Quantum Machine Learning in High Energy Physics

Examples from CERN

QUANTUM
TECHNOLOGY
INITIATIVE

Sofia Vallecorsa
CERN QTI Coordinator
CERN

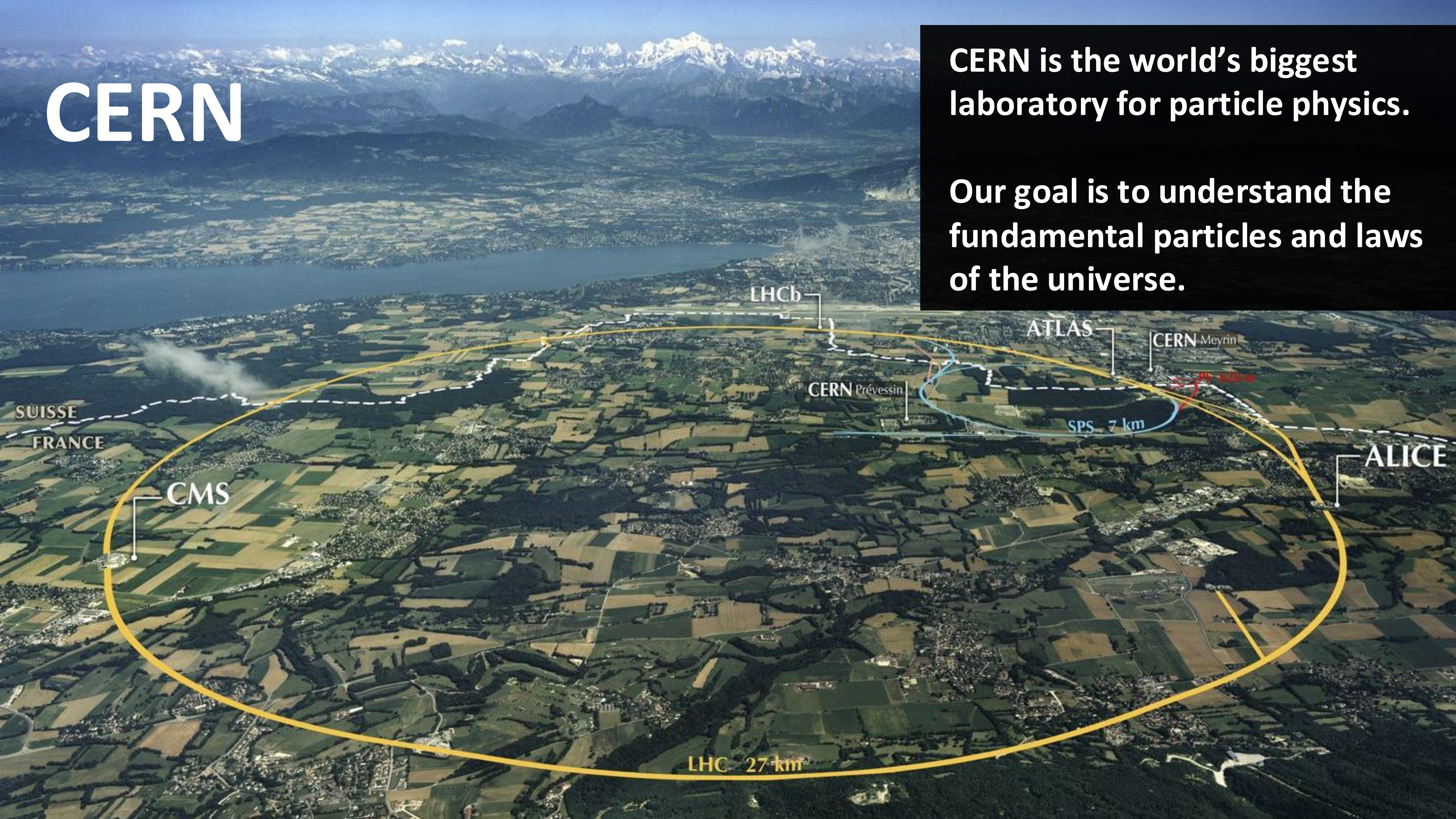
Content

- The **CERN Quantum Technology Initiative**
- Research program in **Quantum Computing**
- **A short detour**
- Introduction **Quantum Machine Learning**
- **Example results**:
  - Kernel – based classification
  - Quantum Guided Compression
  - Anomaly Detection
  - Symmetries and geometric QML
  - Reinforcement Learning
- **The end**

CERN

CERN is the world's biggest laboratory for particle physics.

Our goal is to understand the fundamental particles and laws of the universe.



CERN Quantum Technology Initiative

Launched January 2020

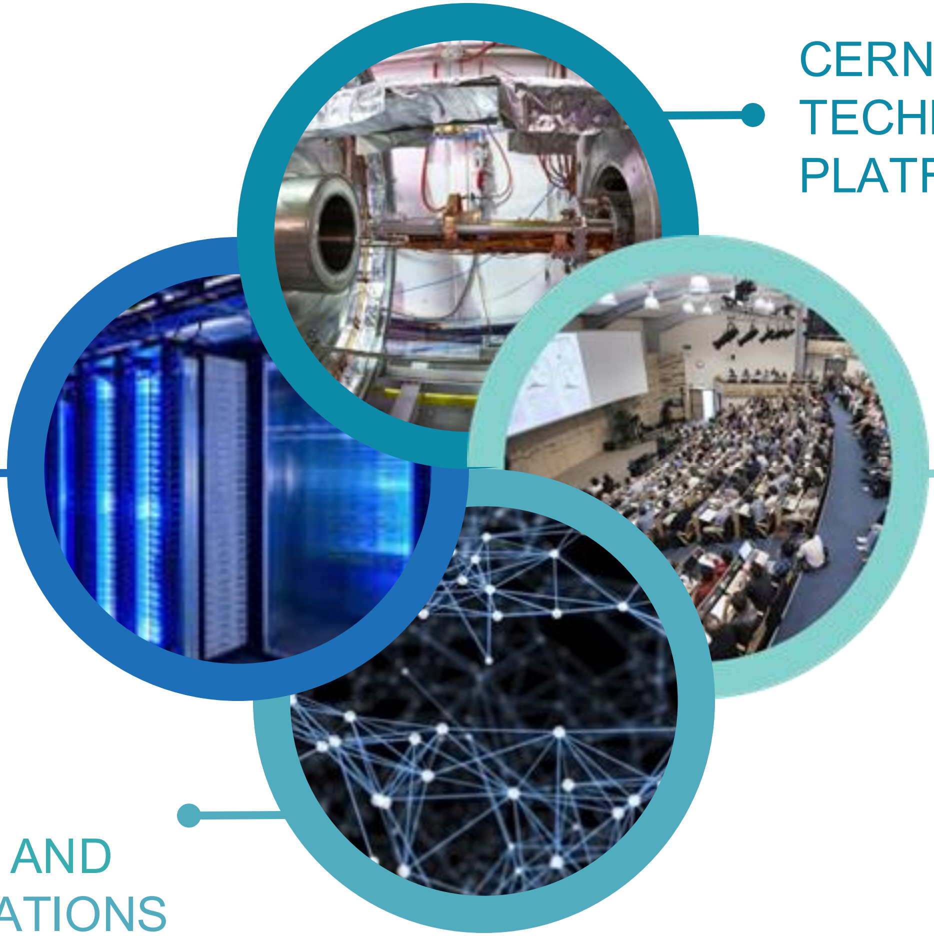
How can future quantum technologies contribute to CERN's scientific mission?

How can CERN's technologies and expertise contribute to the quantum revolution?

QTI Roadmap:
https://doi.org/10.5281/zenodo.5553774

https://quantum.cern

Our areas of research



CERN QUANTUM TECHNOLOGY PLATFORMS

HYBRID QUANTUM COMPUTING AND ALGORITHMS

COLLABORATION FOR IMPACT

QUANTUM NETWORKS AND COMMUNICATIONS

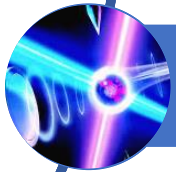
The QTI Hub: A collaboration framework for QTI

The QTI Hub creates a community of partners investigating the differt areas of quantum technologies.

Enable **access to diverse quantum technology and services**

Provide a **unified framework for all collaborative projects** QTI is setting up with multiple partners.

Establish a clear **separation between commercial relationships and R&D collaborations**



Facilitate follow-up and ensure **more efficient coordination of projects** also across departments.

Allow for **multiples approaches to IP protection** according to CERN policies.

QTI Objectives

Integrate quantum computers within HEP computing model

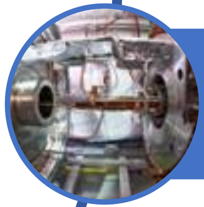
Develop **hybrid algorithms** for realistic applications;

Contribute to **infrastructure development**

Make CERN a node of the future European network infrastructure

Design Quantum Network demonstrators incorporating **White Rabbit** for time synchronization;

Characterize **performance of communication protocols** in realistic use cases



Play a major role in the development of next generation detectors for fundamental physics

Develop **superconducting RF cavities** for sensing and computing applications;

Significant contribution to **ECFA DRD5** program

Join the broader quantum ecosystem to multiply impact

Setup **co-development partnerships** with companies, institutes and other entities.

Hybrid Quantum Computing

Sustain integration of quantum computing within HEP computing model

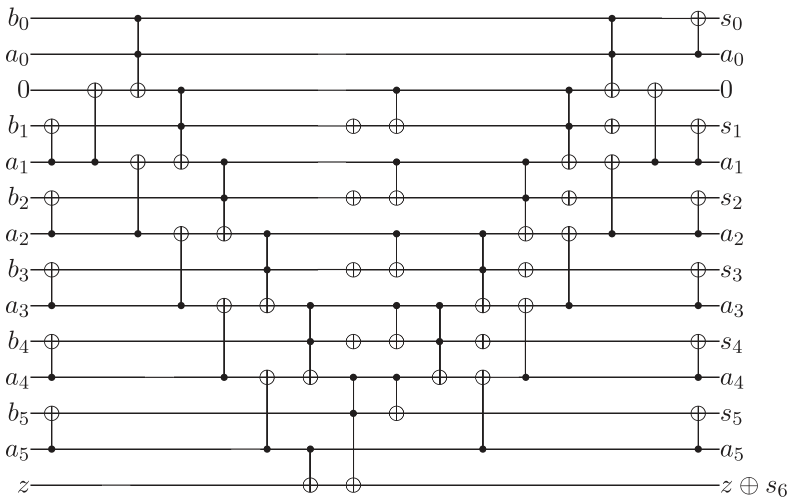
- Develop quantum algorithms and Quantum Machine Learning
- Understand the performance of near-term quantum infrastructure in hybrid setups (HPC + QC, ..)
- Study scaling toward fault tolerant

Most of these developments are common to areas beyond HEP

https://home.cern/science/computing. Image credit: CERN

Main Quantum Computing Paradigms

Gate-based quantum computers

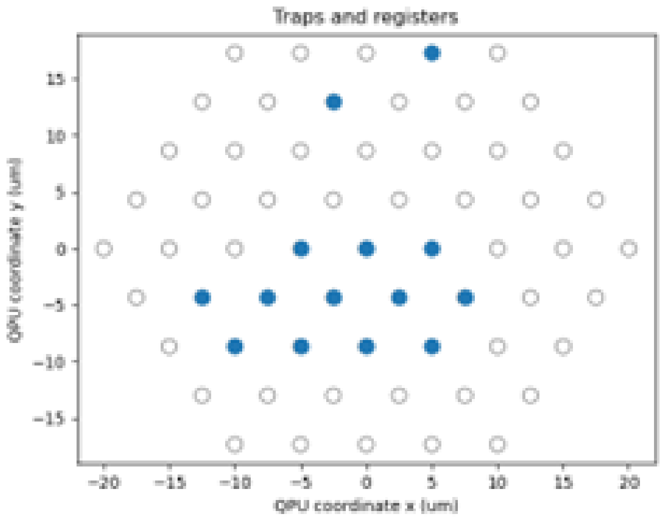


The ripple-carry adder for $n = 6$.

Solve task with an algorithm containing a **series of quantum gates**, implementing any unitary transformation

IBM QUANDELA

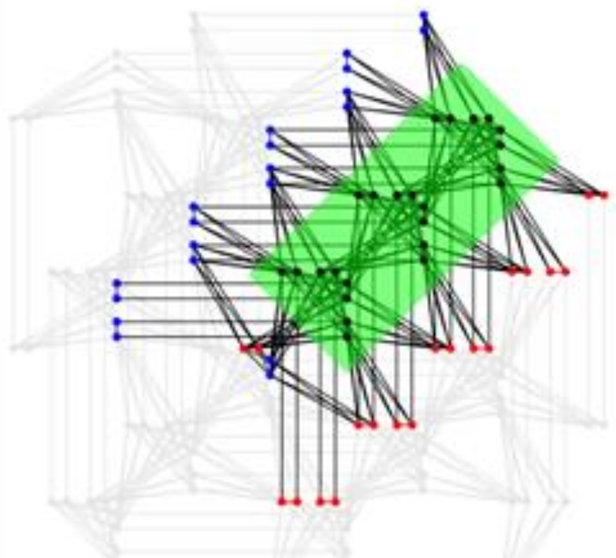
Analog quantum simulators



Embed task in a graph & solve Ising or QUBO formulation, using dynamic qubit positioning but no or poor local qubit control

Pasqal QuEra

Quantum annealers

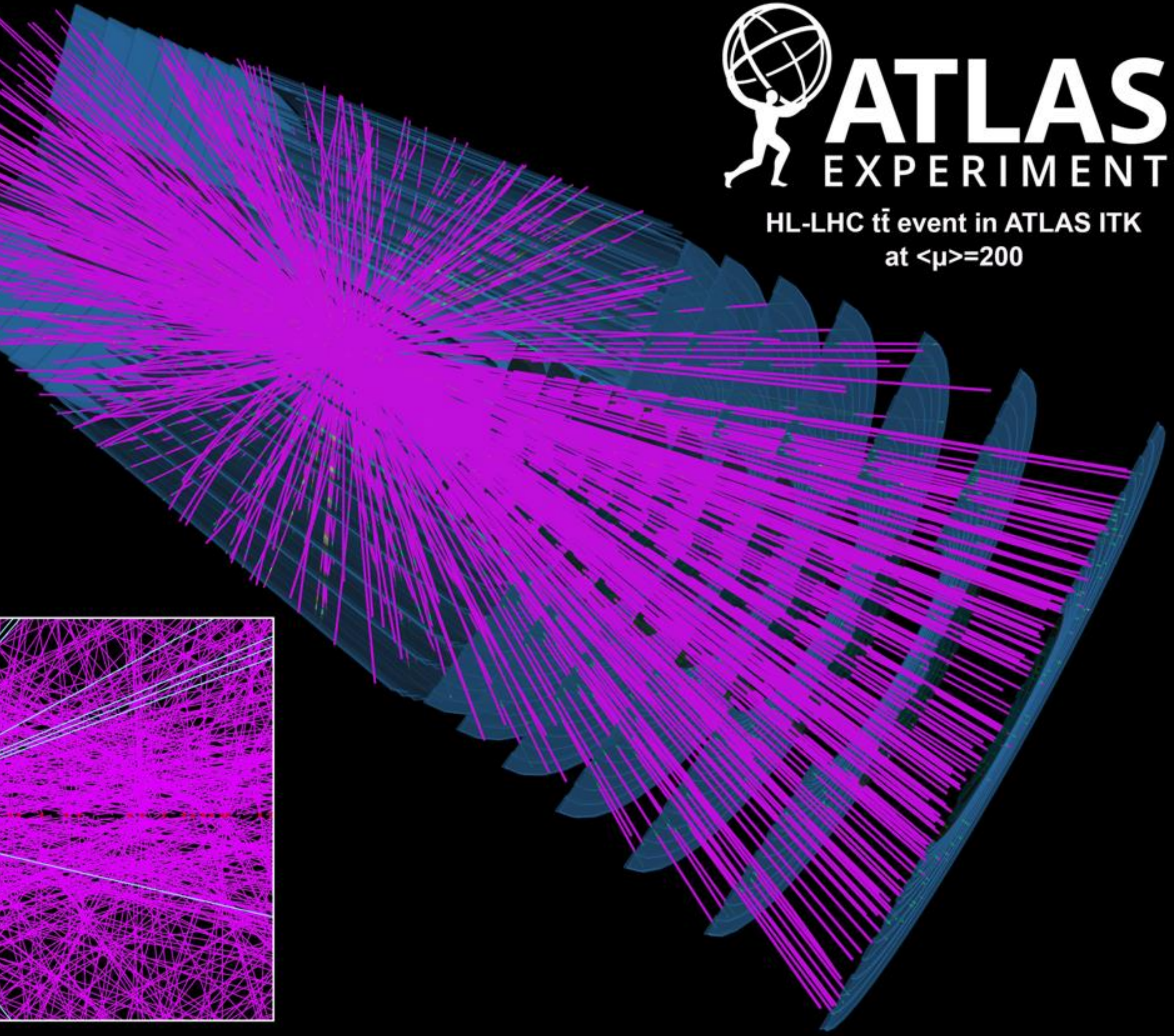


Embed task in a Binary Quadratic Model & solve Ising or QUBO problems, using static qubit connectivity and local control

