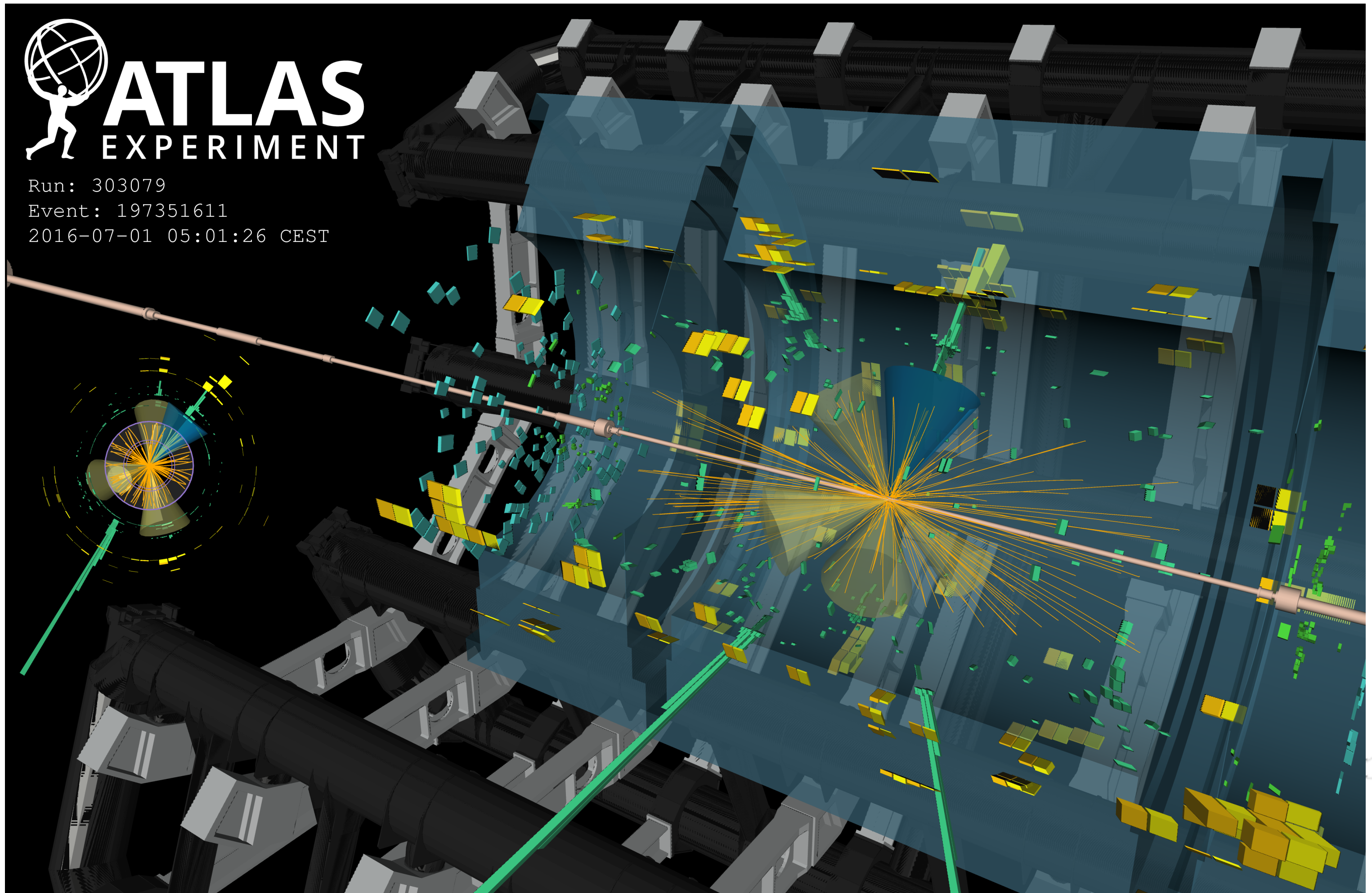
Unsupervised Machine Learning for New Physics Searches

Michael Spannowsky

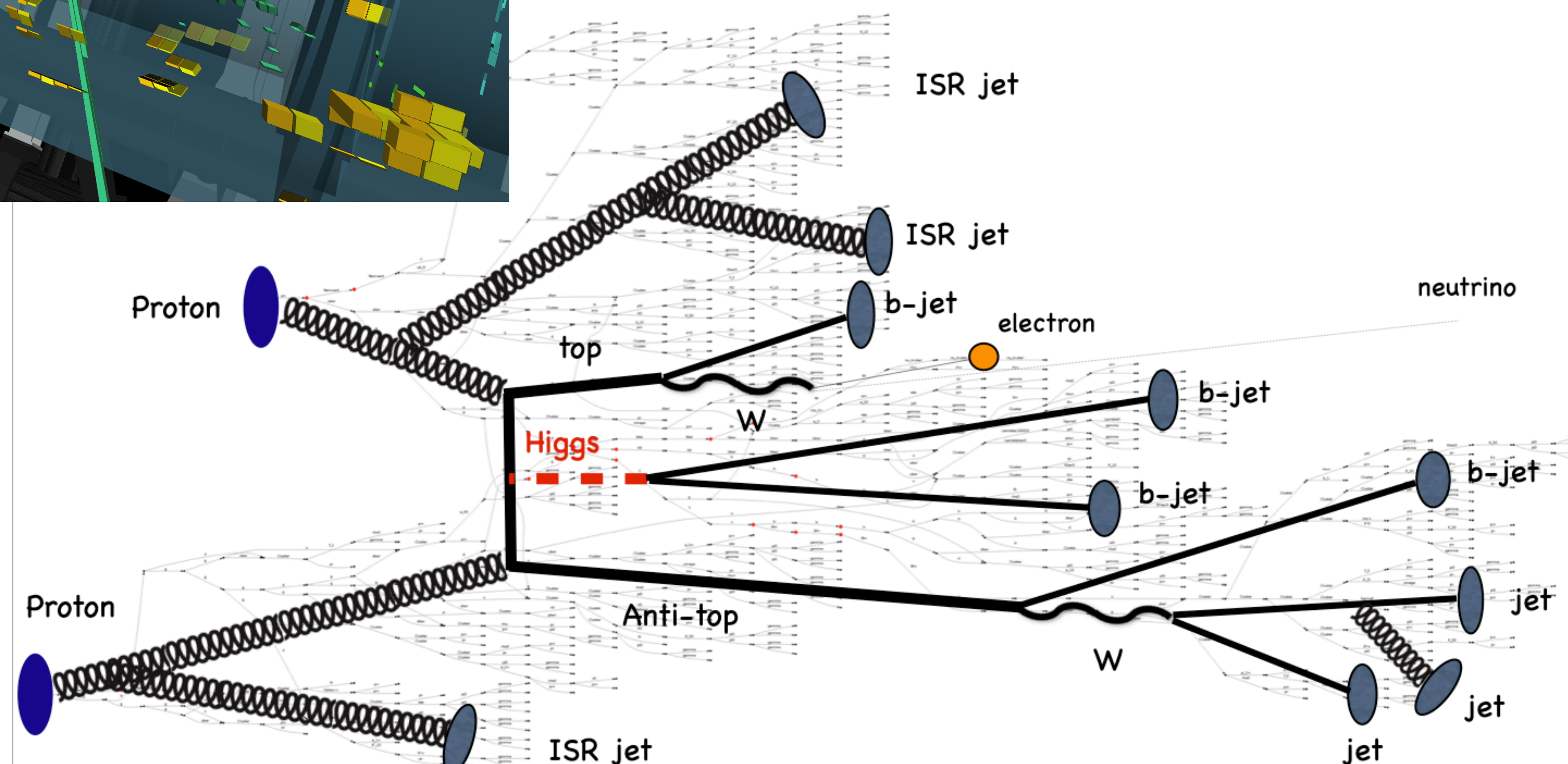
IPPP, Durham University

Complex events LHC

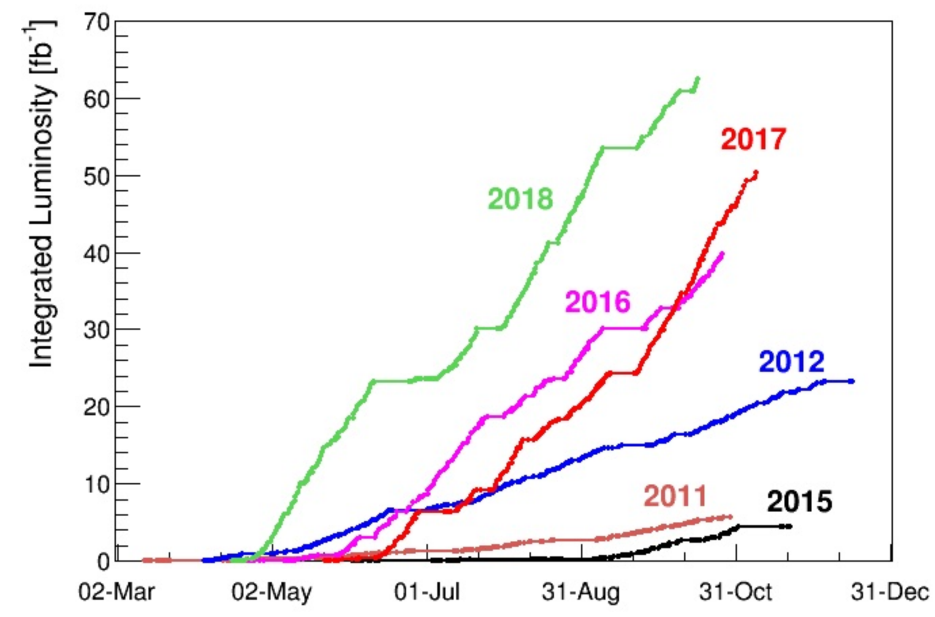


Candidate event tth ATLAS

Theoretical picture of underlying dynamics



Big Data at the LHC



→ ATLAS/CMS 200 events/s passing triggers

→ ATLAS/CMS 2 PB/year of data

High-Energy Physics

Tremendous amount of highly complex data

However, theoretically very precise description of data

Ideal interplay

Machine Learning

Highly performant data analysis techniques

Large number of techniques well suited to physical properties (RNN=time/ordering, CNN=space/orientation)

Hitoshi Murayama:

