Project Pitch

PHYS 591000 Spring 2021

Project Overview

Week 11

- Higgs Production
  - Classification
- Classical 2D Ising model
  - Classification

Week 12

- ML4Pions
  - Regression
- Charged Particle Tracking
  - Classification
- Dijet Generative Model
  - Generative model, Classification

Week 13

- Top Jets
  - Classification
- Electron showers with Emulsion detector
  - Regression, Classification

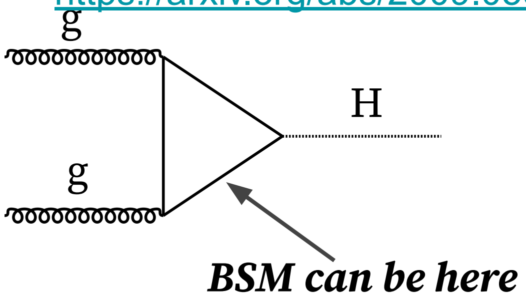
Higgs Production

Credit: Yi-Lun (Alan) Chung

https://arxiv.org/abs/2009.05930

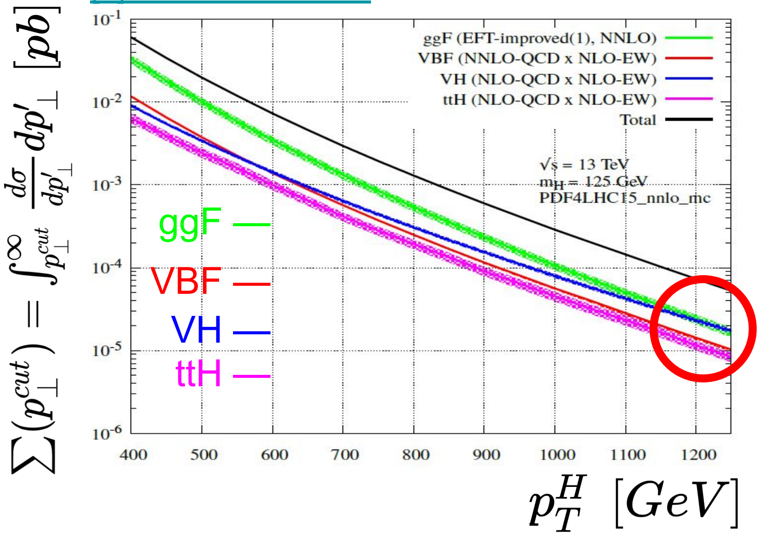
Motivation

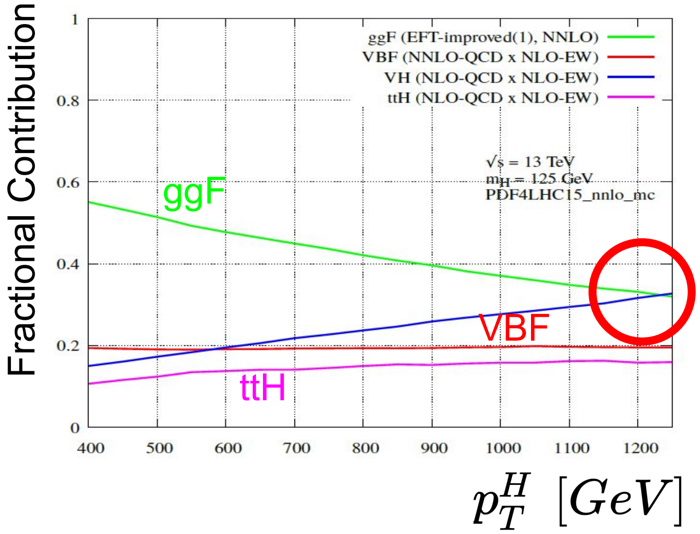
- High pT Higgs from
  - the SM Higgs, e.g. ggF
  - Beyond the Standard Model



- Many Higgs productions other than ggH could be substantial in the boosted region.

[1] arXiv:2005.07762

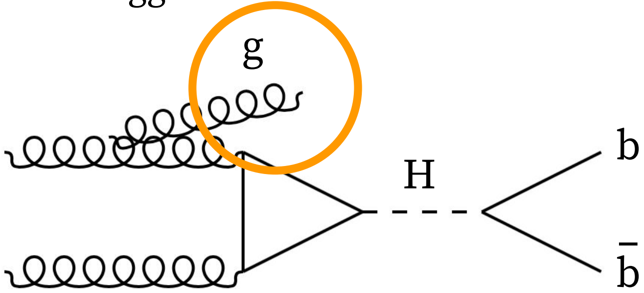




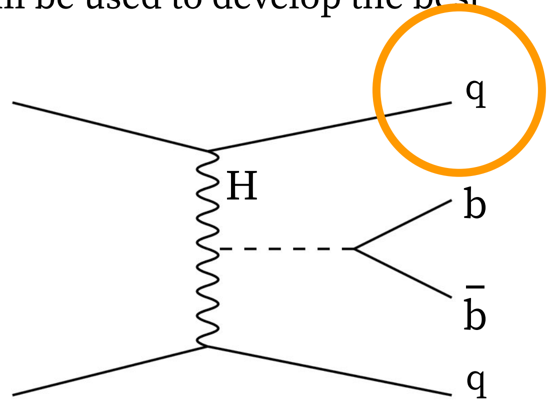
Leading Non-Higgs Jet

- The **leading non-Higgs jet** substructure and global event topology will be used to develop the best boosted Higgs classifier.

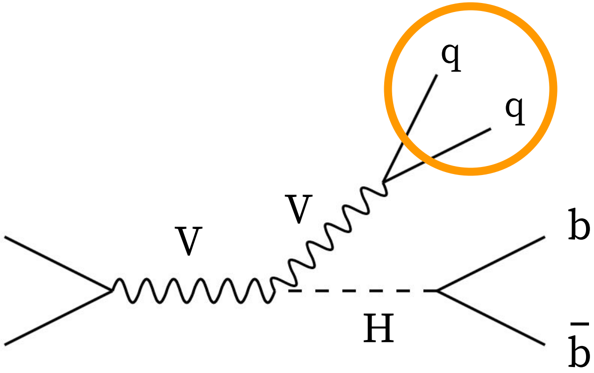
ggF



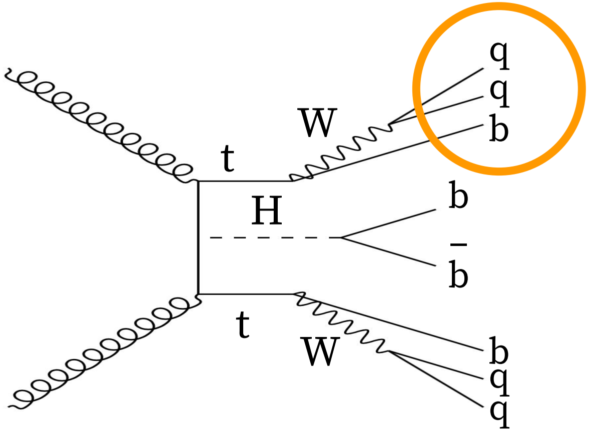
VBF



VH



ttH



Data Structure-Constituent's Table

- **Kaggle:** https://www.kaggle.com/t/c6f28ed7a3c44570af674e426c2bcfba
- **Data:**
  - High-level features: [M_j, η_j, $|\Delta\eta_{jj}|$, M_{jj}, girth, central integral jet shape Ψ]
  - Low-level features: pt, eta, phi, rel_eta, rel_phi, jet_index, process, and label of the constituents.
    - rel_eta = $\eta_{constituent} - \eta_{jet}$; rel_phi=$\phi_{constituent} - \phi_{jet}$
  - There are 35000 events for each production for training.
  - We label the ggF process to be 0, VBF process to be 1, VH process to be 2, and ttH process to be 3.
- **References:**
  - arXiv:1507.00508
  - arXiv:1807.10768
  - arXiv:2009.05930

Classical 2D Ising Model

Credit: Daw-Wei Wang

Classical 2D Ising Model

https://en.wikipedia.org/wiki/Ising_model

1. Ising Model is a fundamentally important model discussed in condensed matter physics, because it is the simplest nontrivial model with exact solution available in 1D and 2D system. The latter shows a phase transition in the thermodynamic limit.

2. The general quantum many-body Hamiltonian is like following

$$\hat{H} = -J\sum_{\langle i,j\rangle}\hat{\sigma}_i^z\hat{\sigma}_j^z - B\sum_i \hat{\sigma}_i^x$$

, where J and B are spin exchange energy and external field. $\hat{\sigma}_i^{x,y,z}$ is Pauli matrix at a lattice site, i. <*i,j*>means the pair of the nearest neighboring sites.

