Physics searches with quantum classifiers and anomaly detection at the LHC

CERN openlab Technical Workshop, 23 March 2022

Vasilis Belis (PhD at ETH, CERN)

Collaborators:

ETH: P. Odagiu, F. Reiter, G. Dissertori.

CERN: K. Wozniak, M. Pierini, M. Grossi, S. Vallecorsa.

IBM Zurich: P. Barkoutsos, I. Tavernelli.

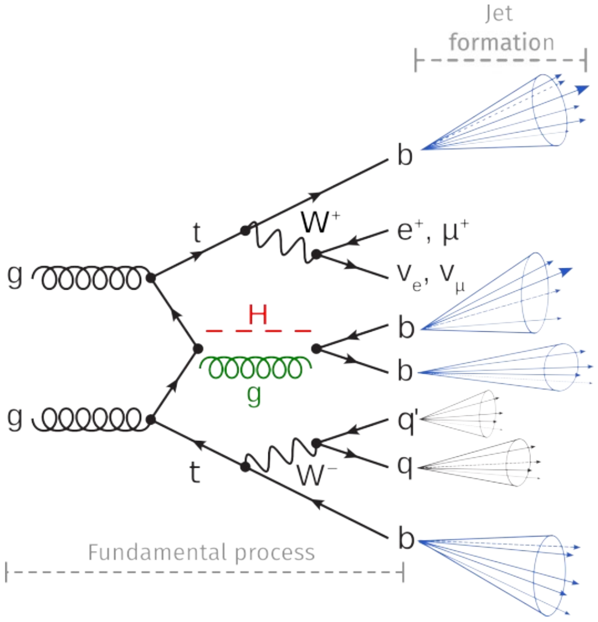
ETH zürich

QUANTUM TECHNOLOGY INITIATIVE

Examples of studied physics processes

Classification - Supervised

Higgs boson production $t\bar{t}H(b\bar{b})$

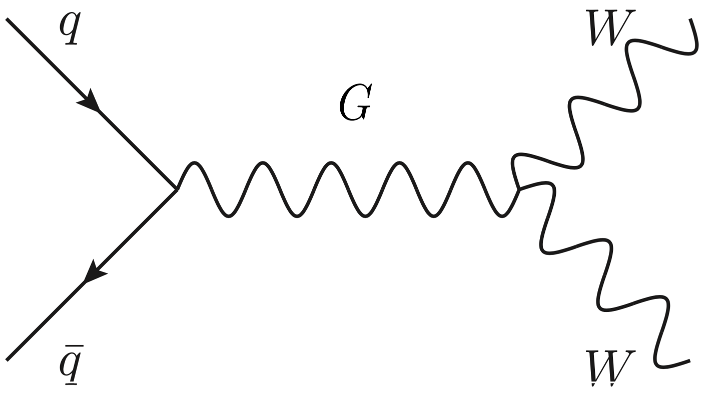


Model-dependent search: New physics **signature known.**

$$n^{\text{features}} = \underbrace{7 \times 8}_{\text{jets}} + \underbrace{1 \times 7}_{\text{lepton}} + \underbrace{1 \times 4}_{\text{MET}} = 67$$

Anomaly detection – (Un)Supervised

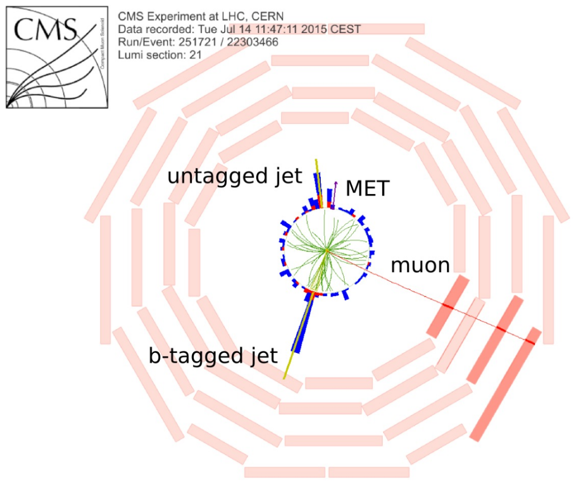
Graviton production (mJJ spectrum)



