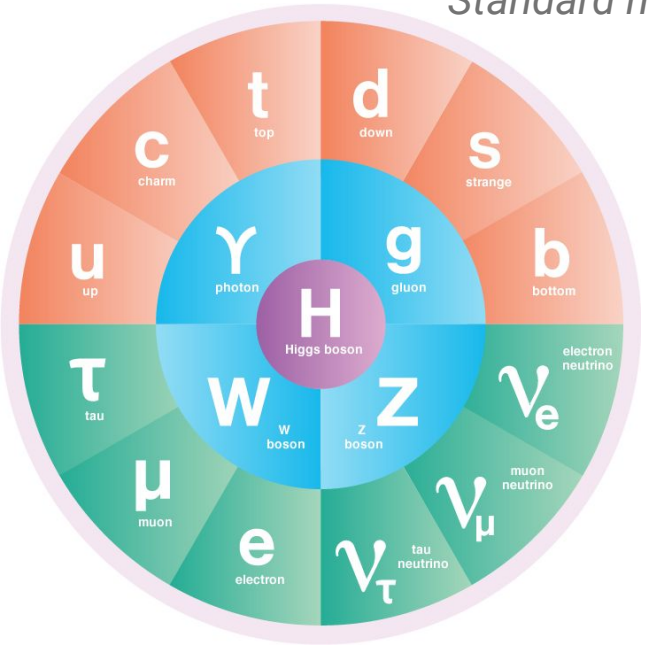
Machine Learning for High pT Physics

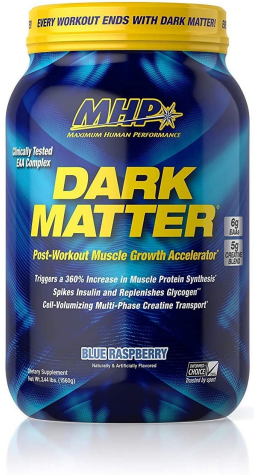
NIKHEF Topical ML Lectures, June 2022

Johnny Raine, University of Geneva

Introduction

Standard model is complete!





But plenty of missing pieces for a complete theory

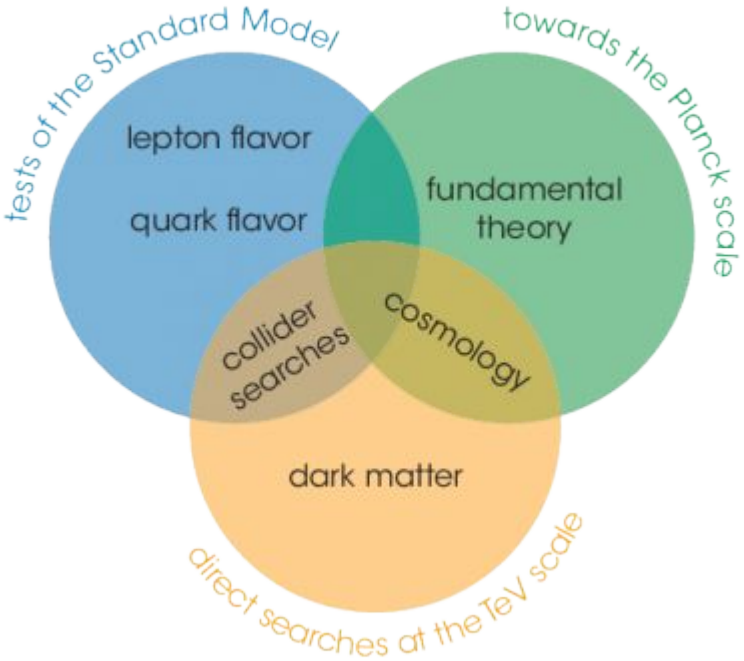
Introduction

Aims at the LHC (ATLAS + CMS)

Find the Higgs boson ✅

Observe and test SM processes ⭕

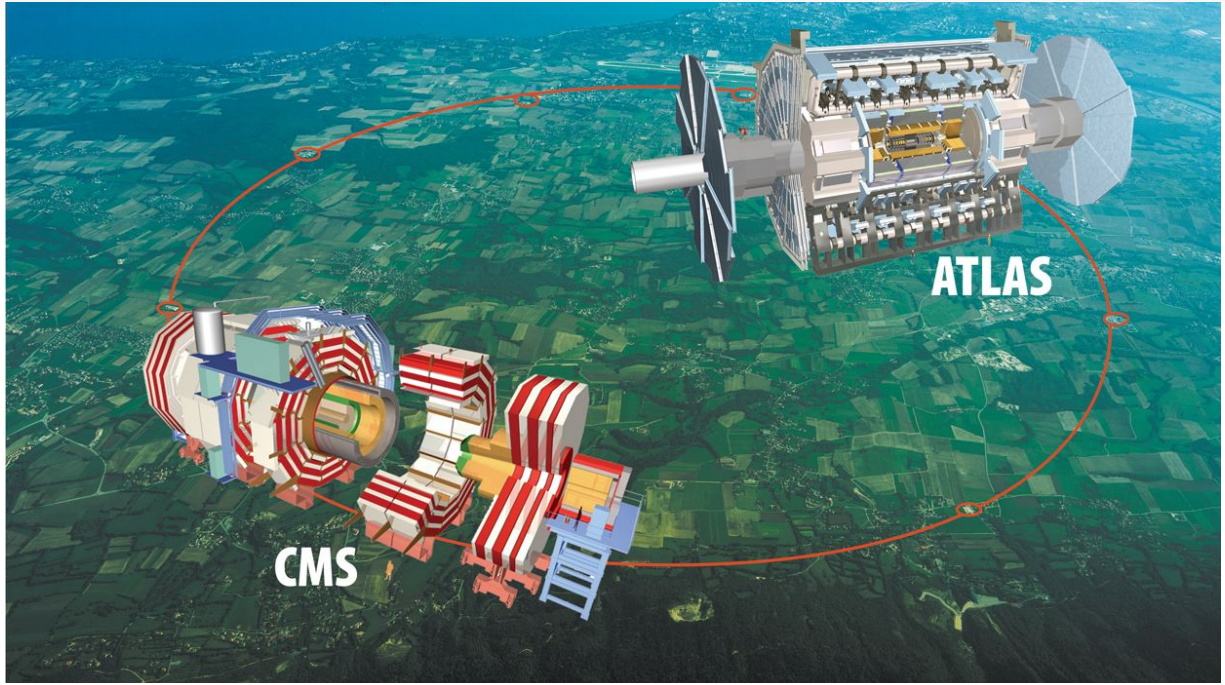
Fill in the missing pieces of the puzzle ❌



ATLAS and CMS

Energy frontier

Intensity frontier



General physics programme

Multipurpose detectors

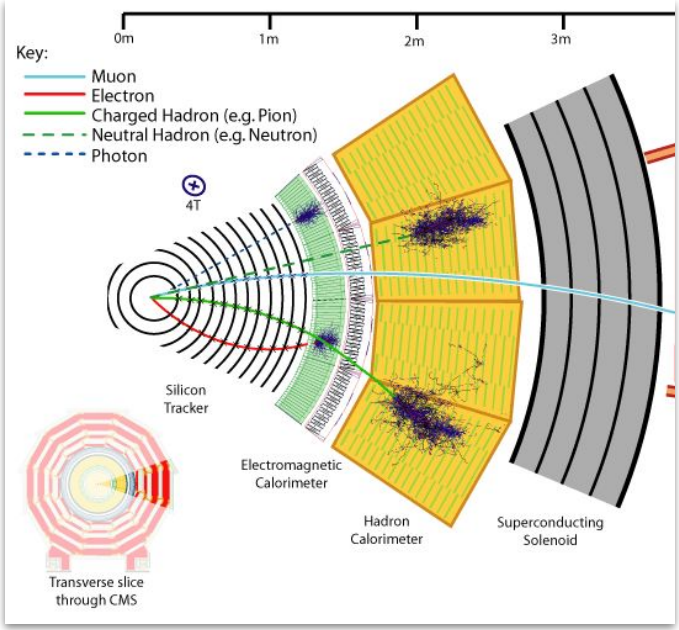
ATLAS and CMS

Record physics objects with detector subsystems

EM showers, hadron showers and tracks become: Electrons, muons, jets, (hadronic taus), missing transverse momentum

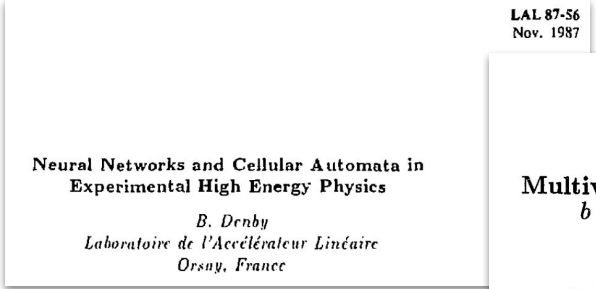
Dependent on accurate **object reco and ID**

Analyses built on object selections, measuring rates and properties at event (collision) level



ML in HEP

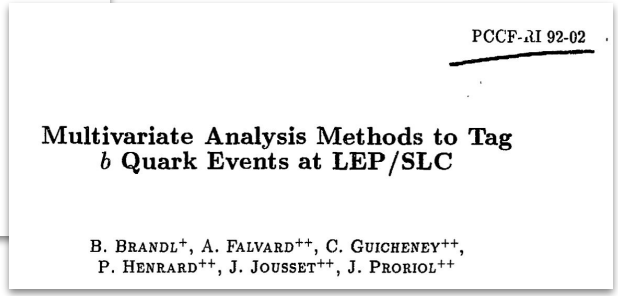
Although recent surge in popularity, ML applied in HEP under different guises for a long time


LAL 87-56
Nov. 1987

Neural Networks and Cellular Automata in Experimental High Energy Physics

B. Denby
Laboratoire de l'Accélérateur Linéaire
Orsay, France

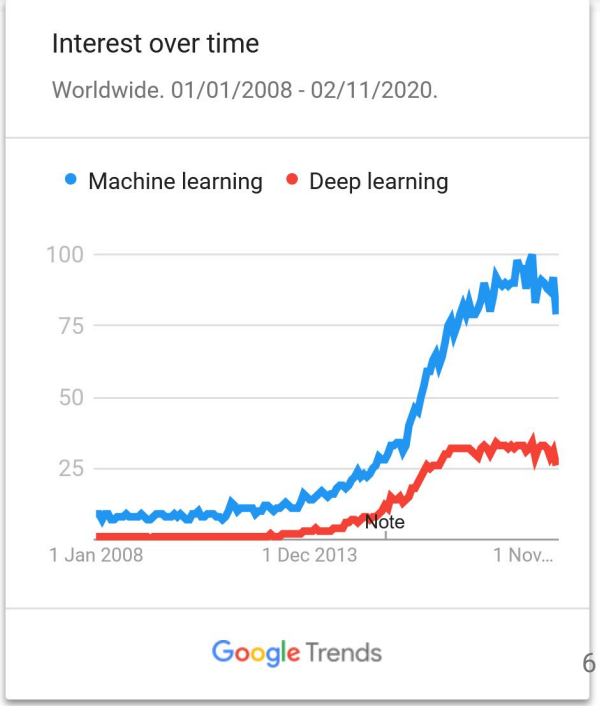
Track reconstruction


PCCF-RI 92-02

Multivariate Analysis Methods to Tag *b* Quark Events at LEP/SLC

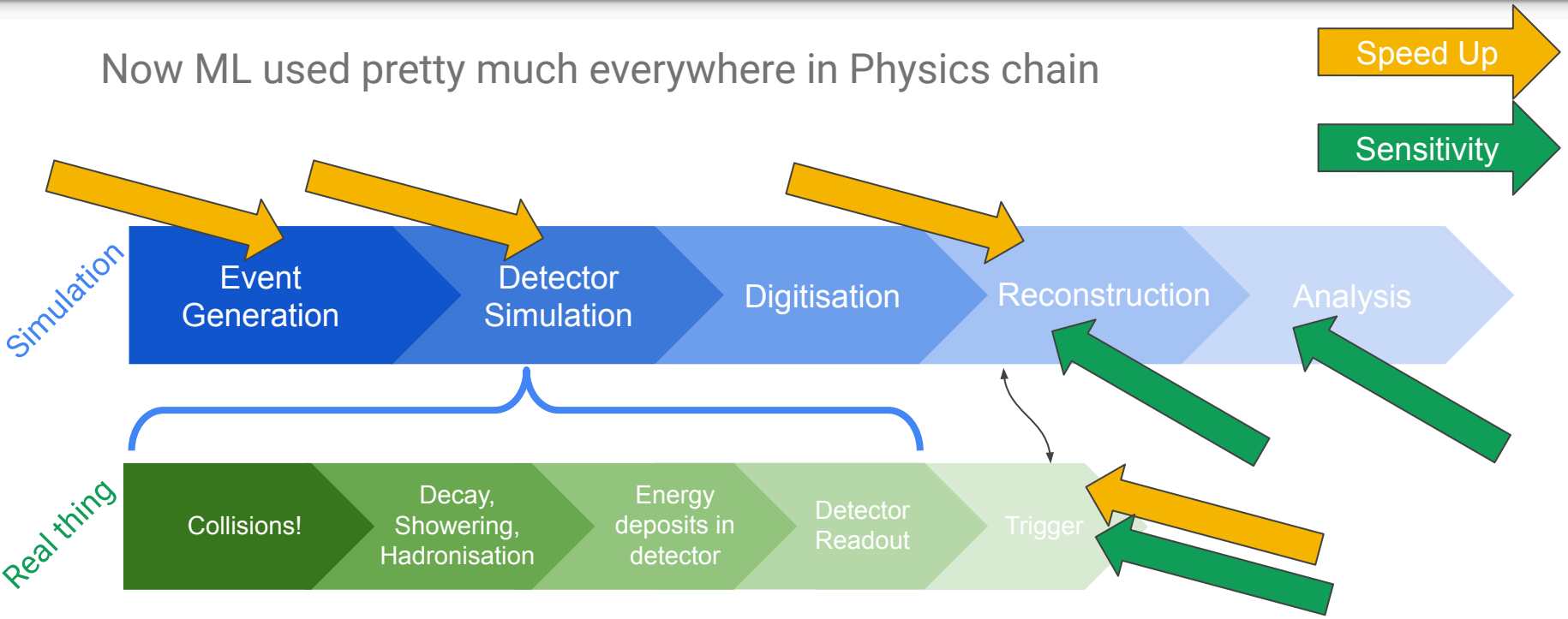
B. Brandl+, A. Falvard++, C. Guicheney++,
P. Henrard++, J. Jousset++, J. Proriol++

Object ID

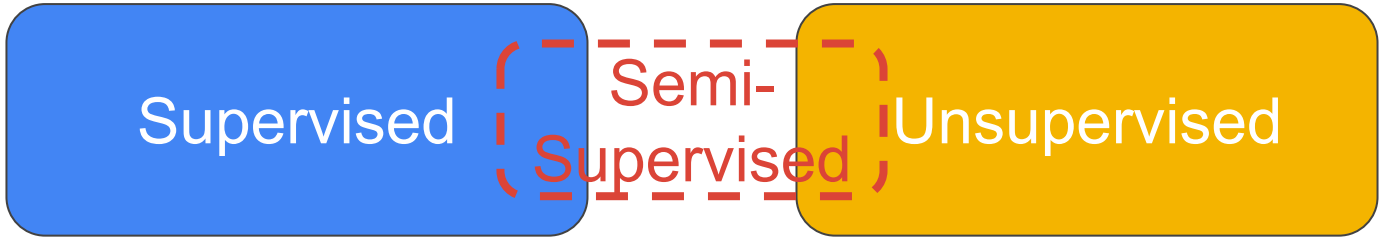


ML in ATLAS and CMS

Now ML used pretty much everywhere in Physics chain



ML in HEP



Reinforcement Learning

Detector design?

