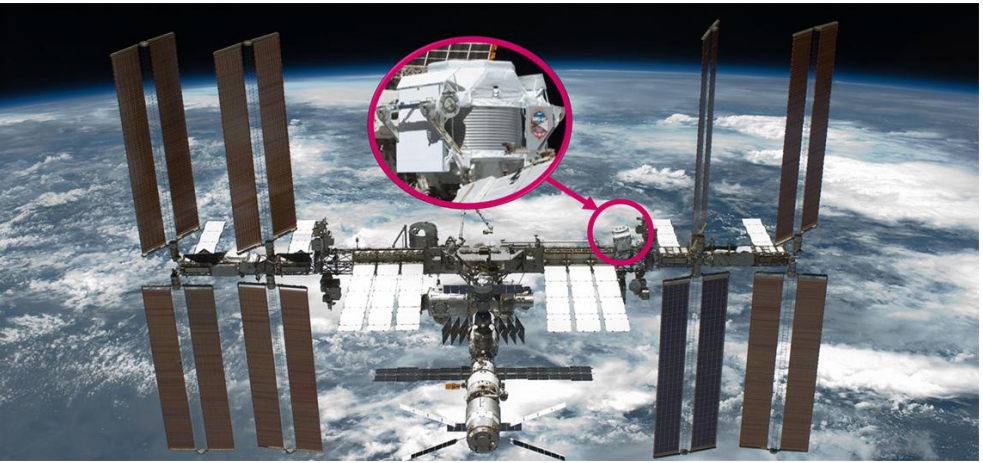
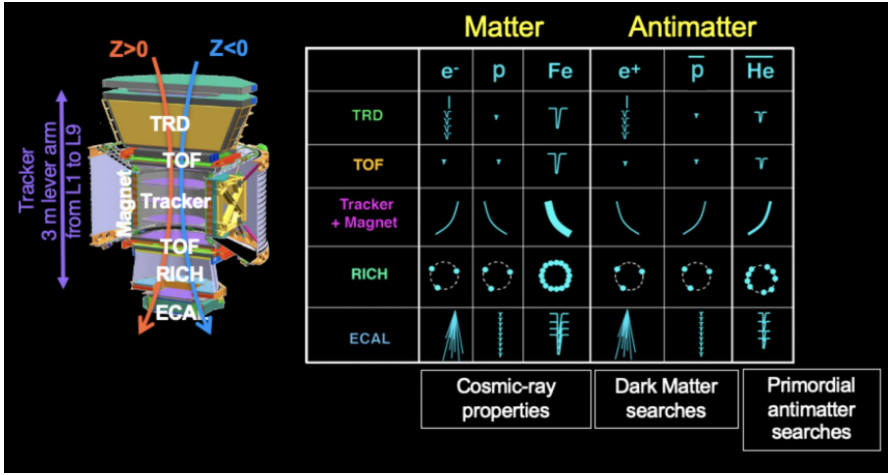
Machine-learning analysis of cosmic-ray nuclei data from the AMS-02 experiment on the International Space Station (ISS)

Shahid Khan



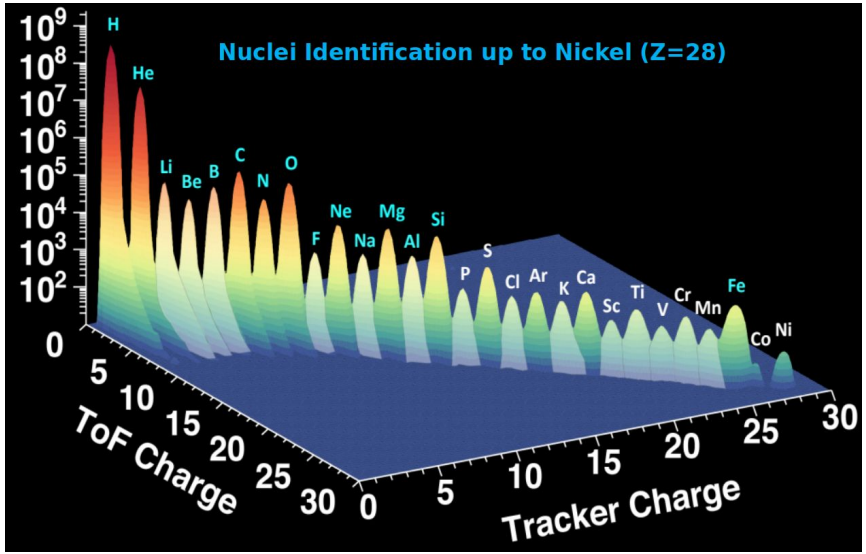
~7000 kg AMS-02 on the ISS



AMS-02 detector components and their use cases for the identification of various charges

