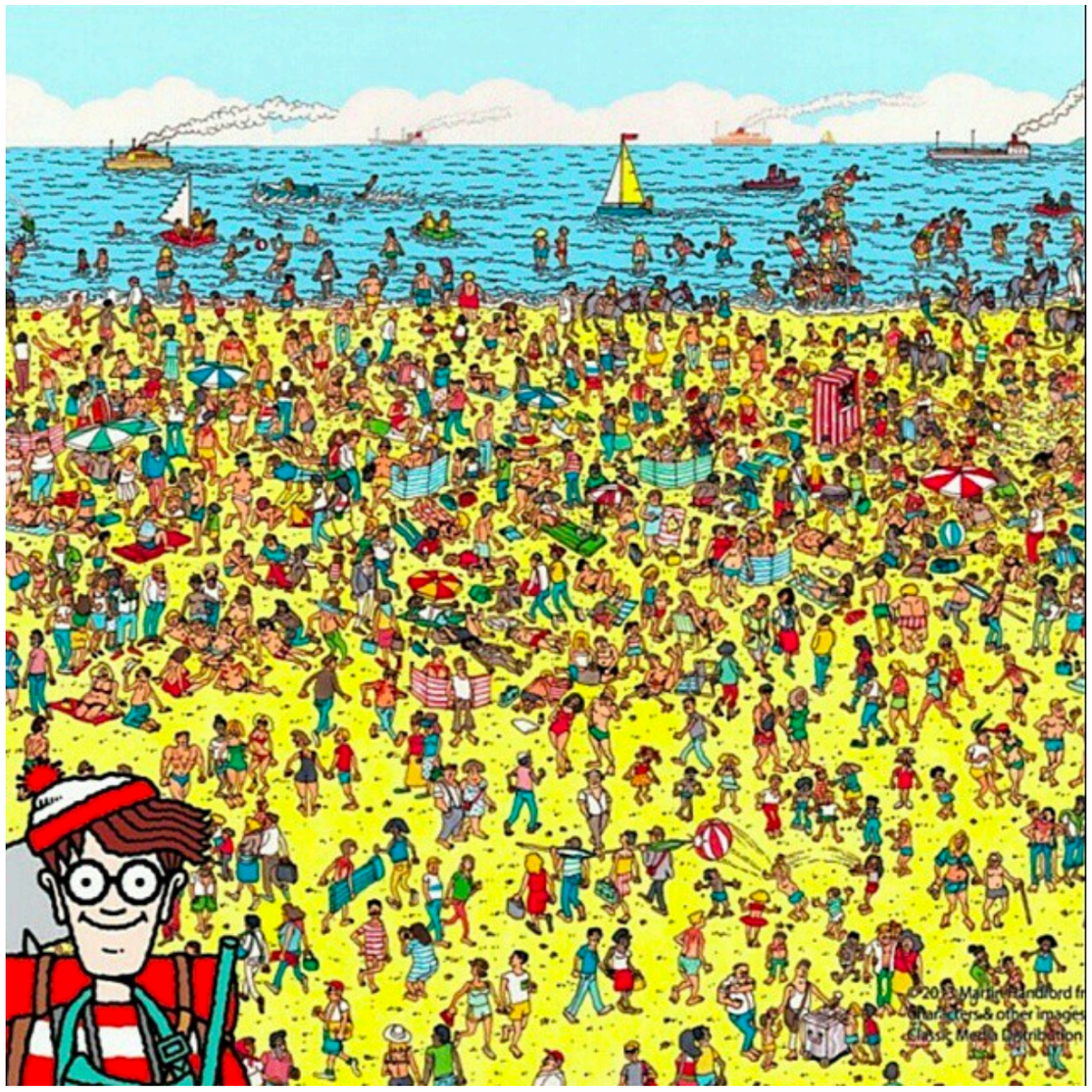


Anomalies? in Open Data

P. Harris (MIT)

Open Data+LHC Olympics

- Given the title and precursor to this talk
  - This talk will focus on Open Data Analyses strategies
    - In particular we are going to focus on anomaly detection
    - Present this in the context of the LHC Olympics 202
- Additionally I will discuss open data presentation
  - Discuss some of my experiences working with open data

LHC Olympics 2020

- https://lhco2020.github.io/homepage/

LHC Olympics 2020

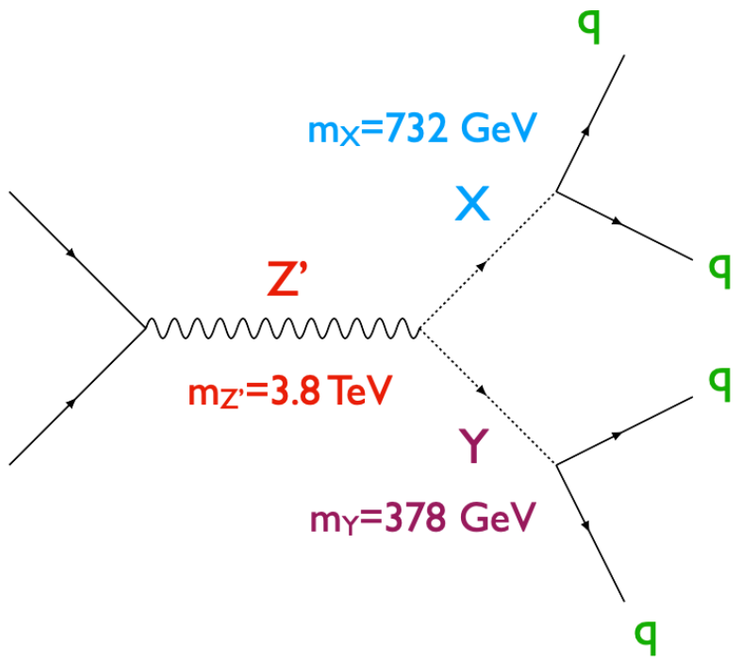
- Over the past year there was a competition
- In this setup a signal was hidden in pseudo data
  - The challenge was to “Find the hidden signal”
    - Emulate a realistic analysis as much as possible
    - Challenge : use deep learning to find an anomaly
- A number of different strategies are used for this approach
  - We will review the core concepts of these strategies

hep-ph/2101.08320

Olympic Data

- Strategy of the olympics:
  - Take a strange signal and hide it in toy data
- There were 3 black boxes split to emulate true data

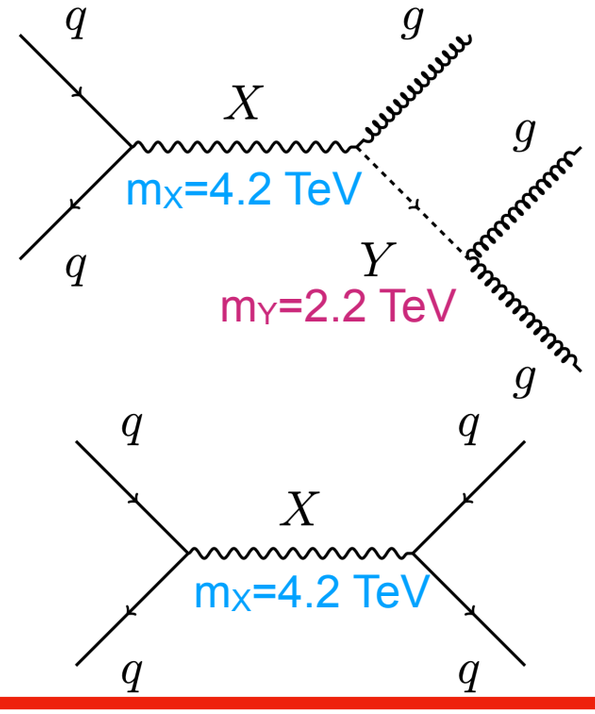
Black Box 1



Black Box 2

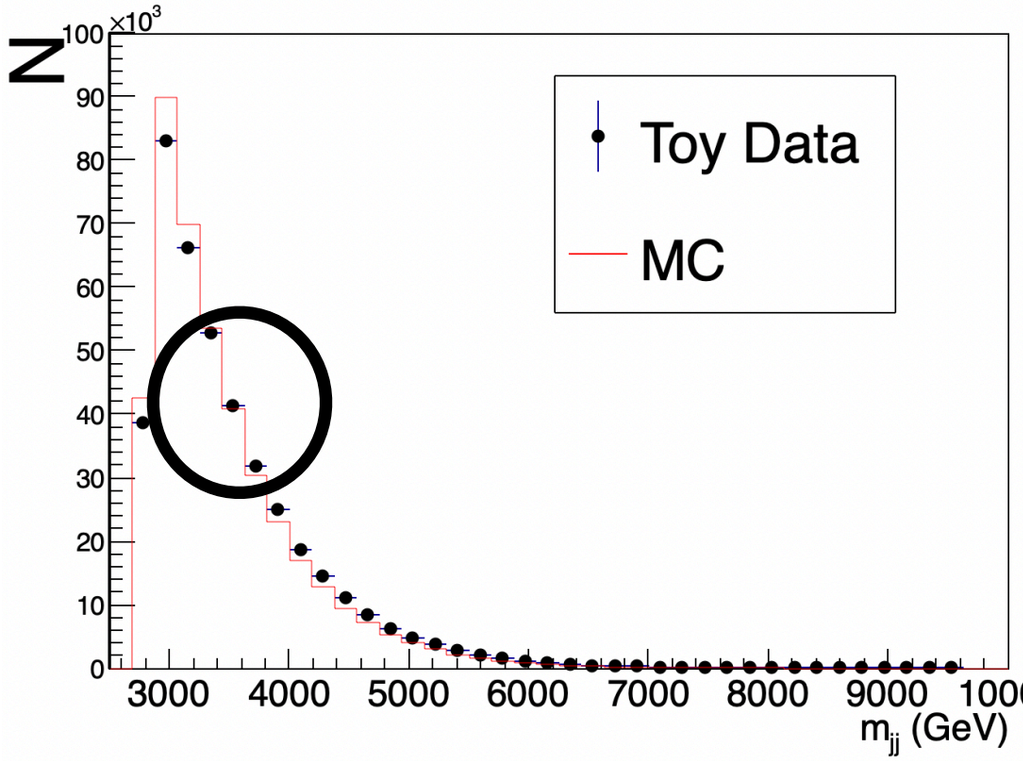
Nothing!

Black Box 3



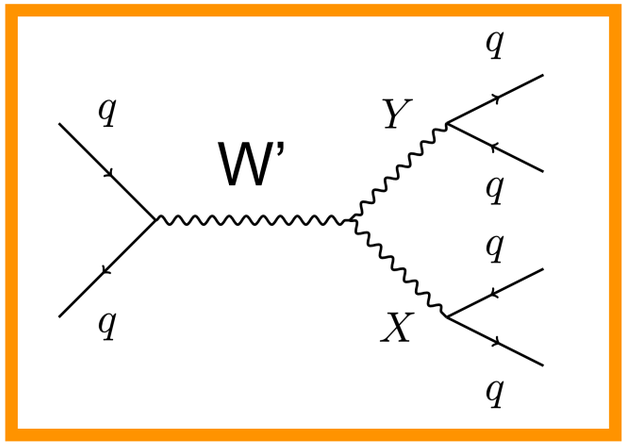
Olympic Simulation

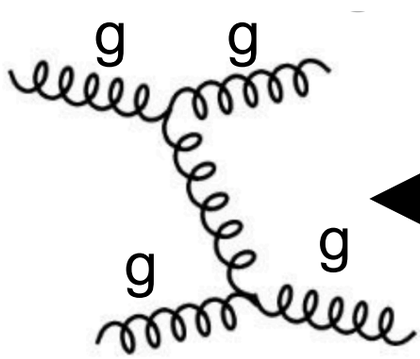
Samples



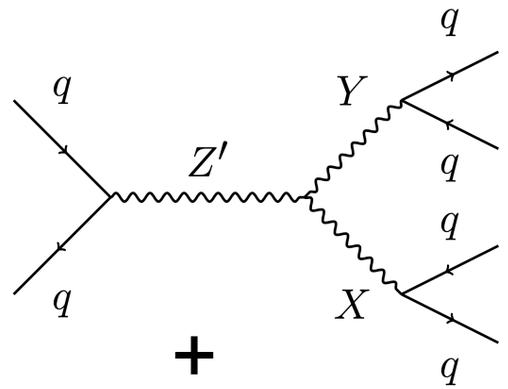
Data and simulation shower parameters lave large differences

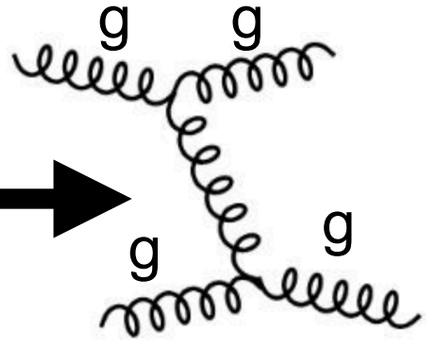
Simulation Samples





Toy Black Box





Drastically Different Simulation Parameters

- Aim was to emulate a real search as much as as possible
  - Simulation and Toy Data are released

Data Format

- Data released in h5 format
  - Standard python format using h5py and pandas
  - Easy to process tools that allow for quick turnaround

Entrée [1]:
```
import numpy as np
import matplotlib.pyplot as plt
import h5py
import pandas as pd
```

Entrée [2]:
```
file='events_anomalydetection_Z_XY_qqq.h5'
#f_sig = h5py.File(file,'r')
pd.read_hdf(file)
```

Out[2]:

	0	1	2	3	4	5	6	7	8	9	...	2091	2092	2093	2094	2095	2096
0	18.283588	-0.903479	0.060979	.316431	-0.784941	-0.008755	9.464178	-0.812918	-0.037386	4.578035	...	0.0	0.0	0.0	0.0	0.0	0.0
1	17.661003	-0.446288	-1.379160	.478683	-0.458125	-1.373650	8.452606	-0.455308	-1.375457	99.440353	...	0.0	0.0	0.0	0.0	0.0	0.0

Particle #1 Particle #2 Particle #3

Search for Non-Standard Sources of Parity Violation in Jets at $\sqrt{s} = 8$ TeV with CMS Open Data

Christopher G. Lester[a] **Matthias Schott**[b,c]

[a] *Cavendish Laboratory, University of Cambridge, UK*
[b] *Massachusetts Institute of Technology, Cambridge, USA*
[c] *Johannes Gutenberg-University, Mainz, Germany*

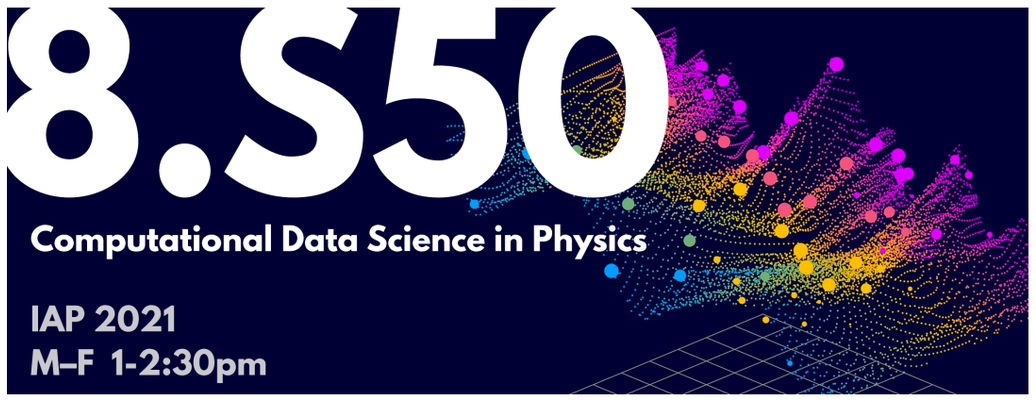
E-mail: lester@hep.phy.cam.ac.uk, matthias.schott@cern.ch

Opportunities and Challenges of Sta
Production Cross Section Measurem
Proton–Proton Collisions at $\sqrt{s}$=8 TeV using CMS Open Data

Aram Apyan[a] **William Cuozzo**[b] **Markus Klute**[b] **Yoshihiro Saito**[b] **Matthias Schott**[1b,c] **Bereket Sintayehu**[b]

[a] *Fermilab, USA*
[b] *Massachusetts Institute of Technology, Cambridge, USA*
[c] *Johannes Gutenberg-University, Mainz, Germany*

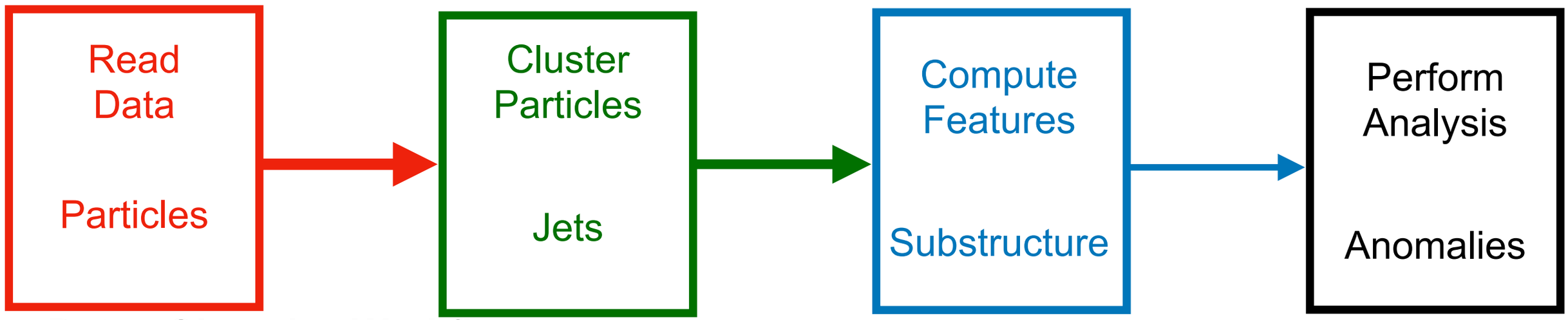
E-mail: matthias.schott@cern.ch



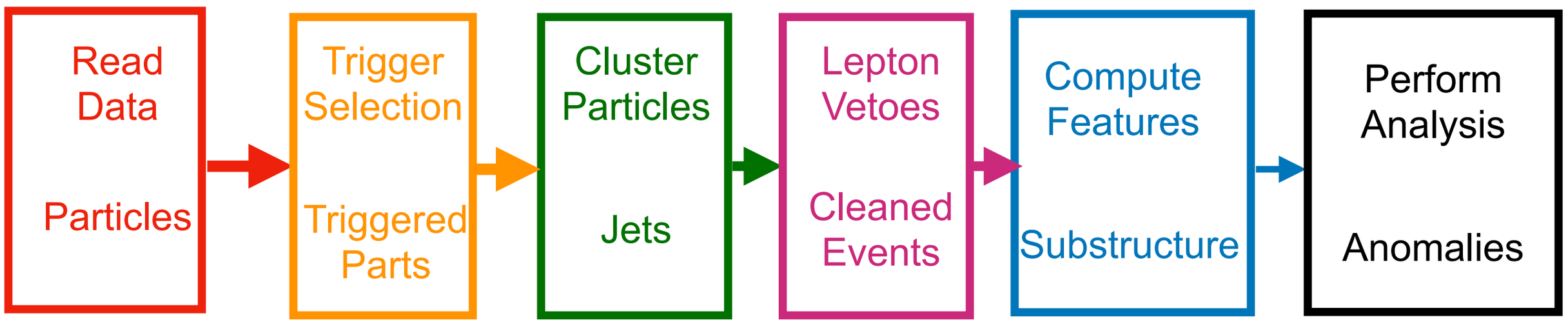
An Aside on Open Data

Processing Data

- To get from particles to analysis follow standard tool flow



Toy Data : Olympics Workflow



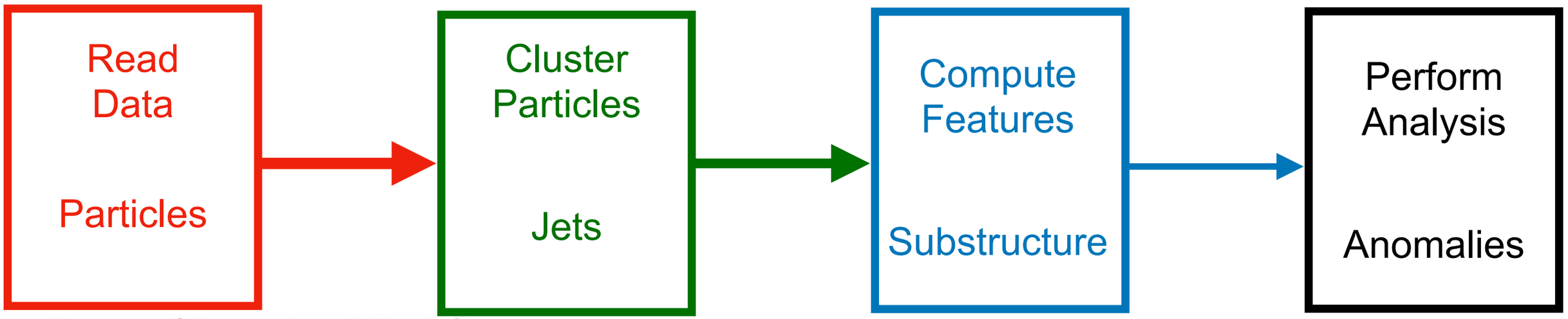
Real Data : Minimum Workflow

Why the extra steps?

- Going to real data a number of effects need to be considered
  - Data needs to pass a well defined/measured trigger
    - Bias or inclusive selection can introduce peaks
      - Sample needs to be close to pure QCD to emulate toy data
        - Processes like ttbar, W+jets will contribute significantly
- In reality, there are several more steps
  - Above steps constitute a minimum to emulate olympics

Processing Data

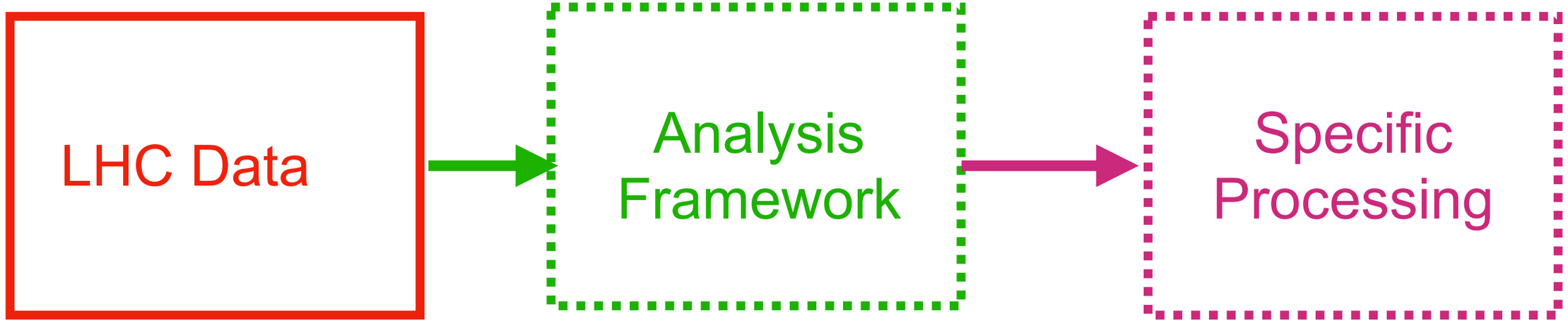
- To get from particles to analysis follow standard tool flow



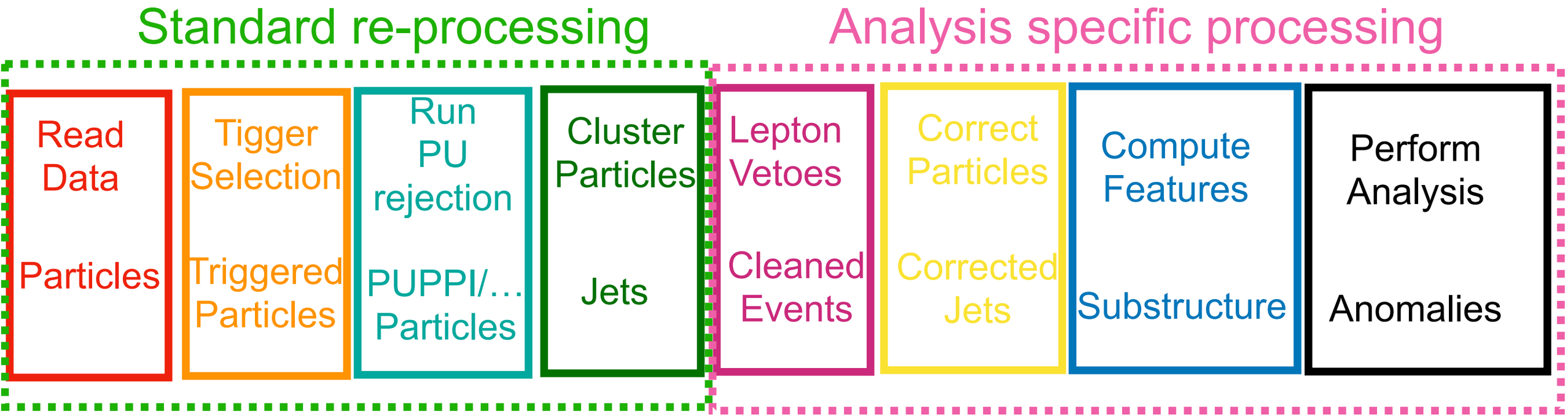
Toy Data : Olympics Workflow

Real Data : Minimum Workflow

How is this usually done?



