

Quantum Machine Learning in High Energy Physics



QUANTUM
TECHNOLOGY
INITIATIVE

Sofia Vallecorsa

AI & Quantum Research - CERN IT

March 27th 2023

Outline

1. **Introduction**
2. **CERN Quantum Technology Initiative**
3. **Quantum Machine Learning and Applications at CERN**
4. **Anomaly Detection**
5. **Beam Optimisation in linear accelerators**
6. **Improving robustness**
 - Stabilizing training on NISQ
 - Quantum data
7. **Summary**



Not just hype, but ...

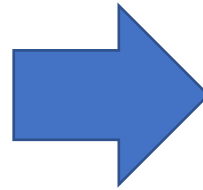
- **Quantum computing evolves rapidly with new improvements and new applications published almost every day**
 - Together with theoretical understanding of their behavior
- **Quantum algorithms are not yet ready to replace classical computing on realistic, large scale problems**
- **(NISQ) hardware limitations have an effect on :**
 - The **size of the problems** we can solve
 - The **complexity of the algorithms** we can implement
 - The **stability** of the results
- **Impact time scale and size is driven by hardware roadmaps**
 - Need **research on algorithms** to evolve in parallel, to accelerate reach of fault-tolerant regime (co-development)



Quantum algorithms and applications

Quantum effects improve and accelerate complex algorithms

- Sampling, searches and optimization
- Linear algebra and machine learning
- Cryptography and communication



Many potential applications in HEP:

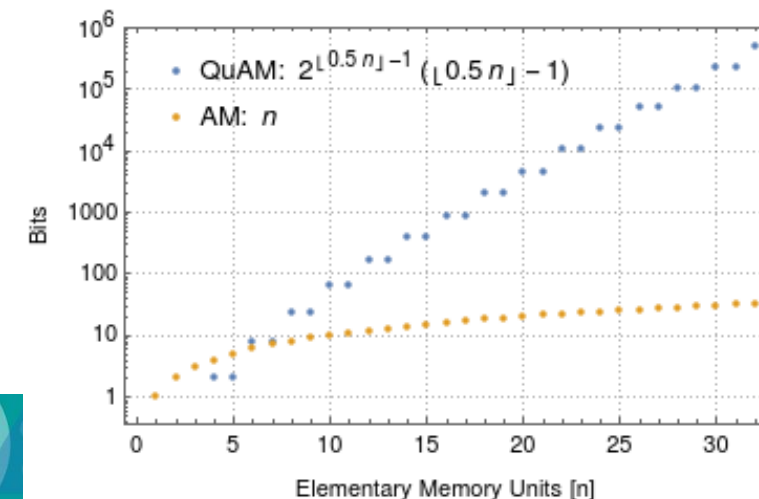
- Monte Carlo and Event Generation
- Quantum Simulation
- Pattern Recognition
- QML

Challenges:

Re-think algorithms design

Fairly design classical benchmarks

Ex.: Exponential data compression with a Quantum Associative memory



Shapoval, Illya, and Paolo Calafiura.
"Quantum associative memory in
HEP track pattern recognition." *EPJ
Web of Conferences*. Vol. 214. EDP
Sciences, 2019



Quantum Computing at CERN



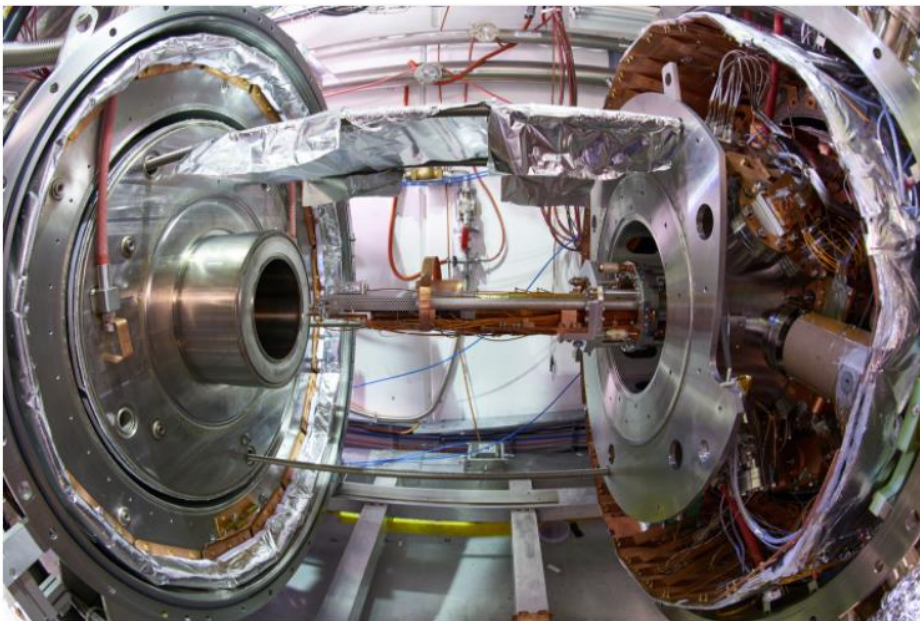
The CERN Quantum Technology Initiative

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)

CERN established the QTI in 2020

- Roadmap in 2021
- Publicly available on Zenodo

<https://doi.org/10.5281/zenodo.5553774>

T1 - Scientific and Technical Development and Capacity Building

T2 - Co-development

APPLICATIONS | NEWS

CERN unveils roadmap for quantum technology

4 November 2021



Credit: CERN

T3 - Community Building

T4 - Integration with national and international initiatives and programmes

Scientific Objectives



- Assess the **areas of potential quantum advantage** in HEP (QML, classification, anomaly detection, tracking)
- Develop **common libraries of algorithms, methods, tools**; benchmark as technology evolves
- Collaborate to the development of shared, **hybrid classic-quantum infrastructures**

Computing & Algorithms



- Identify and develop techniques for **quantum simulation** in collider physics, QCD, cosmology within and beyond the SM
- Co-develop quantum computing and sensing approaches by providing **theoretical foundations** to the identifications of the areas of interest

Simulation & Theory



- Develop and promote **expertise in quantum sensing** in low- and high-energy physics applications
- Develop quantum sensing approaches with emphasis on **low-energy particle physics measurements**
- Assess **novel technologies and materials** for HEP applications

Sensing, Metrology & Materials



- **Co-develop CERN technologies relevant to quantum infrastructures** (time synch, frequency distribution, lasers)
- Contribute to the **deployment and validation of quantum infrastructures**
- Assess requirements and **impact of quantum communication on computing applications** (security, privacy)

Communications & Networks

Quantum Computing Objectives at CERN



- Assess the **areas of potential quantum advantage** in HEP (QML, classification, anomaly detection, tracking)
- Develop **common libraries of algorithms, methods, tools**; benchmark as technology evolves
- Collaborate to the development of shared, **hybrid classic-quantum infrastructures**

Computing & Algorithms

Set baseline for **prioritisation** and **systematisation**

- **Quantum Machine Learning**
 - Relatively loose definition
 - Variational approach / Robustness to noise
- **Algorithms beyond QML**

Formal approach to algorithms, methods, error characterisation and correction

Test different hardware

- Semi-conductors, ions, ... (IBM, Rigetti, IonQ,...)
- Photonic (Xanadu), Annealer (D-Wave)
- Quantum-inspired (Fujitsu digital, Toshiba SBM)



Quantum Machine Learning



Studying ~~Deep~~ Learning in physics

Quantum Machine

- **High quality labelled training data** from realistic MC simulation
- Large **experimental datasets**
- Interestingly **structured data** at multiple scales
- Detailed understanding of **systematic uncertainties**

M. Erdmann, J. Glombitza, G. Kasieczka, U. Klemradt, Deep Learning for physics research



High Energy Physics use cases

Multiple applications:

- Simulation
- Anomaly Detection and trigger
- Binary Classification and data analysis
- Reconstruction: Tracking, Calorimetry and Jets
- Engineering: Reinforcement Learning for beams steering in the accelerator sector

Different use cases can have different requirements:

Fast inference vs Real time training capability vs Fast training for large optimizations

Classical Deep Learning review: <https://iml-wg.github.io/HEPML-LivingReview/>

Quantum Advantage for QML

QML: Quantum computing to "improve" ML

Different advantage definitions

Runtime speedup

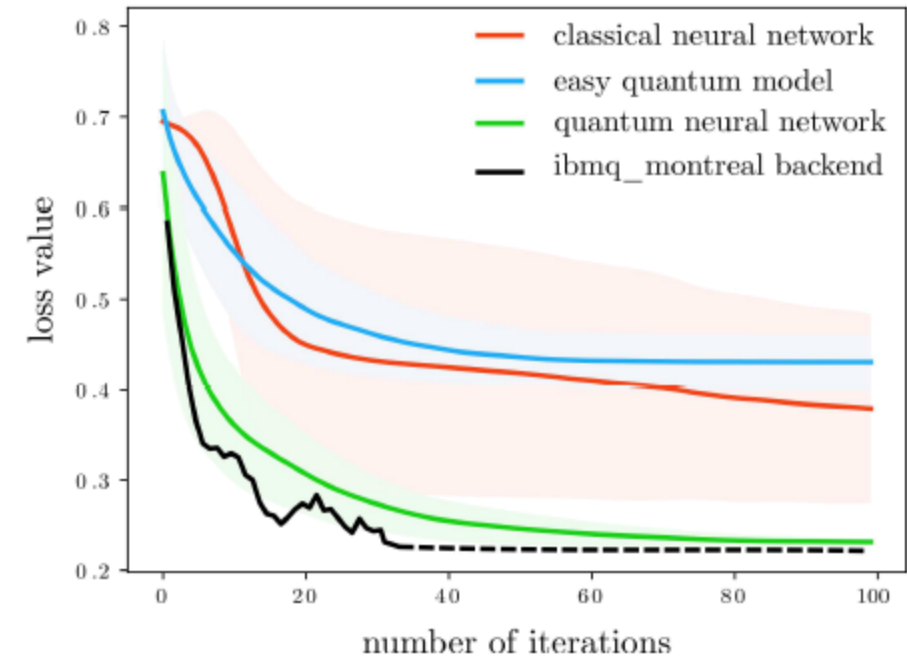
Sample complexity

Representational power

Classical Intractability: a quantum algorithm that cannot be efficiently simulated classically

- No established recipe for classical data
- Need to use the **whole exponential advantage** in Hilbert space, but **will it converge ?**

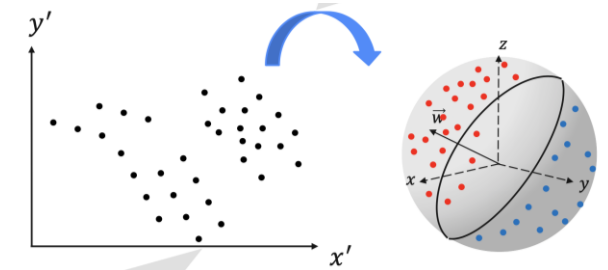
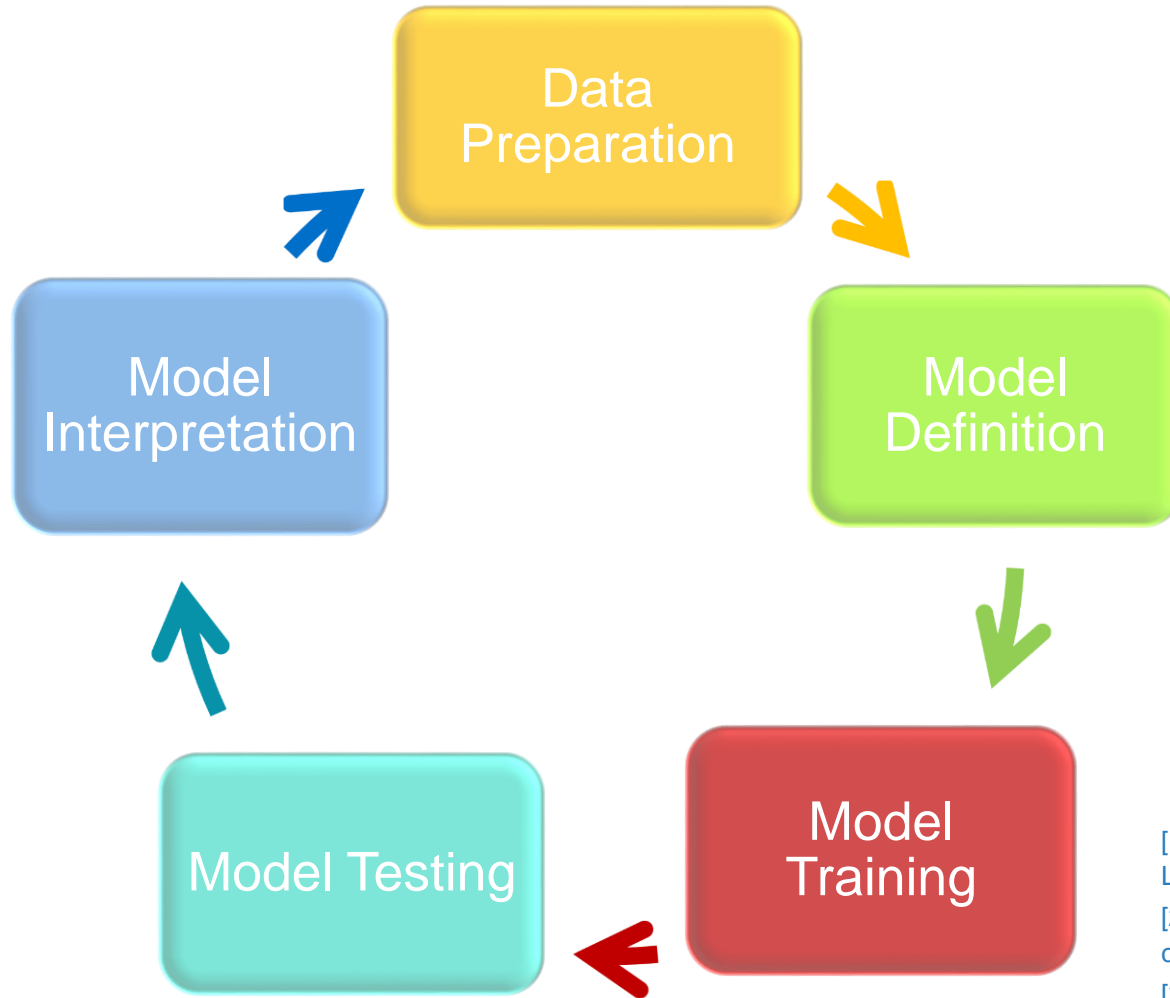
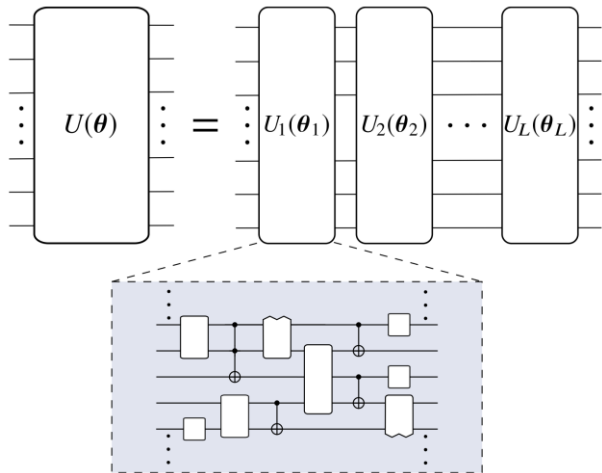
(Algorithm expressivity vs convergence and generalization)



QML Lifecycle

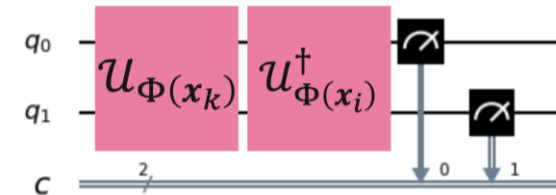
The advantage of many known QML algorithms is impeded by I/O bottleneck

Trainability (BP...)



