Quantum machine learning for anomaly detection at the LHC

Vasilis Belis (ETH Zurich)

March 17th 2023 | CERN openlab Technical Workshop

Outline

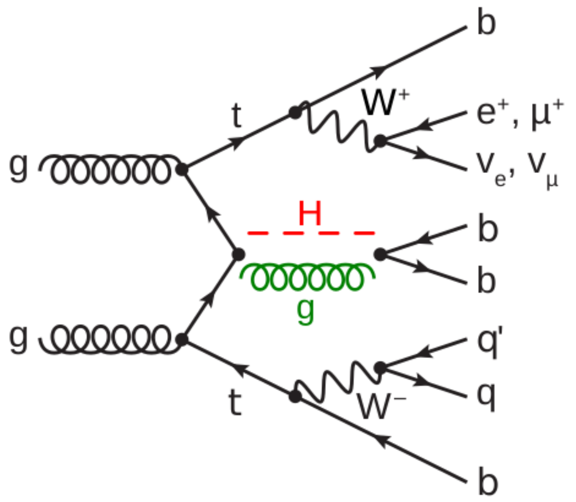
1. Model-independent searches and anomaly detection at the LHC.
   - Motivation
   - Unsupervised learning and anomaly detection

2. Quantum computing and machine learning.
   - Motivation

3. Quantum anomaly detection results
   - Detection of Gravitons and new Scalar bosons
   - Benchmark against classical counterpart.
   - Hardware Run.

Conventional searches at the LHC

Define *signal* and *background*.

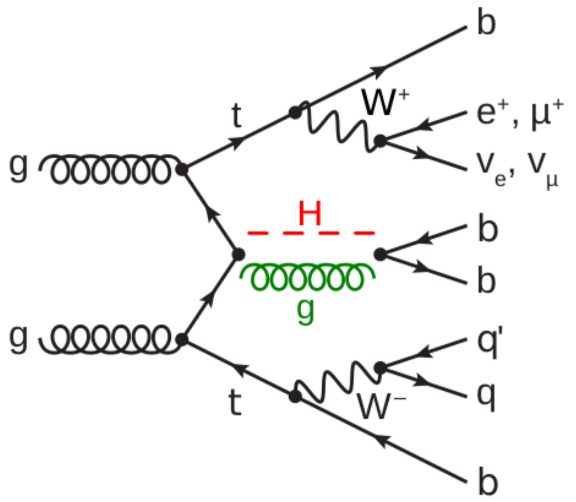


e.g.: $t\bar{t}H(b\bar{b})$ process at leading order in the semi-leptonic channel.

VB, et al., **Higgs analysis with quantum classifiers**
EPJ Web Conf., 251 (2021) 03070

Conventional searches at the LHC

Define *signal* and *background*.



e.g.: $t\bar{t}H(b\bar{b})$ process at leading order in the semi-leptonic channel.

VB, et al., **Higgs analysis with quantum classifiers**
EPJ Web Conf., 251 (2021) 03070

Model-dependent searches of Beyond Standard Model (BSM) physics

Define....

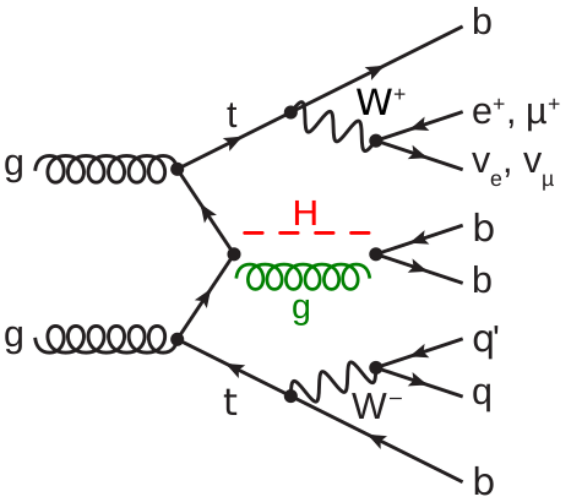
...analysis objects

Jets, Leptons, MET, ...

...signal region

Conventional searches at the LHC

Define *signal* and *background.*



e.g.: $t\bar{t}H(b\bar{b})$ process at leading order in the semi-leptonic channel.

VB, et al., **Higgs analysis with quantum classifiers**
EPJ Web Conf., 251 (2021) 03070

Model-dependent searches of Beyond Standard Model (BSM) physics

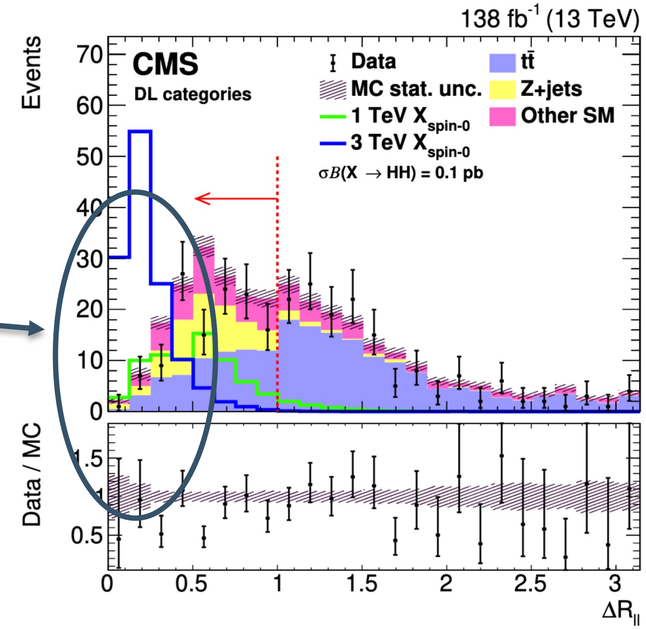
Define....

[10.1007/JHEP05(2022)005]

...analysis objects

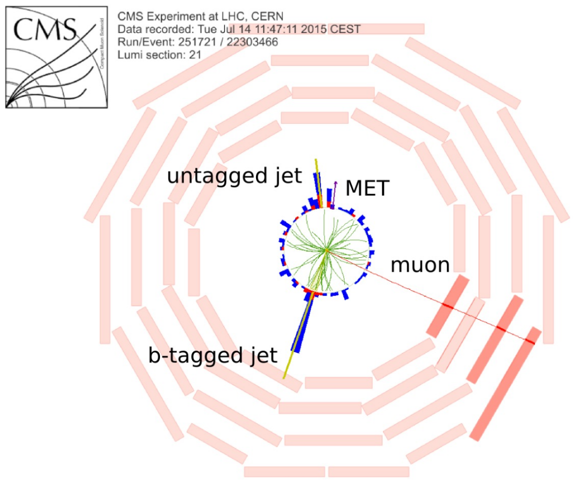
Jets, Leptons, MET, ...

...signal region



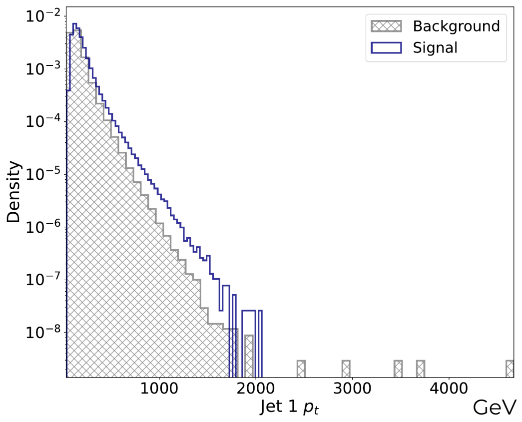
...discriminative variables of interest
For cuts or MVA classification.

Typical workflow: Model-dependent searches



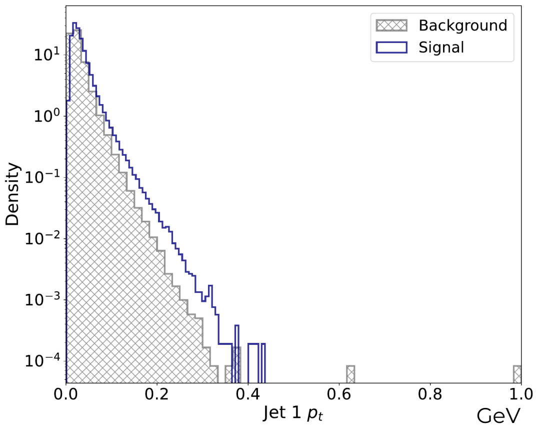
MC Simulation

compute prob. distributions →



•••

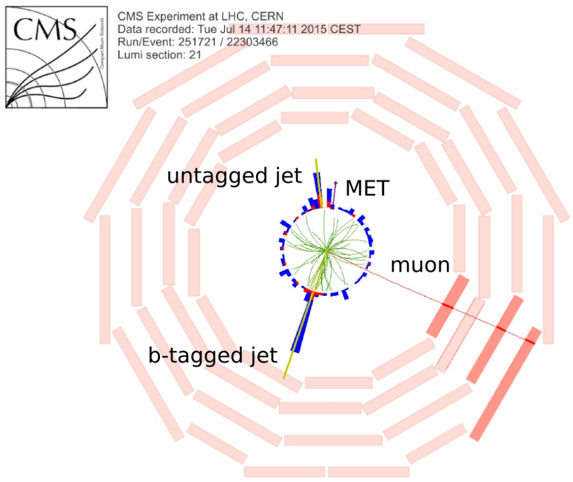
apply cuts and normalise →



•••

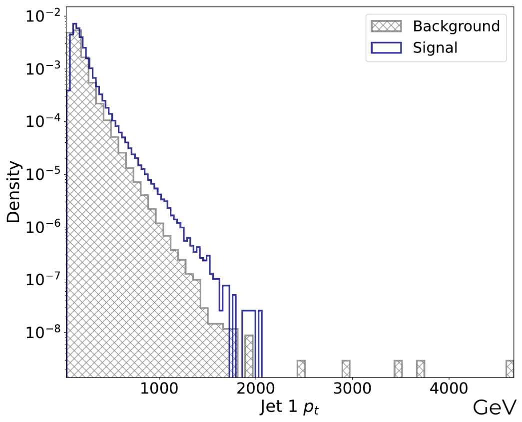
Input for classifiers models

Typical workflow: Model-dependent searches

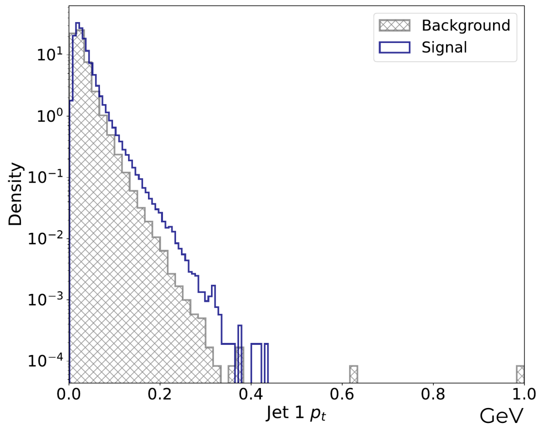


MC Simulation

compute prob. distributions



apply cuts and normalise

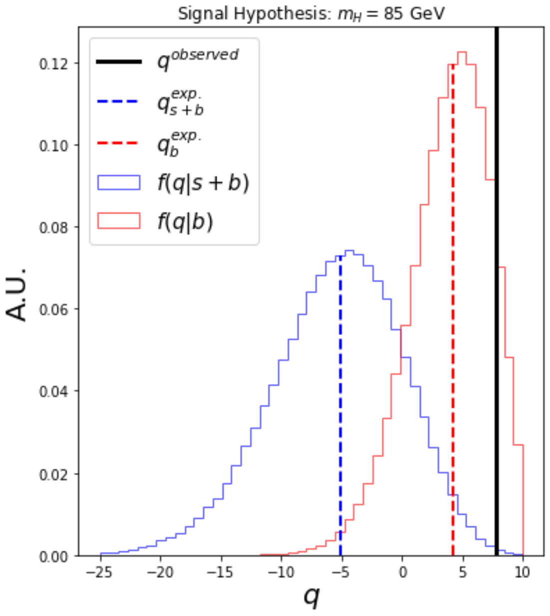


Input for classifiers models

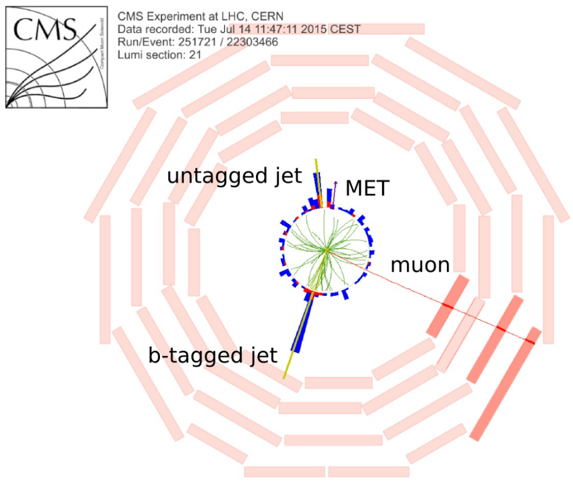
Classifier models:

cut-based or Machine Learning (BDT, NN, graph-net, etc.)

