The Role of Data in Projected Quantum Kernels

The Higgs Boson Discrimination

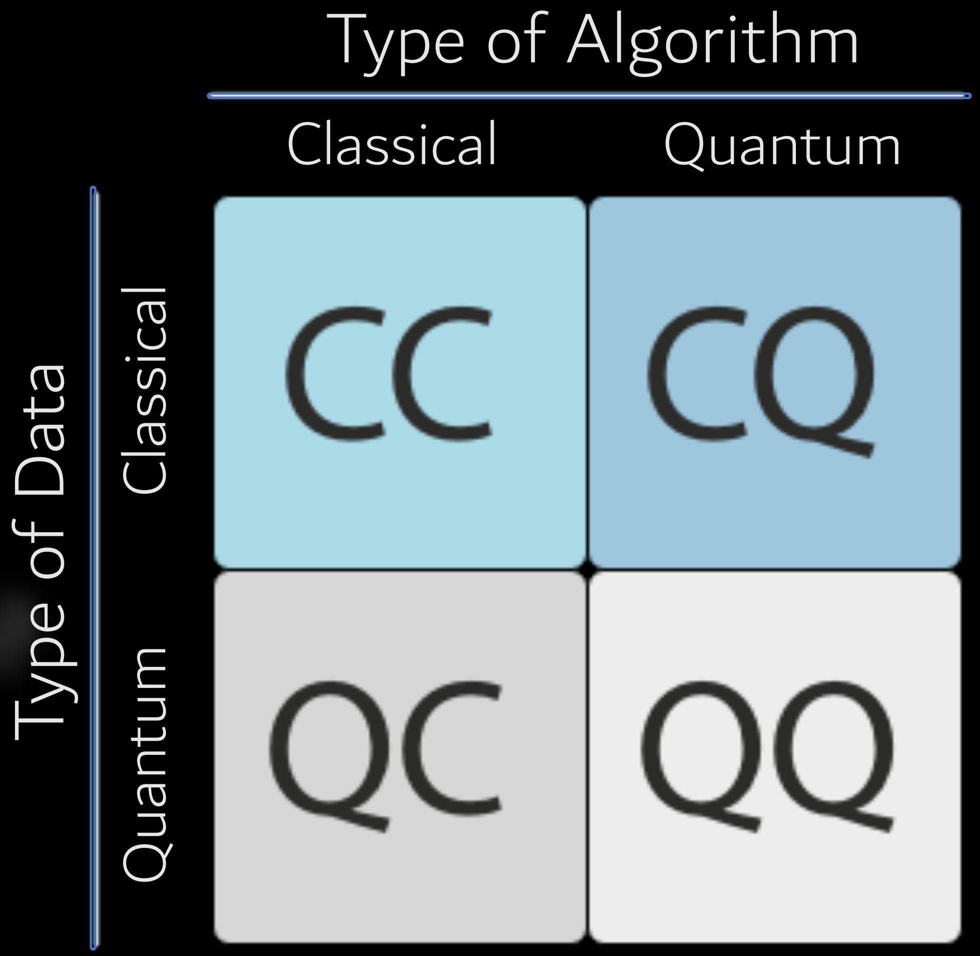
QUANTUM
TECHNOLOGY
INITIATIVE

Francesco Di Marcantonio, **Sofia Vallecorsa**, Michele Grossi

QML in HEP

- Does it make sense to use QML in HEP?

- How do we understand when it is *useful* ?
- Which are the QML models we can leverage?

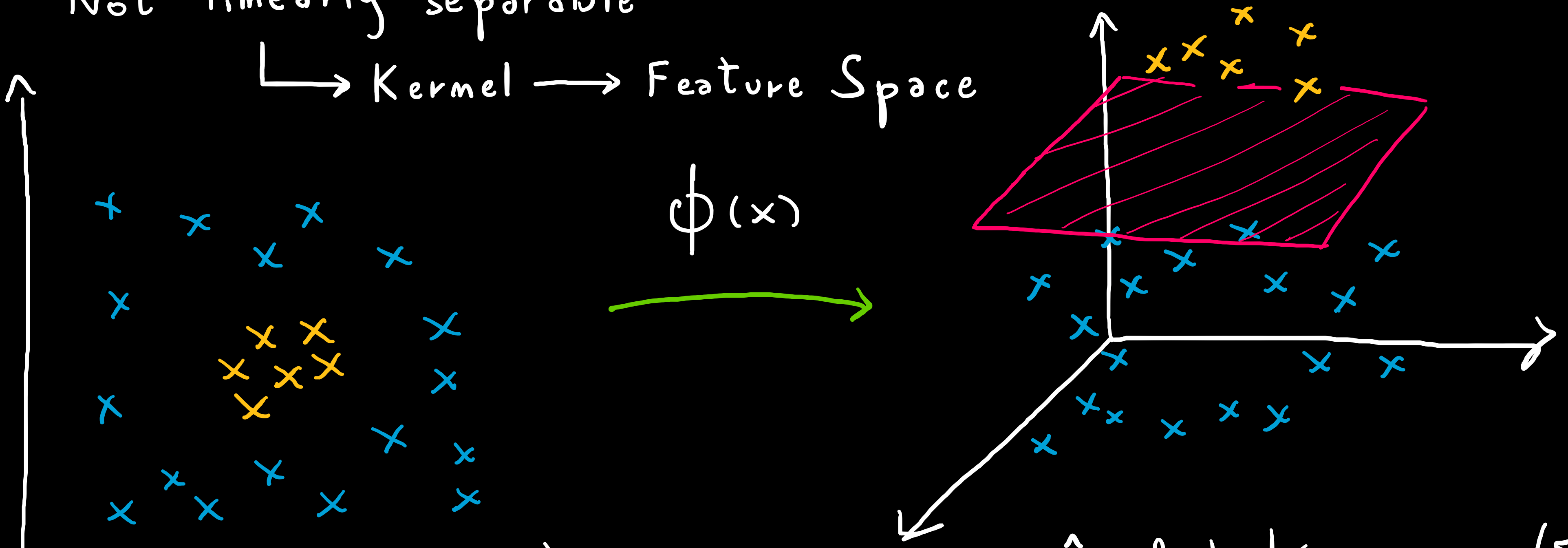


Classical Kernel Methods

Ex. Support Vector Machine (SVM)