3. In our example, we consider classical version with B=0, so that the spin has only two values ($\sigma^z = \pm 1$). At zero temperature, the system energy and magnetization for any give spin configuration, $\{\sigma\}$, are respectively

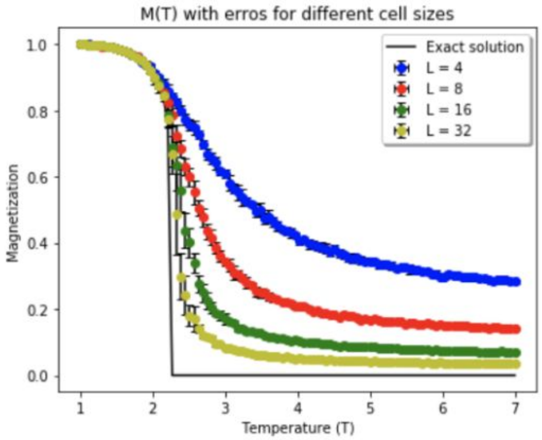
$$E[\{\sigma\}] = -J\sum_{\langle i,j\rangle}\sigma_i^z\sigma_j^z \qquad M[\{\sigma\}] = \frac{1}{N}\sum_i \sigma_i^z$$

4. At finite temperature, the partition function in canonical ensemble and the average magnetization (order parameter) are given by

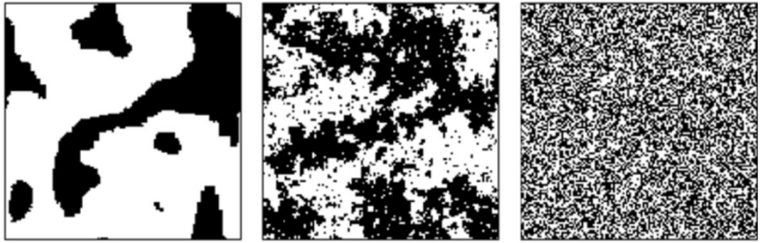
$$Z(T) = \sum_{\{\sigma\}} e^{-E[\{\sigma\}]/T} \qquad M(T) = \frac{1}{Z}\sum_{\{\sigma\}} e^{-E[\{\sigma\}]/T} M[\{\sigma\}]$$

5. For 2D classical Ising, it is known to have a finite temperature phase transition in the thermal dynamic limit. The order parameter (magnetization, M) will have a discontinuous change from a finite value (ferromagnetic) as $T<T_c$ to zero (paramagnetic) as $T>T_c$. See for example

[Jonathan Leban 2D Ising Model]



6. Now the task is to check if we could use pattern recognition of machine learning to predict the critical temperature for such a phase transition from the spin configuration directly. The spin configuration should be quiet uniform in ferromagnetic (FM) phase and randomly fluctuated in paramagnetic (PM) phase.



$T<T_c$ (FM) $\quad T=T_c \quad$ $T>T_c$ (PM)

[Kitzbichler, et. al. PLoS Computational Biology (2009)]

Data Structure

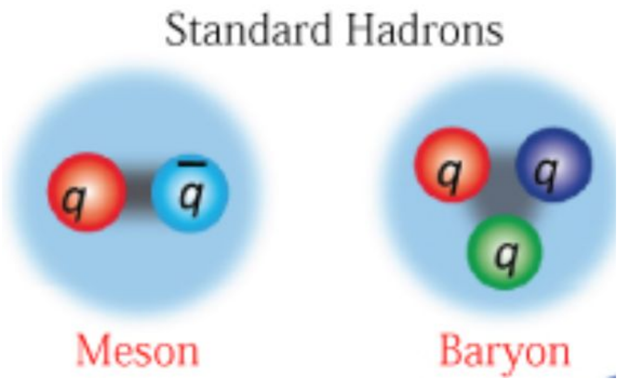
- Kaggle: https://www.kaggle.com/t/e34d404cbba14173ae4729597f2ae43c
- **x_data:**
  - spin configuration for a system with 10*10 system size and periodic boundary condition, using 1 & -1 to represent up/down.
  - x_data have 1001 different spin configurations in different temperature. Temperature is stored in data with key=[temperature]. The critical temperature $T_c \sim 2.38$ corresponding to the maximum point of magnetic susceptibility.
  - 100 ensembles for each temperature.
- **y_data :**
  - one-hot labels to show which is paramagnetism or ferromagnetic. 0 for paramagnetism and 1 for ferromagnetic.

Project Pitch II

PHYS 591000 Spring 2021

ML4Pions

Hadronic Shower

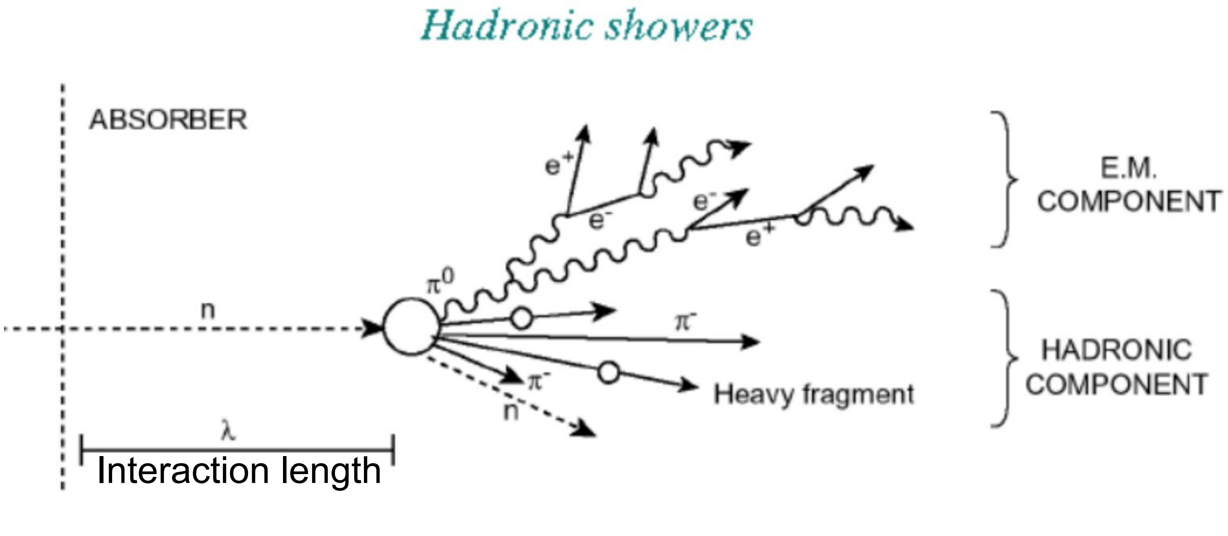


E.g. pions

π^+: $u\bar{d}$
π^0: $u\bar{u}$ or $d\bar{d}$
π^-: $d\bar{u}$

E.g. protons

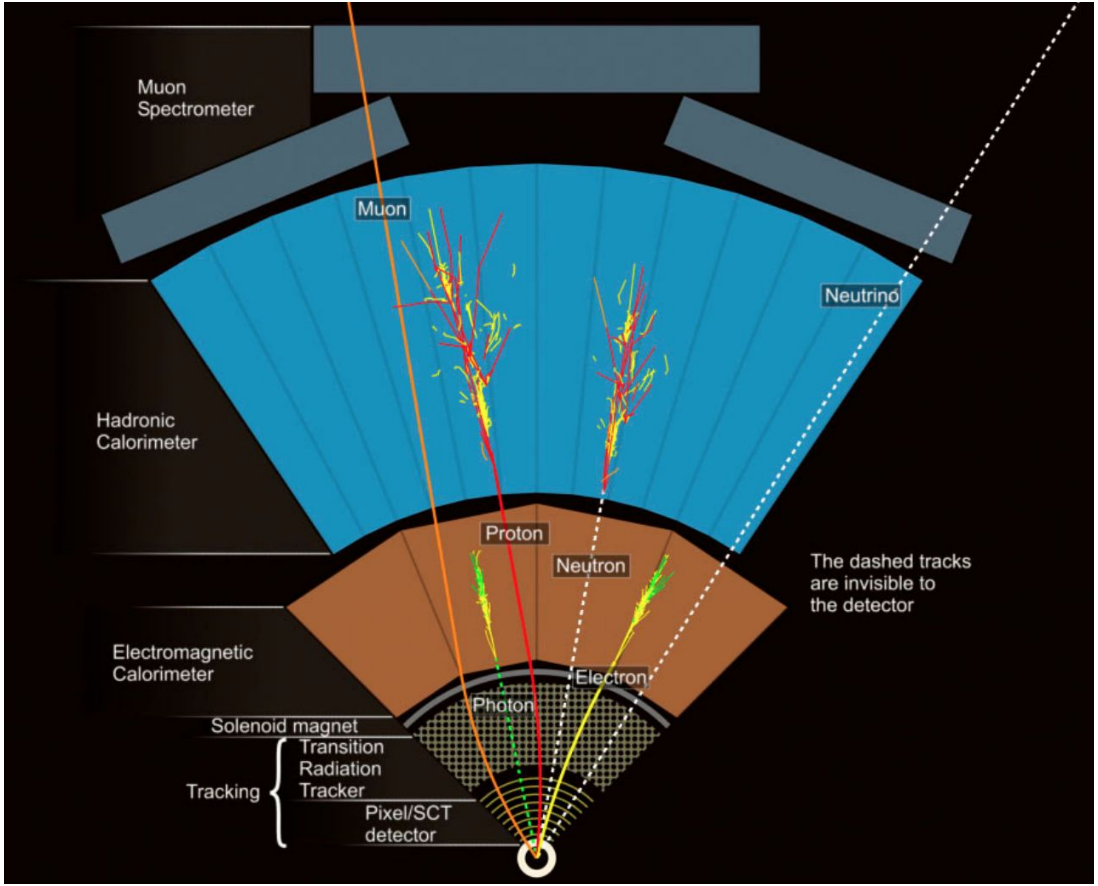
p (uud)
n (udd)



