Università degli Studi di Ferrara

DOCTORAL COURSE IN

PHYSICS

CYCLE XXXVI

COORDINATOR

Prof. Luppi Eleonora

Study of cold nuclear matter effects in fixed-target collisions and charged hadron identification performance at the LHCb experiment

Scientific-Disciplinary Sector (SDS): FIS/01

Candidate
Shinichi Okamura

Supervisors
Prof. Roberto Calabrese
Prof. Massimiliano Fiorini
Prof. Luciano Libero Pappalardo

Years 2020/2024

Abstract

The investigation of charged-particle production in hadronic collisions is crucial for understanding strong interactions governed by Quantum Chromodynamics (QCD). While perturbative QCD effectively describes hard interactions, the soft regime remains less understood. This work presents the measurement of single and double ratios of charged light-hadron production in PbNe and pNe collisions, collected by the LHCb experiment in Run 2 at a centre of mass energy per nucleon $\sqrt{s_{NN}} = 69$ GeV. Significant enhancements have been observed in baryon-meson ratios in PbNe data, suggesting the potential onset of the Cronin Effect for $p_T > 1$ GeV/c. The results are compared with EPOS simulations, revealing substantial differences. These findings can provide inputs for simulation tuning, enhancing the current understanding of Cold Nuclear Matter effects and related QCD phenomena. One of the main challenges of the analysis is related to charged-hadron identification performance, which primarily relies on the LHCb Ring Imaging Cherenkov system. The LHCb upgrade, designed to operate at an instantaneous luminosity increased by a factor of five compared to previous data-taking, introduces several enhancements in particle identification performance with upgraded RICH detectors. This improvement can extend the possibilities for fixed-target and heavy-ion measurements at LHCb, allowing for future studies in higher centrality and a broader kinematic region. In this context, the development of calibration tools plays a central role in optimising the performance of the upgraded LHCb detector. A software framework has been developed to comprehensively study the particle identification (PID) performance. The results of charged-hadron identification performance with early Run 3 pp data are presented, and the first study on PID performance for fixed-target and heavy-ion collisions is also included. Despite the excellent early performance of the RICH systems, there is room for optimisation to attain the ultimate PID performance. In this regard, online monitoring of the PID performance curves serves as a crucial metric for fine-tuning RICH detectors across various settings. A dedicated section of this work focuses on projections and the requirements for the implementation of online monitoring for PID performance.

Acknowledgements

Desidero esprimere la mia sincera gratitudine per il Prof. Roberto Calabrese, senza il quale questo percorso di dottorato non sarebbe stato possibile. Provo una profonda stima nei suoi confronti e lo ringrazio per la sua guida preziosa durante tutto il mio percorso accademico. Vorrei ringraziare profondamente il Prof. Massimiliano Fiorini per avermi sempre coinvolto in una vasta gamma di attività di ricerca e avermi offerto numerose opportunità di crescita che sono state fondamentali per il mio sviluppo sia accademico che personale. Desidero esprimere i miei sinceri ringraziamenti anche al Prof. Luciano Libero Pappalardo. È stata una figura di riferimento cruciale e lo ringrazio per l'impegno e la cura con cui mi ha seguito durante tutto il percorso di dottorato.

Vorrei offrire un ringraziamento speciale a Silvia, Giovanni e Saverio per tutto il supporto che mi hanno dato e per essere stati dei punti di riferimento durante l'ultimo anno di dottorato. Provo una grandissima stima nei loro confronti e li reputo delle persone eccezionali non solo dal punto di vista professionale ma soprattutto dal punto di vista umano. Ringrazio sinceramente anche tutti i colleghi di Unife, in particolare Edoardo, Riccardo, Viola, Carmen, Niccolo, Marco, e Capriot per la loro compagnia preziosa e il loro incoraggiamento nel corso degli anni.

Vorrei ringraziare dal profondo del cuore Mamma, Papà e Daso. Vi ringrazio per tutti gli sforzi che avete fatto per avermi dato la possibilità di intraprendere questo percorso, per avermi sempre dato il massimo sostegno e per avermi costantemente incoraggiato a fare del mio meglio. Grazie di cuore per aver sempre creduto in me.

Ringrazio di cuore Linda, per aver condiviso con me i momenti di gioia ma soprattutto per avermi supportato fino in fondo durante i momenti difficili con pazienza, affetto e amore. Grazie per esserci sempre e per tutto ciò che hai fatto per me. Mi sento estremamente fortunato ad averti accanto.

Un grazie di cuore a Federica, Pietro e Lori Malentacca per il loro affetto, per i bellissimi momenti passati assieme, e per avermi fatto sentire come a casa. Grazie agli amici del CERN, in particolare Giorgia, Sara, Chiara, Scarabot e Andre Villa per la loro serenità, saggezza, sportività, simpatia e per aver condiviso il sogno degli EFS. Non smettete mai di volare!

Ringrazio sinceramente anche gli amici di Ferrara, in particolare Pippo, Diego, Fernet, Nico, Cori, Zanfo e Gambe per la loro compattezza e per essermi sempre stati vicino nonostante la lontananza.

Grazie di cuore a tutti. Sono convinto che non ce l'avrei fatta senza tutti voi e sono estremamente grato per il vostro sostegno costante.

Contents

Chapter 1

Introduction

The LHCb experiment at CERN stands as a major experiment located along the LHC. Distinguished by its single-arm spectrometer design, the LHCb detector sets itself apart from the other LHC experiments. Achieving precise Particle Identification (PID) is crucial for Charge Parity (CP) violation studies and the exploration of b-quark decays, forming the core objectives of the experiment. LHCb is specifically optimised for identifying and differentiating light hadrons. Identification of charged hadrons at LHCb is provided by two Ring Imaging Cherenkov (RICH) detectors: RICH1, located upstream of the magnet and employing a C_4F_{10} gas radiator, and RICH2, located downstream of the magnet and employing a CF_4 gas radiator.

Throughout its operational history, the LHCb experiment has expanded significantly its physics reach beyond the initial objectives, and currently can also be operated in a fixed-target mode, utilising collisions between LHC beams and gas targets. In 2011, the LHCb experiment has been equipped with the System for Measuring Overlap with Gas (SMOG) with the original scope of providing precise measurement of the LHC luminosity. The flexibility of the detector and its operation, coupled with the capability to introduce noble gases into the LHC beam pipe using SMOG, enabled in 2015 a pioneering fixed-target program, exploiting collisions between LHC beams and gaseous targets. The unique combination of forward coverage, precision vertexing, excellent particle identification, and a fast data acquisition system provides LHCb with distinctive capabilities for studying both small and large collision systems. Notably, exploring proton-nucleus collision data in an unexplored energy range facilitates distinct measurements of light-hadron production with various target systems.

Investigating the production of charged particles in hadronic collisions is essential for exploring the properties of the strong interaction encoded by Quantum Chromodynamics (QCD). At the LHC, high-energy collisions generate charged particles through both soft and hard interactions, corresponding to small and large momentum exchanges between the interacting partons of the hadrons, respectively. While perturbative QCD (pQCD) effectively describes hard interactions, the soft regime remains less understood, relying on phenomenological considerations. The exploration of the hard regime, characterised by charged particles with high transverse momentum (p_T), provides valuable insights into the physics of heavy-ion collisions. Analysing modifications in the charged-particle production rate in proton-nucleus (pA) collisions with respect to

pp collisions, allows for the study of various cold nuclear matter (CNM) effects. Recent observations of collective fluid-like phenomena in small systems suggest dynamics beyond traditional CNM effects, indicative of a quark-gluon plasma. These modifications, particularly concerning charged particles, are often linked to initial-state effects parameterised in terms of Nuclear Parton Distribution Functions (nPDFs). Other nuclear effects involve initial- or final-state multiple scatterings of incoming and outgoing partons, potentially manifesting in an enhancement of particle production at large p_T (Cronin effect). Models based on parton saturation, occurring at low values of the parton momentum fraction (x) and involving heavy nuclei, provide an alternative approach. In this regime, QCD dynamics can be described by the Color-Glass Condensate (CGC) effective field theory.

Experimental input is crucial, especially as charged particles from soft interactions dominate even at LHC energies, influencing models and generators for both hadron collider and cosmic ray physics. This work aims to study cold nuclear matter effects in two different collision systems, offering new opportunities to investigate various CNM effects. Despite the LHCb unique capabilities to investigate both small and large collision systems, one of the main challenges of the analysis is related to detector performance, especially in particle identification at high centrality, where the CNM effects should be enhanced. This is due to the expected degradation in PID performance with increasing event multiplicity.

The upgrade of the LHCb detector, which has been installed during the second long shutdown (LS2) of the LHC, is geared towards enhancing data collection by implementing a fivefold increase in instantaneous luminosity compared to previous data-taking periods. Specifically, to withstand this substantial increase, the optical system of RICH1 has undergone a comprehensive redesign. Moreover, the entire opto-electronics chain of both RICH1 and RICH2 has been modified to enable detector readout at 40 MHz and provide single-photon counting capabilities with a repetition fraction of up to 100 MHz/cm^2. This new configuration is anticipated to enable the discrimination between various species of long-lived charged hadrons with momenta ranging from 2.6 GeV/c up to beyond 100 GeV/c in events where the average number of visible interactions in LHCb is up to 7, thereby restoring the excellent performance achieved in Run 1 and Run 2.

Ensuring high performance in charged-particle identification becomes a crucial aspect of the LHCb physics program in Run 3, encompassing pp, fixed-target, and heavy-ion collisions. Consequently, the development of calibration tools will play a central role in optimising the performance of the upgraded LHCb detector to maximise the outcomes of physics measurements. The second part of this work is dedicated to a detailed study on charged-hadron identification performance, where the separation power between pions, kaons, and protons is evaluated using control samples collected by the LHCb experiment during the initial data-taking periods in 2022 and 2023.

The thesis first introduces the necessary theoretical foundation of inclusive charged-particle production in pA collisions in Chapter 2. This chapter describes the basics of QCD, the quantum field theory of strong interactions, and provides an overview of various phenomenological approaches aimed at comprehending inclusive charged-particle production in proton-ion collisions.

Chapter 3 delves into the LHCb experimental environment, focusing on the aspects and performance of the detector relevant to the study of prompt charged-hadron production in fixed-target collisions. Subsequently, Chapter 4 addresses the study of cold nuclear matter effects trough the measurement of charged light-hadron production in pNe and PbNe collisions at $\sqrt{s_{NN}} = 69$ GeV, covering all the steps from data selection to the calculation of particle ratios and comparison with EPOS simulation. The upgrade of the LHCb detector is introduced in Chapter 5, with a particular focus on the improvements made to the particle identification system. Finally, Chapter 6 details the performance of charged-particle identification in Run 3, describing the relevant algorithms and showcasing the results for pp, fixed-target, and heavy-ion collisions.

Chapter 2

Physics motivation

This chapter provides a concise overview of the theoretical foundation underlying the research activities documented in this thesis. The Standard Model of particle physics is introduced in Section 2.1. Section 2.2 delves into the fundamentals of Quantum Chromodynamics, the theory governing strong interactions. Section 2.3 offers an overview of various phenomenological approaches aimed at understanding inclusive charged-particle production in proton-ion collisions. Finally, Section 2.5 addresses the application of Monte Carlo generators to describe particle production in hadronic collisions.

2.1 The Standard Model

The Standard Model (SM) of particle physics[1, 2, 3], describes three out of the four fundamental forces as physical manifestations of local symmetries in a gauge theory, constructed as the product of three groups:

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \tag{2.1}$$

where the $SU(3)_C$ group is associated with the strong interaction, while $SU(2)_L \otimes U(1)_Y$ is responsible for the electroweak sector, i.e., the combined electromagnetic and weak interactions. The fourth fundamental force, gravity, is not described by the SM.

The SM defines three particle fields with the following characteristics:

- **Spin-1 particles**, which carry the fundamental interactions: the photon (γ) for electromagnetism, the $W^{\pm}$ and Z bosons for the weak force, and gluons for the strong force.
- **Spin-$\frac{1}{2}$ particles**, which are the fundamental constituents of matter. These are divided into leptons, which interact through gravitational, weak, and electromagnetic forces (when charged), and quarks, which are also subject to the strong force.
- **Spin-0 particle**, the Higgs boson, produced by an excitation of the Higgs field, which was recently discovered at CERN [4, 5]. The coupling of quarks, leptons, and spin-1 bosons with the Higgs field is proportional to their mass.