Not linearly separable

└→ Kernel ⟶ Feature Space



$\phi(x_i)\,\phi(s)$

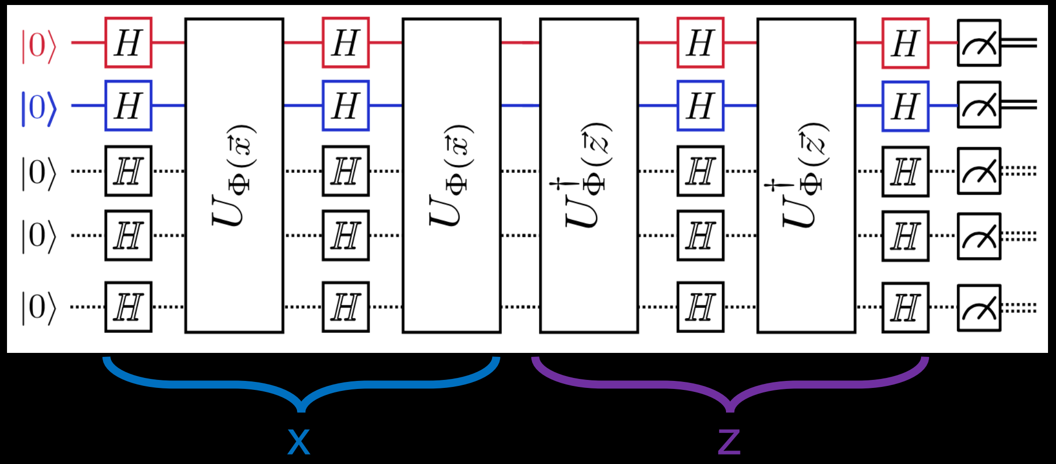
↑

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

Dual Form

Quantum SVM (QSVM)

- 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

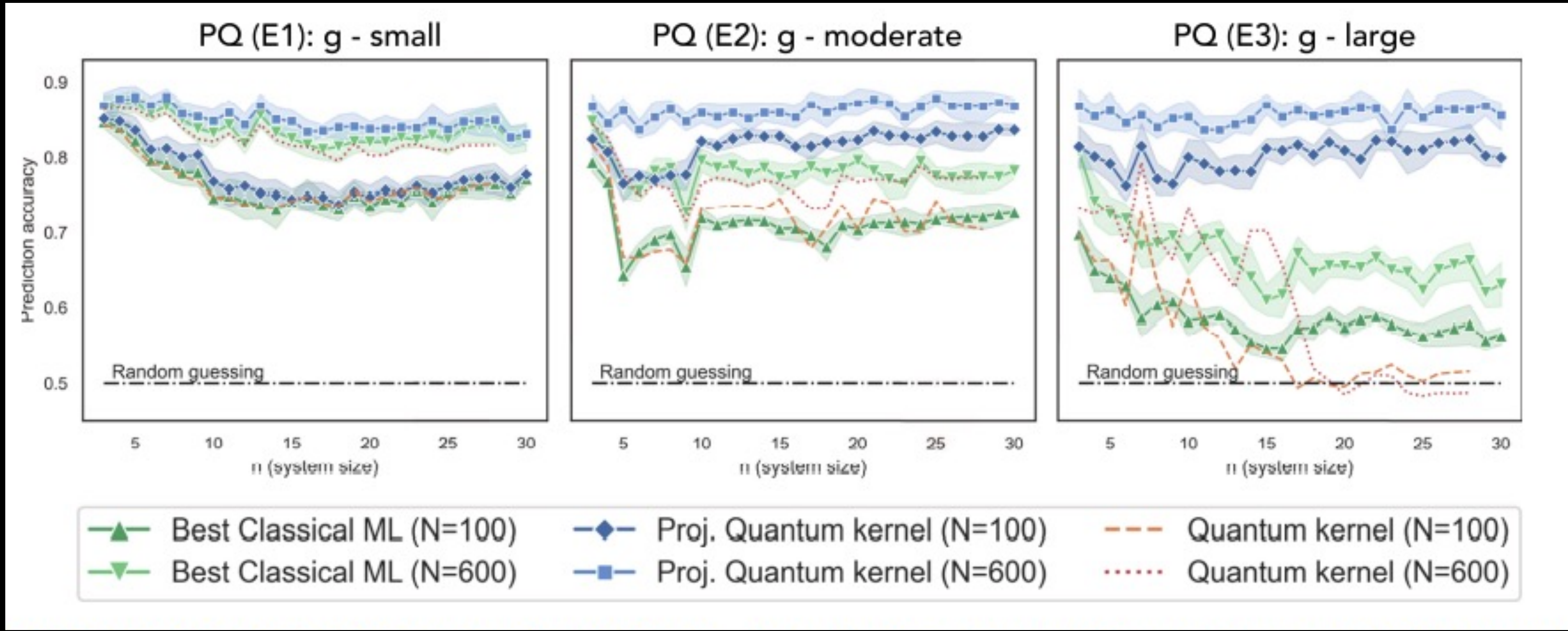
$$k^{lp}(x_i, x_j) = \sum_{k=1}^{m} \frac{Tr[\rho_k(x_i)\rho_k(x_j)]}{m}$$

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

- Improved generalizion while keeping features into states classically hard

- ***g**$_{CQ}$:* geometric difference between classical and quantum embeddings

Huang *et al.* Propose methodology to assess potential advantage according based on clomplexity analysis of data set and model



