# Quantum anomaly detection in the latent space of proton collision events at the LHC

**Vasilis Belis**[1] [arXiv:2301.10780]

K. Woźniak[2,3], E. Puljak[2,4], P. Barkoutsos[5,*], G. Dissertori[1], M. Grossi[2], M. Pierini[2], F. Reiter[1], I. Tavernelli[5], S. Vallecorsa[2]

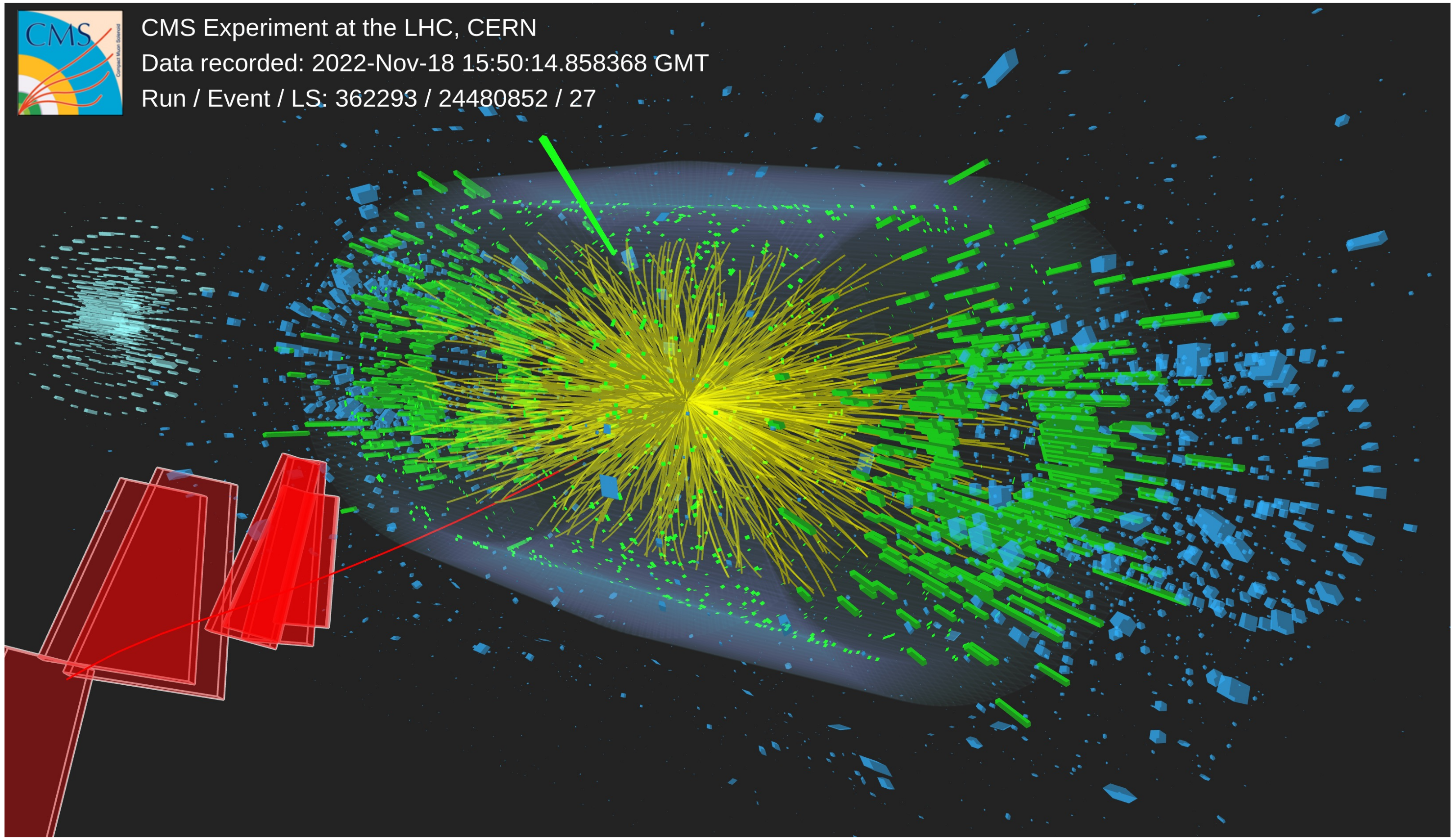
*currently Pasqal

23 November, CERN

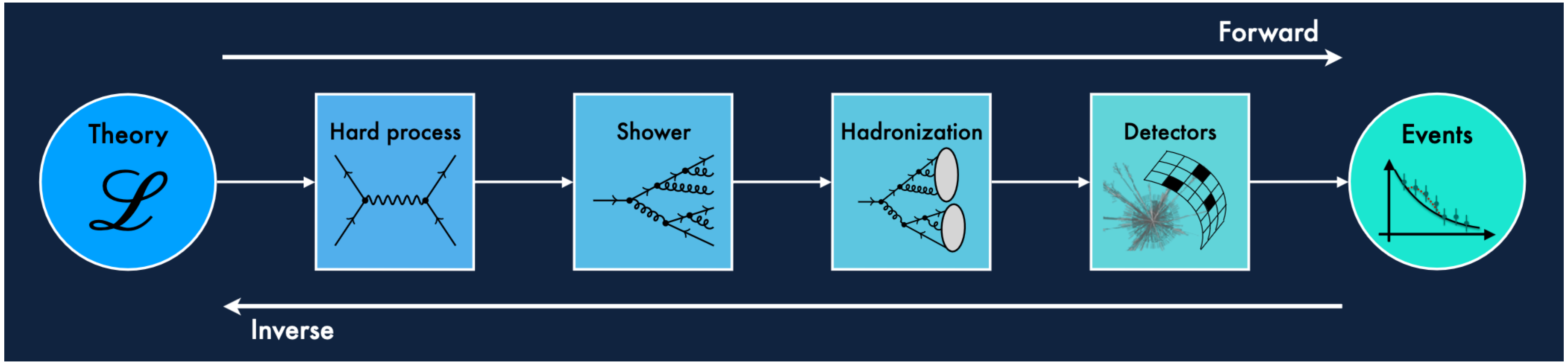
QTML 2023

# No cats and dogs exploited to explain this work

# No cats and dogs exploited to explain this work



# Why QML for HEP?

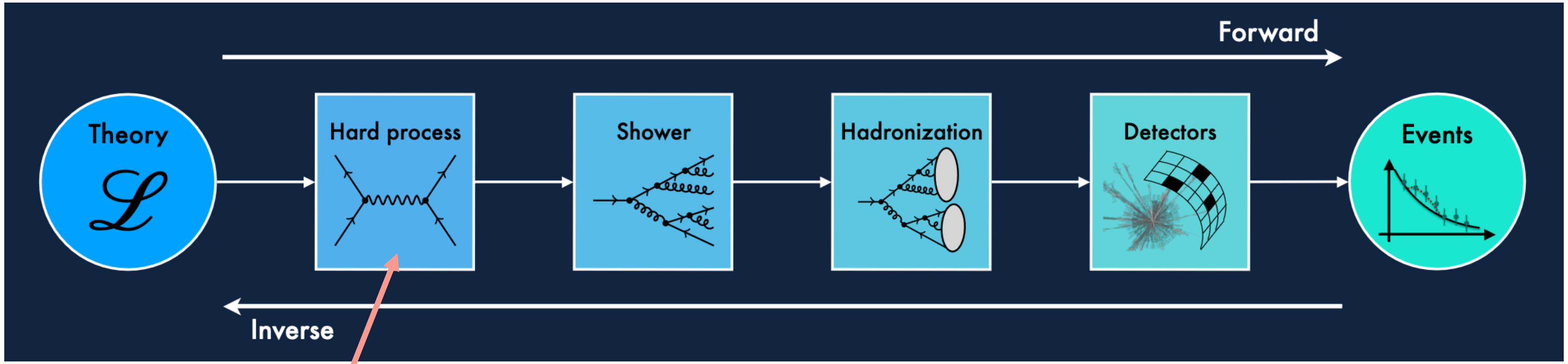



**Fundamental motivation**

Utilise information and correlations inherent in HEP data.

Exploit “quantum remnants” in data.

# Why QML for HEP?



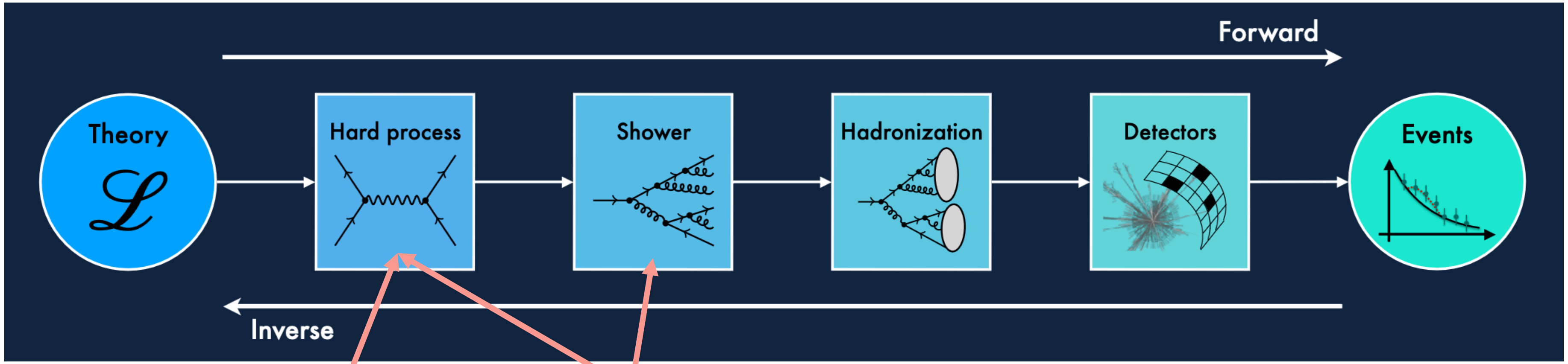

entanglement [1703.02989]

## Fundamental motivation

Utilise information and correlations inherent in HEP data.

Exploit "quantum remnants" in data.

# Why QML for HEP?




interference [2110.10112]

entanglement [1703.02989]

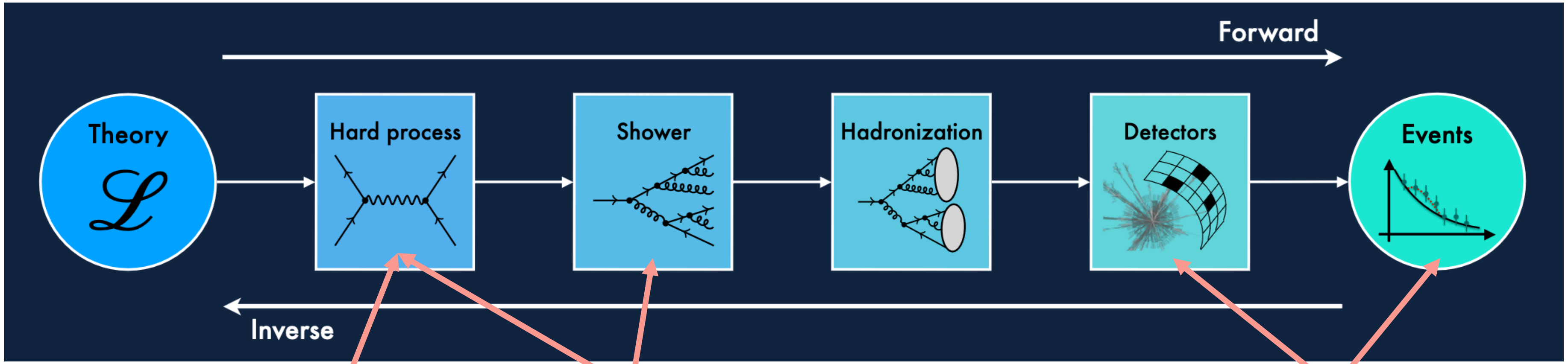
spin correlations [1907.03729]

**Fundamental motivation**

Utilise information and correlations inherent in HEP data.

Exploit “quantum remnants” in data.

# Why QML for HEP?




interference [2110.10112]  Bell inequalities [2102.11883, 2203.05582]

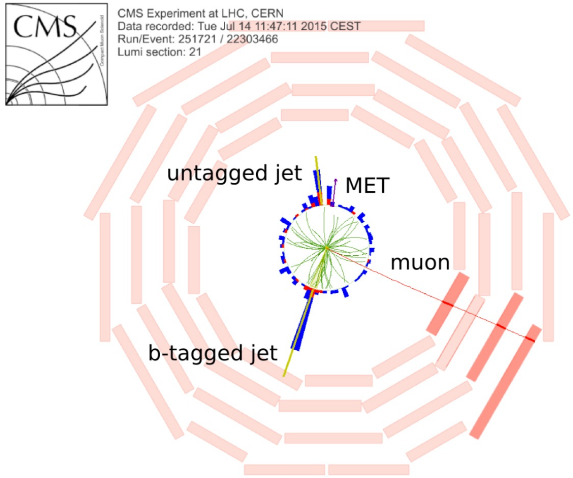
entanglement [1703.02989]  spin correlations [1907.03729]

**Fundamental motivation**

Utilise information and correlations inherent in HEP data.

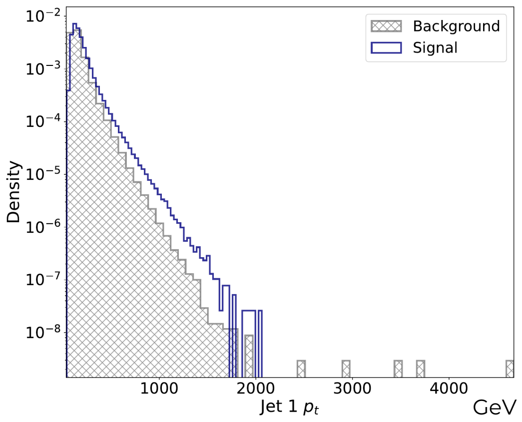
Exploit “quantum remnants” in data.

# Typical workflow: Model-dependent searches

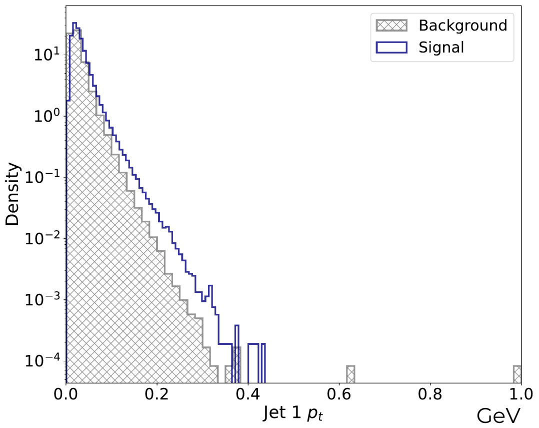



**Theory + Simulation**

compute prob. distributions

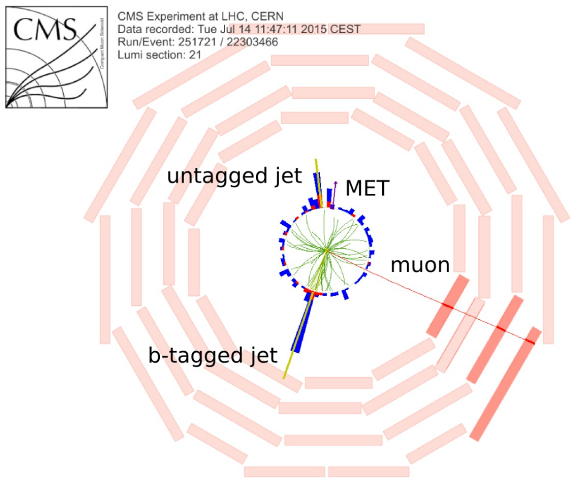



apply cuts and normalise



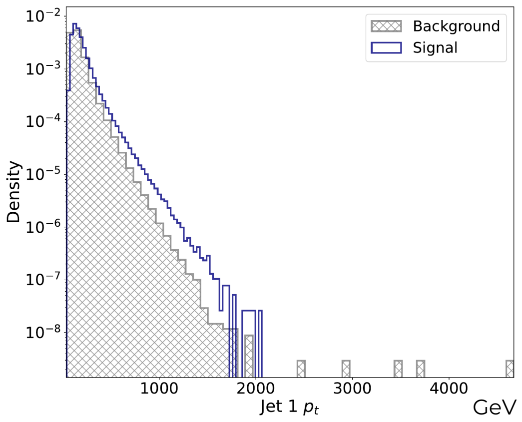

**Input for classifiers models**

# Typical workflow: Model-dependent searches

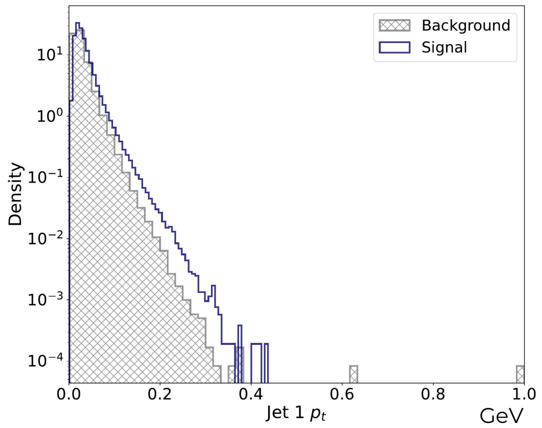



**Theory + Simulation**

compute prob. distributions




apply cuts and normalise




**Input for classifiers models**

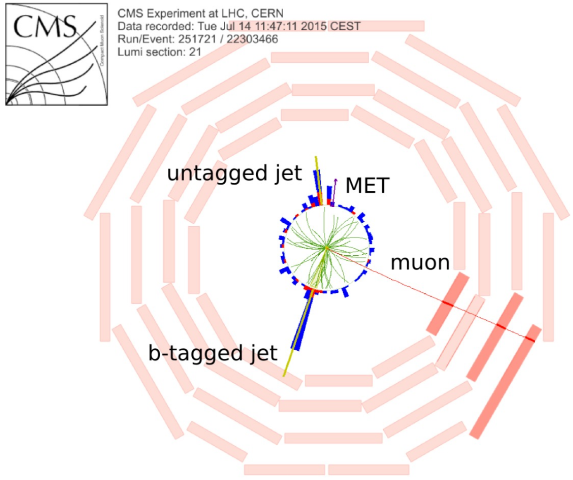
Classifier models

produce the *test statistic*

Likelihood ratio between different hypotheses

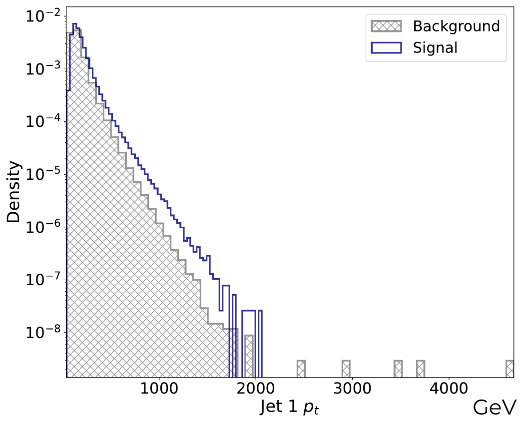
$$q = -2 \log \frac{\mathcal{L}(x|S)}{\mathcal{L}(x|B)}$$

