# Quantum Machine Learning in High Energy Physics

*Examples from CERN*

QUANTUM TECHNOLOGY INITIATIVE

**Sofia Vallecorsa**

AI & Quantum Research - CERN IT

April **26th** 2023

# Outline

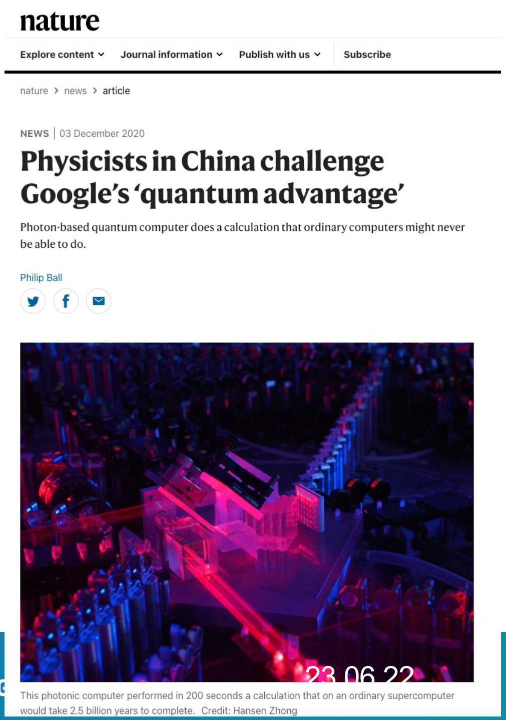
1. **Introduction: the CERN Quantum Technology Initiative**
2. **Quantum Machine Learning and Applications at CERN**
3. **Anomaly Detection**
4. **Beam Optimisation in linear accelerators**
5. **Improving robustness**
   - Stabilizing training on NISQ
   - Quantum data
6. **Summary**

# Hype and Potential...

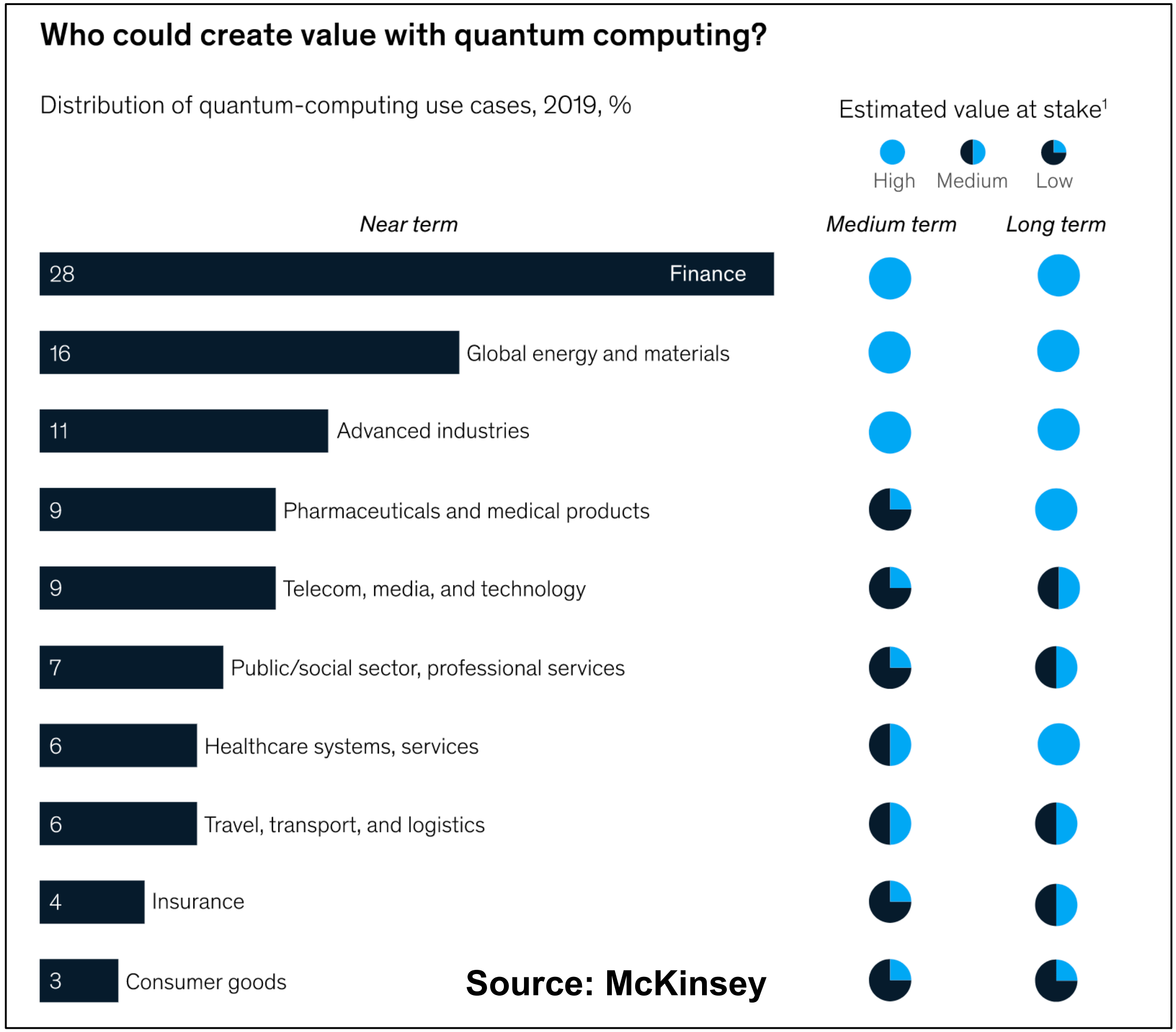
## 2019: Google


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

## 2020: Hefei National Lab


nature
Explore content Journal information Publish with us Subscribe
nature > news > article
NEWS | 03 December 2020
**Physicists in China challenge Google's 'quantum advantage'**
Photon-based quantum computer does a calculation that ordinary computers might never be able to do.
Philip Ball
This photonic computer performed in 200 seconds a calculation that on an ordinary supercomputer would take 2.5 billion years to complete. Credit: Hansen Zhong

https://www.nature.com/articles/d41586-020-03434-7




QC use cases in different sectors: the situation in 2019 with the estimated **medium** (2025) **and long** (2035) **term impact**.

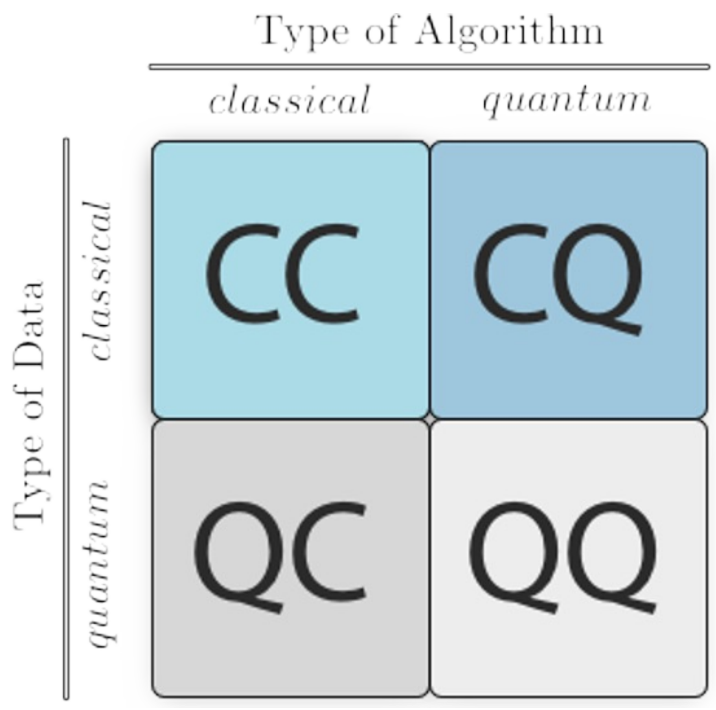
# Quantum Technologies


| | Superconducting loops | Trapped ions | Silicon quantum dots | Topological qubits | Diamond vacancies |
|---|---|---|---|---|---|
| | A resistance-free current oscillates back and forth around a circuit loop. An injected microwave signal excites the current into super-position states. | Electrically charged atoms, or ions, have quantum energies that depend on the location of electrons. Tuned lasers cool and trap the ions, and put them in superposition states. | These "artificial atoms" are made by adding an electron to a small piece of pure silicon. Microwaves control the electron's quantum state. | Quasiparticles can be seen in the behavior of electrons channeled through semi-conductor structures.Their braided paths can encode quantum information. | A nitrogen atom and a vacancy add an electron to a diamond lattice. Its quantum spin state, along with those of nearby carbon nuclei, can be controlled with light. |
| **Number entangled** | 9 | 14 | 2 | N/A | 6 |
| **Company support** | Google, IBM, Quantum Circuits | ionQ | Intel | Microsoft, Bell Labs | Quantum Diamond Technologies |
| **Pros** | Fast working. Build on existing semiconductor industry. | Very stable. Highest achieved gate fidelities. | Stable. Build on existing semiconductor industry. | Greatly reduce errors. | Can operate at room temperature. |
| **Cons** | Collapse easily and must be kept cold. | Slow operation. Many lasers are needed. | Only a few entangled. Must be kept cold. | Existence not yet confirmed. | Difficult to entangle. |

# Algorithms & Applications

**Quantum effects (superposition entanglement, no-cloning theorem, … ) improve and accelerate complex algorithms**

- Efficient **sampling**, **searches** and **optimization**
- Linear algebra, matrices and machine learning
- New algorithms/methods for **cryptography** and **communication**

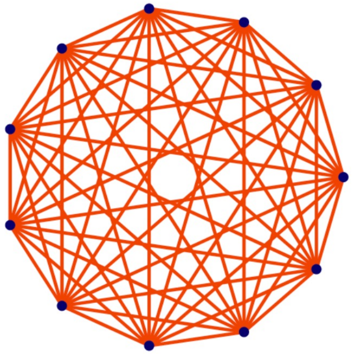



**Challenge is re-thinking algorithms design and define fair benchmarking and comparison to classical algorithms**

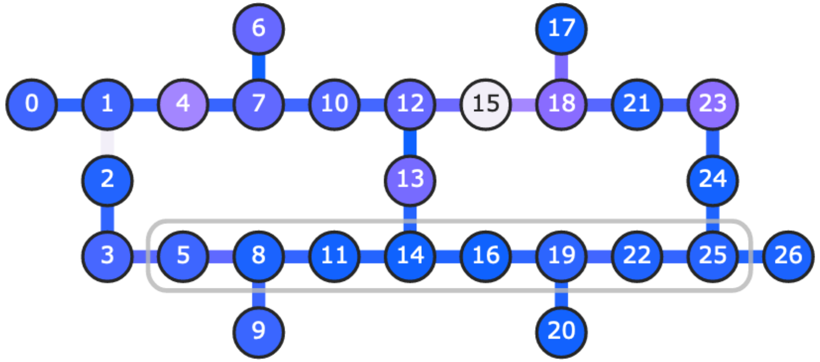
# Noisy Intermediate-Scale Quantum devices

- 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”

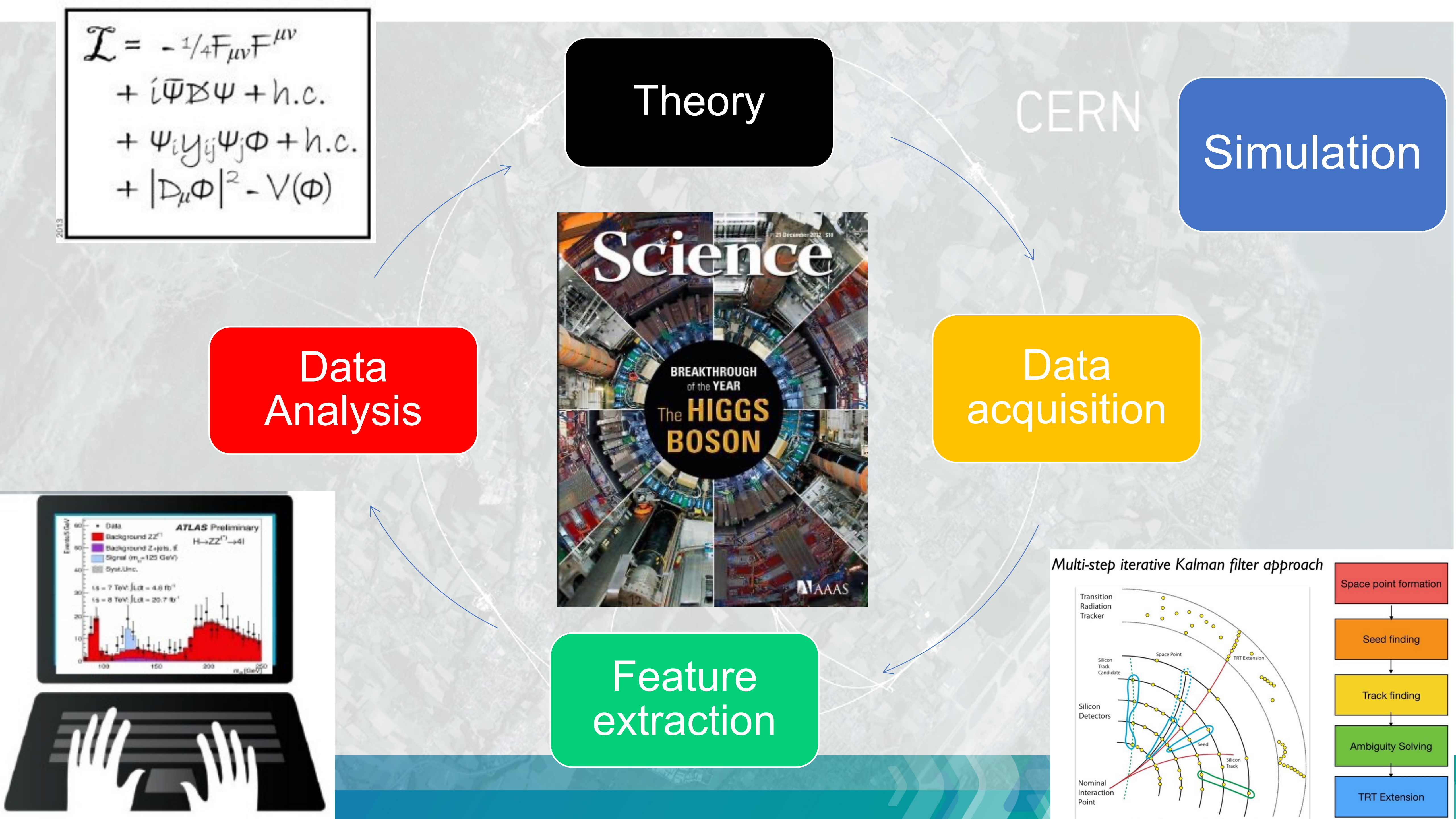
**Trapped ion technology:** *ionQ*
with all-to-all connectivity

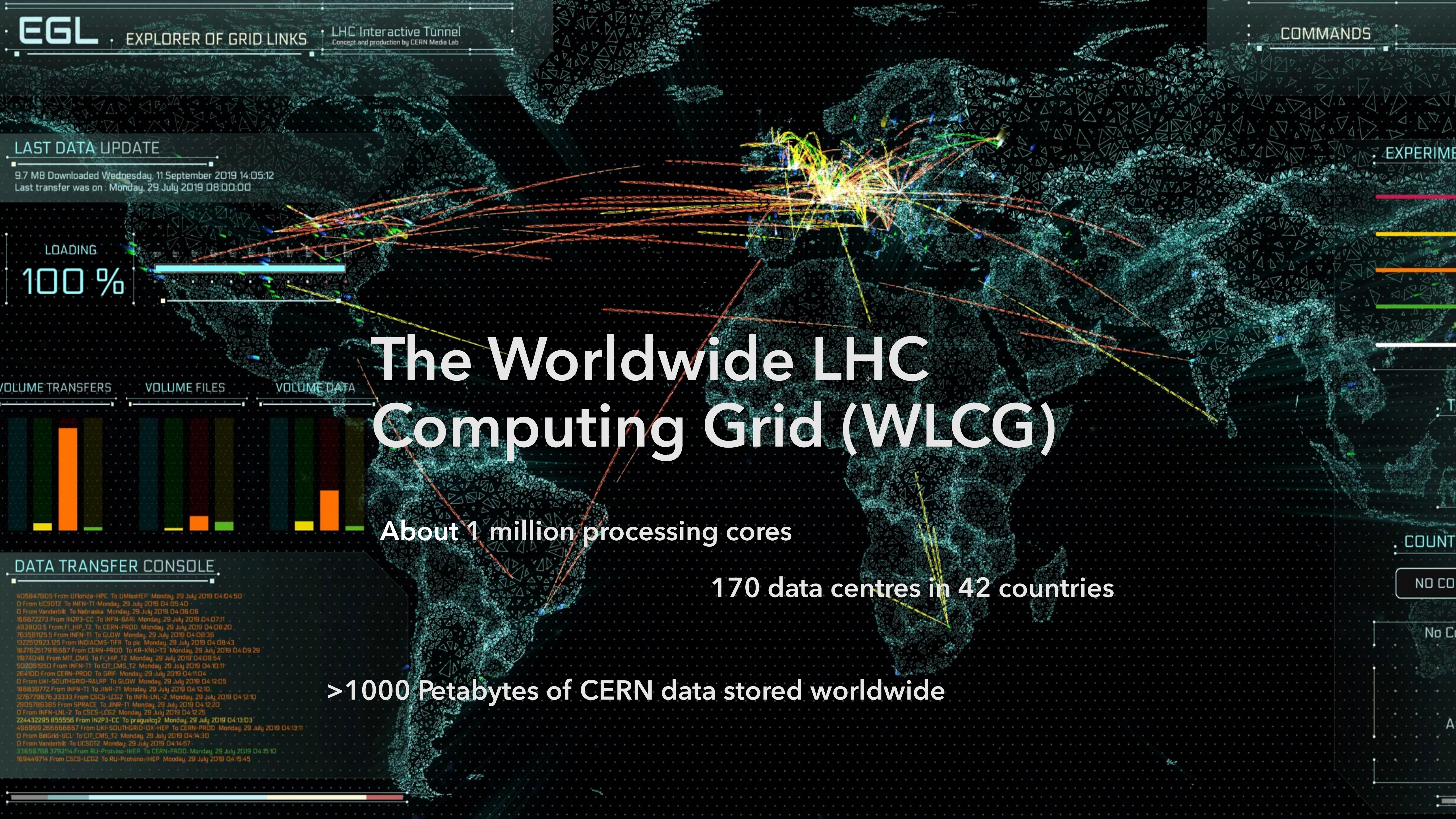


**Semiconducting transmon qubits**:
*IBM Toronto*



# HL-LHC computing challenges


$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\Psi}\not{D}\Psi + h.c. + \Psi_i y_{ij}\Psi_j\Phi + h.c. + |D_\mu\Phi|^2 - V(\Phi)$
Theory
CERN
Simulation
Science
BREAKTHROUGH of the YEAR
The HIGGS BOSON
AAAS
Data Analysis
Data acquisition
Feature extraction
ATLAS Preliminary
Data
Multi-step iterative Kalman filter approach
Transition Radiation Tracker
Space Point
TRT Extension
Silicon Detectors
Seed
Nominal Interaction Point
Space point formation
Seed finding
Track finding
Ambiguity Solving
TRT Extension




# The Worldwide LHC Computing Grid (WLCG)

About 1 million processing cores

170 data centres in 42 countries

>1000 Petabytes of CERN data stored worldwide

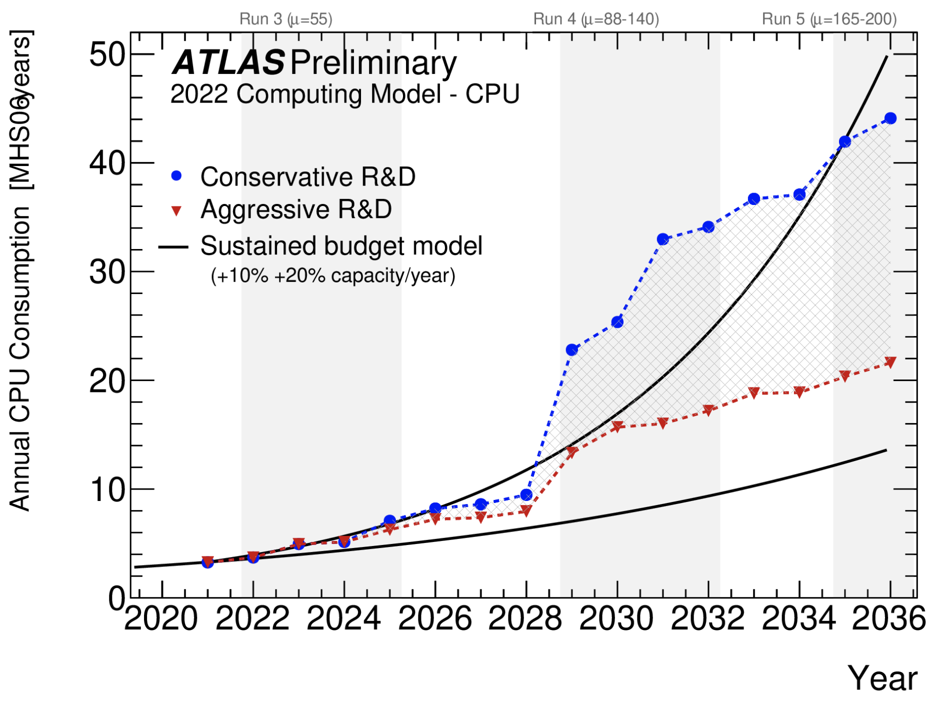
# HL-LHC: Computing Challenges

**High Luminosity LHC can do physics @ unprecedented level of precision**

**Higgs** : measure fermions and bosons couplings at % level

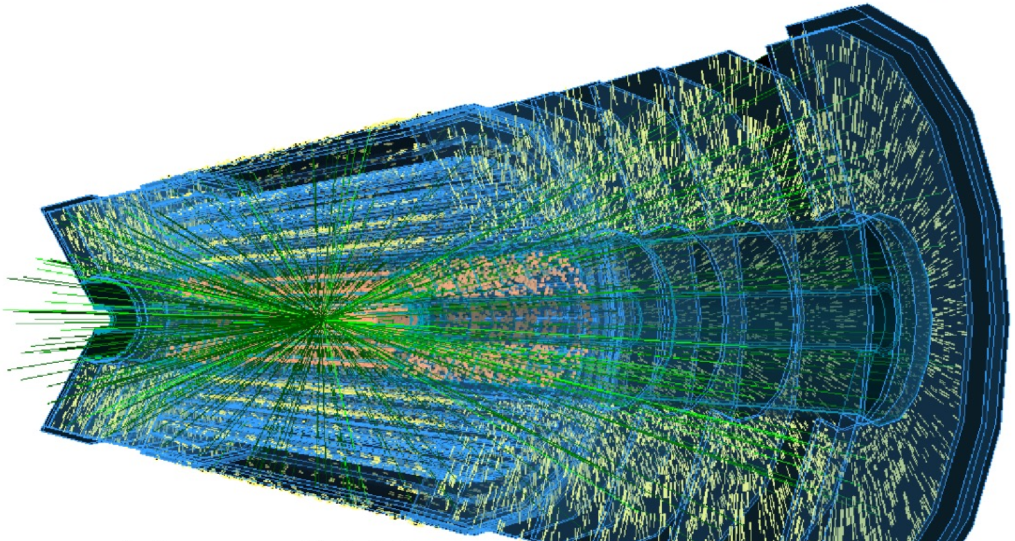
**Electro-weak sector**, **top** quark, **multi-bosons** states