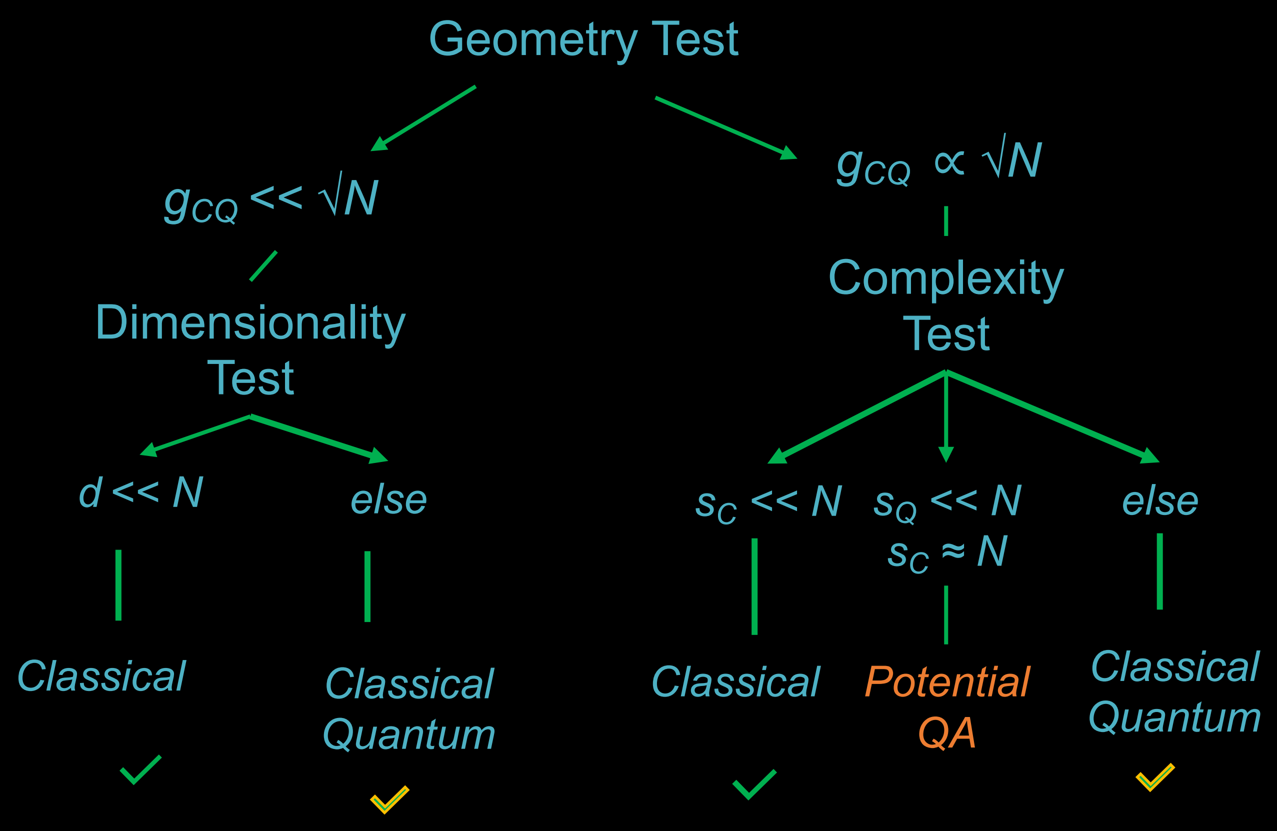
Huang, Hsin-Yuan, et al. "Power of data in quantum machine learning." *Nature communications* 12.1 (2021): 2631.

Quantum Advantage

Define an upper bound on classical and quantum kernels prediction error

$$\mathbb{E}_{\mathbf{x}}|h(\mathbf{x}) - y(\mathbf{x})| \leq \mathcal{O}\left(\sqrt{\frac{s_{K,\lambda}(N)}{N}}\right)$$

- N training events
- ***g_{CQ}****:* geometric difference between classical and quantum embeddings
- ***S* (N)**: model complexity
- **d** *:* feature space dimension

