Quantum Machine Learning

Opportunities and challenges in HEP

CERN | QUANTUM TECHNOLOGY INITIATIVE

Sofia Vallecorsa

CERN QTI Coordinator

Agenda

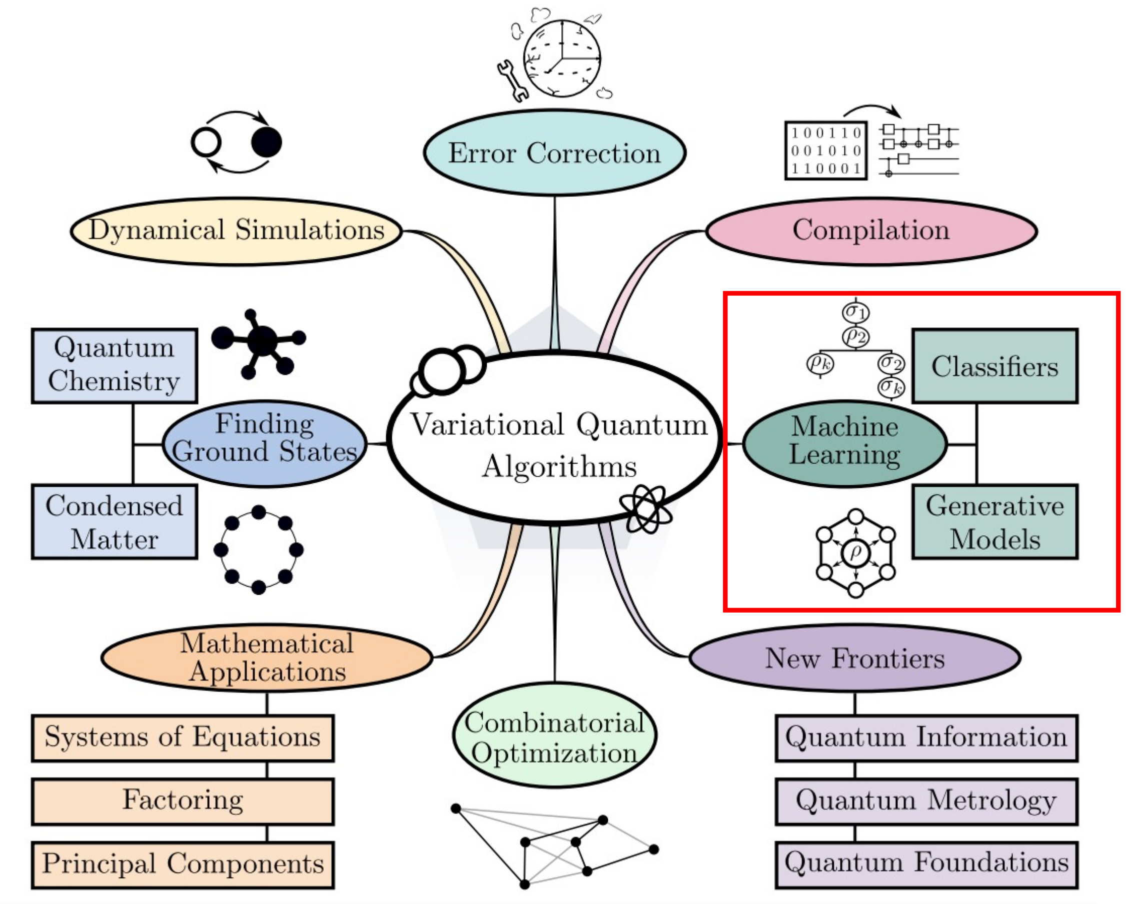
- Part 1: Quantum Computing for Machine Learning
- Part 2: Quantum Machine Learning for HEP

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 Lyfecycle

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

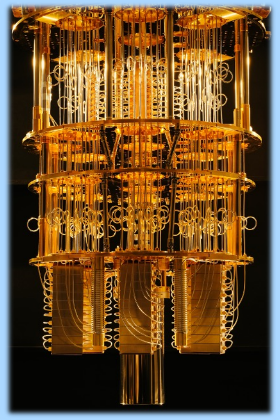
Data Preparation

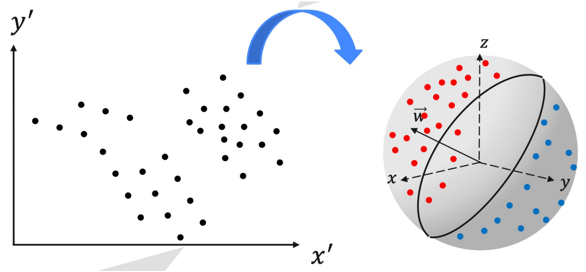
Model Definition

Model Training

Model Testing

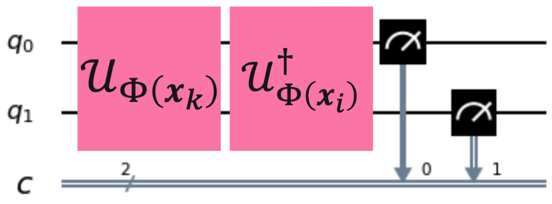
Model Interpretation



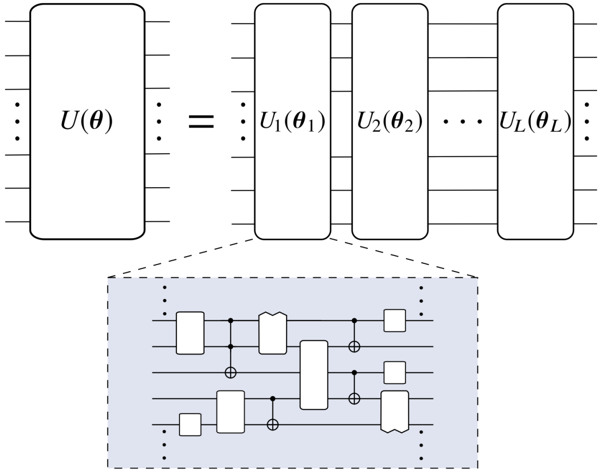


Data Reduction
Data Encoding [1,2,3]

Read Out



Trainability (BP…)



[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

Models

Variational algorithms (ex. QNN)

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

Data Embedding can be **learned**

Ansatz design can leverage data symmetries[1]

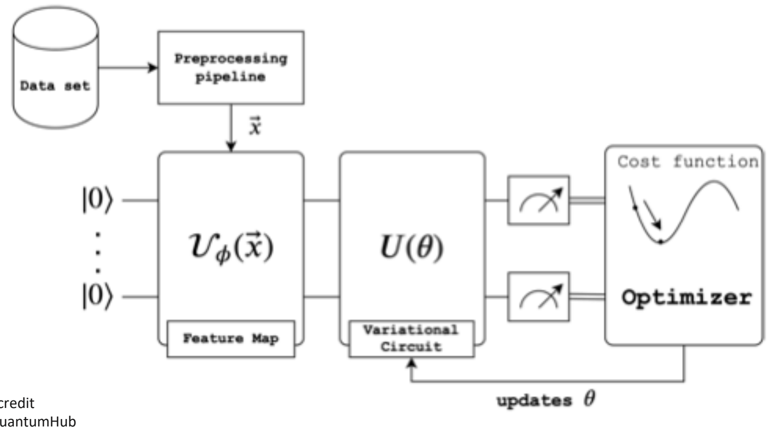


Image credit SwissQuantumHub

Representer theorem:

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

Explicit models exhibit **better generalization**

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]

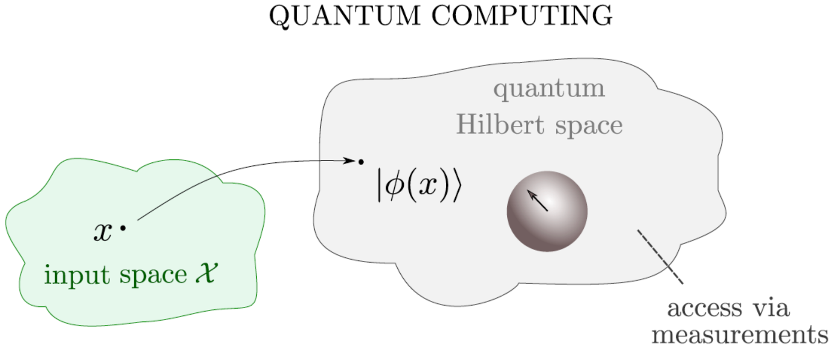


Image credit M. Schuld

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).

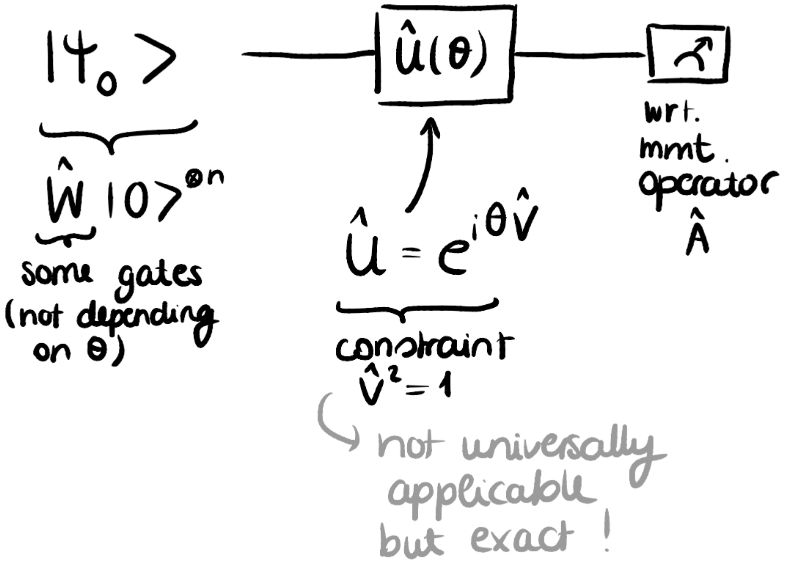
Parameter optimization

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

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

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/

See C. Rieger's summer students lecture

Parameter optimization

$$\theta_{k+1} \leftarrow \theta_k - a_k \hat{g}_k(\hat{\theta}_k)$$

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

$$\text{"} \; y(\theta) = f(\theta) + \varepsilon \; \text{"}$$

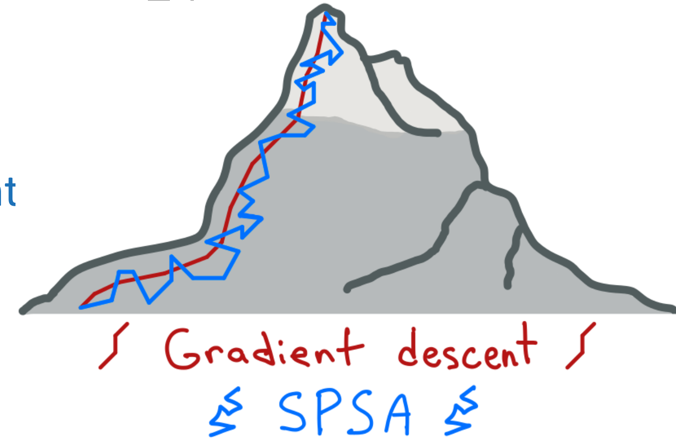
ε: random output perturbation

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

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

Iterative update rule comparable to classical stochastic gradient descent

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



Challenges when using Parametrized Quantum Circuits

- Efficient **data handling** and data **embedding**
- Find balance: **Generalization** and **representational power** vs. **Convergence**
  - 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
- ….

Gradients decay and Model Convergence

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

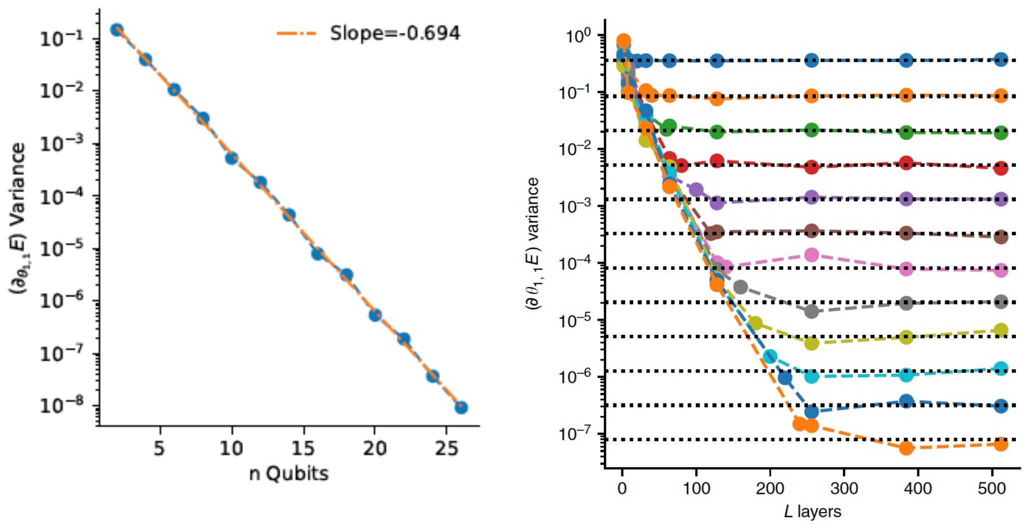
- Convergence still possible if gradients consistent between batches.

