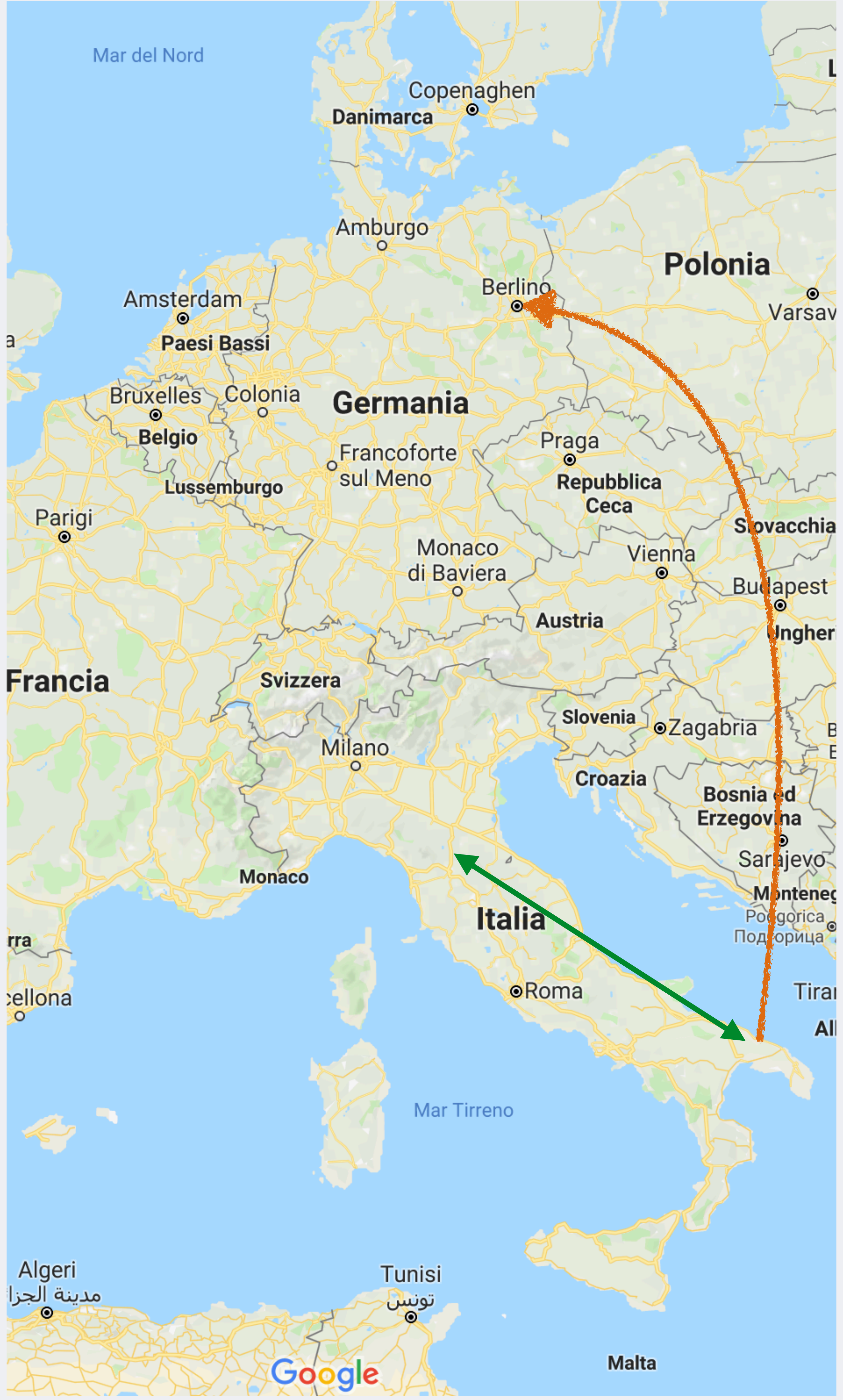
MACHINE LEARNING IN DAMPE

SIMONE GARRAPPA

XVI SEMINAR ON SOFTWARE FOR NUCLEAR, SUBNUCLEAR AND APPLIED PHYSICS

ALGHERO (ITALY), MAY 30 2019

INTRODUCING MYSELF



Now: Hunting cosmic neutrino sources with IceCube and Fermi-LAT @DESY Zeuthen

Before: Particle identification and gamma-ray astronomy with DAMPE @ University of Bari and University of Perugia

OUTLINE

PARTICLE IDENTIFICATION IN A SPACE-BASED PARTICLE PHYSICS EXPERIMENT

- The Dark Matter Particle Explorer - Mission and scientific goals
- The DAMPE detector
- Deep learning and Machine learning for gamma-ray selection in DAMPE:
  - Neural networks for pattern recognition
  - Tree-based classifiers for multivariate analysis

THE DAMPE COLLABORATION

- CHINA
  - –Purple Mountain Observatory, CAS, Nanjing
  - –Institute of High Energy Physics, CAS, Beijing
  - –National Space Science Center, CAS, Beijing
  - –University of Science and Technology of China, Hefei
  - –Institute of Modern Physics, CAS, Lanzhou

- ITALY
  - –INFN Perugia and University of Perugia
  - –INFN Bari and University of Bari
  - –INFN Lecce and University of Salento
  - –GSSI Gran Sasso Science Institute

- SWITZERLAND
  - –University of Geneva

Thanks to the DAMPE Collaboration for supporting this talk and for allowing the presentation of these results!

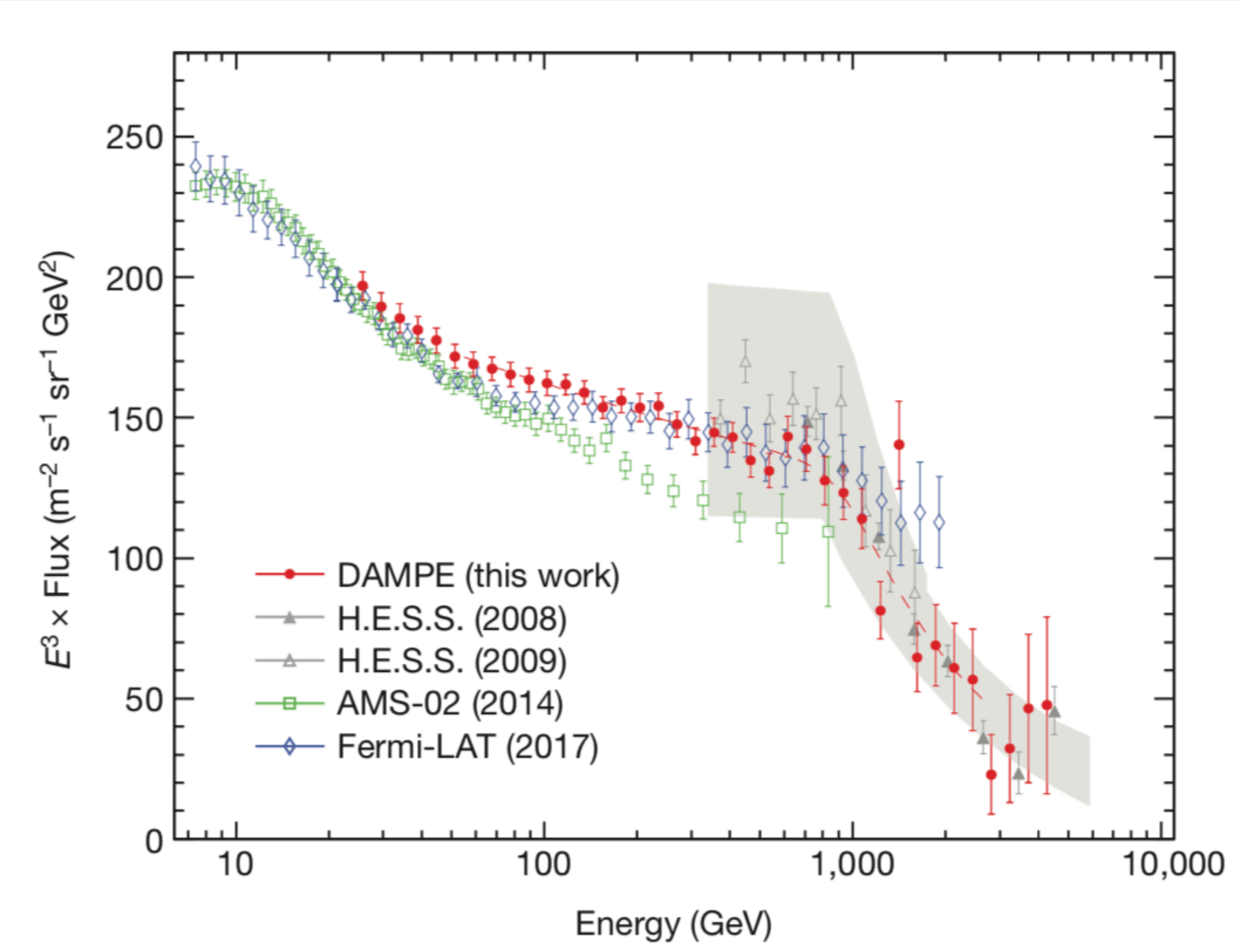
THE DAMPE MISSION

- Satellite launched on Dec. 17, 2015 from the Jiuquan Satellite Center (China)
  - Total payload: 1900 kg
  - Detector payload: 1300 kg
- Polar Sun-synchronous orbit:
  - Altitude: 500 km
  - Inclination: 97.4°
  - Period: 95 min
- Satellite renamed "Wukong" after launch

PHYSICS GOALS

HIGH ENERGY PARTICLE DETECTION IN SPACE

- Study of the cosmic-ray electrons and positrons spectrum
- Study of cosmic ray protons and nuclei:
  - spectrum and composition
- High energy gamma-ray astronomy and photon spectra
- Search for dark matter signatures in lepton and photon spectra



LETTER

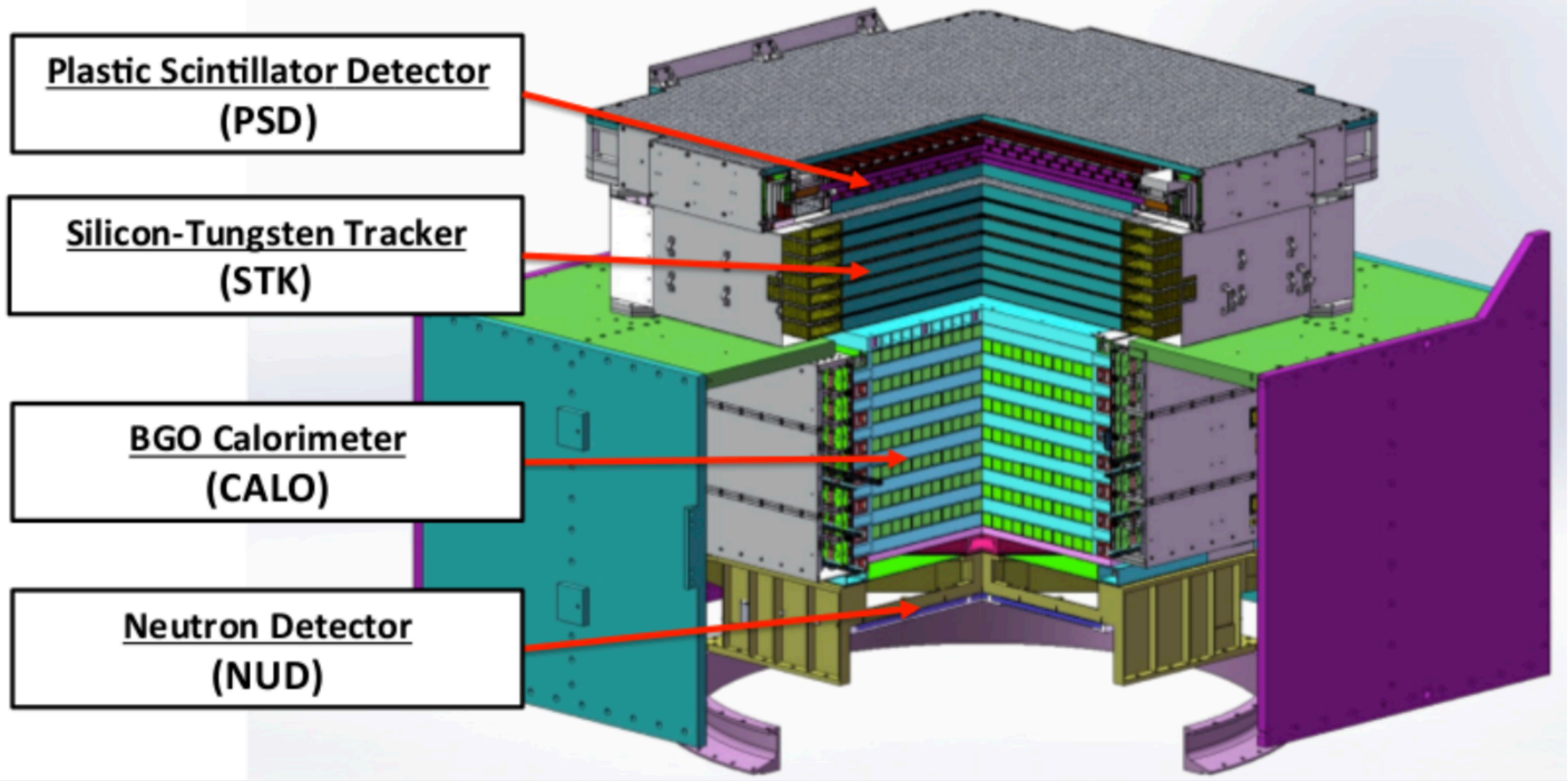
doi:10.1038/nature24475

Direct detection of a break in the teraelectronvolt cosmic-ray spectrum of electrons and positrons

DAMPE Collaboration*

INSTRUMENT DESIGN

- Measuring VHE cosmic-rays
  - e^+/e^- and gamma-rays in the range 1 GeV - 10 TeV
  - Cosmic-ray nuclei in the range 50 GeV - 100 TeV
- Goals:
  - Search for dark matter signals with e^+/e^- and gamma-rays
  - Study of cosmic-ray spectra and composition
  - HE gamma-ray astronomy

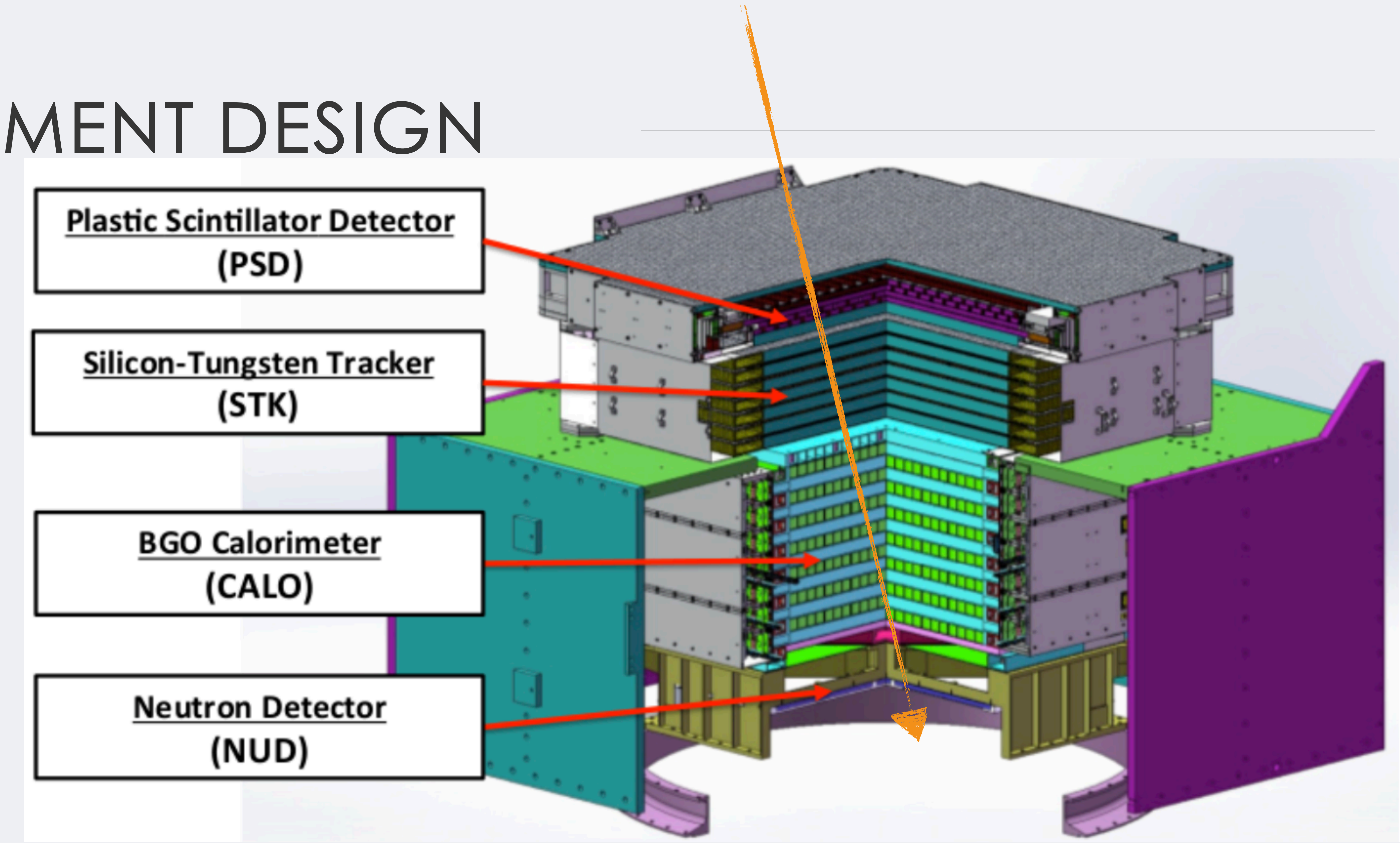


	DAMPE	AMS-02	Fermi LAT
e/γ Energy res.@100 GeV (%)	<1.5	3	10
e/γ Angular res.@100 GeV (deg.)	<0.2	0.3	0.1
e/p discrimination	$>10^5$	10^5 - 10^6	10^3
Calorimeter thickness (X_0)	32	17	8.6
Geometrical accep. (m^2sr)	0.3	0.09	1