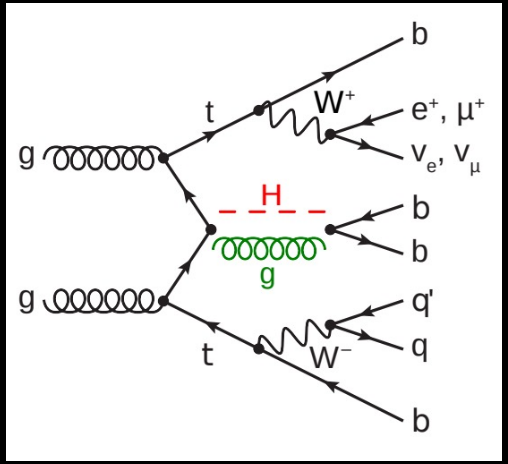


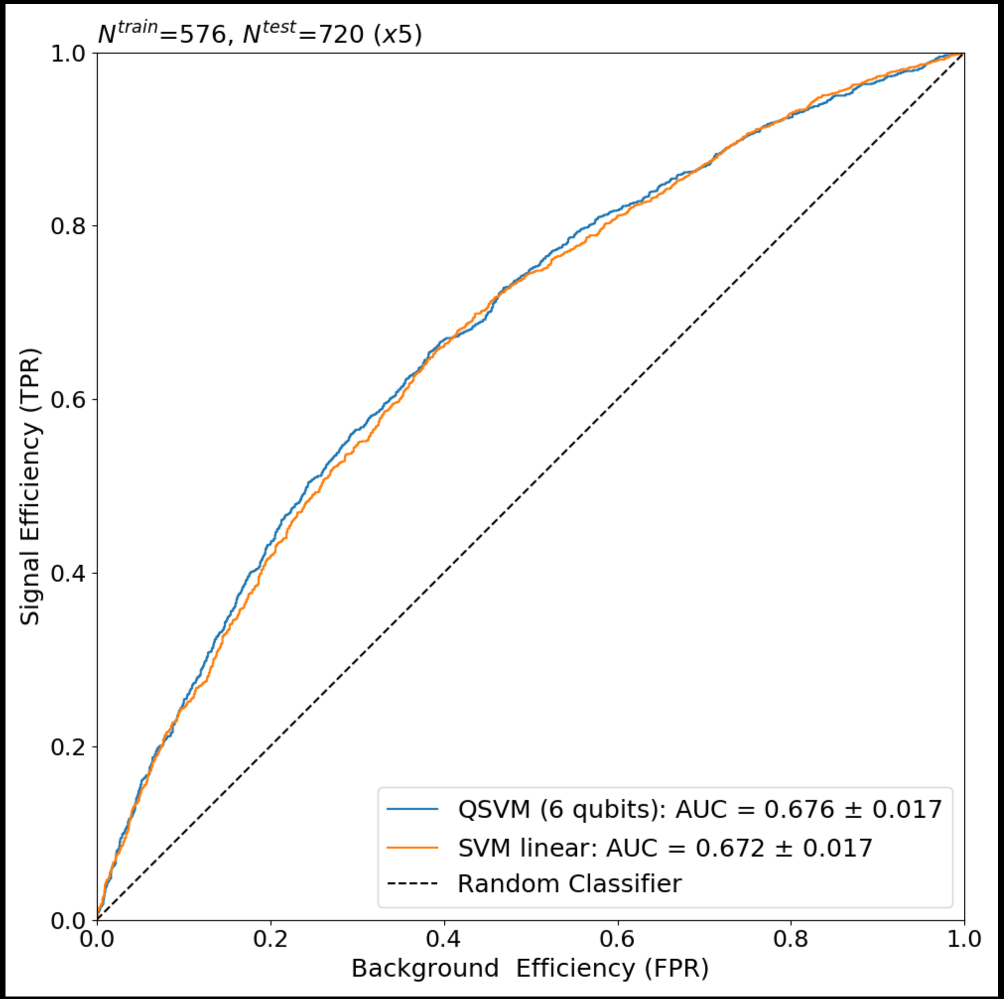
Constraints:

- **Encoding** maps of classical and quantum kernels
- **Data structure:** distribution function or dimensionality
- **Hyperparameters** choices

Higgs classification

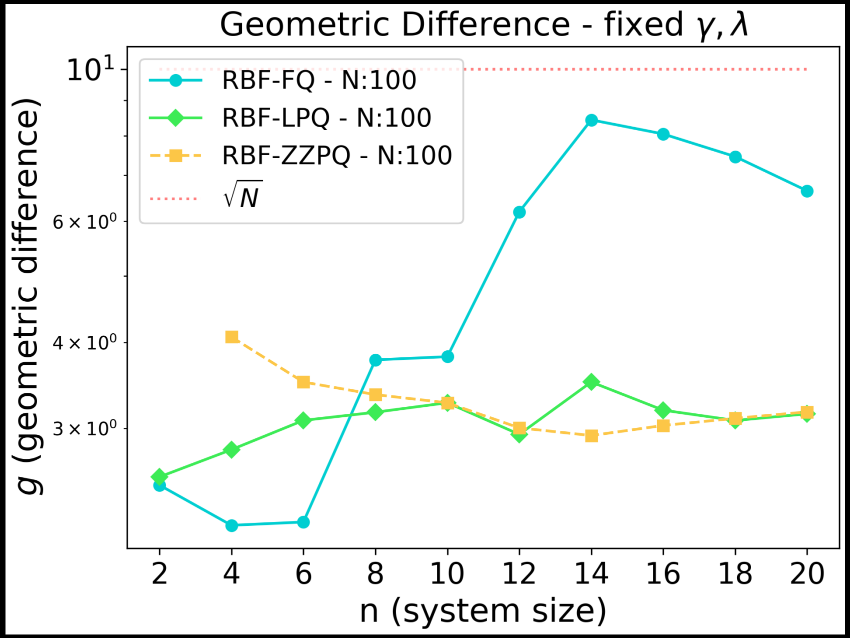
Quantum Support Vector Machine for the *ttH(bb)* event classification[5]