**New physics** searches, **dark matter**, etc..




**Large amount of data, O(100) simultaneous collisions, high granularity detectors will require:**

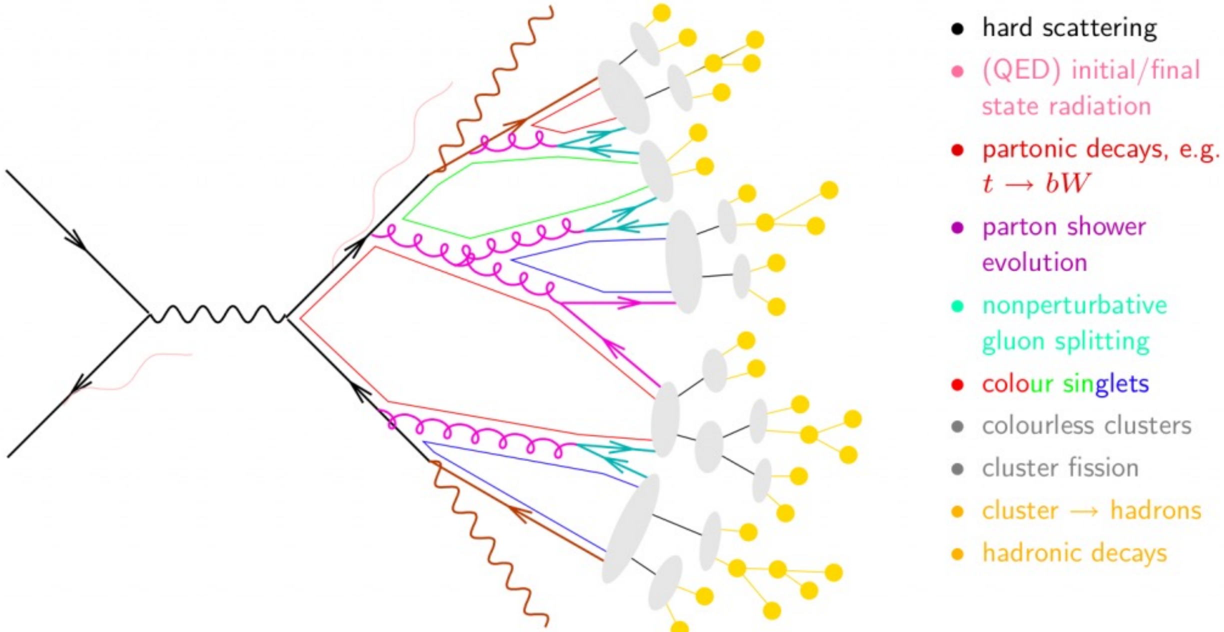
- **Equally accurate theoretical predictions**: improved theory calculations, faster Monte Carlo simulation
- **Fast and accurate analysis methods** (AI-based?)



**ITk**: new ATLAS inner detector for HL-LHC

S. Campana et al. arXiv:2203.07237

# Theory and Simulation




**Quantum Field Theory[1]**

- Lattice QCD, Sign problems

**Parton showering**

**Event Generation & Cross section integration**

**Phase space sampling scales exponentially with number of final state particles[2]**

- HL-LHC @ 3 $10^{-3}$ $fb^{-1}$ will have percent-level precision @$N_{jet}$ = 9
- Need comparable (higher-order) MC
- $N_{jet}$ increases with center-of-mass energy

**Precision studies at FCC**

**Time and memory usage (Sherpa 3.x.y + HDF5) (H. Schulz 2018)**

| Process $W^{+}+$ | 5j | 6j* | 7j* | 8j† |
|---|---|---|---|---|
| RAM Usage | 189 MB | 484 MB | 1.32 GB | 1.32 GB |
| Init/startup time | 3m5s / 1s | 24m52s / 5s | 3h6m / 18s | 5h55m / 20s |
| Integration time | 128×4h38m | 256×13h53m | 512×19h0m | 1024×23h8m |
| MC uncertainty | 1.0% | 0.99% | 2.38% | 4.68% |
| Unweighting eff | $9.56 \cdot 10^{-5}$ | $7.66 \cdot 10^{-5}$ | $7.20 \cdot 10^{-5}$ | $7.51 \cdot 10^{-5}$ |
| 10k evts | 24h 40m | 2d 11h | 10d 15h | 78d 1h |

Numbers generated on dual 8-core Intel® Xeon® E5-2660 @ 2.20GHz
*,† Number of quarks limited to ≤6/4

1 D. Grabowska's presentation at the CERN QTI workshop ([https://indico.cern.ch/event/1098355](https://indico.cern.ch/event/1098355))
2 arxiv:1905.05120

# Quantum Machine Learning

… or Quantum Computing to "improve" classical ML

# 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

- 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

**Major challenges:**

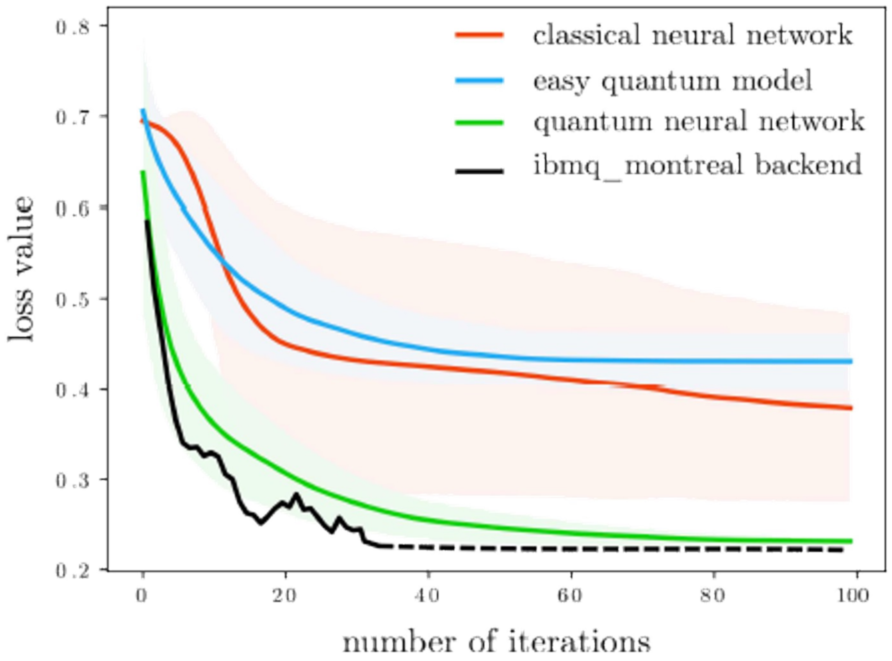
Defining fair **benchmarks**

Processing **large data sets**

Different **computational requirements**

# Quantum Advantage for QML

Abbas, Amira, et al. "**The power of quantum neural networks**." *Nature Computational Science* 1.6 (2021): 403-409.




**Different advantage definitions**

Runtime speedup

Sample complexity

Representational power

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

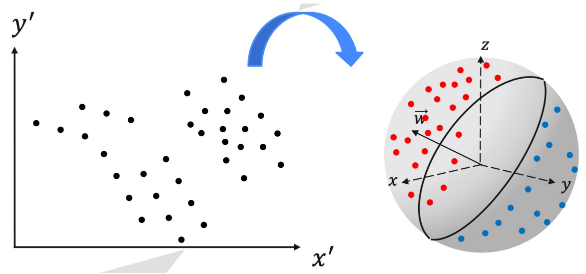
- No established recipe for classical data
- Need to use the **whole exponential advantage** in Hilbert space, but **will it converge** ?

(Algorithm expressivity vs convergence and generalization)

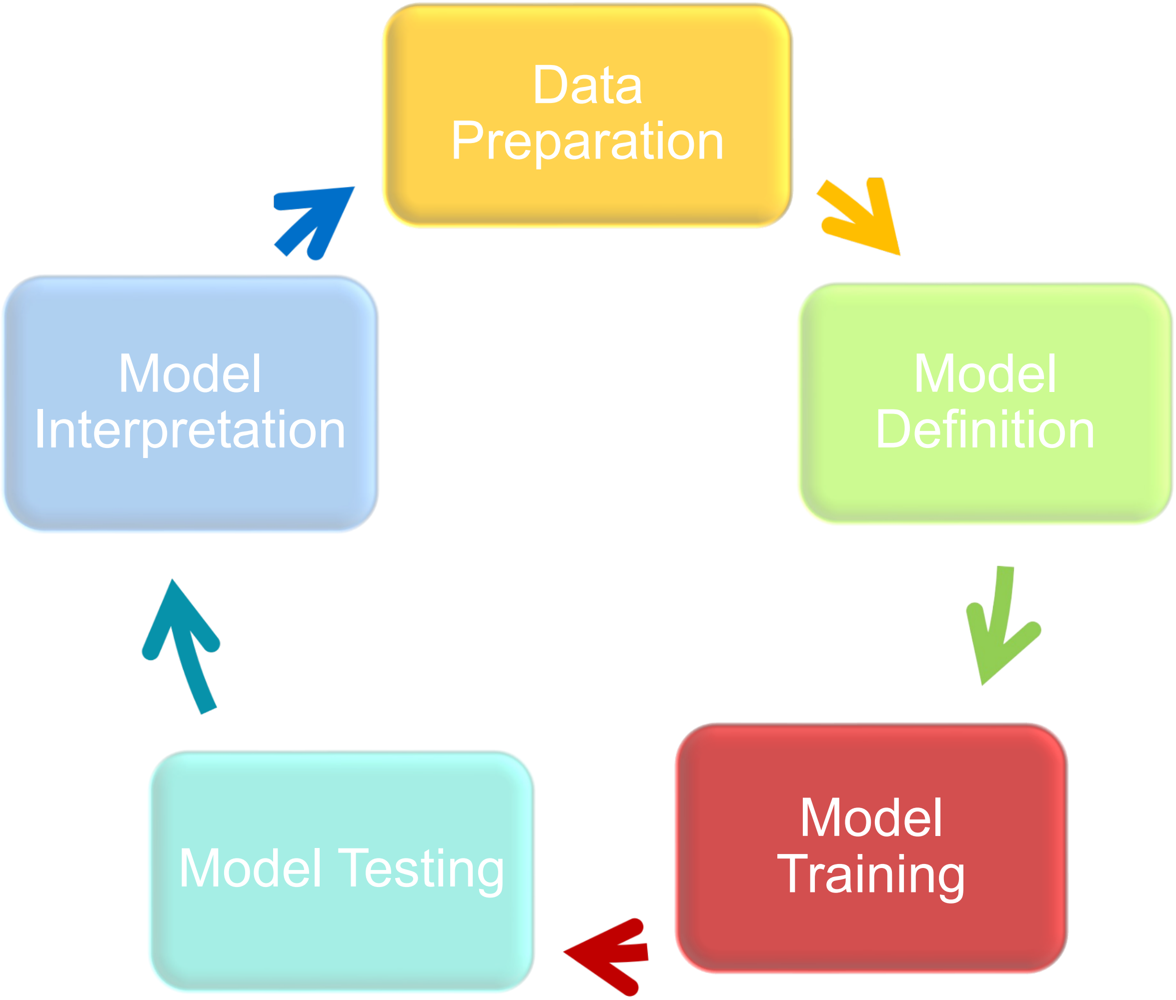
Kübler, Jonas, Simon Buchholz, and Bernhard Schölkopf. "**The inductive bias of quantum kernels.**" Advances in Neural Information Processing Systems 34 (2021).
Huang, HY., Broughton, M., Mohseni, M. et al. **Power of data in quantum machine learning**. Nat Commun 12, 2631 (2021). https://doi.org/10.1038/s41467-021-22539-9

# (Quantum) ML Lifecycle

Data Reduction
Data Encoding

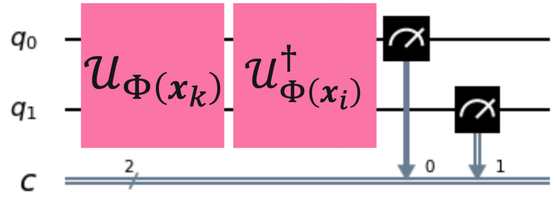



Trainability (BP…)
Generalization




Variational
Kernel based

Read Out and
measurement "shots"




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

# Model definition

## *Variational algorithms*

Parametric ansatz

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

**Data Embedding** can be **learned**

**Ansatz design can leverage data symmetries**[1]

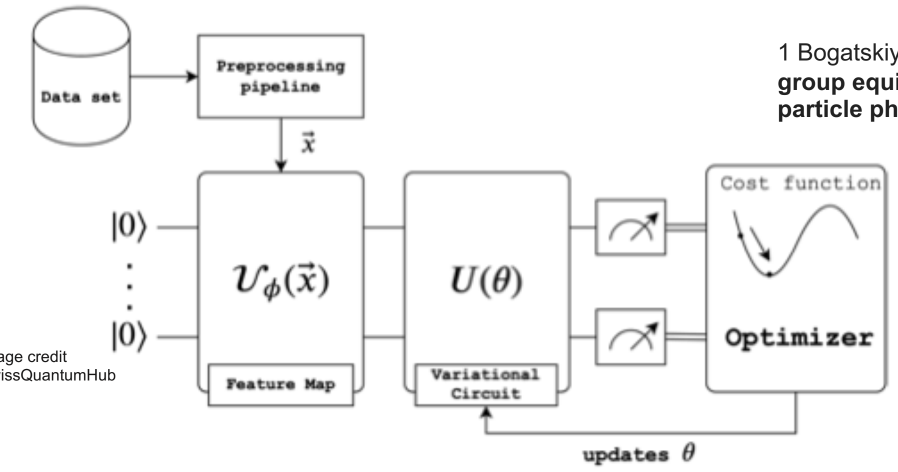



Image credit SwissQuantumHub

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

## *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).

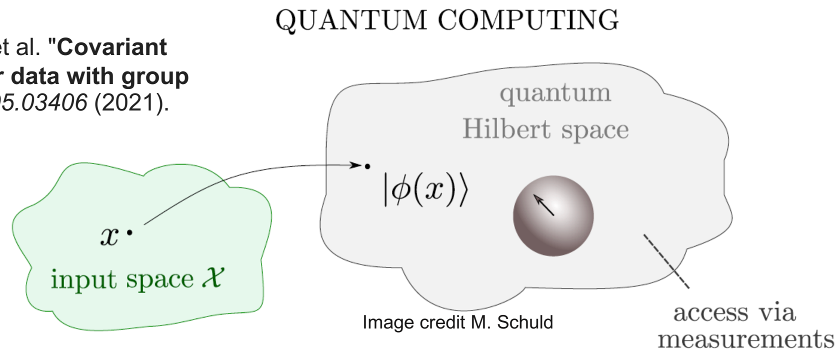



Image credit M. Schuld

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).

# Example QML applications

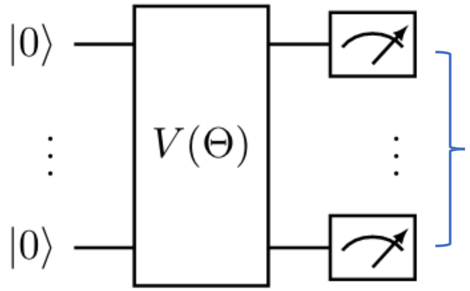
Event generation and Simulation

# Quantum Generative Models

Delgado and Hamilton, arXiv:2203.03578 (2022)
Zoufal, et al., *npj Quantum Inf* **5**, 103 (2019)
Leadbeater et al., *Entropy* **2021**, *23*, 1281.
Amin, et al. *Physical Review X* 8.2 (2018): 021050.

## QCBM

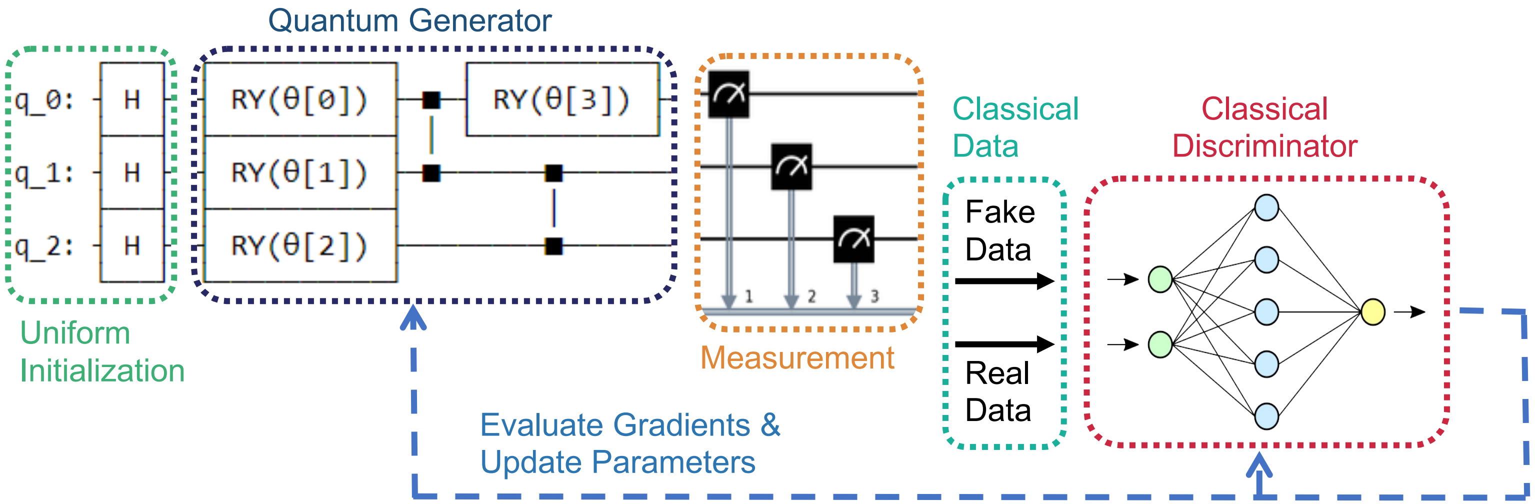
**Sample** variational pure state $|\psi(\theta)\rangle$ by projective measurement through **Born rule**: $\mathbf{p_\theta}(\mathbf{x}) = |\langle \mathbf{x}|\boldsymbol{\psi}(\boldsymbol{\theta})\rangle|^2$ .




n dimensional binary strings map to $2^n$ bins of the discretized dataset.

