



# UNIVERSITY OF TRENTO

DOCTORATE OF NATIONAL INTEREST ON SPACE SCIENCE AND  
TECHNOLOGY

PHYSICS DEPARTMENT, UNIVERSITY OF TRENTO

---

## Combined Machine Learning Analysis for $\overline{He}$ Search with the AMS-02 Experiment

---

*Author*

Francesco ROSSI

*Supervisor*

Professor Paolo ZUCCON

June 16, 2026







# Declaration of Autorship

I hereby declare that, except where specific reference is made to the work of others, the contents of this thesis are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text.

Francesco Rossi

June, 2026



# Abstract

Heavy anti-nuclei, such as  $\overline{He}$ , are unlikely to be formed during cosmic rays (CRs) propagation, as supported by measurements from the PHENIX and ALICE collaborations. The precise measurement of cosmic-ray antinuclei is a powerful probe for possible identification of Dark Matter (DM) signatures, and in combination with measurements of other cosmic-ray species, can help constrain cosmic-ray production and propagation in the Galaxy. The first-time discovery of cosmic-ray  $\overline{He}$  has been identified as a potentially promising detection signature related to DM interactions or to a primordial origin, due to its extremely low astrophysical backgrounds, especially towards low energies.

This thesis presents an  $\overline{He}$  search, performed using 12.5 years of data collected by the AMS-02 experiment. Initially, an event selection for events with  $|Z| = 2$  is applied, after requiring a negative reconstructed rigidity, this analysis exploits four Deep Neural Networks (DNNs), two supervised classifiers and two unsupervised autoencoders, to search for possible  $\overline{He}$  candidates in data. The DNNs are independent and orthogonal, allowing different working points to be combined to improve background suppression. The analysis follows a fully data-driven workflow while applying the same strategy independently to data and Monte Carlo (MC) simulation.

The dominant background consists of charge-confused helium nuclei, which is resonant in the same mass region as the possible signal. As a consequence, no control region can be defined directly from data, and the background estimate relies entirely on MC simulation. The search for  $\overline{He}$  is performed by looking for an excess in the number of events with a reconstructed mass in the range  $[2, 5] \text{ GeV}/c^2$ , compatible with  ${}^3\text{He}$  and  ${}^4\text{He}$  masses, within  $E_K/N \in [0.65, 31.4] \text{ GeV}/N$  for  $\overline{{}^3\text{He}}$  and  $E_K/N \in [0.41, 23.3] \text{ GeV}/N$  for  $\overline{{}^4\text{He}}$ . Up to 12  $\overline{He}$  candidates have been found, resulting in a mild tension with the background-only hypothesis at the level of  $1\text{-}2\sigma$ , depending on the chosen working points on the DNNs outputs. No statistically significant excess has been found; therefore, two exclusion limits using the CLs method on the  $\frac{\overline{{}^3\text{He}}}{\overline{He}}$  and  $\frac{\overline{{}^4\text{He}}}{\overline{He}}$  ratios are presented and discussed.



# Acknowledgments

I would like to express my gratitude to my supervisor, Prof. Paolo Zuccon, who was always available to discuss the latest developments of this work. Throughout my PhD, he provided constructive suggestions and generously shared his expertise in both software development and detector physics, drawing on his long-standing experience within the AMS-02 collaboration. Equally important, he granted me a high degree of autonomy and freedom, allowing me to develop and pursue my own research ideas.

I am also deeply grateful to many members of the AMS-02 collaboration who contributed to the development of this thesis. In particular, I would like to thank Dr Vitali Choutko and the members of the CIEMAT group: Dr Jorge Casaus, Dr Francesca Giovacchini, Dr Miguel Ángel Velasco, and Iñaki Rodríguez. They closely followed the progress of the analysis and helped me achieve substantial improvements through their insightful comments and expertise. Dr Vitali Choutko deserves special recognition as he was always willing to discuss the status of the analysis, helping to identify subtle issues, refine the methodology, and improve the quality of the work through his critical insights. Through the time and knowledge generously shared by these researchers, I have gained valuable experience in data-analysis techniques, cosmic-ray physics, and AMS-02.

In addition to all the people who have supported me throughout my PhD, I would also like to extend my sincere gratitude to those who, without any apparent reason, attempted to hinder this journey. Thanks to them, I came to appreciate the truth of the saying, "the best opportunities arise in times of crisis," which, in my experience, has proven remarkably accurate.

Last but certainly not least, I would like to thank my wife, Greta Brianti. I am deeply grateful for her unwavering support throughout my PhD, for the countless discussions about the work presented in this thesis, and for the innumerable occasions on which she helped me navigate both academic and personal challenges. Most importantly, she always believed in me, especially when my own confidence wavered. Words cannot adequately express the depth of my gratitude towards her. Beyond being one of the most brilliant researchers I have ever met, she possesses a rare combination of adventurism, bravery, and creativity that I deeply admire. Every day, I am grateful to share my life with such an unbowed woman, always ready to face new challenges, pooling ideas, and continuously learning together.



# Contents

|          |                                                                          |           |
|----------|--------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Review of antimatter searches in space</b>                            | <b>11</b> |
| 1.1      | The Standard Model of particle physics . . . . .                         | 11        |
| 1.2      | Introduction to observational cosmology . . . . .                        | 13        |
| 1.2.1    | General Relativity and the Friedmann equations . . . . .                 | 13        |
| 1.2.2    | Solution of the Friedmann equations . . . . .                            | 16        |
| 1.2.3    | The Big Bang model and a brief thermal history of the Universe . . . . . | 16        |
| 1.3      | Big Bang Nucleosynthesis . . . . .                                       | 18        |
| 1.4      | Matter Antimatter asymmetry . . . . .                                    | 23        |
| 1.4.1    | Baryogenesis and the Sakharov principles . . . . .                       | 23        |
| 1.5      | Antinuclei formation . . . . .                                           | 26        |
| 1.5.1    | Coalescence . . . . .                                                    | 27        |
| 1.5.2    | Thermal model . . . . .                                                  | 28        |
| 1.6      | Recent searches for cosmic antinuclei . . . . .                          | 29        |
| <b>2</b> | <b>The Alpha Magnetic Spectrometer, AMS-02</b>                           | <b>34</b> |
| 2.1      | The AMS-02 detector on the ISS . . . . .                                 | 35        |
| 2.1.1    | The Permanent Magnet . . . . .                                           | 36        |
| 2.1.2    | Silicon Tracker Detector . . . . .                                       | 38        |
| 2.1.3    | Time-of-Flight System . . . . .                                          | 45        |
| 2.1.4    | Anti Coincidence Counter . . . . .                                       | 47        |
| 2.1.5    | Rich Imaging Cherenkov Detector . . . . .                                | 48        |
| 2.1.6    | Transition Radiation Detector . . . . .                                  | 49        |
| 2.1.7    | Electromagnetic Calorimeter . . . . .                                    | 52        |
| 2.2      | The velocity resolution of AMS-02 . . . . .                              | 53        |
| 2.3      | The mass resolution of AMS-02 . . . . .                                  | 57        |
| 2.4      | The Trigger Logic . . . . .                                              | 59        |
| 2.4.1    | Fast Trigger . . . . .                                                   | 59        |
| 2.4.2    | The Level-1 Trigger . . . . .                                            | 59        |
| 2.4.3    | Data Acquisition System . . . . .                                        | 60        |
| 2.4.4    | Flight data, livetime and trigger rate . . . . .                         | 61        |

|          |                                                                                     |            |
|----------|-------------------------------------------------------------------------------------|------------|
| <b>3</b> | <b>Event selection and Monte Carlo simulation</b>                                   | <b>64</b>  |
| 3.1      | Event selection for identification of $He$ events . . . . .                         | 64         |
| 3.1.1    | Geomagnetic environment, timestamp RTI and trigger requirements . . . .             | 64         |
| 3.1.2    | Track pattern and number of reconstructed tracks . . . . .                          | 65         |
| 3.1.3    | Tracker charge identification . . . . .                                             | 66         |
| 3.1.4    | Reconstructed track quality selection . . . . .                                     | 68         |
| 3.1.5    | Coherence between different reconstructed rigidities . . . . .                      | 68         |
| 3.1.6    | Event geometry and TOF reconstruction . . . . .                                     | 68         |
| 3.1.7    | TOF charge identification . . . . .                                                 | 70         |
| 3.1.8    | Requirements on the reconstructed velocity . . . . .                                | 70         |
| 3.1.9    | ISS data selection summary . . . . .                                                | 72         |
| 3.2      | Monte Carlo simulation . . . . .                                                    | 75         |
| 3.2.1    | Monte Carlo normalisation . . . . .                                                 | 75         |
| <b>4</b> | <b>Overview of the AMS-02 data analysis for <math>\overline{He}</math> searches</b> | <b>80</b>  |
| 4.1      | Signal and expected background sources . . . . .                                    | 80         |
| 4.1.1    | Expected backgrounds . . . . .                                                      | 83         |
| 4.2      | Introduction to Machine Learning . . . . .                                          | 86         |
| 4.2.1    | Feed-Forward Multi-Layer Perceptron . . . . .                                       | 88         |
| 4.2.2    | Optimisation of the weights and back propagation . . . . .                          | 89         |
| 4.2.3    | Supervised learning and classification task . . . . .                               | 90         |
| 4.2.4    | Unsupervised learning and anomaly detection . . . . .                               | 91         |
| 4.3      | Supervised Classifier . . . . .                                                     | 92         |
| 4.3.1    | Task description and label definition . . . . .                                     | 92         |
| 4.3.2    | Input variables studies . . . . .                                                   | 95         |
| 4.3.3    | Dataset preprocessing, training and validation samples . . . . .                    | 98         |
| 4.3.4    | Network structure and performance . . . . .                                         | 99         |
| 4.4      | Unsupervised Autoencoder . . . . .                                                  | 107        |
| 4.4.1    | Task description . . . . .                                                          | 107        |
| 4.4.2    | Input variables studies . . . . .                                                   | 108        |
| 4.4.3    | Dataset preprocessing, training and validation samples . . . . .                    | 111        |
| 4.4.4    | Network structure and performance . . . . .                                         | 111        |
| 4.5      | Candidate identification strategy . . . . .                                         | 115        |
| 4.6      | Orthogonality . . . . .                                                             | 120        |
| <b>5</b> | <b>Results</b>                                                                      | <b>123</b> |
| 5.1      | Signal and control region definition . . . . .                                      | 123        |
| 5.1.1    | Expected electron background . . . . .                                              | 126        |
| 5.2      | The CLs method and the likelihood definition . . . . .                              | 129        |
| 5.2.1    | Likelihood definition . . . . .                                                     | 130        |



|          |                                                       |            |
|----------|-------------------------------------------------------|------------|
| 5.2.2    | Implementation of the CLs method . . . . .            | 131        |
| 5.3      | Exclusion limits on the signal yield . . . . .        | 133        |
| 5.3.1    | TOF . . . . .                                         | 134        |
| 5.3.2    | RICH-NaF . . . . .                                    | 136        |
| 5.3.3    | RICH-AGL . . . . .                                    | 138        |
| 5.4      | Exclusion limits on the ratio anti-He/He . . . . .    | 142        |
| 5.5      | Candidate events . . . . .                            | 145        |
| <b>6</b> | <b>Conclusions</b>                                    | <b>149</b> |
| <b>A</b> | <b>Deep Neural Networks variables</b>                 | <b>152</b> |
| A.1      | Supervised classifiers . . . . .                      | 152        |
| A.2      | Unsupervised Autoencoders . . . . .                   | 159        |
| <b>B</b> | <b>Deep Neural Networks outputs and orthogonality</b> | <b>163</b> |
| B.1      | Supervised classifiers . . . . .                      | 163        |
| B.2      | Unsupervised Autoencoders . . . . .                   | 178        |
| B.3      | Orthogonality . . . . .                               | 186        |
| <b>C</b> | <b>Expected electron background</b>                   | <b>196</b> |
| <b>D</b> | <b>Exclusion limits on the ratio anti-He/He</b>       | <b>199</b> |

# Chapter 1

## Review of antimatter searches in space

The theoretical motivations behind the work discussed in this Thesis are presented in this Chapter. After a brief introduction to the Standard Model of particle physics (reported from [1]) and observational Cosmology, I focus on the Big Bang model and on the Big Bang Nucleosynthesis to outline the experimental evidence on how light nuclei were formed during the first minutes of the early Universe. I briefly discuss the matter-antimatter asymmetry puzzle and the Sakharov principles. Afterwards, I expand on the formation mechanisms of antinuclei and on the comparison between some recent estimates of the expected antihelium fluxes near the Earth and the corresponding (expected) sensitivity of the AMS-02 experiment.

### 1.1 The Standard Model of particle physics

The Standard Model (SM) is a quantum field theory that describes the elementary constituents of matter and their fundamental interactions (electromagnetic, weak and strong interactions, while gravity is not included) [2, 3, 4, 5, 6, 7]. Particles are represented as excitation of quantum fields that interact with each other: the SM describes their dynamics at energy scales relevant for the subatomic world. Particles and their interactions are described in a Lagrangian formalism. Local gauge symmetry, i.e. the invariance of the Lagrangian under space-time-dependent transformations applied to the phases of fields, is a key concept. Interaction terms appear in the free-field Lagrangian after requiring it to be invariant under local gauge symmetries. The SM is based on the symmetry group

$$SU_C(3) \otimes SU_L(2) \otimes U_Y(1),$$

where  $SU_C(3)$  is the standard unitary group that describes the strong interactions (quantum chromodynamics, QCD), and  $C$  stands for the color charge;  $SU_L(2) \otimes U_Y(1)$  is the product of groups that describe the combination of the weak and electromagnetic interactions (electroweak, EW),

with  $L$  standing for *left* and  $Y$  standing for *hypercharge*.<sup>1</sup> Spin-1 particles, called *gauge bosons* as they correspond to the generators of the symmetry group, mediate the interactions. Strong interactions are mediated by eight massless particles called *gluons*: they carry a charge called *colour*, which can be of three kinds (red, green or blue). Weak interactions are mediated by two charged massive bosons,  $W^\pm$ , and a neutral massive boson,  $Z^0$ . Electromagnetic interactions occur between particles carrying electric charge and are mediated by a neutral massless boson, the photon  $\gamma$  [1].

Particles acquire mass via the interaction with the Higgs field, which is mediated by a spin-0 particle, the Higgs boson. The Higgs potential has a typical mexican-hat shape, where the potential minimum does not preserve the global symmetry of the Lagrangian, causing a spontaneous symmetry breaking. The vacuum state is found to be charged under the generators of the SM Lagrangian symmetry group. The interaction with this charged ground state forces the particles to acquire a mass [5, 6].

Matter particles correspond to excitations of spin- $\frac{1}{2}$  fields and are called *fermions*. Their masses, or, equivalently, their couplings with the Higgs field, are free parameters. Each fermion is also associated with an antiparticle of the same mass but opposite internal quantum numbers. Fermions are further classified into two classes, quarks and leptons, each organised in three weak-isospin doublets. Quark doublets are composed of an up-type quark, with charge  $\frac{2}{3}e$ , and a down-type quark, with charge  $-\frac{1}{3}e$ ,

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix} .$$

They couple with both the strong and electroweak interactions. Each quark has color and it is distinguished by its *flavour*, a quantum number which comes in six varieties and which is conserved in the electromagnetic and strong interactions but not in the weak interactions. Due to colour confinement, free quarks are not observable. They are only observed in their colourless bound states, which include mesons, composed of a quark and an anti-quark, and baryons, composed of three quarks (or three antiquarks). Baryons are assigned a quantum number, called baryon number, found to be conserved even if no symmetry of the Lagrangian requires that [1].

Lepton doublets are composed of each by a massless neutral particle, called neutrino, and a massive particle with electric charge  $-e$ ;

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} .$$

They couple only with the electroweak interaction. Each lepton has a lepton family quantum number; their sum in a process, called global lepton number, is found to be conserved in all interactions, although no symmetry of the dynamics prescribes that. Individual lepton numbers are not conserved only for neutrinos that can oscillate between states of different lepton flavour. Gauge bosons, quarks, and leptons constitute all the fundamental interactions and particles of the

---

<sup>1</sup>Only particles with *left* chirality are influenced by the weak interaction.

SM, which are sketched in Figure 1.1 [1].

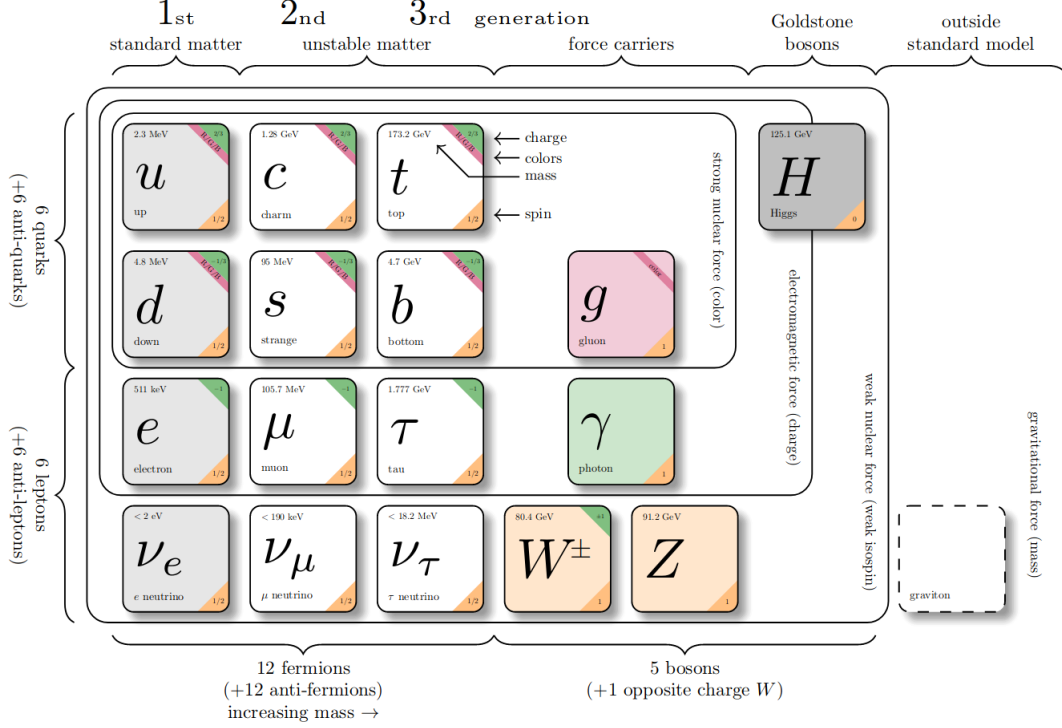


Figure 1.1: Scheme of particles and interactions in the Standard Model.

## 1.2 Introduction to observational cosmology

The current description of the Big Bang Nucleosynthesis (BBN) assumes spacetime characterised by General Relativity, the  $\Lambda$ CDM cosmology, and microphysics characterised by the Standard Model particles and interactions, with 3 light neutrino species [8]. In this section, I will briefly introduce some notation and key concepts related to the current knowledge of the BBN.

### 1.2.1 General Relativity and the Friedmann equations

Our Universe, observed at the cosmological scale  $\geq 10\text{-}10^2$  Mpc, is homogeneous and isotropic. These observations lead to the Cosmological Principle, stating that all spatial positions in the Universe are equivalent. Homogeneity and isotropy lead to a space-time metric of Robertson-Walker (adopting  $c = 1$  throughout):

$$ds^2 = dt^2 - R^2(t) \left[ \frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2) \right] \quad (1.1)$$

Rescaling the radial coordinate allows the  $k$  parameter to take only discrete values, i.e.  $+1$   $-1$  or  $0$  corresponding to a closed, open or flat geometry of our Universe. The coordinate  $r$  is dimensionless, since the dimensions are carried by the cosmological scale factor  $R(t)$ . The latter determines proper distances in terms of comoving coordinates. An alternative notation is  $a(t) =$

$\frac{R(t)}{R_0}$ , where  $R_0 = R(t_0)$  is the present scale factor.

Another important concept is the cosmological redshift, a direct consequence of the Hubble expansion [9]. The Hubble law states that the velocity of a distant object drifting away from the observer increases with its distance:

$$v = H_0 D \quad (1.2)$$

where  $v$  is the recessional velocity,  $H_0$  is the Hubble parameter at the current time, and  $D$  is the proper distance from the galaxy to the observer. As seen in Figure 1.2, this is experimental evidence of an expanding Universe. A local observer, detecting light from a distant source, sees a redshift in frequency. Starting from equation 1.1, it is possible to write:

$$1 + z = \frac{R_S}{R_O} = \frac{a(t_S)}{a(t_O)} \quad (1.3)$$

where  $z$  is the redshift,  $R_S(a(t_S))$  and  $R_O(a(t_O))$  are the (dimensionless) scale factor at the source and at the observer position.

The cosmological equations of motion are derived from Einstein's equation [10]:

$$\mathcal{R}_{\mu\nu} - \frac{1}{2}g_{\mu\nu}\mathcal{R} = 8\pi G_N T_{\mu\nu} + \Lambda g_{\mu\nu} \quad (1.4)$$

where  $\mathcal{R}_{\mu\nu}$  is Ricci's tensor,  $g_{\mu\nu}$  is the metric tensor described by equation 1.1,  $\mathcal{R}$  is the Ricci's scalar,  $G_N$  is the gravitational constant,  $T_{\mu\nu}$  is the stress-energy tensor and  $\Lambda$  is the cosmological constant. Treating the matter content of the Universe as a perfect fluid, the stress energy tensor can be written as:

$$T_{\mu\nu} = -p g_{\mu\nu} + (p + \rho) u_\mu u_\nu \quad (1.5)$$

where  $p$  is the isotropic pressure,  $\rho$  is the energy density and  $u = (1, 0, 0, 0)$  is the velocity vector for the isotropic fluid in the co-moving coordinates [11]. Using the assumption of a perfect fluid as a source, Einstein's equation leads to the two Friedmann equations:

$$H^2 = \left(\frac{\dot{R}}{R}\right)^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G_N \rho}{3} - \frac{k}{R^2} + \frac{\Lambda}{3} \quad (1.6)$$

$$\frac{\ddot{R}}{R} = \frac{\Lambda}{3} - \frac{4\pi G_N}{3}(\rho + 3p) \quad (1.7)$$

The third Friedmann equation is obtained from the energy conservation principle  $T_{;\mu}^{\mu\nu} = 0$ , leading to:

$$\dot{\rho} = -3H(\rho + p) \quad (1.8)$$

Here,  $H(t)$  is the Hubble parameter. The first equation 1.6 shows the relation between the expansion rate  $H$ , the energy density  $\rho$  and the curvature  $k/R^2$ . The second equation 1.7 tells whether the expansion accelerates or decelerates. The third equation 1.8 describes how the energy density evolves during the expansion of the Universe [11].

Starting from the Friedmann equations, it is useful to define a critical density such that, assuming a null cosmological constant ( $\Lambda = 0$ ), the curvature of the Universe is zero (flat Universe):

$$\rho_c = \frac{3H^2}{8\pi G_N} = 1.05 \times 10^{-5} h^2 \text{ GeV cm}^{-3} \quad (1.9)$$

where  $h$  is the scaled Hubble parameter defined as:  $H = 100 h [\text{km s}^{-1} \text{ Mpc}^{-1}]$  or  $H^{-1} = 2998 h^{-1} [\text{Mpc}]$ . Starting from the critical density (equation 1.9) the  $\Omega_{tot}$  parameter is defined as  $\Omega_{tot} = \rho/\rho_c$ . With this notation, the first Friedmann equation can be written as:

$$\frac{k}{R^2} = H^2(\Omega_{tot} - 1) \quad (1.10)$$

showing that when  $\Omega_{tot} > 1$ ,  $k = +1$  and the Universe is closed, when  $\Omega_{tot} < 1$ ,  $k = -1$  and the Universe is open, when  $\Omega_{tot} = 1$ ,  $k = 0$  and the Universe is flat. The contributions to the total density can be divided into three terms, evaluated at the present day:  $\Omega_m$ , the density of non-relativistic particles,  $\Omega_r$ , the density of relativistic particles, and the vacuum energy density  $\Omega_v = \frac{\Lambda}{3H^2}$ . In this way equation 1.6 can be written as follows:

$$\frac{k}{R^2} = H_0^2 (\Omega_m + \Omega_r + \Omega_v - 1) \quad (1.11)$$

where the subscript 0 indicates present-day values [11]. From equation 1.11, it is clearly visible how the sum of densities in matter, relativistic particles, and vacuum affects the curvature of the Universe. Current data indicate that vacuum energy is the largest contributor to the cosmological density budget, with  $\Omega_v = 0.685 \pm 0.007$ ,  $\Omega_m = 0.315 \pm 0.007$ ,  $\Omega_r = 0$  (assuming  $k = 0$ ) [12].

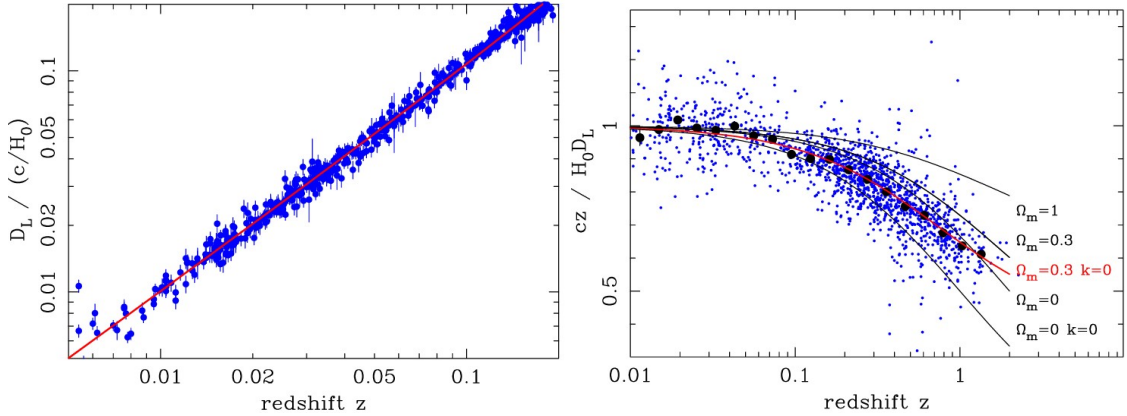


Figure 1.2: The type Ia supernova Hubble diagram, based on 1650 available supernova distances [11]. (Left) For small redshifts  $z \ll 1$ , the large-scale Hubble flow is linear and uniform. (Right) On an expanded scale, the Hubble law becomes nonlinear ( $\Omega_r$  is assumed to be 0). Larger points with errors show median values in redshift bins. Comparison with the prediction of Friedmann models favors a vacuum-dominated Universe [11].

### 1.2.2 Solution of the Friedmann equations

It is common in the literature to obtain the solution of the Friedmann equation considering simplified cases where a single component, described by an equation of state parameter  $\omega = p/\rho$ , dominates the energy density [11]. Considering a general equation of state parameter for a single component  $\omega$ , assumed to be constant, equation 1.8 can be written as  $\dot{\rho} = -3(1 + \omega)\rho\dot{R}/R$  and integrated yields:

$$\rho \propto R^{-3(1+\omega)} \quad (1.12)$$

For  $\omega \neq -1$ , inserting the result in equation 1.6, neglecting the curvature and cosmological constant term and integrating returns:

$$R(t) \propto t^{\frac{2}{3(1+\omega)}} \quad (1.13)$$

#### A radiation-dominated Universe

In the early hot and dense Universe, is coherent assume an equation of state corresponding to a gas of radiation for which  $\omega = 1/3$  [11]. In this case, equation 1.12 becomes  $\rho \propto R^{-4}$ . Indeed, the number density scales as  $R^{-3}$ , but due to cosmological redshift the particle energy is decreasing as  $E \propto R^{-1}$ . Substituting  $\omega = 1/3$  into equation 1.13:

$$R(t) \propto t^{\frac{1}{2}}; \quad H = \frac{1}{2t} \quad (1.14)$$

#### A matter-dominated Universe

At late times, non-relativistic matter dominates the energy density and can be treated as a pressureless ( $p = 0 \rightarrow \omega = 0$ ) gas. This assumption leads to  $\rho \propto R^{-3}$ , from equation 1.12 and if  $k = 0$  the result is:

$$R(t) \propto t^{\frac{2}{3}}; \quad H = \frac{2}{3t} \quad (1.15)$$

#### A Universe dominated by vacuum energy

Assuming a dominant source of vacuum energy,  $V_0$ , it acts as a cosmological constant with  $\Lambda = 8\pi G_N V_0$  and equation of state  $\omega = -1$ . Neglecting the curvature term in the Friedman equation leads to a simple solution corresponding to an exponential expansion of the Universe:

$$R(t) \propto e^{\sqrt{\frac{\Lambda}{3}}t} \quad (1.16)$$

The vacuum energy is also commonly referred to as "dark energy" [11].

### 1.2.3 The Big Bang model and a brief thermal history of the Universe

The origin of the current Big Bang theory began in the 1940s [13], when for the first time an early hot and dense Universe was proposed, including the subsequent expansion and cooling, to account for the current element observed abundances. As a direct consequence of this model, a

relic background radiation with a temperature of a few kelvin was expected. Such background radiation is currently known as the Cosmic Microwave Background (CMB) [14], and it was the first confirmation of the Big Bang model's ability to describe the early Universe.

The timeline of thermodynamic events in the Universe, within the  $\Lambda$ CMD model, presents several stages that can be briefly summarised as follows [15]:

1. Pre-inflationary cosmology ( $T > 10^{19}$  GeV): General Relativity does not hold at this scale, and quantum corrections are required.
2. Inflation: during this epoch, the expansion of the Universe accelerated exponentially. The conditions of adiabaticity and thermodynamic equilibrium are not valid.
3. End of inflation and start of the radiation-dominated Universe: radiation and other elementary particles are produced. The Universe is at the thermodynamic equilibrium while it cools adiabatically.
4. Grand Unification and phase transition ( $T \sim 10^{15}$ - $10^{17}$  GeV): all the forces were united, and different theories try to describe this epoch under the name of Grand Unified Theories (GUTs). The GUTs try to unify the electromagnetic, electroweak and strong interactions into a gauge group that should be a valid symmetry at extremely high energies [15]. At lower energies, the theory undergoes a hierarchy of spontaneous symmetry breakings (SSBs) into subgroups, producing the picture of the Standard Model. From the cosmological point of view, each SSB should correspond to a phase transition during the evolution of the Universe.
5. Electroweak phase transition ( $T \sim 100$  GeV): the  $W^\pm$  and  $Z^0$  bosons, among other particles, acquire mass via the Higgs mechanism.
6. Quantum-Chromo Dynamics (QCD) phase transition ( $T \sim 150$  MeV): quarks lose their asymptotic freedom (valid at high energies), resulting in the Quark Gluon Plasma (QGP) becoming a hadron gas. Combinations of three quarks (baryons) and two quarks (mesons) are formed.
7. Big Bang Nucleosynthesis, neutrino decoupling, and electron-positron annihilation ( $T \in [10, 0.1]$  MeV): at  $T \simeq 1$  MeV the neutron-to-proton ratio freezes out, remaining fixed. After the deuterium bottleneck (more details in Section 1.3, at  $T \simeq 0.1$  MeV the synthesis of light nuclei begins. The decoupling of neutrinos occurs at  $T \sim 1$  MeV. Previously, neutrinos were in thermal equilibrium through weak interactions like  $\bar{\nu}\nu \leftrightarrow e^+e^-$ . Regarding the electrons, they are no more relativistic at  $T \sim 0.5$  MeV.
8. Photon decoupling and recombination ( $T \sim 0.2$  MeV- $0.3$  MeV): photons and matter are no longer in thermal equilibrium, thus the interactions between photons and electrons, protons and nuclei slow down compared with the expansion rate of the Universe. During this era, electrons and protons are able to form neutral hydrogen. The decoupling of radiation from matter transforms the Universe from opaque to transparent with respect to the radiation.



The photons of the CMB provide information on the early Universe up to the recombination epoch.

9. The Dark Age: no stars have yet formed, due to the expansion of the Universe, the CMB photons are shifted towards infrared.
10. Formation of the first stars, reionisation and present time: the first stars are born. The stellar action, through high-energy photons, ionises again some of the matter present, shaping the Universe we know today.

### 1.3 Big Bang Nucleosynthesis

The Big Bang Nucleosynthesis (BBN), being based on well-understood Standard Model physics, is a reliable probe of the early Universe [16]. BBN predicts the abundances of light elements D,  $^3\text{He}$ ,  $^4\text{He}$  and  $^7\text{Li}$  produced within the first three minutes after the Big Bang. The predicted values are in good agreement with experimental data, validating the standard hot Big Bang cosmology. BBN provides constraints on possible deviations from the standard cosmology, and on new physics beyond the Standard Model [17, 18, 19].

The synthesis of light elements is sensitive to physical conditions in the radiation-dominated era at the temperature  $T \sim 1 \text{ MeV}$ , corresponding with an age of  $t \sim 1 \text{ s}$ . At higher temperatures, weak interactions were in thermal equilibrium, thus fixing the ratio of the neutron and proton number densities to  $\frac{n}{p} = e^{-\frac{\Delta m}{T}}$ , where  $\Delta m = 1.293 \text{ MeV}$  is the neutron-proton mass difference. As the temperature dropped, the neutron-proton inter-conversion rate per nucleon  $\Gamma_{n \leftrightarrow p}$  fell faster than the Hubble expansion rate. Indeed, the relation between the temperature and these two quantities is:

$$\Gamma_{n \leftrightarrow p} \sim G_F^2 T^5, \quad H \sim \sqrt{g_* G_N} T^2 \quad (1.17)$$

where  $G_F$  is the Fermi constant,  $G_N$  is the gravitational constant, and  $g_*$  counts the number of relativistic particle species determining the energy density of the radiation. The result is a departure from the chemical equilibrium, called freeze-out. The freeze-out occurred at a certain temperature:

$$T_{fr} \sim \left( g_* \frac{G_N}{G_F^4} \right)^{1/6} \simeq 1 \text{ MeV} \quad (1.18)$$

As a consequence, the neutron fraction at the freeze-out is  $\frac{n}{p} = e^{-\Delta m/T_{fr}} \simeq \frac{1}{6}$ , thus the fraction is sensitive to every known interaction. Indeed,  $\Delta m$  is determined by both strong and electromagnetic interactions, whilst  $T_{fr}$  depends on the weak and gravitational interactions. The sensitivity to the Hubble expansion rate affords a probe of the number of relativistic neutrino species. Afterwards, the neutrons were free to decay, dropping the fraction  $\frac{n}{p} \simeq \frac{1}{7}$  by the time nuclear reaction began. The rates of these reactions are affected by the density of baryons (nucleons), which is

usually expressed normalised to the photon density as:

$$\eta = \frac{n_b}{n_\gamma} \quad (1.19)$$

Considering the standard BBN within the frame of the Standard Model of particle physics, i.e. three relativistic neutrino species, the only free parameter remaining is  $\eta$ .

The nucleosynthesis chain begins with the formation of deuterium in the process  $p(n, \gamma)D$ , where a proton captures a neutron, emits a photon, resulting in deuterium. However, the production of this nucleus is slowed down by the high number density of photons, which can interact with the deuterium through photo-dissociation, destroying it. Indeed, the formation starts well after the temperature drops below the binding energy of deuterium  $\Delta_D = 2.23 \text{ MeV}$ . The deuterium production speeds up for temperatures lower than  $T \simeq 0.1 \text{ MeV}$ . This value is a direct consequence of the number of photons per baryon above the deuterium photo-dissociation threshold, i.e.  $\eta^{-1}e^{-\frac{\Delta_D}{T}}$ , which is less than 1 for  $T \simeq 0.1 \text{ MeV}$ . This process is called the deuterium bottleneck; afterwards, the production of light nuclei continues through reactions such as  $D(p, \gamma)^3\text{He}$  and  $^3\text{He}(D, p)^4\text{He}$ .

Nearly all neutrons end up bound in the most stable light element  $^4\text{He}$ . Heavier nuclei do not form because of the absence of stable nuclei with mass numbers 5 and 8, avoiding nucleosynthesis via  $n^4\text{He}$ ,  $p^4\text{He}$  or  $^4\text{He}^4\text{He}$ , and the large Coulomb barriers for reactions such as  $^3\text{He}(^4\text{He}, \gamma)^7\text{Li}$  and  $^3\text{He}(^4\text{He}, \gamma)^7\text{Be}$ . Hence, the primordial mass fraction of  $^4\text{He}$  is  $Y_p = \rho(^4\text{He})/\rho_b$  can be estimated as:

$$Y_p = \frac{4N_{\text{He}}}{N_p + N_n} = \frac{4(N_n/2)}{N_p + N_n} = \frac{2(n/p)}{1 + (n/p)} \simeq 0.25 \quad (1.20)$$

where  $N_{\text{He}}$ ,  $N_p$  and  $N_n$  are the numbers of  $^4\text{He}$ , proton and neutrons. Nuclear rates are not really important in determining the primordial abundance of  $^4\text{He}$ , though they are extremely relevant in determining the abundances of the other light elements, such as D and  $^3\text{He}$  and  $^7\text{Li}$  [11].

The BBN theory predicts the abundance of light nuclei as a function of the baryon density  $\eta$ . Those abundances are fixed at  $t \sim 180 \text{ s}$  after the Big Bang. Experimental observations are taken at current days, while the stellar nucleosynthesis is active, producing heavy elements, such as C, N, O, and Fe. In literature, these elements are called "metals". Some stars, like supernove can expel some of their material during the last phases of their life, resulting in an alteration of the primordial abundances of light elements. To measure primordial abundance, astrophysical sites with low metal abundances are used. Within these sites, the amount of light nuclei is expected to be closer to its primordial value.

It is important to underline that BBN is the only significant source of deuterium, which is destroyed during the stellar activity [20]. Observation of the primordial D value is obtained in distant and chemically unprocessed matter, with minimal stellar activity [20, 21]. Any detection provides a lower limit to primordial D/H<sub>p</sub> and an upper limit on  $\eta_{10} = \eta \times 10^{10}$ .

The primordial  $^4\text{He}$  abundance is determined through recombination emission lines of He and H in the metal-poor extragalactic HII (ionised) regions, blue compact galaxies at low redshift. The

data show that stellar production is positively correlated with metal production, so extrapolating to zero metallicity gives the primordial  ${}^4\text{He}$  abundance [22]. Here, metallicity is the mass fraction of all elements heavier than helium. Experimentally it is measured as a direct comparison with the Sun value:  $\text{Fe}/\text{H} = \log_{10} \left( \frac{N_{\text{Fe}}}{N_{\text{H}}} \right) - \log_{10} \left( \frac{N_{\text{Fe}}}{N_{\text{H}}} \right)_{\text{sun}}$ , where  $N_{\text{Fe}}$ ,  $N_{\text{H}}$  are numbers of iron and hydrogen nuclei in the star.

The best-suited objects for the determination of primordial  ${}^7\text{Li}$  are metal-poor stars in the Galactic halo, which have metallicities going down to  $10^{-6}$  of the Solar value [23]. The observed primordial  ${}^7\text{Li}$  abundance should be considered a lower bound rather than a measure. In fact, some stellar processes may lead to a  ${}^7\text{Li}$  depletion by a factor of  $\sim 1.8$  [11].

The observations of the primordial abundance of  ${}^3\text{He}$  are obtained from the Solar system and from solar-metallicity HII regions in the Galaxy [24]. Unfortunately, inferring the primordial abundance is problematic due to conflicts between stellar nucleosynthesis models (for  ${}^3\text{He}$ ) and experimental data. Consequently, compared with the other light nuclei, the primordial abundance of  ${}^3\text{He}$  has the poorest observational determination, and in the literature it is considered inappropriate as a cosmological probe [11].

To summarise, some recent experimental results on the primordial abundance of D,  ${}^4\text{He}$  and  ${}^7\text{Li}$  [11]:

$$\frac{\text{D}}{\text{H}_p} = (25.47 \pm 0.29) \times 10^{-6} \quad (1.21)$$

$$Y_p = 0.245 \pm 0.003 \quad (1.22)$$

$$\frac{\text{Li}}{\text{H}_p} = (1.6 \pm 0.3) \times 10^{-10} \quad (1.23)$$

The Figure 1.3 shows the elemental abundances as a function of  $\eta_{10}$ . The  ${}^4\text{He}$  curve reflects mainly the  $2\sigma$  uncertainty in the neutron lifetime. The curves for D,  ${}^3\text{He}$  and  ${}^7\text{Li}$  correspond to  $2\sigma$  uncertainties in nuclear cross sections. The boxes show the observed primordial abundances with their corresponding uncertainties.

The  $\eta$  ranges spanned by the boxes in Figure 1.3 do not overlap, and the discrepancy corresponds to a factor  $\sim 2$ . Specifically, the lithium data corresponds to an  $\eta$  range inconsistent with D/H and  ${}^4\text{He}$  abundances. This difference is called the lithium problem. This problem may be related to difficulties in determining the primordial lithium abundance or hint at some omission in the BBN theory. Taking into account only the D/H and  ${}^4\text{He}$ , the two abundances agree, yielding

$$\eta^{BBN} \times 10^{10} = \eta_{10}^{BBN} = 6.040 \pm 0.118 \quad (1.24)$$

$$\Omega_b^{BBN} h^2 = 0.02205 \pm 0.00043 \quad (1.25)$$

Despite the lithium problem, which remains an unresolved issue in BBN, using well-established physics and the BBN theory, it is possible to predict light element abundances spanning nine orders of magnitude, extrapolating back to  $t \sim 1$  s. This is a non-trivial success for standard cosmology and provides important constraints on the early Universe. Equation 1.25 provides

insight into the matter budget of our Universe. First of all, being  $\Omega_b^{BBN} \ll 1$ , baryons are only a small fraction of the total energy-mass density. Furthermore, the cosmic density of optically luminous matter is  $\Omega_{lum} \simeq 0.0024h^{-1}$  [26], so that most of the baryons are probably in the form of a diffuse intergalactic medium resulting in  $\Omega_b \gg \Omega_{lum}$  [27]. Finally, given  $\Omega_m \sim 0.3$ , most of the matter in the Universe is dark and in a non-baryonic form; for this reason, it is commonly called Dark Matter (DM).

The BBN prediction of  $\Omega_b$  can be tested through precision measurement of the CMB temperature fluctuations. Specifically, it is possible to estimate  $\eta$  by measuring the amplitudes of the acoustic peaks in the CMB angular spectrum [28]. Therefore, it is possible to compare two  $\eta$  measurements obtained using different physics at two separated epochs of our Universe. The standard cosmology predicts no change in  $\eta$  between BBN and CMB decoupling, thus the comparison between  $\eta^{BBN}$  and  $\eta^{CMB}$  is a key test of the theory. The results obtained with the data collected by the Planck spacecraft, without BBN theory [29, 12], yields:

$$\Omega_b^{CMB} = 0.02230 \pm 0.00021 \quad (1.26)$$

$$\eta_{10}^{CMB} = 6.104 \pm 0.058 \quad (1.27)$$

As shown in Figure 1.3, the CMB estimate of the baryon density is consistent with the BBN range. The precise measurement of the baryon density using the CMB can be used as input to BBN calculations. In this way, within the context of the Standard Model, BBN becomes a zero-parameter theory, and the light element abundances are completely determined within the uncertainties of  $\eta^{CMB}$  and the BBN theoretical errors.

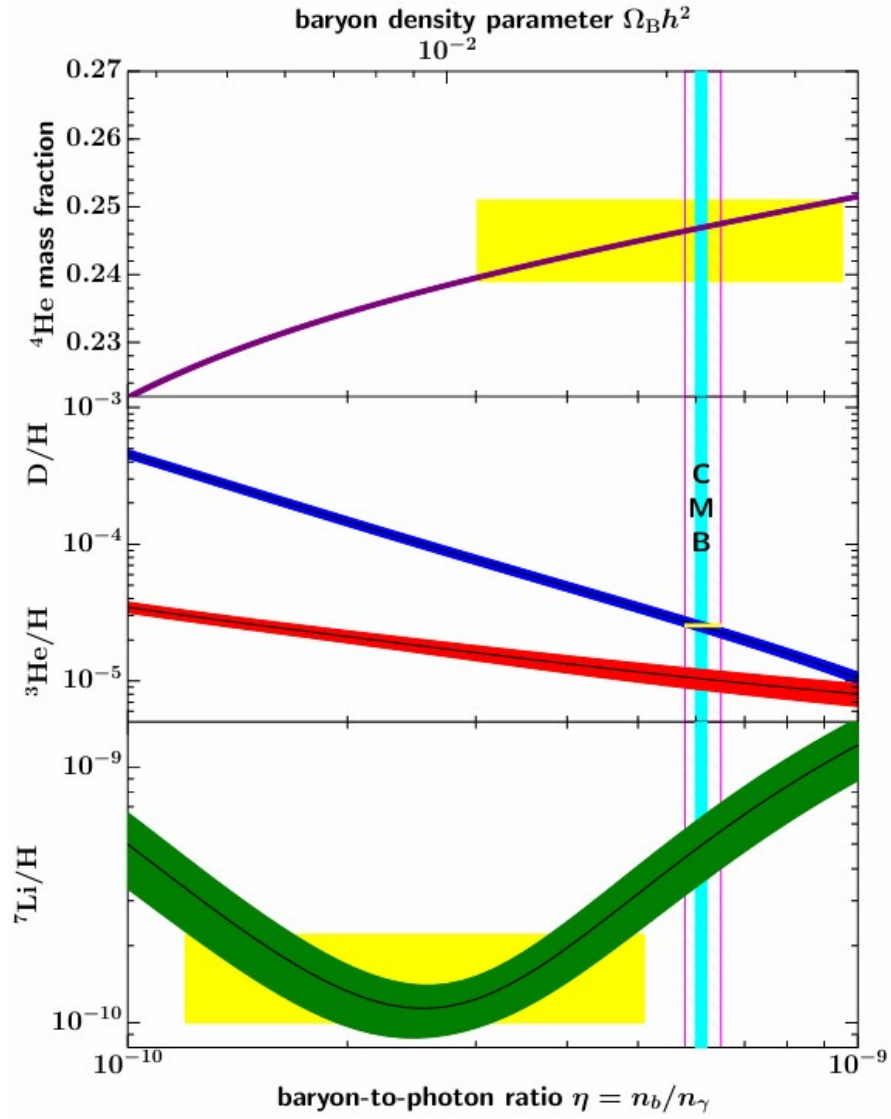


Figure 1.3: The primordial abundances of  ${}^4\text{He}$ ,  $\text{D}$ ,  ${}^3\text{He}$  and  ${}^7\text{Li}$  as predicted by the standard model of BBN. The bands show the 95% CL range [25]. Boxes indicate the observed light element abundances. The narrow vertical band indicates the CMB measure of the cosmic baryon density, while the wider band indicates the BBN  $\text{D}+{}^4\text{He}$  concordance range (both at 95% CL) [11].

The  ${}^4\text{He}$  abundance is proportional to the  $\frac{n}{p}$  ratio when the weak-interaction rate falls behind the Hubble expansion at  $T_{fr} \sim 1 \text{ MeV}$ . The presence of additional neutrino flavours at this time increases  $g_*$ , hence the expansion rate, leading to an increment of  $T_{fr}$ ,  $\frac{n}{p}$ , and therefore  $Y_p$  [30]. In the Standard Model at  $T = 1 \text{ MeV}$ ,  $g_* = 5.5 \pm \frac{7}{4}N_\nu$ , where  $N_\nu$  is the number of neutrino flavours. The helium curves in Figure 1.3 are obtained assuming  $N_\nu = 3$ . Alternatively, possible beyond-the-Standard-Model contributions might change the expansion rate during the BBN, and different models can be tested by comparing their predicted abundances with the experimental data. The BBN and the CMB constraints on  $N_\nu$  are independent and of similar precision, so it is possible to limit any change in this parameter between the nucleosynthesis and the recombination [31].

BBN provides an effective test for cosmological viability of ideas concerning physics beyond the Standard Model. Although baryogenesis and inflation must have occurred at higher temperatures and the Standard Model provides a precise description up to the Fermi scale, no observable relics of these epochs have been identified, either theoretically or experimentally. Therefore, it is not currently possible to reach boundaries at higher temperatures like  $T \sim \Lambda_{QCD}$  at the quark-hadron transition of the electroweak symmetry breaking at  $T \sim \frac{1}{\sqrt{G_F}}$ . BBN marks the boundary between the established and the speculative Big Bang cosmology [11].

## 1.4 Matter Antimatter asymmetry

The law of physics treats matter and antimatter almost symmetrically, leading to a balance in the early Universe where an equal amount of matter and antimatter should have been produced, resulting in a Universe with a vanishing baryon number. Some imbalance must have been present, since BBN assumes a Universe with a net baryon number different from zero and correctly predicts the abundance of light elements like D,  ${}^3\text{He}$  and  ${}^4\text{He}$ . The amount of asymmetry is quantified by the  $\eta$  parameter. Indeed,  $\eta = 0$  would be the result of a purely matter-antimatter-symmetric Universe, whilst the BBN and CMB independent measurements yield  $\eta \simeq 6 \times 10^{-10}$ .

### 1.4.1 Baryogenesis and the Sakharov principles

In 1967, A.D. Sakharov first proposed that the observed baryon density of the Universe need not arise from unnatural initial conditions, but could instead be explained by underlying microphysical laws [15]. Suggesting that an initially symmetric Universe could have evolved into the current observed asymmetry. The necessary conditions to generate the observed asymmetry are three [32]:

1. Baryon number (B) violation: starting from a symmetric Universe, baryon number violation must take place for the Universe to evolve into a state in which  $\eta$  is not null.
2. Violation of both C (charge conjugation symmetry) and CP (composition of parity and C): if C and CP were satisfied, baryon number-violating interactions would produce the same

excess rate of baryon and antibaryon. Therefore, the total baryon number would be kept zero, resulting in no preference for matter or antimatter.

3. Departure from equilibrium: assuming the thermal equilibrium, it is possible to compute the average value of the baryon number operator  $B$  (which is an odd operator under both  $C$  and  $CP$  transformations):

$$\begin{aligned}\langle B \rangle_T &= \text{Tr}(e^{-\beta H} B) = \text{Tr} \left[ (CPT)(CPT)^{-1} e^{-\beta H} B \right] \\ &= \text{Tr} \left[ e^{-\beta H} (CPT)^{-1} B (CPT) \right] = -\text{Tr}(e^{-\beta H} B),\end{aligned}$$

where  $H$  is the Hamiltonian (which commutes with the  $CPT$  operator) and  $\beta = \frac{1}{T}$ . From the previous equation, the final result is  $\langle B \rangle_T = 0$ ; thus, if the thermal equilibrium holds, no generation of net baryon number can occur [33].

The first two criteria are of a quantum nature, the third is gravitational and makes the connection between microscopic and macroscopic scales. The Sakharov criteria are the basic pillars to investigate the mechanism underlying baryogenesis.

The Standard Model qualitatively meet all Sakharov criteria [34]:

- The sum of baryon and lepton number ( $B + L$ ) is conserved when the temperature is low, i.e.  $T \ll 10^{12} \text{ GeV}$ . At higher temperatures  $T > 10^{12} \text{ GeV}$ , such as those present in the early Universe, the Standard Model allow the violation of  $B + L$  through fields called sphalerons [15, 35].
- Both  $C$  and  $CP$  can be violated within the Standard Model frame. The charge conjugation parity is violated through the weak interactions and  $CP$  via the Cabibbo-Kobayashi-Maskawa mechanism [36, 37].
- For a low value of the Higgs boson mass ( $m_H < 70 \text{ GeV}$ ), the expansion of the Universe could have led to a large deviation from thermal equilibrium, satisfying also the third Sakharov criterion. Curiously, the  $125 \text{ GeV}$  Higgs boson is too heavy to cause a hard (first-order) phase transition; as a consequence, the plasma of the early Universe stays very close to thermal equilibrium.

In the end, the Standard Model of Particle Physics predicts an asymmetry in the early Universe of  $\eta^{SM} \simeq 10^{-20}$ , which falls short by 10 order of magnitude with respect to the measured value obtained through BBN and CMB, i.e.  $\eta^{BBN/CMB} \sim 6 \times 10^{-10}$ .

Within the modern cosmological context, the theoretical process that occurred in the early stage of the Universe, leading to an excess of matter over antimatter, is called baryogenesis. Although the Standard Model is not able to explain the matter-antimatter asymmetry, over the years, various mechanisms have been proposed, such as electroweak [38], leptogenesis [39], supersymmetric [40], and GUT baryogenesis [41].

Based on the Standard Model and cosmological predictions and observations, the quantity of antimatter in our Universe is negligible. Indeed, antiparticles are observed mainly through precise experiments such as those at the colliders or by experiments measuring CRs. Concerning the antimatter presence in CRs, positrons and antiprotons have been precisely measured, providing interesting insight into the possible origins of these antiparticles. These particles are believed to be created by primary CR nuclei (accelerated by supernova-driven shock wave [42]) impinging onto the interstellar medium (ISM), i.e. to be of secondary origin. Some deviations from these predictions have been observed, hinting at a possible primary component [43].

In the case of  $e^+$ , a significant high-energy excess has been seen by different experiments, such as PAMELA, FermiLat and AMS-02. Two main sources are under investigation to explain the excess: DM and pulsars [44, 45, 46, 47, 48, 49, 50]. The Figure 1.4 shows the positron flux measured by AMS-02 along with a fit function based on two terms: a diffuse and a source term. Similarly, a possible excess in the AMS-02 low-energy antiproton spectrum, consistent with 20-80 GeV DM, has been discussed [51, 52]. Interestingly, this excess is consistent with some of the models indicated by the excess of the gamma-rays from the Galactic centre [53]. However, the significance of this possible DM signal depends on the interpretation of the theoretical and experimental uncertainties, given that the excess is at the level of  $\sim 10\%$  of the total flux. In particular, energy correlations and background from standard astrophysical processes are among the most important sources of uncertainty [54].

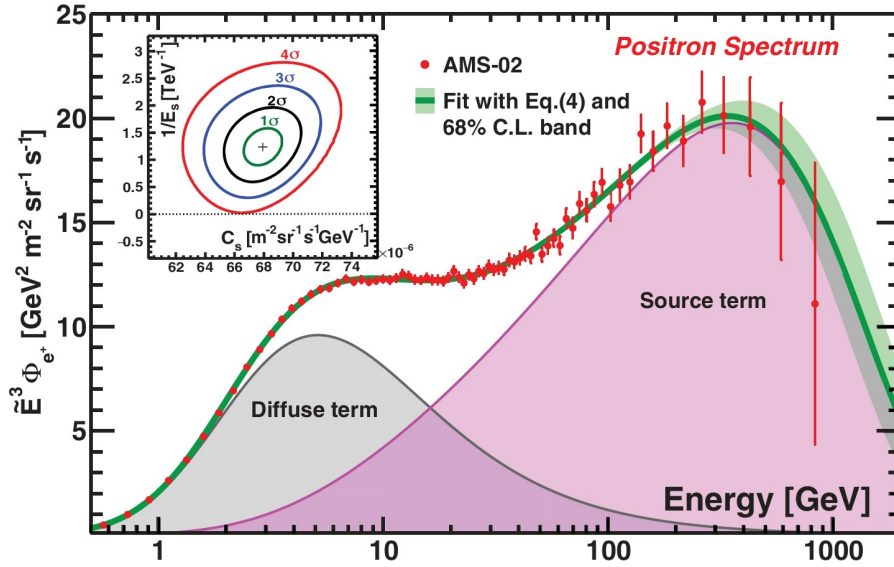


Figure 1.4: The equation  $\Phi_{e^+}(E) = \frac{E^2}{\hat{E}^2} [C_d(\hat{E}/E_1)^{\gamma_d} + C_s(\hat{E}/E_2)^{\gamma_s} \exp(-\hat{E}/E_s)]$  has been used to fit the  $e^+$  flux for  $E \in [0.5, 10^3]$  GeV (green line) together with the 68% C.L. interval (green band). The exponential cutoff of the source term is  $810^{+310}_{-180}$  GeV from the fit. The  $e^+$  flux has been scaled by  $\hat{E}^3$ . The source term is represented by the magenta area, and the diffuse term contribution by the grey area. The insert shows the projection of the regions of different significances of  $1/E_s$  measurement onto the plane of parameters  $1/E_s - C_s$  [50].



## 1.5 Antinuclei formation

Searches for antimatter in CRs do not limit themselves to  $\bar{p}$  and  $e^+$ ; hence, many theoretical and experimental efforts are devoted to detecting antinuclei. Antinuclei are possible clean probe of DM annihilations, especially at low energies (below a few GeV) [54, 43, 55]. The first detection of antihelium in space might be related to new physics, i.e. DM annihilations [56, 57, 58] or other sources of primary CR, such as antimatter stars or clouds [59, 60]. In addition to these possible sources of antihelium in space, the collision between primary CR and the ISM can produce secondary heavy antinuclei. The prediction of  $\overline{He}$  nuclei fluxes at Earth is the result of two main processes: the production due to primary CRs colliding in the ISM and the propagation of CRs in the galactic magnetic field modulated by the impact of the Sun. The latter is described by a diffusion equation:

$$\frac{\partial f}{\partial t} - K(E) \cdot \nabla^2 f + \partial_z(\text{sign}(z)fV_{conv}) = Q - 2h\delta(z)\Gamma_{ann}f \quad (1.28)$$

where  $f = dN/dE$  is the energy derivative of the number density  $N$ . The second term on the left hand side describes the diffusion modulated with a diffusion coefficient  $K(E) = K_0\beta(R/(GV))^\delta$ , where  $K_0$  is a normalisation factor [61],  $\delta$  is the spectral index of the diffusion constant,  $\beta = v/c$  and  $R$  is the particle rigidity. The third term describes convection with a velocity  $V_{conv}$ . On the right-hand side,  $Q$  is the source term and the second term accounts for losses due to annihilation rate  $\Gamma_{ann}$ , with  $h$  is the half-height of the Galaxy disc [61]. The production mechanism of antinuclei is a key ingredient in determining the source term  $Q$  required in equation 1.28.

The production of antinuclei is strictly related to the availability of their constituents, i.e. antinucleons. As a consequence, any production mechanism of heavy antinuclei must also produce antinucleons. The yield is higher for single antinucleons, whilst bigger antinuclei are more difficult to assemble. Approximately, every additional antinucleon reduces the production yield by a factor  $10^3$ . For this reason, a correlation exists between the measurement of cosmic-ray antiprotons and heavy antinuclei. Nowadays, two main approaches exist: the production of antinuclei via freeze-out from a quark-gluon plasma (called the statistical thermal model) and the production through coalescence of individual antinucleons [62]. Both models describe data recorded by different experiments. The fixed target experiment NA49 measured  $Pb$ - $Pb$  interactions at 20-158 GeV/c per nucleon and presented results compatible with both the statistical and the coalescence model [63]. At the Large Hadron Collider (LHC), the ALICE experiment performed several measurements using  $p$ - $p$  ( $\sqrt{s} = 0.9, 2.76, 7, 13$  [TeV]),  $p$ - $Pb$  ( $\sqrt{s}$  per nucleon = 2.76, 5.02 [TeV/n]) and  $Pb$ - $Pb$  ( $\sqrt{s}$  per nucleon = 5.02 [TeV/n]) collisions. The thermal model reproduces the  $Pb$ - $Pb$  results, whilst  $p$ - $Pb$  can be explained with both the models [64, 65, 66, 67, 68, 69, 70]. The thermal model overestimates the deuteron-to-proton ratio for  $p$ - $p$  collisions. Currently, the formation of heavy antinuclei remains incompletely understood [71, 72], allowing for a wide span of predicted fluxes all compatible with the observed  $\bar{p}$  flux [54].

### 1.5.1 Coalescence

The coalescence model describes the fusion of an antiproton and an antineutron into an antideuteron with the simple assumption that any  $\bar{p}$  and  $\bar{n}$  pair within a sphere of radius  $p_0$  in momentum space will coalesce and produce an antinucleus [73, 74] as reported by equation 1.29:

$$|\vec{k}_{\bar{p}} - \vec{k}_{\bar{n}}| < 2p_0 \quad (1.29)$$

where  $\vec{k}_{\bar{p}}$  and  $\vec{k}_{\bar{n}}$  are momentum vectors in the centre of mass reference frame. Following this assumption, as an example, the  $\bar{d}$  spectrum is given by:

$$\gamma_{\bar{d}} \frac{d^3 N_{\bar{d}}}{dp_{\bar{d}}^3} = \frac{4\pi}{3} p_0^3 \left( \gamma_{\bar{p}} \frac{d^3 N_{\bar{p}}}{dp_{\bar{p}}^3} \right) \left( \gamma_{\bar{n}} \frac{d^3 N_{\bar{n}}}{dp_{\bar{n}}^3} \right) \quad (1.30)$$

where  $p_i$  and  $\frac{dN_i}{dp_i}$  are the momentum and the differential yield per event of the  $i$ -th particle ( $\bar{d}$  for antideuterons,  $\bar{p}$  for antiprotons and  $\bar{n}$  for antineutrons). The  $p_0$  parameter is called the coalescence momentum and is a free parameter that can be extrapolated from the data. For more complex antinuclei, a slightly different parametrisation is used, adding the  $B_A$  parameter:

$$B_A = \frac{A}{m_p^{A-1}} \left( \frac{4\pi}{3} p_0^3 \right)^{A-1} \quad (1.31)$$

with  $m_p$  the proton mass. A single experiment can determine the value of the coalescence momentum ( $p_0$ ) by measuring the  $\bar{p}$  and  $\bar{d}$  spectra. The coalescence momentum describes the difference in momentum of the coalescing nucleons, but also several different effects.

The coalescence model is not a nuclear-physics model; it is a statistical approach able to reproduce the observed data. Therefore, effects such as energy conservation and spin alignment are not directly included in the model but are accounted for as uncertainties in the coalescence momentum. Consequently, it is challenging to distinguish the coalescence process from other effects and fully comprehend the antinuclei formation mechanism. The coalescence momentum must be considered as a function of the kinetic energy [71, 72]. The increase is especially steep in the region most relevant for cosmic rays, up to 100 GV. For  $\overline{He}$  the value of the coalescence momentum is around 350 MeV/c [75, 76].

A technical difficulty arises when Monte Carlo hadronic generators are involved. These generators are unable to produce antinuclei. The state-of-the-art technique is to apply a coalescence condition on a per-event basis. For example, to produce an  $\bar{d}$ , the momentum difference of each possible  $\bar{p}\bar{n}$  couple is computed in the centre-of-mass frame. If the difference is smaller than  $p_0$ , an  $\bar{d}$  is formed with a momentum equal to the sum of the constituents, following the coalescence condition reported in equation 1.29. To account for the binding energy, the total energy of the antinucleon is calculated from the total momentum of the nucleus and the rest mass value present in the PDG. The implementation of a production mechanism for  $\overline{^3He}$  and  $\overline{^4He}$  is slightly different. Indeed, two different approaches are currently considered:

- Simultaneous coalescence:  $N$  antinucleons simultaneously coalesce to produce an  $N$ -particle antinucleus. In this scenario, each  $\bar{p}\bar{n}$  pair must fulfil the condition of equation 1.29. Thus, to produce  $\overline{^4\text{He}}$ , two  $\bar{p}$  and two  $\bar{n}$  are selected, and for the 6 possible pairs the condition in equation 1.29 is evaluated [72].
- Iterated coalescence: single antinucleons are iteratively added to a multi-nucleon state if they fulfil the two-particle coalescence condition. For example, an  $\bar{d}$  is further evaluated for the production of  $\overline{^3\text{He}}$ , by pairing it with the remaining  $\bar{p}$  in the event [72].

A pictorial representation of the coalescence

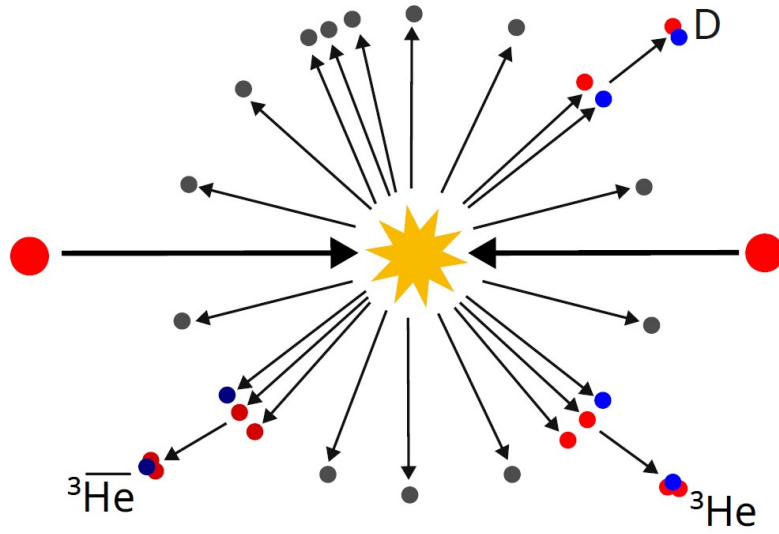


Figure 1.5: Pictorial illustration of the coalescence mechanism. The interaction between two particles (like the collision of a cosmic-ray proton with the ISM) can produce secondary particles such as protons, neutrons, and their antiparticles. If these secondaries are sufficiently close in both space and momentum, they can coalesce into nuclei or antinuclei [77].

### 1.5.2 Thermal model

In the thermal model, the antinuclei yields, as well as all the hadrons, are determined by the thermodynamical equilibrium without a specific reference to a production mechanism [62, 54]. The emission source is called fireball, a system of hot and dense matter created at the early collision stage. The fireball is assumed to be in local thermal and chemical equilibrium. The abundances of the light nuclei are fixed at the chemical freeze-out with a temperature of  $T_{chem} = (156 \pm 4)$  MeV [69, 78]:

$$\frac{dN}{dy} \propto \exp\left(-\frac{m}{T_{chem}}\right) \quad (1.32)$$

with  $\frac{dN}{dy}$  the particle rapidity density for a particle species with mass  $m$ . Surprisingly, light nuclei seem to behave as all other hadrons. This model reproduces the light-nuclei production rate in  $p$ - $p$  and  $p$ - $Pb$  collisions. In this case, instead of the grand-canonical implementation used for  $A$ - $A$  collisions, a canonical ensemble has to be used, where the baryon number, charge, and strangeness

must be conserved [68, 65, 66].

## 1.6 Recent searches for cosmic antinuclei

The expected  $\bar{p}$  and antinuclei fluxes due to the interaction of CRs with the ISM, taking into account the uncertainties associated with the propagation and the coalescence momentum, are shown in Figure 1.6. It is relevant to note how each additional antinucleon reduces the expected flux at Earth by different orders of magnitude.

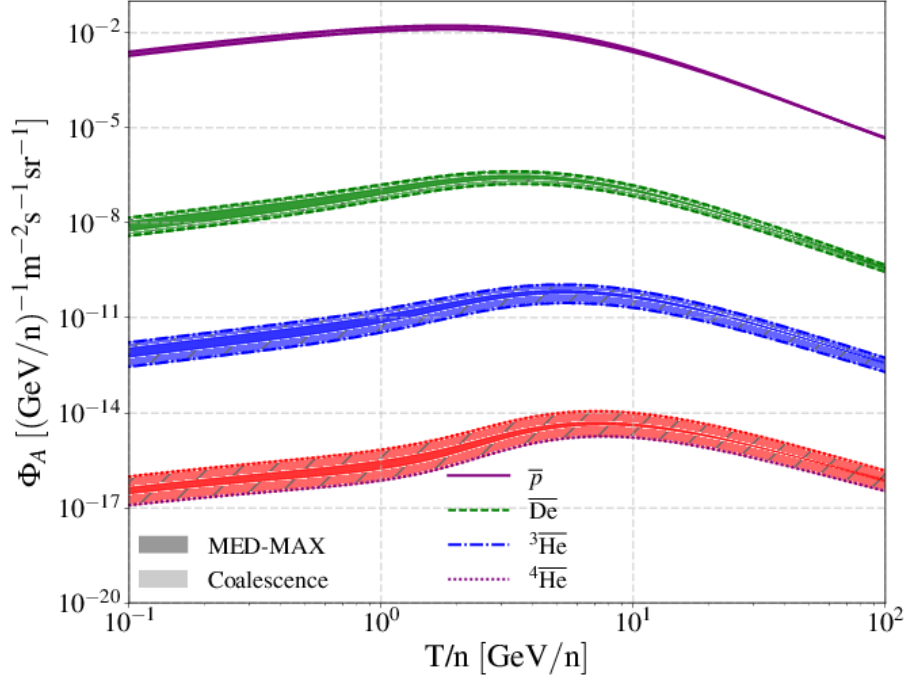


Figure 1.6: Predicted secondary flux of  $\bar{p}$ ,  $\bar{d}$ ,  ${}^3\bar{\text{He}}$  and  ${}^4\bar{\text{He}}$  showing the uncertainty associated with the propagation and the coalescence momentum [43].

In addition to the source term related to the interaction of CRs with the ISM, there is also the possibility that the secondary antihelium nuclei interact inelastically on ISM without annihilating and losing a significant fraction of their energy. In literature, these antinuclei are called tertiary [61]. The prediction of fluxes due to a possible DM signal, the secondary, and tertiary sources for  ${}^3\bar{\text{He}}$  are presented in Figure 1.7. Also, the expected sensitivity of AMS-02 for antihelium searches is displayed for a 5 and 13 years of data taking. In a conservative approach, interpreting the potential excess in antiprotons measured by AMS-02 as an upper limit on DM annihilation implies that the fluxes shown in Figure 1.7 should be considered as estimates of the highest possible fluxes without violating  $\bar{p}$  data [61]. The secondary flux is suppressed below 8 GeV/n and the tertiary component does not contribute to the background for DM searches. Clearly, the interpretation of a potential antihelium candidate depends on the energy range of the observation. If candidates were observed below 1 GeV/n, it would be a hint for DM, whilst candidates at energies above a few GeV/n would point towards a secondary flux component. Indeed, under an optimistic sce-

nario characterised by a large coalescence momentum, the secondary  ${}^3\overline{\text{He}}$  flux lies only a factor 2 below the expected AMS-02 sensitivity after 13 years, at a kinetic energy per nucleon of 30 GeV/n. On the contrary, the DM signal considered in [61] is always below the AMS-02 sensitivity.

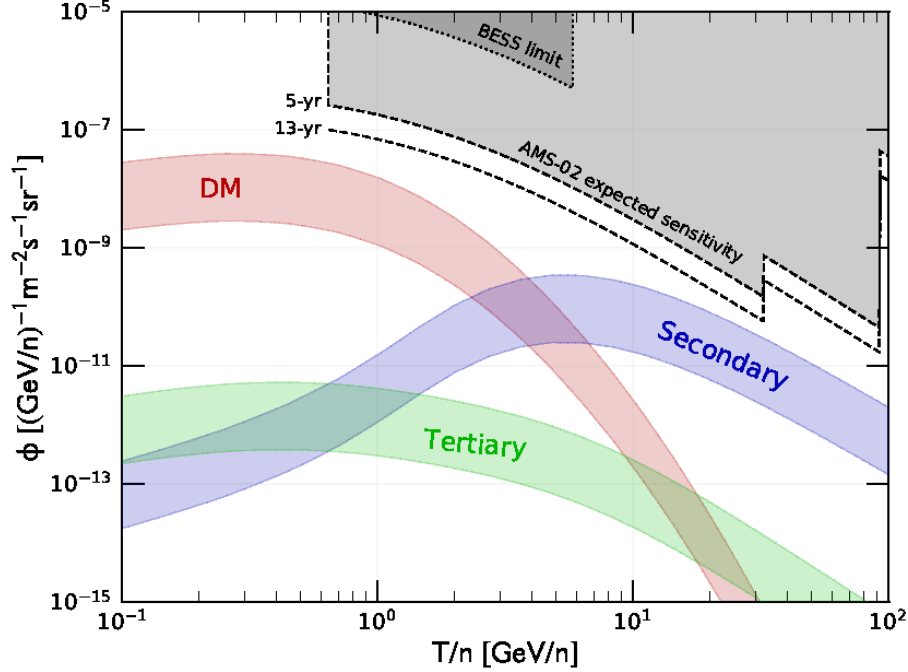


Figure 1.7: Standard astrophysical (secondary and tertiary) flux of antihelium in comparison to a potential DM signal. The bans show the uncertainty on the coalescence process,  $p_c$  spanning from 160 MeV to 248 MeV. The BESS limit (95% CL) [79] and AMS-02 sensitivity (95% CL) scaled from to 5 and 13 years on the antihelium-to-helium flux ratio are transformed to an antihelium flux sensitivity by the measured AMS-02 helium flux. All lines correspond to a force-field solar modulation potential of  $\phi = 600$  MeV [61].

Currently, there is no experimental evidence of antinuclei in space. BBN predicts zero antinuclei production during nucleosynthesis, and some studies showed that an overall  $B = 0$  Universe cannot exist; thus, no large (of size compatible with the observable Universe) antimatter structures are present. If these antimatter islands existed, some distortions in the CMB spectra and in the cosmic diffuse gamma-ray background should be present, due to the annihilations occurring at the boundaries of those regions [80]. Despite this, it is not possible to exclude the existence of small and localised gatherings of antimatter [81, 43].

In recent years, the AMS-02 collaboration has reported the possible discovery of several ( $\sim 10$ ) antihelium events with a reconstructed mass compatible with  ${}^3\text{He}$  and  ${}^4\text{He}$ , with  $Z = -2$  and  $|R| < 50$  GV. It is worth noting that the antihelium detection has not yet been confirmed. Although these results are not published, they have encouraged significant theoretical work on the implications for DM models and predicted  $\bar{p}$  and  $\bar{d}$  fluxes. Data taking, analyses, and interpretation on these possible candidates are still ongoing, and this work is only a part of the work currently done by the AMS-02 collaboration. Even though these candidates are tentative, the detection of antihelium among CRs may question the current understanding of BBN or reveal

something unforeseen in the realm of astrophysical objects, such as anticlouds or antistars [81, 43].

There are three main possible origins for the AMS-02 antihelium candidates: secondary production, DM annihilations, and the existence of anti-clouds or anti-stars in our Galaxy. It is challenging to explain the AMS-02 candidates as a consequence of a pure secondary component, especially for  ${}^4\overline{\text{He}}$  nuclei. The expected sensitivity of the experiment is always above the expected fluxes as seen in the Figures 1.7 and 1.8 showing the results of two independent works [43, 61].

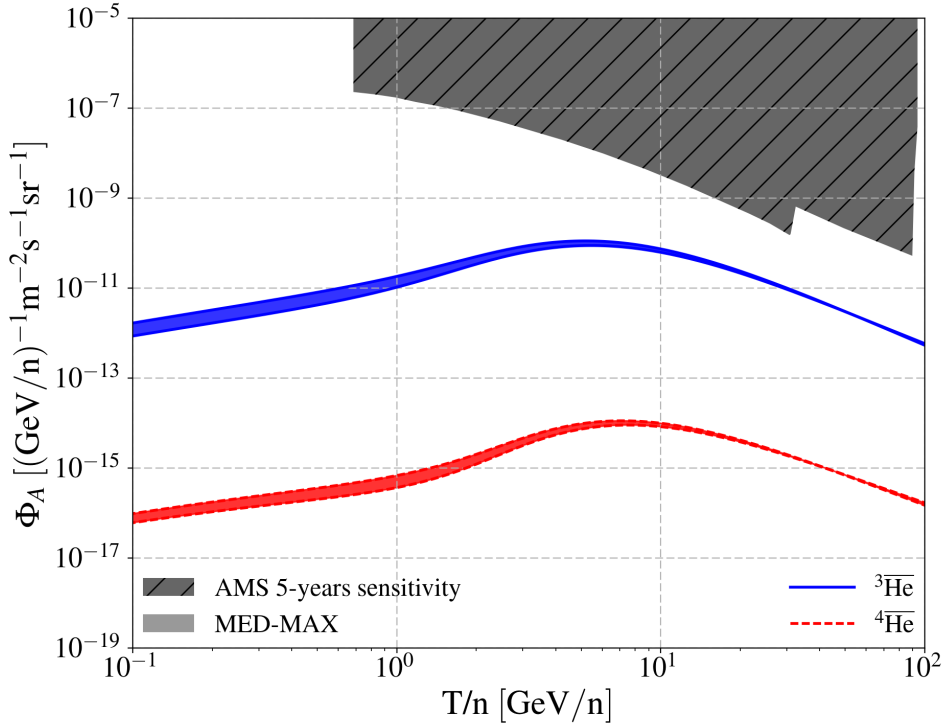


Figure 1.8: Predicted secondary flux of  ${}^3\overline{\text{He}}$  and  ${}^4\overline{\text{He}}$  using the upper limit on the coalescence momentum deduced from the ALICE experiment together with the expected sensitivity from AMS-02 (5 years). The bands show the uncertainty due to the propagation model [43]. The fluxes are the same as those presented in Figure 1.6.

The  ${}^3\overline{\text{He}}$  calculation yields a flux that is a factor  $\sim 30$  below what is measured by AMS-02 [43]. Uncertainties in the propagation are unlikely to be responsible for the mismatch. To boost the secondary flux, the coalescence factor  $B_3$  (see equation 1.31) must increase by the same factor. The ALICE experiment has measured the coalescence factor from pp collisions with centre of mass energies from 0.9 to 7 TeV [64, 65, 66]. The collisions between CRs and the ISM occur at smaller energies (hundreds of GeV), however, there are arguments against a large increase of the coalescence factor at low energy [43]:

- within the energy range covered by ALICE, the coalescence factor is close to constant
- previous measurements of  $B_2$  and  $B_3$  factors from heavy ion collisions with beam energies between 0.4 and 2.1 GeV/n, albeit probed at a much lower centre-of-mass energy than the

case of ALICE, have found the coalescence momentum to lie between 0.130-0.187 GeV for helium-3. This result is smaller than what was found by ALICE [82].

- theoretically, the rate of coalescence of nucleons should be higher at high energies rather than at low energies. Indeed, high-energy particles colliding will create jets of particles with an angular opening smaller than that of low-energy collisions. This will increase the correlation of nucleons within the shower, also increasing the probability of antinuclei production. Additionally, the probability of many nucleons in a low-energy collision is suppressed by the phase space.

As an alternative approach, the column density of interstellar matter crossed by CRs (grammage) could be higher than expected. In this scenario, all secondary particles should be affected, producing excesses in other CRs species. In conclusion, seems unlikely that the AMS-02 antihelium candidates can be explained as a pure secondary component.

Antihelium nuclei arising from the annihilation of DM particles suffer similar limitations to the secondary production, as the estimate of the DM source term depends on the coalescence momentum. Indeed, similarly to secondary production, a hierarchical relation between the fluxes of  $\bar{p}$ ,  $\bar{d}$ ,  ${}^3\bar{\text{He}}$  and  ${}^4\bar{\text{He}}$  is expected. Thus, if DM is responsible for the AMS-02 events, it seems unlikely to observe  ${}^4\bar{\text{He}}$  without seeing also  $\bar{d}$  and an excess in the antiproton flux [61, 43].

The discovery of a single  ${}^4\bar{\text{He}}$  nucleus is challenging to explain in terms of known physics. The possibility that extended local regions made of antimatter have survived, from the early Universe in our Galaxy, may be able to explain the detection of  ${}^4\bar{\text{He}}$  in space. The antimatter regions can be present as anti-cloud or antistars [43]. Any hypothetical anti-cloud survives in a matter dominated Universe only if it contains a negligible amount of matter. It is possible to estimate the matter content considering two cases: the antimatter is in a cold  $T \sim \mathcal{O}(30)$  K or a hot ionised cloud  $T \sim \mathcal{O}(10^6)$  K. The corresponding estimates are independent of the size and density of the antimatter regions and yield [43]:

$$n_p^{\text{cold}} < 3.5 \times 10^{-8} \text{ cm}^{-3} \quad (1.33)$$

$$n_p^{\text{hot}} < 6.1 \times 10^{-5} \text{ cm}^{-3} \quad (1.34)$$

Similar estimates can be provided for the early Universe, resulting in even stronger requirements. These would require an exotic mechanism allowing the separation of matter and antimatter domains all along the cosmological evolution, making the existence of such anti-clouds quite improbable. Additionally, gamma rays provide strong constraints on this scenario due to the absence of spectral features close to the proton, positron and pion masses [43]. Anticlouds could produce a measurable flux of  $\bar{p}$  and  $\bar{d}$ . Most of the parameter space evades current  $\bar{p}$  constraints, but experiments like GAPS could be sensitive to this scenario [55]. Within the third Fermi Large Area Telescope source Catalogue [83], some unassociated point sources can be anti-clouds experiencing annihilations due to CRs propagating through them. Alternatively, they can also be antistars which experience annihilations as they propagate within the ISM. Depending on the

acceleration mechanism, which is currently unknown, it is conceivable that a nearby ( $\sim 1$  pc) anti-star might be responsible for the AMS-02 observation [43].



## Chapter 2

# The Alpha Magnetic Spectrometer, AMS-02



Figure 2.1: AMS is mounted on the ISS with a 12 degree angle to the zenith to prevent the rotating ISS solar arrays from being within the AMS field of view [84].