Hitoshi Murayama:

- SUSY ⟶ Superpartners

Hitoshi Murayama:

- SUSY ⟶ Superpartners
- GUTs ⟶ Z′, W′

Hitoshi Murayama:

- SUSY ⟶ Superpartners
- GUTs ⟶ Z′, W′
- Ex-Dim ⟶ KK-tower

Hitoshi Murayama:

- SUSY ⟶ Superpartners
- GUTs ⟶ Z′, W′
- Ex-Dim ⟶ KK-tower
- Composite ⟶ top partner

Hitoshi Murayama:

- SUSY ⟶ Superpartners
- GUTs ⟶ Z′, W′
- Ex-Dim ⟶ KK-tower
- Composite ⟶ top partner
- Higgs portal/reps ⟶ scalars

Hitoshi Murayama:

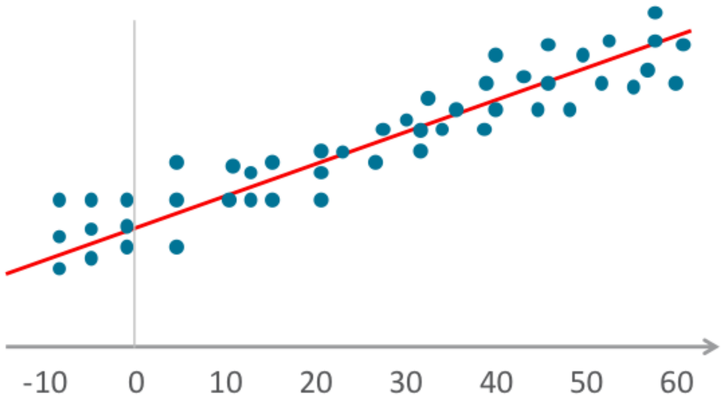
- SUSY ⟶ Superpartners
- GUTs ⟶ Z′, W′
- Ex-Dim ⟶ KK-tower
- Composite ⟶ top partner
- Higgs portal/reps. ⟶ scalars
- Not thought of ⟶ ??

Most models predict TEV scale resonances

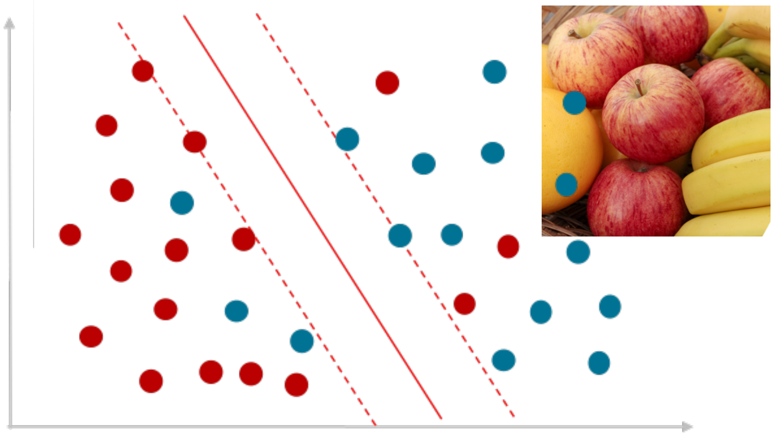
Often similar but not identical features -> difficult to exhaust in bottom-up approach

Supervised

Regression



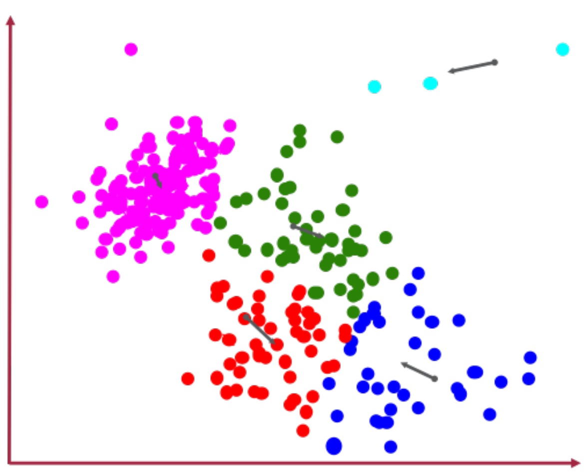
Classification



Fine-grained small net

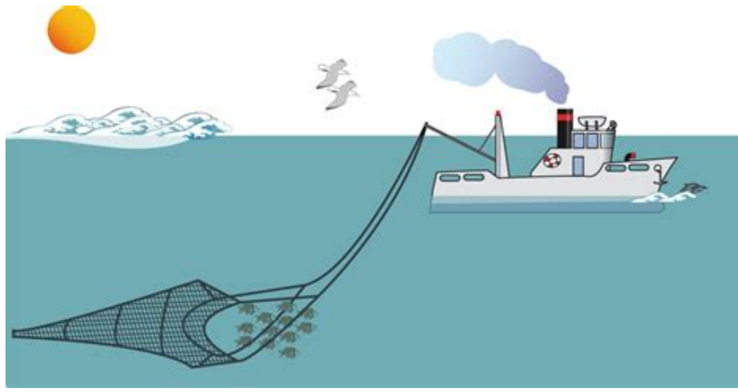
Unsupervised

Clustering

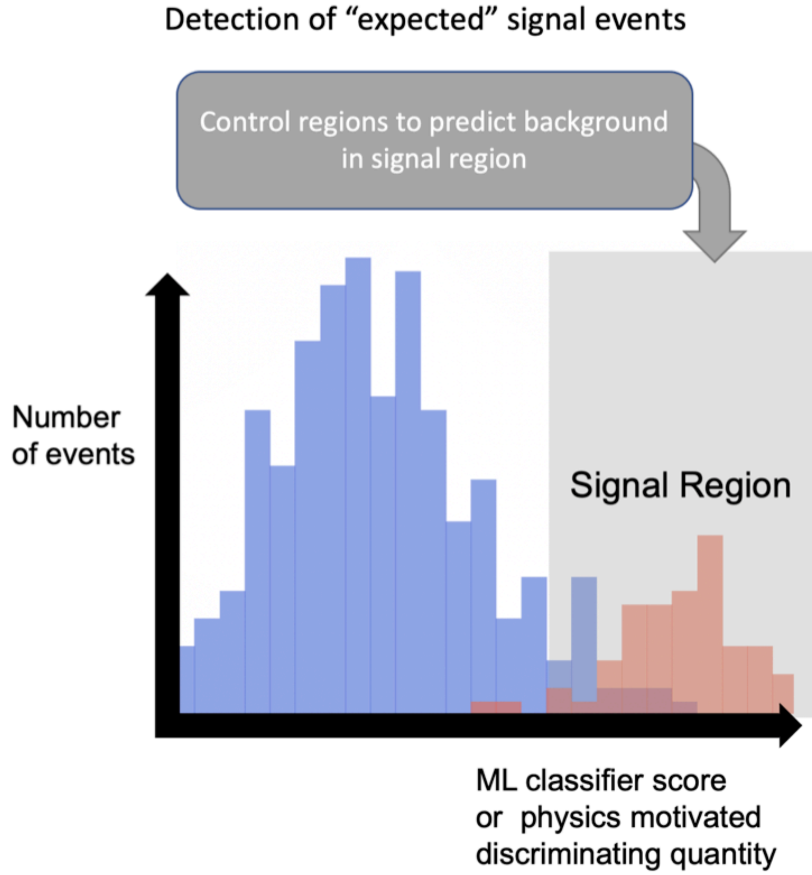


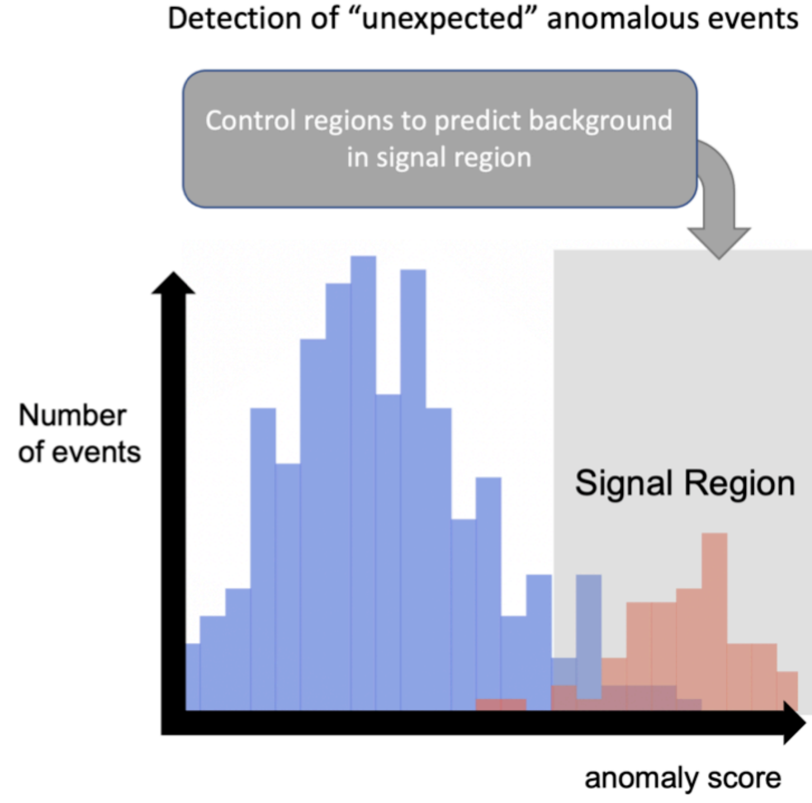
Autoencoder

Large net

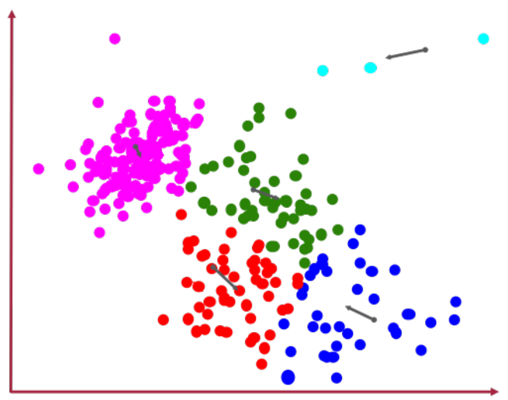


Anomaly detection vs classification





[Aarrested et al '21]

