QML in the latent space of HEP events

Kinga Anna Woźniak, Vasilis Belis, Ema Puljak, Maurizio Pierini, Sofia Vallecorsa, Michele Grossi, Günther Dissertori, Panagiotis Barkoutsos and Ivano Tavernelli

Introduction: The Challenge

Application: Searches at LHC

Finding Signal in a dataset dominated by Background

Look in mJJ resonance spectrum

2 Scenarios:

	Supervised	Unsupervised
Truth	known	unknown
Training Data	MC with signal model	Data
Search	model-dependent	model-agnostic

Goal:

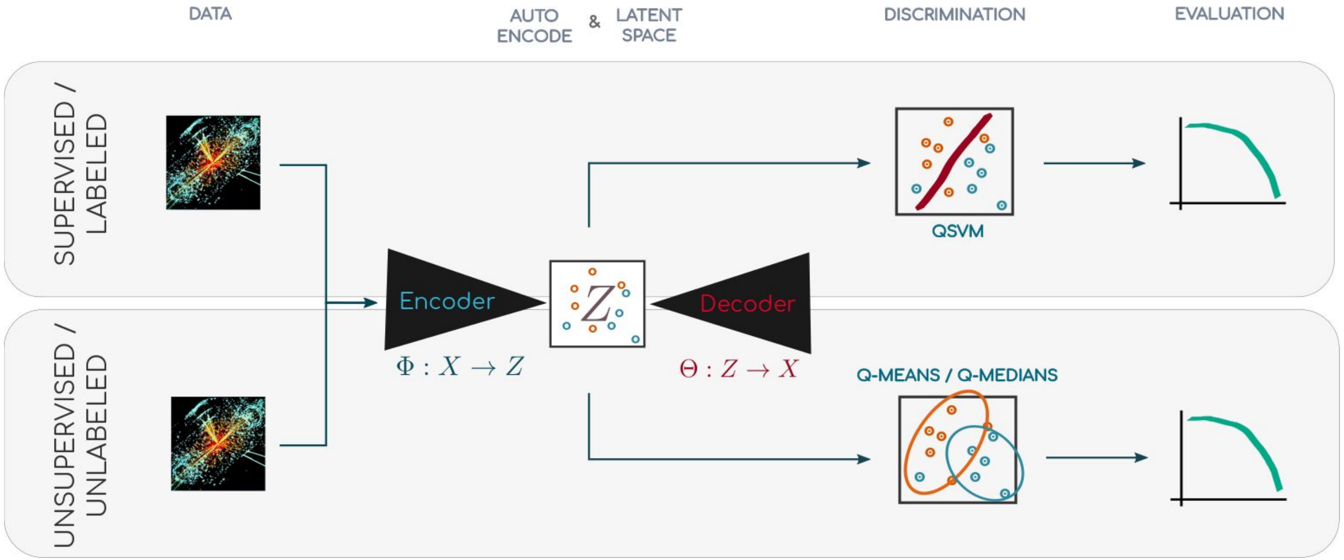
- Compare classic to quantum algorithm performance
- Study impact of latent dimension and training size

Rationale for Quantum:

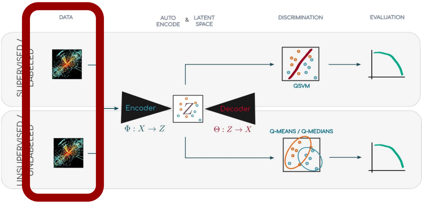
Improved accuracy (because data intrinsically quantum, quantum can find patterns that classic can't)

Introduction: Workflow

- Data: Reduce dimensionality of input to make treatable by noisy quantum computers through Autoencoder
- Algorithms:
  1. SVM Classification for supervised scenario
  2. K-Means / K-Medians for unsupervised scenario
- Evaluation: Signal- vs Background-Accuracy



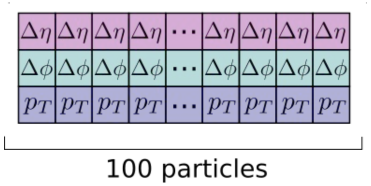
Data & Quantum Embedding



Input:

Dijet Events

Particle list (Δη, Δϕ, ρt)



Autoencoder Training:

Define data sideband (dominated by BG) as |Δη| > 1.4

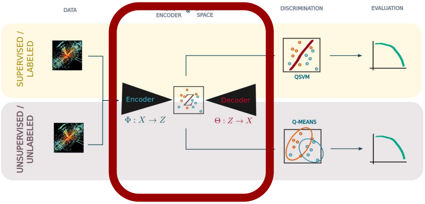
Encoding inputs into quantum state

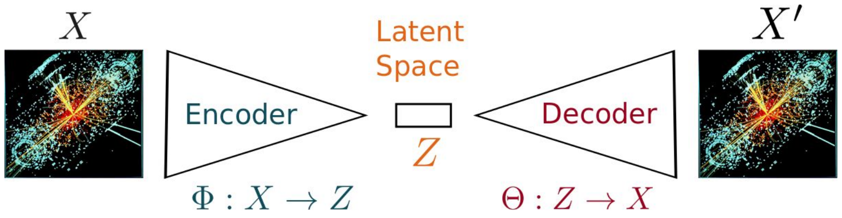
- Amplitude encoding (Q-means and Q-medians)
- Dense angle encoding (QSVM)

Training and Testing

- AE train: QCD **sideband** (2M events)
- Clustering train: QCD **signalregion**
- Clustering test: QCD **signalregion** (10K events)