The Alpha Magnetic Spectrometer (AMS) is a state-of-the-art particle physics detector operating on board the International Space Station (ISS). AMS was launched as a payload on the Space Shuttle Endeavour on the 16<sup>th</sup> of May 2011. The experiment has been acquiring data since the 19<sup>th</sup> of May 2011 and will continue to operate up to the end of the ISS activities. The main goal of AMS-02 is to measure the flux of the different components of the galactic cosmic rays in a rigidity range from a fraction of GV up to a few TV. These measurements allow to search for:

- primordial complex antimatter such as  $\bar{d}$  and  $\overline{He}$ .
- indirect detection of Dark Matter through the precise measurement of the anti-particle fluxes

as  $e^+$  and  $\bar{p}$ .

- measure with unprecedented precision the particles and nuclei present in the CRs spectra, providing a precious insight into the production, acceleration and propagation of CRs in the galaxy.
- measure the effects of short and long-term solar activity on the cosmic ray fluxes.

The data used in this work have been collected with the AMS-02 experiment during 12.5 years of operation. In the following Chapter, I introduce the main features of the AMS-02 detector, including the data acquisition and transmission systems.

## 2.1 The AMS-02 detector on the ISS

The detector is installed at the Upper Payload Attach Point on the ISS S2 Truss with a relative inclination of  $12^\circ$  to the zenith of the station. The ISS orbits at an altitude ranging from 370 to 460 km, with an orbital inclination of  $51.6^\circ$ . The velocity of the station is more or less constant around 7.7 km/s, resulting in a orbit period of approximately 93 min [85]. AMS-02 is a large area ( $\sim \text{m}^2$ ), large field of view ( $\sim 45^\circ$ ) multi TV magnetic spectrometer with a total weight of 7.5 t and a total volume of  $5 \times 4 \times 3 \text{ m}^3$ . The power budget of the payload is only 2 kW despite the large number of readout channels ( $\sim 5 \times 10^5$ ). AMS-02 has several subsystems, allowing for the redundant measurement of physical quantities such as momentum  $p$ , charge  $Z$ , and velocity  $\beta = \frac{v}{c}$  of the particles traversing the detector. The most relevant components of AMS-02 are (see Figure 2.2):

- Permanent Magnet: with a magnetic field of  $B = 0.14 \text{ T}$  and a bending power of  $BL^2 = 0.15 \text{ T} \cdot \text{m}^2$ , it bends charged particles in a different direction depending on their charge, allowing for determining the charge sign.
- Silicon Tracker Detector (TRK): consisting of nine silicon layers, seven inside the magnet bore and two outside of it. These layers measure the position and the deposited energy of the particle traversing the detector, allowing for the reconstruction of its curved trajectory and the measurement of rigidity ( $\frac{p}{Z}$ ), charge sign and absolute value.
- Time of Flight Detector (TOF): made up of four planes of plastic scintillators positioned in orthogonal directions, two situated above and two below the magnet. It is used to detect when a particle traverses the detector, and consequently, it triggers the data acquisition system. It determines the direction of the particle traversing the detector, and it measures both the particle charge and velocity.
- Anti-Coincidence Counter (ACC): a system of curved plastic scintillators inside the magnet bore, it surrounds seven layers of the Silicon Tracker, providing a veto for sideways-going particles.

- Ring-Imaging Cherenkov Detector (RICH): based on the Cherenkov radiation emission [86], it measures particle velocity and absolute charge. The subdetector is located below the lower TOF.
- Transition Radiation Detector (TRD): located on the AMS upper part, separates light and heavy particles. It is crucial to discriminate  $e^+$  from  $p$  and  $e^-$  from  $\bar{p}$ .
- Electromagnetic Calorimeter (ECAL): placed at the bottom of AMS, it measures the remaining energy of electrons and photons.

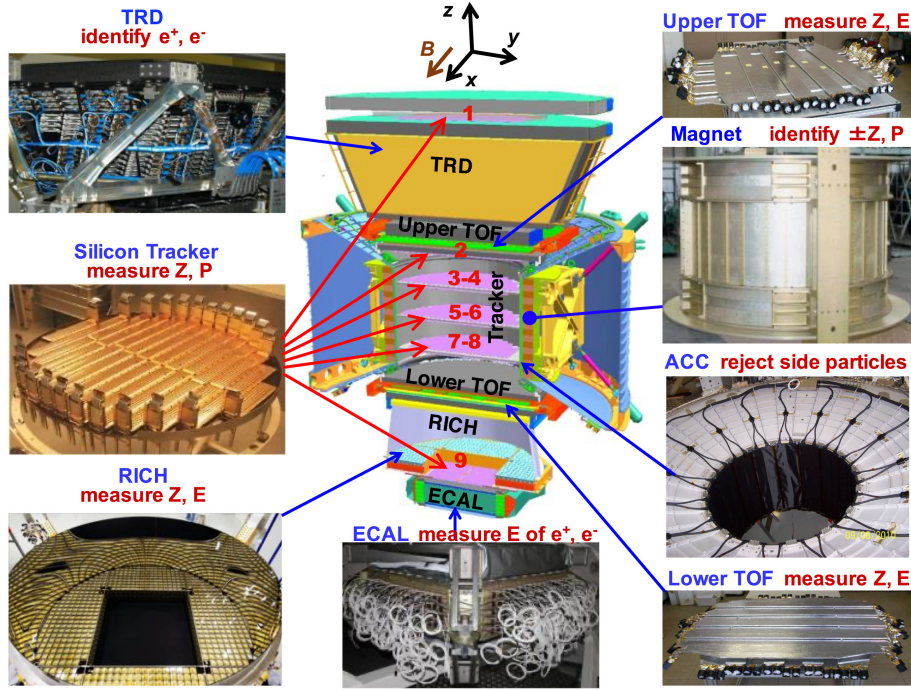


Figure 2.2: The AMS detector and its subdetectors. The spectrometer measures the particles and nuclei charge ( $Z$ ), energy ( $E$ ) and momentum ( $P$ ) or rigidity ( $R = p/Z$ ), which are measured independently by the Tracker, TOF, RICH and ECAL. The ACC counters are used as a veto against particles entering AMS from the side. The AMS coordinate system, concentric with the magnet, is also shown. The  $X$  axis is parallel to the main component of the magnetic field, and the  $z$  axis is pointing vertically [84].

### 2.1.1 The Permanent Magnet

The magnet bends the trajectories of charged particles. The curvature of the trajectory is measured through the sagitta  $s$ . Under the assumption of  $s$  being much smaller than the curvature radius  $R$ , the following relation holds:

$$s \approx \frac{L^2}{8R} = \frac{0.3BL^2c}{8R_T e}, \quad (2.1)$$

where  $L$  is the length of the region over which the field bends the particle trajectory,  $B$  is the magnetic field expressed in T,  $R_T$  is the transverse rigidity (with respect to the magnetic field) in GV,  $c$  is the speed of light, and  $e$  is the electron charge. The product  $BL^2$ , being proportional to  $s$ , is used to quantify the spectrometer resolution and is called the bending or analysing power of the

magnet. The equation 2.1 expresses the relation between the sagitta and the rigidity (momentum) of the curved particle. The magnet is an essential component of a spectrometer, since it is necessary to measure the momentum and the charge sign of the incoming particles. Indeed, negative charges are bent in an opposite direction to positive charges.

The permanent magnet of AMS-02 is the same as that used in the test flight of AMS-01, it is composed of 64 high-quality Nd-Fe-B sectors, and each one of these sectors has 100 blocks measuring  $5.08 \times 5.08 \times 2.54 \text{ cm}^3$  [87] [88, 89]. The left image of figure 2.3 shows the arrangement of the field directions of the 64 sectors, producing a dipole field of 0.14 T at the centre of the magnet, a bending power of  $BL^2 = 0.15 \text{ Tm}^2$ , and a negligible dipole moment outside of it [88]. Indeed, the field outside the magnet drops, becoming less than 3 G at a distance of 2 m. This aspect is crucial to avoid a torque effect on the ISS due to the interaction with Earth's magnetic field of 0.5 G.

The permanent magnet is allocated in the centre of AMS-02, assembled in a cylindrical shell structure of 0.8 m height, an inner diameter of 1.1 m and an outer diameter of 1.29 m. The geometric acceptance is  $0.82 \text{ m}^2\text{sr}$  and the weight of the magnet without the supporting structure is around 1900 kg [88, 87, 89]. The magnet is shown on the right image of Figure 2.3.

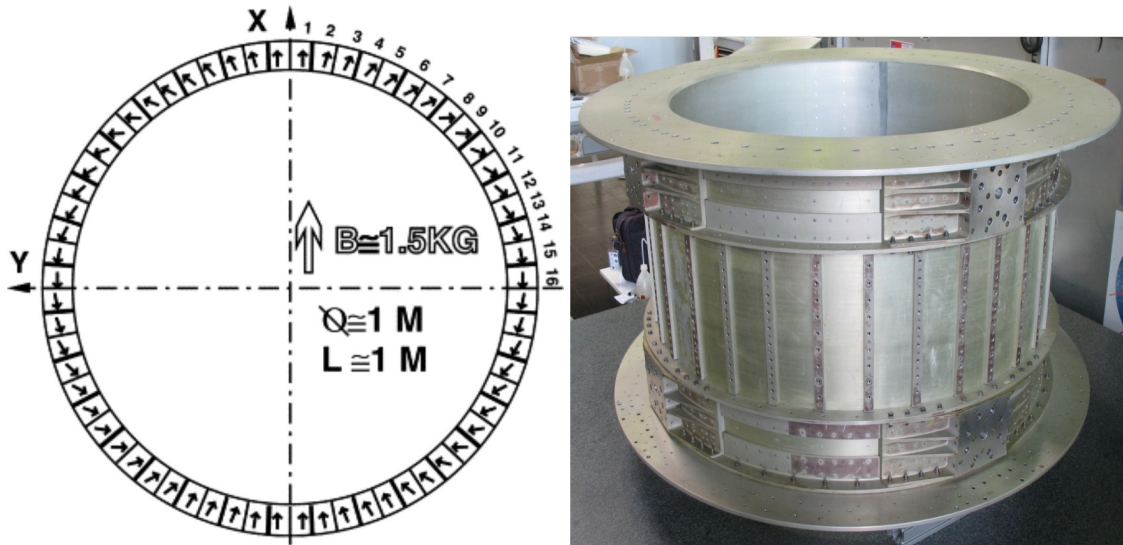


Figure 2.3: (Left) The orientation of the magnetic field of the AMS-02 permanent magnet sectors [88]. (Right) The permanent magnet used by AMS-01 and AMS-02. Three full-size magnets were made, one for space qualification, one for destructive test, and one for flight [88].

The local reference system of AMS is established inside the magnet, where the Y direction aligns with the bending plane, whilst the generated magnetic field is along the X direction. To ensure no sensible degradation has occurred since the AMS-01 mission in June 1998, the magnetic field has been accurately mapped again in 2010 [87]. The right plot of Figure 2.4 shows the comparison between the two measurements as a function of the z-coordinate. The left plot of Figure 2.4 shows the variation along the  $\phi$  angle and the radius in the midplane of the magnet. Both the measurements quantify a degradation lower than 1%, corresponding to the accuracy of the original measurements performed before the AMS-01 mission [87].



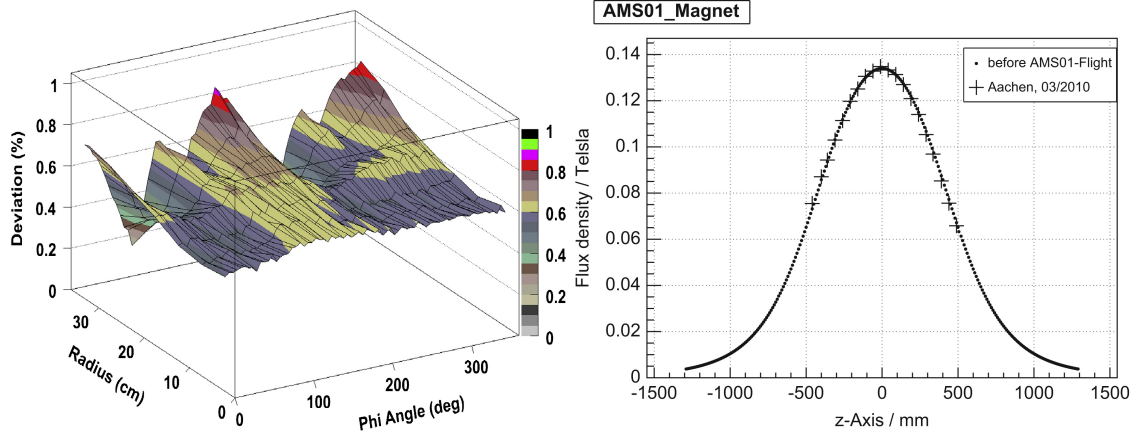


Figure 2.4: (Left) The plot shows the variation of the magnetic field of the AMS-01 permanent magnet in the midplane over 12 years. The accuracy of the original measurement was 1% [87]. (Right) The plot shows the measured magnetic field along the z-axis before the AMS-01 flight (dotted line) and the measurements performed in Aachen before the AMS-02 launch (crosses) [87].

### 2.1.2 Silicon Tracker Detector

The Silicon Tracker Detector (TRK) is a key subdetector of AMS-02. The TRK has two main goals: to provide a precise measurement of the particle position and to measure the energy loss rate inside a thin material. The precise measurement of the position is fundamental to reconstruct the particle trajectory and to measure the sagitta, and as reported in equation 2.1, the particle rigidity. The rigidity, denoted as  $R$ , is measured by the tracker through the subsequent measurements of the position on each plane. Additionally, the tracker measures the energy loss  $\Delta E$  on each plane, which is proportional to the square of the incoming particle's charge. Indeed, every TRK plane provides an independent measurement of the charge. The redundancy of this measurement allows for the reduction of the fluctuations of a single measurement, and it is instrumental in distinguishing between the different species of charged cosmic rays. Combining the curvature, the orientation of the track (upward or downward provided by the TOF), the charge absolute value, and the magnetic field map, it is possible to measure the momentum ( $p = Z \cdot R$ ) of the particle and determine the sign of its charge.

The silicon tracker is made of 2,284 double-sided microstrip silicon sensors of size  $72 \times 41 \times 0.3 \text{ mm}^3$ . These sensors are grouped into 192 ladders, resulting in 200,000 readout channels. The ladders are arranged over 9 Silicon layers, named L1 to L9, which build the Silicon Tracker Detector (TRK) of AMS-02. These layers are separated in 6 planes along the Z coordinate, 3 made by three single layers and 3 composed of 2 silicon layers, one on each side of the plane. The first plane is the L1 on top of AMS-02, the second layer is just above the magnet bore; three planes are inside the magnet bore and then the last plane is just above the electromagnetic calorimeter and is called L9. The TRK layers are visible as magenta planes in figure 2.2. The layers L2 to L8 are named the inner tracker, since these layers are placed inside the magnet bore. To guarantee stability, the tracker planes are held by a honeycomb carbon fibre structure.

The readout electronics produce 150 W of heat, removed by a mechanically pumped  $\text{CO}_2$  two-

phase cooling system. Initially, the system was the Tracker Thermal Control System (TTCS). Afterwards, between November 2019 and January 2020, the system was replaced during a series of Extra Vehicular Activities (EVAs) with the Upgraded Tracker Thermal Pump System (UTTPS). UTTPS will allow the TRK to run during the entire ISS lifetime.

Each tracker layer has a spatial resolution better than  $10\text{ }\mu\text{m}$  in the bending direction ( $Y$ ) and  $30\text{ }\mu\text{m}$  in the non-bending direction ( $X$ ) [90]. The distance between L1 and L9 is 3 m, providing a lever arm to measure rigidity up to 2 TV for  $Z = 1$  particles and up to 3 TV for  $Z > 1$ .

### Track reconstruction

After the calibration of the TRK, which happens near the equator, it is possible to select strips that have a significant signal-to-noise ratio. A group of neighbouring strips, with a non-negligible signal, is called a cluster. A cluster is a unidimensional object with a defined position and energy deposition. Clusters collected on the two sides of the same ladder are combined into 3-dimensional hits. The track-finding algorithm uses these hits to reconstruct the particle's trajectory, identifying several track segments within the inner tracker, which are then combined into a single track utilising a track quality estimator similar to a  $\chi^2$  test [91]. After finding the track, the particle rigidity is obtained by fitting the trajectory in the magnetic field with the cluster coordinates associated with the track. Positively charged cosmic rays, travelling downgoing ( $\beta_{TOF} > 0$ ), will bend towards positive  $y$  coordinates due to the presence of the permanent magnet, whilst negatively charged cosmic rays will be bent in the opposite direction. The sign of the charge is determined by measuring the curvature direction. Figure 2.5 shows the different stages of the track reconstruction.

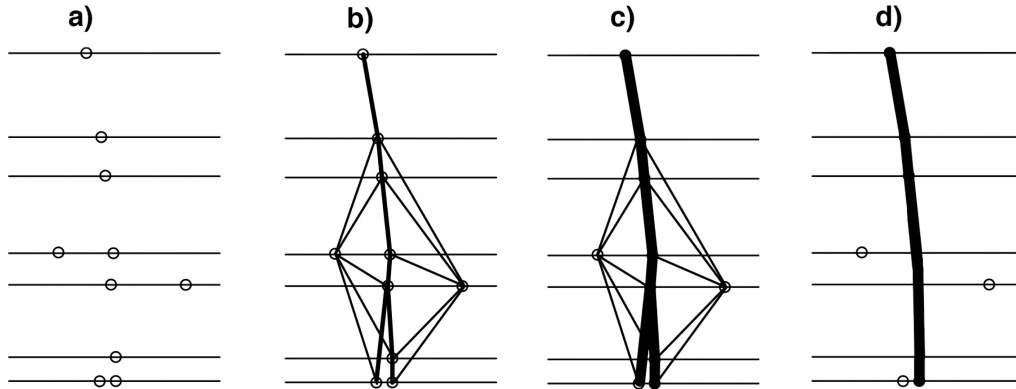


Figure 2.5: Schematics of the track finding algorithm. (a) Event with original hits. (b) Event with all the track segments. (c) Correct segments chosen based on the minimisation of angles between segments. (d) Choose one track as the result of using a  $\chi^2$ -like track estimator [91].

The rigidity measurement takes into account corrections due to multiple scattering of charged particles and energy losses inside the detector. The reliability of the track-finding algorithm is fundamental to correctly reconstruct events with delta-ray emission and secondary tracks due to interactions inside the detector material [91].

The track-finding algorithm identifies four different tracker patterns with different maximum lever arms:

- **Inner:** only hits between L2 and L8 are used in the fit, the Field Of View (FOV) with respect to AMS zenith direction is  $40^\circ$  and the lever arm is 80 cm.
- **L1+Inner:** only hits between L1 and L8 are used in the fit,  $30^\circ$  of FOV and a lever arm of 180 cm.
- **Inner+L9:** only hits between L2 and L9 are used in the fit,  $35^\circ$  of FOV, and a lever arm of 200 cm.
- **L1+Inner+L9 (Full Span):** all available hits on each TRK layer are used in the fit,  $25^\circ$  of FOV, maximum lever arm of 300 cm.

Figure 2.6 shows a pictorial representation of the four different track patterns.

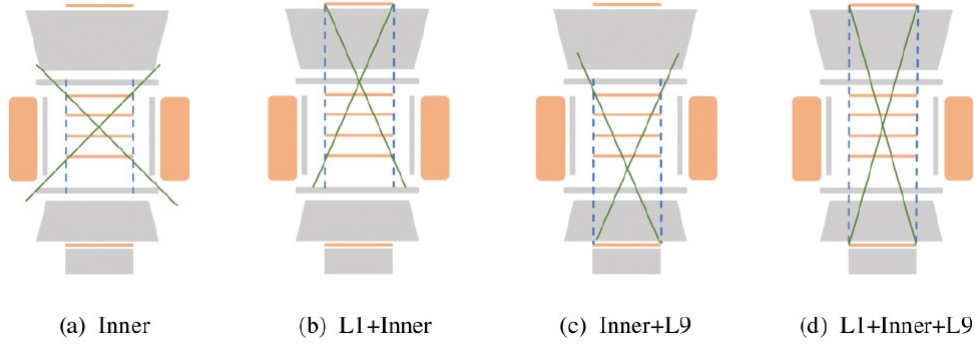


Figure 2.6: The four track patterns of the AMS-02 Silicon Tracker. (a) Inner, (b) L1+Inner, (c) Inner+L9, (d) L1+Inner+L9 or Full Span [92].

The rigidity resolution, i.e.  $\Delta R/R$ , is estimated from Monte Carlo (MC) simulations, since the initial momentum of the particle is known. Comparing the measured value with the expected one, the resolution is obtained by fitting the residual distributions. The inverse of rigidity is compared directly with the MC information, because it is the quantity directly measured by the tracker and also because this variable has a Gaussian-like shape. The rigidity resolution is extracted starting from the  $\frac{1}{R_{REC}} - \frac{1}{R_{GEN}}$  distribution obtained using logarithmic binning. The bulk of the distribution has a Gaussian-like shape, whilst the tails deviate from this behaviour. The tails can be described using a crystal-ball function or an exponential modified Gaussian. The width of the distribution corresponds to  $\sigma_{1/R}$ , and using error propagation can be converted to an uncertainty on rigidity:

$$\sigma_R = \left| \frac{d(1/R)}{dR} \right| \sigma_{1/R} = R^2 \sigma_{1/R} \quad (2.2)$$

and the rigidity resolution can be expressed as  $\frac{\sigma_R}{R} = R \sigma_{1/R}$ .

Figure 2.7 shows the rigidity resolution for helium obtained from the Monte Carlo simulation [93]; the proportionality between the resolution and the rigidity is clearly visible. At low rigidity ( $R < 10$  GV), the resolution is affected by multiple scattering, while after a flattening at higher rigidities, it increases again as theoretically expected. The Maximum-Detectable Rigidity (MDR) is defined as the rigidity whose relative error is equal to 100%. Different tracker patterns correspond to

different MDRs, due to the lever arms. The MDR value for helium with the inner pattern is 400 GV and is particularly interesting for this work, since all the analysis is carried out considering rigidities reconstructed with this track pattern, which corresponds to the maximum acceptance of  $\sim 0.5 \text{ m}^2 \text{ sr}$ . This is really important when searching for rare events such as  $\overline{He}$  candidates.

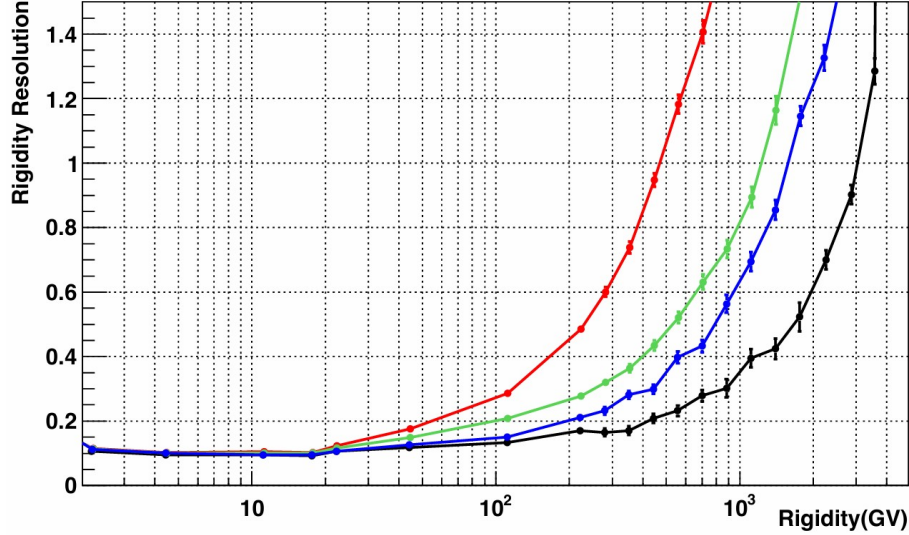


Figure 2.7: Tracker rigidity resolutions for Helium estimated from the Monte Carlo simulation. The four curves represent the tracker different spans: inner (red), inner-L1 (green), Inner-L9 (blue) and full span (black) [93].

The TRK digitalisation model has been tuned using the data collected from the test beam and ISS data. This tuning ensures that the simulation reproduces the TRK point resolution for protons and helium. The upper plot of Figure 2.8 shows the inverse rigidity measured by the inner TRK using both Monte Carlo simulation and data from a test beam with 400 GeV/c protons. The lower plot of the same figure shows that the good agreement also holds for a different momentum range. In this case, ISS data are used, and the comparison is between rigidity measured by the TRK between L1-L8 and L2-L9.

### Tracker rigidity scale

One of the most relevant systematic errors in the measurement of the cosmic ray fluxes performed by AMS-02 is the uncertainty on the absolute rigidity scale. This systematic error is related to the alignment procedure of the TRK.

On the ground, AMS-02 has been tested with particles beam at the CERN Super Proton Synchrotron accelerator. These tests allowed to measure the alignment parameter of the TRK, as shown in the upper plot in Figure 2.8. During the transportation, the installation of AMS-02 on the ISS, or due to the out-gassing or thermal expansions, these parameters may have changed. To account for these effects, the alignment of AMS-02 is carried out in orbit using cosmic rays. In particular, L1 and L9 undergo a dynamic alignment every 2 minutes, using extrapolations from the inner TRK reaching a precision better than  $5 \mu\text{m}$  for both layers [94, 95]. After determining the alignment parameters, a further check of the rigidity scale is carried out. Specifically, the inverse



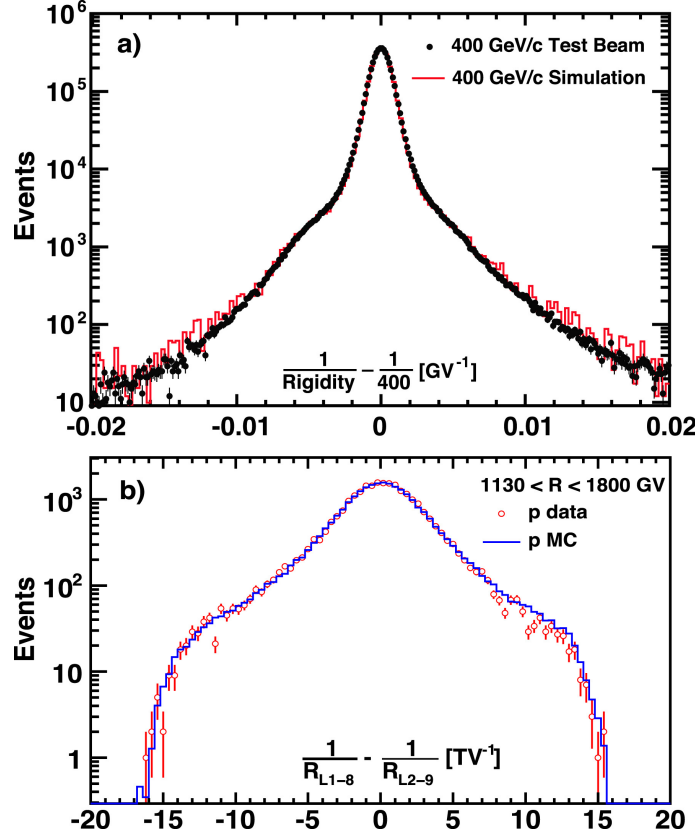


Figure 2.8: (a) Comparison between data and the Monte Carlo simulation of the inverse rigidity measured by the tracker for 400 GeV test beam protons. (b) The difference of the inverse rigidities measured with the upper (Layers L1–L8) and the lower (Layers L2–L9) parts of the tracker for the cosmic ray proton data collected on the ISS and for the Monte Carlo simulation in the rigidity range [1130 – 1800] GV. [84].

absolute rigidity ( $1/|R|$ ), measured by the TRK, is compared with the inverse energy ( $1/E$ ), measured by the ECAL. The comparison uses positrons and electrons from 2 to 300 GeV, and allows to correct possible residual misalignments as shown in Figure 2.9.

The absolute rigidity scale may carry a time-dependency. To determine the time-dependent rigidity scale shift correction, the  $e^\pm$  data from the first 7 years of operation are divided into four time intervals. For each interval, the rigidity absolute scale correction is computed and the results are shown in Figure 2.10. The accuracy of the rigidity scale bias is within  $0.034 \text{ TV}^{-1}$ , resulting in a displacement of the layers from L2 to L8 lower than  $0.2 \mu\text{m}$  [96].

### Tracker charge measurement

Charged particles traversing the silicon layers of the TRK ionise the medium, producing electron-hole pairs that drift in the silicon, inducing a signal on strips around the particle impact position. The strip amplitudes have to be corrected for several effects, such as:

- the dependence on the particle impact position
- the dependence on the particle inclination
- non-linearities in the readout electronics

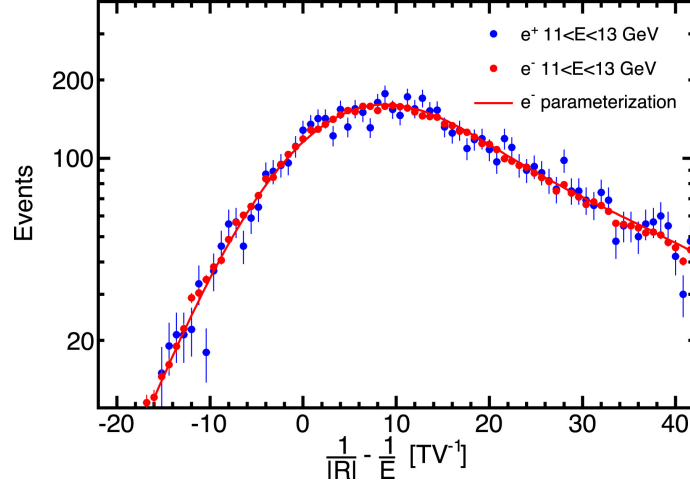


Figure 2.9: The left plot shows the  $1/|R| - 1/E$  distribution for the energy bin 11 to 13 GeV for electron (red points) and positron (blue points). The continuous line is the parametrisation for electrons defined as the sum of a Gaussian and an exponential modified Gaussian [84].

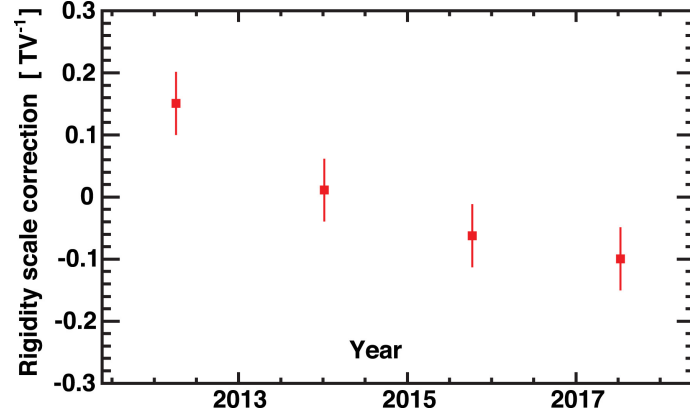


Figure 2.10: On the right, the time dependence of the rigidity scale correction from May 2011 to May 2018 [84].

- thermal environmental changes occurring in time
- the dependence of the energy loss on the velocity according to the Bethe-Block formula.

The Bethe-Block formula states that the average energy loss is given by:

$$-\langle \frac{dE}{dx} \rangle = K \frac{z^2}{\beta^2} \langle \frac{Z}{A} \rangle \left[ \frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 T_{max}}{I^2} - \beta^2 - \delta(\beta\gamma) \right] \quad (2.3)$$

where  $K = 0.307075 \text{ MeVg}^{-1}\text{cm}^2$  is a constant,  $z$ ,  $\beta$ ,  $\gamma$ , and  $T_{max}$  are the particle charge, velocity, Lorentz factor and maximum energy transferable in a collision with an electron. While  $\langle \frac{Z}{A} \rangle$ ,  $I$  and  $\delta$  are the average charge-to-mass ratio, mean excitation potential and density effect correction of the traversed medium. After all the corrections, the nuclei charge measurement is built by combining the signal of all individual strips from both sides  $y$  and  $x$ . Figure 2.11 shows the charge distribution obtained using only the information provided by the inner tracker. It is clearly visible how the cosmic ray species are well separated up to  $Z = 28$  (Nickel).

The TRK charge resolution as a function of the measured charges is reported in Figure 2.12; for

helium the charge resolution is  $\Delta Z / Z = 0.035$  [97].

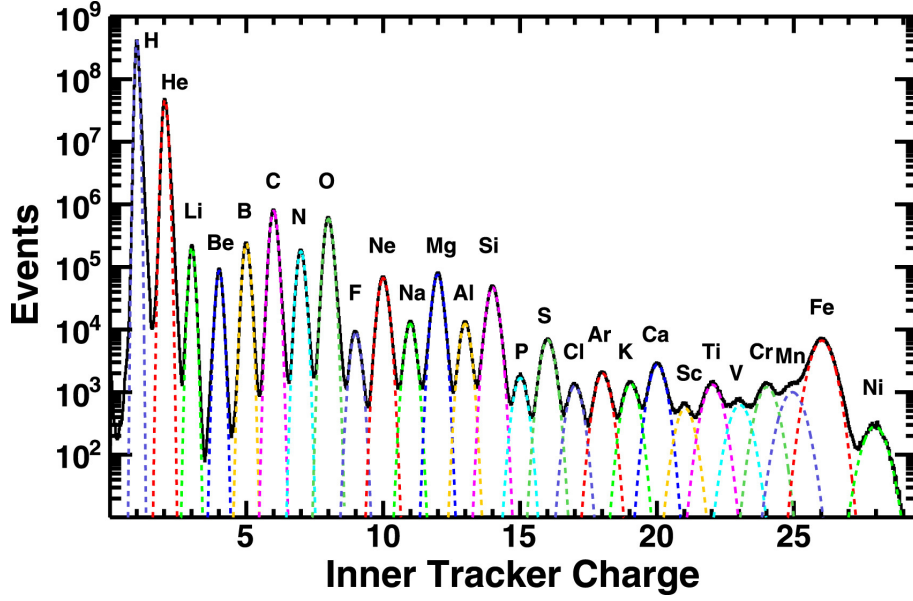


Figure 2.11: The inner tracker charge distribution for the measured cosmic ray nuclei with  $R > 7$  GV. Dashed lines are Gaussian fits for each nuclear charge. The charge identification is done solely by the inner tracker [97] [84].

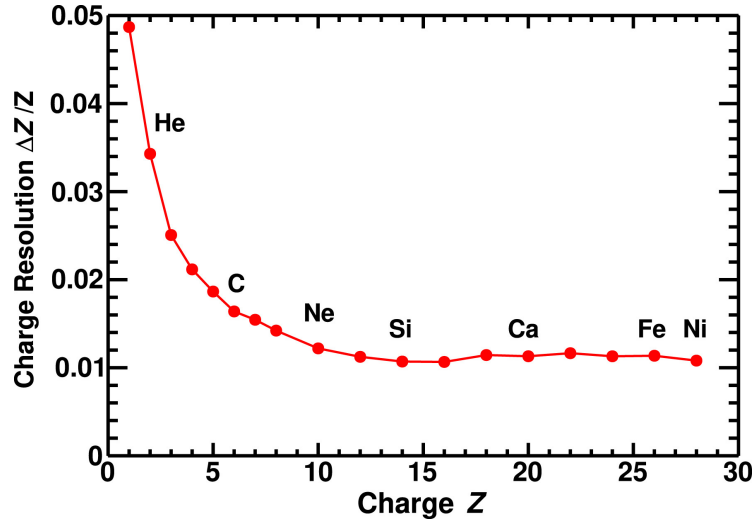


Figure 2.12: The inner tracker (layers L2-L8 combined) charge resolution  $\Delta Z / Z$  for different charges [84]. A detailed explanation of the charge measurement is available here [97].

### 2.1.3 Time-of-Flight System

The Time-of-Flight (TOF) system of AMS measures accurately the velocity and the charge of cosmic rays traversing the detector. Moreover, it determines the direction of the particle (down or upgoing), and it triggers the data acquisition with a high efficiency, approaching 100% [98].

The TOF measures the velocity  $\beta = \frac{v}{c}$  of charged particles by measuring the time of flight between the upper and lower TOF, and it uses a straight line or the fitted track model for measuring the distance travelled, i.e.  $\beta_{TOF} = \frac{\Delta s}{c \cdot \Delta t}$ . The velocity resolution improves with higher charges, starting from 4% for  $Z = 1$ , to 1.2% for  $Z \geq 6$ , corresponding to a time resolution of 48 ps. The time resolution distinguishes between downgoing and upgoing particles at the  $10^{-9}$  level. The left part of Figure 2.13 shows the velocity resolution of the TOF system as a function of the particle charge. The right plot of Figure 2.13 shows the velocity resolution of TOF for *He* nuclei with  $R > 20$  GV, obtained from the Monte Carlo simulation.

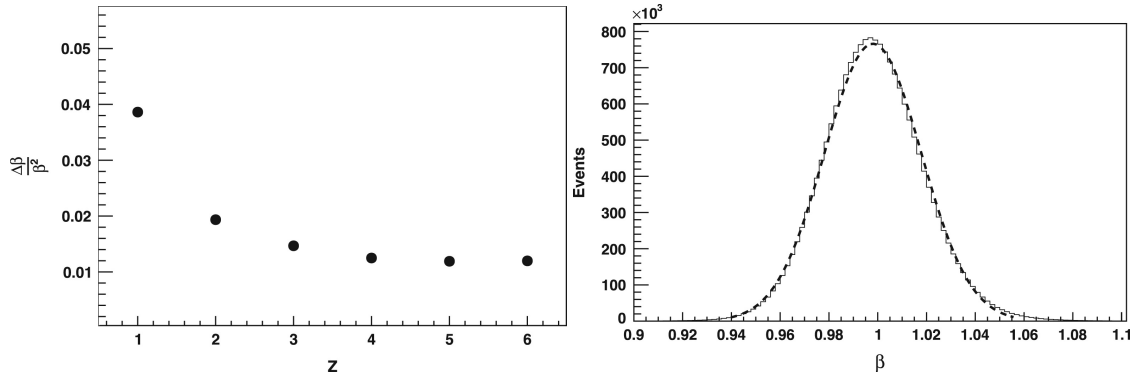


Figure 2.13: (Left) The TOF velocity resolution as a function of the particle charge  $Z$ . (Right) The measured TOF velocity distribution for *He* nuclei with  $R > 20$  GV; the dashed line is a Gaussian fit with a standard deviation of 2% [98].

The TOF detector measures the charge of cosmic rays by detecting the scintillation light produced by ionisation due to the passage of a charged particle. The emitted light is not linearly proportional to the energy deposition; indeed, the number of photons saturates with increasing  $dE/dx$ . The relation between the deposited energy  $\epsilon = \frac{dE}{dx}$  and the light yield of the scintillator  $\frac{dL}{dx}$ , i.e. number of photons per unit of length, is described by the modified Birks law [99, 100]:

$$\frac{dL}{dx} = \epsilon \cdot S \cdot \frac{1}{(1 + k_B \epsilon \cdot \epsilon^2)} \quad (2.4)$$

The signals from PMTs on the anodes and dynodes are utilised in different ranges: anodes for  $1 \leq Z \leq 3$ , anodes and dynodes for  $4 \leq Z \leq 8$  and dynodes alone for  $8 \leq Z \leq 30$  due to their different dynamic range.

The TOF is used as a trigger for charged particles, requesting a coincidence from all four planes. The signals from anodes are compared to three thresholds, used in the level-1 trigger:

- **Low Threshold (LT):** corresponds to 20% of the Minimum-Ionising Particle (MIP) signal.
- **High Threshold (HT):** corresponds to 60% of the MIP signal and is used for  $Z \geq 1$  trigger.

- **Super-High Threshold (SHT):** corresponds to 400% of the MIP signal and is used for ions trigger  $Z \geq 2$ .

The Time Of Flight (TOF) detector is composed of the upper TOF, above the magnet bore, and the lower TOF, below the magnet bore, as shown in Figure 2.2. Both have two layers of EJ-200 plastic scintillators, and the vertical distance between the upper and lower TOF is around 120 cm. Each layer comprises 8 or 10 scintillating paddles, 1 cm thick and varying lengths ranging from 117 to 134 cm. The paddles are alternately oriented in the  $X$  and  $Y$  directions, with different shapes. Inner paddles are rectangular and 12 cm wide, outer paddles are trapezoidal and have a width ranging from 18 to 26 cm. Figure 2.14 shows the top view of the upper (left) and lower (right) TOF. To guarantee a near 100% trigger efficiency, nearby counters are overlapped by 0.5 cm [98]. The scintillation light is read on both ends of each paddle by 2 or 3 Photo-Multiplier Tubes (PMTs), connected through plexiglass light guides to the paddles. These two time measurements provide information on the position and the crossing time of the particle without knowing the length of the paddle. The shape of the light guides minimises the effects of the magnetic field on the PMT signal. In total, there are 144 PMTs in the upper and lower TOF, collecting light from 34 paddles.

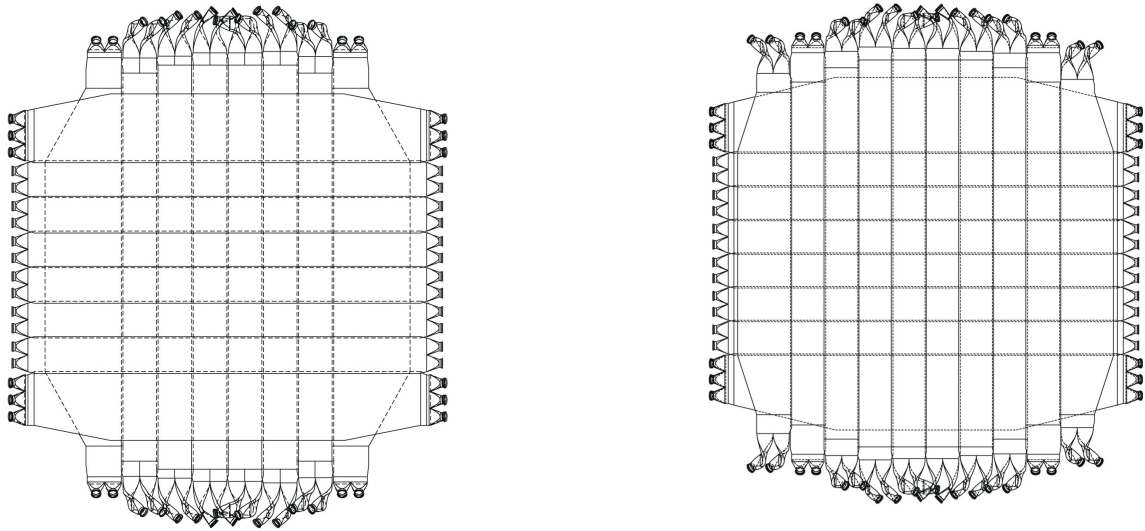


Figure 2.14: The left image shows the upper TOF plane, the right image shows the lower TOF plane [98].

The Figure 2.15 displays the two-dimensional distribution obtained from the charge measurements performed with the silicon tracker and the TOF system. Their combined response provides a clear separation of the cosmic-ray species up to charge  $Z = 28$ .

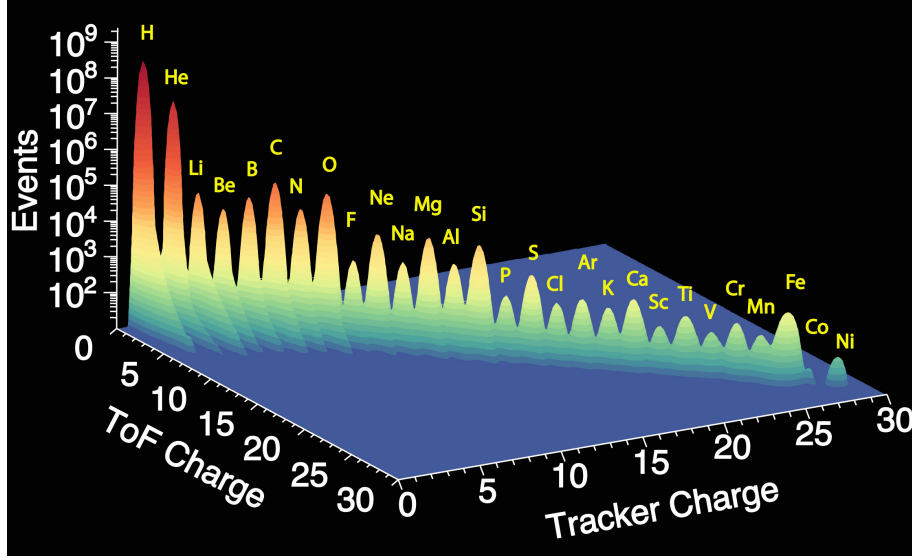


Figure 2.15: The charge measurement by the TOF and the Tracker [84].

#### 2.1.4 Anti Coincidence Counter

The anticoincidence (ACC) detects particles entering or leaving the inner tracker volume transversely and works as a veto. Indeed, particles entering AMS from the sides or secondary interactions within the detector can distort measurements of charge and rigidity. Moreover, during periods of intense fluxes, e.g. in the South Atlantic Anomaly, the trigger rate must be reduced. The ACC is placed within the magnet bore and surrounds the 7 layers of the inner tracker system, as shown on the left of Figure 2.16. The system must be hermetic, efficient and reliable.

The ACC cylinder is made out of 16 scintillation panels (Bicron BC-414) with a thickness of 0.8 mm. A charged particle, traversing the scintillators, loses energy due to ionisation processes and the scintillating material emits light in the UV range,  $\lambda \sim 400$  nm. This light is absorbed by wavelength-shifting fibres (WLS, Kuraray Y-11(200)M) embedded into the panels. The light with the new wavelength of 480 nm is carried with fibre cables to photomultiplier tubes (PMT, Hamamatsu R5946) with a corresponding quantum efficiency around 20% [101]. The scintillators are assembled to form a cylinder with a height of 0.83 m and a diameter of 1.1 m. To maintain the hermiticity, the counters have a tongue and a groove along the vertical edges such that particles crossing this area are detected simultaneously by two panels. The panels are divided into couples; each couple is read out by the same two PMTs, one on top and one on the bottom of the magnet bore. This setup guarantees redundancy and a reduction of the overall weight. After many years of operation in space, the ACC systems have an efficiency close to 99.999% [101].

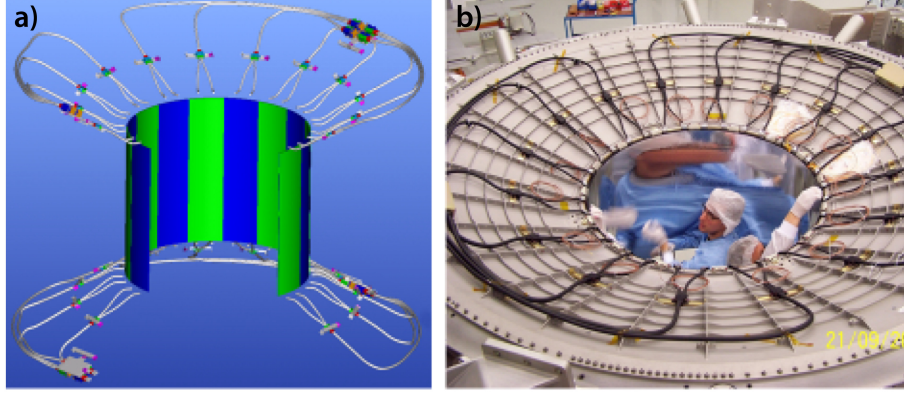


Figure 2.16: (Left) Cutaway view of the ACC counters (blue and green colored panels) and their fibre light guide collection system, and (Right) installation of the ACC in the magnet bore. The fibre light guide collection system (black cables) is also visible [84] [101].

### 2.1.5 Rich Imaging Cherenkov Detector

The Rich Imaging Cherenkov Detector (RICH) provides a precise measurement of the velocity and electric charge of the incoming particle. The velocity and the charge are measured from the emission of Cherenkov photons and the reconstruction of the corresponding ring. The Cherenkov radiation is emitted by charged particles, travelling in a medium with a refractive index  $n$ , with a velocity  $\beta$  higher than the light velocity in that medium [86]. The measurement of the emission angle ( $\theta_C$ ) is related to the particle velocity by the expression:  $\beta = \frac{1}{n \cdot \cos \theta_C}$ . The charge measurement relies on the electric signal associated with the photon ring, thus the number of photo-electrons ( $N_{pe}$ ) according to  $Z^2 \propto \frac{N_{pe}}{\epsilon} \cdot \frac{1}{\sin^2 \theta_C}$ . Where  $\epsilon$  is an overall efficiency, taking into account the ring acceptance, the photon detection efficiency, and possible scattering or absorption effects. This formulation is a direct consequence of the Frank-Tamm formula [102].

The RICH is placed between the lower TOF and the electromagnetic calorimeter (ECAL), and is composed of a radiator plane, an expansion volume surrounded by a conical reflector and a detection plane [103]. The left image of Figure 2.17 shows a cut view of the RICH. The central part of the radiator plane is made out of 16 sodium fluoride (NaF) tiles, measuring  $8.5 \times 8.5 \times 0.5 \text{ cm}^3$  and with a refractive index of  $n = 1.334$ . The aerogel (AGL) surrounds the NaF, and the rest of the plane is covered with 92 AGL tiles, each measuring  $11.5 \times 11.5 \times 2.5 \text{ cm}^3$  and having a refractive index of  $n = 1.05$ . The two different refractive indices result in two velocity thresholds,  $\beta_{th}^{NaF} = 0.75$  and  $\beta_{th}^{AGL} = 0.953$ . The relative dimensions of the radiators are around 1 : 9, so around 90% of the RICH events cross the AGL region [103][104]. The expansion volume has a height of 47 cm and is covered by a conical reflector. The latter redirects photons within the detection plane, thereby improving overall detection efficiency.

Lastly, the detection plane is covered with a matrix of 680 pixelised Hamamatsu PMTs (R7600-M16) coupled to solid plastic light guides. Each PMT has 16 pixels, each one with a size of  $8.5 \times 8.5 \text{ mm}^2$ . The right image of Figure 2.17 shows a single PMT and the corresponding light guides. Charge particles crossing the light guides produce a second signal, in addition to the Cherenkov emission, due to ionisation. This kind of signal is concentrated in a few pixels and is



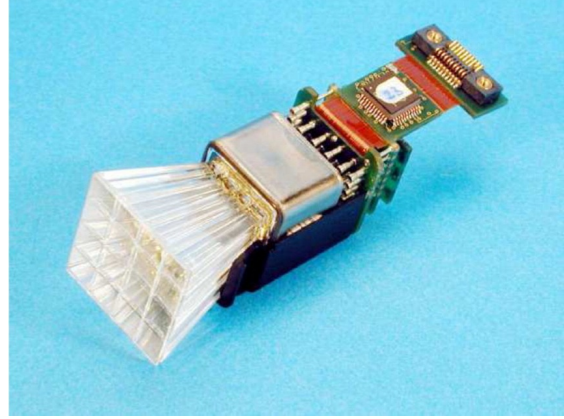
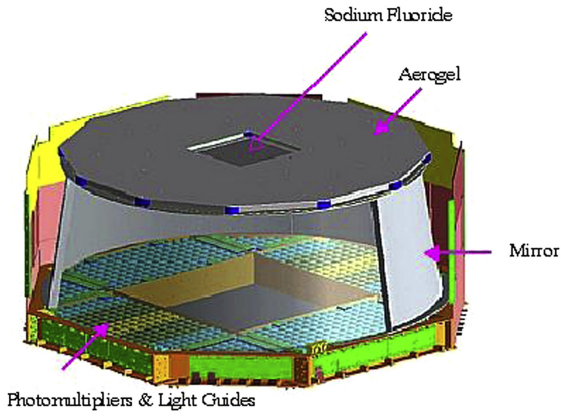


Figure 2.17: (Left) View of the AMS-02 RICH detector. From top to bottom, the radiator plane, the conical reflector and the detection plane. (Right) Individual detection cell including the light guide, the photomultiplier, the cell housing and the front-end electronics [103, 104].

much stronger than the Cherenkov photons. The PMTs are organised into four rectangular grids and four triangular grids, forming an octagonal configuration. An empty area of  $64 \times 64 \text{ cm}^2$ , corresponding to the size of the ECAL, is placed at the centre of the detection plane [103]. In this way, a travelling particle can reach the calorimeter without any further interaction.

After the launch of AMS-02 in 2011, several studies using flight data have been performed [104, 105]. Specifically, thanks to an accurate knowledge of the refractive indices, the velocity resolution for helium is  $\sigma_\beta = 0.8 \times 10^{-3}$  and the charge resolution is  $\sigma_Z = 0.3$  [104]. Figure 2.18 shows the velocity resolution on the left, and the charge resolution on the right as a function of the electric charge of different nuclei.

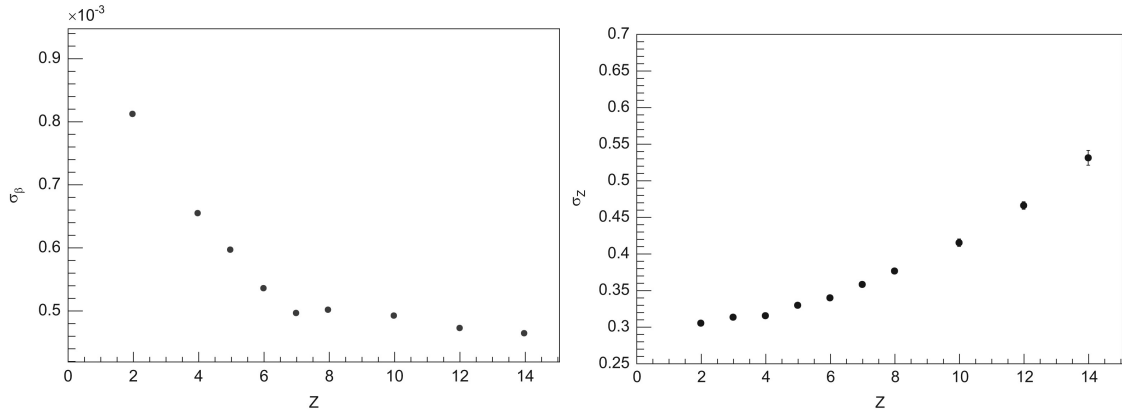


Figure 2.18: (Left) Velocity resolution as a function of the ion charge  $Z$ . (Right) Charge resolution as a function of ion charge  $Z$ . [104].

### 2.1.6 Transition Radiation Detector

The Transition Radiation Detector (TRD) is located at the top of AMS, as shown in Figure 2.2. The purpose of this subdetector is to identify electrons and positrons using the transition radiation, while rejecting protons and antiprotons. The TRD is a fundamental subsystem to perform indirect searches for dark matter candidates, measuring precisely the spectra of  $e^+$  and  $\bar{p}$ . The



transition radiation is an electromagnetic radiation emitted by highly relativistic charged particles traversing material with different refractive indices. Specifically, the amount of transition radiation emitted at the boundary of the two materials is proportional to the Lorentz factor  $\gamma = \frac{E}{mc^2}$  of the charged particle.

The Transition Radiation Detector (TRD) consists of 5248 proportional tubes of 6 mm diameter with a maximum length of 2 m. The tubes are assembled in 328 modules, each one with 16 tubes. The modules are divided into 20 layers, 12 along the y-axis and 8 along the x-axis. Two groups of four layers are placed above and below the 12 layers parallel to the y-axis [106]. This configuration provides a 3D reconstruction of the charge particle trajectory. Moreover, measuring the energy loss rate of charged nuclei traversing the detector, TRD provides an independent estimate of their charge [106]. The left picture of Figure 2.19 shows the TRD and its support structure in carbon fibre. The low coefficient of thermal expansion ensures the minimum relative movement of the TRD components with temperature variations. As shown in the right picture of Figure 2.19, each layer is separated by a fibre fleece radiator (LRP375BK) [106]. Such fiber is 20 mm thick and has a density of  $0.06 \text{ g/cm}^3$ . Each tube is filled with a mixture of two gases: 90% of Xe and 10% of  $\text{CO}_2$ . The Xenon captures the X-rays produced by the transition radiation in the radiator, whilst Carbon Dioxide stabilises the functioning of the tubes [106]. The ionisation signal is detected through an anode wire inside each tube. The gases for the tubes are provided thanks to two different tanks, one with 5 kg of  $\text{CO}_2$  and the second with 49 kg of Xe. After 14 years of smooth operation, the measured leak rate is dominated by the Carbon Dioxide with a loss of 0.47 g/day. These values ensure a total lifetime of the TRD longer than 20 years, until 2035.

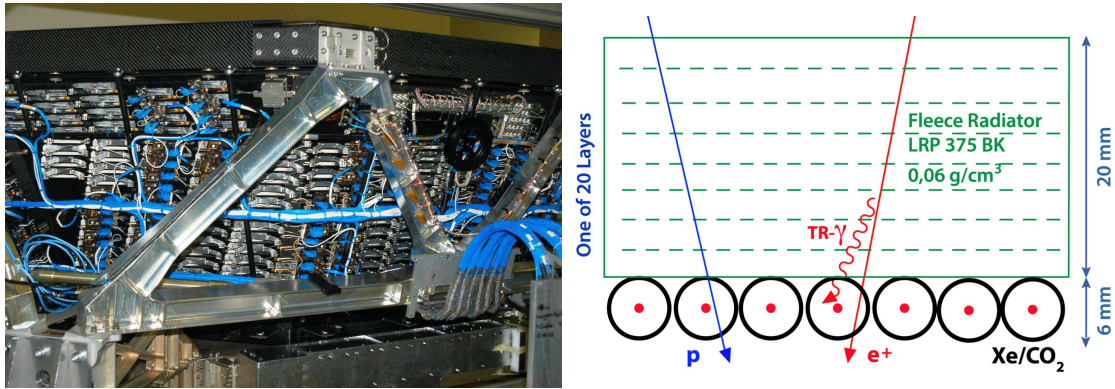


Figure 2.19: (Left) Image of Transition Radiation detector. (Right) A schematic representation of the ionisation losses detection from charge particles; the blue line represents protons, whilst the red arrow represents positrons [106, 84].

The proportionality between the number of transition-radiation photons and the Lorentz  $\gamma$  factor can be exploited by differentiating particles with different masses. As seen in Figure 2.20, where the signal of a single TRD tube is shown, the different  $\gamma$  values result in different distribution shapes. The protons (black dots) have a lower Most Probable Value (MPV), the electrons (red dots) have a higher MPV and a much higher tail, whilst helium nuclei show the highest MPV and a lower tail compared to electrons.

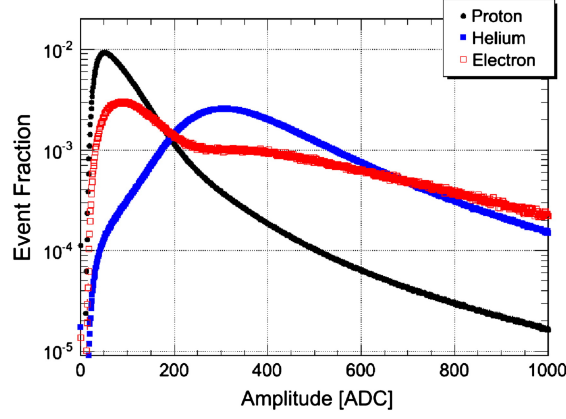


Figure 2.20: Single straw tube energy deposition spectra for proton, electron and helium events on clean single tracks [106].

The responses of the twenty layers are combined to build a statistical estimator  $TRD_{Lkh}$ . In the image of the estimator, two different peaks are clearly visible,  $e^\pm$  peak at  $TRD_{Lkh} = 0.35$  while  $p/\bar{p}$  peak at  $TRD_{Lkh} = 1$ . Selecting a cut value (working point) of 0.6, 90% of  $e^\pm$  are preserved whilst the rejection power against  $p/\bar{p}$  is higher then  $10^3$  up to 200 GeV/c. To keep the rejection even at the highest energies ( $> 200$  GeV), the cut is tighter with a loss of efficiency to 65% of  $e^\pm$ .

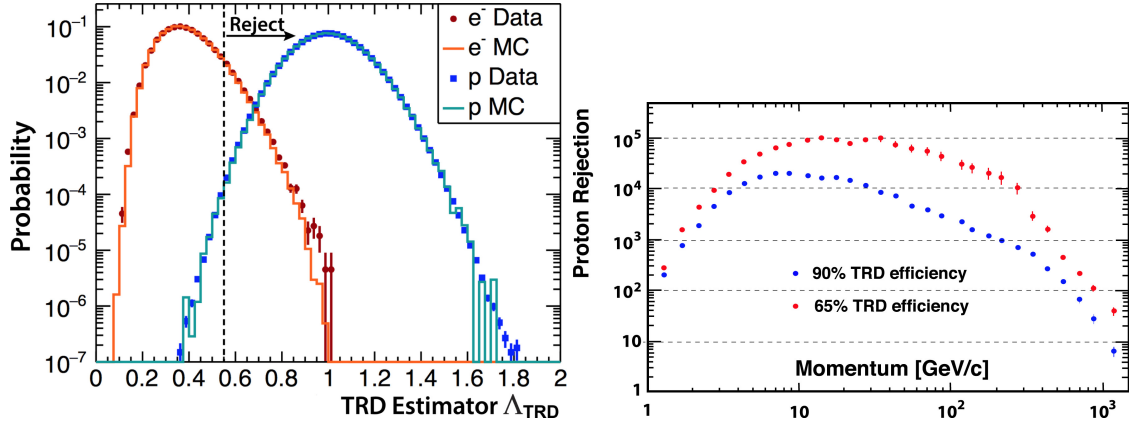


Figure 2.21: (Left) Distribution of the TRD estimator for cosmic rays protons (blue) and electrons (red) with  $R \in [10, 100]$  GV together with the corresponding Monte Carlo simulation. The vertical dashed line corresponds to a 90% efficiency cut. (Right) Proton rejection power provided by the TRD, as a function of rigidity, from ISS data at 90% and 65% electron and positron efficiency [106] [84].

### 2.1.7 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECAL) measures the energy deposited by electromagnetic particles and distinguishes between electron-like particles and hadrons. Combined with the other subdetectors of AMS-02, the total electron/hadron discrimination is higher than  $10^6$  [107]. The ECAL is a fine-grained lead-scintillating fibres sampling calorimeter that allows a precise three-dimensional imaging of the electromagnetic shower development. Thanks to these imaging capabilities, the shower shape of protons and electrons is quite different. Another difference is the ratio  $E/p$ , where  $E$  is the energy released in the ECAL and  $p$  is the momentum measured by the tracker. For electrons and positrons, which are well contained by the ECAL, the ratio is close to unity. On the other hand, protons are likely to escape the ECAL, producing a lower  $E/p$  value. Exploiting these differences, an ECAL estimator based on a multivariate approach uses Boosted Decision Trees (BDTs) to separate electrons and positrons from protons.

The ECAL contains lead and scintillating fibres. It has an active area of  $64.8 \times 64.8 \text{ cm}^2$ , a thickness of 16.6 cm, and an overall weight of 638 kg (including the support structure and the readout electronics). The detector has 9 substructures called superlayers. Each superlayer is 1.85 cm thick and made of 11 grooved 1 mm thin lead foils interleaved with layers of 1 mm diameter scintillating fibres glued to the foils by optical epoxy. The structure has a lead-fibre-glue volume composition of 1 : 0.57 : 0.15, resulting in an average density of  $6.8 \pm 0.2 \text{ g/cm}^3$ . The fibres run in a single direction for each superlayer. The left image of Figure 2.22 shows a cut view of a single superlayer. Stacking superlayers alternatively in orthogonal directions provides the imaging capabilities of the ECAL. Specifically, 5 superlayers have the fibres parallel to the x-axis, and 4 superlayers are parallel to the y-axis [107]. Each superlayer is read out by 36 four anode Hamamatsu R7600-00-M4 PMTs. The PMTs are placed alternatively at the opposite ends of the fibres, avoiding any dead areas. Each anode covers an active area of  $9 \times 9 \text{ mm}^2$ , defined as a "cell". Figure 2.22 shows a superlayer and the corresponding PMTs disposition. In total, the ECAL is divided into 1296 cells, read by 324 PMTs, divided into 18 longitudinal layers (two per superlayer), each one having 72 cells. This segmentation provides a high granularity, enabling a precise sampling of the longitudinal and transverse shower development.

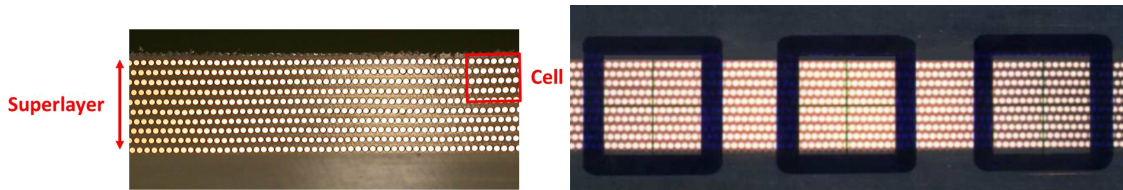


Figure 2.22: (Left) ECAL structure and cell dimensions. Each fibre has a diameter of 1 mm. (Right) Picture of a superlayer with the PMT positioning. Half of the PMTs are on the opposite side to avoid dead areas. [107] [84].

The ECAL has an equivalent thickness corresponding to 17 radiation length ( $X_0$ ) and 0.7 hadronic length ( $\lambda$ ) [107, 108]. Electromagnetic showers are well contained for electrons and positrons up to a few TeV, whilst protons escape the detector, generating partially contained hadronic showers.

After correcting for any lateral or rear energy leakage, the energy linearity is maintained within 1% up to 300 GeV [107]. The energy resolution  $\frac{\sigma_E}{E}$  was determined using measurements obtained during test beam [107, 109] and can be parametrised as:

$$\frac{\sigma_E}{E} = \sqrt{\frac{0.104^2}{E} + 0.014^2} \quad (2.5)$$

. Figure 2.23 shows the good agreement between data and the provided parametrisation.

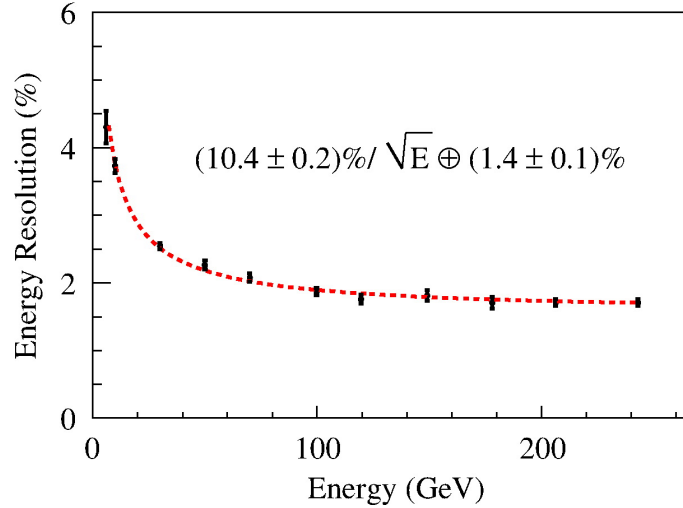


Figure 2.23: ECAL energy resolution for different test beam energies [107] [84].

## 2.2 The velocity resolution of AMS-02

AMS-02 has three different subdetectors used to measure the velocity of the incoming particles: the TOF, RICH-NaF and RICH-AGL. The TOF measures velocity up to  $\beta \sim 0.96$  for  $He$  nuclei, so the RICH covers the high velocity range up to  $R \sim 50$  GV for  $^4He$ .

The velocity resolutions of the TOF, RICH-NaF and RICH-AGL have been evaluated for both  $^3He$  and  $^4He$ . To compare the results, the residuals are obtained in bins of the relativistic factor  $\gamma - 1$  which is proportional to the kinetic energy per nucleon. Indeed, the kinetic energy per nucleon is derived starting from the relativistic kinetic energy  $E_k = (\gamma - 1)mc^2$  and expressing the nucleon mass as follows:

$$m = Zm_p + Nm_n - E_b \approx A \cdot \text{a.m.u.} \quad (2.6)$$

where  $m_p$  and  $m_n$  are the proton and neutron masses,  $Z$  is the atomic mass number,  $N$  is the number of neutrons,  $E_b$  is the binding energy,  $A$  is the mass number and a.m.u. is the atomic mass unit corresponding to  $\sim 0.931$  GeV/c<sup>2</sup>. Thus, the kinetic energy per nucleon can be written as:

$$\frac{E_k}{A} = 0.931(\gamma - 1) [\text{GeV/n}] \quad (2.7)$$

The velocity resolution is obtained from the MC simulation, evaluating the velocity residuals as:

$\frac{\beta_{MEAS} - \beta_{GEN}}{\beta_{GEN}}$ . Where  $\beta_{GEN}$  is the true velocity of the particle right above the detector of interest and  $\beta_{MEAS}$  is the measured velocity. The distributions of residuals have a Gaussian core, with the tails departing from Gaussian behaviour. Performing a fit on the residuals, the width of the distribution, which corresponds to the velocity resolution, is estimated. The chosen fit function is a combination of a Gaussian ( $G(x)$ ) and an Exponentially Modified Gaussian ( $EMG(x)$ ), where the two components can be expressed as follows:

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma_g} \exp\left(-\frac{(x - \mu_g)^2}{2\sigma_g^2}\right) \quad (2.8)$$

$$EMG(x) = \frac{\lambda}{2} \exp\left[\frac{\lambda}{2}(2\mu + \lambda\sigma^2 + 2x)\right] \text{erfc}\left(\frac{\mu + \lambda\sigma^2 + x}{\sqrt{2}\sigma}\right) \quad (2.9)$$

resulting in:

$$f(x) = A_g G(x) + A_e EMG(x) \quad (2.10)$$

where  $\mu_g$  and  $\sigma_g$  are the mean and standard deviation of the Gaussian,  $\text{erfc}$  is the complementary error function,  $A_g$  and  $A_e$  are two normalisation parameters. The EMG describes the sum of two independent normal and exponential random variables, thus an EMG random variable  $Z$  can be expressed as  $Z = X + Y$ , where  $X$  is Gaussian with mean  $\mu$  and variance  $\sigma$ , and  $Y$  is exponential of rate  $\lambda$ . The fit function has 7 free parameters to be determined during the fit procedure.

The Figure 2.24 shows the residual from  ${}^4\text{He}$  events passing through the RICH-AGL, in the following bin of  $\gamma_{GEN} - 1 \in [12.375, 12.500]$ .

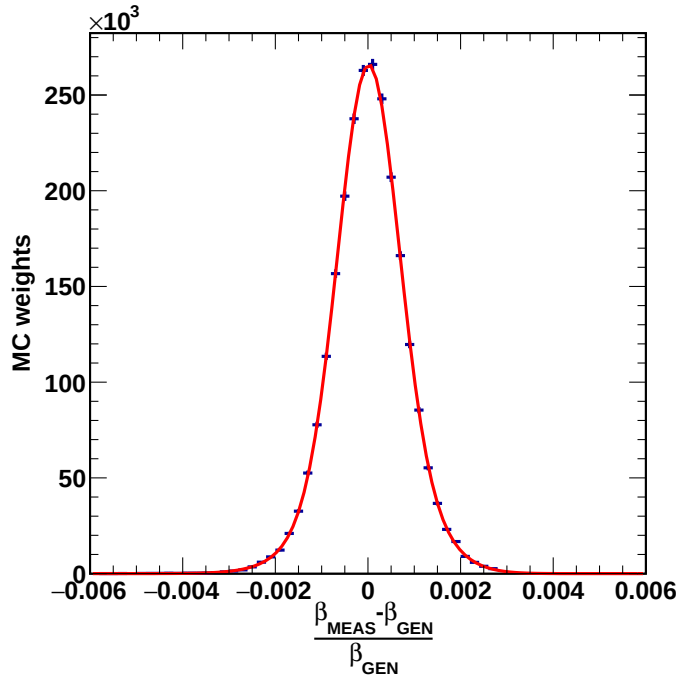


Figure 2.24: AGl pull for  $\gamma - 1 = 12.4$  and the corresponding fit function reported in equation 2.10.

### TOF $\beta$ resolution

The migration matrix for the TOF detector is shown on the left of Figure 2.25. The x-axis contains the MC-generated  $\beta_{GEN}$  right above the TOF, and the y-axis contains the MC-measured  $\beta_{MEAS}$  minus 1. The horizontal red dashed line corresponds to  $\beta_{MEAS} = 1$ , the black continuous lines correspond to  $3\sigma$  quantiles. When the upper black line meets the red line, the TOF is no longer capable of distinguishing subluminal particles from massless particles. Consequently, the TOF velocity measurements are reliable up to  $\beta \sim 0.96$ . The velocity resolution obtained by applying the fitting procedure with the function of equation 2.10 for  ${}^3\text{He}$  and  ${}^4\text{He}$  is shown on the right plot of Figure 2.25. The resolution for each isotope is similar, resulting in almost constant TOF  $\sigma_\beta \sim 1.5\%$ .

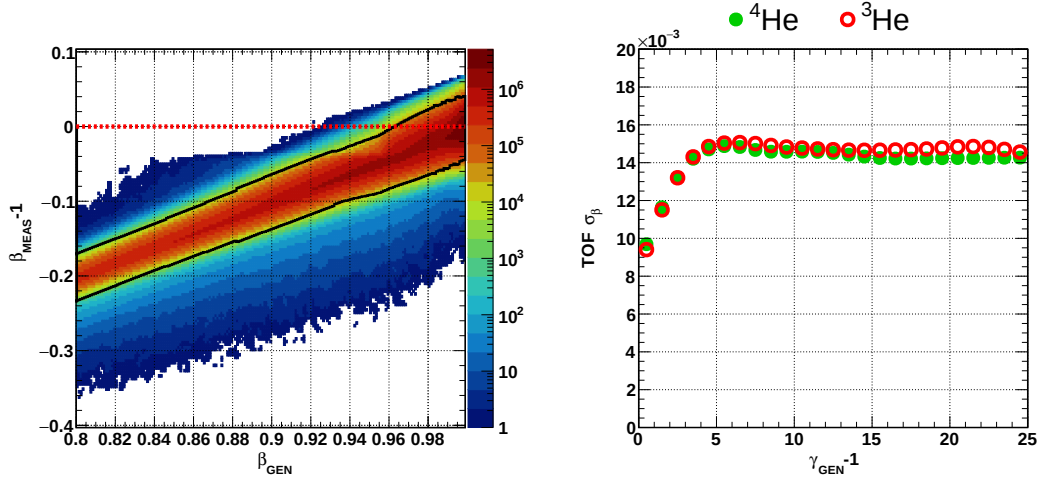


Figure 2.25: (Left) Velocity migration matrix for the TOF, only  ${}^4\text{He}$  events are displayed. The x-axis contains the MC-generated  $\beta_{GEN}$  right above the TOF, and the y-axis contains the MC-measured  $\beta_{MEAS}$  minus 1. The horizontal red dashed line corresponds to  $\beta_{MEAS} = 1$ , the black continuous lines correspond to  $3\sigma$  quantiles. (Right) TOF velocity resolution for  ${}^3\text{He}$  and  ${}^4\text{He}$  nuclei.

### RICH-NaF and RICH-AGL $\beta$ resolution

As previously described for the TOF, the migration matrices for the RICH-NaF and RICH-AGL detectors are shown on the left of Figure 2.26 and Figure 2.27. The upper limit where the NaF velocity measurements are reliable is up to  $\beta \sim 0.993$ , whilst the limit for AGL is around  $\beta \sim 0.998$ .

The right plots of Figure 2.26 and Figure 2.27 show the velocity resolution for RICH-NaF and RICH-AGL for each helium isotope, i.e.  ${}^3\text{He}$  and  ${}^4\text{He}$ . The resolution for each isotope is similar, resulting in almost constant NaF  $\sigma_\beta \sim 0.27\%$  and an AGL  $\sigma_\beta \sim 0.08\%$ .



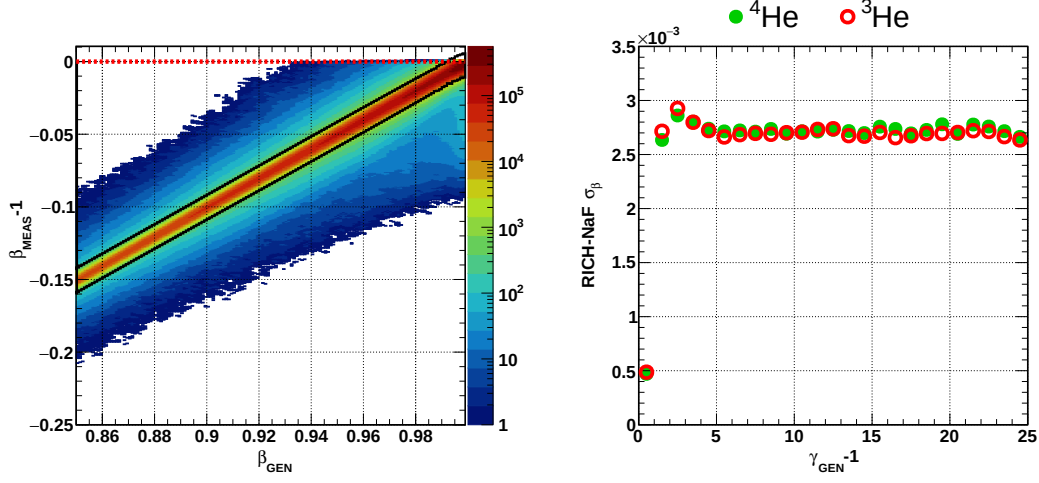


Figure 2.26: (Left) RICH-NaF velocity migration matrix, only  ${}^4\text{He}$  events are displayed. The x-axis contains the MC-generated  $\beta_{\text{GEN}}$  right above the NaF, and the y-axis contains the MC-measured  $\beta_{\text{MEAS}}$  minus 1. The horizontal red dashed line corresponds to  $\beta_{\text{MEAS}} = 1$ , the black continuous lines correspond to  $3\sigma$  quantiles. (Right) RICH-NaF velocity resolution for  ${}^3\text{He}$  and  ${}^4\text{He}$  ions.

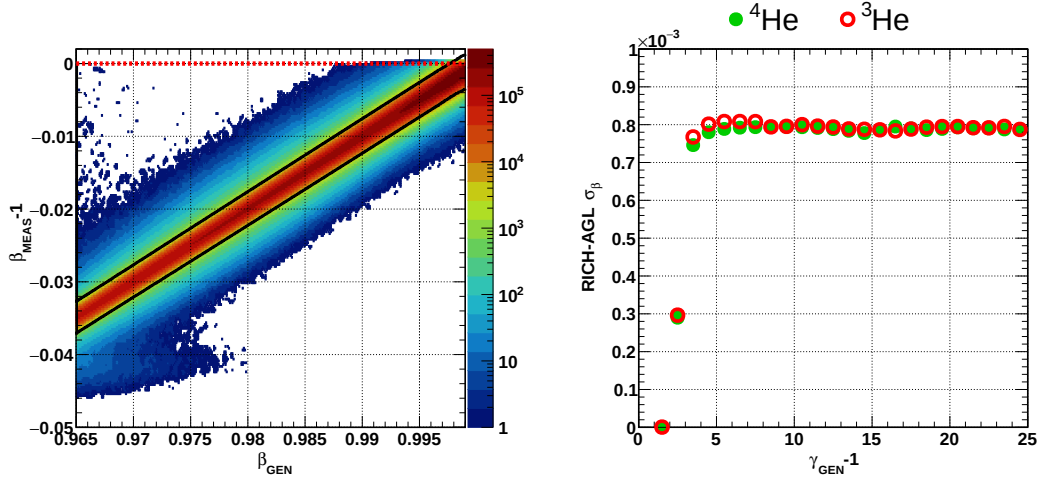


Figure 2.27: (Left) RICH-AGL velocity migration matrix, only  ${}^4\text{He}$  events are displayed. The x-axis contains the MC-generated  $\beta_{\text{GEN}}$  right above the AGL, and the y-axis contains the MC-measured  $\beta_{\text{MEAS}}$  minus 1. The horizontal red dashed line corresponds to  $\beta_{\text{MEAS}} = 1$ , the black continuous lines correspond to  $3\sigma$  quantiles. (Right) RICH-AGL velocity resolution for  ${}^3\text{He}$  and  ${}^4\text{He}$  ions

## 2.3 The mass resolution of AMS-02

An important variable to evaluate the goodness of a possible  $\overline{He}$  candidate is the reconstructed mass. Indeed, a well-reconstructed signal candidate must have a mass compatible with the  ${}^3He$  or  ${}^4He$  masses, corresponding to:  $m_{{}^3He} = 2.81 \text{ GeV}/c^2$  and  $m_{{}^4He} = 3.72 \text{ GeV}/c^2$ . The AMS-02 detector reconstructs the mass of the incoming particle by combining the measurements of rigidity and velocity. Starting from the relativistic expression of the momentum  $|p| = m\beta\gamma$ , the mass can be expressed as a function of rigidity and velocity:

$$m = \frac{p}{\beta \cdot \gamma} = \frac{R \cdot Z}{\beta \cdot \gamma} \quad (2.11)$$

where  $\gamma = \frac{1}{\sqrt{1-\beta^2}}$ . The error corresponding to the reconstructed mass is:

$$\sigma_m = m \cdot \sqrt{\left(\frac{\sigma_R}{R}\right)^2 + \gamma^4 \cdot \left(\frac{\sigma_\beta}{\beta}\right)^2} \quad (2.12)$$

The relative error on the measured rigidity is generally larger compared to the relative error on velocity; the presence of the factor  $\gamma^4$  makes the two terms comparable at higher rigidities.