# Typical workflow: Model-dependent searches

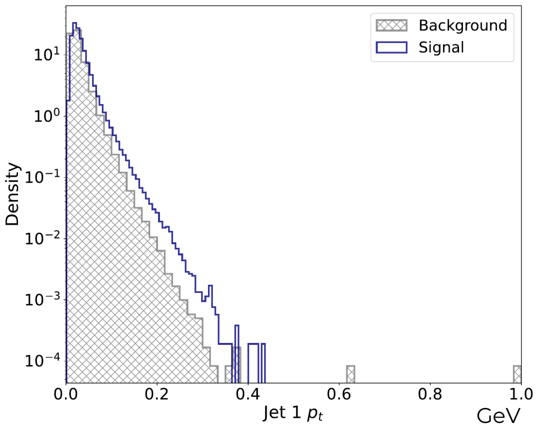



**Theory + Simulation**

compute prob. distributions




apply cuts and normalise



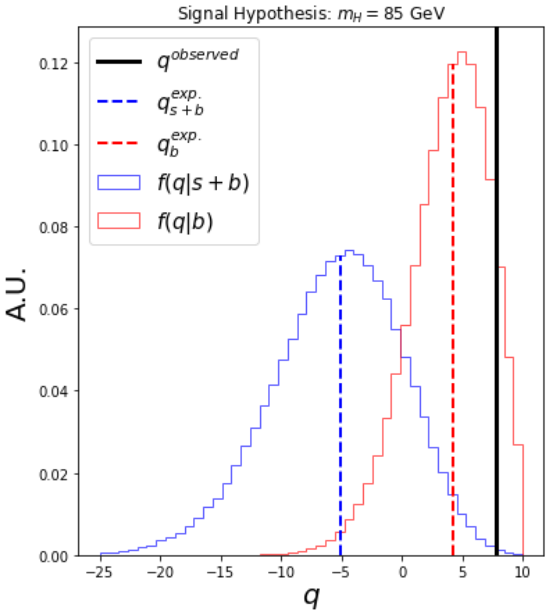

**Input for classifiers models**

Classifier models

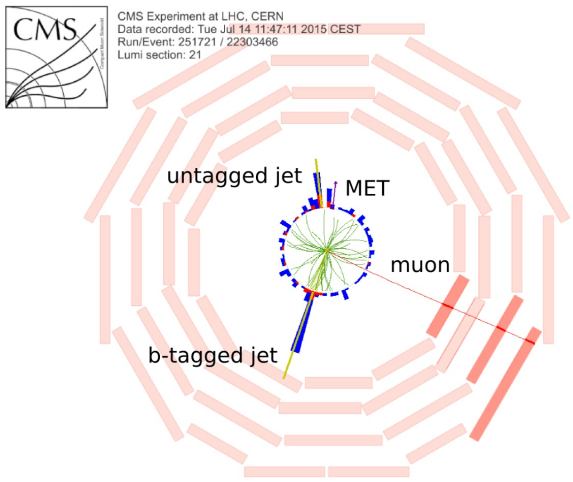
produce the *test statistic*

Likelihood ratio between different hypotheses

$$q = -2 \log \frac{\mathcal{L}(x|S)}{\mathcal{L}(x|B)}$$

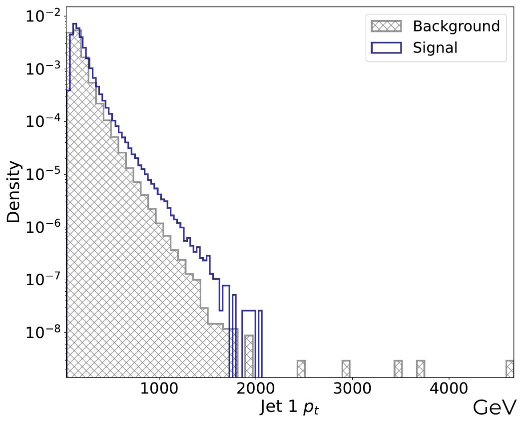


# Typical workflow: Model-dependent searches

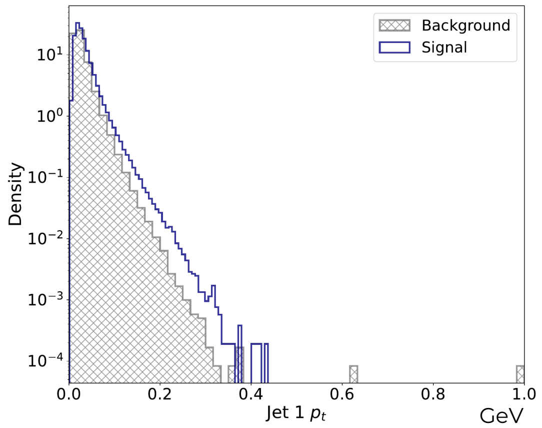



**Theory + Simulation**

compute prob. distributions




apply cuts and normalise



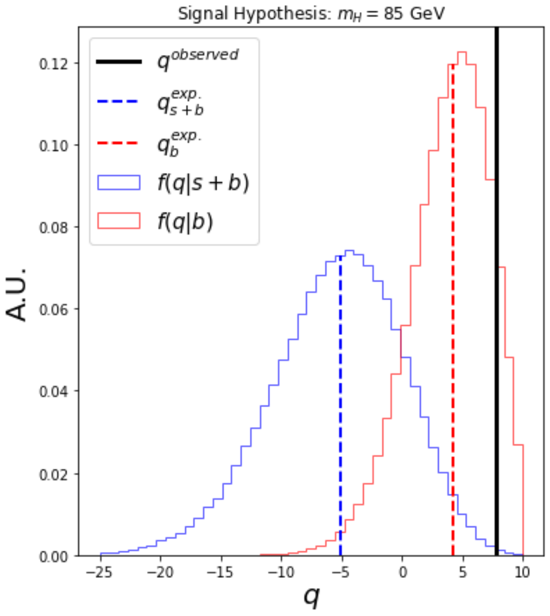

**Input for classifiers models**

Classifier models

produce the *test statistic*

Likelihood ratio between different hypotheses

$$q = -2 \log \frac{\mathcal{L}(x|S)}{\mathcal{L}(x|B)}$$




**What if you do not know the signal?**

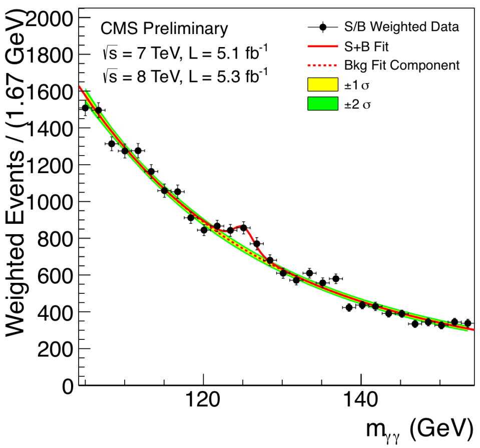
# Motivating model-independent searches

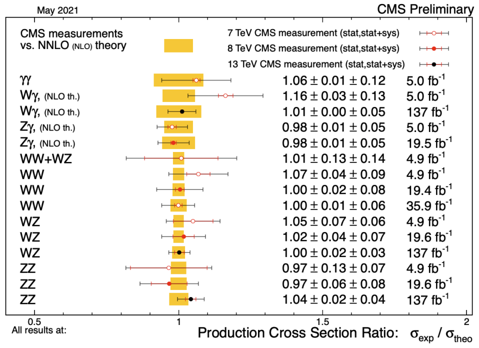
Bias is **not necessarily bad**. It can be **great!**

Higher bias ⟶ lower variance of the statistical model

[Cramér–Rao bound]

[10.1016/j.physletb.2012.08.02]




Summary SM measurements @ CMS [link]

# Motivating model-independent searches

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

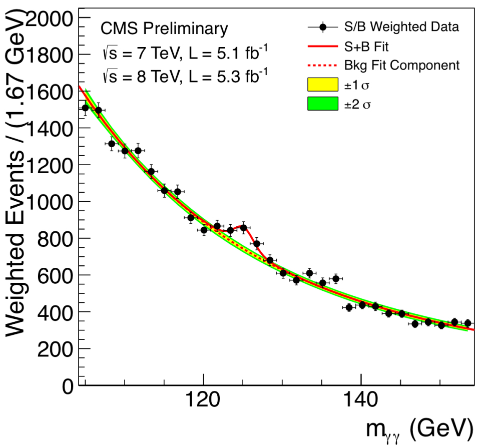
Higher bias ⟶ lower variance of the statistical model

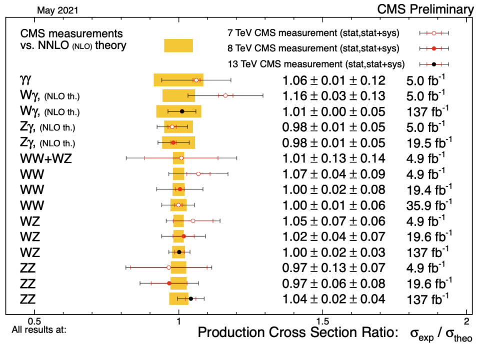
[Cramér–Rao bound]

So far, new-physics searches at the LHC:

assume SM + choose to test a signal hypothesis….

[10.1016/j.physletb.2012.08.02]

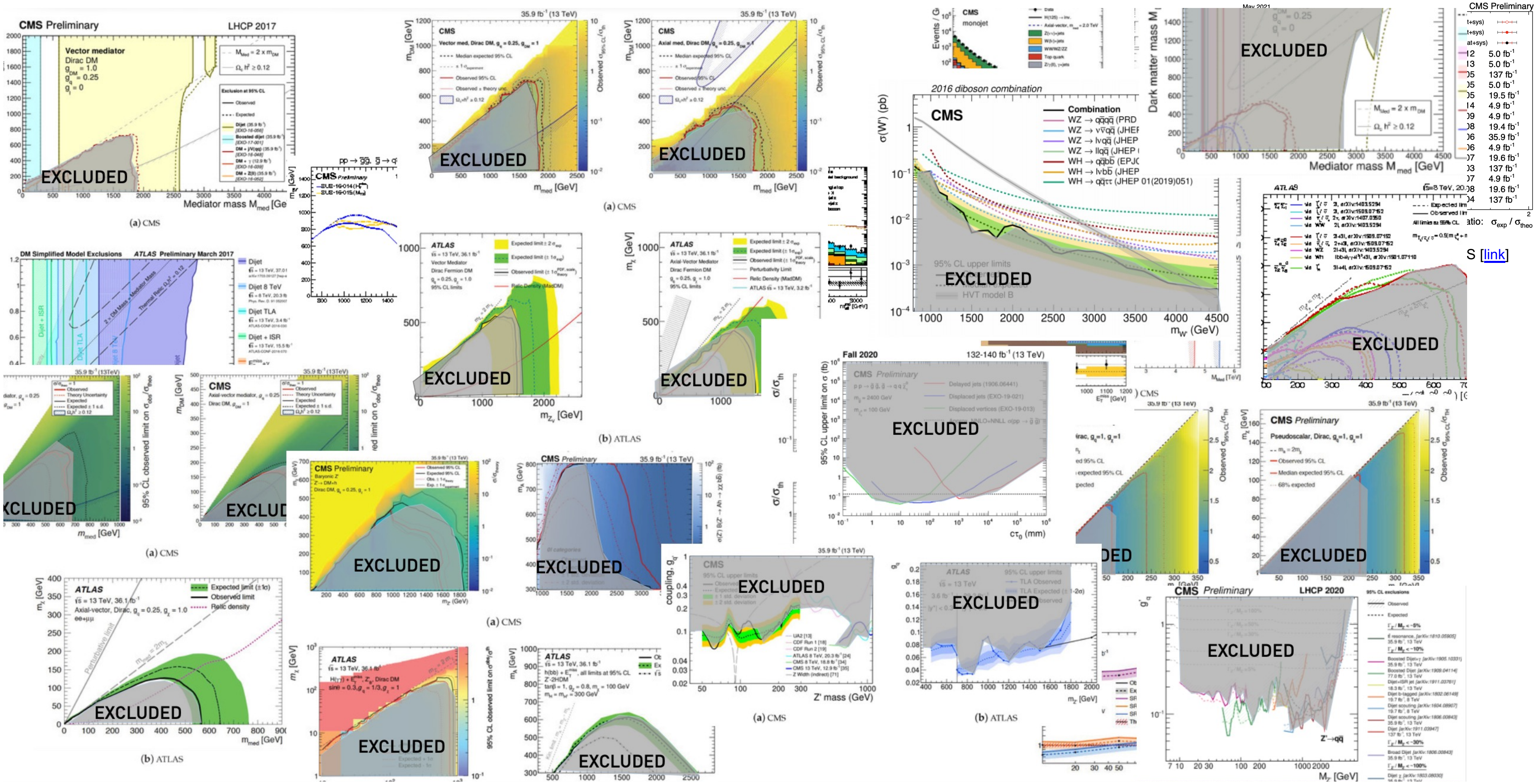



Summary SM measurements @ CMS [link]

# Motivating model-independent searches

[10.1016/j.physletb.2012.08.02]



# Motivating model-independent searches

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

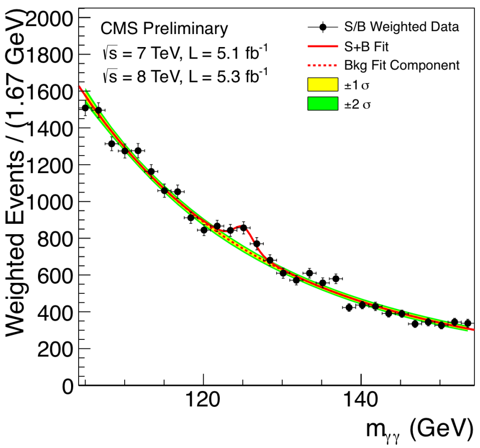
Higher bias ⟶ lower variance of the statistical model

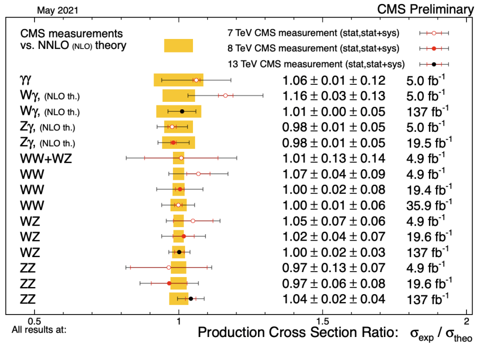
[Cramér–Rao bound]

So far, new-physics searches at the LHC:

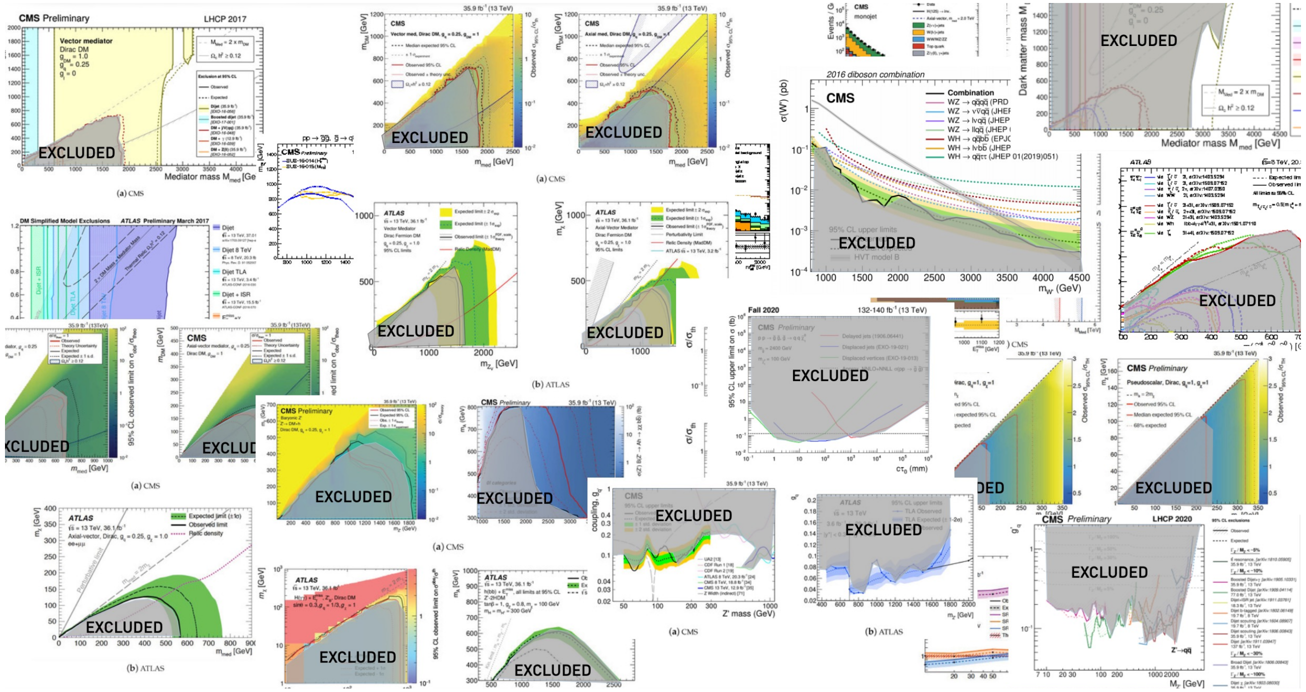
assume SM + choose to test a signal hypothesis….

[10.1016/j.physletb.2012.08.02]




Summary SM measurements @ CMS [link]



# Motivating model-independent searches

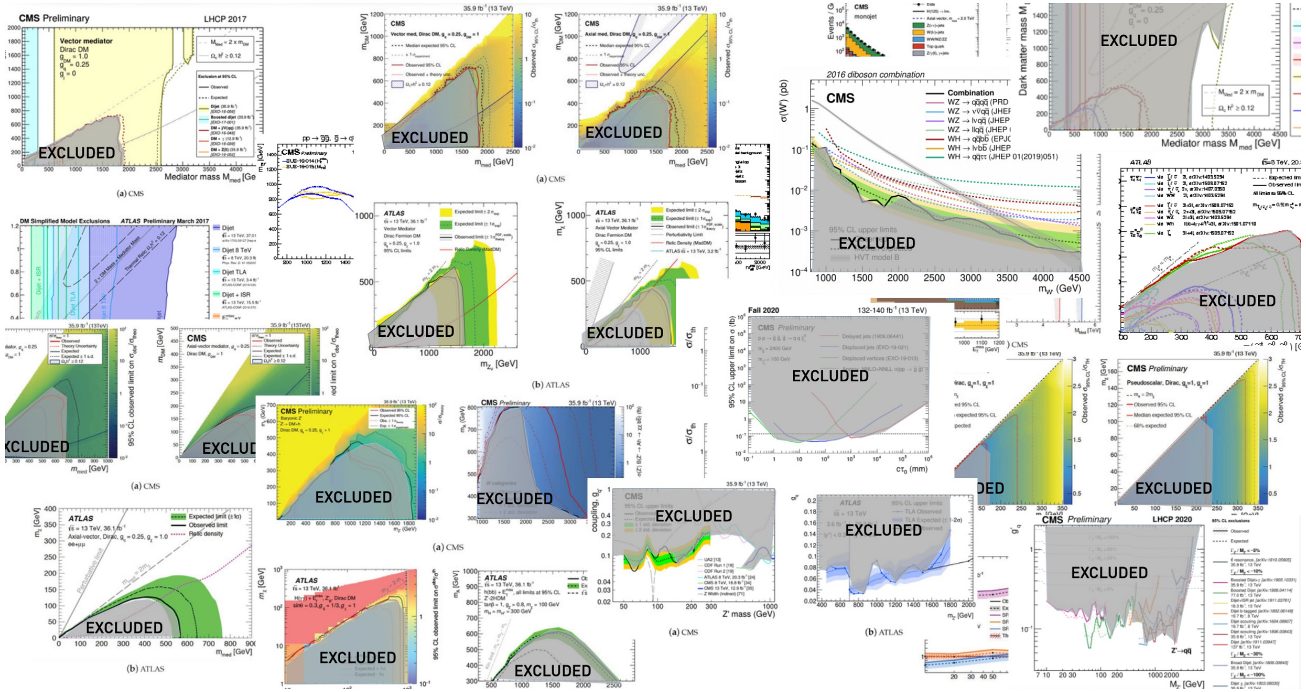
Bias is **not necessarily bad**.  It can be **great!**

Higher bias ⟶ lower variance of the statistical model

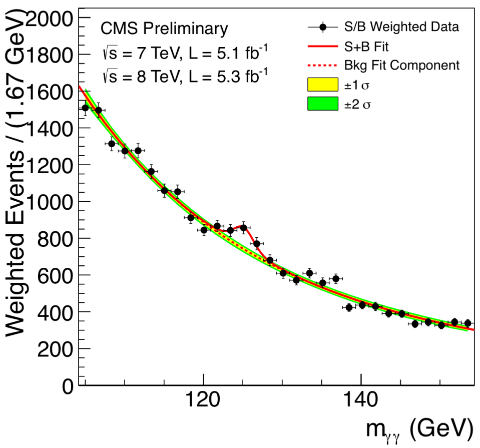
[Cramér–Rao bound]

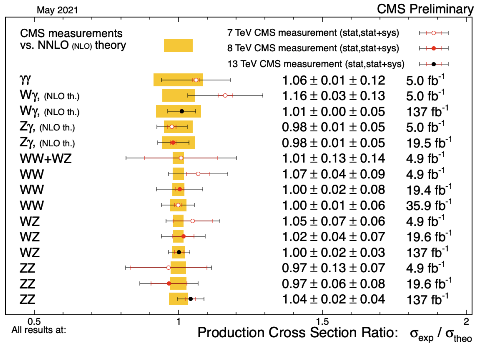
So far, new-physics searches at the LHC:

assume SM + choose to test a 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.

