Università degli Studi di Ferrara

DOCTORAL COURSE IN

PHYSICS

CYCLE XXXV

COORDINATOR

Prof. Luppi Eleonora

Light Hadron Production measurement with LHCb fixed-target data and LHCb RICH system commissioning for Run3

Scientific-Disciplinary Sector (SDS): FIS/01

Ph.D. Candidate:
Edoardo Franzoso

Supervisor:
Prof. Massimiliano Fiorini

Co-Supervisors:
Prof. Luciano Libero Pappalardo
Prof. Marcin Chrząszcz

Years 2019/2023

Contents

Abstract

The LHCb experiment located at CERN is one of the major experiments at the LHC accelerator. The LHCb detector is unique among the other detectors due to its single-arm spectrometer design. An excellent Particle Identification (PID) is among the necessary requirements to achieve the precision needed for CP violation studies and b-quarks decays, which constitute the main original physics cases of the experiment. Provided with a Ring Imaging Cherenkov (RICH) system, LHCb is optimized to identify and discriminate light hadrons (π, K and p). Throughout its lifetime, the LHCb physics reach has extended substantially beyond what was originally planned and the detector is currently operating also in a fixed-target mode exploiting collisions of the LHC beams with a gas targets. The gas is injected into the beam pipe by means of the SMOG system, providing a unique fixed-target facility at the highest available beam energy. The combination of an excellent PID with the collection of proton-nucleus collision data in an unexplored energy range opens the possibility to perform light-hadron production measurements with distinctive target systems. Charged particle production in hadronic collisions is a fundamental observable for studying the properties of the strong interaction, described by quantum chromodynamics (QCD), giving the possibility to probe and study the so-called Cold Nuclear Matter (CNM) effects. By studying the medium-induced modifications of light-hadron production in proton-nucleus collisions (p-A) relative to the naive proton-proton collisions (p-p), information on the nontrivial QCD dynamics can be retrieved. These CNM effects can cause the suppression of the production cross sections and the modification of the particle spectra in a regime where the quark-gluon-plasma (QGP) formation is not expected.
In recent years, the LHCb detector has undergone a major upgrade to better exploit the luminosity delivered by the LHC with the goal to operate safely and efficiently at the luminosity of $2\times10^{33}\text{cm}^{-2}\text{s}^{-1}$, five times higher than the past run conditions. This required an upgrade of most of the LHCb subdetectors and of the read-out system to cope with the 40 MHz rate and the higher level of radiation foreseen for the incoming Run3, started in 2022. In particular, the RICH detectors have been completely renewed after long and successful upgrade and commissioning activities. For the RICH upgrade, the former Hybrid Photon Detectors (HPDs) have been replaced by Multi Anode Photon Multipliers (MaPMTs), the optics of the RICH nearest to the beam collision have been modified, and the electronics replaced. The new front-end electronics are based on the CLARO chip, FPGAs digital board, and Giga Bit Transceiver (GBT) chip for data transmission and front-end configuration. The modifications to the RICH system aim to maintain and improve excellent PID performance. The RICH system has been almost fully commissioned and has been operated during stable beam collisions delivered by LHC throughout 2022. In this thesis, I present my contribution to the Run2 LHCb fixed-target

program, which involves the development of a novel PID tool designed as a Neural Network-based Gaussian Mixture Model making use of the RICH system variables, and a light hadron production analysis using the data collected with He and Ar targets in Run2. Additionally, my activities related to the RICH system Upgrade for the Run3 are also described. They comprehends the validation of the opto-electronic chain and the quality assurance of the photomultipliers and Front-End electronics. Also, the work of my last year of PhD in the LHCb control room for the RICH system commissioning is covered in which I've contributed to achieve a stable and reliable RICH system able to operate and collect data in Run3.

The thesis content is structured as follows: an introduction chapter gives the general theoretical context and motivations. In Chap.1 an overview of the LHCb experiment is given and the design and subdetectors in the Run2 configuration are described in detail. The SMOG system is introduced in Chap.2, as well as the luminosity determination in LHCb. The entire Chap.3 is dedicated to the Upgrade of the RICH detector System for Run3, covering the full opto-electronic chain. The following chapter, Chap.4, is entirely devoted to my last year's activities, the RICH detectors system commissioning in the LHCb control room and Point 8. The last chapters cover my analysis work dedicated to fixed-target data collected with the SMOG system in Run2, which relies extensively on the PID capabilities of the LHCb detector and of the RICH system, specifically. A unique approach to exploit the PID information given by the RICH is discussed in Chap.5 and its application to the SMOG fixed-target data is discussed. This new technique consists of completely data-driven modeling of the PID response through a Neural Network-based Gaussian Mixture Model, and has been validated and published in a paper. The last chapter (Chap.6) describes the light hadron production analysis to search for CNM effect in fixed-target LHCb data. The analysis studies the production π, K and p in pAr and pHe collisions at a center of mass energy $\sqrt{s}$ = 110 GeV, making use of the PID tool developed before. The final results are shown in Chap. 7.

Introduction and Physics Motivations

0.1 Theory of Strong Interactions

Gell-Mann and Zweig introduced the quark model in 1963 to explain the increasing complexity of stable hadrons and hadronic resonances discovered in particle accelerators during the 1950s and 1960s. At that time, it was known that the strong interactions conserved a quantum number called strangeness and also that isospin was a very good symmetry. The quark model was first introduced to give a simple realization of this symmetry. Mesons are described as bound states of quarks and antiquarks. In contrast, baryons are composed of three quarks, while antibaryons consist of three antiquarks. The spectrum of hadrons known at that time could be explained by just three types of quarks (up, down, and strange, also referred to as u, d, and s). In addition, three other quark flavours are known today: charm, bottom, and top (c, b, and t). Protons and neutrons are made up of uud and udd quarks, respectively. It is necessary to assign an electric charge to the u (2/3) and d (-1/3) quarks in order to explain the charge of these baryons. The light pseudo-scalar mesons π^+ and π^- have quark content $u\bar{d}$ and $d\bar{u}$. The π^0, on the other hand, is a linear combination of $u\bar{u}$ and $d\bar{d}$. The three π mesons form an isospin triplet. In pseudo-scalar mesons with non-zero strangeness, the strange quark or its antiquark must be included. For instance, K^+ and K^0 are composed by $u\bar{s}$ and $d\bar{s}$, so the strange quark must have charge -1/3. Quantum Chromodynamics (QCD) is the established theory of strong interactions and is part of the Standard Model (SM). Specifically, the QCD is a non-abelian gauge field theory described by a local symmetry under the SU(3) symmetry group, and the associated quantum number, called color, can assume one of three values defined as red (r), green (g) and blue (b). Each quark carries a color charge, whereas antiquarks carry anti-colors. In contrast to Quantum Electrodynamics (QED), being a non-abelian theory, QCD implies that the bosons mediating the interaction, called gluons, carry a color-anticolor charge pair. The peculiarities arising from this property of QCD are the self-coupling of gluons and the asymptotic freedom, which causes interactions between quarks and gluons to become asymptotically weaker as the energy scale increases and the corresponding length scale decreases. On the other hand, at low energies, the interaction becomes strong, leading to the confinement of quarks and gluons within composite hadrons, so that only color-neutral combinations (color singlets) can be observed. The dimensionless coupling constant, useful to characterize the strength of the strong interactions, is given by

$$\alpha_s(Q^2) = \frac{4\pi}{(11 - \frac{2}{3}n_f)\ln\left(Q^2/\Lambda_{\mathrm{QCD}}\right)} \tag{1}$$

where Q^2 is the momentum transferred in the interaction, n_f is the number of quark flavors and the $\Lambda_{\mathrm{QCD}}(\sim 200)$ GeV set the threshold for a perturbative treatment of interactions [1], implying that strong interactions become weak if probed over short times (or at high energies). Not every pp collision can be studied in a perturbative regime. In fact, most pp collisions are soft: they produce a multitude of hadrons largely in the same direction as the original protons. Experimental findings indicate a smooth transition between "small" collision systems (proton-proton, pp, and proton-nucleus, pA) and "large" collision systems (nucleus-nucleus, AA). Large Hadron Collider (LHC) and Relativistic Heavy Ion Collider (RHIC) findings are particularly significant because some of the involved observables were traditionally regarded as signs that ultra-relativistic AA collisions produce a deconfined partonic matter close to thermodynamic equilibrium - Quark Gluon Plasma (QGP). The QGP is a state of matter that exists at extremely high temperatures and/or densities. This state is considered to consist of asymptotically free quarks and gluons, which are confined inside hadrons in normal conditions.

Historically, the role of p+A collisions has been essential for the interpretation of the results of heavy-ion collisions, since to fully comprehend the latter, it is necessary first to understand the effects of normal nuclear matter. The production of hard probes and charged particles is modified from p+p collisions when the collision involves nuclei. These effects are collectively referred to as Cold Nuclear Matter (CNM) effects.

0.1.1 nPDFs modification

Perturbative calculations of cross-sections in QCD are generally factorized into parton density probability functions and a hard scattering cross-section. Parton density probability functions are described in terms of Parton Distribution Functions (PDFs), which typically depend on the longitudinal momentum fraction (x) [2] and energy scale (Q^2) of the quarks and gluons inside a free proton. Phenomenological parametrizations are used to obtain PDFs based on data from Deep Inelastic Scattering (DIS) experiments and Drell-Yan (DY) production studies. Quark and antiquark distributions are directly probed by DIS and DY, whereas gluon distributions are indirectly probed. Gluon distributions are therefore less constrained than quark distributions. Nuclear DIS experiments have confirmed that the PDFs of nucleons bound in nuclei are modified, leading to nuclear parton distribution functions (nPDFs) [3]. A ratio is typically used to quantify these effects:

$$R_i^A\left(x,Q^2\right) = \frac{f_i^A\left(x,Q^2\right)}{f_i\left(x,Q^2\right)}\,, \tag{2}$$

where $f_i^A\left(x,Q^2\right)$ denotes the PDF for a parton with flavour i of a nucleon bound in a nucleus A. The structure of the ratio as a function of the x variable is shown in Fig.3.

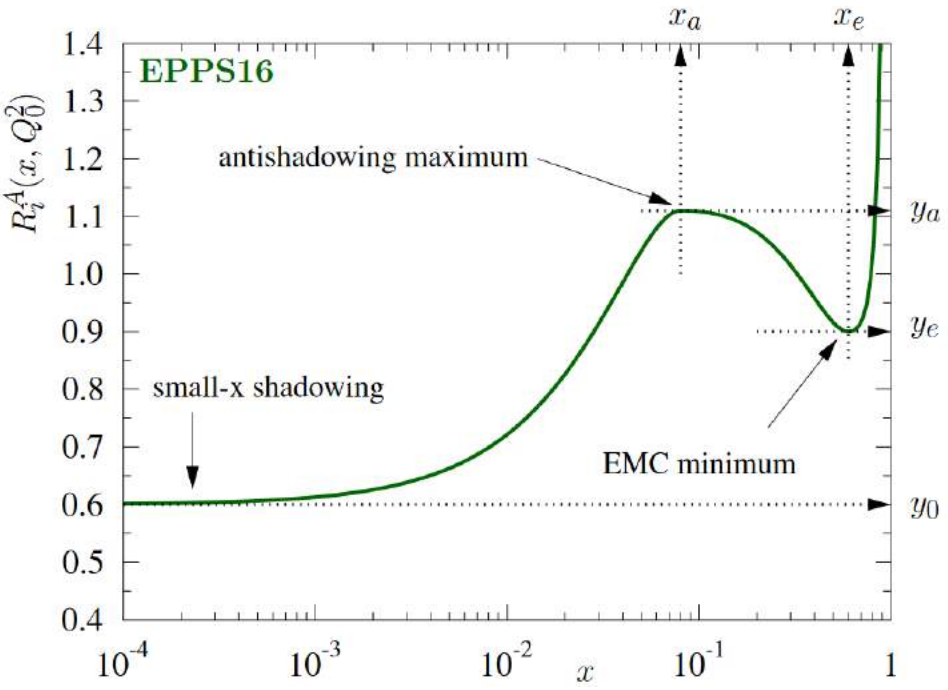


Figure 1: A sketch of the modification of the parton distribution function in nuclei.

The different regions are:

- shadowing: the low-x (x<0.03) region where the ratio is less than $R_i^A\left(x,Q^2\right)<1$;
- anti-shadowing: the intermediate-x region (0.03<x<0.3) having an opposite behaviour of the ratio $R_i^A\left(x,Q^2\right)>1$;
- EMC region: the higher x region (0.3<x<0.7). Here $R_i^A\left(x,Q^2\right)<1$;
- Fermi-motion: the highest x region where $R_i^A\left(x,Q^2\right)>1$ due to Fermi motion of the nucleons in the nucleus.

Based on the perturbative calculation of a hard scattering process, the nPDF's are assumed to be factorizable. The factorization for the free-proton case has proven to be very successful, but some concern has been raised about the factorization of the nPDF, since multiple scatterings may occur with different nucleons. A parameterization of the available data is necessary to include this effect in the nPDF's. Parameterizations are available in the form of nPDF sets. A few examples are nDSg, EKS98 [4], and EPS09 [5]. The most commonly used nPDF set is EPPS16 [6], as it provides not only the fit to the most recent LHC data but also an estimate of the distribution uncertainty. An example of the nPDF distributions is shown in Fig. 2.

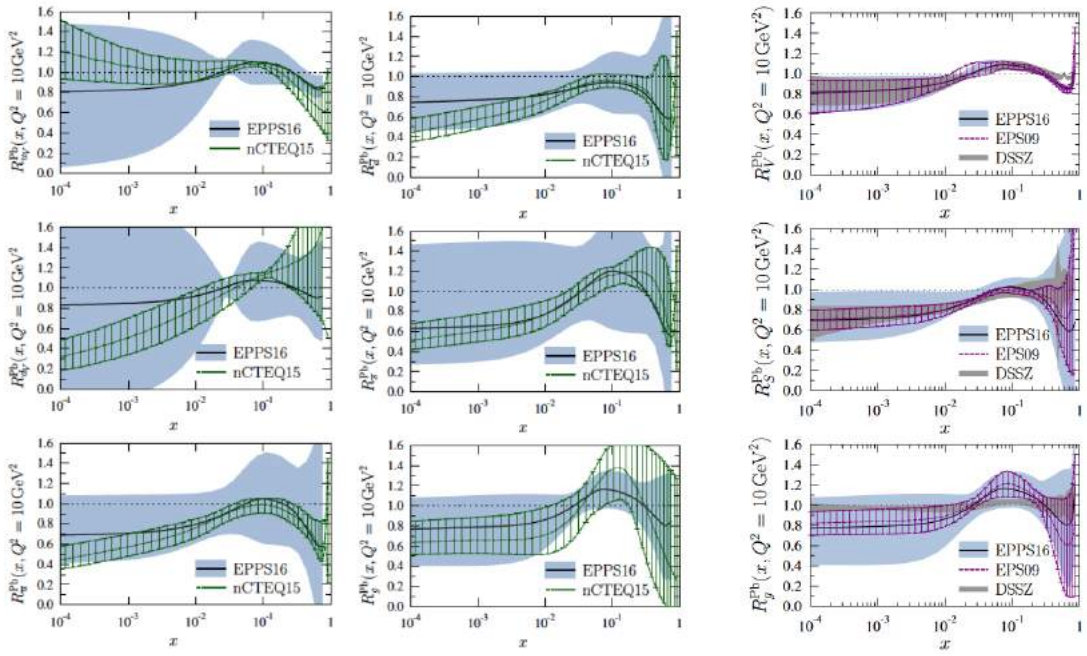


Figure 2: Comparison between the nuclear modification factors from the nCTEQ15 [7] and EPPS16 sets (left and middle plots) and from the EPPS16, EPS09 and DSSZ [8] sets (right).

0.1.2 Cronin Effect

The modification of the transverse momentum distributions of particles produced in p+A collisions compared to scaled p+p collisions was observed for the first time in 1974 [9]-[10]. A suppression of hadron production has been observed at low p_T (below a few GeV/c), while an enhancement is observed at intermediate p_T (around 2-4 GeV/c). The Cronin effect is commonly attributed to multiple elastic scattering of the incoming parton before a hard collision occurs [11]. The Cronin effect has been observed and studied recently using RHIC data [12]-[13]-[14]-[15] and various theoretical efforts make predictions that have to be validated by data [11]-[16]-[17]-[18]. According to measurements conducted at RHIC, the Cronin peak's evolution is highly dependent on the collision's centrality and weakly dependent on the collision energy. Studies of low-x regions at LHC energies [19] in collider mode show that shadowing effects become more significant compared to multiple scattering and suppress the enhancement. Additionally, the hadron mass determines the magnitude of the effect, with increasing trends with larger masses.

0.1.3 Radiative Energy Loss

Radiative energy loss is caused by the exchange of gluons with the medium [21]. Losses of energy can occur at two different levels, the initial state and the final state. In initial-state energy loss, an incoming parton emits gluons during its traversal of the medium before hard scattering occurs. Consequently, the parton distribution changes relative to p+p collisions as the incoming partons x decreases. Final-state energy loss occurs when an outgoing particle, such as a quark, gluon, or preformed meson, radiates energy. As a result, the particle's x is also decreased.

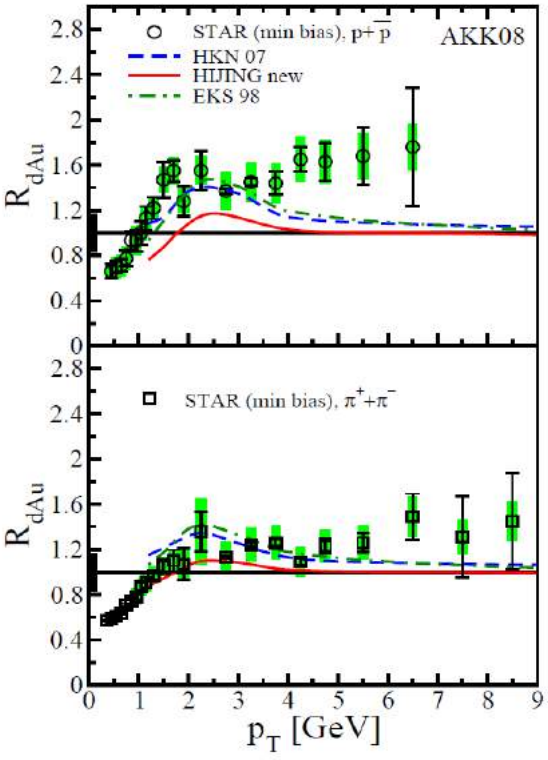


Figure 3: The mini-biased Cronin ratio R for charged π and p at RHIC energy $\sqrt{s}$ = 200 GeV at midrapidity (|y|< 0.5). Data from [20].

0.1.4 Gluon Saturation

The density of gluons starts to dominate the density of quarks at smaller values of the longitudinal momentum fraction, x (or at higher momentum transfer Q^2) [22]-[23]. Saturation of the gluon distribution occurs at some scale, and nonlinear corrections to the normal evolution equations dominate in this regime. It is possible to employ weak coupling techniques if the saturation scale is larger than the perturbative scale (Qs >> QCD). In order to explain this small-x behavior of QCD, the Color Glass Condensate (CGC) field theory was developed. As opposed to initial state energy loss and the Cronin effect, nuclear shadowing and gluon saturation are not completely dissimilar and clearly separated since they both explain the modifications of parton distributions based on Q^2 and x.
Noteworthy, all the effects described above are predicted and observed to be more pronounced for heavier nuclei, as a consequence of the larger volume of cold nuclear matter experienced by the incoming parton.

0.2 Hadron production in hadronic collisions

A key observable for examining the characteristics of the strong interaction under the control of quantum chromodynamics (QCD) is the production of charged particles in hadronic collisions. Charged hadrons can be produced in soft and hard interactions, which correspond to modest and significant momentum transfers between the interacting partons of the hadrons in high-energy col-

lisions. While perturbative QCD (pQCD) can be used to characterize hard interactions, predictions for the soft regime rely primarily on phenomenological considerations. The study of charged particle production as a function of the 'hardness' of the interaction, expressed using the transverse momentum of the particles with respect to the beam axis, is of particular interest when different colliding nuclei are considered. The charged particle production modification between light and heavy systems can be modeled by the various CNM effects described above. CNM effects in charged particle production have been observed by ALICE [19, 24, 25], PHENIX at RHIC [26, 20, 12] and HERMES at DESY [27, 28]. These effects can be probed using the datasets collected by the LHCb experiment in the fixed-target configuration. More targets, different from the ones already studied by other experiments, are available to directly compare 'lighter' and 'heavier' nuclear systems in a fixed target configuration, covering the central-backward rapidity region in the cm frame and high Bjorken-x values. The goal of the present analysis is the measurement of inclusive prompt-charged particles in fixed-target proton-nucleus collisions. Specifically, the search for CNM effects reported in this work uses the pHe and pAr datasets collected at a center-of-mass energy of 110 GeV by the LHCb experiment with the SMOG device. The analysis uses events triggered with minimum bias conditions to study modifications of the charged particle production rate using different target nuclei which can be modeled assuming various underlying CNM effects. The light hadron production rate is studied mainly as a function of the hardness of the interaction which corresponds to charged particles transverse momentum (p_T) with respect to the axis of the colliding hadrons, but also the dependence on the pseudorapidity η of the charged particle and on the multiplicity of the event is investigated. A precise luminosity determination is not provided for these datasets, so, of main interest for this study, are the ratios between different hadron types, since no luminosity measurement is needed. The aim of the analysis is to exploit the unique kinematic range of the fixed-target data collected with SMOG to investigated the light hadron production yields ratios in pHe and pAr for high x-Bjorken and backward-central pseudorapidity region in the centre of mass frame ($-2.8 < \eta^* < 0.2$) where nuclear PDFs are modified by the antishadowing and EMC effect. The PID capabilities of the RICH detector system in LHCb are used to discriminate the hadron components in the mixture of charged tracks produced in the minimum bias events, also studying the production yields as a function of the electric charge. These data will fill a gap in the dataset available from other experiments, as shown in Fig. 4.

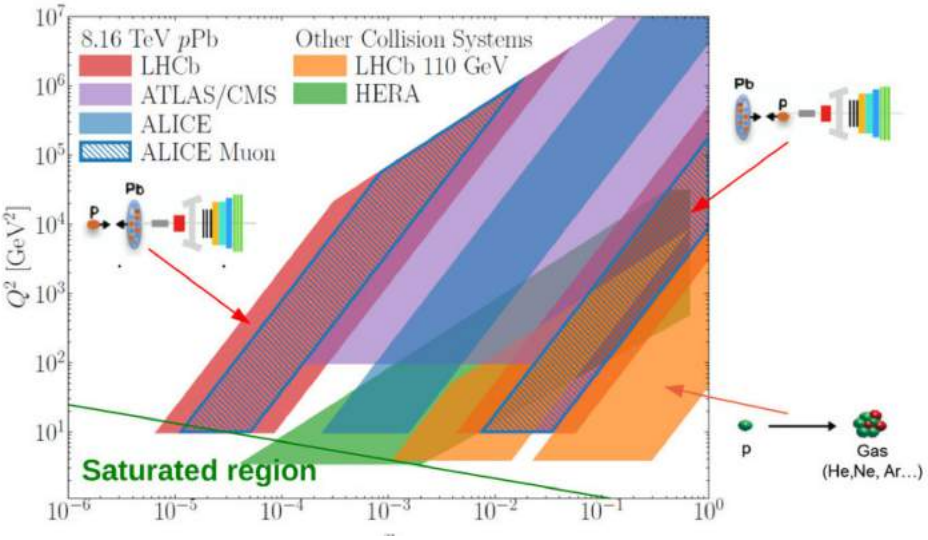


Figure 4: Comparison between the accessible kinematic regions to LHCb in its collider and fixed-target configurations, other experiments are also shown. The coverage is expressed as a function of the Q^2 parton-parton invariant mass and the Bjorken-x value of the nucleon in the nuclear target.

Chapter 1

The LHCb experiment at LHC

In this chapter, an outline of the experimental context is given. The Large Hadron Collider is described in Sec.1.1, with a focus on the LHCb experiment (Sec.1.2). The LHCb detector is presented in detail based on the Run2 configuration with all the subdetectors.

1.1 LHC

The Large Hadron Collider Beauty (LHCb) experiment at the European Center for Nuclear Research (CERN) is one of the main experiments primarily dedicated to the study of flavor physics. Specifically, LHCb is optimized for studies of CP violation in the B-meson sector, as well as rare b-quark hadrons decay by exploiting the large cross-section of b quark production at the Large Hadron Collider (LHC). As a result of these studies, not only the standard model (SM) of particle physics is exhaustively tested and its parameters precisely measured, but also new physics effects (NP) are indirectly explored. The LHC [29] at CERN is a superconducting circular particle accelerator used for accelerating protons and heavy ions. With its two 27 km long rings, which span the Swiss-French border between Geneva Airport and the Jura Mountains, LHC is the most powerful collider ever created, with a nominal center-of-mass energy of 14 TeV and an instantaneous luminosity of $10^{34}\ cm^{-2}s^{-1}$. The LHC is designed not only to provide proton-proton collisions but also lead-ion collisions with a center-of-mass energy of 2.76 TeV per nucleon and a peak of $10^{27}\ cm^{-2}s^{-1}$.

Protons are produced by removing electrons from hydrogen atoms with an electric field. As shown in Fig. 1.1, the particles are accelerated in various steps. In the first step, protons are accelerated to 50 MeV using a linear particle accelerator called LINAC 2. After being accelerated to 1.4 GeV in the Proton Synchrotron Booster (PSB), the protons are injected into the Proton Synchrotron (PS), when they reach an energy of 26 GeV. Before being injected into the LHC ring, the Super Proton Synchrotron (SPS) increases their energy to 450 GeV. After this stage, the bunches of up to 1.15×10^{11} protons are injected into the LHC and further accelerated up to 7 TeV. Proton-proton collisions are the main focus of the LHC's physics program. Heavy-ion collisions are, however, conducted during dedicated periods.

LHC Timing

A radio frequency cavity at the LHC accelerates the beam by 485 keV per revolution. The frequency is related to the energy and it increases by ~ 85 Hz between 450 GeV and 6.5 TeV. Beam and cavities cannot be out of sync, otherwise, the beam will be spread out and lost. The same synchronous clocks (for bunches and orbits) are transmitted to the experiments to perform time alignment. Two consecutive bursts due to bunch crossing occur every 25 ns, which translates into a collision rate of 40 MHz. The effective collision frequency is about 30 MHz due to dead time and delays associated with beam dump and filling schemes.

In the LHC, beams are focused and held in a circular orbit by superconducting dipoles and quadrupoles magnets. Additional trajectory adjustments can be made using sextupoles, octupoles, and decapoles. For each dipole, an electric current of 11.700 A generates a magnetic field with a nominal strength of 8.33 T. Based on niobium-titanium (NbTi) Rutherford cables, the magnetic system operates at 2 K using superfluid helium. Because of the limited space in the tunnel, one cryogenic structure is used in a single cryostat for both proton rings. Nevertheless, this twin-bore design requires opposing magnetic fields to maintain two proton beams simultaneously. Parallel orbits are maintained between the two beams, bringing them together only at the interaction points (IP) in correspondence with the four experiments. The first collisions have been performed in 2010 at an energy of 3.5 TeV per beam. The data collection period from 2010 to 2013 is referred to as Run 1. LHC operations were suspended after Run 1 for detector maintenance and upgrades. Collisions with a center-of-mass energy of 13 TeV resumed in 2015. In 2015, Run 2 began, and it concluded in 2018. During the period named Run 2 (2015-2018) an energy of 6.5 TeV per beam (13 TeV total center of mass energy), was reached. At the end of 2018, LHC entered a three-year shutdown period during which many experiments performed upgrades. Figure 1.3 shows the integrated luminosity accumulated by the LHCb experiment during each data-taking year. The integrated luminosity collected by LHCb is between 87 and 94 % of the total luminosity delivered each year. In 2022, LHC operations restarted for the Run 3.

1.1.1 Luminosity

A variable that indicates the accelerator's ability to deliver collisions to experiments is the instantaneous luminosity $\mathscr{L}$.

$$\frac{dN_{event}}{dt} = \mathscr{L} \cdot \sigma_{event}. \tag{1.1}$$

The general expression for a collider with two circulating beams is given by

$$\mathscr{L} \sim 2\, c\, cos^2\alpha\, n_b\, N_1\, N_2\, f_{rev} \int \rho_1(x,y,z,t)\, \rho_2(x,y,z,t)\, dx\, dy\, dz\, dt \tag{1.2}$$

where α is the beams crossing angle, f_{rev} the revolution frequency, n_b the number of circulating bunches, N the number of particles per bunch and ρ the normalized particle density distributions per bunch. Assuming a Gaussian shape of the bunches in the x and y directions and a negligible

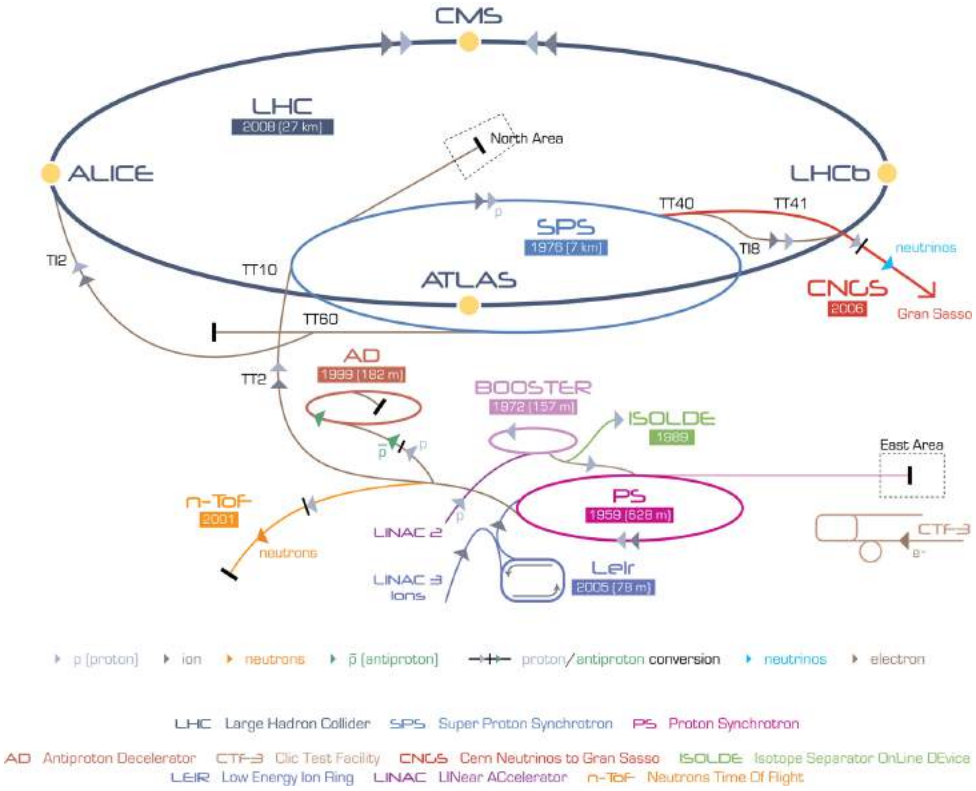


Figure 1.1: The CERN accelerator complex.

crossing angle the luminosity can be written as

$$\mathscr{L} = \frac{N_1 N_2 n_b f_{rev}}{4\pi\sigma_x\sigma_y} \tag{1.3}$$

with $\sigma_{x,y}$ the widths of the bunch Gaussian shape in the transverse plane. Due to beam collisions and scattering within the beam pipe material and residual gas atoms, the luminosity decreases with a lifetime of $\mathcal{O}$(10h).

1.1.2 LHC experiments

To study the products of high-energy collisions, the beams interact at four different points along the ring, where the largest experiments are located. The Fig. 1.1 illustrates in clockwise order:

- ATLAS: *A Toroidal LHC ApparatuS* [30] a general-purpose detector that aims to measure precision SM, search for and study the Higgs boson, and gain insight into new physics mechanisms. With a length of 44 meters, a diameter of 25 meters, and a weight of 7000 tons, it is the biggest experiment at the LHC.

- ALICE: a large ion collider experiment [31] that explores the initial state of matter through ion-ion collisions. Quark-gluon plasmas require high energy densities, which are achieved by colliding lead ions with $\sqrt{s} = 2.67$ TeV at a peak luminosity of $\mathscr{L} = 10^{27}$ (lumi).

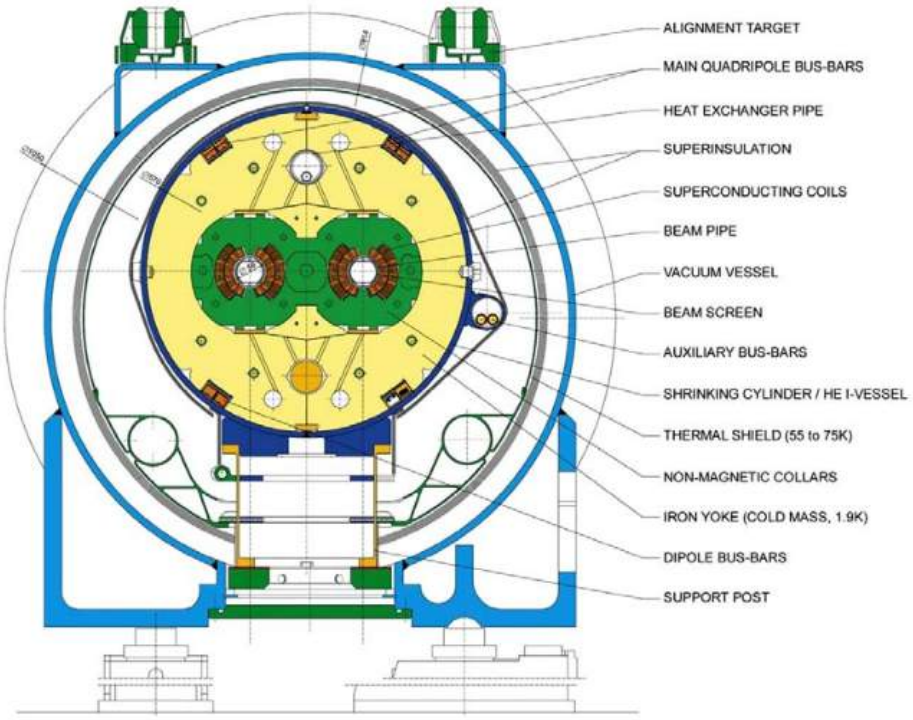


Figure 1.2: Section of the LHC superconducting dipole magnet.

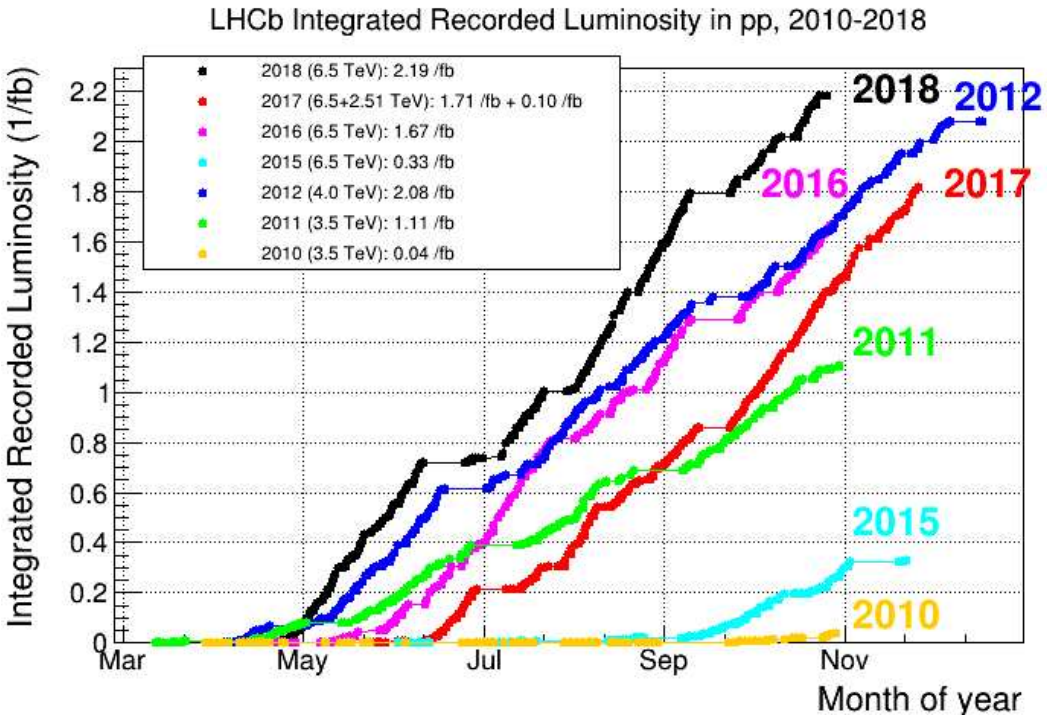


Figure 1.3: Integrated luminosity of pp collisions recorded by the LHCb experiment in each data-taking year.

- CMS: *Compact Muon Solenoid* [32] is a general-purpose detector like ATLAS. An internal magnetic field of 3.8 Tesla is generated by the superconducting solenoid, which is generally equal to 10^5 times the magnetic field of the Earth. In terms of its size, CMS is compact due to its relatively small dimensions compared to its weight: about 14.000 tonnes for a diameter of 15 meters and a length of 28.7 meters.

- LHCb: a description of *LHC-beauty* can be found in the following section: 1.2.

1.2 LHCb experiment

LHCb is a single-arm spectrometer [33] with a forward angular acceptance that covers the range 10-300 (250) mrad in the bending (non-bending) plane. The geometry of the detector is designed to study high-energy processes $b\bar{b}$ hadrons production [34]. At high energies, the cross-section of $b\bar{b}$ pairs is large for small polar angles with respect to the direction of the beam axis, as shown in Fig. 1.4. LHCb, therefore, contains only one arm, unlike the ATLAS and CMS detectors that cover the central region, also referred to as General Detector Purpose (GDP). The LHCb detector covers a pseudorapidity (η) range of $2 < \eta < 5$. A high-precision tracking system measures the momentum of charged particles, with relative uncertainties ranging from 0.5% at the lowest measurable momentum to 1.0% at 200 GeV. The impact parameter, that is the minimum distance of a track to a primary vertex, is measured with a resolution of $(15 + 29/p_T)\ \mu m$, where p_T is the transverse (with respect to the beam) momentum component (in GeV).

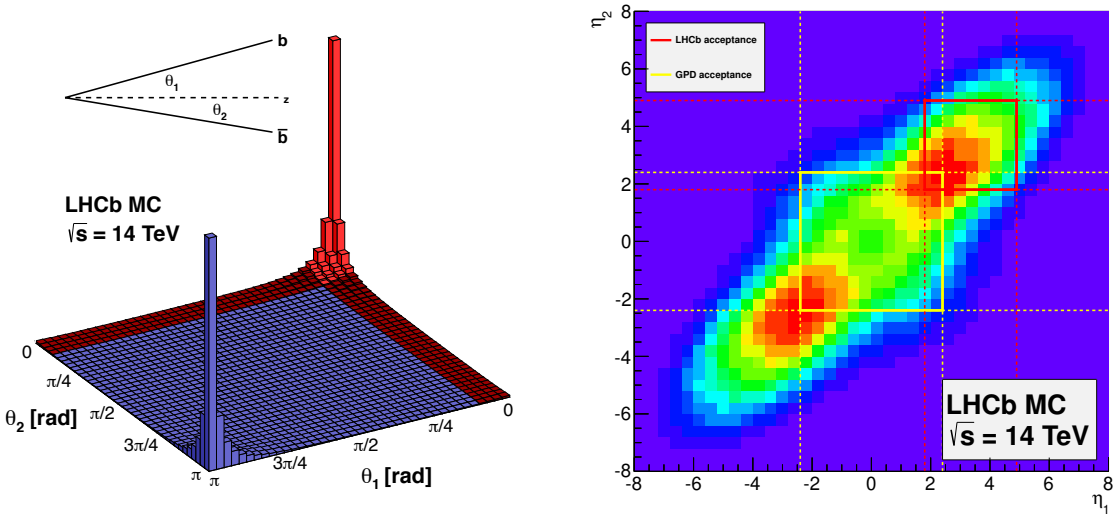


Figure 1.4: The production cross-section of $b\bar{b}$ pairs is shown on the left as a function of their polar angle with respect to the beam axis, assuming a center-of-mass energy of 14 TeV for pp collisions. Acceptance of LHCb is highlighted in red. On the right, a two-dimensional pseudorapidity plot of $b\bar{b}$ production phase space in simulated pp collisions at $\sqrt{s} = 14$ TeV, the LHCb acceptance is highlighted in the red square while the General Purpose Detector (GDP) acceptance is shown in yellow.

Fig 1.5 shows a schematic representation of the experimental apparatus from the side. There are

two halves, referred to as the detector A- and C-sides, of most detector subsystems, which can be moved horizontally for assembly and maintenance purposes as well as to access the beam-pipe.

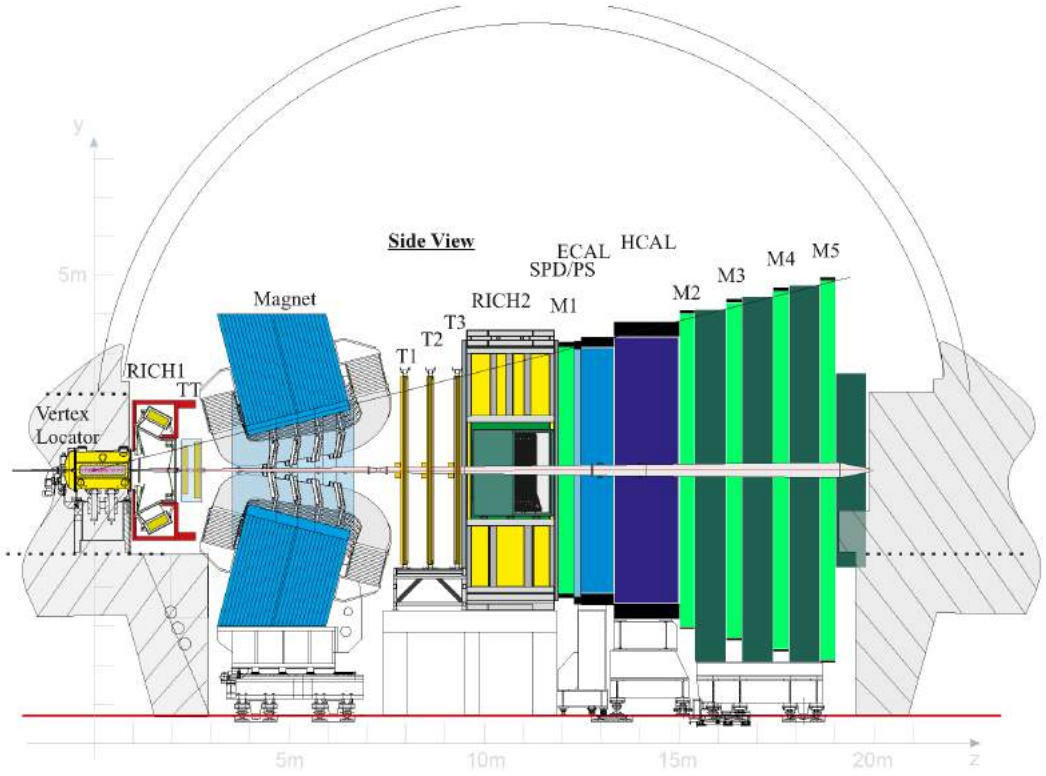


Figure 1.5: Side view of the LHCb detector in the non-bending $y-z$ plane with its subdetectors along the beam $z-$axis.

1.2.1 Tracking system

The LHCb tracking system consists of VELO, TT, and three tracking stations (T1-T3) [35]. As charged particles traverse the detector material, the system provides their trajectory. A warm dipole magnet bends the trajectory of charged particles, allowing their momentum to be measured.

VELO

As part of the LHC vacuum vessel, the VELO (VErtex LOcator) benefits form the absence of material budget in front of it, achieving the best vertex resolution possible [36]. In addition to measuring the trajectories of the charged particles, it measures the positions of the primary vertices of the pp collisions and the displaced secondary vertices. The VELO consists of 46 semicircular silicon modules arranged perpendicularly to the beam direction over a length of 1 m, as shown in Fig. 1.6. Except for the first four modules, which consist of one type of radiation-resistant sensor, the remaining modules all consist of two radiation-resistant sensors with a thickness of 300 microns, each with a sensitive area between 8.2 mm and 41.9 mm from the beam. Sensor pairs are specialized for measuring radial

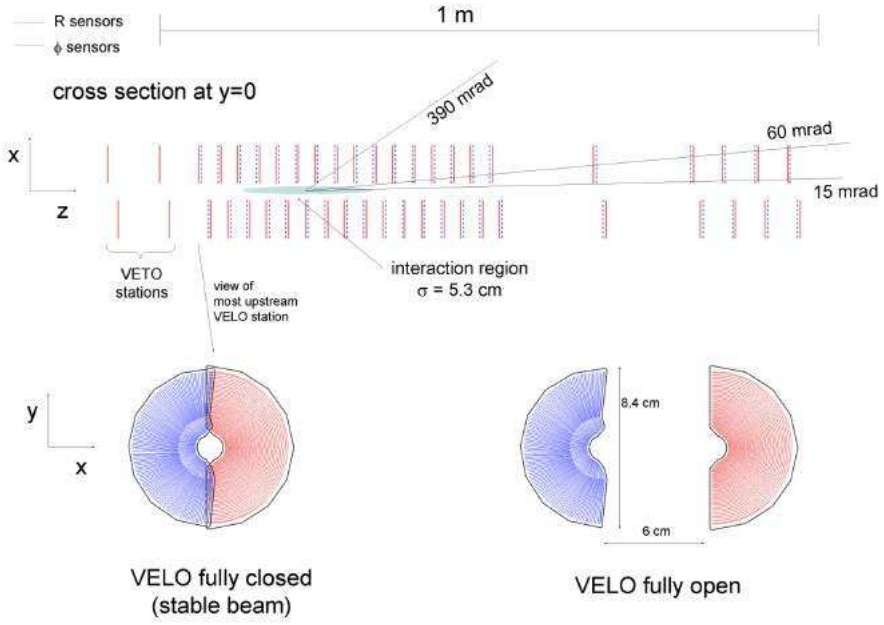


Figure 1.6: Schematics view of VELO system. Lower figure: VELO modules in closed and open position.

distances from beams with semicircular-shaped strips for R sensors or azimuthal angles with strips oriented approximately in radial directions for ϕ sensors. VELO sensors are mounted on a remote-controllable positioning system, which allows them to be moved in order to prevent damage to the sensors when the beams are not stable, since the distance between the sensitive area and the beam is smaller than necessary for safety. Figures illustrating VELO open and VELO closed are shown in fig. 1.6. A silicon strip provides the measurement of radial coordinates for the first four VELO modules, referred to as veto or pile-up, which are used to determine how many particles are generated in a bunch-bunch collision, an input for luminosity measurements. Additionally, these sensors improve the performance of the primary vertex reconstruction by taking into account also tracks with negative pseudorapidity. Within LHCb acceptance, every track passes through at least three modules. At the optimal track angle, the tracks reconstructed in VELO are used to calculate PV with the best spatial resolution of about 4 μm. The Impact Parameter, which represents the distance between the track's point of closest approach and the PV itself, is measured with a resolution of 44 μm for particles with transverse momentum of 1 GeV and 15 μm for particles with larger transverse momentum. In order to shield the VELO from the effects of electromagnetic fields generated by the high-frequency beam structure, the two VELO halves are mounted in an $AlMg_3$ box (RF-box) that separates the vacuum on the LHC beam pipe from the vacuum on the VELO. In order to minimize multiple scattering of the tracks, the thickness and material were chosen to minimize these effects. Multiple scattering degrades the resolution of the impact parameter significantly. The two boxes have a highly corrugated shape, which allows the two halves of the detector to overlap in the closed configuration of the VELO, as shown in Fig.1.7.

The Tracker Turicensis (TT) covers the full LHCb angular acceptance and consists of four planar

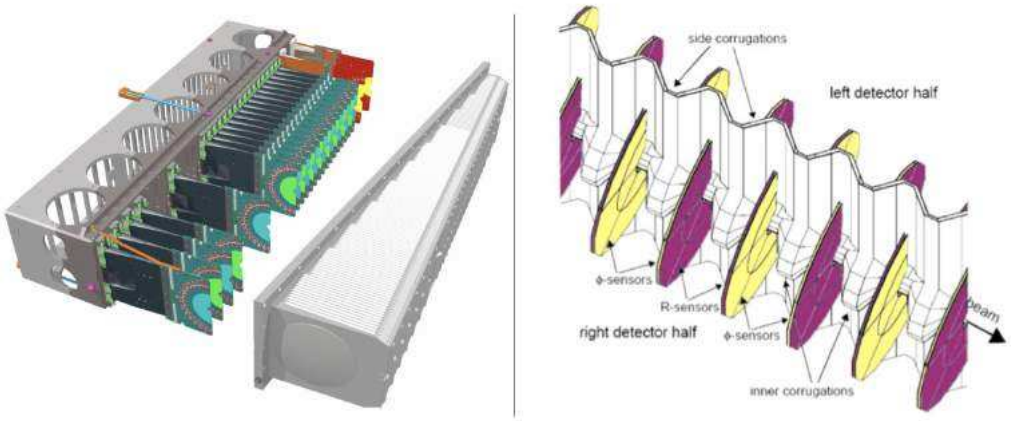


Figure 1.7: (Left) Module support and (Right) inner view of the RF-boxes, with the detector halves in the fully closed position. The edges of the boxes are cut away to show the overlap between the sensors of the two halves. R- and ϕ-sensors are colored in yellow and purple, respectively.

layers of silicon microstrips arranged in two pairs, TTa and TTb, respectively. About 30 cm separates the stations along z. All the layers consist of 500 μm thick, 9.64 cm wide and 9.44 cm long silicon sensors carrying 512 readout strips each one with a pitch of 183 μm. Aiming to minimize ambiguities in the reconstruction of tracks, the first and last layers are arranged in vertical strips measuring the x coordinate, while the two central layers are rotated by 5 degrees by $\pm$ 5° (fig. 1.8). TT provides a rough

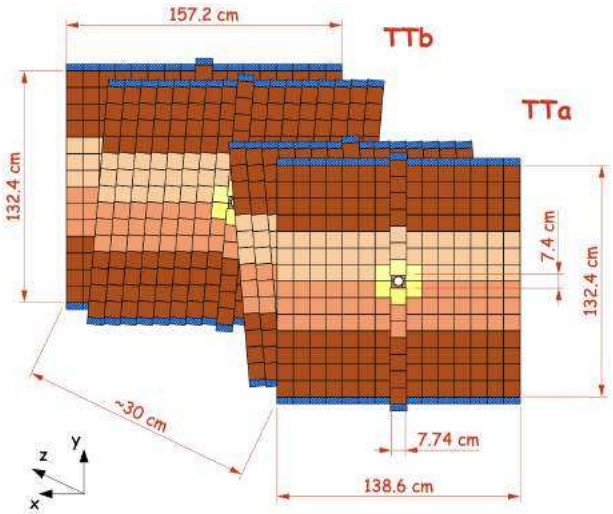


Figure 1.8: Scheme of the TT subdetector. Different readout sectors are indicated by different shadings.

measure of particle momenta without having to reconstruct the tracks downstream of the magnet since the tails of the dipole magnetic field extend into the area where the TT is installed. As a result of this feature, it is feasible to predict the rough trajectories of the tracks as they approach the T stations, which will reduce the number of upstream-downstream track combinations that need to be fitted

during the reconstruction of collision events.

Magnet

The LHCb magnet [37], as illustrated in the Fig.1.9, consists of two coils arranged in an angle that matches the acceptance angle of the LHCb detector. In Fig.1.10, the evolution of the magnetic field profile with Z is shown in comparison with the positions of the tracking detectors. Approximately 4 Tm of bending power is generated, with a peak intensity of 1,1 T directed toward the y-axis. Periodically, the magnetic field polarity is reversed by inverting the current in the magnet.

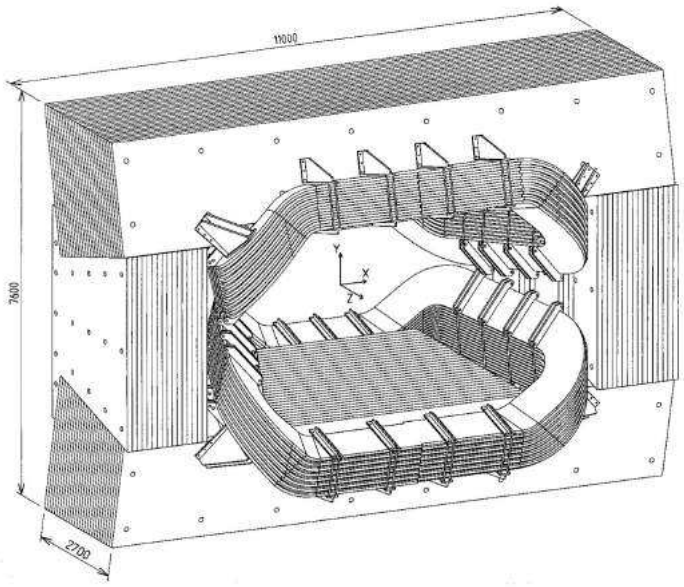


Figure 1.9: Layout of the LHCb magnet

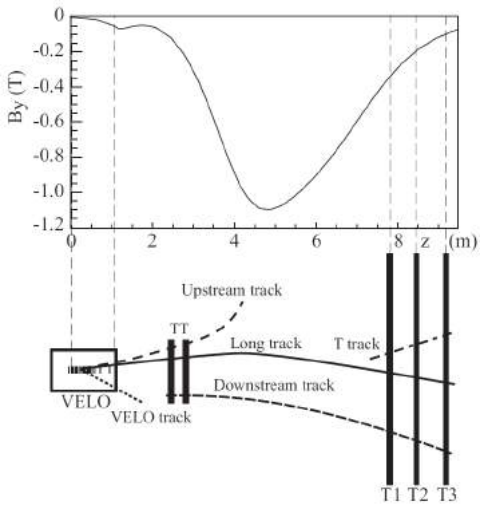


Figure 1.10: Evolution of the magnetic field strength as a function of the z coordinate.

T–stations

Three tracking stations (T1-T3) are installed downstream of the magnet and cover an area of approximately (6×5) m^2. They are divided into two parts: "Inner Tracker" [38] (fig. 1.11) and "Outer Tracker" [39] (fig. 1.12) depending on their distance from the beam line. As in the TT, the "Inner Tracker" is implemented with four silicon microstrip planes, while the "Outer Tracker" is constructed from four rows of staggered straw-tube drift chamber detectors that operate in the proportional-counter regime and cover the larger polar angles. The design is driven by the requirement that the detector occupancy is lower than 10% and the single-hit detection efficiency in each layer is greater than 99%.

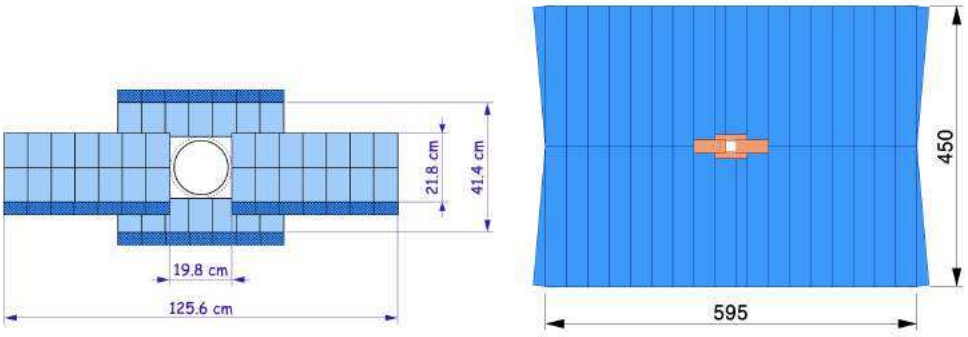


Figure 1.11: Scheme of the IT subdetector: (Left) front view of a layer of the inner tracker, (Right) front view of one of the T–stations. The inner tracker is drawn in orange and the outer tracker in blue [**siliconTracker**].

1.2.2 Tracking performances

Tracks that hit all three tracking detectors (VELO, TT, T-stations) are called long tracks. There is a widespread use of long tracks in physics analyses due to their high momentum resolution (0.5-1.0% for momenta within 20 and 100 GeV/c). Additionally, long tracks can provide a good resolution of impact parameters with respect to a PV position, parametrized as $\sigma_{IP} \sim 15 + 29/p_T$ μm with p_T measured in GeV/c, but this also depends on the number of tracks used for reconstruction. Due to the absence of VELO hits, downstream tracks are reconstructed using only TT and T-station hits. These tracks result from the decay of long-lived particles, including K^0_S and Λ. In general, VELO and TT hits are the only tools used to reconstruct the upstream tracks since they are low-momentum and deviated from the T-station's acceptance. Based on the small curvature generated by the residual magnetic field in the TT region, the momentum can be assigned, though with a low degree of accuracy of 15%. It is useful to reconstruct the primary vertices with the help of VELO tracks. These tracks contain only hits information in the VELO and they correspond to particles generated with a high angle of incidence. Secondary interaction mainly generates T-tracks, which contain hits only at T-stations. Tracks reconstructed based on a random combination of hits are identified as *ghost*. Multivariate algorithms have been developed to quantify the probability that a track is a ghost. The long tracks are reconstructed first, then the downstream and upstream tracks are reconstructed using the remaining

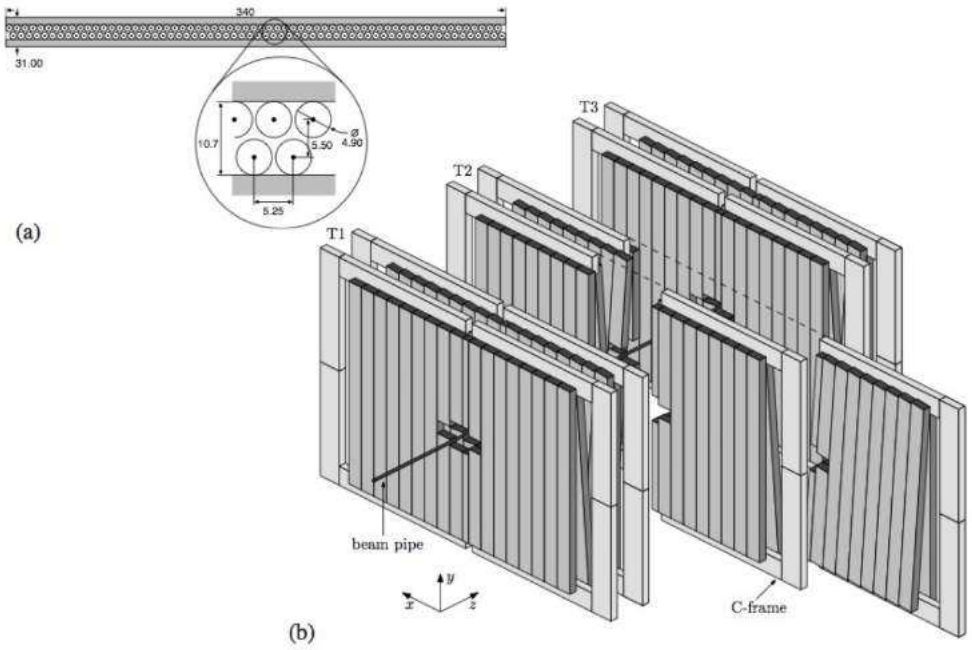


Figure 1.12: Scheme of the OT subdetector. (a) Section of an OT detection plane. (b) Arrangement of OT straw-tubes chambers modules in planes and stations. Each station can be opened for maintenance, as shown for the second station [**outerTracker**].

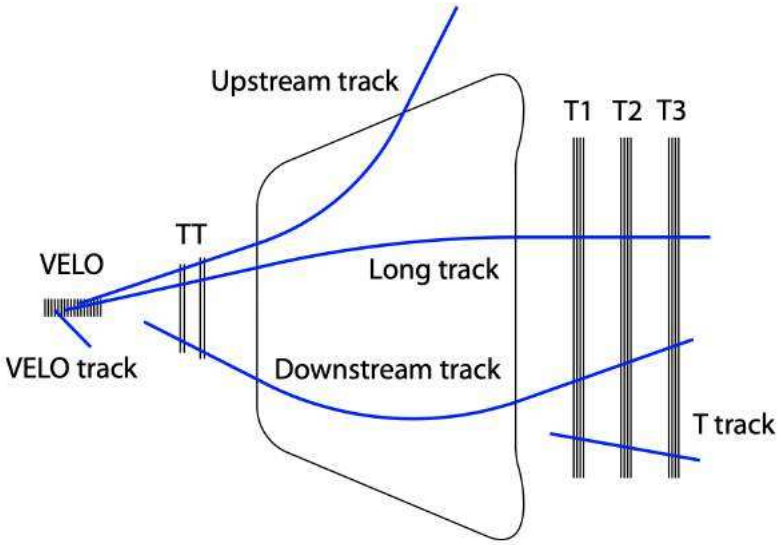


Figure 1.13: Representation of different types of tracks in the LHCb detector.

tracks. This procedure can be broken down into three steps:

- By identifying patterns within the VELO hits, segments are created that correspond to straight lines among the hits. Tracks from VELO are the seeds for long tracks. Following the hypothesis of a parabolic trajectory, the possible hits in T2 are calculated by creating a segment matching

hits in T1 and T3. If the hits are found, they are added to the segments in the VELO and then combined with the seed segments in the T stations.

- A Kalman filter is applied to track candidates in order to accurately measure momentum and charge associated with particles leaving the track while accounting for scattering and ionization effects.
- Checking how many hits are shared between two tracks in order to remove duplicate tracks. The quality of a track is evaluated by measuring the $\chi^2/ndof$ in the kalman filter.

Figure 1.14 shows the tracking efficiency as a function of the momentum in 2011 and 2012.

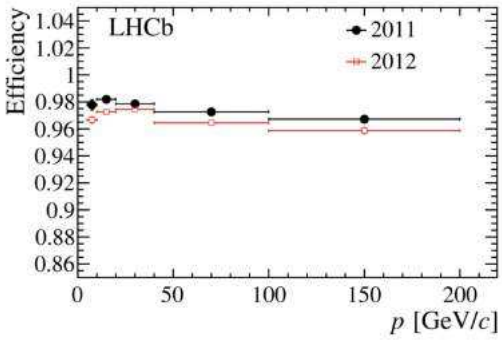


Figure 1.14: Tracking efficiency as a function of the momentum. The tracking efficiency represents the probability for a particle to be reconstructed using hits along its trajectory.

1.2.3 Particle identification

As a charged particle's momentum is measured by the tracking system, its mass can be calculated once its velocity or energy is known. It is crucial to determine the particle species in the interest momentum range (2-450 GeV) using Particle IDentification (PID) strategies. The PID relies on information from the RICH detectors, the two calorimeters and the muon detector system.

RICH Detector System

A key component of the LHCb flavour physics program is charged hadrons discrimination, namely the distinction between pions, kaons, and protons. The RICH system [40] provides hadron PID in LHCb in the 2.6 - 100 GeV momentum range, and it is essential for measurements in LHCb with Run 1 and 2 data [41]. The RICH system makes it possible to differentiate between final states of otherwise identical topologies, such as $B^0_{(s)} \to \pi^+\pi^-$, $K^+\pi^-$, K^+K^- decay modes; greatly reduce the combinatorial background in decay modes involving hadrons in the final state, such as $B^0_s \to \Phi\Phi$, where $\Phi \to K^+K^-$, which would be impractically large without PID requirements. The RICH PID information is exploited also at the HLT2 level trigger to reduce combinatorial background.
Two detectors, RICH1 and RICH2, constitute the RICH system. Both detectors use a system of spherical and planar mirrors to focus the Cherenkov photons produced inside fluorocarbon gaseous radiators outside the LHCb acceptance and onto the photon detection planes.

RICH1 covers the entire spectrometer angular acceptance of 25–300 mrad and is positioned upstream of the LHCb dipole magnet. The size of the beam pipe upstream of the magnet determines the low angle acceptance limit. The low-intermediate momentum range, 2-40 GeV/c, is covered by RICH1. The angular acceptance range of the RICH2 detector, which is located downstream of the LHCb dipole, is 15-120 mrad. RICH2 is optimized to cover the high momentum range, 15-100 GeV/c, populated by particles that are more likely to leave tracks at smaller angles.

To minimize its physical size, RICH1 is positioned as near as feasible to the LHCb interaction point. The detector is positioned right after the VELO, and its gas enclosure is directly sealed to the VELO vacuum tank to reduce the material budget. The RICH1 exit window is made of a low-mass carbon fiber/foam material. RICH2 is positioned downstream of the magnet, immediately following the tracking system. Similar to the RICH1 exit window, the RICH2 entrance and exit windows are foam sandwich constructions that are covered in carbon fiber and aluminum, respectively.

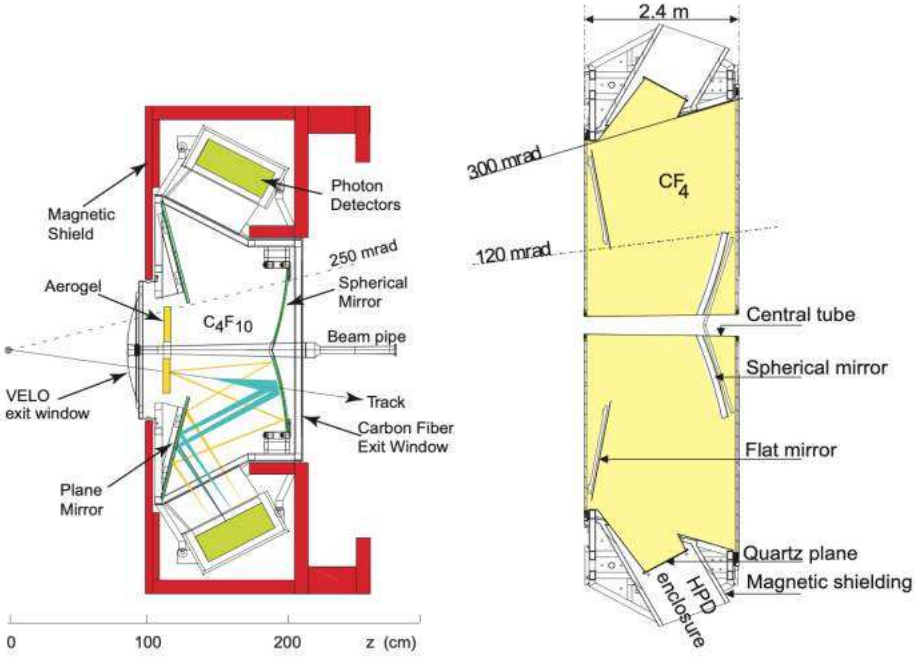


Figure 1.15: Schematic side views of (Left) RICH1 and (Right) RICH2 detectors.

A hybrid photon detector (HPD) array was used in Run1 and Run2 to detect photons at wavelengths between 200 and 600 nanometers. A system of spherical and planar mirrors redirects the photons to the HPDs, which are placed inside iron boxes to shield them from magnetic fields. Particle identity is determined in the RICH system by matching tracks (produced by VELO and TT) to hits on the photodetector planes and selecting the most likely candidate particle. By utilizing the Cherenkov angle, the RICH pattern recognition method fits the associated photon hits of the charged particle tracks with the expected hits pattern via the maximum likelihood method. A typical event in RICH1 is shown in fig 1.17, the number of photoelectrons detected, and as a consequence, the average number of hits per reconstructed ring, depends on the radiator material and track momentum. In RICH1 both small diameter, densely populated rings from the C_4F_{10} (hits $\leq$ 30 and radius $\simeq$ 5 cm) and large sparsely populated rings from the aerogel, before its removal, (hits $\leq$ 6 and radius $\simeq$ 15 cm)

can be formed.

In fig. 1.16 (Left) is shown how the Cherenkov angle saturates at the value of $\theta_c = arccos(1/n)$ by considering different radiators. Fig.4.31.16 (Right) shows the Cherenkov angle based on particle momentum using data from C_4F_{10} for particles detected within LHCb acceptance.

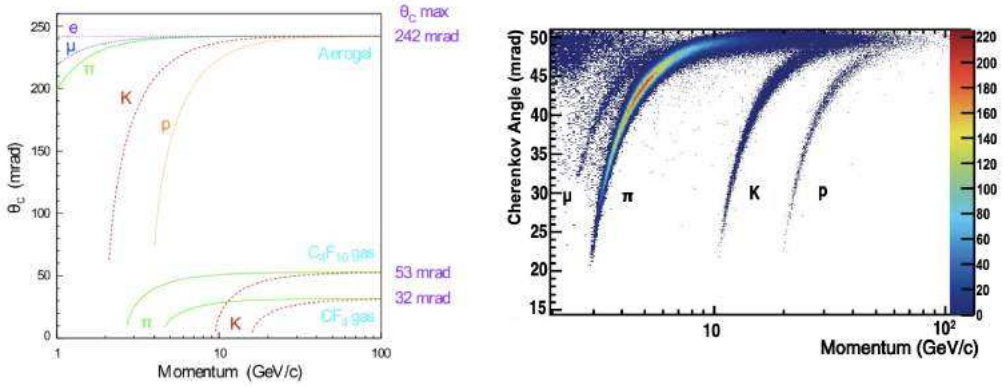


Figure 1.16: (Left) Cherenkov angle as a function of particle momentum for the RICH radiators and for different particle masses and (Right) for isolated tracks, whose Cherenkov ring does not overlap with any other ring, in the C_4F_{10} radiator.

In physics analyses, a particle identification classifier is defined using a RICH likelihood function combined with information from the calorimeter and muon systems. The difference between two particle hypotheses, h_1 and h_2, is typically expressed as a log-likelihood difference DLL (also commonly called PID):

$$DLL_{h1,h2} = log\frac{\mathscr{L}(h_1)}{\mathscr{L}(h_2)}$$

An example is DLLp, which represents the log difference between the proton and the pion hypotheses, whereas DLLK represents the log difference between the kaon and the pion hypotheses.

The main figure of merit to evaluate the RICH detector performance is the Cherenkov angle resolution. In both RICH detectors, the single photon resolution is determined using tracks with high momentum. The difference between the reconstructed and expected Cherenkov angles forms a peak around zero when photons emitted from the tracks are considered for the prediction. A background of random noise or photons from other tracks appears below the peak. Obtaining the resolutions involves fitting the correct matched photons with a Gaussian distribution and modeling the background as a second-order polynomial. As a result, the width of the Gaussian distribution defines the Cherenkov angle resolution (Fig.1.18). Based on Run 2 data, Fig. 1.19 shows how track momentum affects PID efficiency and misidentification rate. A loose DLL cut is applied for high signal efficiency, and a tight cut for good background rejection.

Calorimetric system

Calorimetric measurements of electrons, photons, and hadrons are performed downstream of the RICH2. Four subsystems make up the entire Calorimetric system [42], which is divided into two modules.The system also selects candidates with high transverse energy for the first Level trigger (L0). The

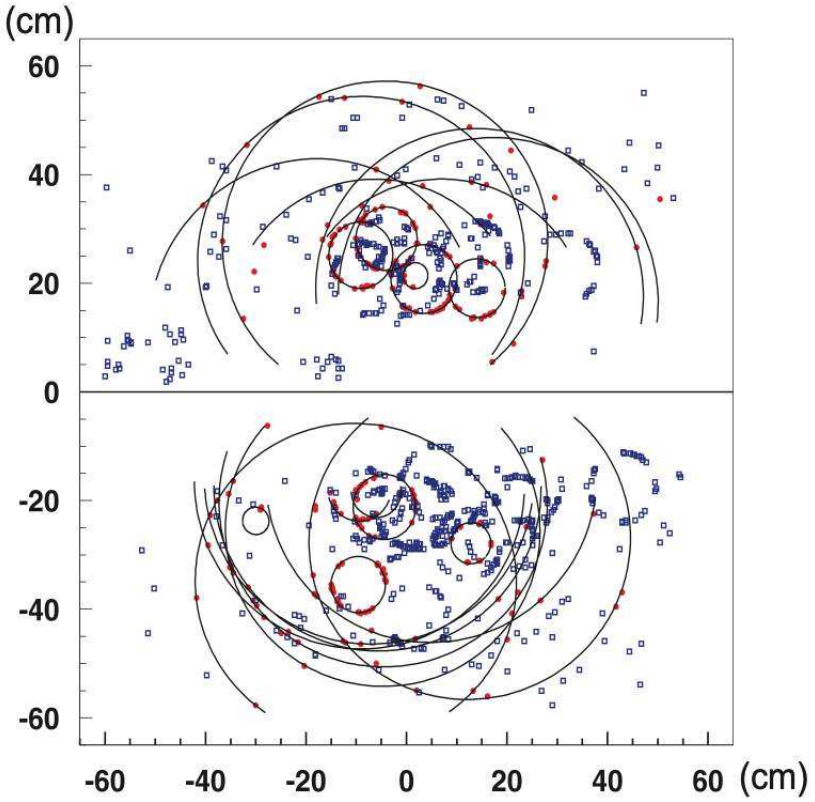


Figure 1.17: Typical LHCb event in RICH1 with Cherenkov rings interpolation. The small (large) ring radii in RICH1 originate from the C4F10 (aerogel) radiator. The solid red rings indicate particles passing through the whole detector and the blue dotted rings indicate the particles for which no match is found with tracks in the VELO and TT detectors.

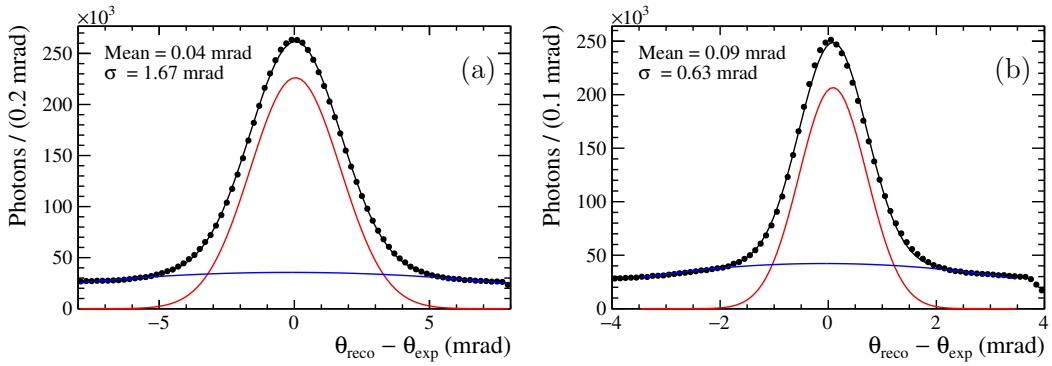


Figure 1.18: Example of photon distributions detected in low track-multiplicity events in RICH1 and RICH2. The total distribution (black) is overlaid with a Gaussian signal component (red), a polynomial background component (blue), and the Gaussian signal component (red).

Scintillating Pad Detector (SPD) and the PreShower (PS) are the first calorimetric modules. A layer of lead is interleaved between two scintillator pads with active surfaces 7.6 meters wide and 6.2 meters high, corresponding to 2.5 radiation lengths, X_0, and ~ 0.06 hadronic interaction lengths, λ_I. By us-

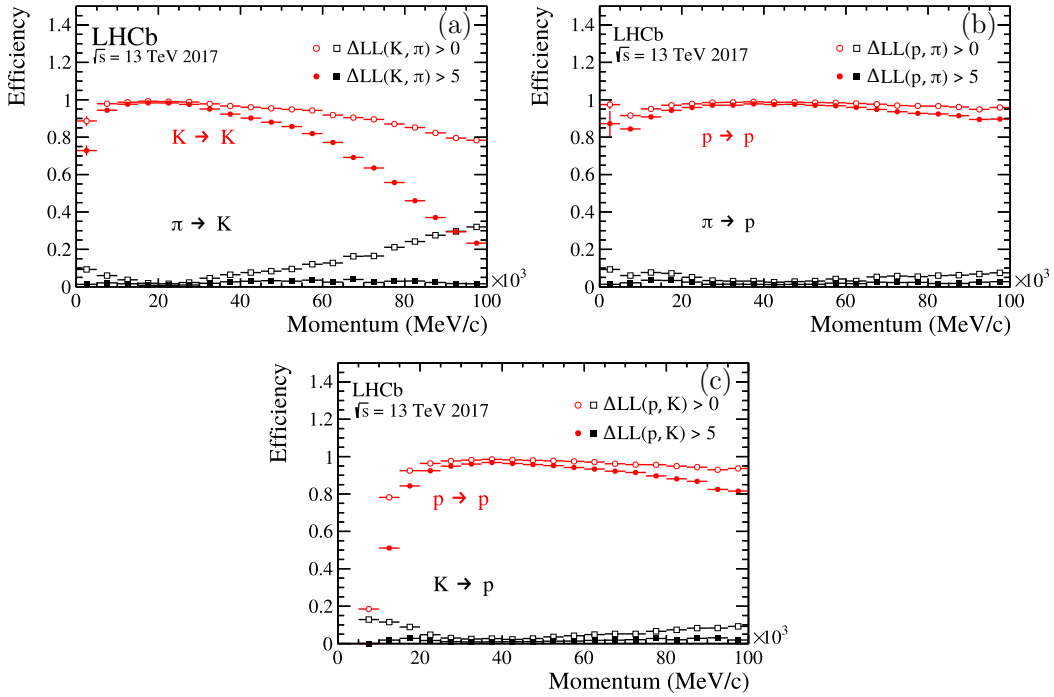


Figure 1.19: Selection efficiency of kaons (a), protons (b and c) with associated leakage associated with misidentifying pions (a and b) and kaons (c). In this case, two selections have been made, a loose selection (hollow circles) and a tight selection (solid circles).

ing SPD, it is possible to distinguish photons from electrons since photons do not deposit energy upstream of the lead layer. The energy deposits in the various detectors from different types of particles are shown in fig. 1.21. The Electromagnetic Calorimeter (ECAL) and Hadronic Calorimeter (HCAL) are then assembled with alternating tiles of plastic scintillator and absorber material. In addition to covering the full LHCb angular acceptance, the ECAL measures 7.8 m in width and 6.3 m in height and is located 12.5 m downstream of the point of interaction. In this case, the absorber is made of lead, with a thickness equal to $25X_0$ and $1.2\lambda_{int}$. The energy resolution achieved is

$$\frac{\sigma(E)}{E} = 1\% \bigoplus \frac{10\%}{\sqrt{E/GeV}} \tag{1.4}$$

The first term is due to a constant contribution, while the second term is due to statistical uncertainty about the energy deposit. HCAL measures 8.4 meters by 6.8 meters and is located 13.3 meters downstream of the interaction point. This design covers only 5.6 λ_I due to space limitations. The energy resolution achieved is

$$\frac{\sigma(E)}{E} = 9\% \bigoplus \frac{69\%}{\sqrt{E/GeV}} \tag{1.5}$$

Photomultipliers read all detectors using wave-shifting fibers to transmit light. Scintillator tiles are segmented according to distance from the beam pipe in SPD, PS, and ECAL. The inner region has the highest occupancy, as shown in Fig. 1.20. Instead of leaving hits in the SPD station, the photons interact with the lead converter, creating an electromagnetic shower in the PS and ECAL. At the PS

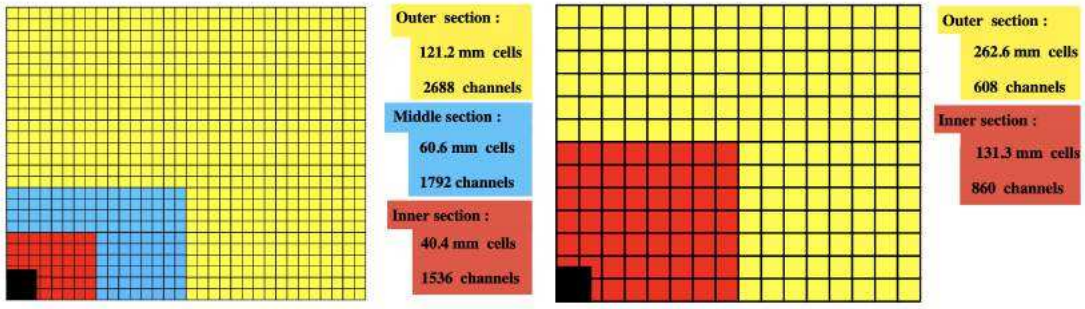


Figure 1.20: Segmentation of one quadrant (left) of the SPD, PS and ECAL and (right) of the HCAL. The black sector, corresponding to the beam pipe, is outside of the LHCb acceptance.[**lhcbdetect**]

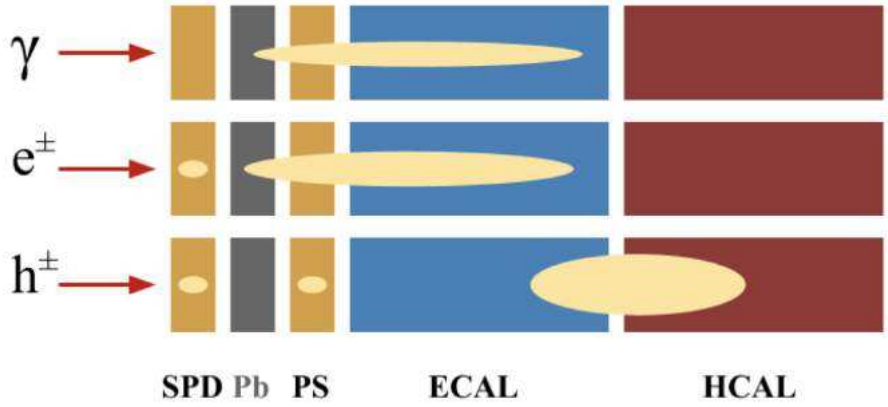


Figure 1.21: Representation of typical electromagnetic and hadronic showers and their interaction with the different detectors of the calorimetric system.

and ECAL levels, electrons behave similarly to photons, but they also leave hits on the SPD detector. Neutral hadrons can be merged or resolved depending on whether they are reconstructed from a single cluster in ECAL or from two photons. Hadrons are typically Minimum Ionising Particles (MIPs) in the SPD, PS, and ECAL (in the latter, the energy deposit can be extremely variable).

Muon System

A muon detection system is used to detect muons and provide event selection information to the experiment's L0 trigger [43]. A total of five rectangular stations (M1-M5) compose the system: M1 is located in the highest rate region and is equipped with triple Gas Electron Multipliers (GEMs); M2-M4 are equipped with Multi Wire Proportional Chambers (MWPCs).

A total of 1380 chambers occupy about 435 square meters. Calorimeter station M1 provides information for p_T trigger measurements and is located in front of the calorimeters. For selecting penetrating muons, 80 cm thick iron absorbers are installed between stations M2 to M5 downstream of the calorimeters. In order for a muon to cross the five regions, its momentum must be at least 6 GeV/c. Each of the five stations provides a projective geometry. R1 to R4 are the four regions within each

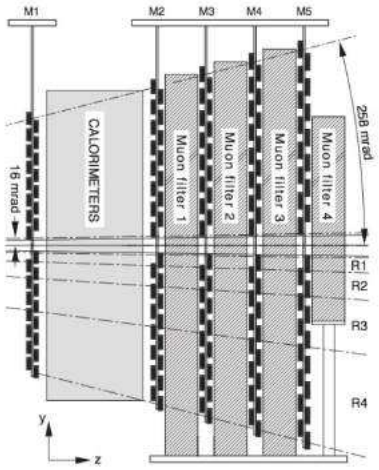


Figure 1.22: Side view of the muon system along the beam axis.

station, arranged in increasing distance from the beam pipe. There is a ratio of 1:2:4:8 between the segmentation scales of the regions R1, R2, R3, and R4. Channel occupancy is comparable in each of a station's four regions as a result of this geometry.

1.2.4 Triggers and online data selection

LHC collides particles with a rate of 40 MHz, if every event has to be stored it implies a flow of $\sim$ 1 TB of data per second so filters and triggers are required [44]. To manage that data flow from the detector, LHCb has a system of triggers [45] both hardware and software to select which events to store.

The Level 0 trigger reduces the signal rate to 1 MHz and it's hardware-based. The calorimeters and the muon front-end electronics send a signal to arrays of FPGA to process data at 40 MHz and select hadrons and lepton with large transverse momentum (in Run 1 for example $p_t > 3.7\ GeV$ for hadrons, $p_t > 3.7\ GeV$ for photons or electrons and $p_t > 1.76\ GeV$ for muons). An event with too many tracks is difficult to reconstruct so a requirement on the number of tracks is also included.

An event accepted by L0 is passed to an Event Filter Farm where software selection is performed. High-Level Trigger (HLT) in the Event Filter Farm is software based and divided into two main stages. The first stage filter process (HLT1) is based on a partial event reconstruction and refines L0 decisions. HLT1 reconstructs track segments in the VELO and it extrapolates the ones with a high impact parameter, the track segments are then matched with the information given by the tracking stations looking for high transverse momentum tracks. The next stage filter process (HTL2) runs asynchronously with respect to HTL1 and uses particle identification information provided by RICH and tracking detectors on the data accepted from HLT1 ($\sim$ 100 Hz). Events accepted by the trigger are sent to the storage system with a rate of $2-5\ KHz$ for later offline analysis. A schematic view of LHCb data flow can be seen in Fig. 1.23. The detector alignment and calibration was implemented online starting with Run2 (Fig.1.24) so that the HLT2 software trigger could perform offline quality reconstruction events. The experimental results were substantially improved by reducing systematics relating to online/offline

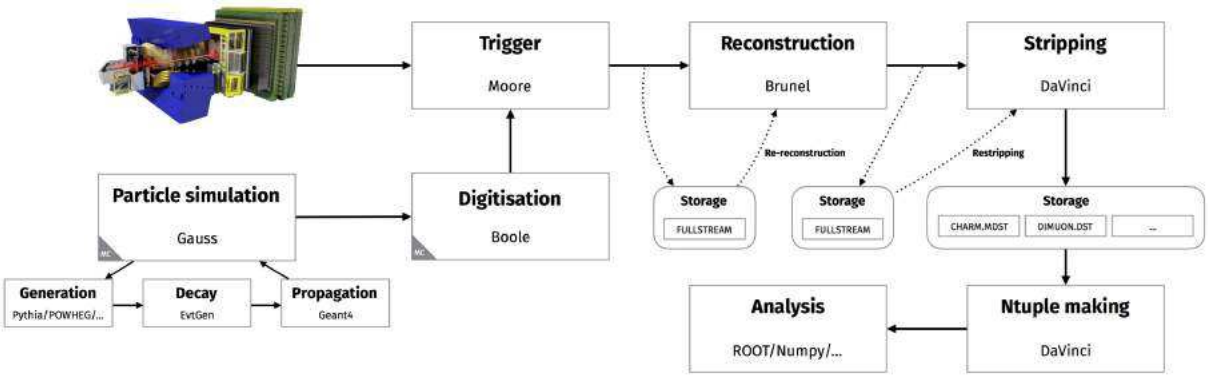


Figure 1.23: LHCb data flow scheme

reconstruction differences and removing data reprocessing.

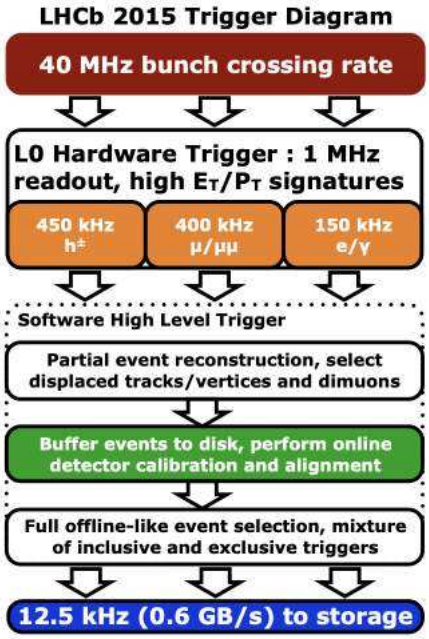


Figure 1.24: Trigger scheme of the LHCb experiment after the Run2 (2015–2018) improvements.

1.3 LHCb Upgrade 1

By the end of the Run2 data-taking period, the main limitation for the LHCb detector was determined by the readout trigger architecture which limited the running luminosity. To allow data operation at an increased luminosity of $2 \times 10^{33}\ cm^{-2}s^{-1}$, LHCb has been upgraded enabling a detector readout running at 40 MHz instead of the previous 1 MHz with the aim to collect at least 50 fb^{-1} of data by the end of 2028.

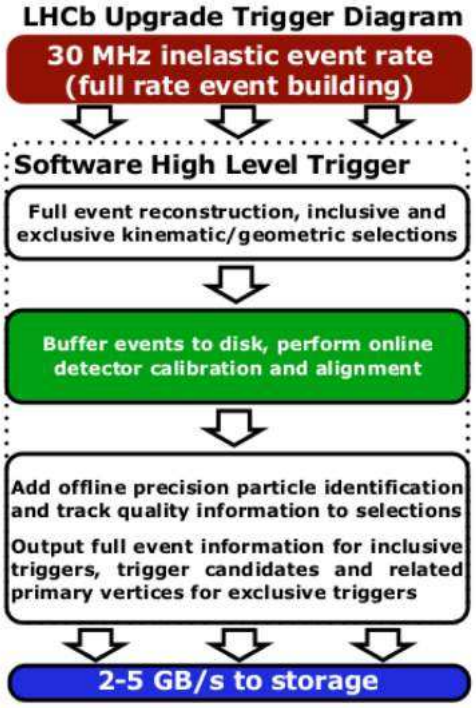


Figure 1.25: Run3 trigger scheme of the LHCb experiment.

For the Run3, the hardware trigger is replaced by a fully software-based which uses information from the calorimeter system and muon with an output of 15-30 MHz [46] running on GPUs [47]. The front-end electronics is also sending data synchronously at 40 MHz to the back-end electronics. Several PCIe40 cards will be used for data readout, slow control, and fast control. Data will be buffered to disk, and the software trigger will perform the first selection and reconstruction at a rate of 2-5 GB/s (see Fig.1.25).

1.3.1 The new LHCb detector layout

The upgrade requires higher radiation tolerance and granularity, particularly for the tracking sub-detectors.

As part of the upgrade, the VELO is be placed closer to the beam axis, the radiation lengths is be reduced from 4.6% to 1.7%, and the strip technology is replaced by pixels [48]. This will result in a 40% improvement in the impact parameter resolution. In place of the existing tracking stations, an Upstream Tracker (UT) is located upstream of the magnet, and a Scintillating Fiber Tracker (SciFi) is placed downstream [44].

The silicon strip technology is mounted on four tracking layers constituting the UT. In the Scifi, twelve layers, each with 2.4 meters long scintillating fibres with a diameter of 250 micrometers, are arranged vertically. The fibres are read out by Silicon Photon Multipliers placed on top and bottom of the layers cooled to -40 degrees Celsius to reduce radiation damage and dark noise. The upgrade of RICH system is discussed in the next chapter. There are no major upgrades to the calorimeters or muon chambers apart from the replacement of the front-end electronics and the removal of the

station used in the old L0 trigger. The Run3 version of the LHCb experiment is shown in fig. 1.26.

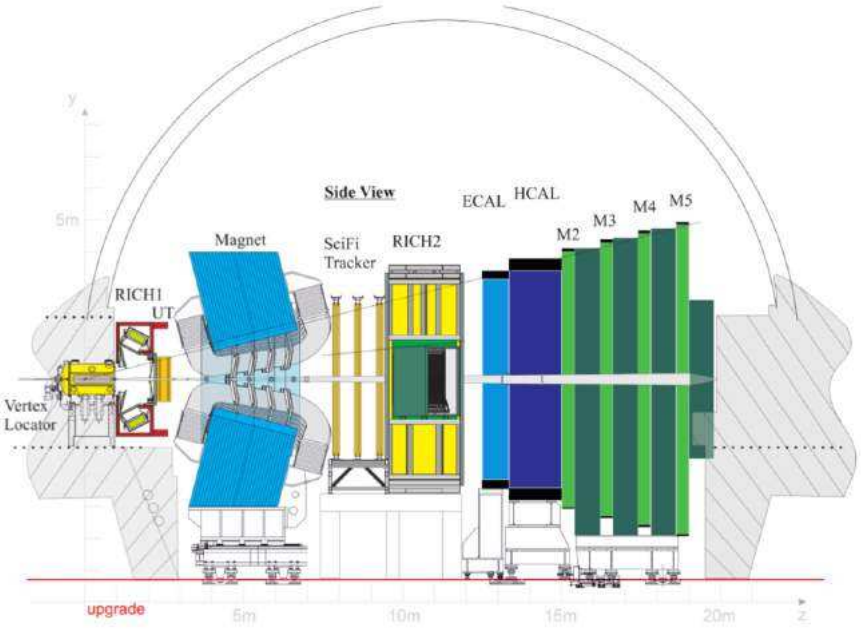


Figure 1.26: Side view of the upgraded LHCb experiment.

Chapter 2

Luminosity at LHCb and the SMOG system

The luminosity in LHCb has a major role in affecting the detector performance and it has to be controlled carefully during data-taking with a procedure called luminosity leveling. In this chapter, the luminosity operations in LHCb are covered, being important to present the studies I performed in 2022, at the beginning of Run3, to adapt the RICH detector as a luminometer for luminosity monitoring (described in Chap.4). The luminosity also introduces the SMOG system, a unique LHCb gas system injection that was originally designed to perform absolute luminosity measurements through the Beam Gas Imaging technique. The scope of SMOG, on the other hand, extends beyond its application for luminosity: since 2015 provides the unique possibility to perform fixed-target measurements in LHCb by injecting noble gases into the beam pipe. Data are then collected in a beam-empty configuration, where protons (or Pb) collide with the injected He, Ar or Ne gas, exploiting the forward single-arm design of the LHCb detector. On these fixed-target data relies a lot of my PhD work presented in the Chap.5-6.