Data and MC Mixture

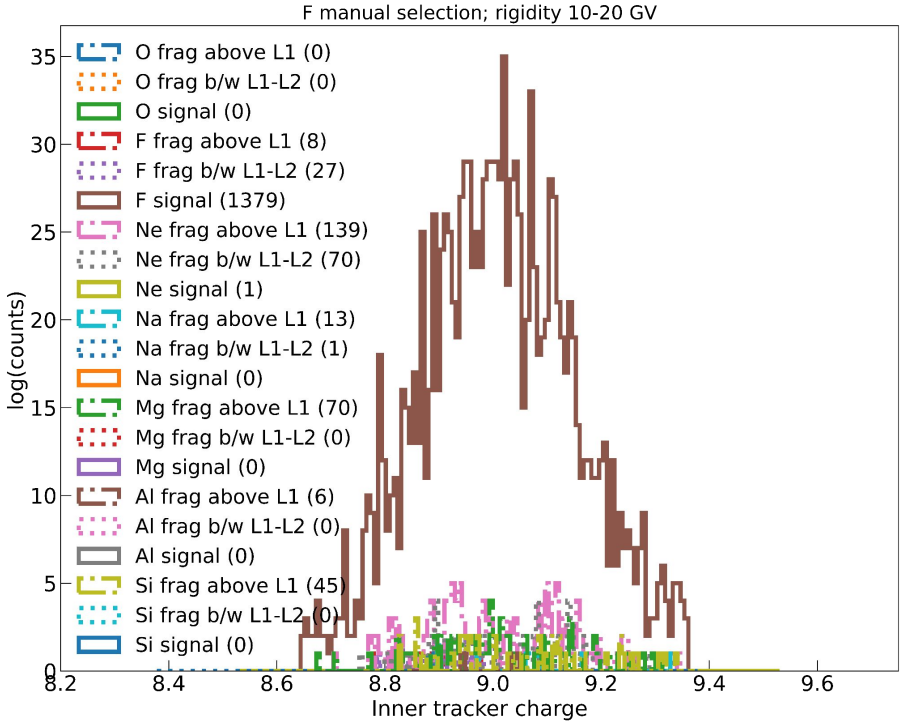
- Around 100 variables from tracker, TOF and TRD
- MC simulations and AMS-02 ISS experimental data with tight manual selections applied have been mixed to create train (equal abundance of all nuclei) and test (natural abundance of nuclei) data sets
- For each nuclei species 3 different classes are created
    - Fragmented above L1
    - Fragmented between L1 and L2
    - Non-fragmented and those which fragment between L2 and L8
- Benchmarking with ML algorithms: Multi-layered perception, Convolutional Neural Networks, Transformers, and XGBoost
- XGBoost showed better performance because of the tabular nature of the data. source:



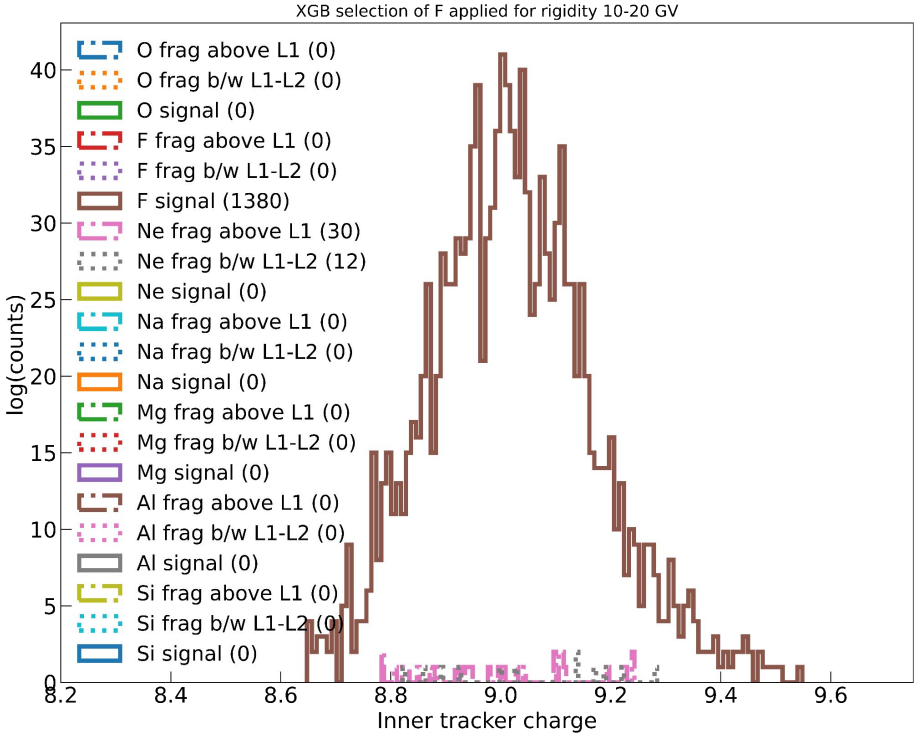
Nuclei identification with tracker, TOF

Comparison of fragments in the selections

- The manually selected Fluorine candidates have contamination from heavier nuclei
- For ML selected Fluorine the contamination mostly comes from the neighboring Ne



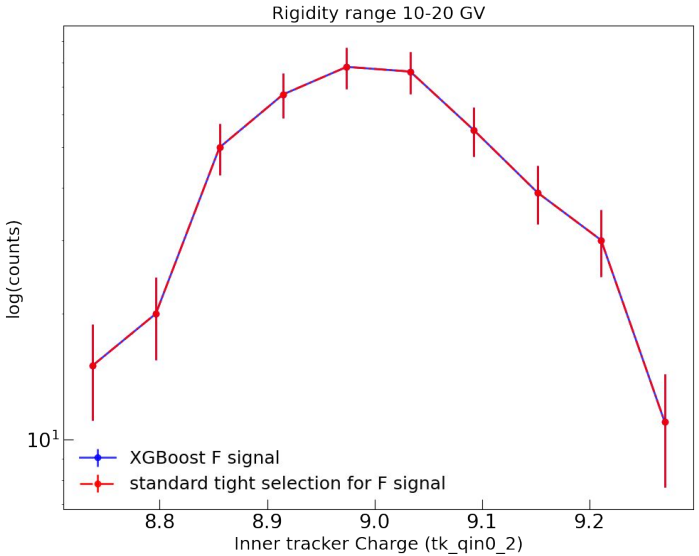
Manually optimized selection



XGBoost optimized selection

Model deployment on AMS-02 ISS data

- Test data, applied tight standard selection on AMS-02 experimental data from ISS, is taken here
- ML selections are applied and the performance of hypercube selection and ML selection is compared here



Summary

- ML performance was tested for nuclei selection of the AMS-02 simulated data and ISS data with tight selections applied
- Comparison between standard AMS-02 standard selection and ML model for Fluorine
  - At the same efficiency of standard selection, higher purity was achieved with ML model
  - ML based selection can be varied easily and is non-linear

