Nicolò Cibrario

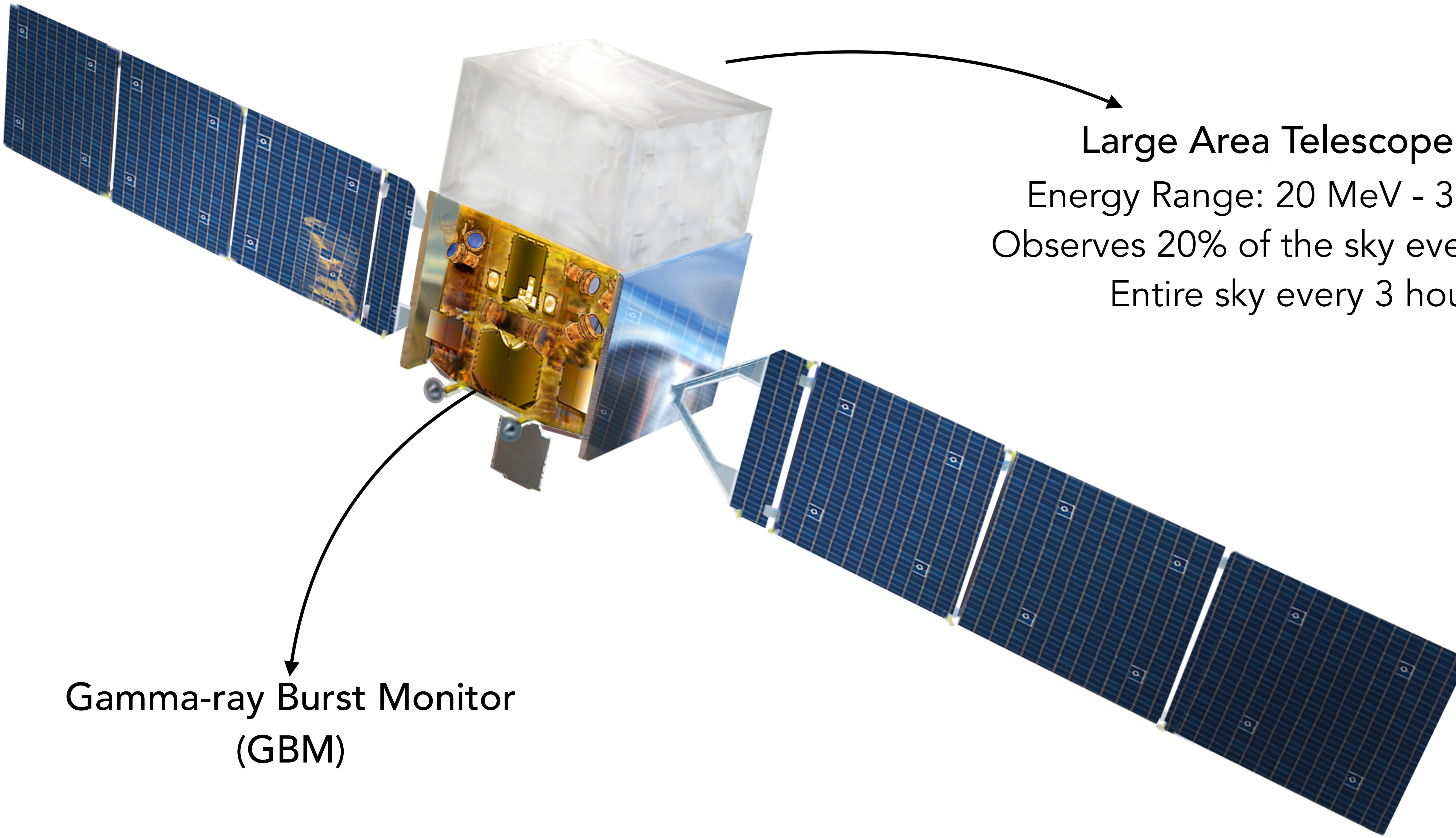
Machine Learning for the measurement of the Cosmic-Ray Electron Spectrum with Fermi Large Area Telescope

UNIVERSITÀ DEGLI STUDI DI TORINO

INFN Istituto Nazionale di Fisica Nucleare

Fermi Gamma-Ray Space Telescope

Launch Date: 11 June 2008



Large Area Telescope (LAT)

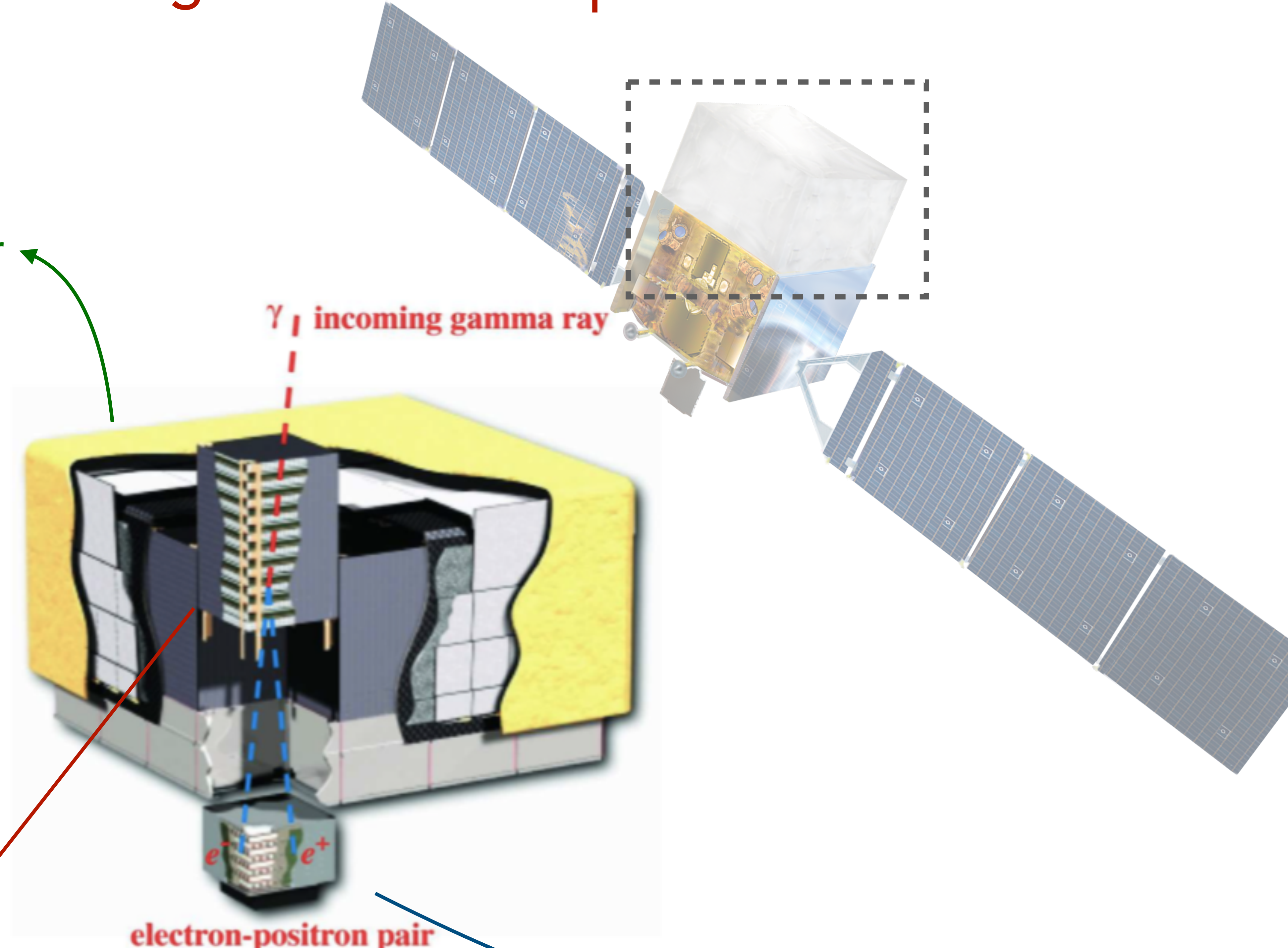
Energy Range: 20 MeV - 300 GeV

Observes 20% of the sky every instant

Entire sky every 3 hours

Gamma-ray Burst Monitor (GBM)

Large Area Telescope



Anticoincidence Detector

Plastic scintillators

Segmented tiles

Tracker-Converter

18 Silicon planes for tracking

16 Tungsten planes for converting

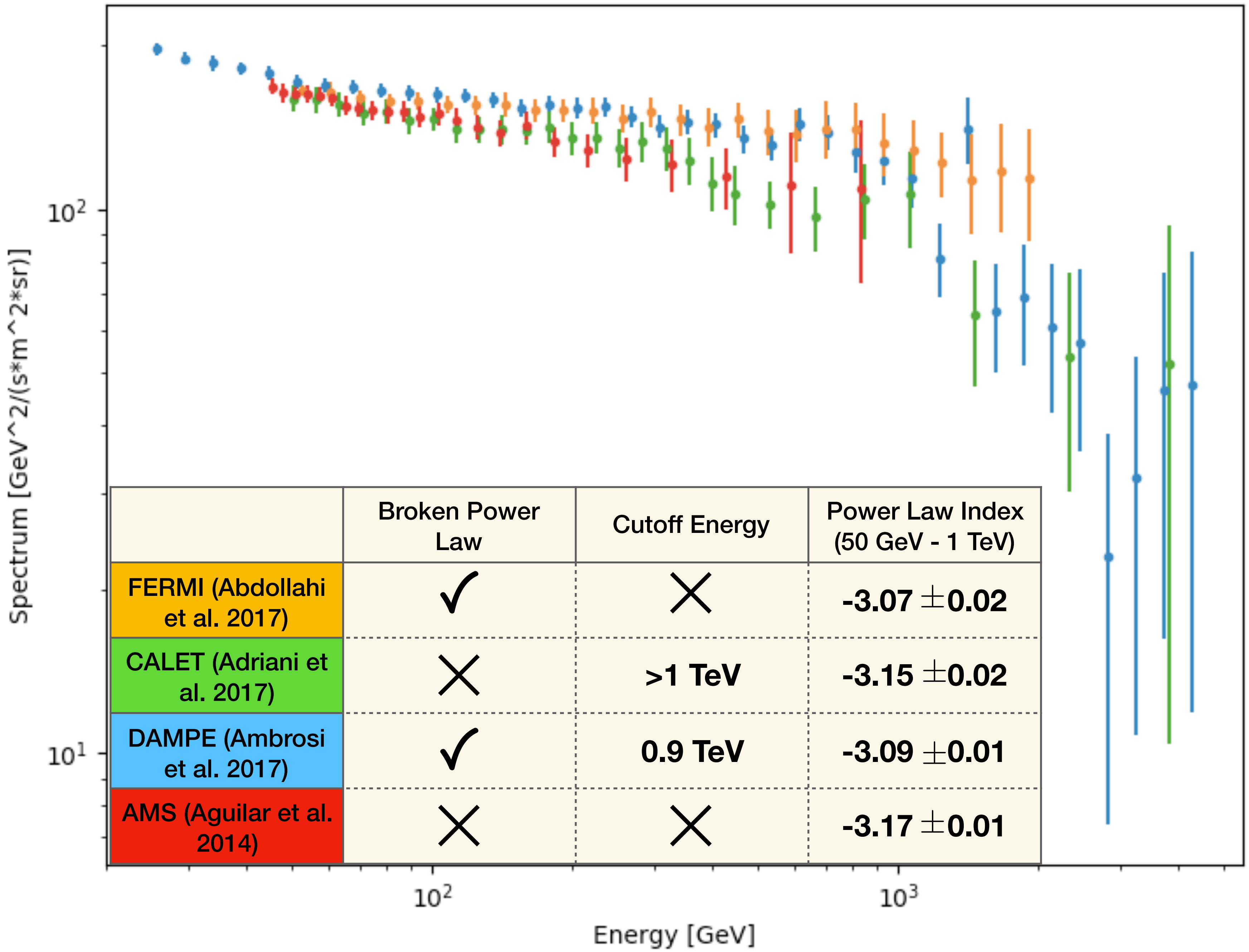
Calorimeter

96 CsI(TI) Crystals

Shower development imaging

Energy deposition

Latest CRE Spectra



	Broken Power Law	Cutoff Energy	Power Law Index (50 GeV - 1 TeV)
FERMI (Abdollahi et al. 2017)	✓	✗	**-3.07 $\pm$0.02**
CALET (Adriani et al. 2017)	✗	**>1 TeV**	**-3.15 $\pm$0.02**
DAMPE (Ambrosi et al. 2017)	✓	**0.9 TeV**	**-3.09 $\pm$0.01**
AMS (Aguilar et al. 2014)	✗	✗	**-3.17 $\pm$0.01**

Systematic uncertainties:

Background Contamination

Data - Monte Carlo simulations agreement

Reconstruction of the energy

Absolute scale of energy

Aim of the work

ML Technique for 2017 analysis: *Boosted Decision Trees (+MC/data correction)*

ML Techniques for this work: *Neural Networks & Unsupervised Learning*

Possible advantages are:

Neural Networks:

They could find different patterns and correlations in data compared to the BDTs: better separation between electrons and protons?

Unsupervised Learning:

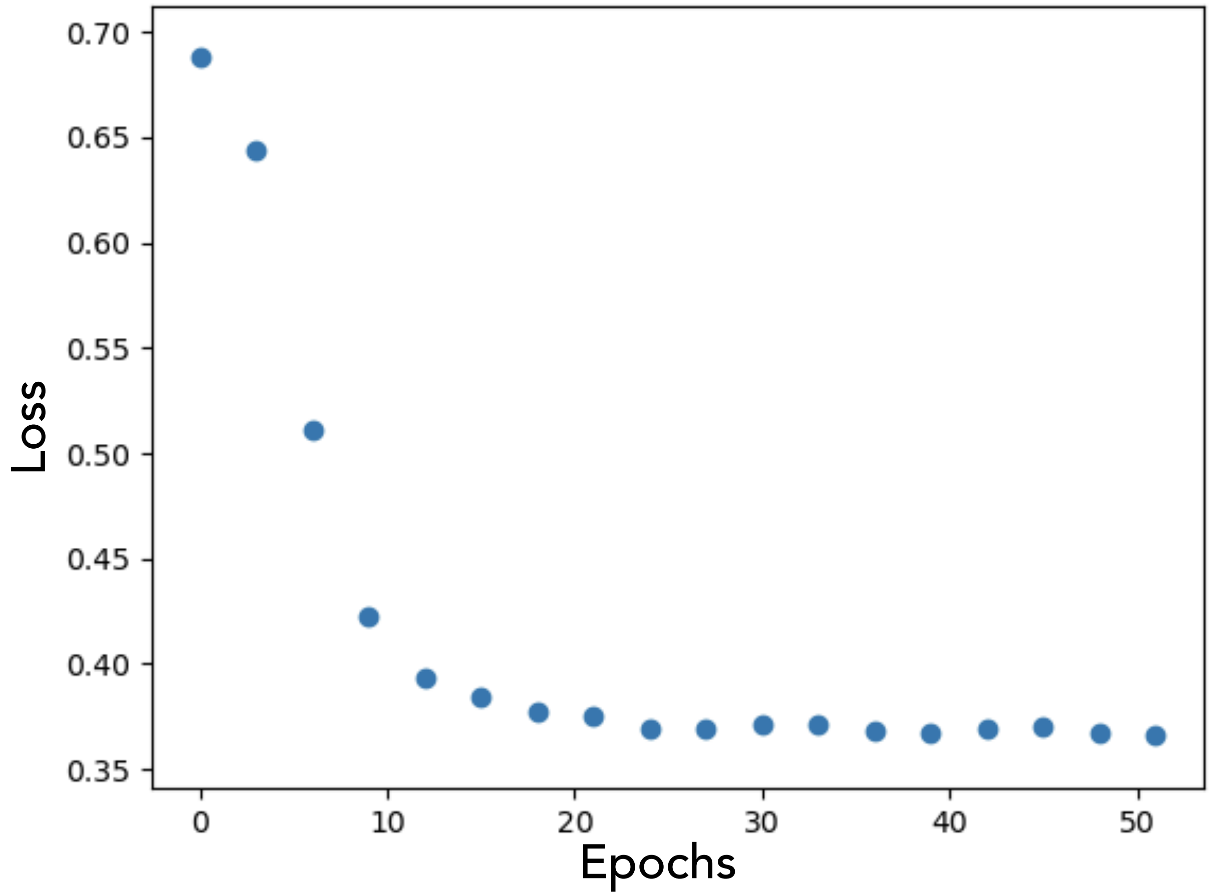
It would allow to avoid uncertainties linked to the MC simulations

1st Part

Neural Networks

Neural Networks: Overview

Forward Pass



Backward Pass

Key words:

Layers: *Made up of neurons*

Activation function: *Activation of the single neuron*

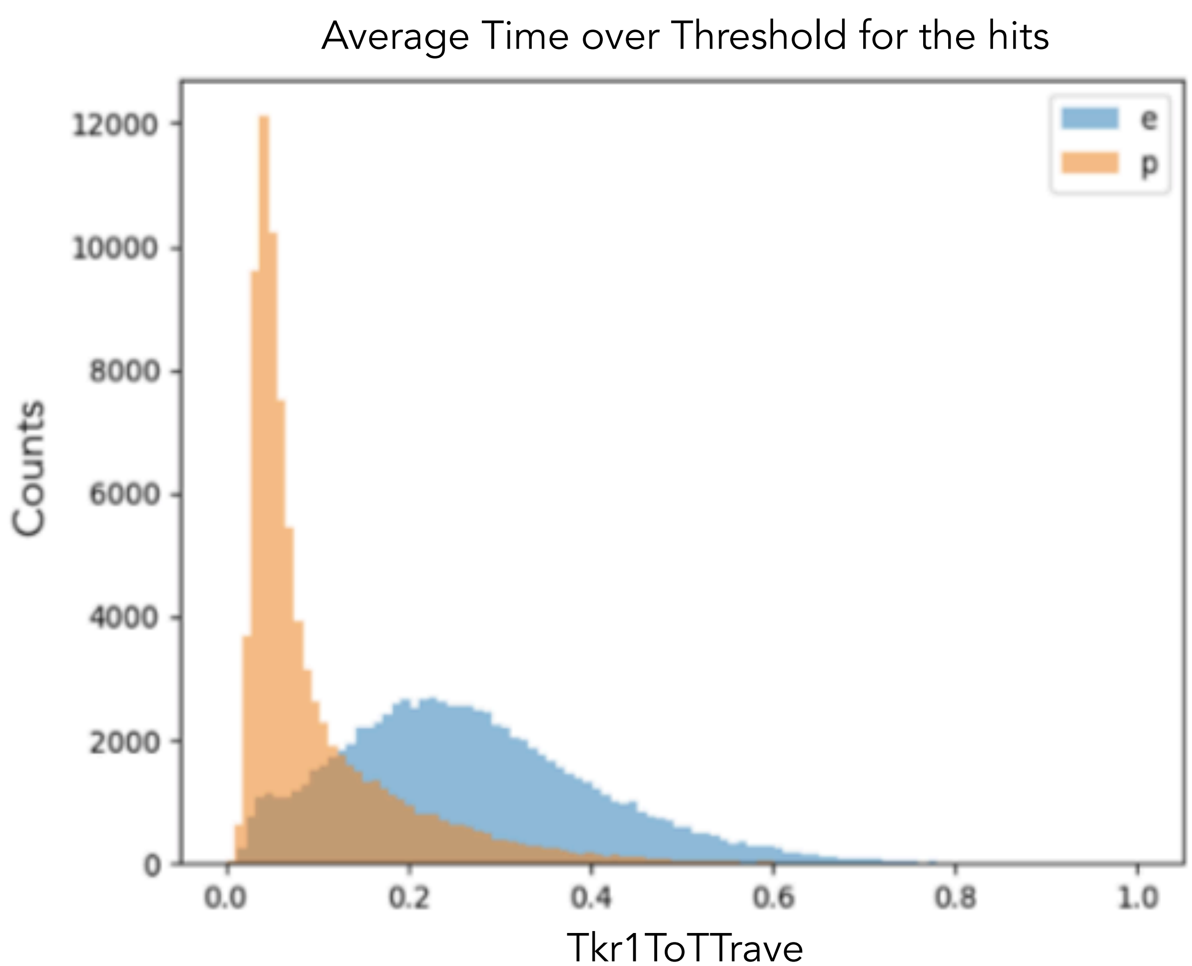
Loss: *Error between the output and the expected value*

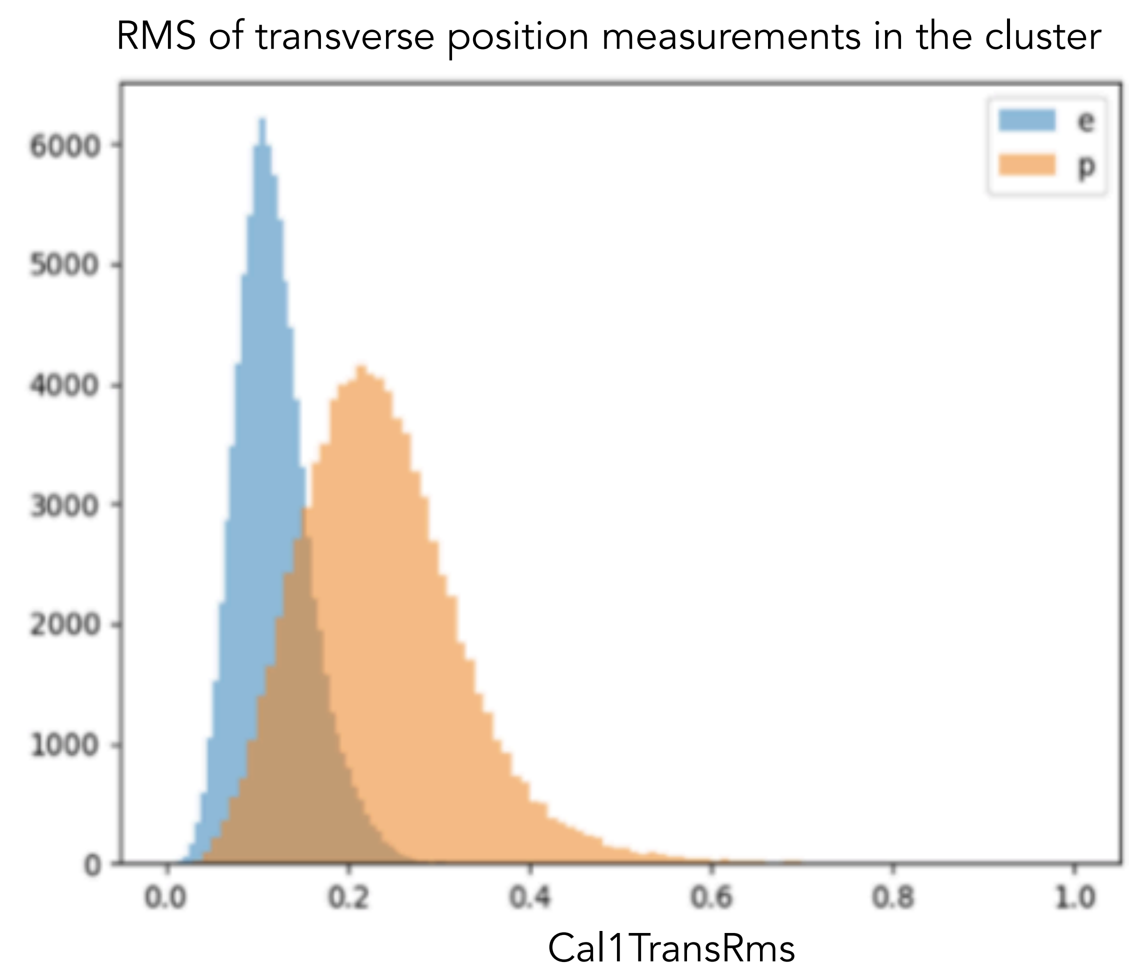
Optimizer: *Tool used to modify the weights*

Variables Selection

Preliminary cuts ⟶ Only protons and electrons in the dataset

A subset of variables is needed: highlight the differences between electrons and protons events.