2.1 Luminosity

A fundamental parameter contributing to the performance of high-energy physics experiments is the *instantaneous luminosity* $\mathscr{L}$. It is a measurement of the number of collisions that can be produced in the detector per cm^2 and per second, and corresponds to the ratio between the event rate R, i.e. the number of events detected in a fixed time interval, and the interaction cross section σ.
At LHC experiments an important parameter related to the luminosity, that can be measured directly, is the average number of visible pp interactions per bunch crossing μ. It can be expressed as:

$$\mu = \frac{\mathscr{L} \cdot \sigma_{bb\,vis}}{N_{bunches} \cdot \nu_{LHC}} = \frac{\mathscr{L}}{N_{bunches} \cdot 0.177 \cdot 10^{30}} \,.$$

where $\sigma_{bb\,vis}$ is the visible cross-section for pp collision in LHCb estimated equal to 63.6 mb and ν_{LHC} = 11.245 kHz is the LHC frequency. Another important running parameter is the *pile-up*, i.e. the average number of pp interactions in visible events. Assuming a Poisson distribution for the visible events, the pile-up is calculated after zero suppression as:

$$pile-up\,(\mu_{vis}) = \frac{\mu}{1-e^{-\mu}} \,.$$

The instantaneous luminosity in 2010 was 10^{28} cm^{-2}s^{-1} and the pile-up was nearly zero. During LHC Run I a luminosity of 10^{32} cm^{-2}s^{-1} with a pile-up of $\mu_{vis} \sim 2.5$ was reached. Despite the highest luminosity in 2010 being already 75% of the LHCb design luminosity (2×10^{32} cm^{-2}s^{-1}), the pile-up was larger than the design value by ~ 0.4 due to the low number of bunches in LHC. In such conditions, the trigger and reconstruction processes worked efficiently and the increased detector occupancy due to the larger pile-up didn't affect the physics performance. With a bunch spacing of 50 ns, the machine was gradually able to handle about 1300 bunches in the first months of 2011. During the year, as the number of bunches increased, the pile-up decreased and LHCb took the majority of the data at a luminosity of 3.5×10^{32}cm^{-2}s^{-1}. Compared to the design luminosity, this value was 1.75 times higher. Therefore, in 2012 it was decided to introduce a luminosity leveling procedure in order to compensate, in contrast to ATLAS and CMS. During this procedure, beam overlaps in the vicinity of the LHCb interaction point are adjusted. A maximum overlap with head-on beams is only achieved after 15 hours of taking data and the instantaneous luminosity can be maintained constant to within 5%. Using this method minimizes the effects of luminosity decay on LHCb and reduces systematic uncertainties associated with detector occupancy changes during a fill. For a detailed overview of the LHCb luminosity in Run2 see [49]. The procedure is shown in Fig. 2.1. The luminosity leveling

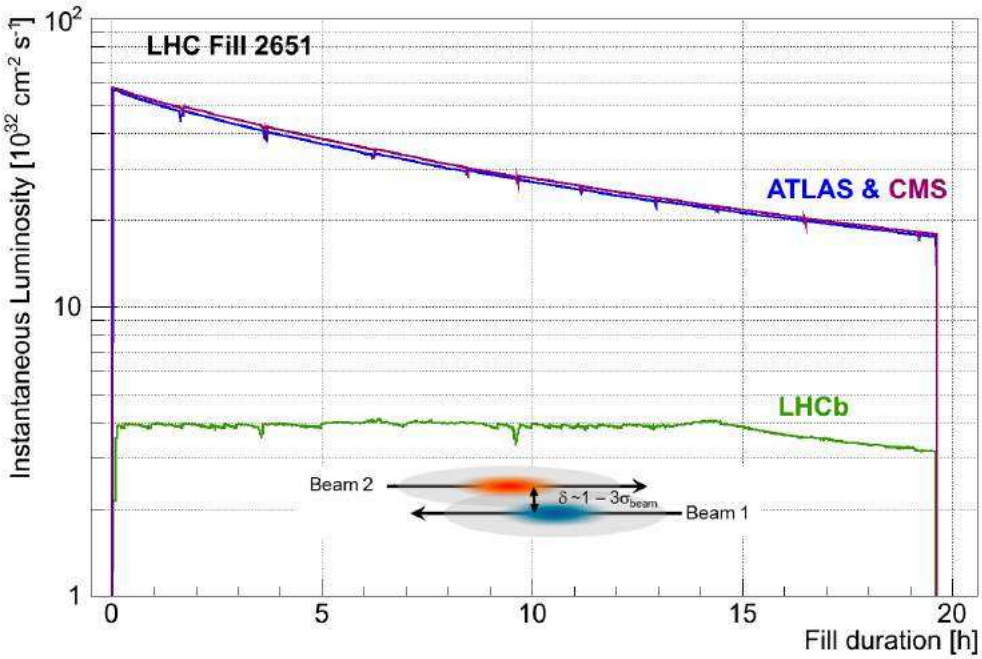


Figure 2.1: Time dependence of the instantaneous luminosity for ATLAS, CMS and LHCb during LHC fill 2651. After ramping to the desired value of 4×10^{32}cm^{-2}s^{-1} for LHCb, the luminosity is kept stable in a range of 5% for about 15 hours by adjusting the transversal beam overlap. The difference in luminosity towards the end of the fill between ATLAS, CMS and LHCb is due to the difference in the final focusing at the collision points.

requires real-time feedback from the experiment. As a result, the beams are continuously steered in order to keep a constant luminosity value and optimize tracking performance. The instantaneous

luminosity $\mathscr{L}$ for each colliding bunch pair can be defined as:

$$\mathscr{L} = \frac{\mu_{\text{inel}}\, f_r}{\sigma_{\text{inel}}} = \frac{\mu_{vis} f_r}{\sigma_{\text{vis}}} \text{ with } \begin{array}{l} \mu_{vis} = \varepsilon\mu \\ \sigma_{vis} = \varepsilon\sigma \end{array} \tag{2.1}$$

where the parameters are defined as follows:

f_r = LHC revolution frequency (11245 Hz);

σ_{inel} = total inelastic pp cross-section ;

μ_{inel} = number of inelastic pp collision per bunch crossing ;

ε = acceptance x efficiency of luminosity detector ;

μ_{vis} = visible collision per bunch crossing ;

σ_{vis} = effective cross-section .

The luminosity is measured using a proxy variable, for which the visible collisions and the effective cross-section are defined, referred to as luminometers (or luminosity counters). Specifically, the μ_{vis} is a relative luminosity measurement and it consists in counting the visible interactions for the particular luminometer, and the σ_{vis} is the absolute luminosity measurement.

2.1.1 Luminosity Counters

The instantaneous luminosities vary over the course of a fill (which typically lasts many hours) depending on factors like bunch populations and beam transverse emittances. Thus, luminosity measurements should be carried out in a short enough time interval to ensure that these quantities remain relatively stable such to provide both a reliable feedback for the leveling process and a precise instantaneous luminosity determination. Accurate cross-section measurement of a physics process requires a precise determination of the relative luminosity. In general, any interaction rate measured as a result of the interaction of the beams can be proportional to the relative luminosity (this may not be true if other factors are involved, such as the electronic dead time during the digitization of the signal). The luminosity counters are variables whose rate is easy to measure and are recorded simultaneously with physics data to determine the relative luminosity during data-taking. The determination of the counter's interaction rate as a function of its cross-section is achieved during dedicated calibration fills (such as the VDM scans, described below). Each luminosity counter is based on the response of an individual subdetector. At LHCb, luminosity counters based on the VELO detector response (number of Tracks and reconstructed Vertices) have been shown to exhibit the greatest stability over long periods of time in Run2. The methods employed to obtain a relative luminosity measurement from a counter allow to estimate of μ_{vis}:

- linearity method: requires to obtain a mean number of interactions in a certain time interval. This is the most simple approach and requires the linearity of the counter response with luminosity in the full luminosity range:

$$\langle \mu_{vis} \rangle \propto \frac{\sum_i n_i}{N}$$

- Log-0 method: if a Poisson distribution for the number of counts per event is assumed, the number of empty events (defined with respect to a threshold which is counter-dependent) can exploit the relation:

$$\langle \mu_{vis} \rangle \propto log\, P(0) \propto -log \frac{N_0}{N}$$

  This method begins to fail or lose accuracy when the luminosity is high and the empty events decrease.

To perform luminosity leveling at IP8, LHC needs a precision of less than 5% on real-time μ_{vis}. For Runs 1 and 2, this measurement was primarily determined by measuring the transverse energy deposition in the calorimeters at the hardware trigger stage. For Run 3, LHCb switched to a software-only trigger system, removing the hardware-based system in favor of increased flexibility. In addition to the obvious computing challenges, it implies that a new solution is required for the control of luminosity and a fast, reliable feedback system for LHC. A dedicated detector, called PLUME (Probe for LUminosity MEasurement) [50], has been installed upstream of VELO and is now being tested on real collisions since May 2022. PLUME is the main LHCb luminometer and provides real-time μ_{vis} measurements to LHC for the luminosity leveling.

A method for absolute luminosity determination can be classified as either direct or indirect (from measuring the production rate of a process with a very well-known cross-section). In direct methods, luminosity is determined by measuring the parameters of colliding beams. In LHCb, the absolute luminosity calibration is determined by two direct methods: the van der Meer scan method (VDM) and the beam-gas imaging method (BGI), which takes advantage of the unique capabilities of the LHCb experiment. Through the VDM method, the colliding beams are moved in both transverse directions with high precision, and a relative rate is measured by scanning the overlap integrals at different beam positions. Using BGI, the angles, positions, and shapes of individual beams are measured without displacing them. Instead, the vertices of interactions between beam particles and gas nuclei are reconstructed in the primary vacuum around the nominal interaction region. A systematic uncertainty in beam overlap determination limits the precision of luminosity calibration in the LHCb experiment. VDM and BGI have different systematic uncertainties and, for the final result, the VDM and BGI calibration measurements are averaged. Due to the fact that absolute luminosity calibration can only be performed during specific running periods, a relative normalization method needs to be applied to the results of the absolute luminosity calibration over the entire period of data collection. This is accomplished by using several observables, each corresponding to a visible cross-section. Measurements of absolute luminosity during specific data-taking periods are employed to calibrate the cross-section for each variable. A calibrated visible cross-section is used to derive the integrated luminosity for a given period of time.

2.1.2 van der Meer Method

In the van der Meer method, colliding beams are moved in transverse planes (Fig. 2.2 and the rates are measured at different positions. In addition to the instantaneous luminosity, the observed

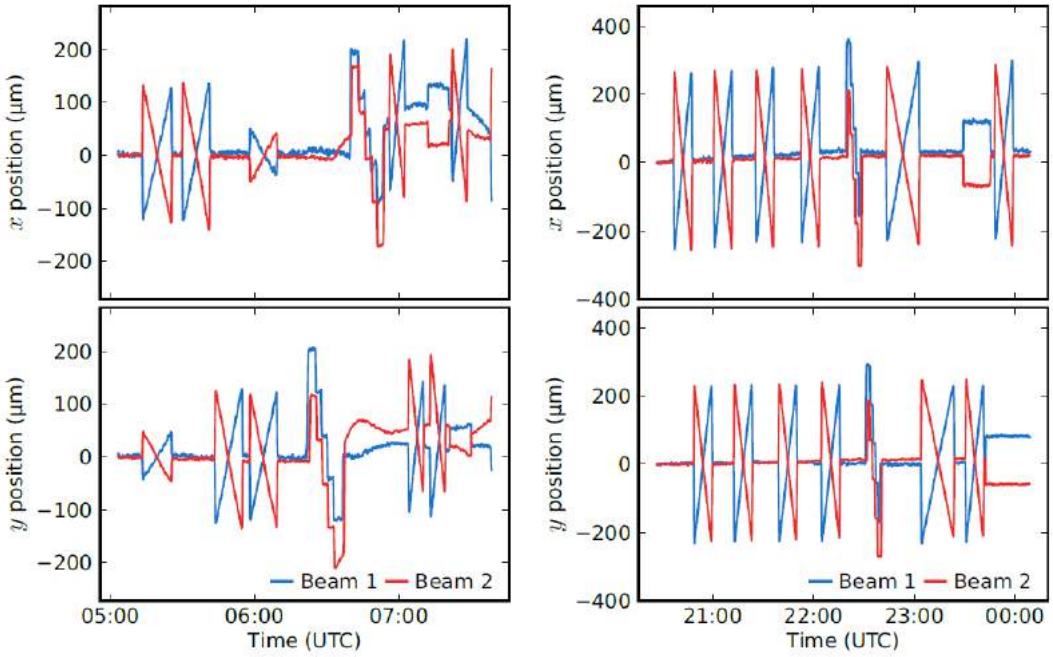


Figure 2.2: The beam positions around the LHCb interaction point recorded by the beam position monitors during VDM sessions. Values are set to zero at the beginning. The top and bottom images show horizontal and vertical positions, respectively. In the scans the beams are moved symmetrically.

interaction rate, R, depends on the orthogonal transverse separation Δx and Δy between the beams. From the observation of the interaction rate and the knowledge of the LHC revolution frequency, it is possible to calculate the average number of visible interactions per bunch crossing. Based on the assumption that the bunch density distributions can be factorized into x and y, it is only necessary to measure μ as functions of the position in the transverse plane. In order to implement this technique, an accurate calibration of the absolute length scale (LSC) is required. There may be corrections introduced at the percentage level during this procedure. A further systematic uncertainty is the non-reproducibility observed between repeated scans. Other corrections to these measurements arise from the beam-beam force, which can cause deflection of the beams when they are displaced. Furthermore, in Run 1, it was demonstrated that the assumption of independent x and y profiles introduces additional systematics.

2.1.3 Beam-Gas Imaging Method

The BGI technique [51] relies on the LHCb experiment's precision vertexing capability, specifically provided by the VELO subdetector. With no beam displacement during the measurement, systematic errors are uncorrelated with those of the VDM approach. This enables a more accurate measurement of luminosity as demonstrated by LHC Run 1. The BGI approach does not need to correct for beam-beam effects or other beam-steering effects. Additionally, it enables LHCb to take data parasitically during other experiments' vdM scans without disrupting their measurements by moving beams. The BGI method relies on precise and accurate position measurements of the reconstructed primary vertex resulting from collisions between the beam particles and the nuclei of the rest gas atoms in the

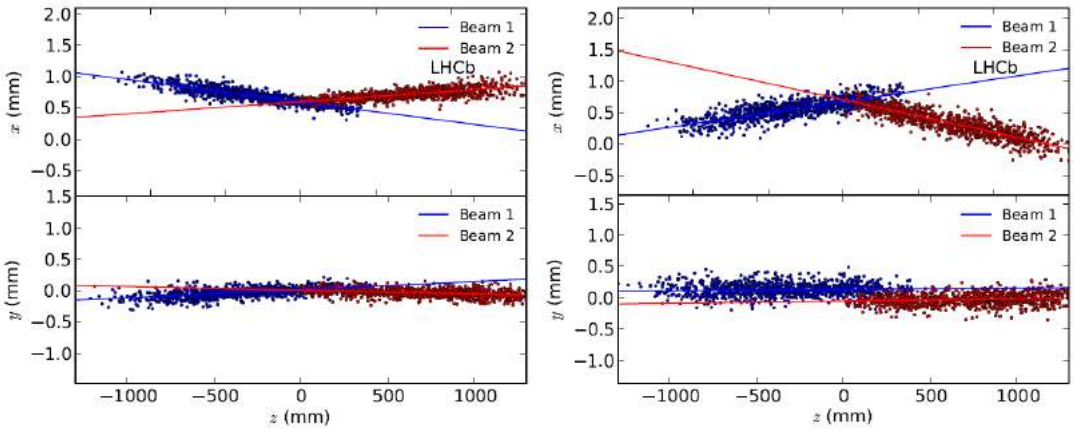


Figure 2.3: Position of beam-gas vertices projected in the xz and yz planes.

beam pipe (Fig. 2.3). In addition, the crossing angle should be precisely determined, as it will easily affect the luminosity produced by colliding beams by up to 20%. Also, a precise fit shape is necessary to describe the profile of charged particle bunches inside the LHC. The simple Gaussian model had been found insufficient to achieve the required precision, and a double Gaussian shape was preferred instead. Transverse beam distributions are also subject to nonlinear correlations, which have been addressed through a factorizable parameter. VdM analysis also addresses these issues, and results from one approach can be used in the other. BGI was used in LHCb run 2 to measure the beam positions during scans as part of vdM luminosity measurements. LHCb has also developed a system of gas injection known as the System for Measuring the Overlap with Gas or SMOG. The BGI technique uses the SMOG system to enhance the number of beam-gas interactions during calibration runs of the detector, where beam profiles can be measured using the interactions of the beam with the injected gas.

2.2 The SMOG System

The SMOG system allows the controlled injection of a limited number of gas species (He, Ne and Ar) to increase the pressure in the interaction region by several orders of magnitude, from $\sim 10-9$ mbar to $\sim 8-7$ mbar. The main components of the SMOG gas injection system are shown in Fig. 2.4. A high-pressure bottle refurbishes the gas and delivers it towards the VELO detector through a 4 m-long tube with $3 \cdot 10^{-5}$ mbar$\cdot l/s$. Four cold-cathode (Penning type) and one hot filament (Bayard-Alpert type) ionization gauges are located around the VELO detector to measure the injected pressure. Extraction takes place at $z = 20$ m from the nominal interaction point, where LHC vacuum pumps are located. The SMOG system was initially developed to increase the number of beam-gas interactions with the detector during luminosity calibration runs but it has also been repurposed for use as part of a rich program of fixed-target physics [52] to produce measurements that are complementary both to the flavour physics program and to other fields covered by pp collisions data, at unique kinematic conditions.

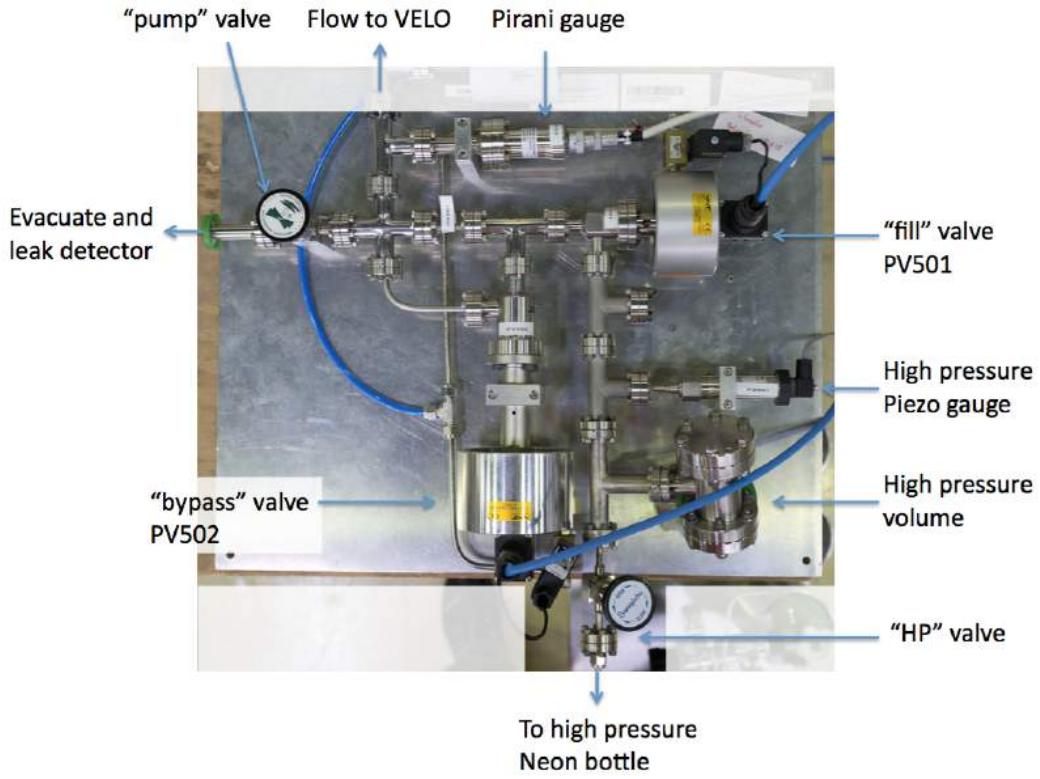


Figure 2.4: The SMOG system employed to inject gas into the beam pipe.

2.2.1 Fixed-Target Data

A pioneering fixed-target physics program began in 2015, with a fully-instrumented detector covering the forward direction, excellent track reconstruction and particle identification capabilities, and the ability to inject noble gases into the vacuum of the LHC accelerator. The LHC Run2 enabled the collection of samples with different center-of-mass energies and collision systems, with both proton and lead beams colliding on helium, argon, and neon gas targets (in Fig. 2.5 a summary of the collected samples). Data samples collected for fixed-target acquisition in 2015 and 2016 were limited in size due to short dedicated periods of data-taking. The SMOG device was injected continuously and simultaneously with a special pp run with $\sqrt{s} = 5$ TeV at the end of 2017. The pNe events were collected using a beam-empty collision type, with the beam approaching the detector from the VELO side. The resulting pNe recorded number of protons on target is about two orders of magnitude larger than the other fixed-target samples. However, the data-taking conditions implied a non-negligible contamination between pNe fixed-target events and pp events. Lastly, PbNe samples were acquired in 2018 using the same center-of-mass energy per nucleon as pNe samples, giving the possibility to look for nuclear matter effects. For all samples, the number of protons on target (POT) is used as a reference since no precise gauges are available to measure the injected gas pressure and no direct measurements of luminosity could be made. The target particle areal density can be defined as a function of the temperature and pressure according to the perfect gas law. For the nominal SMOG gas pressure

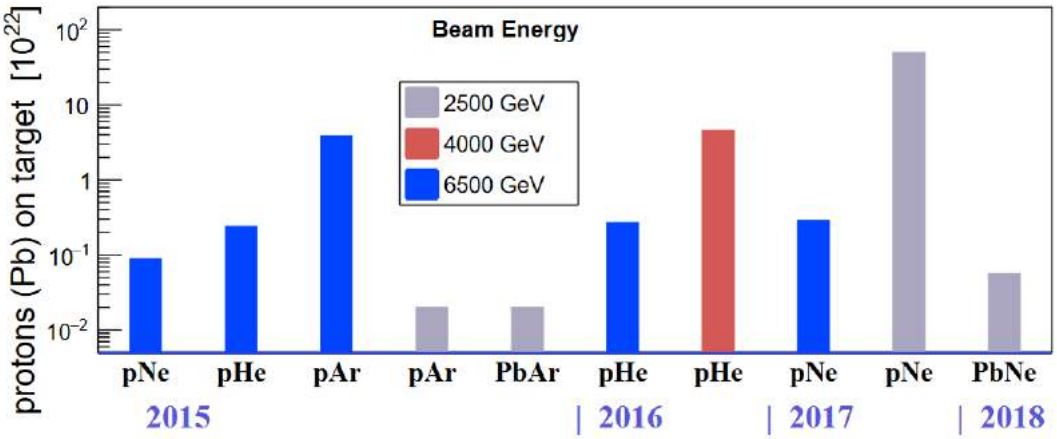


Figure 2.5: The data size is expressed as delivered protons (ions) on target (POT). For a nominal SMOG pressure of $2 \times 10-7$ mbar, 10^{22} POT correspond to an integrated luminosity of about 5 nb^{-1} per meter of gas. However, pressure and efficiency vary in different data-taking.

of $2 \cdot 10^{-7}$ mbar the integrated luminosity is expressed as:

$$\int \mathscr{L} dt \sim 5\text{nb}^{-1} \times \frac{pot}{10^{22}} \times \frac{p_{\text{gas}}}{2 \cdot 10^{-7}\text{mbar}} \times \frac{\Delta z}{1\text{ m}} \tag{2.2}$$

2.2.2 Physics Opportunities

Fixed-target configurations result in nucleon-nucleon collisions with center-of-mass energies between 41 and 115 GeV when proton or ion beams with energy per nucleon E_N between 0.9 and 7 TeV are delivered by LHC:

$$\sqrt{s_{NN}} = \sqrt{2E_N M_N c^2} \in [29, 115]\text{GeV}\,. \tag{2.3}$$

The rapidity of the center of mass with respect to the laboratory frame is

$$y_{CM} \sim \text{arcsinh}\left(\sqrt{E_N/2M_N c^2}\right) \in [3.8, 4.8]\,, \tag{2.4}$$

indicating that the detector covers particles produced at both backward and central rapidities in the center-of-mass frame. Thus, observing particles with large fractions of the target nucleon's longitudinal momentum in c.m. frame, usually expressed as Feynman-x variables, is possible:

$$x_F = \frac{p_L^*}{|max(p_L^*)|} \simeq \frac{2}{\sqrt{s_{NN}}} \sqrt{\left(Mc^2\right)^2 + \left(p_{\text{T}} c\right)^2} \cdot \sinh\left(y^*\right) \simeq x_1 - x_2, \tag{2.5}$$

where M is the particle mass, p_T its transverse momentum, p_L^* and y^* the longitudinal momentum and rapidity in the c.m. frame. Here, x_1 and x_2 indicate the Bjorken-x values of the beam proton and the target nucleon, respectively. A large negative x_F value corresponds to a large Bjorken-x value in the target nucleon, which can only be reached in beam-beam collisions at a much large Q^2 (Fig.2.6). Various results have been produced using fixed target data collected in Run2 [53], [54],[55], [56],[57].

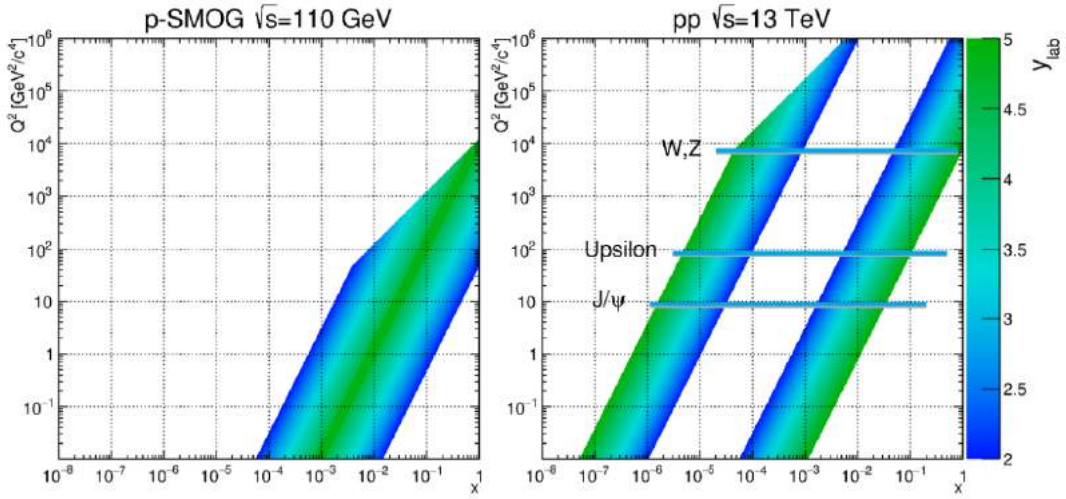


Figure 2.6: Comparison between the kinematic region accessible to the LHCb experiment in the Q^2-target Bjorken-x plane for fixed-target (left) and collider (right) mode with a 6.5 TeV proton beam energy. The region at high-x and low-Q^2 is accessible at a lower energy scale using fixed target collisions.

Chapter 3

The LHCb RICH Detector System Upgrade

This chapter describes the LHCb RICH Detector System Upgrade. The work on the RICH is a main part of my PhD activities that cover different topics: from the Elementary Cell quality assurance, in which the LHCb Ferrara group had a major role in developing the test station and protocol, to the commissioning of the RICH detectors in the LHCb control room, the place where I spent most of my working time in my last PhD year. In Sec.3.1 the upgrade design and the motivation which led to a new RICH detector system are presented and, in the following chapters, every aspect of the upgrade is covered in greater detail, from the photomultiplier to the customized front-end electronics. In Sec.3.7-3.8 a first glimpse at the new RICH detectors is shown, being the last part of the commissioning before the installation at P8. In Sec.3.11 the main measurements performed to calibrate the electronics and achieve a correct configuration of the front end are illustrated.

3.1 The RICH Upgrade for Run3

At the end of 2018, the second long shutdown started at LHC. An LHCb upgrade to achieve a luminosity of $2 \times 10^{33}\ cm^2\ s^{-1}$ and a readout data rate of 40 MHz (without L0 trigger) has been performed. All sub-detectors have been upgraded to satisfy the new operating conditions.
During Run 1 and 2 (before LS2) the read-out rate was limited to 1 MHz after the L0 trigger, the new working condition at 40 MHz impose to modify front-end electronics of all sub-detectors. The interesting event selection will be performed by a fully software-based trigger on a PC farm to enhance event selection efficiency.
The overall structure of the RICH detector will remain unchanged during the upgrade with few modifications in RICH 1 and the adoption of new photon detectors, switching from Hybrid Photo Detectors (HPDs) to Multi-Anode PhotoMultiplier Tubes (MaPMTs) [58]. The MaPMT is coupled with the readout electronics in a fundamental unit called Elementary Cell (EC) described in Sec. 3.3.

3.1.1 RICH 1

RICH1 is the Cherenkov detector closer to the interaction point and it covers the full LHCb angular acceptance (300 $mrad \times 250\ mrad$) and it is filled with C_4F_{10} gas, the gas vessel has a volume of $\sim 2 \times 3 \times 1\ m^3$. The light produced by particles passing through the radiator is focused by spherical

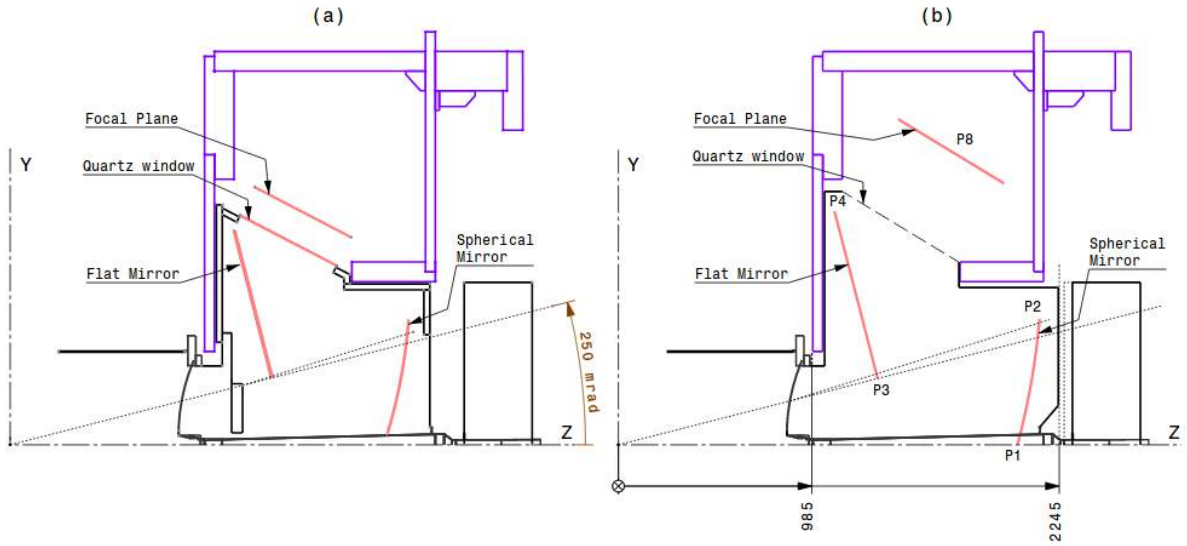


Figure 3.1: RICH 1 optical geometries before (a) and after (b) the upgrade

mirrors (made of a thin light-weight carbon fiber composite) and reflected by plane mirrors (made of SIMAX glass) in the directions of the B-shielded photo-detectors.
The occupancy of 30% on RICH 1 is the upper limit given by simulations and experience, values above 30% will degrade the PID performance. With the current RICH 1 configuration and the expected luminosity of $\mathscr{L} \sim 2\times10^{33} cm^{-2}s^{-1}$, occupancy will increase up to 50%. In order to reduce this value in the Run3 environment, some changes in the optical layout of RICH 1 have taken place. In particular, the position of the focal plane has been adjusted and the radius of curvature of the spherical mirrors increased to obtain a magnification of the Cherenkov rings in the plane of the detectors (and hence decrease occupancy), as shown in Fig.3.1.

In RICH 1 all the HPDs detectors are replaced by Hamamatsu R13742 (1×1 inches, R type) MaPMTs each with a 64-pixel matrix. The upgraded detectors are conceived to be modular systems to facilitate the maintenance and independence of single components. The fundamental element (for both RICH 1 and RICH 2) is the Elementary Cell, the EC comprehends MaPMTs and Front End electronics.

3.1.2 RICH 2

RICH 2 is placed downstream of the spectrometer magnet, filled with CF_4, and covers an angular acceptance of 120×100 mrad with dimensions of $7\times7\times2\ m^3$. Cherenkov light is directed with mirrors into the detector plane as in RICH 1. In the case of RICH 2 Hamamatsu R13742 (1×1 inches) MaPMTs for high occupancy region and Hamamatsu R13743 (2×2 inches, H type) MaPMTs for low occupancy region will be installed.

Table 3.1: Parameters of radiator materials in RICH

Parameter	C_4F_{10}	CF_4
$Length\ [cm]$	110	167
n	1.0014	1.0005
θ_c^{max}	53	32
$\pi\ p_{th}\ [GeV]$	2.6	4.4
$K\ p_{th}\ [GeV]$	9.3	15.6
$p\ p_{th}\ [GeV]$	17.7	29.7

3.2 Particle Identification in RICH

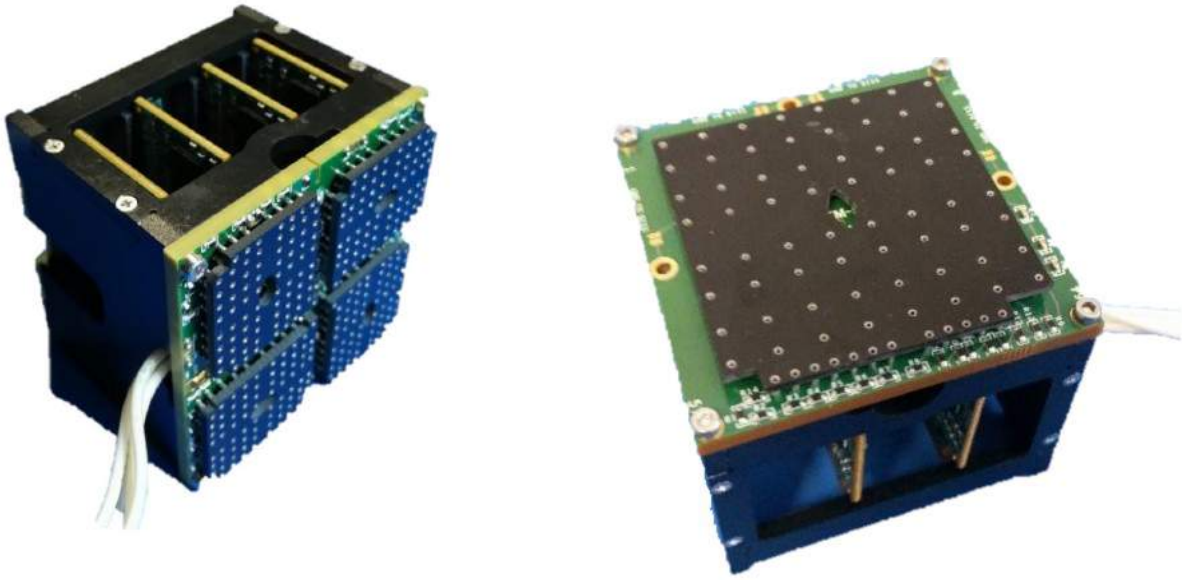
RICH must address the Particle Identification (PID) of particles in a common momentum range, especially in the case of pions and kaons. To achieve the identification, particle direction and momentum are known by the tracking system. In the usual condition of high occupancy, Cherenkov rings cannot be directly identified so for each mass hypothesis ($\pi^{\pm}$, $e^{\pm}$, $\mu^{\pm}$, $K^{\pm}$) and a known track a ring is constructed, the likelihood of each test ring is evaluated against the photon hits on the detectors. With the constructed likelihood and the information from calorimeters and muon system for each track, PID criteria are defined and used in the analysis.
The performance of the RICH detector is measured by looking at known decays that can be reconstructed using kinematic requirements like $K_s^0 \to \pi^+\pi^-$, $\Lambda \to p\pi^-$ and $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$. From these decays, the identification efficiencies and misidentification rates can be calculated as a function of the track momentum.
RICH 1 covers the range $1.5 - 50GeV$, while RICH 2 allows a $\pi - K$ separation up to $100GeV$ (radiator characteristics are reported in Tab. 3.1).
The Cherenkov angle reconstructed from the light emitted by particles has a resolution expressed in mrad and the main contributions are due to:

- $Chromatic\ effect$ caused by the variation of the refractive index of the radiator as a function of the photon wavelength
- $Emission\ point$, angle appear with a spread caused by mirror aberration and physical constraints on the mirrors forcing them to be a bit offset respect to the best focal plane. Also, the angle is smeared under the assumption that all photons are emitted in the midpoint of the track
- $Pixels$ connected to the granularity of the detectors.
- $Track$, an intrinsic error associated with the track of the particle

In Tab.3.2 are reported the resolutions associated with their sources for HPDs and MaPMTs, an improvement is expected in the chromatic error (because MaPMTs quantum efficiency peaks at higher wavelength respect to HPDs), in the pixel resolution (no Point Spread Function contribution) and in the emission point thanks to the improved optics.

Table 3.2: Resolution of the Cherenkov expressed in mrad ring with each contibution.

Source	**HPD - RICH 1**	**MaPMT - RICH 1**	**HPD - RICH 2**	**MaPMT - RICH 2**
Chromatic	0.84	0.58	0.48	0.31
Pixel	0.60	0.44	0.19	0.19
Emission point	0.76	0.37	0.27	0.27
Total	1.70	0.78	0.65	0.45



(a) (b)

Figure 3.2: Elementary Cell of type R (A) and H (B)

3.3 The Elementary Cell

The fundamental modules in which the MaPMTs are grouped and coupled to the electronics is called Elementary Cell (EC) (3.2), the ECs will be mounted on the photosensitive planes of the RICH detectors.

Each EC is composed by :

- BaseBoard: it provides mechanical support for the MaPMTs and also the voltage divider bias (Fig. 3.3a). The BaseBoard has also a low thermal impedance path to dissipate the heat toward the aluminum case of the EC. The BaseBoard R hosts 4 MaPMT of the R-type, and the BaseBoard H hosts one MaPMT of the H-type. Signals from PMTs are directed through the BaseBoard to the CLARO, an integrated circuit ASIC CMOS chip described in Section 3.4.
- Front End Boards (FEBs): 8 CLARO chips are placed on a single FEB. EC-R and EC-H have four and two FEBs respectively. Due to the low power consumption of the chips, no cooling system is needed for a FEB. In Fig. 3.3b a FEB is inserted in the BaseBoard.

(a) (b) (c)

Figure 3.3: Elementary Cell R components.

- BackBoard (Fig. 3.3c): it's needed to connect FEBs and digital boards
  The digital board menage digital signal processing and data transmission.

An exploded view of an EC-R is shown in Fig. 3.4.

3.4 CLARO chip

The CLARO chip [59] is an 8-channel custom integrated circuit ASIC CMOS chip (Fig. 3.5) designed for the readout of multi-anode photomultiplier and build to satisfy the MaPMTs requirements with new LHCb luminosity of 2×10^{33} cm^{-2}s^{-1}. CLARO has low power consumption ($\sim$ 1 mW channel), high radiation resistance, and it can perform photon counting at 40 MHz.
Each CLARO channel consists of an analog charge-sensitive amplifier and a discriminator (Fig. 3.6). A current pulse is converted into a voltage signal and compared with a threshold, the channel returns a binary output related to the signal exceeding the threshold. The CLARO is programmable with 64 threshold levels, 4 attenuation levels, and a single offset of approximately 32 threshold levels that can be applied. All these features allow performing measurements on the input signal and studying one photon peak, noise effects, and set for each channel threshold to maximize the signal-over-noise ratio [60]. The possibility to configure the settings of each channel independently is demanded due to the high variability in the gain of individual pixels in a MaPMT: typically the response can differ by a factor of three and can be compensated by setting attenuation and threshold of the CLARO channels.

MaPMTs generate a charged signal of 10^6 e$^-$ from single photon conversion which is fed to the CLARO input, each channel of the CLARO takes as input the signal of a single anode of MaPMT, in general, the gain is expected to vary up to a factor 3 in the same MaPMT with a fixed voltage for all anodes, so the CLARO is designed to process signals ranging from 30 ke^- up to 20 Me$^-$.
The expected maximum occupancy for RICH detectors from simulations is 25 % or an average of 10^7 hits/s at 40 MHz. In order to minimize inefficiency and pile-up, CLARO recovers the baseline after a signal in 25 ns.

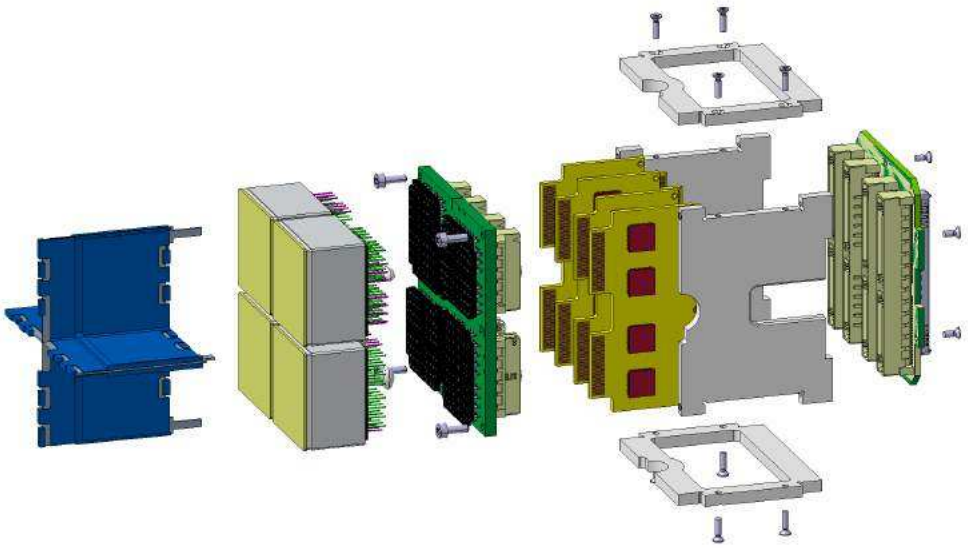


Figure 3.4: EC-R components. From left to right: magnetic shielding, MaPMTs, BaseBoard, FEBs, aluminum case, BackBoard

Figure 3.5: The CLARO8v3 chip

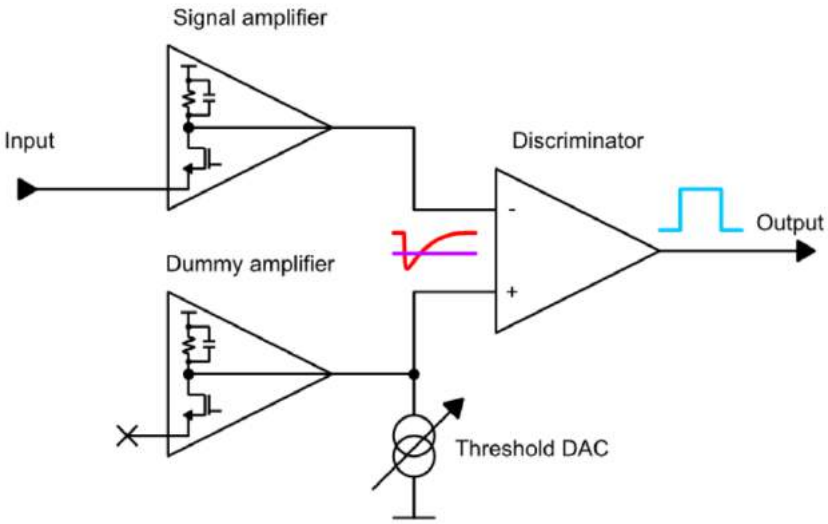


Figure 3.6: Schematic diagram of a CLARO channel. The input is given by the charge collected at the anode, this is compared with a set threshold resulting in a binary output

3.4.1 Configuration Register

CLARO configuration is stored in a 128-bit register (Tab. 3.3) and the writable bits are protected against Single Event Upset (SEU) caused by radiation thanks to the CLARO Triple Modular Redundancy (TMR) which can correct single bit flips events and increase the SEU counter every time a SEU condition is corrected.

Channel configuration bits

The bit assignment for settings the parameters of a CLARO channel are reported in Tab. 3.4.
The specific features of each bit are:

- offset bit: when it's high all the thresholds are shifted by 32 steps, in practice, this feature allows to set negative threshold values. This implies that, nominally, the threshold obtained with offset bit 1 and threshold 32 corresponds to the one given by offset 0 and threshold 0.
- test pulse: it can enable or disable the test pulse injection from the test pulse injection circuit of the CLARO
- hysteresis: enable the hysteresis, when high an output of one threshold is added at the signal amplifier output to prevent oscillation in channels with a threshold set too close to the baseline (as long as the peak-to-peak noise of the baseline is smaller than one threshold step)
- enable bit: when is low the channel doesn't read signals
- attenuation bits : the two bits allow to set four different attenuation values (00 = gain 1, 0.1 = gain 1/2, 10 = gain 1/4 and 11 = gain 1/8)
- threshold bits: a threshold in the range 0-63 can be set in binary value with these 6 bits

Table 3.3: The Configuration Register bit assignment for a CLARO Chip

Bit Position	Description	Read Only / Read-Write	Width
127:120	status of latches	RO	8
119:116	reserved field	RO	4
115:104	SEU counts	RO	12
103:96	configuration bits for common features	RW	8
95:84	channel 7 configuration bits	RW	12
83:72	channel 6 configuration bits	RW	12
71:60	channel 5 configuration bits	RW	12
59:48	channel 4 configuration bits	RW	12
47:36	channel 3 configuration bits	RW	12
35:24	channel 2 configuration bits	RW	12
23:12	channel 1 configuration bits	RW	12
11:0	channel 0 configuration bits	RW	12

Table 3.4: The configuration bits for a CLARO channel

Bit Position	Description	Read Only / Read-Write	Width
11	offset bit	RW	1
10	test pulse (1 disabled)	RW	1
9	hysteresys for the comparator	RW	1
8	channel enable bit	RW	1
7:6	attenuation setting	RW	2
0:5	threshold setting	RW	6

3.5 MaPMT

Multi Anode PhotoMultiplier Tubes are going to replace HPDs in the RICH detectors. The chosen MaPMT models are produced by Hamamatsu and are

- 1” R13742 (R-type), custom modification of the MaPMTs R11625
- 2” R13743 (H-type), custom modification of the MaPMTs R12699

They both are made by an 8×8 pixel matrix (Fig. 3.7), the area of a pixel is $\sim 3 \times 3 mm^2$ for R-type and $\sim 6 \times 6 mm^2$ for H-type, and they’re geometry allow for a close packing ratio.

MaPMTs have a super-bialkali photocathode and a UV glass window. The main advantages of MaPMTs are

- the low noise, so the ability to detect single photon signals in the UV region (where Cherenkov radiation is most intense)
- a fast time response
- a large active area coupled with a pixel granularity

In particular, MaPMTs with respect to HPDs have a higher Quantum Efficiency in the green wavelength region, this feature provides a reduction of the chromatic error due to the lower value of the slope in the refractive index n as a function of the wavelength (as can be seen in Fig. 3.8).

Figure 3.7: The two MaPMT models: R13742 (R-type) will be mounted in RICH 1 and RICH 2 (left), R13743 (H-type) will be mounted in RICH 2 only (right)

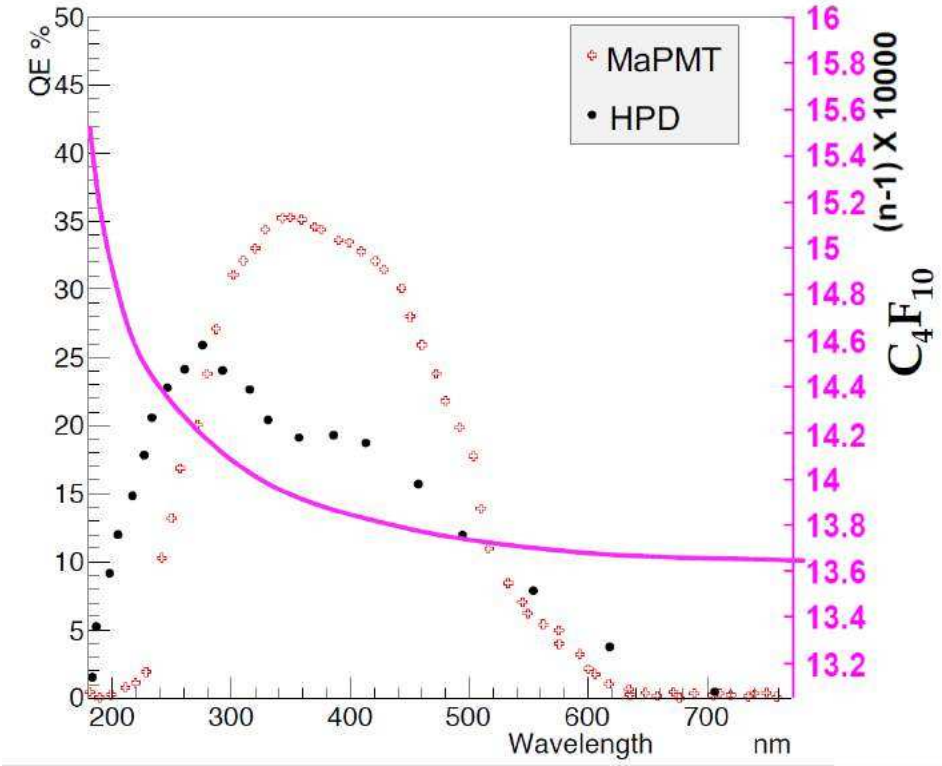


Figure 3.8: QE of MaPMT and HPD comparison and variation of the refractive index (purple) in RICH 1

When a photon hits the bialkali photocathode it can produce a photo-electron with a certain probability given by the Quantum Efficiency. The photo-electron is accelerated by an electric field and focused on the first dynode. When an e^- hits a dynode, an electron multiplication takes place due to secondary emission, all the electrons will be focused to the subsequent dynode (12 in total for R-type, 10 for H-type) up to the anode where the total charge is collected.
The characterization and quality assurance of the MaPMTs from Hamamatsu has been constantly performed during the commissioning process [61]-[62]. In the MaPMTs every pixel has its own dynodes chain, this will result in a spread of the gain of the pixel even for a single MaPMT. The gain in a single MaPMT can vary between single pixels up to a factor of 3. The typical bias voltage applied is 1 kV, the negative HV is applied to the photocathode and the ground to the anode. Every pixel of the MaPMT will operate in a single photon regime and the typical gain is $\sim 2\text{Me}^-$ at 1000 V.
The dark count rate at room temperature is expected to be $\sim$ 10 Hz per pixel at room temperature. The increment of the temperature will result in higher counts since electrons will have more thermal energy and it increases the probability to escape from dynodes or photocathode surfaces generating an avalanche with the consequent signal. The rate per pixel has to be lower than 1 kH, otherwise, the MaPMT would have been rejected during the Photo Detector Quality Assurance.

3.5.1 Magnetic Field Dependence

MaPMTs have been tested with high magnetic fields to study the dependence of efficiency on B. The PMT behavior is affected by a magnetic field because it can change the electron trajectory between anodes, inducing a degradation of efficiency and gain. In RICH detectors the B field can reach 25 G in RICH 1 and $\sim$ 5 G in RICH 2.
In order to assure the correct response by the pixels of the MaPMTs a cross-shaped local shielding has been developed, this shielding structure is necessary for RICH 1 only and it allows to have a full performance recovery for central pixels of MaPMT and an efficiency $>$ 85% for edge pixels (these are strongly affected being near the biasing pins).

3.5.2 Ageing

The PMT ageing is a process that deteriorates PMT performances after long periods of light exposure. This condition is particularly important in high energy physics due to the high photon rates. The performance deterioration is caused by a reduction of the photocathode efficiency, an increase in dark count rates, and a variation in the gain of the photomultiplier.
The most critical effect is the gain variation, this is caused by the deterioration of the dynodes resulting in a lower number of electrons produced by secondary emission. Operating in a single photon regime, the gain variation can be estimated by the change of the single photon peak position during several hours of illumination.
The gain variation in the MaPMTs expressed as a percentage of the single photon peak can be $\sim$ 10% during an acquisition time of 3000 hours, equivalent to two years of LHCb RICH operation (effective operation time over 2011 and 2012 amounted, on average, to 1500 hours). This effect can be compensated by adjusting the high voltage.

3.5.3 Quality Assurance

An extensive quality assurance campaign has been carried out to validate, ensure minimum specifications and characterize the full photo-detection chain. The major campaigns, before the commissioning of the column at CERN and the installation of the RICH detectors at P8, have been:

- Photodetector QA (PDQA) in Padova and Edinburgh
- CLARO chips and control system of Front-End Boards (FEB) and Back Board (BkB) in Ferrara and of Base Boards (BBs) in Genova and by Studioemme
- Elementary Cell QA (ECQA) in Ferrara and Edinburgh

The Ferrara LHCb Group has been responsible for the quality assurance procedure of the CLARO, of the FEBs production, and, together with the Edinburgh group, of the ECs quality assurance (ECQA). The ECQA procedure has validated the reliability and conformity of the components assembled to constitute an EC. All the ECs needed to pass an established test protocol. Four testing stations have been deployed between Ferrara and Edinburgh. Each station consists of a light-tight box that hosts the ECs during testing. A DAQ control software, developed in LabView, allows to control of the HV supply and a LED driver, to configure the electronics (Digital Boards and CLARO chips), and to run an automated test on the ECs. The validation of an EC consists of measurements that verify the conformity of FEB and CLARO, and also monitor dark count rates and afterpulses of the MaPMTs installed on the ECs. In total, more than 1200 EC have been characterized by the ECQA setups and have been shipped to CERN before the beginning of 2020. Now the ECs are mounted in the RICH photodetector planes. The data acquired during the ECQA have been organized in a database and are usually used as a reference in the commissioning stages.

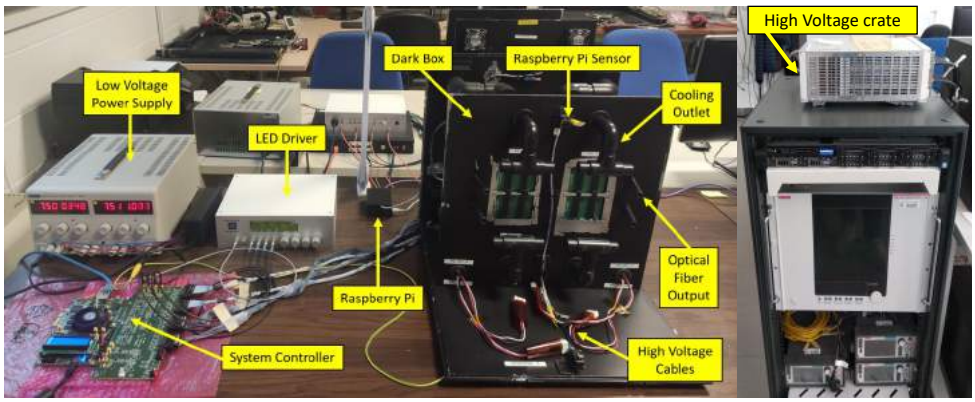


Figure 3.9: The ECQA test setup present in Ferrara.

3.6 Photon detector module digital boards

In the FEBs and PDMDBs, the signals coming from the MaPMTs channels are processed. By transporting the digitized signals away from the high radiation area of the detector, the PDMDB ensures that photon detection efficiency is preserved without introducing dead time. In addition, the layout

of the photon detection plane, the confined space, the magnetic field, the radiation environment, and compatibility with the common data acquisition infrastructure for all LHCb sub-detectors all influence the design of the board. FPGAs are used for a variety of reasons, including the flexibility they provide in formatting data and interfacing with different electrical signaling standards on the front-end ASICs and giga-bit optical links (GBTs). PDMDBs are used in two variations according to the granularity of the photon detector planes (Fig. 3.10). Two PDMDB-Rs are arranged back-to-back, each connected to a group of four EC-Rs, whereas one PMDB-H is combined with four EC-Hs. The module consisting of four ECs and one or two PDMDBs is called Photon Detector Module (PDM). For

Figure 3.10: Picture of (Left) PDMDB-R and (Right) PDMDB-H.

fast and slow control data exchange between the PDMDB and the LHCb global online system, each PDMDB has a TCM (Trigger and Control Module). In the TCM, there is a VTRX (optical transceiver) and a master GBTX (gigabit optical link) chip, which operates in bidirectional forward error correction mode, and through the GBT links, the LHC clock can be distributed synchronously to the FE electronics. In addition, it implements the different protocols required for configuring the PDM by utilizing the GBT Slow Control Adapter system (GBT-SCA) connected to the GBTX sc e-port running at 80 MHz. DTMs are plug-in modules that provide high-speed transmission interfaces for PDMs receiving data from FPGAs. The number of FPGAs and DTMs on each board is the same, three on PDMDB-R and two on PDMDB-H. Optical links consist of VTTX optical transmitters, each connected to a GBTX ASIC for transmission. In one GBTX, the 40 MHz reference clock comes from the TCM. It generates a 160 MHz clock that is routed to the FPGA. In the second GBTX, the 40 MHz clock is routed locally and serves as the reference clock. The data transmission e-links in DTM GBTX that are connected to the

FPGA operate at 160 MHz. ECs, TCMs, and DTMs communicate via the PDMDB motherboard. The board also incorporates CERN FeastMP-CLP DCDC converters for the local power regulation of the FPGAs and the active components on the ECs, TCM, and DTMs. The only active components on the motherboard, apart from the DCDC converters, are the FPGAs that receive the 2.5 V LVCMOS digital GBT frame and the packing of the bits into the frame is performed in the FPGA logic.

As PDMDBs are large heat sources, they are cooled by aluminum plates which provide thermal coupling between the components and the internal structure of the column. FPGAs, GBTs, and VTTXs are equipped with thermal pads to ensure not only thermal isolation but also the required thickness between the PDMDB and the aluminium plate.

3.7 Photon detector columns

There are two types of PDMs: PDMs-R and PDMs-H, depending on their position in the RICH. The first consists of four ECs-R and two back-to-back PDMDBs-R and it's installed in RICH1 and in the central region of RICH2. A PDMDB-H along with ECs-H are installed in the peripheral parts of RICH2. Six PDMs are assembled on a T-shaped aluminum structural element, referred as T-bar. The length is about 1.6 m, the width 55 mm and the depth 40 cm. In fig. 3.11 the CAD view of the T-bar and a fully populated RICH2 column is shown.

The RICH1 columns contain 22 ECs instead of the 24 originally planned, as a convenience for the installation, handling, and maintenance of the columns, and due to the low hit rate expected in the corresponding regions. In the same way, the column placed upstream of each MaPMT plane, which detects Cherenkov photons produced by particles with the lowest pseudorapidity, has 20 ECs. The RICH2 columns, which have two PDM- R and four PDM-H, are completely populated. Cooling is provided by circulating 3M Novec 649 into ducts embedded in the T-bar spine to ensure that the MaPMTs stay below 30°C. As a result of the geometry and the flow of cooling, the temperature distribution is uniform.

Each column requires services such as Low Voltage (LV), High Voltage (HV), data and TFC (Timing and Fast Control)/ECS (Experimental Control System) optical fibers, as well as monitoring devices for DCS (Detector "slow" Control System) and DSS (Detector Safety System). As a result of the different orientations of columns, two different designs have been produced, one horizontal and one vertical. The PDMDB requires a primary supply of about 7V and a secondary supply of 2.5V to allow each board to be powered via GPIO, and a fully configured PDMDB-R (-H) consumes approximately 2.5A (1.5A) of current. LV supply for RICH1 columns is provided by two Wiener Maraton channels, while LV supply for RICH2 columns is provided by one channel. Dedicated distribution boards, located at one end of the column, provide the 2.5 V supply by means of DCDC regulators. For each column, HV supplies are provided by floating ground A1538DN CAEN boards. Voltages and ground are distributed from one PDM to the four ECs via a dedicated HV distribution board.

Columns are electrically insulated: LV ground is electrically connected to detector safety ground by copper braid to prevent ground loops. The PDMDBs have two TFC/ECS. Moreover, PDMDB-R has six fibers while PDMDB-H has four fibers to transport data to the TELL40 boards and control the PDMDBs. A total of 12 fibers connect the optical links to the PDMDB's individual connectors.

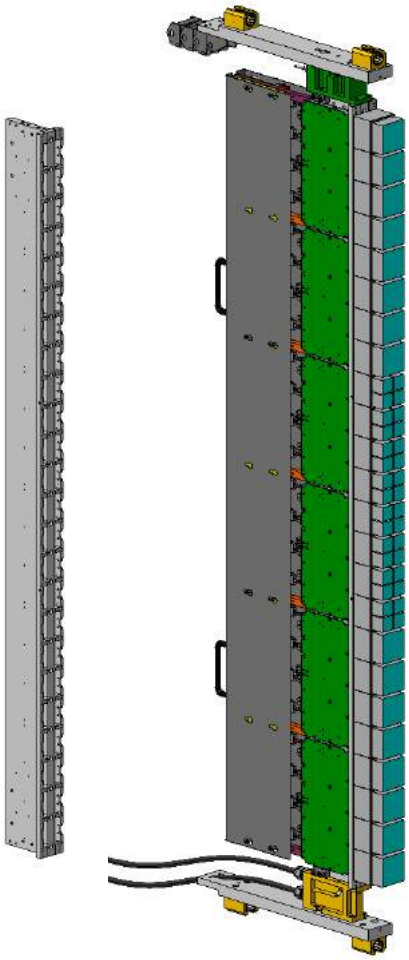


Figure 3.11: (Left) CAD view of the T-bar. (Right): CAD view of the RICH2 column populated with the photon detector modules.

3.8 Photon detectors regions

A RICH1 plane is composed of an array of 11 columns arranged side by side. RICH1 as a whole consists of two planes (22 columns in total), placed above and below the beam pipe. In each column are allocated the optical fibers, the low-voltage distribution the built-in supports for the services to the MaPMTs, and the associated electronics.

Arrays are mounted on rails so that they are easy to remove for maintenance, and are aligned on the MaPMT chassis at the correct angle. Fig. 3.12 shows both the Up and the Down chassis installed in the RICH1.

A RICH2 plane is made of aluminum structures hosting twelve columns arranged side by side, as shown in fig. 3.13.

This type of structure is called a rack, and it is installed into the magnetic shields on the LHCb A- and C-side almost perpendicular to the beam at the LHC. In addition, these racks have rails that enable the columns to move perpendicular to the focal plane. Two cooling manifolds are located under each rack: one to distribute coolant to each column in parallel, and one to collect the fluid. A

Figure 3.12: Support structure for the RICH1 columns, the MaPMT chassis, showing the rails and alignment structures. The chassis is mounted to the soft-iron magnetic shielding that surrounds the MaPMT region.

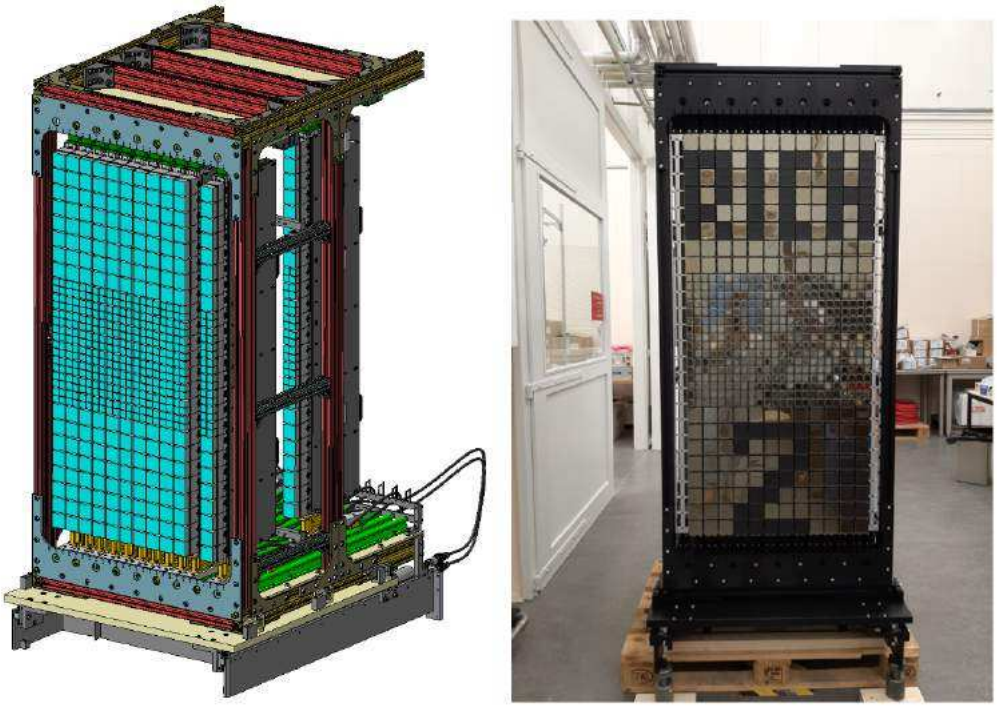


Figure 3.13: CAD view (Left) and photograph (Right) of the fully assembled and commissioned RICH2 A-side photon detectors array.

patch panel is located above the racks and is connected to the optical fibers and electric cables on the upper side of the racks. LV, HV, DSS, TFC/ECS, DCS, DSS, and data transfer are all connected to the patch panel, which is the interface where all services are connected. Nitrogen is permanently flushed into the enclosure to ensure a dry atmosphere with a good dielectric environment, and to minimize the risk of condensation.

3.9 Mirrors

In order to focus Cherenkov light on the photon detector planes, four quadrants of carbon fiber mirrors are arranged around the beamline. As part of the RICH1 upgrade, eight rectangular mirrors of 370 mm×440 mm each have been manufactured and assembled. They have been placed above and below the beamline, outside of the detector acceptance. Each mirror is made of carbon fiber with aluminum and magnesium fluoride coatings optimized for Cherenkov photon incidence angles resulting from upgraded RICH1 optics and for MaPMT quantum efficiency. Mirrors are mounted to rigid aluminum support frames with polycarbonate mounts in their centers.

3.10 Experimental Control System, monitoring and dataflow

The RICH ECS is responsible for setting up, controlling, monitoring, archiving, and operating the RICH subdetectors within the JCOP framework provided by CERN. The RICH ECS is built on top of the WinCC-OA Supervisory Control And Data Acquisition (SCADA) tool, and is divided into three domains: the High Voltage system (HV), the data acquisition system (DAQ) and the DCS. Using a Finite State Machine (FSM), the domains are integrated into a system that takes automatic actions like switching on/off the low voltage and high voltage systems. During operation, the electronics, detector environment, and cooling system are constantly monitored to ensure that the system is safe. A number of temperature and humidity sensors have been placed on the BaseBoards, BackBoards, FEBs, cooling manifolds, and in the photon detector enclosure. Using the SCA chip, located on the PDMDBs, and the Embedded Local Monitoring Board (ELMB), located in the gas enclosure, it is possible to measure the temperature both during operation and when the detectors are switched off. Using a current of 100 μA, the SCA achieves a resolution of about 0.5°C. The sensors and current sources were calibrated in the laboratory and then mapped in DSC WinCC-OA. The temperature sensors can be read out without any operator action through the use of a safe switch-on procedure, where the SCA chip is automatically configured upon power-up. A safe state (OFF) is restored if automatic configuration fails. Finally, a few sensors are connected to the Detector Safety System (DSS), a PLC-based system with many redundancies, which determines the safety of detectors. Also, the DCS records the temperature and pressure of the Cherenkov radiators in the Conditions Database. This is used by the LHCb event reconstruction index to calculate the refractive index of the gas radiator.

Electronics in the Front End (FE) and Back End (BE) require live configuration to work properly and establish predictable dataflows. The control system is segmented into subsystems comprised of a PDMDB, SOL40, and TELL40, and capable of scaling up to thousands of FE and hundreds of BE devices. The SOL40 transmits 40MHz clocks and TFC commands via 48 optical links to the PDMDBs through their TCMs. An SCA ASIC and a GBTx decode all commands, such as writing registers and generating test pulses. Unidirectional optical links are used by the TELL40 to receive FE data from the PDMDB. Through the reading of the Received-Signal-Strength-Indicator (RSSI) and of the Master GBTx Forward-Error-Correction (FEC) counters, TFC links are monitored to detect receiver-side issues. With the TFC commands, communication is enabled, temperature sensors are initialized, firmware is loaded onto the FPGAs, and thresholds are set for CLARO.

3.11 Photon detectors and frontend electronics calibration

The single photon detection efficiency of the RICH system is determined by intrinsic properties of the MaPMT such as the photocathode quantum efficiency, the collection efficiency at the first dynode, and the single photon gain. Detection efficiency is also enhanced by the CLARO signal digitization via a programmable threshold. To minimize inefficiencies caused by threshold setting, monitor the variation of single photon gain with time and aging, and verify the stability of CLARO channels, calibration procedures using DAC scans are implemented.

3.11.1 DAC Scan

If a known charge is injected in the CLARO, an analysis of the CLARO response as a function of the injected charge can be performed to observe the behavior of a single channel of the chip with a set threshold.
The injected signal is generated by applying tension from a 12-bit DAC with a reference voltage of 5 V (the nominal minimum output step is 1.2207 mV) to the internal CLARO charge injection capacitance with a typical value of 640 fF (the process spread of the test capacitor has a sigma, according to the CLARO manual, of 6% and can be trusted within about 3 sigma, so 18%) and a conversion factor of 1 Me^- / 250 mV. Roughly the expected charge injected per DAC step, according to the nominal values, is 4.9 $ke^-/step$.
For each DAC step, a fixed number of signals N are sent to the CLARO input, the transition from 0 counts to N counts is observed in a counts/DAC step plot as an error function. The transition point obtained for each channel of the chip depends on the configured threshold, from the manual a threshold step is expected between 20 ke^- and 45 ke^- with no attenuation.

Fit Function

The observed data are the result of the convolution between the underlying normal distribution due to the noise of the system and a step function.
The data from a DAC scan are fitted using the formula

$$f(x_{tr},\sigma) = \frac{N}{2}[1 + Erf(\frac{x - x_{tr}}{\sqrt{2}\,\sigma})] \tag{3.1}$$

where N is the total number of pulses, x_{tr} the transition point and σ the noise. The noise σ of the transition is related to the standard deviation of the integrated noise distribution. An example of a fitted DAC scan is shown in Fig. 3.14.

For each set of attenuation and offset bits, the threshold DAC code can be converted into the corresponding charge by repeating the DAC scan at different thresholds. Fig. 3.15 illustrates DAC scans performed on a RICH2 column acquiring thousands of events while keeping the HV off and not illuminating the system. There is linearity between the threshold setting and injected charge for all of the attenuation and offset bits considered.

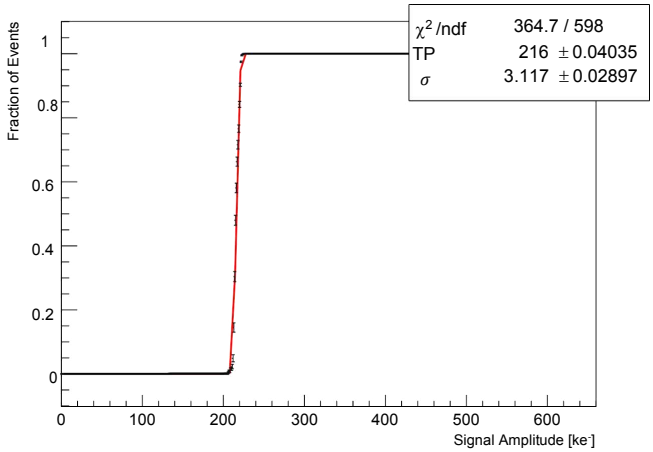


Figure 3.14: S-Curve distribution for a CLARO channel. The parameters of interest are obtained by fitting the curve.

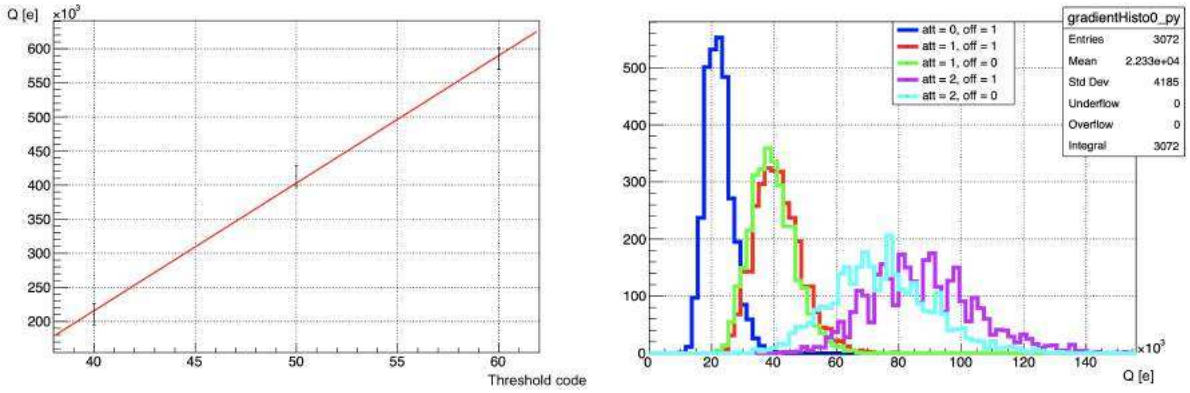


Figure 3.15: (Left) Calibration of a single CLARO channel with offset bit enabled and no attenuation. The charge corresponding to a threshold DAC code (th) is determined by the linear relation $Q = Q_0 + Q_{\text{th}} \cdot \text{th}$. (Right) Distribution of the charges corresponding to one threshold step (Q_{th}) for a RICH2 column in different conditions of offset and attenuation.

3.11.2 Threshold Scan

To maximize single-photon efficiency, threshold scans are performed to find the optimal threshold, attenuation, and offset bits. As opposed to the DAC scan test, this test is performed with the MaPMT powered on and using an external signal source. An approximately constant amplitude CLARO input signal, from the MaPMT operating in a single-photon regime, is fed into the chip when the channel threshold changes with attenuation, offset, or HV. The final distribution, see fig. 3.16 corresponds to the integral of the single photoelectron spectrum, and the parameter of interest is the Working Point (WP) of the CLARO channel, *i.e.* the minimum between the pedestal and the first single photoelectron peak.

To analyze the variation of the MaPMT gain with voltage, threshold scans are performed at 900, 950, and 1050 V. As the voltage is increased, working points should shift to higher threshold values. The optimal operating conditions for identifying the pedestal for every channel were found to be zero

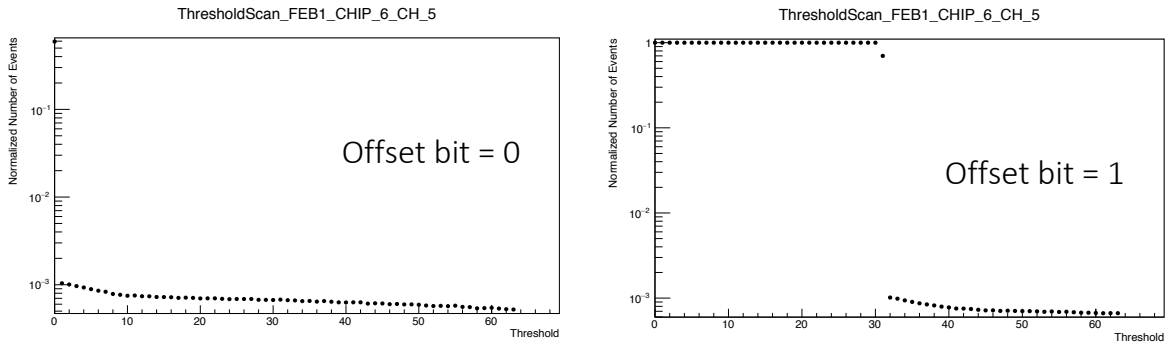


Figure 3.16: Threshold scan on an EC-R channel with offset bit disabled and enabled

attenuation and enabling the offset bit.

In fig. 3.17 is shown the distribution of the threshold settings for the RICH2 channels, converted in absolute charge as determined through DAC scan for the RICH2 channels.

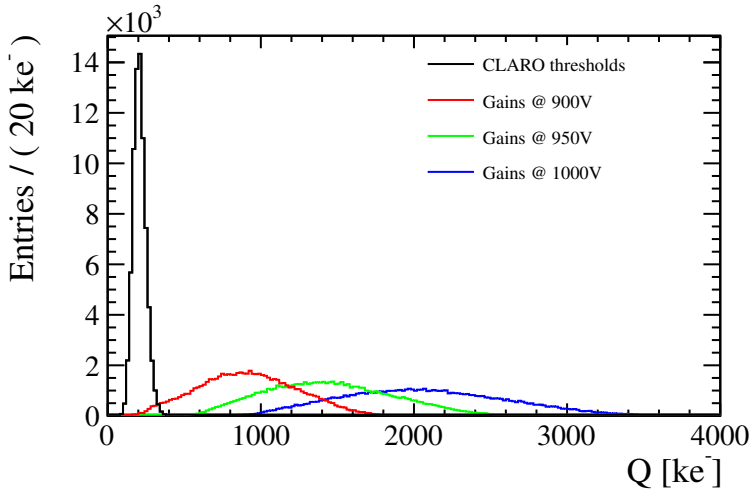


Figure 3.17: Distribution of RICH2 CLARO thresholds converted in absolute charge (black), with a mean of 207.58 ± 0.16 ke^- and a standard deviation of 39.64 ± 0.10 ke^-, reflecting the spread of the pedestals distribution. The threshold settings can be compared to the pixel gains at 900 V (red), 950 V (green), and 1000 V (blue).

As threshold scans provide the integral pulse height spectrum, they are also able to estimate single-photon peaks for each channel, allowing monitoring of gain variations with MaPMT ageing.

3.11.3 Dark Count Rates

In a PMT the dark counts are signals not generated by a photon event but due to a spontaneous electron emission, most likely caused by thermionic emission from the photocathode. Dark count signals are usually single photon events and their frequency depends on the area and the material of the PMT.

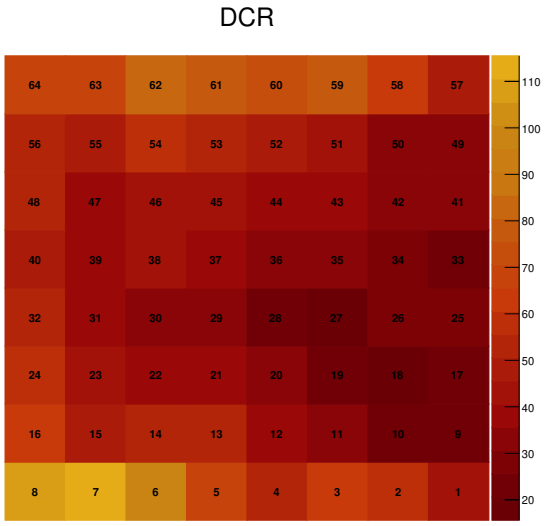


Figure 3.18: DCR rates for a MaPMT of the EC-R at 950 V, every channel has the threshold set to 7 and offset bit 0. The z-axis represents the dark count rates in Hz. The pixel number is written in the pixel center.

The DCR in a time interval Δt is defined as:

$$DCR = \frac{Number\,of\,events}{\Delta t}.$$

DCR measurements allow the identification of noisy anodes and the uniformity of the MaPMT (Fig. 3.18).

3.11.4 Signal Induced Noise

Signal-Induced Noise (SIN) in the MAPMTs was discovered during the quality assurance process [63] and manifests itself as a noise hit arriving typically between a few hundred nanoseconds and a few microseconds after a signal hit in the same pixel (Fig.3.19). In most cases, these processes are caused by electron interactions leading to the ionization of residual gases inside tubes, which generates feedback ions to the photocathode, or by other sources of signal-correlated noise, such as light emissions. SIN intensity information can be obtained by calculating the SIN ratio, defined as

$$SIN_{ratio} = \frac{noise}{signal + noise}$$

or by computing the signal to noise ratio, defined as

$$S/N_{ratio} = \frac{signal}{noise}.$$

The SIN contribution is not uniform across the MaPMT, but it is located mainly in the external pixels, close to the HV biasing pins (Fig.3.20).

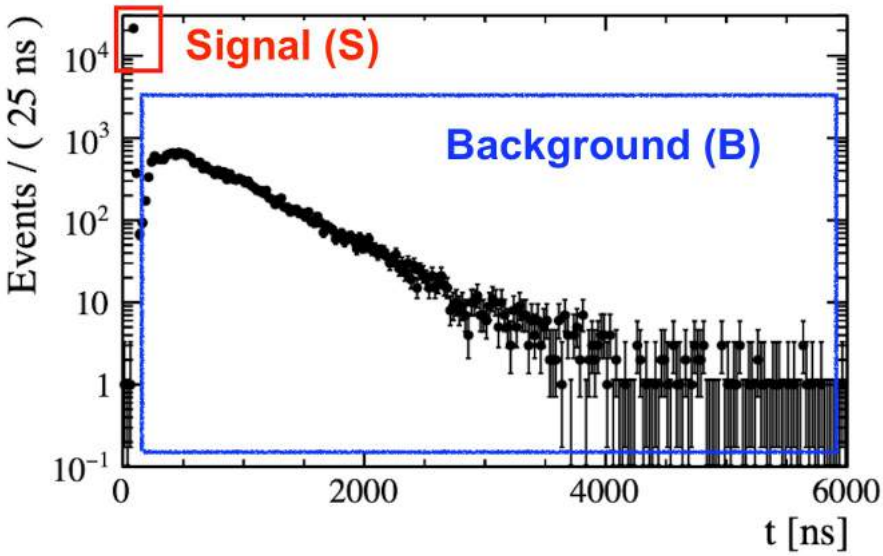


Figure 3.19: SIN spectrum: in red the Signal and in blue is unlighted the background contribution given by the SIN.

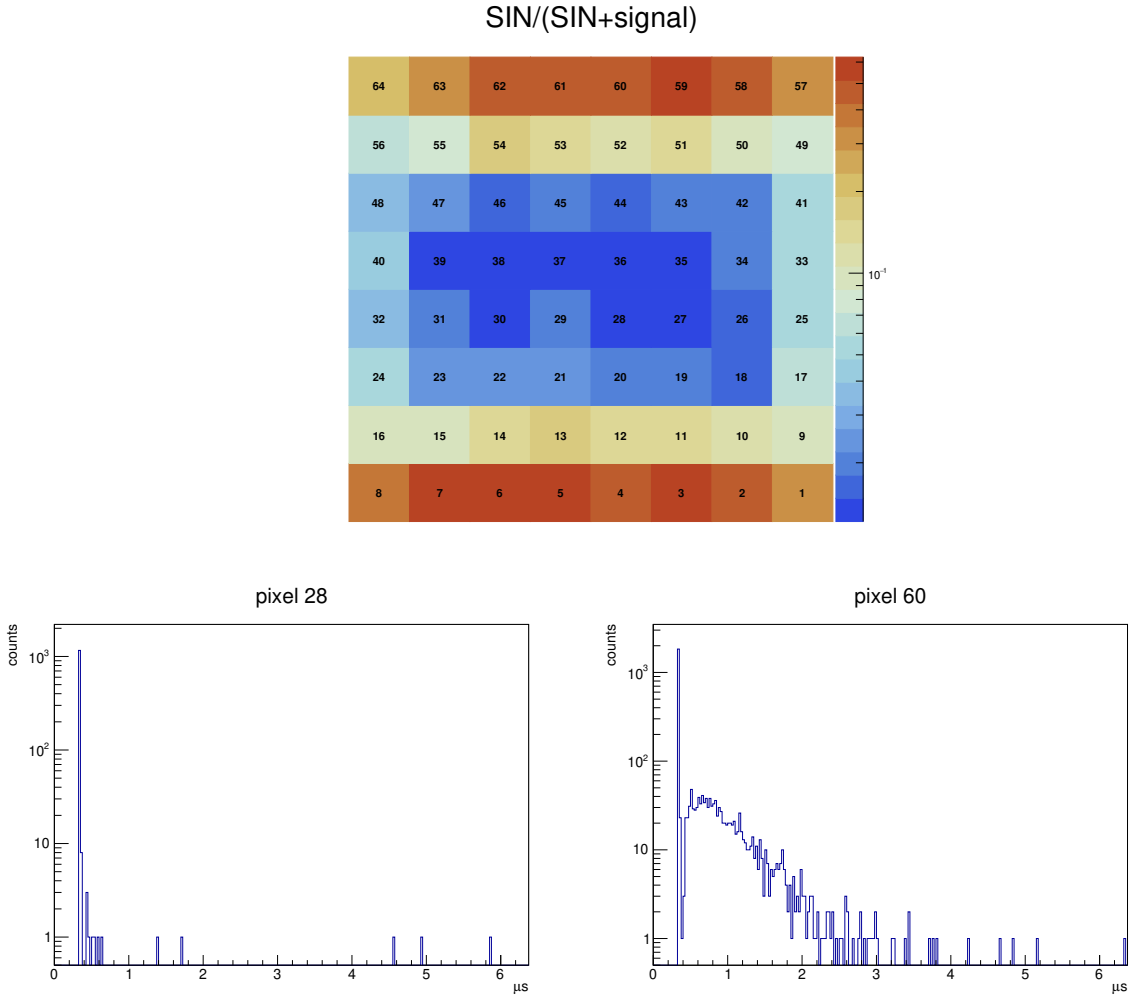


Figure 3.20: Top: fraction of the SIN signal for each pixel of a MaPMT-R. Bottom: SIN distribution for two pixels, one at the center of the MaPMT and low SIN (left) and one exhibiting a high SIN background.

Chapter 4

The LHCb RICH Detector System Commissioning

The upgraded RICH detectors were assembled at CERN during 2020/2021. RICH2 was installed in early 2021 at the pit and commissioned throughout 2021. RICH1 was installed between the end of 2021 and the beginning of 2022. Commissioning of the RICH detector system has been an extensive task performed in consecutive steps, from column assembly in the labs to the full detector characterization with LHC colliding beams. My last year of PhD was devoted to the RICH commissioning in the LHCb control room. In this chapter, the main aspects of the work done to achieve a performing RICH detector system for the Run3 are highlighted starting from the beginnig of 2022. At the end of 2021, the RICH2 was already installed in the cavern and participated successfully in the pilot LHC test beam of October 2022. The RICH1 detector, on the other hand, was fully installed only at the beginning of the next year, when I joined the commissioning activities in the LHCb control room. In Sec.4.1 the Experimental control system is introduced in detail. The principle of operation and the philosophy behind the LHCB control system is discussed with a focus on how to operate the RICH detectors. The Sec.4.3 covers the main activity of the last year of commissioning, the time alignment procedure. Time aligning the detector with the Cherenkov photons from bunch crossing collisions at P8 is a mandatory requirement for efficient data-taking and, in commissioning times, requires dedication for the data-taking, coding, debugging and, finally, validation of the results. The time alignment is composed of two main steps, coarse time alignment and fine time alignment, both are covered in this chapter. Finally, in Sec.4.5 I present my work related to the application of RICH variables for luminosity monitoring. The RICH detectors can provide variables to monitor the luminosity delivered by LHC at the ECS level (the anode currents of the MaPMTs), and at the HLT level. For the latter case, simulation studies have been performed to study the linearity of the hits in the photodetector plane with respect to luminosity. The studies have been validated by real data acquired in a luminosity scan at the end of 2023.

4.1 Front-End Electronics Control and Monitoring for the LHCb Upgrade

To implement the required 40 MHz readout of collision data [64], the LHCb experiment is using a radiation-resistant chipset called Gigabit Transceiver (GBT) for the readout as well as for slow control, monitoring and synchronization. The chipset consists of the GBTx (handling the data readout) and GBT-SCA (configuring and monitoring FE electronics (FEE)). The FEE is controlled by FPGA-based electronics boards using bidirectional optical links interfacing with the control system with PCIe bus, these boards are called PCIe40. The PCIe40 is logically distinguished into three flavors:

- SODIN: readout supervisor board. It distributes the LHC clock, generates synchronous and asynchronous commands to the FE, and transmits data for the event building.
- SOL40: FE interface boards. It is responsible to interface TFC (Timing and Fast Control) and ECS (Experimental Control System) with the Front-End boards. The SOL40 transmits timing and clock sent by SODIN into the optical link to the FEE and, at the same time, allows the ECS to control and monitor the FEE using the same link.
- TELL40: readout board. It is responsible for the readout of the subdetector FE data into the full LHCb DAQ structure.

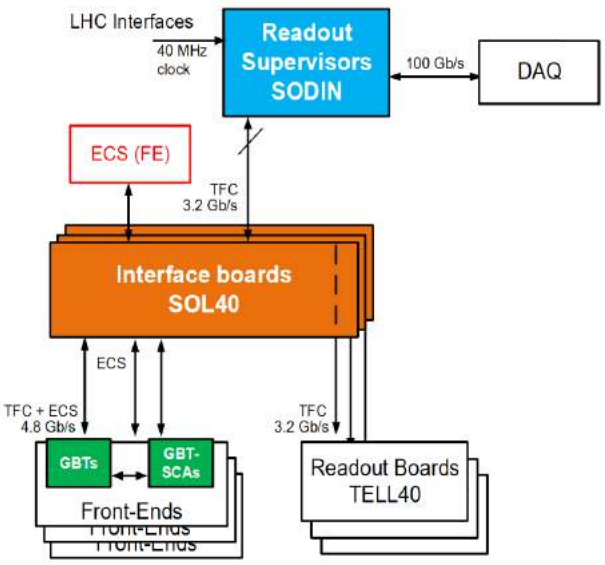


Figure 4.1: The TFC, ECS and Data acquisition architecture of the PCIe40 boards.

The user operates the boards through the Experimental Control System (ECS) which is described in the following section.

4.2 The Experimental Control System

The (ECS) is the framework inside LHCb that controls, monitors, archives, operates and runs LHCb experiments. As part of the JCOP framework, the ECS provides a uniform and homogeneous

structure and also includes communication mechanisms with the Data Acquisition/Trigger systems, a schematic view of the ECS scope is given in Fig.4.2.

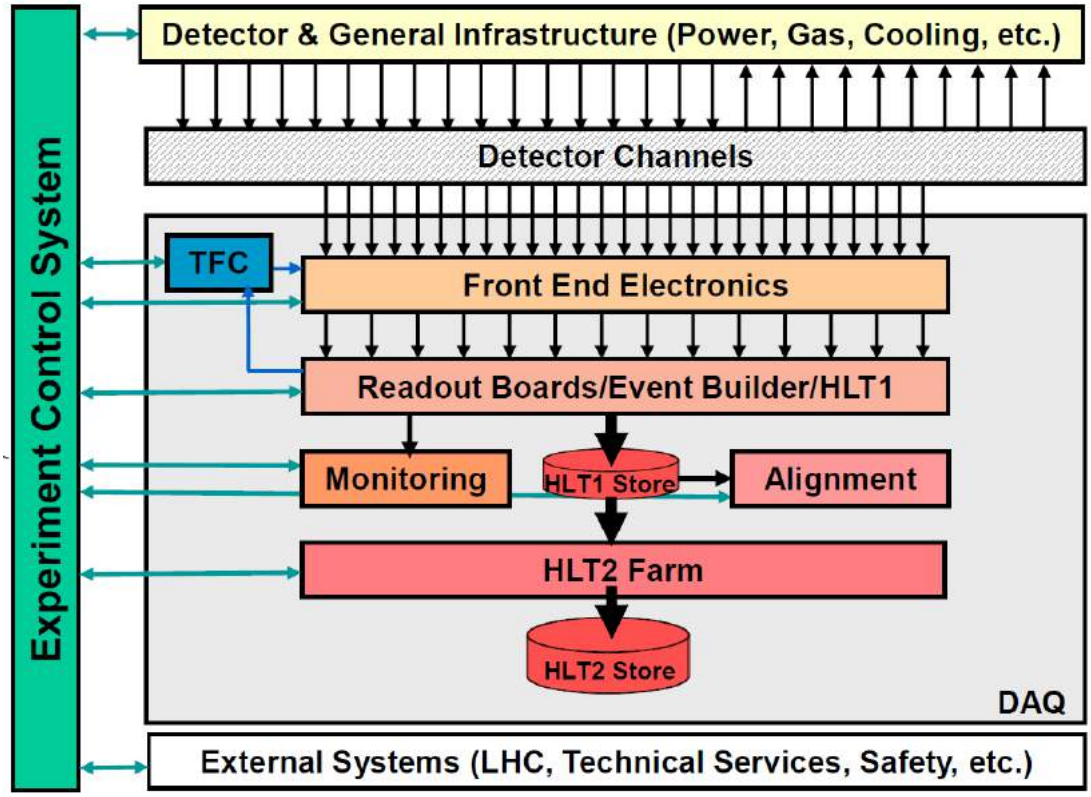


Figure 4.2: The ECS role inside the whole LHCb experiment. The ECS controls and monitor not only the LHCb-specific physical devices, but also supervise the monitoring, reconstruction, alignment and timing tasks as well as the communication with an external systems like LHC and safety.

The JCOP framework is built on top of the WinCC-OA Supervisory Control And Data Acquisition (SCADA) tool and it standardizes the development of the whole LHC control system, ensuring uniform solutions and allowing common tasks to be developed coherently for all the experiments and, at the same time, simplifying the communication between the different projects. Within LHCb, the ECS, given its complexity, follows a tree structure. Each subdetector has its own ECS which is further split into different partitions which tackle different tasks and supervise a specific part of the subdetector.

Control systems are developed using WinCC-OA, a device-oriented software. Within this framework, a structure of variables is defined as Datapoint Type, and an instance that is saved in memory is called Datapoint. In a given Datapoint, individual variables (such as the value read, the hardware address, a range of values outside of which an alarm is raised) can be stored and they are referred to as Datapoint elements. An HV channel is, for example, declared as a Datapoint, and its structure is defined by a Datapoint Type. The JCOP framework describes in this case the structure of the Datapoint Type to have a uniform Control System and is integrated into the WinCC OA project. Each Datapoint can also be represented as a Finite State Machine (FSM).

4.2.1 The Architecture of the LHCb Experiment control system

A hierarchical, tree-like structure can be used to represent the structure of LHCb subdetectors, subsystems, and hardware components. As a result, sub-detectors and sub-systems can be controlled independently during calibration and testing, but can be managed coherently during physics data taking. The nodes in the tree behave like FSM objects and can only be in one state at a time. In a hierarchical structure, commands only propagate downward, while states propagate upwards. In other words, a node can only send commands to its children and not to its parents, and its state is only determined by the states of its children. In the tree-like structure Device Unit (DU) and Control Units (CU) are defined :

- DU: the leaves of the tree structure and do not implement logic behavior. The DU interface directly with the lower-level components (both hardware and software) and they send the commands to perform an action as well as receive the device data. If the command is successful, or the data are as expected, the state of the DU is set.
- CU: logical decision units able to send commands to children, the logic is evaluated in a state transition phase. The FSM defines the logic and it is followed when a command is sent from an operator or from the parent units.

Within the JCOP framework, the FSM tool is built using SMI++ [65]. In the SMI++ framework, objects behave as finite-state machines. In addition to representing real entities like hardware devices and software tasks, these objects can also represent abstract subsystems. The rules are user-defined and propagated to facilitate automation and error recovery. Distributed SMI++ objects run on a variety of platforms through a transparent communication system called DIM (Distributed Information Management). In particular, the FSM toolkit allows to detach and operate every object acting as FSM independently from the control hierarchy. The software implements also the partitioning feature, meaning that it allows individual sub-systems to be controlled independently and concurrently with the other objects acting as FSMs. This is required for calibration and testing purposes. If a child is partitioned out, or excluded, the state is ignored by the parents in every decision process, the commands are not sent to the child and the child state does not influence the top states, also the child itself can be used freely (taken by another parent or acting as a FSM). The LHCb experiment uses the SMI++ component of the framework for monitoring and control of their Detector Control Systems (DCS), for controlling the Data Acquisition System (DAQ) and for the automation of the complete experiment (Fig4.3).