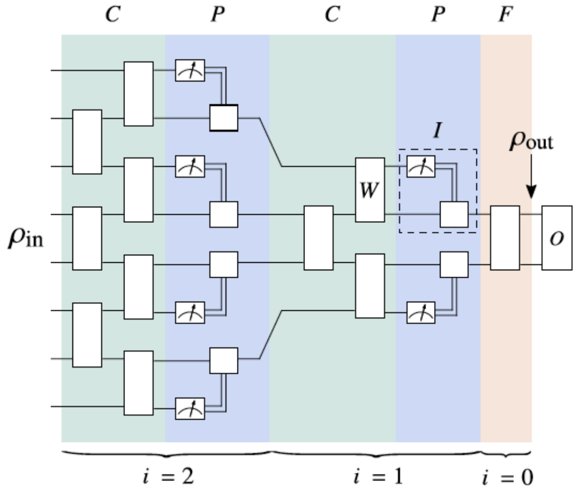
Quantum gradient decay exponentially in the number of qubits (number of graph paths is exponential in the number of gates)

- 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



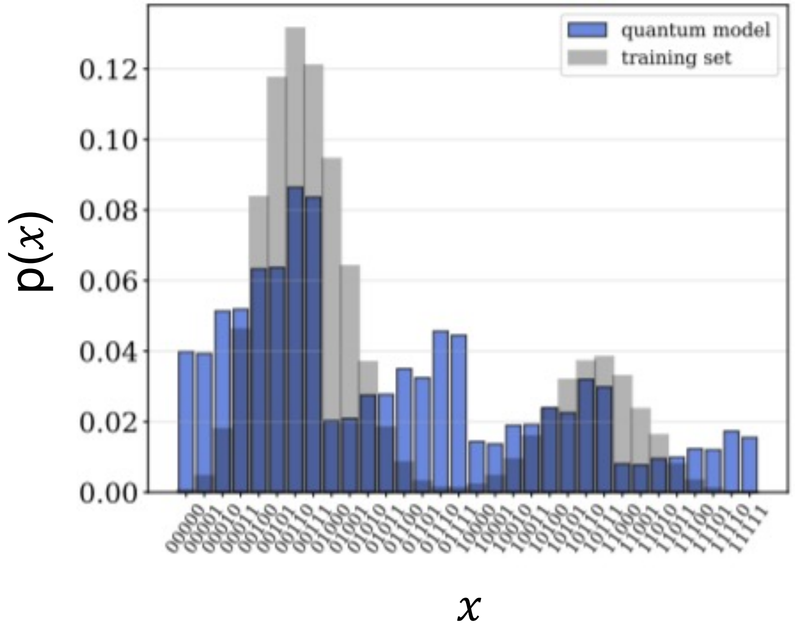


Quantum Convolutional NN:
Convolution: general SU(4)
Pooling: reduces numer of qubits

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

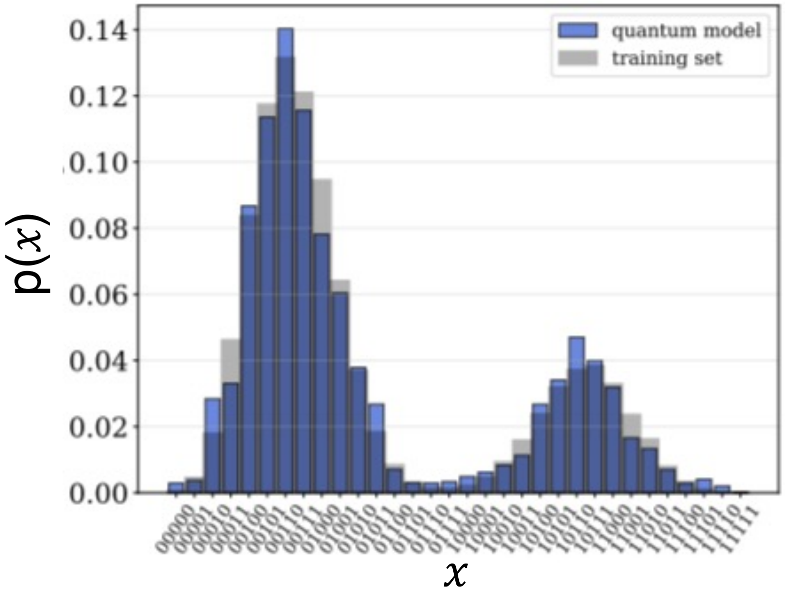
Generative Models

Learn probability distribution that best describes a data set



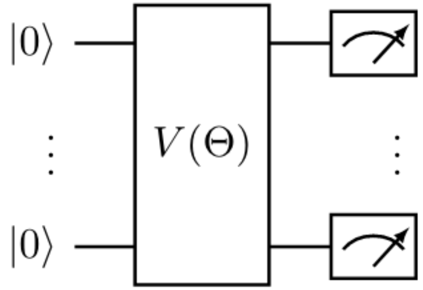
Deeper circuits learn better representations..

Or don't they ??



Quantum Circuit Born Machine

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



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

Implicit and Explicit Models

Classified according to whether or not they **have access to the propability distribution function**

Explicit Models have access to PDF in polynomial time

- Use explicit losses that are defined by probabilities
- Ex. TN or autoregressive models

Implicit models do not have access to PDF. Can sample from it

- Use implicit losses built on samples
- Ex. GAN, QBM, VAE… QCBM…

Explicit

$$\sum_{\boldsymbol{x}} f\left(\tilde{p}(\boldsymbol{x}), \tilde{q}_{\boldsymbol{\theta}}(\boldsymbol{x})\right)$$

Ex. KL Divergence

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

Implicit

$$\underset{\boldsymbol{x},\boldsymbol{y}}{\mathbb{E}}\left[g(\boldsymbol{x}, \boldsymbol{y})\right]$$

Ex. MMD

Strong impact on trainability for generic models

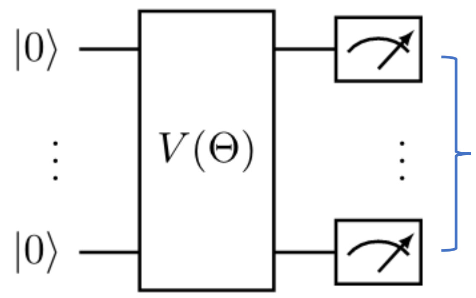
$$\mathrm{MMD}(\mathbb{P}_r, \mathbb{P}_g) = \left(\mathbb{E}_{\substack{\mathbf{x}_r, \mathbf{x}'_r \sim \mathbb{P}_r, \\ \mathbf{x}_g, \mathbf{x}'_g \sim \mathbb{P}_g}}\left[k(\mathbf{x}_r, \mathbf{x}'_r) - 2k(\mathbf{x}_r, \mathbf{x}_g) + k(\mathbf{x}_g, \mathbf{x}'_g)\right]\right)^{\frac{1}{2}}$$

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

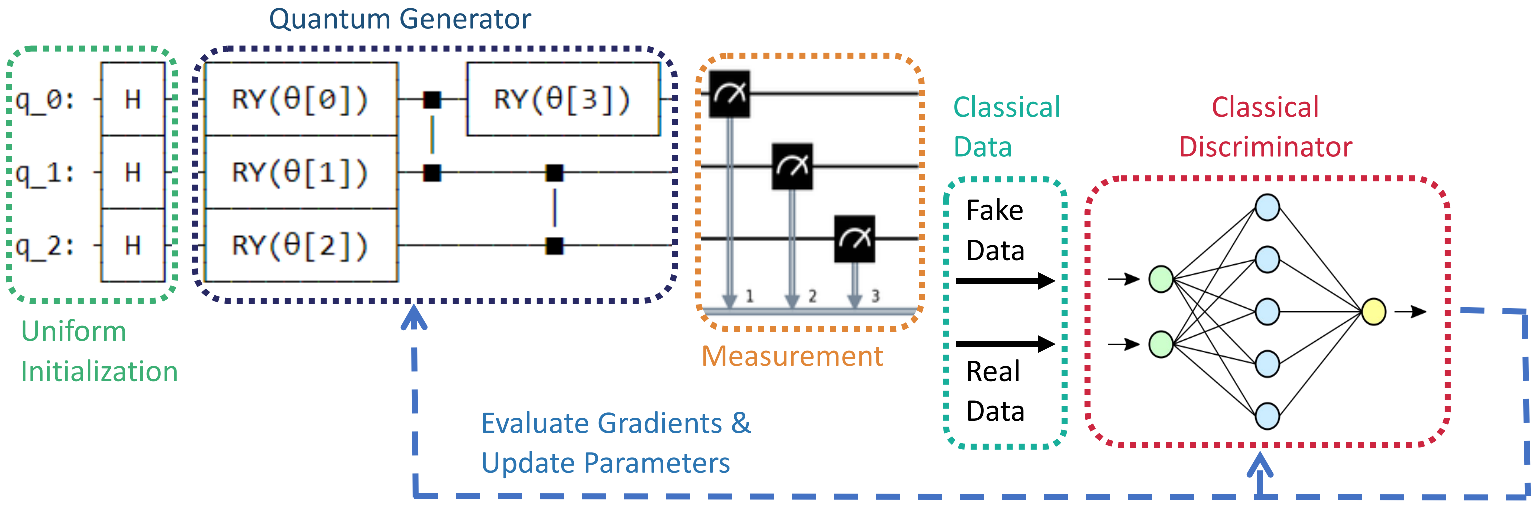
Sample variational pure state $|\psi(\theta)\rangle$ by projective measurement through **Born rule**: $\mathbf{p}_\theta(\mathbf{x}) = |\langle \mathbf{x}|\boldsymbol{\psi}(\boldsymbol{\theta})\rangle|^2$.



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)

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

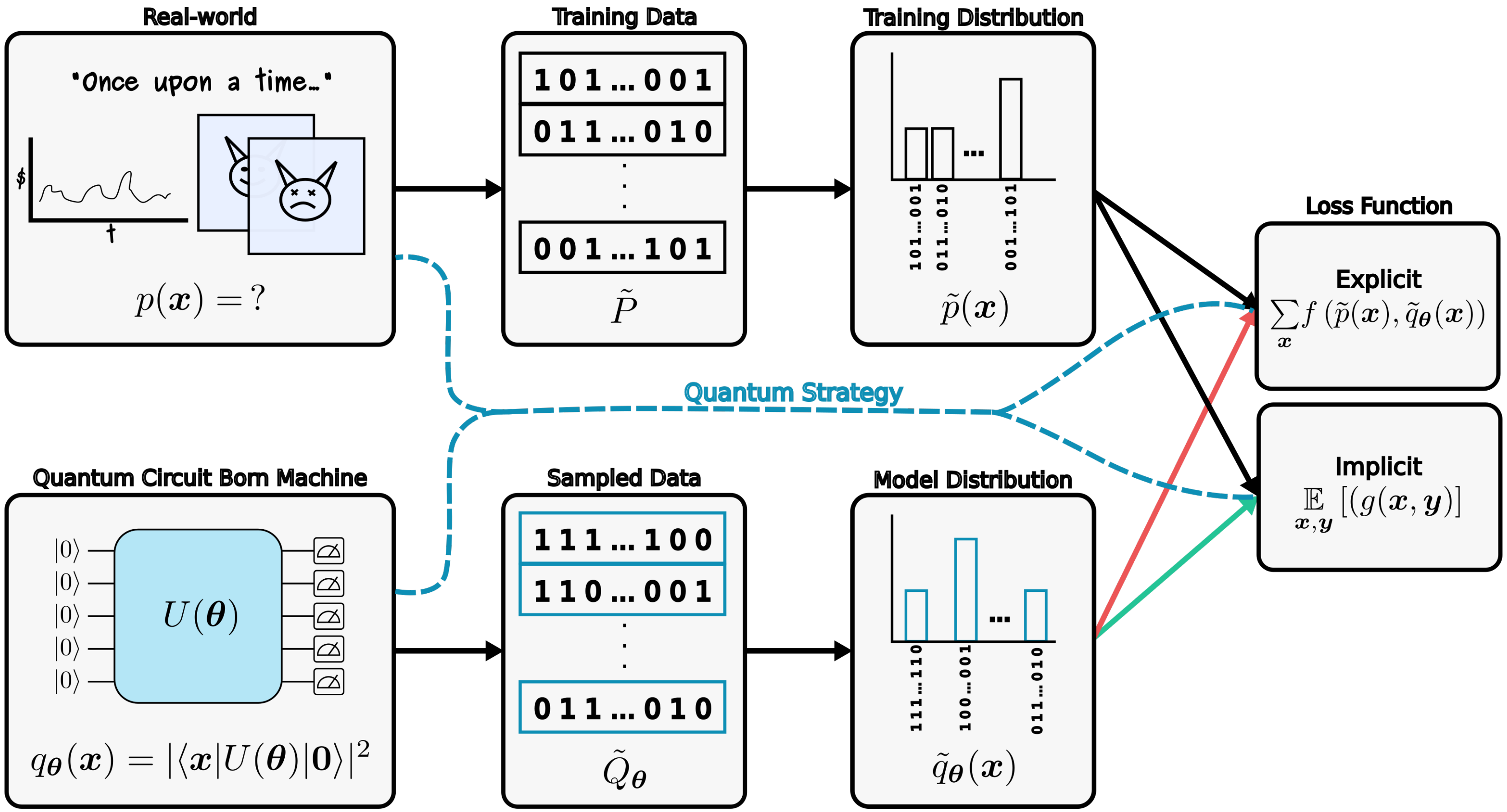
Typical metrics:

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

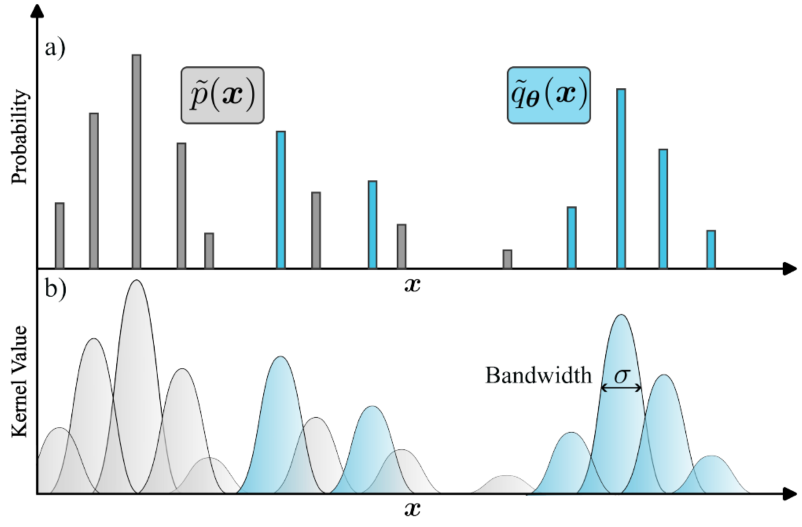
$$\mathrm{MMD}(\mathbb{P}_r, \mathbb{P}_g) = \left(\mathbb{E}_{\substack{\mathbf{x}_r, \mathbf{x}_r' \sim \mathbb{P}_r, \\ \mathbf{x}_g, \mathbf{x}_g' \sim \mathbb{P}_g}} \left[k(\mathbf{x}_r, \mathbf{x}_r') - 2k(\mathbf{x}_r, \mathbf{x}_g) + k(\mathbf{x}_g, \mathbf{x}_g')\right]\right)^{\frac{1}{2}}$$

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

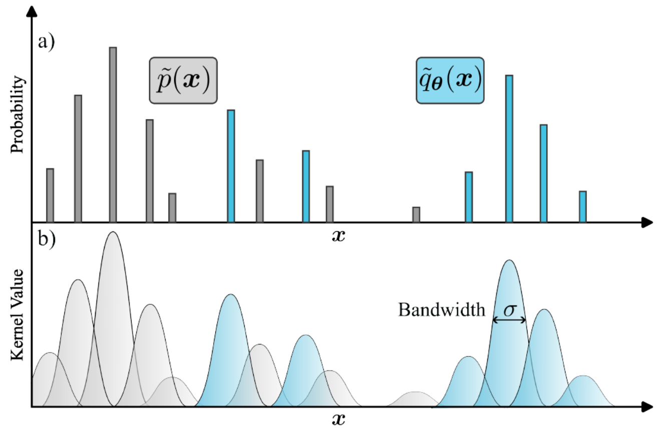
Generative QML and trainability barriers



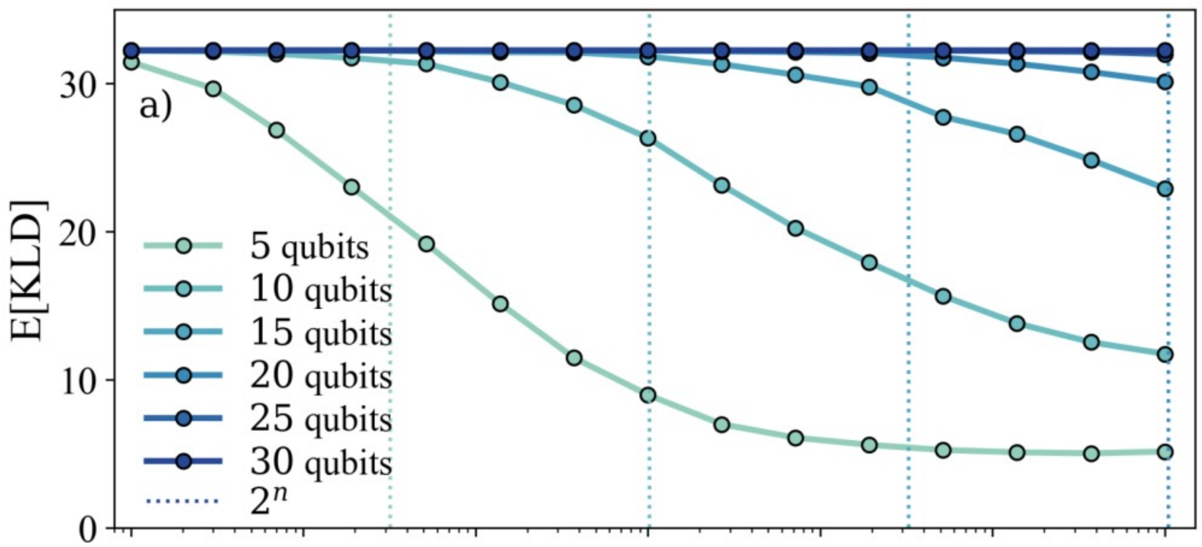
exponentially larger number of shots is required to keep accuracy of explicit losses