produce the *test statistic*

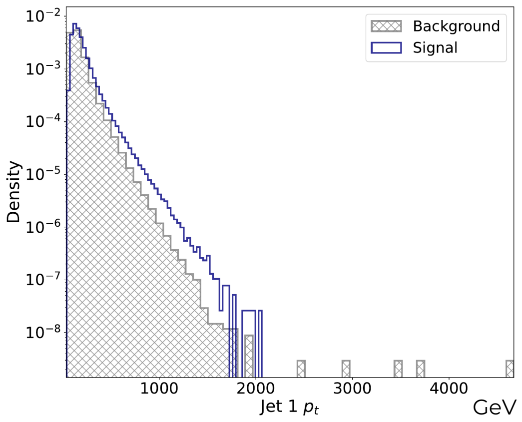


Typical workflow: Model-dependent searches

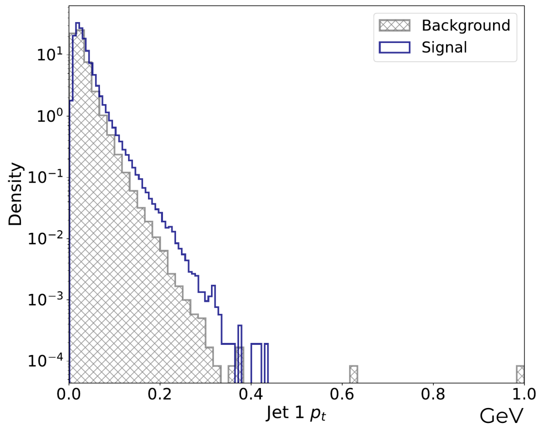


MC Simulation

compute prob. distributions



apply cuts and normalise

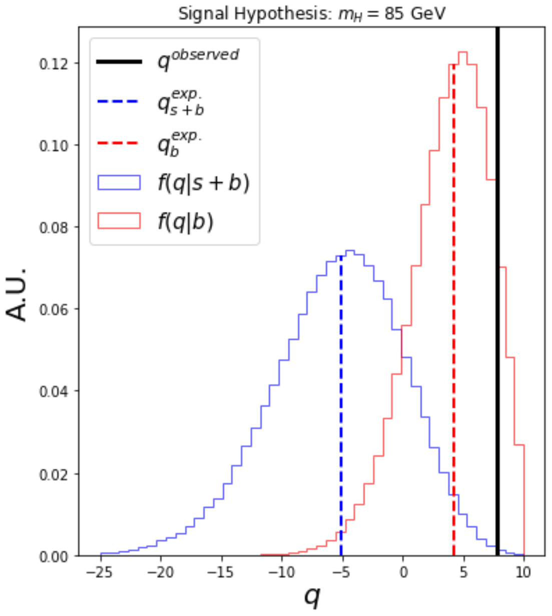


Input for classifiers models

Classifier models:

cut-based or Machine Learning (BDT, NN, graph-net, etc.)

produce the *test statistic*



What if you do not know the signal?

Motivating model-independent searches

Bias is **not necessarily bad**. It can be **great!**

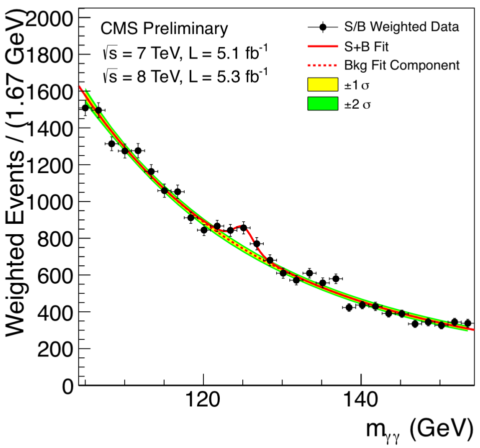
Higher bias $\longrightarrow$ lower variance of the statistical model

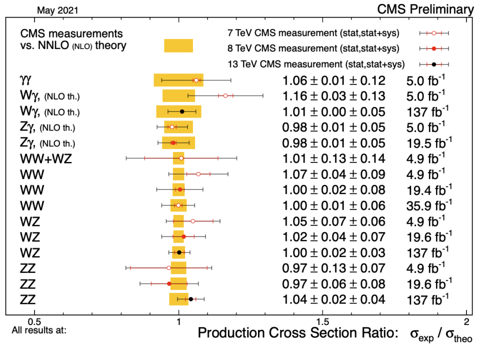
[Cramér–Rao bound]

So far, new-physics searches at the LHC:

assume SM + signal hypothesis….

[10.1016/j.physletb.2012.08.02]

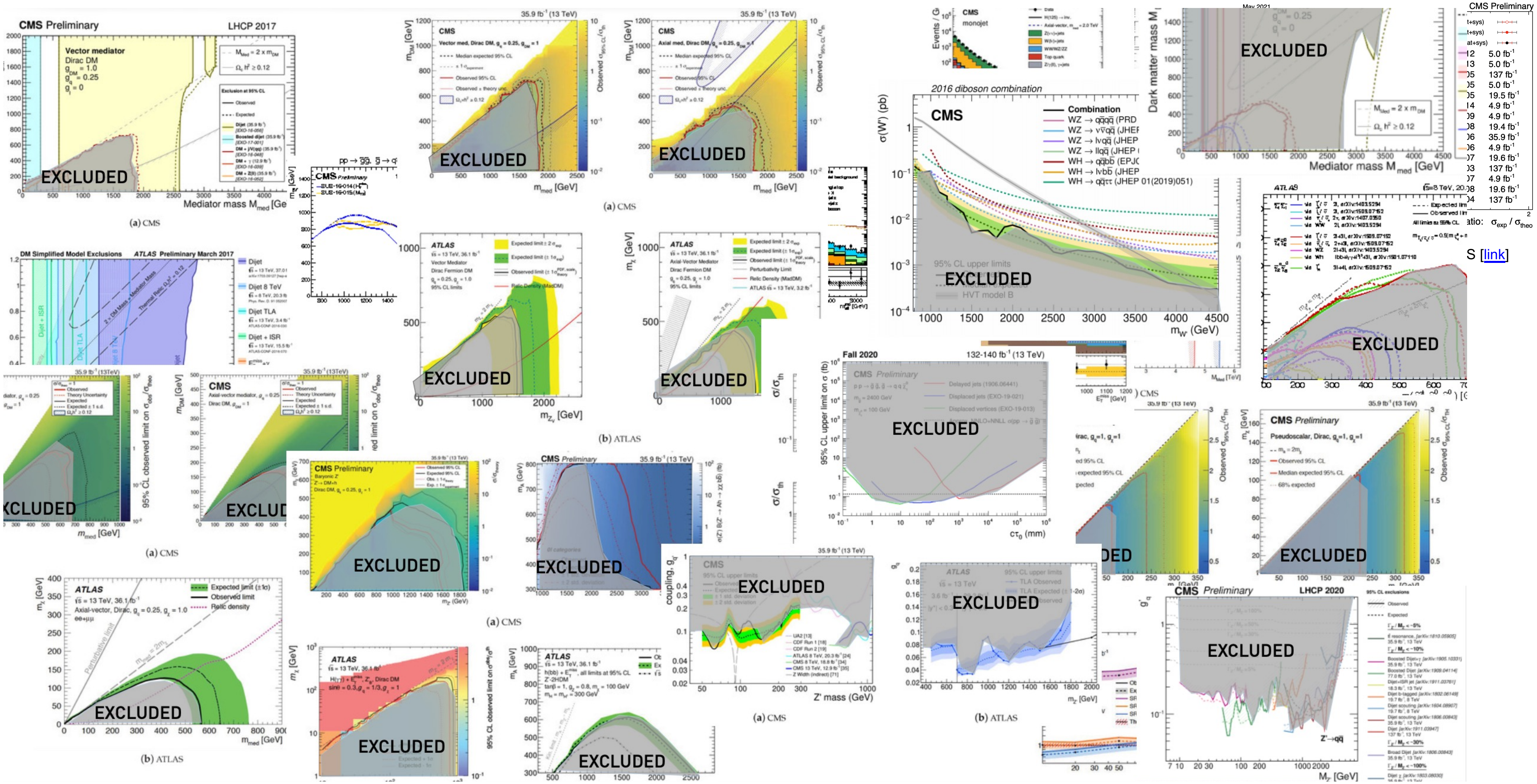




Summary SM measurements @ CMS [link]

Motivating model-independent searches

[10.1016/j.physletb.2012.08.02]



Motivating model-independent searches

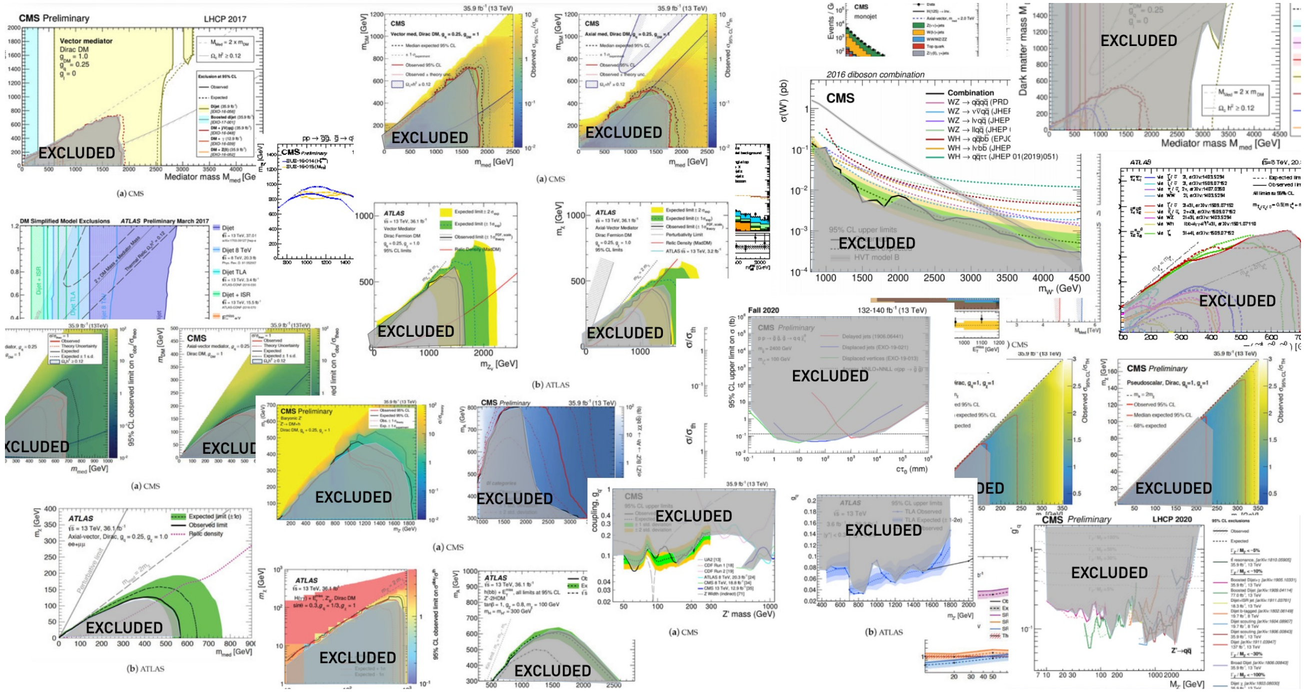
Bias is **not necessarily bad**. It can be **great!**

Higher bias ⟶ lower variance of the statistical model

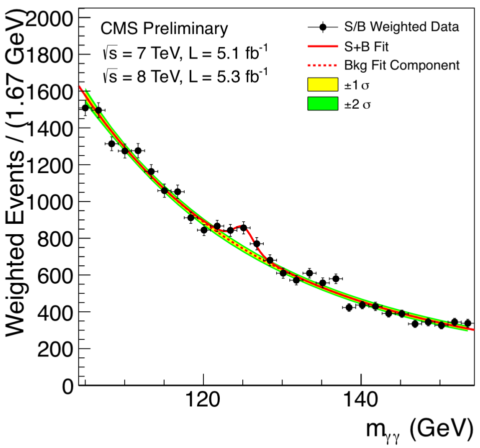
[Cramér–Rao bound]

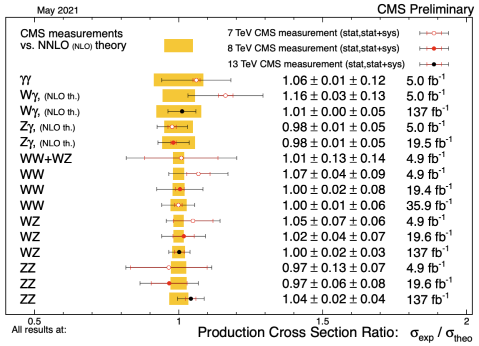
So far, new-physics searches at the LHC:

assume SM + signal hypothesis



[10.1016/j.physletb.2012.08.02]



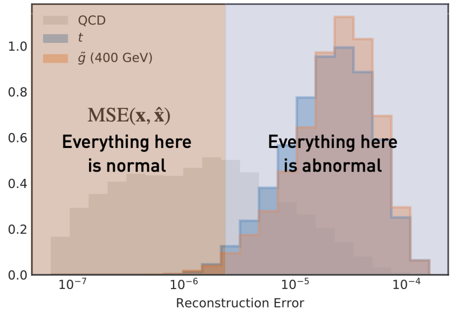


Summary SM measurements @ CMS [link]

What if you don't know where to look for new-physics?