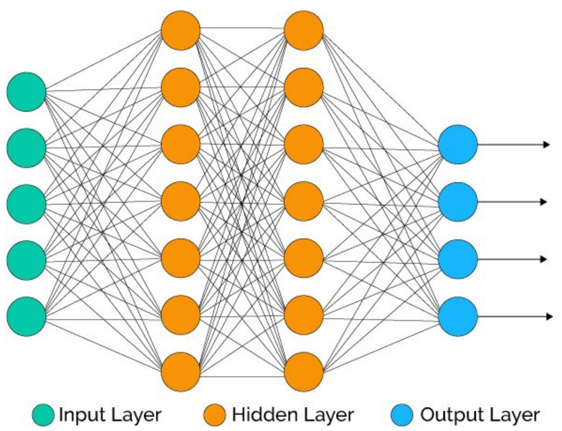


First attempt: variables used in 2017 analysis

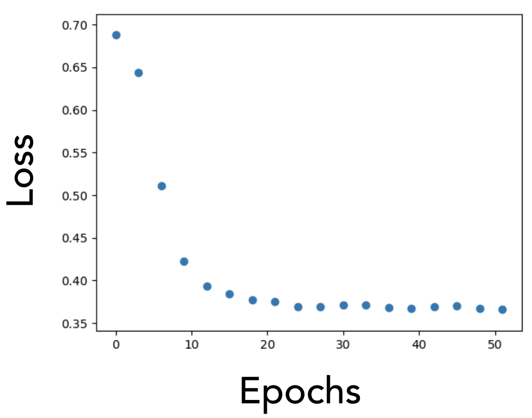
MC simulations datasets: training and testing the network

MC simulations were divided into a *Trainset (50%)*, an *Evalset (25%)*, and a *Testset (25%)*.

Trainset — Forward Pass → / ← Backward Pass ⟶ Used to train the network



Evalset — Forward Pass → ⟶ Used to evaluate the loss



Testset — Forward Pass → ⟶ Used to test the performances of the model

Features tuning

MC data were preprocessed before feeding the network.

Energy bin selection

40 energy bins between 20 and 4034 GeV.

↓

Arrival direction bin selection

Only one bin, $0.5<\cos\theta<1$, where θ is the angle between the event direction and the LAT boresight.

↓

Normalization

All variables are scaled in order to be ranged between 0 and 1.

How to avoid over-fitting?

Two techniques are used in the algorithms:

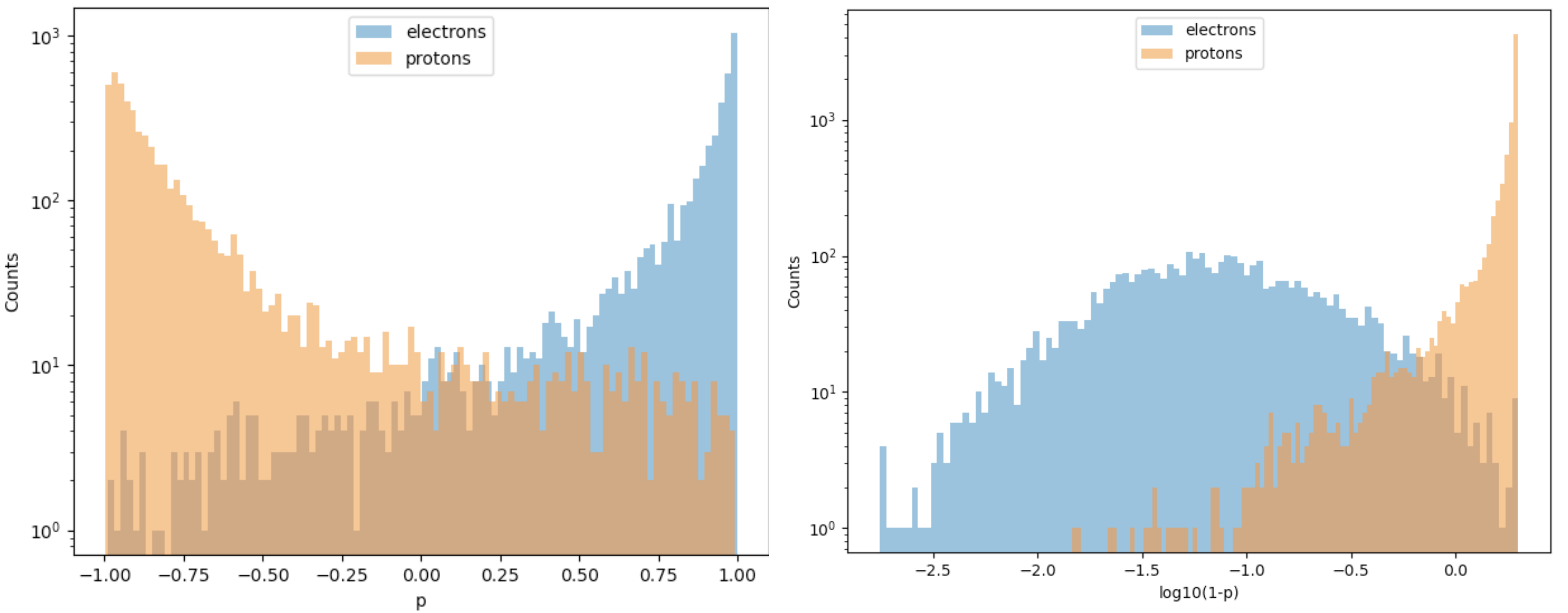
Dropout & Early Stopping

Output of the network

The output of the network is a value of p, ranging from -1 (indicating protons-like events) to 1 (indicating electrons-like events).

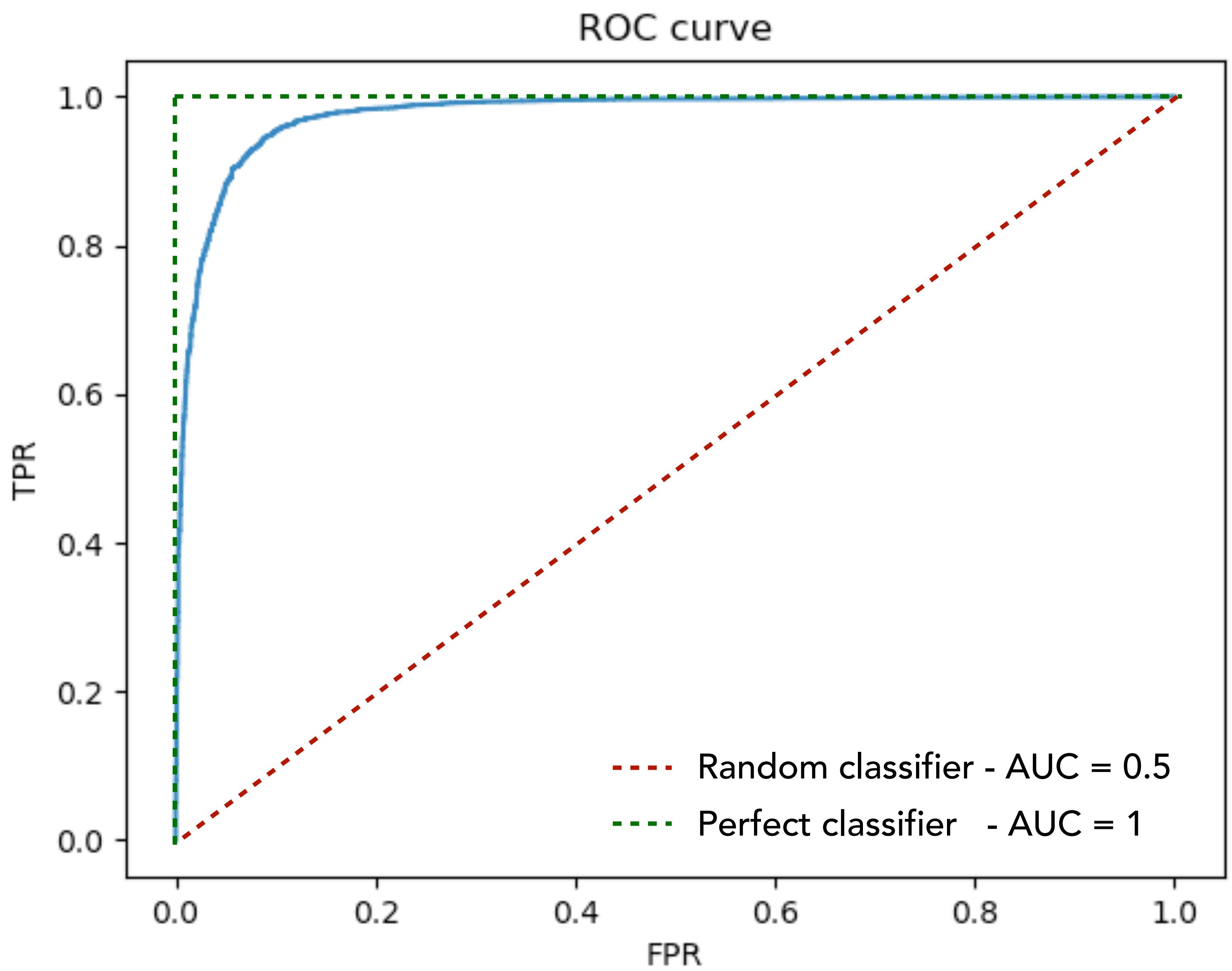
Log(1-p) is calculated, to highlight the region where electrons and protons overlap.

p ⟶ Log(1-p)

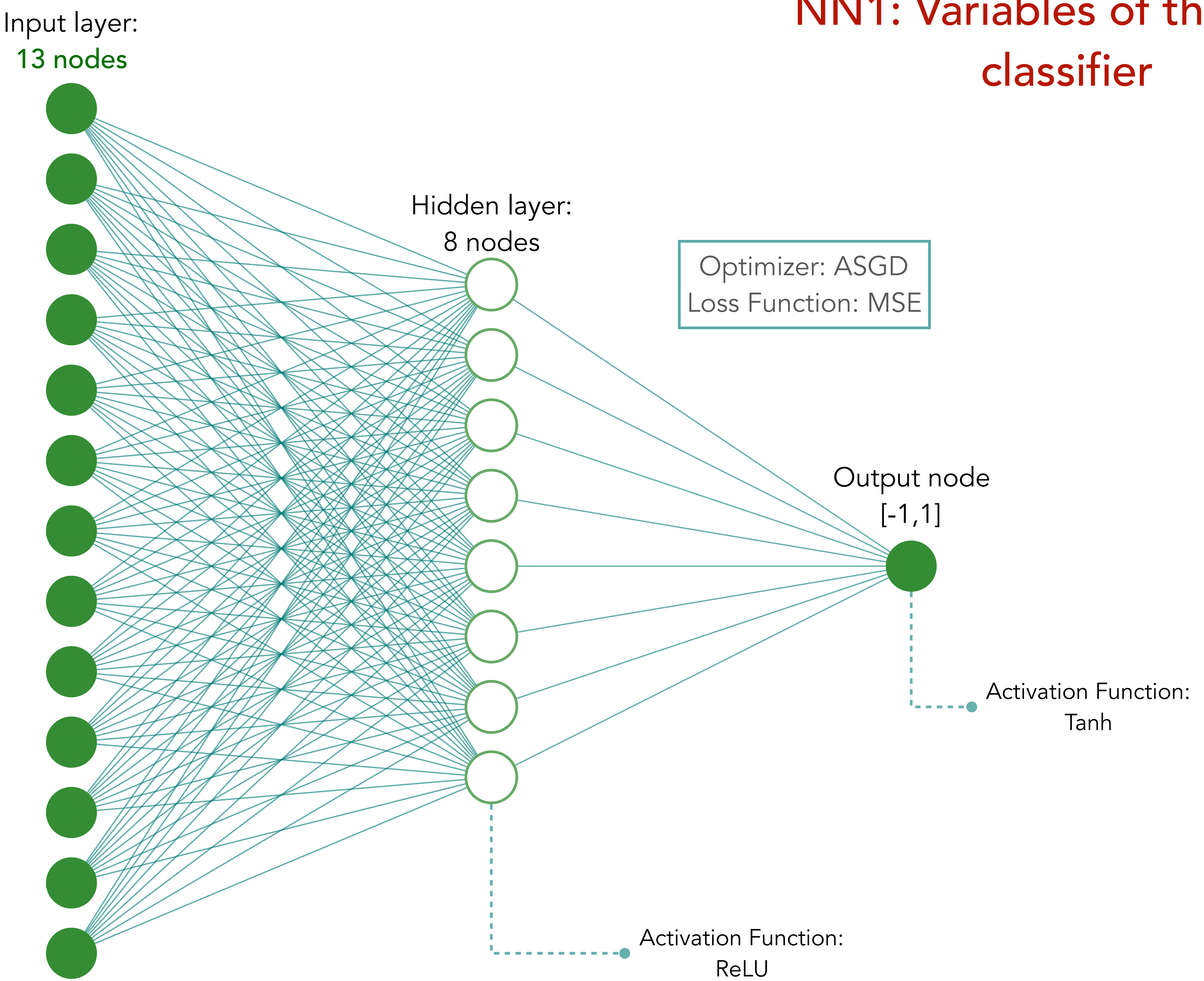


Evaluation of the network

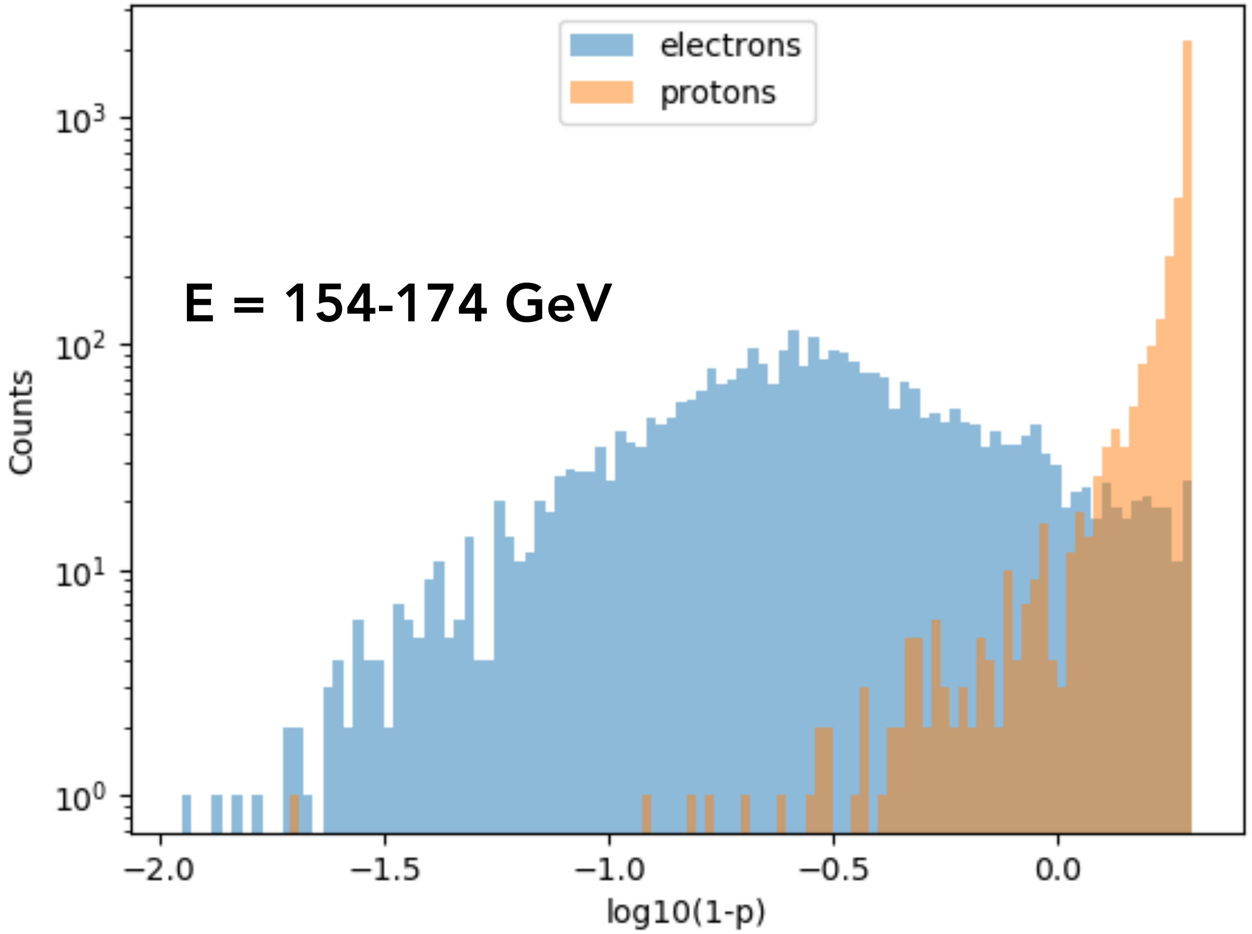
To compare different models, **ROC (Receiver Operating Characteristic) curve**, i.e. *True Positive Rate (TPR)* as a function of *False Positive Rate (FPR)* is shown, and **AUC (Area Under the Curve)** is calculated.