Model-agnostic search: New physics **signature unknown.**

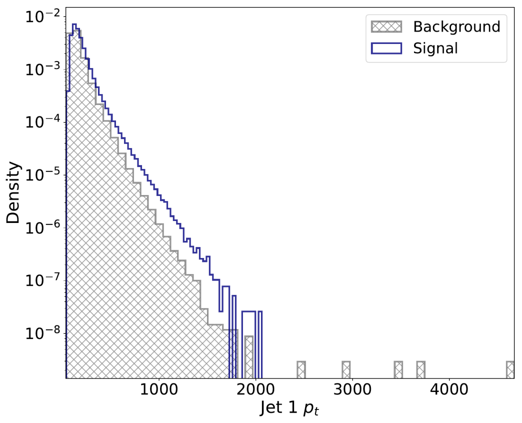
$$n^{\text{features}} = \underbrace{100}_{\text{particles}} \times \underbrace{3}_{\text{features}} = 300$$

Workflow and motivation



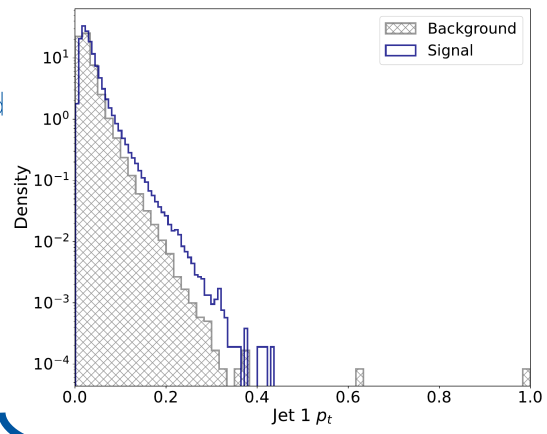
MC Simulation

compute prob. distributions →



•••

apply cuts and normalise →



•••

Input to statistical models

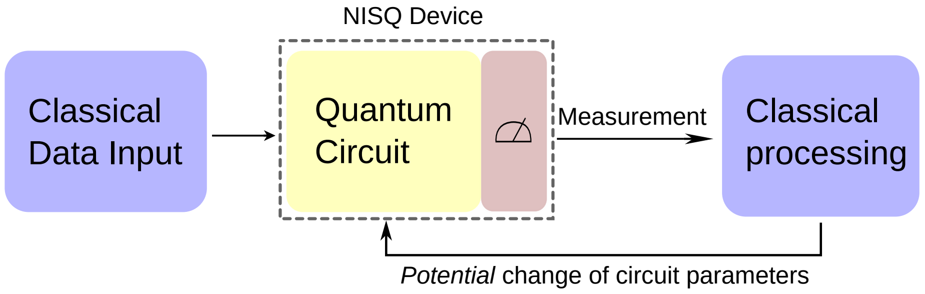
Why quantum machine learning for particle physics?

Fundamental motivation: can quantum models utilise the quantum correlations inherent in HEP data leading to performance advantages?

- Find inductive bias based on prior knowledge on the data generation (quantum process for HEP data).

- If the bias can be constructed and is classically difficult to simulate → quantum advantage.

Hybrid Quantum-Classical algorithms



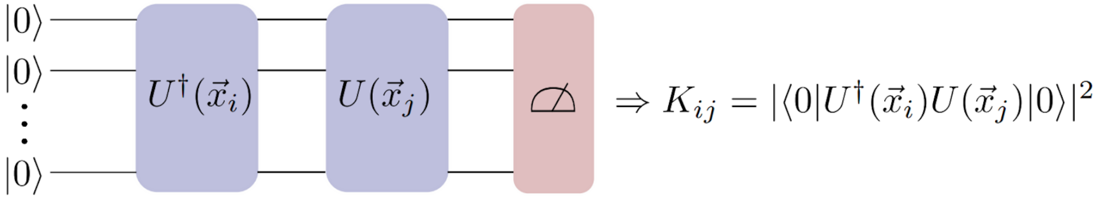
Noisy Intermediate Scale Quantum (NISQ) devices:

- Circuit width: limited number of qubits (superconducting qubits at IBM up to 127).