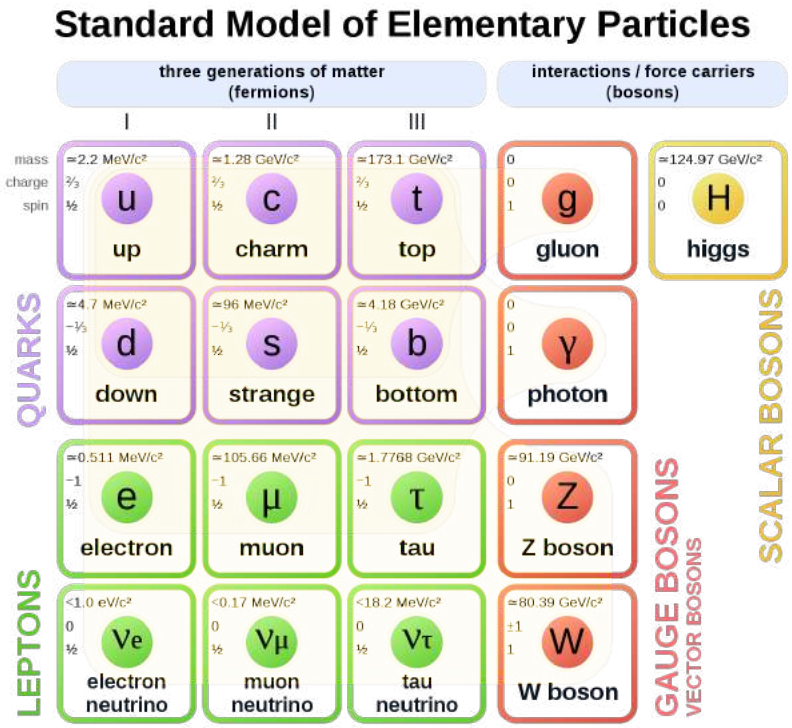


Figure 2.1: Diagram illustrating the fundamental particles as defined by the Standard Model of Particle Physics. Each particle is labelled with its corresponding mass, electric charge, and spin.

A scheme of all the particles described by the SM is presented in Figure 2.1. For each particle, a respective anti-particle exists with all additive quantum numbers, such as the electric charge, of opposite sign.

Leptons are organised into three families, each comprising a negatively-charged and a neutral particle, distinguished by a quantum number known as lepton flavour. The lightest charged lepton is the stable electron (e^-). The other two, the muon (μ^-) and tauon (τ^-), share the same electric charge as the electron but have higher masses and decay into lighter states. For each lepton, there exists a corresponding neutrino: ν_e, ν_μ, ν_τ.

Quarks are also organised into three families based on their hadronic flavour. The up (u) and down (d) quarks of the first family determine the quantum numbers of protons (with a uud combination) and neutrons (with a udd combination), constituting ordinary matter. Quark-antiquark pairs of all families are virtually formed in nucleons through gluon splitting. The two other families include four flavours: strange (s), charm (c), bottom (b), and top (t), arranged in increasing mass order.

Composite particles fall into two main categories:

- **Mesons**: composed of a quark-antiquark pair, all of which are unstable;
- **Baryons**: composed of three (anti)quarks. Among these, the proton and its antiparticle have never been observed to decay.

These categories are distinguished by baryon number, an additive quantum number defined

as $B = \frac{1}{3}(n_q - n_{\bar{q}})$, where n_q ($n_{\bar{q}}$) is the number of constituent (anti)quarks. Mesons have a zero baryon number, while baryons and anti-baryons have $B = \pm 1$. Recently, states predicted by theory that involve the combination of two quarks and two antiquarks (tetraquarks) or a quark-antiquark pair and three (anti)quarks (pentaquarks) have been observed, particularly by the LHCb collaboration [6] at CERN.

2.2 Quantum Chromodynamics

Quantum ChromoDynamics (QCD) [7] is the theory that describes the strong interactions among hadrons. Similar to Quantum Electrodynamics (QED), QCD is a gauge field theory, based on the assumption of a symmetry that leaves the theory unchanged. In the case of QCD, this symmetry is postulated to be exact and is described by an $SU(3)_C$ group.

The properties of the strong interaction are of particular interest to contextualise the present work, as this force governs the behaviour of hadrons. The Lagrangian of QCD is given by:

$$\mathscr{L} = -\frac{1}{4}\sum_{A=a}^{8} F^{A\mu\nu}F^{A}_{\mu\nu} + \sum_{j=1}^{n_f} \bar{q}_j \left(i \not{D} - m_j\right) q_j \ , \tag{2.2}$$

where q_j represents the quark fields (of n_f different flavours) with mass m_j and $\not{D} = D_\mu \gamma^\mu$, where γ^μ are the Dirac γ-matrices and D_μ is the covariant derivative:

$$D_\mu = \partial_\mu - i e_s \sum_A t^A g^A_\mu \tag{2.3}$$

where e_s is the gauge coupling, g^A_μ corresponds to the gluon fields (with A running from 1 to $N_c^2 - 1 = 8$, i.e., eight kinds of gluons), and t^A matrices correspond to eight 3×3 matrices that are the generators of the SU(3) colour group. These generators fulfill the relation $\left[t^A, t^B\right] = i C_{ABC} t^C$, where C_{ABC} are the complete asymmetric structure constants of SU(3). In the first term of Equation 2.2, $F^A_{\mu\nu}$ is the field tensor, describing the dynamics of the gluon field, and is given by

$$F^A_{\mu\nu} = \partial_\mu g^A_\nu - \partial_\nu g^A_\mu - e_s C_{ABC} g^B_\mu g^C_\nu . \tag{2.4}$$

The terms of the QCD Lagrangian in Equation 2.2 describe the potential couplings among the coloured particles. The possible vertices in QCD include: gluon-quark, 3-gluon and 4-gluon vertices. The 3-gluon and 4-gluon vertices are not present in other sectors of the SM like for QED. This distinction accounts for many differences between strong and electromagnetic interactions.

The quantity e_s (or $\alpha_s = e_s^2/(4\pi)$) represents the QCD coupling constant, and is the only fundamental parameter in QCD alongside quark masses. Despite being termed the "QCD coupling constant," α_s is not constant for every vertex but depends on the transferred four-momentum squared, Q^2, between the interacting particles. Figure 2.2 illustrates the dependence of α_s on Q, as measured in various experiments within the range $1 < Q < 10^3$ GeV/c [8]. The behaviour of α_s governs two fundamental phenomena in QCD: asymptotic freedom and confinement. As Q becomes large, α_s decreases, resulting in a weaker strong interaction. In the limit $Q \to \infty$, quarks

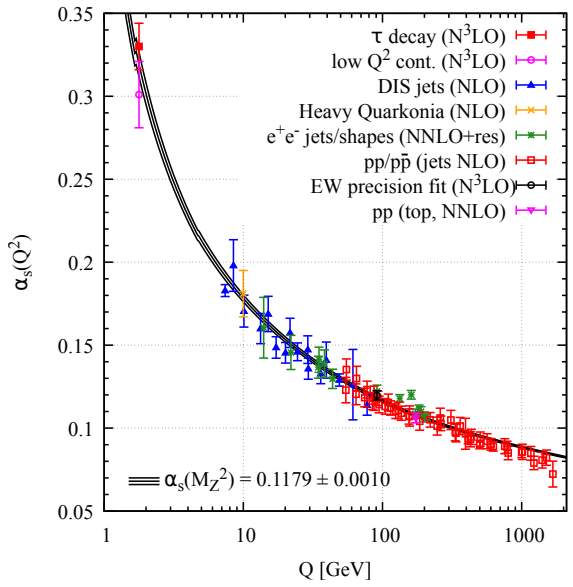


Figure 2.2: Dependence of α_s on the energy scale Q from various measurements [8]. The degree of perturbation theory used to extract α_s is indicated in brackets (NLO: next-to-leading order; NNLO: next-to-next-to-leading order; NNLO + res.: NNLO matched to a resummed calculation; N3LO: next-to-NNLO).

and gluons experience no strong interaction. This characteristic, known as asymptotic freedom, is a common feature in gauge theories based on a non-commuting group of symmetry. Since the coupling asymptotically decreases, QCD-related quantities can be computed using perturbation theory in this regime.

Conversely, for low Q, α_s grows rapidly. At values of Q on the order of $Q \lesssim \Lambda_{\mathrm{QCD}}$, where Λ_{QCD} is an energy scale on the order of a few hundred MeV, the interaction becomes large, and perturbation theory becomes inapplicable. In this regime, the QCD interaction potential between colour charges increases linearly with the distance between the charges. This non-perturbative behaviour is addressed in lattice QCD, which reformulates QCD on a discrete spacetime [9].

The main consequence is that particles with colour charges will appear in the physical spectrum as composite objects which are colour-neutral, known as hadrons. The simplest type of hadron is a meson, composed of a colour neutral $q\bar{q}$ pair. More complex examples of hadrons are baryons, consisting of qqq or $\bar{q}\bar{q}\bar{q}$ states with compensating colour charges. Mesons and baryons constitute the majority of the known hadron spectra. The hadron structure is highly complex, involving not only valence quarks but also a multitude of quark-antiquark pairs and gluons in constant interaction. These components play a crucial role, contributing significantly to the overall hadron mass. Consequently, when an elementary particle undergoes inelastic interaction with a hadron, it interacts with one of its elementary components. In this context, they are referred to as partons, carrying a portion of the total hadron momentum.

The associated quantum number, known as colour, can take one of three values conventionally denoted as red (r), green (g), and blue (b). Each quark carries a colour charge, while antiquarks possess anti-colours ($\bar{r}$, $\bar{g}$, and $\bar{b}$), and gluons carry a colour-anticolour charge pair.

Consequently, quarks in mesons exhibit the same colour-anticolour combination, while baryons (anti-baryons) can only be formed by quarks (antiquarks) with all different (anti-)colours. Due to colour confinement, only colour-neutral combinations (colour singlets) can be observed.

2.2.1 Perturbative QCD

One of the primary applications of QCD is the computation of various particle production cross-sections in particle colliders. The perturbative QCD (pQCD) approach [7] involves expanding the cross-section in powers of the constant α_s. This method is valid for hard processes where $Q^2 \gg \Lambda^2_{\text{QCD}}$, ensuring that $\alpha_s \ll 1$.

In the context of this thesis, one of the most relevant processes is hadronic collisions, with proton-proton interactions being among the simplest cases. However, these interactions are already challenging due to the composite nature of the proton arising from the property of confinement, as explained earlier. Inherently, the proton's structure involves non-perturbative processes that cannot be addressed using pQCD. Nevertheless, the collinear factorisation theorem [10] can be employed to compute production cross-sections of particles produced in hard processes where $Q^2 \gg \Lambda^2_{\text{QCD}}$. The cross-section for the process $h_1 + h_2 \to k + X$, at scales μ^2 and Q^2, is given by the expression:

$$\mathrm{d}\sigma^{h_1+h_2\to k+X}\left(\mu^2, Q^2\right) = \sum_{i,j,X'} f_{i/h_1}\left(x_1, Q^2\right) \otimes f_{i/h_2}\left(x_2, Q^2\right) \otimes \mathrm{d}\hat{\sigma}^{ij\to k+X'}\left(\mu^2, Q^2\right) , \tag{2.5}$$

where, k indicates a hard final-state parton, and $f_{i/h_1}\left(x_1, Q^2\right)$ and $f_{i/h_2}\left(x_2, Q^2\right)$ represent the parton distribution functions (PDFs). They describe the number density distribution of partons $i(j)$ in hadrons $h_1\,(h_2)$ at a momentum fraction $x_1\,(x_2)$ and a factorisation scale Q^2. The momentum fraction x of a parton within a proton is defined as the fraction of the proton momentum carried by the parton.

The term $\mathrm{d}\hat{\sigma}^{ij\to k+X'}\left(\mu^2, Q^2\right)$ corresponds to the production cross-section from the process where partons i and j from h_1 and h_2 interact, producing the elementary particle k and additional products X'. This includes interaction diagrams up to the targeted order of the prediction. The equation is summed over i, j and X', considering all combinations of partons and possible products. The dependence on μ is related to the normalisation scale. The differential cross-section in Equation (2.5) needs to be integrated over the values of x_1 and x_2 contributing to $\mathrm{d}\hat{\sigma}^{ij\to k+X'}\left(\mu^2, Q^2\right)$ for a given final state with particular kinematics.

The PDFs characterise the parton composition of the proton as a function of $\left(x, Q^2\right)$. These non-perturbative entities are universal and undergo scale Q^2 evolution, driven perturbatively by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [11, 12]. Presently, they cannot be computed directly from QCD first principles and are determined through global analyses of data, including deep inelastic lepton-nucleon scattering and related hard-scattering processes initiated by nucleons. For an up-to-date overview of PDFs, refer to Reference [13]. As an illustration, Figure 2.3 displays the parton distribution functions from the CT18 analysis [14] at $Q = 2$ GeV and $Q = 100$ GeV.