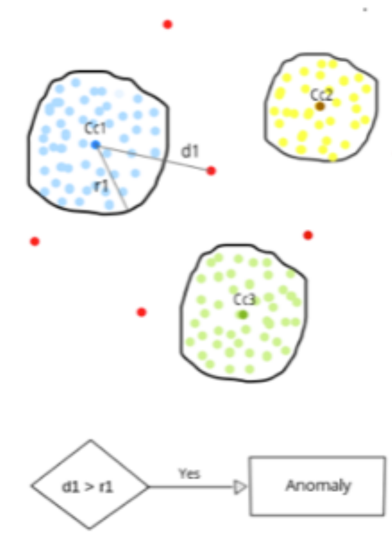


Need to be able to say what is anomalous

individual event vs entire sample

Anomaly detection vs classification

Many different approaches to perform anomaly detection:

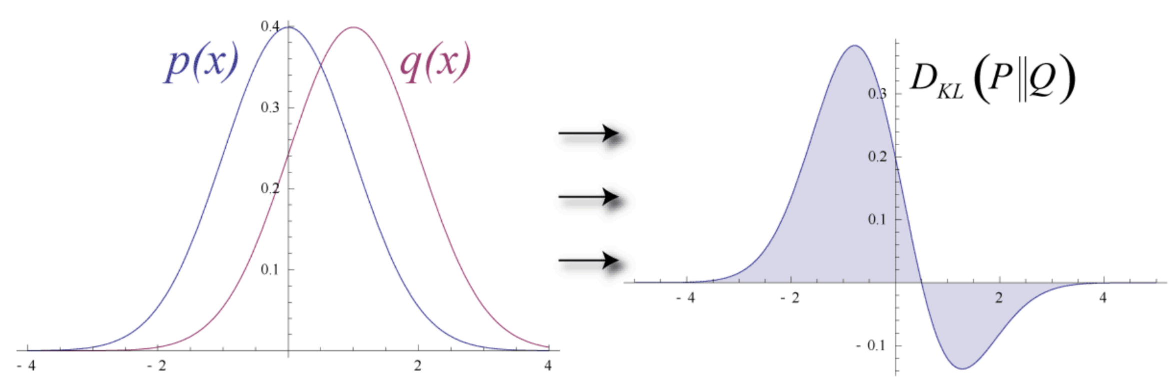


- K-means distance $$\arg\min_{\mathbf{c}} \sum_{i=1}^{k} \sum_{\mathbf{x}\in c_i} d(\mathbf{x}, \mu_i) = \arg\min_{\mathbf{c}} \sum_{i=1}^{k} \sum_{\mathbf{x}\in c_i} \|\mathbf{x} - \mu_i\|_2^2$$
- PCA (SVD)
- Kernel density estimation

And many different ways to assign anomaly score, e.g.

- Means Square Error $$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$
- Kullback-Leibler divergence

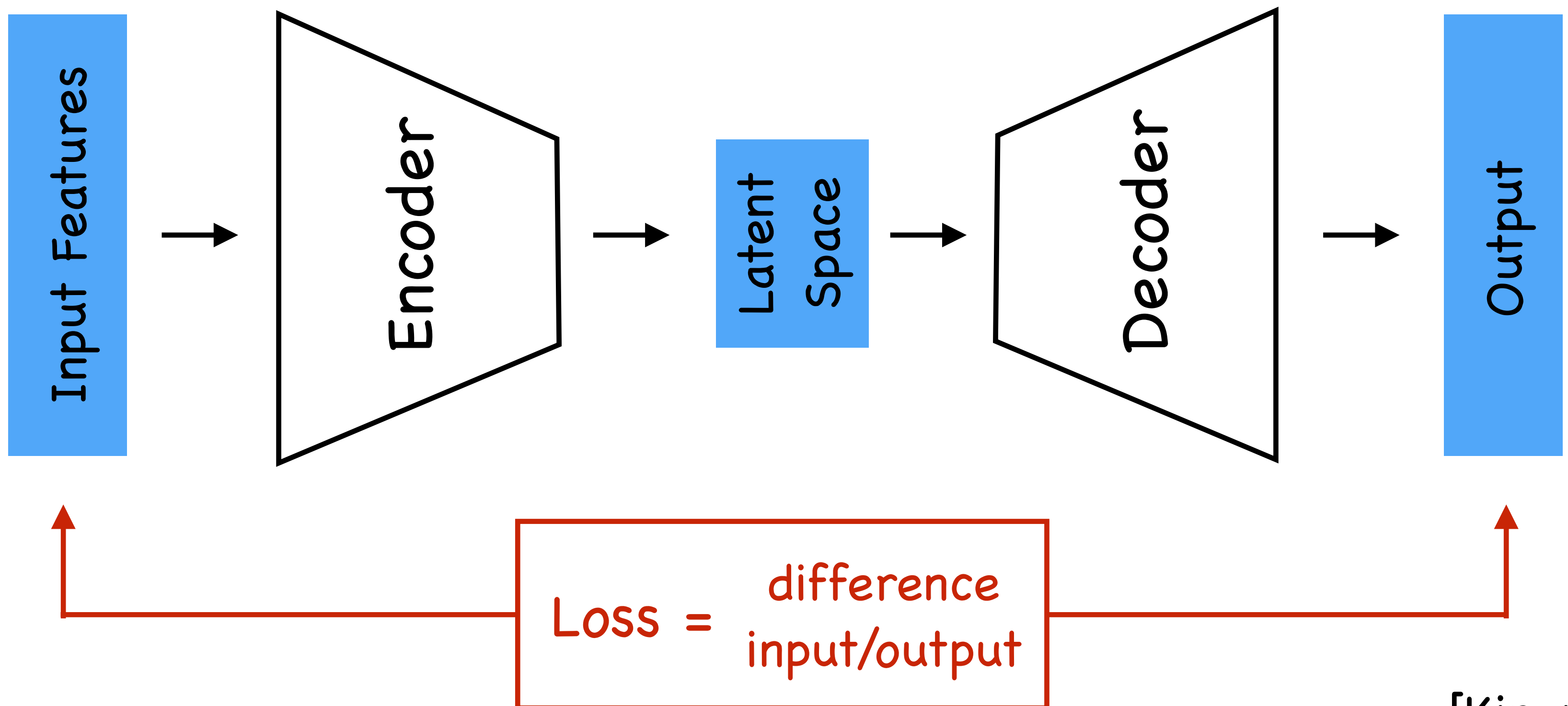
  Measures the difference between two probability distributions



- Also possible to combine for anomaly score

Autoencoder

Most popular NN-based anomaly detection method



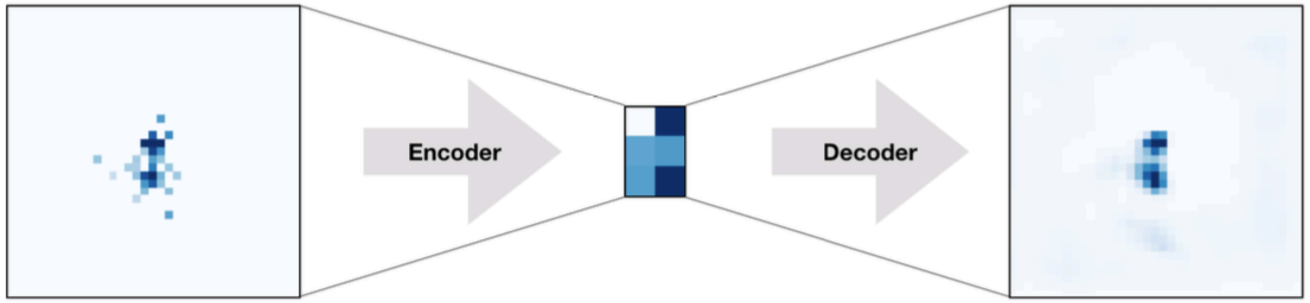
[Kingma, Welling '13]

- in first step input is encoded into information bottleneck
- between input/output layer and bottleneck can be several hidden layers (conv./deep NNs) -> highly non-linear
- after bottleneck decoding step
- Reconstructed output is then compared with input via loss-function (often MSE)
- NN is trained such that input and output high degree of similarity

Convolutional Autoencoder

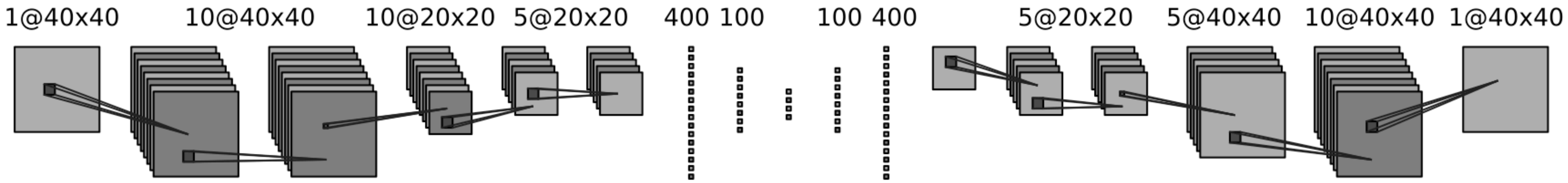
[Heimel et al '18]

[Farina, Nakai, Shih '18]