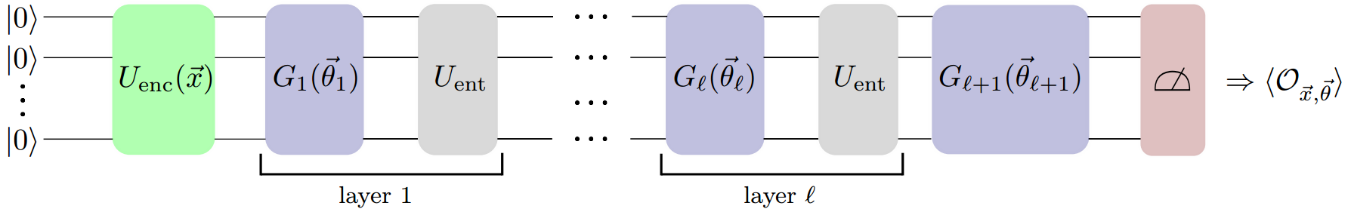
- Circuit depth: limited number of operations per qubit (small decoherence times).

QML models for classification (blueprints)

Quantum Support Vector Machines (QSVM)



Variational Quantum Circuits (VQC)
Quantum "Neural Networks"

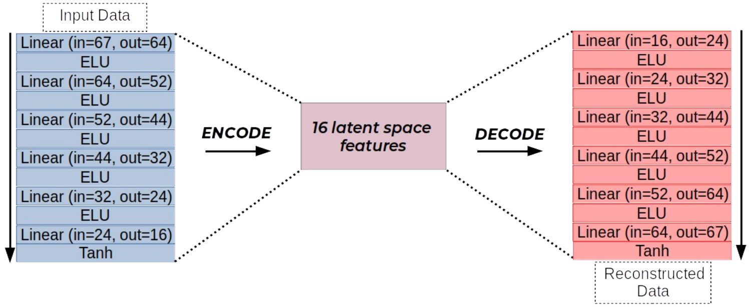


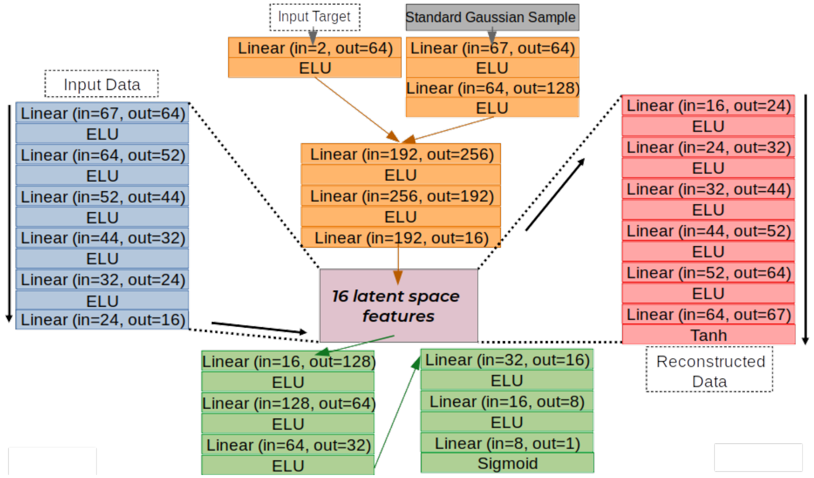
To accommodate for NISQ limitations, feature reduction is needed: (Hybrid) Autoencoder models.