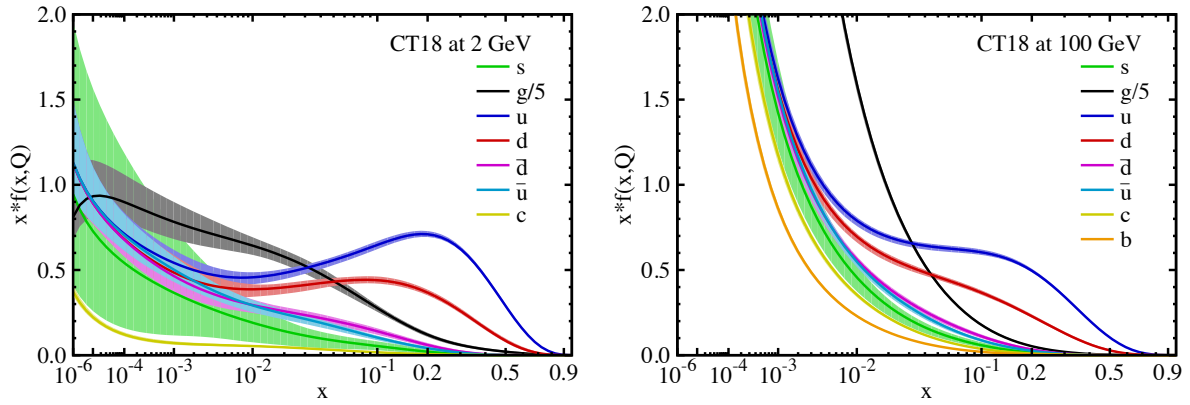


Figure 2.3: Parton distribution functions from the CT18 analysis [14] at $Q = 2$ GeV (left) and $Q = 100$ GeV (right), considering natural units, for $u, \bar{u}, d, \bar{d}, s = \bar{s}, b = \bar{b}$ and g. The gluon PDF is scaled as $g(x,Q)/5$ in all instances, and the charm distribution $c(x,Q)$ is perturbatively generated by evolving from $Q_0 = 1.3$ and 1.4 GeV. The band around each contribution represents the estimated uncertainty.

The focus of this thesis is on the study of inclusive hadron production. While Equation 2.5 represents the cross-section for an elementary particle, it is not directly applicable to hadron production due to the composite nature of hadrons. Nevertheless, new partons produced in the hard scattering process can undergo hadronization, resulting in colourless hadrons that are detectable. The probability for a produced parton to hadronize into a hadron is encapsulated in the parton-to-hadron fragmentation functions (FFs) [15]. Thus, the final inclusive hadron cross-section is obtained by convoluting the hard parton spectra from Equation 2.5 with the FFs:

$$\mathrm{d}\sigma^{h_1+h_2\to h+X}\left(\mu^2, Q^2, Q_F^2\right) = \sum_k \mathrm{d}\sigma^{h_1+h_2\to k+X}\left(\mu^2, Q^2, Q_F^2\right) \otimes D_{h/k}\left(z, Q_F^2\right) \ , \tag{2.6}$$

where, z describes the momentum fraction carried away by the hadron h from the parent parton k, $D_{h/k}\left(z, Q_F^2\right)$ is the parton-to-hadron fragmentation function, and Q_F^2 is the fragmentation scale. Similar to Equation 2.5, obtaining inclusive cross-sections for particles with specific kinematics requires integration over the contributing x_1 and x_2. Similar to PDFs, FFs are generally determined through global analyses of experimental data. For an updated review on FFs, refer to Reference [16].

2.2.2 The QCD phase diagram

The property of asymptotic freedom, as explained earlier, implies the existence of a new high-temperature phase of weakly interacting quarks and gluons, known as quark-gluon plasma (QGP) [17]. This state of hadronic matter arises when the temperature and the energy density are sufficiently high, allowing thermal interactions between quarks and gluons to occur at $Q^2 \gg {\Lambda_{\mathrm{QCD}}}^2$, resulting in the asymptotic freedom.

An illustration of the current understanding of the QCD phase diagram is presented in Figure 2.4. The various phases of strongly interacting matter are illustrated as a function of temper-

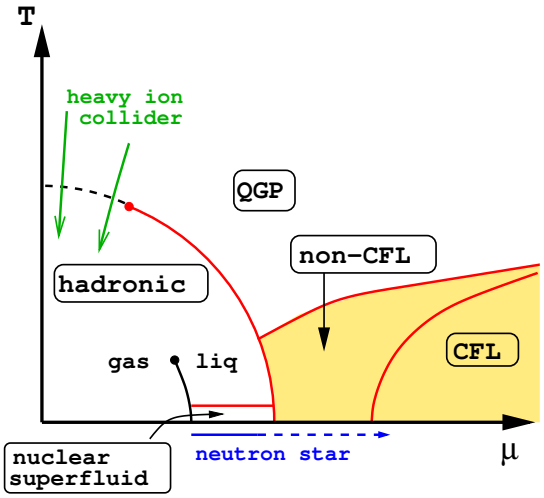


Figure 2.4: Schematic phase diagram of QCD as a function of temperature T and baryon chemical potential μ [17].

ature T and baryon chemical potential μ, describing the amount of energy that can be absorbed or released due to a change of baryon number through an imbalance between quarks and antiquarks. At zero temperature and chemical potential, quark interactions are dominated by large distances, and the coupling α_s is large. Consequently, quarks and gluons are confined within hadrons, with a mass of the order of $\Lambda_{\mathrm{QCD}} \sim 200$ MeV. For example, the proton has a mass of $m_p \approx 935$ MeV. At very high temperatures ($T \gg 170$ MeV), quarks and gluons possess thermal momentum $p \sim T \gg \Lambda_{\mathrm{QCD}}$, resulting in weak interactions among themselves, forming a plasma-like state of colour charges known as Quark Gluon Plasma (QGP). As colour confinement is absent in the QGP, a phase transition separates this deconfined state from the hadronic gas.

The QGP is presumed to have naturally formed in the early universe when it was a few microseconds old, and the temperature was above Λ_{QCD}, rendering it too hot for hadron formation. Additionally, the study of QGP and the phase transition can be conducted through ultrarelativistic collisions of heavy ions. Collisions of AuAu at the Relativistic Heavy Ion Collider (RHIC) reached a nucleon-nucleon center-of-mass energy of $\sqrt{s_{\mathrm{NN}}} = 500$ GeV in the late 2000s, while at the Large Hadron Collider (LHC), PbPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV were performed in the early 2010s.

2.3 Proton-nucleus collisions phenomenology

In high-energy collisions between nuclei, hard-scattered partons fragment into collimated jets of hadrons. In the process color-neutral particles are created with lower momenta than the original parton. The confinement of these particles within a specific angular region, termed a jet cone, is due to the scale dependence of the strong coupling, which suppresses large-angle radiation. As the energetic partons forming the jet traverse the nuclear medium, they lose energy predominantly through gluon radiation and partially through collisional energy loss.

This phenomenon of partonic energy dissipation is known as jet quenching [18, 19, 20].

The modification of parton showers within the medium leads to changes in the hadron spectra at high transverse momentum p_{T}. A quantitative method for studying this medium modification effects consists in extracting a ratio of p_{T} spectra measured in nucleus-nucleus (AA) and proton-proton (pp) collisions at equivalent centre-of-mass energy per nucleon ($\sqrt{s_{NN}}$). These ratios are denoted as nuclear modification factors. Due to the low likelihood of hard scattering events compared to soft processes, it is anticipated that a given nucleon experiences, at most, one hard collision in a nuclear collision event. Therefore, hard scattering processes at the nucleon-nucleon level generally scale with the number N_{coll} of binary nucleon-nucleon collisions.

Specifically, the nuclear modification factor is defined as:

$$R_{\mathrm{AA}}\left(p_{\mathrm{T}}, y\right) = \frac{\mathrm{d}^2 N_{\mathrm{AA}}/\mathrm{d}y\ \mathrm{d}p_{\mathrm{T}}}{\langle T_{\mathrm{AA}}\rangle\, \mathrm{d}^2\sigma_{\mathrm{pp}}^{\mathrm{INEL}}/\mathrm{d}y\ \mathrm{d}p_{\mathrm{T}}} = \frac{1}{\langle N_{\mathrm{coll}}\rangle}\frac{\mathrm{d}N_{\mathrm{AA}}\left(p_{\mathrm{T}}, y\right)}{\mathrm{d}N_{\mathrm{pp}}\left(p_{\mathrm{T}}, y\right)}\,, \tag{2.7}$$

where N_{AA} and $\sigma_{\mathrm{pp}}^{\mathrm{INEL}}$ represent the particle yield and the inelastic (INEL) cross section in AA and pp collisions, respectively. The nuclear overlap function $\langle T_{\mathrm{AA}}\rangle$ is determined from the Glauber model and it is related to the average number of binary nucleon-nucleon collisions and the inelastic nucleon-nucleon cross section through the relation: $\langle T_{\mathrm{AA}}\rangle = \langle N_{\mathrm{coll}}\rangle / \sigma_{\mathrm{INEL}}^{\mathrm{NN}}$ [21].

In the absence of nuclear effects, R_{AA} is, by definition, expected to be equal to unity for hard processes exhibiting binary collision scaling. Binary scaling can be broken due to initial-state effects in nuclei or final-state effects occurring in AA collisions. Any deviation from unity in R_{AA} indicates suppression or enhancement of hadron production in AA collisions compared to pp collisions.

At the LHC, jet quenching was observed experimentally in PbPb collisions first by the ATLAS experiment [22], followed by the experiments ALICE [23] and CMS [24]. In Figure 2.5, R_{AA} is shown for different collision centralities in PbPb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV measured by the ALICE collaboration [25, 26]. In peripheral collisions ($60-80\%$ centrality), a weaker suppression and a flattening behaviour of R_{AA} approaching unity is observed for unidentified charged particles (black markers). In contrast, for most central ($0-5\%$) collisions, strong suppression ($R_{\mathrm{AA}} \ll 1$) is observed, indicative of jet quenching. The measured suppression is larger, by about 40% at $p_{\mathrm{T}} = 10$ GeV/c, than that observed at RHIC [27] due to the higher energy density reached at the LHC.

In Figure 2.5 measurements for identified light charged hadrons are also shown. The trend of R_{AA} for identified particles provides more details about the in-medium interactions of partons fragmenting into hadrons. At high p_{T} (> 10 GeV/c), R_{AA} for all particle species is equally suppressed, suggesting that particle ratios resemble those of jets in vacuum, and the medium does not alter the hadrochemistry of the leading particle of the quenched jet. For $p_{\mathrm{T}} < 10$ GeV/c and all centralities, protons are less suppressed compared to pions and kaons, attributed to effects arising from collective radial flow.

However, these measurements alone do not allow us to determine whether the observed

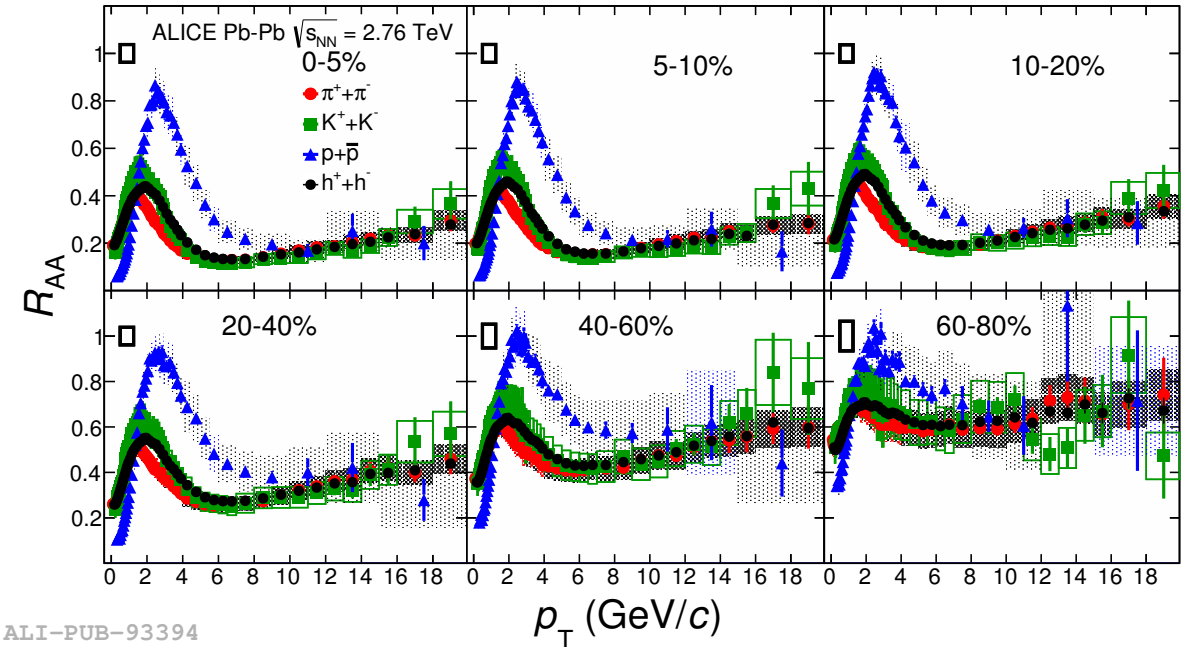


Figure 2.5: The nuclear modification factor R_{AA} as a function of transverse momentum for different particle species and collision centralities in PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV measured by the ALICE collaboration [26].

energy loss is an initial-state or final-state effect. To distinguish between the two, nuclear modification factors must be measured in pPb collisions as well. The corresponding modification factor for inclusive charged particles measured in pPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV by ALICE [25, 26] is shown in Figure 2.6. While initial-state effects are significant at low and intermediate p_{T}, it is established that for $p_{\mathrm{T}} \gtrsim 10$ GeV/c, R_{pPb} is consistent with unity. Thus, the observed suppression in PbPb collisions is attributable to final-state effects such as jet quenching.