# Motivating model-independent searches

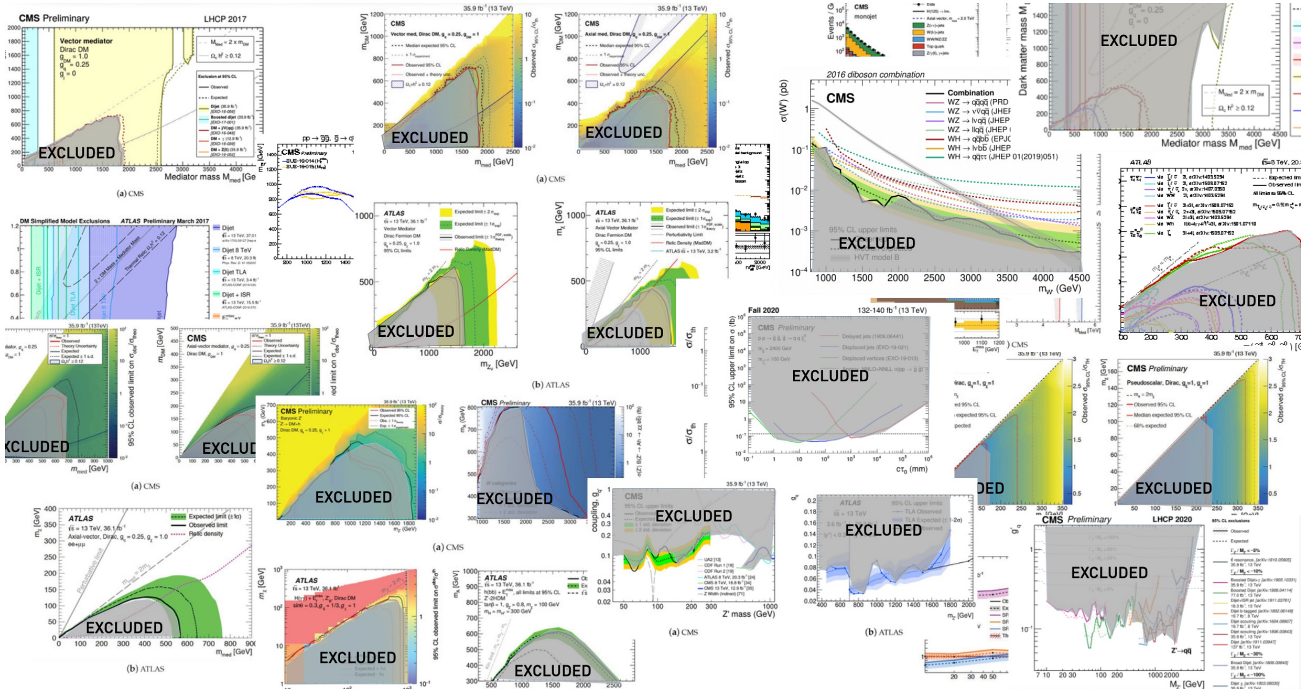
Bias is **not necessarily bad**. It can be **great!**

Higher bias ⟶ lower variance of the statistical model

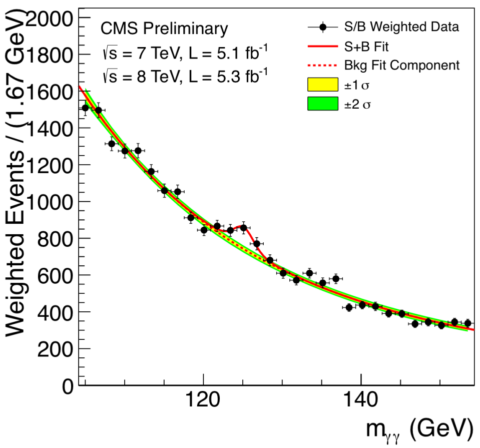
[Cramér–Rao bound]

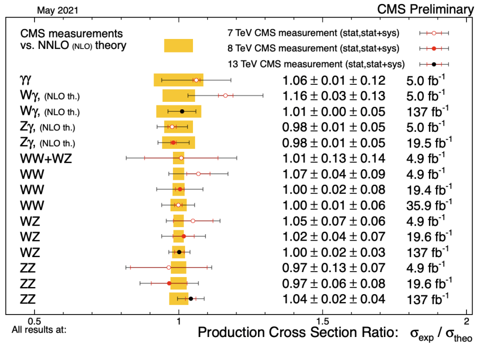
So far, new-physics searches at the LHC:

assume SM + choose to test a signal hypothesis….



[10.1016/j.physletb.2012.08.02]



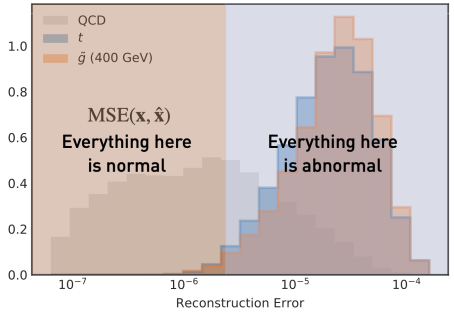




Summary SM measurements @ CMS [link]

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

Look at nature with minimal bias.

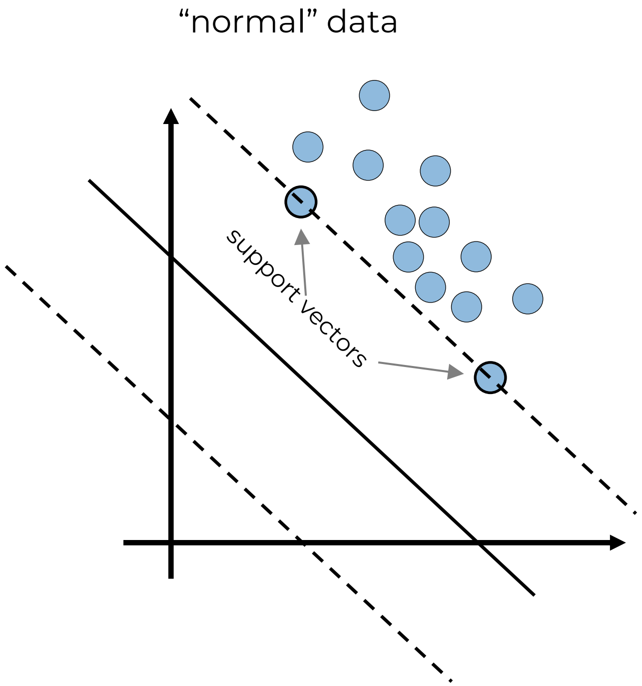
One possible solution:
**Anomaly detection**




**Unsupervised ML models**

ETH zürich

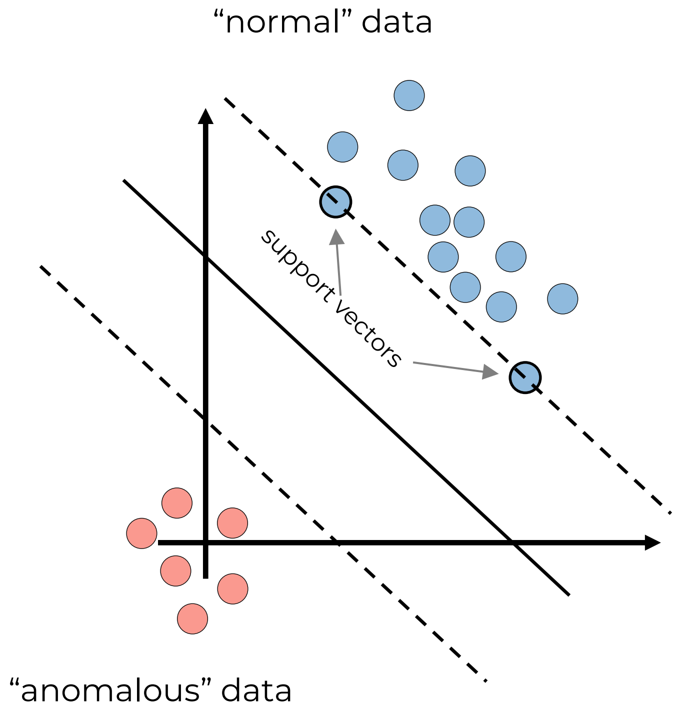
# Unsupervised quantum kernel machines




$$\min_{w\in\mathcal{H},\, \xi\in\mathbb{R}^{\mathrm{N}},\, \rho\in\mathbb{R}} \quad \frac{1}{2}||w||^2 + \frac{1}{\nu \mathrm{N}}\sum_i \xi_i - \rho$$

$$\text{subject to} \quad w\cdot\phi(x_i) \geq \rho - \xi_i,\ \xi_i \geq 0,\ \forall i$$

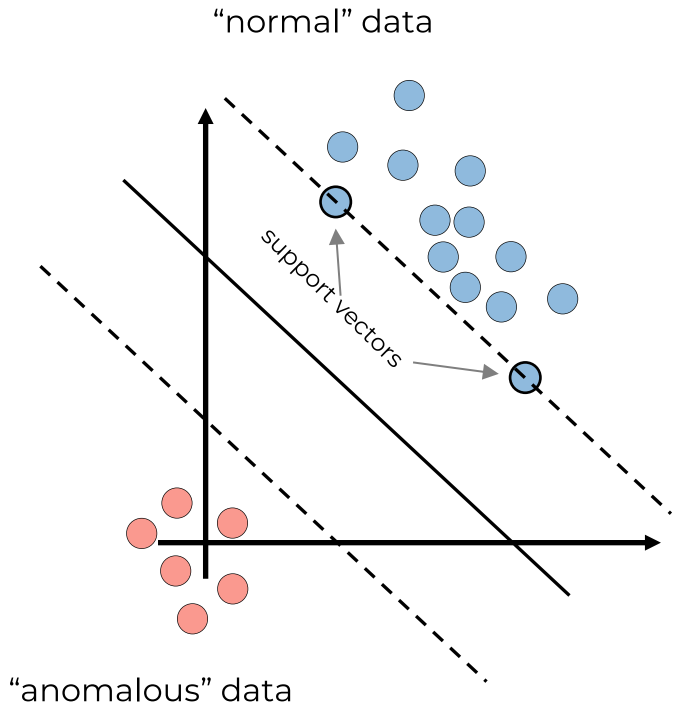
# Unsupervised quantum kernel machines

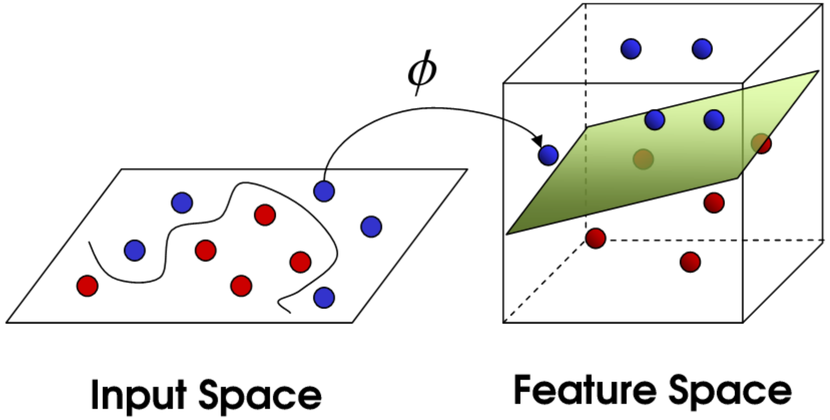



$$\min_{w\in\mathcal{H},\,\xi\in\mathbb{R}^{\mathrm{N}},\,\rho\in\mathbb{R}} \quad \frac{1}{2}||w||^2 + \frac{1}{\nu\mathrm{N}}\sum_i \xi_i - \rho$$

$$\text{subject to} \quad w\cdot\phi(x_i) \geq \rho - \xi_i,\ \xi_i \geq 0,\ \forall i$$

# Unsupervised quantum kernel machines



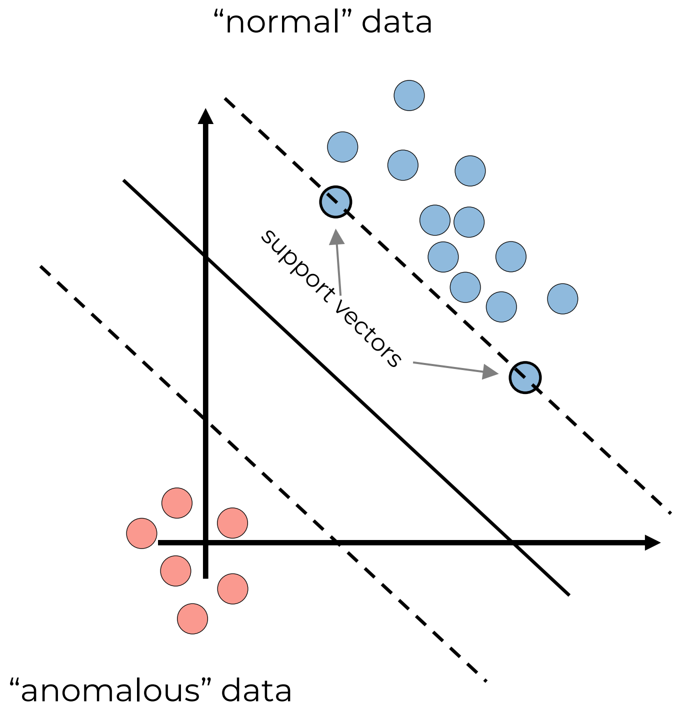

$$\min_{w\in\mathcal{H},\,\xi\in\mathbb{R}^{\mathrm{N}},\,\rho\in\mathbb{R}} \quad \frac{1}{2}||w||^2 + \frac{1}{\nu\mathrm{N}}\sum_i \xi_i - \rho$$

$$\text{subject to} \quad w\cdot\phi(x_i) \geq \rho - \xi_i,\; \xi_i \geq 0,\; \forall i$$

Kernel trick:

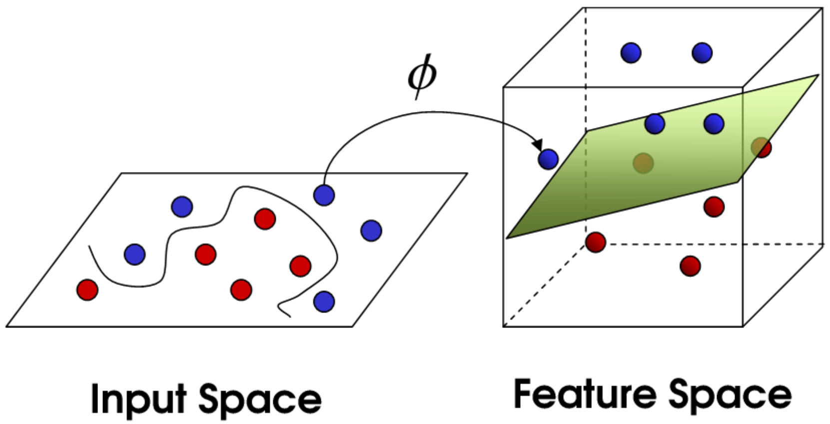
$$(\vec{x}_i\cdot\vec{x}_j) \mapsto k(\vec{x}_i\cdot\vec{x}_j) = \phi(\vec{x}_i)\cdot\phi(\vec{x}_j)$$

# Unsupervised quantum kernel machines



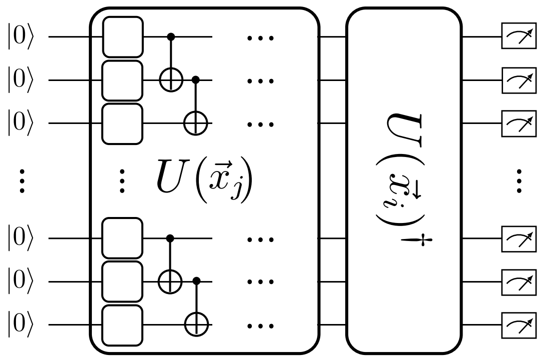

$$\min_{w\in\mathcal{H},\,\xi\in\mathbb{R}^N,\,\rho\in\mathbb{R}} \quad \frac{1}{2}||w||^2 + \frac{1}{\nu \mathrm{N}}\sum_i \xi_i - \rho$$

$$\text{subject to} \quad w\cdot\phi(x_i) \geq \rho - \xi_i,\ \xi_i \geq 0,\ \forall i$$




Kernel trick:

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

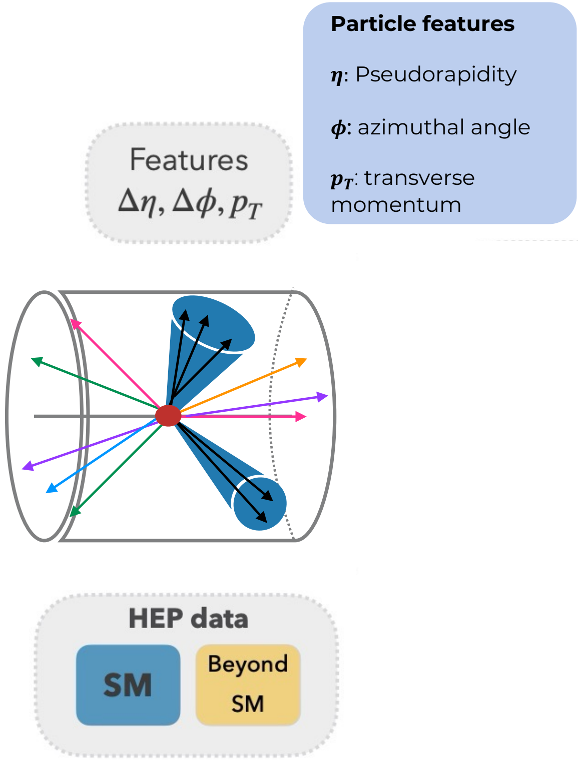



$$\Rightarrow \mathrm{K}_{ij} = |\,\langle 0|U^\dagger(\vec{x}_i)U(\vec{x}_j)|0\rangle\,|^2$$

# Results

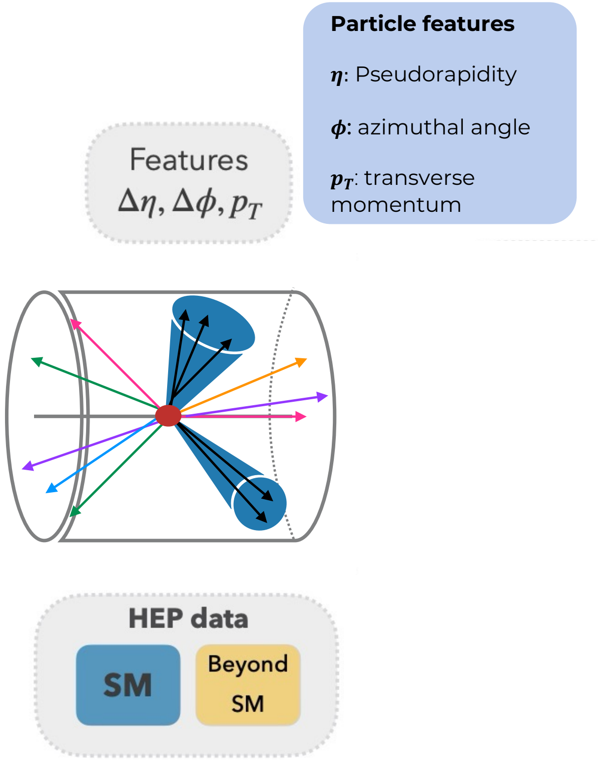
Finding new-physics in dijet events with QML

