Anomaly Detection for Experimental Physics

EMFCSC International School of Subnuclear Physics

Gravity and Matter in the Subnuclear World

58th Course. Erice 15-24 June 2022

Eric A. Moreno (MIT)

Who am I?

- Not a theorist!
- Physics & ML Dual PhD - my passion lies at the intersection of both
- I work on:
  - Jet tagging with novel ML architectures (CMS)
  - Gravitational-wave detection (LIGO)
  - Spiking neural networks (CERN, Intel)
  - Level 1 Trigger system – extreme data reduction (CMS)
  - Dark matter experiment at Fermilab (DarkQuest)
  - Generalized intelligence models - deep metric learning
  - Collider concepts - stay tuned!

Who am I?

- Not a theorist!
- Physics & ML Dual PhD - my passion lies at the intersection of both
- I work on:
  - Jet tagging with novel ML architectures (CMS)
  - Gravitational-wave detection (LIGO)
  - Spiking neural networks (CERN, Intel)
  - Level 1 Trigger system – extreme data reduction (CMS)
  - Dark matter experiment at Fermilab (DarkQuest)
  - Generalized intelligence models - deep metric learning
  - Collider concepts - stay tuned!

Three-Step Rabbit Hole

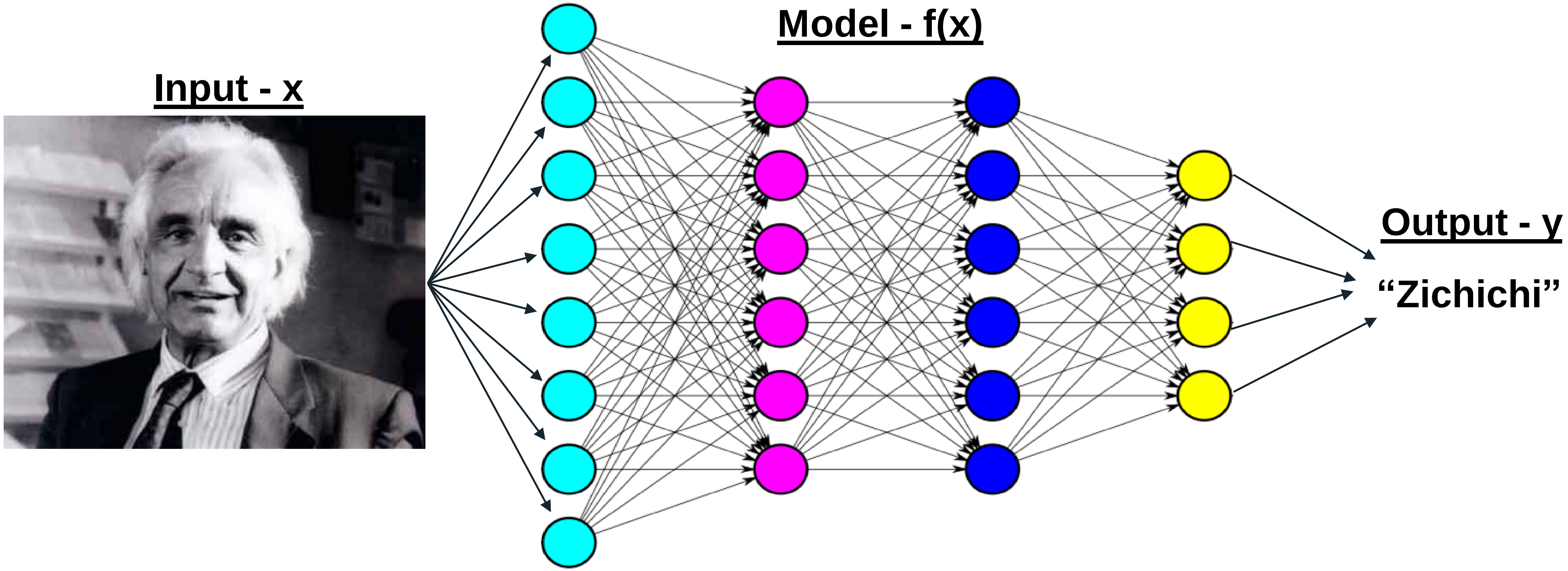
I want to present you with 3 (increasingly complicated) ways* to detect anomalies in your experiment:

1. Autoencoder anomaly detection
2. Quasi-Anomalous Knowledge (QUAK)
3. Latent-Space deep metric learning

*Not all my work - in collaboration with P. Harris (MIT), S.E. Park (MIT), M. Yunis (MIT), D. Rankin (MIT), M. Pierini (CERN)

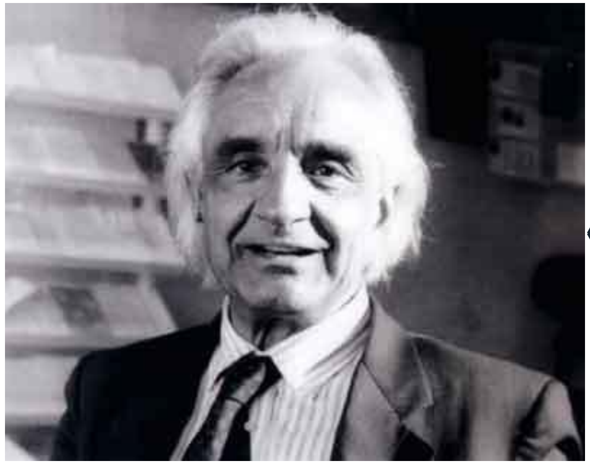
Intro to Machine Learning

Intro to Machine Learning

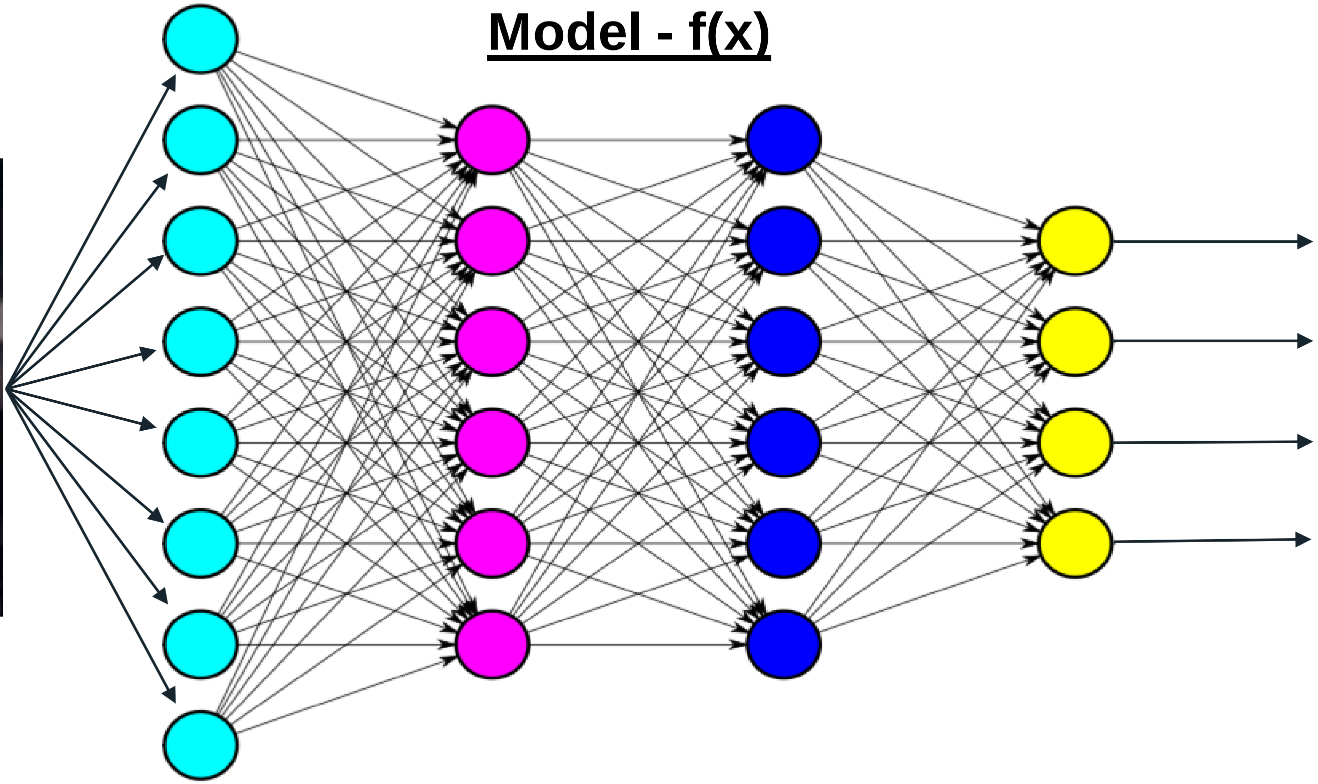


Intro to Machine Learning

Input - x



Model - f(x)



Output - y

25% “Dirac”

25% “Zichichi”

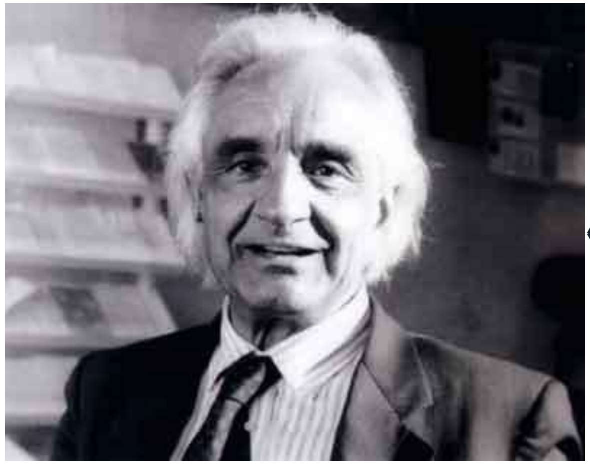
25% “Majorana”

25% “Einstein”

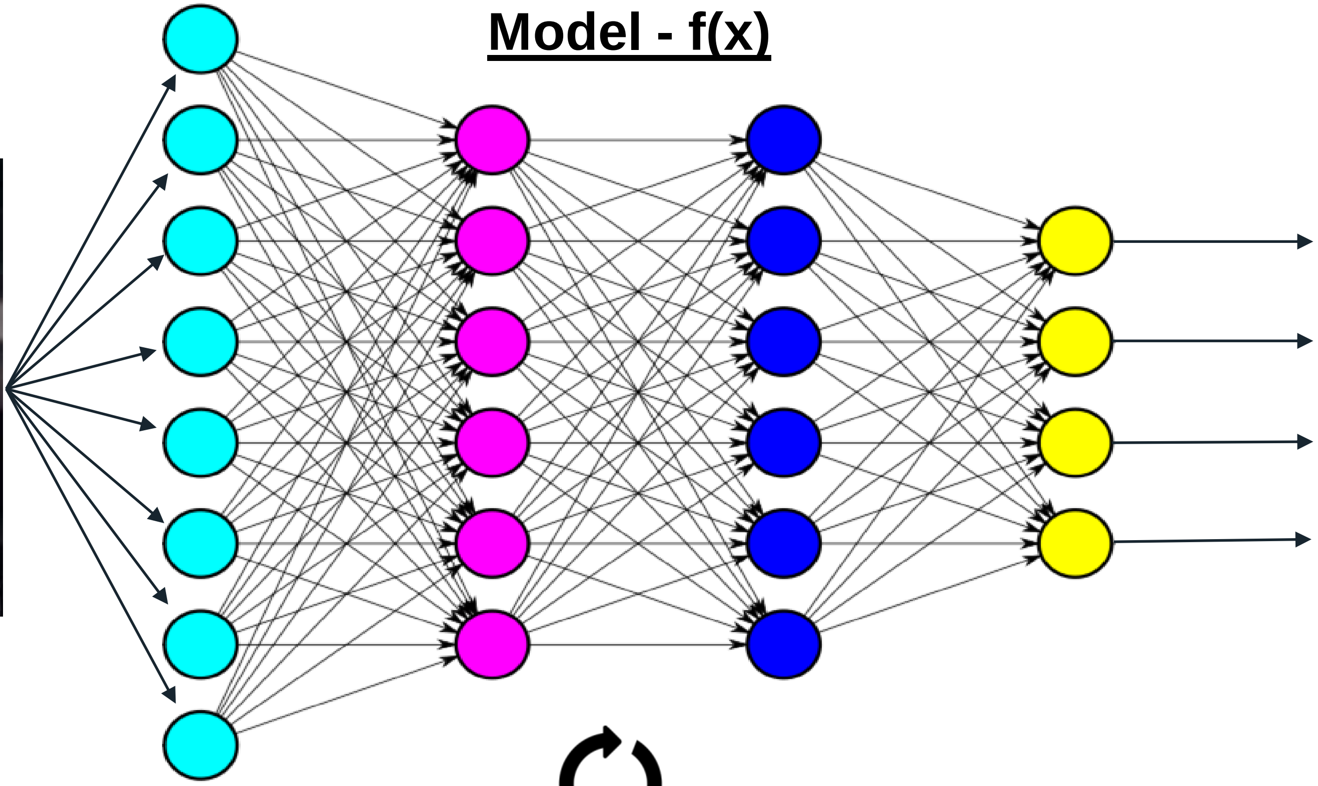
Weights θ at each of these connections encode the network’s knowledge. Weights are continually updated using gradient descent $\theta - \eta \frac{dL}{d\theta} \rightarrow \theta$

Intro to Machine Learning

Input - x



Model - f(x)



Output - y

8% “Dirac”

63% “Zichichi”

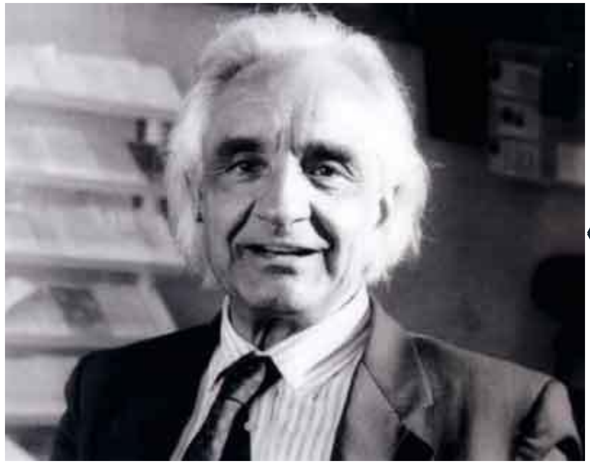
13% “Majorana”

36% “Einstein”

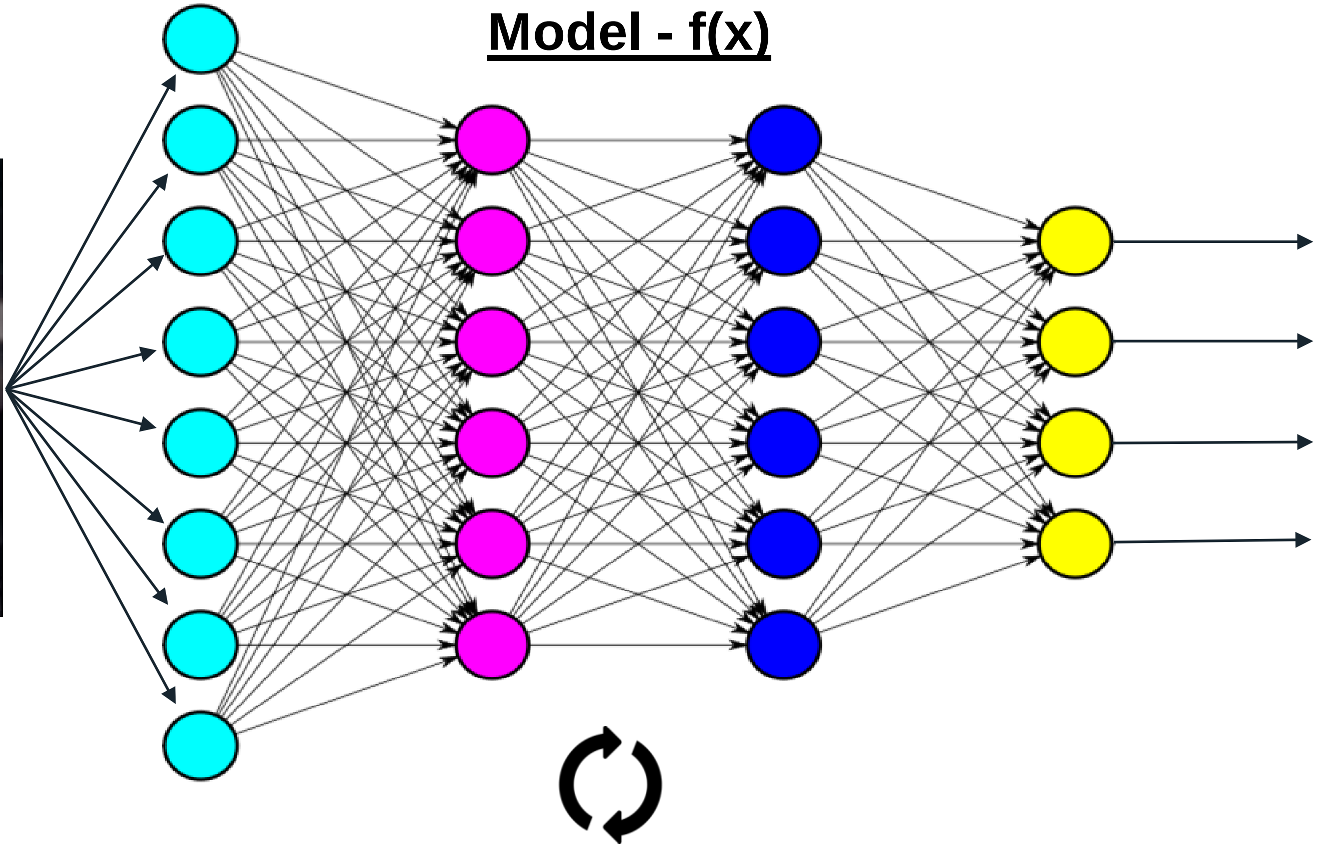
$$\theta - \eta \frac{dL}{d\theta} \rightarrow \theta$$

Intro to Machine Learning

Input - x



Model - f(x)



$$\theta - \eta\frac{dL}{d\theta} \rightarrow \theta$$

Output - y

1% "Dirac"