The Figures 2.28 and 2.29 show the bidimensional plot of rigidity (on the x-axis) and mass (on the y-axis), respectively, for the TOF, RICH-NaF (left) and RICH-AGL(right). All mass distributions show a MC composition of  ${}^4He$  and  ${}^3He$  and broaden at higher rigidities. The diagonal empty regions in each plot correspond to the rigidity threshold required to reach the corresponding subdetector and the Cherenkov velocity thresholds, depending on the refractive index of the material. For example, the TOF has a lower mass resolution (see Figure 2.28), being reliable up to rigidities of  $\sim 7 \text{ GV}$ . The RICH-NaF, as already described in Section 2.1.5, has a Cherenkov velocity threshold of  $\beta_{th}^{NaF} = 0.75$ , corresponding to a rigidity of  $\sim 1.6 \text{ GV}$  for  ${}^3He$  and  $\sim 2.1 \text{ GV}$  for  ${}^4He$ . Due to the low number of Cherenkov photons emitted, the actual rigidity required to have a measurable ring is slightly higher, around  $R \sim 2.5\text{-}3 \text{ GV}$ . The RICH-AGL, has a  $\beta$  threshold of  $\beta_{th}^{AGL} = 0.953$  (see Section 2.1.5), corresponding to a rigidity of  $\sim 4.42 \text{ GV}$  for  ${}^3He$  and  $\sim 5.86 \text{ GV}$  for  ${}^4He$ . As for RICH-NaF, also in this case the rigidity threshold is higher in order to correctly reconstruct the Cherenkov ring, i.e.  $R \sim 6\text{-}7 \text{ GV}$ . The RICH-NaF can be used to measure the mass of helium nuclei up to  $R \sim 20 \text{ GV}$ , whilst the RICH-AGL is reliable up to  $R \sim 50 \text{ GV}$  as visible in Figure 2.29.

After taking into account the experimental resolution, for each one of the three AMS-02 subdetectors capable of measuring the particle velocity, most of the well-reconstructed helium events lay inside the following interval  $m_{REC} \in [2, 5] \text{ GeV}/c^2$ .



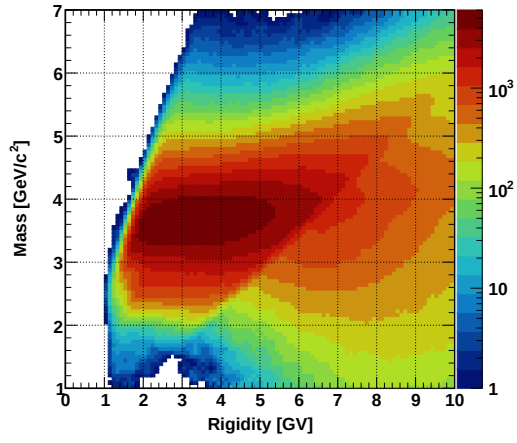


Figure 2.28: TOF mass resolution

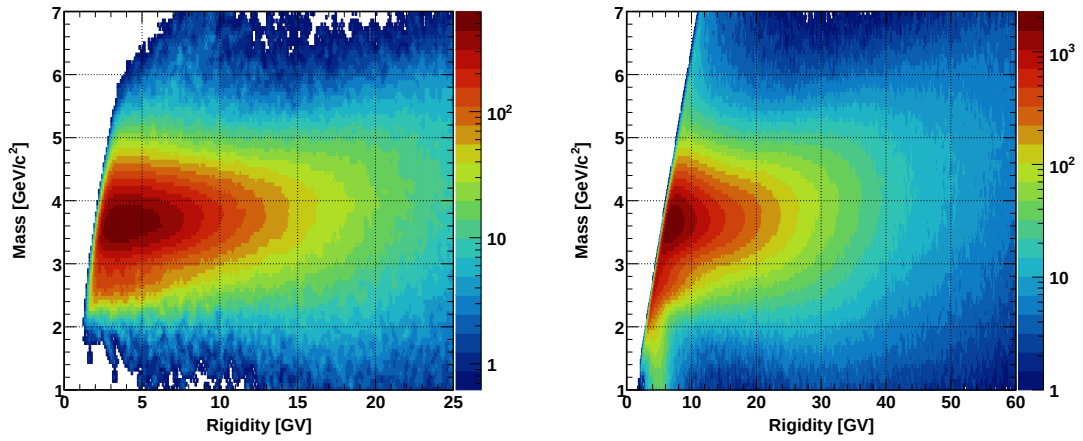


Figure 2.29: (Left) NaF mass resolution. (Right) AGL mass resolution

## 2.4 The Trigger Logic

The trigger logic aims to store any relevant event passing through the active elements of the detector and discards all other events. The decision-making process must be rapid to minimise the dead time of AMS-02. Indeed, every time an event is processed and eventually stored, the electronic is occupied and unable to collect other events. The AMS-02 trigger logic is designed to detect all species of charged cosmic rays with high efficiency. The system is divided into two parts: the Fast Trigger (FT) and the Level-1 trigger. The FT uses information from the TOF and the ECAL; if at least one of three different conditions is met, a fixed time of 1  $\mu$ s is used to combine more complex information from the TOF, ACC and ECAL. Afterwards, a Level-1 trigger signal is created, and the AMS-02 data acquisition is started, or the event is discarded.

### 2.4.1 Fast Trigger

The FT uses information obtained from TOF and ECAL. Three different conditions can fire a FT signal:

- FTC: simultaneous occurrence, within a time window of 240 ns, of signals in any three out of four TOF planes, each one with a pulse height greater than 0.5 times a  $Z = 1$  MIP deposit. This condition applies to all charged CRs [110].
- FTE: energy deposition in the ECAL, indicating an electromagnetic particle with an energy higher than 1 GeV. This criterion is specific for  $e^\pm$  and  $\gamma$ [109].
- FTZ: coincidence of signals from all four TOF planes, within a time window of 640 ns, each with a pulse height greater than 3.5 a  $Z = 1$  MIP. Due to the extended time windows and the higher threshold, this condition takes care of slow particles [110].

### 2.4.2 The Level-1 Trigger

The Level-1 trigger generates 7 sub-triggers combining information from the TOF, ECAL and ACC:

- Single Charge ( $Z = 1$ ): 4/4 TOF planes with a pulse height higher than 0.5 MIP and no ACC hits.
- Normal Ions ( $Z \geq 2$ ): 4/4 TOF planes with a pulse height higher than 3.5 MIP and less than 5 ACC hits. The condition on the ACC increases the efficiency due to gamma ray emission. After February 2016, the number of ACC has been moved to 8, to further improve the efficiency for very-high  $Z$  nuclei.
- Slow Ions ( $Z \geq 2$ ): 4/4 TOF planes with a pulse height higher than 3.5 MIP with an extended time window of 640 ns.

- Electrons: 4/4 TOF planes with a pulse height higher than 0.5 MIP and energy deposition inside the ECAL above a threshold in both directions. The ACC is disabled once the ECAL trigger fires, avoiding any possible self-veto scenario due to backsplash particles emerging from the shower in the calorimeter.
- Photons: ECAL shower above a certain threshold and a shower angle within the geometric acceptance of AMS-02.
- Unbiased charged particle: requires 3/4 planes signals from the TOF above the 0.5 MIP. It is used to estimate the trigger efficiency. Due to the loose requirement, a prescale factor of 1/100 is applied.
- Unbiased ECAL trigger: requires an ECAL shower above a certain threshold, used to estimate the ECAL trigger efficiency with a prescale factor of 1/1000.

### 2.4.3 Data Acquisition System

The AMS-02 Data Acquisition System (DAQ) is based on a tree structure of data collection modules, which collect data from approximately 300'000 analogue channels from all AMS-02 sub-detectors. It consists of over 300 computational nodes based on ADSP-2187L Digital Signal Processors and a Main DAQ Computer (JMDC) based on PPC750 Processor [111, 112]. Among these nodes, 264 are dedicated to gathering signals from the front-end electronics of each subdetector of AMS and performing data reduction. They are named xDR, where x depends on the specific subdetector (E for ECAL, R for RICH, T for Tracker, S for TOF and ACC, U for TRD) and DR stands for Data Reduction. The signals from the xDRs are collected by 14 Low Level DAQ computers (JINF). Trigger signals from TOF and ACC are gathered in 8 SDR nodes and transmitted to the JLV1 nodes, where the Level 1 trigger is generated. The High Level DAQ computers (JINJ) receive all the signals from JINF, SDR and JLV1. They transfer the data to the main computer of AMS (JMDC). The latter uses all the available information and determines whether an event is interesting or not [113].

The DAQ system is connected through a wired link protocol called AMSWire [112, 113]. The protocol has a data transfer rate of 100 Mbits/s, ensuring low latency and power consumption. The JMDCs contain a large buffer, called JBUX, which can store data for approximately one day. When communication between the ISS and the ground is interrupted, the buffer is used. In this way, the recording of data continues smoothly. The DAQ system operates on a 23 minute time interval. Each interval is called a run, resulting in 4 runs per a single ISS orbit. During the operational cycle, the calibrations of sub-detectors are carried out every 2 runs, while the ISS is close to the equator [113, 112]. A schematic representation of the DAQ system is reported in Figure 2.30.

To transmit the data from the ISS to the ground, the NASA Tracking and Data Relay Satellites (TDRS) are used. These satellites send data to NASA's White Sands Test Facility in New Mexico.

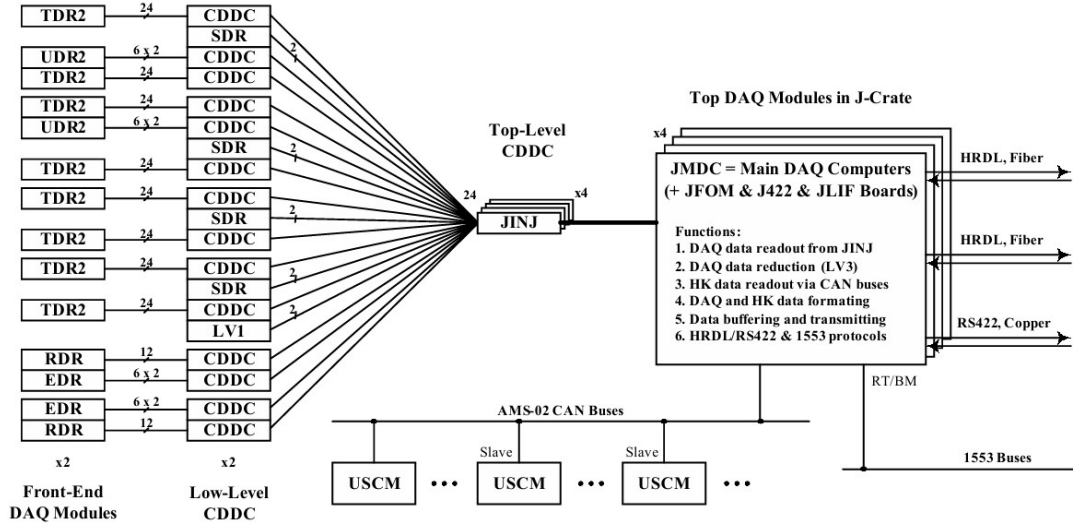


Figure 2.30: AMS DAQ system [112].

Afterwards, data are transferred to NASA's Payload Operations Integration Centre (POIC) at the Marshall Space Flight Centre (MSFC) in Alabama. In the end, data are sent to the AMS Payload Operations Control Centre (POCC) at CERN and the Asia POCC. In the last two locations, the operation of each subdetector and the data collections are continuously monitored 24/7 throughout the whole year.

#### 2.4.4 Flight data, livetime and trigger rate

After the launch in 2011, the DAQ system has acquired data steadily. A few exceptions occurred between October and November 2014, and for small periods between August 2018 and November 2019. During the first time interval, the thermal control system of the tracker was checked by carrying out specific operations. The second time interval is the result of a non-standard tracker configuration caused by issues with the thermal system. These problems have been resolved by installing the Upgraded Tracker Thermal Pump System (UTTPS) during four Extra Vehicular Activities (EVAs) between November 2019 and January 2020.

The DAQ system requires some time to complete the operations mentioned above. The level 1 trigger, the data reduction, and transmission occur within approximately 220  $\mu$ s and during this time, the detector is unable to collect new data. The DAQ status is evaluated in 20 ns samples, and the livetime is the ratio between the number of samples where the detector was not busy and the total number of samples. This calculation is performed every second. Essentially, the livetime is the DAQ efficiency and measures the amount of time during which the detector is acquiring data.

Recorded runs, useful for a scientific purpose, include only those seconds when the detector was in nominal operating conditions, pointing within 40 deg of the local zenith, with a livetime higher than 50%, and the ISS is outside the SAA. The AMS-02 acquisition rate depends heavily on the ISS orbit. This is clearly visible in the Figures 2.32 and 2.33, showing the mean trigger rate and

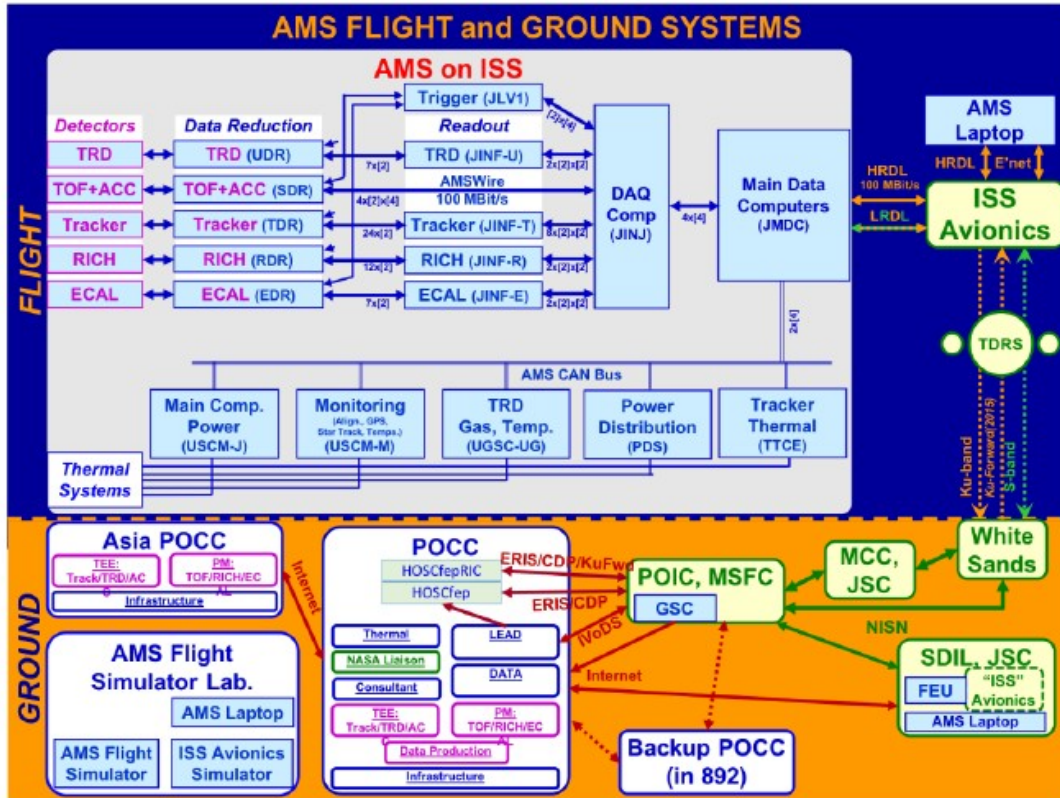


Figure 2.31: The data flow (clockwise from upper left) from the AMS subdetectors, through the AMS main computer, ISS avionics, TDRS satellites, down to the White Sands ground station, to MSFC, and in the end AMS-POCC at CERN and in Taiwan [92]. The commands follow the reverse path.

livetime as a function of latitude and longitude. The number of particles hitting AMS-02 is directly related to the Earth's magnetic field, which deflects or traps particles with a rigidity lower than a certain threshold, depending on the orbit position, the direction and the charge sign of the incoming particle. The threshold is called the geomagnetic cutoff [114]. The geomagnetic cutoff is estimated by backtracing particles from the top of AMS-02 out of 50 Earth's radii using the IGRF magnetic model [115]. Any particle with a rigidity below the cutoff cannot be a galactic cosmic ray.

The Van Allen radiation belt are areas where the magnetic field captures charged particles. The area of the inner Val Allen belt closest to the Earth's surface is the South Atlantic Anomaly (SAA). Within this region, the number of particles reaching AMS increases up to saturation [116].

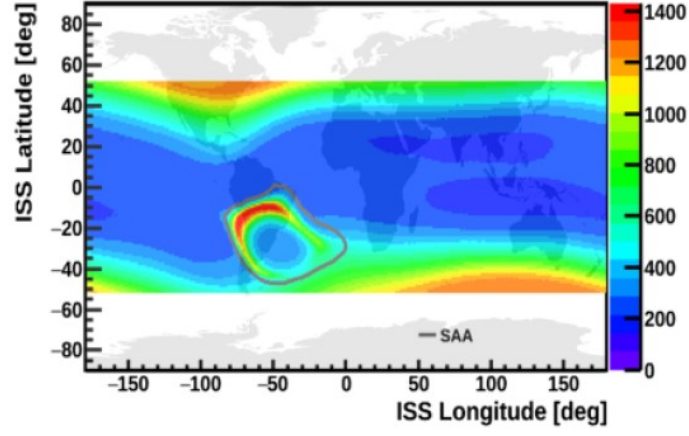


Figure 2.32: Mean trigger rate as a function of latitude and longitude. Near higher latitudes, higher trigger rates result in DAQ saturation. The same behaviour is observed near and within the SAA [92].

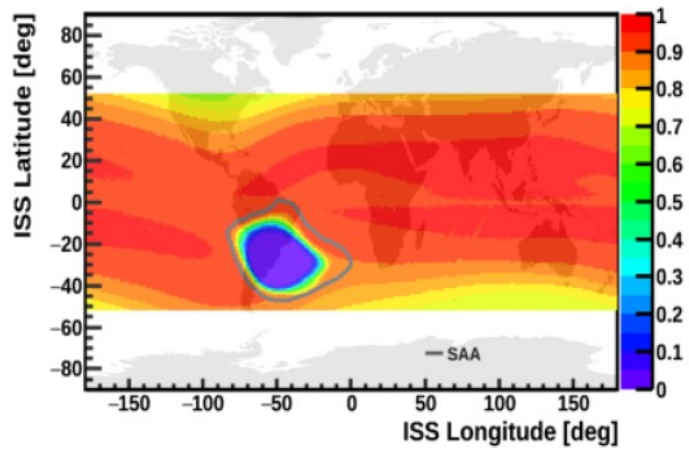


Figure 2.33: Mean livetime as a function of latitude and longitude. Near higher latitudes, lower livetime results in DAQ saturation. The same behaviour is observed near and within the SAA [92].

## Chapter 3

# Event selection and Monte Carlo simulation

The starting point of this analysis is the selection of events where a downgoing particle is reconstructed with charge  $|Z| = 2$ , and it has a well-measured rigidity. Those events should correspond to He nuclei entering AMS-02; however, given the detector limits, the sign of the charge is sometimes reconstructed as negative. These misreconstructed events represent the most relevant background to the search for anti-He signals in the CR flux.

In this Chapter, we present a set of selection criteria applied to both the ISS data and the Monte Carlo simulation to obtain the starting samples for our analysis. I also present how we setup the AMS-02 Monte Carlo simulation and how we obtain an absolute normalisation of the simulated statistics, to be able to compare the ISS data directly to the simulated ones.

### 3.1 Event selection for identification of *He* events

The data sample used in this analysis corresponds to 12.5 years of AMS-02 data taking on the ISS, corresponding to  $\sim 239 \times 10^9$  recorded events.

The first selections concern the data quality and the AMS-02 conditions during the orbits, and as a consequence, are only meaningful for the ISS data.

#### 3.1.1 Geomagnetic environment, timestamp RTI and trigger requirements

The (AMS) Real-Time Information (RTI) is a special data format, which stores on a time-based database, the most relevant information about the detector position, attitude and status averaged over 1-second time intervals.

RTI contains the latitude and longitude of the ISS station, the AMS pointing direction, the AMS DAQ livetime fraction<sup>1</sup>, and the number of triggers occurring in a second.

---

<sup>1</sup>The measurement of the livetime fraction has been discussed in Section 2.4.4

In this preliminary selection, no requirement on rigidity above the geomagnetic cutoff has been applied to better study the possible background from misreconstructed helium events.

Here is a summary of the (pre)selections applied based on information from RTI:

- ISS not in the SAA: the seconds when AMS-02 is within the SAA are not included, the region is removed, applying a geographic cut on the latitude-longitude plane.
- Livetime fraction  $> 0.5$ : the livetime fraction measures, in a one-second time interval, the DAQ efficiency. The seconds with a livetime smaller than 0.5 are disregarded.
- Zenith angle  $< 40^\circ$ : seldom, AMS can deviate significantly from its original orientation  $\sim 12^\circ$  from the zenith, due to ISS operations. All seconds when AMS was pointing more than  $40^\circ$  away from the zenith are removed.

As previously discussed in Section 2.4.2, the AMS-02 electronics acquire an event when one or more different trigger conditions are satisfied. Good charge 2 events correspond to events satisfying the requirements of at least one between the normal and the slow ion triggers<sup>2</sup>.

Another important requirement is that the measured particle is entering AMS-02 from the top and crosses the magnet downward. In fact, an upgoing He nucleus will bend in the tracker as a downgoing anti-He nucleus, and this would be a tedious background source.

The TOF system measures the time of flight of a charged particle between its 2 planes, separated by about 1.2 m, with a resolution of the order of 180 ps. Since the minimum time of flight for a relativistic particle is of the order of  $\sim 4$  ns, the possibility of misreconstructing an upward-going particle as a down-going particle is totally negligible, also for a very rare search like the anti-He one.

### 3.1.2 Track pattern and number of reconstructed tracks

To achieve the highest possible sensitivity for anti-He, we want to take advantage of the largest possible acceptance ( $\sim 0.5 \text{ m}^2\text{sr}$ ). In AMS-02, this is obtained considering events with tracks reconstructed in the inner tracker, without requiring them to have a hit in Layer 1 (top of the instrument) or Layer 9 (in front of the calorimeter).

The rigidity resolution for charge  $|Z| = 2$  particles with a track reconstructed in the inner tracker corresponds to the Maximum Detectable Rigidity (MDR) of  $\sim 400$  GV. To validate an anti-He candidate event, we want to use an independent velocity measurement to determine whether the particle's mass is consistent with that of a helium nucleus. Using the TOF and the RICH detectors, we can have a reliable velocity measurement up to the equivalent rigidity of  $\sim 50$  GV. In summary, then the rigidity resolution provided by the inner tracker is enough for our purposes. The number of reconstructed tracks is an indicator of interactions occurring within the detector. Additionally, detector noise may generate ghost tracks. Nuclei like helium have a higher probability to produce  $\delta$  rays (energy secondary electrons) compared to  $Z = 1$  particles. Taking into

---

<sup>2</sup>The details of both triggers are reported in Section 2.4.2



account possible  $\delta$  rays, only events with at most 2 reconstructed tracks are included. The number of reconstructed tracks within the inner tracker for data,  $^4\text{He}$  and  $^3\text{He}$  MC are displayed in Figure 3.1.

To ensure a well-reconstructed track, even if no additional hits on the external layers are required, a request for the track pattern is applied:

- Number of reconstructed tracks  $\leq 2$ .
- Track pattern: a hit, on both bending (Y) and non-bending direction(X), on the layer  $L_2$  must always be present. Moreover, at least 6/7 hits along the Y coordinate and 4 hits along the X coordinate must be present, with at least one hit on each inner tracker plane.

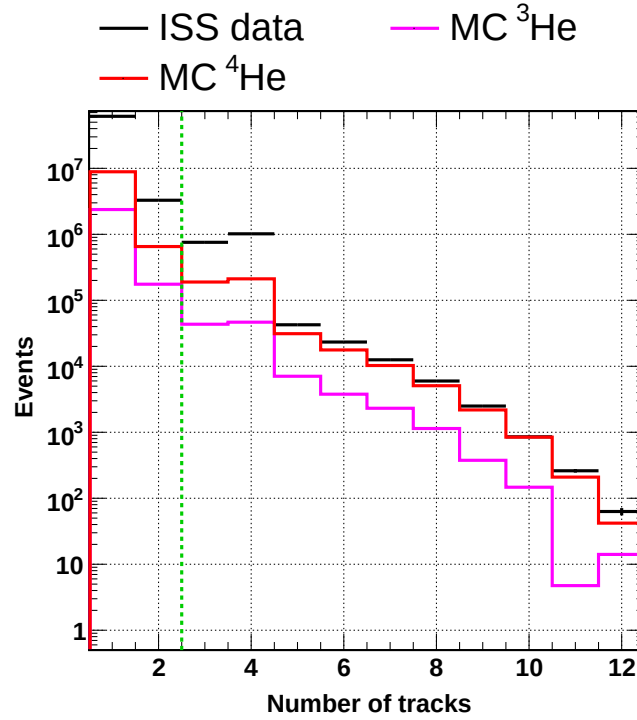


Figure 3.1: Number of reconstructed tracks within the inner tracker. The black histogram is ISS data before any other selection, the continuous lines are  $^4\text{He}$  (red) and  $^3\text{He}$  (magenta) MC used to show the corresponding distribution for both helium isotopes. The green vertical line shows the chosen cut value. More details are available in the text and in table 3.1.

### 3.1.3 Tracker charge identification

The measured particle charge is a very effective variable to select a good sample of  $|Z| = 2$  events, and this measurement is performed by almost all the AMS-02 subdetectors.

The charge measurement obtained in the inner tracker is the most accurate. Indeed, the inner tracker measures the charge of the traversing particle independently in each of its 7 silicon layers. This redundancy allows for high-precision measurement as presented in Figure 2.11.

The charge resolution of the inner tracker allows us to efficiently select a pure sample of  $|Z| = 2$  events, using this information alone.

The used charge event selection is:

- Inner tracker charge  $Z_{inner} \in [1.7, 2.4]$ : the requirement rejects the proton tail and avoids any possible contribution from Lithium.

The distributions of the inner tracker reconstructed charge for data,  ${}^4\text{He}$  and  ${}^3\text{He}$  MC are displayed in Figure 3.2. Using the MC simulation, we verified that contamination from the fragmentation of  $Z > 2$  nuclei in the upper part of AMS-02 is negligible.

Cosmic electrons, due to their low mass value  $m_{e^-} \sim 0.5 \text{ MeV}/c^2$ , are highly relativistic and may leave an energy deposit in the silicon larger than 1 m.i.p., in extreme cases, this may get close to a charge 2 energy deposit and introduce a contamination of the sample.

Indeed, as expressed in equation 2.3, particles with high  $\beta\gamma$  suffer larger energy losses that might lead to the misidentification of an electron as a  $|Z| = 2$  particle. The possible presence of  $e^-$  contamination will be further discussed in Section 4.1.

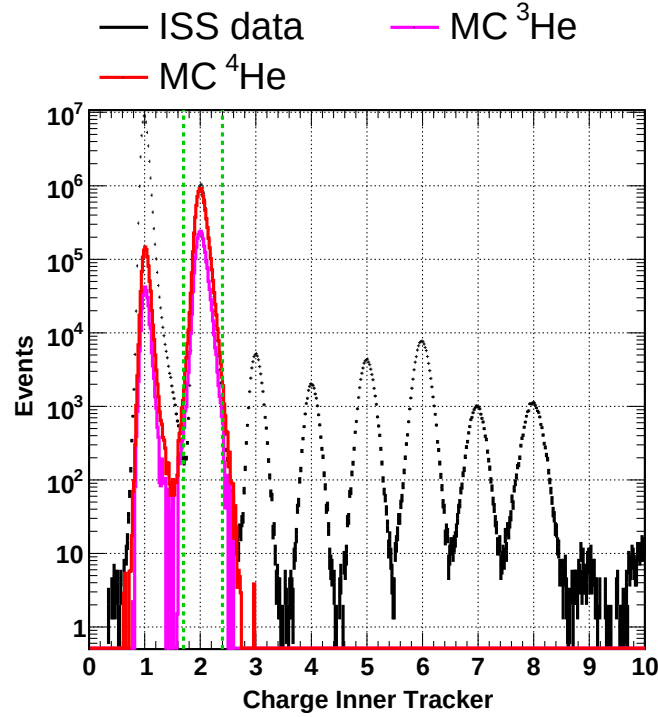


Figure 3.2: Distribution of charge absolute value measured by the inner tracker. The black dots are ISS data before any other selection, the continuous lines are  ${}^4\text{He}$  (red) and  ${}^3\text{He}$  (magenta) MC used to show the corresponding distribution for both helium isotopes. The green vertical line shows the chosen cut values. The selection on the inner tracker is a very effective way to select a  $|Z| = 2$  sample. More details are available in the text and in table 3.1.

### 3.1.4 Reconstructed track quality selection

Further constraints are used to select well-reconstructed tracks. Indeed, previous works within the AMS collaboration have shown that most of the charge-confused events occur at the edge of the inner tracker layers. A charged-confused event is a particle that is wrongly reconstructed with the opposite charge sign. Since charge-confused helium nuclei are the main background of this analysis, as will be described in the Section 4.1.1, a fiducial volume cut is used.

The quality of the track reconstruction is quantified by two normalised  $\chi^2$ , one for each coordinate:  $Y$  (bending direction) and  $X$  (non-bending direction). Cutting on these variables reduces the contamination of badly reconstructed tracks. The distributions of the two  $\chi^2$  for ISS data, after the selection on the inner tracker charge, are shown in the plots of Figure 3.3.

The details of the three cuts are reported here:

- Inner tracker fiducial volume: for each layer of the inner tracker, a cut on the absolute value of the bending coordinate ( $Y$ ) and on the radius ( $R = \sqrt{x^2 + y^2}$ ) of the track hit is applied. The cut values are:
  - $L_2$ :  $|Y| < 40 \text{ cm}$   $|R| < 62, \text{ cm}$
  - $L_3, L_4$ :  $|Y| < 44 \text{ cm}$   $|R| < 46, \text{ cm}$
  - $L_5, L_6$ :  $|Y| < 36 \text{ cm}$   $|R| < 46, \text{ cm}$
  - $L_7, L_8$ :  $|Y| < 44 \text{ cm}$   $|R| < 46, \text{ cm}$
- $\chi_Y^2/\text{d.o.f.} \leq 10$ : quantifies the quality of the reconstructed track along the bending coordinate.
- $\chi_X^2/\text{d.o.f.} \leq 10$ : rejects possible large-angle scattering occurring within the detector (see Section 4.1.1).

### 3.1.5 Coherence between different reconstructed rigidities

The rigidity measured by the inner tracker is compared with the rigidities measured by the upper part ( $R_{UH}$ , layers  $L2$ - $L_6$ ) and lower part ( $R_{LH}$ , layers  $L3$ - $L_8$ ) of the inner tracker. All three rigidities must have the same sign, i.e. the charge sign of the incoming particle must be the same. Despite the correlation between the three quantities, this comparison highlights possible interactions occurring within the detector. Additionally, a lower threshold on  $R_{inner}$  is used ( $|R_{inner}| < 1.2 \text{ GV}$ ) for two reasons: remove low-energy particles that suffer large energy losses and curvatures stopping inside of AMS, and to avoid a rigidity region where the expected background increases, as shown in Figure 3.8.

### 3.1.6 Event geometry and TOF reconstruction

The events used in this analysis satisfy the following cuts:

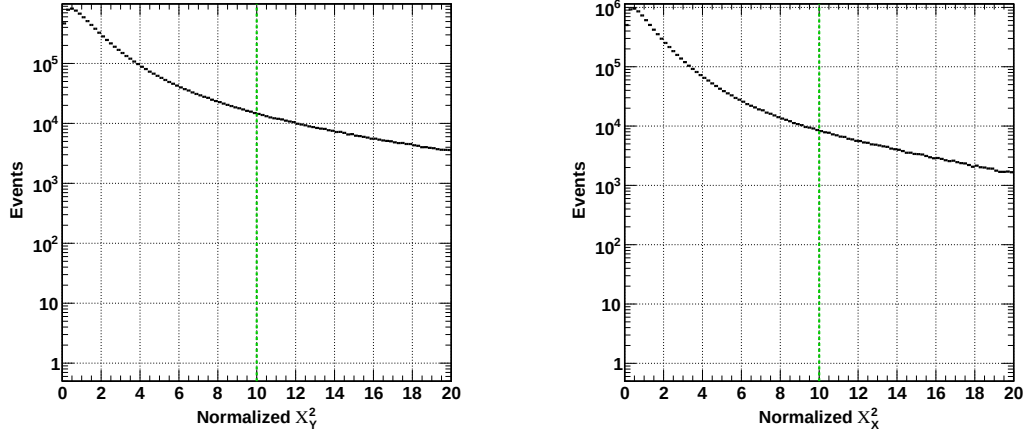


Figure 3.3: (Left) ISS data distribution of the track normalised  $\chi^2$  along the bending direction (Y). (Right) ISS data distribution of the track normalised  $\chi^2$  along the non-bending direction (X). Both distributions are obtained after the charge selection on the inner tracker; thus, the data sample displayed is mainly composed of helium nuclei. More details are available in the text and in table 3.1.

- 4/4 TOF layers: the incoming particle must leave a signal in all 4 layers of the TOF.
- $\beta_{TOF} > 0.3$ : the incoming particle direction must be downgoing, and the threshold rejects low-energy particles which will not go through the detector due to large curvature and energy losses.
- $\chi_{time}^2 / \text{d.o.f.} < 10$

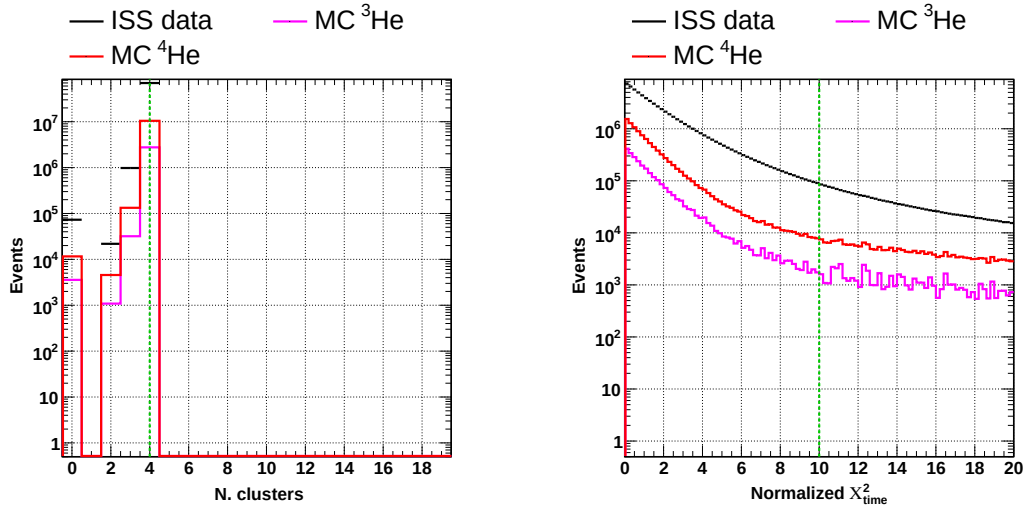


Figure 3.4: (Left) Number of TOF layers crossed. (Right) Normalised temporal chi-square from TOF. The black dots are ISS data before any other selection, the continuous lines are  $^4\text{He}$  (red) and  $^3\text{He}$  (magenta) MC used to show the corresponding distribution for both helium isotopes. The cut value is shown by the green vertical line. More details are available in the text and in table 3.1.

### 3.1.7 TOF charge identification

After the selection on the inner tracker charge, two further cuts are applied on the reconstructed charge obtained from the upper (UTOF) and lower (LTOF) TOF. The request on the upper TOF allows for reducing the fragmentation background above the inner tracker due to  $Z > 2$  nuclei, such as  $Li$  or  $C$ . The lower TOF helps to reduce possible events where a  $\delta$  ray is emitted and traverses the RICH detector. The charge distributions, after the selection on the inner tracker charge, for ISS data,  $^4He$  and  $^3He$  MC of the upper and lower TOF are presented in Figure 3.5. The good agreement between simulation and data for the upper TOF charge shows that the contamination due to fragmentation above the inner tracker is negligible.

The selected charge thresholds are:

- $Z_{UTOF} \in [1.5, 3.0]$
- $Z_{LTOF} \geq 1.5$ .

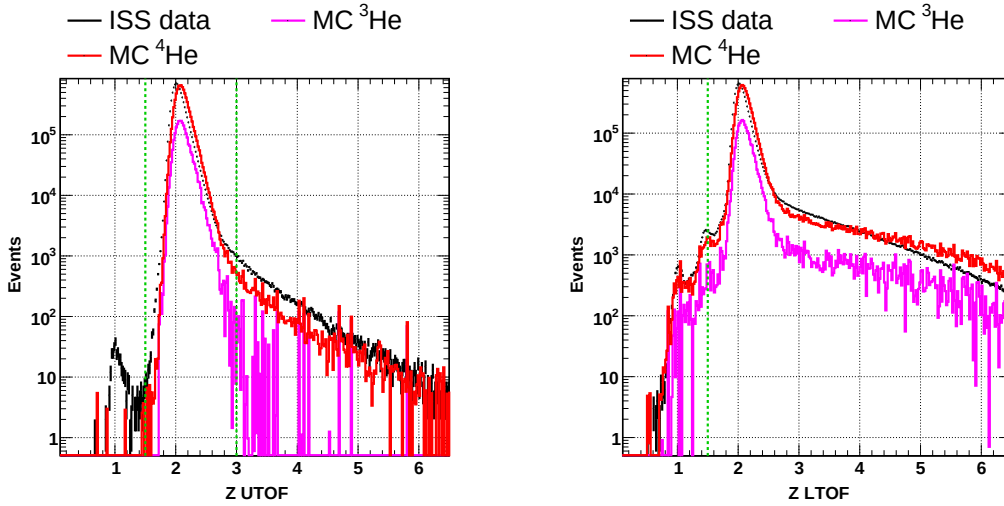


Figure 3.5: Distribution of charge absolute value measured by the upper (right) and lower (left) TOF. The black dots are ISS data, the continuous lines are  $^4He$  (red) and  $^3He$  (magenta) MC. The green vertical line shows the chosen cut values. More details are available in the text and in table 3.1.

### 3.1.8 Requirements on the reconstructed velocity

The goal of this work is to search for possible antihelium candidates by looking at the reconstructed mass of each candidate. As explained in Section 2.3, the mass is measured starting from the rigidity and the particle velocity. Therefore, requirements on the measured velocity are applied to select well-reconstructed events. The cuts described in Section 3.1.6 select well-reconstructed  $\beta$  measured by the TOF. For the RICH measurements, which are not always present, the following criteria are applied:

- $\beta_{RICH} < 1$

- $|\beta_{RICH} - \beta_{TOF}| \leq 0.06$ : the compatibility between the TOF and RICH  $\beta$  must be within  $4\sigma_{TOF}$  for suppressing possible delta rays traversing the RICH detector.
- number of RICH hits used for the ring reconstruction  $> 3$ .
- if  $\beta_{TOF} \geq 0.96$  then RICH must be present, otherwise the event is discarded. The 0.96 corresponds to the maximum  $\beta$  value at which the TOF provides a reliable measurement.

The distribution for ISS data, after the charge selection on the inner tracker, upper and lower TOF, of the compatibility between the velocity measurements of TOF and RICH, and the  $\beta$  measurements for events crossing the RICH-NaF (left) and the RICH-AGL (right) are displayed in Figures 3.6 and 3.7.

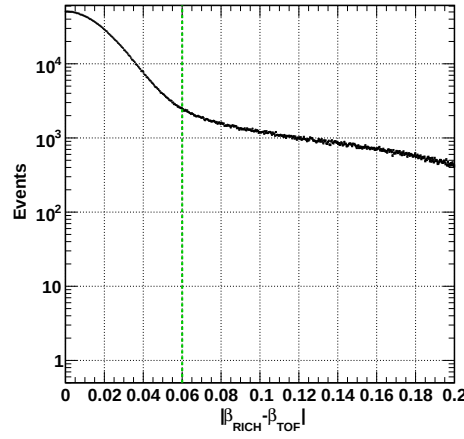


Figure 3.6: The plot shows the compatibility between the velocities measured by TOF and RICH. The green lines show the cut values. The events displayed are obtained after the charge selection on the inner tracker, the upper and lower TOF; thus, the data sample displayed is mainly composed of helium nuclei. More details are available in the text and in table 3.1.

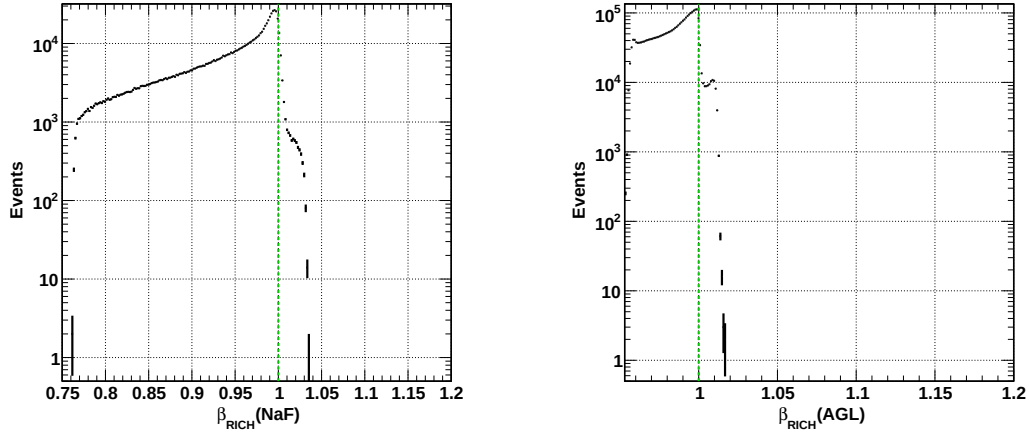


Figure 3.7: Distribution of the measured RICH velocity ( $\beta$ ) for ISS data passing through the NaF (left) and the AGL (right) radiator. The green lines show the cut values. Both distributions are obtained after the charge selection on the inner tracker, the upper and lower TOF; thus, the data sample displayed is mainly composed of helium nuclei. More details are available in the text and in table 3.1.

### 3.1.9 ISS data selection summary

Applying the described event selection to ISS data, two samples are obtained: one with positive rigidities ( $Z = +2$ ) and one with negative rigidities ( $Z = -2$ ). In total, around  $2.4 \times 10^9$  positive helium events and  $5.7 \times 10^4$  events with negative rigidities have been selected.

The Figure 3.8 shows the obtained spectra for  $Z = -2$  events for 12.5 years of data collection and helium simulation, as a function of the absolute value of rigidity. The spectrum has a maximum at  $|R_{inner}| \sim 400$  GV, corresponding to the MDR of the inner tracker. For lower rigidities, the spectrum decreases up to  $\sim 2$  GV and then increases again. The rise at low rigidities is due to high-curvature events, which are not cosmic rays but rather particles trapped by Earth's magnetic field. These events can be rejected by applying an additional selection criterion on the geomagnetic cutoff that is not included in the event selection, in order to better study the charge-confused helium background.

Thanks to the MC normalisation procedure, which will be described in the next Section (see Section 3.2.1), the simulation shows a good agreement at low rigidities. This result is very important for the analysis workflow, because it allows us to use the MC as a control region in the statistical analysis performed to estimate the significance of any possible excess in data<sup>3</sup>. At the same time, for  $|R_{inner}| > 100$  GV, there is a shift between data and MC. The deviation at high rigidities does not affect the results of this analysis, as the mass resolution of AMS-02 allows for searching for antihelium in the following rigidity interval  $|R_{inner}| \in [2, 50]$  GV using the TOF and RICH (NaF and AGL) subdetectors.

As visible in Figure 3.8, ISS data and helium simulation have an overall good agreement, showing that almost the totality of the events are charged-confused helium nuclei. The Monte Carlo

<sup>3</sup>Detailed discussions are reported in Chapter 4, Section 4.5, and in Chapter 5.

simulation is a key tool to study the charged-confused background and its possible sources. Any possible antihelium candidates lie below a considerable background and, as will be presented in the next Chapter, machine learning algorithms can help to reveal a possible signal hint.

A summary of the event selection applied to ISS data and MC simulation is presented in Table 3.1.

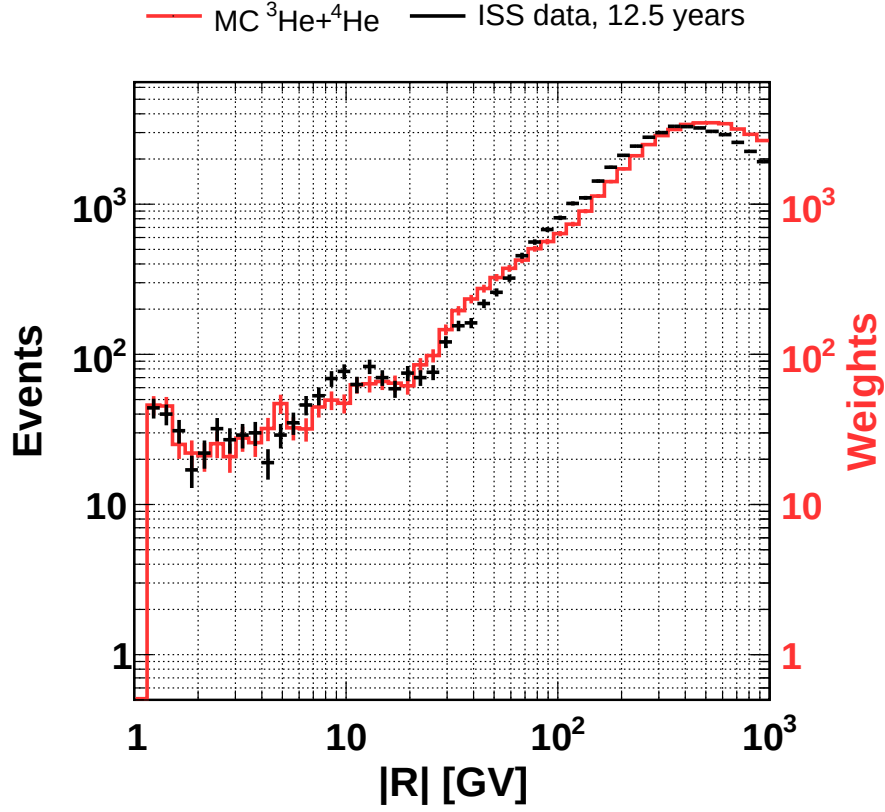


Figure 3.8: Events distribution as a function of rigidity, after the event selection described in Section 3.1. The ISS data of 12.5 years are the black dots, the composition of  $^3\text{He}$  and  $^4\text{He}$  MC is the continuous red line. The MC has been reweighted following the procedure described in Section 3.2.1. The agreement between data and simulation is good apart from a deviation at high rigidities  $|R_{\text{inner}}| > 100$  GV that does not affect the results of this analysis as explained in Chapter 4.



| Detector                                                                                     | Applied cut                                                                                                                                                                                                                                                                                                                    |
|----------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| RTI selections                                                                               |                                                                                                                                                                                                                                                                                                                                |
|                                                                                              | ISS not in the SAA<br>Livetime > 0.5 s<br>Zenith angle < 40°                                                                                                                                                                                                                                                                   |
| TOF                                                                                          | 4/4 TOF layers<br>$\beta_{TOF} > 0.3$                                                                                                                                                                                                                                                                                          |
| TRK                                                                                          | $\chi^2_{time}/\text{d.o.f.} < 10$<br>Number of tracks $\leq 2$<br>$L_2 \& (L_3 L_4) \& (L_5 L_6) \& (L_7 L_8)$<br>6/7 y-coordinate hits<br>4/7 x-coordinate hits<br>$Z_{inner} \in [1.7, 2.4]$<br>inner tracker fiducial volume<br>$\chi^2_Y(\text{inner})/\text{d.o.f.} < 10$<br>$\chi^2_X(\text{inner})/\text{d.o.f.} < 10$ |
| UTOF                                                                                         | $Z_{UTOF} \in [1.5, 3.0]$                                                                                                                                                                                                                                                                                                      |
| LTOF                                                                                         | $Z_{LTOF} > 1.5$                                                                                                                                                                                                                                                                                                               |
| TRK                                                                                          | $\left(\frac{R_{inner}}{R_{UH}} > 0\right) \& \left(\frac{R_{inner}}{R_{LH}} > 0\right)$                                                                                                                                                                                                                                       |
| TOF-RICH                                                                                     | $ R_{inner}  > 1.2 \text{ GV}$<br>if $\beta_{TOF} > 0.96 \rightarrow \text{RICH}$                                                                                                                                                                                                                                              |
| if RICH event is present                                                                     |                                                                                                                                                                                                                                                                                                                                |
| RICH                                                                                         | $\beta_{RICH} < 1$<br>$ \beta_{RICH} - \beta_{TOF}  \leq 0.06$<br>Ring number of hits $\geq 4$                                                                                                                                                                                                                                 |
| $(R_{inner} > 0) \& (R_{UH} > 0) \& (R_{LH} > 0) \rightarrow \text{prescale factor} = 1/100$ |                                                                                                                                                                                                                                                                                                                                |

Table 3.1: Summary of the event selection used for ISS data and MC simulation (excluding the RTI cuts). A detailed description of each cut is discussed in Section 3.1.

## 3.2 Monte Carlo simulation

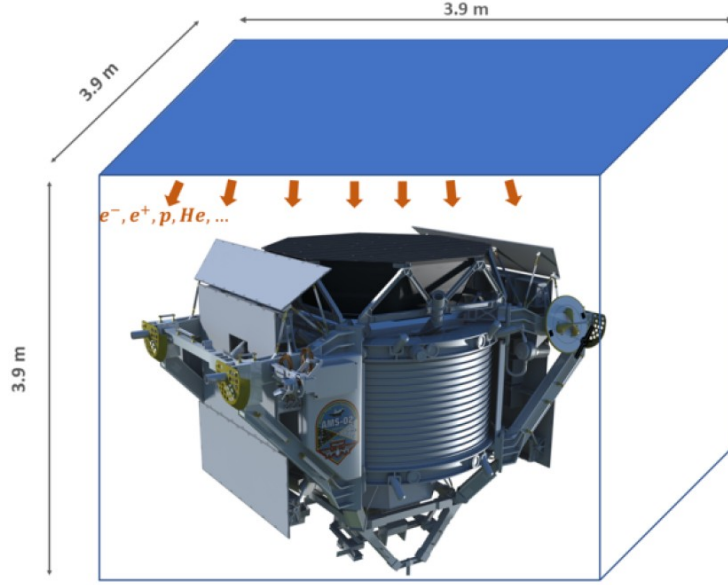


Figure 3.9: Generation scheme of the AMS-02 Monte Carlo simulation [92].

The AMS software is used for both data reconstruction and Monte Carlo (MC) event simulation. The simulation process relies on the GEANT4 package [117, 118, 119]. The simulation includes a detailed representation of the detector, including its geometry, material composition and the interaction between particles and the detector itself. The simulated particles are generated in an isotropic manner from the upper plane of a cube measuring  $3.9 \times 3.9 \times 3.9 \text{ m}^3$ , which contains the detector as shown in Figure 3.9. Each generated particle has a known rigidity, randomly extracted from a uniform distribution with respect to the natural logarithm of rigidity. Thus, the distribution of generated rigidity has a spectral index of  $-1$ , i.e. the number of events  $N$  in a rigidity bin  $[R_1, R_2]$  is:

$$N_{[R_1, R_2]} = \int_{R_1}^{R_2} \frac{dN}{dR} dR = \int_{R_1}^{R_2} \frac{dR}{R} = k \ln \left( \frac{R_2}{R_1} \right) \quad (3.1)$$

where  $k$  is a constant. This approach ensures sufficient statistics at high rigidities, reducing the required computational time if the true spectral index of CRs were used, i.e.  $\frac{dN}{dR} \propto R^{-2.7}$ . As a result, reweighting techniques must be applied to reproduce the experimental data collected in space.

### 3.2.1 Monte Carlo normalisation

In an analysis searching for rare events, such as antihelium nuclei, it is fundamental to quantify the statistics of the MC used and compare it with the available statistics of the data. Moreover, it is key to normalise the Monte Carlo to reproduce the rigidity spectra observed in the data and be able to compare the simulation directly with the collected data. The MC normalisation has been

performed starting from the fluxes measured by AMS [84, 120] and computing for each rigidity bin a MC equivalent exposure time following this relation:

$$T_i^{MC} = \frac{N_i^{GEN}}{A^{GEN} \cdot \Phi_i^{INT}} = \frac{N_i^{GEN}}{A^{GEN} \cdot (\Phi_i \cdot \Delta R_i)} \quad (3.2)$$

where the subscript  $i$  indicates the  $i$ -th rigidity bin,  $N_i^{GEN}$  is the number of generated events within the given rigidity bin,  $A^{GEN}$  is the acceptance of the generation plane (i.e.  $A^{GEN} = 3.9^2 \cdot \pi [\text{m}^2\text{sr}]$ ),  $\Phi_i$  is the average flux in the bin published by AMS ( $\Phi_i [\text{m}^{-2}\text{s}^{-1}\text{sr}^{-1}\text{GV}^{-1}]$ ) and  $\Delta R_i$  is the width of the  $i$ -th rigidity bin. Equation 3.2 allows for reweighting the MC and comparing it directly with the exposure time measured by AMS. The weights are computed for each rigidity bin using:

$$w_i = \frac{T_i^{ISS}}{T_i^{MC}} \quad (3.3)$$

with the same notation as before, and where  $T_i^{ISS}$  is the exposure time in the  $i$ -th rigidity bin. Figure 3.10 shows the MC equivalent time for  $e^-$ ,  $^3\text{He}$  and  $^4\text{He}$  together with the exposure time for positive and negative rigidities. The exposure times differ slightly at low rigidities ( $R \leq 30 \text{ GV}$ ) due to the geomagnetic cutoff, which is a charge-sign-dependent physical effect caused by the magnetic field of the Earth. The rigidity modulation of the AMS-02 exposure time can be extracted from the RTI database.

The maximum exposure time is  $2.67 \times 10^8 [\text{s}]$  from May 2011 to November 2023, equivalent to 12.5 years of data tacking. The vertical jumps in the  $^4\text{He}$  histograms correspond to different MC production, each one with a different generated rigidity interval. The MC with the higher statistic is the  $^4\text{He}$  MC, with a time equivalent always higher than the time exposure of the data. The simulations of electrons and  $^3\text{He}$  have less statistics, especially at lower rigidities, but their contribution is subdominant compared to  $^4\text{He}$ .

The weights obtained using the equation 3.2 and 3.3 are shown in the Figures 3.11 and 3.12. The available MC statistic for  $^4\text{He}$  (left plot in the figure) is at the same order of magnitude as the data around 10 GV, whilst for higher rigidities it is around twice the collected ISS data. On the other hand, the statistics of  $^3\text{He}$  (central plot) and electrons (right plot) are respectively  $\frac{1}{70}$  and  $\frac{1}{10}$  of the available data at  $R_{GEN} = 5 \text{ GV}$  and  $3 \text{ GV}$ .

The normalisation of the Monte Carlo (MC) simulation is performed using the most recent published results from the AMS-02 collaborations. The helium flux [84] has been used to reweight both the  $^3\text{He}$  and  $^4\text{He}$  MC. The isotopic contribution has been taken into account using the measurement of the  $^3\text{He}/^4\text{He}$  ratio [84, 120] and multiplying the weight obtained from the equation 3.3 by the corresponding isotope fraction at a given rigidity. Starting from the publication of the AMS collaboration, the ratio  $\frac{^3\text{He}}{^4\text{He}}$  is described by the following parametric expression:

$$\frac{^3\text{He}}{^4\text{He}} = C \left( \frac{R}{4 \text{ GV}} \right)^\Delta = w(\text{ISO})_i^{^3\text{He}} \quad (3.4)$$

a fit to the data above  $R = 4 \text{ GV}$  returns:  $C = 0.1476 \pm 0.0004$  and  $\Delta = -0.294 \pm 0.004$ . Below

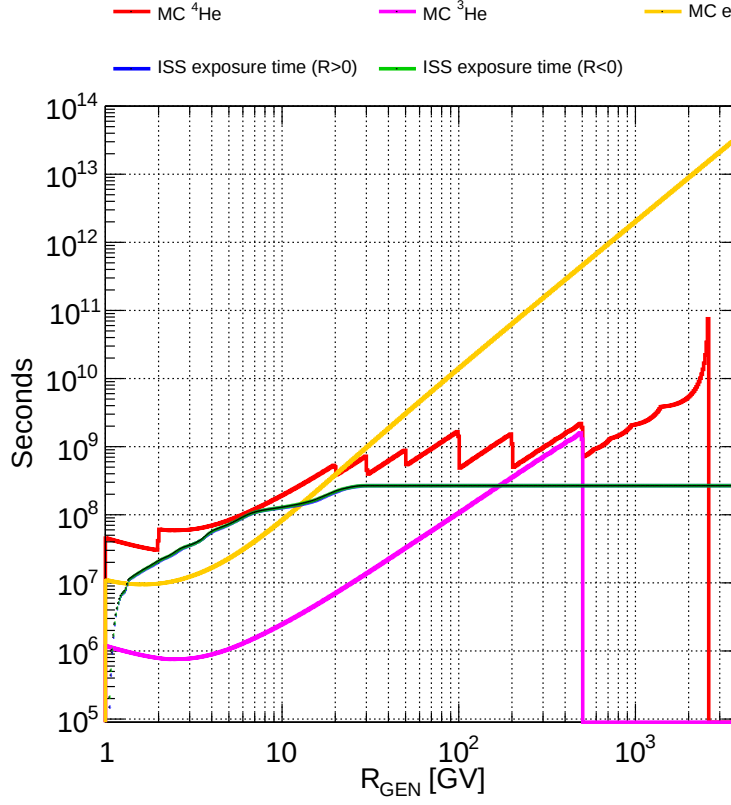


Figure 3.10: Monte Carlo equivalent exposure time for  $^4\text{He}$ ,  $^3\text{He}$  and electrons as a function of the generated rigidity. The distributions are obtained using the equation 3.2 and compared with the exposure time of data for positive and negative rigidities.

4 GV the ratio changes with time and an average value of the spectral index is  $\langle\Delta\rangle = 0.21 \pm 0.02$ . The measured isotopic composition of the helium flux, as a function of rigidity, is shown in Figure 3.13. For rigidities higher than  $\sim 20$  GV, the same parametric description presented in equation 3.4 is used, assuming no change in the isotopic ratio. In conclusion, the weights for the two helium isotopes are:

$$w_i^{^3\text{He}} = w_i \times w(\text{ISO})_i^{^3\text{He}} \quad (3.5)$$

$$w_i^{^4\text{He}} = w_i \times \left(1 - w(\text{ISO})_i^{^3\text{He}}\right) = w_i \times w(\text{ISO})_i^{^4\text{He}} \quad (3.6)$$

$$(3.7)$$

where  $w_i$  is obtained from equation 3.3,  $w(\text{ISO})_i^{^3\text{He}}$  and  $w(\text{ISO})_i^{^4\text{He}}$  are respectively the fraction of  $^3\text{He}$  and  $^4\text{He}$  in the  $i$ -th rigidity bin.

Similarly, the  $e^-$  MC has been reweighted using the parametric description used to fit AMS-02 data. Indeed, within the energy range  $[0.5, 1400]$  GeV, the electron flux is described by the sum of

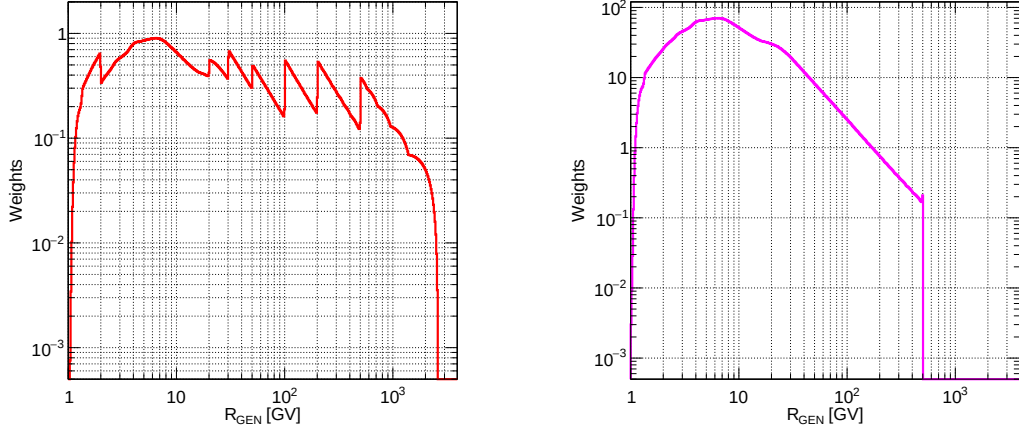


Figure 3.11: Weights distributions, obtained using equation 3.3, for  $^4\text{He}$ (left),  $^3\text{He}$  (right) MC.

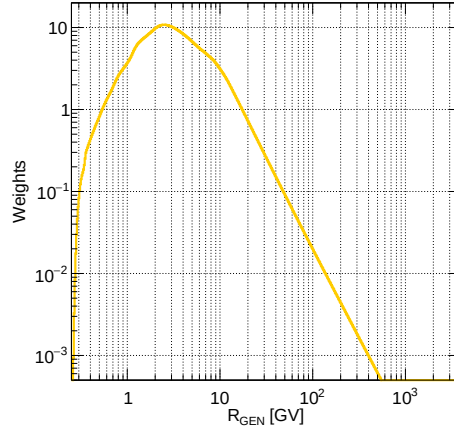


Figure 3.12: Weights distributions, obtained using equation 3.3, for electrons MC.

two power laws:

$$\Phi_{e^-}(E) = \frac{E^2}{\hat{E}} \left[ 1 + \left( \frac{\hat{E}}{E_t} \Delta\gamma_t \right) \right]^{-1} \left[ C_a \left( \frac{\hat{E}}{E_a} \right)^{\gamma_a} + C_b \left( \frac{\hat{E}}{E_b} \right)^{\gamma_b} \right] \quad (3.8)$$

The solar modulation effects are taken into account using the force-field approximation [121, 122] and defining  $\hat{E} = E + \phi_{e^-}$ , where  $\phi_{e^-}$  is the effective modulation potential. The term  $\left[ 1 + \left( \frac{\hat{E}}{E_t} \Delta\gamma_t \right) \right]^{-1}$  has vanishing impact on the flux behavior at energies above  $E_t$ . The constant  $E_a$  is set to 20 GeV to minimise the correlation between parameters  $C_a$  and  $\gamma_a$ ; the constant  $E_b$  is chosen to be 300 GeV to minimise the correlation between  $C_b$  and  $\gamma_b$ . The fit to the data yields:  $E_t = 3.94 \pm 0.21$  [GeV],  $\gamma_t = -2.14 \pm 0.09$ ,  $C_a = (1.13 \pm 0.08) \times 10^{-2}$  [ $\text{m}^{-2}\text{sr}^{-1}\text{s}^{-1}\text{GeV}^{-1}$ ],  $\gamma_a = -4.31 \pm 0.13$ ,  $C_b = (3.96 \pm 0.04) \times 10^{-6}$  [ $\text{m}^{-2}\text{sr}^{-1}\text{s}^{-1}\text{GeV}^{-1}$ ],  $\gamma_b = -3.14 \pm 0.02$  and  $\phi_{e^-} = 0.87 \pm 0.12$  [GeV] [84].

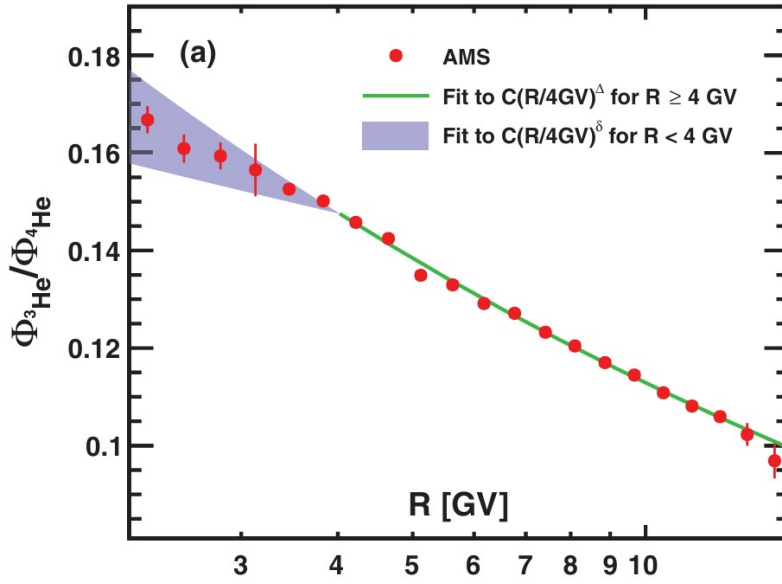


Figure 3.13: The time-averaged  $\frac{^3\text{He}}{^4\text{He}}$  ratio as a function of rigidity  $R$  (red points) with statistical and uncorrelated systematic errors added in quadrature. The solid green curve shows a single power law fit using equation 3.4 above 4 GV. The shaded blue area indicates the results of a single power law fit to the same equation for 21 time periods between May 2011 and November 2017 [120, 84].

## Chapter 4

# Overview of the AMS-02 data analysis for $\overline{He}$ searches

In this Chapter, an overview of the analysis workflow is discussed. In the first section, the expected background sources and the main characteristics of the signal are presented. The peculiarities of each sample are directly compared with the event selection discussed in Chapter 3. Afterwards, I introduce some relevant concepts of Machine Learning, such as the loss function, the Feed-Forward Multi-Layer Perceptron, Deep Neural Networks (DNNs), and backpropagation. I describe two specific approaches: a classification task based on supervised learning and an anomaly detection task based on unsupervised learning. In the second section, I describe the two Machine Learning approaches developed in this work. For each approach, I describe the rationale for its use, then focus on the structure and software used to build each algorithm. Both DNNs are based on the PyTorch library v.2.6 [123]. The input feature selection procedure and the training process are discussed in detail, referring to the contents introduced in the first part of this Chapter. I present the performance of each Neural Network (NN) individually, then I expand on the combination of the two approaches and how to use it in the search for  $\overline{He}$  events in CRs. In the last section, the workflow of the analysis is described, focusing on the background suppression strategy with four different Deep Neural Networks.

### 4.1 Signal and expected background sources

In Chapter 1, the production mechanisms of antinuclei (see Section 1.5), especially antihelium, along with some recent estimates for antihelium searches near the Earth (Section 1.6), have been presented and discussed. Antinuclei are unlikely to be produced during the propagation of Cosmic Rays (CRs), and the expected fluxes due to astrophysical sources are suppressed. Independent works from the PHENIX [124] and the ALICE [76] collaboration have confirmed this.  $\overline{He}$  is a possible golden channel, as any enhanced detection of such heavy antinuclei might be related to Dark Matter (DM) annihilations or some primordial origin. A rough estimate of the  $\overline{He}$  to

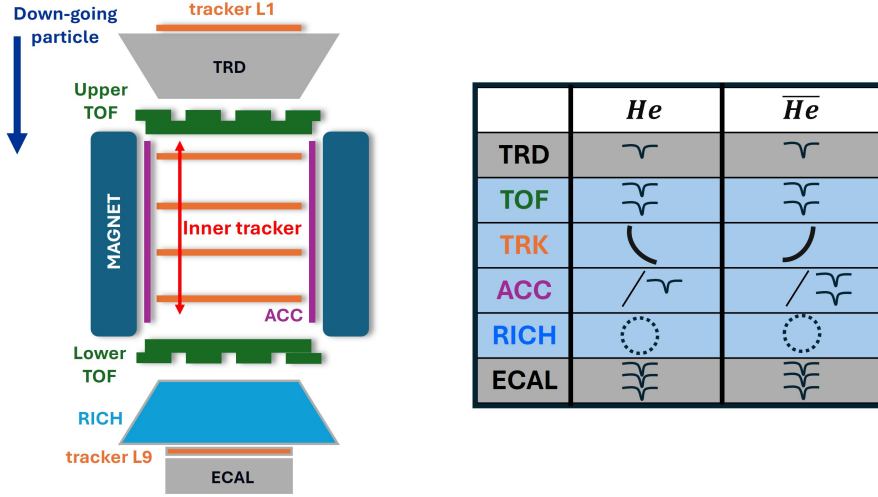


Figure 4.1: The left image shows a pictorial cut view of the AMS-02 detector [125]. The right table shows how particles and antiparticles interact within each subsystem.

$He$  production ratio is  $\sim 10^{-9}$ ; dedicated experiments must possess high charge discrimination capability to enhance their sensitivity to  $\overline{He}$ .

The Figure 4.1 presents a pictorial cut view of the AMS-02 detector, a table showing how helium and antihelium nuclei are expected to interact within each subdetector, and the coloured lines indicate the subsystems used in this analysis. This work aims to identify any antihelium candidates corresponding to a clean event, with a negative reconstructed rigidity, a precise measurement of the velocity and a reconstructed mass compatible with those of  ${}^3He$  and  ${}^4He$  nuclei. Such an alleged  $\overline{He}$  should go through almost all AMS-02 up to the RICH. Looking at the table in Figure 4.1, excluding antihelium events annihilating inside the detector,  $He$  and  $\overline{He}$  act almost identically apart from the curvature of the reconstructed track.

Since any possible signal yield is expected to be very small, Machine Learning algorithms have been used to suppress the background and uncover eventual signal candidates. As will be further discussed in this Chapter, these algorithms must be trained and tuned. This work relies on a fully data-driven approach, applying the same strategy independently to data and Monte Carlo simulations.

The event selection described in Chapter 3, Section 3.1, has been applied to MC samples and ISS data obtaining four different samples: two with positive rigidities, i.e.  $Z = +2$  (one for ISS data and one for MC), and two with negative rigidities corresponding to  $Z = -2$  (one for ISS and one for MC). The  $Z = +2$  samples are dominated by well-reconstructed events, as can be seen from the reconstructed mass distribution as a function of rigidity for ISS data shown in the Figures 4.2 and 4.3. For NaF and AGL (Figure 4.3), the diagonal empty regions correspond to the Cherenkov velocity threshold depending from the refractive index of the material. The same plots but for a combination of  ${}^3He$  and  ${}^4He$  MC are reported in Figure 2.28 and Figure 2.29.

Since the goal of this work is to identify possible antihelium candidates corresponding to a clean event going through the AMS-02 detector up to the RICH, positive rigidity samples represent a



good benchmark for possible well-reconstructed antihelium candidates. For this reason, the two  $R_{inner} > 0$  samples (one for ISS data and one for MC) are used to train the Machine Learning algorithms. The models trained on MC are not applied to data and vice versa. This choice avoids possible MC-data disagreements arising from, for example, small discrepancies in the detector material budget description that might be relevant for rare interactions occurring inside the detector. Additionally, Monte Carlo simulations might be limited by the lack of experimental data regarding relevant cross sections. Nevertheless, a data-driven approach is complementary with respect to a Monte Carlo-based analysis and assuming a true antihelium signal, both strategies should be able to identify it.

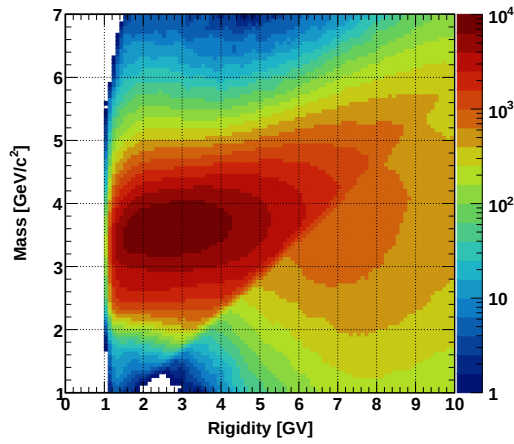


Figure 4.2: ISS data: TOF reconstructed mass as a function of the measured rigidity for the  $R > 0$  sample

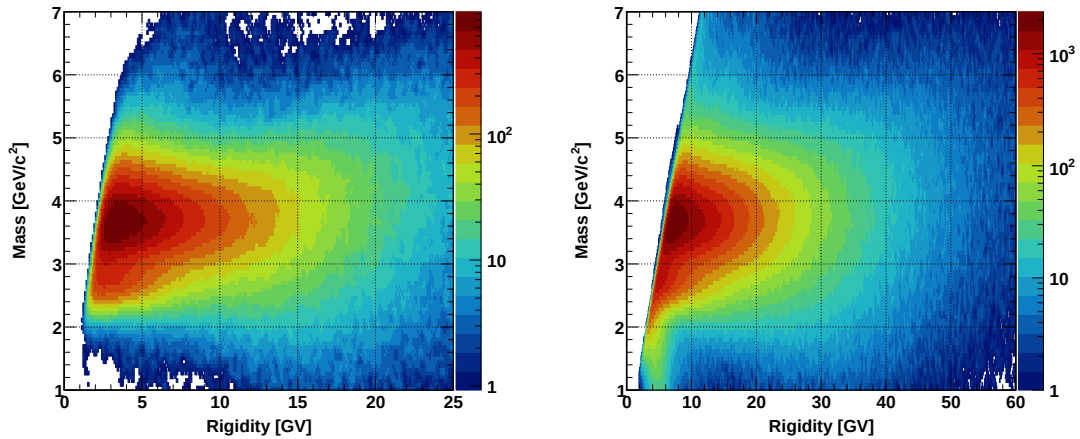


Figure 4.3: ISS data: NaF (left) and AGL (right) reconstructed mass as a function of the measured rigidity for the  $R > 0$  sample.

### 4.1.1 Expected backgrounds

The signal we search for is extremely faint; thus, a meticulous study of the expected backgrounds has been carried out. Considering the composition of different particles and nuclei composing the Cosmic Ray spectra, only species closer to  $He$  nuclei can be sources of background events searching for  $\overline{He}$  events. In particular, the considered species are: protons,  $^4He$  and  $^3He$  nuclei, electrons,  $Li$  and  $C$  nuclei.

Protons are the most abundant particles in CRs, constituting  $\sim 90\%$  of the total, being nine times more abundant than  $He$  nuclei. Despite the large amount of protons in CRs, they constitute a negligible background for  $\overline{He}$  searches, mainly due to the good charge separation provided by the inner tracker, exploiting the proportionality between energy losses and charge; i.e.  $dE/dX \propto Z^2$ . Indeed, the inner tracker charge distribution shown in Figure 3.2 presents a good agreement between data and helium nuclei Monte Carlo, proving a negligible proton contamination.

$He$  nuclei are the second most abundant species in CRs, composing  $\sim 9\%$  of the total and, as will be explained in the next sections, are the main component of the expected background.

Electrons are the most abundant particles in CRs with a negative charge, although they build up less than 1% of the total. In particular, cosmic electrons, due to their low mass value  $m_{e^-} \sim 0.5 \text{ MeV}/c^2$ , are highly relativistic and particles with high  $\beta\gamma$  (see equation 2.3) suffer larger energy losses that might lead to the misidentification of an electron as a  $|Z| = 2$  particle. Thanks to the negative charge sign, the reconstructed track leads to a negative curvature, potentially resulting in a false  $\overline{He}$  candidate. The pollution due to electrons has been studied using MC simulations, and Figure 4.4 shows the electron contamination after the event selection reported in Section 3.1. The plot shows the distribution of the inner tracker charge for simulated electrons. In total,  $\sim 1 \times 10^5$  events survive the event selection, corresponding to  $\sim 550$  weighted events; among them, only 5.7% have a reconstructed charge higher than 1.9. The events have been reweighted following the procedure described in Section 3.2.1. Even if the abundance of electrons is at least one order of magnitude less than  $He$  nuclei, they still must be considered due to the rarity of expected signal events.

Heavier nuclei, such as  $Li$  and  $C$ , are a negligible background due to their low abundances and thanks to the multiple redundant charge measurements performed within the AMS detector (see Figures 3.2 and 3.5). Indeed,  $C$  nuclei are up to 28 times less abundant than  $He$  nuclei, whilst  $Li$  nuclei are even less abundant,  $\sim 200$  times. In conclusion, the main background component, to any  $\overline{He}$  detection, is due to  $He$  nuclei, but  $e^-$  must also be taken into account as a second-order correction.

#### Charged confused $He$ events

To investigate how  $He$  nuclei can be misidentified as  $\overline{He}$ , a MC sample of  $^3He$  and  $^4He$  nuclei has been used. Selecting only events with negative reconstructed rigidities, three different sources of charge confusion have been identified: spillover,  $^4He \rightarrow ^3He$  inelastic and elastic interactions.

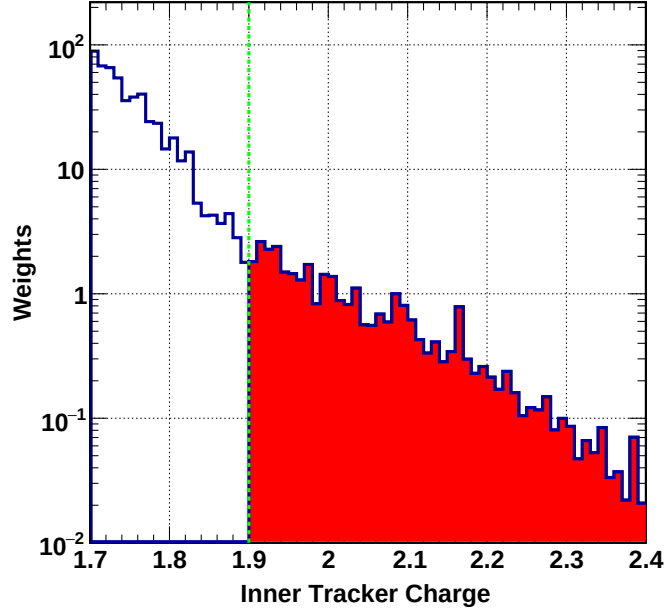


Figure 4.4: Distribution of the inner tracker charge for simulated electrons, after applying each cut of the event selection described in Chapter 3. The events have been reweighted following the procedure described in Section 3.2.1. In total,  $\sim 1 \times 10^5$  unweighted events survive the event selection, corresponding to  $\sim 550$  weighted events; among them, only 5.7% have a reconstructed charge higher than 1.9.

### Spillover

Spillover is a consequence of the finite resolution of the silicon tracker. The tracker measures the inverse of rigidity ( $\frac{1}{R} \propto \text{sagitta}$ ), the mean value of  $\frac{1}{R}$  [GeV $^{-1}$ ] distribution shifts towards zero when the generated rigidities increase, causing the tail of the distribution to spill over negative values. The bulk of the distributions can be approximated as Gaussians, the corresponding width coincides with the tracker resolution and is related to the probability of reconstructing the rigidity with the wrong sign. The plot of Figure 4.5 shows the shift and the widening of the  $\frac{1}{R}$  [GeV $^{-1}$ ] distribution for three different values of generated rigidities, obtained using a combination of  $^3\text{He}$  and  $^4\text{He}$  MC events. Spillover events are correctly reconstructed with a flipped charge sign due to the finite tracker point resolution, resulting in an irreducible background. The spillover contribution grows with rigidity, becoming the dominant one for rigidities close to the MDR or higher, depending on the track pattern considered (see Section 2.1.2), since the rigidity resolution improves with the number of used hits in the track fit. The spillover limits the rigidity range to be used in the analysis; otherwise, the charge-confused background will overwhelm any potential signal. The MDR for the inner track pattern is  $\sim 400$  GV, which is higher than the maximum rigidity usable to obtain a reliable reconstructed mass for helium-like particles, which corresponds to  $\sim 50$  GV.