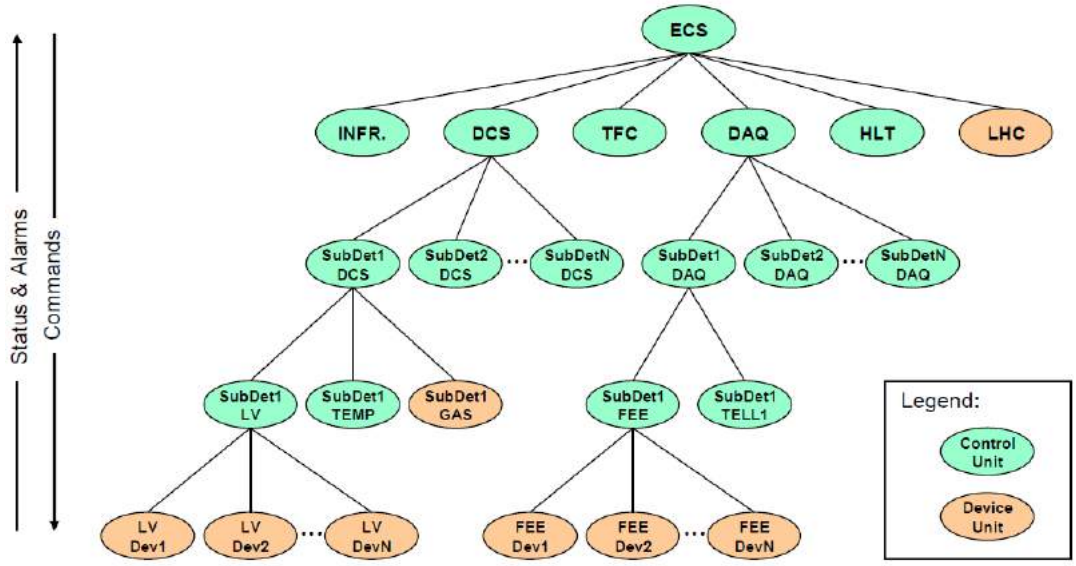


Figure 4.3: The schematic view of the LHCb control hierarchy based on SMI++

To operate the LHCb controls, the FSM is divided into two main panels (Fig.4.4):

- RunControl: integrates all the subdetectors, and handles the DAQ and dataflow. It is used to configure the system and start/stop the data-taking.
- BigBrother: manages the equipment related to the LHC state. It controls the HV of all the sub-detectors, the RunControl and the VELO closure and opening.

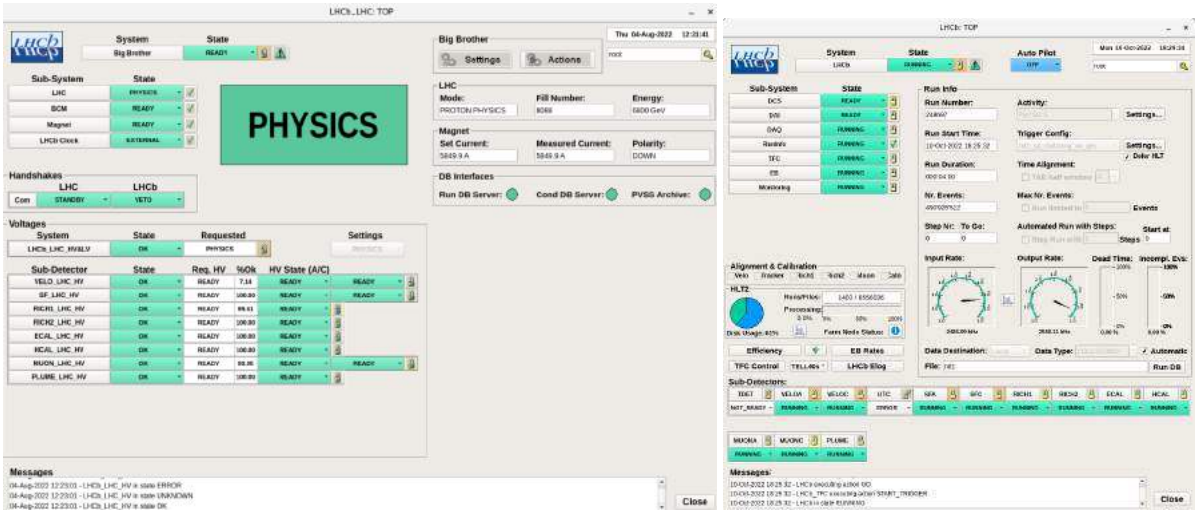


Figure 4.4: Left: the BigBrother panel, it receives the LHC state and executes automatic actions based on such state. RIght: panel of the RunControl with all the subdetectors included.

4.2.2 The RICH Upgrade ECS

The domain controlling and summarizing the state of the RICH detectors is the RICH Experiment COntrol System, called ECS (Fig.4.6). The RICH ECS is supported by different WinCC-OA projects

deployed as a distributed system and running in different virtual machines. The projects and their connections are shown in Fig.4.5. The RICH ECS comprehends three partitions that are specifically developed for the RICH sub-system inside LHCb (HV,DCS,DAQ) and the others are in common with other subdetectors:

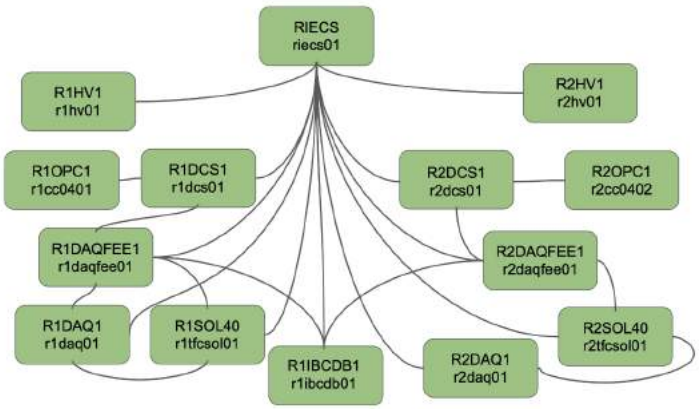


Figure 4.5: RICH specific WinCC-OA projects. Each project manages a different partition in the RICH ECS.

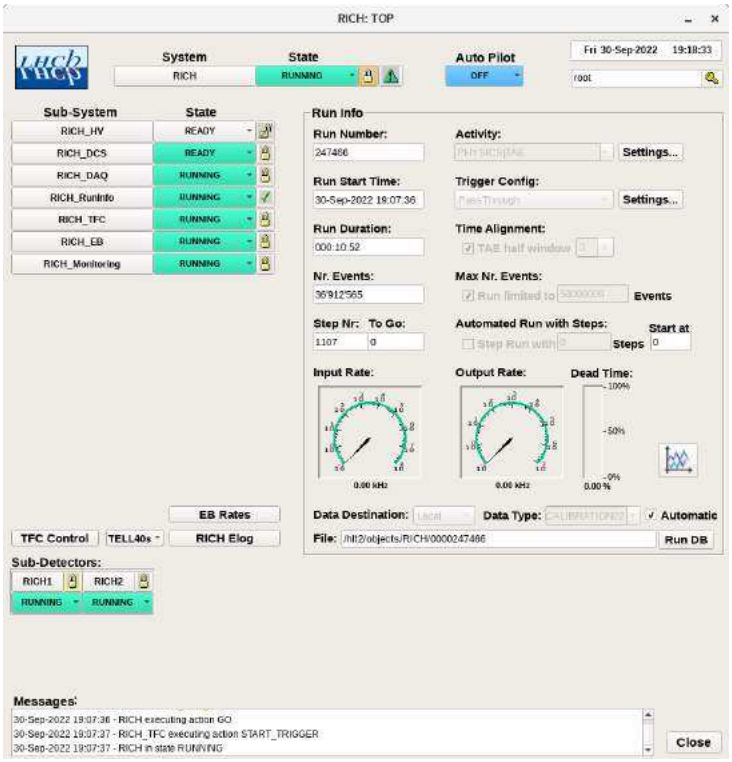


Figure 4.6: Main panel of the RICH FSM during a local data-taking.

- TFC (Timing and Fast Control) : the TFC partition gives access to the TFC panel which is of main importance for data taking and testing. From the point of view of TFC, the timing information refers to the bunch crossing ID (BXID or Bunch ID), this BXID is generated within the

FE modules and increments synchronously with the LHC machine clock of 40.079 MHz. As a consequence, the LHC bunch structure consists of 3564 bunches spaced by 24.95ns. From the panel the time trigger configuration can be set. The sub-panels most useful for commissioning purposes (Fig.4.7) are the following:

- Enable: enables the triggers. For the scope of the work explained in this thesis, the triggers of interest and the relevant features for the work presented are:
  - Periodic Trigger 1: trigger on the first isolated BXID with colliding bunches, the BXID value is automatically detected by the trigger
  - Periodic Trigger 2: the same as before, but identifies the second isolated BXID in the LHC filling scheme
  - Calibration Trigger A (CalibA): triggers connected to the laser illumination system in the RICH enclosures
  - Random Generator and Random Trigger A: trigger used to open randomly an acquisition window. Useful for DCR studied and to identify the BXIDs with collisions.
  - Timing Alignment Event (TAE): enables the possibility to open a window in units of BXID around the main BXID in which fires the selected trigger
- Triggers config: sets the rate of the random trigger, and the BXID for the Periodic Triggers and Calibration Trigger A.
- Trigger Masks: used to enable the TAE window for the required trigger by the make TAE flag
- Commands 2 table: set the TAE half-window

The TFC partition also includes the SOL40s, for timing studies the link00 of one SOL40 is shown in Fig.4.8.

- EB: the Event Builder partition controls and monitors the servers hosting the TELL40 and GPUs. The EB is also responsible for handling the dataflow.
- Monitoring: it manages the tasks for the online monitoring of the subdetectors. A subset of events is collected and organized in plots to check the quality of the data and the performance of the detector. The monitoring is useful to have fast feedback during the data taking, for example, mis-configured FE boards in the RICH can be directly detected when the hitmap is visualized.
- HV: the HV partition controls and monitors the HV power supplies. It follows the state shown in Fig.4.9 to control the ramp up and ramp down of the HV. When in global, the HV partition is integrated into the HV domain of LHCb, in the BigBrother FSM.
- DAQ: the DAQ partition controls the Front-End (FE) boards, the PDMDBs, and the LHCb readout boards related to the RICHes back-end (BE), the TELL40s. The DAQ contains the controls and status of the TELL40s and the FEE. In the DAQ panels, the state of the FEE is monitored

(to check the correct configuration) as well as the TELL40 links. A link correctly transmitting/receiving information from the data center is locked and a link losing the lock prevents the correct functioning of the element of the detector. There are ~ 1500 data links in RICH1 and ~ 1000 data links in RICH2. The master links distributing the clock through the SOL40 are ~ 250 in RICH1 and ~ 150 in RICH2. The communication between the GBT servers and the corresponding WinCC projects (RxDAQFEE1, RxDAQ1 and RxSOL40, x=[1,2]) is done using managers grouping several devices (PDMDBs, TELL40s and SOL40s) together. The SOL40 GBT servers (one per SOL40) are running on the servers hosting the SOL40 cards and the TELL40 GBT servers (one per TELL40) are running on the EB servers hosting the TELL40 cards

- DCS: the DCS partition manages all the slow control equipment, meaning monitoring and control of the low voltage (LV), temperature, pressure, and humidity. DCS must always be in state READY, meaning all the sensors are in the correct range of values and LV is ON. The Safety FSM is integrated in the DCS and uses DCS sensors as inputs and it switches off LV and HV of a single column if a temperature is too high. In the worst scenario it switches off LV and HV of all the columns of one MaPMT enclosure if the cooling flow stops or if the temperature of the box or the cooling are too high. The DSS is the last resource of safety in case of issues during network failures: it is designed to be a fully hardwired system, for instance, thermos witches are present on each column, triggering the power off of the D3 racks in case of high temperatures.

Figure 4.8: Panel of the link00 of one SOL40. Of particular interest for timing studies is the SOL40→TELL40 delay.

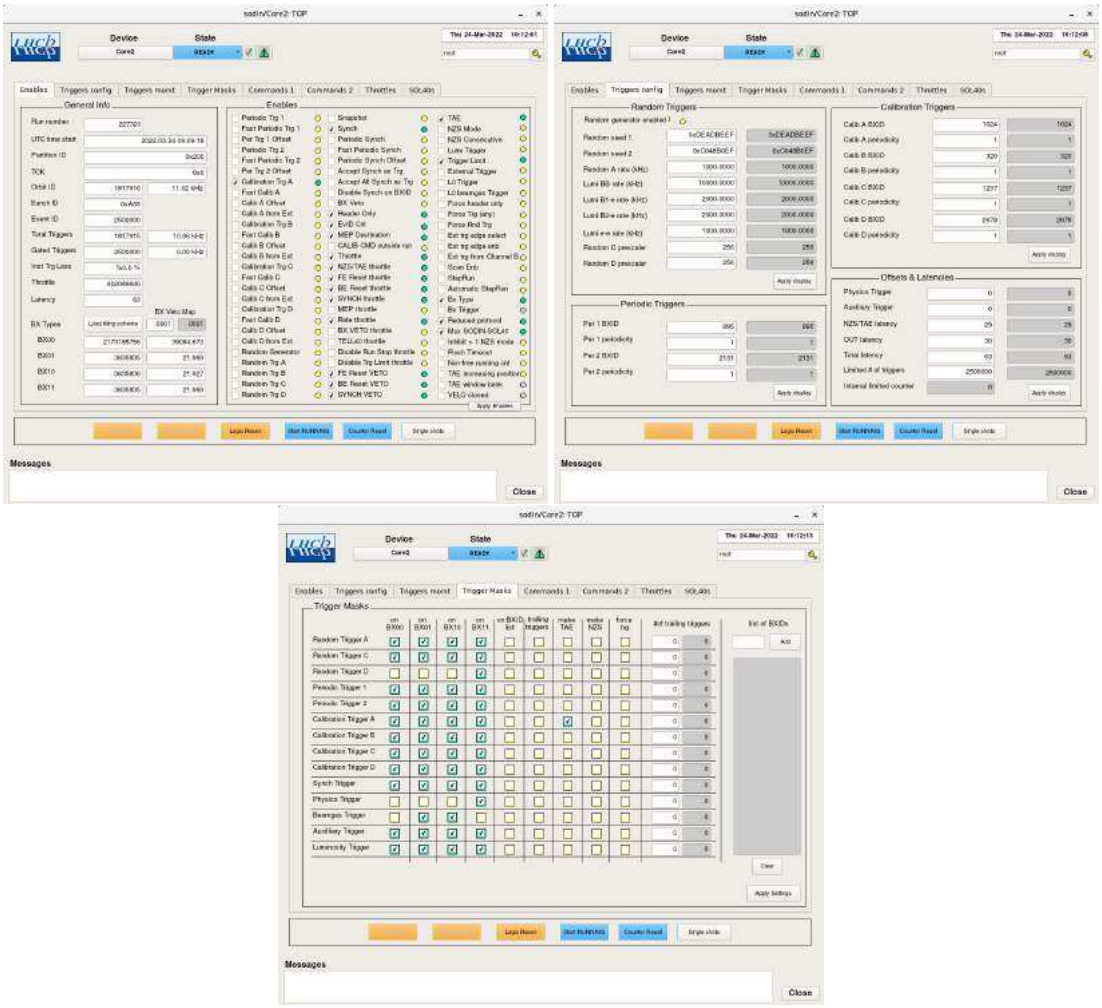


Figure 4.7: Relevant panel of the TFC partition. From left to right: Enables panel enabling the CalibA trigger with TAE, Trigger config panel and Trigger Masks panel.

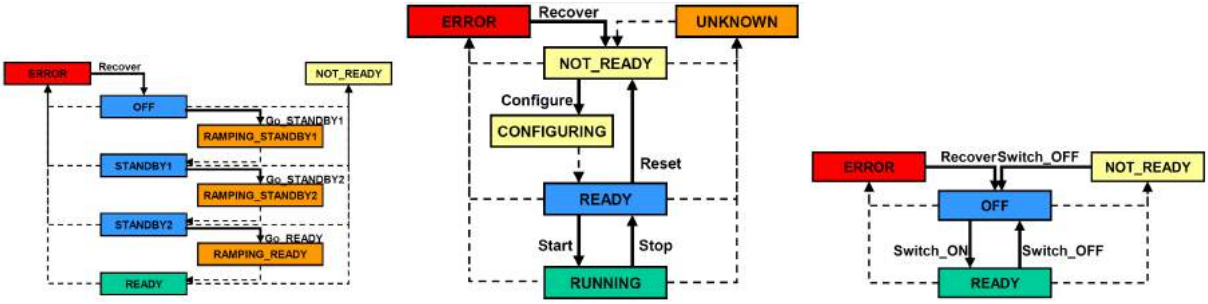


Figure 4.9: Logical states of HV (left), DAQ (center) and DCS (right) domains.

The state of LHCb and all the subdetectors strongly depend by the LHC state. The LHC cycle during injection is shown in Fig.4.10. For the RICH detectors, at the injection, the HV is switched off to prevent damages due to faulty injections which will generate a lot of muons going through the RICH planes. LHC can inject only if the HV state of the RICH is OFF, meaning, only at this point the RICH state influences the whole LHC. During RAMP the HV is switched ON, the details of the switch on varies between RICH1 and RICH2, since in RICH1 also the powering of the last dynode is implemented. At this point, the FE should be correctly configured and at SQUEEZE some signal is already visible in the RICHes.

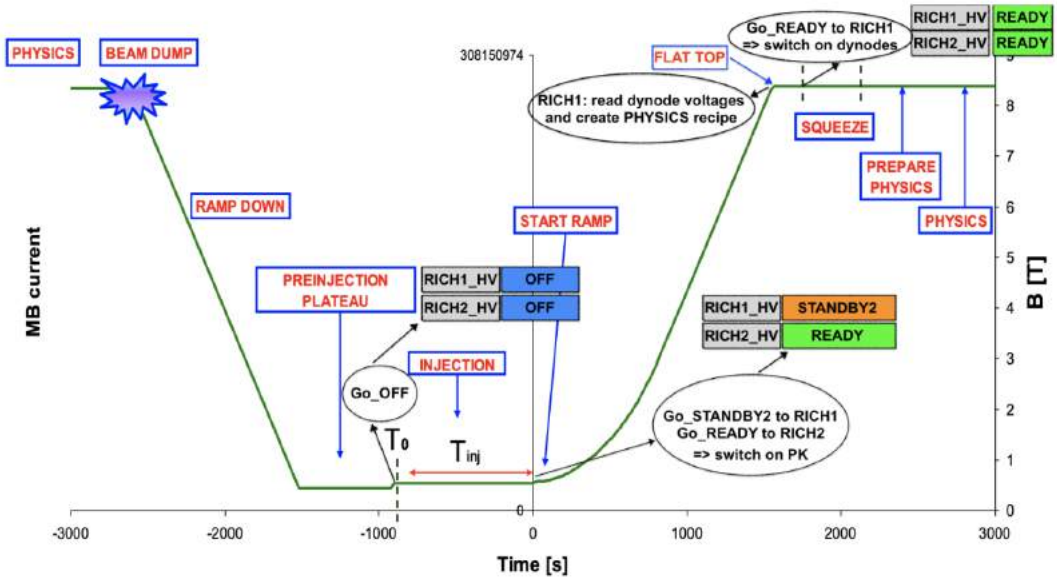


Figure 4.10: The typical LHC PHYSICS cycle. The beam is injected at 450 GeV, then the ramp-up reaches the nominal energy of 6.8 TeV. This is achieved by increasing B in the magnet and currents (frequency) in the accelerating cavities. In the SQUEEZE state the beam is focused in the transverse plane, making the beam smaller in correspondence with IP8, where LHCb is located. The data are acquired in the PHYSICS state since at this point all the subdetectors are ready. At the end of the fill, the BEAM DUMP state gets rid of the beam. The RICH states are also indicated.

4.2.3 Operating the RICHes

The RICH can be operated through the FSM in two ways:

- global control: the RICH partition is owned by RunControl. TFC and EB partitions are shared between the included subdetectors. During LHCb the data-taking the RICH should always be included in global.
- local control: the local mode is used for special runs, commissioning and calibration studies. The RICHes are run standalone and the RICH operator has full control. The measurements, test and calibration described in this chapter have been performed with the RICHes on local control, the structure of the FSM in local control is shown in Fig.4.11.

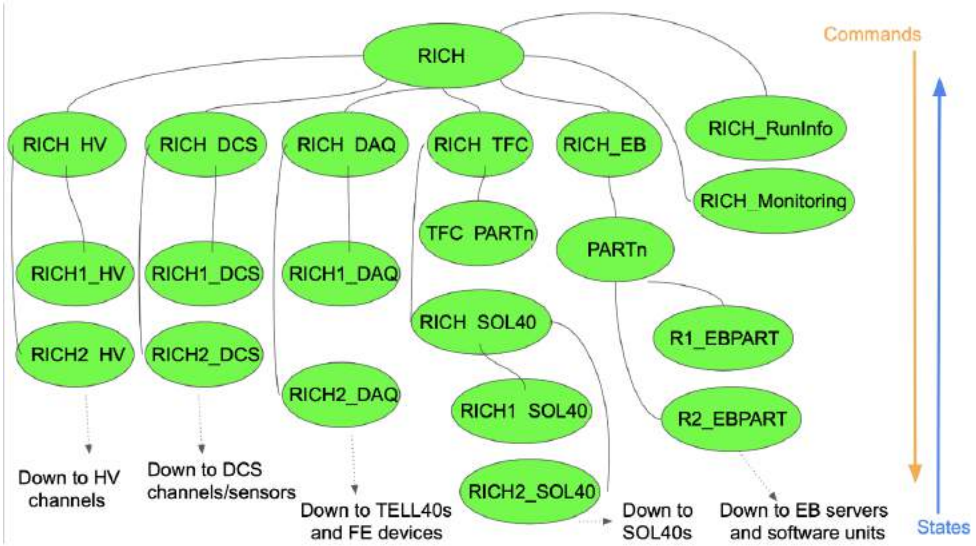


Figure 4.11: A part of the RICH FSM tree structure when operating in local control. Everything is owned by the RICH operator. When on global control, RunControl owns the subpartitions, and the resources are shared between the subdetectors included.

In a year of RICH commissioning, the detector was operated mainly in local control by experts for testing and debugging purposes. The stages of commissioning comprehend:

- data-link mapping validation at the single channel level
- solving various issues preventing the data-link and master-link loss of lock
- replacement of faulty FE components in the LHCb cavern
- debugging of the FE and BE firmware
- first RICH1 HV switch on procedure
- validation of the CLARO thresholds values from column commissioning and HV tuning
- definition the complete time alignment procedure and its validation

In the following parts, the most interesting steps are covered, focusing on time alignment, being a core activity of the last year of commissioning.

Mapping validation

The mapping between the internal ECS data-link channel numbering conventions to locate a pixel on the RICH plane has been validated by drawing a known pattern on the hitmaps with HV off. This is done by changing the configuration of the CLARO chip registers as shown in Fig.4.12. The channels which are used to visualize the known pattern are configured with a low threshold value (7) and the offset enabled. Those channels are, as a consequence, configured with the threshold under the noise pedestal and always trigger a signal at the CLARO output. The remaining pixels on the plane are disabled by configuring the CLARO channels with the highest threshold value (63) and offset bit disabled.

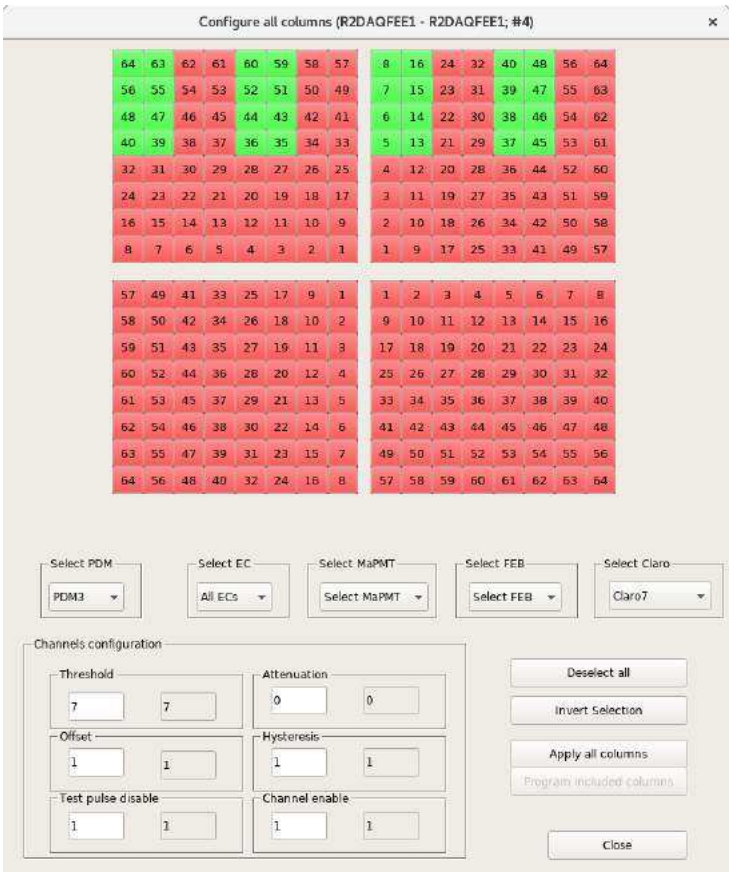


Figure 4.12: The panel used to configure the register of the CLAROs and set the parameters for each individual channel of the chip.

The validation patterns obtained after configuring the CLAROs are shown in Fig.4.13 and the patterns are as expected.

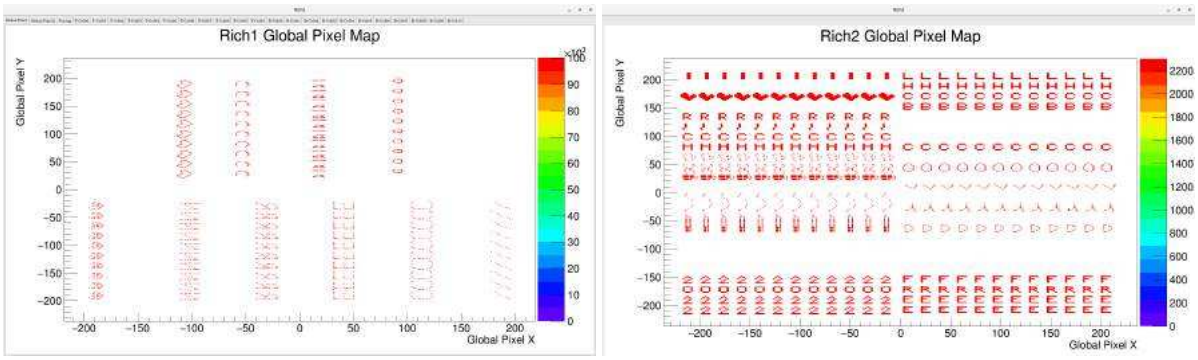


Figure 4.13: Validation of the patterns drawn by configuring individual CLARO channels from the WinCC project.

RICH1 switch on

The HV of RICH1 was switched on for the first time a in few steps. As a preliminary switch on, only the external PDMs of the outer columns were switched on, this allowed excluding major light leaks in the RICH1 box enclosure. The HV was then raised in steps (700V, 850 V and 1000V) after checking the current values and the voltage of the last dynodes. After a preliminary run taken with a small part of RICH1 switched on, the HV was turned on in the full RICH1 detector at 1000V, the hitmap obtained with the latter configuration is shown in Fig.4.14. From the hitmap, the DCR contribution can be clearly seen. The MaPMTs which exhibited better DCR and SIN were placed in the high occupancy regions which show a lower DCR contribution, even if the HV was not yet tuned for each PDM.

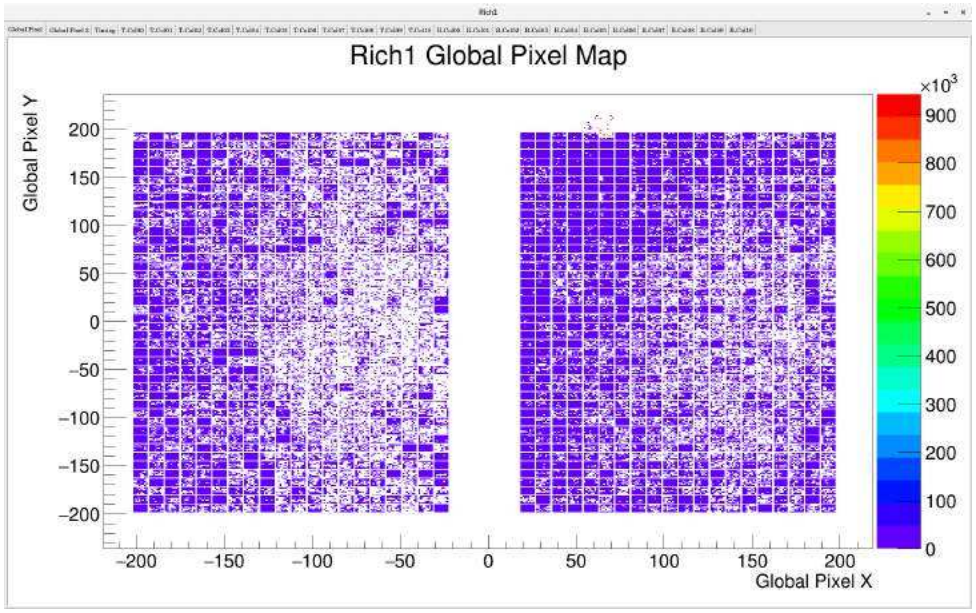


Figure 4.14: RICH1 hitmap obtained with HV at 1000V and no light sources. The hits are due to the noise contribution from DCR and SIN. The planes are rotated with respect the physical orientation, with the high occupancy regions to the right.

After the first switch on, the DCR contribution is expected to be high since the MaPMTs were off for a while and exposed to an external light source throughout the transportation and installation. To

obtain reliable DCR values, they were kept at 1 kV for a night, and then an HV at 1050 V was set for 7 hours (this process is referred as HV training). In Fig.4.15 are shown the pixels that exhibit a DCR higher than 1 kHz at each of described stages, a clear improvement is visible and further validates MaPMTs the reshuffling and grouping.

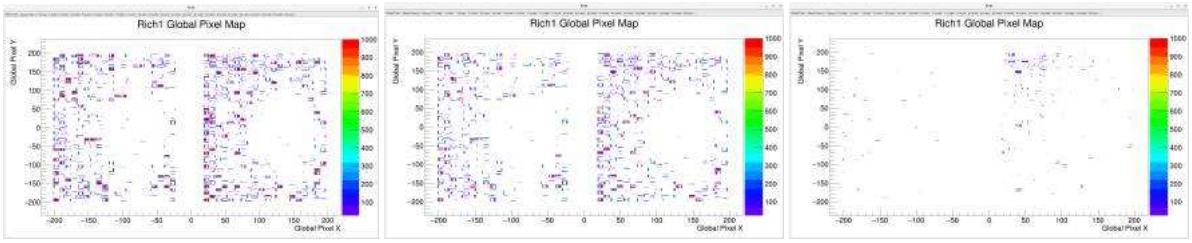


Figure 4.15: RICH1 hitmaps during the HV training process. Only the pixels above the limit threshold of 1 kHz are shown. The remaining noisy pixels are due to CLARO threshold not being set precisely, FE being badly configured or more HV training required.

Also, RICH2 has been kept at 1050 V over one weekend to perform an HV training after the final closing of the MaPMT and gas enclosures. In the dark counts hitmaps only very few pixels above the limit of 1 kHz, as expected (Fig.4.16).

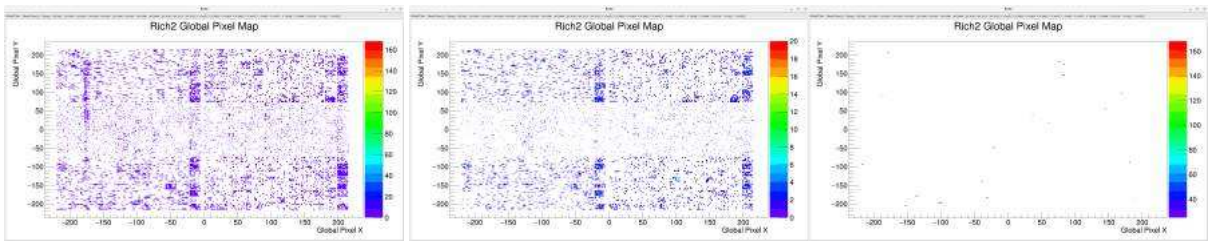


Figure 4.16: RICH2 hitmaps after the HV training process. On the left, the hitmap with the HV set to 1 kV for each PDM. The hitmap with the tuned values of the HV based on the MapMTs gain is shown in the center, and, on the right, the same distribution with the pixels above 1 kHz highlighted.

4.3 Time Alignment

The Cherenkov signal is required to be aligned in a precise BXID: in optimal conditions and stable operation, the data-taking is going to use the Fast Periodic Trigger and will acquire data at 40 MHz. Every BXID should be considered a different bunch crossing since, with complex filling schemes delivered by LHC, trains of colliding bunches are present. In trains of colliding bunches, the collisions happen in adjacent BXIDs, so the contamination from the leading and trailing BXIDs should be minimized while maximizing the detection efficiency.

The primary goal of the time alignment procedure is to synchronize in one BXID the signal generated by a bunch crossing. This is achieved by setting the proper time delays at the SOL40→TELL40 level and to individual TELL40 link (coarse time alignment) but also by applying a proper time gate in the 25 ns windows for the CLARO signal detection in each channel of a TELL40 link (fine time alignment). The time alignment studies and calibration requires the presence of a bunch crossing in

an isolated BXID and the determination of such BXID. The procedure is applied independently for RICH1 and RICH2. To study the time alignment, some 2d histograms are useful:

- Hits vs Bunch ID : in this plot the spread of the signal in the Bunch ID is shown. Each Bunch ID corresponds to a time interval of 25 ns between two consecutive LHC cycles. The plot is populated in correspondence of the Bunch ID in which an activity with a certain number of hits is seen during data-taking.
- Hits vs Slot ID: the same plot as before, but shown in a more compact way after the transformation SLOT ID = $abs(BunchID - 1024)\%23$.
- Bunch ID (or Slot ID) vs Bit ID: for each bit of the data links, the Slot ID is filled when a signal is seen in the corresponding data-link. When the detector is correctly aligned, the signals lie in the same Bunch ID (Slot ID). The difference between Bunch ID and Slot ID visualization is shown in Fig.4.17.

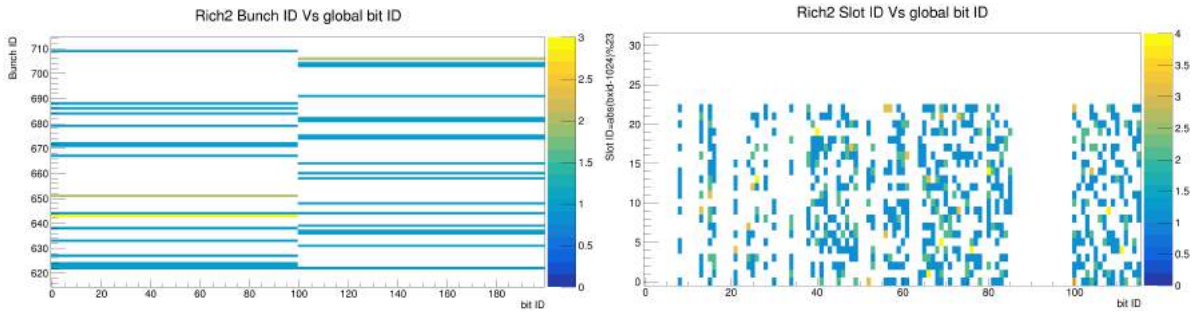


Figure 4.17: Zoomed histograms showing activity vs BitID. On the right, the Bunch ID is plotted against the BitID. The transformation from BXID and Bunch ID is applied to simplify the visualization for the time aligned and it's reported on the y-axis. Every x-bin corresponds to a channel of a data-link (a link has 86 channels, but 100 bins are dedicated to each link, the lasts being empty). On the left, the BXID is plotted against the BitID, but the bits are grouped per data-link, so an x-bin covers a range that is wide 100 units. The former is useful for the fine time alignment studies, the latter for the coarse time alignment.

Finding the BXID of the signal

A signal generated by the CLAROs can be observed in BXIDs, meaning the time intervals of 25 ns are defined by the 40 MHz SODIN clock. When acquiring data, the BXID corresponding to isolated bunches may not be known in advance. To search for activity in the BXID an acquisition enabling the Random Trigger is required. By triggering randomly, it is possible to populate the histogram Hits vs Bunch ID and visualize the BXID with a significant number of hits (Fig.4.18). Once the BXID of the collision is known, the correspondent value can be set to the trigger. The Periodic Trigger will open an acquisition window at the given BXID. If the RICH is not fully aligned in one BXID (the usual situation if the time aligned is not yet applied), the Timing Alignment Event (TAE) half-windows parameter should be enabled to open a window larger than 25 ns. The TAE half-window is a parameter useful for alignment studies and tests when LHC delivers isolated bunch crossing. In Fig.4.19 an acquisition with the two Periodic Triggers enabled and centered on the collisions BXIDs is shown.

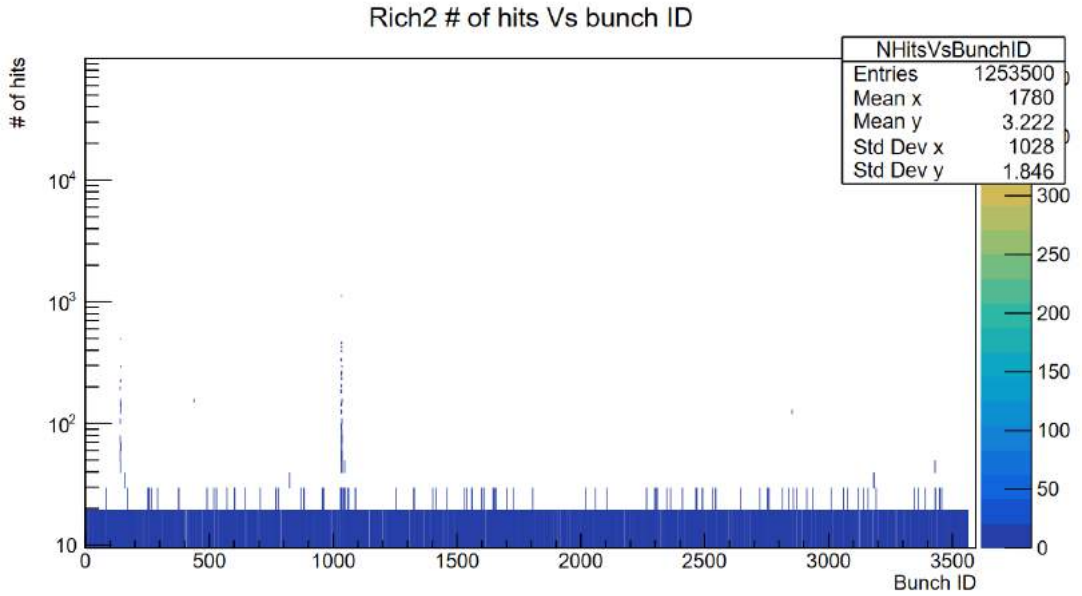


Figure 4.18: Acquisition with random trigger for RICH2. The activity vs the BXID is shown. The trigger fills randomly each BXID value with the number of hits in the event, two colliding bunches are clearly distinguished from the background.

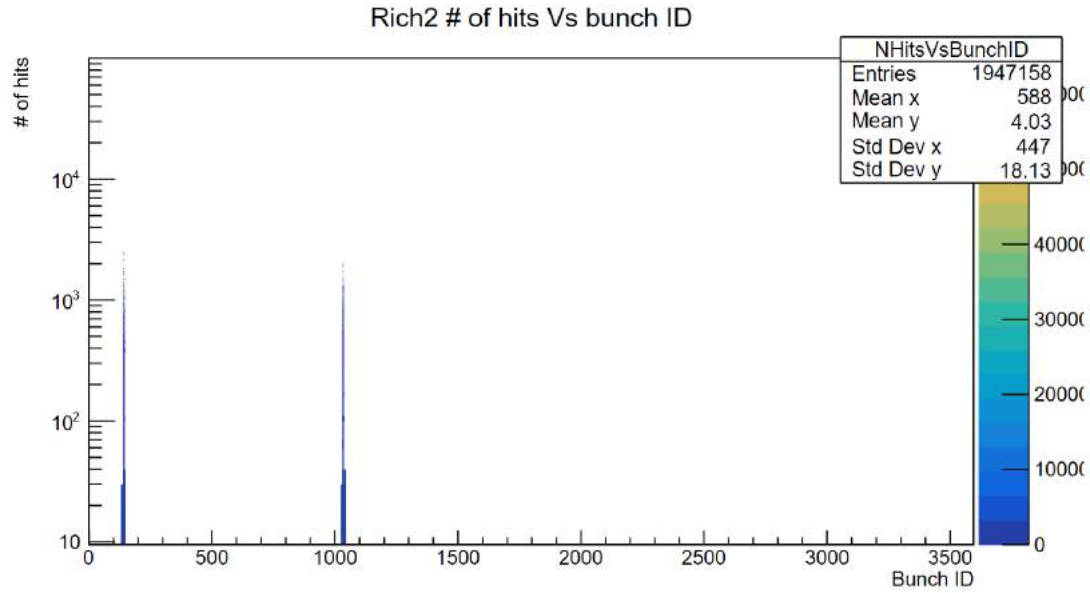


Figure 4.19: Acquisition for RICH2 with the two periodic triggers centered on the BXIDs of the delivered collision at P8. The activity vs the BXID is shown. The TAE half-window is set to 11, so for each trigger, a window of 23 BXIDs is opened and centered on the BXID set for the period trigger.

Coarse Time Alignment

The coarse time alignment consists in setting the correct delays in the SOL40s (meaning the delay between a SOL40 and its TELL40s, in which the clock is distributed) and in every single TELL40 datalink. The delay, in both SOL40 and TELL40, is set in steps of BXID.

Firstly, the TELL40s delays applied. In the default configuration, all the delays are not set (zero delay configuration). It is not expected, without setting the proper TELL40 to SOL40 delays for each data-link, to observe the signals in the same BXID, as can be seen in Fig.4.20, showing a configuration of RICH1 and RICH2 not aligned. This misalignment can be caused by a few factors, for example, the geometry of the detector (time differences between the inner and outer regions of a RICH plane), the FEE timing variations and the length of the fibers carrying the signal from the pit to the data center. The signal can be spread between more BXIDs, if a time gating finer than 25 ns is not applied. Indeed, in Fig.4.20, the signal of some data-links (or bit IDs) fills two Slot IDs, whose counts are higher than the background.

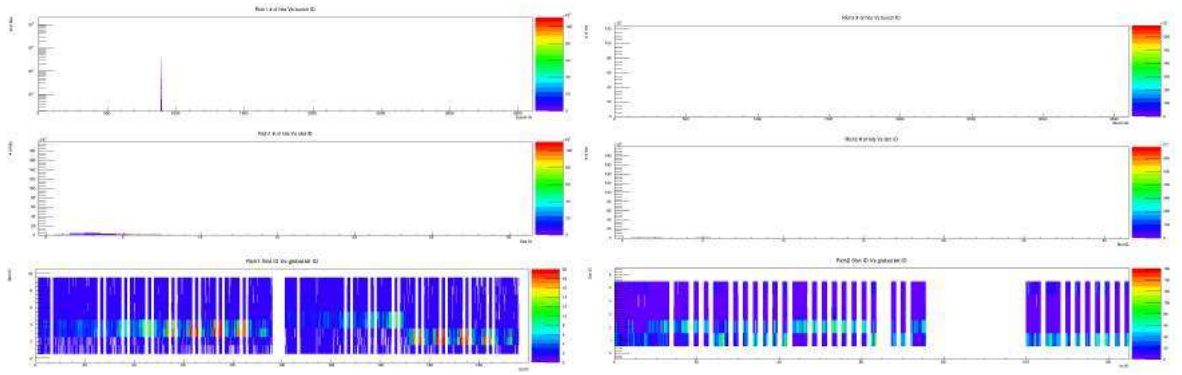


Figure 4.20: Timing information for both RICHes not aligned with the beam. This is evident by looking at the bottom plots, where the signal is not aligned in at least one Slot ID

The main procedure for the TELL40 interlink alignment with the beam is the following:

- acquire a Run with Periodic Trigger 1 and a TAE half-window large enough to include the full signal in near BXIDs. Process the run and analyze a histogram like the one in Fig. 4.20 from the decoded data. The analysis is implemented using the following logic, for each RICH detector:
  - cycle through the first half of the links (one-to-one correspondence with the x bins of the histogram), covering one side (U-side for RICH1 and A-side for RICH2)
  - project each bin (link) on the BXID axis and, from the projected distribution, identify the BXID with the higher number of counts for the given link
  - when the optimized BXIDs are known for each link, align all the links on the lowest BXID from the list by calculating the required shifts. The RICH side will be aligned on the lowest BXID found in the iteration.
  - the result linking the BXID shifts to apply and the TELL40 name and register location address are written in a file loaded by the FSM to set the TELL40 delays.
- After setting the delays from the FSM, a restart of pvss control and a reset of the TELL40-DAQ partition to enable the changes.
- acquire a Run with Periodic Trigger 1 and a TAE half-window. Now, most of the signal should lie in the BXIDs identified before for each side of the RICH detector.

- since the sides of a RICH detector are aligned independently, they may be optimized for two different BXID values. To shift a whole side, it is necessary to act on the SOL40− >TELL40 delay in Link00 of the SOL40 driving the TELL40s of the side under consideration. The same might be required to shift the whole RICH1 or RICH2 to the BXID set in the Periodic Trigger.
- once the detector is aligned in the same BXID of the Periodic Trigger, the TAE half-window can be reduced or disabled.

In Fig.4.22 is shown the RICH2 detector after applying the interlink alignment. In Fig.4.21 another visualization using the BXID vs number of hits of the coarse time alignment procedure.

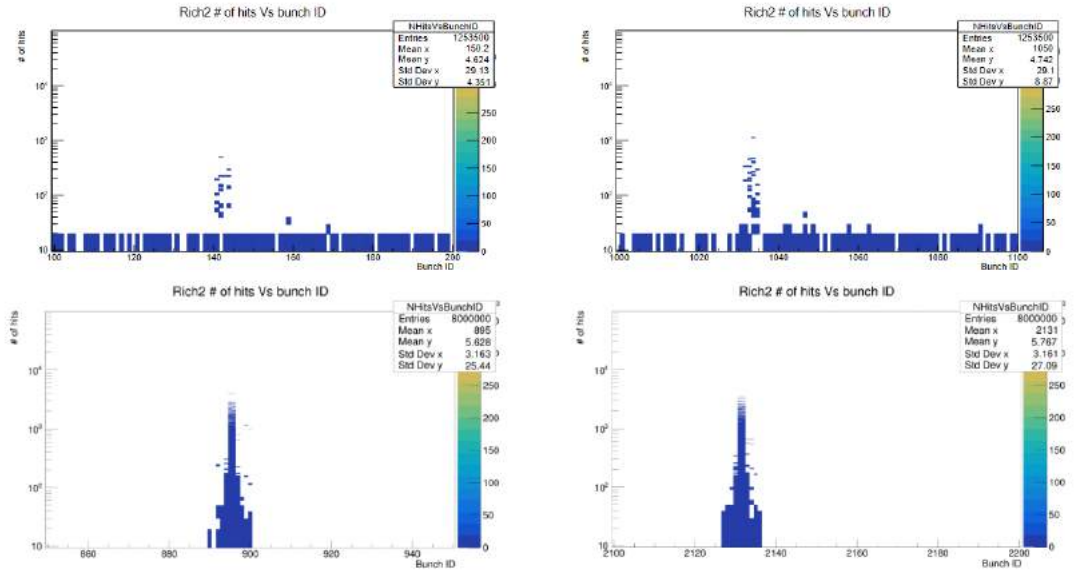


Figure 4.21: Activity vs BXID for RICH2 considering two colliding bunches. In the top row, the Random Trigger was used and the BXIDs with the collisions activity are observed, the two top plots show two colliding bunches after zooming into the interesting BXID region. The signal is clearly spread across at least 50 ns. On the bottom, two plots are different colliding bunches after the interlink alignment, The acquisition has the Period Trigger 1 and 2 enabled, and also the TAE half-window. The delays have been applied so that the central BXID contains the majority of the signal and is centered with respect to the full-time window opened with a trigger.

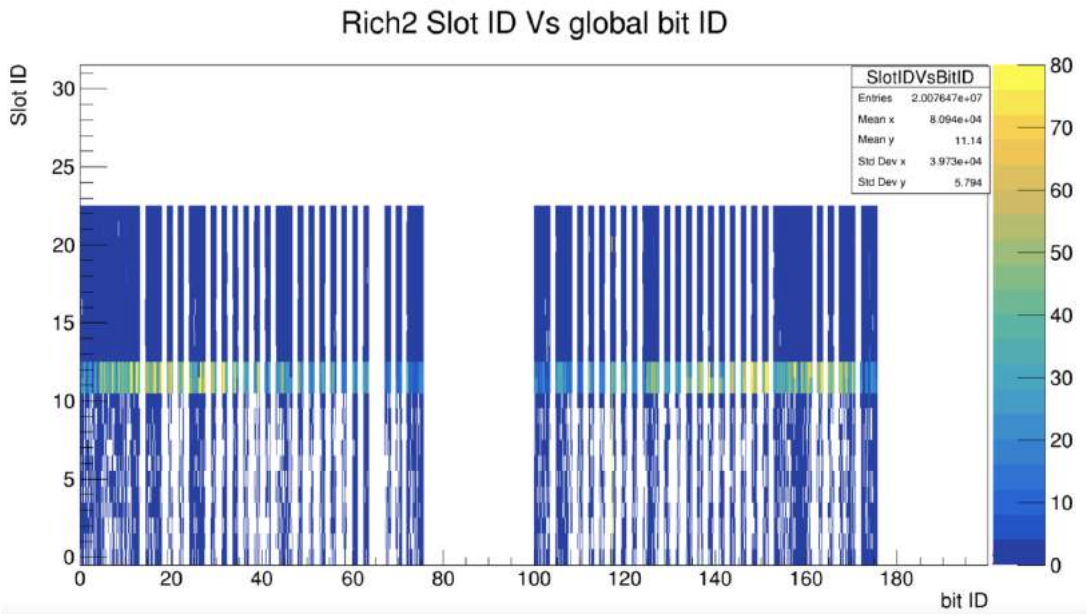


Figure 4.22: Aligned RICH2 after setting the TELL40 links delays with 2 colliding bunches delivered at LHCb. Data have been acquired with TAE half-windows = 11. Each link has been aligned having the BXID with the higher number of signals in the same BXID of the colliding bunches (corresponding to slotID = 11) and centered with respect to the 23-BXID window opened in the acquisition. The counts for slotID < 11 are due to dark count, for slotID>12 there are dark counts and SIN contribution. The signal lies in two BXIDs (slotID = 11 and 12).

Fine Time Alignment

The CLARO output is asynchronous with respect to the LHC clock and it is synchronized in the PDMDB FPGA. Each DTM on the FPGA has a 160 MHz clock and samples the CLARO signal at a 320 MHz bit rate using both clock edges, shifting the sampled data into an 8-bit shift register. A 256-bit lookup table is used to detect specific patterns arriving from the CLARO channel and output the resulting bit at the 40MHz system clock (Fig.4.23). In order to compensate for variations in signal latency across the photon detector plane, the clock phase can be adjusted per PDMDB or per FPGA by using the bit-slip feature of the IO deserializer on the FPGA, meaning to offset the 8-bit window in steps of one bit (at 320 MBit/s it corresponds to a time shift of 3.125 ns). The latency to be compensated is the result of the particle and photon time-of-flight through the detector, the MaPMT response time, the CLARO response time, and the time walk. The clock distribution for a DTM in a PDMDB is shown in Fig.4.24.

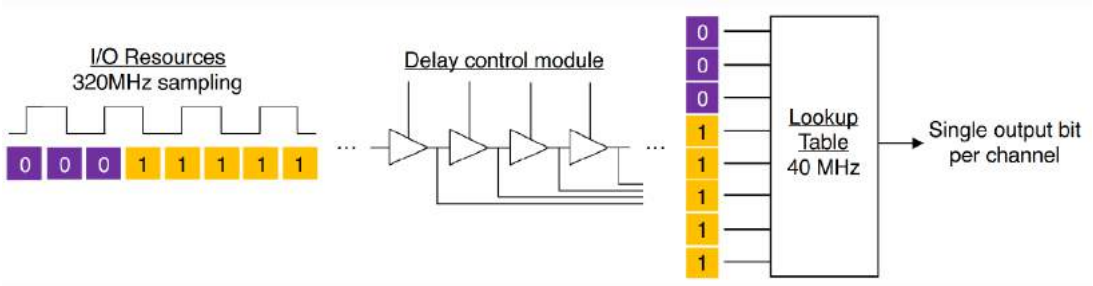


Figure 4.23: Schematic of the sampling of a CLARO channel at 320 MBit/s.

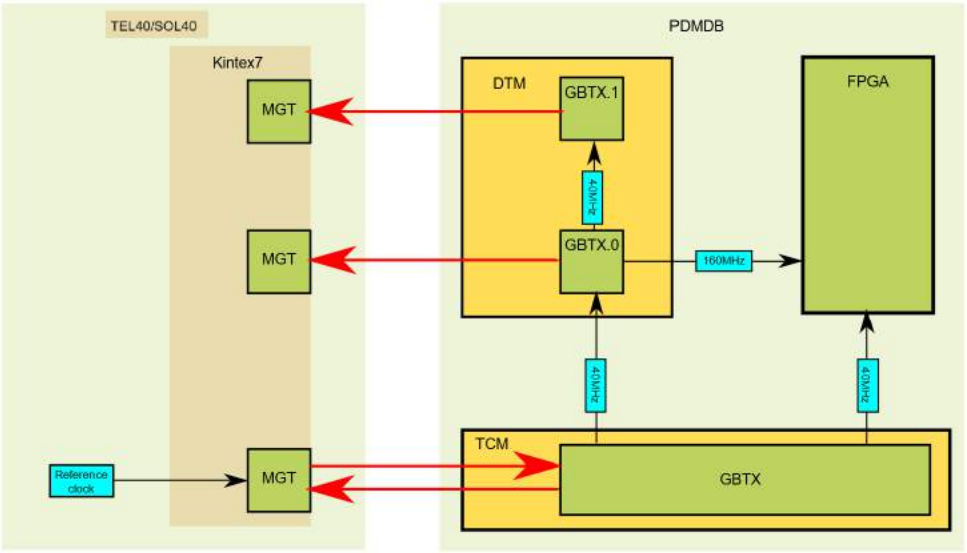


Figure 4.24: Schematic of the clock distribution starting from the reference clock. Red arrows indicate serial links which carry both clock and data. The PDMBD is connected to the Multi-Gigabit Transceiver (MGT) on the SOL40 or TELL40 boards.

The fine time alignment consists in applying a signal latching scheme by the lookup table and gate in a few ns to maximize detection efficiency while reducing out-of-time background. In practice, the BXID time window of 25 ns is divided into 8 minislot of 3.125 ns. Two different logic have been used during commissioning :

- level detection (level[0,1,2,3,4,5,6,7]): a signal is considered if the CLARO output exceeds the threshold in the given minislot
- edge detection (edge[01,12,23,34,45,56,67]): a signal is considered if the CLARO output exceeds the threshold in the given minislot but it doesn't in the previous one

The ideal logic to minimize background is to identify the signal edge (so apply the edge detection). From the MaPMT signals a time walk of less than 3.125 ns on the leading edge is expected. However, the gain dependence (based on the HV supplied) in combination with the PV time spread in the experiment (adding a jitter of ~ 0.5 ns FWHM) may require a 6.25 ns gate in practice. For this reason, also a double-edge detection logic (edge[012,123,234,345,456,567]) is implemented. The fine time

alignment needs to address not only the links as a whole, but the individual bits which correspond to a CLARO channel. The procedure to apply the fine time alignment is the following:

- for each minislot in the edge or level logic:
  - apply the minislot from the FSM panel to select the correspondent latching scheme for all the links
  - once the FEE is configured, take a Run long, at least, 1M events with a time window of five BXIDs (TAE half-window = 2)
- decode the data and analyze the output by looking at the SlotID vs BitID plot (Fig.4.17)
  - for each channel (BitID) the counts vs time plot is reconstructed, the time considered is binned in minislot (3.125 ns) and covers a number of BXIDs equal to the time window opened in the data acquisition
  - in case of level logic, an error function is fitted to the data to extract an estimate of the rising time of the channel
  - for the edge logic, the two adjacent minislots with the highest number of counts are detected
- the output of the analysis is then mapped from the TELL40 WinCC naming convention into the one used in the data center to address the correct links for each 6.25 ns edge detection found.

The level minislot scan is useful to have an estimate of the rising time of the signal, this allows to evaluate the possibility to apply a 3.125 ns edge detection or a 6.25 ns edge detection. The edge minislot scan returns the minislot in which the edge detection is most effective and, eventually the near minislot which contains the remaining signal to apply a 6.25 ns edge detection.

The full procedure to align the signal in one BXID has been first validated with signals generated from the FEE (specifically in the CLARO chip). Using the test pulse at the maximum amplitude (DAC code 255) and threshold 30 for all the RICH1 channels, a Run using the CalibA trigger enables the generation of the pulsed signal in the CLARO channels. The results of a minislot scan with test pulse are shown in Fig.4.25 for level detection in single channels or Link/FPGA and in Fig.4.26 for a RICH2 column for both level and edge detection scans.

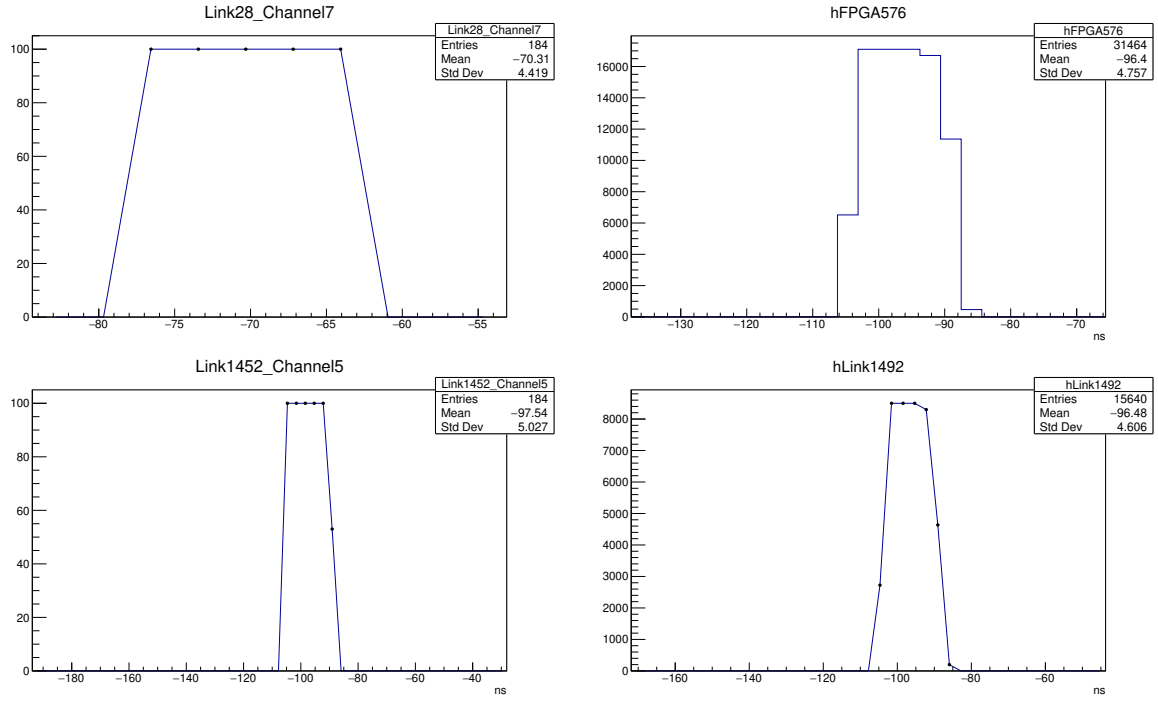


Figure 4.25: Examples of level detection scan for RICH1 (top row) and RICH2 (bottom row). The rising time is steep for the single channels.

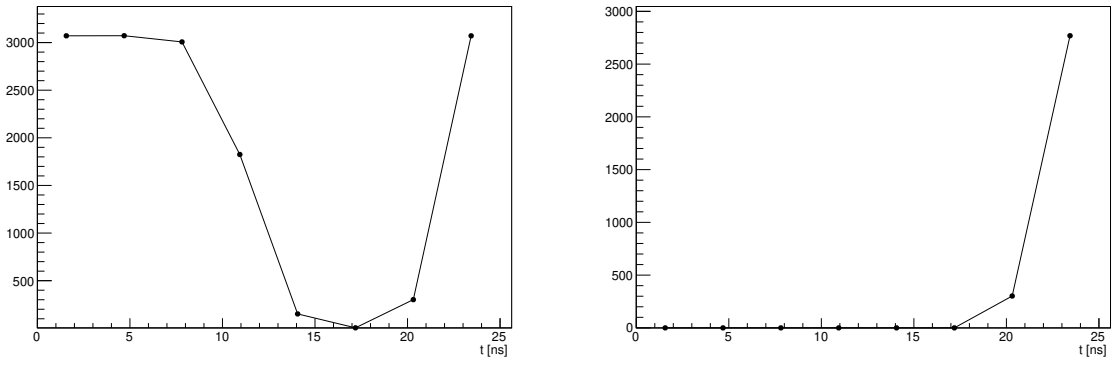


Figure 4.26: Level detection (left) and edge detection (right) scans for a RICH2 column (A01) using the test pulse. The behavior is consistent between the two scans. In the level scan, most of the channels are detected in level7, and a few of them trigger the level detection at level6. They satisfy the level requirement in level0 of the trailing BXID and the signals are mostly gone at level4, so the width of the generated test pulse signal is less than 25 ns. The level detection is coherent with the level shape since all the channels fire between edge6 and edge7 (the last two points).

To achieve the alignment of the whole detector in one BXID, the proper level detection was set in both RICHes, constraining the test pulse signal of each link channel in one BXID. The coarse time alignment was then applied to each link, the result of the operation is shown in Fig.4.27 for RICH1.

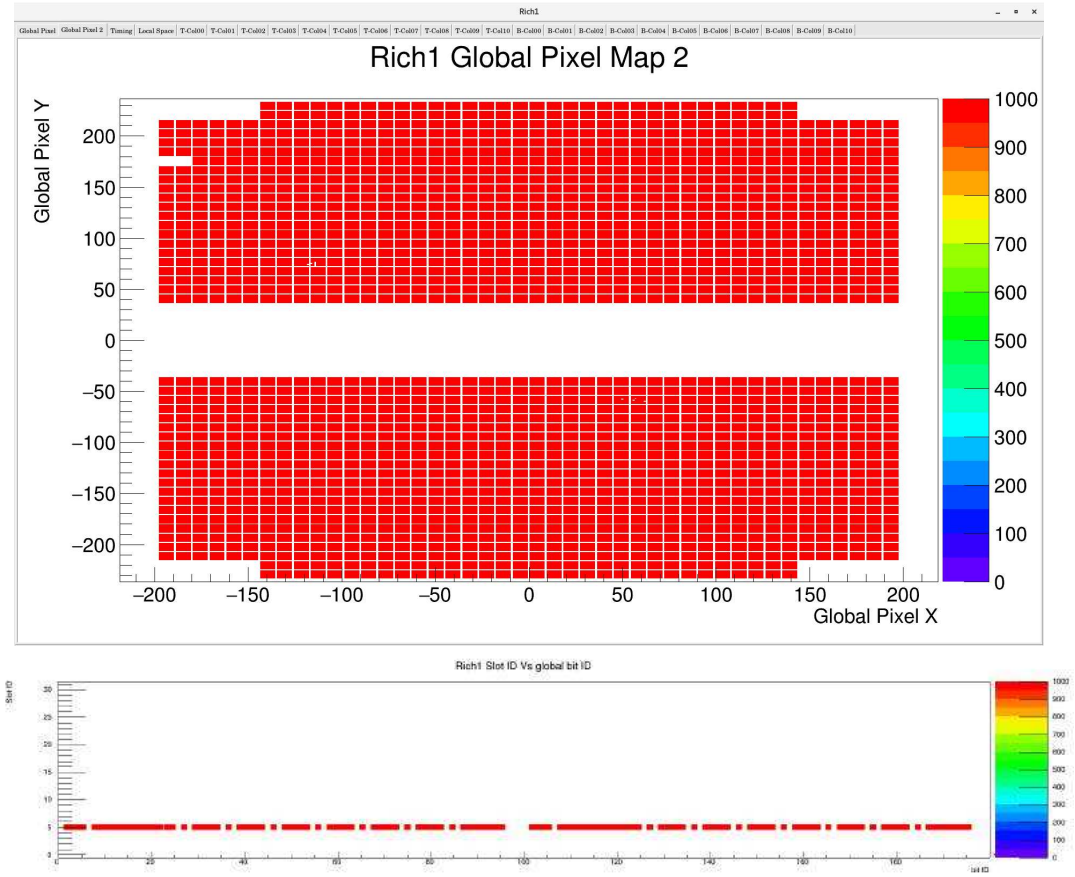


Figure 4.27: On the top, RICH1 hitmap with all the pixels firing due to the internal CLARO test pulse. On the bottom, the result of the applied interlink alignment. No TAE half-window was set in the acquisition, each trigger opened a window of just 25 ns. The minislot selected before the interlink alignment is level1.

After the validation with test pulses, the fine alignment with Cherenkov light from stable beam collisions has been studied. Some examples of level scan results are shown in Fig.4.28 for RICH1 and in Fig.4.29 for RICH2. The width of the rising edges implies that a 6.25 ns edge detection is mandatory to have good efficiency and not discard a significant amount of the signal.

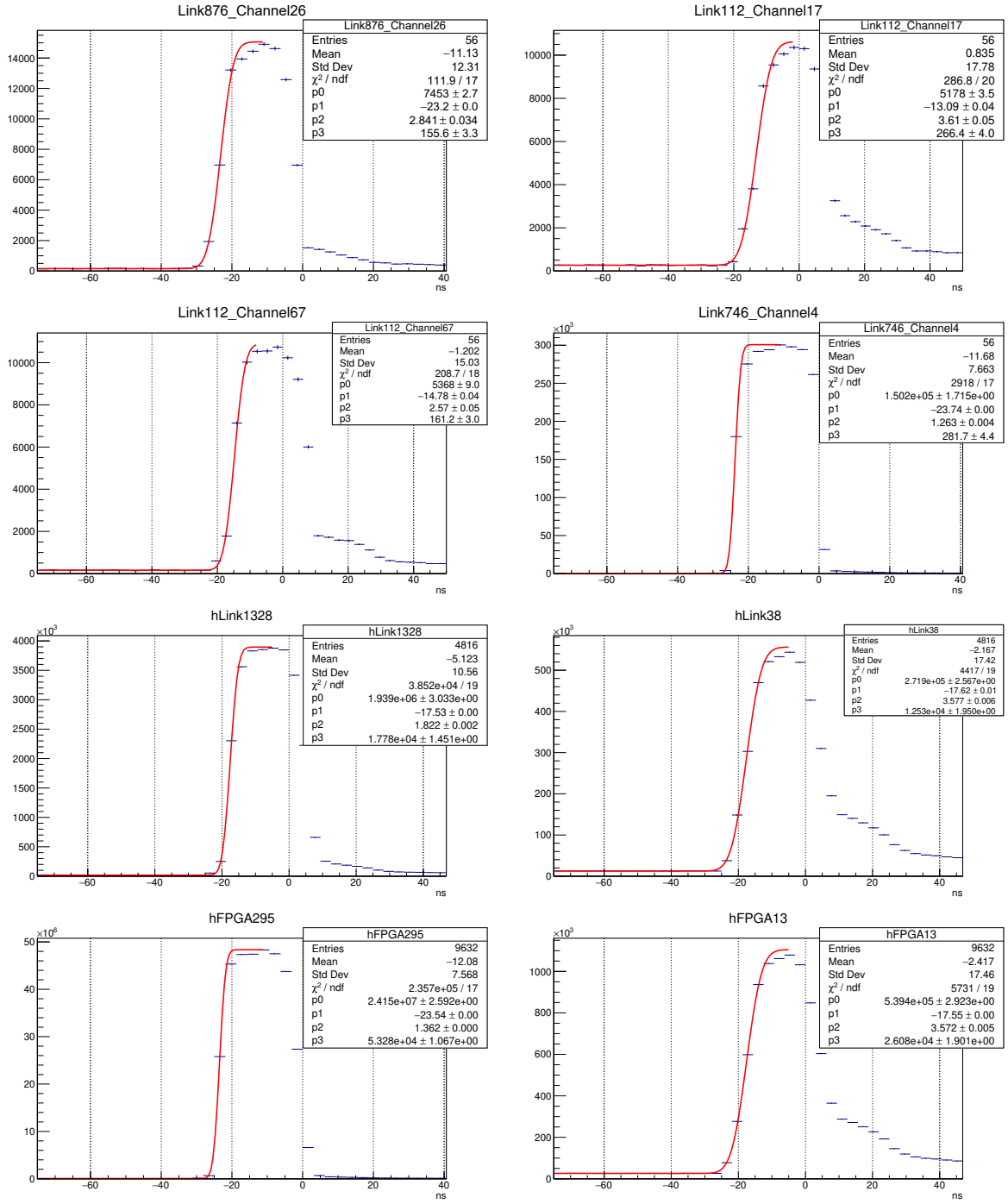


Figure 4.28: Profile of some RICH1 channels response in the level detection minislot scan. On the first two rows, four channels are shown. A TAE half-window of 3 was set (56 entries in the histogram = 8 minislots × 7 BXIDs). Given that a bin is one minislot wide (3.125 ns), the level detection provides and estimate of the width of the CLARO signal and of the rising time. The latter can be roughly estimated by the σ parameter of the Erf function (p2 in the plots above), the background is modeled as a constant (p3). Already in RICH1, the rising time implies the necessity to apply a 6.25 ns edge detection logic due to the non-uniformity across all the channels. On the last two rows, the same plot and Erf fit is shown, but for two links (all the channels are considered) and for two FPGA (two links) exhibiting different rising times.

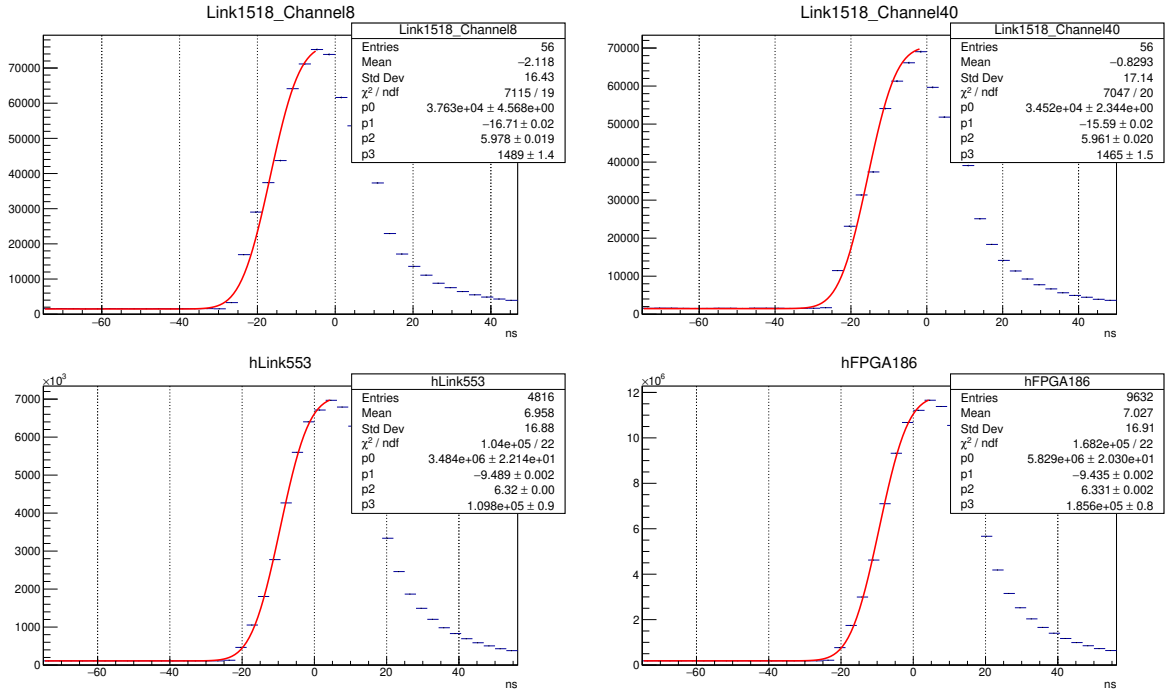


Figure 4.29: RICH2 channels response in a level detection minislot scan. On the top row, two channels are shown. The fit is an Error function modeling the rising edge and a constant for the background, as in the RICH1 channels. Especially in the RICH2 channels, a 6.25 ns edge detection logic is mandatory, being the width of the fit around 6 ns. On the bottom row a whole link and one FPGA are shown.

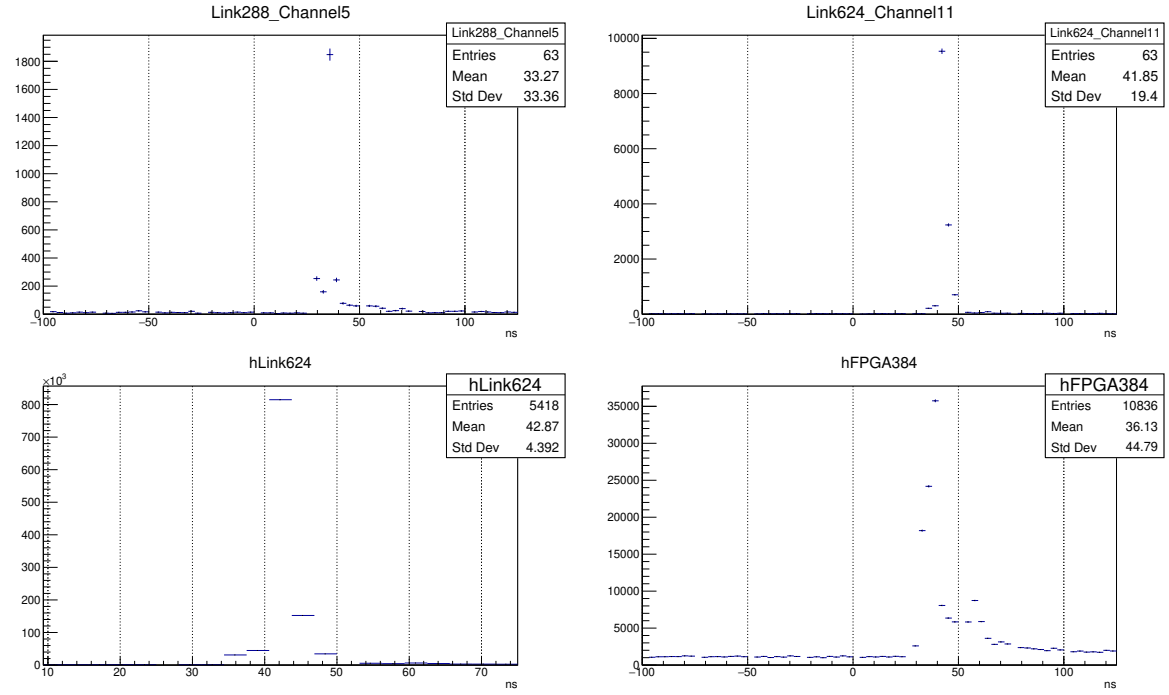


Figure 4.30: Examples of edge detection scan in RICH1. A bin corresponds to a minislot of 3.125 ns and the full-time range opened with the TAE is studied. In the case of edge detection, since the CLARO signal is sampled using as a reference the 40 MHz clock defining the BXID, one bin is empty at the BXID change, corresponding to the transition between minislot 7 and minislot 0 at the trailing BXID. Depending on the link, some background is observed in a few minislots before or after the one collecting the signal (the peak of the histograms).

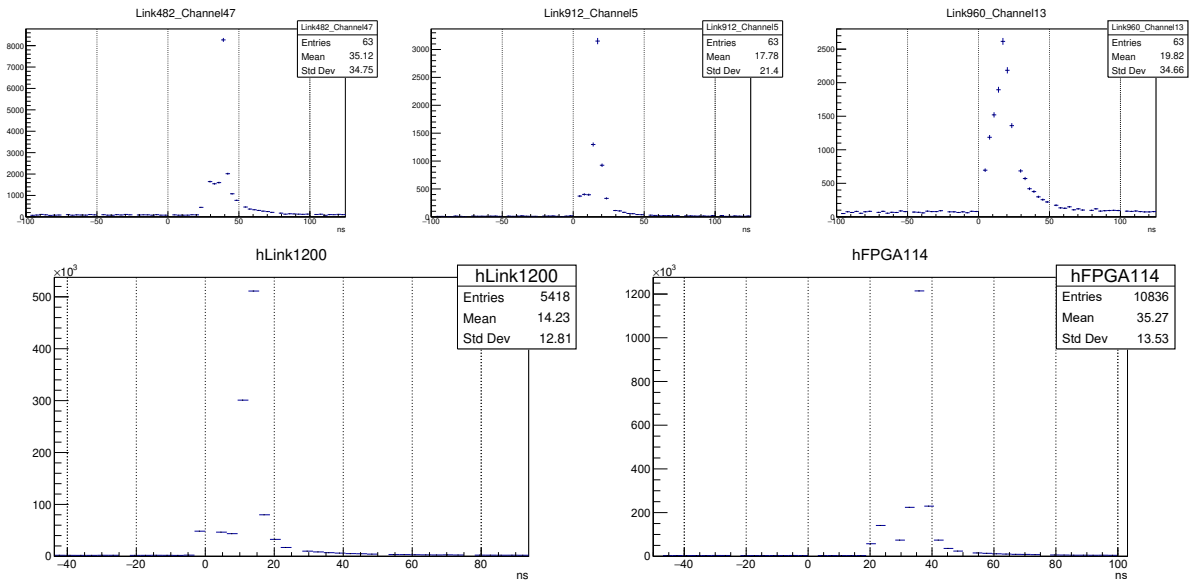


Figure 4.31: Examples of edge detection scan in RICH2.

A first attempt to perform and alignment of RICH1 using an edge detection of 6.25 ns has not been successful since in some channels the signal was not detected in any minislot (Fig.4.32). When the

two minislots with the higher number of counts have been identified with the edge detection scan at 3.125 ns and concatenate with an 'OR' logic to achieve a 6.25 ns edge detection, some holes in the RICH1 hitmaps are visible. This is explained by the fact that most of the signal lies on the edge transitions between two BXIDs, so in a hypothetical edge70 detection. The scan has been used to shift the clock phases (40 and 160 MHz) of the FPGAs having the corresponding channels with a transition at the edges of the 25 ns window (edge 0-1 and edge 6-7). The clocks have been shifted by 12.5 ns towards the center of the BXID window, in Fig.4.33 the hitmaps before the shift and in Fig.4.34 the hitmaps after the clock shift.

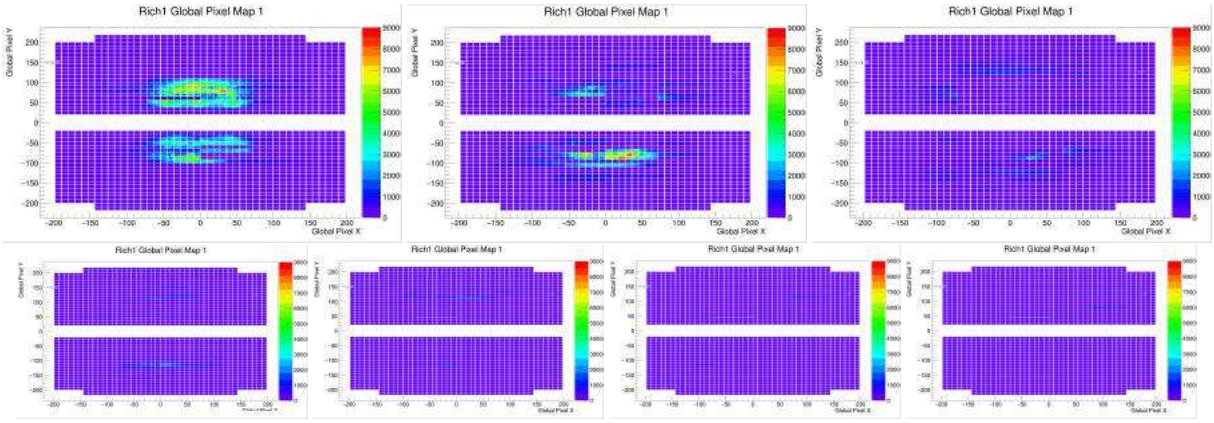


Figure 4.32: Examples of edge detection scan in RICH1.

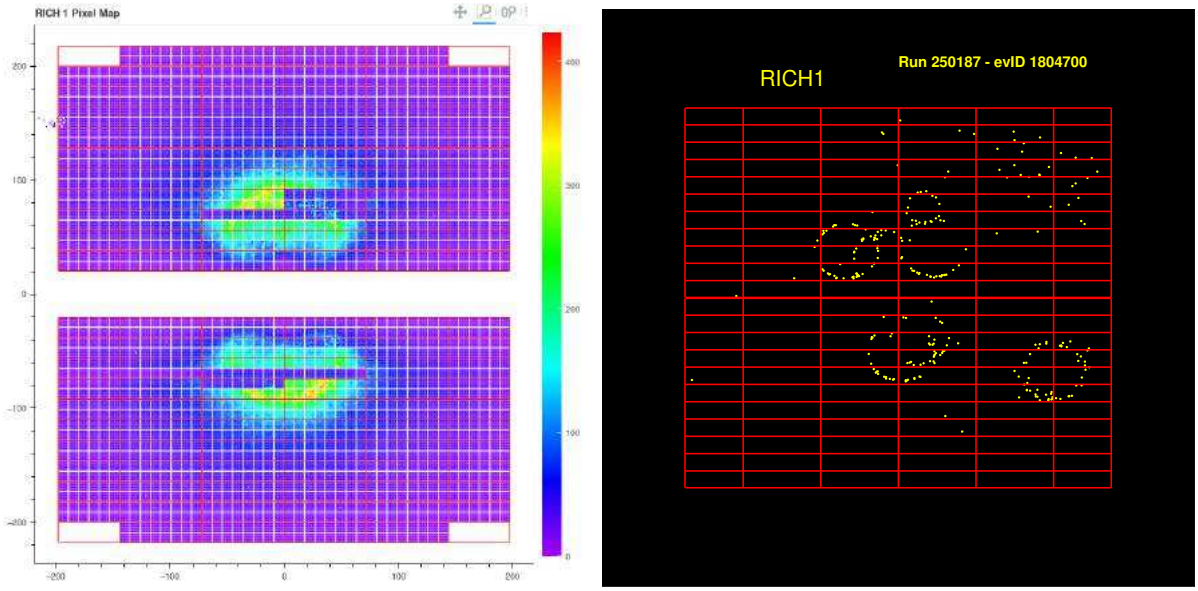


Figure 4.33: On the top, a run with the OR of edges detection (the darker regions are the ones having transitions between minislots 7 and 0). On the bottom, a single event display of RICH1, the hits which constitutes the rings are clearly missing in the off-time PDMs. The red lines define the region of each PDM.

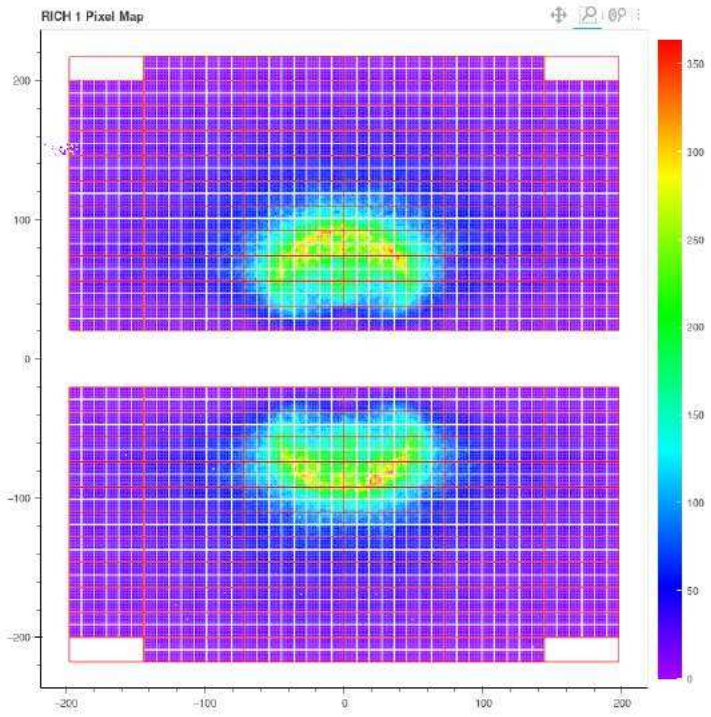


Figure 4.34: A run with the OR of edges detection after the application of the clock phase shift (no darker regions anymore).

4.4 Current status

After a year of commissioning at P8, the RICH detectors are able to operate at 40 MHz, as by LHCb Upgrade design, and with further background suppression using a 6.25 edge detection. The coarse time alignment procedure has been refined and it can be performed at every fill. In Figs.4.35-4.36 taken form the monitoring page, both the RICH detectors operate after time alignment with an edge detection at 6.25 ns. In Fig.4.37 a visualization of single events with the RICHes in optimal conditions, the rings generated by the charged track going though the radiators are clearly visible.

The data collected at the end of 2022 allows a very preliminary estimate of Cherenkov angle resolution, meaning only a preliminary photon detector panel alignment is performed with very preliminary tracking performance, without any mirror alignment and refractive index calibration. Despite the preliminary status, RICH1 already has a better resolution that its Run2 predecessor and RICH2 is approaching the expected performance 4.38.

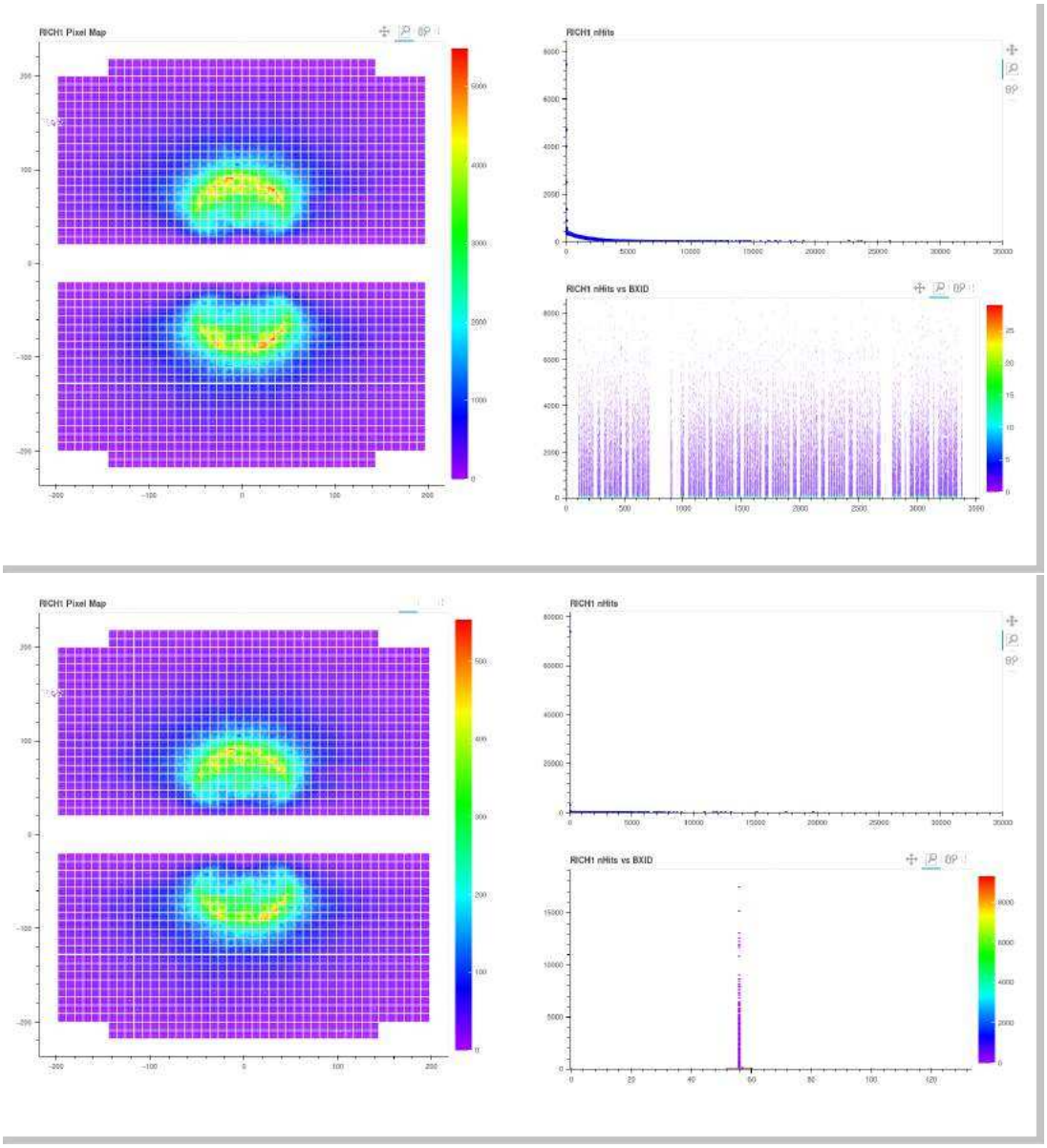


Figure 4.35: Improved timing in RICH1 detector. Top: Run with Random Trigger on bunch-bunch collisions. In this fill, trains of colliding bunches were delivered to P8, but also isolated bunch crossing are presents. The BXIDs with activity are visible in the bottom-right plot. Bottom: Run acquired with Period Trigger on isolated bunch-bunch BXID.

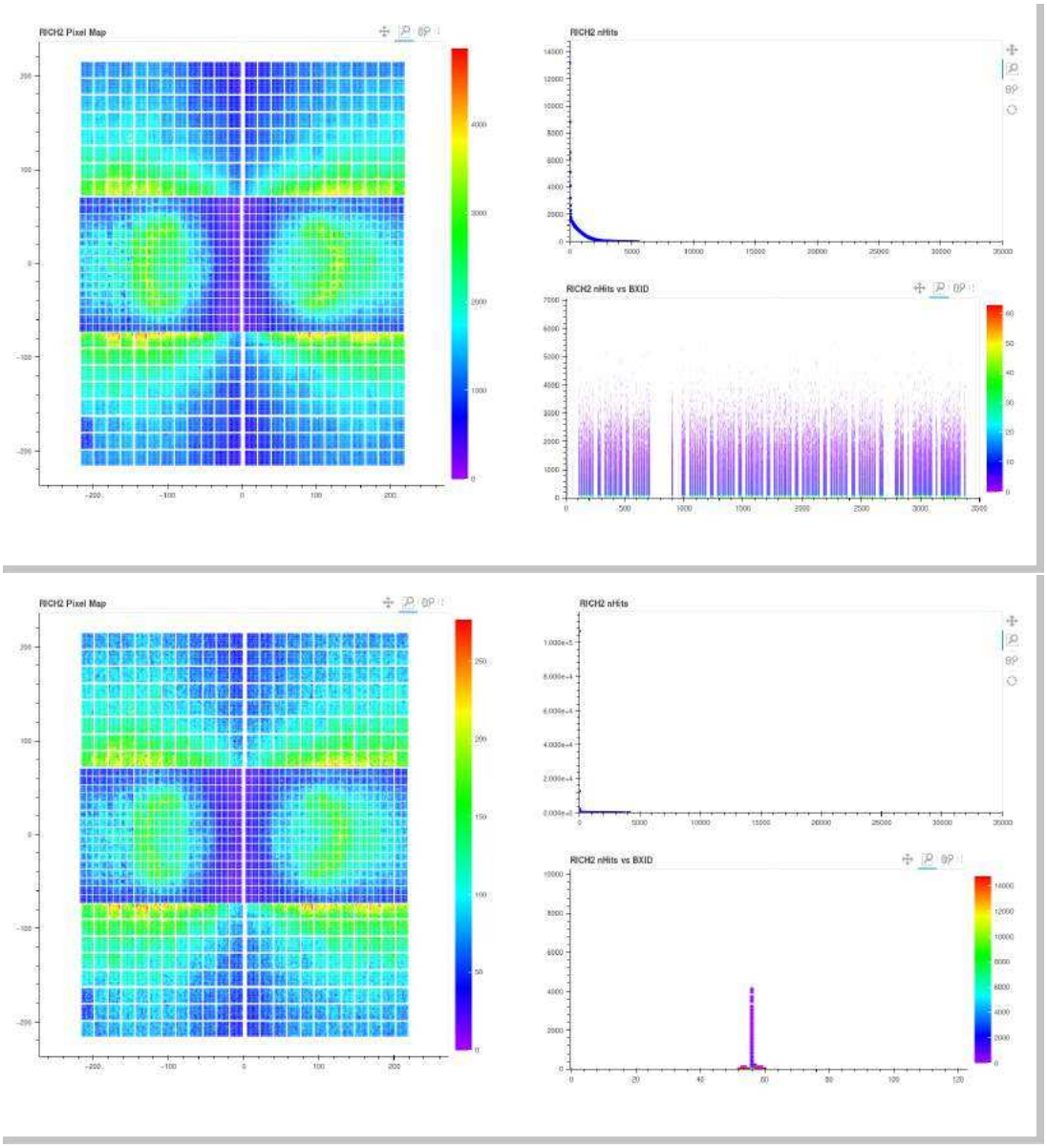


Figure 4.36: Improved timing in RICH2 detector. Top: Run with Random Trigger on bunch-bunch collisions. The BXIDs with activity are visible in the bottom-right plot. Bottom: Run acquired with Period Trigger on isolated bunch-bunch BXID.