INSTRUMENT DESIGN

- Measuring VHE cosmic-rays
  - e^+/e^- and gamma-rays in the range 1 GeV - 10 TeV
  - Cosmic-ray nuclei in the range 50 GeV - 100 TeV

- Goals:
  - Search for dark matter signals with e^+/e^- and gamma-rays
  - Study of cosmic-ray spectra and composition
  - HE gamma-ray astronomy

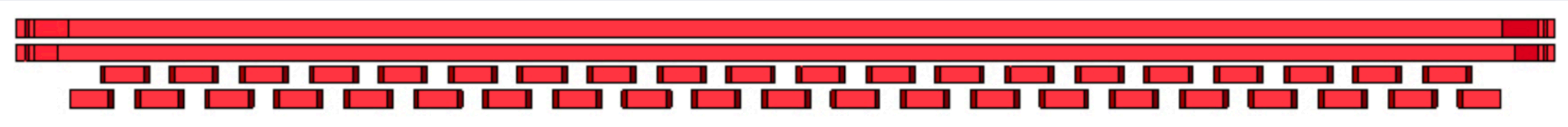


	DAMPE	AMS-02	Fermi LAT
e/γ Energy res.@100 GeV (%)	<1.5	3	10
e/γ Angular res.@100 GeV (deg.)	<0.2	0.3	0.1
e/p discrimination	$>10^5$	10^5 - 10^6	10^3
Calorimeter thickness (X_0)	32	17	8.6
Geometrical accep. (m^2sr)	0.3	0.09	1

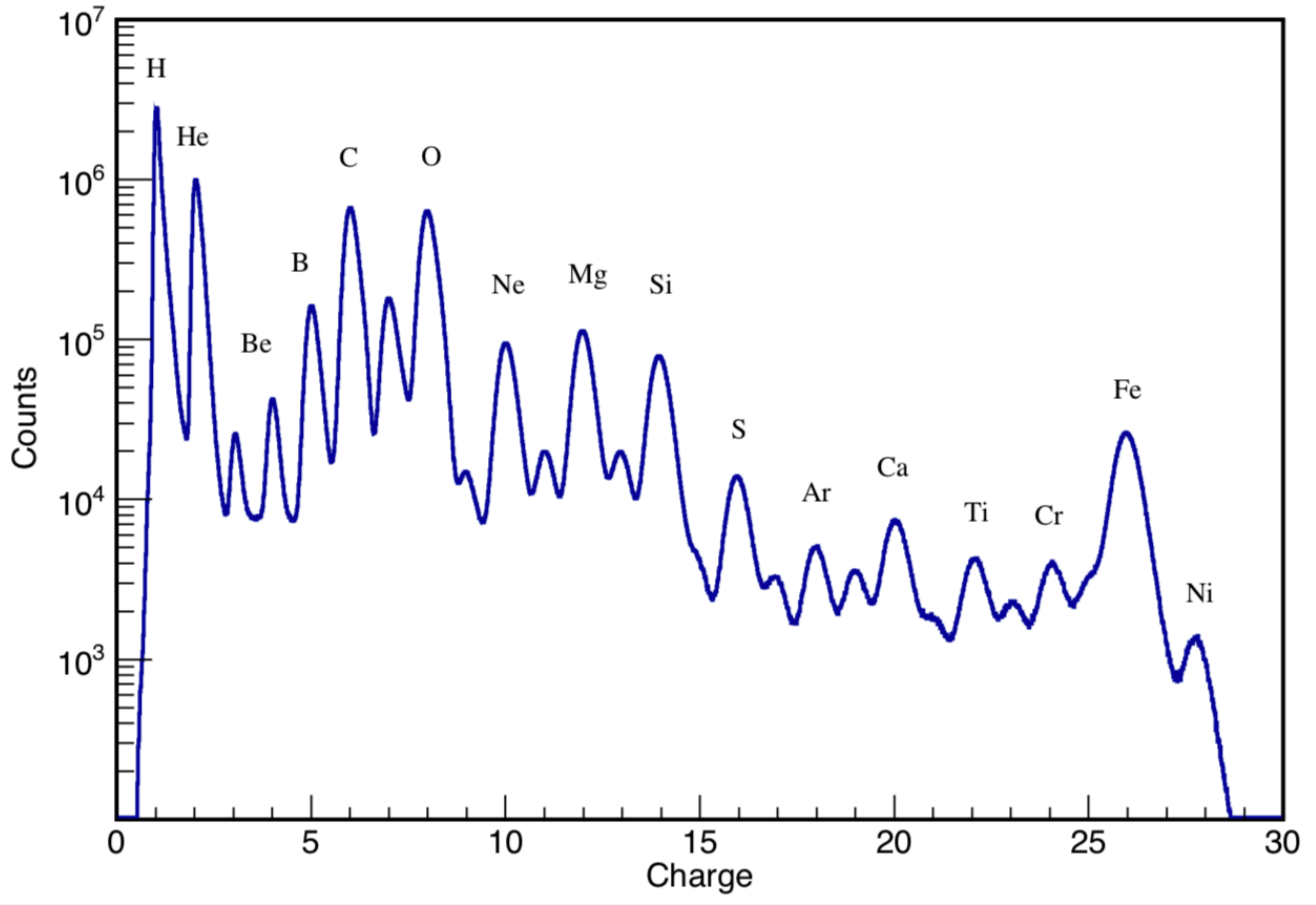
PLASTIC SCINTILLATOR DETECTOR (PSD)

- Dual role:
  - Anti-coincidence detector
  - Measure the charge of the incident particles

- Geometry and performance:
  - 2 layers, 82 bars (XZ and YZ view)
  - Active area: 82cm x 82cm
  - Efficiency single module ⩾0.95 for MIPs
  - Position resolution ⩽ 2cm
  - Charge resolution:
    - 13.7% (Z = 1)
    - 30% (Z = 26)



https://arxiv.org/pdf/1810.10784.pdf

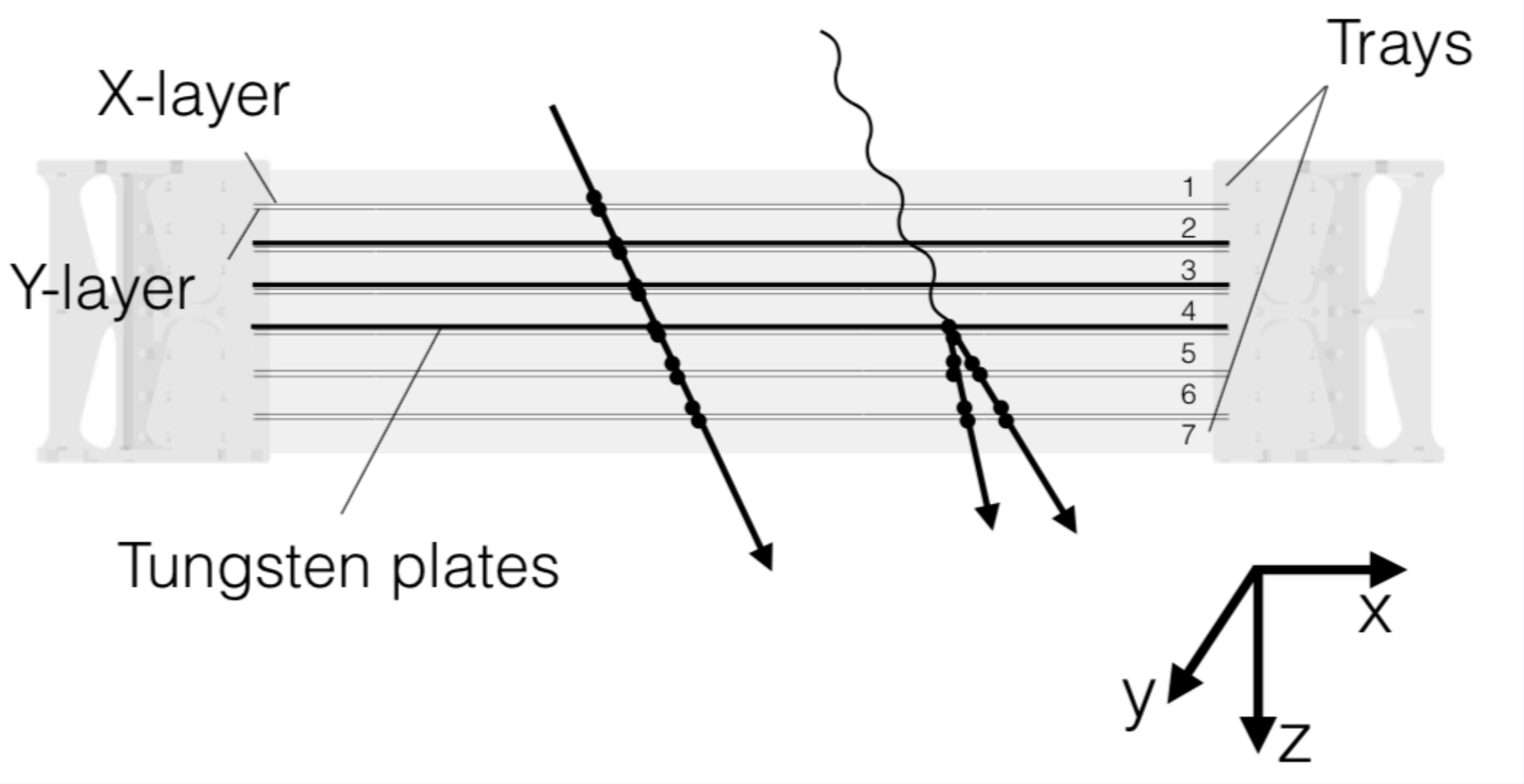


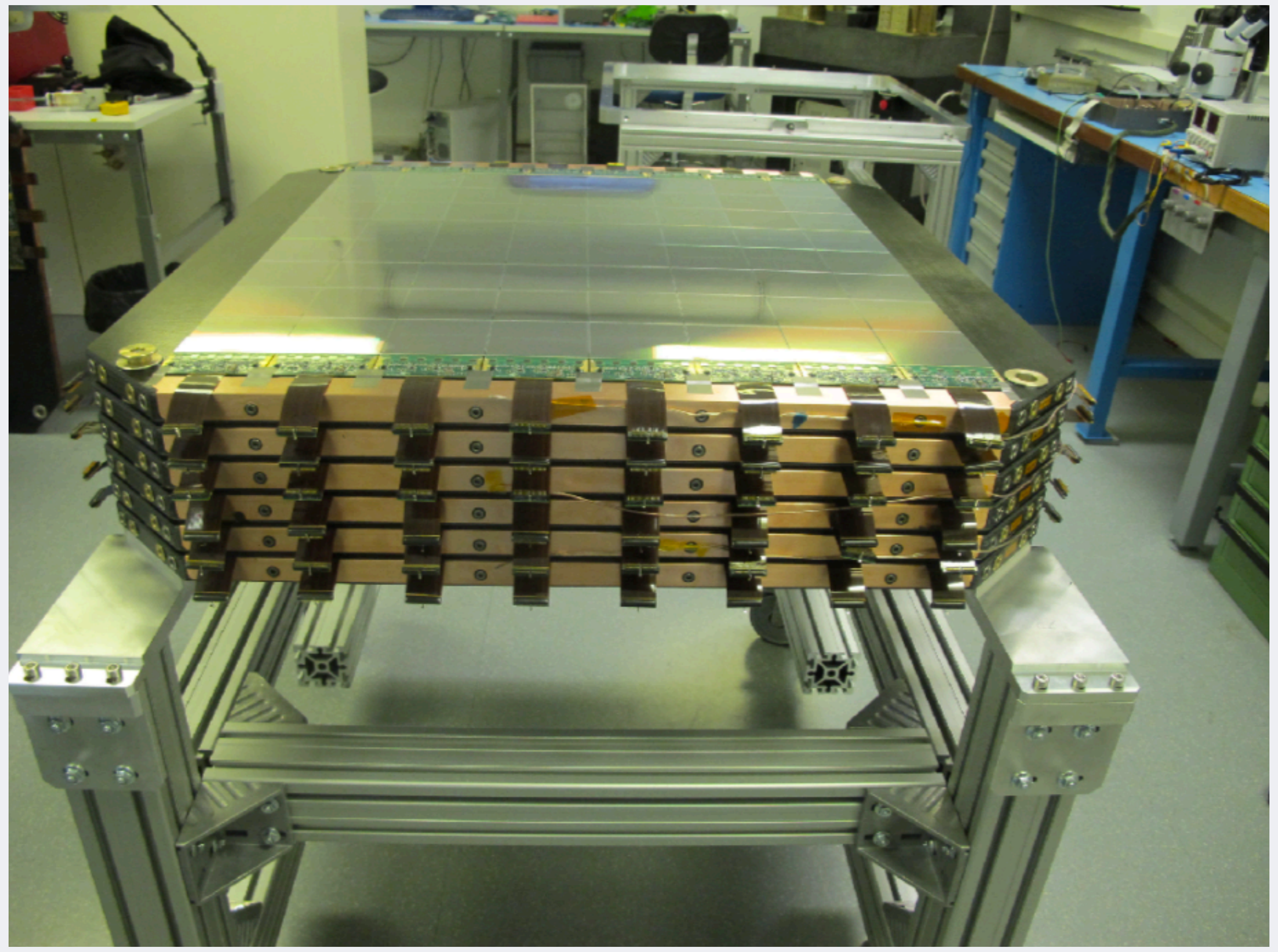
SILICON TRACKER

https://arxiv.org/pdf/1806.10355.pdf

- Principal tasks:
  - Precise track reconstruction
  - Measure the charge of the incident particles
  - Photon conversion to e^+/e^- pairs

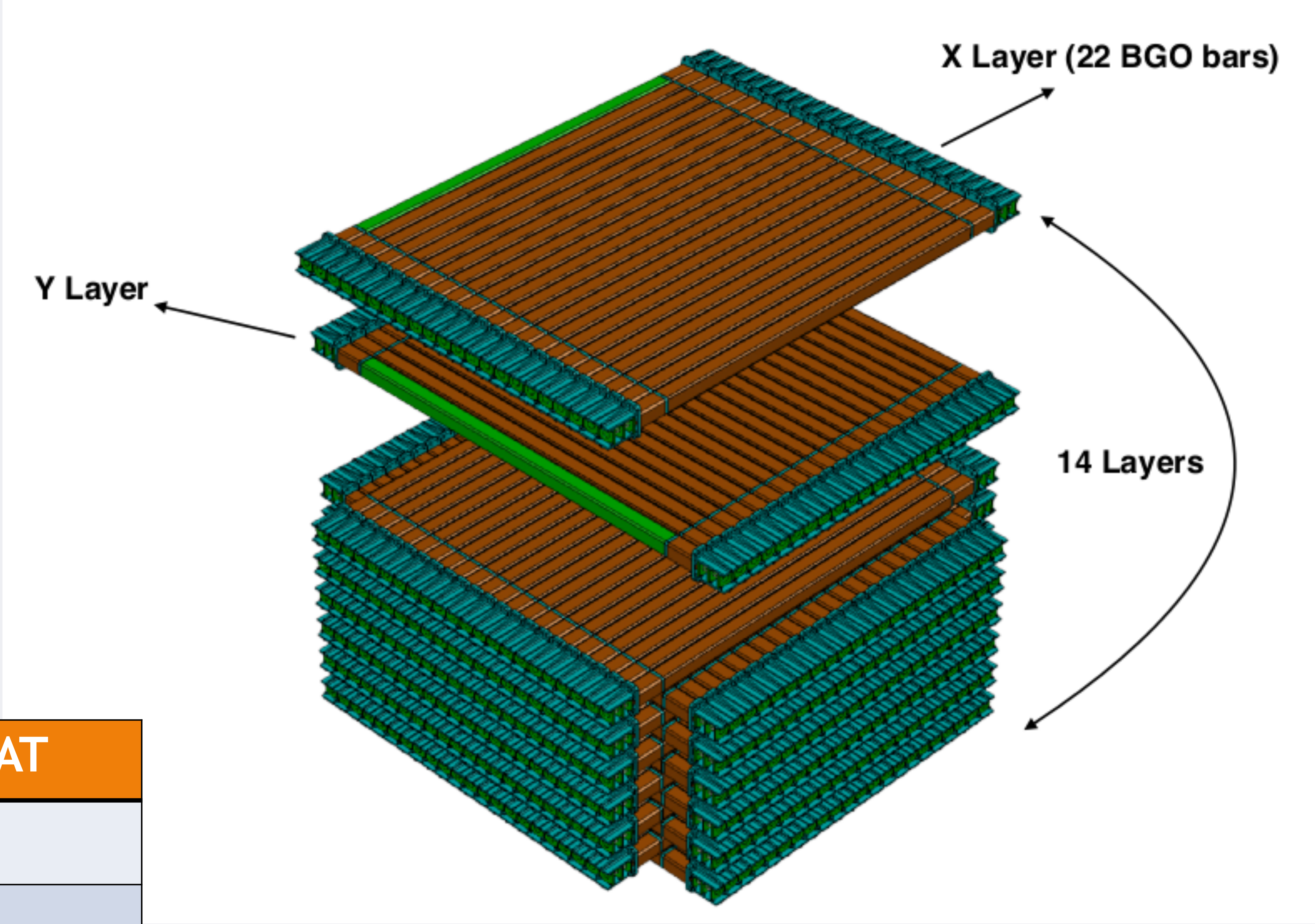
- Geometry and performance:
  - 6 tracking planes, double layer (XZ and YZ)
  - 3 Tungsten plates (total of ~1 X_0)
  - Active area: 0.55 m^2 x 12 layers
  - Spatial resolution ⩽80 μm