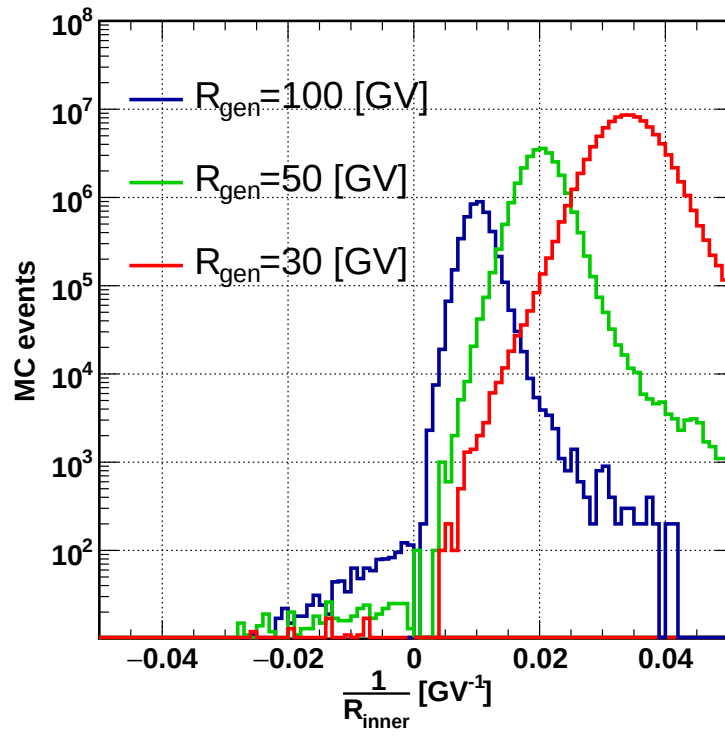


Figure 4.5: Distributions of the measured inverse rigidity ( $1/R \propto \text{sagitta}$ ) for simulated  $^3\text{He}$  and  $^4\text{He}$  events with different generated rigidities. The bulk of the distributions are close to Gaussian distributions, with the width representing the spatial resolution of the silicon tracker. For higher rigidities, the sagitta moves towards zero, and part of the distribution “spills over” the negative values due to the finite tracker resolution.

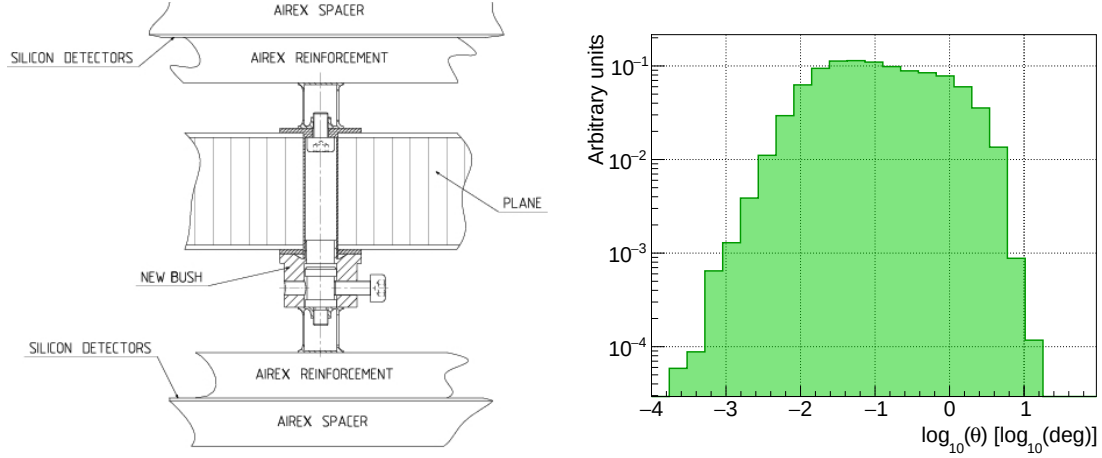


Figure 4.6: Logarithmic value of the primary particle scattering angle due to an elastic interaction. The selected helium events interact elastically within the silicon tracker, resulting in a negative measured rigidity.

### Interactions inside AMS

Inelastic and elastic interactions affect the rigidity measurement if they occur within the considered track pattern, i.e. the inner tracker. Despite the low material budget between the L2-L8 tracker layers, there are some small dense material spots corresponding to aluminium support jig blocks of ladders. The positions of these blocks are located using ISS data by selecting helium events interacting within the inner tracker. A schematic view of the ladder fixing jig, also called tracker foot, is displayed on the left side of Figure 4.6. Helium events going through one or more tracker feet have a higher probability of interacting.

In the  ${}^4\text{He} \rightarrow {}^3\text{He}$  interaction, the primary nucleus loses one neutron and might receive a kink along its path while remaining a  $Z = 2$  particle. The overall effect may lead to reconstructing a negative rigidity, misidentifying a true  $\text{He}$  particle as a  $\overline{\text{He}}$ . Inelastic interactions can be identified by a higher energy deposition in the tracker layers, resulting in a higher reconstructed charge.

Elastic interactions, such as large-angle scattering (i.e. Coulomb scattering), may occur on the supporting structure of the inner tracker and result in a  $\text{He}$  nuclei following an apparently negative curvature and mimicking an  $\overline{\text{He}}$ . These interactions are not easily identified, as the recoiling nucleus might not leave a signal within the silicon sensors. The right plot of Figure 4.6 shows the distribution of the scattering angle for simulated helium nuclei interacting elastically inside the inner tracker. The bulk of the distribution corresponds to small angles, but the tail reaches up to  $\sim 10^\circ$ - $15^\circ$ .

## 4.2 Introduction to Machine Learning

Machine Learning is a wide field with a long history in particle physics, going back to the early 1990s. The topic is divided into learning paradigms, supervised and unsupervised learning. These are not the only existing paradigms, but we will limit ourselves to these since they are the

most relevant ones for this work. Each paradigm contains subcategories called tasks; for example, classification tasks are examples of supervised learning. Over the past decade, this field has experienced significant growth, leading to the development of a diverse range of approaches and techniques. Currently, the Particle Data Group [11] has a dedicated section for Machine Learning within the chapter “Mathematical Tools”. Despite the size of the topic, some common aspects are transversal and will be introduced in the following paragraphs. One of the first paragraphs of the section stands:

Despite the connotations of machine learning and artificial intelligence as a mysterious and radical departure from traditional approaches, we stress that machine learning has a mathematical formulation that is closely tied to statistics, the calculus of variations, approximation theory, and optimal control theory [11].

Let us assume we have some initial values, called input features, represented by a vector  $x \in R^d$ . These vectors can either be Monte Carlo (MC) values or true measurements. Assuming we are using a Monte Carlo simulation, we can build a label  $y$  based on true information. For example, if the task is to distinguish between two particles like protons and electrons, the label is a single number, i.e.  $y = 1$  for protons and  $y = 0$  for electrons [11].

The ordered sequence of input features and labels constitutes a training sample  $\{x_i, y_i\}_{i=1, \dots, n}$ . Given the presence of a label, this is a supervised problem. In a generic classification task, the aim is to find a function  $f : x \rightarrow y$  capable of predicting the label value on each sample. A neural network is essentially a family of parametric functions  $f_\phi : R^d \rightarrow R$ , where  $\phi$  indicates the internal parameter of the network. The best parametric function  $f_\phi$  is obtained through a search called training. During the training, the parameters  $\phi$  are modified and updated to find the best combination. The procedure is carried on through a loss function  $\mathcal{L}(f_\phi(x), y)$ , as for example  $\mathcal{L}_{sq}(f_\phi(x), y) = (y - f_\phi(x))^2$ . Given a loss function, it is possible to compute its average on the training sample of size  $n$ :

$$\mathcal{R}_{emp}(f_\phi) = \frac{1}{n} \sum_{i=1}^n \mathcal{L}(f_\phi(x), y) \quad (4.1)$$

where  $\mathcal{R}_{emp}(f_\phi)$  is called the empirical risk. During the training, the empirical risk is minimised numerically. In literature, with a little abuse of notation, this process is also called minimising the training loss. The minimisation is carried out through a gradient descent, which iteratively adjusts the parameters of the network according to:

$$\phi^{t+1} = \phi^t - \lambda \nabla_\phi \mathcal{R}_{emp}(f_\phi) \quad (4.2)$$

where  $\lambda$  is a parameter called learning rate.

After the minimisation, a specific set of parameters  $\hat{\phi}$  is selected. Given the parameters  $\hat{\phi}$ , the neural network is a fixed function  $f : x \rightarrow y$ , also called a model. The model is tested on an independent dataset called validation or test dataset. Usually, the empirical error on the test dataset

is larger than the empirical error obtained from the training sample. Large differences indicate overfitting, meaning the model does not generalise well to unseen data. The generalisation of a model is measured through its empirical risk. Changes in the network architecture, in the loss function definition and different learning rate values may reduce the generalisation error, achieving a better performance. Once trained, the model can be applied to any sample.

Selecting a threshold on the output of the network, also called the score, an enriched sample with a certain label is selected. Such a threshold is usually referred to as the working point. Given a working point, the efficiency of the selection, the fake rates and purity of the sample can be evaluated (also from MC). Such information is summarised in the receiver operating characteristic curve or ROC curve. The ROC curve summarises the tradeoff between true positive rate (TPR) and false positive rate (FPR). The characterisation of the efficiency and the rejection of a given working point requires labelled data. In particle physics, labelled data are built from MC simulations; however, such simulations are imperfect copies of reality. Discrepancies between the training sample and the real data sample may lead to a domain or distribution shift, leading to a less-than-optimal solution of the minimisation problem. Moreover, mismatches between the simulated data used to evaluate the performance of the algorithm and the unlabeled data generate systematic uncertainties [11].

In this work, two Deep Neural Networks (DNNs) based on supervised learning and two DNNs based on unsupervised learning have been used to suppress the expected background events and enhance any possible signal candidate. A full explanation of the Networks, their structure and how they have been trained is reported in the Sections 4.3 and 4.4. The DNNs used are Feed-Forward Multi-Layer Perceptrons; their functioning and main principles are explained in the next section.

#### 4.2.1 Feed-Forward Multi-Layer Perceptron

The Multi-Layer Perceptron is a specific DNN composed of  $L$  layers such that:

$$f = f^{(L)} \circ \dots \circ f^{(1)} \quad (4.3)$$

the  $l^{th}$  layer is a function  $f^l : \mathcal{R}^{d_{l-1}} \rightarrow \mathcal{R}^{d_l}$  [11]. A unit producing a single scalar value within the  $l^{th}$  layer is called a neuron. For  $l < L$  the  $f^l$  are called hidden layers. The number of neurons, corresponding to the dimension of the codomain, is referred to as the width of the hidden layers. A single layer of an MLP [11] has the following form:

$$f^{(l)}(u) = \sigma^{(l)} \left( W^{(l)}u + b^{(l)} \right), \quad (4.4)$$

where  $W^{(l)} \in \mathcal{R}^{d_l \times d_{l-1}}$  is the weight matrix,  $b^{(l)} \in \mathcal{R}^{d_l}$  is the biases vector,  $u \in \mathcal{R}^{d_{l-1}}$  is the input vector from the previous layer,  $W^{(l)}u$  is the matrix-vector product, and  $\sigma^{(l)}$  is a non-linear activation function applied element-wise. The parameters of a MLP comprise the collection of weights and biases, thus  $\phi = (W^{(1)}, \dots, W^{(L)}, b^{(1)}, \dots, b^{(L)})$ . The activation functions are nonlinear, a

common choice for a DNN, also used in this work, is the Rectified Linear Unit (ReLU), defined as:

$$\sigma_{ReLU}(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.5)$$

The computational cost is low, and the gradient does not vanish for  $x \in (0, +\infty)$  [126]. Another example of an activation function is the sigmoid:

$$\sigma_{sigmoid}(x) = \frac{1}{1 + \exp(-x)} \quad (4.6)$$

This function is bounded between  $(0, 1)$  and is useful at the last layer for classification tasks. MLPs are good approximators, indeed, one of the first approximation theorems in the theory of NNs showed that, even with a single hidden layer ( $L = 2$ ), an MLP can approximate any continuous function using an activation function different from a polynomial and with a sufficiently large depth  $d_1$  [127]. However, it is more convenient to increase the depth of the network, instead of increasing the depth of a single layer. This approach has been proven to be more efficient, reducing the number of parameters required to approximate a given function [128].

#### 4.2.2 Optimisation of the weights and back propagation

The weights of a DNN are updated during training. Given a model  $f(x, \theta)$  and a loss function  $\mathcal{L}(x, \theta)$ , where  $x$  are data and  $\theta$  are the parameters of the model. The optimisation of their values is done iteratively, evaluating the gradient of the loss function [11]. Indeed, following equation 4.2 the weight at the  $t^{th}$  step are:

$$\theta_t = \theta_{t-1} - \lambda \nabla_{\theta} \mathcal{L}(x, \theta) \quad (4.7)$$

Here  $\lambda$  is the learning rate, as introduced in equation 4.2. At the first step, the parameters  $\theta_0$  are initiated randomly with a given seed for reproducibility. Defining  $\delta_{\theta} = \theta_t - \theta_{t-1}$  and considering  $\delta(\nabla_{\theta} \mathcal{L}(x, \theta))$ :

$$\delta(\nabla_{\theta} \mathcal{L}(x, \theta)) \approx \delta_{\theta} \cdot \nabla_{\theta} \mathcal{L}(x, \theta) = -\lambda |\nabla_{\theta} \mathcal{L}(x, \theta)|^2 \quad (4.8)$$

showing how this approach leads to minimising the loss function. The approximation in equation 4.8 holds if the learning rate is low. Instead of computing the exact gradient term, it is possible to approximate it using a subset of the data called batch

$$\nabla_{\theta} E_{\hat{p}(x)} \mathcal{L} \approx \frac{1}{N} \sum_i^N \nabla_{\theta} \mathcal{L}_i, \quad (4.9)$$

where  $\nabla_{\theta} E_{\hat{p}(x)} \mathcal{L}$  is the average number of the gradient, and  $N$  is the number of events in the batch. The batch is a random sample of the dataset. The weight optimisation follows three main steps: define the batches, approximate the gradient on a single batch, update the parameter through



equation 4.7. The batches are not independent, thus a model may overfit. For this reason, it is fundamental to compare the DNN performance on the validation dataset, which is an independent dataset. The approximation in equation 4.9 reduce the computation cost of evaluating the gradient.

Gradient descent is a building block of more advanced optimisation algorithms, improving the convergence rate. An improved approach is called momentum-based optimisation. It is based on a physics perspective of a viscous fluid in an external potential. One updates the current velocity with the force (potential gradient), followed by an update in position based on velocity. Therefore, the approach uses previous gradients along with the current one to compute a running average of the gradient. The average takes into account a forgetting factor, controlling how far back the average goes. Another improvement consists of making the learning rate dependent on the single  $\theta_i$ ; for example, considering the  $(\nabla_{\theta_i} \mathcal{L})^2$ . Evaluating the running average of this quantity, it is possible to reduce the learning rate in the directions with a large average  $(\nabla_{\theta_i} \mathcal{L})^2$ , consequently reducing fluctuations along the same direction. An example of an advanced optimisation algorithm is ADAM (Adaptive Momentum Estimation) [129]. This algorithm combines the momentum and normalisation approach, computing the running averages of both the gradient and the gradient norm squared, with two separate forgetting factors [11, 129].

Concretely, the term  $\nabla_{\theta} \mathcal{L} = \nabla \mathcal{L}(f(x, \theta))$  requires computing the partial derivatives with respect to each parameter  $\theta_i$ . Assuming  $f$  a composite model  $f = f_n(f_{n-1}(\dots, \theta_{n-1}), \theta_n)$ , with each  $f_{i:1,n}$  differentiable the following chain rule holds:

$$\nabla_{\theta_i} \mathcal{L} = \frac{\partial \mathcal{L}(f(x, \theta))}{\partial \theta_i} = \frac{\partial \mathcal{L}}{\partial x_n} \cdot \frac{\partial x_n}{\partial x_{n-1}} \dots \frac{x_i}{\theta_i}, \quad (4.10)$$

where  $x_n$  is the output of the  $n^{th}$  composite function  $f_n$  [11]. The computation requires the gradient at all steps, following the reverse order of the model composition. For this reason, the technique is called back propagation.

In conclusion, in this work, the ADAM optimiser [129], based on the gradient descent approach and the backpropagation, is used during the training of the four DNNs to update the weights and optimise the performances of the models.

### 4.2.3 Supervised learning and classification task

Supervised learning refers to a class of problems where the training dataset is  $\{x_i, y_i\}_{i=1, \dots, n}$  with  $x_i$  the input features and  $y_i$  the labels. It is assumed that  $(x_i, y_i)$  are identically and independently distributed (i.i.d) as  $p(x, y)$ , which is not explicitly known. The loss function can be expressed as follows:  $\mathcal{L}(y, f(x))$ . The task is to predict a finite number of class labels  $y \in \mathcal{Y}$  starting from the input features vector  $x \in \mathcal{X}$ . When the number of classes is equal to two, the task is called binary classification. Generally speaking, a classifier is a model  $g : \mathcal{X} \rightarrow \mathcal{R}^{|\mathcal{Y}|}$ , predicting continuous probabilities, i.e.  $g(x)$ , for each class, followed by an operation that chooses a discrete label  $y \in \mathcal{Y}$ . The latter operation can be a fixed threshold or  $f(x) = \text{argmax} g(x)$ . A standard loss function used

for both binary and multiclass classification tasks is the cross entropy loss:

$$\mathcal{L}_{CE}(y, f(x)) = - \sum_{c \in |\mathcal{Y}|} \mathbf{1}(y = c) \log(f_c(x)) \quad (4.11)$$

where  $\mathbf{1}$  is the indicator function (1 if  $y = c$  and 0 otherwise),  $f_c(x) = \tilde{f}(y = c|x) = p(y = c|x)$ , with  $\tilde{f}(y = c|x)$  an approximation of the true probability  $p(y|x)$ , and  $f : \mathcal{X} \rightarrow \mathcal{R}^{|\mathcal{Y}|}$ . The indicator function picks out the term in the sum for the corresponding class label  $y$ . It is worth noting that:  $f_c(x) \geq 0$  and  $\sum_c f_c(x) = 1$ . Equation 4.11 for a binary classification task simplifies to the binary cross entropy equation:

$$\mathcal{L}_{BCE}(y, f(x)) = - \sum_{i=0, \dots, N} y_i \cdot \log(f(x_i)) + (1 - y_i) \cdot \log(1 - f(x_i)) \quad (4.12)$$

where  $i$  is running on the number of events contained in a subset of the sample called batch.

#### 4.2.4 Unsupervised learning and anomaly detection

Unsupervised learning refers to unlabelled training dataset,  $x_{i=0, \dots, n}$  where  $x_i \in \mathcal{X}$  are the input features. It is assumed that  $(x_i)$  are identically and independently distributed (i.i.d) as  $p(x)$ , which is not explicitly known. The loss function can be expressed as  $\mathcal{L}(x, f(x))$ . Unsupervised learning can be used to find a low-dimensional summary statistic containing the most relevant information from the data. Such dimensionality reduction can either be lossless or not. Lossless compression allows for maintaining all the original information in a smaller dimensionality, since the original data can always be recovered starting from this smaller space. On the other hand, an example of data representation with loss of information is likelihood ratios used for hypothesis testing.

Autoencoders find a nonlinear transformation for data dimensionality reduction:

$$f = g \circ e : \mathcal{X} \rightarrow \mathcal{X} \quad (4.13)$$

where  $e : \mathcal{X} \rightarrow \mathcal{Z}$  is the encoder and  $g : \mathcal{Z} \rightarrow \mathcal{X}$  is the decoder. The dimensionality of  $\mathcal{Z}$  is less than  $\mathcal{X}$ , and  $z = e(x)$  is a compressed representation of the input. The intermediate space  $\mathcal{Z}$  is called latent space and its dimension is an important hyperparameter that must be tuned. Hypothetically, given an initial dimensionality  $d$  for a large enough latent space dimension  $k$  and enough parameters in both the encoder and decoder architecture, the data reduction can be lossless. On the other hand, if  $k$  is too small or the model  $e$  and  $g$  are not flexible enough, reconstruction error will occur during the process. Therefore, the reconstruction error defined in the following equation is a natural choice as a loss function for an autoencoder:

$$\mathcal{L}_{MSE}(x, f(x)) = \frac{(x - f(x))^2}{N_f}, \quad (4.14)$$

where  $x$  is the input features vector,  $f(x)$  is the output of the autoencoder, and  $N_F$  is the total

number of input features.

The structure of an autoencoder is useful for detecting anomalies in an unlabelled dataset. Generally, anomalies are out-of-distribution events (OOD) that can be identified thanks to their different reconstruction error.

## 4.3 Supervised Classifier

### 4.3.1 Task description and label definition

Supervised learning relies on the definition of a label, a variable that separates the training sample into two or more classes. Sometimes, the label is obtained directly from MC true information, even though such a choice is not mandatory. Indeed, data-driven approaches rely on high-level variables, thus measured or reconstructed variables. In this work, the supervised classifier label is based on a measured variable, defined as:

$$l_{CL} = \frac{R_{UH} - R_{LH}}{R_{UH} + R_{LH}}, \quad (4.15)$$

where  $R_{UH}$  is the measured rigidity using the upper-half part of the inner tracker, and  $R_{LH}$  is the rigidity measured with the lower-half part. The inner tracker has seven silicon layers,  $L2-L8$ .  $R_{UH}$  is obtained from the information provided by the first five layers  $L2-L6$ ; whilst  $R_{LH}$  relies on the last six layers  $L3-L8$ . The distribution is expected to be asymmetrical, because  $R_{UH}$  has four layers in the magnetic field, while  $R_{LH}$  has 7 layers, resulting in a better resolution for the curvature measurement. The difference over the sum of these two rigidities returns an estimate of the compatibility between the two measurements. The variable is constrained between  $[-1, 1]$ , well-reconstructed rigidities cluster close to 0, whereas bad-reconstructed rigidities scatter towards higher values, i.e.  $\left| \frac{R_{UH} - R_{LH}}{R_{UH} + R_{LH}} \right| > 0.5$ .

The figures 4.7 and 4.8 show the  $l_{CL}$  as a function of the logarithmic values of the reconstructed rigidity with the inner-only span. The left plots show positive rigidities, which are prescaled by a factor of 100, while the plot on the right show negatives rigidities. For  $R_{INNER} > 0$ , the distribution is well confined between  $[-0.2, 0.2]$ , with a selection efficiency higher than 90% for rigidities up to  $R \sim 15$  GV. The distributions is broader for  $R_{INNER} < 3$  GV and for  $R_{INNER} > 50$  GV. In these intervals, the rigidity resolution is getting worse due to multiple scattering at low  $R$ , and due to the reduction of the overall curvature at higher  $R$ .

The compatibility between two rigidities is sensitive to spillover events. For  $R > 100$  GV, the bidimensional plot is quite uniform, since for rigidities closer to the MDR, the spillover events become more abundant. Since spillover events are random statistical fluctuations, the uniformity of the distribution reflects the equal probability of observing opposite or coordinated fluctuations. For spillover events, the compatibility is close to 0 only if the statistical fluctuations are synchronous, thus choosing a bandwidth of 0.4 around 0, since the label is constrained within  $[-1, 1]$ , only the

20% of the total events are selected. In this work, well-reconstructed rigidities are defined as:

$$l_{CL} = \frac{R_{UH} - R_{LH}}{R_{UH} + R_{LH}} \in [-0.2, 0.2], \quad (4.16)$$

whilst bad reconstructed rigidities are defined as:

$$|l_{CL}| = \left| \frac{R_{UH} - R_{LH}}{R_{UH} + R_{LH}} \right| > 0.2 \quad (4.17)$$

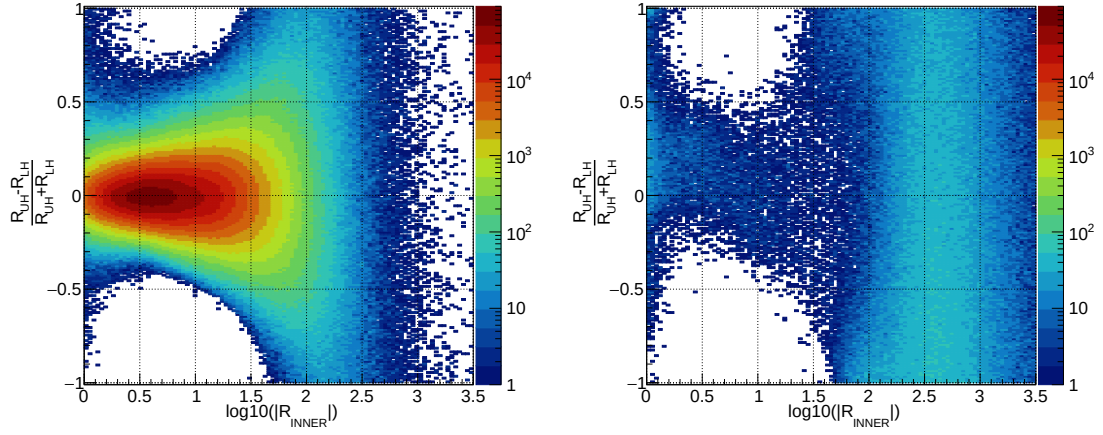


Figure 4.7: The rigidity dependence of the classifier label for AMS-02 data, positive rigidities are on the left, negative rigidities are on the right.

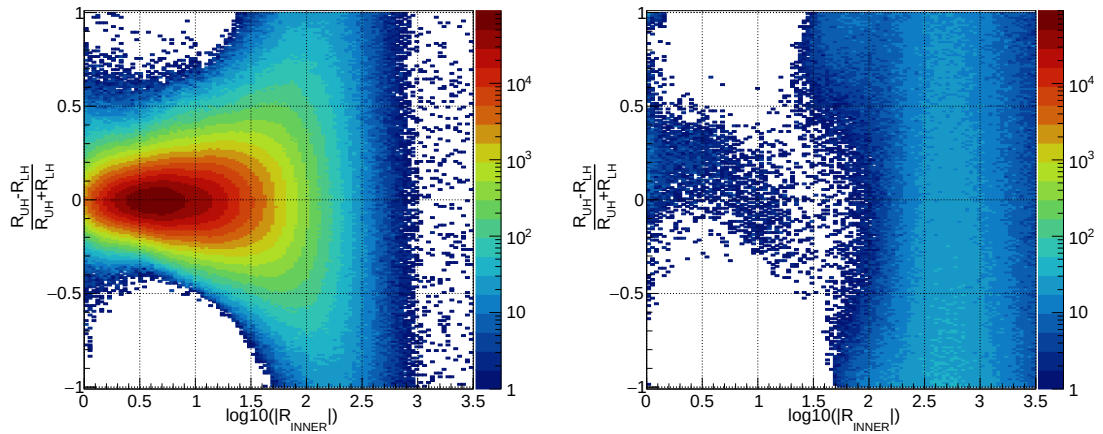


Figure 4.8: The rigidity dependence of the classifier label for  $^4\text{He}$  and  $^3\text{He}$  MC simulation, positive rigidities are on the left, negative rigidities are on the right.

The definitions reported in equations 4.16 and 4.17 are rigidity-dependent. The dependency is plotted in Figure 4.9, where the efficiency of the selection on  $\left| \frac{R_{UH} - R_{LH}}{R_{UH} + R_{LH}} \right| < 0.2$ , i.e. the ratio between the number of surviving events after the cut and the total number of events in a rigidity

bin, is shown as a function of the reconstructed rigidity. For positive rigidities, the efficiency is evaluated every bin of  $\log_{10}(R_{INNER})$ , for negative rigidities, the efficiency is integrated on a 5 bin interval due to the different statistic. The efficiency, considering the positive rigidities sample, is above 90% for  $R < 15$  GV, then it smoothly decreases to 20% for  $R > 300$  GV. Looking at the negative rigidities sample, the efficiency rises to 70% at  $|R| \sim 5$  GV, then it drops to 10% at  $|R| \sim 60$  GV, afterwards it rises to 20% for  $|R| > 300$  GV.

The task of the supervised classifier is to distinguish between the two classes defined by the label definition specified by equation 4.16 and equation 4.17, i.e. well and bad reconstructed rigidities. To achieve the goal, the classifier is trained only on the positive rigidities samples. The chosen sample is composed of helium nuclei, and it is dominated by well-reconstructed rigidities, as visible in Figure 4.9, even though the class of badly reconstructed tracks (mainly spillover events) becomes gradually more relevant for rigidities higher than 30 GV. Training a DNN on the positive rigidity samples (one for ISS data, and one for  $^3\text{He}$  and  $^4\text{He}$  MC) provides a classifier capable of recognising spillover from well-reconstructed events. After the training, the algorithm can be applied to the negative rigidities sample and used to reduce the expected charged-confused background described in Section 4.1.

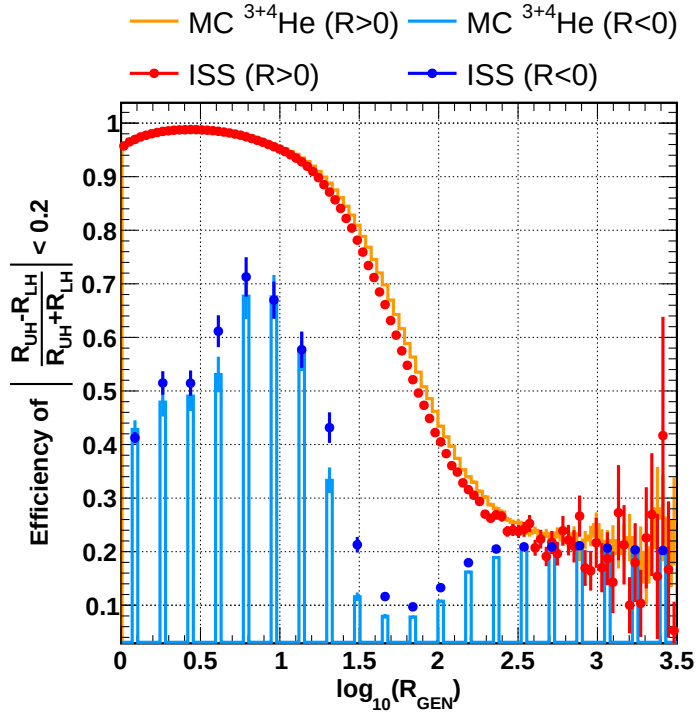


Figure 4.9: The rigidity dependence of the classifier label for data and MC simulation ( $^3\text{He}$  and  $^4\text{He}$ ), divided into positive and negative reconstructed rigidities.

### 4.3.2 Input variables studies

The chosen label identifies two classes: the well and the bad reconstructed rigidities. It is crucial to select the classifier input variables to allow for the DNN to distinguish among them. Such variables, also called features, must provide information on the classes defined by the label. The feature distributions for well and poorly reconstructed datasets must be as different as possible, or in other words, the variable must possess a good separation power. To quantify the separation power of different variables, the CatBoost classifier algorithm has been used [130].

CatBoost, standing for Categorical Boosting, is a ML algorithm that implements gradient boosting on decision trees. The CatBoost classifier has been trained on the same sample as the supervised classifier, i.e. the events with positive charge. The dataset is divided into  $k = 4$  groups, also named folds, and the classification model is trained on  $k - 1$  folds and validated on the remaining one. During the training, the CatBoost classifier minimises a weighted loss function defined as follows:

$$\mathcal{L} = \sum_i w_i \cdot l(y_i, p_i) \quad (4.18)$$

where the sum runs on all the events in a fold,  $w_i$  is the Monte Carlo weight of the  $i$ -th event computed following the procedure described in Section 3.2.1, and  $l(y_i, p_i)$  is the log-loss defined as:

$$l(y_i, p_i) = -(y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \quad (4.19)$$

To avoid overfitting, the training loop is blocked if the validation loss does not improve for 20 epochs. At the end of the training procedure, the model returns the importance of each input variable. The importance quantifies how much each variables contribute to the prediction power of the model, thus reflecting complex non-linear relationships. In the end, all the importances are aggregated across the fold, and the process is repeated  $k$  times, changing the validation fold. Comparing the importance, it is possible to rank the features from the most relevant to the least; for this reason, in the literature, the procedure is called feature ranking.

In total, the importance of 58 reconstructed quantities, provided from different subdetectors as TRK, TOF, ACC and RICH, has been quantified. The CatBoost algorithm has been tested both on data and on MC independently. The results are shown in figure 4.10 for MC, and in figure 4.11 for ISS data. Both plots show the importance of the features on the horizontal axis, while the names of each variable are on the vertical axis. To validate the result, a random feature has been injected into the sample; since the CatBoost algorithm is not able to extract any information from such a variable to fulfil the task, the corresponding importance should be low. For MC, in Figure 4.10, the random feature has an importance smaller than  $10^{-6}$ , which is compatible with the value obtained from ISS data displayed in Figure 4.11.

After the features ranking, 23 features are used as input variables to the two supervised classifiers  $CL_{MC}$  and  $CL_{ISS}$  and the full list is reported in the table 4.1. The plots of each input feature, divided into the two different classes, are available in Appendix A.1.

| Detector                   | Variable                                                |
|----------------------------|---------------------------------------------------------|
| TRK                        | $L_2$ Y residual                                        |
|                            | $L_3$ Y and X residuals                                 |
|                            | $L_4$ Y residual                                        |
|                            | $L_5$ Y and X residuals                                 |
|                            | $L_6$ Y and X residuals                                 |
|                            | $L_7$ Y residual                                        |
|                            | Number of X and Y hit inside inner                      |
|                            | $\frac{R_{INNER}-R_{INNER-L1}}{R_{INNER}+R_{INNER-L1}}$ |
|                            | MIN and MAX Z on $L_2$                                  |
|                            | MIN and MAX Z on $L_3-L_4$                              |
|                            | MIN and MAX Z on $L_5-L_6$                              |
| MIN and MAX Z on $L_7-L_8$ |                                                         |
| RICH                       | Measured PE                                             |
|                            | Z                                                       |
|                            | Number of PMTs in the ring                              |

Table 4.1: The 23 input features for both the supervised classifiers  $CL_{ISS}$  and  $CL_{MC}$ .

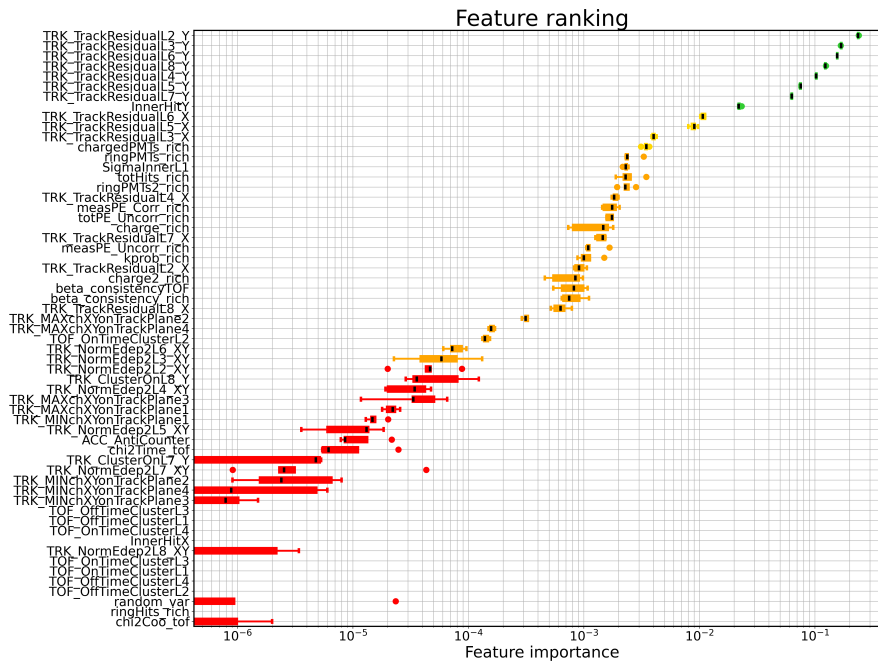


Figure 4.10: Features ranking for MC sample with positive rigidities.

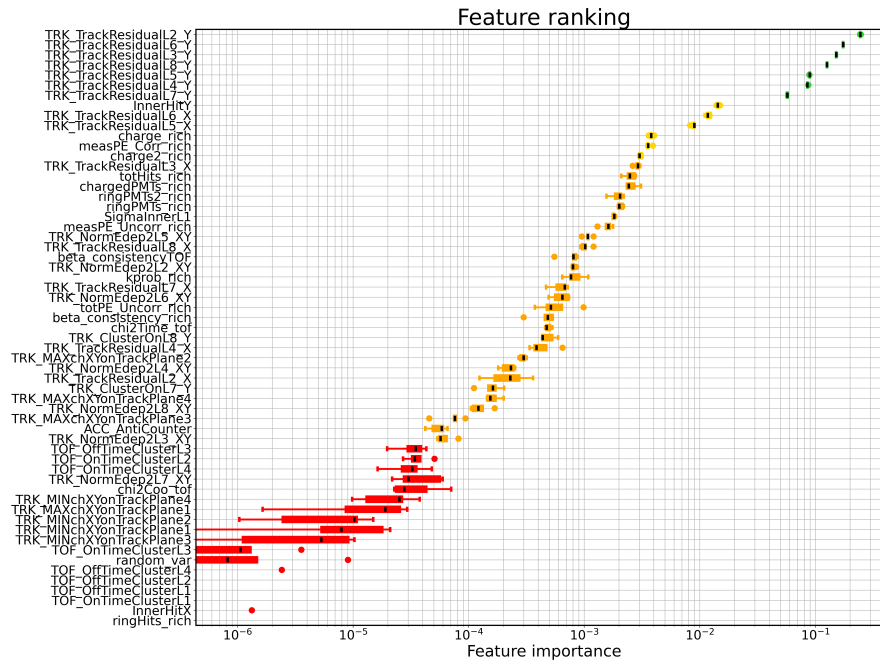


Figure 4.11: Features ranking for ISS-data sample with positive rigidities.



### 4.3.3 Dataset preprocessing, training and validation samples

As previously explained in section 4.2.3, the input data are divided into two parts: the training dataset and the validation dataset. In this work, 70% of the overall available statistics are used as the training dataset, while the remaining 30% is used as the validation dataset. These values are common in the current literature as described here [131, 132, 133].

Following a data-driven approach, two different supervised classifiers have been trained:  $CL_{ISS}$  and  $CL_{MC}$ . The first is trained on AMS-02 data, whilst the latter is trained using a combination of  $^3\text{He}$  and  $^4\text{He}$  MC events. The model trained on MC is not applied to ISS data and vice versa. As discussed in Section 4.1, both samples are trained using only positive rigidities ( $Z = +2$ ) and afterwards applied to the negative rigidities samples.

The dataset must be handled carefully before training through data preprocessing. The performance of the ML algorithms depends strongly on how the data are handled before the training. The weights of a DNN are updated according to the gradient estimation of the chosen loss function, and input features with different scales may affect the gradient calculation, modifying the speed of convergence and the stability throughout the training. The standardisation of the dataset ensures that all the features contribute equally to the loss function, resulting in a smoother multi-dimensional surface, improving the optimisation, stability and generalisation during the training. Without a scale transformation, large absolute values may dominate the model response even if they are not the most important features.

The list of variables reported in table 4.1 are not always available for each event. Indeed, for example, the track pattern can be different, i.e. the number of hits used to reconstruct the track and their locations change; the RICH variables may not be present if any particle travels across the subdetector. These missing values must be treated thoroughly, with an approach commonly applied in Machine Learning literature called masking [134, 135, 136].

Masking missing information prevents the DNN from using non-physical information during the training process. DNNs have a fixed structure; therefore, every event is associated with an input vector with 23 values. When an input value is not available, an arbitrary value is used as a placeholder to maintain the right dimensionality; for this analysis, the chosen value is  $-9999$ . This approach allows for training a single model for different event geometries, reducing the overall complexity without loss of generalisation. The variables reported in table 4.1 have different units of measure, meaning different scales of values. It is important to transform all these variables to the same order of magnitude; otherwise, a Feed Forward Multilayer Perceptron may introduce biases due to the different sizes of each node output. The applied transformation is:

$$\tilde{x}_i = \frac{x_i - \mu_{x_i}}{\sigma_{x_i}}, \quad (4.20)$$

where  $\tilde{x}_i$  is the transformed  $i$ -th value,  $x_i$  is the real input value,  $\mu_{x_i}$  and  $\sigma_{x_i}$  are the mean and standard deviation of the  $i$ -th input feature. The transformation reported in equation 4.20 transforms each input feature into a normally distributed variable, with 0 mean and  $\sigma = 1$ . The

mean  $\mu_{x_i}$  and the standard deviation  $\sigma_{x_i}$  are evaluated only from the training dataset. In this way, no information from the validation dataset is used in the training.

Using the masked array available in the numpy library, the masked values do not affect either the mean or the standard deviation calculation [137]. Moreover, the masked values are left unchanged by the transformation, avoiding unphysical biases in the estimation of the distribution parameters and also to recognise masked values during the training.

The computed means and sigmas are then stored and used as a benchmark to apply the same transformation in equation 4.20 on the validation dataset and on all the future datasets the model will be applied to. This procedure has been validated in numerous available works, as in [131, 132, 133, 134, 135, 136].

#### 4.3.4 Network structure and performance

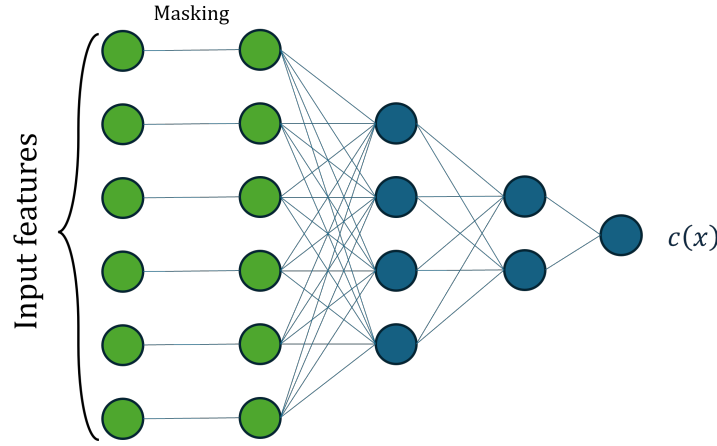


Figure 4.12: Pictorial representation of the two supervised classifiers developed for this analysis:  $CL_{ISS}$  and  $CL_{MC}$ . Both models have a first masking layer that prevents the use of non-physical information during the training.

The pictorial representation of the supervised classifier structure in Figure 4.12 shows that the starting dimensionality of the DNNs is equal to the number of input features, and then it shrinks to a single number, which is the output of the classifier. This scalar is a number between 0 and 1, corresponding to the probability that the input event belongs to the well-reconstructed or to the badly reconstructed rigidities. The exact number of layers, nodes, and the value of the hyperparameters are presented in Appendix B.1.

The first layer<sup>1</sup> of each supervised classifier treats the input feature values differently, depending on whether they are masked values or not. The input vector is multiplied by a mask vector, where the latter assumes only two values: 0 or 1. The null values are assigned in correspondence with the masked values, while the unity values leave the corresponding features unaffected. After the masking, the result is a vector with a dimension equal to the number of input features. This vector undergoes a linear transformation as described in equation 4.4, where each element is multiplied

<sup>1</sup>The description of the main principles of DNN, i.e. a feed-forward multilayer perceptron, has been reported in section 4.2.1.

by the weight matrix, then added to elements of the bias vector, and finally modified under the action of the activation function. The result of the linear transformation is still a vector with the dimension equal to the number of features. The resulting vector is multiplied by the same mask vector used for the input vector, setting to 0 all the elements obtained starting from missing, i.e. non-physical, information. In the end, this procedure prevents any leakage of information due to the missing values into the second layer of the NN.

The two classifiers ( $CL_{ISS}$  and  $CL_{MC}$ ) have the same number of input features and the same structure; the full description of the hyperparameters of the two supervised classifiers is reported in the Appendix B.1.

The classifier trained on MC ( $CL_{MC}$ ) has been trained for a total of 100 epochs, using as a loss function a weighted binary cross-entropy described by equation 4.12. The mean values on each batch for the training and validation loss are shown in the Figure 4.14. Initially, both losses decrease, so the model learns to distinguish between the two classes defined by the label in the equations 4.16 and 4.17. After the  $\sim 50^{th}$  epoch, the two losses rise towards higher values, showing a small divergence around the  $90^{th}$  epoch. This behaviour shows that the gradient descent of the loss function finds a first minimum, around the  $50^{th}$  epoch, then it shifts towards other local minima. The model is saved at the best epoch, i.e. at the minimum value of the validation loss, which corresponds to the best minimum found during the training, so that the rise of the loss functions is not particularly relevant.

A common metric to evaluate the performance of a classifier is the accuracy, defined as:

$$Accuracy = \frac{TP + TN}{TOT}, \quad (4.21)$$

where  $TP$  is the number of true positives,  $TN$  it the number of true negatives, and  $TOT$  is the total number of events. The accuracy mean value on each batch as a function of the training epochs, evaluated on the validation sample, is shown in the left plot of figure 4.14. The accuracy, after an initial rise, remains stable throughout the training, reaching values close to 97%. The ROC curve<sup>2</sup> for the  $CL_{MC}$  classifier is shown in the right plot of figure 4.14, and the Area Under the Curve (AUC) is 0.9946.

Analogously, Figure 4.16 and Figure 4.15 summarises the performance of  $CL_{ISS}$ . The  $CL_{ISS}$  model has been trained for 150 epochs, using as a loss function a binary cross-entropy described by equation 4.12. The mean accuracy value, both for the training and the validation loss, is shown in the left plot of Figure 4.15. The overall behaviour of the function is similar, even though the validation loss is slightly higher than the training loss. The two functions fall during the training, hence the model is correctly learning the task. It is possible to exclude any relevant overfitting effect during the training, since there is no rise in either loss. The mean accuracy as a function of the training epoch evaluated on the validation dataset is shown in the left plot of Figure 4.16. The overall behaviour is similar to  $CL_{MC}$ , with a stable accuracy throughout the training procedure

---

<sup>2</sup>The meaning of such a curve is reported in section 4.2

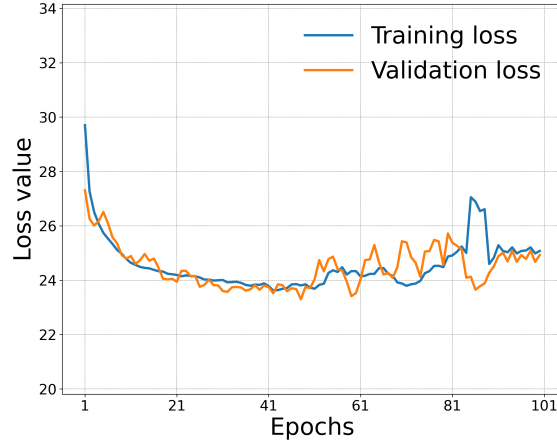


Figure 4.13: Values of the training and validation loss functions of  $CL_{MC}$ , which are weighted BCE loss (see equation 4.12).

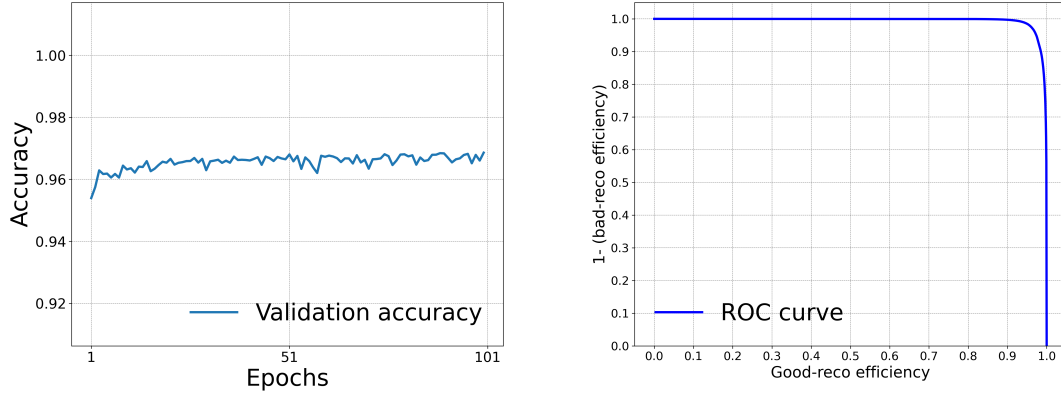


Figure 4.14: (Left) Accuracy of  $CL_{MC}$  as a function of the training epochs. ROC curve of the Monte Carlo classifier, with an AUC = 0.9946.

with values close to 98%. The ROC curve, shown in the right plot of figure 4.16, has an AUC of 0.9965, similar to the performance achieved by  $CL_{MC}$ .

The outputs of the two classifiers, also called scores, are rigidity dependent; this is a direct consequence of the chosen input features (Table 4.1) and the label definition (see equations 4.16 and 4.17). As a consequence, the outputs have been plotted in rigidity bins, and some of them, showing events with positive rigidities from the validation samples, are shown from Figure 4.17 to Figure 4.19. The remaining plots are presented in Appendix B.1. The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

The scores, in addition to the division into rigidity bins, have also been divided depending on which velocity measurements are available for each event. Given the event selection described in Chapter 3, all the events have a  $\beta_{TOF}$  but around the  $\sim 90\%$  crossing the RICH go through the AGL region.

The Figure 4.17 shows the distribution of the  $CL_{MC}$  (left plot) and  $CL_{ISS}$  (right plot) score for events traversing the TOF and with a reconstructed rigidity between  $[1.92, 4.02]$  GV. Similarly,

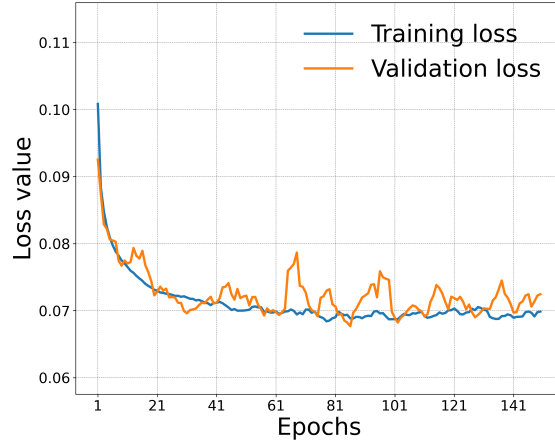


Figure 4.15: Values of the training and validation loss functions of  $CL_{ISS}$ , which are BCE loss (see equation 4.12).

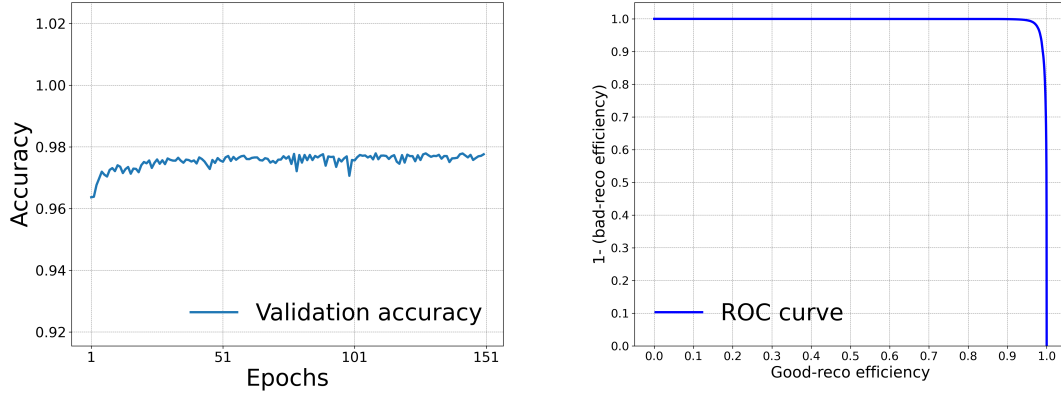


Figure 4.16: (Left) Accuracy of  $CL_{ISS}$  as a function of the training epochs. ROC curve of the data classifier, with an AUC = 0.9965

Figure 4.18 shows the score for events crossing the RICH-NaF with a reconstructed rigidity between  $[4.02, 6.47]$  GV, whilst Figure 4.19 presents the score for events going through the RICH-AGL having  $R_{inner} \in [6.47, 9.26]$  GV. All the events displayed have a positive reconstructed rigidity. For each classifier, in all three cases, two clear peaks are visible, one close to 0 for events belonging to the poorly reconstructed rigidities, and the other close to 1 for well-reconstructed events.

The Figures 4.20, 4.21, and 4.22 present the scores obtained applying the two classifiers to the negative rigidity samples. The  $CL_{ISS}$  algorithm was used on the ISS data  $Z = -2$  sample, whilst  $CL_{MC}$  was used on the MC sample of charged-confused helium nuclei. In this case, the distributions are discrete and less continuous with respect to the positive rigidities samples. Figure 4.20 shows the distribution of the  $CL_{MC}$  (left plot) and  $CL_{ISS}$  (right plot) score for events traversing the TOF and with a negative reconstructed rigidity between  $[-1.92, -4.02]$  GV. Analogously, Figure 4.21 shows the score for events crossing the RICH-NaF with a negative reconstructed rigidity between  $[-4.02, -6.47]$  GV, whilst Figure 4.22 presents the score for events going through the RICH-AGL having  $R_{inner} \in [-6.47, -9.26]$  GV. In each plot, almost the entire distribution is in-

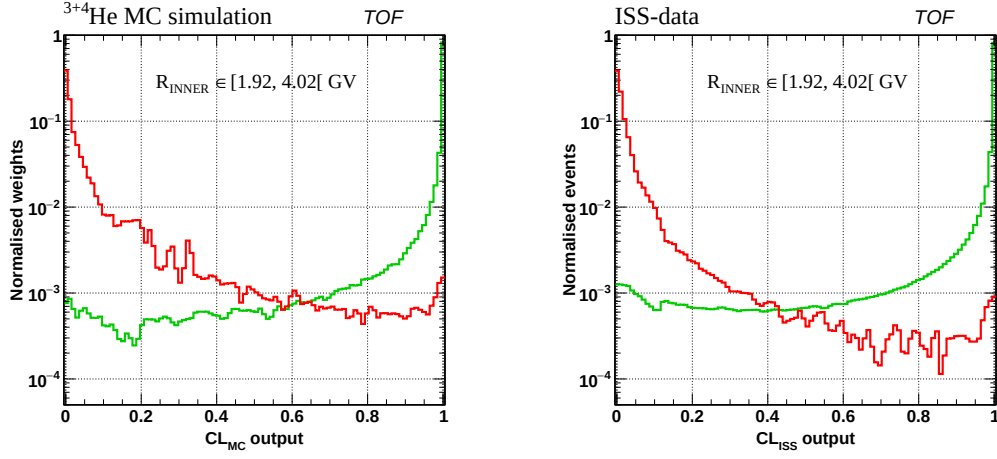


Figure 4.17: Classifier scores for events with a positive reconstructed rigidity between  $[1.92, 4.02]$  GV and a reconstructed  $\beta_{\text{TOF}}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

cluded within the first bins close to 0, meaning that both  $CL_{\text{ISS}}$  and  $CL_{\text{MC}}$  recognise the majority of the  $Z = -2$  events as poorly reconstructed events. Interestingly, both for MC and AMS-02 data, few events have a score close to 1, i.e. the algorithms recognise them as events with a well-reconstructed rigidity.

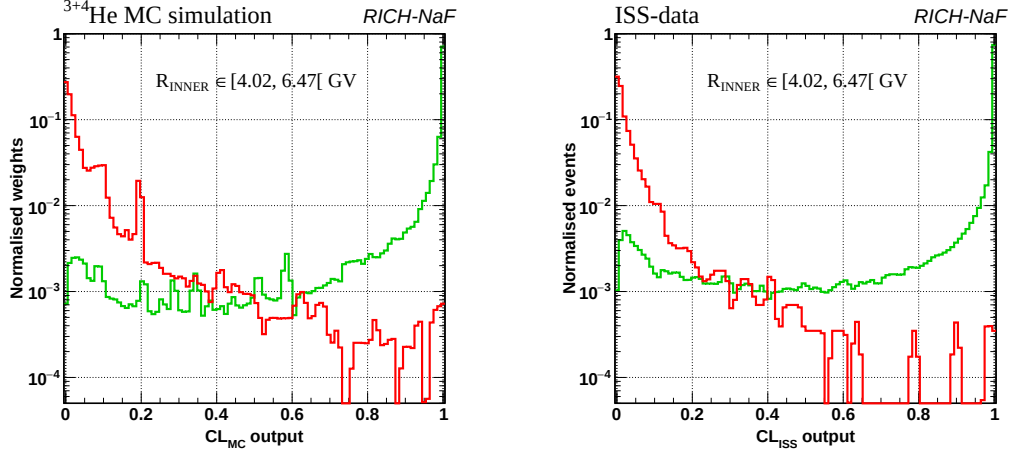


Figure 4.18: Classifier scores for events with a positive reconstructed rigidity between  $[4.02, 6.47]$  GV and a reconstructed  $\beta_{NaF}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

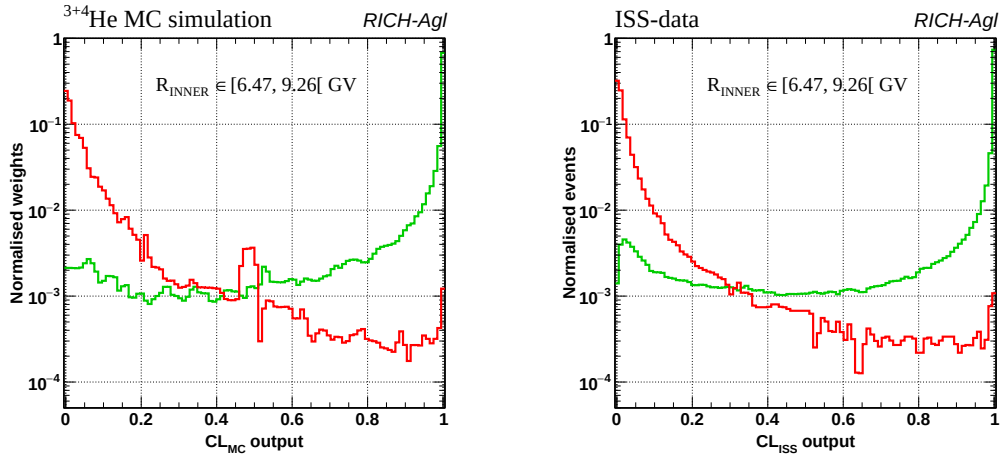


Figure 4.19: Classifier scores for events with a positive reconstructed rigidity between  $[6.47, 9.26]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

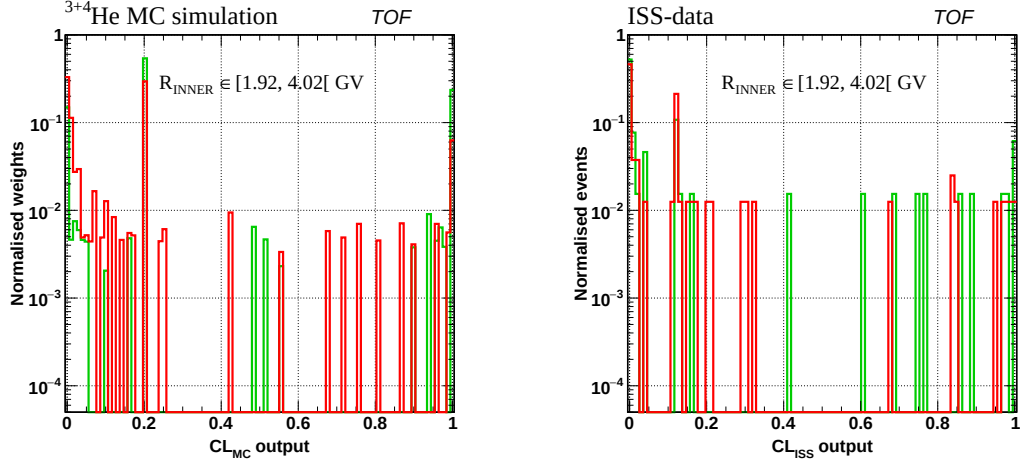


Figure 4.20: Classifier scores for events with a negative reconstructed rigidity between  $[-4.02, -1.92]$  GV and a reconstructed  $\beta_{\text{TOF}}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

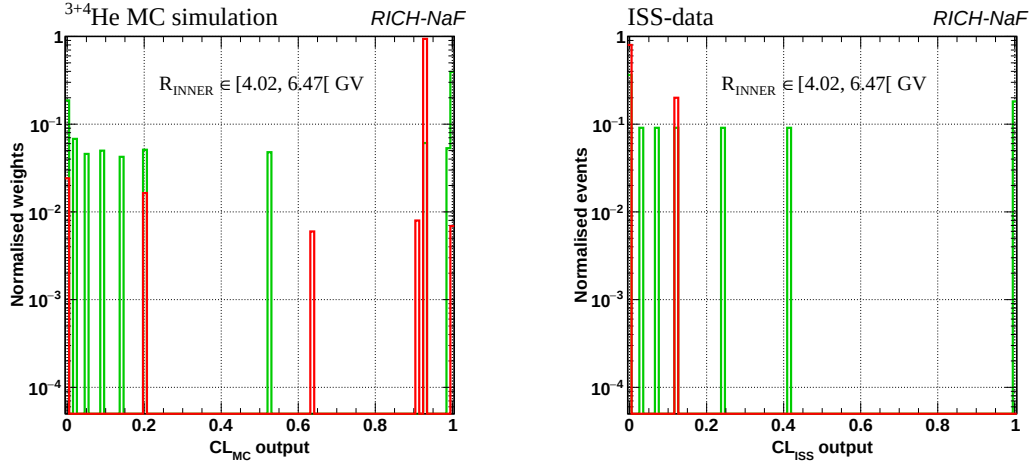


Figure 4.21: Classifier scores for events with a negative reconstructed rigidity between  $[-6.47, -4.02]$  GV and a reconstructed  $\beta_{\text{NaF}}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.



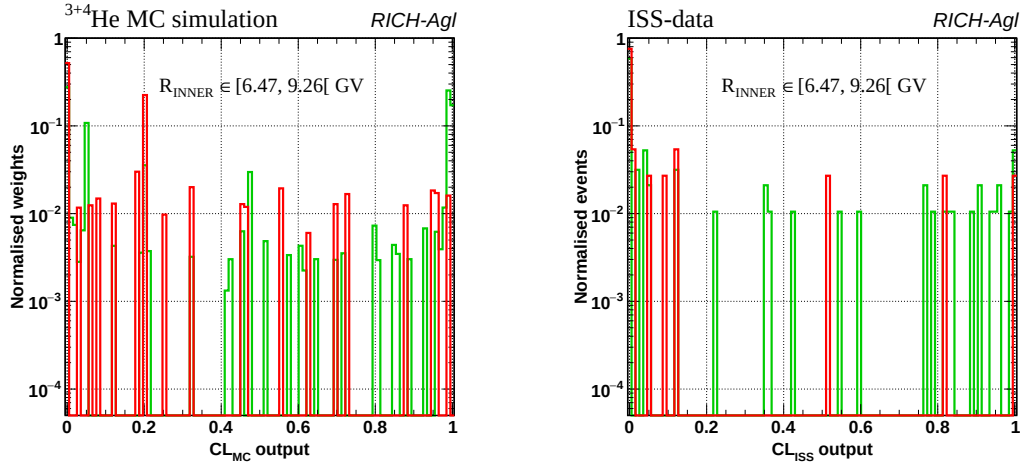


Figure 4.22: Classifier scores for events with a negative reconstructed rigidity between  $[-9.26, -6.47]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

## 4.4 Unsupervised Autoencoder

### 4.4.1 Task description

This approach relies on an unlabelled dataset, and during the training, through the minimisation of a mean square error loss function (see equation 4.14), the NN learns a significant representation of the data with a lower dimensionality. In this work, unsupervised autoencoders are used for an anomaly detection task; where anomalies are defined as out-of-distribution events (OOD), i.e. unusual events within a certain sample or population of events.

The two autoencoders  $AE_{ISS}$  and  $AE_{MC}$  developed for this analysis are trained and validated, as for  $CL_{ISS}$  and  $CL_{MC}$ , on positive rigidity samples. Well-reconstructed  $He$  nuclei events dominate these samples, as discussed in Section 4.1; thus, the models learn a significant representation of these events with a lower dimensionality.

After the training, the models are applied to the negative rigidity samples, which are dominated by background events (Section 4.1). These background events have an average poor quality, resulting in very different distributions compared with well-reconstructed events. Since the domains of the validation and test samples are different, the algorithms will fail on the majority of the events with a negative reconstructed rigidity. The repeated failures cause a different distribution of the reconstruction error, compared with the validation ( $Z = +2$ ) sample. Following this logic, the anomalous events within the test sample ( $Z = -2$ ) are those having a small reconstruction error and residing in the tail of the corresponding distribution. Indeed, such events will have a reconstruction error similar to well-reconstructed events, considerably smaller compared to the bulk of the background-dominated sample.

Instead of directly comparing the reconstruction error, the following transformation is applied to obtain the anomaly score ( $AS$ ):

$$AS = \frac{1}{1 + \exp^{-MSE}}, \quad MSE = \frac{(X - Y)^2}{N_F}, \quad (4.22)$$

where  $X$  is the input vector,  $Y$  is the output of one autoencoder,  $N_F$  is the number of input features,  $MSE$  is the mean square error, and the  $AS$  is the sigmoid function evaluated on the  $MSE$ , also called anomaly score. The anomaly score is always comprised between 0.5 and 1, since the  $MSE$  is positive defined. The lower the  $MSE$ , the closer to 0.5 is the anomaly score. Applying the autoencoders on well-reconstructed events should return an anomaly score close to 0.5, whilst poorly reconstructed events will have a much larger  $MSE$  value, meaning an average anomaly score close to or equal to 1.

Only the events within the test sample ( $R < 0$ ) having an anomaly score in the same interval of the validation ( $R > 0$ ) sample are considered anomalous events within the scope of this analysis. Specifically, we define an anomalous event, a  $Z = -2$ ,  $R_{INNER} < 0$  event with an anomaly score similar to the majority of  $Z = +2$ ,  $R_{INNER} > 0$  events. Where, similar stands for contained within a certain interval corresponding to a certain quantile of the anomaly score distribution of

well-reconstructed events.

#### 4.4.2 Input variables studies

Ranking the input features requires a labelled dataset; it is not straightforward to replicate the same study discussed in Section 4.3.2 within the unsupervised paradigm. As previously discussed, the task fulfilled by the autoencoders can be divided into two main points:

- minimise the median value of the anomaly score distribution on the training and validation datasets.
- locate anomalous events within the negative rigidities sample.

In a naive way, this task can be translated into a classification task, where the autoencoders have to distinguish between the  $R > 0$  and  $R < 0$  classes. In this way, it is possible to evaluate a ranking of the input feature similarly to what has been done for the supervised classifiers  $CL_{ISS}$  and  $CL_{MC}$ .

The Catboost classifier has been trained using the loss function defined in equation 4.18, with a sample containing both positive and negative rigidities. Applying the same procedure described in Section 4.3.2, the importances of 58 reconstructed physical quantities have been obtained. The variables with the highest importance correspond to the features presenting the most separation between the two classes. Indeed, selecting the most relevant features indicated by the CatBoost classifier [130], the two models learn to reconstruct the features with the highest discrepancies between positive and negative rigidities. This approach aims to increase the rejection power of the autoencoders with respect to the expected background sources of the negative rigidity samples.

The results obtained from the MC are shown in figure 4.23, whilst the results from ISS-data are shown in figure 4.24. To validate the results, a random variable has been injected into the sample: since the Catboost classifier is not able to extract any information from this variable, its importance should be low. The results obtained for the two Catboost classifiers, one on ISS data and the other on MC, are similar, yielding an importance of the random variable close to  $\sim 5 \times 10^{-4}$ . The total number of input variables is 30, considering reconstructed quantities provided by three different subdetectors: TRK, ACC and TOF. The full list of all the input features of  $AE_{ISS}$  and  $AE_{MC}$  is presented in table 4.2.

| Detector | Variable                                                |
|----------|---------------------------------------------------------|
| TRK      | $L_2$ Y residual                                        |
|          | $L_3$ Y and X residuals                                 |
|          | $L_4$ Y residual                                        |
|          | $L_5$ Y and X residuals                                 |
|          | $L_6$ Y residual                                        |
|          | $L_7$ Y and X residuals                                 |
|          | $L_8$ Y and X residuals                                 |
|          | $\frac{R_{INNER-L1}-R_{INNER}}{R_{INNER-L1}+R_{INNER}}$ |
|          | $\frac{R_{UH}-R_{LH}}{R_{UH}+R_{LH}}$                   |
|          | NormEdep2L5 <sub>XY</sub>                               |
|          | NormEdep2L6 <sub>XY</sub>                               |
|          | NormEdep2L7 <sub>XY</sub>                               |
|          | NormEdep2L8 <sub>XY</sub>                               |
|          | Number of X and Y hit inside inner                      |
|          | MIN and MAX Z on $L_2$                                  |
|          | MIN and MAX Z on $L_3$ - $L_4$                          |
|          | MIN and MAX Z on $L_5$ - $L_6$                          |
|          | MIN and MAX Z on $L_7$ - $L_8$                          |
| ACC      | Number of counters                                      |
| TOF      | Number of on-time clusters $L_3$                        |
|          | Number of on-time clusters $L_4$                        |

Table 4.2: The 30 input features for both the unsupervised autoencoders  $AE_{ISS}$  and  $AE_{MC}$ .

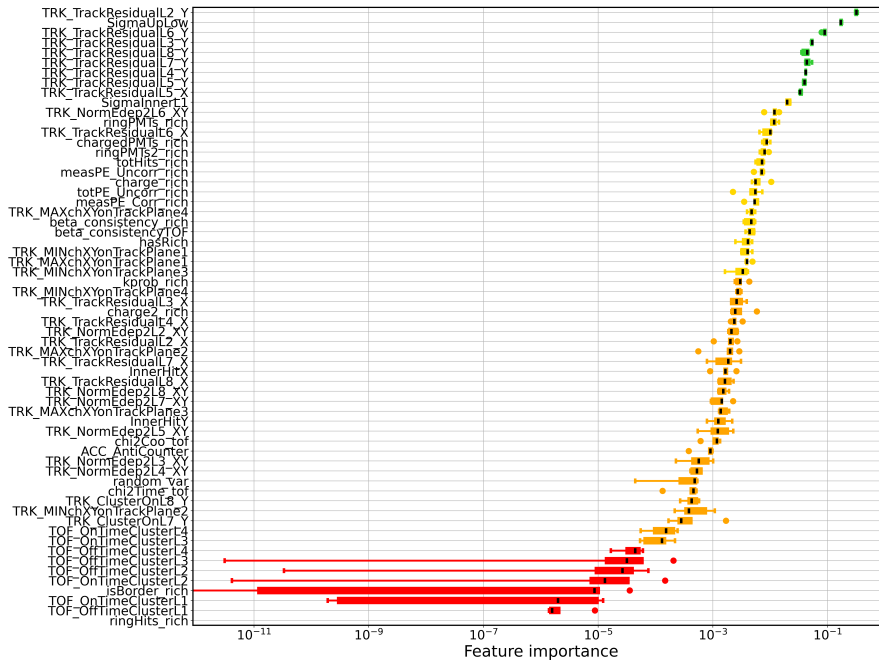


Figure 4.23: Features ranking for MC sample with both positive and negative rigidities, obtained using the rigidity sign as label.

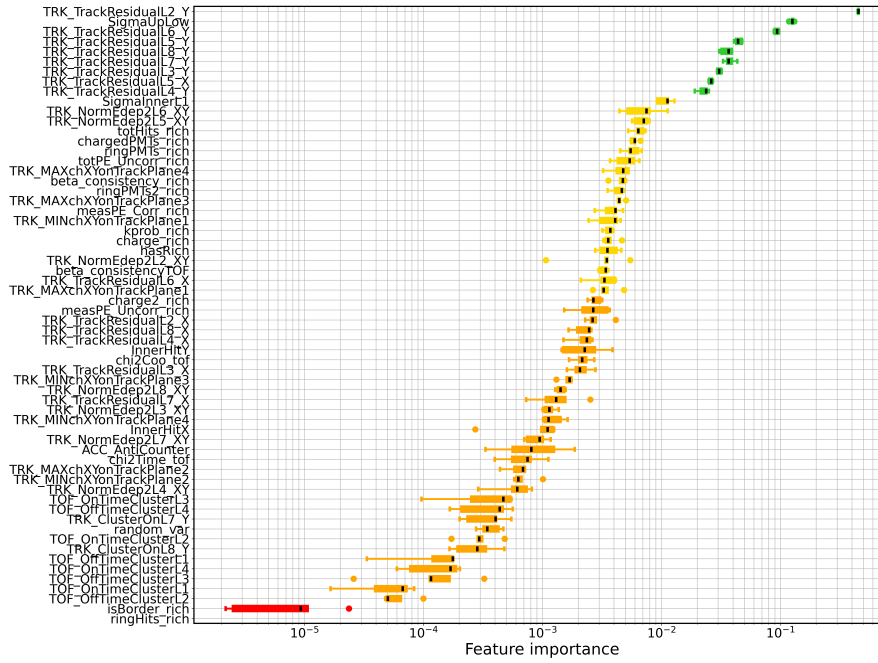


Figure 4.24: Features ranking for ISS sample with both positive and negative rigidities, obtained using the rigidity sign as label.

### 4.4.3 Dataset preprocessing, training and validation samples

The samples of positive rigidities are divided into two samples, the training sample contains 70% of the total statistic, whilst the validation sample is the remaining 30% [131, 132, 133]. Following a data-driven approach, two different autoencoders have been trained:  $AE_{ISS}$  and  $AE_{MC}$ . The first is trained on ISS data, the latter is trained using a combination of  $^3\text{He}$  and  $^4\text{He}$  simulated events. The algorithm trained on the MC is not applied on data and vice versa.

An additional cut on the absolute value of rigidity is applied to the two training samples. Indeed, only  $|R_{inner}| < 100$  GV are considered, where 100 GV is close to the MDR of the inner tracker span but still higher than maximum value to obtain a reliable mass measurement, i.e.  $R_{inner} \sim 50$  GV. This request is not applied to the supervised classifier, because the label definition is sensitive to spillover events, as explained in section 4.3.1, and cutting on the reconstructed rigidity of the inner tracker will reduce the available statistics for this irreducible background.

The data preprocessing follows the same logic used for the supervised classifiers and reported in the section 4.3.3. The variables listed in table 4.2 are not always present in each event. Analogously to the supervised classifiers case, the missing values are replaced by a placeholder value and successively masked, to prevent the autoencoders ( $AE_{ISS}$  and  $AE_{MC}$ ) from using non-physical information during the training. In this way, a single model can be trained using different event geometries, without loss of generalisation. To transform all input features to the same order of magnitude, the transformation described by equation 4.20 is applied. As a result, all the features are scaled to a normally distributed variable, with 0 mean and  $\sigma = 1$ . The masked values are not used and not affected by the transformation, avoiding possible biases and keeping them recognisable during the training.

### 4.4.4 Network structure and performance

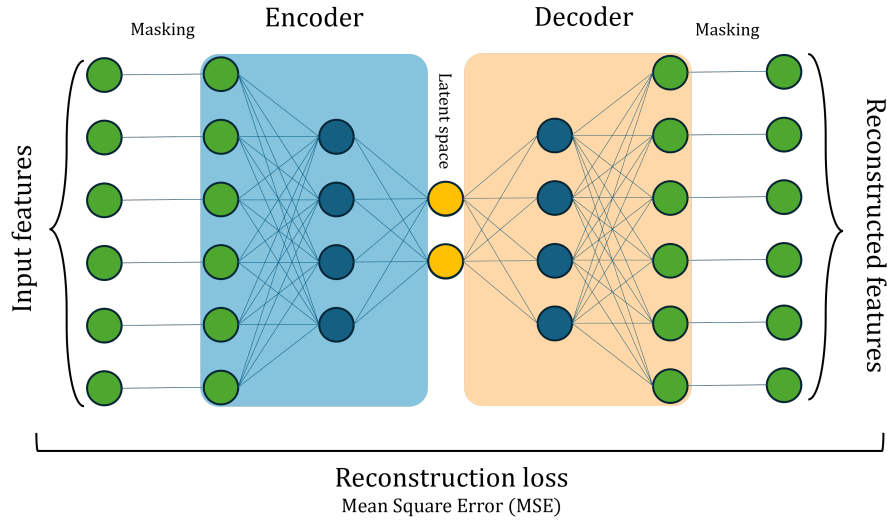


Figure 4.25: Pictorial representation of the two unsupervised autoencoders developed for this analysis:  $AE_{ISS}$  and  $AE_{MC}$ . Both the DNNs have an initial masking layer that prevents the use of non-physical information during the training.

The cartoon in Figure 4.25 shows a schematic representation of the two autoencoders, which have the same structure. The initial dimensionality of the autoencoders is equal to the number of input features, then it is reduced to the latent space dimension (yellow dots), afterwards, unlike the supervised classifiers, it returns to its initial value. The dimension of the latent space is a hyperparameter that must be tuned during training. The output of the two autoencoders are vectors with the same dimension as the inputs; comparing them, it is possible to compute the MSE.

A complete description of the number of layers, nodes and the values of all the hyperparameters are reported in Appendix B.2. The first layer<sup>3</sup> of each autoencoder applies a mask to the vector of the input features, following the same logic explained in section 4.3.4 for the supervised classifiers, with the only difference concerning the number of nodes, due to the different number of input features.

The autoencoder  $AE_{MC}$  has been trained for 100 epochs, whilst the model trained on ISS data ( $AE_{ISS}$ ) has been trained for 150 epochs. The mean values of the training and validation loss functions are presented in Figure 4.26, the left plot refers to  $AE_{MC}$ , whilst the right plot refers to  $AE_{ISS}$ . The blue lines are the training losses, and the orange lines are the validation losses. The loss function for  $AE_{MC}$  is a weighted MSE, while the  $AE_{ISS}$  uses an unweighted MSE loss described by equation 4.14. In both cases, the loss functions decrease along the training epochs, showing that the two algorithms are learning a meaningful representation of the training samples. The validation losses show similar behaviour in comparison with the training losses, thus the models are correctly generalising between the training and the validation sample, i.e. there is no clear evidence of overfitting occurring during the training.

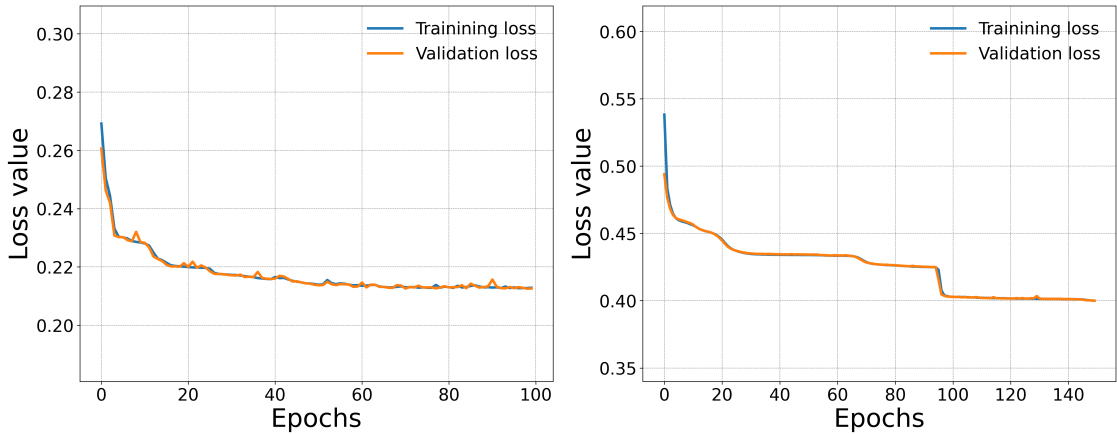


Figure 4.26: Training and validation loss function for Monte Carlo (left) and ISS(right) autencoders

The outputs of the two autoencoders, i.e. the anomaly scores, present a rigidity dependence mainly due to the chosen input features, as the track residuals, which are reported in Table 4.2. The anomaly scores are plotted in rigidity bins and are also divided depending on the velocity measurements available for each event. The plots of the anomaly score for three different rigidity bins, one for TOF, RICH-NaF and RICH-AGL, are displayed from Figure 4.27 to Figure 4.29. All

<sup>3</sup>See section 4.2.1 for a description of the main principles of DNN, i.e. a feed-forward multilayer perceptron.

the other rigidity bins are presented in Appendix B.2. The green histograms correspond to the positive rigidity samples, while the red histogram show events with a negative reconstructed rigidity.

The plots in Figure 4.27 show the anomaly scores distributions for  $AE_{MC}$  (left) and  $AE_{ISS}$  (right) for events with a reconstructed rigidity between  $|R| \in [1.92, 4.02]$  GV and a reconstructed  $\beta_{TOF}$ . Analogously, Figure 4.28 show events with a reconstructed  $\beta_{NaF}$  and rigidity between  $|R| \in [4.02, 6.47]$  GV, and Figure 4.29 show events with a measured rigidity in the interval  $|R| \in [6.47, 9.26]$  GV and going through the RICH-AGL.

The anomaly score distributions of the  $R > 0$  samples are all shifted towards small anomaly scores, i.e.  $\sim 0.6$ - $0.5$  with a tail reaching values close to 1. On the other hand, the same distribution for the  $R < 0$  samples appears flatter or with a peak at high values, i.e.  $> 0.9$ . As expected, the majority of the negative rigidities are poorly reconstructed by the two autoencoders. The events with an anomaly score  $\leq 0.6$  correspond to anomalous events, as described in Section 4.4.1. Among these events, those identified by  $AE_{MC}$  are charge-confused helium nuclei that constitute the background for the  $\overline{He}$  search; on the other hand, the events identified by  $AE_{ISS}$  may include possible signal candidates, although charge-confused background events can also be present within the sample.

