2022 CERN openlab Technical Workshop

QML
Connecting the dots

Quantum machine learning questions

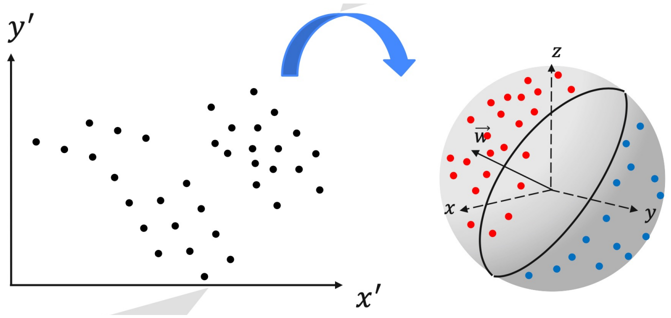
- **Classical intractability**: what useful problems can we solve on a quantum computer that we cannot on a classical computer?
- **Innovation**: what new algorithms can we come up with?
- **Computational complexity**: how can we obtain certain speedups?
- Can quantum **advantage** be proved with QML?

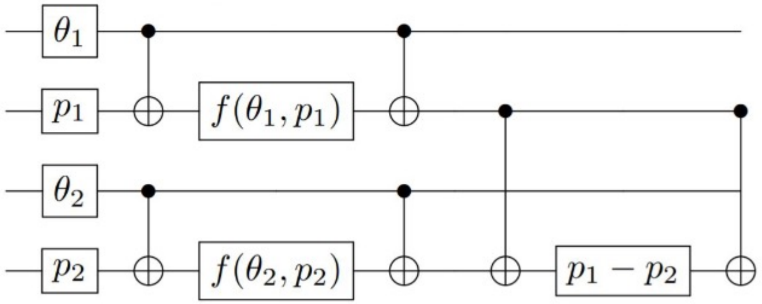
Characterize Quantum Advantage

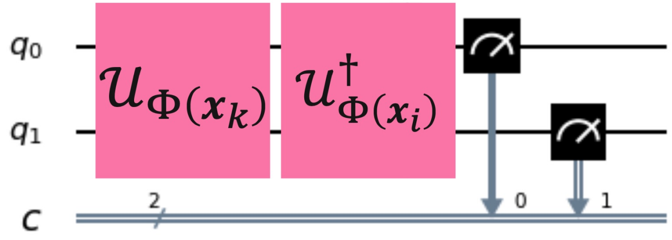
- Classical machine learning models can often compete or outperform existing quantum models even on data sets generated by quantum evolution, especially at large system sizes
- Large quantum Hilbert space in existing quantum models can result in significantly inferior prediction performance compared to classical machines: expressivity of QML hinder generalization
- We need a methodology for assessing the potential for quantum advantage in prediction on learning tasks
- Are there alternative research questions beyond the goal of beating classical machine learning?

Lifecycle of a Quantum Classification Problem [6,7,8]

- *Features reduction/extraction (classic step: PCA, AE, etc…)*
- Data Encoding
- Feature Map / ansatz definition for kernel and variational methods
- Read Out (Different technologies, NISQ, Error…)







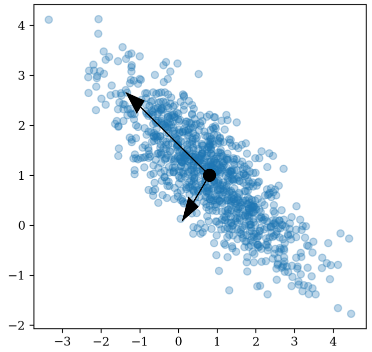
[6] Robust data encodings for quantum classifiers, Ryan LaRose and Brian Coyle, Phys. Rev. A 102, 032420

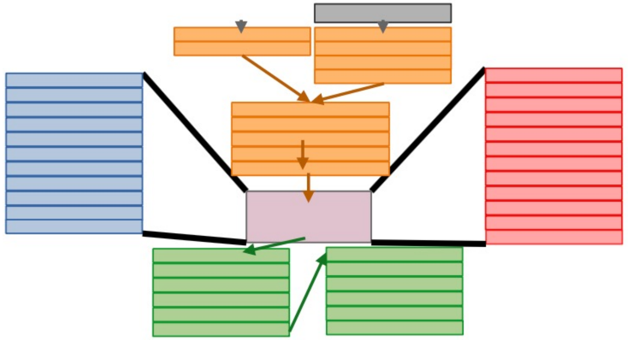
[7] Quantum convolutional neural network for classical data classification, https://arxiv.org/pdf/2108.00661.pdf

[8] Quantum Support Vector Machines for Continuum Suppression in B Meson Decays, https://arxiv.org/abs/2103.12257

Lifecycle of a Quantum Classification Problem

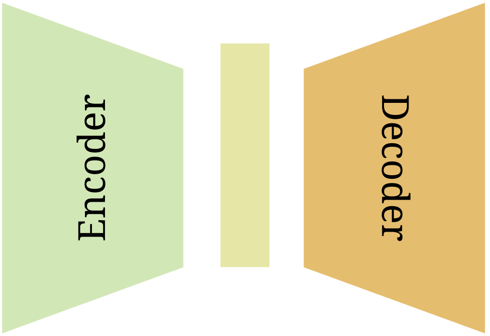
- *Features extractions/reduction (classic step: PCA, AE, etc…)*

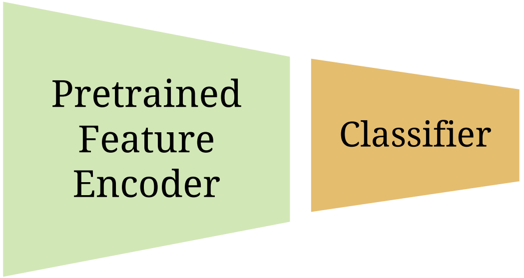




Autoencoders

1) Principal Components Analysis (PCA)
2) Convolutional Autoencoder
3) Wasserstein Variational Autoencoder
4) Transformers
5) Transfer Learning





Lifecycle of a Quantum Classification Problem

- *Features reduction/extraction (classic step: PCA, AE, etc…)*

- Data Encoding → how much is it important?

1) Amplitude Encoding

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

+ Exponential compression in number of qubits

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

− Polynomial number of gates

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

Lifecycle of a Quantum Classification Problem

- *Features reduction/extraction (classic step: PCA, AE, etc…)*

- **Data** Encoding → how much is it important?