2.4 Cold Nuclear Matter Effects

Particle collisions provide an intuitive method to explore the structure of matter and the fundamental interactions among its constituents. Experimental observations indicate a gradual transition between "small" collision systems (proton-proton, pp, and proton-nucleus, pA) and "large" collision systems (nucleus-nucleus, AA). The investigation of pA collisions at the LHC serves as a baseline for AA collisions, opening doors to new physics opportunities [29], such as exploring the partonic structure of heavy nuclei. Modifications of the particle production rate in pA collisions relative to binary-scaled proton-proton collisions are commonly referred to as Cold Nuclear Matter (CNM) effects. Understanding these effects is of utmost importance for a correct interpretation of ultra-relativistic heavy-ion collisions, where genuine QGP signals may be masked by concurring CNM effects.

The subsequent discussion focuses on inclusive prompt charged-particle production, often termed inclusive charged-hadron production, given that hadrons ($\pi^{\pm}, K^{\pm}, {}^{(-)}p$) constitute the

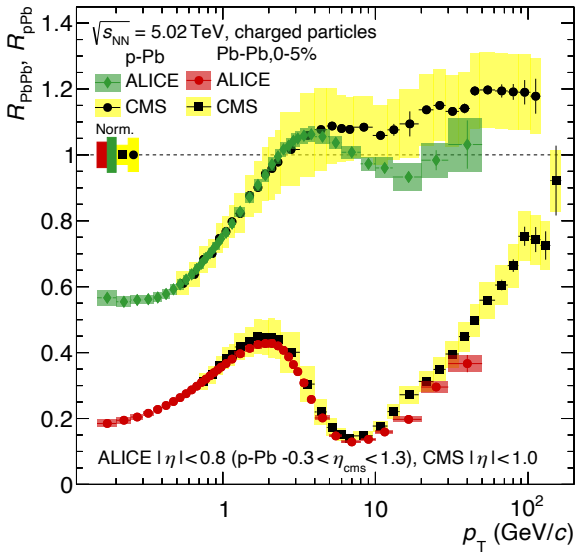


Figure 2.6: A comparison of the nuclear modification factors for central (0-5%) PbPb and pPb collisions measured by the ALICE and the CMS collaborations [28].

majority of final-state particles in a hadronic collision. Since CNM effects originate from modifications of parton distributions within nuclei, they may vary with respect to (x, Q^2), similar to the parton distribution functions of the proton (see Figure 2.3). Therefore, it is crucial to investigate these effects across the (x, Q^2) range. As these phenomena are essentially non-perturbative, experimental input is indispensable. Hence, the coverage of different LHC and RHIC experiments in (x, Q^2) needs careful examination.

However, relating final-state particle kinematics, usually given in terms of variables like pseudorapidity (η) and transverse momentum (p_T), to the (x, Q^2) of the partons within the nuclei undergoing hard scattering is not straightforward. An approach involves considering LO $2 \rightarrow 2$ kinematics [30]. In this context, a way to approximate x in the nuclei is given by

$$x \approx \frac{m_T}{\sqrt{s_{NN}}} e^{-\eta}, \tag{2.8}$$

where $m_T = \sqrt{m^2 + p_T^2}$ is the transverse mass of the produced hadron. Note that for asymmetric pA collisions, positive η is taken with respect to the direction of the proton beam. Equation 2.8 suggests that lower values of x can be probed with higher $\sqrt{s_{NN}}$ and more forward η. Figure 2.7 illustrates the coverage of major heavy-ion experiments at LHC and HERA. The figure also highlights the coverage of the LHCb experiment in fixed-target mode, capable of probing high x-Bjorken and the backward-central pseudorapidity region in the center-of-mass frame ($-2.8 < \eta^* < 0.2$). The values of x are obtained using Equation 2.8, and Q^2 is approximated as the transverse mass m_T. An average hadron mass of $m = 255$ MeV/c^2 has been considered based on the proportion of π, K and p in the EPOS generator at low p_T.

At the LHC, the ALICE [28, 31], CMS [32, 33], and ATLAS [34] collaborations have measured prompt charged-particle production at $\sqrt{s_{NN}} = 5.02$ TeV in the central region. At RHIC,

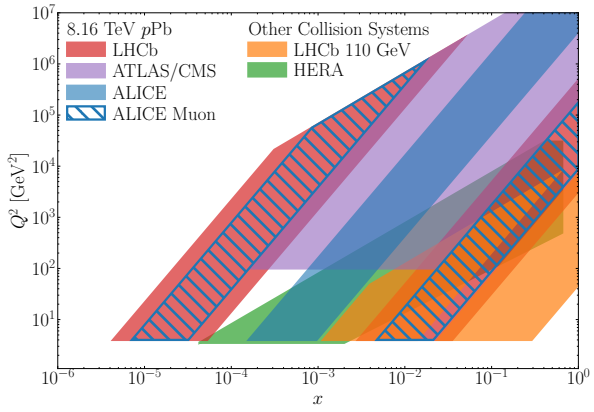


Figure 2.7: Accessible range in the (x, Q^2) plane of the LHCb fixed-target configuration (orange area) compared to other existing experiments.

measurements at more forward pseudorapidities have been performed by the BRAHMS [35] and PHENIX [36] collaborations with dAu collisions at $\sqrt{s_{\mathrm{NN}}} = 200$ GeV. However, for RHIC measurements at $\sqrt{s_{\mathrm{NN}}} = 200$ GeV, the reached x is not as low as for LHC measurements. In the following sections, various theoretical approaches describing charged-particle production in pA collisions will be presented.

2.4.1 Nuclear parton distribution functions

Perturbative calculations of cross-sections in QCD are typically factorised into parton density probability functions and a hard scattering cross-section. These parton density probability functions are expressed in terms of Parton Distribution Functions (PDFs), which depend on the longitudinal momentum fraction (x) [37] and the energy scale (Q^2) of quarks and gluons within a free proton.

Phenomenological parametrisations are employed to derive PDFs based on data from Deep Inelastic Scattering (DIS) experiments and Drell-Yan (DY) production studies. Quark and anti-quark distributions are directly probed by DIS and DY, while gluon distributions are indirectly probed. Consequently, gluon distributions are less constrained compared to quark distributions. DIS experiments with nuclear targets have confirmed modifications to the PDFs of nucleons bound in nuclei, as described by nuclear parton distribution functions (nPDFs) [38]. A ratio is commonly used to quantify these effects:

$$R_i^A\left(x, Q^2\right) = \frac{f_i^A\left(x, Q^2\right)}{f_i\left(x, Q^2\right)}\,, \tag{2.9}$$

where $f_i^A\left(x, Q^2\right)$ represents the PDF for a parton with flavour i of a nucleon bound in a nucleus A, and $f_i\left(x, Q^2\right)$ is the ordinary pdf of a free nucleon.

To illustrate the structure of the initial parametrisation, Figure 2.8 depicts the fit function employed in the EPPS16 analysis [39]. This function reveals four primary regions characterising the behaviour of $R_i^A(x, Q^2)$ [30], corresponding to:

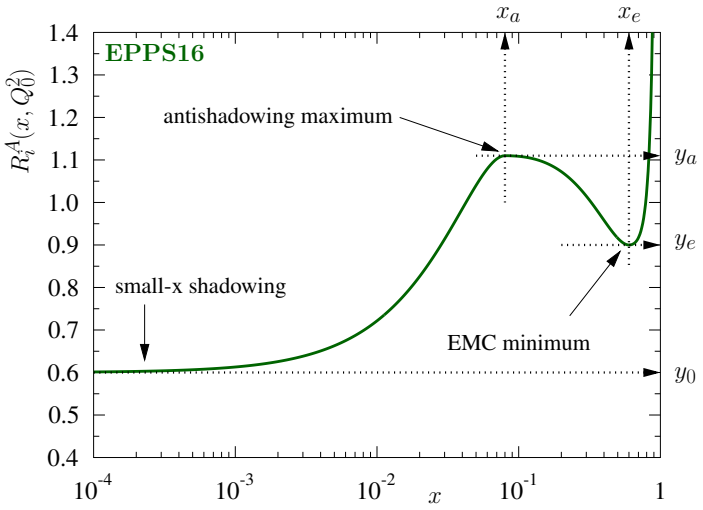


Figure 2.8: Fit function of nPDF set EPPS16 at the parametrization scale Q_0^2 [39].

- $R_i^A(x, Q^2) > 1$ for $x > 0.8$: the Fermi motion region;
- $R_i^A(x, Q^2) < 1$ for $0.3 < x < 0.8$: the EMC region;
- $R_i^A(x, Q^2) > 1$ for $0.04 < x < 0.3$: the antishadowing region;
- $R_i^A(x, Q^2) < 1$ for $x < 0.1$: the shadowing region.

For proton PDFs, a large number of experimental data, particularly from HERA and the Tevatron, are available to constrain the global fits performed at Leading Order (LO), Next-to-Leading Order (NLO), or Next-to-Next-to-Leading Order (NNLO) calculations. Various sets of free proton PDFs are available, including the widely used general-purpose CT14 [40], which incorporates data from the LHC, as well as updated information from the Tevatron and HERA experiments.

In contrast, the availability of experimental data for nPDFs in the perturbative regime ($Q^2 \geq 1$ GeV^2), particularly for the region where $x \leq 0.01$, is considerably limited. This lack of data leads to significant uncertainties in nPDFs relevant for LHC kinematics. Some of the commonly used sets of global fits for nPDFs at NLO include EPS09 [41], HKN07 [42], nDS [43], and DSSZ [44]. These sets are calibrated with data from charged-lepton DIS with fixed nuclear targets, DY processes in proton-nucleus collisions, and also include hadronic final-state results from pA (dA) collisions. Nuclear PDFs are often presented through their nuclear modification $R_i^A(x, Q^2)$, as defined in Equation 2.9, to emphasise the effect from the nuclei. Each parton flavour i is represented by a parameterised function, with some parameters fixed and others determined based on certain assumptions. QCD global analyses of experimental data allow for the parametrisation of nuclear PDFs and their dependence on Q^2 values and atomic mass number. For instance, the widely used EPS09 parametrisation of PDFs is derived from an NLO pQCD analysis incorporating three distinct experimental inputs: charged-lepton DIS off nuclei, dilepton production in the DY process, and inclusive pion production in dAu collisions observed at RHIC. The resulting nuclear modifications for Pb at the initial scale $Q^2 = 1.69$ GeV^2 are shown

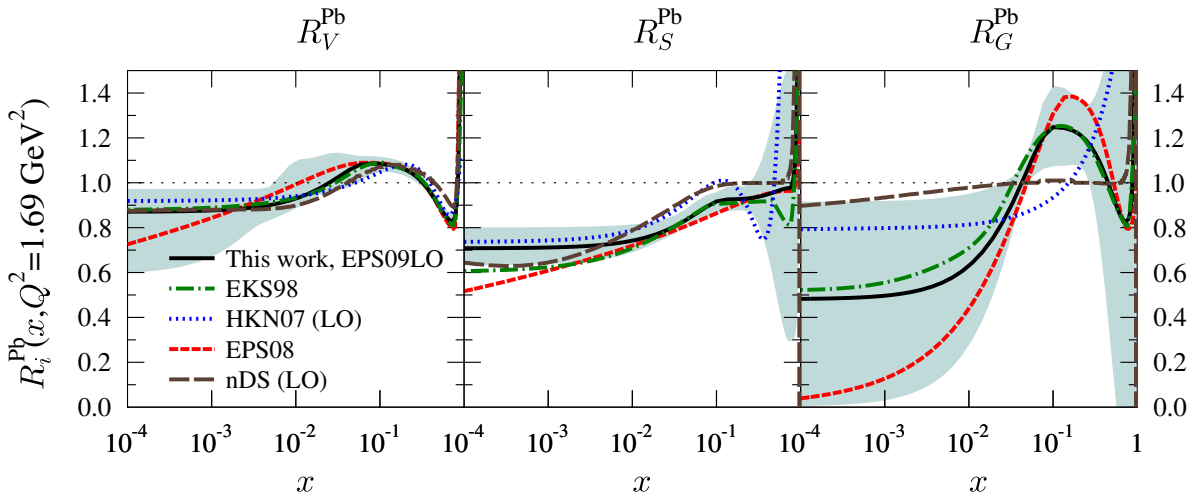


Figure 2.9: Comparative analysis of the average modifications for valence and sea quarks, and for gluons, at $Q^2 = 1.69$ GeV^2 for Pb nucleus, as derived from LO global DGLAP analyses including EKS98 [45, 46], EKPS [47], nDS [43], HKN07 [48], and EPS09LO [41].

in Figure 2.9, alongside parametrisations from other analyses [41]. Nuclear modifications R_v^{Pb}, R_s^{Pb}, and R_g^{Pb} correspond to valence quarks, sea quarks and antiquarks, and gluons, respectively.