# Anomaly detection with quantum machine learning




**Background:** QCD dijet events. 600 features per event

# Anomaly detection with quantum machine learning

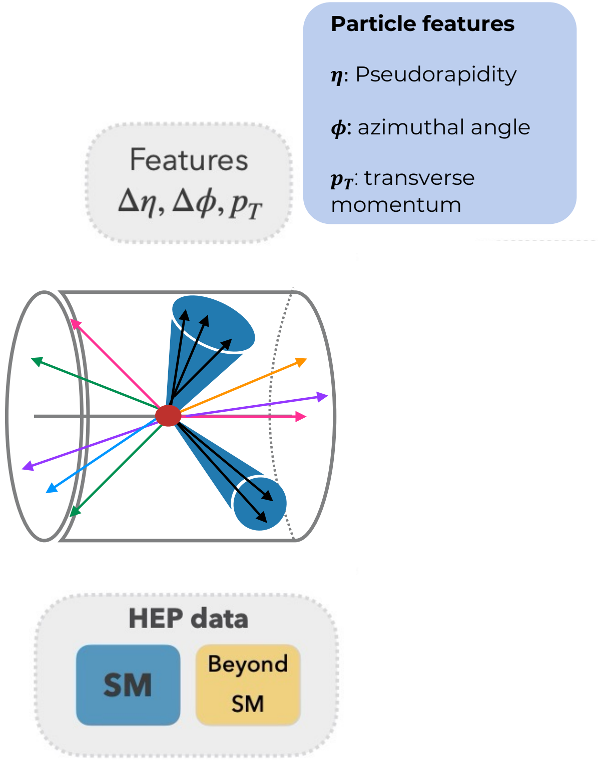



**Background:** QCD dijet events. 600 features per event

**BSM anomalies:** Graviton & New Scalar Boson

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

# Anomaly detection with quantum machine learning

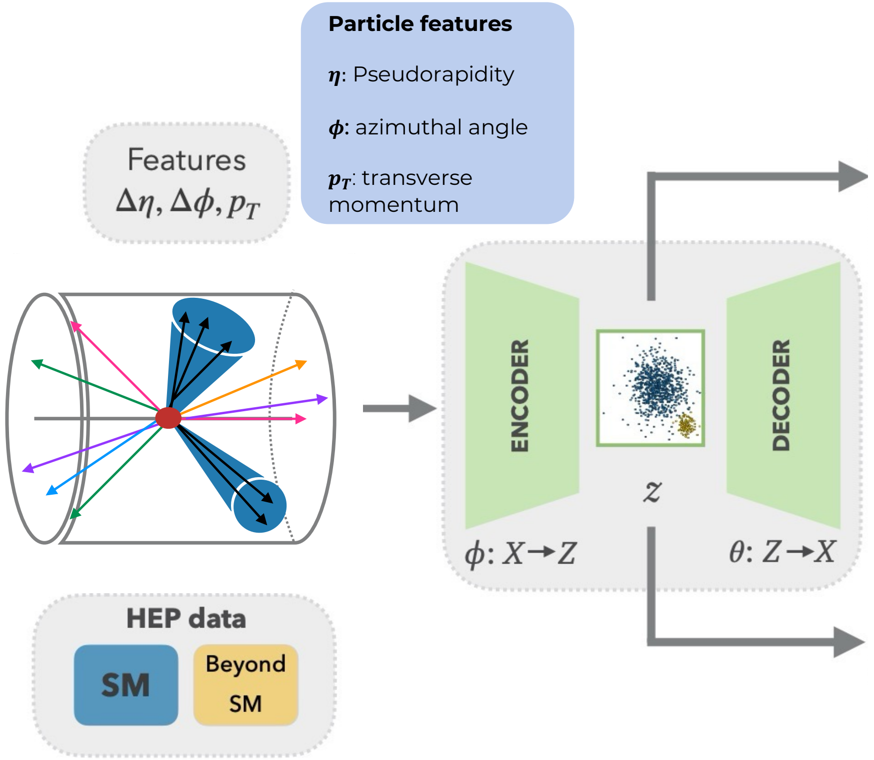



**Background:** QCD dijet events. 600 features per event $\longrightarrow$ Too many for current hardware.

**BSM anomalies:** Graviton **&** New Scalar Boson

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

# Anomaly detection with quantum machine learning

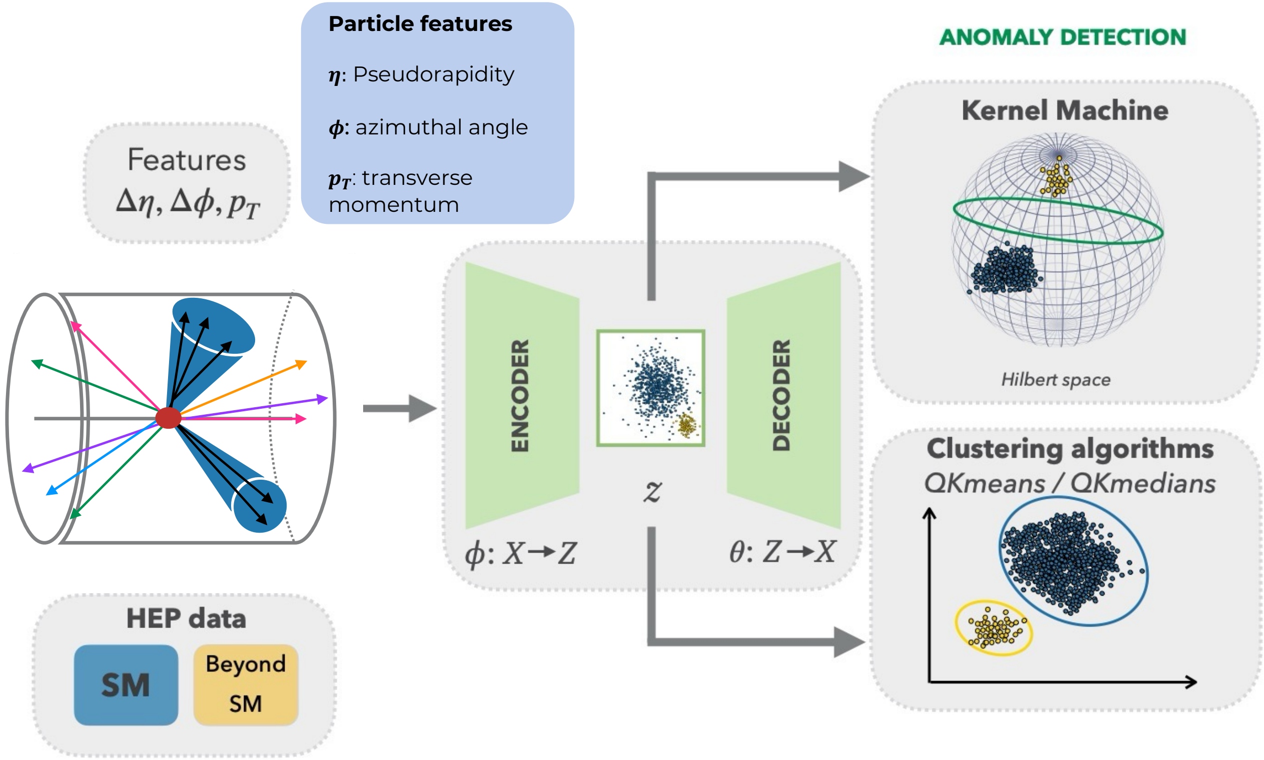



**Background:** QCD dijet events. 600 features per event $\longrightarrow$ Too many for current hardware.

**BSM anomalies:** Graviton & New Scalar Boson

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

# Anomaly detection with quantum machine learning

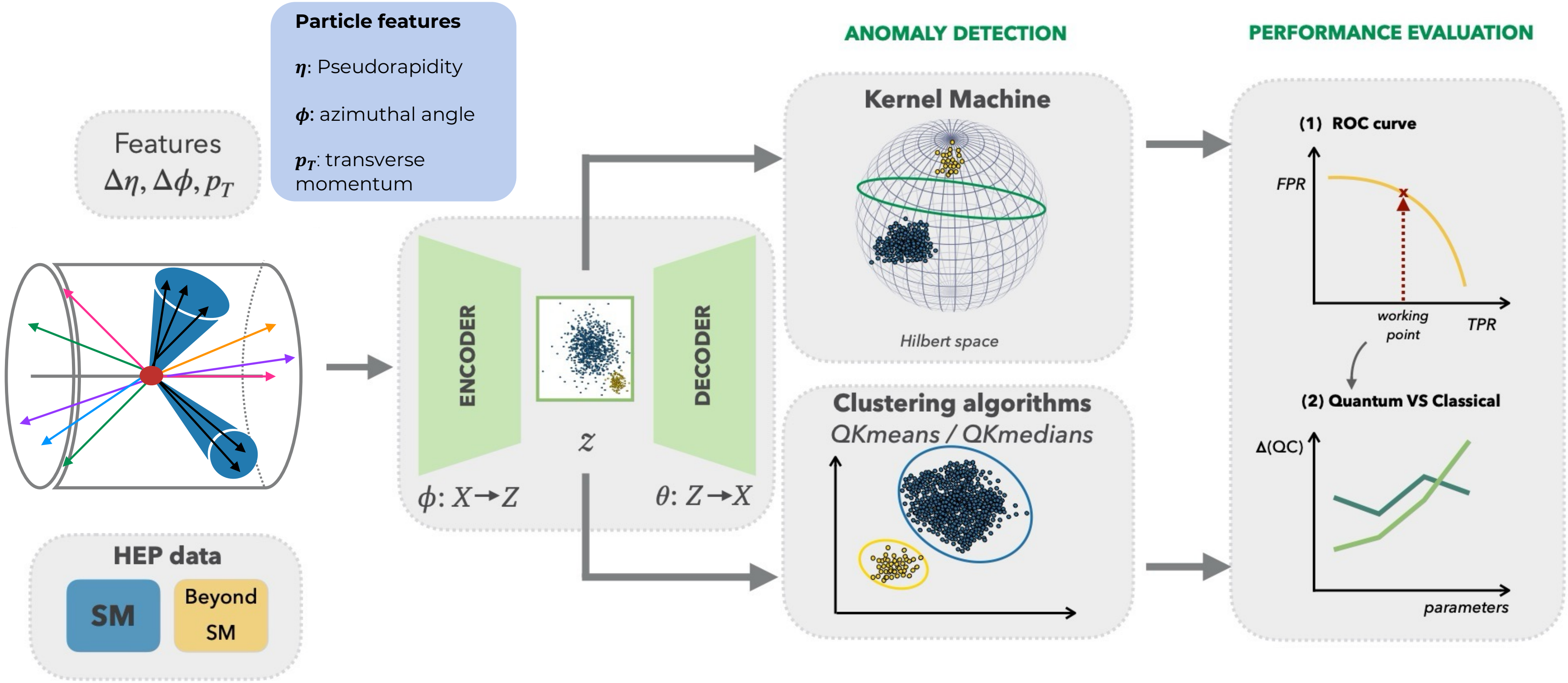



**Background:** QCD dijet events. 600 features per event $\longrightarrow$ Too many for current hardware.

**BSM anomalies:** Graviton **&** New Scalar Boson

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

# Anomaly detection with quantum machine learning

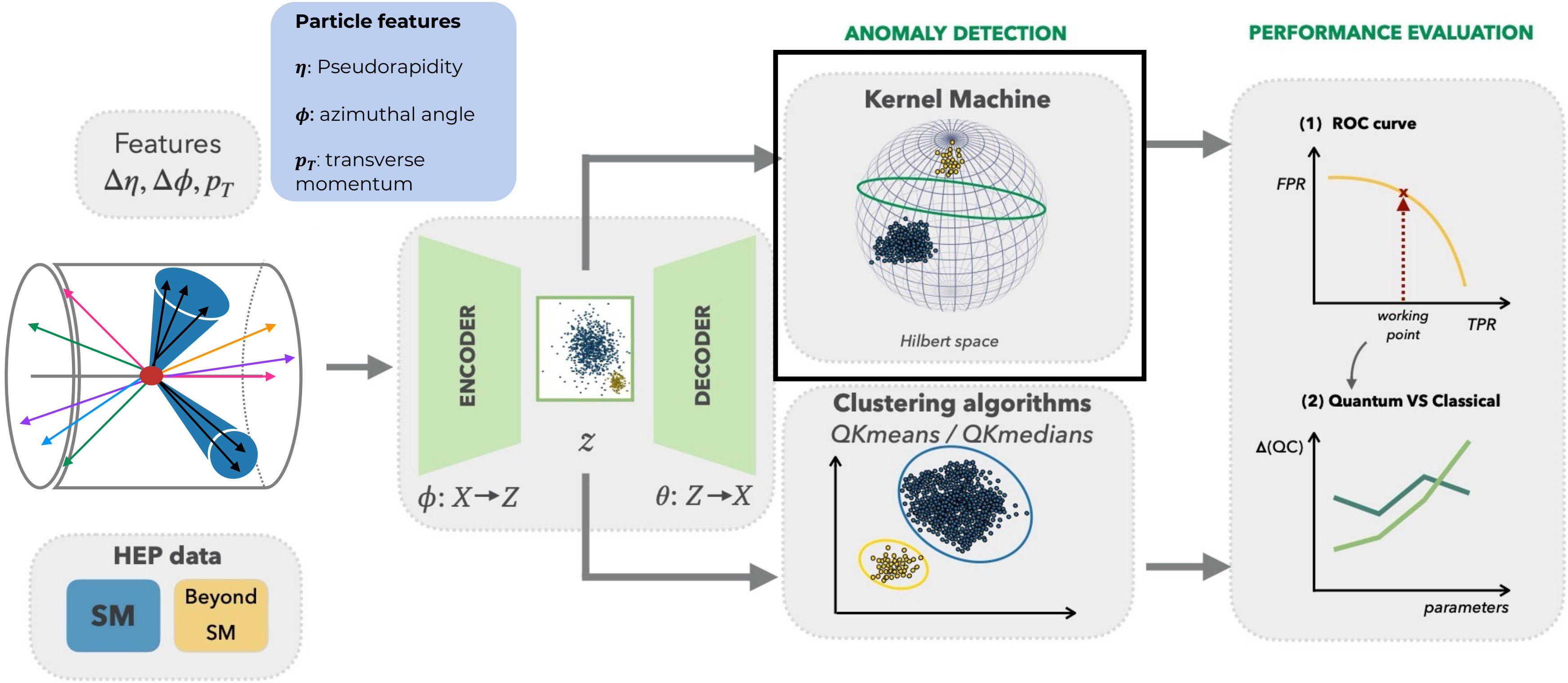



**Background:** QCD dijet events. 600 features per event $\longrightarrow$ Too many for current hardware.

**BSM anomalies:** Graviton & New Scalar Boson

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

# Anomaly detection with quantum machine learning

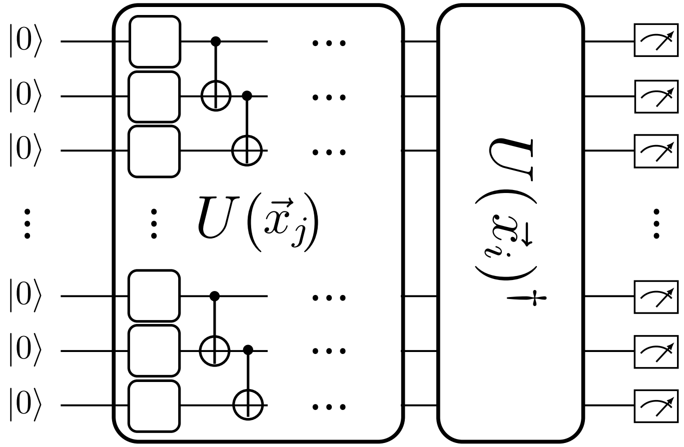