Autoencoding for Dimensionality Reduction





Originally designed to compress and decompress inputs, passing through bottleneck (latent space)

Idea

Make AE learn how to compress BG, it will fail when seeing SIG event (reconstruction error)

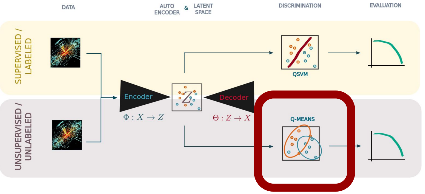
Architecture: Convolutional + Dense Layers

Latent Activation: tanh, dimensionality variation $Z \in \mathbb{R}^8, \mathbb{R}^{16}, \mathbb{R}^{32}$

Loss Metric: **Chamfer-Loss** / Pairwise distance

$$L_R = \sum_{i \in input} \min_j((x^{(i)} - x^{(j)})^2) + \sum_{j \in output} \min_i((x^{(j)} - x^{(i)})^2)$$

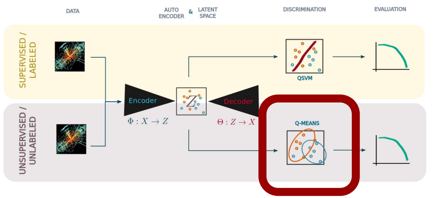
Unsupervised Clustering: Q-MEANS



Algorithm in 3 parts:

1) **Quantum** distance calculation: distance to cluster
2) **Quantum** minimization (Grover / Duerr & Hoyer): closest cluster assignment
3) New cluster center calculation (**classic**)

Unsupervised Clustering: Q-MEANS



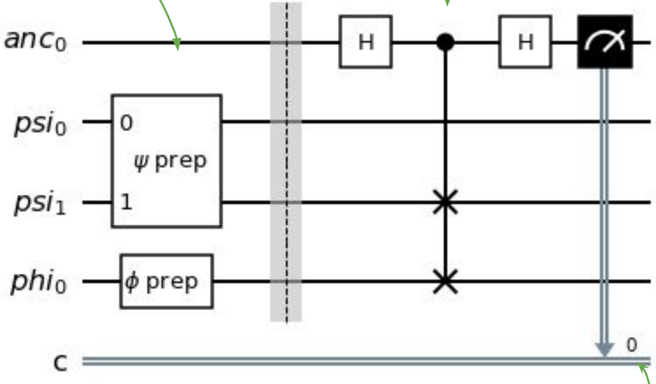
1) Quantum **distance calculation**: distance to cluster

Prepare 2 quantum states

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0,u\rangle + |1,v\rangle)$$

$$|\phi\rangle = \frac{1}{\sqrt{Z}}(|u|\,|0\rangle - |v|\,|1\rangle)$$

$$Z = |u|^2 + |v|^2$$



Do Swap Test

$$|x_0\rangle = |0,a,b\rangle$$

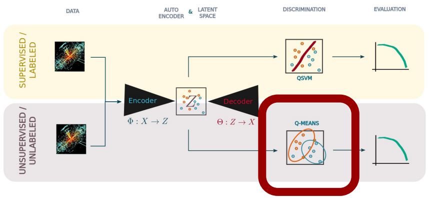
$$|x_1\rangle = \frac{1}{\sqrt{2}}(|0,a,b\rangle + |1,a,b\rangle)$$

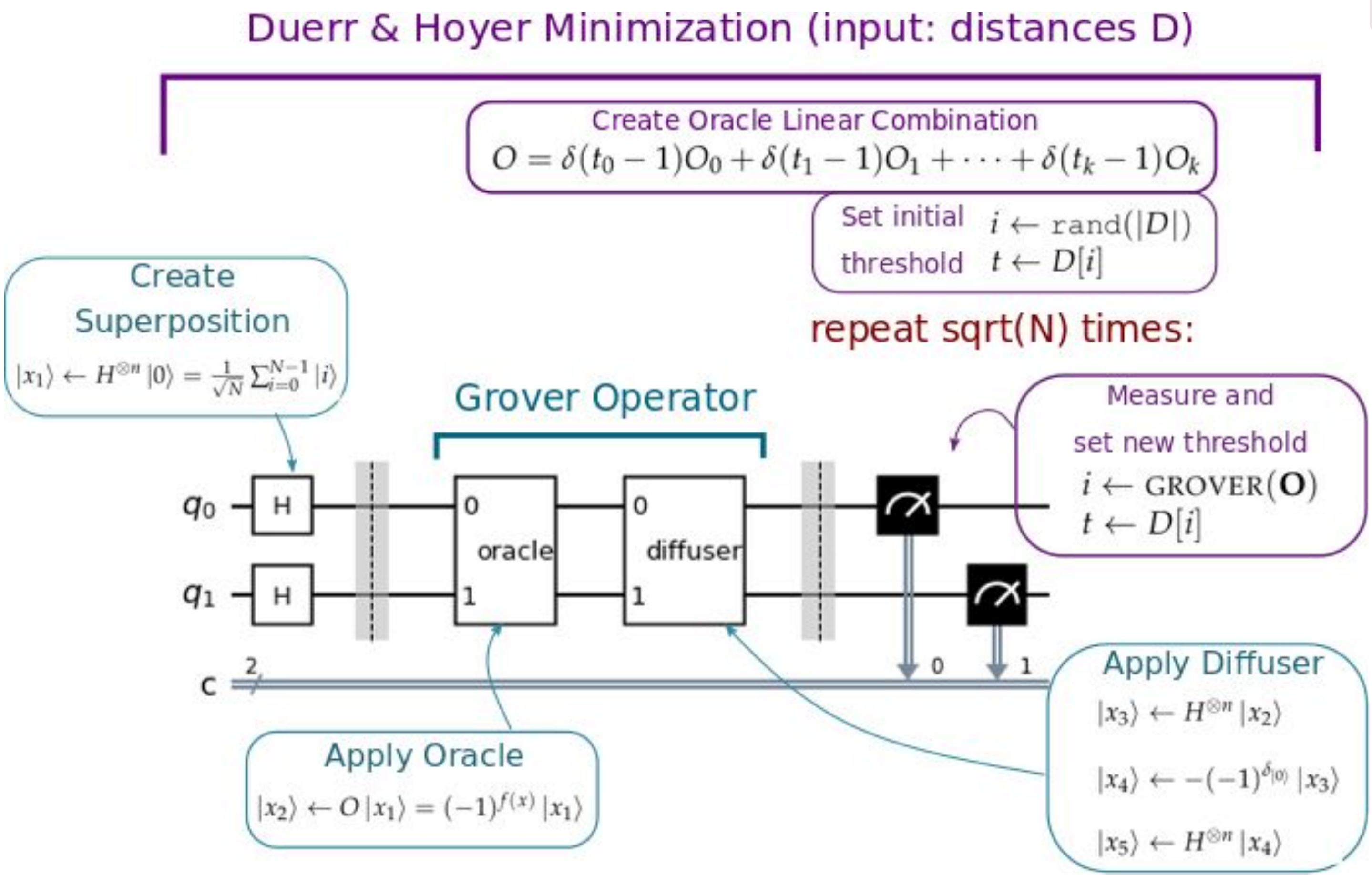
$$|x_2\rangle = \frac{1}{\sqrt{2}}(|0,a,b\rangle + |1,b,a\rangle)$$

