Search for ultra-high energy neutrinos in the background of UHECR

Abha Khakurdikar

Co-authors: Prof. Jörg R. Hörandel and Dr. Washington R. Carvalho Jr.

10th International Workshop on

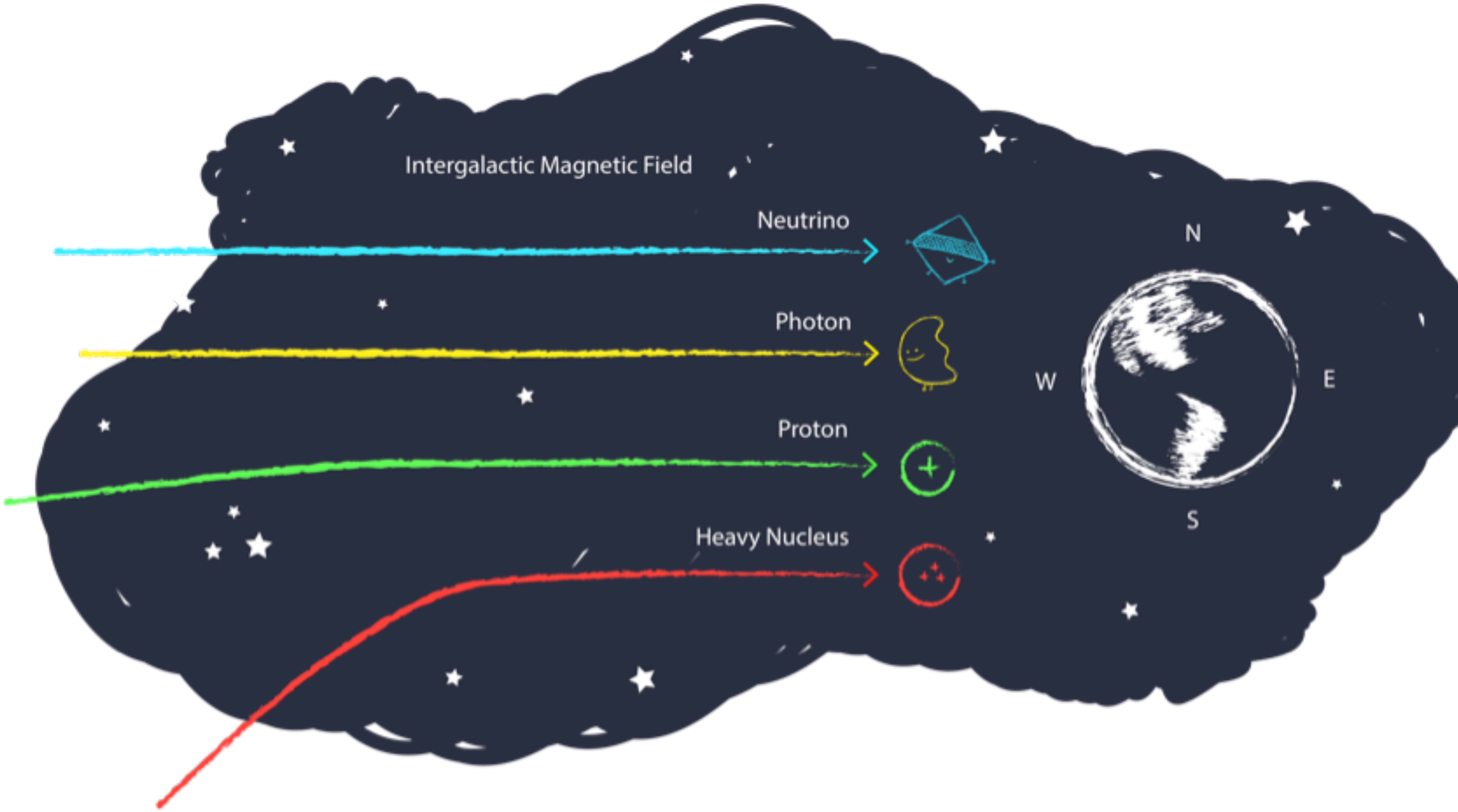
Acoustic and Radio EeV Neutrino Detection Activities

11-14 June 2024

Radboud Universiteit

Why ultra-high energy neutrinos ?

- Key role in understanding the origin of the ultra-high energy cosmic rays (UHECR).
- Their observation should open a new window to the universe, otherwise hidden by large amounts of matter.
- Neutrinos are not deviated by the magnetic field -> point back to the sources.
- In the EeV range, neutrinos are expected to be produced in the same sources where the UHECR are thought to be accelerated.

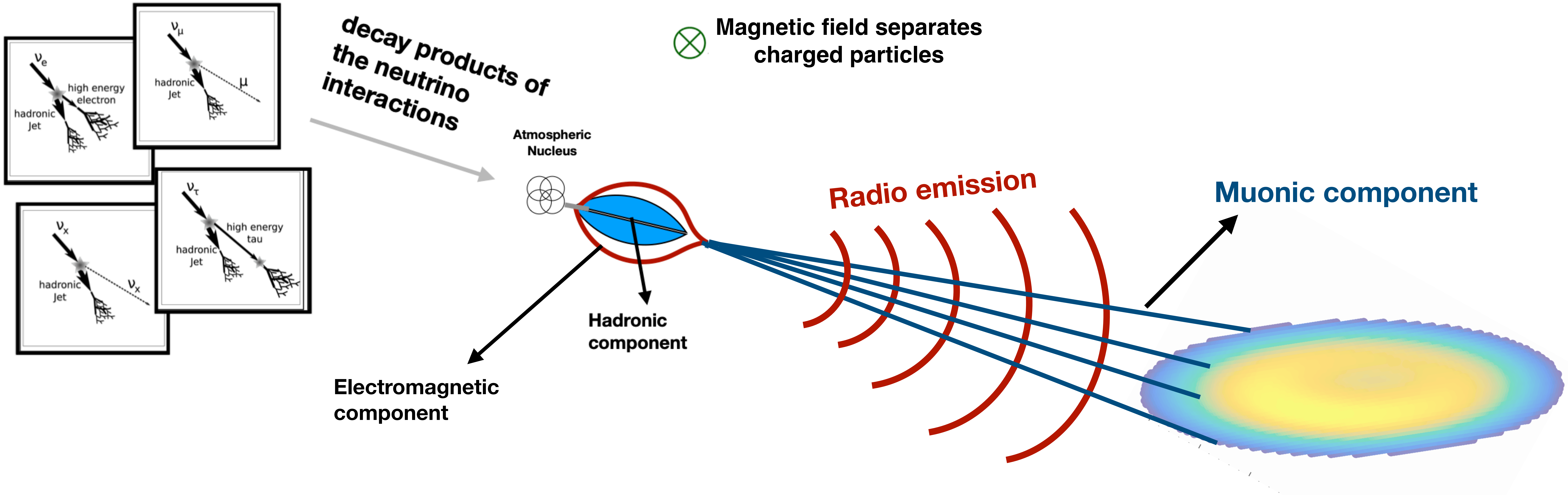


Source: https://subarutelescope.org/subaru20anniv/assets/files/Shigeru_Yoshida.pdf

Contact: a.khakurdikar@astro.ru.nl

Ultra-high energy neutrinos

- All flavors of neutrinos with Energy E ≥ 1 EeV
- Highly inclined extensive air showers (EAS) induced by the **neutrinos** : $75^\circ \leq \theta \leq 85^\circ$



Schematic diagram of the development of pipeline to detect the air showers induced by the the ultra-high energetic neutrinos