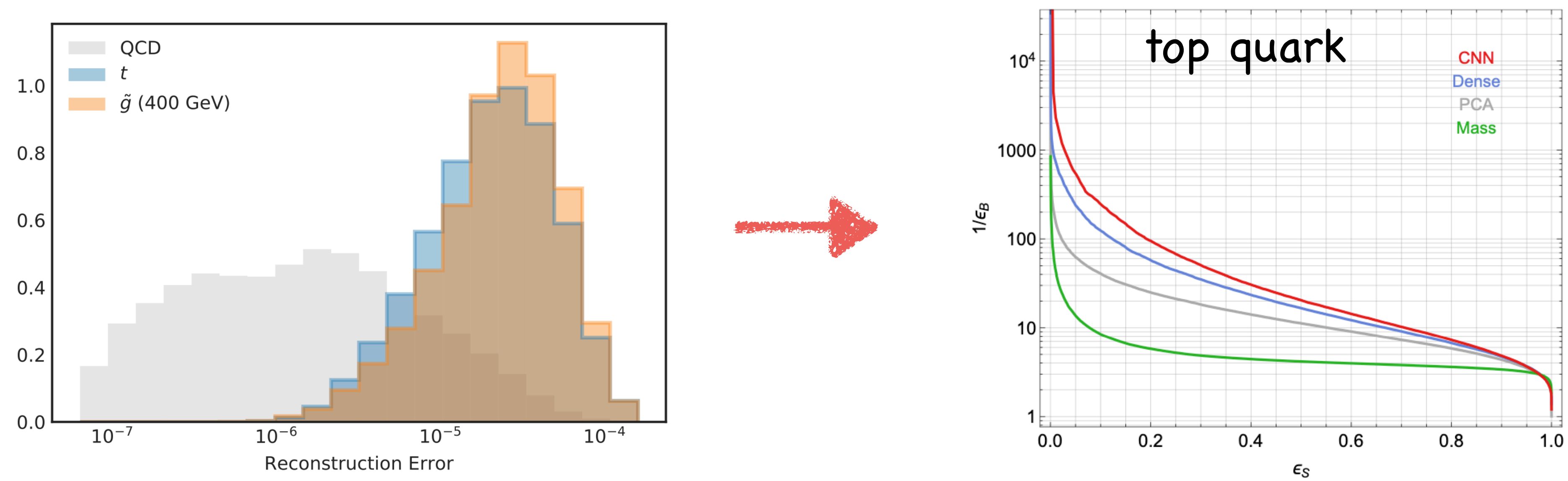


- Combination of CNN with Autoencoder has shown very good performance in jet anomaly detection

- CNN is space/orientation aware information compressor

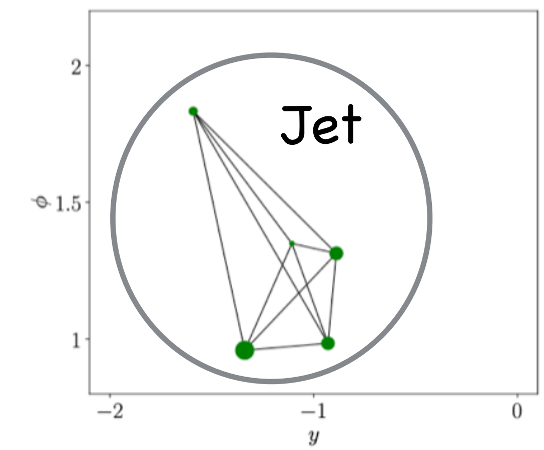


- CNN is space/orientation aware information compressor

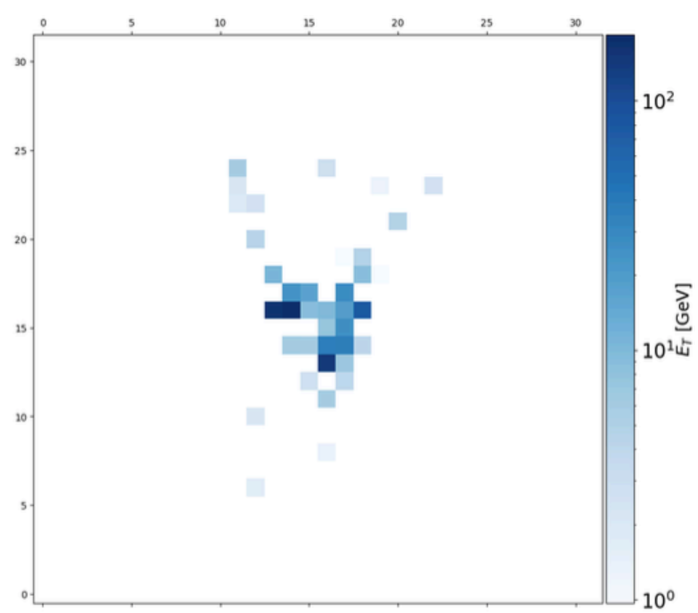


Graph neural network autoencoder

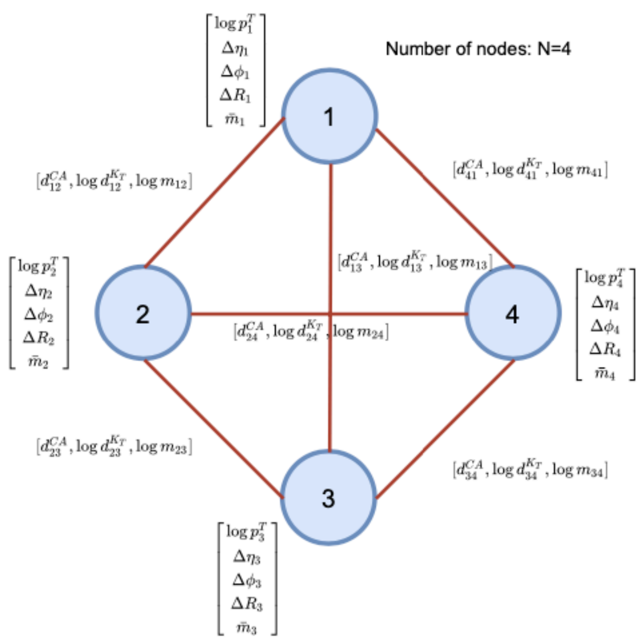
- Graphs are powerful ways of representing data
- Graph: Models set of objects (nodes) and their relationship (edges)



Jet image - CNN

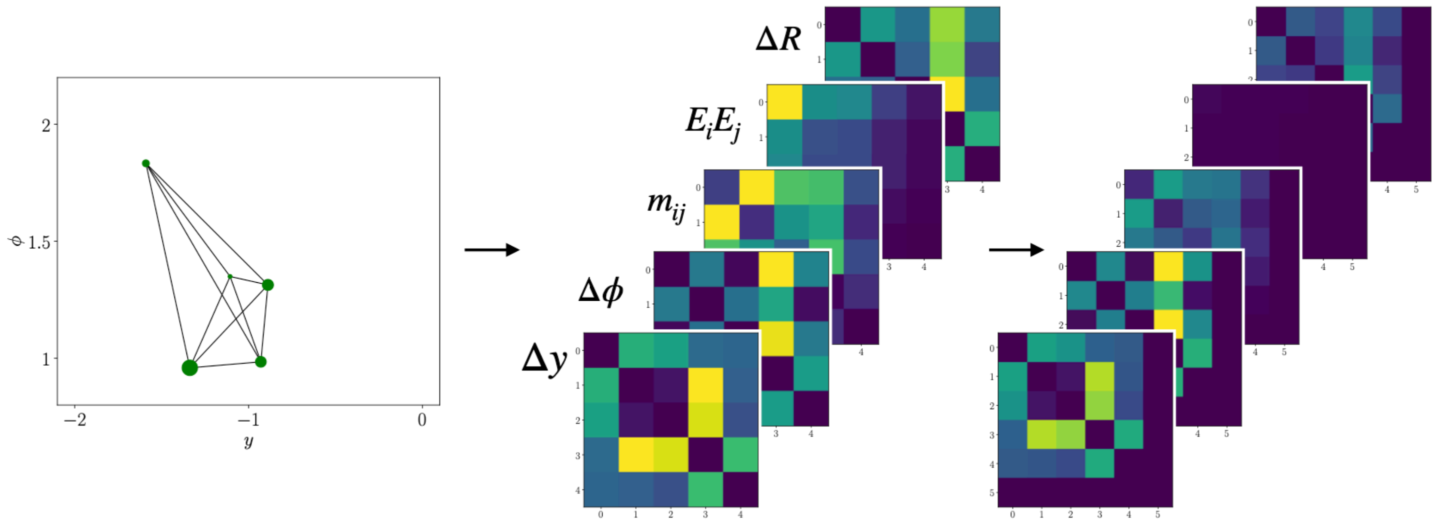


Jet graph - CNN



- Restricted to Euclidean features
- Extremely sparse and comp. wasteful
- Difficult going D>2 -> more wasteful
- fixed length vector

- Domain can be chosen suitable for problem
- Easily extended to D>2
- Variable length vector no problem

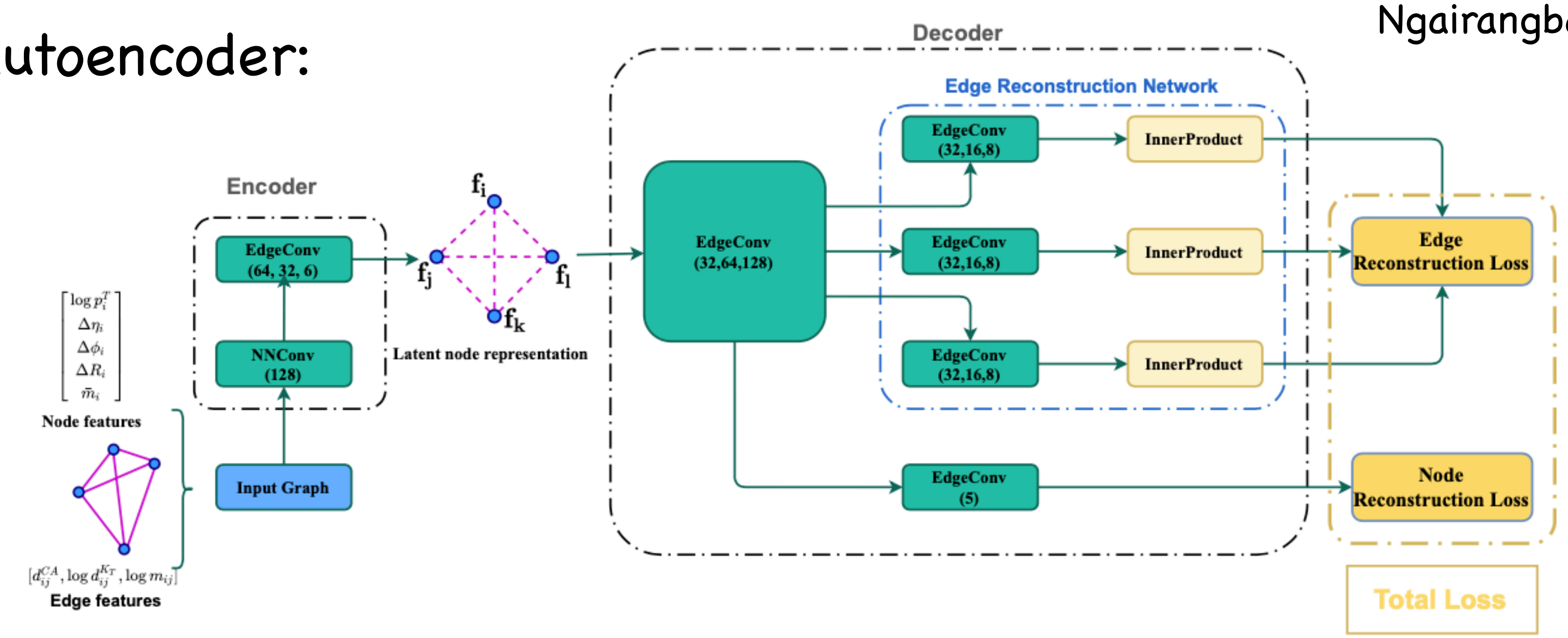


- Graph specified by node and edge features -> adjacency matrix

[Atkinson, Bhardwaj, Englert, Ngairangbam, MS '21]