T.S.Virdee, Proc. of the 1998 European School of High-Energy Physics, CERN 99-04

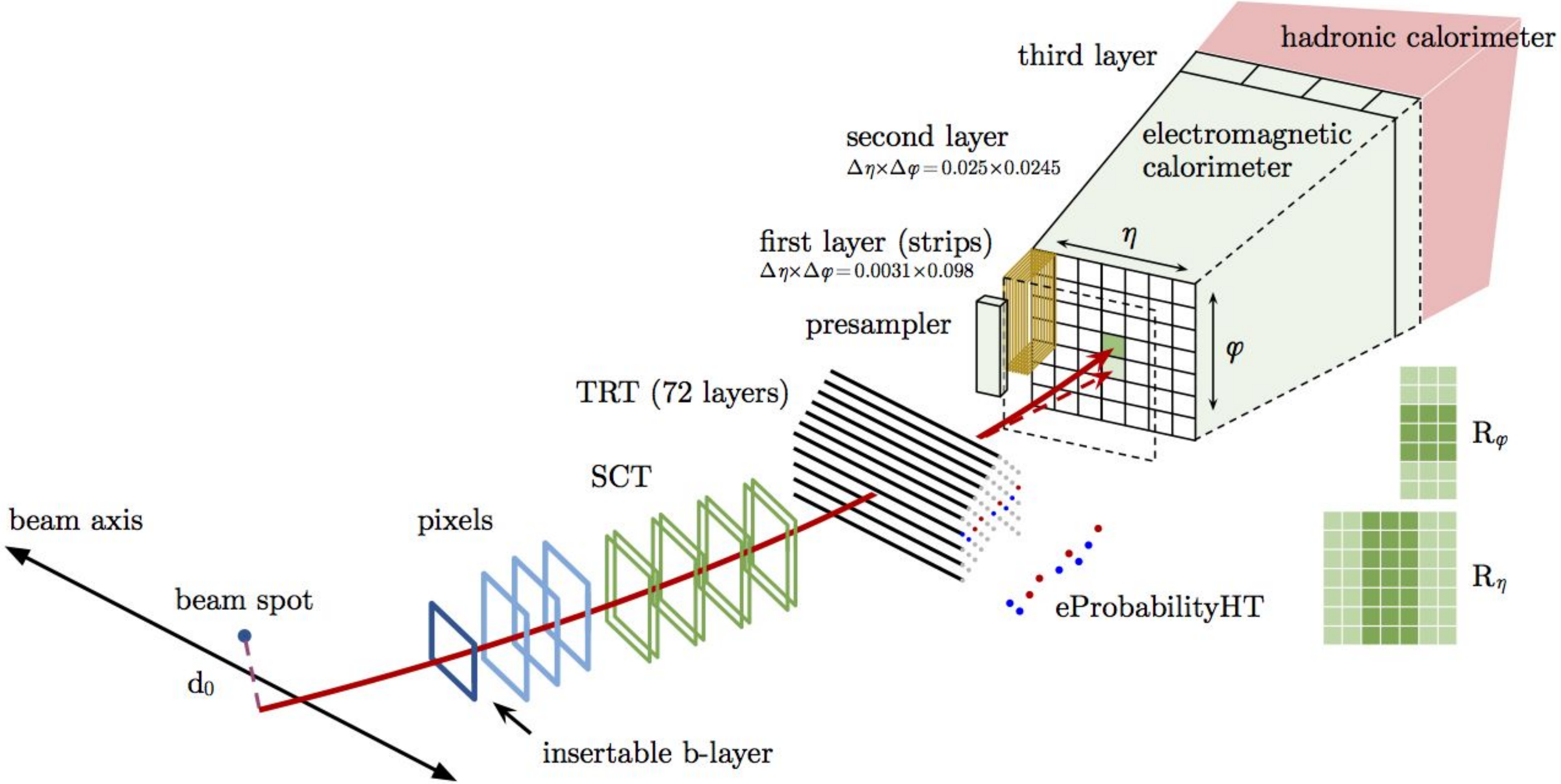
[phys.org]

Particle Interactions with Detectors



ATL-PHYS-PUB-2020-018

Particle Detectors



https://www.hep.shef.ac.uk/

Calorimeter

1603.02934

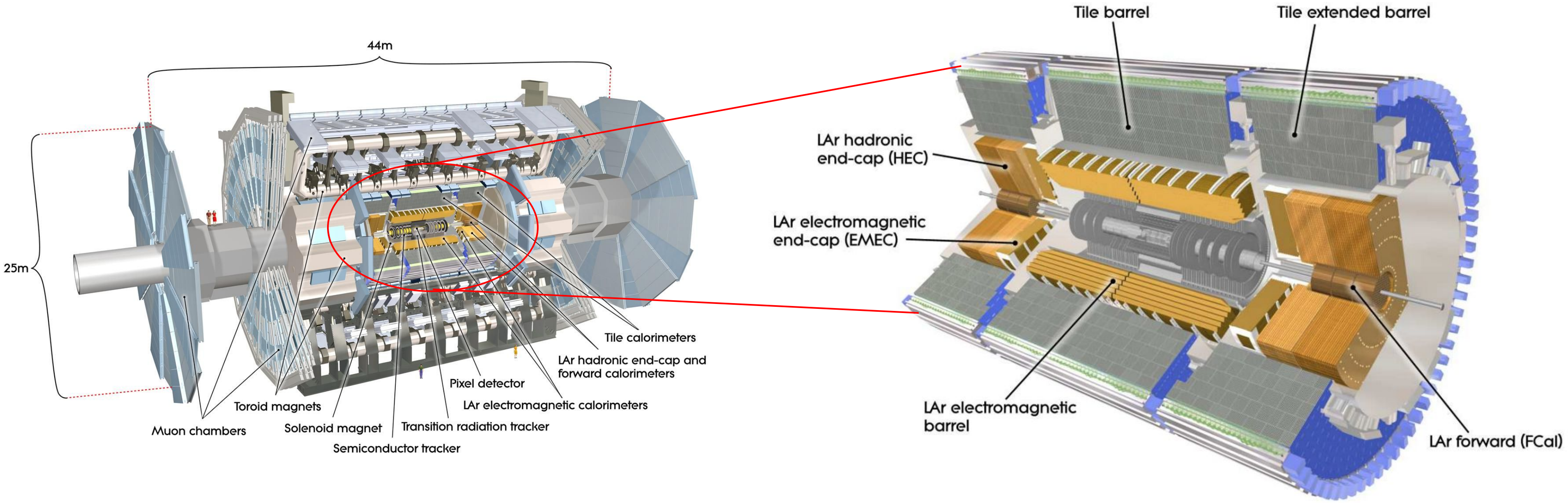
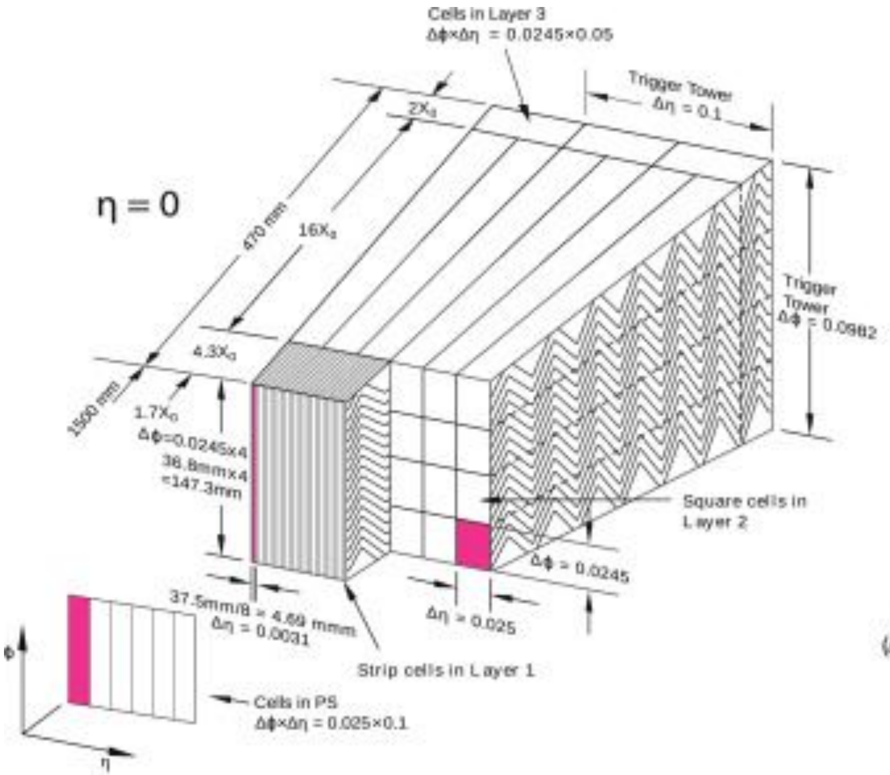
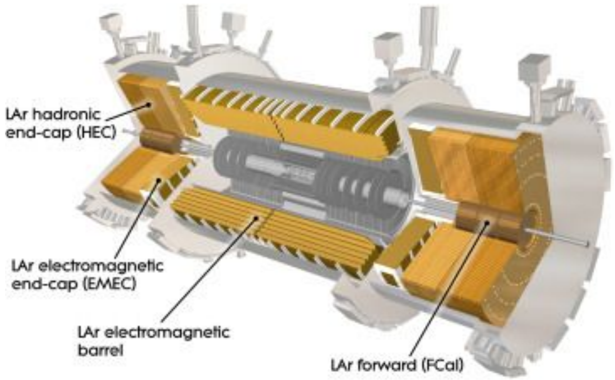


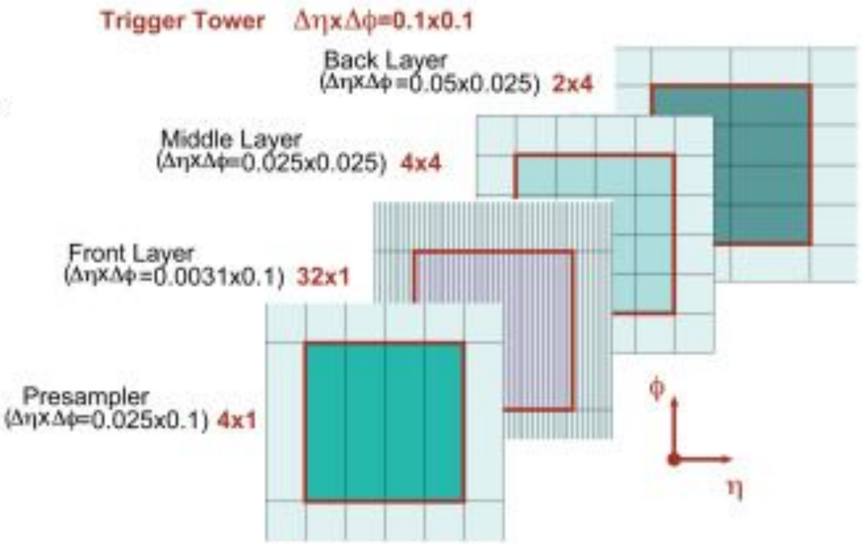
Figure 1: Cutaway view on the ATLAS calorimeter system.