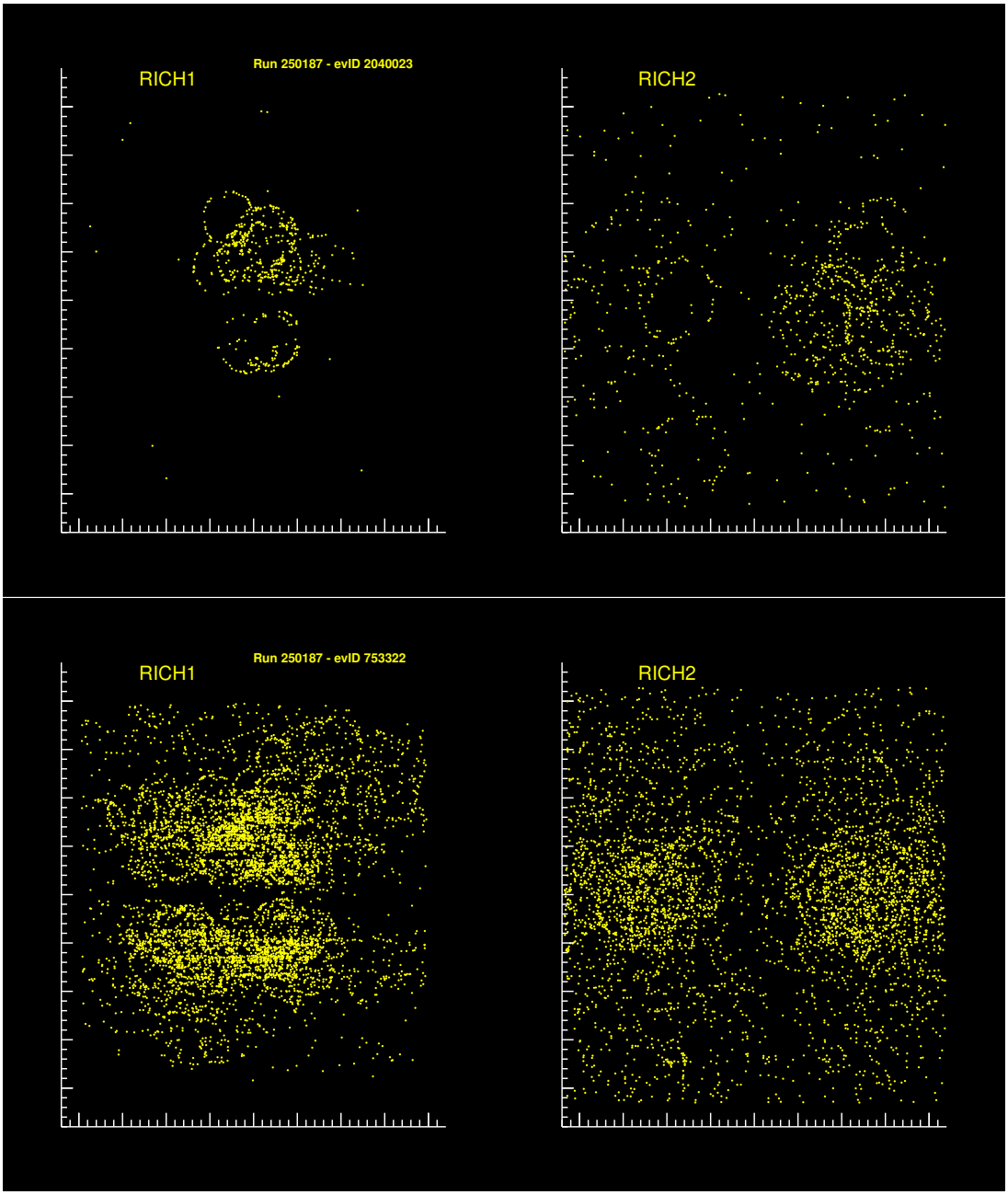


Figure 4.37: Single event visualization in a low luminosity environment (top) with $\mu = 1.1$ and a more standard Run3-luminosity with $\mu \sim 5$ (bottom). The pixel size is enlarged for better visualization.

4.5 Luminosity Monitoring

To perform luminosity leveling at IP8, LHC needs a precision of less than 5% on real-time μ_{vis}. For Runs 1 and 2, this measurement was primarily determined by measuring the transverse energy deposition in the calorimeters at the hardware trigger stage. For Run 3, LHCb switched to a software-only trigger system, removing the hardware-based system in favor of increased flexibility. In addition to the obvious computing challenges, it implies that a new solution is required for the control of

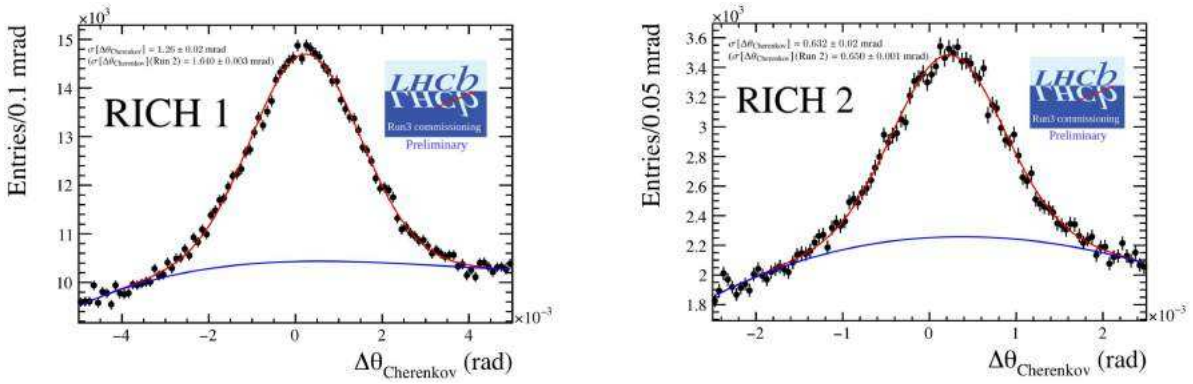


Figure 4.38: The Cherenkov angle resolution obtained with Run3 data at the end of 2022. The resolution for the Run2 RICH system is also reported in the plots. Despite the early stage of data-taking and being a preliminary results, the RICH1 already performers better than the Run2 version and RICH2 is approaching the expected performance.

luminosity and a fast, reliable feedback system for LHC. A dedicated detector, called PLUME (Probe for LUminosity MEasurement), has been installed upstream of VELO and is now being tested on real collisions since May 2022. In order to level the luminosity, PLUME is the main LHCb luminometer and provides real-time μ_{vis} measurements to LHC.

Relative luminosity measurement

The main characteristics of a luminosity counter are linearity (response vs μ_{vis}) and stability (efficiency constant over time). The number of VELO reconstructed tracks and vertices exhibited the best offline precision in Runs 1 and 2, other luminometers have been used for calibration, systematics, and cross-checks [66]. In Run 1, 1.16% accuracy was achieved and Run 2 data are being analyzed. In Run 3, a study of the luminosity-proportional properties of many observables is being carried out in order to utilize them as luminometers. For instance, PLUME coincidences, VELO track and vertices counts, energy deposited in the calorimeters, and the occupancy of various detectors is being investigated. The requirement of many counters is justified by the possibility to cross-check their stability in time and to correct for various efficiency dependences and evaluate systematics.

In the LHCb dataflow, two kind of luminosity determination can be defined:

- Online Luminosity
  - needed for luminosity leveling,reporting of quantities to both LHC and LHCb (center and width of the luminous region, beam shapes, integrated and delivered luminosity)
  - hierarchy of counters: the main one (PLUME) provides μ, and the rest are used for monitoring, cross-checking and back-up for PLUME
  - requires a precision: < 10%
  - the luminosity counters are available via ECS and are mainly used for online monitoring purposes since they are based on minimum bias conditions and operate independently of the high-level reconstruction sequence

- Offline Luminosity
  - needed for physics analyses
  - most stable counter (VELO tracks for pp collisions in case of Run 1 and 2) used for lumi determination, the rest for cross-checks, corrections and systematics
  - requires best possible precision (for pp reached < 2%, for PbPb ~4%)
  - counters are provided at the HLT trigger level using a 30 kHz random trigger shared as 70:15:10:5 in bb:be:eb:ee events, respectively, and aimed at providing high-precision offline luminosity measurement

4.5.1 Simulation HLT Studies

The investigation of possible variables which can be used as a counter for the RICH detectors has been carried out from simulations. Centrally MC samples have been produced to perform the studies for the lumi counters. The samples were generated for different values of inelastic pp collisions ν (without spillover) per bunch crossing, for both Magnet up and magnet down configurations:

- ν = 1.9 without spillover
- ν = 3.8 without spillover
- nominal ν = 7.6 without spillover
- ν = 14.3 without spillover
- ν = 21.5 without spillover
- ν = 28.6 without spillover
- ν = 35.8 without spillover

Also, samples with different positions of the luminous region along the X, Y, and Z axis are available. For the RICH detector, the variables more easily accessible are the number of hits in the detectors, and the response has been studied extensively to validate the linearity with respect to the luminosity in the configurations available.

Occupancy and linearity

As a first step, the occupancy of the detector has been evaluated as a function of the ν parameter. In Figs. 4.39 and 4.40, the occupancy of the RICH detectors is shown when increasing the number of collisions. The nominal working point for the LHCb Run 3 corresponds to a ν value of 7.6. Under this condition, the overall occupancy, especially for RICH1 central region, is under control and below the 30 % which is required for a good PID performance of the detector. It is interesting to observe the SIN noise in the pixels on the edges of MaPMTs when the ν reaches higher values, this behavior is evident in particular in the outer region of RICH1, where the MaPMTs with high SIN noise are installed.

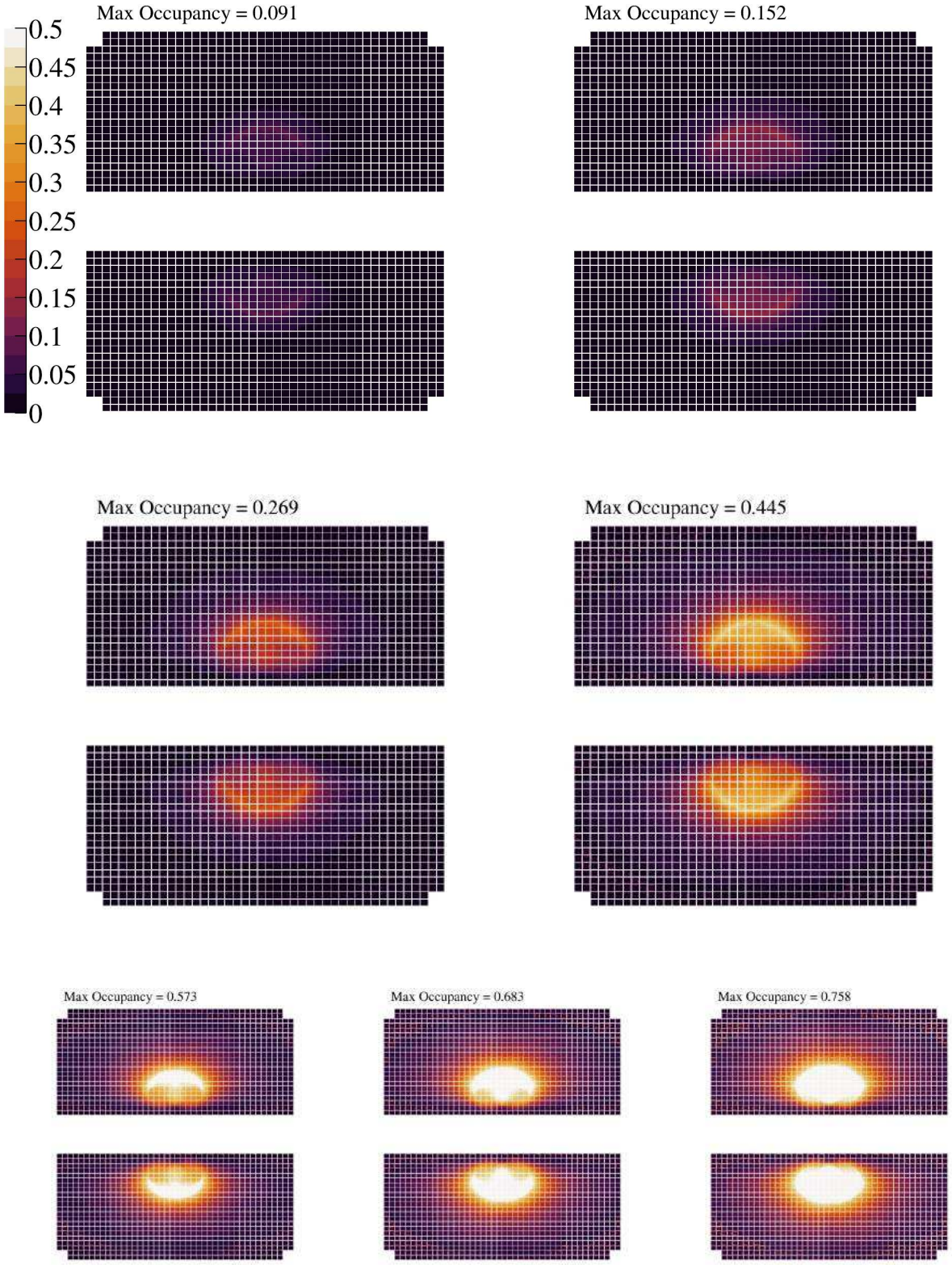


Figure 4.39: Occupancy of the RICH1 detector obtained from simulations with ν = 1.9, 3.8, 7.6, 14.3, 21.5, 28.6, 35.8. The Z-axis range is maintained constant to observe the increment of the occupancy and saturates at 50%.

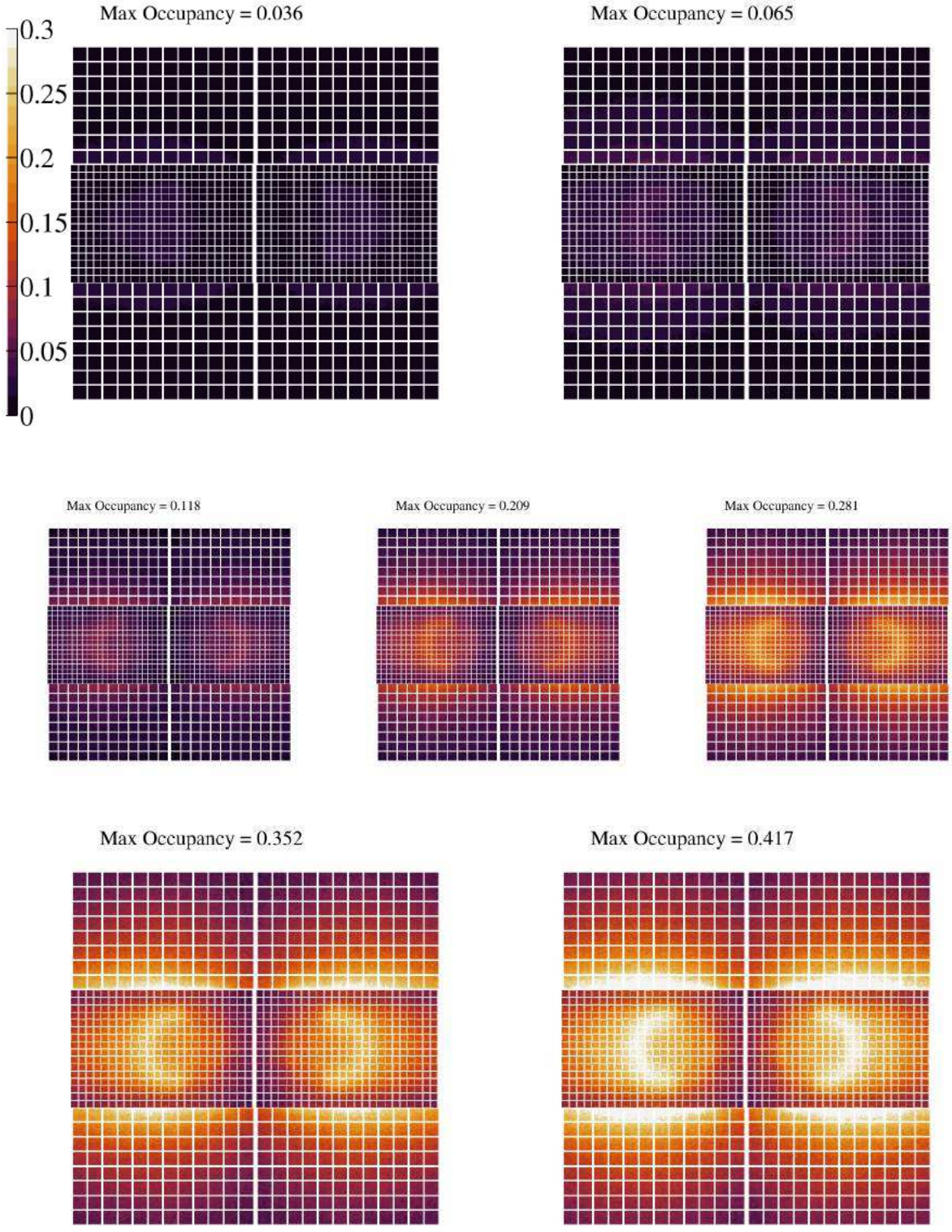


Figure 4.40: Occupancy of the RICH2 detector obtained from simulations with ν = 1.9, 3.8, 7.6, 14.3, 21.5, 28.6, 35.8. The Z-axis range is maintained constant to observe the increment of the occupancy and saturates at 30%.

The RICH candidates' HLT luminosity counters under consideration are the hit in the detector. As the main requirement, the counter should scale linearly with the luminosity, so with ν. The linearity is not expected to be satisfied in the regions where the probability of having more than one hit in a pixel per event is significant since, in this scenario, a saturation of the number of hits is observed. The reason for this loss of linearity is mainly due to the events where two or more photons hit the same MaPMT pixel in a short time interval and, in the digitization of the signal, only the first photon is counted. For each ν value, the distribution of the number of hits is studied in different regions of the RICH detectors and the mean of the distribution is used to evaluate the linearity as a function of ν. Different regions of the RICH system are considered and selected based on the average occupancy (Figs 4.41-4.42-4.43-4.44-4.45-4.46). As expected, the region where the linearity is lost is mainly the central region of RICH1, where the occupancy is higher. RICH2 has, on average a much lower occupancy, and its number of hits number is approximately linear with respect ν. In conclusion, the study confirms that RICH2, and in particular the low occupancy regions, are the ones that can be considered to obtain a stable luminosity counter, with the outer H-type region of RICH2 as the main candidate.

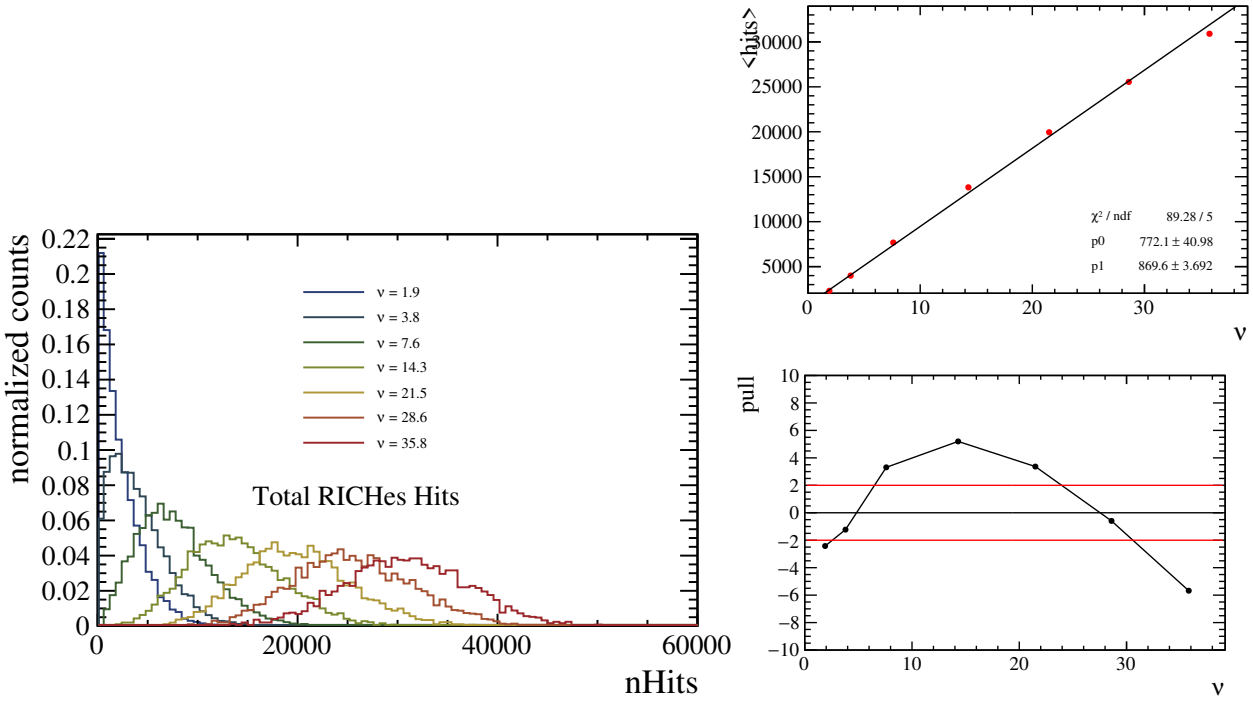


Figure 4.41: Linearity with respect ν of the whole RICH system hits, including both RICH1 and RICH2. On the left, the hits distributions for each ν value. On the right, the linear fit with the mean of the distributions and the pull.

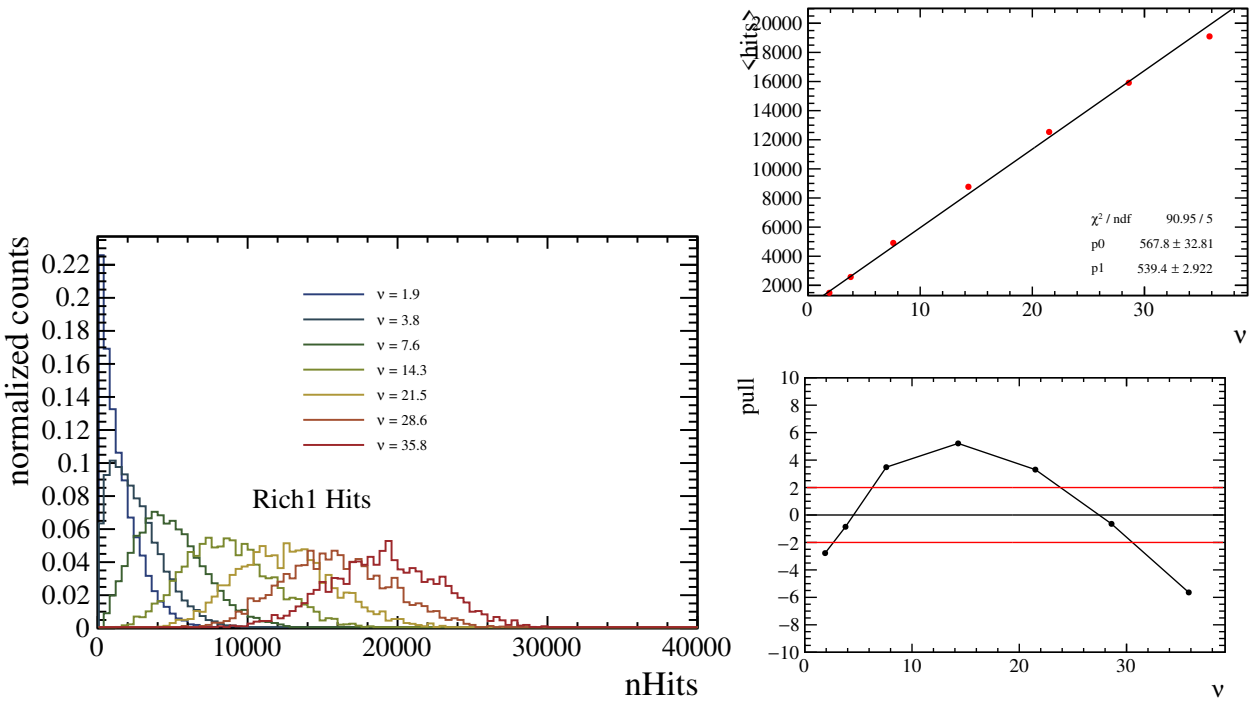


Figure 4.42: Linearity with respect ν of RICH1 hits. On the left, the hits distributions for each ν value. On the right, the linear fit with the mean of the distributions and the pull.

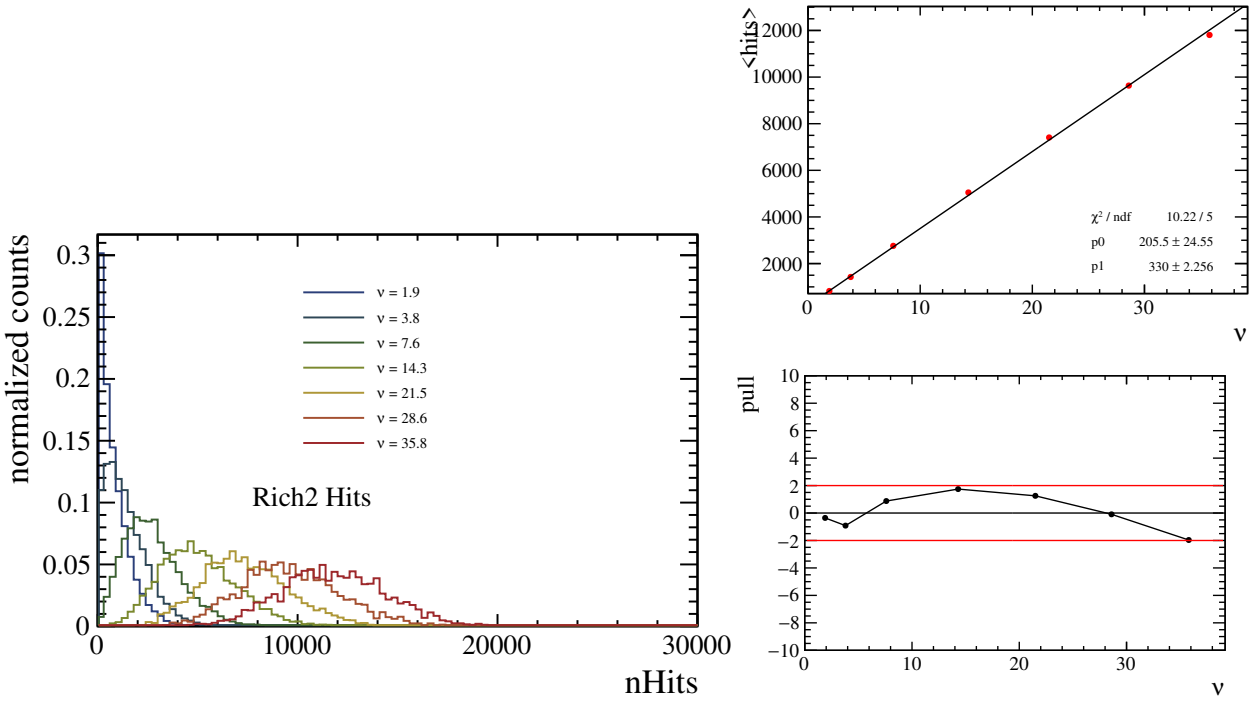


Figure 4.43: Linearity with respect ν of RICH2 hits. On the left, the hits distributions for each ν value. On the right, the linear fit with the mean of the distributions and the pull.

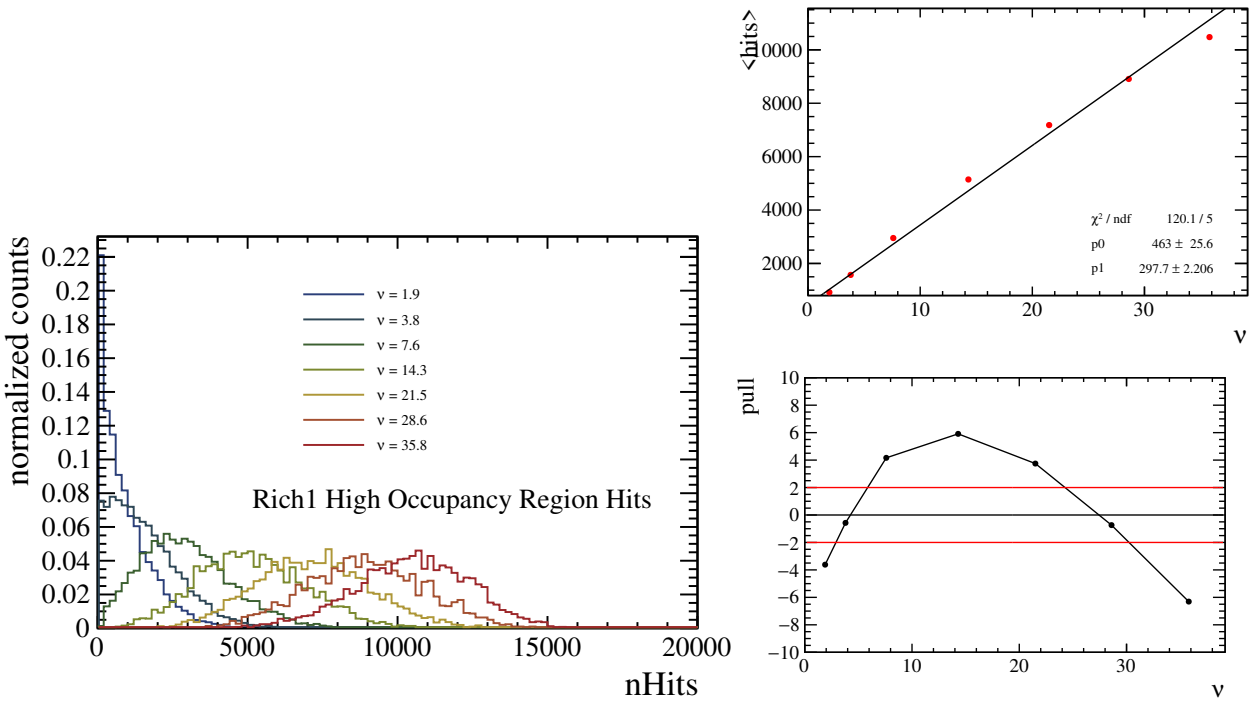


Figure 4.44: Linearity with respect ν of hits in the central RICH1 region, where the occupancy is higher. On the left, the hits distributions for each ν value. On the right, the linear fit with the mean of the distributions and the pull.

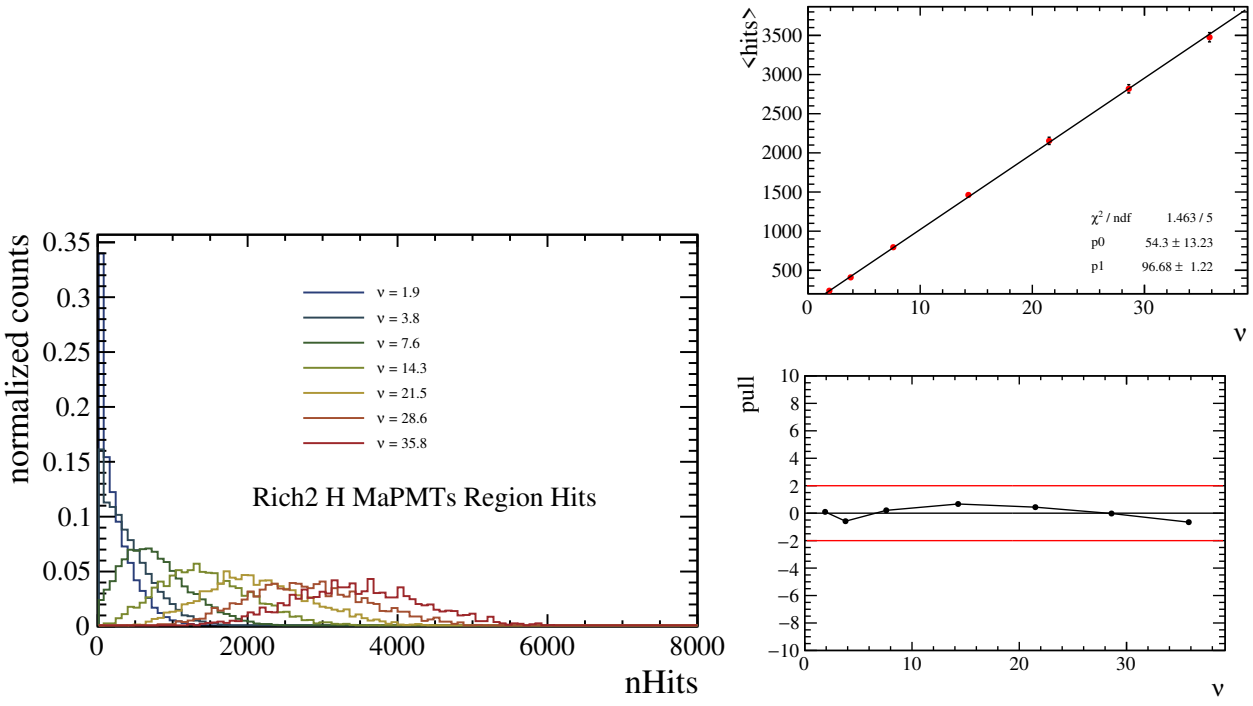


Figure 4.45: Linearity with respect ν of hits in the other regions of RICH2, populated with ECH. On the left, the hits distributions for each ν value. On the right, the linear fit with the mean of the distributions and the pull.

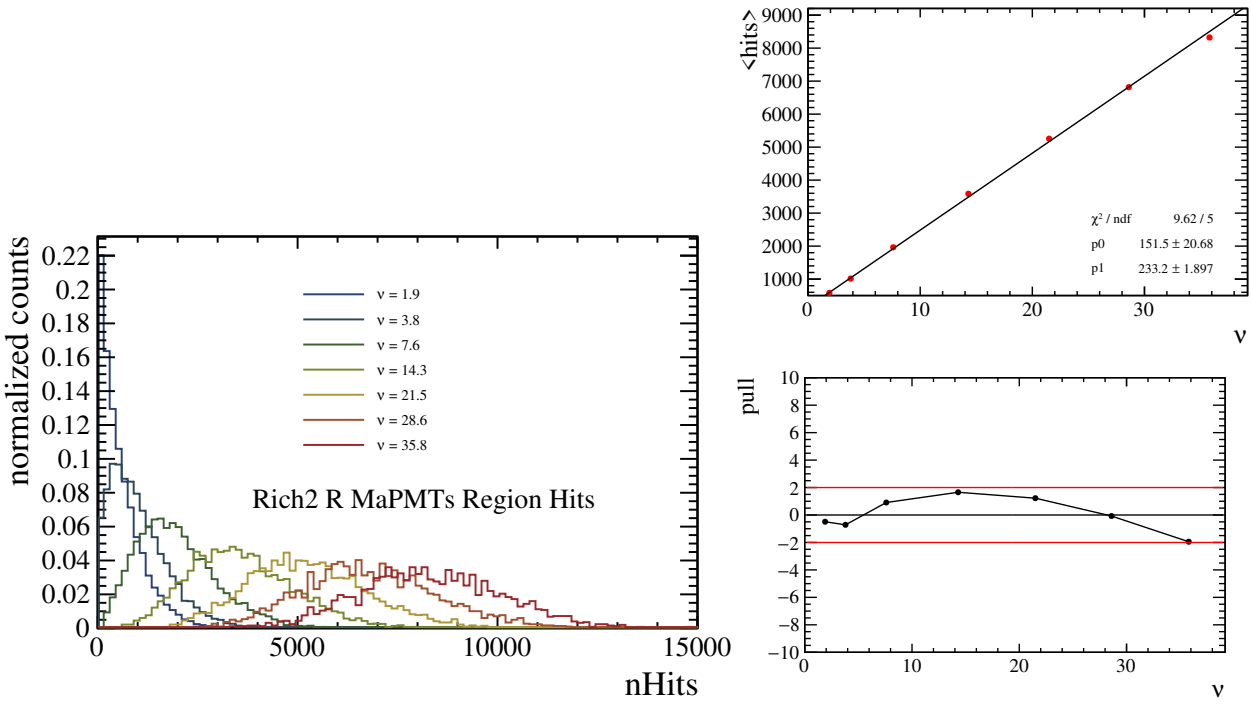


Figure 4.46: Linearity with respect ν of hits in the central regions of RICH2, populated with ECR. On the left, the hits distributions for each ν value. On the right, the linear fit with the mean of the distributions and the pull.

Magnet Comparison

The simulation has been produced with both magnet Up and Down configurations. The difference is studied in the nominal ν datasets and quantified with the variation of the occupancy in the RICH planes. The effect of the magnet is to separate the charged tracks along the X (horizontal) axis in the plane perpendicular to the beam direction. This is visualized as a tiny variation in the occupancy on the left and right sides of RICH1 and RICH2, but, overall, it does not significantly affect the distribution of the hits.

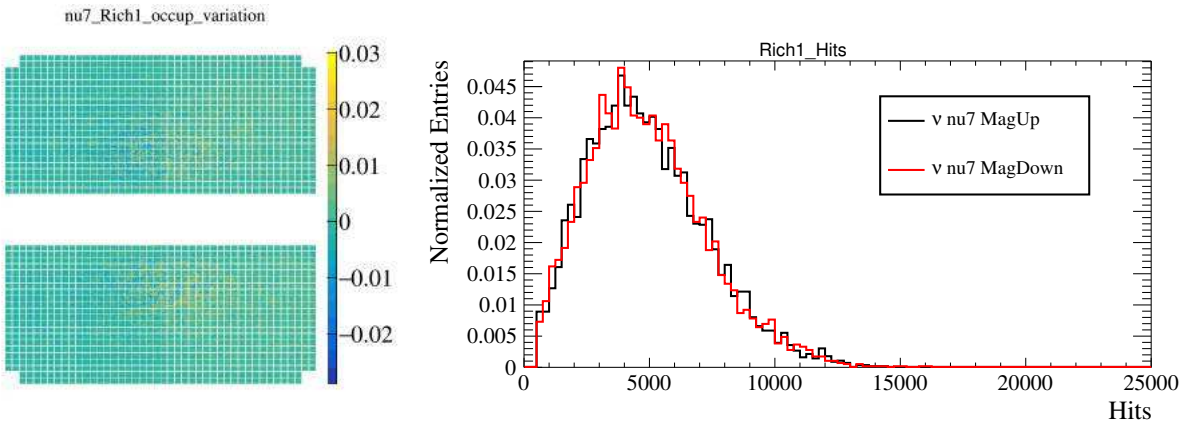


Figure 4.47: Effect of the magnetic field inversion in RICH1. On the left, the difference in the occupancy of the planes in each pixel. On the right, the normalized distribution of the hits.

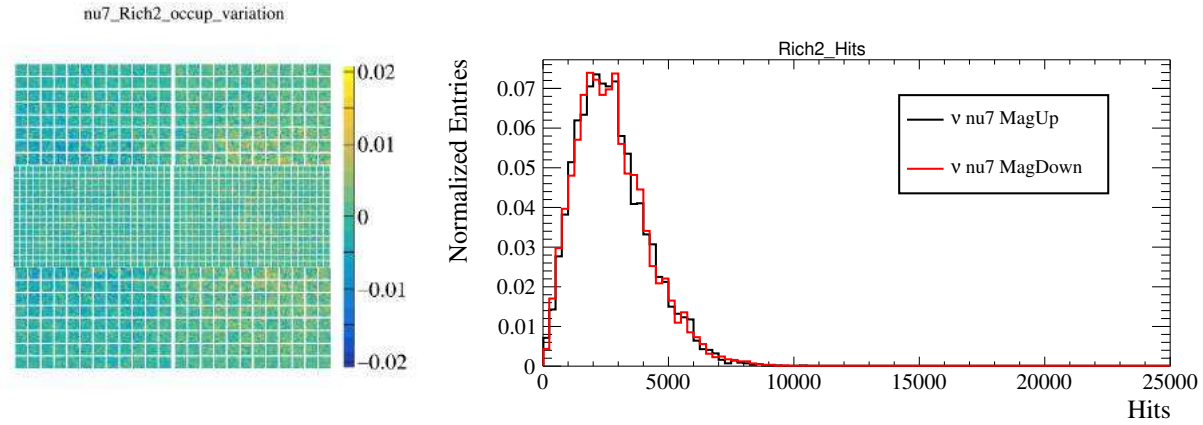


Figure 4.48: Effect of the magnetic field inversion in RICH2. On the left, the difference in the occupancy of the planes in each pixel. On the right, the normalized distribution of the hits.

Luminous region variation

Simulation samples have been produced also varying the luminous region in the X-Y plane and Z direction. Along the beam direction, the variation is in the range $[-120\ mm, 120\ mm]$, and along the X and Y axis the range $[-1.5\ mm, 1.5\ mm]$ is covered. The RICH1 detector, being closer to the interaction point, is more sensitive to the variation of the luminous region, as can be seen in Figs. 4.49-4.51-4.53, especially along the z (beam) direction, where also the hit distribution is modified. RICH2, on the contrary, is not affected significantly and has a more stable response in terms of uniformity of occupancy and hit distribution (Figs. 4.50-4.52-4.54).

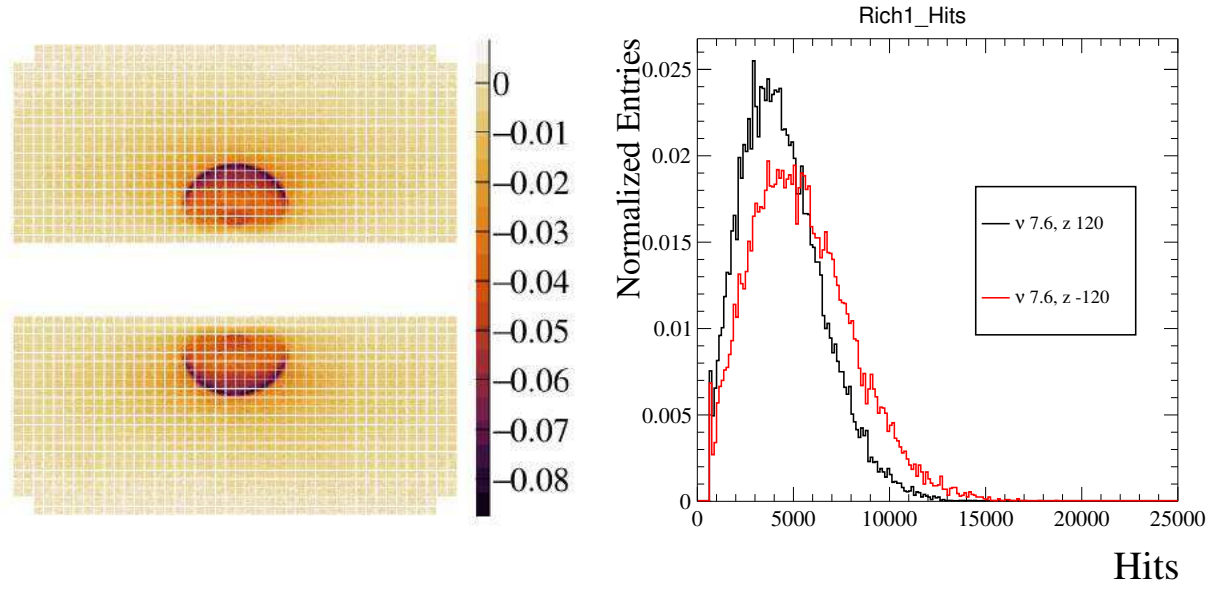


Figure 4.49: Variation of the occupancy in the RICH1 detector between $z = 120\ mm$ and $z = -120\ mm$. The z position of the luminous region can affect significantly the RICH1 response, also the hit distribution is modified.

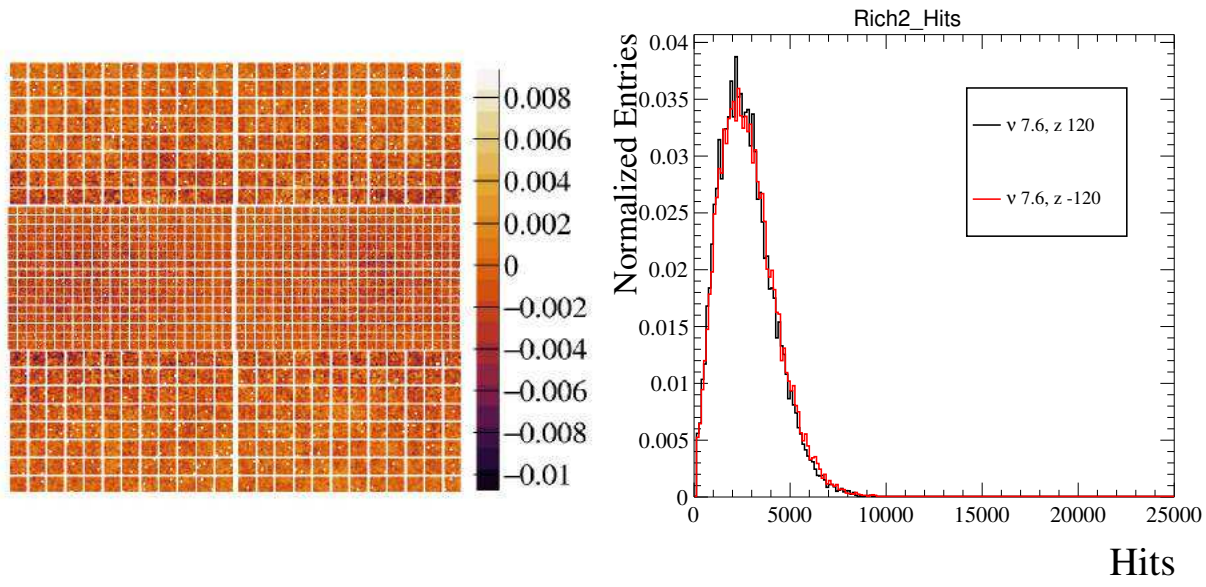


Figure 4.50: Variation of the occupancy in the RICH2 detector between $z = 120\ mm$ and $z = -120\ mm$. In the RICH2, the variation along the Z axis does not affect the response of the detector.

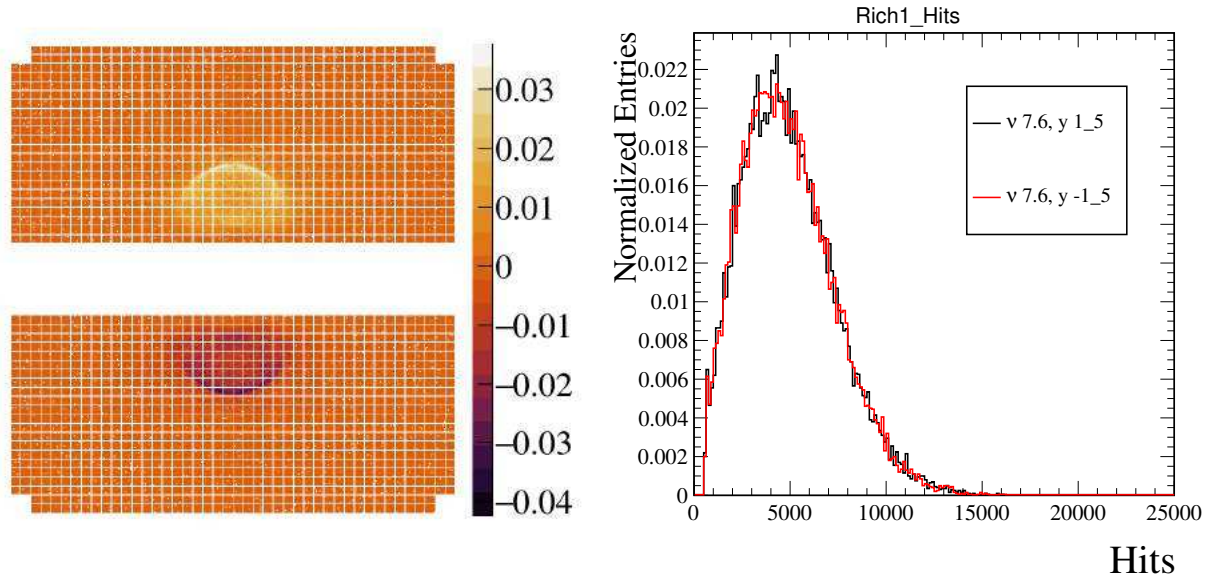


Figure 4.51: Variation of the occupancy in the RICH1 detector between $y = 1.5\ mm$ and $y = -1.5\ mm$. The effect is noticeable in the plane hitmap, the hit distribution is not affected.

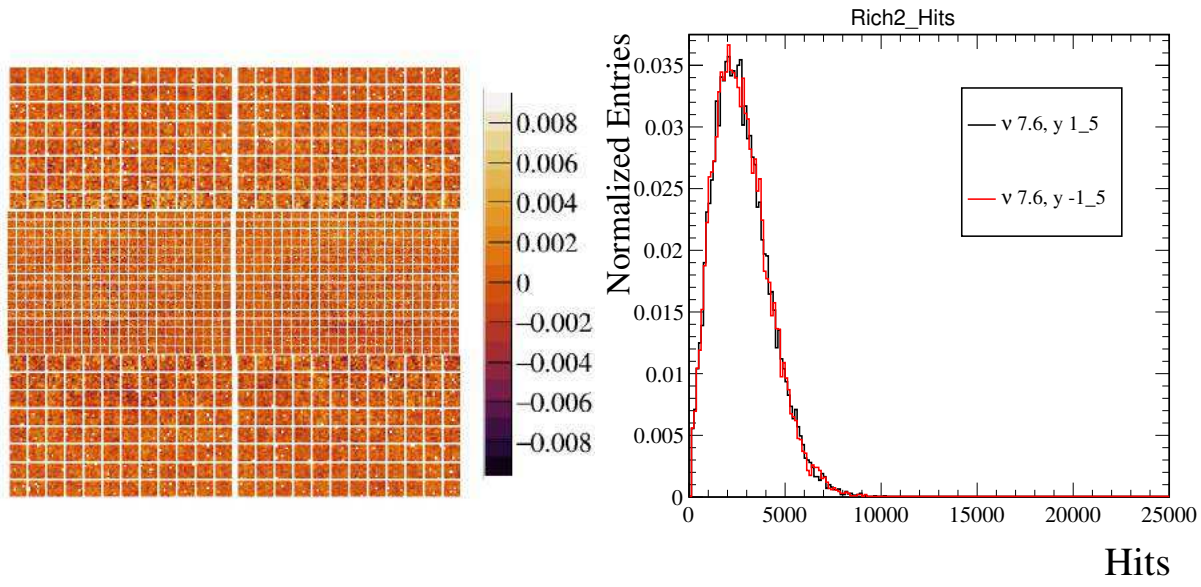


Figure 4.52: Variation of the occupancy in the RICH2 detector between $y = 1.5\ mm$ and $y = -1.5\ mm$.

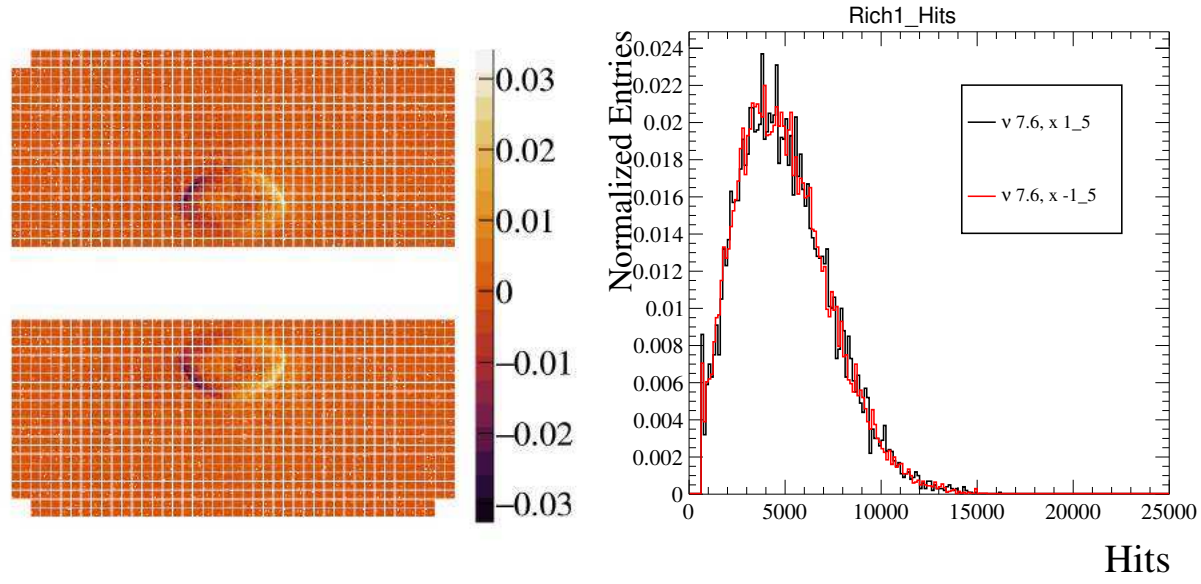


Figure 4.53: Variation of the occupancy in the RICH1 detector between $x = 1.5\ mm$ and $x = -1.5\ mm$.

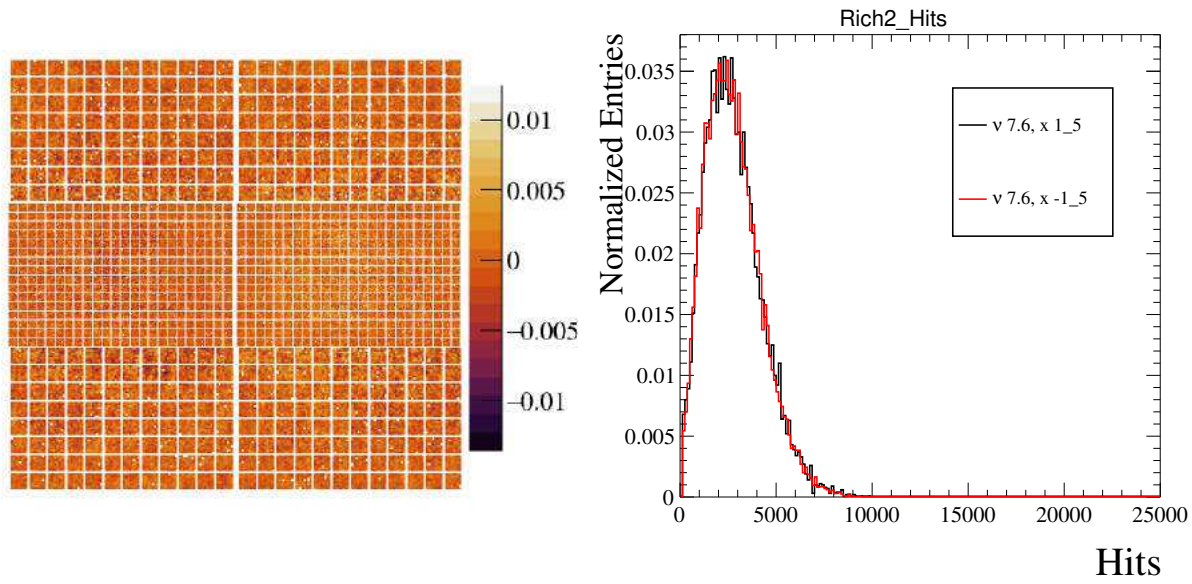


Figure 4.54: Variation of the occupancy in the RICH2 detector between $x = 1.5\ mm$ and $x = -1.5\ mm$.

4.5.2 RICH ECS luminosity counters

On average, the number of Cherenkov photons arriving at the RICH photon detection planes is proportional to the number of charged tracks in the event and therefore to the number of visible interactions in LHCb. The anode current of one MaPMT can be parameterized as:

$$i_A = N_\gamma\, \epsilon\, G\, e\, \nu_{bb} + \delta \tag{4.1}$$

where:

- N_γ is the average number of Cherenkov photons hitting one MaPMT
- ϵ is the photon detection efficiency given by the MaPMT geometrical efficiency (*i.e.* the active area fraction), the quantum efficiency and the photoelectron collection efficiency at the first dynode;
- G is the single photon gain;
- e is the electric charge;
- $\nu_{bb} = N_{bb} \times \nu_{\mathrm{LHC}}$ is the collision rate, where N_{bb} is the number of colliding bunches in LHCb and $\nu_{\mathrm{LHC}} = 11245$ Hz;
- δ is the contribution to the anode current given by other sources such as dark noise and beam-gas events that can be considered negligible with respect to the first term.

Given the presence of ν_{bb} and since $N_\gamma \propto \mu_{\mathrm{LHCb}}$, the anode current is a proxy for the LHCb instantaneous luminosity.

The RICH currents are constantly monitored using the ECS. The reading of the current is available for each channel of the CAEN modules and a single channel is connected to one PDM in the RICH1 or RICH2 plane. The photocathode current (I_{PK}, bias supply current) read from the MaPMTs is expected to have a close linear relationship with the number of photons hitting the photocathode and the anode current, under the assumption of the constancy of the MaPMTs gain. The relation is not straightforward if the MapMT gain and increase in not fully controlled. This is the case for the high occupancy region of RICH1 in the nominal Run 3 data-taking conditions. In particular, the gain and the operating voltage of the MaPMTs in the hot region of RICH1 must be chosen in order to not exceed an anode current of 100 μA.

Dynode biasing scheme

An increase in the current flowing through the dynode chain can change the potential difference between the dynodes. In particular, detailed studies were performed with SPICE models on the stability of the 11th and 12th dynodes. A reduction of the dynodes voltages implies a degradation of the collection efficiency at the anode and of the photoelectron multiplication. To mitigate this effect, operate with a stable gain, and preserve a known relation between the bias supply current and the anode current, a dynode bias mode has been implemented in RICH1. It consists in powering the last dynode (Dy12) in the MaPMT-R. Powering Dy12 implies the following relation between the power supply currents and the anode current:

$$I_A = I_{PK} - (V_{Dy12}/R\Omega) + I_{Dy12} \tag{4.2}$$

where V_{Dy12} is the constant bias supplied to the last dynode and R is the voltage divider resistance (typically 16/0.3 MΩ for one photon detector module that groups sixteen MaPMTs). At the cost of doubling the HV channels, it can be assured to operate MaPMTs in a controlled (stable gain) and safe way (below maximum ratings of 100 μA average anode current/MaPMT). But this comes with another advantage, having a stable gain and linear relation between I_{PK} and I_A allows to use the $I_{\text{LumiRICH}} = i_{PK} + i_{Dy12}$ currents as luminosity counters, since they are expected to scale linearly also with the luminosity delivered to LHCb in the region which is most affected by the occupancy of the events, the RICH1 central region. So, in principle, $I_{\text{LumiRICH}} \propto \mu \cdot f_r \cdot N_b$. The validation of the RICH currents as luminosity counters have been studied in various μ scans throughout the 2022 commissioning year. During certain fills, LHC leveled LHCb to reach different values of μ_{vis} following the real-time feedback from PLUME, the LHCb main luminometer. The RICH system took part to these μ scans with the following configurations for RICH1 (the Dy12 powering in RICH2 is not needed):

- Dy12 monitoring mode: the bias supply was applied to the photocathode, without the powering of the last dynode. The variables more sensitive to the luminosity variations in this configuration were the power supply currents I_{PK}
- Dy12 bias mode: the last dynode of the MaPMT-R is biased. The applied bias keeps the interstage voltage constant. If no bias voltage is applied, the voltage drops significantly for high photocathode currents. The constant voltage value is determined by measuring the interstage

voltage at the last dynodes without any light falling onto the photocathode. When the last dynode is biased, the power supply current is less sensitive to luminosity variations and it is not used anymore, instead, I_A and I_{Dy12} are expected to change sensibly. This HV recipe has been applied since July 2022.

The implementation of the last dynode powering follows this logic :

- OFF → STANDBY2 (LHC state: RAMP): Go STANDBY2, apply the voltage settings of the PHYSICS recipe to the photocathode channels, the dynodes remain off;
- STANDBY2 → READY (LHC state: SQUEEZE): Go READY, set the state of the photocathode channels to READY, read the monitored voltage value of the dynodes and switch on them at that value;
- READY → OFF (LHC state: INJECTION): switch off the dynodes, wait for them to be off, switch off the photocathodes

The HV values for the dynodes are set at every FLAT TOP by reading the dynode-monitored values at that point.

μ scans

In the first μ scan performed by LHCb, the luminosity monitoring variables used by the RICHes were the I_{PK} measurements. The target μ values were $\mu = (3.3, 2.3, 1.5, 1.1)$. LHC leveled the beams accordingly in P8, following the values provided by PLUME (Fig. 4.55). In Fig.4.56 the trends of the monitored RICH values are shown.

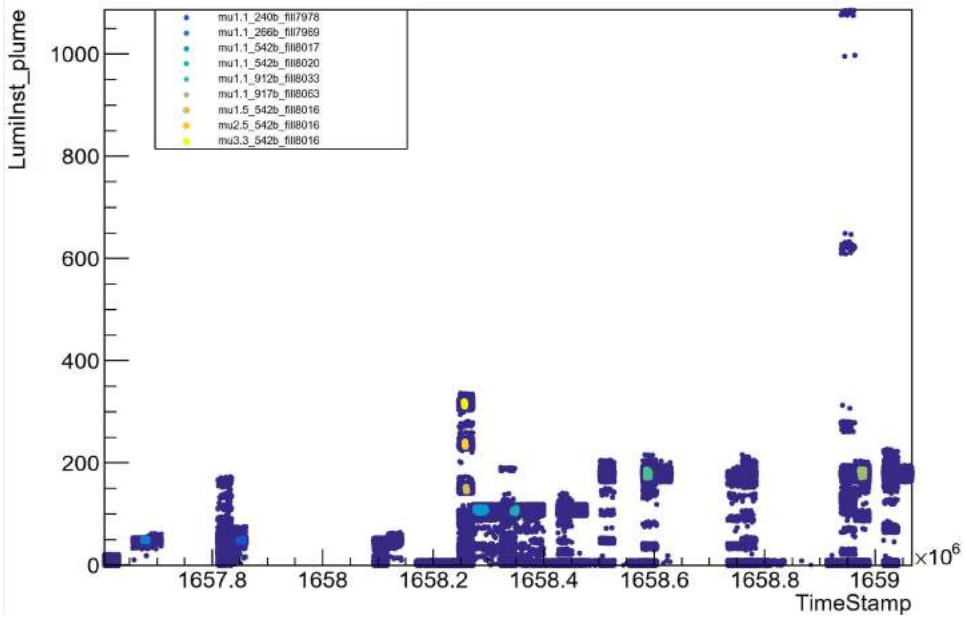


Figure 4.55: Luminosity provided by PLUME vs the Timestamp in the July data taking. The steps in the luminosity due to the μ scan are visible, compared to the default $\mu = 1.1$ which was set in that period. The luminosity is measured in a time interval of ~ 3 s. This leads to a proportionality of the luminosity value with the number of colliding bunches (reported in the legend).

Figure 4.56: Trends of the July μ scan. The last dynode was not powered yet. The variable more sensitive to the luminosity variation is I_{PK} (PK_imon in the legend). The voltage of the last dynode exhibits a drop of ~ 2 V at the highest luminosity reached in the scan. The voltages and currents are shown for 2 PDMs in the central region of RICH1. Also, the currents from the LLD detector positioned in RICH2 see the light from the pp collisions, as shown in the brown and yellow lines.

The powering of the last dynode was implemented for RICH1 in the following μ scan in August.

The current trends are sown in Fig.4.57.

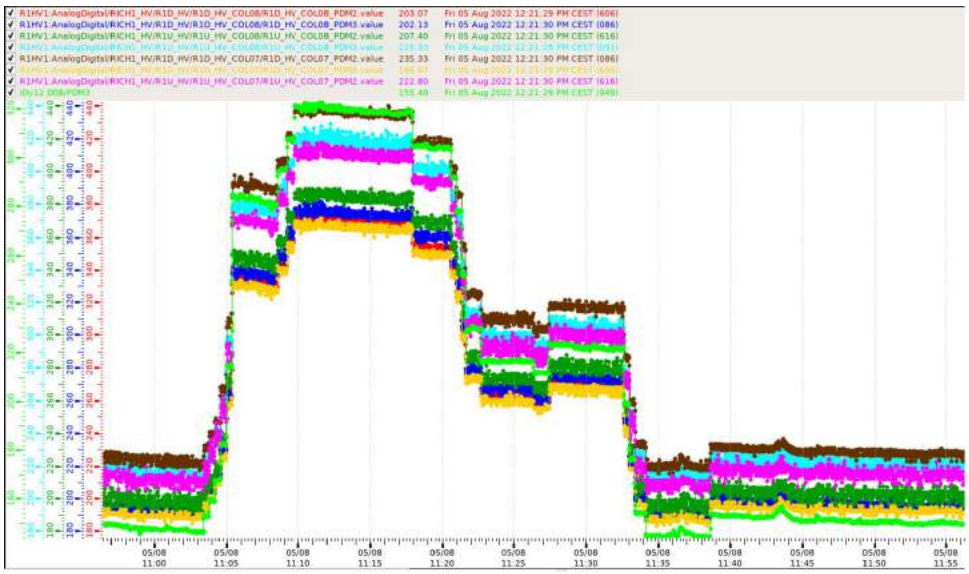


Figure 4.57: Trends of the August μ scan. The last dynode was biased. The variable more sensitive to the luminosity variation is I_{Dy12} and I_A. A value of $\mu \sim 2.6$ was reached between 11:10 and 11:17, $\mu \sim 1.7$ was leveled around 11:27. At the end the μ value was set to 1.1, at $\sim$ 11:40.

One last μ scan was taken at the end of 2022 with 1735 colliding bunches, before the LHC winter shutdown. Data were acquired with the RICH included in global data-taking and with the 6.25 ns time gating configuration. From this scan, it was possible to select bunch-crossing events requiring also the luminosity trigger to have unbiased datasets. Based on the simulation studies, the hits in the RICHes observed in the data acquired during the μ scan and the I_{LumiRICH} currents are compared to self-validate the ECS and HLT luminosity counters for the RICH detectors. In Figs.4.59-4.60 the linear fits of the I_{LumiRICH} proxy variable of one PDM in the high occupancy region of RICH1 against the hits from the μ scan in different detector regions (the same selected for the simulation study). The μ estimates are given by the PLUME detector and are overestimated by $\sim$ 10-20 % at high μ, due to the ongoing calibration of PLUME. However, the fit performance can be roughly compared to the ones in Figs.4.41-4.46. The luminosity range covered by simulation is much broader (reaching higher ν values), but the fit performance validates the excellent linearity of the RICH2 hits, on the other way, the number of hits seems to saturate in the high occupancy region of RICH1.

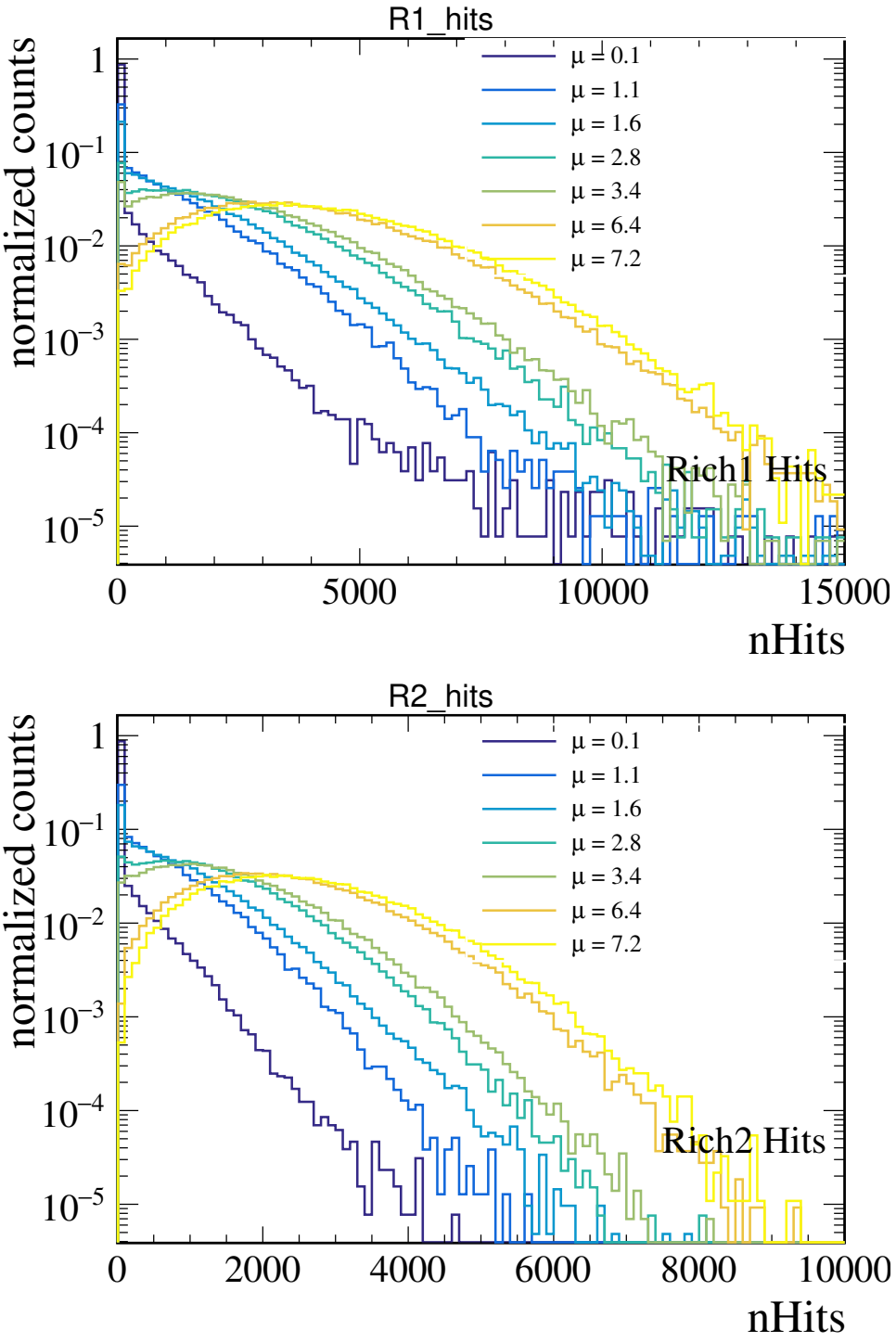


Figure 4.58: Hits distributions for RICH1 and RICH2 in the Runs acquired during the μ scan.

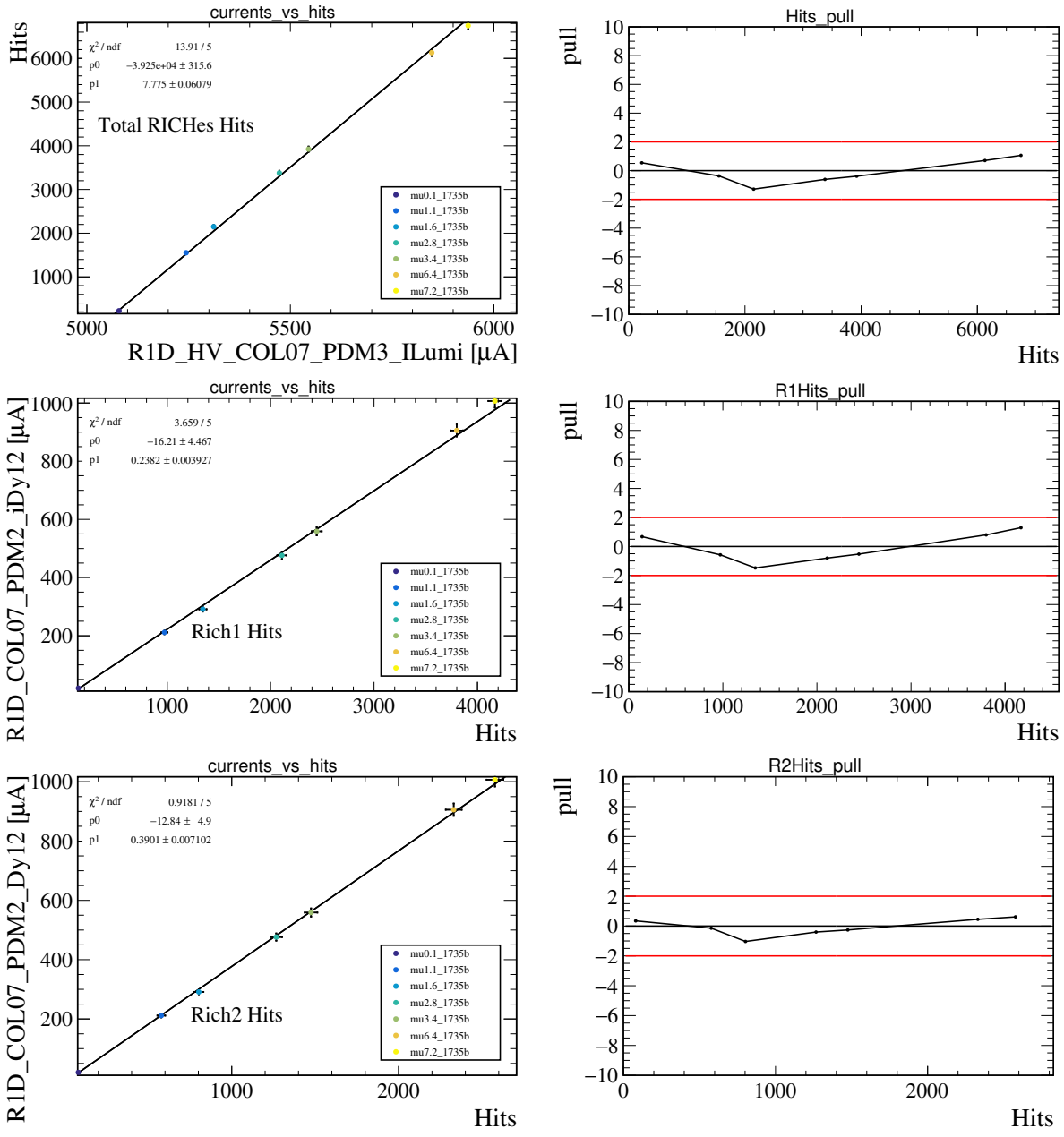


Figure 4.59: Linearity relation between one I_{LumiRICH} reading from an high occupancy region of RICH1 and the hits in the RICHes (top), RICH1 (middle), and RICH2 (bottom).

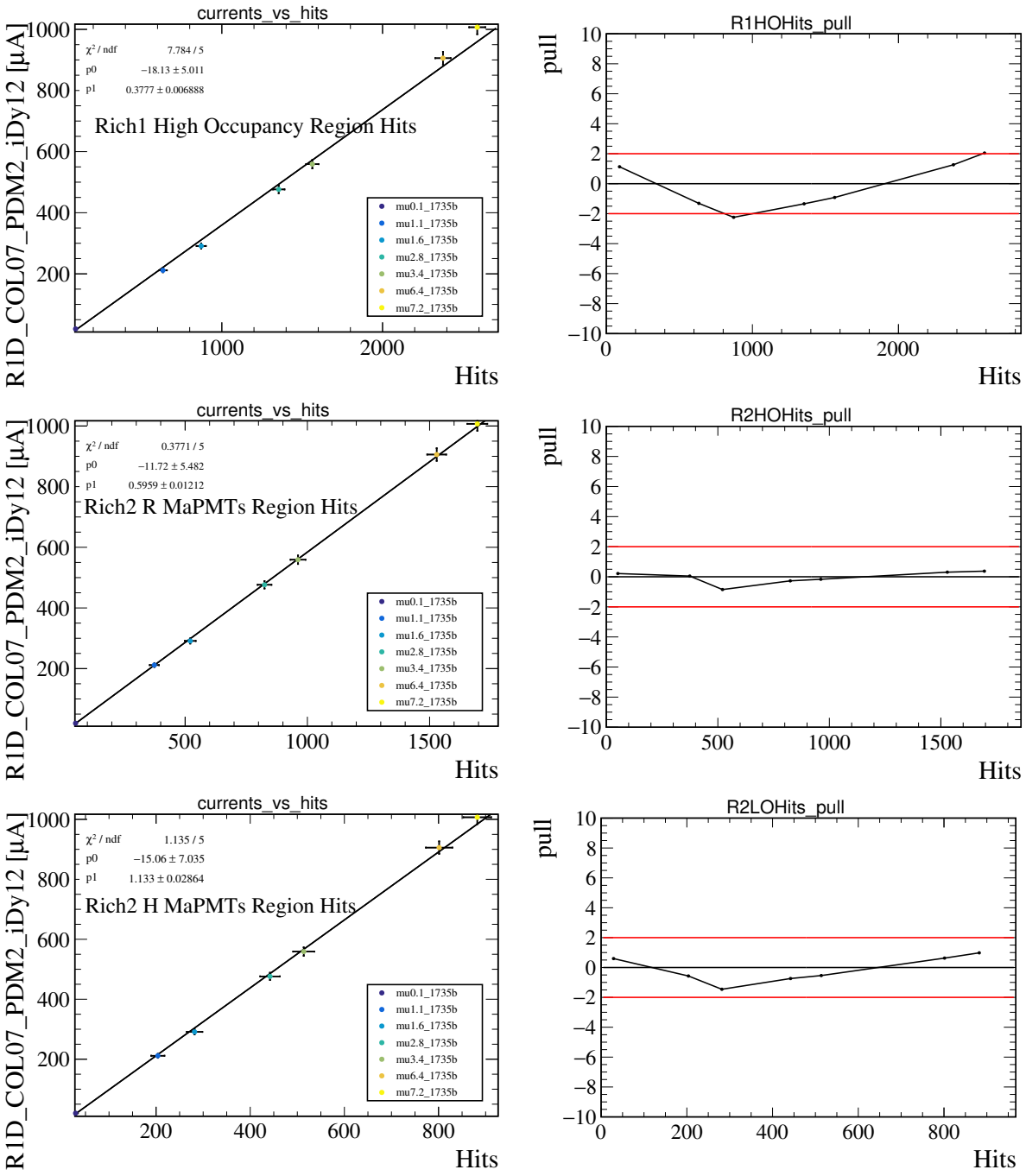


Figure 4.60: Linearity relation between one I_{LumiRICH} reading from a high occupancy region of RICH1 and the hits in selected regions of RICH1 and RICH2, the same used for simulation studies.

Chapter 5

Neural Network-based Gaussian Mixture Model for Particle Identification

Coherently with my major activity involving the RICH detector system upgrade for Run3, my analysis work depends abundantly on the PID of LHCb, and especially on the RICH detectors. Even during Run2, many analyses rely on the excellent PID performance of the RICH detectors. The PID information has been used extensively also in fixed-target analysis and the peculiar features of these data (considering also the low statistics of the available samples) often require customized approaches since the pp data might be not useful for fixed-target studies. Following the experience acquired by the collaboration with the antiproton production analysis [53], a novel strategy for the PID has been developed and applied to fixed-target data. I took part in the effort to develop and validate this new and fully data-driven technique based on a Neural Network-based Gaussian Mixture Model to parameterize the PID response of the RICH detectors. A paper has been already published following the development of this new technique [67]. In this chapter, the model is described in detail, from the calibration sample selection to the training process and validation of the results.

5.1 Particle Identification Modelling

The particle identification for charged hadrons is based on the information provided by the RICH detector reconstruction and it is obtained by combining many experimental observables. The datasets under study are mainly a mixture of π, K, and p, and to extract the relative abundances for each particle type it is necessary to predict the probability density function (pdf) of the PID variables. This task is usually challenging in high-energy physics experiments since the dependence of PID response has a complex dependence on the experimental features. The detailed simulation does not provide the required accuracy for the PID pdfs determination and it has been proven not to be completely reliable in such use cases due to the intrinsic imperfection of the simulated process or the prohibitive amount of simulated data needed to cover the full parameters space. For this reason, usually, data-driven methods are preferred when the statistics of the calibration samples is sufficient to perform the study with the required precision and to cover the same parameter space of the dataset under investigation. Unfortunately, this is not the case for the SMOG datasets collected in RUN2, due to the

lack of statistics of all samples but the pNe one at $\sqrt{s}$ = 68.5 GeV. Therefore, a novel approach has been developed [67]. With this approach, the PID classifiers are modeled using a Neural-Network-defined Gaussian Mixture Model which predicts the pdfs of such classifiers in the multivariate space spanned by the relevant experimental features that affect the detector response, such as the kinematic properties of the particle under consideration and the occupancy of the event.

5.1.1 Training dataset

As already mentioned, the standard approaches to extract the PID templates in similar analysis use the detailed simulation information or look at either specific calibration channels: $\overline{\Lambda} \to \overline{p}\pi^+$, $K^0_S \to \pi^-\pi^+$ and $\phi \to K^-K^+$. The latter strategy is usually used to evaluate the PID performance in pp collisions [68]. Unfortunately, the overlap of the features which affect the PID performance is small between pp and fixed target datasets. In particular, the occupancy variables such as the number of hits in the RICH detectors and also the PV determination (a few cm around the interaction point for pp collisions and about a meter for beam-gas collisions) differ substantially. Only a sufficiently large dataset has been collected during Run2 with beam-gas collisions: the proton-neon (pNe) sample at a center of mass of $\sqrt{s_{NN}}$ = 68 GeV recorded in 2017. This relatively large sample can indeed be exploited to reconstruct the calibration lines with significant statistics and obtain pure candidates with features compatible with the p-gas particle kinematics and event characteristics. The $\overline{\Lambda} \to \overline{p}\pi^+$, $K^0_S \to \pi^-\pi^+$ and $\phi \to K^-K^+$ lines are selected by the stripping applied to the pNe dataset(details are shown in Tab. 5.1). In addition to the stripping selection, further requirements are applied:

- on decay daughter (π, K or p):
  - $p \in [12, 100]\ GeV$
  - $p_T \in [400, 4000]\ MeV]$
  - $OWNPV_Z \in [-700, 100]\ mm$
  - hasRich
- on the event variables:
  - nBackTracks == 0
  - PUHits == 0 when $OWNPV_Z \in [-150, 100]\ mm$

The additional cuts are intended to optimize the overlap in the phase space and of the feature distributions of the daughter particles with the candidates from the pHe and pAr datasets. Moreover, the pNe dataset under consideration has been collected concurrently with pp data. This implies a contamination from pp collisions even after the requirement of beam-empty bunch crossing (corresponding to beam-gas events). The strict requirements on the nBackTracks and the PUHits cut in the region around the nominal pp interaction point are intended to reduce greatly this contamination.

Table 5.1: Calibration lines stripping selection. The selection is applied to obtain a high purity π, p and K samples from the calibration lines. This selection implies requirements on the quality of the vertex fit (χ^2_{vtx}) and position ($vtx_z < 2200\ mm$) as well as track quality criteria to be satisfied.

Line	Decay	Selection
π line	$K^0_S \to \pi^-\pi^+$	$p_\pi > 2\ GeV$, track $\chi^2/ndf < 5$, $\chi^2_{IP} > 25$, $M_{\pi\pi} < 1\ GeV$, $\chi^2_{vtx} < 16$, $vtx_z < 2200\ mm$, $\lvert M_{\pi\pi} - M(K^0_S)\rvert < 50\ MeV$, Λ veto
p line	$\overline{\Lambda} \to \overline{p}\pi^+$	$p_{\pi,p} > 2\ GeV$, track $\chi^2/ndf < 5$, $\chi^2_{IP} > 25$, $M_{\pi p} < 1.5\ GeV$, $\chi^2_{vtx} < 16$, $vtx_z < 2200\ mm$, $\lvert M_{\pi\pi} - M(\Lambda)\rvert < 25\ MeV$, K^0_S veto
K line	$\phi \to K^-K^+$	$p_K > 2\ GeV$, track $\chi^2/ndf < 5$, on tag K : ghostprob< 0.025 and $ProbNN_K > 0.75$, doca $\chi^2 < 10$, $\chi^2_{vtx} < 16$, $vtx_z < 2200\ mm$, $\lvert M_{KK} - M(\phi)\rvert < 20\ MeV$

5.1.2 K^0_S and $\overline{\Lambda}$ lines

Considering the K^0_S, the stripping selects a pure pion sample using only kinematics constraints. Only few % of the sample is background as can be seen by the invariant mass $M_{\pi\pi}$, see Fig.5.1. Also, the background is further reduced by the application of the Armenteros-Podolanski selection [69]. The same procedure is applied to the $\overline{\Lambda}$ decay to select protons since, also in this case, the background is almost negligible.

5.1.3 ϕ line

For the ϕ calibration line, the background is not negligible and the sWeights method [70] is applied to extract the signal component. The fitted distribution of the reconstructed ϕ invariant mass is shown in Fig.5.3. The signal component is described by a Voigtian, a convolution between the theoretical Breti-Wigner of the ϕ resonance and a gaussian which models the detector resolution, and the background by a third-order Chebychev polynomial. The pdfs resulting from the binned fit are then used to calculate the sWeights through the COWs method [71] which is preferred to the standard sPlot technique since it is more accurate. The weights profile is extracted and shown in Fig.5.4. The variable to evaluate the weighted sample is also shown, the signal and background components are obtained by applying the weights.

5.2 Gaussian Mixture Model

The concept behind the employed method is to describe the marginal pdf of the LHCb PID classifiers in terms of a Gaussian Mixture Model:

$$x_p \sim \sum_{j=1}^{N_{g,p}} \alpha_{j,p}(\underline{\theta})\mathcal{G}(x, \mu_{j,p}(\underline{\theta}), \sigma_{j,p}(\underline{\theta})) \tag{5.1}$$

where the index p refers to the particle species, $\mathcal{G}$ is a Gaussian distribution with mean μ and covariance matrix parameters σ, and N_g is the number of Gaussian functions in the model. The whole

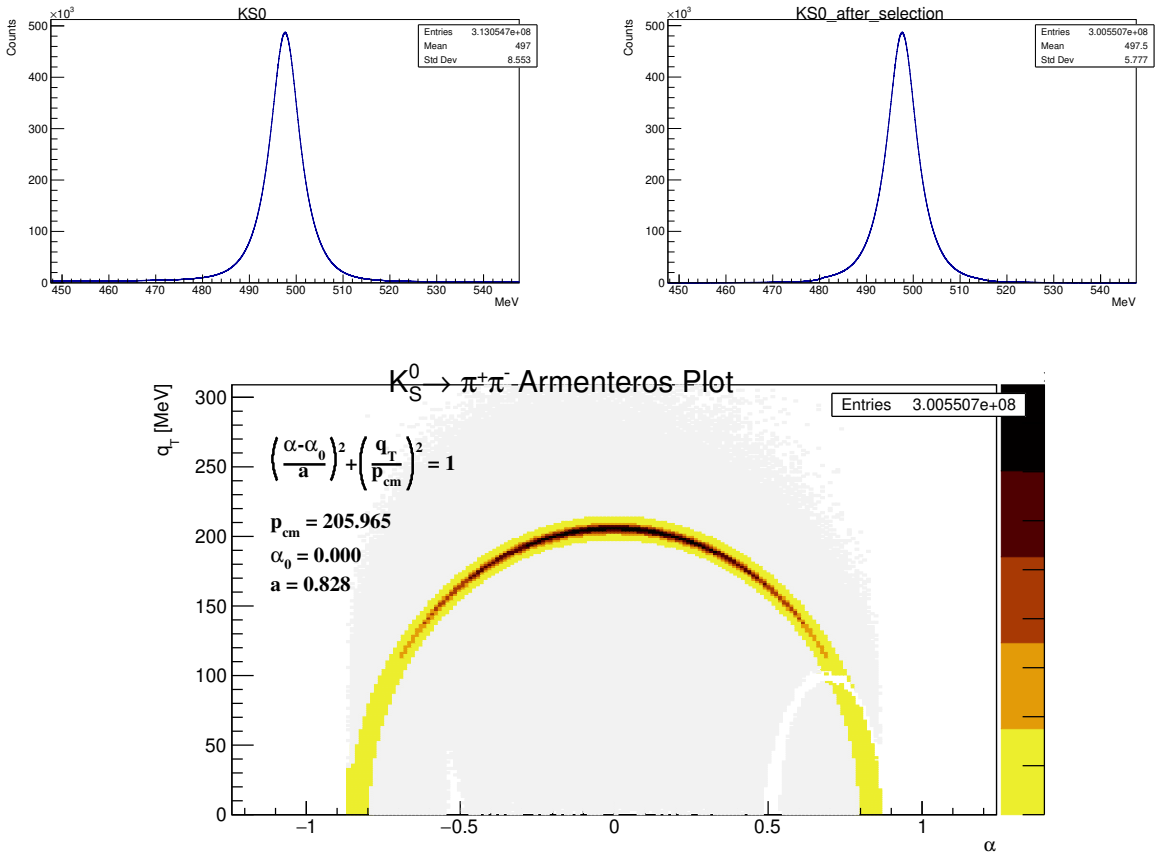


Figure 5.1: On top, $M_{\pi\pi}$ invariant mass before (left) and after(right) the Armenteros selection. In the bottom plot, the Armenteros plot is shown, and the candidates accepted by the selection are highlighted.

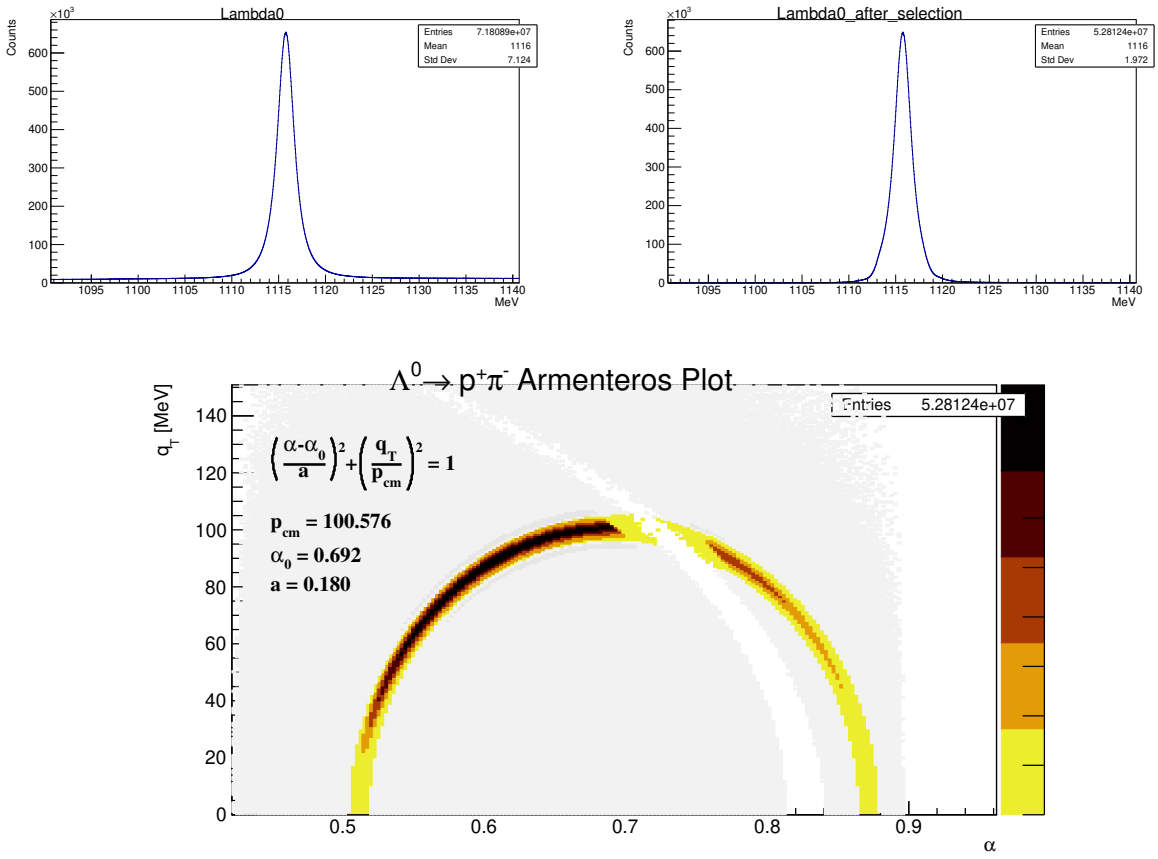


Figure 5.2: On top, $M_{\pi p}$ invariant mass before (left) and after(right) the Armenteros selection. In the bottom plot, the Armenteros plot for Λ^0 decay.

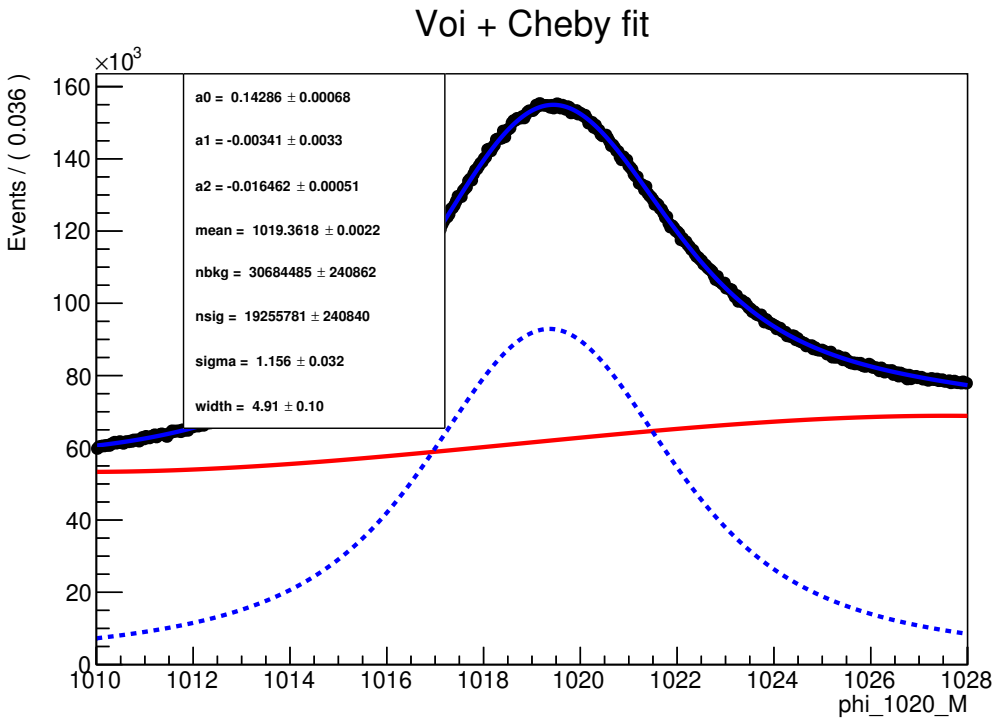


Figure 5.3: KK invariant mass overlayed with the fit to extract the signal component.

model depends on a set of features $\underline{\theta}$. The application is intended on a bi-dimensional distribution of PID classifiers (for example $\mathrm{DLL}_{p,\pi}$ and $\mathrm{DLL}_{p,K}$) Therefore, their distribution is modeled according to Eq. 5.1 replacing the Gaussian with a two-dimensional multivariate normal distribution:

$$\underline{x}_p \sim \sum_{j=1}^{N_{g,p}} \alpha_{j,p}(\underline{\theta}) \frac{\exp(-\frac{1}{2}(\underline{x}_p - \underline{\mu}_{j,p}(\underline{\theta}))^T \Sigma_{j,p}^{-1}(\underline{\theta})\,(\underline{x}_p - \underline{\mu}_{j,p}(\underline{\theta})))}{2\pi\sqrt{\det(\Sigma_{j,p}(\underline{\theta}))}} \tag{5.2}$$

where the vectors $\underline{x}$ represent the two PID classifiers ($\mathrm{DLL}_{p,\pi}$, $\mathrm{DLL}_{p,K}$) and Σ is their covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}. \tag{5.3}$$

The parameters of the model (α, μ, Σ) are predicted by a set of Multi-Layer Perceptron (MLP) neural networks. These neural networks are trained using the calibration lines described above from the pNe sample. During the training process, the loss function is defined as the negative log-likelihood of the data evaluated on the predicted pdfs:

$$\mathscr{L} = -\sum_{i=1}^{n_p} w_i \log\left[\sum_{j=1}^{N_{g,p}} \alpha_{j,p}(\underline{\theta}_i)\mathscr{G}(x_i, \mu_{j,p}(\underline{\theta}_i), \sigma_{j,p}(\underline{\theta}_i))\right]. \tag{5.4}$$

where w_i is the weight associated to each event. The parameter is required in case of the Φ line to take into account the sWeights.

