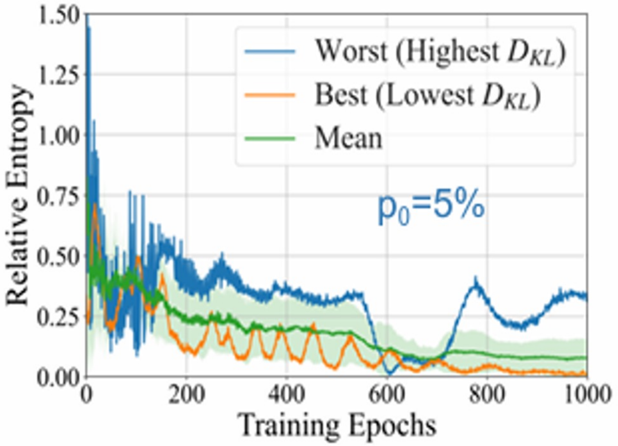
Quantum Computing

Quantum Sensing

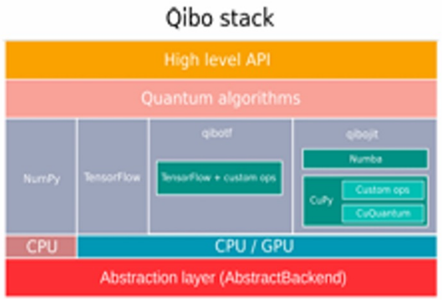
Quantum Communication

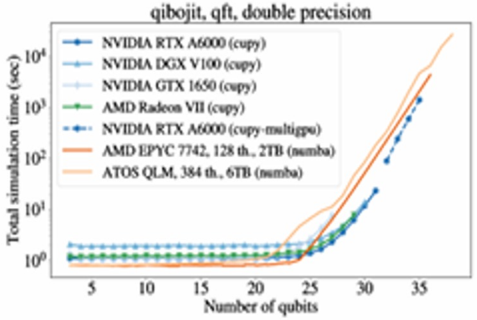
QC @CERN

Borras, Kerstin, et al. "**Impact of quantum noise on the training of quantum Generative Adversarial Networks.**" *arXiv preprint arXiv:2203.01007* (2022).

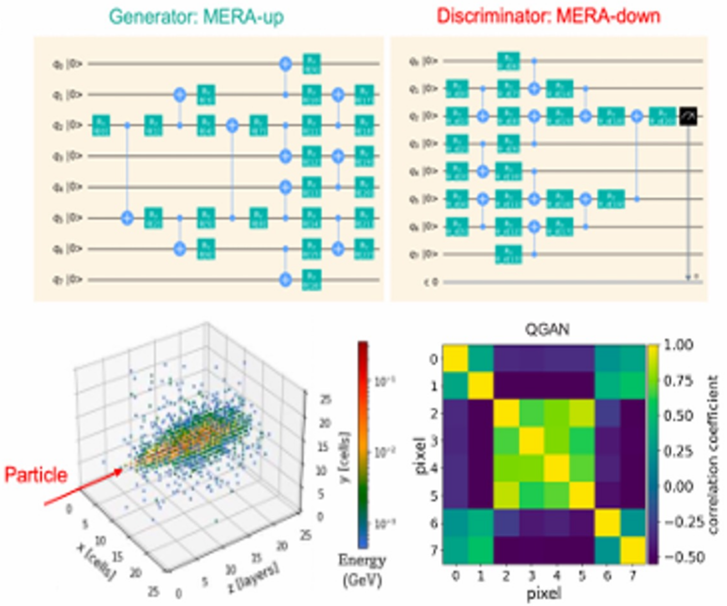


E.Stavros et all., **Quantum simulation with just-in-time compilation**, Quantum 2022

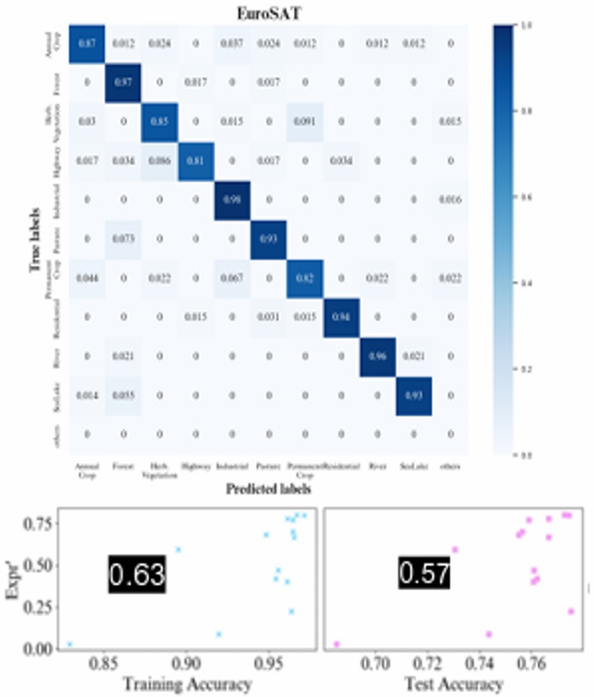




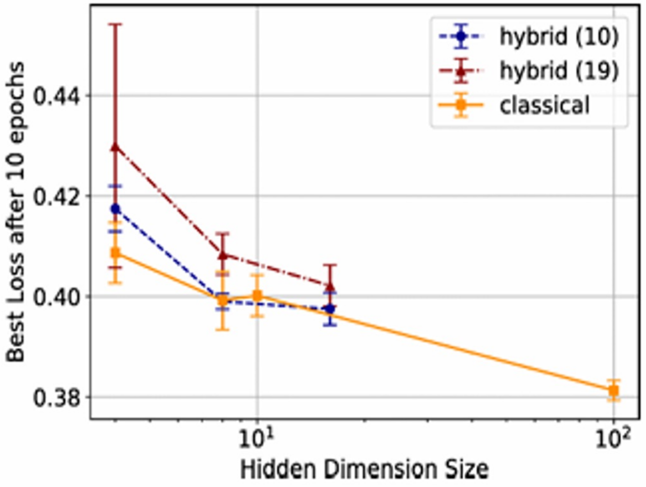
F.Rehm, **Full Quantum GAN Model for HEP Detector Simulations**, ACAT22



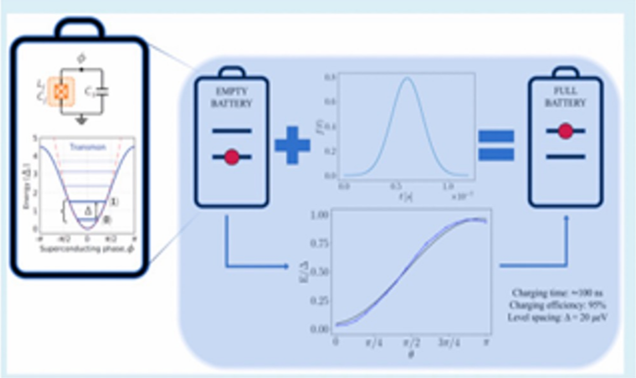
S.Chang, et all, **Hybrid Quantum-Classical Networks for Reconstruction and Classification of Earth Observation Images**, ACAT22



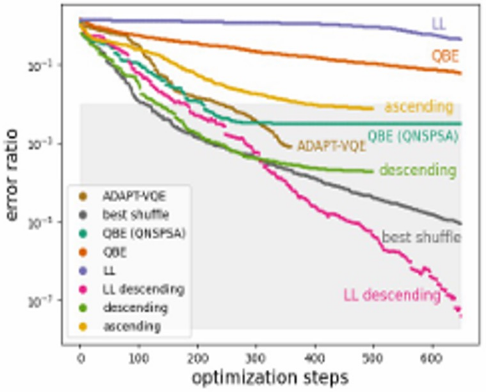
Tüysüz, Cenk, et al. "**Hybrid quantum classical graph neural networks for particle track reconstruction.**" *Quantum Machine Intelligence* 3.2 (2021): 1-20.



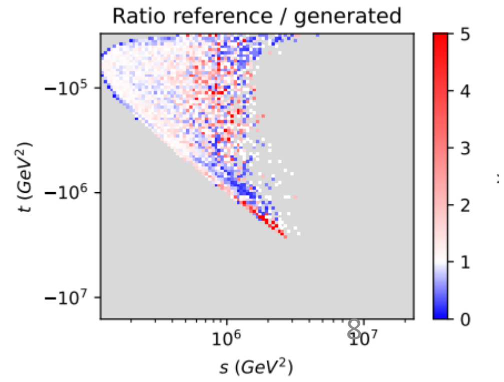
G. Gemme, M. Grossi et al, **IBM Quantum Platforms: A Quantum Battery Perspective**, *Batteries 8, 43 (2022)*



O. Kiss, **Quantum computing of the 6Li nucleus via ordered unitary coupled cluster**, 10.1103/PhysRevC.106.034325



Bravo-Prieto, Carlos, et al. "**Style-based quantum generative adversarial networks for Monte Carlo events.**" *Quantum 2022*

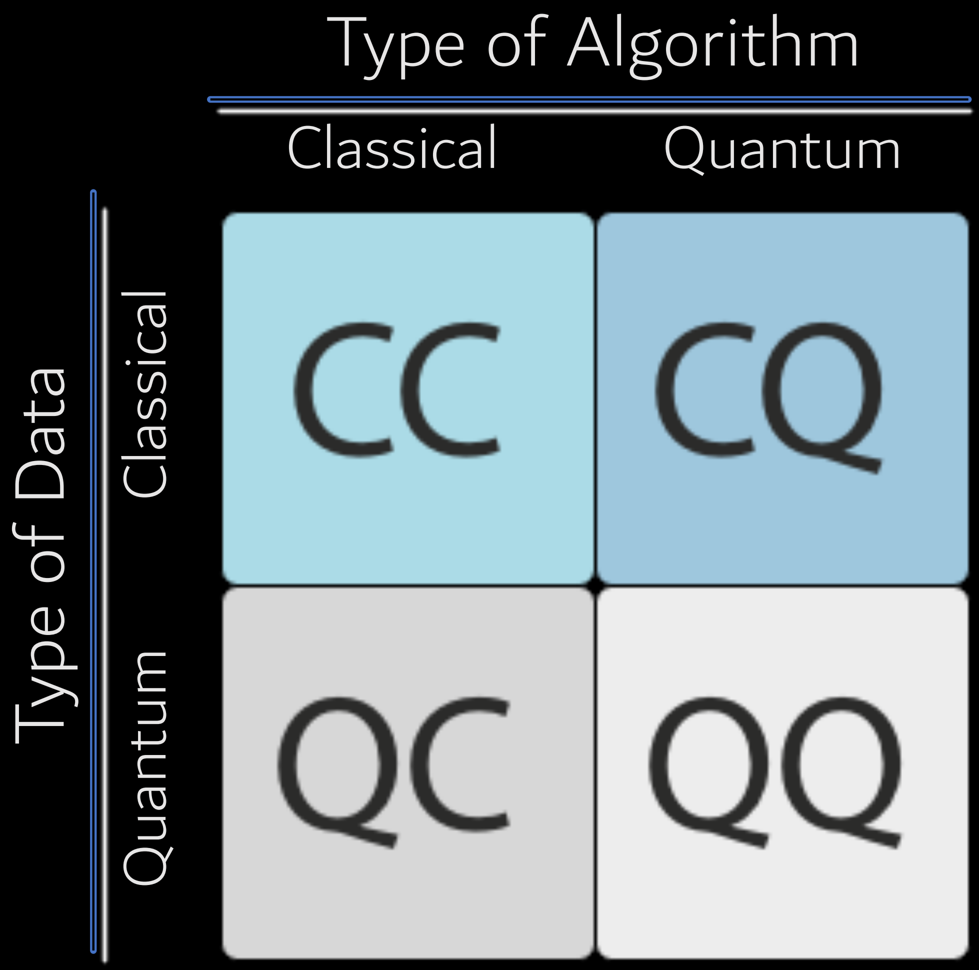


Quantum Machine Learning :

Some basic concepts

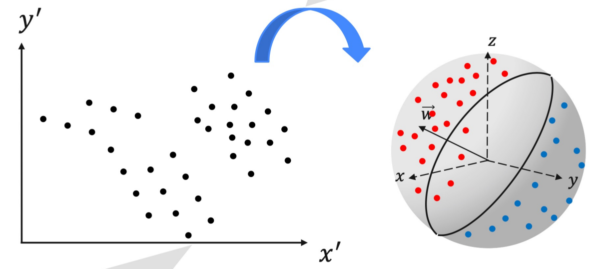
QML in HEP

- Does it make sense to use QML in HEP?
- How do we understand when it is *useful* ?
- Which are the QML models we can leverage?



		Type of Algorithm	
		Classical	Quantum
Type of Data	Classical	CC	CQ
	Quantum	QC	QQ

(Quantum) ML Lifecycle



Data Embedding

Data Preparation

Model Definition

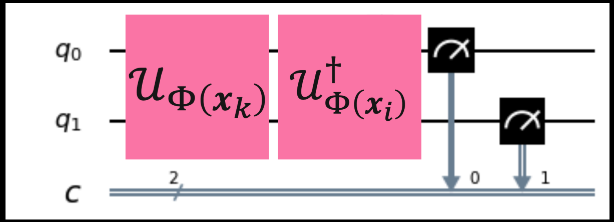
Model Training

Model Testing

Model Interpretation

Adapt classical learning models to quantum space

Readout and measurement "shots"



The advantage of many known QML algorithms is impeded today by I/O bottleneck

Quantum embedding for classical data

Compromise between **exponential compression and circuit depth**

Ex: **Amplitude Encoding**

$$|\phi(x)\rangle = \frac{1}{\|x\|}\sum_{i=0}^{N} x_i|i\rangle$$

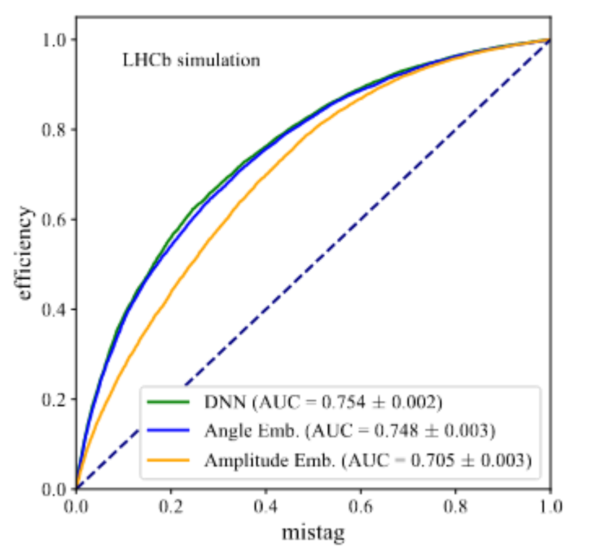
+ Exponential compression

$n_{qubit} \propto O(\log(N))$

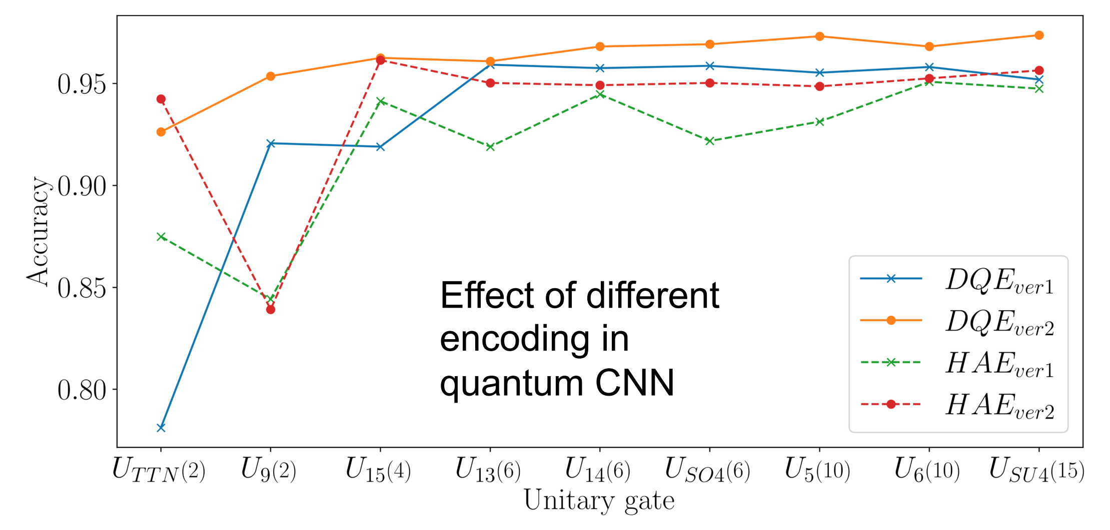
− Polynomial number of gates

$n_{gate} \propto O(poly(N))$

Gianelle, A., Koppenburg, P., Lucchesi, D. *et al.* **Quantum Machine Learning for *b*-jet charge identification**. *J. High Energ. Phys.* **2022,** 14 (2022). https://doi.org/10.1007/JHEP08(2022)014