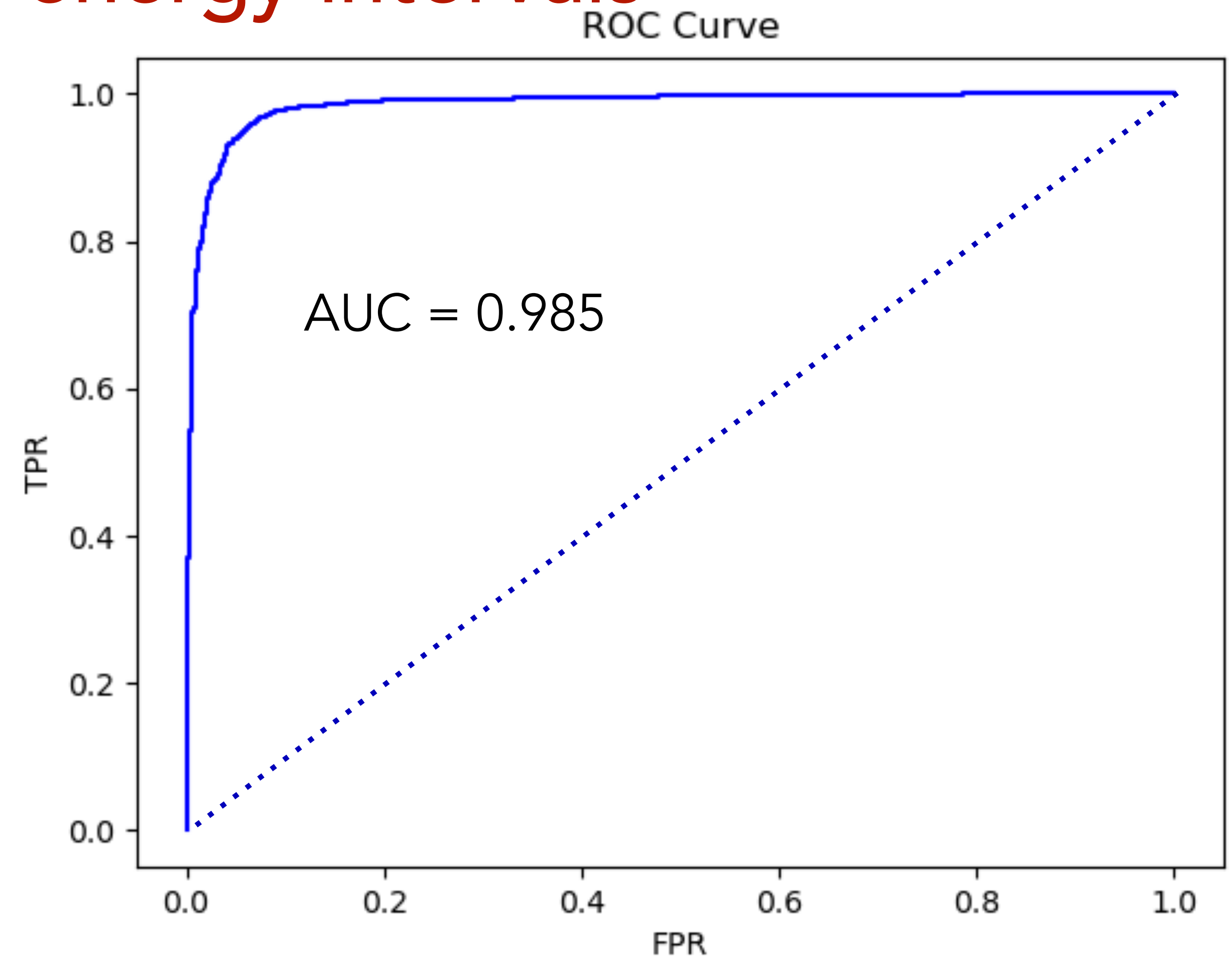


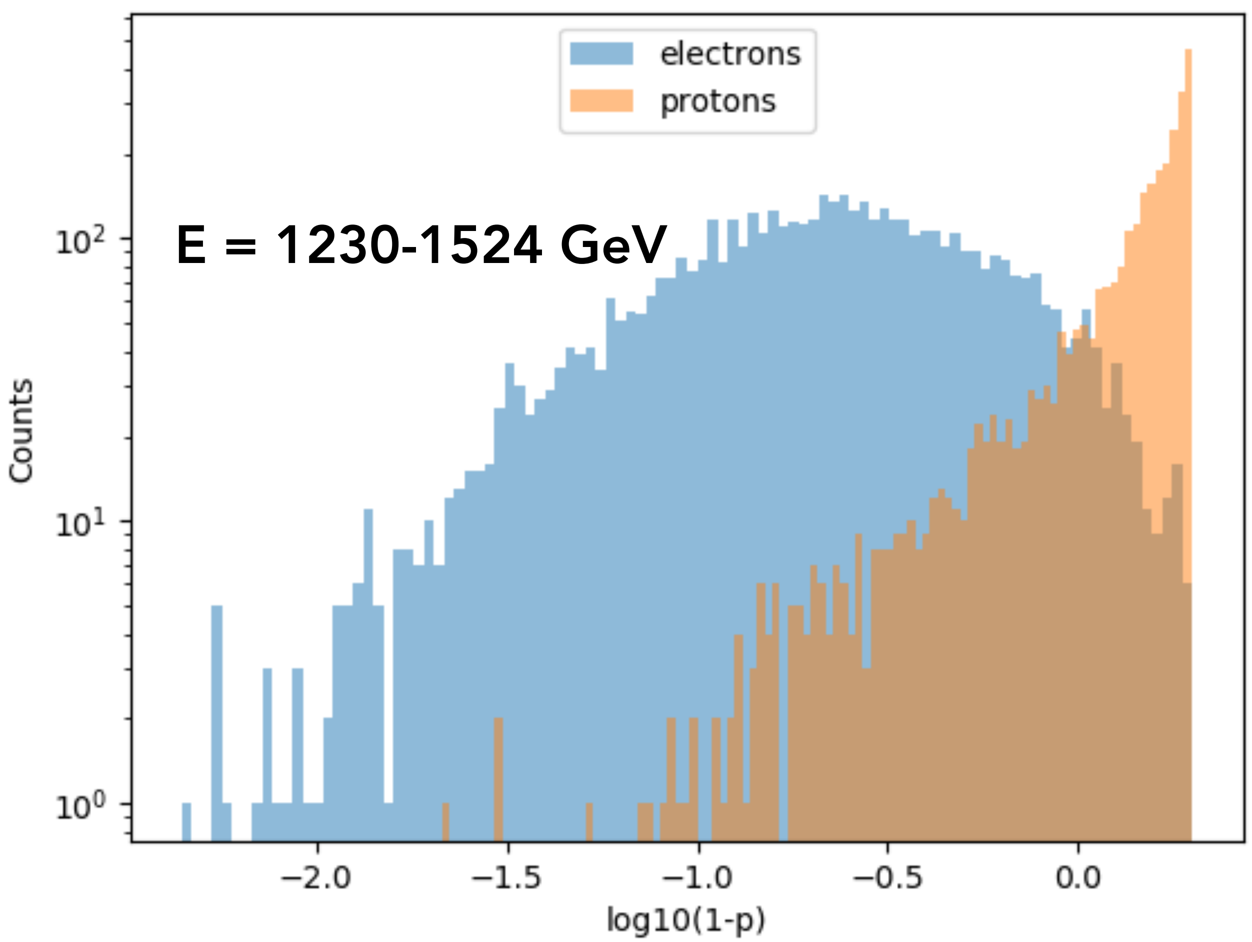
NN1: Variables of the 2017 classifier

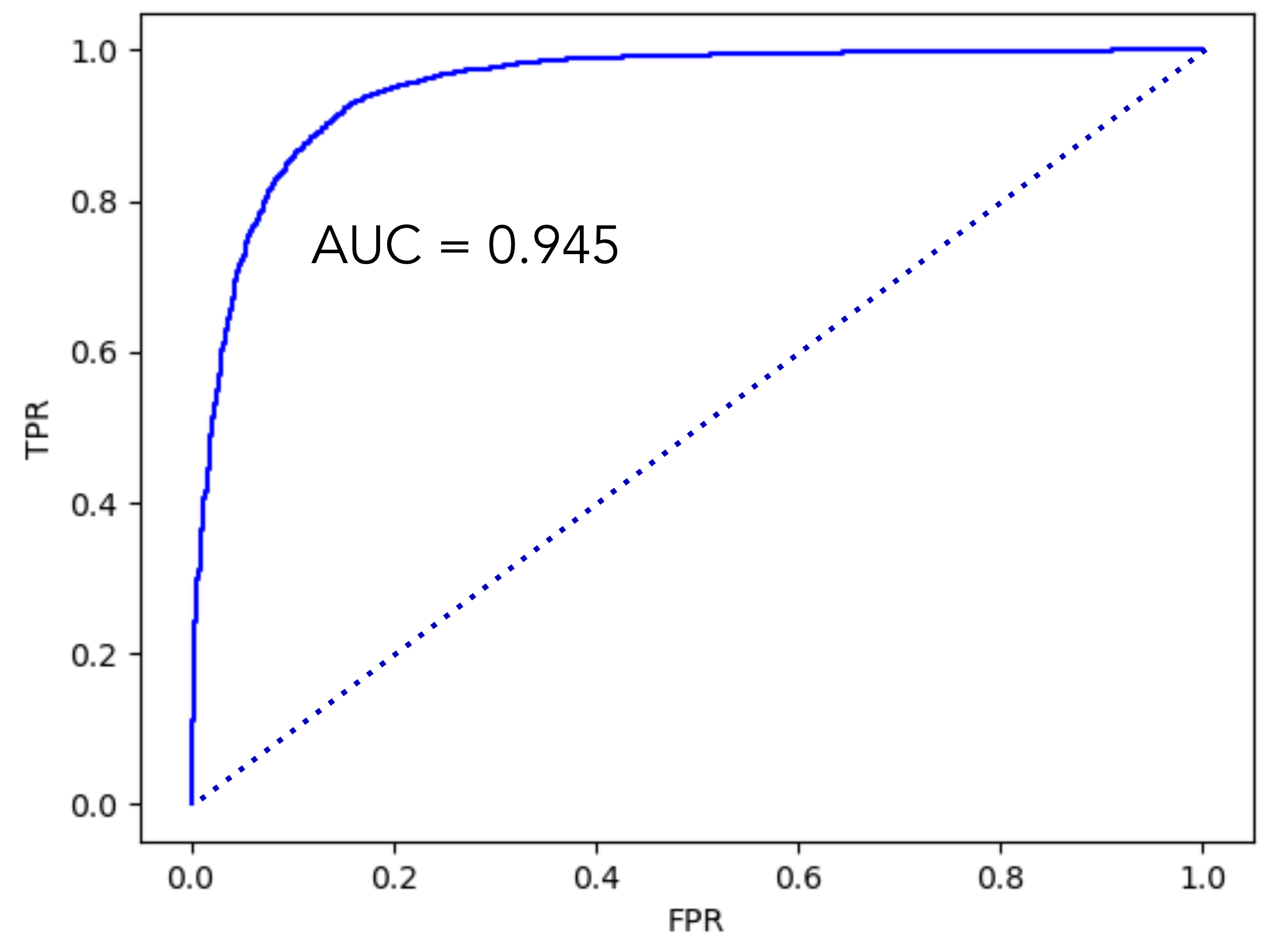


Output for two energy intervals

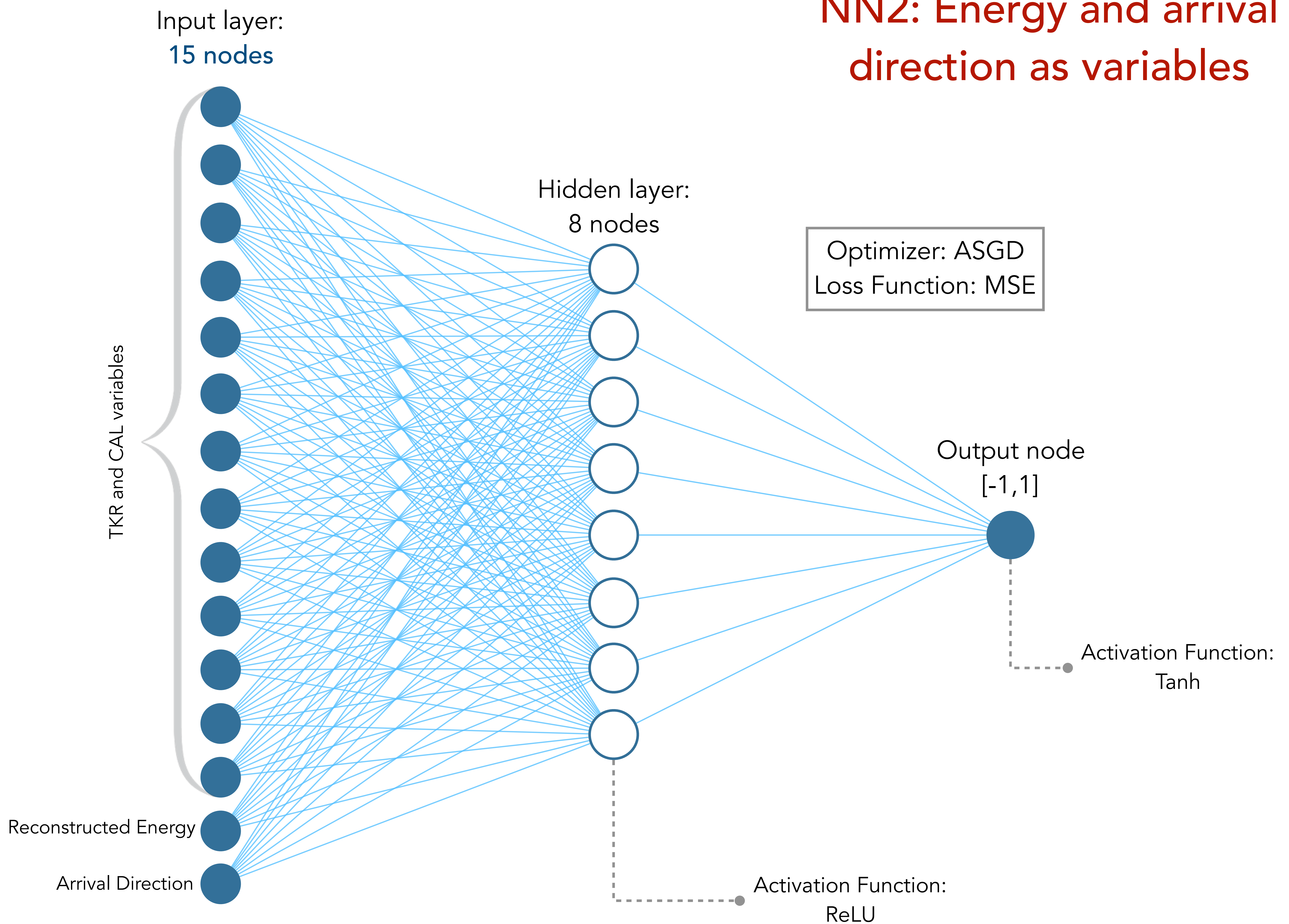




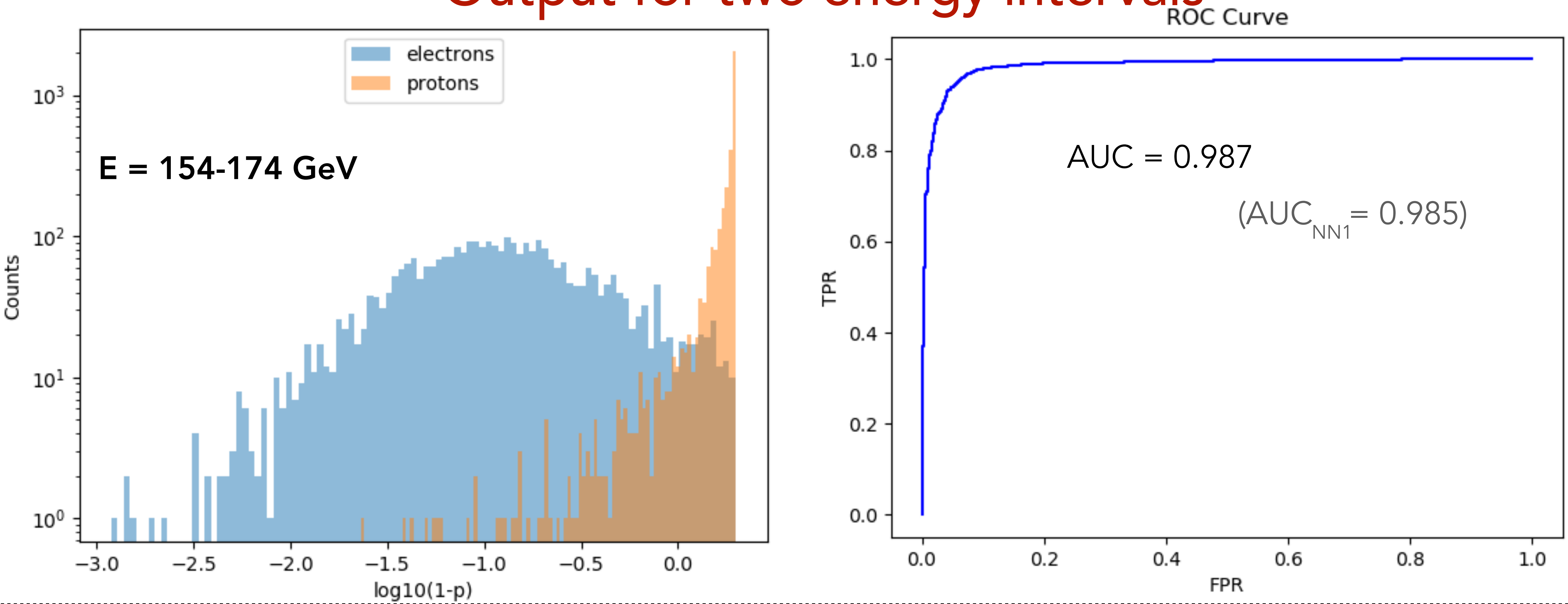


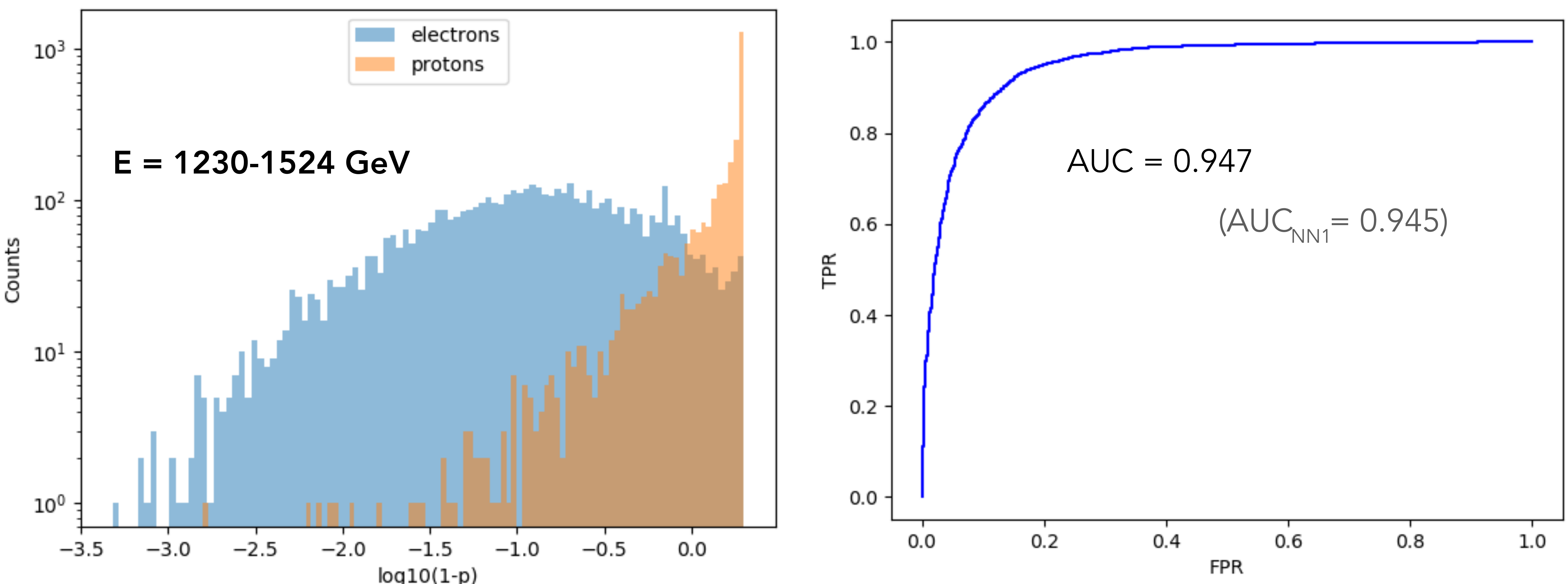


NN2: Energy and arrival direction as variables

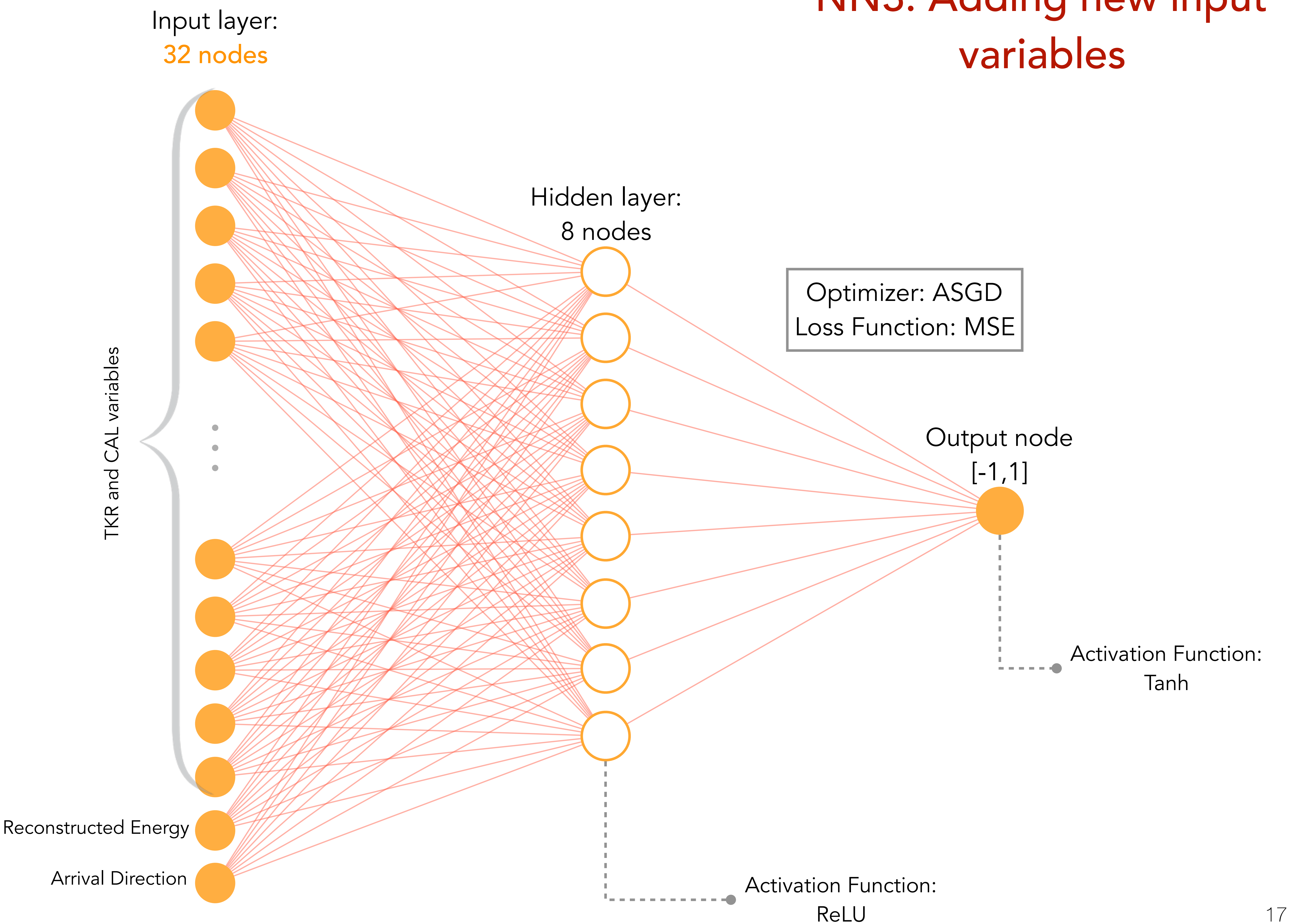


Output for two energy intervals

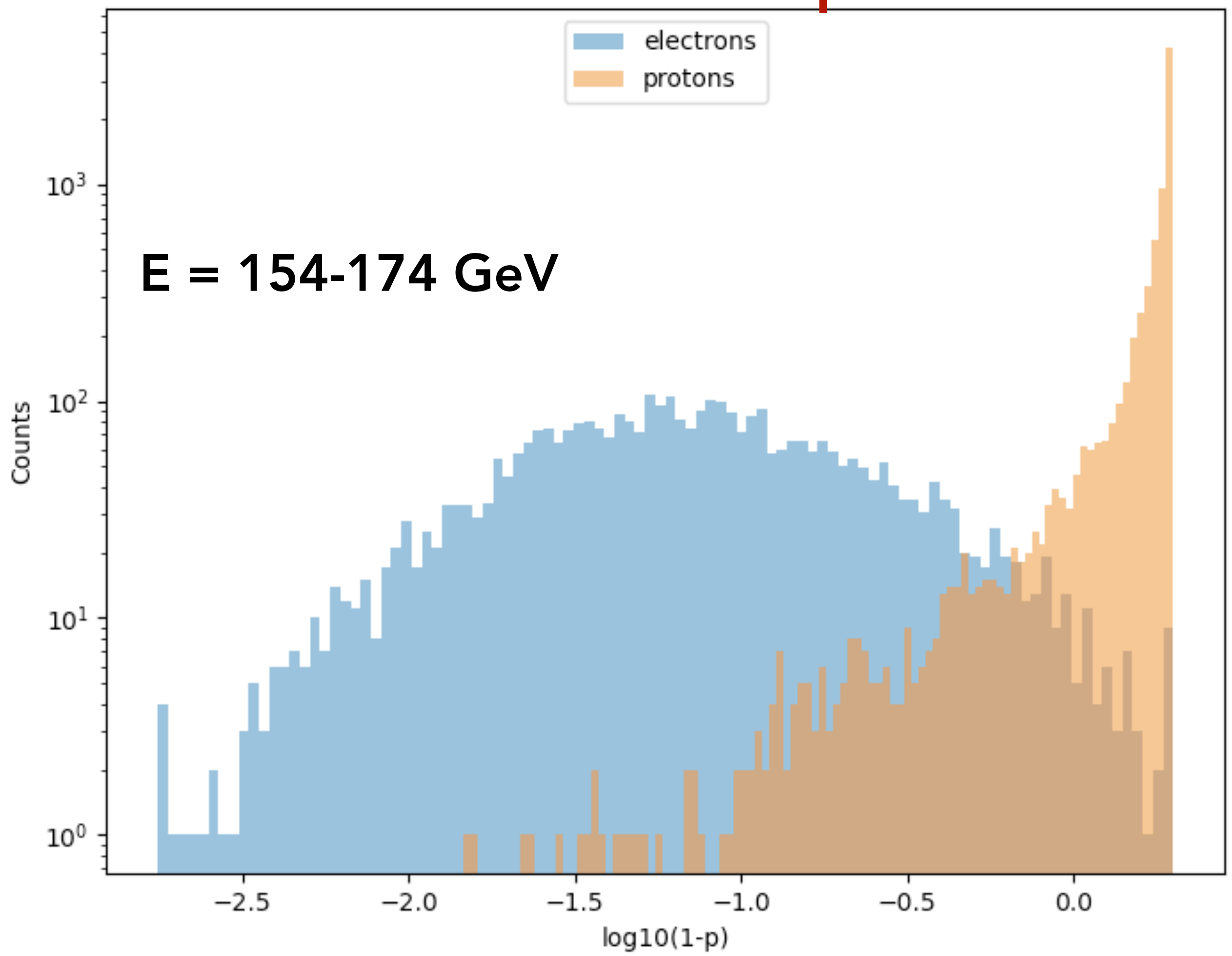


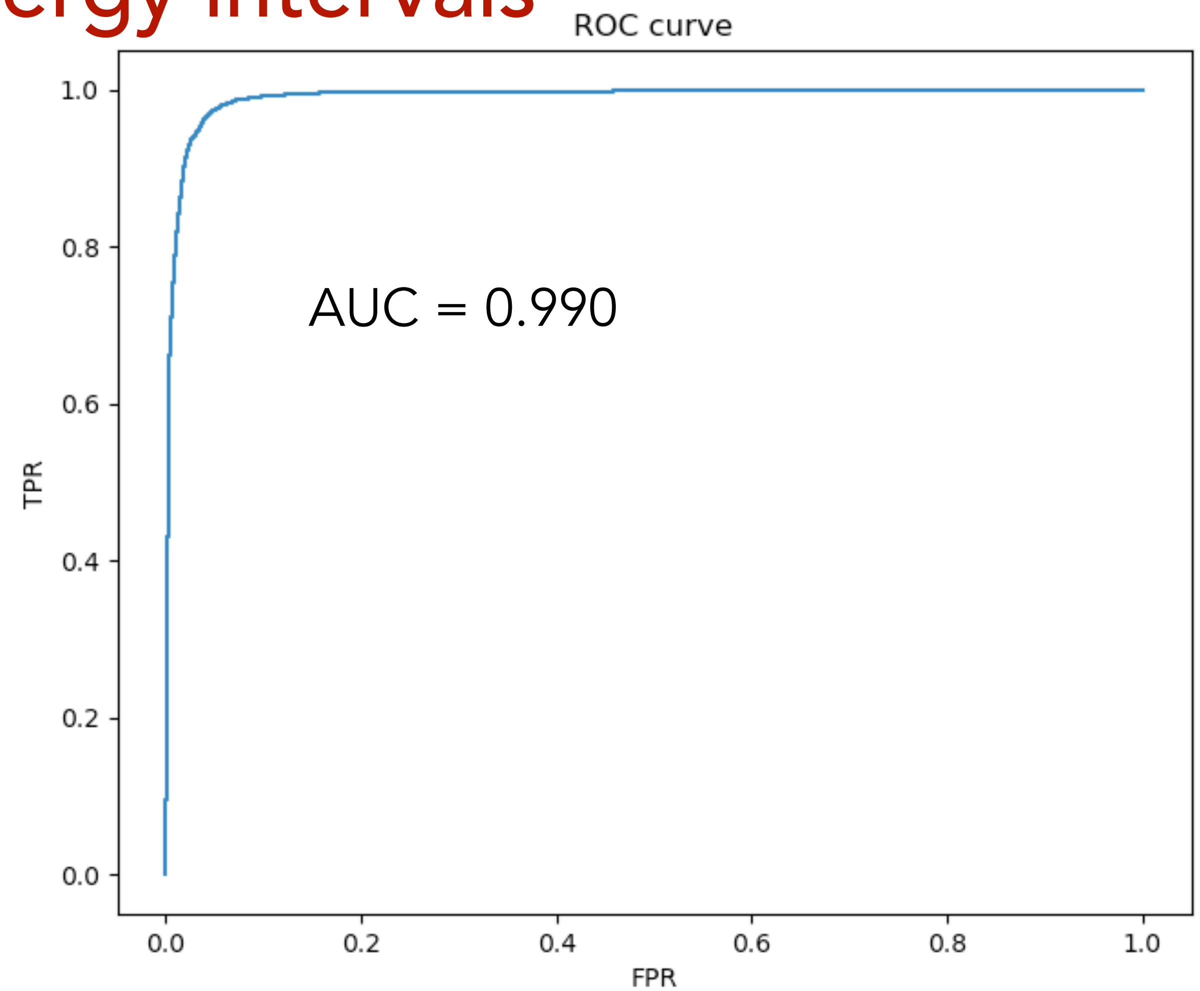


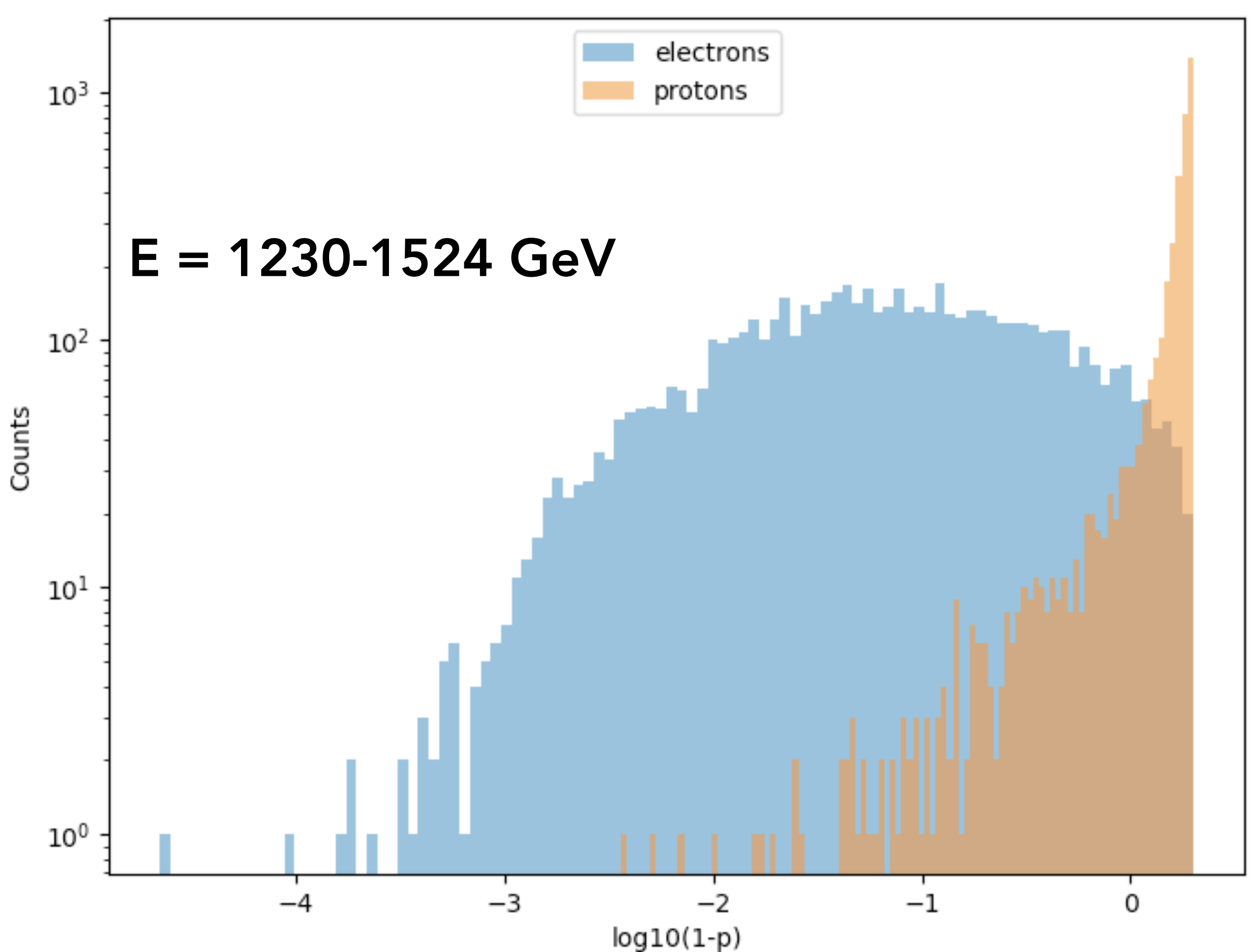
NN3: Adding new input variables