Look at nature with minimal bias.

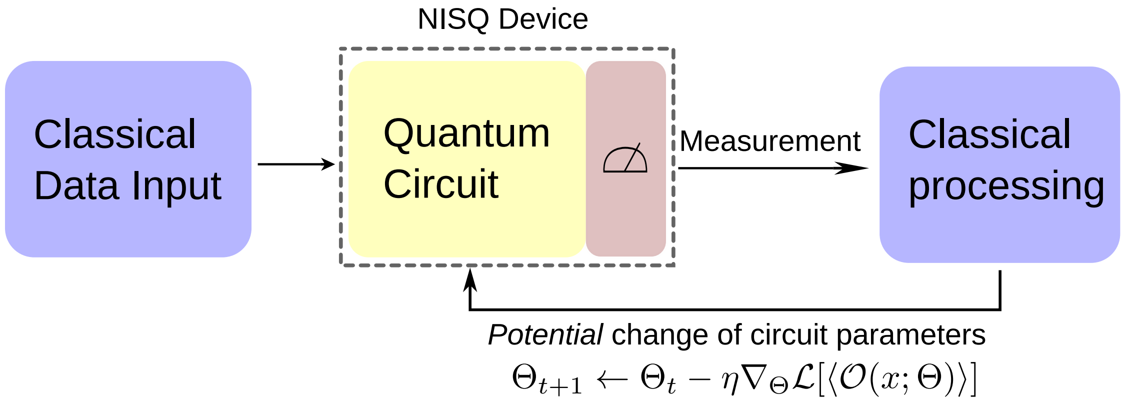
One possible solution:
Anomaly detection (ML/DL)



[Credit: Thea Årrestad; IPA Colloquium]

Quantum Machine Learning

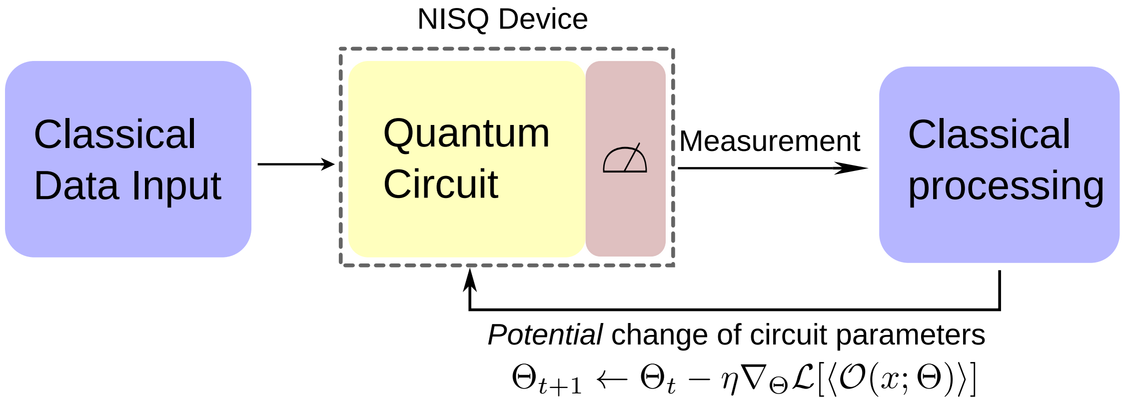
Hybrid quantum-classical algorithms



Noisy intermediate scale quantum devices

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

Hybrid quantum-classical algorithms

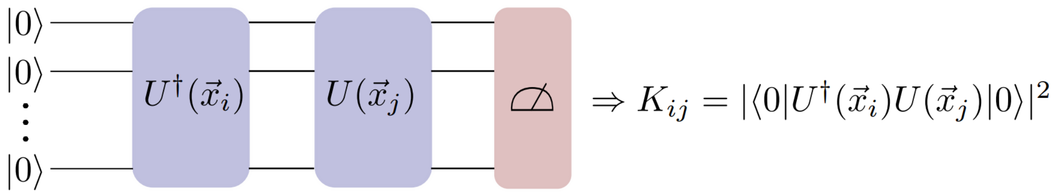


Noisy intermediate scale quantum devices

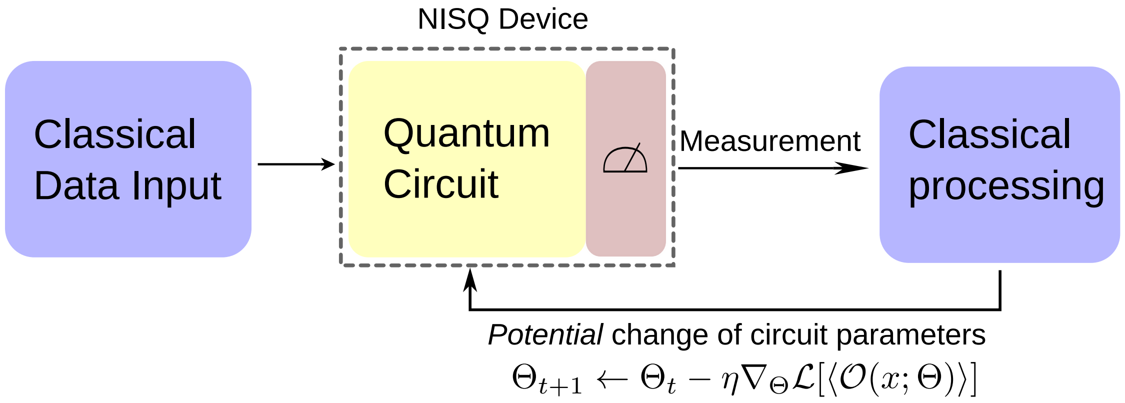
- Circuit width: limited number of qubits.
- Circuit depth: limited number of operations per qubit (small decoherence times).
- Hardware noise.

Quantum Machine Learning (QML) models for classification

<u>Kernel methods</u>
Quantum Support Vector Machines



Hybrid quantum-classical algorithms

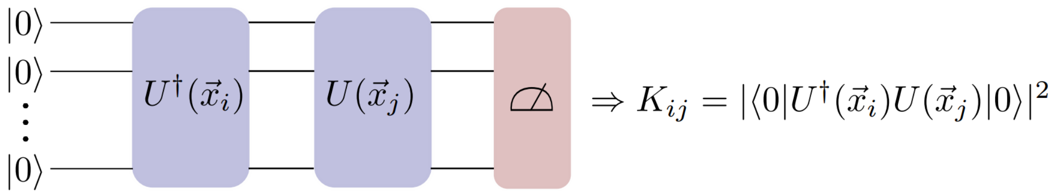


Noisy intermediate scale quantum devices

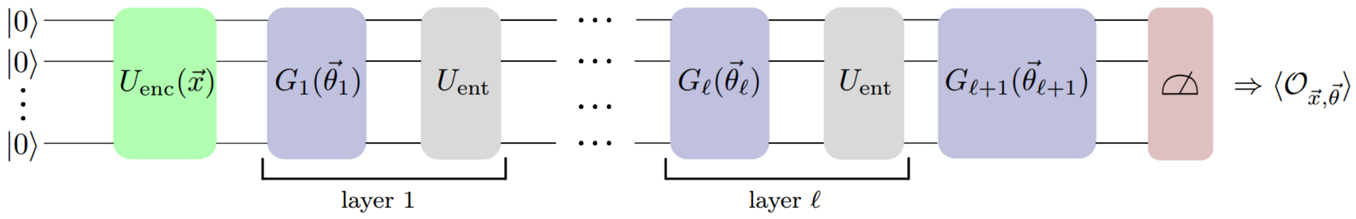
- Circuit width: limited number of qubits.
- Circuit depth: limited number of operations per qubit (small decoherence times).
- Hardware noise.

Quantum Machine Learning (QML) models for classification

<u>Kernel methods</u>
Quantum Support Vector Machines



<u>Variational/Parametrised Quantum Circuits</u>
Quantum Neural Networks



Current hardware limitations: feature reduction presently needed for realistic datasets.

Motivation

Why quantum machine learning? Why for HEP?

Practical and exploratory answer

Investigate a new set of ML techniques to assess for advantages. Why not?

Motivation

Why quantum machine learning? Why for HEP?

Practical and exploratory answer

Investigate a new set of ML techniques to assess for advantages. Why not?

Fundamental motivation

Potentially, utilise the information and correlations (quantum remnants) inherent in HEP data? performance advantages?

Motivation

Why quantum machine learning? Why for HEP?

Practical and exploratory answer

Investigate a new set of ML techniques to assess for advantages. Why not?

Fundamental motivation

Potentially, utilise the information and correlations (quantum remnants) inherent in HEP data? performance advantages?

Theoretical results

Generalisation with few data, computational speed-ups, uncover correlations unrecognisable to classical methods

[M. Caro et al., Nature Communications 13, 4919 (2022)]
[A. Abbas et al., Nature Computational Science 1, 403 (2021)]
[Y. Liu et al., Nature Physics 17, 1013 (2021)]
[H. Huang et al., Nature Communications 12, 2631 (2021)]
[H . Huang et al.,, Science 376, 1182 (2022)]
[N. Pirnay et al., arXiv: 2212.08678 (2022)]
Among others...

Results

Finding new-physics in dijet events with QML

Identifying new-physics with quantum models

Anomaly detection with quantum machine learning

Background: QCD multi-jet events. $n^{\text{features}} = 300$ per jet $\longrightarrow$ Too many for current hardware

$G \to W^- W^+$ $\quad$ $A \to HZ \to ZZZ$

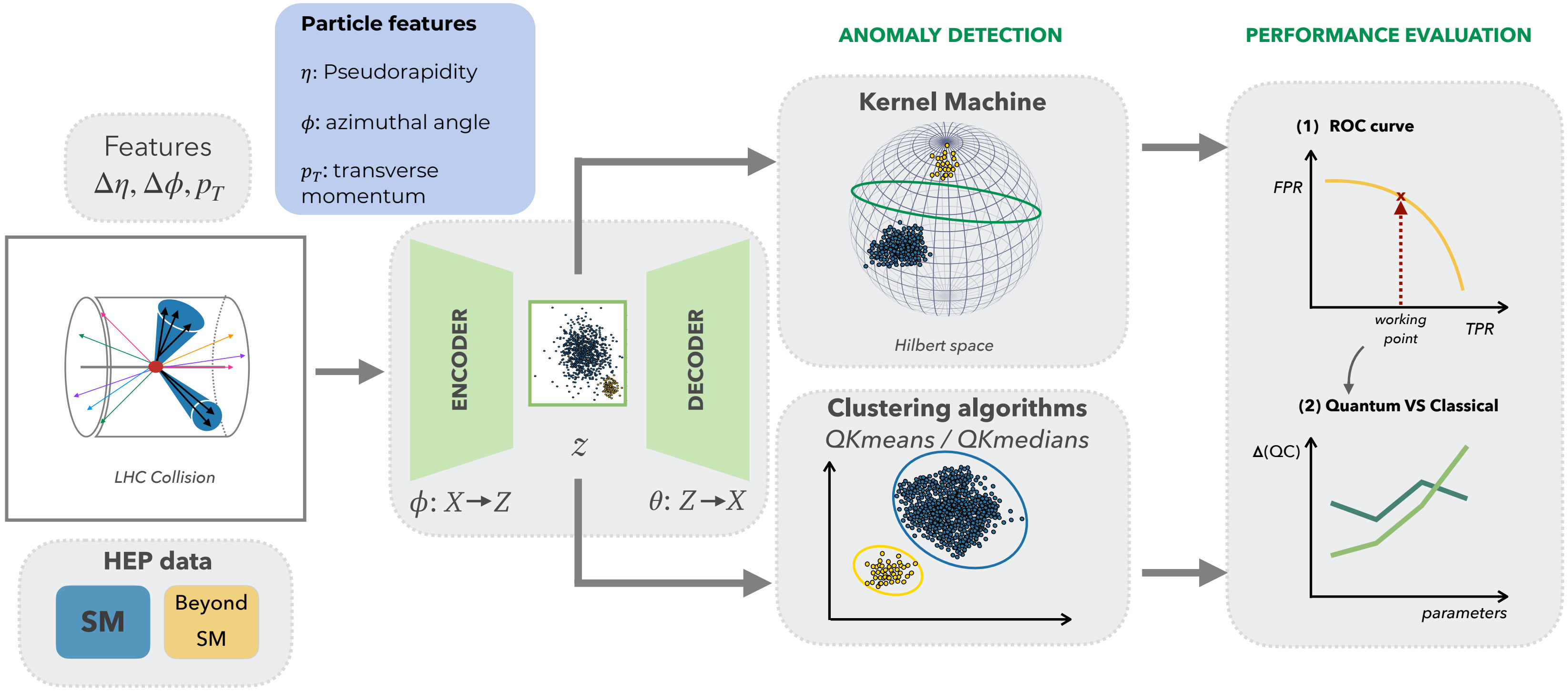
Tested BSM anomalies: Graviton **&** New Scalar Boson $\longrightarrow$ Multi-jet final state

Identifying new-physics with quantum models

Anomaly detection with quantum machine learning

Background: QCD multi-jet events. $n^{\text{features}} = 300$ per jet ⟶ Too many for current hardware.