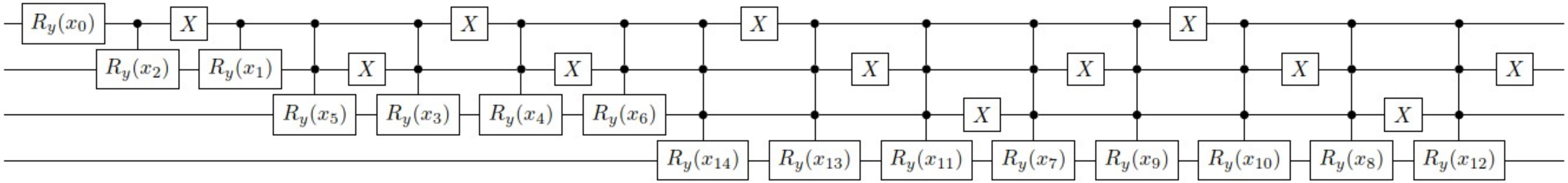
3) Hybrid Angle Encoding

→ Encode $b \times 2^m$ values into $b \times m$ qubits

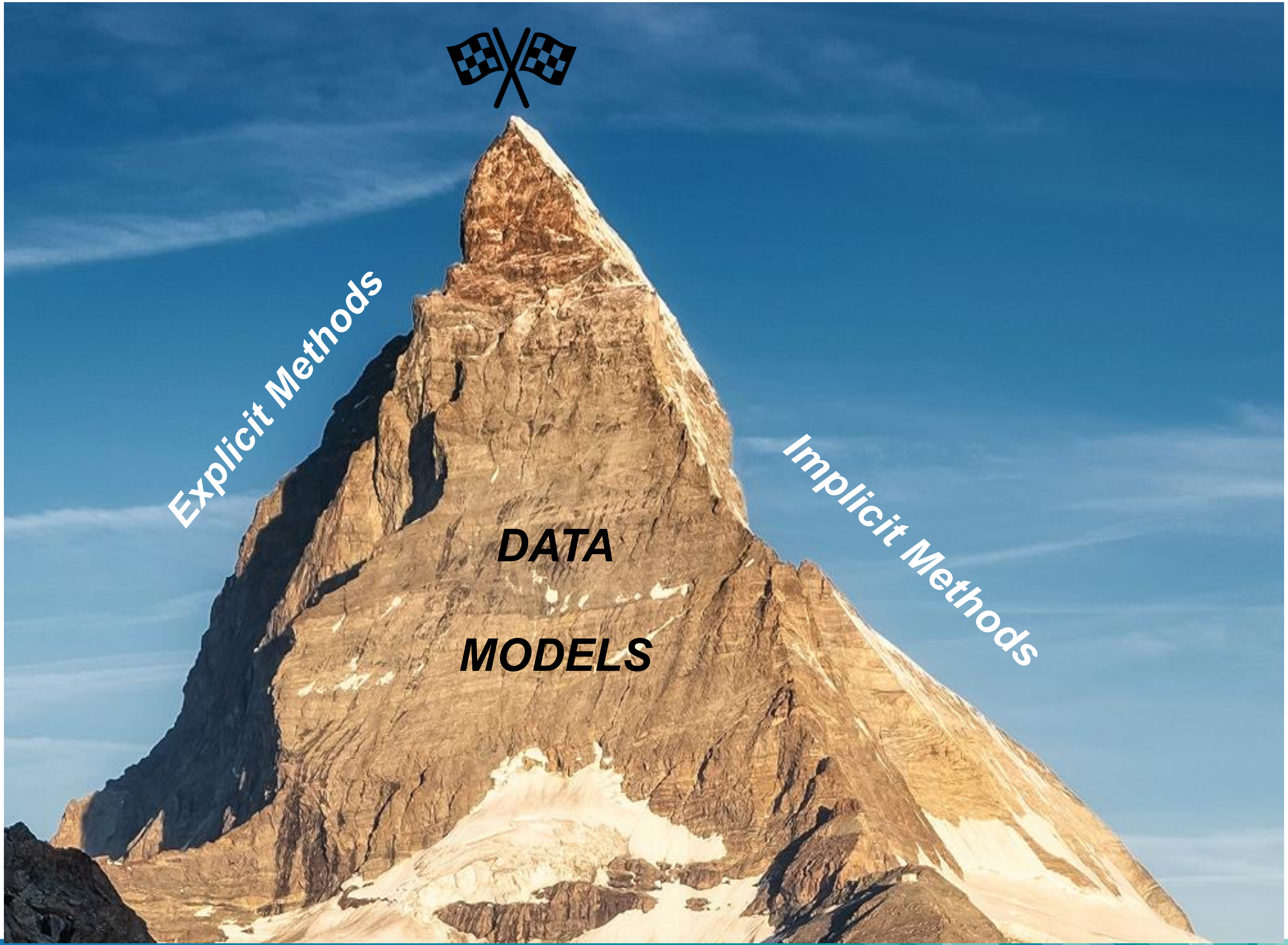
$$|\phi(x)\rangle = \otimes_{k=1}^{b} \left(\sum_{i=1}^{2^m} \prod_{j=0}^{m-1} \cos^{1-i_j}(x_{g(j),k}) \sin^{i_j}(x_{g(j),k}) |i\rangle_k \right.$$

\+ Compromise between Amplitude and Qubit Encoding

− Still requires too many two qubit gates



→ the choice of mapping (when exact) has no impact on the learning performance of the resulting (explicit) model, …it does impact its kernelization [4]