Data Reduction
Data Encoding [1,2,3]

Read Out



- [1] Robust data encodings for quantum classifiers, Ryan LaRose and Brian Coyle, Phys. Rev. A 102, 032420
- [2] Quantum convolutional neural network for classical data classification, <https://arxiv.org/pdf/2108.00661.pdf>
- [3] Quantum Support Vector Machines for Continuum Suppression in B Meson Decays, <https://arxiv.org/abs/2103.12257>

Model definition

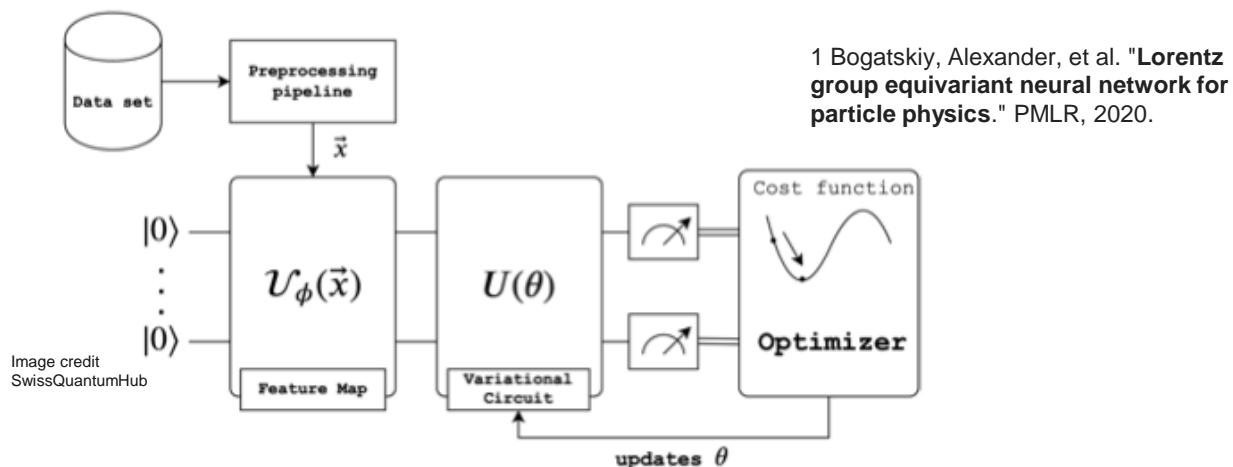
Variational algorithms

Parametric ansatz

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

Data Embedding can be learned

Ansatz design can leverage data symmetries¹



Kernel methods

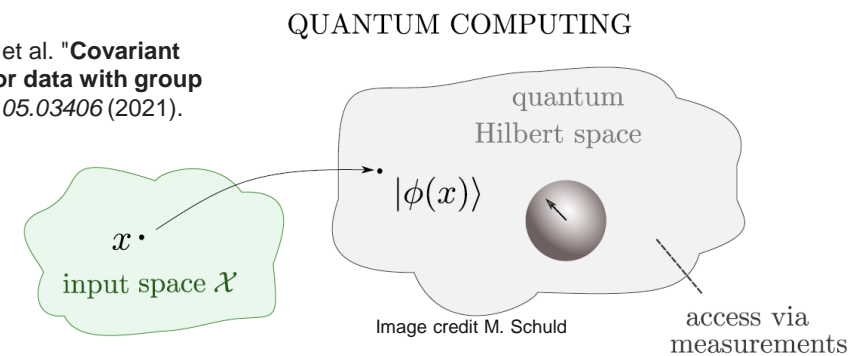
Feature maps as quantum kernels

Use classical **kernel-based training**

- **Convex** losses
- Compute pair-wise distances in N_{data}

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

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



Representer theorem: implicit models achieve **better accuracy**³

Explicit models exhibit **better generalization** performance

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

Model Convergence and Barren Plateau

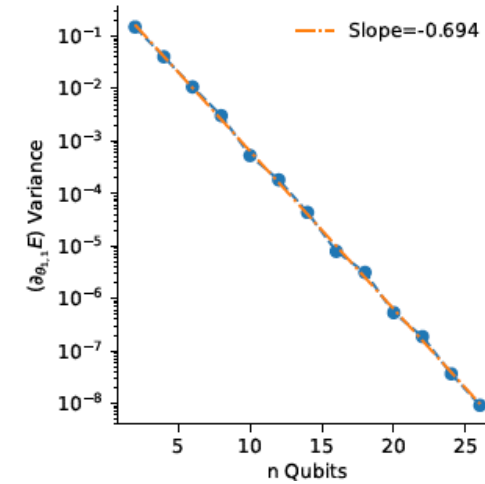
The size of the Hilbert space requires compromises between **expressivity**, **convergence** and **generalization**

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

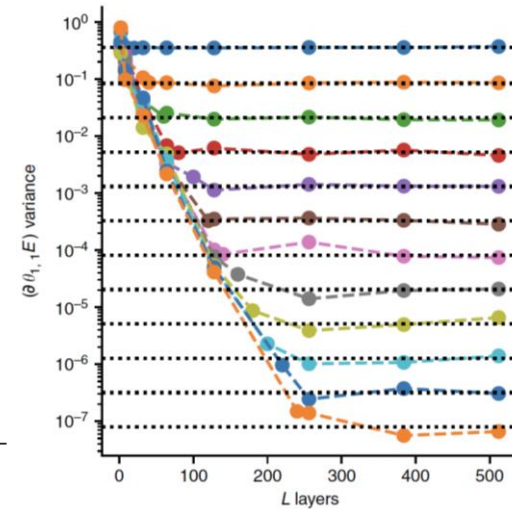
- Convergence still possible if gradients consistent between batches.

Quantum gradient decay exponentially in the number of qubits

- 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))

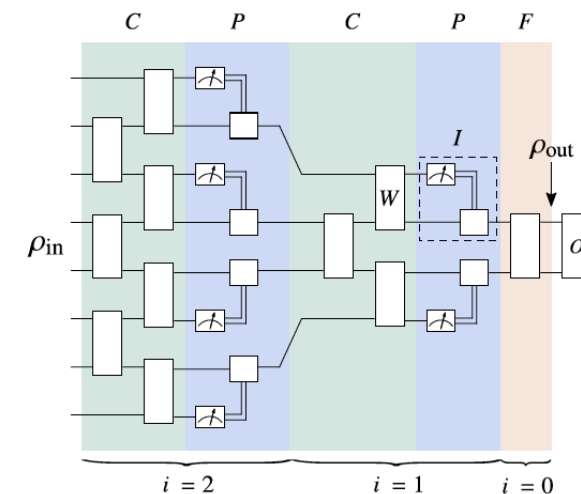
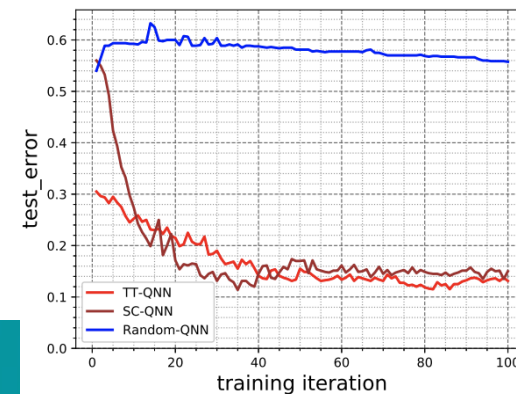


J. McClean *et al.*, arXiv:1803.11173



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

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



Kernel trainability and kernel concentration

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

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

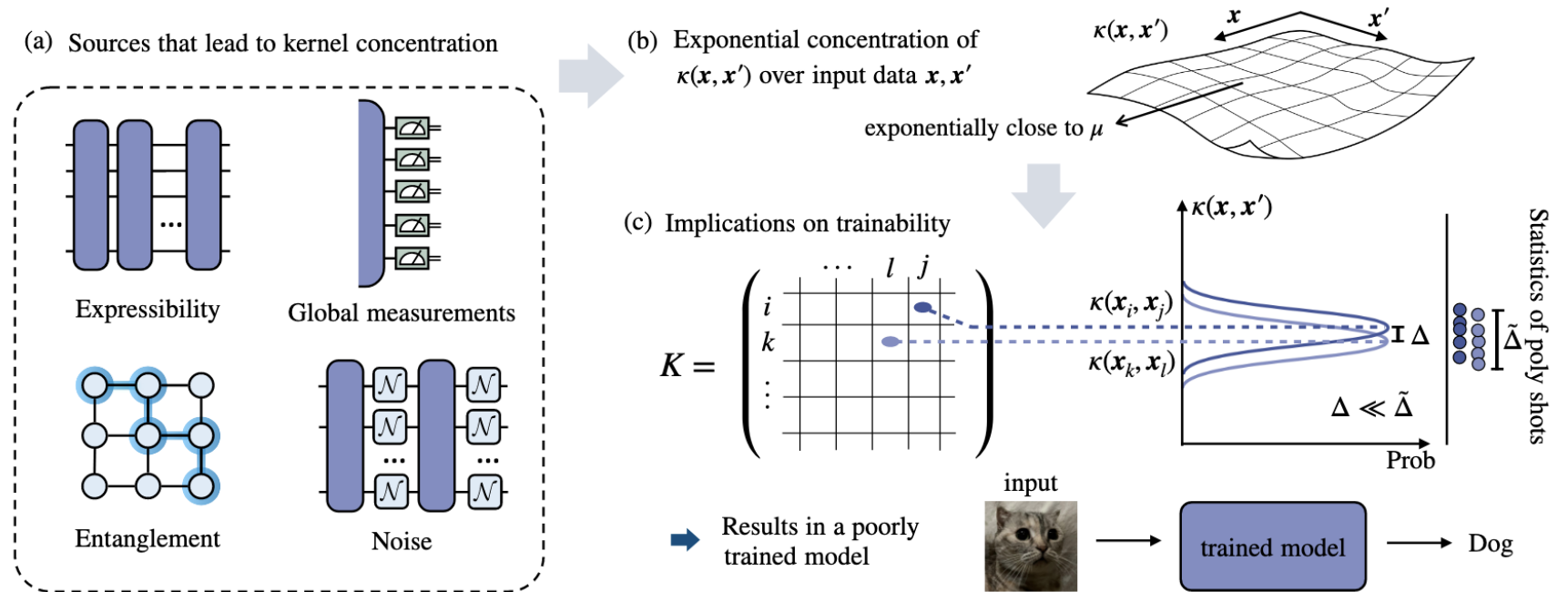


Figure 1. **Kernel concentration and its implications on trainability:** The exponential concentration (in the number of qubits n) of quantum kernels $\kappa(\mathbf{x}, \mathbf{x}')$, over all possible input data pairs \mathbf{x}, \mathbf{x}' , can be seen to stem from the difficulty of information 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 $\Delta \ll \tilde{\Delta}$, $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ cannot be resolved from some other $\kappa(\mathbf{x}_k, \mathbf{x}_l)$, leading to a poorly trained model.

Study kernel trainability in our AD model (arxiv:2208.11060)

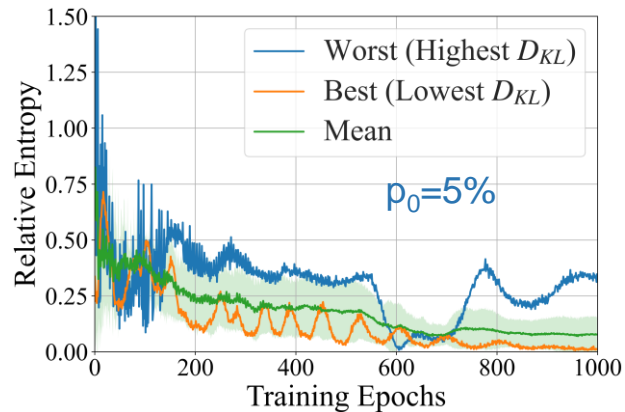
Our results so far..

- ⊕ **Multiple QML prototypes for different applications**
We can build expressive models and we can train them 😊
- ⊕ **Increasing level of precision**
- ⊕ **Robustness against noise ?**
- ⊖ **Scale is still a problem on current quantum hardware**
Complex data pre-processing
- ⊖ **Generalization**
- ⊖ **Empirical results → Need theoretical grounding**

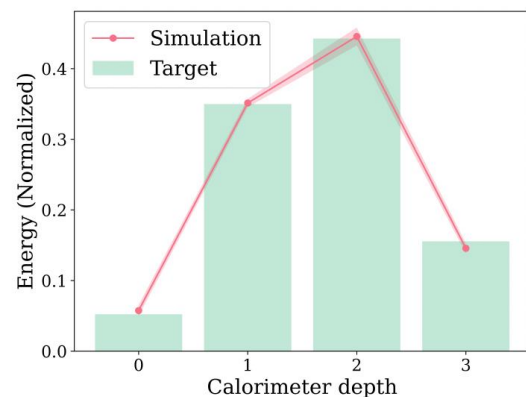


QML at CERN

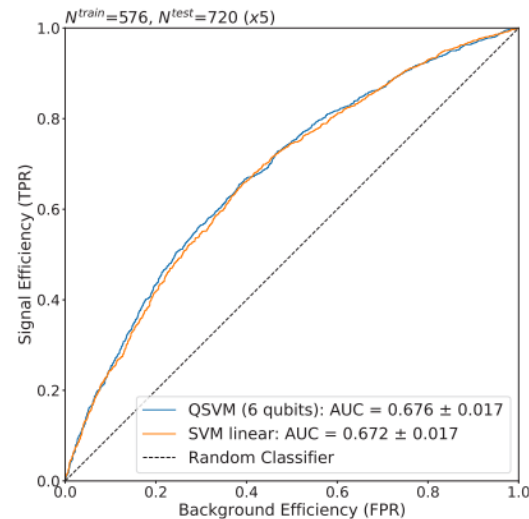
Borras, Kerstin, et al. "Impact of quantum noise on the training of quantum Generative Adversarial Networks." *arXiv preprint arXiv:2203.01007* (2022).