Tested BSM anomalies: Graviton ($G \to W^- W^+$) & New Scalar Boson ($A \to HZ \to ZZZ$) ⟶ Multi-jet final state



Suitable metric for anomaly detection

Background rejection @ working point

$$\varepsilon_{\mathrm{b}}^{-1}(\varepsilon_s; \mathcal{M})$$

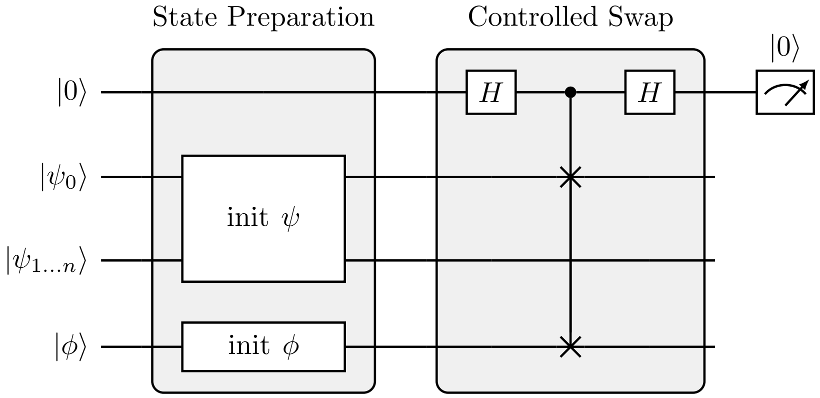
Compare models

$$\Delta_{\mathrm{QC}}(\varepsilon_s) = \frac{\varepsilon_{\mathrm{b}}^{-1}(\varepsilon_s; Q)}{\varepsilon_{\mathrm{b}}^{-1}(\varepsilon_s; C)}$$

Quantum clustering for anomaly detection

Construct clusters in the Hilbert space

Quantum distance calculation from clusters

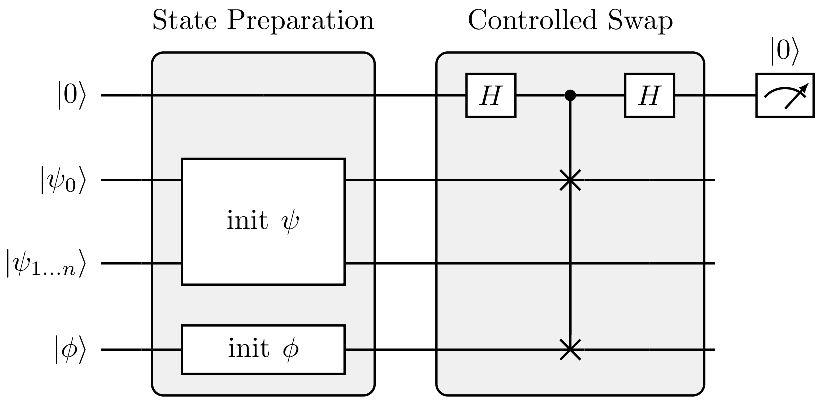


Minimise the distance with **quantum** (QK-means) or hybrid/**classical** (QK-medians) optimisation algorithms

Quantum clustering for anomaly detection

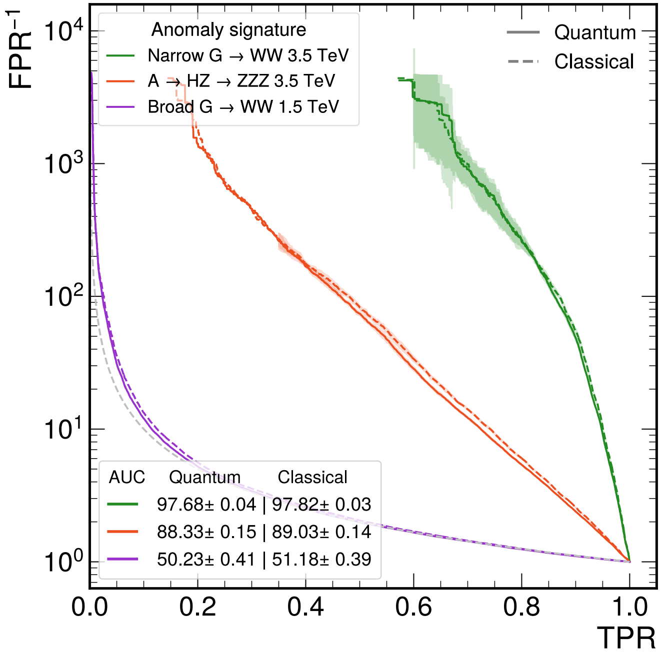
Construct clusters in the Hilbert space

Quantum distance calculation from clusters

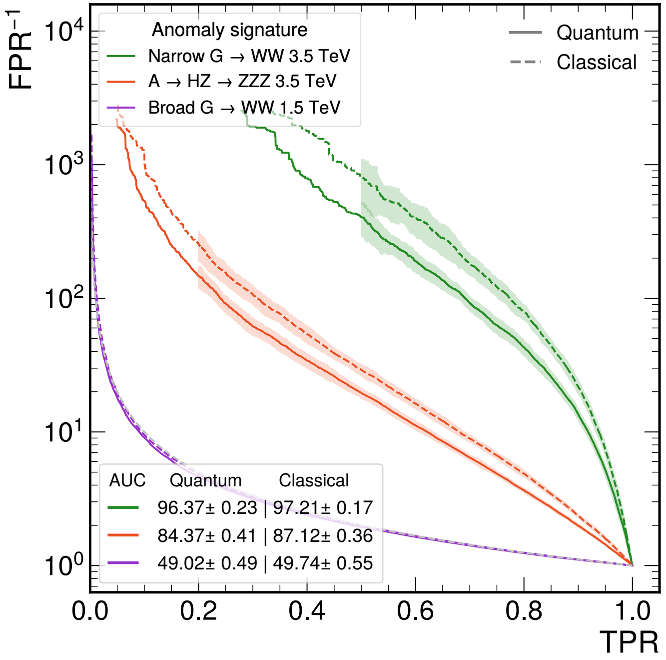


Minimise the distance with **quantum** (QK-means) or hybrid/**classical** (QK-medians) optimisation algorithms

Quantum K-medians



Quantum K-means



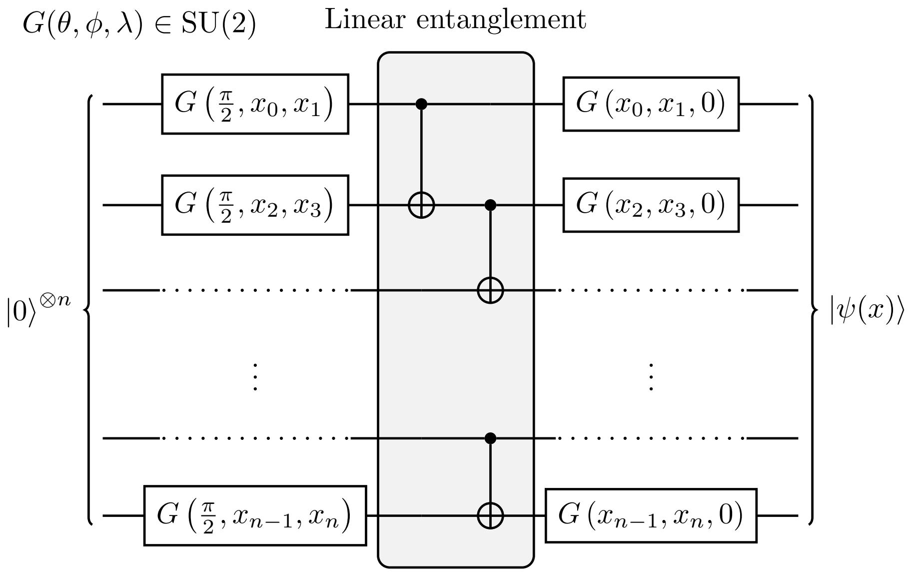
[K.A. Wozniak*, VB*, E. Puljak*, et al., arXiv: 2301.10780]

Quantum and classical anomaly detection has similar performance.

Kernel-based quantum anomaly detection

Unsupervised quantum kernel machine $\mathrm{K}_{ij} = |\langle 0|U^\dagger(\vec{x}_i)U(\vec{x}_j)|0\rangle|^2$

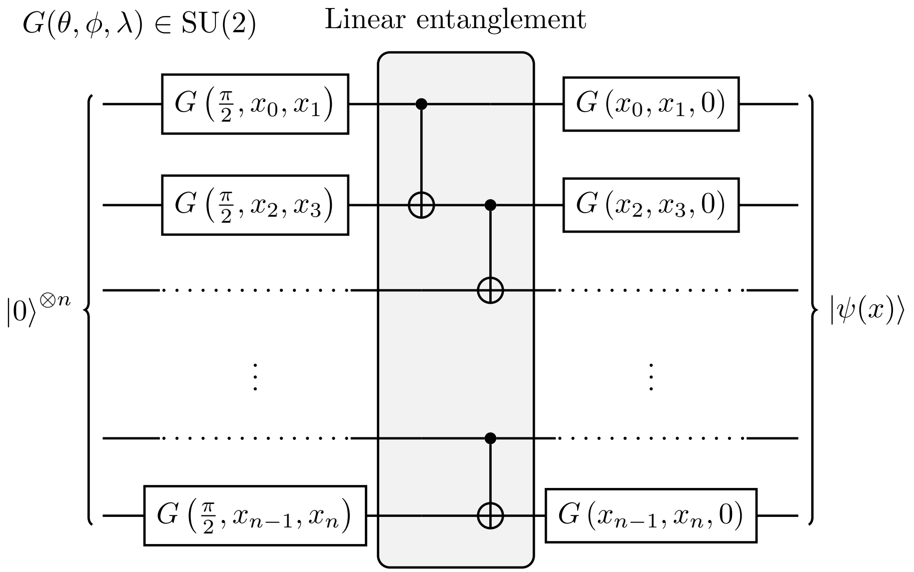
Designed data encoding circuit



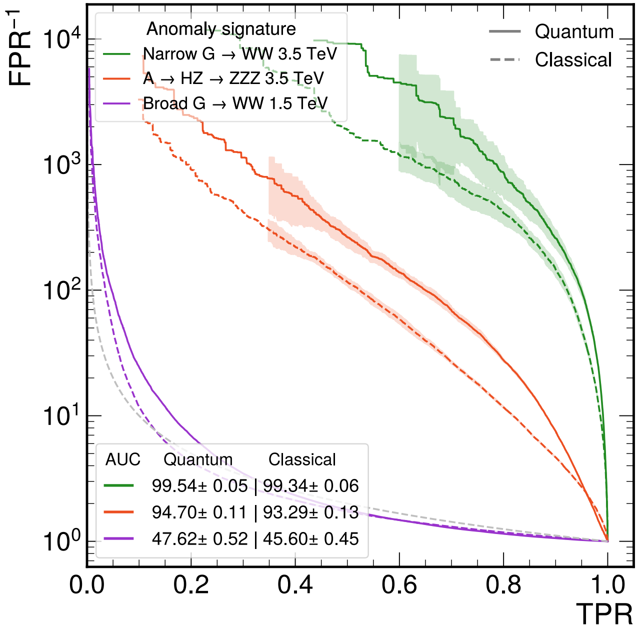
Kernel-based quantum anomaly detection

Unsupervised quantum kernel machine $\mathrm{K}_{ij} = |\langle 0|U^\dagger(\vec{x}_i)U(\vec{x}_j)|0\rangle|^2$