2.4.2 Saturation and Colour Glass Condensate

At low values of x, the gluon density experiences a substantial increase, as illustrated in Figure 2.3. The rapid growth of the gluon distribution at a fixed Q^2, with $x \to 0$, is described by the BFKL equation [49, 50, 51]. The initial observation of this behaviour in the proton PDF was made by HERA [52]. The stability of the theory dictates that gluons reach a maximum occupation number at a specific saturation momentum $Q_s(x)$ [53, 54]. The saturation phenomenon is depicted in Figure 2.10. For nuclei, saturation occurs at higher values of x due to the Lorentz contraction of the nuclear parton density in the probe rest frame.

In this regime, QCD dynamics become non-linear and non-perturbative, and the Colour Glass Condensate (CGC) effective field theory [55, 56] is utilised. The CGC theory is applicable to pA collisions, predicting particle production cross-sections for kinematics dominated by the low x contribution. Generally, models based on this approach anticipate a suppression of charged-particle production in pA collisions compared to pp, indicating strong shadowing at very low x. For inclusive charged-particle production at $\sqrt{s_{NN}} = 5$ TeV, several predictions within the CGC framework have been proposed [57, 58, 59].

2.4.3 The Cronin Effect

In 1975, an increase in the production of inclusive hadrons with high transverse momentum ($p_T \gtrsim 1.5$ GeV/c) was observed in pA collisions compared to scaled pp collisions [61]. Subsequent similar enhancements were observed at central pseudorapidity at RHIC [62, 63]. As an example, the top panel of Figure 2.11 illustrates R_{dA} for inclusive charged particles, alongside

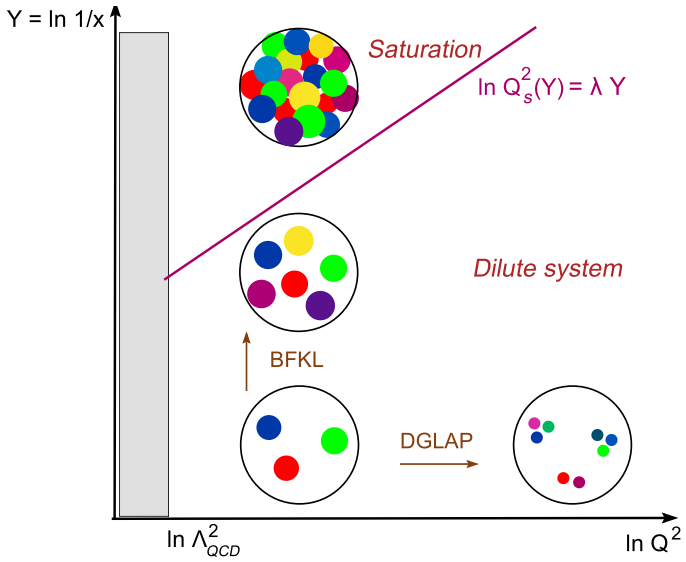


Figure 2.10: Evolution of the partonic density is depicted for DGLAP (with fixed x as $Q^2 \to \infty$) and BFKL (with fixed x as $Q^2 \to \infty$), illustrating the eventual saturation of the gluon density at $Q_s(x)$ [60].

R_{AA} observed in central AuAu collisions, while the lower panel compares with π^0. The data suggest an increase in inclusive charged particle production at $p_T > 2$ GeV/c, in dAu collisions. This enhancement is commonly referred to as the "Cronin effect". The π^0 data imply a smaller enhancement for pions compared to inclusive charged particles at $p_T = 2-4$ GeV/c.

Figure 2.12 presents the nuclear modification factor in pPb collisions compared to both central (0–5% centrality) and peripheral (70–80% centrality) PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV [31]. R_{pPb} remains close to unity for $p_T > 2$ GeV/c, suggesting that the significant suppression observed in central PbPb collisions at the LHC does not arise from an initial-state effect but rather reflects characteristics of the hot matter generated in collisions of heavy ions, as explained in Section 2.3. In dAu collisions at $\sqrt{s_{NN}} = 200$ GeV, $R_{d\text{Au}}$ peaked around 1.4 for charged hadrons with p_T between 3 and 5 GeV/c. Recent measurements indicate a diminished magnitude of the Cronin effect at the LHC [31, 34]. Typically, this enhancement is notable for $p_T \gtrsim 1.5$ GeV/c and diminishes with increasing p_T. Interestingly, in charged-hadron production, protons exhibit a more pronounced enhancement compared to pions and kaons [64, 65].

Explanations for this phenomenon often invoke multiple interactions in the large x regime, where the object experiencing the re-scattering can be either the incoming or outgoing parton [66, 67]. The contribution to the charged-particle cross-section from multiple scatterings can be computed using pQCD techniques, incorporating additional partonic processes. Recent calculations [68] shed light on the enhancement observed for muons from heavy-flavour hadron decays at both the LHC and RHIC [69, 70], as well as for inclusive hadron production at RHIC [36]. However, a definitive and clear explanation for the Cronin effect remains elusive. Alternative approaches to clarify the Cronin effect are based on final-state recombination of soft and shower partons [71], demonstrating the ability to reproduce RHIC data.

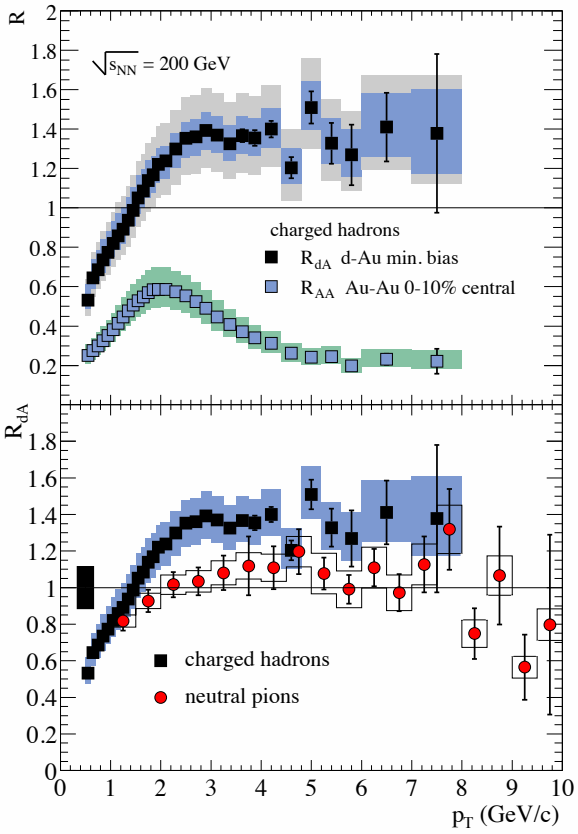


Figure 2.11: Top: Nuclear modification factor R_{dA} for $\left(h^{+}+h^{-}\right)/2$ in minimum bias dAu compared to R_{AA} in the 10% most central AuAu collisions. Inner bands show systematic errors which can vary with p_T, and outer bands include also the normalisation uncertainty. Bottom: Comparison of R_{dA} for $\left(h^{+}+h^{-}\right)/2$ and the average of the π^0 measurements in dAu. The bar at the left indicates the systematic uncertainty in common for the charged and π^0 measurements [62].

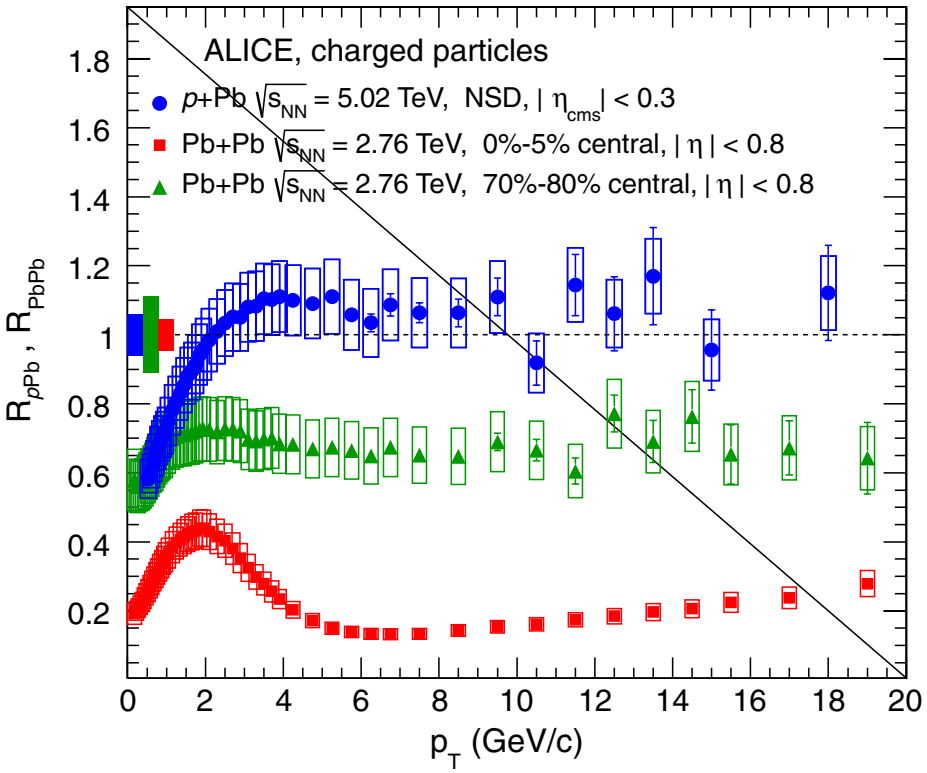


Figure 2.12: The nuclear modification factor of charged particles as a function of transverse momentum in minimum bias (NSD) pPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. The data for $|\eta_{\text{c.m.s.}}| < 0.3$ are compared to measurements in central $(0-5\%$ centrality) and peripheral $(70-80\%)$ PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV [31].

2.5 Monte Carlo Generators

Although QCD provides a comprehensive description of strong interactions, its predictive power diminishes in the nonperturbative regime at low Q^2. Many processes in hadronic collisions are inherently soft and, thus, non-perturbative. Even in the case of hard interactions, they are intricately connected to other non-perturbative processes, such as hadronisation. Despite these challenges, obtaining accurate descriptions of the final state of hadron collisions is crucial for studies at the LHC. A common solution lies in the use of MC generators, based on models that integrate the probabilities of various phenomena to reproduce the final state of hadronic collisions [72]. Constructing a MC generator involves defining the probabilities of processes, with hard processes computed accurately from first principles, while soft processes rely on models.

The use of MC generators is pivotal for the physics program development in hadron colliders like the LHC. Monte Carlo predictions simulate the signatures of processes of interest, representing known SM processes or New Physics signatures, along with signal-like signatures originating from other SM processes that may interfere with the observation or background. Moreover, MC generators play a crucial role beyond collider physics; they are widely employed in cosmic ray physics to model the interaction of high-energy cosmic rays with the atmosphere, producing extensive air showers [73]. These showers are detected by experiments like the Pierre Auger Observatory [74], providing a means to measure cosmic rays with energies exceeding 10^{15} eV. Currently, the dominant source of systematic uncertainty in interpreting these measurements is associated with the reliability of simulations, especially those for hadronic interactions [75, 76]. Furthermore, inconsistencies between measurements and simulations in the number of muons produced in such cascades, especially at forward rapidities in LHC energies, highlight the need for improved simulation accuracy [77].