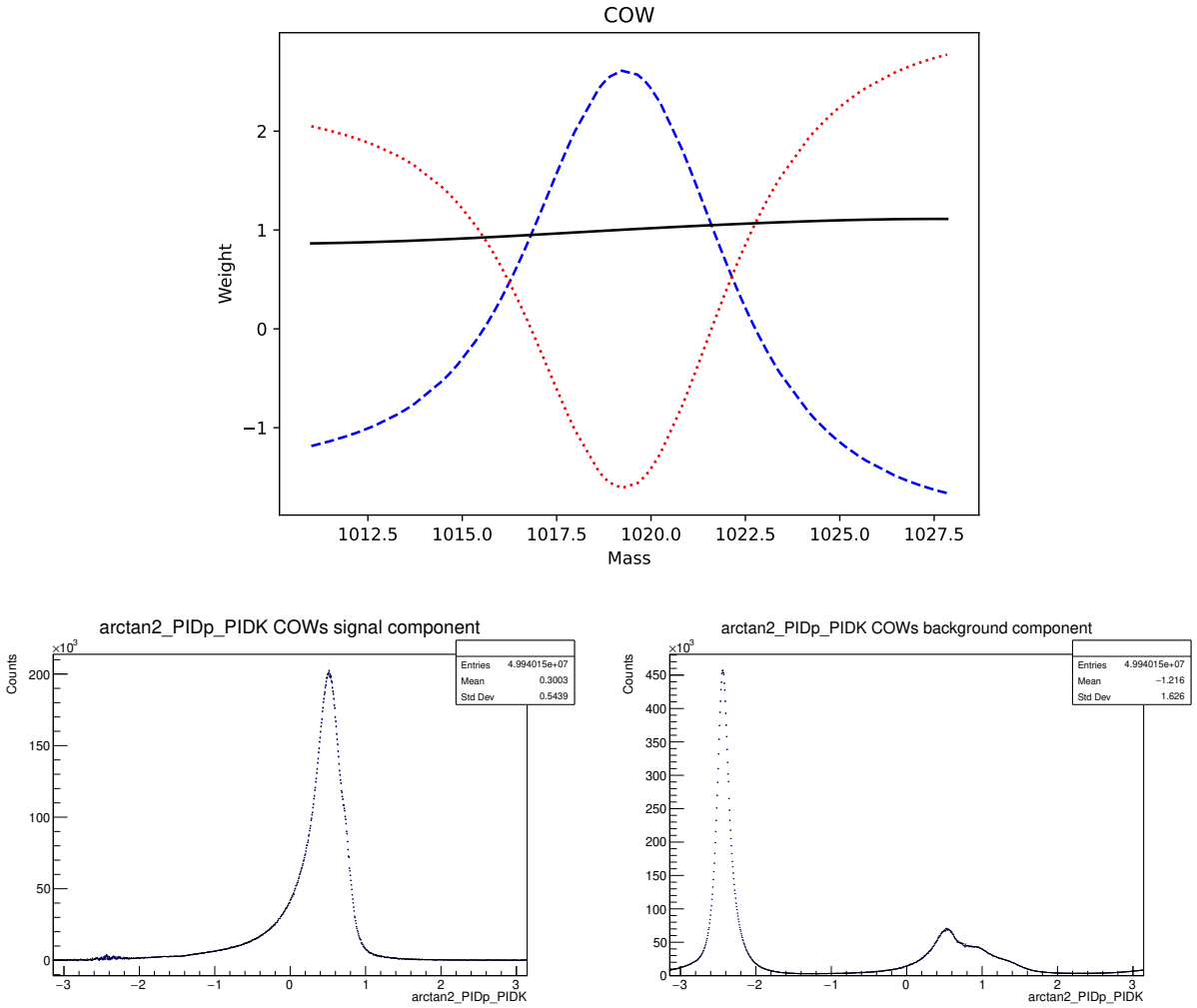


Figure 5.4: Top: the profile of the sWeights for the signal (blue) and the background (red) component as a function of the invariant mass. Bottom: a combination of the PID variables plotted after applying the signal (left) and background (right) sWeights. For the signal component, only the K contribution survives with negligible contamination from π. The background is dominated by π and p as expected.

5.2.1 Neural Network structure

The aim of the network is to predict the probability density function of the PID classifiers given a set of specific features $\underline{\theta}$ of the particle and the event. The dimensionality of the method can be extended to more the one dimensional, depending on the number of PID classifiers. In this application, the model predicts a 2D pdf distribution in the PIDp and PIDK space. The 2D pdf is a function with an exponential-like tail and be described accurately by a large enough number of multivariate normal distributions. The parameters to predict for each multivariate normal distribution are:

- the mean μ
- the covariance matrix

additionally, the normalization α of each multivariate normal distribution which composes the gaussian mixture model is also predicted by a neural network. In total, four sub-networks predict the four parameters (mean, three covariance matrix values and normalization coefficient) require to construct the GMM and allow the negative loss function evaluation. The sub-nets are made of 2 dense hidden layers, both $tanh$ activated, and one final external layer to match the correct dimensions. The hyperparameters to be set for each sub-net are the number of multivariate normal distributions that compose the GMM, N_G, and the number of nodes in the hidden layers. Overall, the dimensions of the layers are defined for each parameter:

- μ sub-net : $(N_f, nNodes) \cdot (nNodes, nNodes) \cdot (nNodes, N_G \times N_t)$
- σ_d sub-net : $(N_f, nNodes) \cdot (nNodes, nNodes) \cdot (nNodes, N_G \times N_t)$
- σ_c sub-net : $(N_f, nNodes) \cdot (nNodes, nNodes) \cdot (nNodes, N_G)$
- α sub-net : $(N_f, nNodes) \cdot (nNodes, nNodes) \cdot (nNodes, N_G)$

where σ are the values required to compute the covariance matrix using the Cholesky decomposition of the covariance, N_f is the number of input features and N_t is the number of PID target variables.

5.2.2 Training

The hyperparameters set for the training of the algorithm are:

- the number of iterations in the training process
- the batch size: the number of training events to be considered for each iteration of the minimization procedure
- the initial learning rate, which quantifies the step in the update of the network parameters for each parameter. The learning rate decreases linearly during the training by up to one order of magnitude smaller at the end of the process.

Table 5.3: Values of the input parameters for the model training of the three calibration channels.

Input parameter	$K^0_S \to \pi^-\pi^+$	$\overline{\Lambda} \to \overline{p}\pi^+$	$\phi \to K^-K^+$
Number of Gaussians	64	20	64
Number of NN nodes	128	128	128
Starting learning rate	10^{-3}	10^{-4}	$5 \cdot 10^{-6}$
Batch size	10000 events	10000 events	20000 events

The training is performed using a mini-batches gradient descent optimized with the RMSProp algorithm.

The GMM parameters are defined as

$$\begin{cases} \mu = I_\mu + f \cdot N_\mu \\ \sigma_d = log(1 + e^{I_{\sigma_d} + f \cdot N_{\sigma_d}}) \\ \sigma_c = I_{\sigma_c} + f \cdot N_{\sigma_c} \\ \alpha = log(1 + e^{I_\alpha + f \cdot N_\alpha}) \end{cases} \tag{5.5}$$

where I is the network initialization value and N is the output of the sub-net, both defined for each GMM parameter. The initialization I is chosen at the beginning of the training and the values are allowed to change during the backpropagation [72]. The initialization is set by sampling from the intervals:

- $\left[\langle \underline{x} \rangle - \frac{1}{2}\sqrt{\langle \underline{x}^2 \rangle - \langle \underline{x} \rangle^2}, \langle \underline{x} \rangle + \frac{1}{2}\sqrt{\langle \underline{x}^2 \rangle - \langle \underline{x} \rangle^2}\right]$ for the mean values $\underline{\mu}$, being $\langle \cdot \rangle$ the average value. For the ϕ calibration channel, the background-subtracted distributions are considered;
- $[-1, 1]$ for the correlation-related σ_c parameter;
- $\left[\frac{1}{10}\sqrt{\langle \underline{x}^2 \rangle - \langle \underline{x} \rangle^2}, \sqrt{\langle \underline{x}^2 \rangle - \langle \underline{x} \rangle^2}\right]$ for the widths-related σ_{d1} and σ_{d2} parameters;
- $[-1, 1]$ for the component weights α_j.

The component weights α values are also processed by a softmax function. The variable f is a flag that is used in the training to enable or disable the sub-nets and it's implemented to speed up the training that is split into two steps. In the first one, a raw estimate of the parameter is achieved by setting the flag $f = 0$, meaning that the sub-nets are not considered in the process and the few iterations only optimize the initialization values for each GMM parameter. In the second step, f is set to 1, the actual training happens and the weights and biases of the sub-nets change to improve the log-likelihood and minimize the loss function as defined in Eq.5.4. The covariance matrix is then obtained from the following triangular matrix :

$$L = \begin{pmatrix} \sigma_{d1} & 0 \\ \sigma_c & \sigma_{d2} \end{pmatrix}$$

and the covariance matrix is given by $\Sigma = LL^T$.

An example of the loss function minimization results is shown in Fig.5.5 for the three pNe calibration lines.

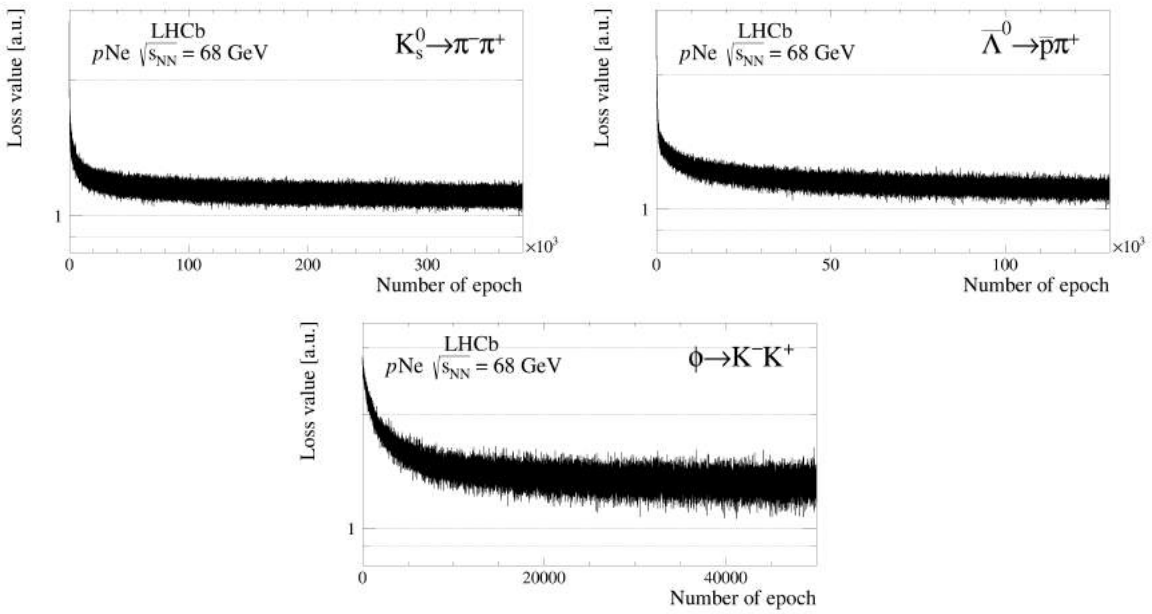


Figure 5.5: Evolution of the loss function with the number of iterations for the (top left) pion, (top right) proton and (bottom) kaon calibration channels.

5.2.3 PreProcessing

The features given as input to the neural network (NN) and the PID variables to predict are preprocessed to ease the learning process. Different transformations are applied between input features and target variables since they play a different role in the learning process. The preprocessing uses the scikit-learn open-source machine-learning library [73]. For the target PID features, a linear transformation is applied to map their values in the range $[0,1]$, but preserve the shape of the distribution, since it has to be predicted by the GMM algorithm. This is implemented using the MinMaxScaler function. The Quantile Transformer algorithm is applied to the input features to transform their pdf into a Gaussian Distribution with $\mu = 0$ and $\sigma = 1$. This transformation is feasible since their pdf is not required to be reproducible by the model. The preprocessing allows the NN layer for a standardized initialization and speeds up the numerical estimation of the derivatives in the loss function minimization procedure.

5.2.4 Overtraining Checks

The model is not meant for a classification problem and is not prone to overtraining by construction. Some checks are however needed to check that the pdfs modeling is smooth as a function of the input features since the overtraining would imply rapid oscillations of the GMM parameters instead to adapt to the statistical fluctuations of the training sample. An example of the expected and observed smooth behaviour is shown in Fig.5.8.

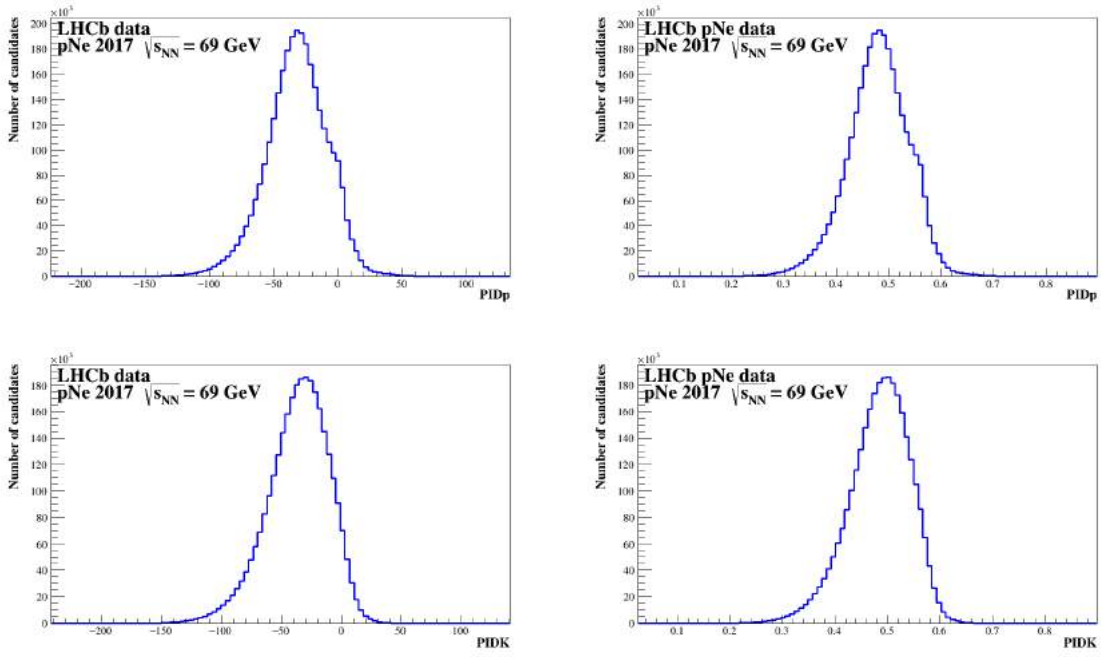


Figure 5.6: MixMaxScaler preprocessing applied to the PID target variables. On the left, original distributions, and on the right, distributions after the preprocessing.

5.2.5 Model Validation

The model is validated by comparing the GMM prediction with the training samples after the overtraining checks. The training dataset is binned with respect to the input features. For each bin the model predicts the pdfs in the 2D PID space for each particle of the dataset, given the input features. To validate the overall pdf distribution for a dataset of size N, a few values are sampled from each pdf predicted by the GMM given the element N_i and its features. The total pdf for the dataset is then obtained by the overall distribution of the sampled values. The results of the validation process are shown in a few features bins for $K^0_S \to \pi^-\pi^+$ (Fig.5.9), $\overline{\Lambda} \to \overline{p}\pi^+$ (Fig.5.10) and $\phi \to K^-K^+$ (Fig.5.11) lines.

5.2.6 Application to the pHe and pAr datasets

After the validation on the same training dataset, the pNe dataset, the GMM algorithm prediction has been applied and validated against two other lower-statistics datasets collect during Run2, pAr and pHe, with a nucleon center-of-mass energy of $\sqrt{s_{NN}}$ = 110 GeV. These datasets differ from the pNe used to train the networks due to the different collision energy and have different feature distribution shapes since the number of nucleons per target nuclei is also different. Assuming the same data-taking conditions, the change in the feature distribution should be accounted for by the GMM algorithms and the three models, trained on the $K^0_S \to \pi^-\pi^+$, $\overline{\Lambda} \to \overline{p}\pi^+$ and $\phi \to K^-K^+$ lines of the pNe data, are expected to generate the pdfs for π, K, and p, respectively. The procedure to extrapolate the pdfs is the same as applied for the validation on pNe (Sec.5.2.5), but the pHe and pAr features are given as input to the networks. Data-driven templates are compared with the detailed simulations

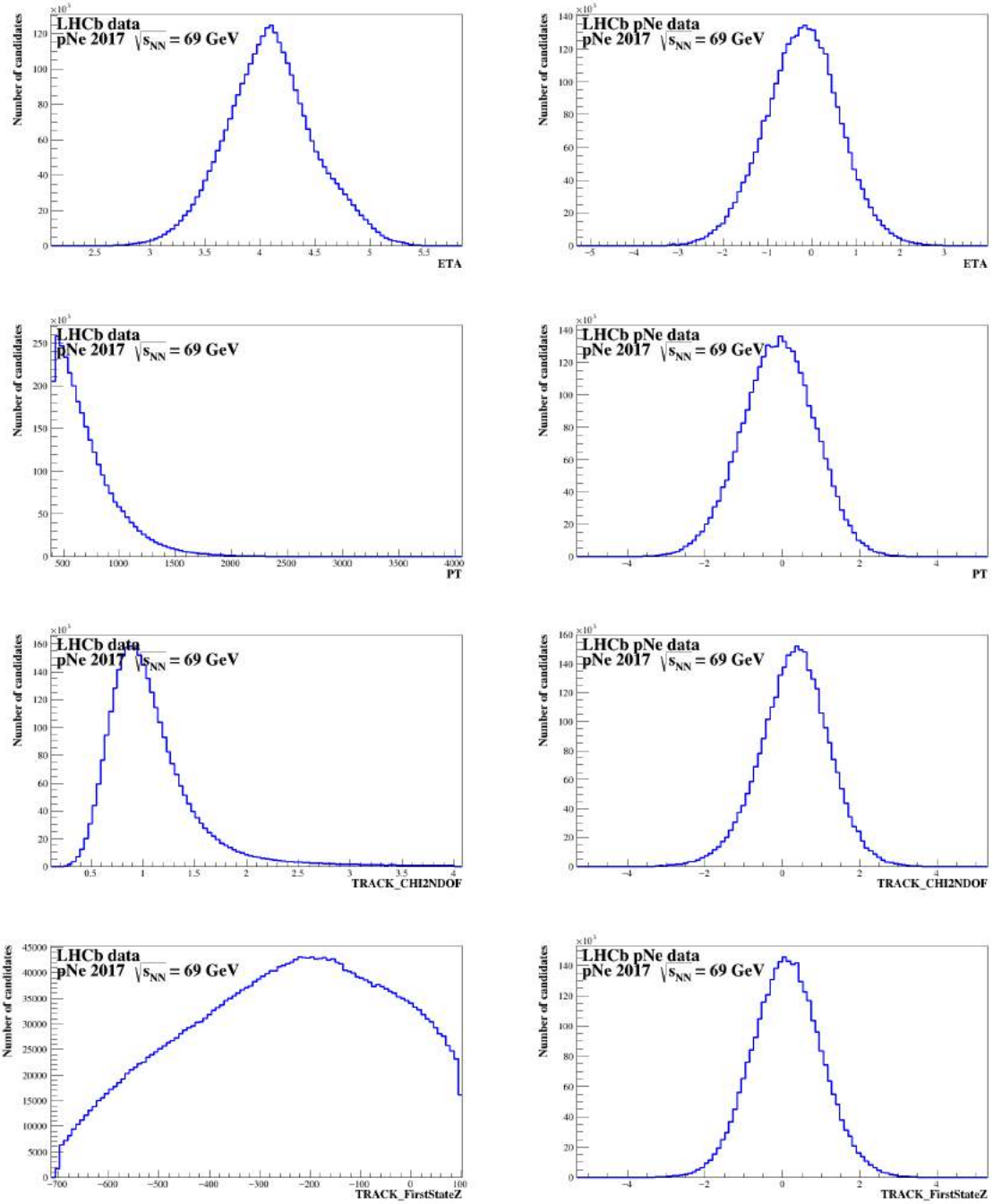


Figure 5.7: Feature distributions (left) preprocessed using the Quantile Transformer algorithm (right)

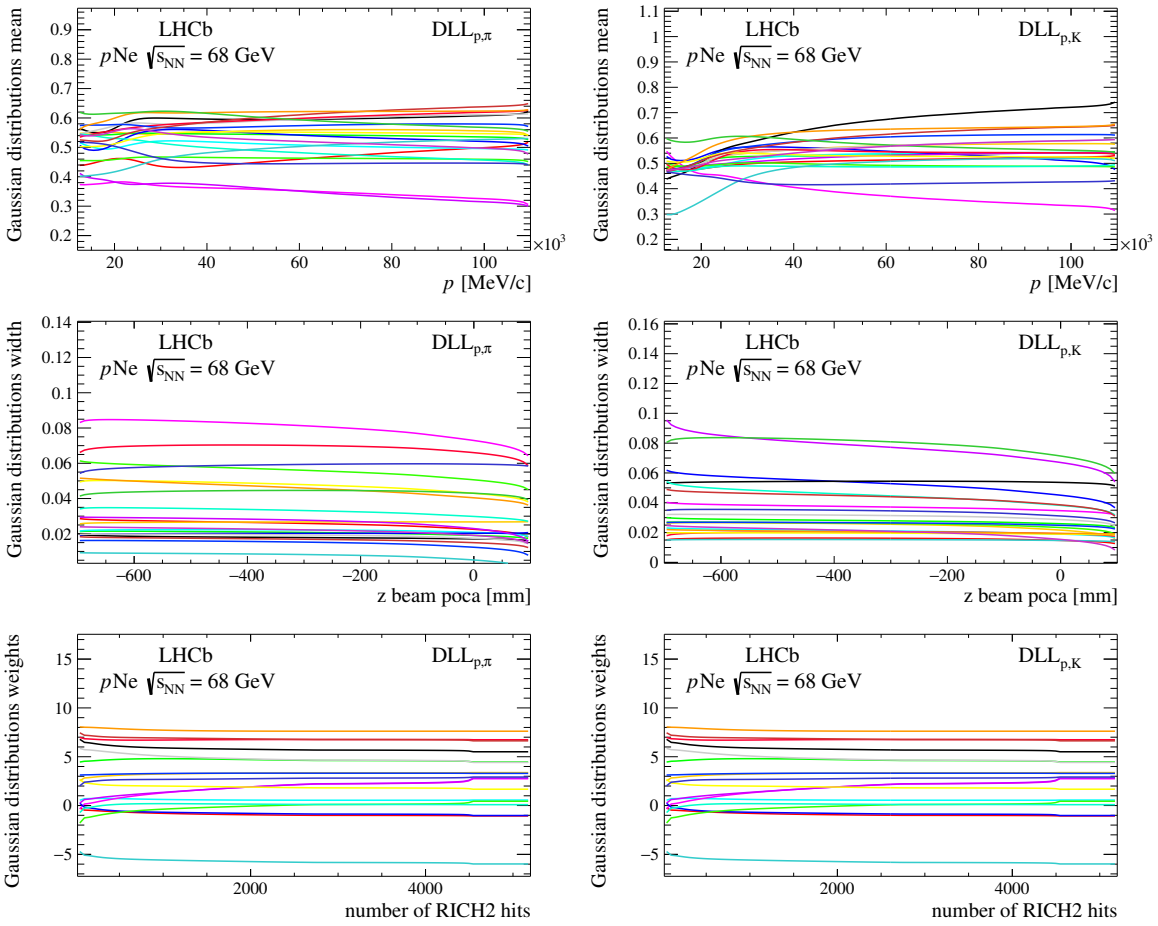


Figure 5.8: Check against overtraining reported as an example for the $\overline{\Lambda} \to \overline{p}\pi^+$ calibration channel from the evolution of (top) the Gaussian mean values, (middle) widths and (bottom) weights for (left) the DLL$_{p,\pi}$ and (right) the DLL$_{p,K}$ target variables as a function of the particle momentum, z coordinate of the track position of closest approach to the beam and number of hits in the RICH2 subdetector. Different colors represent the different Gaussian components considered in the model.

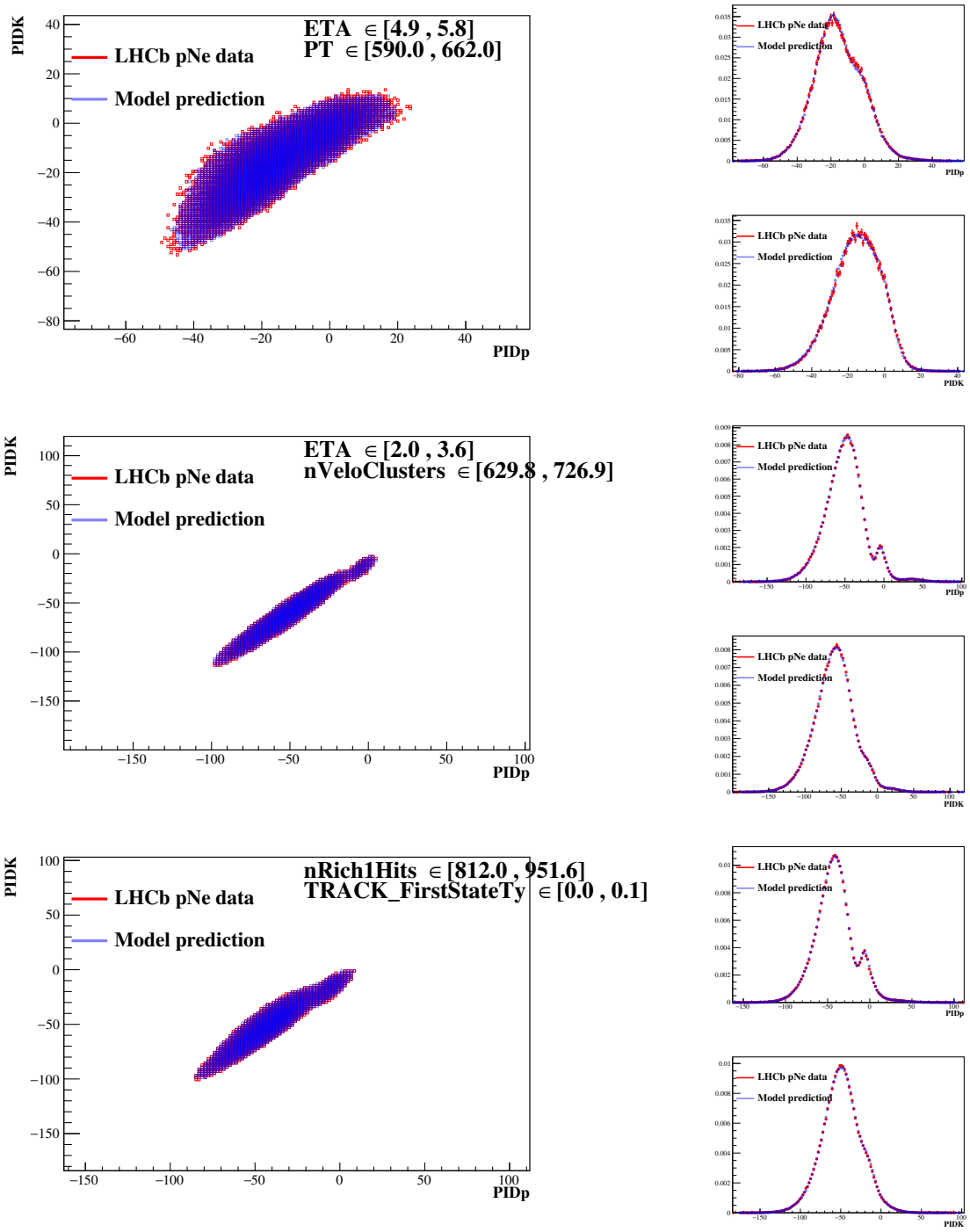


Figure 5.9: Validation of the $K^0_S \to \pi^-\pi^+$ line in selected bins of the feature space. Left: comparison between data and model predicted pdfs in the PID space. Right: the pdfs are projected into the PIDp and PIDK axis.

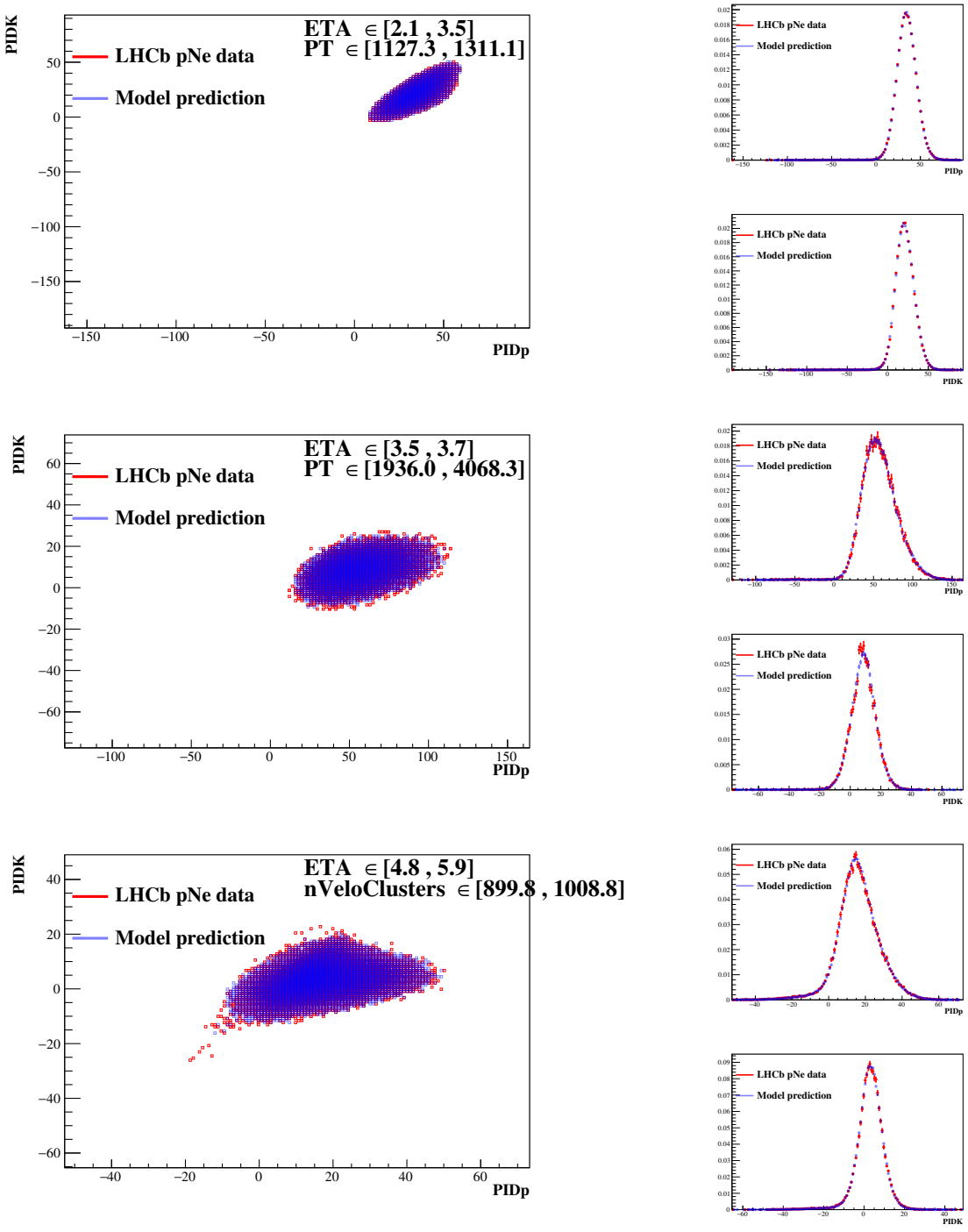


Figure 5.10: Validation of the $\overline{\Lambda} \to \overline{p}\pi^+$ line in selected bins of the feature space. Left: comparison between data and model predicted pdfs in the PID space. Right: the pdfs are projected into the PIDp and PIDK axis

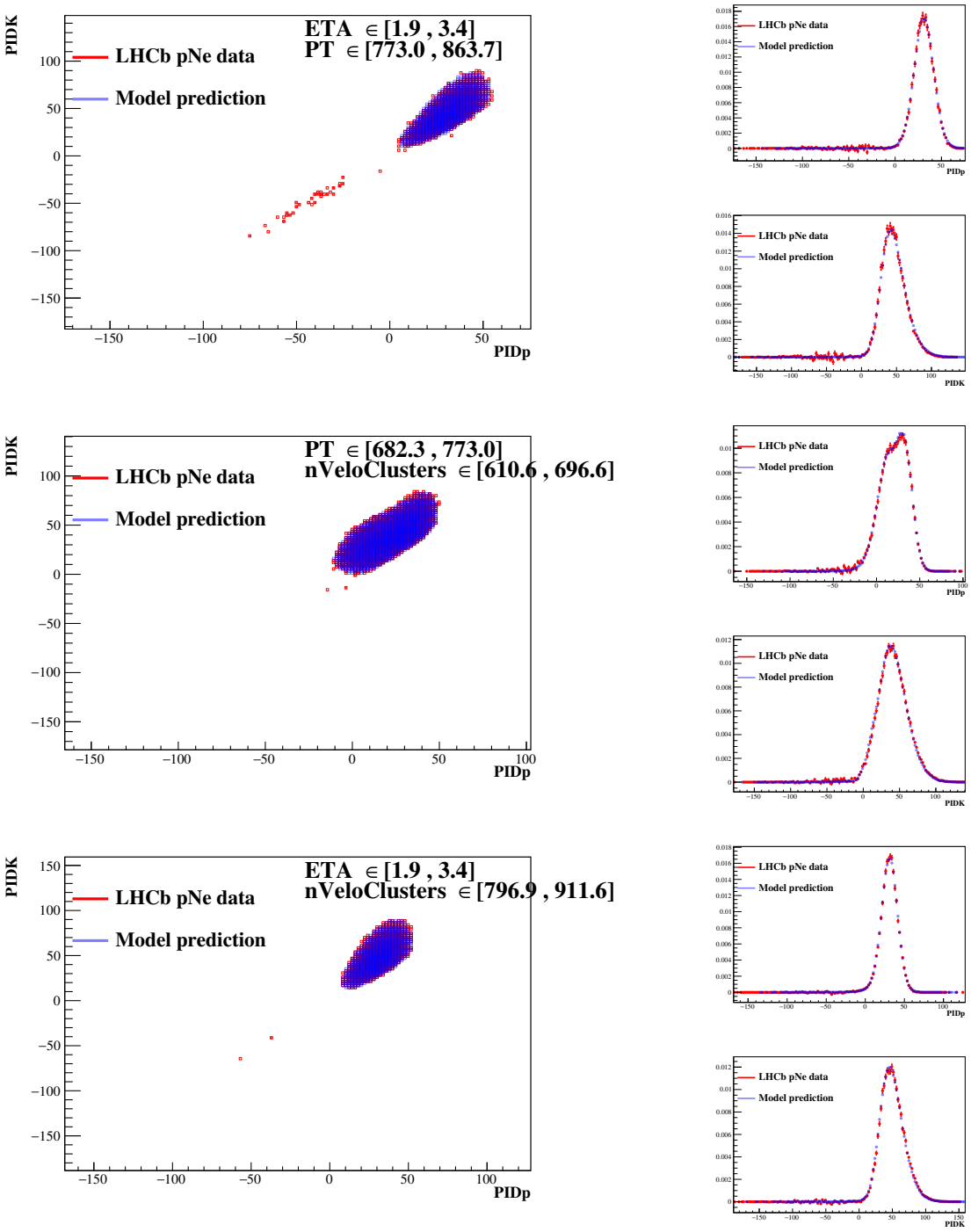


Figure 5.11: Validation of the $\phi \rightarrow K^{-}K^{+}$ line in selected bins of the feature space. Left: comparison between data and model predicted pdfs in the PID space. Right: the pdfs are projected into the PIDp and PIDK axis. The data shown for this line are weighted to observe the signal contribution using the sWeights, this implies the statistical fluctuations that can be observed when the data pdfs reach negative values.

and the results obtained with respect to $\mathrm{DLL}_{p,\pi}$ and $\mathrm{DLL}_{p,K}$ variables for the pHe sample. $\mathrm{DLL}_{p,\pi}$ and $\mathrm{DLL}_{K,\pi}$ variables are used for the pAr sample, to observe the performance by varying the target space. Simulated data samples are generated with EPOS-LHC, and particle interactions with the detector and its response are implemented using the GEANT4 toolkit.

Reweighting techniques can be applied to simulation to correct for the small discrepancies observed with data. The reweight procedure was applied to the simulated pHe sample in two steps, firstly reweighting the detector occupancy (using $nSPDHits$) and then the tracks p_{T}. For the pAr sample, a boosted decision tree algorithm is used to determine a single event weight depending on p_{T} and the number of hits in the SPD and RICH1 subdetectors.

A simulation-based template is obtained for each particle species on every bind after reweighting. In addition, a fourth particle category is considered (and can be obtained from simulation only): ghost tracks reconstructed from hits of different simulated particles and accounting for only a few percent of candidate tracks. It is only possible to obtain the corresponding template through simulation, and it is also used in the data-driven model. Using either the simulation-based or the data-driven approach, the templates are finally fitted to the pHe (Fig. 5.13) and pAr (Fig. 5.16) data. In this study, kinematic intervals are used to fit a maximum likelihood distribution over a two-dimensional target variable distribution leaving as free parameters the relative abundances of each particle species. The simulation-based templates and fit projections to pHe data are shown in Fig.5.12 for the same kinematic intervals. The fit projections for five other kinematic intervals for the pAr sample can be seen in Fig.5.15). Simulation-based and data-driven fit results are directly compared in Figs.5.14-5.17. The data-driven method predicts the PID classifier distribution more accurately than the full-simulation method. This model appears to be less biased due to the larger training dataset and the use of smooth functions. The Kolmogorov-Smirnov distances between the fitted distribution and the actual data distribution are measured in kinematic bins to better quantify the fit quality between the two sets of templates. The difference between the values obtained with the simulation-based and data-driven templates is shown in Figs. 5.18 for the pHe (top) and pAr (bottom) data samples, respectively. The observation of positive differences in most of the bins in both cases demosntrates that the templates produced with the GMM method describe the data better than those based on detailed simulations.

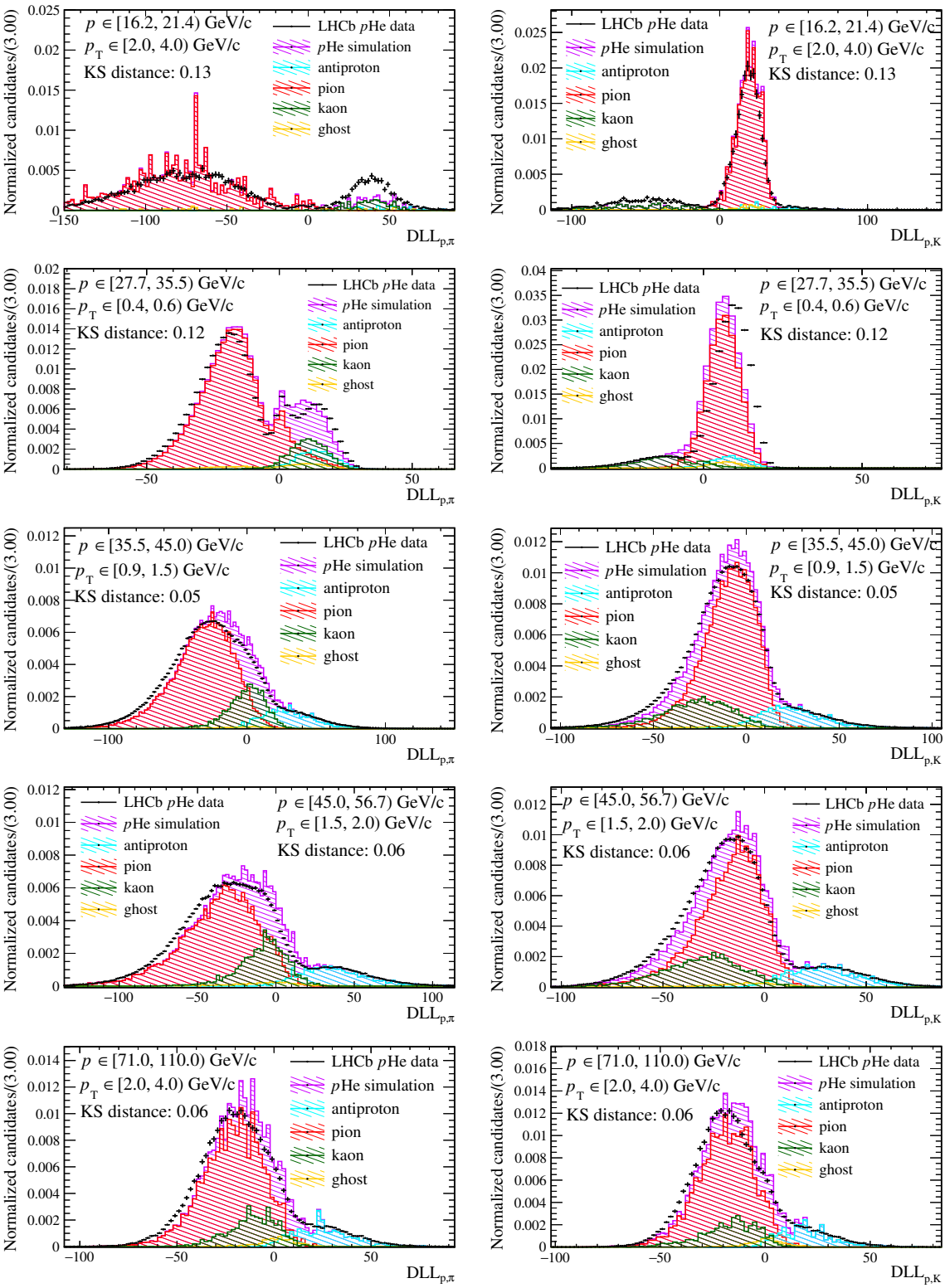


Figure 5.12: Projections onto the (left) $\mathrm{DLL}_{p,\pi}$ and (right) $\mathrm{DLL}_{p,K}$ axes of the fit to pHe data employing simulation-based templates in five momentum, transverse momentum intervals.

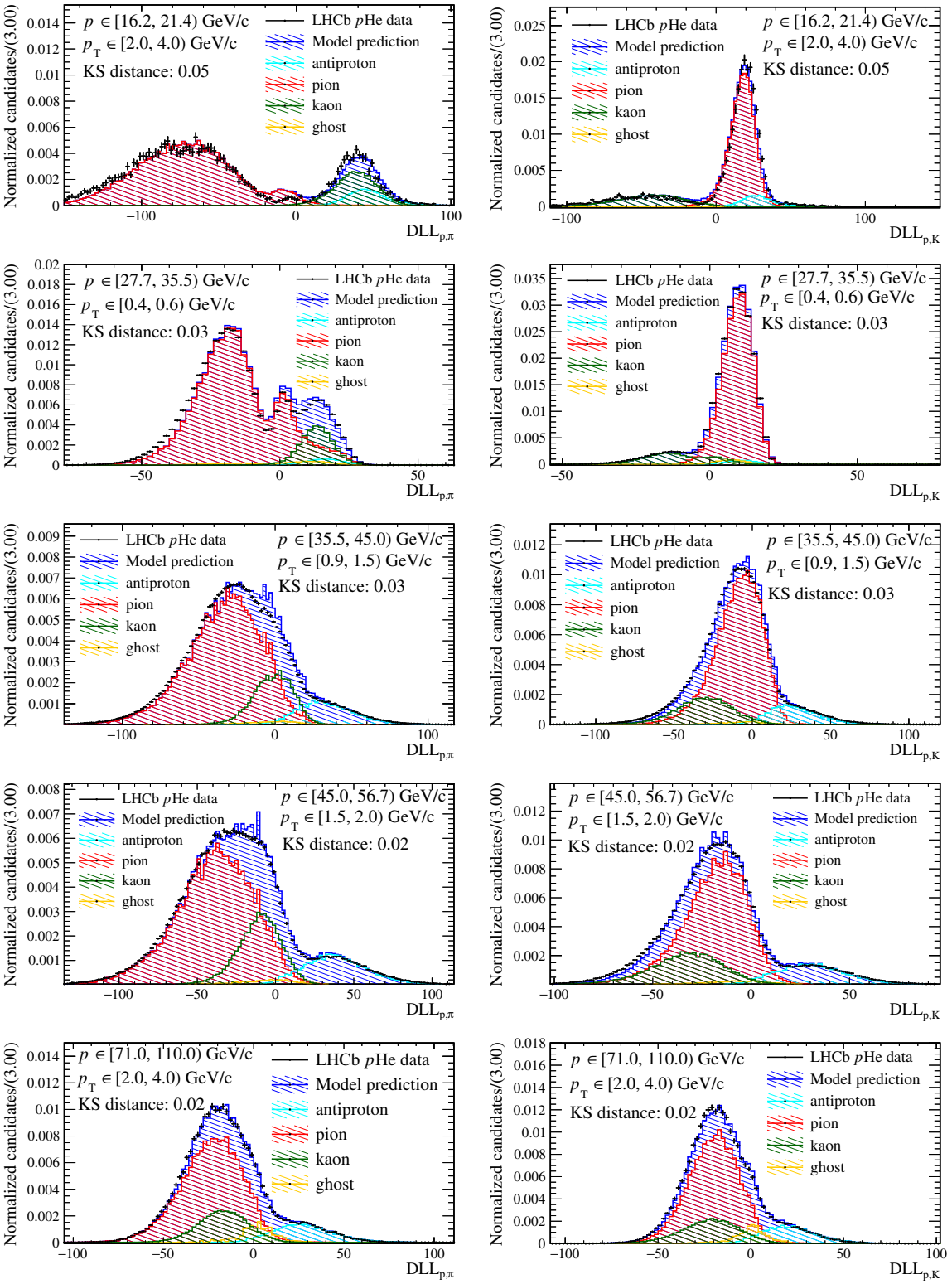


Figure 5.13: Projections onto the (left) $\mathrm{DLL}_{p,\pi}$ and (right) $\mathrm{DLL}_{p,K}$ axes of the fit to pHe data employing the GMM data-driven templates in five momentum and transverse momentum intervals.

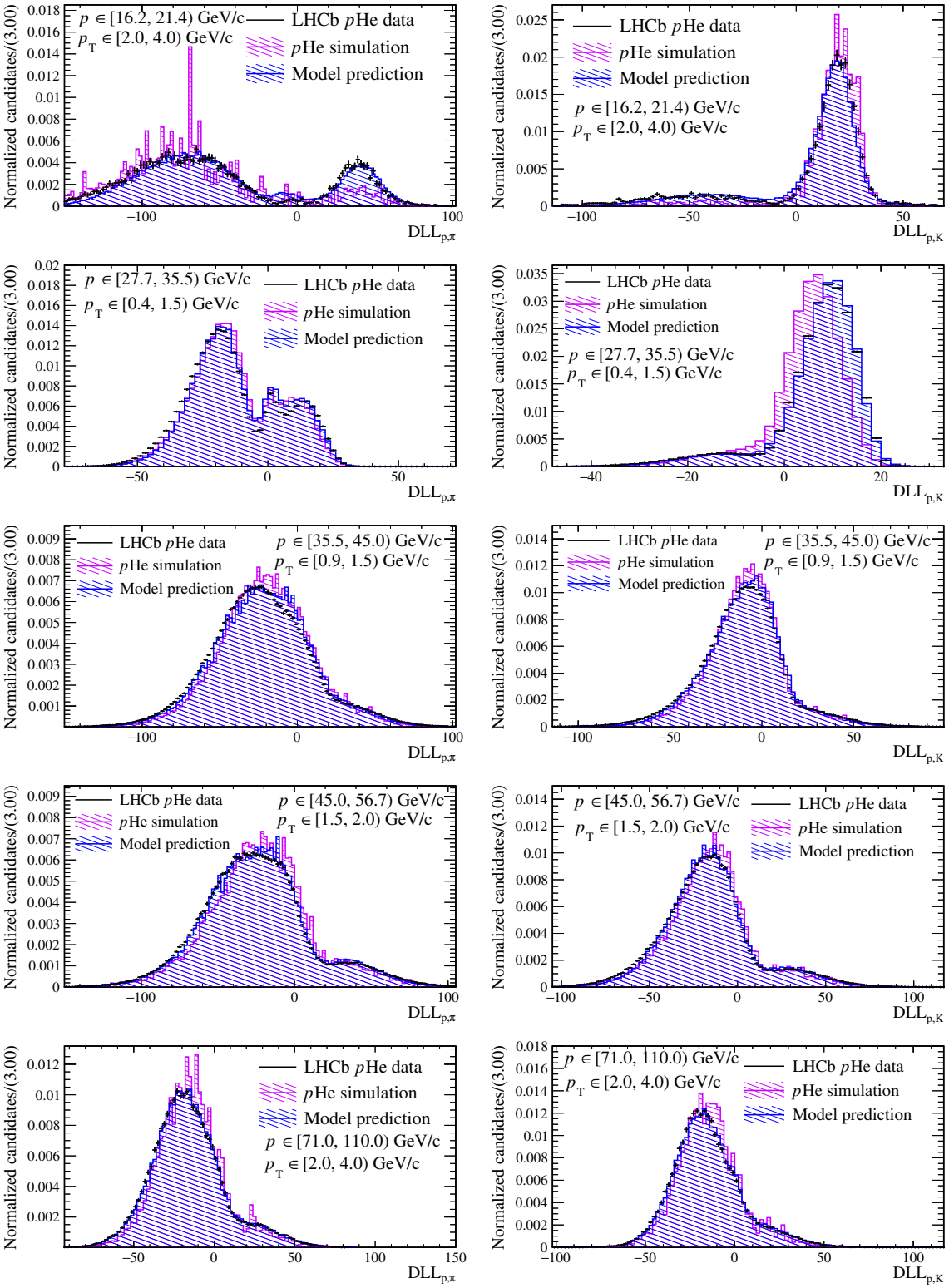


Figure 5.14: Comparison of the (left) $\mathrm{DLL}_{p,\pi}$ and (right) $\mathrm{DLL}_{p,K}$ distributions in pHe data modelled with simulation (violet) and data-based (blue) templates in five momentum, transverse momentum intervals.

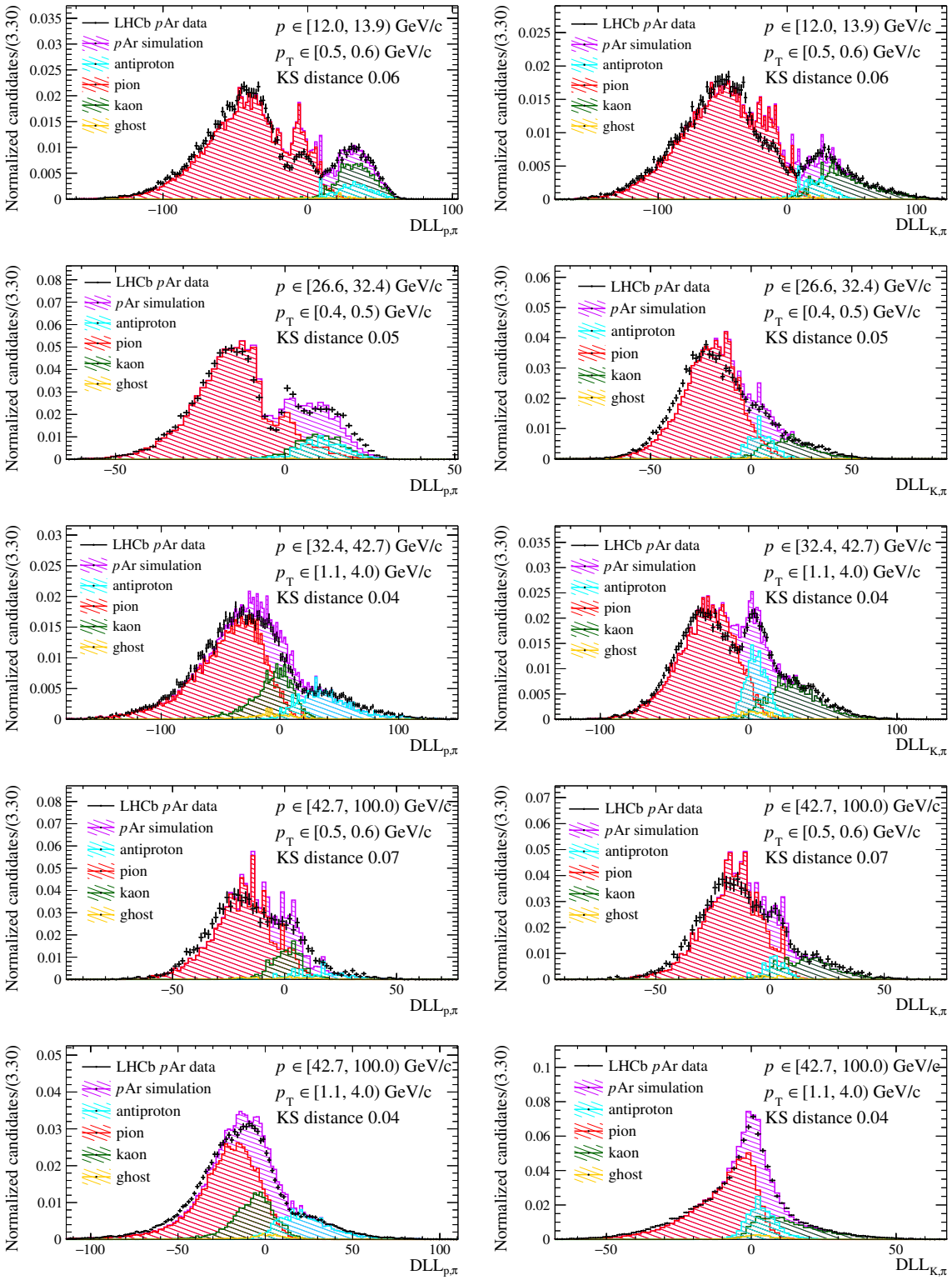


Figure 5.15: Projections onto the (left) $\mathrm{DLL}_{p,\pi}$ and (right) $\mathrm{DLL}_{K,\pi}$ axes of the fit to pAr data employing the simulation-based templates in five momentum and transverse momentum intervals.

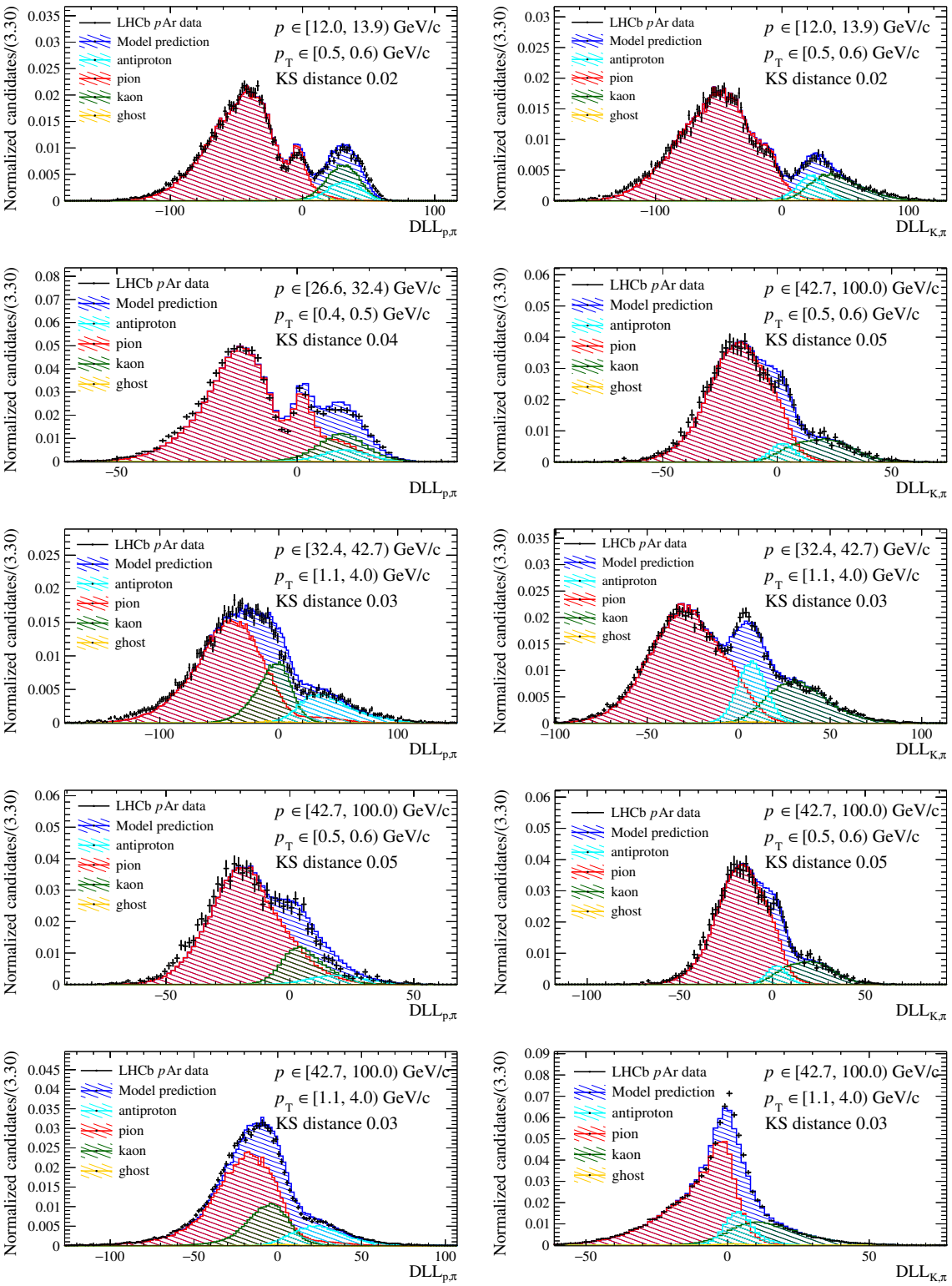


Figure 5.16: Projections onto the (left) $\mathrm{DLL}_{p,\pi}$ and (right) $\mathrm{DLL}_{K,\pi}$ axes of the fit to pAr data employing the GMM data-driven templates in five momentum and transverse momentum intervals.

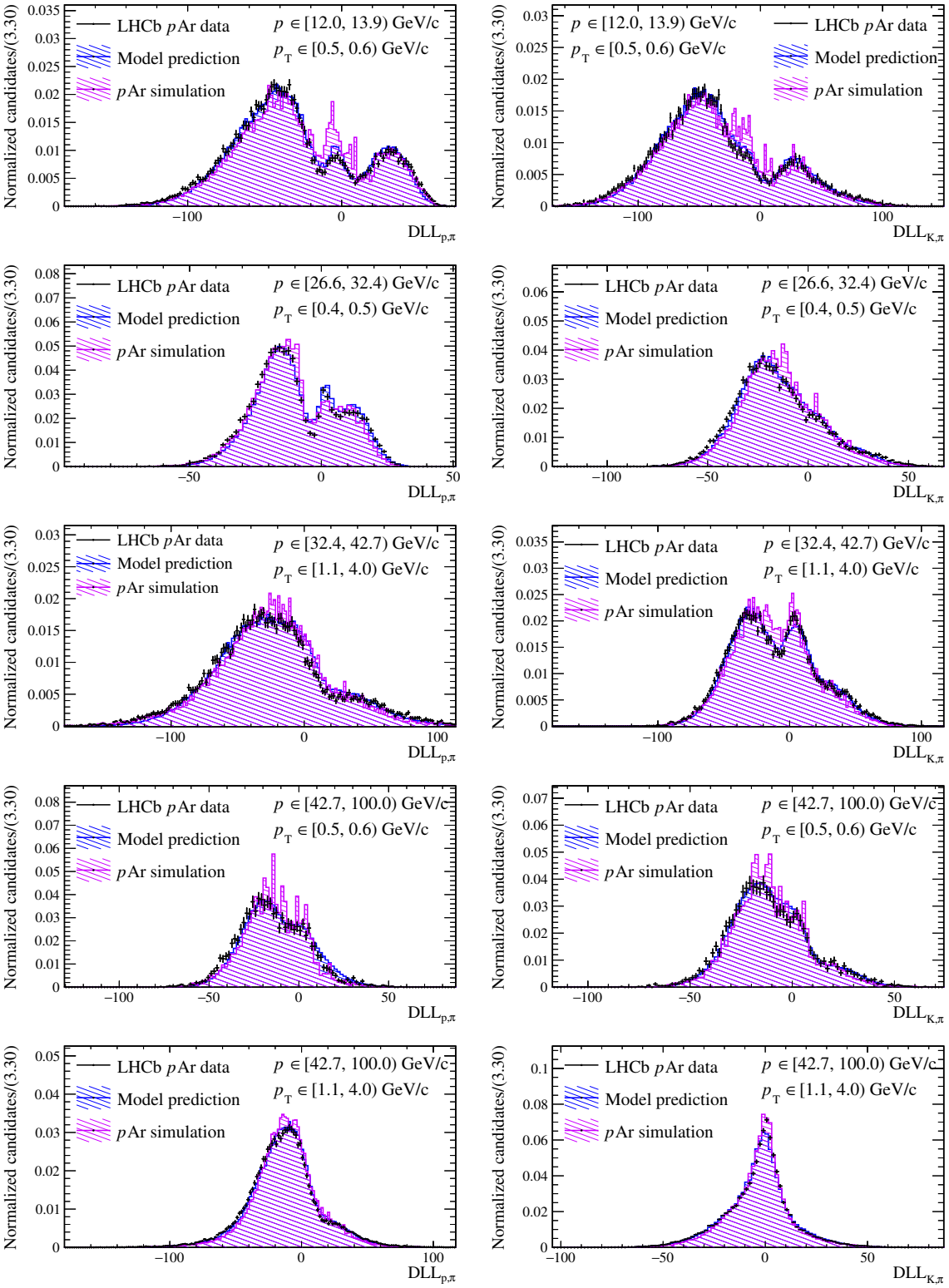


Figure 5.17: Comparison of the (left) $\mathrm{DLL}_{p,\pi}$ and (right) $\mathrm{DLL}_{K,\pi}$ distributions in pAr data modeled with simulation (violet) and data-based (blue) templates.

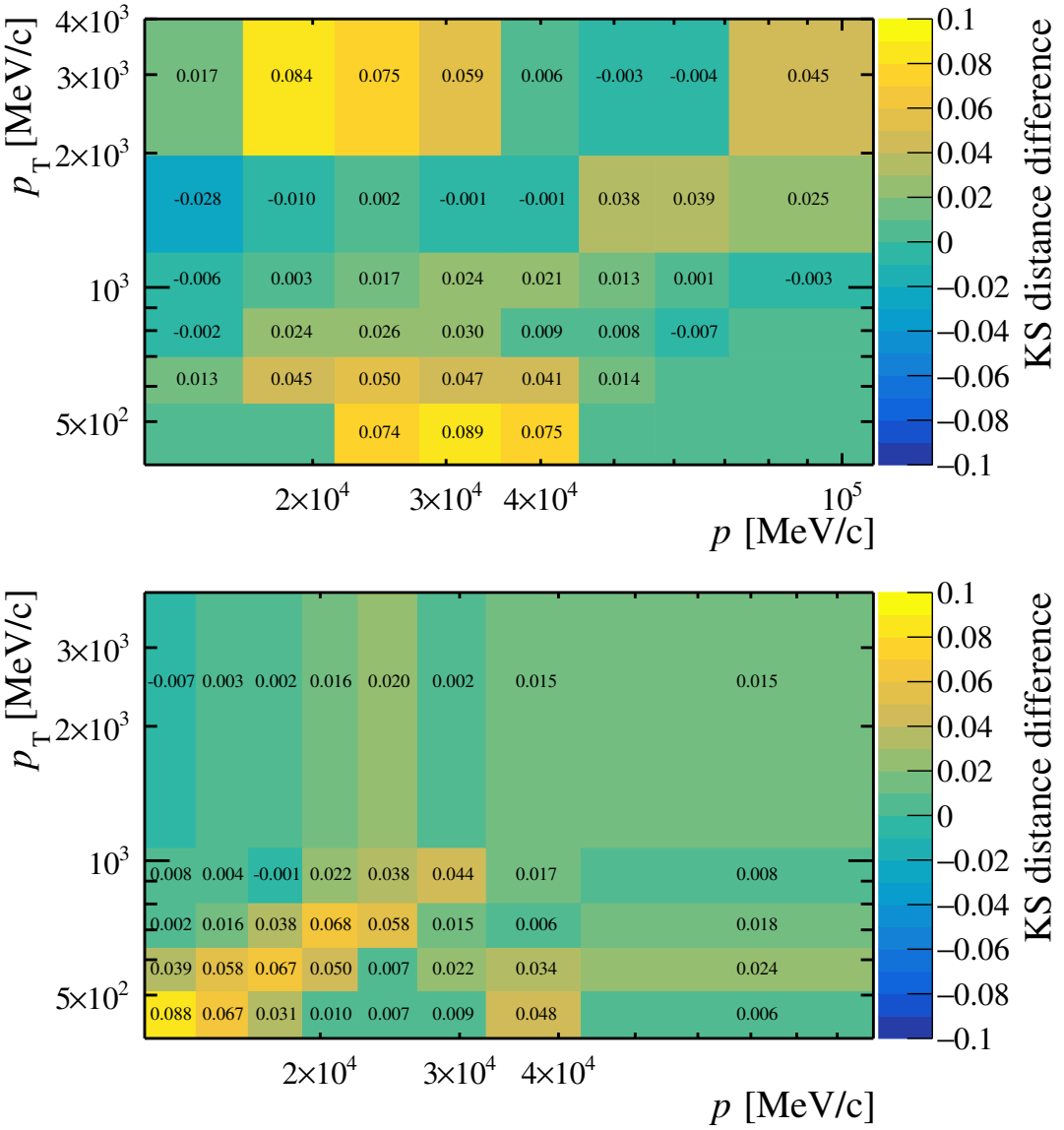


Figure 5.18: Difference of the Kolmogorov-Smirnov distance for the modeling of the bi-dimensional (top) pHe or (bottom) pAr target variables distribution composing templates produced with a full simulation approach or with the machine-learning GMM method.

Chapter 6

Light Hadron Production

My analysis work on light hadron production using fixed-target data is presented here. The analysis makes use of the PID tool described in Chap.5. In this section, an outline of the analysis strategy is given, followed by a description of the relevant measurements and the datasets used in the study. The main analysis steps include the selection of charged tracks, followed by simulation studies and efficiencies evaluation. The templates are then generated using the NN GMM and the data are fitted on the PID space to extract the relative yields for pHe and pAr samples for π, K and p, also selecting the electric charge of the tracks. The systematics uncertainties not canceling in the yield ratios are also estimated. The study is performed as a function of the transverse momentum p_T, the pseudorapidity η and a multiplicity variable proxy (nVeloClusters) by binning the datasets with respect to those variables, also simultaneously, achieving a multi-dimensional extraction of the results.

6.1 Analysis strategy

The goal of the present analysis is the measurement of inclusive prompt-charged particles in fixed-target proton-nucleus collisions. Specifically, the analysis makes use of the pHe and pAr datasets collected at a center-of-mass energy of 110 GeV by the LHCb experiment with the SMOG device. The analysis uses events triggered with minimum bias conditions in beam-empty bunch crossing configuration (only the bunch arriving from the VELO side if filled and trigger the event by colliding on the gas target injected in the beam pipe). In Sec. 6.1.1 the aim of the analysis is discussed more in detail as well as the target measurements. The datasets collected and analyzed are described in Sec. 6.2, including also the selection criteria which are first applied to all charged tracks. The charged particle composition of the analyzed sample consists of a mixture of pions, kaons, and protons. Simulated samples have been produced to estimate the selection efficiencies and to study the prompt condition selection (Sec. 6.4). The efficiencies have been calculated for each hadron type independently, since few differences may occur and are indeed observed, especially for the proton case.
Throughout the analysis, no PID requirement is imposed on candidates' tracks since the fraction of π, K, and p is determined by a bi-dimensional fit in the PID plane using a fully data-driven method based on the NN-based Gaussian Mixture Model discussed previously. The shape of each template used to fit the data is given by the prediction of the NN for the π, K, and p. A template modeling

the ghost tracks is also included and drawn from simulations. In Sec. 6.5 the whole procedure is described more in detail.
The relative abundances for π, K, and p are obtained after binning the datasets in η, p_T, and multiplicity. The binning is defined with respect to the previous variables in one or more dimensions (up to 3 when all the variables are considered simultaneously). For each chosen binning the yield extraction procedure is applied. The main sources of systematics are evaluated in Sec. 6.6. The results are shown in Sec. 7.1.

6.1.1 Physics Observables

The aim of the analysis is to obtain the distributions of light hadrons (π, K and p) as a function of the variables to which CNM effects are more sensitive when light and heavy nuclei are compared. Those variables are the pseudorapidity η of the track, the transverse momentum p_T and a multiplicity variable which can be used to parameterize the centrality of the collision. For the latter, the nVeloClusters variable is used as a multiplicity proxy in this analysis. The distributions are obtained separately for each gas type and for both positive and negative tracks. The luminosity measurement is not yet available for all the SMOG datasets. For the ones considered in this analysis, the luminosity has been determined only for the pHe sample, so the candidates' yield ratios are the measurements of main interest since the luminosity is not required in such cases. The relevant distributions are the following:

- η, p_T, and multiplicity distributions for each particle type (π, K, p) and collision type (pHe and pAr);
- p/π, p/K, k/π charge-separated ratios for each collision type;
- double ratios in 1D, 2D and 3D in η, p_T, and multiplicity.

The CNM effects are revealed and quantified by studying the deviations from unity of double yields ratios defined, e.g. for the 2D case, as:

$$\frac{\frac{d^2N^{pAr\to h+X}}{dp_Td\eta}/\frac{d^2N^{pAr\to h'+X}}{dp_Td\eta}}{\frac{d^2N^{pHe\to h+X}}{dp_Td\eta}/\frac{d^2N^{pHe\to h'+X}}{dp_Td\eta}}, \tag{6.1}$$

where h and h' denote charged hadrons of interest. In the ratio most of the systematic uncertainties as well as the gas purity and luminosity determination cancel.

6.2 Datasets

The datasets were acquired during dedicated fills with very low bias conditions on beam1-empty (BCtype=1) crossings.

Fill ID	Filling scheme	Non coll. Bunches
4495	100ns_54b_42_18_18_18bpi3inj	36
4496	100ns_252b_240_68_87_72bpi5inj	165
4499,4505,4509,4510,4511	100ns_685b_671_665_1_72bpi13inj	684

Table 6.1: Fills delivered by LHC during the pAr data-taking. The number of non colliding bunches, of particular interest for beam-empty buch crossings, are reported.

Fill ID	Filling scheme	Non coll. Bunches
4937	Multi_55b_11_28_16_4bpi14inj	39
4945, 4954	Multi_56b-52b_32_16_8_4bpi14inj	48

Table 6.2: Fills delivered by LHC during the pHe data-taking. The number of non colliding bunches, of particular interest for beam-empty buch crossings, are reported.

pAr dataset

The data were acquired during the special high β^* run in October 2015. The beam energy was 6.5 TeV, corresponding to a center-of-mass energy per nucleon of $\sqrt{s_{NN}}$ = 110 GeV. The selected fills are listed in Tab. 6.1. In total, 1477 M events have been recorded with an integrated luminosity of approximately 3 nb^{-1}. The trigger line used to select MB events is Hlt1MBMicroBiasVelo which requires at least one reconstructed track in the VELO and runs with a prescale 0.01.

pHe dataset

The pHe dataset was collected in 2016 during VdM dedicated fills with a beam energy of 6.5 TeV. The Fills are reported in Tab. 6.2. The total number of recorded events is 1400 M, with an integrated luminosity of $\sim 1.7\ nb^{-1}$. Events are selected based on the MicroBiasVelo HLT1 trigger requiring at least one reconstructed VELO track. The line was prescaled with different values during the data acquisition: 1, 0.5, and 0.1.

6.3 Charged tracks selection

The Runs used in the analysis are shown in Tab.6.3. The candidates are selected by considering charged tracks in the events. From the tracks, all charged particles are reconstructed as pions using the StdAllNoPIDsPions definition. The StdAllNoPIDsPions selection uses the charged tracks in the events with some minimum quality requirements and assign them the mass hypothesis of a pion. The baseline assumption that each candidate is a π is used only to retrive a particle object from the reconstruction. The real abundances for each particle hypothesis (π, K, p) are extracted independently from the initial assumption, as explained in Sec .6.5. The StdAllNoPIDsPions candidates satisfy the following criteria:

- track $\frac{\chi^2}{ndof} < 4$: a loose requirement on the quality of the track fitting during the reconstruction;
- ghost probability < 0.4: excludes most tracks that are reconstructed from random hits in the detector;

Gas	Good Runs
Ar	165892, 165893, 165894, 165895, 165976, 165993, 165994, 165996, 166007, 166008, 166062, 166063, 166064, 166065, 166066, 166067, 166089, 166090, 166091, 166092, 166093, 166094, 166095, 166096, 166097, 166111, 166112, 166113, 166114, 166134, 166135, 166136, 166137
He	174625, 174626, 174627, 174629, 174630, 174631, 174632, 174633, 174634, 174635, 174636, 174637, 174638, 174705, 174708, 174710, 174711, 174713, 174714, 174715, 174716, 174717, 174718, 174719, 174720, 174721, 174722, 174723, 174724, 174727, 174728, 174729, 174730, 174731, 174732, 174733, 174734, 174735, 174736, 174737, 174738, 174740, 174741, 174742, 174743, 174744, 174745, 174746, 174748, 174749, 174750, 174751, 175290, 175292, 175294, 175296, 175298, 175299, 175302

Table 6.3: The list of good runs which passed the data quality requirements for both pHe and pAr datasets.

- the track is a long track: only the tracks which leave hits in all the tracking subdetectors are considered. The long tracks are, in general, the tracks with the best quality and better momentum and position resolution.

Other selection criteria are applied in cascade to the StdAllNoPIDsPions candidates:

- Global Event cuts
  - at least one reconstructed PV in the event in the range $PV_Z \in [-700, 100]$ mm. This range is defined using the antiproton analysis as a reference since the luminosity is already available and might be used. Along the transverse plane, the requirement is $PV_{[X,Y]} \in [-20, 20]$ mm.
  - nBackTracks < 4: from the simulation, a negligible fraction of events have backtracks ≥ 4. The presence of such events in data implies contamination of *splash events*, where beam-gas collisions occur several meters upstream of the VELO and produce events with high detector occupancy. This contamination is removed with this nBackTracks requirement (see Fig. 6.1).
- Track cuts
  - $p \in [12, 100]$ GeV/c: the momentum cut is chosen to ensure a good PID response from the RICH detectors. The lower limit is set to achieve a K identification, which is mandatory to separate the K abundance from the p one. The upper limit is slightly lower than the nominal RICH2 coverage.
  - $p_T \in [400, 4000]$ MeV/c: also the transverse momentum range is defined to cover the RICH acceptance. Lower p_T values imply a worse K-p separation and a small difference between the PID response of the two types of particle.

– hasRich == 1: the track left signal in the RICH detectors and has associated Cherenkov photons.

– $\chi^2_{IP} < 12$: cut for the prompt requirement, the particle track is consistent with the collision vertex. The χ^2_{IP} of a track is defined as the χ^2 difference of the PV fit considering or ignoring the track. Higher values are obtained if the track is not consistent with the PV.

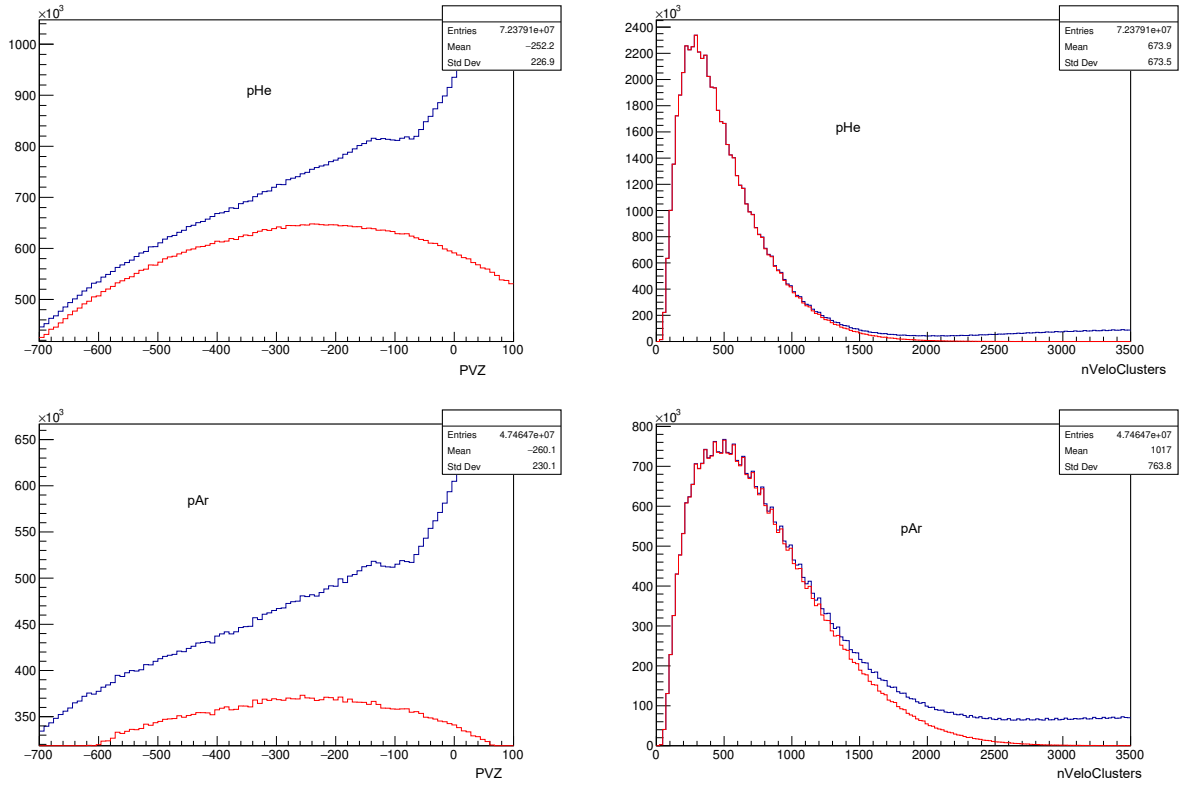


Figure 6.1: Effect of the nBackTracks cut on the PVZ and nVeloClusters distributions for pHe (top) and pAr (bottom). The cut is effectively cleaning the sample from splash events.

6.3.1 Binning definition

The binning is defined for each variable under study: η, p_T, and nVeloClusters. The limits are chosen by imposing a uniform binning in one dimension for each variable independently using a mixture of pAr and pHe candidates. The ranges are the following:

- η: [2.0, 3.56, 3.76, 3.90, 4.03, 4.14, 4.25, 4.37, 4.5, 4.8];
- p_T: [400, 457, 514, 575, 642, 722, 819, 952, 1178, 1500, 2100, 3000, 4000] MeV/c;
- nVeloClusters: [25, 280, 440, 640, 950, 7000] .

6.4 Simulation studies

For the evaluation of the efficiencies and of some systematic uncertainties, Monte Carlo (MC) data emulating the real collision event, as well as a full reconstruction of the generated tracks through a detailed model of the detector, are essential. Minimum bias MC samples are produced using the EPOS [74] generator implemented in Gauss which delivers the best description for ion physics data. For both pHe and pAr samples, 11 M events are generated.

After the digitization, reconstruction, and selection process, the distributions of some variables may differ between data and simulation. The distributions of some relevant variables are compared in Figs. 6.2-6.3.

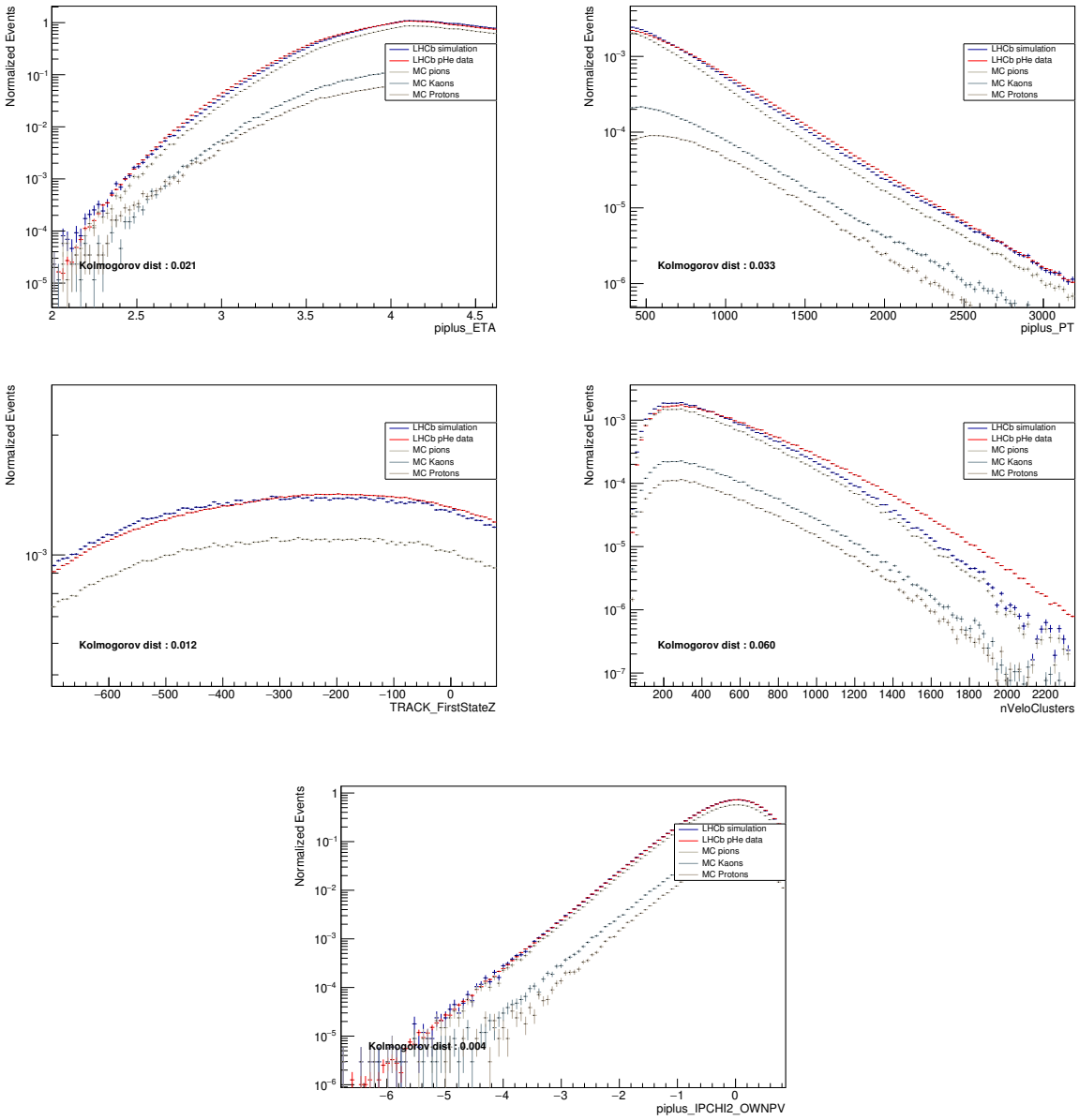


Figure 6.2: pHe simulation comparison with data. The reconstructed charged tracks distributions are compared with the simulation applying applying the track selection to both data and simulation. Only the variables most relevant for the analysis are shown, as well as simulated π, K and p contributions. The nVeloClusters variable is not reproduced with enough precision,also the low p_T region shows a significant difference (not the log scale on the Y axis).

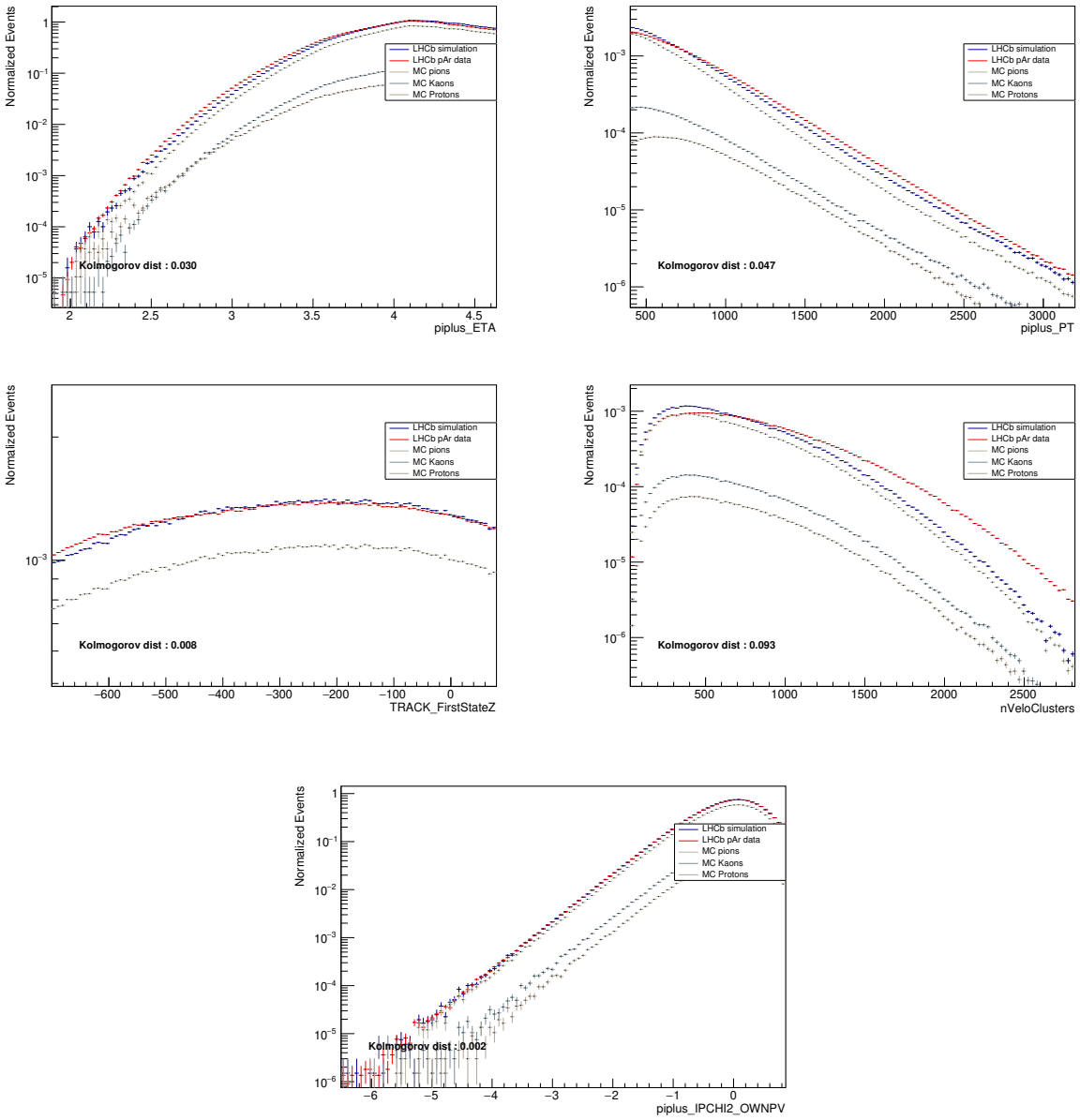


Figure 6.3: pAr simulation comparison with data. The reconstructed charged tracks distributions are compared with the simulation applying applying the track selection to both data and simulation. Only the variables most relevant for the analysis are shown, as well as simulated π, K and p contributions.

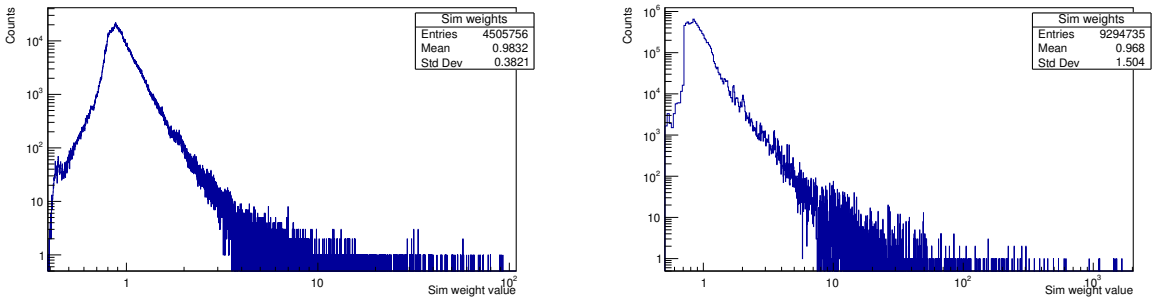


Figure 6.4: Distributions of the weights obtained from the GBReweighter algorithms for pHe (left) and pAr (right). The weights are then used to reweight the simulation.

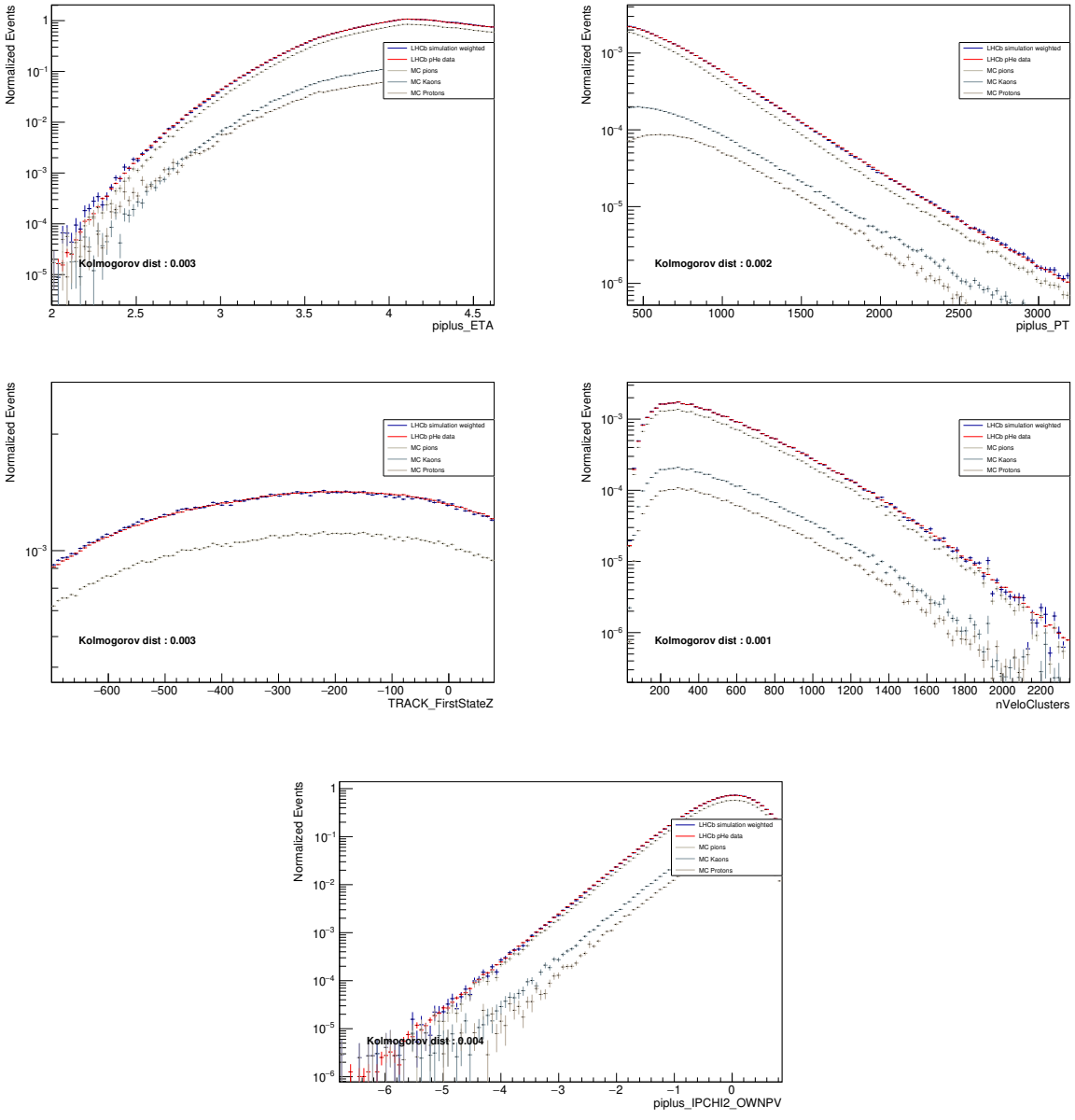


Figure 6.5: pHe simulation comparison with data after the reweighting. The reconstructed charged tracks distributions are compared with the simulation applying applying the track selection to both data and simulation and reweighting the latter as explained in the text. Only the variables most relevant for the analysis are shown, as well as simulated π, K and p contributions.