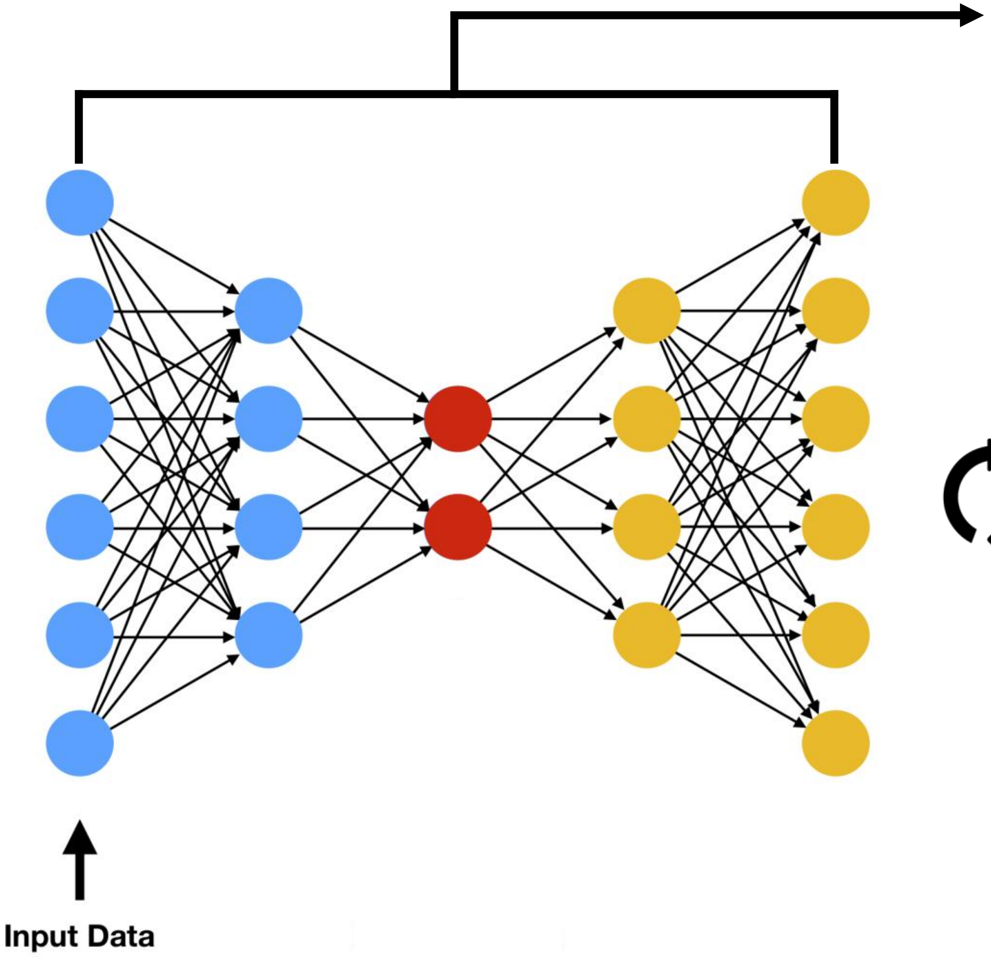
96% "Zichichi"

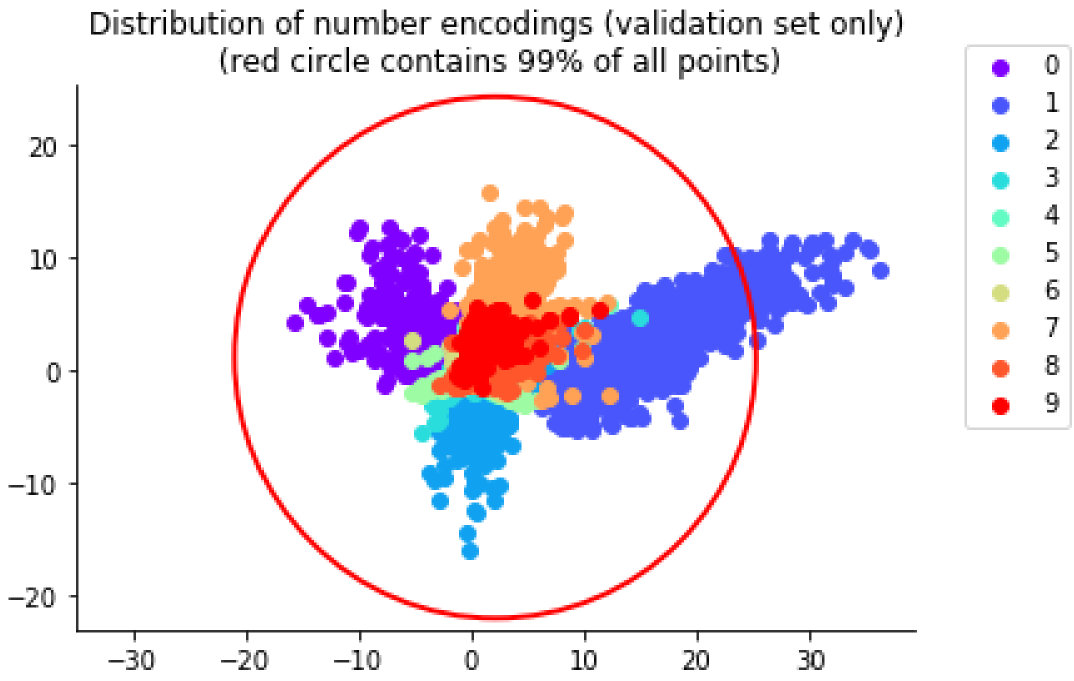
1% "Majorana"

2% "Einstein"

The Autoencoder

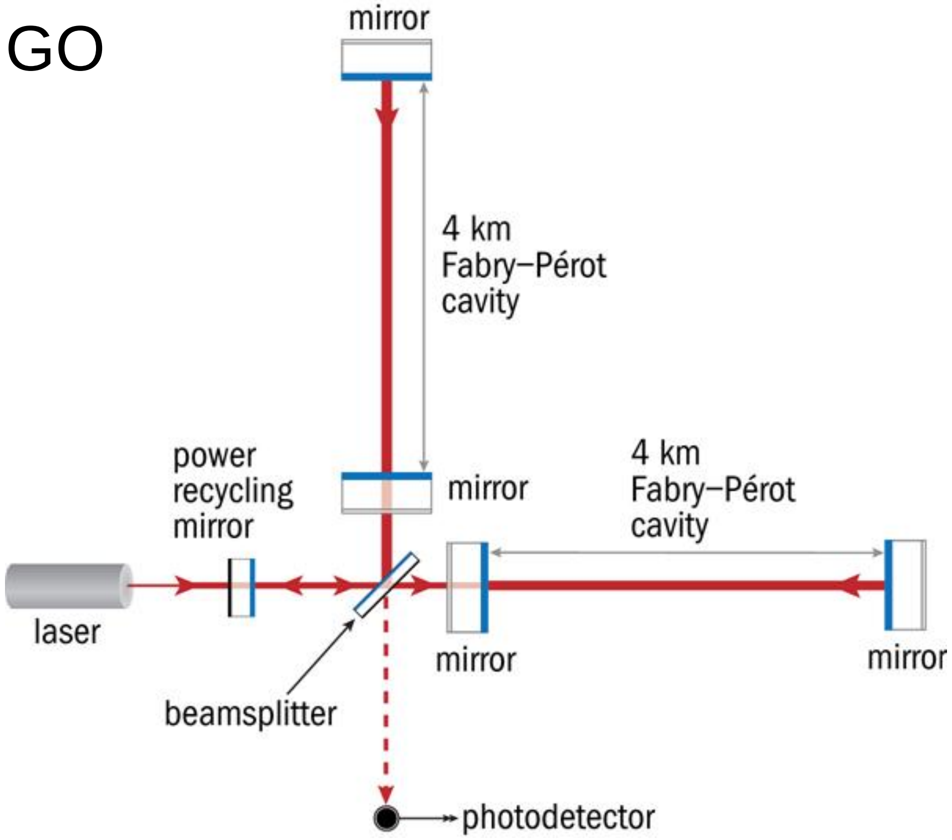
Comparing input and reconstructed data gives a model loss





Introduction to the Step 1

- Detection of gravitational waves (GWs) at LIGO

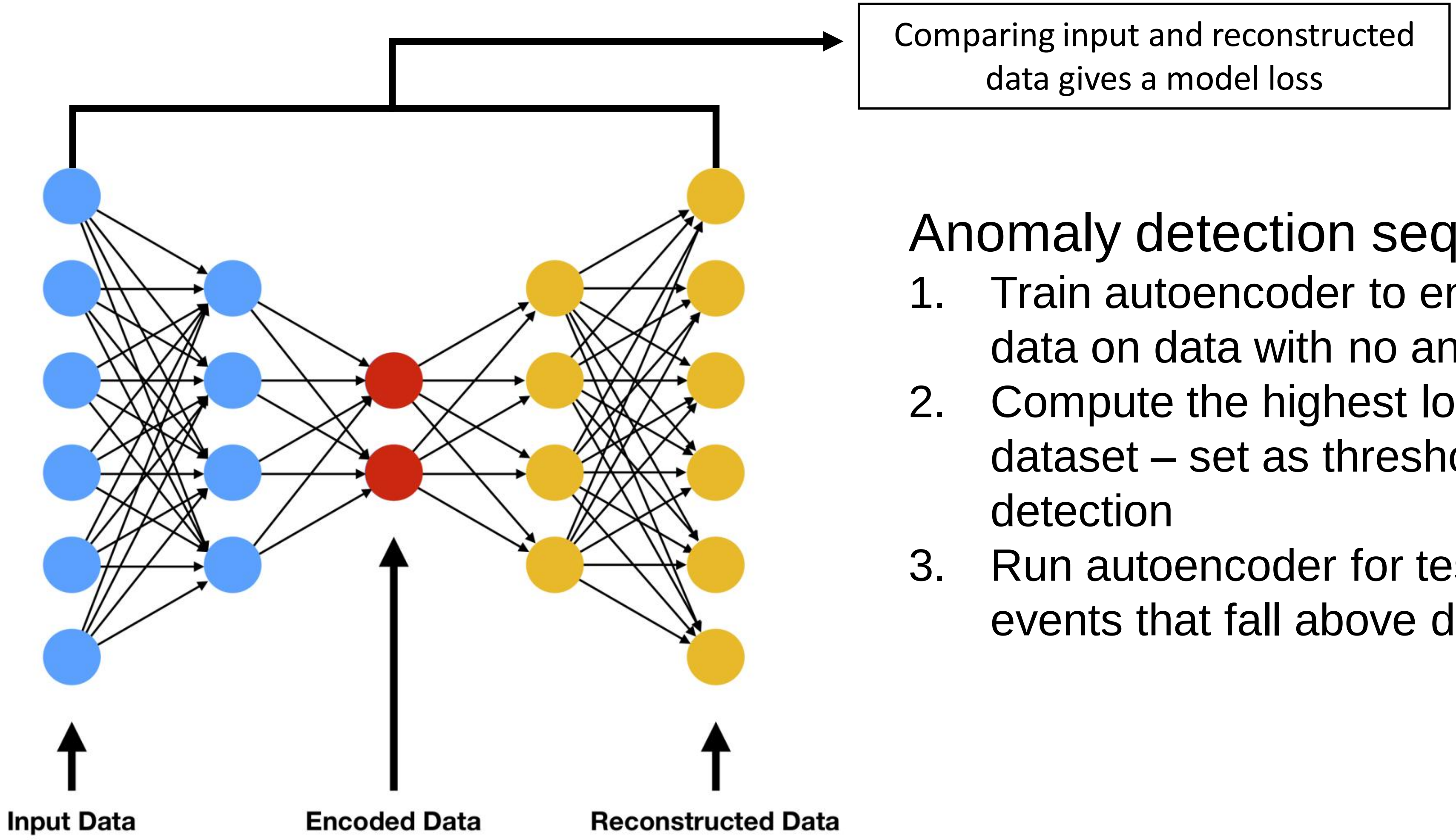


Produces: 1-D time-series strain

Source: Elena Cuoco - Real Time Classifier for transient signals in Gravitational Waves, From raw data to classified triggers

Unsupervised Learning: Detection

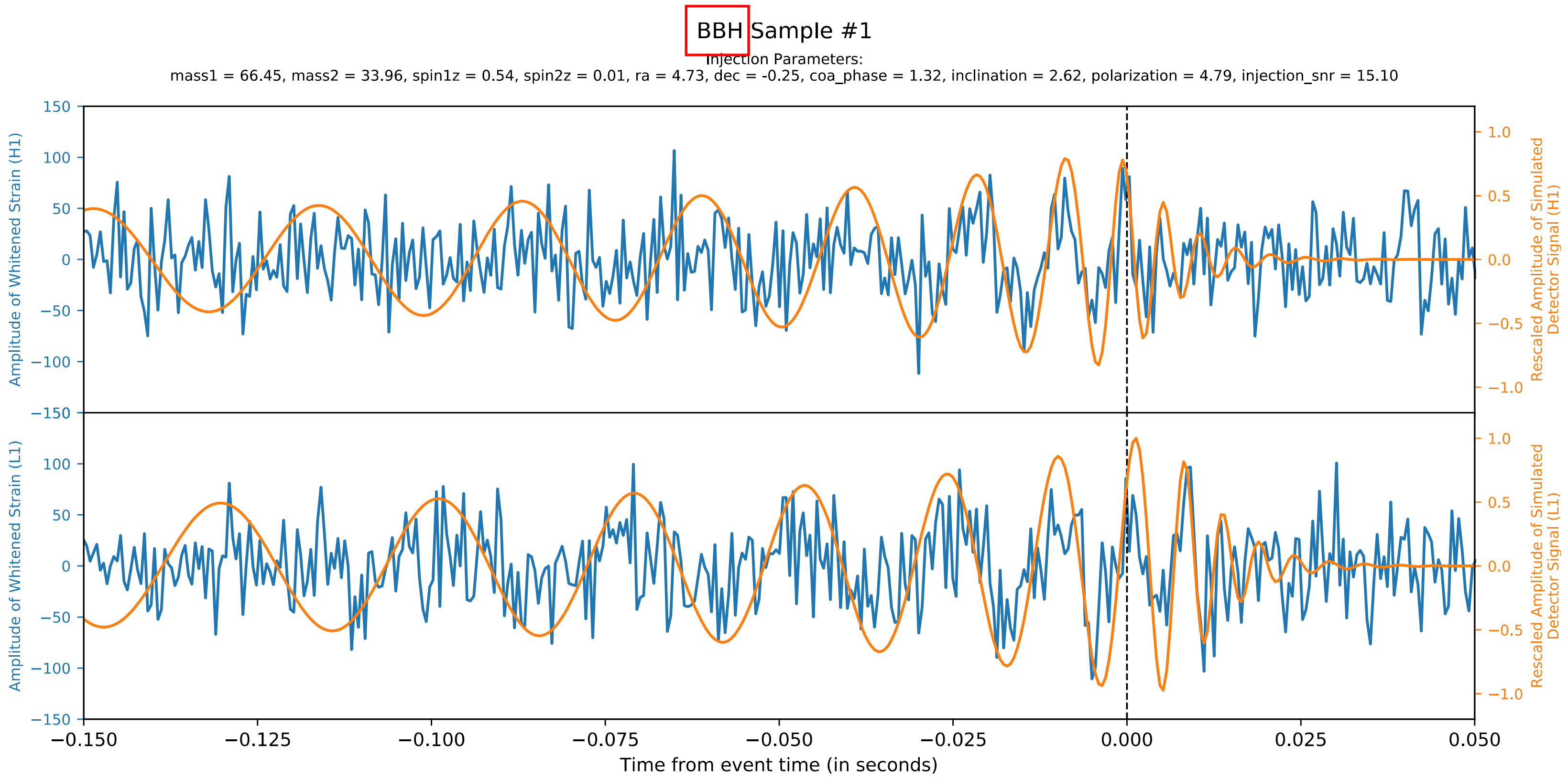
https://github.com/eric-moreno/LSTM-Autoencoder



Anomaly detection sequence:

1. Train autoencoder to encoder and decode data on data with no anomalies.
2. Compute the highest loss on the training dataset – set as threshold for anomalous detection
3. Run autoencoder for test data, identify events that fall above detection threshold

LIGO Dataset



1. Simulates typical detector noise conditions from a PSD
2. Simulates GW waveforms for the following conditions:
   - Binary masses of black hole mergers (BBH) or neutron star mergers (BNS)
   - SNR of 5-20
   - Variable angles in the sky
3. Adds GW strain into noise for signal events
4. Data is whitened, bandpass, and normalized

Source: github.com/timothygebhard/ggwd, https://www.gw-openscience.org/data/

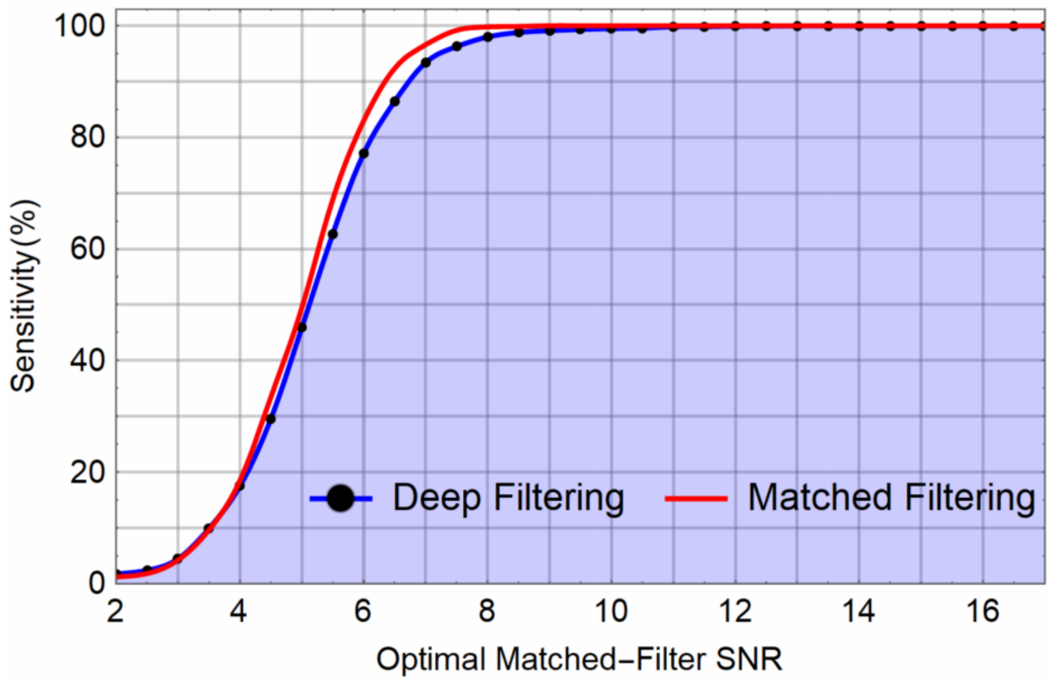
Currently used methods

Matched Filtering

- **Current method** used by LIGO
- Compares incoming GW data to bank of simulated waveforms
- Can only identify GWs that are available in GW banks (no exotic events)