**Background:** QCD dijet events. 600 features per event $\longrightarrow$ Too many for current hardware.

**BSM anomalies:** Graviton **&** New Scalar Boson

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

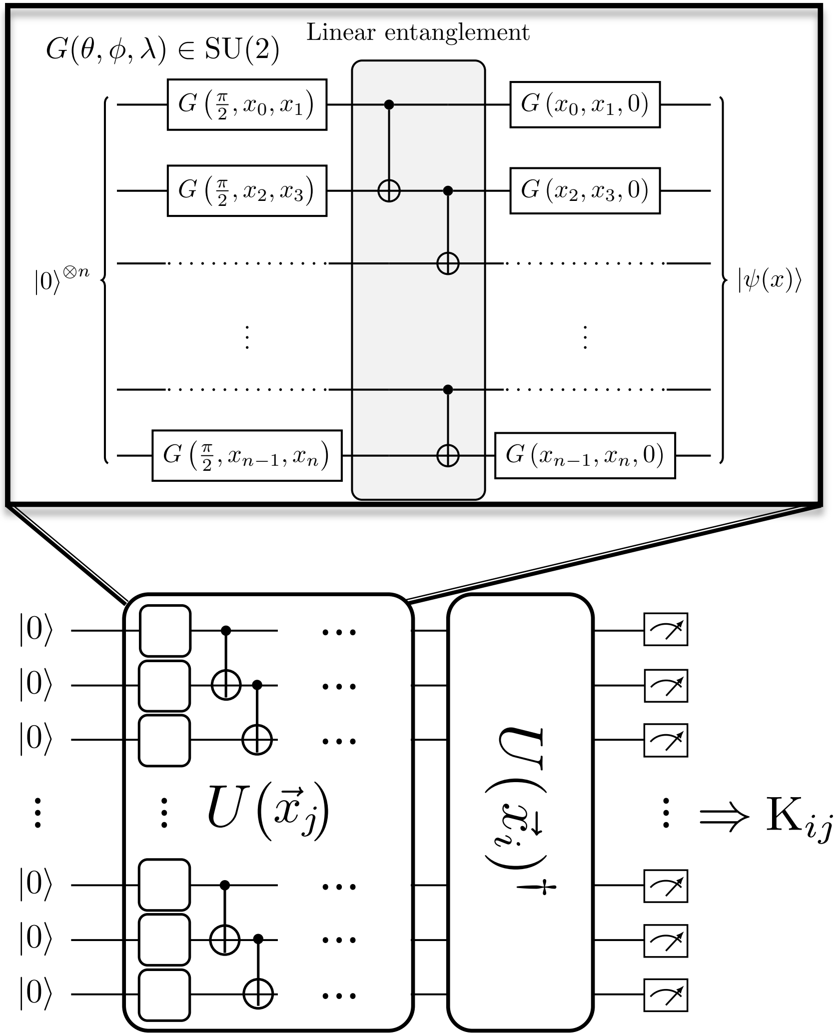
# Unsupervised quantum kernel machine




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

# Unsupervised quantum kernel machine

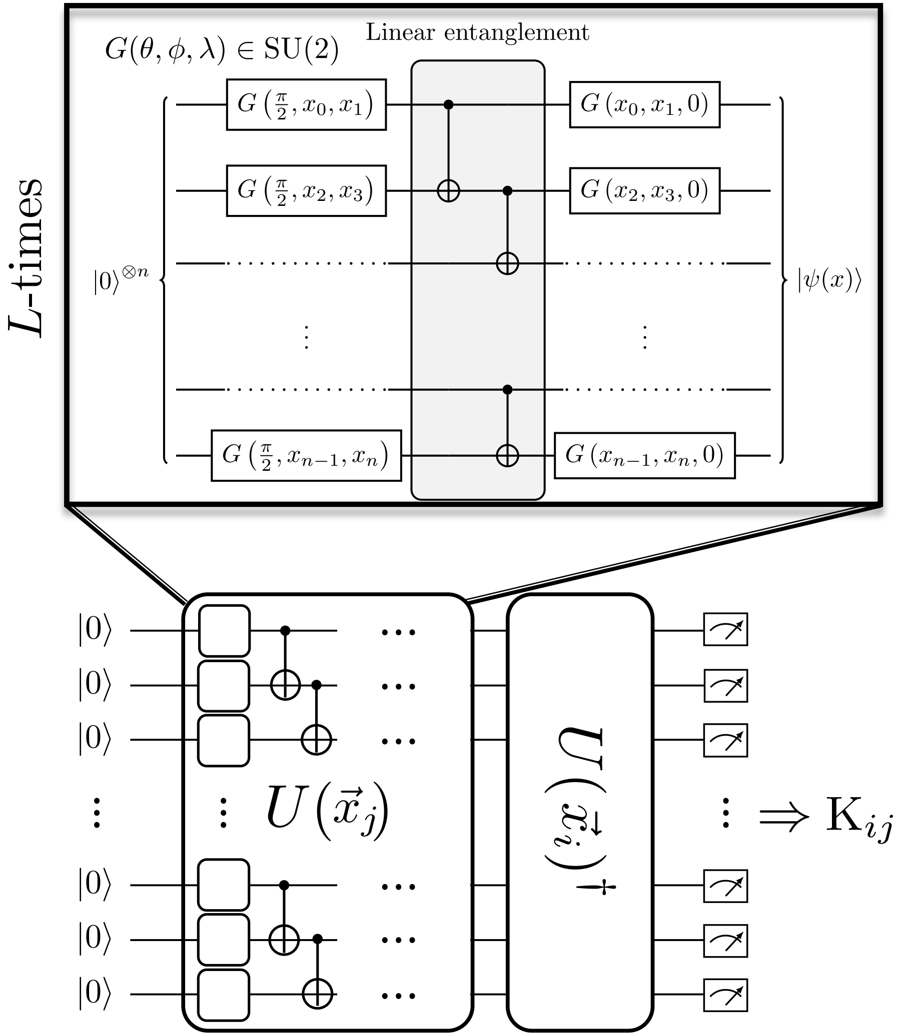
## Designed data encoding circuit




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

# Unsupervised quantum kernel machine

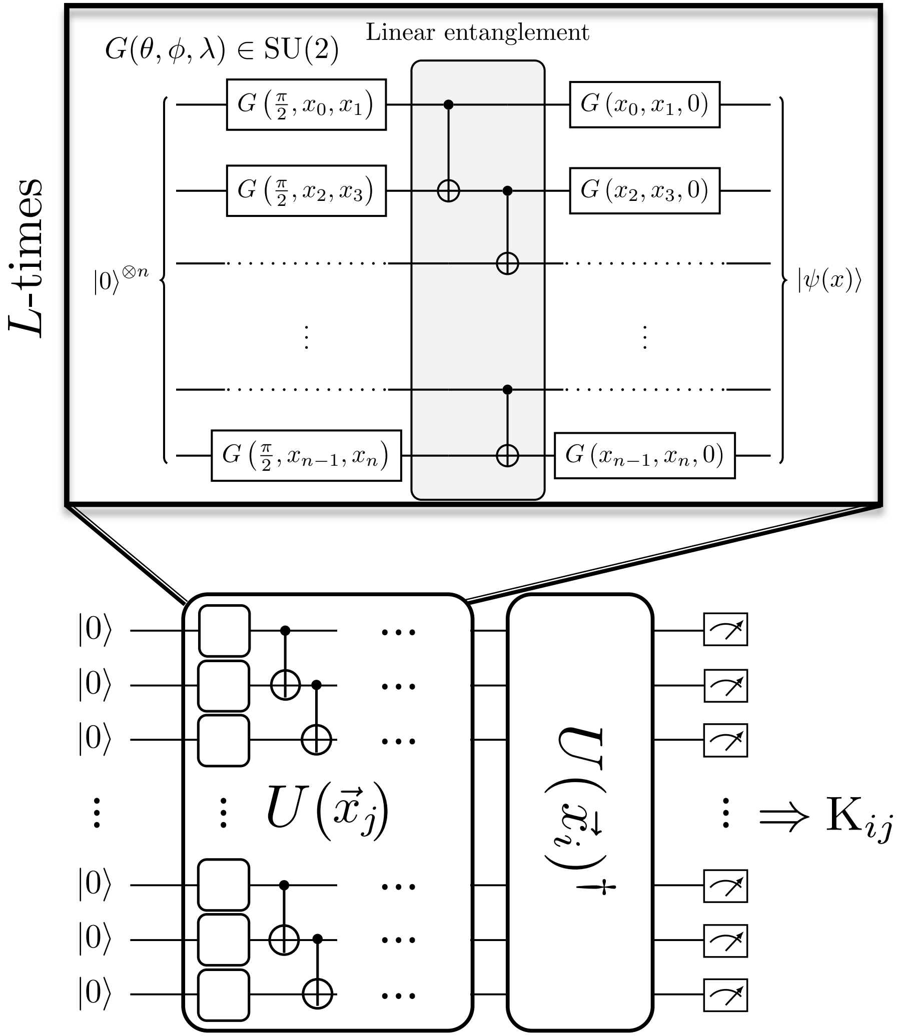
## Designed data encoding circuit




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

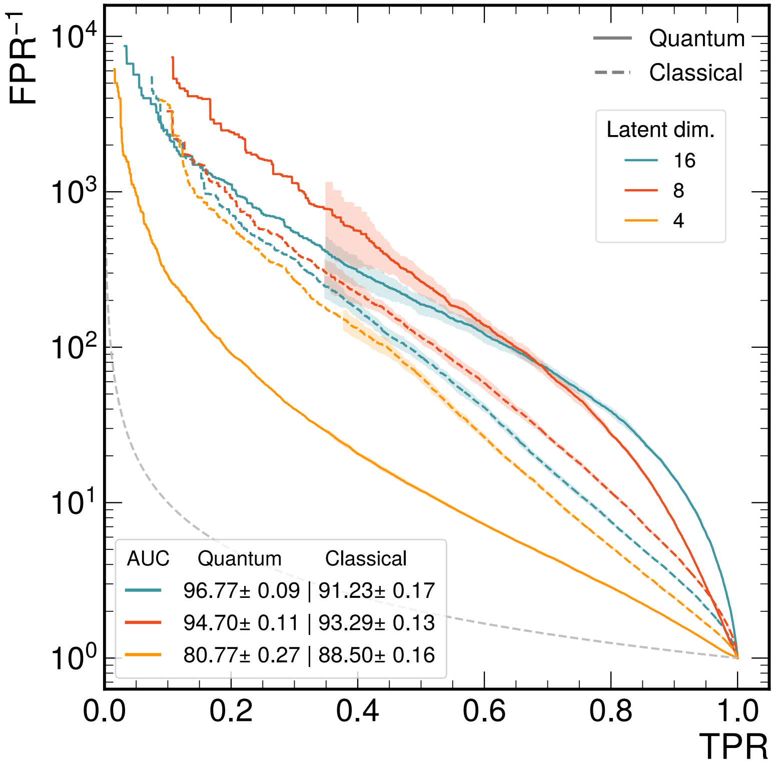
# Unsupervised quantum kernel machine

## Designed data encoding circuit



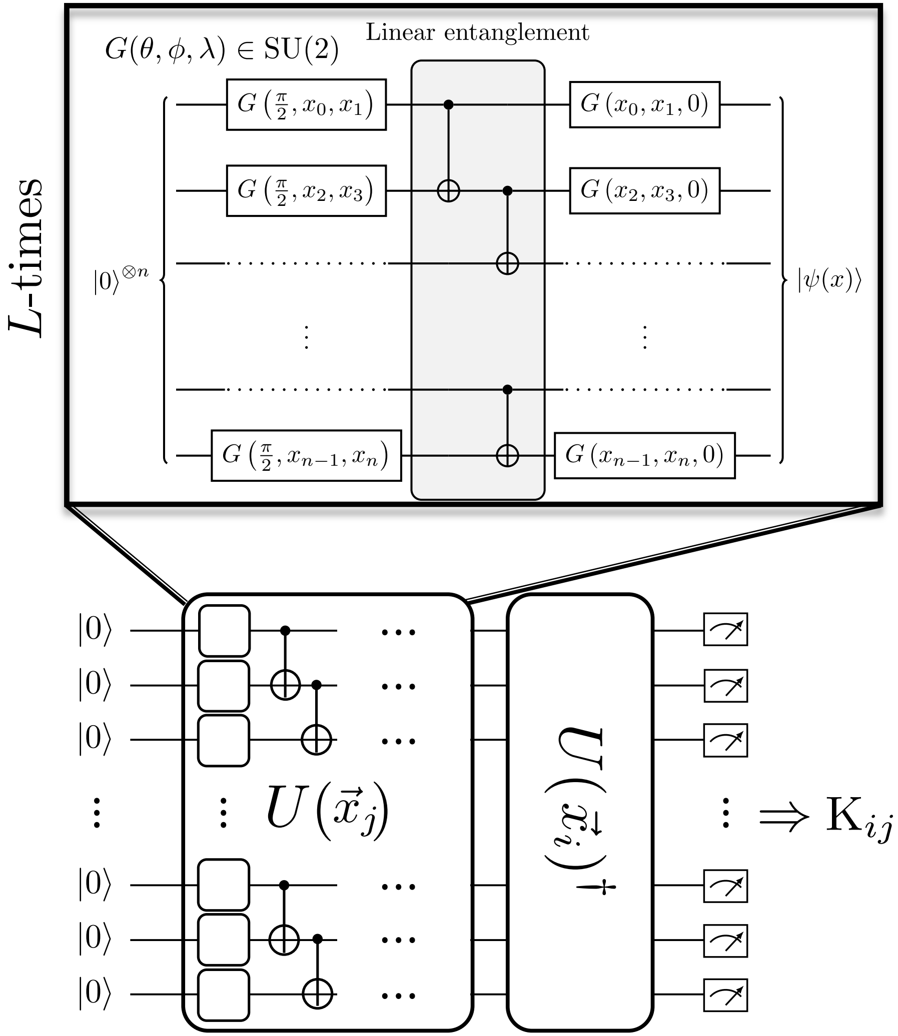

$$\Rightarrow \mathrm{K}_{ij} = |\langle 0|U^\dagger(\vec{x}_i)U(\vec{x}_j)|0\rangle|^2$$

## Different qubit number



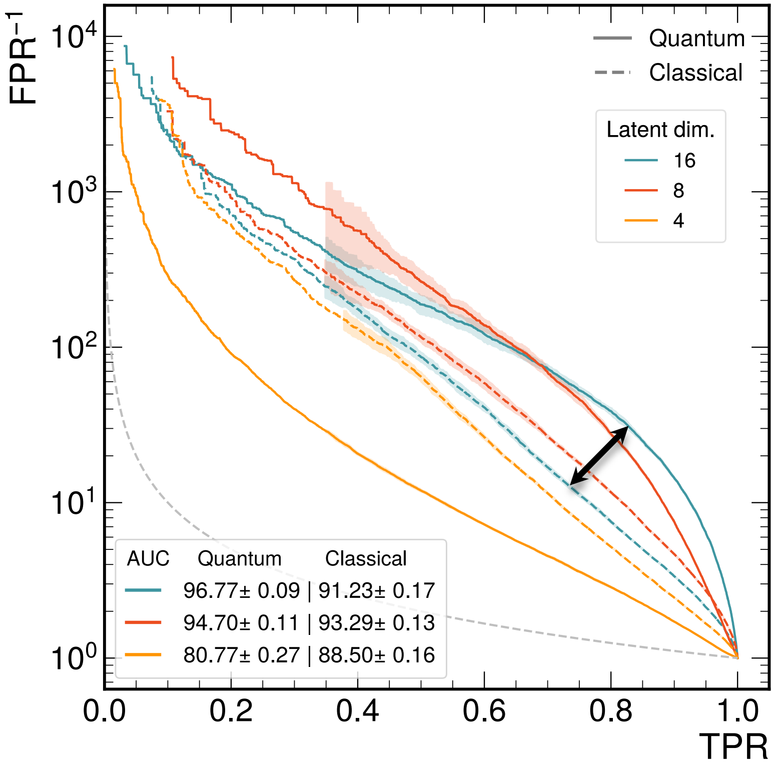
# Unsupervised quantum kernel machine

## Designed data encoding circuit



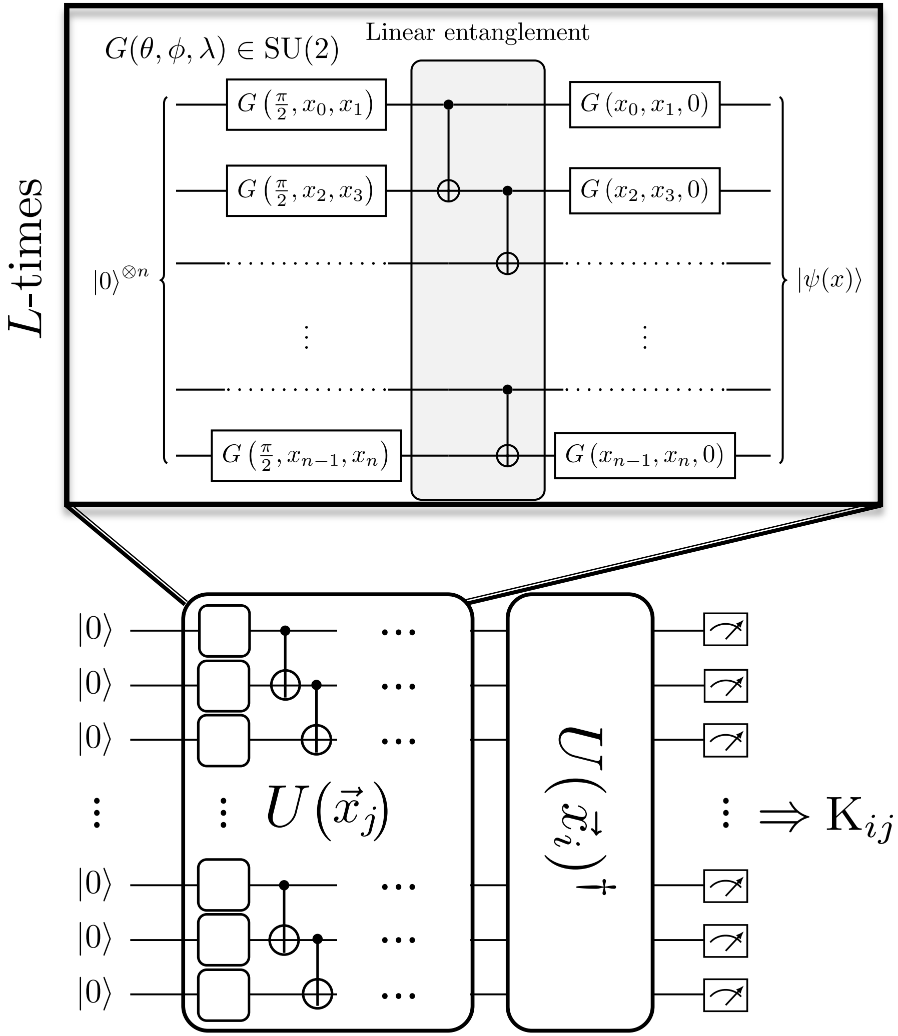

$$\Rightarrow \mathrm{K}_{ij} = |\langle 0|U^\dagger(\vec{x}_i)U(\vec{x}_j)|0\rangle|^2$$

## Different qubit number



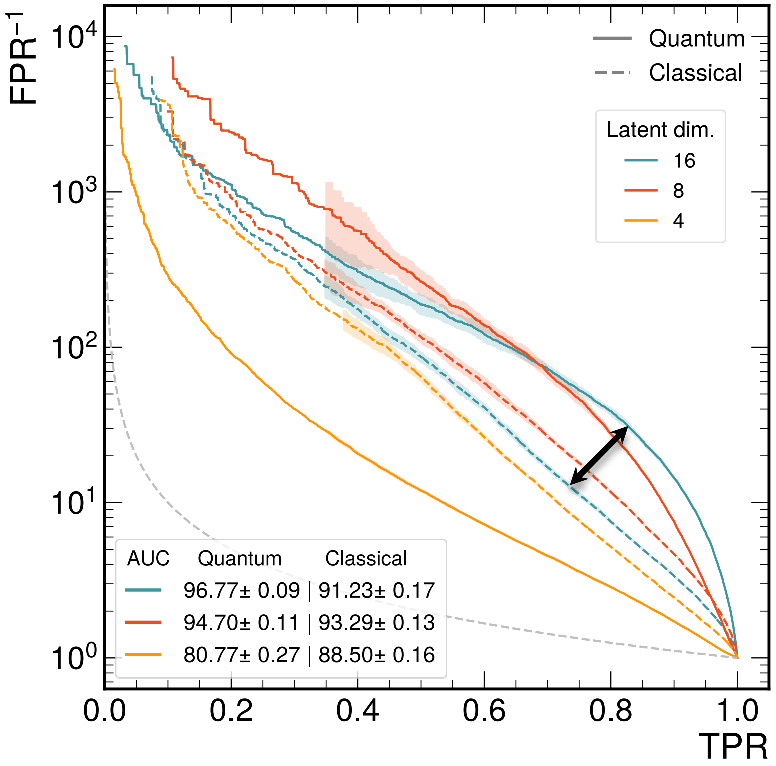
# Unsupervised quantum kernel machine

## Designed data encoding circuit

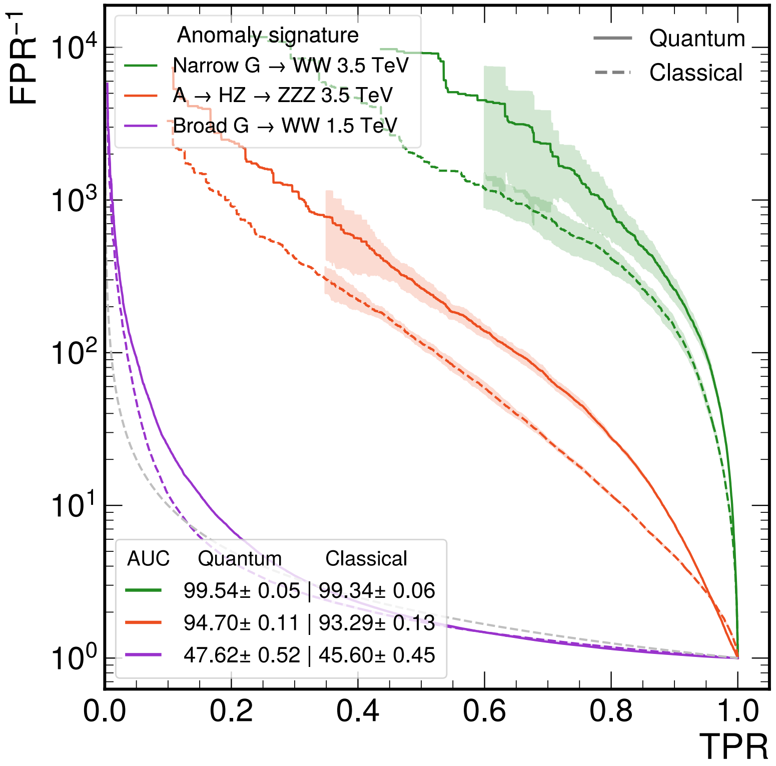



$$\Rightarrow \mathrm{K}_{ij} = |\langle 0|U^\dagger(\vec{x}_i)U(\vec{x}_j)|0\rangle|^2$$

## Different qubit number

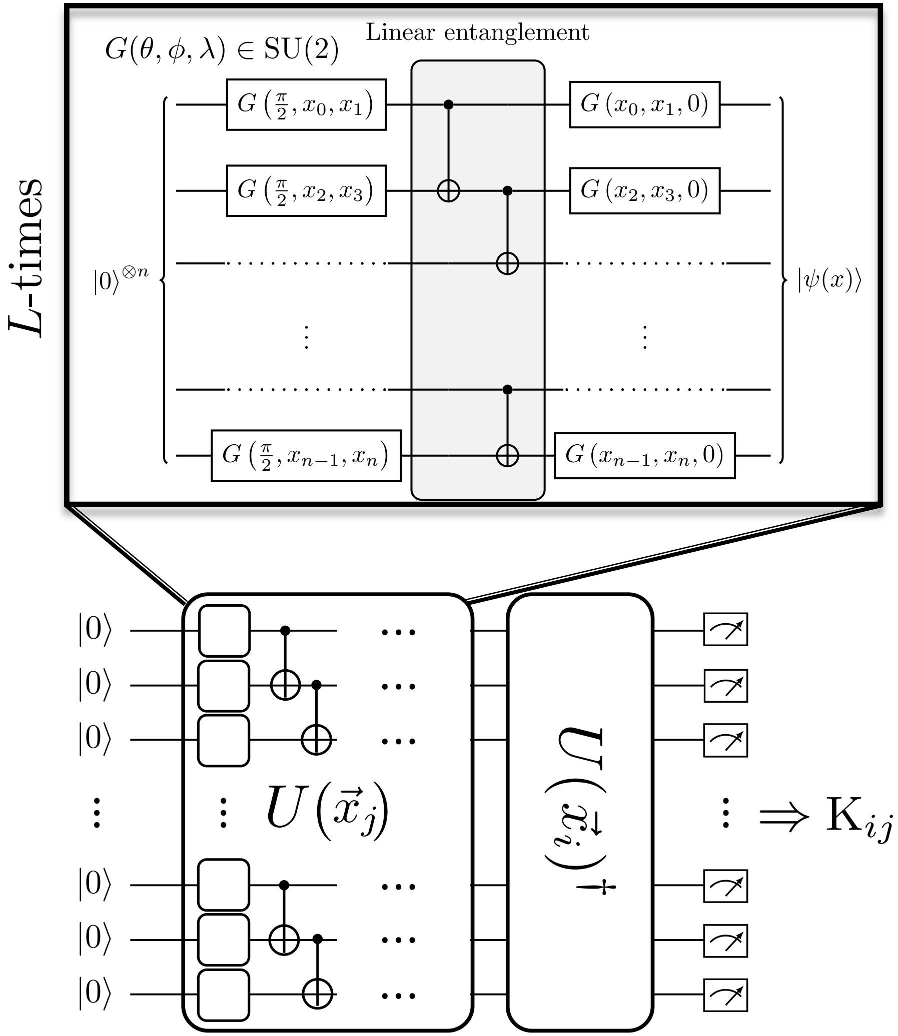



## Different BSM scenarios



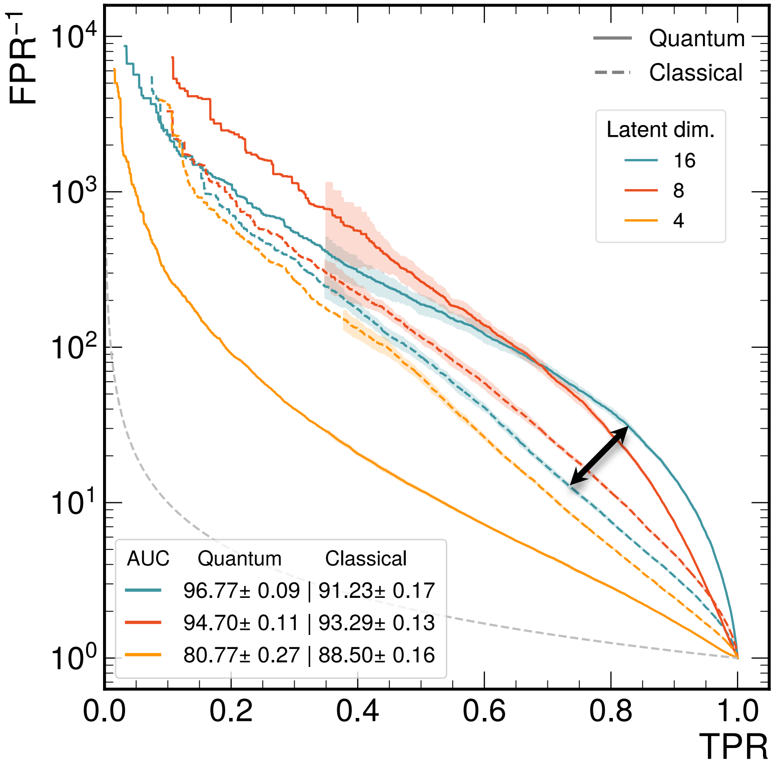
# Unsupervised quantum kernel machine

## Designed data encoding circuit

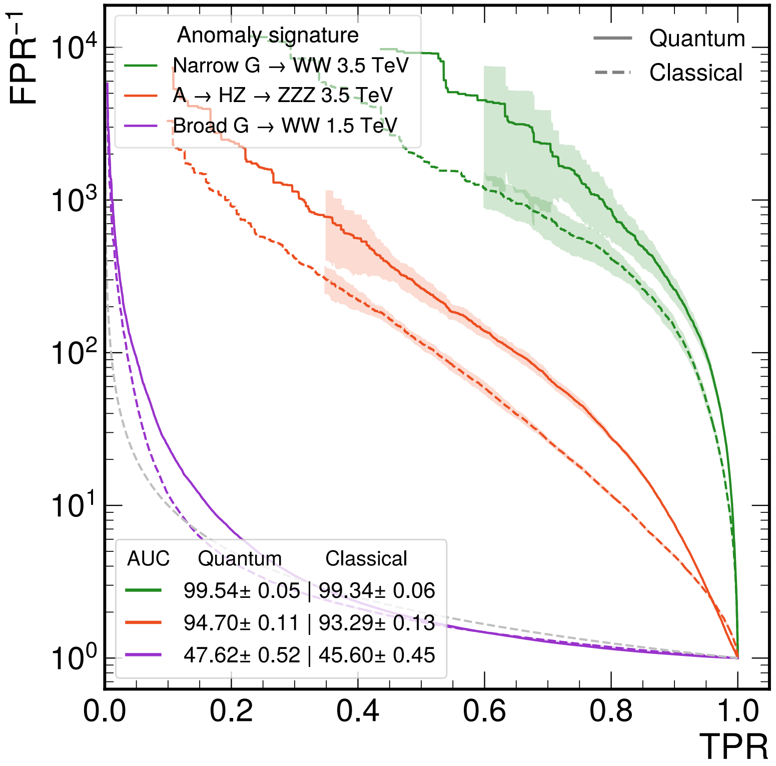



$$\Rightarrow \mathrm{K}_{ij} = |\langle 0|U^\dagger(\vec{x}_i)U(\vec{x}_j)|0\rangle|^2$$