We mostly live here
Use labelled data, known targets

Learn direct from the data
Anomaly detection, clustering
Dimensionality reduction

ML in HEP

HEP applications mostly **fully supervised classification**

- We know our target (**class**, **value**)
- Learn to predict from set of inputs
- In all cases, network is learning to functional approximation $f(y|\{x_i\})$

Classification

Object Tagging

Combinatorics

Analysis Discriminant

Regression

Infer "Observables"

Background Fit

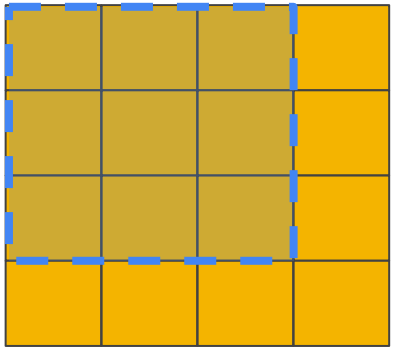
Improve Precision

Overview - Network Design

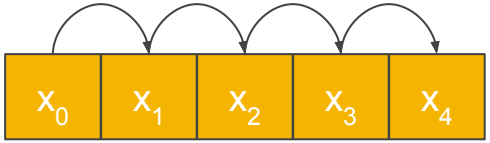
- What kind of inputs are we dealing with - **Structured Data**

- My inputs are independent
- Each block is a different variable
- **Flat input->fully connected layers**

Mostly used



- My inputs have regular spatial separation
- Each block is the same "variable"
- **Convolutional networks (1-3D)**

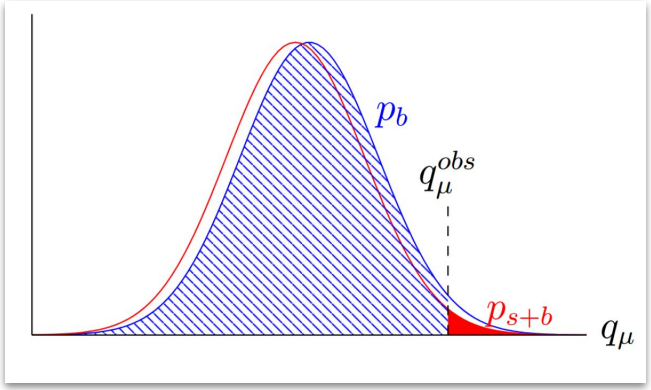


- My inputs come in a sequence
- Each block is the same "variable"
- Logical order with dependence on what comes before/after
- Time/spatial sequence
- **Recurrent networks**

Overview - Analysis

Key part of (most) analyses is hypothesis testing

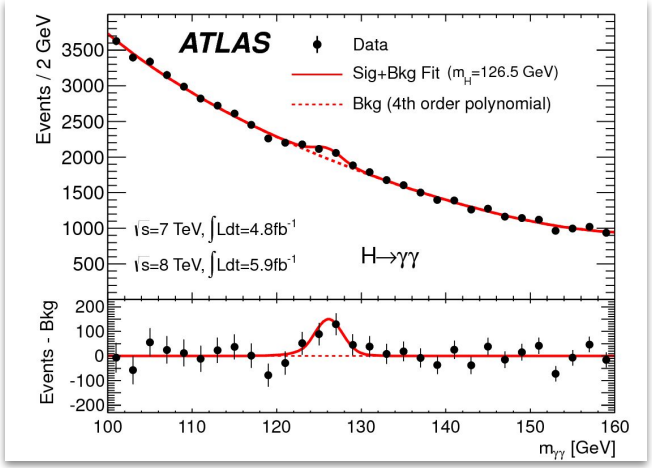
- Does data fit Hypothesis 1 or Hypothesis 2?
  - e.g. Amount of signal on top of background
  - Optimal value for a “free” parameter



Overview - Analysis

Key part of (most) analyses is hypothesis testing

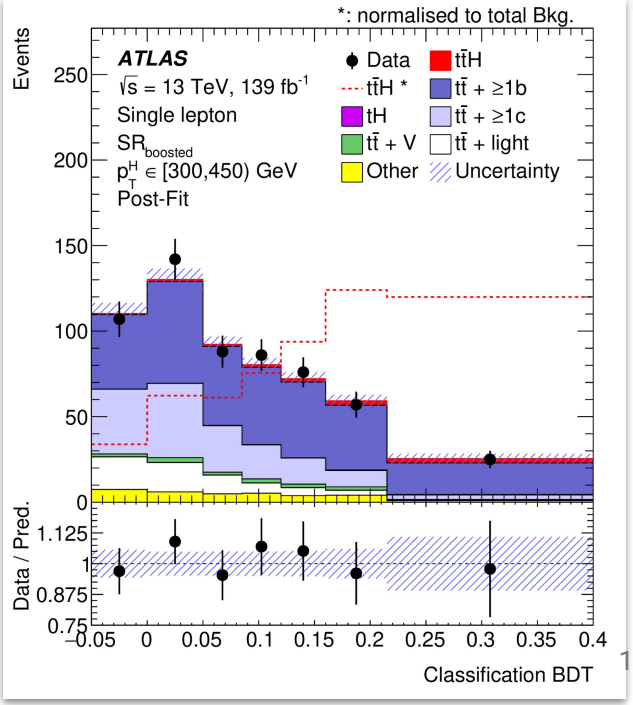
- Does data fit Hypothesis 1 or Hypothesis 2?
  - e.g. Amount of signal on top of background
  - Optimal value for a “free” parameter
- Sometimes use objects directly with limited ML



Overview - Analysis

Key part of (most) analyses is hypothesis testing

- Does data fit Hypothesis 1 or Hypothesis 2?
  - e.g. Amount of signal on top of background
  - Optimal value for a “free” parameter
- Sometimes use objects directly with limited ML
- But other times have very small signal with complex background

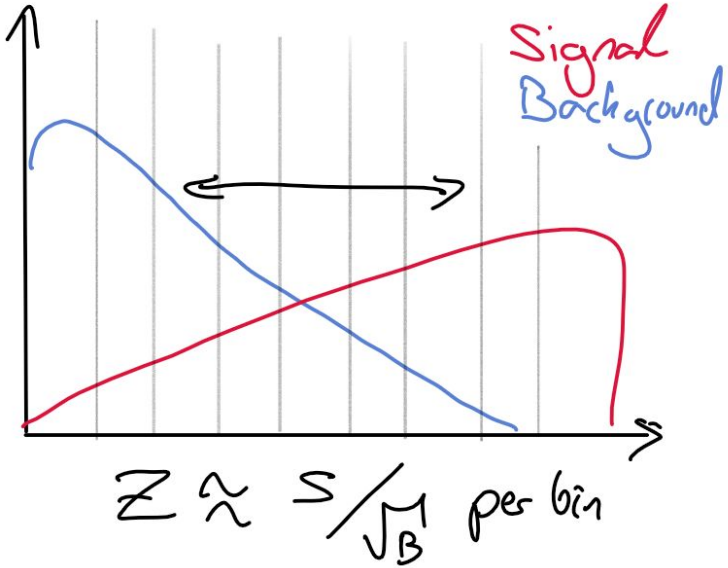


Event classification

Combine many observables with some separation between S and B

Maximise separation of S and B in **one 1D** discriminant

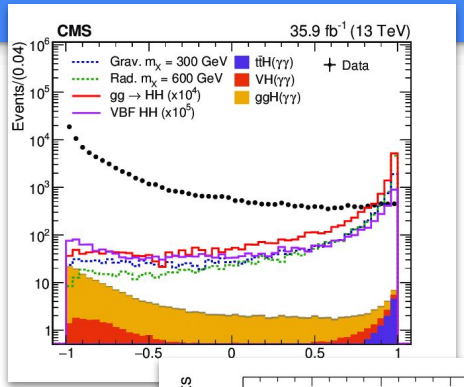
Distribution in profile likelihood fit - max sensitivity using all bins

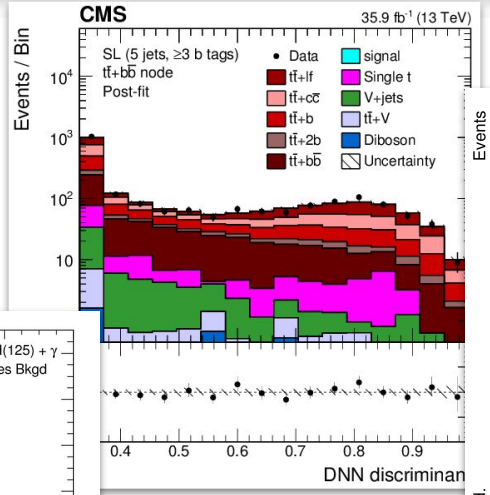


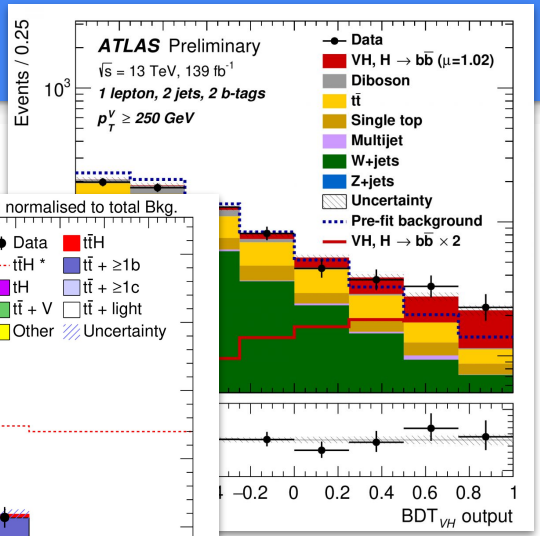
x_0	x_1	x_2	x_3	x_4

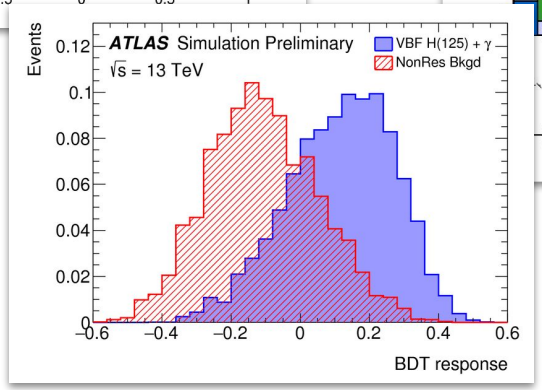
Simple Signal vs Bkg DNN/simpler algos (almost) everywhere

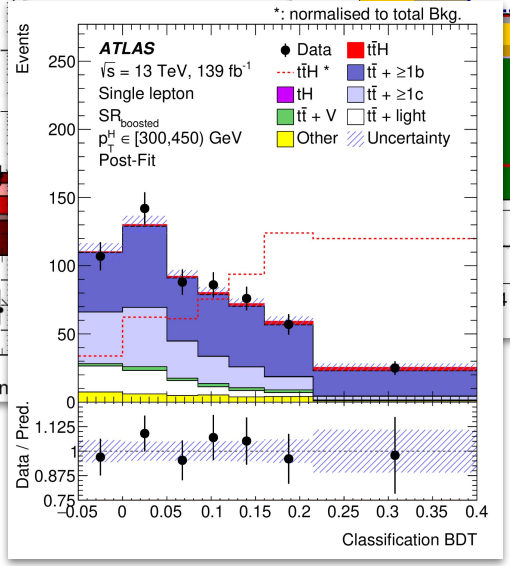
Event classification









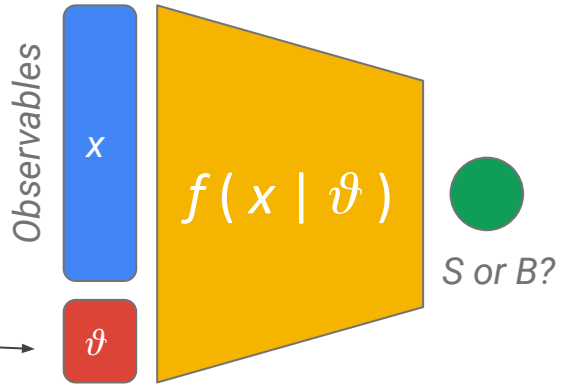


Combine sensitive observables and fit the whole output distribution!

Parametrised Neural Networks

Everything I can measure that's useful

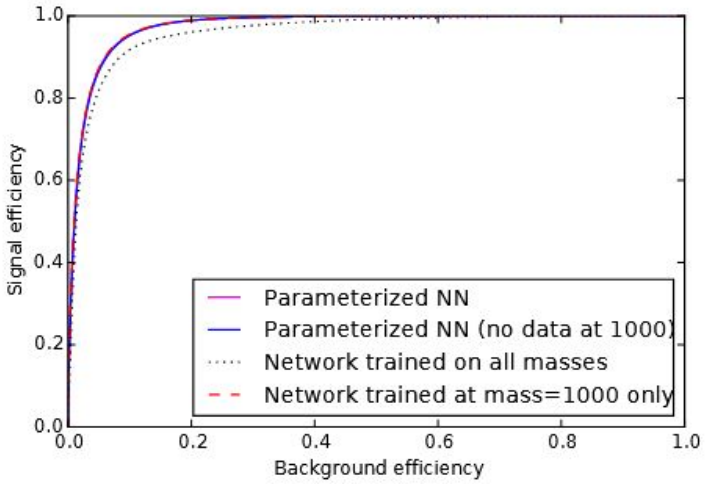
The models/property I'm looking for

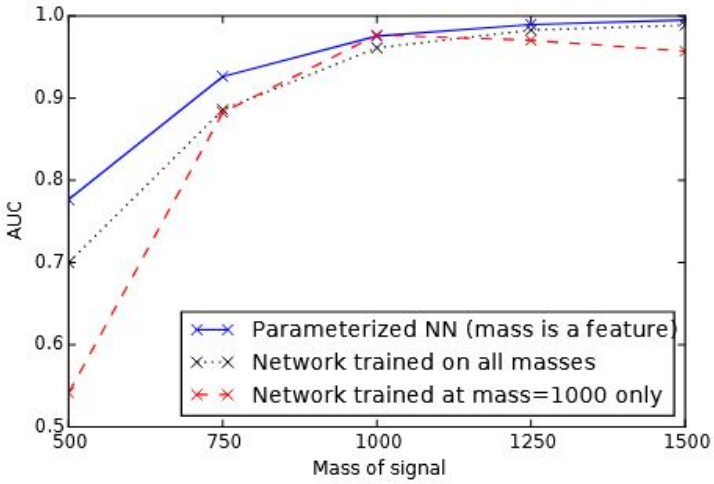


NNs are really good functional approximators - can do conditional functions too!

Baldi et al

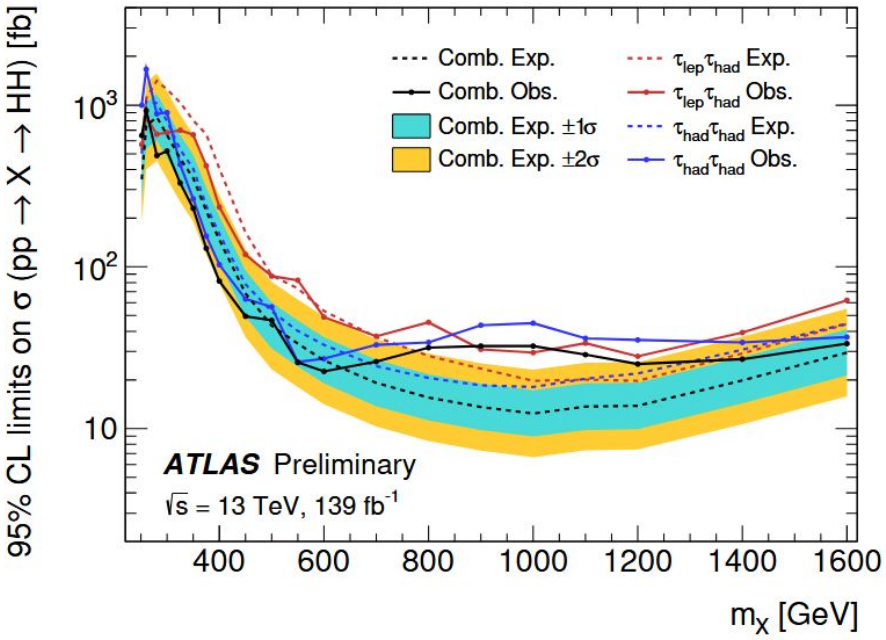
Parametrised Neural Networks





Improved performance and generalisability

Parametrised Neural Networks



PNN used to search for resonant X production

mX used as input alongside event oberservables
Trained on all signals mass points vs background

Same NN used for all 20 points in scan

Old approach needed a new classifier at each mass point!

Event classification

Widest use of ML in ATLAS/CMS but also simplest

Can also use ML to “reconstruct” event - match objects to hypothesised process

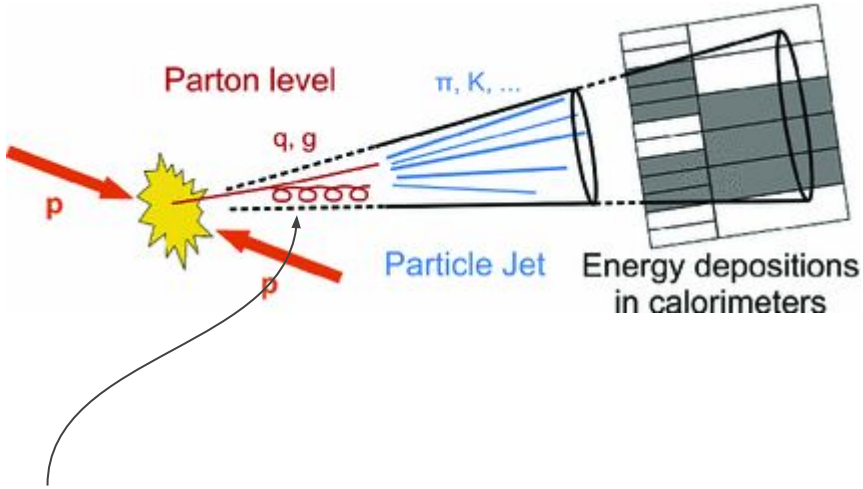
Aim is to enhance sensitivity to signal process/property of interest

Object Tagging

General purpose detector with high energy particles

Only observe energy of stable particles

Everything but muons reconstructed as cone of energy depositions and tracks



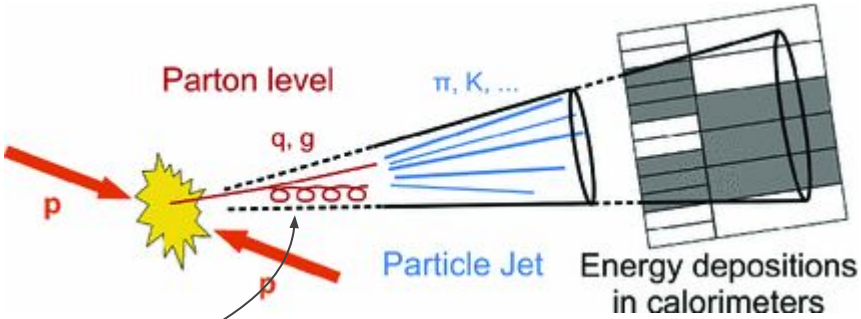
What was this?