- Split is typically done to limit the amount of re-computing



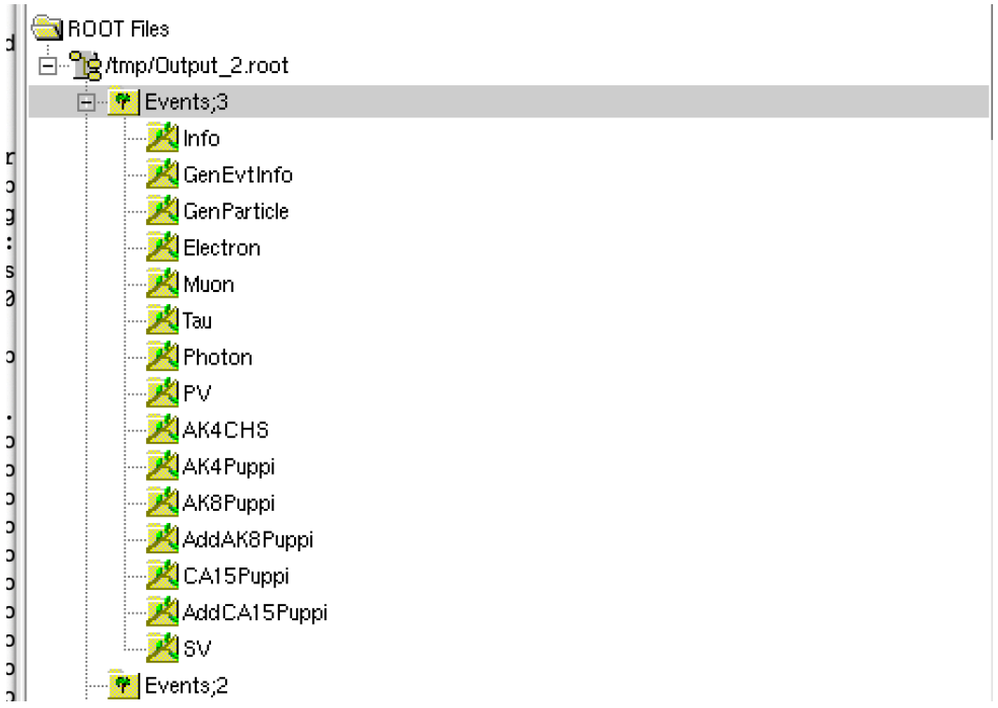
Real Data : Minimum Workflow

Building an Analysis FWK

- Frameworks take a long time to build
  - Complicated steps to follow careful curation of the data
  - Many iterations to avoid bugs in code
  - Data formatting what to keep a complex decision
- When preparing data for open analysis worked to get flat ntuple
- Collaborations have taken steps to centralize this
  - Newer data formats embed standard corrections
  - These data formats starting to be avaible in open data

Towards Regularization

- Bigger biases/corrections eventually embedded in software
  - In CMS: MiniAOD => NanoAOD
  - These are light smaller frameworks that lead to fast analysis
  - Still don’t solve all problems



Other things Lost

- Certain aspects in the data requires insider knowledge
  - Trigger preparation/Trigger biases
  - Which detectors were misfired
    - Details to address these issues are often complicated
- How do you deal with understanding inside knowledge?
  - Talk to others doing data analysis
  - Inside the collaboration many of these are well known

Examples Approaching

- Example sample approaching toy data
  - Special MC simulation sample used for Higgs tagging here
- Discussion on FAIRness of CMS open data here
  - Consensus is that this is close, but could be better
- Samples are are converted to h5 inputs

Dataset semantics

Variable	Type	Description
event_no	*UInt_t*	Event number
npv	*Float_t*	Number of reconstructed primary vertices (PVs)
ntrueInt	*Float_t*	True mean number of the poisson distribution for this event from which the number of interactions in each bunch crossing has been sampled
rho	*Float_t*	Median density (in GeV/A) of pile-up contamination per event; computed from all PF candidates of the event
sample_isQCD	*Int_t*	Boolean that is 1 if the simulated sample corresponds to QCD multijet production

Future of Datasets is the FAIR convention

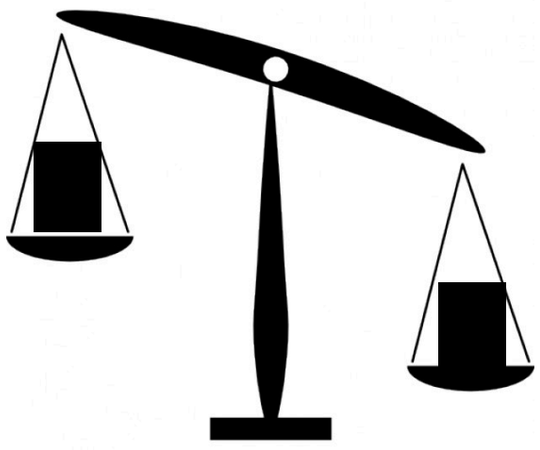
FAIR

- **F**indable
  - Resources easy to find to by both humans+computers
  - Metadata readily available; allows for the discovery of interesting data
- **A**ccessible
  - Resource and metadata can be easily accessed and downloaded
    - Both locally by a human, but also machines using standard protocols
- **I**nteroperability
  - Metadata should be ready to be exchanged, interepreted and combined in a semiautomated way with other datasets by humans and computers
- **R**euseability
  - Data and metadata are sufficiently well described to allow data to be reused
  - Proper citation must be facilitated and conditions should be valid to machines

Anomaly Strategies@LHC

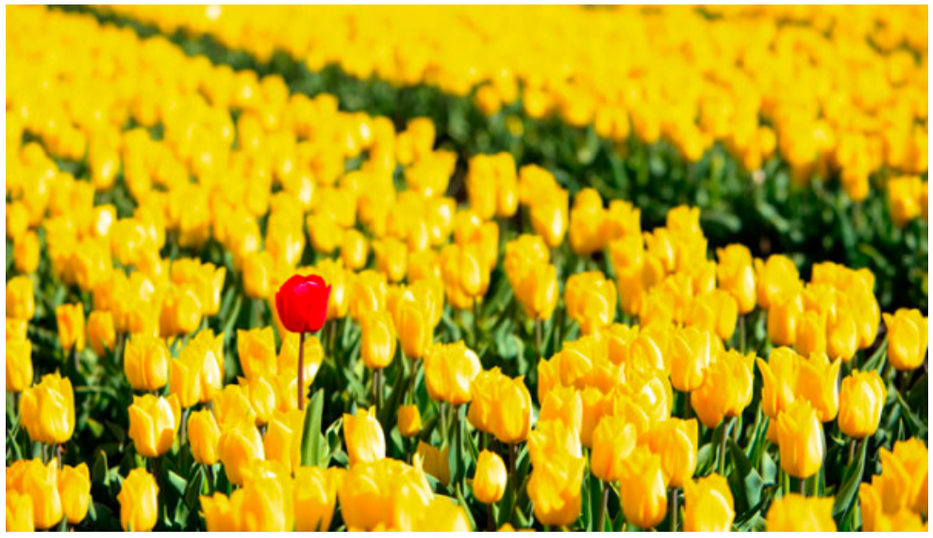
- Anomaly Strategies at LHC fall into two categories

I know regions where new physics does not exist



I want to leverage those regions against other parts of the data to find differences

I know how to predict all collisions



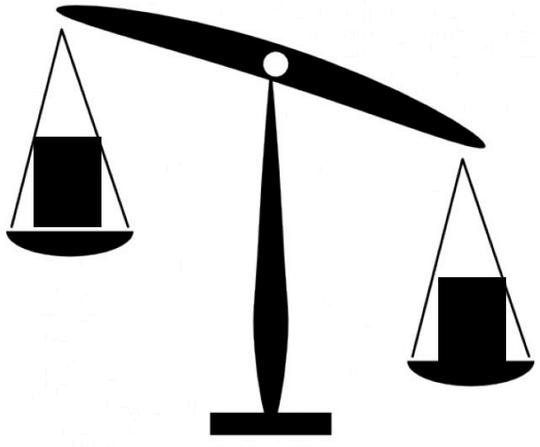
Are there any collisions that I cannot predict?

Anomaly Strategies@LHC

- Anomaly Strategies at LHC fall into two categories

Weakly-Supervised

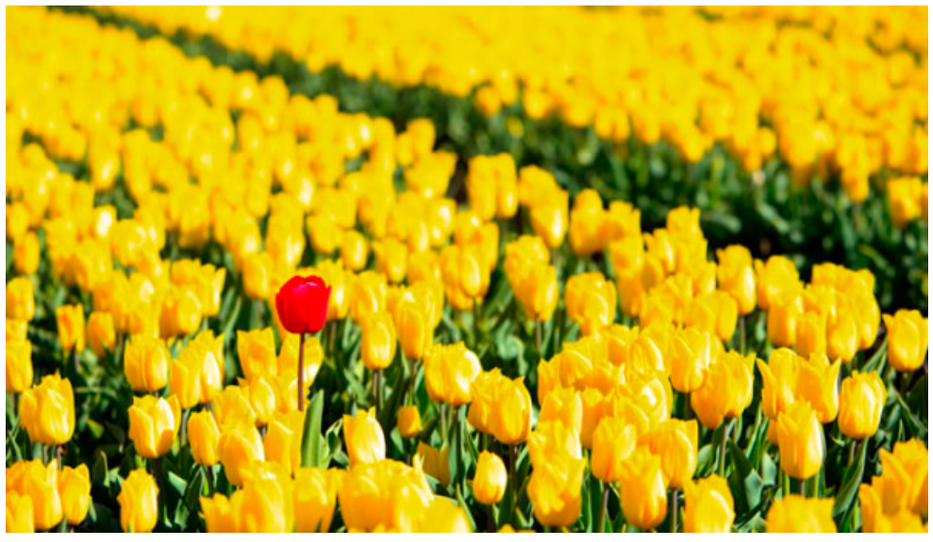
I know regions where new physics does not exist



I want to leverage those regions against other parts of the data to find differences

Autoencoders

I know how to predict all collisions



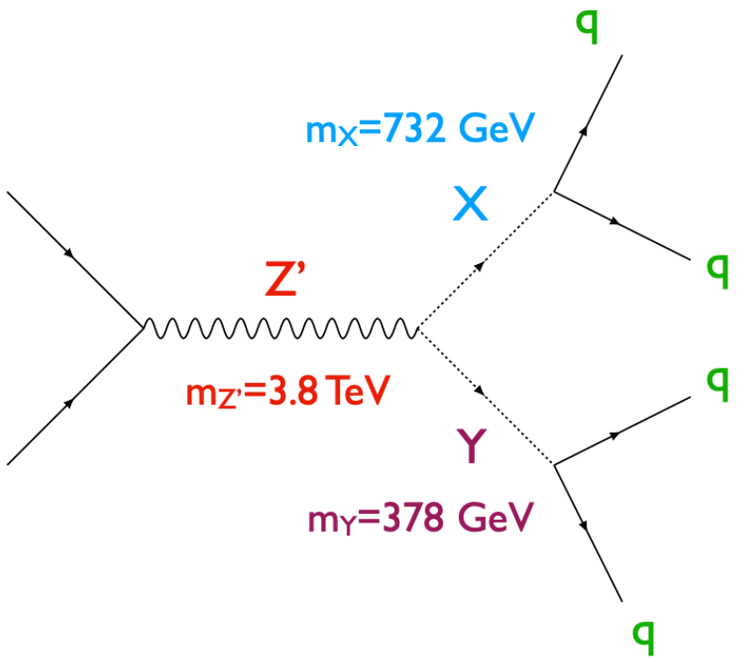
Are there any collisions that I cannot predict?

The LHC Olympics (interspersed with DL concepts)

Results

- We are going to foucs on black box 1

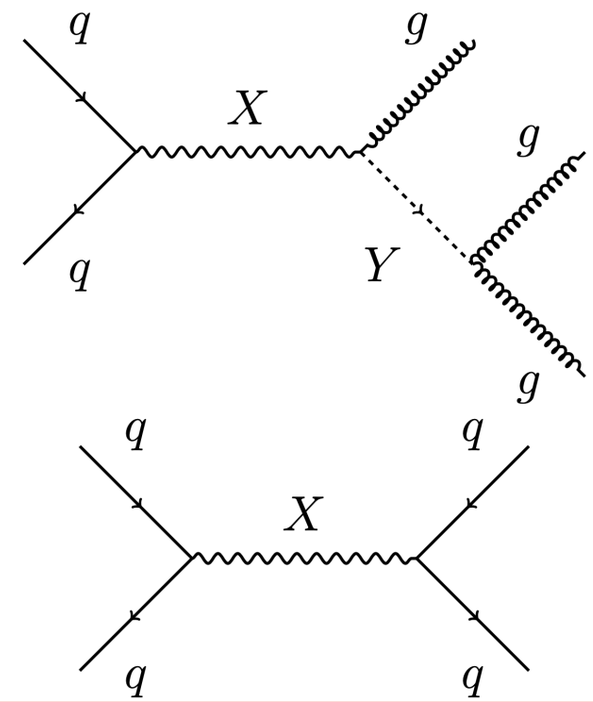
Black Box 1



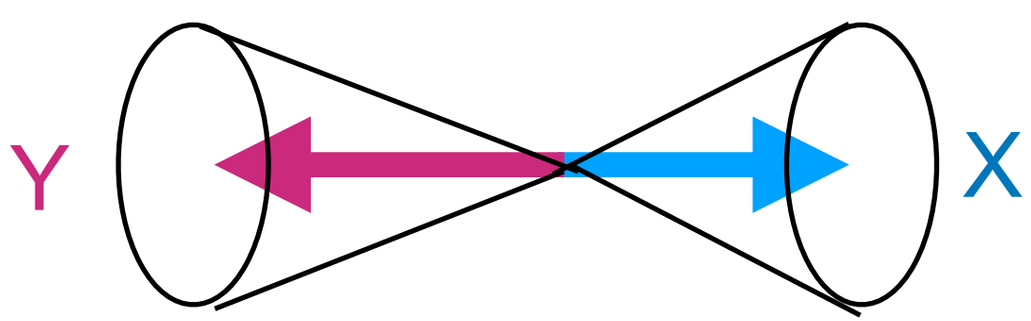
Black Box 2

Nothing!

Black Box 3

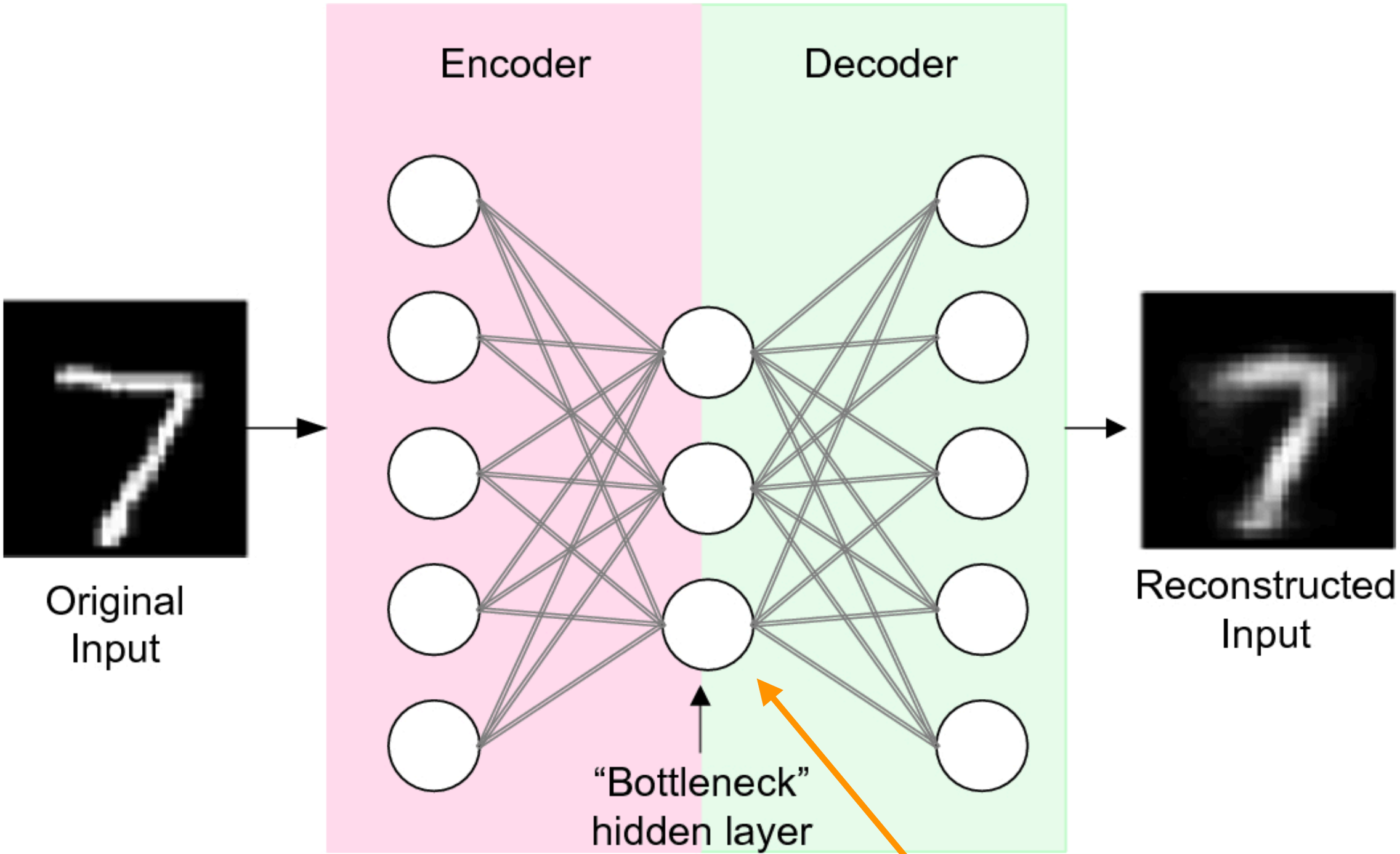


Dijet scenario



Can use all the jet substructure tools as input

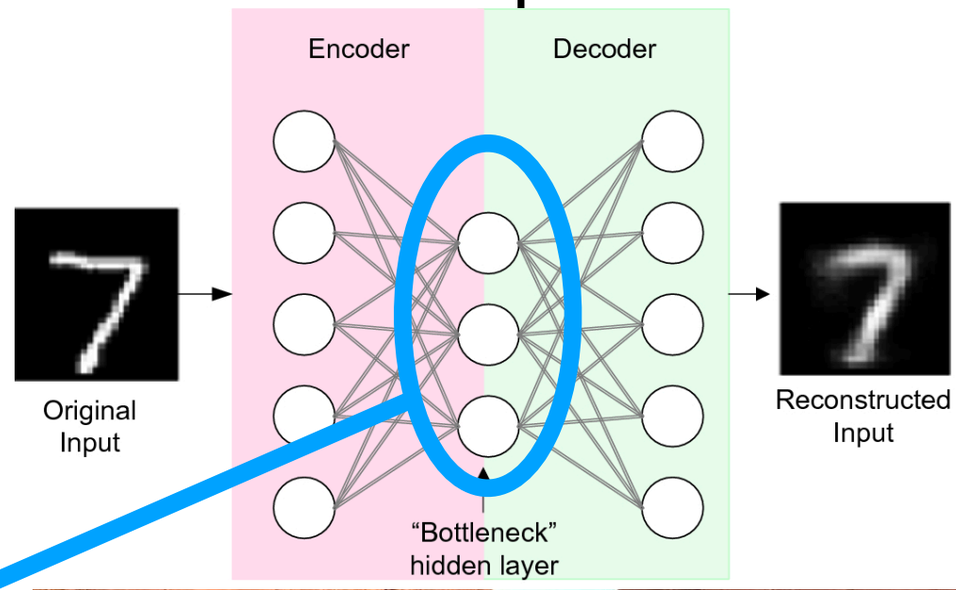
Autoencoders

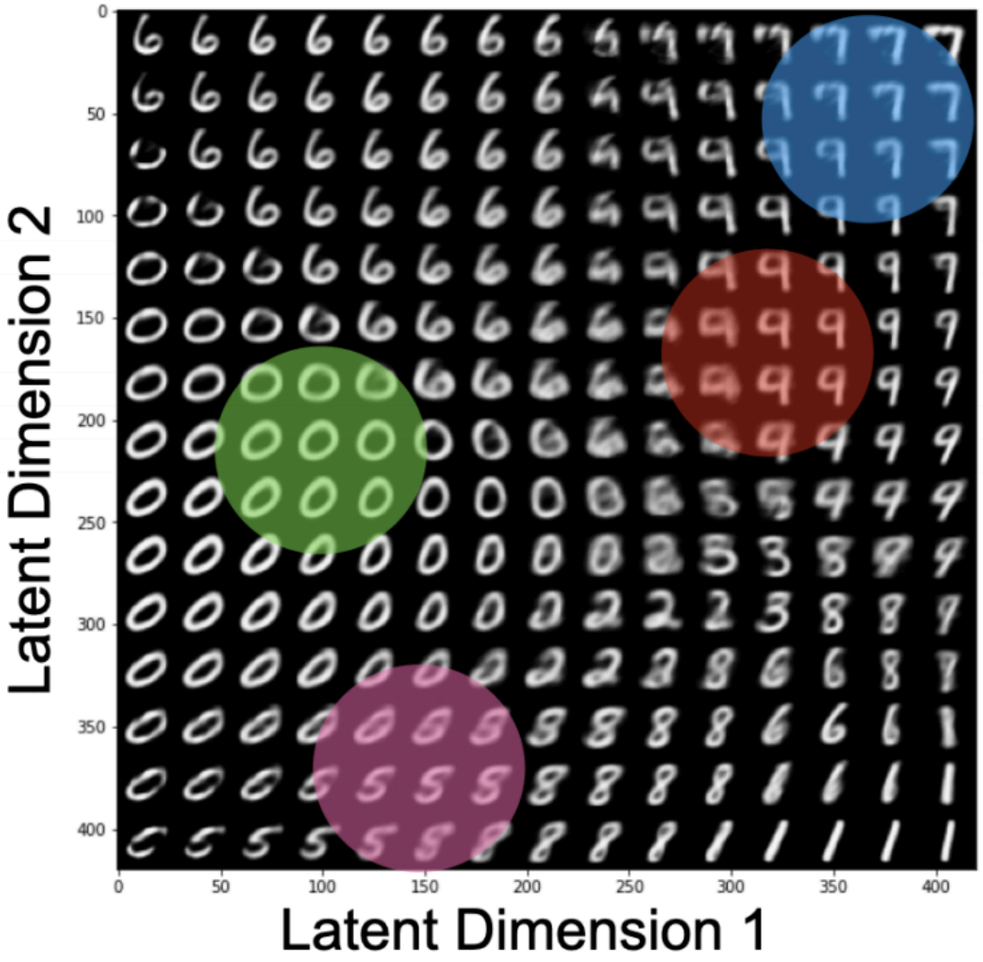


Strategy is to create a space in the middle that embodies all features of physics

The Latent Space

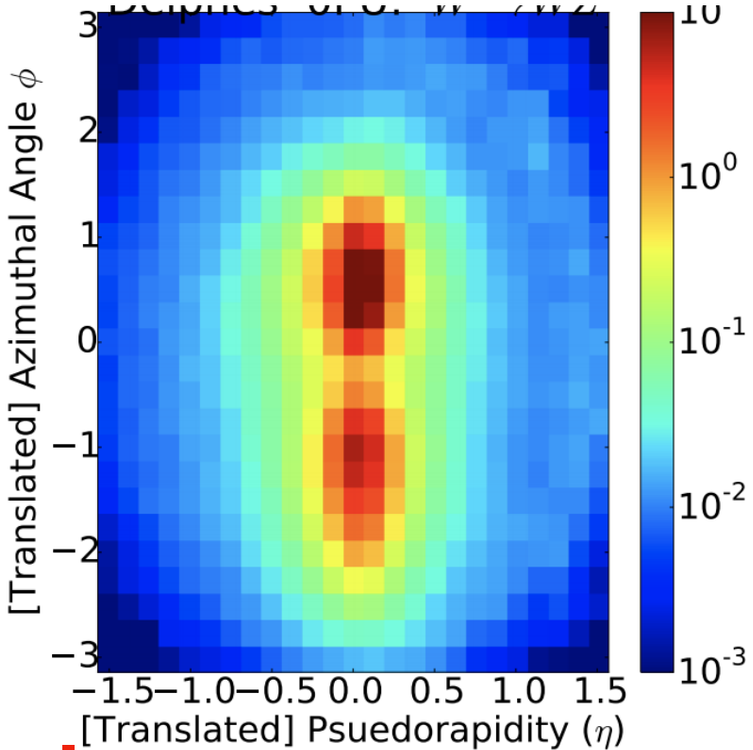
- Deep learning algos tend to focus on the latent space
- What is the latent space?
  - Its whatever you want it to be





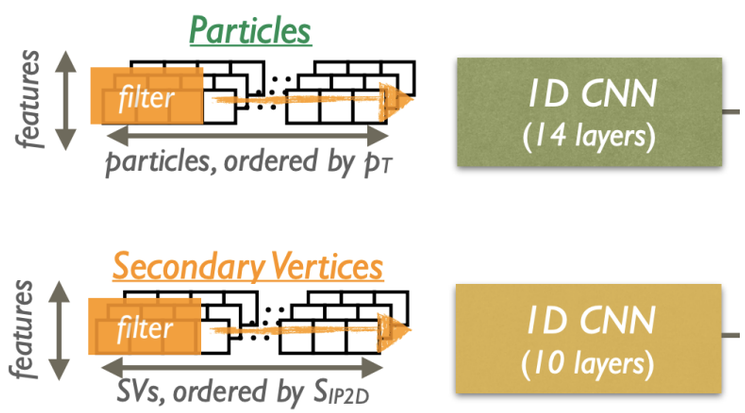
Encoder Progression

2016



Images
(not lorentz invariant)

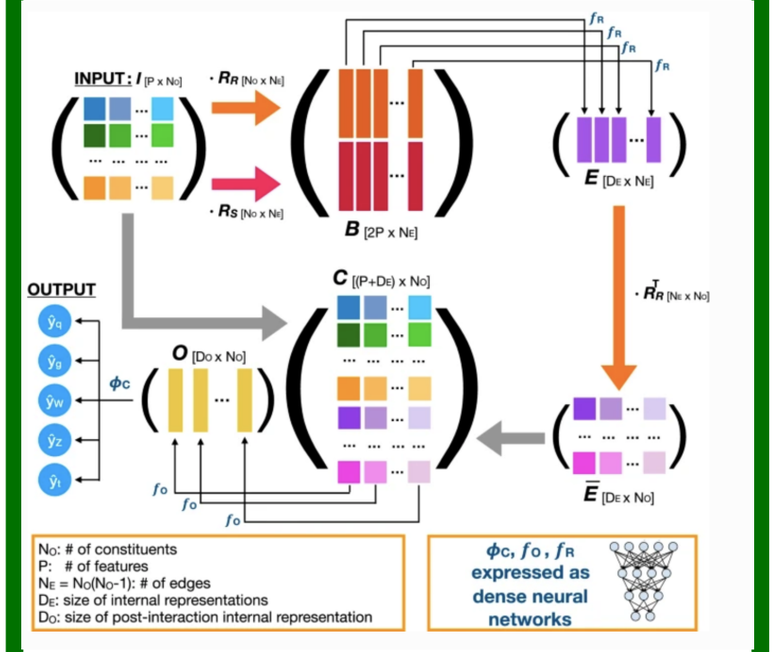
2018



Particles and SVs
with 4-vectors+features

Particles
(limited correlations)

2020



Graphs
(Particles+correlations)