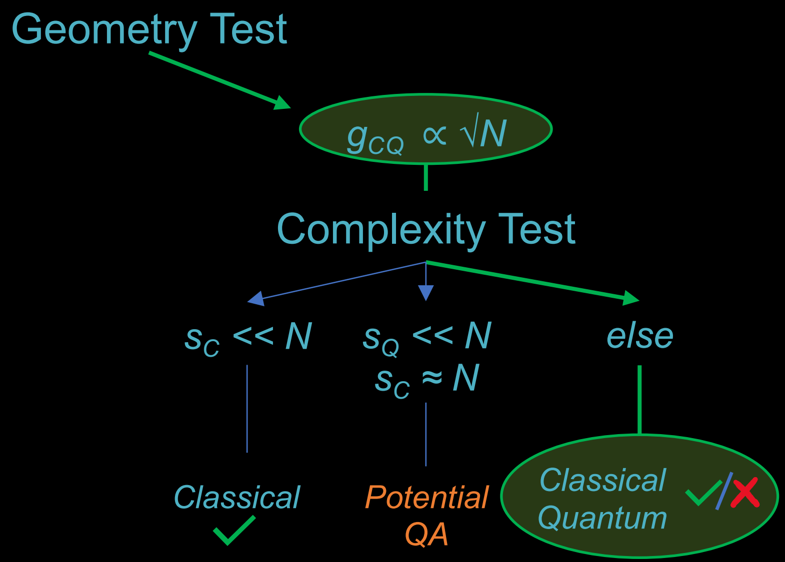


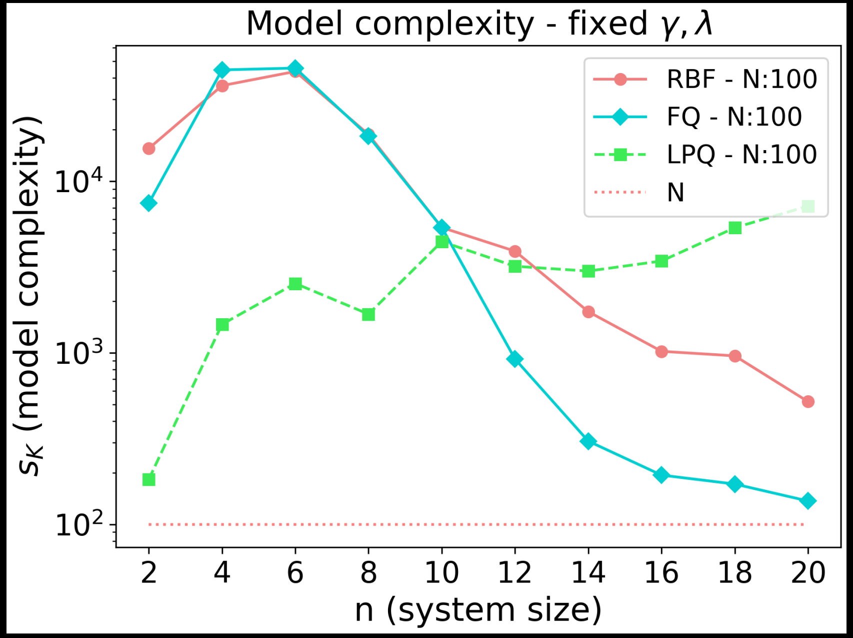


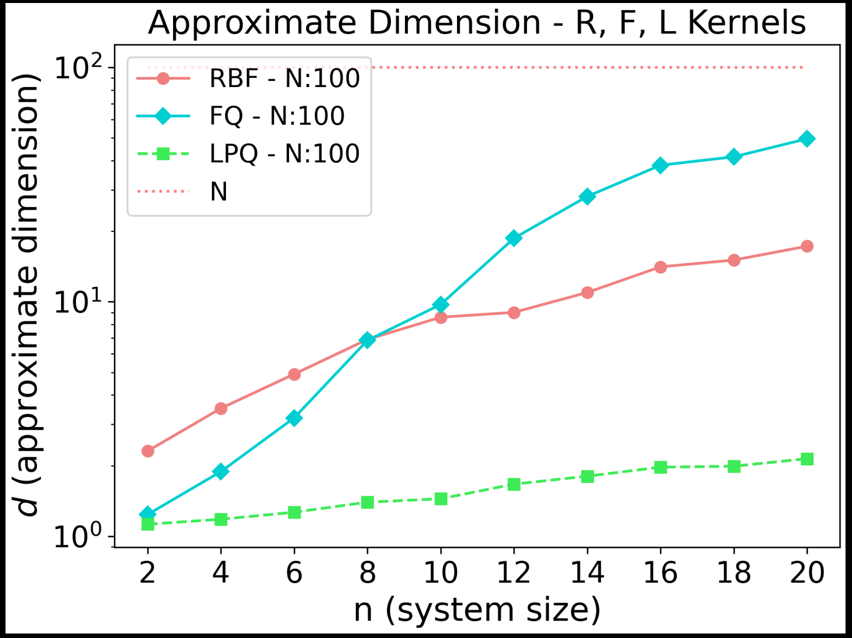
[1] V Belis et al, (2021), *Higgs Analysis with Quantum Classifiers,* EPJ Web Conf

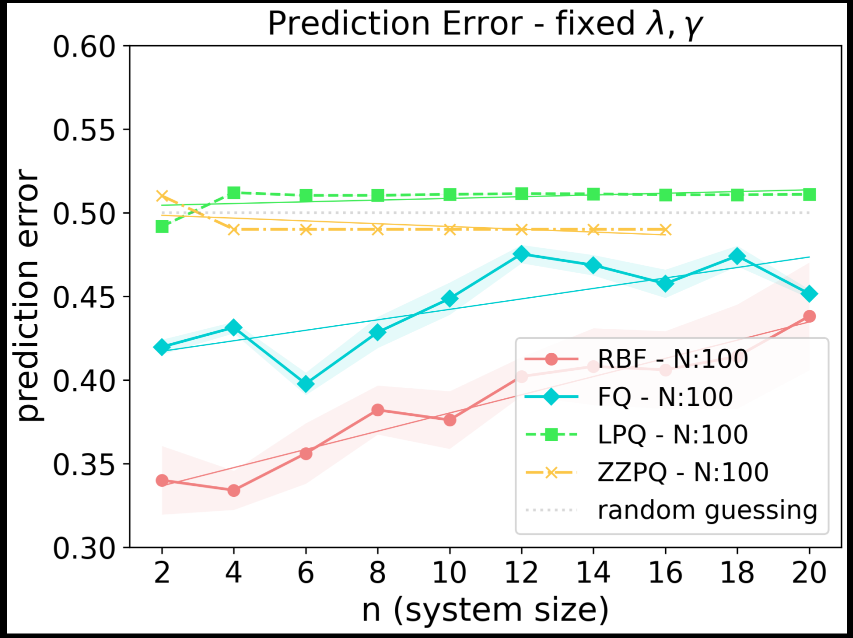
Results


Geometric Difference - fixed γ, λ
RBF-FQ - N:100
RBF-LPQ - N:100
RBF-ZZPQ - N:100
√N
g (geometric difference)
n (system size)


Geometry Test
g_CQ ∝ √N
Complexity Test
s_C << N
s_Q << N
s_C ≈ N
else
Classical
Potential QA
Classical
Quantum