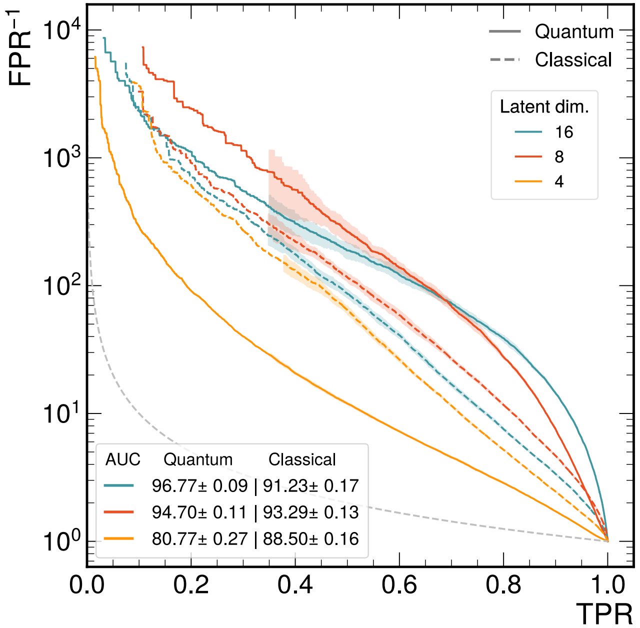
Designed data encoding circuit



Different BSM scenarios



Different qubit number

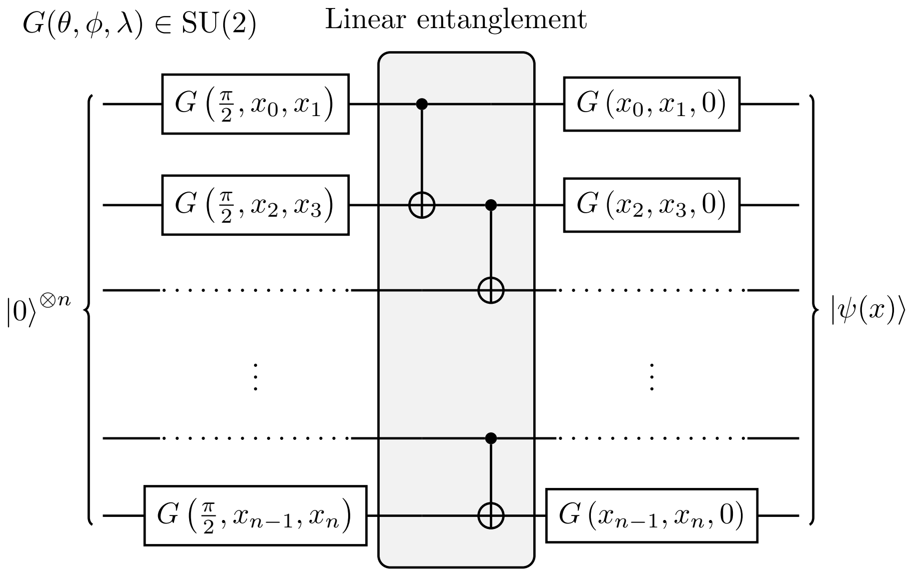


[K.A. Wozniak*, VB*, E. Puljak*, et al., arXiv: 2301.10780]

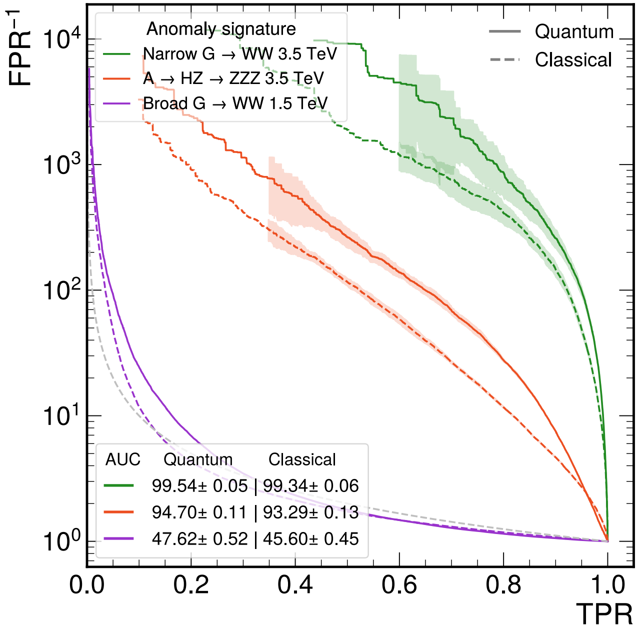
Kernel-based quantum anomaly detection

Unsupervised quantum kernel machine $\mathrm{K}_{ij} = |\langle 0|U^\dagger(\vec{x}_i)U(\vec{x}_j)|0\rangle|^2$

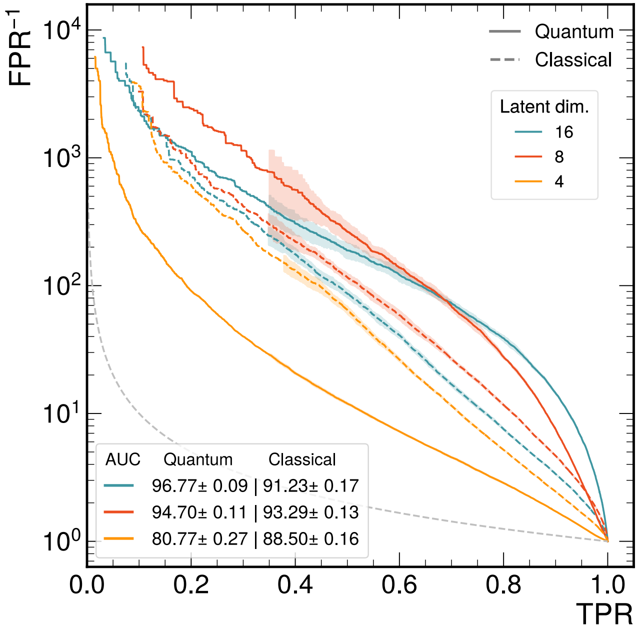
Designed data encoding circuit



Different BSM scenarios



Different qubit number



[K.A. Wozniak*, VB*, E. Puljak*, et al., arXiv: 2301.10780]

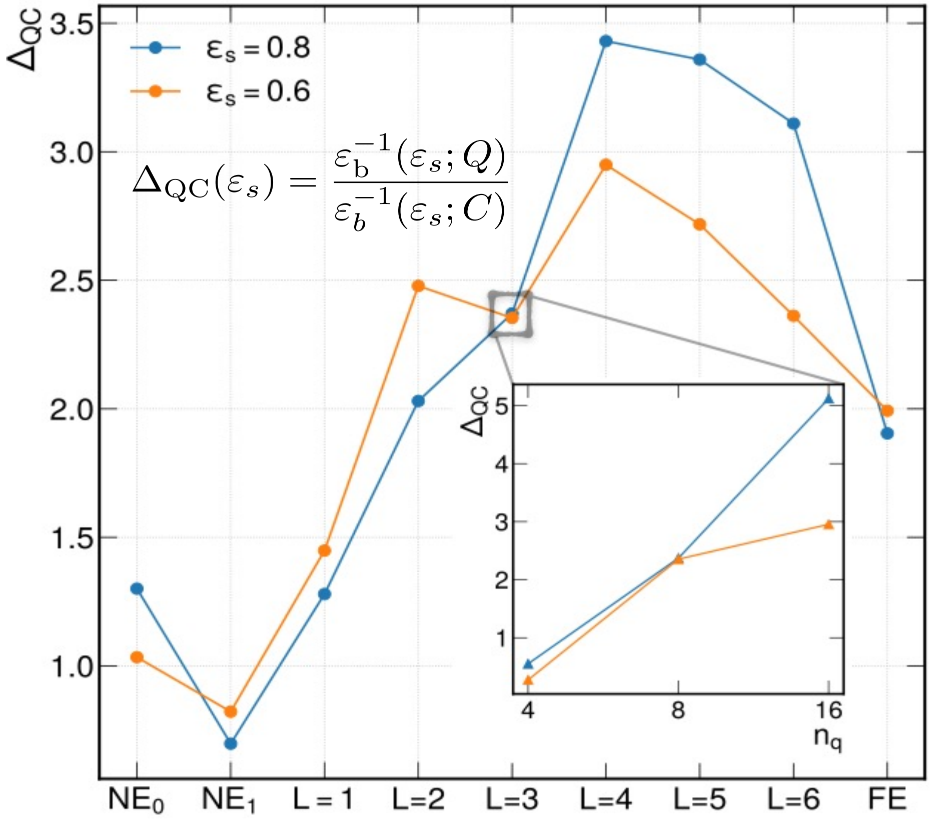
Instance of significant and consistent quantum performance advantage!

Very exciting and first of its kind result (HEP + Anomaly detection)!

Quantum circuit properties vs. performance

[K.A. Wozniak*, VB*, E. Puljak*, et al., arXiv: arXiv:2301.10780]

Performance vs. circuit architectures
Analysing circuit depth (expressibility) and amount entanglement

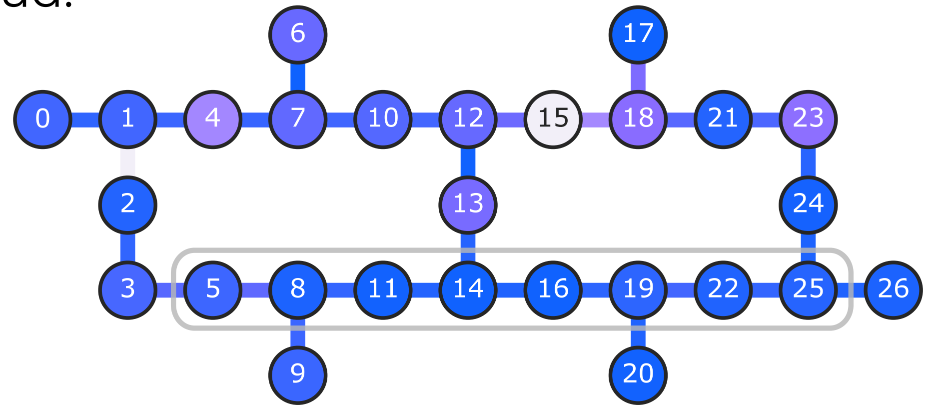


Importance of intrinsically quantum properties of the feature map.

Up to **five times** the performance of the classical model for 16 qubits!

Quantum hardware runs

Submit jobs to a real machine (ibm_toronto) using IBMQ cloud.
(CERN quantum-hub)

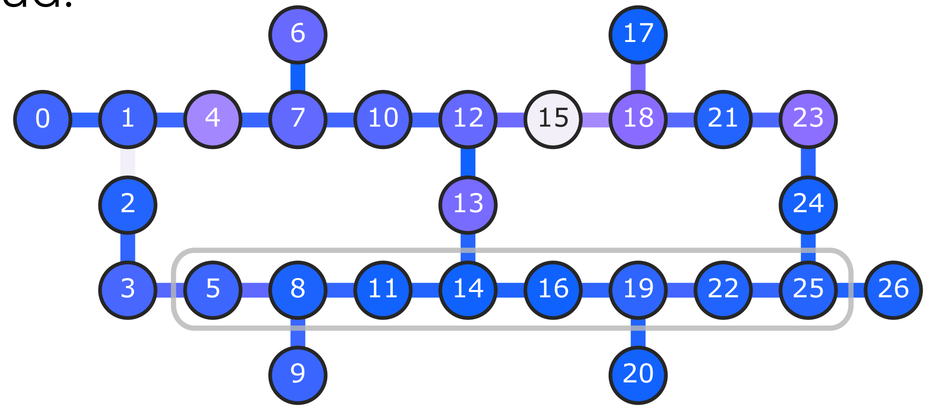


Map algorithm to hardware qubits.

Minimal instance 100 + 100 (train + test) datapoints.

Quantum hardware runs

Submit jobs to a real machine (ibm_toronto) using IBMQ cloud. (CERN quantum-hub)



Map algorithm to hardware qubits.

Minimal instance 100 + 100 (train + test) datapoints.

Kernel Machine Run	AUC	$\langle \mathrm{tr}\rho^2 \rangle$
Hardware $L = 1$	0.844	0.271(6)
Ideal $L = 1$	0.999	1
Hardware $L = 3$	0.997	0.15(2)
Ideal $L = 3$	1.0	1
Classical	0.998	-

Purity of fully mixed state: $1/2^{n_q} \approx 0.39 \times 10^{-2}$
(decoherence = loss of “quantumness”)

$$\langle \mathrm{tr}\rho^2 \rangle = \langle K(x_i, x_i) \rangle$$

$$\rho(x_i) = U(x_i)|0\rangle\langle 0|U^\dagger(x_i)$$

Proposed data encoding circuit realistic and suitable for current devices

Conclusions

Quantum anomaly detection for HEP

Fundamentally different way of data representation and processing.

Model-independent (unsupervised learning) approach for minimally biased searches of new-physics.

Promising results identifying a **significant and consistent advantage** in anomaly detection!

Conclusions

Quantum anomaly detection for HEP

Fundamentally different way of data representation and processing.

Model-independent (unsupervised learning) approach for minimally biased searches of new-physics.

Promising results identifying a **significant and consistent advantage** in anomaly detection!