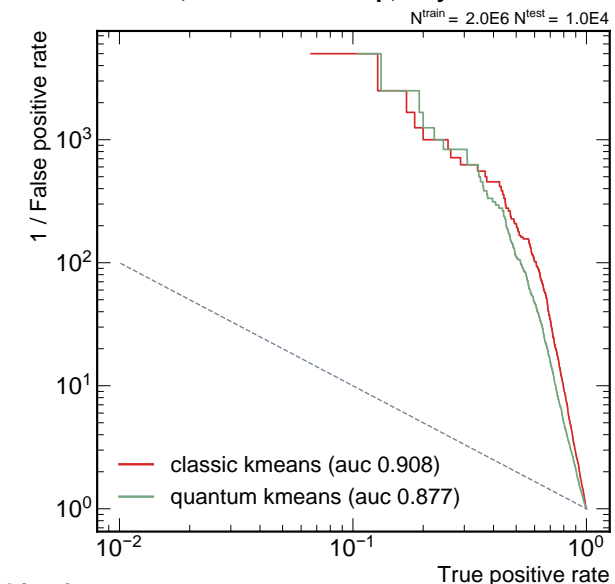
Chang S.Y. et al., **Running the Dual-PQC GAN on Noisy Simulators and Real Quantum Hardware**, QTML2021, ACAT21



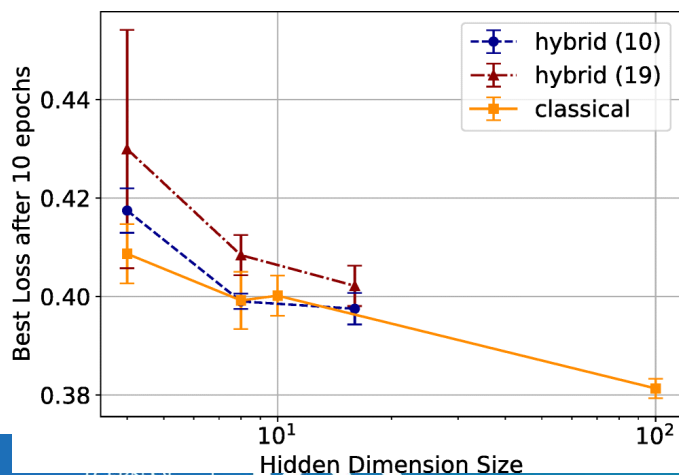
Vasilis Belis, Samuel González-Castillo, Christina Reissel, Sofia Vallecora, Elías F. Combarro, Günther Dissertori, and Florentin Reiter. **Higgs analysis with quantum classifiers**. EPJ Web of Conferences, 251:03070, 2021



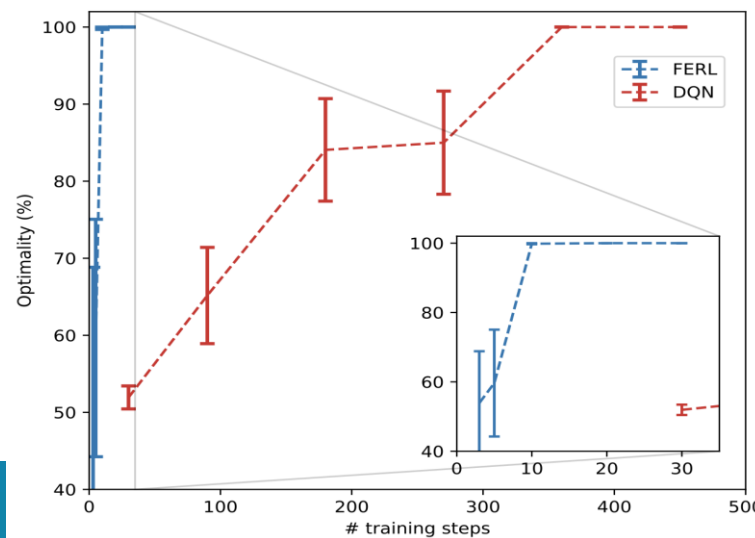
Kinga Wozniak, **Unsupervised clustering for a Randall–Sundrum Graviton at 3.5TeV narrow resonance**, 5th IML workshop, May 2022



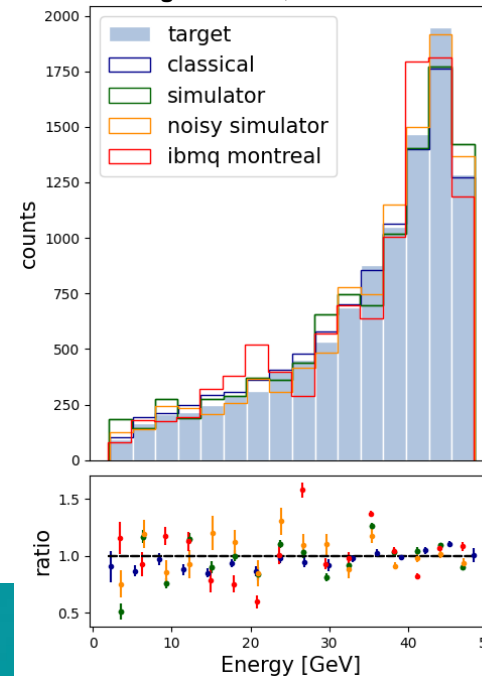
Tüysüz, Cenk, et al. "Hybrid quantum classical graph neural networks for particle track reconstruction." *Quantum Machine Intelligence* 3.2 (2021): 1-20.



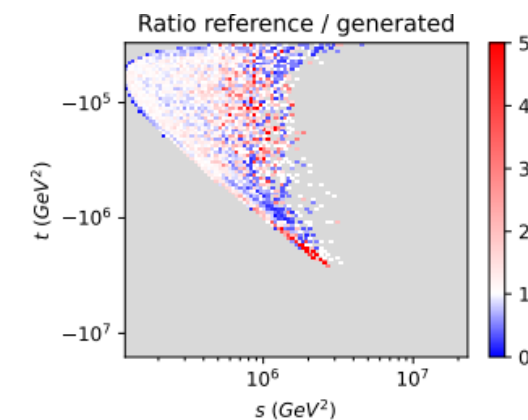
M. Shenk, V. Kain, **Quantum Reinforcement Learning**, BQIT 2021, 2022 CERN openlab Tech Workshop



O. Kiss, **Quantum Born Machine for event generation**, ACAT2021



Bravo-Prieto, Carlos, et al. "Style-based quantum generative adversarial networks for Monte Carlo events." *arXiv preprint arXiv:2110.06933* (2021).





Anomaly Detection



New Physics at the LHC

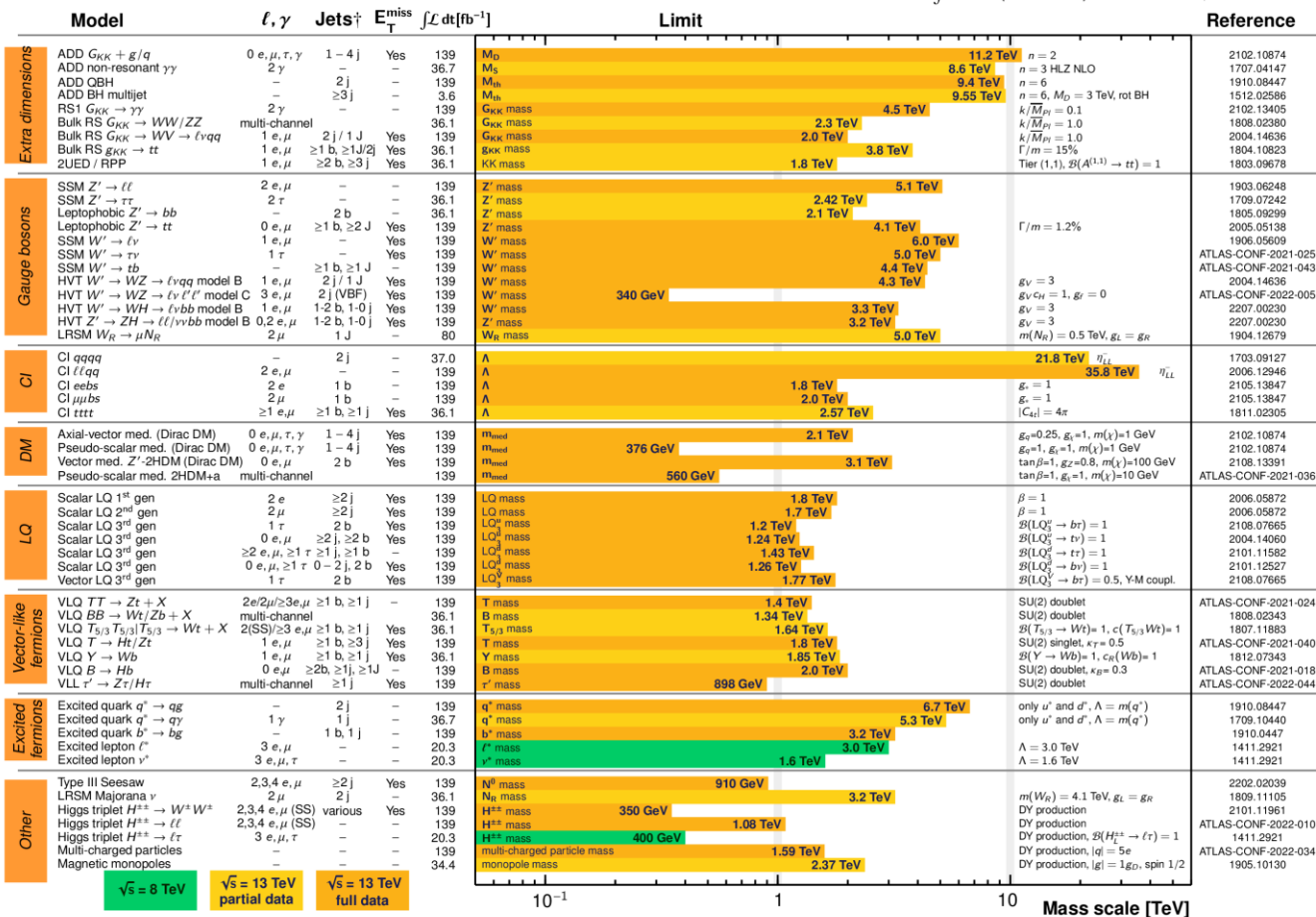
So far only negative results in direct (model dependent) searches

ATLAS Heavy Particle Searches* - 95% CL Upper Exclusion Limits

Status: July 2022

ATLAS Preliminary

$\int \mathcal{L} dt = (3.6 - 139) \text{ fb}^{-1}$ $\sqrt{s} = 8, 13 \text{ TeV}$



How to insure we do not miss potential discoveries?

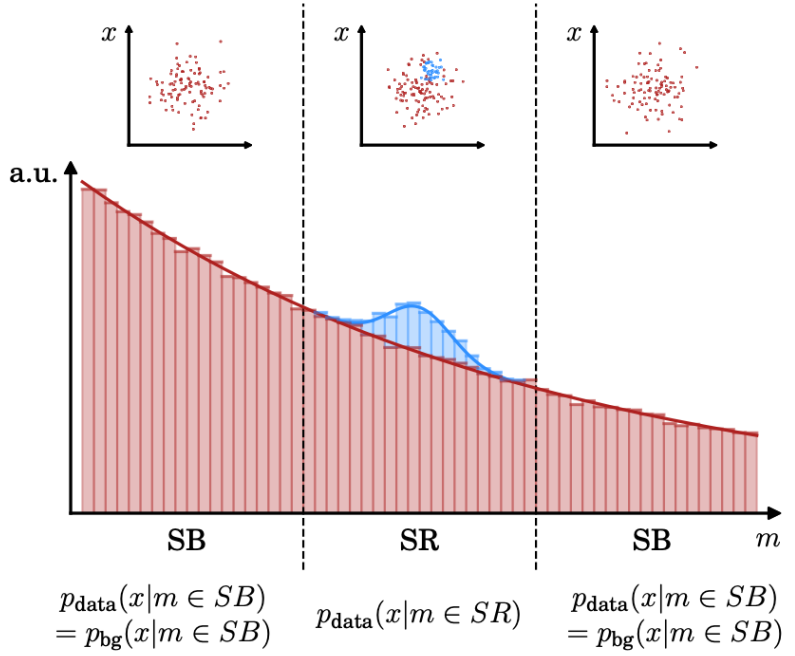
*Only a selection of the available mass limits on new states or phenomena is shown.

†Small-radius (large-radius) jets are denoted by the letter j (J).

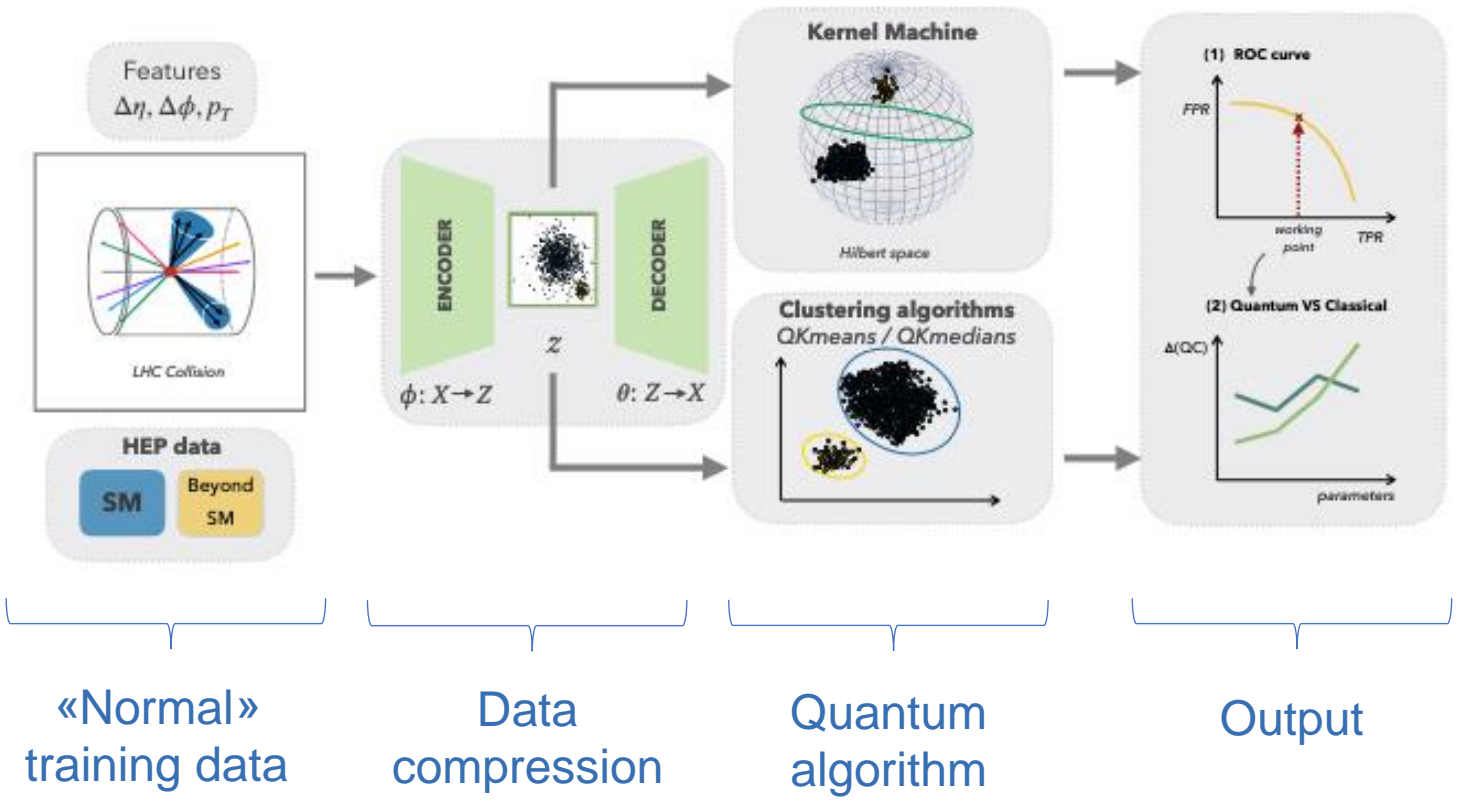
A typical hybrid QML workflow

Anomaly detection can point to new physics at the LHC

Model-agnostic!