Explicit Methods
Implicit Methods
DATA
MODELS

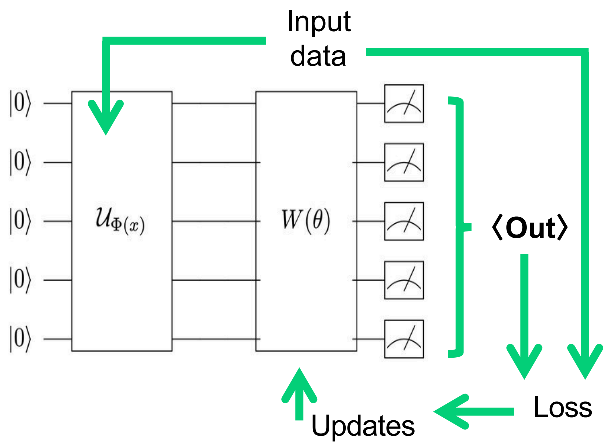
QML implementations

Variational algorithms - EXPLICIT

Parametric ansatz

Can use **gradient-free** methods or **stochastic gradient-descent**

Data Embedding can be **learned**



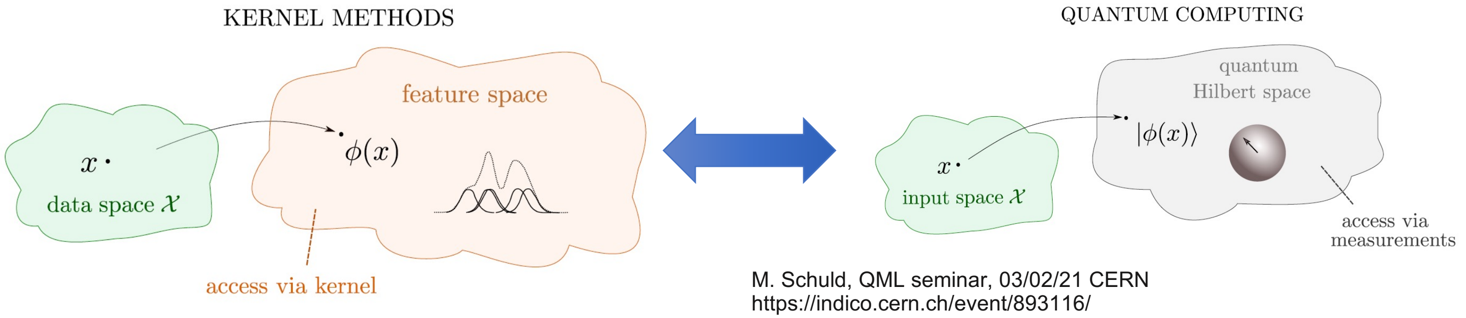
Kernel methods - IMPLICIT

Feature maps as **quantum kernels**

Use classical **kernel-based training**

Convex losses**, global** minimum

Compute pair-wise distances in N_{data}



M. Schuld, QML seminar, 03/02/21 CERN
https://indico.cern.ch/event/893116/

Do they really differ? Where to focus?

→ What is easiest to use/define?

*Data reduction - Data Encoding – Model performance (**accuracy vs generalization**)*

DATA

Model

→ guidance to find a quantum advantage in ML
  → no recipe for obtaining a quantum advantage on a **classical dataset** [1]

→ **Representer theorem:** implicit models can always achieve a smaller labelling error than explicit models on the same training set [4]

[1] The Inductive Bias of Quantum Kernels - https://proceedings.neurips.cc/paper/2021/file/69adc1e107f7f7d035d7baf04342e1ca-Paper.pdf
[4] Quantum Machine Learning Beyond Kernel Methods - https://arxiv.org/abs/2110.13162

*Data reduction - Data Encoding – Model performance (**accuracy vs generalization**)*

DATA

→ many relevant problems benefit from latent space connected with the theory of the specific underlying symmetry group [11]
→ identify a broad class of kernels that relate to learning problems with a particular structural aspect **(group structure –** i.e. discrete log**)** [2]
→ general method using covariant kernel alignment to explores a variety of fiducial states by applying a parameterised quantum circuit

Model

→ guidance to find a quantum advantage in ML
→ no recipe for obtaining a quantum advantage on a **classical dataset**. [1]
→ **Representer theorem:** implicit models can always achieve a smaller labelling error than explicit models on the same training set [4]

*Unless we have a clear idea how the **data** generating process can be described with a quantum computer, we cannot expect an advantage by using a quantum model in place of a classical machine learning model.*

[11] Lorentz Group Equivariant Neural Network for Particle Physics - https://arxiv.org/pdf/2006.04780.pdf
[2] Covariant quantum kernels for data with group structure - https://arxiv.org/abs/2105.03406
[4] Quantum Machine Learning Beyond Kernel Methods - https://arxiv.org/abs/2110.13162

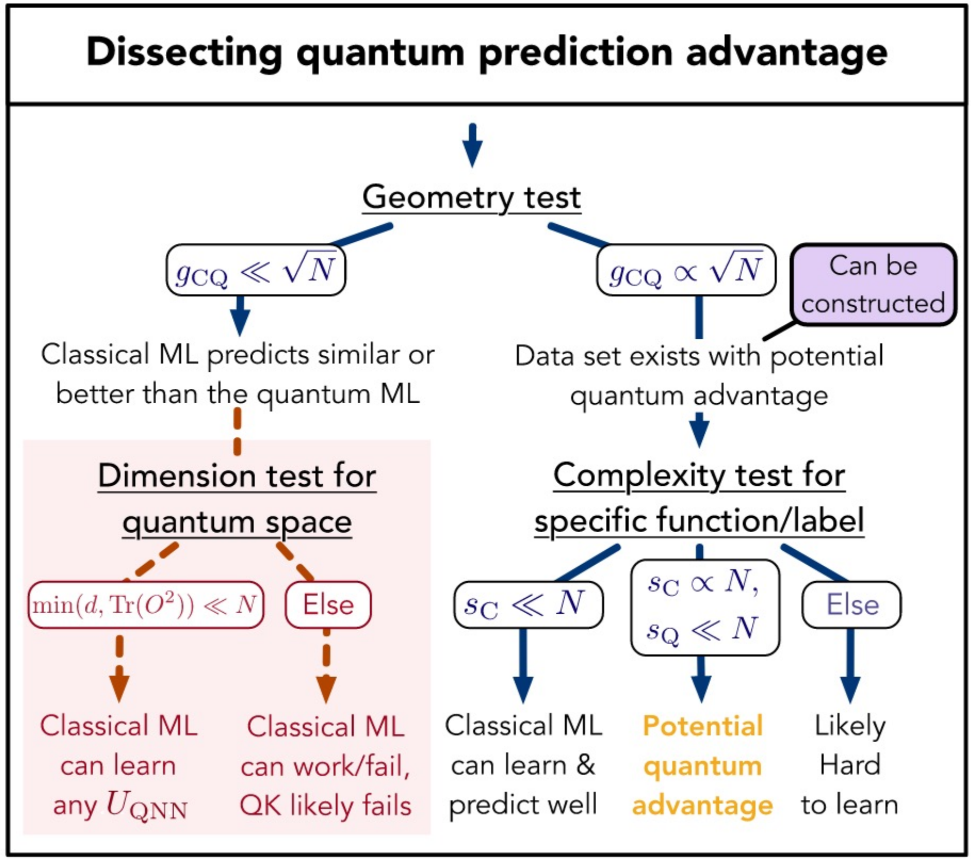
Potential Quantum Advantage

We have to utilize the entire exponential quantum state space otherwise the quantum machine learning model could be simulated efficiently classically.

Native quantum state space to define the kernel function can fail to learn even a simple function when the full exponential quantum state space is being used.