## QGAN

Multiple implementations, mostly classical-quantum hybrid




## QBM

Network of stochastic binary units with a quadratic energy function that follows the Boltzman distribution (Ising Hamiltonian)

$$H = -\sum_a b_a \sigma_a^z - \sum_{a,b} w_{ab} \sigma_a^z \sigma_b^z$$

## Typical metrics:

$$D_{\mathrm{KL}}(P\|Q) = \sum_i P(i) \log\left(\frac{P(i)}{Q(i)}\right)$$

$$\mathrm{MMD}(\mathbb{P}_r, \mathbb{P}_g) = \left(\mathbb{E}_{\substack{\mathbf{x}_r, \mathbf{x}_r' \sim \mathbb{P}_r, \\ \mathbf{x}_g, \mathbf{x}_g' \sim \mathbb{P}_g}} \left[k(\mathbf{x}_r, \mathbf{x}_r') - 2k(\mathbf{x}_r, \mathbf{x}_g) + k(\mathbf{x}_g, \mathbf{x}_g')\right]\right)^{\frac{1}{2}}$$

# qGAN for event generation

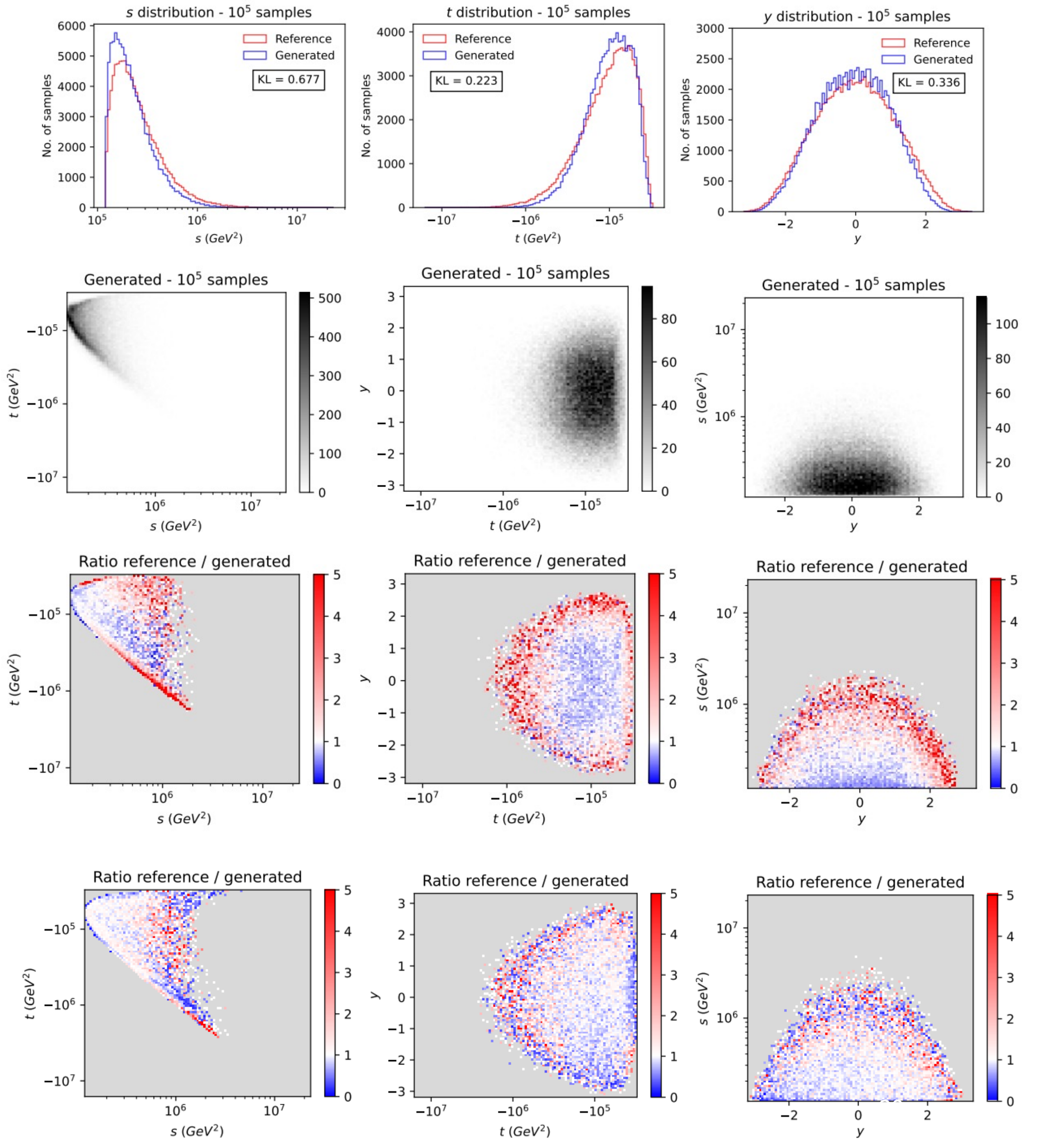
Generate Mandelstam (*s,t*) + *y* variables for **t-tbar production**

Introduce a **style-based** approach

| | $pp \to t\bar{t}$ **LHC events** |
|---|---|
| Qubits | 3 |
| $D_{\text{latent}}$ | 5 |
| Layers | 2 |
| Epochs | $3 \times 10^4$ |
| Training set | $10^4$ |
| Batch size | 128 |
| Parameters | 62 |
| $U_{\text{ent}}$ | 2 sequential $CR_y$ gates |

IBM Q Santiago

Quantum simulator




Bravo-Prieto et al. "**Style-based quantum generative adversarial networks for Monte Carlo events**." Quantum 6, 777 (2022) , *arXiv preprint arXiv:2110.06933* (2021).

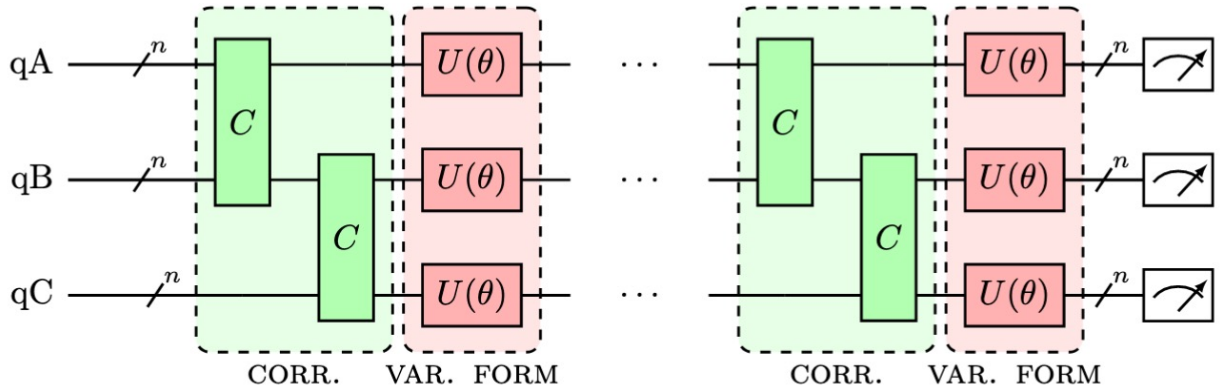
Kiss, Grossi, *et al.*, *Phys. Rev. A* **106**, 022612 (2022)

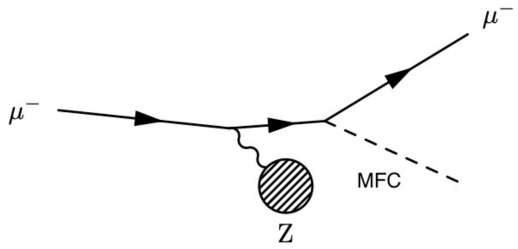
# QCBM for event generation

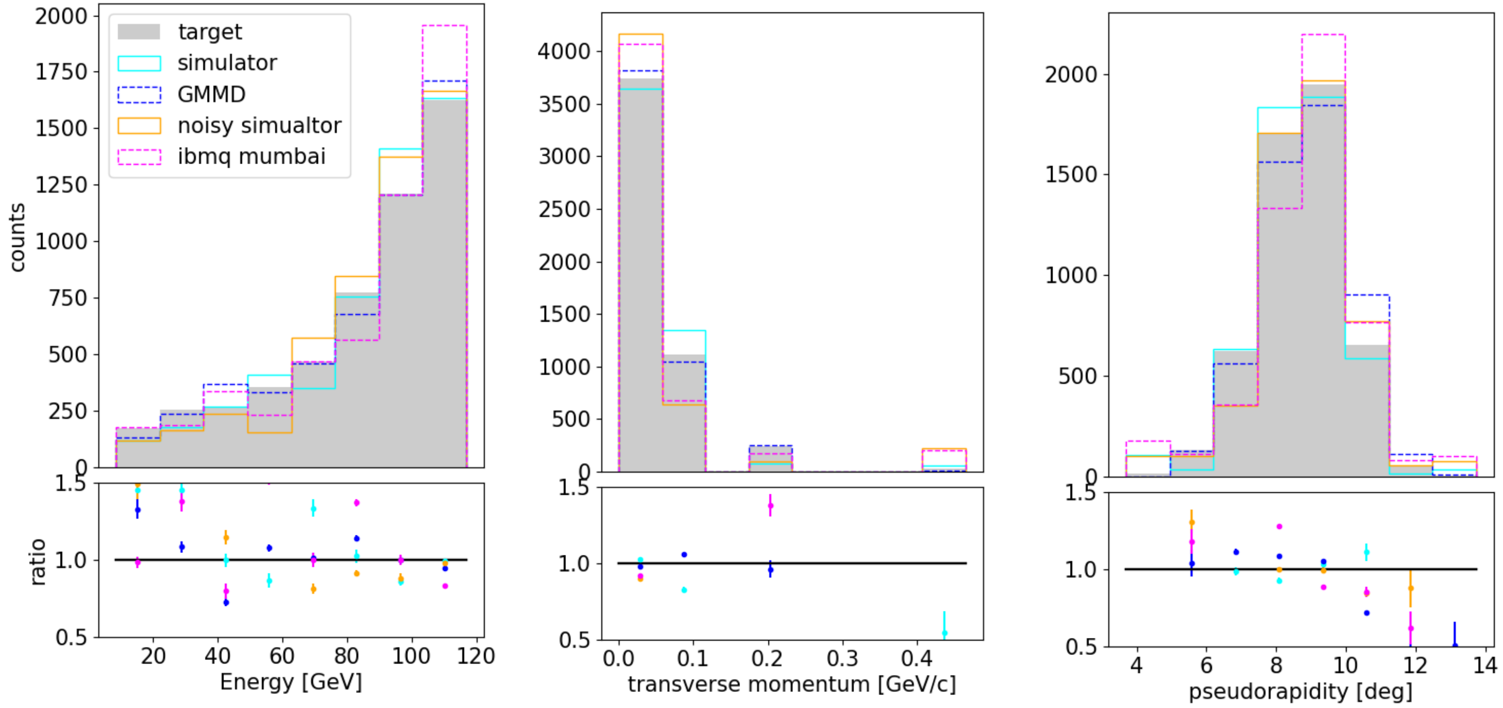
**Muon Force Carriers,** in muon fixed-target experiments (FASER) or muon interactions in calorimeters (ATLAS)[1].

**Generate multivariate distribution (E, $p_t$, η)**

**Maximum Mean Discrepancy for training**



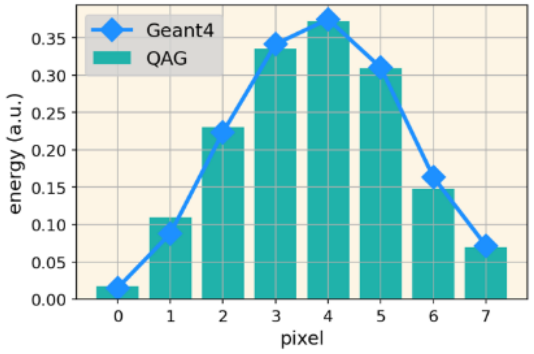



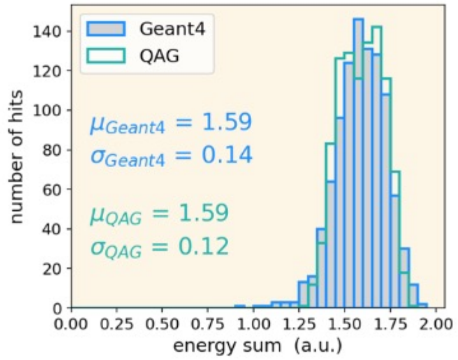

1 Galon, I, Kajamovitz, E et al. "**Searching for muonic forces with the ATLAS detector**". In: Phys. Rev. D 101, 011701 (2020)

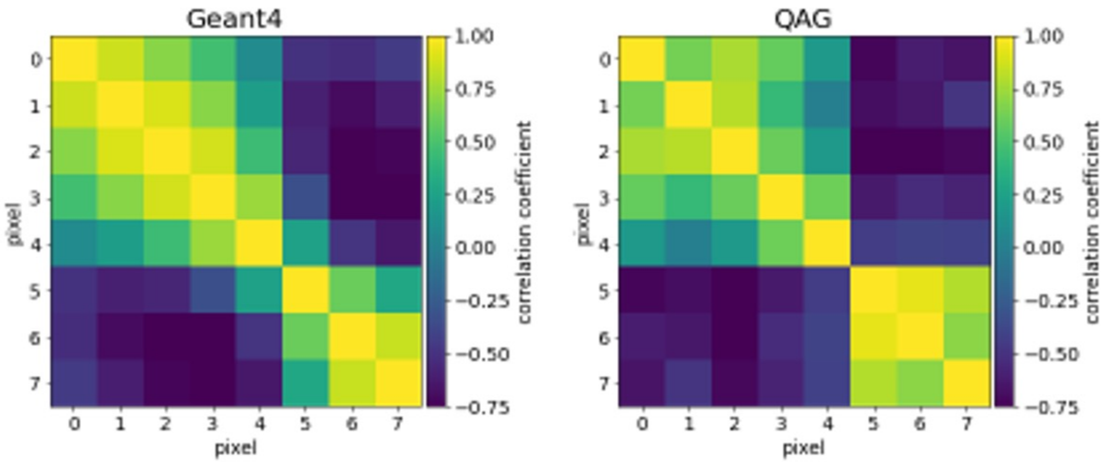
# The case of detector simulation

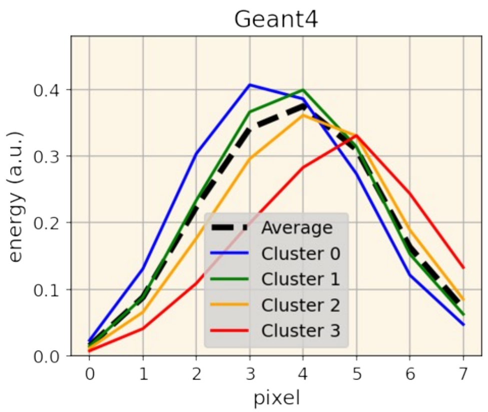
**QML can realistically simulate the energy deposited by particles in a detector**

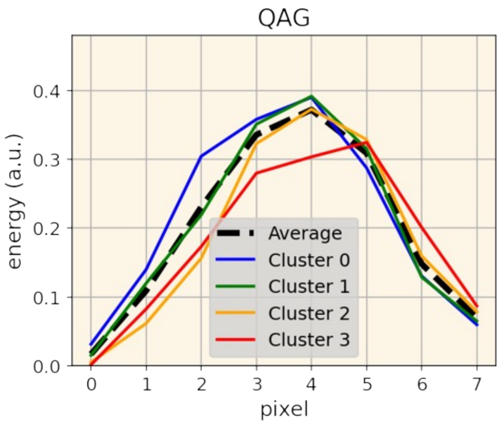
QNN (MMD loss)



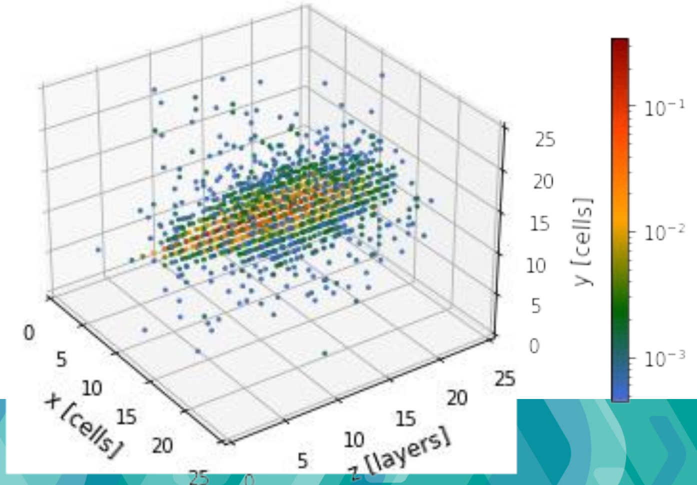




**Scale** is the main problem

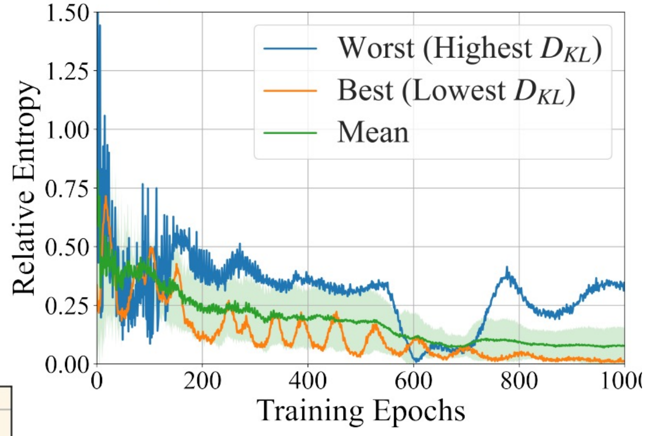
Entirely change the formulation?

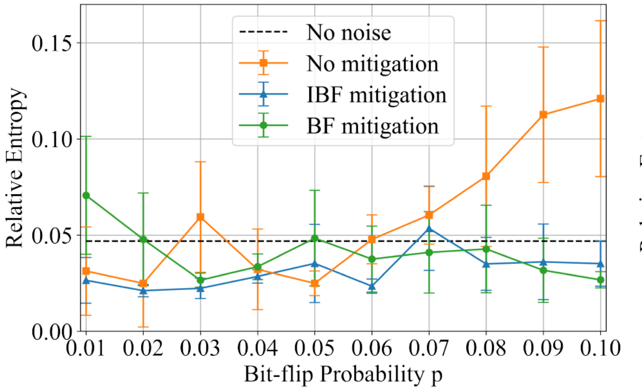


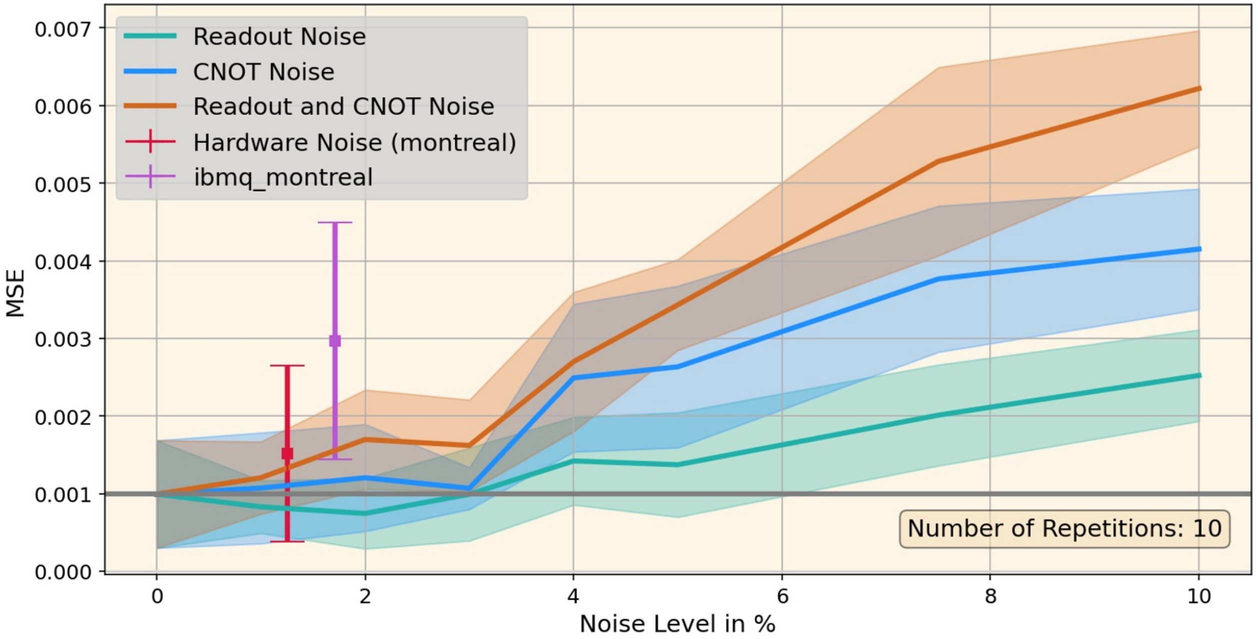
Borras, Kerstin, et al. "**Impact of quantum noise on the training of quantum Generative Adversarial Networks**." ACAT2021, *arXiv preprint arXiv:2203.01007* (2022).

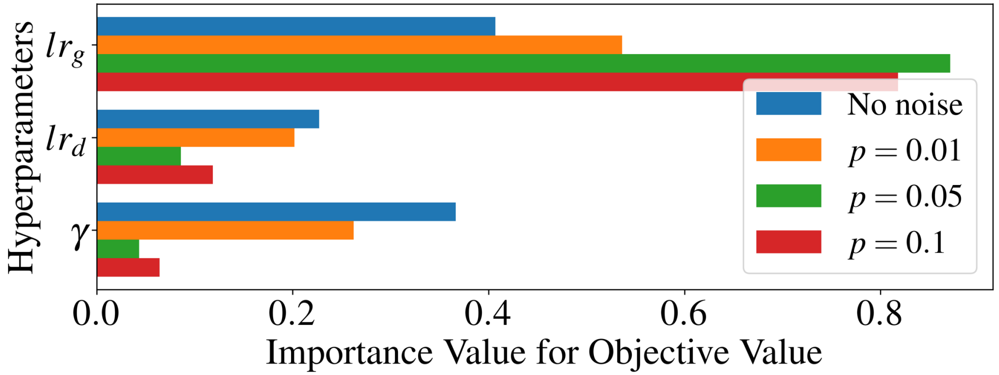
# Robustness against noise

QML training process seems **robust against noise** (error mitigation is needed in extreme cases)



# Example QML applications

Data Processing

# Quantum Data Classification

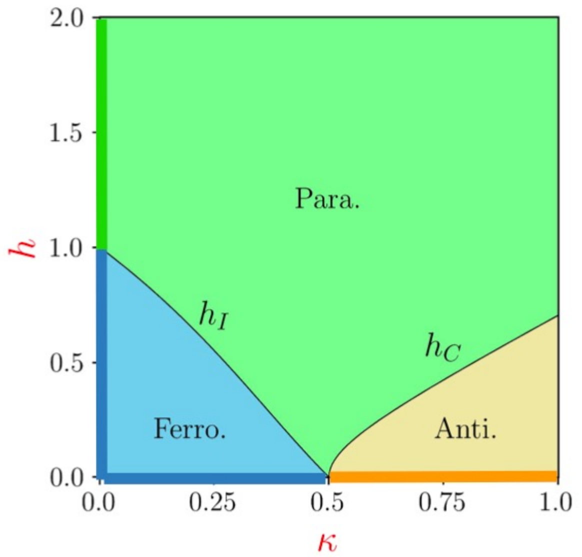
# QML for quantum data: drawing phase diagrams

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)

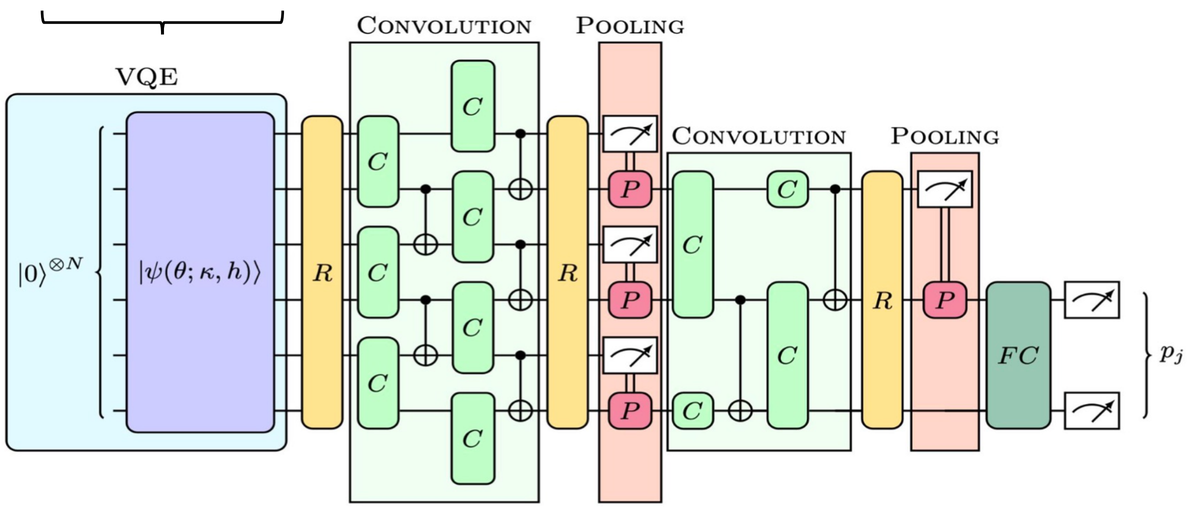
Integrable for $\kappa = 0$ or $h = 0$.