- Input to graph autoencoder:

multiple adjacency matrices for each graph



Total Loss: $L_{auto} = \lambda_{node}\, L_{node} + \lambda_{edge}\, L_{edge}$ with

$\lambda_{node} = 0.3 \qquad \lambda_{edge} = 1$

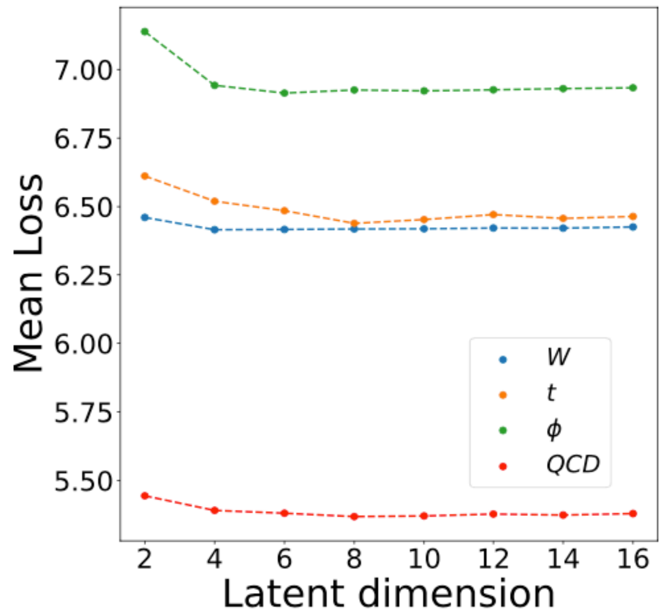
$$L_{node} = \sqrt{\sum_{ia} \frac{(\hat{x}_i^a - x_i^a)^2}{N \times 5}}$$

$$L_{edge} = \sum_a \sqrt{\sum_{ij} \frac{(\hat{A}_{ij}^a - A_{ij}^a)^2}{N \times N}}$$

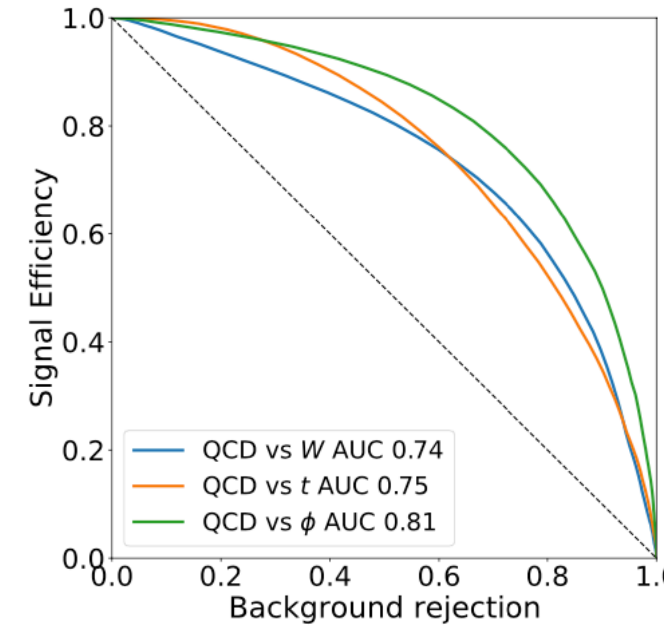
Autoencoders provide two ways for anomaly detection

First optimise for latent space dimension

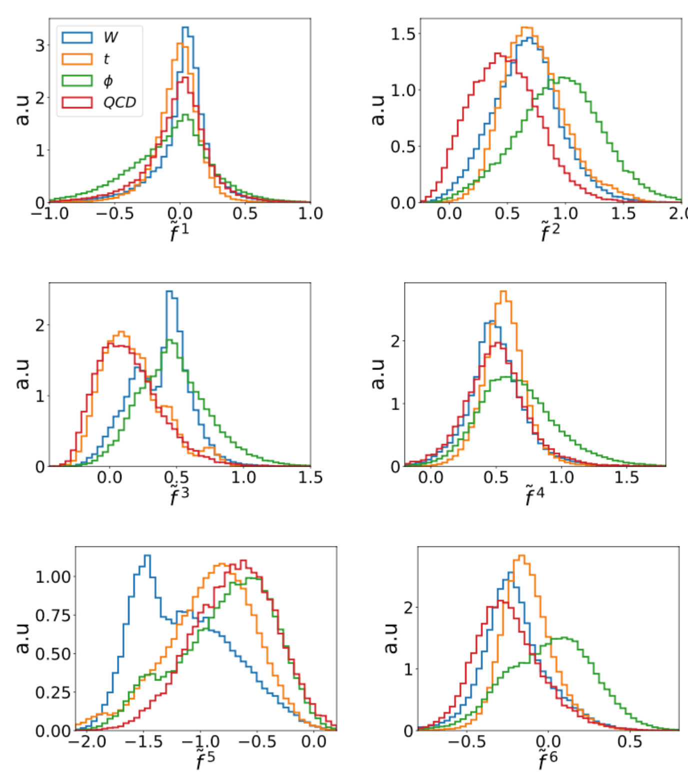
here 6



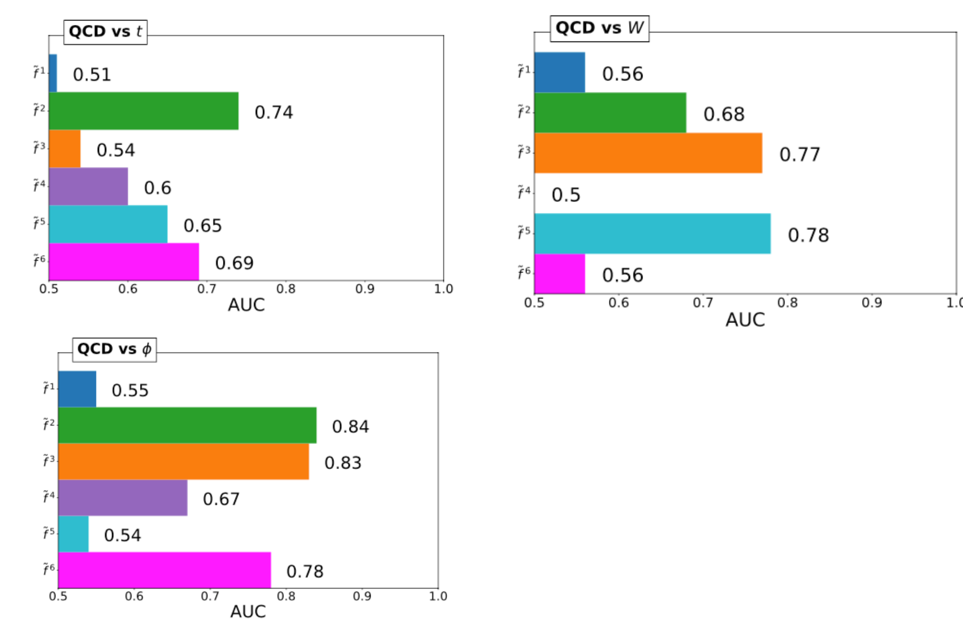
compare input vs output



Latent space contains plenty of information to perform classification task



use latent space



Quantum autoencoder

See talks by
J. Lykken
B. Sanders

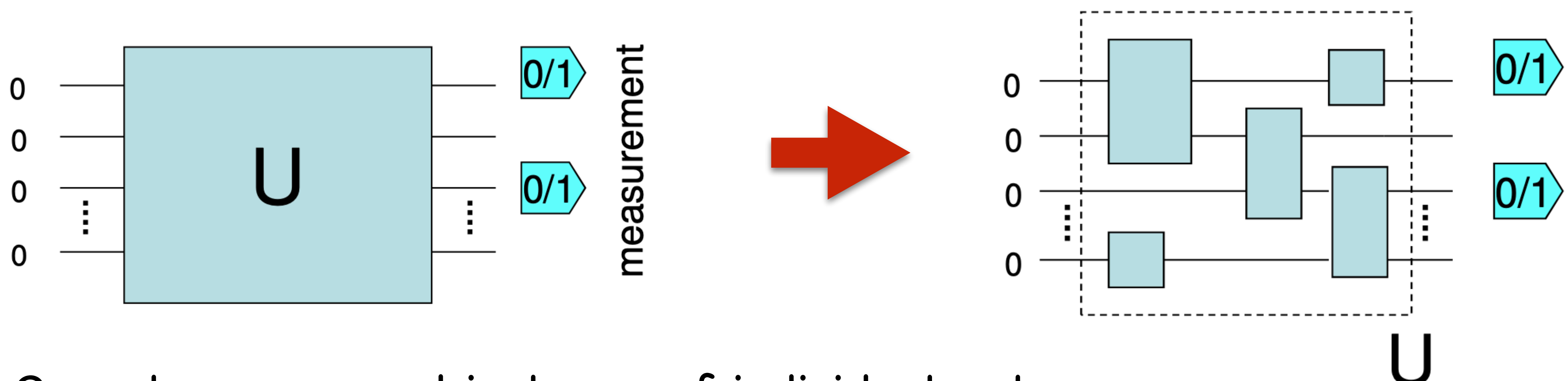
- Quantum algorithms can enhance ML performance
- General structure of any QC algorithm:

operator acts on Hilbert space states $U|x\rangle = |\Psi_1\rangle$

measurement of observable $\hat{U}$ corresponds to exp. value of operator U $\left\langle \hat{U} \right\rangle_\Psi = \frac{\langle \Psi | U | \Psi \rangle}{\langle \Psi | \Psi \rangle}$