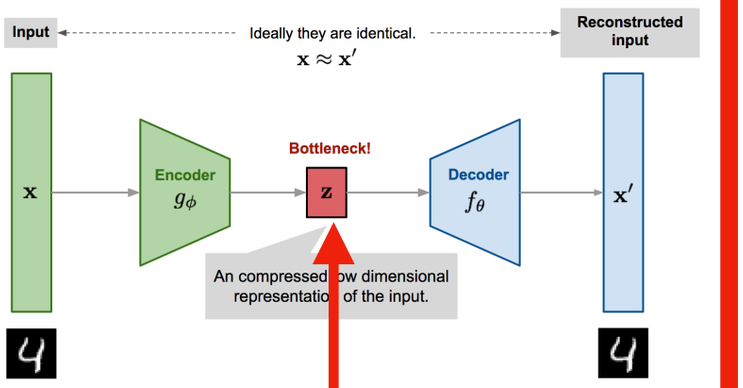
Current collaboration results

Progressively moving towards use of more info

Autoencoder Progression

- Autoencoders are gaining popularity in HEP just now

Dawn of Time

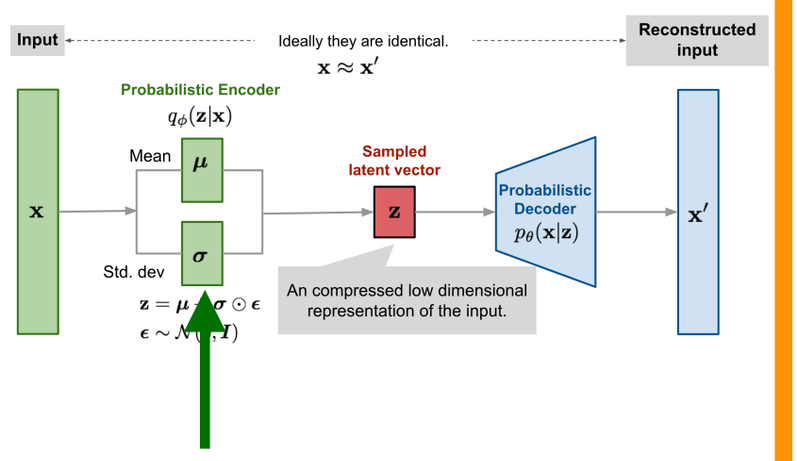


Small latent space that encodes physics

Autoencoder

Not smooth

2015

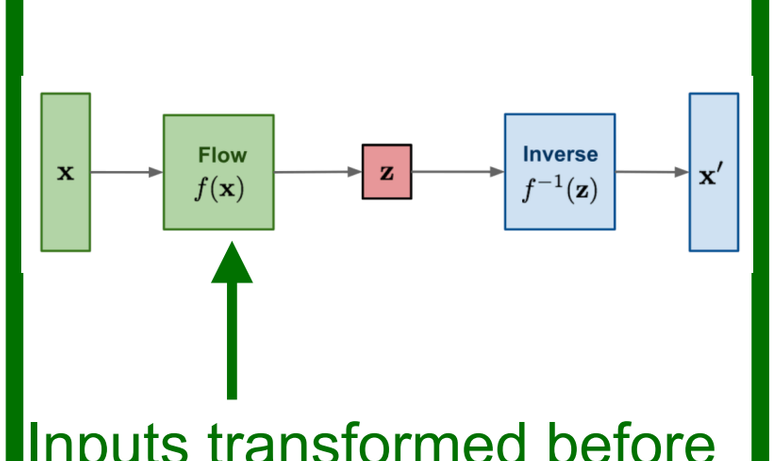


Inputs smeared w/gaussian before latent space

Variational AE/GAN

Smooth AE

2017

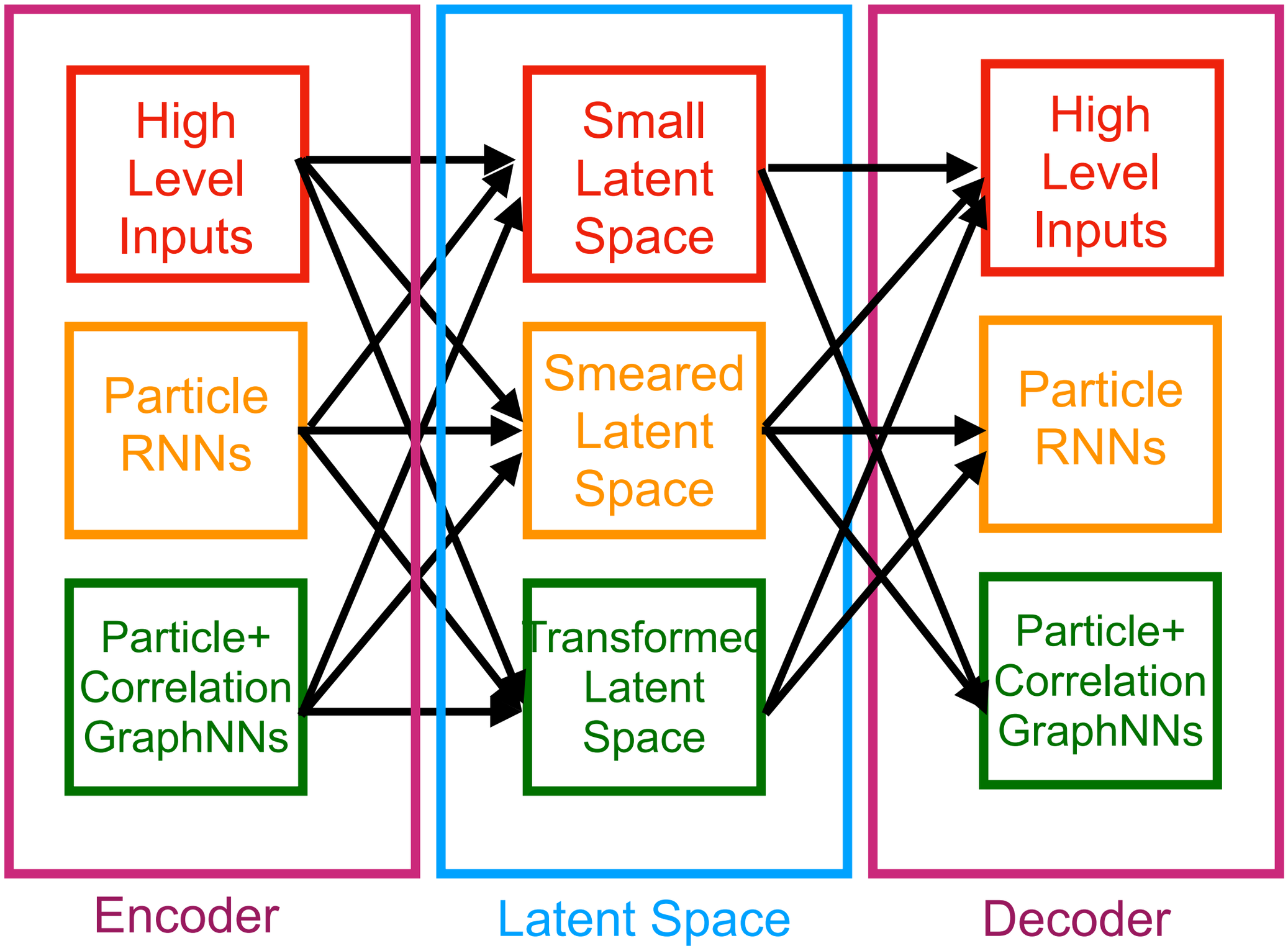


Inputs transformed before entering latent space

Normalizing Flow

Non-Gaussian VAE

Combinations



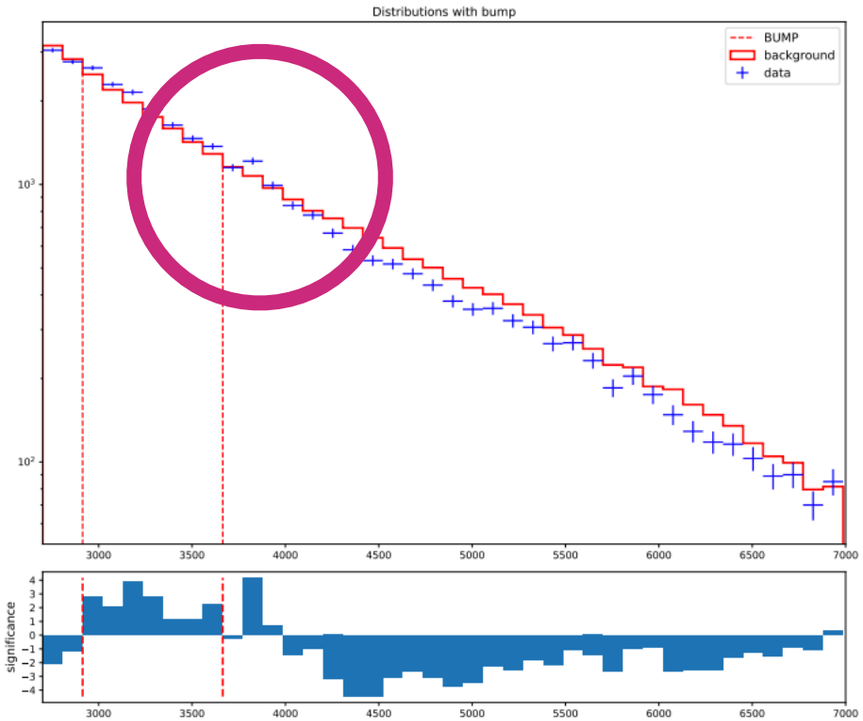
GAN supported AE

Inputs: High Level Features (Nsubjettiness/Jet masses/…)

- Build an auto encoder (AE)
  - Add an GAN to help AE
  - Additionally decorrelate with mass

$$\text{loss}_{\text{AE}} = \text{BC} + \varepsilon \times \text{MED} + \alpha \times \text{DisCo}$$

BB1 Dataset



ıtent space forced to be decorrelated with mass

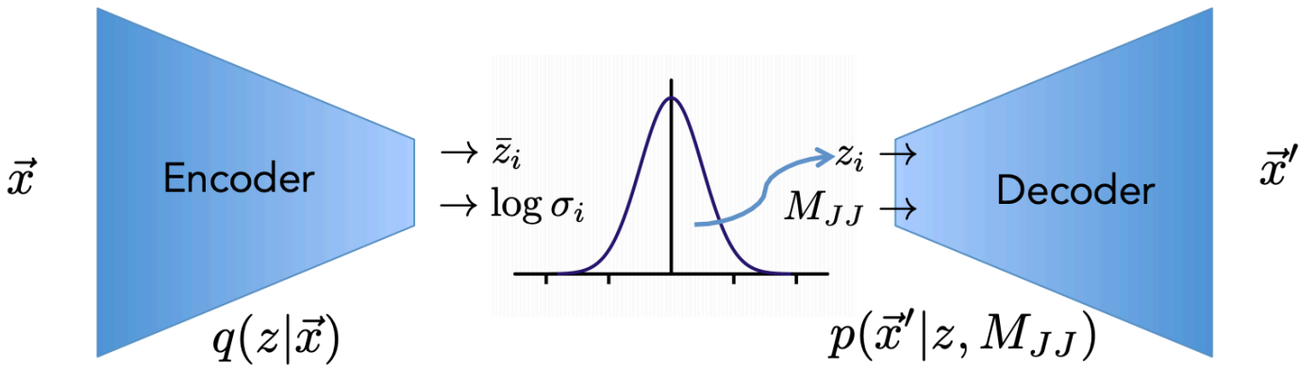
Signal Extraction : Bump Hunter (it Failed)

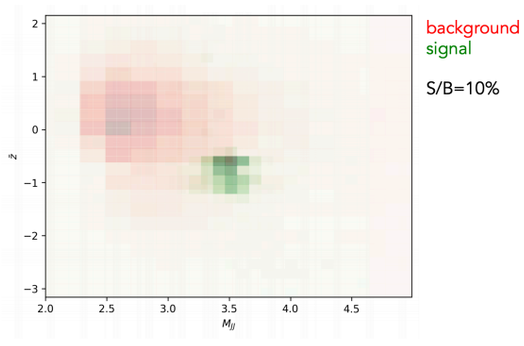
Take away: Mass Decorrelation+Good Simulation needed

BuHuLaSpa

Inputs: High Level Features (Nsubjettiness/Jet masses/…)

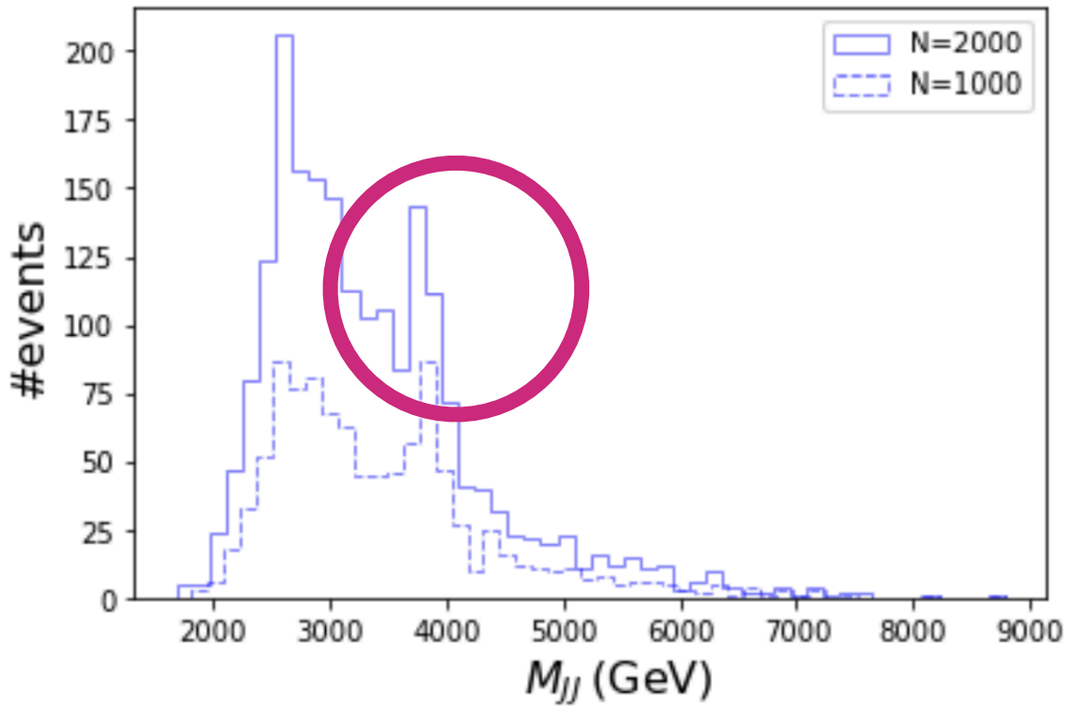
- Bump hunting in the latent space





Encoder output

BB1 Dataset



Autoencoder with 1D latent space
Latent space forced to be decorrelated with mass

$$\mathcal{L} = -D_{\mathrm{KL}}(q_\phi(\vec{z}_i|\vec{x}_i)|p(\vec{z}_i)) + \beta_{\mathrm{reco}} \log p_\theta(\vec{x}_i|\vec{z}_i)$$

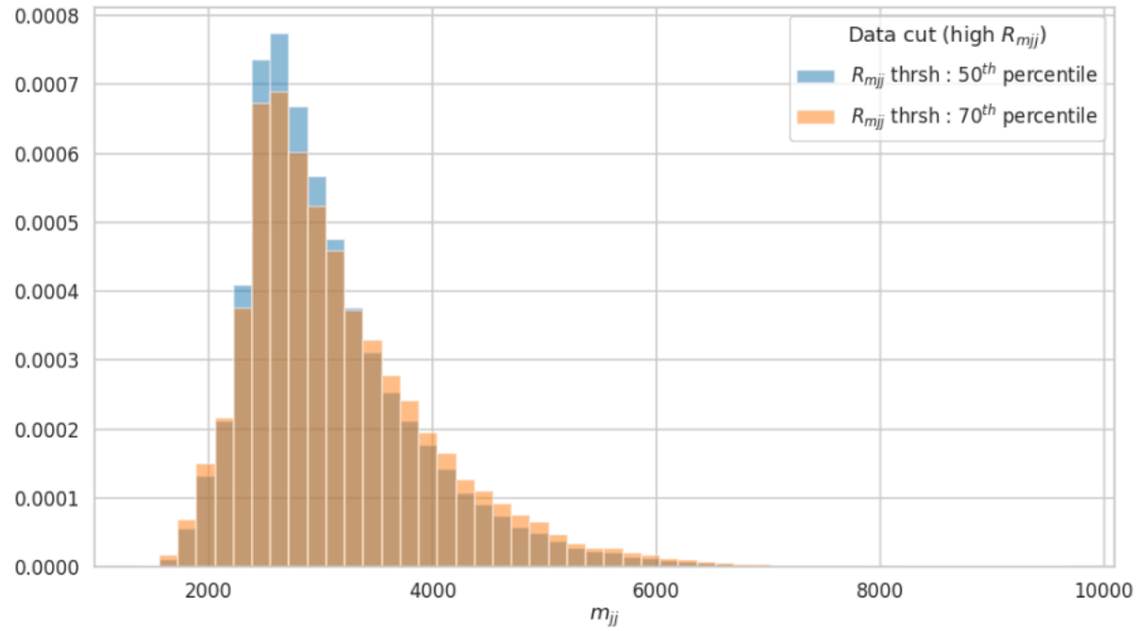
Signal Extraction : None

Take Away:Training is critical to ensure good performance

Normalizing Flow

Inputs: High Level Features (Nsubjettiness/Jet masses/…)

BB1 Dataset



- Use a normalizing flow
  - Cut on high loss
  - Decorrelate loss with mass

$$\mathcal{R}_{m_{jj}}(x) = \frac{||x - g(g^{-1}(x))||^2}{1 + \frac{p_u(g^{-1}(x))}{p_{KDE}(m^x_{jj})}}$$

Cut is too loose (may actually work)

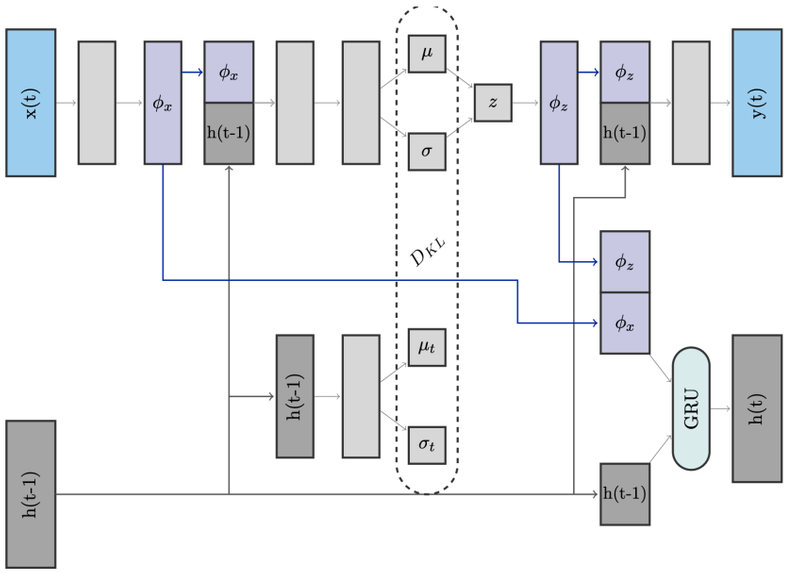
Signal Extraction : None (No signal)

Take Away: single auto encoder even with NF is not enough
too many anomalies (no clear signal)

Particle VAE

Inputs: Particle four vectors of the jet

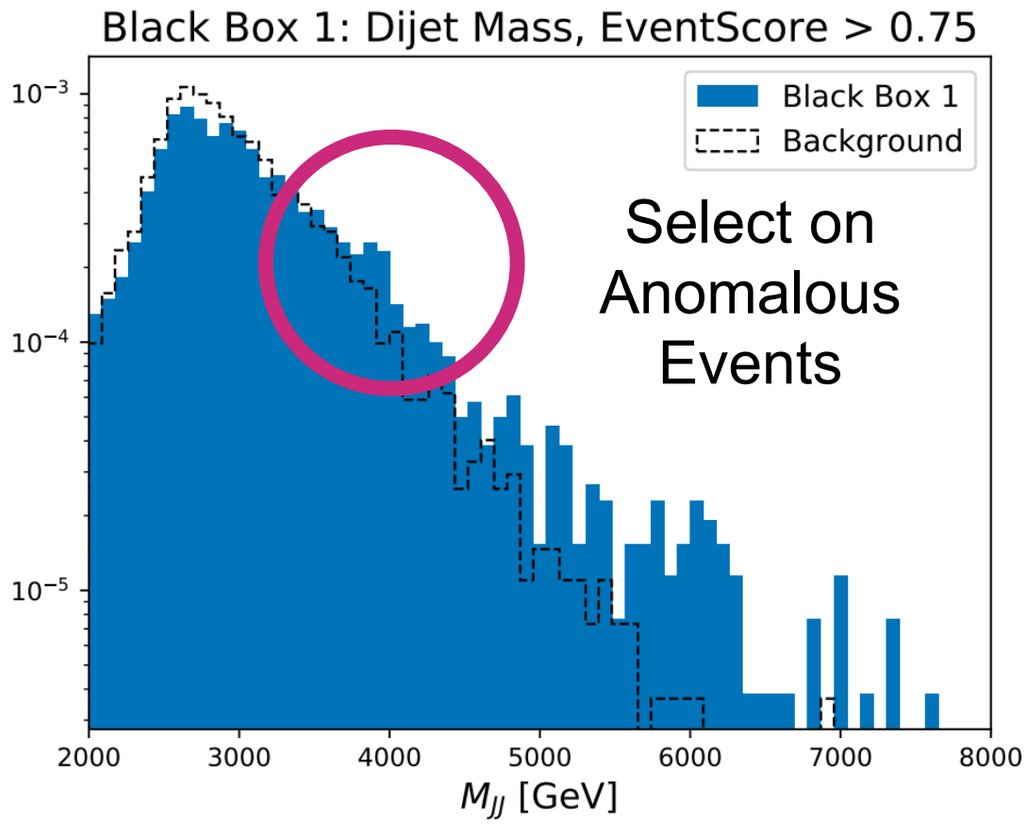
- VAE using particle inputs (RNN)



$$\mathcal{L}(t) = \text{MSE} + 0.1 \times \overline{p_T}(t) D_{\text{KL}}$$

$$\text{Anomaly Score} = 1 - e^{-\overline{D_{\text{KL}}}}$$

BB1 Dataset

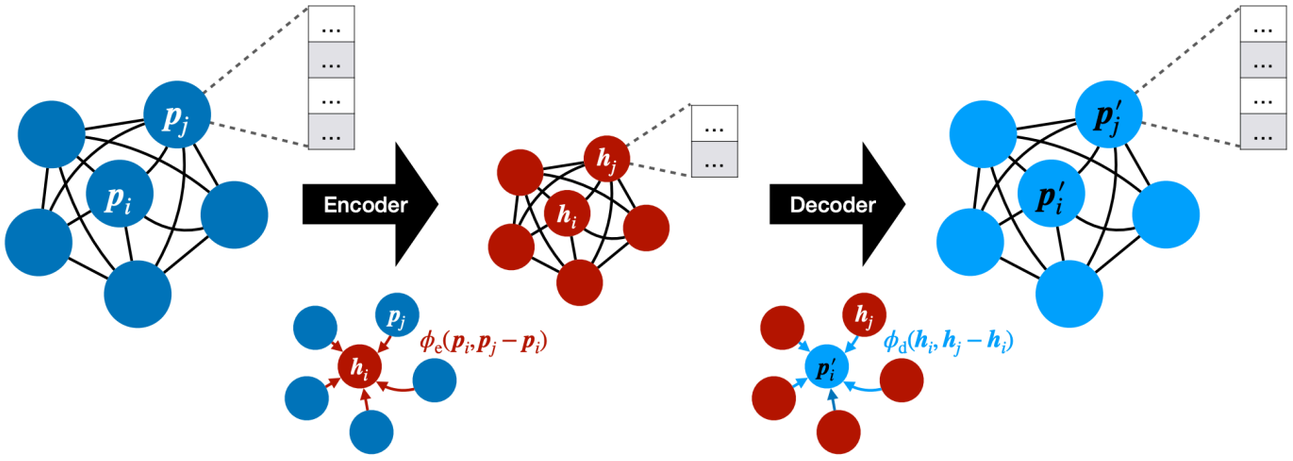


Signal Extraction : None

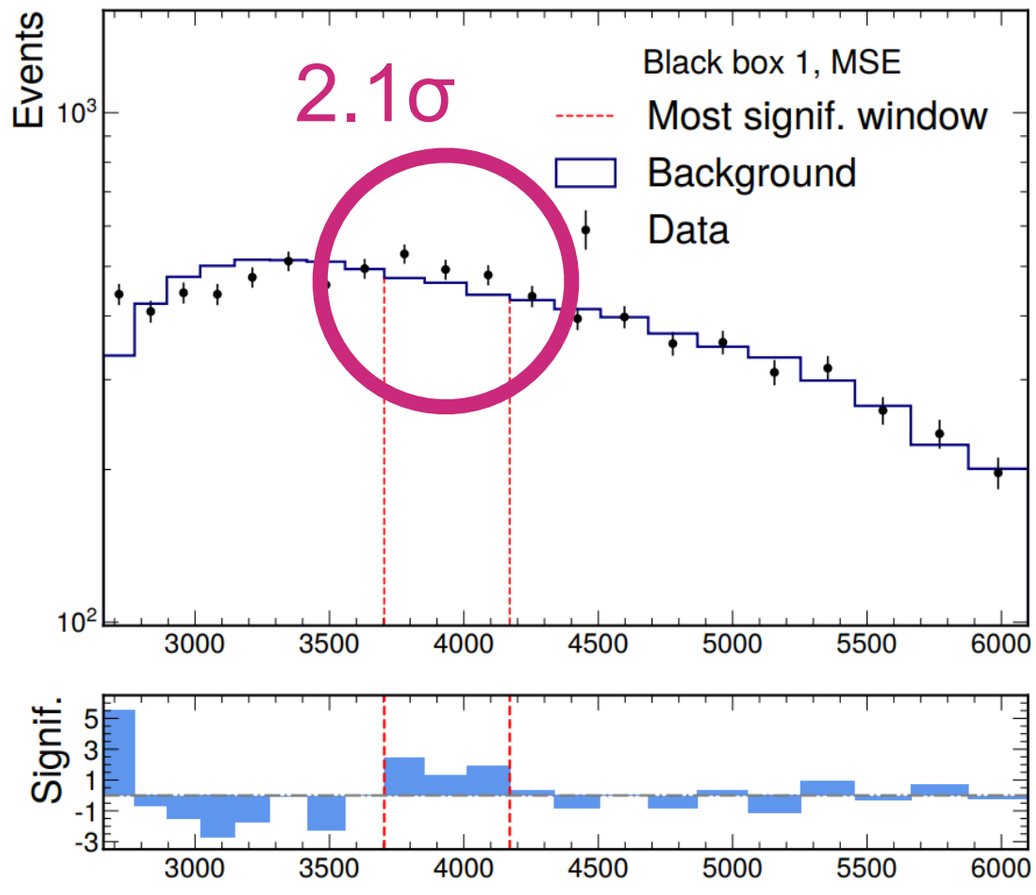
Take Away: Works but preparation of inputs is critical

Particle Graph AE

Inputs: Particle four vectors of the jet (Graph w/correlations)



BB1 Dataset



- Build a GraphNN Autoencoder
  - Try with mean squared error loss

Signal Extraction : Bump Hunter Algo

Take Away: No good handle on loss

Weak Supervision

How do we separate two samples (one with anomalies)

Sample A

VS

Sample B

Difference:

Strategy: Train the data in A agains B

Challenge:

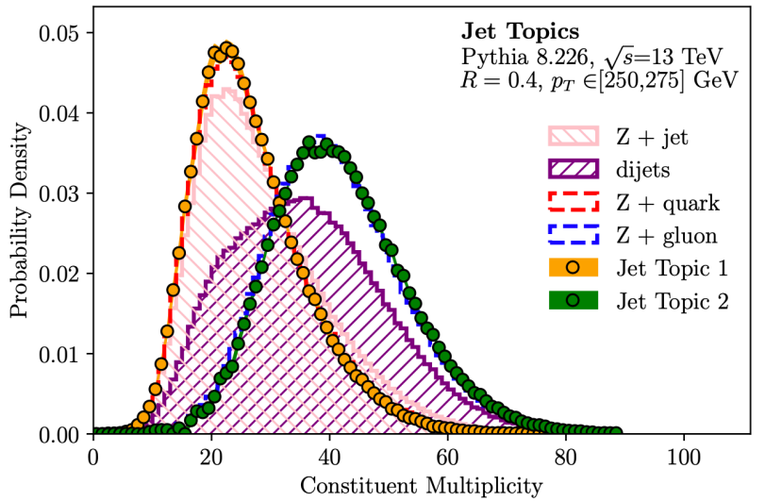
Must all be same in A and B

More realistic example

How do we train samples with variations of populations of an anomaly

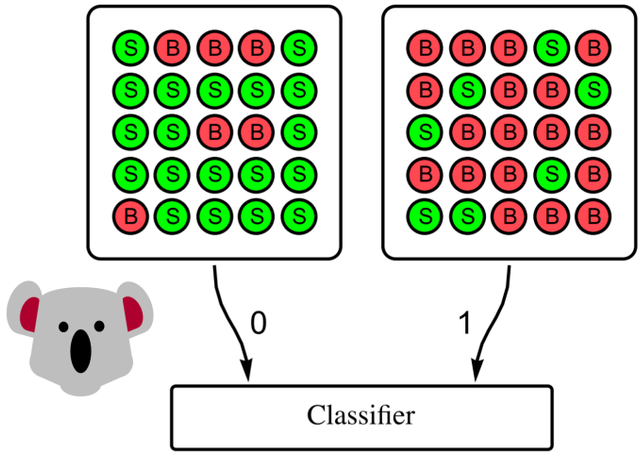
Training Strategies

Topic Modeling/ Clustering



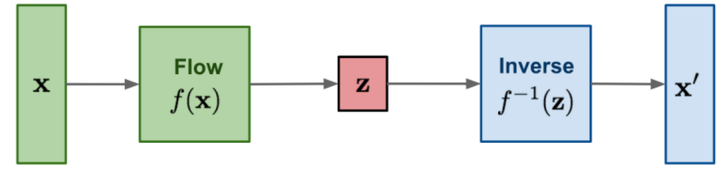
Split a histogram into multiple distributions by looking for separate regions

Classification W/O Labels



Separate out Sample 1 from Sample 2 by hidden signal

Likelihood Discrimination



$$p(x|x_c) = \pi(f_{x_c}(x)) \left| \det \left(\frac{\partial f_{x_c}(x)}{\partial x} \right) \right|$$

$p(x \mid x \in A)$

$p(x \mid x \in B)$

$$R(x|m) = \frac{p_{\text{data}}(x|m)}{p_{\text{background}}(x|m)}$$

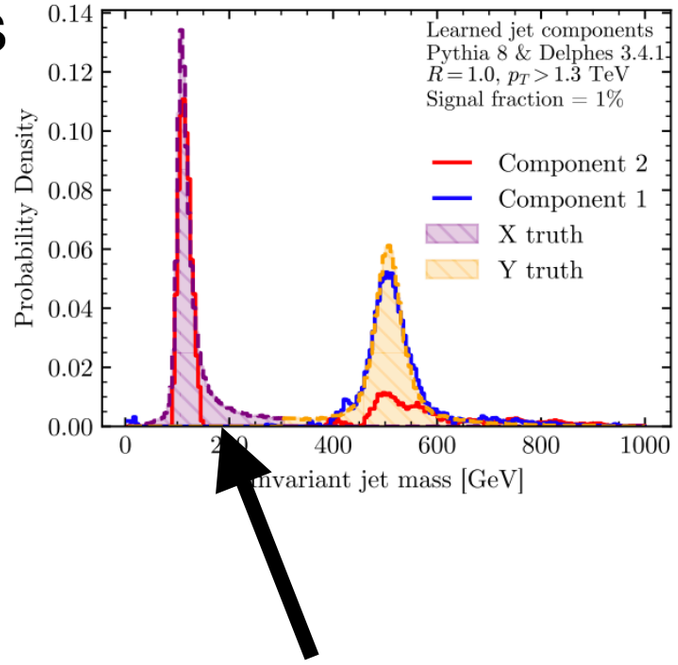
Factorized Topics

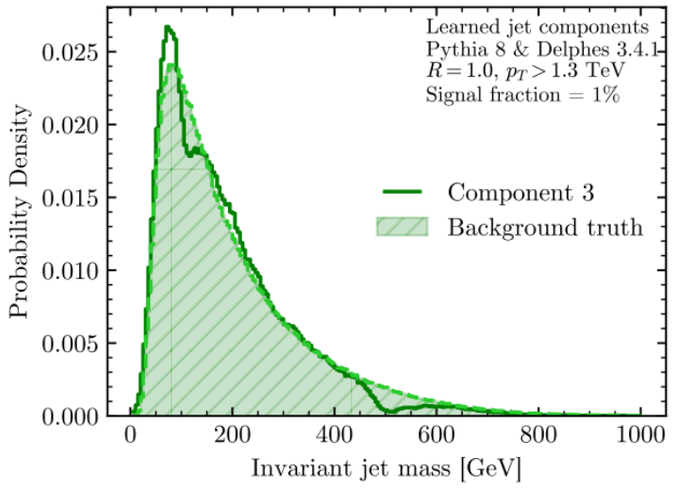
Inputs: Jet mass of each jet

R&D Dataset

- Factorization: each jet mass distributions can be factorized

- QCD composition is the same for leading and subleading





Use leading and trailing jet masses to make “topics”

Solve for the jet mass 1 and 2 that yield 3 distinct categories

Signal Extraction : None (did not work on BB1)

Take Away: Breaks down with small signal

LDA

Inputs: Jet splittings from declustering

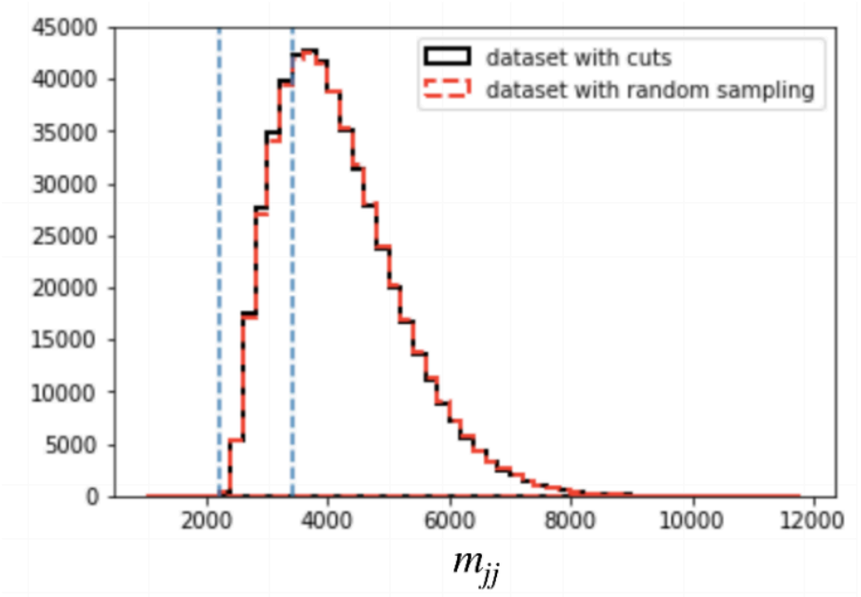
- Latent Dirichlet Allocation (LDA)
  - Decluster jet and use splitting info
  - Construct 2 hypotheses in data
  - LDA minimization to get 2

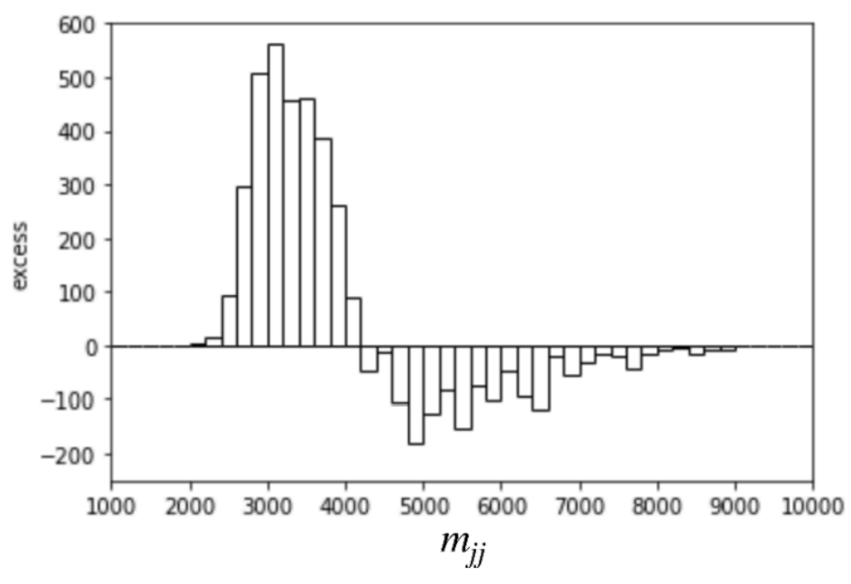
Compute likelihood of two hypothesis to be consistent

$$L(o_1, \ldots, o_N|\alpha) = \prod_{i=1}^{N} \frac{p(o_i|\hat{\beta}_1(\alpha))}{p(o_i|\hat{\beta}_2(\alpha))} .$$

Signal Extraction : None (did not work)

BB1 Dataset





Take Away: LDA benefits from many event observables

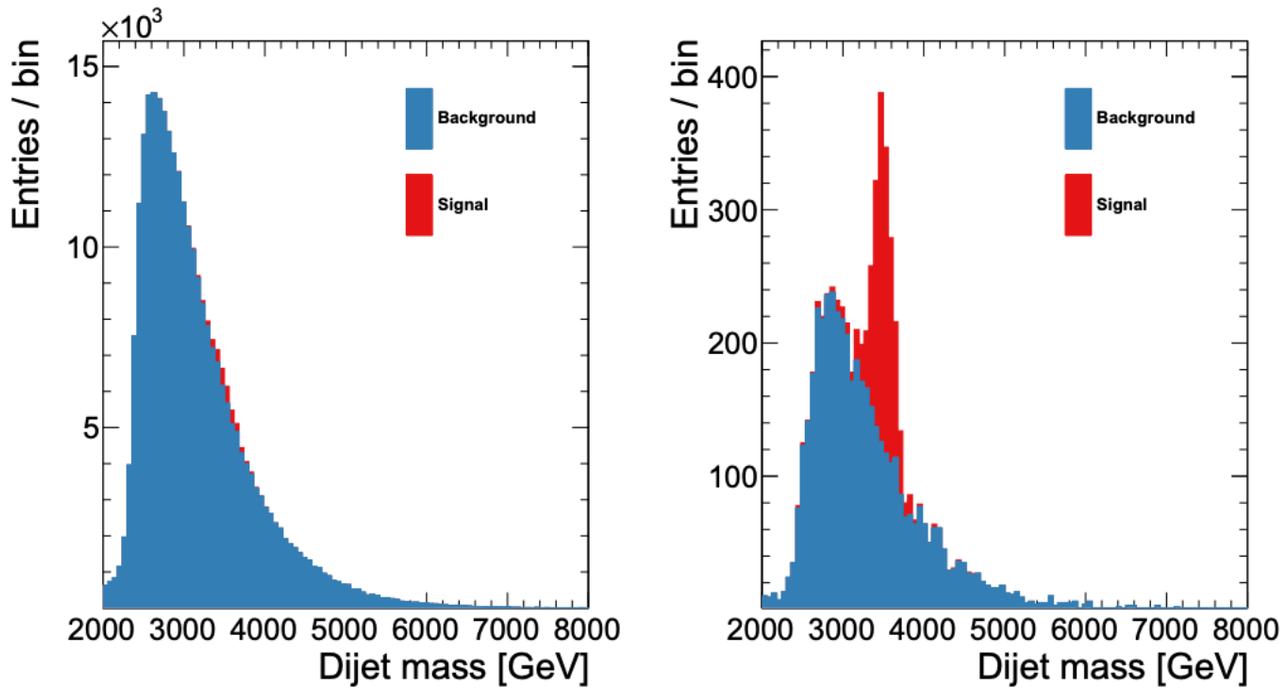
UCluster

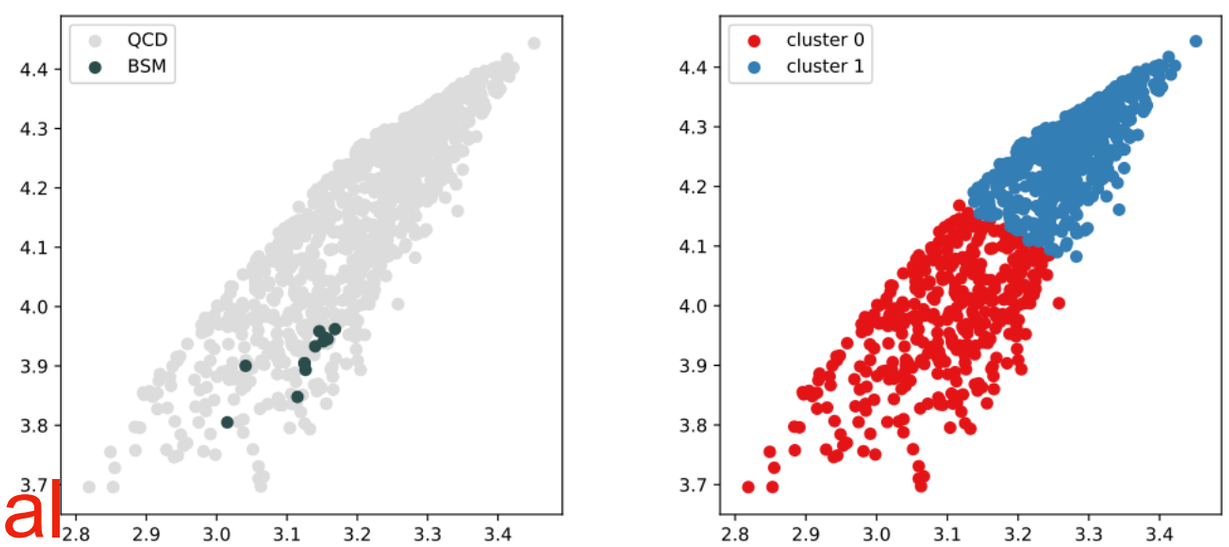
Inputs: Particle Objects

R&D Dataset

Train a supervised network for jet classification

Cluster in the latent space
Scan clusters for anomaly





Signal Extraction : No signal

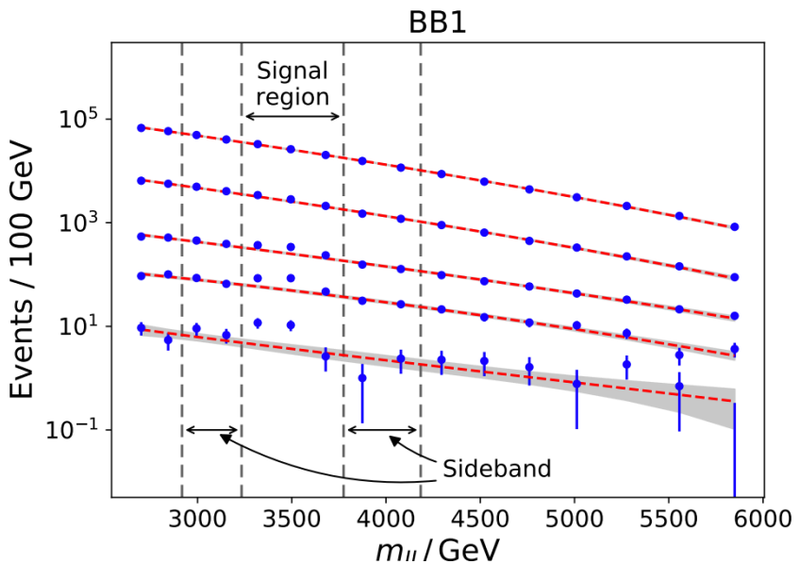
Take Away: Hard with small signal

https://arxiv.org/abs/2010.07106

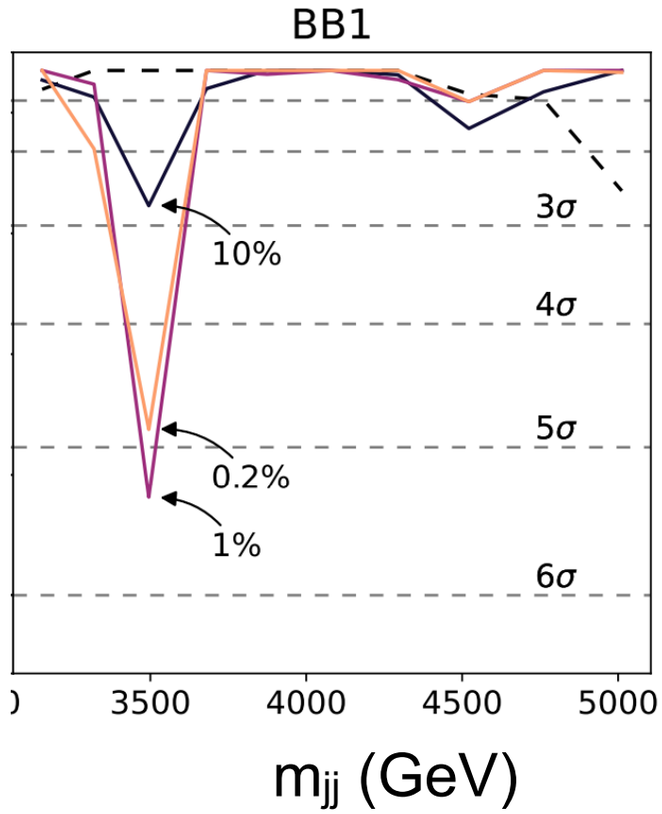
CWOLA

Inputs: High level features

- CWOLA modified from original paper
  - Mass inputs dimensionless



BB1 Dataset



Signal Extraction : Bump fit(5σ)

Take Away: Works but needed to correct dimension

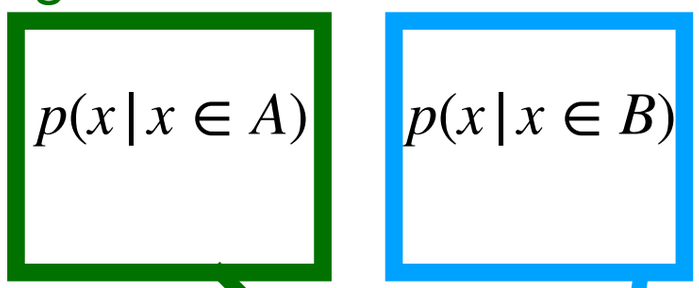
GIS(CWOLA+NF)

Inputs: High level features

GIS normalizing flows trained conditional on the mass distribution

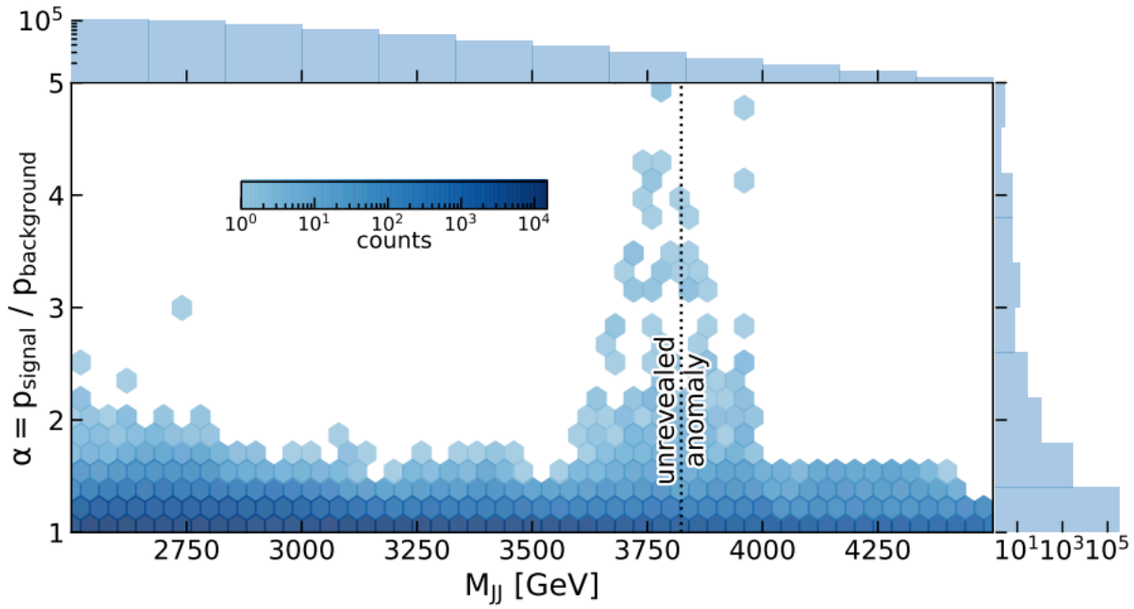
Scan mass window (250 GeV)
Compute likelihood ratio (below)

Signal mass Mass Side band



$$R(x|m) = \frac{p_{\text{data}}(x|m)}{p_{\text{background}}(x|m)}$$

BB1 Dataset



Large and significant signal

Signal Extraction : Note, but large signal

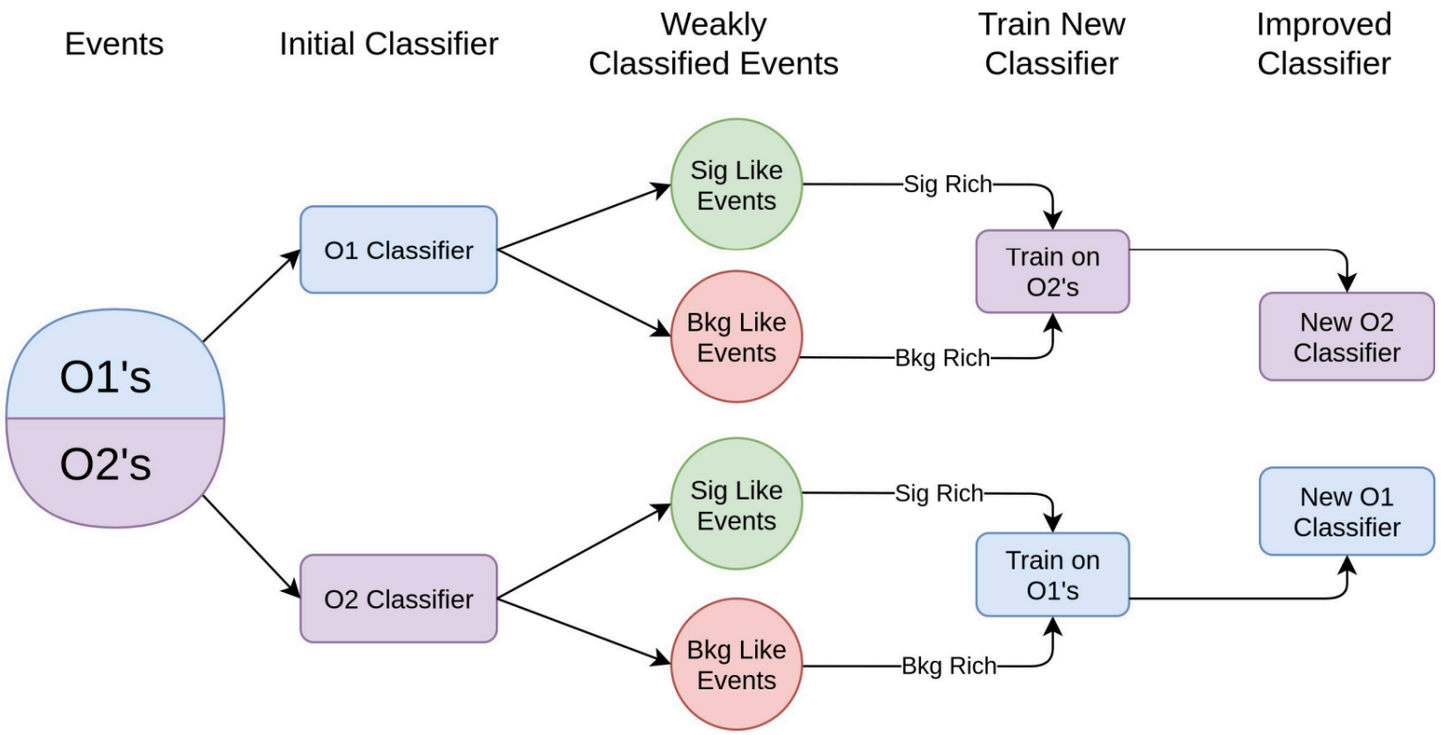
Take Away: Normalizing Flow can help CWOLA style approach

https://arxiv.org/pdf/2001.04990.pdf

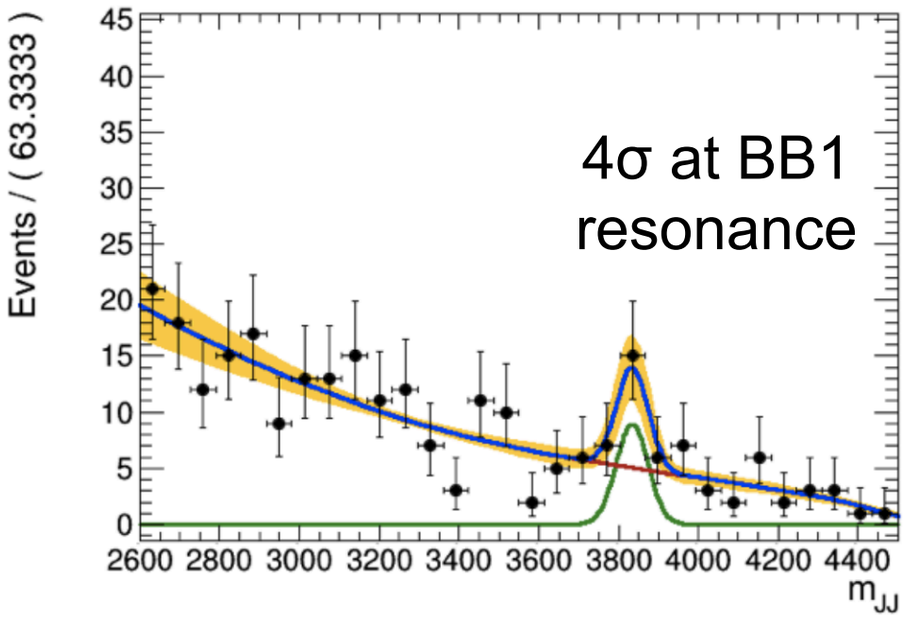
Tag N'Train

Inputs: High level features

Use dijet signature play one jet off the other
Start with an autoencoder on jet to split sample
Run CWOLA on other jet with split sample



BB1 Dataset



Would benefit more from mass decorrelation

Signal Extraction : Bump Fit

Take Away: Avoid mass windows by relying on the different jets

https://arxiv.org/abs/2002.12376

Semi-Supervision

Autoencoder

A large amount of unlabelled data

Supervised Training

A small amount labeled data

- Use supervised training to catch  and not 

(i.e. Find anamalous tulips not anomalous something else in LHC a detector glitch)