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]

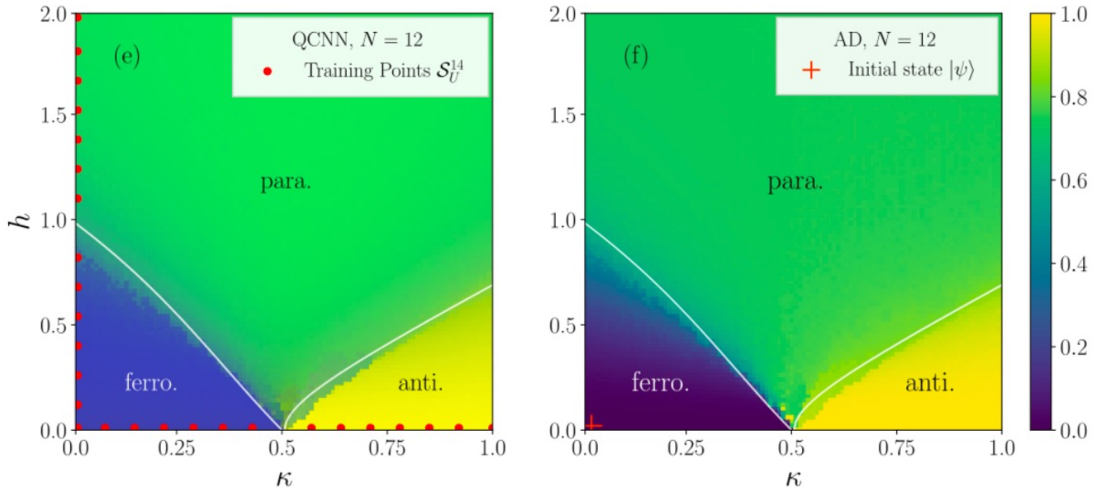
# Setting the stage

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

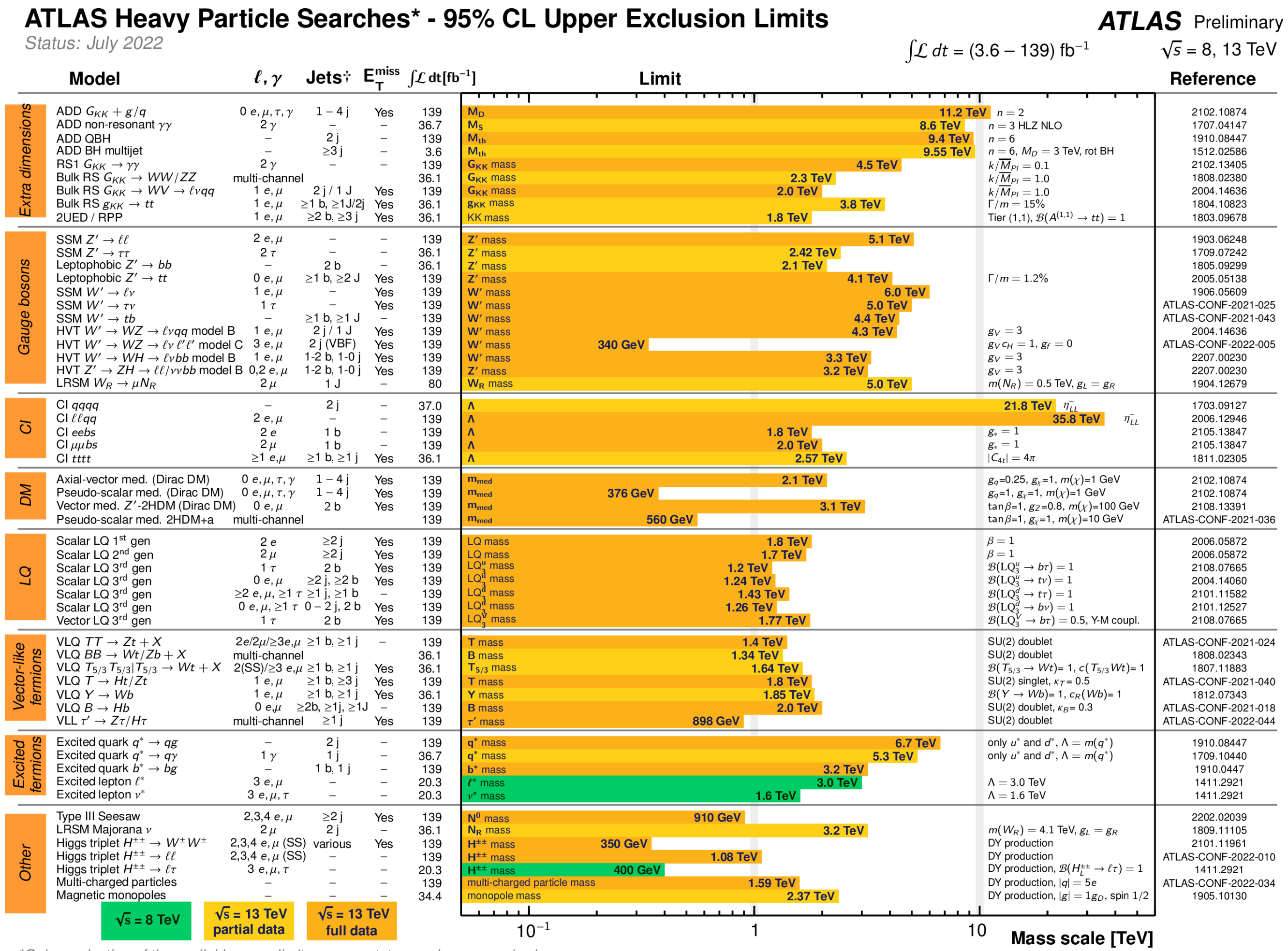
*Monaco, at al. arXiv: 2208.08748 (2022), accepted PRB*

[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)

# Anomaly Detection

# New Physics at the LHC

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




*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).

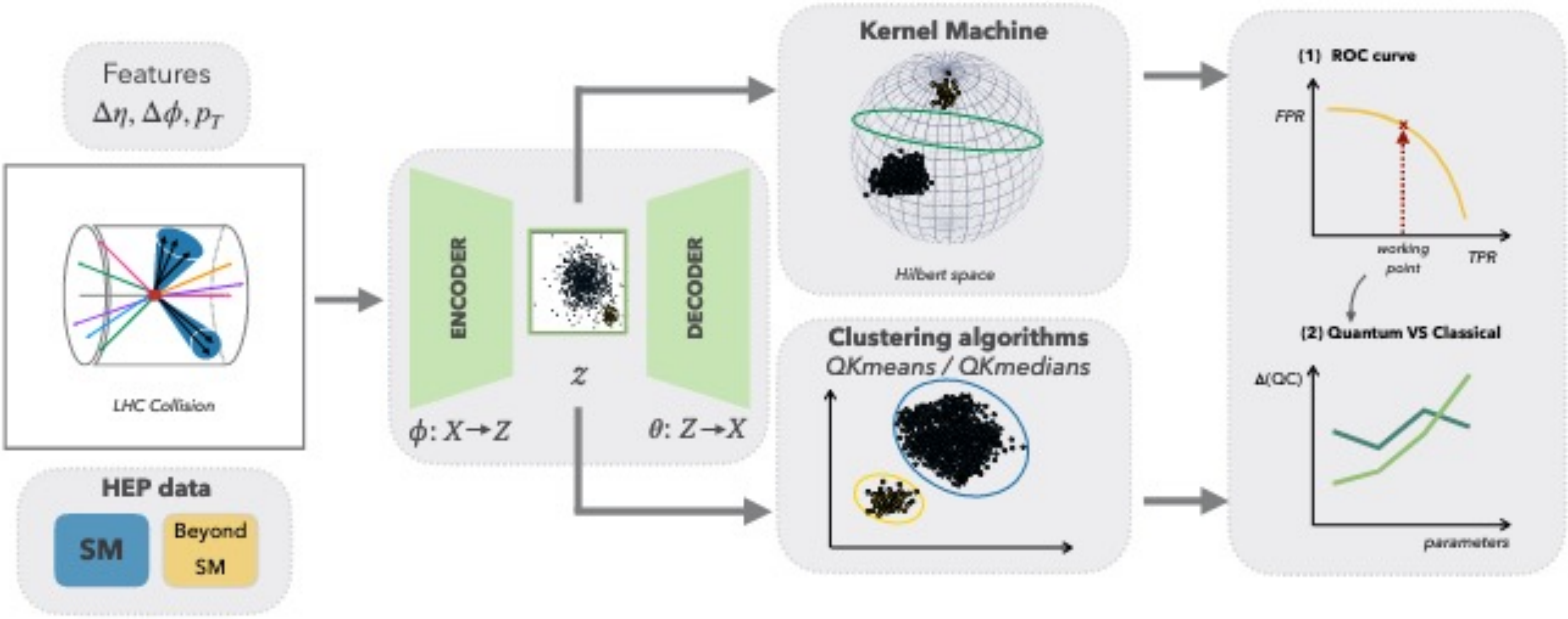
**How to insure we do not miss potential discoveries?**

**Anomaly detection** can point to **new physics**

**Model-agnostic!**

V. Belis *et al.*, **Quantum anomaly detection in the latent space of proton collision events at the LHC**, arxiv:2301.10780

# A typical hybrid QML workflow




«Normal» training data

Data compression

Quantum algorithm

Output

# Results

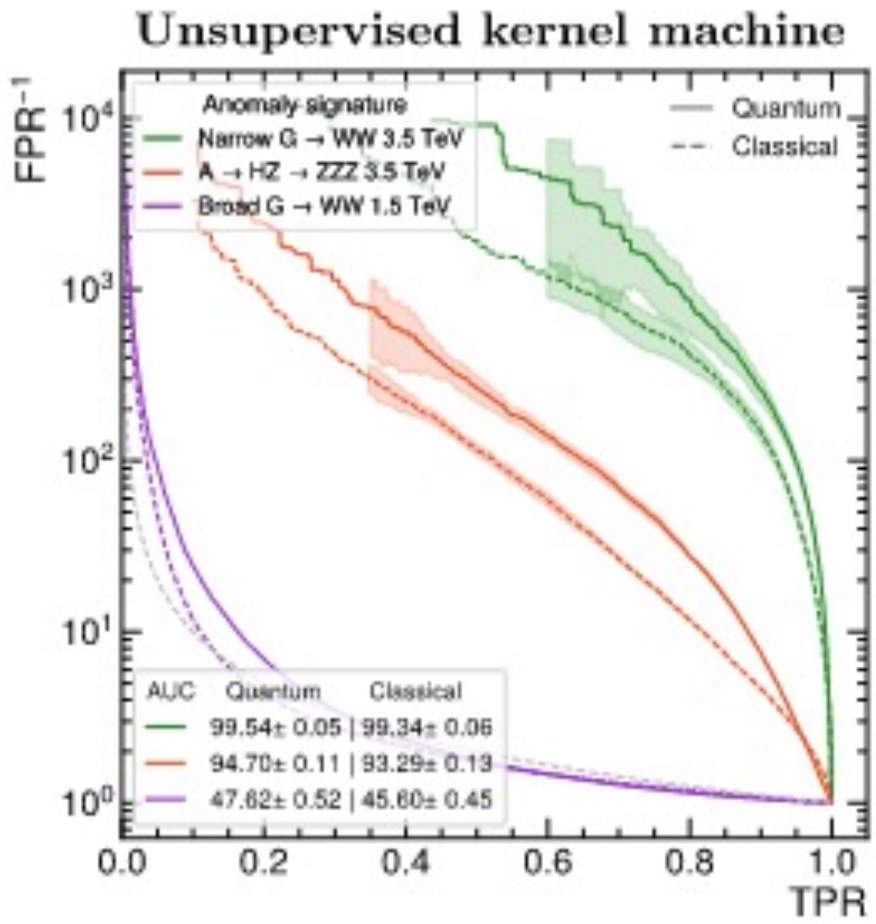
**Comparison to best-performing classical algorithm with similar** complexity trained and tested on the same data

- RBF –based SVM

AUC shows **marginal advantage** for quantum algorithm

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

**Quantum kernel machine works best for more complex physics**



# Characterizing the advantage

Higher is better

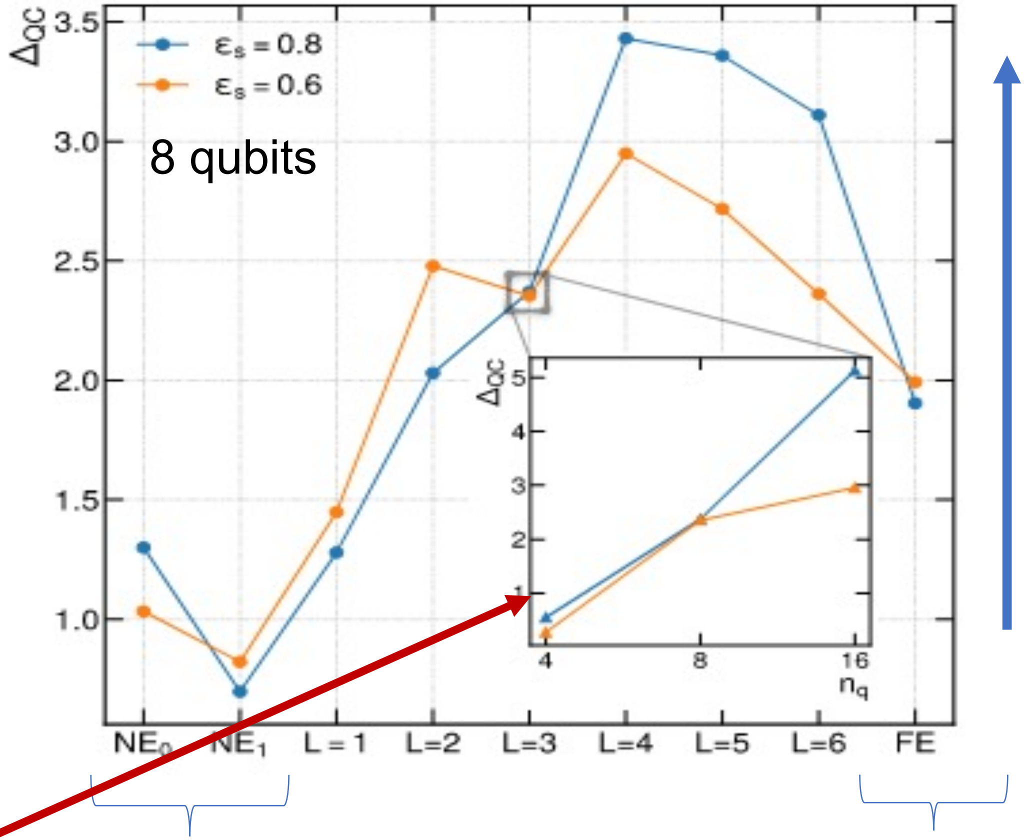
Given **signal and background efficiencies**, $\varepsilon_s$ and $\varepsilon_b$ respectively:

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

**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



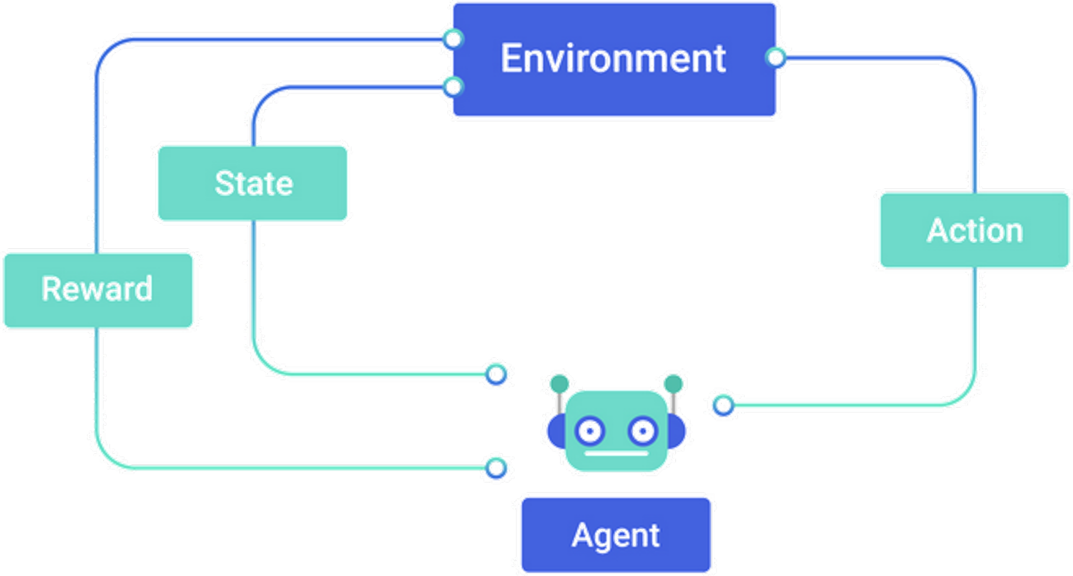

All-to-all entanglement

$NE_0$: encoding layer only
$NE_1$: no entanglement

# Reinforcement learning

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

# Quantum Reinforcement Learning




**Agent interacts with environment**

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

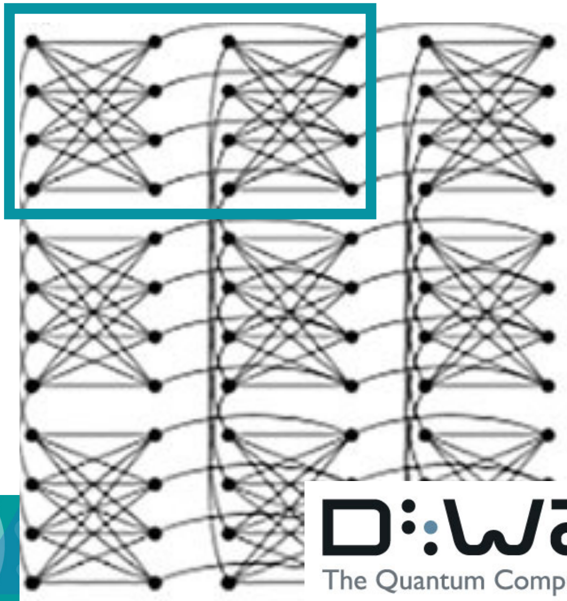
**Expected reward** is estimated by ***value function*** $Q(s,a)$