For more details checkout:

- K.A. Wozniak*, VB*, E. Puljak*, et al., **Quantum anomaly detection in the latent space of proton collision events at the LHC,** arXiv:2301.10780

- J. Shuhmacher, L. Bogia, VB, et al. **Unravelling physics beyond the standard model with classical and quantum anomaly detection,** arXiv: 2301.10787

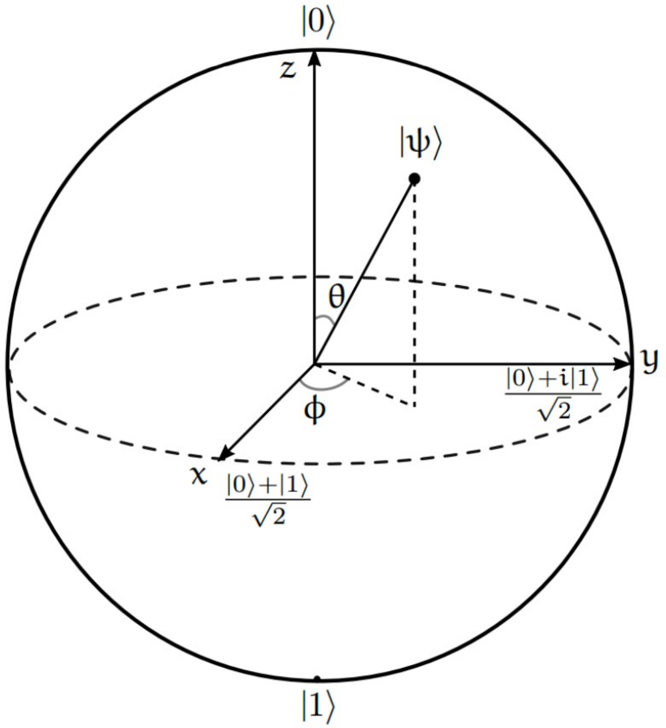
Questions?

Backup slides

Basics of quantum information processing

The qubit:

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle \;\equiv \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle$$

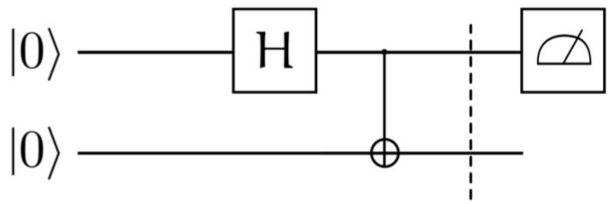


Generic qubit operations (quantum gates) $U = e^{-i\vec{\theta}\cdot\frac{\vec{\sigma}}{2}} \in \mathrm{SU}(2)$:

$$U(\theta,\phi,\lambda) = \begin{pmatrix} \cos\left(\frac{\theta}{2}\right) & -e^{i\lambda}\sin\left(\frac{\theta}{2}\right) \\ e^{i\phi}\sin\left(\frac{\theta}{2}\right) & e^{i(\phi+\lambda)}\cos\left(\frac{\theta}{2}\right) \end{pmatrix}$$

Construct all possible gates from $U(\theta,\phi,\lambda)$

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \equiv U\left(\frac{\pi}{2}, 0, \pi\right)$$



$$|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

Quantum Gate Universality [DiV95]: The above “building blocks” can construct any quantum circuit acting on n qubits, i.e. $\mathrm{SU}(2^n)$, operating on at most *two-qubits* at a time.

Quantum gates and universality

Single qubit gates:

- A generic quantum gate can be decomposed in a series of R_y and R_z [BBC+95]

$$U(\theta,\phi,\lambda) = R_z(\lambda)R_y(\theta)R_z(\phi)$$

Multi-qubit gates:

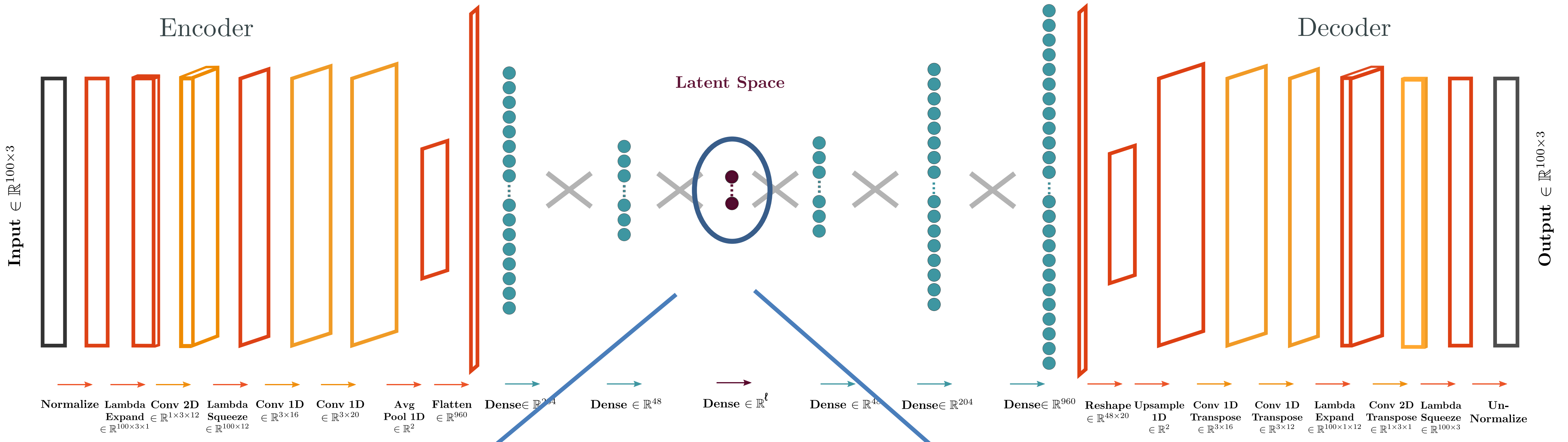
- 2-qubit SWAP and CNOT (Control-X) gates and the 3-qubit Toffolli gate

$$CX = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

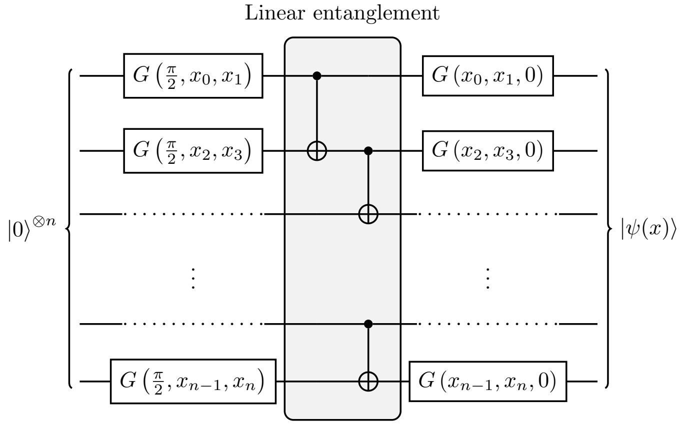
- Any control-U gate can be written as a combination of CX, R_y and R_z gates.

Quantum Gate Universality [DiV95]: The above "building blocks" can construct any quantum circuit acting on n qubits, i.e. SU(2^n), operating on at most *two-qubits* at a time.

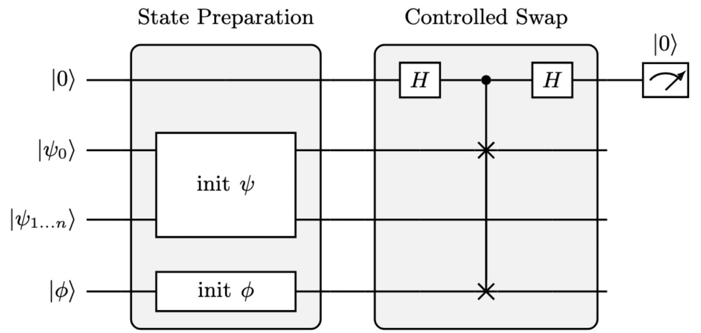
Convolutional autoencoder architecture



Unsupervised kernel machine

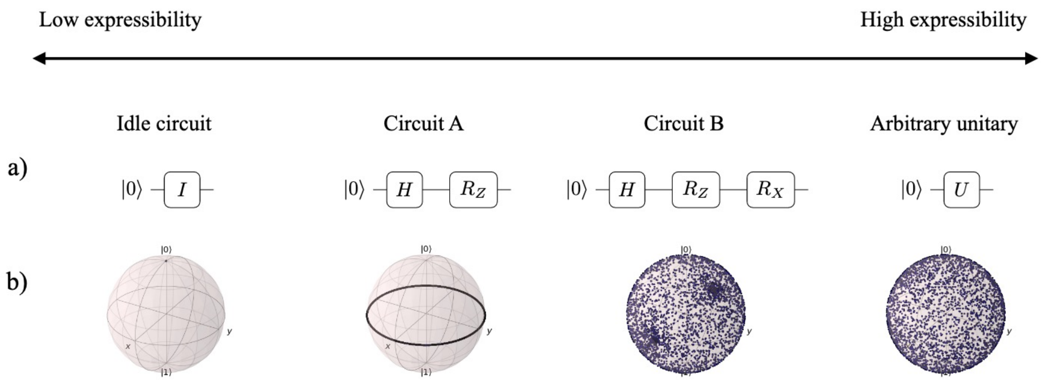


Quantum clustering algorithms

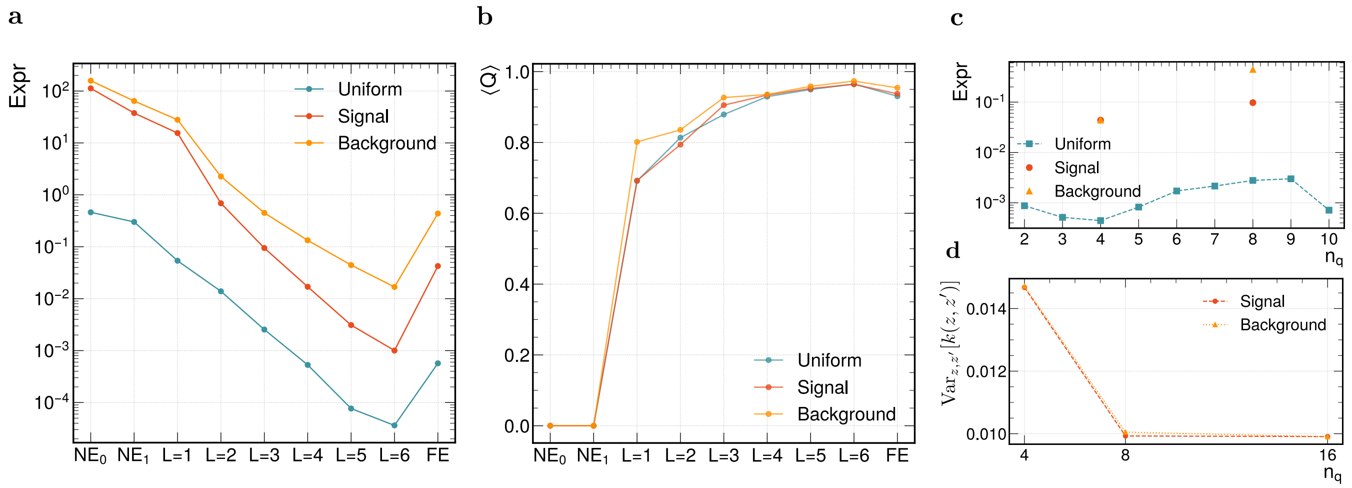


Expressibility and entanglement capability

Expressibility [S. Sim, et al., Adv. Quantum Technol. 2 (2019) 1900070]