Contact: a.khakurdikar@astro.ru.nl

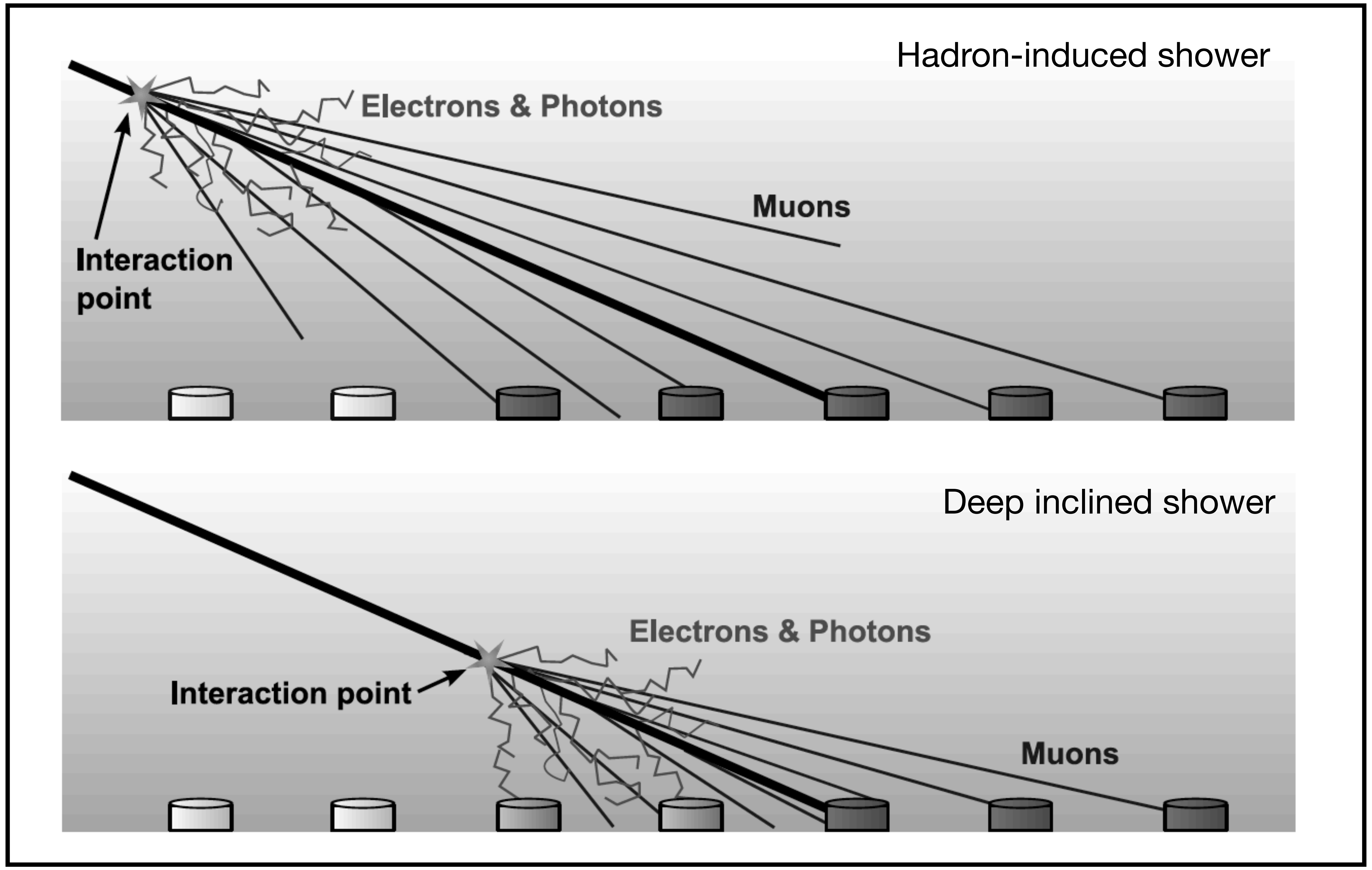
Radboud Universiteit

Identify the neutrino induced EAS

- The main challenge in detecting the UHE neutrinos is to identify a neutrino-induced shower (signal) in the background of showers initiated by UHERCRs.