$$|x_3\rangle = \frac{1}{2}|0\rangle(|a,b\rangle + |b,a\rangle) + \frac{1}{2}|1\rangle(|a,b\rangle - |b,a\rangle)$$

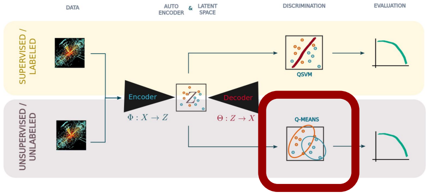
Measure ancilla in zero state $\quad P_{anc}(|0\rangle) = \frac{1}{2} + \frac{1}{2}|\langle\phi|\,|\psi_A\rangle|^2$

Unsupervised Clustering: Q-MEANS





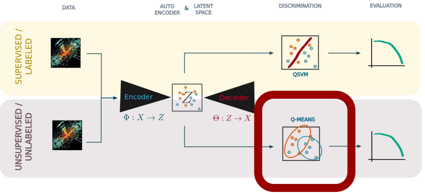
Unsupervised Clustering: Q-MEDIANS



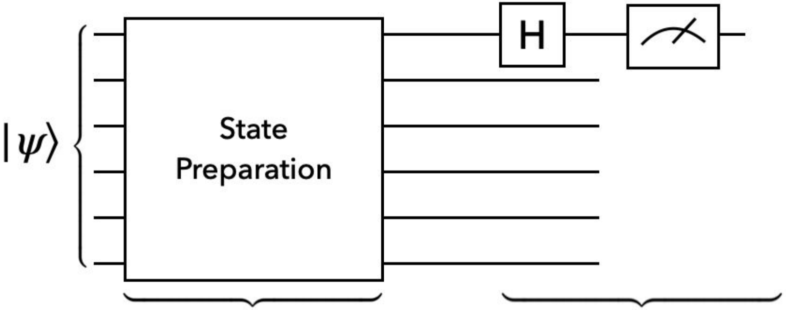
Algorithm in 3 parts:

1) **Quantum** distance calculation: distance to cluster
2) **Classic** minimization to closest cluster
3) Cluster median calculation (**quantum** distance + **classic** heuristics)

Unsupervised Clustering: Q-MEDIANS



1) Quantum **distance calculation**: distance to cluster



classical inputs:

$t = (t_x, t_y)$

$c = (c_x, c_y)$

$$Norm = \sqrt{t_x^2 + t_y^2 + c_x^2 + c_y^2)}$$

$$t_i' = \frac{t_i}{Norm} \quad c_i' = \frac{c_i}{Norm}$$

↓

$$|\psi\rangle = [t_x', t_y', c_x', c_y']$$

$$|\psi_1\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} t_x' \\ t_y' \\ c_x' \\ c_y' \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} t_x' + c_x' \\ t_y' + c_y' \\ t_x' - c_x' \\ t_y' - c_y' \end{pmatrix} \begin{matrix} |00\rangle \\ |01\rangle \\ |10\rangle \\ |11\rangle \end{matrix}$$