A crucial aspect in developing accurate MC generators is tuning. As mentioned earlier, a significant fraction of processes in an event cannot be computed from first principles and must be modelled with phenomenological considerations. These models typically involve several parameters adjusted through comparisons with experimental data. The validation of these models must be global to ensure accurate descriptions of underlying physics mechanisms, moving beyond mere data parametrisation. Since generators encompass various processes, the number of parameters is usually on the order of 15 or more, grouped based on the involved processes and tuned in different stages. The typical order is: first, tuning for hadronisation and final-state fragmentation processes; second, tuning for initial-state parton showers; and third, tuning for multiple parton scattering and beam remnant effects.

Currently, numerous MC generators can replicate pp, pA, and AA collisions. The EPOS model is integrated into the LHCb simulation framework and is utilized in the analysis presented in this thesis. EPOS-LHC [78] is an event generator for pp, pA, and AA collisions, designed to reproduce minimum-bias data. The generator is based on the EPOS 1.99 model [79] which is tuned to reproduce any kind of hadronic interactions from hA to AB where h can be $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ and A or B range from 1 to 210 nucleons. The energy range is from 40 GeV in the laboratory frame to more than 1000 TeV centre-of-mass energy [80].

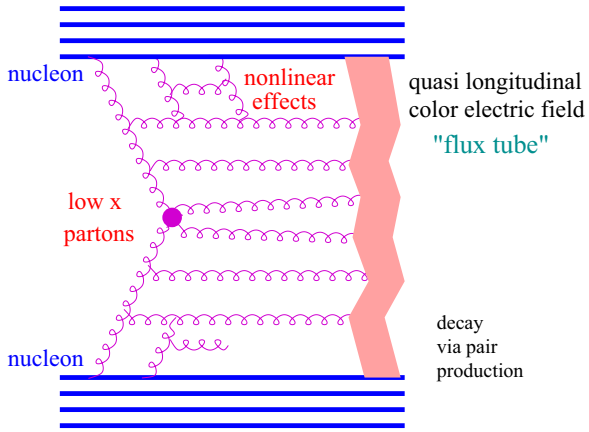


Figure 2.13: Schematic view of an elementary interaction in the EPOS model [80].

2.5.1 Basic principles of EPOS LHC and comparison with LHC data

Nucleus-nucleus scattering, and even proton-proton collisions, involve multiple elementary collisions occurring simultaneously. Each of these elementary scatterings is known as "parton ladder," as depicted in Figure 2.13, also known as a cut Pomeron [81]. A parton ladder represents the evolution of partons from both the projectile and target sides toward the center (at small x). This evolution is governed by an evolution equation, typically the DGLAP equation. It has long been recognised that such a parton ladder can be conceptualised as a quasi-longitudinal colour field, often referred to as a "flux tube" [82], which can conveniently be treated as a relativistic string. This flux tube ultimately decays through the production of quark-antiquark pairs, giving rise to fragments that are identified with hadrons.

It is essential to understand that in EPOS, the initial conditions for hadronisation are based on strings, not on partons. The initial scatterings give rise to string formation, which subsequently fragments into segments, typically identified as hadrons. At an early proper time τ_0, well before hadron formation, the system distinguishes between string segments in dense areas and those in low-density regions. The denser regions are termed "core," while the lower density areas are referred to as "corona" [83]. The corona plays a significant role in various aspects, such as the centrality dependence of all observables in heavy-ions collisions. In each event, a fraction of the string segments undergoes normal hadronisation (forming the corona), while another fraction contributes to the creation of a core through collective hadronisation, as shown in Figure 2.14. The formation of the core takes place with sufficiently high local density of string segments. This condition is typically met in central heavy-ion collisions at RHIC or the LHC, and even at SPS, because of the large number of pairs of nucleons suffering an inelastic interaction.

Several comparisons with EPOS LHC prediction and experimental data have been conducted. In particular, a direct comparison of EPOS LHC simulations with the transverse momentum distribution measured by the ATLAS experiment [84] in minimum bias pp interactions, as presented in Figure 2.15, shows that particles originating from core hadronisation with radial flow dominate the flux around 1–2 GeV/c. This corresponds precisely to the range where a deficit

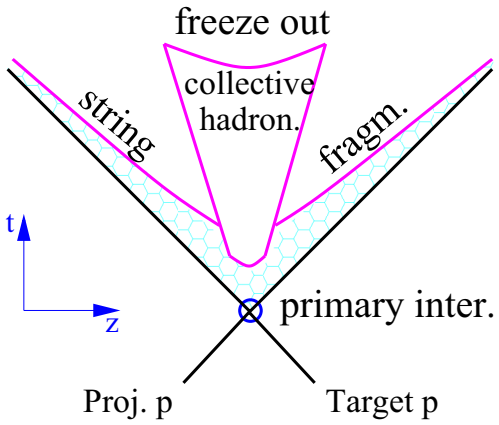


Figure 2.14: Schematic view of the space-time evolution of the particle production in a hadronic interaction in EPOS 1.99. A hyperbola (line) represents particles with the same proper time. The same treatment is used for pp or AB but the collective hadronisation is simplified compared to the full heavy-ion picture [80].

is observed when comparing a model without flow to measured data. The shape of the p_T distribution depends on the parameter p_{cut}, with its optimal value being 1 GeV/c. At higher p_T (>5 GeV/c), particles that are not entirely absorbed into the high-density region post-string fragmentation dominate once again.

This effect is already visible in the minimum bias transverse momentum distribution of identified particles, as illustrated in Figure 2.16 (a). Here, simulations with (solid line) and without (dash-dotted line) core formation can both account for the pion p_T spectrum from CMS data [85]. However, as the hadron mass increases, the deviation between standard hadronisation without flow and the observed data becomes more pronounced, while simulations incorporating collective hadronization yield to better results. Finally, the ratios of kaons to pions and protons to pions as a function of p_T, as shown in Figure 2.16 (b), clearly indicates that the flow effect manifests only above $p_T > 1$ GeV/c in both experimental data and simulations with (solid line) and without (dash-dotted line) core formation.

Core formation, indicating a transverse flow, proves crucial for accurately describing minimum bias pp data. These effects cannot be neglected, especially in pA scattering, where final-state interactions play an even more significant role than in pp interactions. Consequently, any analysis involving particles with $p_T < 5$ GeV/c should be interpreted cautiously, as even pp data may include final-state interactions.

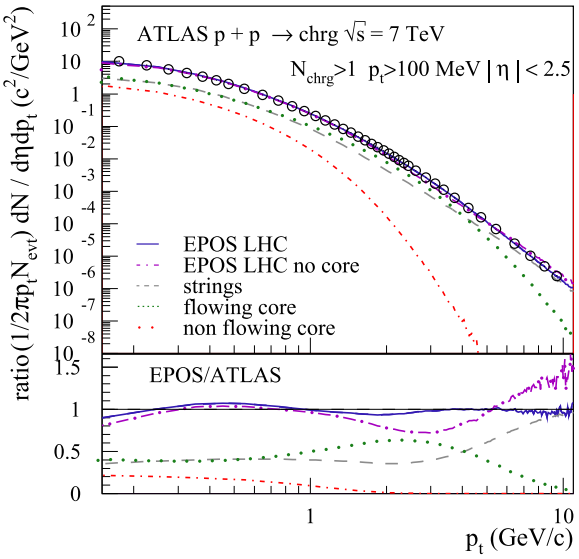
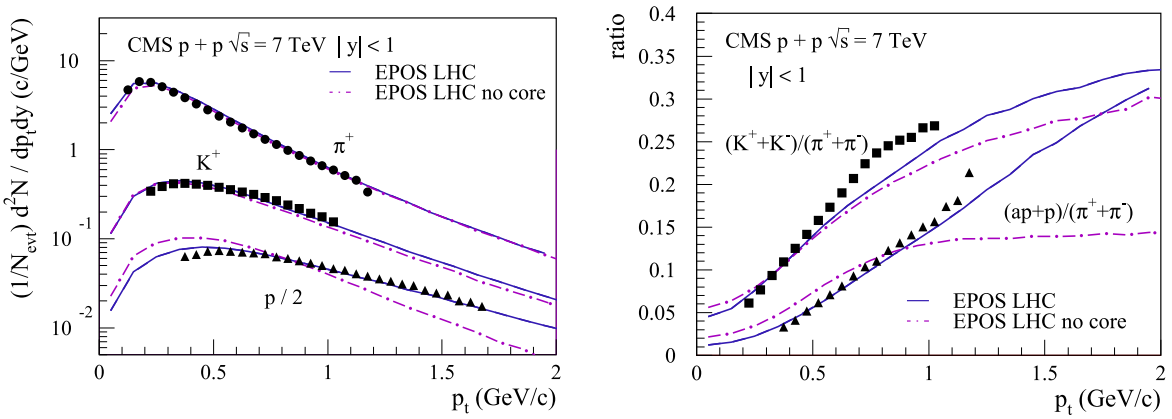


Figure 2.15: Transverse momentum distribution of charged particles with $|\eta| < 2.5$ produced in minimum bias pp collisions at 7 TeV. Simulations are done with EPOS LHC with (solid line) and without (dash-dotted line) core formation. The contributions of particles coming from the core hadronization are shown as a dotted line while particles coming directly from string fragmentation are represented by a dashed line. Points are data from the ATLAS experiment [84].



(a) (b)

Figure 2.16: (a) Transverse momentum distribution of identified particles and (b) ratio of particle yield as a function of transverse momentum for $|y| < 1$ for non-single-diffractive pp scattering at 7 TeV. Simulations are done with EPOS LHC with (solid line) and without (dash-dotted line) core formation [80]. Points are data from the CMS experiment [85].

Chapter 3

The LHCb experiment at CERN

In this chapter, the experimental framework of the research activities discussed in this thesis is introduced. Section 3.1 provides an overview of the CERN complex, with a specific focus on the LHC accelerator. Following this, Section 3.2 introduces and details the detectors comprising the LHCb experiment, along with their respective performances. Lastly, Section 3.3 delves into the discussion of the LHCb fixed-target program during Run 2.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [86] is a circular superconducting particle accelerator designed for accelerating protons and heavy ions and represents a monumental achievement in the field of particle physics. Situated beneath the France-Switzerland border near Geneva, the LHC is located in a 26.7 km long tunnel, 3.8 m wide, at depths ranging from 50 to 175 m. It was designed and constructed by the Conseil Européen pour la Recherche Nucléaire (CERN) to perform hadronic collisions at unprecedented high energies.

The LHC operates by accelerating particles to almost the speed of light. A schematic representation of the CERN accelerator complex is shown in Figure 3.1. This complex involves several key components, including multiple storage rings and smaller accelerators, superconducting magnets to guide and focus the particle beams, and advanced beam diagnostic systems. Within the tunnel, a total of 1232 superconducting dipole magnets are positioned to maintain the particle beams in a circular trajectory. Additionally, 392 quadrupole magnets are used to ensure that the beams remain properly focused. The power required to accelerate the beams is delivered by superconductive radiofrequency (RF) cavities. A total of 8 cavities per beam, each delivering 2 MV at 400 MHz, are used in the LHC.