Object Tagging

General purpose detector with high energy particles

Only observe energy of stable particles

Everything but muons reconstructed as cone of energy depositions and tracks



What was this?

Looking for b- or c-quarks?
Flavour tagging!

Flavour tagging

Lots of interesting physics decays to heavy quarks (b and c)

Top quarks!

BSM most likely in 3rd Generation

Higgs to bb (cc)!

Flavour changing neutral currents

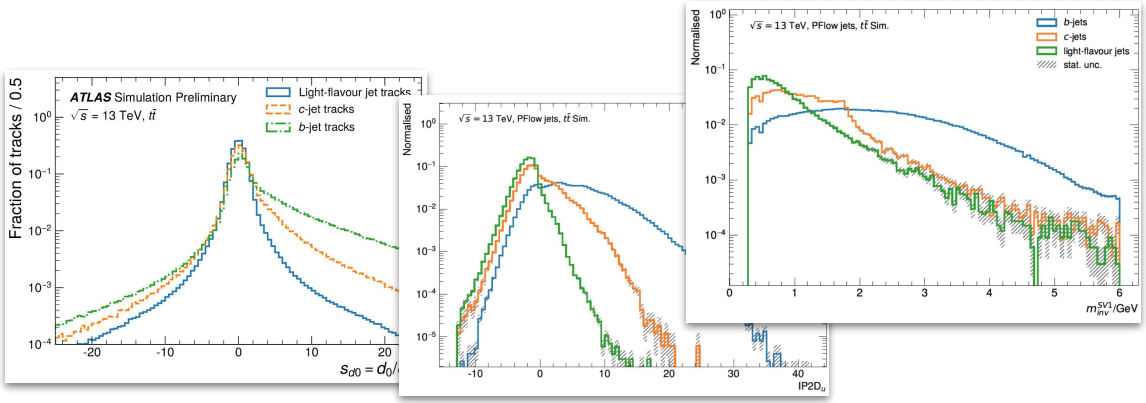
But detector only reconstructs tracks of charged stable particles + energies

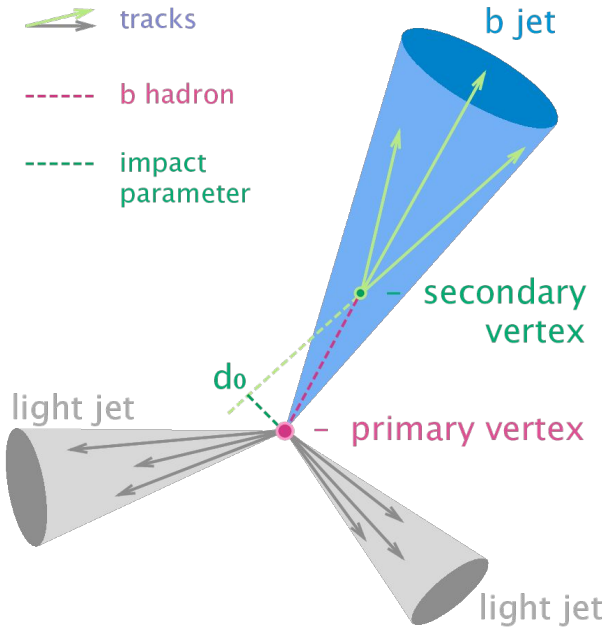
Flavour tagging

Exploit lifetime of B (D) hadrons

- displaced tracks, vertices!

Traditional approach: many variables, combine

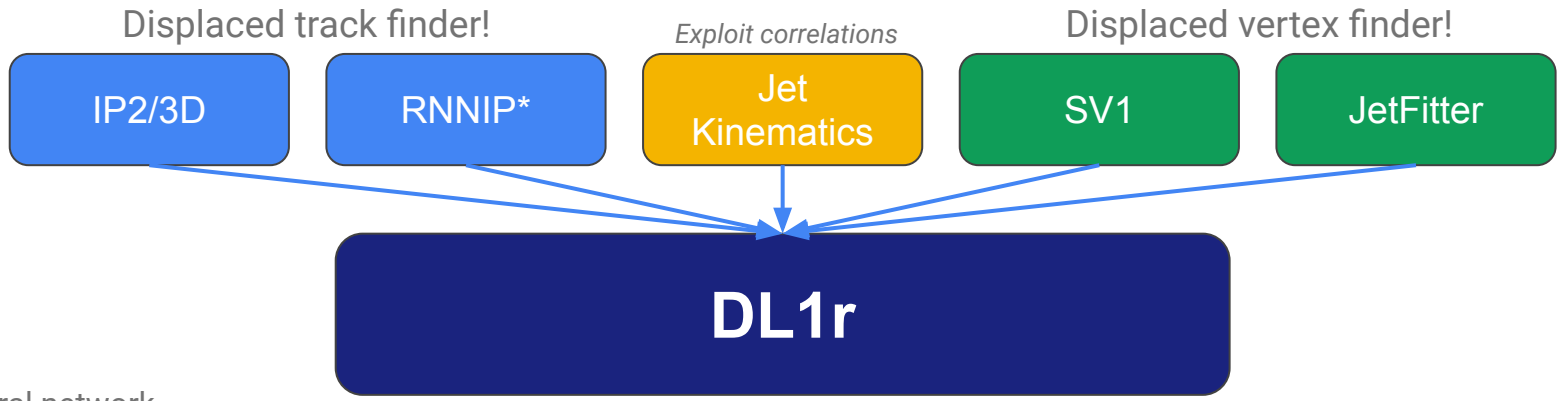




ATLAS Flavour tagging

Takes as input several algorithms, each trained/optimised separately

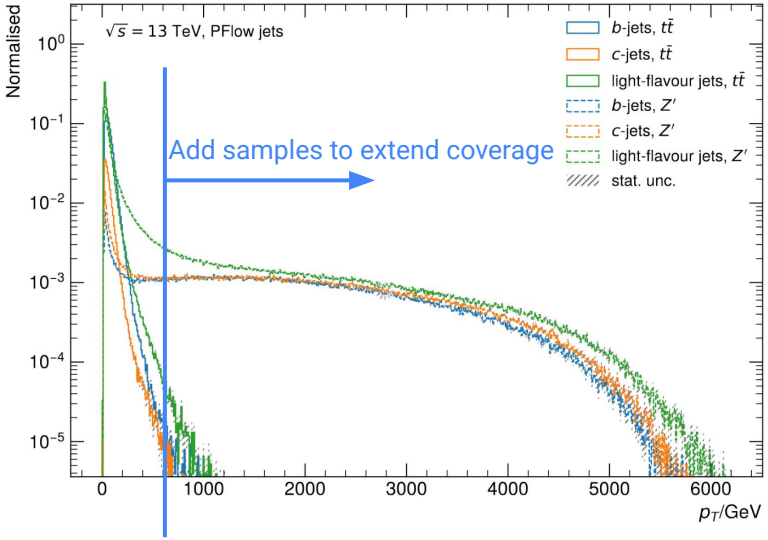
Combine many discriminants and properties with **deep neural network**

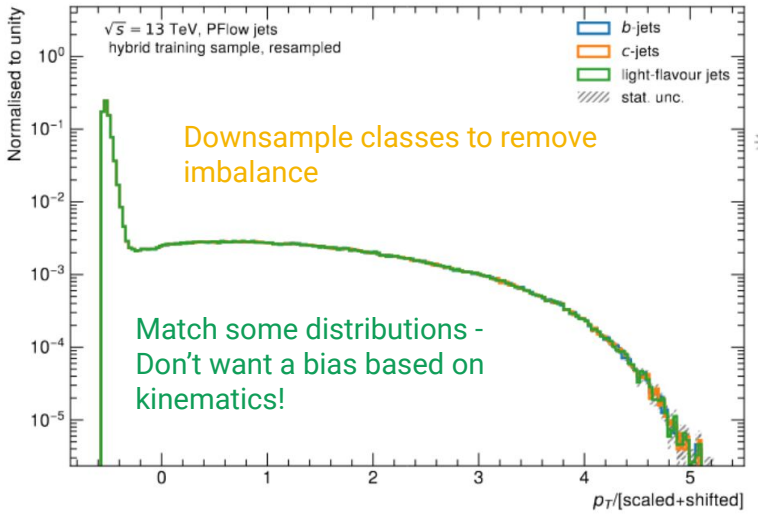


*Also a neural network

Performance across wide phase space

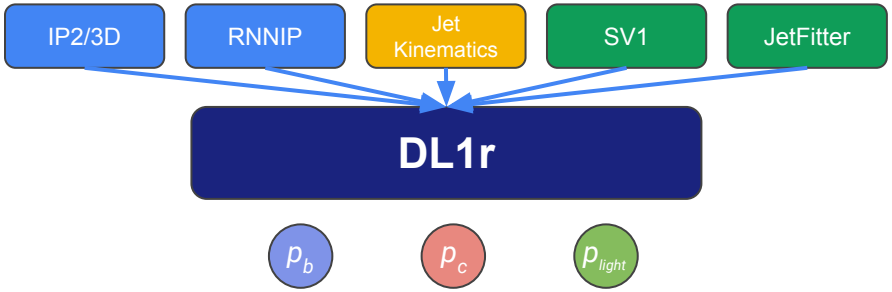
Training data will always introduce a bias - can't perform equally well everywhere without a lot of work!





ATLAS Flavour tagging

Apply to each jet individually - need to define a discrete label!



Not like dog v cat v horse

Taking max not optimal…

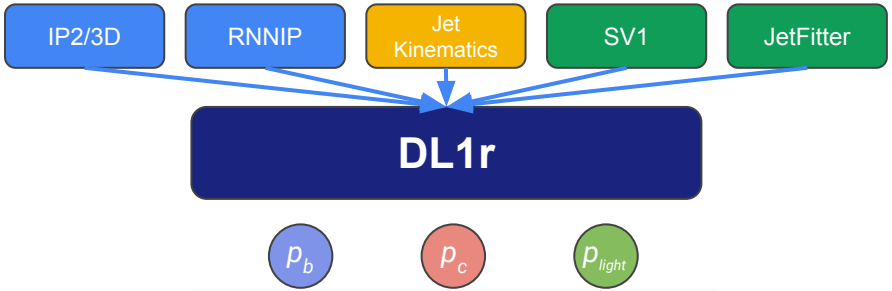
ATLAS Flavour tagging

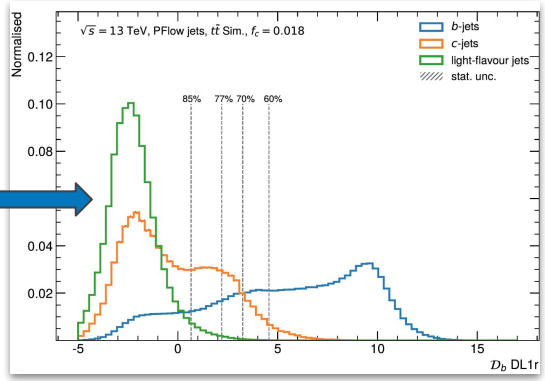
Apply to each jet individually - need to define a discrete label!

- Combine multiclass nodes into discriminant
- Flexibly optimise working points
  - Choose which is signal b or c

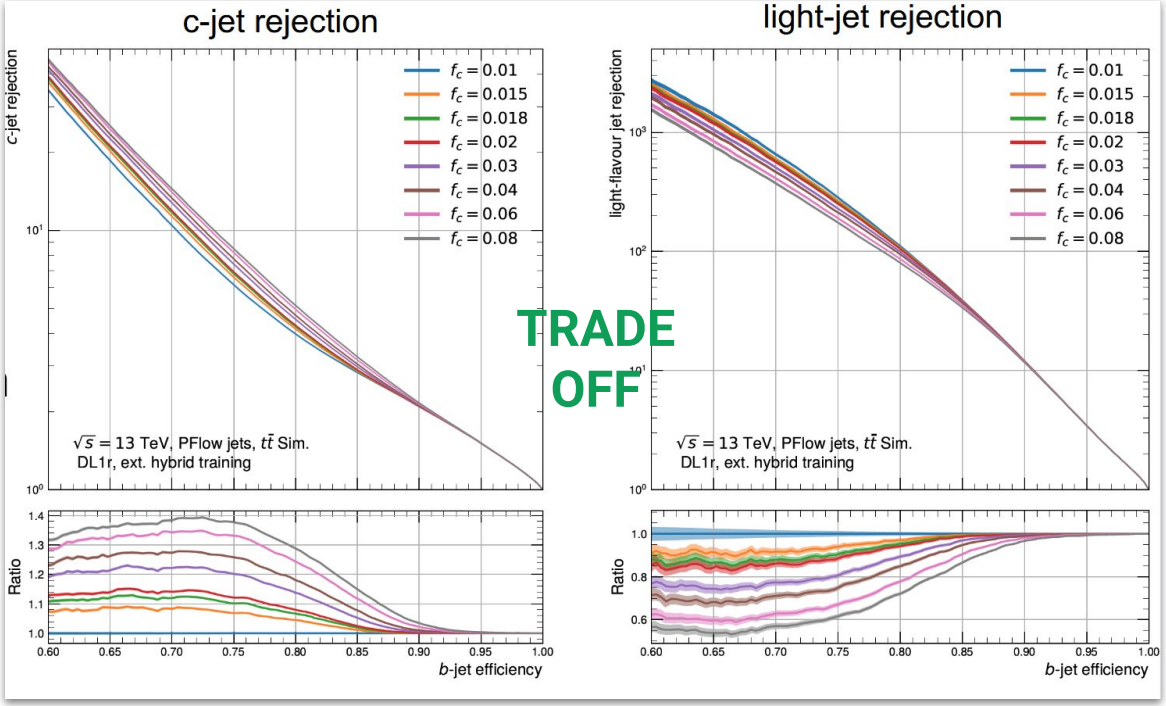
Neyman Pearson Lemma:

$$\mathcal{D} = \frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_{light}}$$





ATLAS Flavour tagging

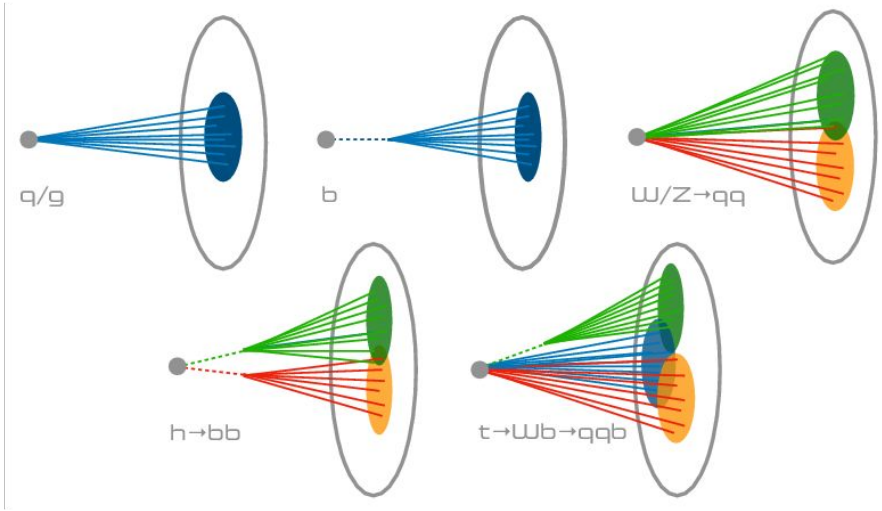


Boosted jet tagging

New physics could be hiding at high pT

Massive particle decay - two boosted particles

But at very high pT can’t resolve individual final state objects!



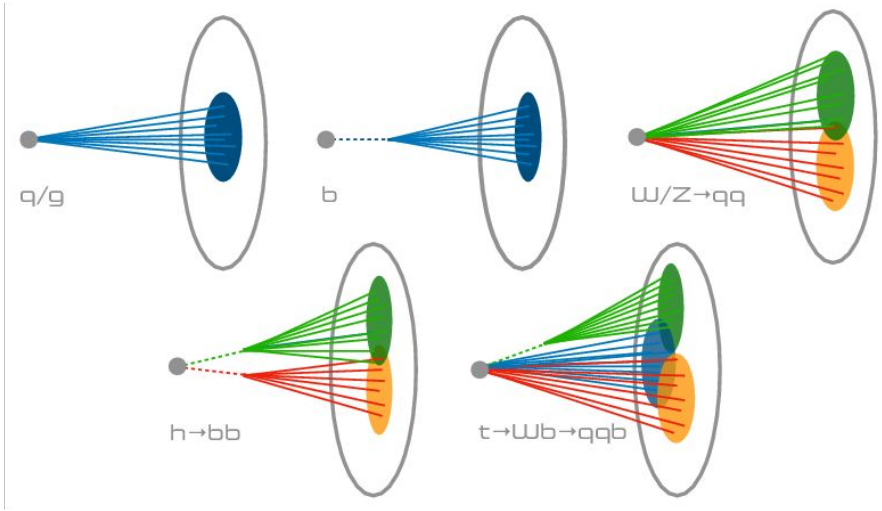
Boosted jet tagging

Very similar story to flavour tagging

Look for structure in “clusters” that build up individual jets

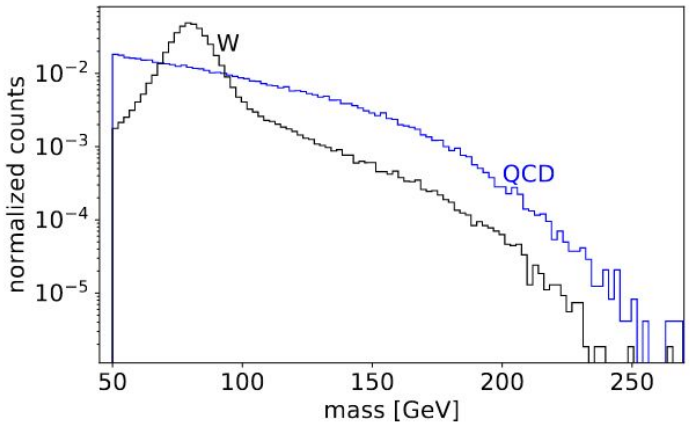
Traditional approach:

- Use physics knowledge
- Build observables
- Combine!