EM component is absorbed and only the muons reach the detector

Significant EM component at the detector level

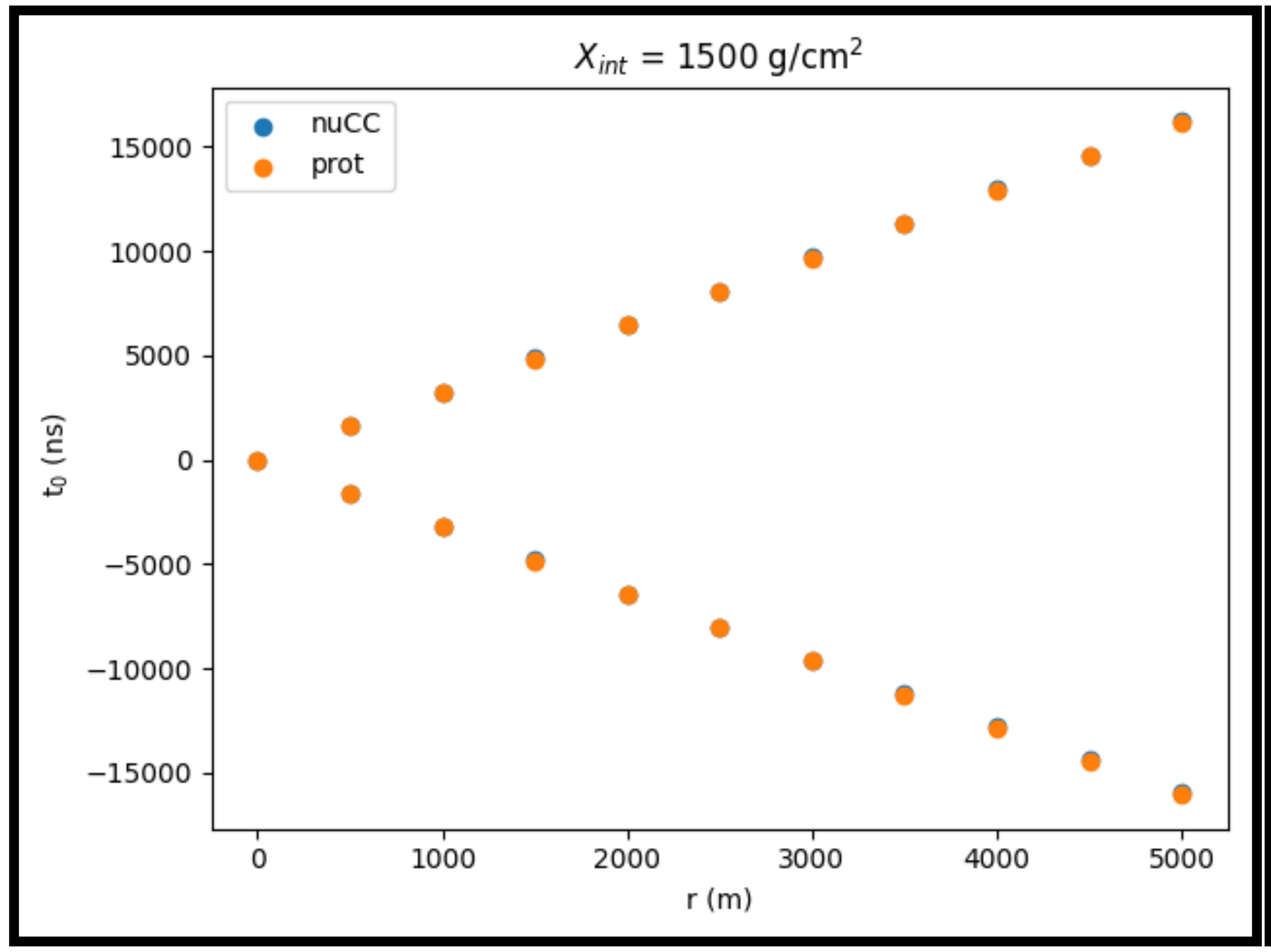


Source: arXiv:1202.1493

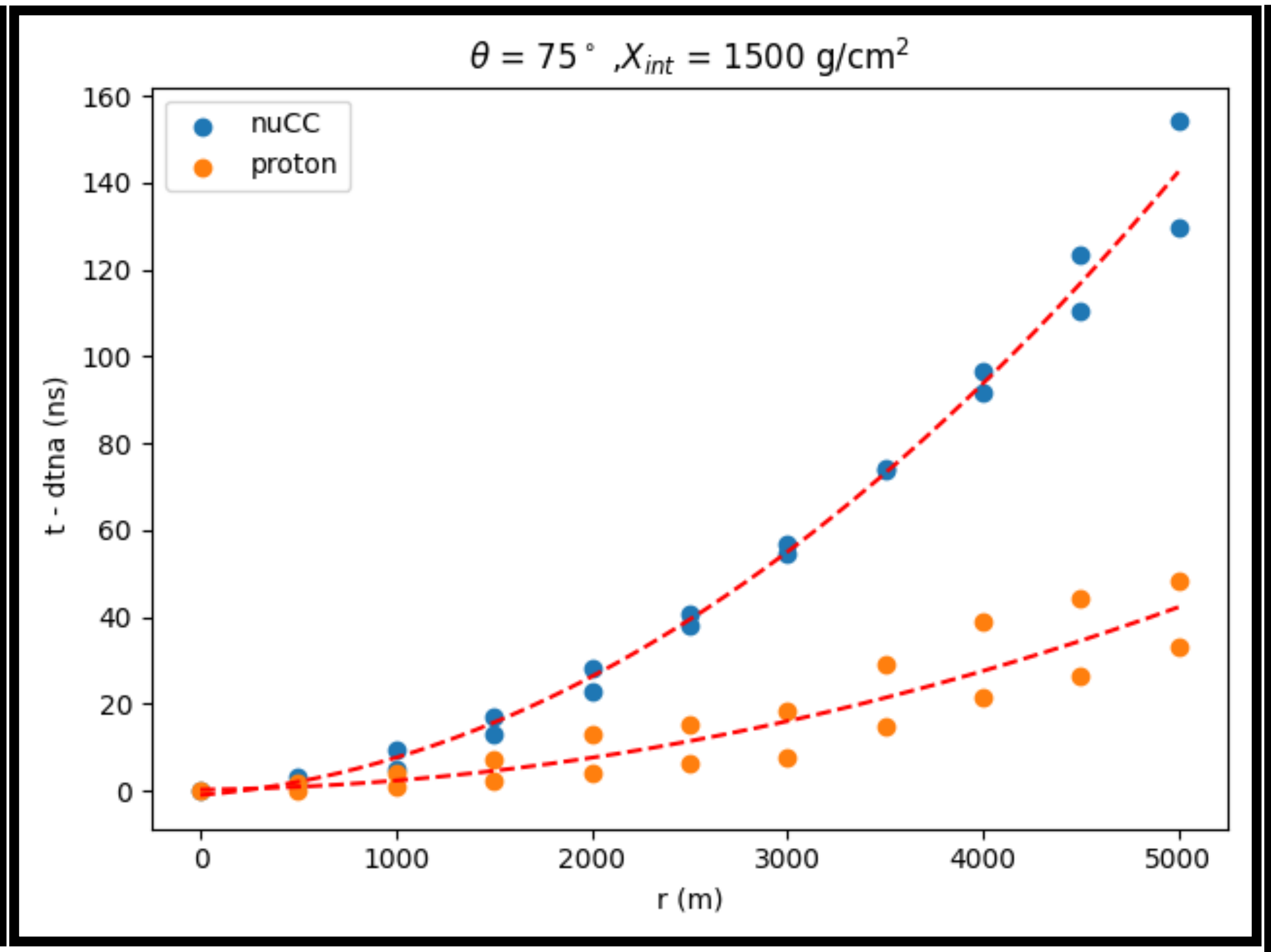
Contact: a.khakurdikar@astro.ru.nl

Radboud Universiteit

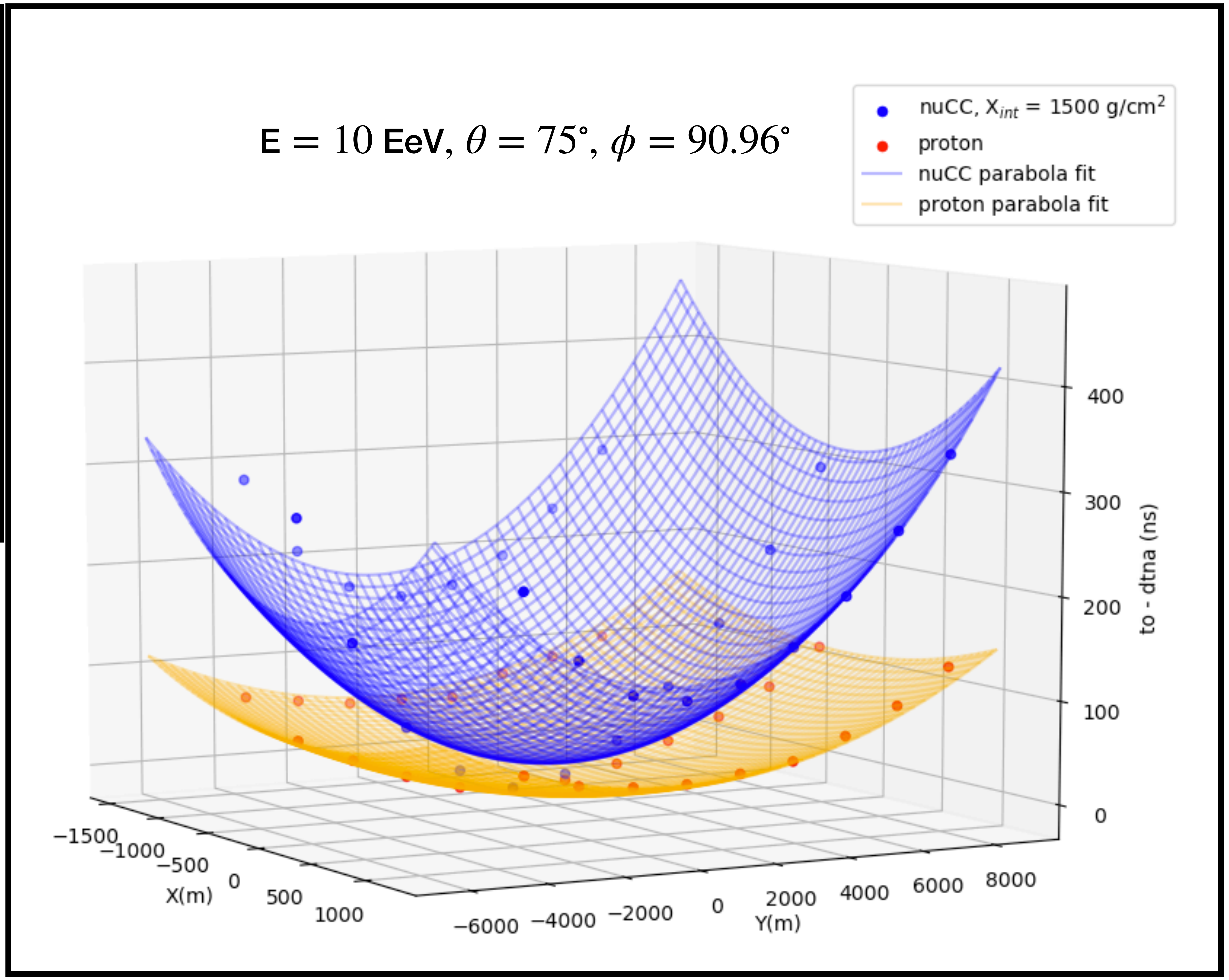
Features for identification



Signal time measured with respect to the distance from the shower core for a single event.



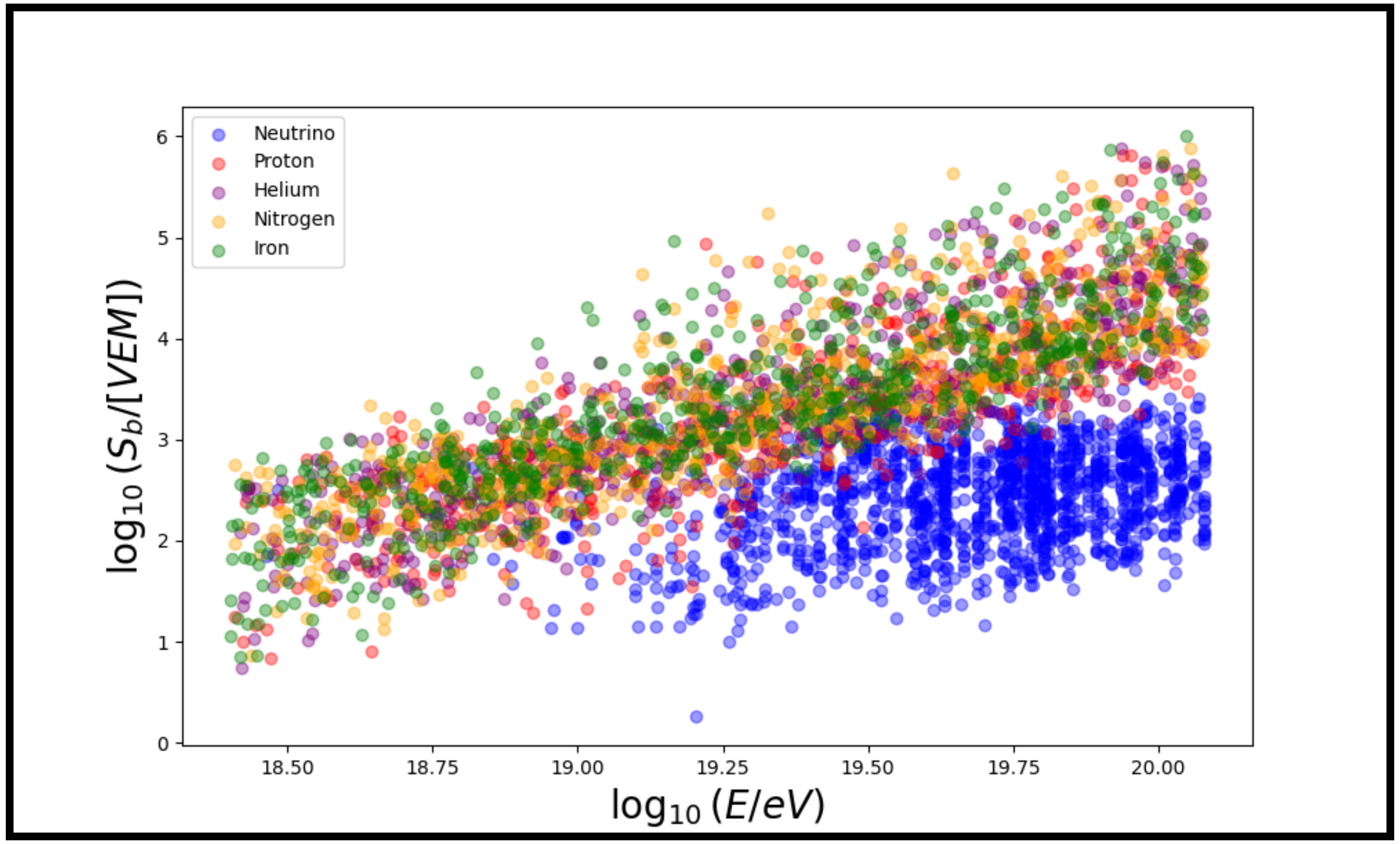
Corrected time signal with respect to the distance from the shower core