- **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**

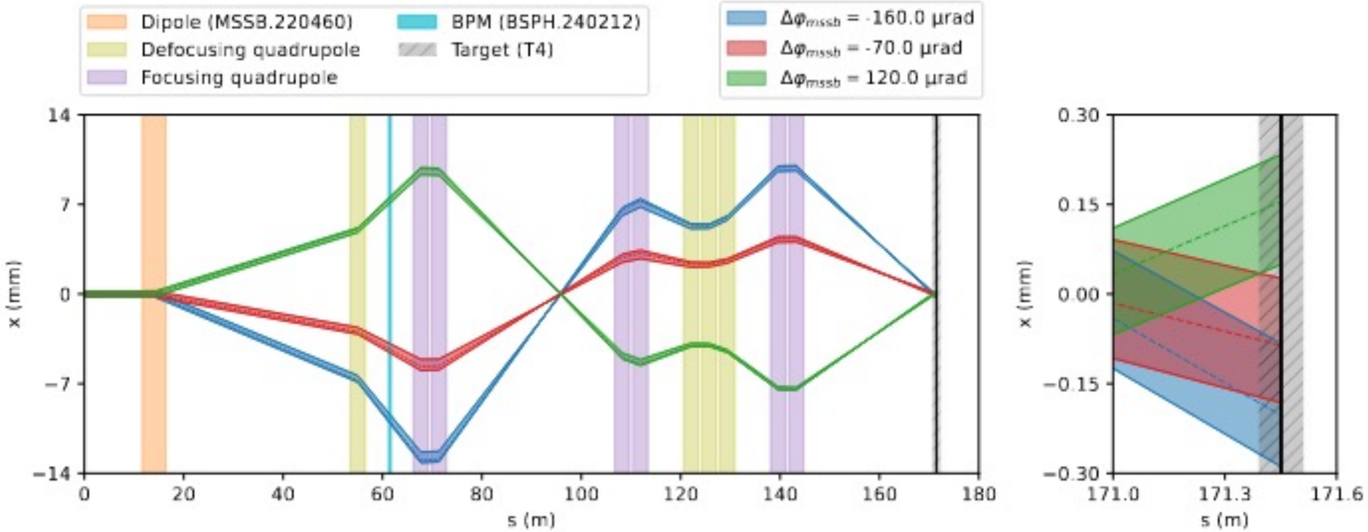


D:WAVE The Quantum Computing Company™

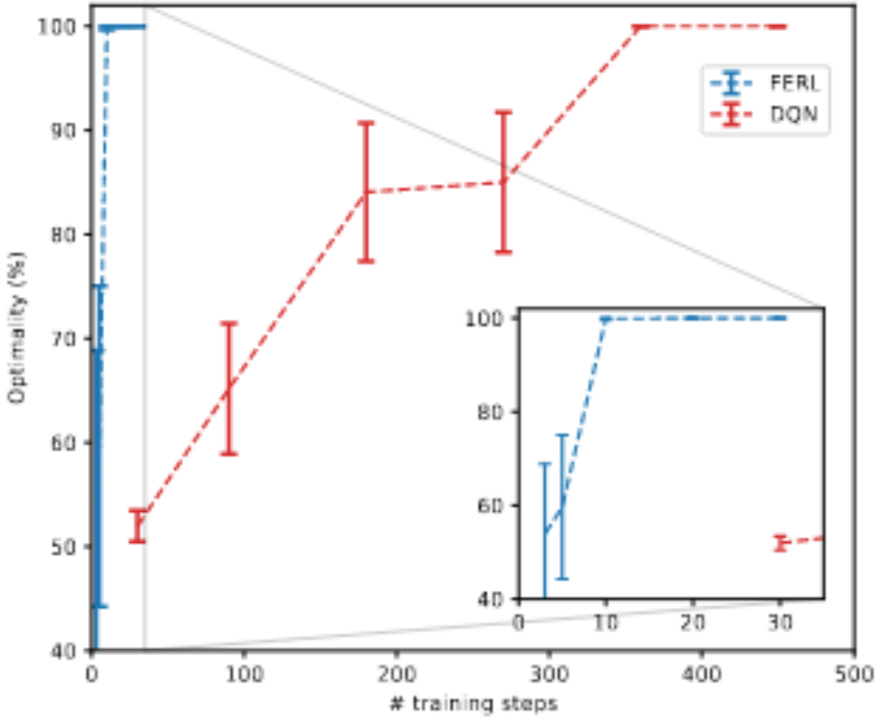
Michael Schenk et al., **Hybrid actor-critic algorithm for quantum reinforcement learning at CERN beam lines**, e-Print: 2209.11044 [quant-ph]

# 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

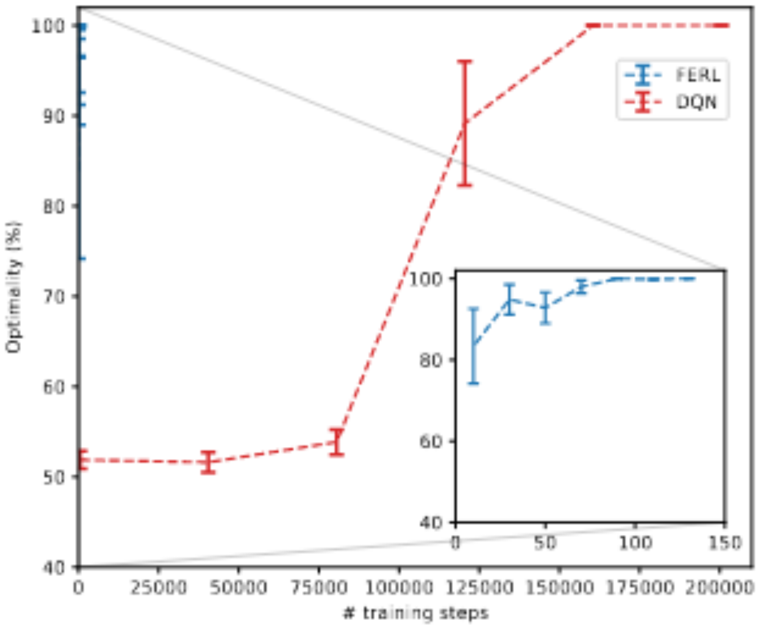



- **Quantum RL massively outperforms** classical Q-learning *(8±2 vs. 320±40 steps with e. r.)*

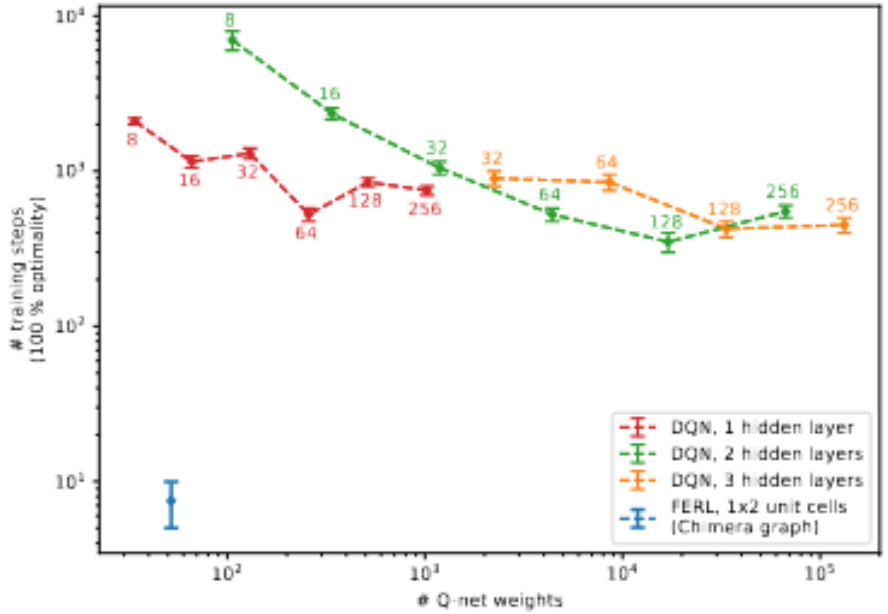


# Convergence and representational power

Without **experience replay**




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




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

# Outlook and Questions

**Quantum Machine Learning is a broad-lively research field**

- Some **preliminary hints** of advantage
- Need more **robust theoretical studies** to interpret experimental results and **build efficient circuits** (physics-based..)
- Need to establish **«fair comparison»** to classical tools on realistic use cases
- **Studying the behviour of trainable systems in the NISQ regime is useful**

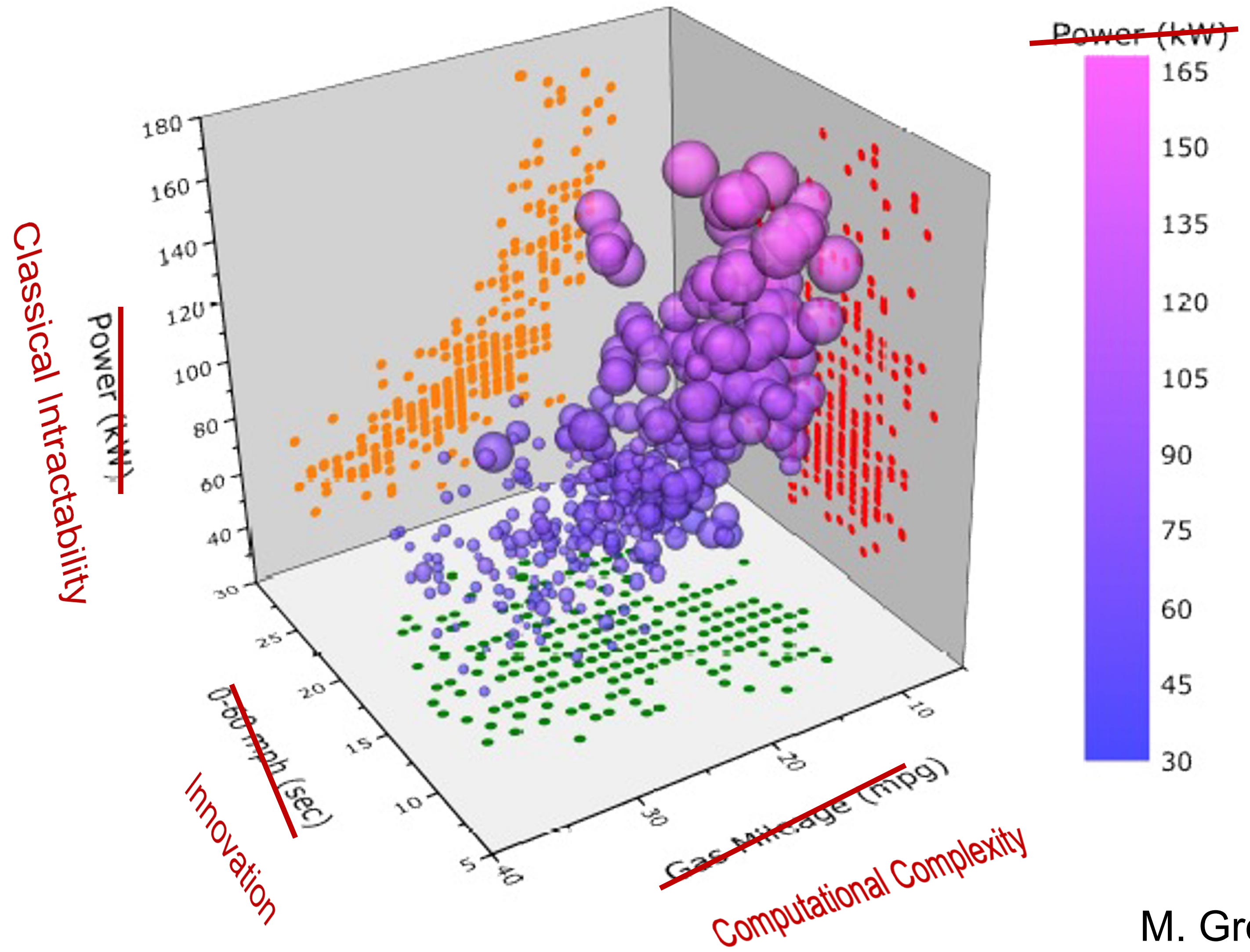
**Can we reduce the impact of data reduction techniques?**

**Can we find the right balance of trainability vs generalization?**

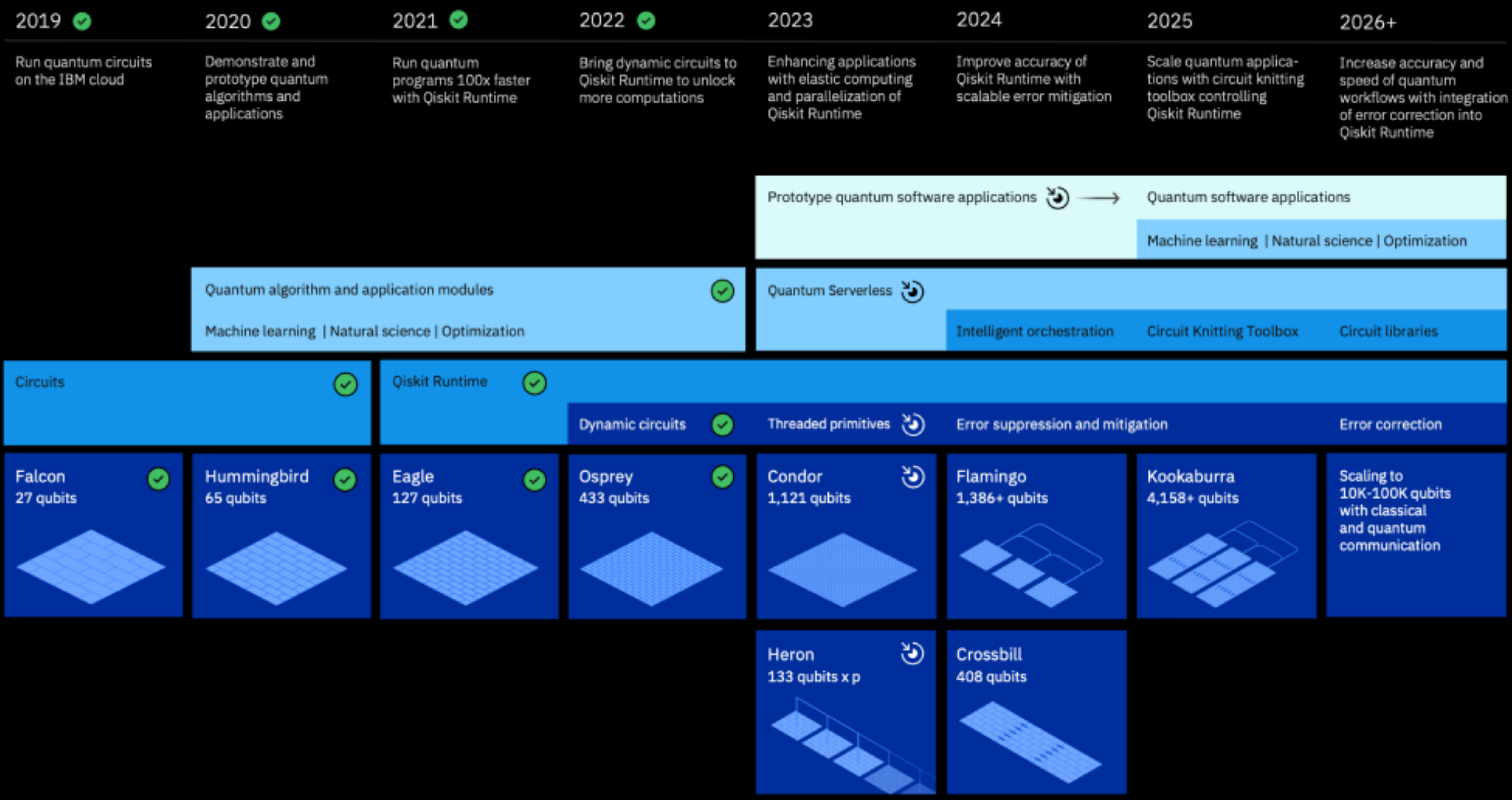
**Can we build a «continuous path» toward fault tolerance?**

# Exclusion Region for QML in HEP?

CERN is formulating a **longer term research plan** dedicated to **understanding impact for High Energy Physics**




M. Grossi, CERN


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
Prototype quantum software applications
Quantum software applications
Machine learning | Natural science | Optimization
Quantum algorithm and application modules
Machine learning | Natural science | Optimization
Quantum Serverless
Intelligent orchestration
Circuit Knitting Toolbox
Circuit libraries
Circuits
Qiskit Runtime
Dynamic circuits
Threaded primitives
Error suppression and mitigation
Error correction
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

# Thank you!

November 20th-24th, 2023
@CERN

Quantum Techniques in Machine Learning

Sofia.Vallecorsa@cern.ch

# Model Convergence and Barren Plateau

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

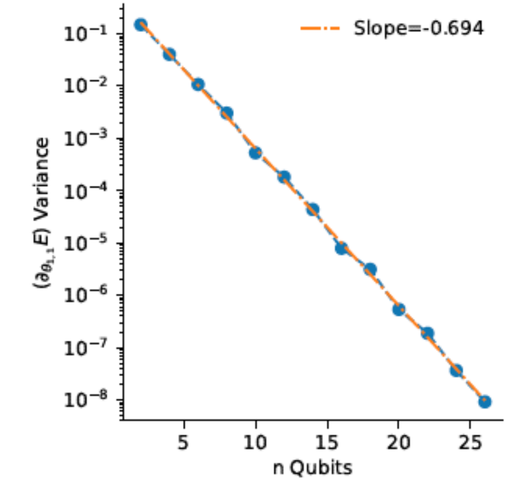
Classical gradients **vanish exponentially** with the number of layers (J. McClean *et al.*, arXiv:1803.11173)

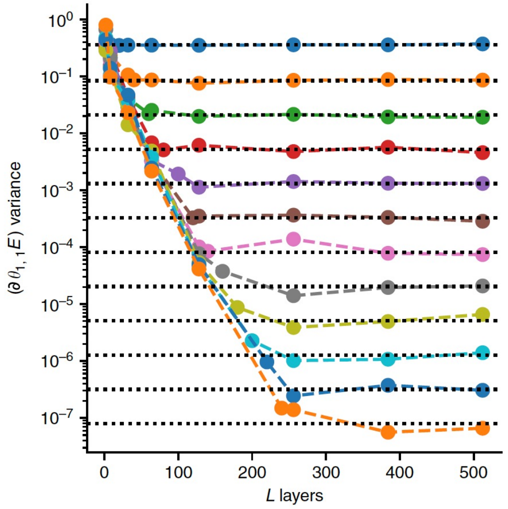
- Convergence still possible if gradients consistent between batches.

**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))

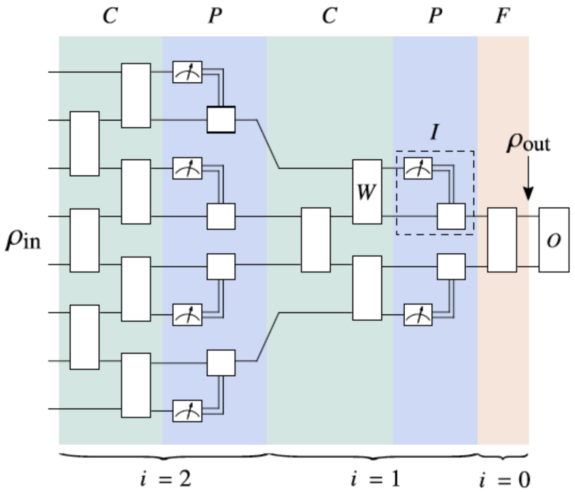
J. McClean *et al.*, arXiv:1803.11173




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


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



# Kernel trainability and kernel concentration

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

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

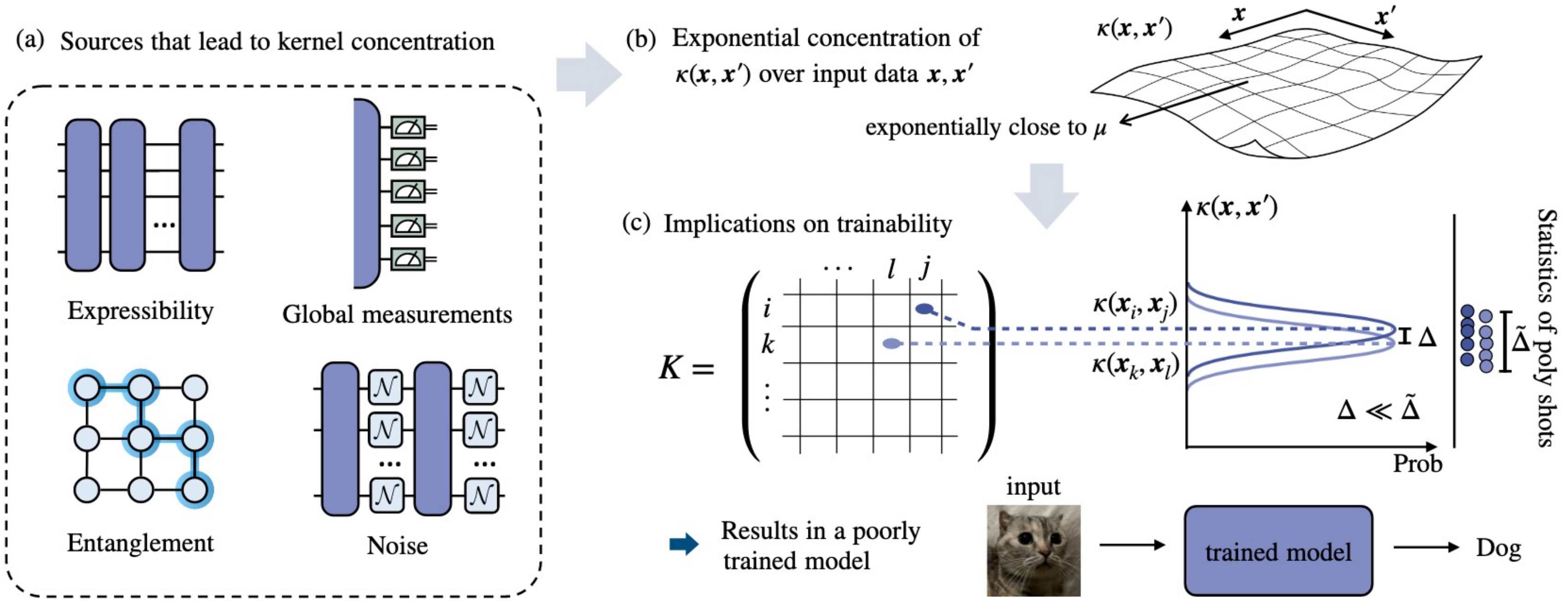
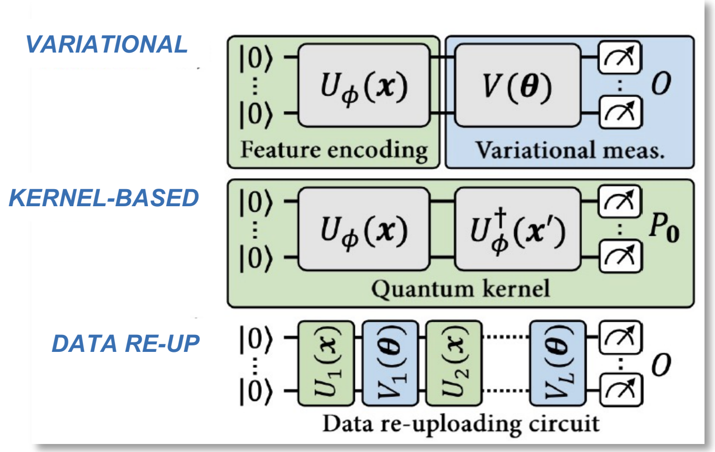



Figure 1. **Kernel concentration and its implications on trainability:** The exponential concentration (in the number of qubits $n$) of quantum kernels $\kappa(\boldsymbol{x}, \boldsymbol{x}')$, over all possible input data pairs $\boldsymbol{x}, \boldsymbol{x}'$, can be seen to stem from the difficulty 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?