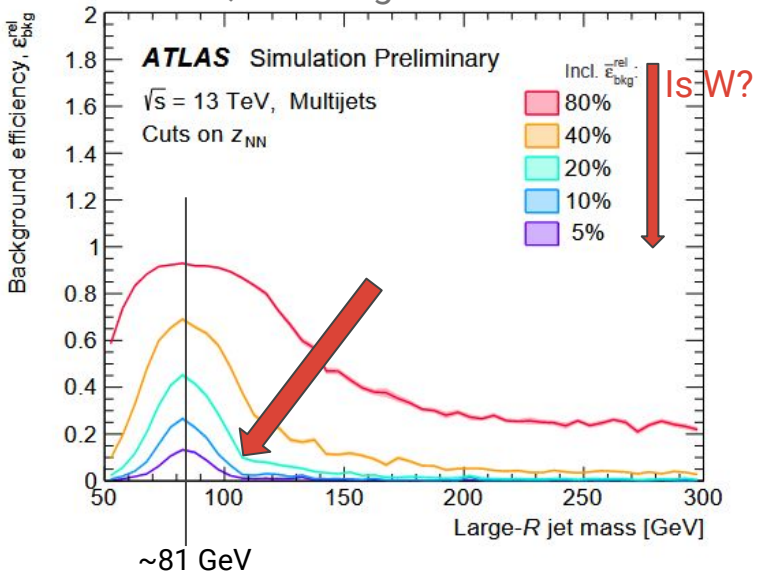


Boosted jet tagging

Input distributions



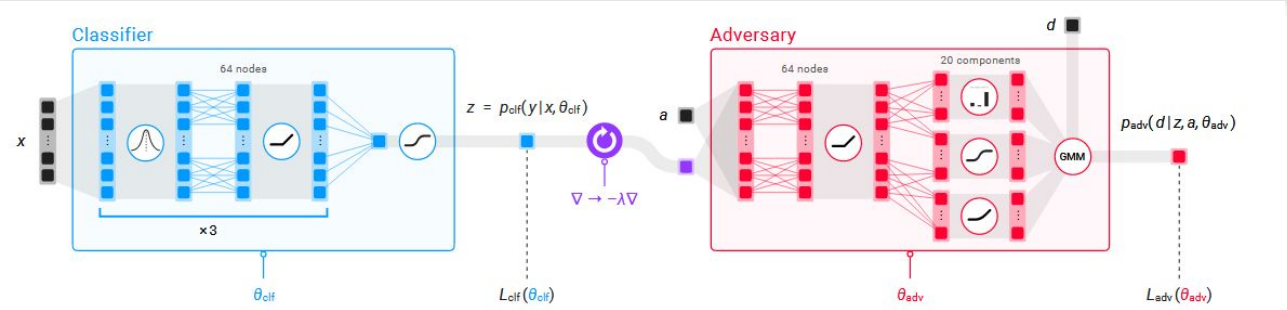
QCD background



Classifier learns to cut on mass!

Not ideal for tagging!

Mass decorrelation with adversaries



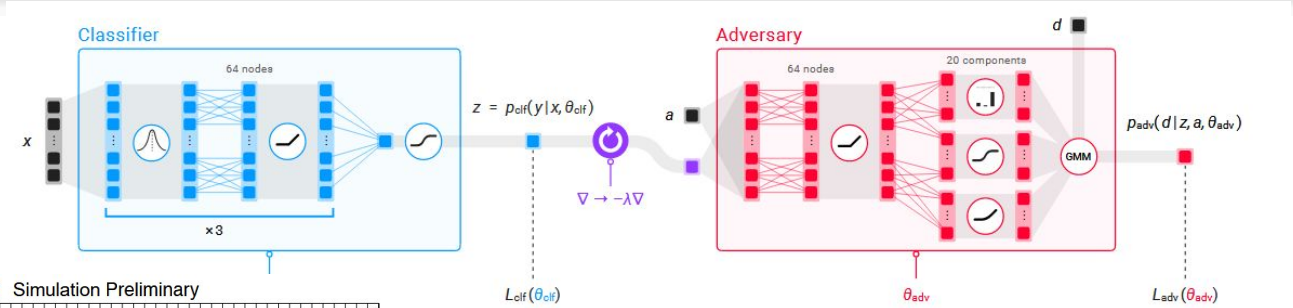
Jet classifier

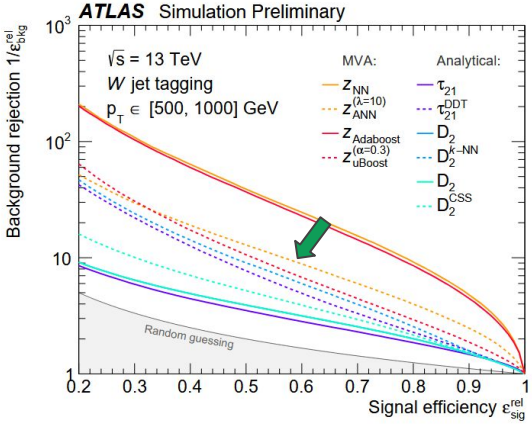
Standard classifier, predict jet class

Adversarial network

Learn the mass from classifier output
Penalty term if it can do well

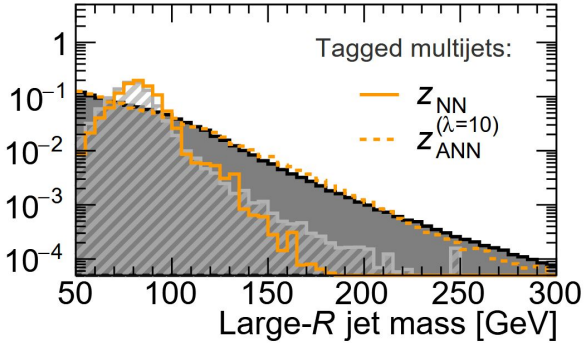
Mass decorrelation with adversaries



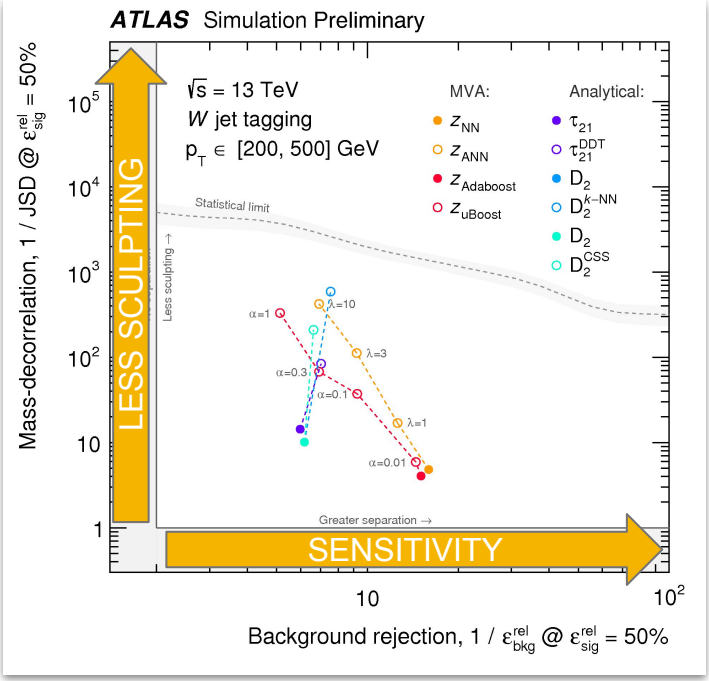


Trade sensitivity

For generalisability



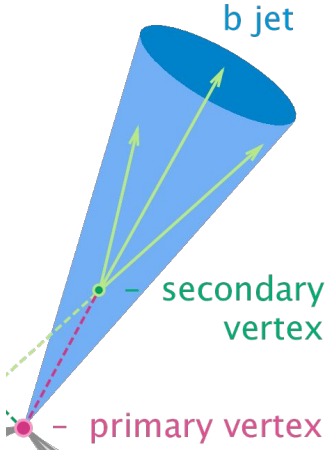
Mass decorrelation!



Input format - smarter networks

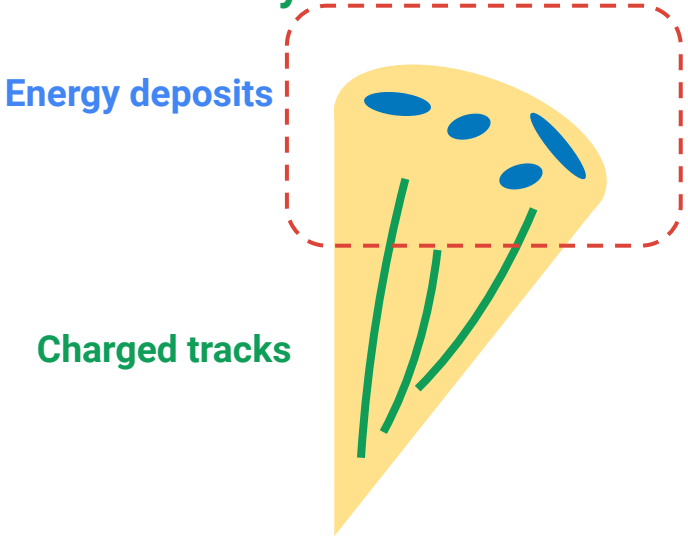
Do we really know the best variables to identify objects?

Collection of tracks matched to a jet

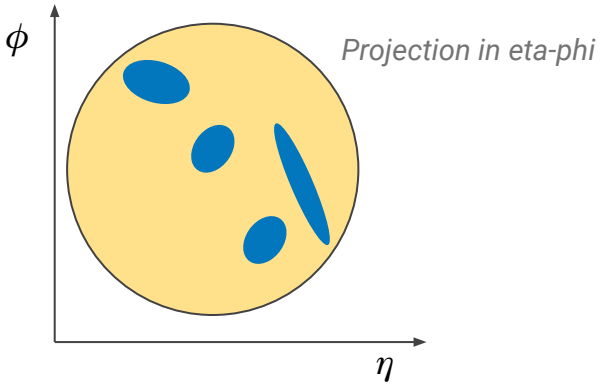


Input format - smarter networks

Do we really know the best variables to identify objects?



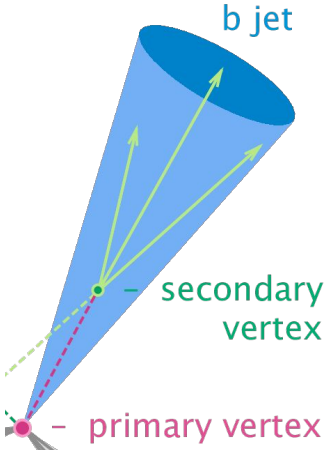
Jet cone with sub-constituents

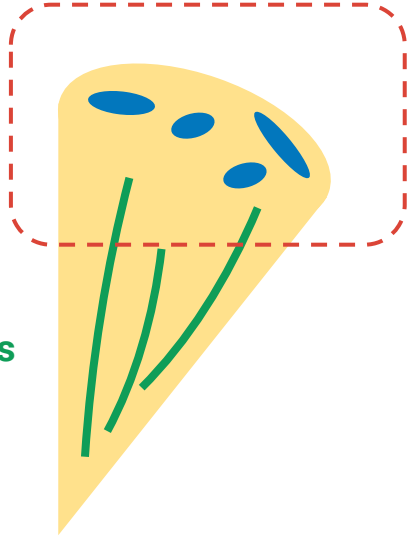


Input format - smarter networks

Do we really know the best variables to identify objects?

Collection of tracks matched to a jet





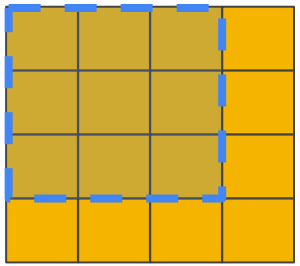
Jet cone with sub-constituents

Lets use these detector objects themselves!

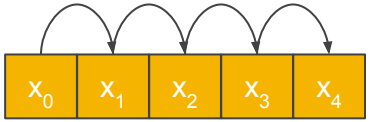
Back to Network Design

What about **unordered data**?

Arbitrary sorting



Forced structure/order
Requires each element to have same meaning



Forced order

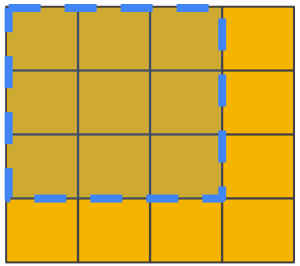
Set of tracks

Set of energy depositions (point cloud)

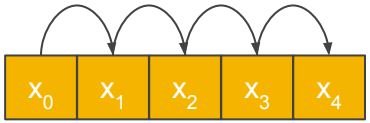
Back to Network Design

- But what about **unordered data**?

Arbitrary sorting

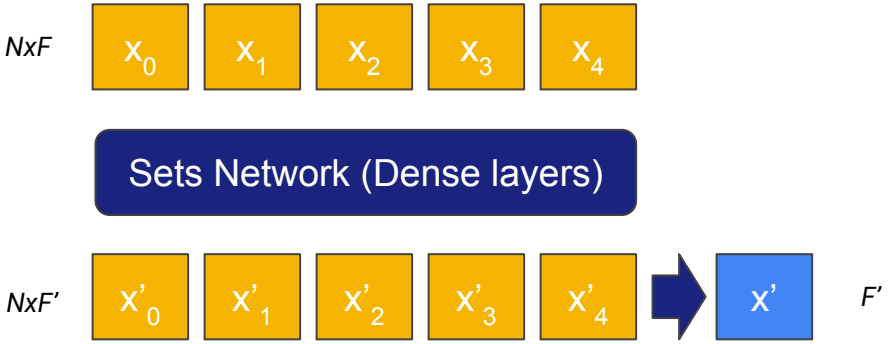


Forced structure/order
Requires each element to have same meaning



Forced order

Deep Sets

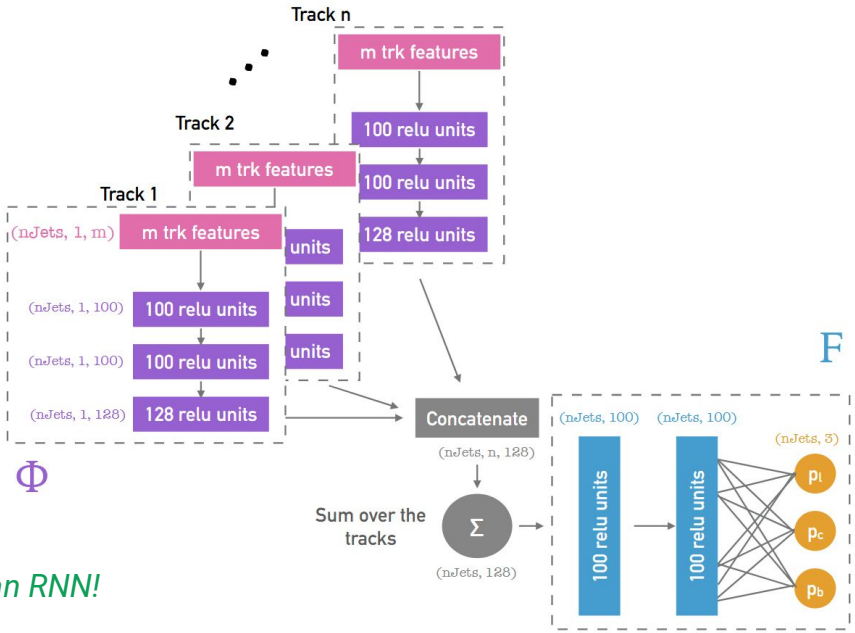


- Apply same operation to each element separately
- Apply pooling to output
- **Permutation invariant**

ATLAS Deep Set for FTag

Replace/supplement handcrafted observables with **Deep Sets over Tracks**

Replaces an RNN with enforced ordering

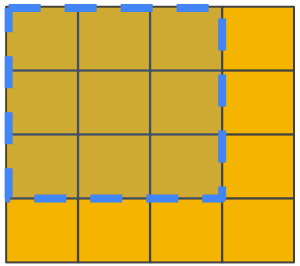


Faster to train, evaluate and better performance than RNN!