EM Calorimeter

1603.02934

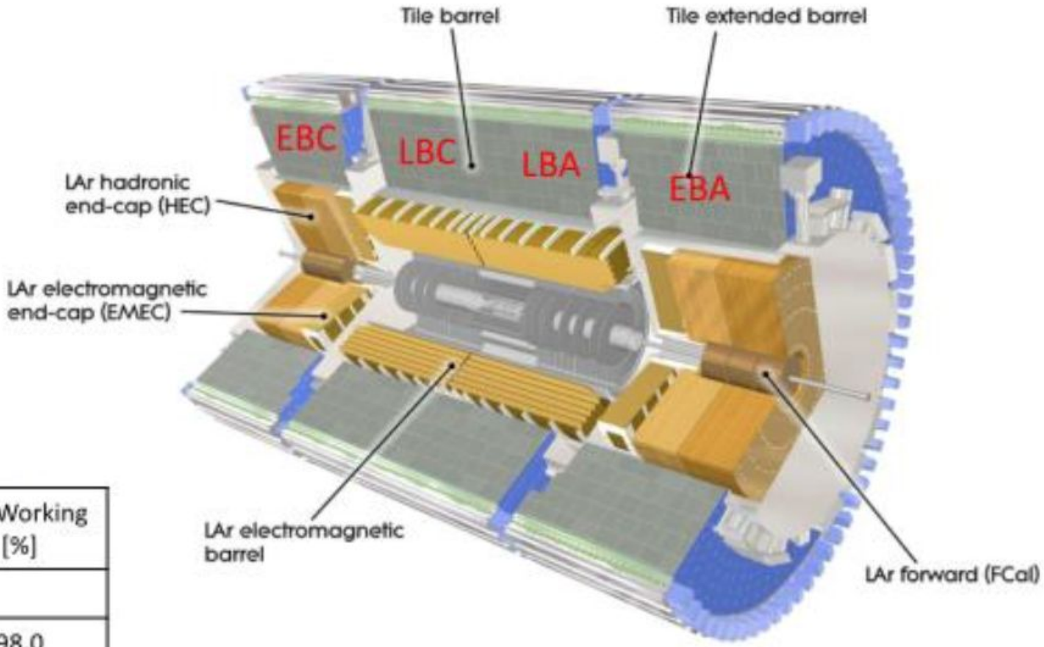






Hadronic Calorimete

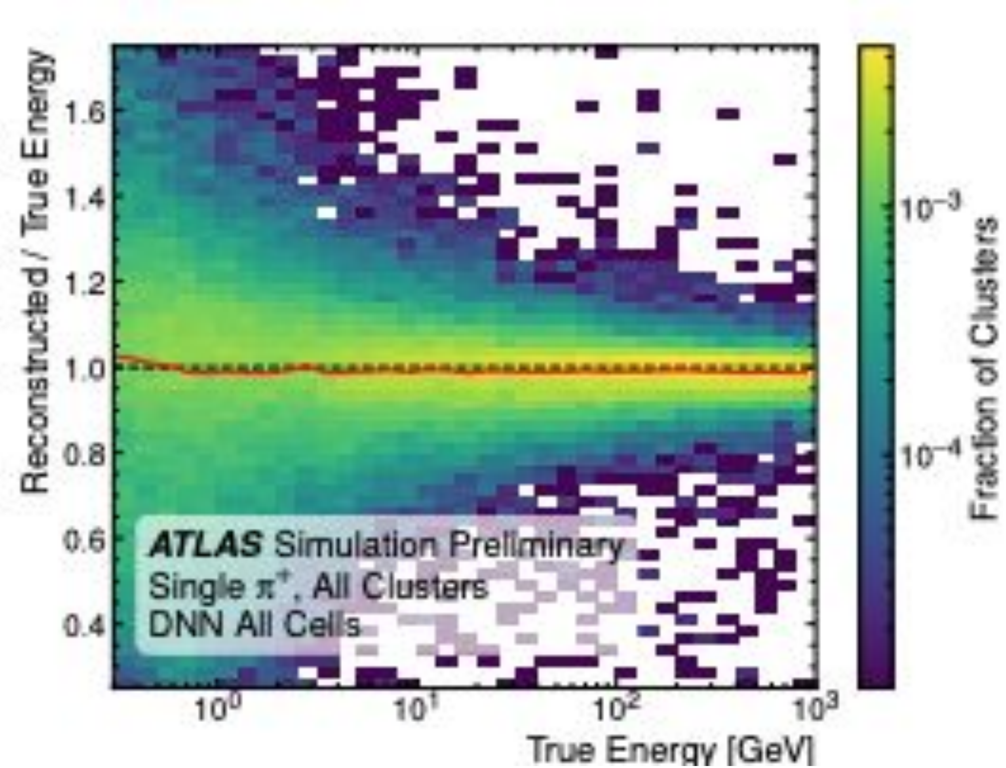
	Absorber	Active medium	Coverage	Readout Channel	Working [%]
EM					
EB, EMEC	Pb	LAr	$\lvert\eta\rvert<3.2$	173k	98.0
Hadronic					
HEC	Cu	LAr	$1.5<\lvert\eta\rvert<1.8$	5.6k	99.9
Fcal	Cu/W	Lar	$3.1<\lvert\eta\rvert<4.9$	3.5k	100
TileCal	Steel	Scintillator	$\lvert\eta\rvert<1.7$	10k	95.6



1603.02934

Energy Regression

ATL-PHYS-PUB-2020-018

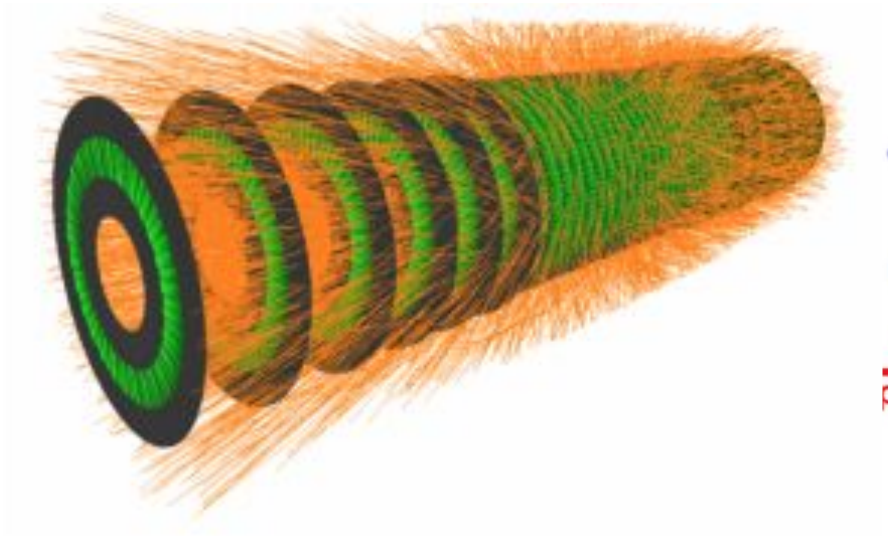


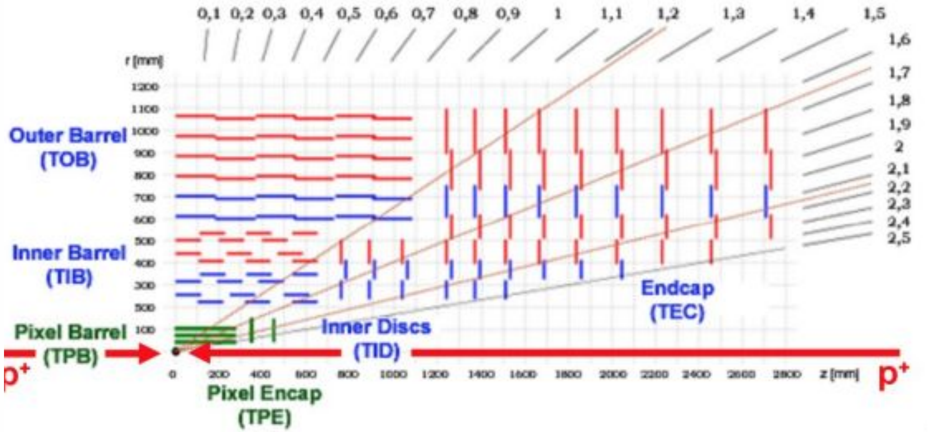
Data Structure

- **Kaggle:** https://www.kaggle.com/t/ca1834422f9e42b69b0940f4a20735cc
- **Features:**
  - **Energy, eta, phi of each detector cell**
- **References:**
  - ATL-PHYS-PUB-2020-018
  - https://www.hep.shef.ac.uk/