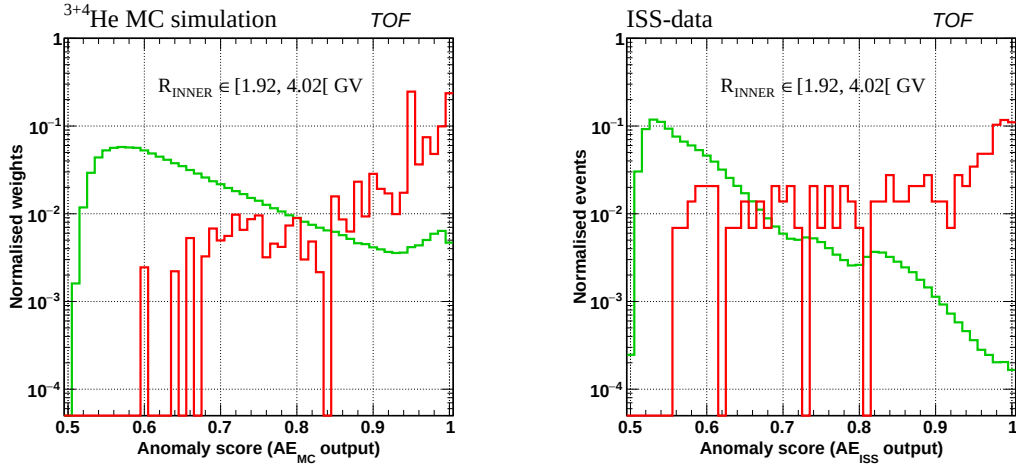


Figure 4.27: Anomaly scores for events with a reconstructed rigidity between  $|R| \in [1.92, 4.02]$  GV and a reconstructed  $\beta_{TOF}$ , obtained using the  $AE_{MC}$  (left) and the  $AE_{ISS}$  (right). The green histogram shows positive rigidities, and the red histogram shows negative rigidities.



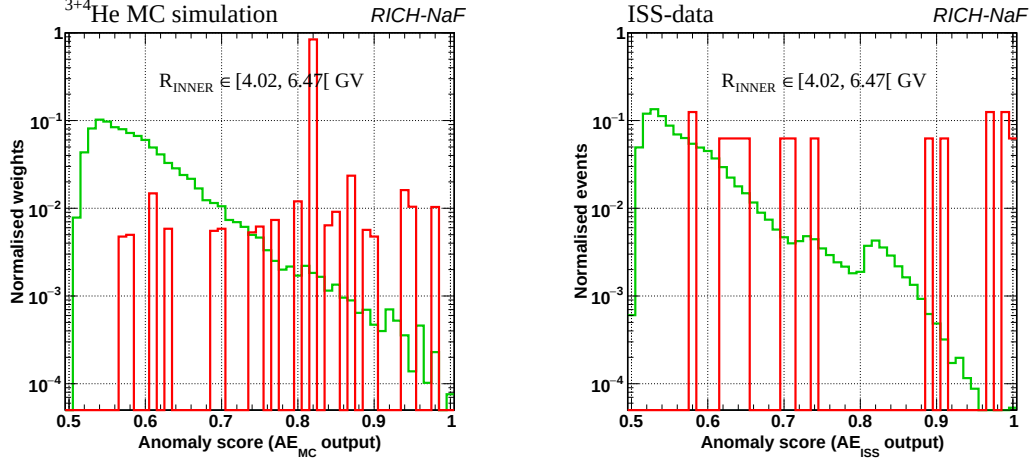


Figure 4.28: Anomaly scores for events with a reconstructed rigidity between  $|R| \in [4.02, 6.47]$  GV and a reconstructed  $\beta_{NaF}$ , obtained using the  $AE_{MC}$  (left) and the  $AE_{ISS}$  (right). The green histogram shows positive rigidities, and the red histogram shows negative rigidities.

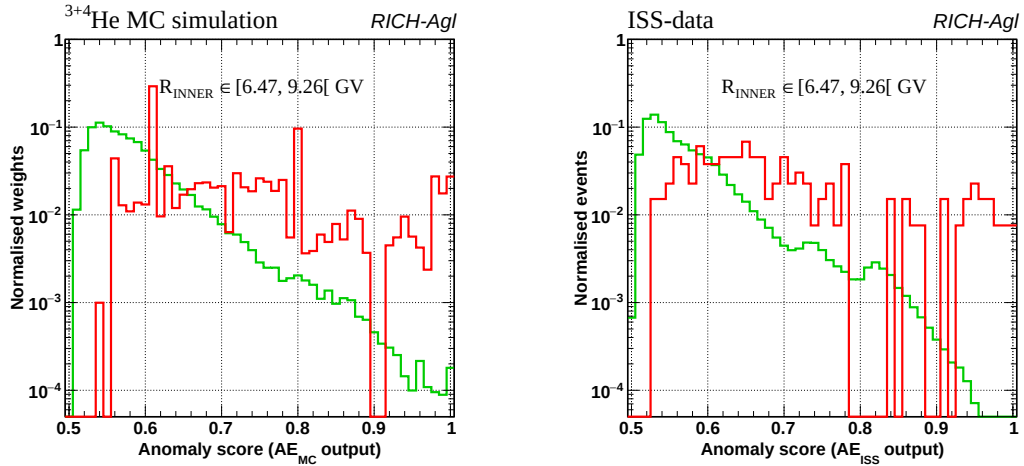


Figure 4.29: Anomaly scores for events with a reconstructed rigidity between  $|R| \in [6.47, 9.26]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the  $AE_{MC}$  (left) and the  $AE_{ISS}$  (right). The green histogram shows positive rigidities, and the red histogram shows negative rigidities.

## 4.5 Candidate identification strategy

In this analysis, four different DNNs have been developed: two supervised classifiers ( $CL_{ISS}$  and  $CL_{MC}$ ) and two unsupervised autoencoders ( $AE_{ISS}$  and  $AE_{MC}$ ). As discussed at the beginning of this Chapter (see Section 4.1), the algorithms trained on MC simulations are not applied to ISS data and the models trained on AMS-02 data are not applied to MC simulations. Machine Learning techniques based on supervised and unsupervised learning improve the background suppression by exploiting the correlations between variables, preserving a good efficiency on possible signal candidates.

The choice of developing two different tipologies of Machine Learning algorithms has the final goal of combining them to reach a higher background suppression. Additionally, two independent DNNs identifying the same events in experimental data provide a more robust hint of possible signal candidates. The candidate identification strategy used in this analysis is based on the assumption of orthogonality, i.e. independence, between the classifiers and the autoencoders. The test performed to confirm the validity of this assumption will be presented and discussed in the next Section.

The  $\overline{He}$  candidates we are looking for are clean events, going through AMS without annihilating, with a negative rigidity, a precise measurement of the velocity and a reconstructed mass compatible with those of  $^3He$  and  $^4He$  ( $m \in [2, 5] \text{ GeV}/c^2$ ). These kinds of events should have a classifier score close to 1 and an anomaly score  $\sim 0.5$ . Assuming that the outputs of the autoencoders and classifiers can be combined, selecting two working points, i.e. cut values, on the output of  $CL_{ISS}$  and  $AE_{ISS}$ , it is possible to identify possible signal candidates within ISS data. Analogously, given two equivalent working points for the models trained on MC, it is possible to estimate the expected background due to charge-confused helium nuclei. How to choose the working points on the DNN scores is not straightforward, since the order of the cuts might be relevant and not commutative.

In this work, the working points are selected based on the definition of two benchmark samples of signal-like events: one for ISS data and one for MC. The reference samples are obtained starting from the positive rigidity datasets, selecting events labelled as well-reconstructed rigidities (based on the definition of the classifier label at equations 4.16 and 4.17), and with a measured rigidity  $\leq 100 \text{ GV}$ . With this definition, the majority of the events have a classifier score close to 1 and a small anomaly score  $\sim 0.5$ , like the possible antihelium candidates we are searching for. Using only events of the benchmark samples, the working points are selected by choosing precise efficiencies on the network outputs. The corresponding cut values identify a signal region on the bidimensional plane composed of the  $CL_{ISS}$  and  $AE_{ISS}$  outputs. Computing the same efficiency values on the MC benchmark sample, two equivalent cut values for  $CL_{MC}$  and  $AE_{MC}$  are obtained. These cuts identify a control region on the plane of the two network outputs; the number of MC weighted events within the region is the estimate of the expected background due to charged-confused helium nuclei.

The Figures 4.30 and 4.31 show an application case of this strategy, considering only events with a measured rigidity within  $|R| \in [6.47, 9.26]$  GV and having a RICH-AGL velocity measurement. The plot in Figure 4.30 show only events with a positive rigidity, tagged as well-reconstructed by the classifier label definition. The models trained on MC are on the left, those trained on AMS data are on the right. The dashed pink lines correspond to the 95% efficiency working points on each DNN output. The shaded areas corresponds to the signal and control regions defined by the working points, with a total efficiency around 90%.

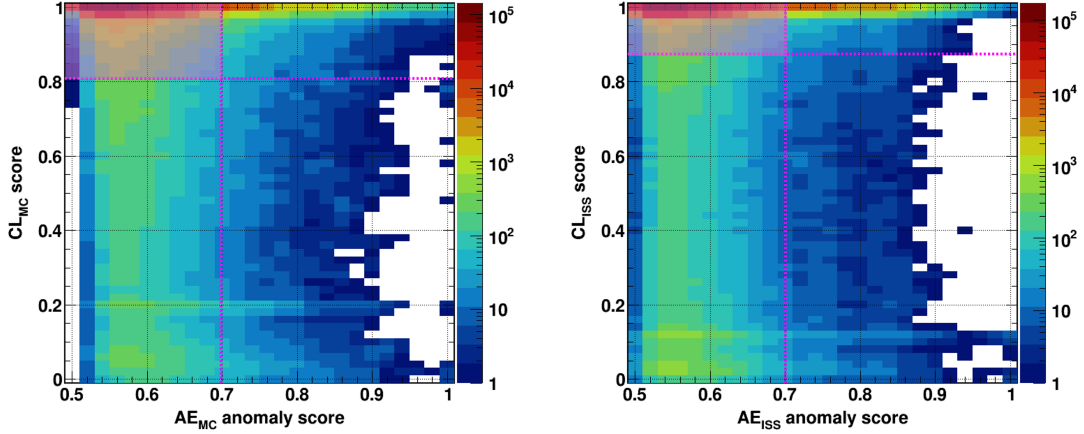


Figure 4.30: The two plots show the bidimensional plane of the DNNs outputs, the models trained on MC are on the left, and the models trained on ISS data are on the right. The events displayed have a positive rigidity within the following interval  $R \in [6.47, 9.26]$  GV, are tagged as well-reconstructed by the classifier label definition and have RICH-AGL velocity measurement. The dashed pink lines show the 95% efficiencies on each score, the shaded area in the top-left corner of the plots identifies the region selected by these working points, corresponding to an overall efficiency around 90%.

The plot in Figure 4.31 show only events with a negative rigidity. As before, the models trained on MC are on the left, those trained on AMS data are on the right. The pink lines and the shaded area are the same as in Figure 4.30. Using the four working points defined on the reference samples, the data and MC events within the shaded areas are selected.

Since no information on the velocity of the particle is provided as input to the DNN, after the selection on the DNN outputs, the mass measurement of the surviving events is an important cross-check to evaluate possible  $\overline{He}$  candidates. Indeed, reliable candidates should have a reconstructed mass compatible with the  $He$  nuclei.

The mass distributions relative to the four samples selected by the 95% efficient working points are presented in Figure 4.32. The agreement between data and MC for  $R > 0$  is good, and both mass distributions are well contained within the  $m \in [2, 5]$  GeV/c<sup>2</sup> interval. The majority of negative rigidity events are clearly background events, since the reconstructed masses are small and not compatible with helium-like particles. For this reason, an additional cut on the reconstructed mass distribution is applied: only  $R < 0$  events with a reconstructed mass between  $m \in [2, 5]$  GeV/c<sup>2</sup> will be considered in this analysis. Adding the mass cut, data and MC are in

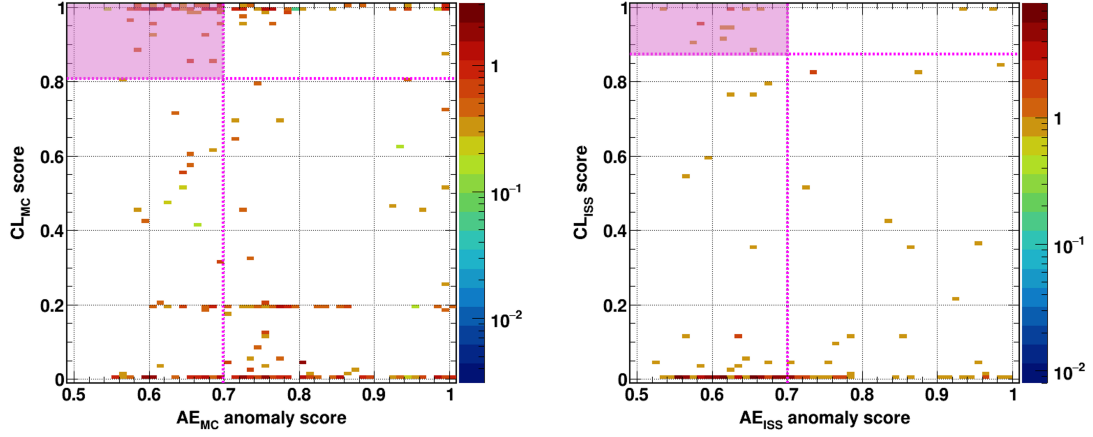


Figure 4.31: The two plots show the bidimensional plane of the DNNs outputs, the models trained on MC are on the left, and the models trained on ISS data are on the right. The events displayed have a negative rigidity within the following interval  $|R| \in [6.47, 9.26]$  GV and have RICH-AGL velocity measurement. The dashed pink lines show the 95% efficient working points for each DNN, the shaded area in the top-left corner of the plots identifies the enhanced signal region with an overall efficiency around 90%. The number of weighted MC events within the shaded region is the estimate of the expected background.

good agreement also for events with  $R < 0$ , yielding respectively:  $4 \pm 2$  and  $7.71 \pm 1.98$  events.

In conclusion, after cutting on the DNN outputs, an additional cut on the mass distribution is applied ( $m \in [2, 5]$  GeV/c<sup>2</sup>). The scheme in Figure 4.33 summarises the candidate identification strategy applied in this work. Since the 95% working points (WPs) are too loose, tighter cut values will be considered in the next Chapter. As discussed in the Sections 4.3.4 and 4.4.4, the scores of the DNNs are rigidity dependent, thus the efficiencies must be evaluated for each rigidity bin and for the three AMS subdetectors which measure the particle velocity: TOF, RICH-NaF and RICH-AGL. The division based on which velocity measurements are available takes into account the different velocity and mass resolution of each subdetector. Due to AMS-02 mass resolution, the reconstructed mass of  $He$  or  $\overline{He}$  events can be measured up to  $R_{rec} \sim 50$  GV. There is also a lower bound due to large energy losses and an increase of the expected background which has been set to 1.92 GV. For these reasons, the  $[1.92, 50]$  GV interval has been divided in 10 rigidity bins corresponding to:

|               |              |              |               |              |              |
|---------------|--------------|--------------|---------------|--------------|--------------|
| Bin number    | 1            | 2            | 3             | 4            | 5            |
| Rigidity [GV] | [1.92, 4.02] | [4.02, 6.47] | [6.47, 9, 26] | [9.26, 13.0] | [13.0, 18.0] |

|               |              |              |              |              |              |
|---------------|--------------|--------------|--------------|--------------|--------------|
| Bin number    | 6            | 7            | 8            | 9            | 10           |
| Rigidity [GV] | [18.0, 24.7] | [24.7, 31.1] | [31.1, 36.1] | [36.1, 41.9] | [41.9, 48.5] |

Table 4.3: The ten rigidity bins used in this analysis to compute the working points on the four algorithms:  $CL_{ISS}$ ,  $CL_{MC}$ ,  $AE_{ISS}$ , and  $AE_{MC}$ .

The rigidity bins are obtained from the AMS-02 published helium flux [84]. Each subdetectors can

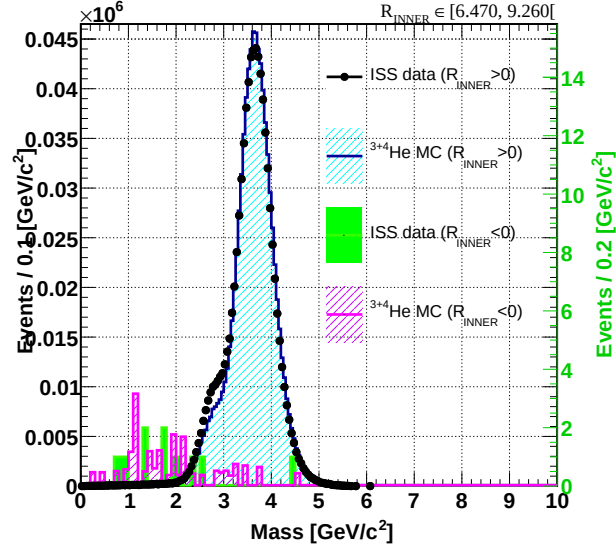


Figure 4.32: Mass distribution for ISS data and  ${}^3\text{He}$ - ${}^4\text{He}$  MC, for  $Z = +2$  and  $Z = -2$  samples after the selection on the four DNNs scores with measured rigidities within  $|R| \in [6.47, 9.26]$  GV. The working points are the same as shown in Figure 4.30, with an overall efficiency of 90%.

be used in a subrange of the chosen rigidity interval, indeed considering the velocity resolution (and the Cherenkov threshold): the TOF is used within the first two bins, the RICH-NaF is used from the 2<sup>nd</sup> bin up to the 5<sup>th</sup> (included), and the RICH-AGL is used from the 3<sup>rd</sup> up to the 10<sup>th</sup> rigidity bin.

For each subdetector, after the choice of the WPs, the cuts on the algorithm scores and on the mass distributions are applied in each rigidity bin. To achieve a more accurate background estimation, all the surviving events are added, returning a rigidity integrated scalar value for three AMS subdetectors: TOF, RICH-NaF and RICH-AGL. Any possible excess is evaluated statistically by comparing the number of remaining events in the data and MC.

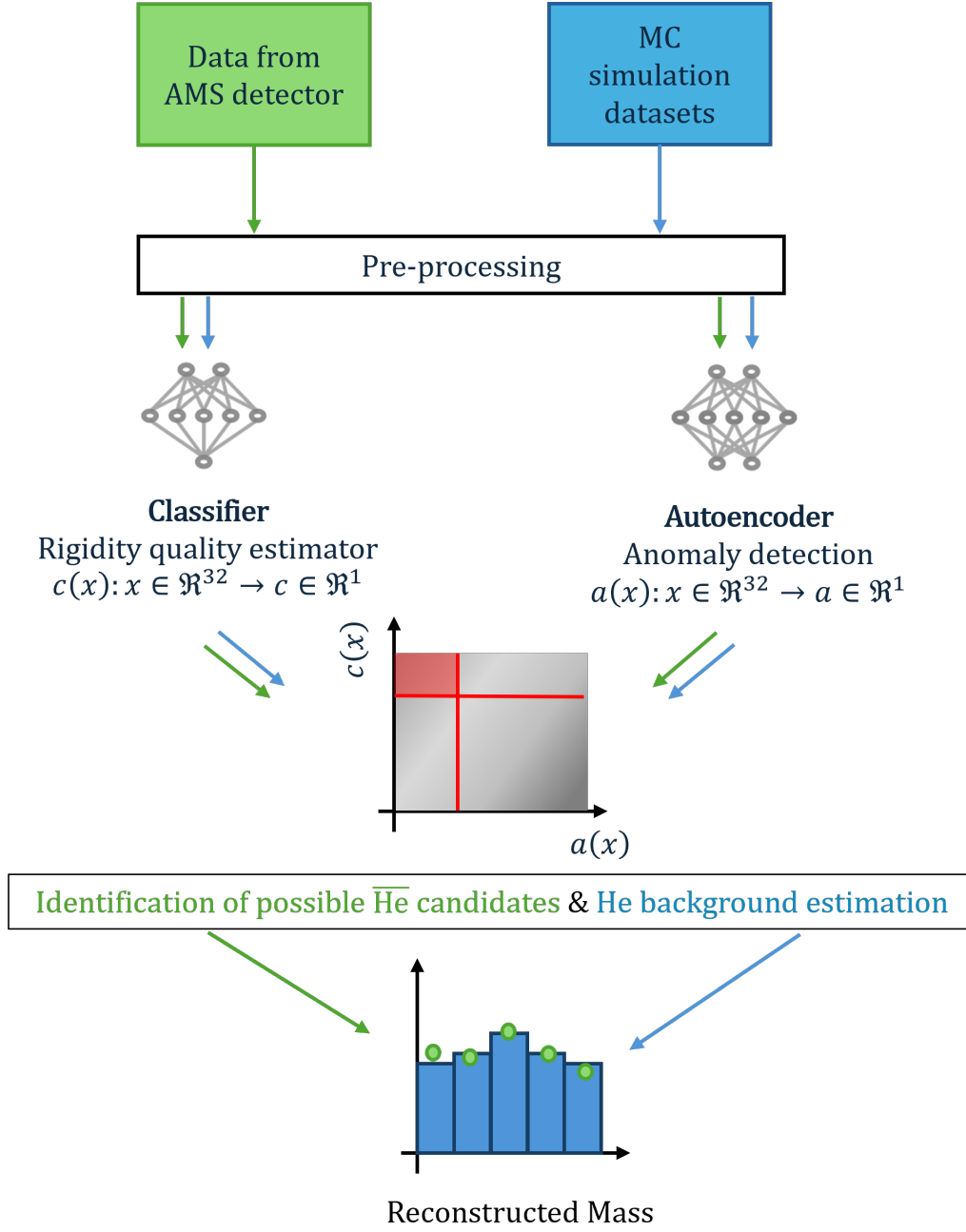


Figure 4.33: Scheme of the analysis workflow, repeated for three different AMS detectors (TOF, RICH-NaF and RICH-AGL) and 10 rigidity bins. Afterwards, the number of surviving events in MC and data is summed between every rigidity bin to achieve a more precise estimate of the expected background.

## 4.6 Orthogonality

The classifier scores can be coupled with the anomaly scores if the networks are independent and orthogonal, up to a reasonable level. Both the DNNs use reconstructed physical quantities measured by the silicon tracker, but the classifiers exploit the information from the RICH, whilst the autoencoders use the TOF and the ACC. Since some of the input variables are identical, there might be a correlation between the scores and the anomaly scores. Despite these similarities, the two types of networks have some key differences regarding the training paradigms, the task to fulfil, and the training datasets. The classifiers need a label to divide the training sample into classes; they use a supervised approach (Section 4.2.3). On the other hand, autoencoders do not require any labels and exploit unsupervised learning (Section 4.2.4). The tasks of the classifiers and the autoencoders can be summarised as follows:

- Task of  $CL_{ISS}$  and  $CL_{MC}$ : distinguish between two classes (well-reconstructed and poorly reconstructed rigidities) defined by the equations 4.16 and 4.17. The training samples contain spillover events, which are the dominant component of the poorly reconstructed rigidities. After the training, the classifiers are sensitive to spillover events and can distinguish them from well-measured rigidities.
- Task of  $AE_{ISS}$  and  $AE_{MC}$ : learn a meaningful representation of the training sample. In other words, compress the input features vector into a smaller latent space and then recover the initial values minimising the losses of information. Using a training sample dominated by well-reconstructed events, the autoencoders learn to recognise these events.

The two DNNs typologies are both trained on the positive rigidities samples, but the training sample of the autoencoders has an additional cut on the measured rigidities maintaining only events with  $R \leq 100$  GV. This cut reduces the percentage of events with a measured rigidity close to the MDR, increasing the relative percentage of well-reconstructed events.

The hypothesis of the orthogonality between the classifier scores and the anomaly scores has been tested assuming that, if the classifier scores can be combined with the autoencoder anomaly score, the distributions of  $CL_{MC}$  ( $CL_{ISS}$ ), considering only events satisfying the definition of the benchmark samples, should not vary while cutting on the anomaly score of  $AE_{MC}$  ( $AE_{ISS}$ ). The plots in the Figures 4.34, 4.35 and 4.36 show three results of this test for all four models developed in this thesis. All the remaining results are reported in Appendix B.3.

In each Figure, the plots show the ratio between the normalised distributions of the classifier ( $CL_{MC}$  or  $CL_{ISS}$ ) output without and with different cuts on the anomaly score (of  $AE_{MC}$  or  $AE_{ISS}$ ). Five different efficiencies on the anomaly score have been tested: 95%, 90%, 85%, 80%, and 67%. The scores are divided into rigidity bins, and in subdetectors, considering the velocity measurements available for the corresponding event.

For all the plots, the classifier score distributions are slightly modified after cutting on the anomaly score, resulting in smaller tails towards 0 and an increased percentage of events with scores close

to 1. The higher discrepancy in the ratio occurs at the next-to-last bin, where the value drops around  $\sim 0.8$ . Apart from that, in most cases the ratio is compatible or close to 1 for most of the bins at high score values. This effect shows that, depending on the efficiency of the anomaly score cut, the number of events wrongly classified as bad-reconstructed rigidities by the classifier is reduced. Interestingly, the two networks present a coherence between their outputs.

In conclusion, there is a negligible correlation between the outputs of the classifiers and the autoencoders, thus cutting first on the anomaly scores preserves within the 5 – 10% the shape of the classifier scores. After the cut on the anomaly score, the distributions of the classifier scores are skewed towards 1, and the magnitude of the effect depends on the hardness of the cut. In other words, evaluating the working points on the efficiencies of the network outputs relative to the benchmark samples, under the assumption of perfect orthogonality between classifiers and autoencoders, might result in an underestimate of the signal efficiency, thus in a conservative approach.

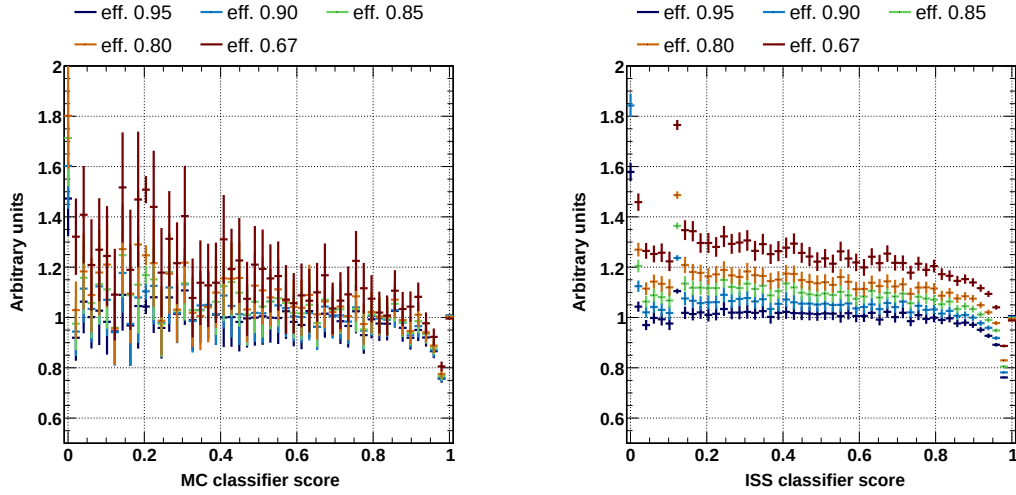


Figure 4.34: Ratios between the score distribution without any cut and the distributions with a cut on the anomaly score for five different cut values. The left plot presents the ratios for the model trained on MC, and the plot on the right contains the results for the models trained on AMS data. The distributions show only events with a measured rigidity in the following bin  $R \in [1.92, 4.02]$ , labelled as well-reconstructed (see equation 4.16), and having a TOF velocity measurement.



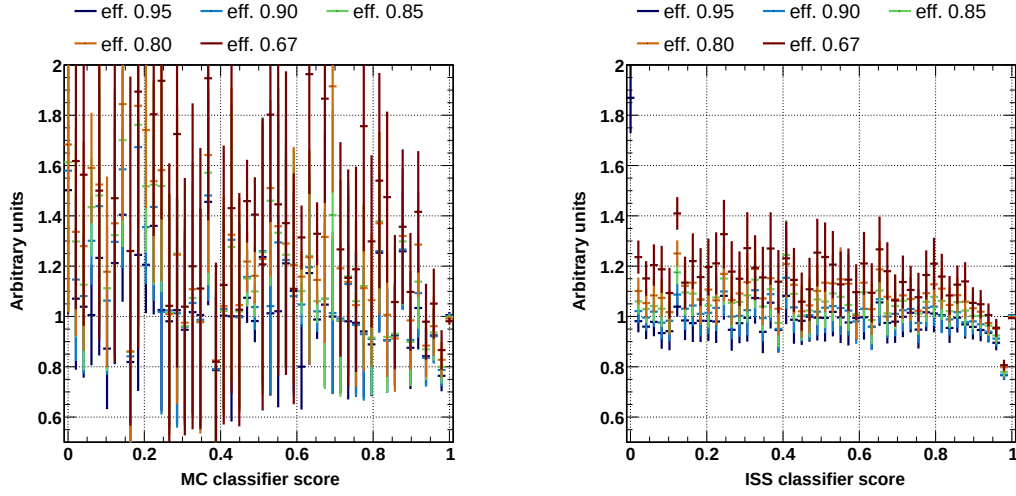


Figure 4.35: Ratios between the score distribution without any cut and the distributions with a cut on the anomaly score for five different cut values. The left plot presents the ratios for the model trained on MC, and the plot on the right contains the results for the models trained on AMS data. The distributions show only events with a measured rigidity in the following bin  $R \in [4.02, 6.47]$ , labelled as well-reconstructed (see equation 4.16), and having a RICH-NaF velocity measurement.

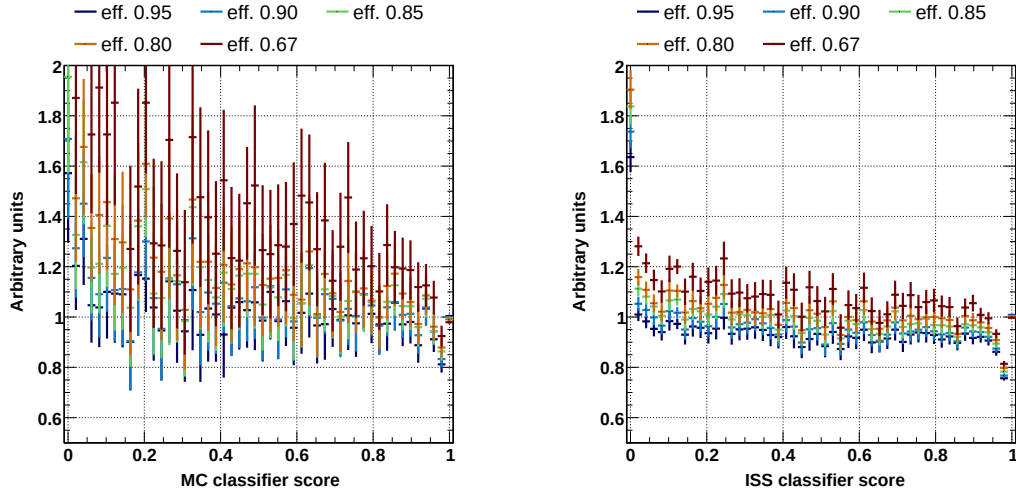


Figure 4.36: Ratios between the score distribution without any cut and the distributions with a cut on the anomaly score for five different cut values. The left plot presents the ratios for the model trained on MC, and the plot on the right contains the results for the models trained on AMS data. The distributions show only events with a measured rigidity in the following bin  $R \in [6.47, 9.26]$ , labelled as well-reconstructed (see equation 4.16), and having a RICH-AGL velocity measurement.

# Chapter 5

## Results

In this Chapter the results of the analysis are discussed and presented. The analysis workflow described in Section 4.5 is applied to three different AMS-02 subdetectors: TOF, RICH-NaF and RICH-AGL. The number of possible antihelium candidates from data and the corresponding background estimates from MC obtained as a function of the chosen working points (WPs). In the first part of this Chapter I describe how I chose three different couples of WPs, and I report the corresponding efficiency values for each one of the four algorithms developed and presented in Chapter 4. To evaluate any possible excess in data, I define a likelihood function, which is used, through the CLs method, to discuss the compatibility of the observed value of the test statistic with the pure background hypothesis. Afterwards, three (observed or expected) exclusion limits for each one of the considered AMS detectors are presented, along with the antihelium candidates found by this work.

### 5.1 Signal and control region definition

The candidate identification strategy described in Section 4.5 has been used to select the WPs and define a signal and a control region as a function of the relevant AMS subdetectors and rigidity bin. For a single rigidity bin, the signal and control regions are defined as follows:

- The signal region is the area on the DNNs outputs plane, obtained for AMS data with  $R < 0$ , selected using the WPs defined on the scores of  $AE_{ISS}$  and  $CL_{ISS}$ .
- The control region is identified using the same efficiencies on the plane of  $AE_{MC}$  and  $CL_{MC}$ , obtained for MC events with  $R < 0$ .

Each signal region is uniquely identified by an ordered pair of numbers, representing the cut efficiencies on the anomaly score and on the classifier score. The order of the cuts is relevant, and cutting first on the  $AE_{ISS}$  anomaly score preserves the orthogonality, as reported in Section 4.6. The four cut values are extracted in the same way, but their values are different since they are obtained from the output of different algorithms. As described in Section 4.5, all the regions are

then integrated in rigidity. In this way, due to the limited available MC statistic, the estimate of the expected background is improved.

The Monte Carlo plays a fundamental role in this analysis, indeed it is not possible to define a control region using only ISS data. The expected background sources (Section 4.1) produce a resonant background identical to the possible signal. The background estimation is possible only by relying on Monte Carlo simulations. Another relevant aspect is the normalisation of the simulated events. The MC re-weighting described in 3.2.1 guarantees the direct comparison between simulated events and ISS data.

In this work, three different ordered pairs of efficiencies have been used: (95%, 50%), (80%, 50%), and (75%, 95%). The first number is the anomaly score efficiency, the second is the classifier score efficiency, and the total signal efficiency is the product of the two, yielding respectively  $95\% \times 50\% = 48\%$ ,  $80\% \times 50\% = 40\%$ , and  $75\% \times 95\% = 71\%$ . The pairs have been chosen without knowing the final results of the analysis, avoiding possible biases.

The Figures 5.1 and 5.2 show the single WPs efficiencies, respectively, on the positive rigidity benchmark samples and the negative rigidity samples. The models trained on MC are on the left, and the models trained on ISS data are on the right. The 95% efficiency has been chosen looking at the orthogonality checks between the networks, as visible in the Figures 4.34, 4.35, and 4.36 (but also in Appendix B.3), the distortion of the classifier scores is reduced for looser cuts on the anomaly score. On the other hand, to remove as many background events as possible, the coupled classifier efficiency is much tighter, i.e. 50%. The second pair follows the same logic but with a tighter cut on the anomaly scores, resulting in a loss of total signal efficiency around 8%. The last pair, unlike the first two, has a tighter cut on the anomaly scores with respect to the classifier scores but a higher overall efficiency.

The results for AMS-02 data and MC for each pair of WPs and the considered three subdetectors (TOF, RICH-NaF and RICH-AGL) are reported in Table 5.1. The analyses regarding the TOF and RICH-NaF do not present any excess, regardless of the chosen WPs. On the other hand, RICH-AGL is more interesting as it shows two cases where the number of events in ISS data is higher than the expected background, having the following corresponding WPs pairs: (95%, 50%) and (80%, 50%).

The third couple of WPs, (75%, 95%) for the RICH-AGL detector, shows a discrepancy between the expected background events from MC and the observed events in data, i.e.  $\sim 27$  against 17. The overestimate of the expected background might be related to the loose classifier score efficiency. Indeed, as visible in Figure 4.19, the distributions of the classifier scores have a sharp peak at 1 for well-reconstructed rigidities and tails being 2-3 order of magnitude smaller. The 95% efficiency cut lies on the tails of the distributions, as shown in Figure 5.1, inheriting a less stable behaviour due to the smaller statistics. On the other hand, a tight cut on the classifier scores efficiency lies in the bulk of the distribution, providing a more reliable result. The same argument does not apply to a loose anomaly score efficiency, since the distributions are broader and the differences between the peaks and the tails are smaller, i.e. around  $\sim 1$  order of magnitude as

visible in the Figures 5.1 and 4.29.

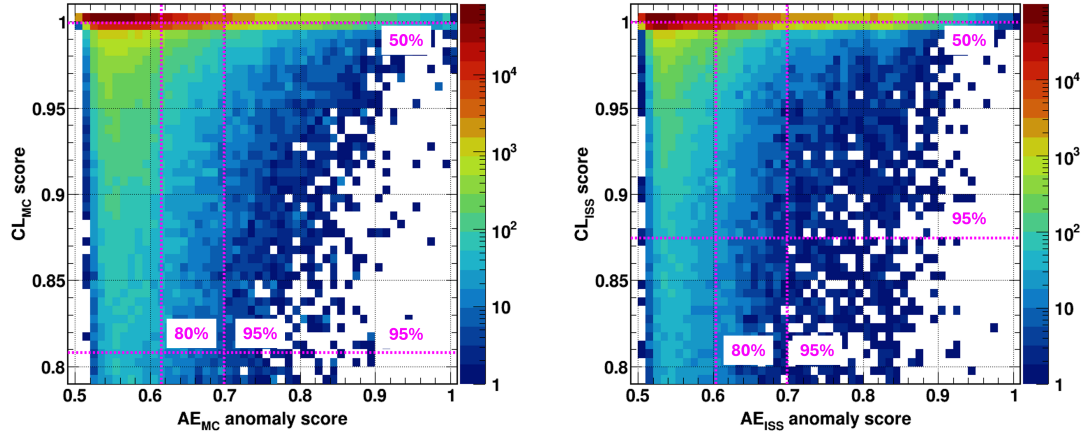


Figure 5.1: The bidimensional planes of the DNNs outputs, MC (left) and data (right). The events displayed have a positive rigidity within  $R \in [6.47, 9.26]$  GV, are tagged as well-reconstructed by the classifier label definition and have RICH-AGL velocity measurement. The dashed pink lines correspond to different WPs, and the efficiencies are shown close by.

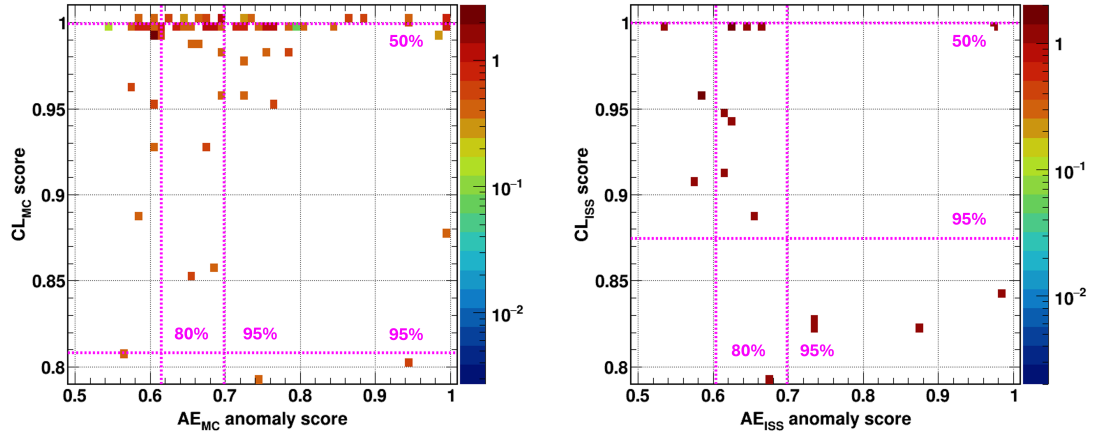


Figure 5.2: The bidimensional planes of the DNNs outputs, MC (left) and data (right). The events displayed have a negative rigidity within  $|R| \in [6.47, 9.26]$  GV, and have RICH-AGL velocity measurement. The dashed pink lines correspond to the same WPs shown in Figure 5.1.

| Detector | WPs( $\varepsilon_{AE}$ , $\varepsilon_{CL}$ ) | ISS data events | MC weighted events | MC events |
|----------|------------------------------------------------|-----------------|--------------------|-----------|
| TOF      | (95%, 50%)                                     | 0               | 2.80               | 5         |
|          | (80%, 50%)                                     | 0               | 2.43               | 4         |
|          | (75%, 95%)                                     | 0               | 2.02               | 3         |
| RICH-NaF | (95%, 50%)                                     | 2               | 2.39               | 5         |
|          | (80%, 50%)                                     | 1               | 2.39               | 5         |
|          | (75%, 95%)                                     | 2               | 2.93               | 6         |
| RICH-AGL | (95%, 50%)                                     | 12              | 8.65               | 21        |
|          | (80%, 50%)                                     | 7               | 5.49               | 15        |
|          | (75%, 95%)                                     | 17              | 27.44              | 71        |

Table 5.1: The number of surviving events in AMS-02 data,  $^3\text{He}$  and  $^4\text{He}$  MC (with and without weights) for the three pairs of WPs and the three subdetectors: TOF, RICH-NaF and RICH-AGL.

### 5.1.1 Expected electron background

As previously discussed in Section 4.1.1, electrons must be considered as possible background events for antihelium searches. Indeed, cosmic electrons are highly relativistic particles which suffer larger energy losses, as a direct consequence of their higher  $\beta\gamma$  as shown by the Bethe-Block expression (see equation 2.3) that might lead to the misidentification of an electron as a  $Z = -2$  particle. The electron pollution has been evaluated using MC simulations, after the cut on the DNN outputs and the selection on the reconstructed mass, i.e.  $m_{rec} \in [2, 5] \text{ GeV}/c^2$ . The results are summarised in Table 5.2 for each chosen pair of WPs and each subdetector.

For the TOF, the background due to electrons is absent independently of the WPs on the DNNs. Considering the RICH-NaF, for two pairs of WPs, i.e. (80%, 50%) and (75%, 95%), no surviving events are found; the third pair (95%, 50%) presents a small contribution of electrons with a reconstructed mass within the selected interval. This population can be removed, requiring a reconstructed charge by the inner tracker higher than 1.9, as shown in the left plot of Figure C.3. The additional cut on the inner tracker charge has a  $\sim 100\%$  efficiency on the surviving events in ISS data and helium MC, as visible from the right plot of Figure 5.4. On the other hand, the analysis based on RICH-AGL presents a slightly higher contamination, which can be neglected by applying the additional cut on the inner tracker reconstructed charge.

The Figures 5.3 and 5.4 show the inner tracker charge distributions for the electron and helium MC for the three different pairs of WPs. In each case, after requiring an inner tracker charge  $> 1.9$ , the electron contamination is at least two orders of magnitude smaller than the expected background due to charged-confused helium nuclei, resulting in an overall negligible contribution. The pair (75%, 95%) has the highest electron pollution, while the other two pairs present a difference of  $\sim 10^{-3}$  between helium and electrons. Similar plots for the TOF and RICH-NaF are reported in Appendix C.

| Detector | WPs( $\epsilon_{AE}$ , $\epsilon_{CL}$ ) | $^{3+4}\text{He}$ MC | $e^-$ MC ( $Z_{TRK} > 1.9$ ) |
|----------|------------------------------------------|----------------------|------------------------------|
| TOF      | (95%, 50%)                               | $2.80 \pm 1.27$      | 0                            |
|          | (80%, 50%)                               | $2.43 \pm 1.16$      | 0                            |
|          | (75%, 95%)                               | $2.02 \pm 1.03$      | 0                            |
| RICH-NaF | (95%, 50%)                               | $2.39 \pm 1.12$      | 0                            |
|          | (80%, 50%)                               | $2.39 \pm 1.12$      | 0                            |
|          | (75%, 95%)                               | $2.93 \pm 1.25$      | 0                            |
| RICH-AGL | (95%, 50%)                               | $8.65 \pm 1.99$      | $7.8 \times 10^{-3}$         |
|          | (80%, 50%)                               | $5.49 \pm 1.47$      | $2.5 \times 10^{-3}$         |
|          | (75%, 95%)                               | $27.44 \pm 3.47$     | 0.31                         |

Table 5.2: The expected number of electron background compared with the contribution of  $^3\text{He}$  and  $^4\text{He}$  MC (considering weighted events) for the three pairs of WPs and the three subdetectors: TOF, RICH-NaF and RICH-AGL.

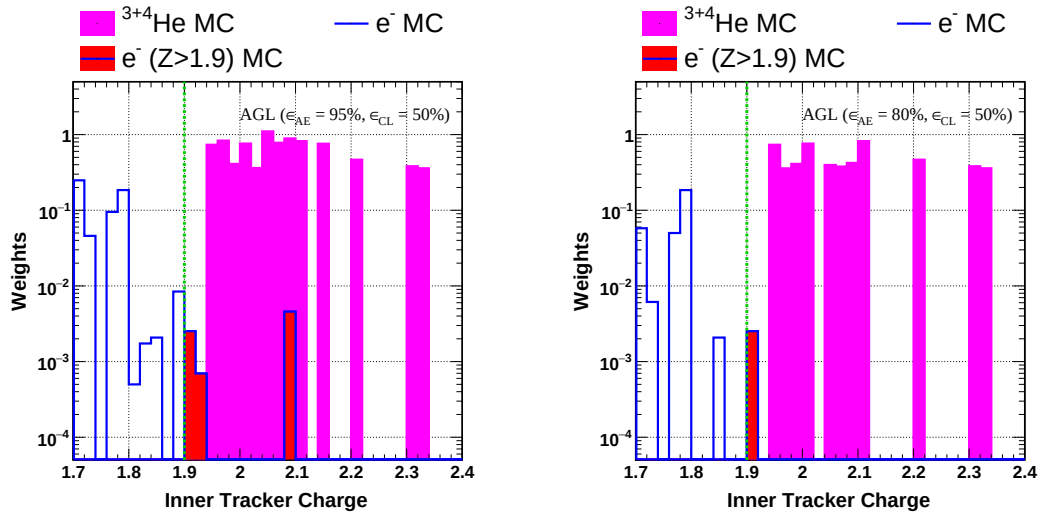


Figure 5.3: Distributions of the inner tracker charge for simulated helium and electron events traversing the RICH-AGL. The left plot shows the result obtained using the (95%, 50%) pair, while the right plot presents the case with the (80%, 50%) pair. The electron contribution can be neglected by requiring  $Z_{TRK} > 1.9$ .

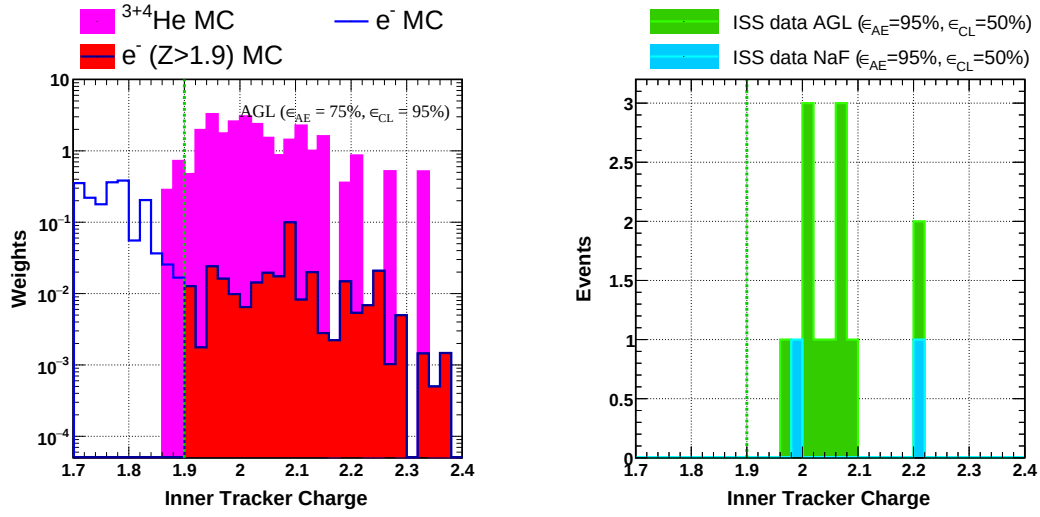


Figure 5.4: (Left) Distributions of the inner tracker charge for simulated helium and electron events traversing the RICH-AGL obtained considering the (75%, 95%) pair. (Right) Distributions of the inner tracker charge for ISS data survivors obtained using the (95%, 50%) pair both for RICH-NaF and RICH-AGL. The electron contribution can be neglected by requiring  $Z_{\text{TRK}} > 1.9$ , which does not affect the possible antihelium candidates found in data.

## 5.2 The CLs method and the likelihood definition

At the beginning of the century, a vivid discussion regarding the results presentation of search experiments occurred, which can be summarised in three main points or approaches:

- Bayesian credible interval: based on a prior probability distribution, use the observed data to constrain one or more Parameters Of Interest (POIs) into credible intervals, through the theorem of Bayes [138].
- Frequentists confidence interval: is a member of a set such that the set has the following property:  $P(\mu) \in [\mu_1, \mu_2] = \alpha$ . Where  $\mu$  is the POI,  $\mu_1, \mu_2$  are functions of the measurement  $x$  and  $\alpha$  is the confidence level. Assuming the true value of the POI to be  $\mu_t$ , the intervals contain  $\mu_t$  in a fraction  $\alpha$  of experiments [139, 140, 141].
- CLs method: is a modified frequentist approach based on the test statistic  $Q = \frac{L_{s+b}}{L_b}$ , where  $L_{s+b}$  is the likelihood under the hypothesis of discovery ( $s \neq 0$ ) and  $L_b$  is the likelihood under the pure background hypothesis. The null hypothesis is the background hypothesis, i.e. data can be understood with the existing physics. The alternative hypothesis requires the introduction of new physics to understand the data [138].

A weakness of frequentist confidence intervals is the possibility of obtaining unintuitive results. The most common example considers two experiments with identical efficiencies and observations, but the experiment with the larger expected background quotes the strongest limit on the physical parameter. Moreover, the experiment with the larger background can exclude false signals up to lower signal rates than the superior experiment [138]. Although all three alternatives have strengths and weaknesses, this analysis is based on the CLs method. There are two main reasons: this method does not require a prior probability distribution, and it's possible to make a statement about the signal+background hypothesis.

The CLs method is based on the likelihood ratio, which is the optimal test statistic for performing a hypothesis test as expressed by the Neyman–Pearson lemma [142]. This analysis aims to make a statement on whether or not data show any deviation from the null hypothesis, i.e. the pure background hypothesis. The test statistic is:

$$Q = -2 \ln \left[ \frac{L_{s+b}}{L_b} \right] = -2 \ln \left[ \frac{L(s \neq 0, \hat{b}(s))}{L(\hat{s}, \hat{b})} \right] + 2 \ln \left[ \frac{L(s = 0, \hat{b}(0))}{L(\hat{s}, \hat{b})} \right] \quad (5.1)$$

where  $L$  is the likelihood function,  $s$  is the signal rate (the POI),  $b$  is a generic nuisance parameter,  $\hat{s}$  and  $\hat{b}$  are the unconditional Maximum Likelihood Estimators (MLEs), and  $\hat{b}(s)$  is the conditioned MLE. If the null hypothesis (only background) is true, the test statistic moves towards higher values. Therefore, it is interesting to evaluate the p-value defined as:

$$CL_B = \int_{-\infty}^{Q^{OBS}} f(Q|B) dQ. \quad (5.2)$$



where  $f(Q|B)$  is the probability density function of the test statistic  $Q$  under the background-only hypothesis. On the other hand, if the alternative hypothesis is true (signal + background), the test statistic moves in the opposite direction; analogously, the interesting p-value is:

$$CL_{S+B} = \int_{Q^{OBS}}^{+\infty} f(Q|S+B) dQ. \quad (5.3)$$

where  $f(Q|S+B)$  is the probability density function of the test statistic  $Q$  under the signal+background hypothesis. The decision whether reject or not the signal+background hypothesis, for a given value of the POI  $s$ , is taken using the ratio of the p-values:

$$CL_S = \frac{CL_{S+B}}{1 - CL_B}. \quad (5.4)$$

The observed value of  $CL_B$  is a test of the signal significance, i.e. an observation lower than  $2.8 \times 10^{-7}$  results in a  $5\sigma$  deviation from the background-only hypothesis and a discovery. In literature, a  $S+B$  hypothesis is rejected if  $CL_S < 0.05$ , identifying an exclusion interval. At any point in the exclusion interval, the probability of having falsely excluded a true signal hypothesis is less than 5% [138]. The method does not allow one to claim anything regarding the unexcluded region within the parameter space.

### 5.2.1 Likelihood definition

The likelihood definition is based on the so-called On-Off problem. This problem arises when the existence of a source of a new signal must be proven. For example, Li and Ma in the 80's described the problem for  $\gamma$  rays sources, where experimental data must be analysed to determine the confidence level of a candidate source. Meaning, verify whether the excess in counts of photons is due to a genuine source rather than to a background fluctuation [143]. The problem becomes more complex if nuisance parameters are present and if the number of counts is small.

In the context of this analysis, the problem can be described as follows:

- An experiment measures  $N$  events, assumed to follow a Poisson distribution  $\text{Pois}(N|s+b)$ . The mean of the distribution is  $E[N] = s+b$ , with  $s$  signal rate and  $b$  background rate. The Parameter Of Interest (POI) is  $s$ , and  $b$  is a nuisance parameter.
- A control measurement is performed to constrain the nuisance parameter  $b$ . This measurement is also Poissonian  $\text{Pois}(M|\tau b)$ . The mean of the distribution is  $E[M] = \tau b$  and the parameter  $\tau$  is assumed to be known.

Under these assumption the Likelihood is the product of two Poissonians distributions:

$$L(s+b) = \frac{(s+b)^N}{N!} e^{-(s+b)} \cdot \frac{(\tau b)^M}{M!} e^{-\tau b} \quad (5.5)$$

$N$  is the number of events, with  $R_{INNER} < 0$ , observed in ISS data after the cut on the DNNs scores lying within the mass window of  $He$  nuclei ( $m \in [2, 5] \text{ GeV}/c^2$ ).  $M$  is the number of MC

counts obtained in the same way as  $N$ , but using the model trained on Monte Carlo simulations of  ${}^3\text{He}$  and  ${}^4\text{He}$ . The parameter  $\tau$  scales the number of background events from the signal region to the control region. As described in the Section 5.1, in this work, the control region relies only on Monte Carlo; therefore,  $\tau$  is related to the number of simulated events and the amount of available data. More specifically:

$$\tau = \frac{M}{\sum_{i=1}^M w_i} \quad (5.6)$$

where  $\sum_{i=1}^M w_i$  is the sum of the weights of the single MC events after the cut on the DNNs scores, lying within the mass window of  $\text{He}$  nuclei ( $m \in [2, 5] \text{ GeV}/c^2$ ).

Given the likelihood definition (Equation 5.5), the expressions of the unconditioned and conditioned Maximum Likelihood Estimator (MLE) are obtained by computing the first and second partial derivatives:  $\frac{\partial L}{\partial s}$ ,  $\frac{\partial L}{\partial b}$ ,  $\frac{\partial^2 L}{\partial s^2}$ , and  $\frac{\partial^2 L}{\partial s \partial b}$ . The unconditional MLE are:

$$\hat{s} = N + \frac{M}{\tau}, \hat{b} = \frac{M}{\tau} \quad (5.7)$$

whilst the conditioned MLE of  $b$ , assuming  $s$  fixed, is:

$$\hat{b} = \frac{N + M - (1 + \tau)s}{2(1 + \tau)} + \sqrt{\frac{(N + M - (1 + \tau)s)^2 + 4(1 + \tau)Ms}{4(1 + \tau)^2}} \quad (5.8)$$

Under the background-only hypothesis ( $s = 0$ ), the last expression becomes:  $\hat{b}(s = 0) = \frac{N+M}{1+\tau}$ .

It must be noted that, with this definition of  $\tau$ , the variance of the MLE  $\hat{b}$  is:  $\text{Var}[\hat{b}] = \frac{1}{\tau^2} \text{Var}[M] = \frac{(\sum_{i=1}^M w_i)^2}{M}$ . Clearly, increasing the size of the Monte Carlo reduces the variance of the nuisance parameter.

## 5.2.2 Implementation of the CLs method

In this analysis, the available statistic is small ( $\sim 10$  events), and the expected background contributions are of the same order of magnitude; therefore, no large sample assumption can be used, and the Probability Density Functions (pdfs) of the test statistic  $f(Q|B)$  and  $f(Q|S+B)$  have been obtained using Monte Carlo pseudo-experiments. Since the true value of the POI, the  $\overline{He}$  signal rate  $s$ , is unknown, the interval  $s \in [2, 22]$  with a step of 0.2 was scanned. Given a fixed value of  $s$ , the procedure followed to obtain the PDFs is:

- For each pair of WPs, the values of  $N^{OBS}$  and  $M^{OBS}$  are obtained and used to compute  $\hat{b}(0) = \frac{N^{OBS} + M^{OBS}}{1 + \tau}$ .
- Generate 50'000 pseudo-experiments under two different hypotheses. The PDF under the background hypothesis is obtained extracting a random  $n_i$  from  $\text{Pois}(\hat{b}(0))$ , a  $m_i$  from  $\text{Pois}(\tau \hat{b}(0))$ , and computing the test statistic  $Q$ . The PDF under the signal+background hypothesis is obtained extracting a random  $n_i$  from  $\text{Pois}(s + \hat{b}(0))$ , a  $m_i$  from  $\text{Pois}(\tau \hat{b}(0))$ , and computing the test statistic  $Q$ .

Afterwards, having determined the PDFs, it is possible to compute the observed value of  $CL_S$  for a given value of the POI.

The Figures 5.5 and 5.6 show three examples of the test statistic PDFs for the two hypotheses. All the plots refer to the (80%, 50%) WP pair, for events crossing the RICH-AGL, and are relative to three different values of the POI:  $s = 2.8$ ,  $s = 6.2$ , and  $s = 8.2$ . The green histograms correspond to the PDFs under the signal+background hypothesis, while the red histograms are the PDFs of the background-only hypothesis. The blue dashed vertical lines show the observed values, computed starting from the measurements of  $N^{OBS}$  and  $M^{OBS}$ . The green-filled area corresponds to the p-value  $CL_{S+B}$ ; the red area coincides with  $CL_B$ . As expected, the two PDFs have a considerable overlap, with the observed value slightly favouring the null hypothesis.

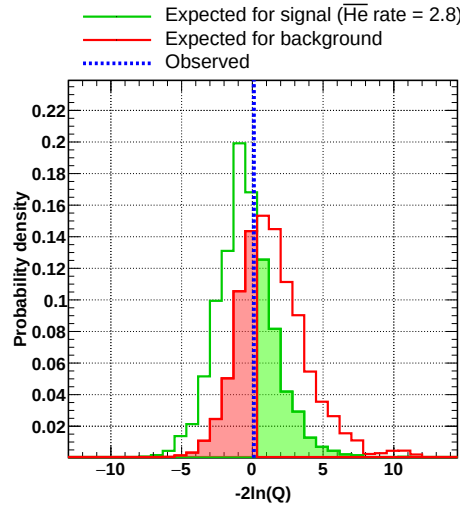


Figure 5.5: Distributions of the test statistic under the signal+background (green histogram) and background only (red histogram) hypotheses together with the observed values for a signal rate of 2.8. The two filled areas correspond to the  $CL_{S+B}$  (green) and  $CL_B$  (red) p-values. The plot refers to the measured values obtained with the (80%, 50%) WP pair, for events crossing the RICH-AGL ( $\epsilon_{AE} = 80\%$ ,  $\epsilon_{CL} = 50\%$ ,  $\epsilon_{SIG} = \epsilon_{AE} \times \epsilon_{CL} = 40\%$ ).

It is interesting to compute the outcome of the  $CL_S$  method for pseudo-experiments with only background, at a fixed value of the POI. Therefore, 50'000 pseudo-experiment have been generated extracting randomly a  $N_i^{EXP}$  from  $\text{Pois}(\hat{b}(0))$ , a  $M_i^{EXP}$  from  $\text{Pois}(\tau\hat{b}(0))$ , where the subscript  $i$  indicates the  $i$ -th pseudo-experiment. Then, for each couple  $(N_i^{EXP}, M_i^{EXP})$ , the corresponding  $CL_S$  value is computed:

$$CL_S^{EXP}(s=0) = \frac{CL_{S+B}^{EXP}}{1 - CL_B^{EXP}} \quad (5.9)$$

The distribution of  $CL_B^{EXP}$ , compared with the observed value of  $CL_b$ , gives an indication of how far out on the tail of the expected background the observed value is. The distribution of  $CL_S^{EXP}$  is particularly useful when the data presents fewer events than the number of expected background events, and the experiment observes an underfluctuation of the background. In these cases, the median expected limit is quoted, instead of the observed  $CL_S$  exclusion limit [138]. This is similar to the ‘sensitivity’ quoted in frequentist confidence intervals, which is the mean confidence limit

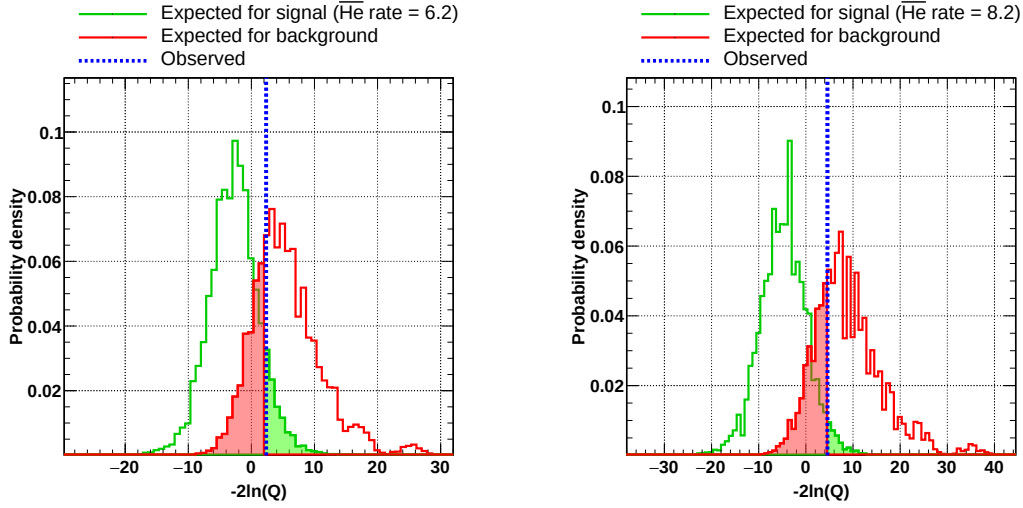


Figure 5.6: Distributions of the test statistic under the signal+background (green histograms) and background only (red histograms) hypotheses together with the observed values for a signal rate of 6.2 (left) and  $s = 8.2$  (right). The two filled areas correspond to the  $CL_{S+B}$  (green) and  $CL_B$  (red) p-values. The plots refer to the measured values obtained with the (80%, 50%) WP pair, for events crossing the RICH-AGL ( $\epsilon_{AE} = 80\%$ ,  $\epsilon_{CL} = 50\%$ ,  $\epsilon_{SIG} = \epsilon_{AE} \times \epsilon_{CL} = 40\%$ ).

expected for an ensemble of background experiments [140, 141].

### 5.3 Exclusion limits on the signal yield

After applying the candidate identification strategy described in Section 4.5, namely cutting on the DNNs outputs and on the reconstructed mass distribution, on MC and data samples with  $R < 0$ , the results reported in Table 5.1 are obtained. For each one of the three AMS subdetectors considered, i.e. TOF, RICH-NaF and RICH-AGL, and for each of the three pairs of the chosen WPs, the observed and expected values of  $CL_S$  and  $CL_B$  are obtained as a function of the signal rate, by applying the implementation of the  $CL_S$  method described in Section 5.2.

The Figures from 5.7 to 5.15 show the results obtained applying the  $CL_S$  method. The left plots show the values of  $CL_S$ , the right plots show the values of  $CL_B$  as a function of the POI, the signal rate. The blue line is the observed value, the red line is the expected value for an ensemble of background-only pseudoexperiments, the green and yellow bands correspond to the 68% ( $\pm 1\sigma$ ) and 95% ( $\pm 2\sigma$ ) quantiles of the ensemble distribution. The black dashed horizontal line in the left plots corresponds to 0.05, a value commonly chosen in literature to determine the observed (expected) exclusion limit [138]. The horizontal lines in the right plots correspond to the conversion of  $CL_b$  into standard deviations through a one-sided Gaussian approximation, where  $CL_B = 2.27 \times 10^{-2}$  indicates  $2\sigma$ ,  $CL_B = 1.35 \times 10^{-3}$  corresponds to  $3\sigma$ , and  $CL_B = 2.8 \times 10^{-7}$  would indicate a  $5\sigma$  discovery.

### 5.3.1 TOF

The TOF is used within the following rigidity interval:  $[1.92, 6.47]$  GV, corresponding to the first two bins in Table 4.3. As reported in Table 5.1, for all the WP pairs selected, no antihelium candidates have been found in AMS data with a reconstructed mass within the range  $m_{rec} \in [2, 5]$  GeV/c<sup>2</sup>. Given the absence of observations, the expected exclusion limits are used as results of the search. The exclusion intervals obtained from the left plots of the Figures 5.7, 5.8, and 5.9 are all equal to  $s = 4.50$ , as also reported in Table 5.4. Therefore, this work excludes with the 95% confidence the presence of an antihelium signal rate higher than 4.5 for reconstructed rigidities within  $[-6.47, -1.92]$ .

| WPs( $\epsilon_{AE}$ , $\epsilon_{CL}$ ) | ISS data ev. | MC weighted ev. | MC ev. | Exclusion limit (exp.) |
|------------------------------------------|--------------|-----------------|--------|------------------------|
| (95%, 50%)                               | 0            | 2.80            | 5      | 4.5                    |
| (80%, 50%)                               | 0            | 2.43            | 4      | 4.5                    |
| (75%, 95%)                               | 0            | 2.02            | 3      | 4.5                    |

Table 5.3: Expected exclusion limits for events crossing the TOF and having a reconstructed rigidity within  $[-6.47, -1.92]$  GV.

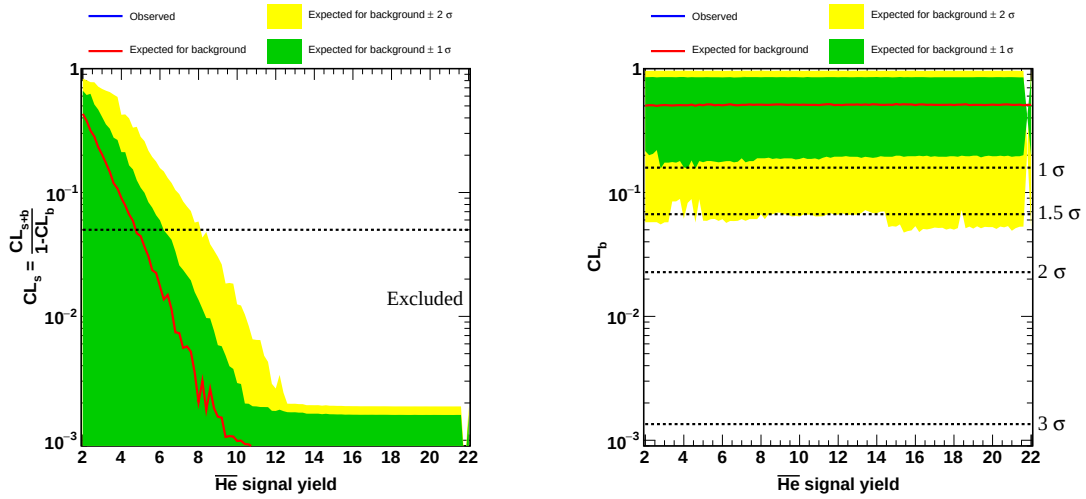


Figure 5.7: (Left) Values of  $CL_s$  (see equation 5.4) as a function of the antihelium signal rate. (Right) Values of  $CL_b$  (see equation 5.2) as a function of the signal rate. The results are obtained for events with  $R \in [-6.47, -1.92]$ , ( $\epsilon_{AE} = 95\%$ ,  $\epsilon_{CL} = 50\%$ ), with a TOF velocity measurement. Since no events have been observed in ISS data, the blue lines corresponding to the observed  $CL_s$  and  $CL_b$  are not displayed.

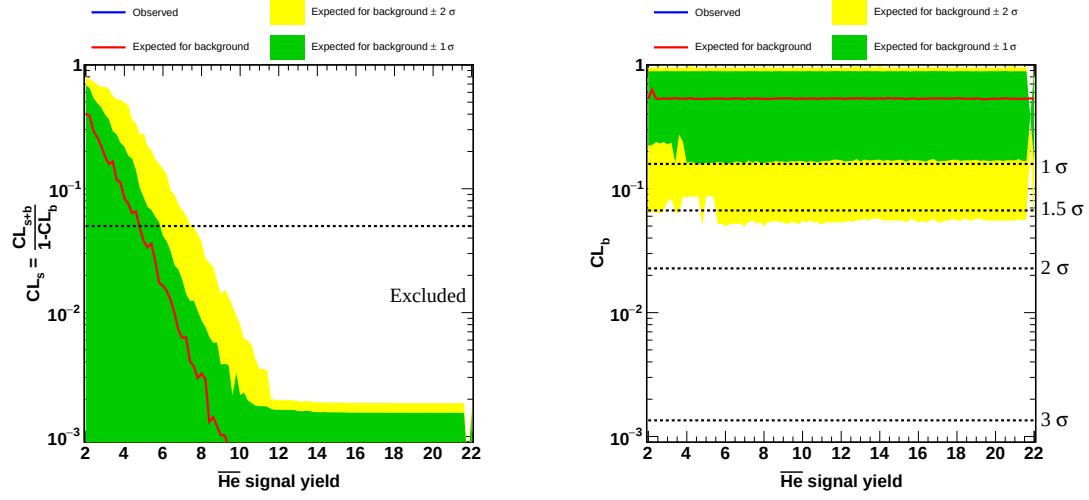


Figure 5.8: (Left) Values of  $CL_s$  (see equation 5.4) as a function of the antihelium signal rate. (Right) Values of  $CL_b$  (see equation 5.2) as a function of the signal rate. The results are obtained for events with  $R \in [-6.47, -1.92]$ , ( $\varepsilon_{AE} = 80\%$ ,  $\varepsilon_{CL} = 50\%$ ), with a TOF velocity measurement. Since no events have been observed in ISS data, the blue lines corresponding to the observed  $CL_s$  and  $CL_b$  are not displayed.

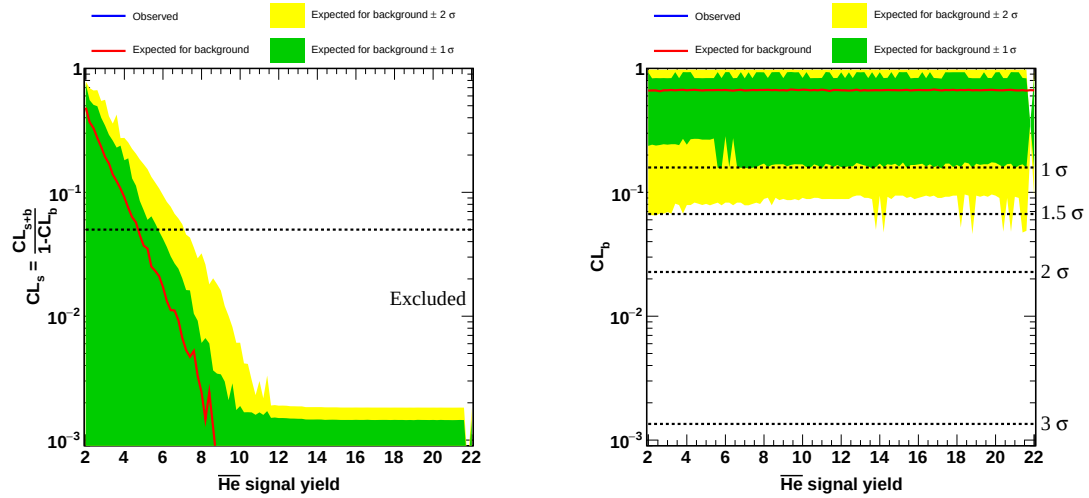


Figure 5.9: (Left) Values of  $CL_s$  (see equation 5.4) as a function of the antihelium signal rate. (Right) Values of  $CL_b$  (see equation 5.2) as a function of the signal rate. The results are obtained for events with  $R \in [-6.47, -1.92]$ , ( $\varepsilon_{AE} = 75\%$ ,  $\varepsilon_{CL} = 95\%$ ), with a TOF velocity measurement. Since no events have been observed in ISS data, the blue lines corresponding to the observed  $CL_s$  and  $CL_b$  are not displayed.

### 5.3.2 RICH-NaF

The RICH-NaF provides a reliable velocity measurement within the following rigidity interval:  $[4.02, 18.0]$  GV, corresponding from the 2<sup>nd</sup> to the 5<sup>th</sup> rigidity bin in Table 4.3. As reported in Table 5.1, for all the WP pairs selected, the number of antihelium candidates, with a reconstructed mass within the range  $m_{rec} \in [2, 5]$  GeV/c<sup>2</sup>, is slightly lower than the expected charged-confused helium nuclei background from MC. No excess has been found compared to the expected background from MC, and the background-only hypothesis describes the observations. For this reason, the expected exclusion limits are used as results of the search and are reported in Table 5.4:

| WPs( $\epsilon_{AE}$ , $\epsilon_{CL}$ ) | ISS data events | MC weighted events | MC events | Exclusion limit (exp.) |
|------------------------------------------|-----------------|--------------------|-----------|------------------------|
| (95%, 50%)                               | 2               | 2.39               | 5         | 4.80                   |
| (80%, 50%)                               | 1               | 2.39               | 5         | 4.54                   |
| (75%, 95%)                               | 2               | 2.93               | 6         | 4.59                   |

Table 5.4: Expected exclusion limits for events crossing the RICH-NaF and having a reconstructed rigidity within  $[-18.0, -4.02]$  GV.

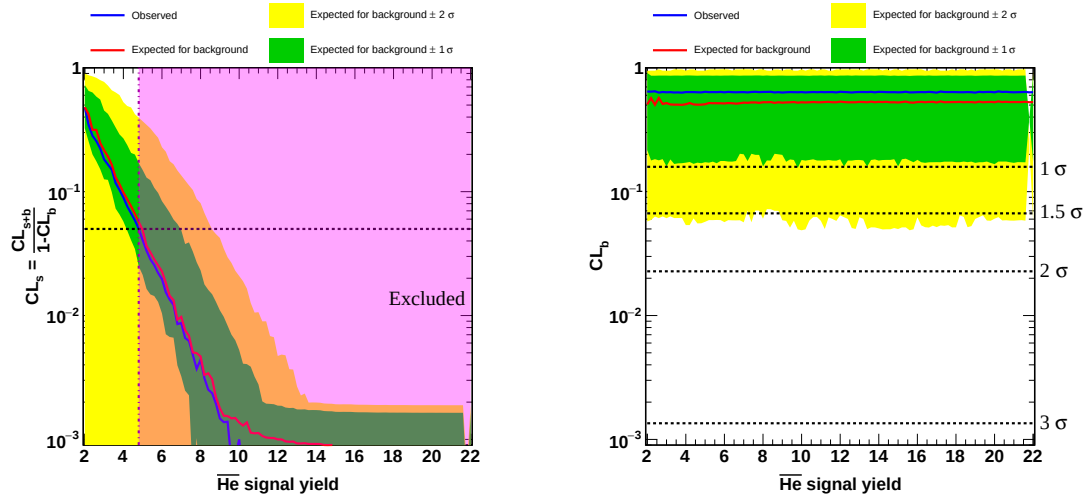


Figure 5.10: (Left) Values of  $CL_s$  (see equation 5.4) as a function of the antihelium signal rate. (Right) Values of  $CL_b$  (see equation 5.2) as a function of the signal rate. The results are obtained for events with  $R \in [-18.0, -4.02]$ , ( $\epsilon_{AE} = 95\%$ ,  $\epsilon_{CL} = 50\%$ ), with a RICH-NaF velocity measurement.

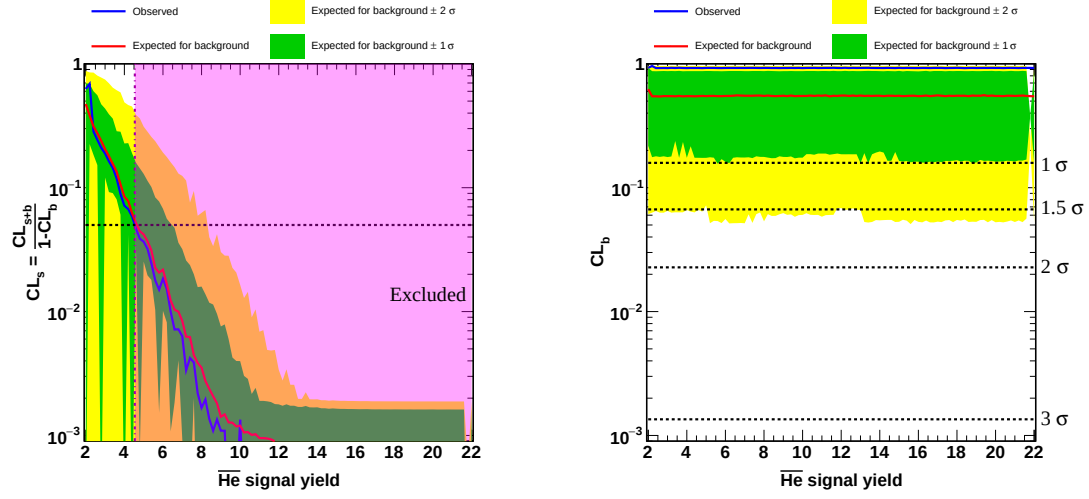


Figure 5.11: (Left) Values of  $CL_s$  (see equation 5.4) as a function of the antihelium signal rate. (Right) Values of  $CL_b$  (see equation 5.2) as a function of the signal rate. The results are obtained for events with  $R \in [-18.0, -4.02]$ , ( $\epsilon_{AE} = 80\%$ ,  $\epsilon_{CL} = 50\%$ ), with a RICH-NaF velocity measurement.

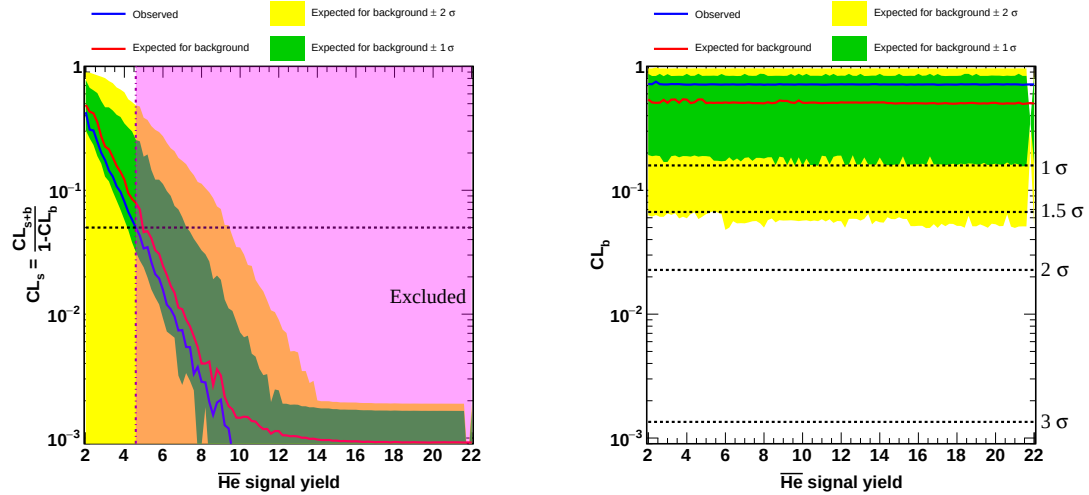


Figure 5.12: (Left) Values of  $CL_s$  (see equation 5.4) as a function of the antihelium signal rate. (Right) Values of  $CL_b$  (see equation 5.2) as a function of the signal rate. The results are obtained for events with  $R \in [-18.0, -4.02]$ , ( $\epsilon_{AE} = 75\%$ ,  $\epsilon_{CL} = 95\%$ ), with a RICH-NaF velocity measurement.