$$P\,|1\rangle = \frac{1}{2}\left[(t_x' - c_x')^2 + (t_y' - c_y')^2\right]$$

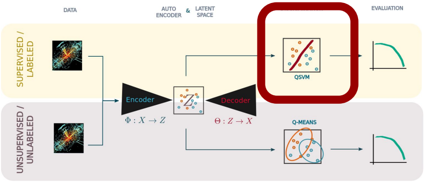
measure Most Significant Qubit

$$dist(t, c) = Norm \cdot \sqrt{2 \cdot P\,|1\rangle}$$

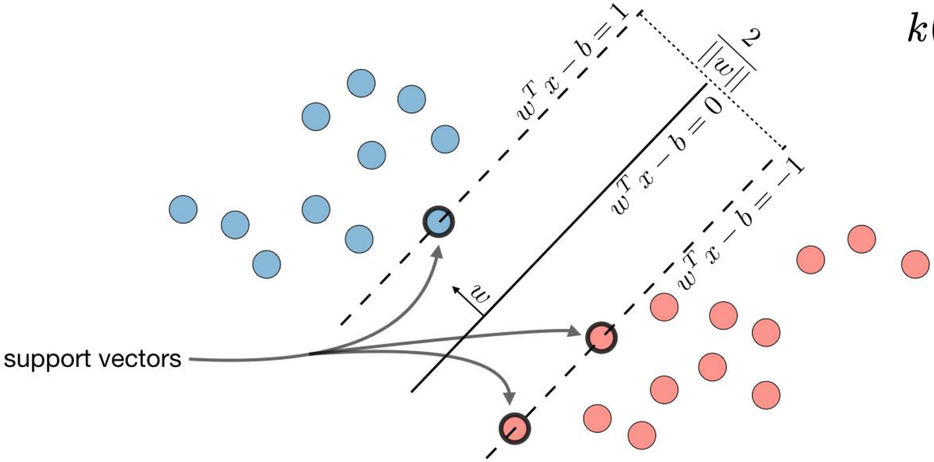
Use destructive interference probabilities to calculate distance

[*] K-Means Clustering on Noisy Intermediate Scale Quantum Computers

Supervised: QSVM Classifier

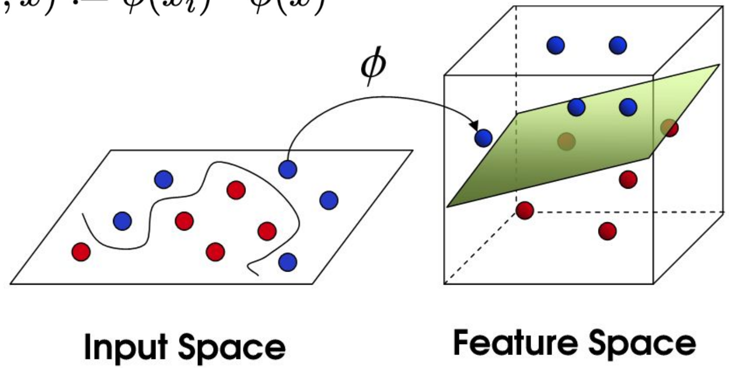


- Supervised training on 600 qcd (background) and 600 G_{RS} (signal) samples.
- Train to find the optimal separating hyperplane ⟶ convex optimisation task.
- Feature maps enable SVM to construct non-linear decision boundaries.



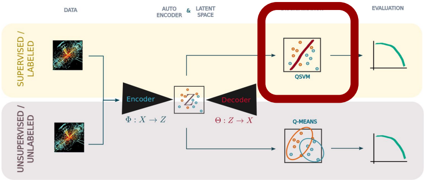
The kernel is defined via the feature map:

$$k(x_i, x) := \phi(x_i)^\dagger \cdot \phi(x)$$

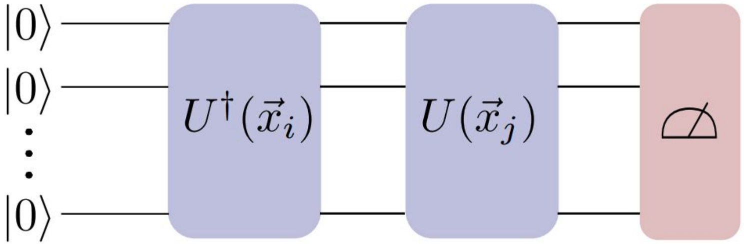


Source: [https://towardsdatascience.com/the-kernel-trick-c98cdbcaeb3f]

Supervised: QSVM Classifier

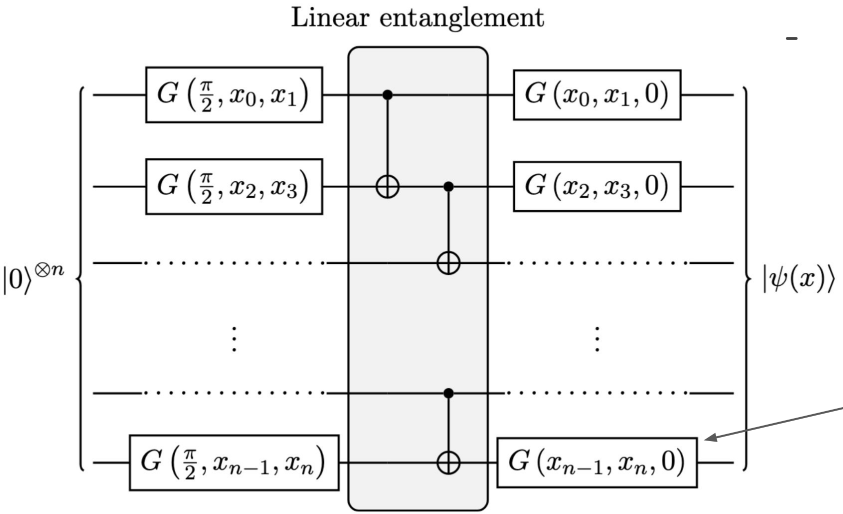


Quantum Kernel



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

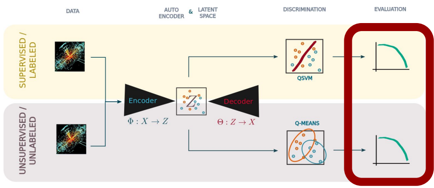
- Quantum kernel is sampled from a quantum device.
- The optimisation of the objective function remains on a classical computer.