BGO CALORIMETER

- 308 $Bi_4Ge_3O_{12}$ crystal bars
  - Size of single bar: 2.5cm x 2.5cm x 60cm
  - Bar cross-section 60cm x 60cm
  - 14 layers of 22 bars
  - Alternate XZ and YZ arrangement
- Thickness of almost 32 X_0
  - Energy resolution (e/γ) @100 GeV: < 1.5%

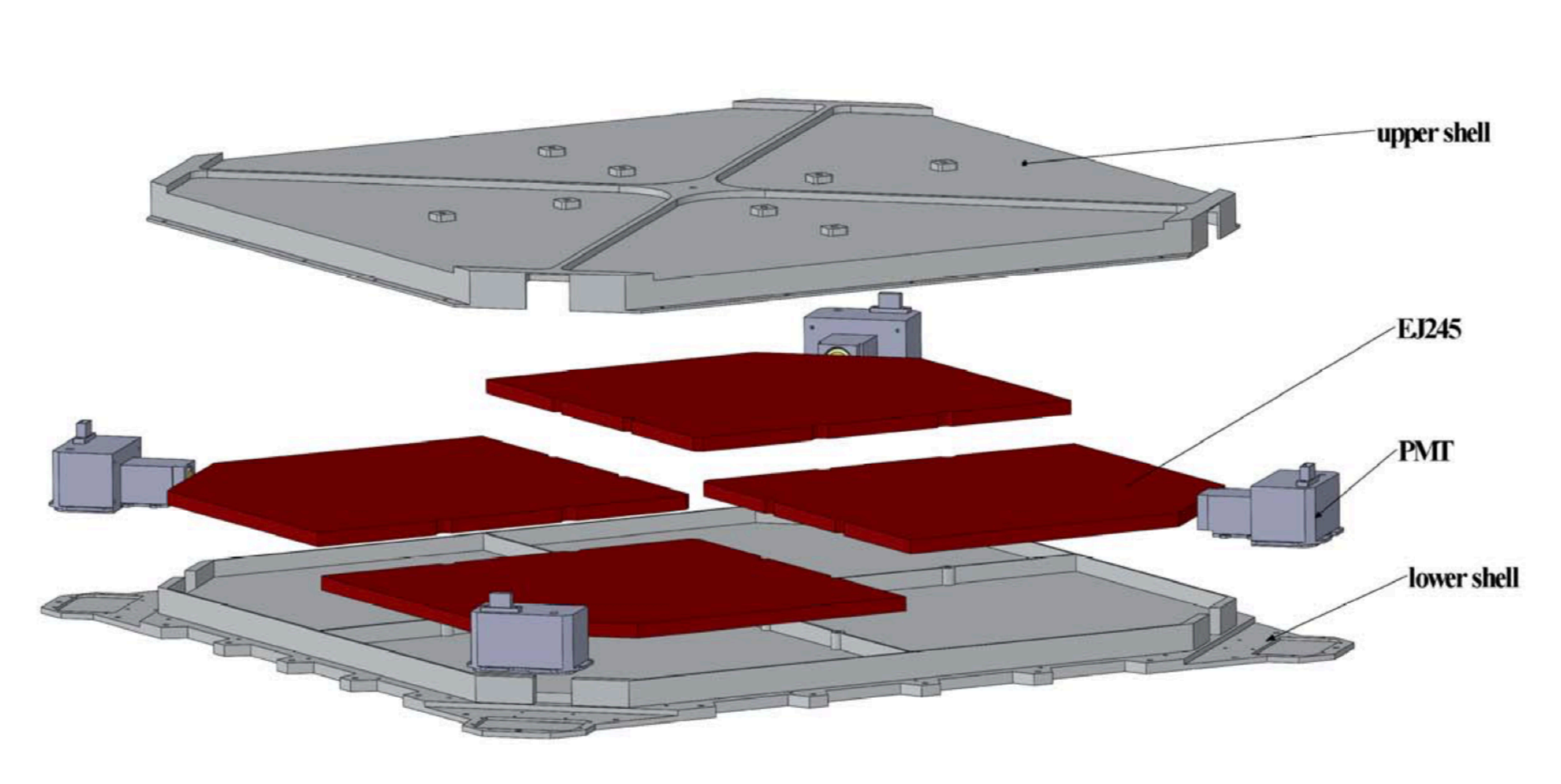


	DAMPE	AMS-02	Fermi LAT
e/γ Energy res.@100 GeV (%)	<1.5	3	10
e/γ Angular res.@100 GeV (deg.)	<0.2	0.3	0.1
e/p discrimination	$>10^5$	10^5 - 10^6	10^3
Calorimeter thickness (X_0)	32	17	8.6
Geometrical accep. (m^2sr)	0.3	0.09	1

NEUTRON DETECTOR

- Principal task:
  - Enhance e/p separation
- Geometry:
  - 4 Boron-doped plastic scintillator plates
  - Dimension (single plate): 19.5 cm x 19.5 cm x 1 cm
- Detection technique:
  - Neutrons that enter the scintillator undergo the capture

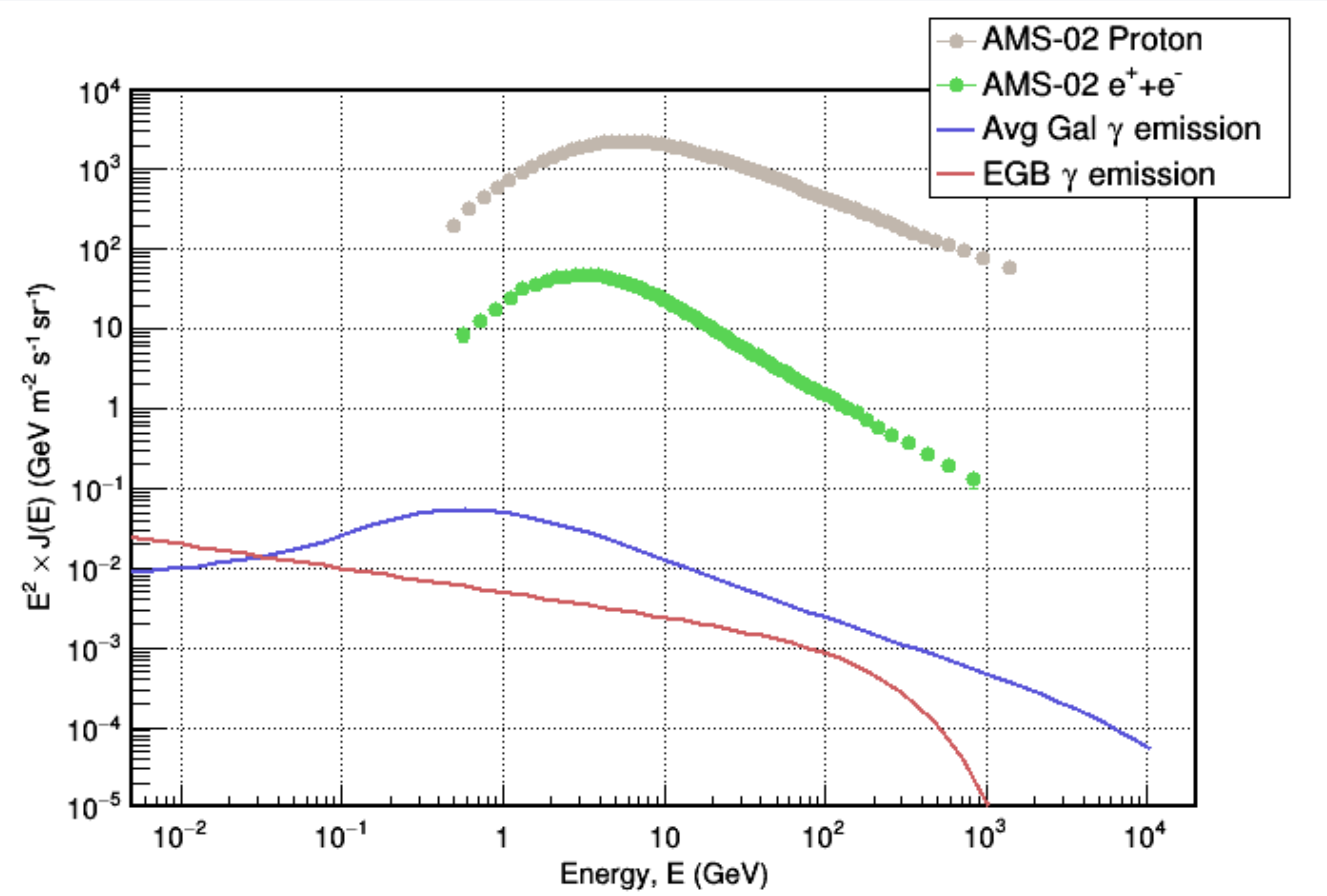
$$^{10}Be + n \rightarrow ^{7}Li + \alpha + \gamma$$



THE GAMMA-RAY SELECTION

CHALLENGES WITH AN HIGH COSMIC-RAY BACKGROUND

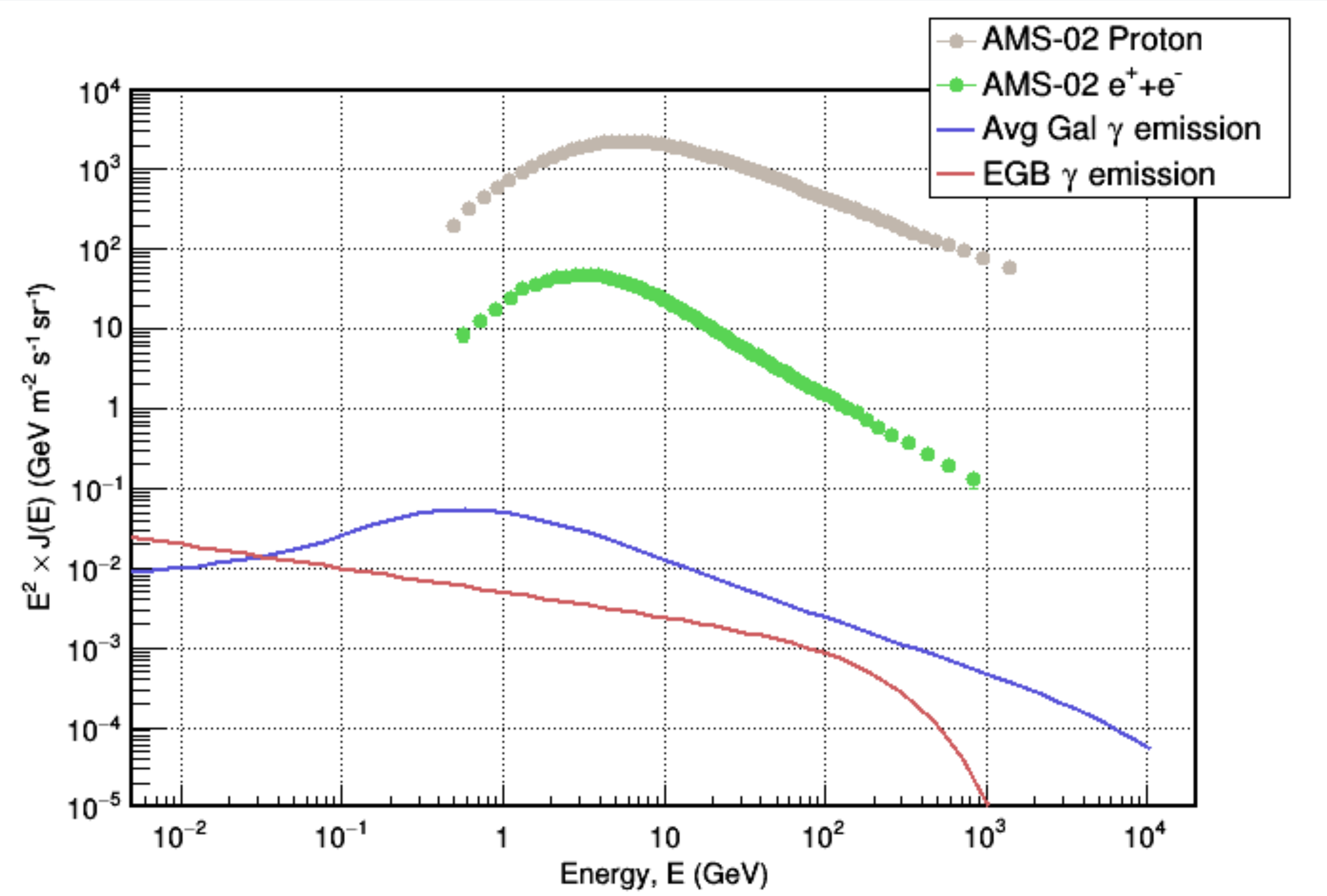
- The main sources of background are cosmic-ray protons and electrons:
  - Protons: 10^5 @ E > 100 GeV
  - Electrons: 10^3 @ E > 100 GeV
- How do we remove such components?
  - Shower profile
  - Interactions in all sub-detectors



THE GAMMA-RAY SELECTION

CHALLENGES WITH AN HIGH COSMIC-RAY BACKGROUND

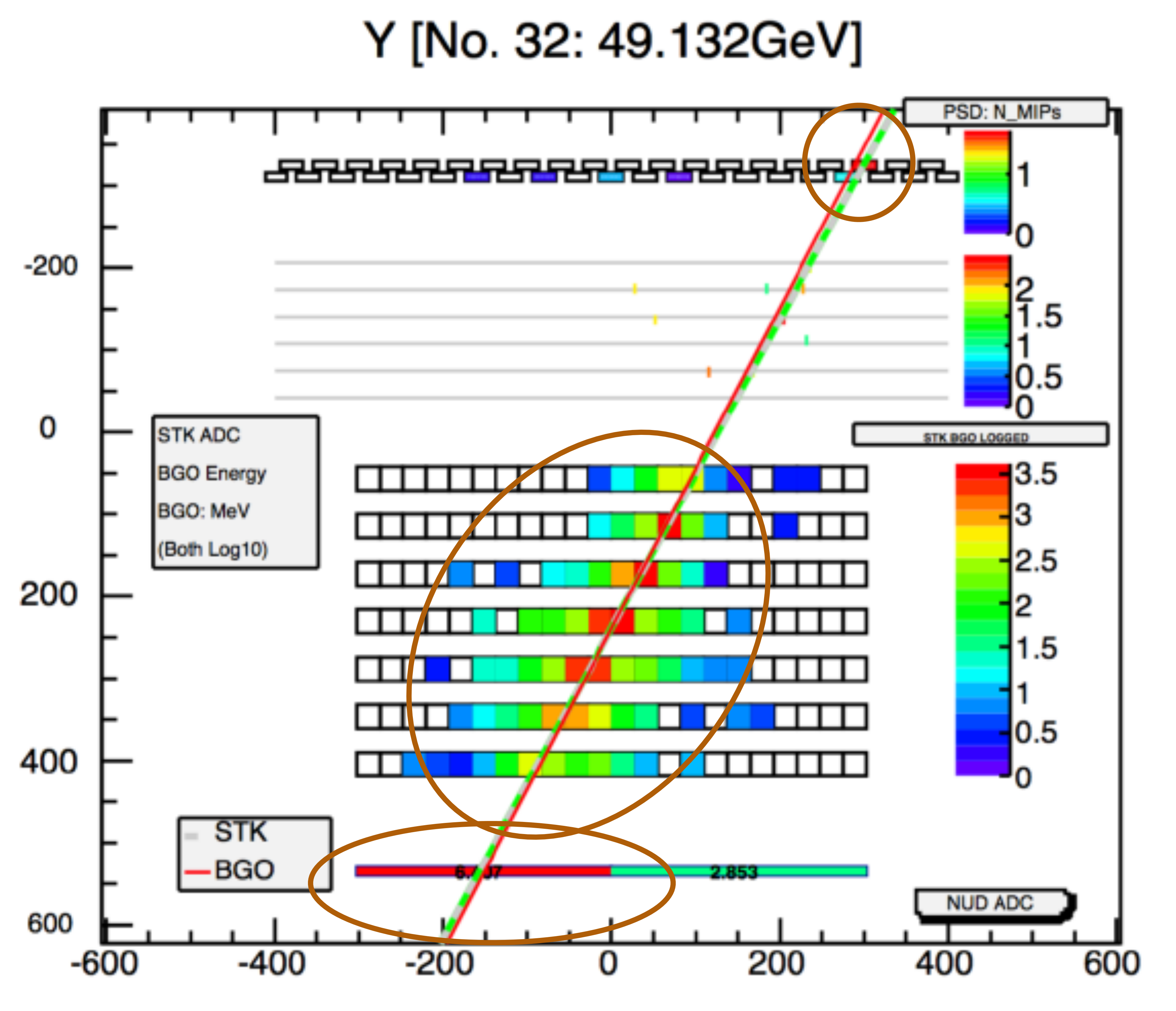
- Each sub-detector contributes in the PID
- **Pre-selection:** geometrical cuts
- **First step:** reject hadronic component (protons, heavier nuclei)
- **Second step:** reject charged component (electrons, remaining protons and nuclei)