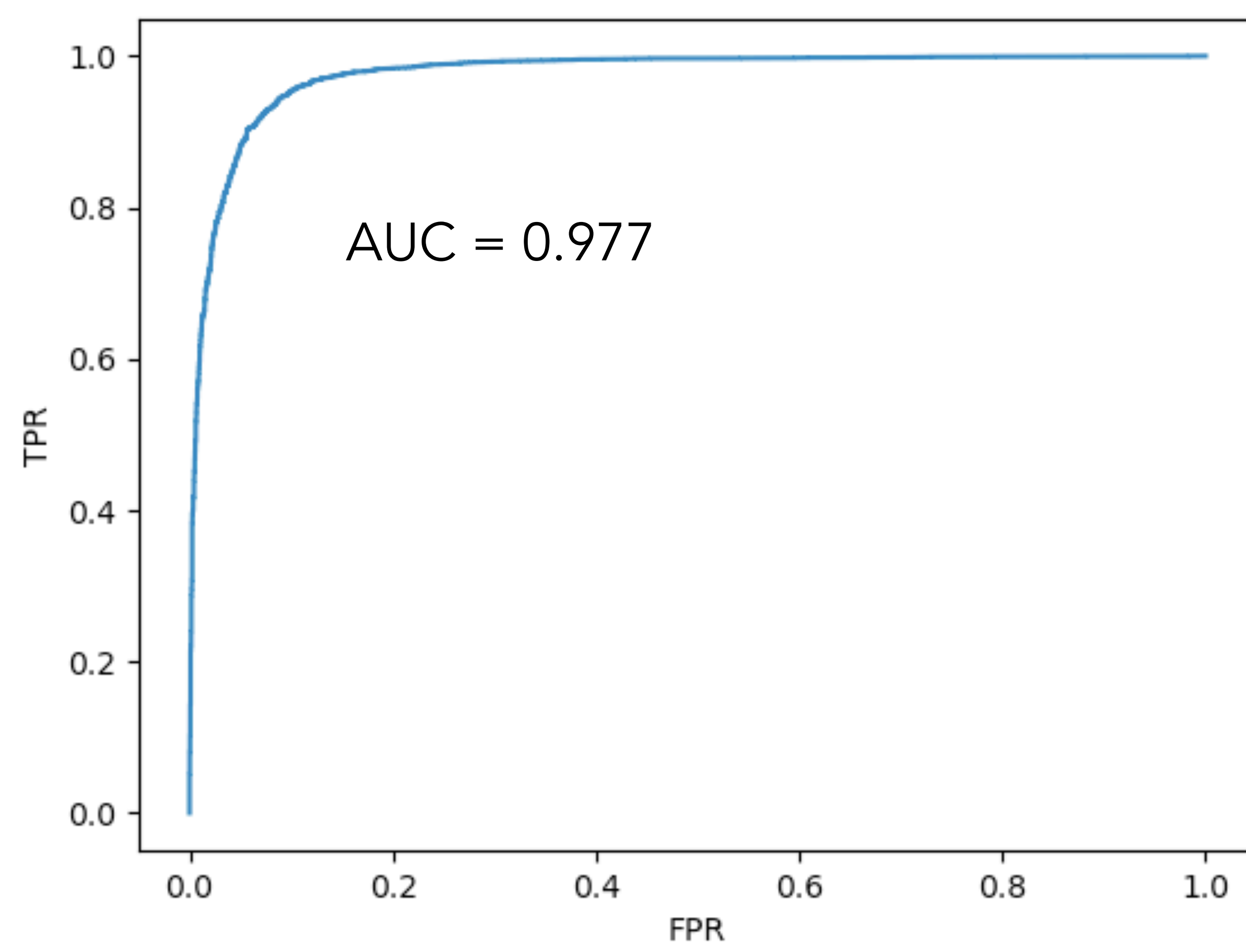


Output for two energy intervals

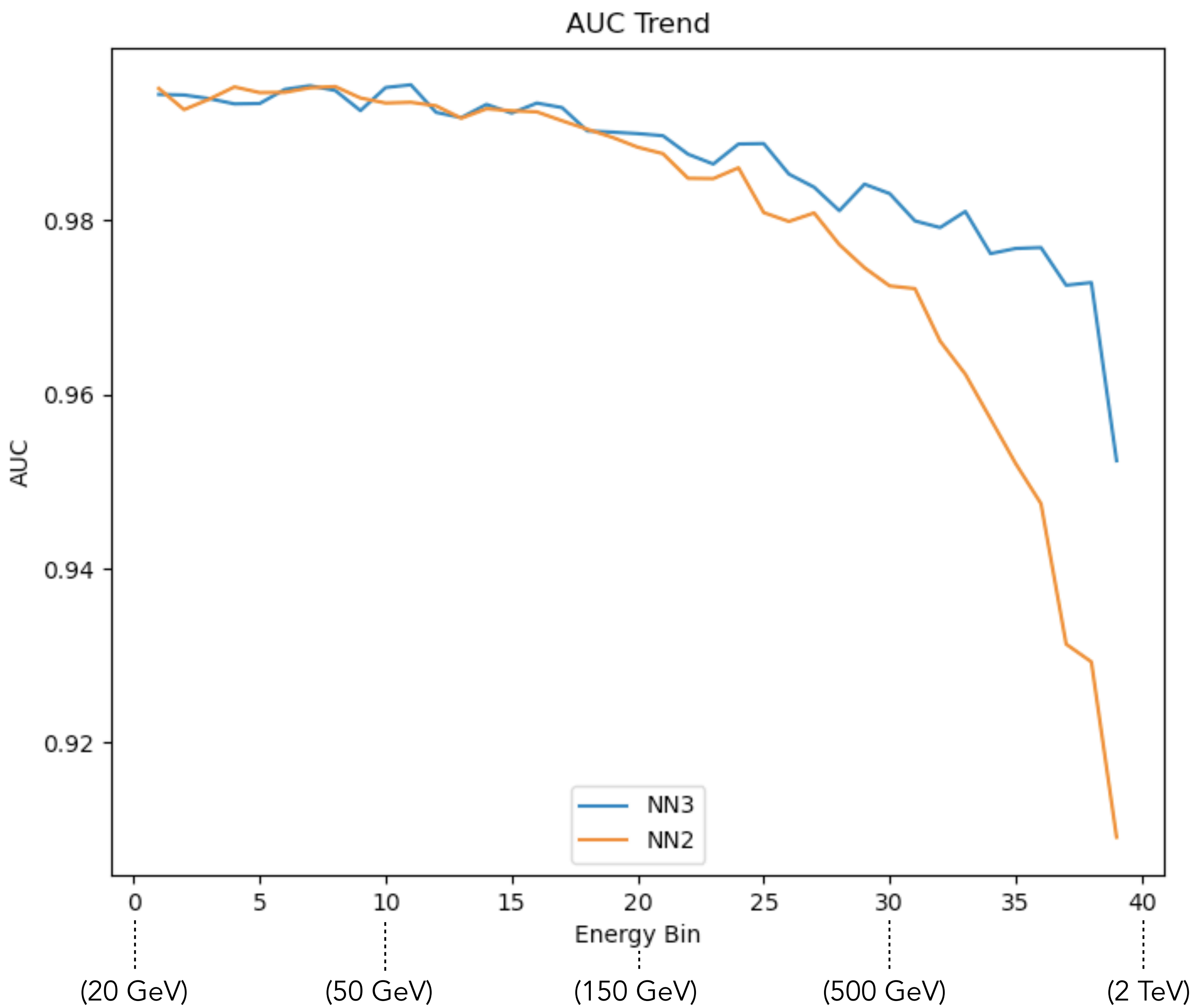








NN2 and NN3 comparison



NN3: Neural network to be used with experimental data

Datasets for CRE spectrum

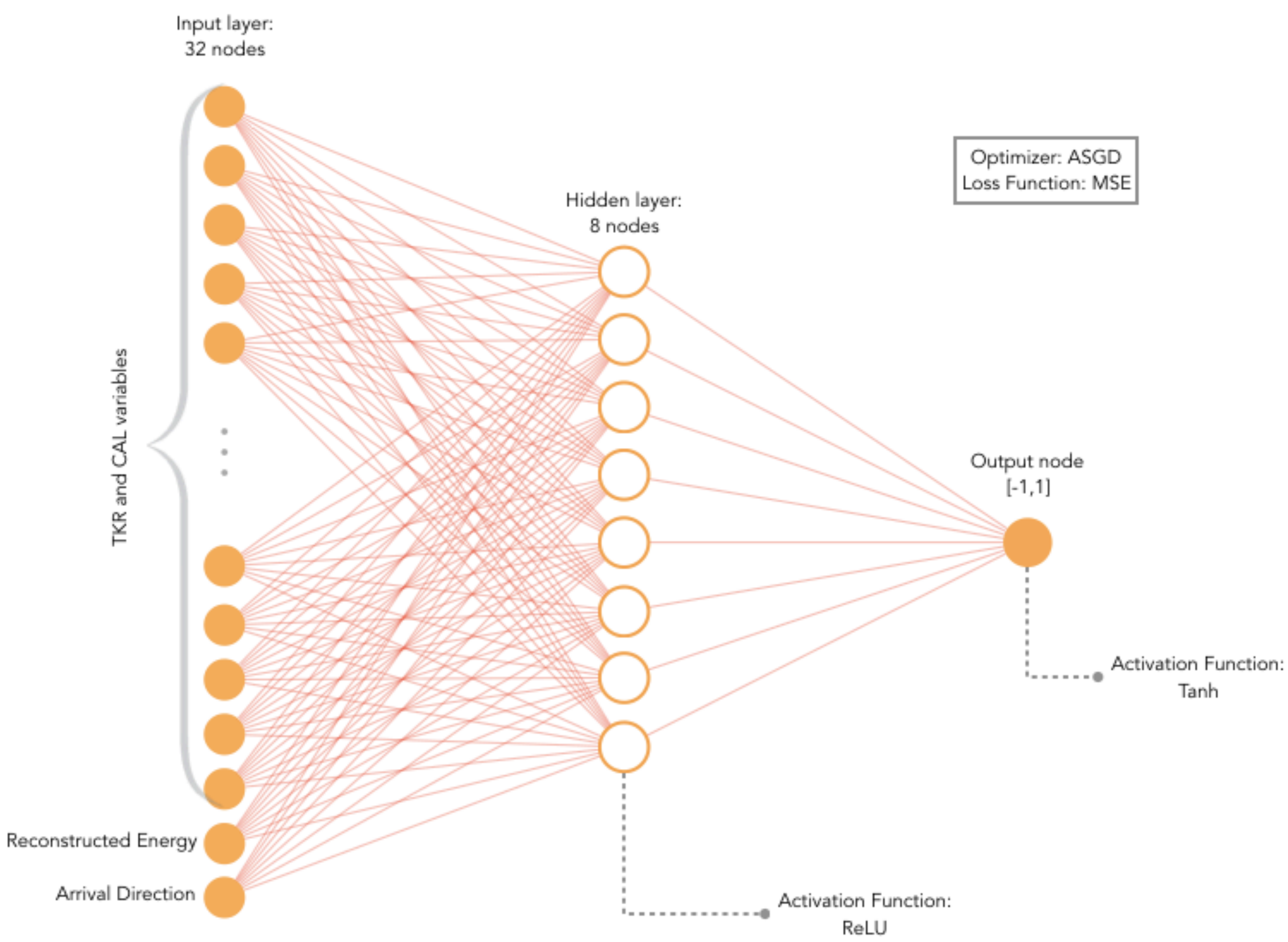
Experimental Data:

5Y of data (Aug 2015 - Aug 2020) -> $4*10^8$ events

Monte Carlo simulations:

MC electrons -> $6*10^5$ events

MC protons -> $9*10^6$ events

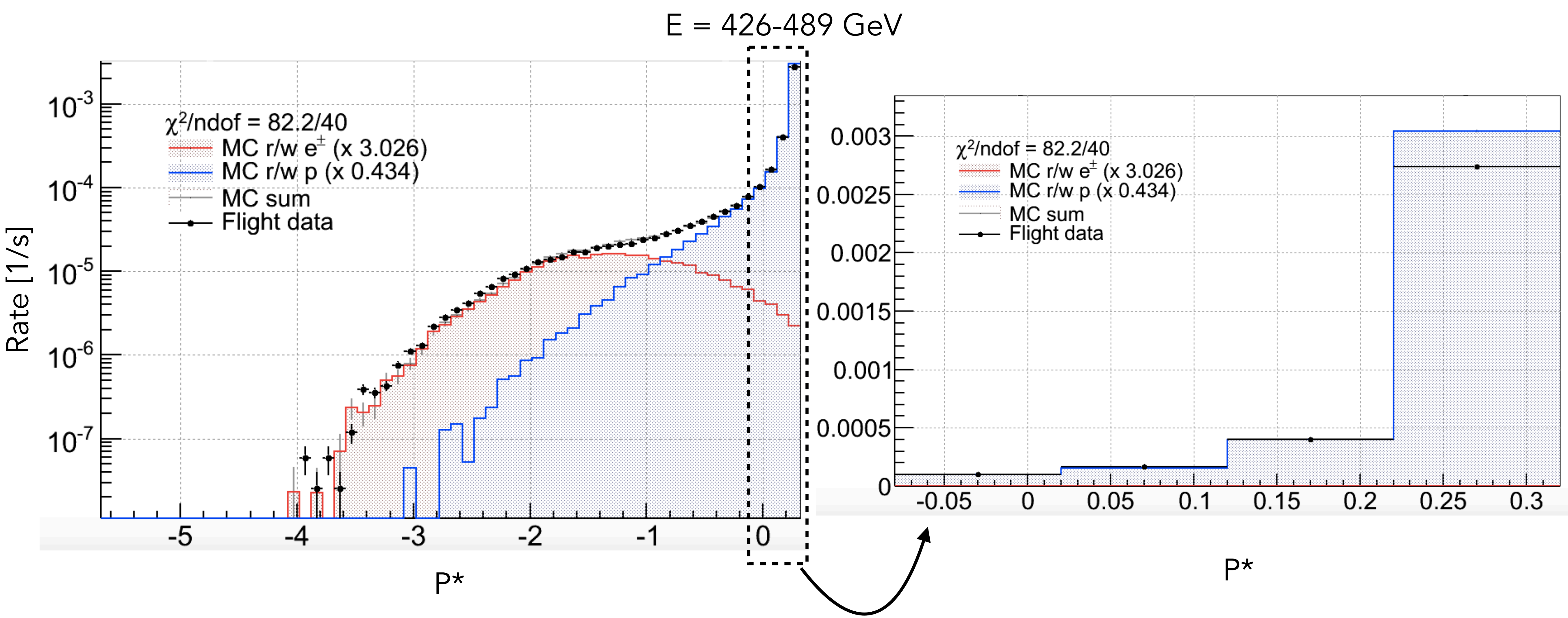


- For each event a value of **P*=log(1-p)** is assigned through NN3

- **Counts** are converted in **rates** using *Livetime*

Template Fit