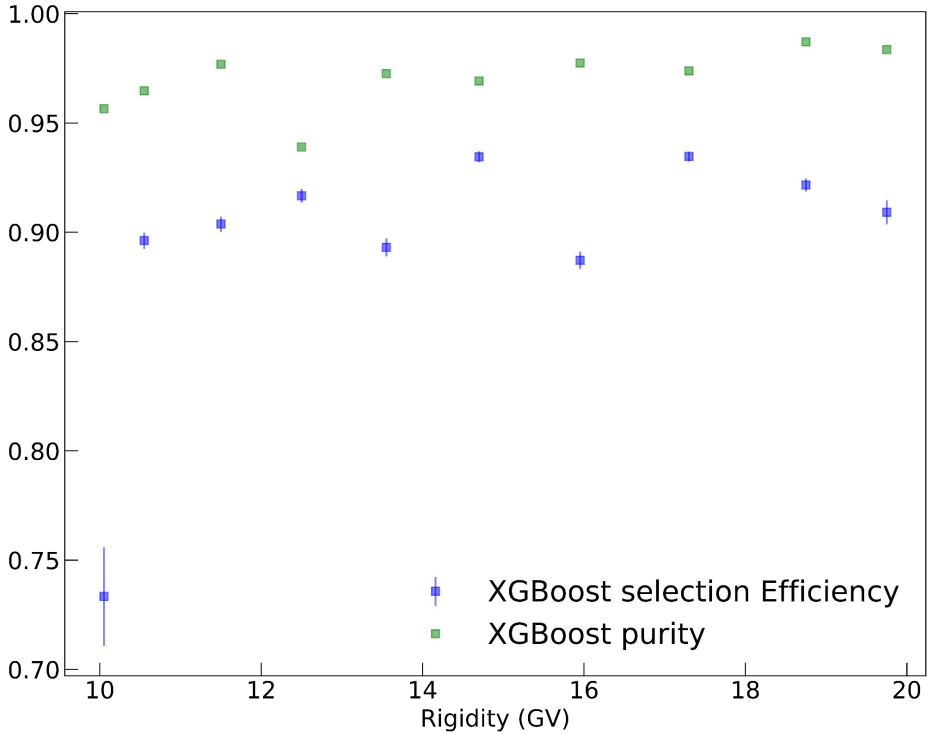
Outlook

- Efficiency correction for selection
- Systematic evaluation

Poster on Thursday

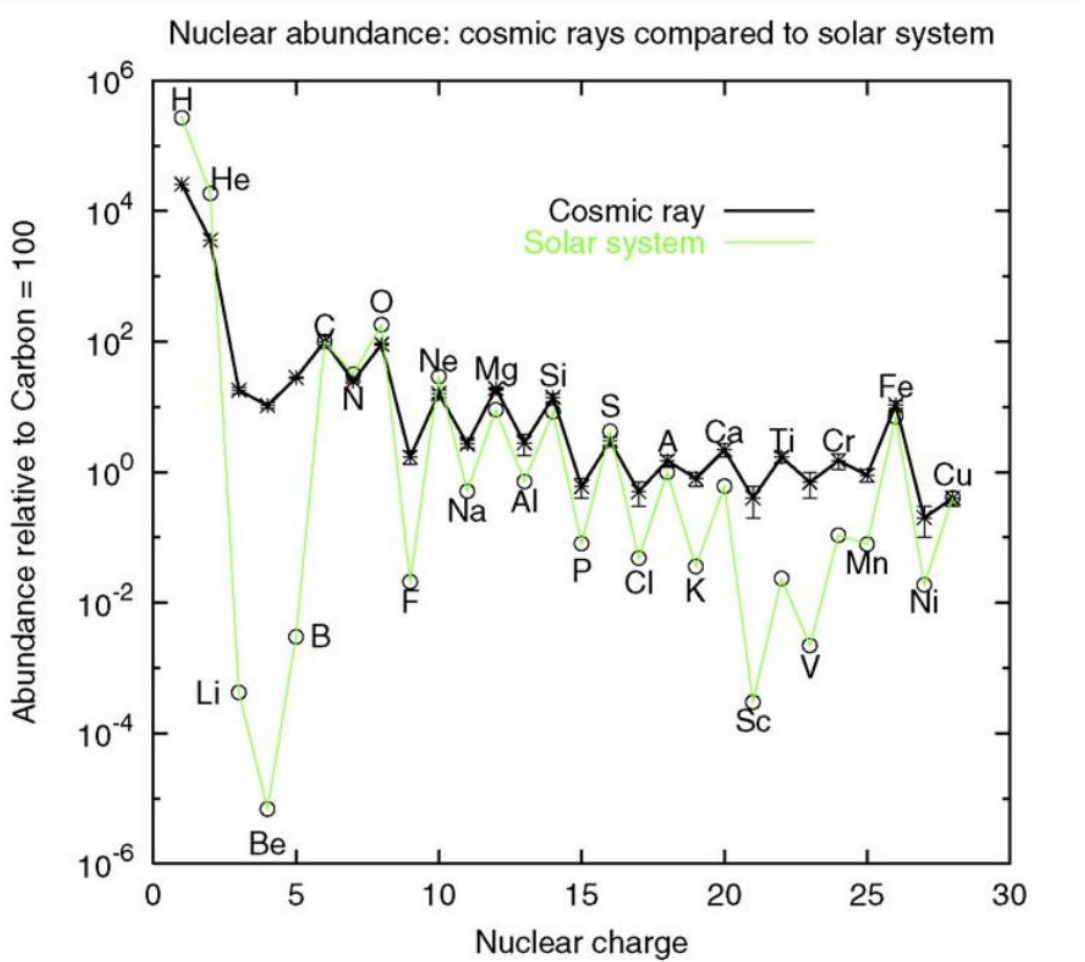
Back up

Bin wise visualization



$$Efficiency = \varepsilon = \frac{signal\ after\ selection}{signal\ before\ selection}$$

$$purity = \frac{signal}{signal + background}$$



https://doi.org/https://doi.org/10.1016/j.nuclphysa.2005.01.024