Advanced feature reduction techniques

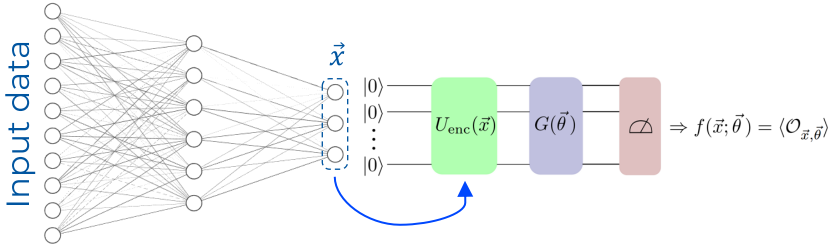
Goal: Reduce the number of features while preserving the discrimination power of the original physical features.

- Investigated the power of conventional ML dimensionality reduction techniques (e.g., manifold learning).
- Conclusion: Deep Learning techniques based on (Hybrid) Autoencoders (physics aware) are superior.





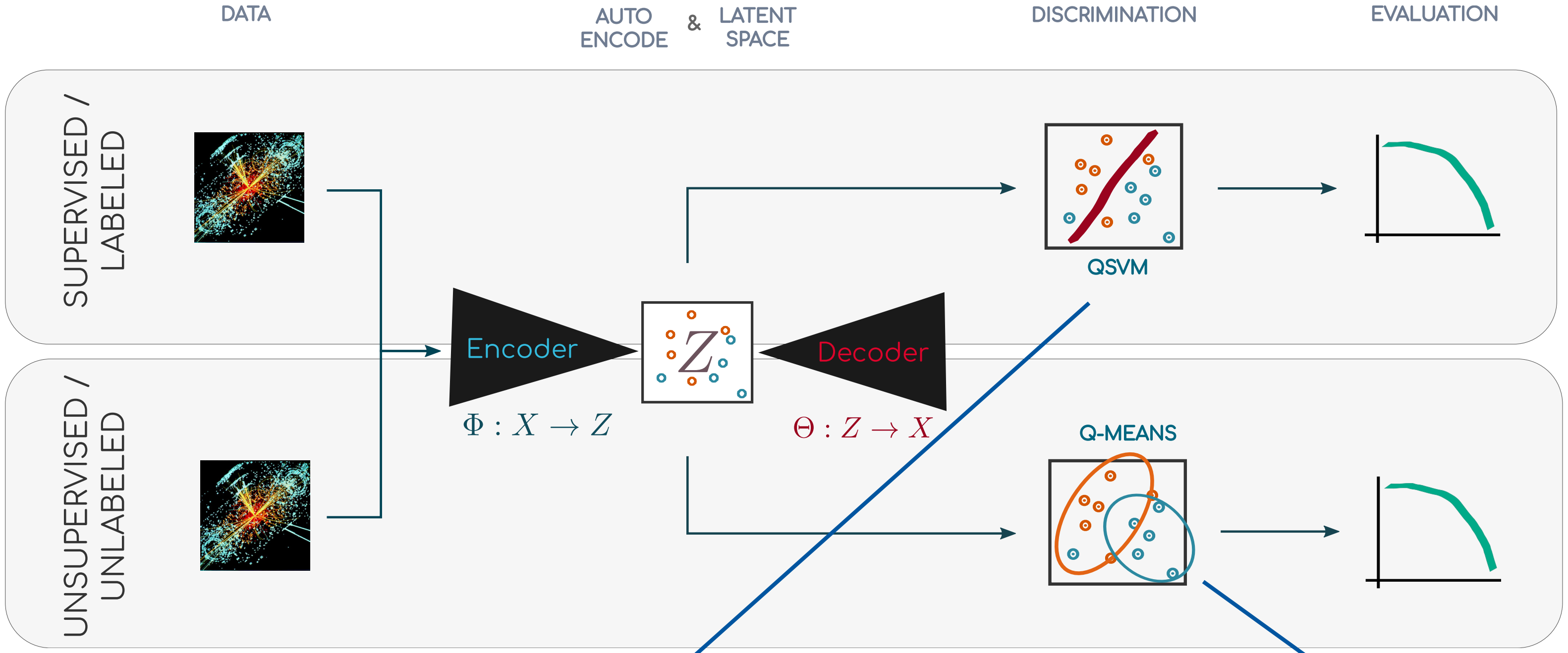
Assess the power of hybrid quantum-classical networks:



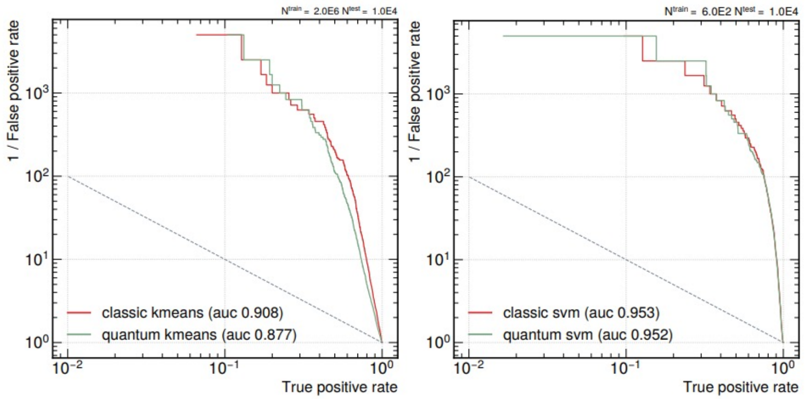
Able to match state-of-the art performance on ideal quantum simulations:

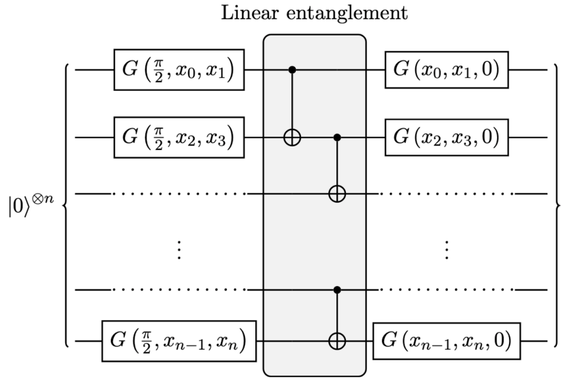
- QSVM on the latent space of particle physics events.

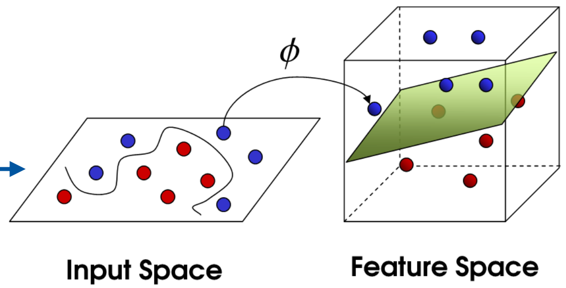
Identifying new physics (graviton) via anomaly detection



Performance







- Distance from cluster calculated between quantum states.
- Minimisation of the distance with a quantum algorithm.

Conclusions and outlook

Summary:

- State-of-the-art performance of the developed hybrid data compression models.
- Feature reduction is crucial, training classical + quantum at the same time yields better results (hybrid VQC) than step-wise training.

Ongoing/Future work:

- Investigate other physics signatures and processes.
- Running the models on real hardware and assess the need for error mitigation.
- Anomaly detection (AD) for model independent searches of new physics using kernel-based models.
- Quantum branches (QSVM+VQC) on developed networks for feature reduction and AD.

Thank you!

More information: *Higgs analysis with quantum classifiers*, EPJ Web Conf., 251 (2021) 03070, https://doi.org/10.1051/epjconf/202125103070, pre-print: arXiv:2104.07692.

In preparation:
- Classification and anomaly detection in the latent space of high energy physics events.
- Hybrid Autoencoder and VQC studies for optimal feature reduction.