- Reduce the dimensionality of RKHS by projection of QK:
  - to limit the expressivity
  - to construct inductive bias → classically hard
    → related to the target function

A possible theoretical discriminator: The power of Data in (Q)ML algorithms [3]

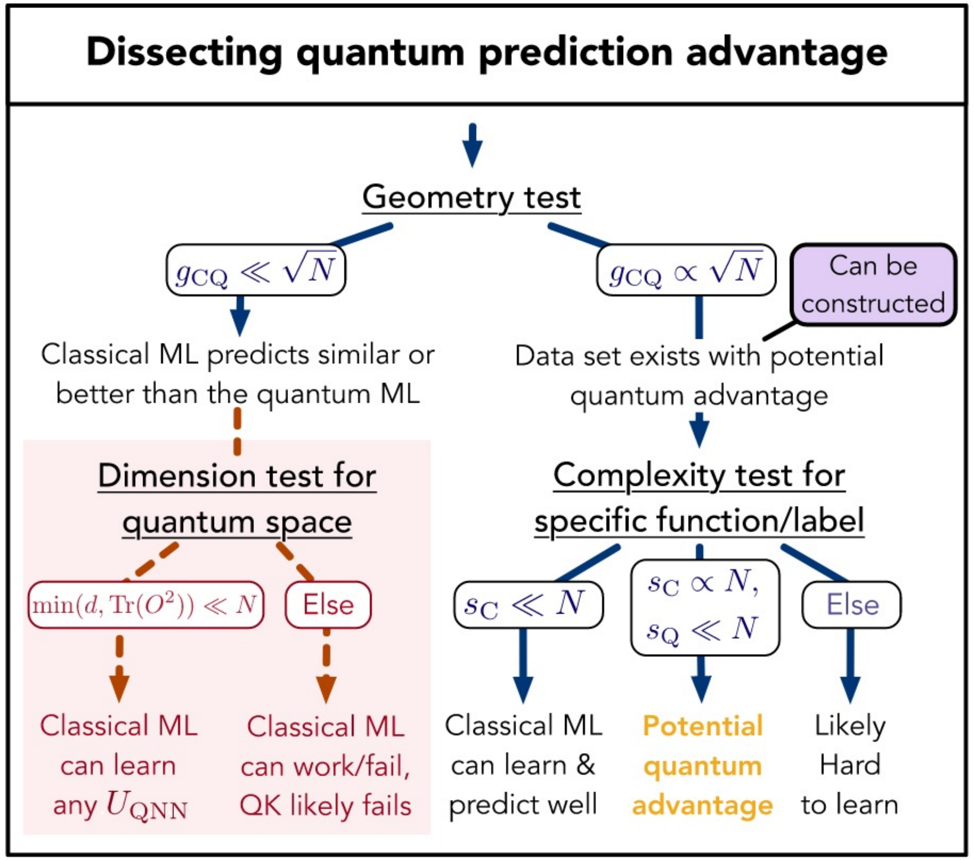


projected quantum kernels still use the exponentially large quantum Hilbert space for evaluation and can be hard to simulate classically

[3] The power of Data in (Q)ML algorithms - https://www.nature.com/articles/s41467-021-22539-9

A possible solution: Projected Quantum Kernel [3]

The projection allows us to reduce to a low-dimensional classical space that can **generalize** better



Theoretically, Projected quantum kernel can learn **any quantum models** with sufficient data.

It can be reconstructed by local randomized measurements using the formalism of **classical shadows.**

[3] The power of Data in (Q)ML algorithms - https://www.nature.com/articles/s41467-021-22539-9

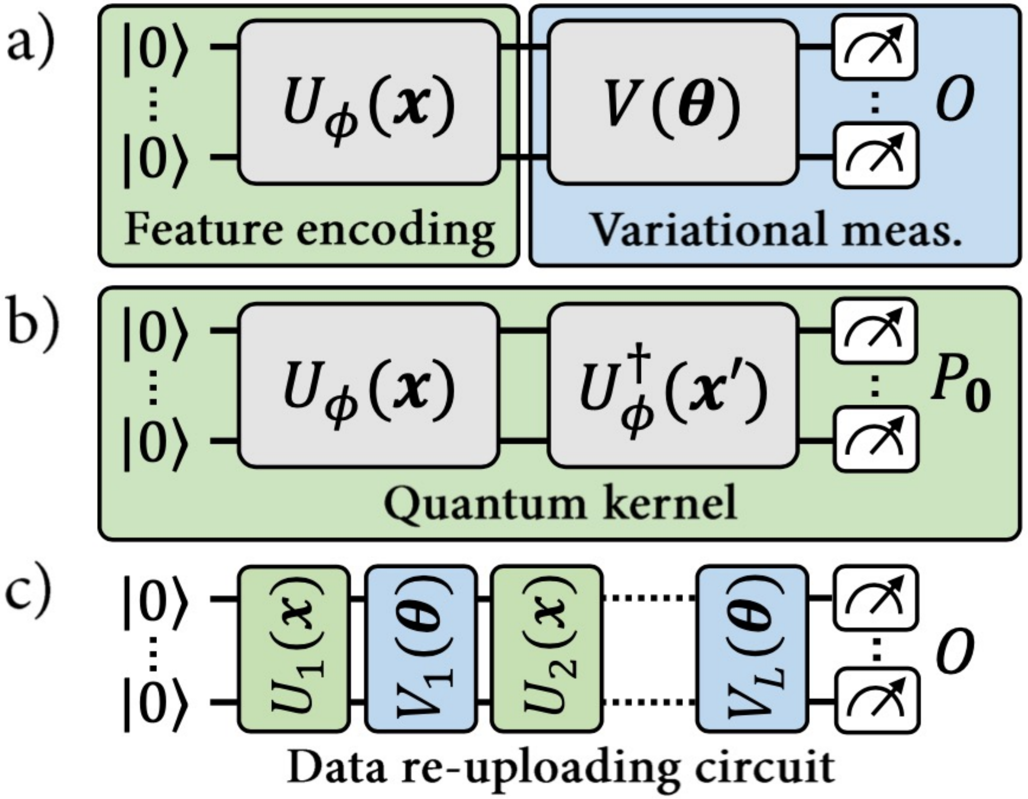
Quantum models

3 types of quantum machine learning models that can be formulated as linear models in quantum feature spaces:

- encoded data point is measured according to a variational observable
- weighted inner products of encoded data points are used to assign labels
- parametrized quantum circuits are universal function approximators, data- encoding layers, interlaid with variational unitaries