Model complexity - fixed γ, λ
RBF - N:100
FQ - N:100
LPQ - N:100
N
s_K (model complexity)
n (system size)

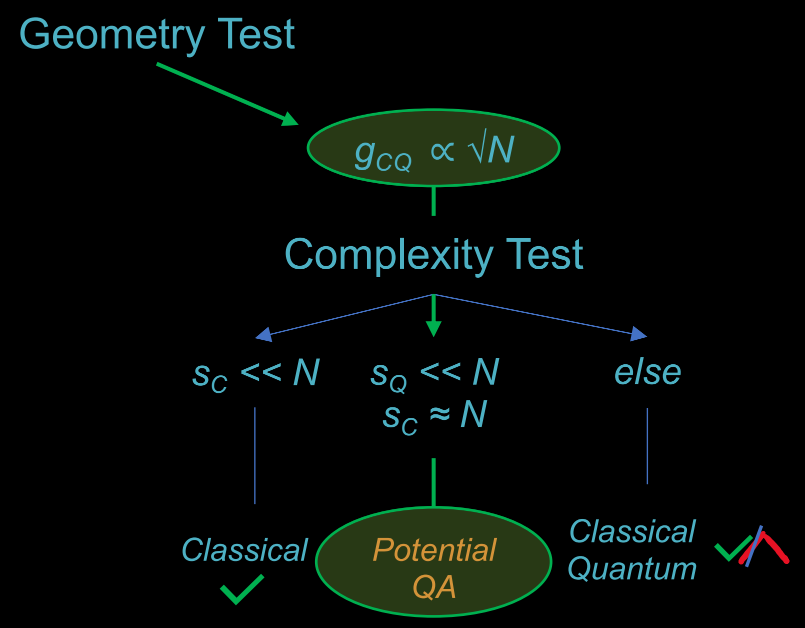

Approximate Dimension - R, F, L Kernels
RBF - N:100
FQ - N:100
LPQ - N:100
N
d (approximate dimension)
n (system size)

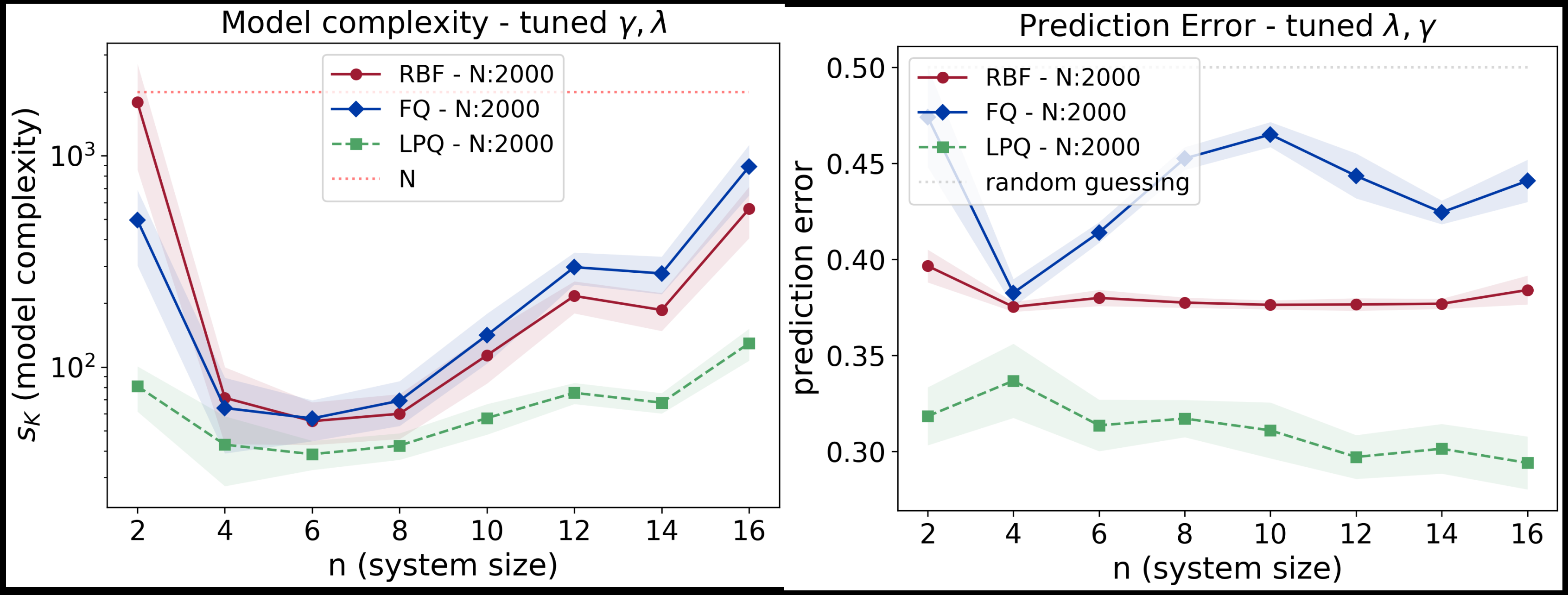

Prediction Error - fixed λ, γ
RBF - N:100
FQ - N:100
LPQ - N:100
ZZPQ - N:100
random guessing
prediction error
n (system size)

Results (tuned LPQ kernels)

Optimized quantum and classical kernels

- g_{CQ} moderate to √N
- s_C and s_Q moderate/comparable to N





Quantum works best!