D-Wave The Quantum Computing Company™

Motivation

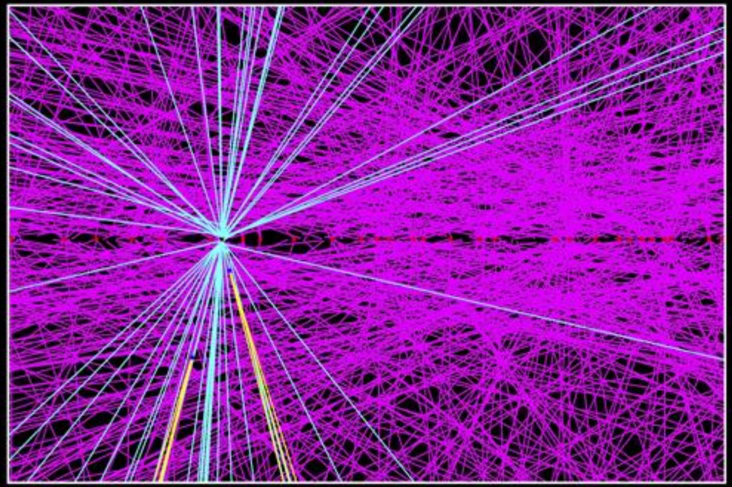
HL-LHC: The curse of dimensionality



ATLAS EXPERIMENT

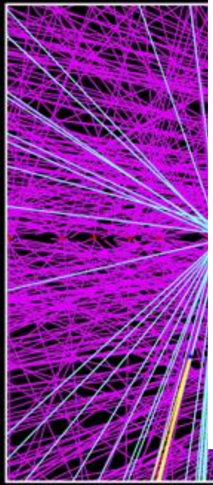
HL-LHC $t\bar{t}$ event in ATLAS ITK
at <μ>=200

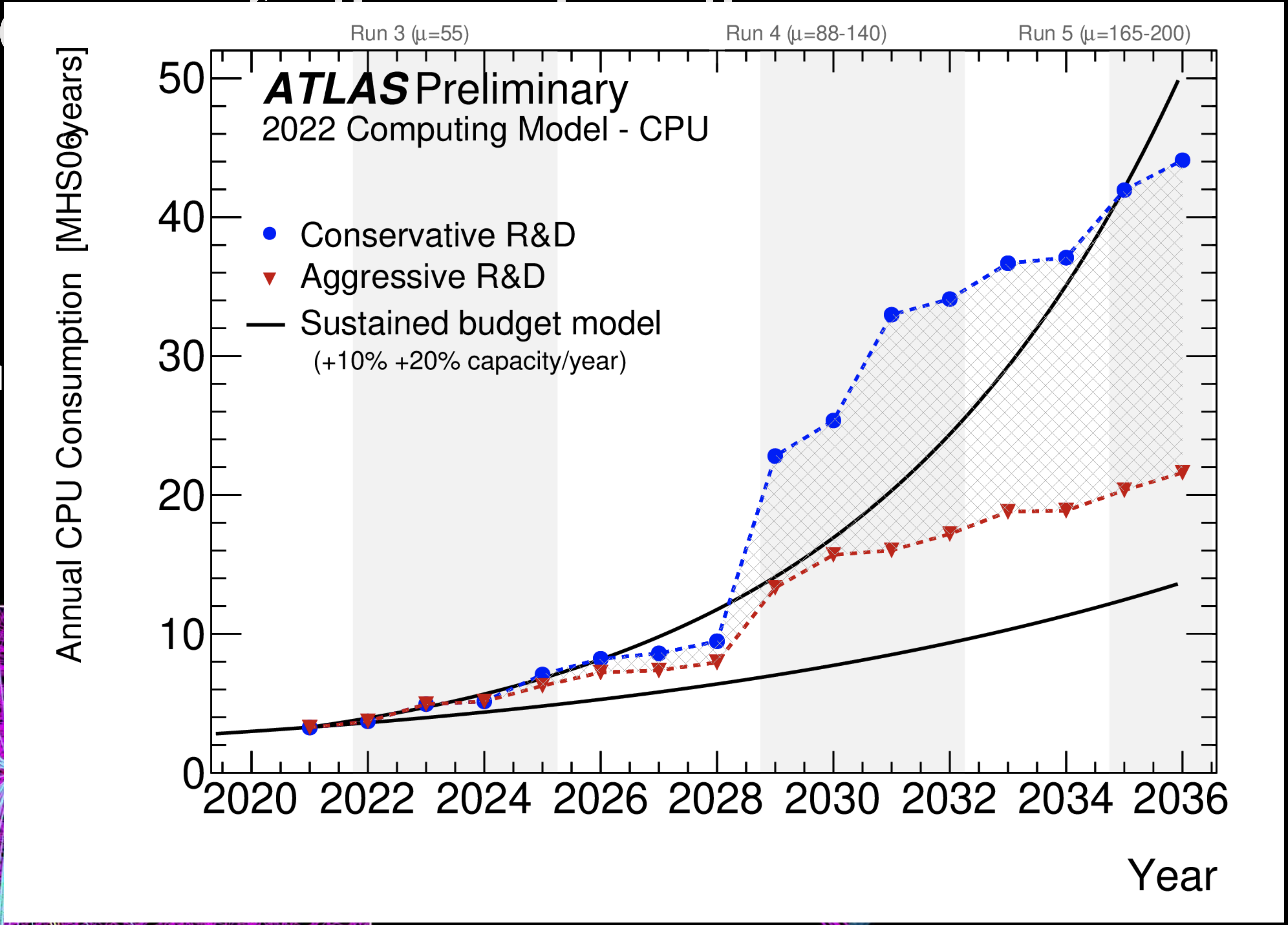
200 simultaneous
collisions!



HL-LHC: The

200 simultaneou
collisions!

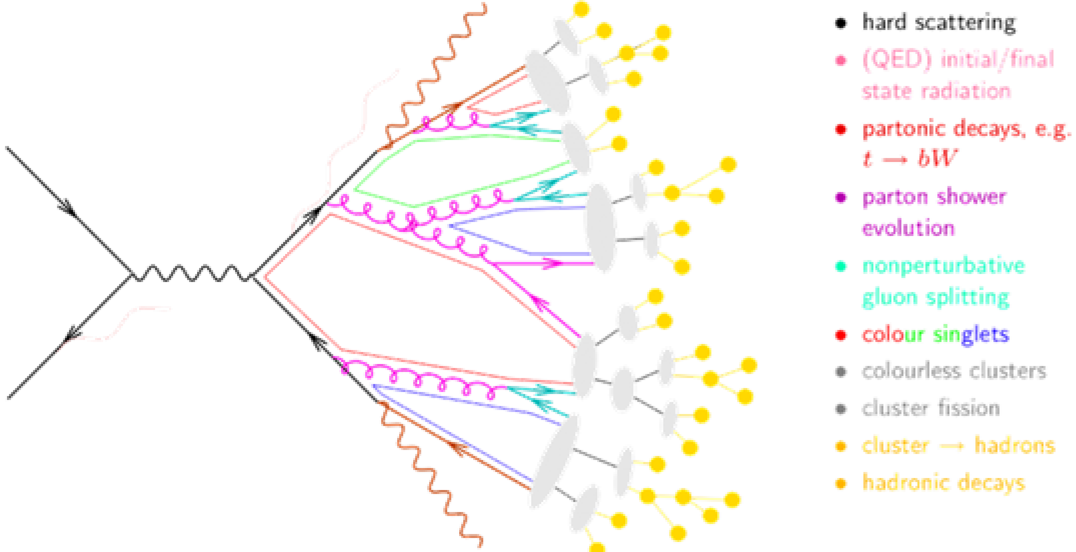




Theory and simulations challenges

- We are interested **in out-of equilibrium and real-time dynamic problems**

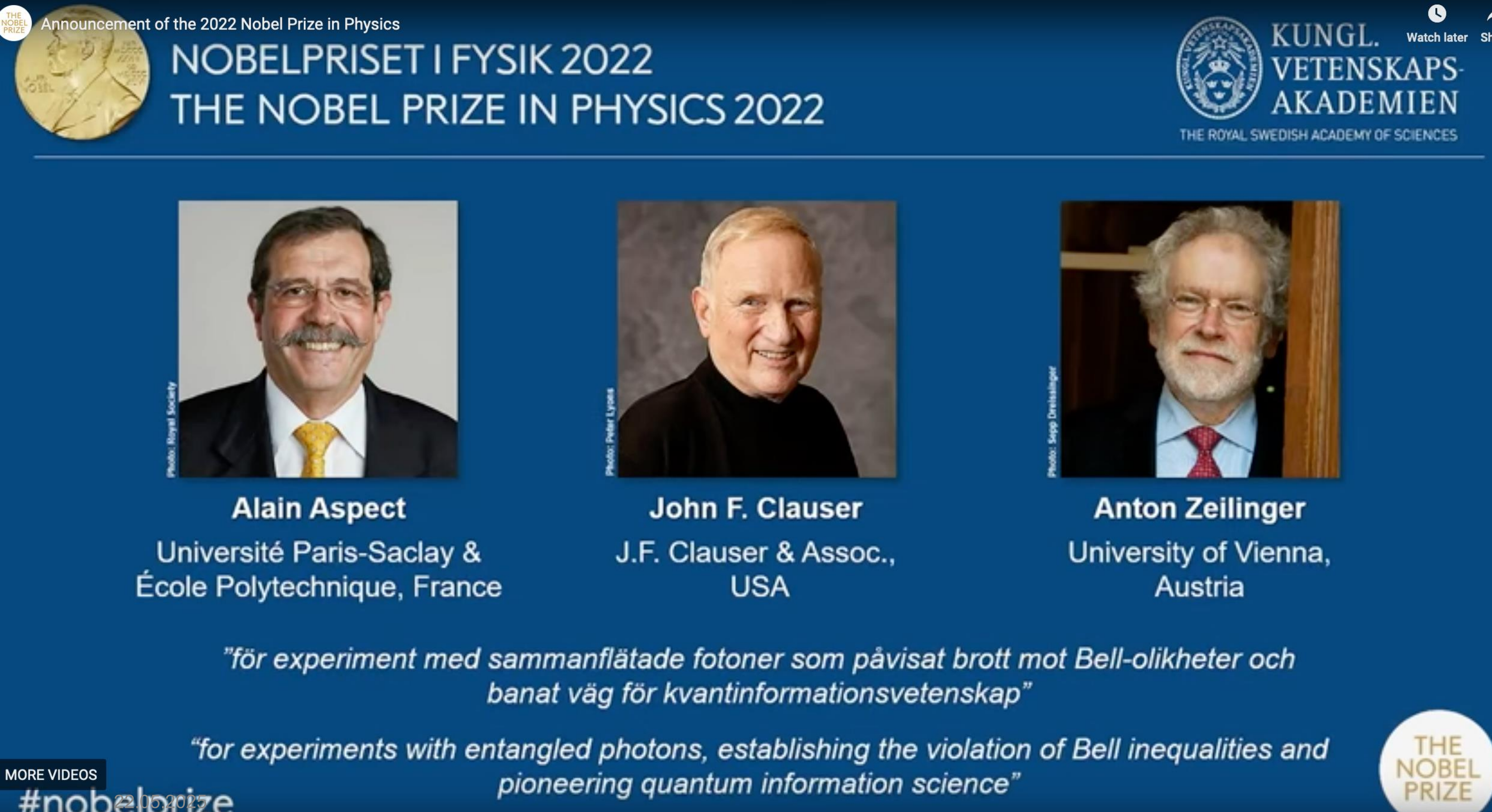
  (scattering, thermalisation or dynamics after quenches)
- **Complex equation of states and phase diagrams (QCD)**
- **Standard Monte Carlo solutions are two expensive or fail entirely**



Why do we think that Quantum Computers could be a solution to data simulation and data analysis in HEP ?