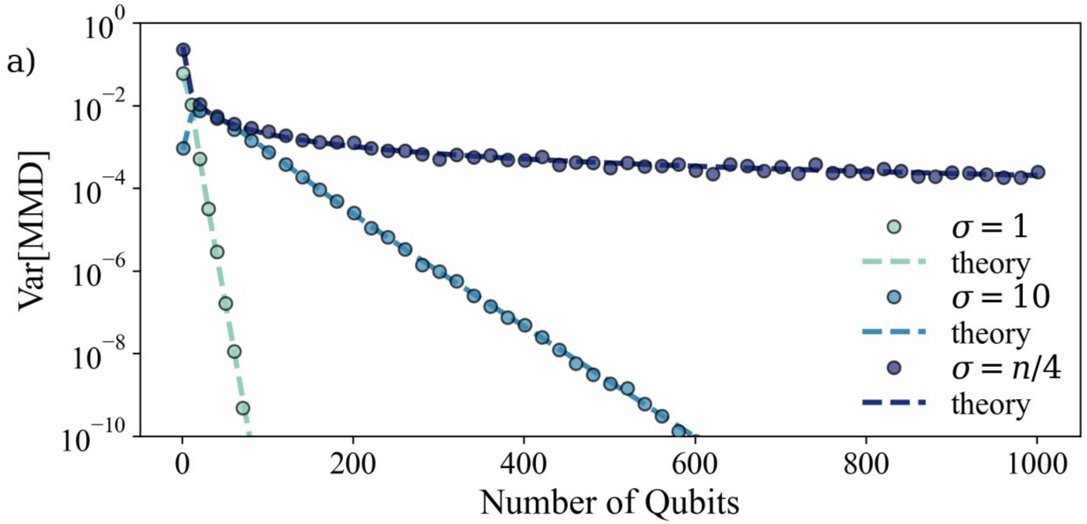
Trainability



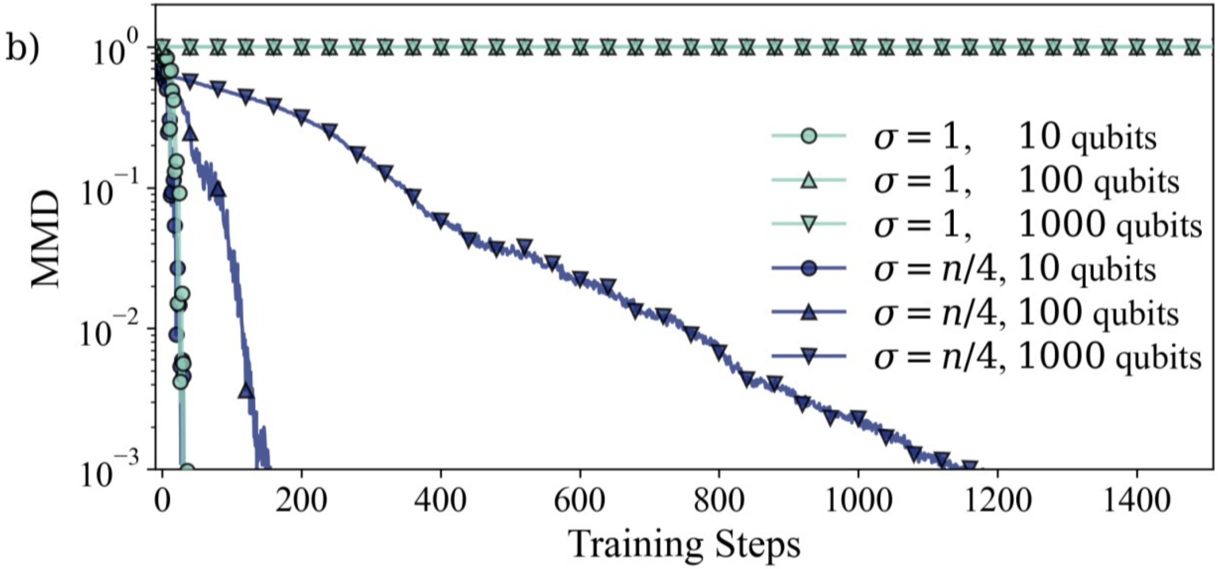
KL: Training loss



MMD: Exact Loss Variance



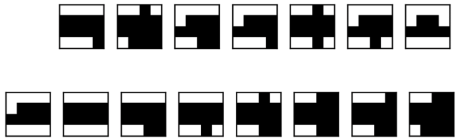
MMD: Training loss

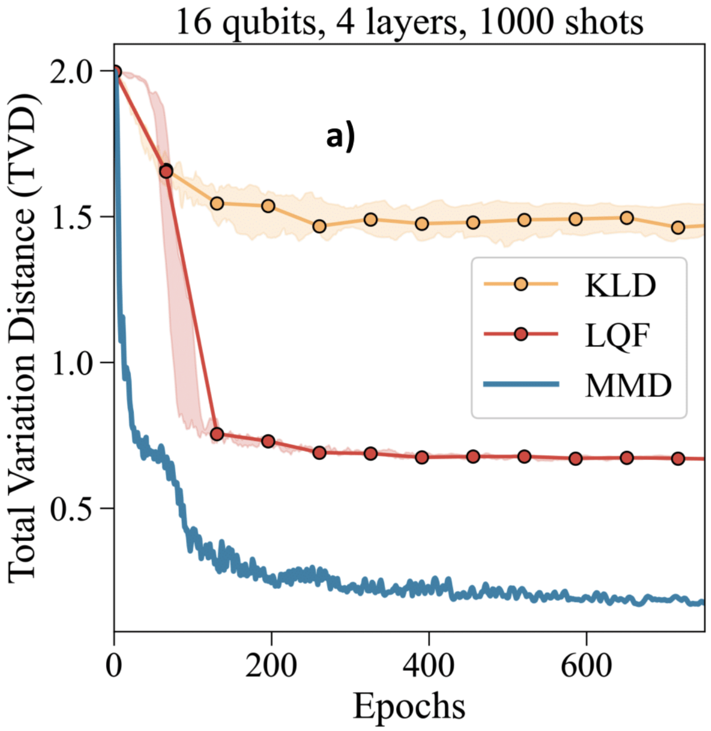


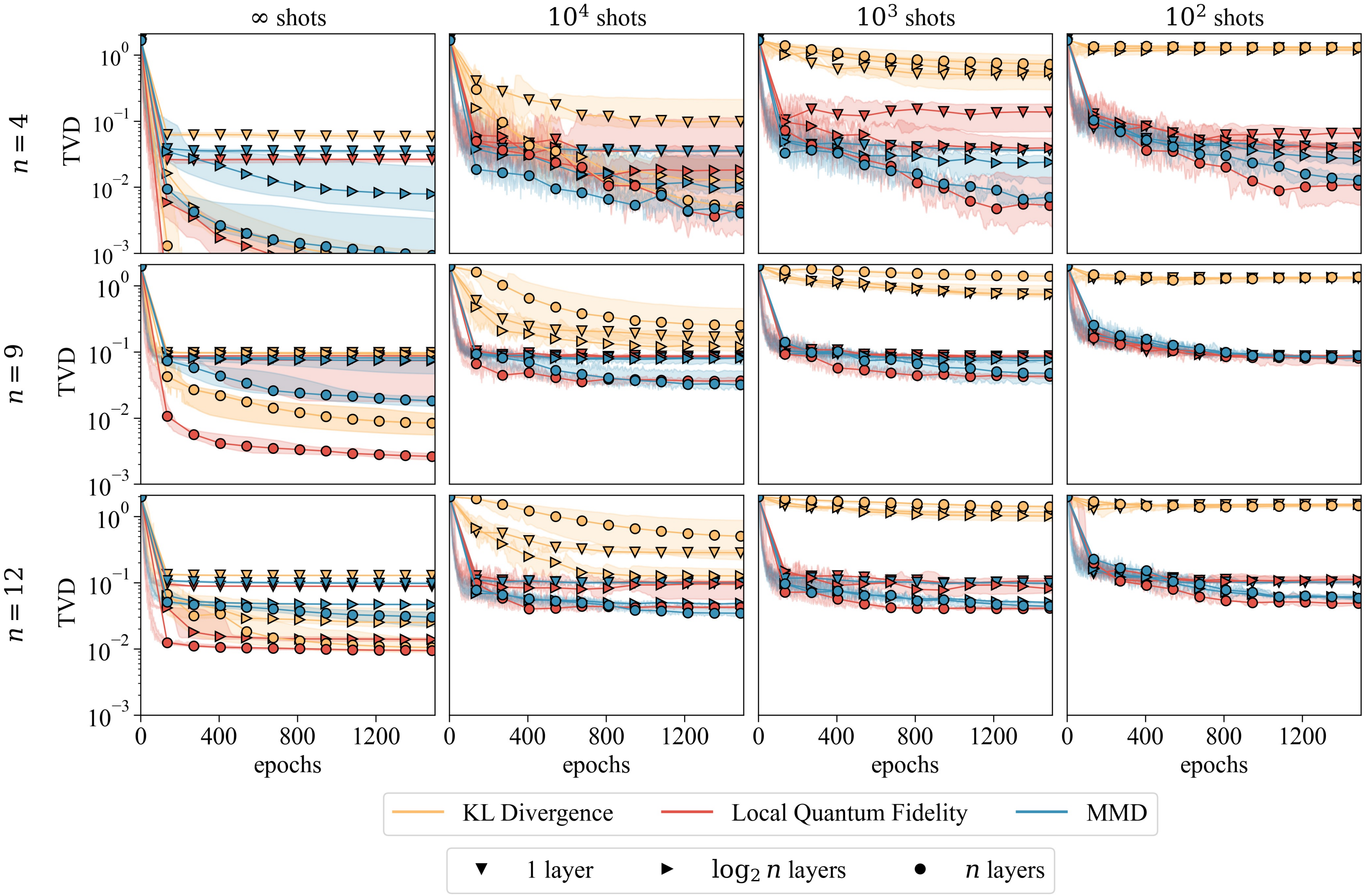
Rudolph, M. S., Lerch, S., Thanasilp, S., Kiss, O., Vallecorsa, S., Grossi, M., & Holmes, Z. (2023). **Trainability barriers and opportunities in quantum generative modeling.** *arXiv:2305.02881*.

Benchmark: QCBM for energy depositions

Sample pictures:



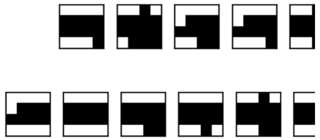


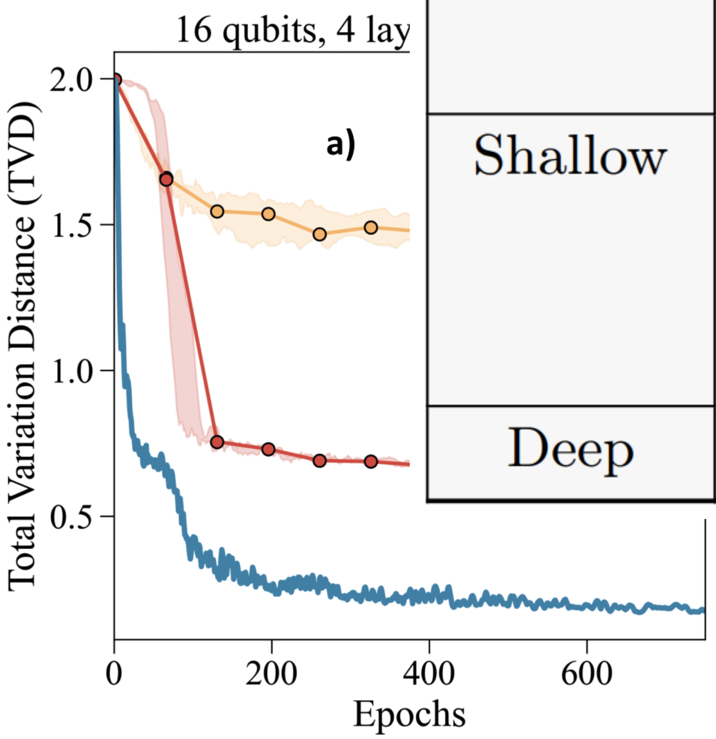


Rudolph, M. S., Lerch, S., Thanasilp, S., Kiss, O., Vallecorsa, S., Grossi, M., & Holmes, Z. (2023). **Trainability barriers and opportunities in quantum generative modeling.** *arXiv:2305.02881*.

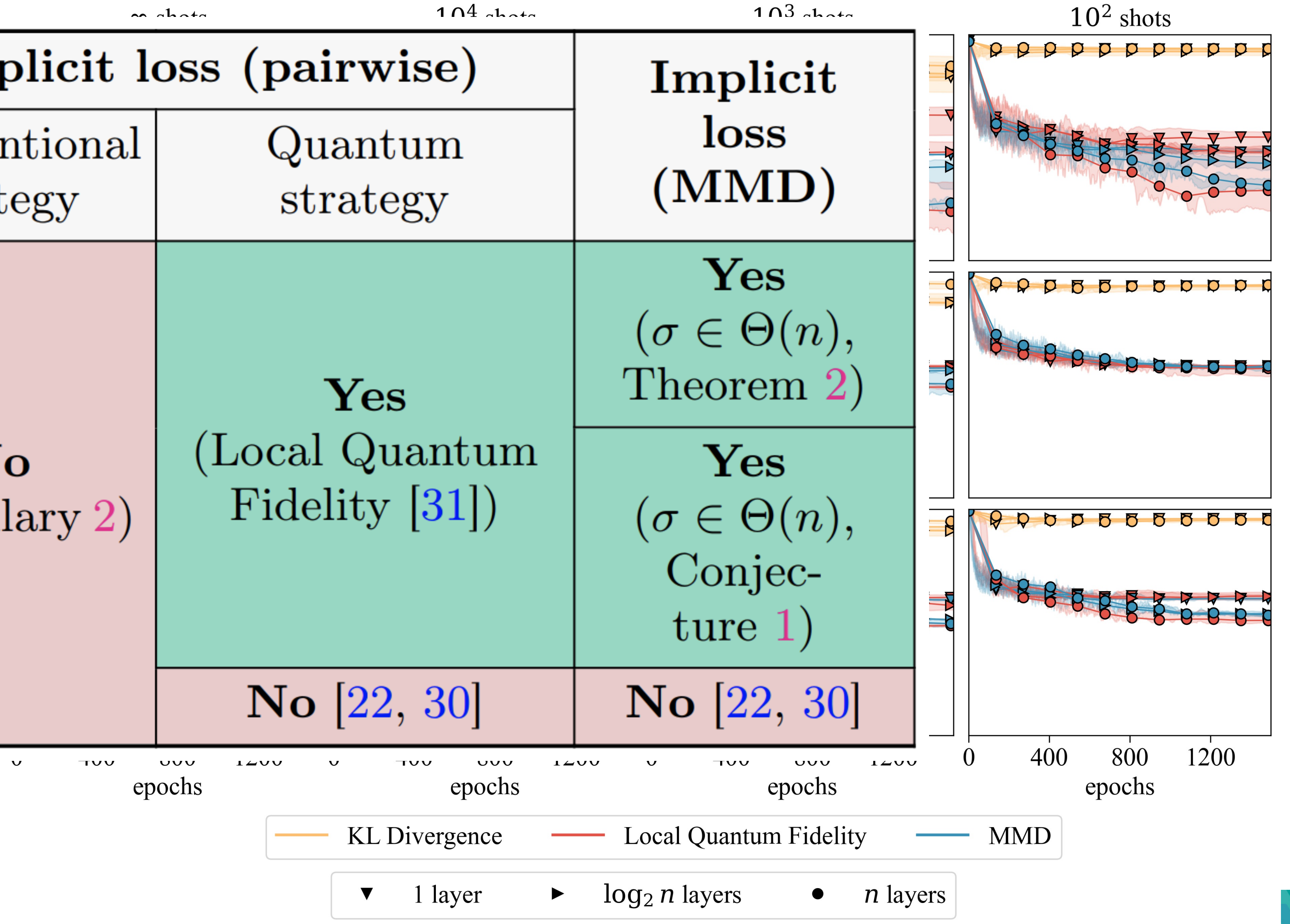
Benchmark: QCBM for energy depositions

Sample pictures:





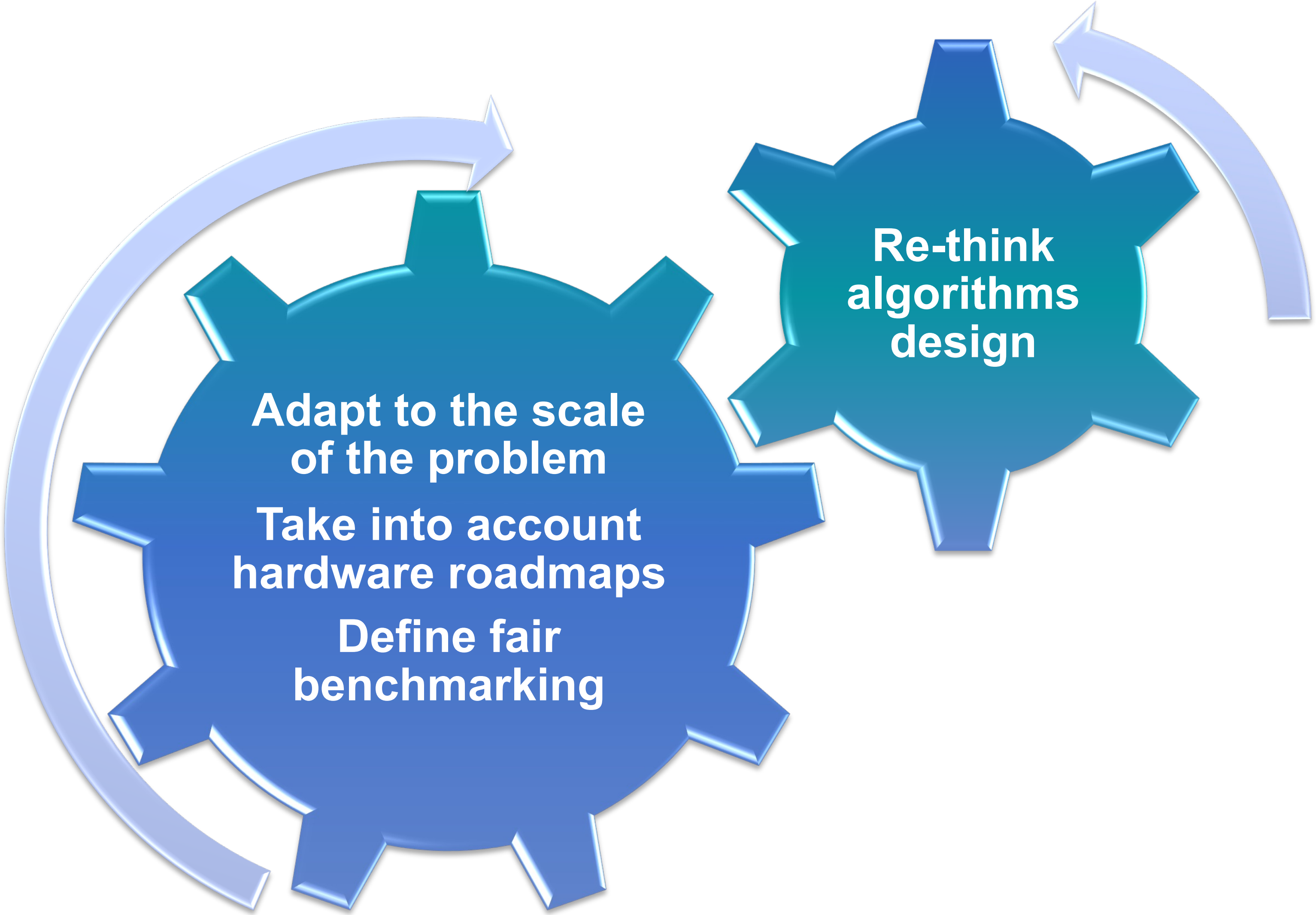
Circuit depth	Explicit loss (pairwise): Conventional strategy	Explicit loss (pairwise): Quantum strategy	Implicit loss (MMD)
Product	**No** (Corollary 2)	**Yes** (Local Quantum Fidelity [31])	**Yes** ($\sigma \in \Theta(n)$, Theorem 2)
Shallow			**Yes** ($\sigma \in \Theta(n)$, Conjecture 1)
Deep		**No** [22, 30]	**No** [22, 30]



Agenda

- Part 1: QC for Quantum Machine Learning
- Part 2: QML for HEP.
- Challenges
  - Input dimensionality
  - Symmetries and data structures
  - Discrete variables

Quantum ML for HEP

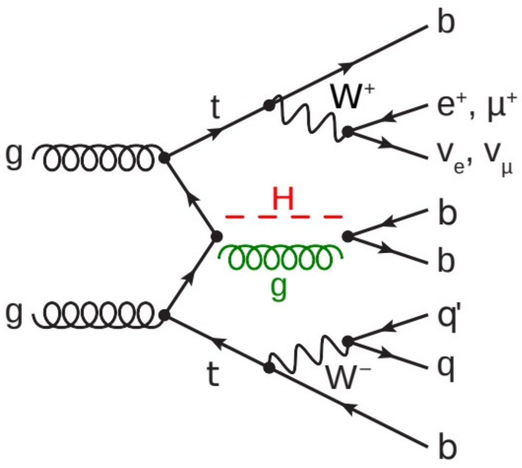


Quantum ML for **realistic data processing** at **next generation colliders**?

Data dimensionality reduction

Analysis setup

Analysis
Discrimination of the signal over the overwhelming background

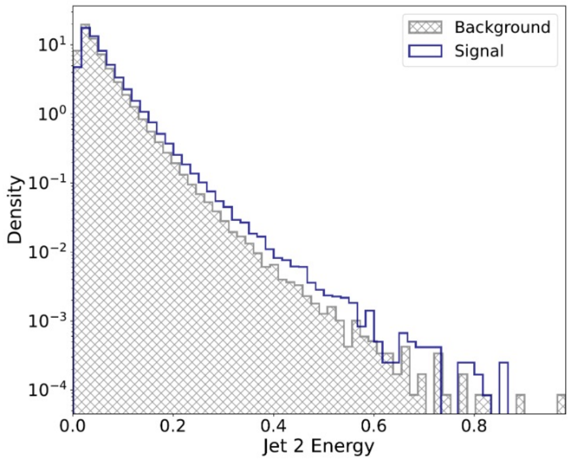


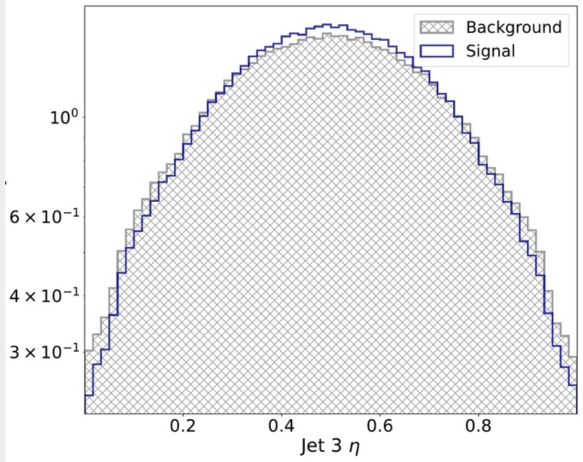
Features

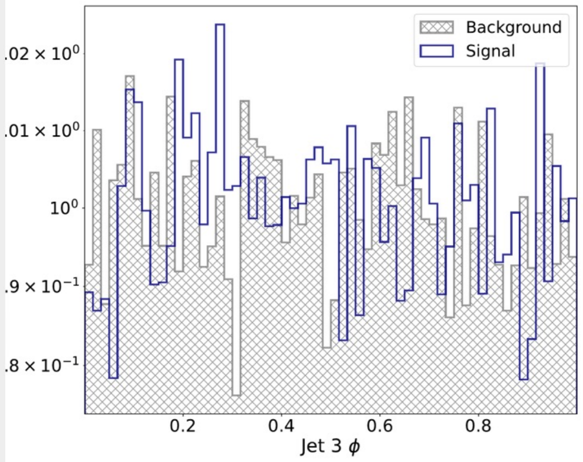
- For the each jet we have 8 features: (pT,η,ϕ,E,b tag,px,py,pz)
- For MET we have 4 features: (pT,px,py,ϕ)
- For the lepton (electron or muon) we have 7

features: (pT,η,ϕ,E,px,py,pz)

#features = 8×7(jets)+7(1lepton)+4(MET) = 67





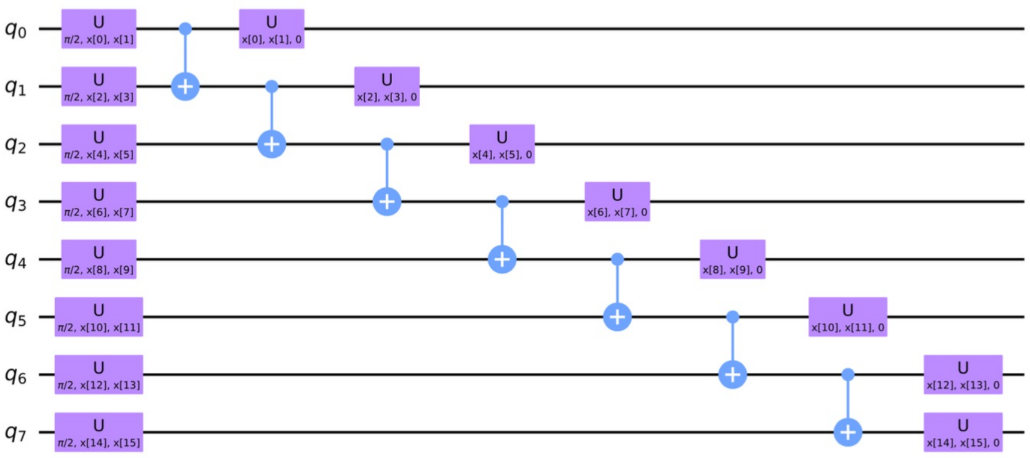


Quantum SVM for Higgs Classification

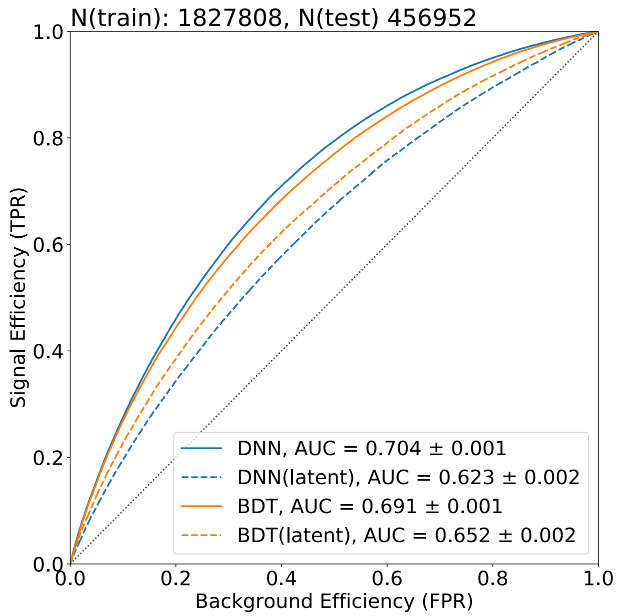
Input dimensionality reduction through an Auto-Encoder projects to a lower dimension latent space (8,16)

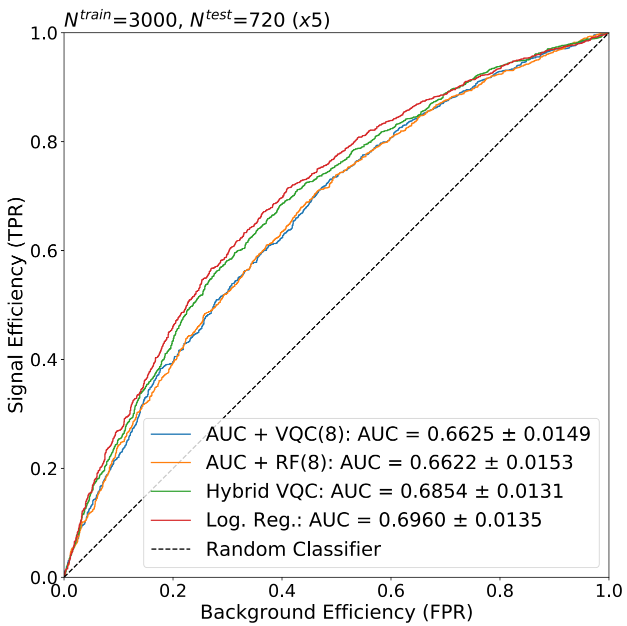
Feature selection + Model	AUC
AUC + QSVM	0.66 ± 0.01
PyTorch AE + QSVM	0.62 ± 0.03
AUC + SVM rbf	0.65 ± 0.01
PyTorch AE + SVM rbf	0.62 ± 0.02
KMeans + SVM rbf	0.61 ± 0.02

Feature selection + Model	AUC
AUC + QSVM	0.68 ± 0.02
AUC + Linear SVM	0.67 ± 0.02
Logistic Regression	0.68 ± 0.02



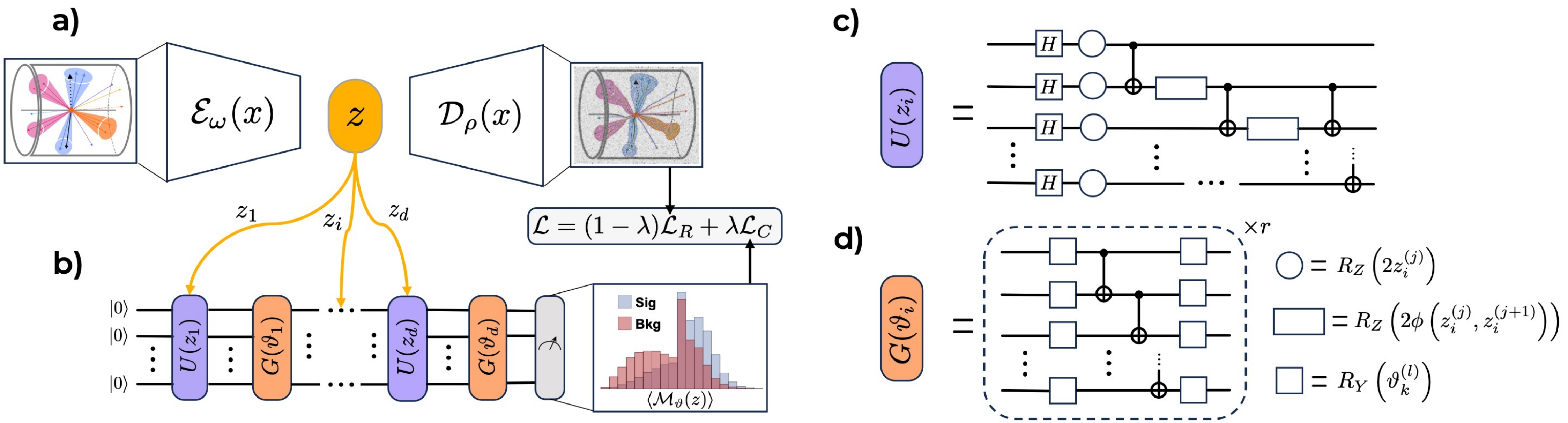
Data encoding circuit serving as feature map for the 8-qubit QSVM implementation.