In order to generate the intense proton and lead (Pb) beams essential for experiments, several steps are involved. Protons are first stripped from hydrogen molecules using an electric field, while lead atoms are obtained by heating 2 cm of pure lead with an electric current. Particles are initially accelerated by linear machines, known as LINACs, which feed the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and Super Proton Synchrotron (SPS), where they are accellerated up to energies of 2 GeV, 26 GeV and 450 GeV, respectively. Subse-

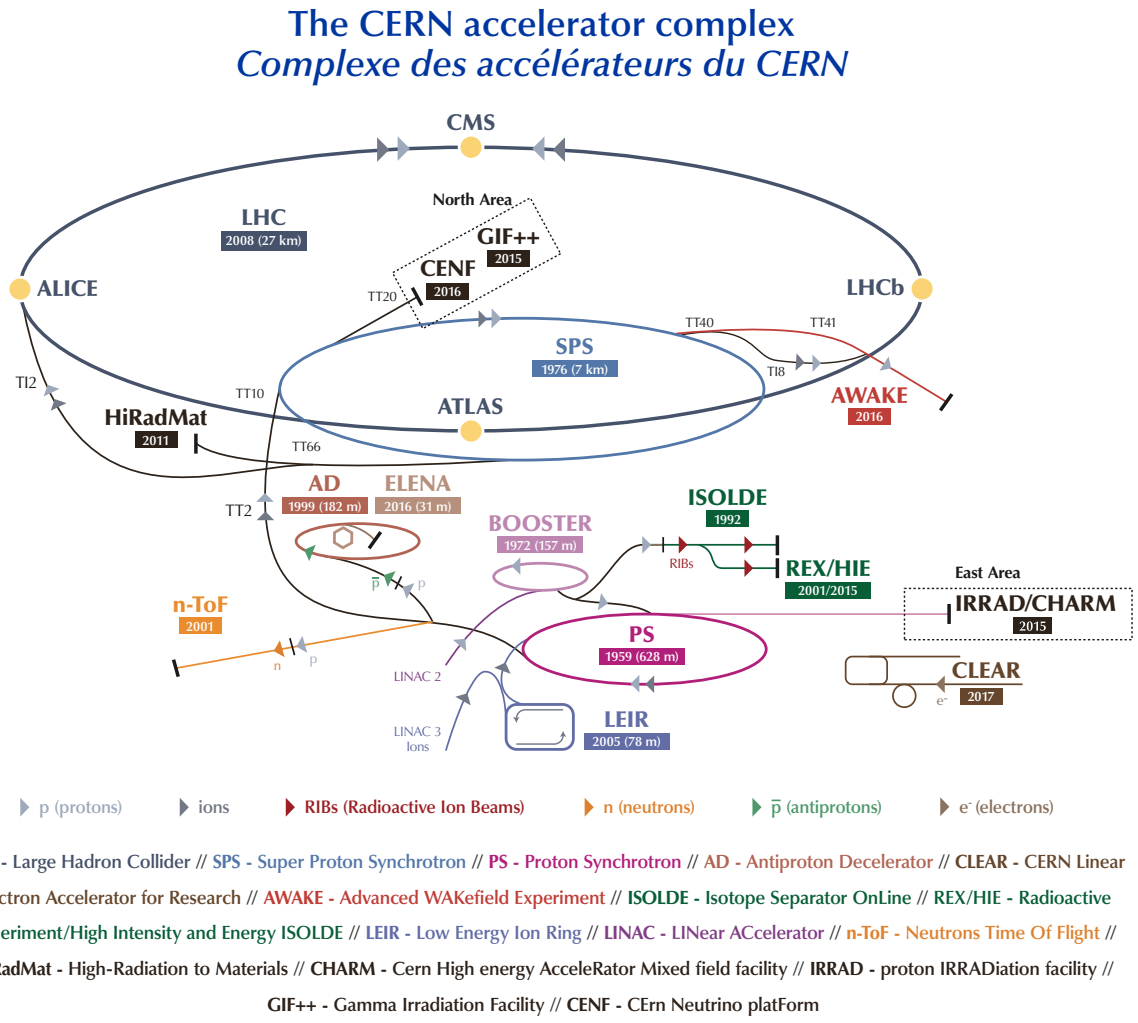


Figure 3.1: Schematic view of the CERN accelerator complex.

quently, they are injected into the LHC and reach the operational energy in about 20 minutes. Within the LHC accelerator, proton or lead ions are bent by a magnetic field with intensities of up to 8 T, generated by superconducting dipole magnets cooled with liquid helium. Particles circulating in the LHC are grouped into bunches, with each bunch separated by a multiple of 25 ns. Other multi-pole magnets are then used to squeeze the beam dimensions at the four collision points, where the major experiments are located:

- ATLAS (A Toroidal LHC ApparatuS) [87] and CMS (Compact Muon Solenoid) [88] are general purpose experiments dedicated to direct searches of particles beyond the SM and to precision measurements of the SM physics at the TeV scale;
- ALICE (A Large Ion Collider Experiment) [89] is conceived for studying the formation and properties of the Quark-Gluon-Plasma (QGP);
- LHCb (Large Hadron Collider beauty) [90] is a general purpose forward spectrometer,

optimised for the study of heavy-flavour physics.

The LHC began operating in 2010 with initial beam energy of 3.5 TeV and a peak instantaneous luminosity of $10^{30} cm^{-2} s^{-1}$ during the Run 1 period (2010-2012). After the first extended shutdown of 2 years, known as LS1, the LHC increased its beam energy to 6.5 TeV, resulting in a total centre-of-mass energy of 13 TeV during the Run 2 phase (2015-2018). The accelerator underwent another prolonged shutdown of 3 years, called LS2, and achieved a centre-of-mass energy of 13.6 TeV in July 2022 with the start of Run 3. The LHC will continue to operate at this energy level until 2026, promising even greater precision and potential for new discoveries.

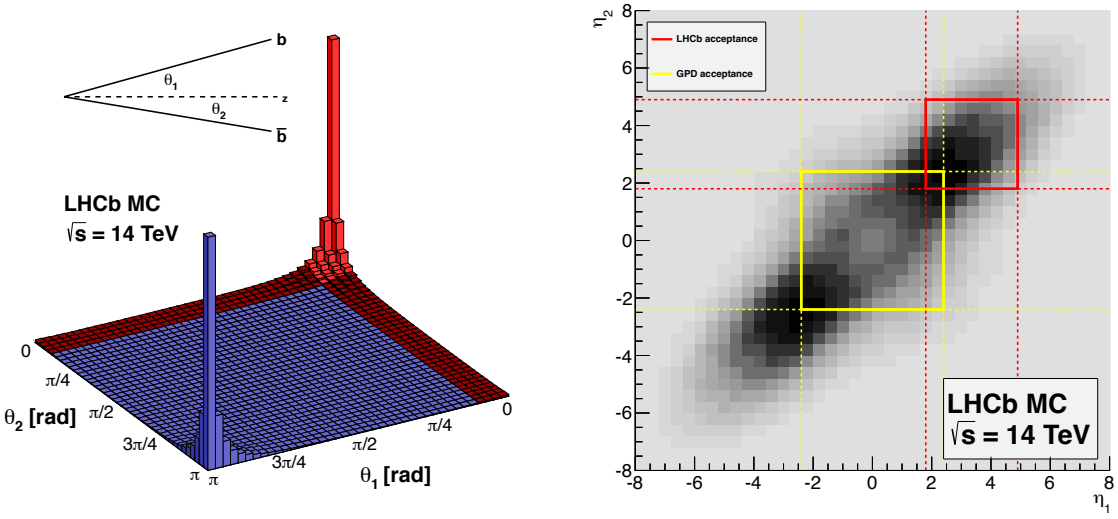
The LHC is expected to reach a higher luminosity with the High-Luminosity Large Hadron Collider (HL-LHC) [91] upgrade at the end of Run 3. The HL-LHC, which should be operational from the beginning of 2029, aims to crank up the performance of the LHC in order to increase the potential for discoveries and to study known mechanisms in greater detail, such as the Higgs boson, and observe rare new phenomena that might reveal themselves.

3.2 The LHCb detector

The LHCb detector [90] is a single-arm forward spectrometer located at Intersection Point 8 (IP8) of the LHC, covering an angular acceptance in the range 10-300 (10-250) mrad in the horizontal (vertical) plane. LHCb was conceived to study CP violating processes to search for new physics in rare decays of beauty and charm hadrons. For this reason, the detector geometry is optimised to study the decays of heavy-flavour hadrons at high energies. These hadrons are primarily produced in a narrow cone that is centred around the beamline. The LHCb detector covers the pseudorapidity (η) range of $2 < \eta < 5$, which corresponds to approximately 24% of the $b\bar{b}$ pairs produced in pp collisions. In comparison, general-purpose detectors like ATLAS and CMS have a geometrical efficiency of around 41%, as illustrated in Figure 3.2. Recently, LHCb has evolved into a general-purpose detector, expanding its capabilities to explore a broader range of physics phenomena, including QED measurements, hadron spectroscopy, heavy ions and much more. The forward geometry is also ideal to the fixed-target configuration where the produced particles are subject to a large Lorentz boost in the laboratory frame.

During the data acquisition periods of Run 1 and Run 2, the LHCb experiment operated with an average instantaneous luminosity of $2 \times 10^{32}\,\mathrm{cm}^{-2} s^{-1}$. To optimise the physics performance, the luminosity was kept constant across an entire LHC fill through adjustments in beam overlap, as displayed in Figure 3.3 (left). This strategy maximises the accumulated integrated luminosity throughout the fill and enables the maintenance of the same trigger configuration, thereby reducing overall systematic uncertainties. The integrated luminosity collected for pp collisions in each data-taking year is reported in Figure 3.3 (right). The peak luminosity reached the value $5 \times 10^{32} cm^{-2} s^{-1}$, which was limited by the detector occupancy and radiation tolerance, and the requirement of an acceptable level of pile-up to ensure reliable detector performance.

A schematic view of the LHCb detector during Run 2 is shown in Figure 3.4. The right-handed coordinate system (O, x, y, z) is centred at the collision point. The x-axis is horizontal,



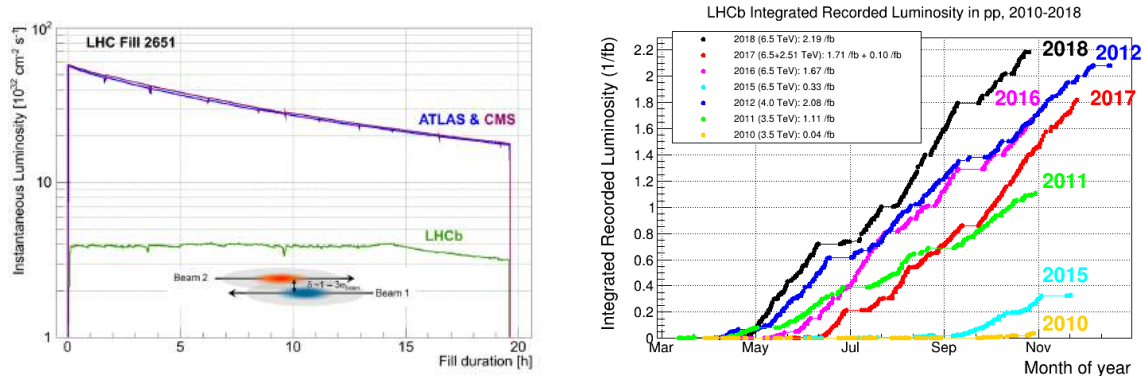
(a) (b)

Figure 3.2: (a) Distribution of the polar angle of the $b\bar{b}$ quark pairs at a collision energy of $\sqrt{s} = 14$ TeV. The red part of the distribution represents the geometrical acceptance of the LHCb detector. (b) Comparison of the geometrical acceptance of produced $b\bar{b}$ pairs between LHCb and the general purpose detectors, ATLAS and CMS [92].

and points from the interaction point towards the outside of the LHC ring. The y-axis is perpendicular to the x-axis and to the beam line pointing upwards and is inclined by 3.601 mrad with respect to the vertical. The z-axis points from the interaction point towards the LHCb detector and is aligned with the beam direction. Tracks produced at LHCb are bent by a dipole magnet with magnetic field lines along the y-direction. The LHCb detector comprises three main systems:

- the **tracking system** is dedicated to determining the three components of particles' momenta and is composed of several sub-detectors, including the VErtex LOcator (VELO), the Tracker Turincensis (TT), the dipole magnet and the T-stations. The momentum of charged particles is determined with a relative uncertainty smaller than 1%. Additionally, a precise measurement of the impact parameter, *i.e.* the minimum distance of a track to a pp collision point (primary vertex), is crucial for the identification of displaced $b-$ and $c-$ quark decay vertices;

- the **particle identification system** plays a crucial role in the flavour physics program. Charged-hadron identification is performed by two Ring Imaging Cherenkov detectors (RICH). RICH 1 is positioned upstream of the magnet and is designed for low-momentum particles ranging from 2 to 60 GeV/c. RICH 2 is located downstream of the magnet and is optimised for particles momenta from 15 to 100 GeV/c. Following RICH 2, a calorimeter system provides information on the energy deposited by electrons, photons, charged and neutral hadrons, along with their spatial coordinates. Lastly, the muon stations, situated in the most downstream region, are utilised to detect muons produced in the collisions or



(a) (b)

Figure 3.3: (a) Evolution of the instantaneous luminosity for LHCb, ATLAS and CMS during LHC fill 2651 [93] (b) Integrated luminosity of pp collisions recorded by the LHCb experiment in each data-taking year.