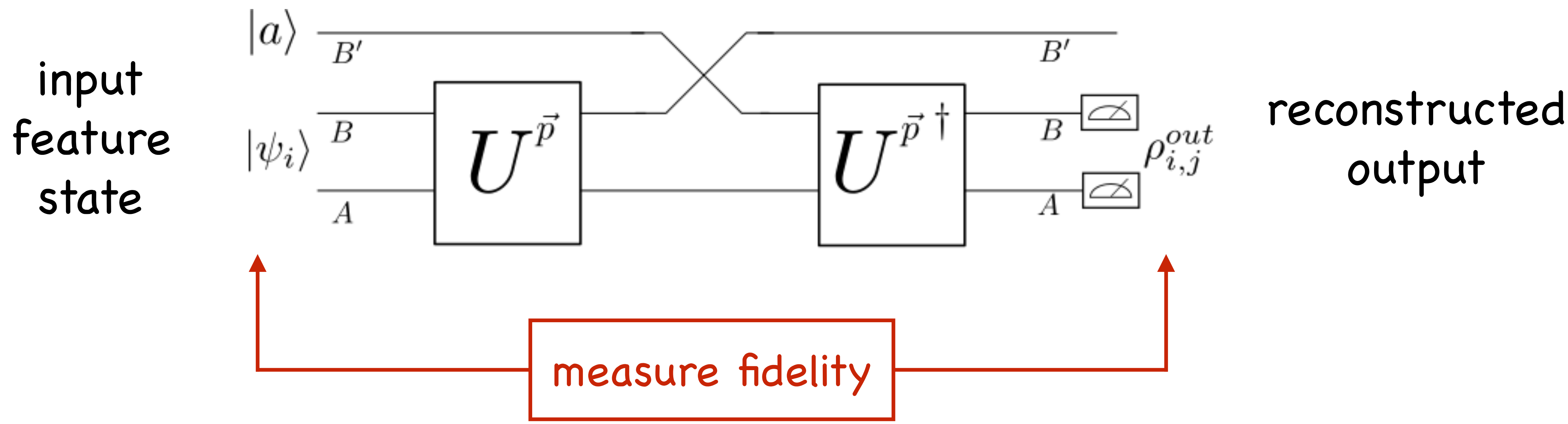
statistical statement need to evaluate often

Need to encode Hilbert space and operator suitable for quantum system

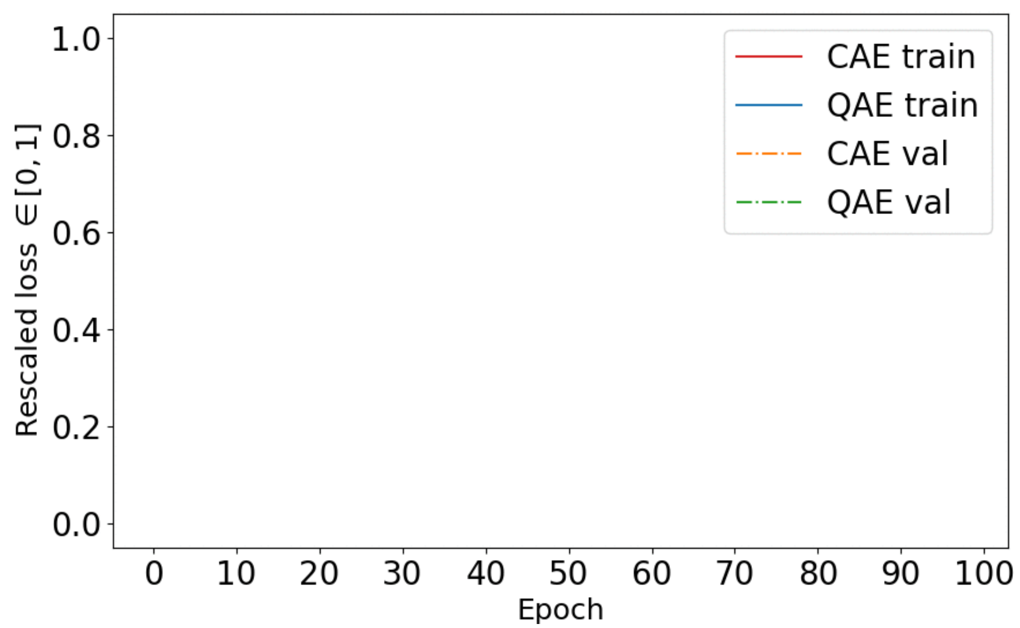


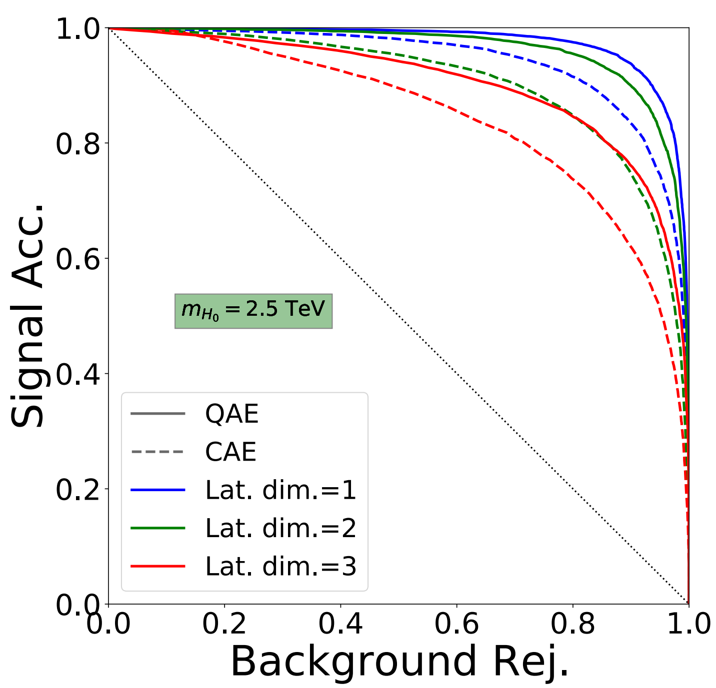
- Operator expressed in terms of individual gates

- Implementation of a quantum autoencoder



- For network training quantum gradient descent is beneficial [Blance, MS '20]
- Find improved performance over classical autoencoder and extremely fast training [Ngairangbam, MS, Takeuchi 'next week]



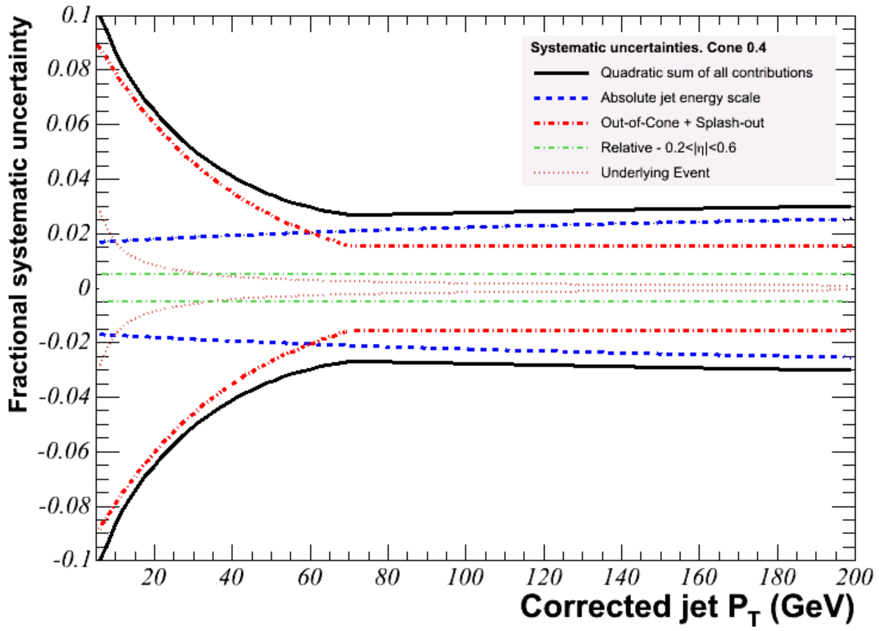


Unsupervised learning prone to un

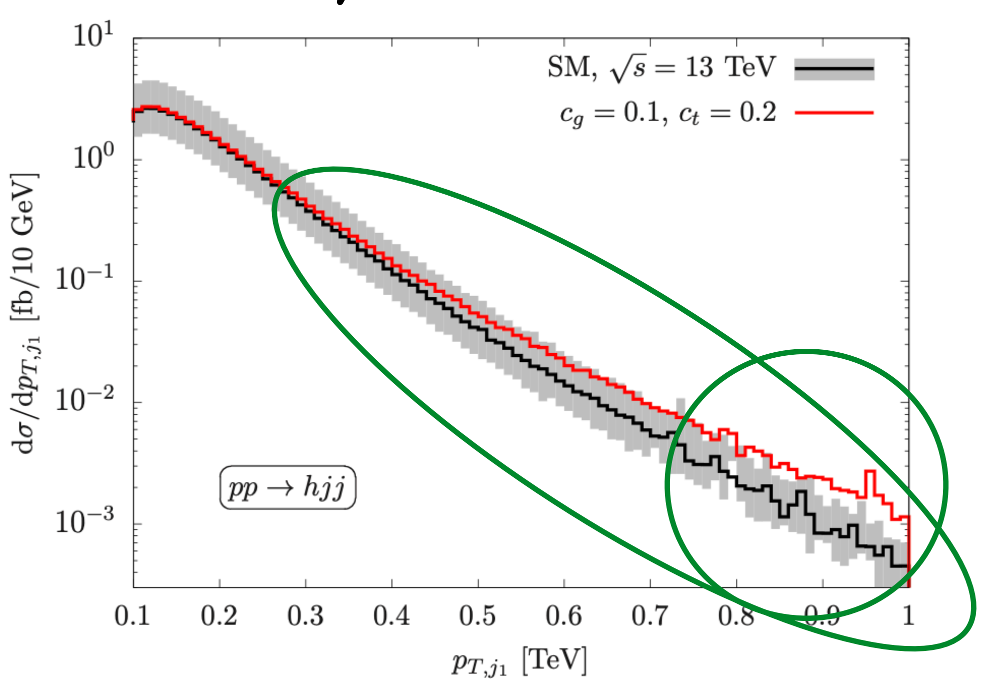
- In general, performance curves etc should often be taken with some grain of salt (data vs pseudo-data)