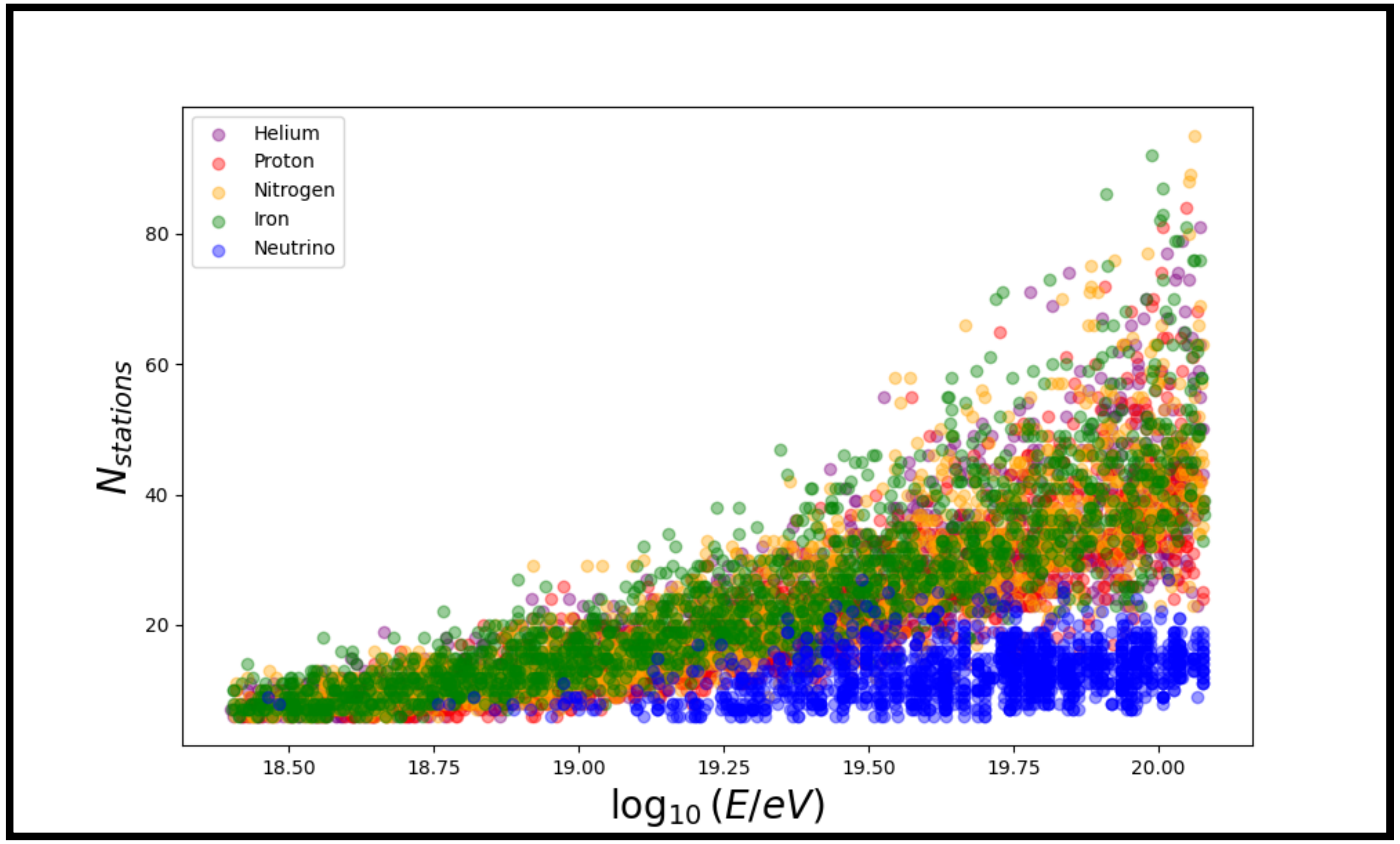


Corrected time signal with respect to the distance from the shower core

- Due to geometrical reasons, the arrival of the first particles at lateral distance r from the axis is expected to be delayed with respect to planar shower front.
- We calculate the time delay and get a structure of shower front from the corrected time signal.
- Neutrino showers are in general much closer to the ground and hence the wavefront will be more curved as compared to hadron induced showers.

Features for identification





Total muon signal in an event

Number of stations per event with detectable muon and radio signal

$$S_b = \sum_i S_i \times \left(\frac{R_i}{3500\ \mathrm{m}}\right)^b$$, S_i is the measured muon signal at a distance R_i

Contact: a.khakurdikar@astro.ru.nl

Radboud Universiteit

Separation of neutrinos from Background

Dataset contains - E, Zenith, fit parameters, Muon signal, Number of stations and primary particle (neutrino or background (p, fe, he, n))

1. Define the features and the target variable:,

Features (X) -> Fit parameters, Sb, Nstat

Target variable (y) -> primary particle (Neutrino or Hadron)

2. Split the data into training and testing sets: 80% training, 20% testing

3. Random Forest Classifier: trains on the training data.

4. Use the trained model to make predictions on the testing set.

5. Calculate the accuracy of the model

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

Simulations available: (1.5 km grid)

Signal-like events: CoREAS showers

- Force Model: Sibyll2.3d
- Type: Electron Neutrino
- CC and NC interactions
- E = 1.0 - 120 EeV
- $\theta = 75^\circ - 85^\circ$, Uniform distribution in $\sin\theta\cos\theta$
- For varying Interaction lengths starting from 100 g/cm^2

Background-like events: CoREAS showers

- Primaries: proton, helium, nitrogen and iron
- E = 1.0 - 120 EeV
- $\theta = 75^\circ - 85^\circ$, Uniform distribution in $\sin\theta\cos\theta$

Contact: a.khakurdikar@astro.ru.nl

Radboud Universiteit

Separation of neutrinos from Background

The DecisionTreeClassifier in scikit-learn typically employs the CART (Classification and Regression Trees) algorithm.

Consider the features: 'a', 'b', 'c', ‘Sb’, ‘Nstat’ and the target variable ‘primary_particle’.

1. **Binary Tree Structure**: The CART algorithm starts with the entire dataset and selects the feature and threshold that minimize the Gini impurity for a binary split. Let's say the algorithm decides to split based on the ‘a' feature.

2. **Decision Node 1**: The decision node checks if 'a' is less than or equal to a certain threshold. If true, the instance goes to the left child node; otherwise, it goes to the right child node.

3. **Child Nodes**: The left child node might represent instances with 'a' <= Threshold, and the right child node represents instances with 'a' > Threshold. The algorithm repeats this process for each child node, selecting features and thresholds that minimize impurity.