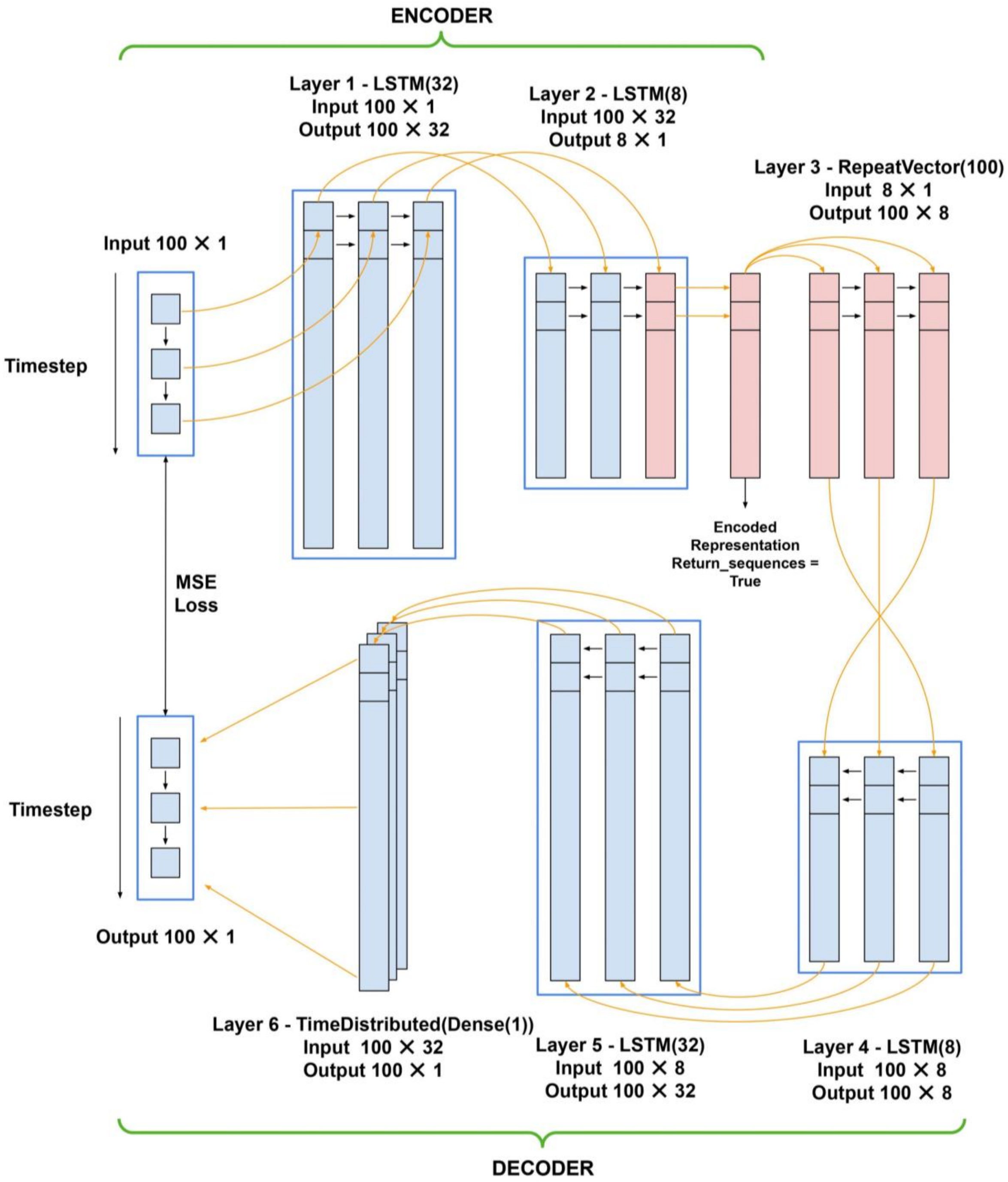
Deep Filtering

- Convolutional Neural Networks (CNNs)
- Take time-series inputs, can determine detections and estimate parameters of events
- Still can miss events that aren't included in training set



Source: https://arxiv.org/pdf/1711.03121.pdf, https://arxiv.org/pdf/2005.06534.pdf

LSTM AE Architecture

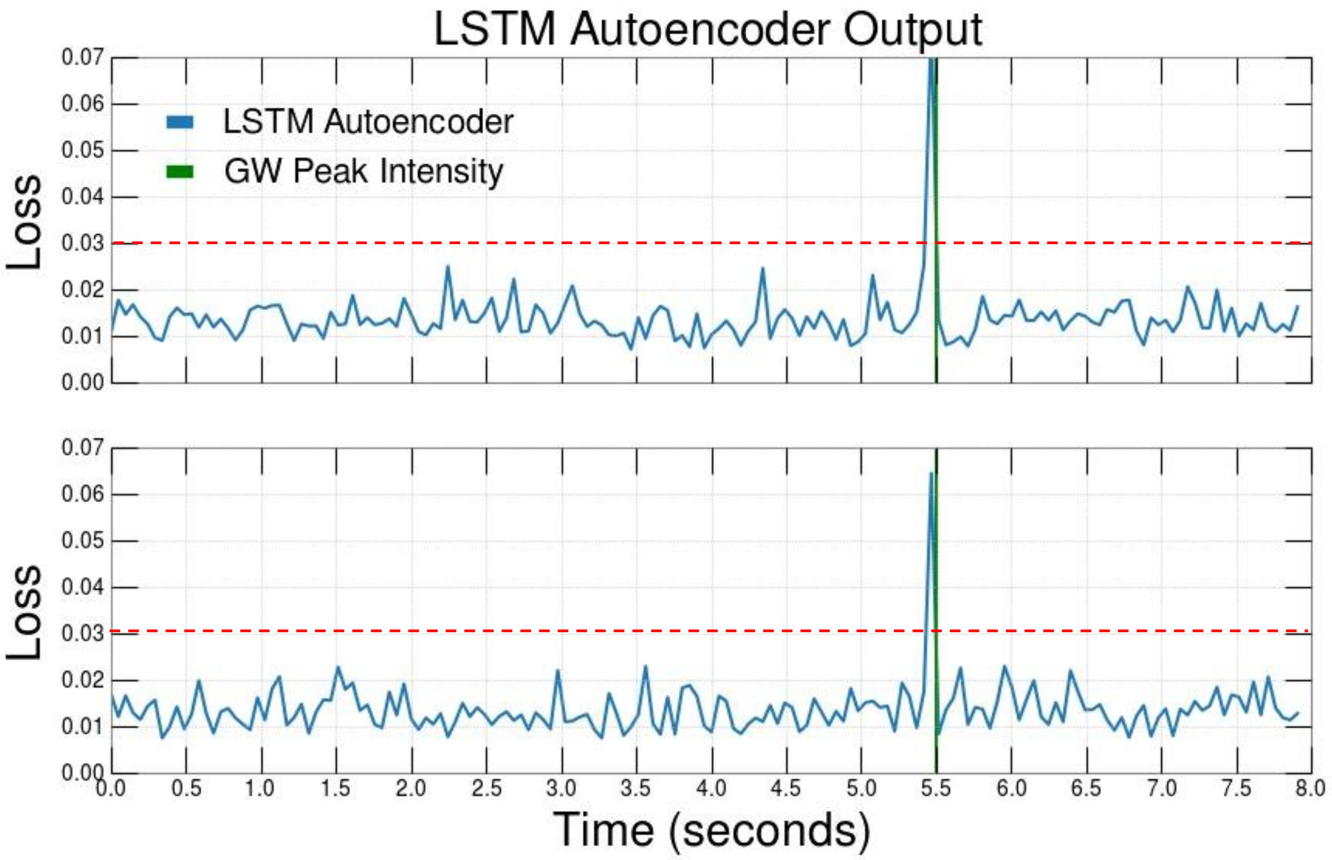


https://github.com/eric-moreno/LSTM-Autoencoder

Event Loss with Autoencoders

2107.12698

- LSTM AE evaluated BBH and BNS events yields promising results
- **Red dotted line represents detection threshold** which can be determined according to FPR
- During training, **AE never receives information about any GW (signal) -> Source Agnostic**

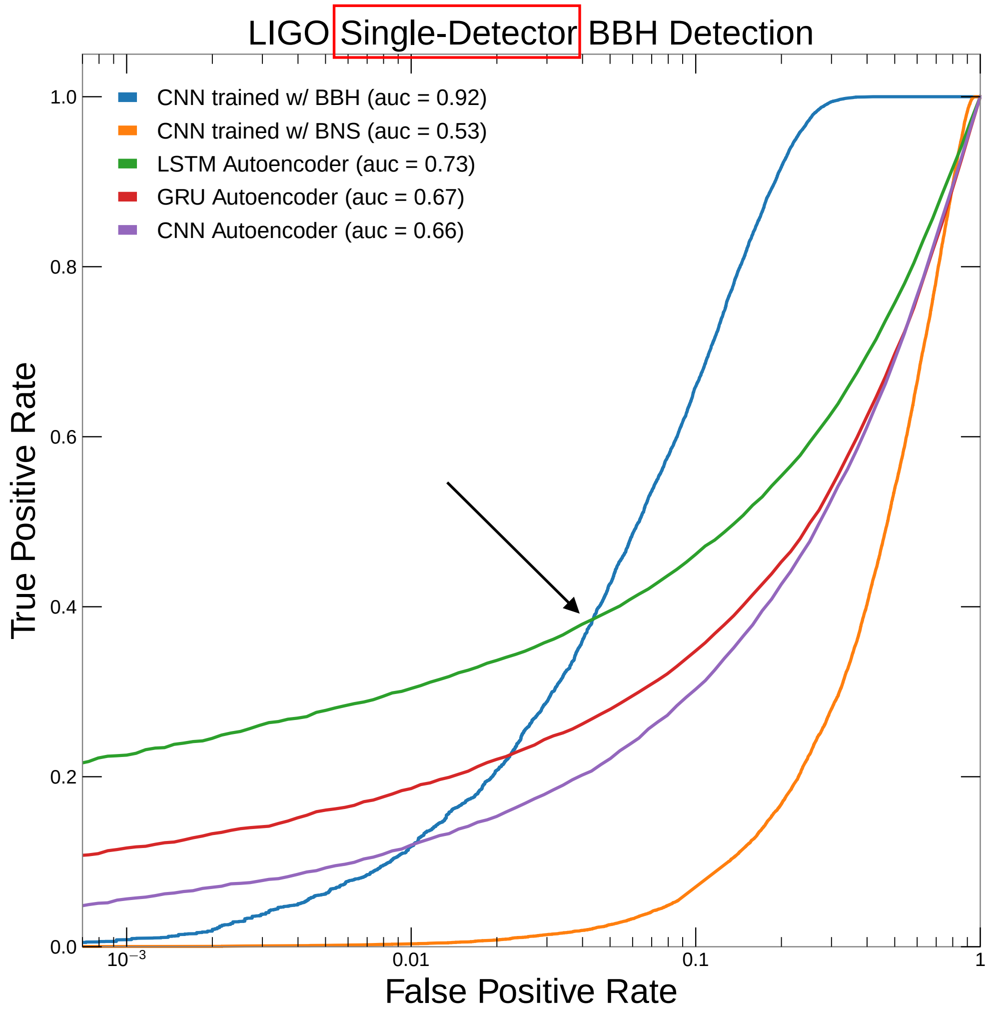


Supervised vs Unsupervised BBH

2107.12698

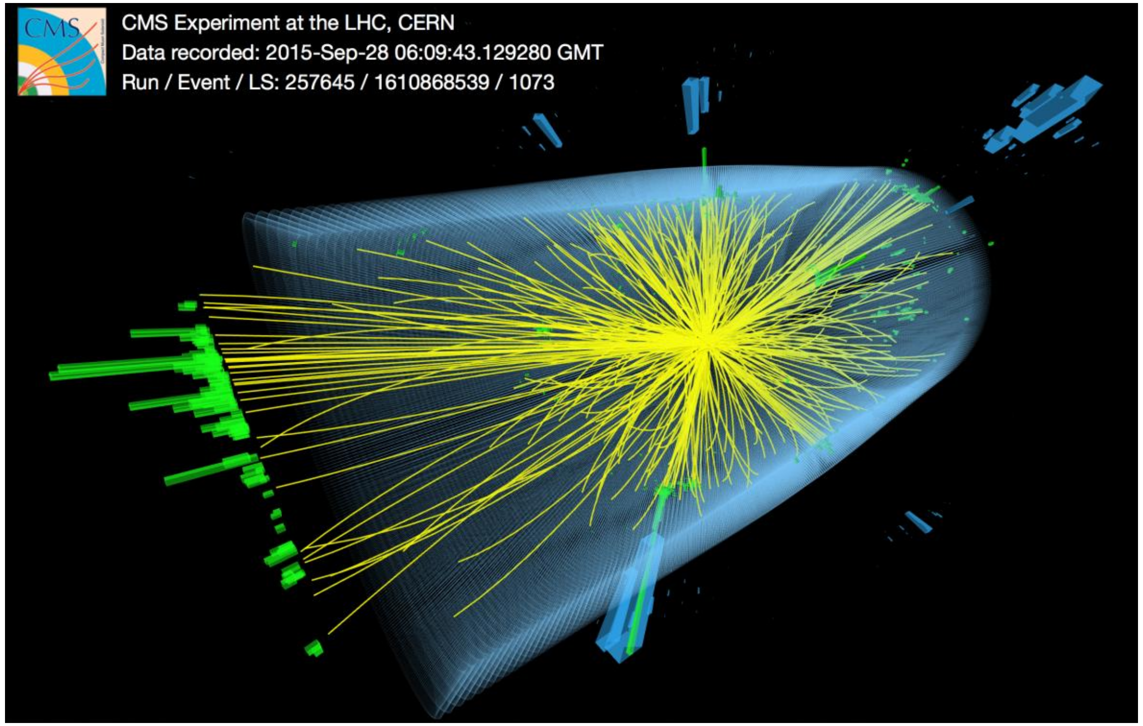
- BBH generated from SEOBNRv4 Approximant
- High mass BH (10–80+ solar masses) produce large amplitude events
- Both autoencoders perform better than supervised models generalized from BNS data
- **Outperforms supervised methods (trained on equivalent length data) at below FPR = 0.04**

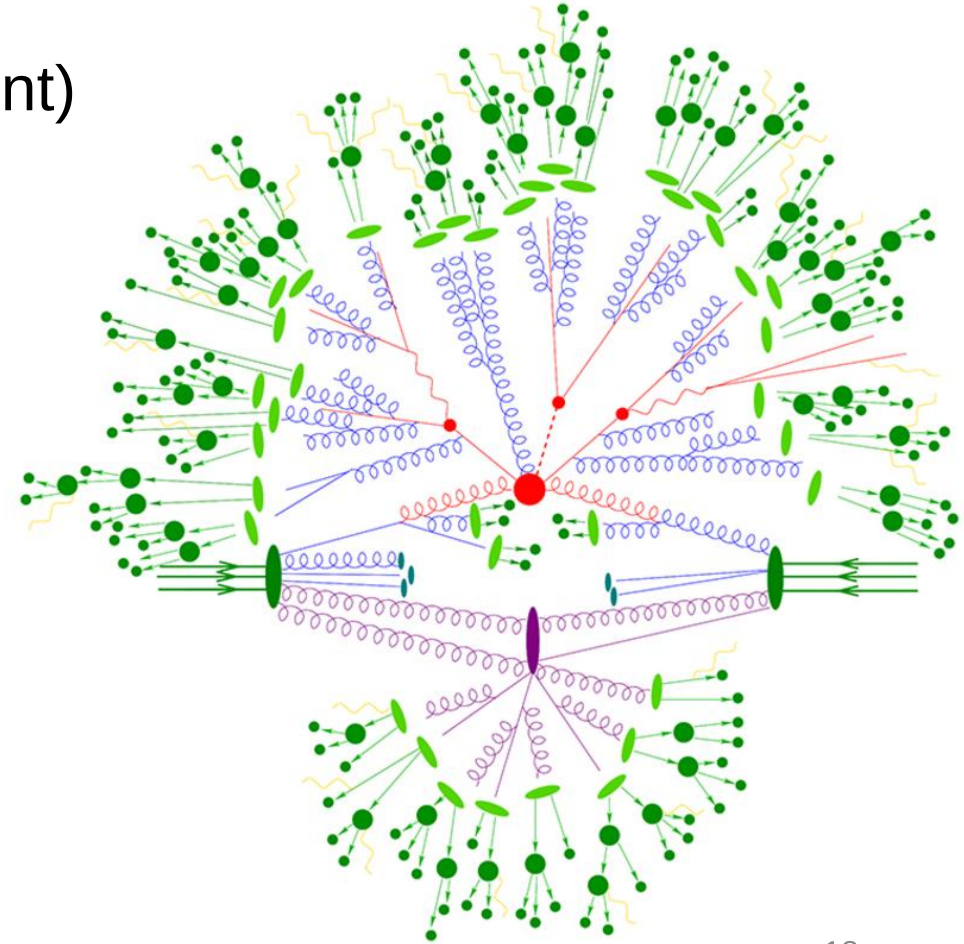
AE can be used for:
- Triggering on high SNR rare events
- Glitch detection within LIGO apparatus
    - Glitches are hard to simulate and more easily identifiable with AE



Introduction to the Step 2

- Anomalous events at the LHC (CMS experiment)



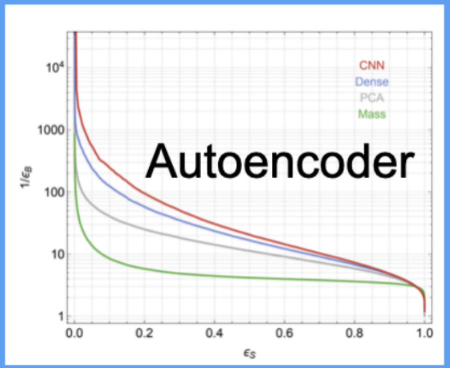


Produces: particle tracks and jet information

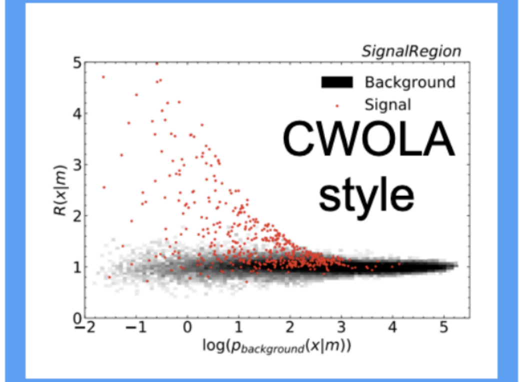
Source: CMS Experiment, CERN

Semi-Supervised Learning

Prior Free → Fully Supervised



Knowledge of Background

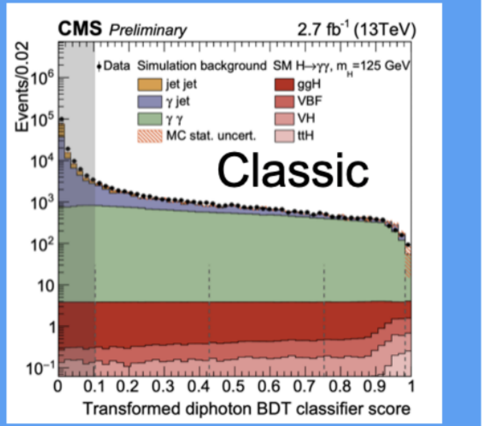


Signal in (Mass) Window

Semi-Supervised Approaches

QUAK

If it quaks like a duck?