CERN

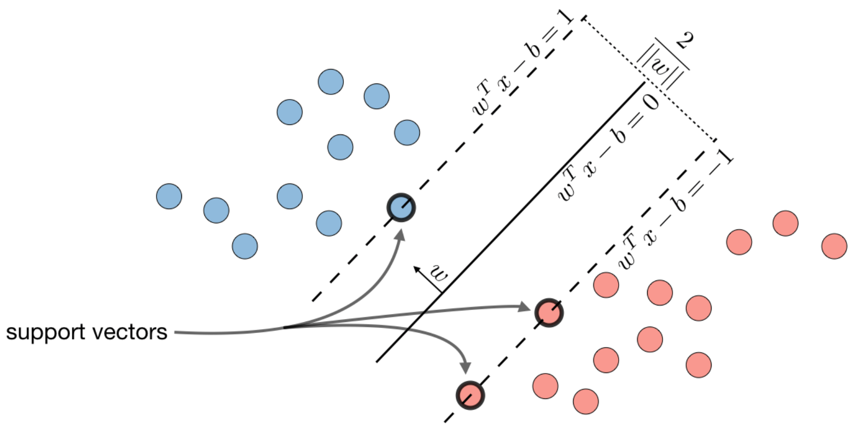
References

[1] Sjöstrand T *et al.* 2015 *Comput. Phys. Commun.* **191** 159–177 (*Preprint* 1410.3012)
[2] de Favereau J *et al.* (DELPHES 3) 2014 *JHEP* **02** 057 (*Preprint* 1307.6346)
[3] Randall L and Sundrum R 1999 *Phys. Rev. Lett.* **83** 3370 (*Preprint* hep-ph/9905221)
[4] Fan H, Su H and Guibas L 2017 *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* p 2463 (*Preprint* 1612.00603)
[5] Lloyd S, Mohseni M and Rebentrost P 2013 Quantum algorithms for supervised and unsupervised machine learning (*Preprint* 1307.0411)
[6] Aïmeur E, Brassard G and Gambs S 2006 *Advances in Artificial Intelligence* ed Lamontagne L and Marchand M (Berlin, Heidelberg: Springer Berlin Heidelberg) pp 431–442 ISBN 978-3-540-34630-2
[7] Durr C and Hoyer P 1999 A quantum algorithm for finding the minimum (*Preprint* quant-ph/9607014)
[8] Boyer M, Brassard G, Høyer P and Tapp A 1998 *Fortschritte der Physik* **46** 493–505 ISSN 1521-3978
[9] Grover L K 1996 A fast quantum mechanical algorithm for database search (*Preprint* quant-ph/9605043)
[10] Boser B E, Guyon I M and Vapnik V N 1992 *Proceedings of the fifth annual workshop on Computational learning theory*
[11] Schuld M and Killoran N 2019 *Physical Review Letters* **122** ISSN 1079-7114 URL http://dx.doi.org/10.1103/PhysRevLett.122.040504
[12] Havlíček V, Córcoles A, Temme K and et al 2019 *Nature* **567** 209–212
[13] LaRose R and Coyle B 2020 *Phys. Rev. A* **102**(3) 032420
[14] Belis V, González-Castillo S, Reissel C, Vallecorsa S, Combarro E, Dissertori G and Reiter F 2021 *EPJ Web Conf.* **251** 03070

CERN

BACK-UP

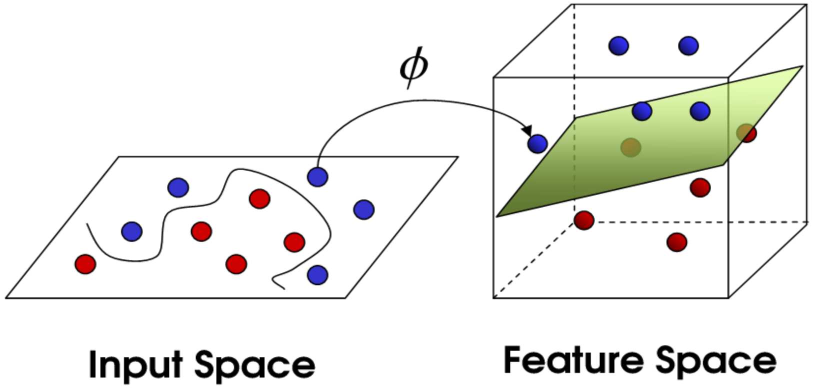
Support Vector Machines



SVM objective function is equivalent to (dual Lagrangian):