Training Strategies

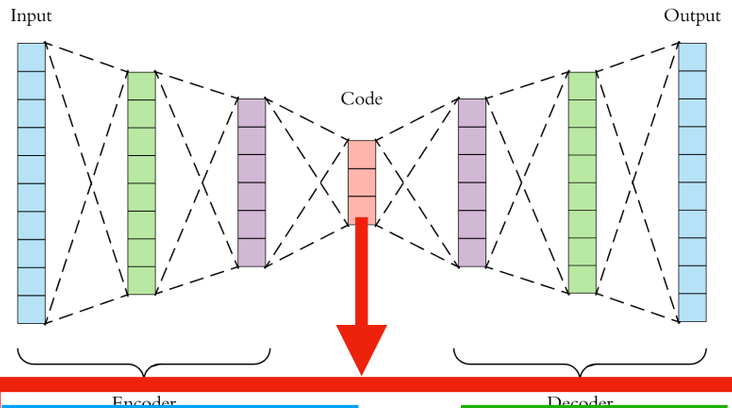
Just do a supervised training

Wrong Signals vs Background

Search for new physics by using an incorrect signal

Use classifier to isolate

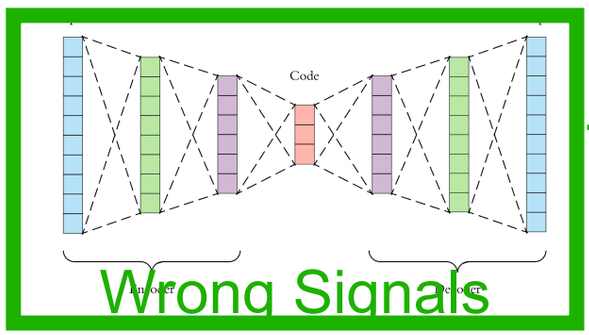
Use the latent space for autoencoder/ supervised

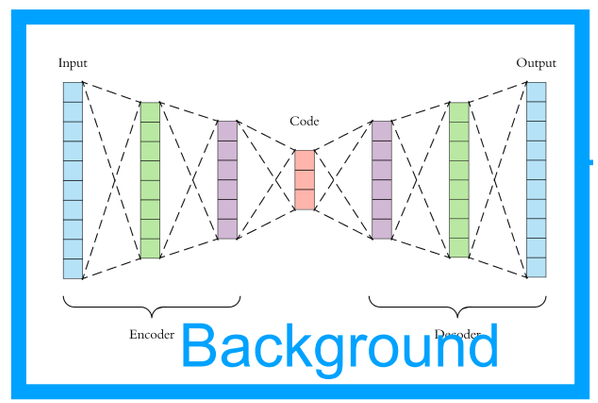


Background vs Wrong Signals

Use classifier loss for search

Construct Space from autoencoders on sig/bkg





Just Training

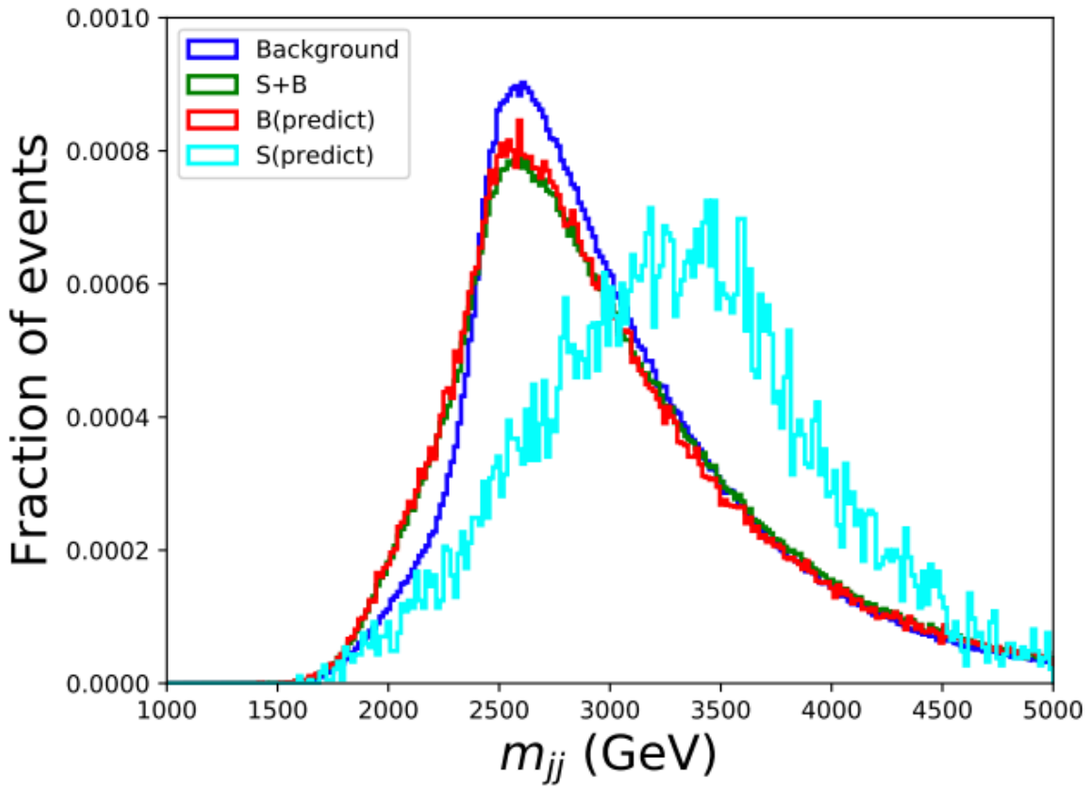
Inputs: High level features

Use R&D dataset and do a fully supervised training

Use the output discriminator

Try to see a signal from that

BB1 Dataset



Two submissions tried

No Significant excess in either

Signal Extraction : None

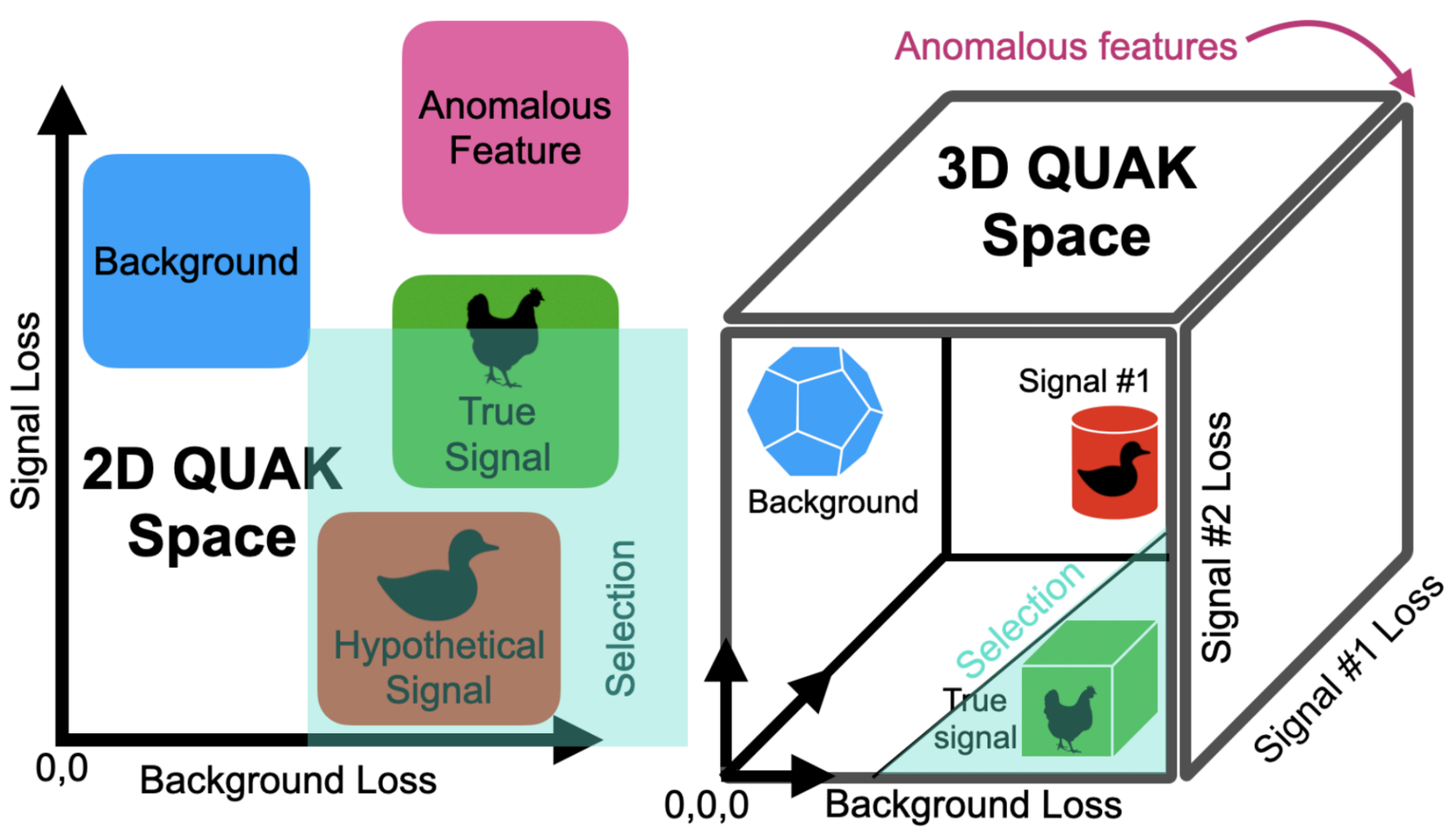
Take Away: Signal needs to be close to the hidden signal

QUasi-Anomalous Knowledge(QUAK)

Strategy: Train autoencoders on background and Signals

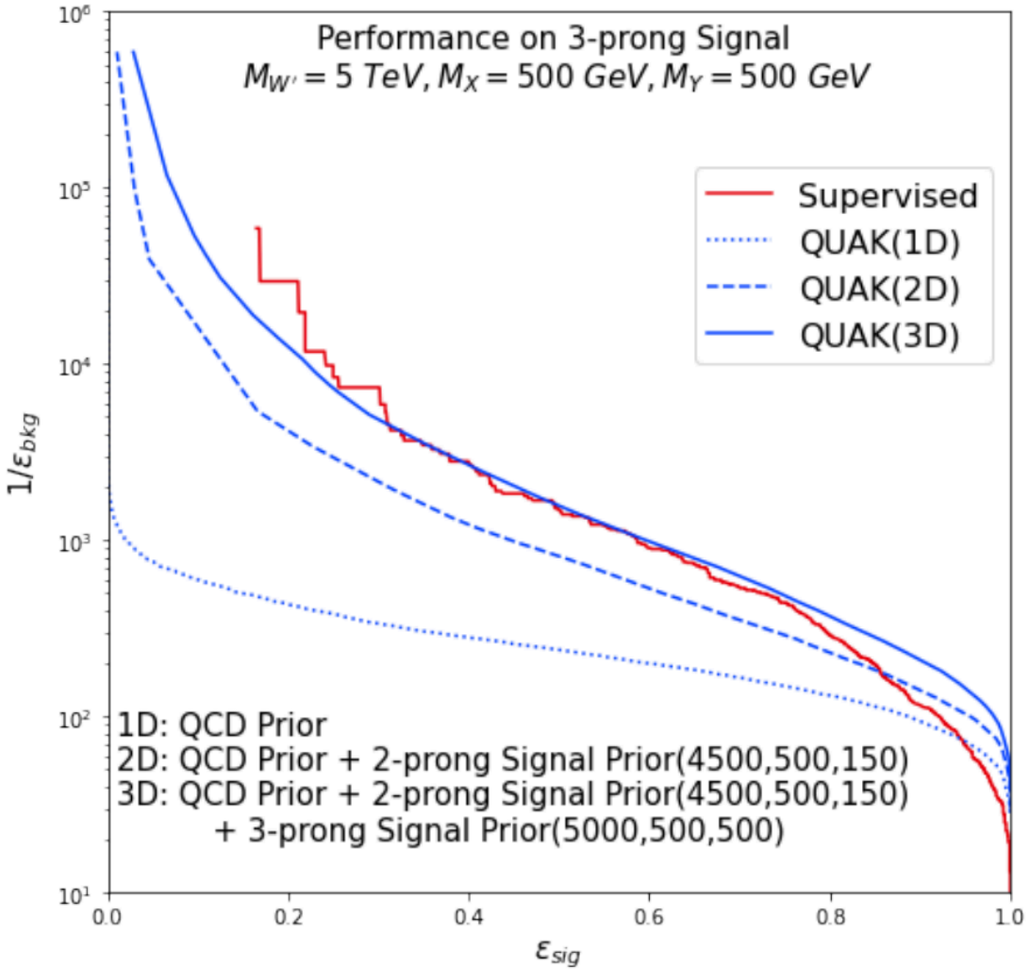
Choose a broad range of signals that capture physics of interest

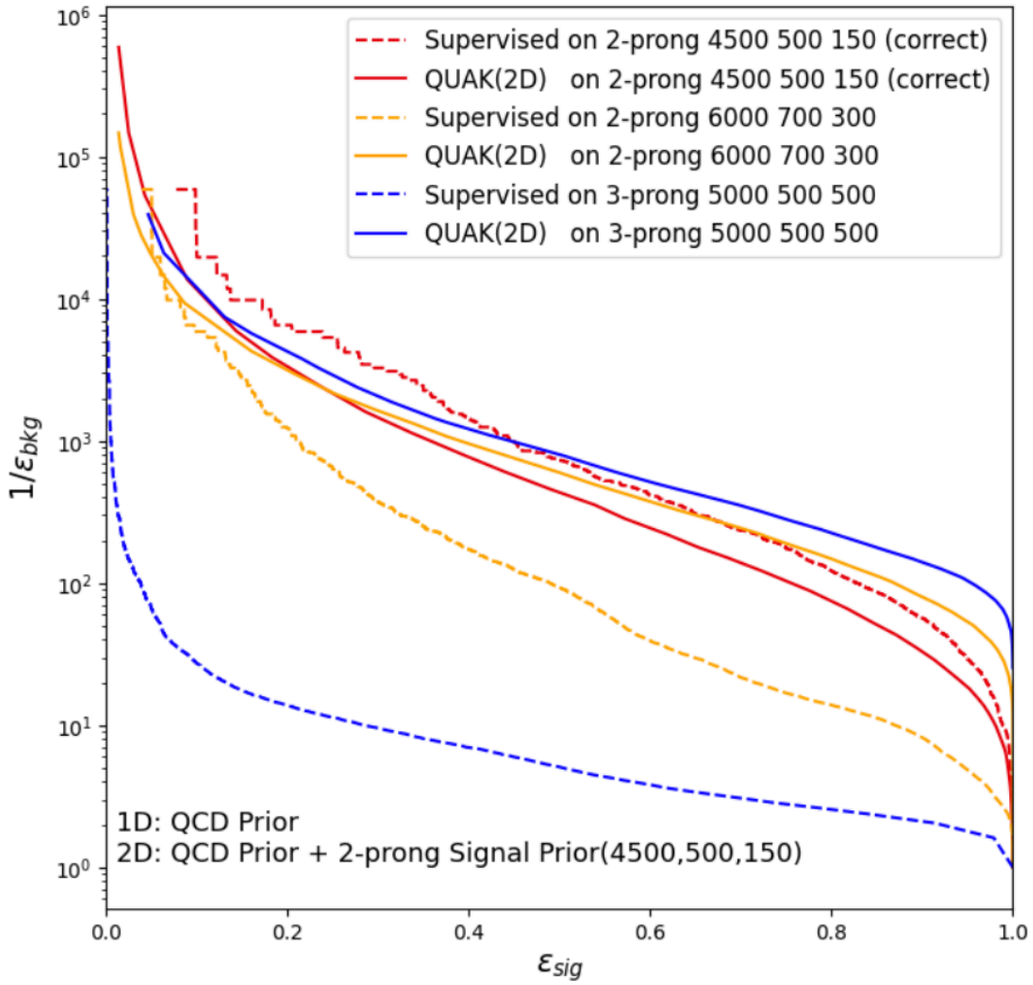
Probe the result space for physics-like anomalies



Use Normalizing Flow Autoencoders

QUAK





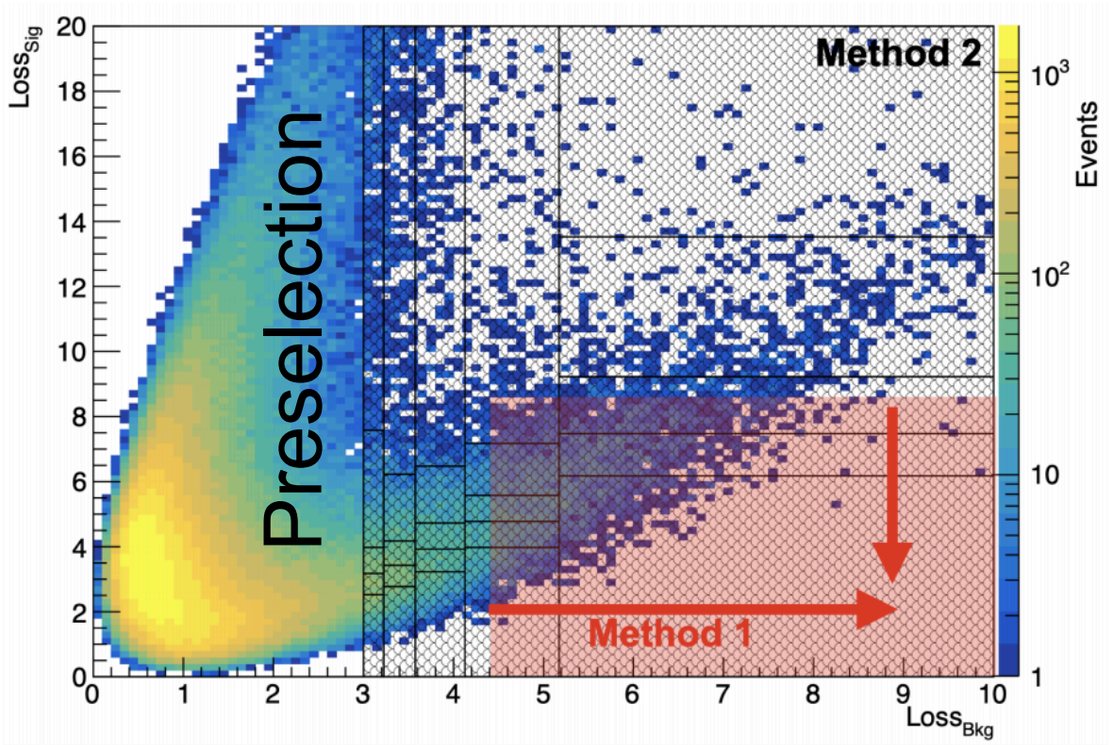
QUAK approaches or beats supervised NNs when signal is similar

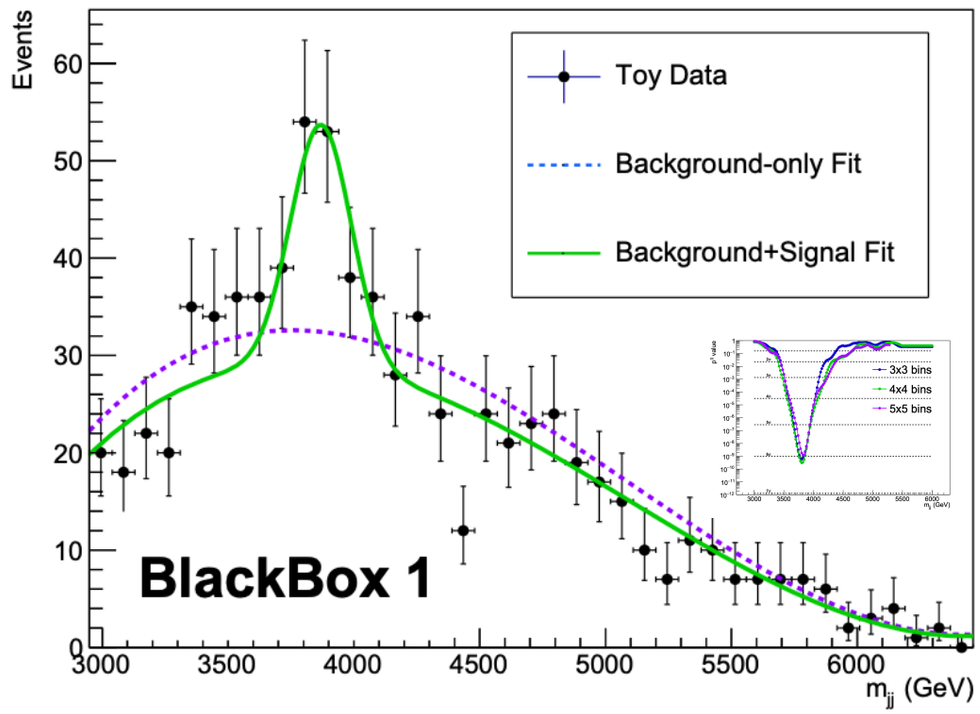
https://arxiv.org/pdf/2011.03550.pdf

QUAK

Inputs: High level features (Nsubjettiness/Jet masses/…)

BB1 Dataset





6.3σ significance

Mass dependence exists, but not large
Due to construction of loss space

Signal Extraction : Bump Fit in categories split across loss space

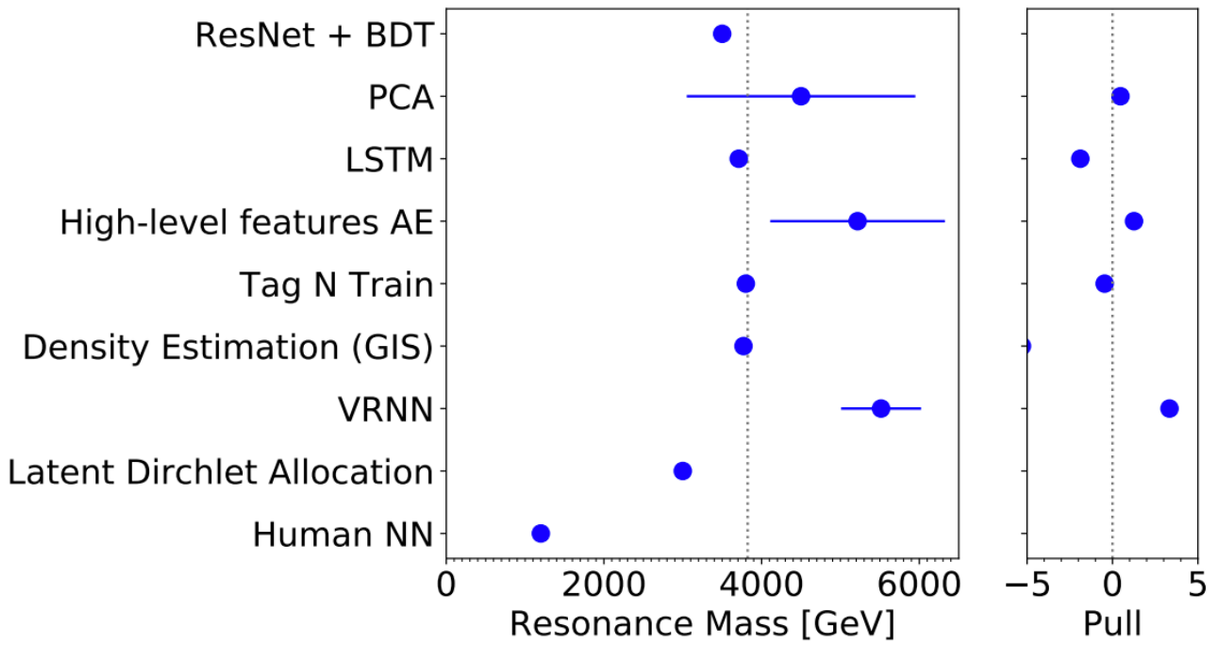
Take Away: Signal Libraries need to be close to hidden signal

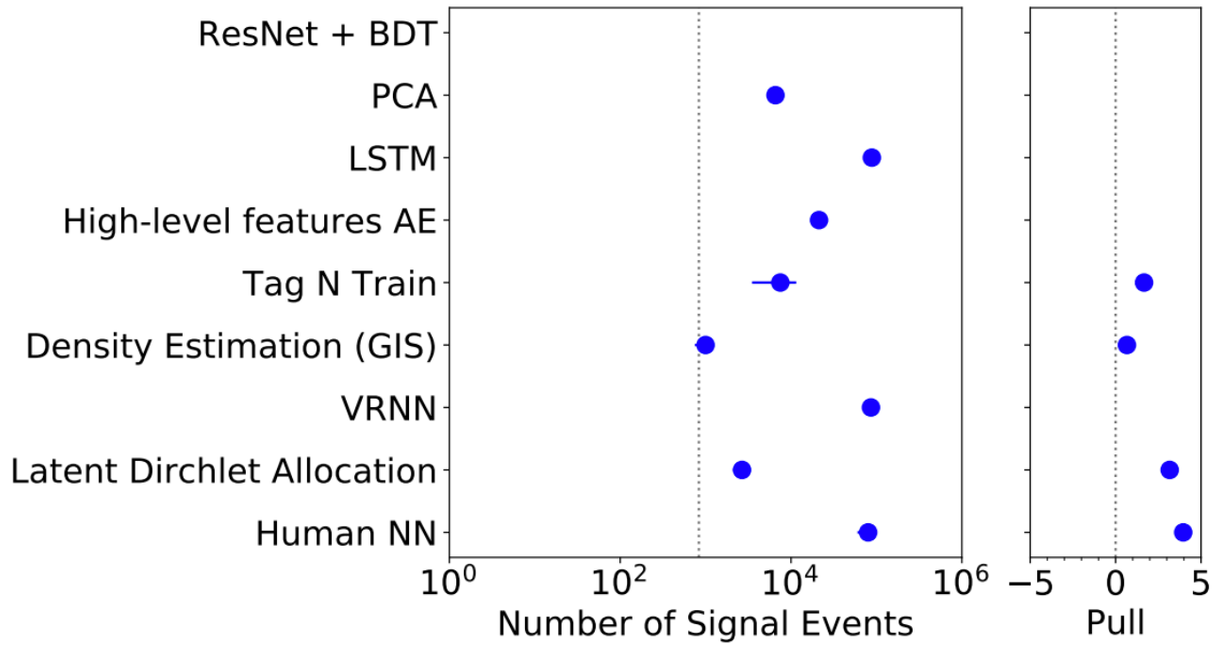
Training on Data

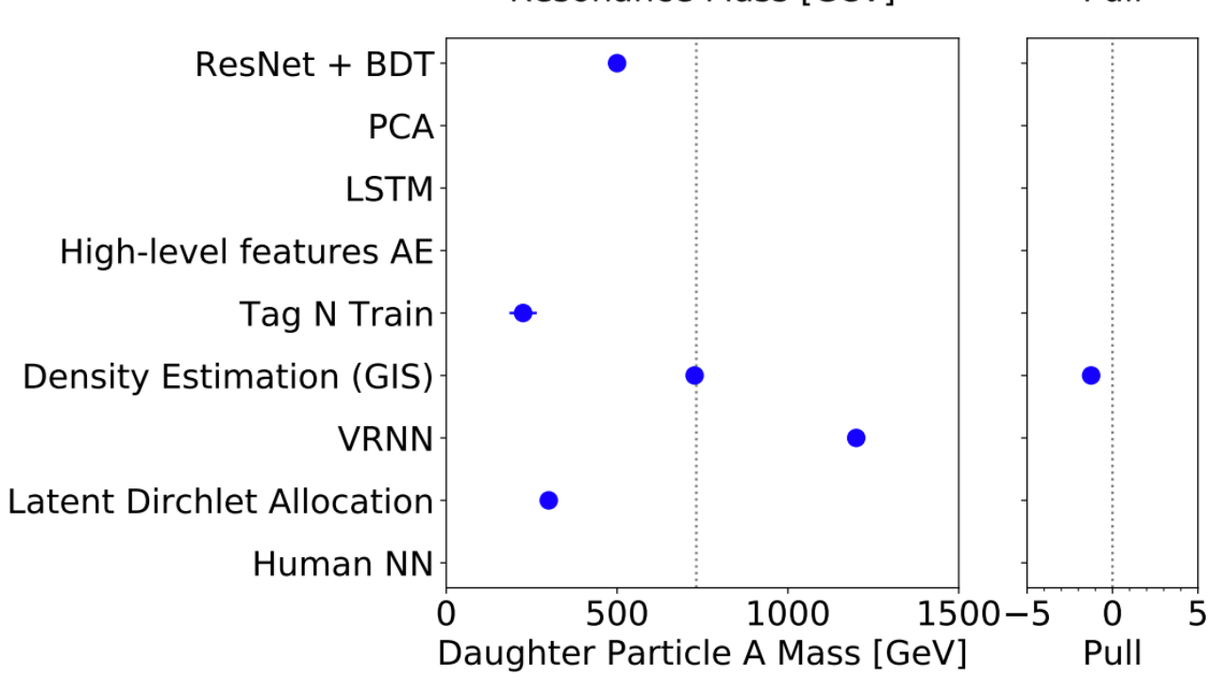
- Generally with anomaly approaches
  - There has been an emphasis to train on data
- Training on data simplifies our ability to process data
  - No need to correct for simulation/data disagreements
  - Regions where data/simulation don’t agree can be probed
  - No fancy methods to probe these regions w/complicated fits
- Training on data throws away some interpretability of result
  - Not clear what features may drive an access