Fully Supervised

Step 1

[2107.12698](2107.12698)

Step 2

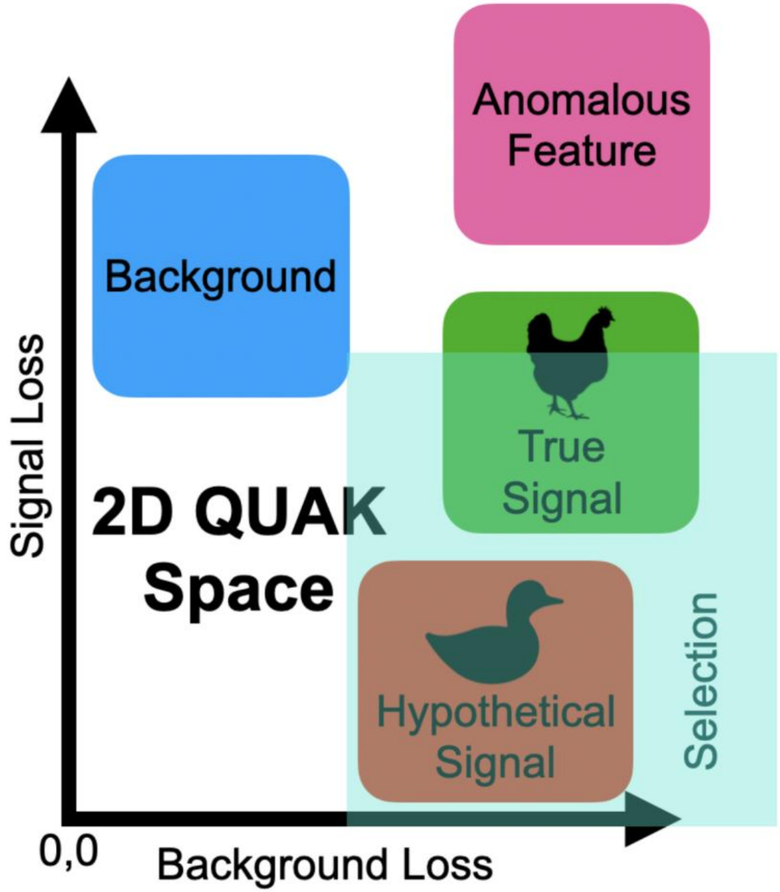
[2011.03550](2011.03550)

Step 0

[1909.12285](1909.12285)
[1908.05318](1908.05318)

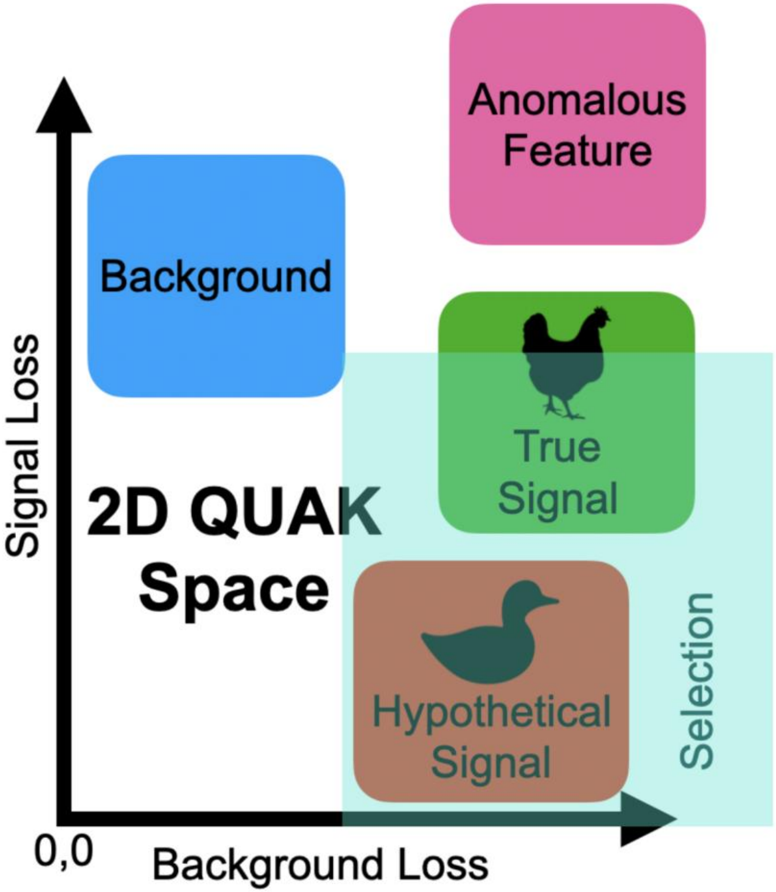
Quasi-Anomalous Knowledge - QUAK

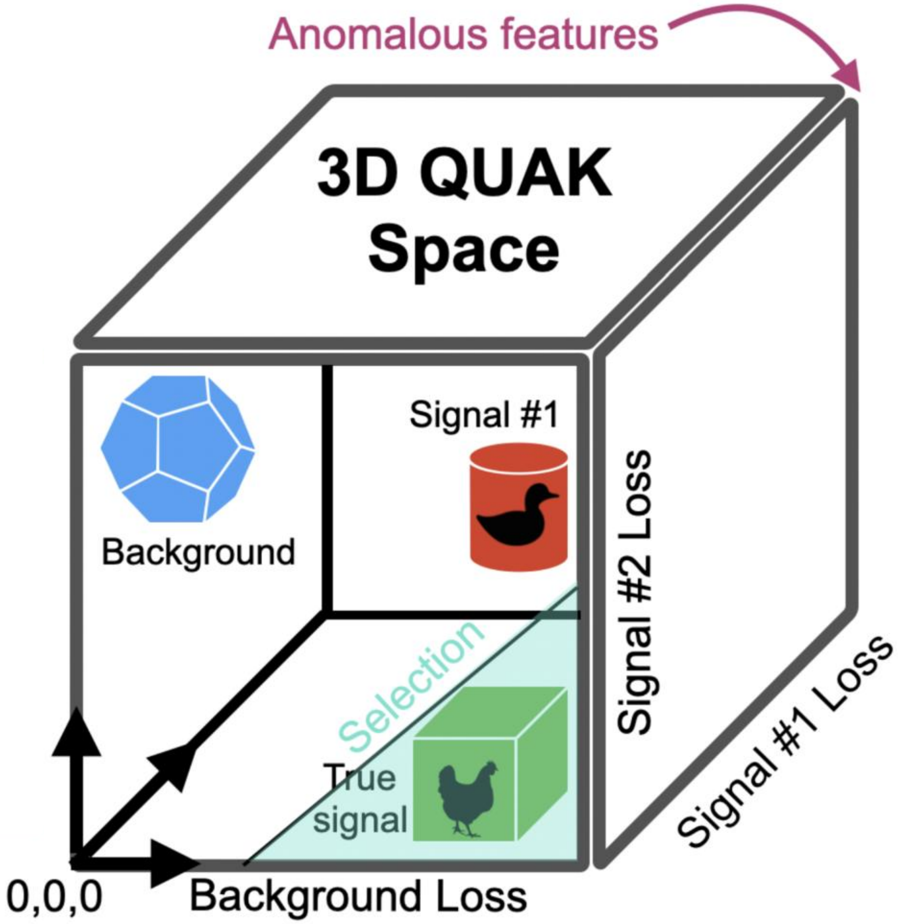
2011.03550



Quasi-Anomalous Knowledge - QUAK

2011.03550

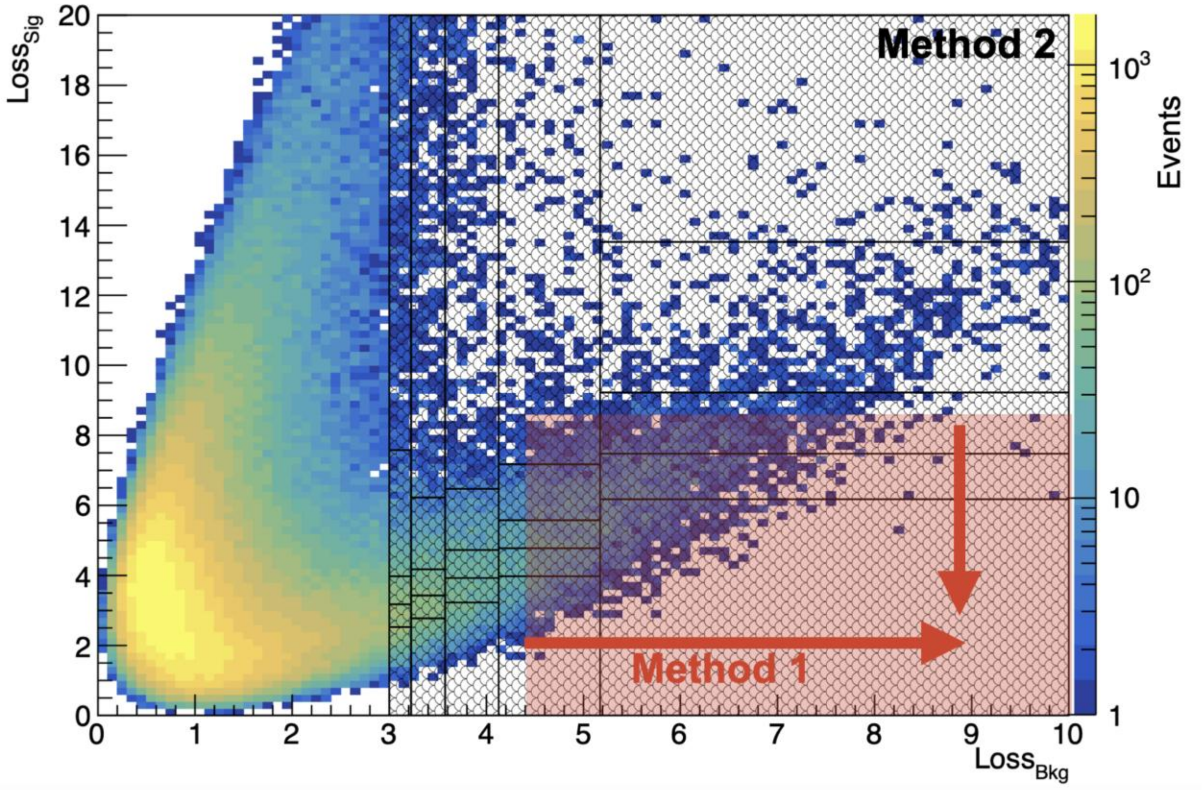




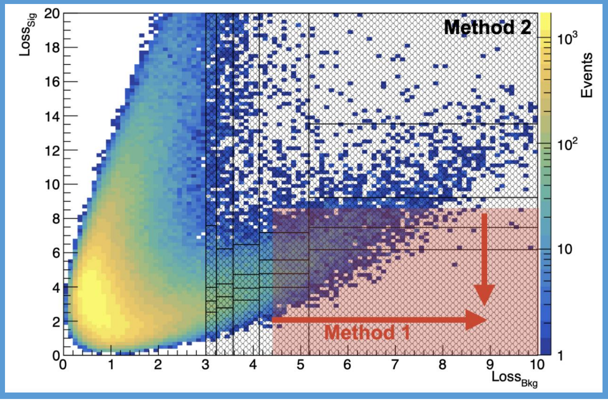
Example: LHC Olympics

2011.03550

- Signal Dataset: W' → XY
  - W' = 3.5 TeV
  - X = 500 GeV
  - Y = 100 GeV
- Background Dataset: 1M simulated QCD dijet events
  - Hidden signal: 900-event W' resonance
  - W' = 3.8 TeV
  - X = 732 GeV
  - Y = 378 GeV

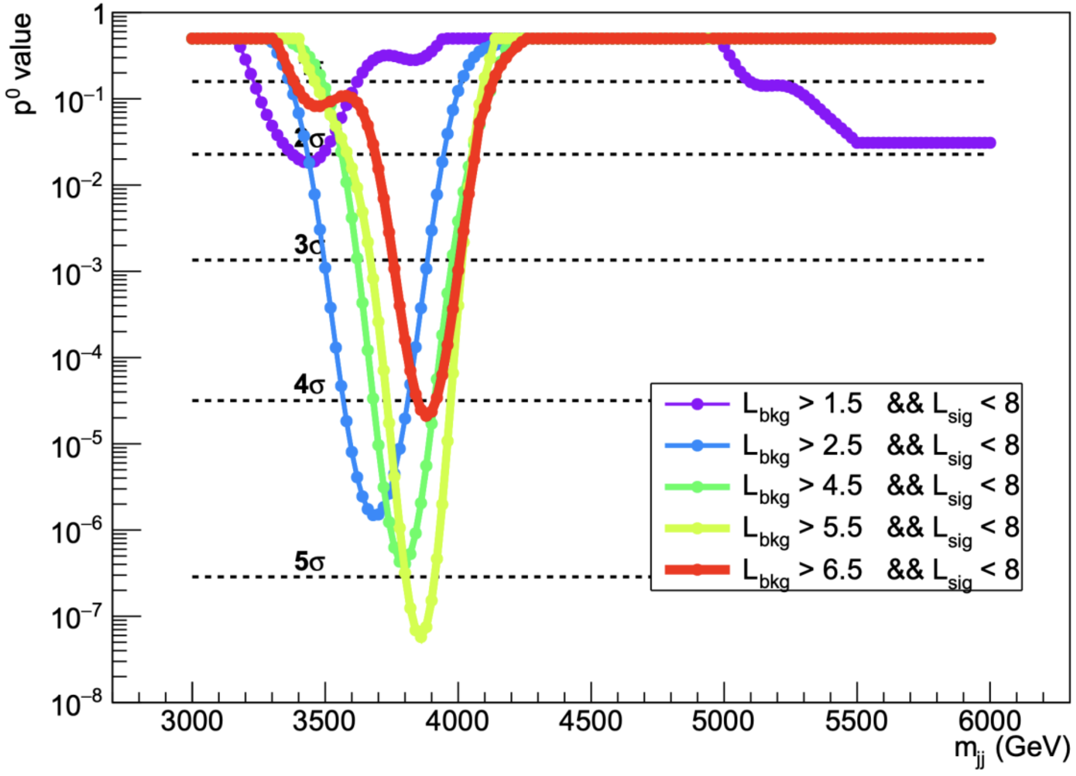


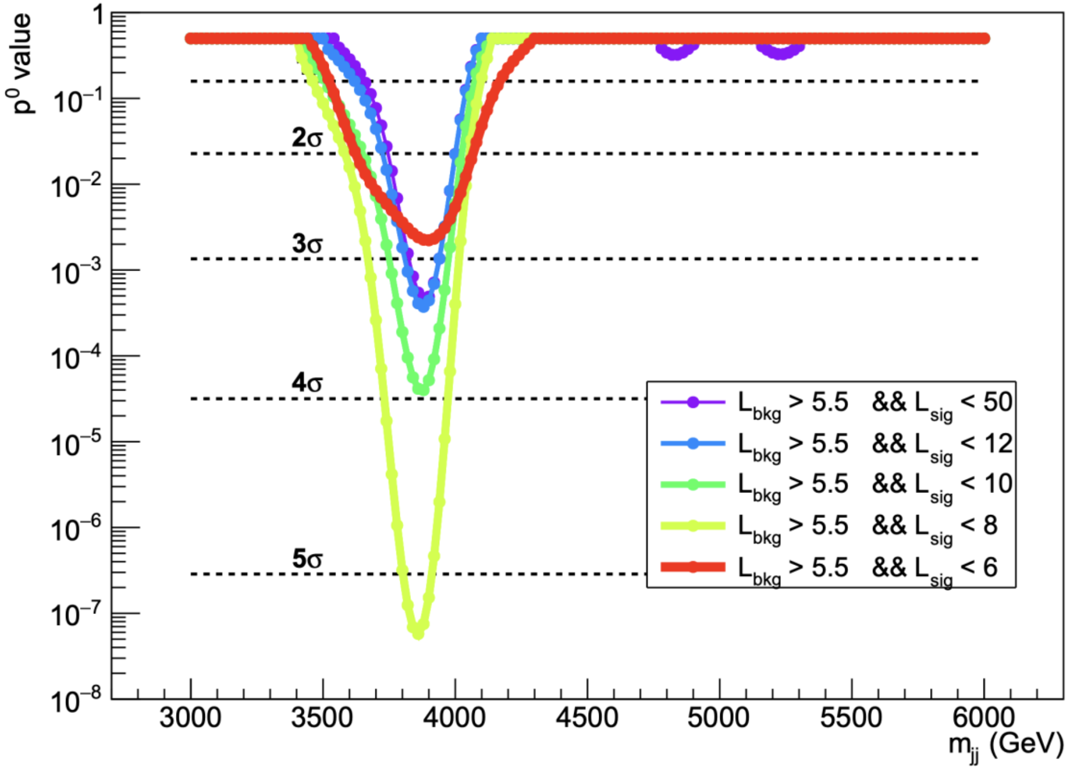
LHC Olympics – Method 1



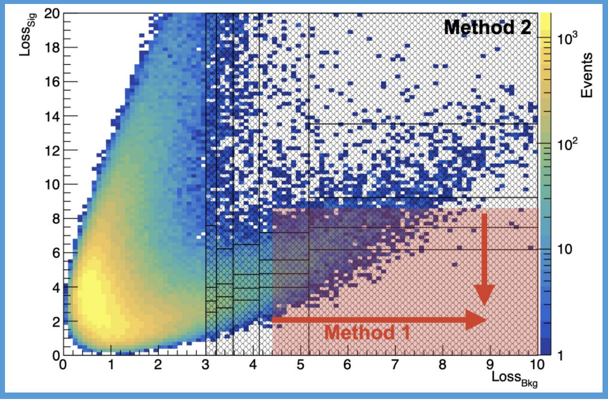
2011.03550

Method 1: Iteratively vary a selection on the signal loss and background loss and select regions of low signal loss and high background loss. Biased analysis method!