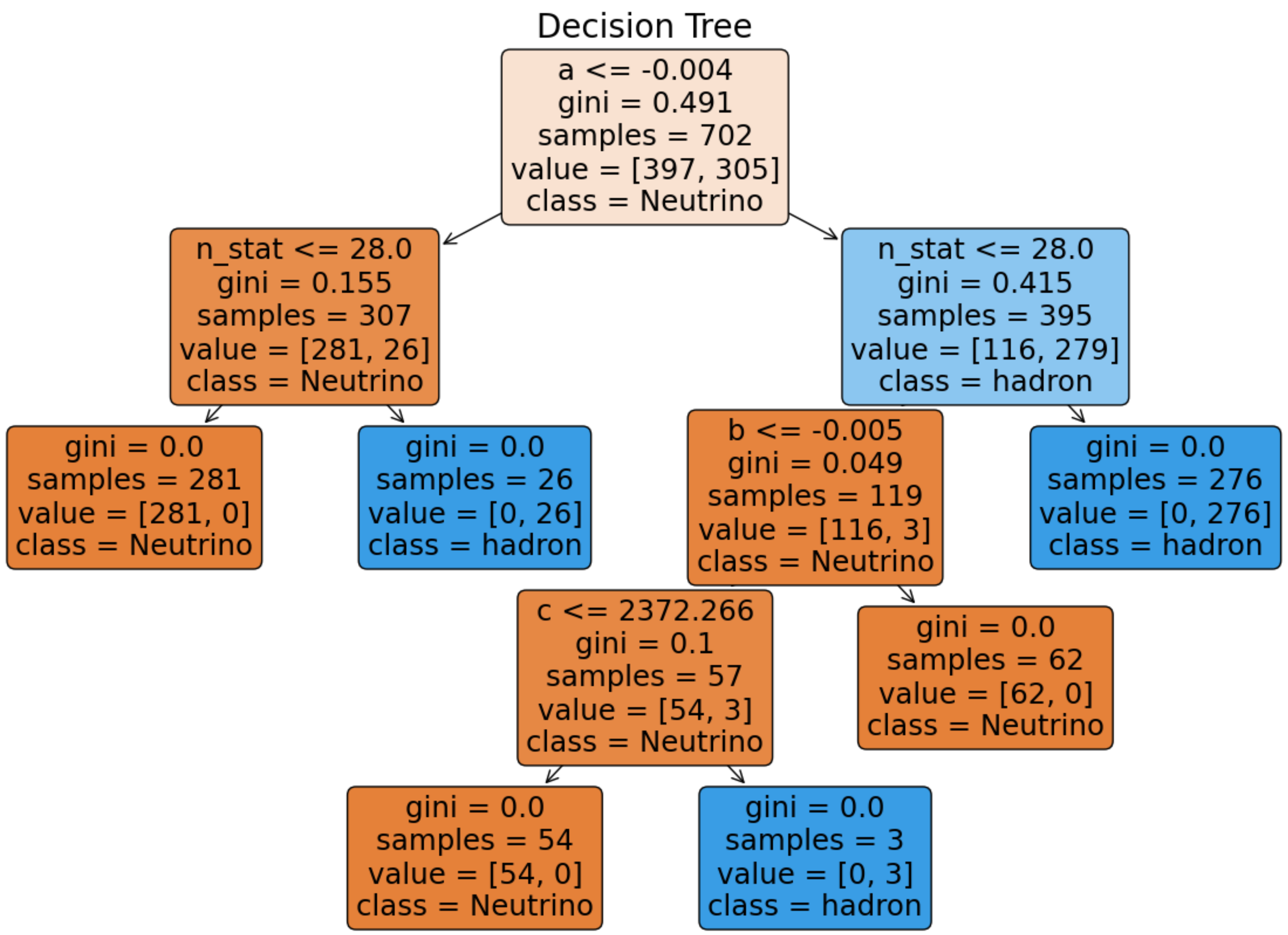
$X_i \leq$ threshold -> left child node, else -> right child node

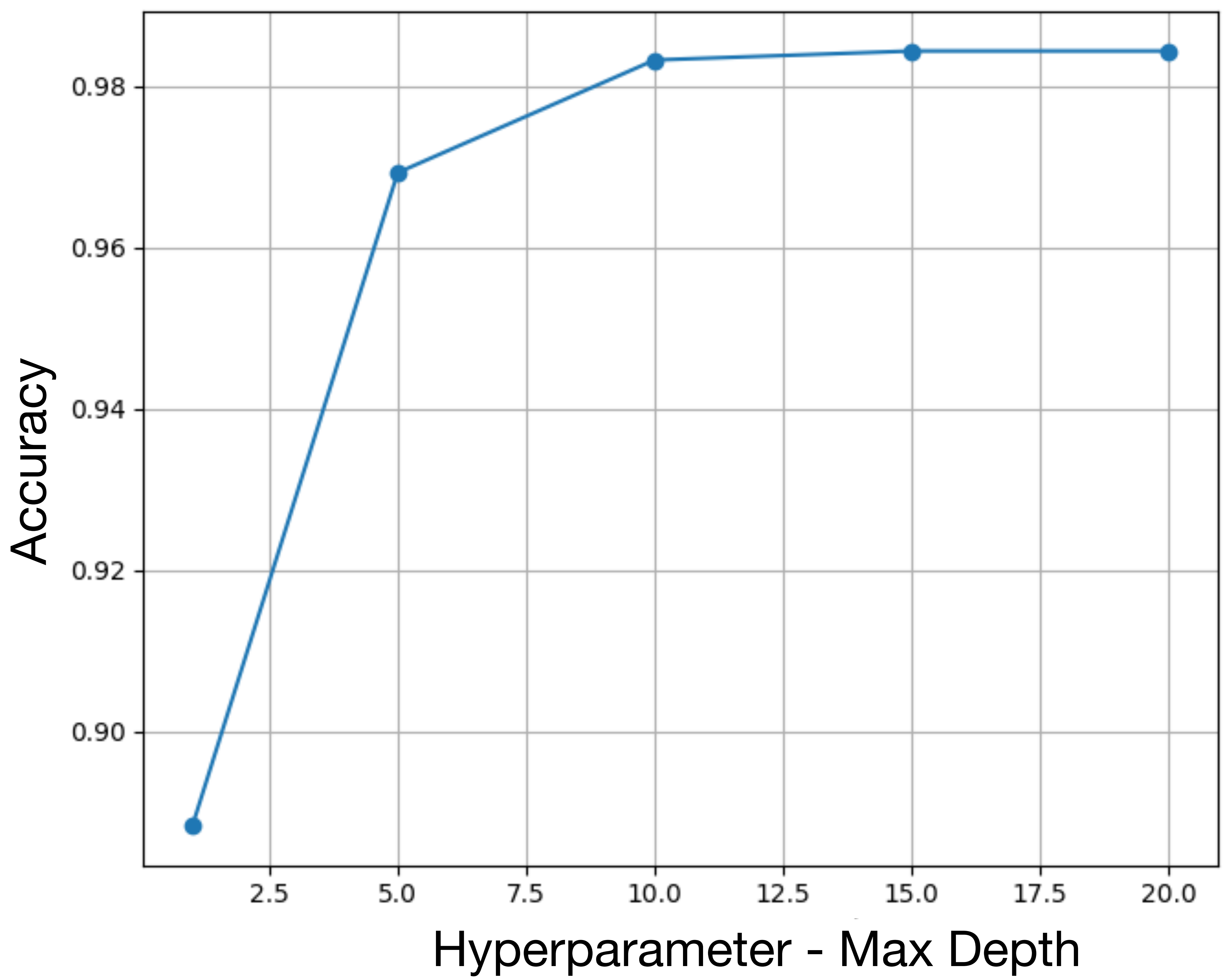
Contact: a.khakurdikar@astro.ru.nl

Radboud Universiteit

Separation of neutrinos from Background

4. **Tree Depth**: The process continues, creating decision nodes at each level of the tree. The depth of the tree depends on factors like the dataset and the stopping criteria (e.g., maximum depth, minimum samples per leaf).





Separation of neutrinos from Background

5. **Leaf Nodes**: When a stopping criterion is met (e.g., maximum depth reached), or if further splitting does not significantly reduce impurity, the algorithm creates leaf nodes. Each leaf node represents a predicted class label.