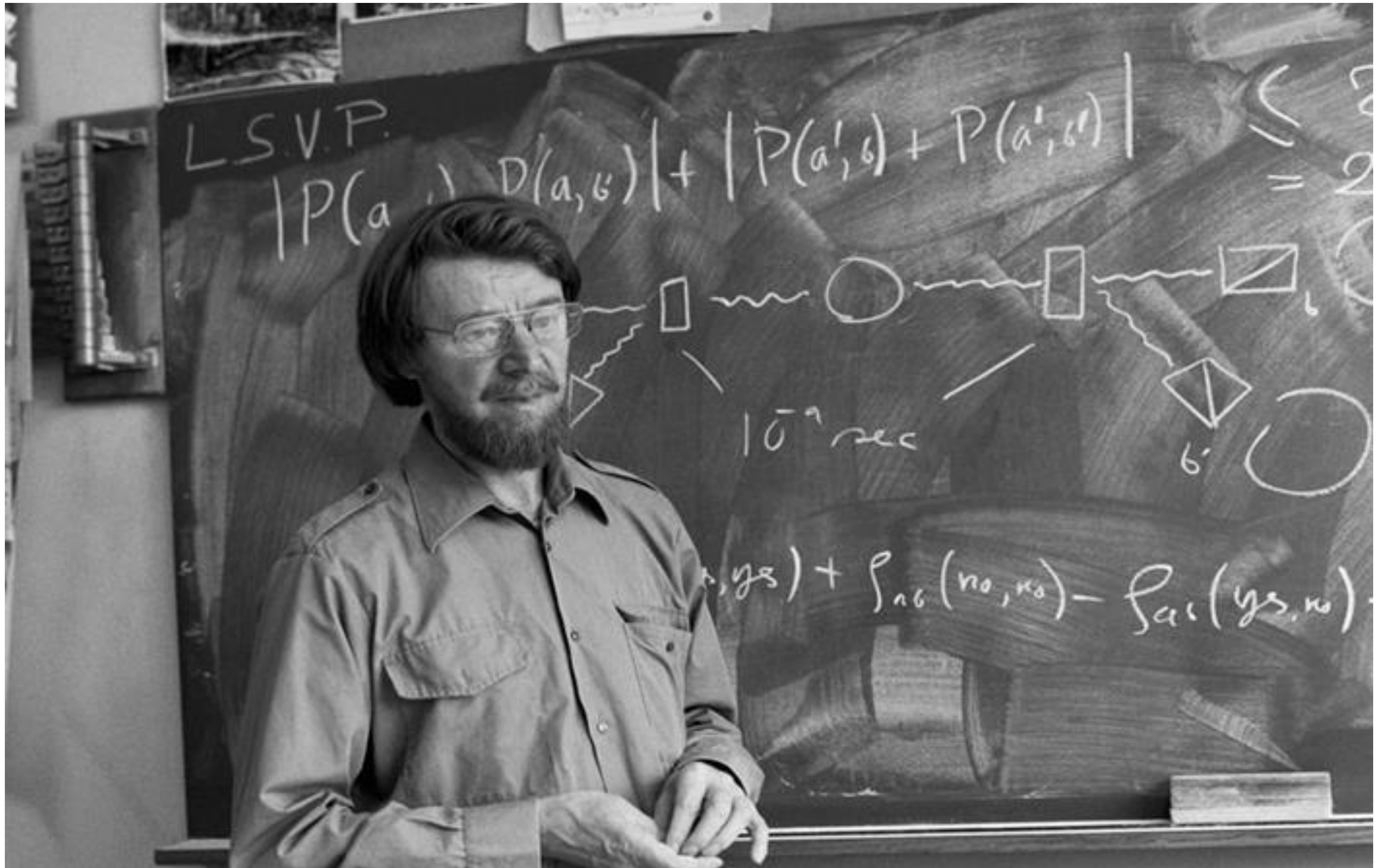
High Energy Physics studies quantum correlations at high energy

A short detour…



https://youtu.be/mtgYG2zsbbQ

The Bell inequalities



John Bell in 1982 at CERN (Image: CERN)

- 1964: **Bell inequalities** prove that no theory based on **local hidden variables** (realism) can reproduce QM results

- Major step confirming the possibility of using distant entangled photons as a quantum information resource

- Create **"artificial" quantum states** for a range of applications (single photons, trapped ions, superconductors, etc.)

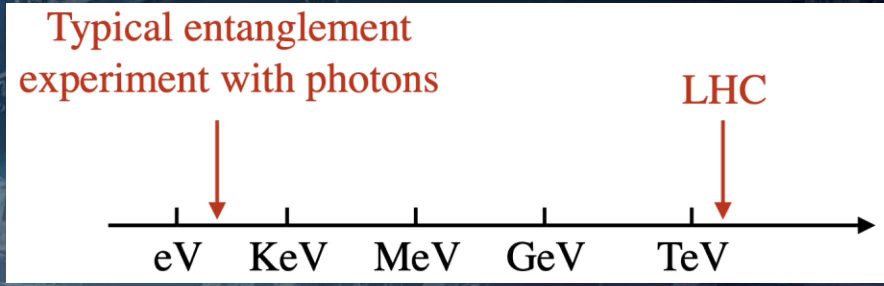
Quantum Technologies

Observation of quantum entanglement in top quark pair production in proton–proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$

The CMS Collaboration

Published 23 October 2024 • © CERN, for the benefit of the CMS Collaboration
Reports on Progress in Physics, Volume 87, Number 11
Citation The CMS Collaboration 2024 *Rep. Prog. Phys.* **87** 117801
DOI 10.1088/1361-6633/ad7e4d

The highest energy observation of quantum entanglement



Nature. 2024 Sep 18;633(8030):542–547. doi: 10.1038/s41586-024-07824-z

Observation of quantum entanglement with top quarks at the ATLAS detector

The ATLAS Collaboration[1]

Author information   Article notes   Copyright and License information

PMCID: PMC11410654 PMID: 39294352

QML concept and examples

Quantum Computing .. A computer science perspective

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 measurement according to Born rule → **stochastic computation ?**

No-cloning theorem → **information security**

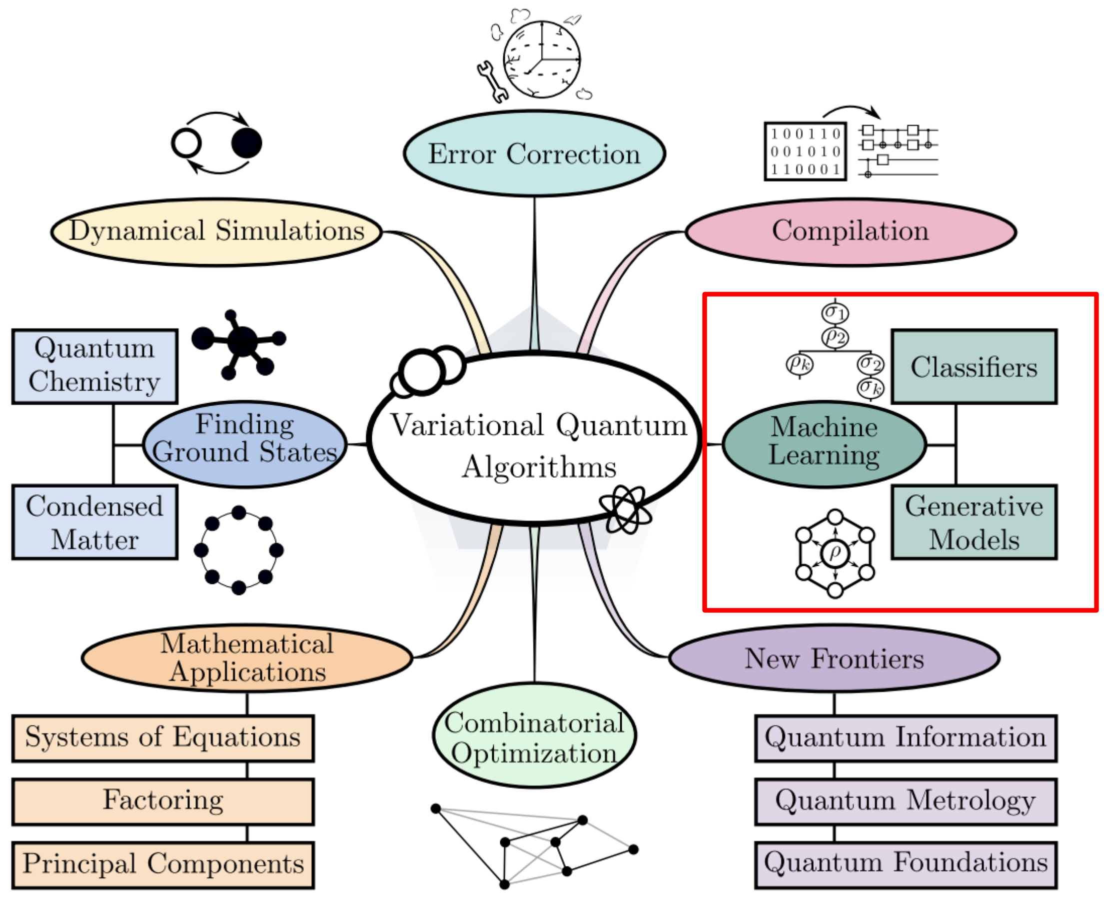
Quantum state coherence and isolation → **computation stability and errors**

Qubit state collapses → **reproducibility?**

QML: Quantum computing to "improve" ML

- Speed-up and complexity
- Sample efficiency
- Representational power
- Energy efficiency???

- Evaluate performance on realistic use cases
- QPU as accelerators within classical infrastructure?

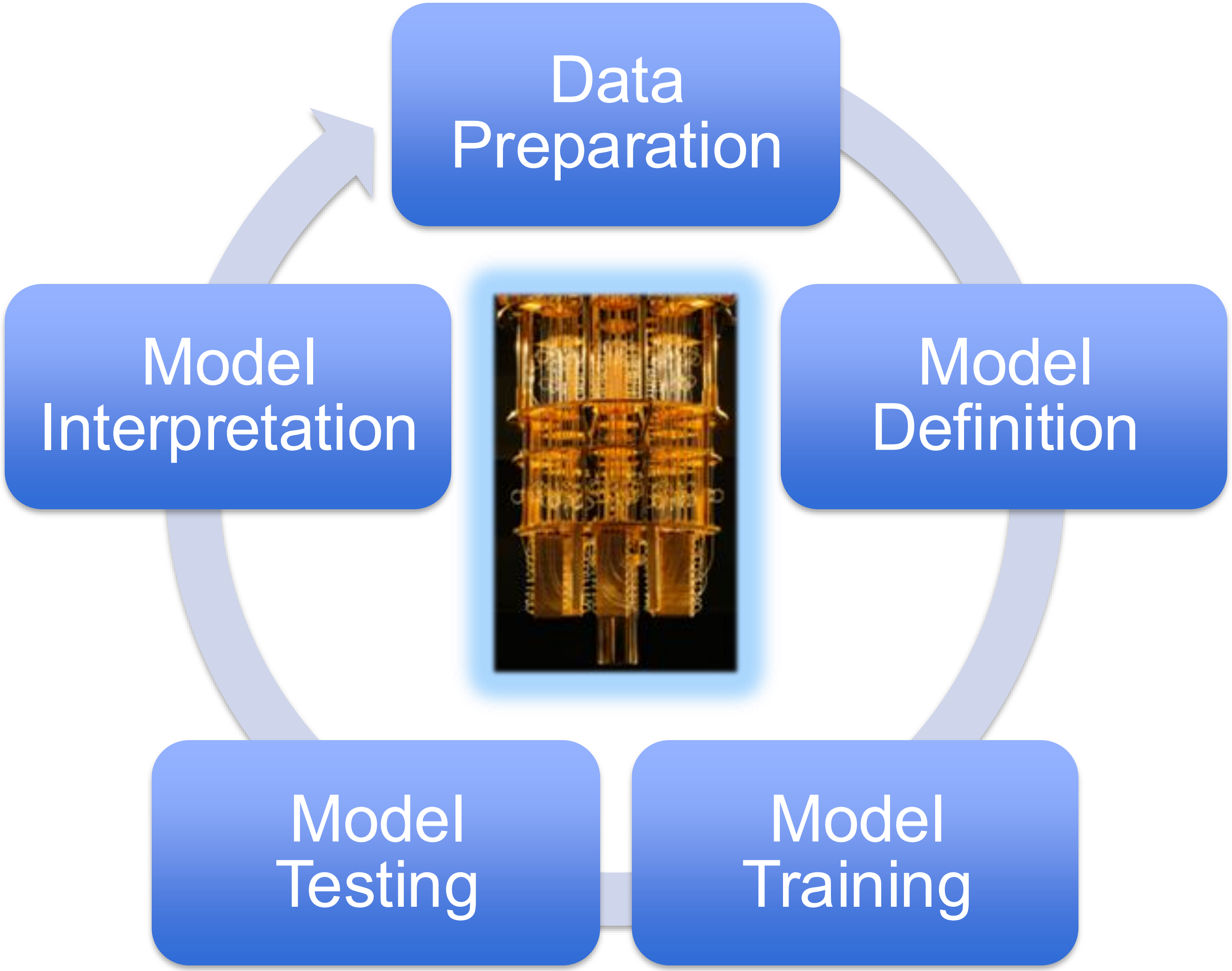


Study **classical intractability**:
Focus on quantum circuits that are **not efficiently simulable classically?**

Cerezo, Marco, *et al*. "*Variational quantum algorithms.*"Nature Reviews Physics3.9 (2021)

Quantum Machine Learning Lifecycle

The quantum advantage of many known QML algorithms is impeded by an input or output bottleneck



Models

Variational algorithms (ex. QNN)

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

Data Embedding can be **learned**

Ansatz design can leverage data symmetries[1]

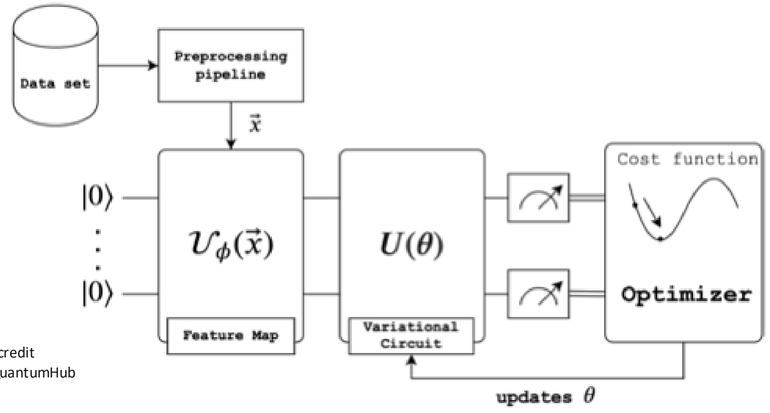


Image credit
Swiss QuantumHub

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]

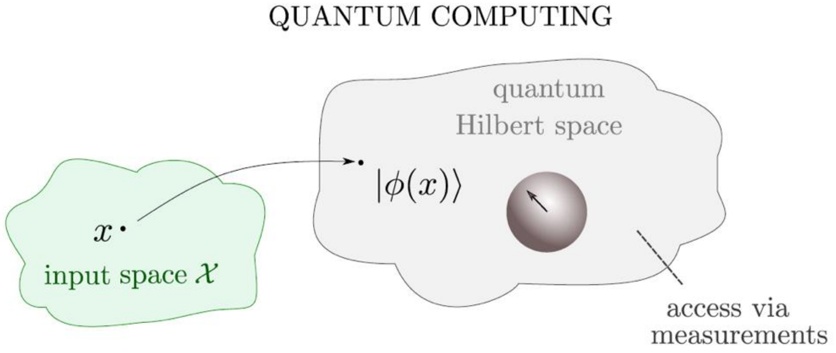


Image credit M. Schuld

Energy-based ML (ex. QBM)

Build networks 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:2110.13162* (2021).

Parameter optimization

$$\theta_{k+1} \leftarrow \theta_k - a_k \underbrace{\hat{g}_k(\hat{\theta}_k)}_{\text{stochastic estimate of } \nabla_\theta f}$$

Simultaneous Perturbation Stochastic Approximation (SPSA) (gradient-free)

If gradient computation not possible, too resource-intensive, or noise-robustness required (slower convergence but fewer function evaluations)

Gradient is approximated by two sampling steps and parameters are perturbed in all directions simultaneously

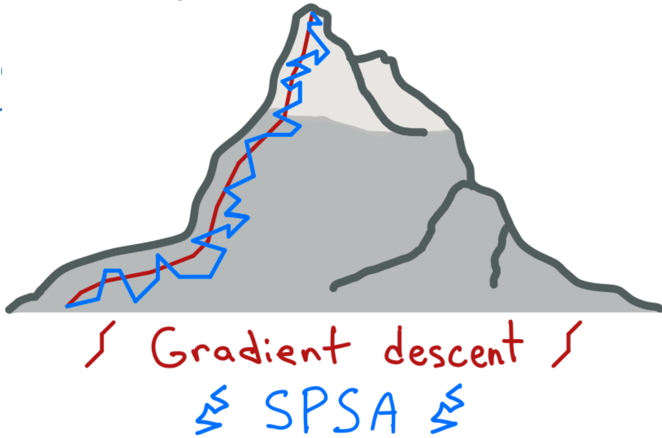
"$y(\theta) = f(\theta) + \varepsilon$" random output perturbation

$$\hat{g}_{ki}(\hat{\theta}_k) = \frac{y(\hat{\theta}_k + c_k \Delta_k) - y(\hat{\theta}_k - c_k \Delta_k)}{2 c_k \Delta_{ki}}$$

$c_k \geq 0$, $\Delta_k = (\Delta_{k_1}, \Delta_{k_2}, \ldots, \Delta_{k_p})^T$ perturbation vector (~ randomly sampled from zero-mean distr.)