Estimate the rate of electrons and protons for each energy bin, **combining the informations of both MC simulations and experimental data.**

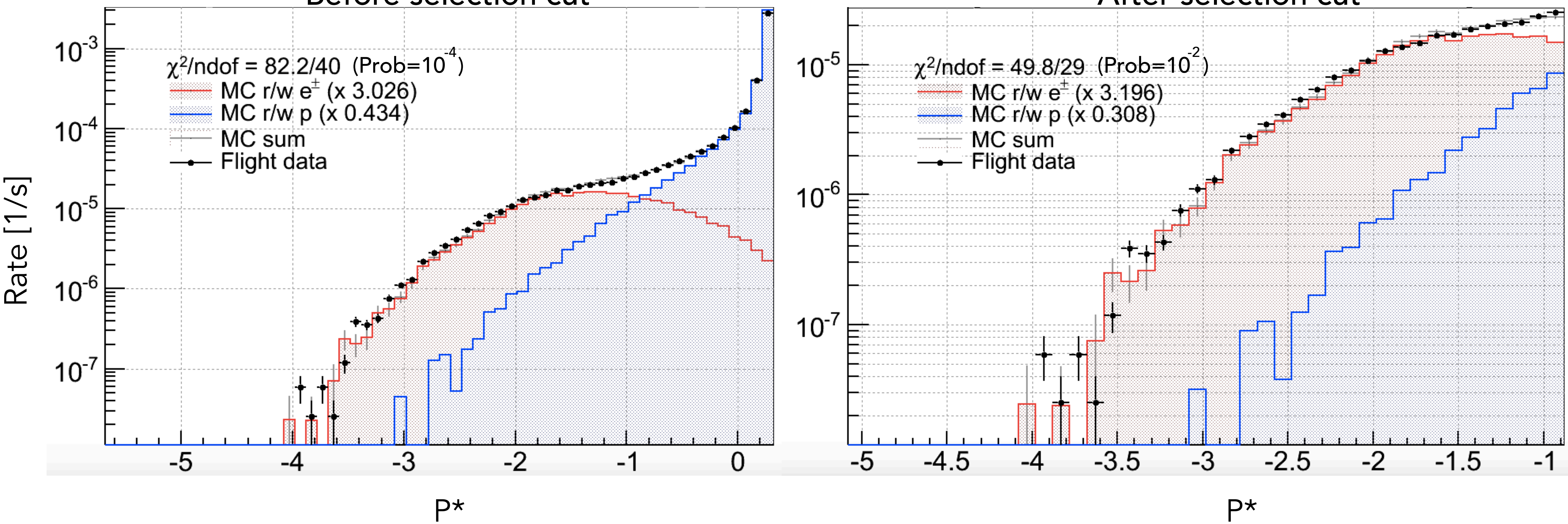


Selection Cut

Selecting a maximum value of P* for computing the template fit, performances can be improved.

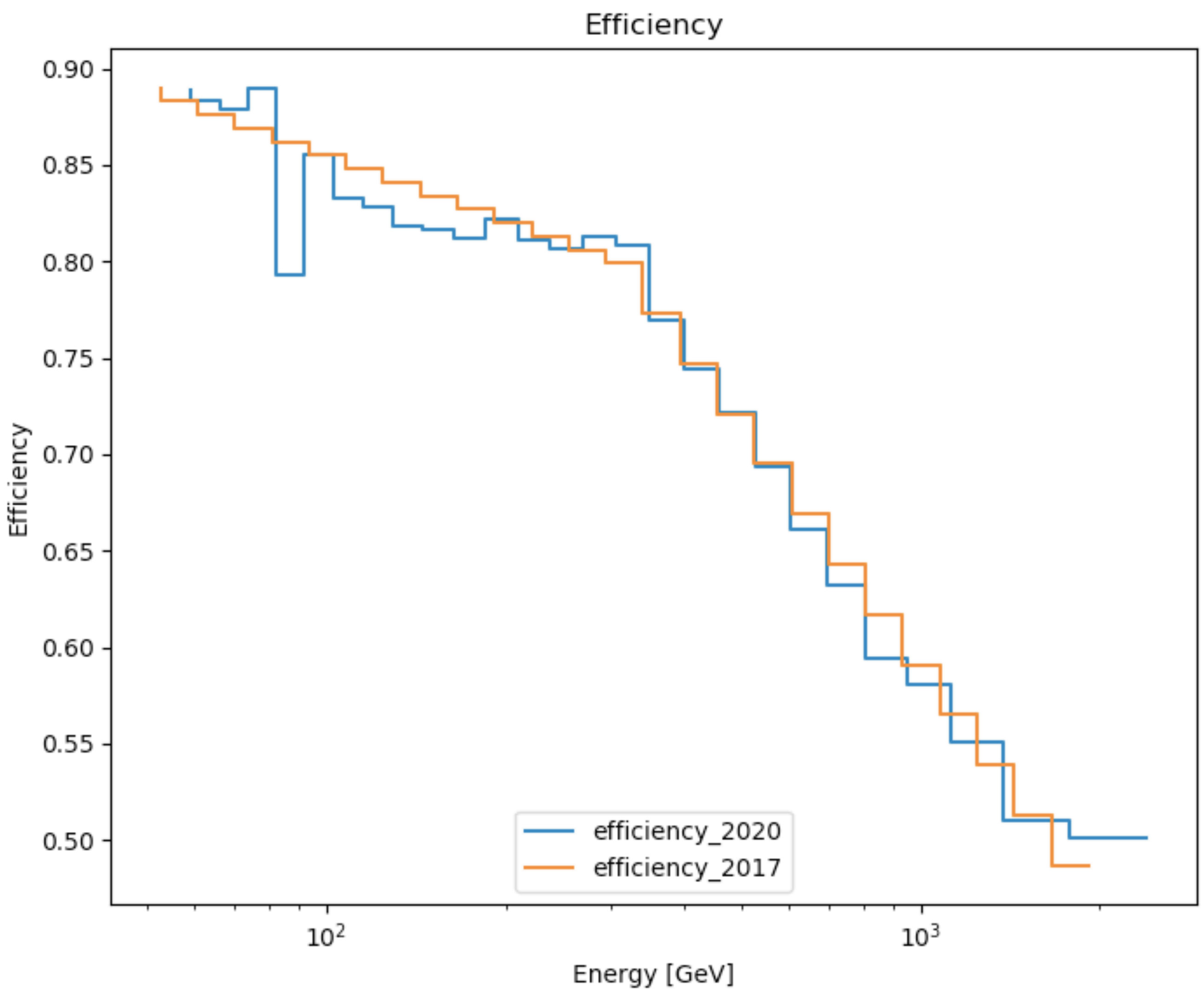
E = 426-489 GeV

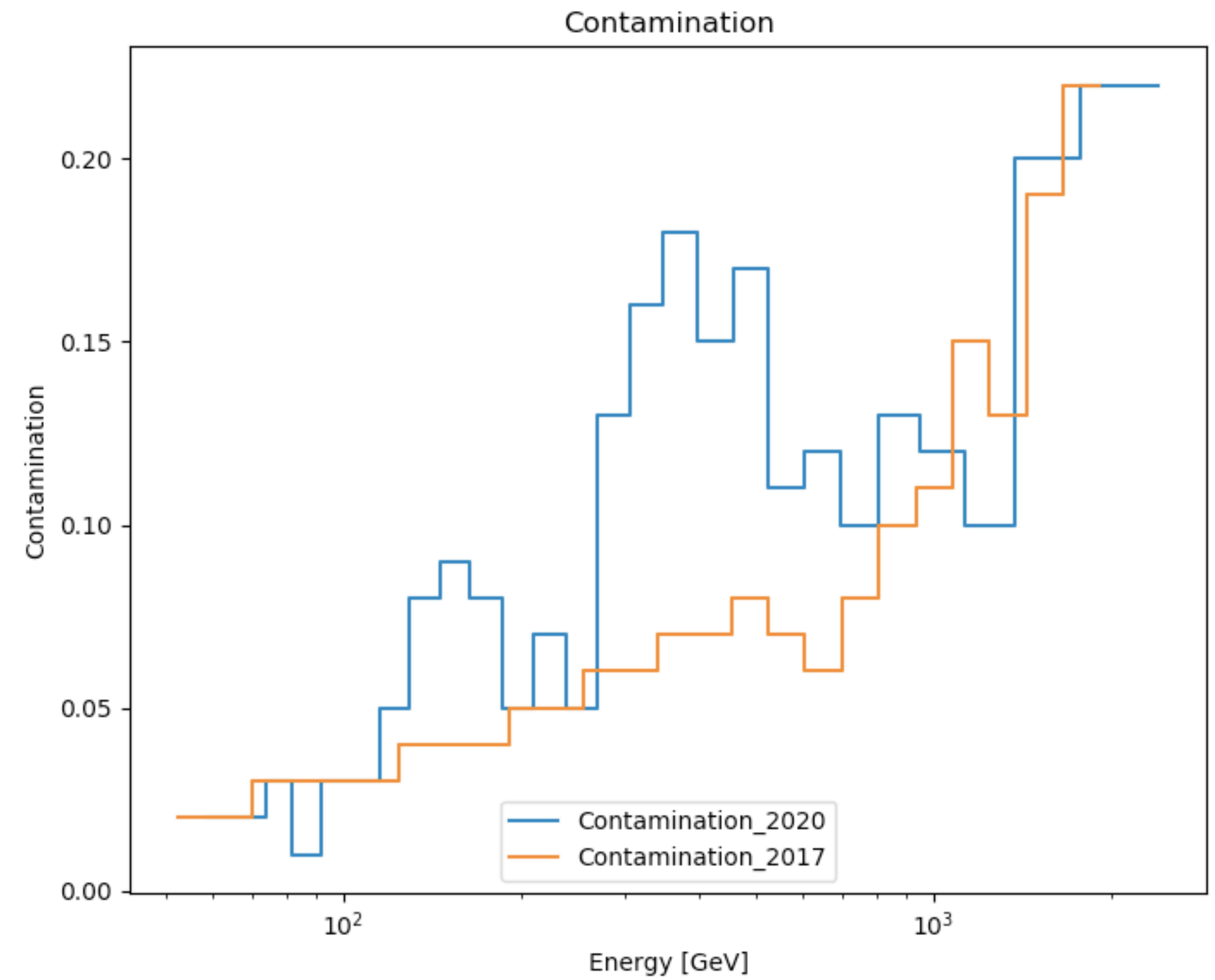
Before selection cut | After selection cut



Efficiency and Contamination

For equal values of efficiency, a comparison between the 2017 analysis and this work is possible looking at the values of contamination