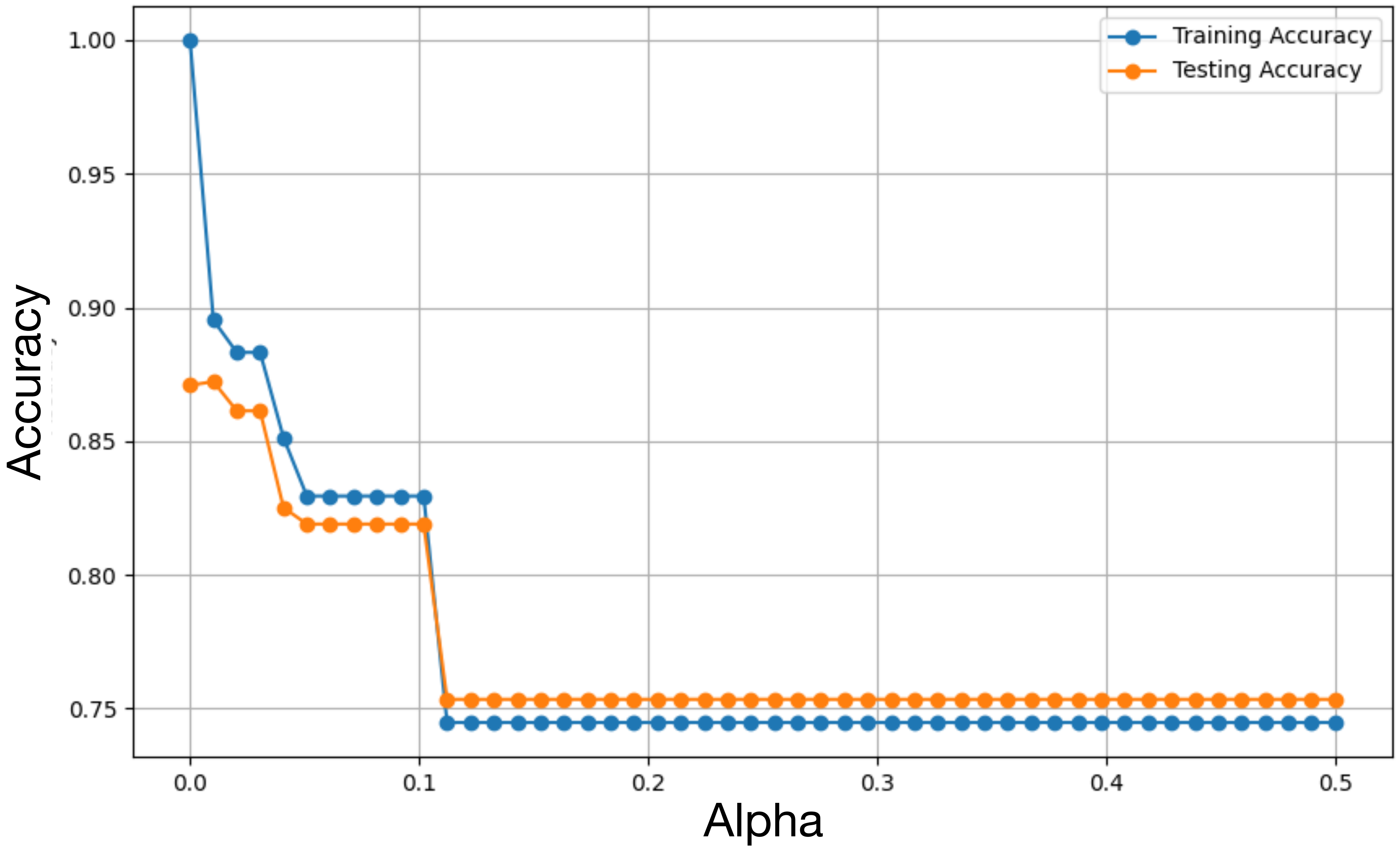
6. **Pruning**: After building the tree, the algorithm might prune branches that do not contribute significantly to impurity reduction. Pruning helps prevent overfitting, ensuring the model generalizes well to new data.

- $R_\alpha(T)$ is the cost complexity criterion of tree T with respect to parameter α

- $R(T)$ is the total impurity measure of tree T

- $|T|$ is the number if terminal nodes (leaves) of tree T

The alpha parameter controls the complexity of the decision tree. A higher value of alpha results in more aggressive pruning, leading to simpler trees with fewer nodes. Conversely, a lower value of alpha allows the tree to grow larger and potentially more complex.



Contact: a.khakurdikar@astro.ru.nl

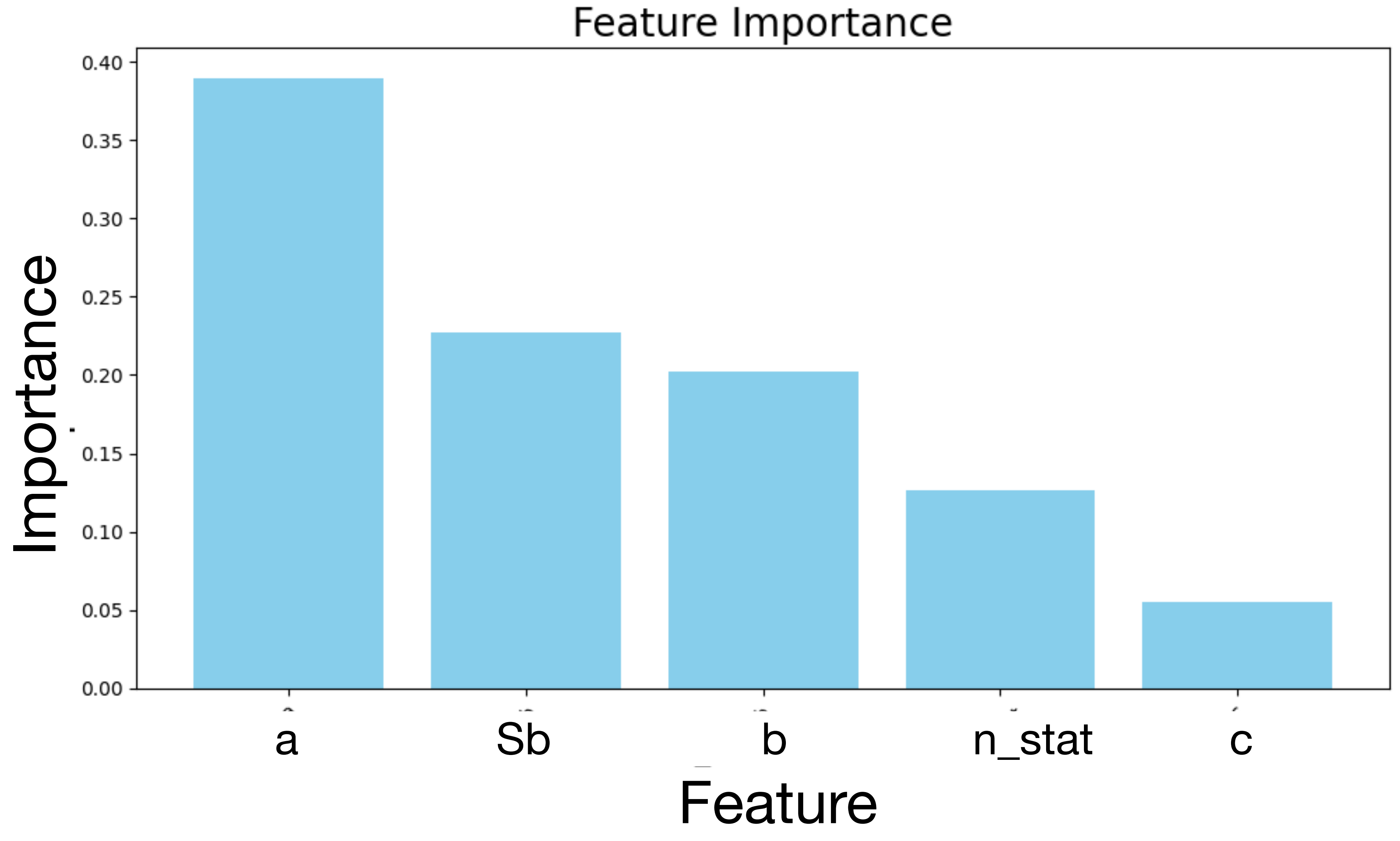
Radboud Universiteit

Separation of neutrinos from Background

7. **Feature Importance**: After training, the model can provide information about the importance of each feature. Feature importance is calculated based on how much each feature contributes to impurity reduction across all decision nodes. In this case, the final decision tree might have decision nodes based on different features and thresholds (not just 'dist'). The tree structure and decision rules would be determined by the specific features and conditions that minimize impurity during the training process. The resulting tree is a set of decision rules that collectively predict the 'primary_particle' class labels for new instances.

$$\text{Gini Importance (feature)} = \sum_{\text{nodes using feature}} \left(\frac{N_{node}}{N_{total}} \times \text{Gini}_{node} - \frac{N_{left}}{N_{node}} \times \text{Gini}_{left} - \frac{N_{right}}{N_{node}} \times \text{Gini}_{right} \right)$$

- N_{node} is the total number of samples in the node
- N_{left} and N_{right} are the number of samples in the left and right child nodes after the split.
- Gini_{node} is the Gini impurity of the node before the split
- Gini_{left} and Gini_{right} are the Gini impurities of the left and right child nodes after the split.