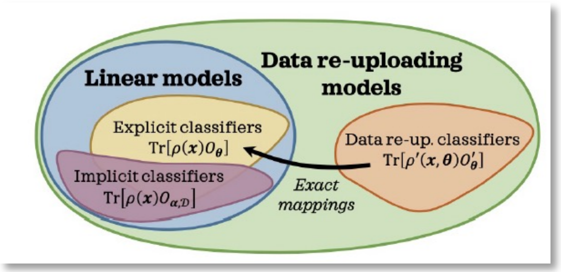


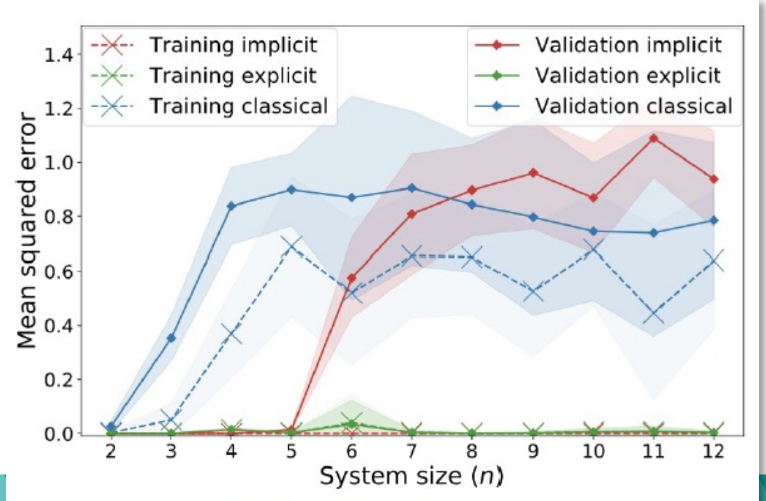

**Characterize models behaviour, similarities among them** and **link to data properties.**

**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).

# Model definition

## *Variational algorithms - EXPLICIT*

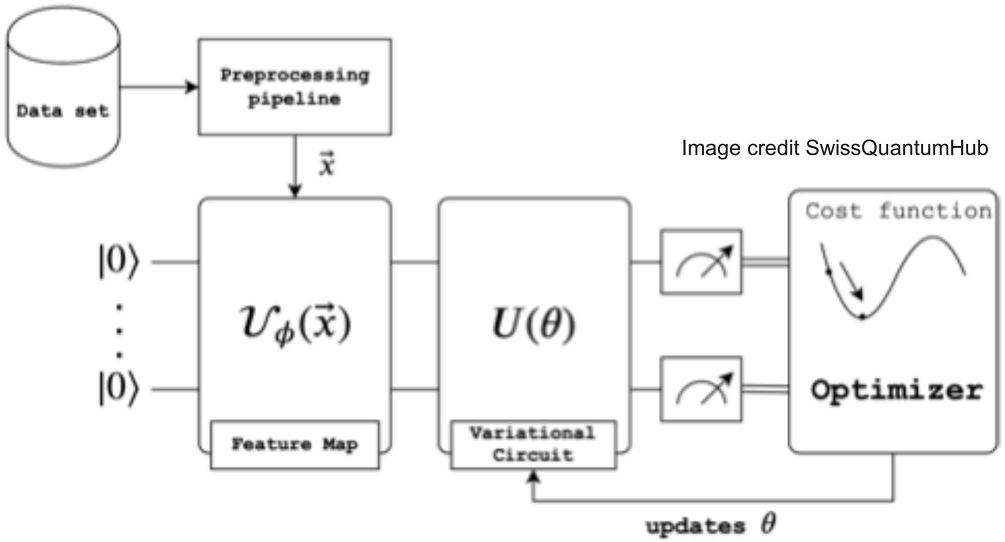



Image credit SwissQuantumHub

Define a **parametric quantum circuit** with trainable parameters $\vartheta$

$$U(x, \vartheta)$$

Given an observable $O$, build a model

$$y(x, \vartheta) = \langle 0 | U^\dagger(x, \vartheta) O U(x, \vartheta) | 0 \rangle$$

- **Trained using gradient-free** or **gradient-based** optimization in a classical loop
- **Data Embedding** $\mathcal{V}_\phi(x)$ can be **learned**
- Improve performance by designing architectures to **leverage data symmetries**[1]
- Aim at quantum circuits that are **hard to simulate classically**

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

# Unsupervised Anomaly Detection

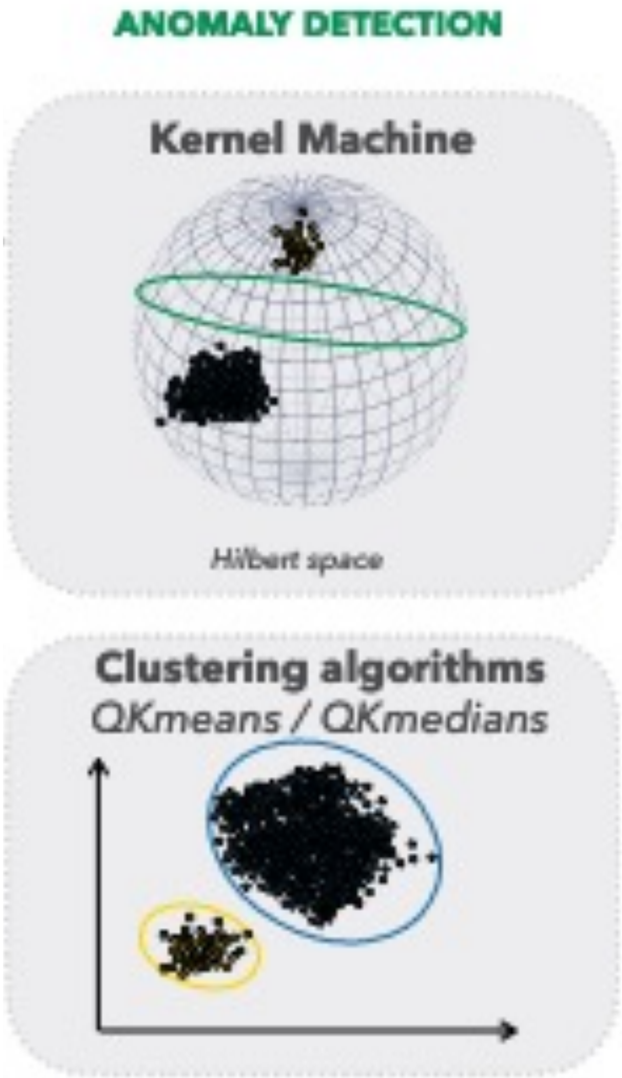
Quantum and classical AD algorithms are trained on QCD multijet events in the **signal region $|\Delta\eta_{jj}| \leq 1.4$.**

Tested on three **representative BSM:**

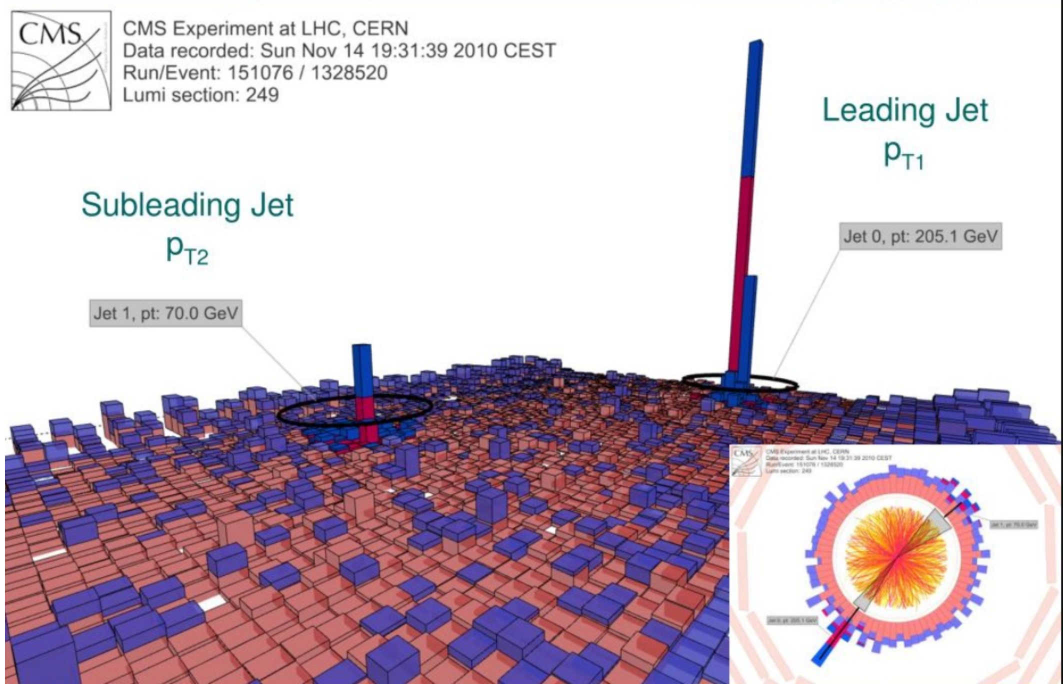
- Narrow Randall-Sundrum gravitons ($G \rightarrow WW$)
- Broad Randall-Sundrum gravitons ($G \rightarrow WW$)
- Scalar boson ($A \rightarrow HZ$, $H \rightarrow ZZ$)

Compare performance of **three unsupervised methods**:

- One-class SVM
- Q-Means clustering
- Q-Medians clustering

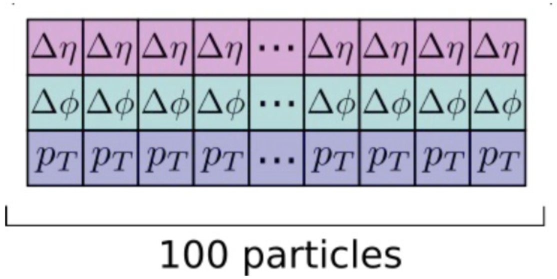


# Standard Model jet data




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

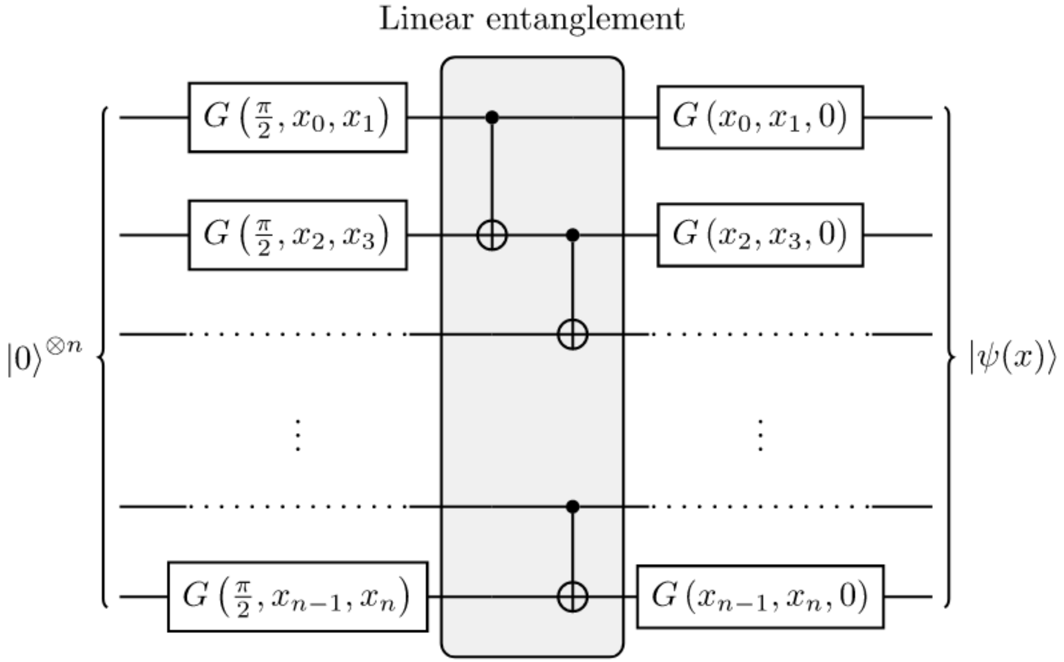
**Convolutional AutoEncoder** compresses particle jet learning the **internal structure**

- Trained on background events

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

# Unsupervised kernel machine

Data Embedding circuit




"Standard" kernel definition

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

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

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

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

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

$\nu \in (0,1)$ 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)

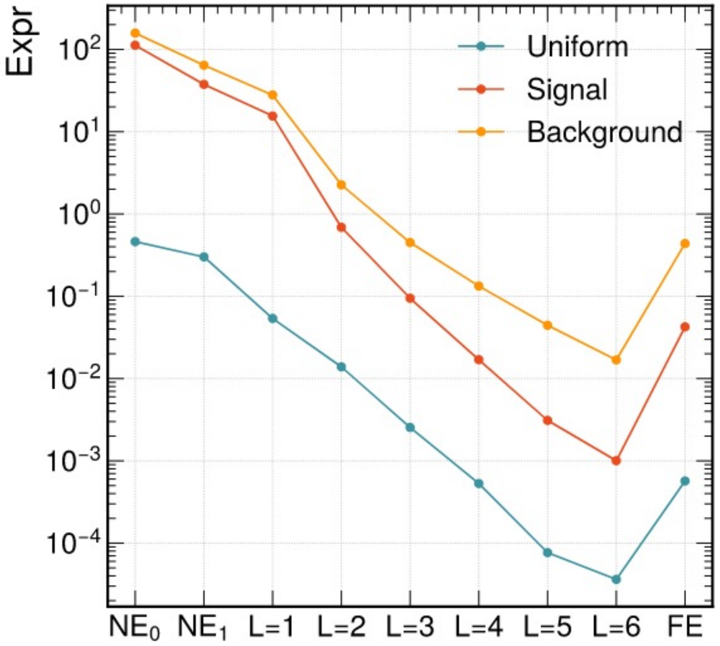
# Unsupervised vector machine kernels

Study **expressibility** of embedding circuit and **variance of the quantum kernel**

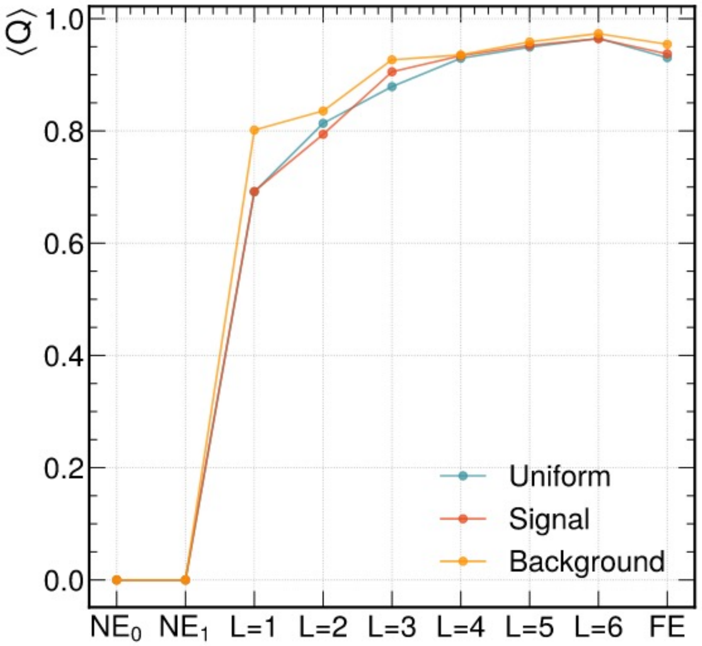
- Expressibility is roughly constant vs $N_q$
- Kernel variance does not decay exponentially

We observe **no exponential concentration** due to expressibility or global measurements.

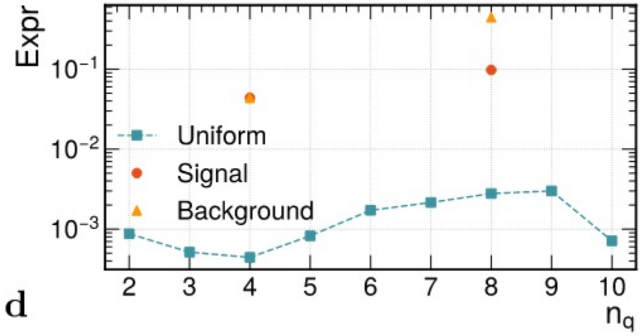
Expressibility as a function of different circuit architectures

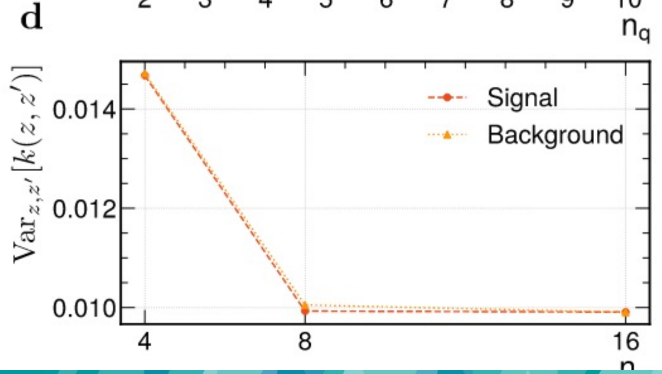



Entanglement capability




Expressibility as a function of the number of qubit



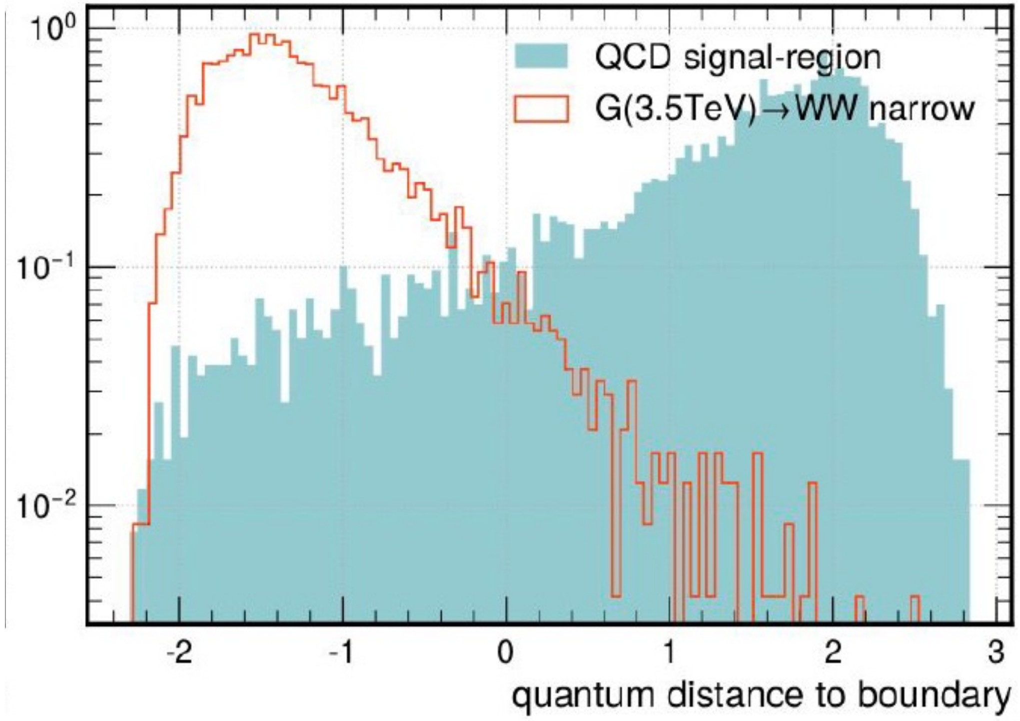



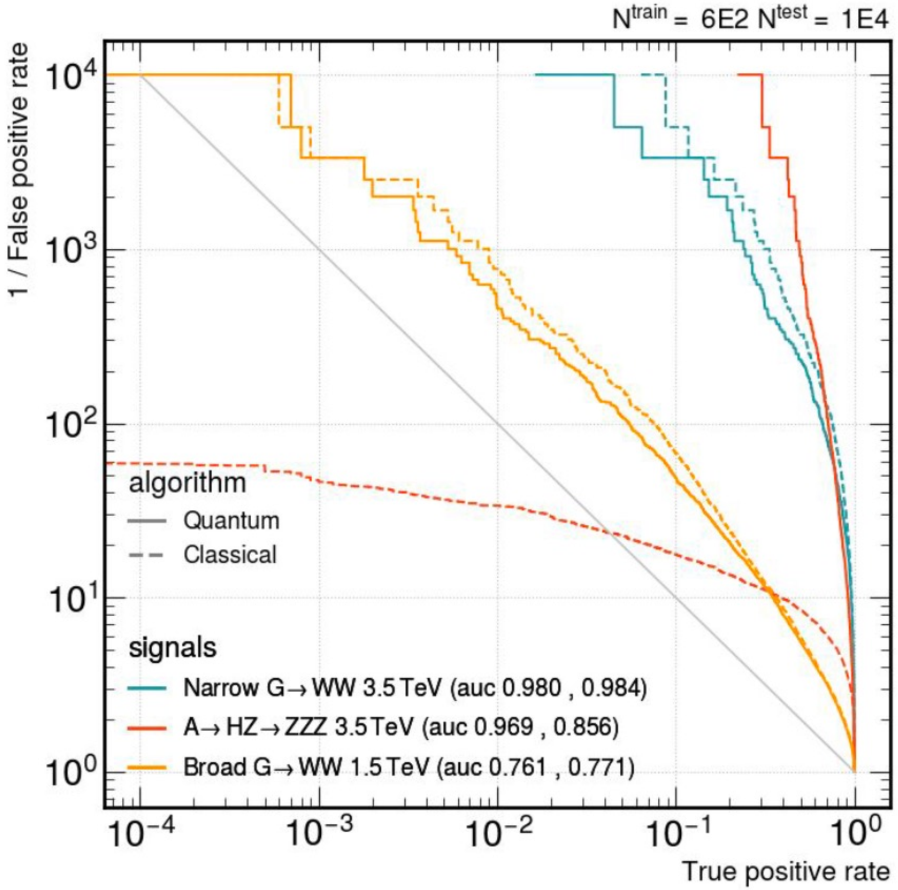

Kernels Variance

# Comparison to Supervised QSVM

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

Our initial study trained a supervised QSVM using the same setup.
Classical SVM outperform quantum