Iterative update rule comparable t classical stochastic gradient descer

https://pennylane.ai/qml/demos/tutorial_spsa

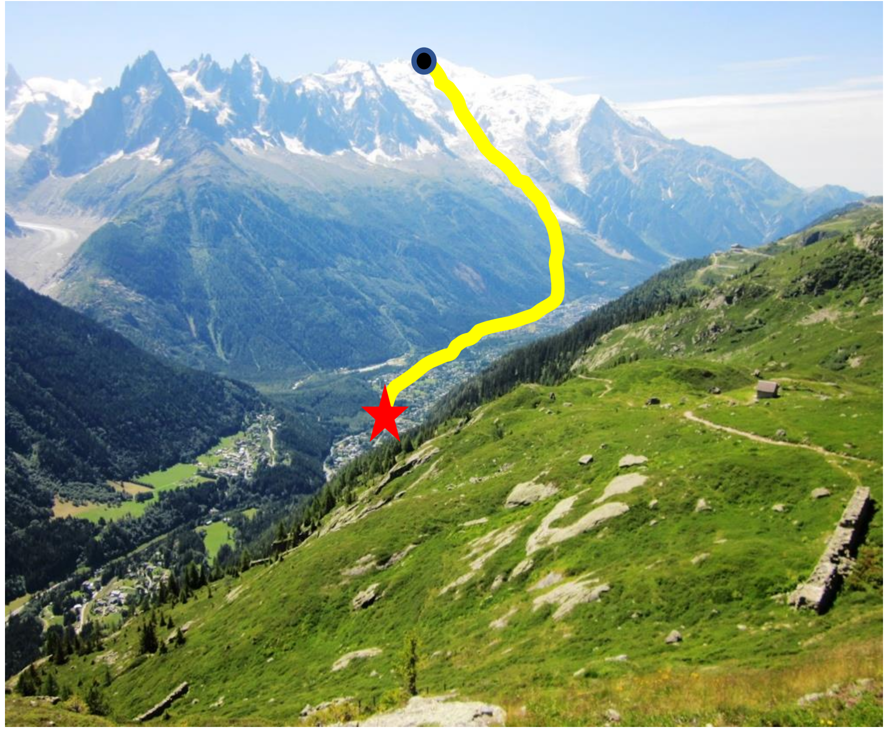


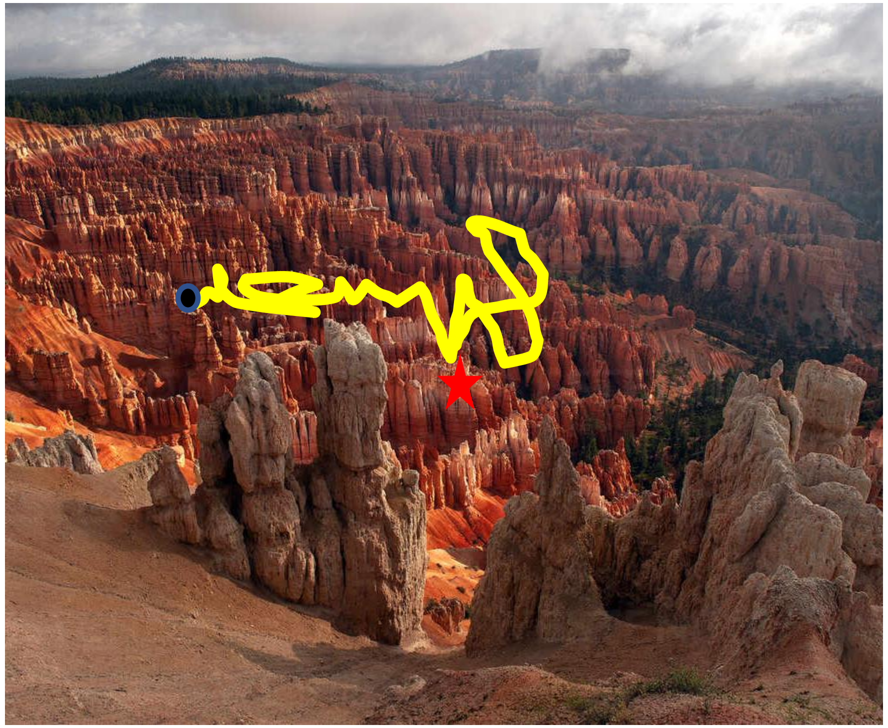
Model convergence in the quantum space

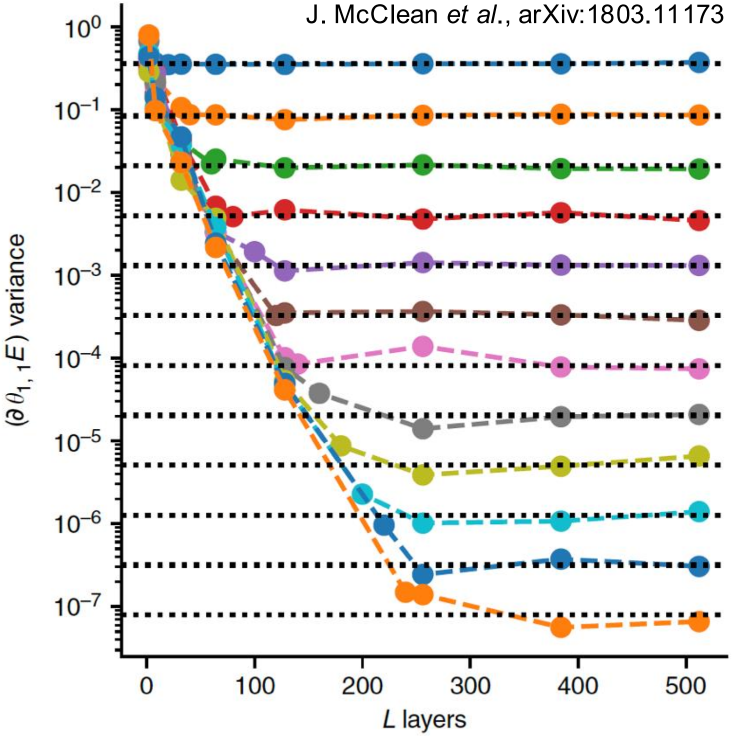
Gradient-based optimization suffers from ”barren plateaus”

Quantum NN are strongly affected

Need **compromise** between “power” and convergence







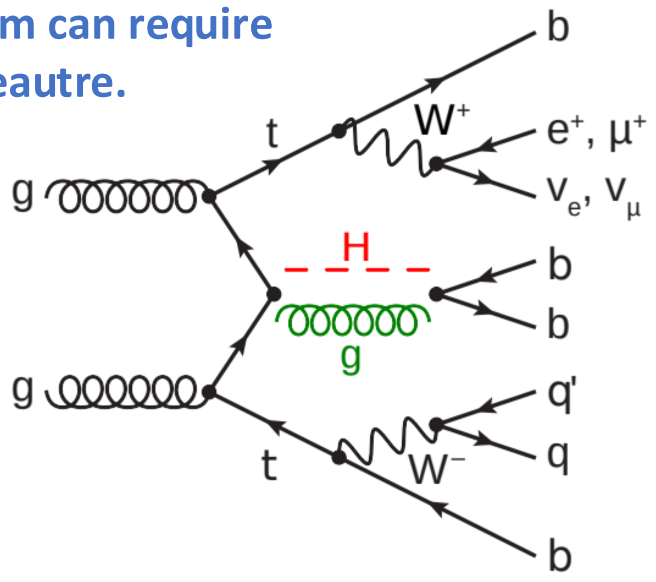
Challenges for QML

- Efficient **data handling** and data **embedding**
- Find balance: **Generalization** and **representational power** vs. **Convergence** and **intractability**
    - Problem of barren plateaus and vanishing gradients in optimization landscape
    - How well can we survey the Hilbert space (expressibility)?

- **Current hardware limitations**
    - Limited number of qubits and connectivity → **data dimensionality reduction**
    - **Quantum Noise Effects** (decoherence, measurement errors or gate-level errors)
    - Efficient interplay between classical and quantum computer
- ….

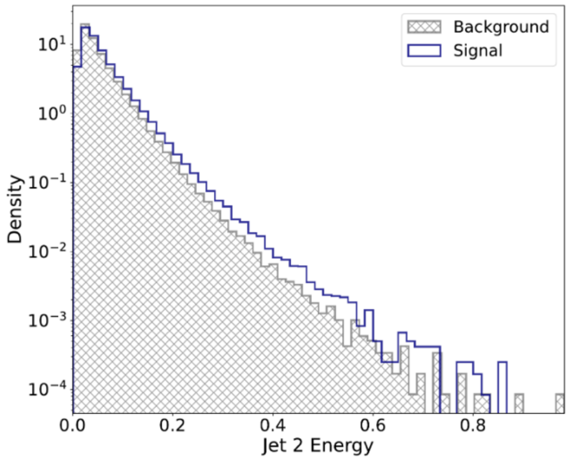
Typical data analysis setup

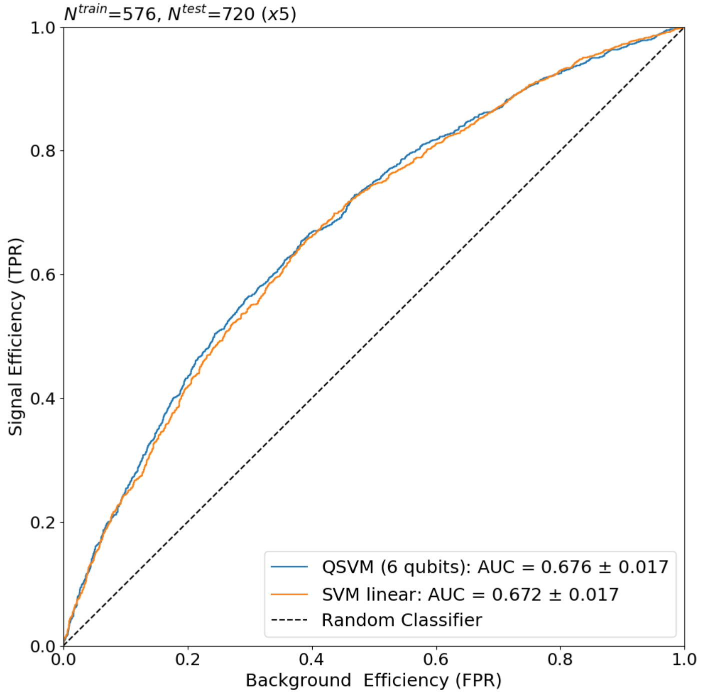
Such a problem can require hundreds of feautre.



Signal vs background discrimination

- Define a number of features that distinguish signal wrt background
- Study and characterise those features
- Build criteria for improving separation



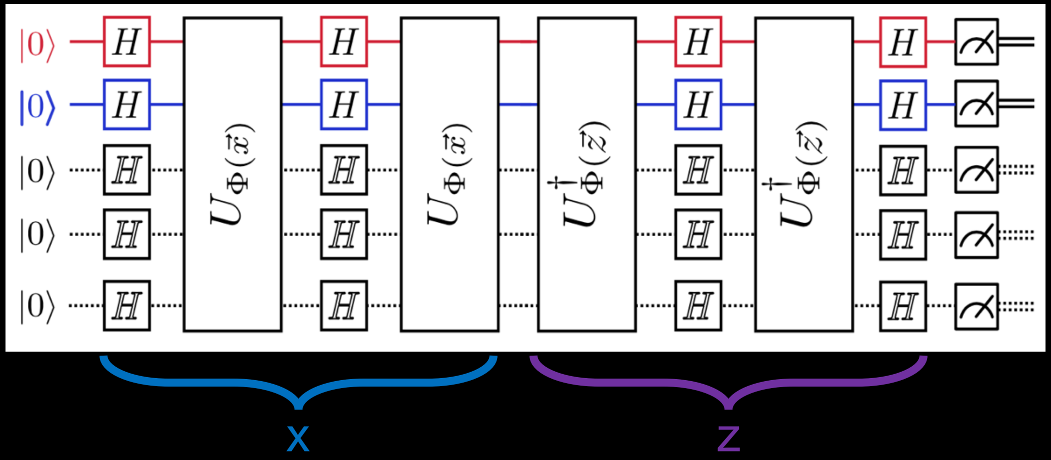


Results are given in terms of ROC curves

Can QML leverage the «exponential advantage ?»