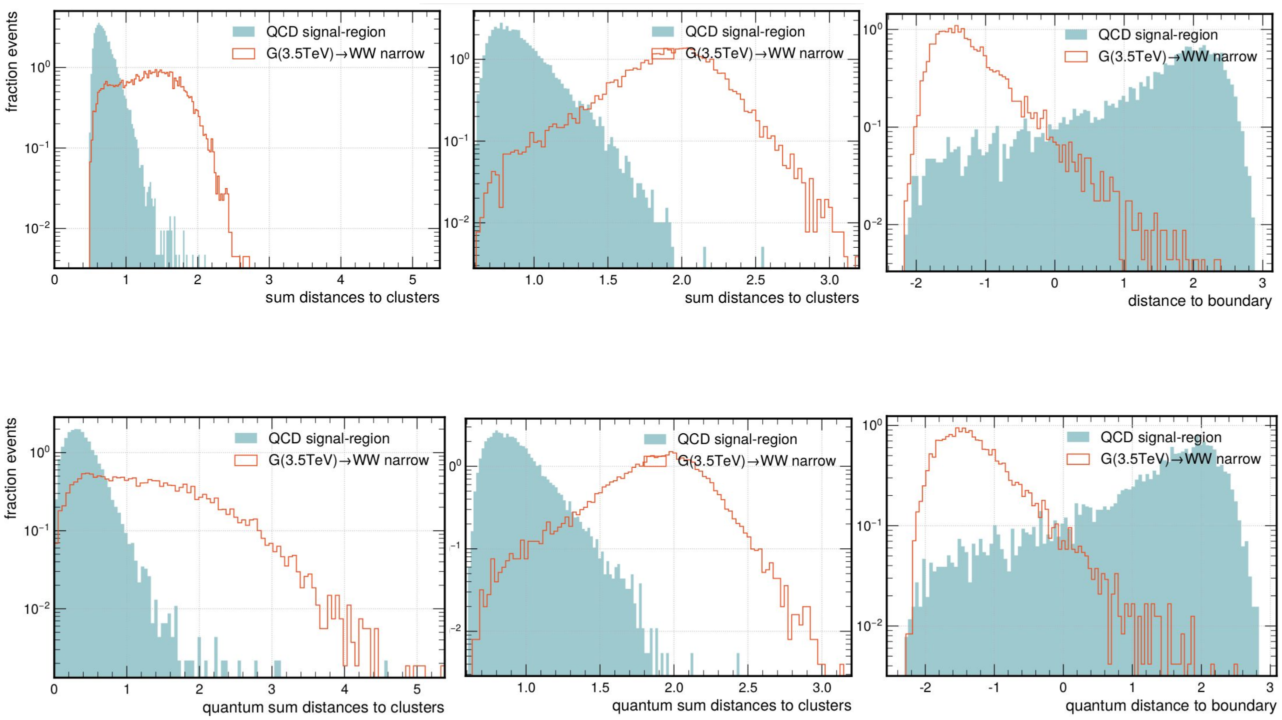
Feature map circuit U(x), for latent dim = 16, and n = 8 qubits.

$G \in \mathrm{SU}(2)$

Discrimination Metric Distributions



Unsupervised (Q)K-means | Unsupervised (Q)K-medians | Supervised (Q)SVM



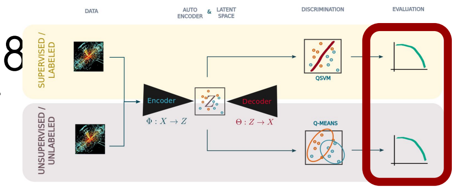
Metric

- QK-means & QK-medians: **Sum squared distance to cluster centers**
- QSVM: **Distance from decision boundary**

Results

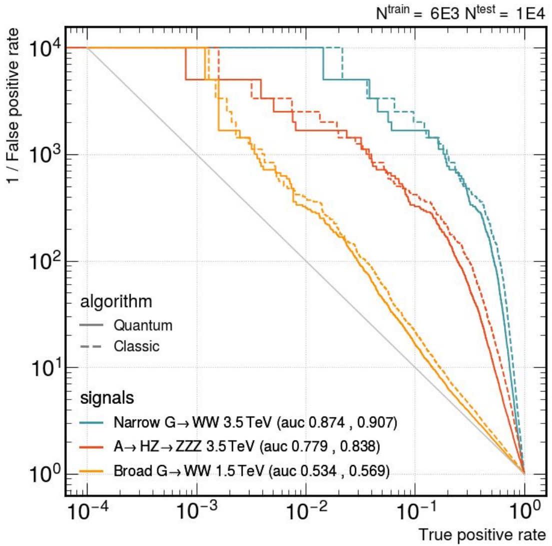
- Good separation of background vs signal
- Set cut-threshold β for signal efficiency
- Cut on tail for QK-clustering (e.g. β > 2)
- Cut on left mode for QSVM (e.g. β < -1)