### 5.3.3 RICH-AGL

The rigidity range considered for the RICH-AGL corresponds to the following rigidity interval:  $[6.47, 48.5]$  GV. As reported in Table 5.1, for two WP pairs, namely  $(95\%, 50\%)$  and  $(80\%, 50\%)$ , a small excess of antihelium candidates with a reconstructed mass within the range  $m_{rec} \in [2, 5]$  GeV/ $c^2$  is observed in data compared with the MC expected value; on the other hand, the  $(75\%, 95\%)$  WPs present a lack of data events with respect to the MC expectations. For these reasons, the observed exclusion limits are quoted for the first two pairs, whilst the expected exclusion limit is reported for the third pair. The observed exclusion limit obtained with the  $(75\%, 95\%)$  pair follows the lower  $2\sigma$  band for the background-only pseudoexperiments, which might hint at a slight overestimate of the expected background for this case. The highest signal significance is slightly lower than  $1\sigma$  and is observed using the  $(95\%, 50\%)$  pair (right plot of Figure 5.13). In conclusion, the background-only hypothesis is in good agreement with the AMS-02 data. The observed and expected exclusion limits for the three pairs are reported in Table 5.5:

| WPs( $\epsilon_{AE}$ , $\epsilon_{CL}$ ) | ISS data ev. | MC weighted ev. | MC ev. | Exclusion limit |
|------------------------------------------|--------------|-----------------|--------|-----------------|
| (95%, 50%)                               | 12           | 8.65            | 21     | 11.60 (obs.)    |
| (80%, 50%)                               | 7            | 5.49            | 15     | 8.21 (obs.)     |
| (75%, 95%)                               | 17           | 27.44           | 71     | 12.80 (exp.)    |

Table 5.5: Expected exclusion limits for events crossing the RICH-AGL and having a reconstructed rigidity within  $[-48.5, -6.47]$  GV.

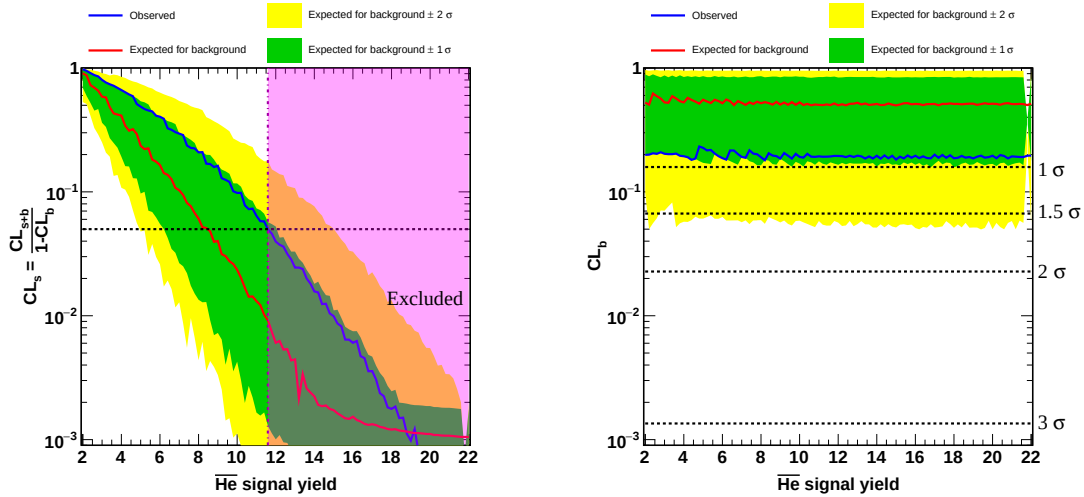


Figure 5.13: (Left) Values of  $CL_s$  (see equation 5.4) as a function of the antihelium signal rate. (Right) Values of  $CL_b$  (see equation 5.2) as a function of the signal rate. The results are obtained for events with  $R \in [-48.5, -6.47]$ ,  $(\epsilon_{AE} = 95\%, \epsilon_{CL} = 50\%)$ , with a RICH-AGL velocity measurement.

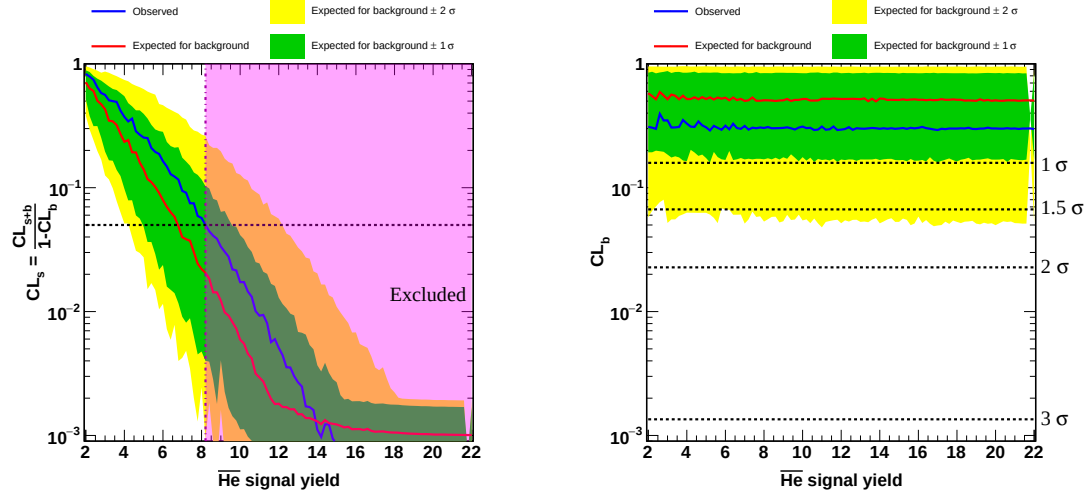


Figure 5.14: (Left) Values of  $CL_s$  (see equation 5.4) as a function of the antihelium signal rate. (Right) Values of  $CL_B$  (see equation 5.2) as a function of the signal rate. The results are obtained for events with  $R \in [-48.5, -6.47]$ ,  $(\varepsilon_{AE} = 80\%, \varepsilon_{CL} = 50\%)$ , with a RICH-AGL velocity measurement.

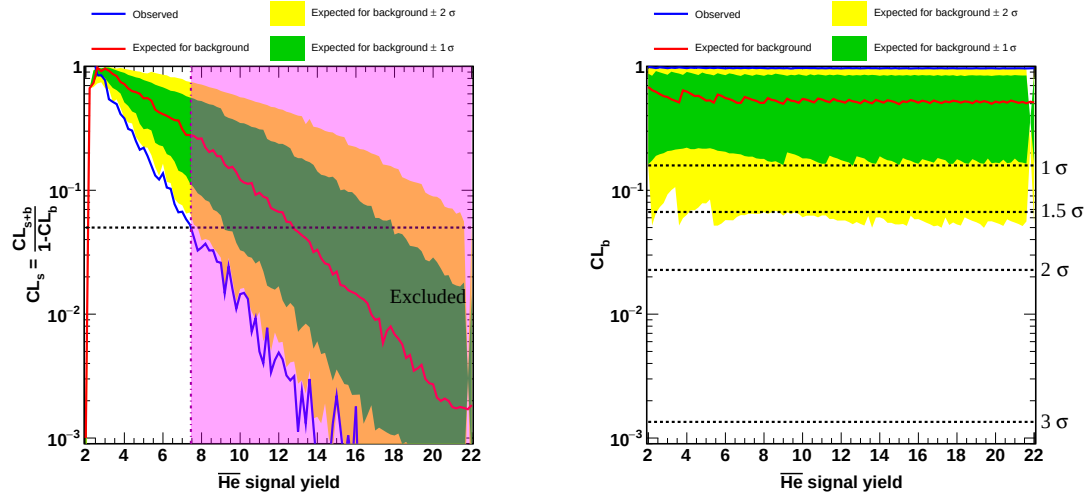


Figure 5.15: (Left) Values of  $CL_s$  (see equation 5.4) as a function of the antihelium signal rate. (Right) Values of  $CL_B$  (see equation 5.2) as a function of the signal rate. The results are obtained for events with  $R \in [-48.5, -6.47]$ ,  $(\varepsilon_{AE} = 75\%, \varepsilon_{CL} = 95\%)$ , with a RICH-AGL velocity measurement.

As previously described in 4.1.1, the support structure of the inner tracker presents some dense material spots corresponding to aluminium support jig blocks of ladders (tracker feet). The tracker feet can be a source of charge-confused background since helium nuclei interacting with them can mimic a true antihelium event. Therefore, it is interesting to add a cut on the minimum distance to the aluminium support jigs as an additional cross-check on the antihelium candidates found.

The jigs can be schematised as squares with a side of  $\sim 0.5$  cm. Requiring a minimum distance from the jigs of 0.6 cm and lowering the threshold on the normalised  $\chi_Y^2$  of the track along the bending direction from 10 to 8 (see the left plot of Figure 3.3), the number of surviving events for the first two WP pairs changes as shown in Table 5.6.

The  $CL_s$  and  $CL_B$  values as a function of the signal rate are presented in the Figures 5.16 and 5.17. The additional cuts enhance the significance of the signal+background hypothesis from  $\sim 1\sigma$  to  $\sim 2\sigma$  for the (95%, 50%) pair and from  $\sim 0.5\sigma$  to  $\sim 1.4\sigma$  for the (80%, 50%) WPs, showing a mild tension with the background-only hypothesis.

| WPs( $\varepsilon_{AE}$ , $\varepsilon_{CL}$ ) | ISS data ev. | MC weighted ev. | MC ev. | Exclusion limit (obs.) |
|------------------------------------------------|--------------|-----------------|--------|------------------------|
| (95%, 50%)                                     | 10           | 4.34            | 11     | 13.01                  |
| (80%, 50%)                                     | 6            | 2.75            | 8      | 9.13                   |

Table 5.6: Expected exclusion limits for events crossing the RICH-AGL, having a reconstructed rigidity within  $[-48.5, -6.47]$  GV, and with a minimum distance from the inner support jigs of 0.6 cm.

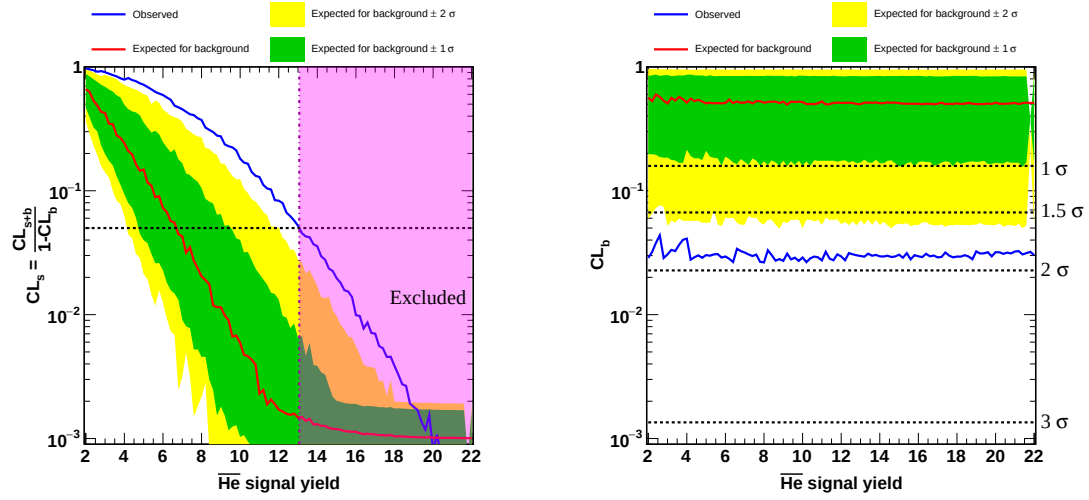


Figure 5.16: (Left) Values of  $CL_s$  (see equation 5.4) as a function of the antihelium signal rate. (Right) Values of  $CL_B$  (see equation 5.2) as a function of the signal rate. The results are obtained for events with  $R \in [-48.5, -6.47]$ , ( $\epsilon_{AE} = 95\%$ ,  $\epsilon_{CL} = 50\%$ ), with a RICH-AGL velocity measurement, and a minimum distance from the inner tracker aluminium support jigs of 0.6 cm.

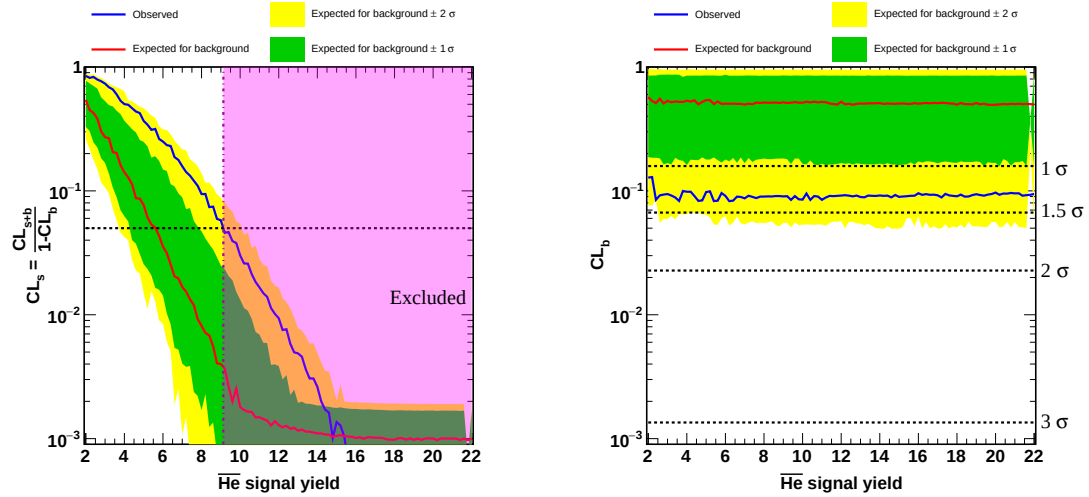


Figure 5.17: (Left) Values of  $CL_s$  (see equation 5.4) as a function of the antihelium signal rate. (Right) Values of  $CL_B$  (see equation 5.2) as a function of the signal rate. The results are obtained for events with  $R \in [-48.5, -6.47]$ , ( $\epsilon_{AE} = 75\%$ ,  $\epsilon_{CL} = 95\%$ ), with a RICH-AGL velocity measurement, and a minimum distance from the inner tracker aluminium support jigs of 0.6 cm.

## 5.4 Exclusion limits on the ratio anti-He/He

The exclusion limits presented in the Tables 5.3, 5.4, and 5.5 have been used to obtain exclusion limits on the ratio  $\frac{\overline{He}}{He}$ . The denominator is obtained by applying the same procedure described in Section 4.5, but considering only well-reconstructed helium nuclei ( $R_{inner} > 0$ ) and with a measured rigidity higher than the maximum rigidity cutoff in the solid angle of AMS. Following the reasons presented in Section 5.1, the (75%, 95%) combination is not used to compute the exclusion limit on the antihelium-to-helium ratio. The results for the subdetectors considered in this analysis are reported in Table 5.7.

| Det.                   | WPs( $\epsilon_{AE}$ , $\epsilon_{CL}$ ) | ISS data | MC weights | MC ev. | Exc. lim.    | Ex. lim. $\frac{\overline{He}}{He}$ |
|------------------------|------------------------------------------|----------|------------|--------|--------------|-------------------------------------|
| TOF                    | (95%, 50%)                               | 0        | 2.80       | 5      | 4.5 (exp.)   | $0.55 \times 10^{-8}$               |
|                        | (80%, 50%)                               | 0        | 2.43       | 4      | 4.5 (exp.)   | $0.64 \times 10^{-8}$               |
| RICH-NaF               | (95%, 50%)                               | 2        | 2.39       | 5      | 4.80 (exp.)  | $4.12 \times 10^{-8}$               |
|                        | (80%, 50%)                               | 1        | 2.39       | 5      | 4.54 (exp.)  | $4.59 \times 10^{-8}$               |
| RICH-AGL<br>+ TRK feet | (95%, 50%)                               | 12       | 8.65       | 21     | 11.60 (obs.) | $1.62 \times 10^{-8}$               |
|                        | (80%, 50%)                               | 7        | 5.49       | 15     | 8.21 (obs.)  | $1.33 \times 10^{-8}$               |
|                        | (95%, 50%)                               | 10       | 4.34       | 11     | 13.07 (obs.) | $2.14 \times 10^{-8}$               |
|                        | (80%, 50%)                               | 6        | 2.75       | 8      | 9.13 (obs.)  | $1.74 \times 10^{-8}$               |

Table 5.7: Results of the antihelium search for three AMS subdetectors: TOF, RICH-NaF and RICH-AGL. The table presents the number of surviving events in AMS-02 data,  ${}^3He$  and  ${}^4He$  MC (with and without weights), the exclusion limits (expected or observed) on the signal rate and the ratio  $\frac{\overline{He}}{He}$ , obtained applying the  $CL_s$  method for two pairs of WPs. The two additional rows for RICH-AGL show the results obtained applying the additional cuts on the minimum distance from the TRK aluminium jigs and on the  $\chi^2_Y$  as described in Section 5.3.3.

The exclusion limits on the ratio  $\frac{\overline{He}}{He}$  are displayed as a function of the kinetic energy per nucleon. Here only the limits obtained using the (95%, 50%) pair is reported, where Figure 5.18 shows the exclusion limit on  $\frac{{}^3\overline{He}}{He}$ , while the Figure 5.19 shows the exclusion limit on  $\frac{{}^4\overline{He}}{He}$ . The two plots are obtained assuming that all the candidates are respectively  ${}^3\overline{He}$  or  ${}^4\overline{He}$ . All the other results are presented in Appendix D.

The exclusion limits from this work are presented with two previous results from the AMS-01 and BESS-Polar II collaborations, yielding respectively:  $\frac{\overline{He}}{He} = 1.1 \times 10^{-6}$  and  $\frac{\overline{He}}{He} = 9.4 \times 10^{-8}$  [144, 79]. The exclusion limits for RICH-AGL and TOF obtained with the two pairs of WP ((95%, 50%) and (80%, 50%)) are between 4 and 17 times smaller than the result from BESS-Polar II. The limits obtained considering the RICH-NaF are lower by a factor of 2 compared with the best previous result.

When the number of observed events in the data is lower than the number of expected background events, the expected exclusion limit is quoted. The expected exclusion limit corresponds to the median exclusion limit obtained under the assumption of a background-only experiment (similar to the sensitivity quoted in frequentist confidence interval), as previously described in Section 5.2.2. On the other hand, when the number of expected events in data is higher than the prevision from MC the observed limit is reported. In each plot, the continuous blue lines show

the observed limits, while the dashed lines represent the expected limits.

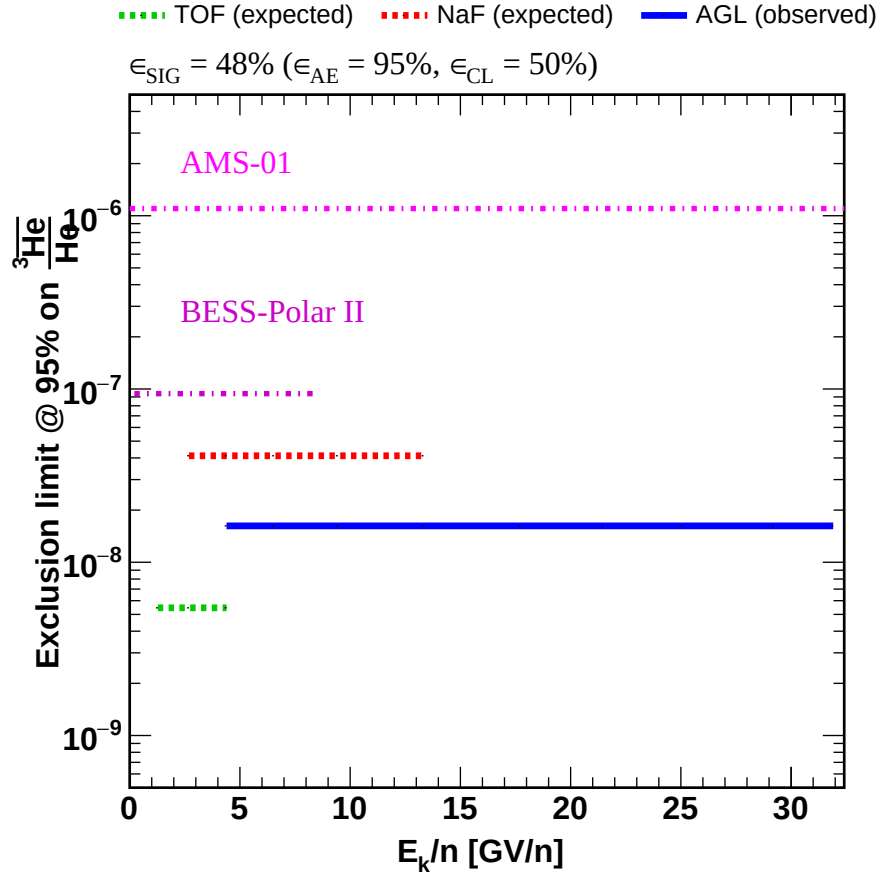


Figure 5.18:  $\frac{^3\overline{\text{He}}}{\text{He}}$  exclusion limit, assuming all the antihelium candidates are  $^3\overline{\text{He}}$ , for the combination ( $\epsilon_{\text{AE}} = 95\%$ ,  $\epsilon_{\text{CL}} = 50\%$ ) ( $\epsilon_{\text{SIG}} = \epsilon_{\text{AE}} \times \epsilon_{\text{CL}} = 48\%$ ) as a function of the kinetic energy per nucleon. The results from this work are presented together with the previous limits from AMS-01 and BESS-Polar II [79, 144].

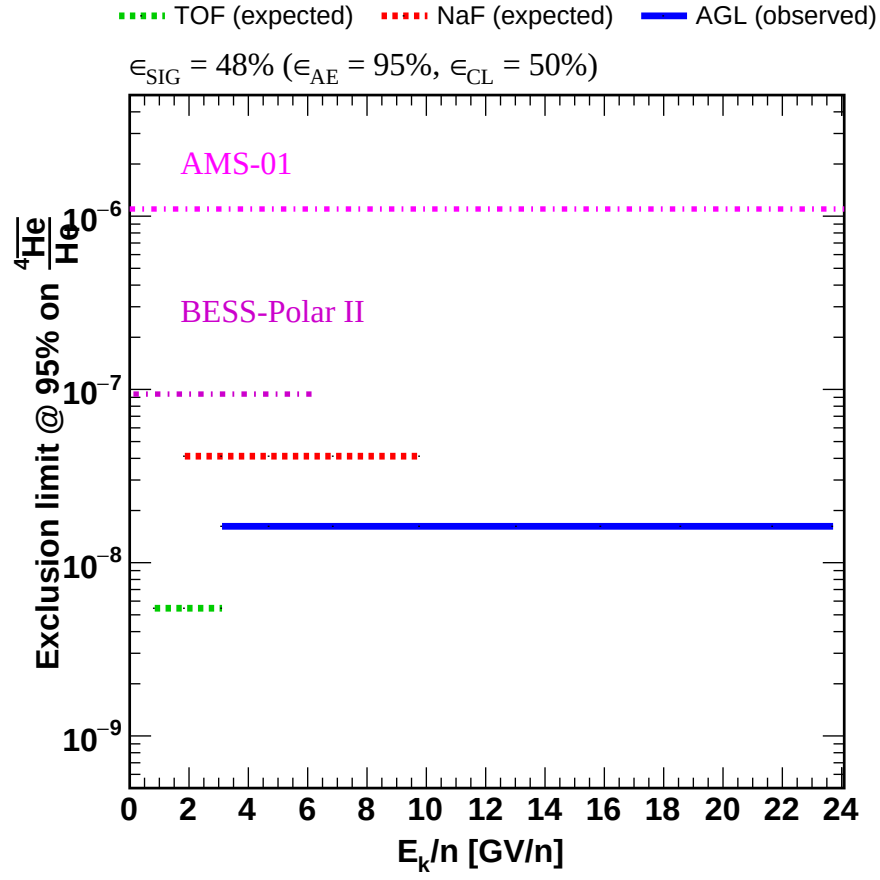


Figure 5.19:  $\frac{^4\overline{\text{He}}}{^4\text{He}}$  exclusion limit, assuming all the antihelium candidates are  $^4\overline{\text{He}}$ , for the combination ( $\epsilon_{\text{AE}} = 95\%$ ,  $\epsilon_{\text{CL}} = 50\%$ ) ( $\epsilon_{\text{SIG}} = \epsilon_{\text{AE}} \times \epsilon_{\text{CL}} = 48\%$ ) as a function of the kinetic energy per nucleon. The results from this work are presented together with the previous limits from AMS-01 and BESS-Polar II [79, 144].

## 5.5 Candidate events

The most interesting antihelium candidates found in this work are those obtained for the (95%, 50%) WPs, crossing the RICH-AGL, that survive the cut on the minimum distance from the tracker feet and have a  $\chi_Y^2 < 8$ . All the selected events have a single reconstructed track, and some of their main characteristics are reported in Table 5.8. Where  $m$  is the reconstructed mass,  $\sigma_m/m$  is the relative error on the mass measurement,  $R_{IN}$  is the measured rigidity using the inner tracker span, IGRF +/- are the maximum IGRF cutoff within the  $40^\circ$  field of view of AMS-02 for positive and negative particles [115],  $\chi_X^2$  and  $\chi_Y^2$  are the track normalised  $\chi^2$  for the non-bending and the bending direction,  $Z_{TRK}$  and RMS are the charge measured by the inner tracker and its root mean squared error,  $\beta_{AGL}$ , and  $\beta_{TOF}$  are the velocity measurements of the RICH-AGL and TOF.

| N. | $m$ [GeV/c <sup>2</sup> ] | $\sigma_m/m$ | $R$ [GV] | IGRF +/- [GV] | $\chi_X^2$ | $\chi_Y^2$ | $Z_{TRK}$ | RMS  | $\beta_{AGL}$ | $\beta_{TOF}$ |
|----|---------------------------|--------------|----------|---------------|------------|------------|-----------|------|---------------|---------------|
| 1  | 3.24                      | 11%          | -11.3    | 3.83/ - 4.1   | 5.0        | 7.8        | 2.20      | 0.18 | 0.9899        | 0.97          |
| 2  | 4.69                      | 27%          | -42.3    | 12.8/ - 17.5  | 9.5        | 5.7        | 2.00      | 0.10 | 0.9985        | 1.02          |
| 3  | 4.54                      | 12%          | -23.0    | 23.3/ - 18.8  | 7.8        | 0.4        | 2.20      | 0.13 | 0.9952        | 0.97          |
| 4  | 2.53                      | 17%          | -18.6    | 15.0/ - 21.8  | 9.3        | 3.9        | 2.08      | 0.11 | 0.9977        | 0.99          |
| 5  | 4.50                      | 20%          | -31.9    | 4.5/ - 4.3    | 7.2        | 1.1        | 2.04      | 0.21 | 0.9975        | 0.98          |
| 6  | 2.11                      | 27%          | -20.6    | 20.0/ - 25.0  | 2.7        | 3.6        | 2.06      | 0.20 | 0.9987        | 1.02          |
| 7  | 2.06                      | 15%          | -13.0    | 14.2/ - 19.9  | 9.9        | 1.8        | 2.04      | 0.13 | 0.9969        | 0.95          |
| 8  | 2.73                      | 12%          | -14.0    | 18.3/ - 23.9  | 8.6        | 6.4        | 2.01      | 0.12 | 0.9953        | 0.97          |
| 9  | 3.11                      | 13%          | -12.6    | 4.3/ - 3.9    | 2.0        | 7.6        | 2.07      | 0.08 | 0.9925        | 0.98          |
| 10 | 4.38                      | 11%          | -15.7    | 3.8/ - 3.7    | 2.4        | 7.9        | 1.96      | 0.16 | 0.9904        | 1.00          |

Table 5.8: Antihelium candidates crossing the RICH-AGL for the (95%, 50%) WPs, with  $\chi_Y^2 < 8$ , and after the cut on the minimum distance from the inner tracker aluminium jigs.

The candidates 1, 2, 5, 9, and 10 have all a measured rigidity exceeding the maximum IGRF cutoffs for both positive and negative particles, indicating a galactic origin. For the remaining events, the exact calculation of the geomagnetic cutoff in the particle direction is needed to verify whether they are trapped particles or cosmic events.

The antihelium candidates 1, 9, and 10 have a  $\chi_Y^2$  close to the cut value, the events 2, 4, 7 and 8 have  $\chi_X^2$  values close to the chosen threshold, which might hint at large-angle scattering interactions.

All the candidates have a measured inner tracker charge higher than 1.9; the events 1 and 3 have a charge measurement slightly higher than the other candidates. The antihelium candidate number 7 has a  $\beta_{TOF}$  too low compared to the reconstructed rigidity and the RICH-AGL measurement.

Interestingly, the candidates 3, 6, and 8 have a measured rigidity in the inner-L1 pattern, which are in good agreement with the rigidities measured using the inner pattern, yielding respectively -23.1, -22.4, and -14.1 [GV]. Using the inner-L1 rigidity to reconstruct the mass provides the same relative errors but slightly higher masses. Indeed, the mass measurements are: 4.55, 2.29, and 2.76 [GeV/c<sup>2</sup>]. The 6<sup>th</sup> candidate has the largest discrepancy between the two mass measurements, although the values remain compatible.

The mass distribution of the antihelium candidates presented in Table 5.8, along with the cor-



responding mass distribution of the surviving events from MC, is presented in Figure 5.20. As shown in Table 5.6, the MC predicts  $4.34 \pm 1.41$  events, whilst 10 antihelium candidates are observed in AMS-02 data. The black histogram shows the mass distribution for helium nuclei obtained from AMS-02 data and is used as a benchmark for the red ( $\overline{\text{He}}$  candidates) and the blue (MC background events) histograms.

The mass values as a function of the measured rigidity are shown in Figure 5.21. All the candidates have reconstructed masses compatible with  ${}^3\text{He}$  or  ${}^4\text{He}$  within 1 or  $2\sigma$ .

The event displays of one antihelium candidate found by this analysis are presented in Figures 5.22 and 5.23.

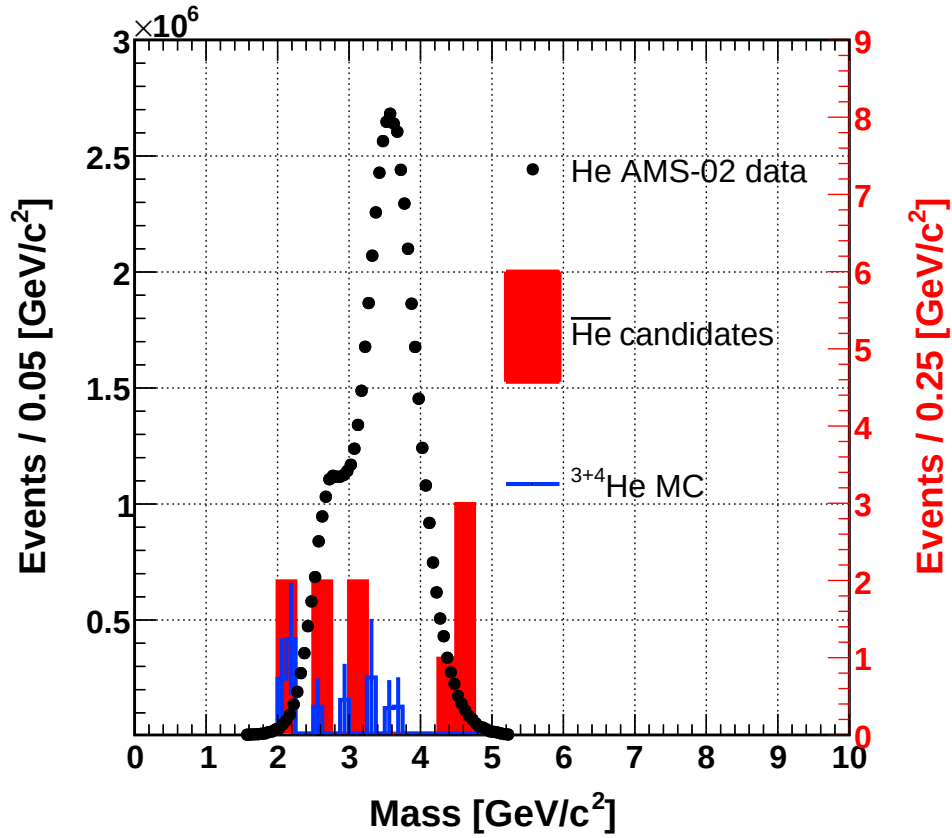


Figure 5.20: Mass distributions of the events selected using the WPs (95%, 50%), after the cut on the reconstructed mass,  $\chi_Y^2 < 8$  and the minimum distance from the tracker feet. The red histogram represents the antihelium candidates; the blue histogram represents the expected background from MC; the black histogram is used as a benchmark and is the mass distribution for ordinary helium nuclei from AMS-02 data.

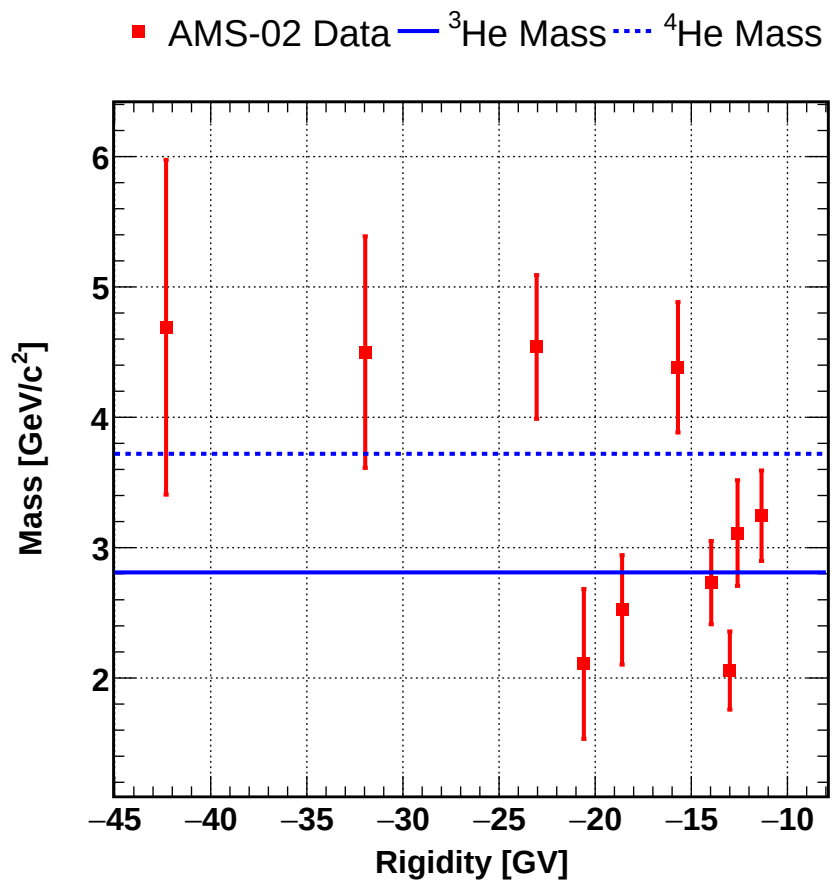


Figure 5.21: Mass of the antihelium candidates as a function of the measured rigidity. The events displayed are the same as those shown in Table 5.8.

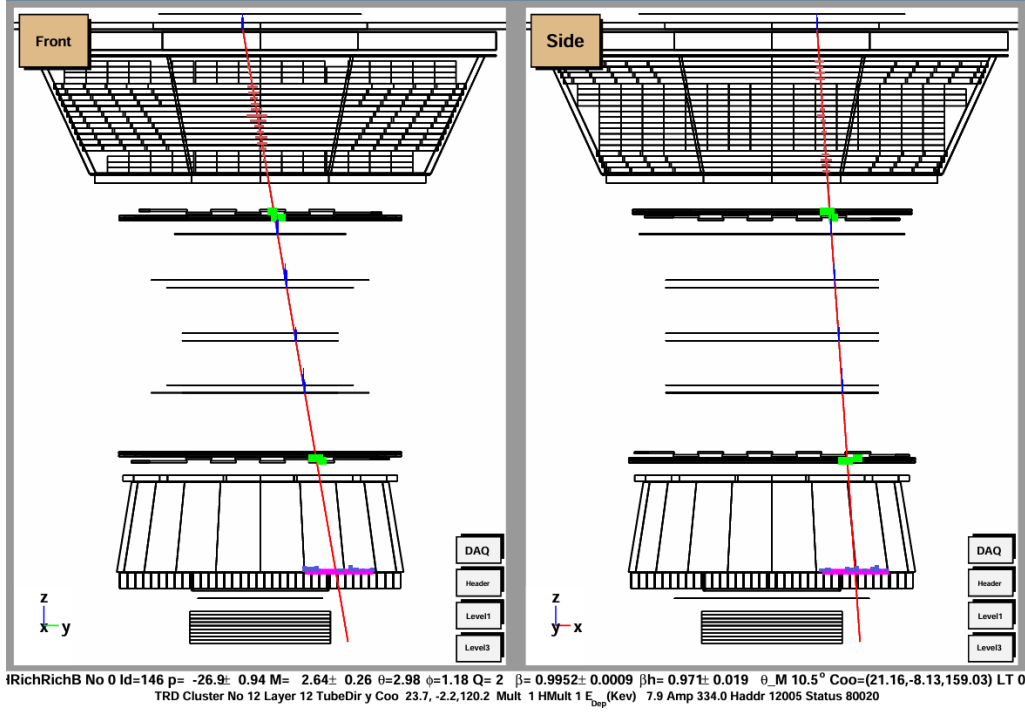


Figure 5.22: AMS-02 side views of the anithelium candidate number 8 (see Table 5.8). The binding plane (Z-Y) is shown in the left plot, the right side shows the non-bending plane (Z-X).

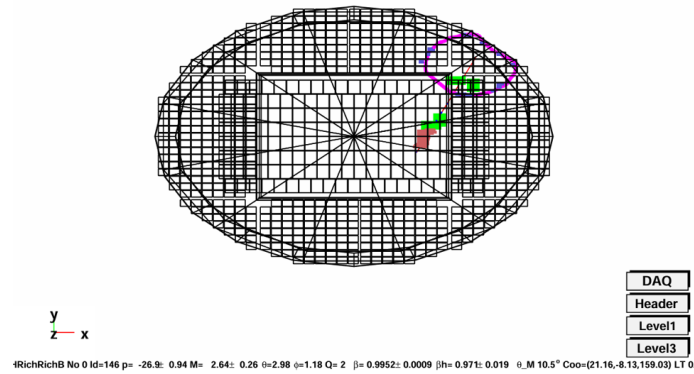


Figure 5.23: AMS-02 top view of the anithelium candidate number 8 (see Table 5.8).

## Chapter 6

# Conclusions

This Thesis presented a search for  $\overline{He}$  in CRs by combining two ML techniques, using 12.5 years of data collected by the AMS-02 experiment. The first-time discovery of cosmic-ray  $\overline{He}$  has been identified as a potentially promising detection signature related to DM interactions or to a primordial origin, due to its extremely low astrophysical backgrounds, especially towards low energies. The expected signal yield is very small, with a rough estimate of the  $\overline{He}$  to  $He$  production ratio of  $\sim 10^{-9}$ , Machine Learning algorithms have been used to suppress the background and uncover eventual signal candidates. Machine Learning techniques based on supervised and unsupervised methods improve background suppression by exploiting correlations among variables, thereby preserving high efficiency for potential signal candidates. Four different Deep Neural Networks (DNNs), two supervised classifiers and two unsupervised autoencoders have been developed to search for possible  $\overline{He}$  candidates in data.

This work applied a fully data-driven approach, applying the same strategy to data and Monte Carlo simulations independently. Initially, an event selection for events with  $|Z| = 2$  has been applied to obtain four different samples: two with positive rigidities, i.e.  $Z = +2$  (one for ISS data and one for MC), and two with negative rigidities corresponding to  $Z = -2$  (one for ISS and one for MC). The two  $R > 0$  samples have been used to train the Machine Learning algorithms. The models trained on MC are not applied to data and vice versa, avoiding possible MC-data disagreements arising from, for example, small discrepancies in the detector material budget description that might be relevant for rare interactions occurring inside the detector. Due to resonant background processes with reconstructed masses indistinguishable from those of the signal candidates, background estimation can only be performed using Monte Carlo simulations. The outputs of the DNNs have been proven to be negligibly correlated; therefore, under the assumption of independence and orthogonality, the scores have been combined to define rigidity-dependent signal and control regions. Due to the finite MC statistic, all regions have been summed, returning two rigidity-integrated count numbers, one for ISS data and one for MC simulation. In this way, the estimate of the expected background is improved.

Since no information on the velocity of the particle is provided as input to the DNNs, after the

selection on the DNN outputs, the mass measurement of the surviving events is used to constrain possible antihelium candidates requiring a reconstructed mass within  $m \in [2, 5] \text{ GeV}/c^2$ . Three different subdetectors of AMS-02 have been considered (TOF, RICH-NaF and RICH-AGL), covering an interval of kinetic energy per nucleon of  $E_K/n \in [0.65, 31.4] \text{ GeV}/N$  for  $\overline{^3\text{He}}$  and  $E_K/n \in [0.41, 23.3] \text{ GeV}/N$  for  $\overline{^4\text{He}}$ . The search for  $\overline{He}$  is performed by looking for an excess in the number of events with a reconstructed mass compatible with  $^3\text{He}$  and  $^4\text{He}$  masses. The results obtained for two different pairs of working points have been presented; up to 12  $\overline{He}$  candidates have been found, resulting in a mild tension with the background-only hypothesis at the level of  $1\text{-}2\sigma$ , depending on the chosen working points on the DNNs outputs. No statistically significant excess has been found; therefore, two exclusion limits using the CLs method on the  $\frac{\overline{^3\text{He}}}{\overline{He}}$  and  $\frac{\overline{^4\text{He}}}{\overline{He}}$  ratios have been obtained for each subdetector and each pair of chosen working points. The exclusion limits obtained with the two pairs of WP ((95%, 50%) and (80%, 50%)) are between 2 and 17 times smaller than the best previously known limit. The results are reported in Table 6.1:

| Det.                   | WPs( $\epsilon_{AE}$ , $\epsilon_{CL}$ ) | ISS data | MC weights | MC ev. | Exc. lim.    | Ex. lim. $\frac{\overline{He}}{He}$ |
|------------------------|------------------------------------------|----------|------------|--------|--------------|-------------------------------------|
| TOF                    | (95%, 50%)                               | 0        | 2.80       | 5      | 4.5 (exp.)   | $0.55 \times 10^{-8}$               |
|                        | (80%, 50%)                               | 0        | 2.43       | 4      | 4.5 (exp.)   | $0.64 \times 10^{-8}$               |
| RICH-NaF               | (95%, 50%)                               | 2        | 2.39       | 5      | 4.80 (exp.)  | $4.12 \times 10^{-8}$               |
|                        | (80%, 50%)                               | 1        | 2.39       | 5      | 4.54 (exp.)  | $4.59 \times 10^{-8}$               |
| RICH-AGL<br>+ TRK feet | (95%, 50%)                               | 12       | 8.65       | 21     | 11.60 (obs.) | $1.62 \times 10^{-8}$               |
|                        | (80%, 50%)                               | 7        | 5.49       | 15     | 8.21 (obs.)  | $1.33 \times 10^{-8}$               |
|                        | (95%, 50%)                               | 10       | 4.34       | 11     | 13.07 (obs.) | $2.14 \times 10^{-8}$               |
|                        | (80%, 50%)                               | 6        | 2.75       | 8      | 9.13 (obs.)  | $1.74 \times 10^{-8}$               |

Table 6.1: Results of the antihelium search for three AMS subdetectors: TOF, RICH-NaF and RICH-AGL. The table presents the number of surviving events in AMS-02 data,  $^3\text{He}$  and  $^4\text{He}$  MC (with and without weights), the exclusion limits (expected or observed) on the signal rate and the ratio  $\frac{\overline{He}}{He}$ , obtained applying the CLs method for two pairs of WPs. The two additional rows for RICH-AGL show the results obtained applying the additional cuts on the minimum distance from the TRK aluminium jigs and on the  $\chi^2_Y$  as described in Section 5.3.3.

This work has led to the identification of up to ten interesting candidate events that satisfy all the selection criteria developed throughout the analysis. The observation of such a number of candidates, in light of the expected background, results in a mild tension with the background-only hypothesis. Considerable effort has been devoted to understanding detector performance and validating the background estimates obtained through MC simulations. Nevertheless, given the rarity of the phenomenon under investigation, even minute residual instrumental effects can play a significant role. For this reason, the reliability of the background estimate is a critical aspect of the analysis.

Given the profound implications that the detection of antihelium would have for cosmic-ray physics, baryogenesis, and the possible existence of antimatter domains in the Universe, these results must be treated with both interest and caution. At present, they do not allow a definitive statement, but they motivate further investigations. Refinements in background modelling, improved background-rejection strategies, and additional cross-checks will be crucial for assessing

the nature of the observed events. Ultimately, a dedicated experimental effort may be required to conclusively determine whether the tension observed in this analysis reflects new physics or residual instrumental effects. The present work represents a step within this broader and ongoing scientific endeavour.

# Appendix A

## Deep Neural Networks variables

### A.1 Supervised classifiers

The distributions of all input features of the two supervised classifiers are presented. Each plot shows two histograms, one for each label value (see equations 4.16 and 4.17), where the well-reconstructed rigidities are the blue (green) histogram and the poorly reconstructed rigidities are the pink (red) histogram for ISS data ( $^3\text{He}$  and  $^4\text{He}$  MC).

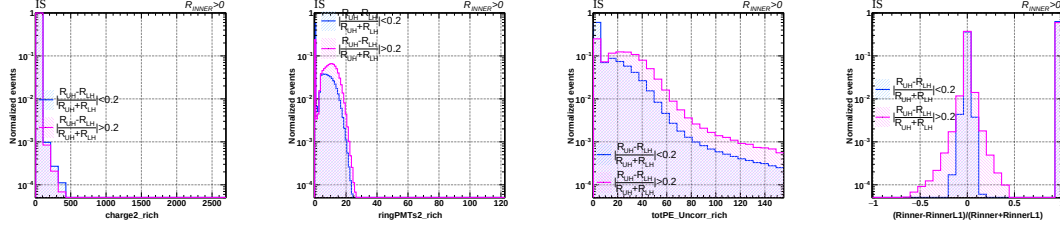


Figure A.1: Positive rigidities in ISS data. (Left) RICH charge; (Center-left) Number of PMTs in RICH ring; (Center-right) Measured photo-electrons in RICH; (Right)  $\frac{R_{INNER-L1}-R_{INNER}}{R_{INNER-L1}+R_{INNER}}$ .

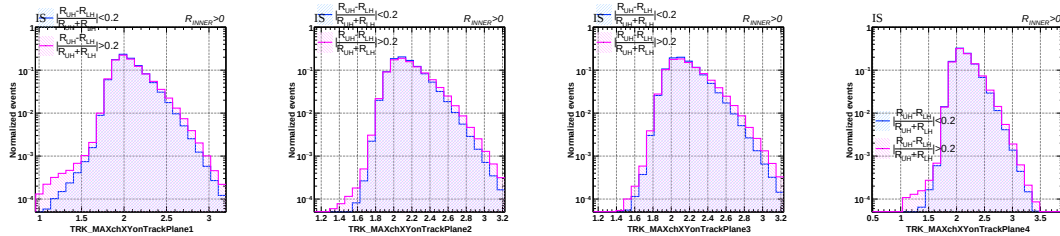


Figure A.2: Positive rigidities in ISS data. (Left) Max charge on inner tracker plane 1; (Center-left) Max charge on inner tracker plane 2; (Center-right) Max charge on inner tracker plane 3; (Right) Max charge on inner tracker plane 4.

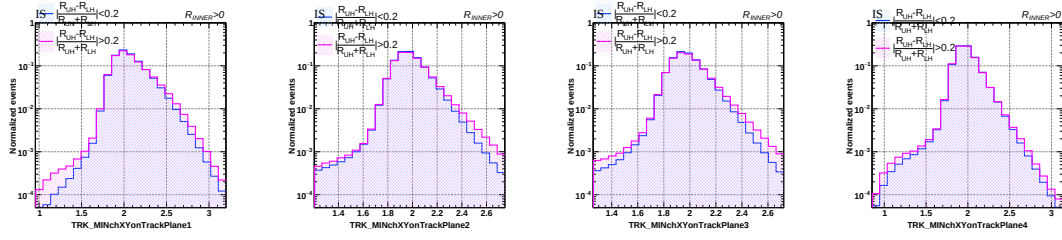


Figure A.3: Positive rigidities in ISS data. (Left) Min charge on inner tracker plane 1; (Center-left) Min charge on inner tracker plane 2; (Center-right) Min charge on inner tracker plane 3; (Right) Min charge on inner tracker plane 4.

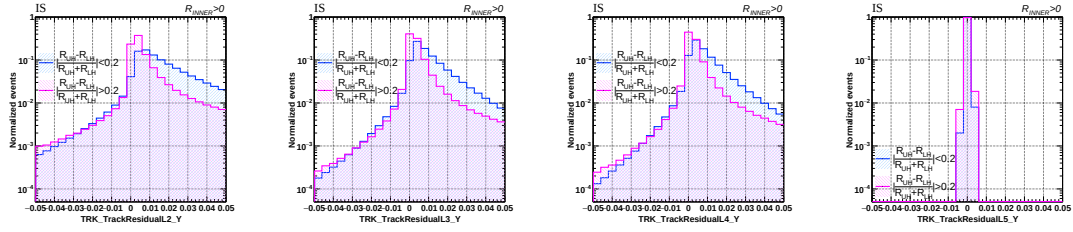


Figure A.4: Positive rigidities in ISS data. (Left) L2 y-residual; (Center-left) L3 y-residual; (Center-right) L4 y-residual; (Right) L5 y-residual.

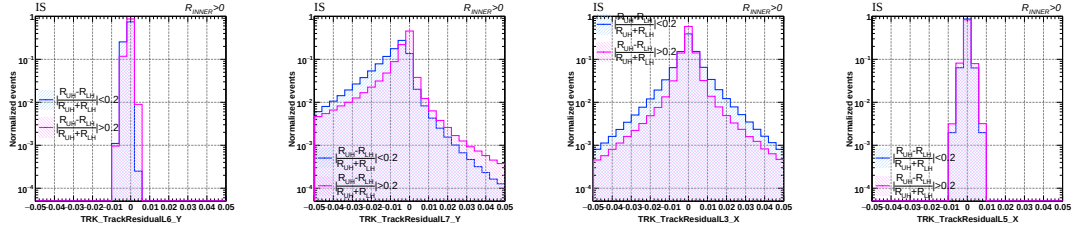


Figure A.5: Positive rigidities in ISS data. (Left) L6 y-residual; (Center-left) L7 y-residual; (Center-right) L3 x-residual; (Right) L5 x-residual.

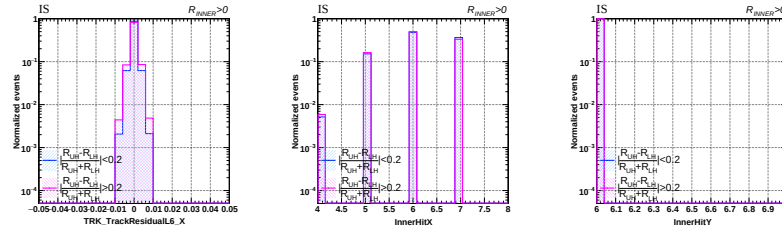


Figure A.6: Positive rigidities in ISS data. (Left) L6 x-residual; (Center) Number of x-hits inside the inner tracker; (Right) Number of y-hits inside the inner tracker.



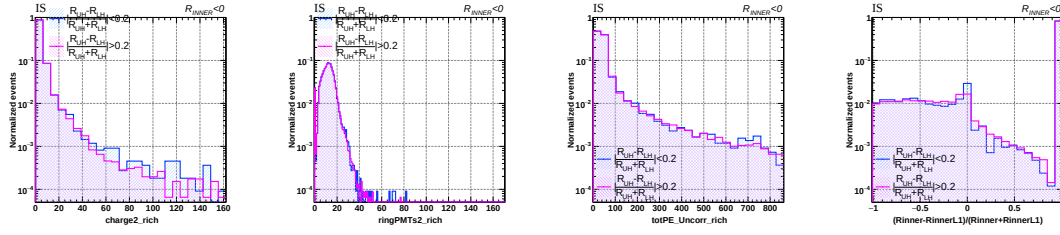


Figure A.7: Negative rigidities in ISS data. (Left) RICH charge; (Center-left) Number of PMTs in RICH ring; (Center-right) Measured photo-electrons in RICH; (Right)  $\frac{R_{INNER-L1}-R_{INNER}}{R_{INNER-L1}+R_{INNER}}$ .

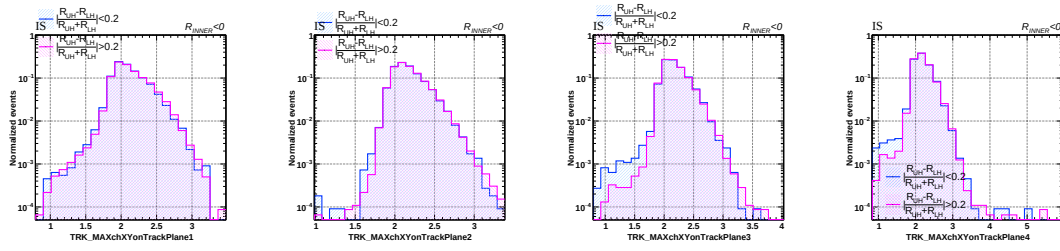


Figure A.8: Negative rigidities in ISS data. (Left) Max charge on inner tracker plane 1; (Center-left) Max charge on inner tracker plane 2; (Center-right) Max charge on inner tracker plane 3; (Right) Max charge on inner tracker plane 4.

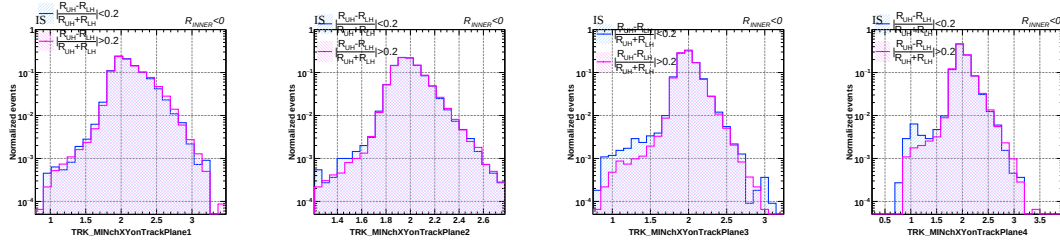


Figure A.9: Negative rigidities in ISS data. (Left) Min charge on inner tracker plane 1; (Center-left) Min charge on inner tracker plane 2; (Center-right) Min charge on inner tracker plane 3; (Right) Min charge on inner tracker plane 4.

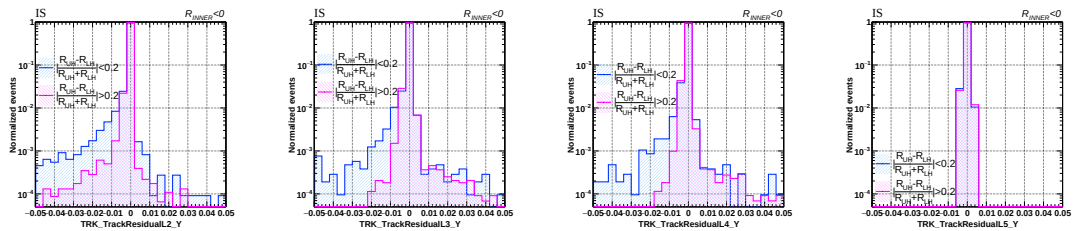


Figure A.10: Negative rigidities in ISS data. (Left) L2 y-residual; (Center-left) L3 y-residual; (Center-right) L4 y-residual; (Right) L5 y-residual.

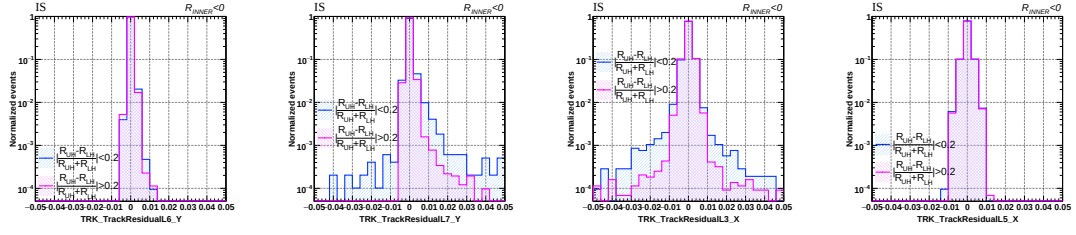


Figure A.11: Negative rigidities in ISS data. (Left) L6 y-residual; (Center-left) L7 y-residual; (Center-right) L3 x-residual; (Right) L5 x-residual.

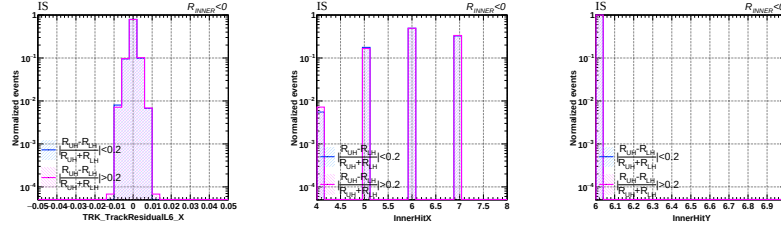


Figure A.12: Negative rigidities in ISS data. (Left) L6 x-residual; (Center) Number of x-hits inside the inner tracker; (Right) Number of y-hits inside the inner tracker.

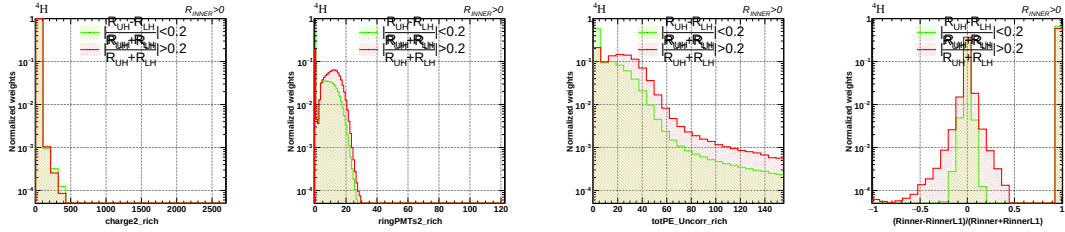


Figure A.13: Positive rigidities in MC. (Left) RICH charge; (Center-left) Number of PMTs in RICH ring; (Center-right) Measured photo-electrons in RICH; (Right)  $\frac{R_{INNER-L1} - R_{INNER}}{R_{INNER-L1} + R_{INNER}}$ .

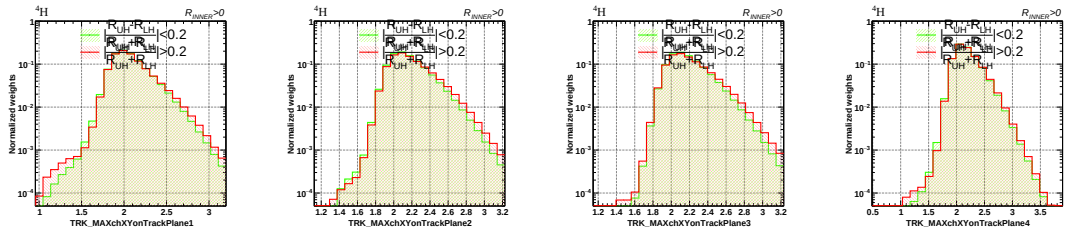


Figure A.14: Positive rigidities in MC. (Left) Max charge on inner tracker plane 1; (Center-left) Max charge on inner tracker plane 2; (Center-right) Max charge on inner tracker plane 3; (Right) Max charge on inner tracker plane 4.

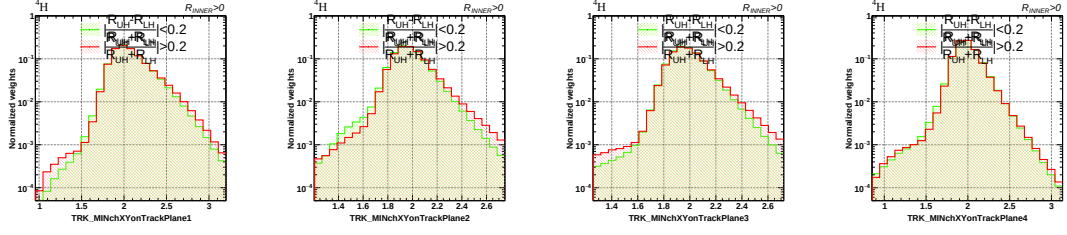


Figure A.15: Positive rigidities in MC. (Left) Min charge on inner tracker plane 1; (Center-left) Min charge on inner tracker plane 2; (Center-right) Min charge on inner tracker plane 3; (Right) Min charge on inner tracker plane 4.

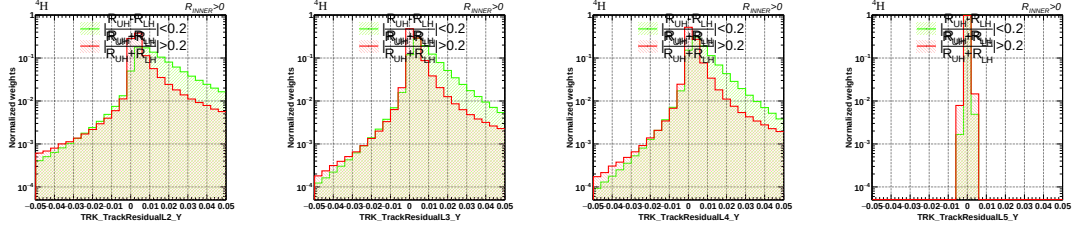


Figure A.16: Positive rigidities in MC. (Left) L2 y-residual; (Center-left) L3 y-residual; (Center-right) L4 y-residual; (Right) L5 y-residual.

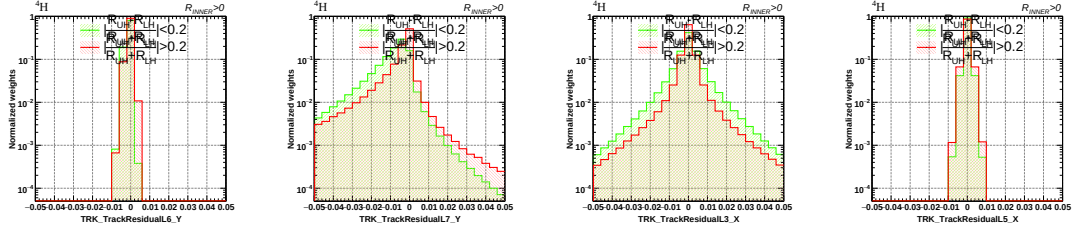


Figure A.17: Positive rigidities in MC (Left) L6 y-residual; (Center-left) L7 y-residual; (Center-right) L3 x-residual; (Right) L5 x-residual.

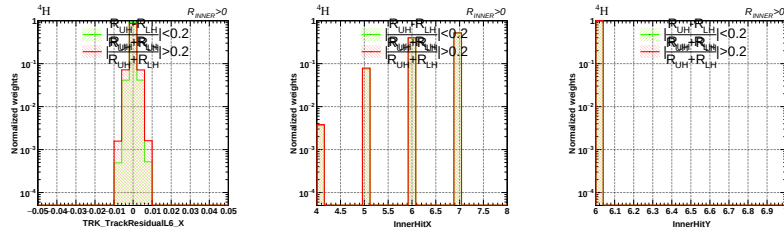


Figure A.18: Positive rigidities in MC (Left) L6 x-residual; (Center) Number of x-hits inside the inner tracker; (Right) Number of y-hits inside the inner tracker.

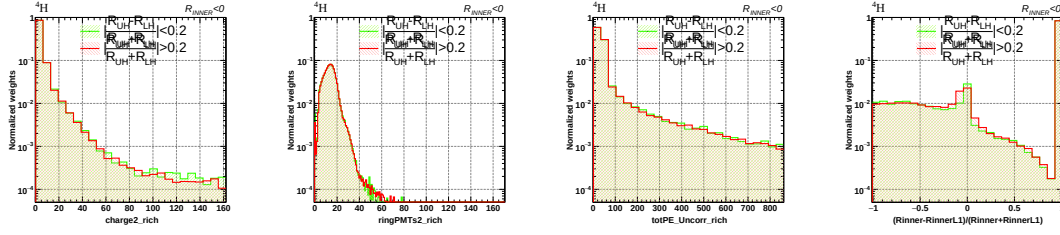


Figure A.19: Positive rigidities in MC. (Left) RICH charge; (Center-left) Number of PMTs in RICH ring; (Center-right) Measured photo-electrons in RICH; (Right)  $\frac{R_{INNER-L1} - R_{INNER}}{R_{INNER-L1} + R_{INNER}}$ .

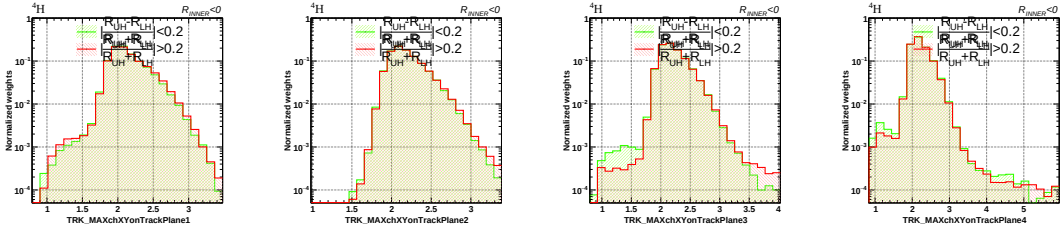


Figure A.20: Positive rigidities in MC. (Left) Max charge on inner tracker plane 1; (Center-left) Max charge on inner tracker plane 2; (Center-right) Max charge on inner tracker plane 3; (Right) Max charge on inner tracker plane 4.

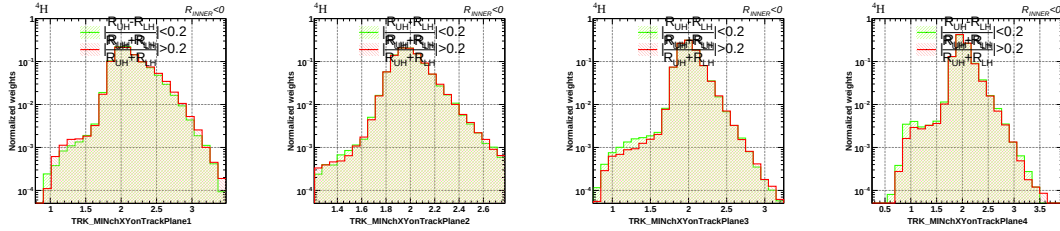


Figure A.21: Positive rigidities in MC. (Left) Min charge on inner tracker plane 1; (Center-left) Min charge on inner tracker plane 2; (Center-right) Min charge on inner tracker plane 3; (Right) Min charge on inner tracker plane 4.

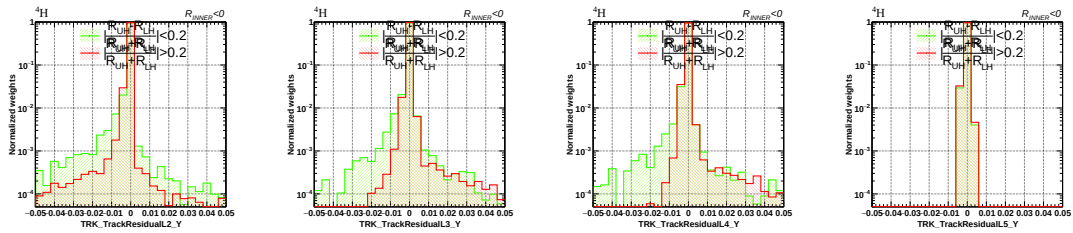


Figure A.22: Positive rigidities in MC. (Left) L2 y-residual; (Center-left) L3 y-residual; (Center-right) L4 y-residual; (Right) L5 y-residual.

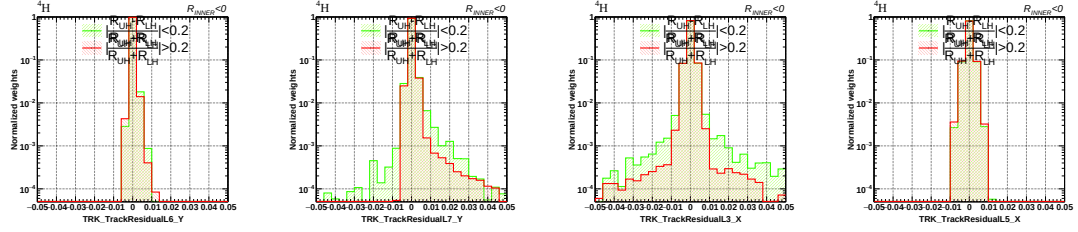


Figure A.23: Positive rigidities in MC (Left) L6 y-residual; (Center-left) L7 y-residual; (Center-right) L3 x-residual; (Right) L5 x-residual.

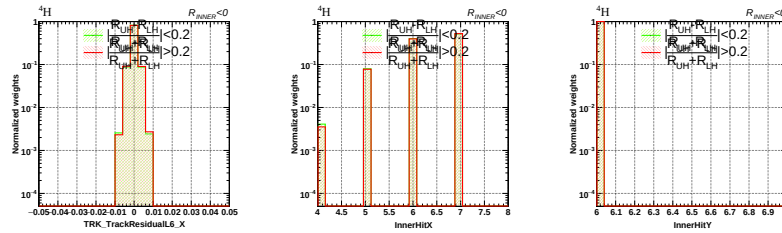


Figure A.24: Positive rigidities in MC (Left) L6 x-residual; (Center) Number of x-hits inside the inner tracker; (Right) Number of y-hits inside the inner tracker.

## A.2 Unsupervised Autoencoders

The distributions of all input features of the two unsupervised autoencoders are presented. Each plot shows two histograms, one for positive rigidities and one for negative rigidities. The blue (green) histograms correspond to  $R > 0$ , whilst the pink (red) histograms correspond to  $R < 0$  for ISS data (MC).

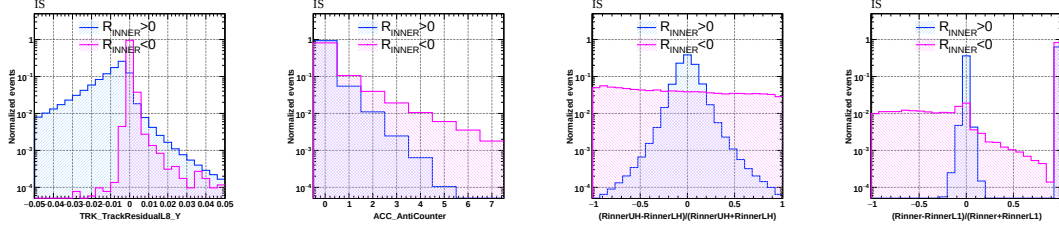


Figure A.25: ISS data. (Left) L8 y-residual; (Center-left) ACC counters; (Center-right)  $\frac{R_{UH}-R_{LH}}{R_{UH}+R_{LH}}$ ; (Right)  $\frac{R_{INNER-L1}-R_{INNER}}{R_{INNER-L1}+R_{INNER}}$ .

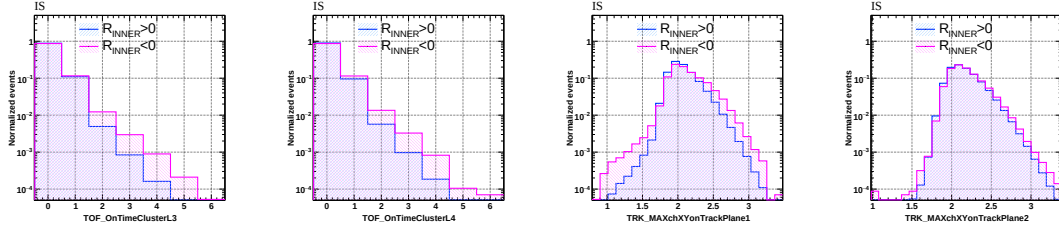


Figure A.26: ISS data. (Left) On time clusters on TOF L3; (Center-left) On time clusters on TOF L4; (Center-right) Max charge on inner tracker plane 1; (Right) Max charge on inner tracker plane 2.

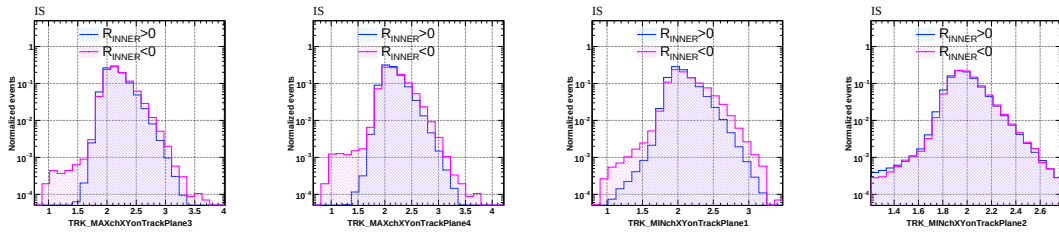


Figure A.27: ISS data. (Left) Max charge on inner tracker plane 3; (Center-left) Max charge on inner tracker plane 4; (Center-right) Min charge on inner tracker plane 1; (Right) Min charge on inner tracker plane 2.

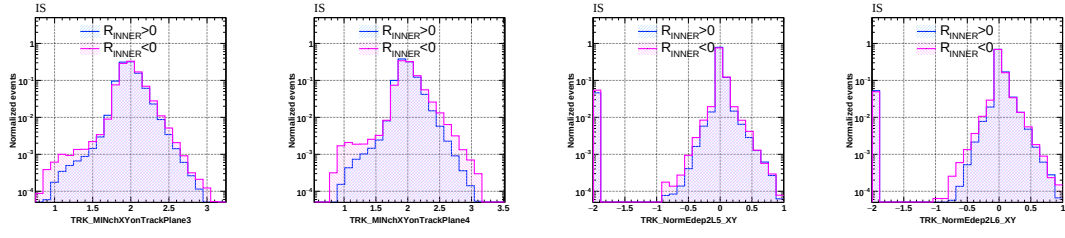


Figure A.28: ISS data. (Left) Min charge on inner tracker plane 3; (Center-left) Min charge on inner tracker plane 4; (Center-right) NormEdep2L5<sub>XY</sub>; (Right) NormEdep2L6<sub>XY</sub>.

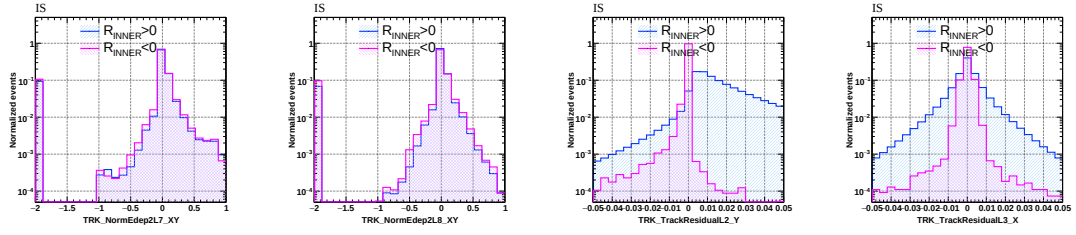


Figure A.29: ISS data. (Left) NormEdep2L7<sub>XY</sub>; (Center-left) NormEdep2L8<sub>XY</sub>; (Center-right) L2 y-residual; (Right) L3 x-residual.

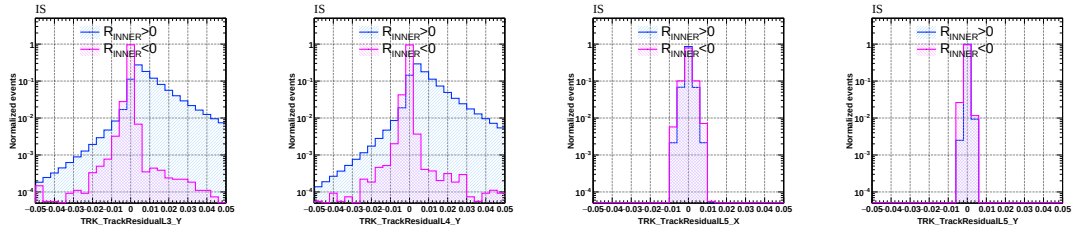


Figure A.30: ISS data. (Left) L3 y-residual; (Center-left) L4 y-residual; (Center-right) L5 x-residual; (Right) L5 y-residual.

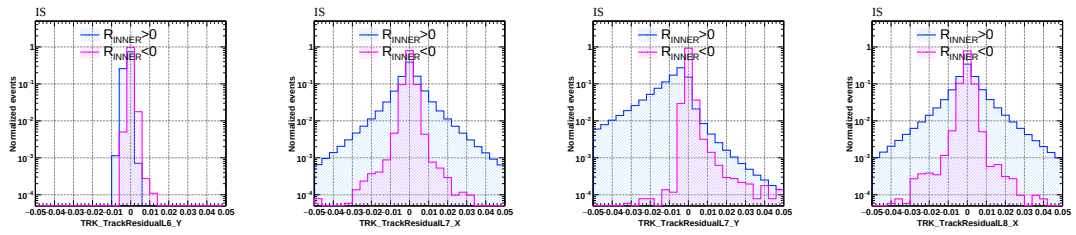


Figure A.31: ISS data. (Left) L6 y-residual; (Center-left) L7 x-residual; (Center-right) L7 y-residual; (Right) L8 x-residual.

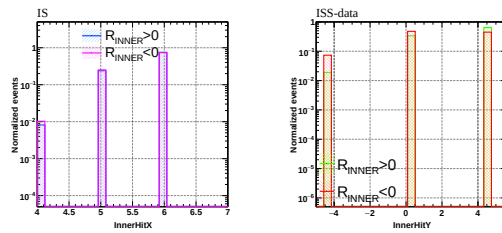


Figure A.32: ISS data. (Left) Number of x-hits inside the inner tracker; (Right) Number of y-hits inside the inner tracker.



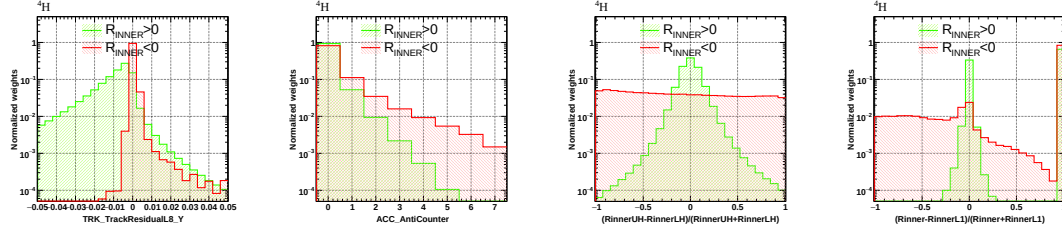


Figure A.33:  $^3\text{He}$  and  $^4\text{He}$  MC. (Left) L8 y-residual; (Center-left) ACC counters; (Center-right)  $\frac{R_{UH}-R_{LH}}{R_{UH}+R_{LH}}$ ; (Right)  $\frac{R_{INNER-L1}-R_{INNER}}{R_{INNER-L1}+R_{INNER}}$ .

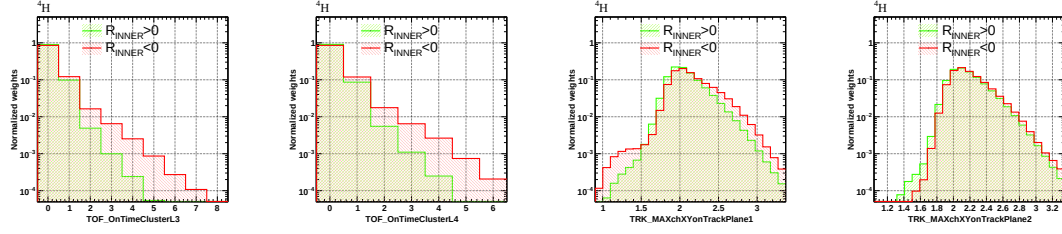


Figure A.34:  $^3\text{He}$  and  $^4\text{He}$  MC. (Left) On time clusters on TOF L3; (Center-left) On time clusters on TOF L4; (Center-right) Max charge on inner tracker plane 1; (Right) Max charge on inner tracker plane 2.

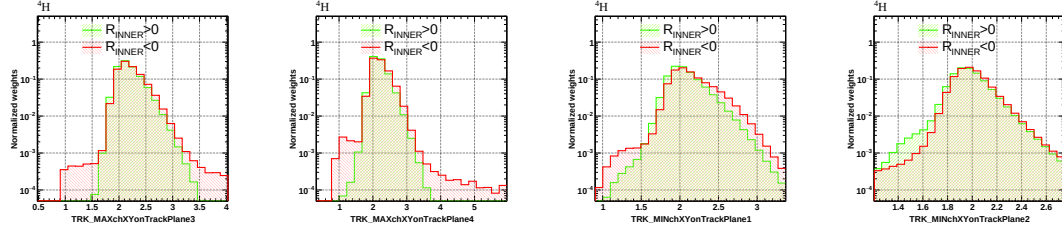


Figure A.35:  $^3\text{He}$  and  $^4\text{He}$  MC. (Left) Max charge on inner tracker plane 3; (Center-left) Max charge on inner tracker plane 4; (Center-right) Min charge on inner tracker plane 1; (Right) Min charge on inner tracker plane 2.