EXPLICIT

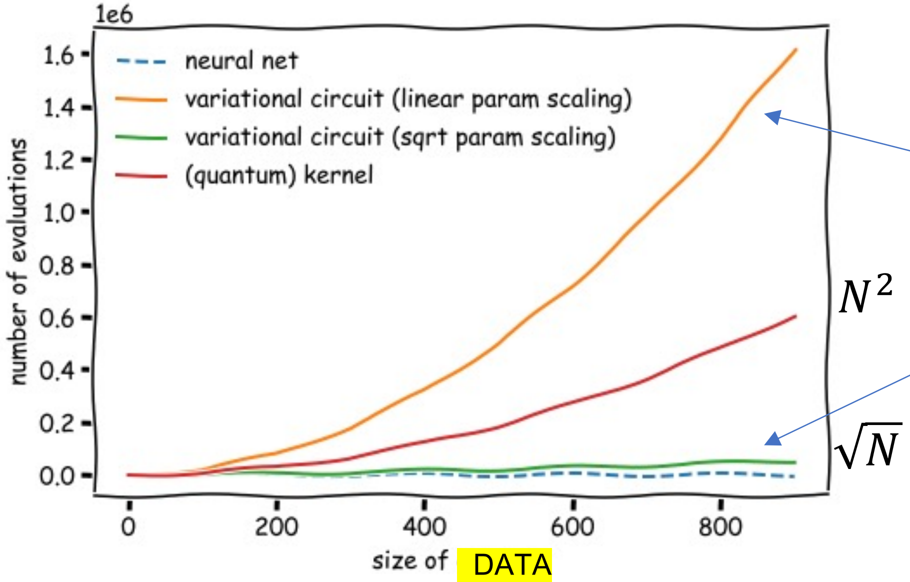
IMPLICIT



[4] Quantum Machine Learning Beyond Kernel Methods - https://arxiv.org/abs/2110.13162

Quantum models comparison

- Models defined and trained **variationally** can exhibit a critically better generalization performance than their kernel formulations
- Explicit models is expected to terminate in *O(M)* optimization steps
- Implicit models can be computed using *O(M^2)* evaluations of inner products on a quantum computer



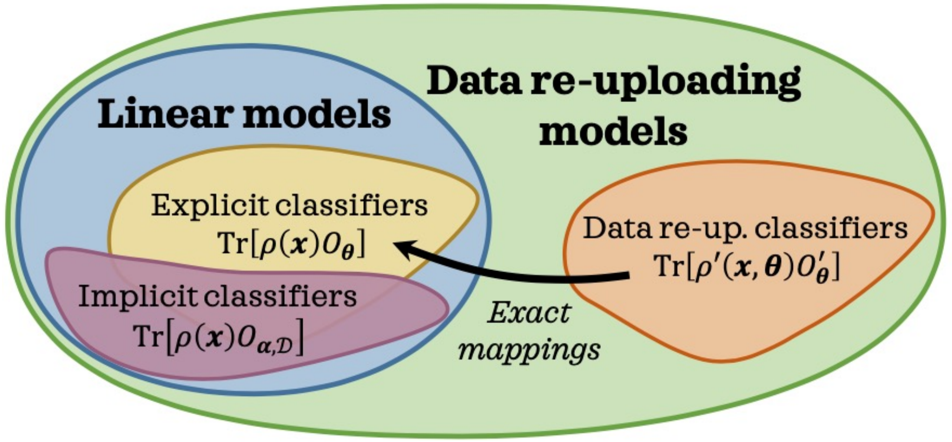
Different scaling with the number of data samples:

- number of parameters of the order of the training set size
- Small number of variational parameters

https://pennylane.ai/qml/demos/tutorial_kernel_based_training.html

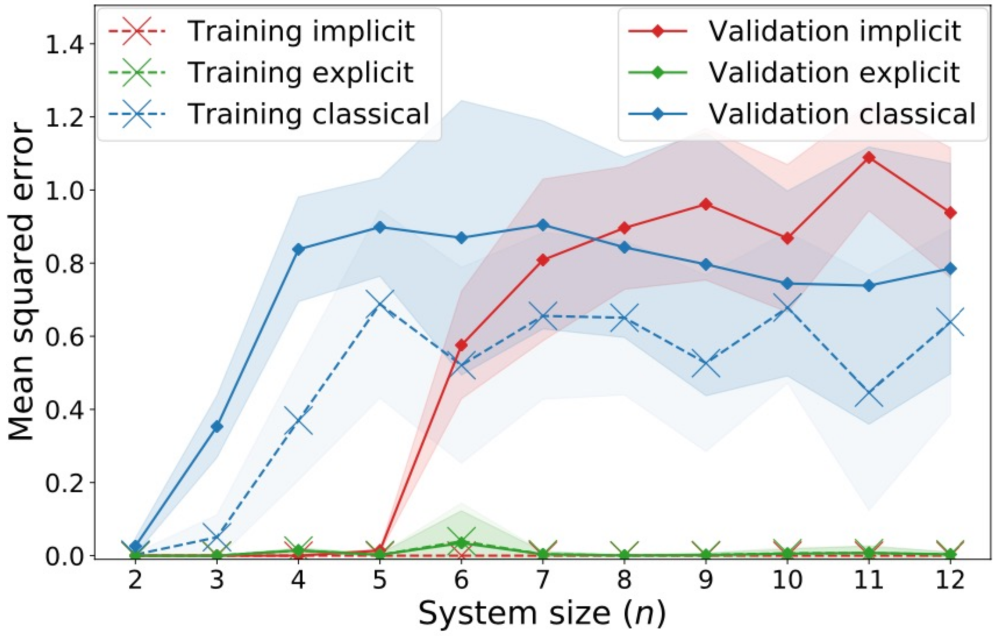
[4] Quantum Machine Learning Beyond Kernel Methods - https://arxiv.org/abs/2110.13162

A unified picture for EX, IM, data re-uploading quantum models



- Data re-uploading circuits can be represented exactly by explicit linear models in larger feature spaces, which allows them to be reformulated as implicit kernel methods
- Kernelizing a linear model also has its advantages, namely generalizing its hypothesis family and turning the learning task into a convex optimization problem
- **Representer theorem:** implicit models can always achieve a smaller labelling error than explicit models on the same training set.
- explicit models exhibit a critically better generalization performance than their kernel formulations