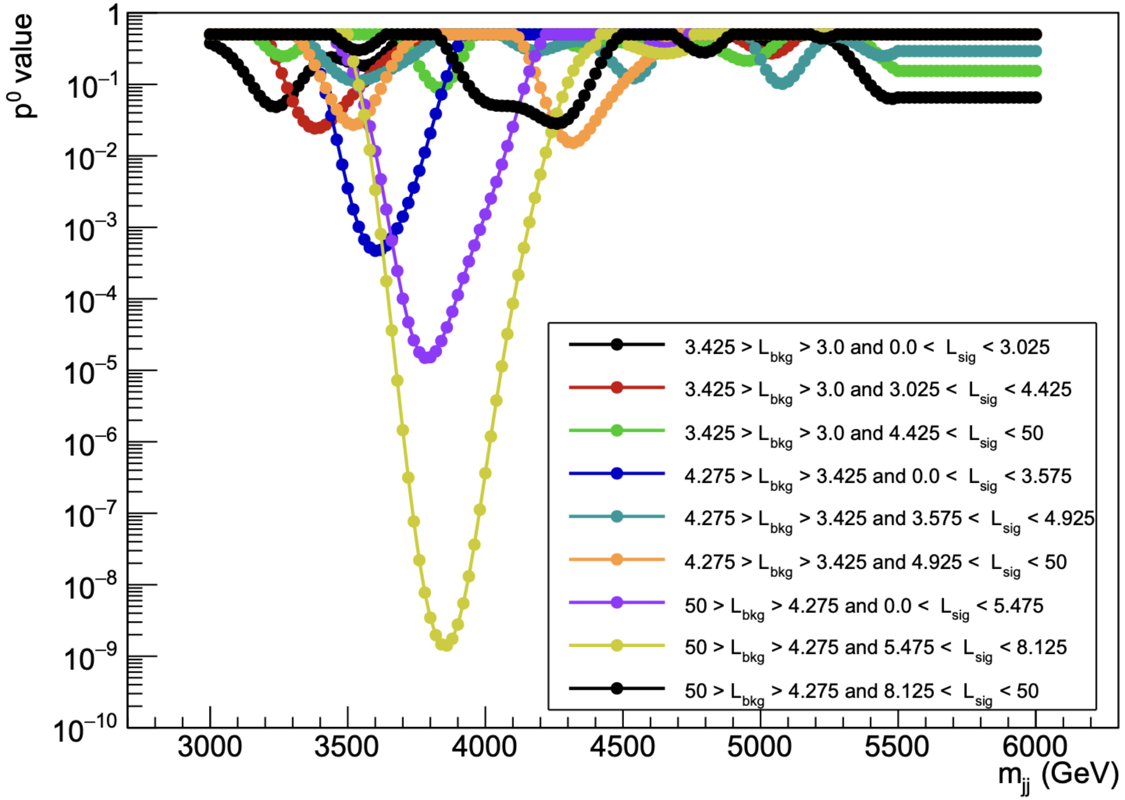


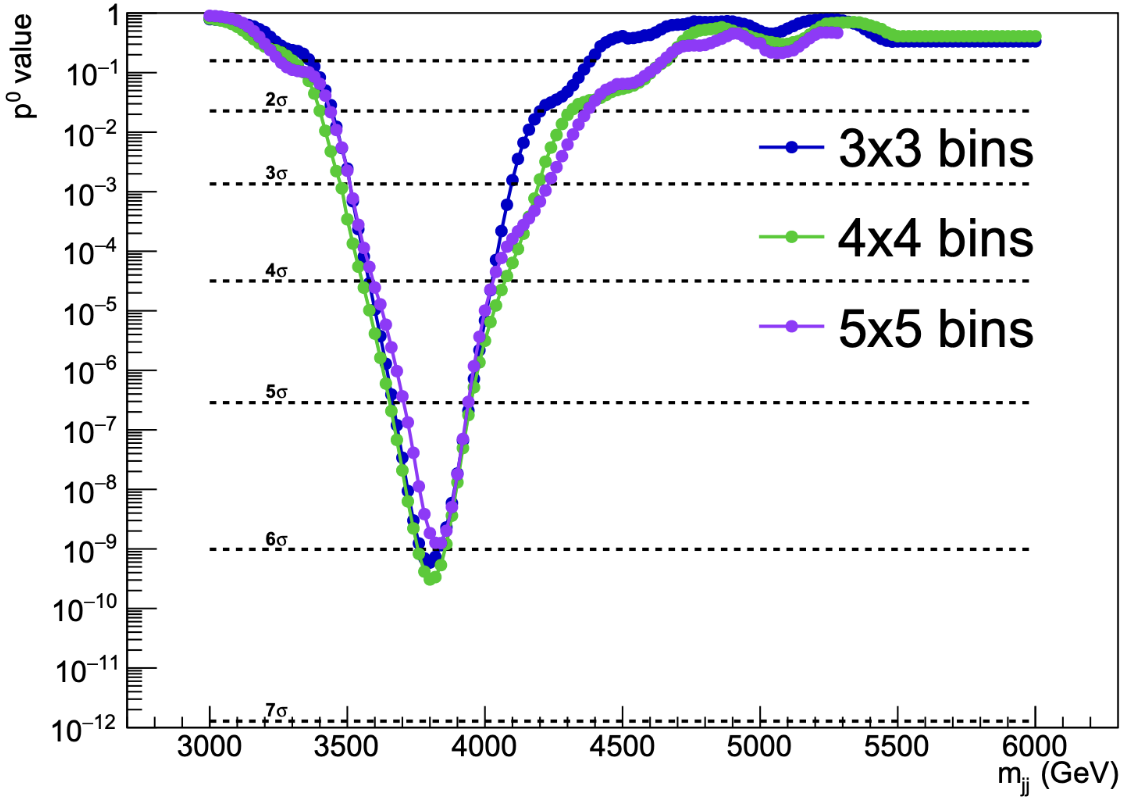
LHC Olympics – Method 2



2011.03550

Method 2: Separate the events by the black shaded boxes shown corresponding roughly to a uniform populations of events within each shaded region

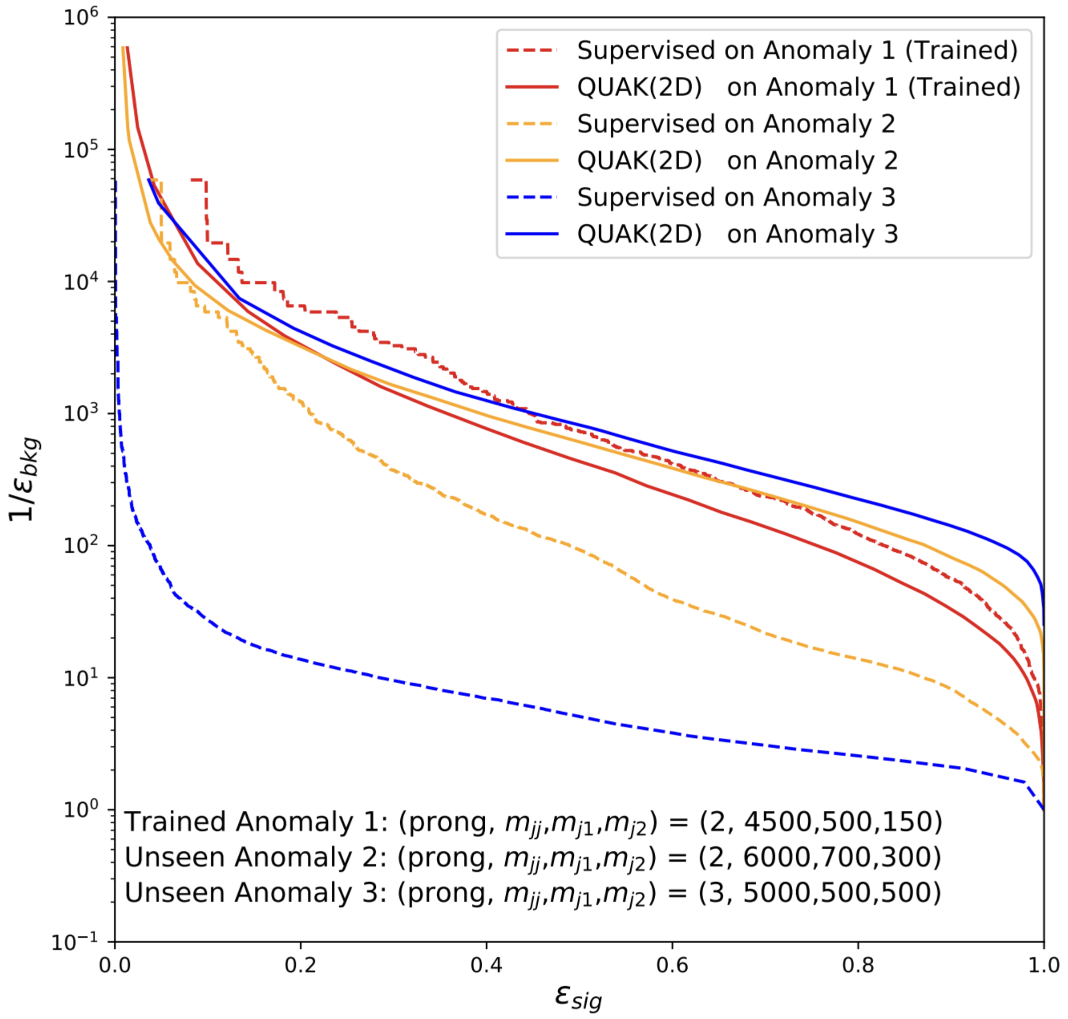




QUAK Anomaly Detection

2011.03550

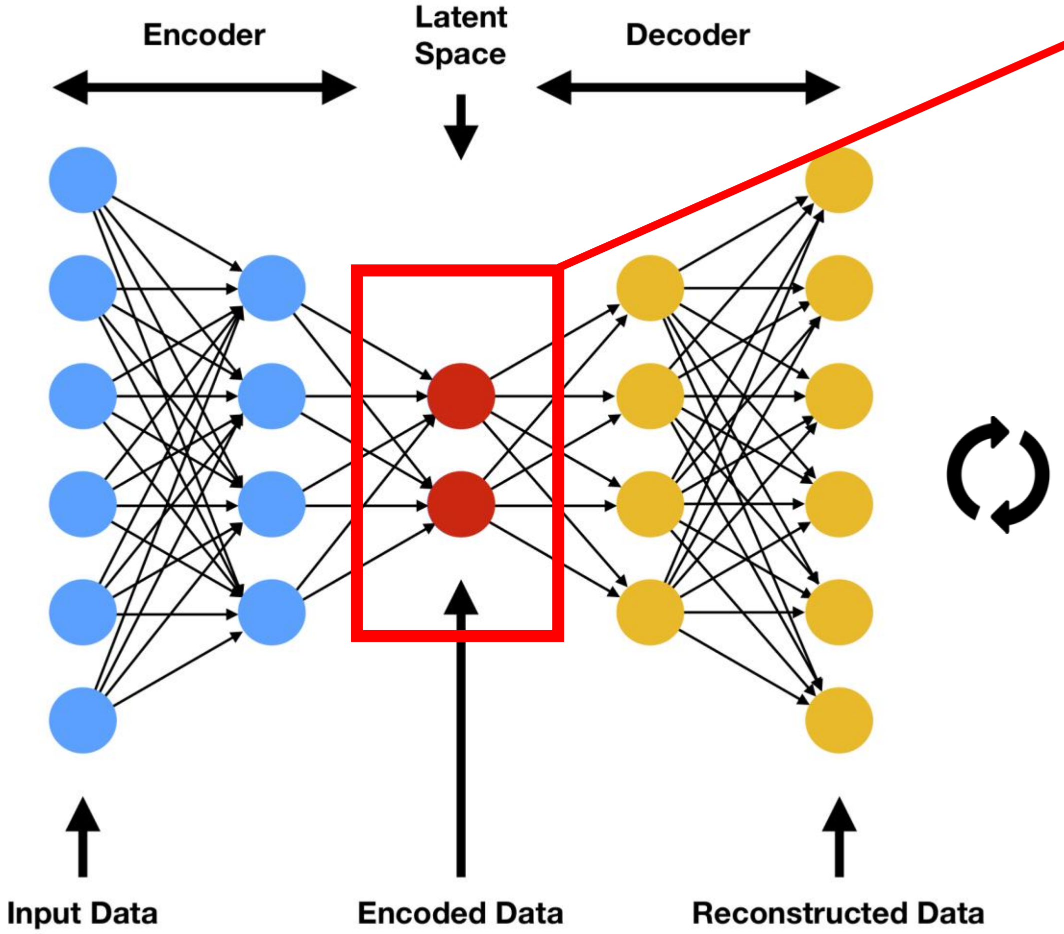
QUAK allows generalization capability beyond supervised algorithms! Can search for events that are similar but not the same as a hypothetical signal.



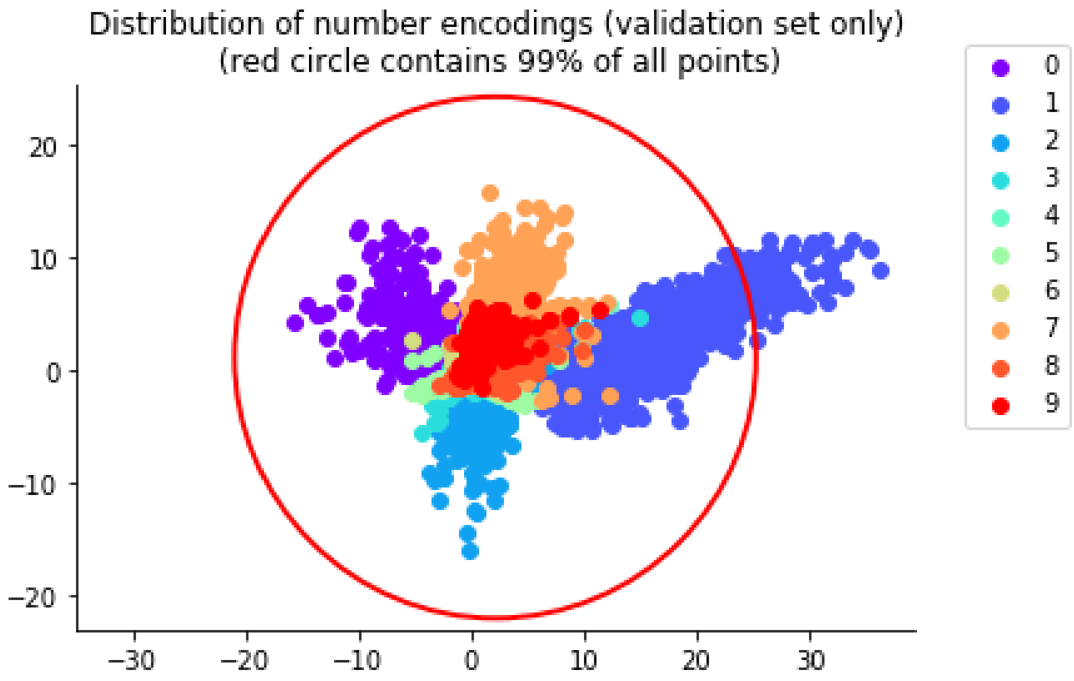
Introduction to the Step 3

Remember how I showed you this slide…

The Autoencoder

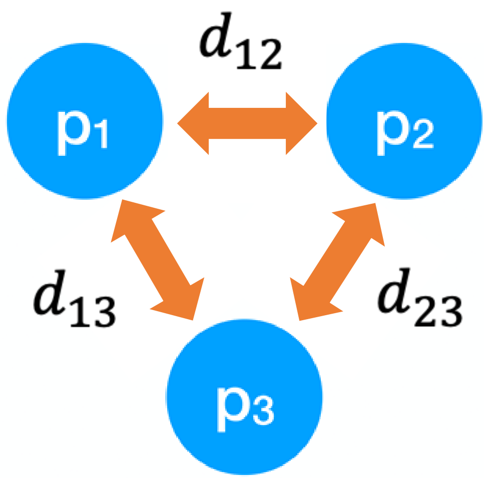


What's happening here?



Introduction to the Step 3

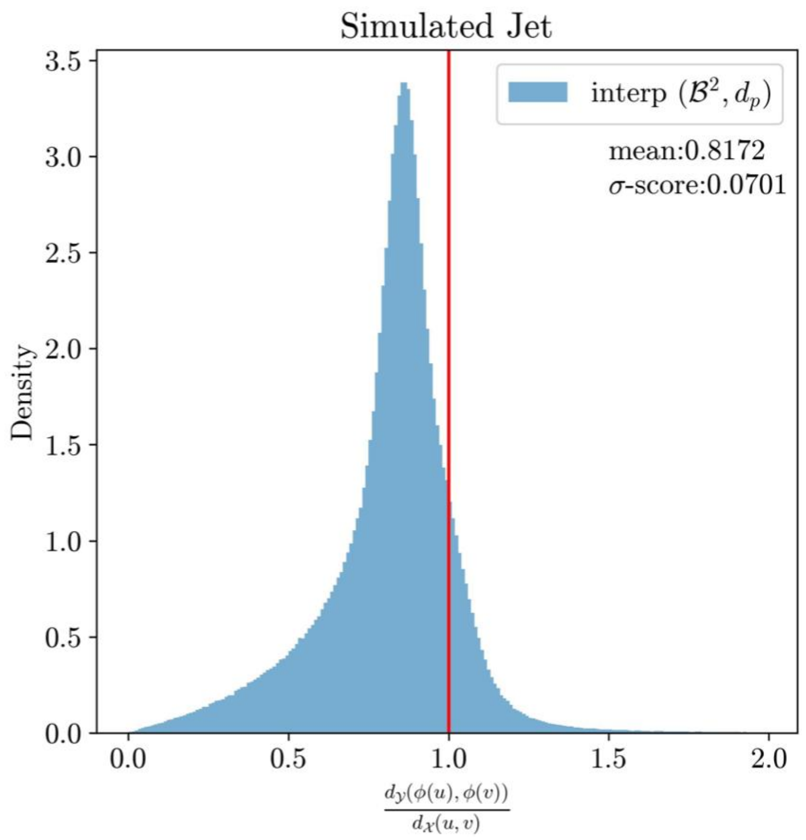
- The autoencoder is constructing a latent space according to some distance metric it has designed.
- What if we take over this metric? We can make the most out of metric space properties of collider events.
- Distance metrics include:
  - Euclidean
  - Hyperbolic
  - Energy Movers Distance – For HEP
  - Power spectral distance? – For LIGO
  - Appropriate metrics can be tailored to the domain!

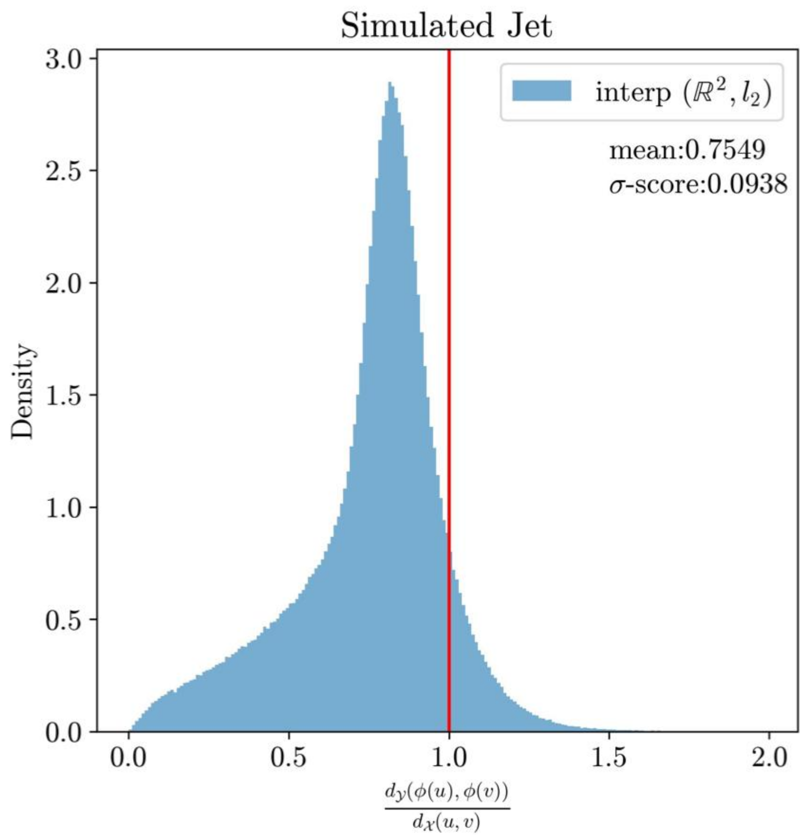


Does it work?