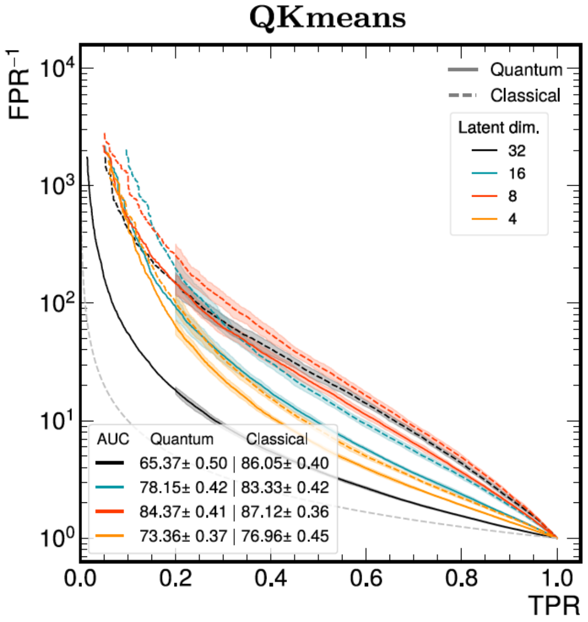


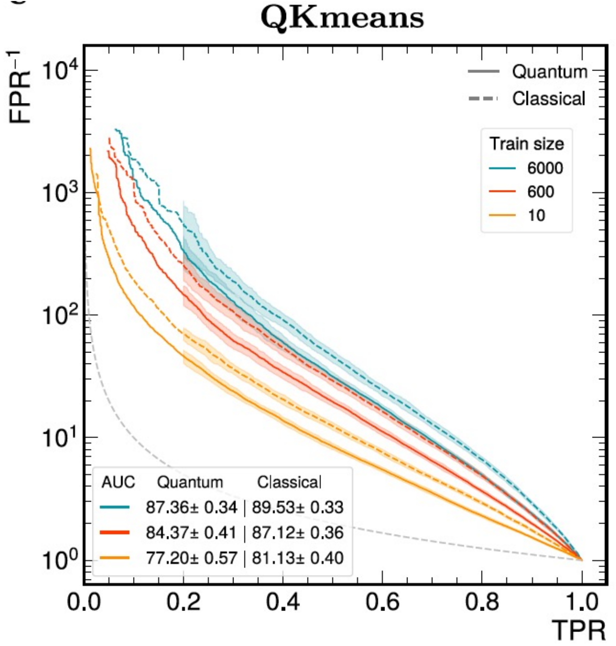



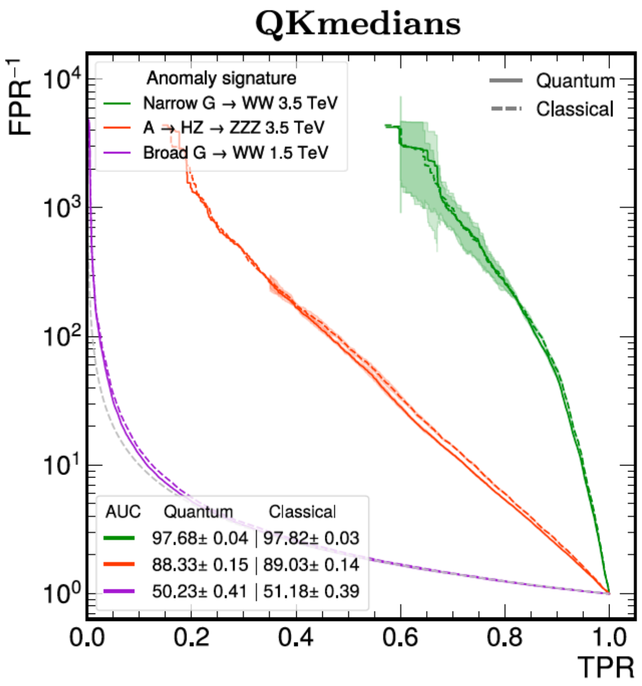
# Comparison to unsupervised clustering

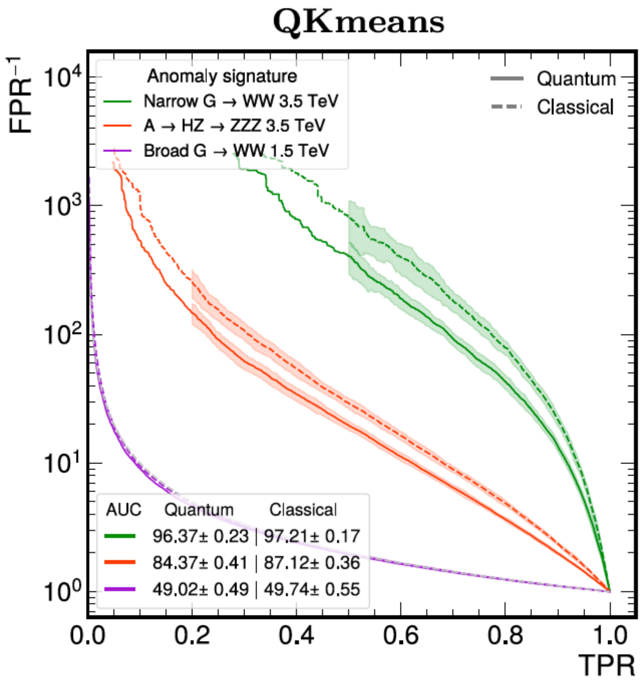
Quantum* clustering algorithms **do not outperform classical** counterpart

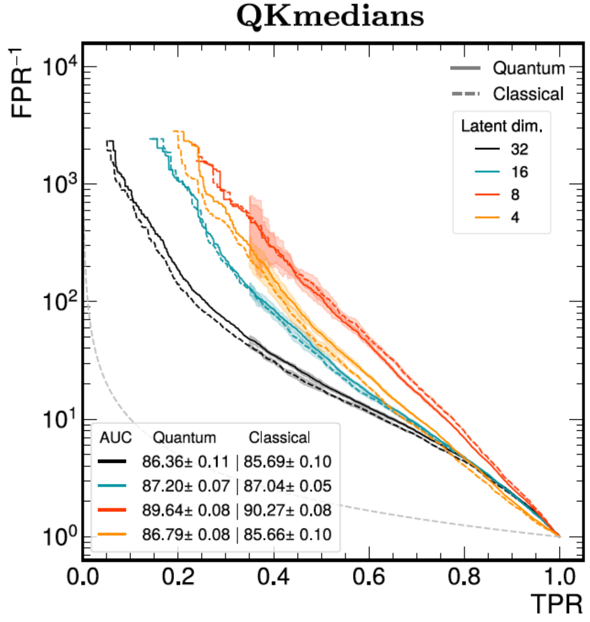
QMEANS performs worst

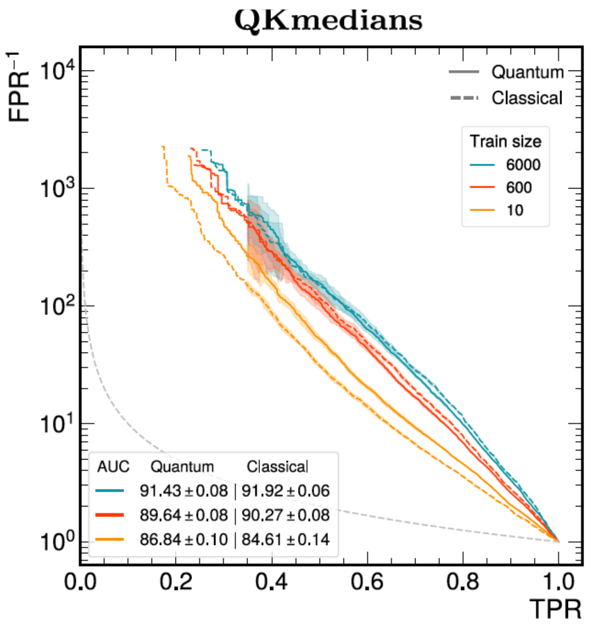


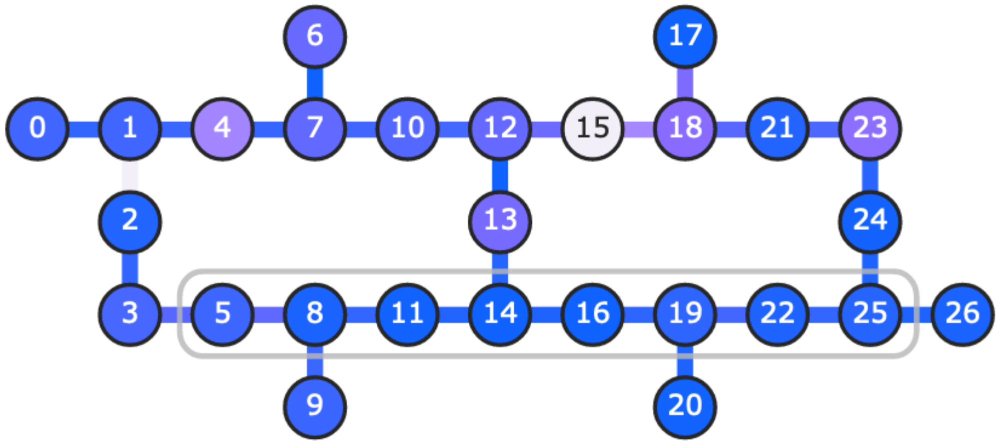



# Preliminary hardware runs




**Stable performance** on Hardware *ibmq_toronto* :

- Design circuit taking **qubits topology** into account
- Use **8 qubits** and native gates
- **Reduced training set size** (100 ) → increased statistical uncertainty
- Use **AUC** (less affected by statistics)
- Monitor **mean purity of states** to verify state coherence during computation
  - Fully mixed state yields a purity of 0.39 $10^{-2}$ ($1/2^n$)

| Kernel Machine Run | AUC | $\langle \mathrm{tr}\rho^2 \rangle$ |
|---|---|---|
| Hardware $L = 1$ | 0.844 | 0.271(6) |
| Ideal $L = 1$ | 0.999 | 1 |
| Hardware $L = 3$ | 0.997 | 0.15(2) |
| Ideal $L = 3$ | 1.0 | 1 |
| Classical | 0.998 | - |

# 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

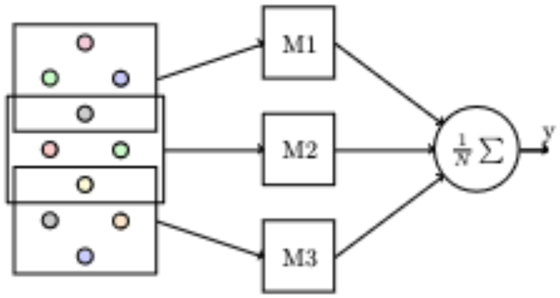
Incudini, M. , *et al*. "**Resource Saving via Ensemble Techniques for Quantum Neural Networks**." *arXiv:2303.11283* (2023).

# Ensembles of quantum neural networks

**NISQ regime affects QML performance. Can we build ensembles?**

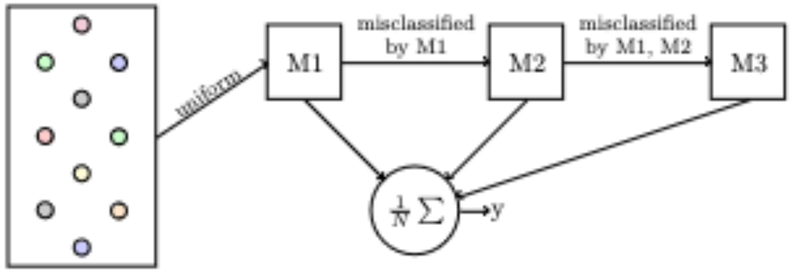
**Bagging**: best for **high variance**; reduces BPs by keeping the feature space limited

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




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

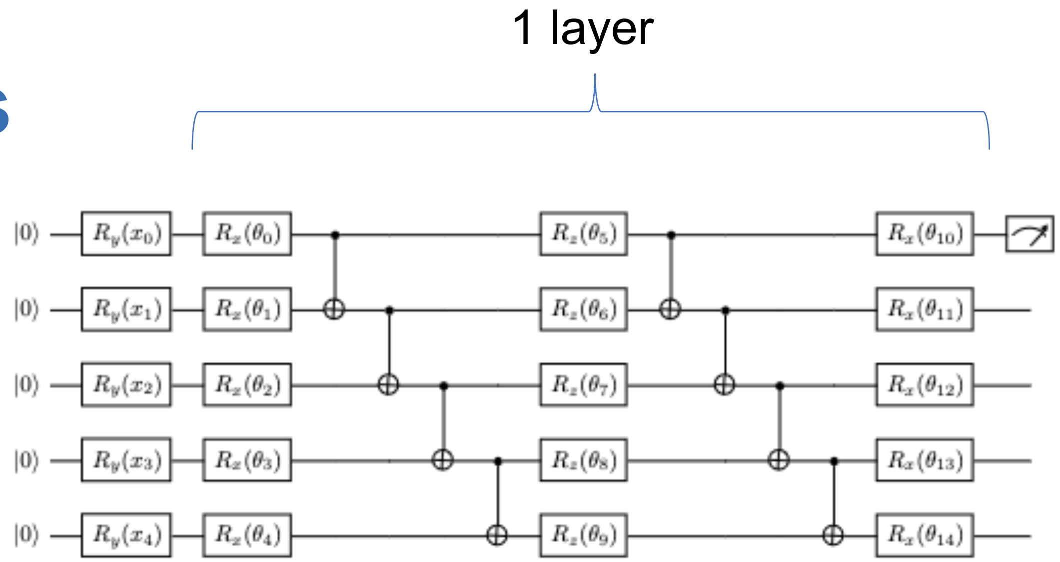
- AdaBoost, 10 repetitions




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



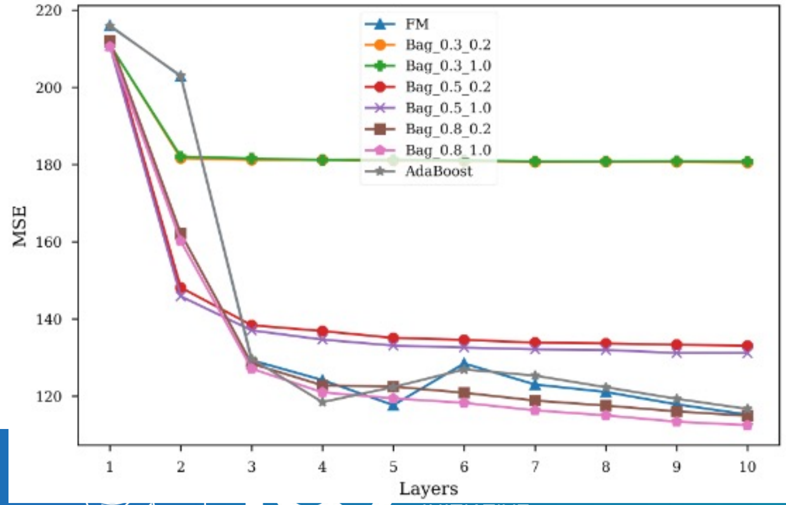

Choose **relatively simple QNN**:

- n qubits = n features
- Ry single rotation gates
- CNOT in linear entanglement
- Local observable ($\sigma_z$)

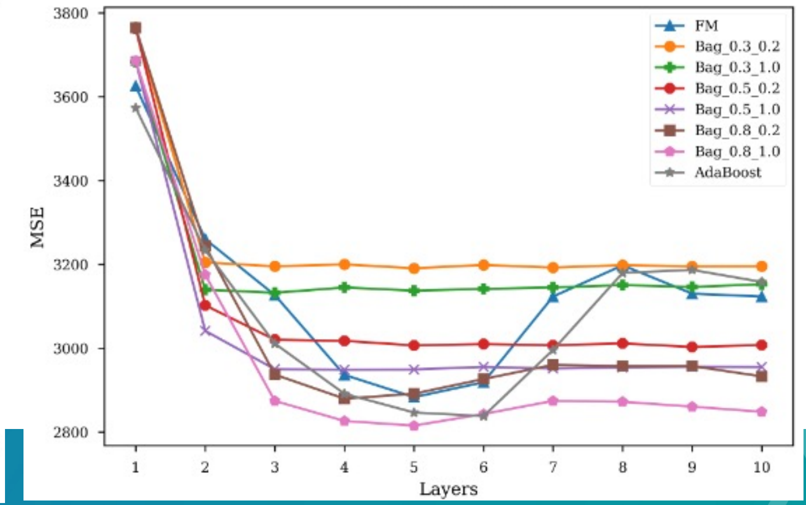
Measure the generalisation error on **test sample** (20 %)

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

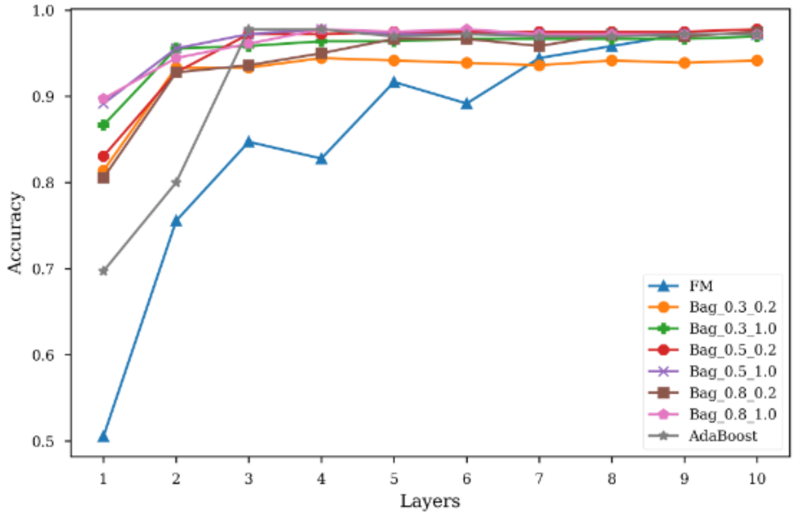
Concrete (MSE)




Diabetes (MSE)

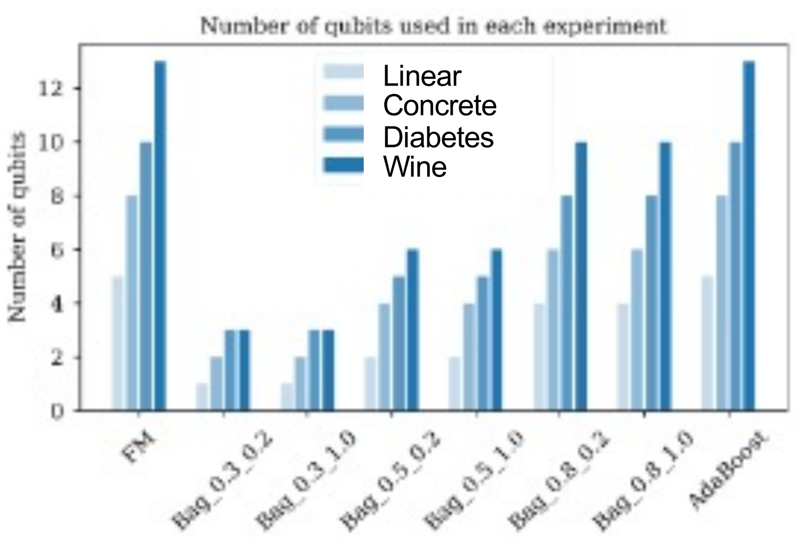


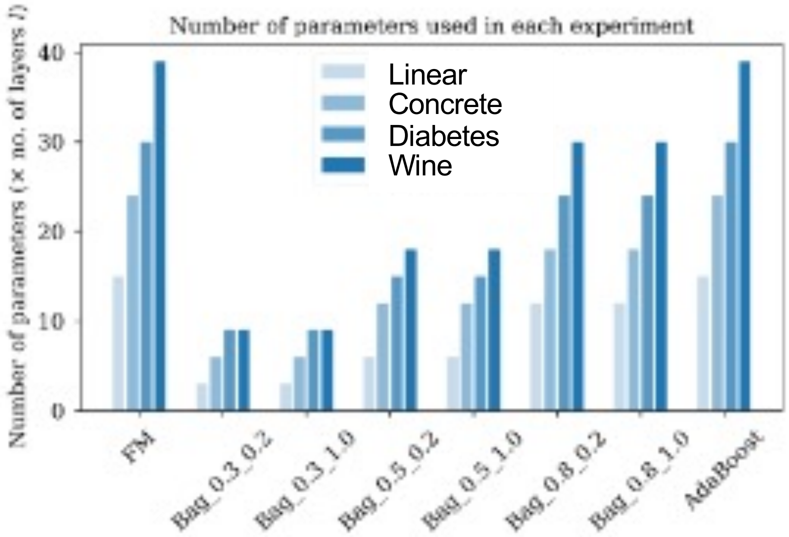

Diabetes (CCE)