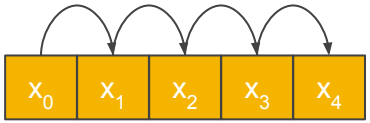
Overview - Network Design

- But what about **unordered spatial data**?

Arbitrary sorting

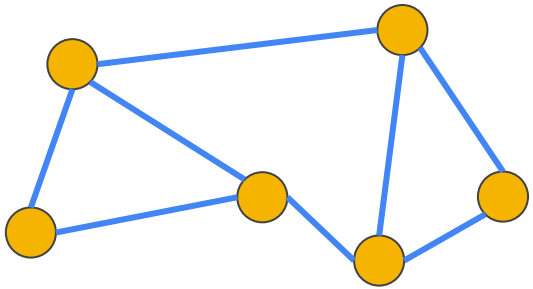


Forced structure
Reduced granularity *or*
Increased dimensionality+sparsity



Forced order

Graph Networks

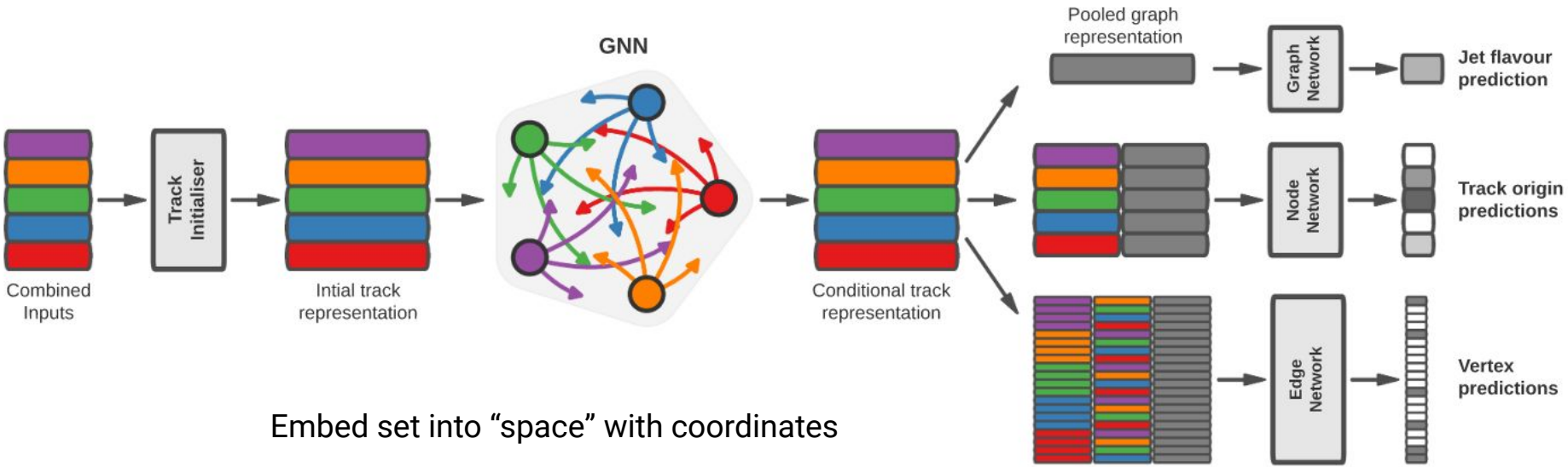


- Operate on nodes and edges
- Update nodes and edge based on connections
- Individual networks work like sets
- **Permutation invariant**
- **Extract maximal relational information**

Cutting edge!

ATLAS Collaboration

ATLAS Graph for FTag



Embed set into “space” with coordinates

Auxiliary tasks

Kasieczka et al

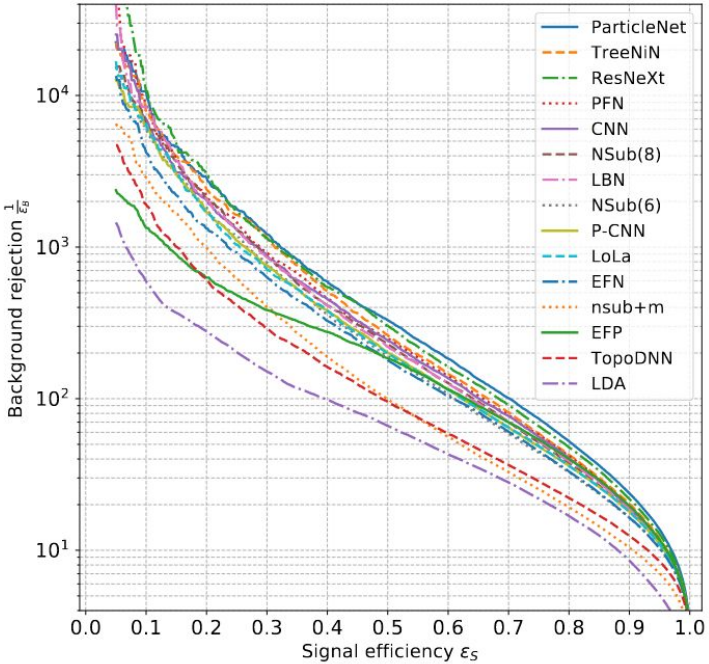
Advanced Networks for Jet Tagging

Convert point cloud to image

Use deep sets

Lorentz invariant layers

Graph networks



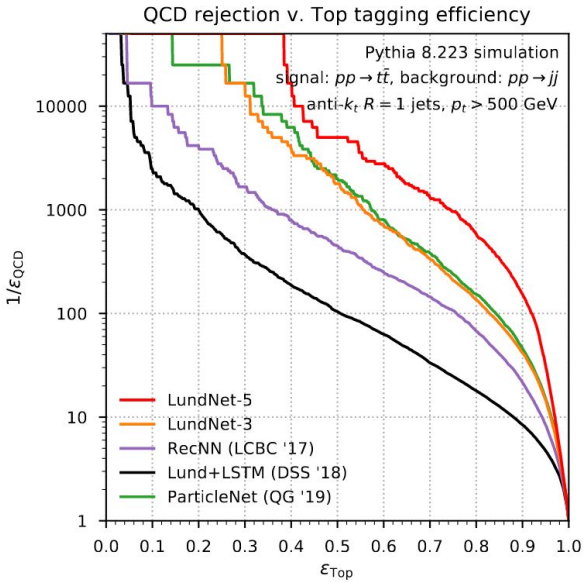
Very active and competitive playground for modern ML!

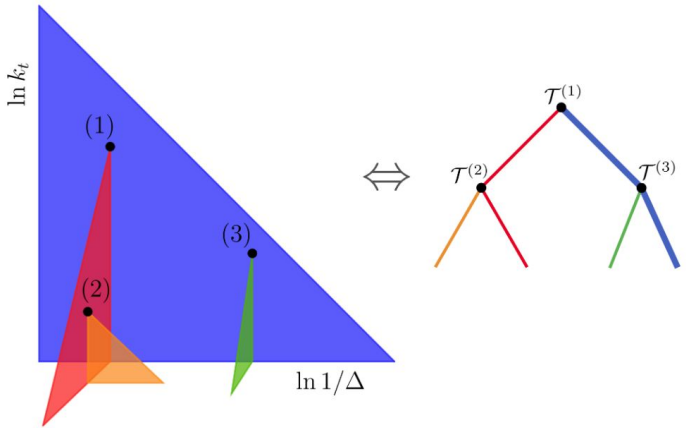
Cutting edge!

Dreyer and Qu

Advanced Networks for Jet Tagging

Not restricted to physical graphs in 2/3D

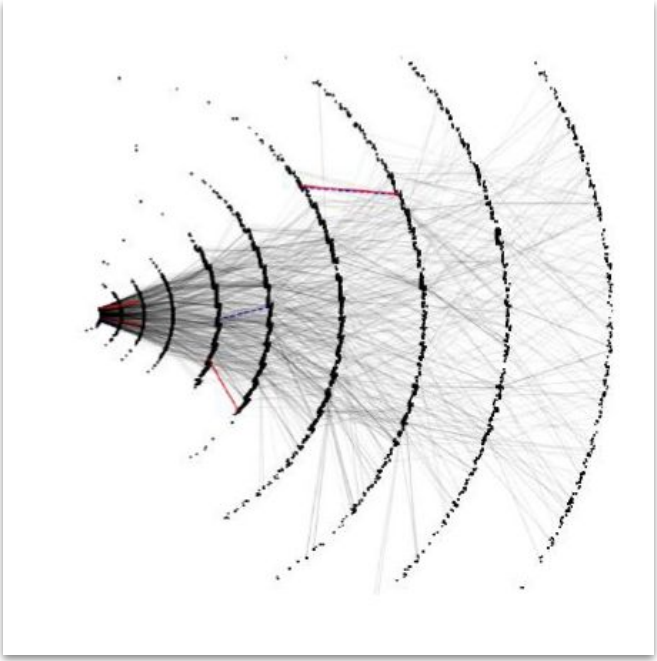




Lund clustering plane for Jets!

Cutting edge!

Graphs for Track Reconstruction



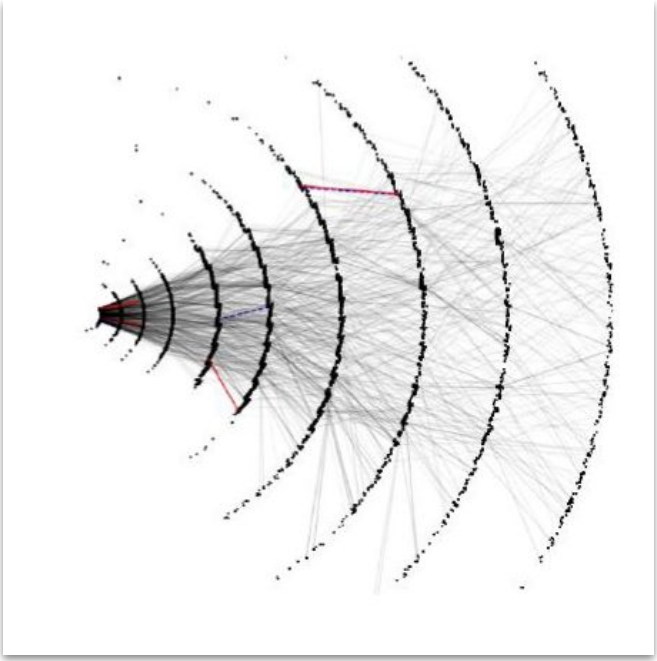
Every hit is a node

Join up the nodes with edges

Identify the “true” edges -> build tracks!

Cutting edge!

Graphs for Track Reconstruction



Every hit is a node

Join up the nodes with edges

Classify the “true” edges -> build tracks!

BUT!

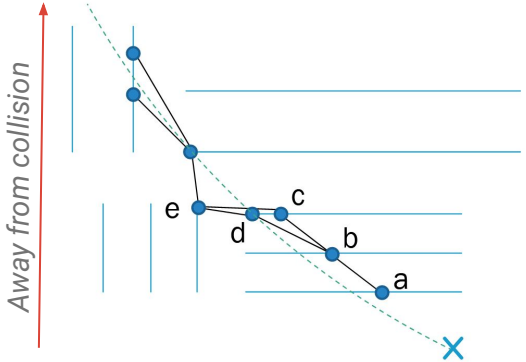
Far too many hits to be efficient connecting them all ($O(N^2)$ edges, N=100k!)

Cutting edge!

ExaTrkX

Graphs for Track Reconstruction

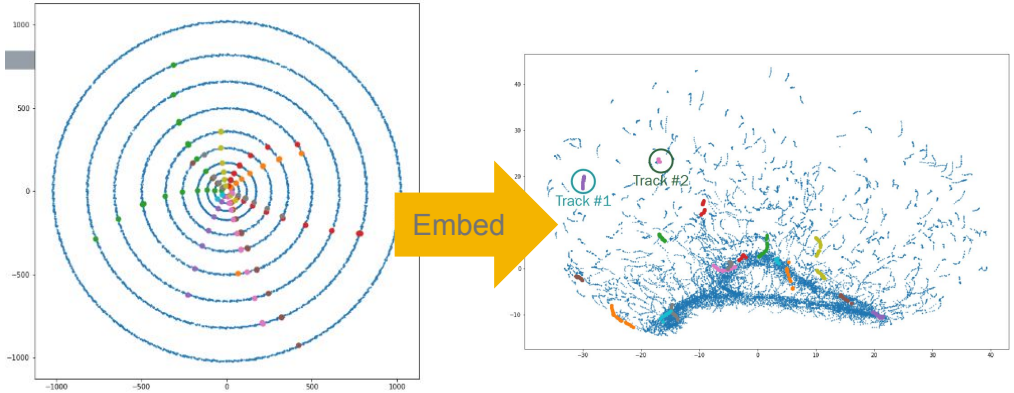
Reduce the complexity of the graph!



Predefine which edges can exist using detector geometry

Embed into abstract space

Track hits should be *localised* - use to define edges

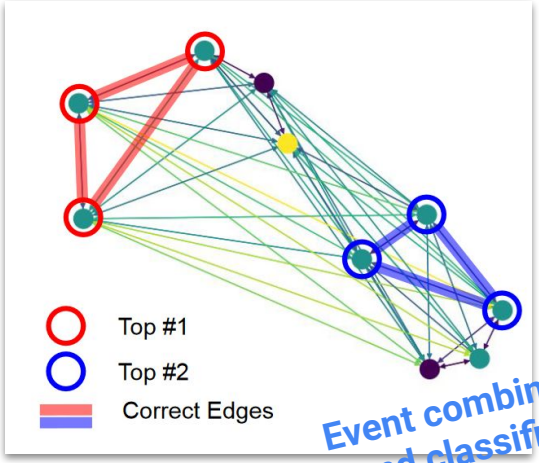


Roberts et al, Shlomi et al , Murnane et al

Graphs in HEP

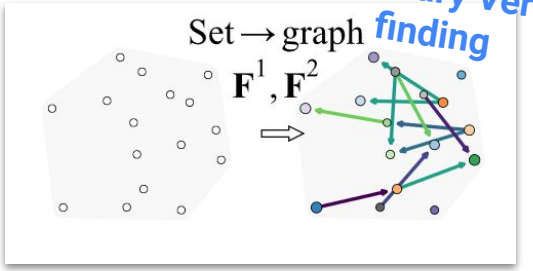
Really starting to become prevalent in HEP

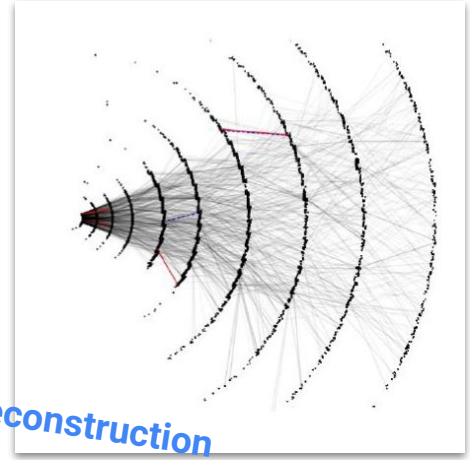
Lots of new developments and growing fast!



Event combinatorics and classification

Secondary Vertex finding





Track reconstruction

Classification

Works great when you know what you are looking for!