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

Guided Quantum Compression

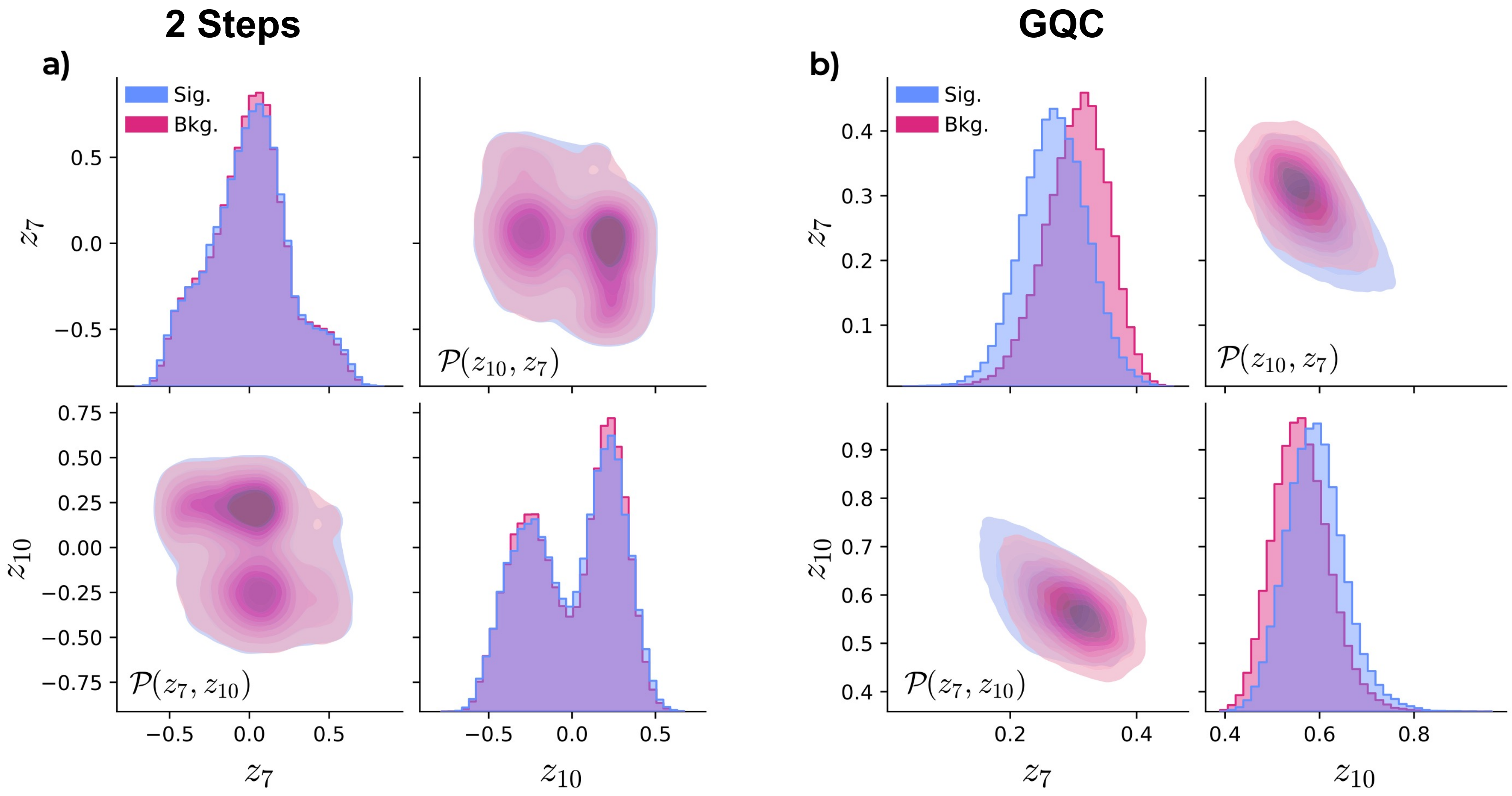


Two hybrid quantum-classical strategies:

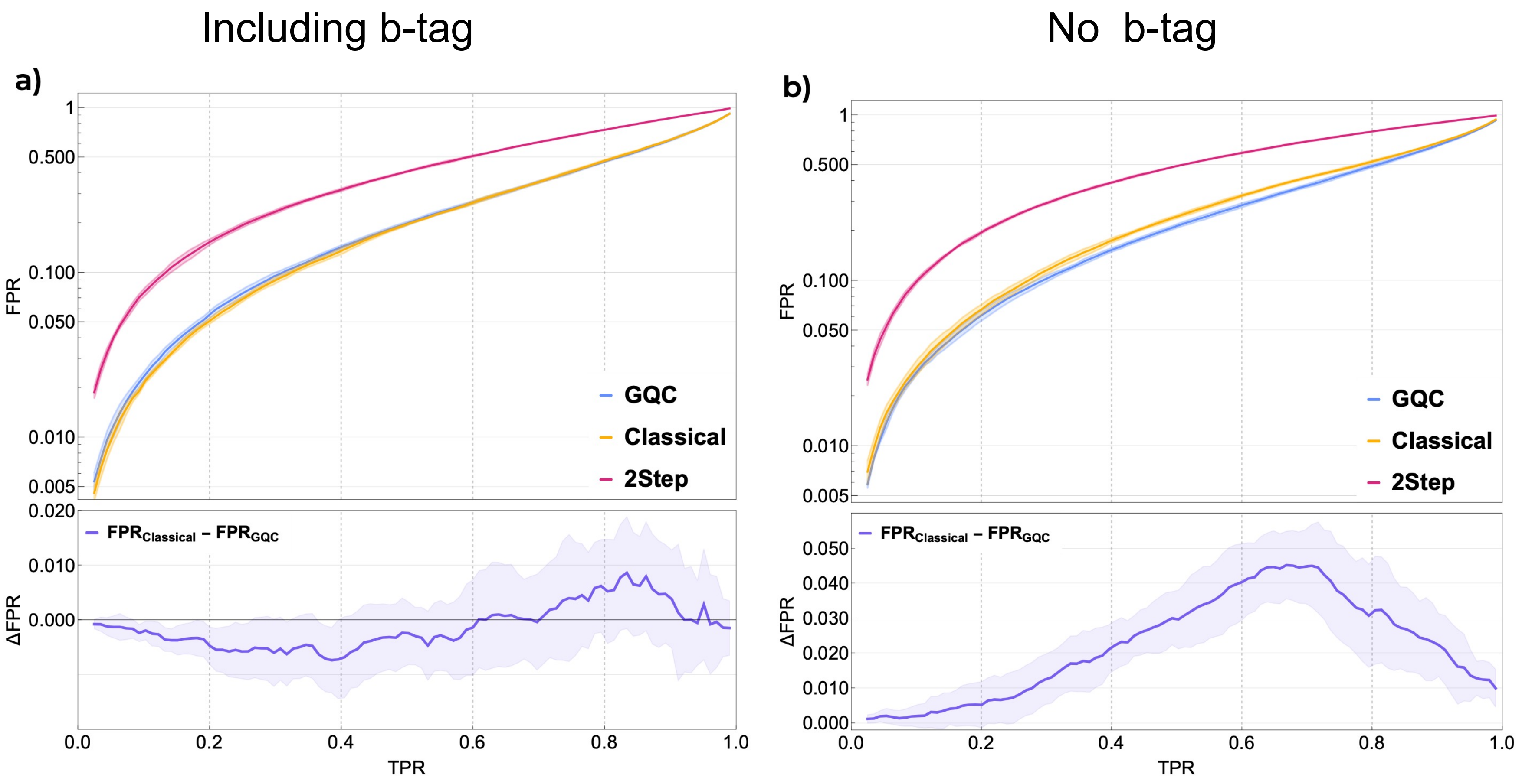
GQC: Joint training

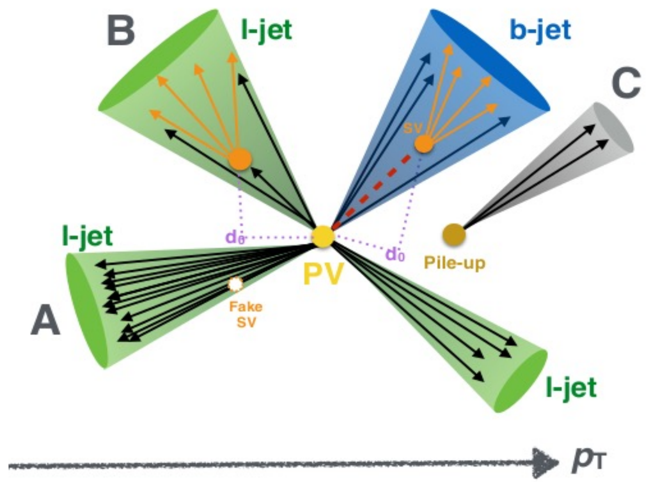
2Steps: The data compression step is independently trained

Latent Space Representation



Results





b-tag features are high level features containing information about the quark content

compression method has significant impact on the classifier performance.

CHALLENGE: DATA COMPRESSION

Leveraging symmetries

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 $R_G(\theta) = \exp(-i\theta G)$:

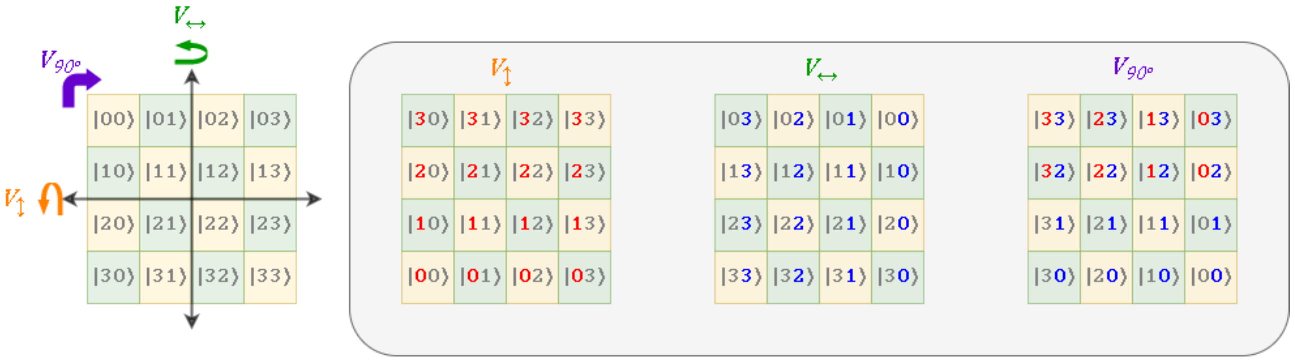
$$[R_G(\theta), V_s[g]] = 0 \quad \leftrightarrow [G, V_s[g]] = 0$$

Invariant Measurement:

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

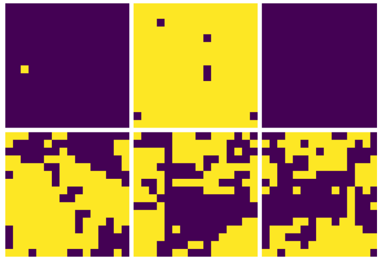
Equivariant Quantum CNN

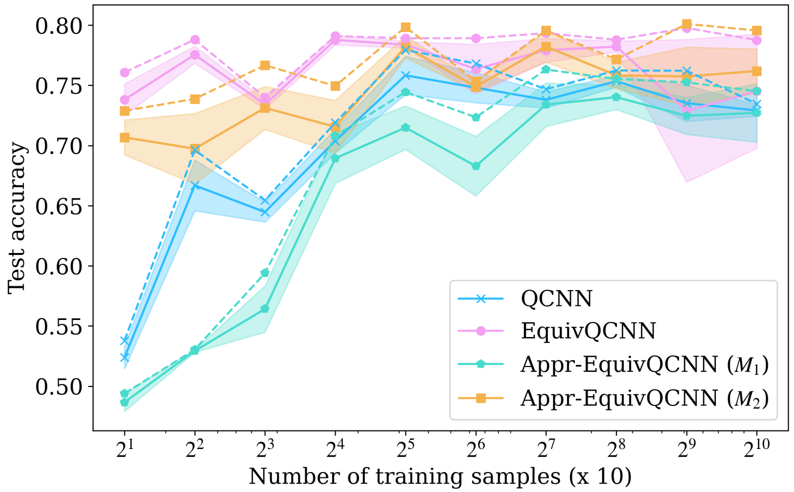
- Construct **equivariant** quantum CNN **under rotational & reflectional symmetry (p4m)**
- Improved generalization power



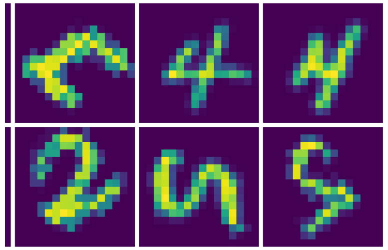
$$\mathcal{H} = -J \sum_{\langle ij \rangle} \sigma_i \sigma_j$$

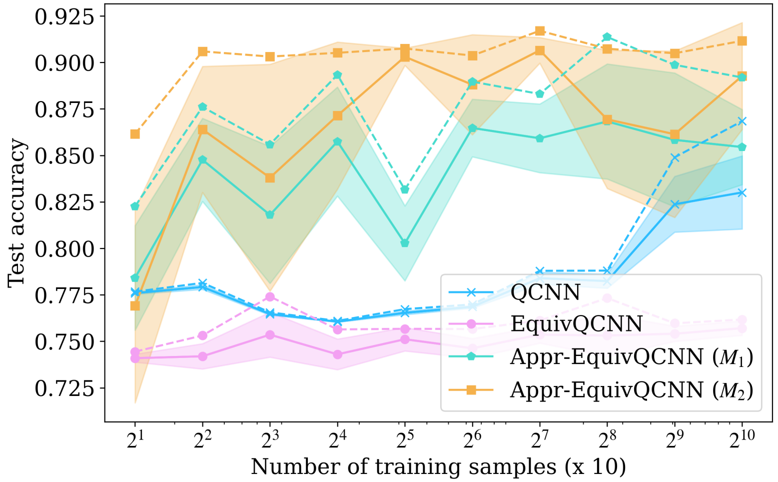
Ising spins phase classification :





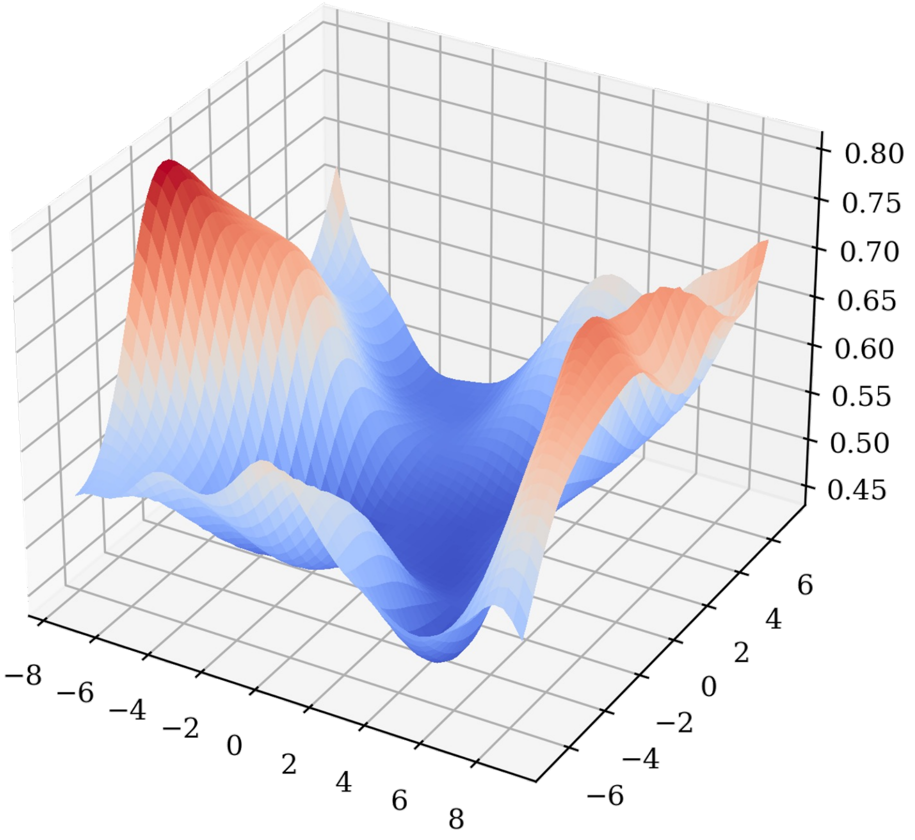
Extended MNIST Image classification: (digits 4,5)



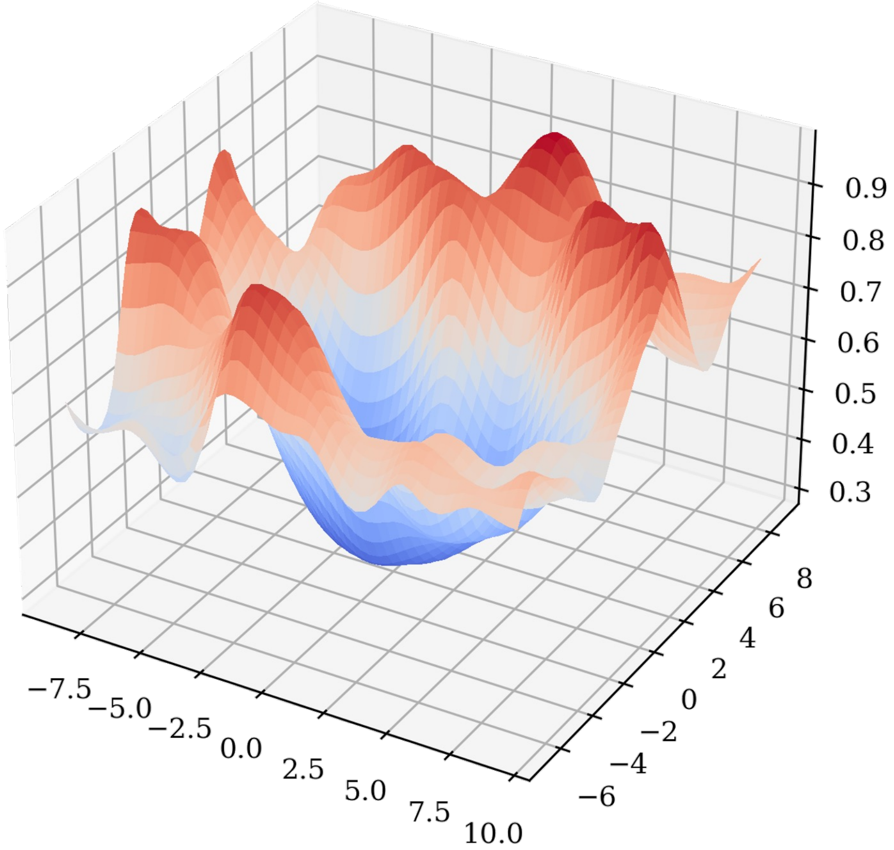


Non-convexity of loss landscape

Loss landscape plotted with orqviz



Non-equivariant QCNN



ApprEquivQCNN

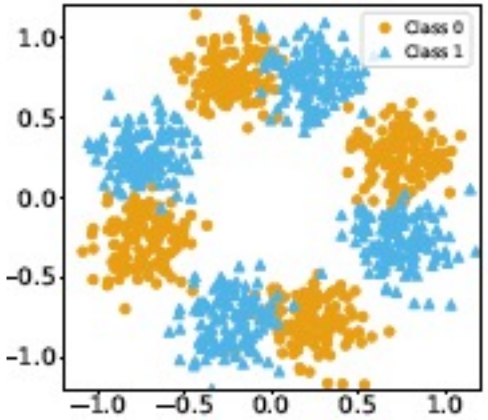
Tüysüz, Cenk, et al. "Symmetry breaking in geometric quantum machine learning in the presence of noise." *arXiv preprint arXiv:2401.10293* (2024).

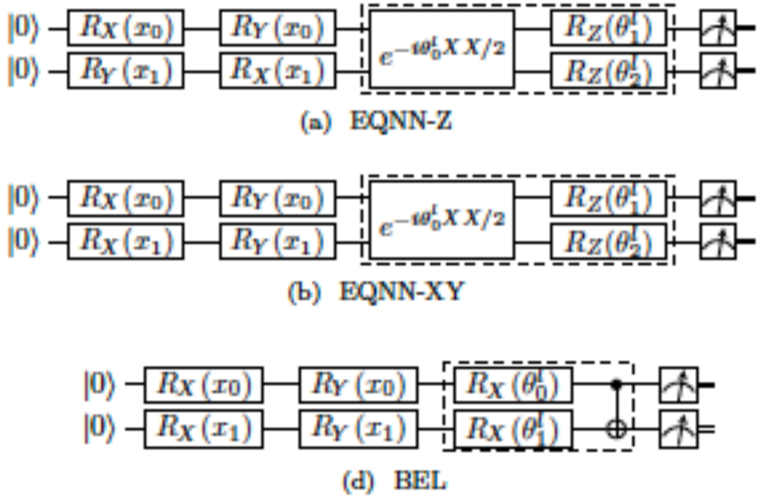
Noise induced symmetry breaking

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

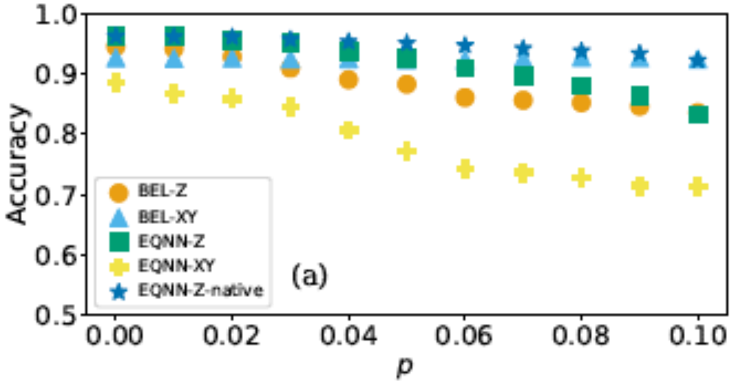
$\mathbb{Z}_2$: $R(\sigma)\cdot(x_i) = -x_i$

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

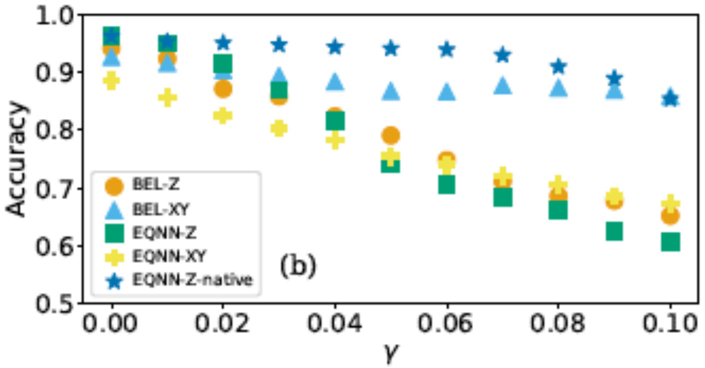




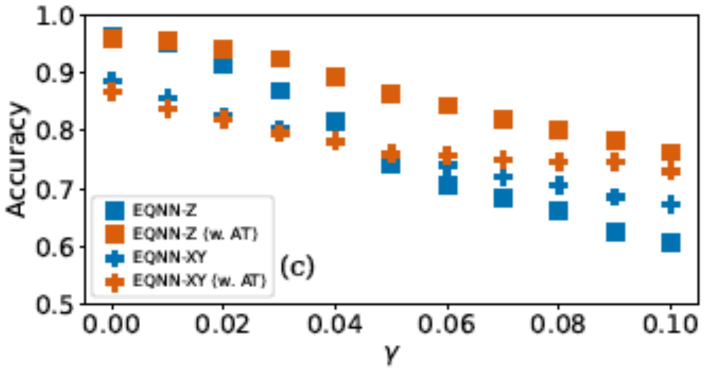
DP should not affect symmetry



EQNN performance drops with AD



Adaptive threshold classification



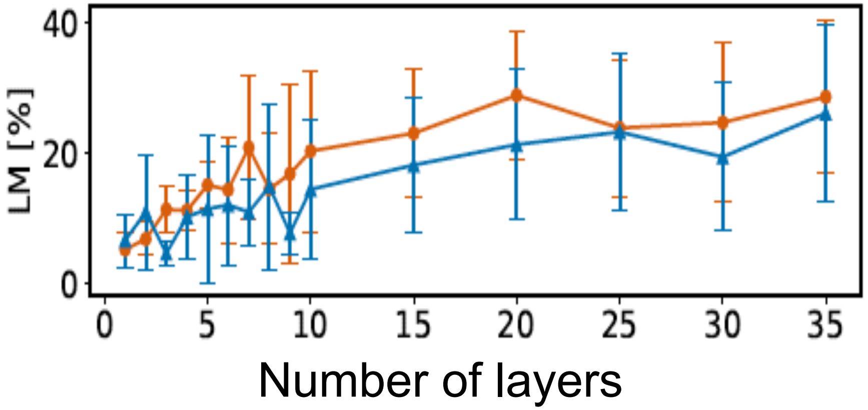
EQNN-Z native: Z_0Z_1 commutes with the AD channel generator, but native gate set is limited on hardware!