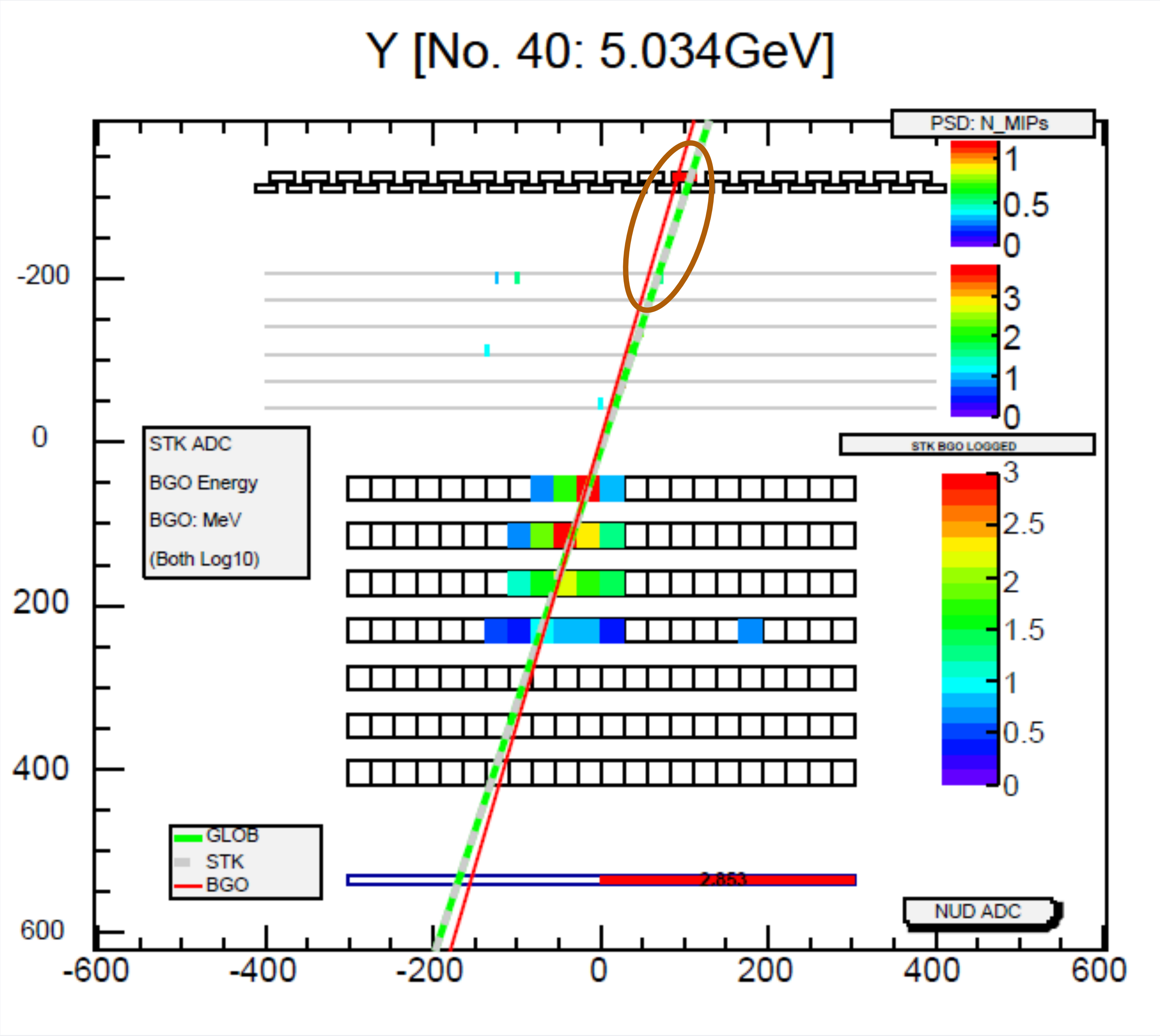
PARTICLE IDENTIFICATION

CHALLENGES WITH AN HIGH COSMIC-RAY BACKGROUND

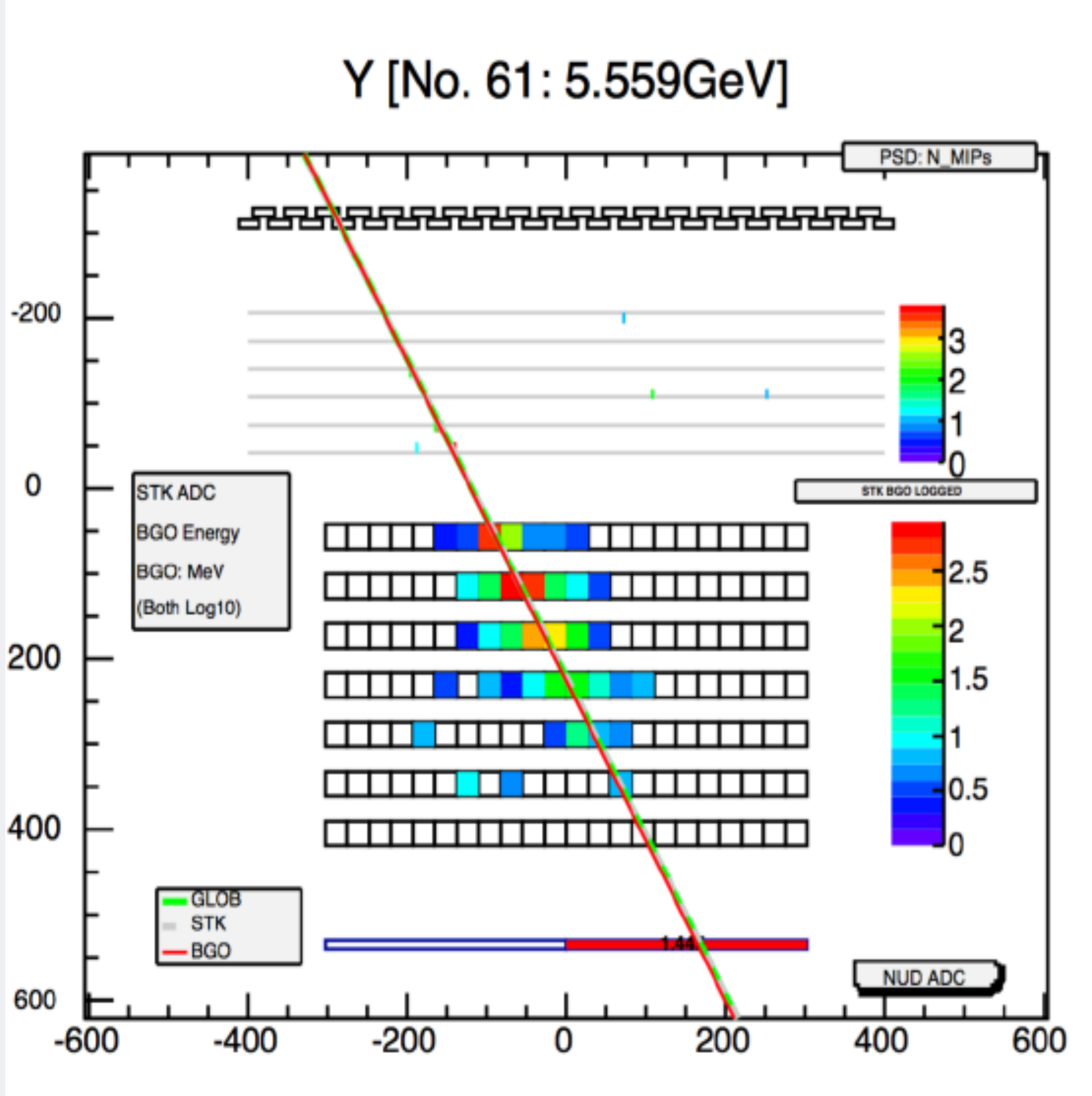
Proton



Electron



Gamma-ray

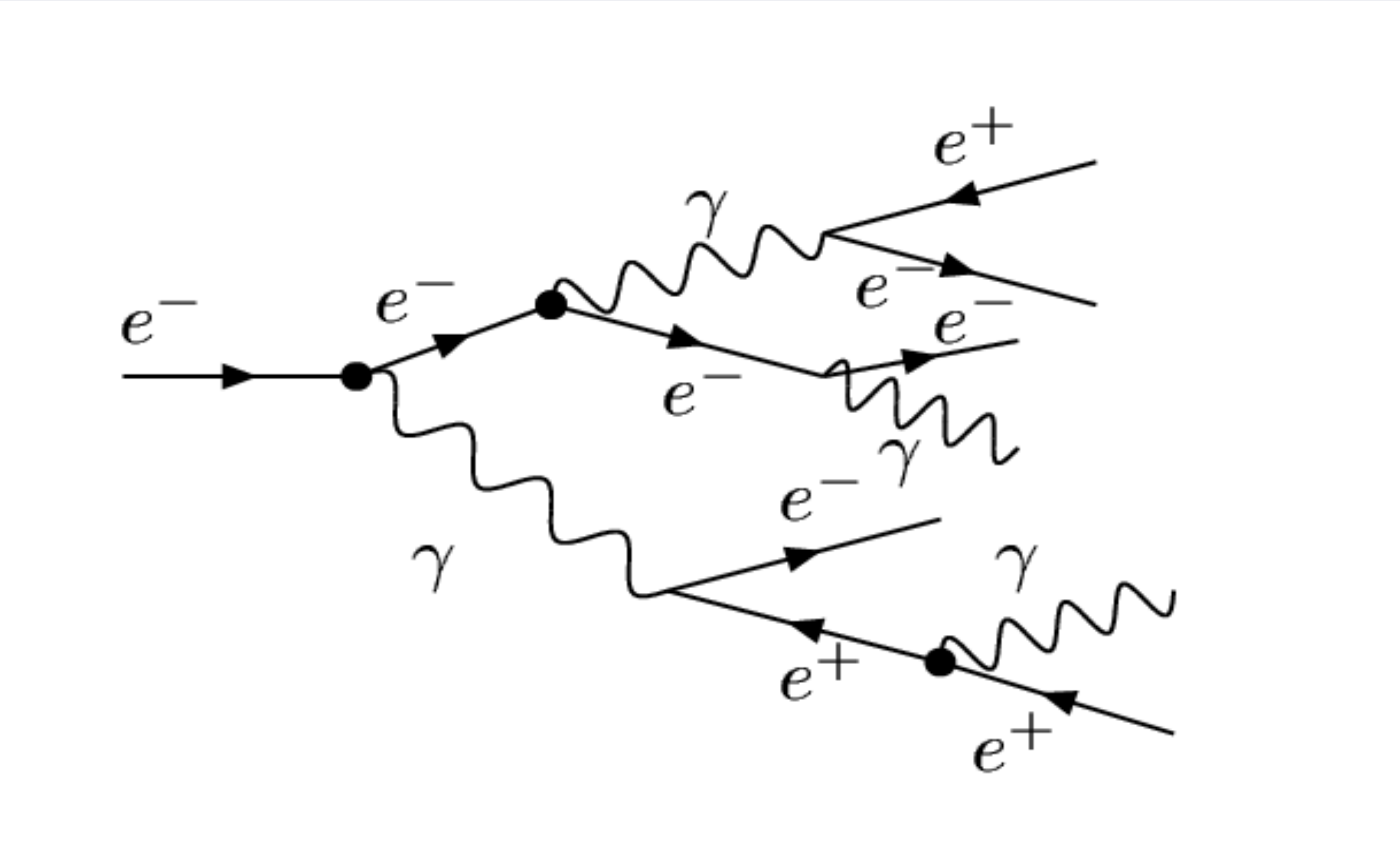


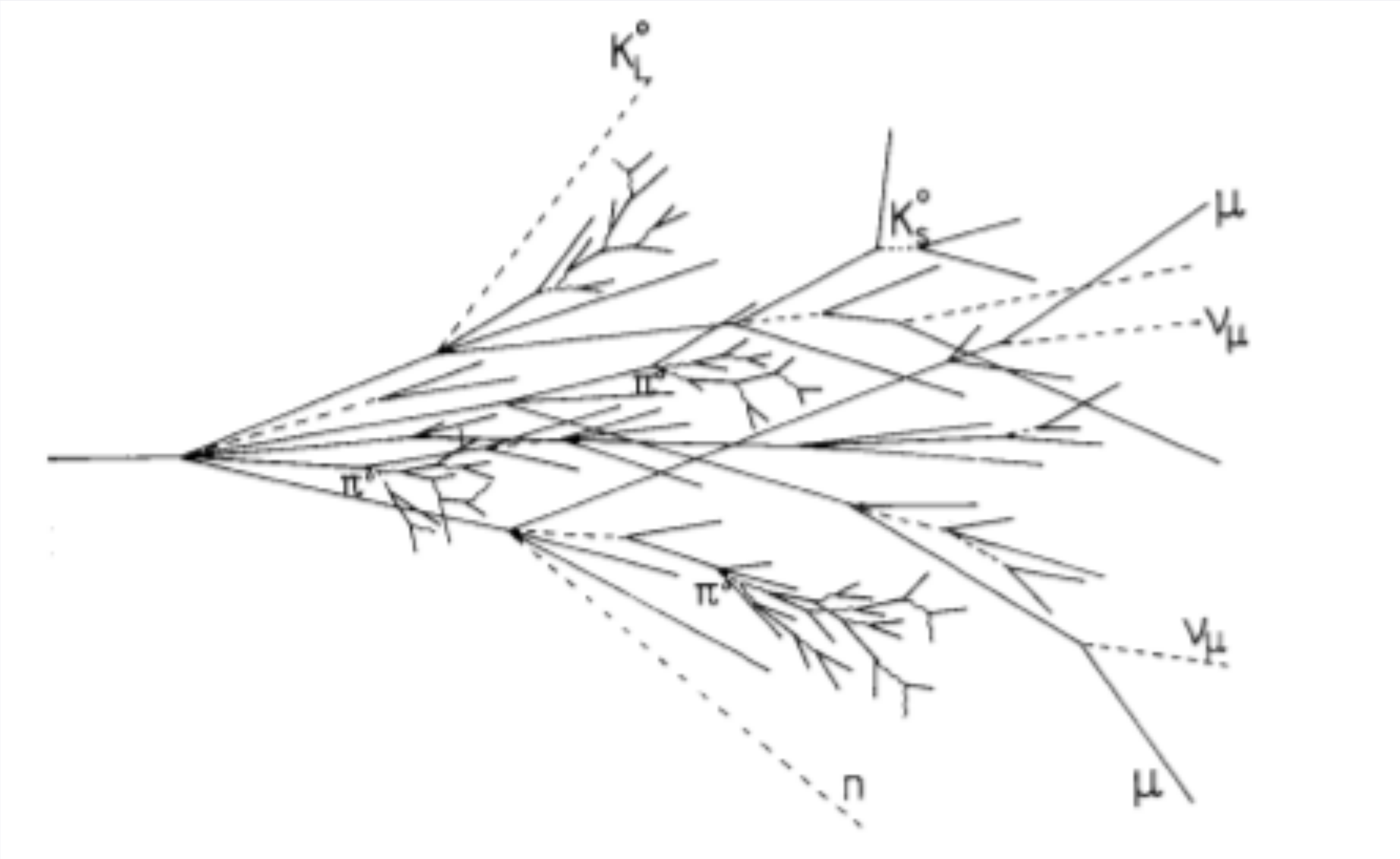
- Step 0: the reconstructed track must cross all sub-detectors

e/p separation

SEPARATE ELECTROMAGNETIC AND HADRONIC COMPONENTS USING SHOWER PROFILE

- Particle showers initiated by electrons (gamma) and protons (nuclei) involve different interactions with the medium
- Projections in the XZ and YZ planes give 2D sections of the shower profile