S.Y. Chang, poster at "**Quantum Tensor Network in Machine Learning**, NeurIPS 2021



Effect of different encoding in quantum CNN

Models

Variational algorithms (ex. QNN)

Gradient-free or **gradient-based** optimization

Data Embedding can be **learned**

Ansatz design can leverage data symmetries[1]

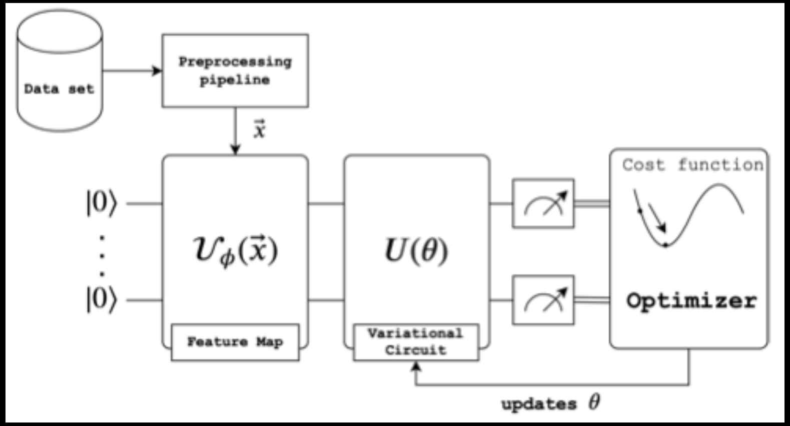


Image credit SwissQuantumHub

Representer theorem:

Implicit models achieve **better accuracy**[3]

Explicit models exhibit **better generalization** performance

Kernel methods (ex. QSVM)

Feature maps as **quantum kernels**

Classical **kernel-based training** (**convex** losses)

Identify classes of kernels that relate to specific data **structures**[2]

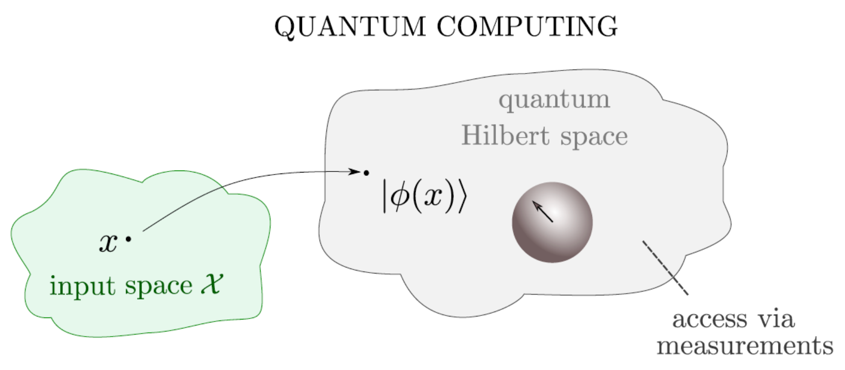


Image credit M. Schuld

Energy-based ML (ex. QBM)

Build network of stochastic binary units and optimise their energy.

QBM has quadratic energy function that follows the Boltzman distribution (Ising Hamiltonian)

---

[1] Bogatskiy, Alexander, et al. "**Lorentz group equivariant neural network for particle physics**." PMLR, 2020.

[2] Glick, Jennifer R., et al. "**Covariant quantum kernels for data with group structure**." *arXiv:2105.03406* (2021).

[3] Jerbi, Sofiene, et al. "**Quantum machine learning beyond kernel methods**." *arXiv preprint arXiv:2110.13162* (2021).

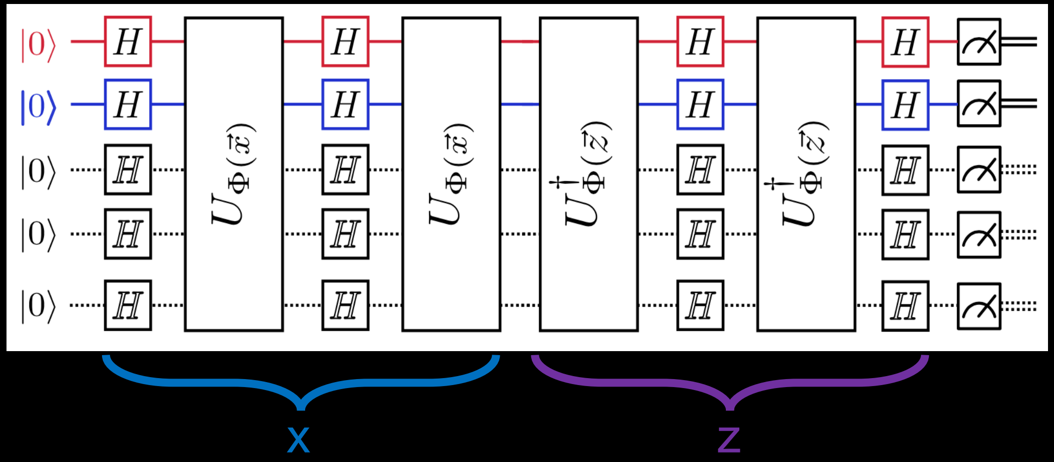
QML Convergence

Classical Intractability & expressivity vs trainability and generalization

F. Di Marcantonio *et al.* , CHEP2023

Quantum embedding and kernel methods

- Create **classically intractable features in the Hilbert space**
- Estimate **Fidelity kernel**
- Use classical training (convex losses)



Hilbert space is exponentially larger

↓

Sparser data

↓

Loss of predictive power

$$\hat{y} = \text{label}(z) = \text{sign}\left(\sum \alpha_i y_i K(x_i, z) + b\right)$$

$$|\langle \Phi(\vec{x}) | \Phi(\vec{z}) \rangle|^2 = |\langle 0^m | U^\dagger_{\Phi(\vec{x})} U_{\Phi(\vec{z})} | 0^m \rangle|^2$$

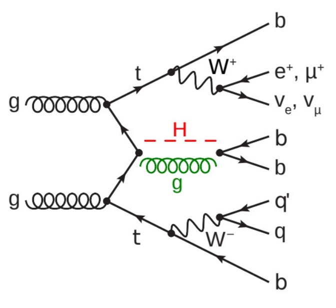
Projected Quantum Kernel

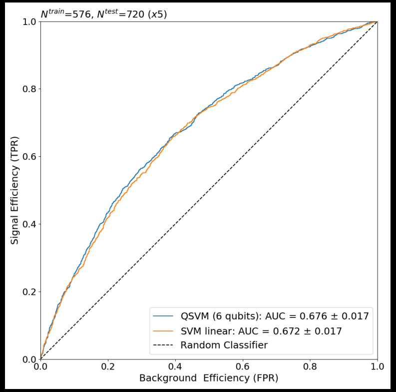
$$k^{lp}(x_i, x_j) = \sum_{k=1}^{m} \frac{Tr[\rho_k(x_i)\,\rho_k(x_j)]}{m}$$

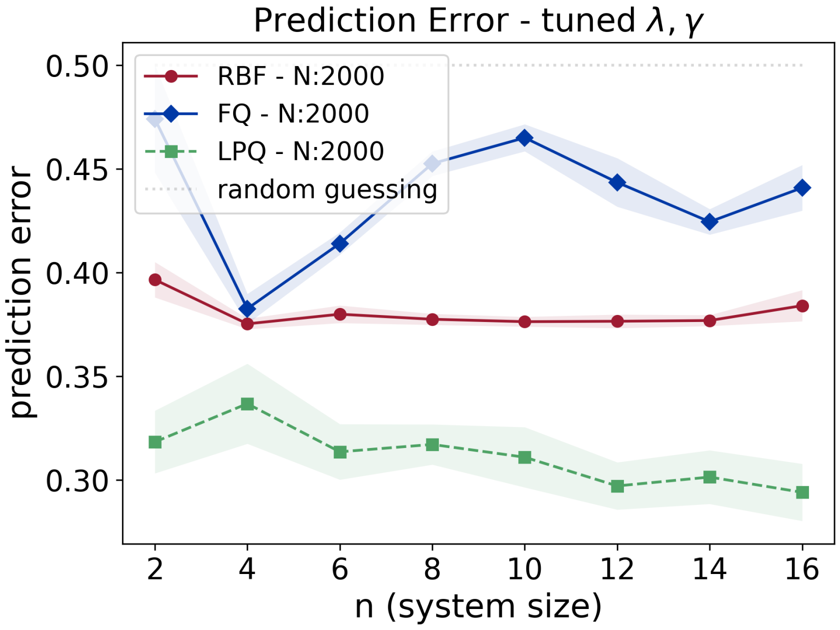
Project quantum kernels to lower dimensionality (i.e. local density matrix)[1]:

- Improved generalizion while keeping features into states classically hard

- Example: ttH(bb) binary classification[2]







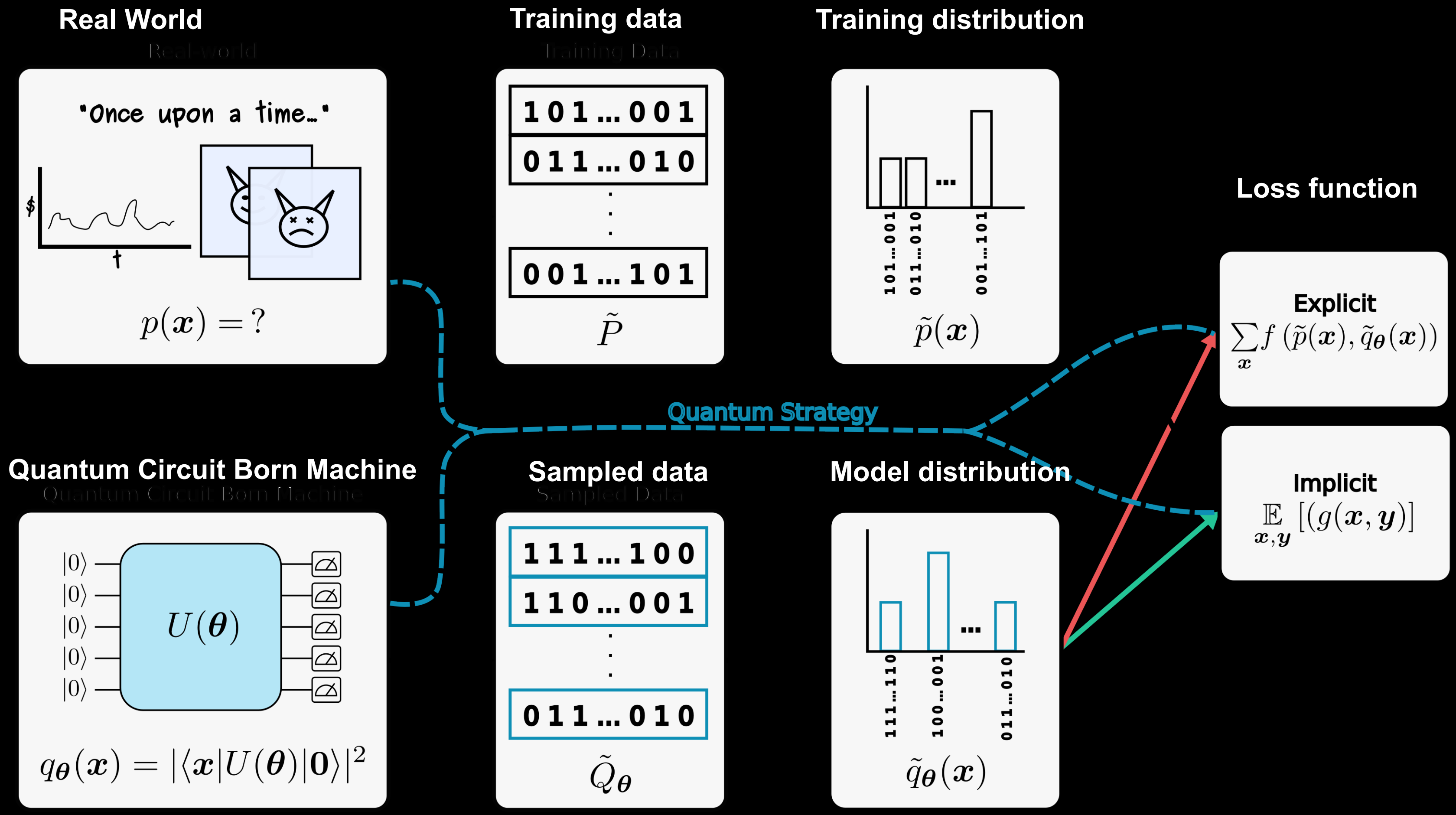
---

[1] Huang, Hsin-Yuan, et al. "Power of data in quantum machine learning." *Nature communications* 12.1 (2021): 2631.
[2] V Belis et al, (2021), *Higgs Analysis with Quantum Classifiers,* EPJ Web Conf

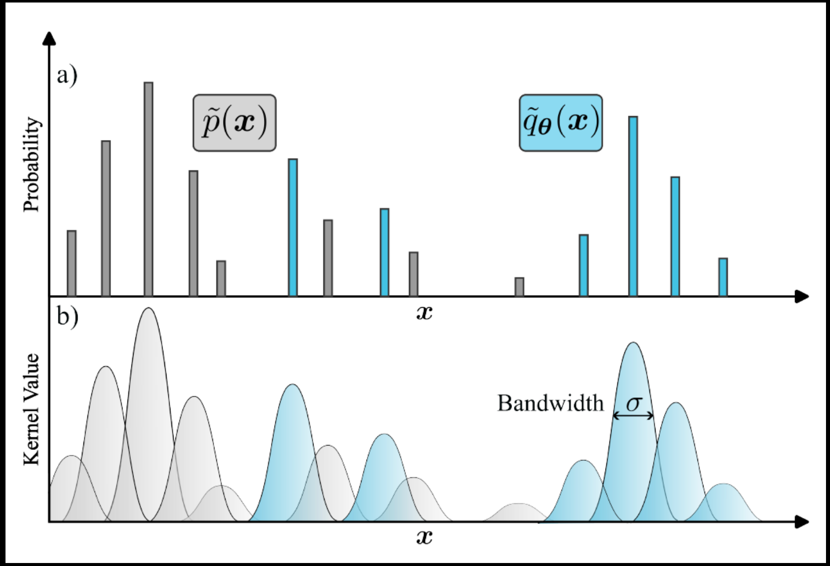
Rudolph, M. S., Lerch, S., Thanasilp, S., Kiss, O., Vallecorsa, S., Grossi, M., & Holmes, Z. (2023). **Trainability barriers and opportunities in quantum generative modeling.** *arXiv:2305.02881*.