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*

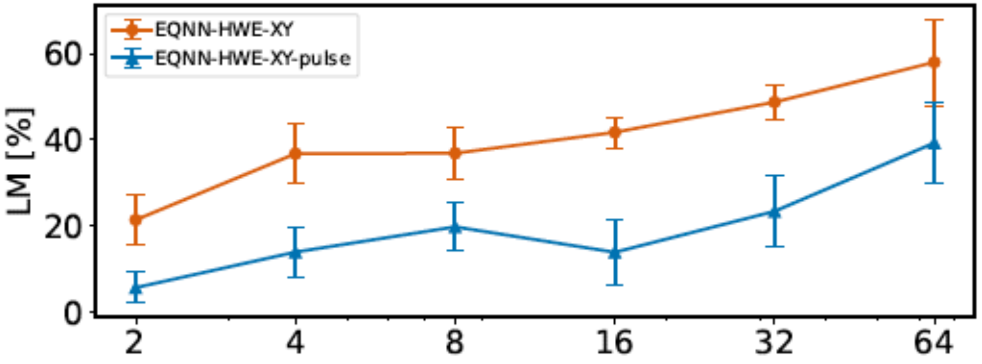
- Confirms AD channel is dominant
- **Symmetry breaking is linear in the number of layers**

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



Number of qubits
(Number of layers $\log_2 N$)

Cea, M., et al. "Exploring the Phase Diagram of the quantum one-dimensional ANNNI model." *arXiv preprint arXiv:2402.11022* (2024).

A full pipeline: QML analysis of quantum data

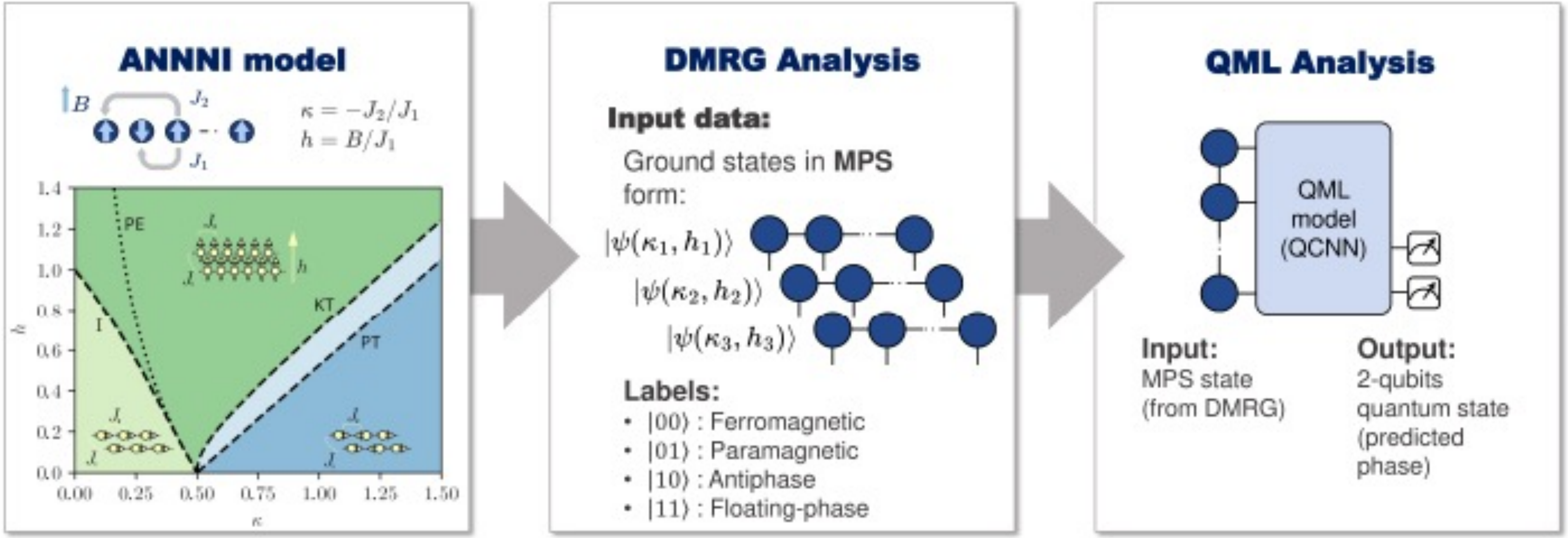
Connect QML and TN as different steps in the phase diagram reconstruction process for a ANNNI model.

Comparison between supervised and unsupervised QML

Use density matrix renormalization group (DMRG) for simulation of one dimensional multi-body systems (training data for the QML algorithms)

- Faster state preparation then VQE

Thanks to the TN characterization of the wave function, we can run a systematic study on QML performance



Phase classification result

Cea, M., et al. "Exploring the Phase Diagram of the quantum one-dimensional ANNNI model." *arXiv preprint arXiv:2402.11022* (2024).

Unsupervised QML

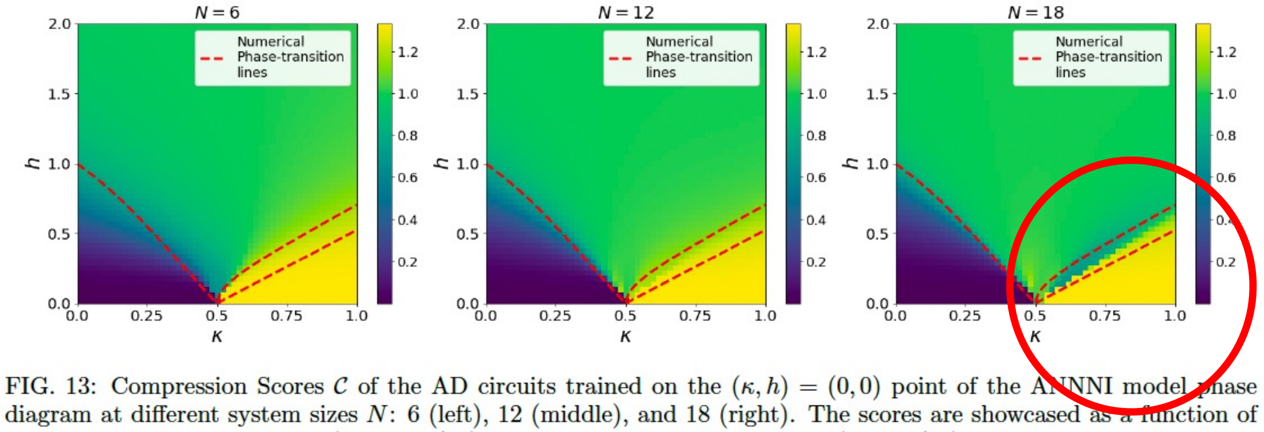


FIG. 13: Compression Scores $\mathcal{C}$ of the AD circuits trained on the $(\kappa, h) = (0, 0)$ point of the ANNNI model phase diagram at different system sizes N: 6 (left), 12 (middle), and 18 (right). The scores are showcased as a function of the interaction strength ratio ($\kappa = -J_2/J_1$) and the external magnetic field ($h = B/J_1$). Lower compression scores indicate better disentanglement of trash qubits from others, as defined by eq. 2.

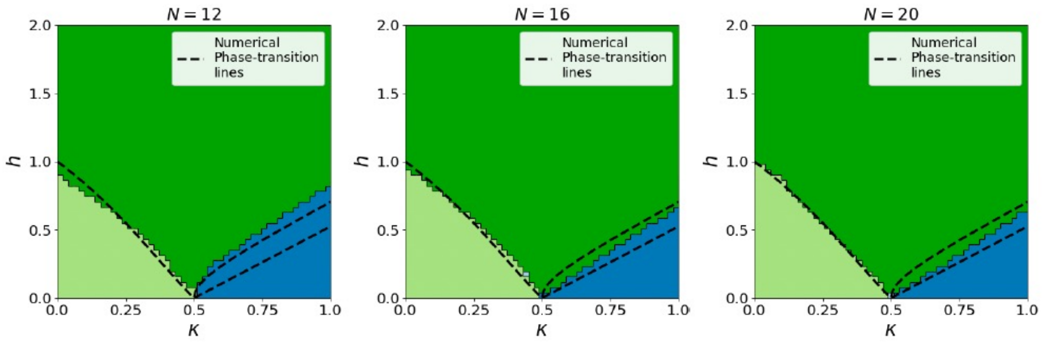


FIG. 11: Predictions of the QCNNs trained on the analytical points of the ANNNI Spin Model at different system sizes: $N = 12$ (left), $N = 16$ (middle), and $N = 20$ (right). Colors represent Ferromagnetic (light green), Paramagnetic (dark green), Antiphase (dark blue), and Floating Phase (light blue) as a function of the external magnetic field ($h = B/J_1$) and interaction strength ratio ($\kappa = -J_2/J_1$) (refer to eq. 1).

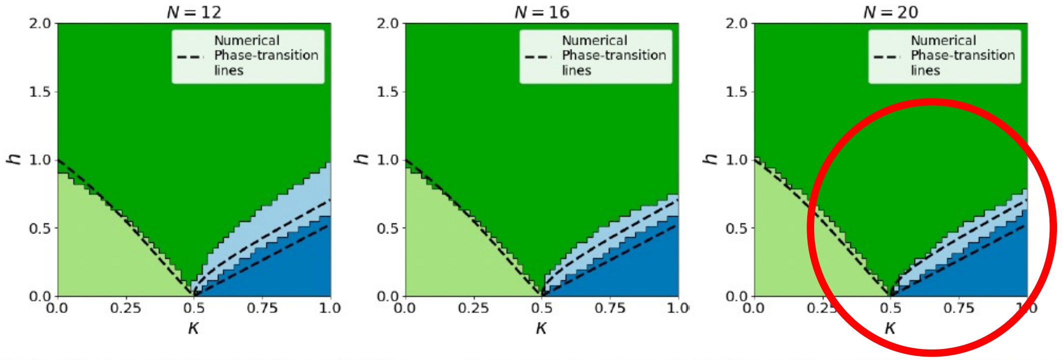


FIG. 12: Predictions of the QCNNs, trained on a subset of points from each phase of the ANNNI model at various system sizes: $N = 12$ (left), $N = 16$ (middle), and $N = 20$ (right). The color scheme indicates Ferromagnetic (light green), Paramagnetic (dark green), Antiphase (dark blue), and Floating Phase (light blue), as a function of the external magnetic field ($h = B/J_1$) and interaction strength ratio ($\kappa = -J_2/J_1$) (refer to eq. 1).

Supervised QML

Supervised QML does not generalize to the floating phase unless explicitly see at training time

With max 20 spins the systems still experiences significant limitations due to its constrained size

Open questions

- Quantum computing offers great opportunties while HEP provides challenging problems
  - **What are the most promising applications?**
  - How do we define performance and validate results on **realistic use cases**?

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

CERN is committed to creating impact on QT research in the coming years

CERN Quantum Technology Initiative

Accelerating Quantum Technology Research and Applications

Thanks!

Lectures and Hands-On at CERN

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

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

- «Introduction to quantum computing », Heather Grey

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

- A set of two hands-on (introduction) sessions part of the 2023 openlab summer student lectures series

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

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