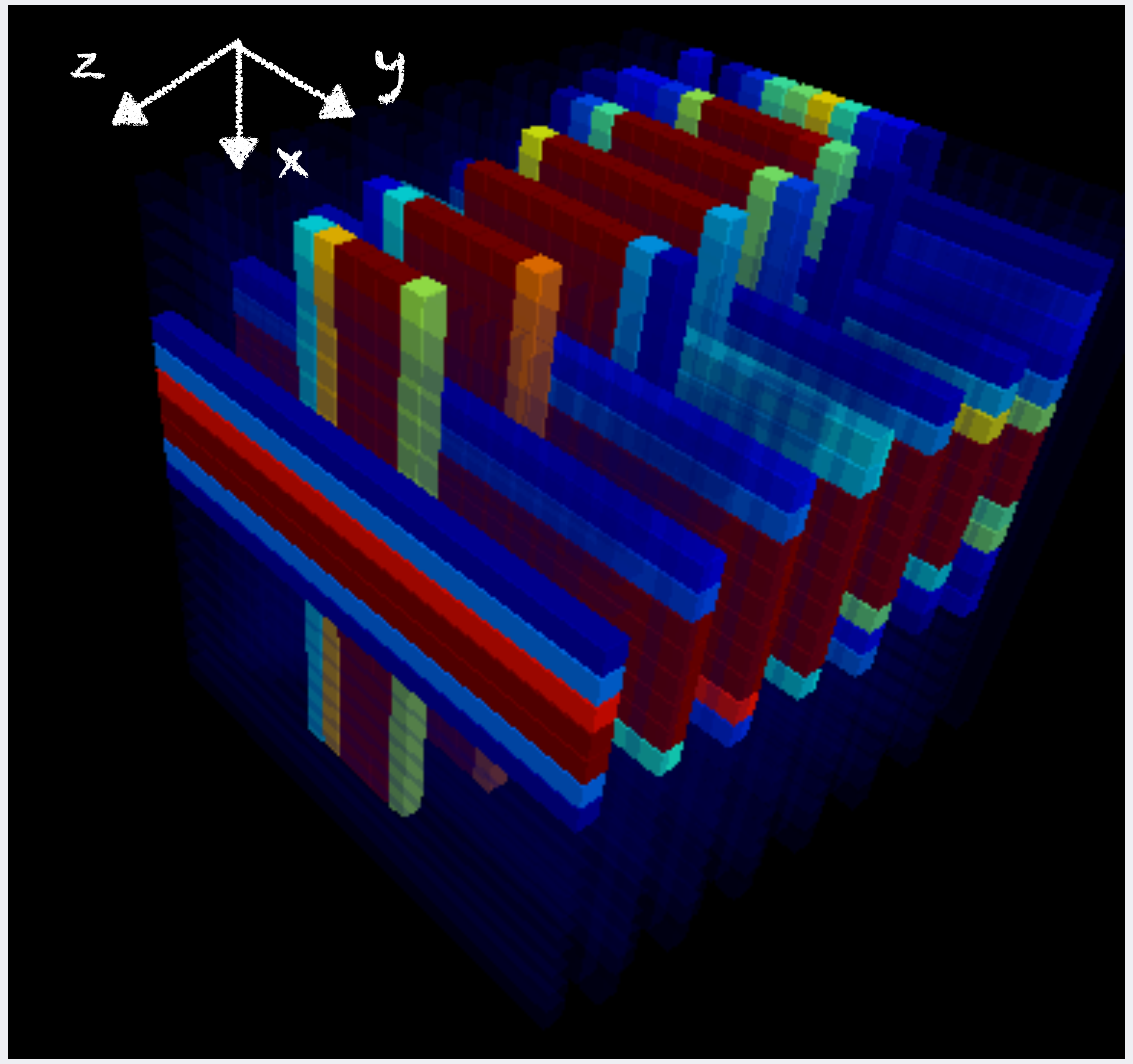




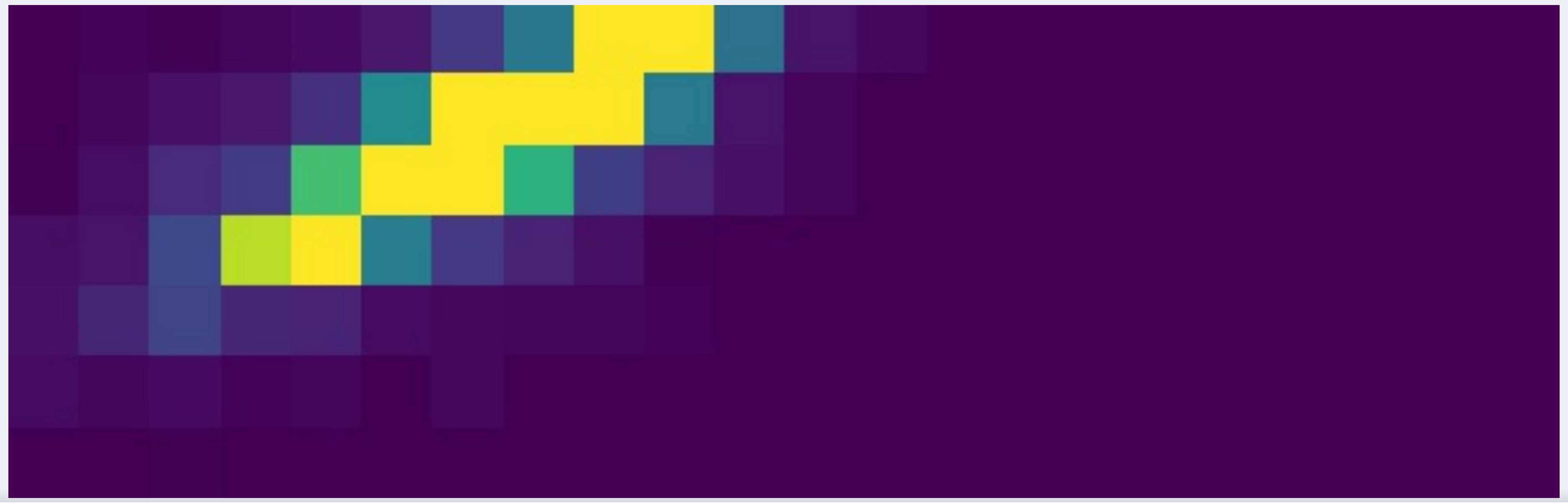
e/p separation

SEPARATE ELECTROMAGNETIC AND HADRONIC COMPONENTS USING SHOWER PROFILE

- The BGO calorimeter is the main sub-detector for e/p separation
- Projections in the XZ and YZ planes give 2D sections of the shower profile
- Can be treated as images, with additional information on the energy deposit
- e/p separation can be treated as a binary classification problem
- How we can make the most of it?



Electron with ~150 GeV deposit (XZ view)

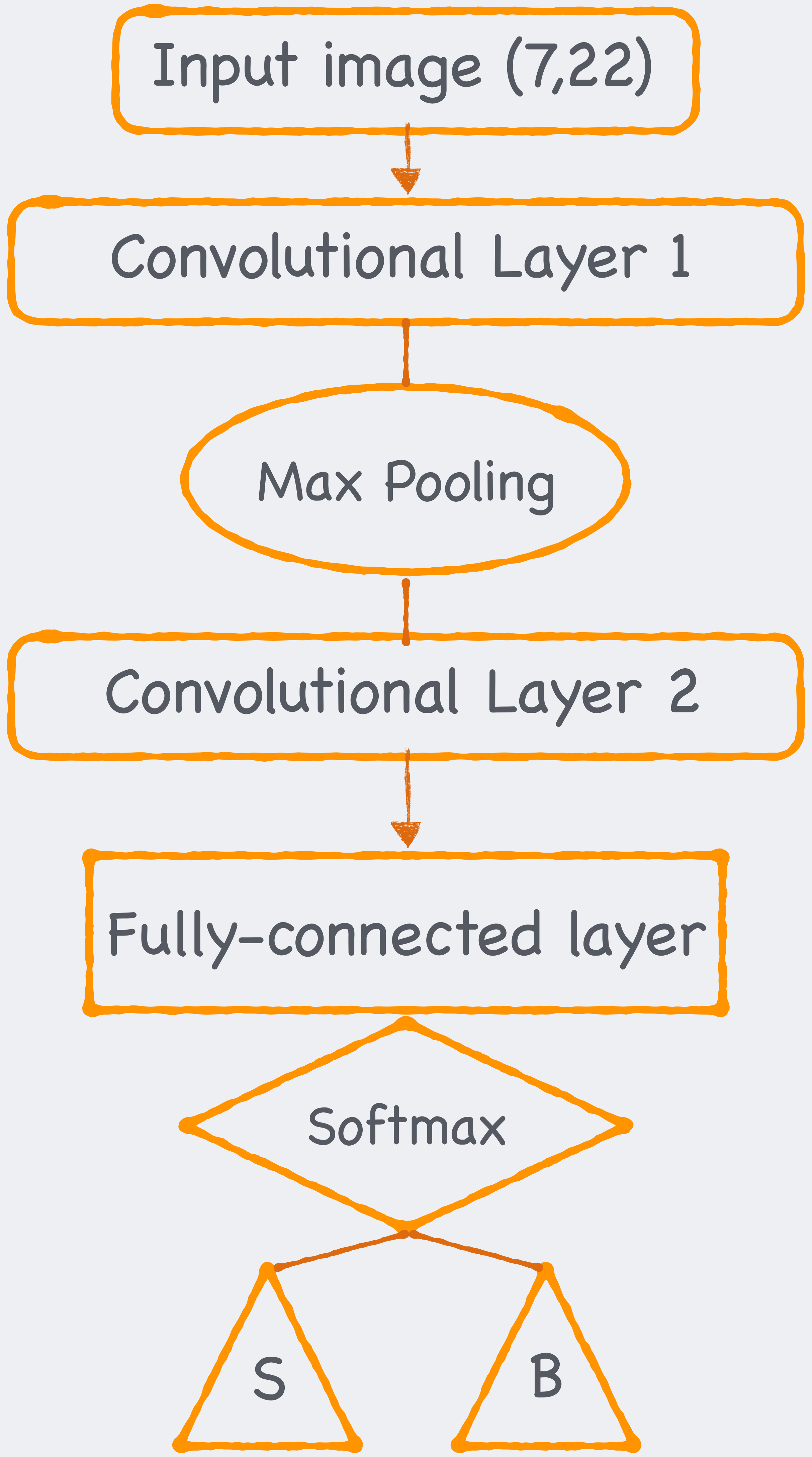


Proton with ~150 GeV deposit (XZ view)

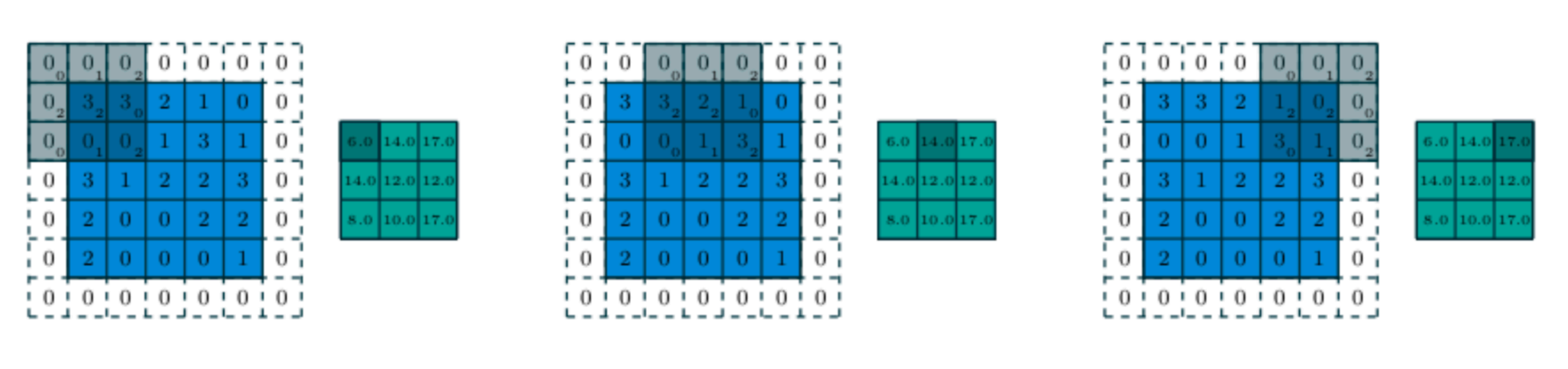
CONVOLUTIONAL NEURAL NETWORKS (CNN)

DEEP LEARNING FOR IMAGE CLASSIFICATION

- CNNs are powerful architectures in deep learning used for image classification, pattern recognition, object detection (and more...)
- Projections in the XZ and YZ planes of the calorimeter are used as images (matrices) with 7x22 pixels (elements)
- As additional information, each pixel is weighted with the value of the deposited energy in the bar
- The network input is an image (matrix) with shape (7,22,1)

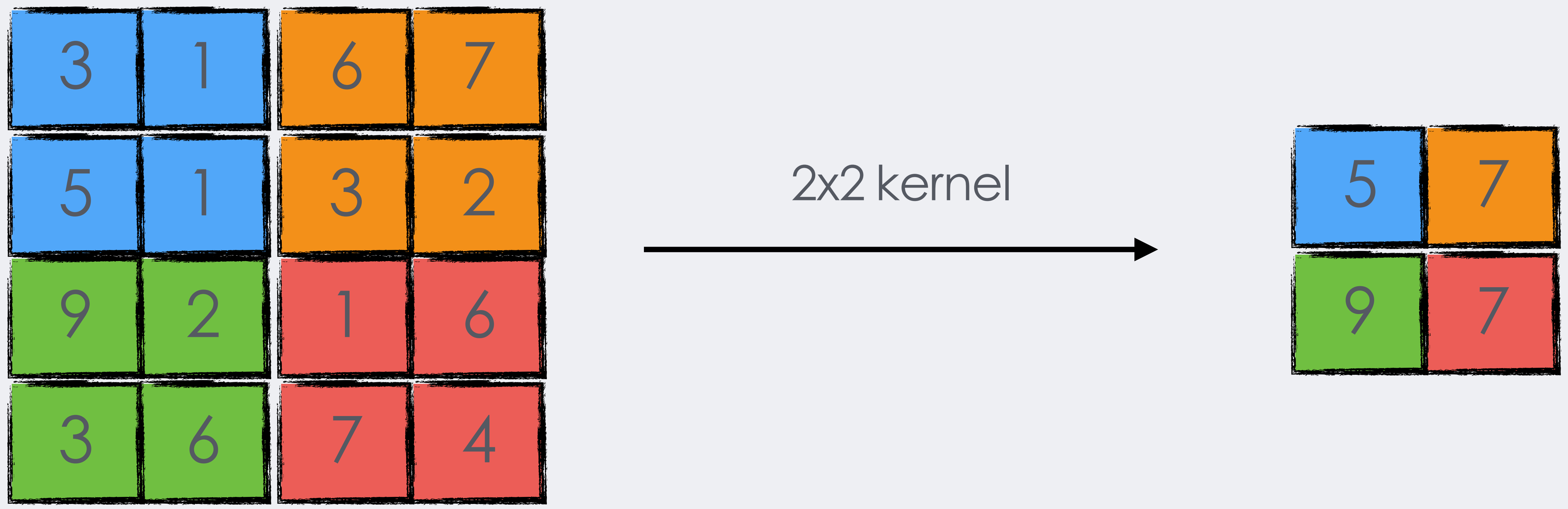


THE CONVOLUTION LAYER



- A kernel (shaded area) slides across the input matrix (light blue). At each step, a linear combination between the kernel elements and the matrix elements is computed, defining the weights for the new feature map
- Some technicalities :
    - Stride: step size for kernel scanning across the matrix
    - Zero padding: a null-weights frame to control the feature-maps size
- After using N kernels, the output is a 4-D tensor with shape (7,22,1,16)

MAX POOLING



- Similar to convolution but performs a downsample of the data object
- Pooling comes in two variants: Max pooling and Average pooling
- Two main advantages:
  - A 2x2 max pooling removes about 75% of network parameters
  - Reduces the pattern always to a region of interest where the information is maximized, providing a basic form of spatial invariance

CNN - FEW MORE INGREDIENTS

- **Fully-connected layer:** at this stage, a flattening operation transforms the 4-D tensor in a 2-D vector. Then the product with a weights matrix downsamples the object to a 2 elements vector

- **Softmax:** outputs from the fully-connected layer are difficult to interpret, so one can apply the function:

$$S(y_i) = \frac{e^{y_i}}{\sum_{j=1}^{J} e^{y_j}}$$

  for J classes, to convert the scores into a probability in the range [0,1].