ROC: Classic vs quantum for latent dim R^8

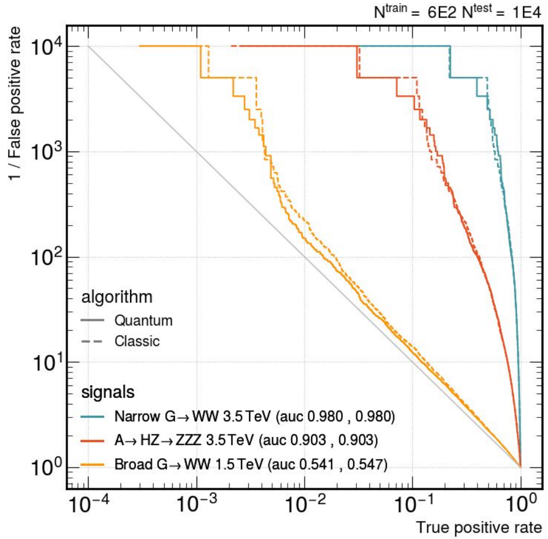


Graviton 1.5TeV (broad), Graviton 3.5TeV (narrow), A to HZ to ZZZ

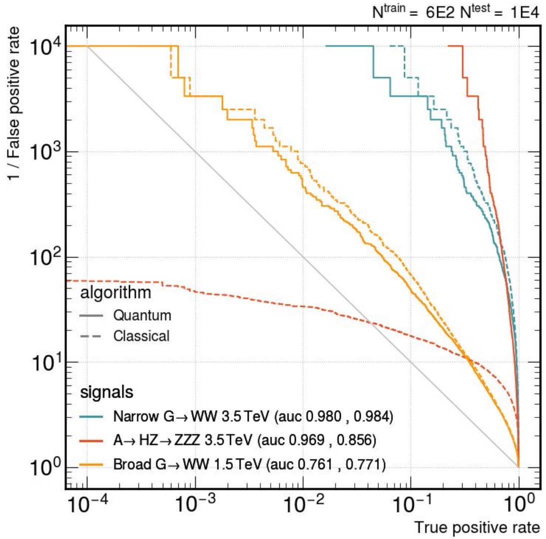
Unsupervised (Q)K-means



Unsupervised (Q)K-medians

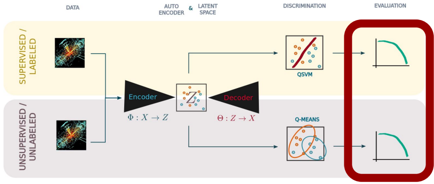


Supervised (Q)SVM



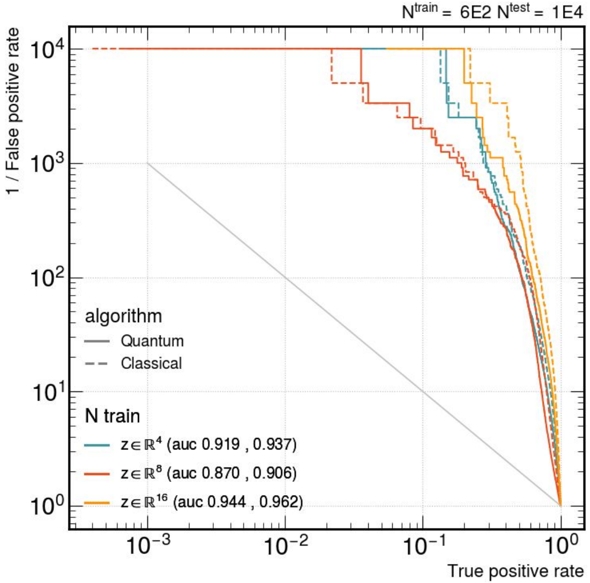
- Big variation in performance with high AUC values ~0.9 for narrow G_RS at 3.5TeV and ~random for broad G_RS at 1.5TeV both QK-means/-medians and QSVM algorithms (consistent with results in purely classic projects)
- Globally, **supervised** model **outperforms unsupervised** model but QK-means/-medians viable approach for solving model-agnostic problems
- performance of **quantum** algorithms is **competitive** when compared to **classical** counterparts

ROC: Impact of latent dimensionality

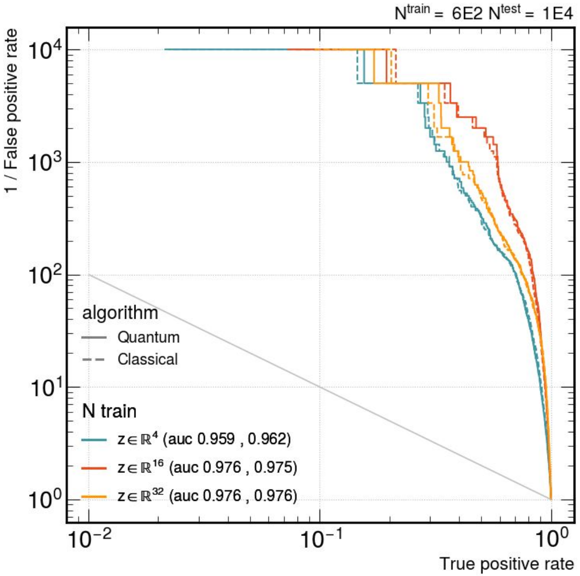


Quantum vs classic algorithm accuracy comparison
for latent dim $\mathbb{R}^4$, $\mathbb{R}^8$, $\mathbb{R}^{16}$ and $\mathbb{R}^{32}$