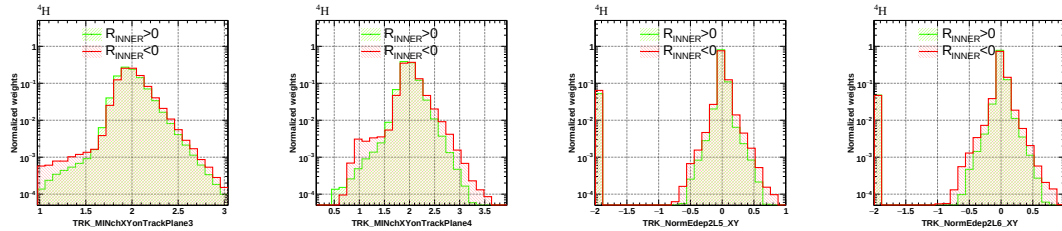


Figure A.36:  $^3\text{He}$  and  $^4\text{He}$  MC. (Left) Min charge on inner tracker plane 3; (Center-left) Min charge on inner tracker plane 4; (Center-right) NormEdep2L5<sub>XY</sub>; (Right) NormEdep2L6<sub>XY</sub>.



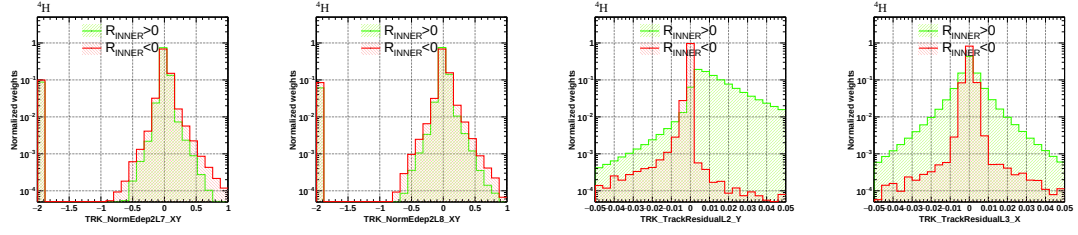


Figure A.37:  $^3\text{He}$  and  $^4\text{He}$  MC. (Left) NormEdep2L7<sub>XY</sub>; (Center-left) NormEdep2L8<sub>XY</sub>; (Center-right) L2 y-residual; (Right) L3 x-residual.

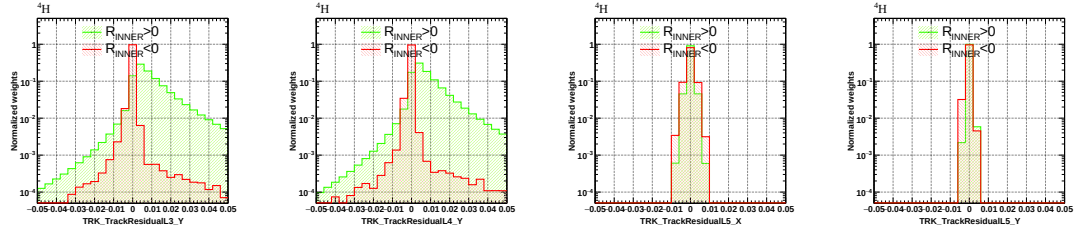


Figure A.38:  $^3\text{He}$  and  $^4\text{He}$  MC. (Left) L3 y-residual; (Center-left) L4 y-residual; (Center-right) L5 x-residual; (Right) L5 y-residual.

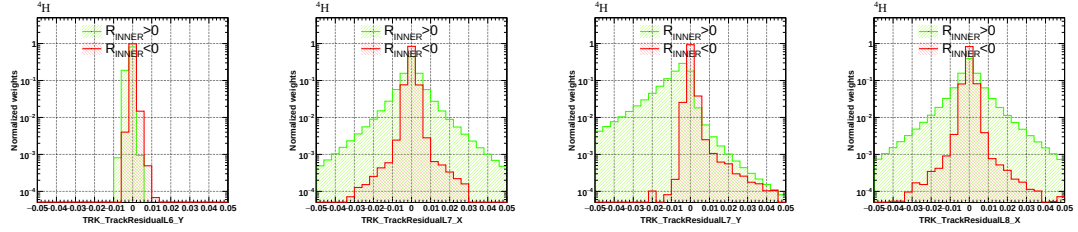


Figure A.39:  $^3\text{He}$  and  $^4\text{He}$  MC. (Left) L6 y-residual; (Center-left) L7 x-residual; (Center-right) L7 y-residual; (Right) L8 x-residual.

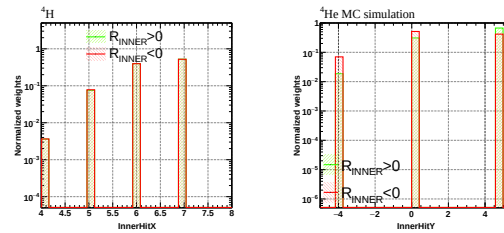


Figure A.40:  $^3\text{He}$  and  $^4\text{He}$  MC. (Left) Number of x-hits inside the inner tracker; (Right) Number of y-hits inside the inner tracker.

## Appendix B

# Deep Neural Networks outputs and orthogonality

### B.1 Supervised classifiers

The supervised classifiers have 4 layers of Feed-forward multilayer perceptrons (see Section 4.2.1), three layers use the Rectified Linear Unit function (ReLU) activation function defined as:

$$\text{ReLU}(x) = \max(0, x). \quad (\text{B.1})$$

and the last layer uses the sigmoid (defined in equation 4.6) as activation function. Each layer applies a linear transformation described by equation 4.4 followed by the activation function applied element-wise. The first layer applies the masking procedure described in Section 4.3.3. Both classifiers ( $CL_{ISS}$  and  $CL_{MC}$ ) use the ADAM optimiser [129], and have the same structure, described in the Table B.1:

| Layer | Type          | Input Dim. | Output Dim. | Activation |
|-------|---------------|------------|-------------|------------|
| 1     | Masked Linear | 23         | 23          | ReLU       |
| 2     | Linear        | 23         | 47          | ReLU       |
| 3     | Linear        | 47         | 31          | ReLU       |
| 4     | Linear        | 31         | 1           | Sigmoid    |

Table B.1: Structure of the two classifiers ( $CL_{ISS}$  and  $CL_{MC}$ ), the output dimension represents the number of nodes in each layer. See the text for more info.

The values of the hyperparameters are reported in Table B.2:

The distributions of the anomaly scores for each classifier, divided into rigidity bins and in velocity detectors, are reported in the next paragraphs.

TOF

| Parameter                | Value ( $CL_{ISS}/CL_{MC}$ ) |
|--------------------------|------------------------------|
| Learning rate            | $(1/1) \times 10^{-2}$       |
| Batch size               | 256/256                      |
| Training epochs          | 150/100                      |
| Training sample (events) | $(1.8/3.4) \times 10^6$      |

Table B.2: The values of the hyperparameters for each one of the two autoencoders ( $AE_{ISS}$  and  $AE_{MC}$ ).

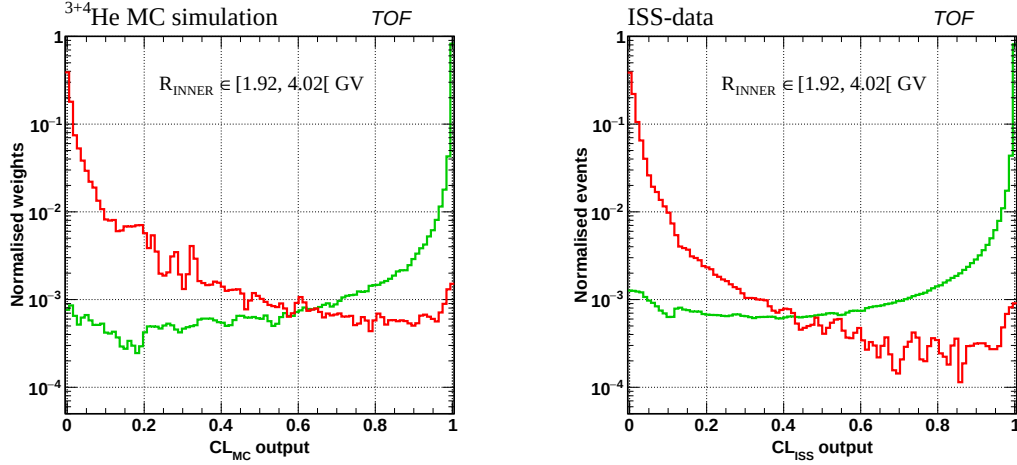


Figure B.1: Classifier scores for events with a positive reconstructed rigidity between  $[1.92, 4.02]$  GV and a reconstructed  $\beta_{TOF}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

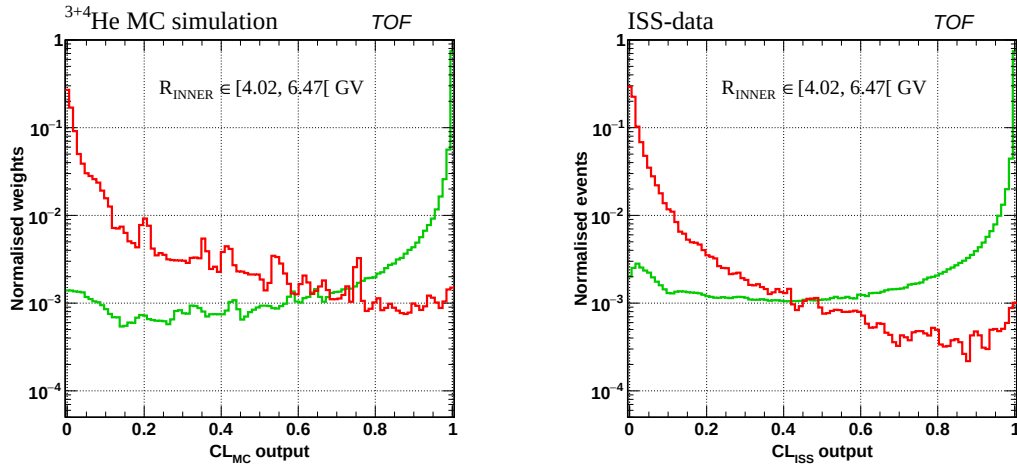


Figure B.2: Classifier scores for events with a positive reconstructed rigidity between  $[4.02, 6.47]$  GV and a reconstructed  $\beta_{TOF}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

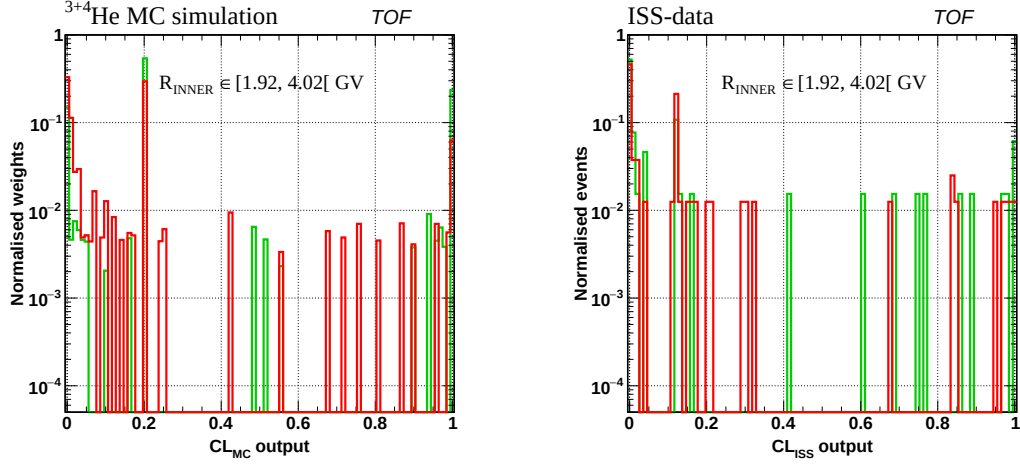


Figure B.3: Classifier scores for events with a negative reconstructed rigidity between  $[1.92, 4.02]$  GV and a reconstructed  $\beta_{\text{TOF}}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

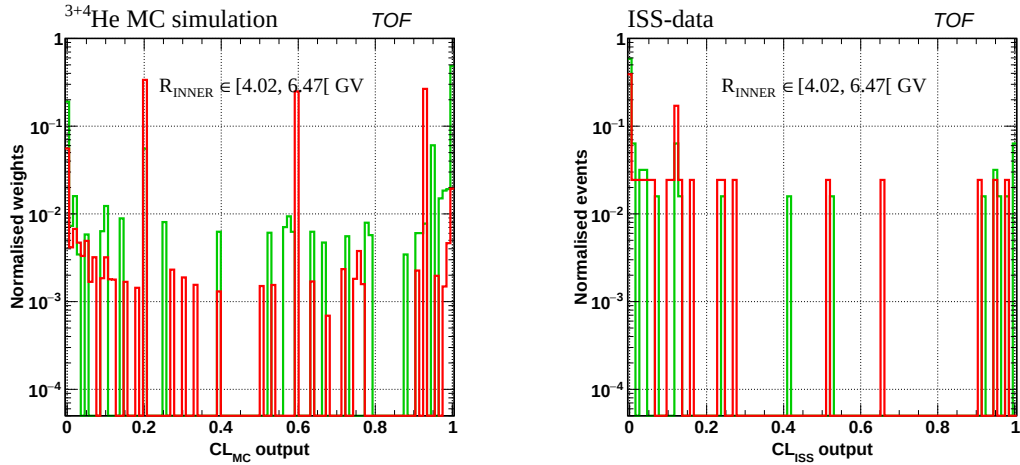


Figure B.4: Classifier scores for events with a negative reconstructed rigidity between  $[4.02, 6.47]$  GV and a reconstructed  $\beta_{\text{TOF}}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

## RICH-NaF

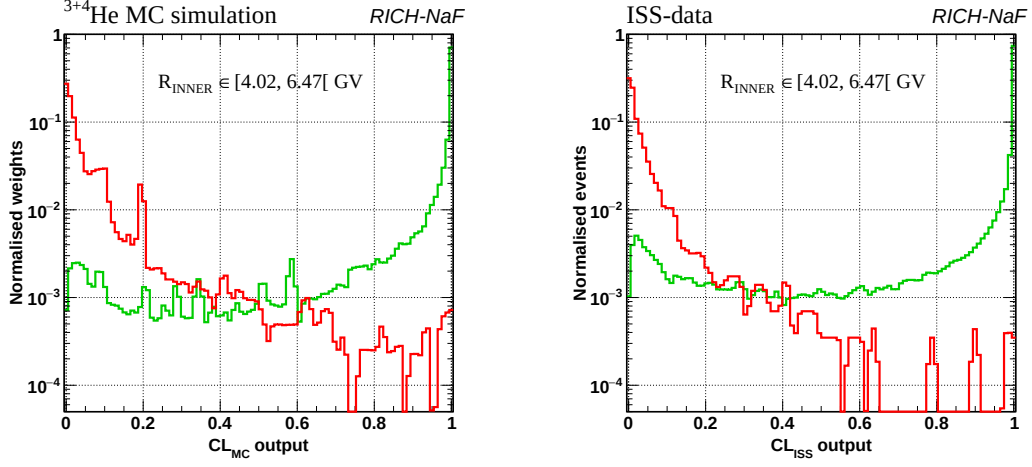


Figure B.5: Classifier scores for events with a positive reconstructed rigidity between  $[4.02, 6.47]$  GV and a reconstructed  $\beta_{\text{NaF}}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

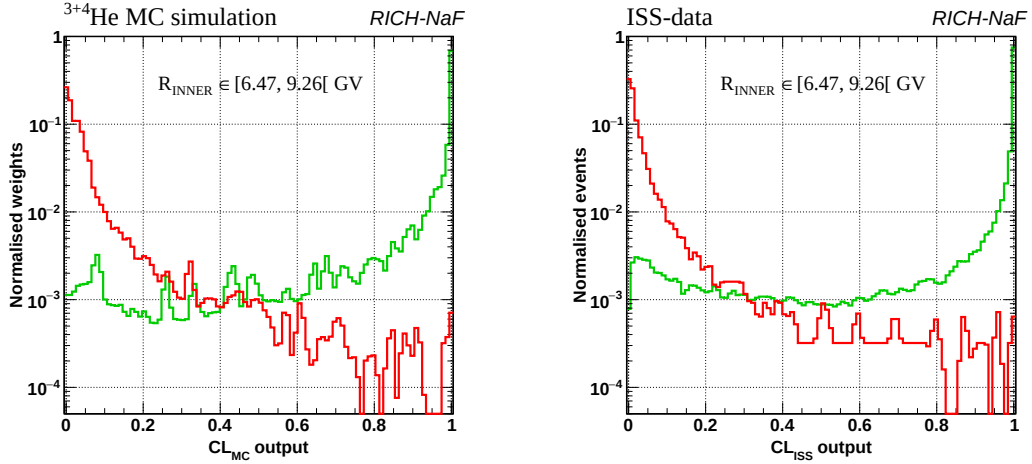


Figure B.6: Classifier scores for events with a positive reconstructed rigidity between  $[6.47, 9.26]$  GV and a reconstructed  $\beta_{\text{NaF}}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

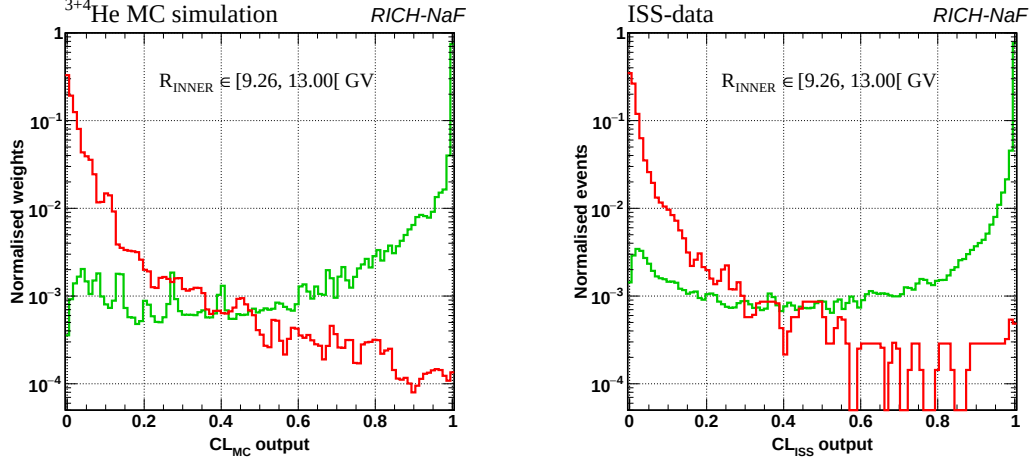


Figure B.7: Classifier scores for events with a positive reconstructed rigidity between  $[9.26, 13.0]$  GV and a reconstructed  $\beta_{\text{NaF}}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

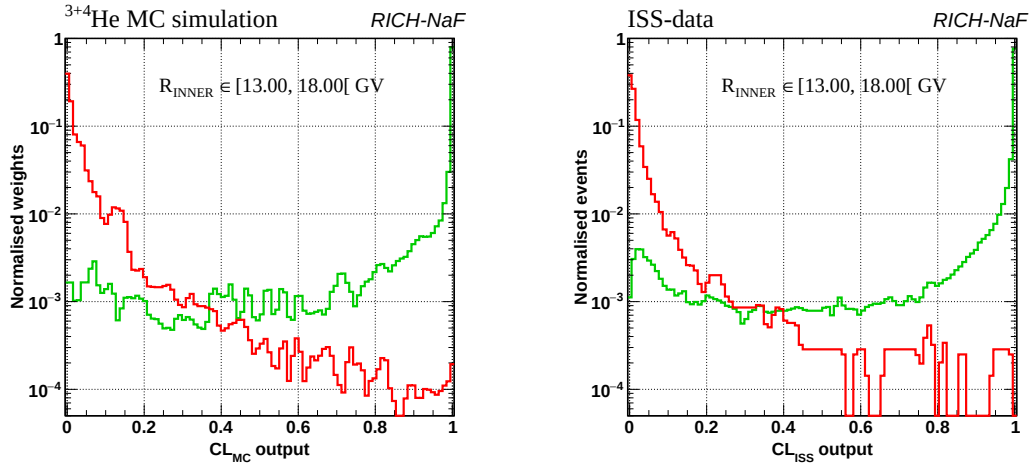


Figure B.8: Classifier scores for events with a positive reconstructed rigidity between  $[13.0, 18.0]$  GV and a reconstructed  $\beta_{\text{NaF}}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

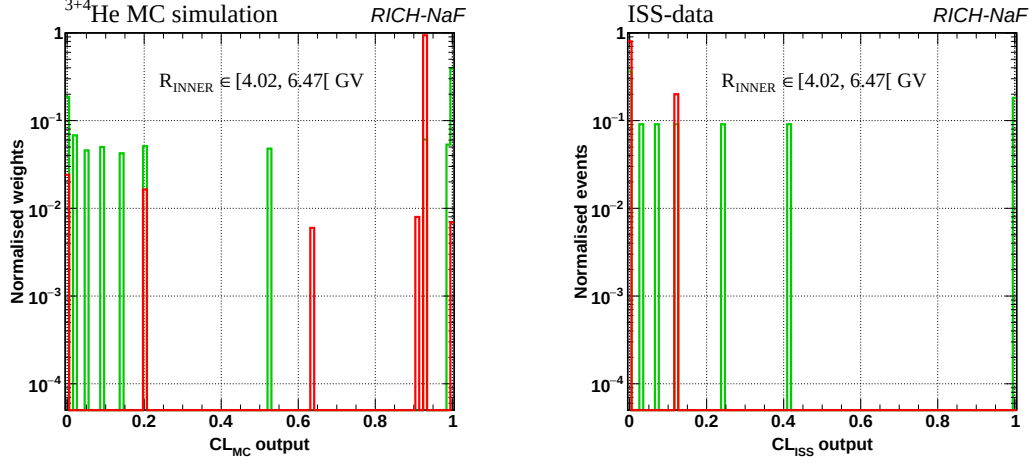


Figure B.9: Classifier scores for events with a negative reconstructed rigidity between  $[4.02, 6.47]$  GV and a reconstructed  $\beta_{\text{NaF}}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

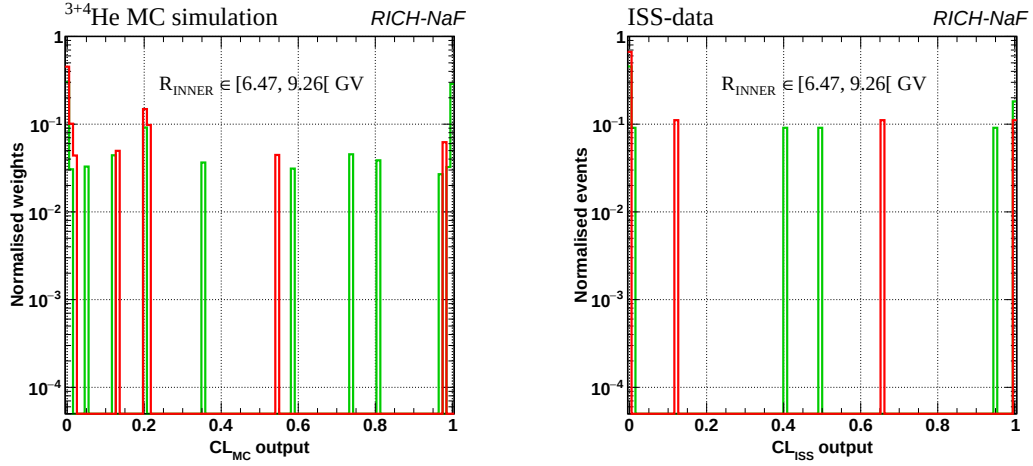


Figure B.10: Classifier scores for events with a negative reconstructed rigidity between  $[6.47, 9.26]$  GV and a reconstructed  $\beta_{\text{NaF}}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

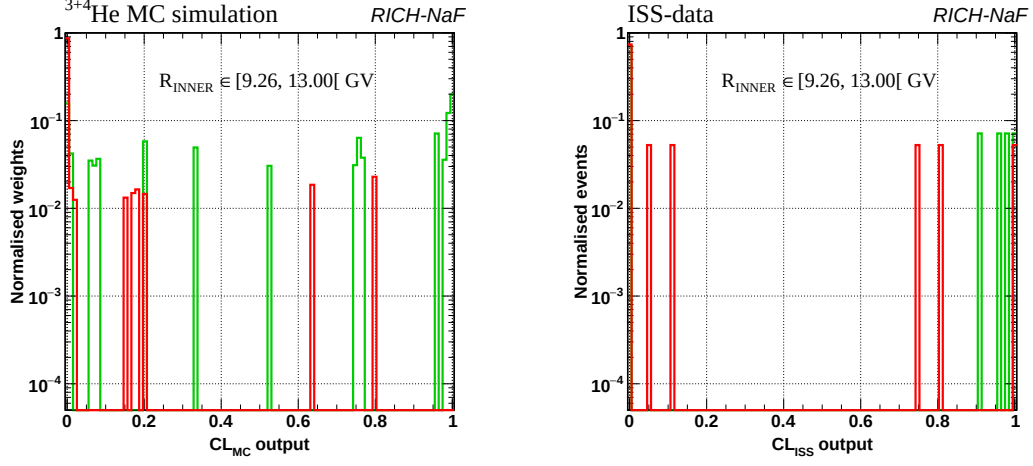


Figure B.11: Classifier scores for events with a negative reconstructed rigidity between  $[9.26, 13.0]$  GV and a reconstructed  $\beta_{NaF}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

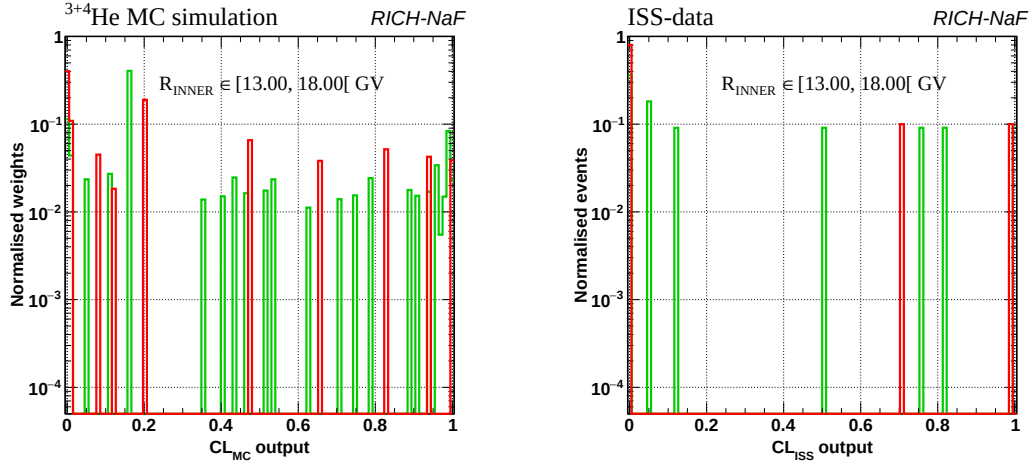


Figure B.12: Classifier scores for events with a negative reconstructed rigidity between  $[13.0, 18.0]$  GV and a reconstructed  $\beta_{NaF}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.



## RICH-AGL

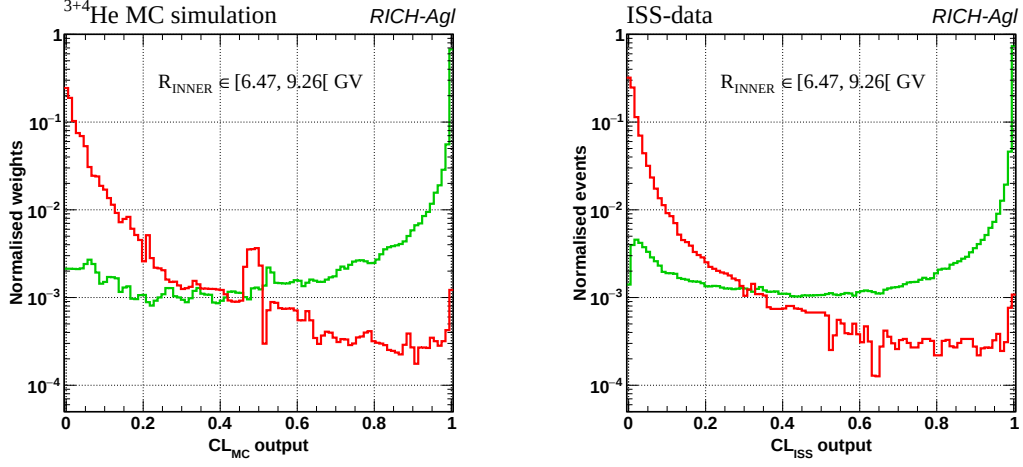


Figure B.13: Classifier scores for events with a positive reconstructed rigidity between  $[6.47, 9.26]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

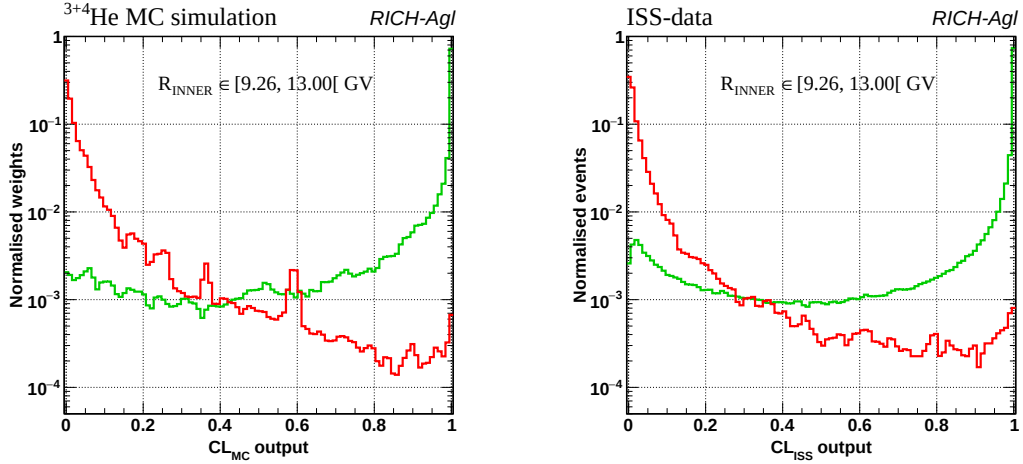


Figure B.14: Classifier scores for events with a positive reconstructed rigidity between  $[9.26, 13.0]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

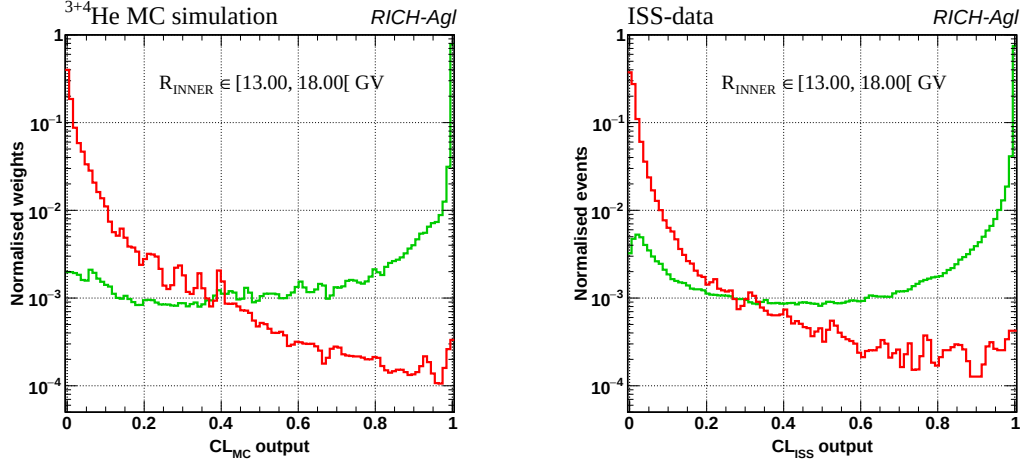


Figure B.15: Classifier scores for events with a positive reconstructed rigidity between  $[13.0, 18.0]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

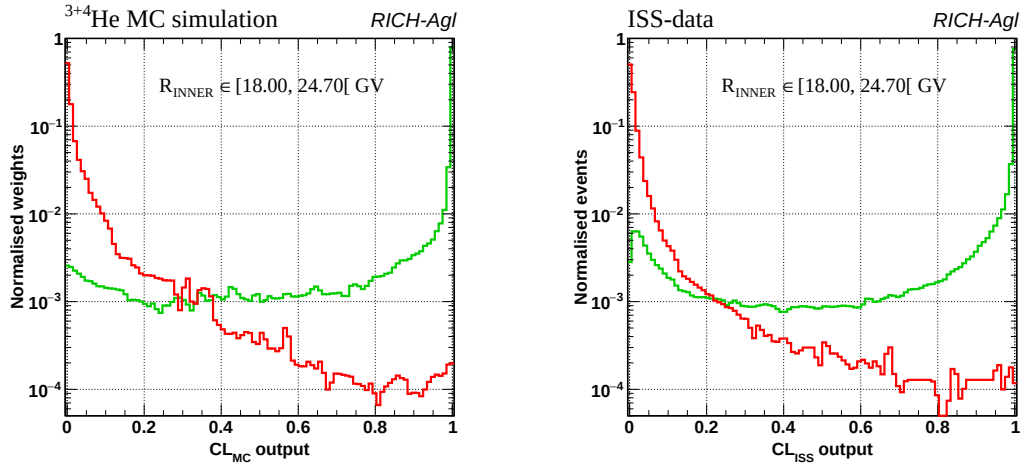


Figure B.16: Classifier scores for events with a positive reconstructed rigidity between  $[18.0, 24.7]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

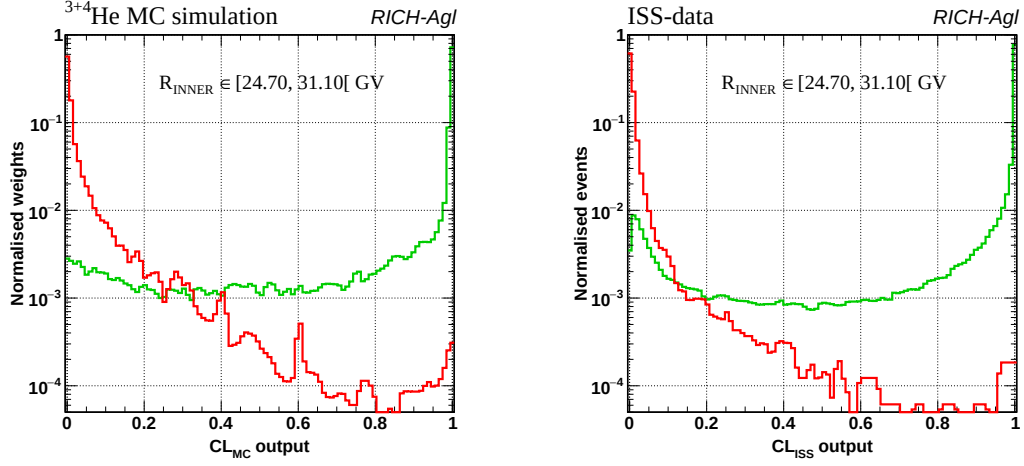


Figure B.17: Classifier scores for events with a positive reconstructed rigidity between  $[24.7, 31.1]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

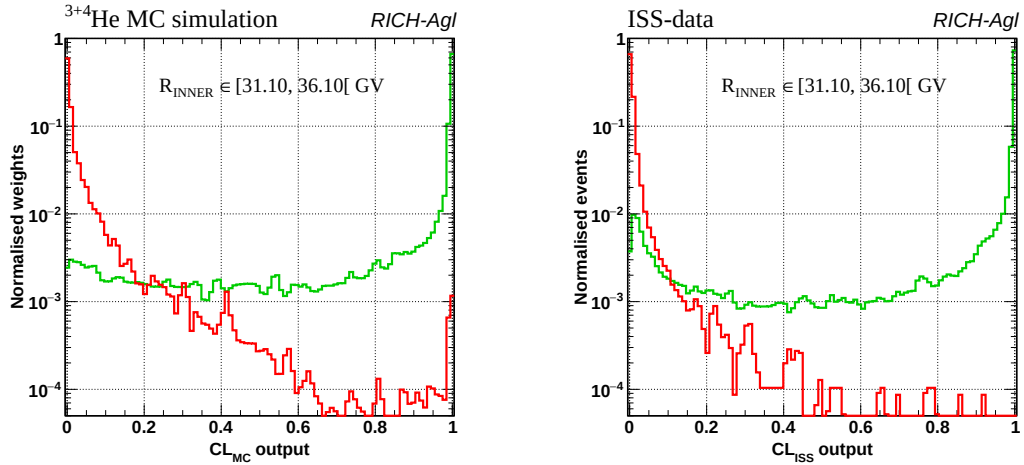


Figure B.18: Classifier scores for events with a positive reconstructed rigidity between  $[31.1, 36.1]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

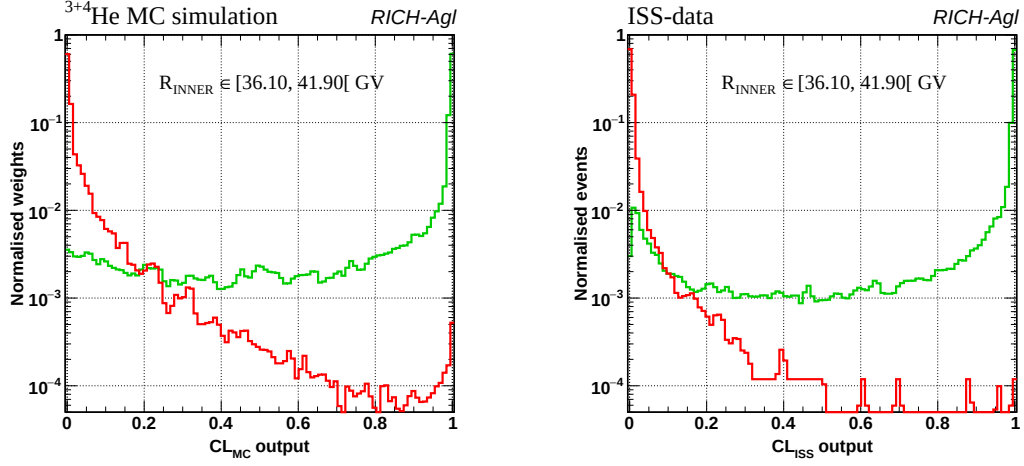


Figure B.19: Classifier scores for events with a positive reconstructed rigidity between  $[36.1, 41.9]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

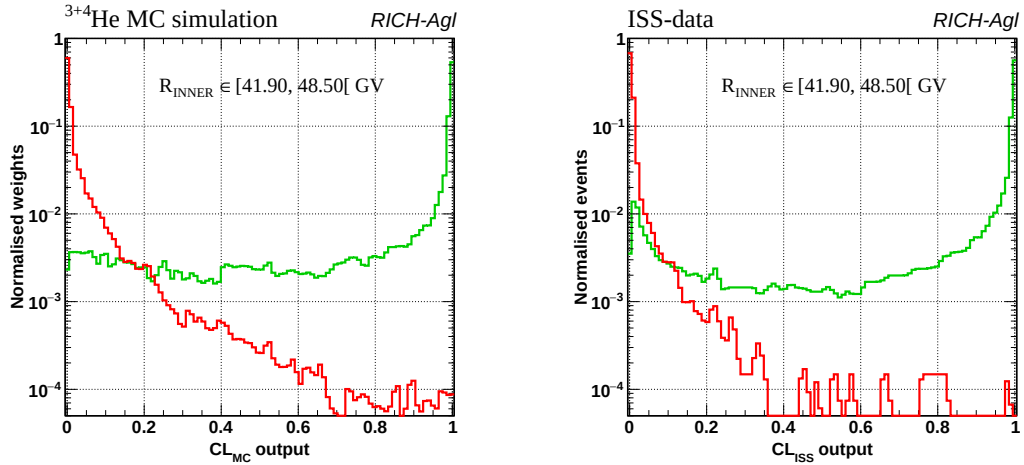


Figure B.20: Classifier scores for events with a positive reconstructed rigidity between  $[41.9, 48.5]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

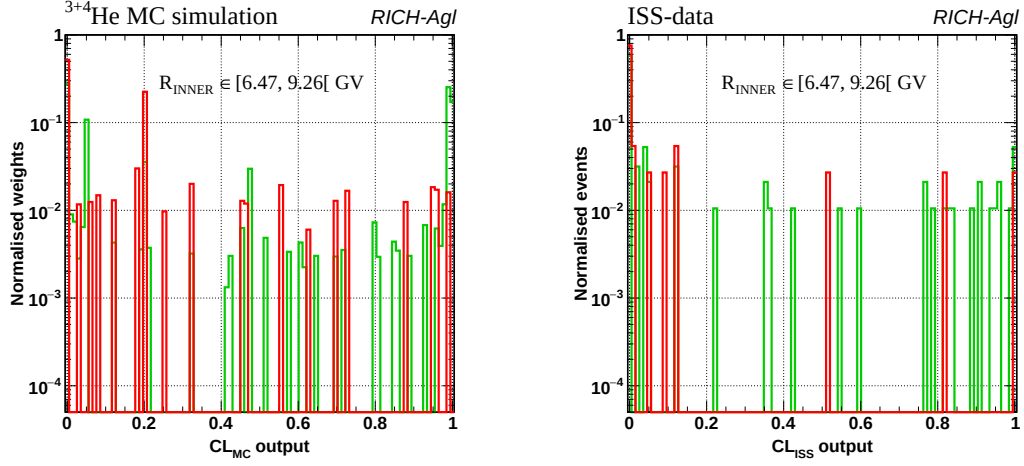


Figure B.21: Classifier scores for events with a negative reconstructed rigidity between  $[6.47, 9.26]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

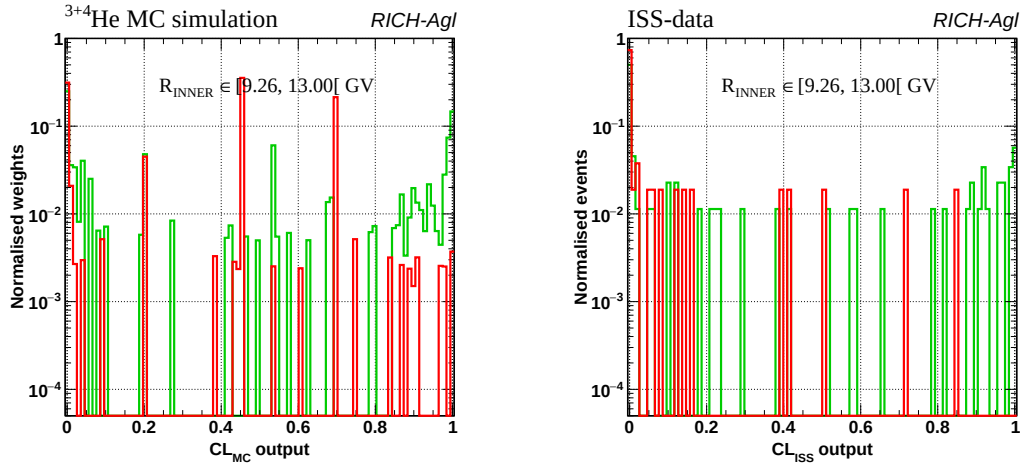


Figure B.22: Classifier scores for events with a negative reconstructed rigidity between  $[9.26, 13.0]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

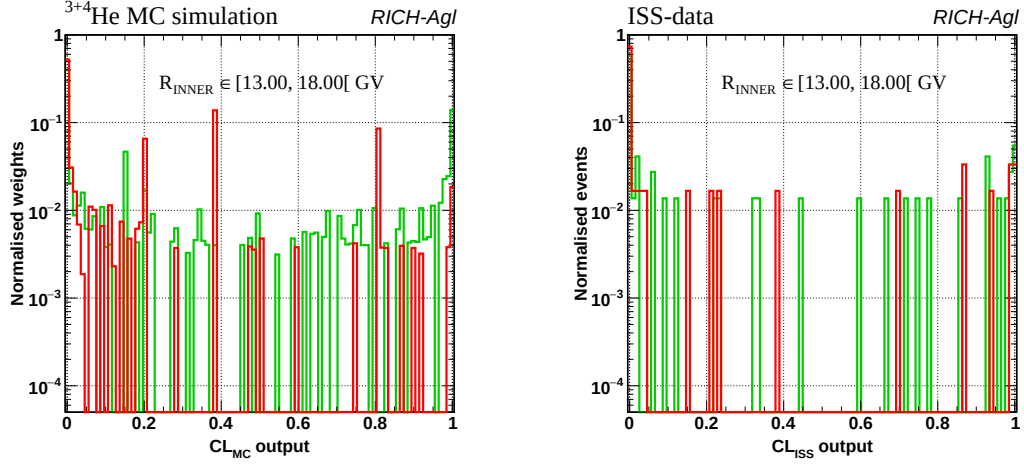


Figure B.23: Classifier scores for events with a negative reconstructed rigidity between  $[13.0, 18.0]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

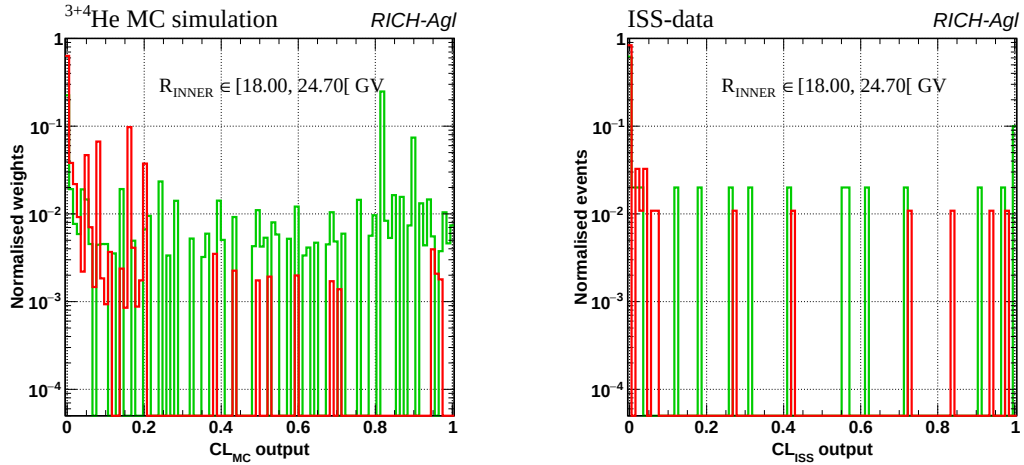


Figure B.24: Classifier scores for events with a negative reconstructed rigidity between  $[18.0, 24.7]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

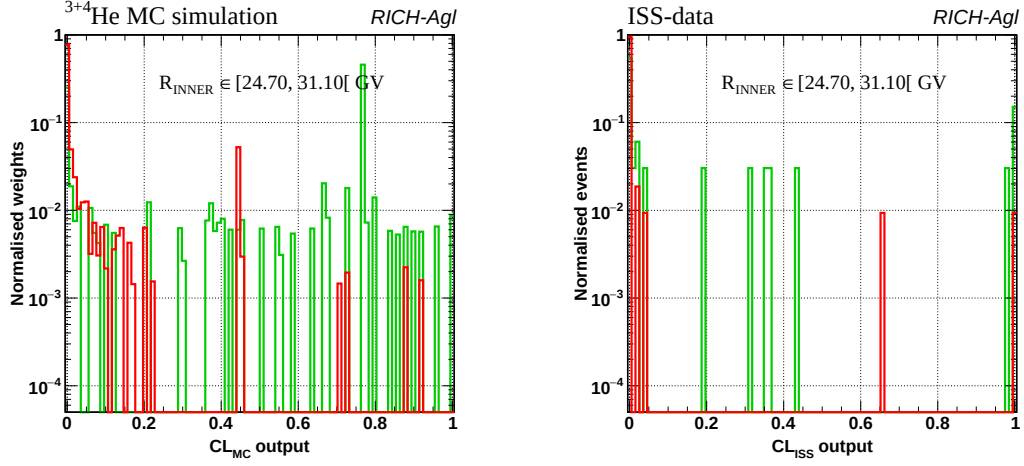


Figure B.25: Classifier scores for events with a negative reconstructed rigidity between  $[24.7, 31.1]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

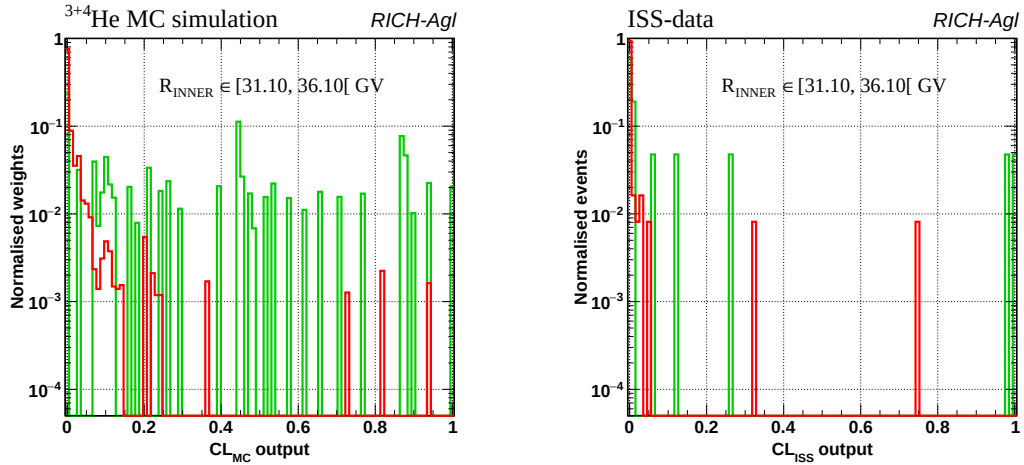


Figure B.26: Classifier scores for events with a negative reconstructed rigidity between  $[31.1, 36.1]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

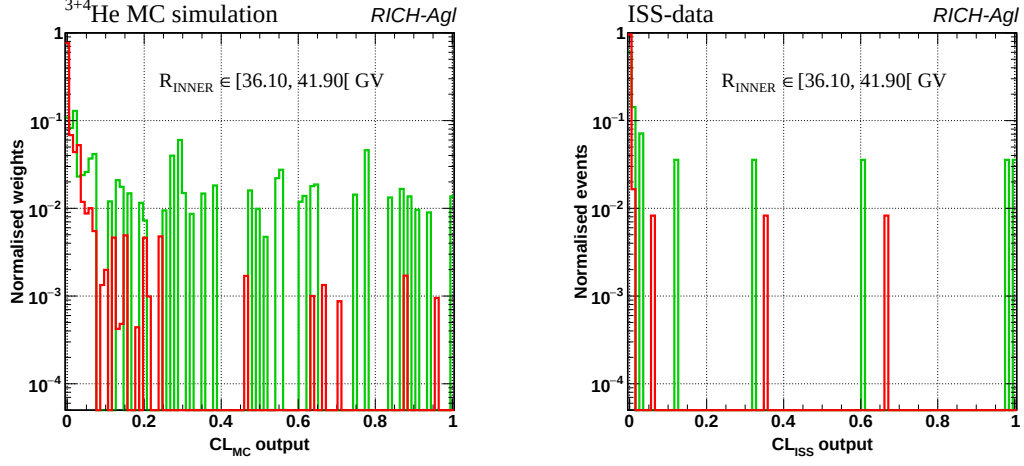


Figure B.27: Classifier scores for events with a negative reconstructed rigidity between  $[36.1, 41.9]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.

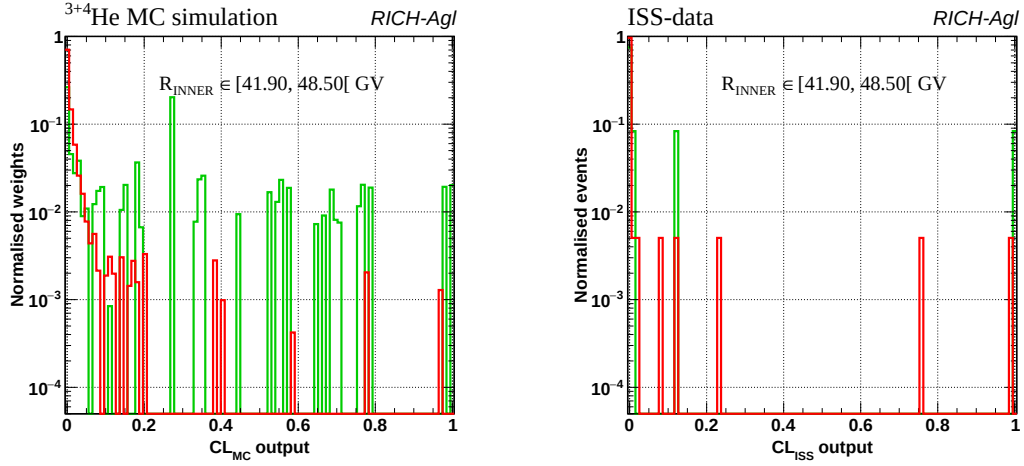


Figure B.28: Classifier scores for events with a negative reconstructed rigidity between  $[41.9, 48.5]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the MC classifier (left) and the ISS classifier (right). The outputs of the DNNs have been divided into the two classes defined by the label, with the green histogram showing well-reconstructed rigidities and the red histogram displaying the poorly reconstructed rigidities.



## B.2 Unsupervised Autoencoders

### DNN hyperparameters

The unsupervised autoencoders have 5 layers of Feed-forward multilayer perceptrons (see Section 4.2.1), and each one of them uses the Rectified Linear Unit function (ReLU) activation function defined as:

$$\text{ReLU}(x) = \max(0, x). \quad (\text{B.2})$$

Each layer applies a linear transformation described by equation 4.4 followed by the activation function applied element-wise. The first layer applies the masking procedure described in Section 4.3.3. Both autoencoders ( $AE_{ISS}$  and  $AE_{MC}$ ) use the ADAM optimiser [129], and have the same symmetric structure, described in the Table B.3:

| Layer | Type          | Input Dim. | Output Dim. | Activation |
|-------|---------------|------------|-------------|------------|
| 1     | Masked Linear | 30         | 30          | ReLU       |
| 2     | Linear        | 30         | 14          | ReLU       |
| 3     | Linear        | 14         | 7           | ReLU       |
| 4     | Linear        | 7          | 14          | ReLU       |
| 5     | Linear        | 14         | 30          | None       |

Table B.3: Structure of the two autoencoders ( $AE_{ISS}$  and  $AE_{MC}$ ), the output dimension represents the number of nodes in each layer. See the text for more info.

The values of the hyperparameters are reported in Table B.4:

| Parameter                | Value ( $AE_{ISS}/AE_{MC}$ ) |
|--------------------------|------------------------------|
| Latent Space             | 7/7                          |
| Learning rate            | $(1/5) \times 10^{-4}$       |
| Batch size               | 256/256                      |
| Training epochs          | 150/100                      |
| Training sample (events) | $(8.4/8.6) \times 10^6$      |

Table B.4: The values of the hyperparameters for each one of the two autoencoders ( $AE_{ISS}$  and  $AE_{MC}$ ).

The distributions of the anomaly scores for each autoencoder, divided into rigidity bins and in velocity detectors, are reported in the next paragraphs.

### TOF

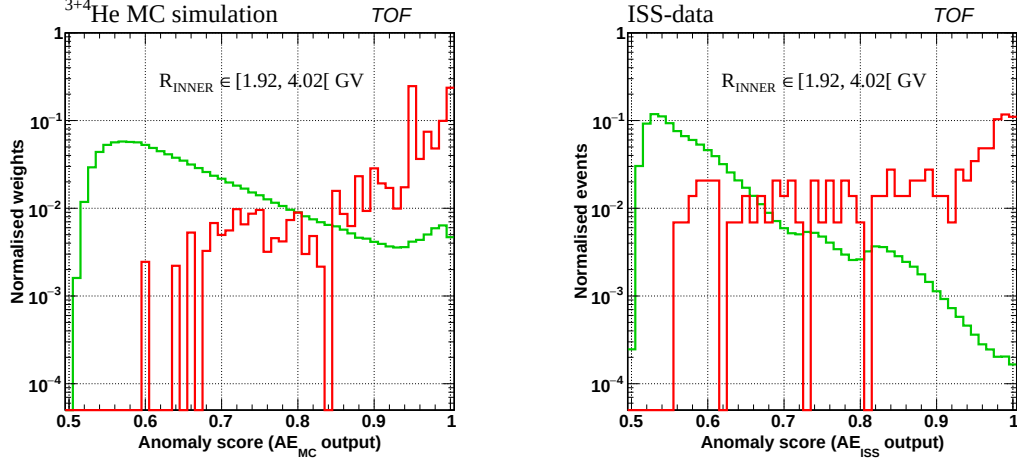


Figure B.29: Anomaly scores for events with a reconstructed rigidity between  $|R| \in [1.92, 4.02]$  GV and a reconstructed  $\beta_{TOF}$ , obtained using the  $AE_{MC}$  (left) and the  $AE_{ISS}$  (right). The green histogram shows positive rigidities, and the red histogram shows negative rigidities.

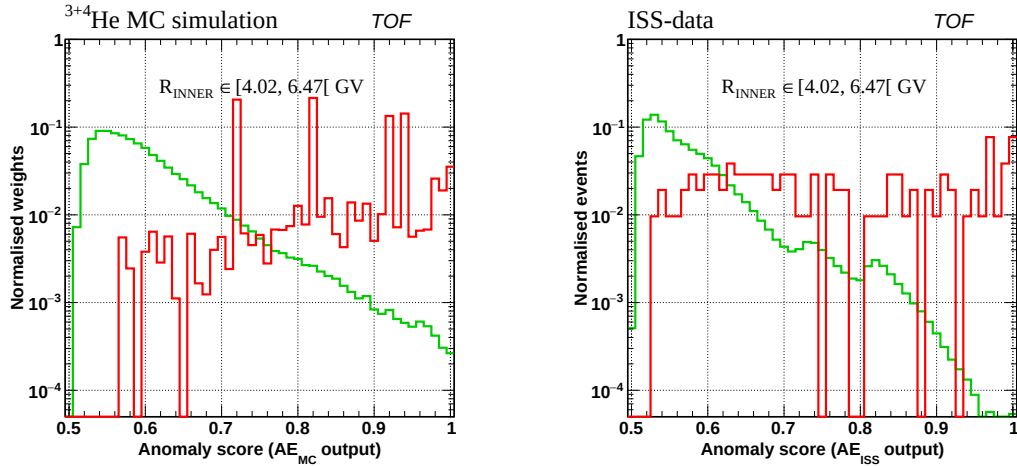


Figure B.30: Anomaly scores for events with a reconstructed rigidity between  $|R| \in [4.02, 6.47]$  GV and a reconstructed  $\beta_{TOF}$ , obtained using the  $AE_{MC}$  (left) and the  $AE_{ISS}$  (right). The green histogram shows positive rigidities, and the red histogram shows negative rigidities.

## RICH-NaF

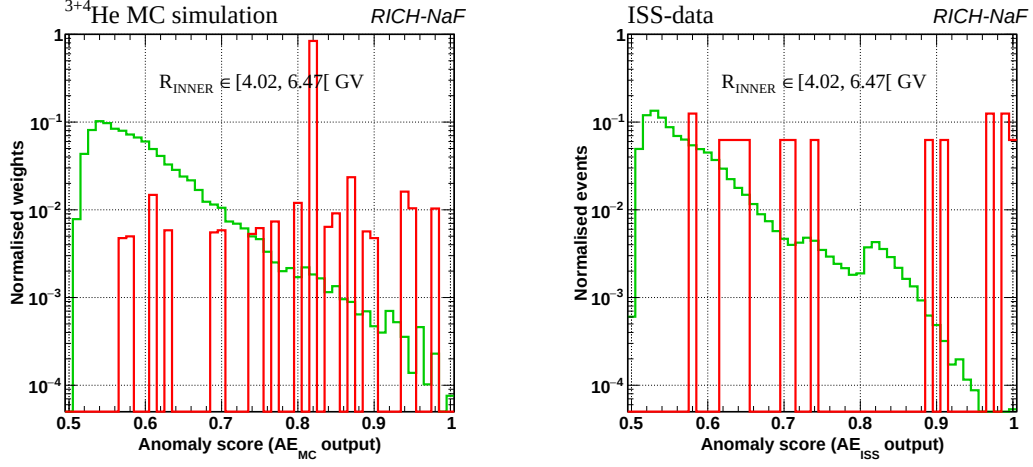


Figure B.31: Anomaly scores for events with a reconstructed rigidity between  $|R| \in [4.02, 6.47]$  GV and a reconstructed  $\beta_{\text{NaF}}$ , obtained using the  $\text{AE}_{\text{MC}}$  (left) and the  $\text{AE}_{\text{ISS}}$  (right). The green histogram shows positive rigidities, and the red histogram shows negative rigidities.

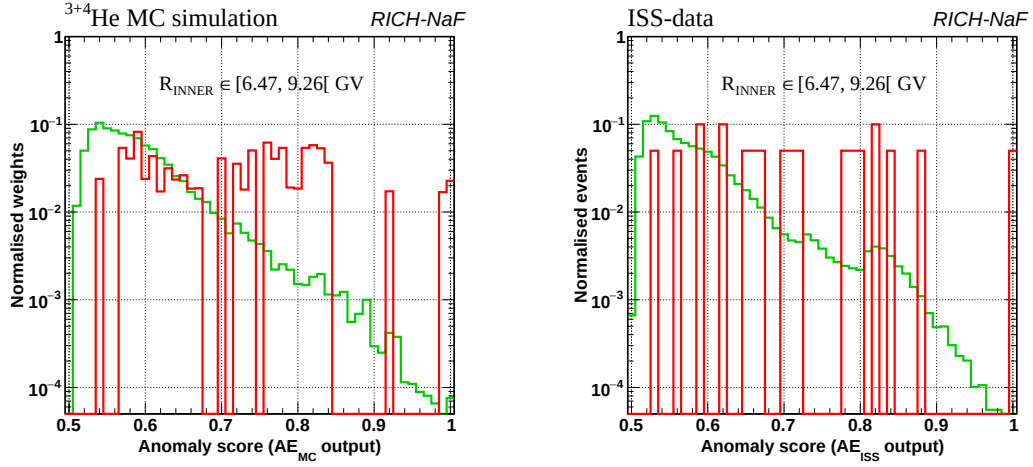


Figure B.32: Anomaly scores for events with a reconstructed rigidity between  $|R| \in [6.47, 9.26]$  GV and a reconstructed  $\beta_{\text{NaF}}$ , obtained using the  $\text{AE}_{\text{MC}}$  (left) and the  $\text{AE}_{\text{ISS}}$  (right). The green histogram shows positive rigidities, and the red histogram shows negative rigidities.

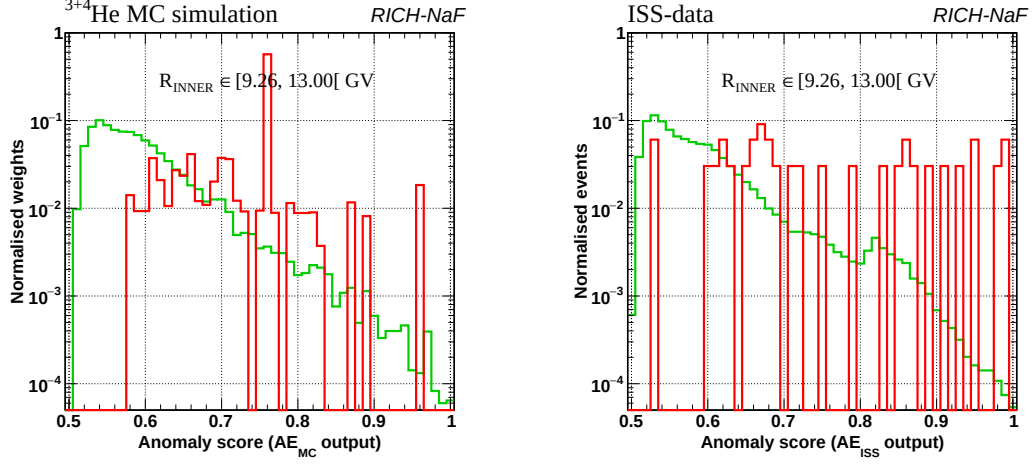


Figure B.33: Anomaly scores for events with a reconstructed rigidity between  $|R| \in [9.26, 13.0]$  GV and a reconstructed  $\beta_{NaF}$ , obtained using the  $AE_{MC}$  (left) and the  $AE_{ISS}$  (right). The green histogram shows positive rigidities, and the red histogram shows negative rigidities.

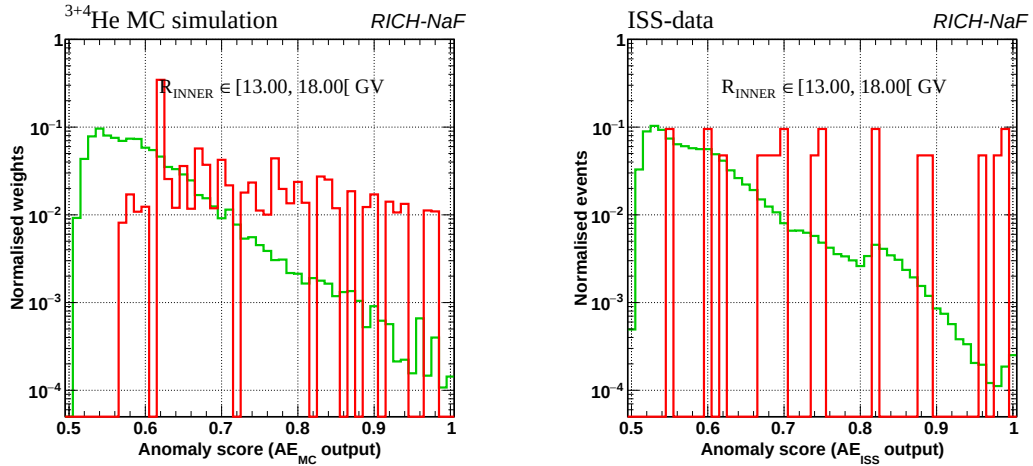


Figure B.34: Anomaly scores for events with a reconstructed rigidity between  $|R| \in [13.0, 18.0]$  GV and a reconstructed  $\beta_{NaF}$ , obtained using the  $AE_{MC}$  (left) and the  $AE_{ISS}$  (right). The green histogram shows positive rigidities, and the red histogram shows negative rigidities.

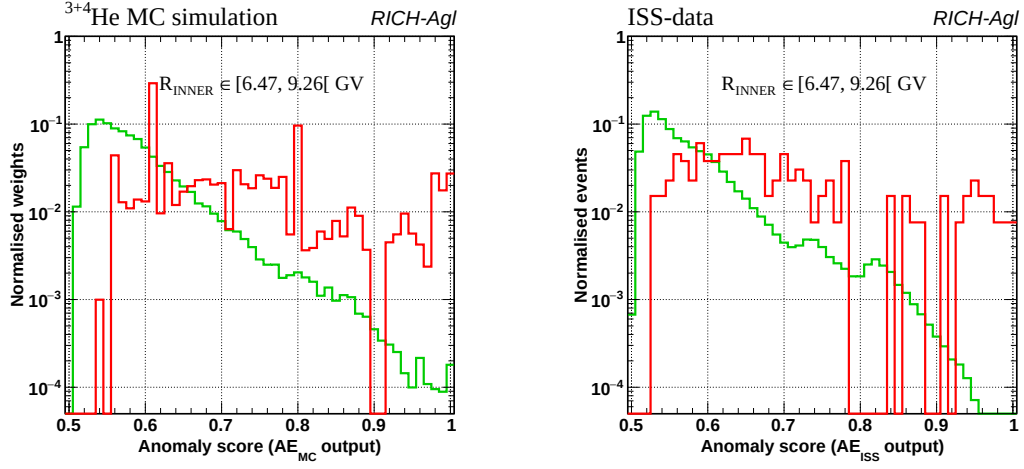


Figure B.35: Anomaly scores for events with a reconstructed rigidity between  $|R| \in [6.47, 9.26]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the  $AE_{MC}$  (left) and the  $AE_{ISS}$  (right). The green histogram shows positive rigidities, and the red histogram shows negative rigidities.

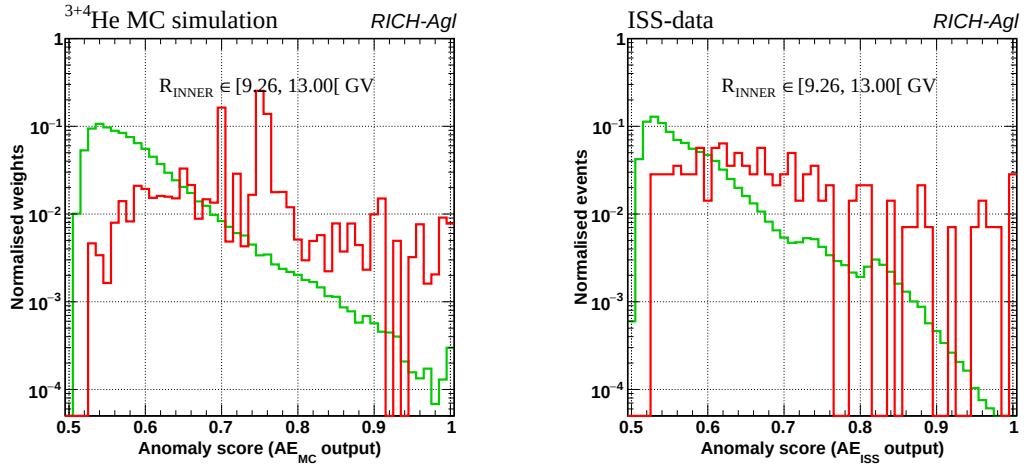


Figure B.36: Anomaly scores for events with a reconstructed rigidity between  $|R| \in [9.26, 13.0]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the  $AE_{MC}$  (left) and the  $AE_{ISS}$  (right). The green histogram shows positive rigidities, and the red histogram shows negative rigidities.

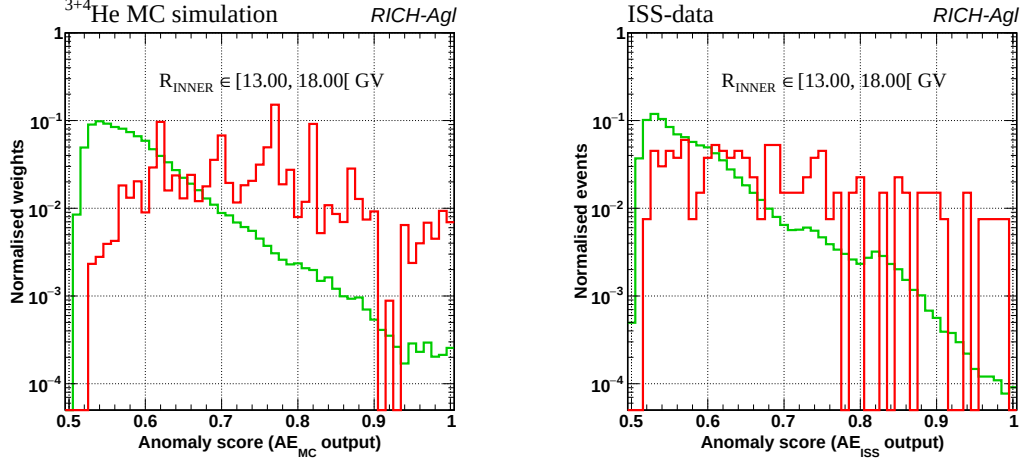


Figure B.37: Anomaly scores for events with a reconstructed rigidity between  $|R| \in [13.0, 18.0]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the  $AE_{MC}$  (left) and the  $AE_{ISS}$  (right). The green histogram shows positive rigidities, and the red histogram shows negative rigidities.

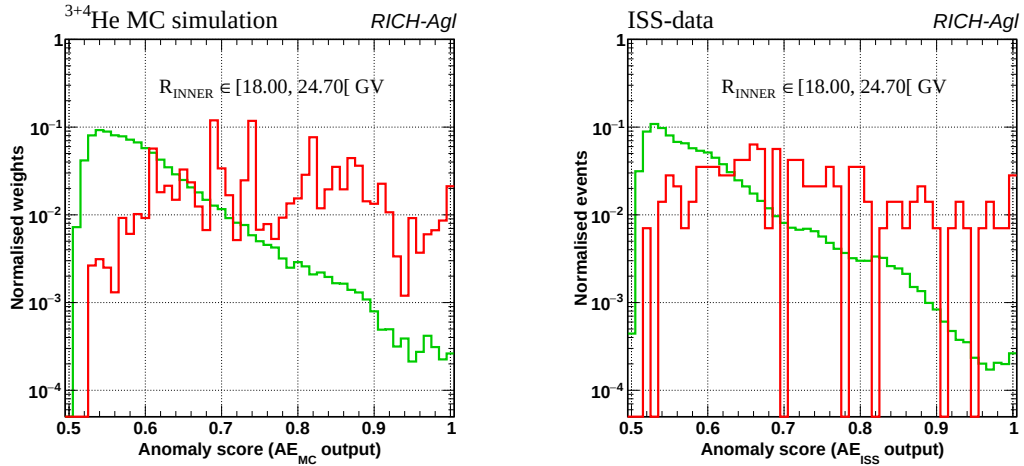


Figure B.38: Anomaly scores for events with a reconstructed rigidity between  $|R| \in [18.0, 24.7]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the  $AE_{MC}$  (left) and the  $AE_{ISS}$  (right). The green histogram shows positive rigidities, and the red histogram shows negative rigidities.

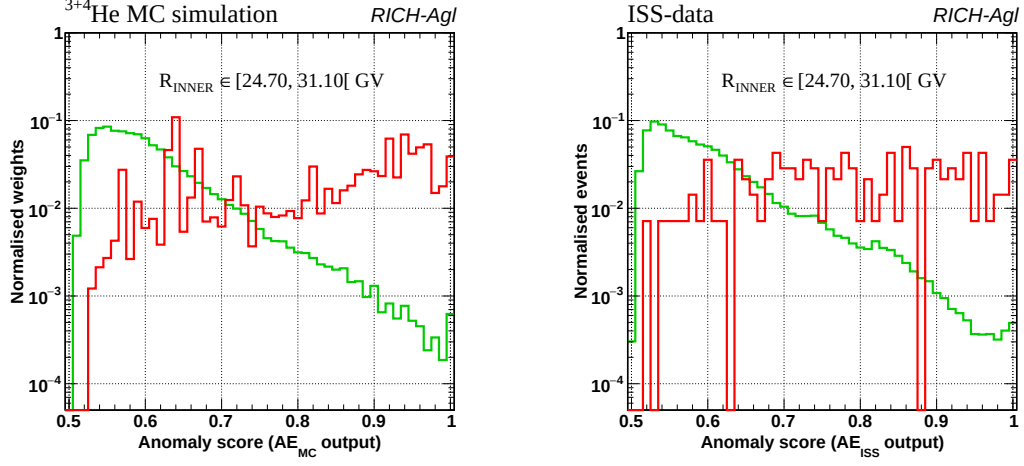


Figure B.39: Anomaly scores for events with a reconstructed rigidity between  $|R| \in [24.7, 31.1]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the  $AE_{MC}$  (left) and the  $AE_{ISS}$  (right). The green histogram shows positive rigidities, and the red histogram shows negative rigidities.

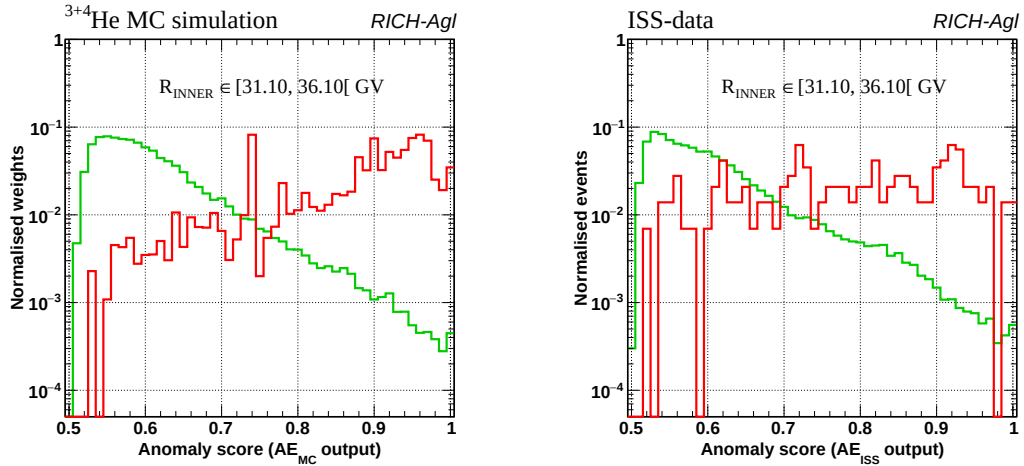


Figure B.40: Anomaly scores for events with a reconstructed rigidity between  $|R| \in [31.1, 36.1]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the  $AE_{MC}$  (left) and the  $AE_{ISS}$  (right). The green histogram shows positive rigidities, and the red histogram shows negative rigidities.

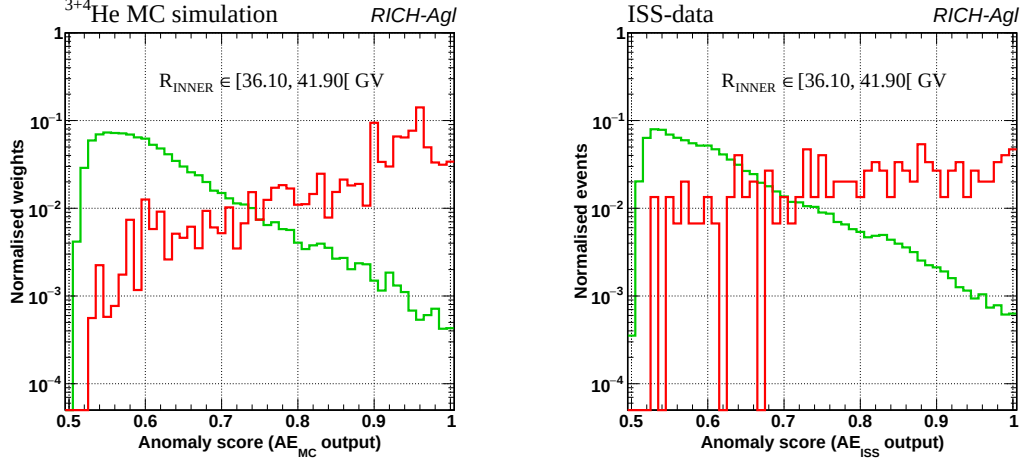


Figure B.41: Anomaly scores for events with a reconstructed rigidity between  $|R| \in [36.1, 41.9]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the  $AE_{MC}$  (left) and the  $AE_{ISS}$  (right). The green histogram shows positive rigidities, and the red histogram shows negative rigidities.

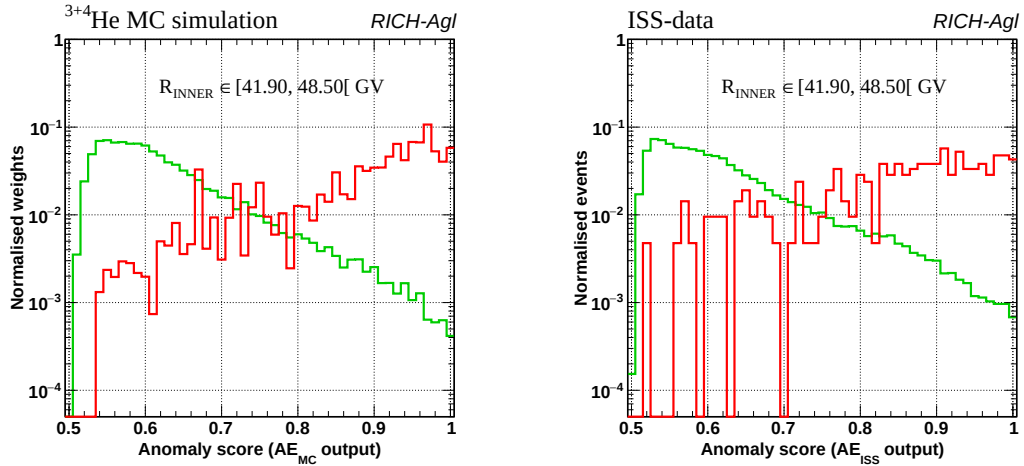


Figure B.42: Anomaly scores for events with a reconstructed rigidity between  $|R| \in [41.9, 48.5]$  GV and a reconstructed  $\beta_{AGL}$ , obtained using the  $AE_{MC}$  (left) and the  $AE_{ISS}$  (right). The green histogram shows positive rigidities, and the red histogram shows negative rigidities.