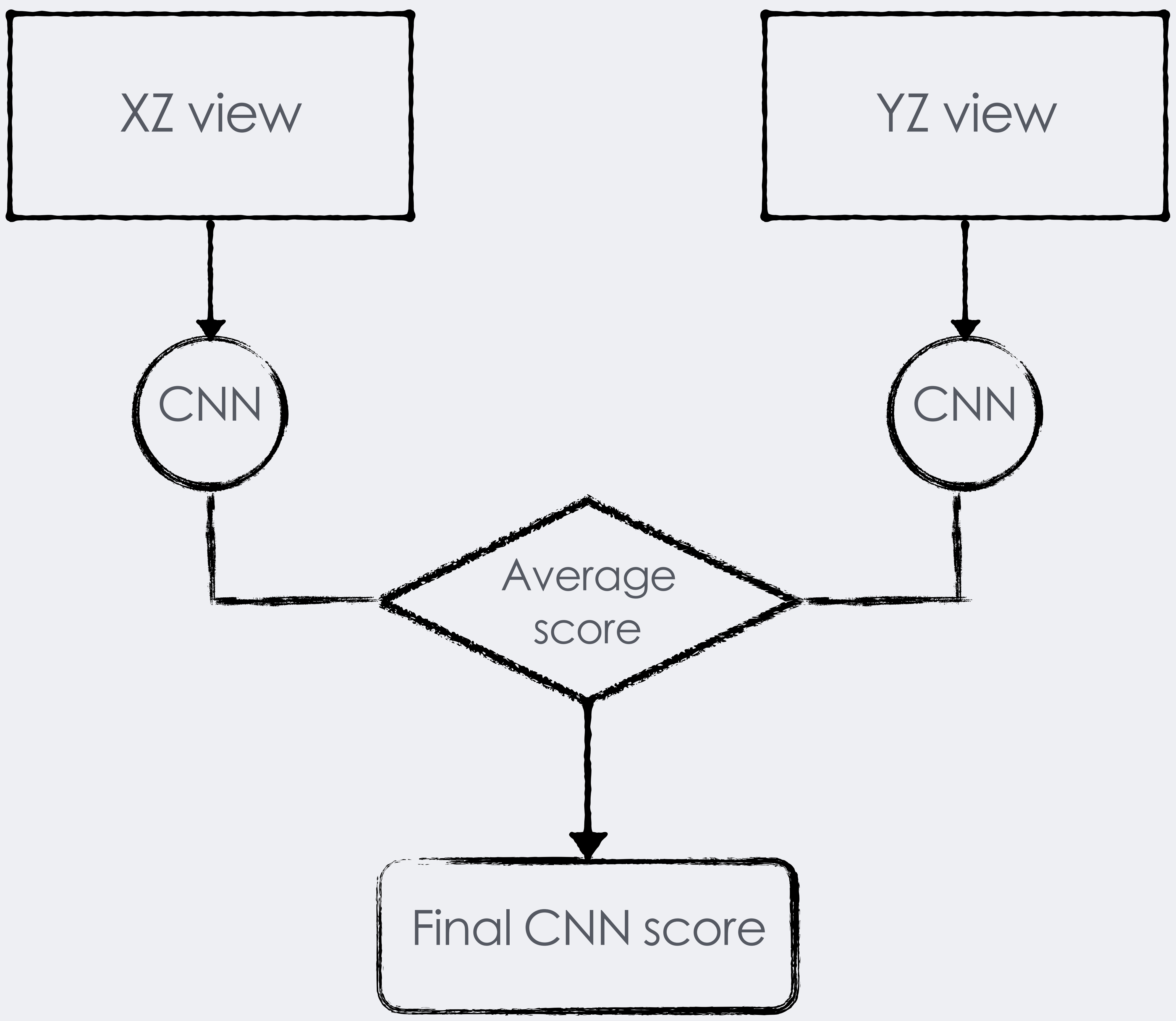
- **Training and optimization:** The CNN is trained on samples of Monte Carlo electrons and protons in the energy range 1 GeV - 10 TeV (up to 100 TeV for protons). The optimization function is the cross-entropy:

$$D(S, L) = -\sum_i L_i log(S_i)$$

- The optimizer algorithm used is ADAM (gradient-descent based)

CNN APPLICATION

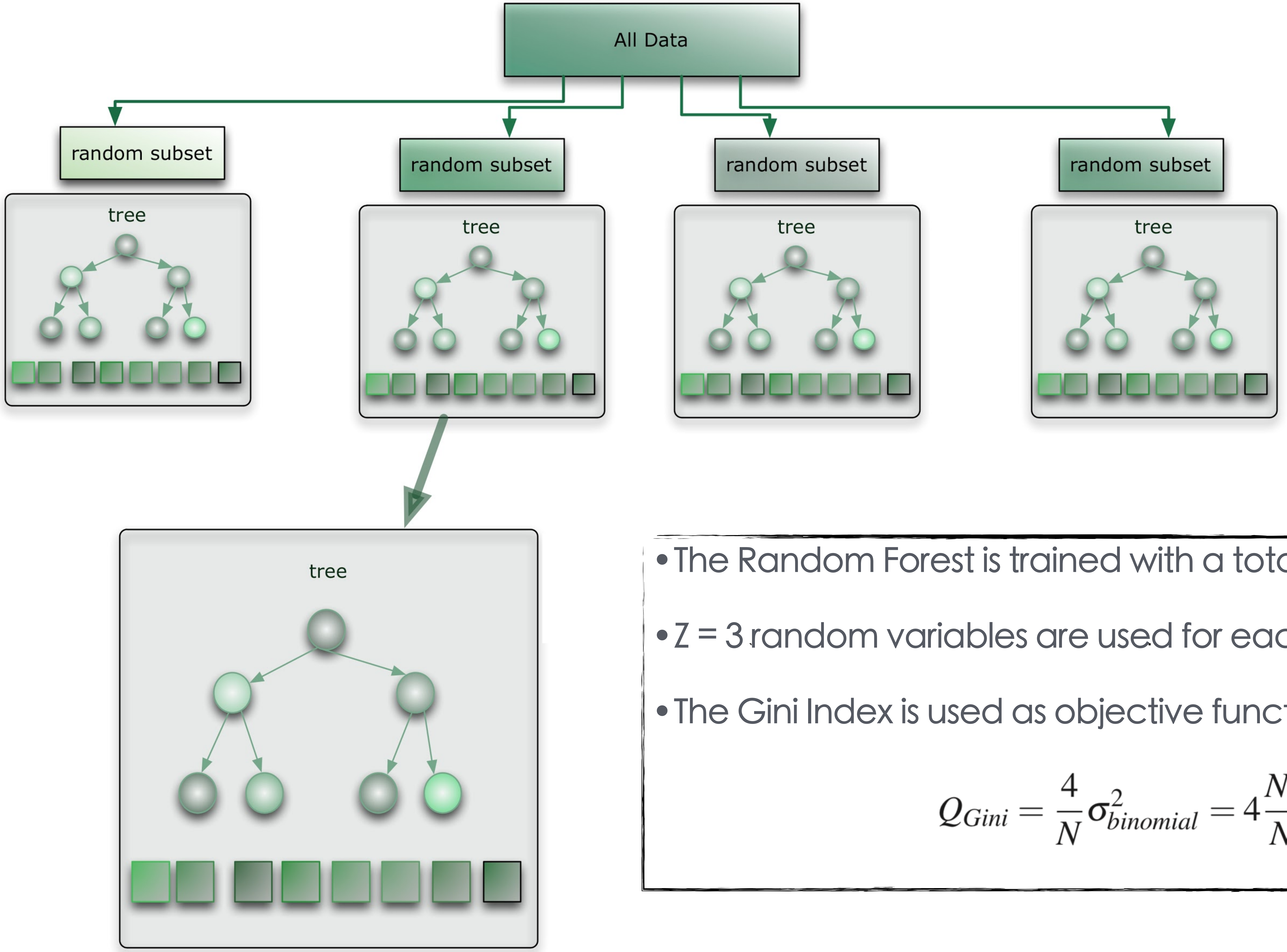
- The trained CNN is used independently with the XZ and YZ view of each event
- The output score obtained from each CNN is combined via majority voting to determine the final event score
- Average voting resulted in a more effective approach than majority voting



Is this enough? No!

- We add another classifier that uses data in a more "classic" way, in order to improve the overall e/p separation

RANDOM FOREST CLASSIFIER



- N trees (estimators) in the forest
- For each tree, a random subset of data is selected
- At each node, a random subsample of z variables is selected
- The variable and cut that maximize the purity of the sample are chosen

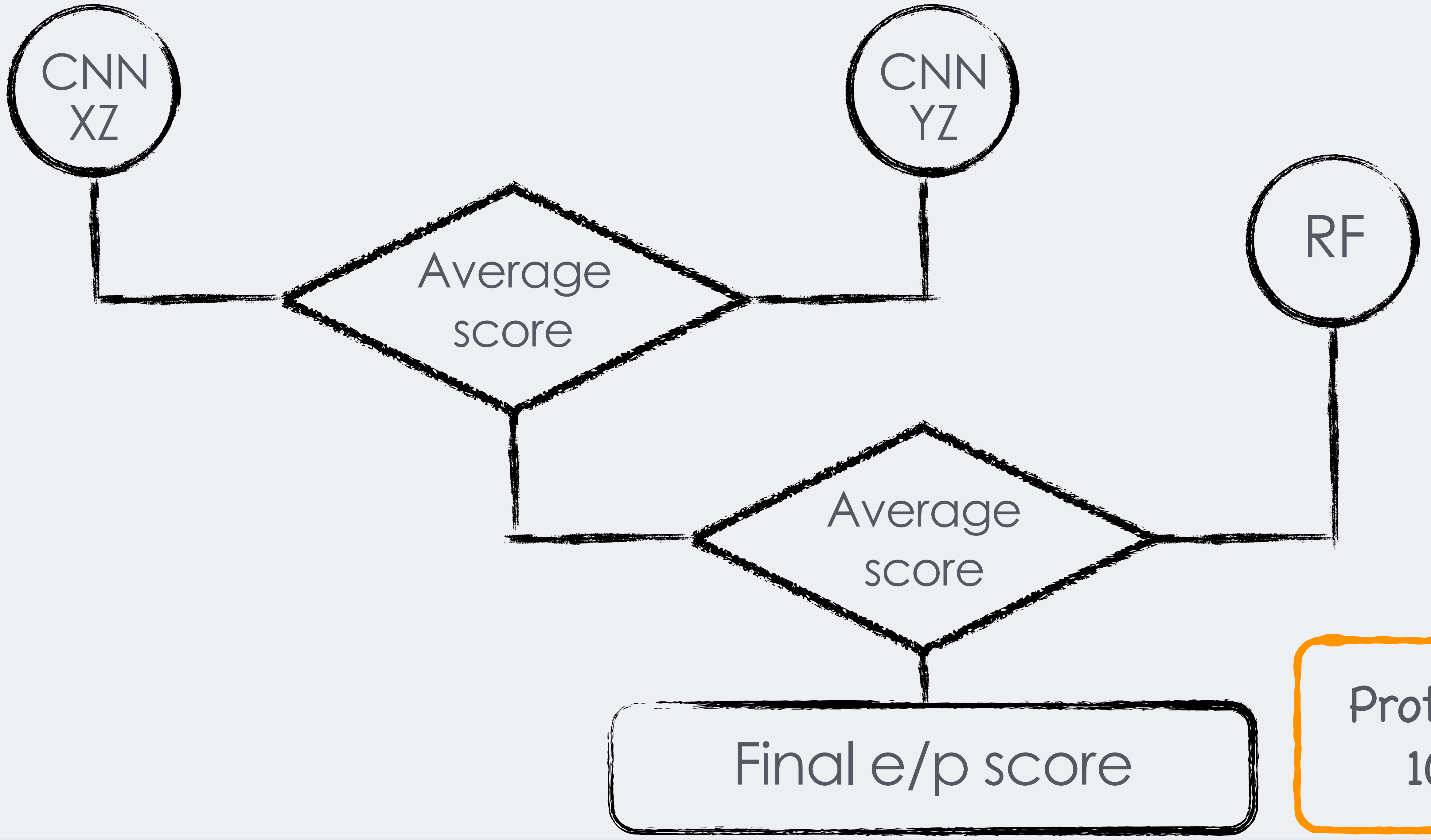
- The Random Forest is trained with a total of N = 1000 estimators
- Z = 3 random variables are used for each node
- The Gini Index is used as objective function to train each estimator node

$$Q_{Gini} = \frac{4}{N}\sigma^2_{binomial} = 4\frac{N_e}{N}\frac{N_p}{N} = 4\frac{N_e(N-N_e)}{N^2} \in [0,1]$$

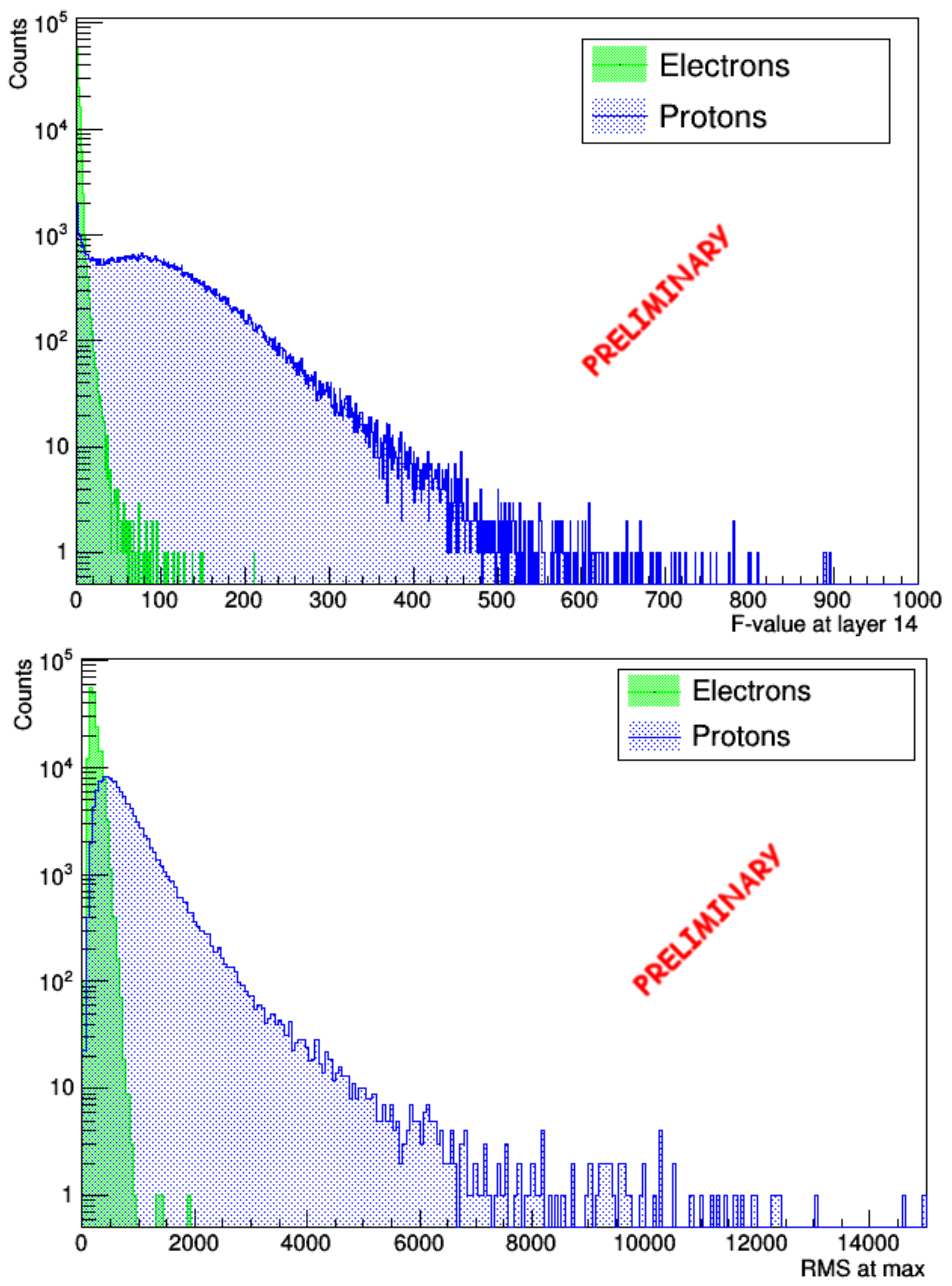
RANDOM FOREST CLASSIFIER

PoS(ICRC2017)764

- A total of 9 variables defined on the BGO calorimeter data are used.
- The energy released in the NUD is added to this variables
- All these variables have been selected with the same RF classifier from a larger sample