Idea from Sangeon Park et. al 1804.03329

Hyperbolic spaces: Better for handling graph / tree structured data, biological sequences

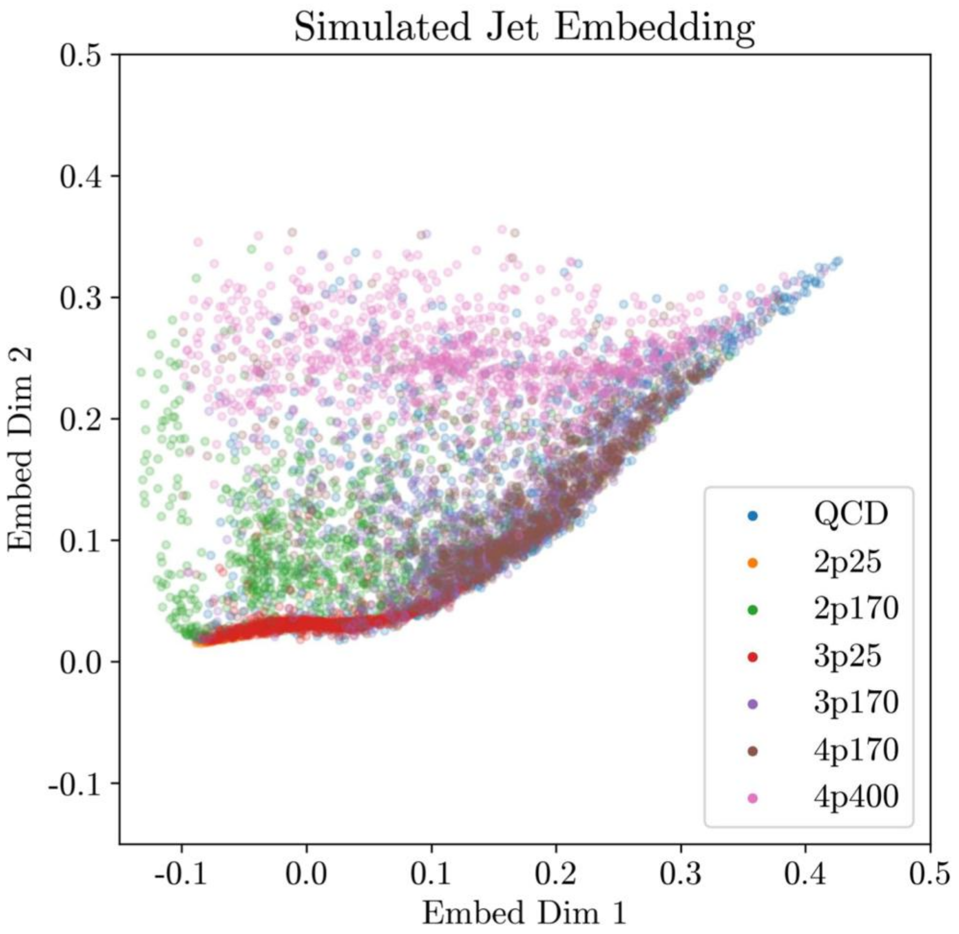


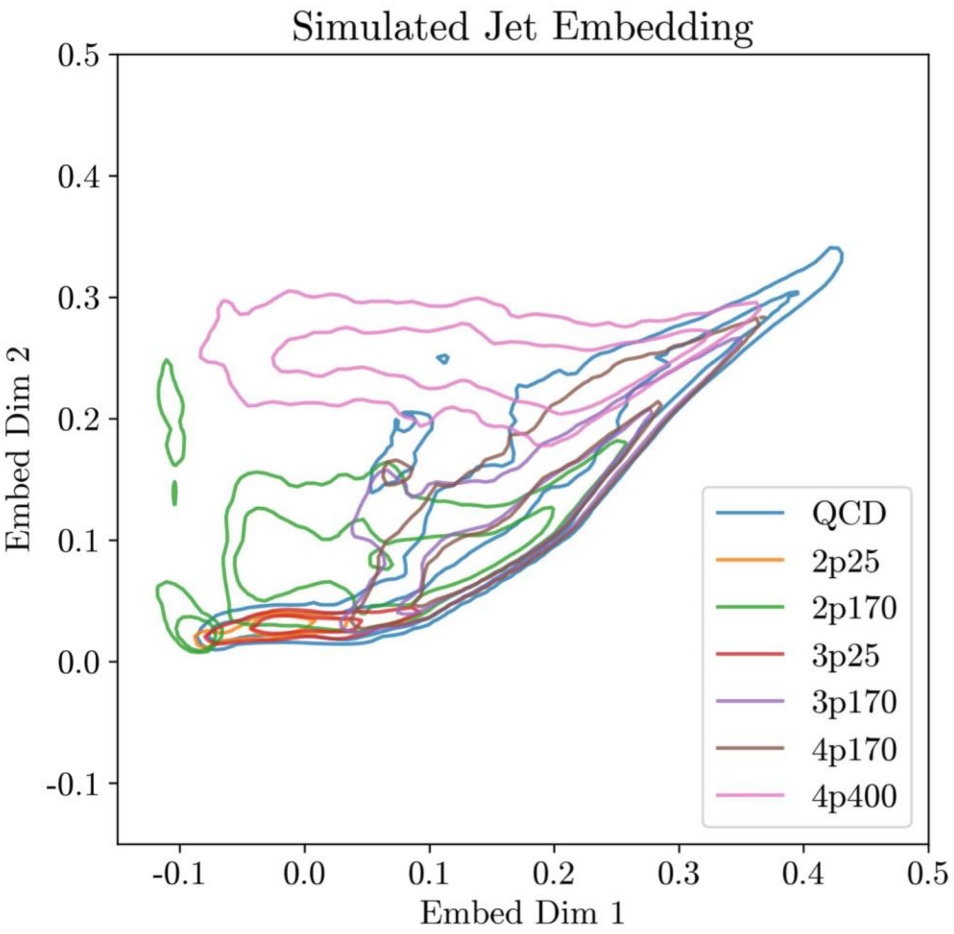


Euclidean spaces: Most common choice, easy to calculate volume

Embedding results – new latent space!

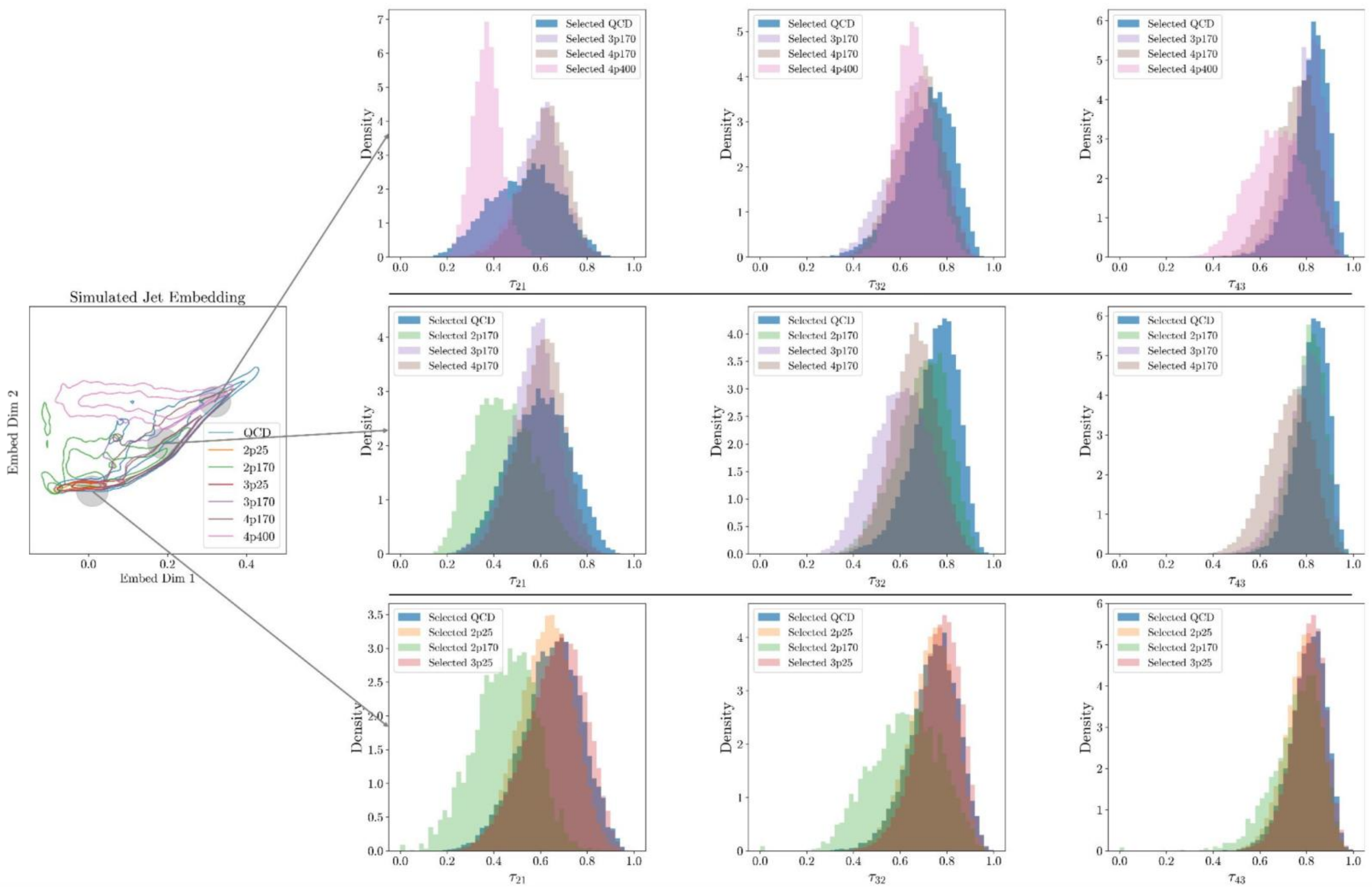
Idea from Sangeon Park et. al





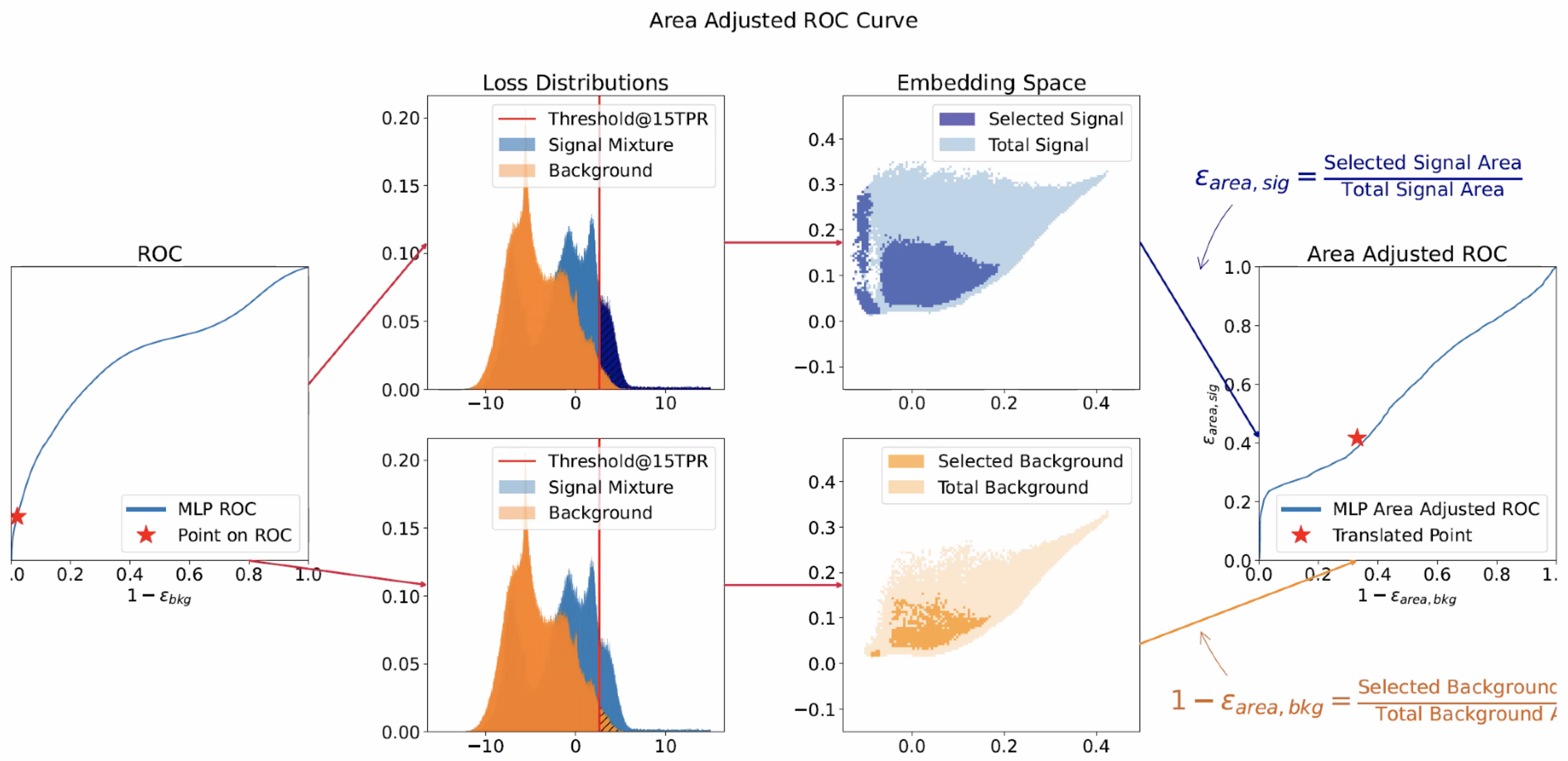
What is it learning?

Idea from Sangeon Park et. al



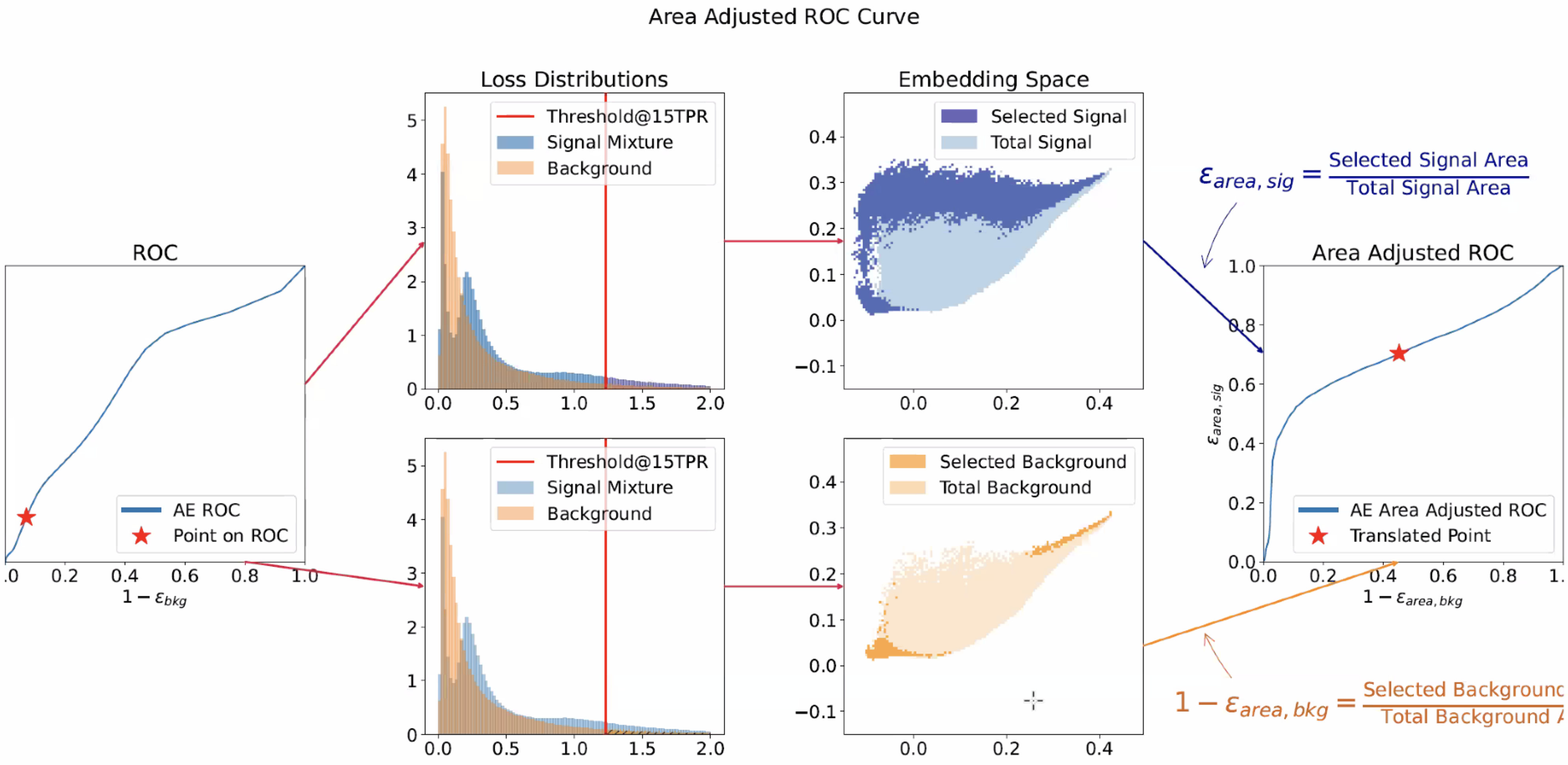
Application – searching algorithmic coverage