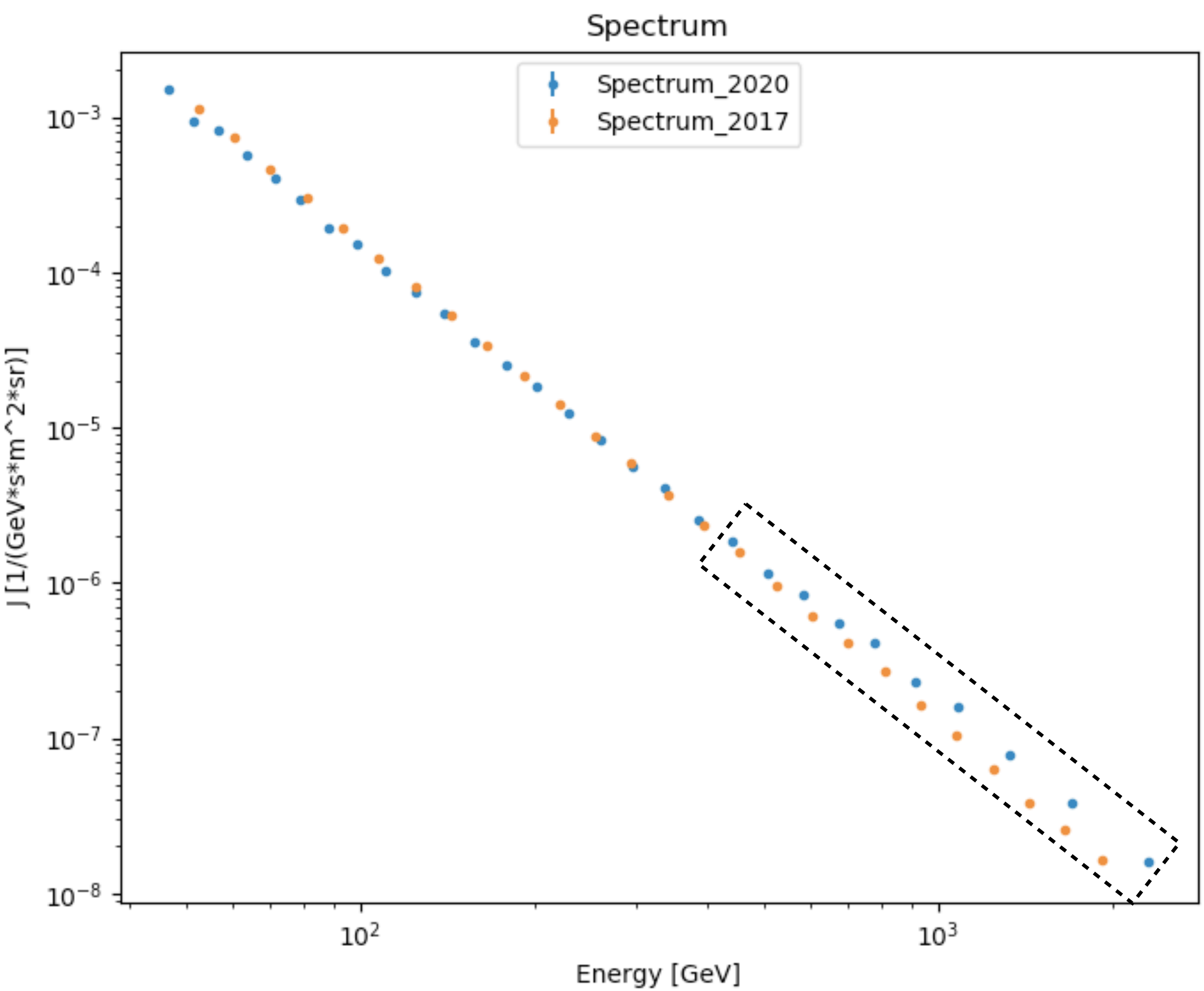


Main possible reason

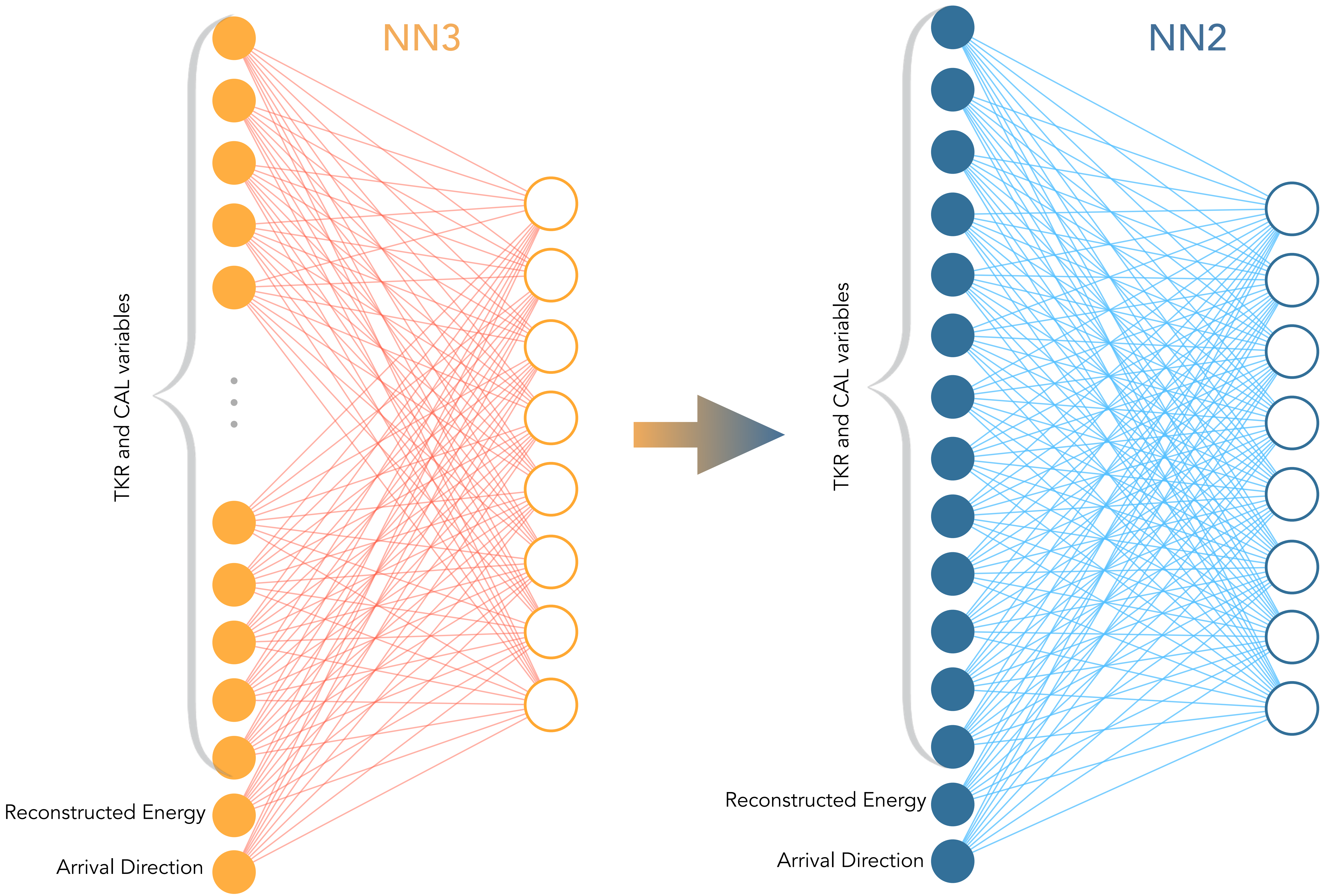
Absence of Data/MC agreement corrections

Final Spectrum

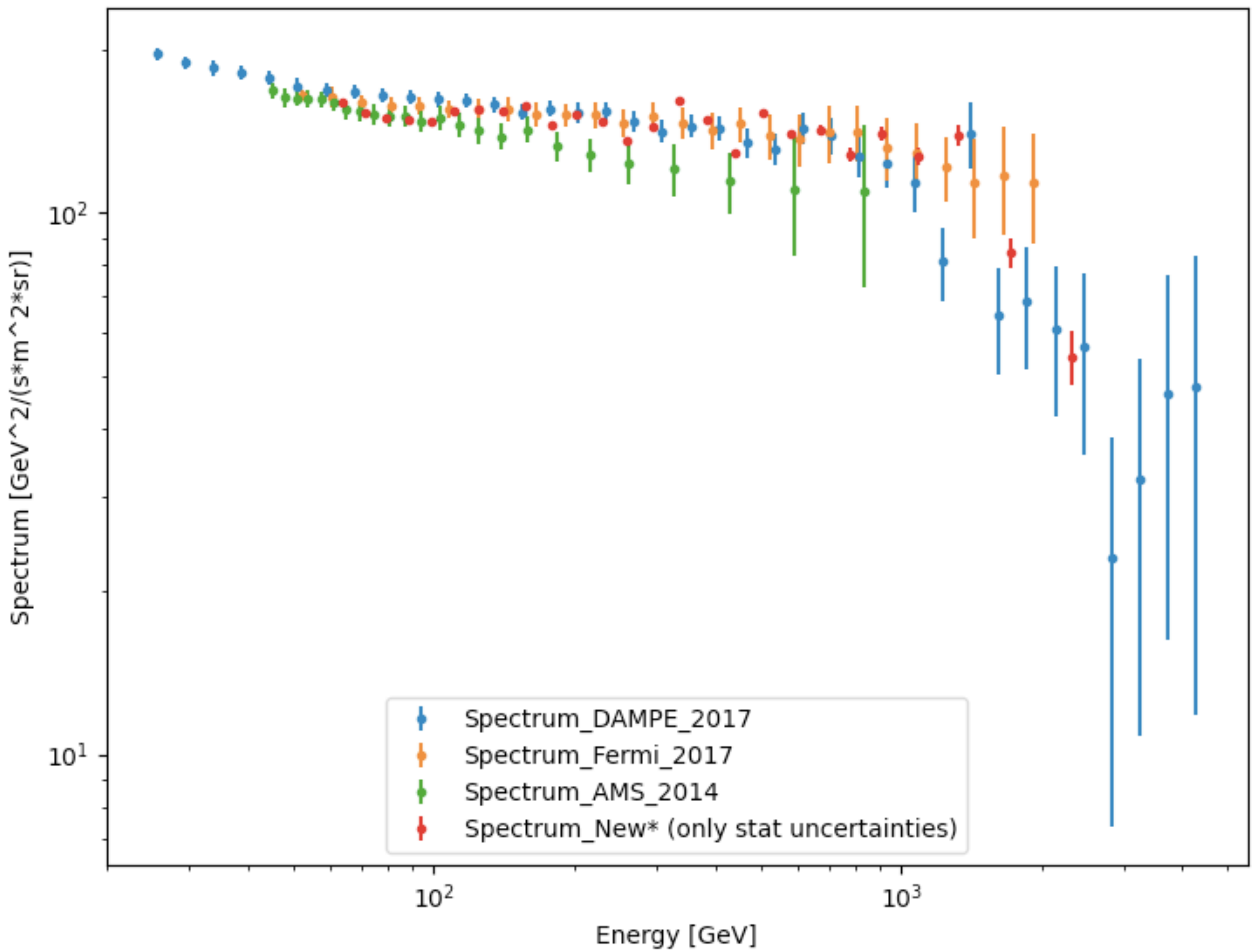
Only Statistical Uncertainties are reported, systematic errors still to be estimated



Changing variables



Preliminary Spectrum

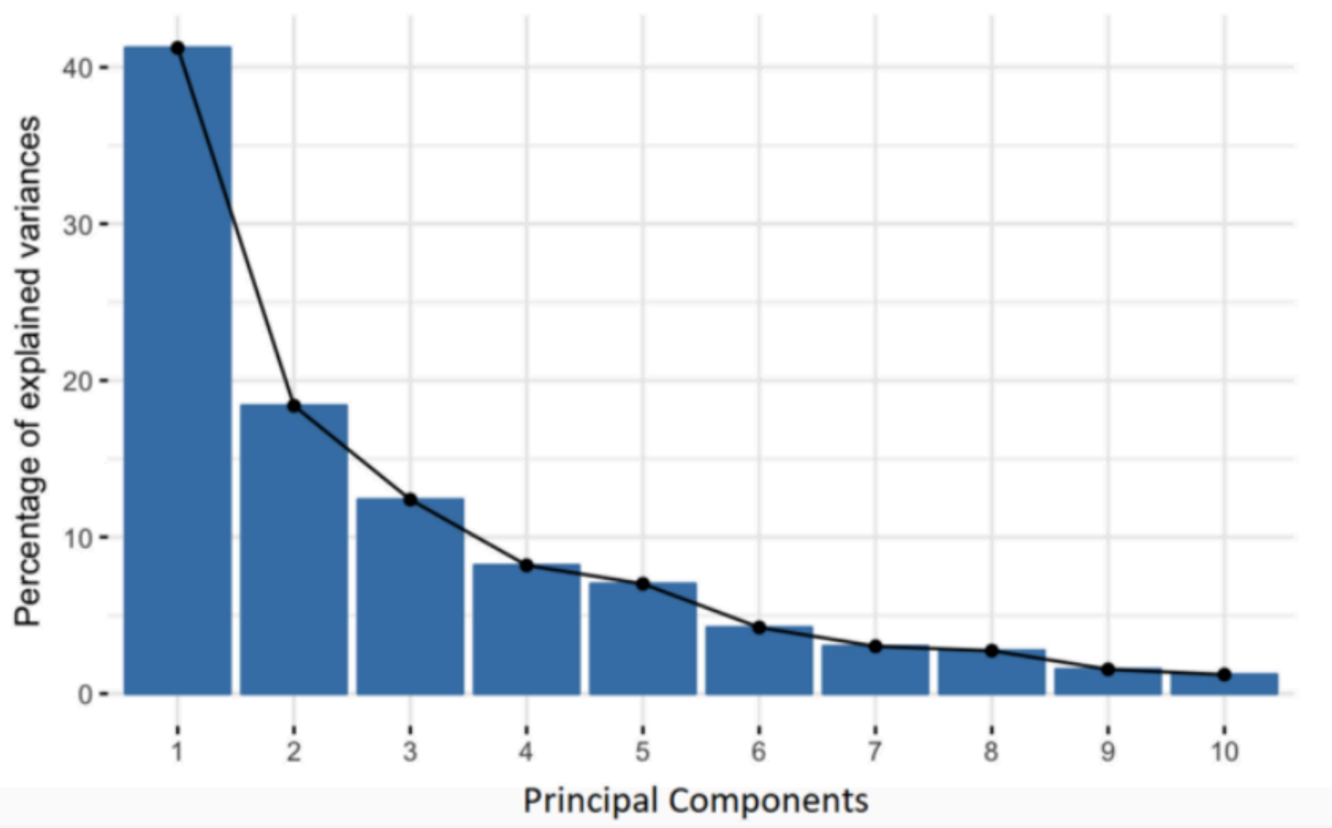


2nd Part

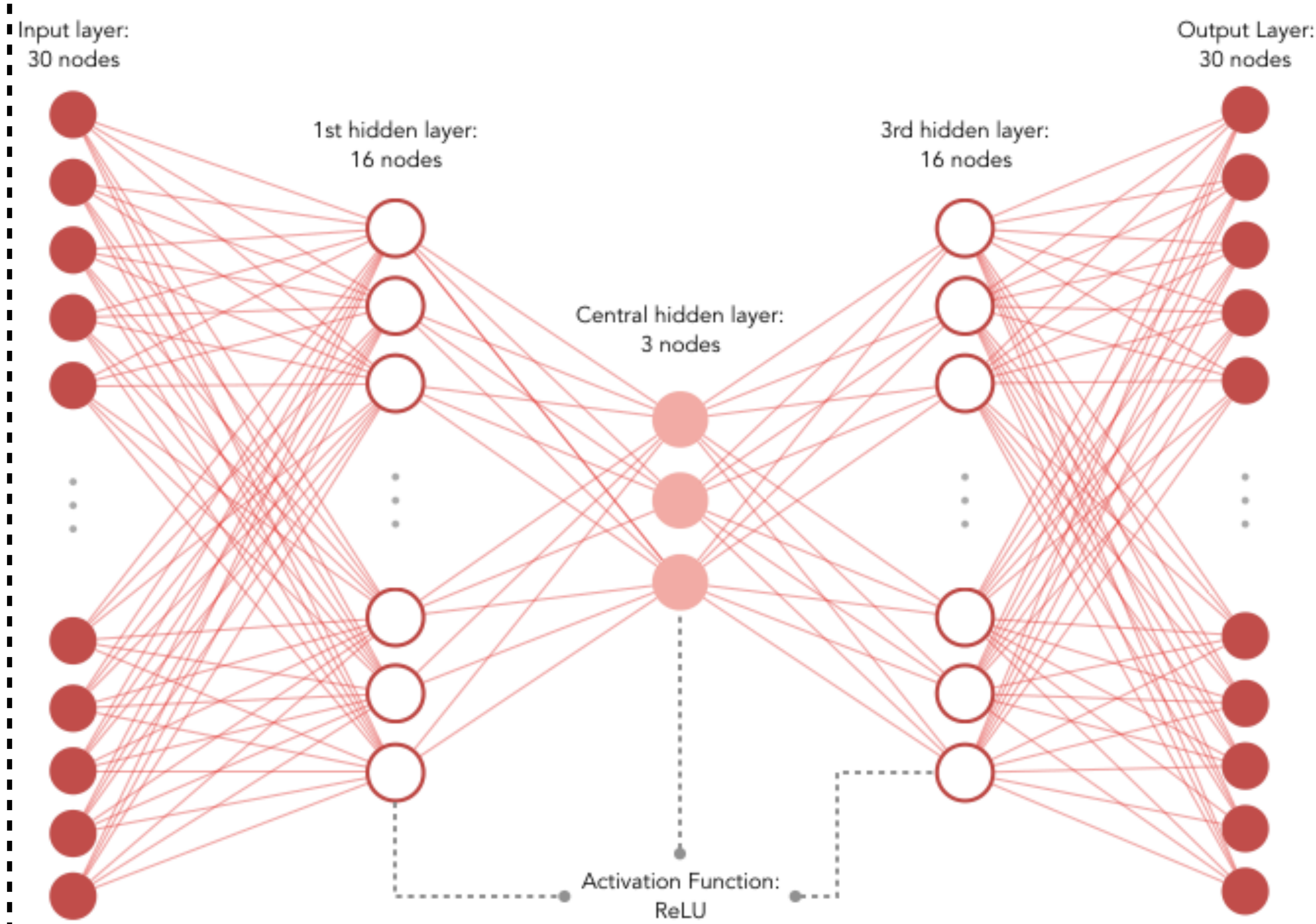
Unsupervised Learning

Dimensionality Reduction

Principal Component Analysis



Autoencoders

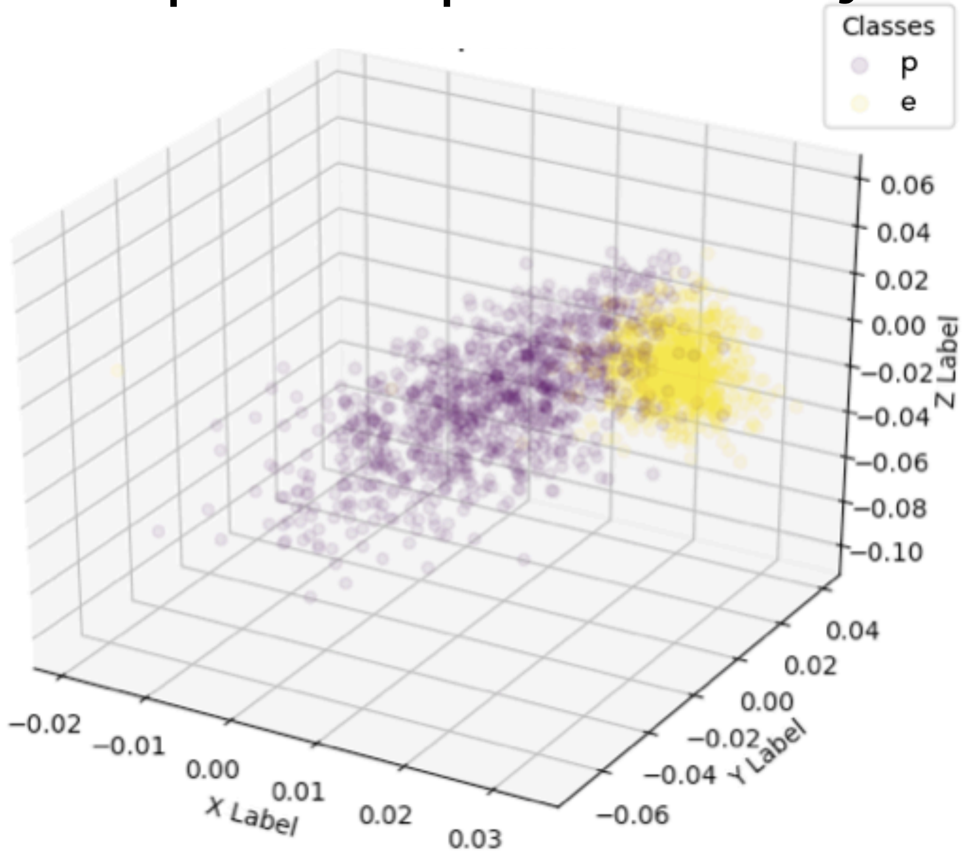


Input space reduced to 3 variables with both methods

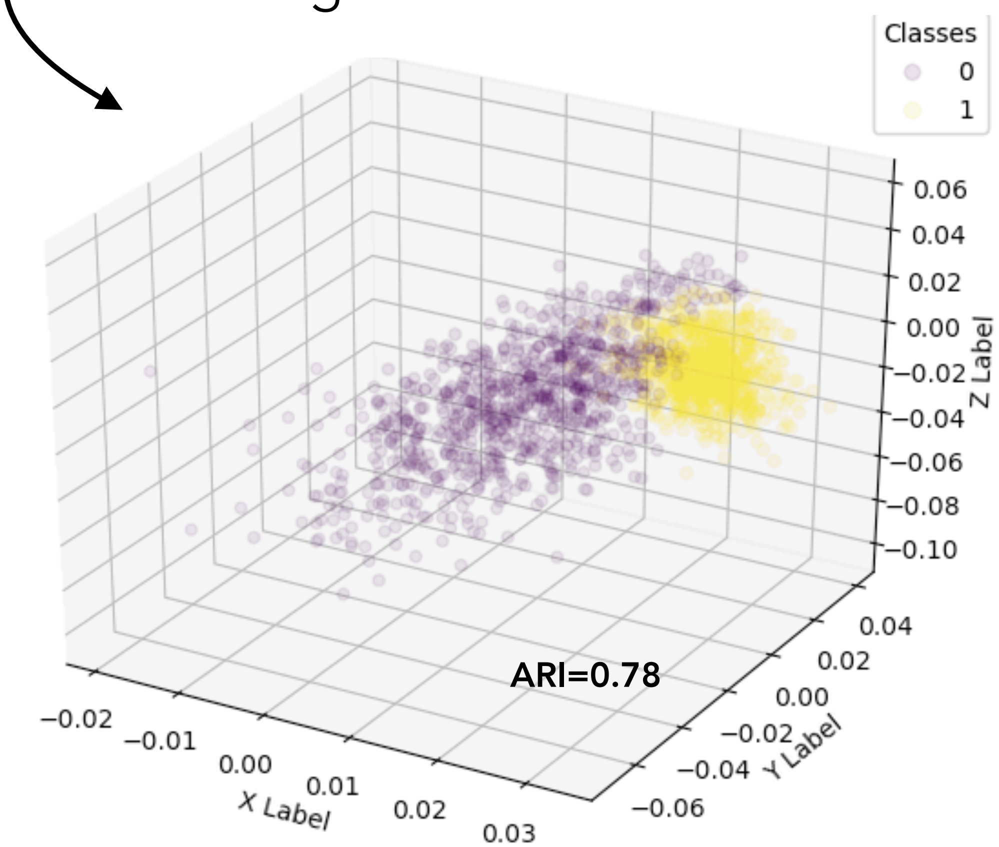
After the dimensionality reduction, a clustering algorithm is applied, and evaluation is performed through the **Adjust Rand Index (ARI)** value