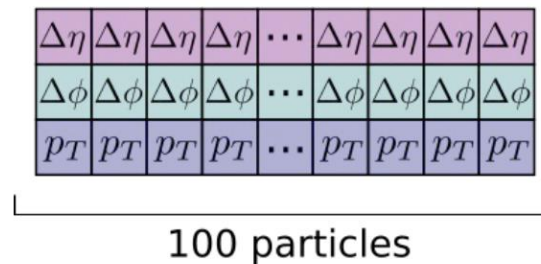
Use a hybrid quantum-classical workflow



Standard Model jet data

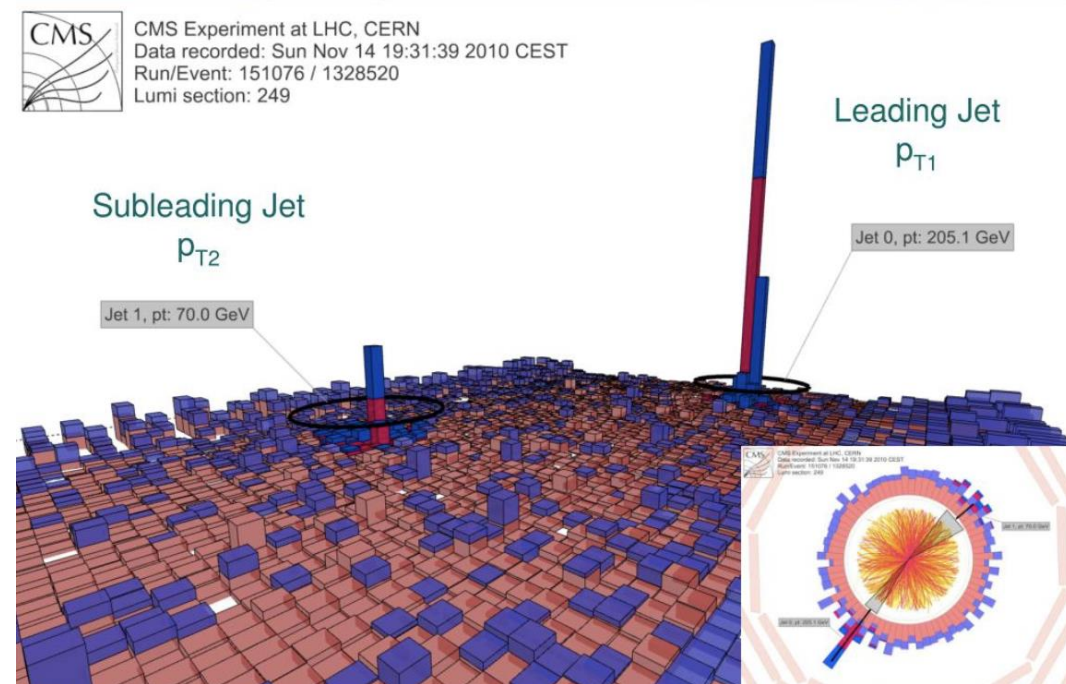
Simulate QCD multijet production at the LHC (64 fb^{-1})

Jet is built of **100 highest- p_T particles** within $\Delta R < 0.8$ from its axis.



Event selection:

- Two jets with $p_T > 200 \text{ GeV}$ and $|\eta| < 2.4$
- $m_{jj} > 1260 \text{ GeV}$ (emulate online selection)
- Each event is represented by its two highest- p_T jets.



Convolutional AutoEncoder compresses particle jet learning the **internal structure**

- Trained on background events

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

Unsupervised kernel machine

“Standard” kernel definition

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

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

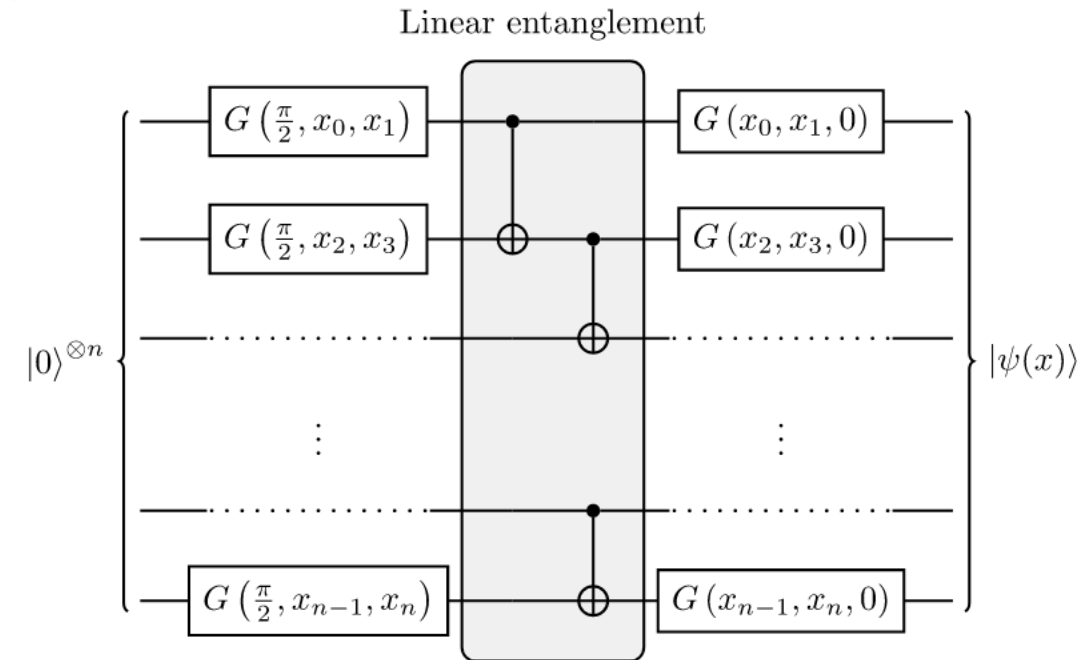
Train a kernel machine to find the hyperplane that **maximizes the distance of the data from the origin of the feature vector space**

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

subject to $w \cdot \Phi(x_i) \geq \rho - \xi_i, \xi_i \geq 0, \forall i.$

$\nu \in (0, 1)$ Is a **upper bound** on the fraction of anomalies in the training data set at 0.01 (at most 1% QCD training data are falsely flagged)

Data Embedding circuit



Results

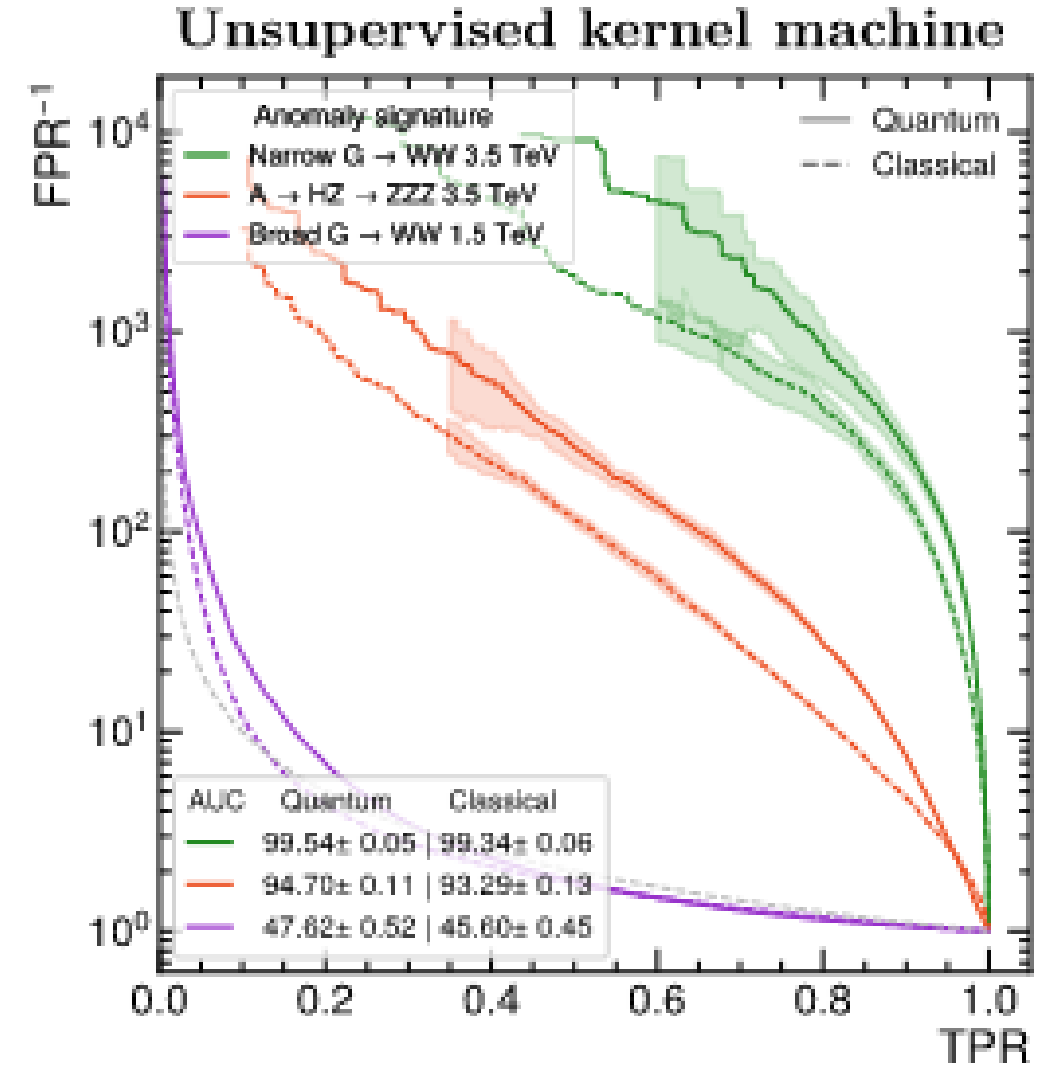
Comparison to best-performing classical algorithm with similar complexity trained and tested on the same data

- RBF –based SVM

AUC shows **marginal advantage** for quantum algorithm

Evaluate performance at **typical working**, where $\epsilon_s = 0.6, 0.8$

Quantum kernel machine works best for more complex physics



Characterizing the advantage

Higher is better

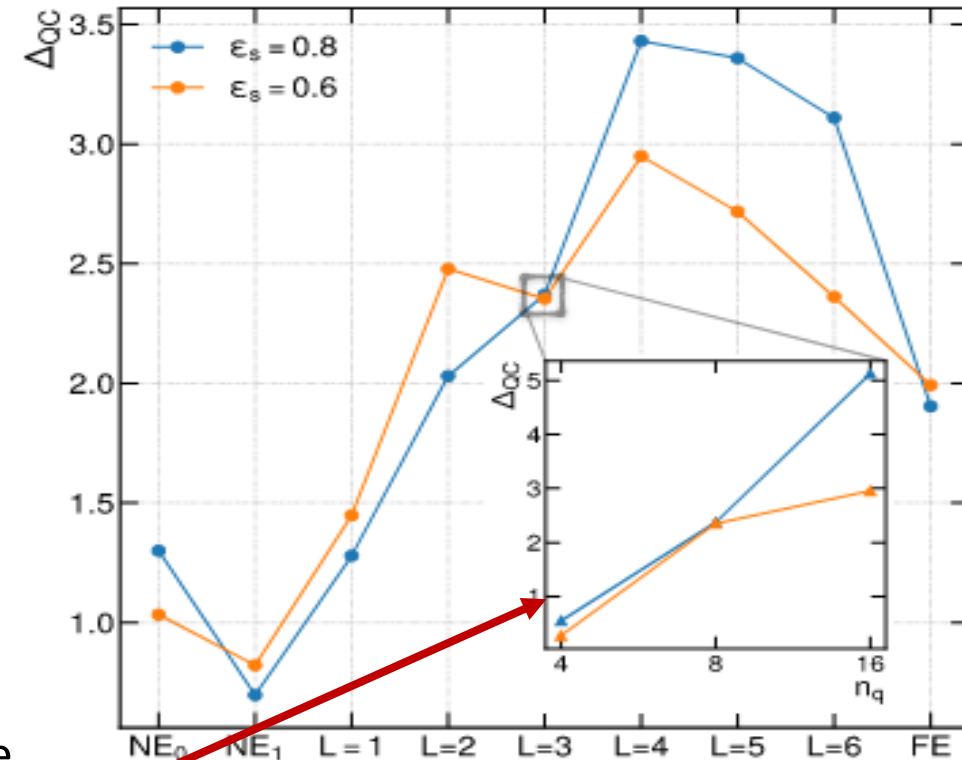
Given **signal and background efficiencies**, ϵ_s and ϵ_b respectively:

$$\Delta_{QC}(\epsilon_s) = \frac{\epsilon_b^{-1}(\epsilon_s; Q)}{\epsilon_b^{-1}(\epsilon_s; C)}$$

Performance advantage is consistent

- Increase in the expressibility and entanglement up to $L=4$ improve performance, reduce it above
- **Full entanglement is not better**