Idea from Sangeon Park et. al



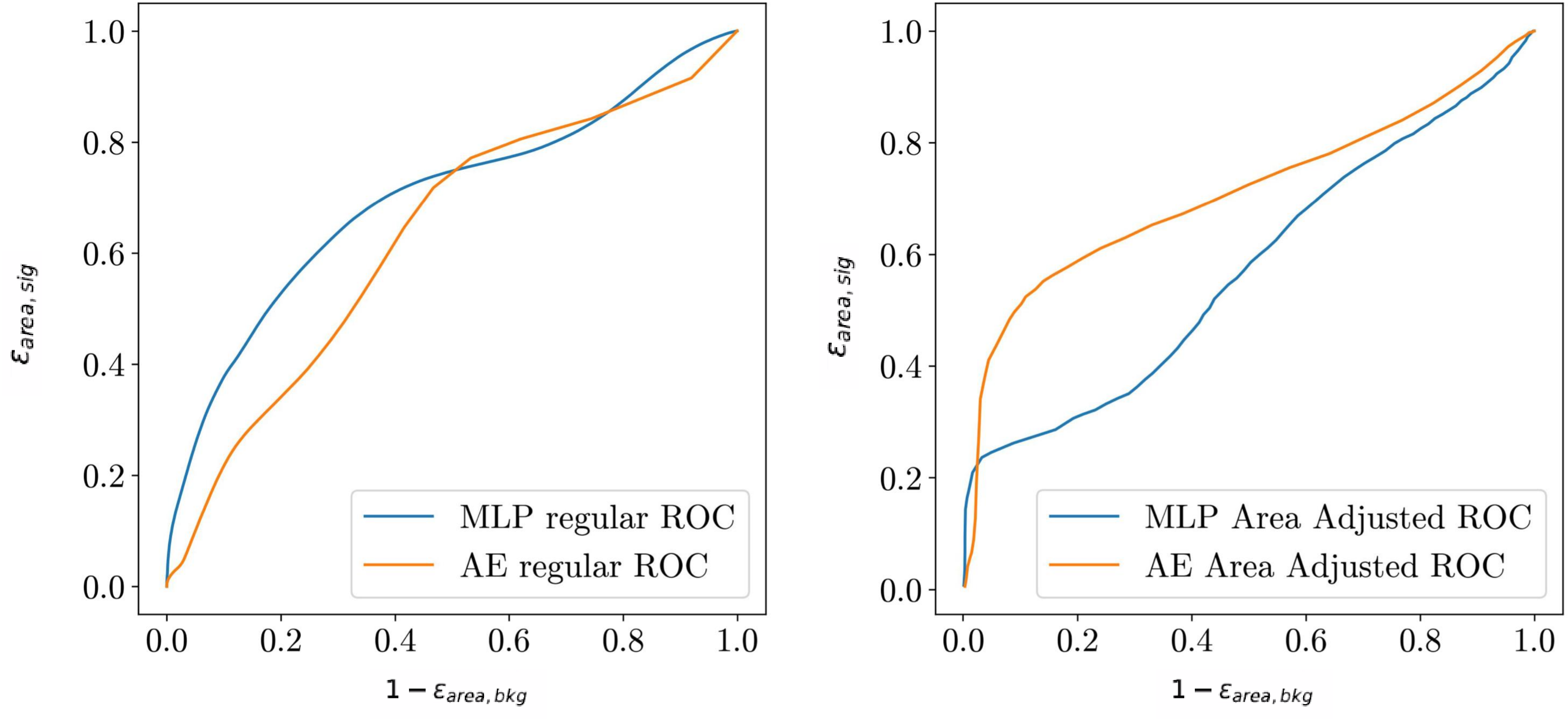
Application – searching algorithmic coverage

Idea from Sangeon Park et. al

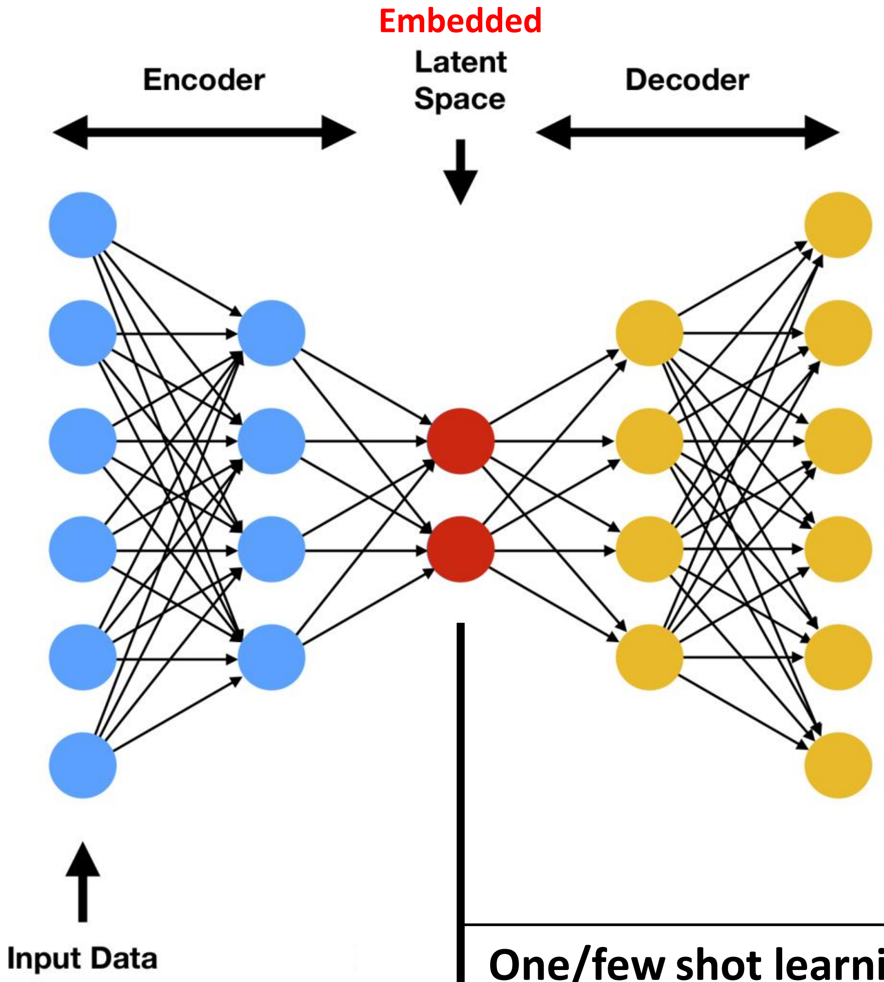


Application – searching algorithmic coverage

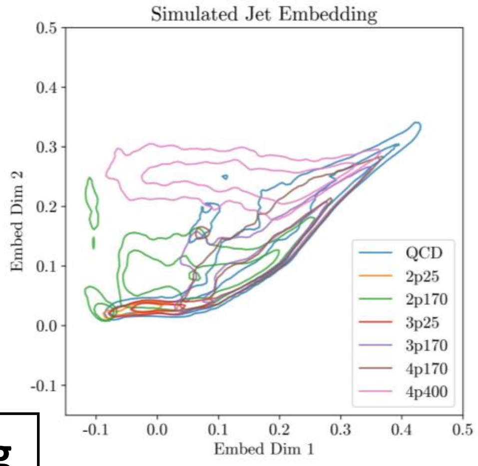
Idea from Sangeon Park et. al



Application – few shot learning



State of the art: Large language models like GPT3 (175 Billion parameter AE model) learn using few-shot learning on their latent spaces [2005.14165](2005.14165)



One/few shot learning

Classification

Conclusion

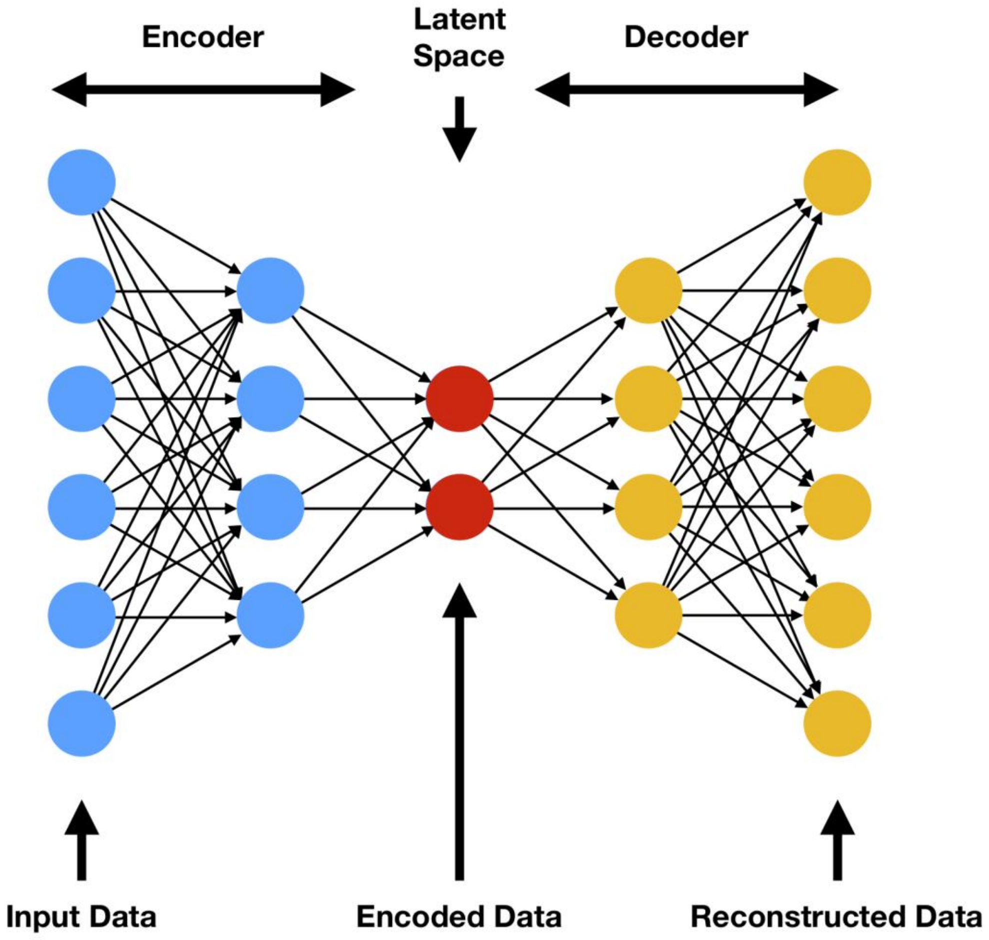
We fell down the rabbit hole and learned that:

1. Anomaly detection is an unsupervised learning tool that can detect exotic events in a variety of different settings
2. The QUAK-space furthers these concepts, allowing for more complex selections and analysis
3. The holy grail of anomaly detection (in my opinion) lies in embedded latent spaces that contain useful physics metrics. This can be used to **test the coverage space** of different analyses, **design analyses** that cover a new phase space, or **perform a classification**!

Thank you for your attention!

Questions?

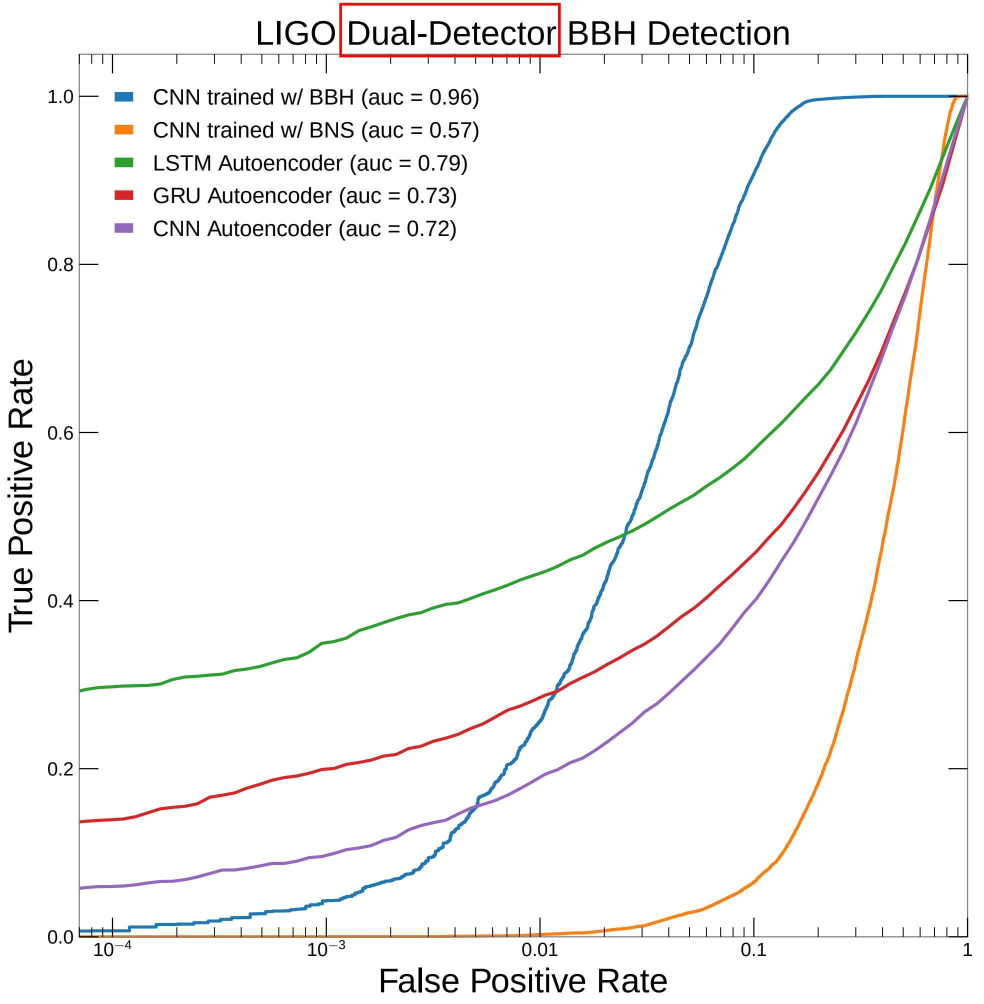
The Autoencoder

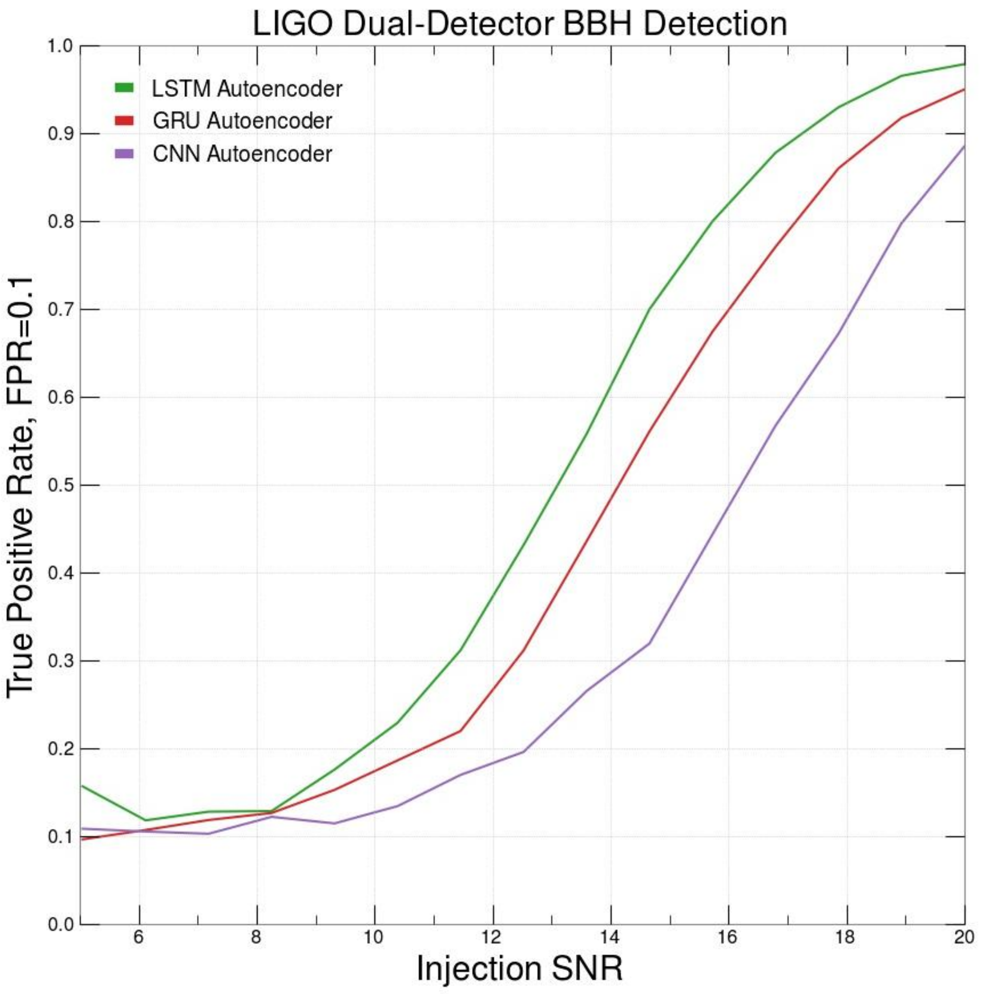


- Encoders and decoders made of:
  - Dense Neural Networks (DNN)
  - Convolutional layers (CNN)
  - Recurrent Neural Networks (RNNs) such as LSTMs or GRUs which are good with dealing with time-dependent data
  - Spiking Neural Networks (interesting proposition!)

Exploiting Dual-Detector Coincidence

2107.12698

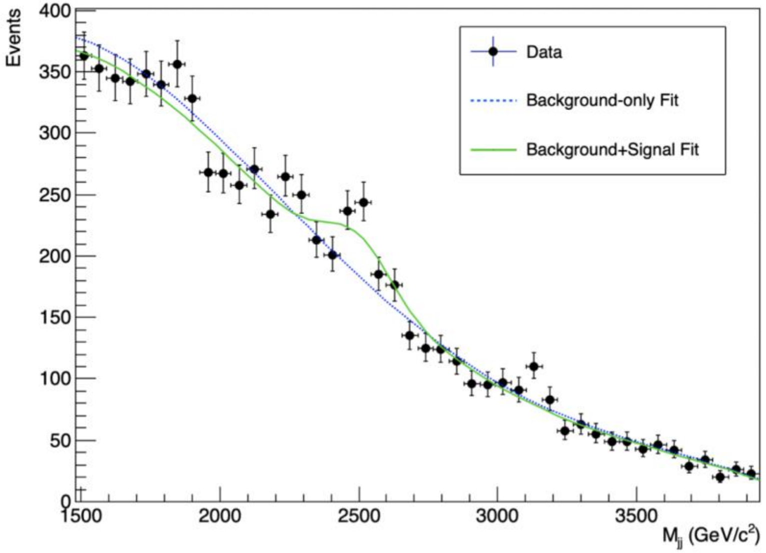




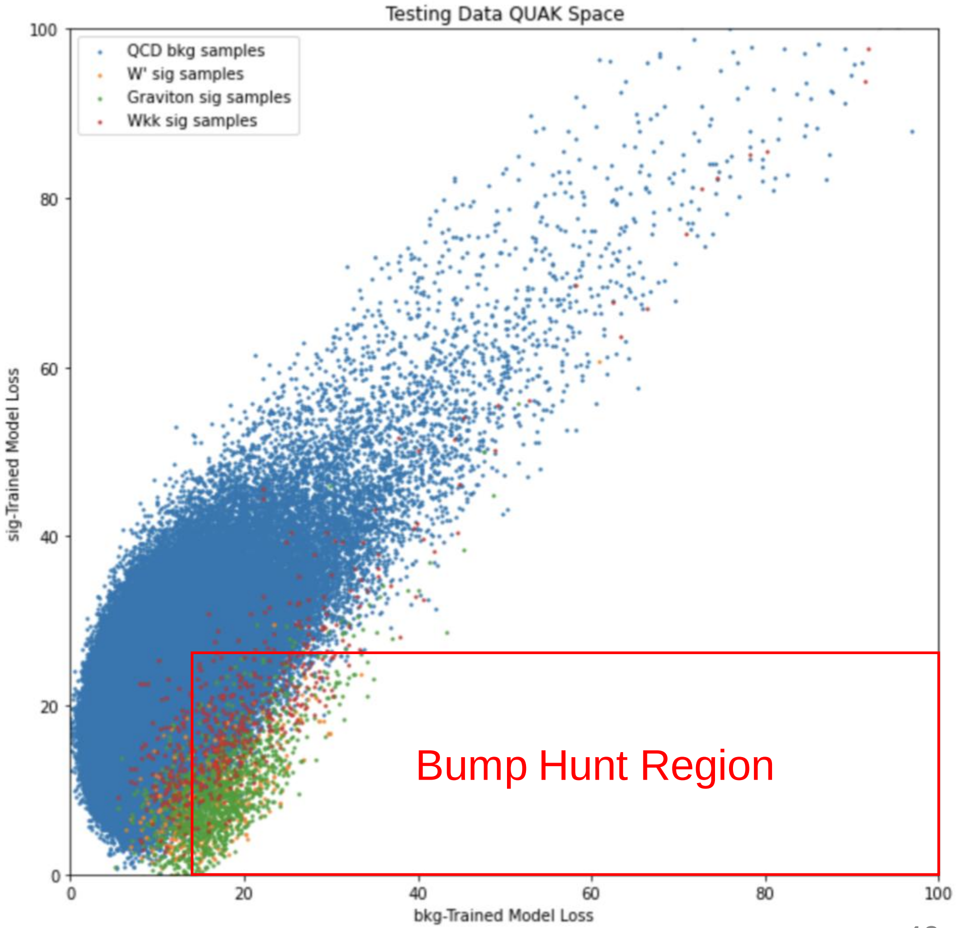
QUAK Anomaly Detection

M. Yunis, Thesis 2022

- Background: QCD
- Trained Signal: $W' \rightarrow WZ$
- Anomaly: $G \rightarrow ZZ$
- Anomaly: $W_{kk} \rightarrow WR \ \rightarrow W + WW$
- $M_W, M_G, M_{W_{kk}} = 2500\ GeV$



4.3 excess
at 2500 GeV!