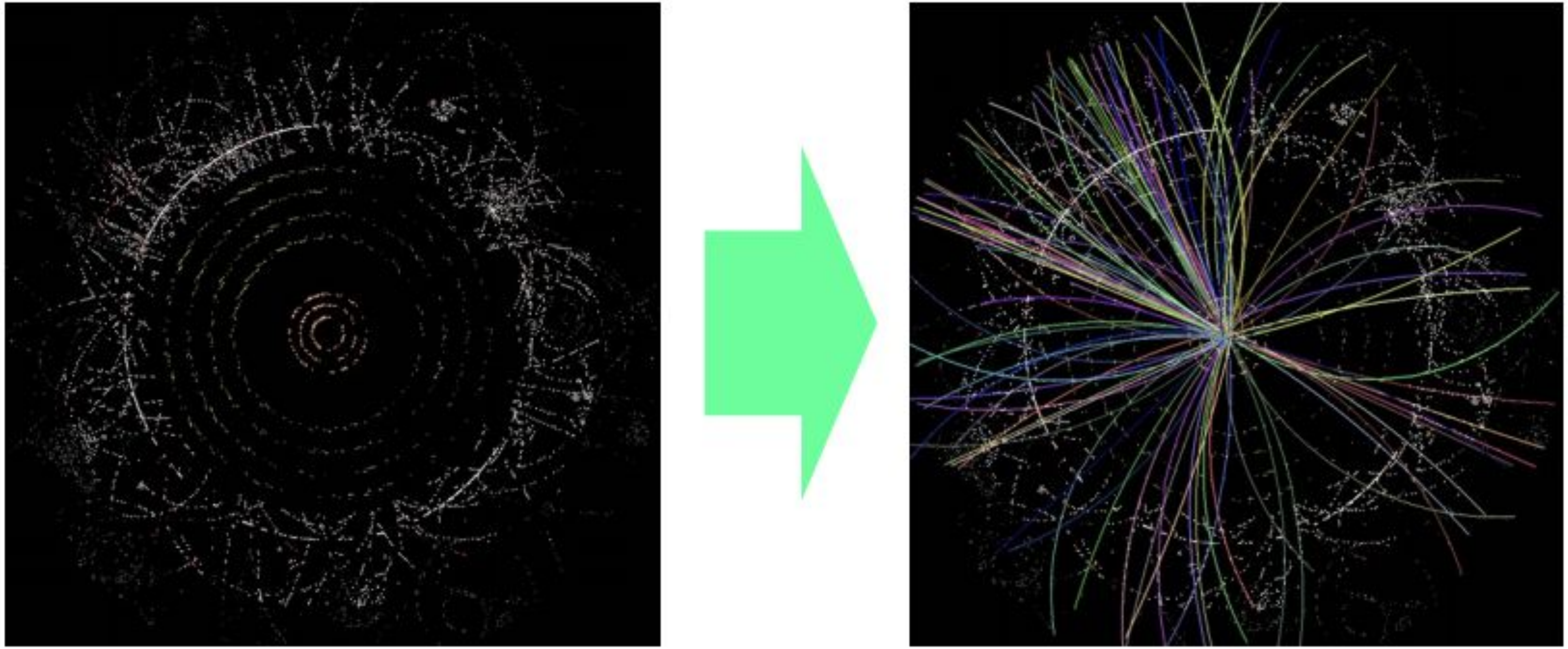
Charged Particle Tracking

Charged Particle Trajectories





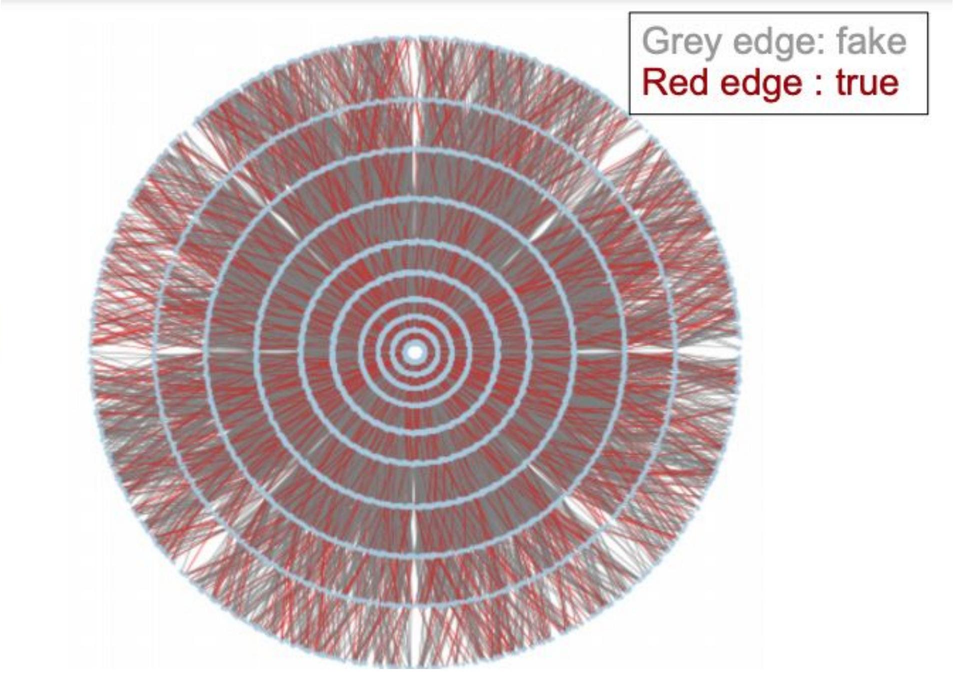
From hits to tracks



From hits ...

... to trajectory & parameters

Edge Classification



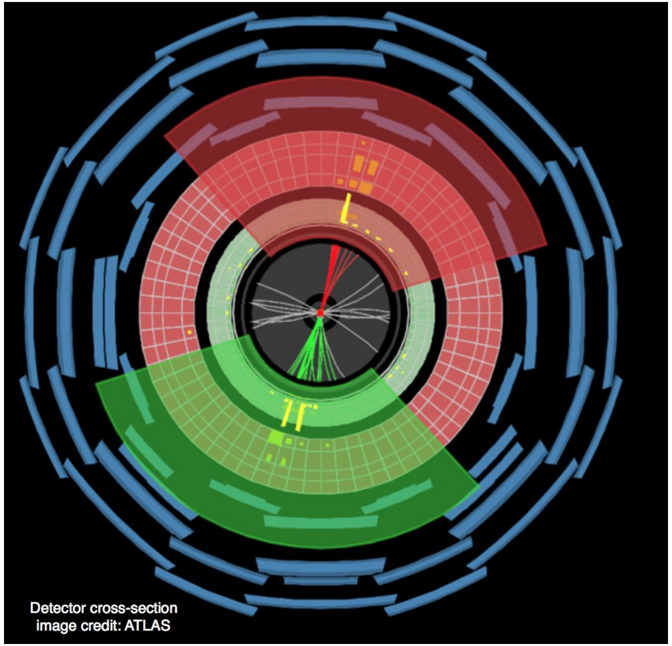
Data Structure

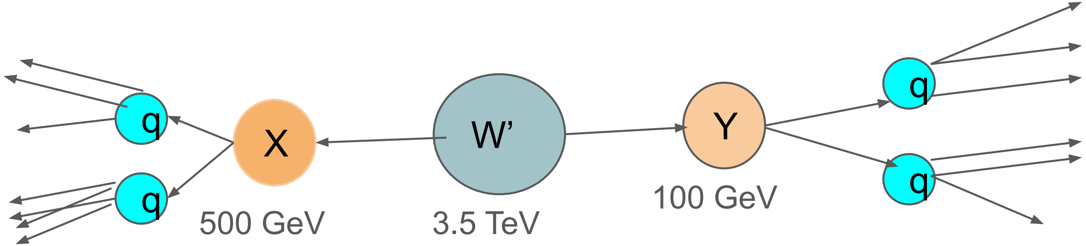
- **Kaggle:** https://www.kaggle.com/t/f43836bbb1e640a4b2fd3f058b8139fd
- References:
    - J. Shlomi, P. Battaglia, J.-R. Vlimant, “Graph Neural Network in Particle Physics” https://arxiv.org/abs/2007.13681
    - Battaglia, et. al. “Relational inductive biases, deep learning, and graph networks” https://arxiv.org/pdf/1806.01261.pdf

Dijet Generative Model

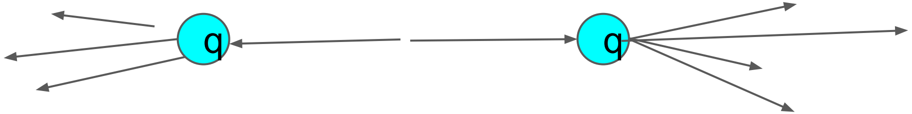
Dijet Events

[LHCO2020]





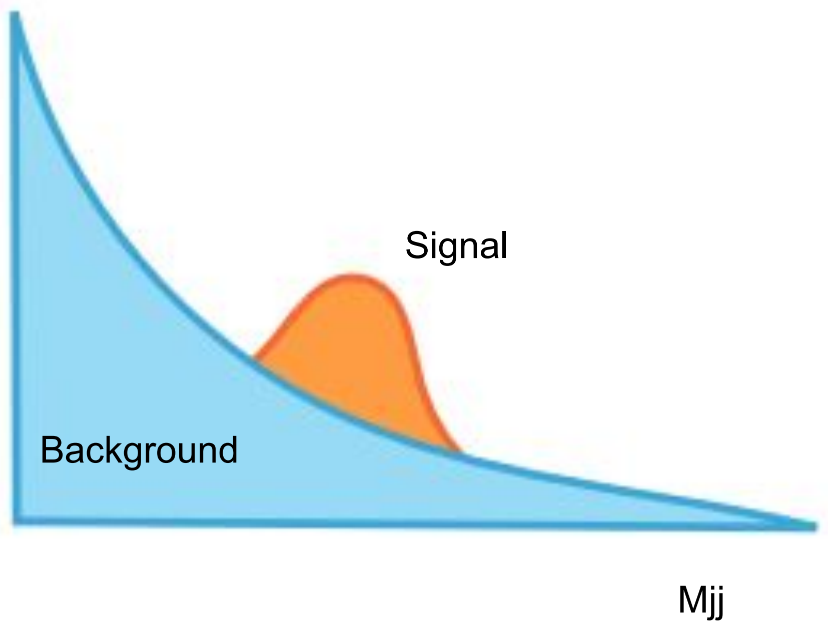
Beyond Standard Model (BSM) Signal



QCD (Quantum Chromodynamics) Background

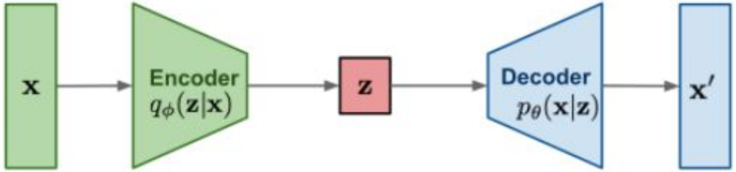
Anti-kT fat-jet R = 1,
pT > 1.2 TeV and |η| < 2.5