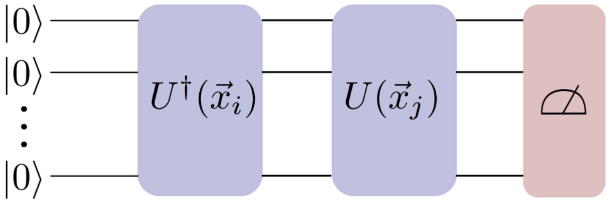
maximize $L(c_1 \dots c_n) = \sum_{i=1}^{n} c_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i c_i (\vec{x}_i \cdot \vec{x}_j) y_j c_j$

subject to $\sum_{i=1}^{n} c_i y_i = 0$, and $0 \leq c_i \leq C$ for all i.



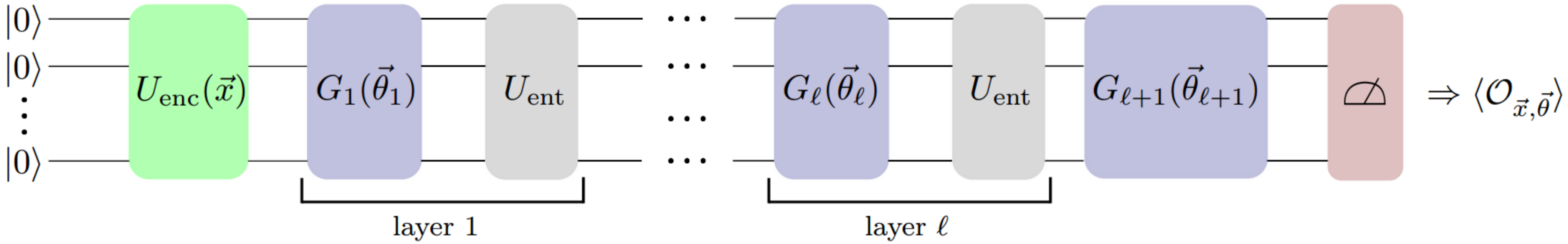
Kernel trick: $(\vec{x}_i \cdot \vec{x}_j) \mapsto k(\vec{x}_i, \vec{x}_j) := \phi(\vec{x}_i) \cdot \phi(\vec{x}_j).$

Make the kernel *quantum*:



$\Rightarrow K_{ij} = |\langle 0 | U^\dagger(\vec{x}_i) U(\vec{x}_j) | 0 \rangle|^2$

Variational Quantum Circuits



- Data embedding circuit (feature map) here is fixed.
- Layers of parametrised quantum gates → trainable parameters.
- Output of the model → expectation value of an observable on the prepared state $|\psi(\vec{x},\vec{\theta})\rangle$
  e.g. measure the first qubit on the computational basis

$$\mathcal{O} = \sigma_z \otimes \mathbb{1} \otimes \mathbb{1} \cdots \otimes \mathbb{1},$$