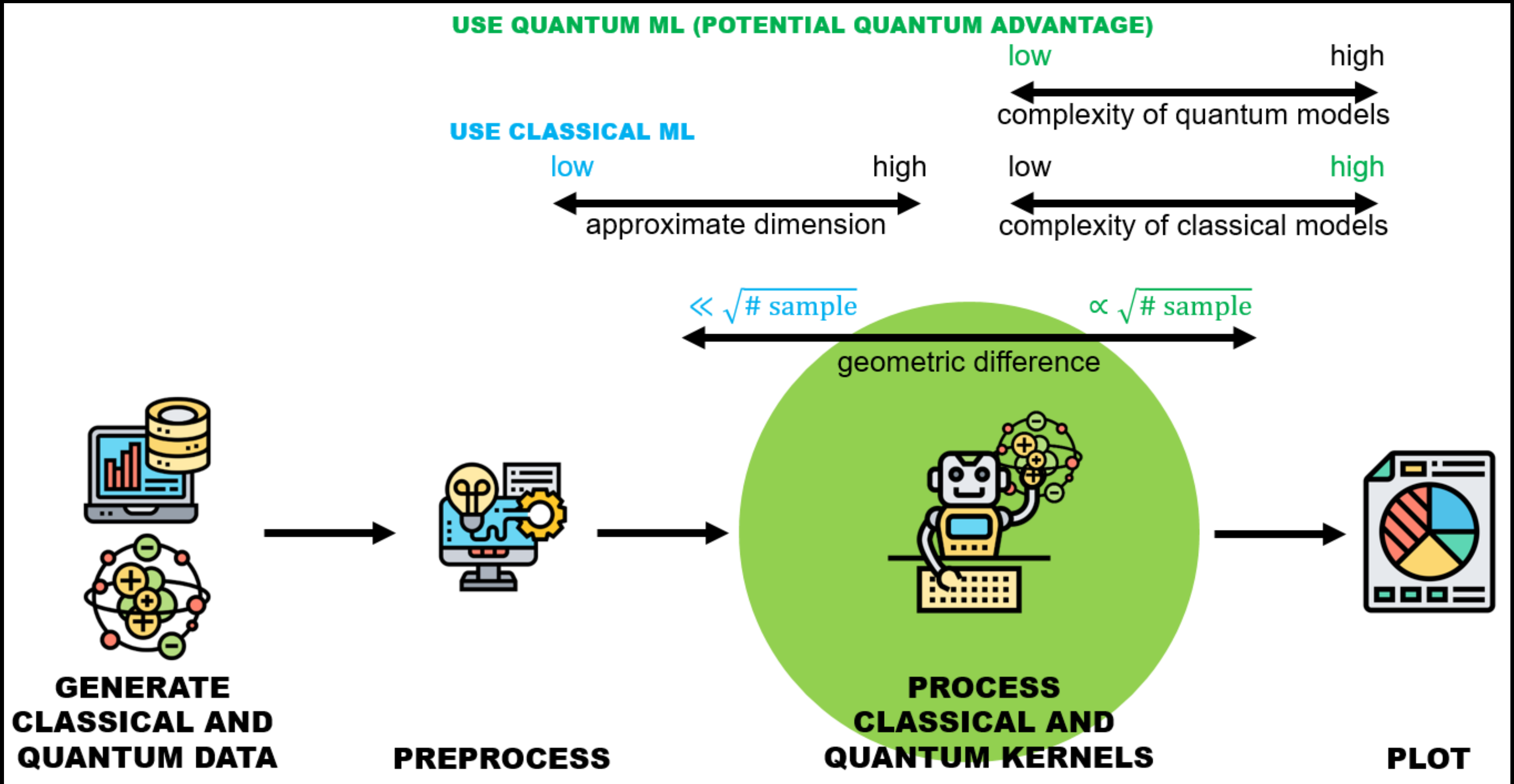
Summary

- Quantum Computing has the potential **to revolutionize many fields** including Machine Learning. However
- **Quantum Computing is not the optimal solution to «everything»**
- **Quantum Advantage is no free lunch**

- Higgs dataset **lives in low dimensional space**.
- We tuned projected kernel parameters to outperform classical RBF kernels
- It is important to **perform systematic studies on models and data properties to design ad-hoc quantum models**

QUASK: Quantum Advantage Seeker With Kernels

A priori methodology to assess quantum advantage according to data and kernels considered.



Di Marcantonio, **MI**, et al. "QuASK--Quantum Advantage Seeker with Kernels." (2022).

Thanks!

Sofia.Vallecorsa@cern.ch

Model Convergence and Barren Plateau

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

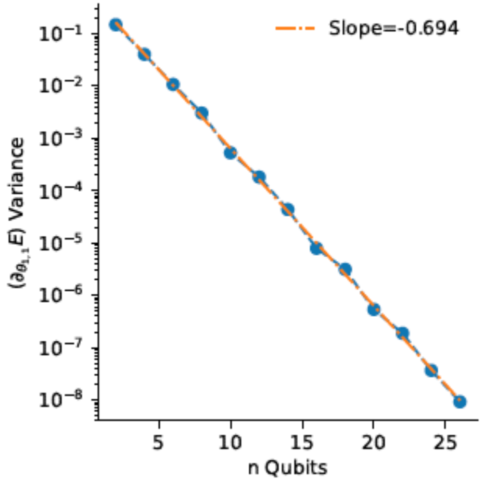
Classical gradients **vanish exponentially** with the number of layers (J. McClean *et al.*, arXiv:1803.11173)

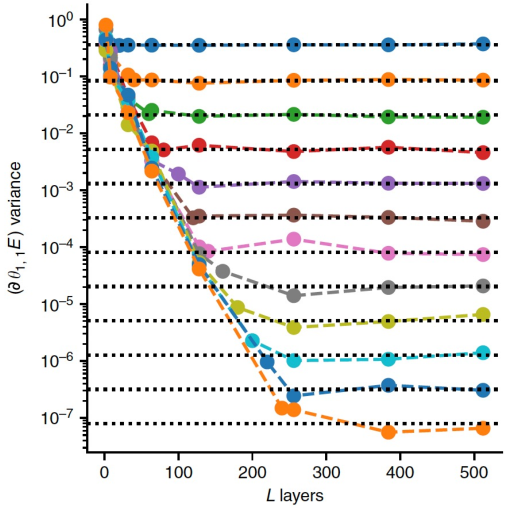
- Convergence still possible if gradients consistent between batches.