Giving structure to physics events

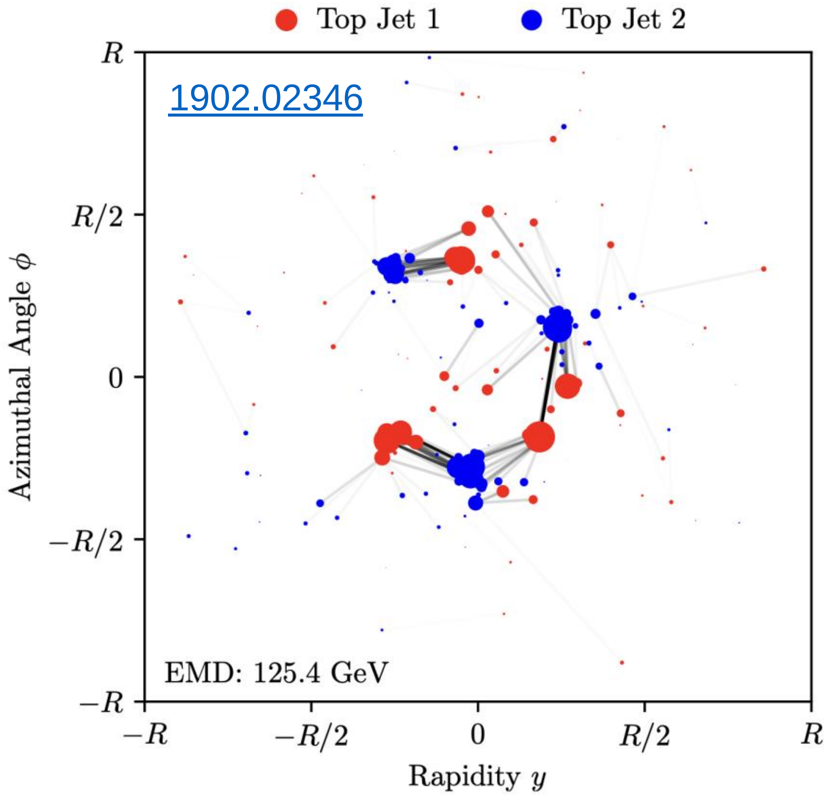
Idea from Sangeon Park et. al

Optimal transport based metric : Move one event to another by moving energy around
Energy Mover's Distance (Komiske, Metodiev, Thaler, 2019) :

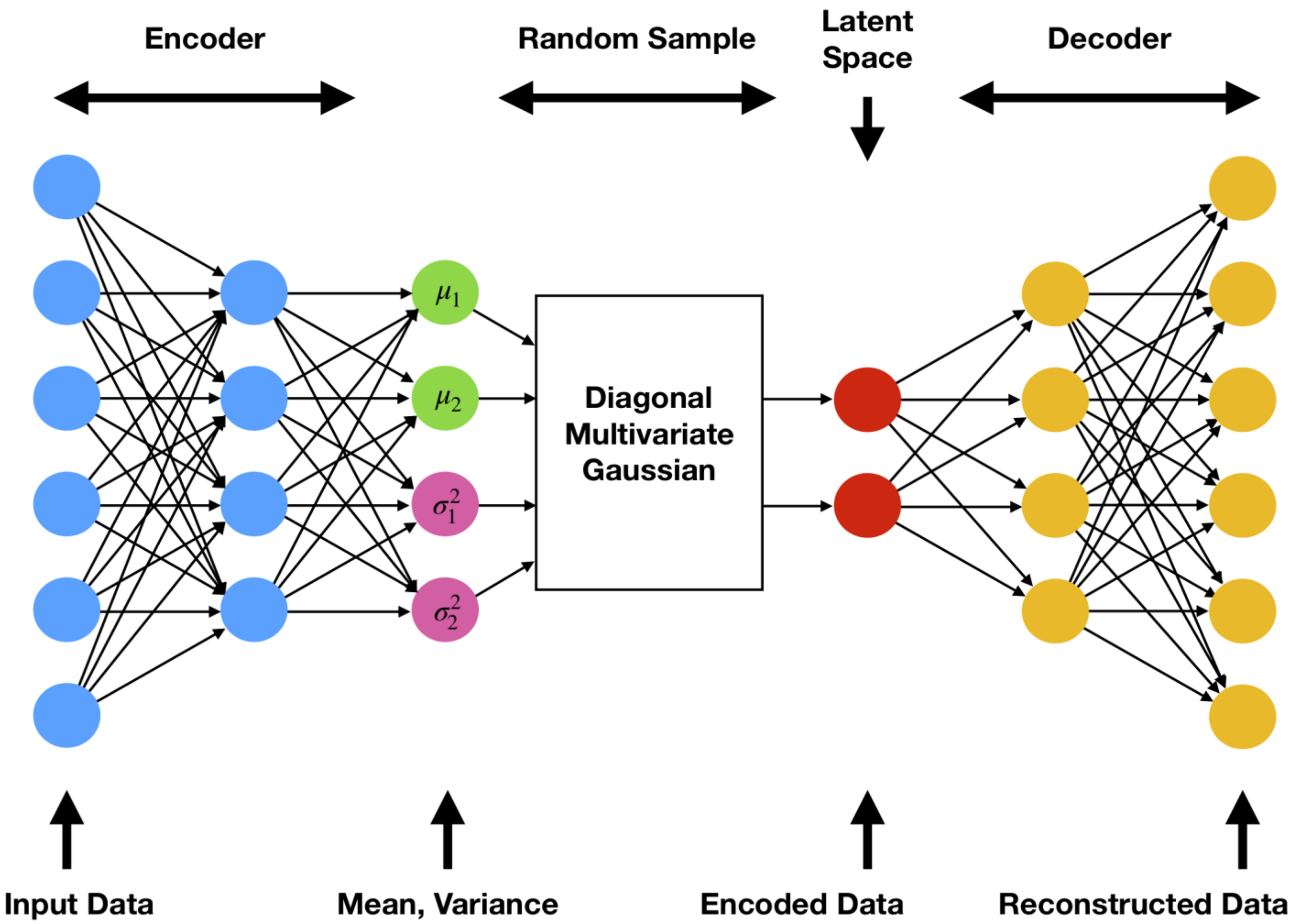
$$\mathrm{EMD}(\mathcal{E},\mathcal{E}') = \min_{\{f_{ij}\}} \sum_{ij} f_{ij} \frac{\theta_{ij}}{R} + \left| \sum_i E_i - \sum_j E'_j \right|,$$

$$f_{ij} \geq 0, \;\; \sum_i f_{ij} \leq E_i, \;\; \sum_i f_{ij} \leq E'_j, \;\; \sum_{ij} f_{ij} = E_{\min}$$

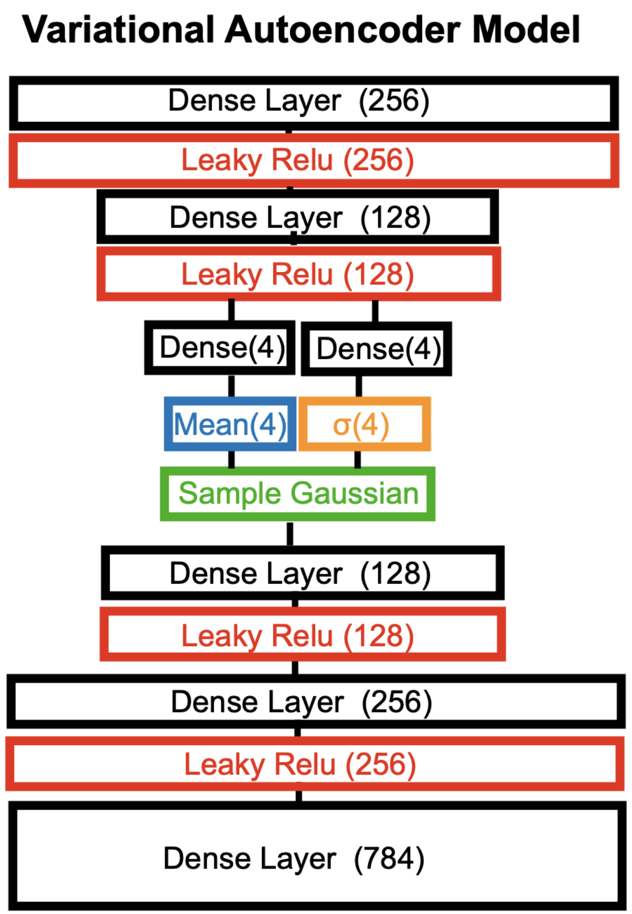
By embedding, we can do a lot of things! Mapping complicated metrics to simpler metrics can give access to powerful algorithmic toolkits, data compression

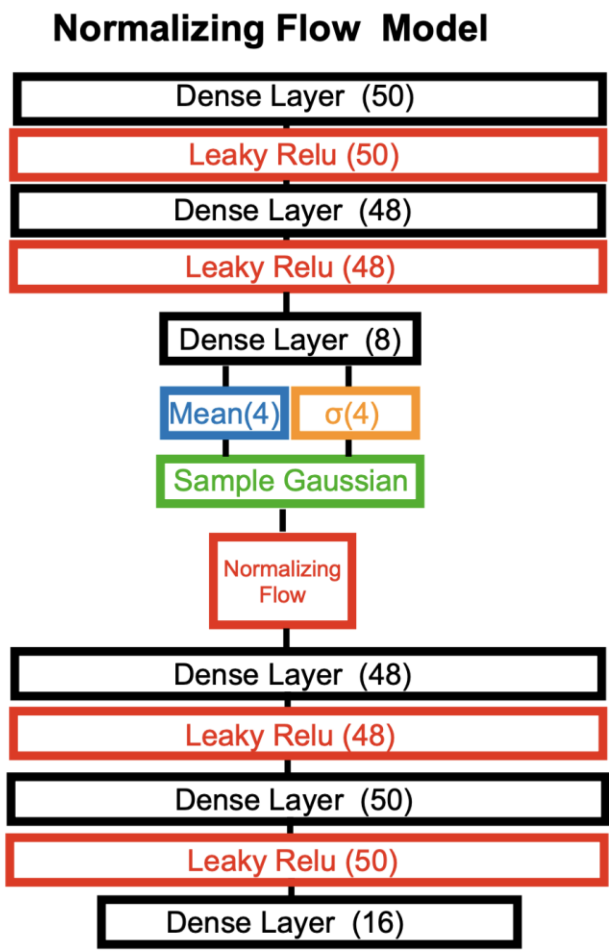


Variational Autoencoder (VAE)



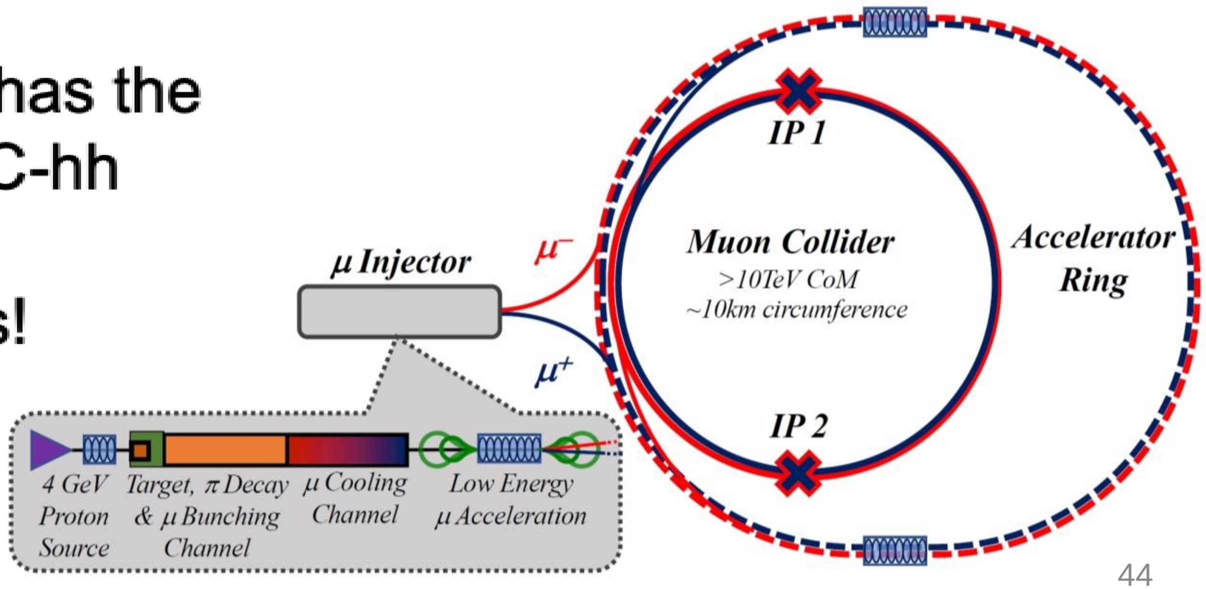
VAE vs Normalizing Flow





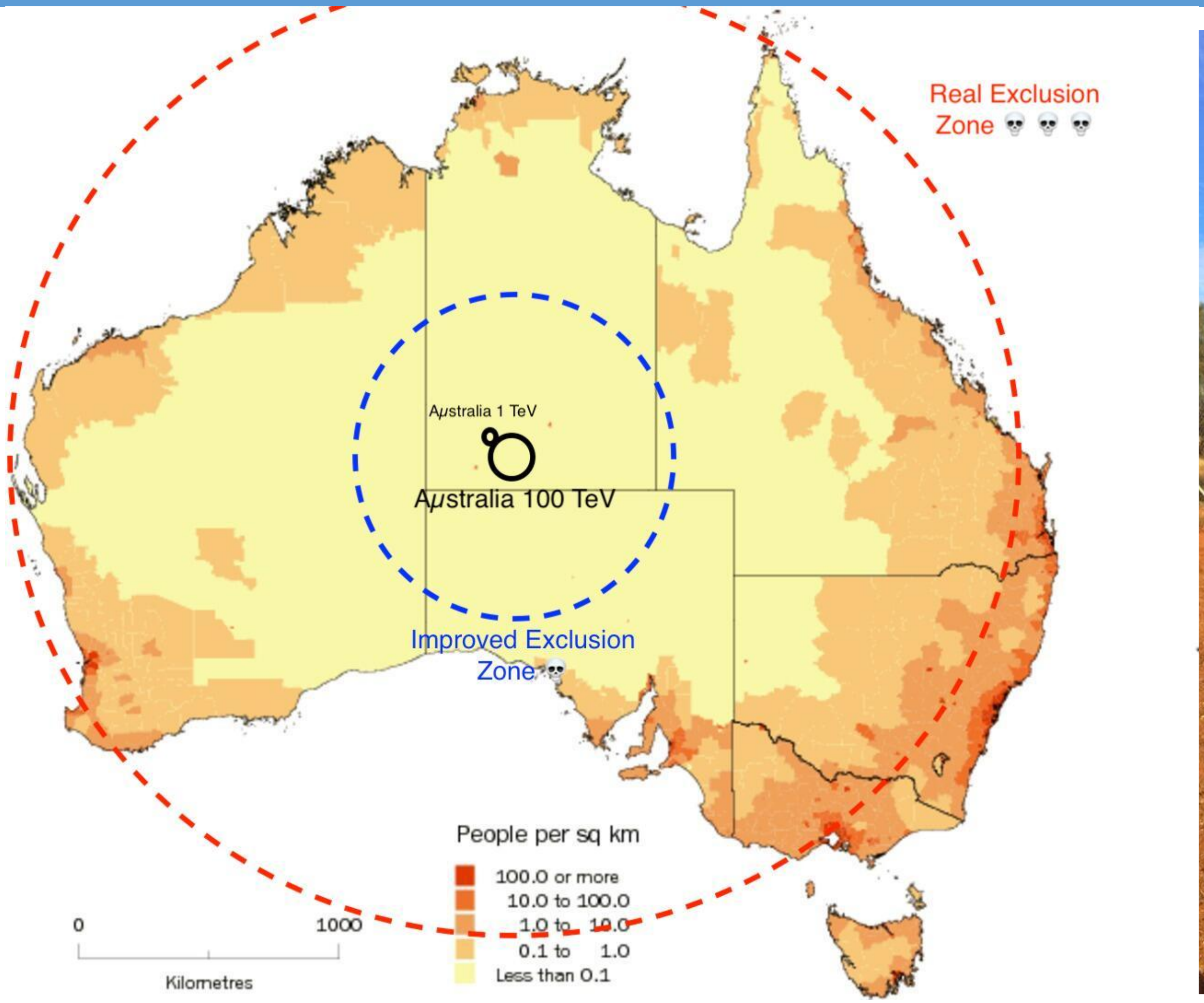
μstralia, Mustralia

- μ+μ- collider is necessary to efficiently explore higher energies!
- Very hard to make a muon beam – would require protons on a target, resulting on pions which decay to muons and need to be refocused and sent down a accelerator/collider – likely circular.
- Cyclotron radiation goes as m^{-4}
- Even a simple 10 TeV μ+μ- collider has the same CoM energy as proposed FCC-hh (100 TeV)
  - Muons are fundamental particles!



Radioactive Kangaroos/Emus

https://accelconf.web.cern.ch/p99/PAPERS/THP52.PDF



Nonnegligible radiation from muons in soil