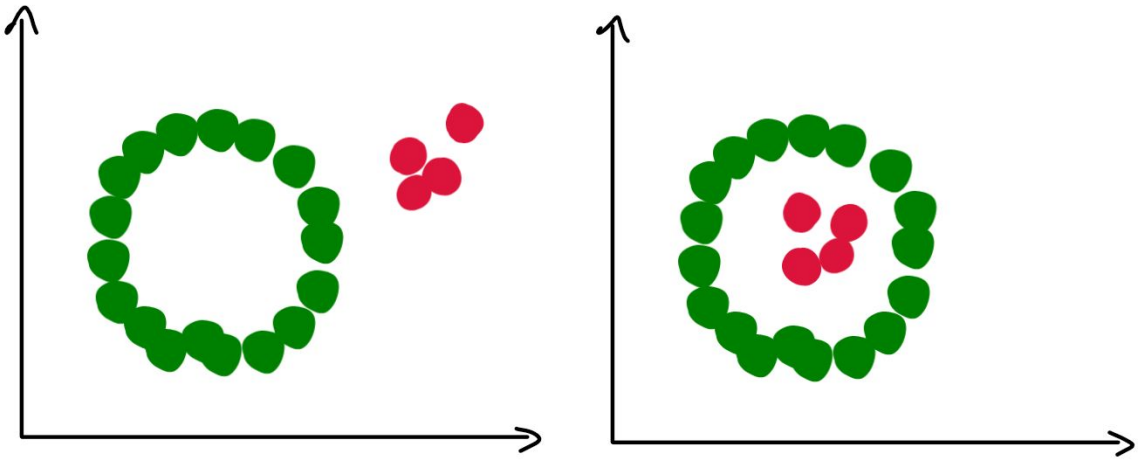
Can build on prior physics knowledge or let networks learn on low level info

Bedrock for physics programme at LHC - particle ID, sensitivity to small signals

But what if we don’t know what we’re looking for…

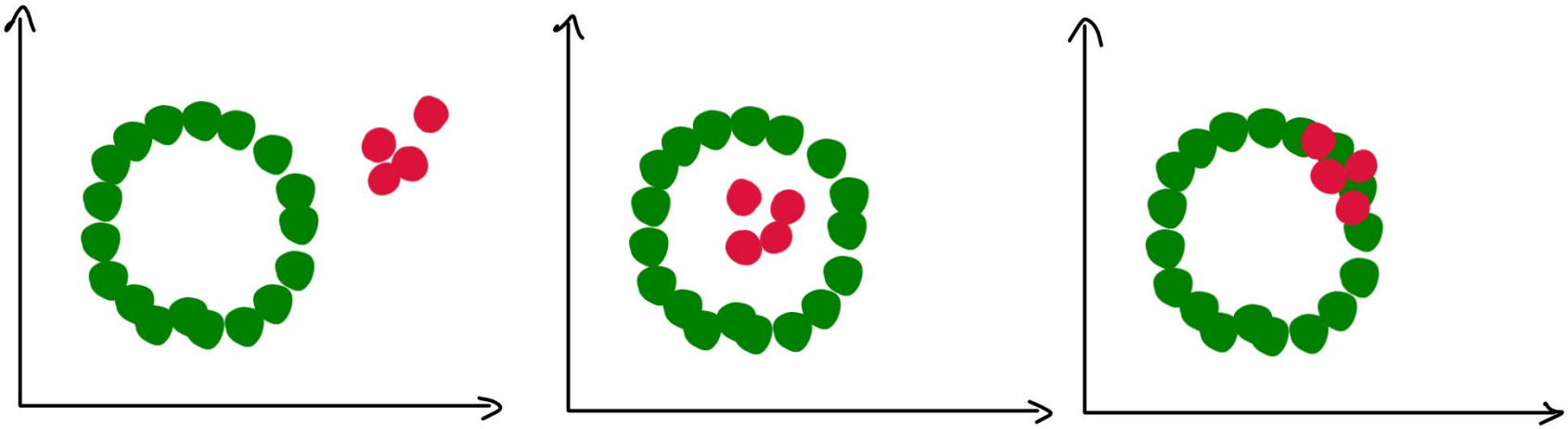
Anomaly detection

Unfortunately this isn’t the case…



Anomaly detection

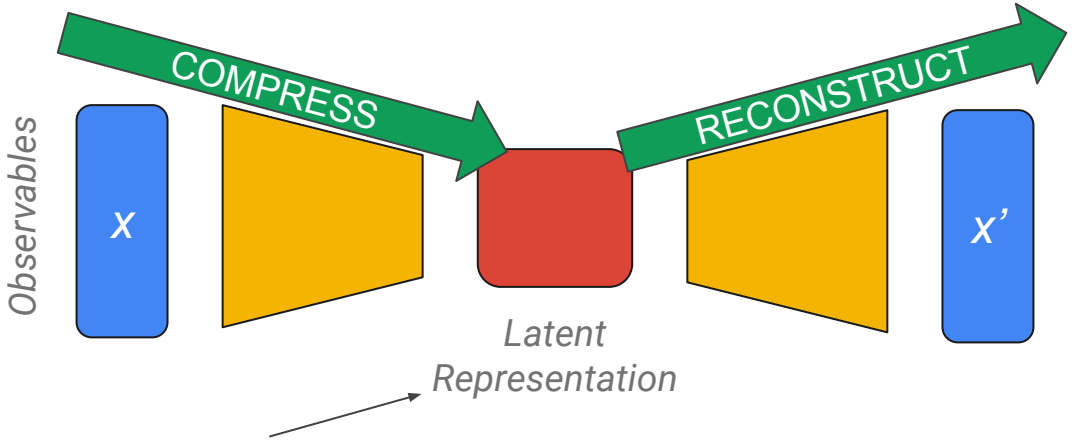
Unfortunately this isn't the case…

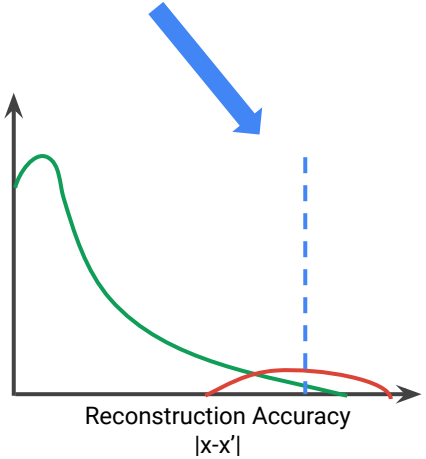


Looking for an unexpected excess, usually something uncommon

Anomaly detection

Can we design a “tagger” to identify something as *weird*

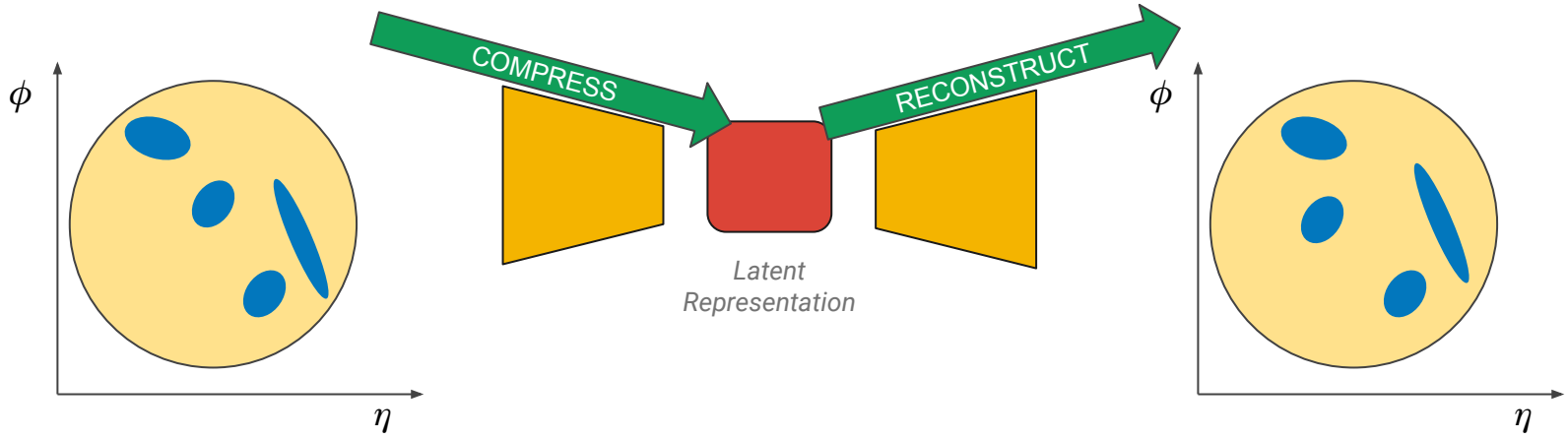




Learn what it is to be *“normal”*

Anomaly detection

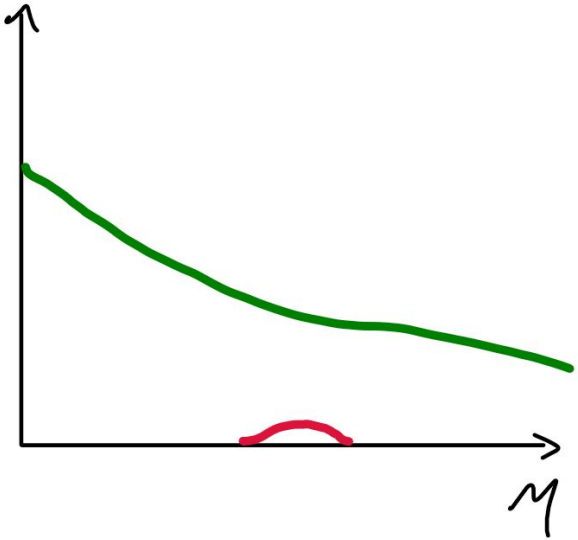
Apply to boosted high pT jet

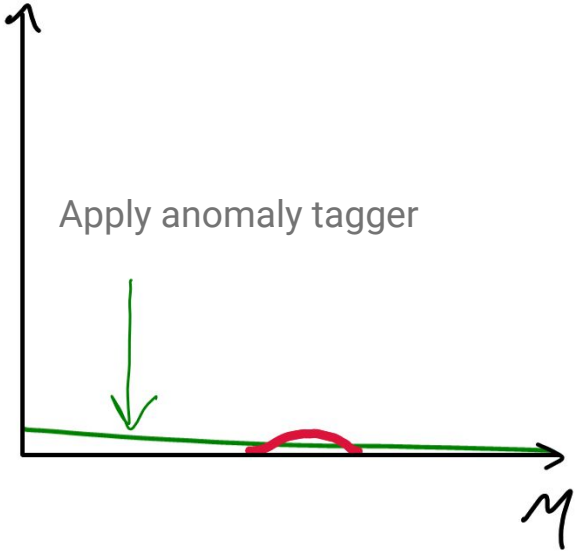


Not looking for top or W but something that isn't QCD!

Anomaly detection

Potential signal hidden under dominant background



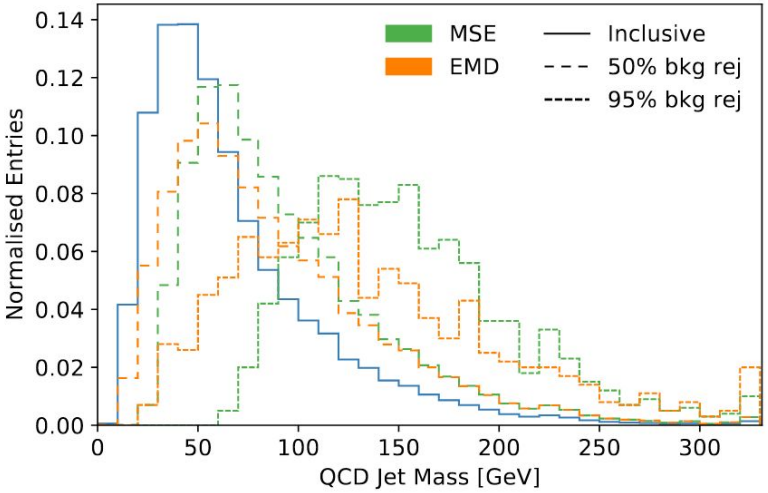


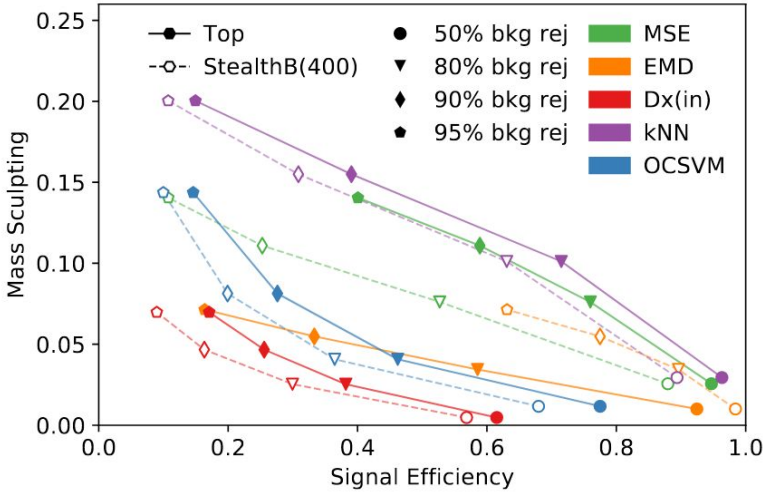
Identify an excess!

Golling et al

Anomaly Detection

Unfortunately without tricks end up just finding “massive = rare = anomalous”





Anomaly Detection

Semi-supervised

Train on background only

Learn where things should be

Use to reject normal, enhance weird

Unsupervised

Train on data

Network groups things

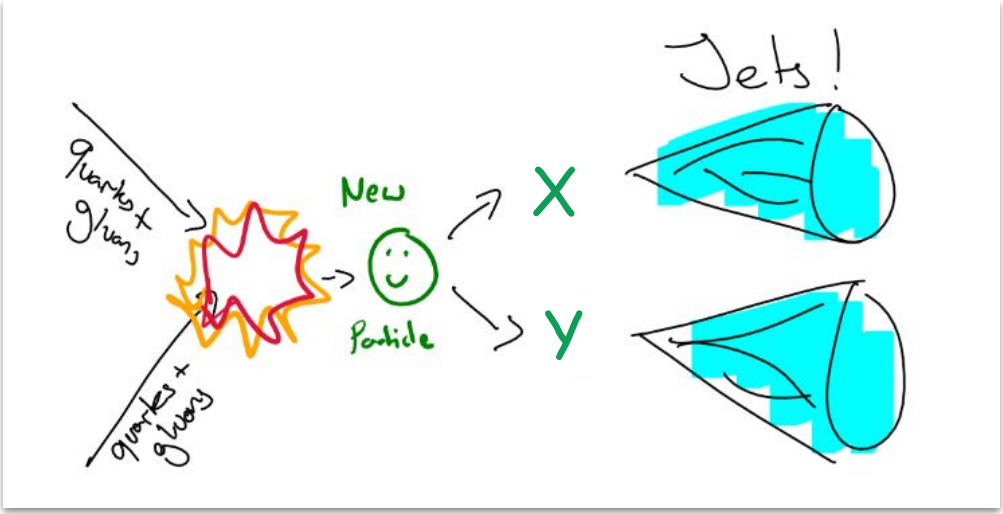
Separate weird from normal

Anomaly Detection

Instead of a general “anomaly tagger” develop upon standard bread and butter

Assume new massive particle

Decays (eventually) to two jets



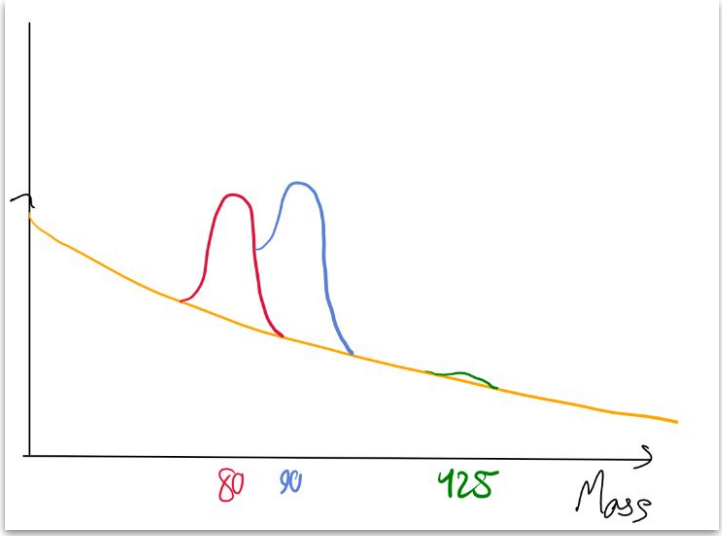
Anomaly Detection

Instead of a general “anomaly tagger” develop upon standard bread and butter

Assume new massive particle

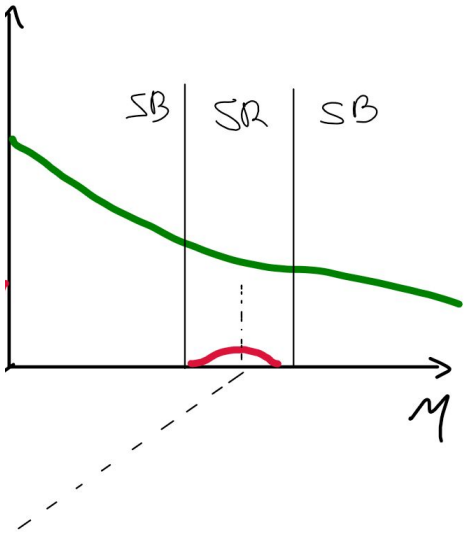
Decays (eventually) to two jets

Look for a bump in sliding window fit



Anomaly Detection - Bump hunt

But what if the bump is dominated by background…



Anomaly Detection - Bump hunt

But what if the bump is dominated by background…

Is there more information we can use than just mass?

If we use ML what data do we train on?

Anomaly Detection - Bump hunt

But what if the bump is dominated by background…

Is there more information we can use than just mass?

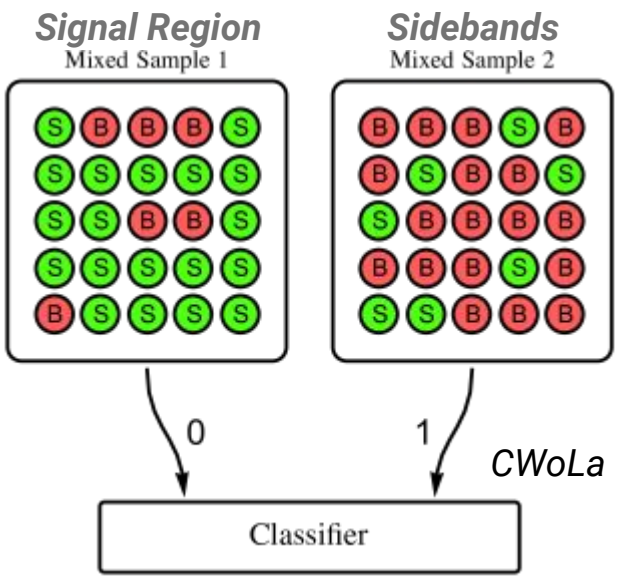
High mass - boosted decays
Look at substructure!