Quantum embedding and kernel methods

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



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

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

Hilbert space is exponentially larger

↓

Sparser data

↓

Loss of predictive power

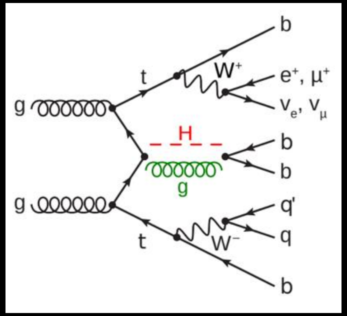
Projected Quantum Kernel

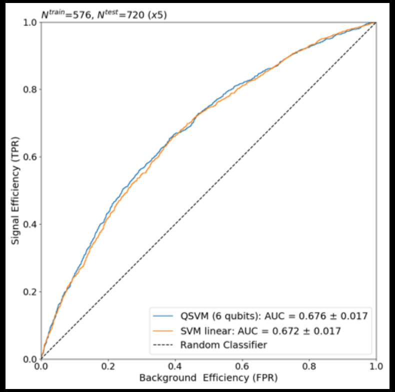
$$k^{lp}(x_i, x_j) = \sum_{k=1}^{m} \frac{\mathrm{Tr}[\rho_k(x_i)\rho_k(x_j)]}{m}$$

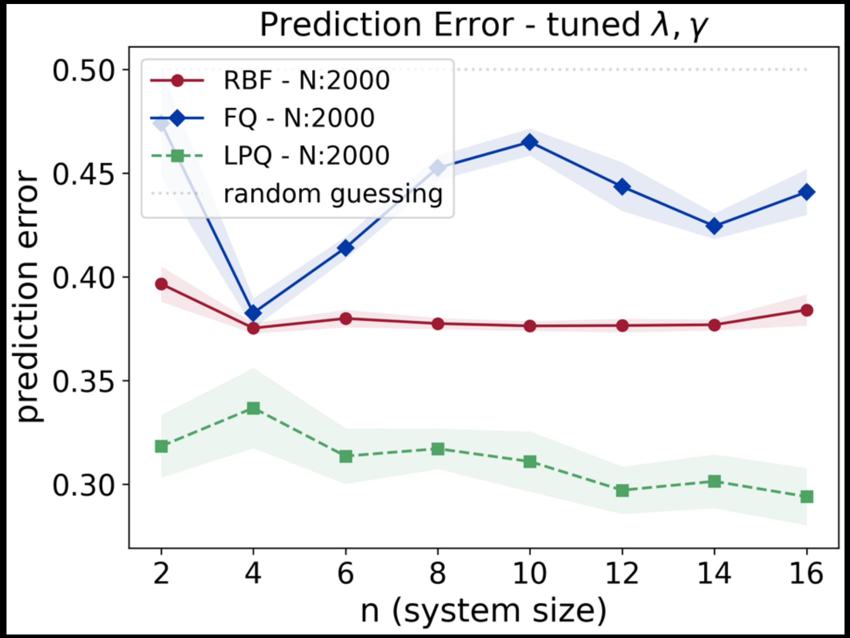
Project quantum kernels to lower dimensionality (i.e. local density matrix)[1]:

- Improved generalizion while keeping features into states classically hard

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





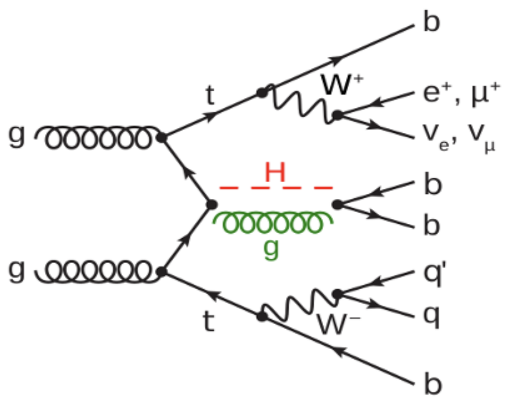


[1]Huang, Hsin-Yuan, et al. "Power of data in quantum machine learning." *Nature communications* 12.1 (2021): 2631.
[2] V Belis et al, (2021), *Higgs Analysis with Quantum Classifiers,* EPJ Web Conf

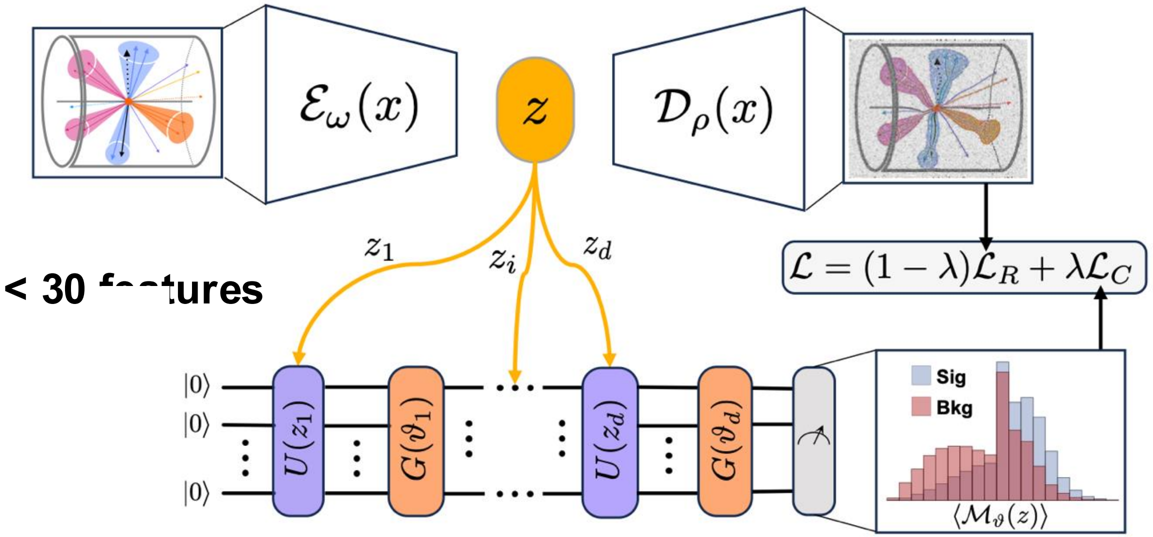
How do we address the limitations of current quantum hardware ?

V. Belis *et al.*, arxiv: 2402.0952

Guided Quantum Compression

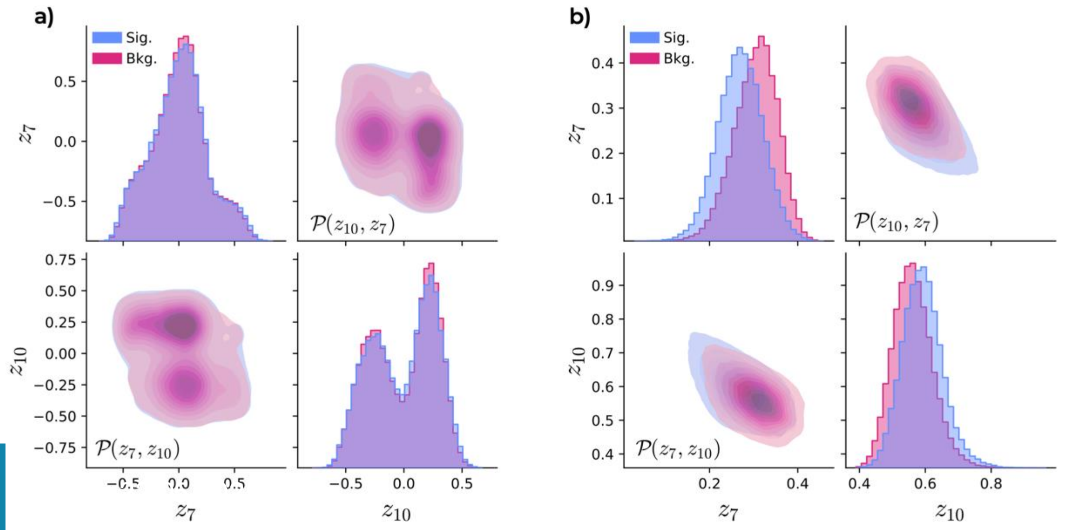


600 features



< 30 features

Two independent steps: Classical preprocessing and quantum classification

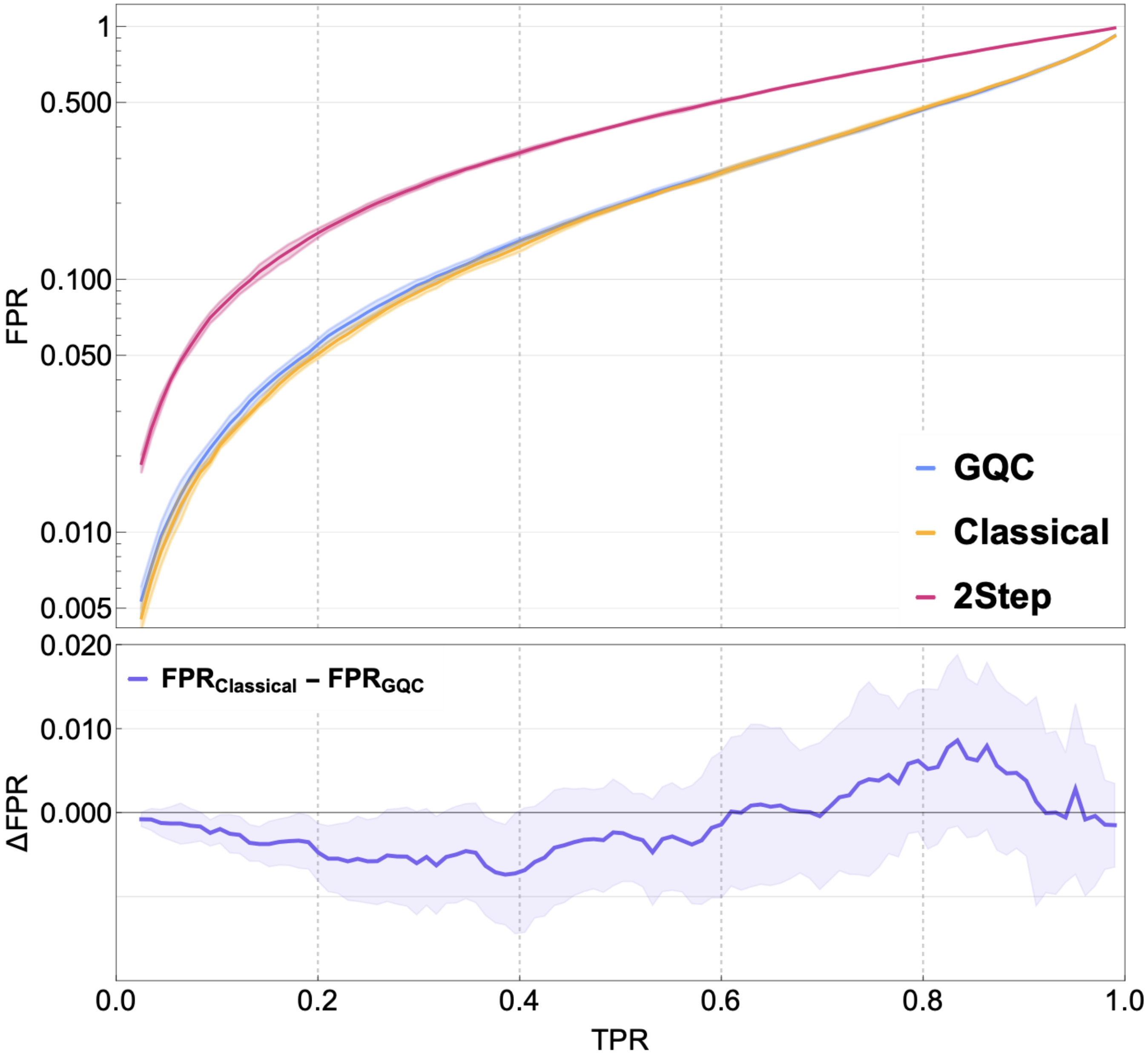


Quantum Guided data compresstion

Result

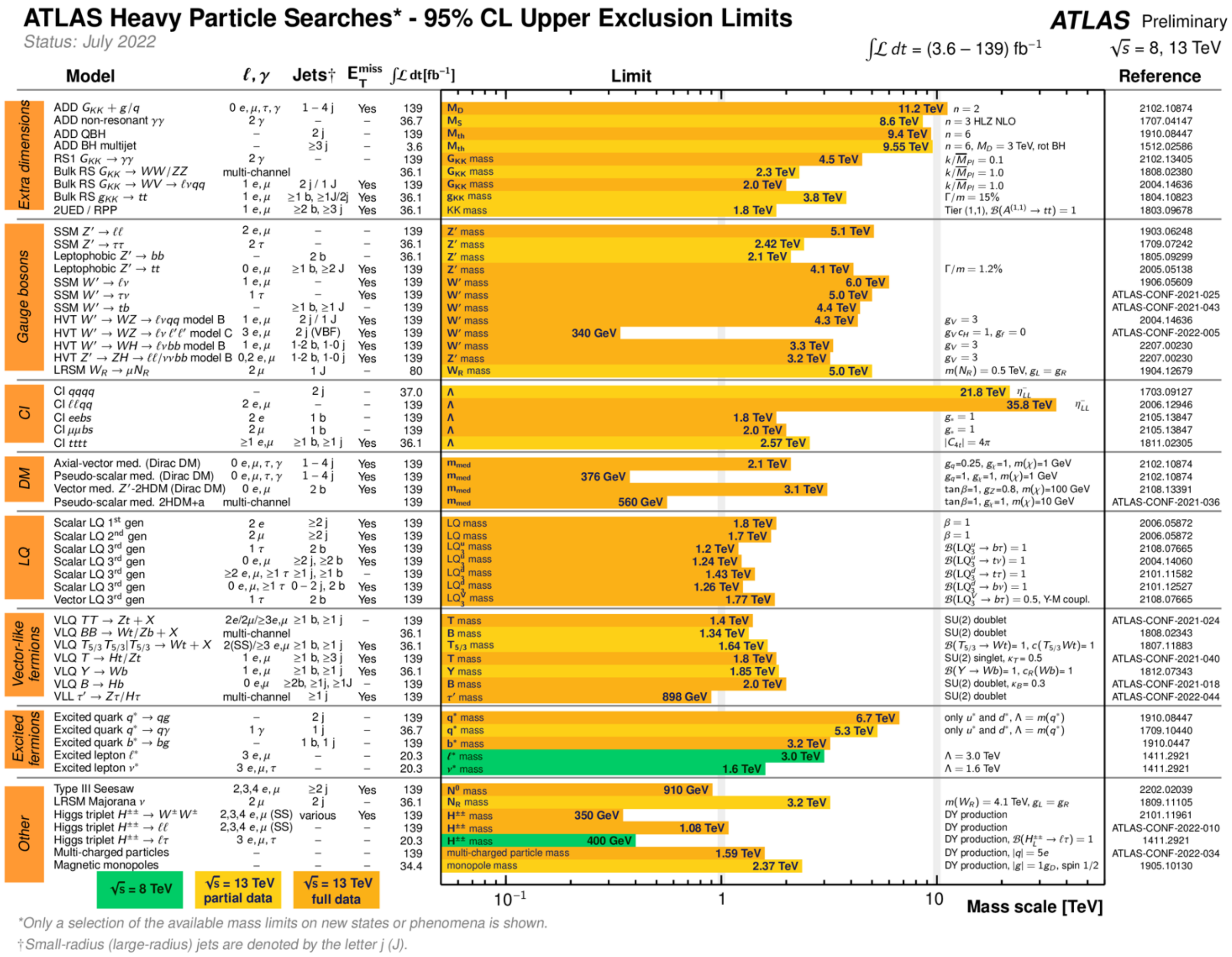
Guided quantum Compression **greatly improves the performance**

We can build efficient hybrid systems for HEP



Does entanglement allow QML to learn more complex distributions?