## Different qubit number



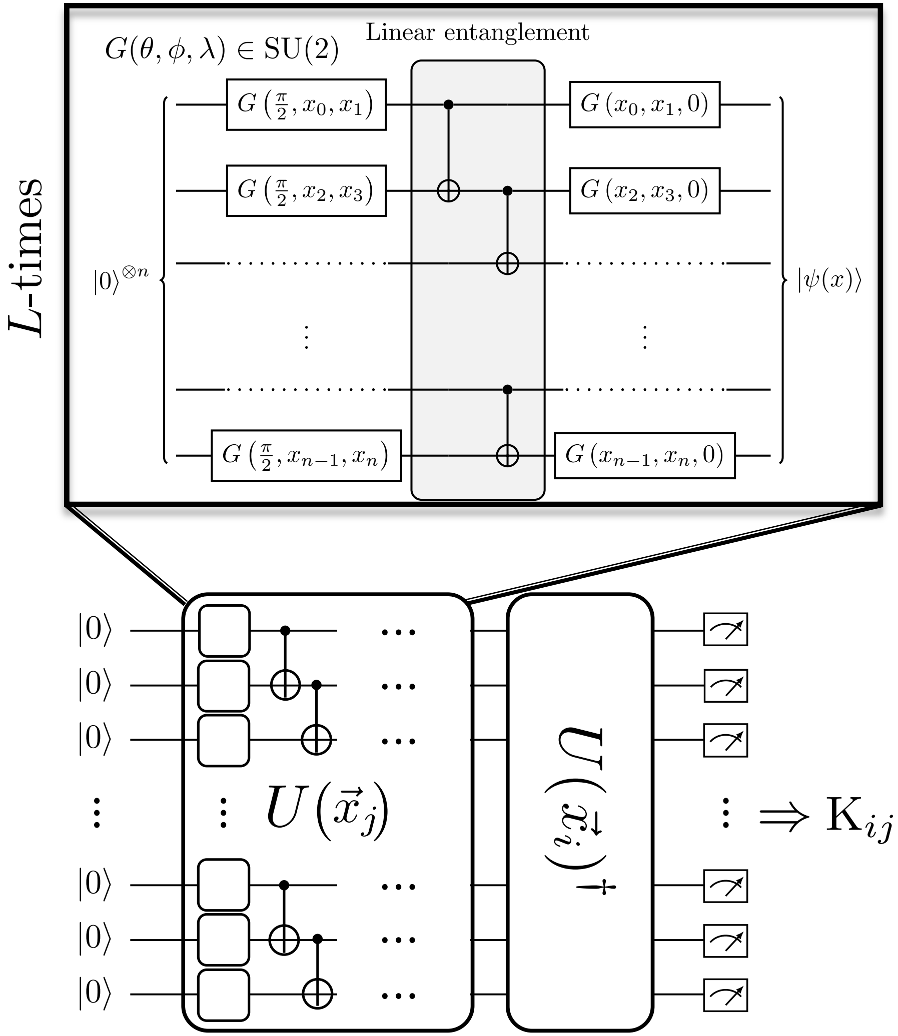

## Different BSM scenarios




**Are quantum effects utilized by the model?**

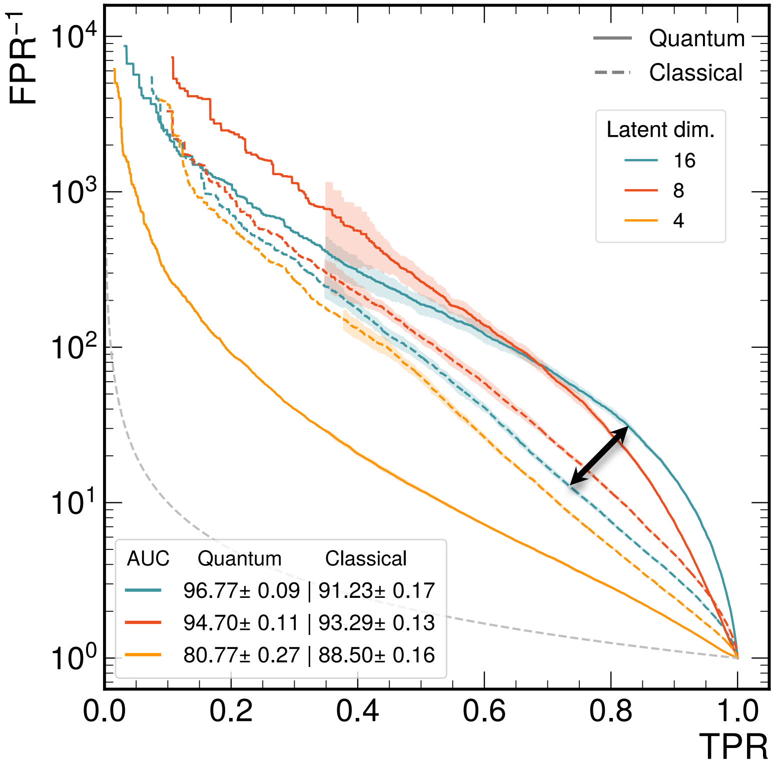
# Unsupervised quantum kernel machine

## Designed data encoding circuit

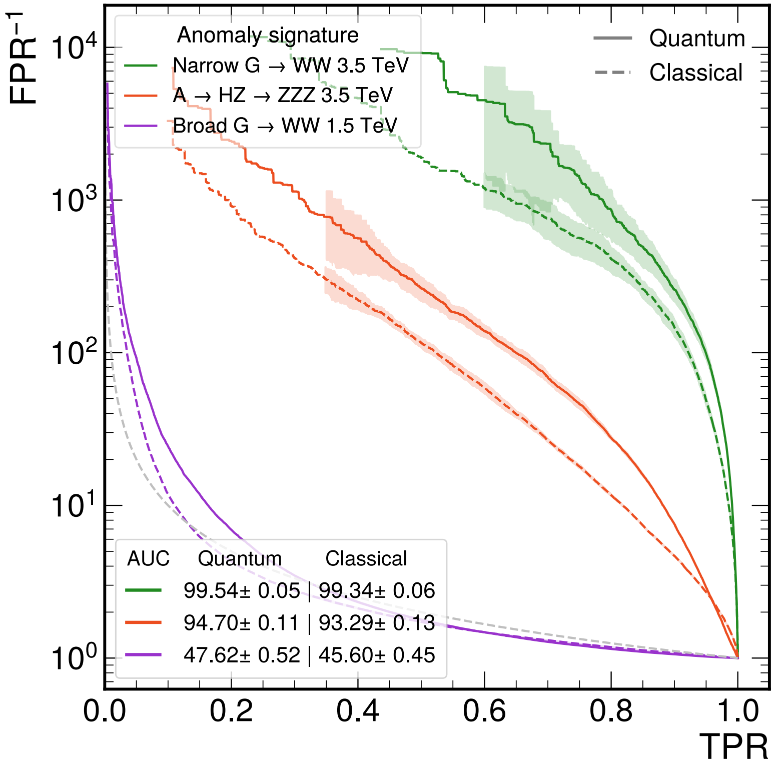



$$\Rightarrow \mathrm{K}_{ij} = |\langle 0|U^\dagger(\vec{x}_i)U(\vec{x}_j)|0\rangle|^2$$

## Different qubit number




## Different BSM scenarios



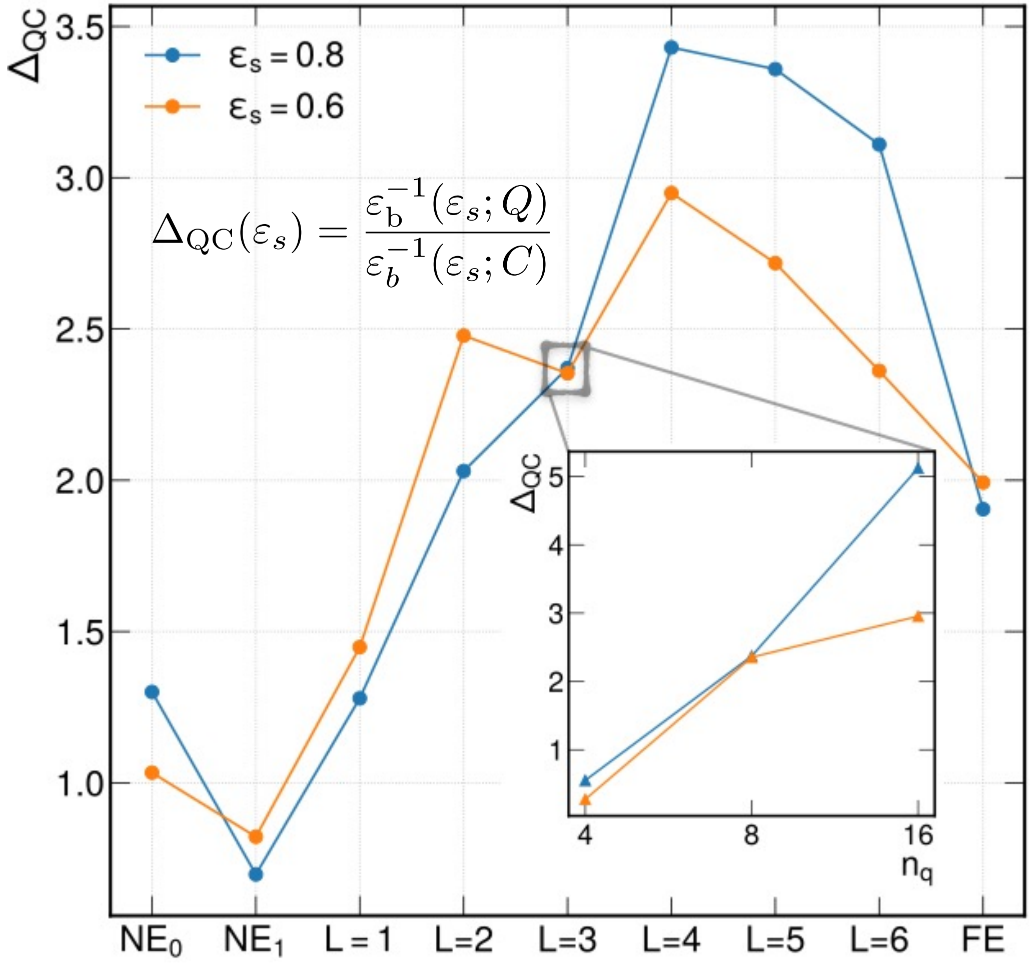

**Are quantum effects utilized by the model?**
**Yes!**

# Quantum circuit properties vs. performance

Importance of intrinsically quantum properties of the feature map.

# Quantum circuit properties vs. performance

Importance of intrinsically quantum properties of the feature map.




More expressive

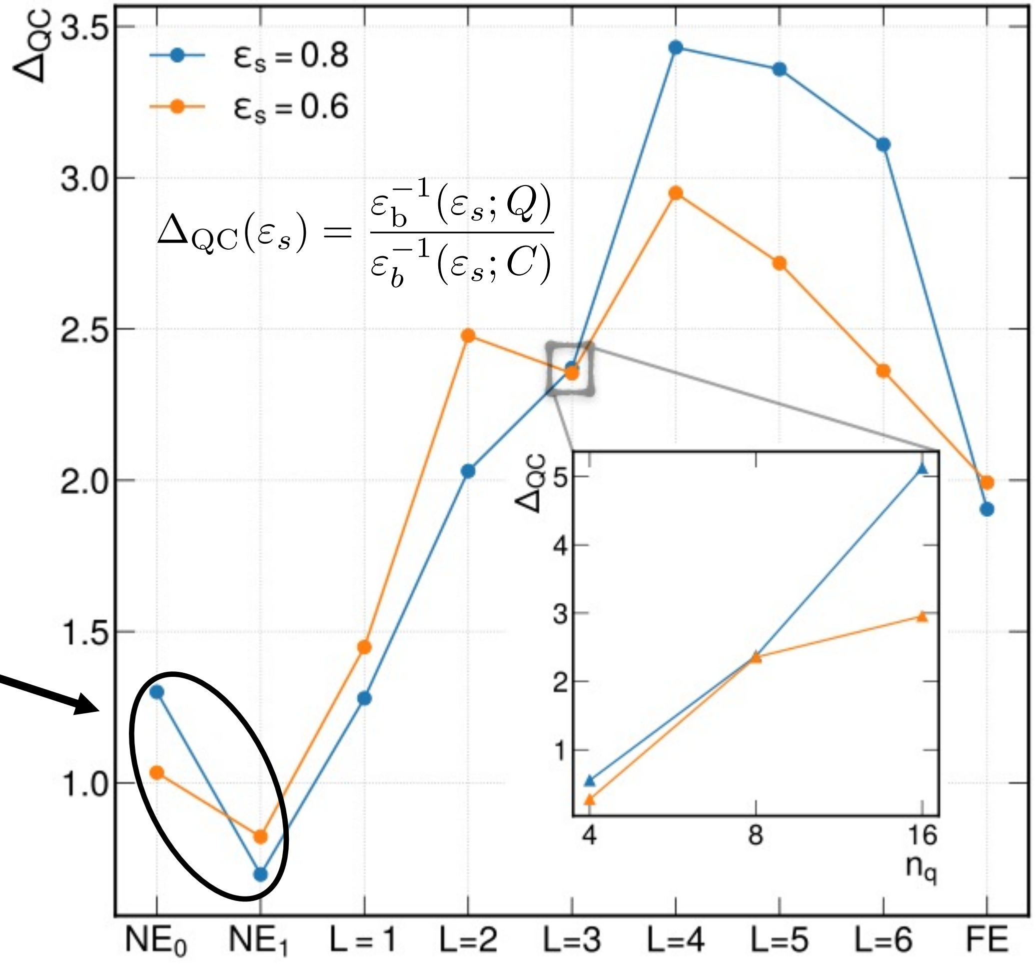
$\varepsilon_s$ Signal efficiency = True Positive Rate (TPR)

$\varepsilon_b^{-1}$ Background Rejection = Inverse of False Positive Rate

# Quantum circuit properties vs. performance

Importance of intrinsically quantum properties of the feature map.

No entanglement




More expressive

$\varepsilon_s$ Signal efficiency = True Positive Rate (TPR)

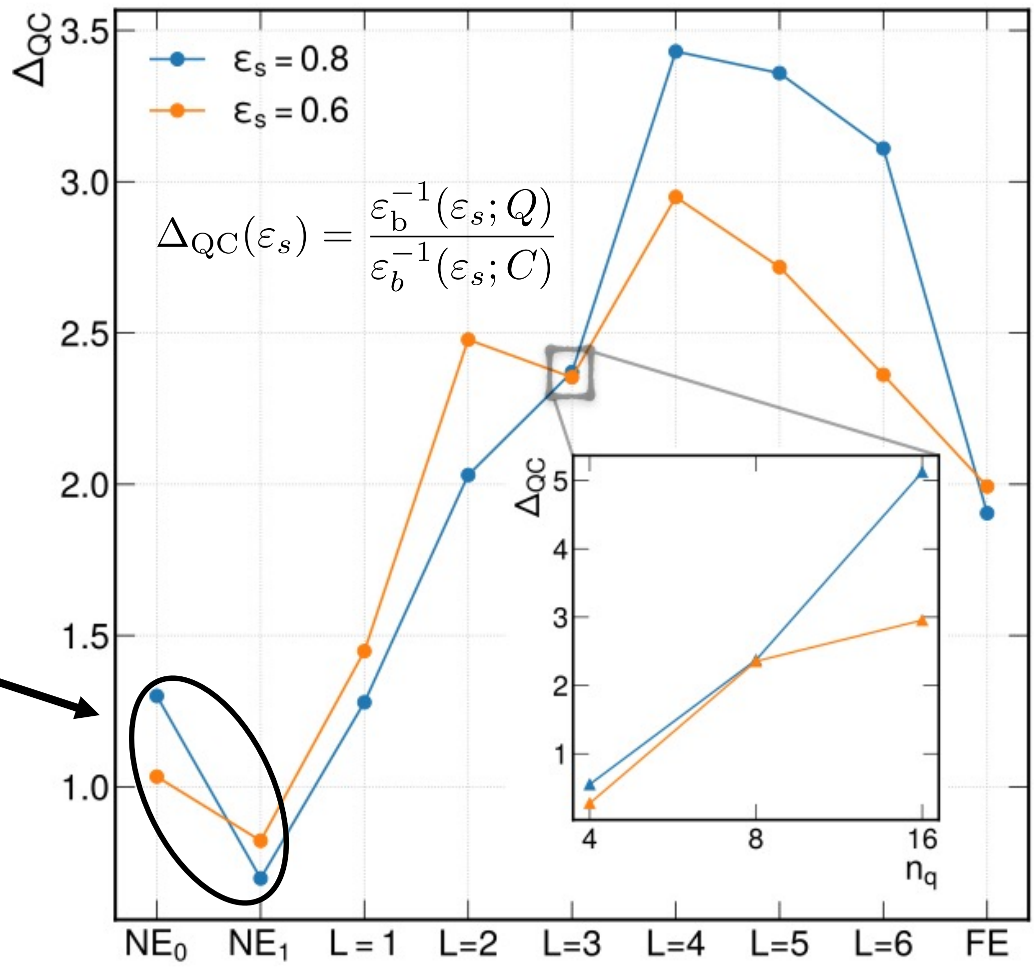
$\varepsilon_b^{-1}$ Background Rejection = Inverse of False Positive Rate

# Quantum circuit properties vs. performance

Importance of intrinsically quantum properties of the feature map.

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

No entanglement




More expressive

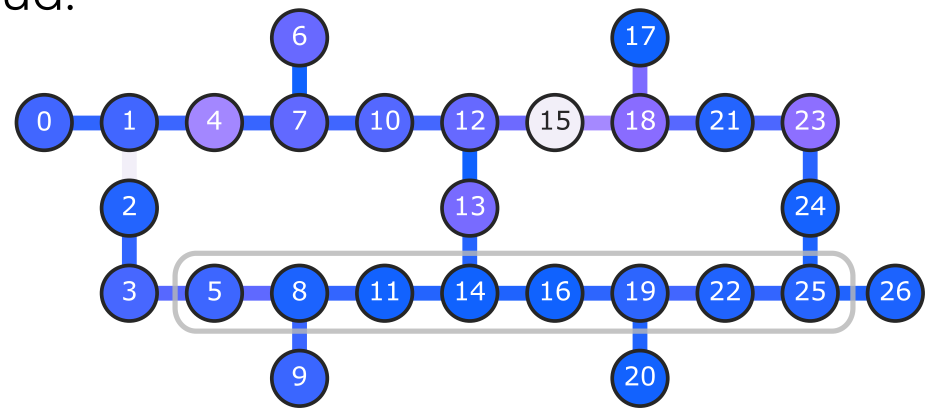
$\varepsilon_s$ Signal efficiency = True Positive Rate (TPR)

$\varepsilon_b^{-1}$ Background Rejection = Inverse of False Positive Rate

# Quantum hardware runs

Submit jobs to a real machine (ibm_toronto) using IBMQ cloud.

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

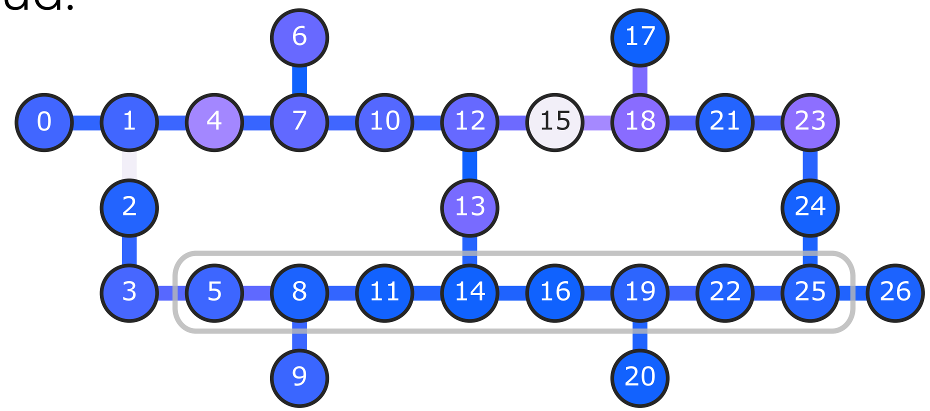



Superconducting qubits connectivity topology

# Quantum hardware runs

Submit jobs to a real machine (ibm_toronto) using IBMQ cloud.

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



Superconducting qubits connectivity topology

| 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: $1/2^{\mathrm{n_q}} \approx 0.39 \times 10^{-2}$

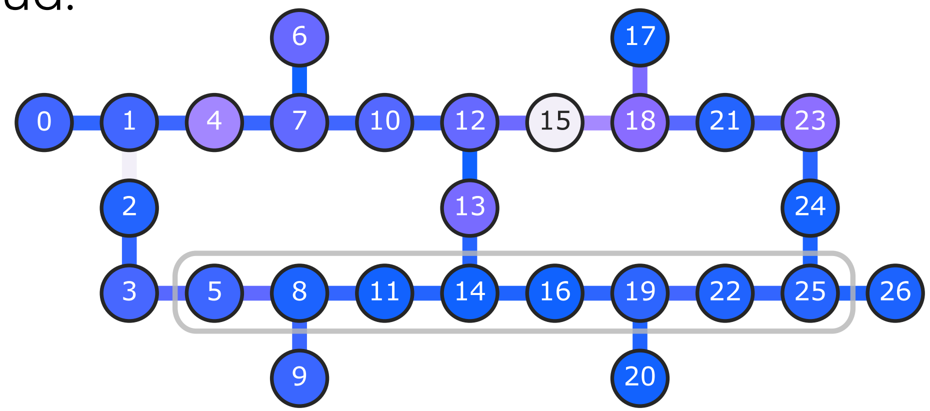
$$\langle \mathrm{tr}\rho^2 \rangle = \langle K(x_i, x_i) \rangle$$

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

# Quantum hardware runs

Submit jobs to a real machine (ibm_toronto) using IBMQ cloud.