Contact: a.khakurdikar@astro.ru.nl

Radboud Universiteit

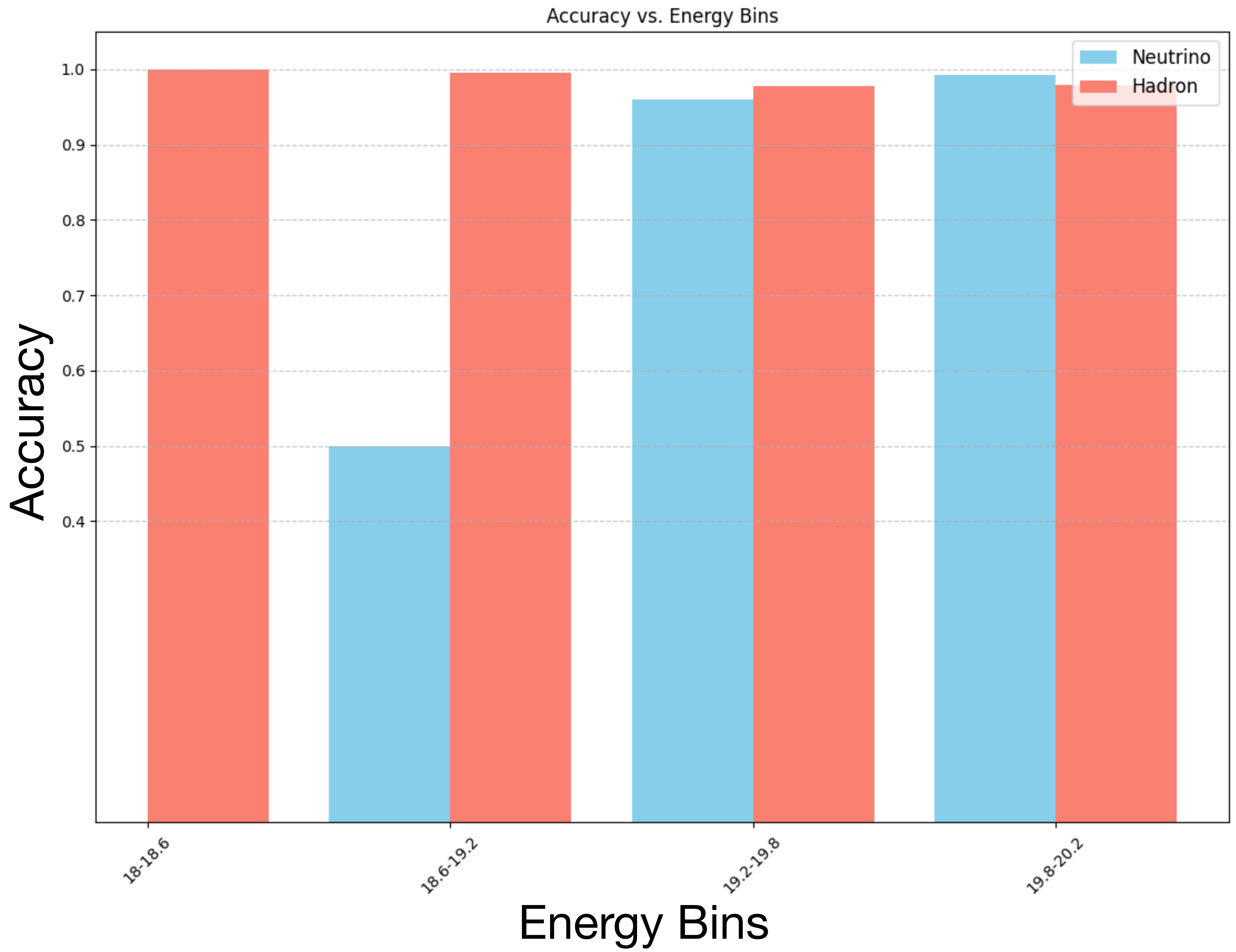
Separation of neutrinos from Background

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{True Positives} + \text{True Negatives} + \text{False Positives} + \text{False Negatives}}$$

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

1. **Elliptical parabola** (Ground plane - $ax^2 + by^2 + cx + dy + exy + f$) : Accuracy = **0.963**
(Features used: a,b,c,d,e,f, Sb, Nstat (min = 6))

2. **Hyperbola** (Shower plane - $a(x^2 - y^2) + bxy + c$): Accuracy = **0.832**
(Features used: a,b,c, Sb, Nstat (min = 4))



*final prediction is made by taking the ensemble of 100 individual trees and optimizing the hyper parameters

Contact: a.khakurdikar@astro.ru.nl

Radboud Universiteit

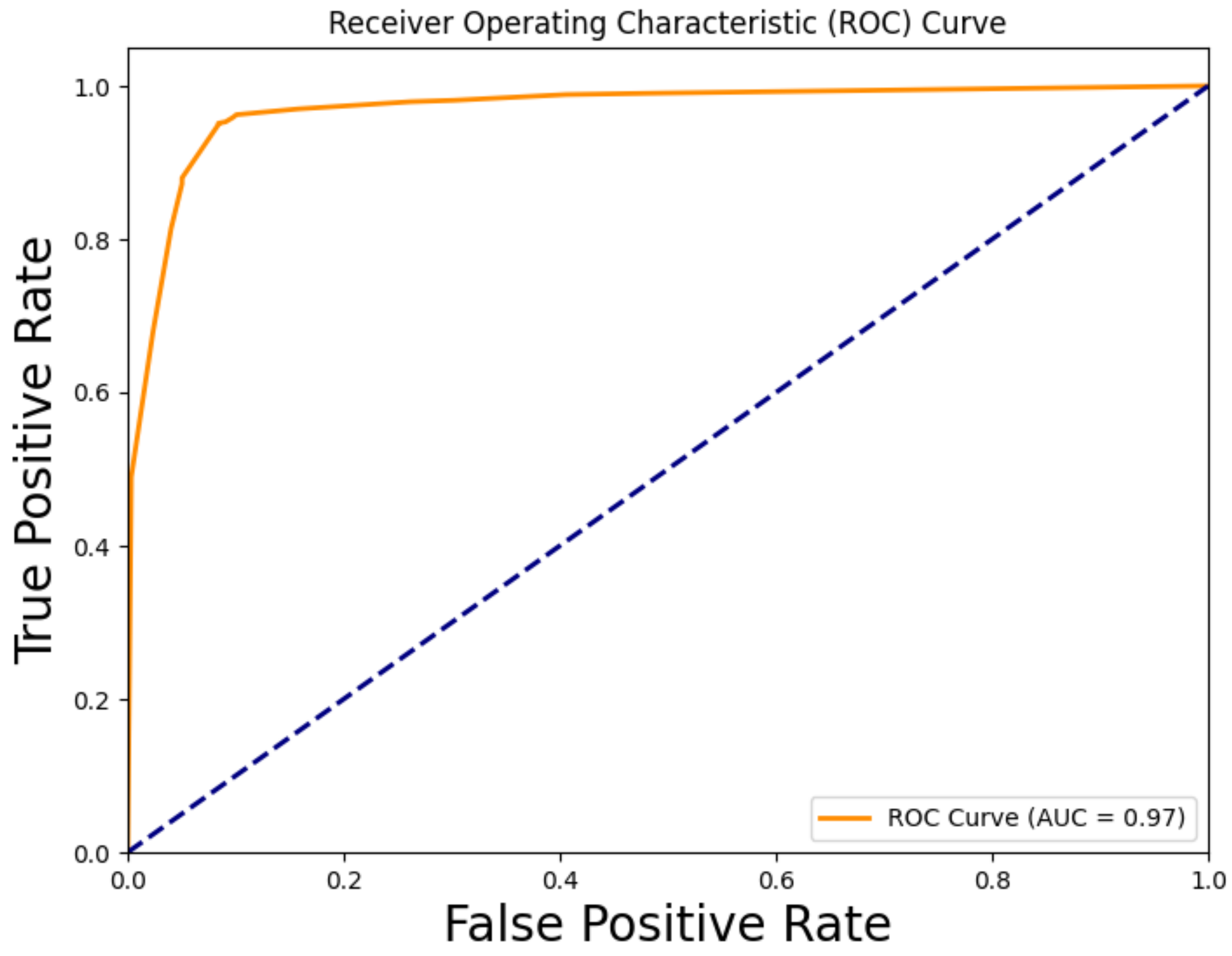
Separation of neutrinos from Background

- ROC curves are typically used for binary classification problems, where the target variable has two classes

- The area under the ROC curve (AUC) is a measure of the classifier's performance. A perfect classifier would have an AUC of 1.0, while a random classifier would have an AUC of 0.5. Generally, a higher AUC value indicates better classifier performance.

TPR - the ratio of true positives to the total number of actual positive instances. It measures the proportion of actual positive instances that are correctly identified by the classifier.

FPR - the ratio of false positives to the total number of actual negative instances. It measures the proportion of actual negative instances that are incorrectly classified as positive by the classifier.



$$TPR = \frac{\text{True Positives}}{\text{True Positives + False Negatives}} \qquad FPR = \frac{\text{False Positives}}{\text{False Positives + True Negatives}}$$

Summary

- Ultra-high energy neutrinos play a key role in understanding the origin of UHECRs.

- The main challenge in detecting the UHE neutrinos is to identify a neutrino-induced shower (signal) in the background of showers initiated by UHERCRs.

- Geometry of the radio footprint of the showers is the main difference in the hadron- induced showers and the neutrino-induced showers

- Supervised Learning tool, such as Random Tree Classifier help to classify the showers induced by neutrinos and hadrons.