Generative QML and trainability barriers

Representation learning: encoding probability distributions



exponentially larger number of shots is required to keep accuracy of explicit losses

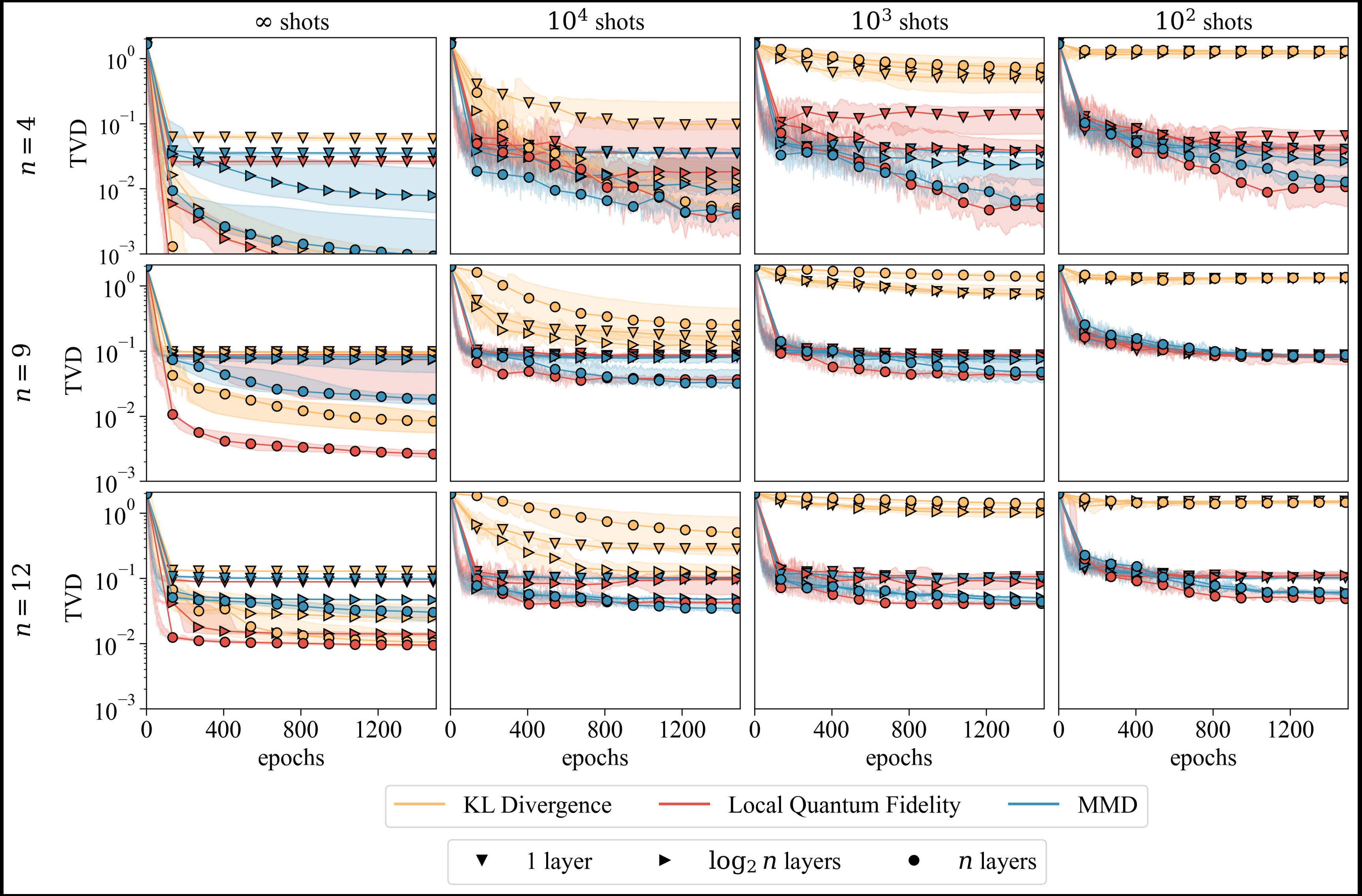


Rudolph, M. S., Lerch, S., Thanasilp, S., Kiss, O., Vallecorsa, S., Grossi, M., & Holmes, Z. (2023). **Trainability barriers and opportunities in quantum generative modeling.** *arXiv:2305.02881*.

Quantum Circuit Born Machine for HEP

QCBM

Sample variational pure state $|\psi(\theta)\rangle$ by projective measurement through **Born rule**: $\mathbf{p_\theta(x)} = |\langle \mathbf{x}|\boldsymbol{\psi}(\boldsymbol{\theta})\rangle|^2$.



Model Convergence and Barren Plateau

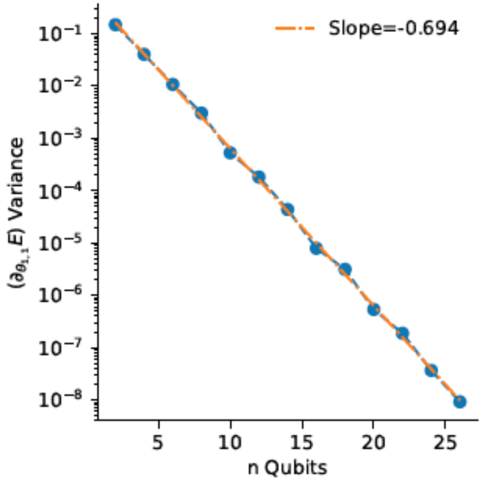
The size of the Hilbert space requires compromises between **expressivity**, **convergence** and **generalization**

Classical gradients **vanish exponentially** with the number of layers (J. McClean *et al.*, arXiv:1803.11173)

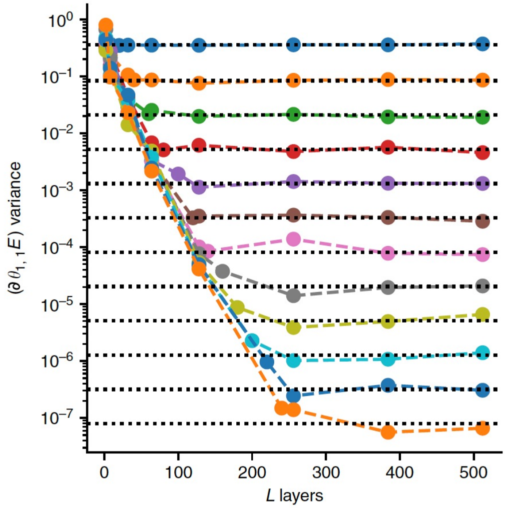
- Convergence still possible if gradients consistent between batches.

Quantum gradient decay exponentially in the number of qubits (number of graph paths is exponential in the number of gates)

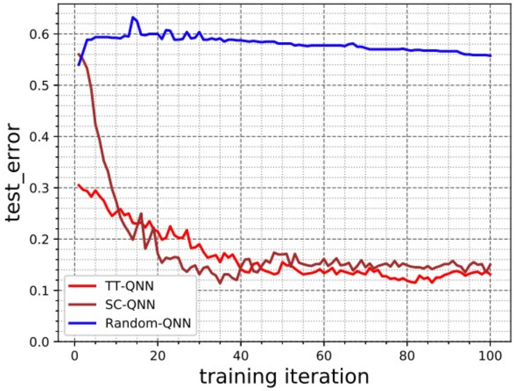
- Random circuit initialization
- Loss function locality in shallow circuits (M. Cerezo *et al.*, arXiv:2001.00550)
- Ansatz choice: TTN, CNN (Zhang *et al.,* arXiv:2011.06258, A Pesah, *et al.*, *Physical Review X* 11.4 (2021): 041011.)
- Noise induced barren plateau (Wang, S *et al.*, Nat Commun 12, 6961 (2021))



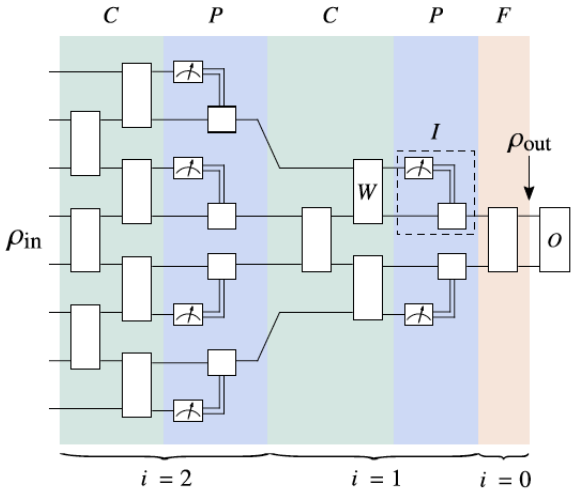
J. McClean *et al.*, arXiv:1803.11173



TTN for MNIST classification (8 qubits), Zhang *et al.,* arXiv:2011.06258

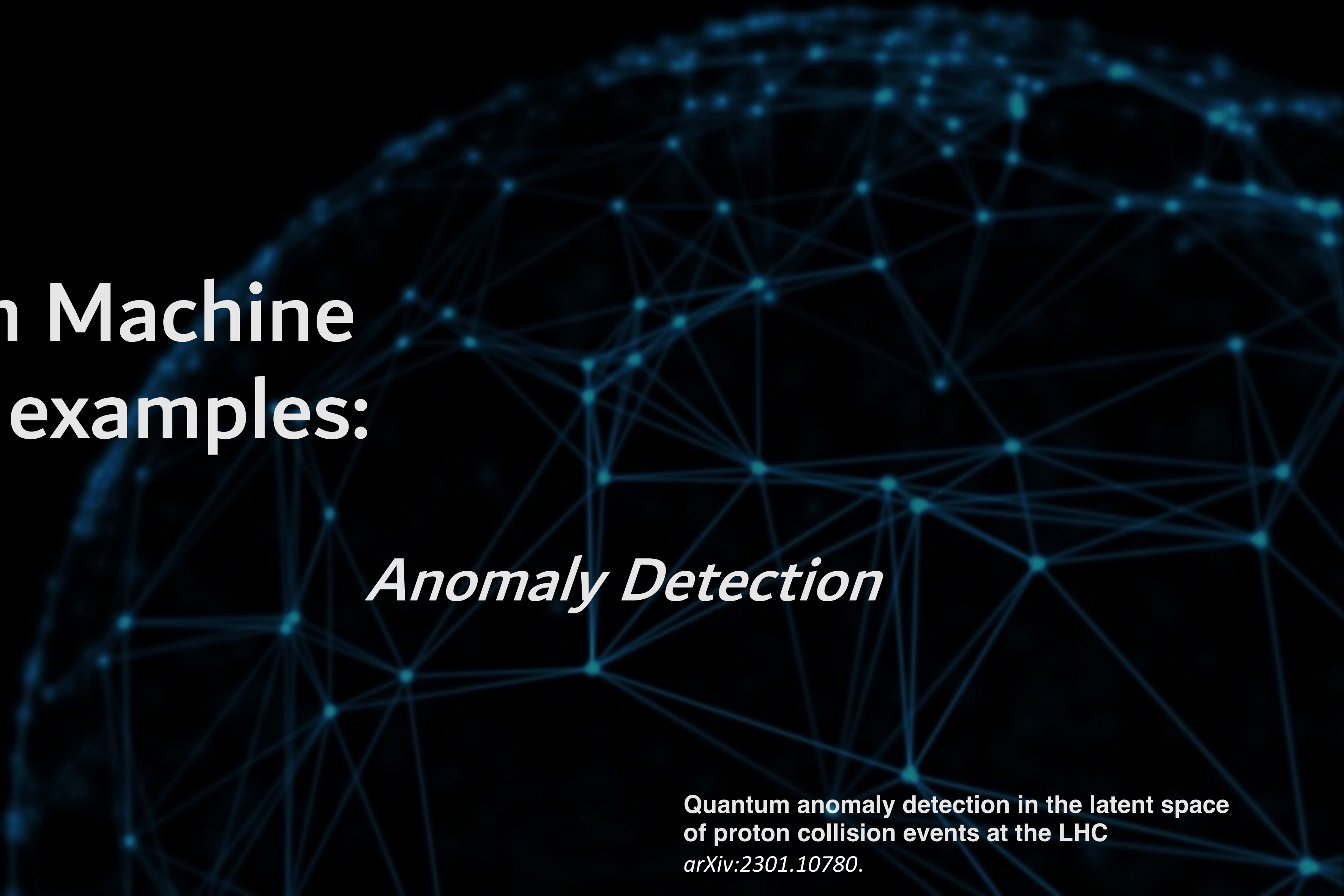


QCNN: A Pesah, *et al.*, *Physical Review X* 11.4 (2021): 041011



Quantum Machine Learning examples:

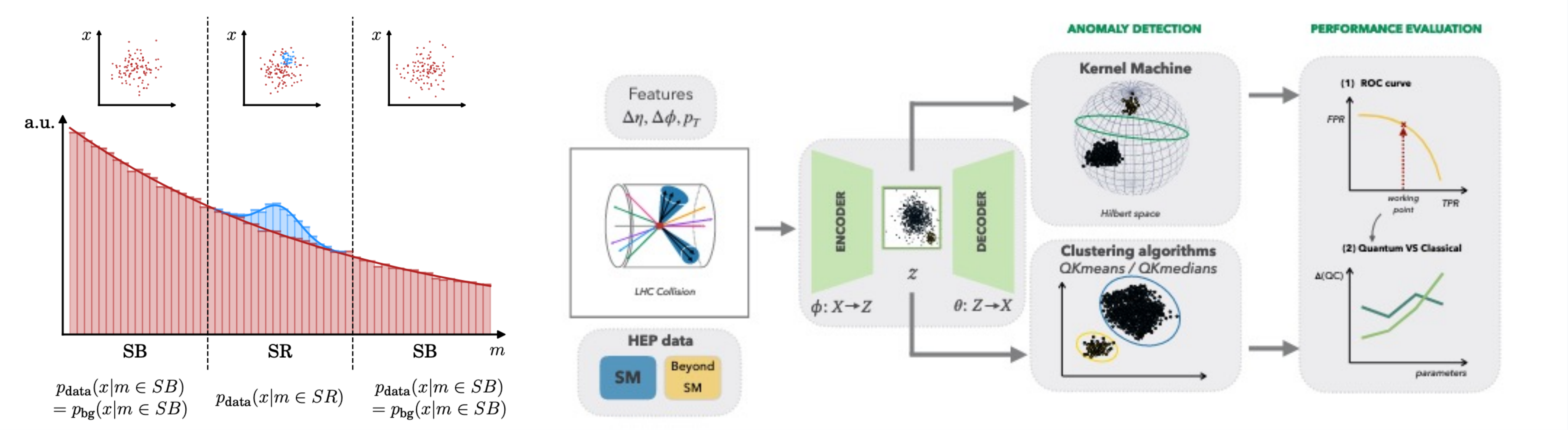
Anomaly Detection



Quantum anomaly detection in the latent space of proton collision events at the LHC
arXiv:2301.10780.