Quantum gradient decay exponentially in the number of qubits

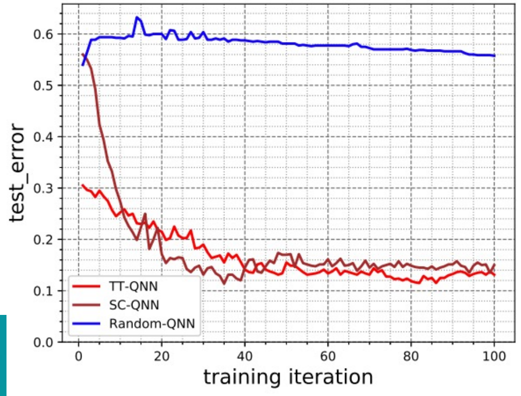
- Random circuit initialization
- Loss function locality in shallow circuits (M. Cerezo *et al.*, arXiv:2001.00550)
- Ansatz choice: TTN, CNN (Zhang *et al.*, arXiv:2011.06258, A Pesah, *et al.*, *Physical Review X* 11.4 (2021): 041011.)
- Noise induced barren plateau (Wang, S *et al.*, Nat Commun 12,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

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

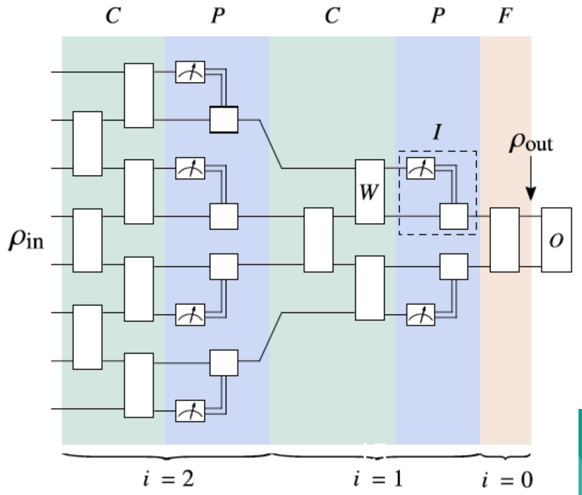




TTN for MNIST classification (8 qubits), Zhang *et al.*, arXiv:2011.06258



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

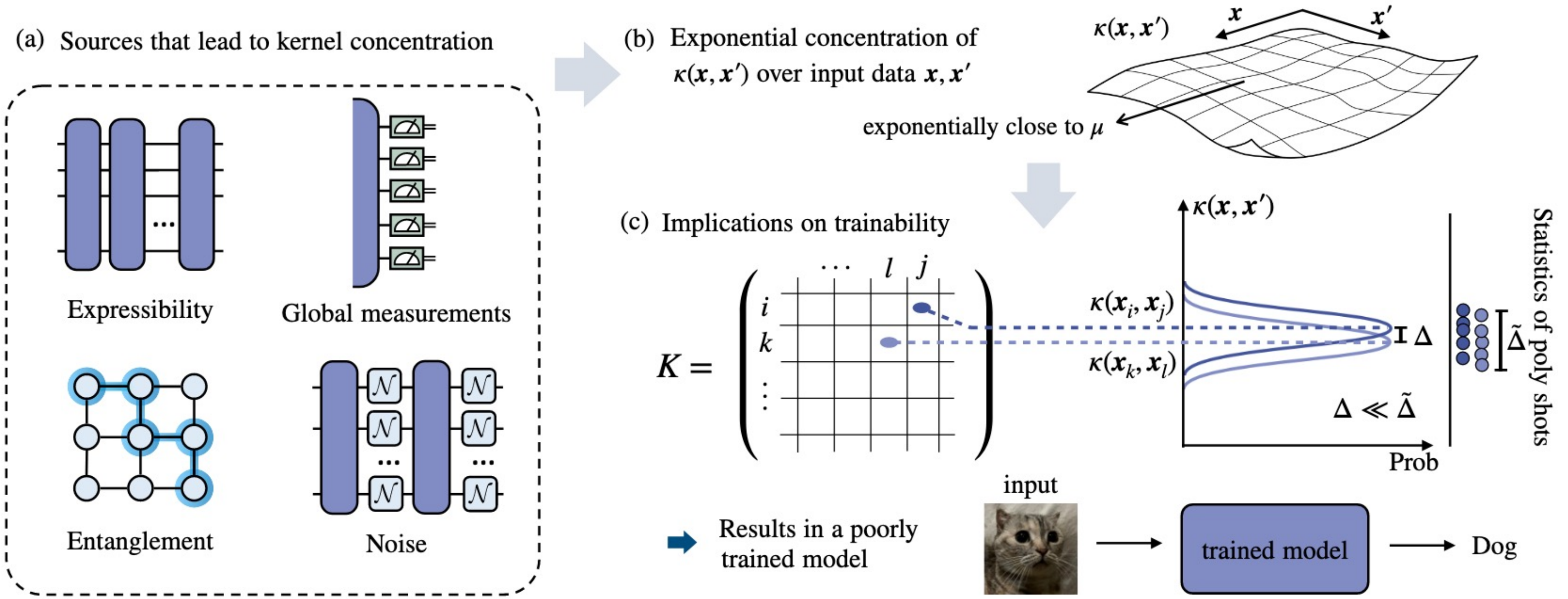


Figure 1. **Kernel concentration and its implications on trainability:** The exponential concentration (in the number of qubits n) of quantum kernels $\kappa(\boldsymbol{x}, \boldsymbol{x}')$, over all possible input data pairs $\boldsymbol{x}, \boldsymbol{x}'$, can be seen to stem from the difficulty 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 AD model (arxiv:2208.11060)