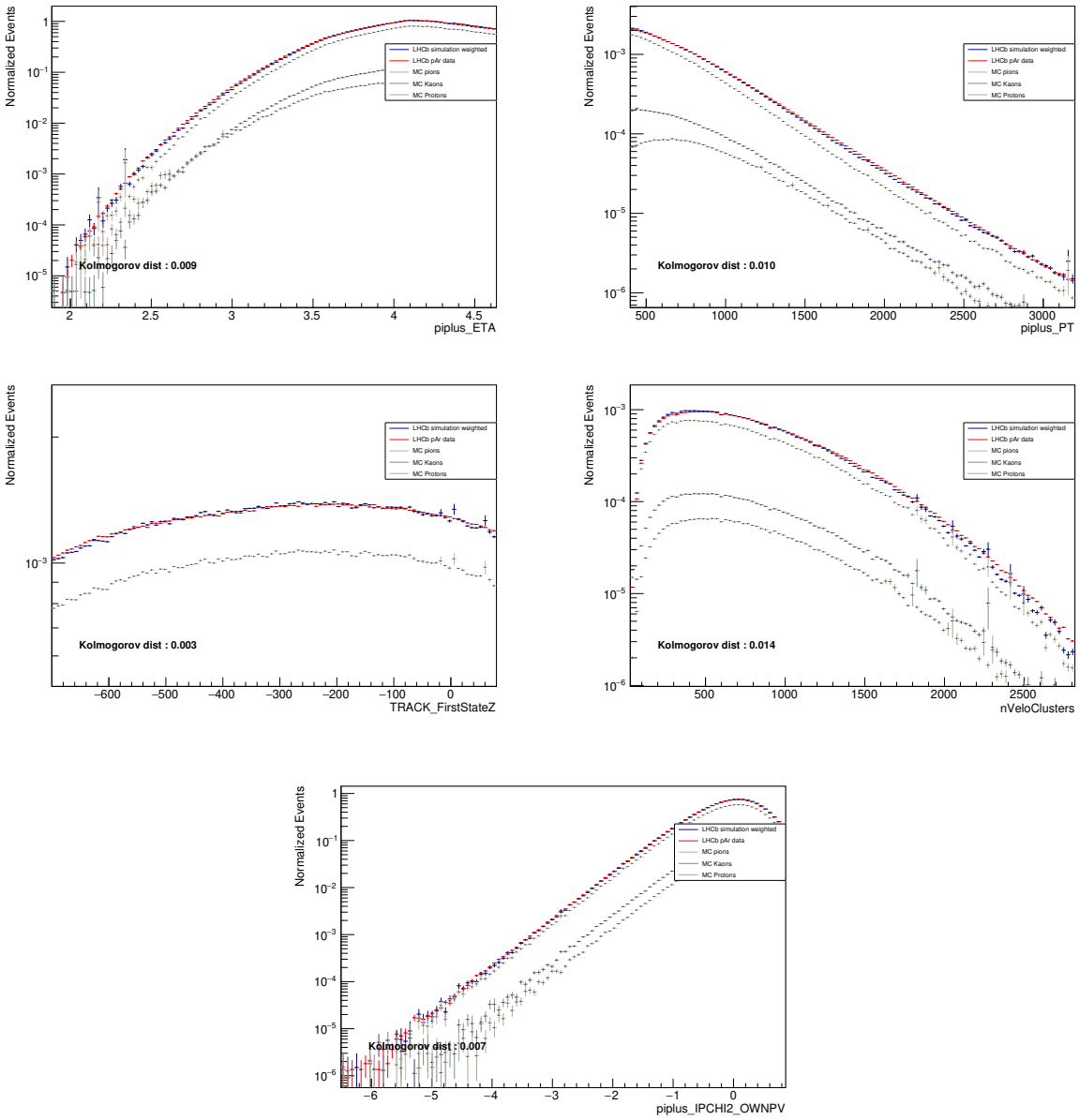


Figure 6.6: pAr simulation comparison with data after the reweighting. The reconstructed charged tracks distributions are compared with the simulation applying applying the track selection to both data and simulation and reweighting the latter as explained in the text. Only the variables most relevant for the analysis are shown, as well as simulated π, K and p contributions.

In order to improve the data-MC agreement, the most relevant distributions are reweighted using a multi-dimensional Gradient Boost Reweighting algorithm [75]. The algorithm is then configured to apply a 4D reweighting and the variable used to reweight the simulation are: η, p_T, nVeloClusters and track PV_Z (also called FistState$_Z$). The PV_Z is included in order to ensure that the distribution of the Z position of the primary vertex in the simulation reflects the real data distribution. The PV$_Z$ is correlated with the occupancy (for which nVeloClusters is the chosen proxy) and a reweighting of nVeloClusters modifies the PV$_Z$ distribution in an undesirable way if it is not included in the reweighting

Table 6.4: Efficiencies definitions and variable cuts. All except σ_{sel} are calculated using the true MC information. σ_{sel}, on the other hand, is calculated using the reconstructed MC tracks.

Efficiency	Definition	Requirement
acceptance σ_a	track in the detector acceptance	Reconstructible>0
tracking σ_t	track is reconstructed	Reconstructed>0
PV reconstruction σ_{PV}	at least one primary vertex reconstructed in the event	nPVs>0
selection σ_{sel}	loose requirements on track quality	reconstructed as a StdAllNoPIDsPions hasRich = 1 $IP\chi^2_{OWNPV} < 12$ $abs(MCPV_Z - OWNPV_Z) < 20\ mm$ $abs(MCPV_{[X,Y]} - OWNPV_{[X,Y]}) < 2\ mm$

process. The parameters of interest such as the number of iterations, learning rate, maximum depth, and minimum number of leaves have been chosen independently for the pHe and pAr datasets to avoid overfitting and a GBReweighter algorithm is trained separately for each dataset. The weights are obtained using a folding over three reweighters, the weights distributions are shown in Fig. 6.4. The distributions after the application of the extracted weights are shown in Figs. 6.5-6.6. A clear improvement is visible and quantified by the Kolmogorov-Smirnov (KS) distance between data and weighted simulation.

6.4.1 Efficiencies

The efficiencies are calculated from the simulation using the same binning applied to the data. The selections are applied in a cascade on simulated data and the efficiencies are reported in Table 6.4.

The total efficiency is then defined as

$$\sigma_{tot} = \sigma_a \cdot \sigma_t \cdot \sigma_{PV} \cdot \sigma_{sel} \tag{6.2}$$

where the various contributions are detailed in Table 6.4.

The 1D efficiencies as a function of η, p_T, or nVeloClusters are shown in Figs. 6.7, 6.8, 6.9 for each hadron type in the pHe simulation. The binning scheme is the same as reported in Sec. 6.3.1. There are small differences between the efficiencies for π, K, and p, the latter being systematically lower. As a consequence, each particle type is treated independently and corrected by its own efficiency. In Fig. 6.10 the total efficiencies in 1D are shown for both samples. The trends for the pHe and pAr datasets are similar, the latter corresponding to a total efficiency slightly lower than the former for all particle types. The 2D efficiency visualization is shown in Fig. 6.11-6.12 for pHe simulation.

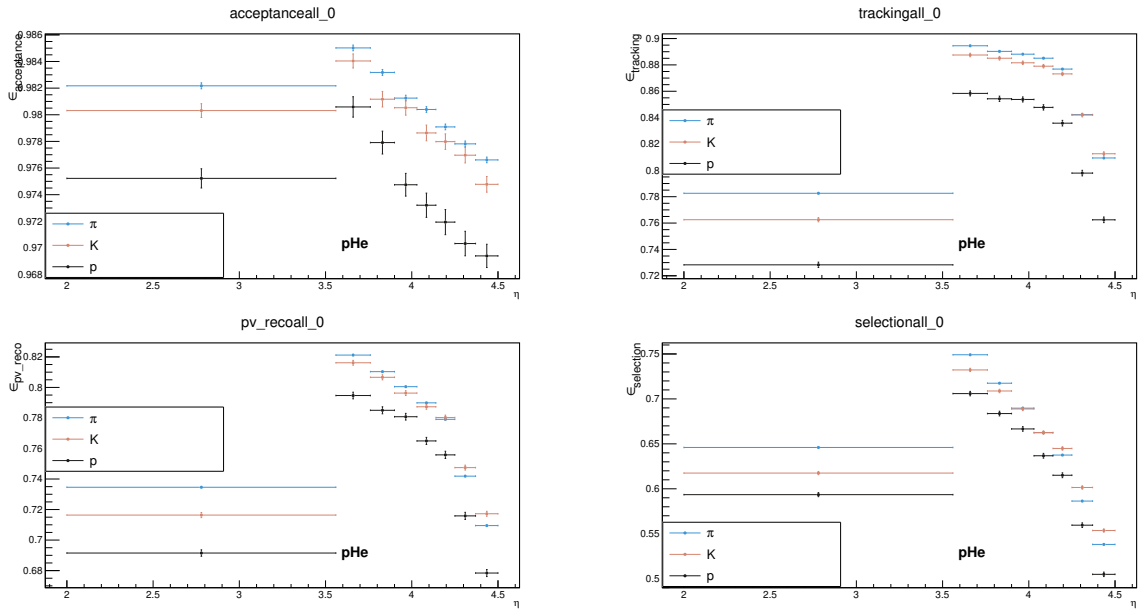


Figure 6.7: 1D efficiencies for π, K and p using the pHe reweighted simulation in bins of η. Top-left: acceptance efficiency for , Kp. Top-right: tracking efficiency. Bottom: pv reconstruction efficiency (left) and selection efficiency (right).

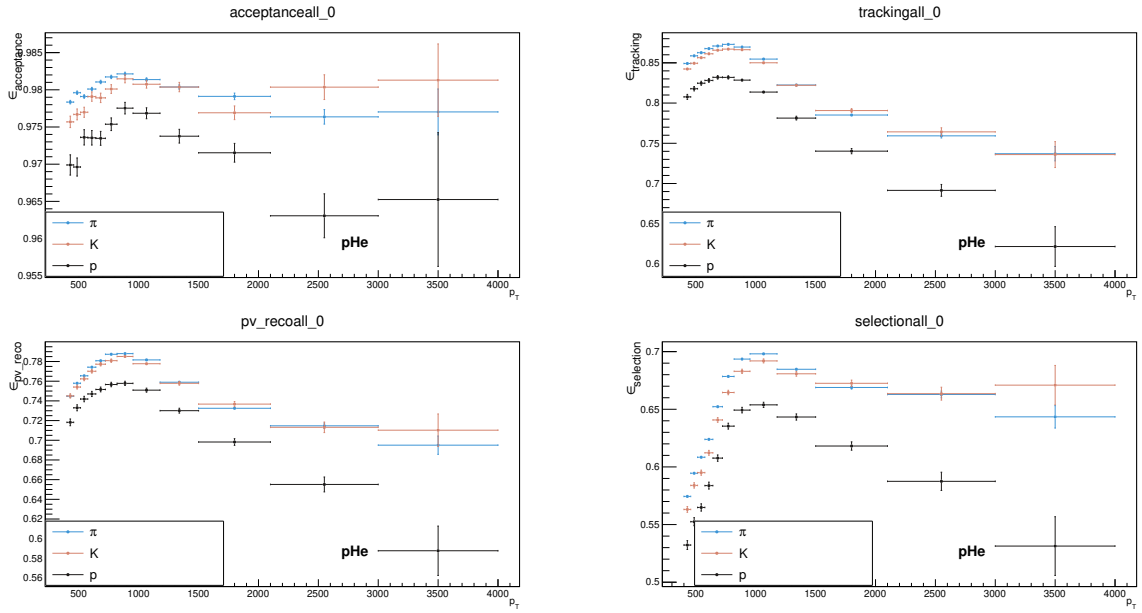


Figure 6.8: 1D efficiencies for π, K and p in pHe weighted simulation in bins of p_T. Top-left: acceptance efficiency for , Kp. Top-right: tracking efficiency. Bottom: pv reconstruction efficiency (left) and selection efficiency (right).

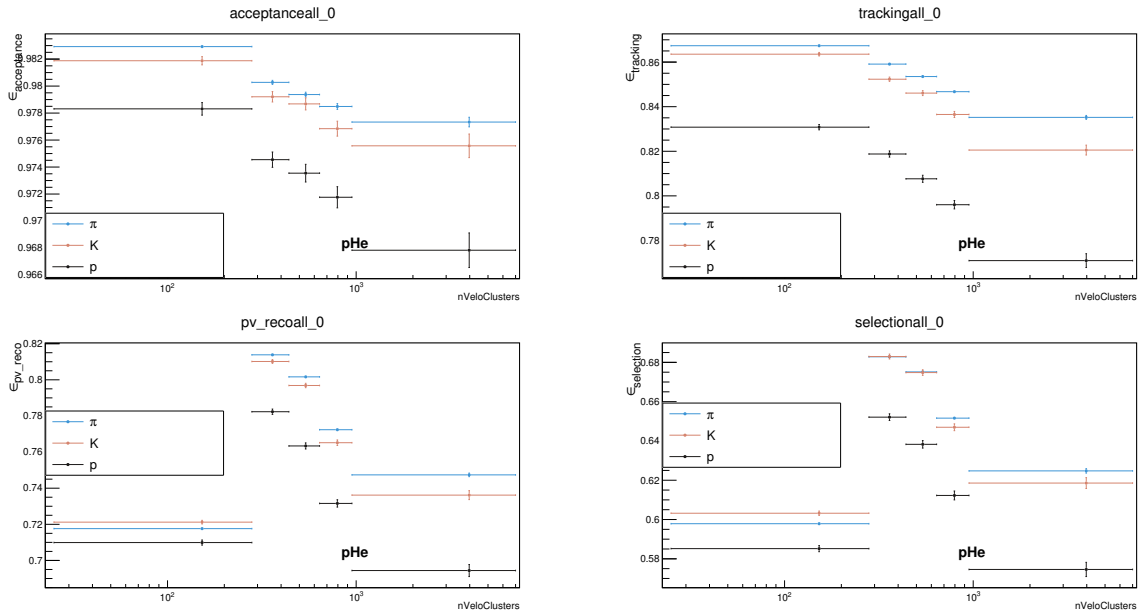


Figure 6.9: 1D efficiencies for π, K and p in pHe weighted simulation in bins of nVeloClusters. Top-left: acceptance efficiency for , Kp. Top-right: tracking efficiency. Bottom: pv reconstruction efficiency (left) and selection efficiency (right).

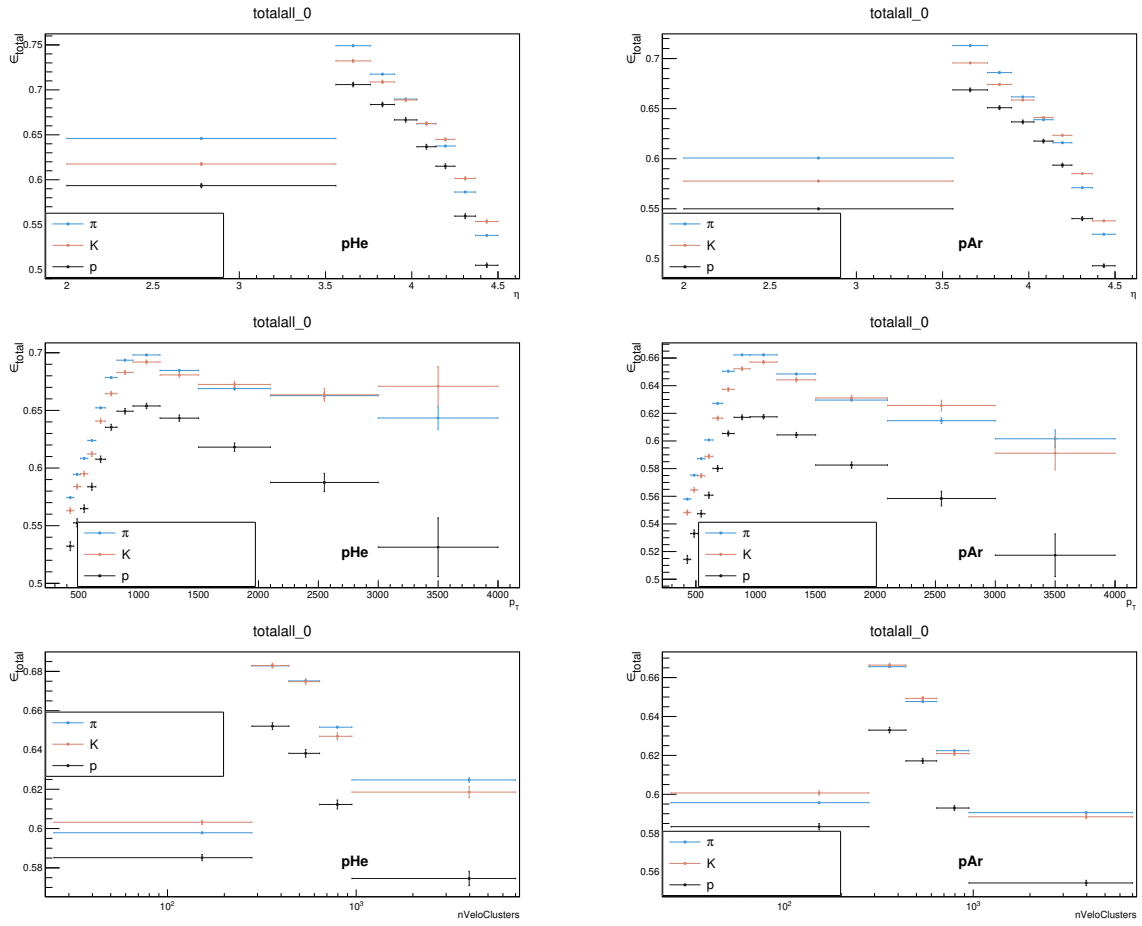


Figure 6.10: Total 1D efficiencies for π, K and p in pHe (left) and pAr (right) weighted simulation in bins of η (top), p_T (middle) and nVeloClusters (bottom).

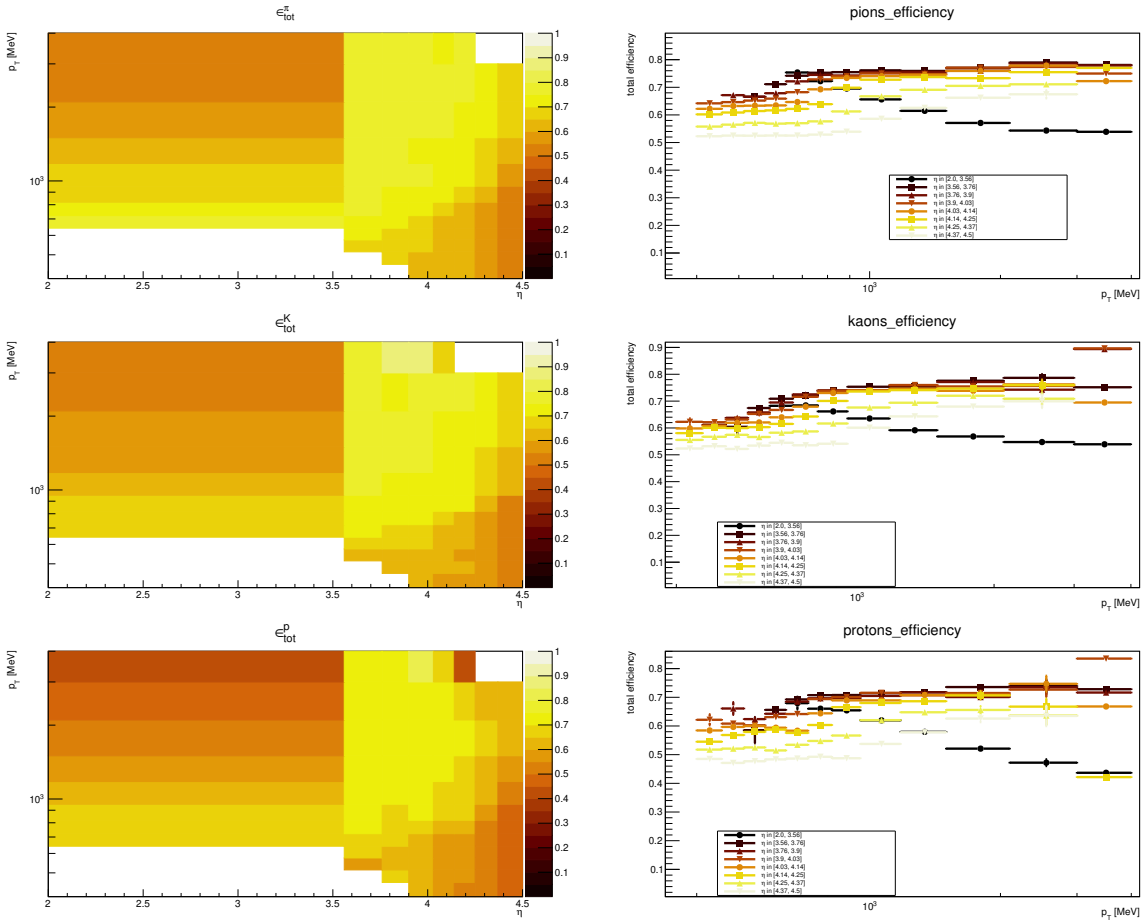


Figure 6.11: 2D efficiencies for each hadron type in pHe in bins of η and p_T obtained with the reweighted simulation.

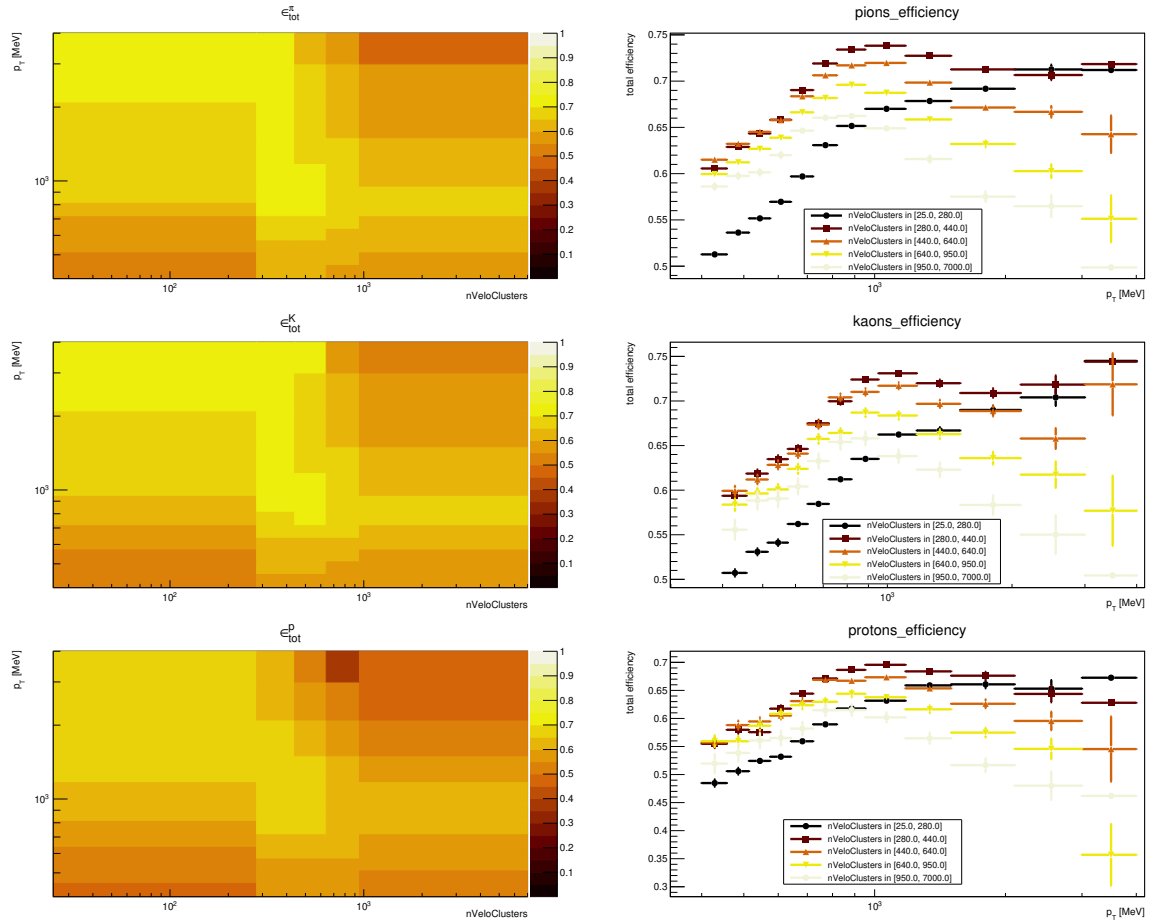


Figure 6.12: 2D efficiencies for each hadron type in pHe in bins of p_T and nVeloClusters obtained with the reweighted simulation.

6.4.2 IP cut and prompt definition

The prompt requirement is studied in the simulation using the IS_PROMPT condition. This variable refers to a track and defines a prompt particle if the mother lifetime is less than 10^{-9} ns. The distribution of prompt components can be disentangled into three main contributions:

- prompt particle;
- non-prompt particle;
- ghost particle: a particle reconstructed due to random hits, but no MC particle is associated to the track.

Each contribution is shown in Fig. 6.13 for tracks passing all the other selections defined previously. The contamination of non-prompt components after the χ^2_{IP} cut is around 1-2%. The ghost component is not considered since it is modeled by a template drawn from simulation during the relative yield extraction.

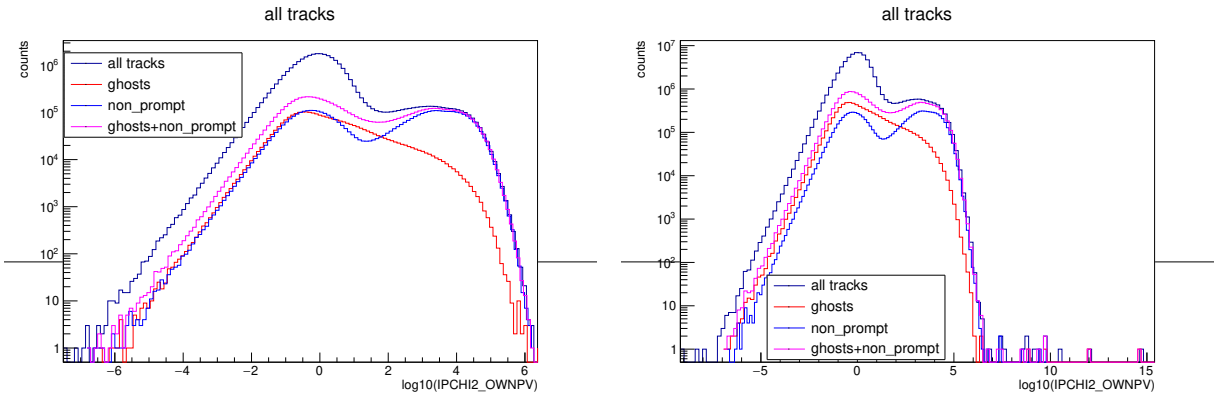


Figure 6.13: Contribution of prompt, non-prompt and ghost tracks as a function of the $\log_{10}$ of χ^2 for pHe (left) and pAr (right).

6.4.3 Trigger efficiency

The events are collected with a minimum bias trigger with very loose requirements. The minimum bias trigger under the beam-empty bunch crossing condition has already been proven fully efficient by studying its performance with a No Bias sample in the pHe data. Since the trigger has the same requirements in pAr, it is safe to assume a fully (i.e. 100%) efficient trigger also for the pAr sample.

6.5 Particle Identification

After applying the selection discussed in Sec. 6.3, the PID classification reduces to separate the pions, kaons, and proton components. Among the LHCb sub-detectors designed for particle identification, the only one capable of proton-pion(kaon) and kaon-pion separation is the RICH system. Although the PID variables consist of the log-likelihood of the PID response of all the sub-detectors, the contribution from Calorimeter and Muon system outputs is negligible for the hadrons of interest in the analysis. The modeling of the PID response in fixed-target data is not consistent with pp data, since the data-taking conditions change considerably (from the events occupancy to the PV range reconstruction). Also, the detailed simulation does not provide a PID that is precise enough to model the PID contributions from data. These issues have been already addressed in Chap.5. The data-driven approach based on the parametrization of the PID marginal pdfs through NN-based Gaussian Mixture Models is then preferred.

6.5.1 Templates Generation

The NN models are available after the training process described in Sec. 5.2.2. For each hadron type, a NN is trained to predict the PID bi-dimensional distribution. The feature variables and the bi-dimensional PID plane used to fit the data is defined at the training step. The NN network trained on the K^0_S decay predicts the PID response for the π and the corresponding template can be obtained. The same principle applies to the GMM network trained on the $\overline{\Lambda}$ for p and the one trained on ϕ for

K. The procedure for the extraction of the template, given the pHe or pAr dataset, consists in applying a binning scheme to the dataset and, on a bin-by-bin basis, repeating the following steps:

- select the data in the given bin interval. For example, in Fig. 6.14 the 2D η-p_T binning has been applied to the pAr dataset.
- preprocess the candidates using the Quantile transformer for the features, the Scikit-learn transformation objects are initialized in the preprocessing step (Sec. 5.2.3), before the network training.
- for each trained NN network (π, K and p):
  - feed the candidates feature variables to the network
  - for each candidate, a pdf in the 2D space is predicted
  - the PID template is obtained by sampling the pdf and applying the inverse MinMaxScaler transformation to the sampled points in the 2D PID plane. The resulting binned distributions are then normalized. The results of such process for each hadron type are shown in Fig .6.15 for the same pAr bin considered in Fig. 6.14.

The binned bi-dimensional distributions obtained are smooth and represent the expected PID response in the 2d plane defined by the target PID variables for the given binned dataset.

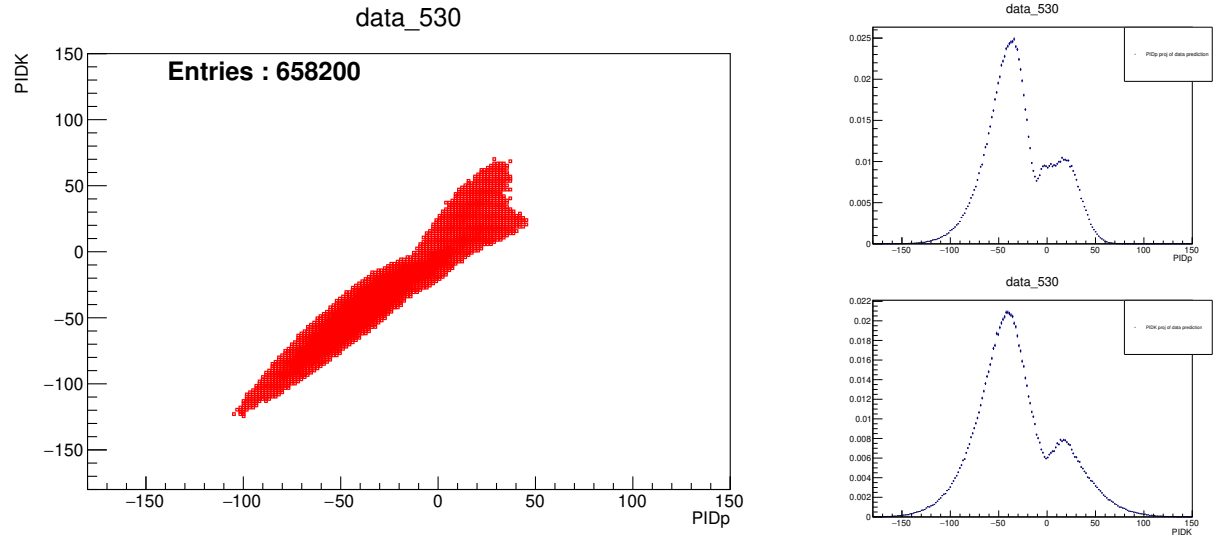


Figure 6.14: Bi-dimendional PID distribution of pAr data in the bin defined by $\eta \in [4, 4.26]$ and $p_T \in [575, 642]$ MeV/c. The plots on the right show the projections onto the two PID variables, separately.

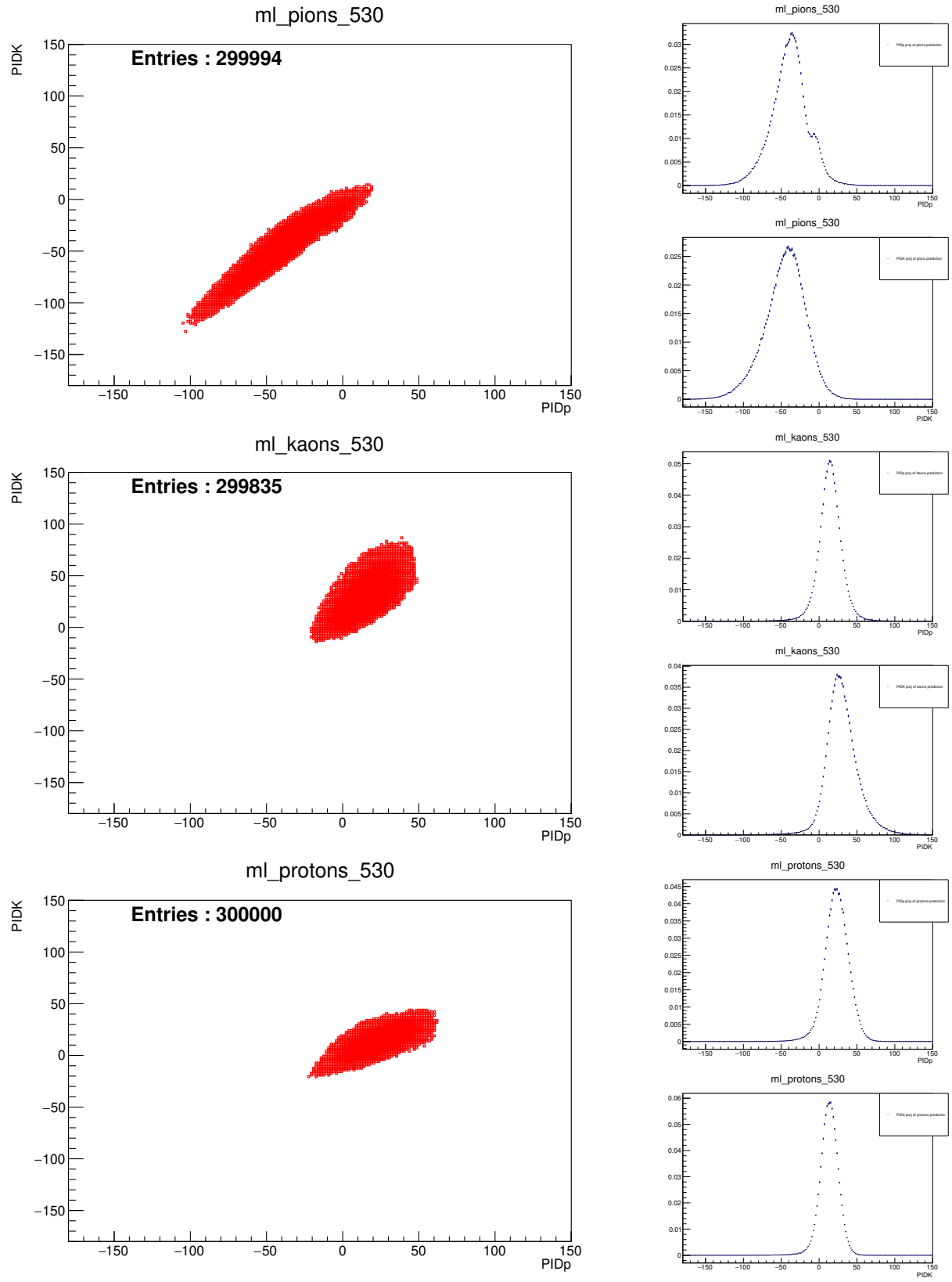


Figure 6.15: Bi-dimendional PID pdfs predicted from pAr data in the same bin shown in Fig. 6.14 and defined by $\eta \in [4, 4.26]$ and $p_T \in [575, 642]$ MeV/c. From top to bottom, the template predicted by the π, K and p GMM, respectively. The right plots show the projections onto the two PID variables.

6.5.2 Yields extraction

The templates are used to fit the PID distributions observed in binned data. The fit is performed on the plane PIDK vs PIDp (or PIDpK) with a joint two-dimensional pdf

$$f(\text{PIDp}, \text{PIDK}) = \sum_h w_h f_h(\text{PIDp}, \text{PIDK}) \, . \tag{6.3}$$

The weights w_h are determined for each bin independently through a maximum likelihood binned fit and represent the relative abundance of the hadron h. The h index runs through π, K, p and ghosts. The ghost template shape is fixed by the simulation prediction, the remaining templates are generated by the NN networks for each hadron type. An example of fits using a 2D binning scheme is shown in Fig. 6.16-6.17. For each fitted bin, the 2D model is shown, as well as the projections along the axis. The Kolmogorov-Smirnov (KS) distance of the model in Eq. 6.3 with respect to the data is calculated after each fit to identify outlier bins with a poor agreement compared to the average fit quality. The yield distribution for each hadron type and the ghost component obtained from the fit is shown in Figs.6.22-6.23 for the pHe dataset in bins of η-p_T and p_T-nvVeloClusters. Only the 2D binning yield distributions are shown since they give a better visualization of the phase-space coverage of the data, but the results are obtained also in 1D (in η, p_T or nVeloClusters) and 3D (datasets are binned with respect all the three variables). The distributions for pAr, not shown, are visually similar.

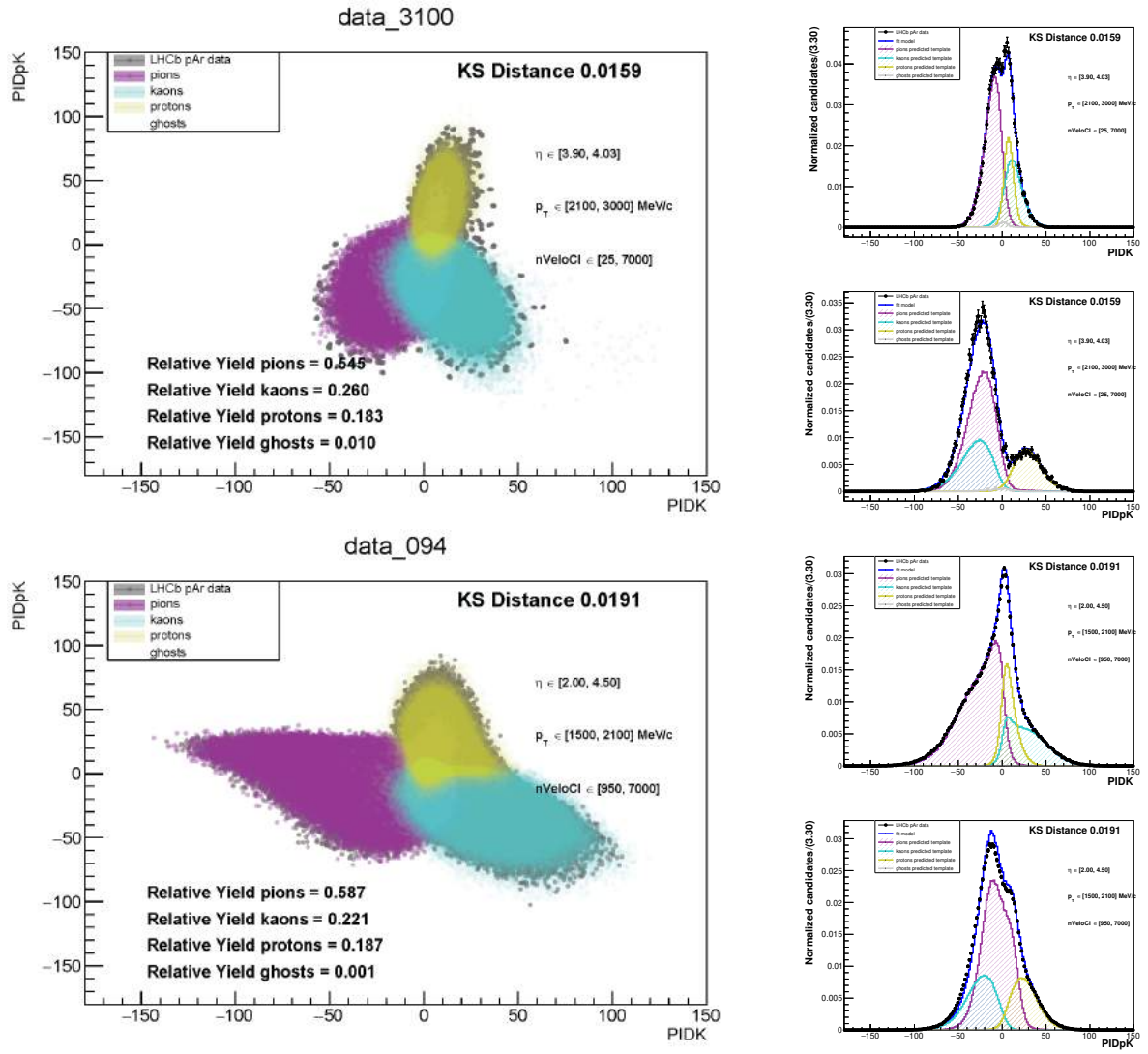


Figure 6.16: Bi-dimendional PID pdfs fits for the pAr dataset using a 2D binning. Top left: fit in one η-p_T bin (integrated over nVeloClusters). Bottom left: fit in a 2D bin in p_T-nVeloClusters (integrated over η). The right plots show the corresponding projections onto the two PID variables, respectively.

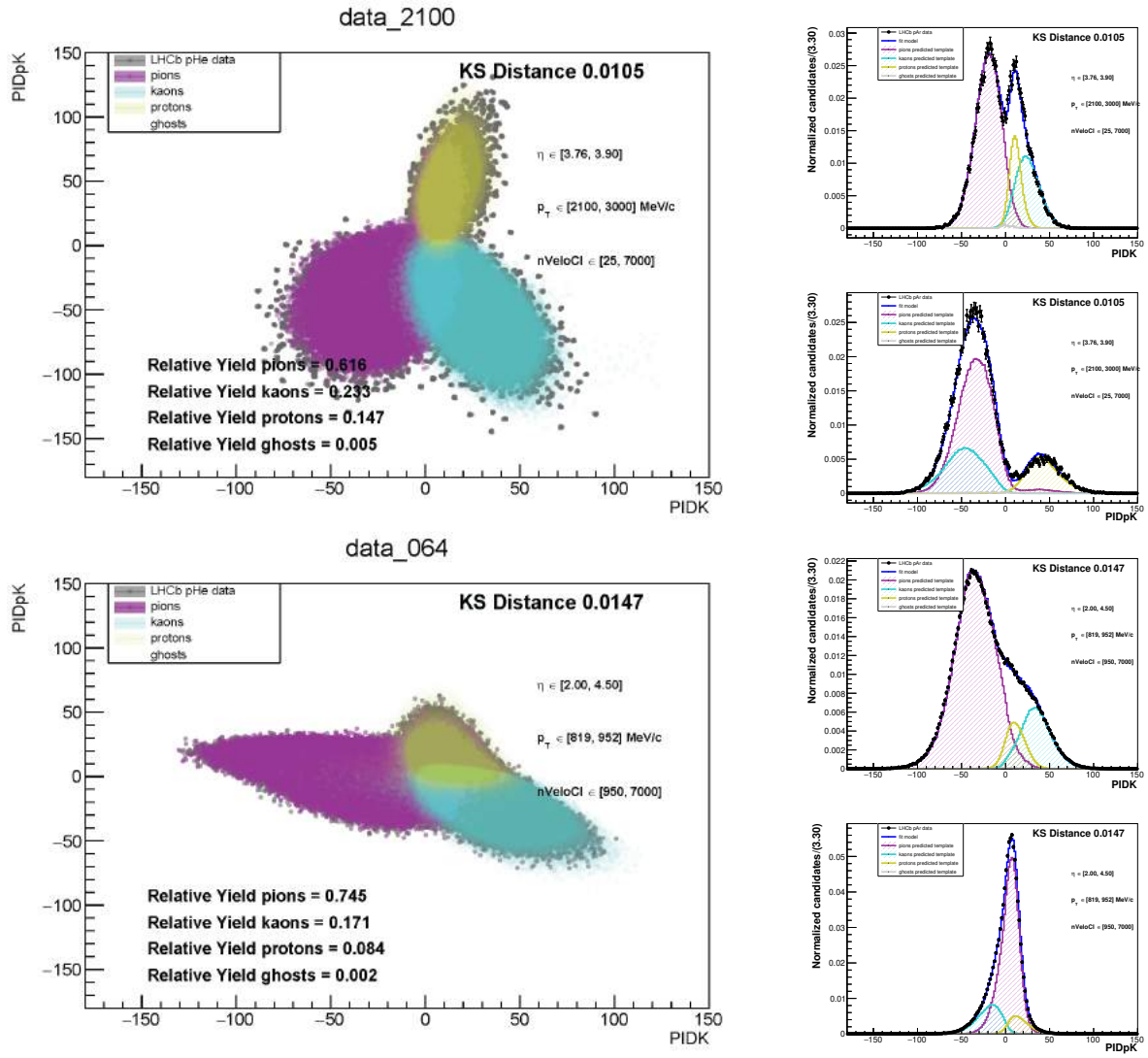


Figure 6.17: Bi-dimendional PID pdfs fits for the pHe dataset using a 2D binning. Top left: fit in one η-p_T bin (integrated over nVeloClusters). Bottom left: fit in a 2D bin in p_T-nVeloClusters (integrated over η). The right plots show the corresponding projections onto the twi PID variables, respectively.

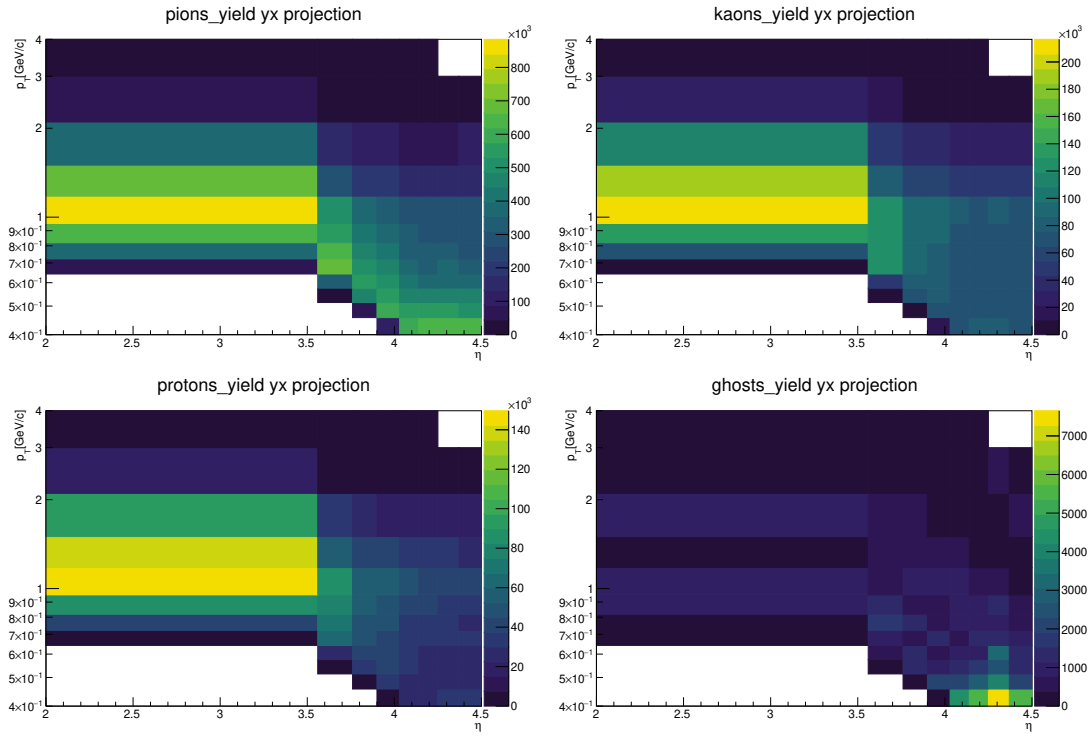


Figure 6.18: Yield of π (top-left), K (top-right), p (bottom-left) and ghosts (bottom-right) in the pHe dataset. Results are shown in the 2D η-p_T binning.

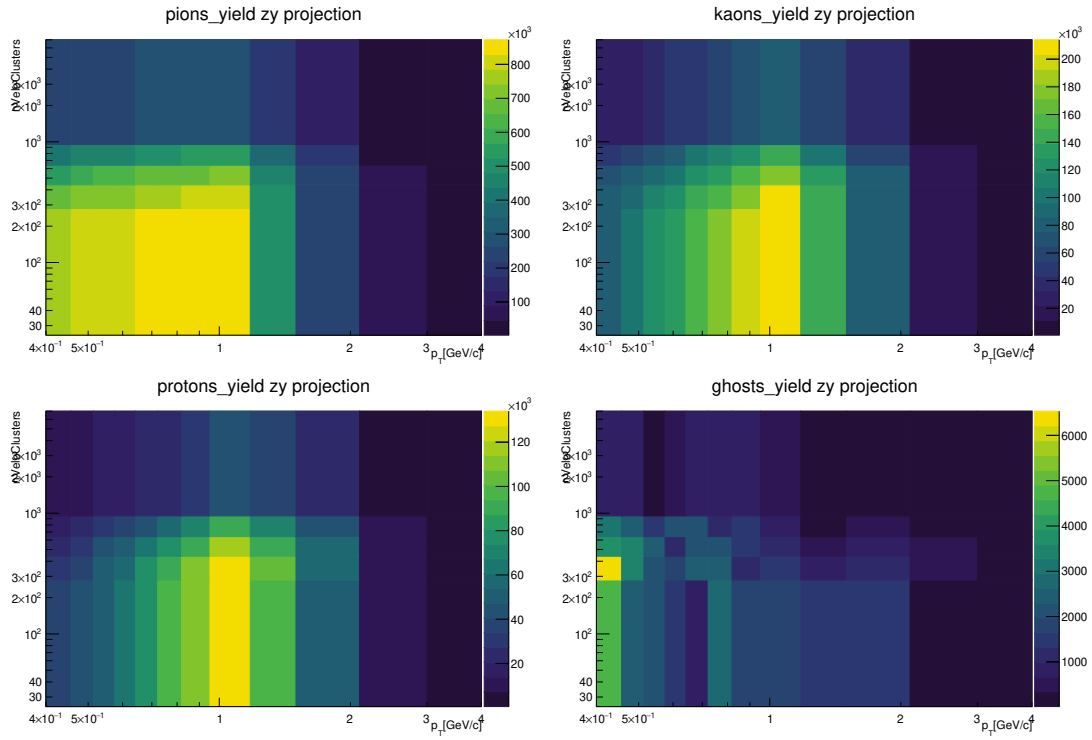


Figure 6.19: Yield of π (top-left), K (top-right), p (bottom-left) and ghosts (bottom-right) in the pHe dataset. Results are shown in the 2D p_T-nVeloClusters binning.

6.5.3 Bin acceptance conditions

Not all bins are accepted and included in the final results. The criteria to accept a bin is defined and applied to every bin independently. The requirements applied to each bin are:

- enough statistics: at least 100 candidates populate the bin;
- a KS distance between model and data in the bi-dimensional PID plane ≤ 0.04. The value is set by considering the mean KS values in the bins with good fit performance. There is a trade-off between the threshold value acceptance for the KS and the PID systematic error associated with the bin, since accepting larger values of KS implies a less defined NN response and relative abundance stability. The KS values per bin in a 2D binning scheme is shown in Fig. 6.20.

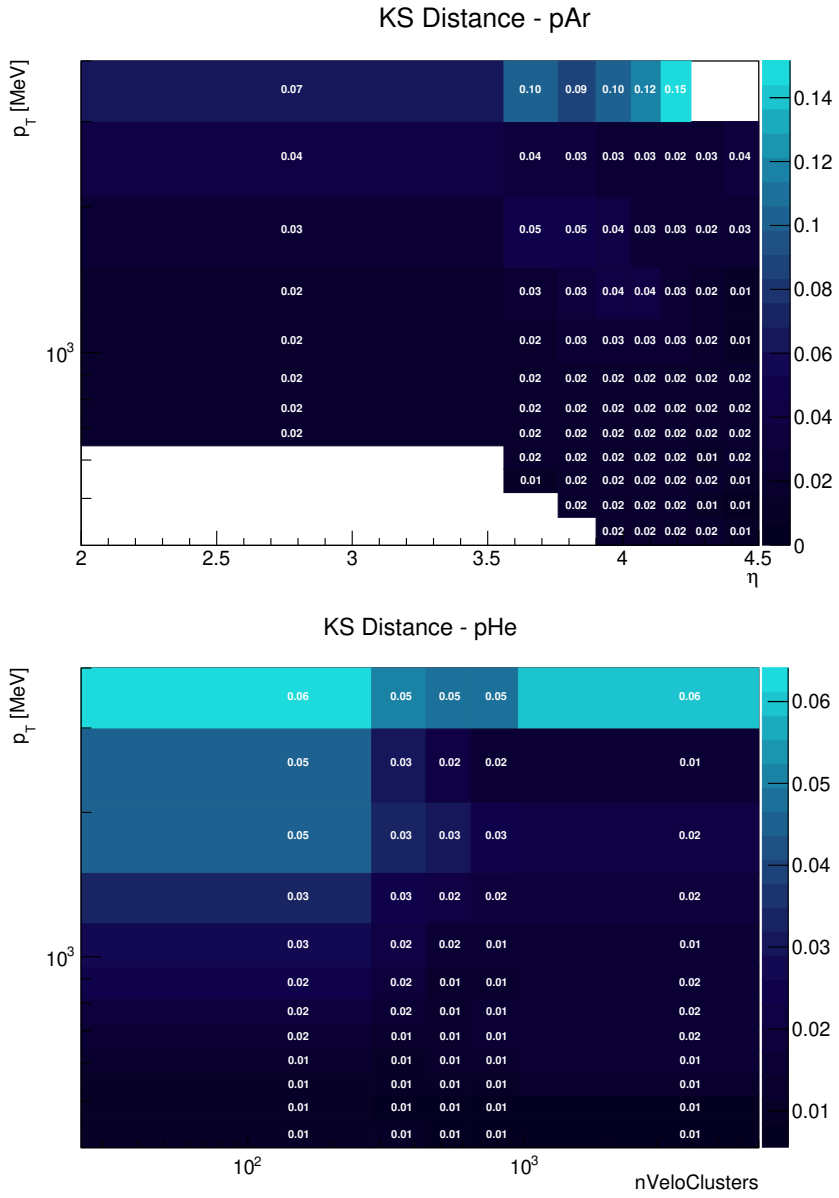


Figure 6.20: KS distances in 2D binning schemes. The values are shown for each bin. At the edges of the covered phase space, the fit quality is worse. On the top, the η-p_T binning is shown for pAr dataset and, on the bottom, the p_T-nVeloClusters binning is shown for pHe dataset.

In addition, the events with $\eta > 4.5$ are excluded from the analysis. The forward region relies on RICH2 for the PID (see Fig. 6.21), and the overlap between the K and p pdfs in the 2D plane is important since the two hypotheses aren't resolved completely and the proton/kaon separation performance is degraded. This is reflected in a worse fit quality and variability of the relative yields between different NN settings, which gives a systematic PID error of about 10%. For these reasons, the high η bins ($\eta > 4.5$) are excluded in the analysis.

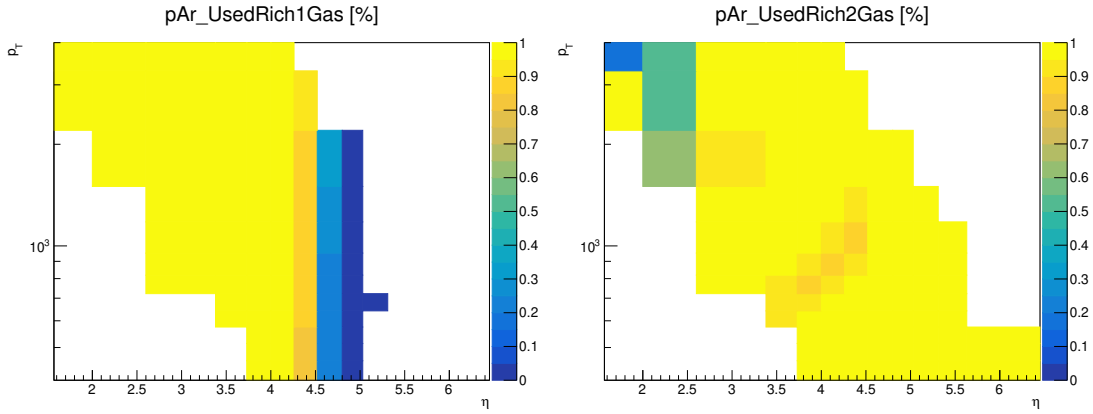


Figure 6.21: Fraction of candidates in the pAr dataset that use the RICH1 information (left) and RICH2 information (right). The same coverage is seen in the pHe data.

Efficiency correction

For the bins passing the mask the candidates yields are corrected by the corresponding efficiencies, which are extracted separately for each bin (Sec. 6.4.1). The raw yield arrays from the fit are masked based on the requirements explained in Sec. 6.5.3 and corrected for the total efficiency which depends by the hadron under consideration. As a consequence, the π, K and p raw yields are divided, respectively, by the π, K and p total efficiencies extracted from the simulation. The result of the masking and efficiency correction is shown in Figs. 6.22-6.23 for the pAr datasets in the usual 2D binning for better visualization.

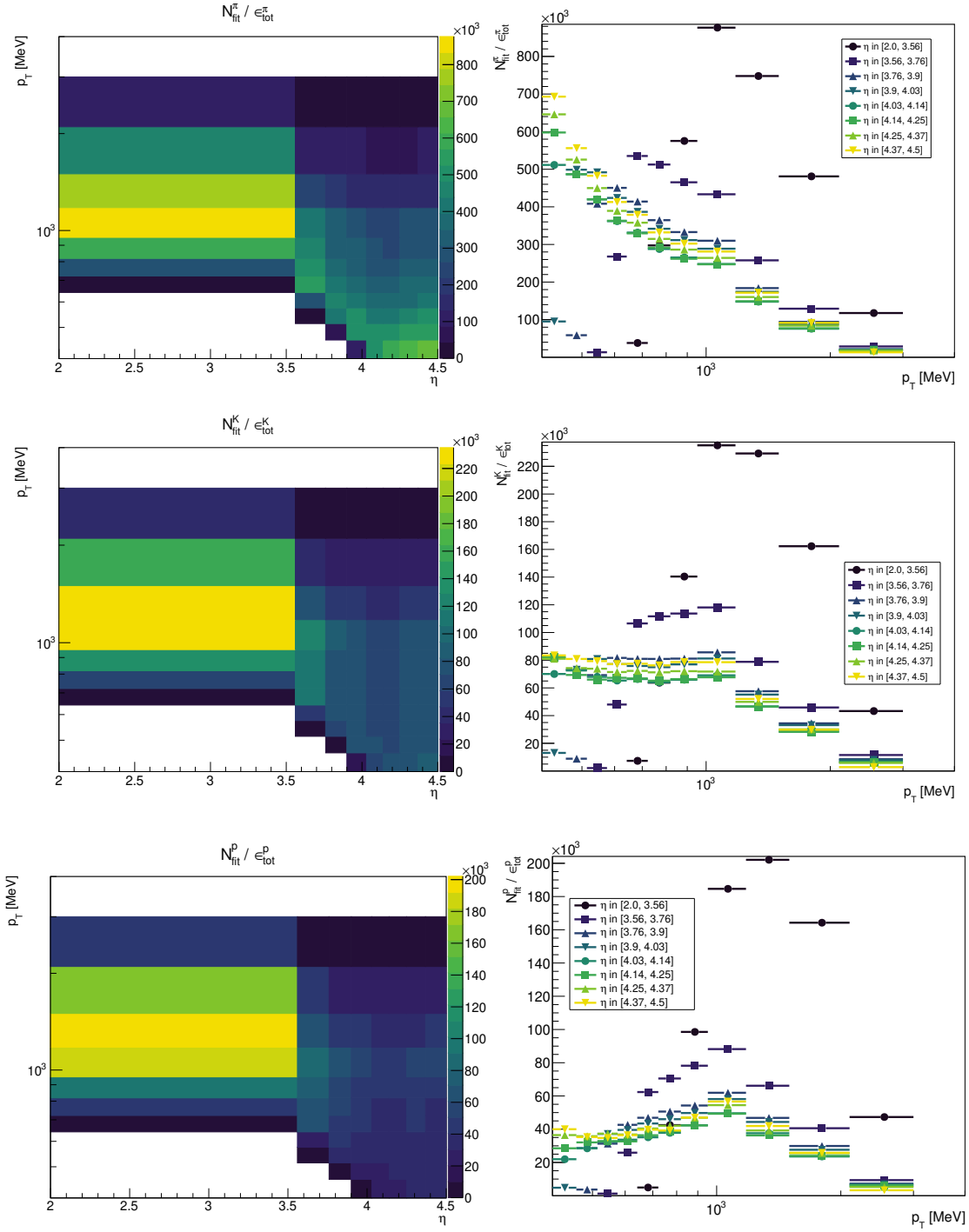


Figure 6.22: Masked and efficiency corrected yield for π (top), K (middle) and p (bottom) in pAr data visualized using the η-p_T binning. On the left the 2D histogram is shown, on the right, the 1D projections are in slices of η.

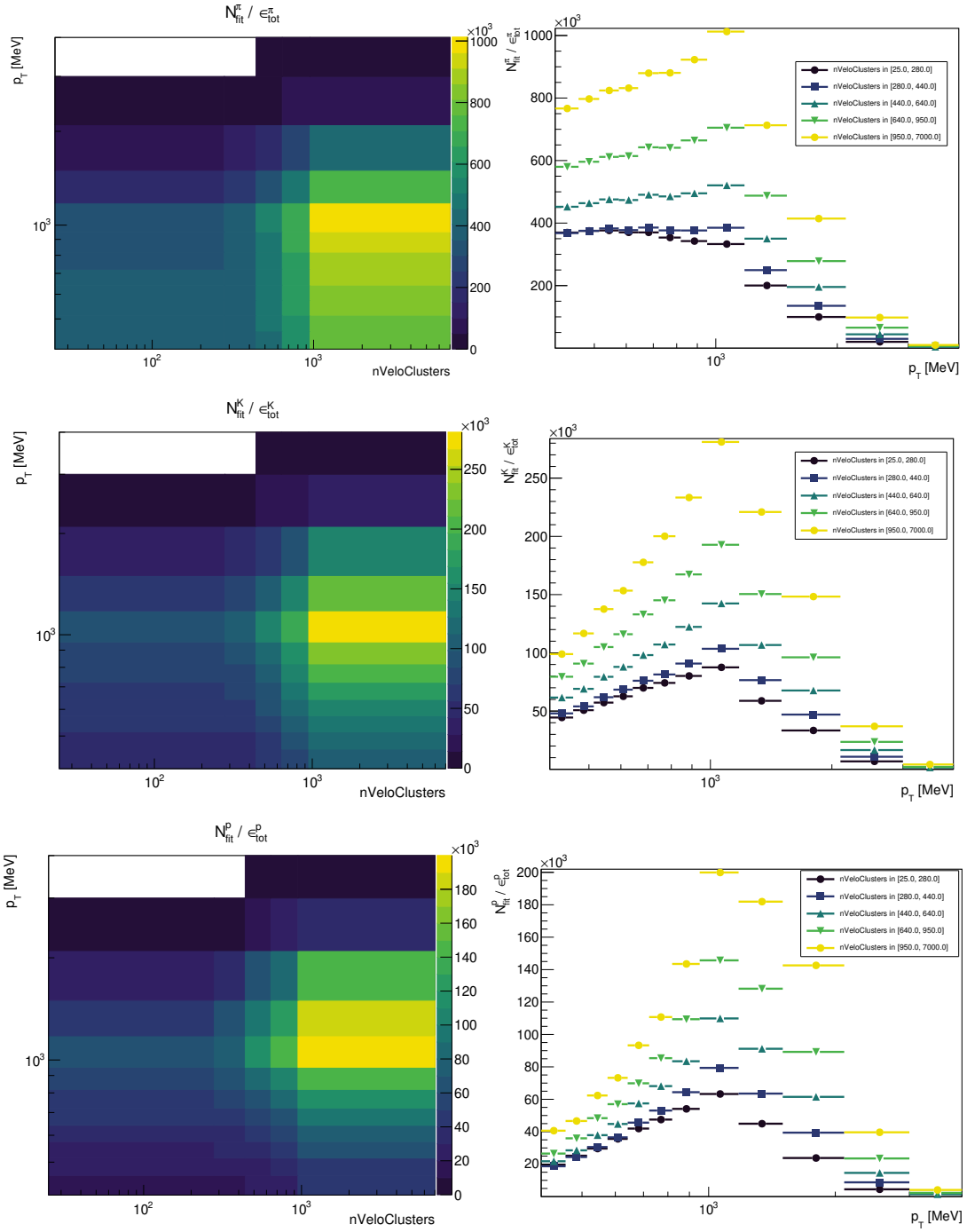


Figure 6.23: Masked and efficiency corrected yield for π (top), K (middle) and p (bottom) in pAr data visualized using the p_T-nVeloClusters binning. On the left the 2D histogram is shown, on the right, the 1D projections are in slices of nVeloClusters.

6.6 Systematics

The systematic error evaluation needs to address the uncertainties which affect the ratios between two hadron types in the same dataset. The main sources of uncertainties which survives the ratios

are:

- NN template estimation: it's important to evaluate the reliability and stability of the neural network response. The shape of the generated templates affects the fit in the bi-dimensional PID space and the yields obtained in each bin. Moreover, the variation of the yield related to the NN template estimate is correlated between the hadron types, since, for example, a modification of the p template implies a variation in the relative yield of the p itself, but also of K and π;
- simulation reweighting: the efficiencies change for each hadron type and this variation has to be controlled in the ratio. The default efficiencies are calculated after reweighting the simulation (Sec. 6.4). A conservative systematic uncertainty is assigned to this procedure by estimating the variation of the yields when the reweighting procedure is not applied;

6.6.1 PID

The extraction of the relative yields from the template fit relies almost entirely on the parametrization of the pdf in the 2D PID plane of the NN trained on the pNe calibration lines. To evaluate this uncertainty, different NN have been trained after varying the network parameters with respect to a reference configuration defined by the following settings for all the 3 networks trained to predict π, K and p templates:

- number of Nodes (nNodes) = 64;
- number of multivariate Gaussian distribution (nGaussians) of the GMM = 32;
- features = η, p_T, nRich1Hits, nRich2Hits, nLongTracks, nVeloClusters, track χ^2/ndof, track FirstStateX Distance Of Closest Approach (doca) to the beam along X-Y-Z, slopes (tx,ty) of the trajectory with respect to x and y;
- target variables = PIDK, PIDpK.

The NNs have been varied from the above configuration according to Tab. 6.6.

The NNs variations are chosen in such a way to evaluate the performance with different networks complexities, by increasing or decreasing the number of Nodes of the hidden layers and the number of Gaussians, thus the total number of NN parameters optimized during the minimization of the loss function in the training process. Also the input features are varied, by introducing the variables usedRich1gas and usedRich2gas. Finally, the target PID variables are varied to include the possible combinations of the log-likelihood hypothesis. The best discrimination in the 2D plane for π is given by the PIDK and PIDp combination since they probe the K and p hypothesis against the π one. Following the same logic, the best K discrimination is achieved with PIDK-PIDpK and the p are better isolated in the 2D plane using PIDp-PIDpK. However, the yields obtained by varying the target 2D PID combinations show that the NN approach for the template extraction and the following relative yields fit are robust enough and no bias is introduced by the target variables choice. The yields as a function of the NN structures listed in Tab.6.6 are shown for a few bins in Fig. 6.24. The yield variation in a given bin depends on the NN structure and on the parameter space covered by the bin itself. No

Table 6.6: Different NN settings trained to evaluate the variation in the yield estimate after the fit. All the three networks used to predict the π, K and p template are trained using the same parameters.

Label	Feature changes	Targets	nNodes	nGaussians
baseline richgas	+ usedRich1gas, usedRich2Gas	PIDK, PIDpK	64	32
lessGauss richgas	+ usedRich1gas, usedRich2Gas	PIDK, PIDpK	64	16
lessGauss	none	PIDK, PIDpK	64	16
lessNodes	none	PIDK, PIDpK	32	32
moreGauss	none	PIDK, PIDpK	64	64
moreNodes richgas	+ usedRich1gas, usedRich2Gas	PIDK, PIDpK	128	32
moreNodes	none	PIDK, PIDpK	128	32
multvars lessGauss	+ nTracks, nSPDHits	PIDK, PIDpK	64	16
multvars richgas	+ nTracks, nSPDHits, usedRich1gas, usedRich2Gas	PIDK, PIDpK	64	32
multvars stdpid richgas	+ nTracks, nSPDHits, usedRich1gas, usedRich2Gas	PIDK, PIDp	64	32
stdpid richgas	+ usedRich1gas, usedRich2Gas	PIDK, PIDp	64	32
stdPID	none	PIDK, PIDp	64	32
pidppk lessGauss	none	PIDp, PIDpK	64	16
pidppk richgas	+ usedRich1gas, usedRich2Gas	PIDp, PIDpK	64	32

particular preference for a NN structure can be observed, so they are all retained for the evaluation of the PID systematic error. The yield in a bin for a hadron type (π, K or p) is assigned as the mean of the yields obtained from the fit by using the templates from the different NN configurations, and the associated systematic PID error is the standard deviation of the distribution of those yields.

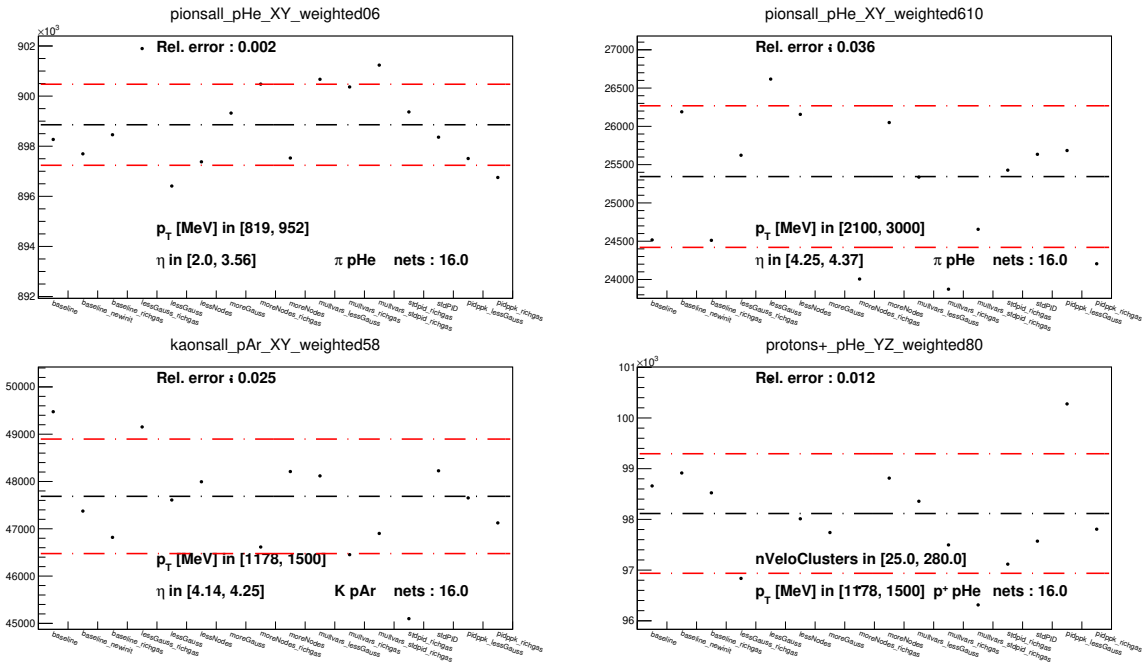


Figure 6.24: Yields obtained for π (top), K (bottom-left) and p (bottom-right) in some bins. Along the X axis, the different NN structures are labeled, and on the Y axis, the corresponding yields extracted from the fit of the templates for each NN structure. The mean of Y-axis values is used as the final yield for the given bin (black line), the standard deviation (red lines at $\pm 1\,\sigma$) of the Y axis values is used to assign a systematic uncertainty related to the PID procedure.

6.6.2 Simulation Reweighting

The efficiencies used to correct the raw yields from the fits are estimated by reweighting the simulation in 4D 6.4. A conservative systematic error associated to the reweighting procedure is evaluated by repeating the yields efficiency correction using the simulation without applying the reweighting. The yields are compared for π, K and p using the two approaches. The relative variation of the yields depends on the bin considered and it might be up to few %, but lower than 1% for most of the pAr bins. In Fig. 6.25 the systematic uncertainty estimated in η-p_T binning is shown for both samples.

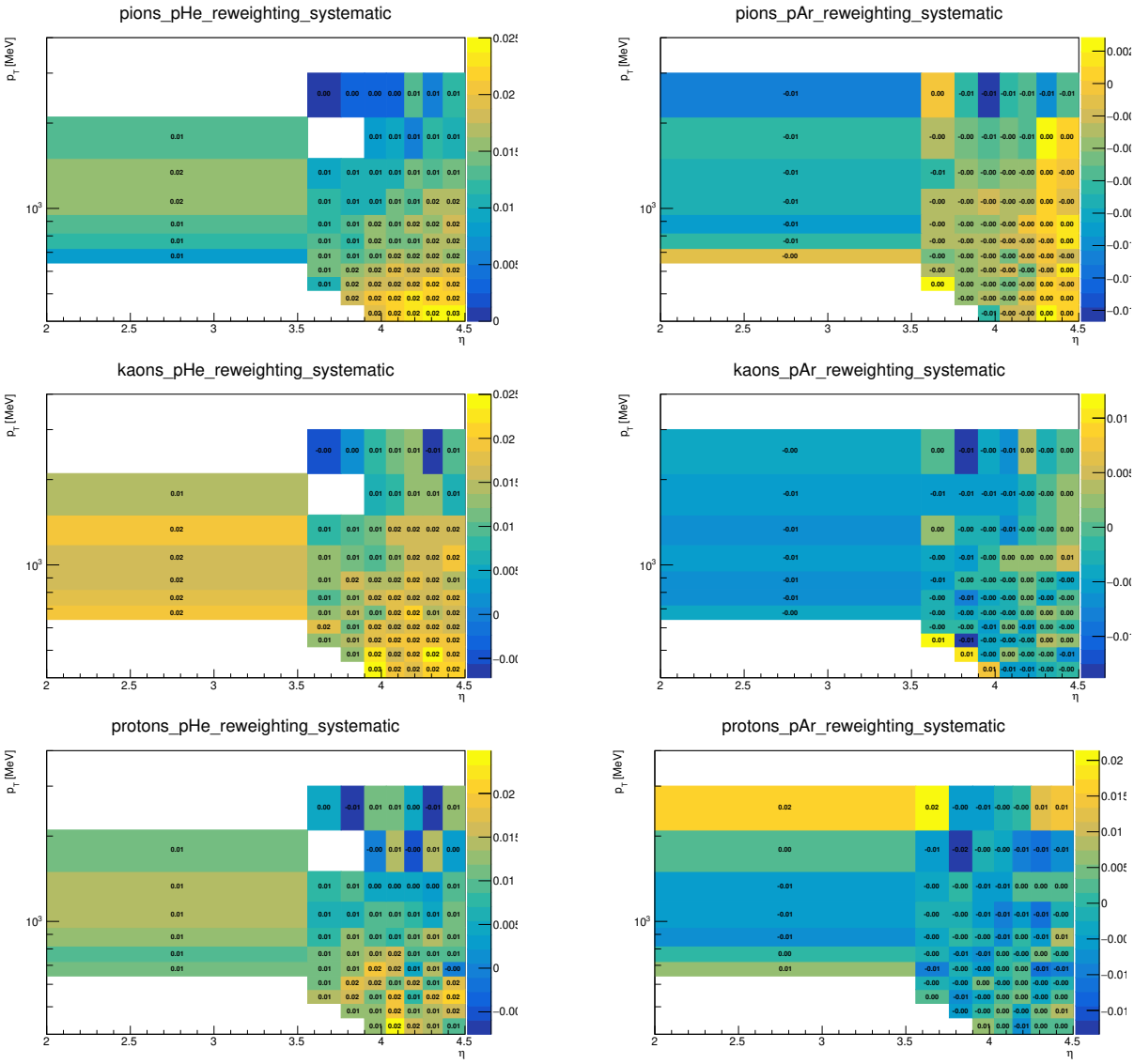


Figure 6.25: Relative error on the particle yields in the η-p_T plane obtained by varying the efficiencies calculation (with and without reweighting the simulation). Only the bins accepted (see Sec. 6.5.3) are shown. From top to bottom, π, K, p yields in pHe (left) and in pAr (right). The pHe simulation is more affected by the reweighting procedure.

6.7 Purity

After the raw yields extraction and efficiency correction, the final yield values are also corrected considering possible contaminations to the prompt yields due to:

- non-prompt particles surviving the χ^2_{IP} cut (Sec. 6.4.2);
- small fraction of $e^{-(+)}$ and $\mu^{+(-)}$, in particular in the bins at low p_T and low multiplicity.

6.7.1 Non-prompt component

A non-prompt component may survive the selections described in Sec. 6.4.2. The contamination from non-prompt particles is estimated from simulation. For the reconstructed and truth-matched π, K and p, the non-prompt component is estimated after applying all the selections described in Sec. 6.3. In Fig. 6.26 the non-prompt fraction in the pHe simulation is shown for each truth-matched hadron type togheter with the relative contamination of non-prompt hadrons in the η-p_T plane. The per-bin yields are corrected by a prompt purity factor calculated using the simulation. Among the three hadron species, the proton is the one which exhibits the highest non-prompt contamination surviving the track selection cuts. A possible explanation for this contamination is protons from decaying Λ and other strange baryons (Σ,Ξ) which do not fall into the prompt definition, having a mean lifetime longer than 10^{-6} ns (see Fig. 6.27).

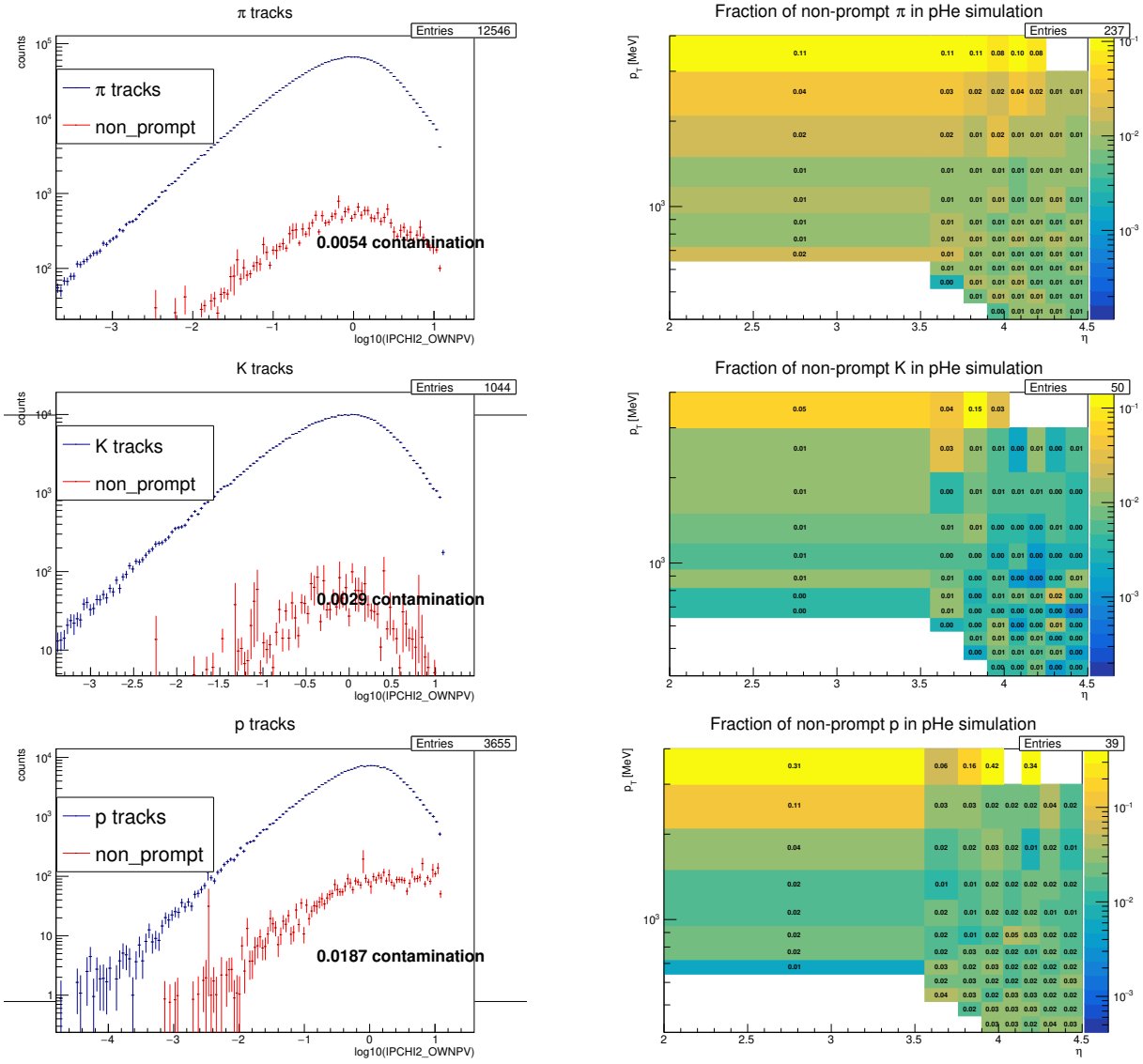


Figure 6.26: Non-prompt fraction after the reconstruction,tracks selection cuts and Montecarlo truth-matching in the pHe simulation for π (top), K (middle) and p (bottom). On the left, the total non-prompt component considering the tracks with both electric charges. On the right, the non-prompt fraction is bins of η-p_T. The results for the pAr dataset are compatible and not shown here.

6.7.2 Fraction of $e^{-(+)}$ and μ

The final correction addresses the electrons and muons which are reconstructed and survives the track selection. The relative yield is studied in each bin by requiring the simulation non-prompt truth condition (but without including the ghost component which is already modeled in the fit). The pHe simulation study is shown in Fig. 6.28. Muons and electrons have a PID response compatible with the π one (this is shown in the left plot of Fig. 6.28), this is expected due to the small difference in mass between π and μ. As a result, the $e^{-(+)}$ and μ contamination correction is applied to the π yields only. The prompt $e^{-(+)}$ and μ fraction is negligible ($\sim 10^{-5}$ of the total reconstructed tracks passing the selection).

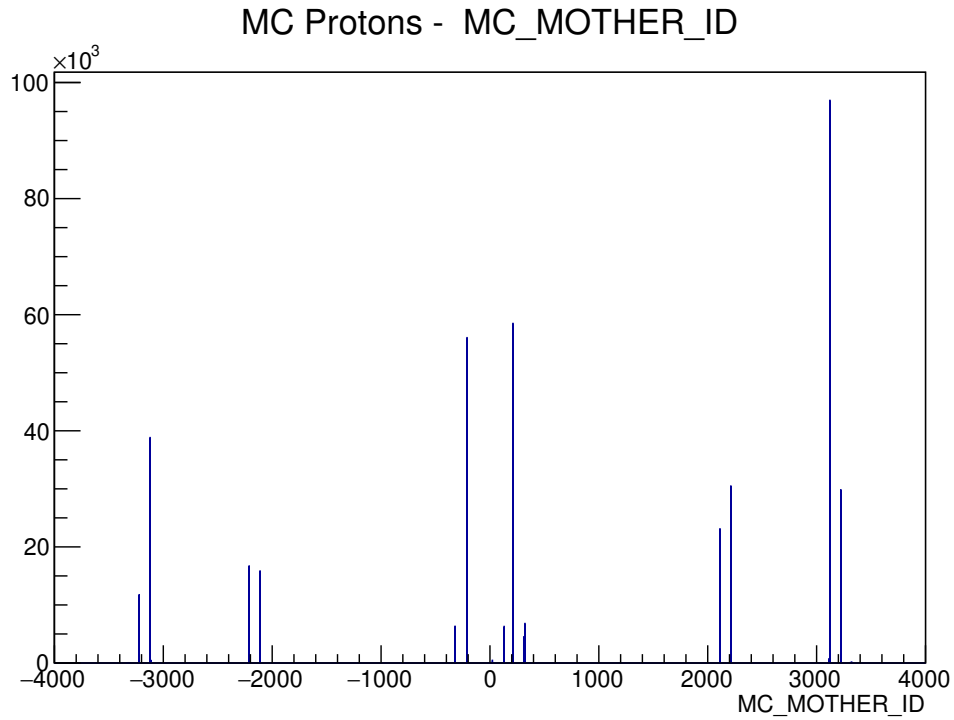


Figure 6.27: Mother True ID information of the non-prompt reconstructed protons in pHe simulation passing the tracks selection. The bins with high counts for values higher of lower than 3000 are due to Λ (ID = 3122) and Σ^+ (ID = 3222) and their antiparticles.

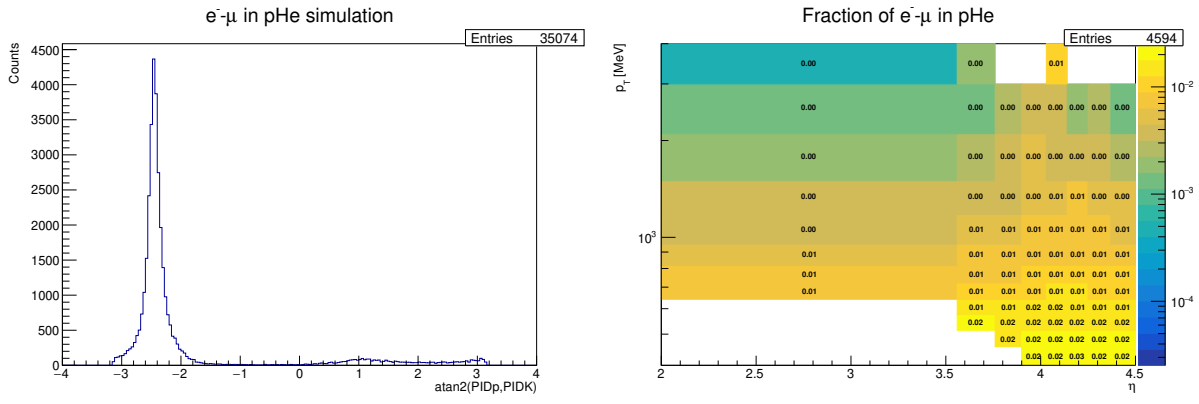


Figure 6.28: Electron and muon contamination in the pHe simulation reconstructed tracks passing the selection. On the left, the atan2(PIDK, PIDp) variable of the truth-matched electron and muons tracks, the main peak corresponds to a PID compatible with the π hypothesis. On the left, the relative yields of e and μ in bins of η-p_T.

Chapter 7

Results

The results of the light hadron production analysis are organized based on the different binning schemes adopted. All the yields and ratios are obtained for each particle type for all the tracks as well as for positive and negative tracks, separately. The majority of the results are plotted against p_T, since it is the most sensitive variable to CNM effects. The 1D results are compared for the different hadron charges (positive vs. negative) and collision types (pAr vs. pHe). In this chapter, only a few selected results are shown. The full collection of results is presented in Appendix A.

The yield ratios of interest are :

- single ratios for a fixed gas target: p/π, K/π and p/K. The heavier hadron is placed at the numerator. The ratios are provided also separately for positively and negatively charged particles;
- double ratios: the single ratios defined above for pHe and pAr are divided to directly compare the two collision systems.

All the error bars comprehends both systematic and statistical errors.

7.1 Single Ratios

7.1.1 1D binning scheme

The simplest 1D binning scheme has been applied to the η and p_T variables. A clear deviation between the pAr and the pHe single ratios is observed along the entire η range, see Fig. 7.1. These results, which are integrated over p_T, indicate a significant difference in the hadron production mechanisms between light (pHe) and heavier (pAr) reaction systems. In particular, all three hadron ratios appear enhanced in pAr compared to pHe collisions. A more interesting feature is revealed when considering the same ratios as a function of p_T, see Fig. 7.2. In this case, at least when considering the p/π and p/K ratios (i.e. the baryon-meson ratios), the difference between the two collisions systems is clearly more pronounced in the high-p_T region ($p_T > 1$ GeV/c). This evidence could suggest the onset of the Cronin effect at $p_T \sim 1$ GeV/c. The observation of a more homogeneous and less pronounced deviation in the case of the K/π ratio compared to that of the p/π and p/K ratios might be related to the so-called *baryon-meson anomaly*, observed by previous experiments. It is also interesting to

compare the p_T dependence of the three ratios separately for the two different charges and different collision systems (see Fig. 7.3). In general, and for both the pHe and pAr systems, the ratios for the positive particles are all significantly higher (by about a factor of 2) than those for the corresponding negatively charged particles. Another relevant difference regards the relative values of the K/π and p/π ratios, which are much closer to each other for the case of positive hadrons, whereas they are observed to be quite different for the case of negative hadrons, with the K^-/π^- ratios about twice as large as the $\bar{p}/\pi^-$ ones. Finally, for both collisions systems, the ratios of negative hadrons appear to reach the maximum at a somewhat (~ 0.5 GeV/c) smaller value of p_T compared to the case of positive hadrons.

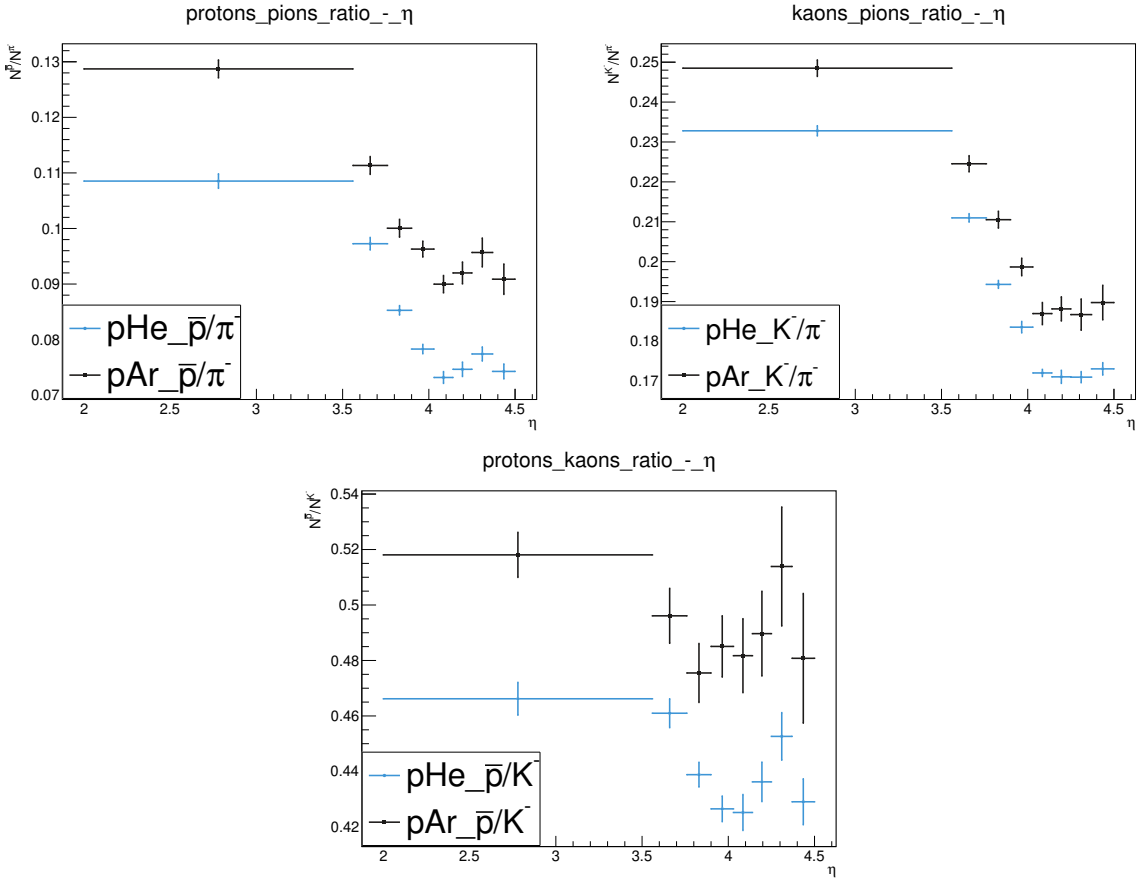


Figure 7.1: Negative charged tracks single ratios as a function of η. The ratio for both gas targets is shown.

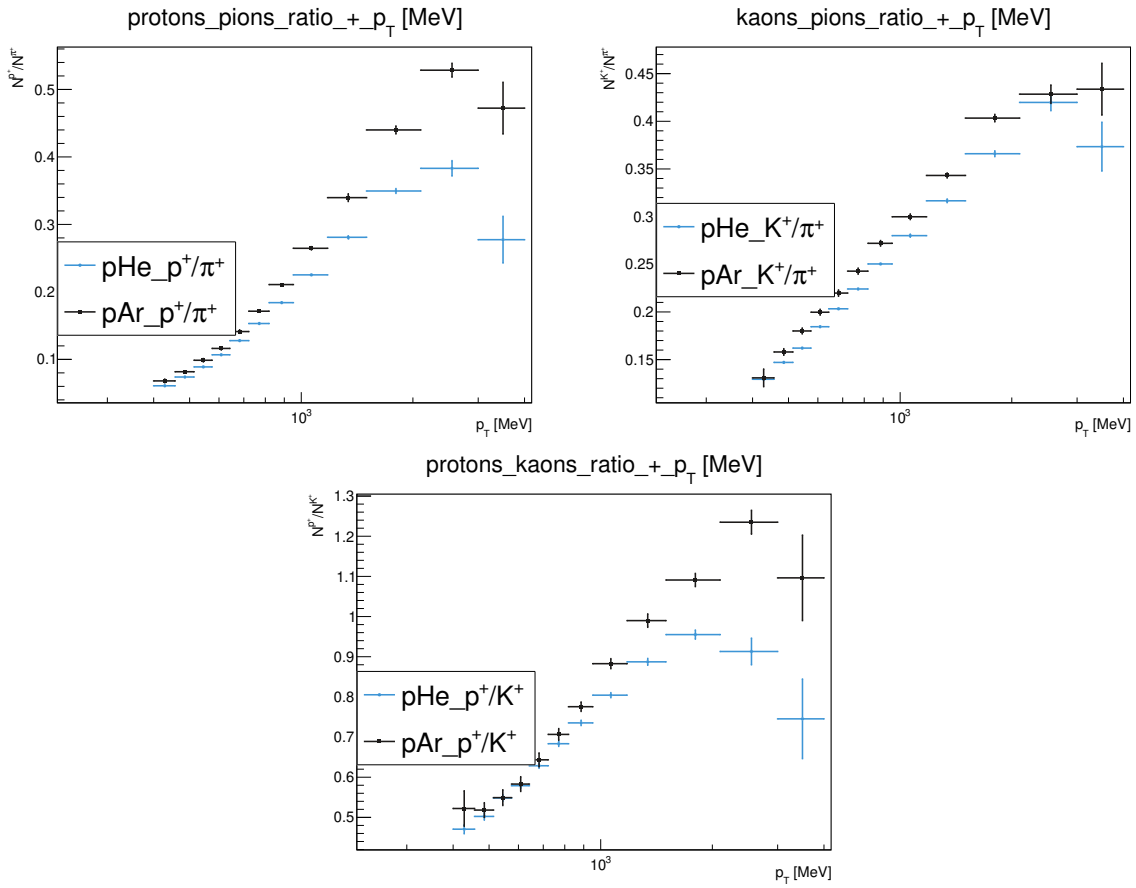


Figure 7.2: Positive charged tracks single ratios as a function of p_T. The ratio for both gas targets is shown.