Unsupervised learning for Anomaly Detection

A typical hybrid QML workflow



Standard Model jets

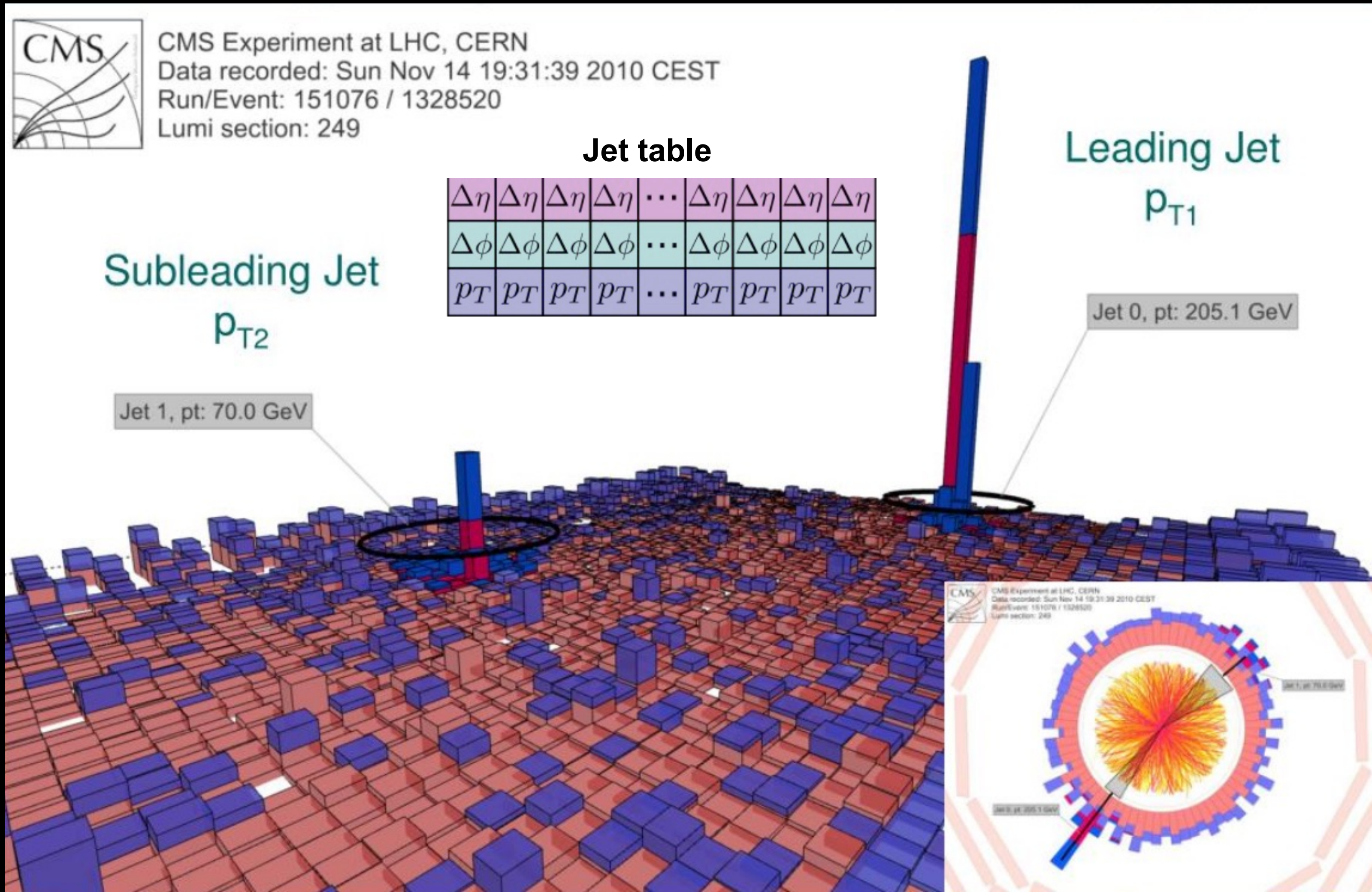
Simulate QCD multi-jets at the LHC

Build jet from 100 highest pt particles

Apply realistic event selection

Convolutional AutoEncoder
learns the jet **internal structure**

$$\mathbb{R}^{300} \rightarrow \mathbb{R}^{\ell}\,, \ell = 4, 8,16$$



Jet table

$\Delta\eta$	$\Delta\eta$	$\Delta\eta$	$\Delta\eta$	$\cdots$	$\Delta\eta$	$\Delta\eta$	$\Delta\eta$	$\Delta\eta$
$\Delta\phi$	$\Delta\phi$	$\Delta\phi$	$\Delta\phi$	$\cdots$	$\Delta\phi$	$\Delta\phi$	$\Delta\phi$	$\Delta\phi$
p_T	p_T	p_T	p_T	$\cdots$	p_T	p_T	p_T	p_T

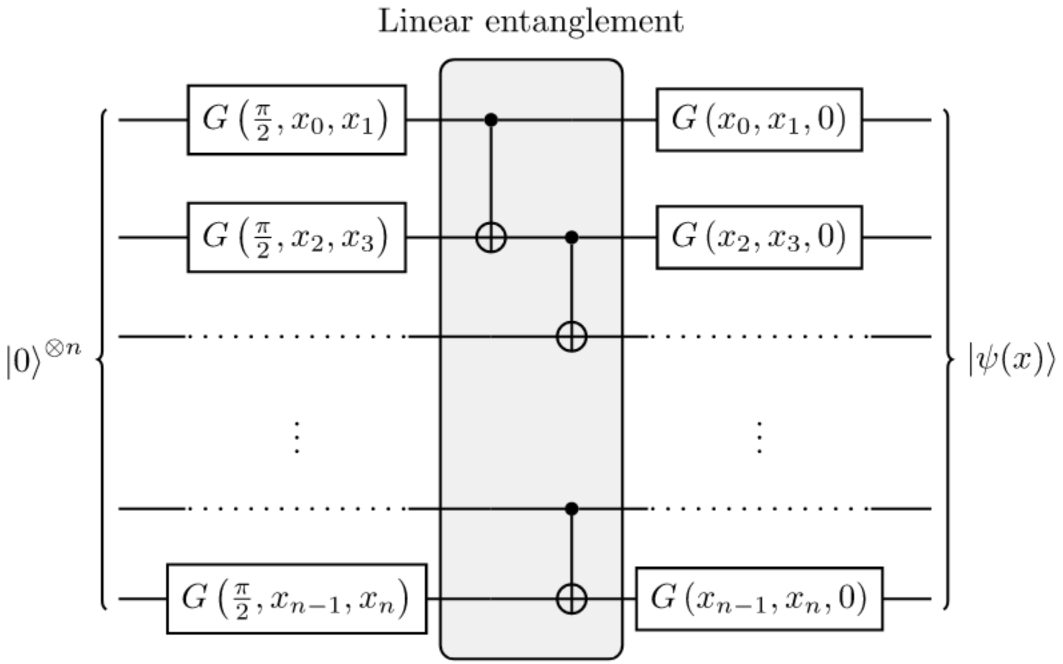
Unsupervised kernel machine

Find the hyperplane that maximizes the distance of the data from the origin of the feature vector space

Upper bound on fraction of anomalies in training data at 0.01 (at most 1% QCD training data are falsely flagged)

$$k(x_i, x_j) := \mathrm{tr}[\rho(x_i)\rho(x_j)] = \left|\langle 0|U^\dagger(x_i)U(x_j)|0\rangle\right|^2$$

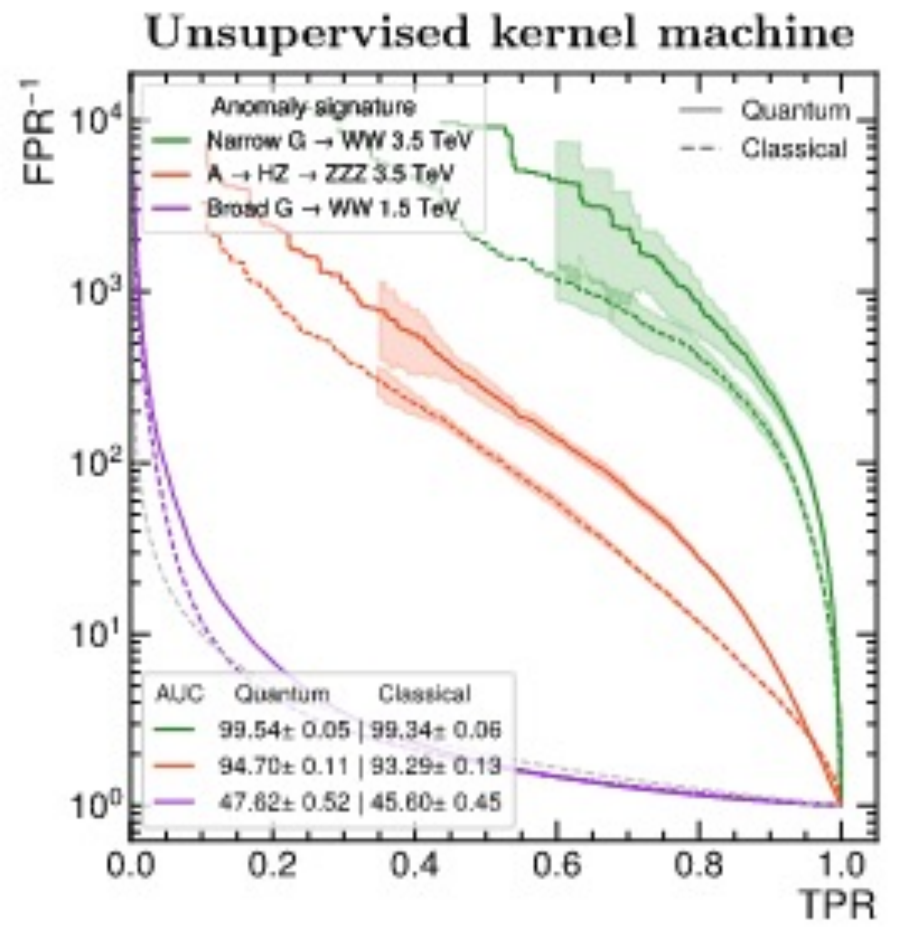
$$\rho(x_i) := U(x_i)\,|0\rangle\,\langle 0|\,U^\dagger(x_i)$$



$$\min_{w\in\mathcal{F},\, \xi\in\mathbb{R}^\ell,\, \rho\in\mathbb{R}} \quad \frac{1}{2}||w||^2 + \frac{1}{\nu\ell}\sum_i \xi_i - \rho$$

$$\text{subject to} \quad w\cdot\Phi(x_i) \geq \rho - \xi_i,\ \xi_i \geq 0,\ \forall i. \quad \nu \in (0,1)$$

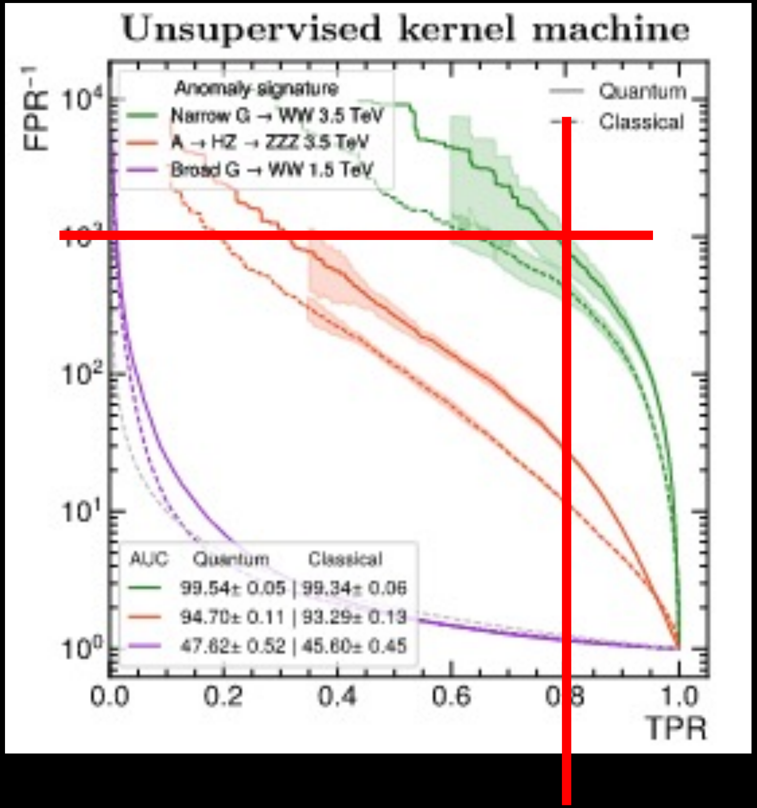
Results



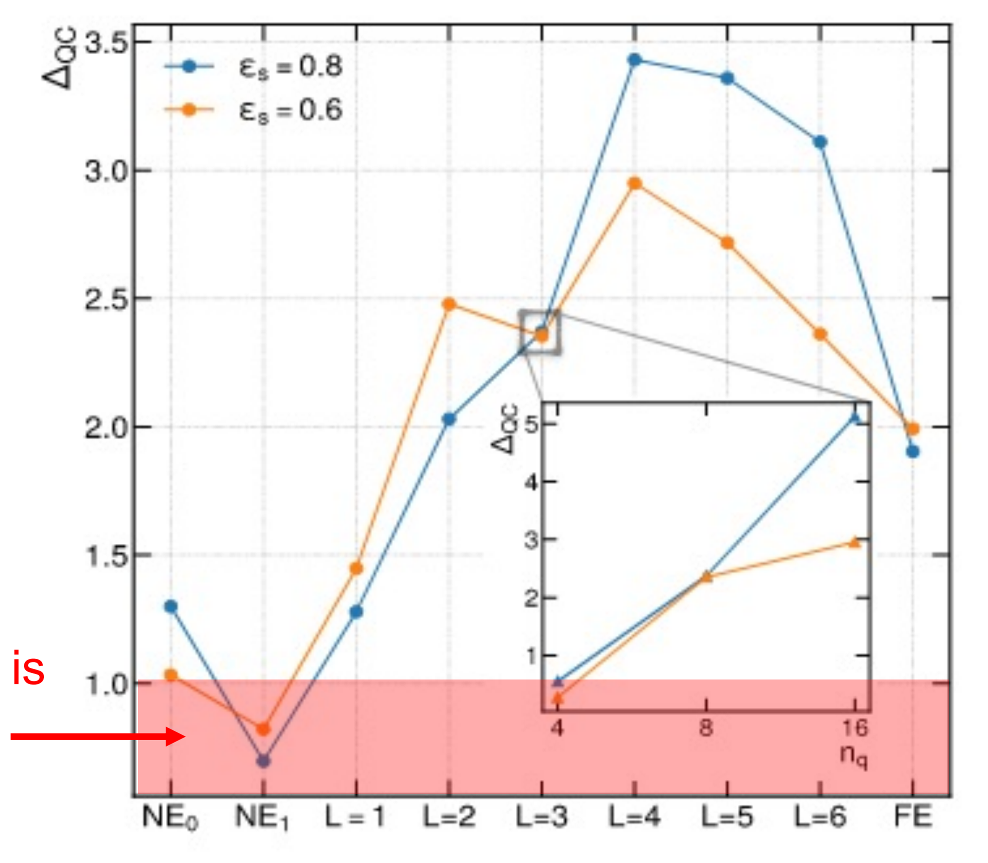
Is this an «advantage» we can use?