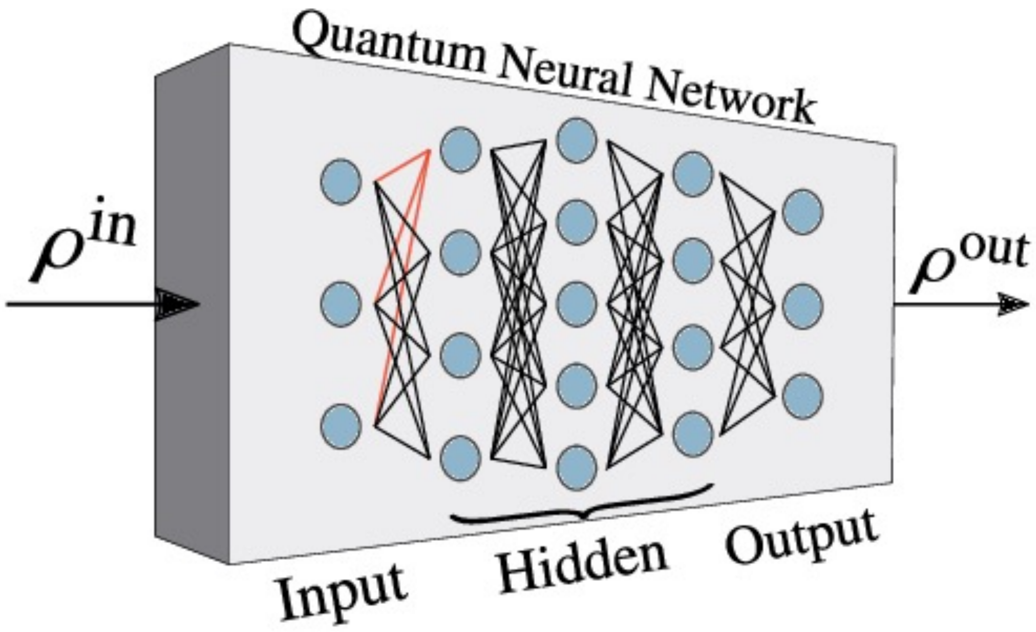
→ *increased expressivity can also be guaranteed to increase generalization performance??*

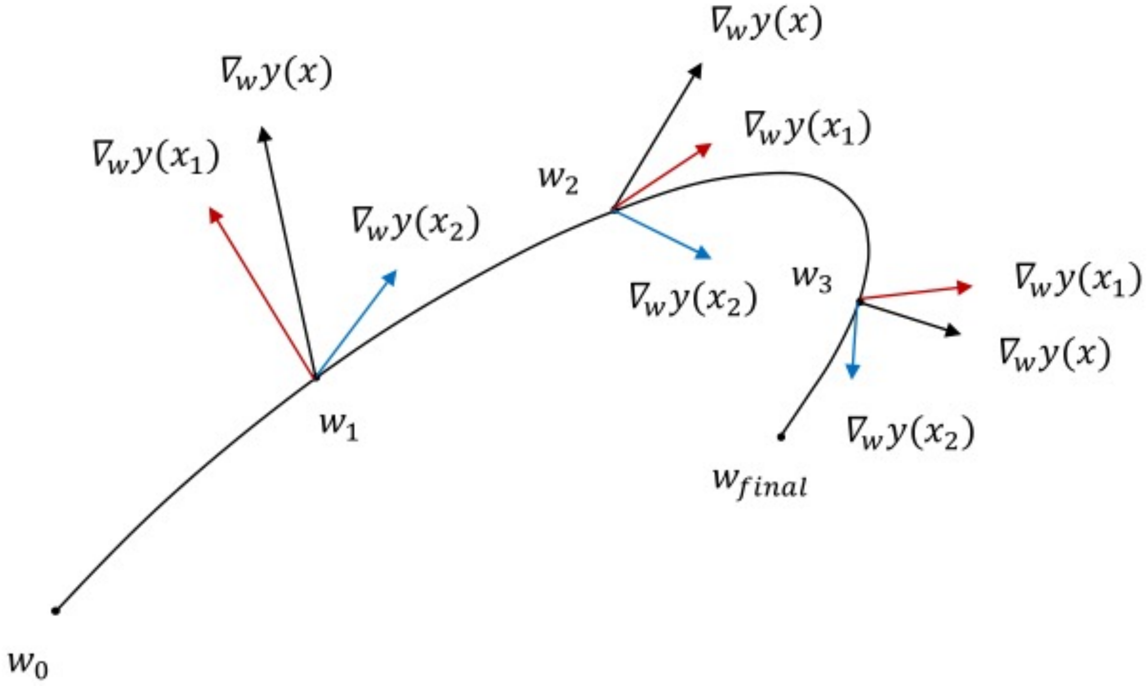


PCA on 28x28 fashion-MNIST dataset, ZZ feature encoding + hardware-efficient variational unitary

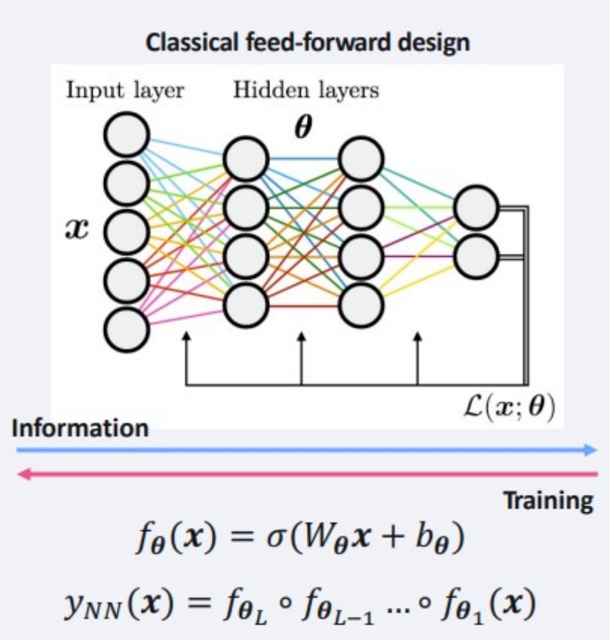
[4] Quantum Machine Learning Beyond Kernel Methods - https://arxiv.org/abs/2110.13162

(Q)NN in depth or in width?