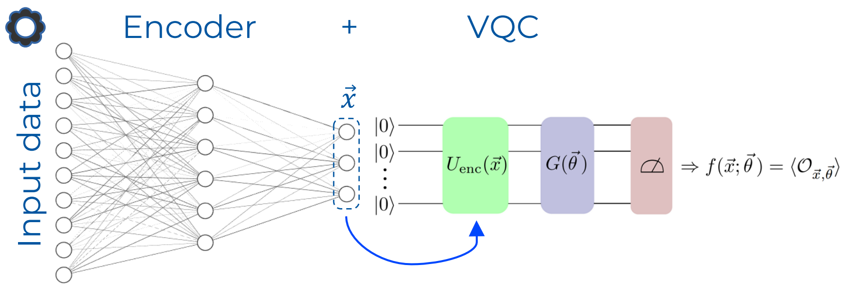
$$f(\vec{x},\vec{\theta}) = \langle \psi(\vec{x},\vec{\theta})|\, \mathcal{O}\, |\psi(\vec{x},\vec{\theta})\rangle \equiv \langle \psi(\vec{x})|\, G^\dagger(\vec{\theta}) \mathcal{O} G(\vec{\theta})\, |\psi(\vec{x})\rangle \equiv \langle \mathcal{O} \rangle_{\vec{x},\vec{\theta}}.$$

- Classification: if $\langle \mathcal{O} \rangle_{\vec{x},\vec{\theta}} > 0$ → signal, otherwise background.

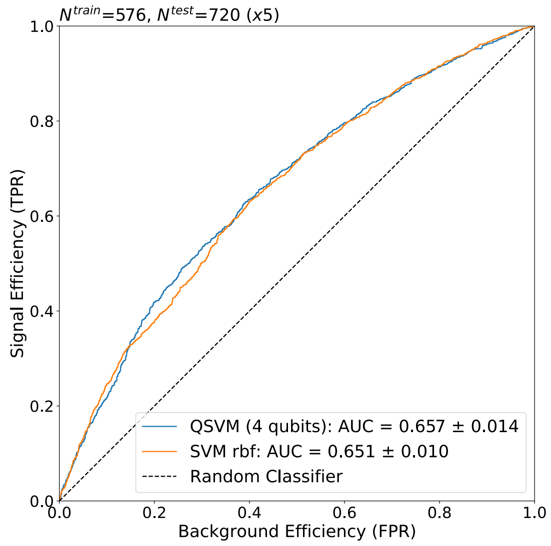
CERN

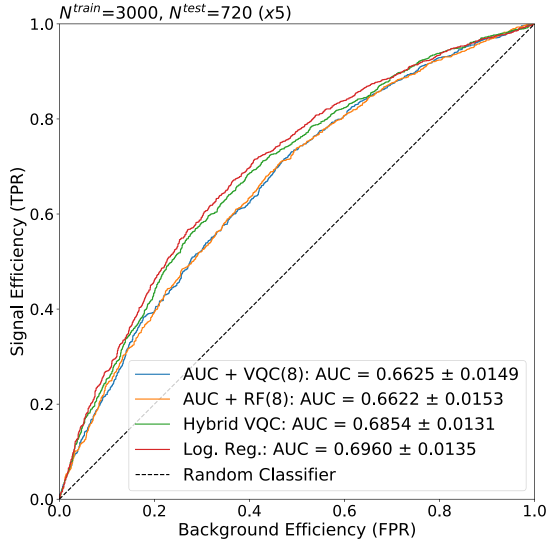
AE, QSVM, and *Hybrid VQC

Autoencoder	HP Optimisation	MSE Loss $\times 10^{-4}$	BCE Loss	Classifier AUC	QSVM AUC
Vanilla	-	4.77	-	-	0.56± 0.01
Variational	MSE	4.49	-	-	0.56± 0.02
Classifier	MSE	5.47	0.63	0.700± 0.001	0.56± 0.02
	BCE	62.97	0.61	0.734± 0.002	0.72± 0.01
Sinkhorn	MSE	9.65	-	-	0.51± 0.01
Sinkclass	MSE	26.41	0.65	0.642± 0.003	0.50± 0.01
	BCE	24.69	0.61	0.734± 0.002	0.74± 0.01



Model	BCE Loss	AUC
Encoder + VQC	0.61	0.702± 0.004





Sinkclass AE shows best performance when considering both *reconstruction power* and *classification power*.

It even matches the classical *state-of-the-art* result!

Results from: https://doi.org/10.1051/epjconf/202125103070

CERN