Unsupervised (Q)K-means

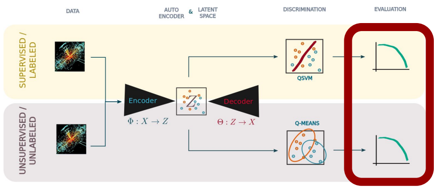


Unsupervised (Q)K-medians



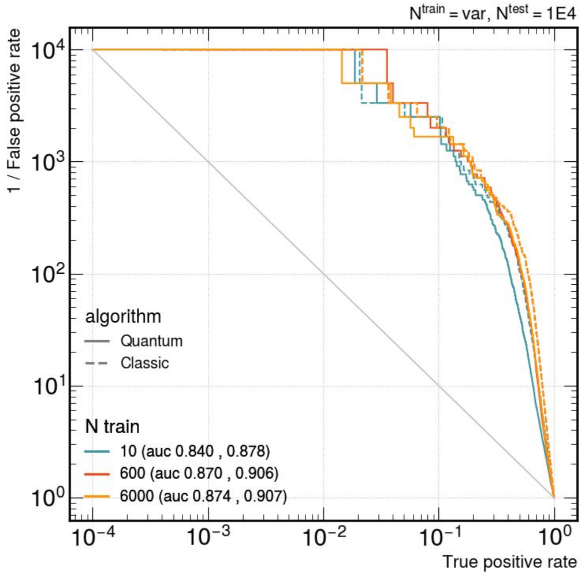
- Sweet spot of latent space dimensionality for QK-medians around $\mathbb{R}^{16}$
- No dramatic drop in performance for very small dimension $\mathbb{R}^4$

ROC: Impact of training size

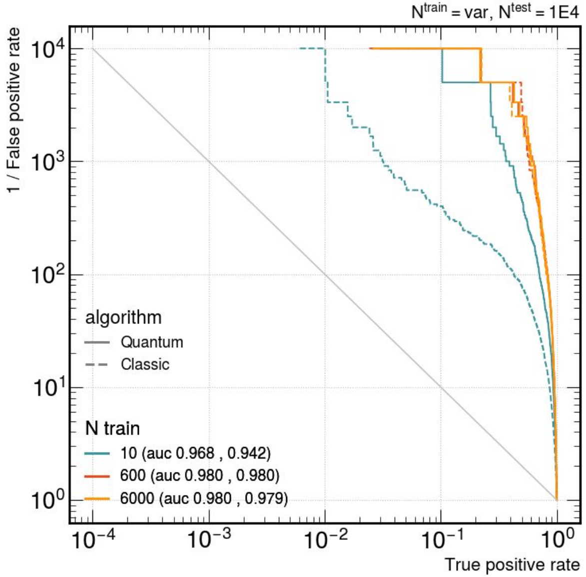


Quantum vs classic algorithm accuracy comparison for training size of **10**, **600** and **6000** training samples (Graviton 3.5 TeV)

Unsupervised (Q)K-means



Unsupervised (Q)K-medians



- Training size has **minor impact** on accuracy
- Quantum and classical are competitive
- **Quantum** algorithm has slight **advantage** in QK-medians approach for very small training size of 10 samples but needs to be investigated further

Conclusion

- We studied a quantum anomaly detector and a quantum classifier operating in a latent space representation of HEP events
- Both, QK-means/-medians and QSVM, proved effective in discriminating background from signal data-sets
- Supervised QSVM method (e.g. model-dependent searches) shows superior results compared to the unsupervised QK-clustering approach (e.g. model-independent searches)
- Performance of quantum algorithms is competitive when compared to their classical counterparts
- Marginal impact of training size on accuracy, further investigate very small sample sizes
- Divergent impact of latent space dimensionality, sweet spot dependent on algorithm choice

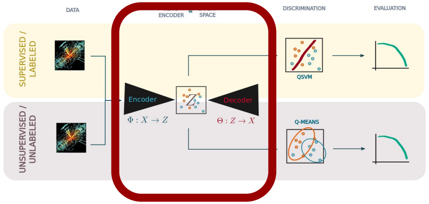
Based on results, we conclude that quantum algorithms are applicable to both, a model-independent and model-dependent analysis and could contribute to extend the sensitivity of the LHC experiments.

References

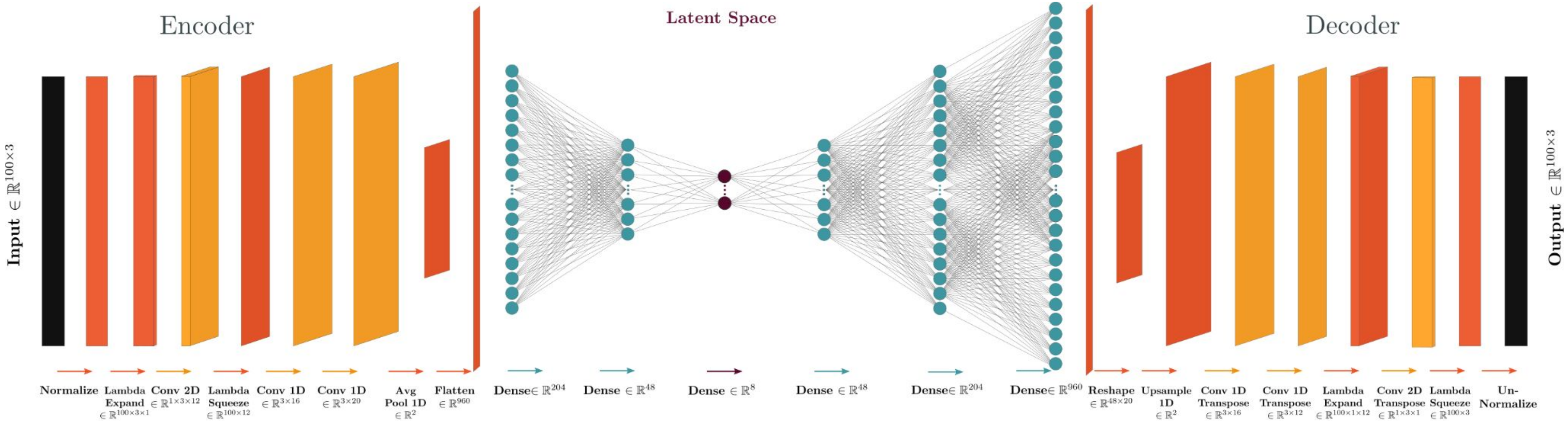
[1] Sjöstrand T *et al.* 2015 *Comput. Phys. Commun.* **191** 159–177 (*Preprint* `1410.3012`)
[2] de Favereau J *et al.* (DELPHES 3) 2014 *JHEP* **02** 057 (*Preprint* `1307.6346`)
[3] Randall L and Sundrum R 1999 *Phys. Rev. Lett.* **83** 3370 (*Preprint* `hep-ph/9905221`)
[4] Fan H, Su H and Guibas 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