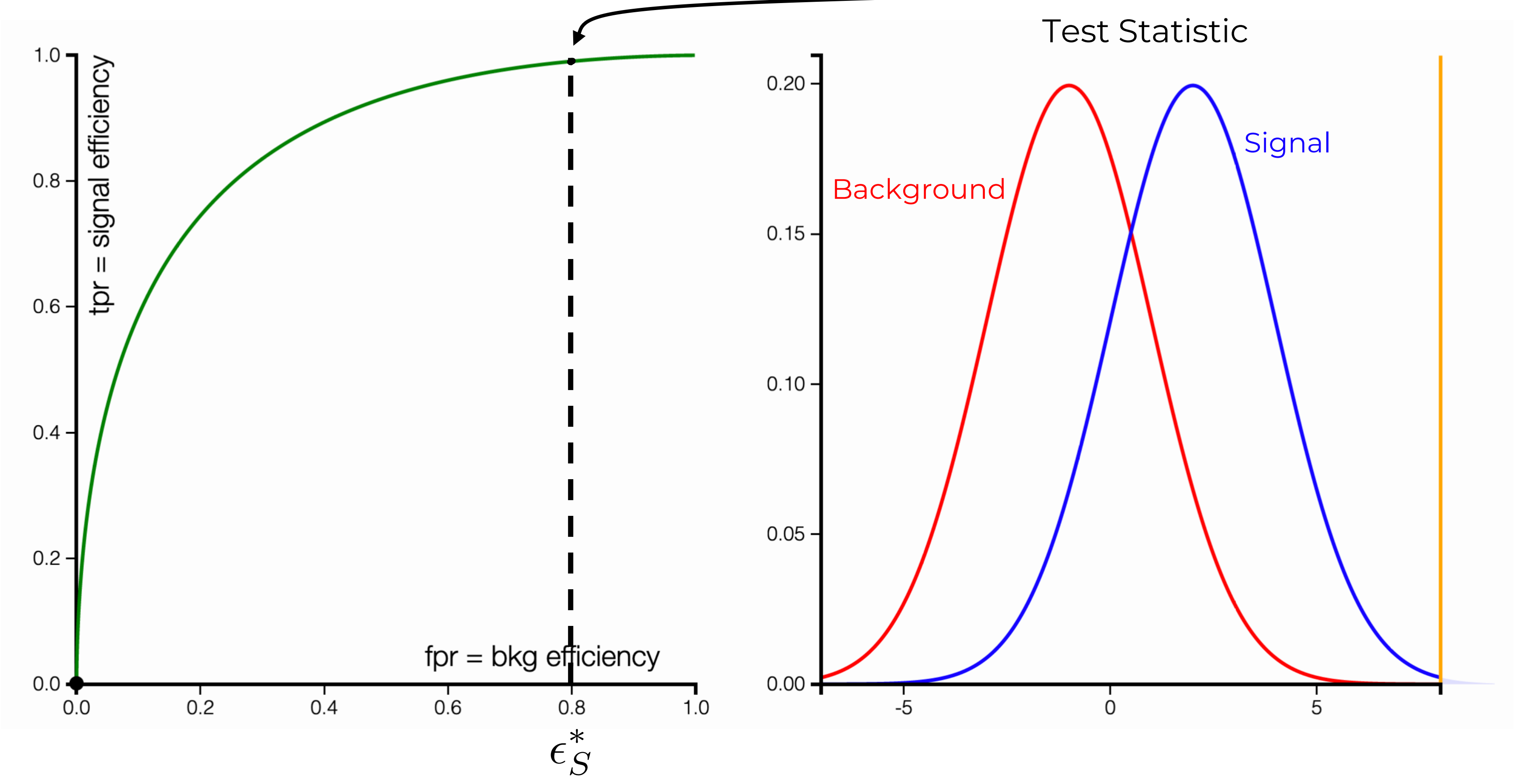
Expressibility & Entanglement capability of our data encoding circuit



[K.A. Wozniak*, **VB***, E. Puljak*, et al., arXiv:2301.10780]

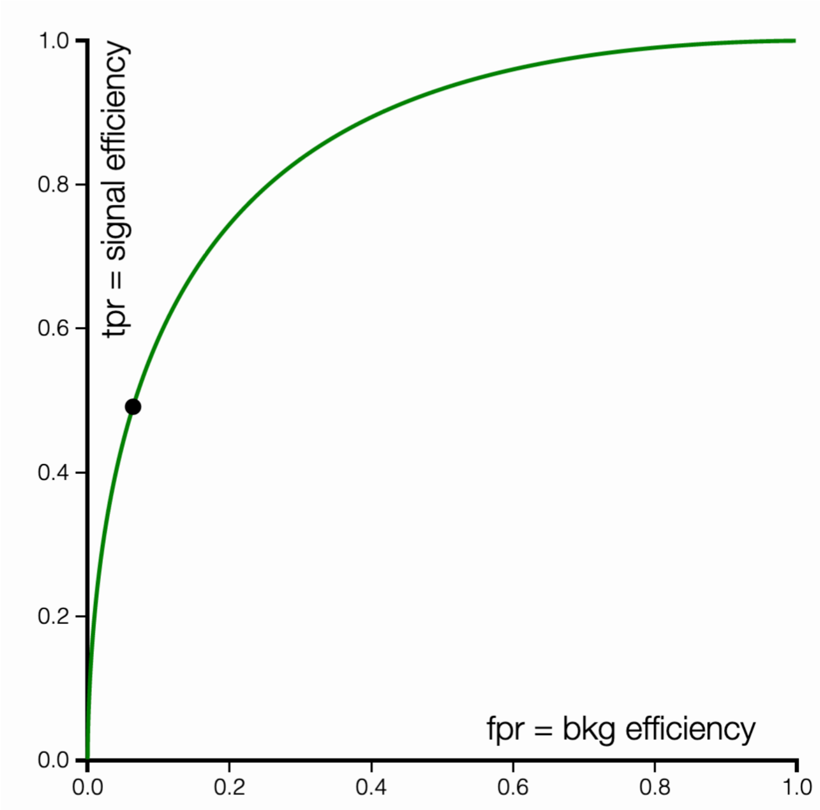
Performance Metric

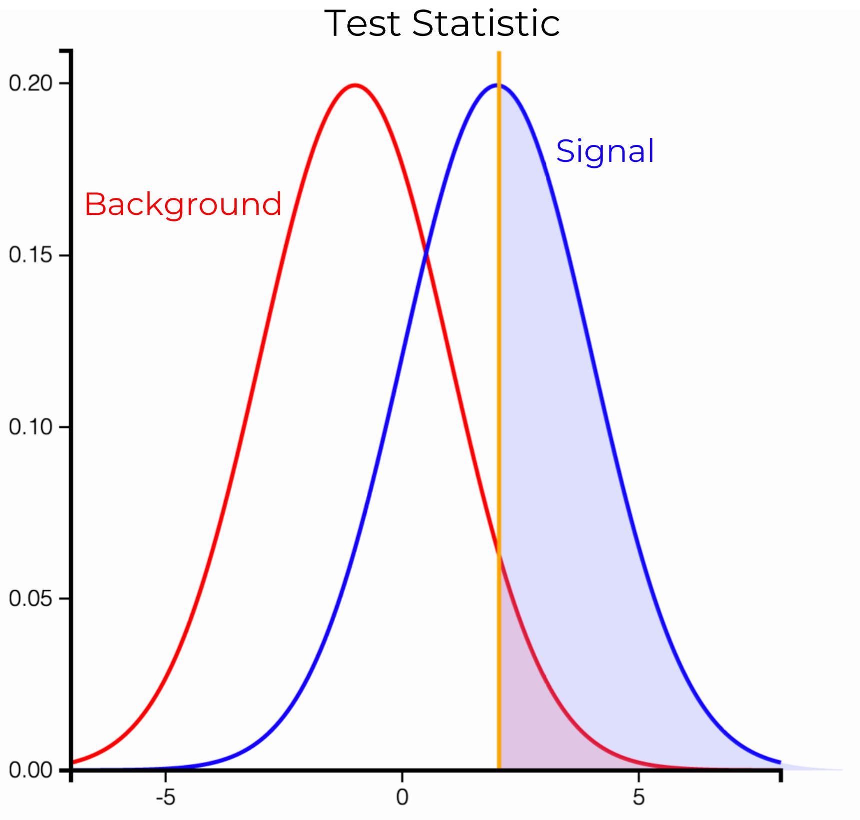
- The normalised data samples are split into training, validation, and testing data sets.
- Classification power metric: Receiver Operating Characteristic (ROC) curve.
- More compact metric: Area Under Curve (AUC) of the ROC curve.
- More practical metric: working point of an analysis $\epsilon_B(\epsilon_S^*)$



Performance Metric

- The normalised data samples are split into training, validation, and testing data sets.
- Classification power metric: Receiver Operating Characteristic (ROC) curve.
- More compact metric: Area Under Curve (AUC) of the ROC curve.
- More practical metric: working point of an analysis $\epsilon_B(\epsilon_S^*)$

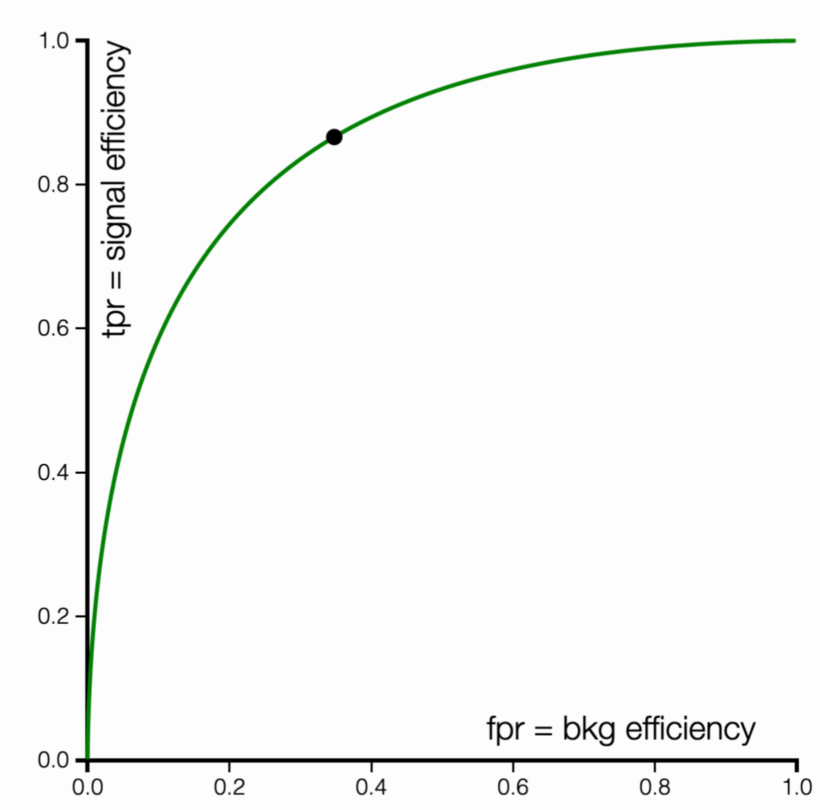


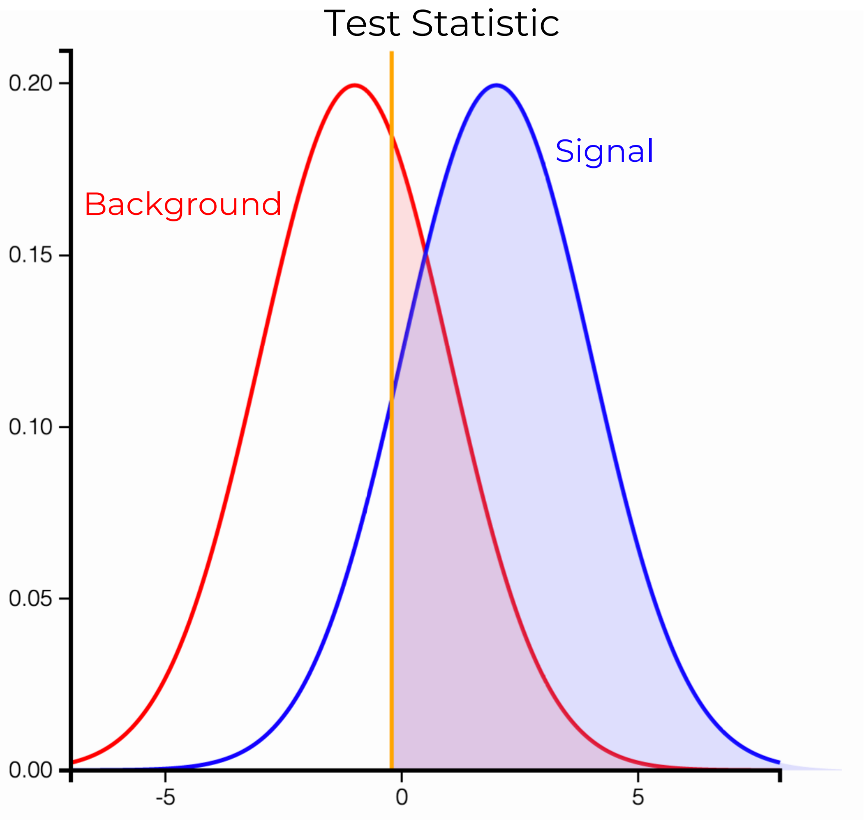


ETH zürich

Performance Metric

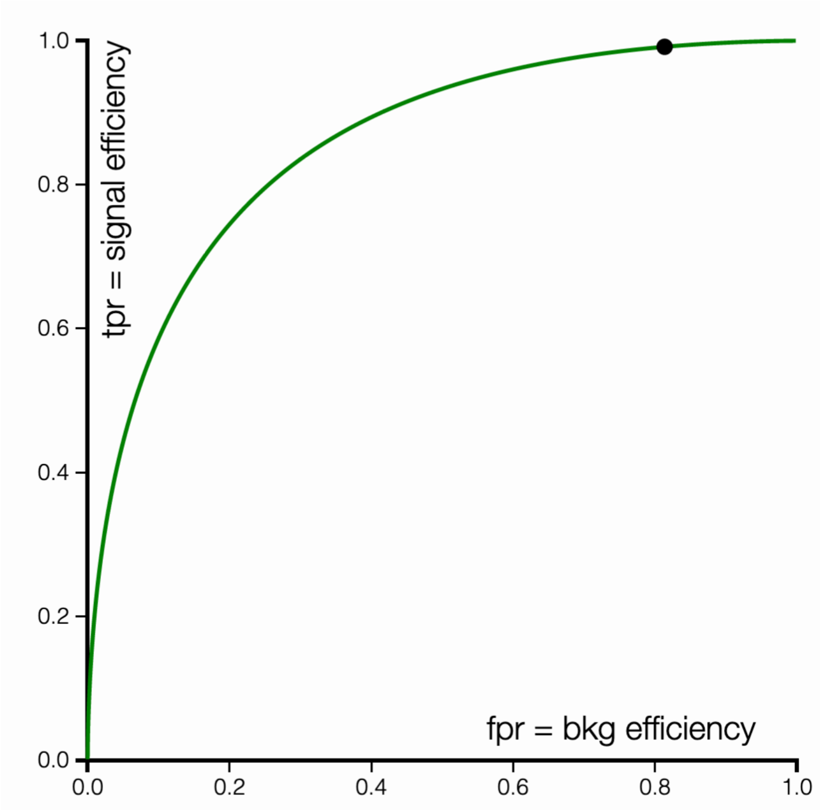
- The normalised data samples are split into training, validation, and testing data sets.
- Classification power metric: Receiver Operating Characteristic (ROC) curve.
- More compact metric: Area Under Curve (AUC) of the ROC curve.
- More practical metric: working point of an analysis $\epsilon_B(\epsilon_S^*)$