Classical is better than 4 qubit QSVM



Full entanglement

NE_0 : 1 layer - No entanglement

NE_1 : no entanglement



**Reinforcement
learning**



Quantum Reinforcement Learning

Agent interacts with environment

- Follow **policy**
- Find policy that **maximizes reward**

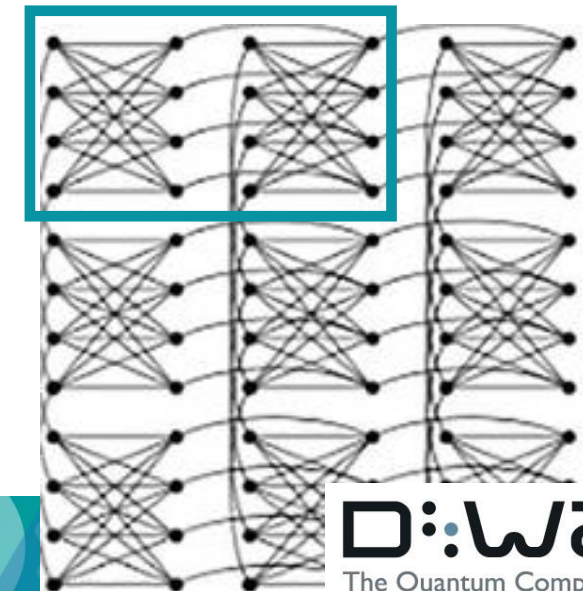
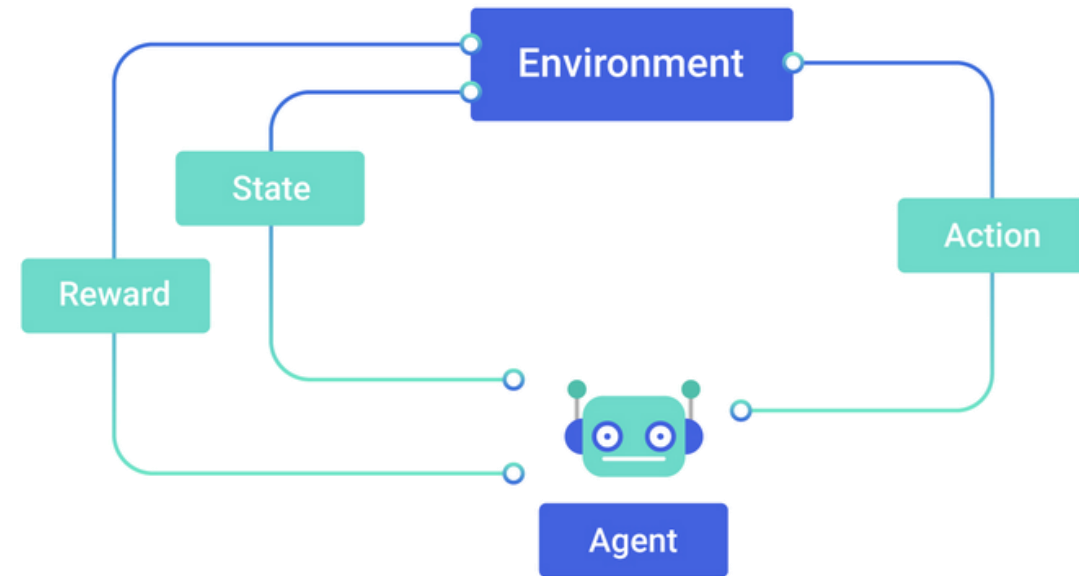
Expected reward is estimated by *value function* $Q(s, a)$

- **DQN:** Deep Q-learning (*NN-based*)
- **FERL:** Free energy-based RL (*clamped Quantum Boltzmann Machine*)

Implement the **quantum NN** on a set of qubits

Quantum computer calculates the **reward as the energy** of the qubit system

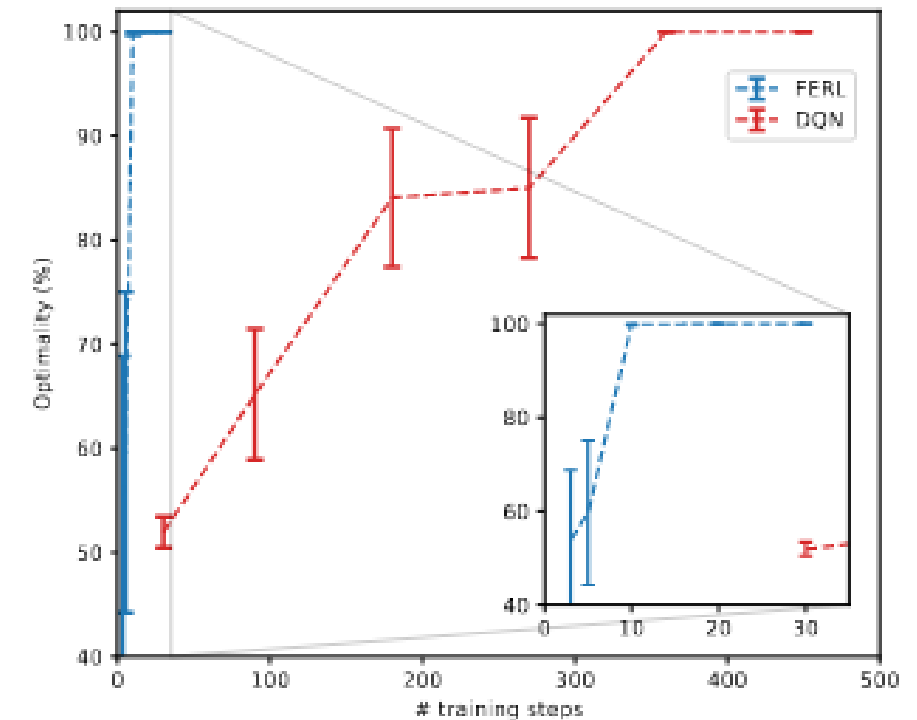
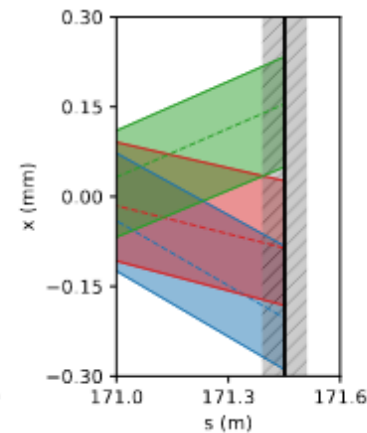
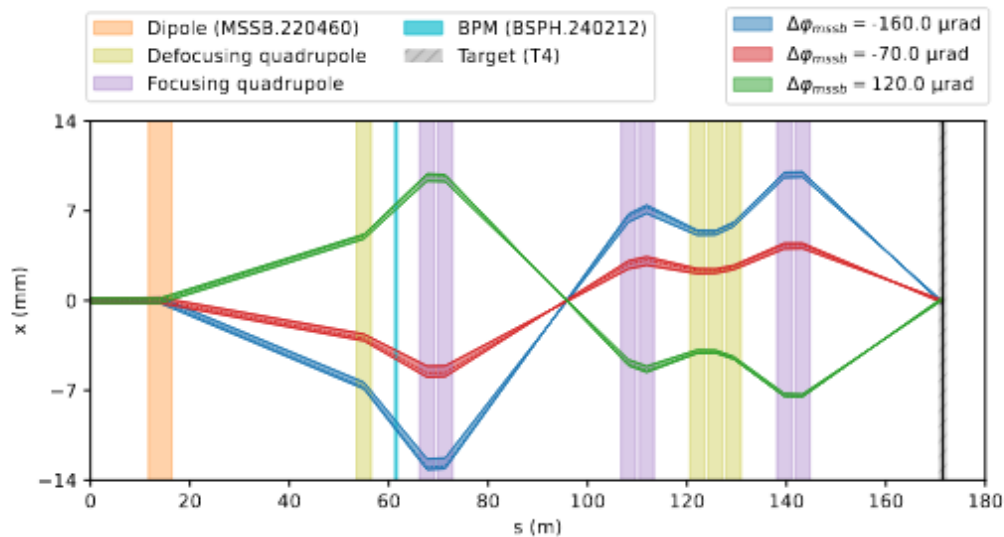
In this framework the **agent is classical**



Beam optimisation in linear accelerators

- **Action:** (discrete) deflection angle
- **State:** (continuous) BPM position
- **Reward:** integrated beam intensity on target
- **Optimality:** fraction of states in which the agent takes the right decision

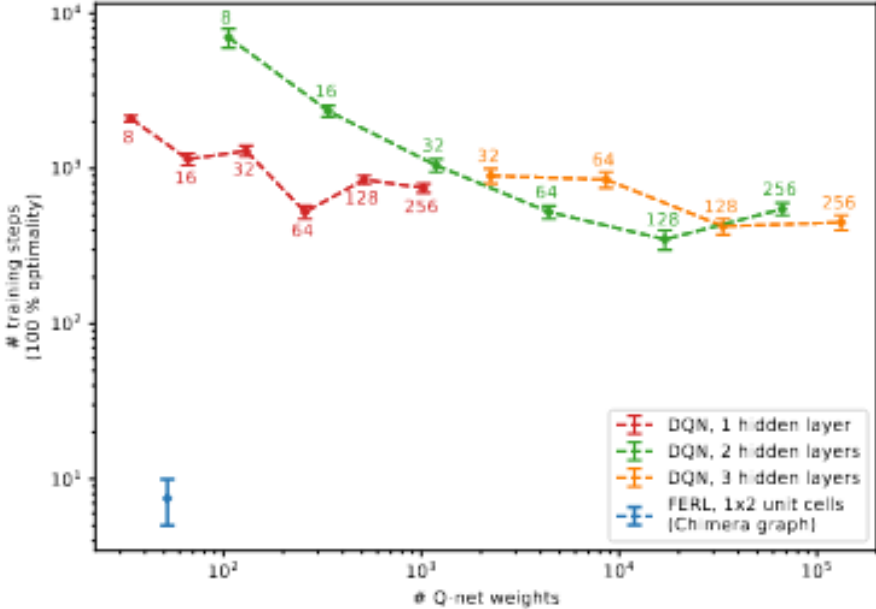
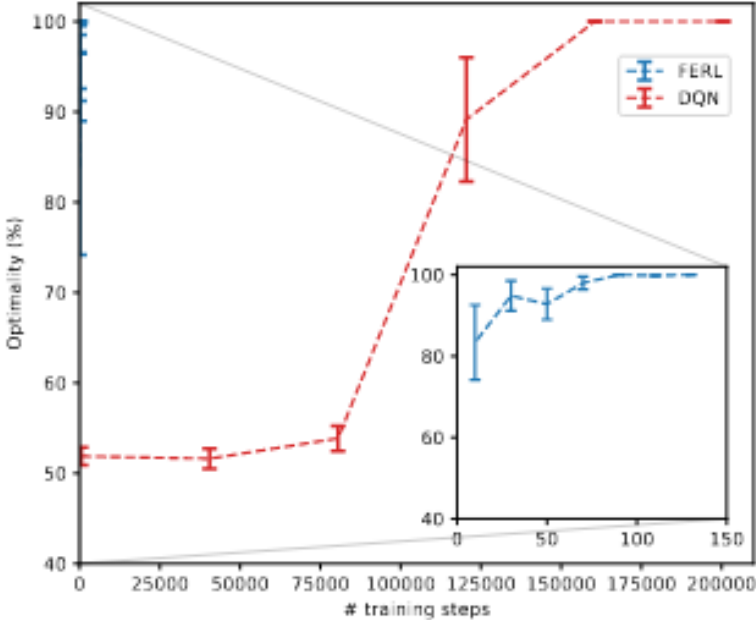
- **Quantum RL massively outperforms** classical Q-learning (8 ± 2 vs. 320 ± 40 steps with *e. r.*)



Convergence and representational power

QRL use cases confirms advantage in terms of **model size** and **training steps**

Without **experience replay**

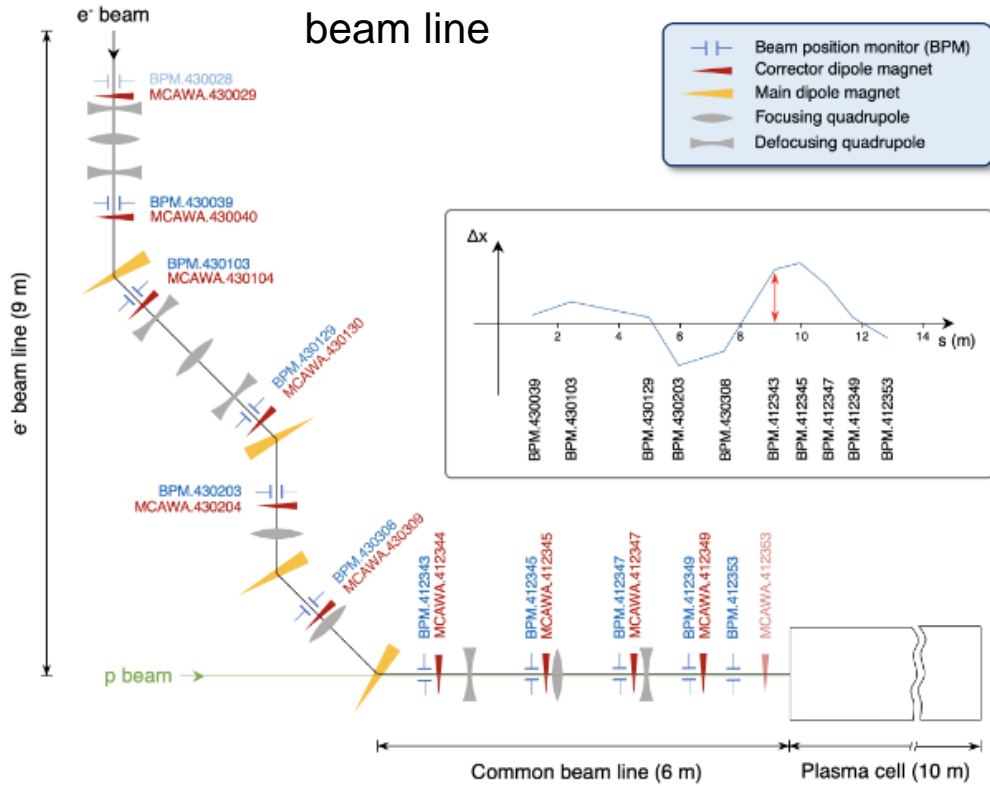


Michael Schenk, Elías F. Combarro, Michele Grossi, Verena Kain, Kevin Shing Bruce Li, Mircea-Marian Popa, Sofia Vallecorsa, **Hybrid actor-critic algorithm for quantum reinforcement learning at CERN beam lines**. arXiv:2209.11044