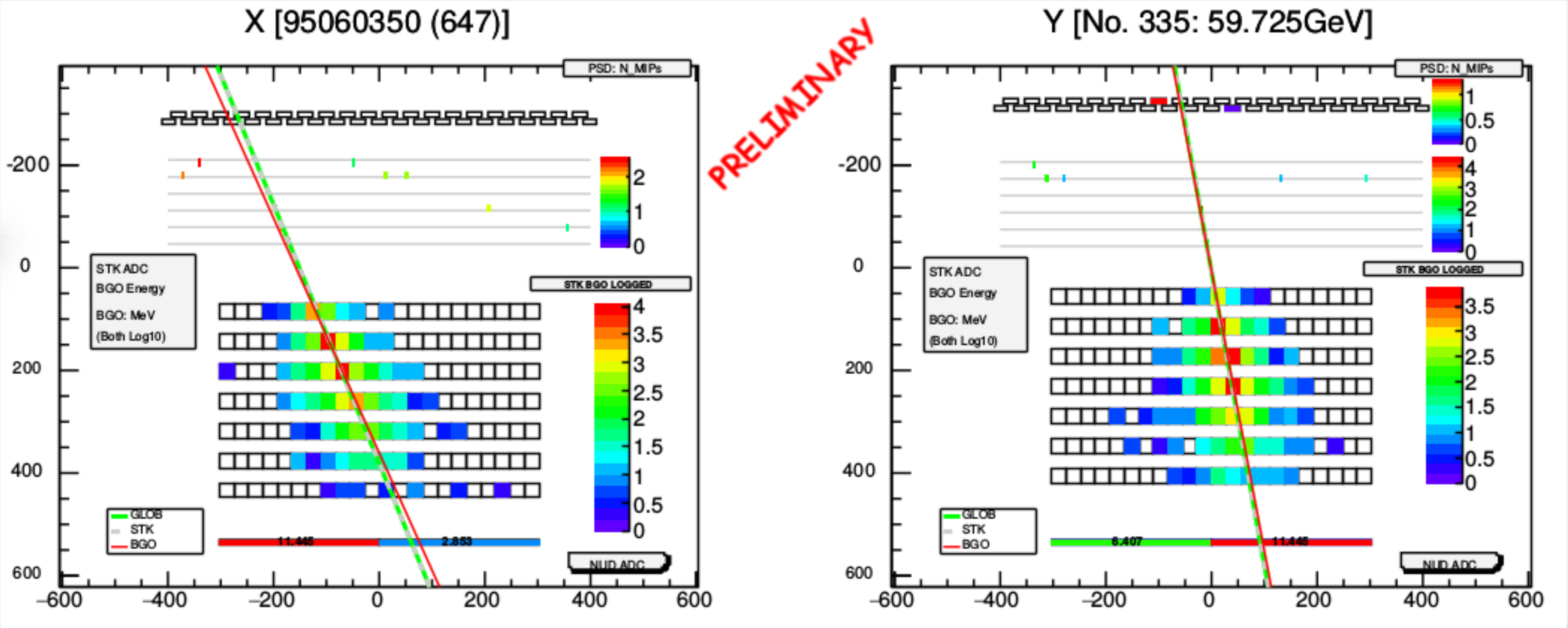
Proton contamination
10^{-4} @ 99% eff.



e/ɣ separation

SEPARATING ELECTRONS AND PHOTONS IN THE ELECTROMAGNETIC COMPONENT

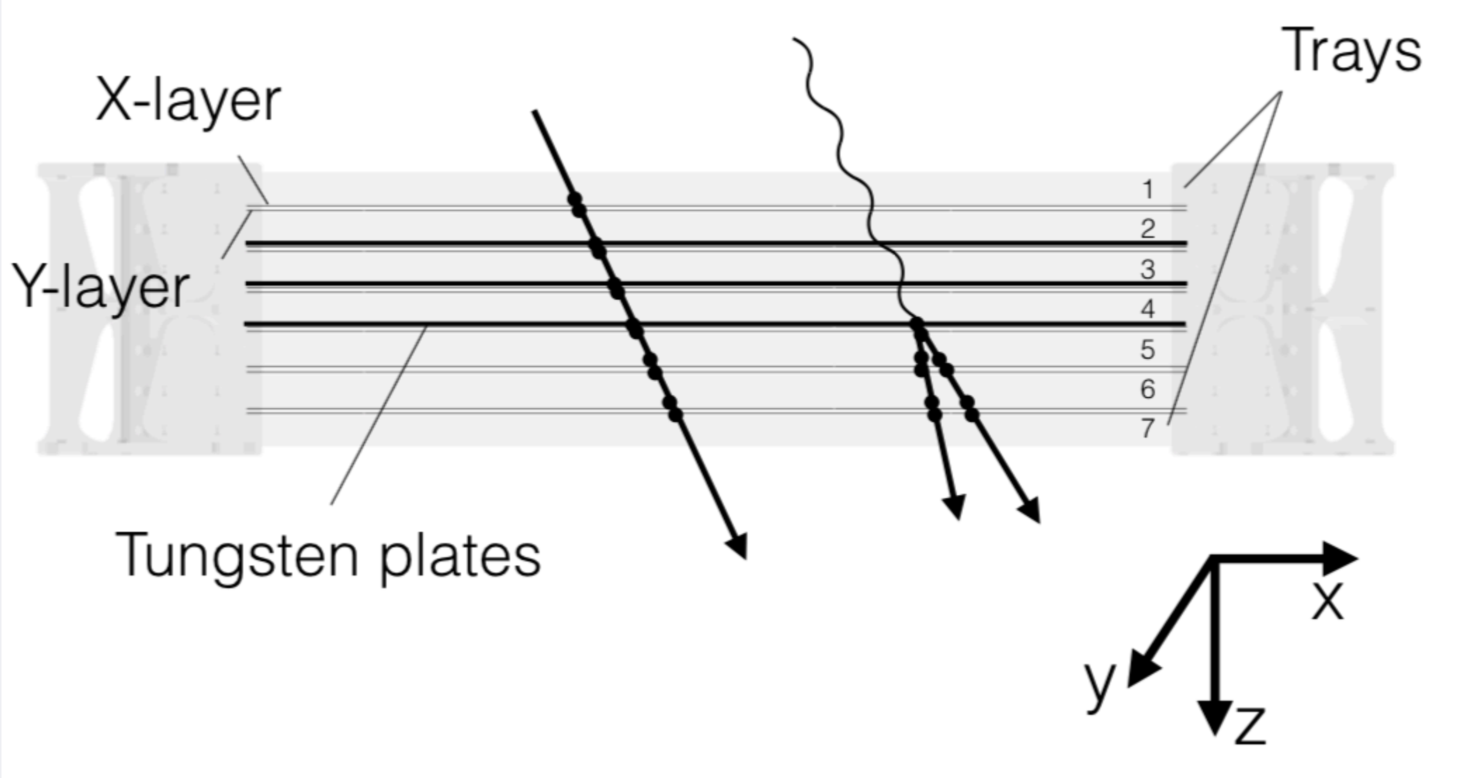
- Once selected the electromagnetic component ($e^-/e^+/\gamma$) we want to remove the charged component
- The PSD and the first layer of the STK can be used as powerful veto system
- One main effect limits the veto efficiency: back-scattering (aka backsplash)

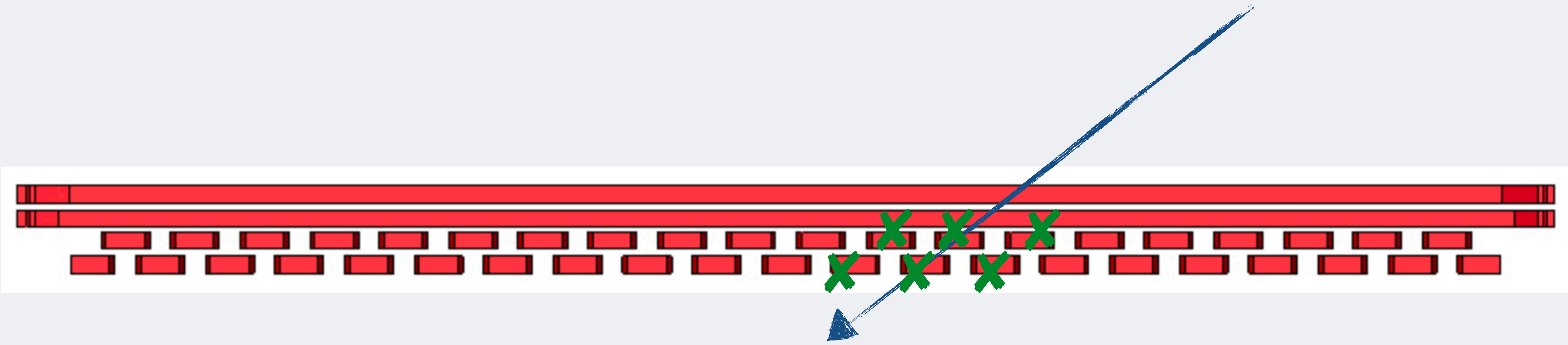


PoS(ICRC2017)764

e/γ separation

SEPARATING ELECTRONS AND PHOTONS IN THE ELECTROMAGNETIC COMPONENT

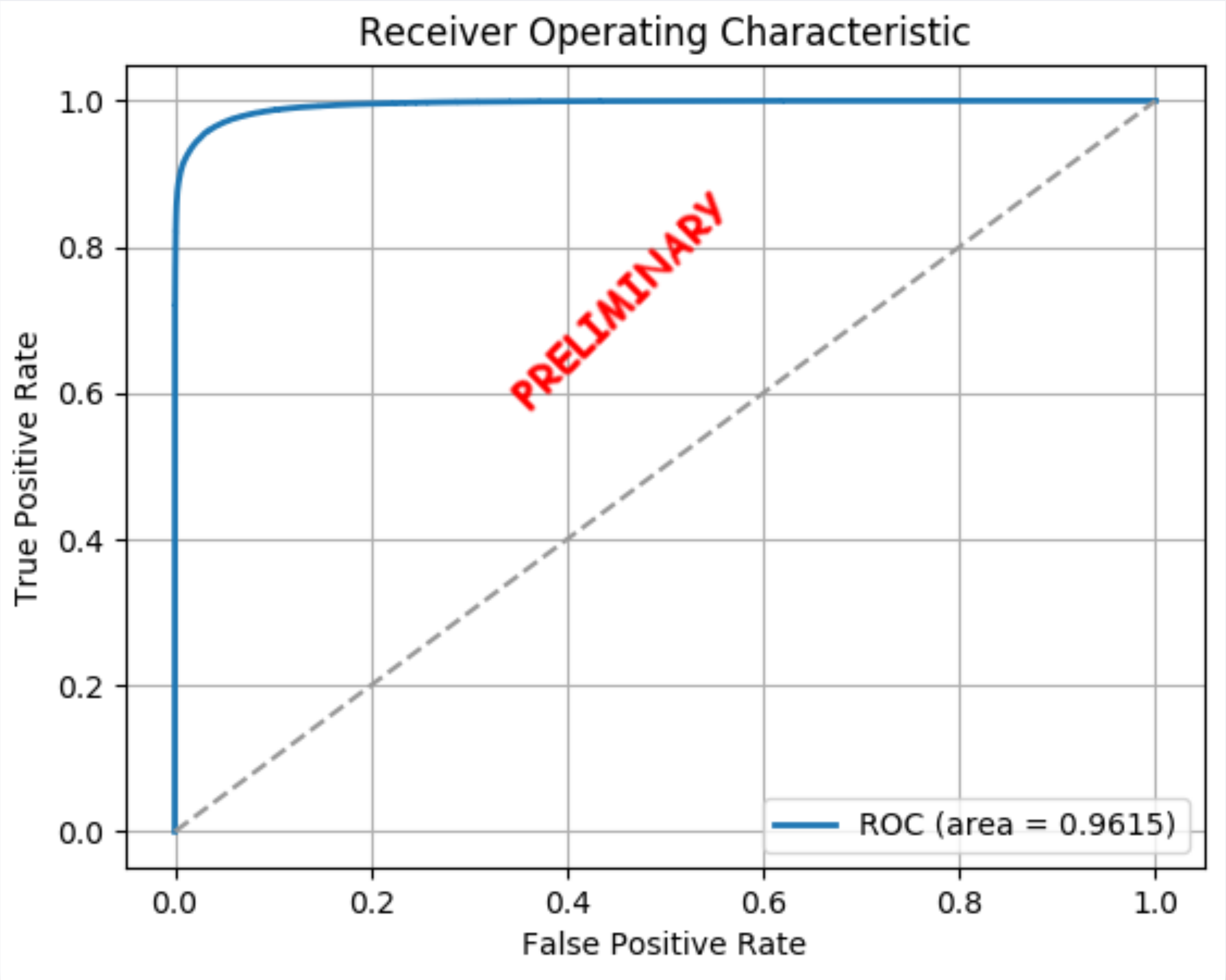


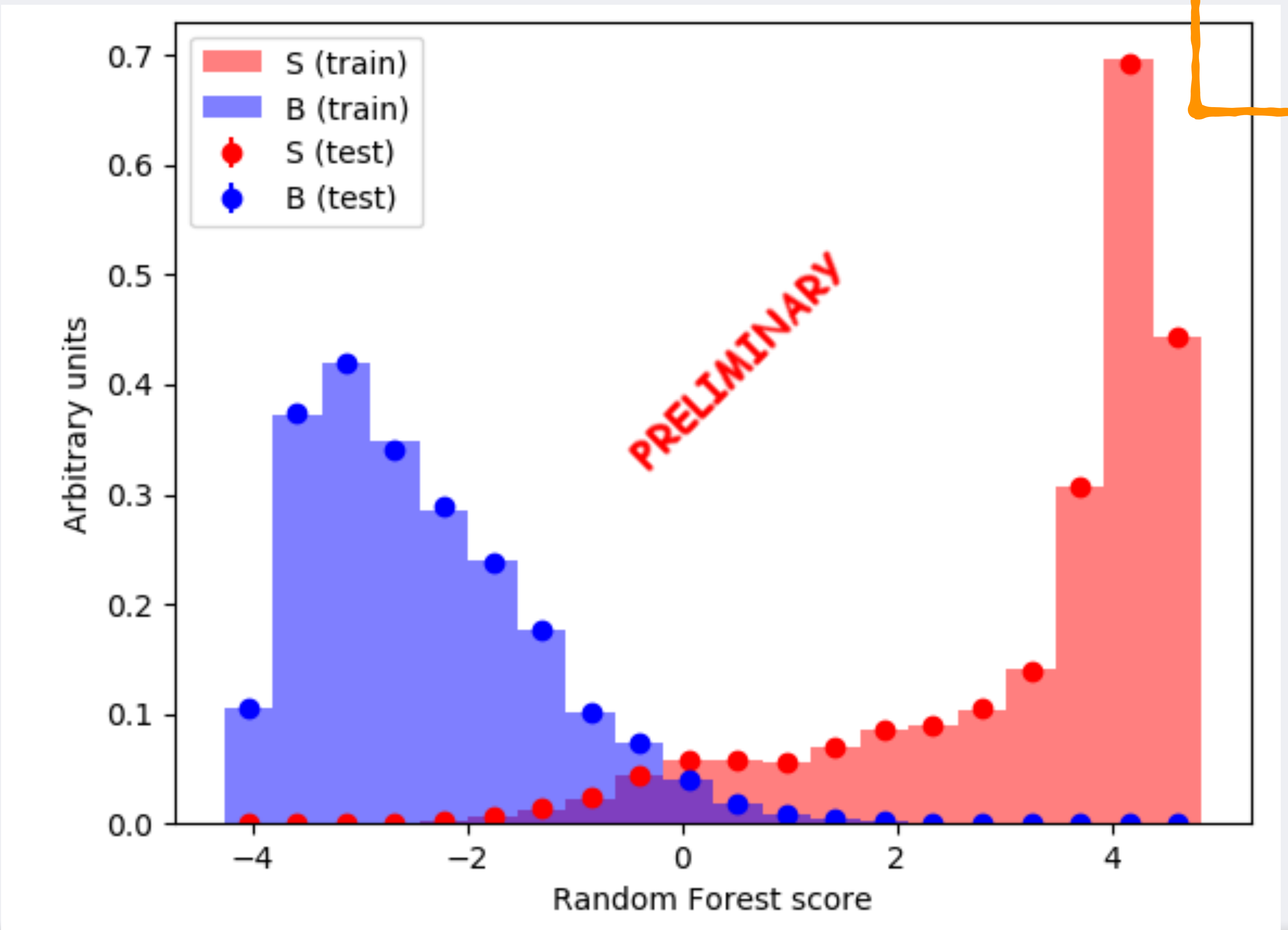


- In the high energy range for the gamma-rays (> 1GeV), the backsplash effect is not negligible
- Simple cut-based analysis may not be very effective —> we introduce another RF classifier for e/γ separation
- A total of 12 variables has been defined using region of interests of the PSD around the particle track, and from the STK.
- Same RF configuration used for the e/p separation