Uncharted High Energy Frontier



No hints of physics beyond the Standard Model

Most searches focus on specific **theoretical models** …

Uncharted High Energy Frontier

Hunt-Lenox Globe, early XVI, NY Public Library

No hints of physics beyond the Standard Model

Most searches focus on specific **theoretical models** …

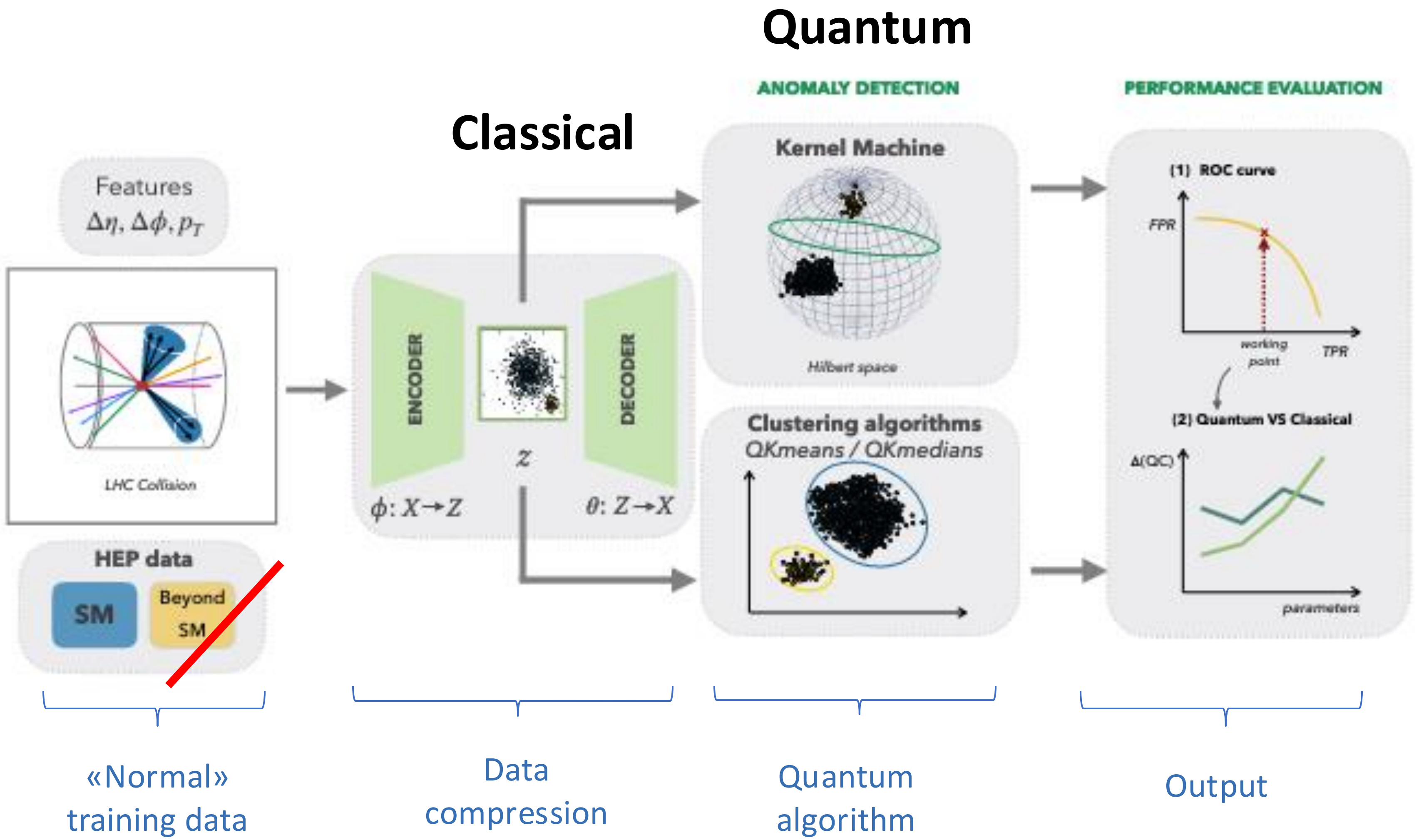
Need to cast the net wider!

… Model Agnostic Anomaly Detection … based on Deep Learning

Anomaly Detection on Quantum Computers

Hybrid implementation:
Use classical data compression

Model Agnostic approach:
- Train using baseline data
- New physics will be flagged as an anomaly



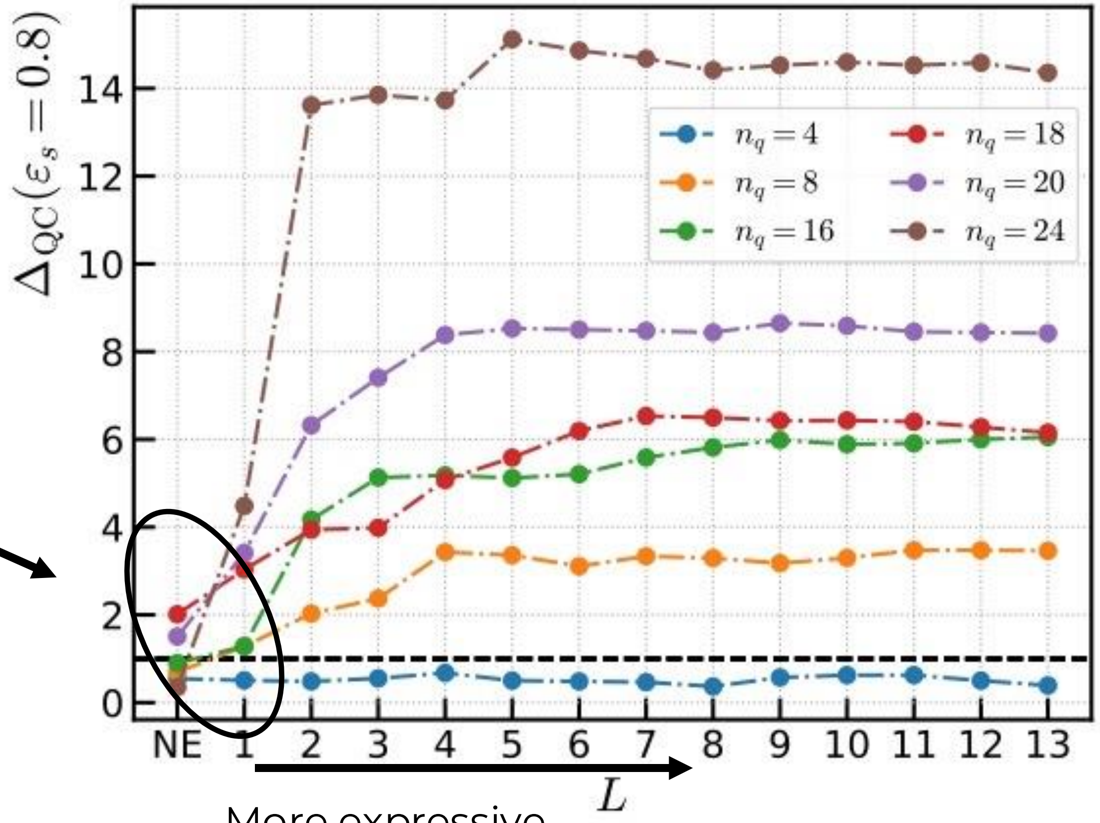
Belis V., GM, et al – *COMMSPHYS-23-1149C*

Performance driven by intrinsically quantum properties!

24 qubits SVM reaches **14x classical model performance**

Quantum/Classical ratio

No entanglement



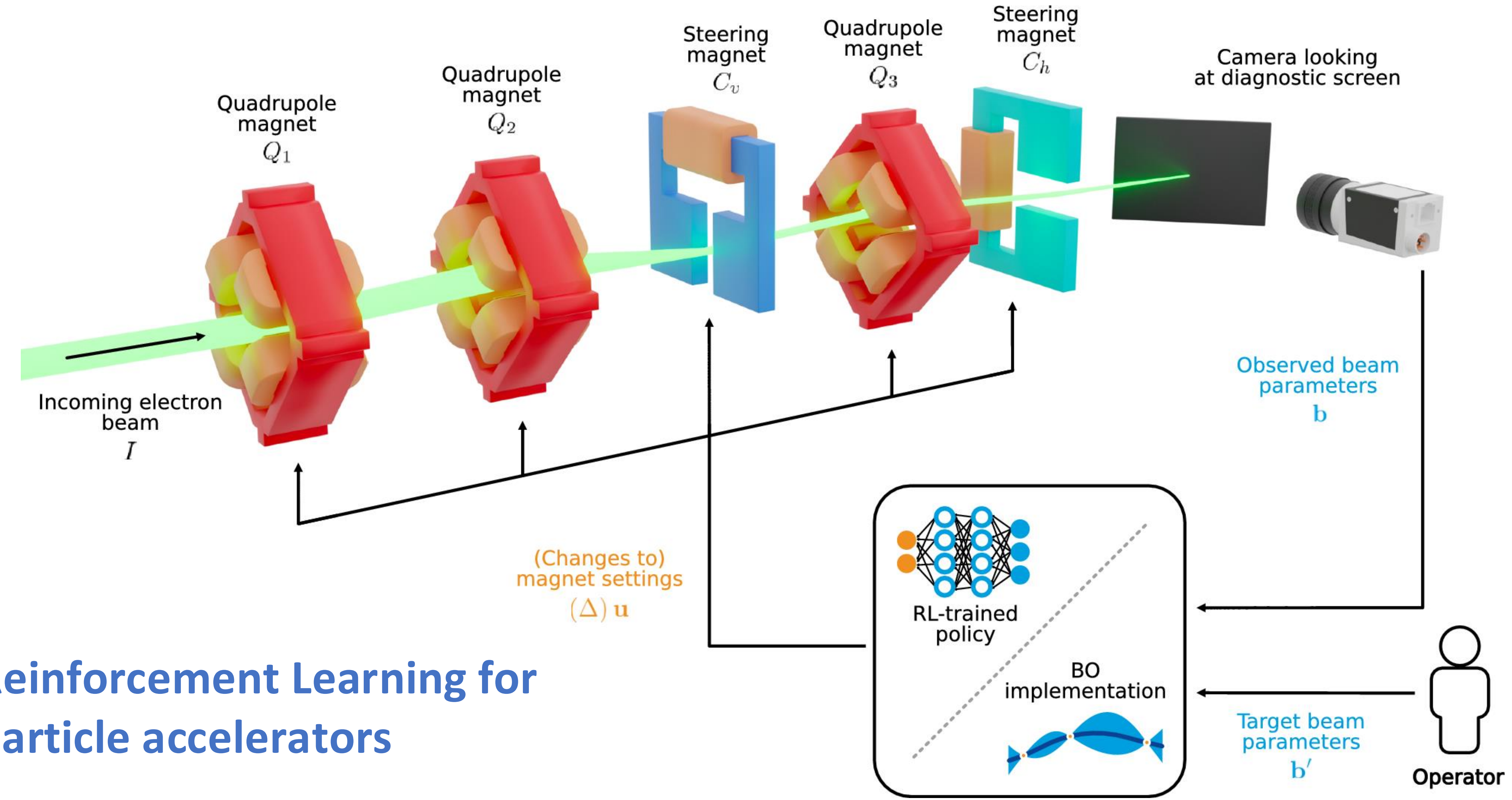
More expressive

This is a simulation. Trend confirmed on IBM Q *Toronto*

Is this evidence for quantum advantage?

Can QML address problems of limited data/resources?

Kaiser, J., Xu, C., Eichler, A. *et al.* **Reinforcement learning-trained optimisers and Bayesian optimisation for online particle accelerator tuning.** *Sci Rep* **14**, 15733 (2024). https://doi.org/10.1038/s41598-024-66263-y



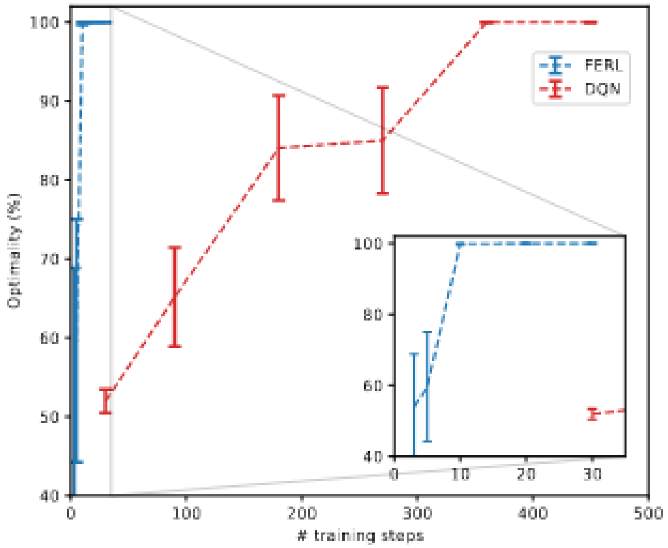
Reinforcement Learning for particle accelerators

Michael Schenk et al., **Hybrid actor-critic algorithm for quantum reinforcement learning at CERN beam lines**, e-Print: 2209.11044, under review «**Quantum Science and Technology**»

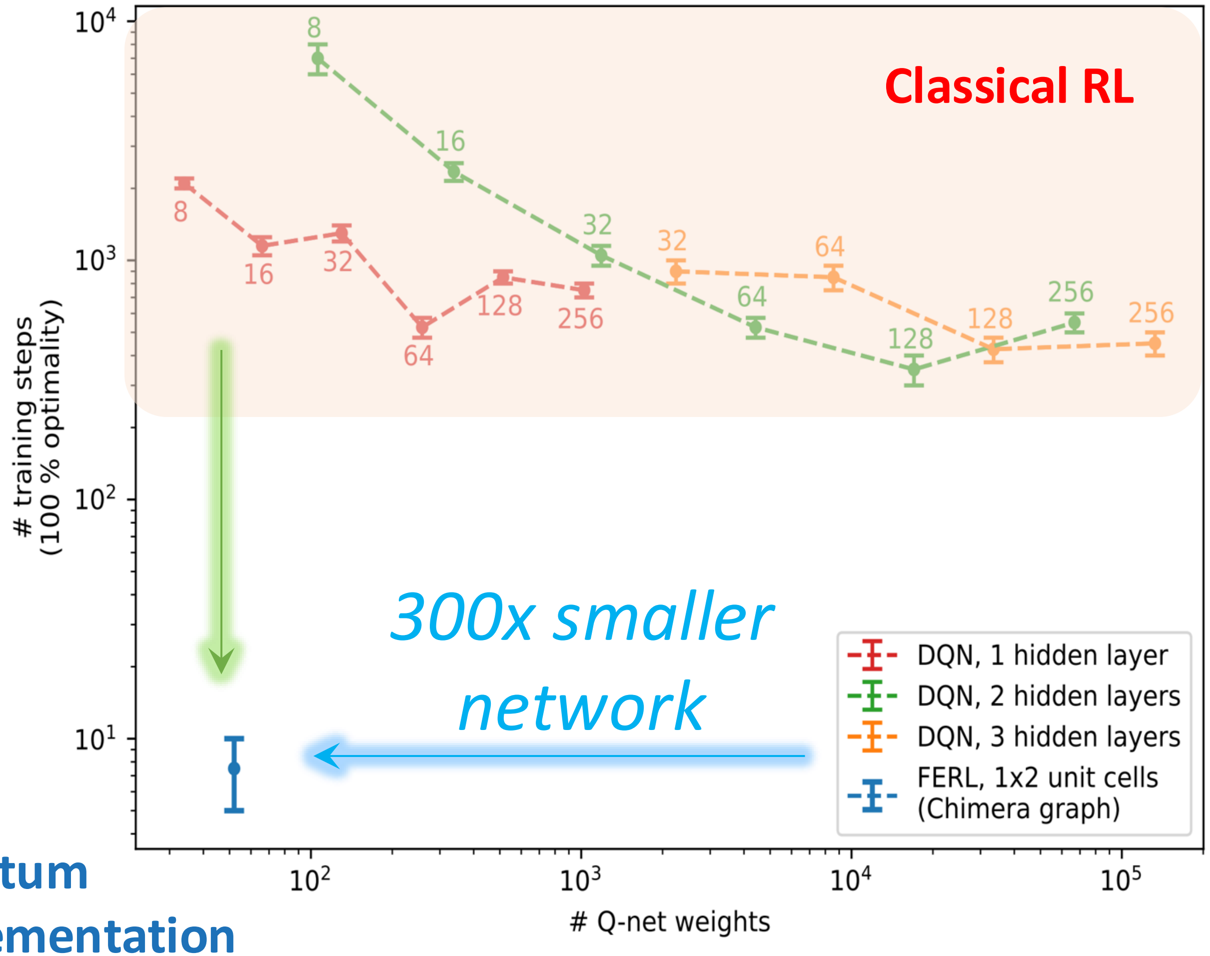
The Quantum Computing Company™

Quantum reinforcement learning

Quantum RL massively outperforms classical algorithm in terms of **model size** and **steps** to convergence