Results for e/p=1

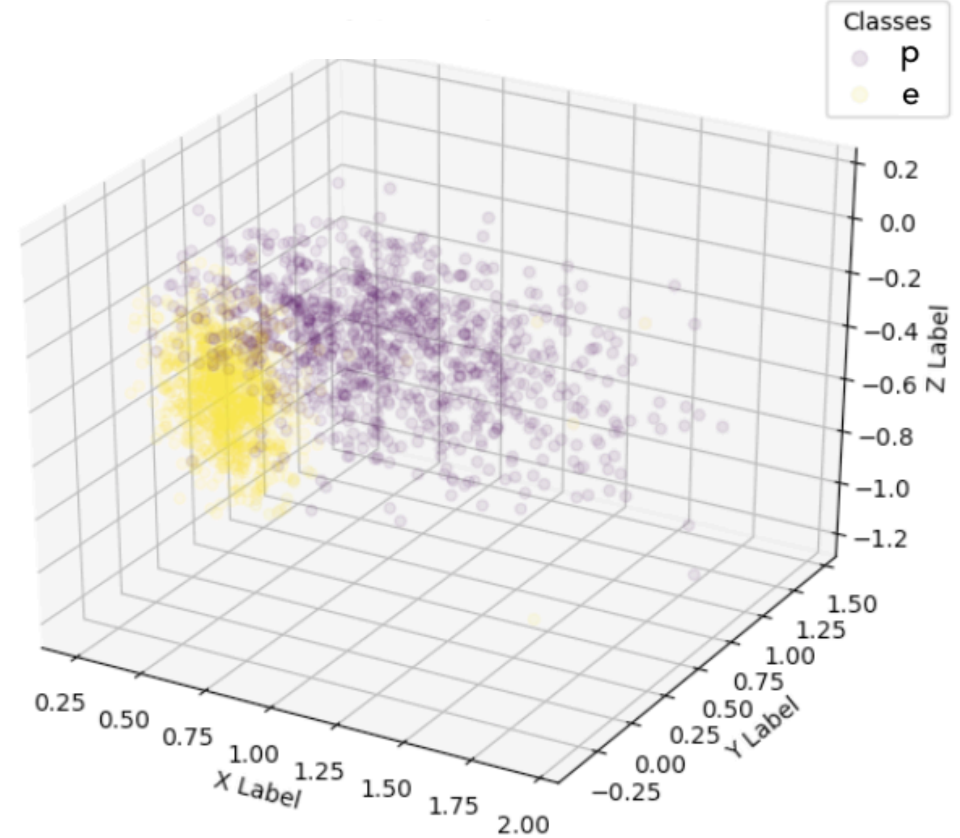
Principal Component Analysis



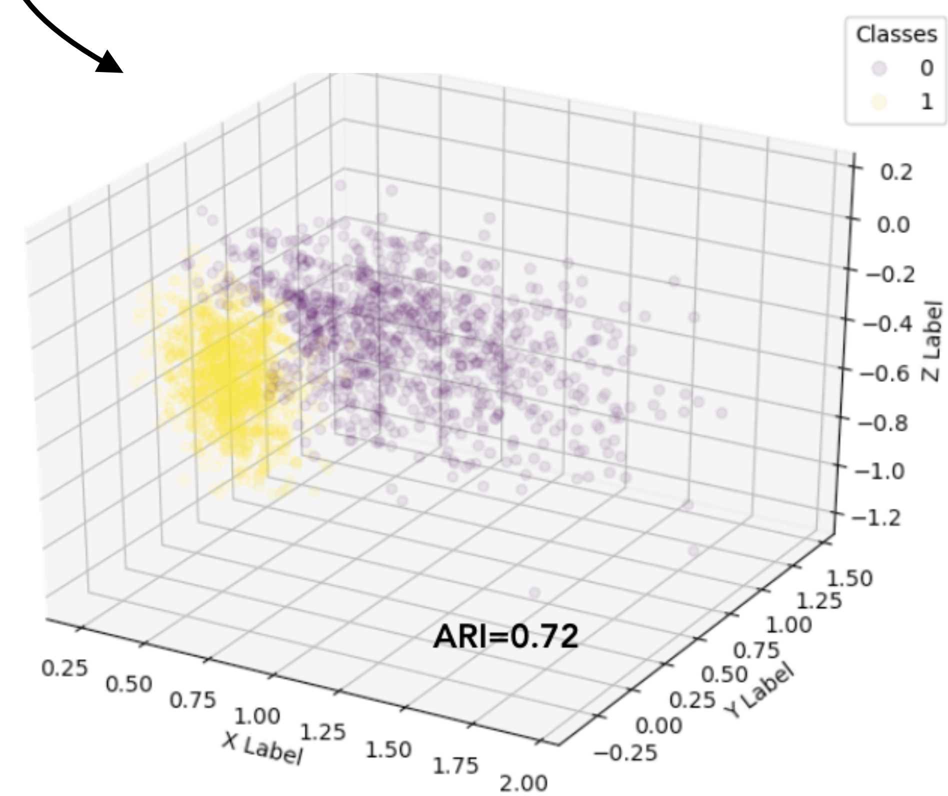
Clustering: *Hierarchical*



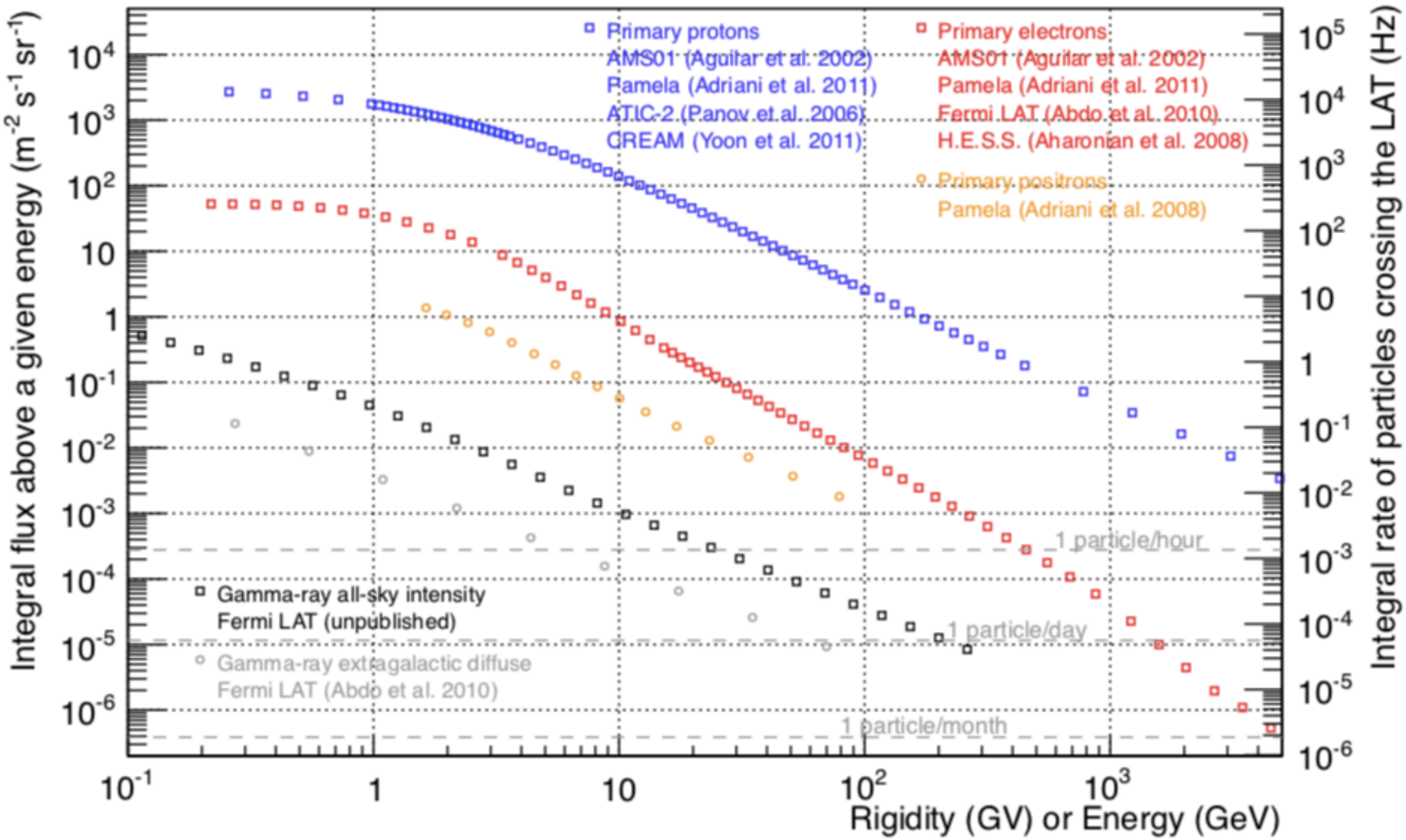
Autoencoders



Clustering: *Hierarchical*



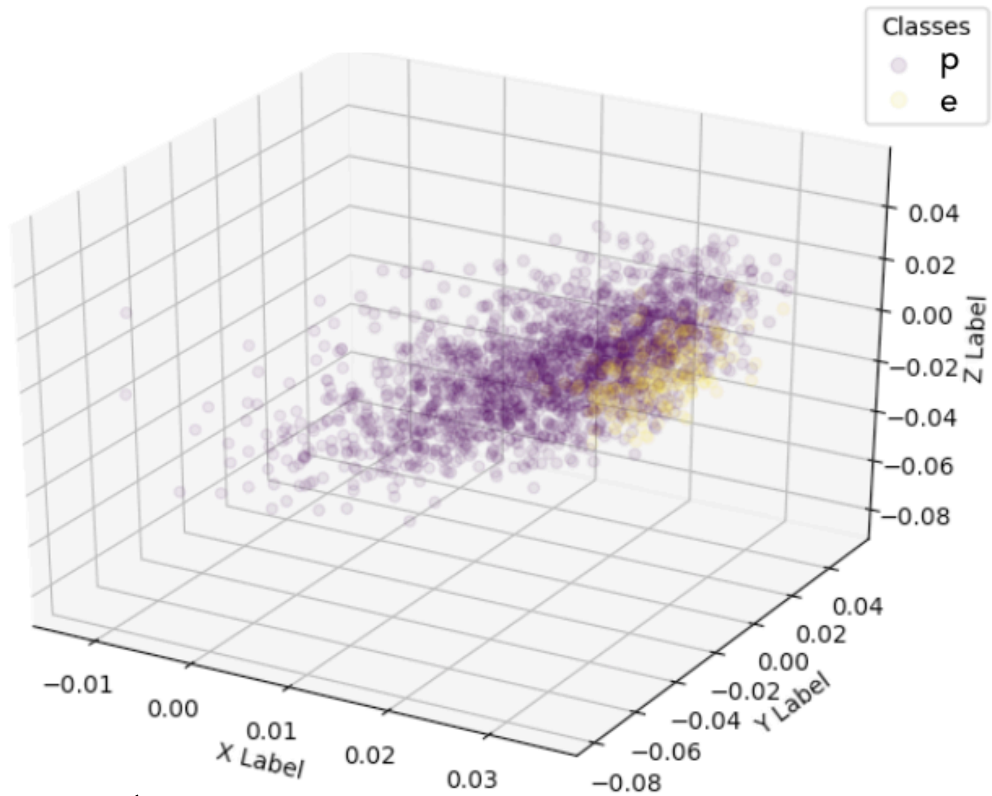
Actual Electrons to Protons Rate



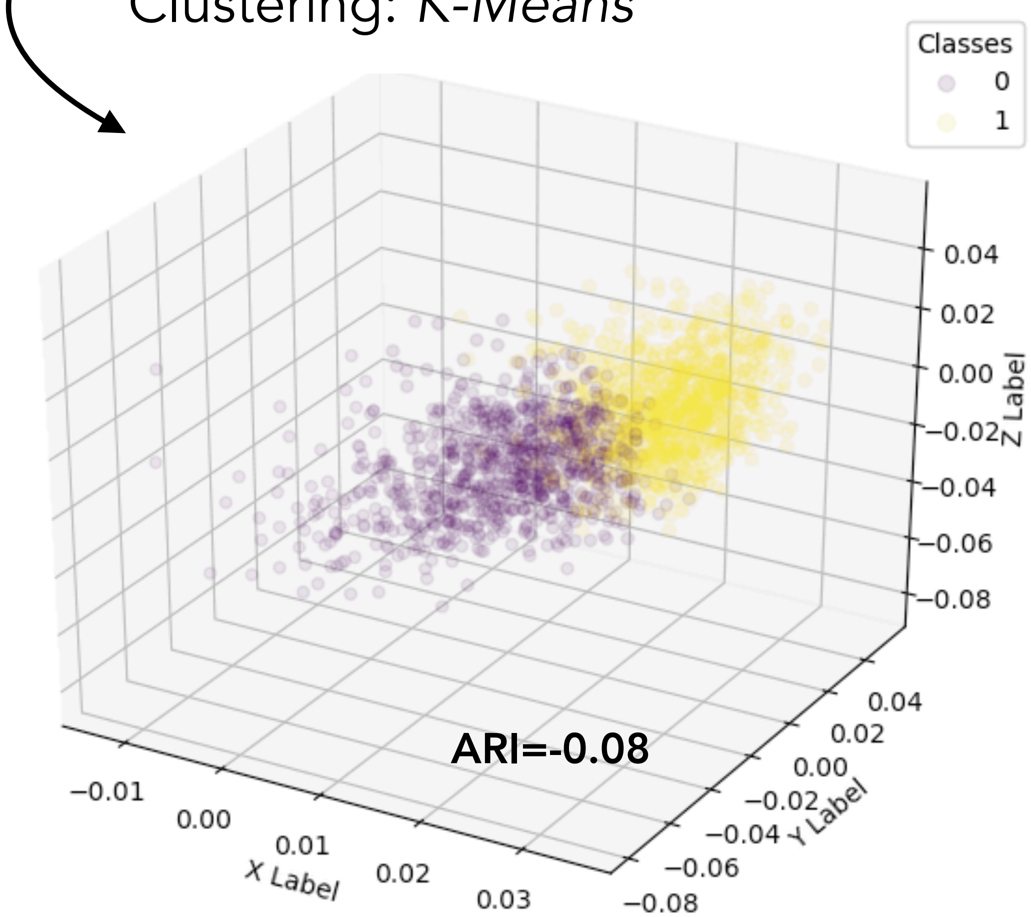
Interval of interest

Results for e/p=1/10

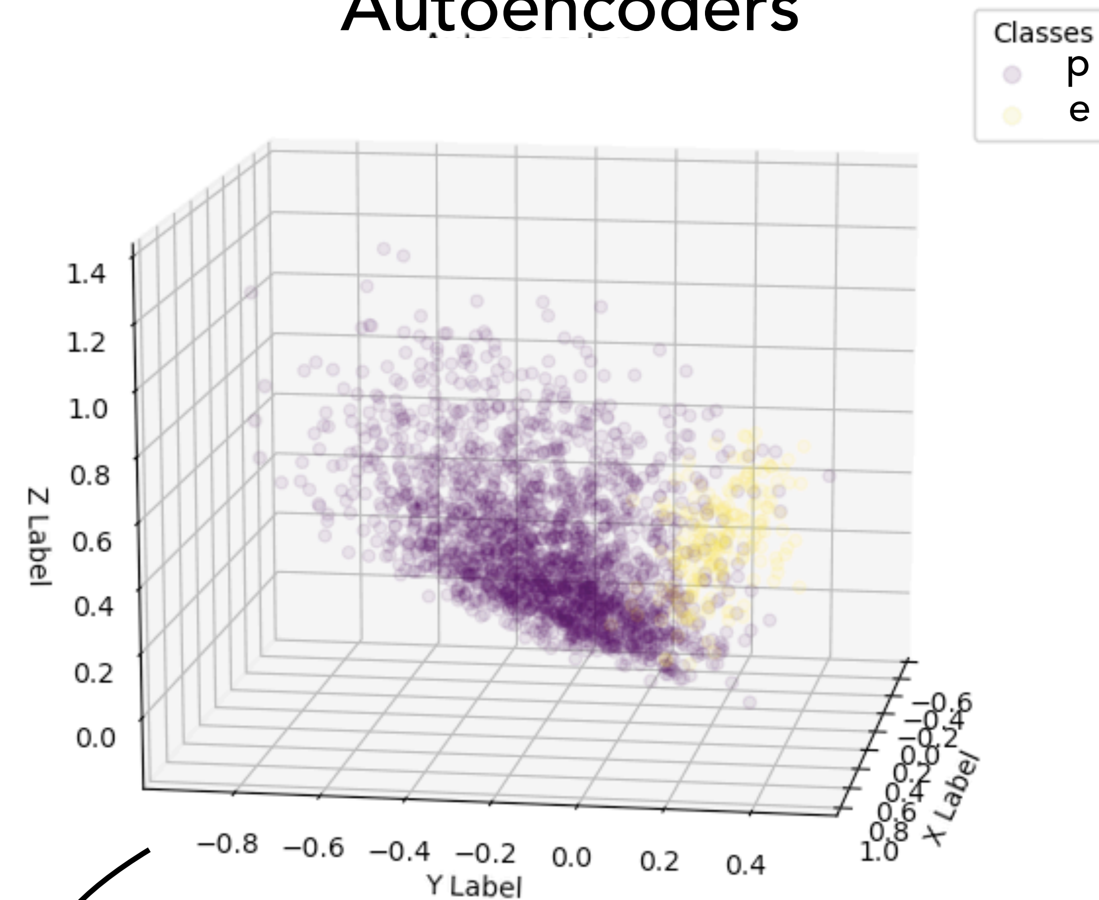
Principal Component Analysis



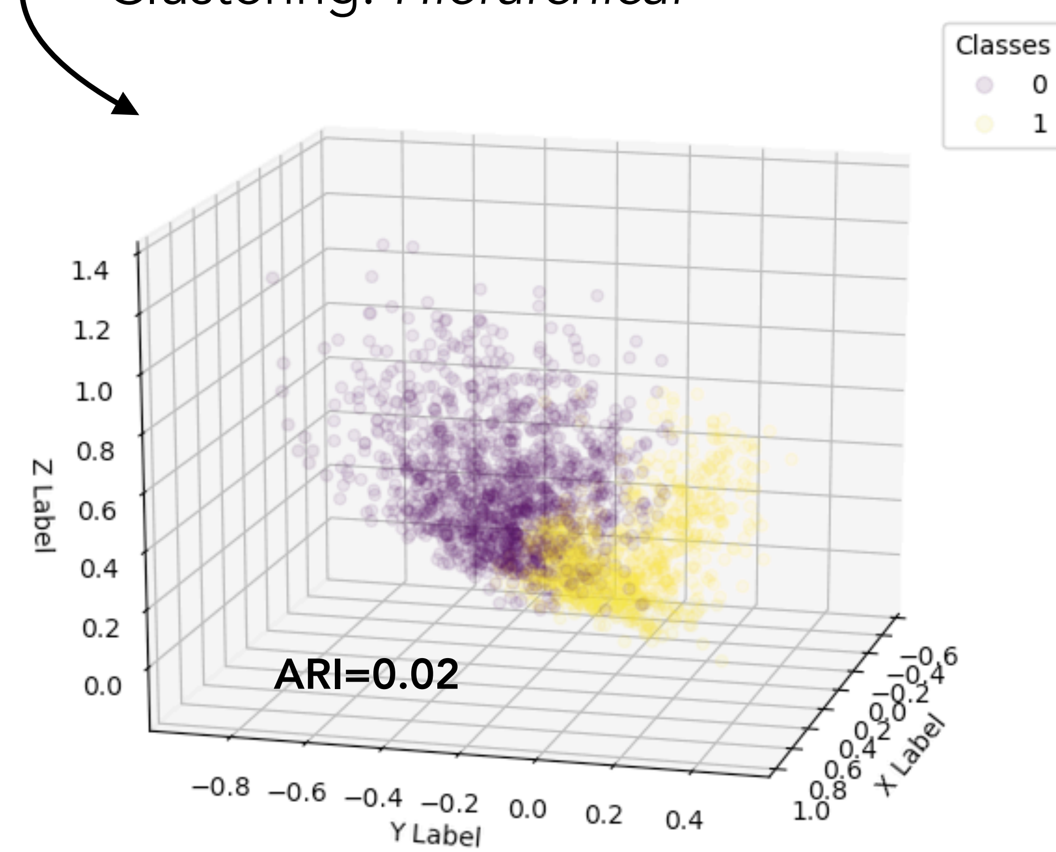
Clustering: *K-Means*



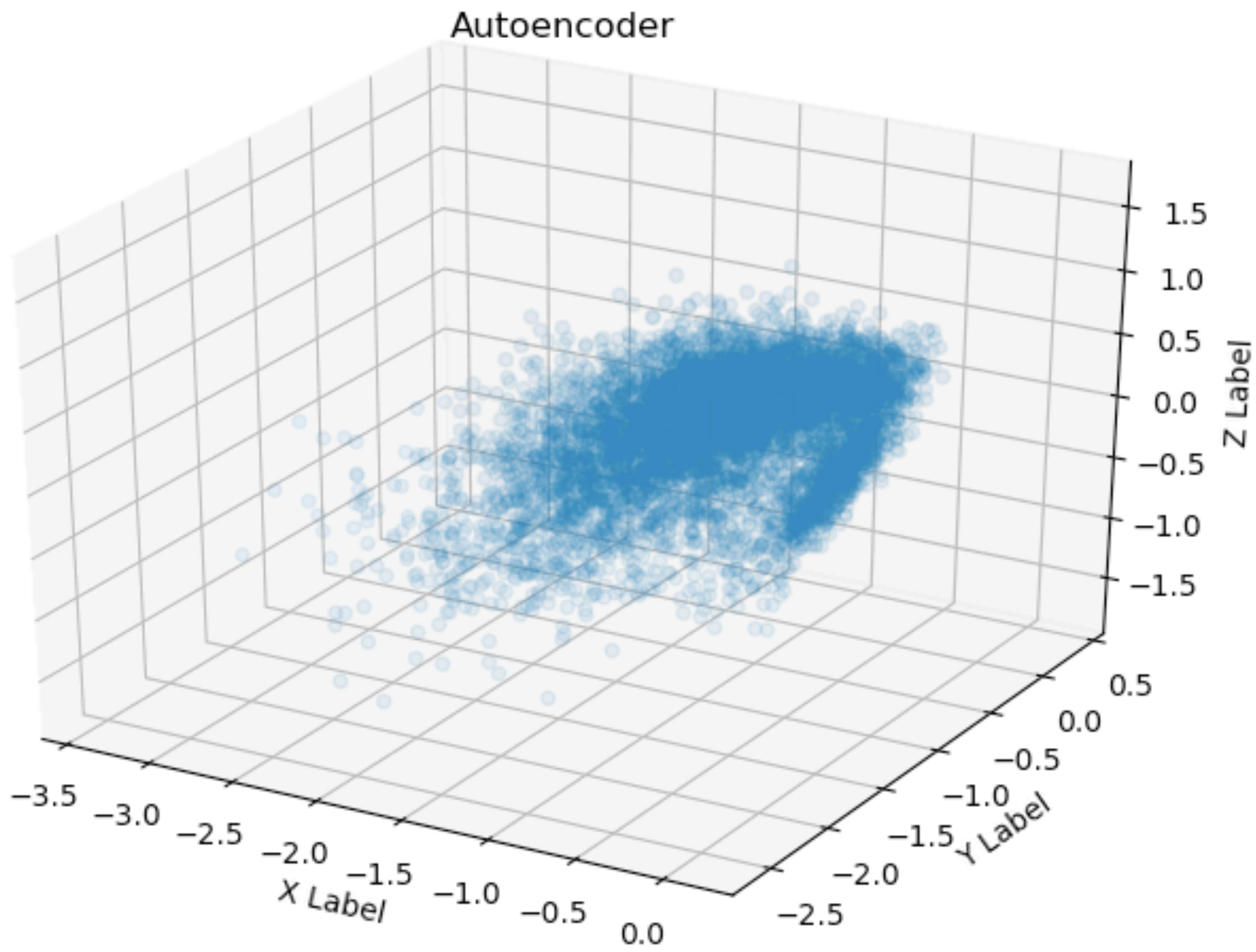
Autoencoders



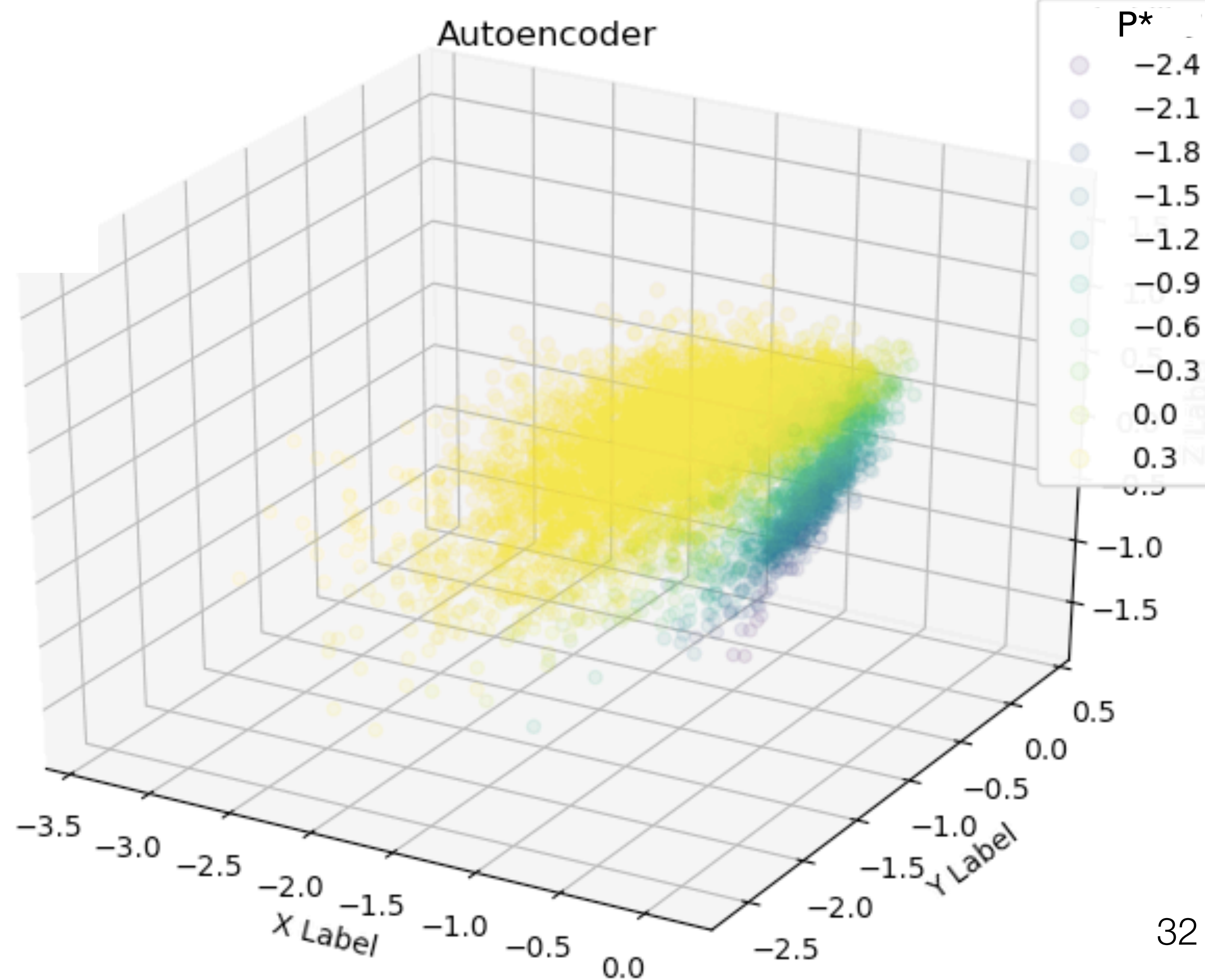
Clustering: *Hierarchical*



Autoencoder with experimental data



Through NN3



Conclusions

Neural Networks

First application of NN to Fermi cosmic-rays data. Preliminary results are promising, but some adjustments are still required.