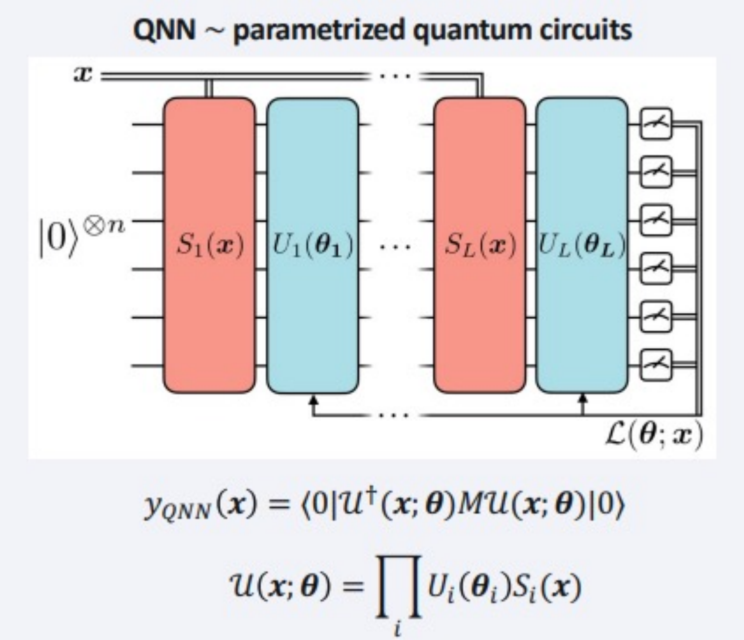
Relation between kernel and DNN in classic world [9] is true also in QC?



- Every finite-width NN trained by l^2 regularized loss functions is approximately a Kernel Machine [9]
- Quantum kernel can learn arbitrarily deep quantum neural network [3]
- Gaussian kernel can learn any QNN [3]

→ Quantum ML based on kernels can be made equivalent to training an infinite depth quantum neural network

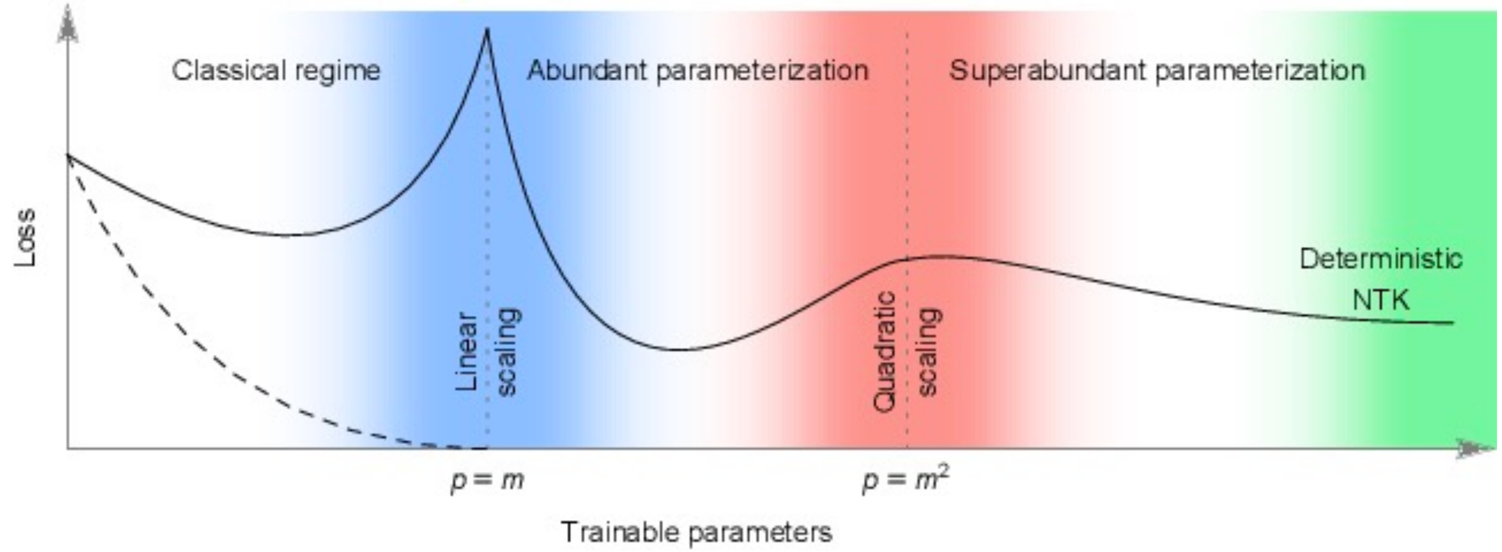
→ Larger models has been shown empirically to work better



[3] The power of Data in (Q)ML algorithms - https://www.nature.com/articles/s41467-021-22539-9

[9] On the Equivalence between Neural Network and Support Vector Machine - https://arxiv.org/abs/2111.06063

Relation between kernel and DNN in classic world [9] is true also in QC?



The Neural Tangent Kernel in High Dimensions: Triple Descent and a Multi-Scale Theory of Generalization

- The dynamics of an infinitely wide neural network (NN) trained by gradient descent can be characterized by Neural Tangent Kernel (NTK) [5]
- Neural tangent kernel shows that training neural networks with large hidden layers is equivalent to training an ML model with a particular kernel [3]
- Equivalence between NN and a broad family of l^2 regularized KMs with finite-width bounds

→ every finite-width NN trained by such regularized loss functions is approximately a KM [9]

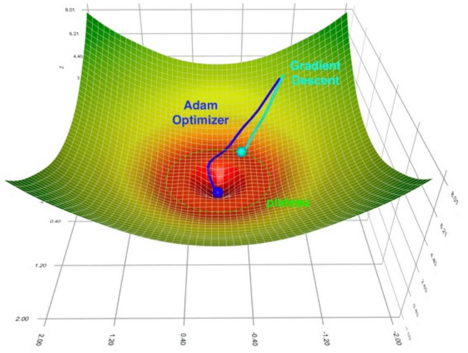
- Quantum NTK: deep parameterized quantum circuit whose representation power and performance are expected to be enhanced. (QNTK is not QSVM)

[3] The power of Data in (Q)ML algorithms - https://www.nature.com/articles/s41467-021-22539-9

[5] Quantum tangent kernel - https://arxiv.org/pdf/2111.02951.pdf [9] On the Equivalence between Neural Network and Support Vector Machine - https://arxiv.org/abs/2111.06063

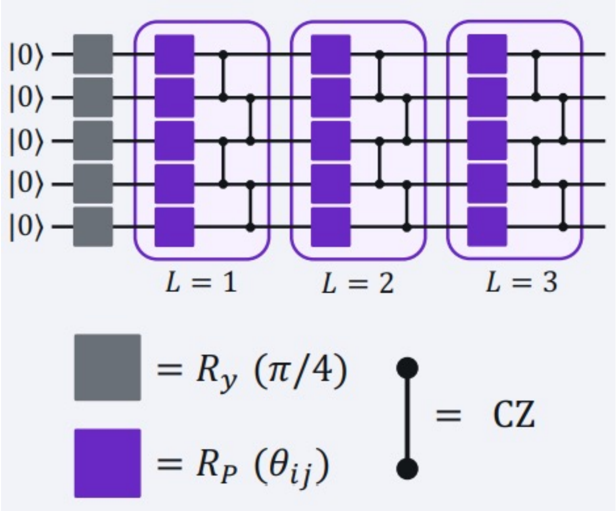
What about barren plateau??