# Bagging brings significant advantage

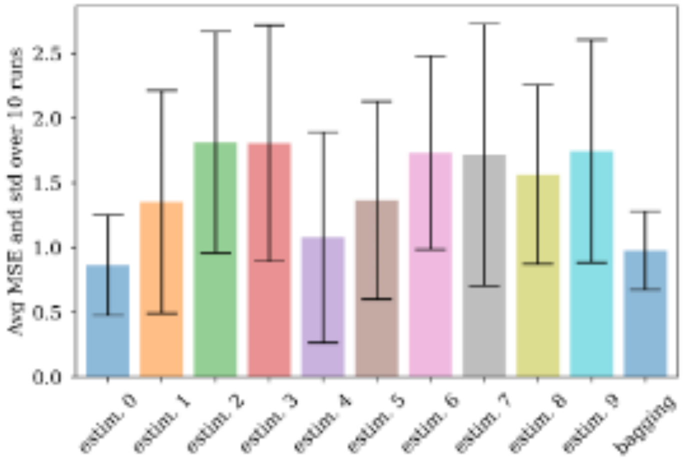
**Reducing resources:**
Best performance for low dimensionality

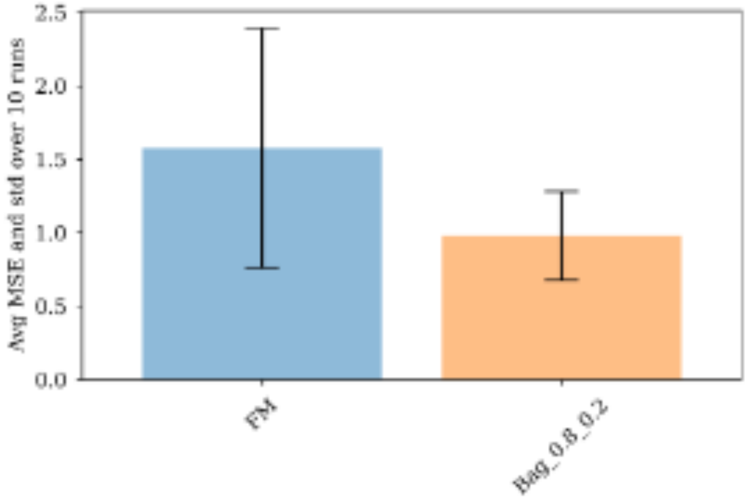


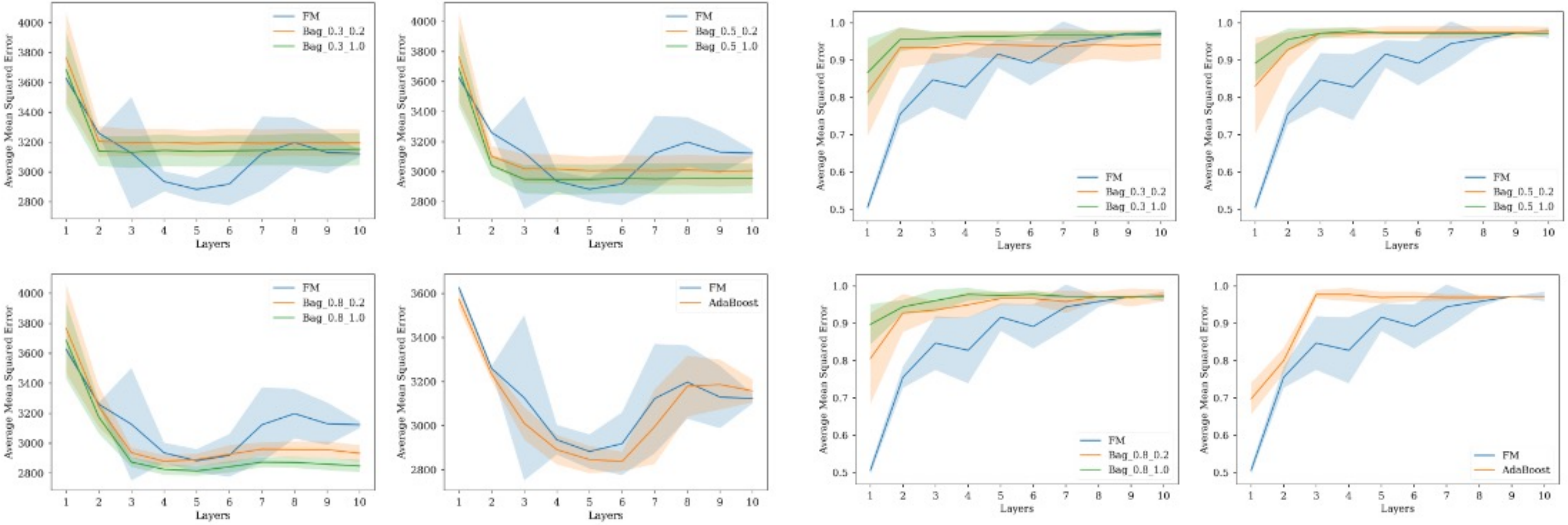

**Robustness against noise:**
Linear regression task on **IBM QPU** (ibm_lagos):

- **Bagging:** 80% features, 20% samples
- **QNN**: 4 qubit, 1 layer



# Uncertainties: Diabetes and Wine dataset



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

# qGAN Benchmarks on hardware

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

- For IBMQ machines, choose the qubits with the lowest

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

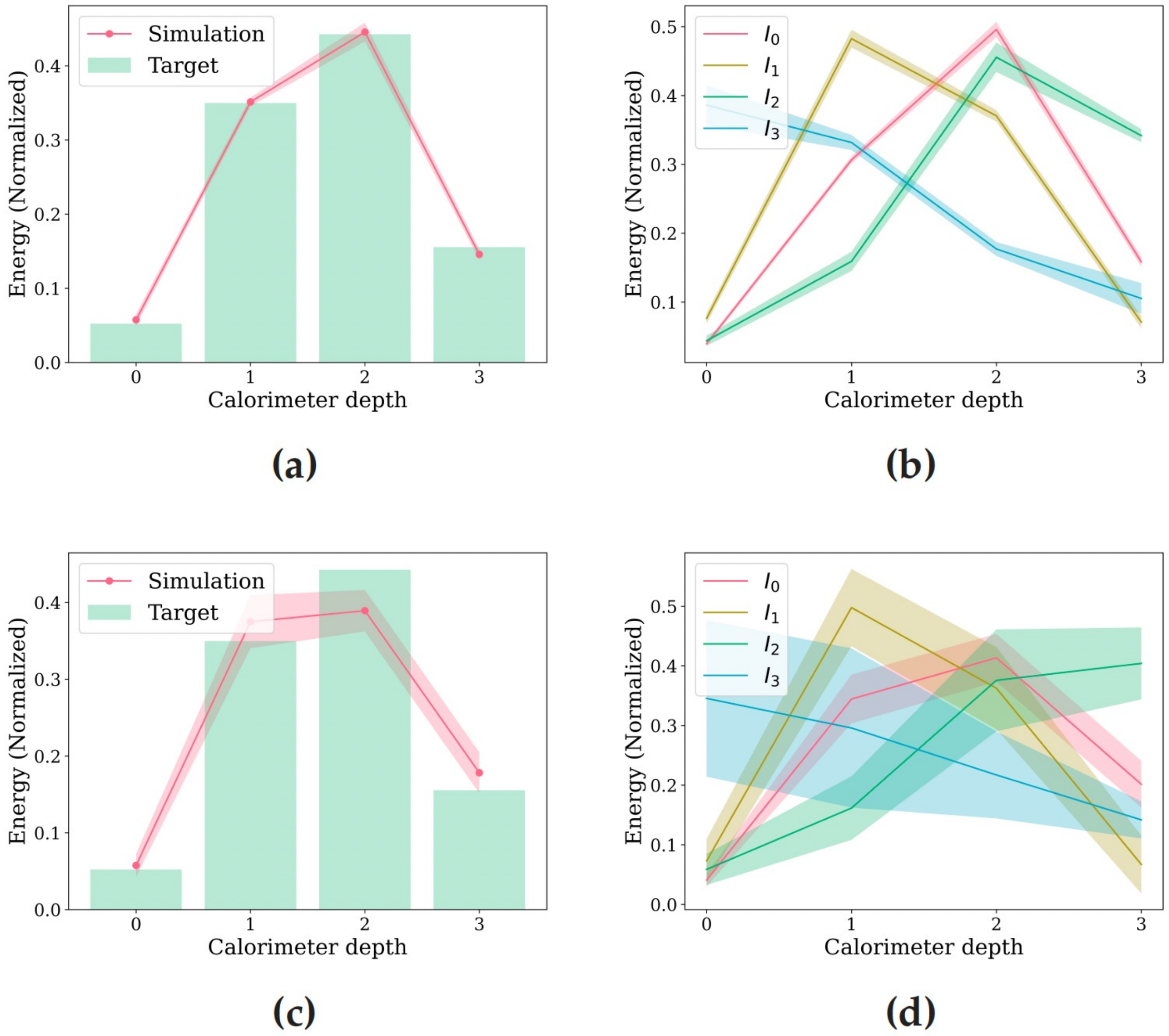
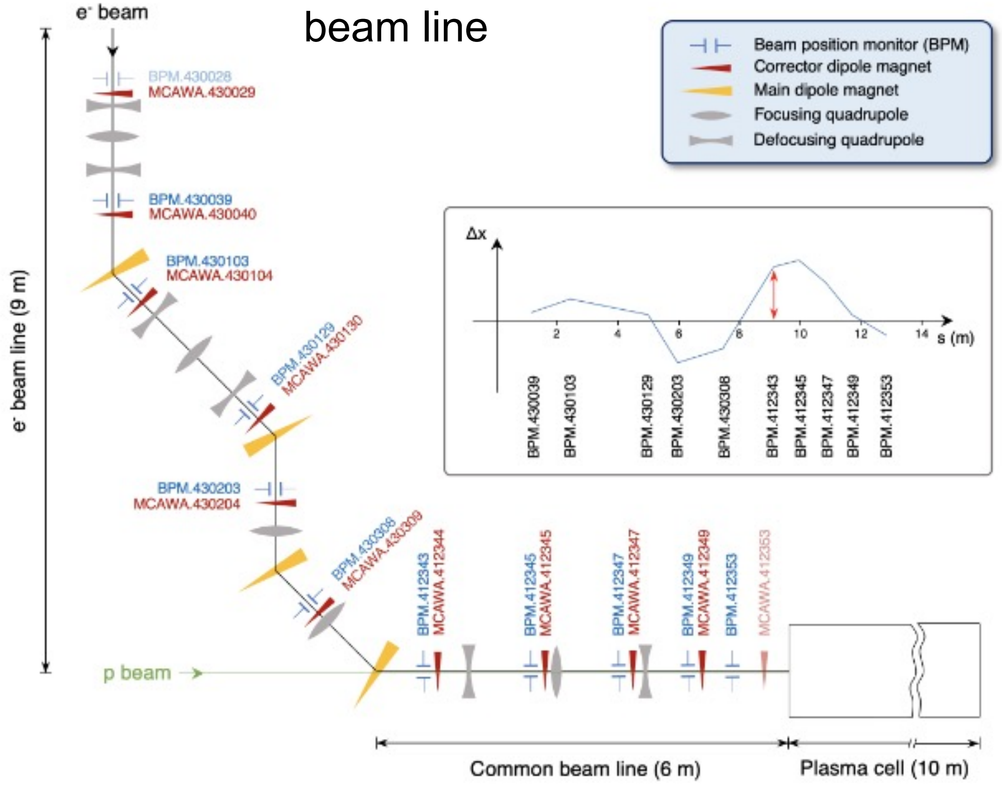
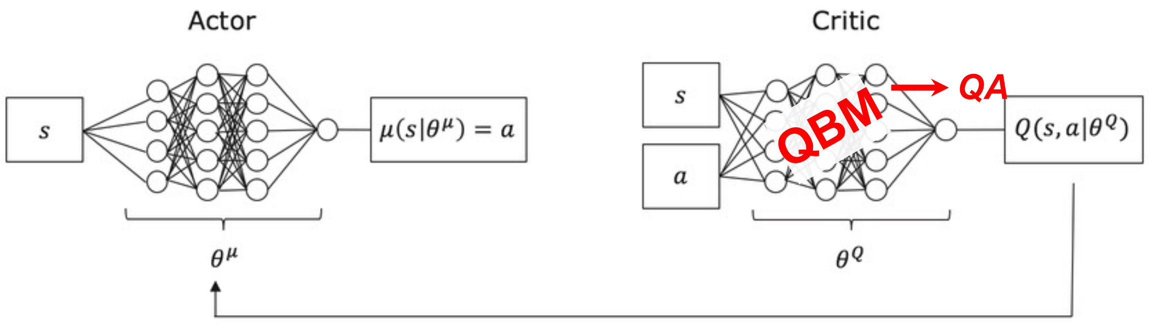



**Figure 4:** Mean (a,c) and individual images (b,d) obtained by inference test on ibmq_jakarta (a,b) and IONQ (c,d).

# CERN AWAKE facility

2GeV electron beam line




**Policy Gradient:** $\nabla_{\theta^\mu}\mu = \mathbb{E}_\mu[\nabla_{\theta^\mu} Q(s, \mu(s|\theta^\mu)|\theta^Q)] = \mathbb{E}_\mu[\nabla_a Q(s, a|\theta^Q) \cdot \nabla_{\theta^\mu}\mu(s|\theta^\mu)]$

**Actor-Critic Q-learning** training D-Wave Advantage

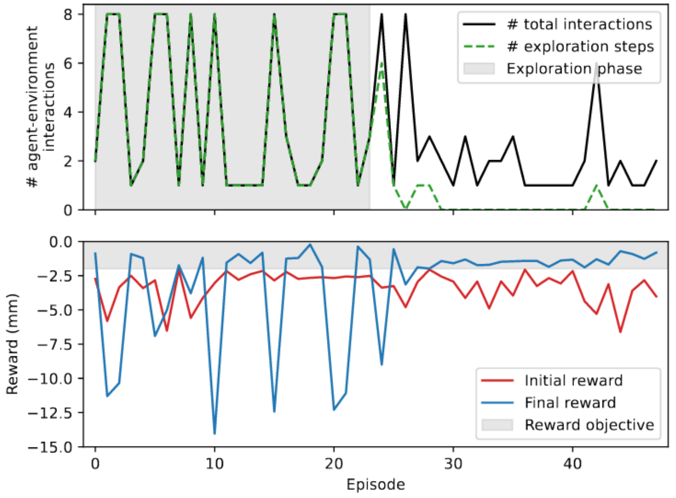
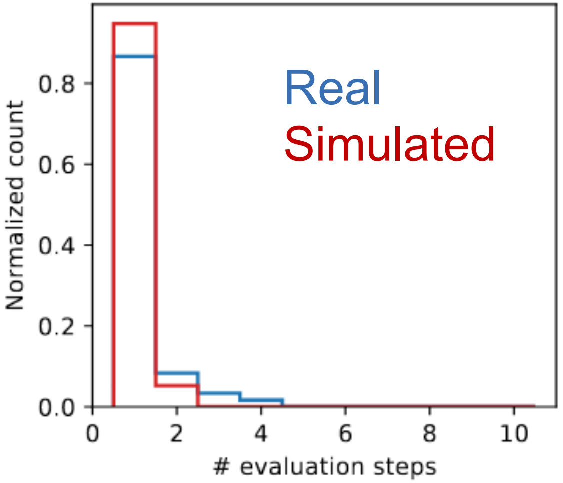



Figure 11: Single RL agent training evolution on D-Wave Advantage Systems using the simulated AWAKE environment with a reward objective of $-2$ mm.

**Successful** evaluation on the real beam-line



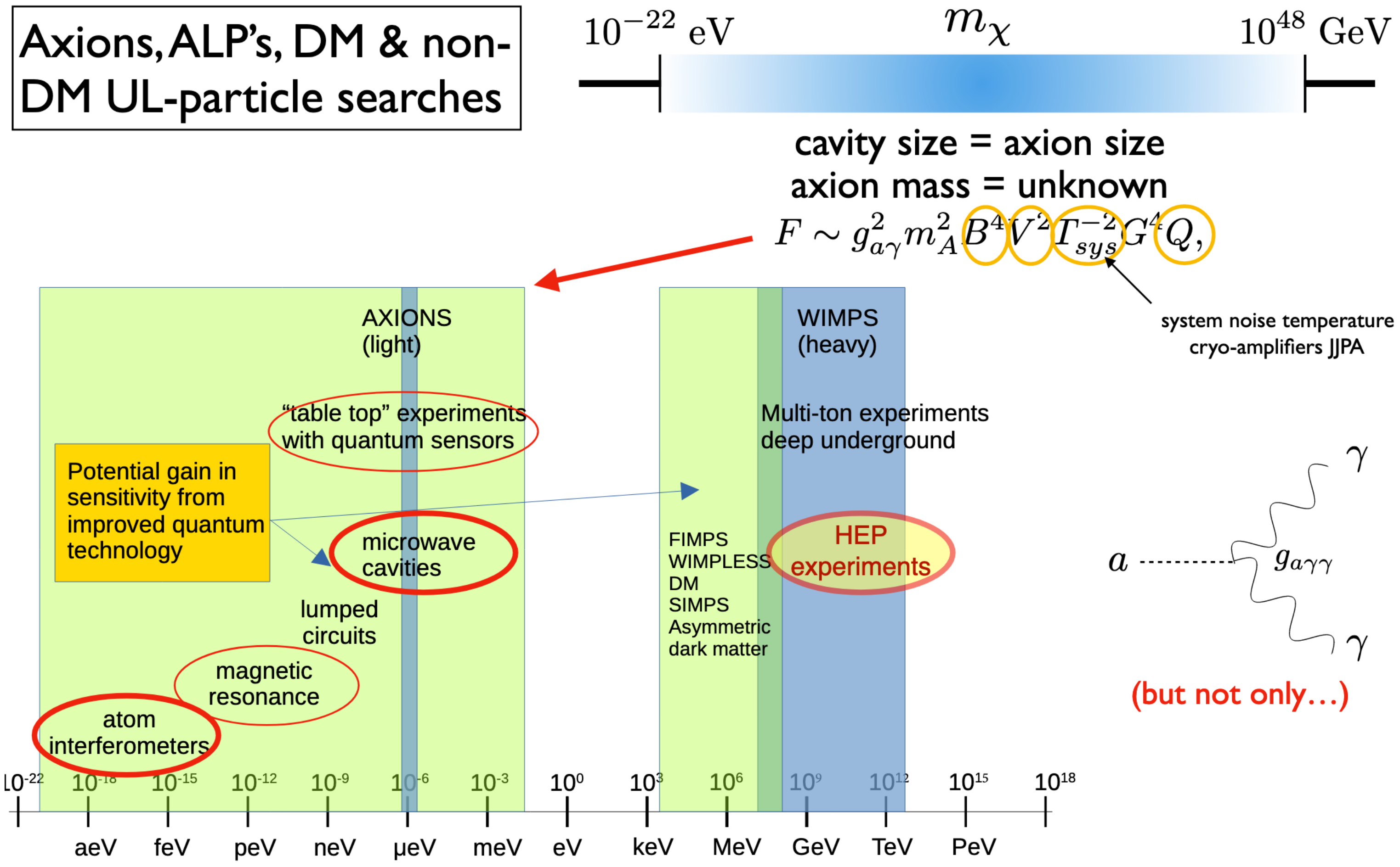

Michael Schenk et al., **Hybrid actor-critic algorithm for quantum reinforcement learning at CERN beam lines**, e-Print: 2209.11044 [quant-ph]

# 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

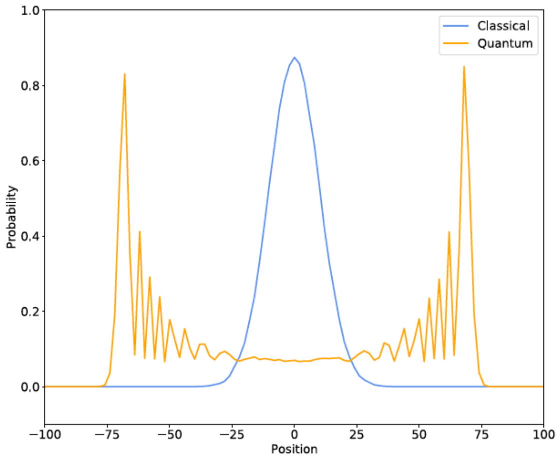
**QFT**: **Focus on computations that are exponentially hard with classical methods. Ex. Sign problems in particle theory**

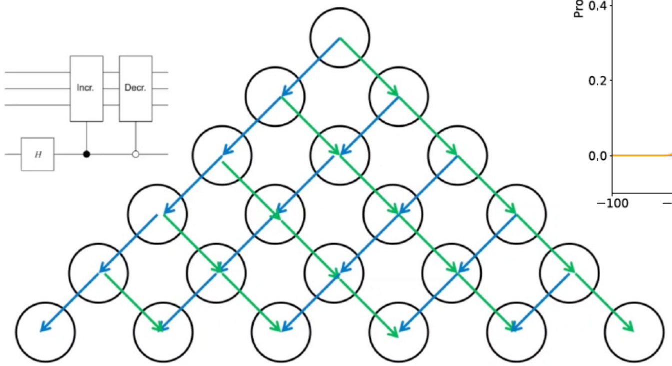
- Dynamical Simulations of Lattice Gauge Theories
- Finite-Density Nuclear Matter
- Challenges related to digitization and truncation of filed representation (on a finite number of quantum states) and redundancy in the Hilbert space[1]

**Cross section integration as quantum amplitude estimation[3]**

**Event generation with quantum generative models or direct simulation**

**Parton showering as quantum random walk[2]**




[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)