Quantum anomaly detection in the latent space of proton collision events at the LHC
arXiv:2301.10780.

In reality....



Classical is better



Higher is better

Increasing entanglement & expressivity

Quantum anomaly detection in the latent space of proton collision events at the LHC
arXiv:2301.10780.

Quantum Machine Learning examples:

Reinforcement Learning

Free-energy based RL (FERL)

RL performance depends **on type of Q-function approximator**

- **Classical Deep Q-learning (DQN)**
  Feed-forward neural net
- **Free-energy based RL (FERL)**
  Quantum Boltzmann machine (QBM)

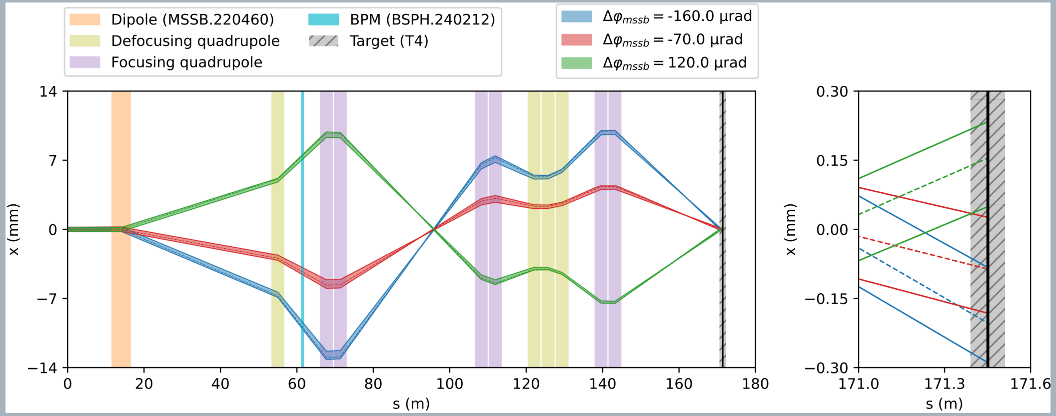
Key concept: sample-efficiency

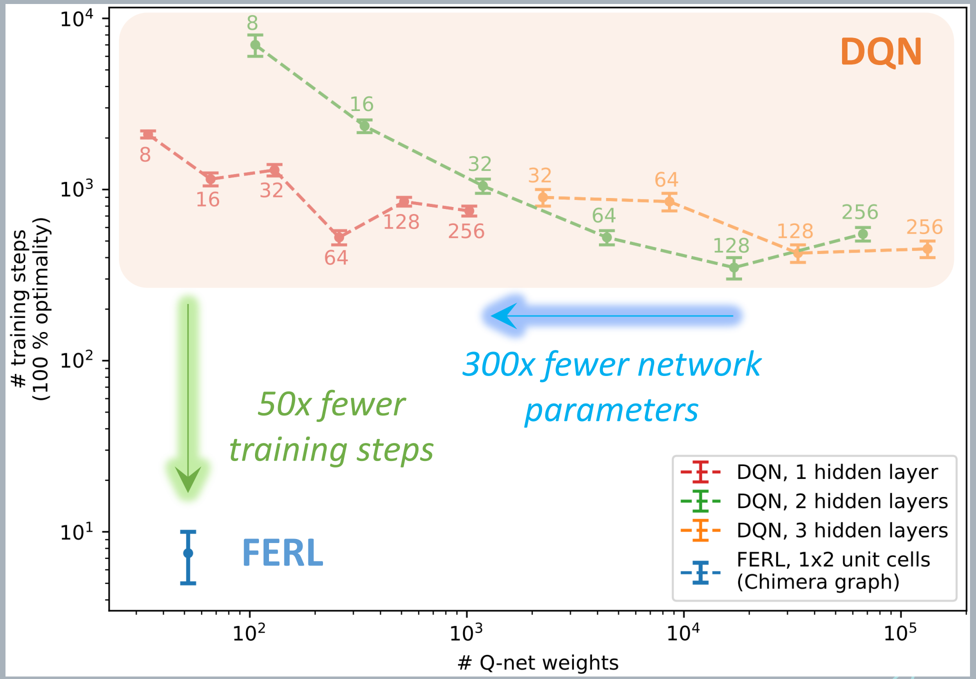
- Relevant for **particle accelerator control** given cost of beam time *(online training)*

Schenk, M *et al.* Hybrid actor-critic algorithm for quantum reinforcement learning at CERN beam lines. *arXiv preprint arXiv:2209.11044.*, CHEP2023

1st study: 1D beam steering

CERN *North Area transfer line (discrete action space)*





DQN

50x fewer training steps

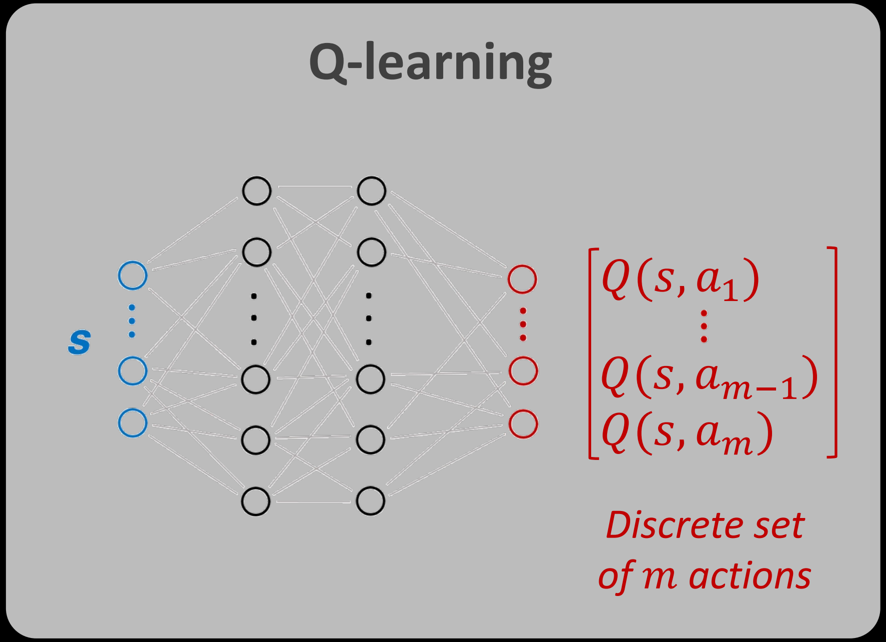
300x fewer network parameters

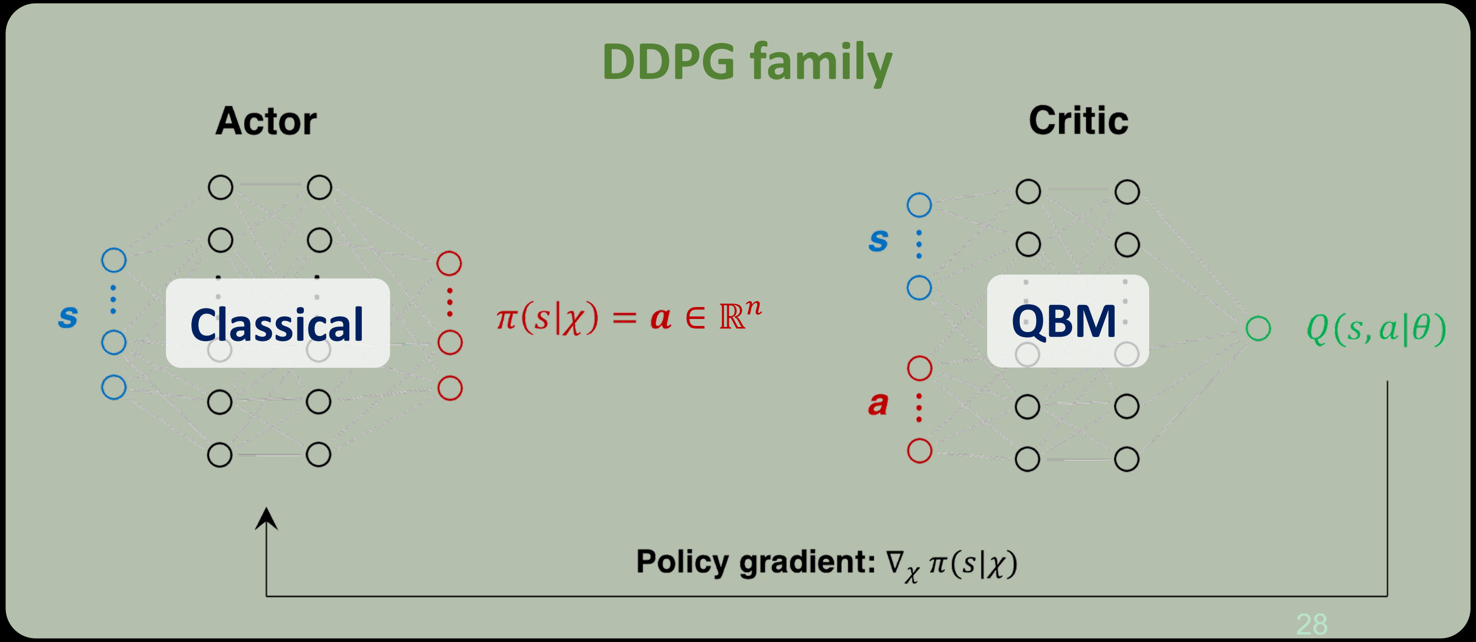
FERL

Developing a hybrid actor-critic scheme

Accelerator optimization requires **continuous action space** ⇨ **develop hybrid actor-critic algorithm**

- **QBM replaces classical critic net**



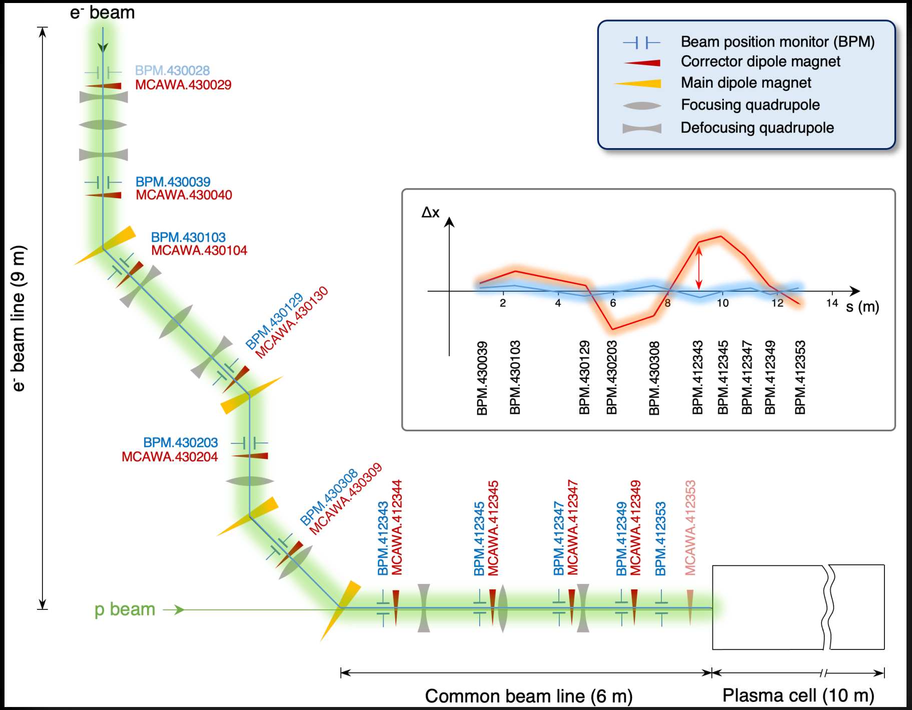


2nd study: 10D continuous beam steering

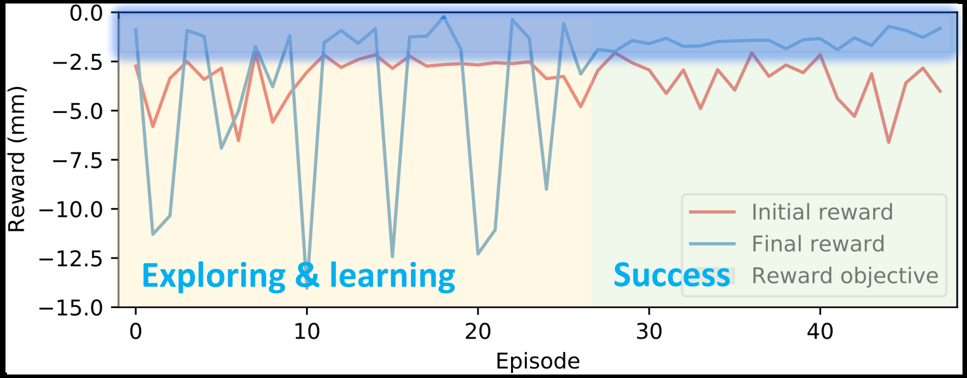
Environment: e- beam line of AWAKE

- **Action:** deflection angles at 10 correctors
- **State:** beam positions at 10 BPMs
- **Objective:** minimize beam trajectory rms
  - ⇨ **reward:** negative rms from 10 BPMs

***AWAKE:** Advanced Proton Driven Plasma Wakefield Acceleration Experiment*



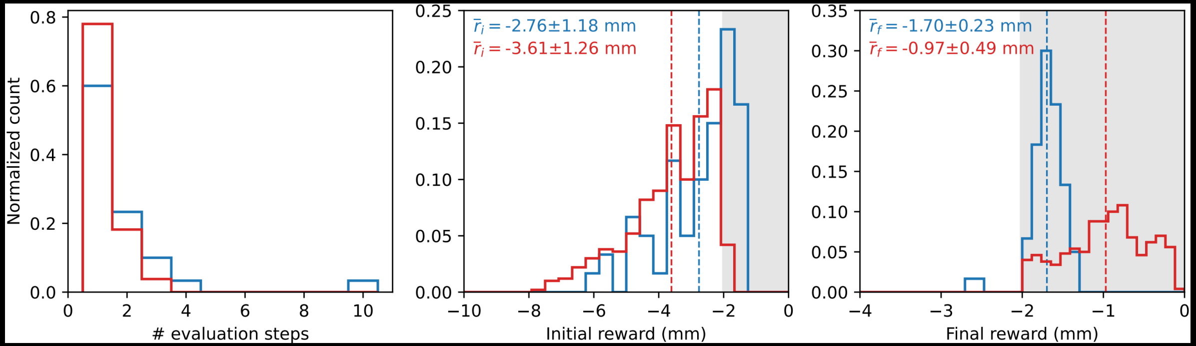
Training: on D-Wave Advantage quantum annealer (QA)