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




Superconducting qubits connectivity topology

| 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: $1/2^{n_q} \approx 0.39 \times 10^{-2}$

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

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

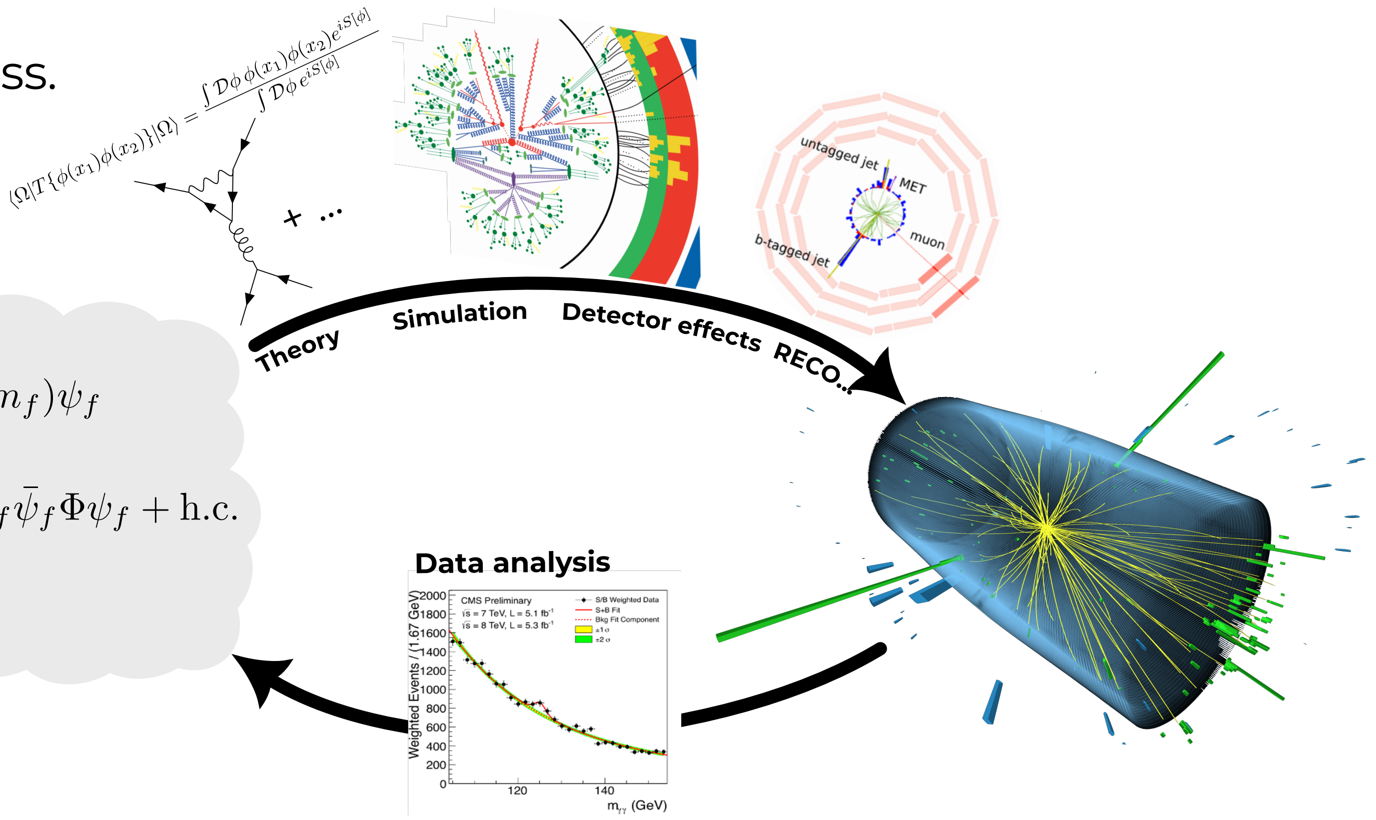
**Proposed data encoding circuit realistic and suitable for current devices**

# Summary

QML **+** HEP : very promising results and interplay (sandbox).

- Utilize data generated by a quantum process.
- New inductive biases?

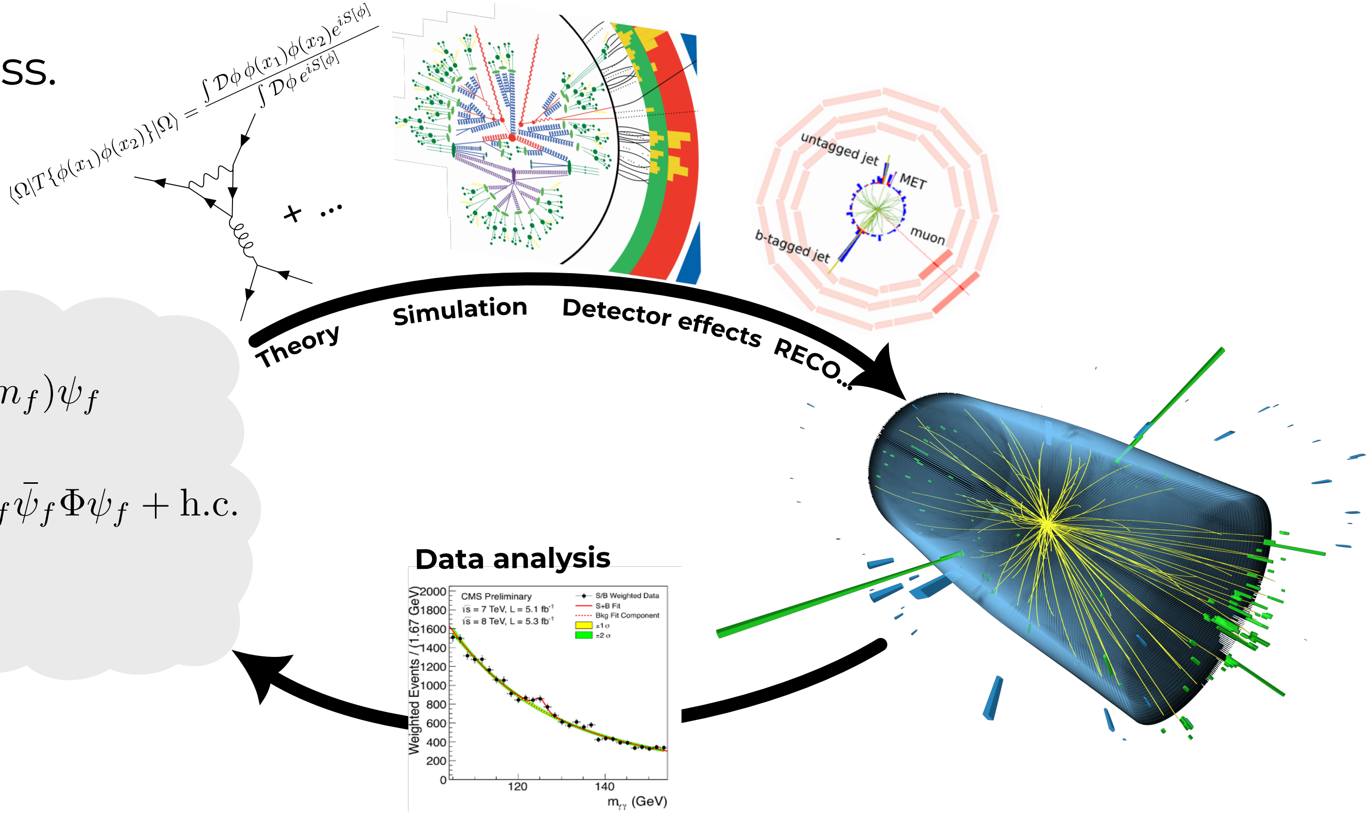
$$\mathcal{L} = -\frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} + \sum_f \bar{\psi}_f(i\gamma^\mu D_\mu - m_f)\psi_f$$
$$+ (D_\mu\Phi)^\dagger(D^\mu\Phi) - V(\Phi) - \sum_f y_f\bar{\psi}_f\Phi\psi_f + \text{h.c.}$$
$$+\,\mathbf{BSM?}$$



# Summary

QML **+** HEP : very promising results and interplay (sandbox).

- Utilize data generated by a quantum process.
- New inductive biases?




$$\mathcal{L} = -\frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} + \sum_f \bar{\psi}_f(i\gamma^\mu D_\mu - m_f)\psi_f$$
$$+ (D_\mu\Phi)^\dagger(D^\mu\Phi) - V(\Phi) - \sum_f y_f\bar{\psi}_f\Phi\psi_f + \text{h.c.}$$
$$+ \textbf{BSM?}$$

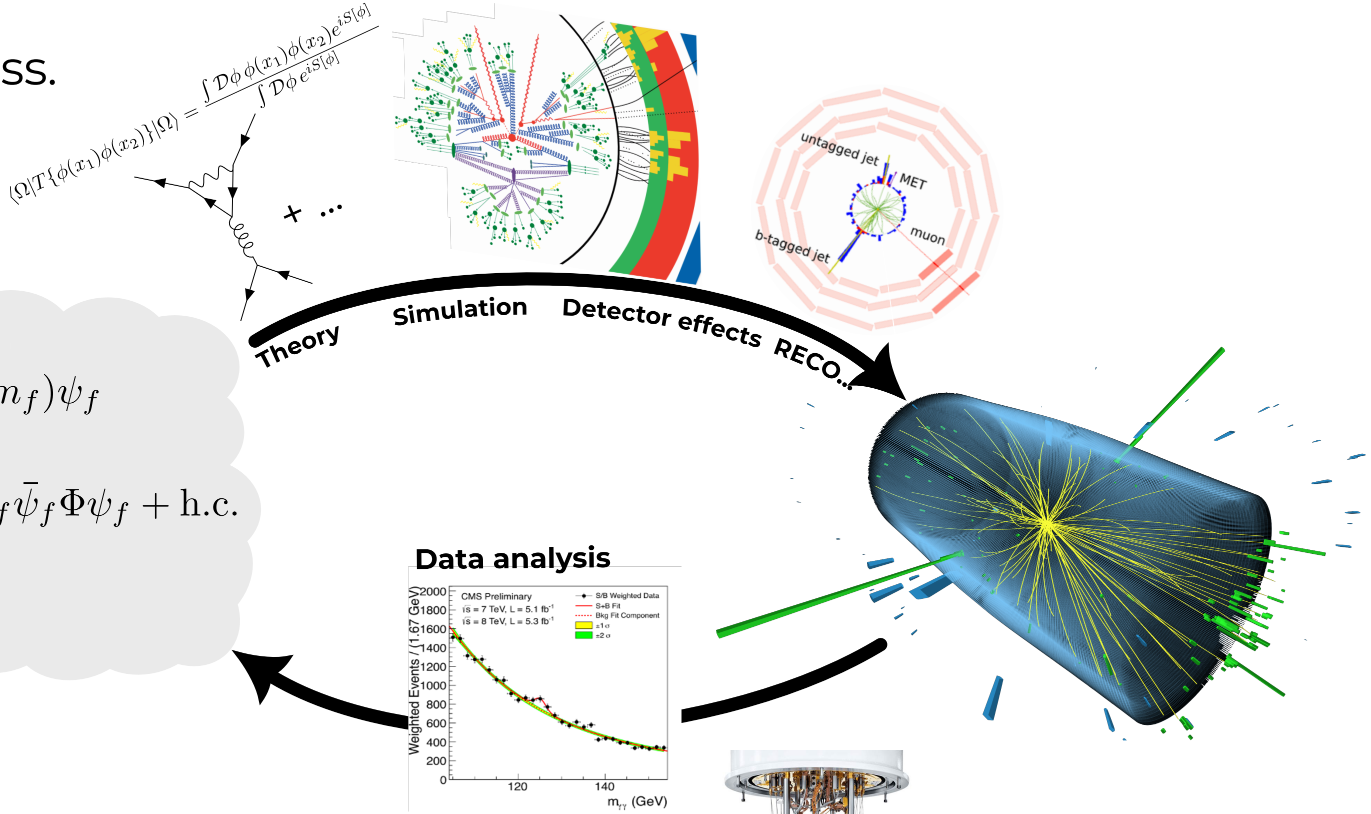
**For more details checkout:**

[K.A. Wozniak*, VB*, E. Puljak* et al., 2301.10780]
[VB, S. González-Castillo, et al., 2104.07692]
[J. Shuhmacher, L. Bogia, VB, et al., 2301.10787]
[T. Muser, E. Zapusek, VB, F. Reiter, 2309.14406]

# Summary

QML **+** HEP : very promising results and interplay (sandbox).

- Utilize data generated by a quantum process.
- New inductive biases?

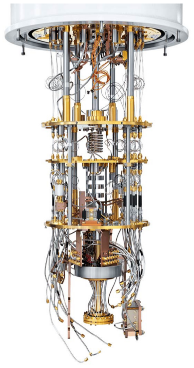


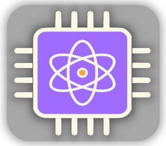

$$\mathcal{L} = -\frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} + \sum_f \bar{\psi}_f(i\gamma^\mu D_\mu - m_f)\psi_f$$
$$+ (D_\mu\Phi)^\dagger(D^\mu\Phi) - V(\Phi) - \sum_f y_f\bar{\psi}_f\Phi\psi_f + \text{h.c.}$$
$$+\, \mathbf{BSM?}$$

**For more details checkout:**

[K.A. Wozniak*, VB*, E. Puljak* et al., 2301.10780]
[VB, S. González-Castillo, et al., 2104.07692]
[J. Shuhmacher, L. Bogia, VB, et al., 2301.10787]
[T. Muser, E. Zapusek, VB, F. Reiter, 2309.14406]

# Any cool ideas?
# Questions?

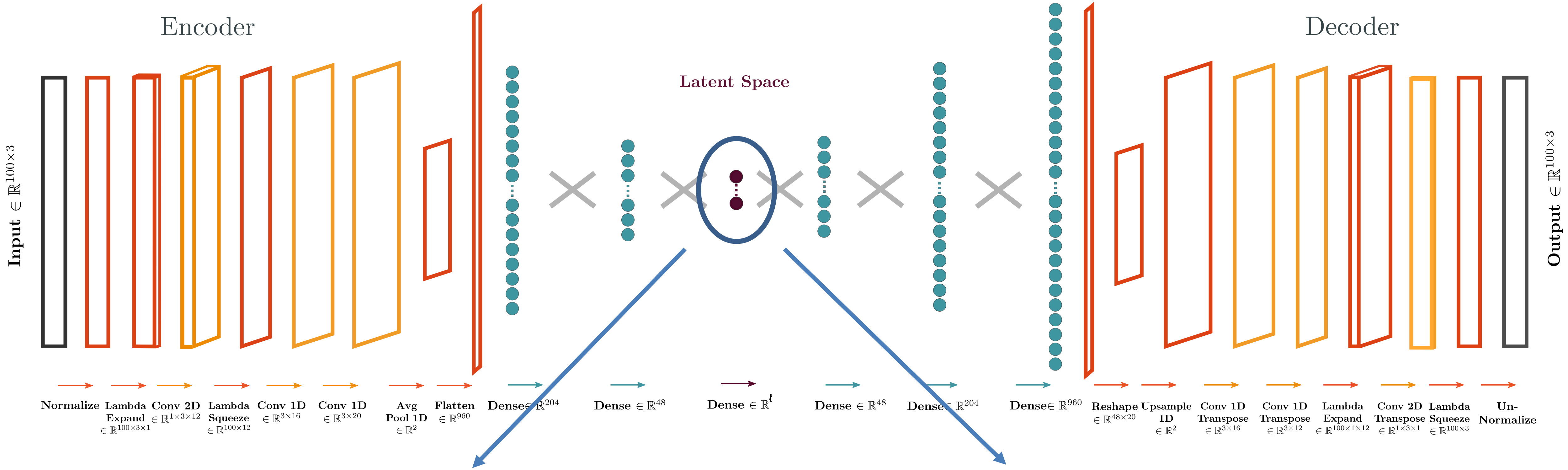




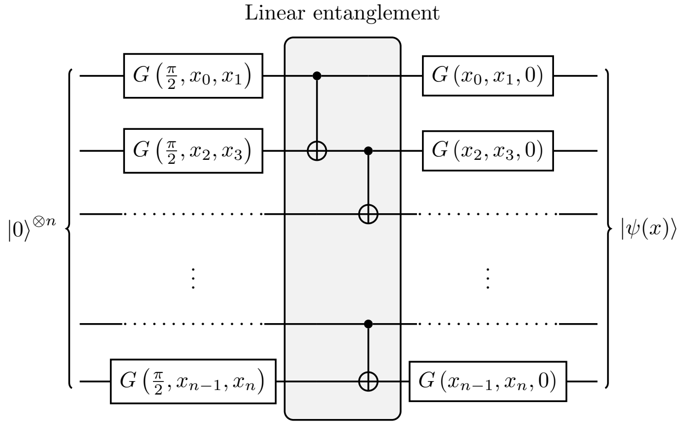
Credit: Rigetti Computing.

# Backup slides

# Convolutional autoencoder architecture



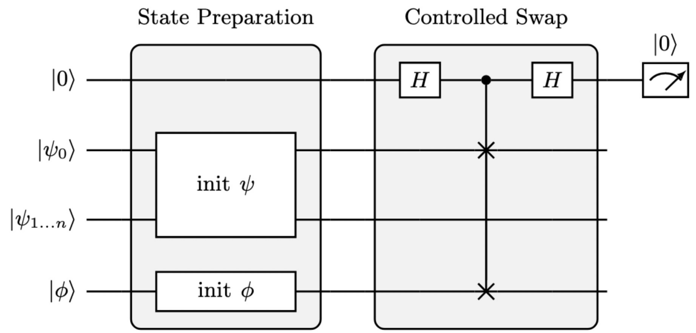

## Unsupervised kernel machine




Training size:

$\mathcal{O}(10^2) - \mathcal{O}(10^3)$

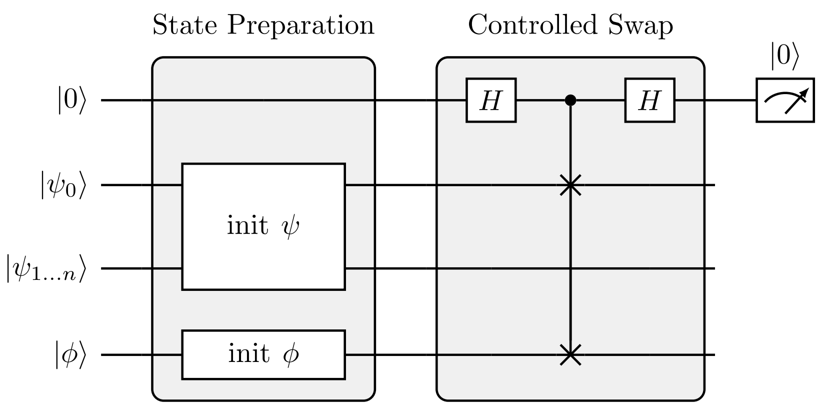
## Quantum clustering algorithms



# Quantum clustering for anomaly detection

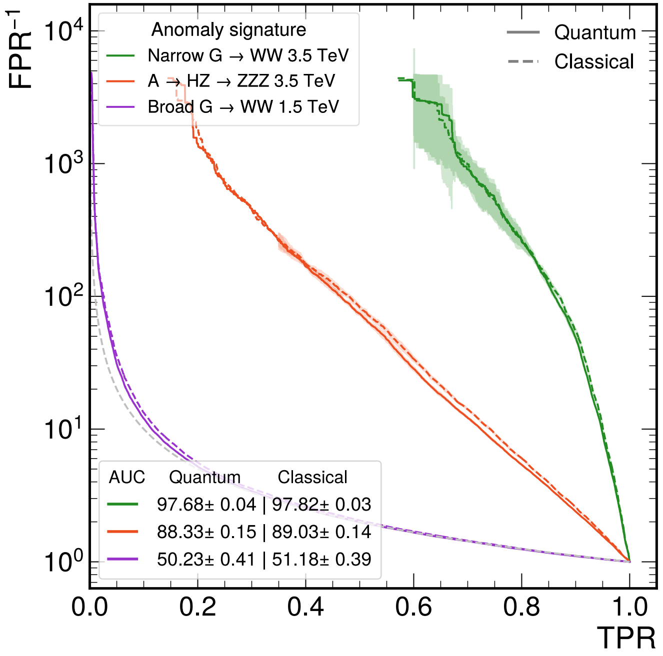
Construct clusters in the Hilbert space

Quantum distance calculation from clusters

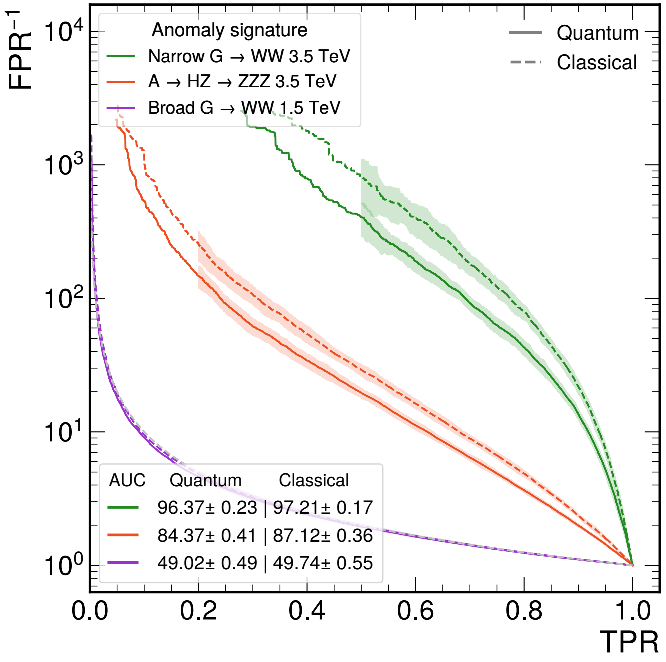



Minimise the distance with **quantum** (QK-means) or hybrid/**classical** (QK-medians) optimisation algorithms

**Quantum K-medians**




**Quantum K-means**

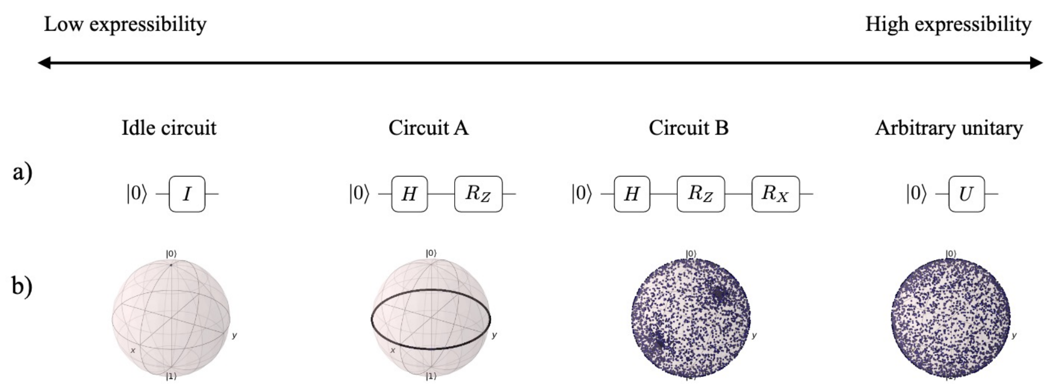



[K.A. Wozniak*, VB*, E. Puljak*, et al., arXiv: 2301.10780]