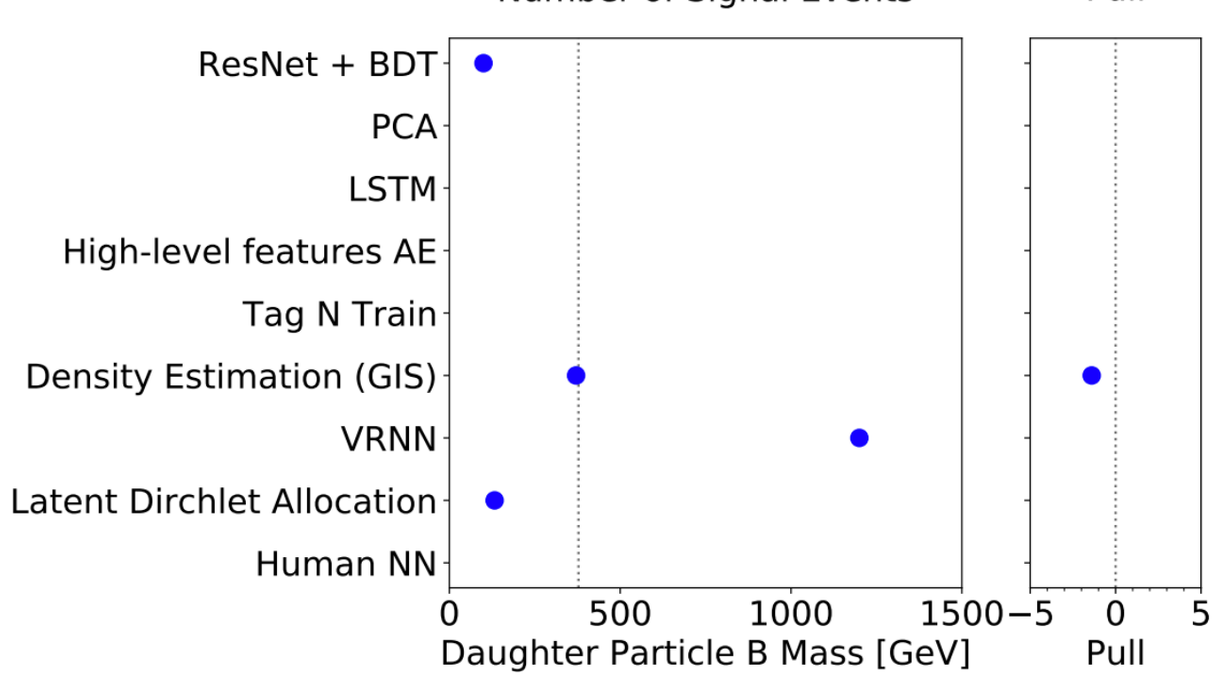
Black Box 1

A fun look at results









- Nobody found an excess in black box 3
- Black box 2 was empty

Black Box 2

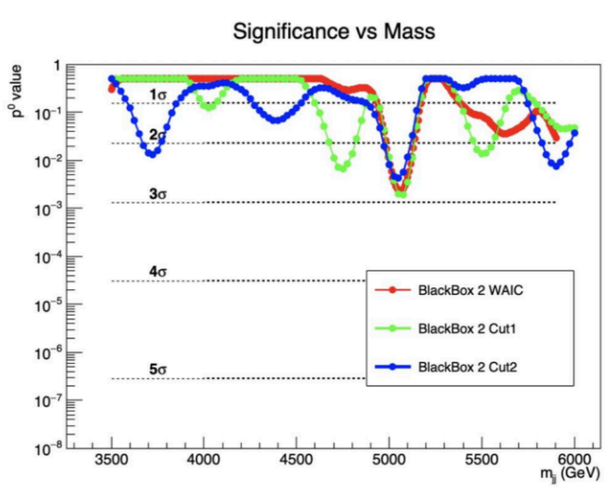
Black Box 2 - Predictions

- **PCA on high-level features (old):**
  A > BC with B > jj and C > jj
  m(A)=**4800** +- 100 GeV, m(B)=725 GeV, m(C)=125 GeV
  p-value / Signal events: 0.00764 / 89
- **VRNN (old):**
  A > BC with B > jj and C > jj
  m(A)=**4422** +- 722 GeV
  p-Value: 0.229181609 / Signal events: < 12k
- **Embedding clustering:**
  Z' resonance with mass **4600** GeV +- 17 GeV decaying to 2 jets
  p-Value: 0.0396 (1.8 sigma) / Event count: 76 +- 28
- **Latent Dirichlet Allocation (old)**

Our method extracts signal descriptions which appear convincing, however the classifier does nc identify a bump in the invariant mass spectra. Without this we were unable to determine that signal was present. The di-jet description extracted consisted of one jet of mass 350-400 GeV an another of mass 150-200 GeV. If the production of these states was non-resonant, we would b unable to find the signal with our method. Or if more than just di-jets were relevant to reconstru the invariant mass, we would also not be able to find it. Otherwise, ***we determine that no signa was present in the data.***

Reminder no signal

- **QUAK:**
  BB2 3sigma local evidence for resonance at ~ 5 TeV



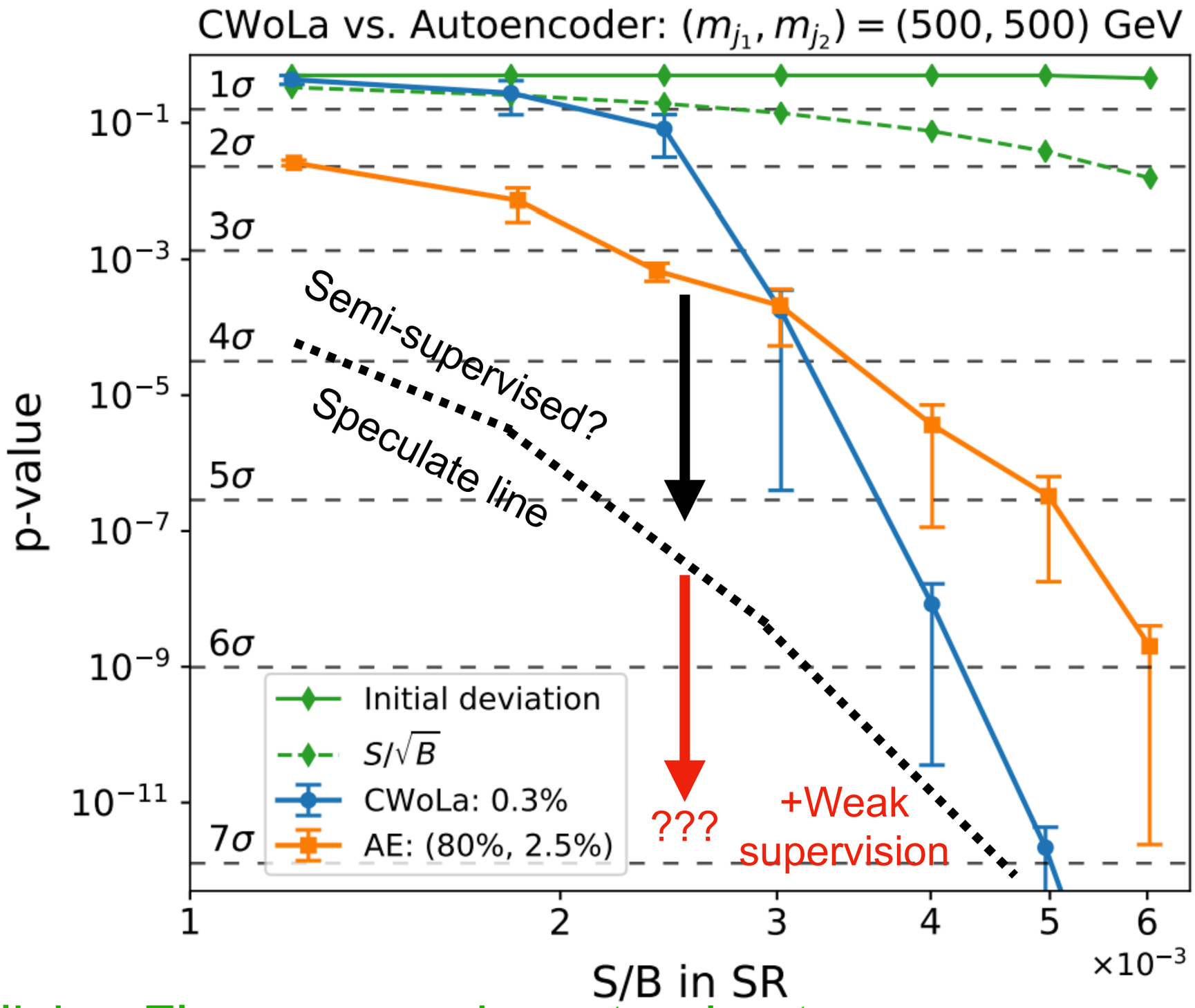
Sang Eon Park, LHCO 2020

- **M-flows and GAN-AE:**
  work in progress (inconclusive)
- **VRNN (new):**
  Hint of an excess at 4.2 TeV

Observations

- There is no catch all solution
  - Many of the best approaches combine multiple ideas
  - A diversity of approaches helps robustness
- LHC Olympics focused on resonant processes
  - Non-resonant processes make background extraction harder
  - Can we deal with complex topologies (such as black box 3)
- Data processing pipeline is assumed to be offline reconsturction
  - Could envision some approach in the triggers
- How can we actually compare sensitivities if we don’t have a model?

Performance Observations



Normalizing Flow approaches stood out

So did Observable based encoding(not sure why)

Preparing for open data

- The LHC Olympics dataset is a great tool to use
  - Sample is tractable and easy to put together
  - Shows the how a group can come together
- It would be fun to merge all of these strategies
  - Perhaps we could make a much stronger discovery?
- These kind of exercises are needed for the future

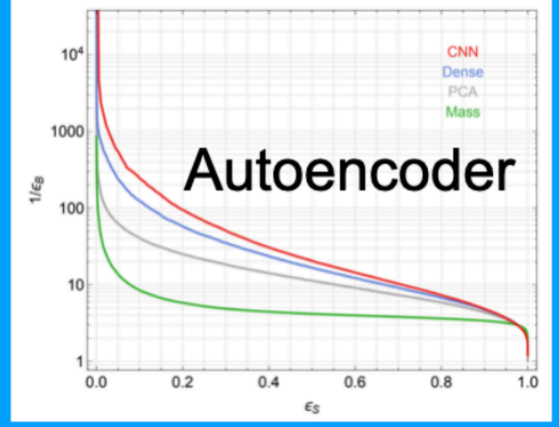
Strategy for future

- Its important to make simple tractable datasets
  - There are a lot of steps towards doing an analysis
  - There is also some black magic that goes in
    - Run your studies by somebody on the experiment
- A broad range of anomaly detection algorithms exist
  - Anomaly detection will not just come from one method
  - The diversity of approaches aids in building a robust discovery

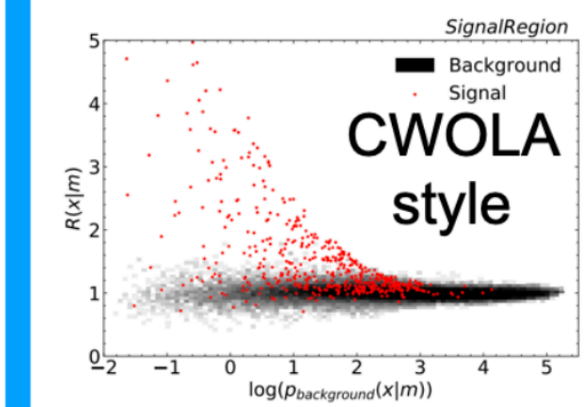
Thanks to the organizers for inviting me!

Playing with Prior

Prior Free → Fully Supervised



Knowledge of Background

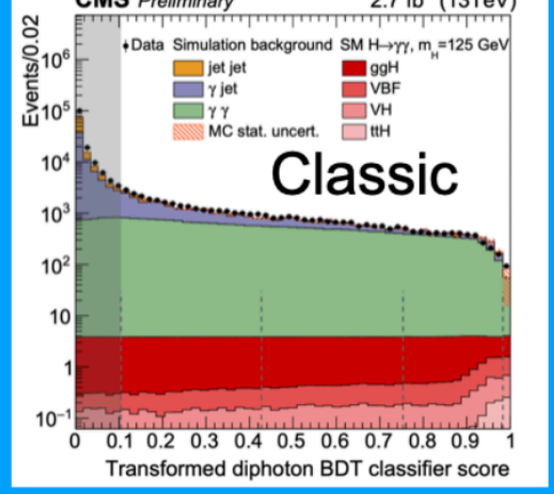


Signal in (Mass) Window

Semi-Supervised Approaches

QUAK

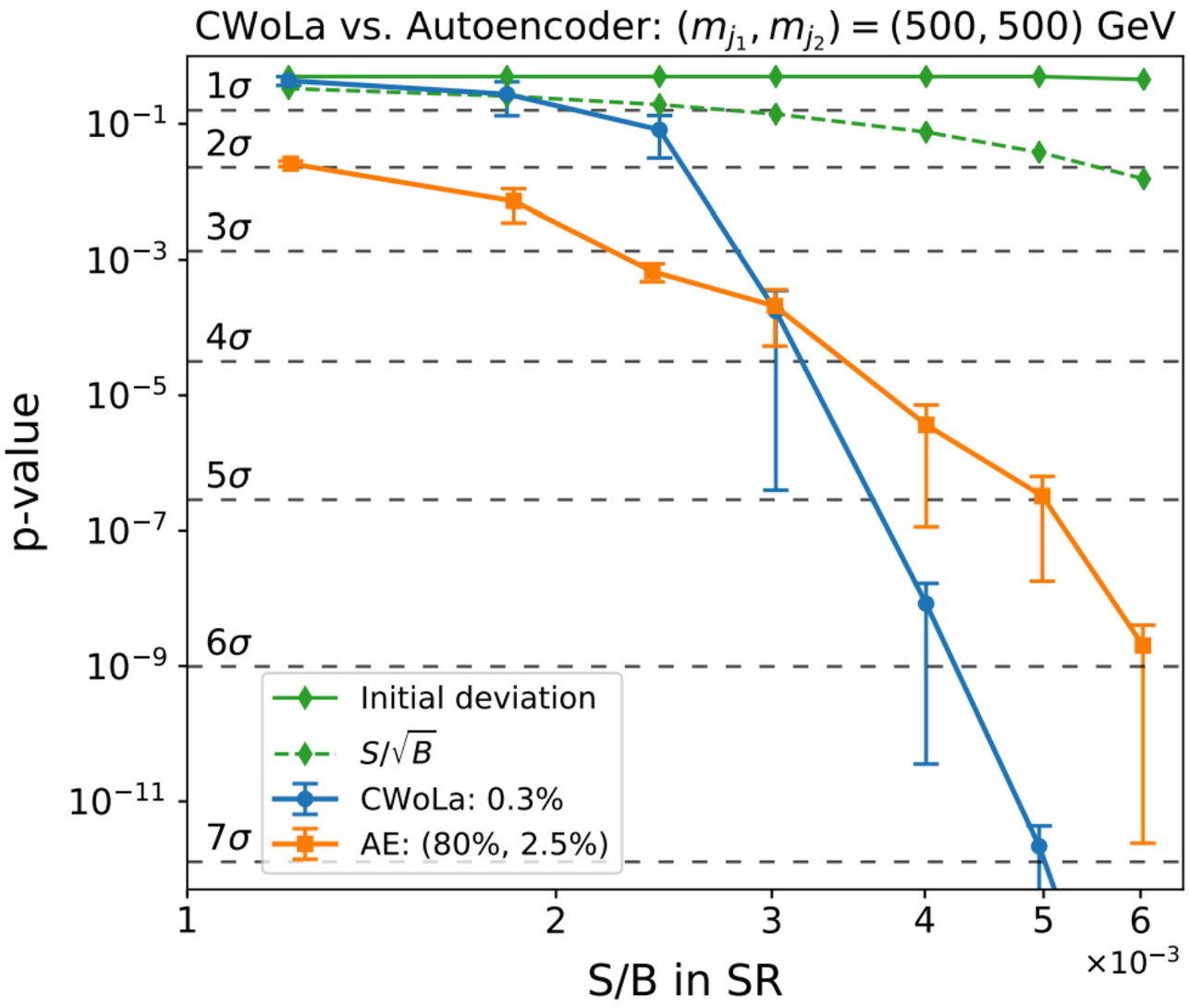
If it quaks like a duck?