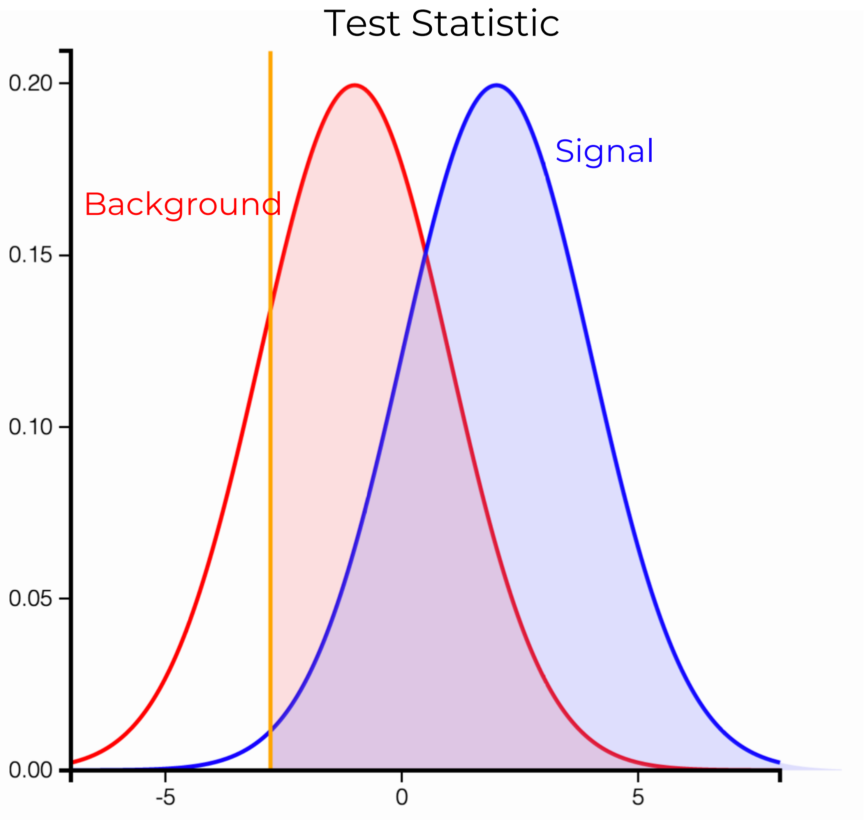




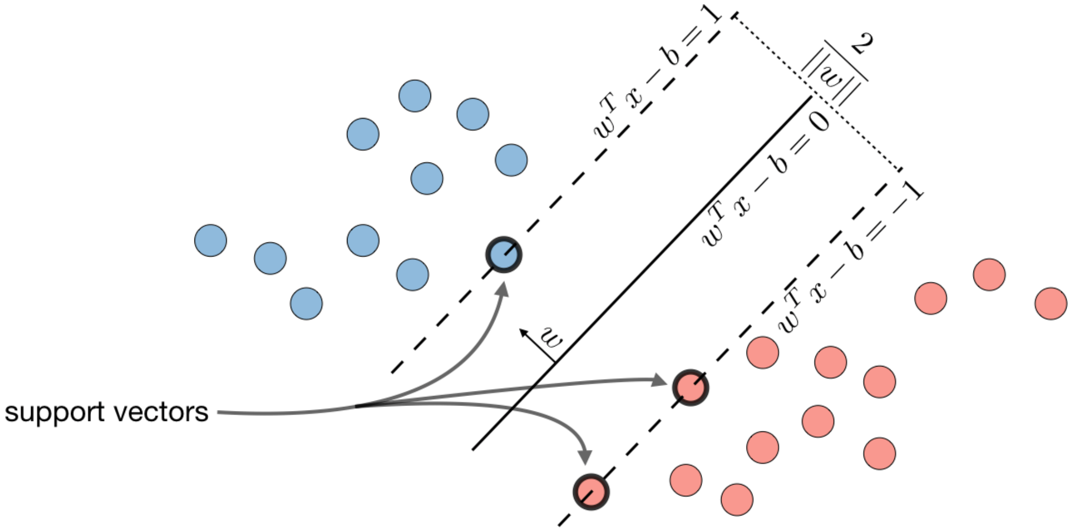
Performance Metric

- The normalised data samples are split into training, validation, and testing data sets.
- Classification power metric: Receiver Operating Characteristic (ROC) curve.
- More compact metric: Area Under Curve (AUC) of the ROC curve.
- More practical metric: working point of an analysis $\epsilon_B(\epsilon_S^*)$





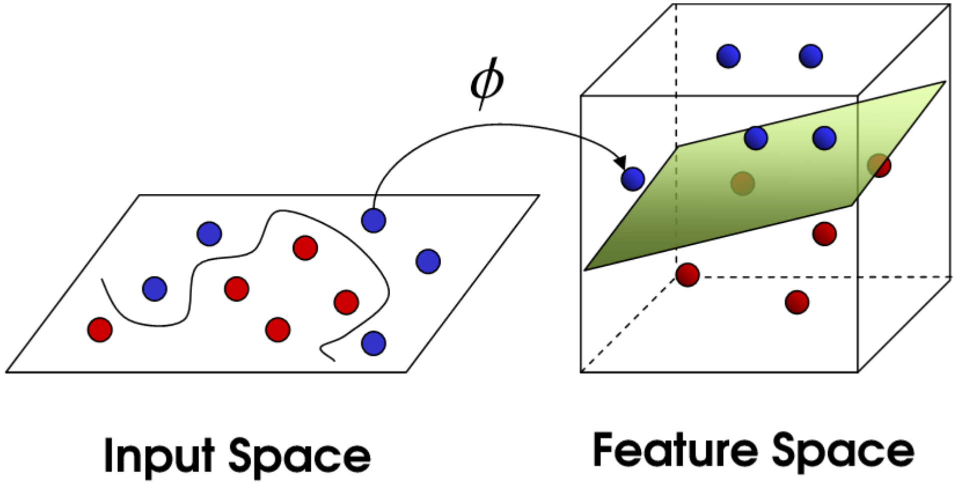
Quantum Support Vector Machines



SVM objective function is equivalent to (dual Lagrangian):

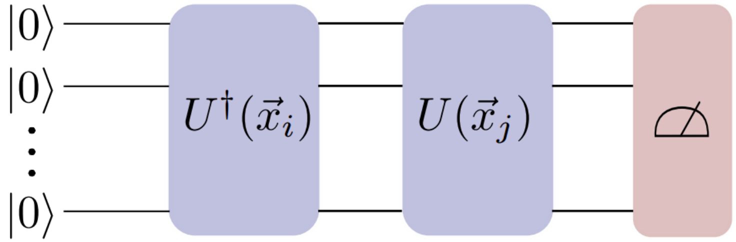
maximize $$L(c_1, \ldots, c_n) = \sum_{i=1}^{n} c_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i c_i (\vec{x}_i \cdot \vec{x}_j) y_j c_j$$

subject to $$\sum_{i=1}^{n} c_i y_i = 0, \text{ and } 0 \leq c_i \leq C, \; \forall i$$



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

Make the kernel *quantum*

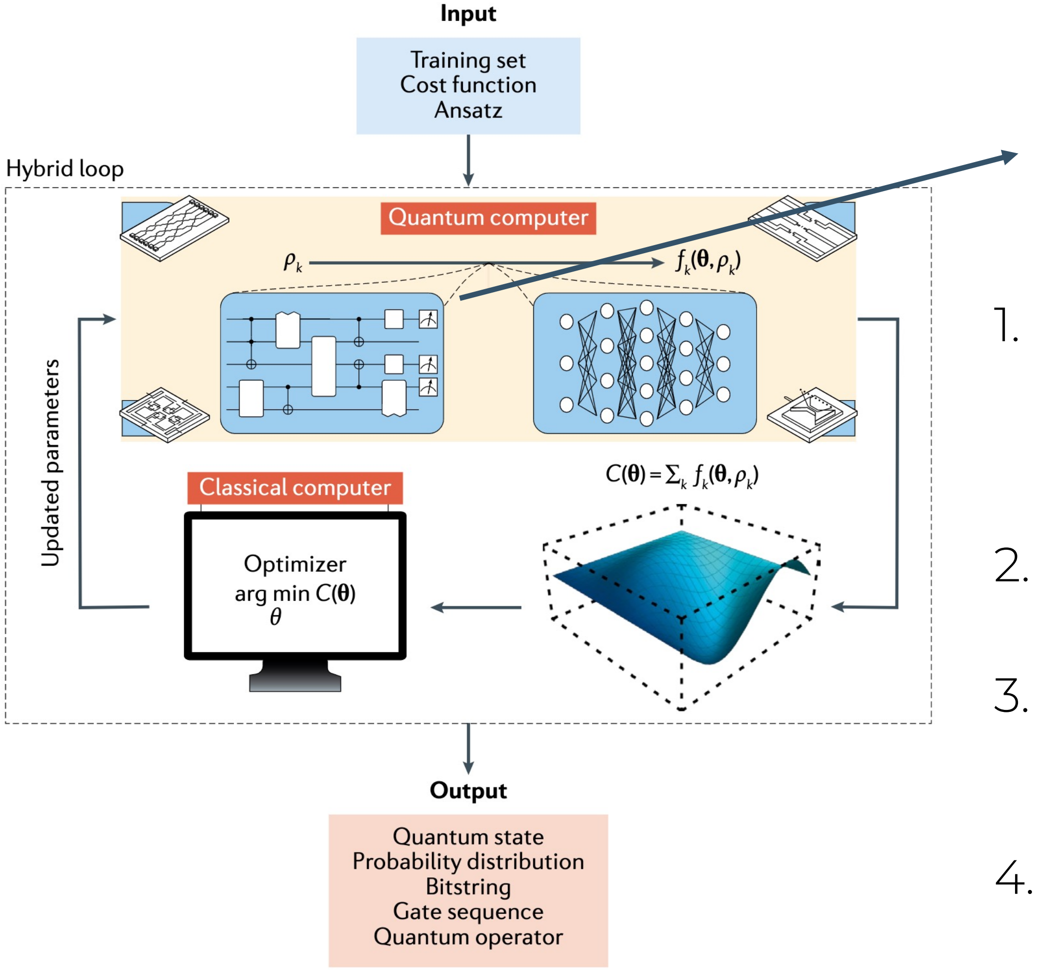


$\Rightarrow K_{ij} = |\langle 0|U^\dagger(\vec{x}_i) U(\vec{x}_j)|0\rangle|^2$

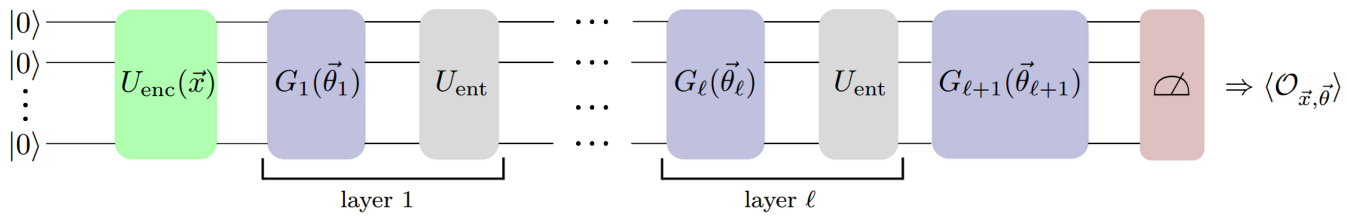
***Can be generalised to unsupervised learning**

Quantum Neural Networks

Variational quantum algorithm workflow



[M. Cerezo, et al. Nat. Rev. Phys. 3, 625–644 (2021)]



1. Choose loss function
   Task dependent: e.g. classification, reconstruction, generative modeling.

2. Embed classical data to circuit.

3. Process quantum state with parametrized quantum gates.

4. Update trainable parameters

$$\Theta_{t+1} \leftarrow \Theta_t - \eta \nabla_\Theta \mathcal{L}[\langle \mathcal{O}(x;\Theta) \rangle]$$