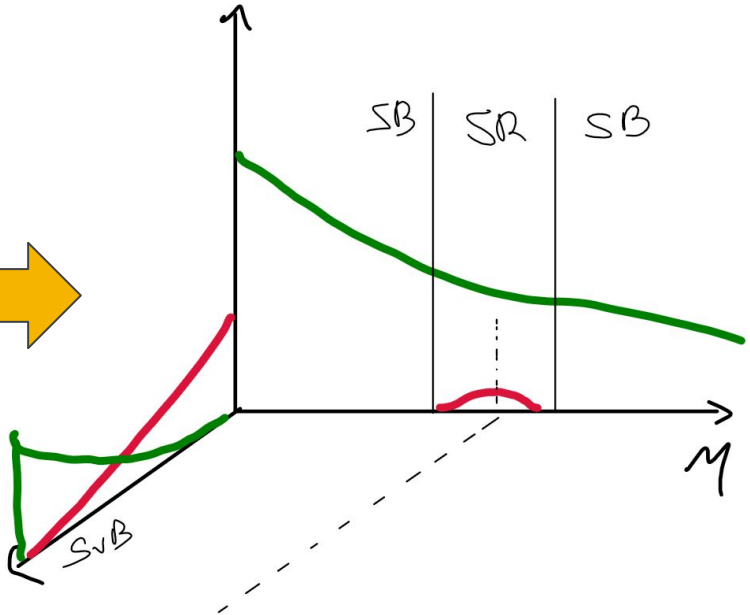
What do we train on?

Control region data
vs
Signal region

Noisy labels - weakly supervised!

Anomaly Detection - Bump hunt

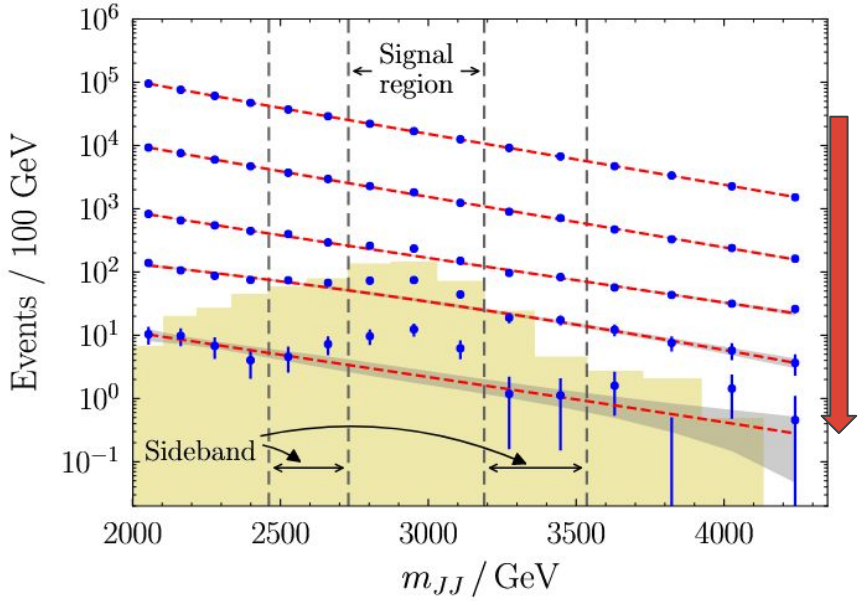


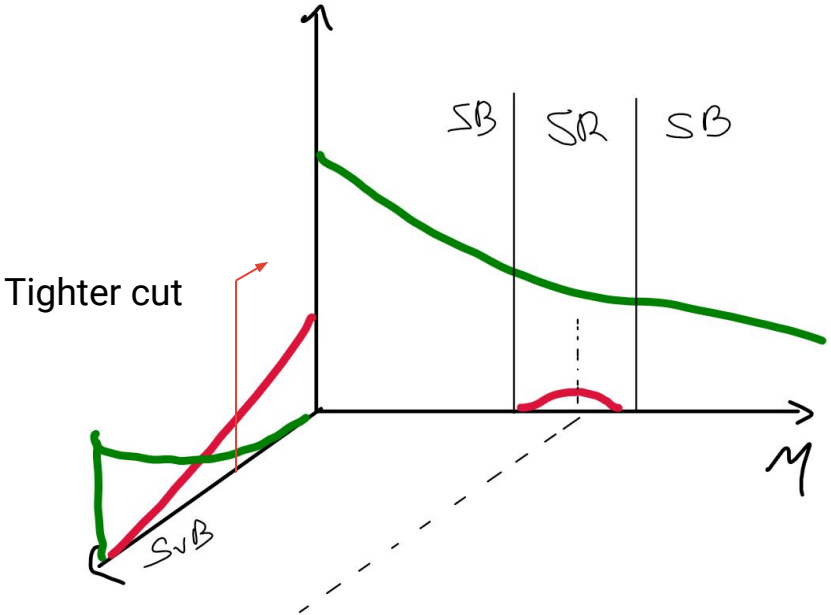


Optimal for S vs B!

Collins et al

Anomaly Detection - Bump hunt





Anomaly Detection - Bump hunt

Works really well unless observables in classifier are *correlated with Mass*

Otherwise:

Classifier learns differences in Mass distribution between the Sidebands and Signal region

Bias!

Cutting edge!

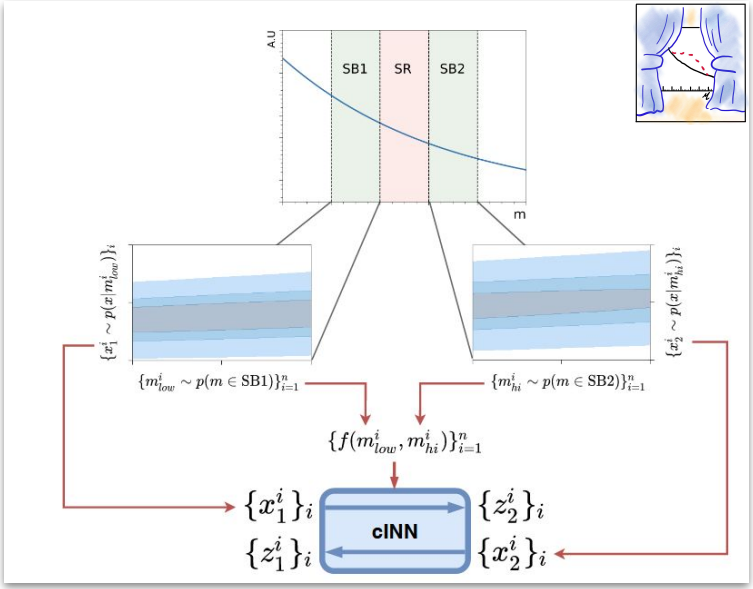
Raine et al

Anomaly Detection - Bump hunt

Works really well unless observables in classifier are *correlated with Mass*

Introduce *CURTAINs*!

Learn to **transform data into signal region** with cINN

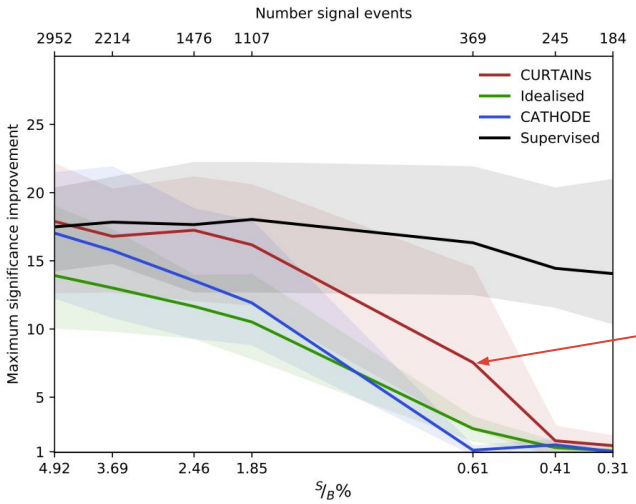


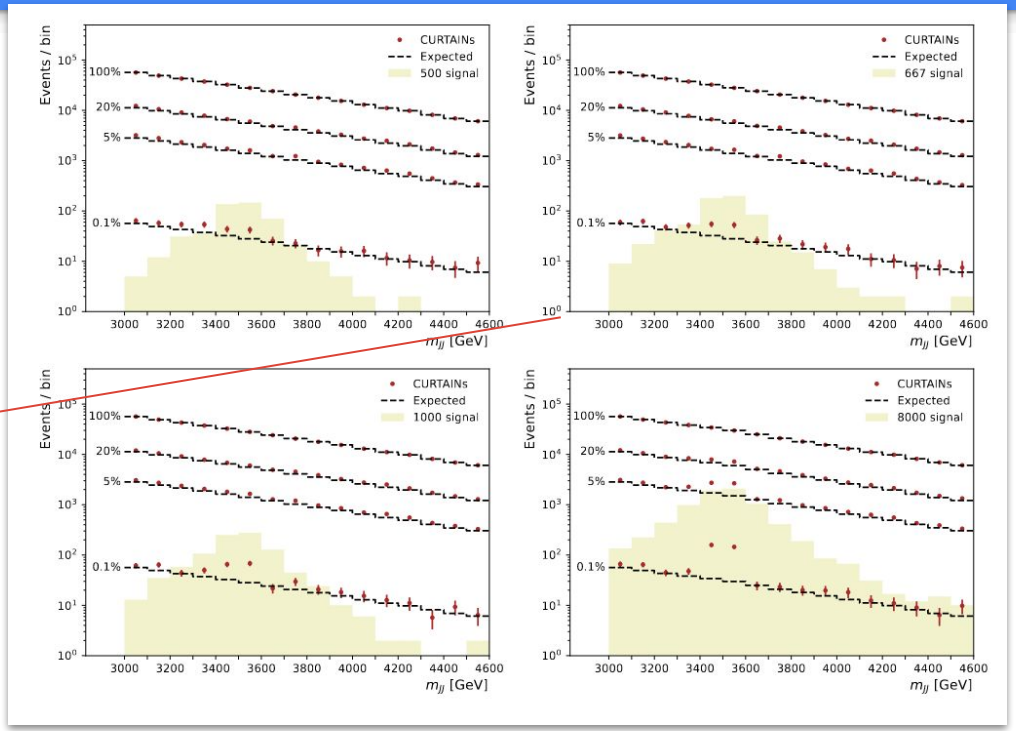
Constructing Unobserved Regions by Transforming Adjacent INtervals

Cutting edge!

Raine et al

Anomaly Detection - Bump hunt





Regression

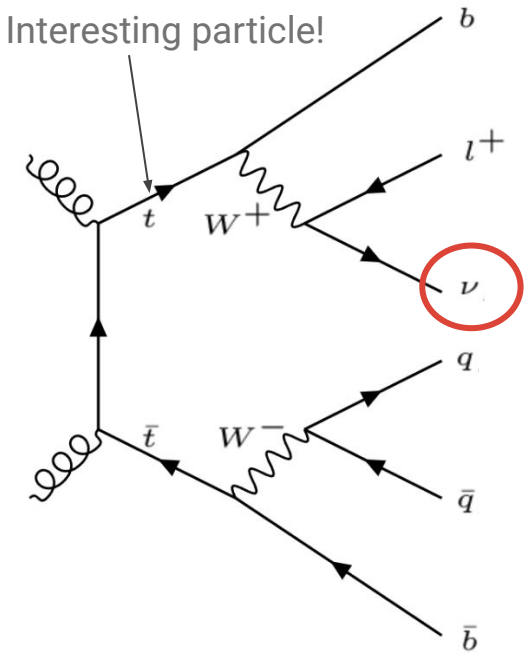
Regression also used in HEP but not as much as classification

e.g. Correction of object energy as recorded by detector

Most of what applied to classification holds here too

- Watch out for biases
- Design your network to match your data

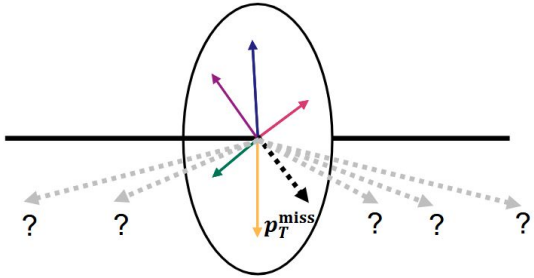
Neutrino Regression



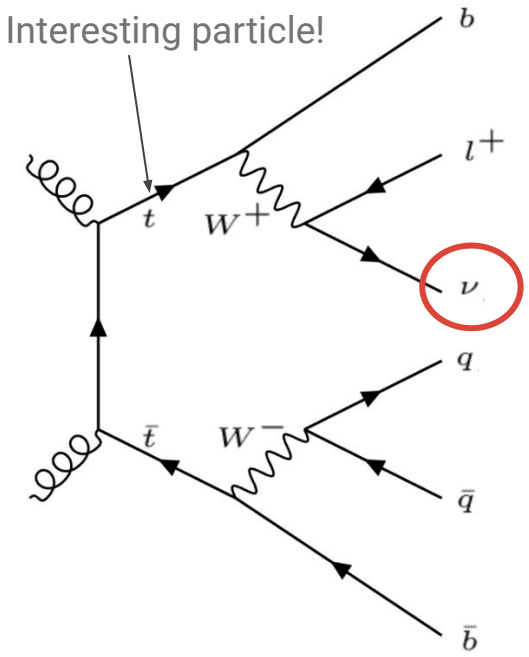
Example: Predicting neutrino momentum

Neutrinos don't interact at all with detector!
No longer able to fully reconstruct **top quark**

Conservation of momentum helps constrain direction perpendicular to collisions

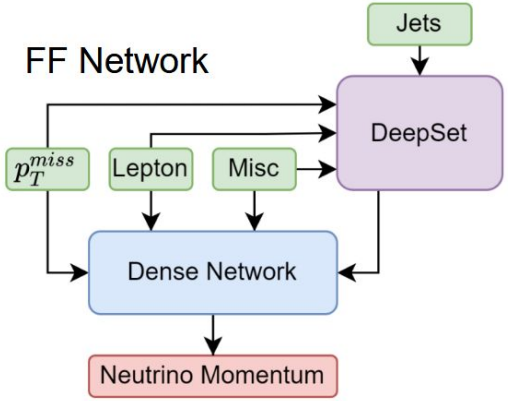


Neutrino Regression

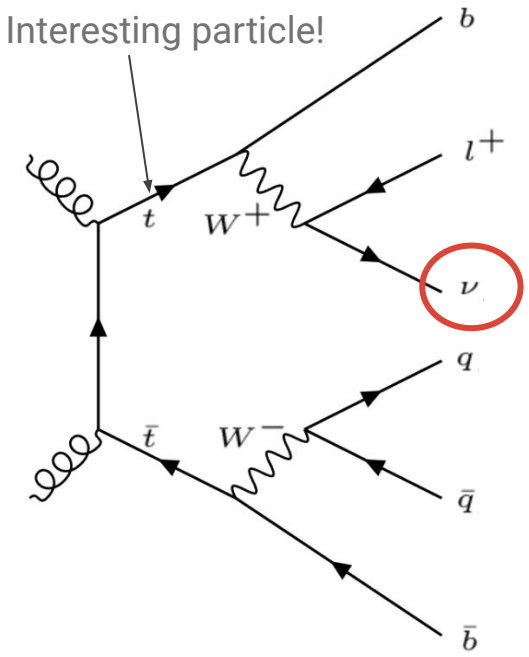


Standard approach uses W mass to constrain potential solution

Use ML to predict neutrino momentum using event observables



Neutrino Regression



Standard approach uses W mass to constrain potential solution

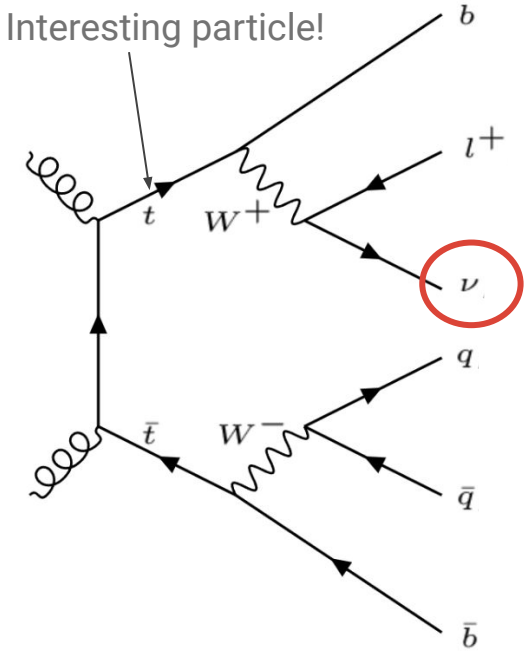
Use ML to predict neutrino momentum using event observables

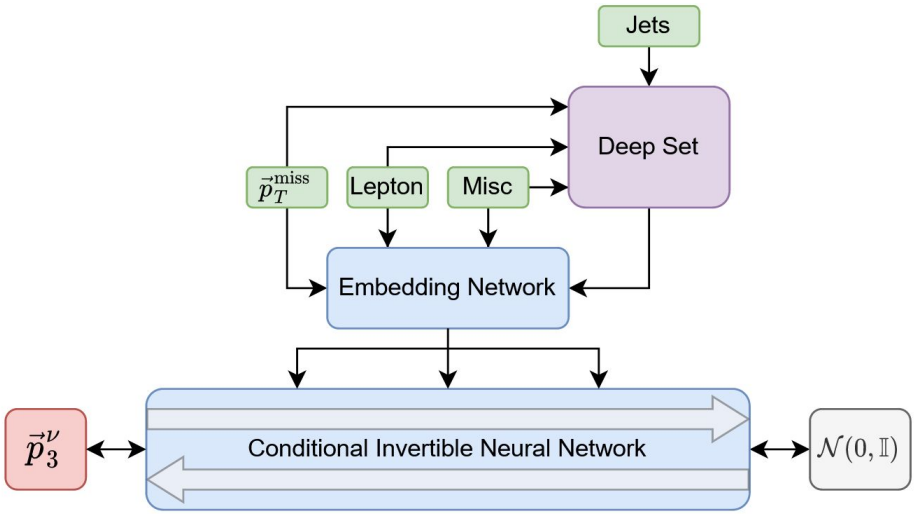
But! Trying to recover a degree of freedom, standard network can't easily handle this…

Cutting edge!