Objective

Evaluation: on actual beam line

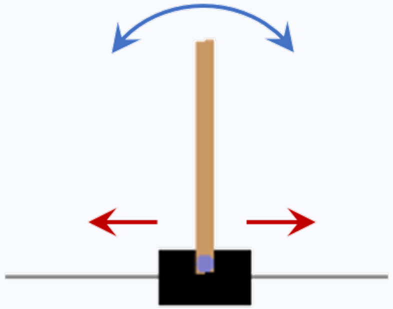
Real vs. simulated QA



- Agent minimizes rms in **1 step in 60 % cases**
- **Hyperparameter** tuning with **simulated** QA

3rd study: Cartpole-v1

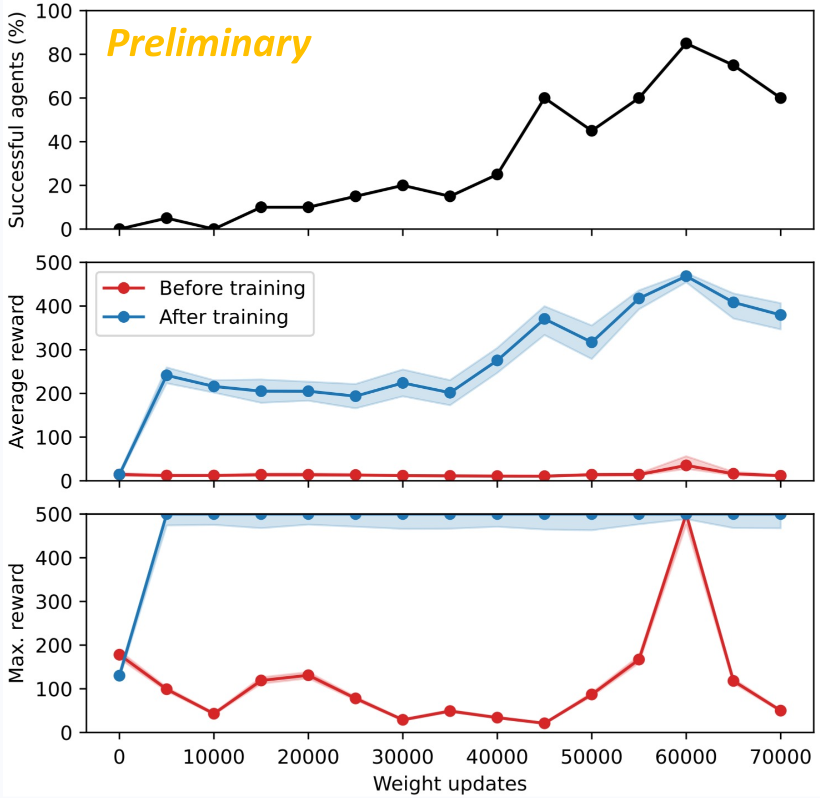
Discrete action problem, non-linear dynamics



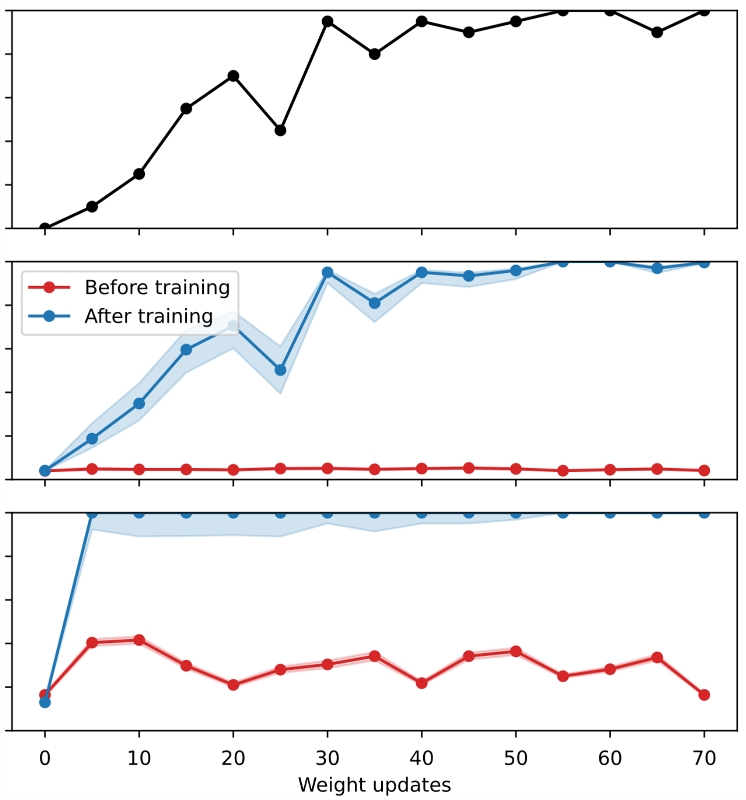
- **Cartpole-v1:** official OpenAI gym env from classic control problems domain
- Continuous state (4D), **discrete action** *(right, left)* problem with **non-linear dynamics**
- Terminate episodes after **max. 500 steps**

- **Big gain in sample-efficiency** and **robustness** for FERL vs DQN

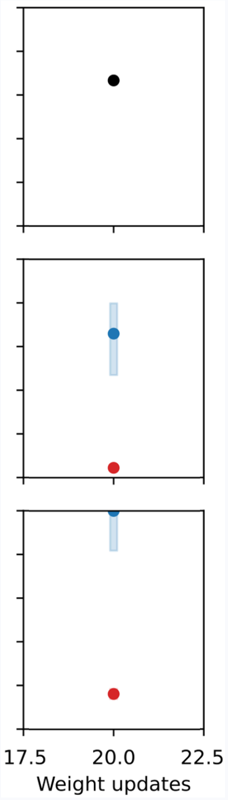
DQN

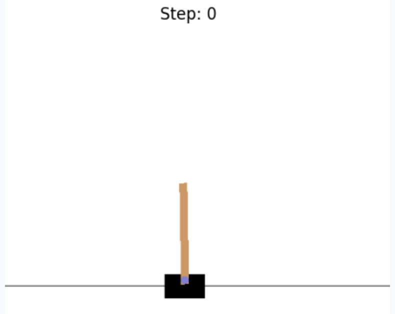


FERL *(simulated QA)*

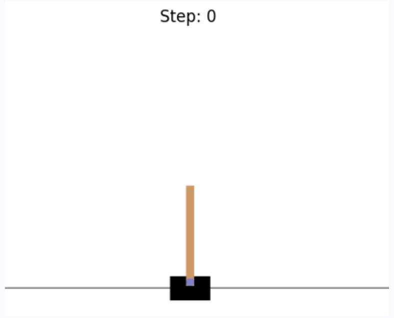


FERL *(trained on D-Wave)*





before



after

Quantum Machine Learning examples:

Phase Transitions identification

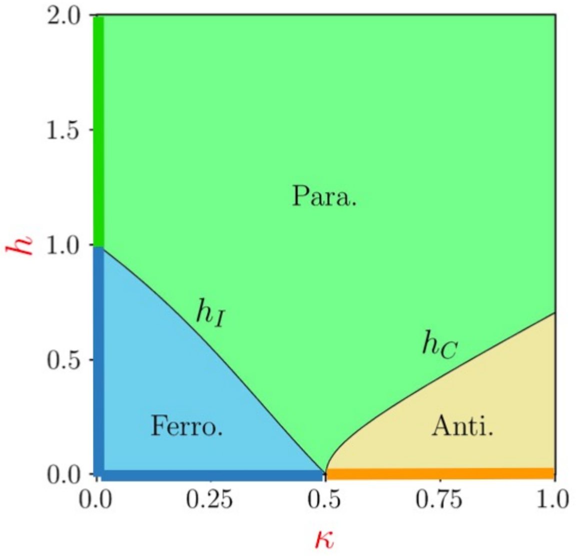
QML for quantum data: drawing phase diagrams

Model: Axial Next Nearest Neighbor Ising (ANNNI) Hamiltonian:

$$H = J \sum_{i=1}^{N} \sigma_x^i \sigma_x^{i+1} - \kappa \sigma_x^i \sigma_x^{i+2} + h \sigma_z^i,$$

Senk, *Physics Reports*, **170**, 4 (1988)

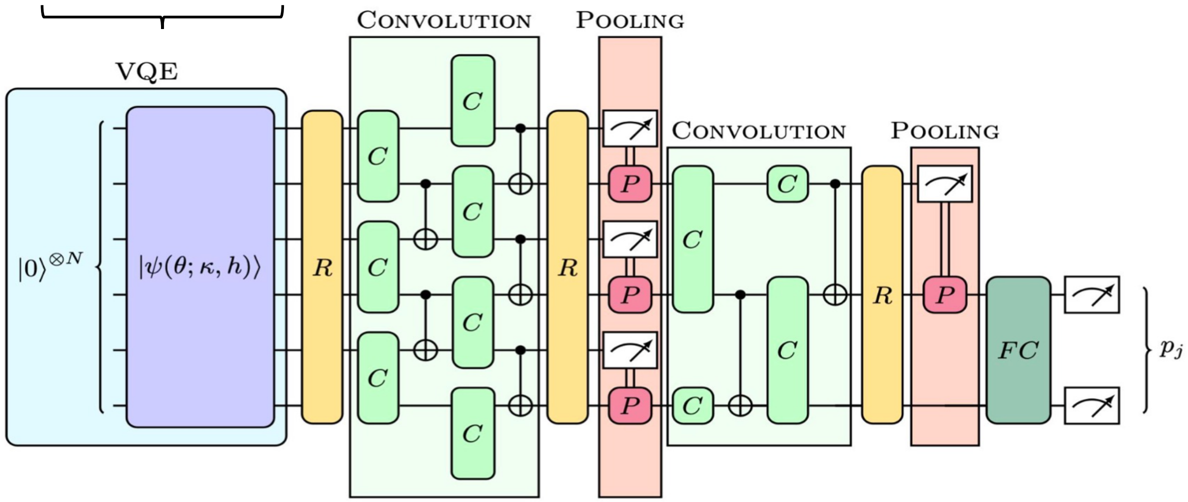
Integrable for $\kappa = 0$ or $h = 0$.



1. **Supervised classification of the ground state** using a convolutional QNN
2. Quantum states are **exponentially hard to save classically**.
3. **Bottleneck** from access to classical training labels (Interpolation does not work)
    - Train in integrable subregions
    - Generalize to a full model[1]

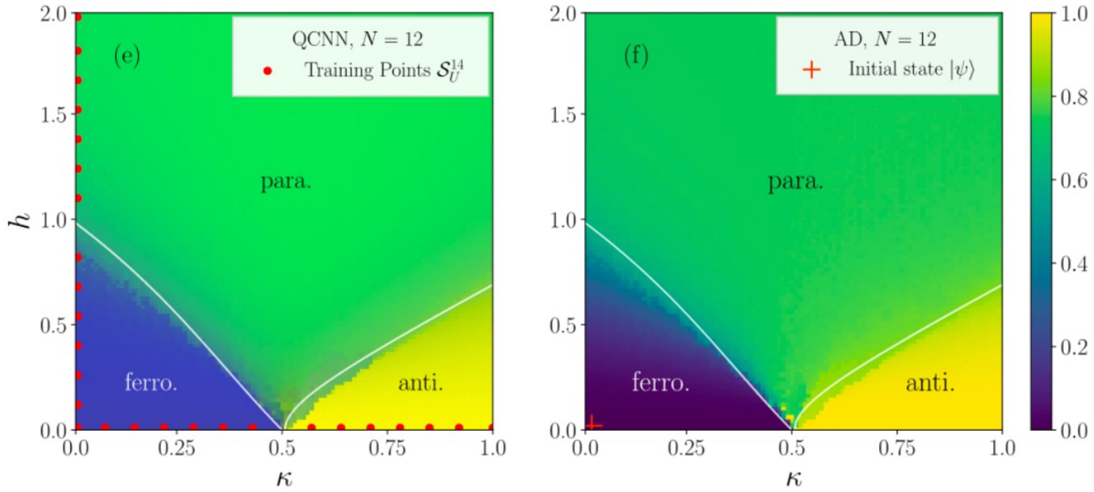
Results

Variational quantum data



QCNN (95%)

Autoencoder[1]



Binary Cross-entropy

Loss: $$\mathcal{L} = -\frac{1}{|\mathcal{S}_X^n|} \sum_{(\kappa,h)\in\mathcal{S}_X^n} \sum_{j=1}^{K} y_j(\kappa,h)\log\left(p_j(\kappa,h)\right)$$

Labels:

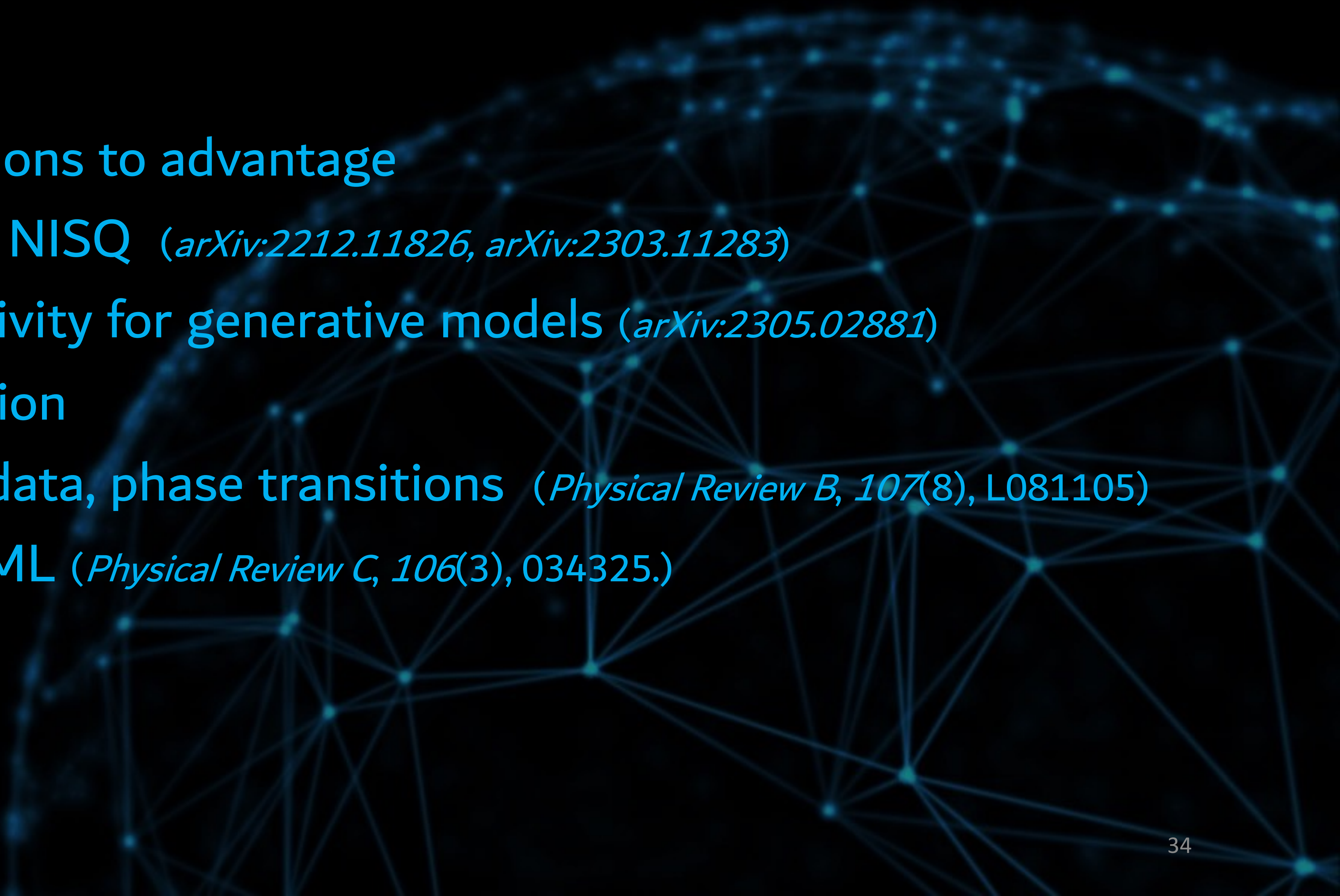
- [0,1] ferromagnetic
- [1,0] antiphase
- [1,1] paramagnetic
- [0,0] trash label