Event Observables

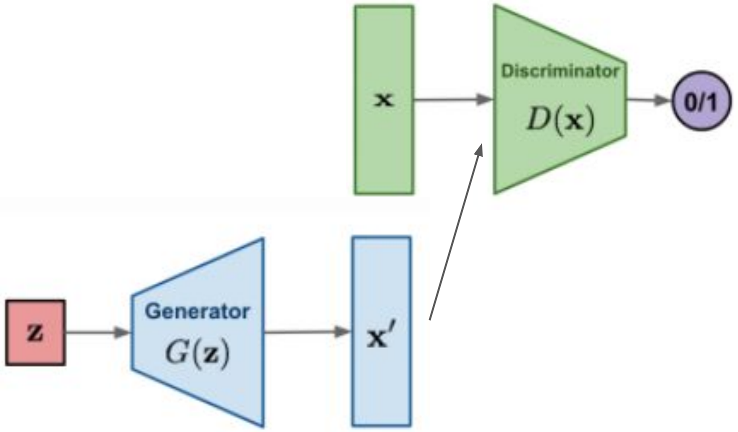


Generative Models

Autoencoders (AE) and
Variational Autoencoders (VAEs)



Generative Adversarial Network (GAN)



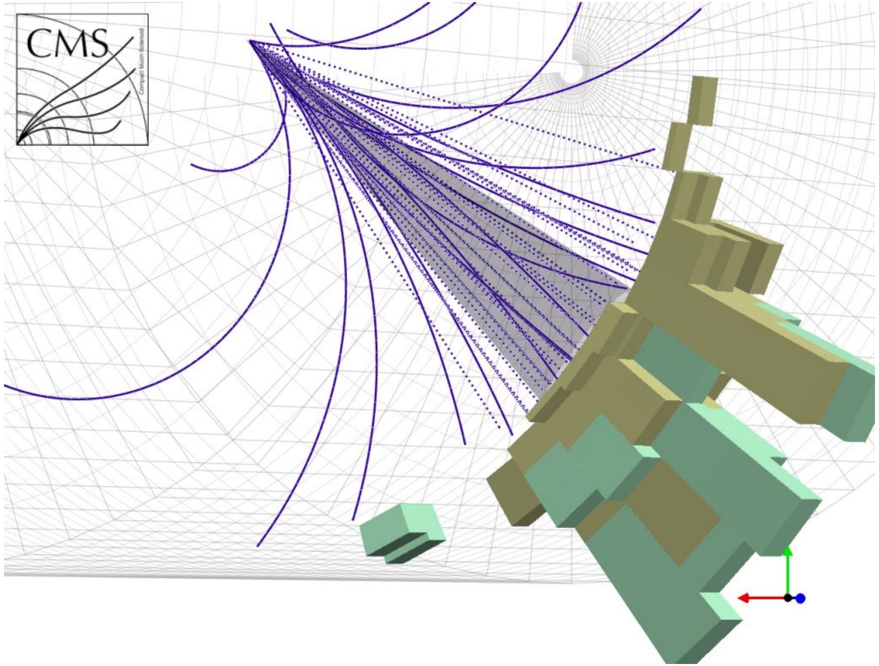
Data Structure

- **Kaggle:** https://www.kaggle.com/t/215a5eedab064c8a9fa5ae68ebfe3c28
- **Data Structure:**
  - It consists of 1M QCD dijet events and 100k W'->XY events, with X->qq and Y->qq. The W', X, and Y masses are 3.5 TeV, 500 GeV and 100 GeV respectively. The events are produced using Pythia8 and Delphes 3.4.1, with no pileup or MPI included. They are selected using a single fat-jet (R=1) trigger with pT threshold of 1.2 TeV.
  - Each dataset contains kinematic variables of each jet (pxj1 ', 'pyj1', 'pzj1', 'mj1') and jet shape variables (e.g. 'tau1j1', 'tau2j1', 'tau3j1').
- **References:**
  - VAE s://arxiv.org/abs/2101.08944
  - DijetGAN arXiv:1903.02433 (code)

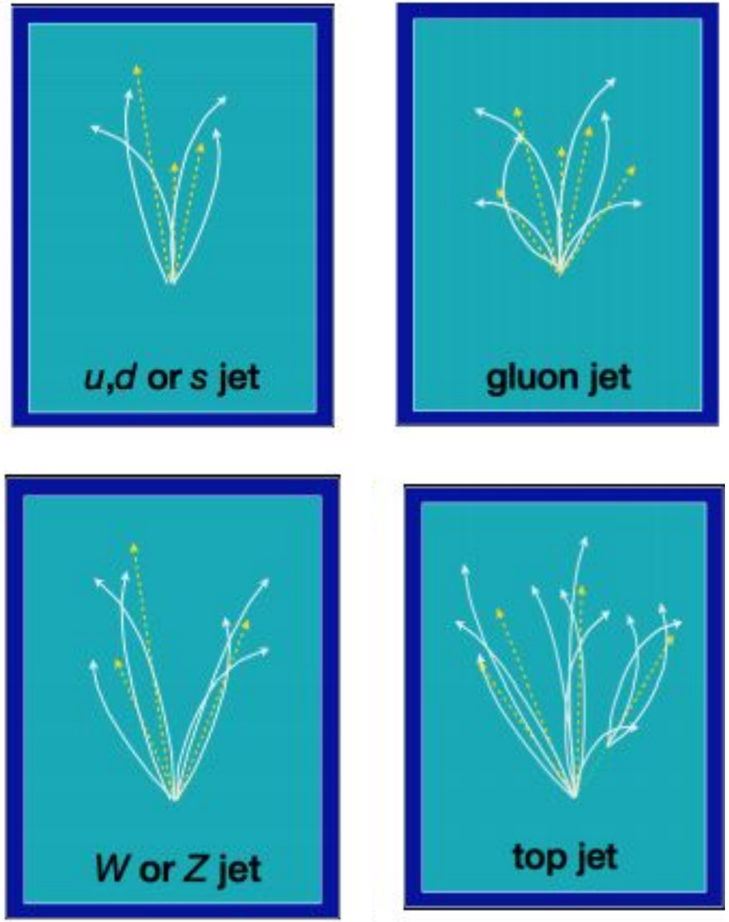
Top Jets

Jet Classification

A jet is composed of many constituents from trackers or calorimeters.



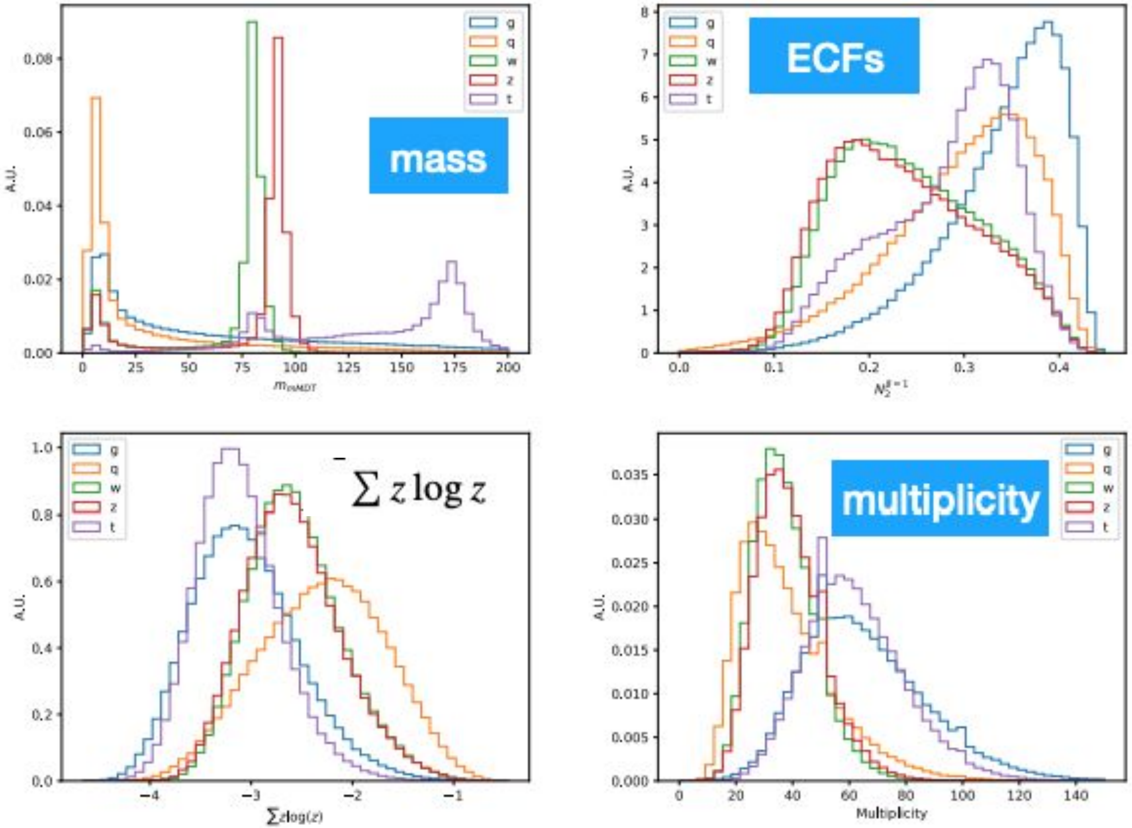
https://www.quantumdiaries.org/tag/jets/