Leigh et al

Neutrino Regression



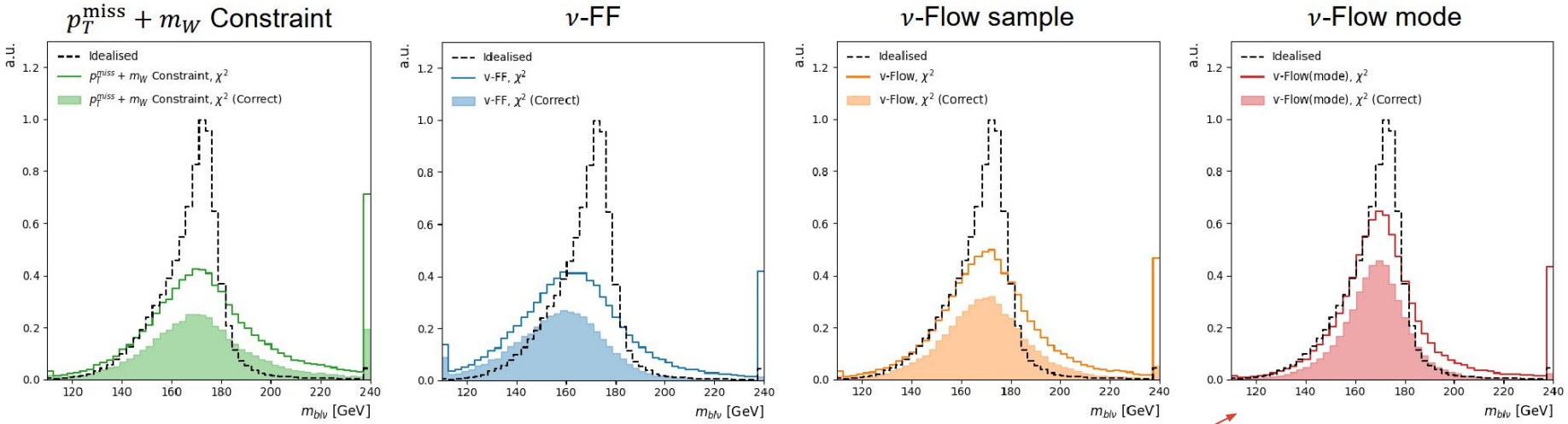


Use networks with built in variation!
Extendable to multiple neutrinos!

Cutting edge!

Leigh et al

Neutrino Regression



Better performance in downstream tasks and better reconstruction of top mass!

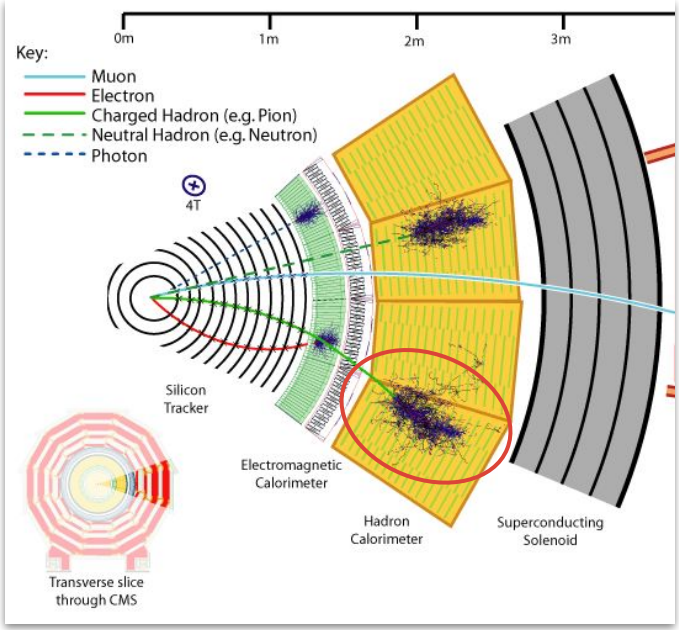
Speed up detector simulation

Detector simulation (calorimetry!) is very slow

Simulate huge number of interactions in detector material

At the end need energy deposits in readout cells of detector

Perfect for generative modelling



High multiplicity of secondary particles in shower

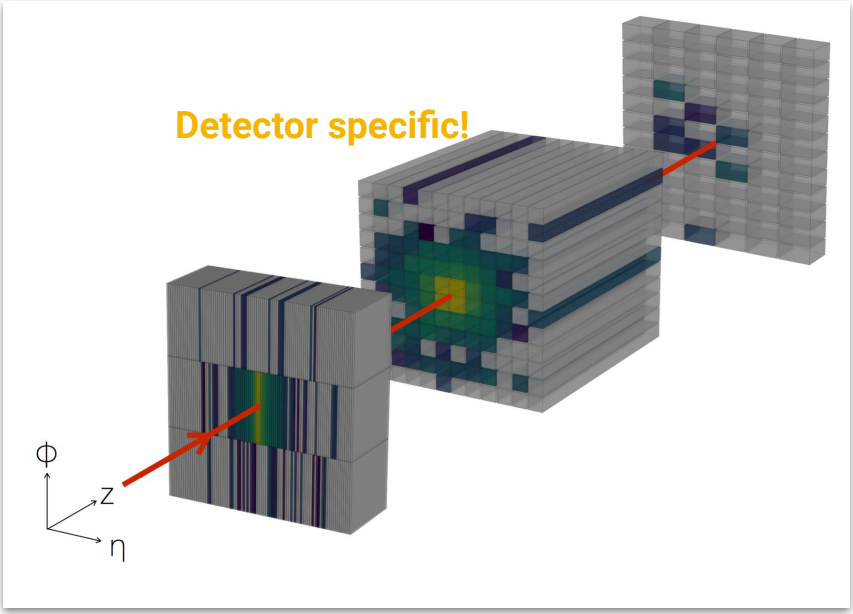
Speed up detector simulation

Individual particle traverses detector

Energy deposited in many cells from secondary particles in shower

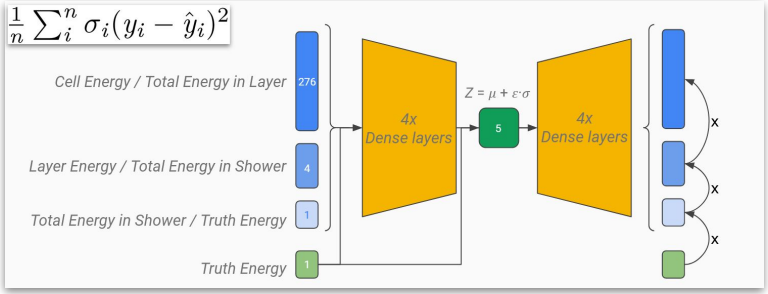
Can build an “image” but

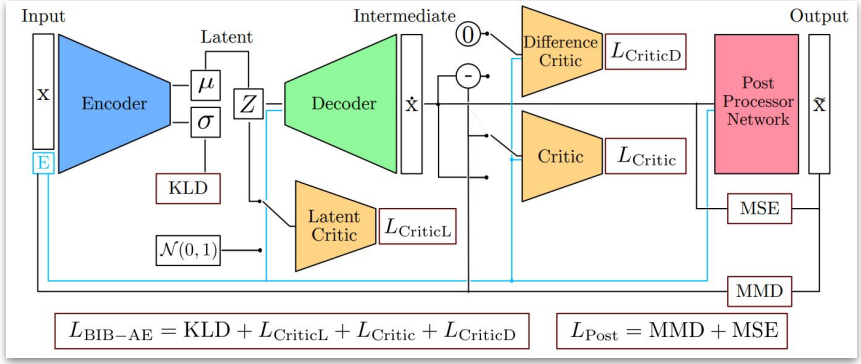
- **High dynamic range** of “pixels”
- Often very **sparse**
- **Stochastic** - same incoming particle results in different shower

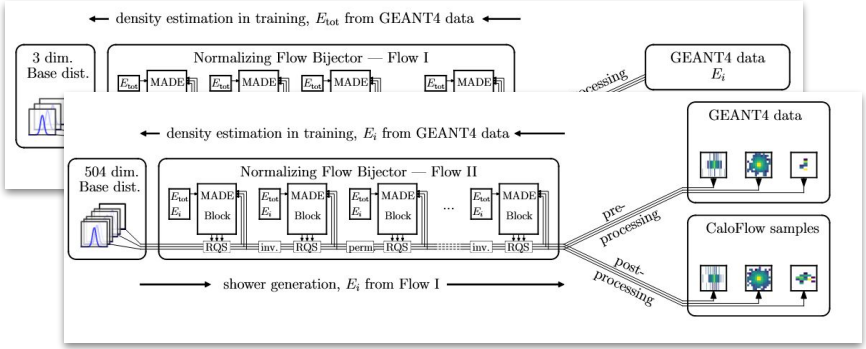


Paganini et al (2017), Krause, Shih (2021), ATLAS Collaboration (2019, 2019), Erdmann et al (2018)

Speed up detector simulation





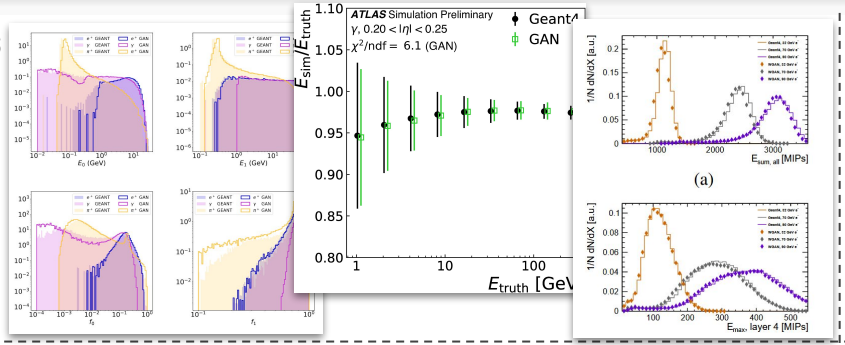


Wide range of models of varying complexity

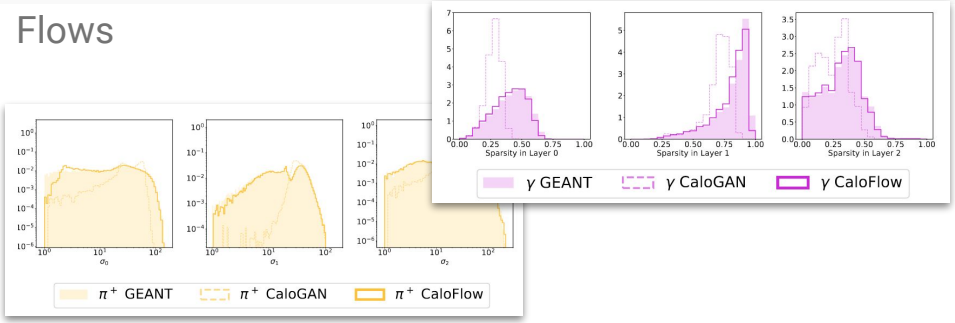
Paganini et al (2017), Krause, Shih (2021), ATLAS Collaboration (2019, 2019), Erdmann et al (2018)

Speed up detector simulation

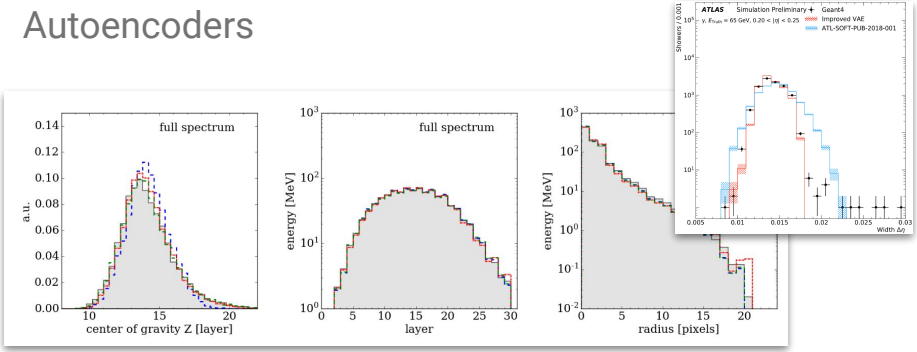
GANs

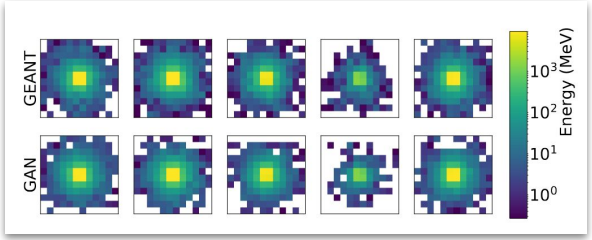


Flows



Autoencoders





Performance measured over lots of distributions
Impossible to judge individual showers

Things I've not had time for

Unfolding and statistical interpretations

Combinatoric solving and event reconstruction

Uncertainty aware approaches

Speeding up Monte Carlo simulation

Multidimensional reweighting

Reconstructing physics objects with GNNs

Region definitions for analyses

Optimal transport for calibrations

Conclusions

Almost every aspect of ML is being employed in HEP

and

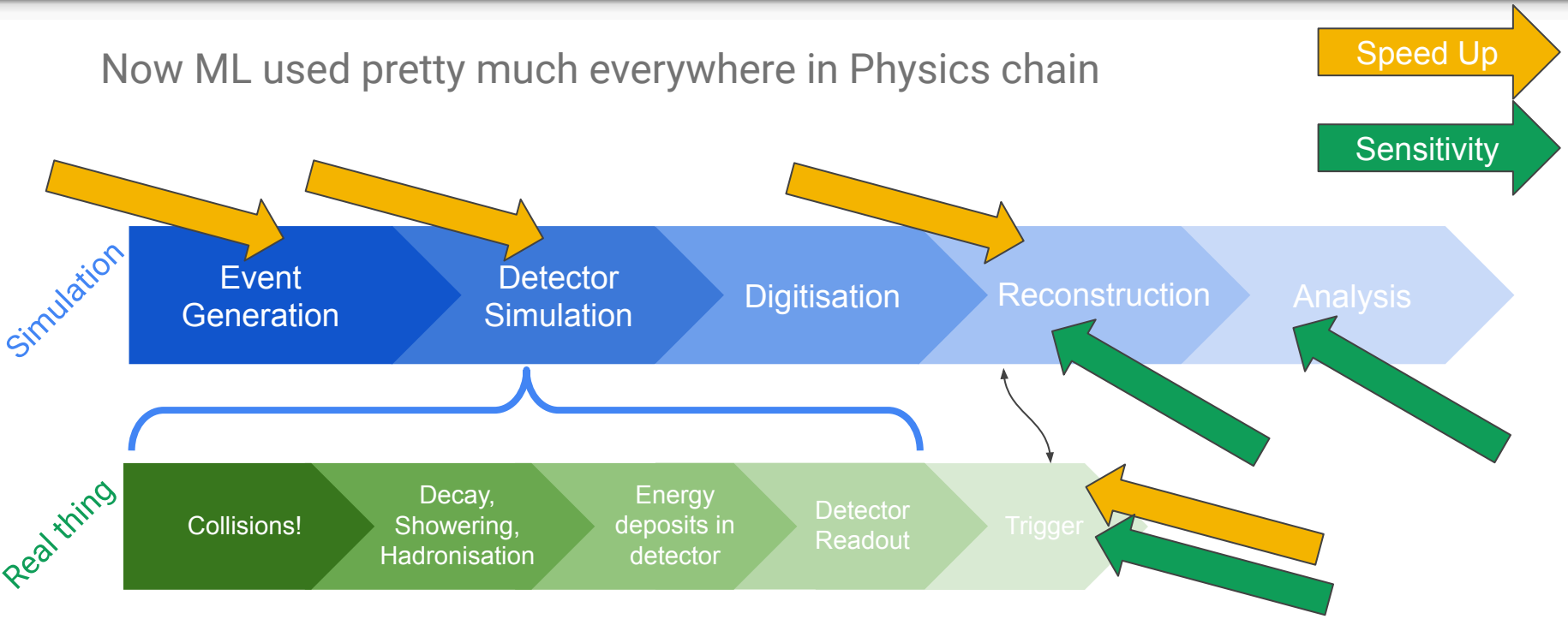
Most areas of HEP are using ML

A lot of complex R&D in ML is done addressing challenges in HEP, to improve sensitivity and reach of collider experiments

Really fun field to be involved in and now is the best time to join the fun!

ATLAS and CMS

Now ML used pretty much everywhere in Physics chain



Tutorial this afternoon

This afternoon we'll return to Identifying high pT jets!

Setting up the multiclass classification

Addressing bias

Combining into binary classifier

You could even use this data with the Anomaly Detection from yesterday instead of classification!
QCD = nominal, W/Z and top both anomalous!