Backup

Autoencoding for Dimensionality Reduction



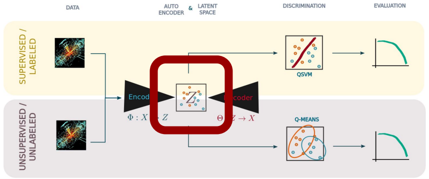
Architecture: Convolutional + Dense Layers

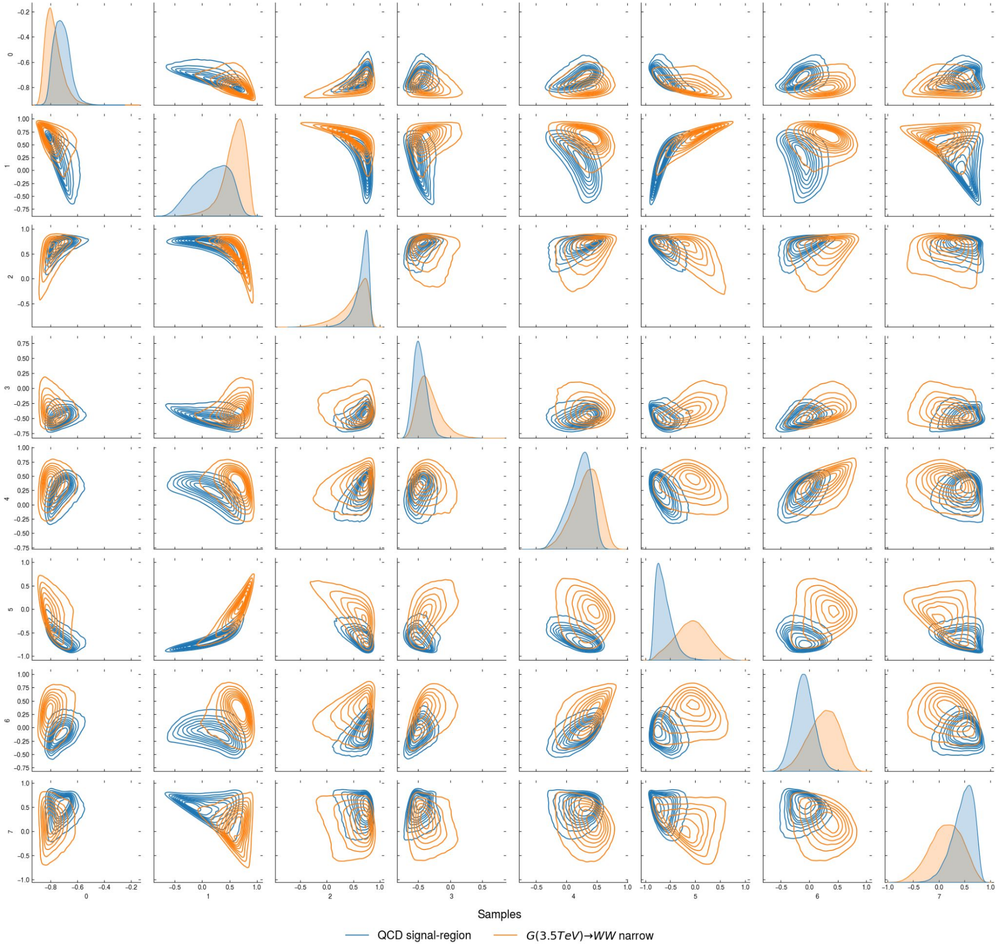


Study impact of latent space dimensionality:

$$Z \in \mathbb{R}^8, \mathbb{R}^{16}, \mathbb{R}^{32}$$

Results: Latent Space Representation





- ➔ Encoder Output
- ➔ $\mathbb{R}^{300} \to \mathbb{R}^{8}$
- ➔ Separation of Background and Signal

Noise in distance calculation

$$
\begin{aligned}
P(|0\rangle) &= |\frac{1}{2}\langle 0|0\rangle (|a,b\rangle + |b,a\rangle) + \frac{1}{2}\langle 0|\,|1\rangle (|a,b\rangle - |b,a\rangle)|^2 \\
&= \frac{1}{4}|(|a,b\rangle + |b,a\rangle)|^2 \\
&= \frac{1}{4}(\langle b|b\rangle \langle a|a\rangle + \langle b|a\rangle \langle a|b\rangle + \langle a|b\rangle \langle b|a\rangle + \langle a|a\rangle \langle b|b\rangle) \\
&= \frac{1}{2} + \frac{1}{2}|\langle a|b\rangle|^2
\end{aligned}
$$