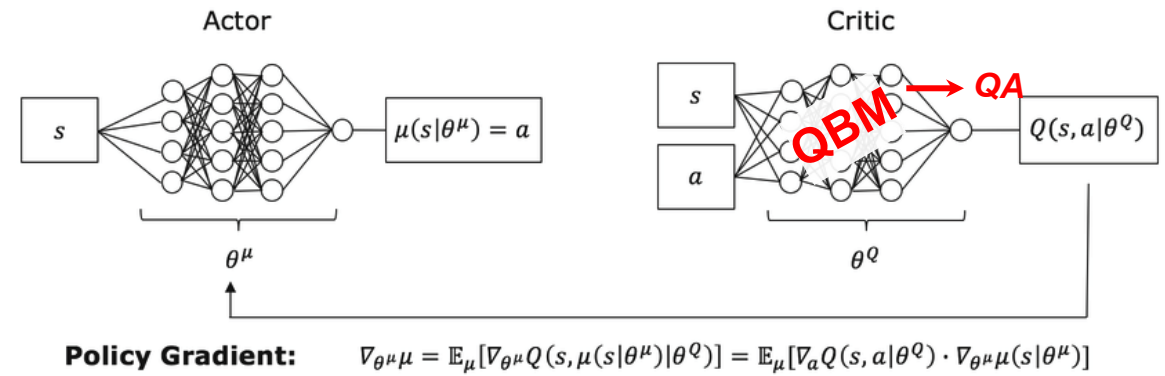


CERN AWAKE facility

2GeV electron beam line



Michael Schenk et al., Hybrid actor-critic algorithm for quantum reinforcement learning at CERN beam lines, e-Print: 2209.11044 [quant-ph]



Actor-Critic Q-learning training D-Wave Advantage

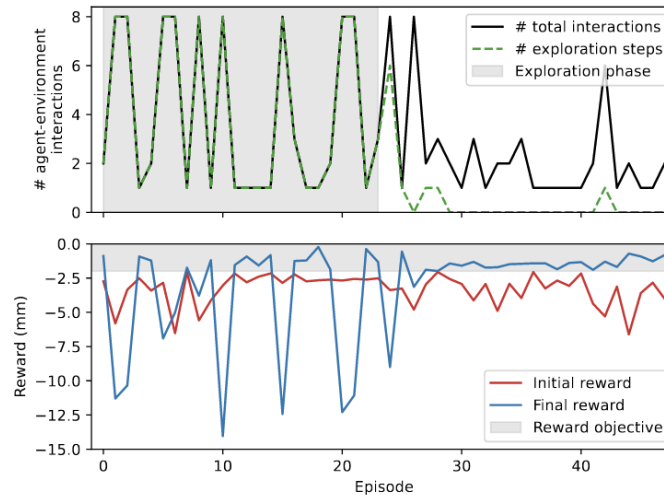
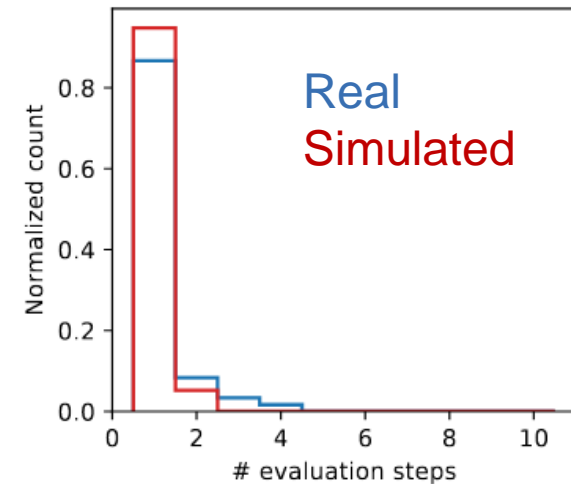


Figure 11: Single RL agent training evolution on D-Wave Advantage Systems using the simulated AWAKE environment with a reward objective of -2 mm.

Successful evaluation on the real beam-line





Improving
robustness



Improving robustness

- **Correlate expected model performance to data set properties**
- **Stabilizing training on NISQ**
- **Trainability vs expressivity robustness studies**
- **Evaluating generalisation**
- **Quantum vs classical data**
- **Algorithms beyond QML**



Improving robustness

- Correlate expected model performance to data set properties
- **Stabilizing training on NISQ**
- Trainability vs expressivity robustness studies
- Evaluating generalisation
- **Quantum vs classical data**
- Algorithms beyond QML

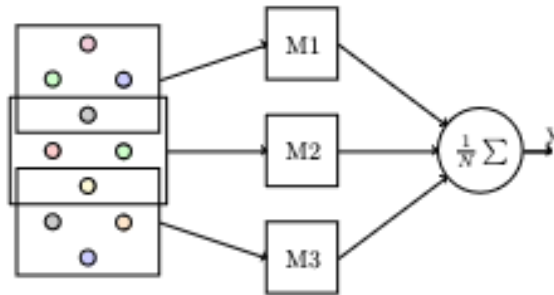


Ensembles of quantum neural networks

NISQ regime affects QML performance. Can we build ensembles?

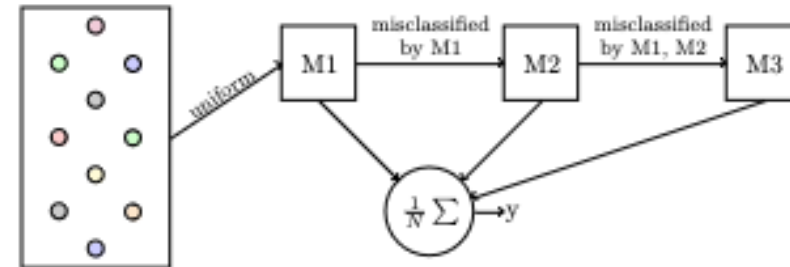
Bagging: best for **high variance**; reduces BPs by keeping the feature space limited

- 10 independently trained instances
- r_f :% of samples, r_n :% features



Boosting: **high bias** models (little sensitivity to subsampling)

- AdaBoost, 10 repetitions



Study **regression** and **classification** tasks in toy and realistic datasets

Dataset	Source	Nature	# Features	# Samples	Task
Linear	-	Synthetic	5	250	Regression
Concrete	UCI	Real-world	8	1030	Regression
Diabetes	Scikit-Learn	Real-world	10	442	Regression
Wine	UCI	Real-world	13	178	Classification

QNN setup and simulated results

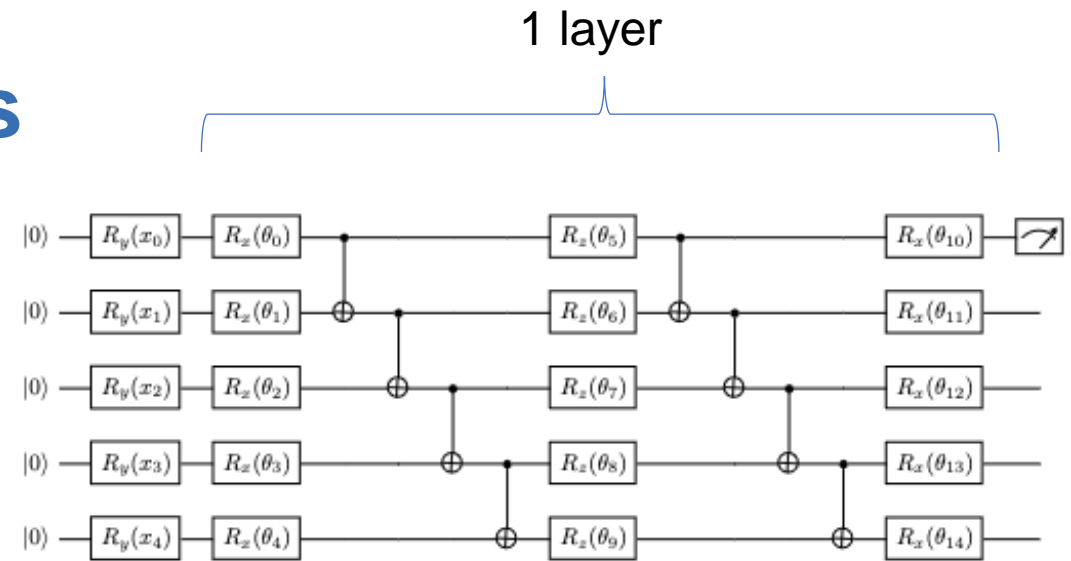
Choose **relatively simple QNN**:

n qubits = n features

Ry single rotation gates

CNOT in linear entanglement

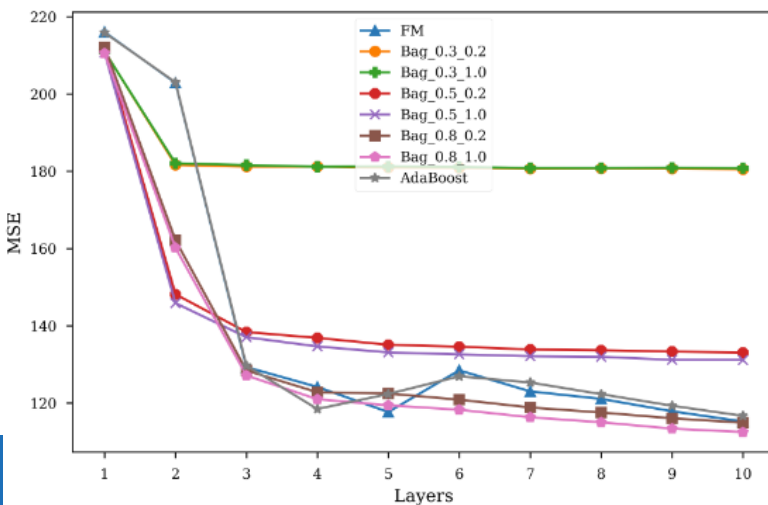
Local observable (σ_z)



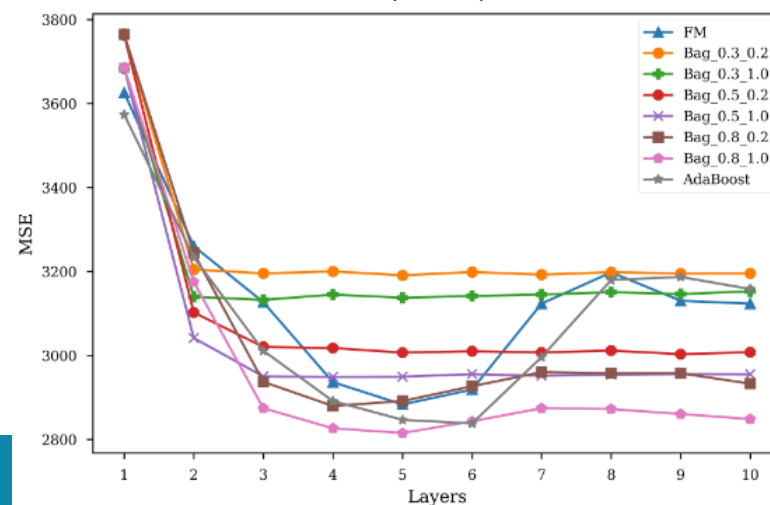
Measure the generalisation error on **test sample (20 %)**

Bagging methods outperform full model and Boosting: **shallower networks, fewer input features**

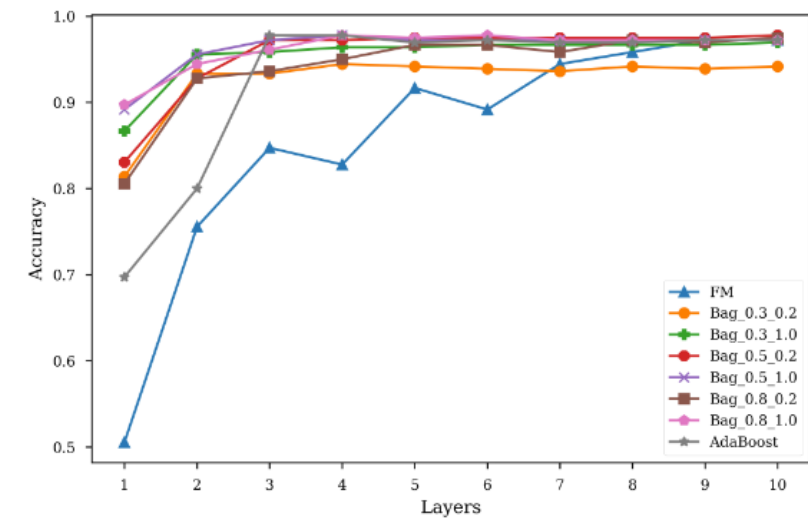
Concrete (MSE)



Diabetes (MSE)

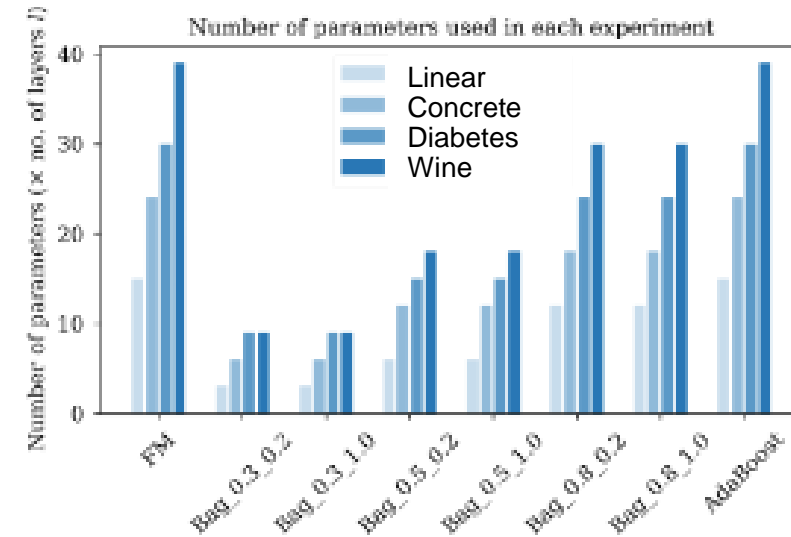
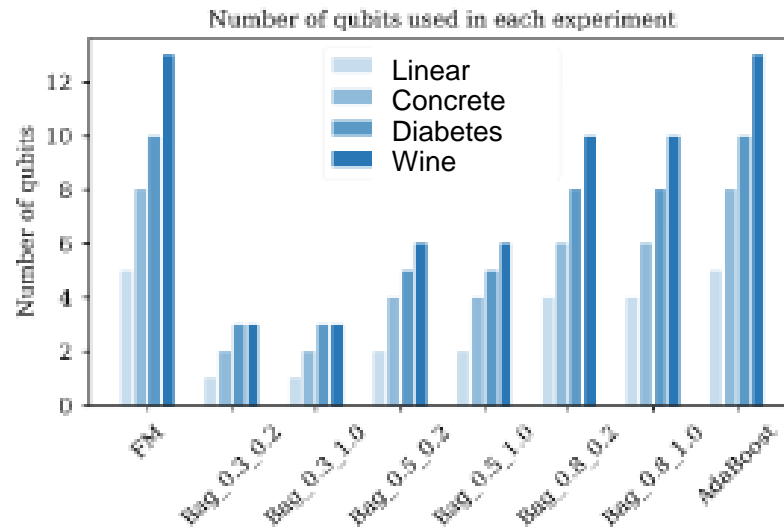


Diabetes (CCE)