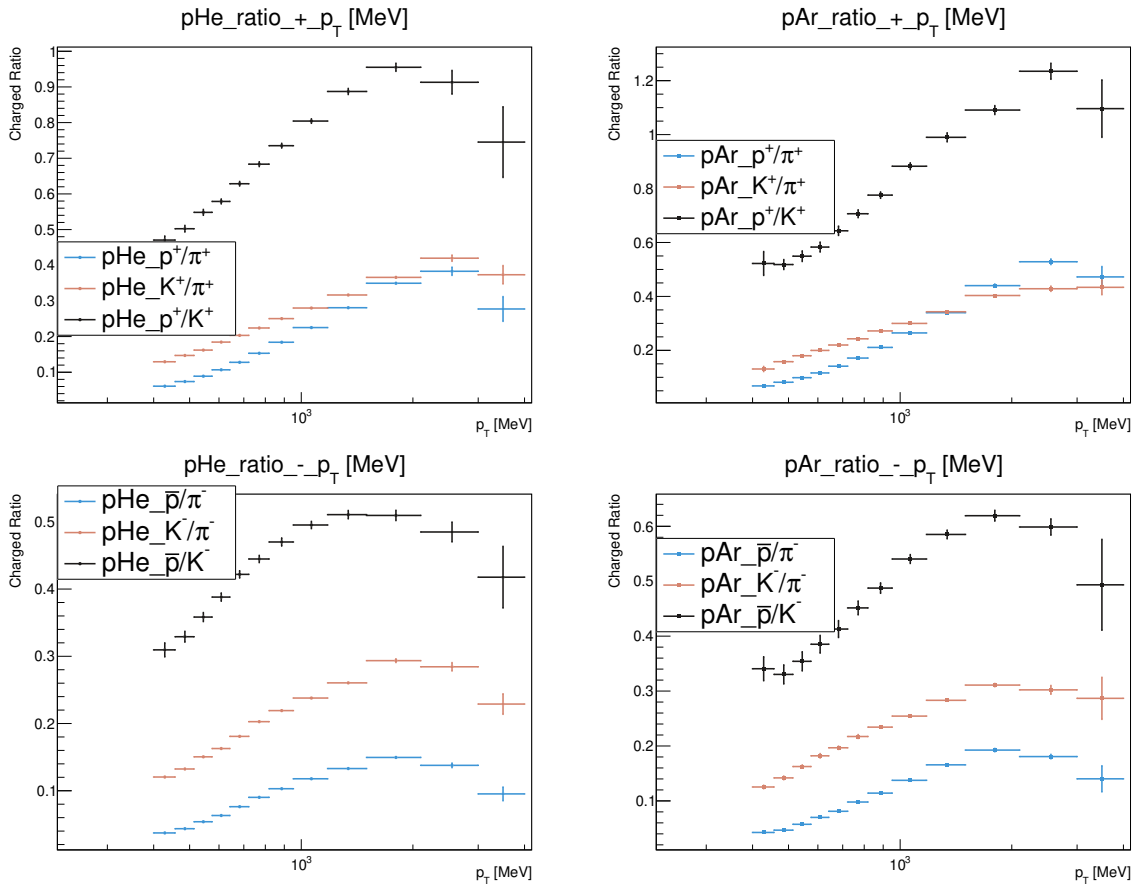


Figure 7.3: Positive and negative charged tracks single ratios as a function of p_T. The trends for every single ratio are shown for the same target.

7.1.2 2D binning scheme

The single-ratios trends are studied against p_T by varying η or the occupancy. In general, ratios involving protons exhibit a clear dependence on both η and nVeloClusters along the entire range of p_T, while the K/π ratios begin to become sensitive to those variables only at $p_T > 2$ GeV, where the uncertainties are also more important (see Fig. 7.4). The positively charged p/π and p/K ratios reach higher values in the low η region, corresponding to the backward region in the center of mass frame. This feature is reversed for the corresponding negatively charged hadron ratios, shown in Fig. 7.6, where one observes that at high p_T the low-η bins correspond to lower values of the ratios, instead. Also the difference between the pHe and pAr samples is interesting, with the latter having a significantly higher abundance of p compared to π or K. A relevant feature of the p/π and p/K single ratios vs. p_T in bins of nVeloClusters (integrated over η) is the different slope observed for the different nVeloClusters bins. Specifically, while at $p_T < 1$ GeV/c high-multiplicity bins correspond to smaller ratios compared to the low-multiplicity ones, the opposite is observed at $p_T > 1$ GeV/c, with a clear cross-over at $p_T \sim 1$ GeV/c, i.e. approximately the same p_T region where the 1D ratios suggest the onset of the Cronin effect. This effect is particularly visible in the p/K ratios for both collision systems (see Fig. 7.5) and also noticeable for the ratios of negatively charged hadrons (see Fig. 7.7)),

although the cross-over seems to occur at higher p_T (around 2 GeV/c) in this case.

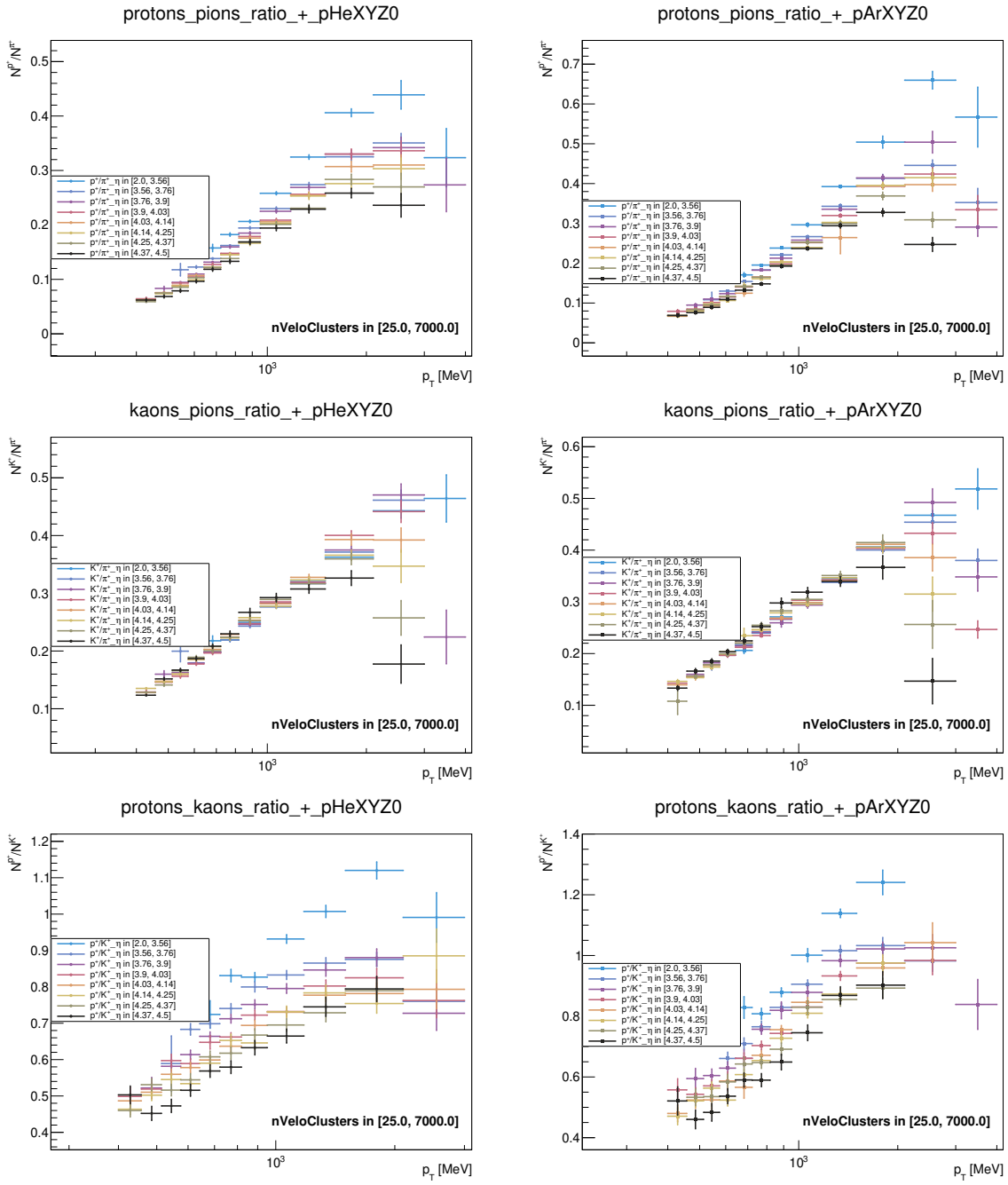


Figure 7.4: Positively charged tracks single ratios in the η-p_T binning scheme as a function of p_T. On the left, the single ratios for the pHe dataset, on the right, the same ratios for pAr.

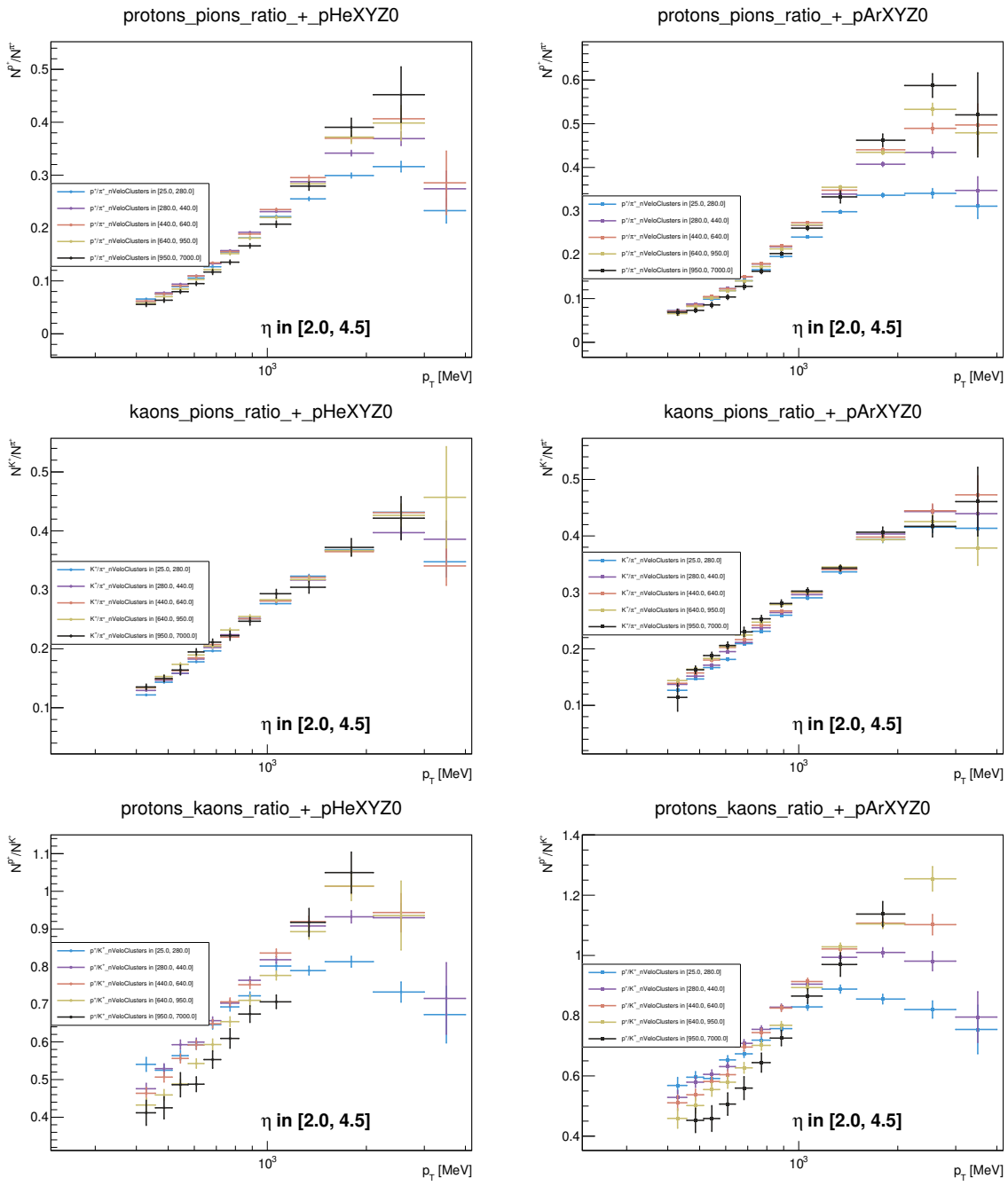


Figure 7.5: Positively charged tracks single ratios in the p_T-nVeloClusters binning scheme as a function of nVeloClusters. On the left, the single ratios for the pHe dataset, on the right, the same ratios for pAr.

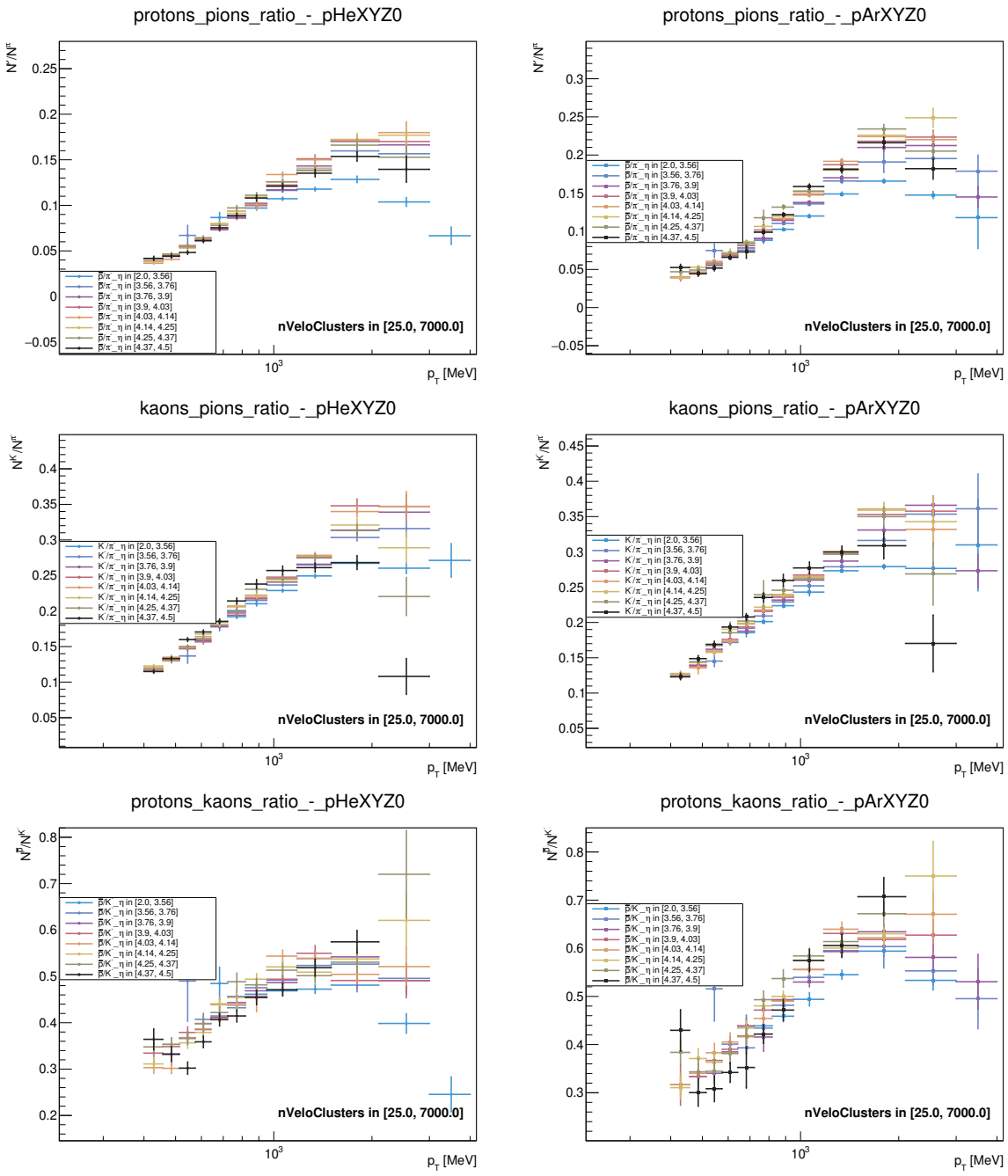


Figure 7.6: Negatively charged tracks single ratios in the p_T-η binning scheme as a function of p_T. On the left, the single ratios for the pHe dataset, on the right, the same ratios for pAr.

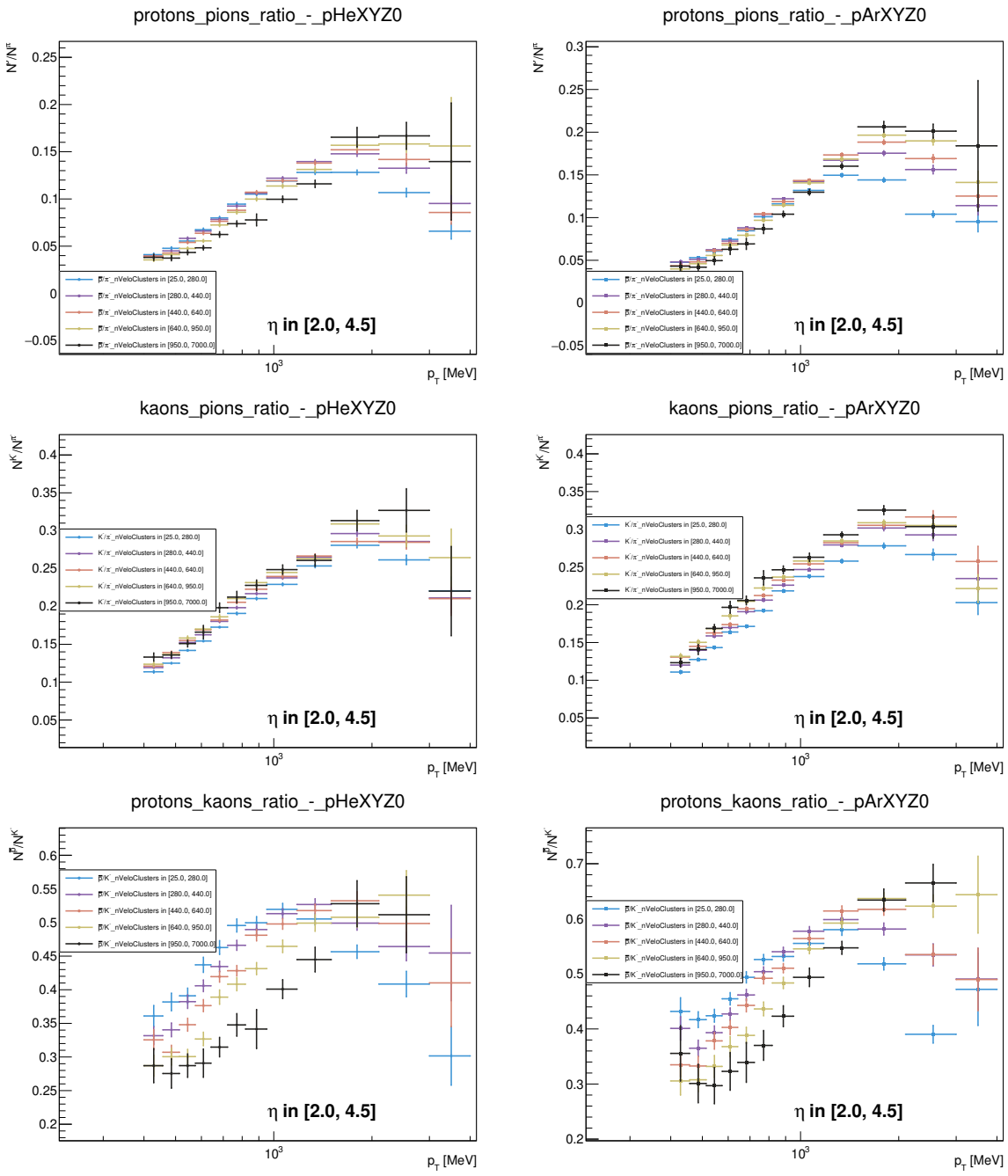


Figure 7.7: Negatively charged tracks single ratios in the p_T-nVeloClusters binning scheme as a function of p_T. On the left, the single ratios for the pHe dataset, on the right, the same ratios for pAr.

7.2 Double Ratios

The double ratios defined in Eq.6.1 allow to directly evaluate the modifications on the relative particle productions (of p and K with respect to π and of p with respect to K) in a heavier system with respect to a lighter one (pAr and pHe, respectively, in the present case) without requiring the knowledge of the integrated luminosities of the two data samples.

The extracted 1D double ratios are found to significantly differ from unity, revealing larger nuclear-

matter effects in pAr than in pHe collisions. This deviation from unity, which is expected due to the much larger volume of the Ar nuclei, strongly depends on p_T. Specifically, the p/π and p/K double ratios exhibit a clear rise with p_T, starting at around 0.7 GeV/c (Fig. 7.8). On the other hand, no clear rise is observed for the K/π double ratio, which instead shows a rather flat distribution above unity for the entire p_T range.

The double ratios for the 2D binning scheme does not shown a clear dependence from η, however, a depence from nVeloCluster can be observed in the p_T-nVeloClusters binning (see Fig.7.9). The remaining ratios are shown in Appendix A. Note that, in the double ratios, the PID systematic uncertainty contributes significantly since the yields of the particles are highly correlated due to the fit strategy. This, combined with the limited statistics of the datasets, results in too large error bars in the 3D binning scheme. For this reason the double rations in 3D are not provided here.

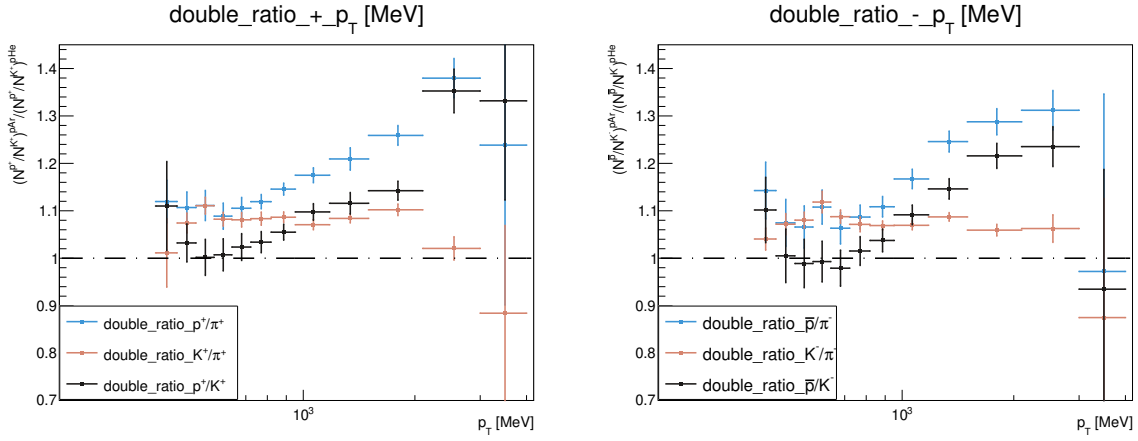


Figure 7.8: Double ratios for positive (left) and negative (right) charged tracks as a function of p_T.

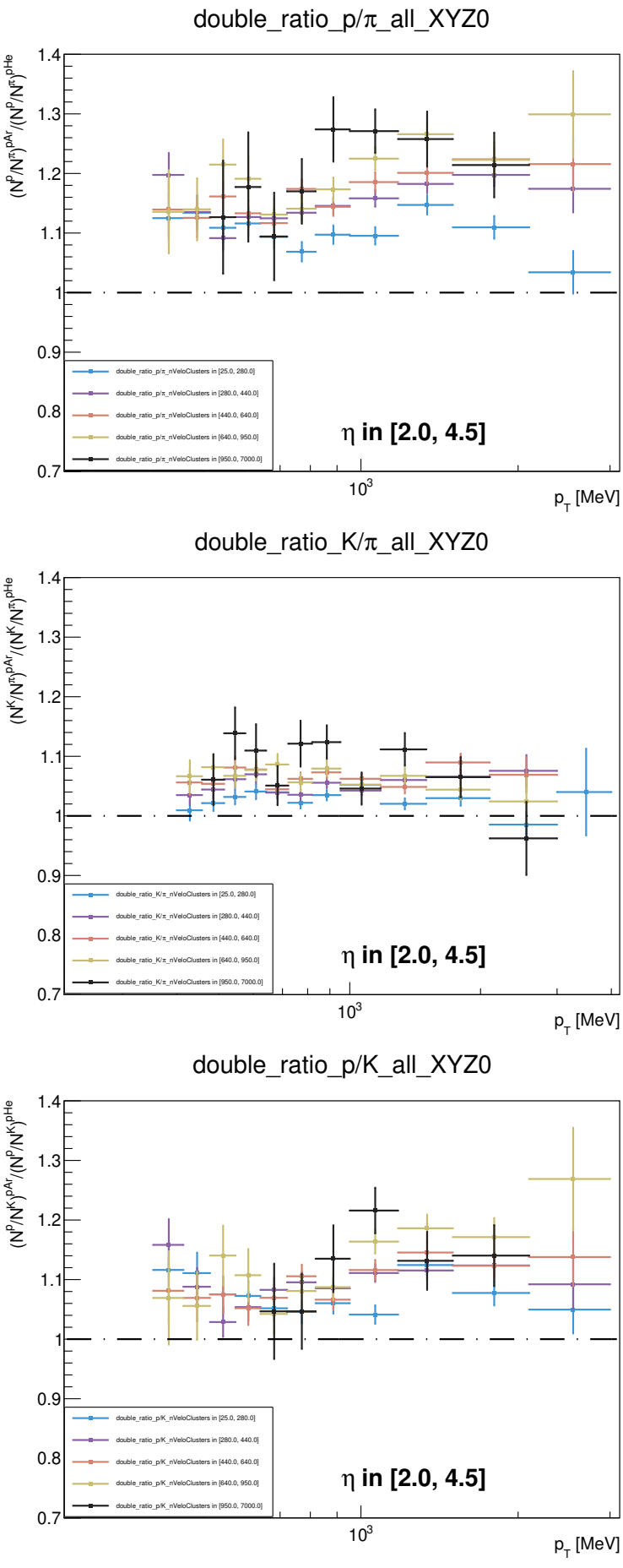


Figure 7.9: Double ratios for particle without charge discrimination as a function of p_T and nVelo-Clusters.

Conclusions

During my PhD I covered different activities related to the LHCb experiment, involving both hardware and software aspects. I contributed to the Rich system Upgrade 1 throughout these years. In the beginning by finalizing the EC quality assurance, giving a strong contribution to the development of the test station, and collecting and organizing the data. The EC tested during the ECQA are now installed in the RICH detectors for Run3. The data collected by the ECQA also gave the opportunity to study in detail the performance of the MaPMTs and a new source of noise (SIN) which contributed to a related paper [63]. The last year of PhD was mostly focused on the commissioning of the RICH detectors in the LHCb control room during Run3. I've worked with the LHCb ECS while contributing to the data-taking and calibration studies for the RICHes. My contribution involved data analysis, debugging and developing configuration and alignment procedures, with a particular focus on coarse and fine time alignment. On behalf of the RICH group, I was responsible for the luminosity studies to exploit the RICH data for online luminosity monitoring and calibration, both at ECS and HLT levels. This work required preliminary simulation studies in the Run3 configuration and real data-taking. The RICH took part in various μ scans, and the latter one was used to temporarily validate the RICH counters and their linearity with respect to luminosity. The RICH can now provide anode currents values and hits in low occupancy regions which can be used for luminosity monitoring. This latter work will be followed by more studies once the LHC restarts in 2023. As a result of the commissioning activities, the RICH detectors, at the end of 2022 and before the LHC winter shutdown, operated in the best configuration achievable, time-aligned, and integrated into the LHCb ECS for global data-taking. The Cherenkov angle resolutions obtained with data collected at the end of 2022, despite the preliminary status, show an improvement in the RICH1 performance compared to its Run2 predecessor and RICH2 already provides comparable values and it is approaching the expected performance.

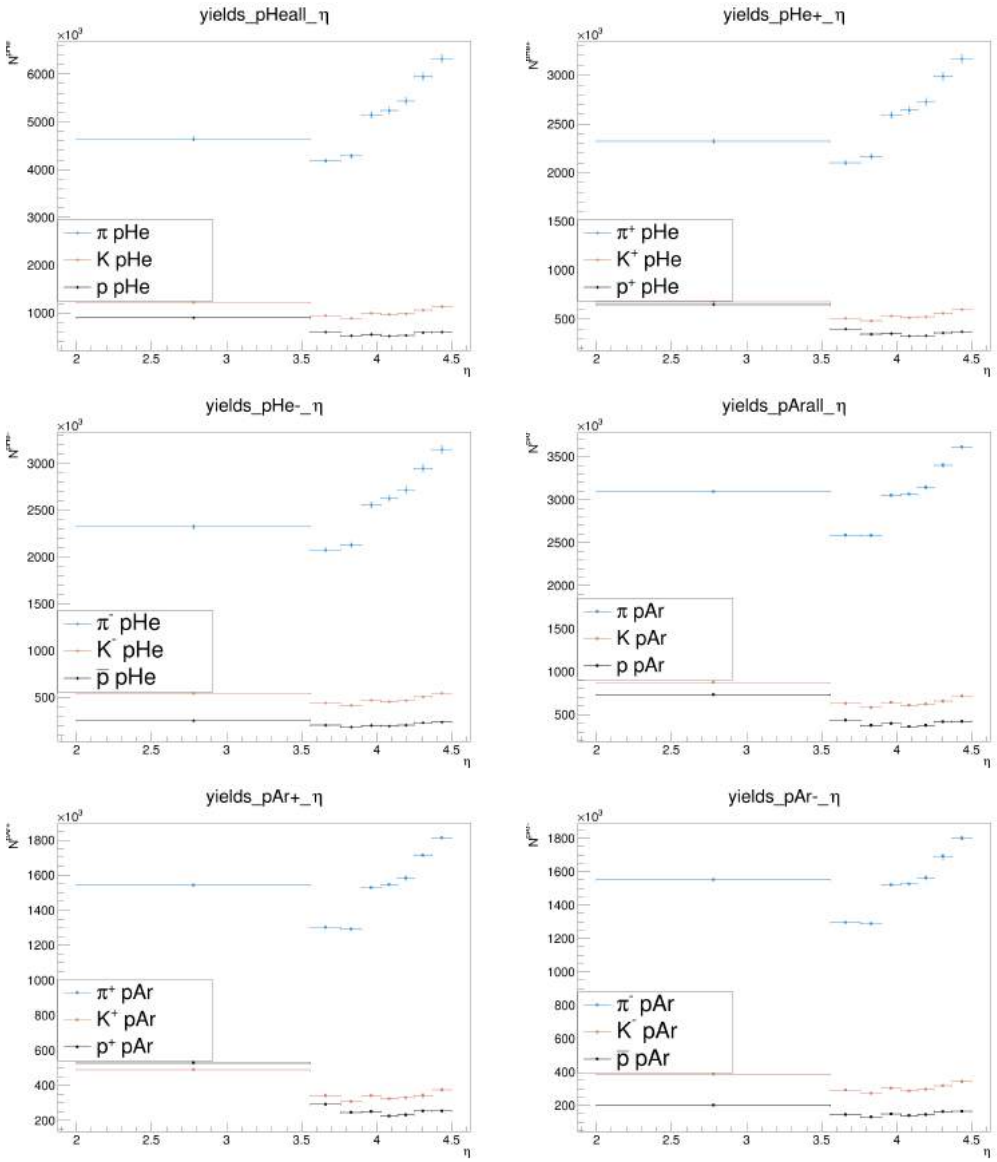
In parallel with the hardware activities for the RICH, I've been involved in studies with fixed target data collected with the SMOG system. The fixed target mode a is a unique configuration in LHCb and also the data presents important peculiarities with respect to the standard pp collider mode data. To overcome the limitations of the small statistics available in Run2, I've contributed to the development of a PID tool using a data-driven NN-based gaussian mixture model to parametrize the PID response of the RICH detectors. A related paper has already been published [67]. The PID of the RICH has been exploited already in a few analyses using SMOG data, including the present analysis on light hadron production. The aim of the analysis is to search for signals of Cold Nuclear Matter effects by comparing the production of prompt light charged hadrons in two different collision system, a light one (pHe) and a heavier one (pAr), at a center-of-mass energy of $\sqrt{110}$ GeV in the backward

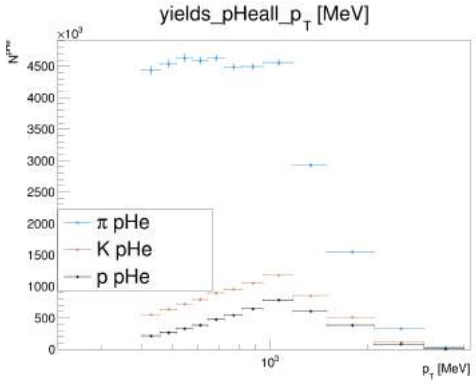
pseudorapidity region in the center-of-mass frame. The main physical observables of the analysis are charge-separated single and double ratios of light hadrons (pions, kaons and protons) as a function of the transverse momentum p_T and the pseudorapidity η of the particle, as well as of a variable sensitive to the multiplicity of the event (nVeloClusters).The difference between the two collisions systems is found to be more pronounced in the high p_T region ($p_T > 1$ GeV), an observation compatible with the onset of the Cronin effect, as already observed by other experiments. The double ratios used to directly compared the two colliding systems differ significantly from unity, revealing larger nuclear-matter effects in pAr than in pHe collisions. The CNM effects are typically studied as a function of p_T or of the event multiplicity, a way to infer the centrality of the collision and, in turn, the number of nucleons involved. A multi-dimensional approach, such as the one adopted in the present analysis, which allows for a simultaneous study of the physical observables as a function of the most relevant variables, has the advantage of providing a more complete information and to reveal features of the data which would otherwise appear diluted or hidden in a mono-dimensional analysis of the same data. Examples are, e.g., the different slopes of the single ratios observed as a function of p_T for different bins of the multiplicity variable and the way the double ratios scale with the multiplicity in the high-p_T region, a feature particularly visible in the case of the p/π double ratios. In the future, these results will be integrated with luminosity measurements and a Glouber model describing the multiple interactions among the colliding nucleons, to obtain the corresponding Nuclear Modification Factors, which can be more directly compared with the available theoretical models.

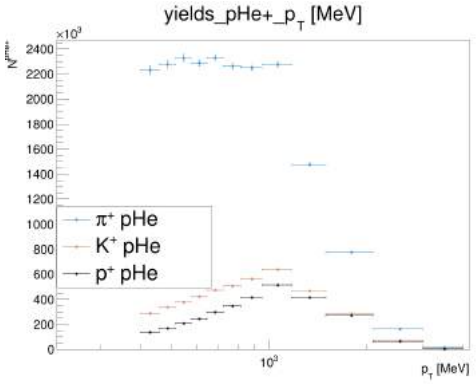
Appendix A

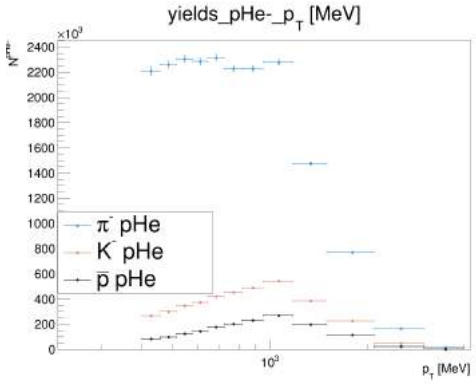
The final result are shown here organized by binning scheme and single or double ratios. The limited statistic of the datasets influence the total uncertainty in the 3D binning scheme, in which case the double ratios are dominated by the errors and thus not shown. All the result are shown as a function of the charge of the hadron, including also the case in which no charge discrimination in applied. These results underline the presence of modification of the light hadron productions due to nuclear effects affecting the pAr dataset in comparison with pHe collisions.

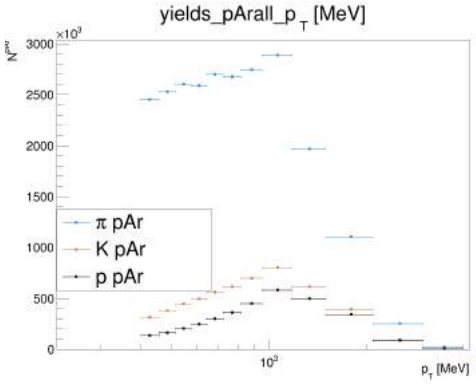
.1 1D yields in η and p_T

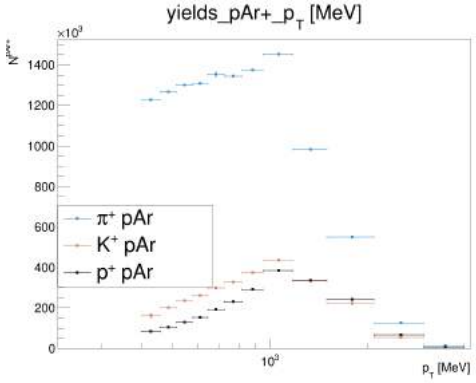


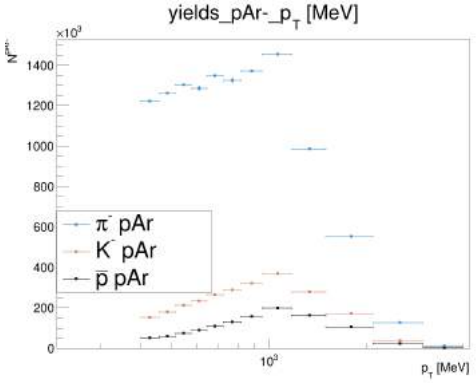

yields_pHeall_p_T [MeV]
π pHe
K pHe
p pHe
p_T [MeV]


yields_pHe+_p_T [MeV]
π⁺ pHe
K⁺ pHe
p⁺ pHe
p_T [MeV]

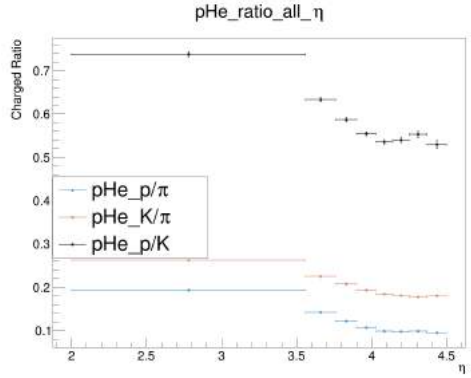

yields_pHe-_p_T [MeV]
π⁻ pHe
K⁻ pHe
p̄ pHe
p_T [MeV]

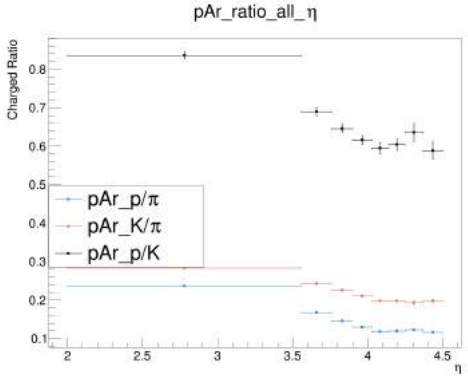

yields_pArall_p_T [MeV]
π pAr
K pAr
p pAr
p_T [MeV]

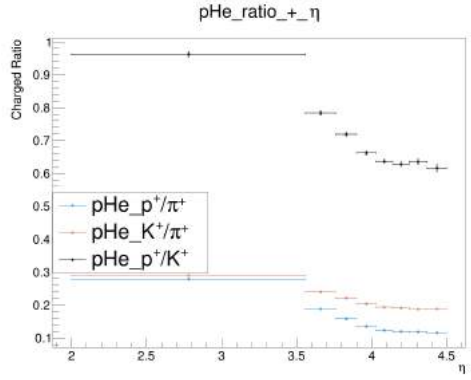

yields_pAr+_p_T [MeV]
π⁺ pAr
K⁺ pAr
p⁺ pAr
p_T [MeV]

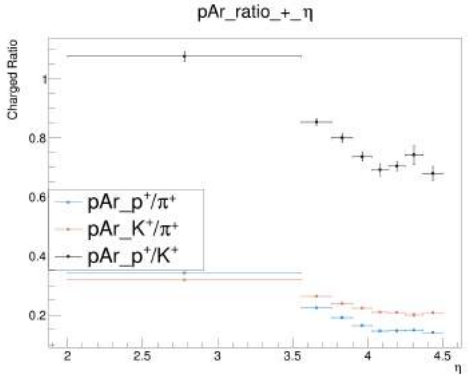

yields_pAr-_p_T [MeV]
π⁻ pAr
K⁻ pAr
p̄ pAr
p_T [MeV]

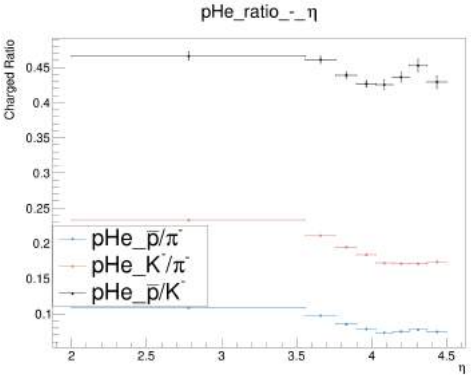
.2 1D single ratios in η and p_T at fixed gas type

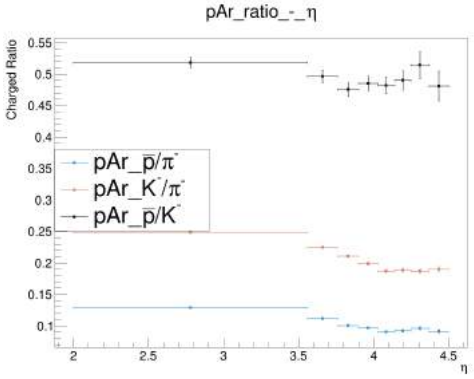


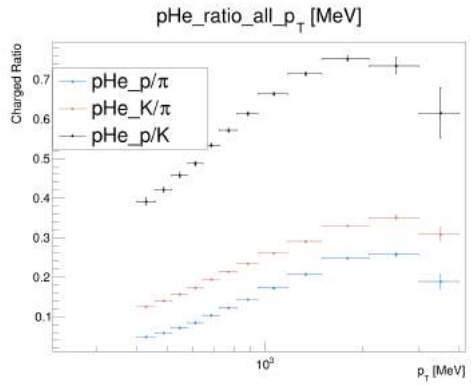


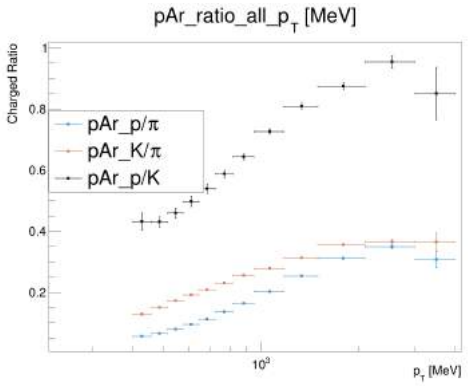


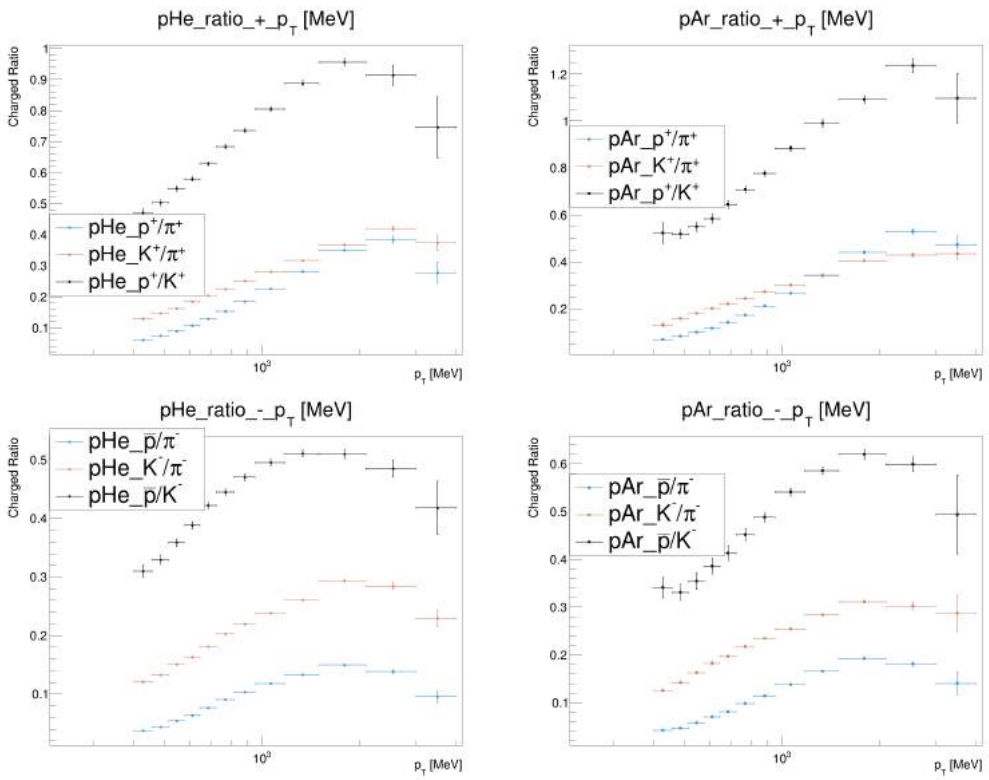




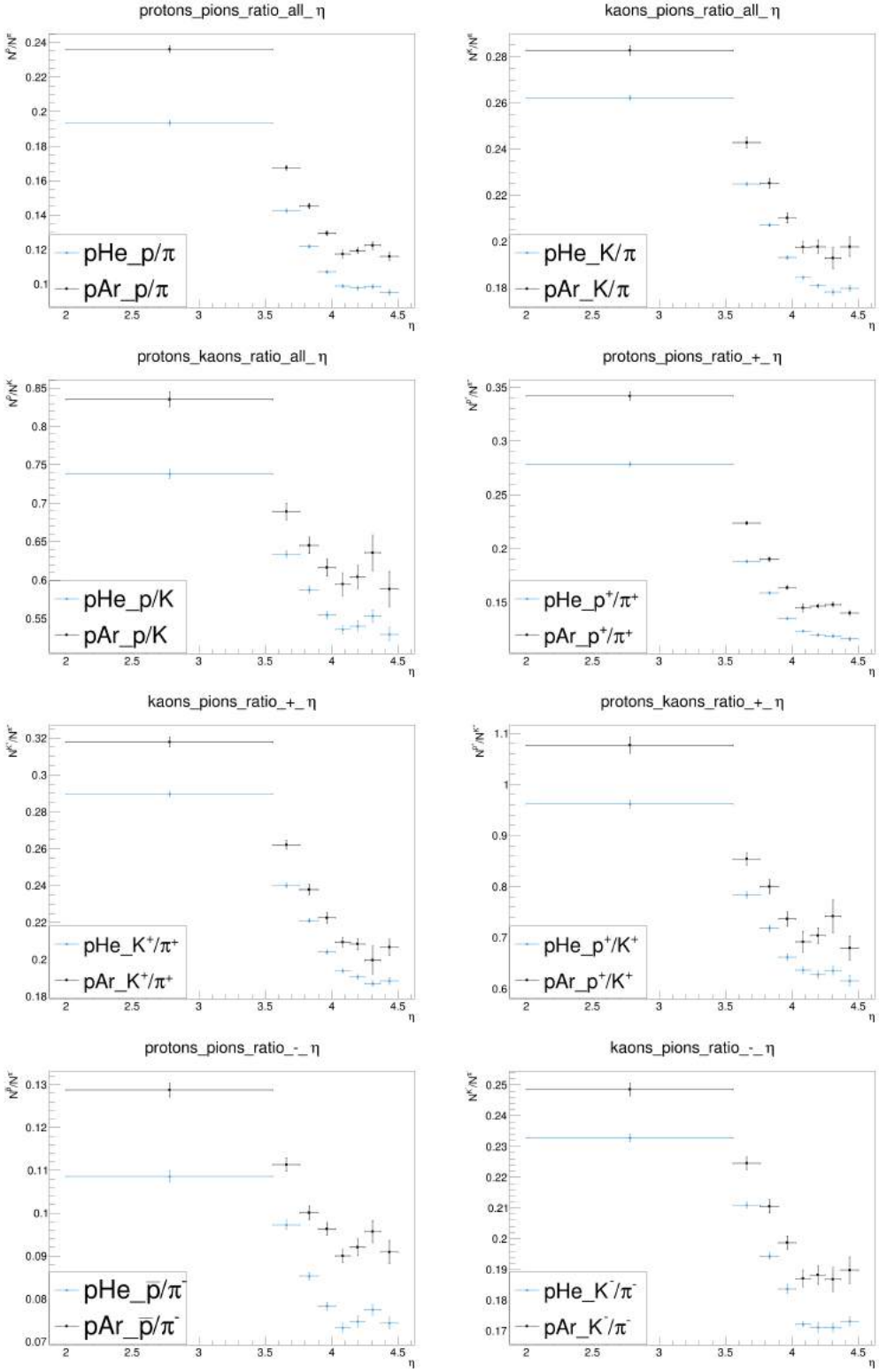


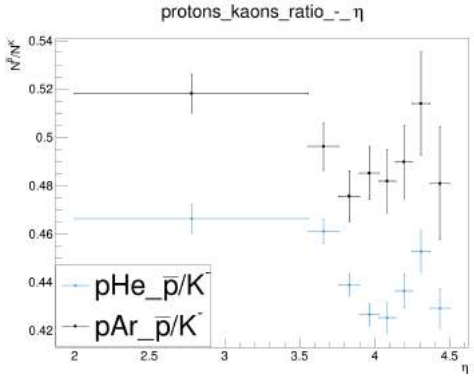


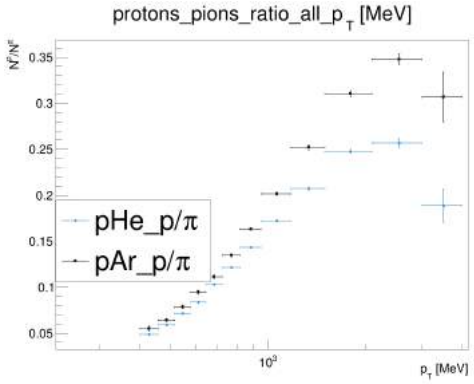


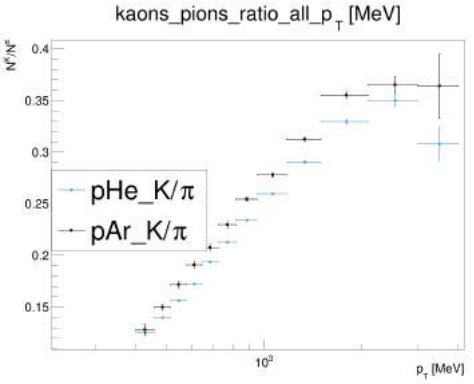

pHe_ratio_+_p_T [MeV]
Charged Ratio
pHe_p+/π+
pHe_K+/π+
pHe_p+/K+
p_T [MeV]
pAr_ratio_+_p_T [MeV]
Charged Ratio
pAr_p+/π+
pAr_K+/π+
pAr_p+/K+
p_T [MeV]
pHe_ratio_-_p_T [MeV]
Charged Ratio
pHe_p̄/π-
pHe_K-/π-
pHe_p̄/K-
p_T [MeV]
pAr_ratio_-_p_T [MeV]
Charged Ratio
pAr_p̄/π-
pAr_K-/π-
pAr_p̄/K-
p_T [MeV]

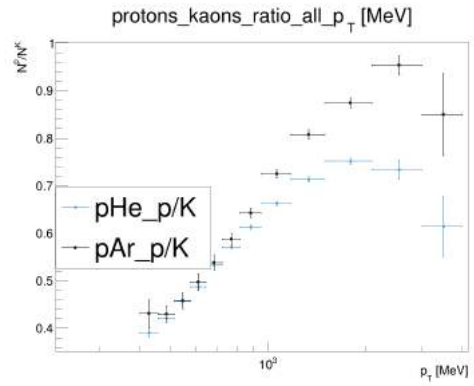
.3 1D single ratios in η and p_T for pHe and pAr comparison

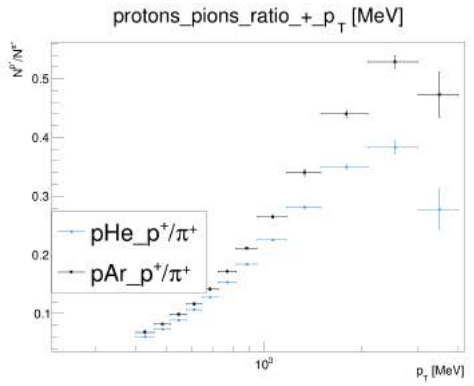


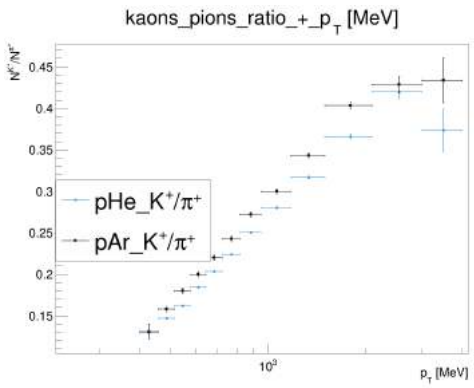

protons_kaons_ratio_-_η
pHe_p̄/K⁻
pAr_p̄/K⁻
η

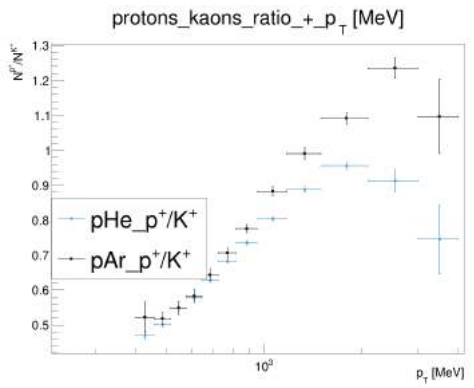

protons_pions_ratio_all_p_T [MeV]
pHe_p/π
pAr_p/π
p_T [MeV]

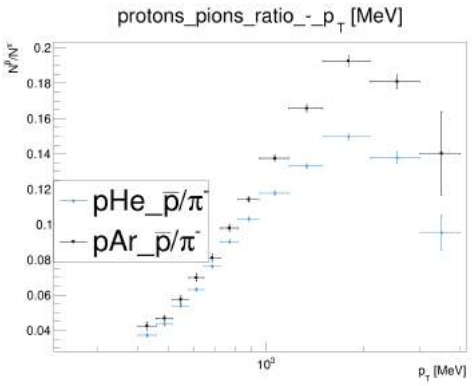

kaons_pions_ratio_all_p_T [MeV]
pHe_K/π
pAr_K/π
p_T [MeV]

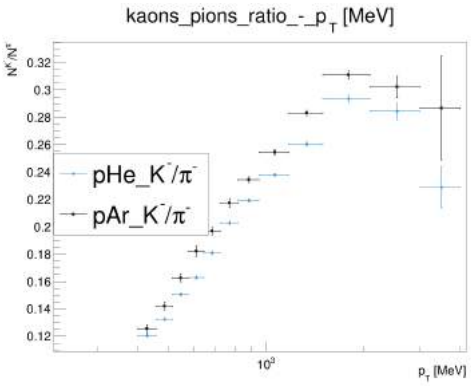

protons_kaons_ratio_all_p_T [MeV]
pHe_p/K
pAr_p/K
p_T [MeV]

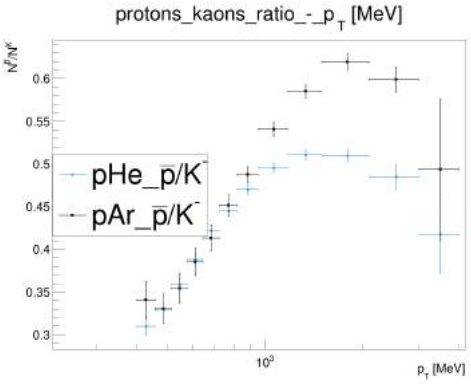

protons_pions_ratio_+_p_T [MeV]
pHe_p⁺/π⁺
pAr_p⁺/π⁺
p_T [MeV]


kaons_pions_ratio_+_p_T [MeV]
pHe_K⁺/π⁺
pAr_K⁺/π⁺
p_T [MeV]

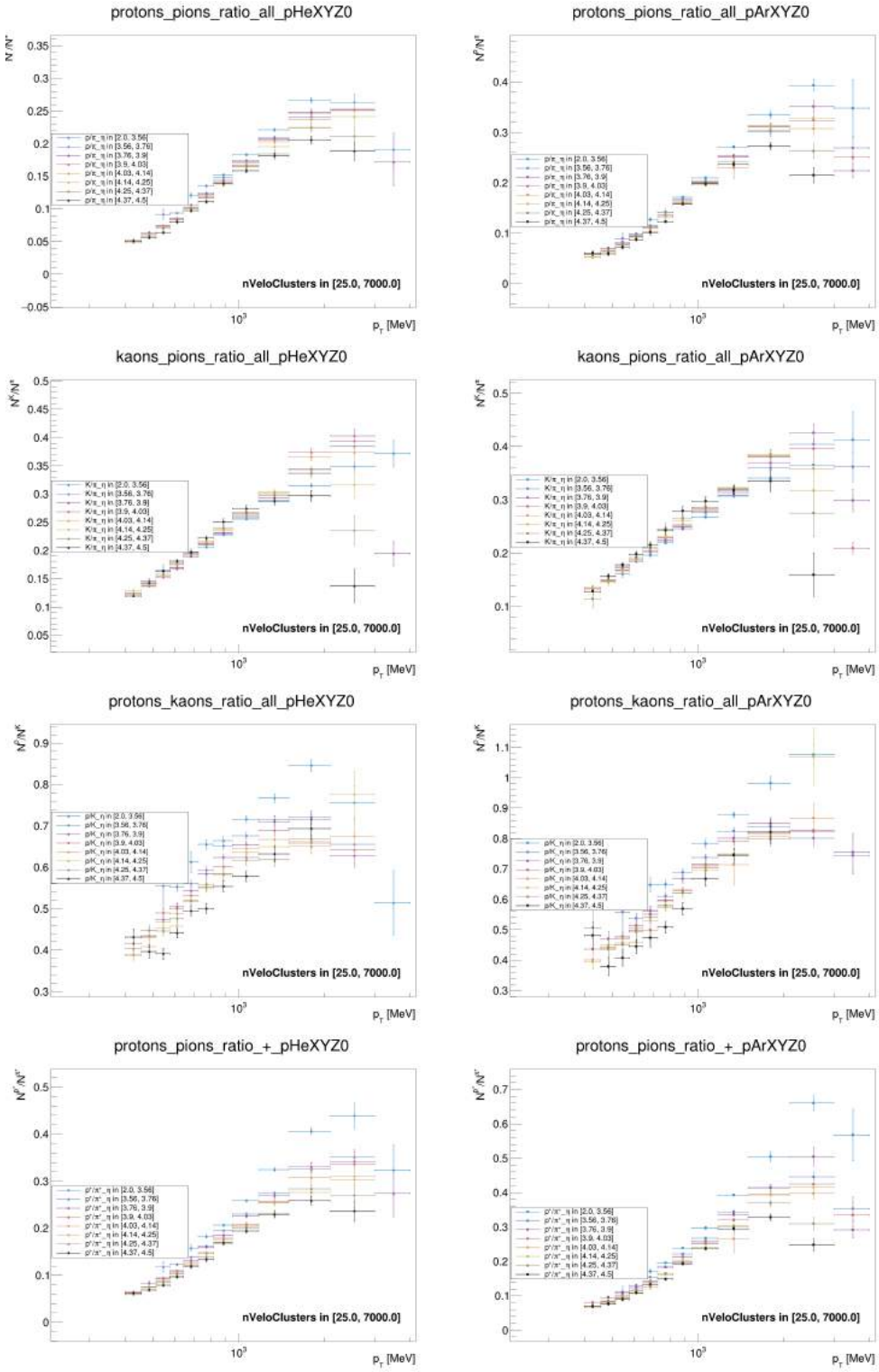

protons_kaons_ratio_+_p_T [MeV]
pHe_p⁺/K⁺
pAr_p⁺/K⁺
p_T [MeV]

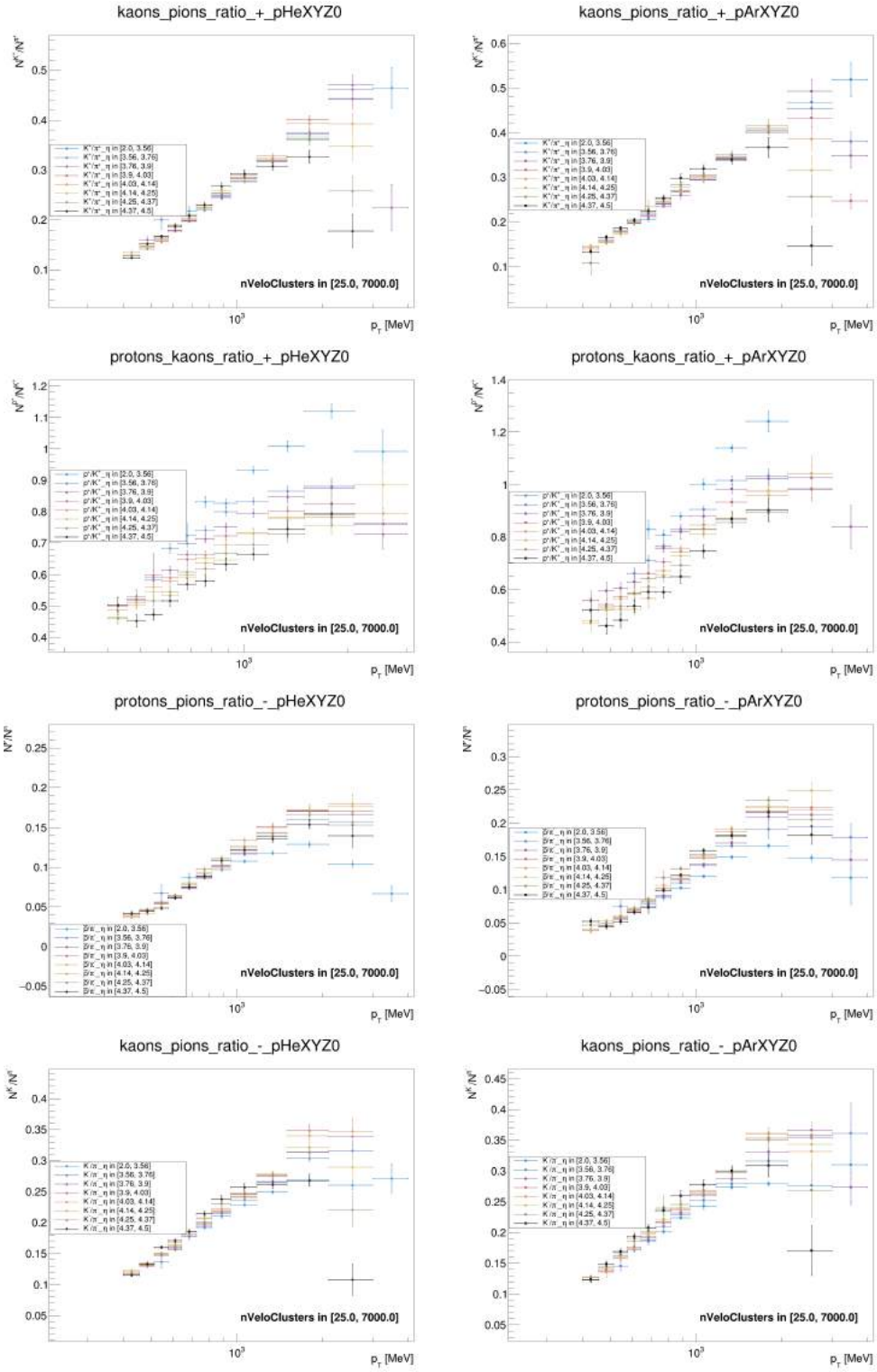

protons_pions_ratio_-_p_T [MeV]
pHe_p̄/π⁻
pAr_p̄/π⁻
p_T [MeV]

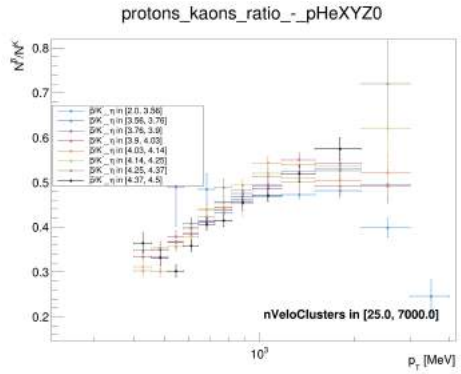

kaons_pions_ratio_-_p_T [MeV]
pHe_K⁻/π⁻
pAr_K⁻/π⁻
p_T [MeV]

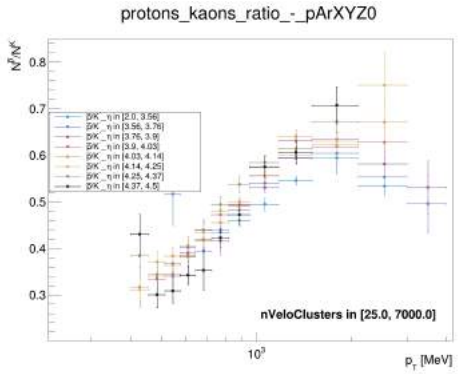

protons_kaons_ratio_-_p_T [MeV]
pHe_p̄/K⁻
pAr_p̄/K⁻
p_T [MeV]

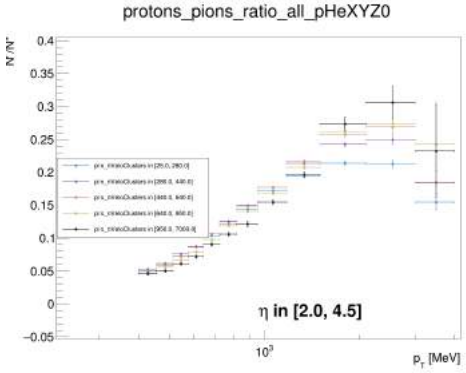
.4 2D single ratios in η-p_T and p_T-nVeloClusters

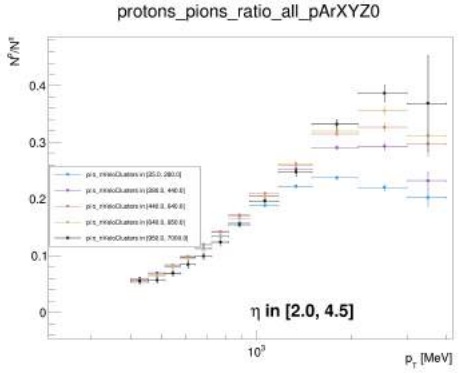


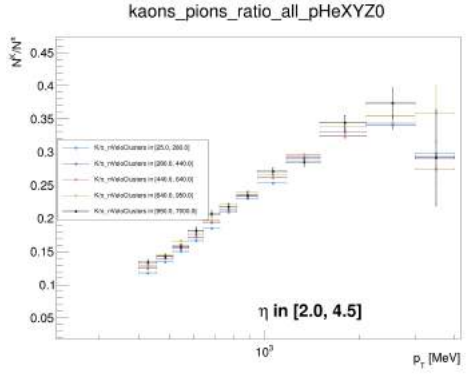

kaons_pions_ratio_+_pHeXYZ0
kaons_pions_ratio_+_pArXYZ0
protons_kaons_ratio_+_pHeXYZ0
protons_kaons_ratio_+_pArXYZ0
protons_pions_ratio_-_pHeXYZ0
protons_pions_ratio_-_pArXYZ0
kaons_pions_ratio_-_pHeXYZ0
kaons_pions_ratio_-_pArXYZ0
nVeloClusters in [25.0, 7000.0]
p_T [MeV]

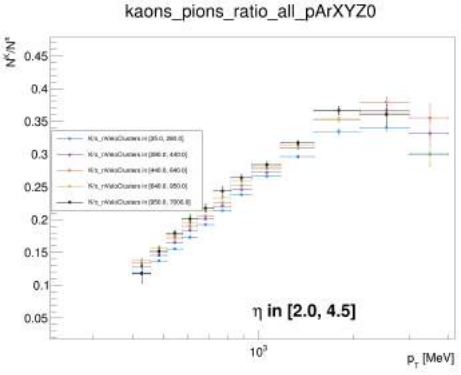

protons_kaons_ratio_-_pHeXYZ0
nVeloClusters in [25.0, 7000.0]
p_T [MeV]

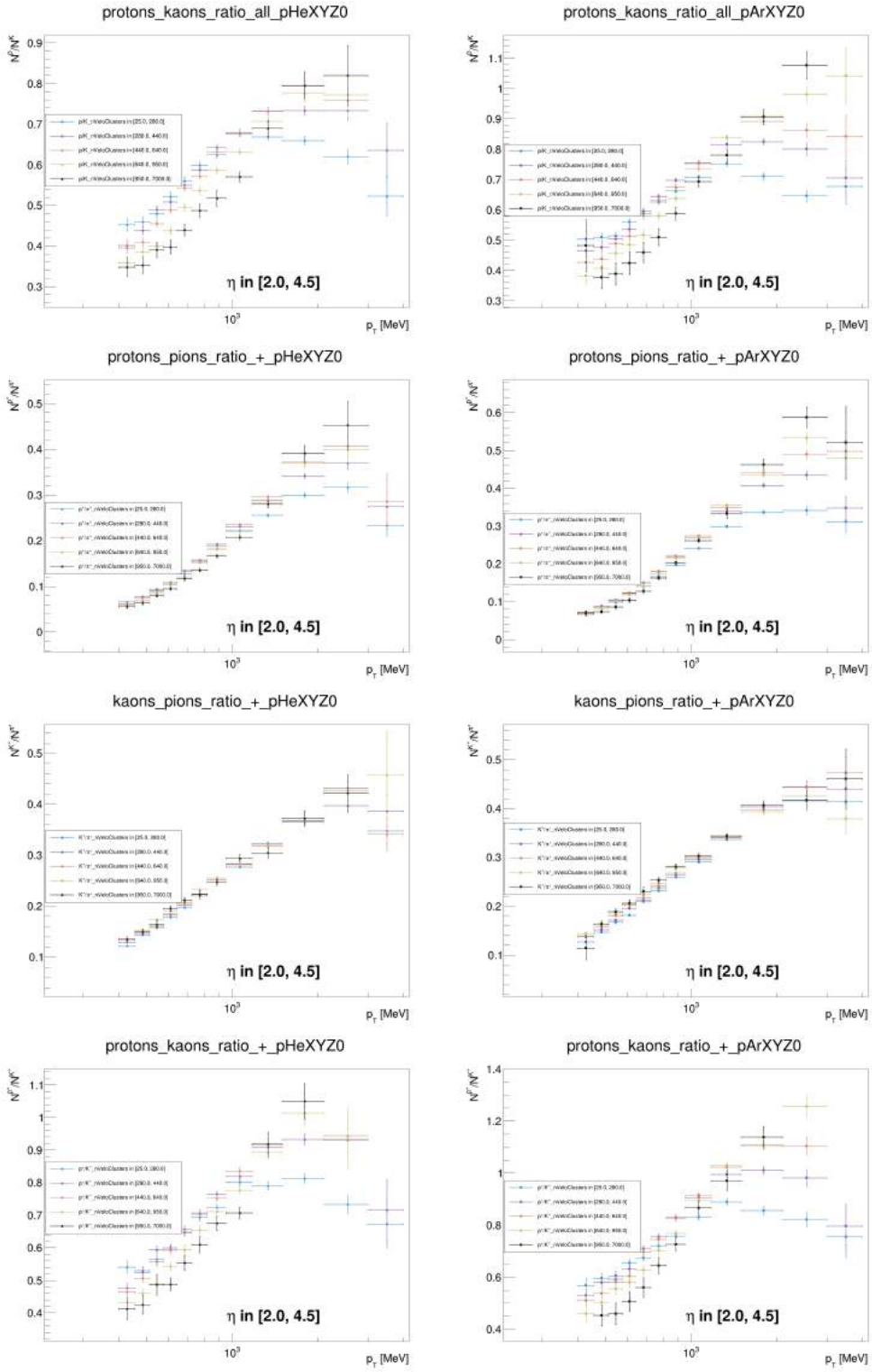

protons_kaons_ratio_-_pArXYZ0
nVeloClusters in [25.0, 7000.0]
p_T [MeV]

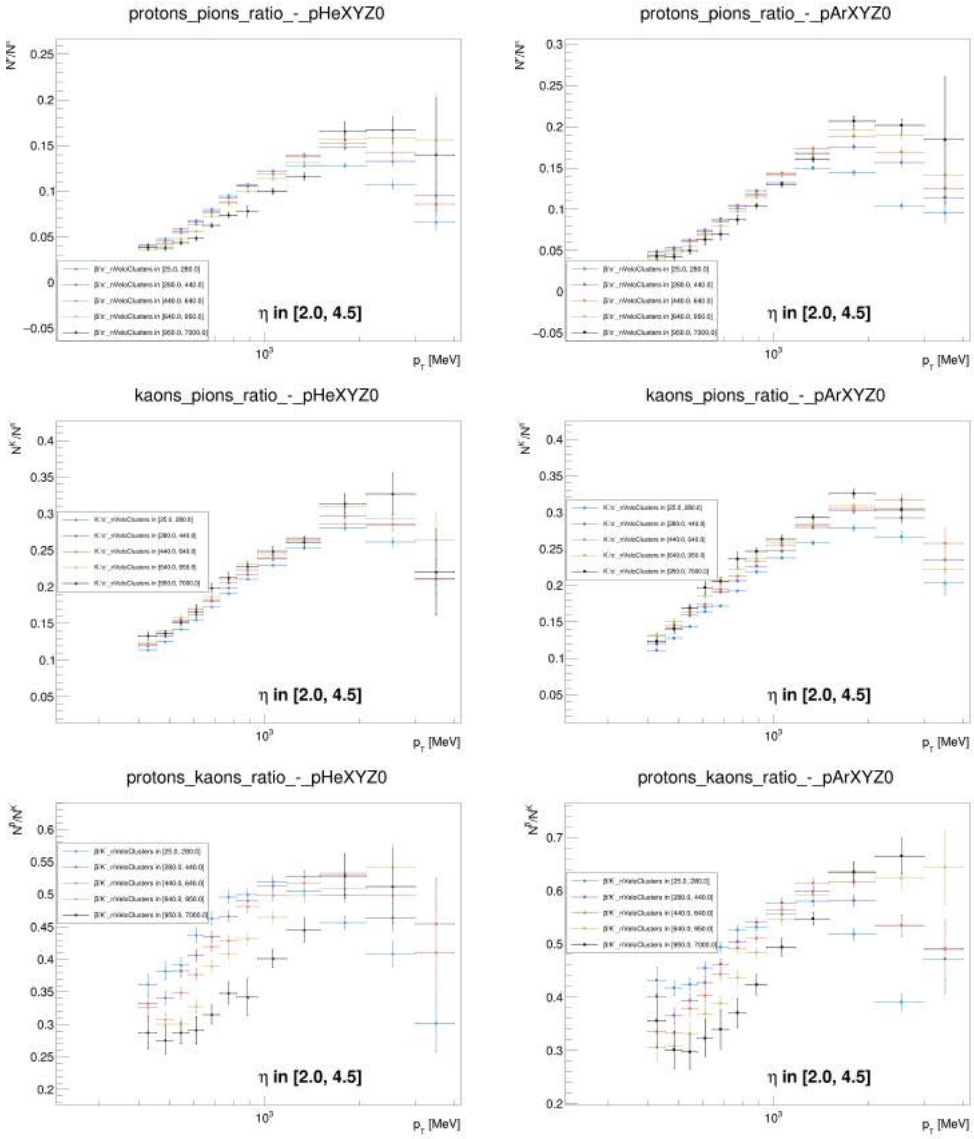

protons_pions_ratio_all_pHeXYZ0
η in [2.0, 4.5]
p_T [MeV]


protons_pions_ratio_all_pArXYZ0
η in [2.0, 4.5]
p_T [MeV]

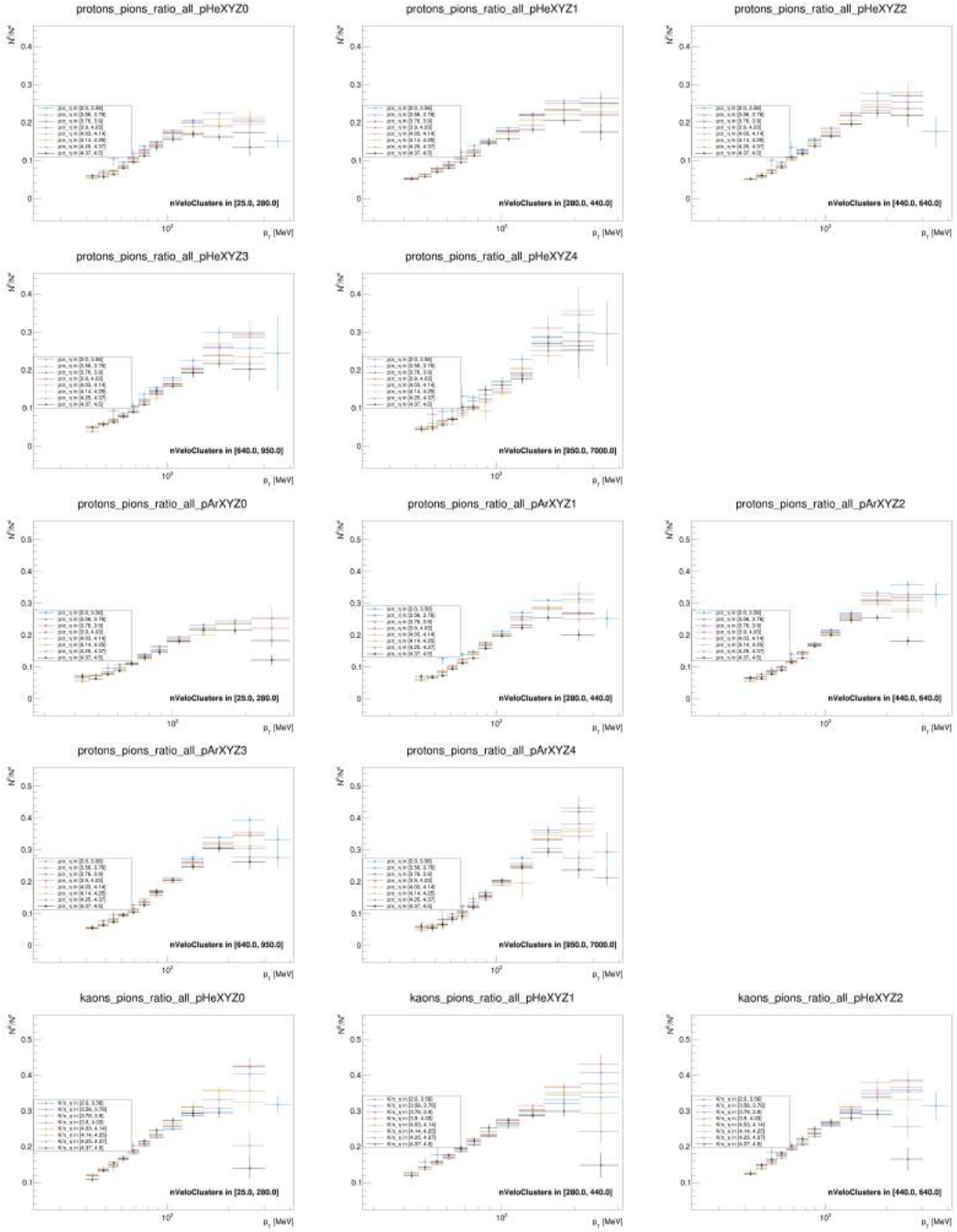

kaons_pions_ratio_all_pHeXYZ0
η in [2.0, 4.5]
p_T [MeV]

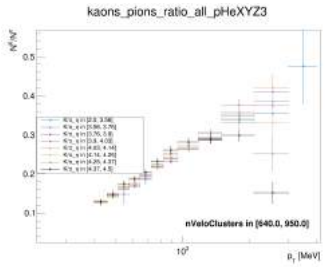

kaons_pions_ratio_all_pArXYZ0
η in [2.0, 4.5]
p_T [MeV]

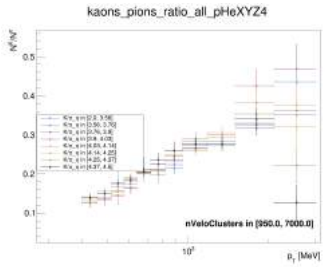

protons_kaons_ratio_all_pHeXYZ0
protons_kaons_ratio_all_pArXYZ0
protons_pions_ratio_+_pHeXYZ0
protons_pions_ratio_+_pArXYZ0
kaons_pions_ratio_+_pHeXYZ0
kaons_pions_ratio_+_pArXYZ0
protons_kaons_ratio_+_pHeXYZ0
protons_kaons_ratio_+_pArXYZ0
η in [2.0, 4.5]
p_T [MeV]

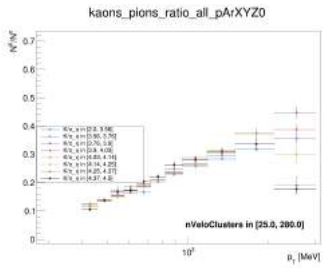

protons_pions_ratio_-_pHeXYZ0
protons_pions_ratio_-_pArXYZ0
kaons_pions_ratio_-_pHeXYZ0
kaons_pions_ratio_-_pArXYZ0
protons_kaons_ratio_-_pHeXYZ0
protons_kaons_ratio_-_pArXYZ0
η in [2.0, 4.5]
p_T [MeV]

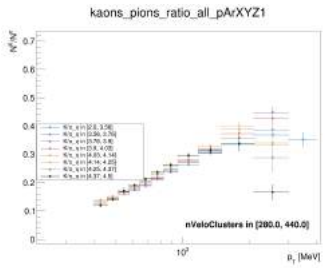
.5 3D single ratios in η-p_T-nVeloClusters

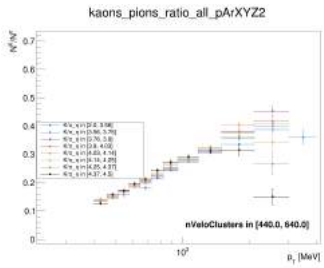


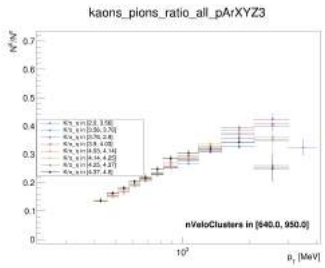

kaons_pions_ratio_all_pHeXYZ3
nVeloClusters in [640.0, 950.0]

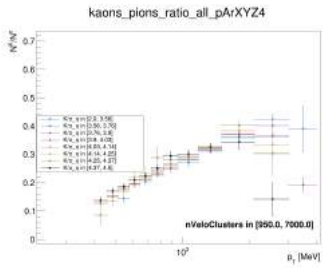

kaons_pions_ratio_all_pHeXYZ4
nVeloClusters in [950.0, 7000.0]

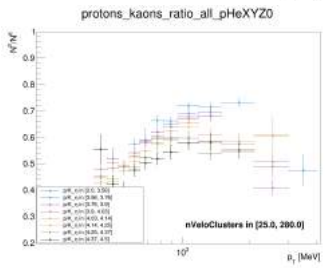

kaons_pions_ratio_all_pArXYZ0
nVeloClusters in [25.0, 280.0]

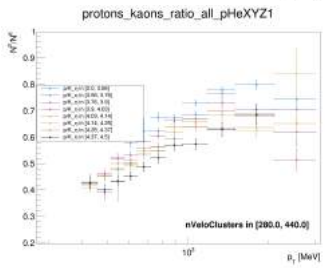

kaons_pions_ratio_all_pArXYZ1
nVeloClusters in [280.0, 440.0]

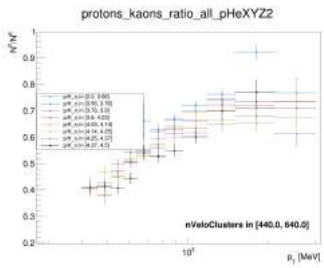

kaons_pions_ratio_all_pArXYZ2
nVeloClusters in [440.0, 640.0]

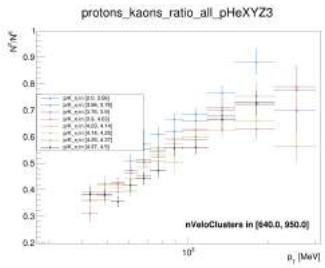

kaons_pions_ratio_all_pArXYZ3
nVeloClusters in [640.0, 950.0]

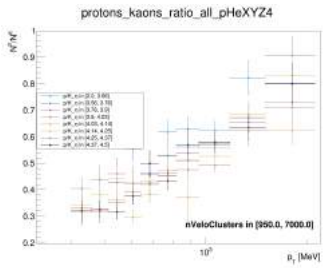

kaons_pions_ratio_all_pArXYZ4
nVeloClusters in [950.0, 7000.0]

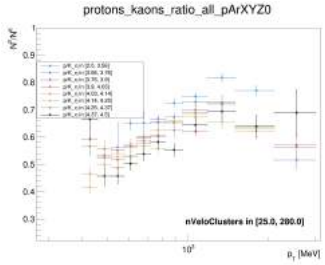

protons_kaons_ratio_all_pHeXYZ0
nVeloClusters in [25.0, 280.0]

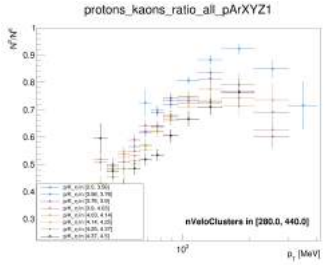

protons_kaons_ratio_all_pHeXYZ1
nVeloClusters in [280.0, 440.0]

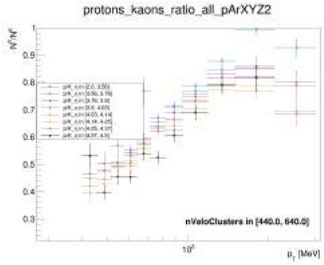

protons_kaons_ratio_all_pHeXYZ2
nVeloClusters in [440.0, 640.0]

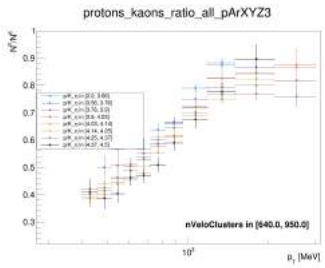

protons_kaons_ratio_all_pHeXYZ3
nVeloClusters in [640.0, 950.0]

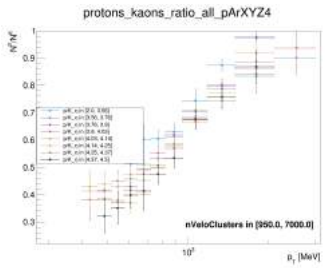

protons_kaons_ratio_all_pHeXYZ4
nVeloClusters in [950.0, 7000.0]

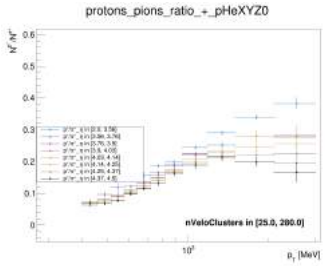

protons_kaons_ratio_all_pArXYZ0
nVeloClusters in [25.0, 280.0]

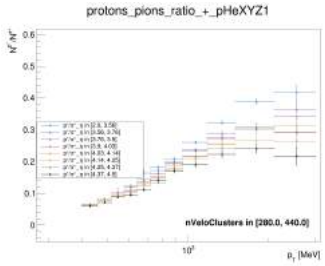

protons_kaons_ratio_all_pArXYZ1
nVeloClusters in [280.0, 440.0]

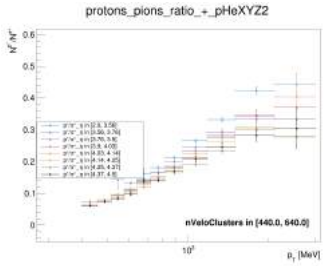

protons_kaons_ratio_all_pArXYZ2
nVeloClusters in [440.0, 640.0]

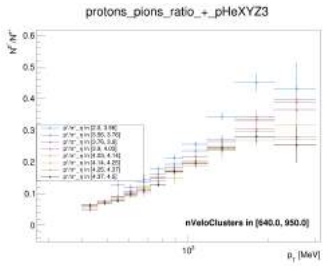

protons_kaons_ratio_all_pArXYZ3
nVeloClusters in [640.0, 950.0]

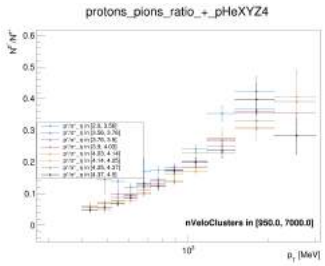

protons_kaons_ratio_all_pArXYZ4
nVeloClusters in [950.0, 7000.0]

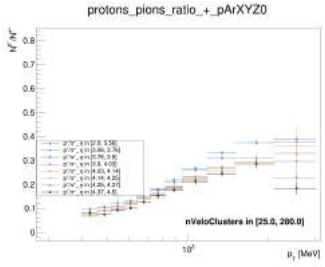

protons_pions_ratio_+_pHeXYZ0
nVeloClusters in [25.0, 280.0]

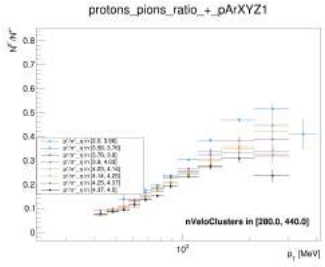

protons_pions_ratio_+_pHeXYZ1
nVeloClusters in [280.0, 440.0]

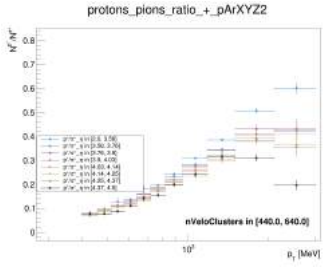

protons_pions_ratio_+_pHeXYZ2
nVeloClusters in [440.0, 640.0]

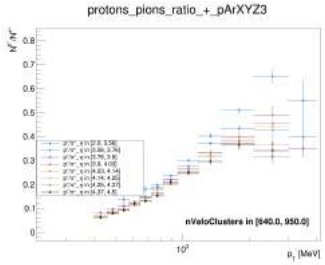

protons_pions_ratio_+_pHeXYZ3
nVeloClusters in [640.0, 950.0]

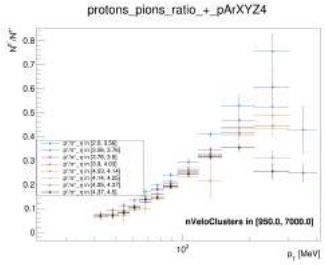

protons_pions_ratio_+_pHeXYZ4
nVeloClusters in [950.0, 7000.0]

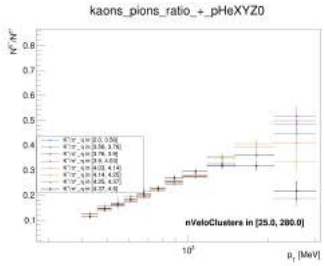

protons_pions_ratio_+_pArXYZ0
nVeloClusters in [25.0, 280.0]

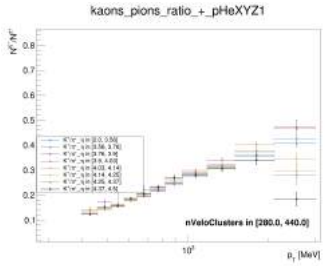

protons_pions_ratio_+_pArXYZ1
nVeloClusters in [280.0, 440.0]

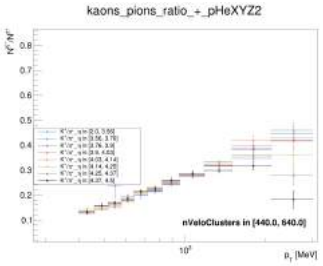

protons_pions_ratio_+_pArXYZ2
nVeloClusters in [440.0, 640.0]

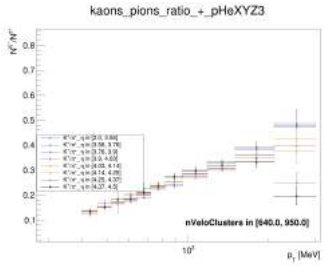

protons_pions_ratio_+_pArXYZ3
nVeloClusters in [640.0, 950.0]

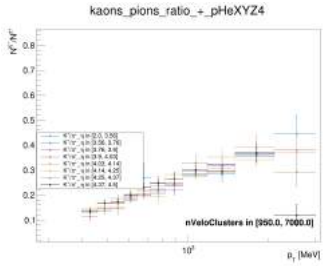

protons_pions_ratio_+_pArXYZ4
nVeloClusters in [950.0, 7000.0]

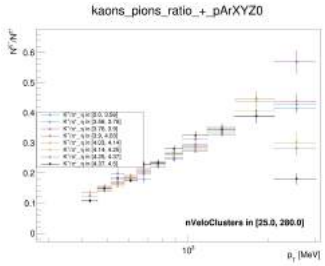

kaons_pions_ratio_+_pHeXYZ0
nVeloClusters in [25.0, 280.0]

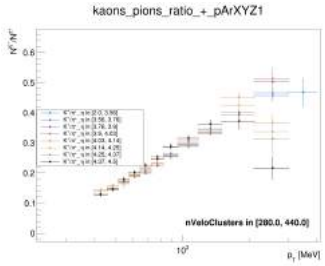

kaons_pions_ratio_+_pHeXYZ1
nVeloClusters in [280.0, 440.0]

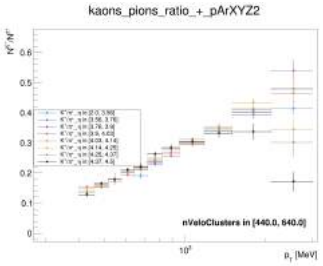

kaons_pions_ratio_+_pHeXYZ2
nVeloClusters in [440.0, 640.0]

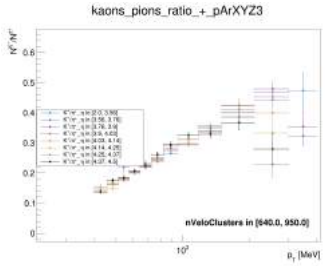

kaons_pions_ratio_+_pHeXYZ3
nVeloClusters in [640.0, 950.0]

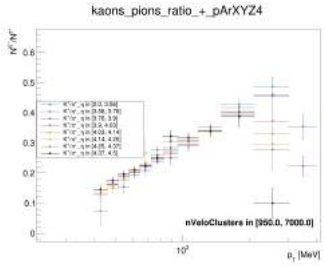

kaons_pions_ratio_+_pHeXYZ4
nVeloClusters in [950.0, 7000.0]

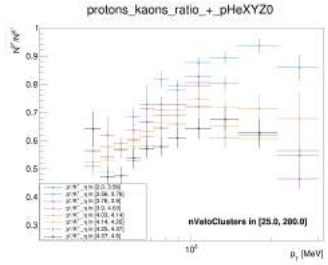

kaons_pions_ratio_+_pArXYZ0
nVeloClusters in [25.0, 280.0]

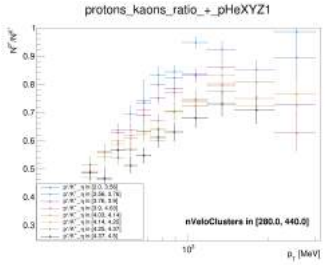

kaons_pions_ratio_+_pArXYZ1
nVeloClusters in [280.0, 440.0]

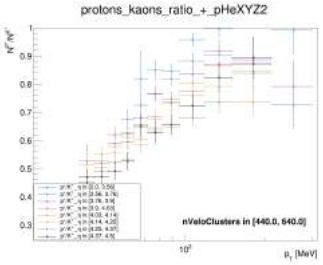

kaons_pions_ratio_+_pArXYZ2
nVeloClusters in [440.0, 640.0]