https://indico.cern.ch/event/817013/

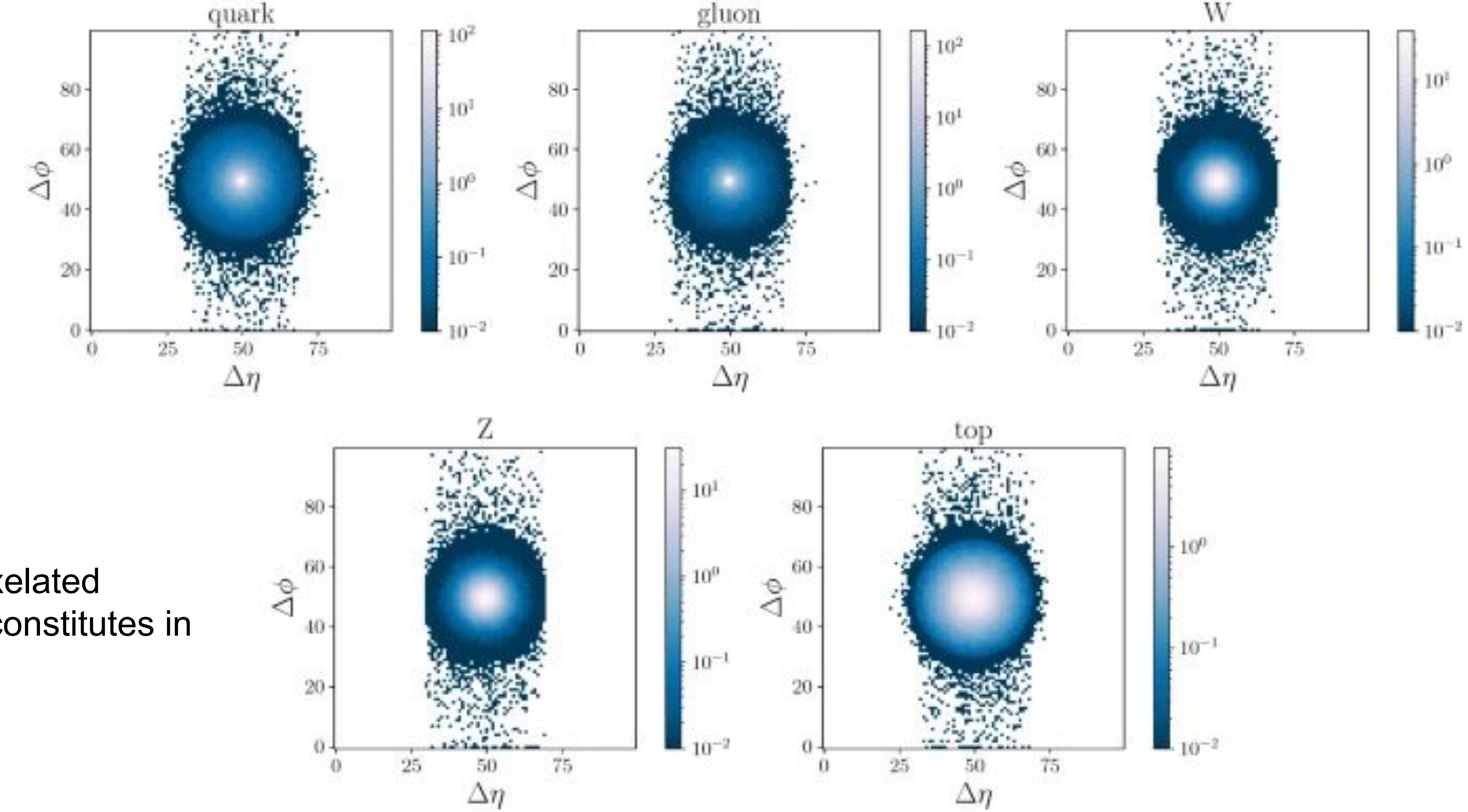
High Level Features

Observables calculated from 4-vector of jet constituents



arxiv:1908.05318

Jet Visions

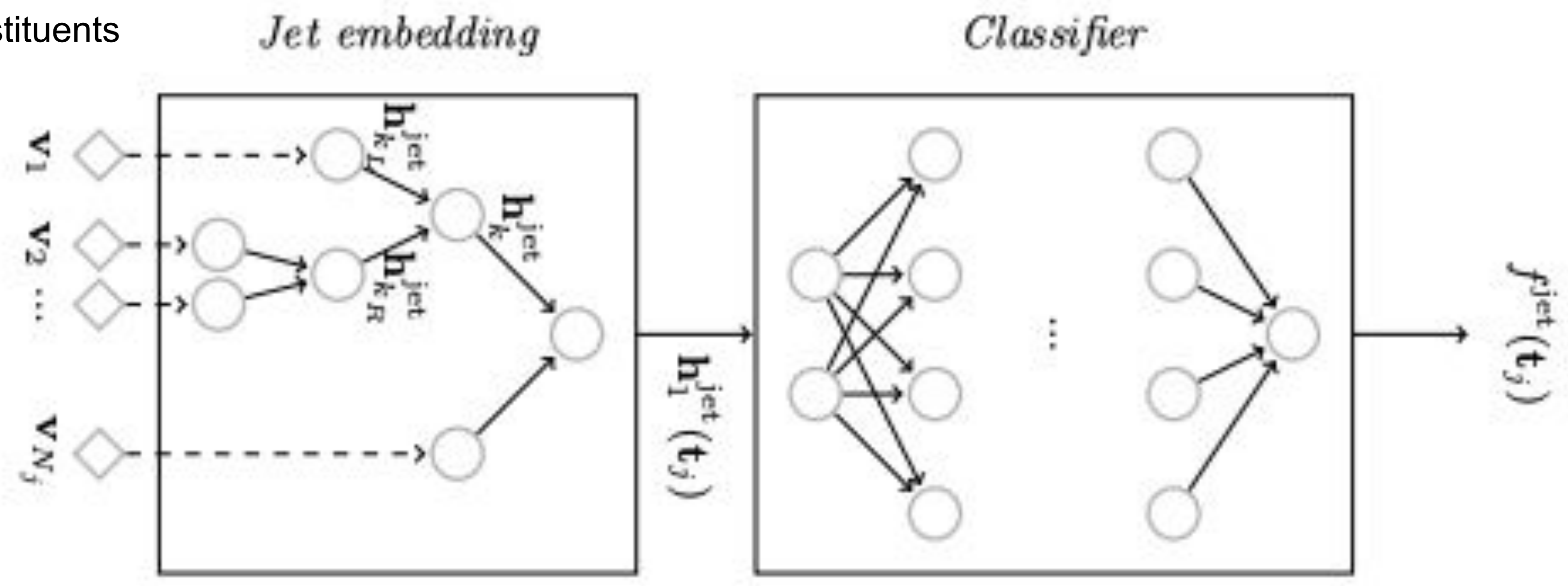


jet images = pixelated versions of jet constitutes in 2D (η, Φ)

Recursive Jet Embedding

arxiv:1702.00748

Jet constituents as input



Data Structure

- **Kaggle:** https://www.kaggle.com/t/edb7826e3b2a422e8f5ff9d0e6df6c48
- **References:**
  - https://arxiv.org/abs/1902.09914
  - https://arxiv.org/abs/1908.05318
  - https://indico.cern.ch/event/817013/

Electron Showers with Emulsion Detector

Credit: John Spencer

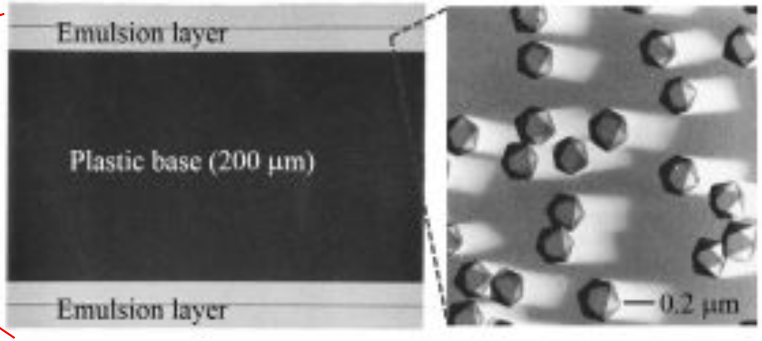
arxiv:/2001.03073.pdf

Emulsion @FASER

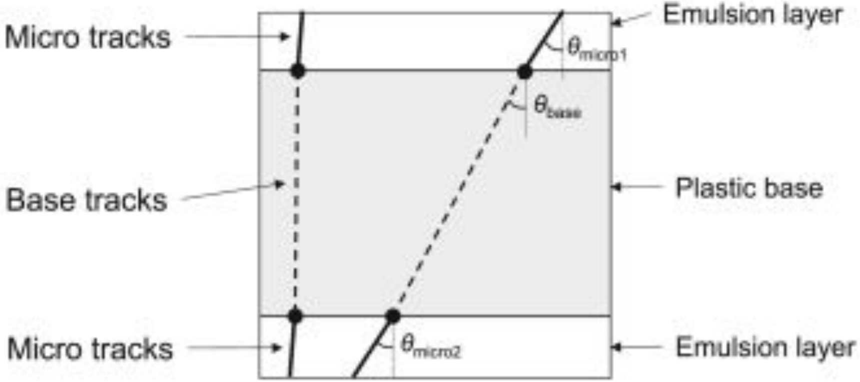
- 770 layers of an emulsion film and 1mm tungsten plate
- 25 cm×30 cm×1.1 m
- 1.1 tons, 220 X0

Emulsion Detector Box

Yosuke ⇒↑

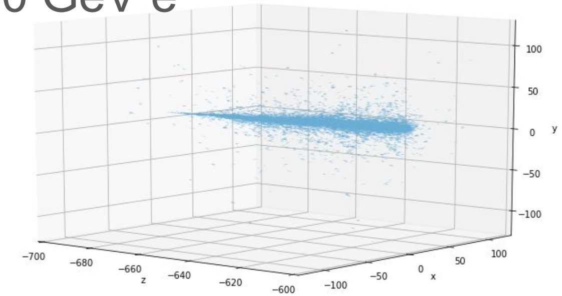


Each emulsion film is composed of two emulsion layers and one plastic base.