- Known uncertainties should be taken into account if possible

syst. exp. uncertainties



theory uncertainties

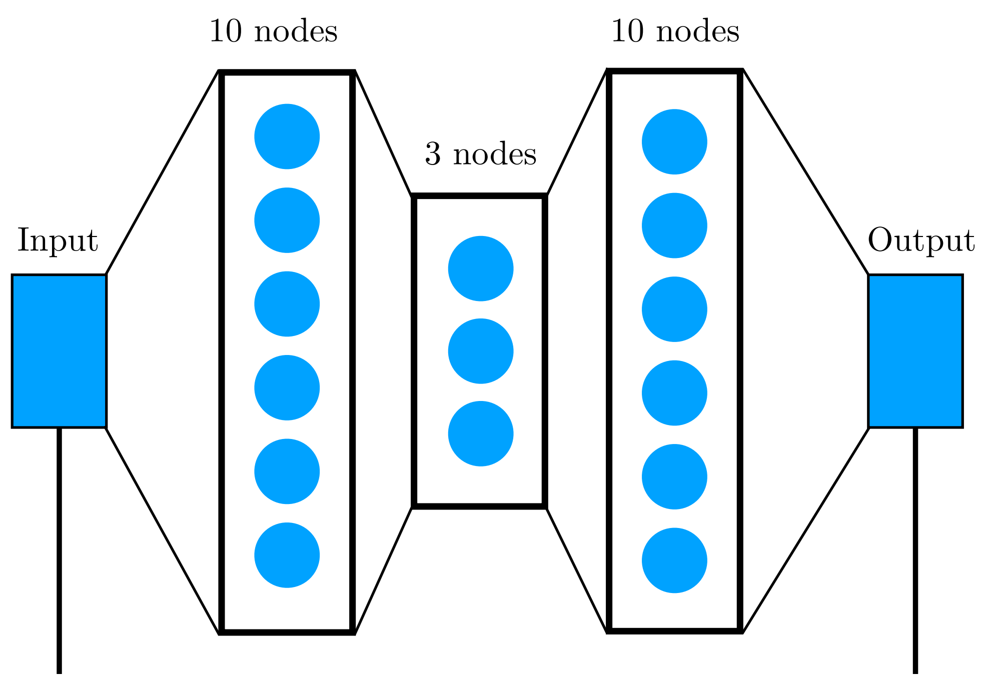


- Adversially trained NNs can help making performance estimates insensitive to uncertainties

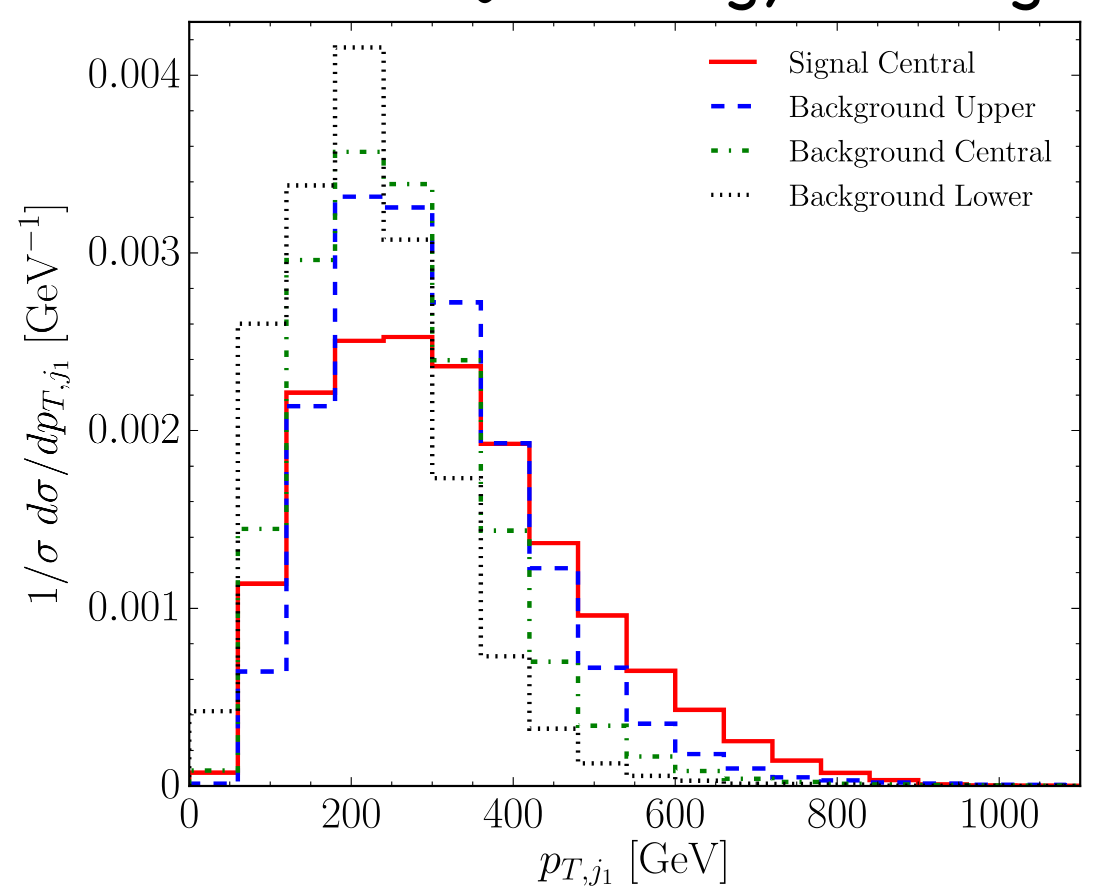
Example for uncertainties for anomaly detection with autoencoder

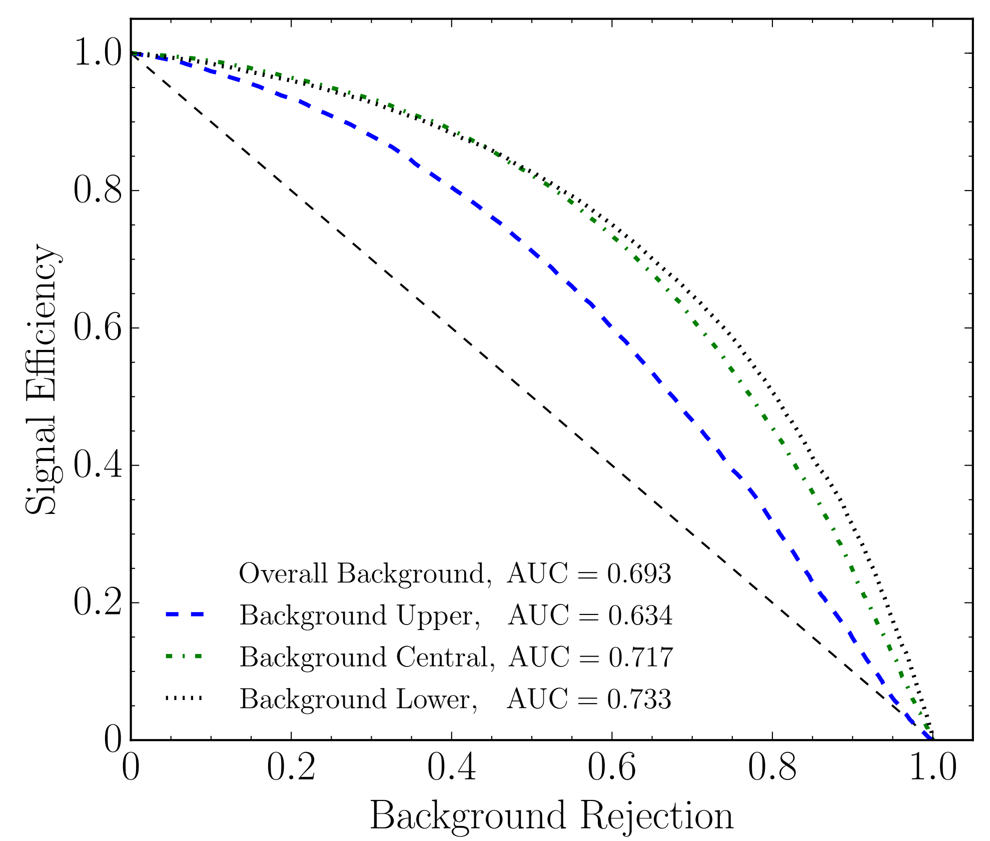
signal: pp->Z'->tt   bkg: pp->tt

- to benchmark sensitivity first without adversarial
- only train on bkg (anomaly detection)
- syst. uncertainties via jet, MET smearing