Fully Supervised

Observation

Inputs: High level features



CWOLA works really well for large signals

But for small signals Autoencoders tend to win

You need enough events in your data to separate them

Signal Extraction : Bump fit

Take Away: Works but needed to correct dimension

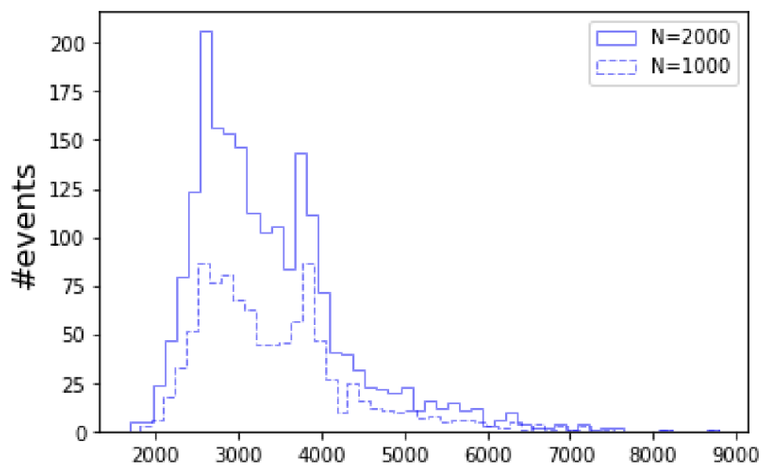
Variation of Encoder

Varying the encoder architecture
can allow for a broad range of possibilities

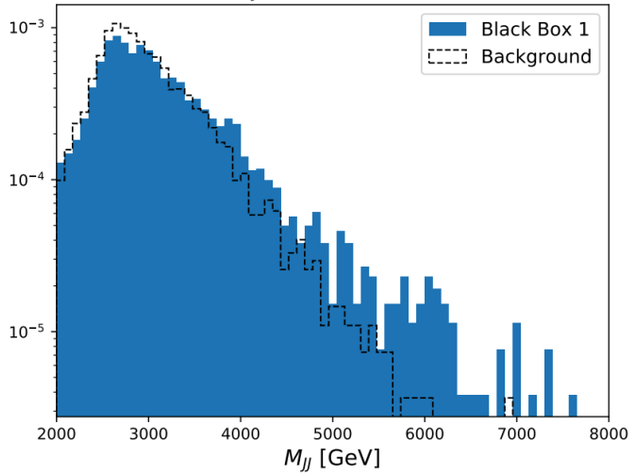
Derived Inputs

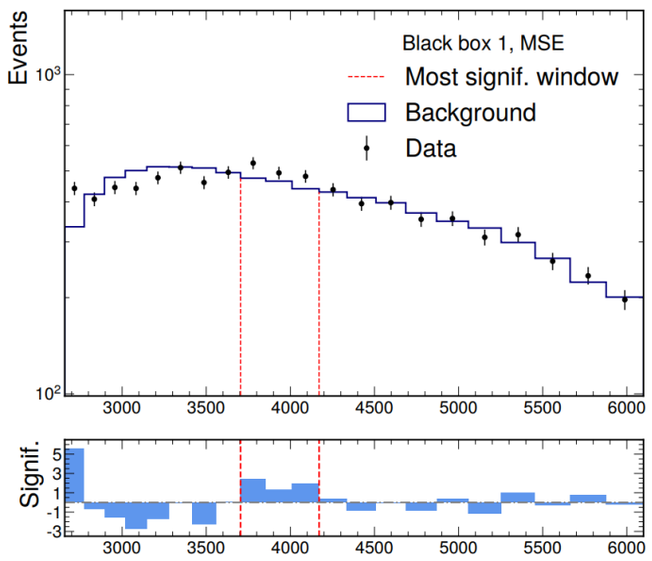
Particle Inputs

Graphs



Black Box 1: Dijet Mass, EventScore > 0.75





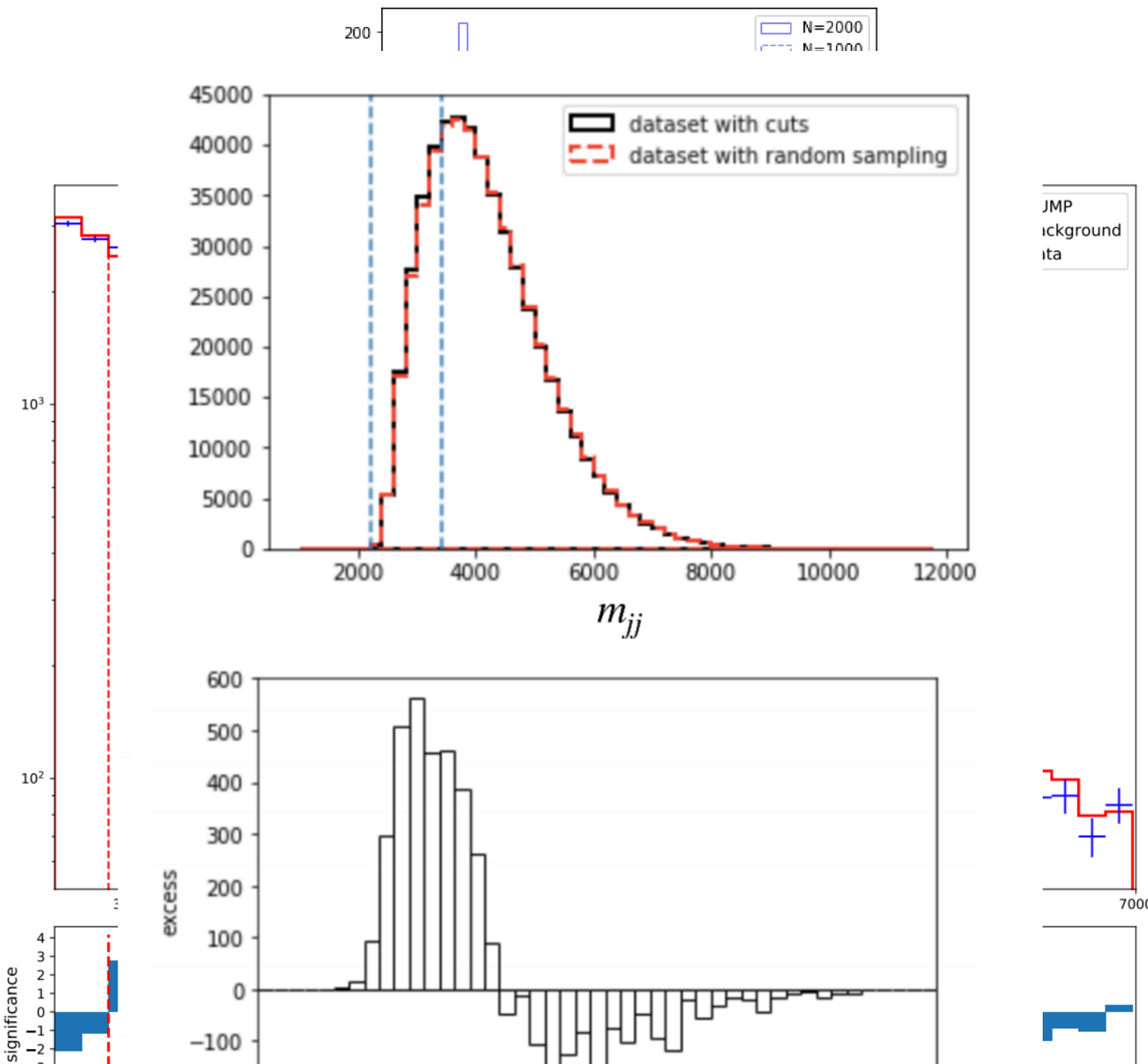
Variation of Architecture

Varying the encoder architecture
can allow for a broad range of possibilities

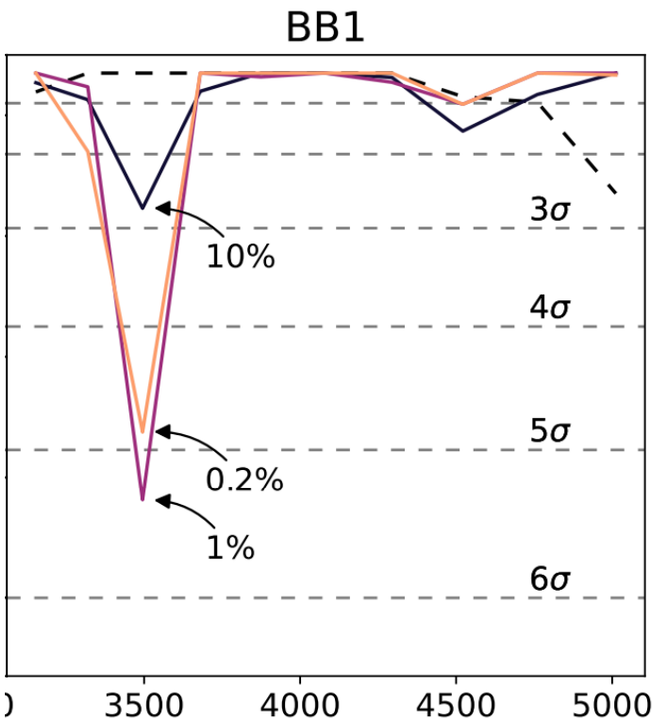
Autoencoder

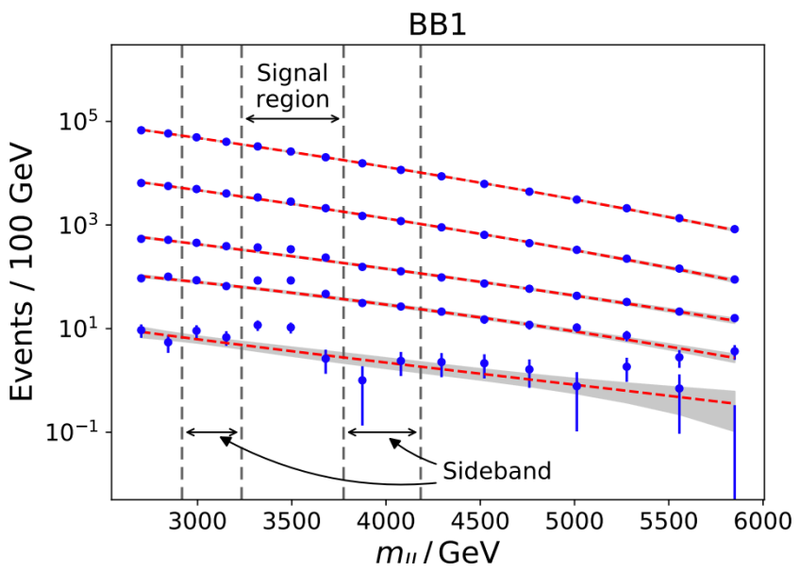
Variational Autoencoder

Normalizing Flow



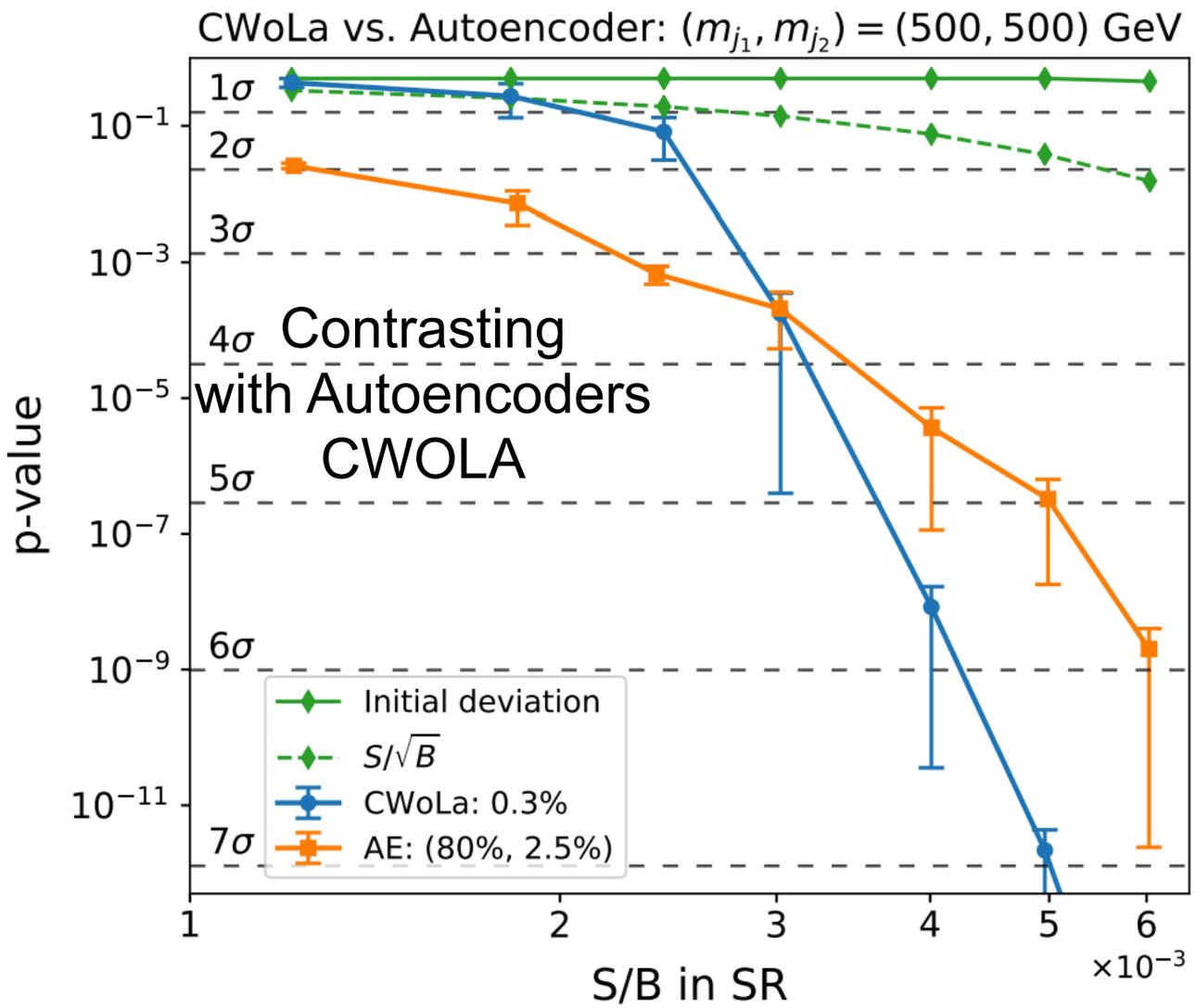
CWOLA style approach





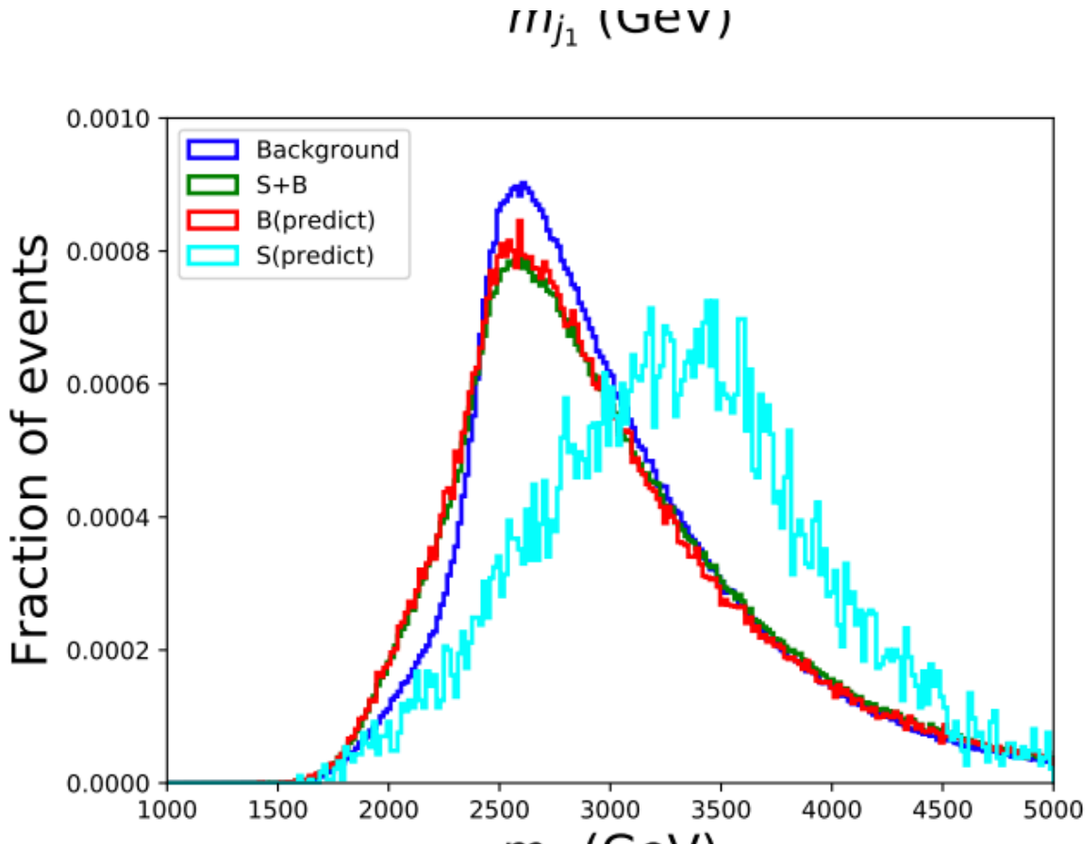
Excess at 3500 instead of 3800

- Running just a training got it to work
  - Was able to observe 5 standard deviations



Method 12:Deep Ensemble

- Use R&D dataset and do a fully supervised training
  - Use the output discriminator
  - Try to see a signal from that
- Try with both a CNN on jet images and BDT on observables



Observation:Low noise robust density estimation is key

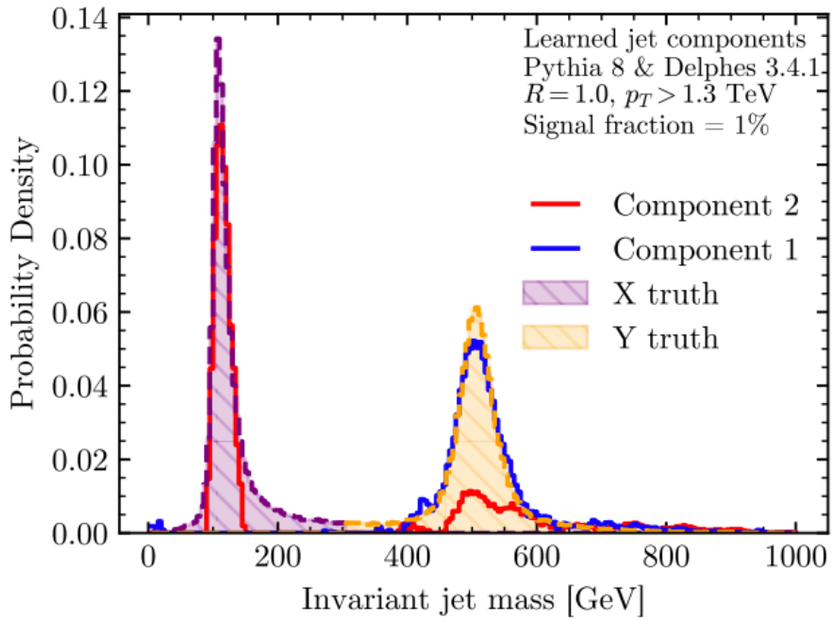
Method13:Factorized Topics

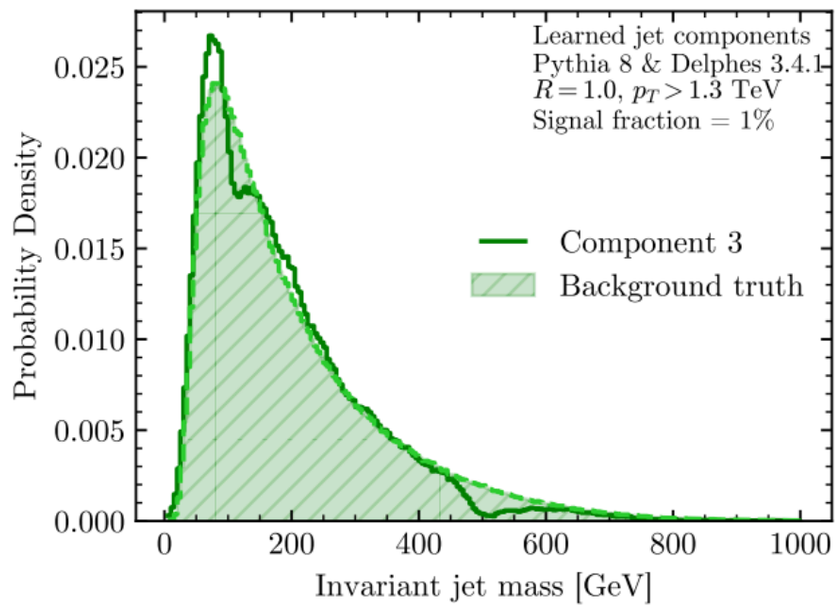
- Sample independence: each jet of a dijet can be treated as independent and for QCD its composition is the same for leading and subleading

- Factorization: jet mass distributions can be factorized

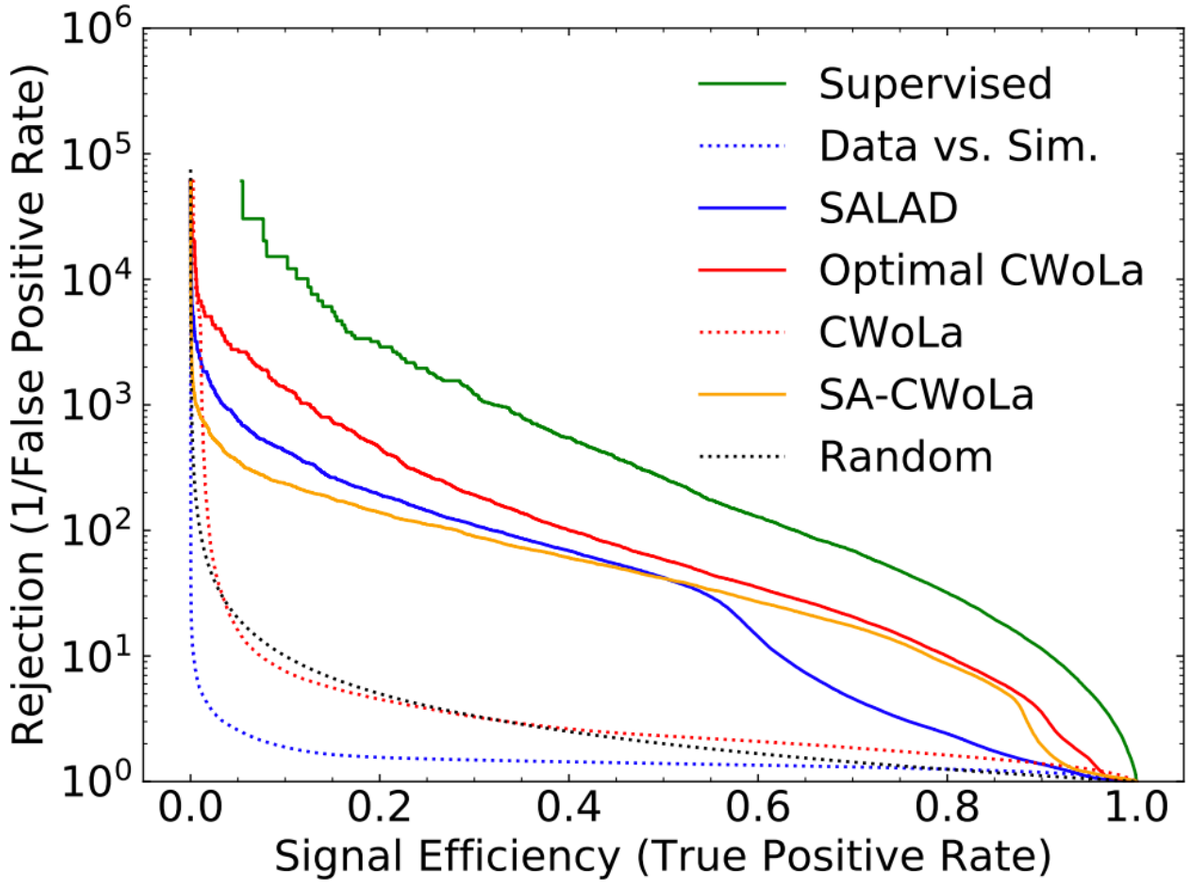
-

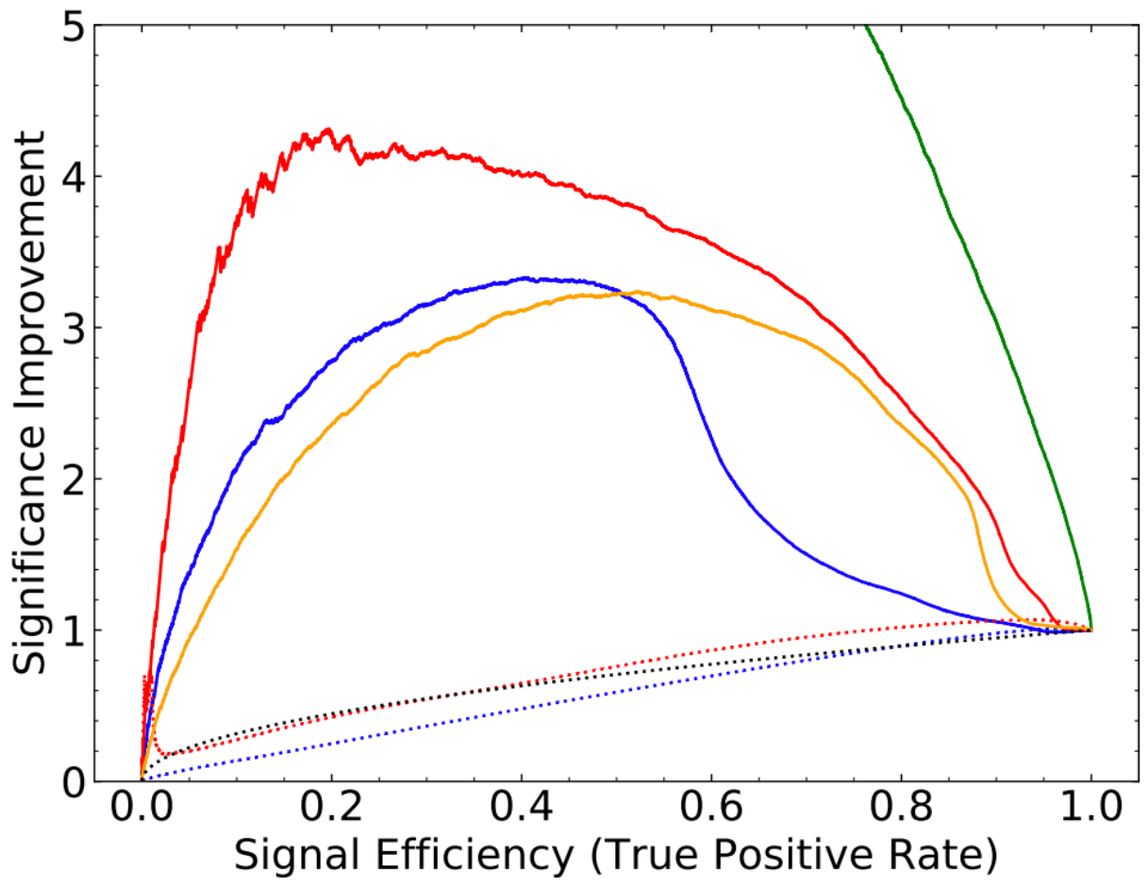
$$\mathcal{L}(\mathbf{x}_1, \mathbf{x}_2) = \frac{f(\text{signal}) \cdot p_{\text{signal}}(\mathbf{x}_1, \mathbf{x}_2)}{f(\text{background}) \cdot p_{\text{background}}(\mathbf{x}_1, \mathbf{x}_2)} = \frac{f(X,Y)\, p_X(\mathbf{x}_1)\, p_Y(\mathbf{x}_2) + f(Y,X)\, p_Y(\mathbf{x}_1)\, p_X(\mathbf{x}_2)}{f(\text{QCD, QCD})\, p_{\text{QCD}}(\mathbf{x}_1)\, p_{\text{QCD}}(\mathbf{x}_2)}$$

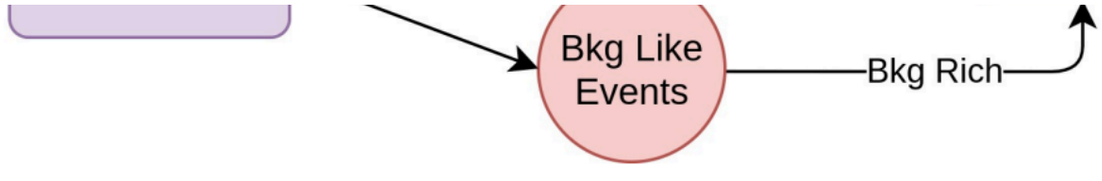




Method 10: Salad+CWOLA



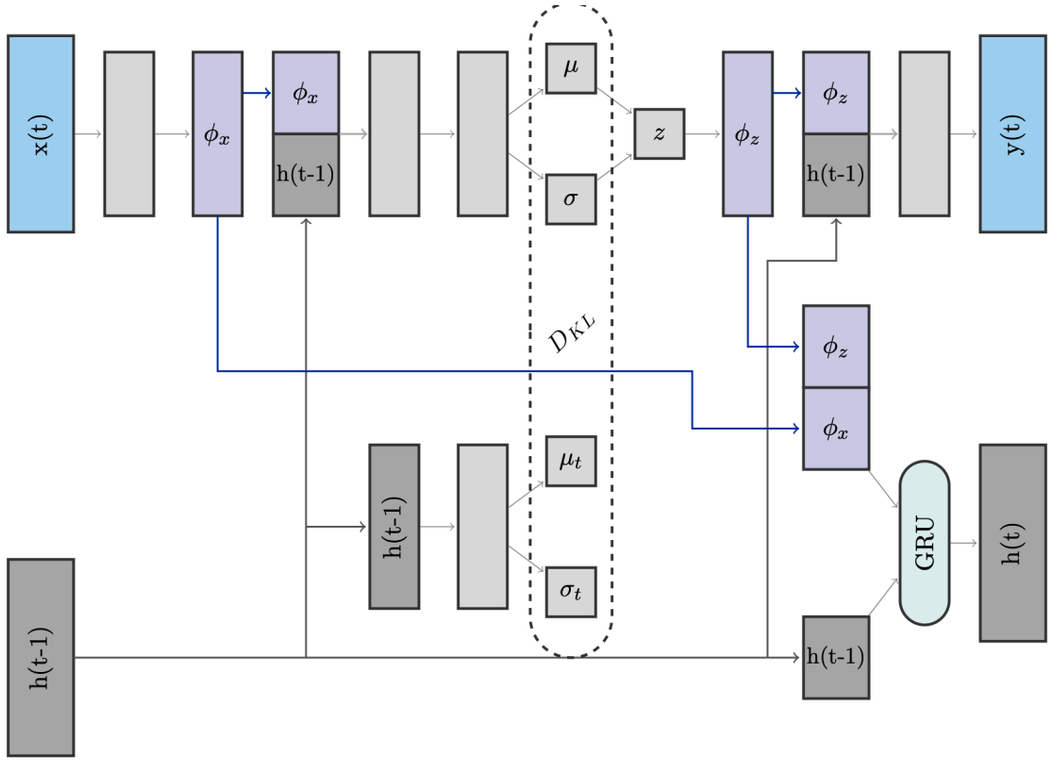


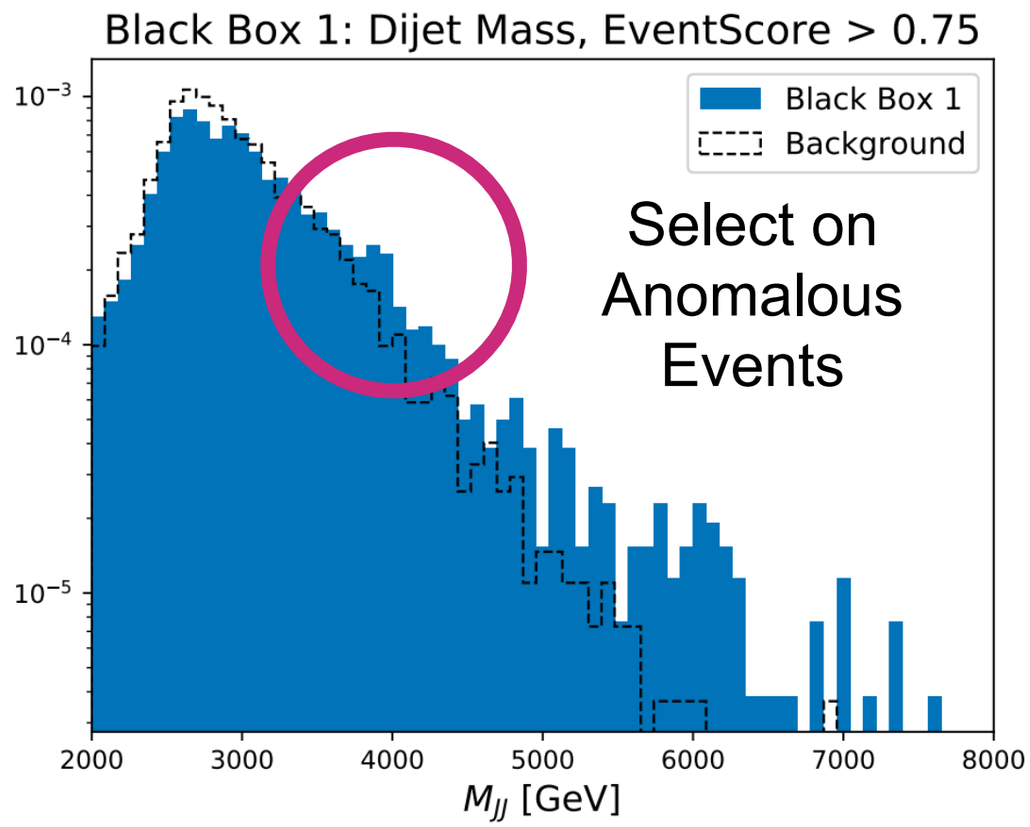


Observation:Works well on jets, some limiations from using jet images
Would benefit more from mass decorrelation

Method1:VRNN

- Variational Autoencoder using particle inputs (RNN)





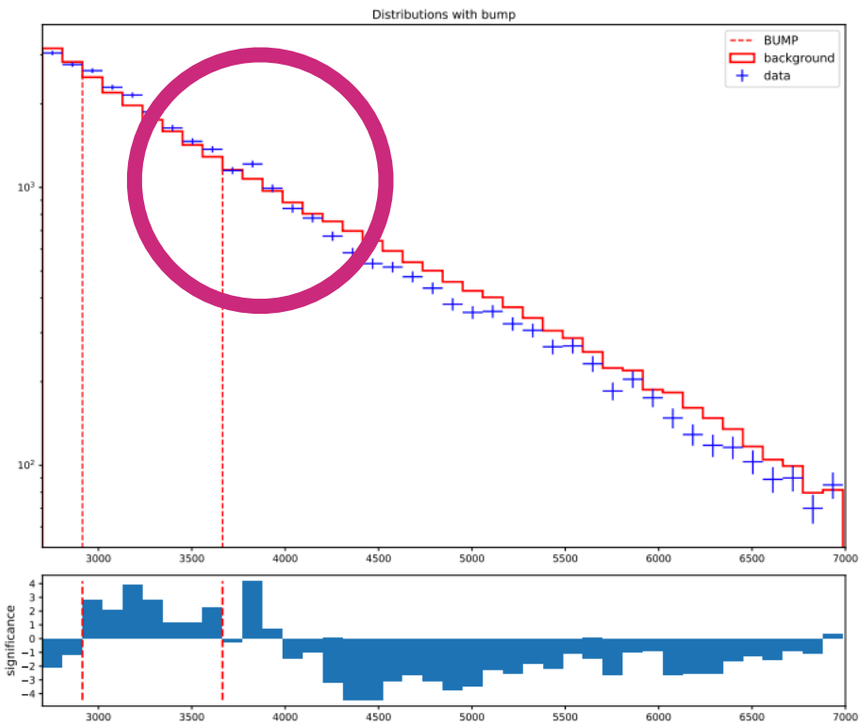
$$\mathcal{L}(t) = \mathrm{MSE} + 0.1 \times \overline{p_T}(t) D_{\mathrm{KL}}$$

$$\text{Anomaly Score} = 1 - e^{-\overline{D_{\mathrm{KL}}}}$$

Observation: Works but preparation of inputs is critical

Method 3:GAN-AE

- Build an auto encoder (AE)
  - Add an GAN to help AE
  - Additionally decorrelate with mass
- Compute a distance (ED) for anom