1. **Out of Distribution** Generalization[2]?
2. Performance increases with the system's size N=6 → N=12).
3. QCNN gives quantitative predictions

[1]Kottman, *et al., Phys. Rev. Research* **3**, 043184 (2021)
[2]M..Caro et al., arxiv:2204.10268, Banchi et all., PRX QUANTUM 2, 040321 (2021)

Monaco, at al. arXiv: 2208.08748 (2022), accepted PRB

Improving Robustness of QML applications



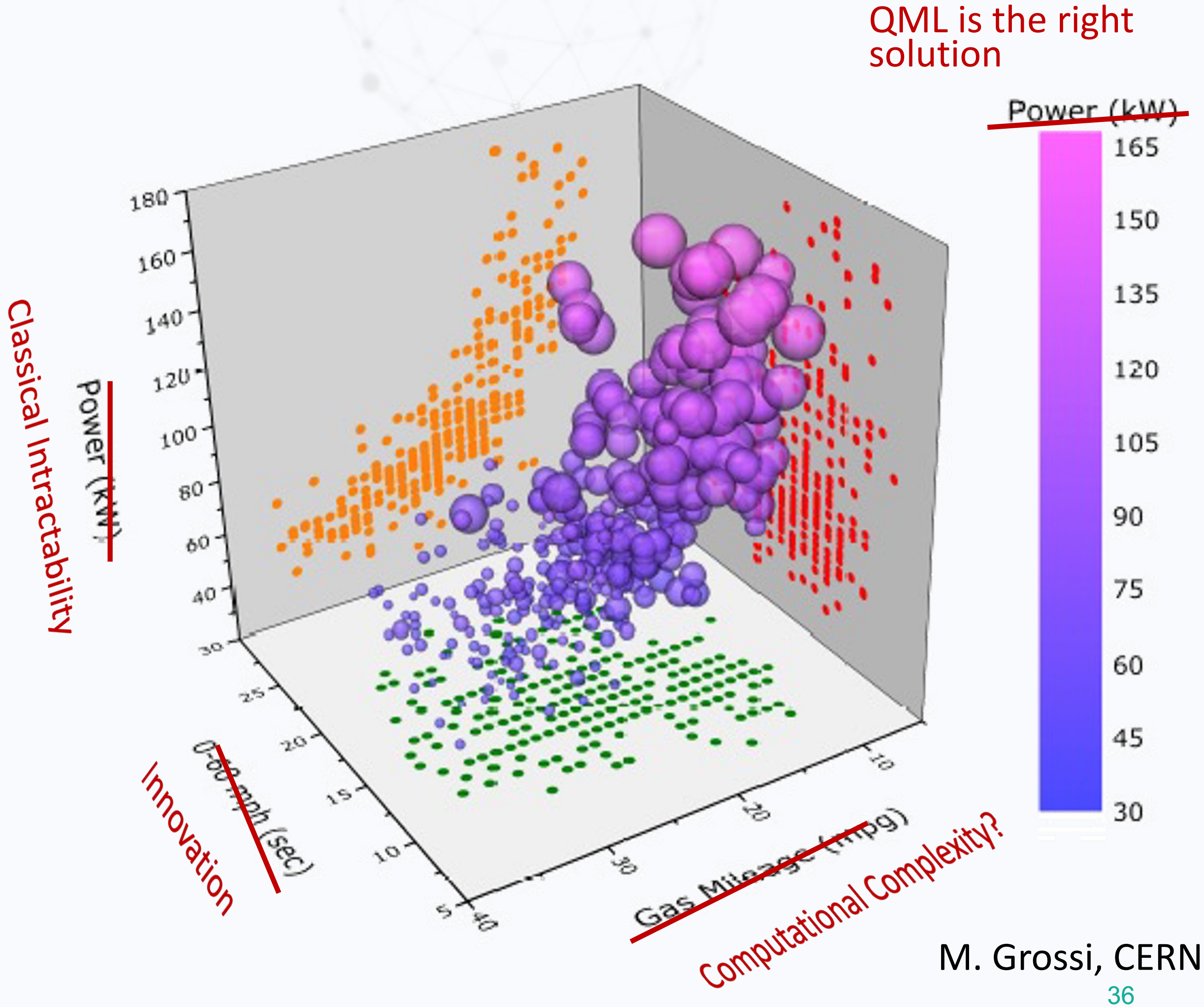
- Understanding conditions to advantage
- Stabilizing training on NISQ (*arXiv:2212.11826, arXiv:2303.11283*)
- Trainability vs expressivity for generative models (*arXiv:2305.02881*)
- Evaluating generalisation
- Quantum vs classical data, phase transitions (*Physical Review B, 107*(8), L081105)
- Algorithms beyond QML (*Physical Review C, 106*(3), 034325.)

Outlook and open questions

- Quantum technolgies could be revolutionary in terms of computing
- HEP provides challenges to Quantum Machine Learning
  - **What are the most promising applications?**
  - How do we define performance and validate results on **realistic use cases**?

- Experimental data has high dimensionality
  - Can we train Quantum Machine Learning algorithms effectively?
  - Can we reduce the impact of **data reduction** techniques?
- Experimental data is shaped by physics laws
  - Can we leverage them to build better algorithms?

QML Exclusion Region **in HEP?**



M. Grossi, CERN

Lectures and Hands-On at CERN

- «A practical Introduction to quantum computing», Elias Combarro

https://indico.cern.ch/event/970903/

- «Introduction to quantum computing », Heather Grey

https://indico.cern.ch/event/870515/

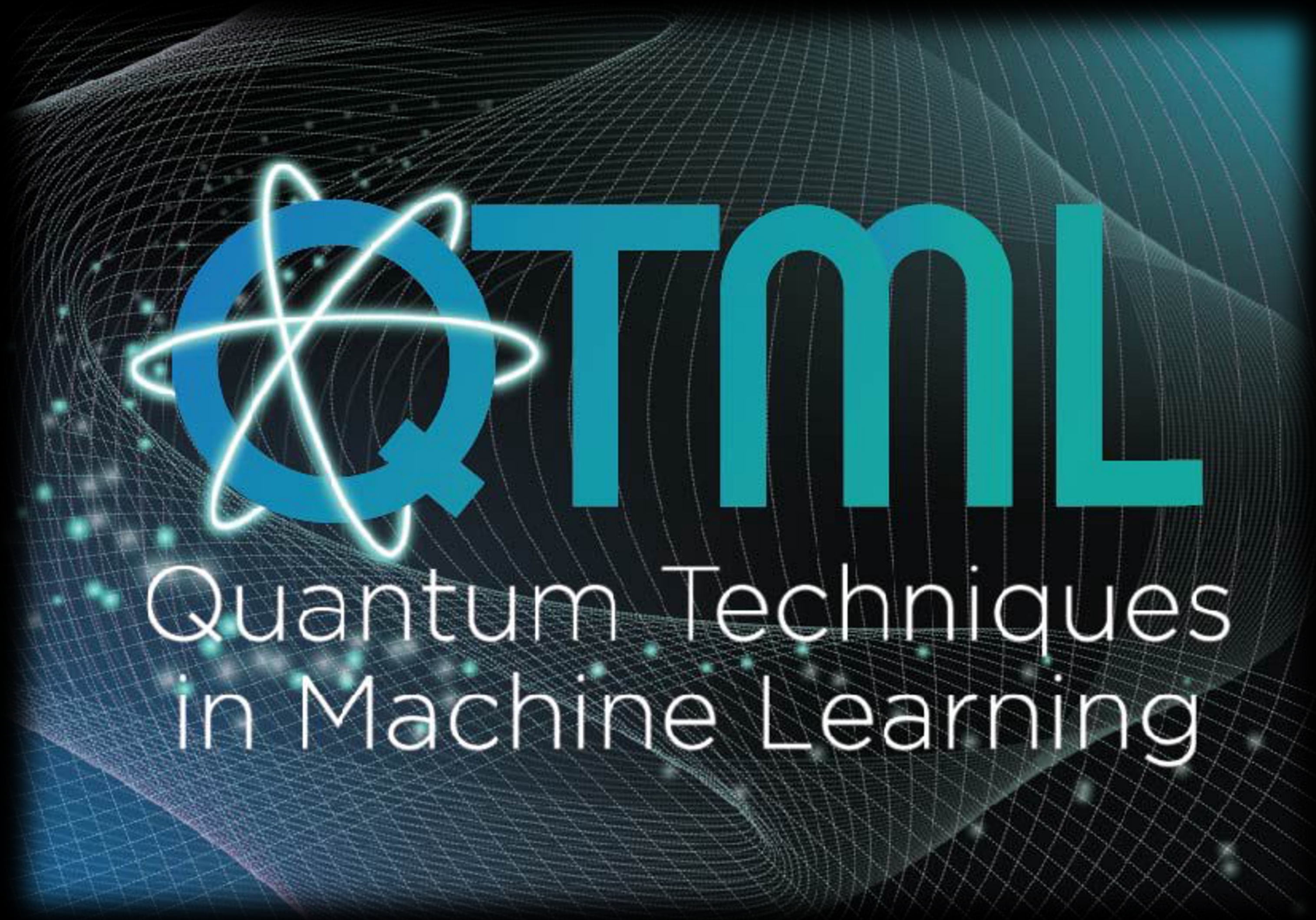
- A set of two hands-on (introduction) sessions for summer students (2023 openlab summer student lectures)

https://indico.cern.ch/event/1293871/

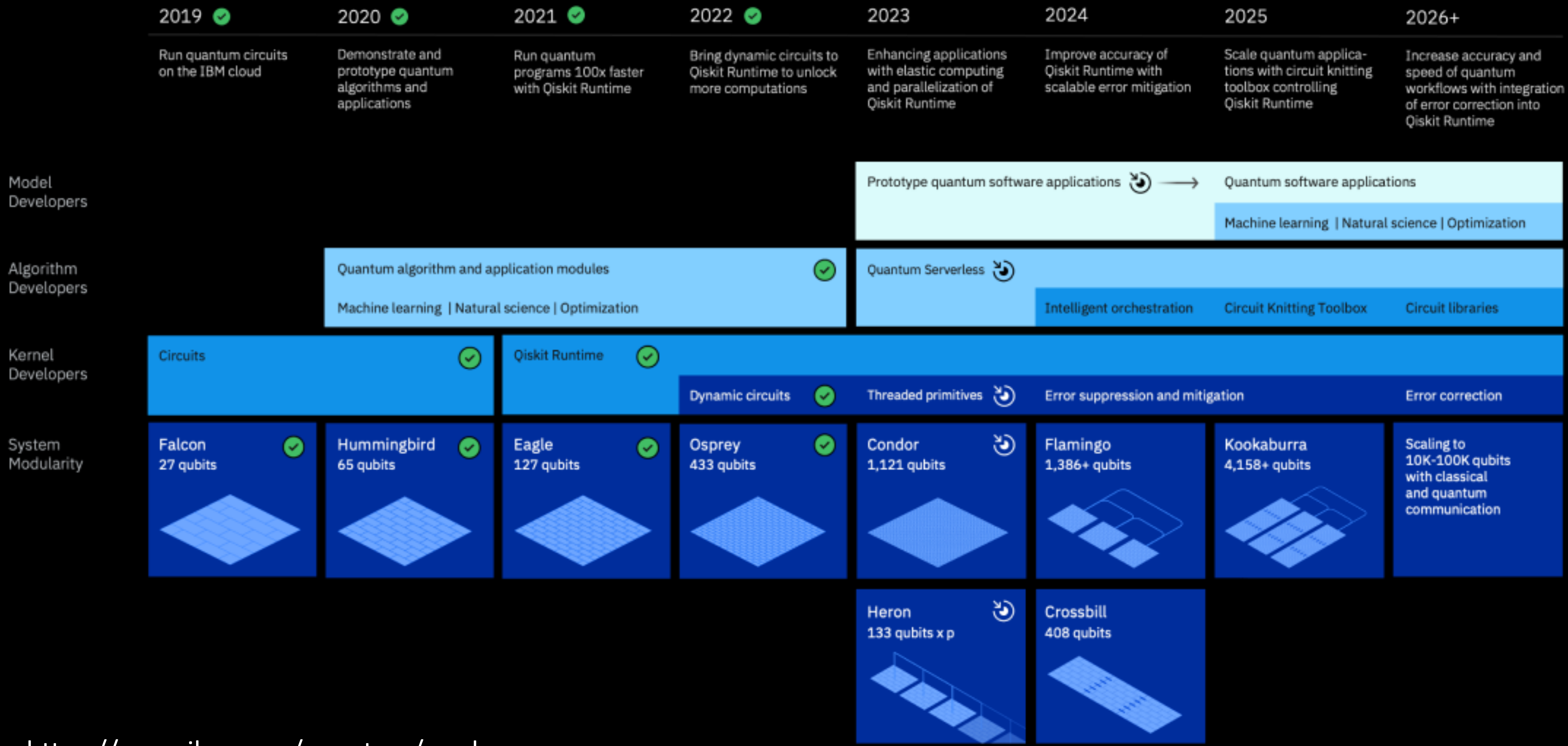
https://indico.cern.ch/event/1293874/

Thank you!