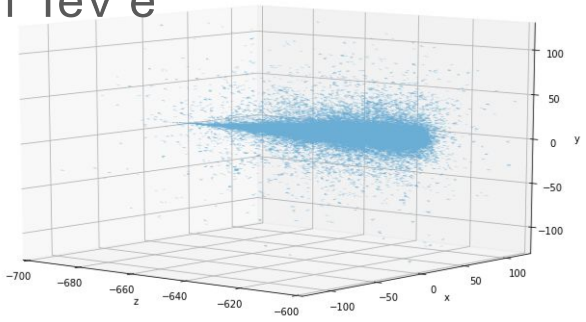


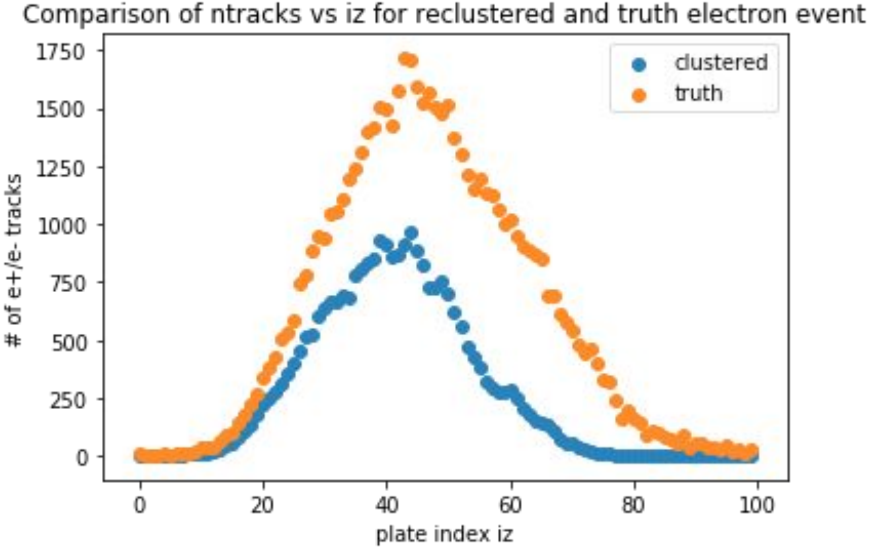
Electron shower

100 GeV e$^-$



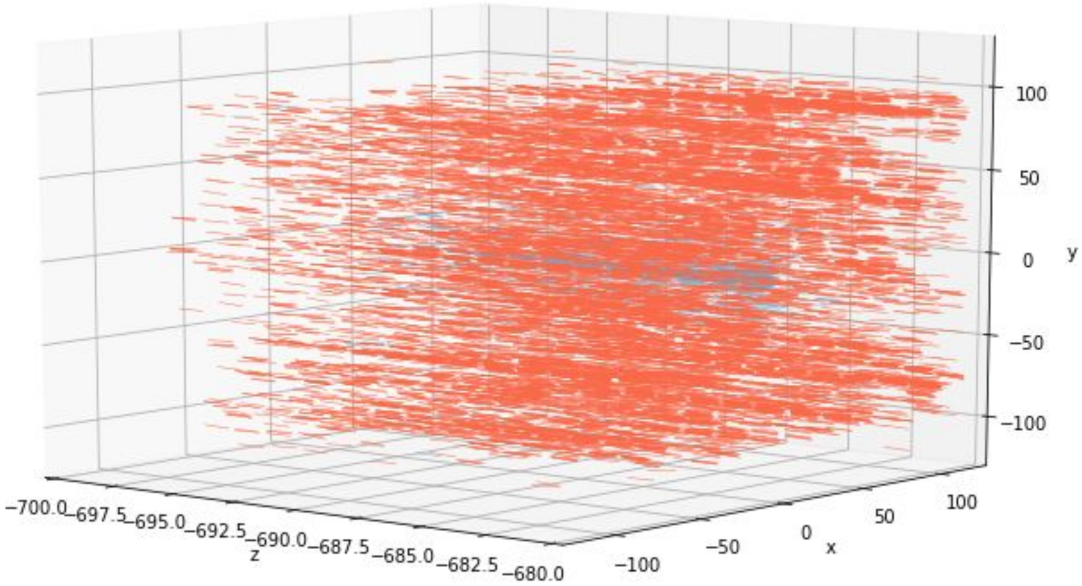
1 TeV e$^-$



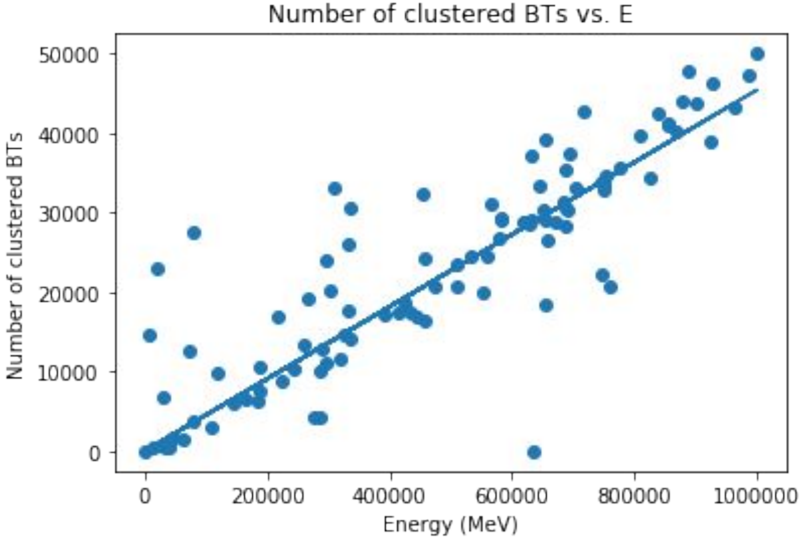


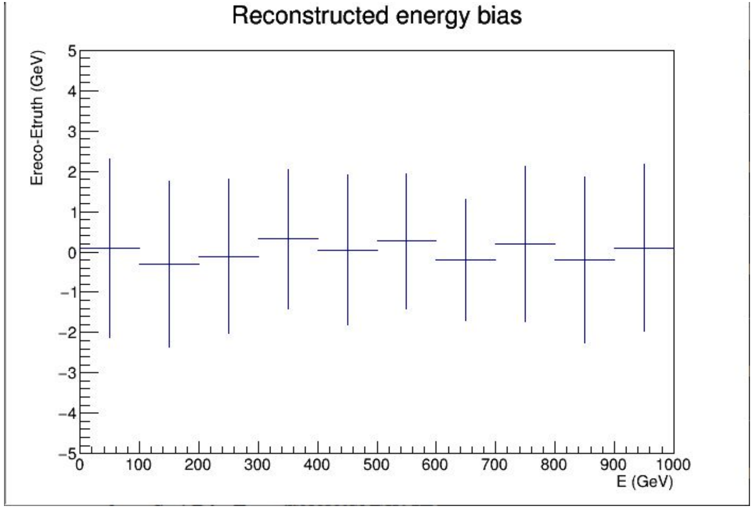
Event display (pileup)

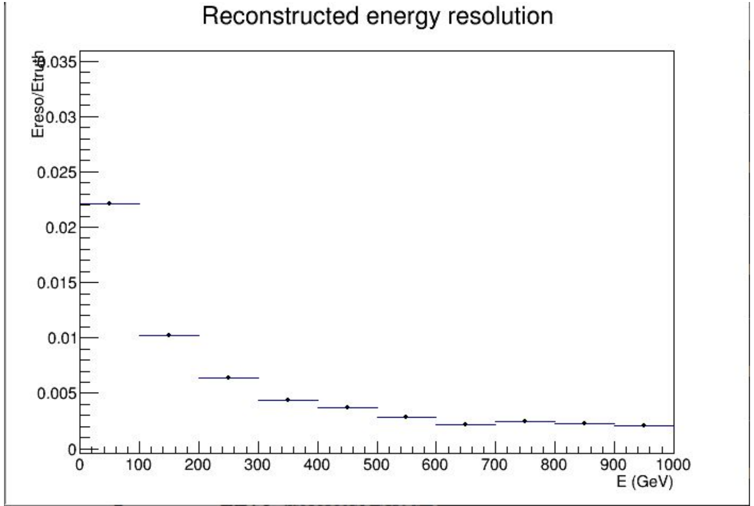
Shown: 1 TeV e^- (blue), 25k muons (red) in first 20mm of pilot detector



Energy Regression







Data Structure

- **Kaggle:** https://www.kaggle.com/t/016e01fc19fb4c90865ddae0f3bed6ee
- **Features:**
  - **Total number of base tracks**
  - **Shower maximum depth**
  - **Number of base tracks at shower maximum**
  - **Number of base tracks per layer**
- **References:**
  - **Matteo Tenti thesis (2012) [URL]**
  - **Fr´ed´eric Juget Calor 2008 proceeding [URL]**
  - **https://faser.web.cern.ch/physics/presentations**

Data Exploration

Data Exploration

1. Data structure, e.g. data shape, label, ...

Data Exploration

1. Data structure, e.g. data shape, label, ...
2. Data visualization, e.g. histogram, correlation, ...

Data Exploration

1. Data structure: e.g. data shape, label, ...
2. Data visualization: e.g. histogram, correlation, ...
3. Discussion: What can we do with this data, e.g. preprocessing, classification, regression, ...

-- Create a notebook for each project and share with instructors and TA's.

-- Answer the three questions above with clearly-labeled sections ("Data structure, Data visualization, Discussion").

Bonus: sharing your answer starting at 12PM