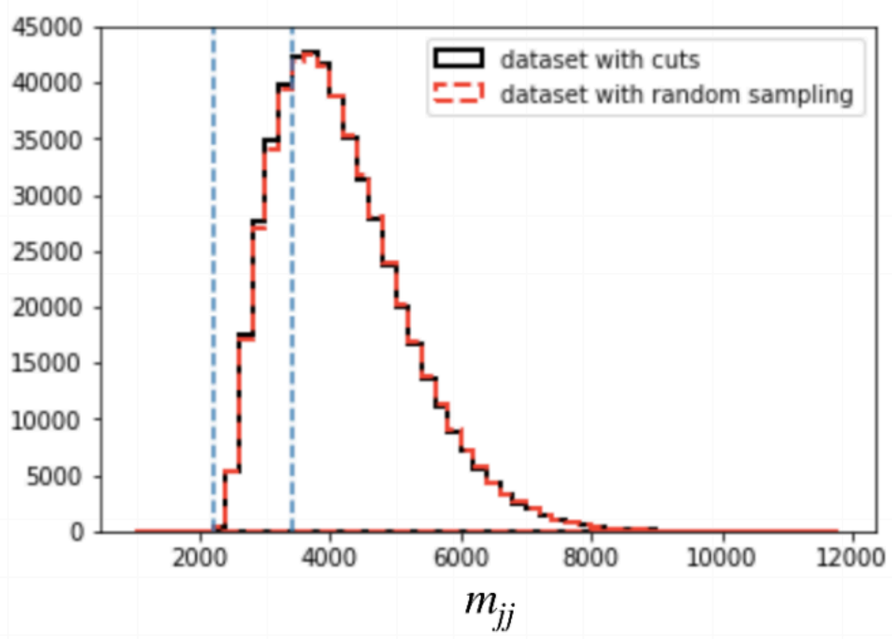
Autoencoder with 10D latent space
Latent space forced to be decorrelated with mass

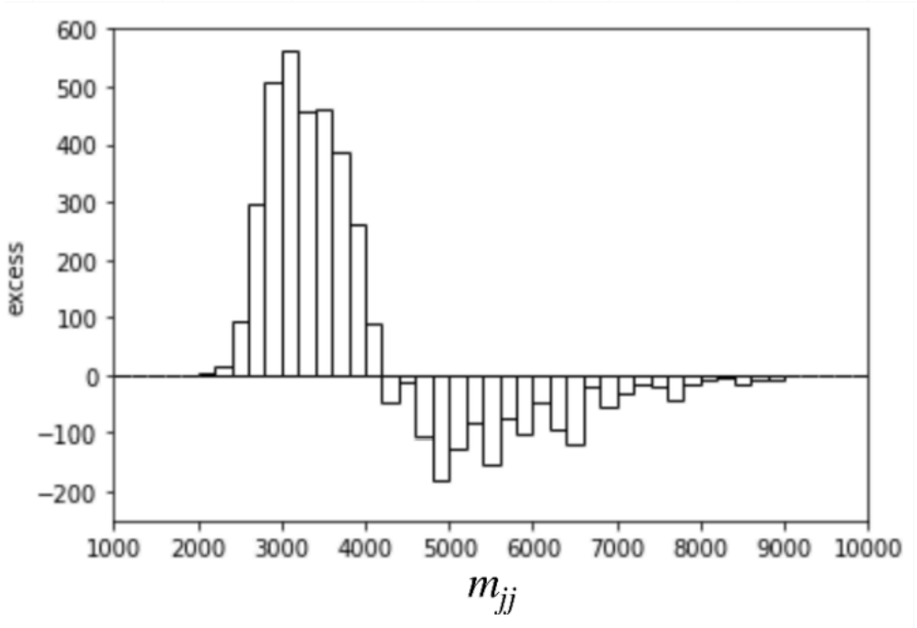
$$\text{loss}_{\text{AE}} = \text{BC} + \varepsilon \times \text{MED} + \alpha \times \text{DisCo}$$

Observation: Mass Decorrelation+Good Simulation needed

Method 4:LDA

- Latent Dirichlet Allocation (LDA)
  - Decluster jet and use splitting info
  - Construct 2 hypotheses in data
  - Generated through LDA approach



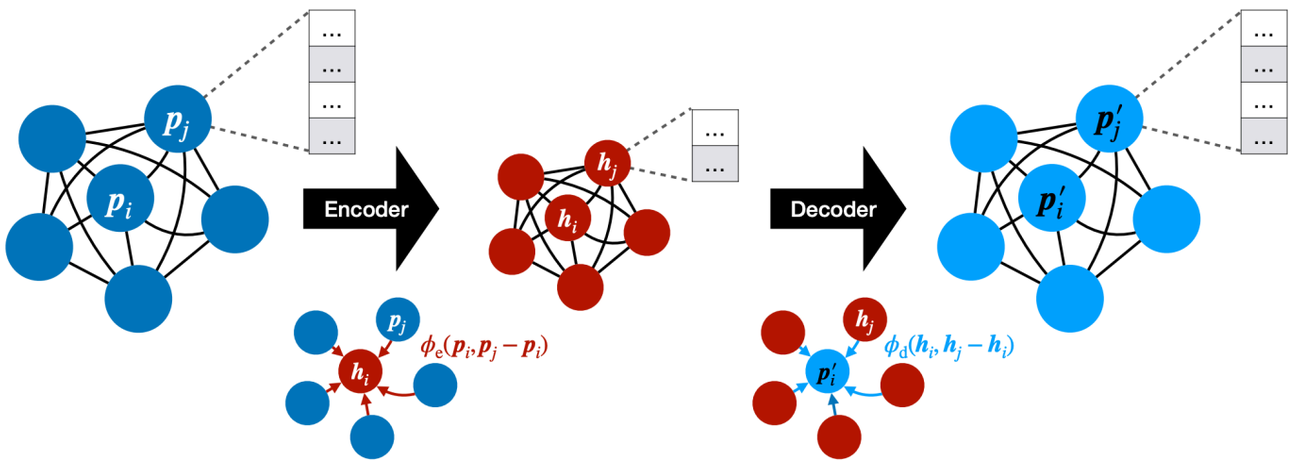


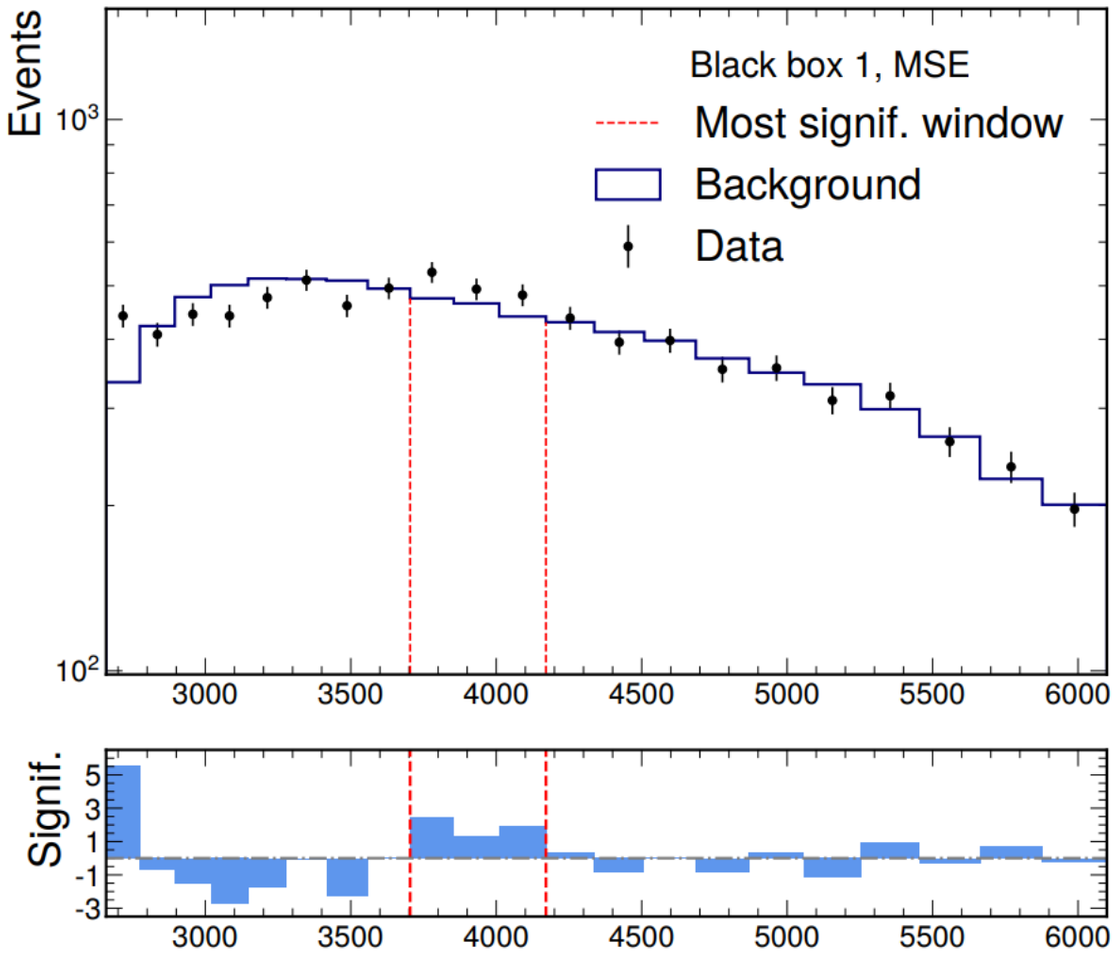
Compute likelihood of two hypoth to be consistent

$$L(o_1, \ldots, o_N|\alpha) = \prod_{i=1}^{N} \frac{p(o_i|\hat{\beta}_1(\alpha))}{p(o_i|\hat{\beta}_2(\alpha))} .$$

Observation: LDA benefits from many observables

Method 5: Particle Graph AE



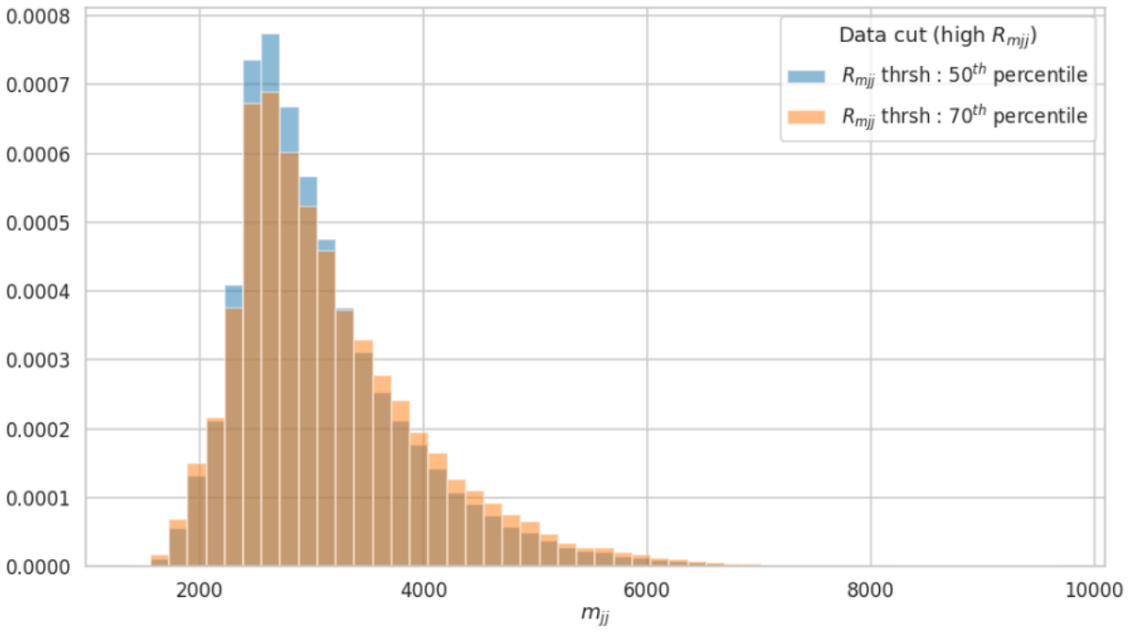


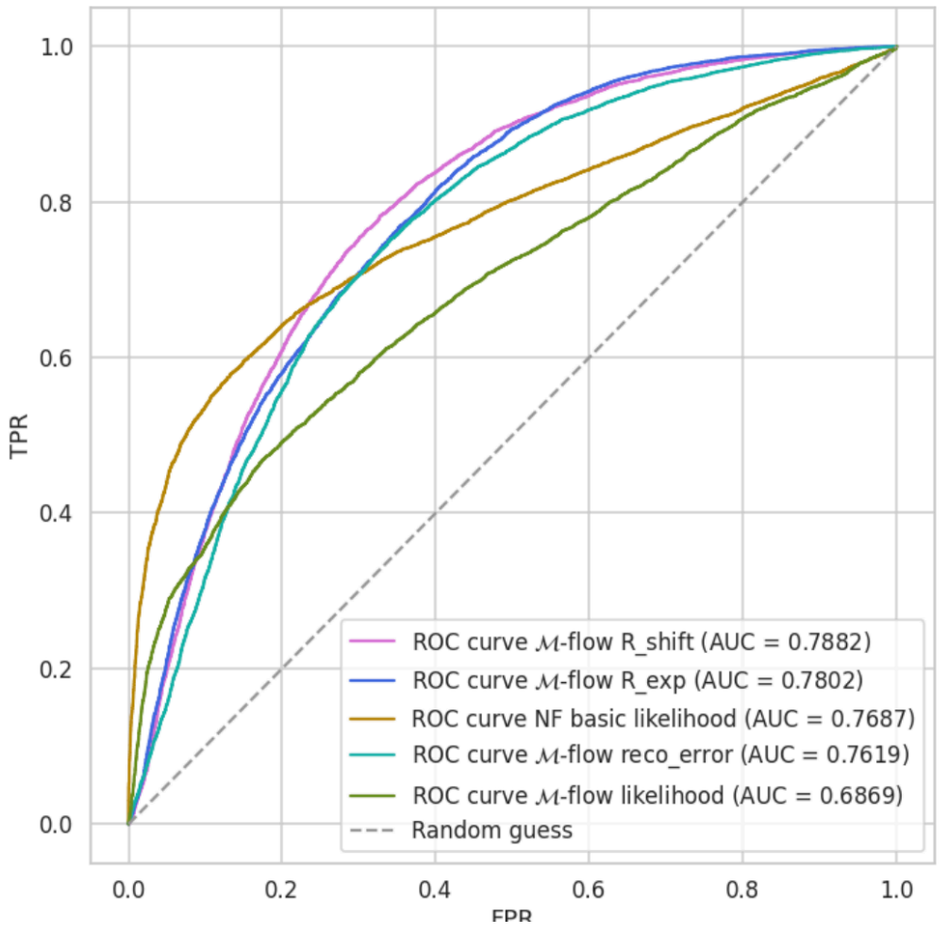
- Build a GraphNN Autoencoder
  - Try with mean squared error loss
  - Try with a permuation invariant loss (robust against physics)

Observation: No good handle on loss

Method 6: Regularized Likelihood

- Use a normalizing flow
  - Cut on high loss
  - Decorrelate loss with mass



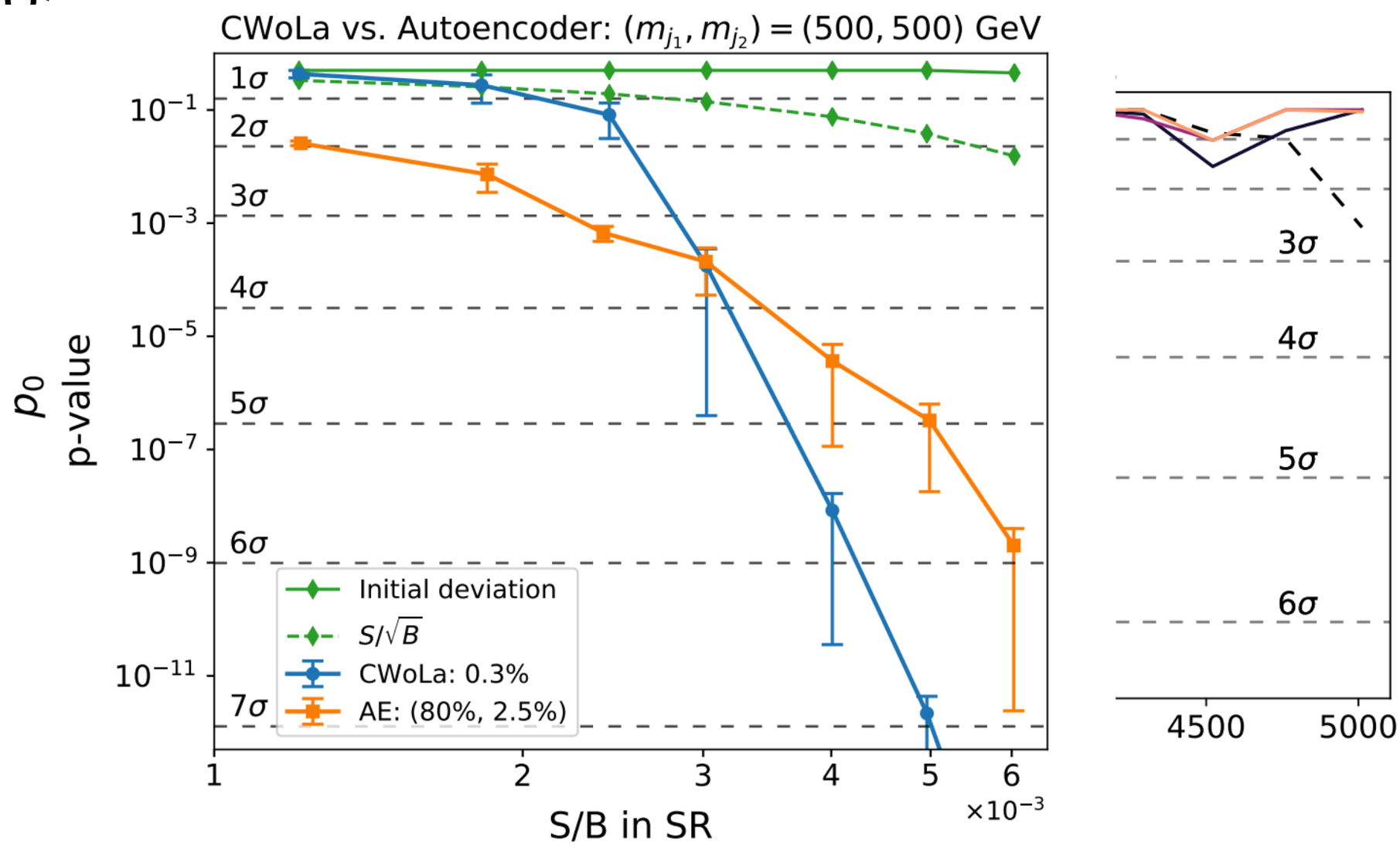


$$\mathcal{R}_{m_{jj}}(x) = \frac{||x - g(g^{-1}(x))||^2}{1 + \frac{p_u(g^{-1}(x))}{p_{KDE}(m^x_{jj})}}$$

Observation: A single auto encoder even with NF is not enough too many anomalies (no clear signal)

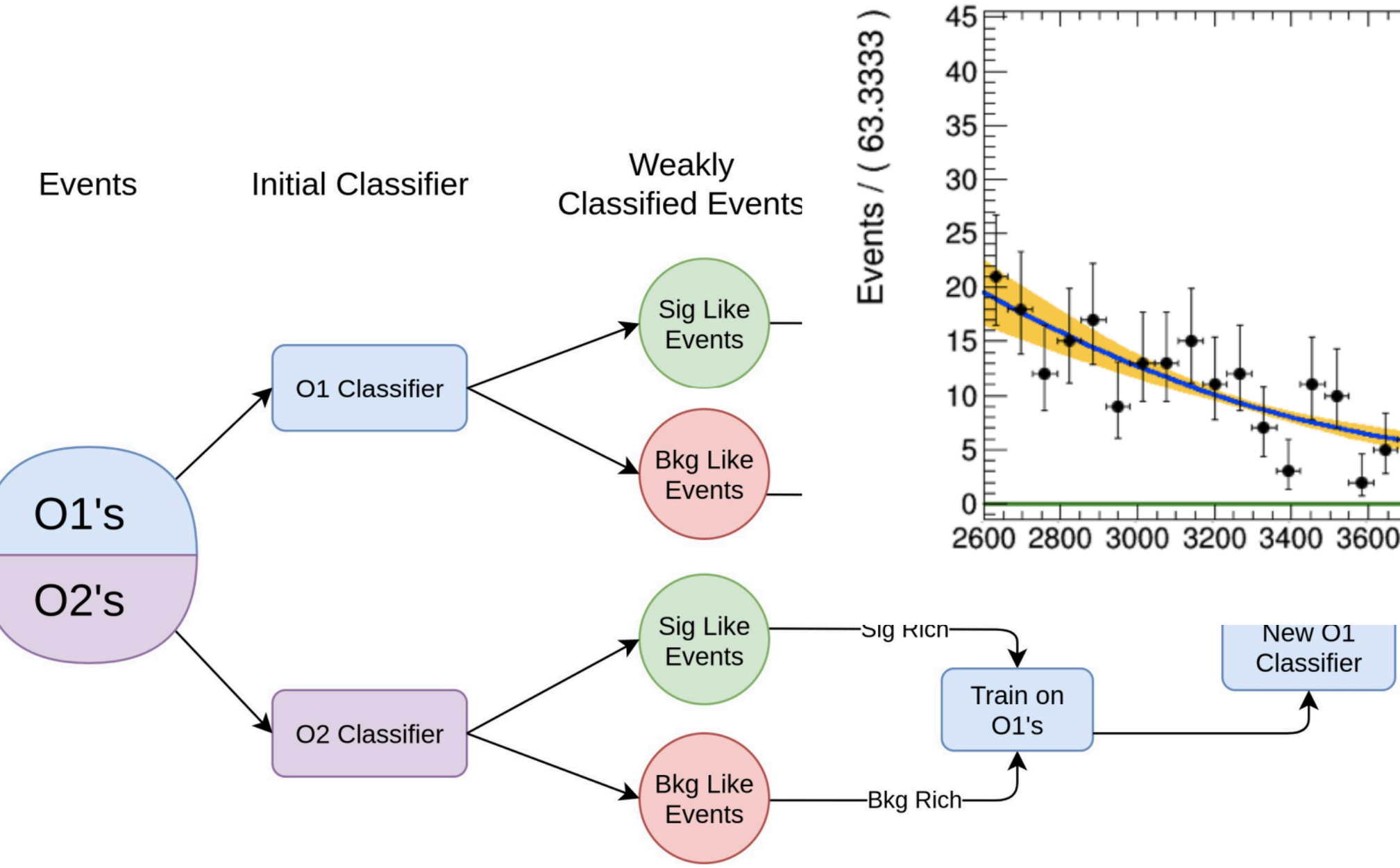
Method 8: CWoLa

- Use a normalizing flow
  - Cut on high loss
  - Decorrelate



Observation:Approach works for single jet resonances

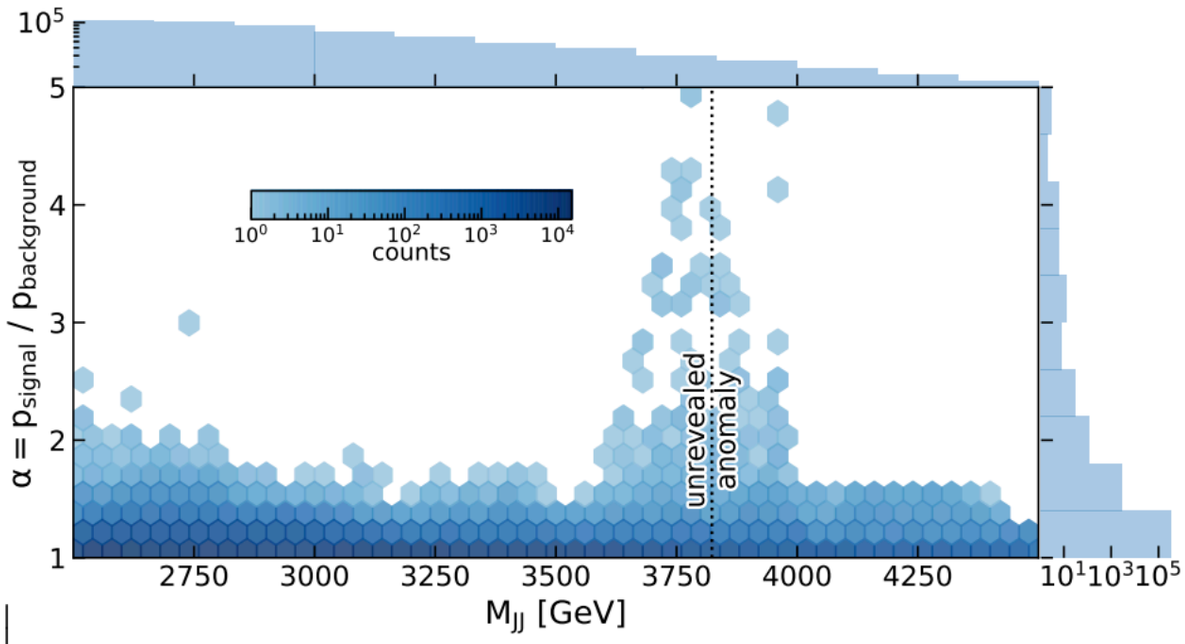
Method 9: Tag N'Train



Observation:Works well on jets, some limiations from using jet images
Would benefit more from mass decorrelation

Method 11:GIS

- Guassian Iterative Slicing
  - Cut on high loss
  - Decorrelate loss with mas



$$p(x|x_c) = \pi(f_{x_c}(x)) \left|\det\left(\frac{\partial f_{x_c}(x)}{\partial x}\right)\right| = \pi(f_{x_c}(x)) \prod_{i=1}^{i=N} \left|\det\left(\frac{\partial f_{x_c,i}(x)}{\partial x}\right)\right|.$$

Observation:Low noise robust density estimation is key

Method14:QUAK