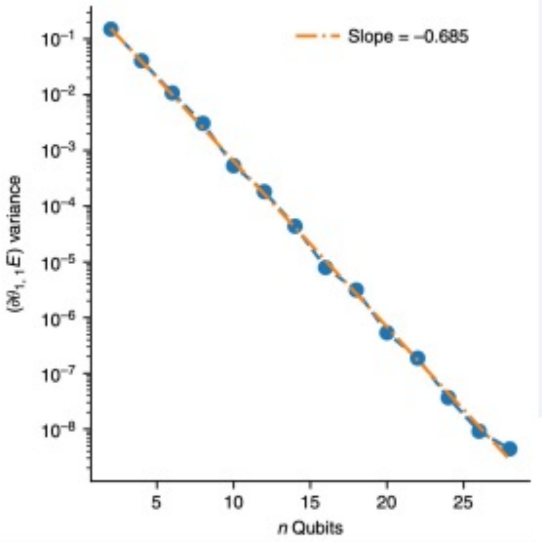
Noisy-induced, entanglement induced, cost function dependent…



Barren plateaus are large regions of the cost function’s parameter space where the variance of the gradient is almost 0 - the cost function landscape is flat.
This means that a variational circuit initialized in one of these areas will be untrainable using any gradient-based algorithm.

Avoiding barren plateaus using classical shadows - https://arxiv.org/pdf/2201.08194.pdf





The variance of the gradient decreases exponentially with the number of qubits

Barren plateaus in quantum neural network training landscapes, https://arxiv.org/abs/1803.11173

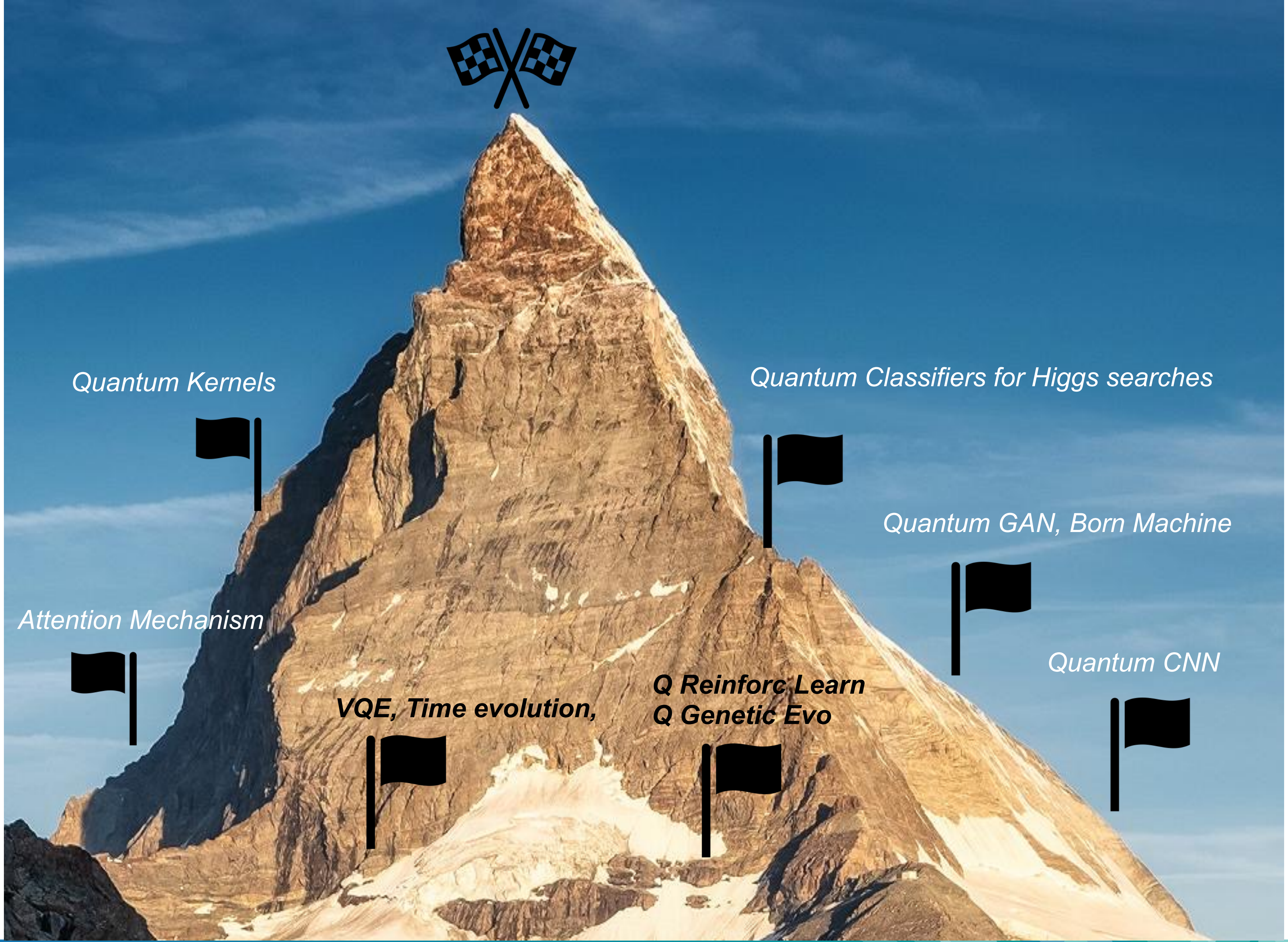
Classical shadow estimation (VQE test)

- efficient diagnosis of the WBP both at the initialization step and during the optimization process of variational parameters
- entanglement induced BPs and BPs for local cost functions are the same
- the algorithm restarts the optimization process with a decreased value learning rate
- avoidance of BPs during the optimization using quantum Fischer information
- classical shadow protocol and the estimation of observables are **stable** with respect to the addition of a small but finite amount of **noise**

Conclusion

- What is the actual power of Representation learning in QML?
- Is there any potential of representation learning for Deep QNN wrt classic DNN?
- Implict vs Explicit is the same for QML and ML?

We are climbing the mountain, placing a flag for each individual contribution at the very frontier of (Q)ML


Quantum Kernels
Quantum Classifiers for Higgs searches
Quantum GAN, Born Machine
Attention Mechanism
Quantum CNN
Q Reinforc Learn
Q Genetic Evo
VQE, Time evolution,

CERN

IQ

QUANTUM
TECHNOLOGY
INITIATIVE