Unsupervised Learning

Two steps needed:

- Reducing the number of protons in the experimental datasets with specific cuts.
- Developing a new algorithm which can separate the clusters with different sizes.

Backup Slides

Dataset Composition

Cuts applied to select only electrons and protons

Trigger Filter

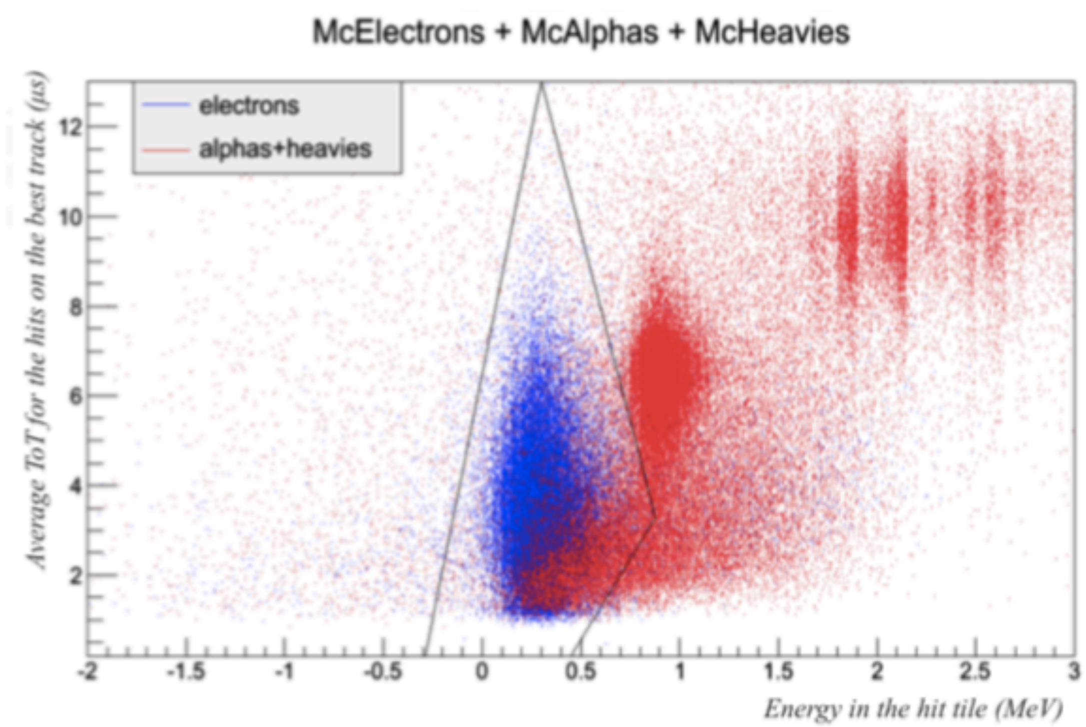
Selection of events that trigger the LAT and pass the on-board gamma.

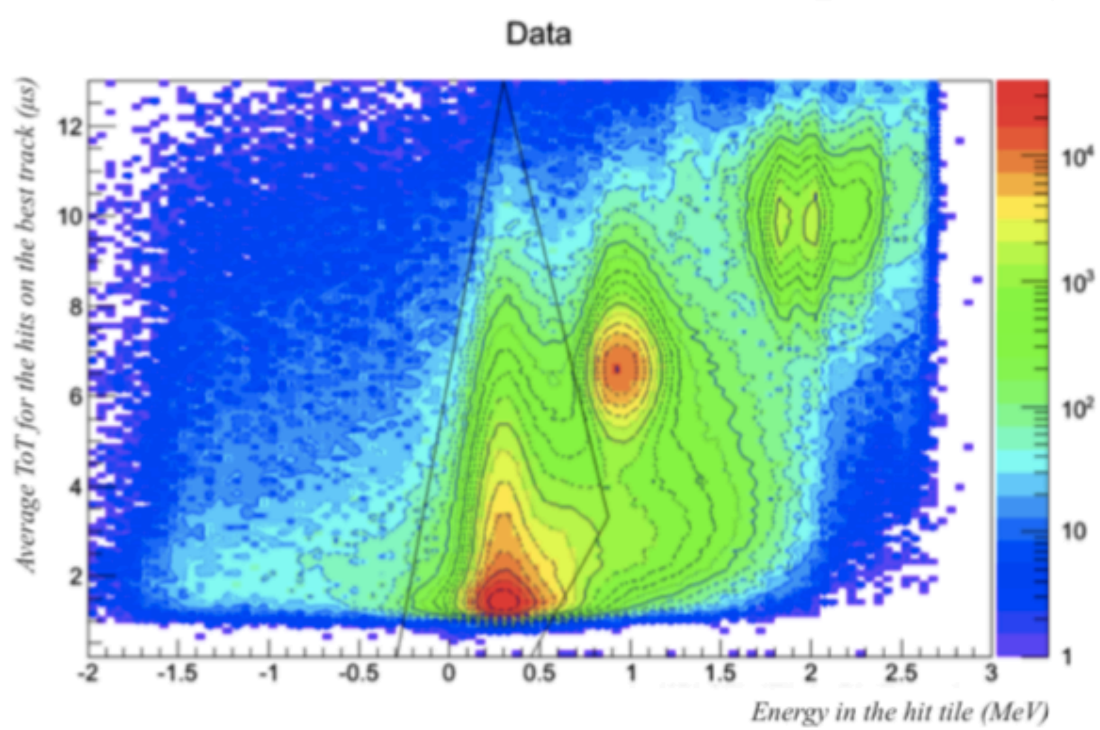
Quality Cut

Selection of events with at least a reconstructed track and a minimal PSF quality.

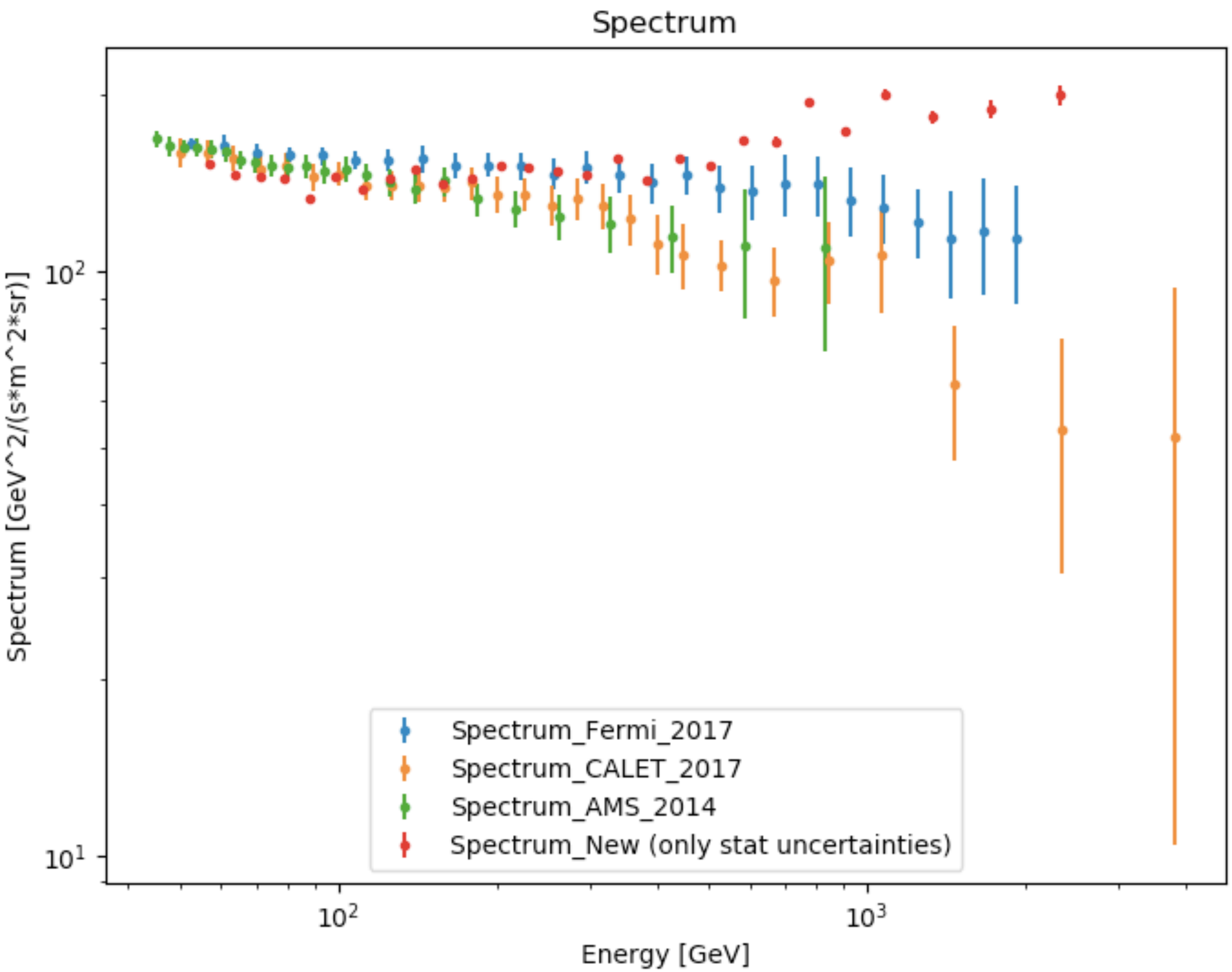
Alpha Cut

Cut removing α and heavies -> α/p =0.003 .



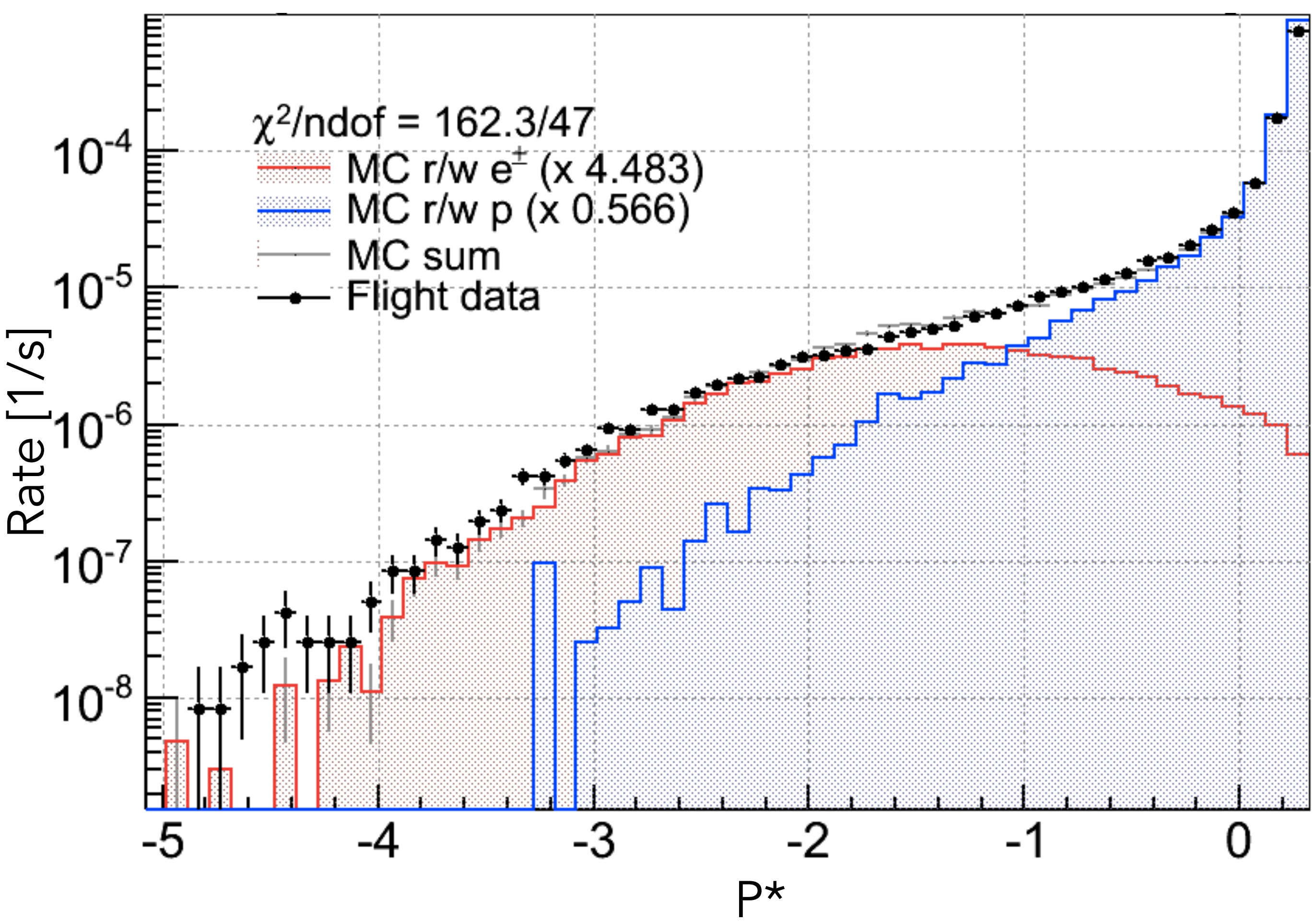


Spectrum E3 NN3



Preliminary comparison

New Analysis



2017 Analysis