Gamma-ray selection - RESULTS

- In the high energy range for gamma-rays (> 1GeV), the backsplash effect is not negligible
- Simple cut-based analysis may not be very effective —> we introduce another RF classifier for e/ɣ separation
- A total of 12 variables has been defined using region of interests of the PSD around the particle track, and from the STK.
- Same RF configuration used for the e/p separation





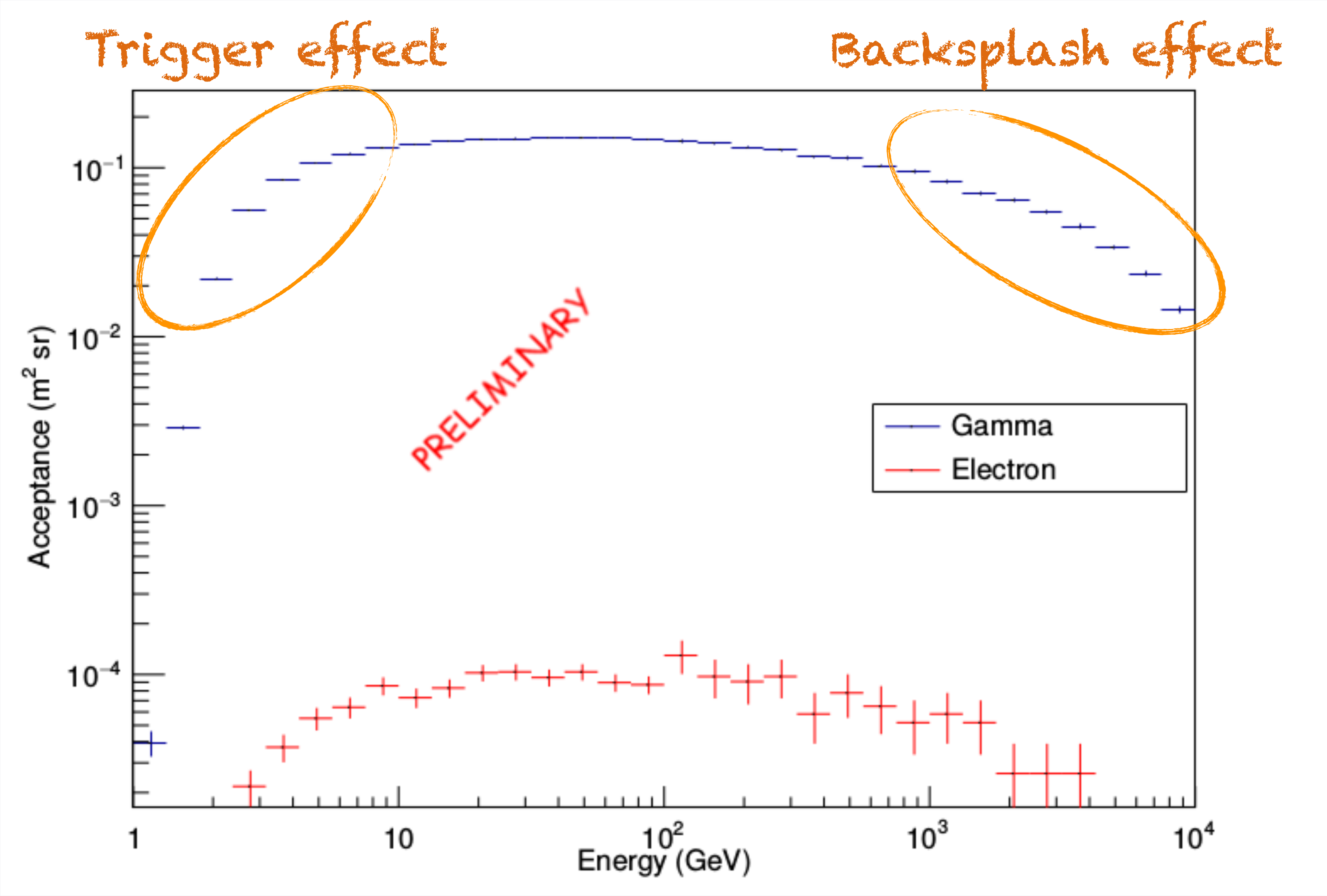
Kolmogorov-Smirnov test

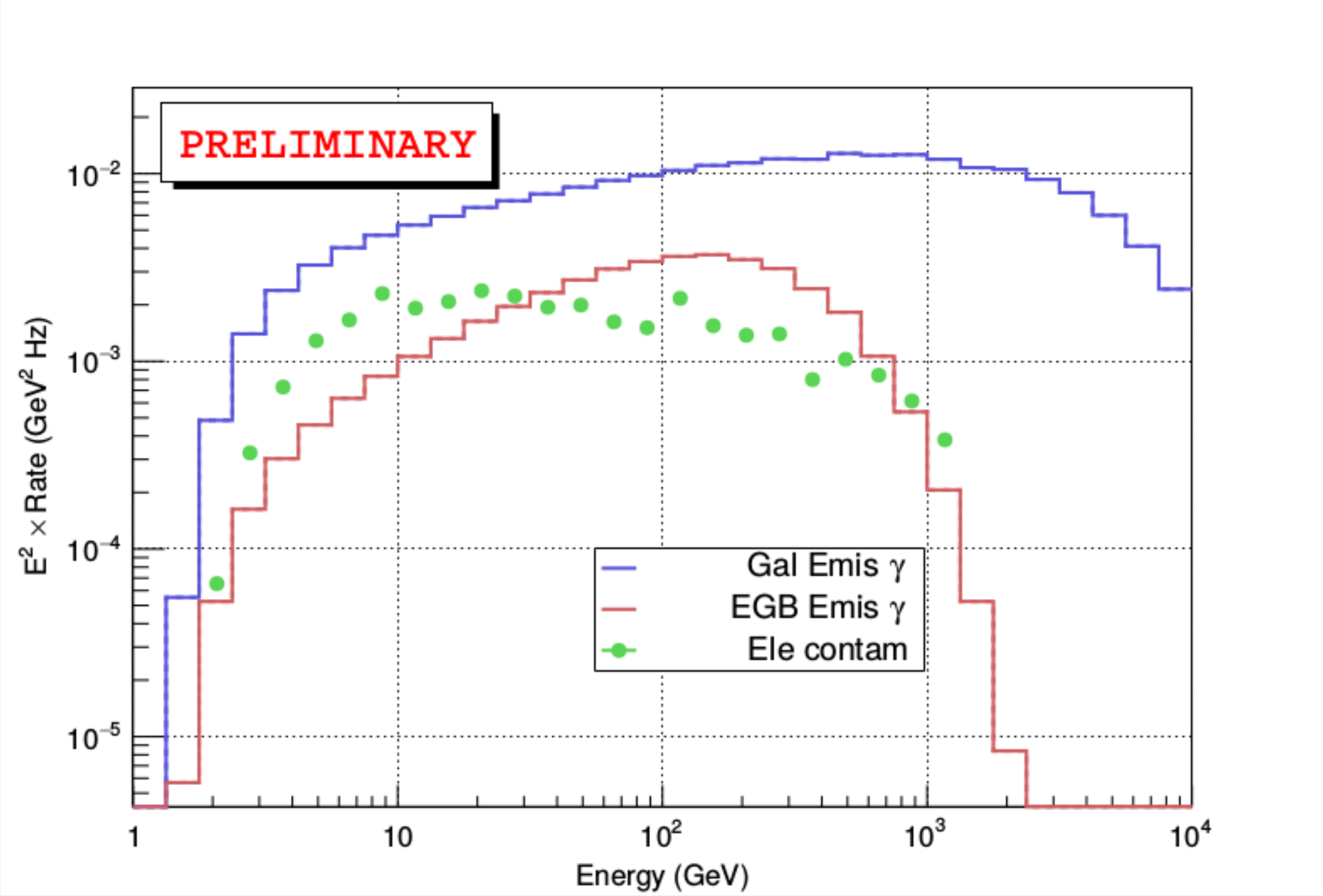
0.72 (S) - 0.76 (B)

PoS(ICRC2017)764

Gamma-ray selection - RESULTS

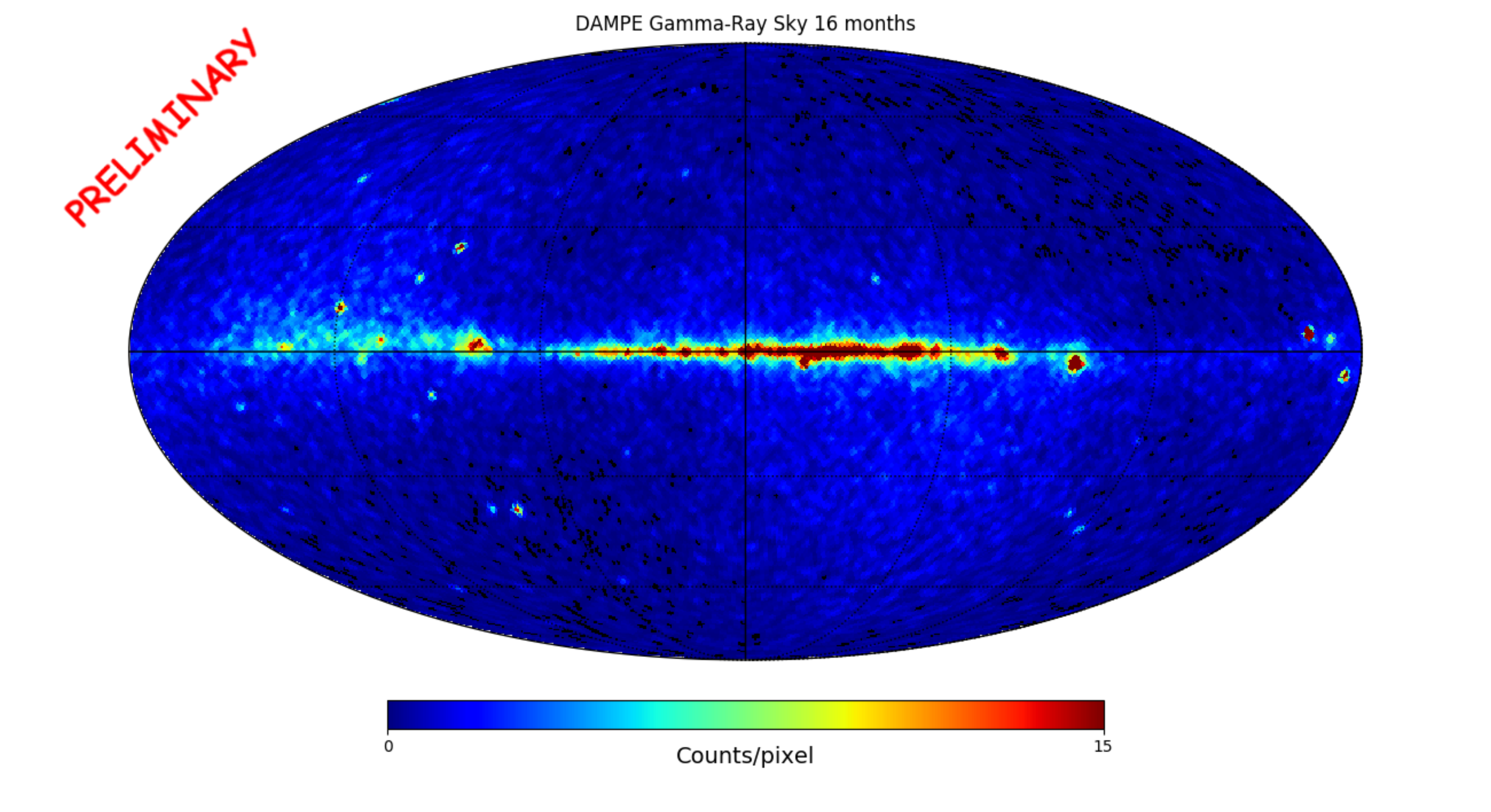
PoS(ICRC2017)764





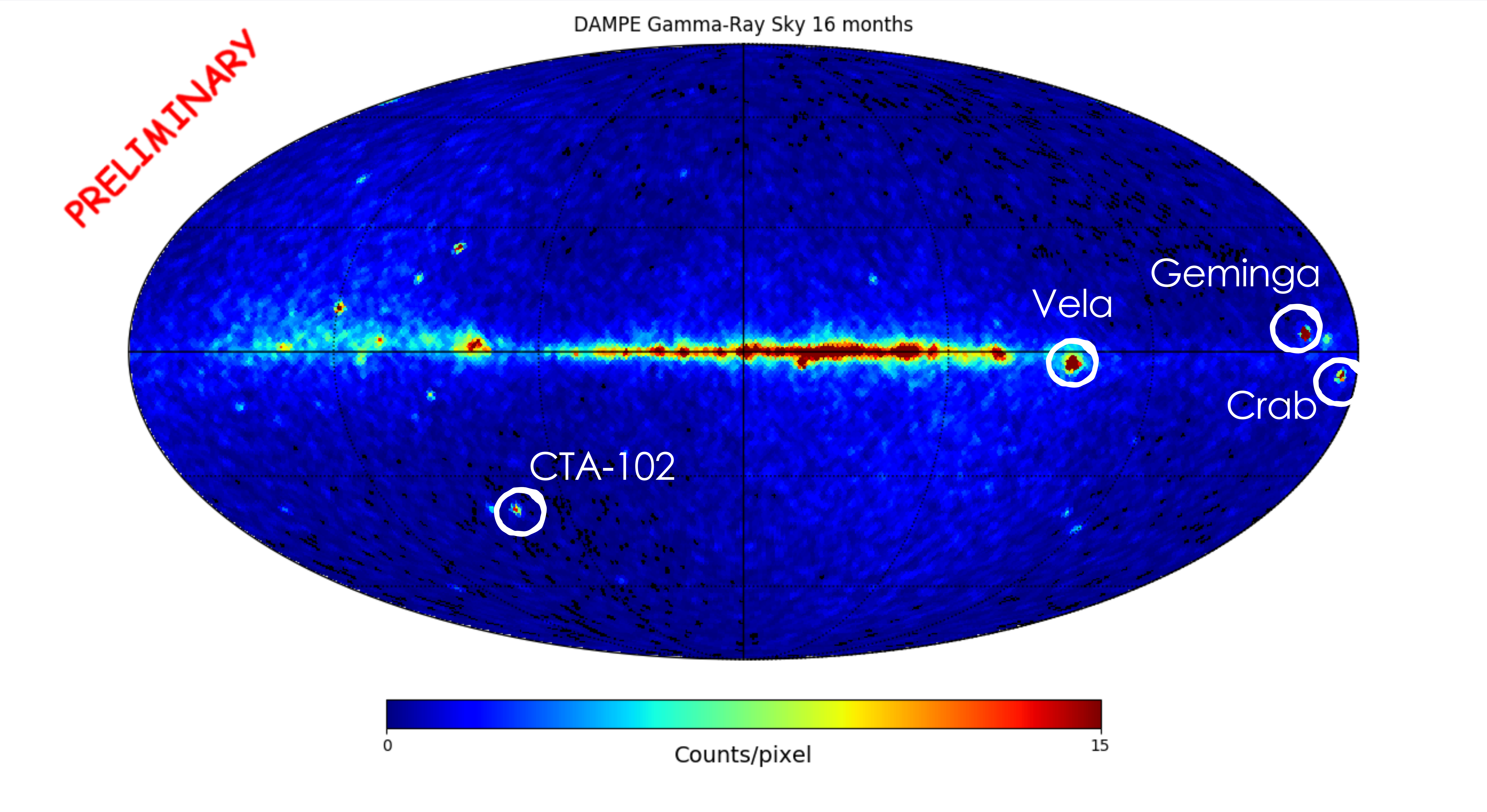
- The thresholds from the overall selection have been tuned to obtain a trade-off where the expected contamination rate from electrons is at the level of the extragalactic isotropic emission
- The proton contamination after the 2 selections is reduced at a level lower than 10^{-7}

Gamma-ray selection - RESULTS



- An overall selection efficiency of 68% with an electron contamination of 10^{-3}-10^{-4} in 1 GeV - 10 TeV
- An average of 110 photons/day observed

Gamma-ray selection - RESULTS



- An overall selection efficiency of 68% with an electron contamination of 10^{-3}-10^{-4} in 1 GeV - 10 TeV
- An average of 110 photons/day observed

Summary

- A complete machine learning-based particle identification pipeline can be implemented in a space-based detector with very good performance
- Each detector can give a peculiar contribute to the selection and multiple different approaches can be combined
- Imaging detectors like the DAMPE BGO calorimeter are very suitable environments to apply pattern recognition deep learning architectures like Convolutional Neural Networks
- Multivariate algorithms can improve the performance of classic cut- based selections
- The ML gamma-ray selection in DAMPE has an high level of purity over the cosmic-rays background components
- DAMPE is a good laboratory for this application, that can be scaled-up to more complex detector systems

Take-home message: Machine Learning is not magic. Identify the physics-driven problem, choose the best approach, and have fun!