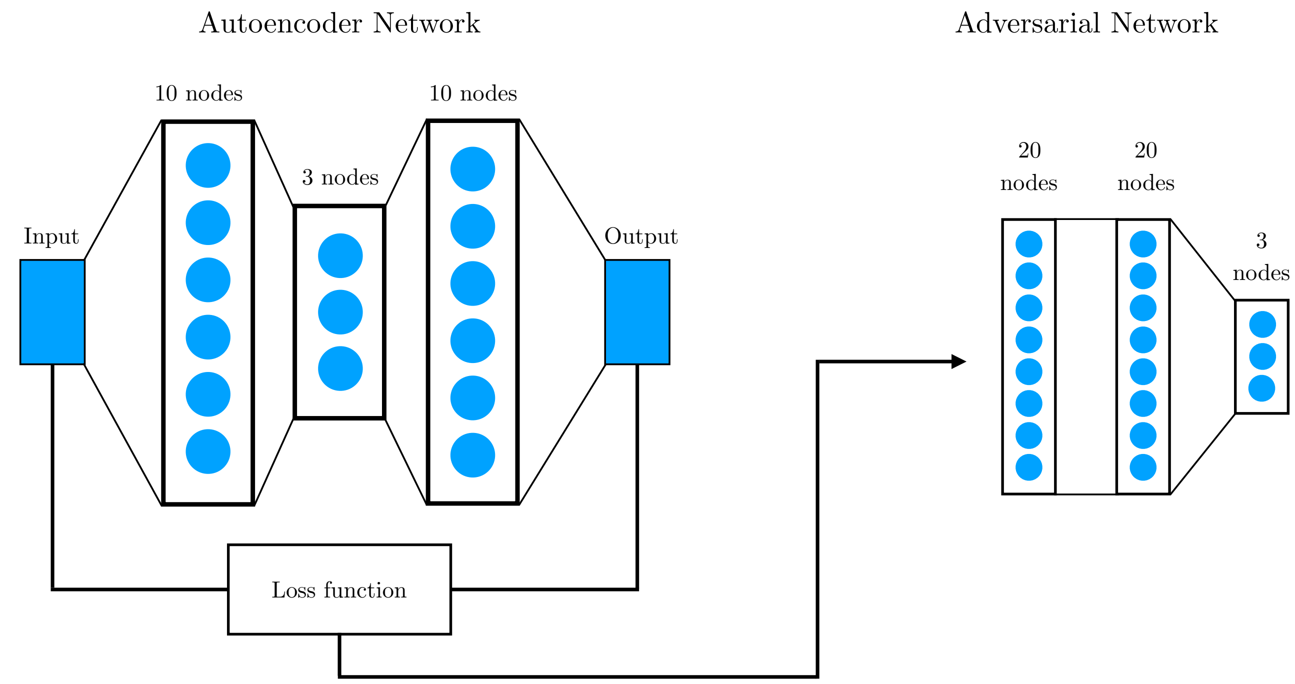
smeared jet energy for bkg



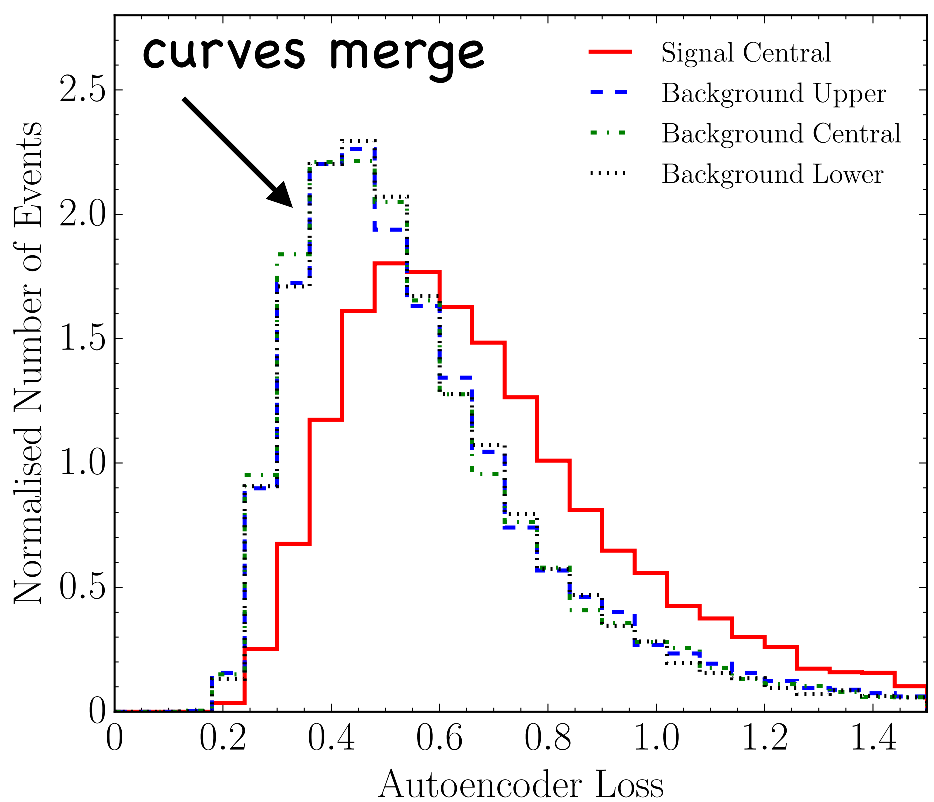


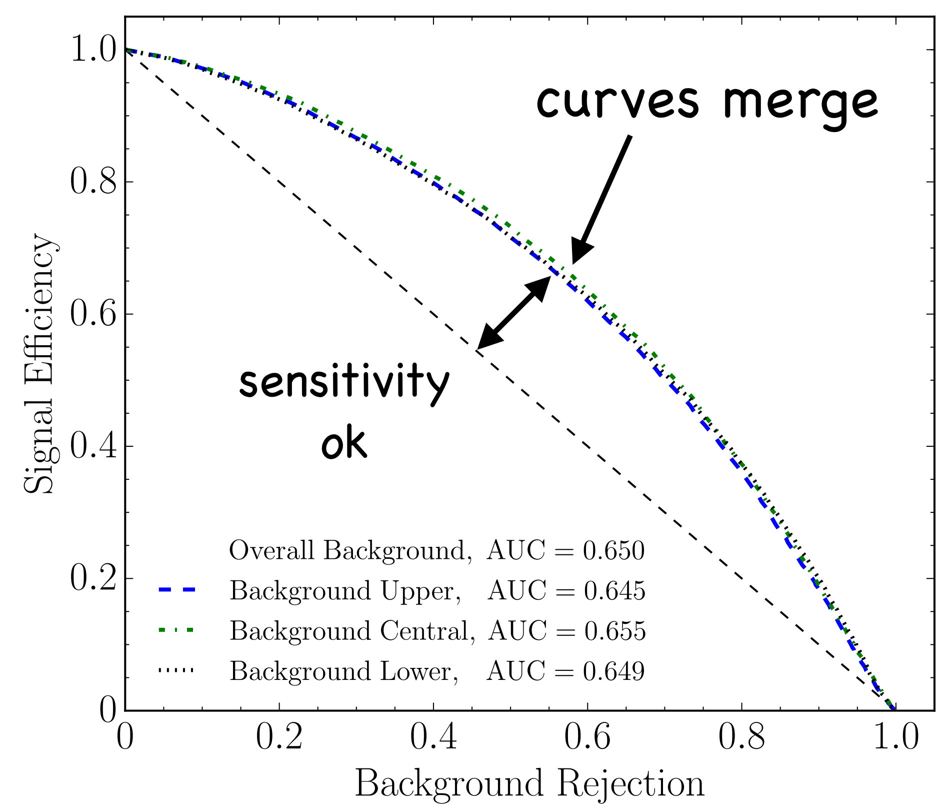
- We can remove dependance on smearing by applying an adversary that will try to classify the direction of smearing

- The two networks are in a zero-sum game - an increase in adversary performance will result in the autoencoder being penalised.

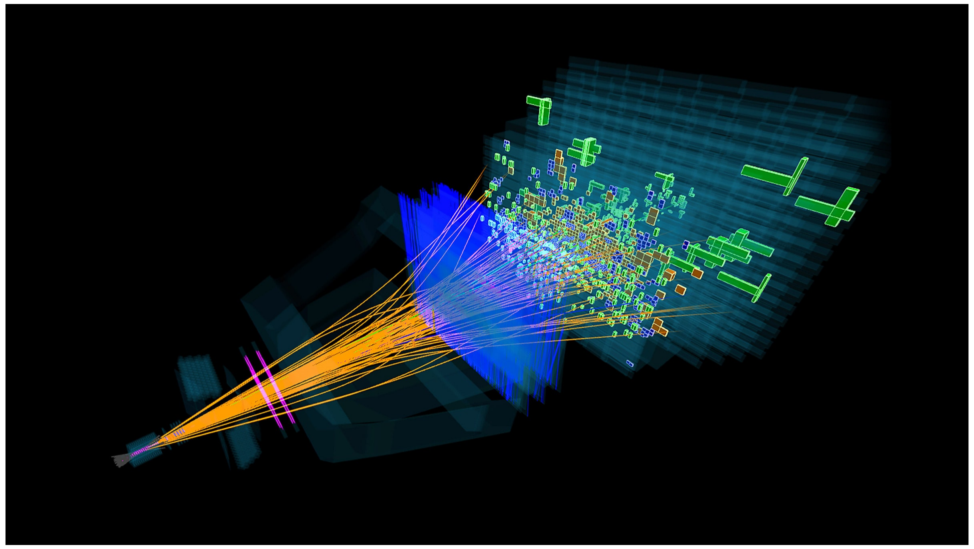


$$\mathcal{L}_{\text{tot}} = \mathcal{L}_{\text{auto}} - \alpha\mathcal{L}_{\text{adv}} \qquad \alpha = 100$$

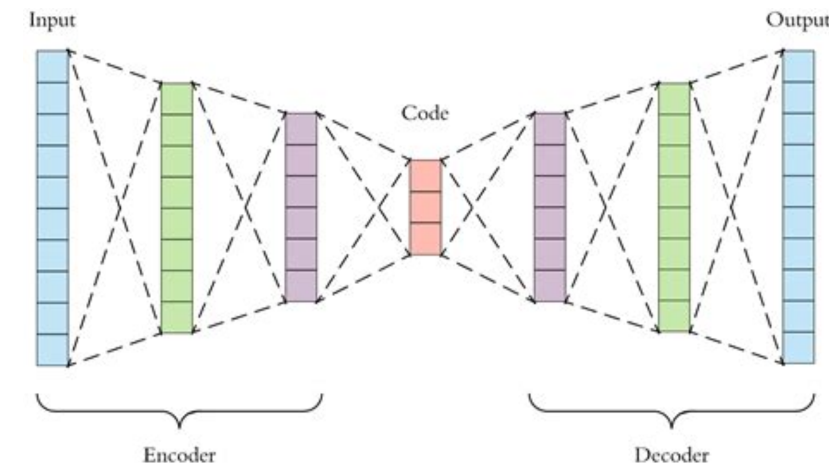




Robust, yet sensitive, anomaly detection with adversarial autoencoders



Summary



- Anomaly detection is important discipline to ensure no new physics is missed at LHC
- Autoencoders are the most popular NN realisation of anomaly detection methods
- Autoencoders can be combined with other network methods to incorporate physics knowledge:

    CNN - spacial knowledge

    RNN - time/orderings

    Adversarials - to desensitise against known unknowns