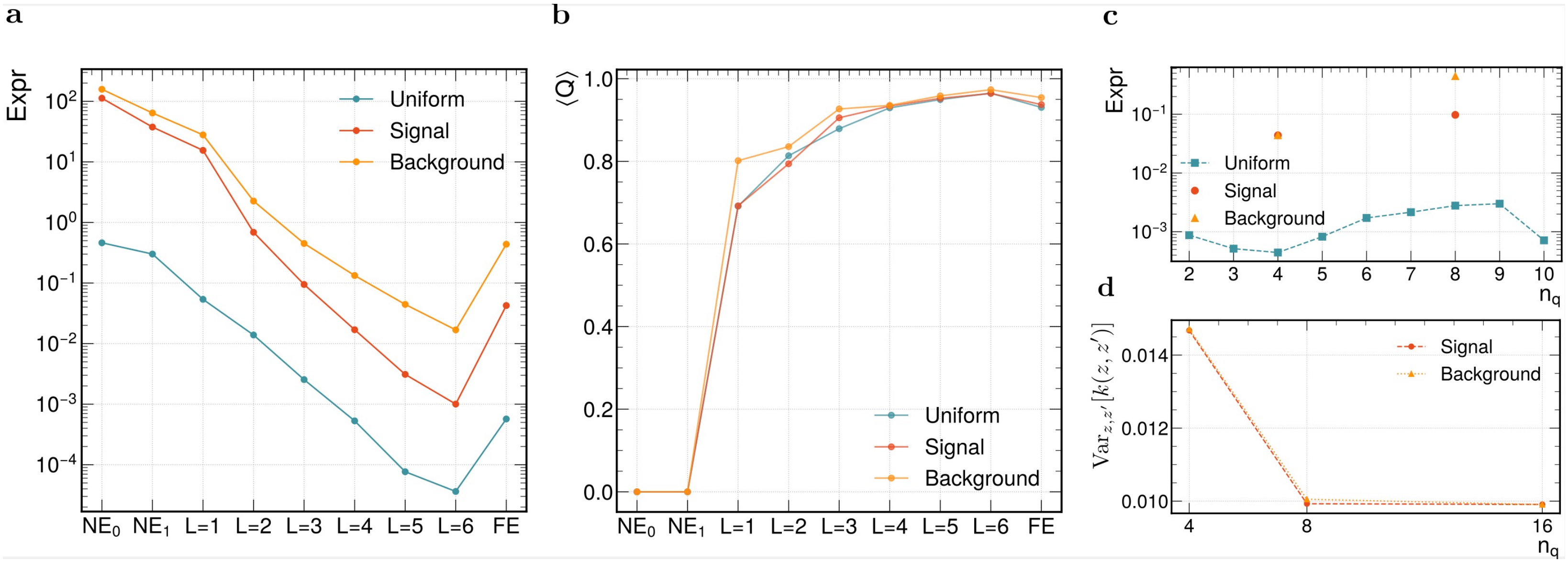
Quantum and classical anomaly detection has similar performance.

# Expressibility and entanglement capability

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

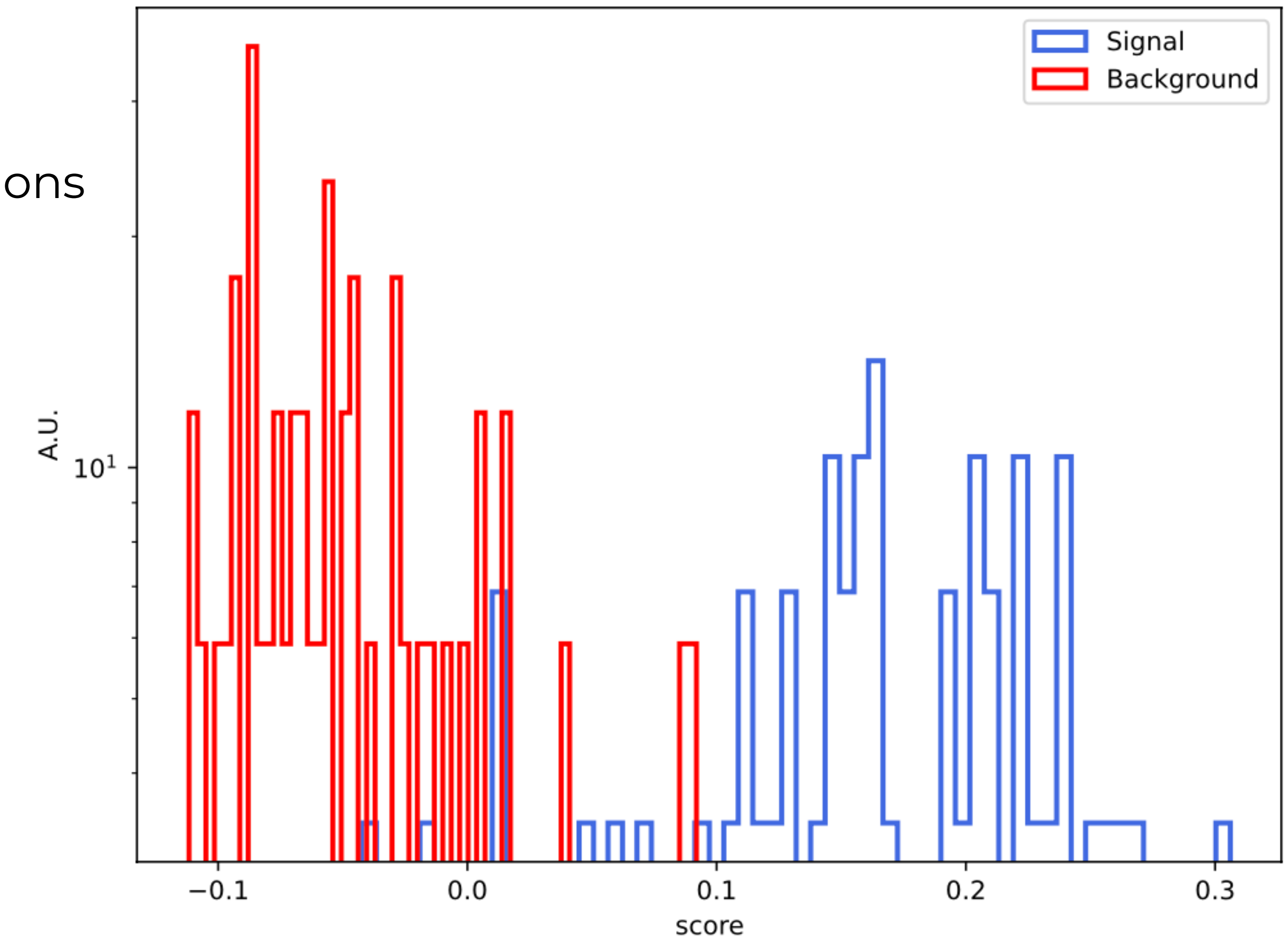



**Expressibility & Entanglement capability** of our data encoding circuit



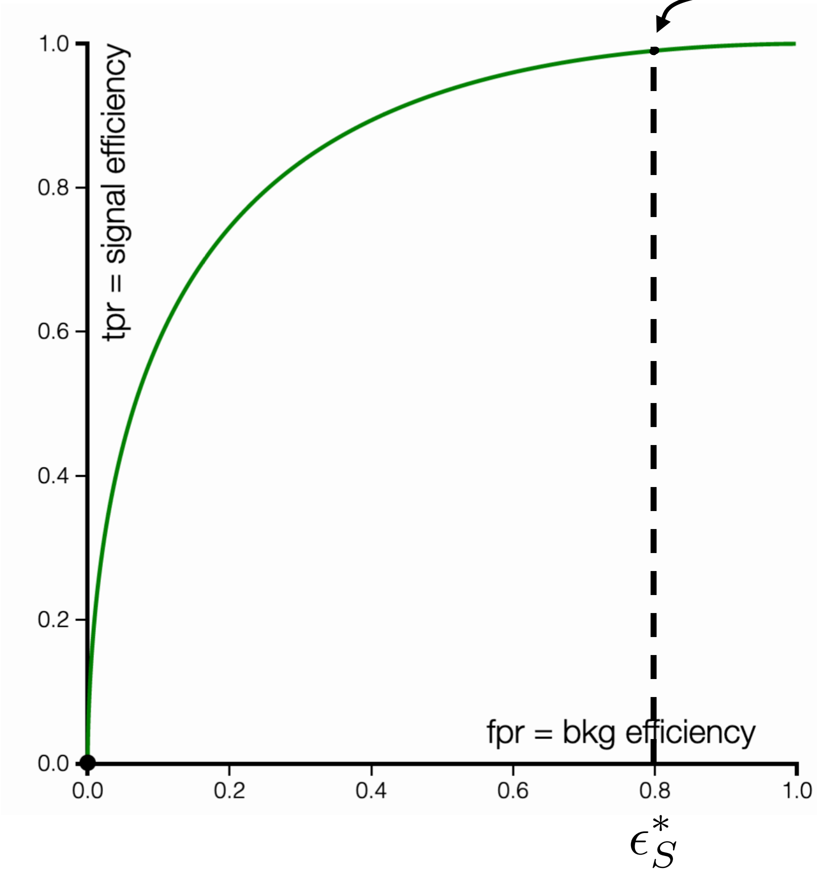
# Sparsity of the distributions in hardware

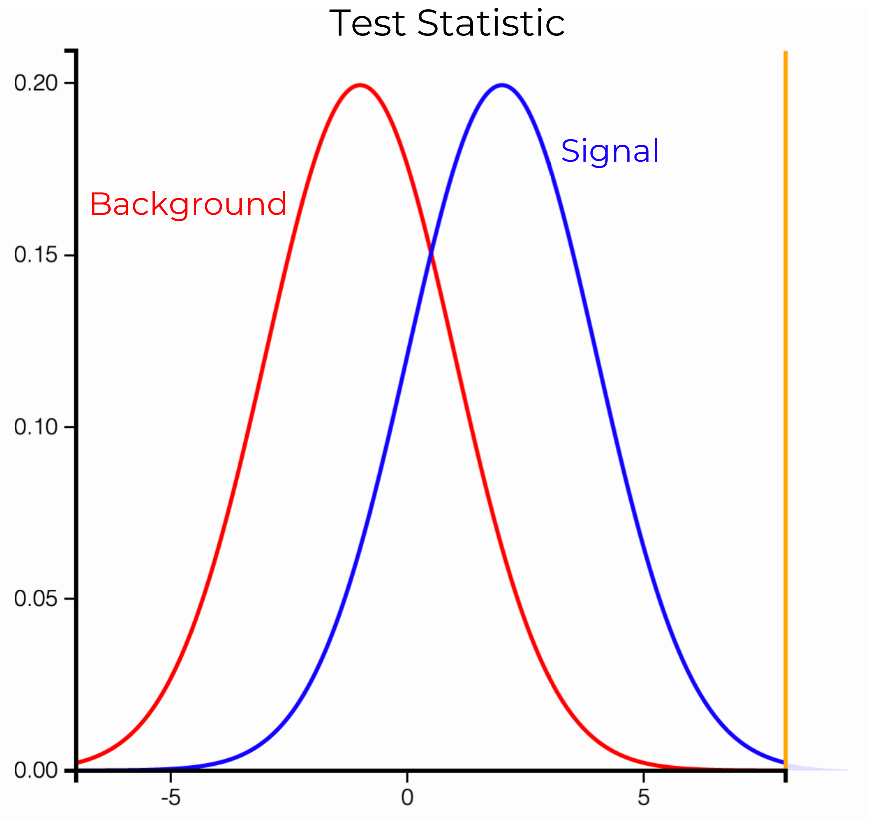
All models have sparse output distributions when training size is low.



# 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 $\epsilon_B(\epsilon_S^*)$

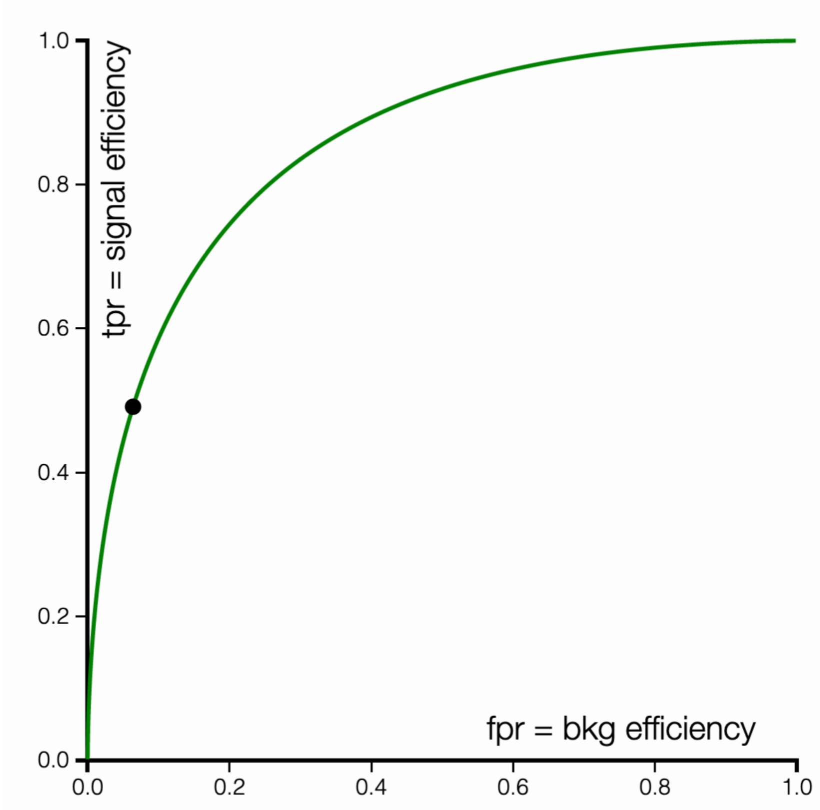


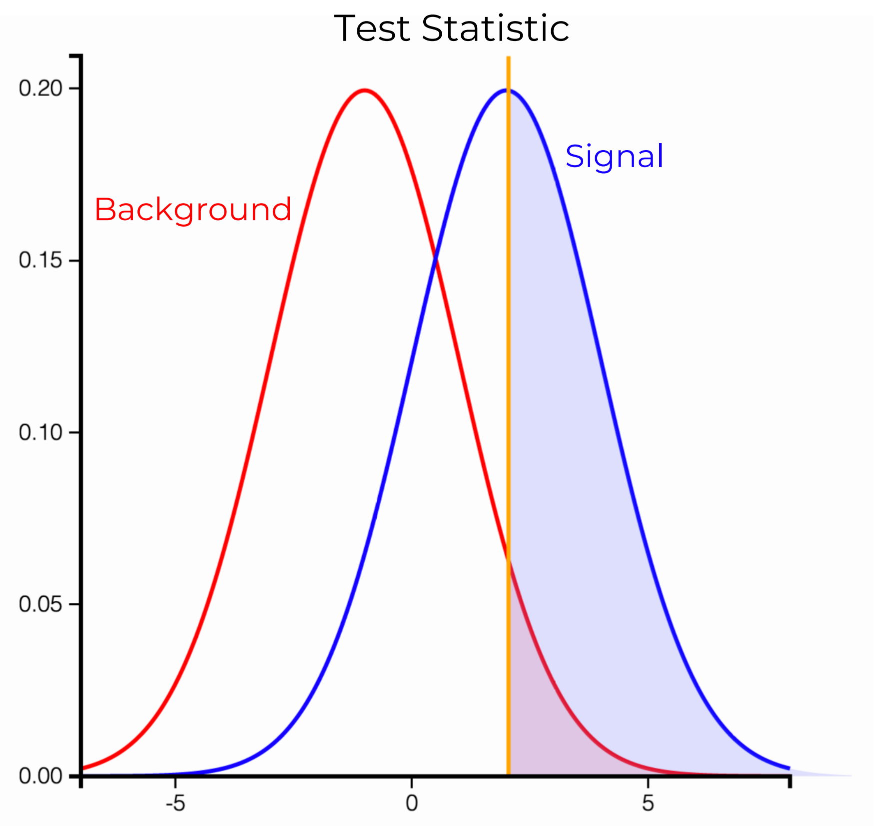




ETH zürich

# Performance Metric

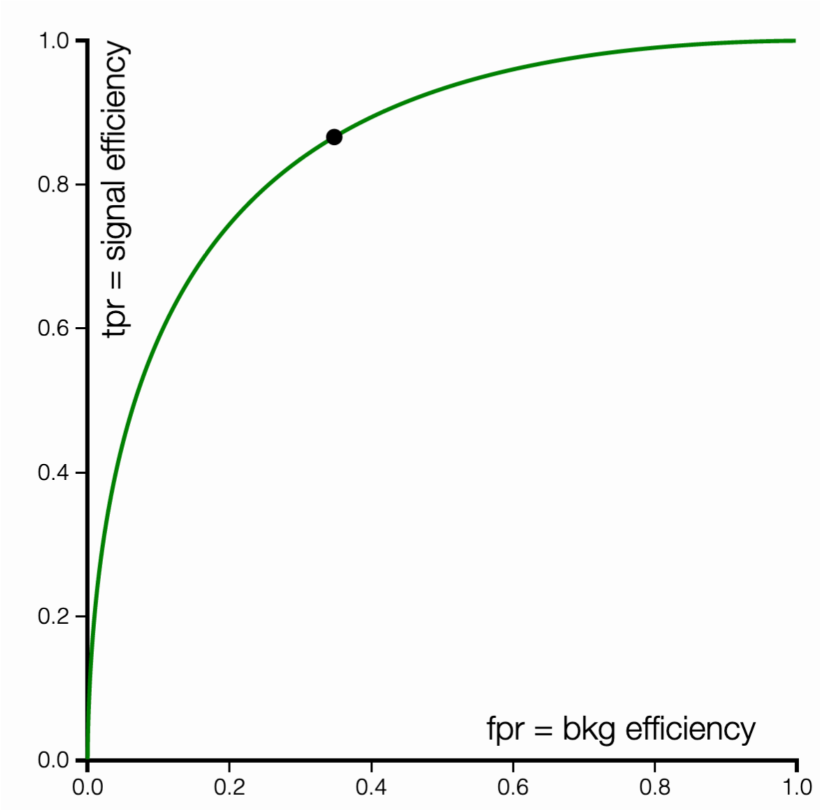
- The normalised data samples are split into training, validation, and testing data sets.
- Classification power metric: Receiver Operating Characteristic (ROC) curve.
- More compact metric: Area Under Curve (AUC) of the ROC curve.
- More practical metric: working point of an analysis $\epsilon_B(\epsilon_S^*)$

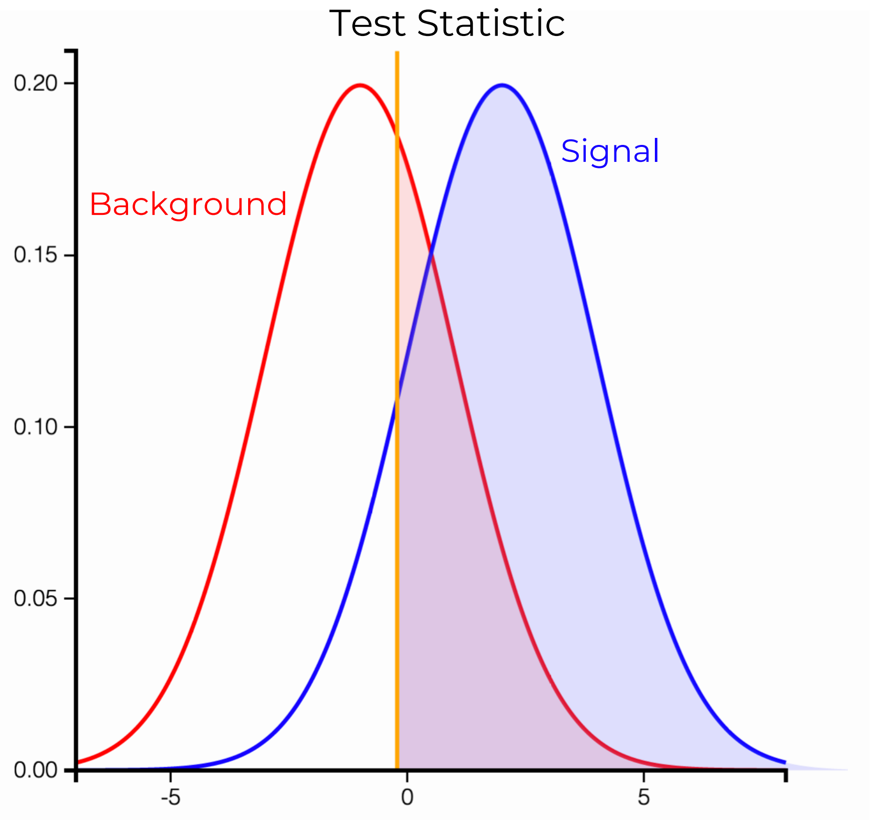


# Performance Metric

- 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 $\epsilon_B(\epsilon_S^*)$



# 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 $\epsilon_B(\epsilon_S^*)$