50x fewer training steps

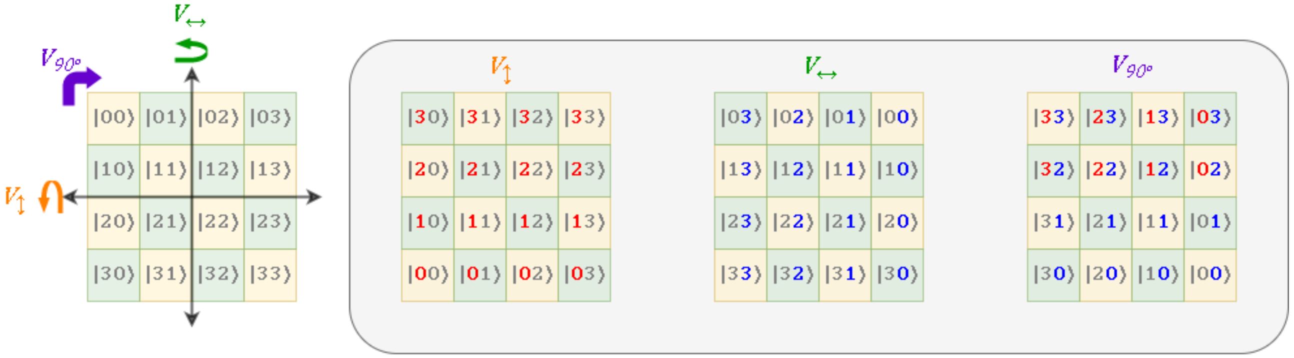


Quantum Implementation

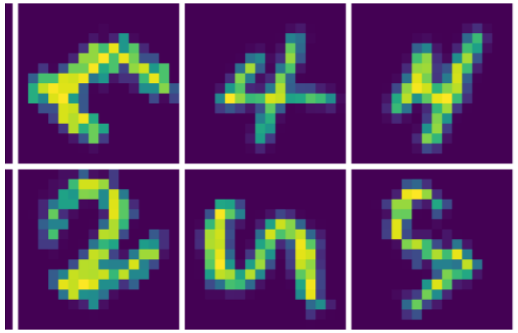
Leveraging symmetries to improve convergence

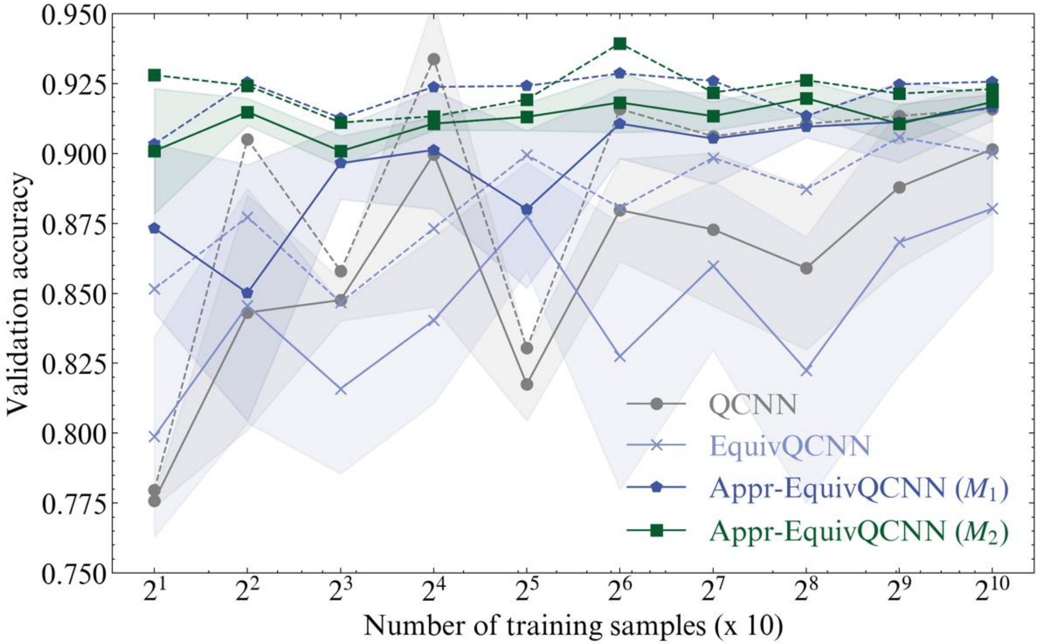
Equivariant Quantum CNN

- Construct **equivariant** quantum CNN **under rotational & reflectional symmetry (p4m)**
- Improved generalization power



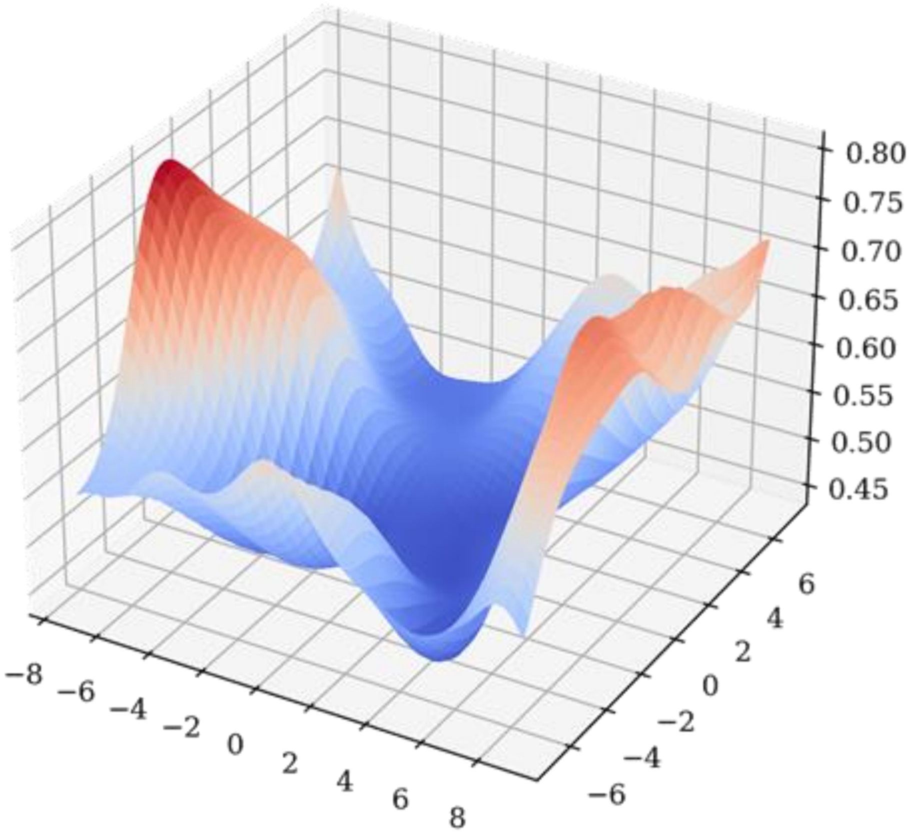
Extended MNIST Image classification:
(digits 4,5)



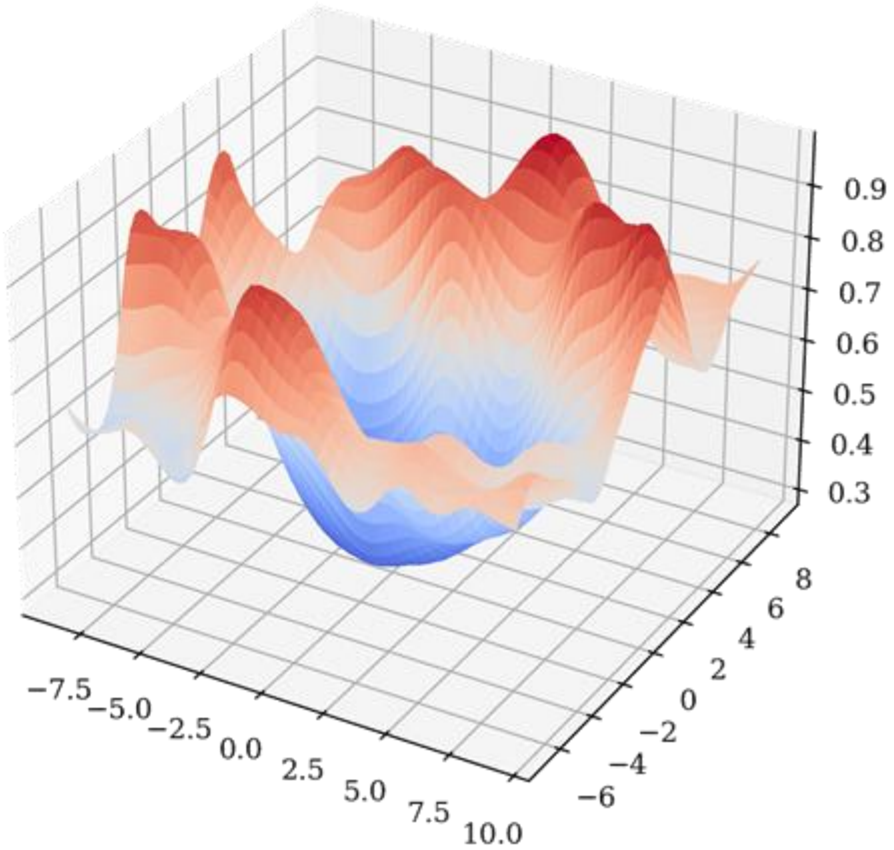


Non-convexity of loss landscape

Loss landscape plotted with orqviz



Non-equivariant QCNN



ApprEquivQCNN

What about running in realistic conditions?

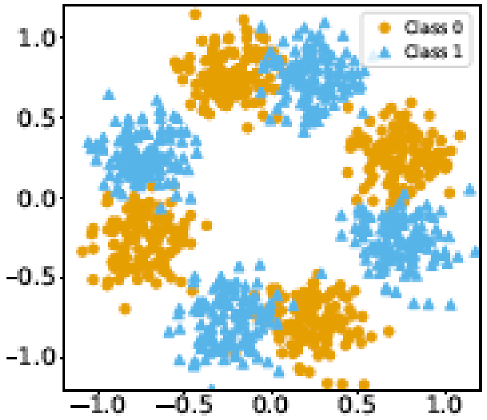
Tüysüz, Cenk, et al. "Symmetry breaking in geometric quantum machine learning in the presence of noise." *arXiv preprint arXiv:2401.10293* (2024).

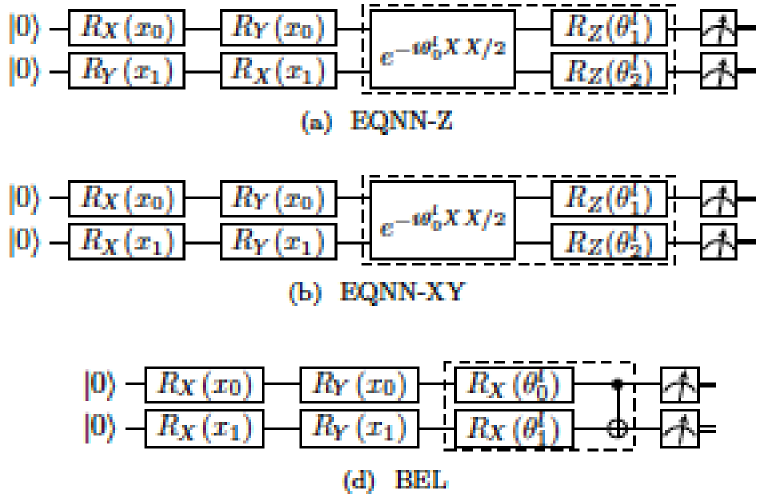
Noise induced symmetry breaking

Noise effects on **EQNN** wrt discrete symmetry groups e.g.

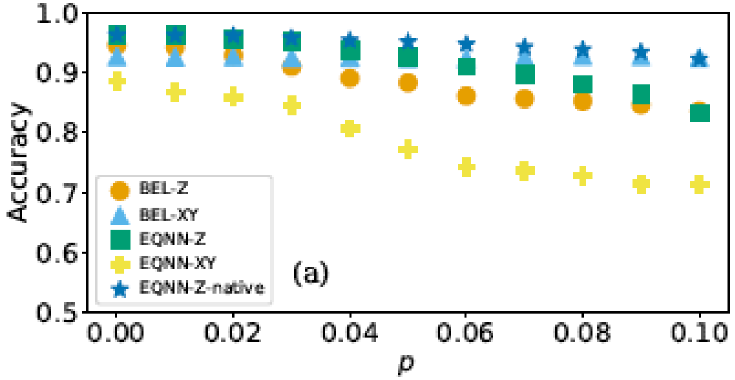
$\boldsymbol{Z_2}$: $R(\sigma)\cdot(x_i) = -x_i$

Bit Flip, Depolarizing (Pauli) and **Amplitude Damping channels**

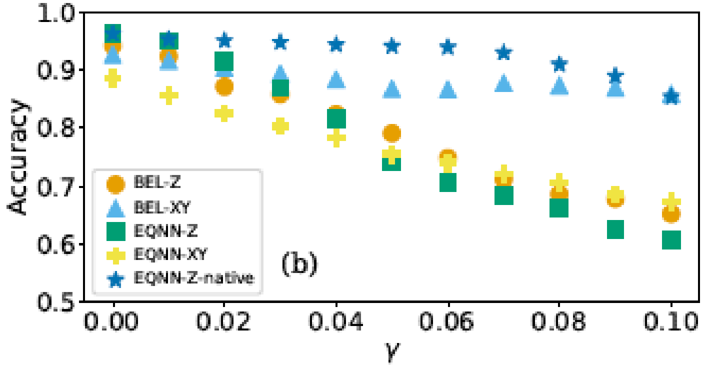




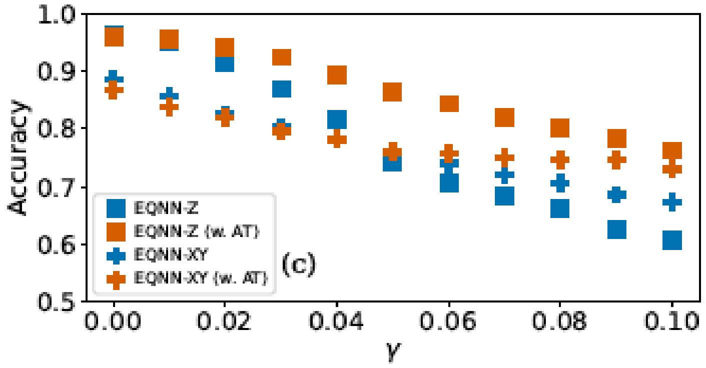
DP should not affect symmetry



EQNN performance drops with AD



Adaptive threshold classification

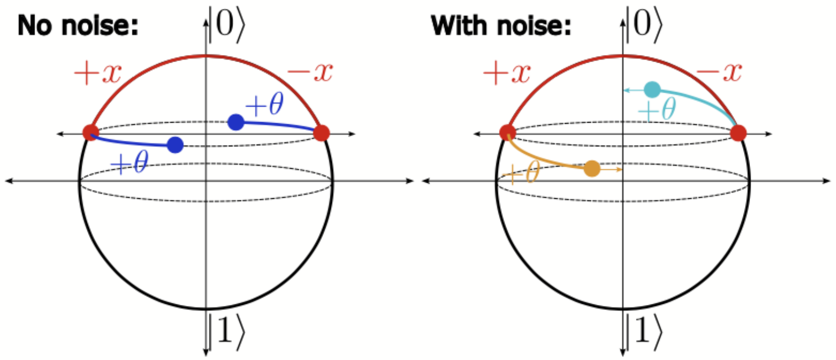


EQNN-Z native: Z_0Z_1 commutes with the AD channel generator, but native gate set is limited on hardware!