November 20th-24th, 2023
@CERN



Sofia.Vallecorsa@cern.ch



	2019 ✓	2020 ✓	2021 ✓	2022 ✓	2023	2024	2025	2026+
	Run quantum circuits on the IBM cloud	Demonstrate and prototype quantum algorithms and applications	Run quantum programs 100x faster with Qiskit Runtime	Bring dynamic circuits to Qiskit Runtime to unlock more computations	Enhancing applications with elastic computing and parallelization of Qiskit Runtime	Improve accuracy of Qiskit Runtime with scalable error mitigation	Scale quantum applications with circuit knitting toolbox controlling Qiskit Runtime	Increase accuracy and speed of quantum workflows with integration of error correction into Qiskit Runtime
Model Developers					Prototype quantum software applications →		Quantum software applications Machine learning | Natural science | Optimization	
Algorithm Developers		Quantum algorithm and application modules ✓ Machine learning | Natural science | Optimization			Quantum Serverless	Intelligent orchestration	Circuit Knitting Toolbox	Circuit libraries
Kernel Developers	Circuits ✓		Qiskit Runtime ✓	Dynamic circuits ✓	Threaded primitives	Error suppression and mitigation		Error correction
System Modularity	Falcon 27 qubits ✓	Hummingbird 65 qubits ✓	Eagle 127 qubits ✓	Osprey 433 qubits ✓	Condor 1,121 qubits	Flamingo 1,386+ qubits	Kookaburra 4,158+ qubits	Scaling to 10K-100K qubits with classical and quantum communication
					Heron 133 qubits x p	Crossbill 408 qubits		

https://www.ibm.com/quantum/roadmap

Kernel trainability and kernel concentration

Kernel values can **concentrate exponentially** around a common value

Need **exponentially larger number of measurements** to resolve

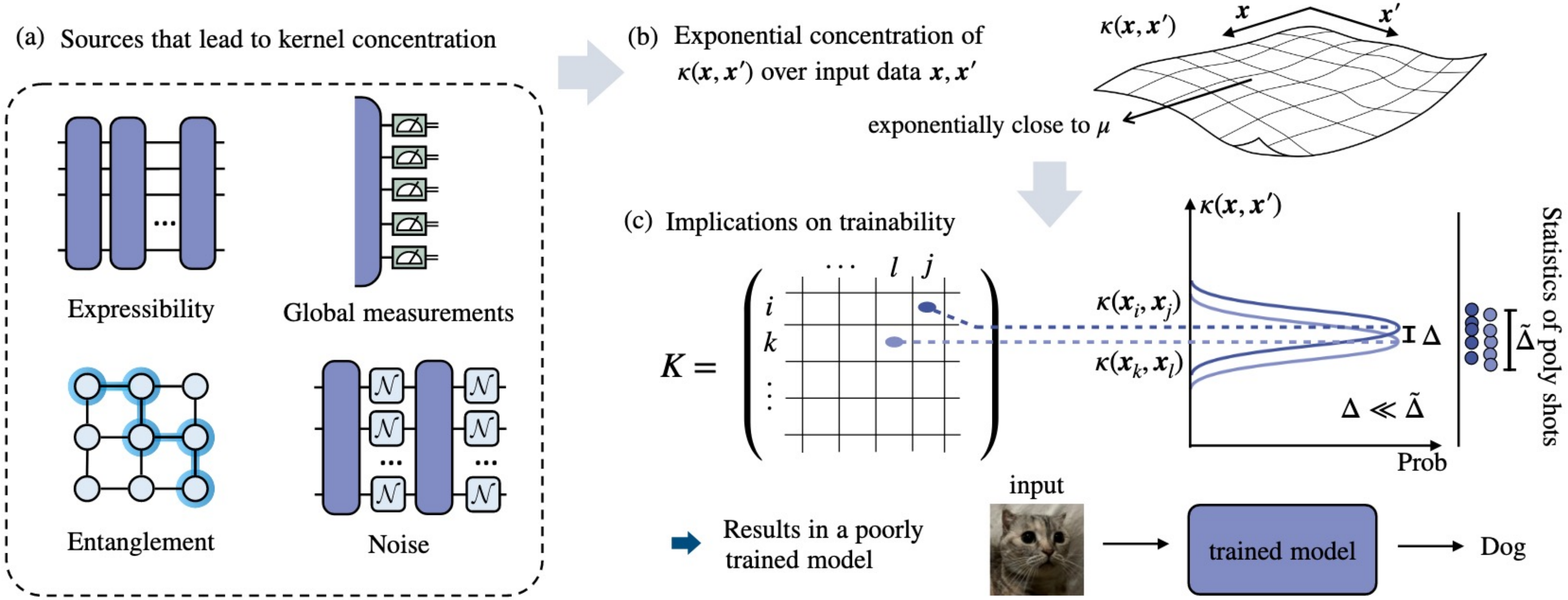
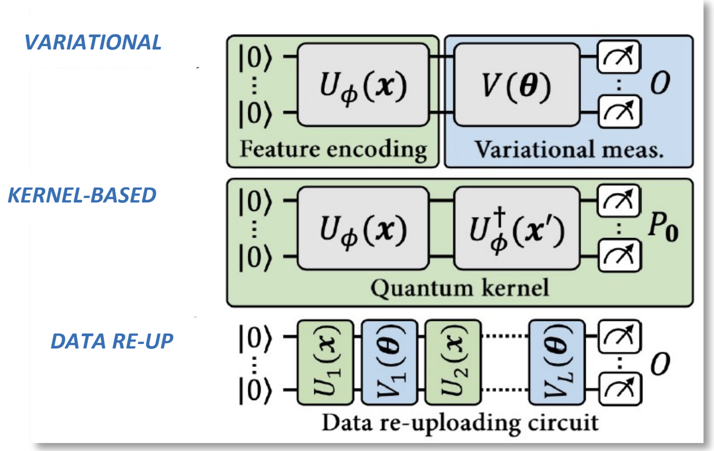


Figure 1. **Kernel concentration and its implications on trainability:** The exponential concentration (in the number of qubits n) of quantum kernels $\kappa(\boldsymbol{x}, \boldsymbol{x}')$, over all possible input data pairs $\boldsymbol{x}, \boldsymbol{x}'$, can be seen to stem from the difficulty of information extraction from data quantum states due to various sources (illustrated in panels (a) and (b)). The kernel concentration has a detrimental impact on the trainability of quantum kernel-based methods. As shown in panel (c), for a polynomial (in n) number of measurement shots, the sampling noise $\tilde{\Delta}$ dominates for large n and, as $\Delta \ll \tilde{\Delta}$, $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$ cannot be resolved from some other $\kappa(\boldsymbol{x}_k, \boldsymbol{x}_l)$, leading to a poorly trained model.

Study kernel trainability in our Anomaly Detection model (arxiv:2208.11060)

Equivalent interpretations?

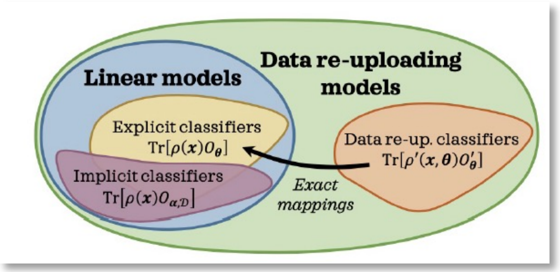


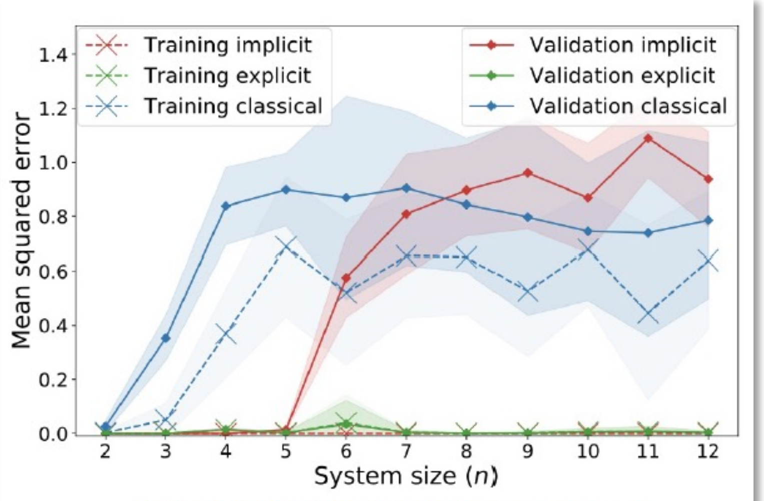
Characterize models behaviour, similarities among them and **link to data properties.**

Ex:

- Data Re-Uploading circuits: alternating data encoding and variational layers.
  - Represented as **explicit linear models** (variational) in larger feature space

  → can be reformulated as **implicit models** (kernel)
- Representer theorem: implicit models achieve **better accuracy**
  - Explicit models exhibit **better generalization** performance





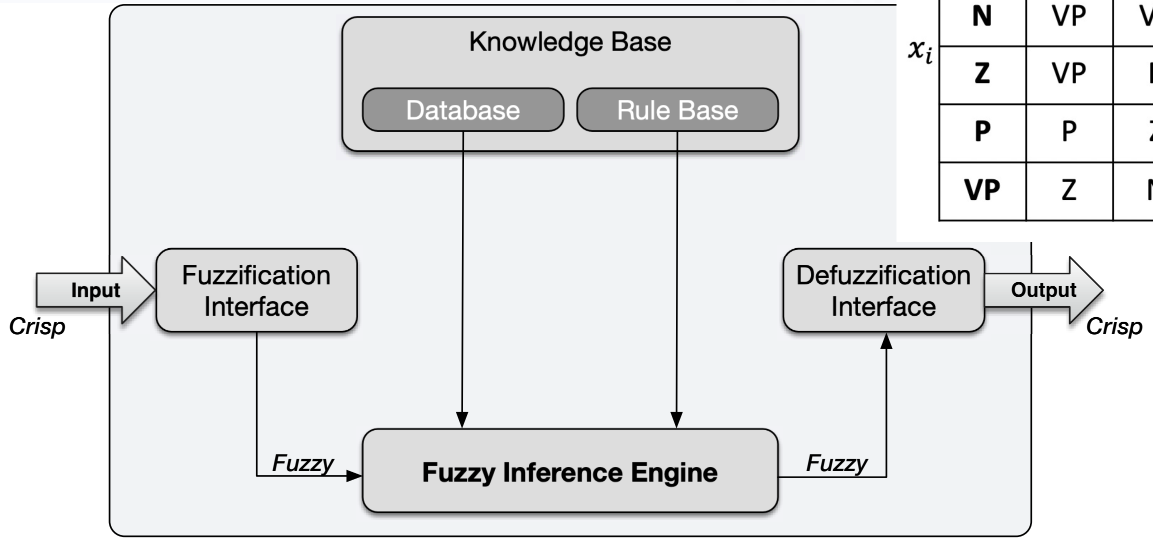
PCA on 28x28 fashion-MNIST dataset, ZZ feature encoding + hardware-efficient variational unitary

Jerbi, Sofiene, et al. "**Quantum machine learning beyond kernel methods.**" *arXiv preprint arXiv:2110.13162* (2021).

1-slide excursion: quantum fuzzy logic controller

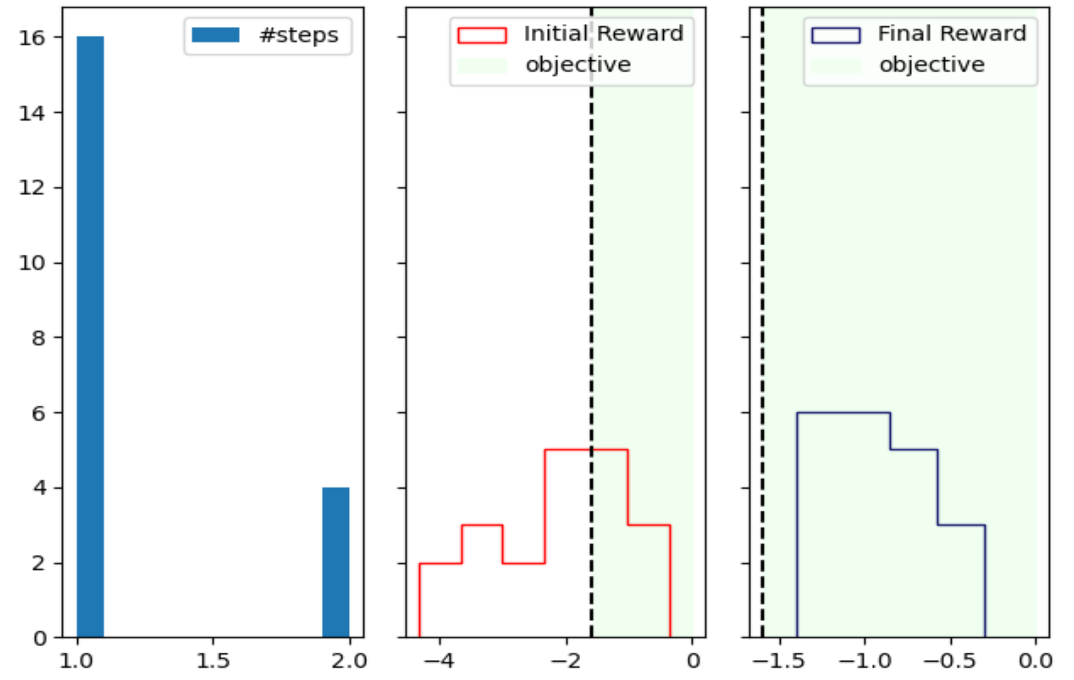
- **Alternative control algorithm** to RL
- **Fuzzy Logic** is used to develop control systems **based on linguistic rules** ⇨ **highly interpretable**
- **Quantum Fuzzy Control System** *(G. Acampora, R. Schiattarella, A. Vitiello)*
  Exploit **exponential advantage** in computing fuzzy rules on quantum computers
- **Successfully evaluated on AWAKE beam line, no training required**

Evaluation: on AWAKE beam line
Objective reached typically in 1 step

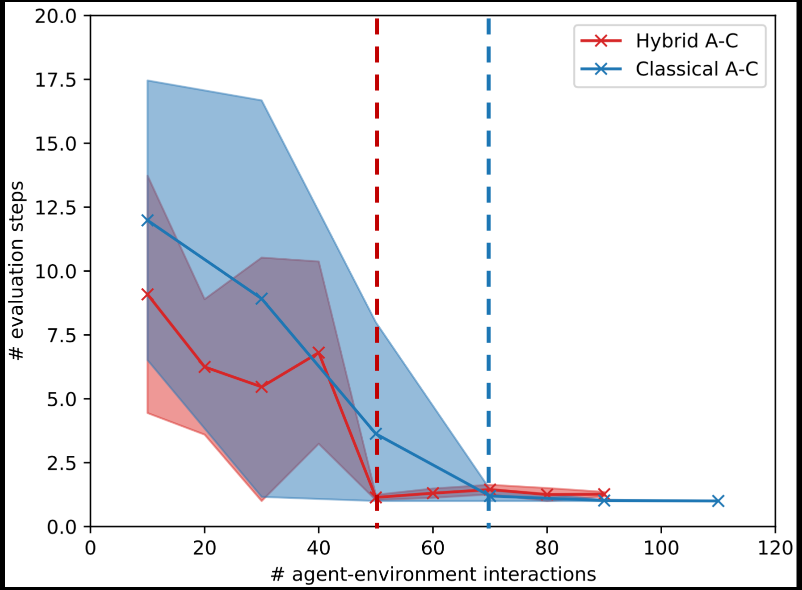


dk_i

x_i	VN	N	Z	P	VP
VN	VP	VP	VP	P	Z
N	VP	VP	P	Z	N
Z	VP	P	Z	N	VN
P	P	Z	N	VN	VN
VP	Z	N	VN	VN	VN



2nd study: 10D continuous beam steering



- **Hybrid actor-critic (A-C) works**
- **Minor improvement in terms of sample efficiency**
  50 vs 70 interactions
- **Very few interactions sufficient** for both approaches
- **Dynamics potentially too simple** (linear)

⇨ Move towards **more complex RL benchmarks**