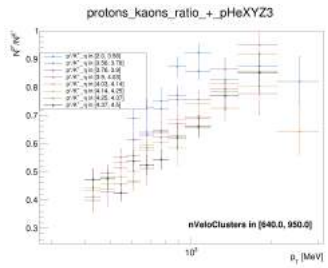

kaons_pions_ratio_+_pArXYZ3
nVeloClusters in [640.0, 950.0]

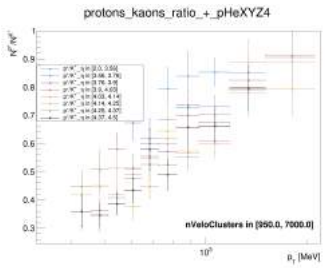

kaons_pions_ratio_+_pArXYZ4
nVeloClusters in [950.0, 7000.0]

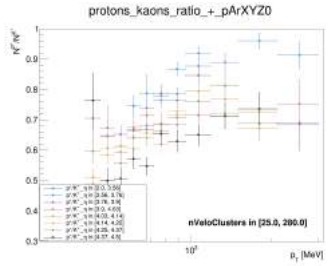

protons_kaons_ratio_+_pHeXYZ0
nVeloClusters in [25.0, 280.0]

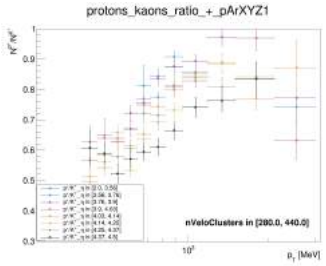

protons_kaons_ratio_+_pHeXYZ1
nVeloClusters in [280.0, 440.0]

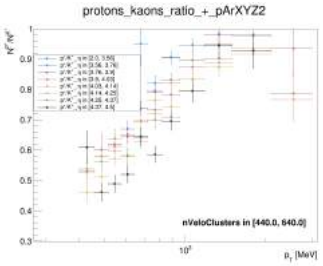

protons_kaons_ratio_+_pHeXYZ2
nVeloClusters in [440.0, 640.0]

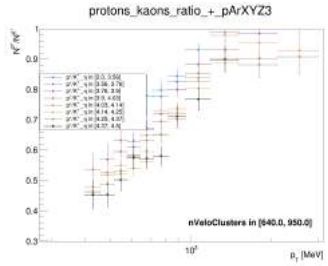

protons_kaons_ratio_+_pHeXYZ3
nVeloClusters in [640.0, 950.0]

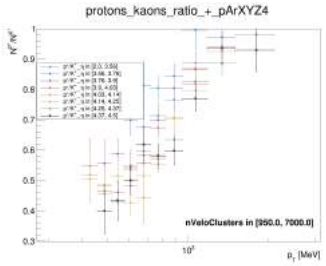

protons_kaons_ratio_+_pHeXYZ4
nVeloClusters in [950.0, 7000.0]

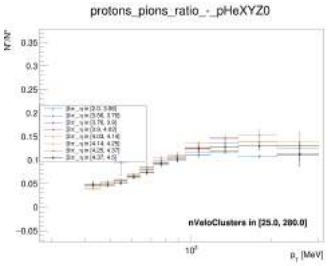

protons_kaons_ratio_+_pArXYZ0
nVeloClusters in [25.0, 280.0]

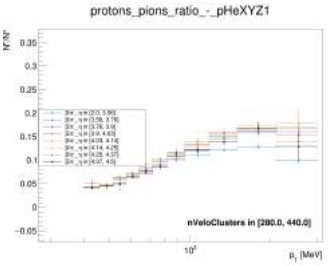

protons_kaons_ratio_+_pArXYZ1
nVeloClusters in [280.0, 440.0]

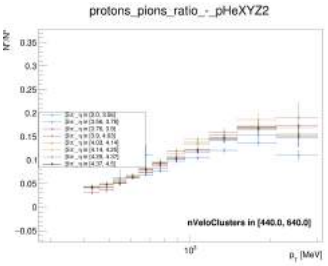

protons_kaons_ratio_+_pArXYZ2
nVeloClusters in [440.0, 640.0]

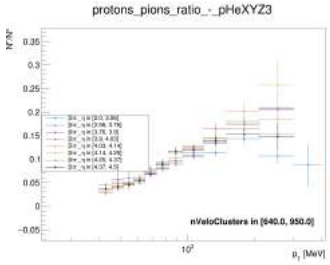

protons_kaons_ratio_+_pArXYZ3
nVeloClusters in [640.0, 950.0]

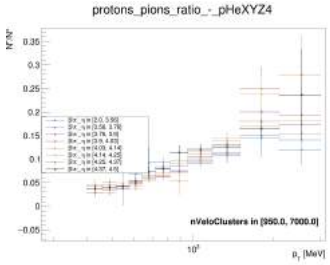

protons_kaons_ratio_+_pArXYZ4
nVeloClusters in [950.0, 7000.0]

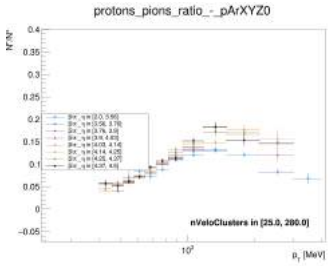

protons_pions_ratio_-_pHeXYZ0
nVeloClusters in [25.0, 280.0]

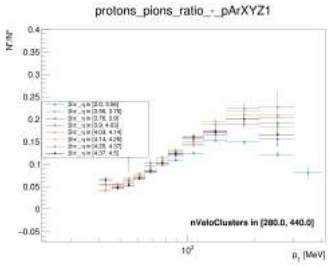

protons_pions_ratio_-_pHeXYZ1
nVeloClusters in [280.0, 440.0]

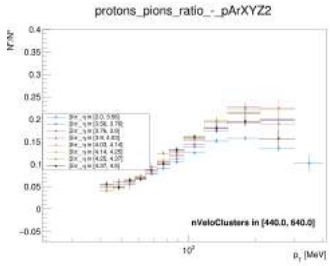

protons_pions_ratio_-_pHeXYZ2
nVeloClusters in [440.0, 640.0]

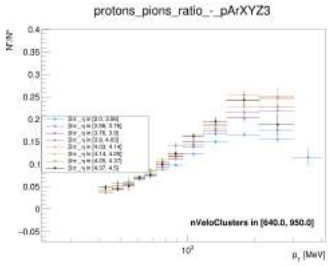

protons_pions_ratio_-_pHeXYZ3
nVeloClusters in [640.0, 950.0]

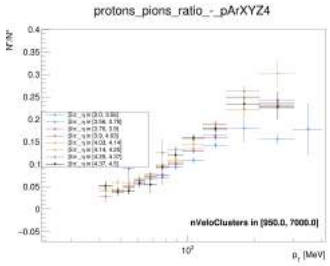

protons_pions_ratio_-_pHeXYZ4
nVeloClusters in [950.0, 7000.0]

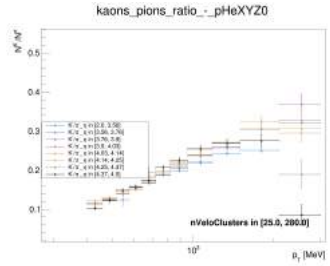

protons_pions_ratio_-_pArXYZ0
nVeloClusters in [25.0, 280.0]

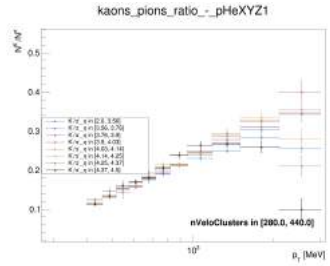

protons_pions_ratio_-_pArXYZ1
nVeloClusters in [280.0, 440.0]

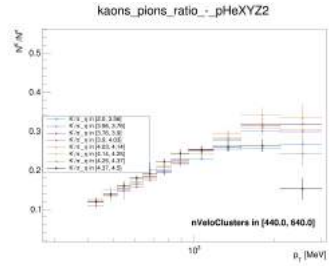

protons_pions_ratio_-_pArXYZ2
nVeloClusters in [440.0, 640.0]

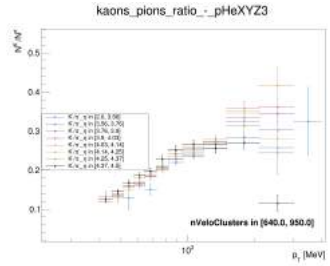

protons_pions_ratio_-_pArXYZ3
nVeloClusters in [640.0, 950.0]

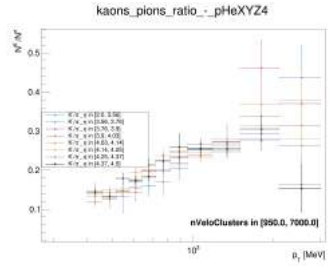

protons_pions_ratio_-_pArXYZ4
nVeloClusters in [950.0, 7000.0]

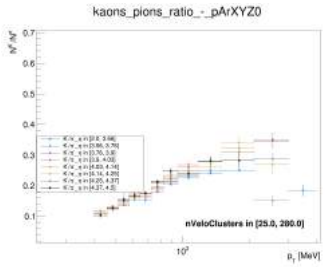

kaons_pions_ratio_-_pHeXYZ0
nVeloClusters in [25.0, 280.0]

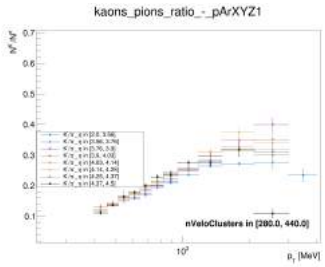

kaons_pions_ratio_-_pHeXYZ1
nVeloClusters in [280.0, 440.0]

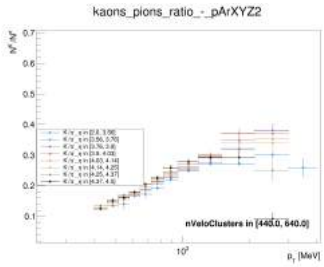

kaons_pions_ratio_-_pHeXYZ2
nVeloClusters in [440.0, 640.0]

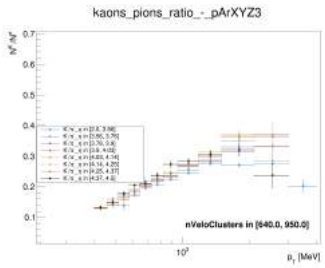

kaons_pions_ratio_-_pHeXYZ3
nVeloClusters in [640.0, 950.0]

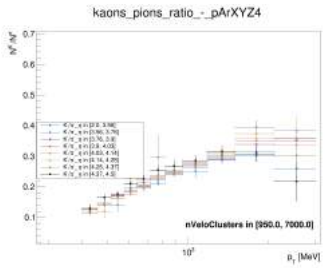

kaons_pions_ratio_-_pHeXYZ4
nVeloClusters in [950.0, 7000.0]

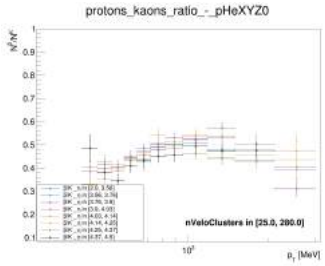

kaons_pions_ratio_-_pArXYZ0
nVeloClusters in [25.0, 280.0]

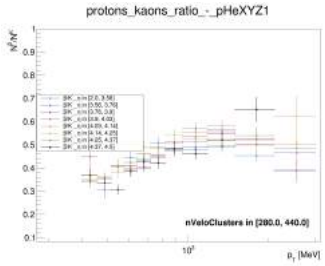

kaons_pions_ratio_-_pArXYZ1
nVeloClusters in [280.0, 440.0]

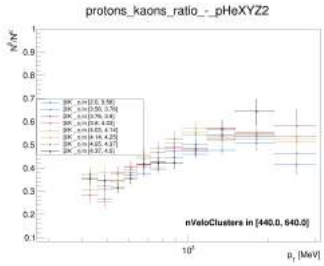

kaons_pions_ratio_-_pArXYZ2
nVeloClusters in [440.0, 640.0]

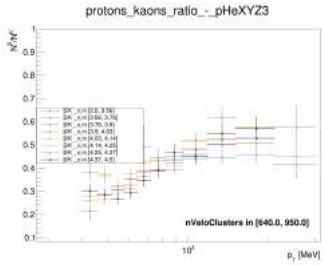

kaons_pions_ratio_-_pArXYZ3
nVeloClusters in [640.0, 950.0]

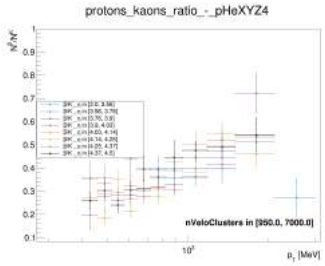

kaons_pions_ratio_-_pArXYZ4
nVeloClusters in [950.0, 7000.0]

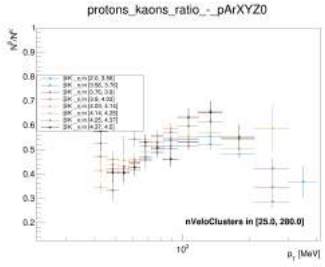

protons_kaons_ratio_-_pHeXYZ0
nVeloClusters in [25.0, 280.0]

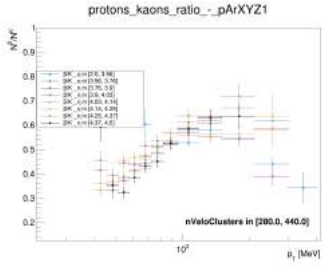

protons_kaons_ratio_-_pHeXYZ1
nVeloClusters in [280.0, 440.0]

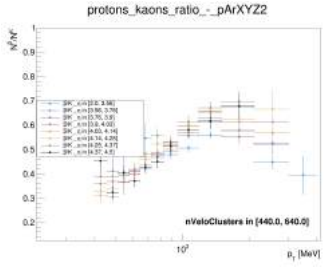

protons_kaons_ratio_-_pHeXYZ2
nVeloClusters in [440.0, 640.0]

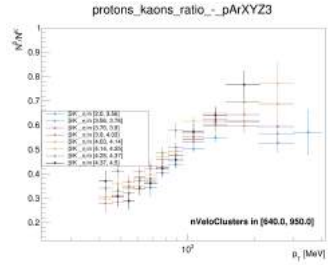

protons_kaons_ratio_-_pHeXYZ3
nVeloClusters in [640.0, 950.0]

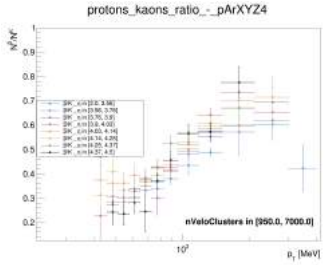

protons_kaons_ratio_-_pHeXYZ4
nVeloClusters in [950.0, 7000.0]

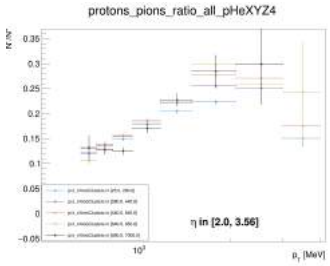

protons_kaons_ratio_-_pArXYZ0
nVeloClusters in [25.0, 280.0]

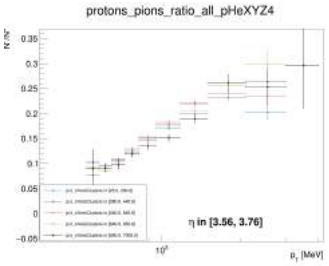

protons_kaons_ratio_-_pArXYZ1
nVeloClusters in [280.0, 440.0]

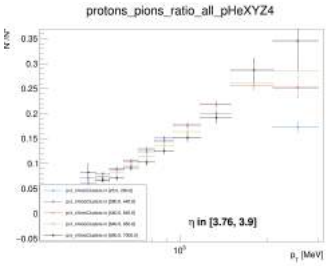

protons_kaons_ratio_-_pArXYZ2
nVeloClusters in [440.0, 640.0]

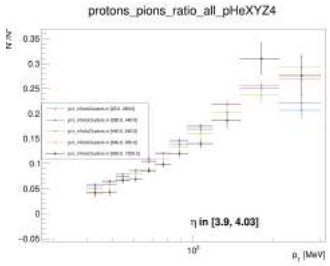

protons_kaons_ratio_+_pArXYZ3
nVeloClusters in [640.0, 950.0]

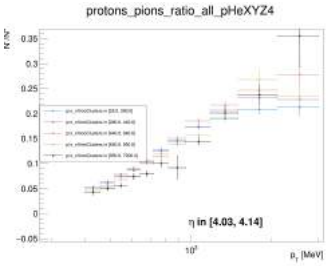

protons_kaons_ratio_+_pArXYZ4
nVeloClusters in [950.0, 7000.0]

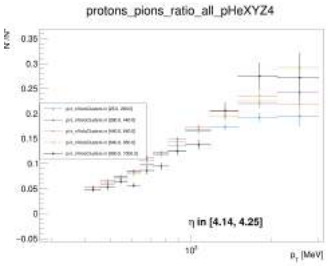

protons_pions_ratio_all_pHeXYZ4
η in [2.0, 3.56]

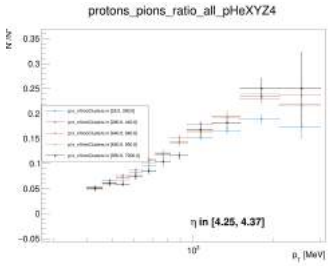

protons_pions_ratio_all_pHeXYZ4
η in [3.56, 3.76]

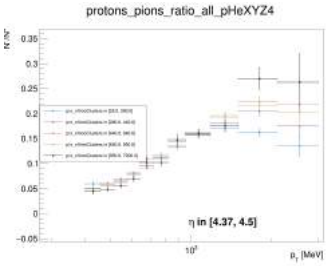

protons_pions_ratio_all_pHeXYZ4
η in [3.76, 3.9]

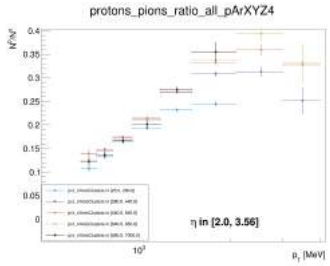

protons_pions_ratio_all_pHeXYZ4
η in [3.9, 4.03]

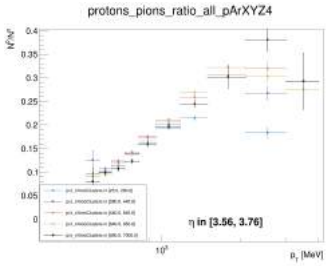

protons_pions_ratio_all_pHeXYZ4
η in [4.03, 4.14]

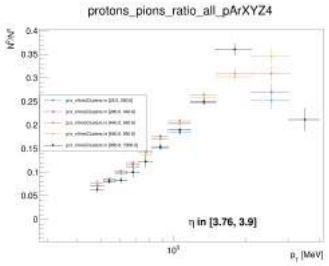

protons_pions_ratio_all_pHeXYZ4
η in [4.14, 4.25]

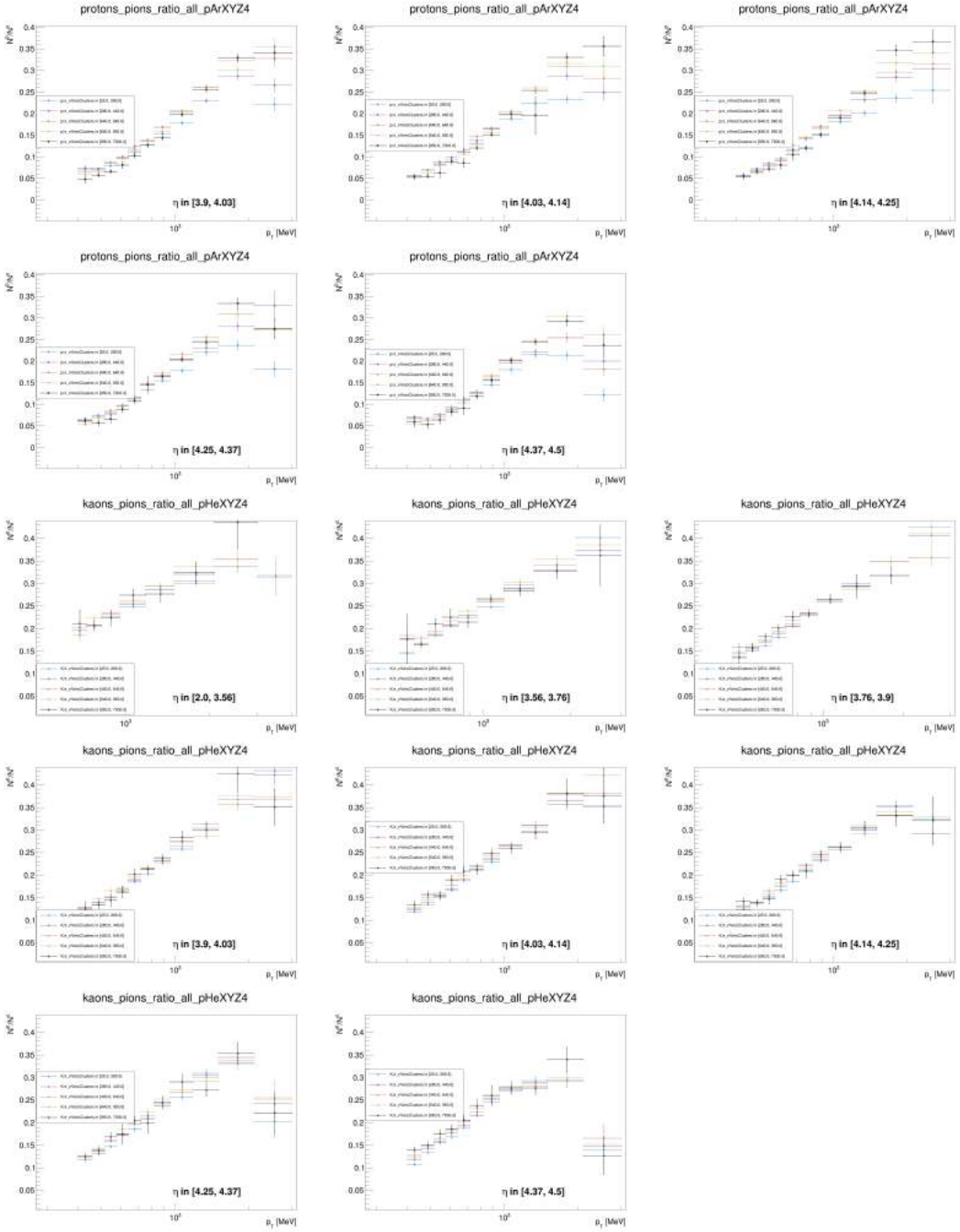

protons_pions_ratio_all_pHeXYZ4
η in [4.25, 4.37]

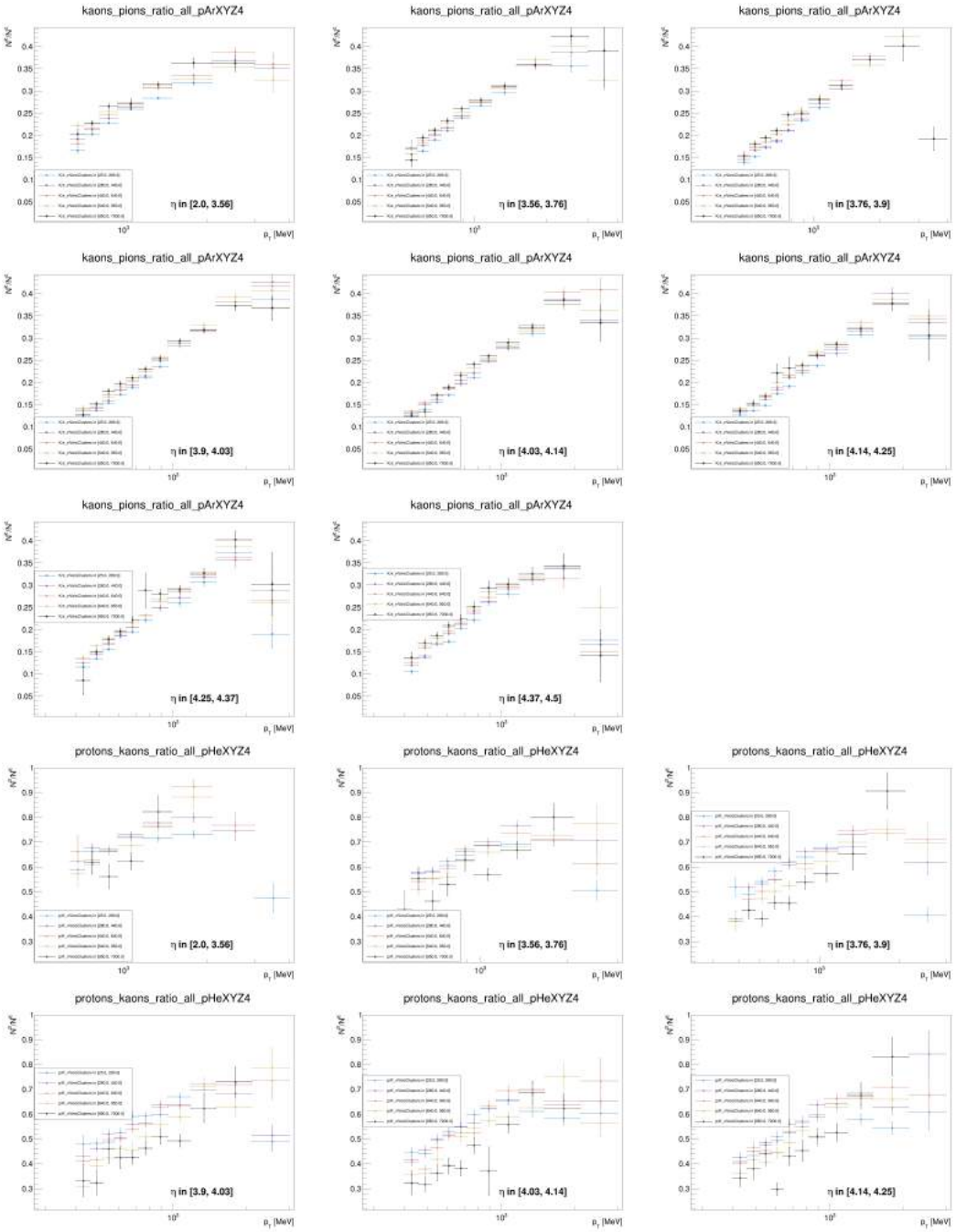

protons_pions_ratio_all_pHeXYZ4
η in [4.37, 4.5]

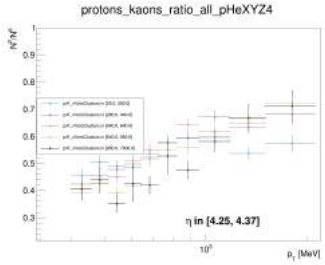

protons_pions_ratio_all_pArXYZ4
η in [2.0, 3.56]

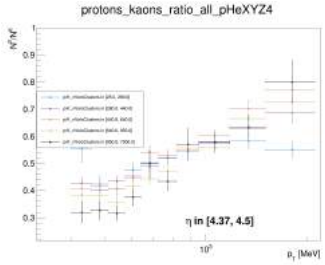

protons_pions_ratio_all_pArXYZ4
η in [3.56, 3.76]

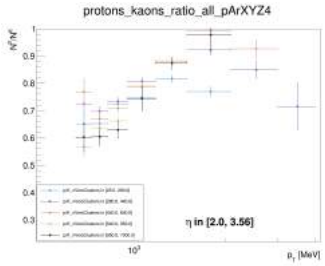

protons_pions_ratio_all_pArXYZ4
η in [3.76, 3.9]

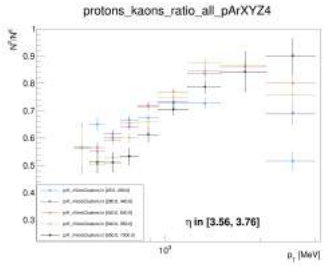

protons_pions_ratio_all_pArXYZ4
η in [3.9, 4.03]
protons_pions_ratio_all_pArXYZ4
η in [4.03, 4.14]
protons_pions_ratio_all_pArXYZ4
η in [4.14, 4.25]
protons_pions_ratio_all_pArXYZ4
η in [4.25, 4.37]
protons_pions_ratio_all_pArXYZ4
η in [4.37, 4.5]
kaons_pions_ratio_all_pHeXYZ4
η in [2.0, 3.56]
kaons_pions_ratio_all_pHeXYZ4
η in [3.56, 3.76]
kaons_pions_ratio_all_pHeXYZ4
η in [3.76, 3.9]
kaons_pions_ratio_all_pHeXYZ4
η in [3.9, 4.03]
kaons_pions_ratio_all_pHeXYZ4
η in [4.03, 4.14]
kaons_pions_ratio_all_pHeXYZ4
η in [4.14, 4.25]
kaons_pions_ratio_all_pHeXYZ4
η in [4.25, 4.37]
kaons_pions_ratio_all_pHeXYZ4
η in [4.37, 4.5]

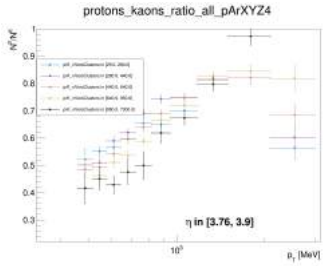

kaons_pions_ratio_all_pArXYZ4
η in [2.0, 3.56]
kaons_pions_ratio_all_pArXYZ4
η in [3.56, 3.76]
kaons_pions_ratio_all_pArXYZ4
η in [3.76, 3.9]
kaons_pions_ratio_all_pArXYZ4
η in [3.9, 4.03]
kaons_pions_ratio_all_pArXYZ4
η in [4.03, 4.14]
kaons_pions_ratio_all_pArXYZ4
η in [4.14, 4.25]
kaons_pions_ratio_all_pArXYZ4
η in [4.25, 4.37]
kaons_pions_ratio_all_pArXYZ4
η in [4.37, 4.5]
protons_kaons_ratio_all_pHeXYZ4
η in [2.0, 3.56]
protons_kaons_ratio_all_pHeXYZ4
η in [3.56, 3.76]
protons_kaons_ratio_all_pHeXYZ4
η in [3.76, 3.9]
protons_kaons_ratio_all_pHeXYZ4
η in [3.9, 4.03]
protons_kaons_ratio_all_pHeXYZ4
η in [4.03, 4.14]
protons_kaons_ratio_all_pHeXYZ4
η in [4.14, 4.25]

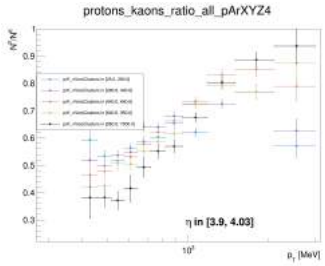

protons_kaons_ratio_all_pHeXYZ4
η in [4.25, 4.37]

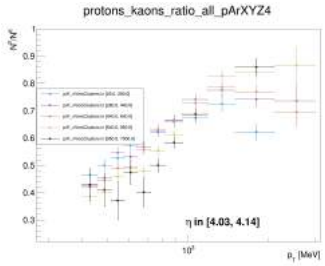

protons_kaons_ratio_all_pHeXYZ4
η in [4.37, 4.5]

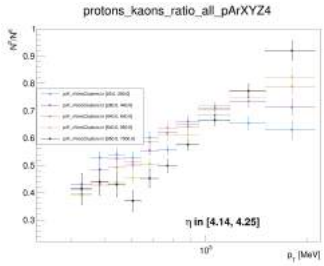

protons_kaons_ratio_all_pArXYZ4
η in [2.0, 3.56]

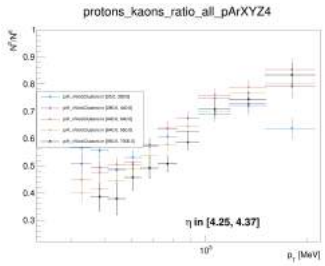

protons_kaons_ratio_all_pArXYZ4
η in [3.56, 3.76]

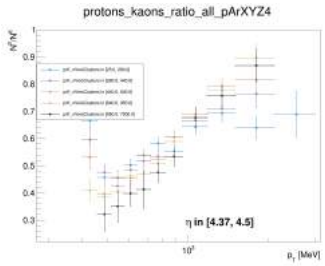

protons_kaons_ratio_all_pArXYZ4
η in [3.76, 3.9]

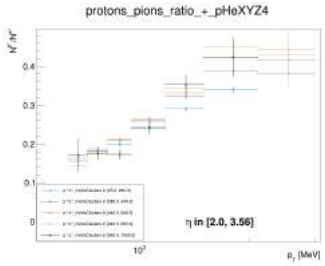

protons_kaons_ratio_all_pArXYZ4
η in [3.9, 4.03]

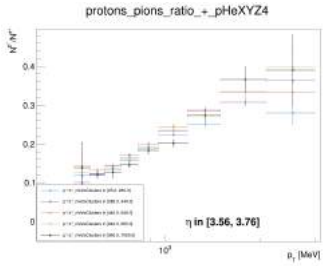

protons_kaons_ratio_all_pArXYZ4
η in [4.03, 4.14]

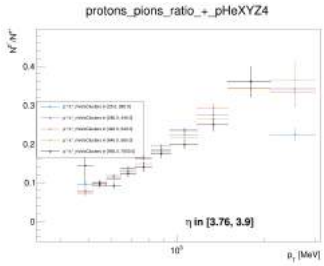

protons_kaons_ratio_all_pArXYZ4
η in [4.14, 4.25]

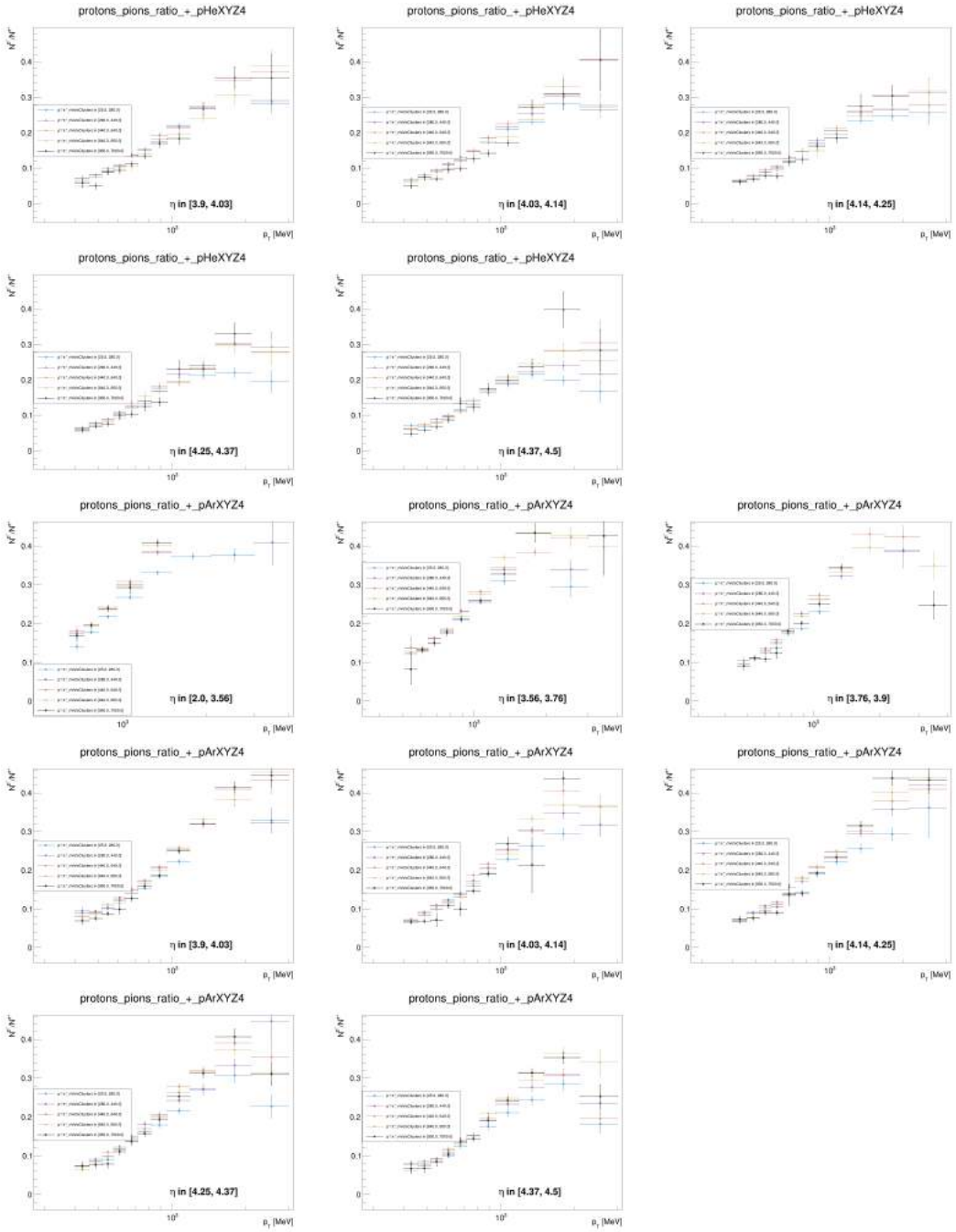

protons_kaons_ratio_all_pArXYZ4
η in [4.25, 4.37]

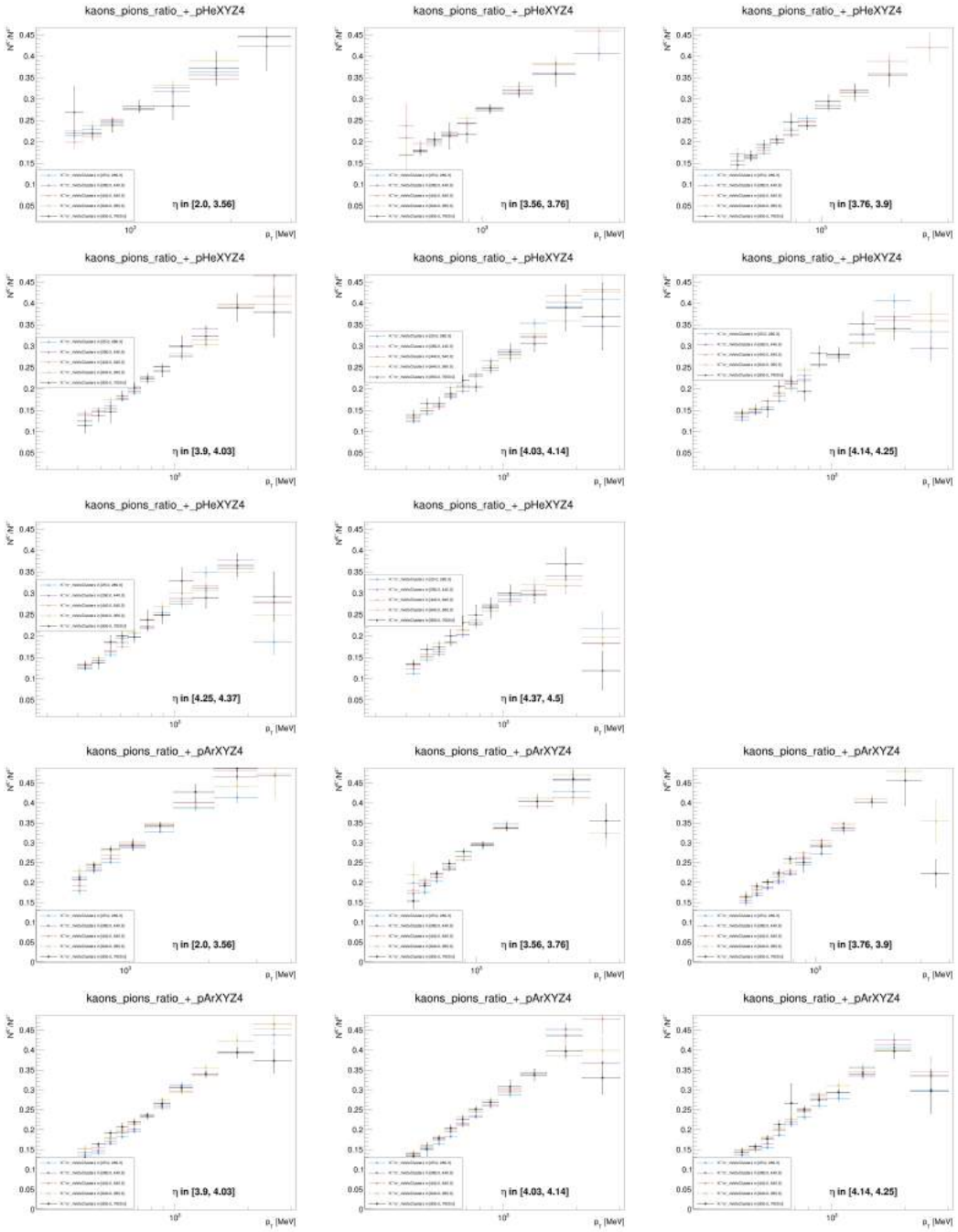

protons_kaons_ratio_all_pArXYZ4
η in [4.37, 4.5]

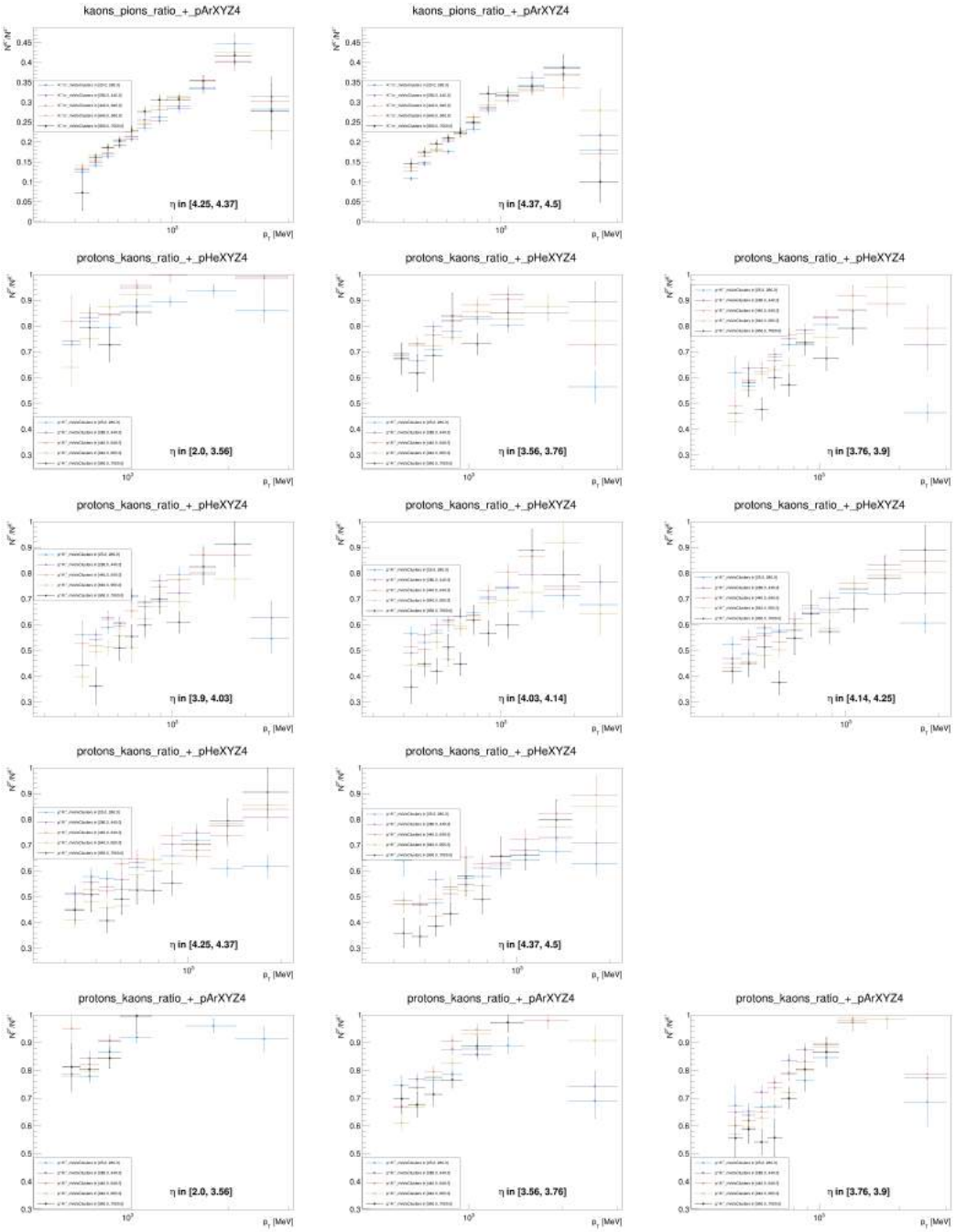

protons_pions_ratio_+_pHeXYZ4
η in [2.0, 3.56]

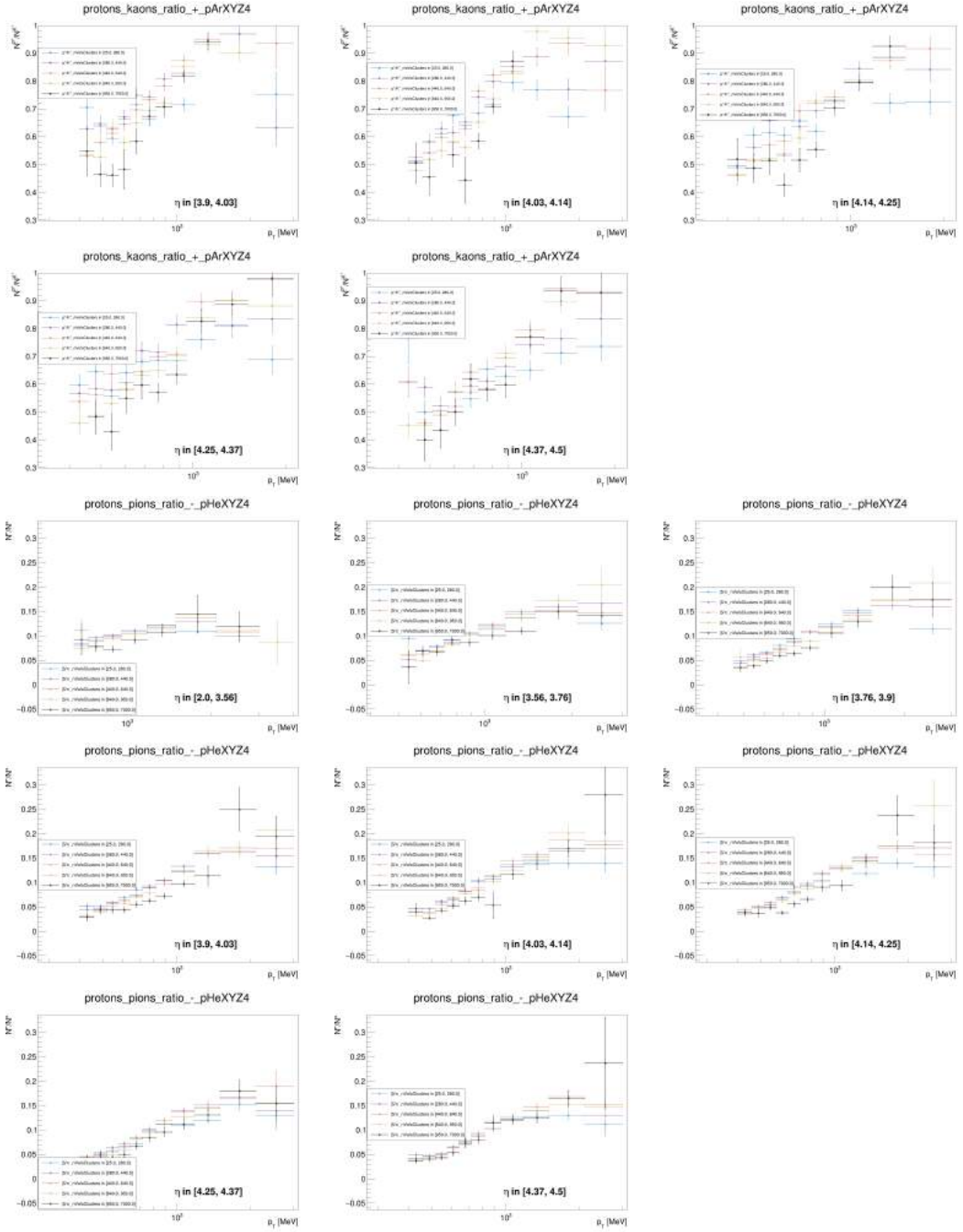

protons_pions_ratio_+_pHeXYZ4
η in [3.56, 3.76]

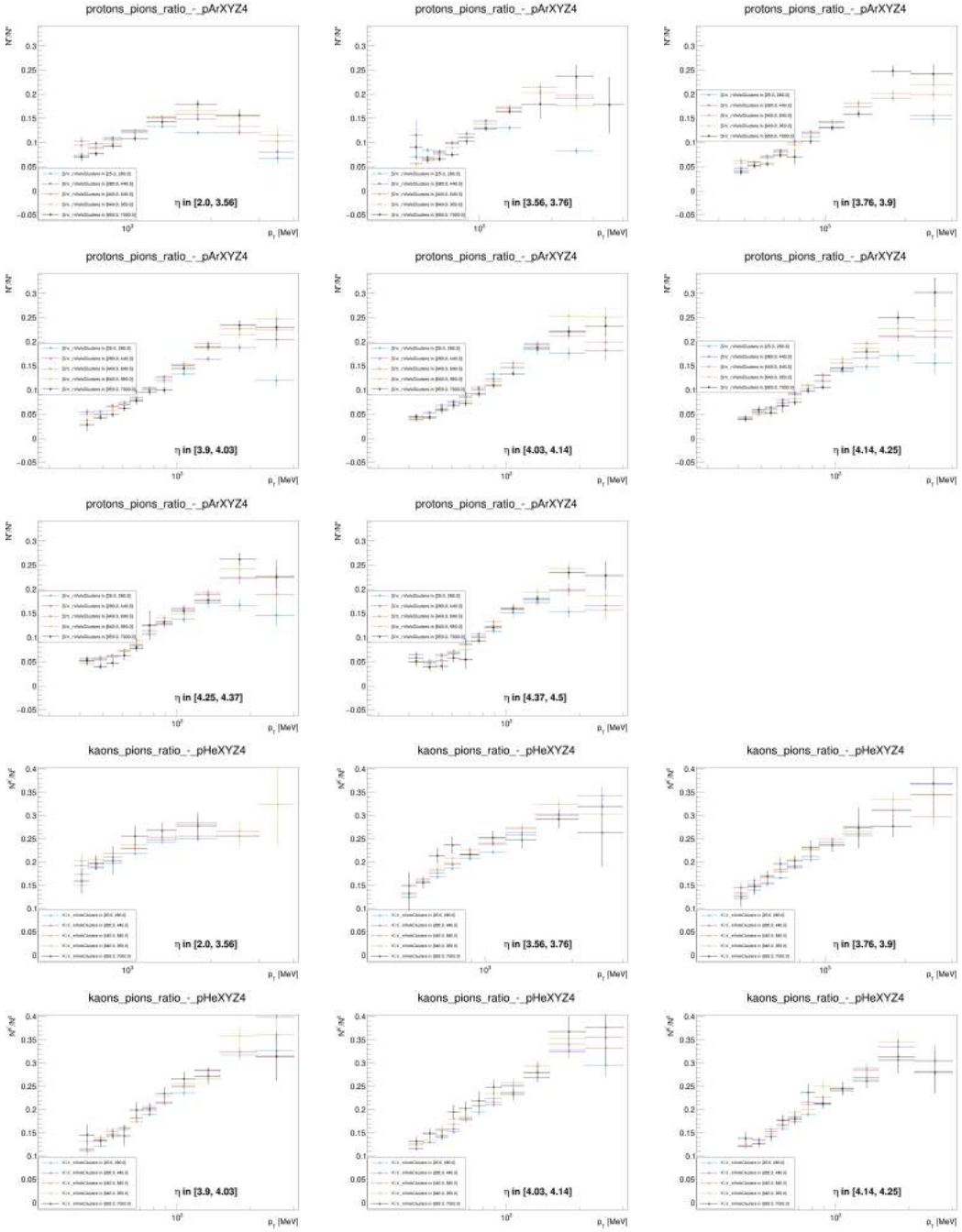

protons_pions_ratio_+_pHeXYZ4
η in [3.76, 3.9]

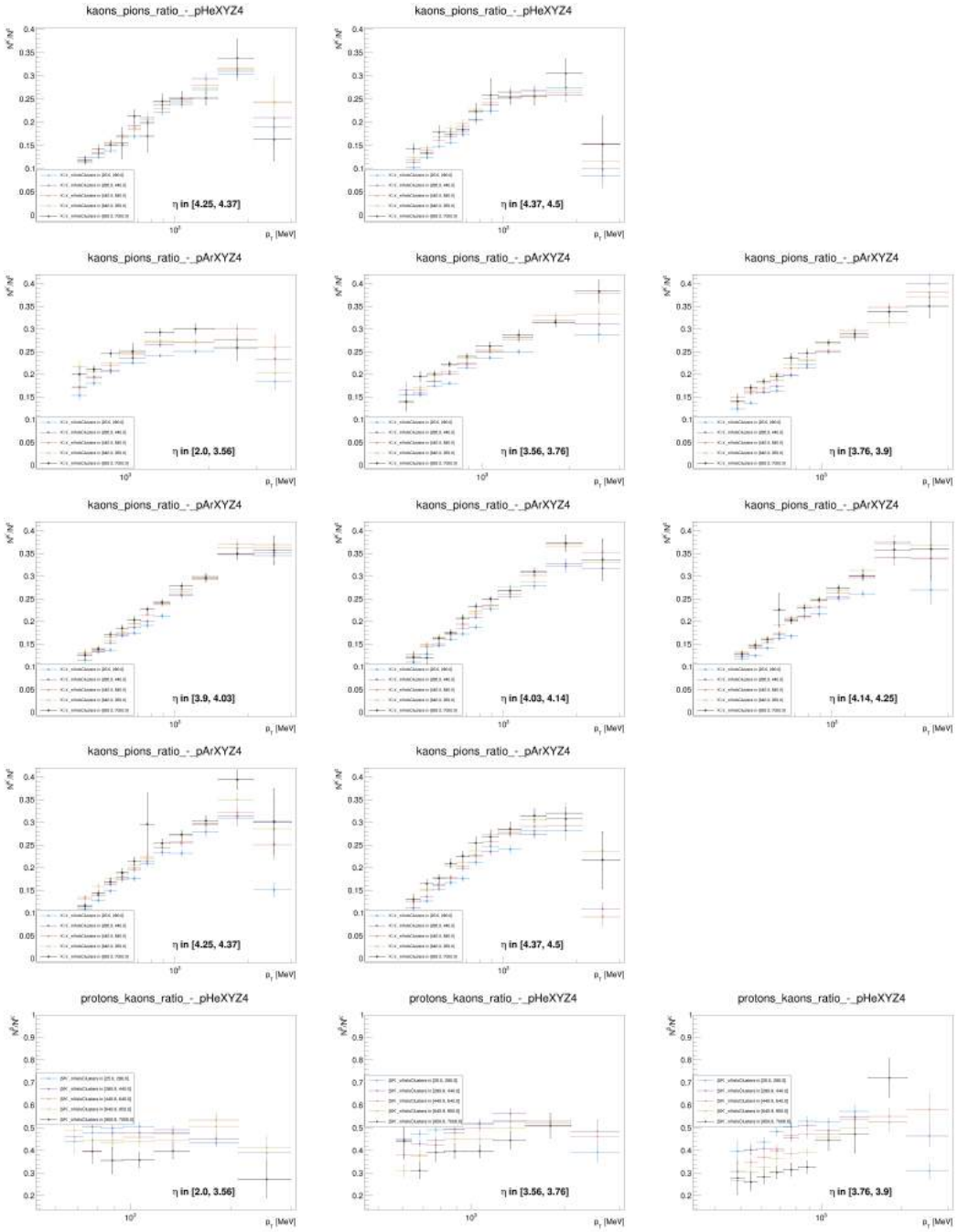

protons_pions_ratio_+_pHeXYZ4
η in [3.9, 4.03]
protons_pions_ratio_+_pHeXYZ4
η in [4.03, 4.14]
protons_pions_ratio_+_pHeXYZ4
η in [4.14, 4.25]
protons_pions_ratio_+_pHeXYZ4
η in [4.25, 4.37]
protons_pions_ratio_+_pHeXYZ4
η in [4.37, 4.5]
protons_pions_ratio_+_pArXYZ4
η in [2.0, 3.56]
protons_pions_ratio_+_pArXYZ4
η in [3.56, 3.76]
protons_pions_ratio_+_pArXYZ4
η in [3.76, 3.9]
protons_pions_ratio_+_pArXYZ4
η in [3.9, 4.03]
protons_pions_ratio_+_pArXYZ4
η in [4.03, 4.14]
protons_pions_ratio_+_pArXYZ4
η in [4.14, 4.25]
protons_pions_ratio_+_pArXYZ4
η in [4.25, 4.37]
protons_pions_ratio_+_pArXYZ4
η in [4.37, 4.5]

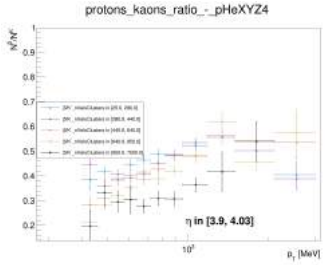

kaons_pions_ratio_+_pHeXYZ4
η in [2.0, 3.56]
kaons_pions_ratio_+_pHeXYZ4
η in [3.56, 3.76]
kaons_pions_ratio_+_pHeXYZ4
η in [3.76, 3.9]
kaons_pions_ratio_+_pHeXYZ4
η in [3.9, 4.03]
kaons_pions_ratio_+_pHeXYZ4
η in [4.03, 4.14]
kaons_pions_ratio_+_pHeXYZ4
η in [4.14, 4.25]
kaons_pions_ratio_+_pHeXYZ4
η in [4.25, 4.37]
kaons_pions_ratio_+_pHeXYZ4
η in [4.37, 4.5]
kaons_pions_ratio_+_pArXYZ4
η in [2.0, 3.56]
kaons_pions_ratio_+_pArXYZ4
η in [3.56, 3.76]
kaons_pions_ratio_+_pArXYZ4
η in [3.76, 3.9]
kaons_pions_ratio_+_pArXYZ4
η in [3.9, 4.03]
kaons_pions_ratio_+_pArXYZ4
η in [4.03, 4.14]
kaons_pions_ratio_+_pArXYZ4
η in [4.14, 4.25]

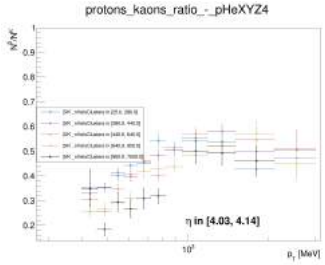

kaons_pions_ratio_+_pArXYZ4
η in [4.25, 4.37]
kaons_pions_ratio_+_pArXYZ4
η in [4.37, 4.5]
protons_kaons_ratio_+_pHeXYZ4
η in [2.0, 3.56]
protons_kaons_ratio_+_pHeXYZ4
η in [3.56, 3.76]
protons_kaons_ratio_+_pHeXYZ4
η in [3.76, 3.9]
protons_kaons_ratio_+_pHeXYZ4
η in [3.9, 4.03]
protons_kaons_ratio_+_pHeXYZ4
η in [4.03, 4.14]
protons_kaons_ratio_+_pHeXYZ4
η in [4.14, 4.25]
protons_kaons_ratio_+_pHeXYZ4
η in [4.25, 4.37]
protons_kaons_ratio_+_pHeXYZ4
η in [4.37, 4.5]
protons_kaons_ratio_+_pArXYZ4
η in [2.0, 3.56]
protons_kaons_ratio_+_pArXYZ4
η in [3.56, 3.76]
protons_kaons_ratio_+_pArXYZ4
η in [3.76, 3.9]

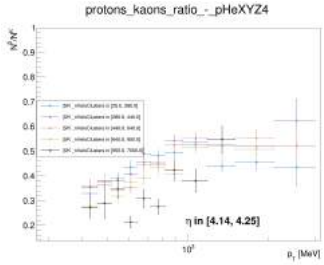

protons_kaons_ratio_+_pArXYZ4
η in [3.9, 4.03]
protons_kaons_ratio_+_pArXYZ4
η in [4.03, 4.14]
protons_kaons_ratio_+_pArXYZ4
η in [4.14, 4.25]
protons_kaons_ratio_+_pArXYZ4
η in [4.25, 4.37]
protons_kaons_ratio_+_pArXYZ4
η in [4.37, 4.5]
protons_pions_ratio_+_pHeXYZ4
η in [2.0, 3.56]
protons_pions_ratio_+_pHeXYZ4
η in [3.56, 3.76]
protons_pions_ratio_+_pHeXYZ4
η in [3.76, 3.9]
protons_pions_ratio_+_pHeXYZ4
η in [3.9, 4.03]
protons_pions_ratio_+_pHeXYZ4
η in [4.03, 4.14]
protons_pions_ratio_+_pHeXYZ4
η in [4.14, 4.25]
protons_pions_ratio_+_pHeXYZ4
η in [4.25, 4.37]
protons_pions_ratio_+_pHeXYZ4
η in [4.37, 4.5]

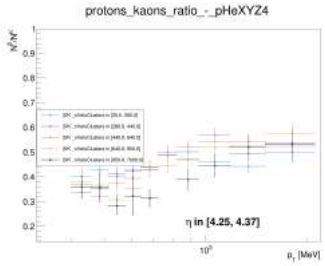

protons_pions_ratio_-_pArXYZ4
η in [2.0, 3.56]
protons_pions_ratio_-_pArXYZ4
η in [3.56, 3.76]
protons_pions_ratio_-_pArXYZ4
η in [3.76, 3.9]
protons_pions_ratio_-_pArXYZ4
η in [3.9, 4.03]
protons_pions_ratio_-_pArXYZ4
η in [4.03, 4.14]
protons_pions_ratio_-_pArXYZ4
η in [4.14, 4.25]
protons_pions_ratio_-_pArXYZ4
η in [4.25, 4.37]
protons_pions_ratio_-_pArXYZ4
η in [4.37, 4.5]
kaons_pions_ratio_-_pHeXYZ4
η in [2.0, 3.56]
kaons_pions_ratio_-_pHeXYZ4
η in [3.56, 3.76]
kaons_pions_ratio_-_pHeXYZ4
η in [3.76, 3.9]
kaons_pions_ratio_-_pHeXYZ4
η in [3.9, 4.03]
kaons_pions_ratio_-_pHeXYZ4
η in [4.03, 4.14]
kaons_pions_ratio_-_pHeXYZ4
η in [4.14, 4.25]

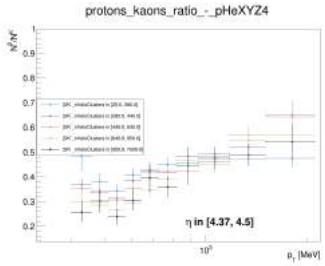

kaons_pions_ratio_-_pHeXYZ4
η in [4.25, 4.37]
kaons_pions_ratio_-_pHeXYZ4
η in [4.37, 4.5]
kaons_pions_ratio_-_pArXYZ4
η in [2.0, 3.56]
kaons_pions_ratio_-_pArXYZ4
η in [3.56, 3.76]
kaons_pions_ratio_-_pArXYZ4
η in [3.76, 3.9]
kaons_pions_ratio_-_pArXYZ4
η in [3.9, 4.03]
kaons_pions_ratio_-_pArXYZ4
η in [4.03, 4.14]
kaons_pions_ratio_-_pArXYZ4
η in [4.14, 4.25]
kaons_pions_ratio_-_pArXYZ4
η in [4.25, 4.37]
kaons_pions_ratio_-_pArXYZ4
η in [4.37, 4.5]
protons_kaons_ratio_-_pHeXYZ4
η in [2.0, 3.56]
protons_kaons_ratio_-_pHeXYZ4
η in [3.56, 3.76]
protons_kaons_ratio_-_pHeXYZ4
η in [3.76, 3.9]

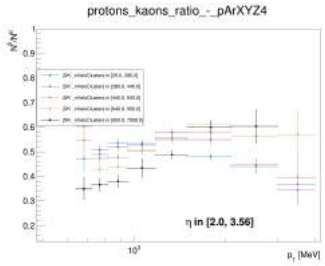

protons_kaons_ratio_-_pHeXYZ4
η in [3.9, 4.03]

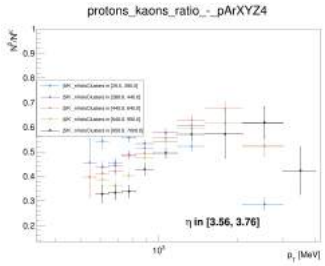

protons_kaons_ratio_-_pHeXYZ4
η in [4.03, 4.14]

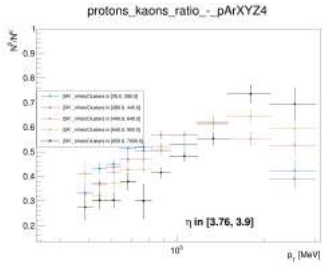

protons_kaons_ratio_-_pHeXYZ4
η in [4.14, 4.25]

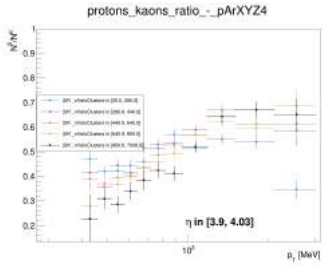

protons_kaons_ratio_-_pHeXYZ4
η in [4.25, 4.37]

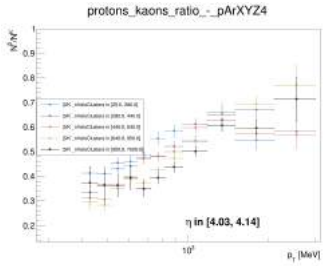

protons_kaons_ratio_-_pHeXYZ4
η in [4.37, 4.5]

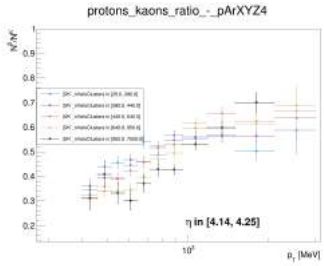

protons_kaons_ratio_-_pArXYZ4
η in [2.0, 3.56]

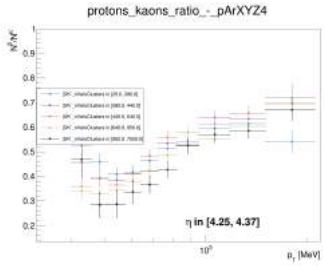

protons_kaons_ratio_-_pArXYZ4
η in [3.56, 3.76]

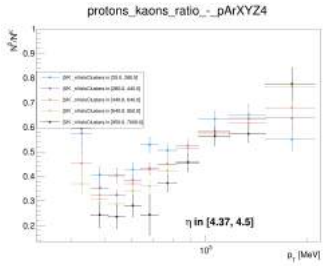

protons_kaons_ratio_-_pArXYZ4
η in [3.76, 3.9]


protons_kaons_ratio_-_pArXYZ4
η in [3.9, 4.03]

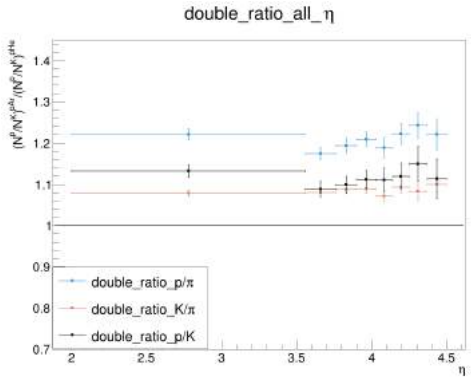

protons_kaons_ratio_-_pArXYZ4
η in [4.03, 4.14]

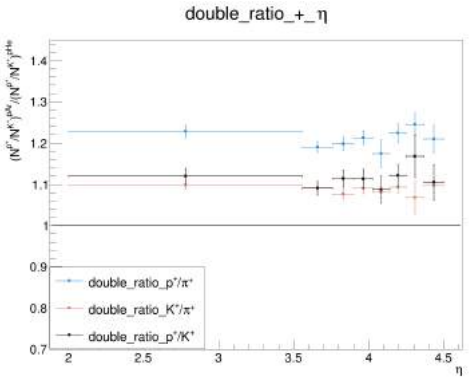

protons_kaons_ratio_-_pArXYZ4
η in [4.14, 4.25]

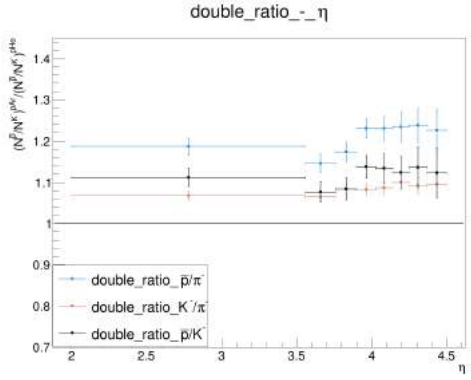

protons_kaons_ratio_-_pArXYZ4
η in [4.25, 4.37]

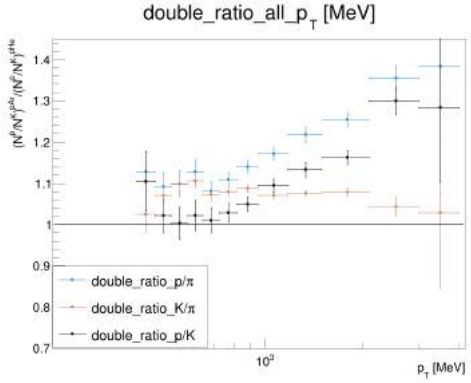

protons_kaons_ratio_-_pArXYZ4
η in [4.37, 4.5]

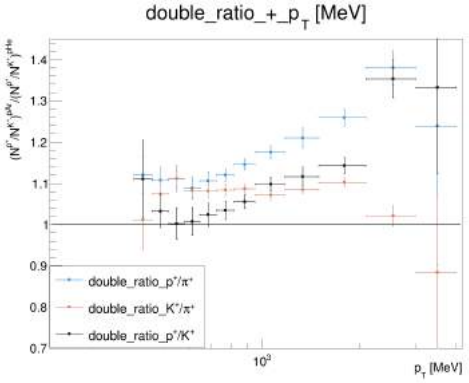
.6 1D double ratios n in η and p_T

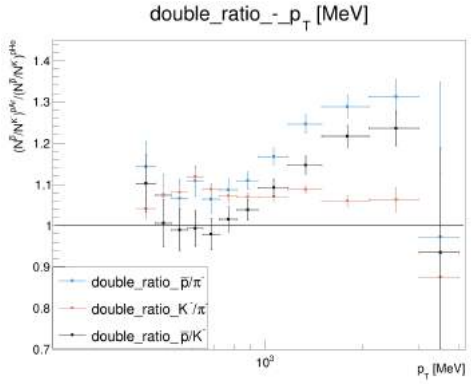




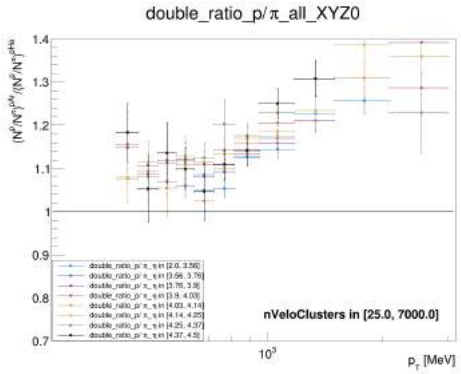


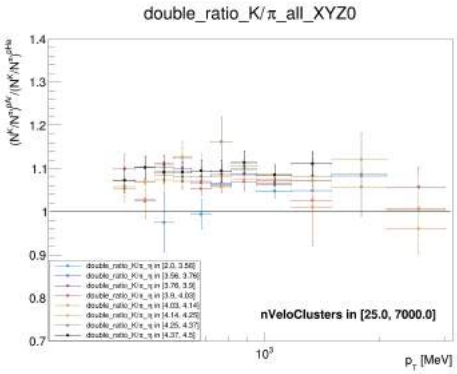


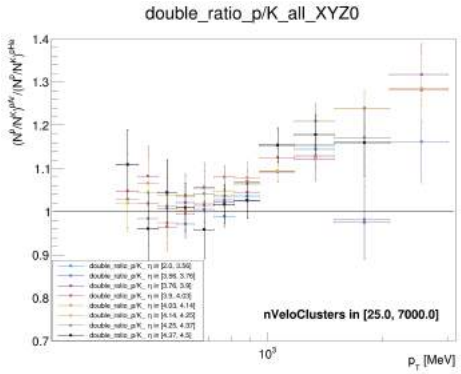


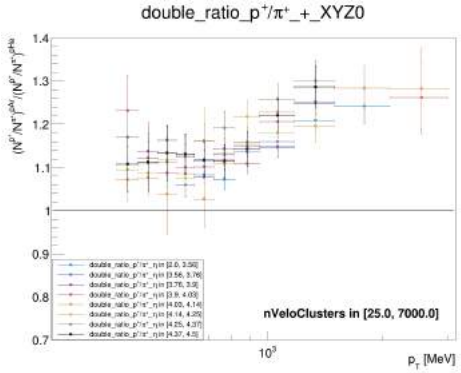


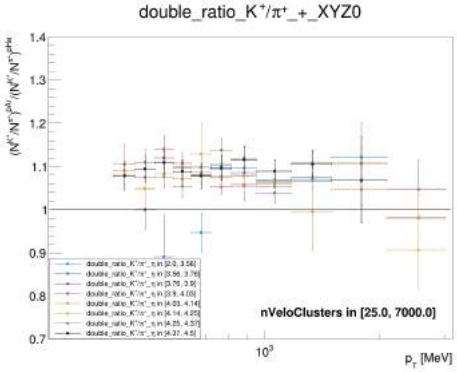
.7 2D double ratios in η-p_T and p_T-nVeloClusters

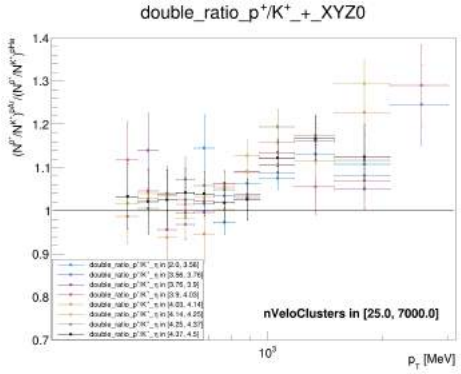


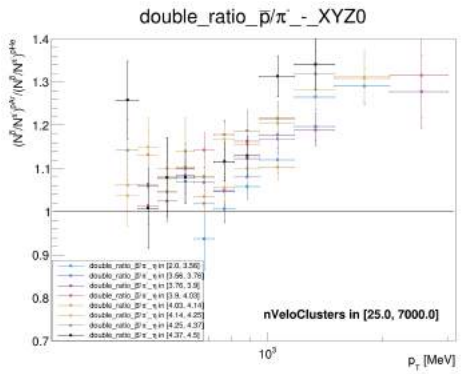


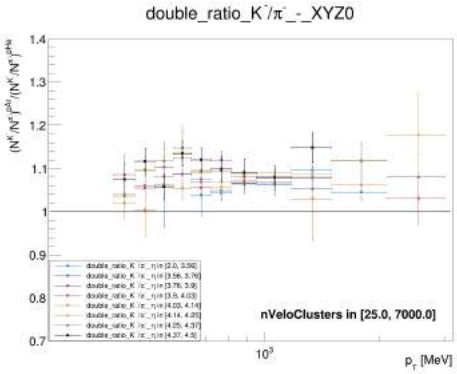


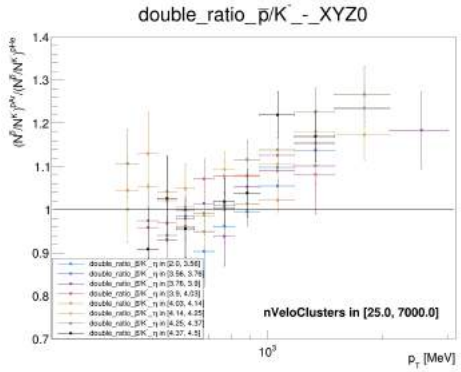


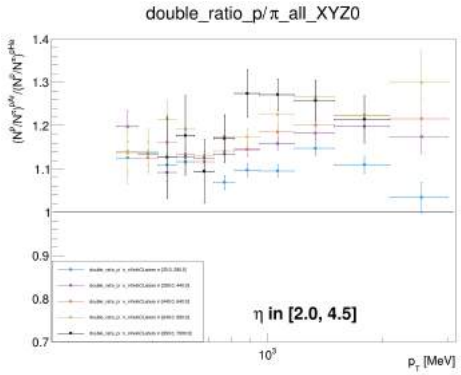


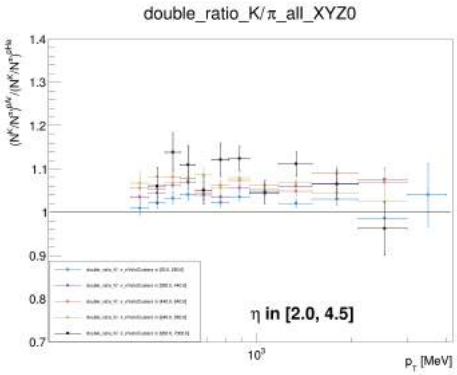


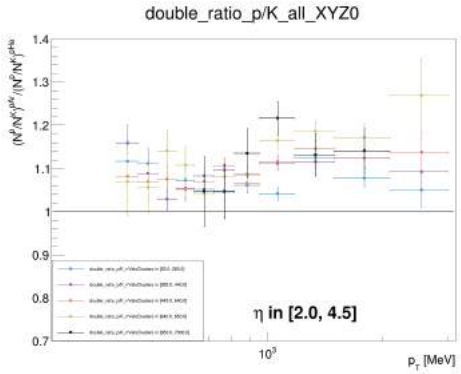

double_ratio_p̄/π⁻_-_XYZ0
nVeloClusters in [25.0, 7000.0]
p_T [MeV]

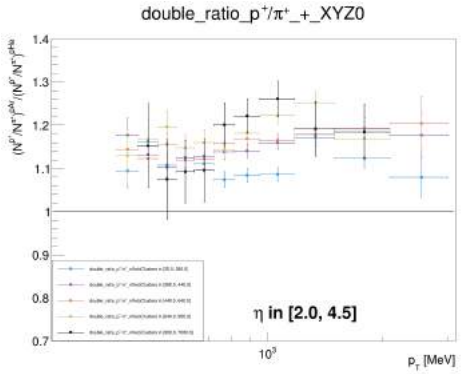

double_ratio_K⁻/π⁻_-_XYZ0
nVeloClusters in [25.0, 7000.0]
p_T [MeV]

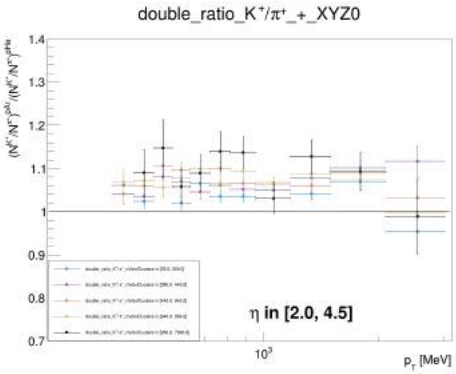

double_ratio_p̄/K⁻_-_XYZ0
nVeloClusters in [25.0, 7000.0]
p_T [MeV]

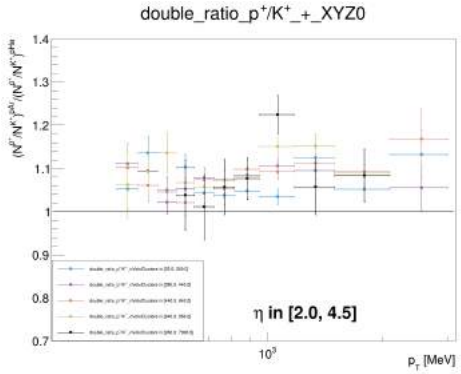

double_ratio_p/π_all_XYZ0
η in [2.0, 4.5]
p_T [MeV]

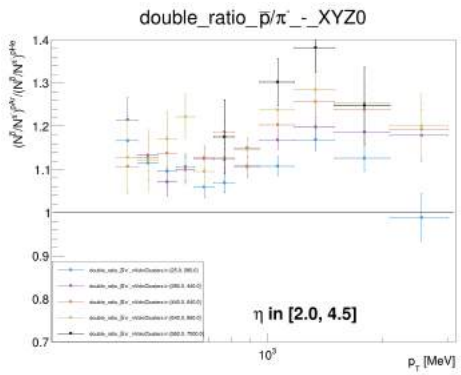

double_ratio_K/π_all_XYZ0
η in [2.0, 4.5]
p_T [MeV]

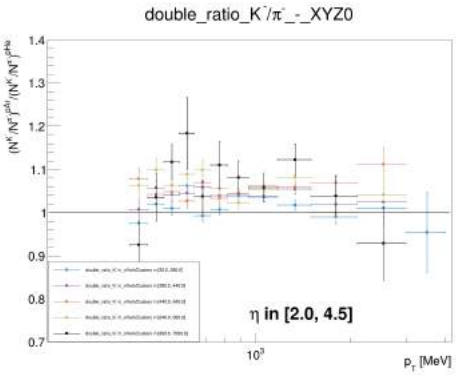

double_ratio_p/K_all_XYZ0
η in [2.0, 4.5]
p_T [MeV]

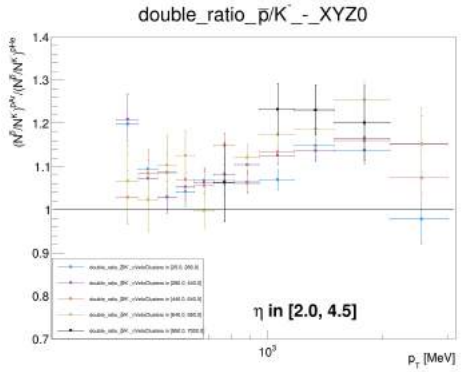

double_ratio_p⁺/π⁺_+_XYZ0
η in [2.0, 4.5]
p_T [MeV]


double_ratio_K⁺/π⁺_+_XYZ0
η in [2.0, 4.5]
p_T [MeV]


double_ratio_p⁺/K⁺_+_XYZ0
η in [2.0, 4.5]
p_T [MeV]


double_ratio_p̄/π⁻_-_XYZ0
η in [2.0, 4.5]
p_T [MeV]


double_ratio_K⁻/π⁻_-_XYZ0
η in [2.0, 4.5]
p_T [MeV]


double_ratio_p̄/K⁻_-_XYZ0
η in [2.0, 4.5]
p_T [MeV]

Bibliography

[1] Alfred H Mueller. *Perturbative quantum chromodynamics*. Vol. 5. World Scientific, 1989.

[2] James D Bjorken. "Asymptotic sum rules at infinite momentum". In: *Physical Review* 179.5 (1969), p. 1547.

[3] Jean-Jacques Aubert et al. "The ratio of the nucleon structure functions F2n for iron and deuterium". In: *Physics Letters B* 123.3-4 (1983), pp. 275–278.

[4] Kari J Eskola, VJ Kolhinen, and CA Salgado. "The scale dependent nuclear effects in parton distributions for practical applications". In: *The European Physical Journal C-Particles and Fields* 9.1 (1999), pp. 61–68.

[5] KJ Eskola, Hannu Paukkunen, and CA Salgado. "EPS09—a new generation of NLO and LO nuclear parton distribution functions". In: *Journal of High Energy Physics* 2009.04 (2009), p. 065.

[6] Kari J Eskola et al. "EPPS16: Nuclear parton distributions with LHC data". In: *The European Physical Journal C* 77 (2017), pp. 1–28.

[7] K Kovařík et al. "nCTEQ15: Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework". In: *Physical Review D* 93.8 (2016), p. 085037.

[8] Daniel de Florian et al. "Global analysis of nuclear parton distributions". In: *Physical Review D* 85.7 (2012), p. 074028.

[9] D Antreasyan et al. "Production of π+ and π- at Large Transverse Momentum in p- p and p- d Collisions at 200, 300, and 400 GeV". In: *Physical Review Letters* 38.3 (1977), p. 112.

[10] JW Cronin et al. "Production of hadrons at large transverse momentum at 200, 300, and 400 GeV". In: *Physical Review D* 11.11 (1975), p. 3105.

[11] BZ Kopeliovich et al. "Cronin effect in hadron production off nuclei". In: *Physical review letters* 88.23 (2002), p. 232303.

[12] C Aidala et al. "Nuclear-modification factor of charged hadrons at forward and backward rapidity in p+ Al and p+ Au collisions at s N N= 200 GeV". In: *Physical review C* 101.3 (2020), p. 034910.

[13] AH Rezaeian and Zhun Lu. "Cronin effect for protons and pions in high-energy pA collisions". In: *Nuclear Physics A* 826.1-2 (2009), pp. 198–210.

[14] I Arsene et al. "Evolution of the Nuclear Modification Factors with Rapidity and Centrality in d+ A u Collisions at N N S= 200 G e V". In: *Physical review letters* 93.24 (2004), p. 242303.

[15] John Adams et al. “Forward neutral pion production in p+ p and d+ Au collisions at s N N= 200 GeV”. In: *Physical review letters* 97.15 (2006), p. 152302.

[16] Stephen Scott Adler et al. “Identified charged particle spectra and yields in Au+ Au collisions at s N N= 200 GeV”. In: *Physical Review C* 69.3 (2004), p. 034909.

[17] BZ Kopeliovich and J Nemchik. “Challenges of high-pT processes on nuclei”. In: *Journal of Physics G: Nuclear and Particle Physics* 38.4 (2011), p. 043101.

[18] BZ Kopeliovich et al. “Quenching of high-p T hadrons: Energy loss versus color transparency”. In: *Physical Review C* 86.5 (2012), p. 054904.

[19] Betty Abelev et al. “Transverse momentum distribution and nuclear modification factor of charged particles in p+ Pb collisions at s N N= 5.02 TeV”. In: *Physical review letters* 110.8 (2013), p. 082302.

[20] John Adams et al. “Identified hadron spectra at large transverse momentum in p+ p and d+ Au collisions at sNN= 200 GeV”. In: *Physics Letters B* 637.3 (2006), pp. 161–169.

[21] Nestor Armesto et al. “Comparison of jet quenching formalisms for a quark-gluon plasma “brick””. In: *Physical Review C* 86.6 (2012), p. 064904.

[22] Alfred H Mueller and Jianwei Qiu. “Gluon recombination and shadowing at small values of x”. In: *Nuclear Physics B* 268.2 (1986), pp. 427–452.

[23] Leonid Vladimirovič Gribov, Eugene M Levin, and Michail G Ryskin. “Semihard processes in QCD”. In: *Physics Reports* 100.1-2 (1983), pp. 1–150.

[24] Gyula Bencedi. “Nuclear modification factor of charged particles and light-flavour hadrons in p–Pb collisions measured by ALICE”. In: *arXiv preprint arXiv:1609.05665* (2016).

[25] Jaroslav Adam et al. “Multiplicity dependence of charged pion, kaon, and (anti) proton production at large transverse momentum in p–Pb collisions at sNN= 5.02 TeV”. In: *Physics Letters B* 760 (2016), pp. 720–735.

[26] Christian Klein-Bösing, PHENIX Collaboration, et al. “PHENIX measurement of high pT particles in Au+ Au and d+ Au collisions at”. In: *Journal of Physics G: Nuclear and Particle Physics* 30.8 (2004), S975.

[27] A Airapetian et al. “Hadronization in semi-inclusive deep-inelastic scattering on nuclei”. In: *Nuclear Physics B* 780.1-2 (2007), pp. 1–27.

[28] A Airapetian et al. “Transverse momentum broadening of hadrons produced in semi-inclusive deep-inelastic scattering on nuclei”. In: *Physics Letters B* 684.2-3 (2010), pp. 114–118.

[29] Lyndon Evans and Philip Bryant. “LHC machine”. In: *Journal of instrumentation* 3.08 (2008), S08001.

[30] Georges Aad et al. “The ATLAS experiment at the CERN large hadron collider”. In: *Jinst* 3 (2008), S08003.

[31] Kenneth Aamodt et al. “The ALICE experiment at the CERN LHC”. In: *Journal of Instrumentation* 3.08 (2008), S08002.

[32] Roman Adolphi et al. “The CMS experiment at the CERN LHC”. In: *Jinst* 803 (2008), S08004.

[33] A Augusto Alves Jr et al. “The LHCb detector at the LHC”. In: *Journal of instrumentation* 3.08 (2008), S08005.

[34] LHCb Collaboration. “LHCb detector performance”. In: *International Journal of Modern Physics A* 30.07 (2015), p. 1530022.

[35] Ralf Bernhard et al. “The LHCb silicon tracker”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 596.1 (2008), pp. 17–20.

[36] Roel Aaij et al. “Performance of the LHCb vertex locator”. In: *Journal of Instrumentation* 9.09 (2014), P09007.

[37] S Amato et al. “LHCb magnet: Technical design report”. In: (2000).

[38] LHCb LHCb. *Inner tracker technical design report*. Tech. rep. CERN-LHCC-2002-029, 2002.

[39] PR Barbosa Marinho et al. “LHCb outer tracker: Technical design report”. In: (2001).

[40] TDR LHCb. *RICH technical design report*. Tech. rep. Citeseer, 2000.

[41] Antonis Papanestis, Carmelo D’Ambrosio, LHCb RICH Collaboration, et al. “Performance of the LHCb RICH detectors during the LHC Run II”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 876 (2017), pp. 221–224.

[42] O Omelaenko et al. *LHCb Calorimeters: Technical Design Report*. Tech. rep. LHCb-TDR-002, 2000.

[43] A Augusto Alves Jr et al. “Performance of the LHCb muon system”. In: *Journal of Instrumentation* 8.02 (2013), P02022.

[44] LHCb Collaboration et al. *LHCb tracker upgrade technical design report*. Tech. rep. 2014.

[45] Albert Puig. *The LHCb trigger in 2011 and 2012*. Tech. rep. 2014.

[46] LHCb Collaboration et al. “LHCb Trigger and Online Upgrade Technical Design Report”. In: (2014).

[47] LHCb Collaboration et al. *LHCb upgrade GPU high level trigger technical design report*. Tech. rep. 2020.

[48] LHCb Collaboration et al. *LHCb VELO upgrade technical design report*. Tech. rep. 2013.

[49] Roel Aaij et al. “Precision luminosity measurements at LHCb”. In: *arXiv preprint arXiv:1410.0149* (2014).

[50] Chris Parkes et al. *LHCb PLUME: Probe for LUminosity MEasurement*. Tech. rep. CERN, 2021.

[51] Colin Barschel. “Precision luminosity measurement at LHCb with beam-gas imaging”. PhD thesis. RWTH Aachen U., 2014.

[52] SJ Brodsky et al. “Physics opportunities of a fixed-target experiment using LHC beams”. In: *Physics Reports* 522.4 (2013), pp. 239–255.

[53] Roel Aaij et al. “Measurement of Antiproton Production in p- He Collisions at s NN= 110 GeV”. In: *Physical Review Letters* 121.22 (2018), p. 222001.

[54] Roel Aaij et al. “First measurement of charm production in its fixed-target configuration at the LHC”. In: *Physical review letters* 122.13 (2019), p. 132002.

[55] LHCb Collaboration et al. “Charmonium production in p Ne collisions at \sqrt {s_ {\rm NN}}= 68.5 GeV”. In: *arXiv preprint arXiv:2211.11645* (2022).

[56] LHCb Collaboration et al. “$J/$\psi and D^ 0$productionin$\sqrt {s_ {\rm NN}} = 68.5 GeV PbNe collisions”. In: *arXiv preprint arXiv:2211.11652* (2022).

[57] LHCb Collaboration et al. “Measurement of antiproton production from antihyperon decays in pHe collisions at \sqrt {s_ {NN}}= 110 GeV”. In: *arXiv preprint arXiv:2205.09009* (2022).

[58] LHCb Collaboration et al. *LHCb PID upgrade technical design report*. Tech. rep. 2013.

[59] L. Cassina A. Cotta Ramusino P. Dorosz M. Fiorini C. Gotti W. Kucewicz M. Maino R. Malaguti G. Pessina M. Baszczyk P. Carniti. “CLARO8V3 EXTENDED MANUAL REV 0.4”. In: *Internal note* (2016).

[60] Lorenzo Cassina. “LHCb RICH Upgrade: an overview of the photon detector and electronic system”. In: *Journal of Instrumentation* 11.01 (2016), p. C01025.

[61] Marta Calvi et al. “Characterization of the Hamamatsu H12700A - 03 and R12699 - 03 multi-anode photomultiplier tubes”. In: *Journal of Instrumentation* 10.09 (2015), P09021.

[62] S Gambetta, LHCb RICH Collaboration, et al. “First results from quality assurance testing of MaPMTs for the LHCb RICH upgrade”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 876 (2017), pp. 206–208.

[63] Mirco Andreotti et al. “Characterisation of signal-induced noise in Hamamatsu R11265 Multi-anode Photomultiplier Tubes”. In: *Journal of Instrumentation* 16.11 (2021), P11030.

[64] Federico Alessio and Richard Jacobsson. *System-level Specifications of the Timing and Fast Control system for the LHCb Upgrade*. Tech. rep. 2014.

[65] B Franek and C Gaspar. “SMI++ object oriented framework used for automation and error recovery in the LHC experiments”. In: *Journal of Physics: Conference Series*. Vol. 219. 2. IOP Publishing. 2010, p. 022031.

[66] Pietro Albicocco et al. “A Method Based on Muon System to Monitor LHCb Luminosity”. In: *Symmetry* 14.5 (2022), p. 860.

[67] Giacomo Graziani et al. “A Neural-Network-defined Gaussian Mixture Model for particle identification applied to the LHCb fixed-target programme”. In: *Journal of Instrumentation* 17.02 (2022), P02018.

[68] Roel Aaij et al. “Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2”. In: *EPJ Techniques and Instrumentation* 6.1 (2019), p. 1.

[69] J Podolanski and R Armenteros. “III. Analysis of V-events”. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 45.360 (1954), pp. 13–30.

[70] Muriel Pivk and Francois R Le Diberder. “Plots: A statistical tool to unfold data distributions”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 555.1-2 (2005), pp. 356–369.

[71] Hans Dembinski et al. “Custom Orthogonal Weight functions (COWs) for event classification”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1040 (2022), p. 167270.

[72] Adam Paszke et al. “Automatic differentiation in pytorch”. In: (2017).

[73] Fabian Pedregosa et al. “Scikit-learn: Machine learning in Python”. In: *the Journal of machine Learning research* 12 (2011), pp. 2825–2830.

[74] T Pierog et al. “EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider”. In: *Physical Review C* 92.3 (2015), p. 034906.

[75] Alex Rogozhnikov. “Reweighting with boosted decision trees”. In: *Journal of Physics: Conference Series*. Vol. 762. 1. IOP Publishing. 2016, p. 012036.