Bagging brings significant advantage

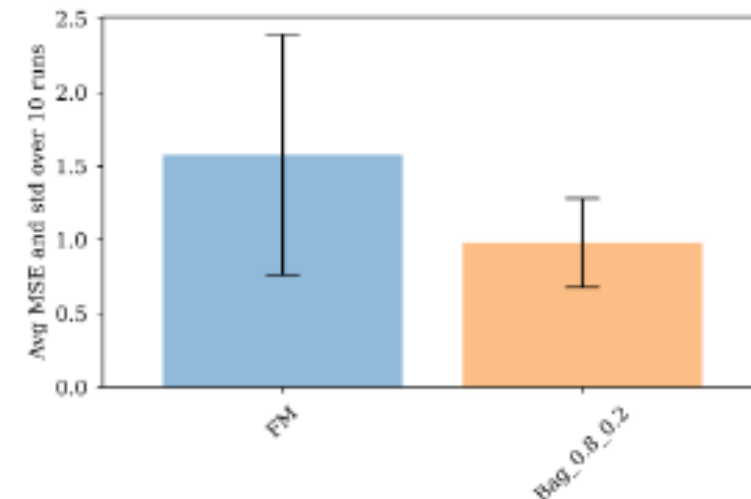
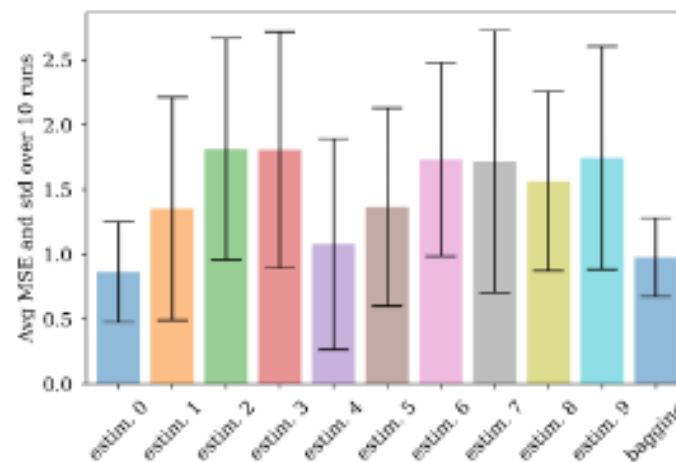
Reducing resources:
Best performance for low dimensionality



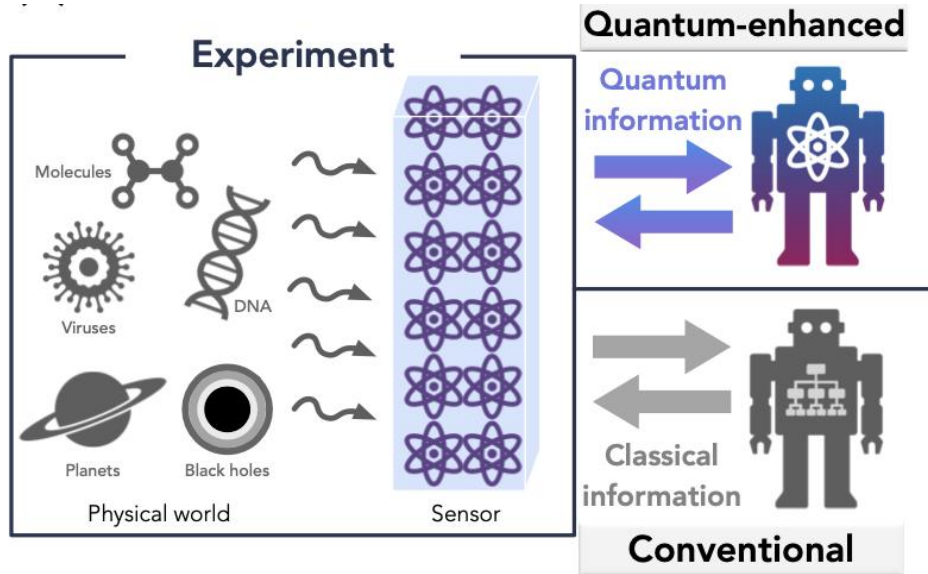
Robustness against noise:
Linear regression task on **IBM QPU**
(ibm_lagos):

Bagging: 80% features, 20% samples

QNN: 4 qubit, 1 layer



Quantum machine learning for quantum data



Huang, *et al.*, *Science* **376**, 6598 (2022)

Work directly with quantum states.

Task: Drawing phase diagrams

1. **Supervised classification** using a convolutional QNN using the groundstates as input data.
2. Advantageous since quantum states are **exponentially hard to save classically**.
3. **Bottleneck**: we need access to classical training labels! Interpolation does not work

Cong, *et al.*, *Nat. Phys.* **15**, 1273–1278 (2019)

Setting the stage

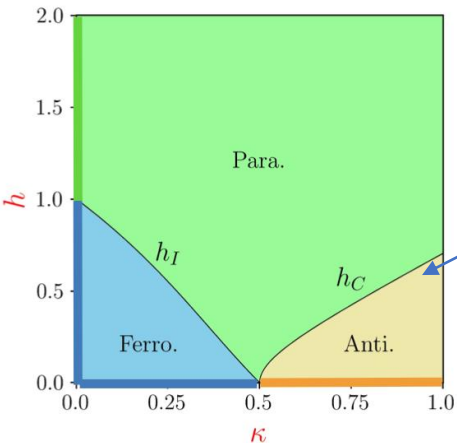
- Train in easy (integrable) subregions
- Generalize to a full model¹
- Model: Axial Next Nearest Neighbor

Ising (ANNNI) Hamiltonian:

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

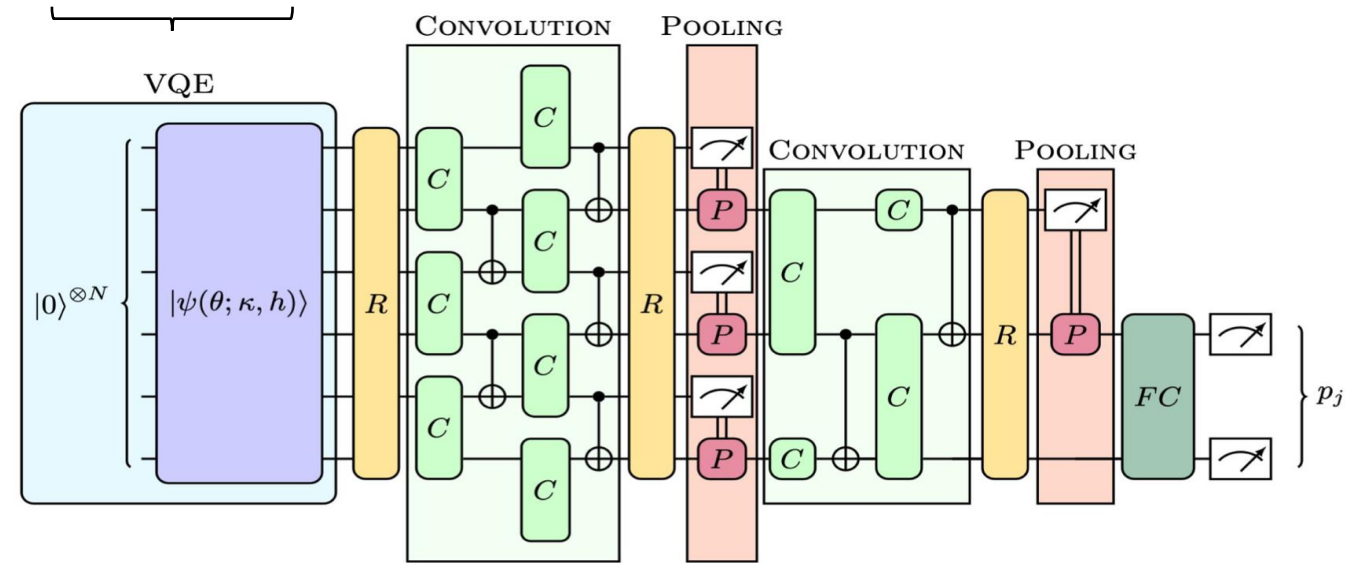
Senk, *Physics Reports*, **170**, 4 (1988)

Which is integrable for $\kappa = 0$ or $h = 0$.



Monte Carlo,
DMRG

Variational quantum data



Binary Cross-entropy

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

Labels:

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

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

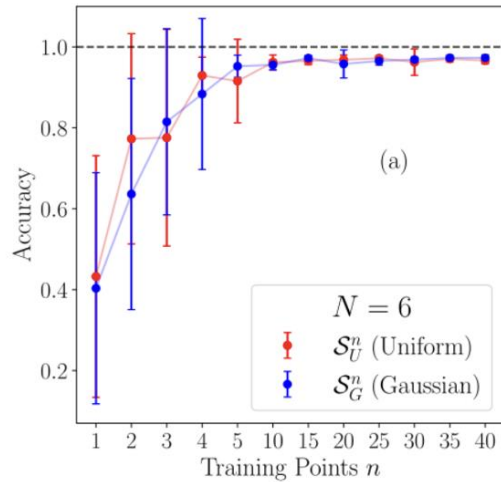
Results

Learn a similarity function between the data.

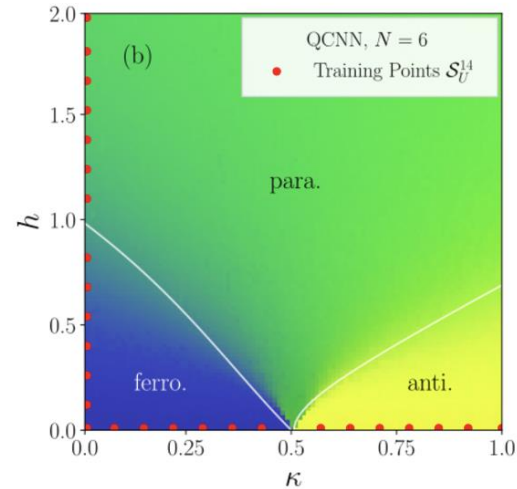
Kottman, *et al.*, *Phys. Rev. Research* **3**, 043184 (2021)

$N = 6$

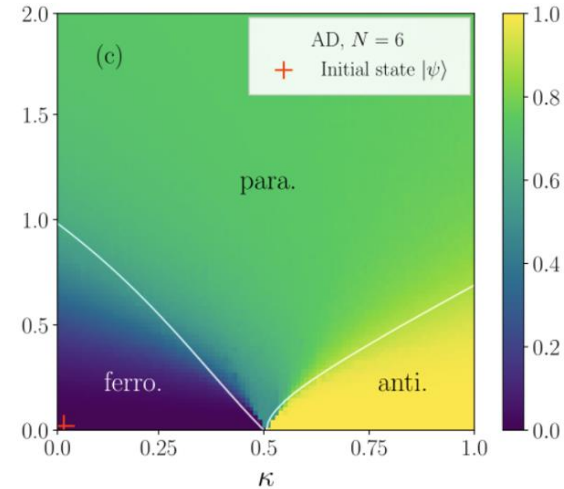
Size of training set



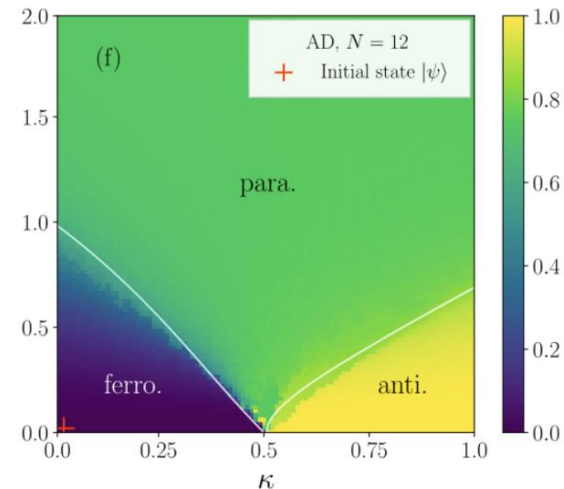
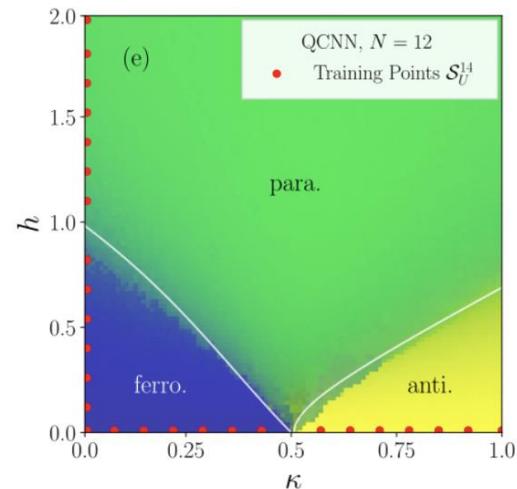
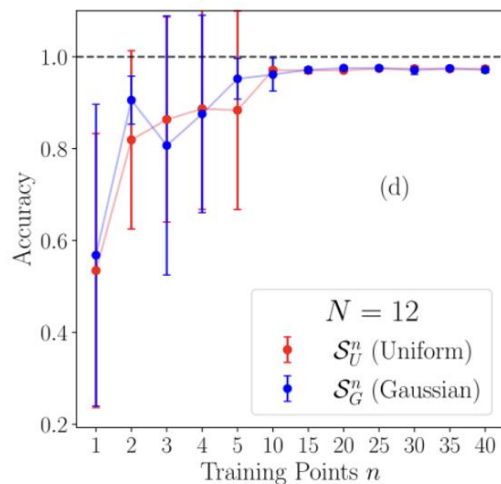
QCNN (95%)



Autoencoder



$N = 12$



1. Out of Distribution Generalization? [M..Caro et al., Out-of-distribution generalization for learning quantum dynamics, <https://arxiv.org/abs/2204.10268>]
2. Performance increases with the system's size.
3. Addresses the bottleneck of needing expensive training labels.
4. QCNN gives quantitative predictions [Banchi et al., Generalization in Quantum Machine Learning: A Quantum Information Standpoint, PRX QUANTUM 2, 040321 (2021)]

Perspective

The CERN QTI is studying impact of Quantum Technologies in High Energy Physics:

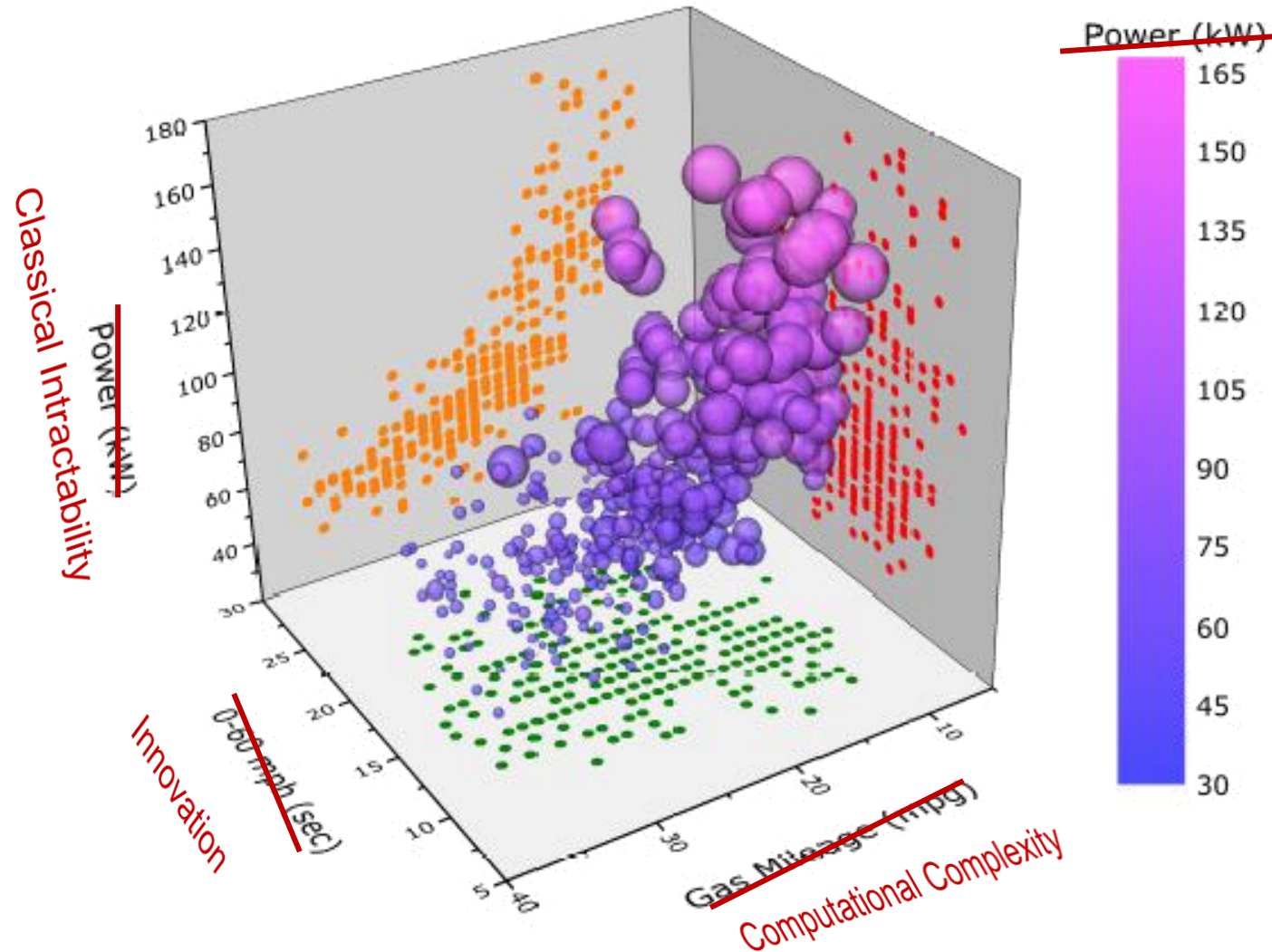
- Some **preliminary hints** of advantage
- So far.. we can do «**as good as classical methods**». In many cases, limitations are hardware-related
- Need more **robust studies** to estimate **performance** and drive **model development**

We are now formulating a **longer term research plan**



Exclusion Region for QML in HEP?

QML is the right solution



Thank you!

November 20th-24th, 2023
@CERN



Sofia.Vallecorsa@cern.ch



QUANTUM
TECHNOLOGY
INITIATIVE

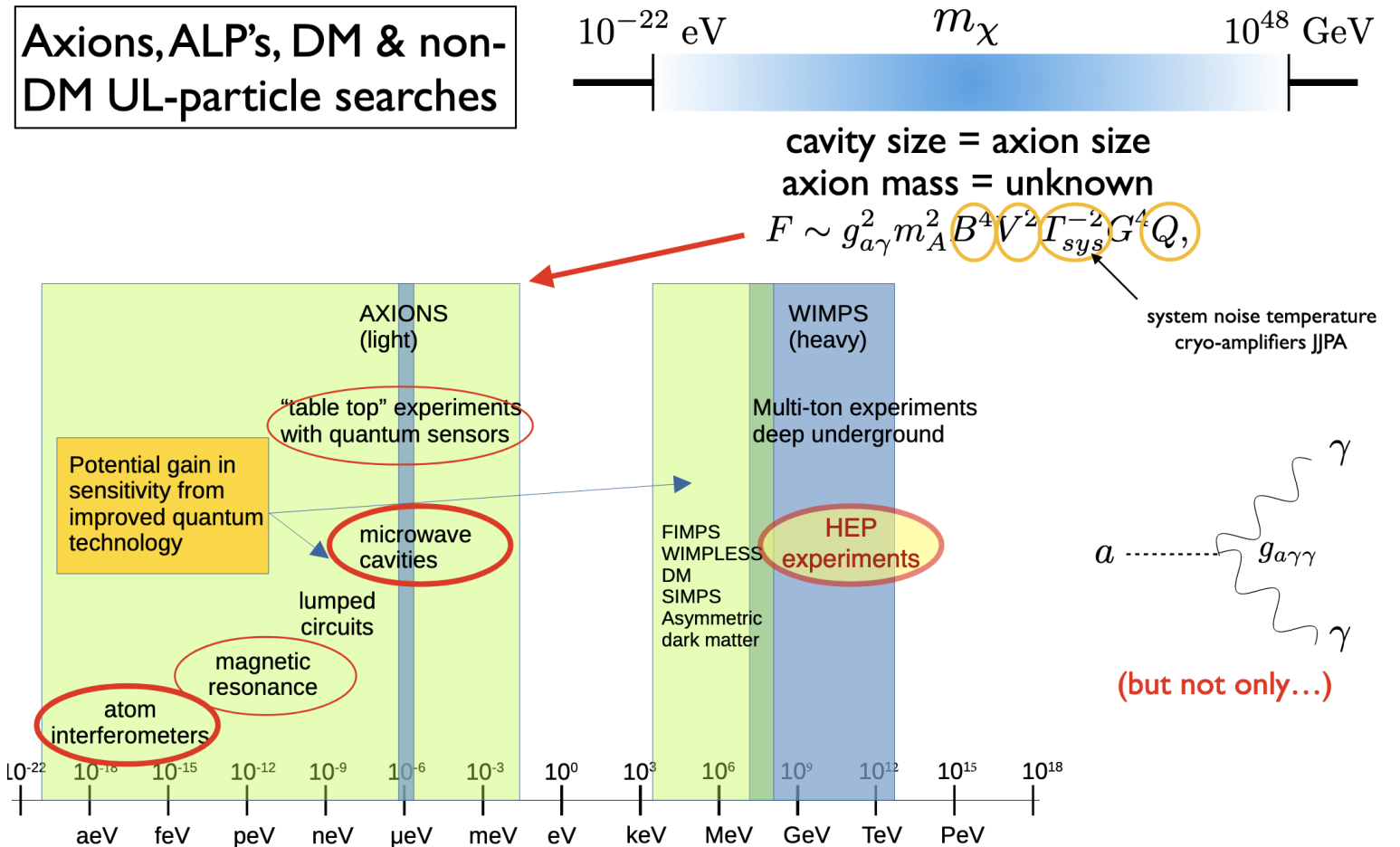


Quantum sensing

Change of quantum state caused by the interaction with an external system:

- transition between superconducting and normal-conducting
- transition of an atom from one state to another
- change of resonant frequency of a system (quantized)

quantum sensors & particle physics: what are we talking about?



Theory and Simulation

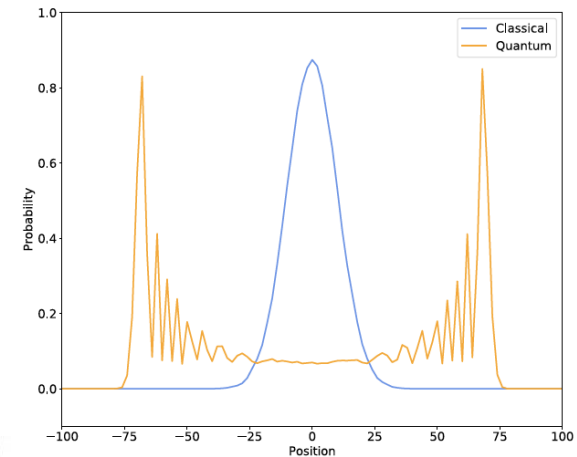
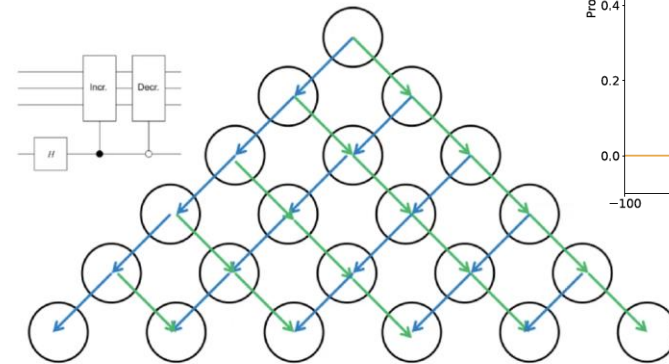
QFT: Focus on computations that are exponentially hard with classical methods. Ex. Sign problems in particle theory

- Dynamical Simulations of Lattice Gauge Theories
- Finite-Density Nuclear Matter
- Challenges related to digitization and truncation of field representation (on a finite number of quantum states) and redundancy in the Hilbert space¹

Cross section integration as quantum amplitude estimation³

Event generation with quantum generative models or direct simulation

Parton showering as quantum random walk²



¹ D. Grabowska's presentation at the CERN QTI workshop (<https://indico.cern.ch/event/1098355>)

² A quantum walk approach to simulating parton showers Khadeejah Bepari, Sarah Malik, Michael Spannowsky, Simon Williams arxiv:2109.13975 and presentation at the CERN QTI workshop (<https://indico.cern.ch/event/1098355>)

³ Agliardi, Gabriele, et al. "Quantum integration of elementary particle processes." *arXiv preprint arXiv:2201.01547* (2022)

qGAN Benchmarks on hardware

Train models using **noisy simulator** and test the inference on **trapped-ion (IONQ) quantum hardware**

- For IBMQ machines, choose the qubits with the lowest

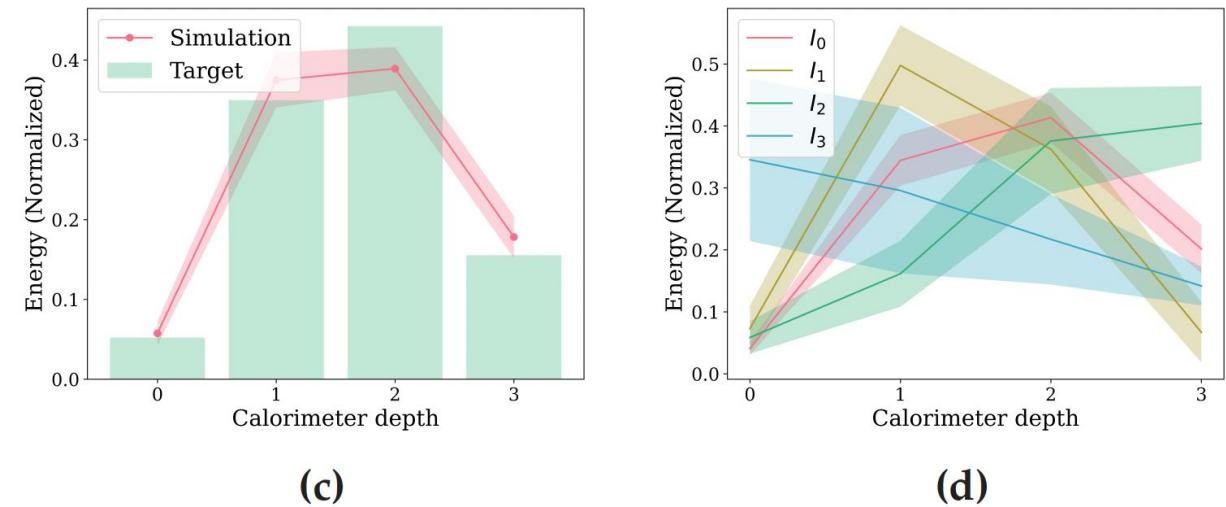
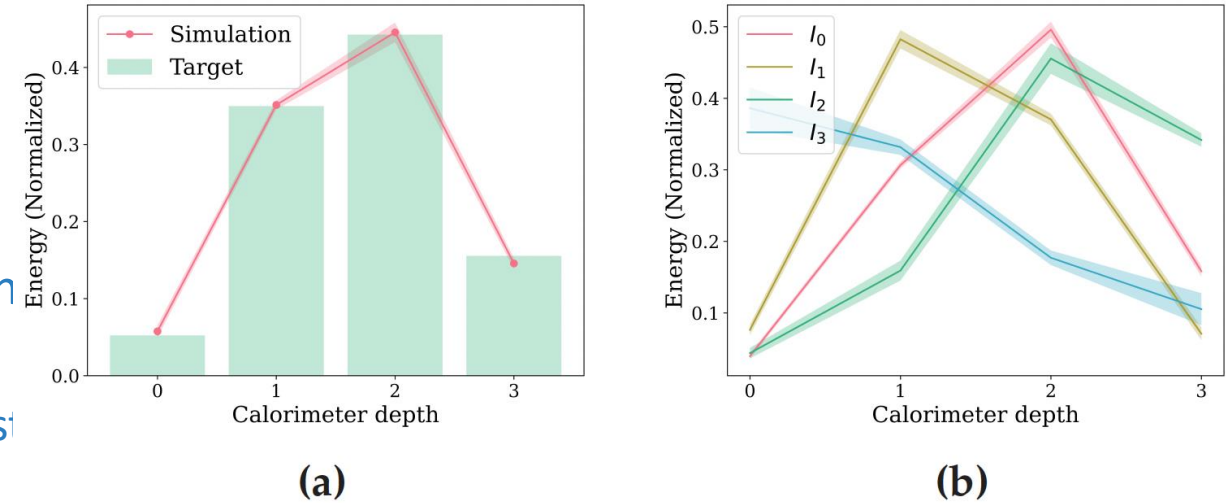


Figure 4: Mean (a,c) and individual images (b,d) obtained by inference test on `ibmq_jakarta` (a,b) and `IONQ` (c,d).

Device	Readout error CX error	$D_{KL}/D_{KL,ind}$ ($\times 10^{-2}$)
<code>ibmq_jakarta</code>	0.028 $1.367 \cdot 10^{-2}$	0.14 ± 0.14 6.49 ± 0.54
<code>ibmq_lagos</code>	0.01 $5.582 \cdot 10^{-3}$	0.26 ± 0.11 6.92 ± 0.71
<code>ibmq_casablanca</code>	0.026 $4.58 \cdot 10^{-2}$	4.03 ± 1.08 6.58 ± 0.81
<code>IONQ</code>	NULL $1.59 \cdot 10^{-2}$	1.24 ± 0.74 10.1 ± 5.6