Contact: a.khakurdikar@astro.ru.nl

Radboud Universiteit

Back-up

t_o: time at which the signal was found, relative to the event start time

$$r_{ant} = \sqrt{x_{ant}^2 + y_{ant}^2}$$

$$\phi_{ant} = \text{atan2(y , x)}$$

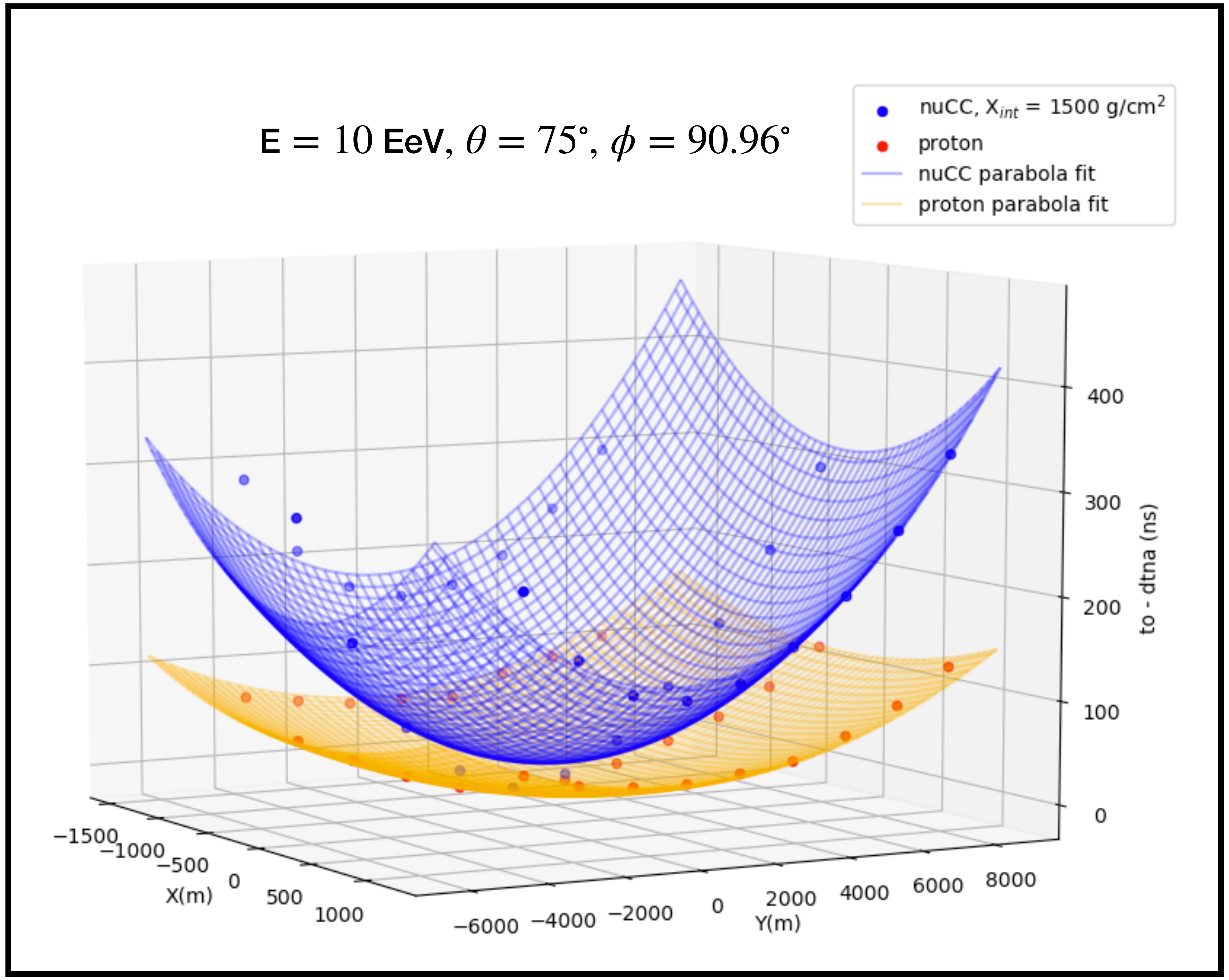
$$\text{angle} = |\,\phi_{MC} - \phi_{ant}\,|$$

$$r_{proj} = r_{ant} * \cos(\text{angle})$$

Delay in time due to the curvature of the air shower front

$$dtna = -\,(r_{proj}\,\sin(\theta_{MC}))\ /\ c$$

- We calculate the time delay and get a structure of shower front from the corrected time signal.



Corrected time signal with respect to the distance from the shower core

Contact: a.khakurdikar@astro.ru.nl

Radboud Universiteit

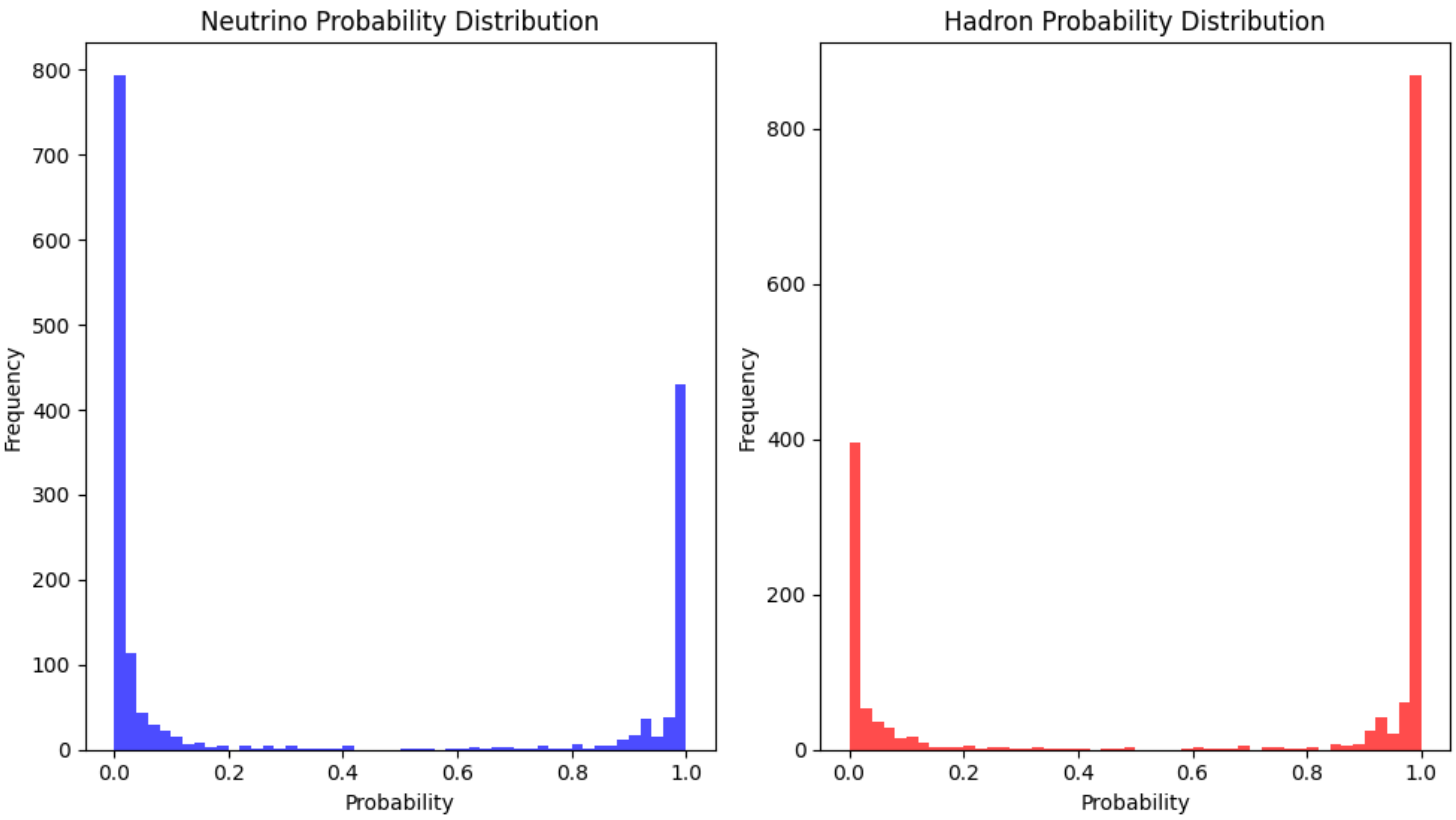
Back-up

If an instance falls into a leaf node N, the predicted probability p of belonging to class 1 is:

$$p = \frac{\text{count}(N,\ \text{class} = 1)}{\text{count}(N)}$$

Where:

- count (N, class = 1) is the number of instances in node N that belongs to class 1
- count (N) is the total number of instances in node N



Contact: a.khakurdikar@astro.ru.nl

Radboud Universiteit