Introducing Adaptive Thresholds

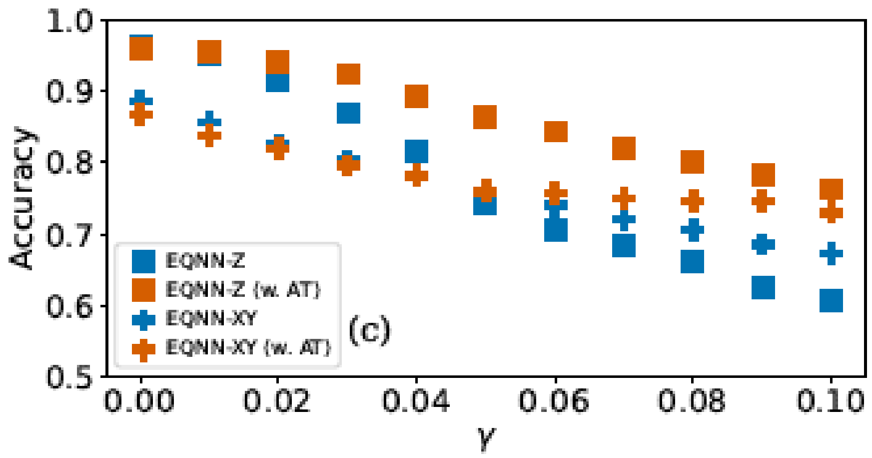
Apart from Noise induced Barren Plateau and exponential concentrations the AD channel exhibits the largest effect on the accuracy performance.

The AD channel shifts the mean of the Z observable: this results in the model having a bias towards one label



Use **adaptive threshold:** computed as the median over the predictions of the training set at every iteration.

Adaptive threshold classification

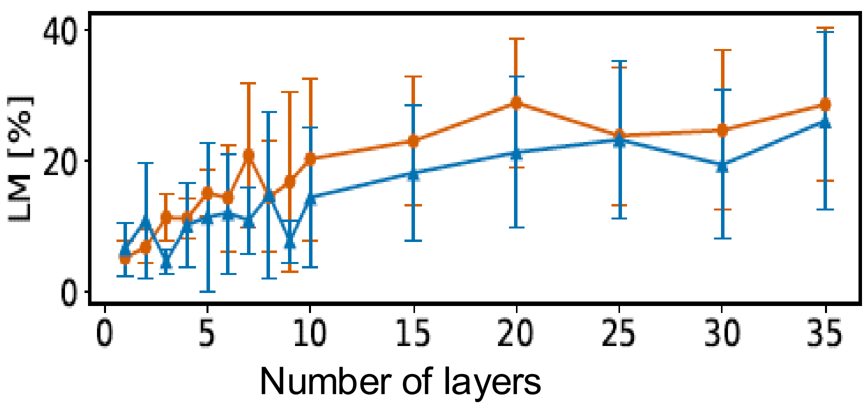


Symmetry breaking on hardware

$$LM = \frac{1}{M}\sum_{i=1}^{M}\frac{(\tau(\hat{y}_i) - \tau(\hat{y}_j))^2}{\tau(\hat{y}_i) + \tau(\hat{y}_j)}$$

Label Misassignment uses adaptive thresholds

ibm_cairo



Tests on *ibm_cairo*

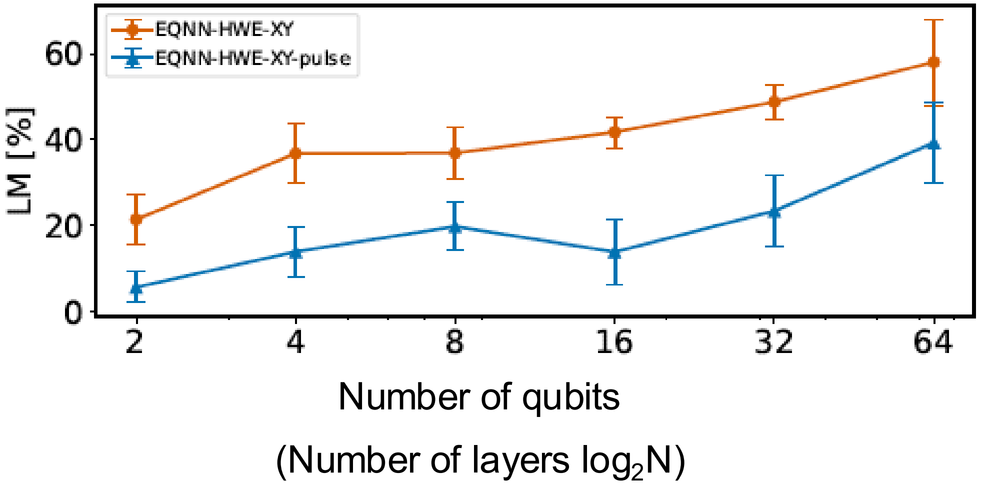
- Confirms AD channel is dominant
- **Symmetry breaking is linear in the number of layers**

Tests on *ibm_cusco* using **hardware efficient ansatz** and **pulse efficient gate** implementation

- create $R_{ZX}(\theta)$ gates by controlling pulses in a continuous way

LM reaches 50% (random) at around 50 qubits

ibm_cairo



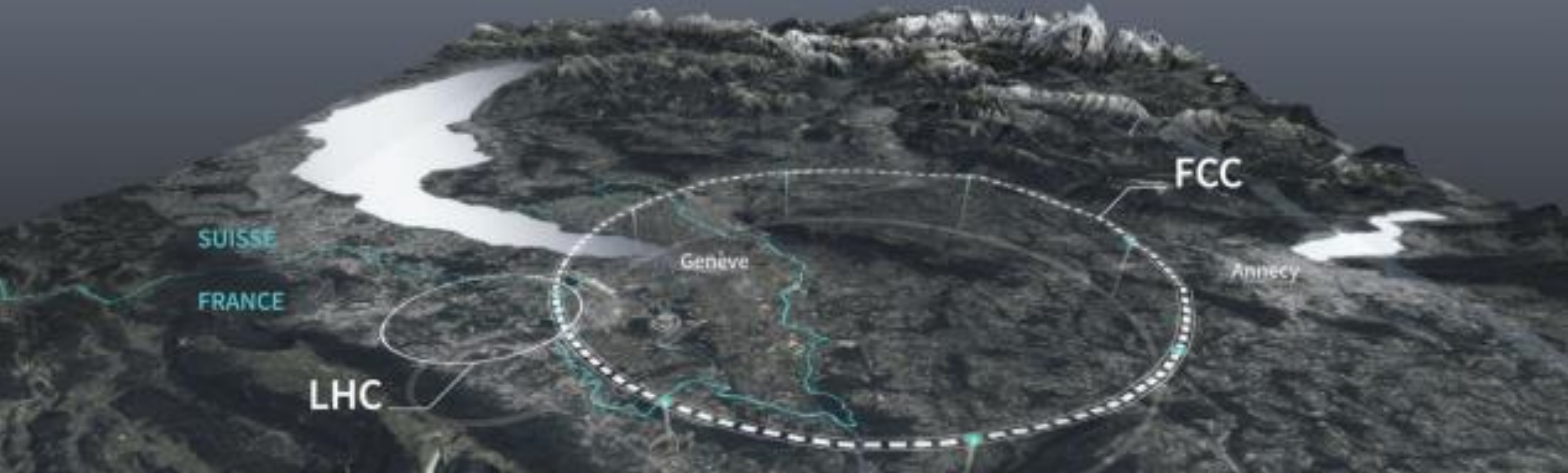
Open questions

- Today's approach to Quantum Machine Learning is variational or kernel based
  - Currently **gradient based optimisation is suboptimal**
  - Can we **train Quantum Machine Learning algorithms effectively**?

- How do we define **advantage** ?
- What is the definition of a fair classical benchmark ?

- Experimental High Energy Physics data has high dimensionality
  - Can **we reduce the impact of data reduction** techniques?
- Experimental High Energy Physics data is shaped by **physics laws**
  - Can we leverage them to build better algorithms?

The Future Circular Collider



CERN is investigating the feasibility of a 91 km circumference collider

Global collaboration: 150 institutes & 30 companies from 34 countries

If approved, it will start operations ≥ 2045
and continue until the end of the century!

Time scale for fault tolerant quantum computing era ?

Thanks !

Geometric Quantum Machine Learning

- Given a data point $x \in \mathcal{X}$ and its label $y \in \mathcal{Y}$
- Estimate the prediction y_θ from observable O: $y_\theta(x) = \langle\psi(x)|\mathcal{U}^\dagger(\theta) O\, \mathcal{U}(\theta)\, |\psi(x)\rangle$
- Given a symmetry group $\mathfrak{G}$ on the data space $\mathcal{X}$
- **$\mathfrak{G}$ – Invariance** : For all $x \in \mathcal{X}$ and $g \in \mathfrak{G}$

$$y_\theta(g[x]) = y_\theta(x)$$

- **Final prediction y_θ is invariant if:**

Equivariant data embedding:

For feature map $\psi: \mathcal{X} \to \mathcal{H}$

$$|\psi(g[x])\rangle = V_s[g]|\psi(x)0\rangle$$

$V_s[g]$ = **Representation** of g on $\mathcal{H}$ induced by ψ

Equivariant ansatz:

For operators generated by a fixed generator G as $R_G(\theta) = \exp(-i\theta G)$:

$$[R_G(\theta), V_s[g]] = 0 \leftrightarrow [G, V_s[g]] = 0$$

Invariant Measurement:

$$V_s^\dagger[g] O V_s[g] = O$$

Gradients decay and Model Convergence

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

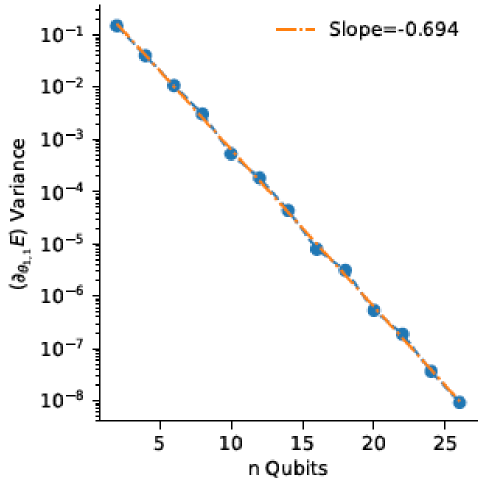
- Convergence still possible if gradients consistent between batches.

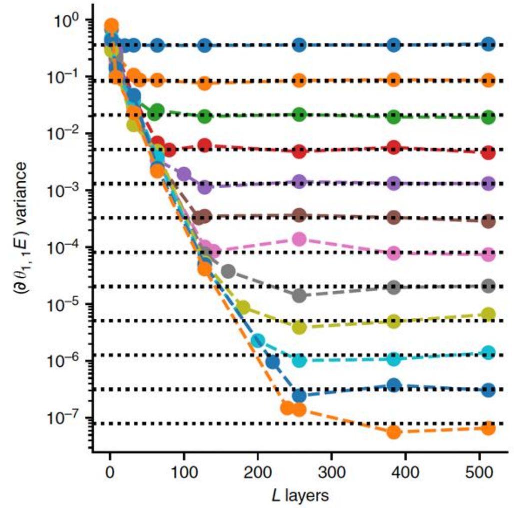
Quantum gradient decay exponentially in the number of qubits (number of graph paths is exponential in the number of gates)

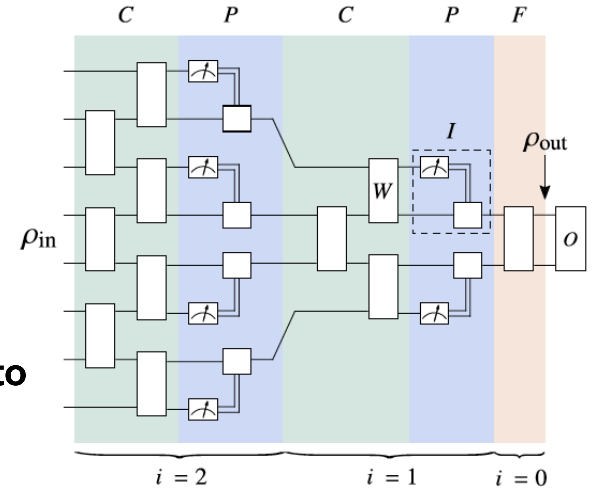
- Random circuit initialization
- Loss function locality in shallow circuits (M. Cerezo *et al.*, arXiv:2001.00550)
- Ansatz choice: TTN, CNN (Zhang *et al.*, arXiv:2011.06258, A Pesah, *et al.*, *Physical Review X* 11.4 (2021): 041011.)
- Noise induced barren plateau (Wang, S *et al.*, Nat Commun 12,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

Large number of measurements: $1/\epsilon^2$ measurements to estimate a cost to precision ϵ

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







Quantum Convolutional NN:
Convolution: general SU(4)
Pooling: reduces numer of qubits

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

See C. Rieger's summer students lecture

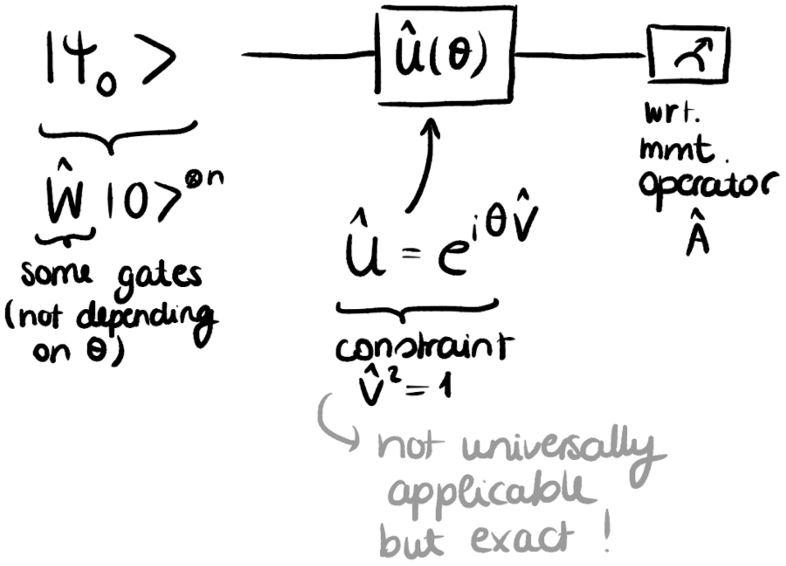
Parameter optimization

$$\theta \rightarrow \theta - \eta \nabla_\theta f$$

$$\uparrow \langle \hat{A}(\theta) \rangle$$

The parameter-shift rule (gradient-based)

→ Compute **partial derivative** of variational circuit parameter θ, alternative to analytical gradient computation and classical finite difference rule (numerical errors and resource cost considerations)



$$\Rightarrow \nabla_\theta \langle \hat{A} \rangle = u\left[\langle \hat{A}(\theta + \tfrac{\pi}{4u}) \rangle - \langle \hat{A}(\theta - \tfrac{\pi}{4u}) \rangle\right]$$

→ Evaluate Quantum Circuit twice at shifted parameters to compute gradient

Source:https://pennylane.ai/qml/demos/tutorial_stochastic_parameter_shift/