## B.3 Orthogonality

TOF

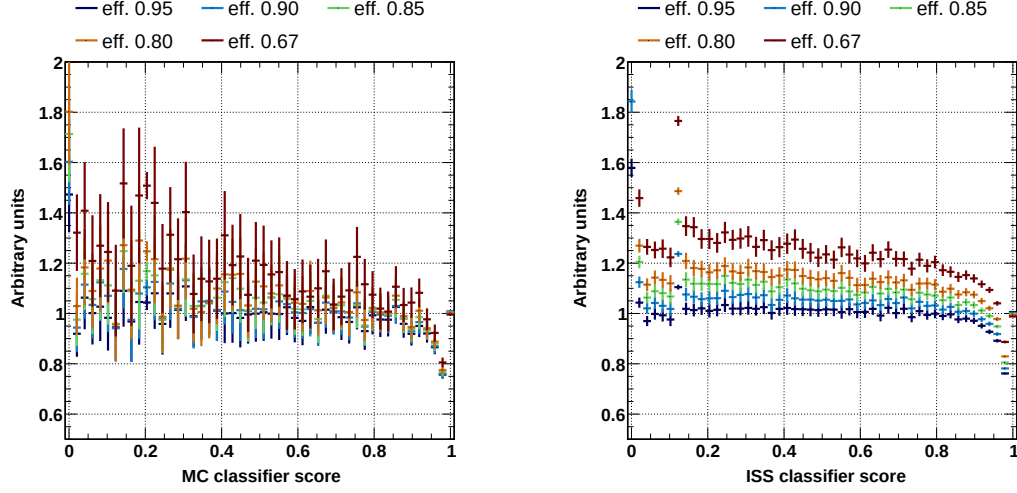


Figure B.43: Ratios between the score distribution without any cut and the distributions with a cut on the anomaly score for five different cut values. The left plot presents the ratios for the model trained on MC, and the plot on the right contains the results for the models trained on AMS data for events with measured rigidity in the following bin  $R \in [1.92, 4.02]$ , labelled as well-reconstructed (see equation 4.16), and having a TOF velocity measurement.

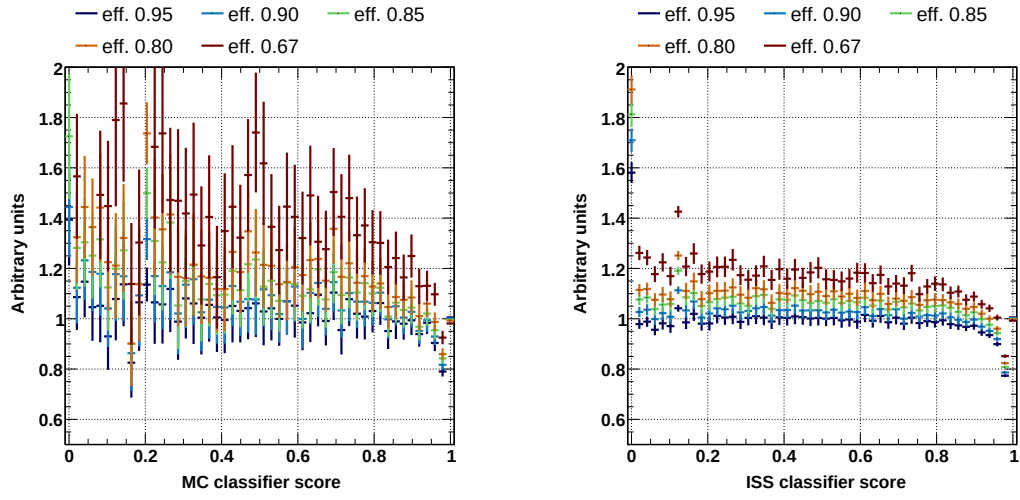


Figure B.44: Ratios between the score distribution without any cut and the distributions with a cut on the anomaly score for five different cut values. The left plot presents the ratios for the model trained on MC, and the plot on the right contains the results for the models trained on AMS data for events with a measured rigidity in the following bin  $R \in [4.02, 6.47]$ , labelled as well-reconstructed (see equation 4.16), and having a TOF velocity measurement.

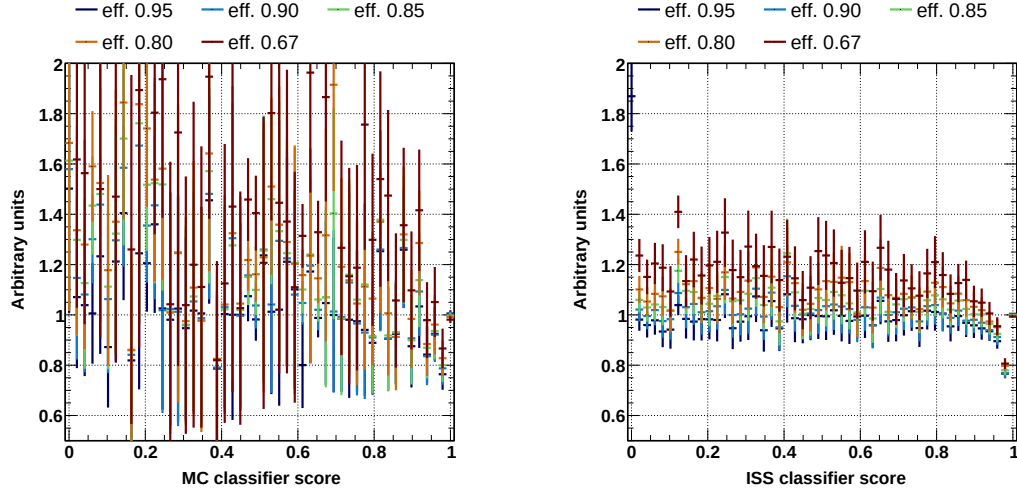


Figure B.45: Ratios between the score distribution without any cut and the distributions with a cut on the anomaly score for five different cut values. The left plot presents the ratios for the model trained on MC, and the plot on the right contains the results for the models trained on AMS data for events with a measured rigidity in the following bin  $R \in [4.02, 6.47]$ , labelled as well-reconstructed (see equation 4.16), and having a RICH-NaF velocity measurement.

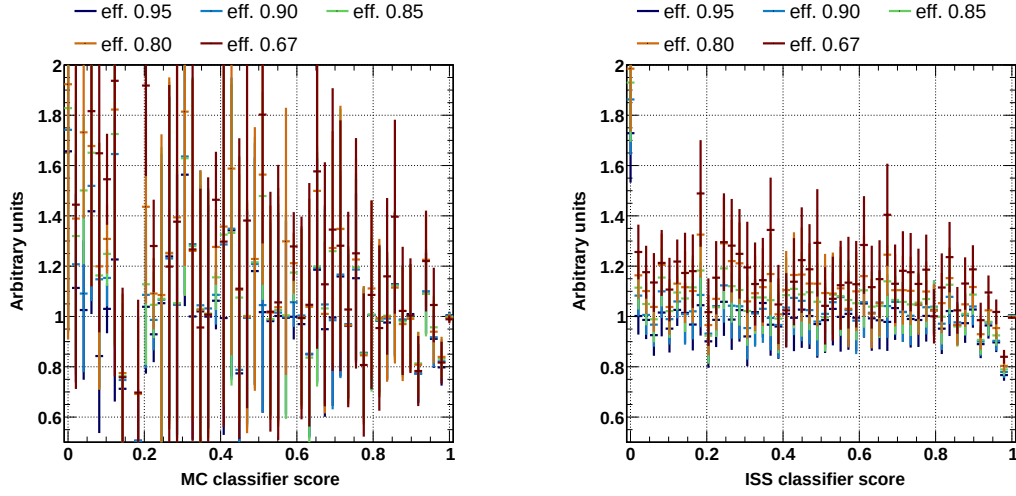


Figure B.46: CRatios between the score distribution without any cut and the distributions with a cut on the anomaly score for five different cut values. The left plot presents the ratios for the model trained on MC, and the plot on the right contains the results for the models trained on AMS data for events with a measured rigidity in the following bin  $R \in [6.47, 9.26]$ , labelled as well-reconstructed (see equation 4.16), and having a RICH-NaF velocity measurement.

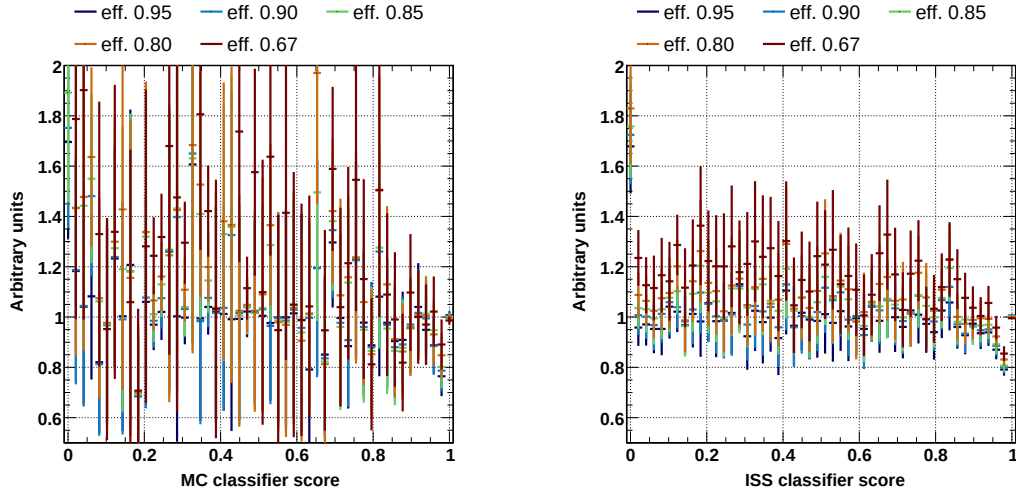


Figure B.47: Ratios between the score distribution without any cut and the distributions with a cut on the anomaly score for five different cut values. The left plot presents the ratios for the model trained on MC, and the plot on the right contains the results for the models trained on AMS data for events with a measured rigidity in the following bin  $R \in [9.26, 13.0]$ , labelled as well-reconstructed (see equation 4.16), and having a RICH-NaF velocity measurement.

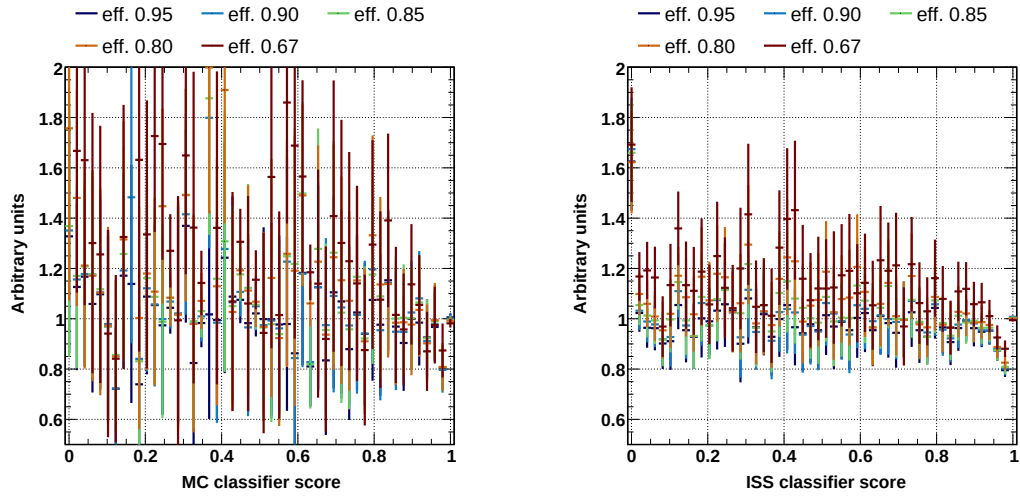


Figure B.48: Ratios between the score distribution without any cut and the distributions with a cut on the anomaly score for five different cut values. The left plot presents the ratios for the model trained on MC, and the plot on the right contains the results for the models trained on AMS data for events with a measured rigidity in the following bin  $R \in [13.0, 18.0]$ , labelled as well-reconstructed (see equation 4.16), and having a RICH-NaF velocity measurement.

## RICH-AGL

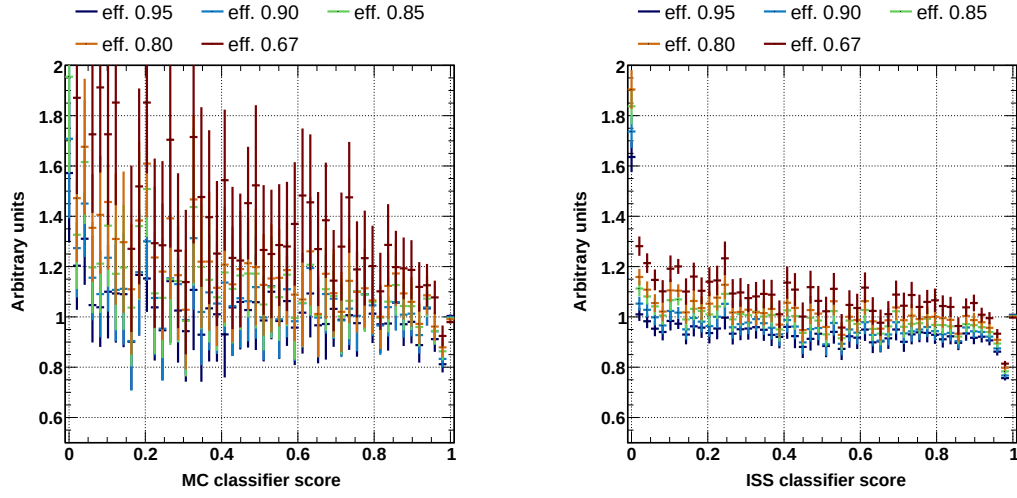


Figure B.49: Ratios between the score distribution without any cut and the distributions with a cut on the anomaly score for five different cut values. The left plot presents the ratios for the model trained on MC, and the plot on the right contains the results for the models trained on AMS data for events with a measured rigidity in the following bin  $R \in [6.47, 9.26]$ , labelled as well-reconstructed (see equation 4.16), and having a RICH-AGL velocity measurement.

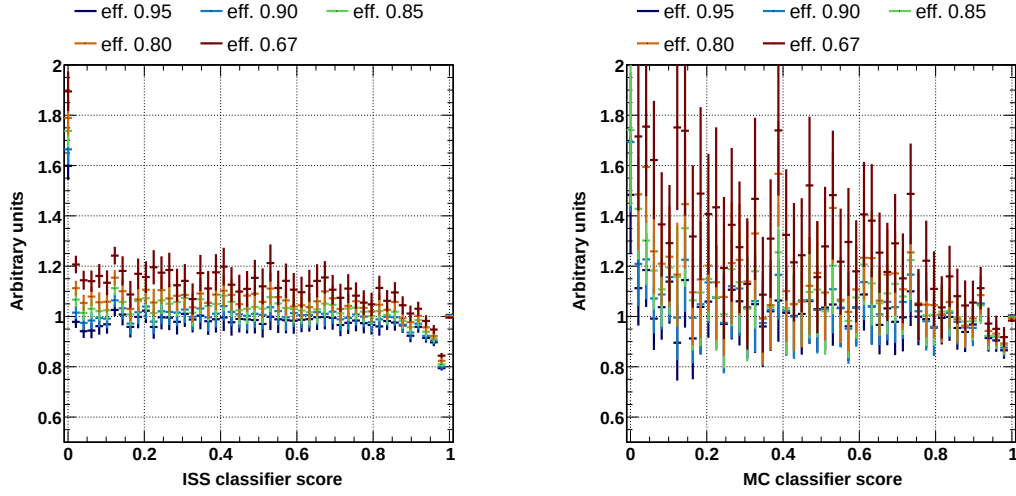


Figure B.50: Ratios between the score distribution without any cut and the distributions with a cut on the anomaly score for five different cut values. The left plot presents the ratios for the model trained on MC, and the plot on the right contains the results for the models trained on AMS data for events with a measured rigidity in the following bin  $R \in [9.26, 13.0]$ , labelled as well-reconstructed (see equation 4.16), and having a RICH-AGL velocity measurement.

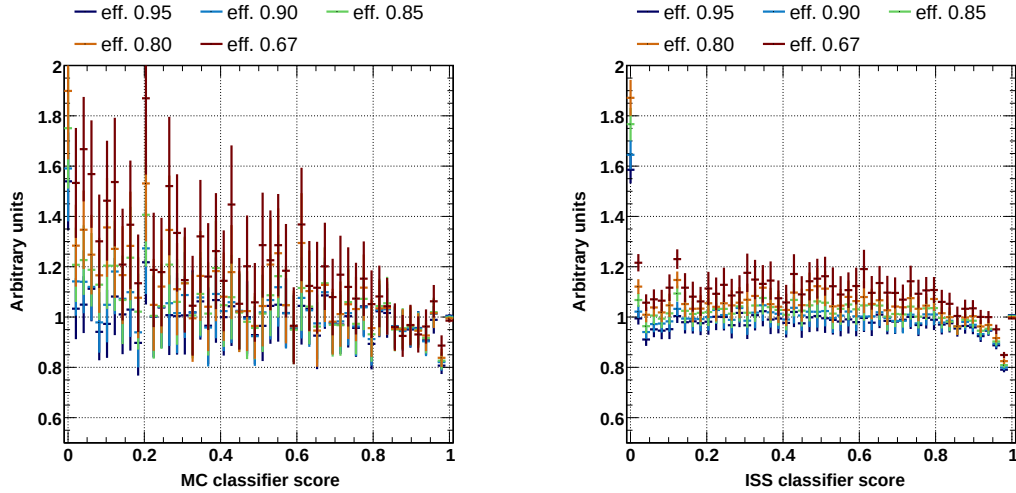


Figure B.51: Ratios between the score distribution without any cut and the distributions with a cut on the anomaly score for five different cut values. The left plot presents the ratios for the model trained on MC, and the plot on the right contains the results for the models trained on AMS data for events with a measured rigidity in the following bin  $R \in [13.0, 18.0]$ , labelled as well-reconstructed (see equation 4.16), and having a RICH-AGL velocity measurement.

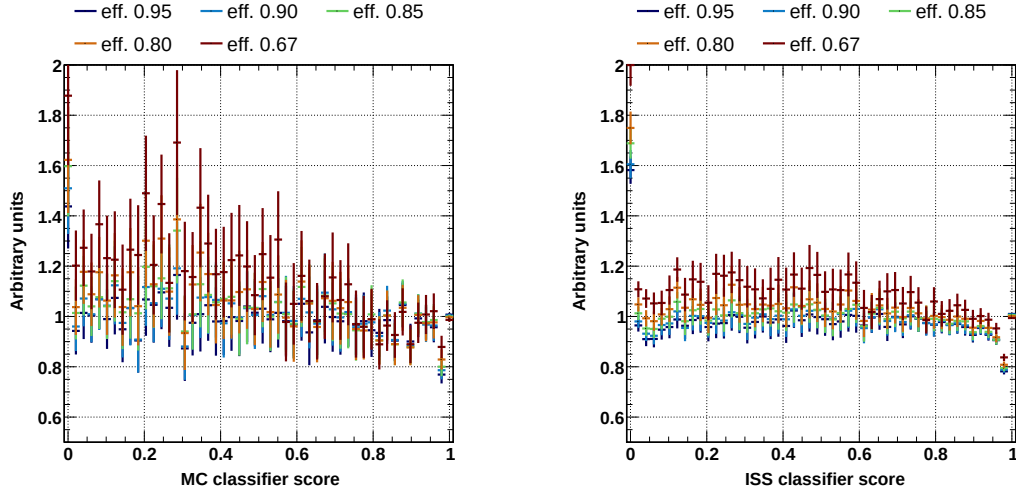


Figure B.52: Ratios between the score distribution without any cut and the distributions with a cut on the anomaly score for five different cut values. The left plot presents the ratios for the model trained on MC, and the plot on the right contains the results for the models trained on AMS data for events with a measured rigidity in the following bin  $R \in [18.0, 24.7]$ , labelled as well-reconstructed (see equation 4.16), and having a RICH-AGL velocity measurement.

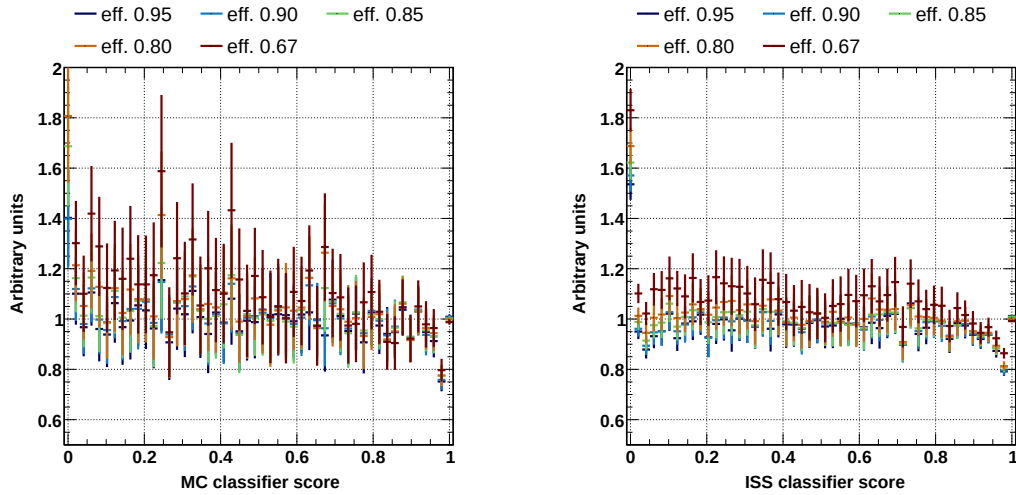


Figure B.53: Ratios between the score distribution without any cut and the distributions with a cut on the anomaly score for five different cut values. The left plot presents the ratios for the model trained on MC, and the plot on the right contains the results for the models trained on AMS data for events with a measured rigidity in the following bin  $R \in [24.7, 31.1]$ , labelled as well-reconstructed (see equation 4.16), and having a RICH-AGL velocity measurement.



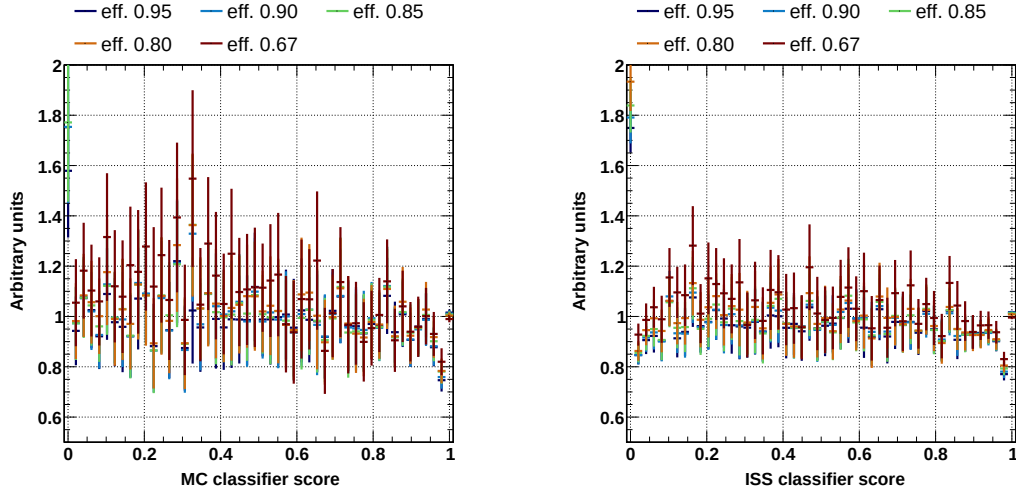


Figure B.54: Ratios between the score distribution without any cut and the distributions with a cut on the anomaly score for five different cut values. The left plot presents the ratios for the model trained on MC, and the plot on the right contains the results for the models trained on AMS data for events with a measured rigidity in the following bin  $R \in [31.1, 36.1]$ , labelled as well-reconstructed (see equation 4.16), and having a RICH-AGL velocity measurement.

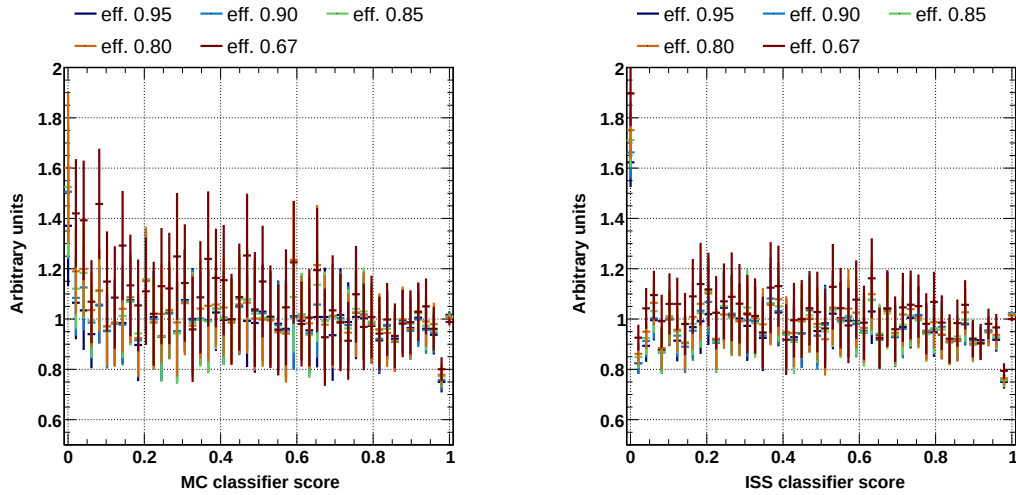


Figure B.55: Ratios between the score distribution without any cut and the distributions with a cut on the anomaly score for five different cut values. The left plot presents the ratios for the model trained on MC, and the plot on the right contains the results for the models trained on AMS data for events with a measured rigidity in the following bin  $R \in [36.1, 41.9]$ , labelled as well-reconstructed (see equation 4.16), and having a RICH-AGL velocity measurement.

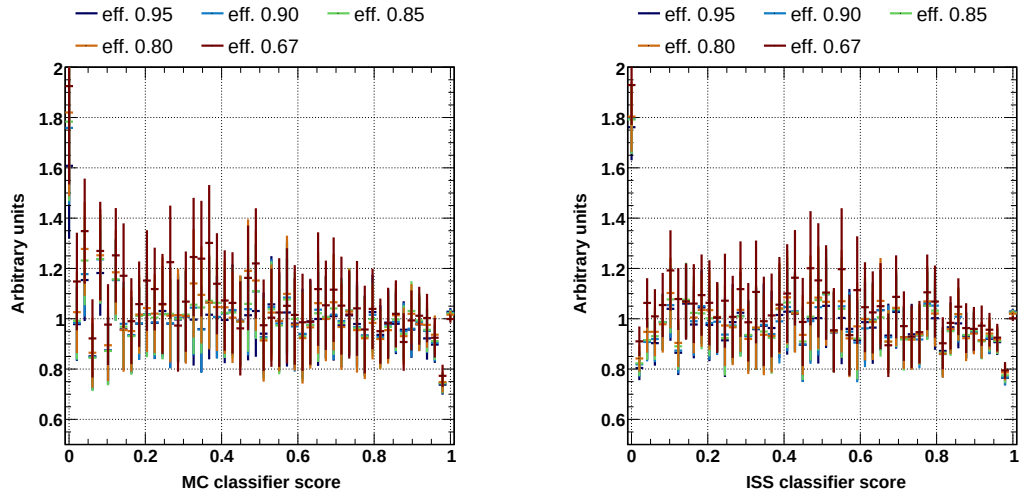


Figure B.56: Ratios between the score distribution without any cut and the distributions with a cut on the anomaly score for five different cut values. The left plot presents the ratios for the model trained on MC, and the plot on the right contains the results for the models trained on AMS data for events with a measured rigidity in the following bin  $R \in [41.9, 48.5]$ , labelled as well-reconstructed (see equation 4.16), and having a RICH-AGL velocity measurement.

## Appendix C

### Expected electron background

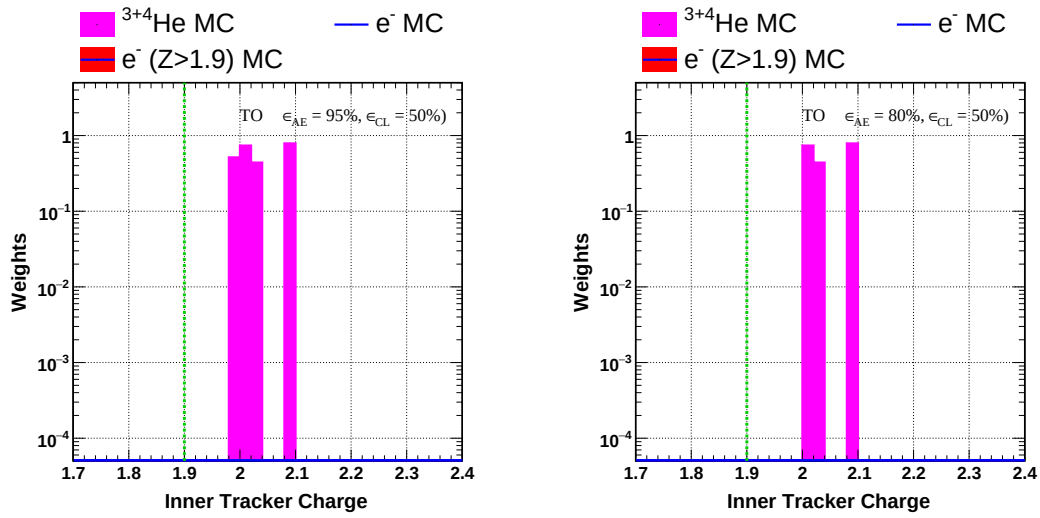


Figure C.1: Distributions of the inner tracker charge for simulated helium and electron events traversing the TOF. The left plot shows the result obtained using the (95%, 50%) pair, while the right plot presents the case with the (80%, 50%) pair. No electron contamination is present.

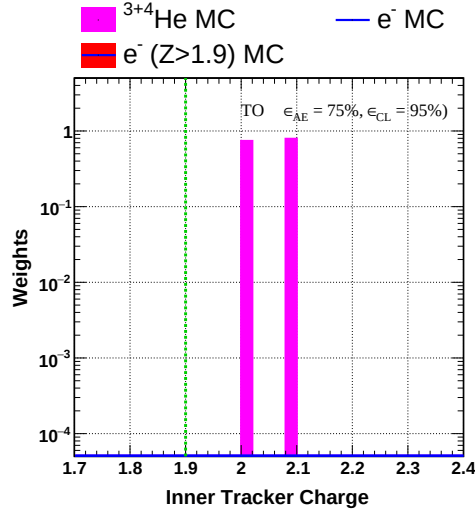


Figure C.2: Distributions of the inner tracker charge for simulated helium and electron events traversing the TOF, obtained using the (75%, 95%) pair. No electron contamination is present.

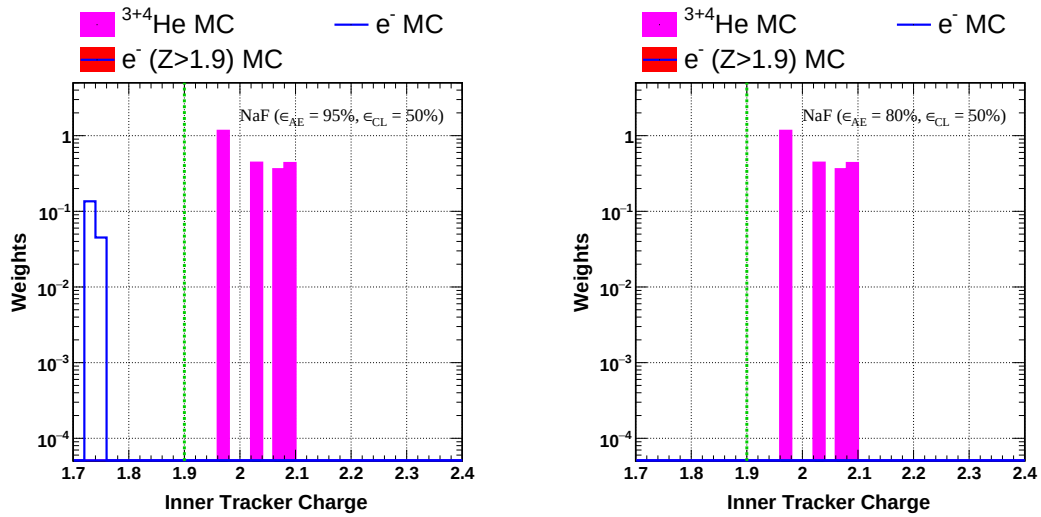


Figure C.3: Distributions of the inner tracker charge for simulated helium and electron events traversing the RICH-NaF. The left plot shows the result obtained using the (95%, 50%) pair, while the right plot presents the case with the (80%, 50%) pair. The electron contribution can be removed by requiring  $Z_{\text{TRK}} > 1.9$ .

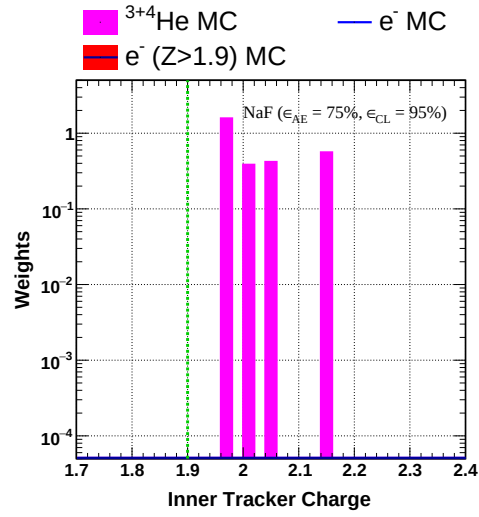


Figure C.4: Distributions of the inner tracker charge for simulated helium and electron events traversing the RICH-NaF, obtained using the (75%, 95%) pair. No electron contamination is present.

## Appendix D

# Exclusion limits on the ratio anti-He/He

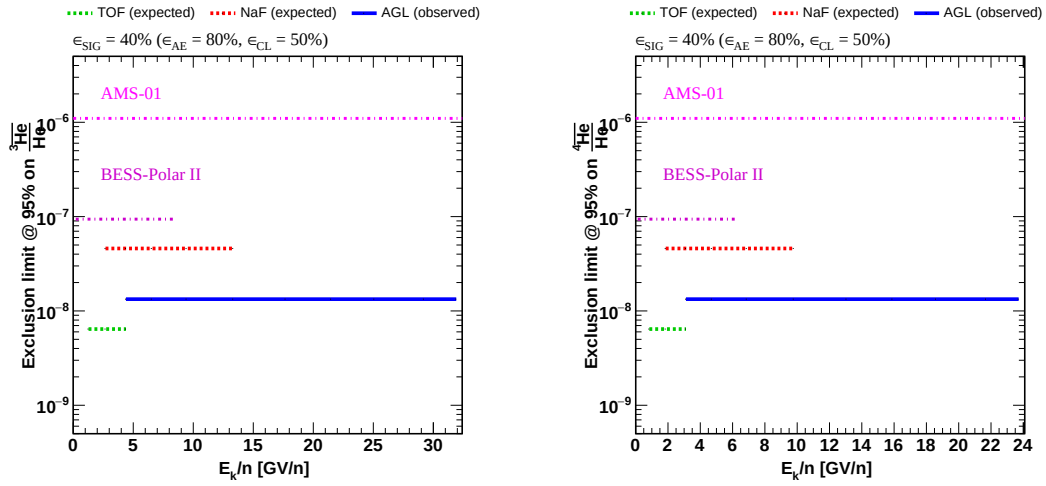


Figure D.1: (Left)  $\frac{{}^3\overline{\text{He}}}{\text{He}}$  exclusion limit, assuming all the candidates are  ${}^3\overline{\text{He}}$ , as a function of the kinetic energy per nucleon. (Right)  $\frac{{}^4\overline{\text{He}}}{\text{He}}$  exclusion limit, assuming all the candidates are  ${}^4\overline{\text{He}}$ , as a function of the kinetic energy per nucleon. Both plots show the results obtained using the combination ( $\epsilon_{\text{AE}} = 80\%$ ,  $\epsilon_{\text{CL}} = 50\%$ ) ( $\epsilon_{\text{SIG}} = \epsilon_{\text{AE}} \times \epsilon_{\text{CL}} = 40\%$ ), which are presented together with the previous limits from AMS-01 and BESS-Polar II [79, 144].

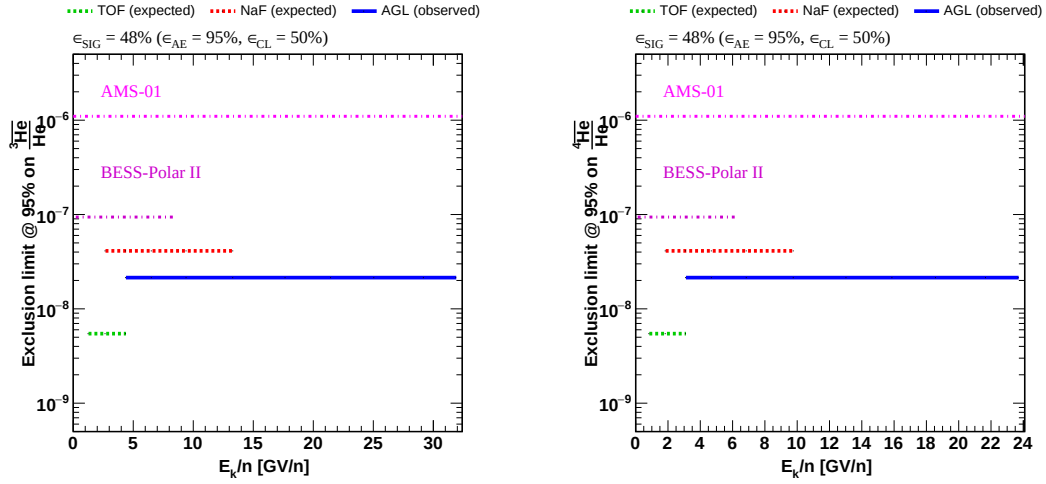


Figure D.2: (Left)  $\frac{^3\text{He}}{\text{He}}$  exclusion limit, assuming all the candidates are  $^3\text{He}$ , as a function of the kinetic energy per nucleon. (Right)  $\frac{^4\text{He}}{\text{He}}$  exclusion limit, assuming all the candidates are  $^4\text{He}$ , as a function of the kinetic energy per nucleon. Both plots show the results obtained using the combination ( $\epsilon_{\text{AE}} = 95\%$ ,  $\epsilon_{\text{CL}} = 50\%$ ) ( $\epsilon_{\text{SIG}} = \epsilon_{\text{AE}} \times \epsilon_{\text{CL}} = 48\%$ ) with the additional cuts on the minimum distance from the track aluminium jigs and  $\chi^2_{\text{Y}}$ , which are presented together with the previous limits from AMS-01 and BESS-Polar II [79, 144].

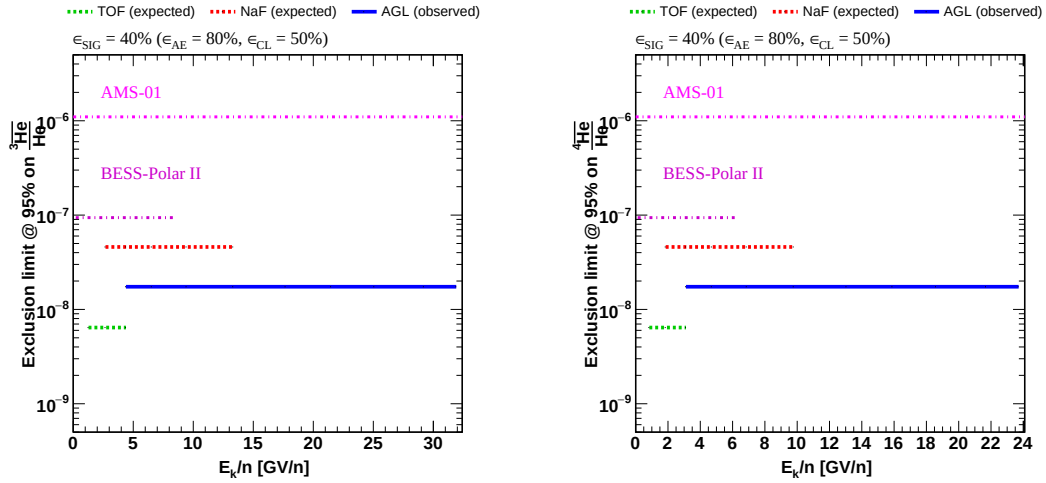


Figure D.3: (Left)  $\frac{^3\text{He}}{\text{He}}$  exclusion limit, assuming all the candidates are  $^3\text{He}$ , as a function of the kinetic energy per nucleon. (Right)  $\frac{^4\text{He}}{\text{He}}$  exclusion limit, assuming all the candidates are  $^4\text{He}$ , as a function of the kinetic energy per nucleon. Both plots show the results obtained using the combination ( $\epsilon_{\text{AE}} = 80\%$ ,  $\epsilon_{\text{CL}} = 50\%$ ) ( $\epsilon_{\text{SIG}} = \epsilon_{\text{AE}} \times \epsilon_{\text{CL}} = 40\%$ ) with the additional cuts on the minimum distance from the track aluminium jigs and  $\chi^2_{\text{Y}}$ , which are presented together with the previous limits from AMS-01 and BESS-Polar II [79, 144].

# Bibliography

- [1] S. Raiz F. Rossi M. Dorigo. *Study of the  $\pi^0$  reconstruction efficiency for the decay of  $B^0$  into pairs of neutral pions at the Belle II experiment*. 2022. URL: [https://docs.belle2.org/pub\\_data/documents/3948/](https://docs.belle2.org/pub_data/documents/3948/) (visited on 10/22/2022).
- [2] Steven Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (21 1967), p. 1264. DOI: 10.1103/PhysRevLett.19.1264.
- [3] Sheldon L. Glashow. “Partial-symmetries of weak interactions”. In: *Nuclear Physics* 22 (1961), p. 579. ISSN: 0029-5582. DOI: [https://doi.org/10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2).
- [4] Abdus Salam. “Weak and Electromagnetic Interactions”. In: *Conf. Proc. C* 680519 (1968), p. 367. DOI: 10.1142/9789812795915\_0034.
- [5] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (1964). Ed. by J.C. Taylor, p. 321. DOI: 10.1103/PhysRevLett.13.321.
- [6] Peter W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys. Rev. Lett.* 13 (1964). Ed. by J.C. Taylor, p. 508. DOI: 10.1103/PhysRevLett.13.508.
- [7] G.S. Guralnik, C.R. Hagen, and T.W.B. Kibble. “Global Conservation Laws and Massless Particles”. In: *Phys. Rev. Lett.* 13 (1964). Ed. by J.C. Taylor, p. 585. DOI: 10.1103/PhysRevLett.13.585.
- [8] Richard H. Cyburt et al. “Big Bang Nucleosynthesis: 2015”. In: *Rev. Mod. Phys.* 88 (2016), p. 015004. DOI: 10.1103/RevModPhys.88.015004. arXiv: 1505.01076 [astro-ph.CO].
- [9] Edwin Hubble. “A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae”. In: *Proceedings of the National Academy of Science* 15.3 (Mar. 1929), pp. 168–173. DOI: 10.1073/pnas.15.3.168.
- [10] Albert Einstein. “The foundation of the general theory of relativity.” In: *Annalen Phys.* 49.7 (1916). Ed. by Jong-Ping Hsu and D. Fine, pp. 769–822. DOI: 10.1002/andp.19163540702.
- [11] S. Navas et al. “Review of particle physics”. In: *Phys. Rev. D* 110.3 (2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001.
- [12] N. Aghanim et al. “Planck 2018 results. VI. Cosmological parameters”. In: *Astron. Astrophys.* 641 (2020). [Erratum: *Astron. Astrophys.* 652, C4 (2021)], A6. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209 [astro-ph.CO].



- [13] R. A. Alpher, H. Bethe, and G. Gamow. “The Origin of Chemical Elements”. In: *Phys. Rev.* 73 (7 Apr. 1948), pp. 803–804. DOI: 10.1103/PhysRev.73.803. URL: <https://link.aps.org/doi/10.1103/PhysRev.73.803>.
- [14] Ruth Durrer. “The cosmic microwave background: the history of its experimental investigation and its significance for cosmology”. In: *Class. Quant. Grav.* 32.12 (2015), p. 124007. DOI: 10.1088/0264-9381/32/12/124007. arXiv: 1506.01907 [astro-ph.CO].
- [15] David S. Pereira et al. “Baryogenesis: A Symmetry Breaking in the Primordial Universe Revisited”. In: *Symmetry* 16.1 (2024), p. 13. DOI: 10.3390/sym16010013. arXiv: 2312.14080 [gr-qc].
- [16] Robert V. Wagoner, William A. Fowler, and F. Hoyle. “On the Synthesis of Elements at Very High Temperatures”. In: 148 (Apr. 1967), p. 3. DOI: 10.1086/149126.
- [17] Karsten Jedamzik and Maxim Pospelov. “Big Bang Nucleosynthesis and Particle Dark Matter”. In: *New J. Phys.* 11 (2009), p. 105028. DOI: 10.1088/1367-2630/11/10/105028. arXiv: 0906.2087 [hep-ph].
- [18] Maxim Pospelov and Josef Pradler. “Big Bang Nucleosynthesis as a Probe of New Physics”. In: *Ann. Rev. Nucl. Part. Sci.* 60 (2010), pp. 539–568. DOI: 10.1146/annurev.nucl.012809.104521. arXiv: 1011.1054 [hep-ph].
- [19] Brian D. Fields. “The primordial lithium problem”. In: *Ann. Rev. Nucl. Part. Sci.* 61 (2011), pp. 47–68. DOI: 10.1146/annurev-nucl-102010-130445. arXiv: 1203.3551 [astro-ph.CO].
- [20] R. I. Epstein, J. M. Lattimer, and D. N. Schramm. “The origin of deuterium”. In: 263 (Sept. 1976), pp. 198–202. DOI: 10.1038/263198a0.
- [21] Freeke van de Voort et al. “On the deuterium abundance and the importance of stellar mass loss in the interstellar and intergalactic medium”. In: *Mon. Not. Roy. Astron. Soc.* 477.1 (2018), pp. 80–92. DOI: 10.1093/mnras/sty591. arXiv: 1704.08254 [astro-ph.GA].
- [22] Y. I. Izotov, T. X. Thuan, and N. G. Guseva. “A new determination of the primordial He abundance using the He I 10830 Å emission line: cosmological implications”. In: *Monthly Notices of the Royal Astronomical Society* 445.1 (Sept. 2014), pp. 778–793. ISSN: 0035-8711. DOI: 10.1093/mnras/stu1771. eprint: <https://academic.oup.com/mnras/article-pdf/445/1/778/18468145/stu1771.pdf>. URL: <https://doi.org/10.1093/mnras/stu1771>.
- [23] David S. Aguado et al. “Back to the Lithium Plateau with J0023+0307 with [Fe/H]<-6”. In: *Astrophys. J. Lett.* 874 (2019), p. L21. DOI: 10.3847/2041-8213/ab1076. arXiv: 1904.04892 [astro-ph.SR].
- [24] Dana S. Balser and T. M. Bania. “Green Bank Telescope Observations of  $^3\text{He}^+$ : H II Regions”. In: 156.6, 280 (Dec. 2018), p. 280. DOI: 10.3847/1538-3881/aaeb2b. arXiv: 1810.09422 [astro-ph.GA].

- [25] Samuel M. Leach et al. “Cosmological parameter estimation and the inflationary cosmology”. In: *Phys. Rev. D* 66 (2002), p. 023515. DOI: 10.1103/PhysRevD.66.023515. arXiv: astro-ph/0202094.
- [26] Masataka Fukugita and P. James E. Peebles. “The Cosmic energy inventory”. In: *Astrophys. J.* 616 (2004), pp. 643–668. DOI: 10.1086/425155. arXiv: astro-ph/0406095.
- [27] Renyue Cen and Jeremiah P. Ostriker. “Where are the baryons?” In: *Astrophys. J.* 514 (1999), pp. 1–6. DOI: 10.1086/306949. arXiv: astro-ph/9806281.
- [28] Gerard Jungman et al. “Cosmological parameter determination with microwave background maps”. In: *Phys. Rev. D* 54 (1996), pp. 1332–1344. DOI: 10.1103/PhysRevD.54.1332. arXiv: astro-ph/9512139.
- [29] R. Adam et al. “Planck 2015 results. X. Diffuse component separation: Foreground maps”. In: *Astron. Astrophys.* 594 (2016), A10. DOI: 10.1051/0004-6361/201525967. arXiv: 1502.01588 [astro-ph.CO].
- [30] Gary Steigman, David N. Schramm, and James E. Gunn. “Cosmological limits to the number of massive leptons”. In: *Physics Letters B* 66.2 (1977), pp. 202–204. ISSN: 0370-2693. DOI: [https://doi.org/10.1016/0370-2693\(77\)90176-9](https://doi.org/10.1016/0370-2693(77)90176-9). URL: <https://www.sciencedirect.com/science/article/pii/0370269377901769>.
- [31] Tsung-Han Yeh et al. “Probing physics beyond the standard model: limits from BBN and the CMB independently and combined”. In: *JCAP* 10 (2022), p. 046. DOI: 10.1088/1475-7516/2022/10/046. arXiv: 2207.13133 [astro-ph.CO].
- [32] A. D. Sakharov. “Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe”. In: *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967), pp. 32–35. DOI: 10.1070/PU1991v034n05ABEH002497.
- [33] Antonio Riotto and Mark Trodden. “Recent progress in baryogenesis”. In: *Ann. Rev. Nucl. Part. Sci.* 49 (1999), pp. 35–75. DOI: 10.1146/annurev.nucl.49.1.35. arXiv: hep-ph/9901362.
- [34] Björn Garbrecht. “Why is there more matter than antimatter? Computational methods for leptogenesis and electroweak baryogenesis”. In: *Progress in Particle and Nuclear Physics* 110 (2020), p. 103727. ISSN: 0146-6410. DOI: <https://doi.org/10.1016/j.ppnp.2019.103727>. URL: <https://www.sciencedirect.com/science/article/pii/S0146641019300626>.
- [35] G. 't Hooft. “Symmetry Breaking through Bell-Jackiw Anomalies”. In: *Phys. Rev. Lett.* 37 (1 July 1976), pp. 8–11. DOI: 10.1103/PhysRevLett.37.8. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.37.8>.
- [36] Nicola Cabibbo. “Unitary Symmetry and Leptonic Decays”. In: *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.10.531>.

- [37] Makoto Kobayashi and Toshihide Maskawa. “CP-Violation in the Renormalizable Theory of Weak Interaction”. In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. ISSN: 0033-068X. DOI: 10.1143/PTP.49.652. eprint: <https://academic.oup.com/ptp/article-pdf/49/2/652/5257692/49-2-652.pdf>. URL: <https://doi.org/10.1143/PTP.49.652>.
- [38] David E. Morrissey and Michael J. Ramsey-Musolf. “Electroweak baryogenesis”. In: *New J. Phys.* 14 (2012), p. 125003. DOI: 10.1088/1367-2630/14/12/125003. arXiv: 1206.2942 [hep-ph].
- [39] Sacha Davidson, Enrico Nardi, and Yosef Nir. “Leptogenesis”. In: *Phys. Rept.* 466 (2008), pp. 105–177. DOI: 10.1016/j.physrep.2008.06.002. arXiv: 0802.2962 [hep-ph].
- [40] Michael Dine, Lisa Randall, and Scott D. Thomas. “Baryogenesis from flat directions of the supersymmetric standard model”. In: *Nucl. Phys. B* 458 (1996), pp. 291–326. DOI: 10.1016/0550-3213(95)00538-2. arXiv: hep-ph/9507453.
- [41] Dietrich Bodeker and Wilfried Buchmuller. “Baryogenesis from the weak scale to the grand unification scale”. In: *Rev. Mod. Phys.* 93.3 (2021), p. 035004. DOI: 10.1103/RevModPhys.93.035004. arXiv: 2009.07294 [hep-ph].
- [42] Enrico Fermi. “On the Origin of the Cosmic Radiation”. In: *Phys. Rev.* 75 (1949), pp. 1169–1174. DOI: 10.1103/PhysRev.75.1169.
- [43] Vivian Poulin et al. “Where do the AMS-02 antihelium events come from?” In: *Phys. Rev. D* 99.2 (2019), p. 023016. DOI: 10.1103/PhysRevD.99.023016. arXiv: 1808.08961 [astro-ph.HE].
- [44] Oscar Adriani et al. “An anomalous positron abundance in cosmic rays with energies 1.5–100 GeV”. In: *Nature* 458 (2009), pp. 607–609. DOI: 10.1038/nature07942. arXiv: 0810.4995 [astro-ph].
- [45] M. Ackermann et al. “Measurement of separate cosmic-ray electron and positron spectra with the Fermi Large Area Telescope”. In: *Phys. Rev. Lett.* 108 (2012), p. 011103. DOI: 10.1103/PhysRevLett.108.011103. arXiv: 1109.0521 [astro-ph.HE].
- [46] M. Aguilar et al. “First Result from the Alpha Magnetic Spectrometer on the International Space Station: Precision Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5–350 GeV”. In: *Phys. Rev. Lett.* 110 (2013), p. 141102. DOI: 10.1103/PhysRevLett.110.141102.
- [47] L. Accardo et al. “High Statistics Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5–500 GeV with the Alpha Magnetic Spectrometer on the International Space Station”. In: *Phys. Rev. Lett.* 113 (2014), p. 121101. DOI: 10.1103/PhysRevLett.113.121101.
- [48] M. Aguilar et al. “Electron and Positron Fluxes in Primary Cosmic Rays Measured with the Alpha Magnetic Spectrometer on the International Space Station”. In: *Phys. Rev. Lett.* 113 (2014), p. 121102. DOI: 10.1103/PhysRevLett.113.121102.

- [49] M. Aguilar et al. “Precision Measurement of the ( $e^+ + e$ ) Flux in Primary Cosmic Rays from 0.5 GeV to 1 TeV with the Alpha Magnetic Spectrometer on the International Space Station”. In: *Phys. Rev. Lett.* 113 (2014), p. 221102. DOI: 10.1103/PhysRevLett.113.221102.
- [50] M. Aguilar et al. “Towards Understanding the Origin of Cosmic-Ray Positrons”. In: *Phys. Rev. Lett.* 122.4 (2019), p. 041102. DOI: 10.1103/PhysRevLett.122.041102.
- [51] M. Aguilar et al. “Antiproton Flux, Antiproton-to-Proton Flux Ratio, and Properties of Elementary Particle Fluxes in Primary Cosmic Rays Measured with the Alpha Magnetic Spectrometer on the International Space Station”. In: *Phys. Rev. Lett.* 117.9 (2016), p. 091103. DOI: 10.1103/PhysRevLett.117.091103.
- [52] Alessandro Cuoco et al. “Scrutinizing the evidence for dark matter in cosmic-ray antiprotons”. In: *Phys. Rev. D* 99.10 (2019), p. 103014. DOI: 10.1103/PhysRevD.99.103014. arXiv: 1903.01472 [astro-ph.HE].
- [53] Ilias Cholis, Tim Linden, and Dan Hooper. “A Robust Excess in the Cosmic-Ray Antiproton Spectrum: Implications for Annihilating Dark Matter”. In: *Phys. Rev. D* 99.10 (2019), p. 103026. DOI: 10.1103/PhysRevD.99.103026. arXiv: 1903.02549 [astro-ph.HE].
- [54] P. von Doetinchem et al. “Cosmic-ray antinuclei as messengers of new physics: status and outlook for the new decade”. In: *Journal of Cosmology and Astroparticle Physics* 2020.08 (Aug. 2020), p. 035. DOI: 10.1088/1475-7516/2020/08/035. URL: <https://doi.org/10.1088/1475-7516/2020/08/035>.
- [55] T. Aramaki et al. “Antideuteron Sensitivity for the GAPS Experiment”. In: *Astropart. Phys.* 74 (2016), pp. 6–13. DOI: 10.1016/j.astropartphys.2015.09.001. arXiv: 1506.02513 [astro-ph.HE].
- [56] Adam Coogan and Stefano Profumo. “Origin of the tentative AMS antihelium events”. In: *Phys. Rev. D* 96.8 (2017), p. 083020. DOI: 10.1103/PhysRevD.96.083020. arXiv: 1705.09664 [astro-ph.HE].
- [57] Marco Cirelli et al. “Anti-helium from Dark Matter annihilations”. In: *JHEP* 08 (2014), p. 009. DOI: 10.1007/JHEP08(2014)009. arXiv: 1401.4017 [hep-ph].
- [58] Eric Carlson et al. “Antihelium from Dark Matter”. In: *Phys. Rev. D* 89.7 (2014), p. 076005. DOI: 10.1103/PhysRevD.89.076005. arXiv: 1401.2461 [hep-ph].
- [59] G. Steigman. “Observational tests of antimatter cosmologies.” In: 14 (Jan. 1976), pp. 339–372. DOI: 10.1146/annurev.aa.14.090176.002011.
- [60] C. Bambi and A.D. Dolgov. “Antimatter in the Milky Way”. In: *Nuclear Physics B* 784.1 (2007), pp. 132–150. ISSN: 0550-3213. DOI: <https://doi.org/10.1016/j.nuclphysb.2007.06.010>. URL: <https://www.sciencedirect.com/science/article/pii/S0550321307004737>.

- [61] Michael Korsmeier, Fiorenza Donato, and Nicolao Fornengo. “Prospects to verify a possible dark matter hint in cosmic antiprotons with antideuterons and antihelium”. In: *Phys. Rev. D* 97 (10 May 2018), p. 103011. DOI: 10 . 1103/PhysRevD . 97 . 103011. URL: <https://link.aps.org/doi/10.1103/PhysRevD.97.103011>.
- [62] Stanislaw Mrowczynski. “Production of light nuclei in the thermal and coalescence models”. In: *Acta Phys. Polon. B* 48 (2017), p. 707. DOI: 10 . 5506/APhysPolB . 48 . 707. arXiv: 1607.02267 [nucl-th].
- [63] T. Anticic et al. “Production of deuterium, tritium, and  $^3\text{He}$  in central Pb + Pb collisions at 20A, 30A, 40A, 80A, and 158A GeV at the CERN Super Proton Synchrotron”. In: *Phys. Rev. C* 94 (4 Oct. 2016), p. 044906. DOI: 10 . 1103/PhysRevC . 94 . 044906. URL: <https://link.aps.org/doi/10.1103/PhysRevC.94.044906>.
- [64] S. Acharya et al. “Production of deuterons, tritons,  $^3\text{He}$  nuclei, and their antinuclei in  $pp$  collisions at  $\sqrt{s} = 0.9, 2.76$ , and 7 TeV”. In: *Phys. Rev. C* 97 (2 Feb. 2018), p. 024615. DOI: 10 . 1103/PhysRevC . 97 . 024615. URL: <https://link.aps.org/doi/10.1103/PhysRevC.97.024615>.
- [65] S. Acharya et al. “Multiplicity dependence of (anti-)deuteron production in  $pp$  collisions at  $\sqrt{s}=7\text{TeV}$ ”. In: *Physics Letters B* 794 (2019), pp. 50–63. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2019.05.028>. URL: <https://www.sciencedirect.com/science/article/pii/S0370269319303387>.
- [66] S. Acharya et al. “(Anti-)deuteron production in  $pp$  collisions at  $\sqrt{s} = 13\text{ TeV}$ ”. In: *Eur. Phys. J. C* 80.9 (2020), p. 889. DOI: 10.1140/epjc/s10052-020-8256-4. arXiv: 2003.03184 [nucl-ex].
- [67] S. Acharya et al. “Production of (anti-) $^3\text{He}$  and (anti-) $^3\text{H}$  in  $p\text{-Pb}$  collisions at  $\sqrt{s_{NN}} = 5.02\text{ TeV}$ ”. In: *Phys. Rev. C* 101 (4 Apr. 2020), p. 044906. DOI: 10 . 1103/PhysRevC . 101 . 044906. URL: <https://link.aps.org/doi/10.1103/PhysRevC.101.044906>.
- [68] S. Acharya et al. “Multiplicity dependence of light (anti-)nuclei production in  $p\text{-Pb}$  collisions at  $\sqrt{s_{NN}}=5.02\text{ TeV}$ ”. In: *Physics Letters B* 800 (2020), p. 135043. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2019.135043>. URL: <https://www.sciencedirect.com/science/article/pii/S0370269319307658>.
- [69] S. Acharya et al. “Production of  $4\text{He}$  and  $\bar{4}\text{He}$  in Pb–Pb collisions at  $\sqrt{s_{NN}}=2.76\text{TeV}$  at the LHC”. In: *Nuclear Physics A* 971 (2018), pp. 1–20. ISSN: 0375-9474. DOI: <https://doi.org/10.1016/j.nuclphysa.2017.12.004>. URL: <https://www.sciencedirect.com/science/article/pii/S0375947417304839>.
- [70] Shreyasi Acharya et al. “Production of charged pions, kaons, and (anti-)protons in Pb-Pb and inelastic  $pp$  collisions at  $\sqrt{s_{NN}} = 5.02\text{ TeV}$ ”. In: *Phys. Rev. C* 101.4 (2020), p. 044907. DOI: 10.1103/PhysRevC.101.044907. arXiv: 1910.07678 [nucl-ex].

- [71] Diego-Mauricio Gomez-Coral et al. “Deuteron and Antideuteron Production Simulation in Cosmic-Ray Interactions”. In: *Phys. Rev. D* 98.2 (2018), p. 023012. DOI: 10.1103/PhysRevD.98.023012. arXiv: 1806.09303 [astro-ph.HE].
- [72] Anirvan Shukla et al. “Large-scale simulations of antihelium production in cosmic-ray interactions”. In: *Phys. Rev. D* 102 (6 Sept. 2020), p. 063004. DOI: 10.1103/PhysRevD.102.063004. URL: <https://link.aps.org/doi/10.1103/PhysRevD.102.063004>.
- [73] A. Schwarzschild and C. Zupancic. “Production of Tritons, Deuterons, Nucleons, and Mesons by 30-GeV Protons on A-1, Be, and Fe Targets”. In: *Phys. Rev.* 129 (1963), pp. 854–862. DOI: 10.1103/PhysRev.129.854.
- [74] H. Sato and K. Yazaki. “On the coalescence model for high energy nuclear reactions”. In: *Physics Letters B* 98.3 (1981), pp. 153–157. ISSN: 0370-2693. DOI: [https://doi.org/10.1016/0370-2693\(81\)90976-X](https://doi.org/10.1016/0370-2693(81)90976-X). URL: <https://www.sciencedirect.com/science/article/pii/037026938190976X>.
- [75] Eric Carlson et al. “Antihelium from dark matter”. In: *Phys. Rev. D* 89 (7 Apr. 2014), p. 076005. DOI: 10.1103/PhysRevD.89.076005. URL: <https://link.aps.org/doi/10.1103/PhysRevD.89.076005>.
- [76] S. Acharya et al. “Measurement of anti- $^3\text{He}$  nuclei absorption in matter and impact on their propagation in the Galaxy”. In: *Nature Physics* 19.1 (Dec. 2022), pp. 61–71. ISSN: 1745-2481. DOI: 10.1038/s41567-022-01804-8. URL: <http://dx.doi.org/10.1038/s41567-022-01804-8>.
- [77] Robin Sonnabend. “Search for antihelium nuclei in cosmic rays with the AMS-02 experiment on the International Space Station”. PhD thesis. RWTH Aachen University, RWTH Aachen U., 2023.
- [78] Francesca Bellini and Alexander P. Kalweit. “Testing production scenarios for (anti-)(hyper-)nuclei and exotica at energies available at the CERN Large Hadron Collider”. In: *Phys. Rev. C* 99 (5 May 2019), p. 054905. DOI: 10.1103/PhysRevC.99.054905. URL: <https://link.aps.org/doi/10.1103/PhysRevC.99.054905>.
- [79] K. Abe et al. “Search for Antihelium with the BESS-Polar Spectrometer”. In: *Physical Review Letters* 108.13 (Mar. 2012). ISSN: 1079-7114. DOI: 10.1103/physrevlett.108.131301. URL: <http://dx.doi.org/10.1103/PhysRevLett.108.131301>.
- [80] Peter von Ballmoos. “Antimatter in the Universe: constraints from gamma-ray astronomy”. In: *Hyperfine Interactions* 228.1–3 (Feb. 2014), pp. 91–100. ISSN: 1572-9540. DOI: 10.1007/s10751-014-1024-9. URL: <http://dx.doi.org/10.1007/s10751-014-1024-9>.
- [81] A. G. Cohen, A. De Rújula, and S. L. Glashow. “A Matter-Antimatter Universe?” In: *The Astrophysical Journal* 495.2 (Mar. 1998), p. 539. DOI: 10.1086/305328. URL: <https://doi.org/10.1086/305328>.

- [82] M.-C. Lemaire et al. "Composite particle emission in relativistic heavy ion collisions". In: *Physics Letters B* 85.1 (1979), pp. 38–42. ISSN: 0370-2693. DOI: [https://doi.org/10.1016/0370-2693\(79\)90772-X](https://doi.org/10.1016/0370-2693(79)90772-X). URL: <https://www.sciencedirect.com/science/article/pii/037026937990772X>.
- [83] F. Acero et al. "Fermi Large Area Telescope Third Source Catalog". In: *Astrophys. J. Suppl.* 218.2 (2015), p. 23. DOI: 10.1088/0067-0049/218/2/23. arXiv: 1501.02003 [astro-ph.HE].
- [84] M. Aguilar et al. "The Alpha Magnetic Spectrometer (AMS) on the international space station: Part II — Results from the first seven years". In: *Physics Reports* 894 (2021). The Alpha Magnetic Spectrometer (AMS) on the International Space Station: Part II - Results from the First Seven Years, pp. 1–116. ISSN: 0370-1573. DOI: <https://doi.org/10.1016/j.physrep.2020.09.003>. URL: <https://www.sciencedirect.com/science/article/pii/S0370157320303434>.
- [85] Gary Kitmacher. *Reference guide to the international space station*. 1st. Number NASA/SP-2006-557, 2006.
- [86] P. A. Cherenkov. "Visible radiation produced by electrons moving in a medium with velocities exceeding that of light". In: *Phys. Rev.* 52.4 (1937), pp. 378–379. DOI: 10.1103/PhysRev.52.378.
- [87] K. Lübelmeyer et al. "Upgrade of the Alpha Magnetic Spectrometer (AMS-02) for long term operation on the International Space Station (ISS)". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 654.1 (2011), pp. 639–648. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2011.06.051>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900211011752>.
- [88] M. Aguilar et al. "The Alpha Magnetic Spectrometer (AMS) on the International Space Station: Part I – results from the test flight on the space shuttle". In: *Physics Reports* 366.6 (2002), pp. 331–405. ISSN: 0370-1573. DOI: [https://doi.org/10.1016/S0370-1573\(02\)00013-3](https://doi.org/10.1016/S0370-1573(02)00013-3). URL: <https://www.sciencedirect.com/science/article/pii/S0370157302000133>.
- [89] S. Ahlen et al. "An antimatter spectrometer in space". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 350.1 (1994), pp. 351–367. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/0168-9002\(94\)91184-3](https://doi.org/10.1016/0168-9002(94)91184-3). URL: <https://www.sciencedirect.com/science/article/pii/0168900294911843>.
- [90] G. Ambrosi et al. "The spatial resolution of the silicon tracker of the Alpha Magnetic Spectrometer". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 869 (2017), pp. 29–37. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2017.07.014>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900217307258>.

- [91] Zhaoyi Qu et al. “New track finding based on cellular automaton for AMS-02 detector”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 869 (2017), pp. 135–140. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2017.07.007>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900217307180>.
- [92] Abhinandan Dass. “Be Isotope Measurement With Ams-02 to Reduce Astrophysical Uncertainty in Dark Matter Models”. PhD thesis. Trento U., 2024.
- [93] G. Ambrosi et al. “AMS-02 Track Reconstruction and Rigidity Measurement”. In: *International Cosmic Ray Conference*. Vol. 33. International Cosmic Ray Conference. Jan. 2013, p. 2306.
- [94] Wolfgang Wallraff. “The AMS-02 tracker alignment system for 3D position control with artificial laser generated straight tracks”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 572.1 (2007). Frontier Detectors for Frontier Physics, pp. 325–327. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2006.10.273>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900206021000>.
- [95] Qi Yan and Vitaly Choutko. “Alignment of the Alpha Magnetic Spectrometer (AMS) in space”. In: *Eur. Phys. J. C* 83 (2023), p. 245. DOI: 10.1140/epjc/s10052-023-11395-0. arXiv: 2303.14072 [hep-ex].
- [96] J. Berdugo et al. “Determination of the rigidity scale of the Alpha Magnetic Spectrometer”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 869 (2017), pp. 10–14. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2017.07.012>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900217307234>.
- [97] Y. Jia et al. “Nuclei charge measurement by the Alpha Magnetic Spectrometer silicon tracker”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 972 (2020), p. 164169. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2020.164169>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900220305659>.
- [98] V. Bindi et al. “Calibration and performance of the AMS-02 time of flight detector in space”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 743 (2014), pp. 22–29. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2014.01.002>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900214000102>.
- [99] John Betteley Birks. *The theory and practice of scintillation counting: International series of monographs in electronics and instrumentation*. Vol. 27. Elsevier, 2013.



- [100] Thomas Pöschl et al. "Measurement of ionization quenching in plastic scintillators". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 988 (2021), p. 164865. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2020.164865>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900220312626>.
- [101] Ph. von Doetinchem et al. "The AMS-02 Anticoincidence Counter". In: *Nuclear Physics B - Proceedings Supplements* 197.1 (2009). 11th Topical Seminar on Innovative Particle and Radiation Detectors (IPRD08), pp. 15–18. ISSN: 0920-5632. DOI: <https://doi.org/10.1016/j.nuclphysbps.2009.10.025>. URL: <https://www.sciencedirect.com/science/article/pii/S0920563209007464>.
- [102] I. M. Frank and I. E. Tamm. "Coherent visible radiation of fast electrons passing through matter". In: *Compt. Rend. Acad. Sci. URSS* 14.3 (1937). Ed. by V. L. Ginzburg, B. M. Boltovskoy, and I. M. Dremin, pp. 109–114. DOI: 10.3367/UFNr.0093.196710o.0388.
- [103] M. Aguilar-Benitez et al. "In-beam aerogel light yield characterization for the AMS RICH detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 614.2 (2010), pp. 237–249. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2009.12.027>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900209023444>.
- [104] F. Giovacchini. "Performance in space of the AMS-02 RICH detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 766 (2014). RICH2013 Proceedings of the Eighth International Workshop on Ring Imaging Cherenkov Detectors Shonan, Kanagawa, Japan, December 2-6, 2013, pp. 57–60. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2014.04.036>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900214004410>.
- [105] Jianan Xiao et al. "The AMS-02 RICH detector: Status, latest results, and physics prospects". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1086 (2026), p. 171364. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2026.171364>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900226000902>.
- [106] Thomas Kirn. "The AMS-02 TRD on the international space station". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 706 (2013). TRDs for the Third Millenium, pp. 43–47. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2012.05.010>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900212004974>.
- [107] C. Adloff et al. "The AMS-02 lead-scintillating fibres Electromagnetic Calorimeter". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 714 (2013), pp. 147–154. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2012.05.010>.

- //doi.org/10.1016/j.nima.2013.02.020. URL: <https://www.sciencedirect.com/science/article/pii/S016890021300199X>.
- [108] F. Cadoux et al. "The AMS-02 electromagnetic calorimeter". In: *Nuclear Physics B - Proceedings Supplements* 113.1 (2002), pp. 159–165. ISSN: 0920-5632. DOI: [https://doi.org/10.1016/S0920-5632\(02\)01836-4](https://doi.org/10.1016/S0920-5632(02)01836-4). URL: <https://www.sciencedirect.com/science/article/pii/S0920563202018364>.
  - [109] G. Gallucci and (for the AMS-02 ECAL group). "Performance of the AMS-02 Electromagnetic Calorimeter in Space". In: *Journal of Physics: Conference Series* 587.1 (Feb. 2015), p. 012028. DOI: 10.1088/1742-6596/587/1/012028. URL: <https://doi.org/10.1088/1742-6596/587/1/012028>.
  - [110] A. Basili et al. "The TOF-ACC flight electronics for the fast trigger and time of flight of the AMS-02 cosmic ray spectrometer". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 707 (2013), pp. 99–113. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2012.12.089>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900212016452>.
  - [111] Florian Hauler. "Development of the Space Qualified Data Acquisition for the Transition Radiation Detector of the AMS-02 Experiment on the International Space Station". PhD thesis. Karlsruhe Institute of Technology (KIT), 2005.
  - [112] X. Cai. "On board computing system for AMS-02 mission". In: (2005). URL: <https://cds.cern.ch/record/962442>.
  - [113] Andrei Kounine and Vladimir Koutsenko. "Flight Software for xDR and JINx nodes in AMS-02". In: ().
  - [114] D.F. Smart and M.A. Shea. "A review of geomagnetic cutoff rigidities for earth-orbiting spacecraft". In: *Advances in Space Research* 36.10 (2005). Solar Wind-Magnetosphere-Ionosphere Dynamics and Radiation Models, pp. 2012–2020. ISSN: 0273-1177. DOI: <https://doi.org/10.1016/j.asr.2004.09.015>. URL: <https://www.sciencedirect.com/science/article/pii/S0273117705001997>.
  - [115] P. Alken et al. "International Geomagnetic Reference Field: the thirteenth generation". In: *Earth Planets Space* 73.49 (2021). Solar Wind-Magnetosphere-Ionosphere Dynamics and Radiation Models. DOI: <https://doi.org/10.1186/s40623-020-01288-x>. URL: <https://link.springer.com/article/10.1186/s40623-020-01288-x#citeas>.
  - [116] Alejandro Reina Conde. "Measurement of Boron, Carbon and Oxygen Fluxes as Function of Time with The Alpha Magnetic Spectrometer". PhD thesis. 2022.
  - [117] S. Agostinelli et al. "Geant4—a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/S0168->

- 9002(03)01368-8. URL: <https://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [118] J. Allison et al. "Geant4 developments and applications". In: *IEEE Transactions on Nuclear Science* 53.1 (2006), pp. 270–278. DOI: 10.1109/TNS.2006.869826.
  - [119] J. Allison et al. "Recent developments in Geant4". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 835 (2016), pp. 186–225. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2016.06.125>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900216306957>.
  - [120] M. Aguilar et al. "Properties of Cosmic Helium Isotopes Measured by the Alpha Magnetic Spectrometer". In: *Phys. Rev. Lett.* 123 (18 Nov. 2019), p. 181102. DOI: 10.1103/PhysRevLett.123.181102. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.123.181102>.
  - [121] L. J. Gleeson and W. I. Axford. "Solar Modulation of Galactic Cosmic Rays". In: 154 (Dec. 1968), p. 1011. DOI: 10.1086/149822.
  - [122] R. A. Caballero-Lopez and H. Moraal. "Limitations of the force field equation to describe cosmic ray modulation". In: *Journal of Geophysical Research (Space Physics)* 109.A1, A01101 (Jan. 2004), A01101. DOI: 10.1029/2003JA010098.
  - [123] Jason Ansel et al. "PyTorch 2: Faster Machine Learning Through Dynamic Python Byte-code Transformation and Graph Compilation". In: *29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24)*. ACM, Apr. 2024. DOI: 10.1145/3620665.3640366. URL: <https://docs.pytorch.org/assets/pytorch2-2.pdf>.
  - [124] S. S. Adler et al. "Deuteron and Antideuteron Production in Au + Au Collisions at  $\sqrt{s_{NN}} = 200$  GeV". In: *Phys. Rev. Lett.* 94 (12 Apr. 2005), p. 122302. DOI: 10.1103/PhysRevLett.94.122302. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.94.122302>.
  - [125] Samuel Ting. "The Alpha Magnetic Spectrometer on the International Space Station". In: *Nuclear Physics B - Proceedings Supplements* 243-244 (2013). Proceedings of the IV International Conference on Particle and Fundamental Physics in Space, pp. 12–24. ISSN: 0920-5632. DOI: <https://doi.org/10.1016/j.nuclphysbps.2013.09.028>. URL: <https://www.sciencedirect.com/science/article/pii/S0920563213005525>.
  - [126] Vinod Nair and Geoffrey Hinton. "Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair". In: vol. 27. June 2010, pp. 807–814.
  - [127] G. Cybenko. "Approximation by superpositions of a sigmoidal function". In: *Mathematics of Control, Signals, and Systems* 2.4 (Dec. 1989), pp. 303–314. DOI: 10.1007/BF02551274. URL: <https://hal.science/hal-03753170>.

- [128] Olivier Delalleau and Yoshua Bengio. “Shallow vs. Deep Sum-Product Networks”. In: *Advances in Neural Information Processing Systems*. Ed. by J. Shawe-Taylor et al. Vol. 24. Curran Associates, Inc., 2011. URL: [https://proceedings.neurips.cc/paper\\_files/paper/2011/file/8e6b42f1644ecb1327dc03ab345e618b-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2011/file/8e6b42f1644ecb1327dc03ab345e618b-Paper.pdf).
- [129] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG]. URL: <https://arxiv.org/abs/1412.6980>.
- [130] Liudmila Prokhorenkova et al. “CatBoost: unbiased boosting with categorical features”. In: *Advances in Neural Information Processing Systems*. Vol. 31. 2018.
- [131] Houda Bichri, Adil Chergui, and Hain Mustapha. “Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets”. In: *International Journal of Advanced Computer Science and Applications* 15 (Jan. 2024). DOI: 10.14569/IJACSA.2024.0150235.
- [132] Thomas Nagler et al. *Reshuffling Resampling Splits Can Improve Generalization of Hyperparameter Optimization*. 2024. arXiv: 2405.15393 [stat.ML]. URL: <https://arxiv.org/abs/2405.15393>.
- [133] Zahra Bami et al. *A New Flexible Train-Test Split Algorithm, an approach for choosing among the Hold-out, K-fold cross-validation, and Hold-out iteration*. 2026. arXiv: 2501.06492 [cs.LG]. URL: <https://arxiv.org/abs/2501.06492>.
- [134] Youran Zhou, Sunil Aryal, and Mohamed Reda Bouadjenek. *Review for Handling Missing Data with special missing mechanism*. 2024. arXiv: 2404.04905 [stat.ME]. URL: <https://arxiv.org/abs/2404.04905>.
- [135] David K. Lim et al. *Unsupervised Imputation of Non-ignorably Missing Data Using Importance-Weighted Autoencoders*. 2022. arXiv: 2101.07357 [cs.LG]. URL: <https://arxiv.org/abs/2101.07357>.
- [136] Simon Tihon et al. *DAEMA: Denoising Autoencoder with Mask Attention*. 2021. arXiv: 2106.16057 [cs.LG]. URL: <https://arxiv.org/abs/2106.16057>.
- [137] Charles R. Harris et al. “Array programming with NumPy”. In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2. URL: <https://doi.org/10.1038/s41586-020-2649-2>.
- [138] Thomas Junk. “Confidence level computation for combining searches with small statistics”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 434.2 (1999), pp. 435–443. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(99\)00498-2](https://doi.org/10.1016/S0168-9002(99)00498-2). URL: <https://www.sciencedirect.com/science/article/pii/S0168900299004982>.
- [139] J. Neyman. “Frequentist Probability and Frequentist Statistics”. In: *Synthese* 36.1 (1977), pp. 97–131. ISSN: 00397857, 15730964. URL: <http://www.jstor.org/stable/20115217> (visited on 12/21/2025).

- [140] Gary J. Feldman and Robert D. Cousins. “Unified approach to the classical statistical analysis of small signals”. In: *Phys. Rev. D* 57 (7 Apr. 1998), pp. 3873–3889. DOI: 10.1103/PhysRevD.57.3873. URL: <https://link.aps.org/doi/10.1103/PhysRevD.57.3873>.
- [141] Robert D. Cousins, James T. Linnemann, and Jordan Tucker. “Evaluation of three methods for calculating statistical significance when incorporating a systematic uncertainty into a test of the background-only hypothesis for a Poisson process”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 595.2 (2008), pp. 480–501. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2008.07.086>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900208010255>.
- [142] Jerzy Neyman and Egon Sharpe Pearson. “IX. On the problem of the most efficient tests of statistical hypotheses”. In: *Philosophical Transactions of the Royal Society of London, Series A: Containing Papers of a Mathematical or Physical Character* 231.694-706 (Feb. 1933), pp. 289–337. ISSN: 0264-3952. DOI: 10.1098/rsta.1933.0009. eprint: <https://royalsocietypublishing.org/rsta/article-pdf/231/694-706/289/240090/rsta.1933.0009.pdf>. URL: <https://doi.org/10.1098/rsta.1933.0009>.
- [143] Ti-Pei Li and YuQian Ma. “Analysis methods for results in gamma-ray astronomy”. In: *The Astrophysical Journal* 272 (Aug. 1983), pp. 317–324. DOI: 10.1086/161295.
- [144] J. Alcaraz et al. “Search for antihelium in cosmic rays”. In: *Physics Letters B* 461.4 (1999), pp. 387–396. ISSN: 0370-2693. DOI: [https://doi.org/10.1016/S0370-2693\(99\)00874-6](https://doi.org/10.1016/S0370-2693(99)00874-6). URL: <https://www.sciencedirect.com/science/article/pii/S0370269399008746>.