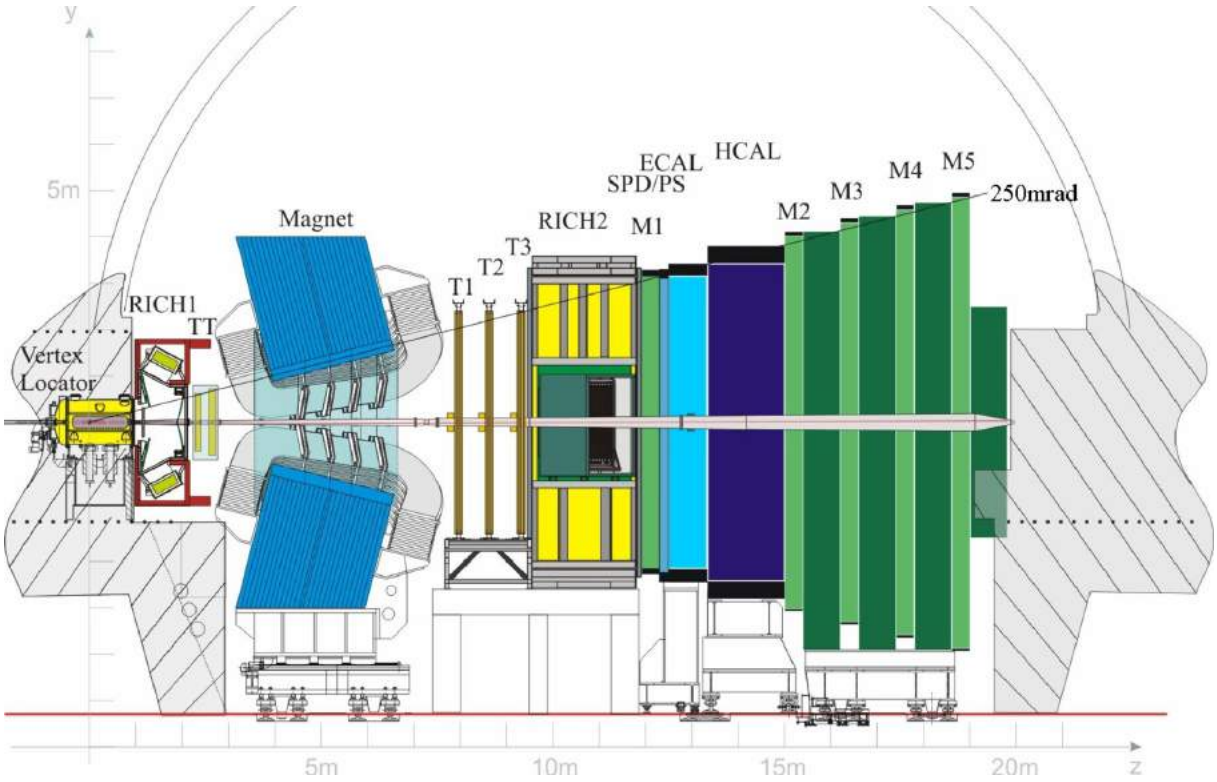


Figure 3.4: Schematic view of the LHCb detector during Run 2 in the non-bending vertical plane $(y-z)$ with the subdetectors highlighted

in the decay of the produce particles;

- the **trigger system** is responsible for selecting events of interest for subsequent physics analyses. Initially, a low-level hardware trigger (L0) rapidly processes partial information from the subdetectors to make an initial selection of interesting events. Subsequently, a high-level software trigger (HLT) searches for more complex signatures.

In the upcoming sections, a more detailed overview of the various LHCb subsystems and their respective performances is presented.

3.2.1 The tracking system

The tracking system aims to reconstruct the trajectories of the particles tracks based on the positions where they interacted with the tracking detectors (hits). The momentum (p) of a particle is derived from its deflection angle after passing through the magnetic field generated by a dipole warm magnet. The first element is the Vertex Locator (VELO), a silicon microstrip detector located around the interaction point, which precisely measures the positions of primary vertices and track Impact Parameters (IP). Another silicon microstrip detector (TT) is located before the dipole magnet, aiming to enhance the momentum resolution of reconstructed tracks and eliminate pairs of tracks that actually belong to the same particle. The tracking system is complemented by the T stations, which, together with VELO information, determine the momentum and flight direction of the particles. The T stations employ diverse technologies for particle detection, including silicon microstrips near the beam pipe and straw-tubes in the outer regions. All tracking detectors share the common characteristics of high spatial resolution (in one or two coordinates) and a low material budget. Particles reconstructed using data from the complete LHCb tracking system exhibit a momentum resolution ranging from 0.5% to 1% within the range of $5 < p < 200\,\text{GeV}/c$.

The Vertex Locator

The VELO [94] is a silicon microstrip detector positioned around the interaction point. Its primary purpose is to precisely measure the coordinates of the Primary Vertex (PV) and Secondary Vertex (SV), to enable the accurate reconstruction of the decay times and impact parameters of heavy flavour hadrons. The VELO covers a length of around 1 meter and consists of 46 double-sided modules arranged perpendicularly with respect to the beam direction, as illustrated in Figure 3.5. Each module comprises two layers of silicon micro-strip sensors, one for measuring the radial distance from the beam axis (R-sensor) and the other for measuring the azimuthal angle (phi-sensor), as depicted in Figure 3.6. The initial four VELO modules, known as the "veto" or "pile-up" modules, are equipped with silicon strips for measuring exclusively the radial coordinate. Their purpose is to estimate the number of primary vertices generated in each bunch-bunch collision, contributing to the luminosity measurement.

The detector is situated inside the LHC vacuum vessel, with the sensors separated from the LHC primary vacuum by only a 300 μm thin aluminium foil (RF foil). This configuration has

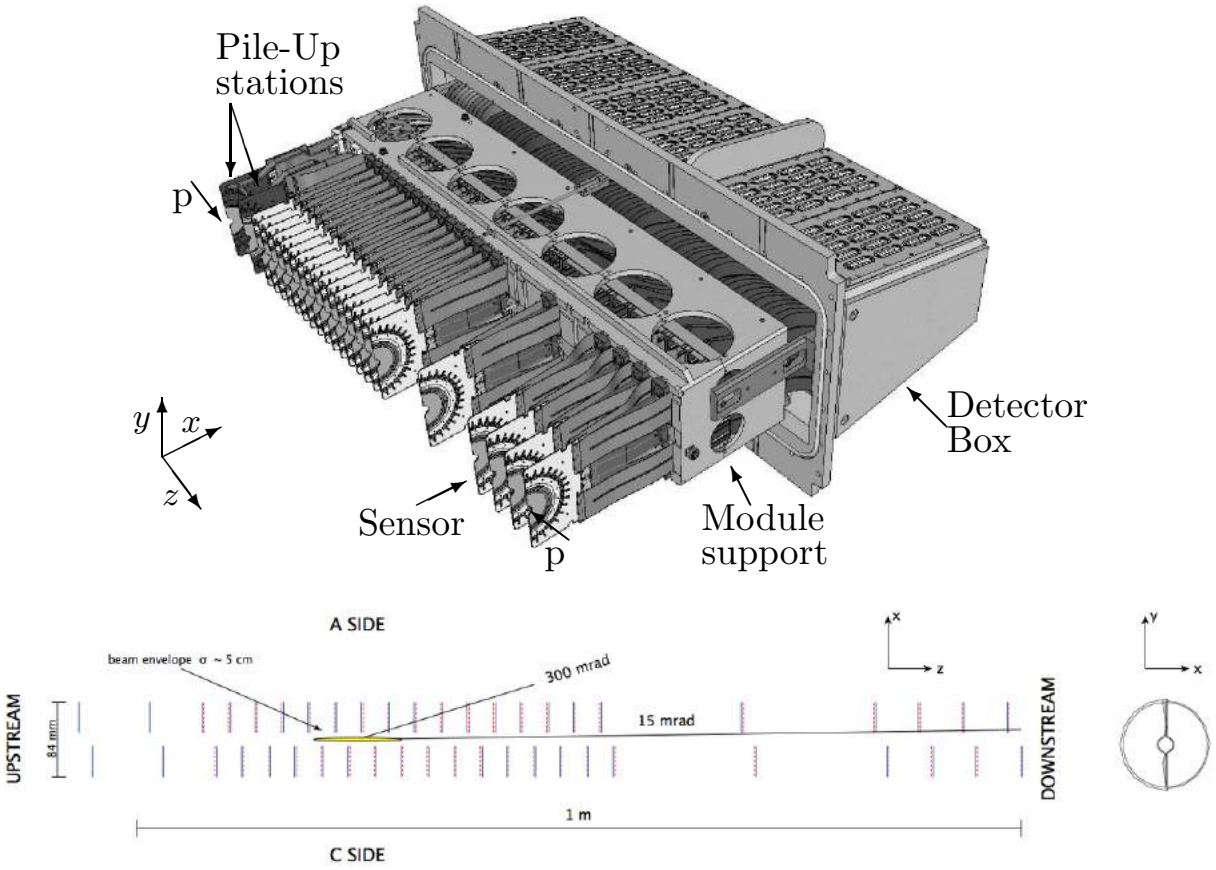
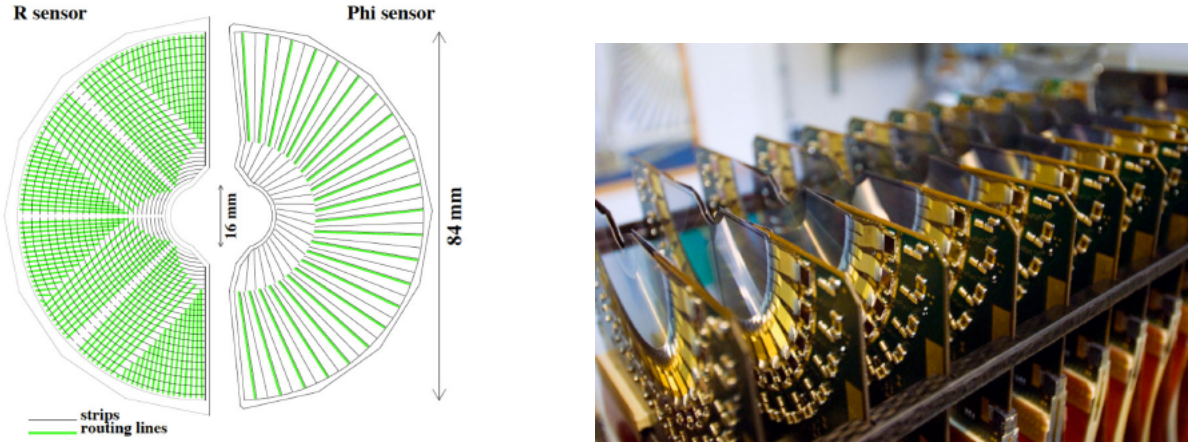


Figure 3.5: (Top) Overview of the VELO A-side. The sensors, module supports and detector box are indicated [95] (Bottom) Schematic view of the VELO sensors along z and cross-section of a VELO sensor in the xy plane [96].

two purposes: protect the sensors from radio-frequency waves generated by the beams while minimising the material that charged particles traverse before reaching the sensors. To protect the sensors from intense radiation exposure, during the injection of LHC beams the two VELO halves are horizontally displaced 29 mm away with respect to the beam position, and they are subsequently moved back to reach the closed position once beam stability is established.

The Tracker Turincensis

The Tracker Turicensis (TT) is placed upstream the magnet and plays an essential role in the overall reconstruction of tracks by providing measurements of particle space coordinates before the magnet. The tracking stations encompass the entire angular acceptance of the LHCb detector and are composed of four planar layers of silicon microstrips, divided into two pairs named TTa and TTb [97], as displayed in Figure 3.7. These layers are separated by a distance of 27 cm along the z-axis. The four detection layers follow a stereo layout denoted as x-u-v-x, featuring vertical strips in the x layers, while the u and v layers have strips rotated at stereo angles of -5 degrees and +5 degrees, respectively. In total, there are 143,000 strips, with lengths ranging from 10 to 40 cm and a pitch of 183 micrometers. These strips are oriented vertically to take advantage of the fact that the tracks are curved in the horizontal plane due to the dipole magnet,



(a) (b)

Figure 3.6: (a) Schematic of the R and Φ sensors of a VELO module (b) Photograph of the assembly of the VELO silicon modules [96].

covering a combined active area of 8 square meters. Shorter strips are utilised in the inner region where particle density is highest, while longer strips are employed in the outer region. Thanks to this configuration, the combination of the u and v measurements allows to extract the y position of the track and provide the 3D information needed for full reconstruction providing an outstanding hit resolution of about 50 μm in the bending plane.

The magnet

The LHCb magnet consists of two coils that are tilted to match the acceptance of the LHCb detector. It has been designed to generate a magnetic field between the upstream and downstream tracking stations, curving the path of charged particles and allowing the determination of the momentum. The magnet produces around 4 Tm of bending power, and this value is known with a relative precision of 10^{-4}. Additionally, the position of the peak of the B-field is determined with an accuracy of a few millimetres. The magnet's polarity is periodically reversed during data collection to investigate potential charge-dependent detection imbalances that may impact the measurements. The magnetic field is precisely assessed across all LHCb areas to consider the impact of any remaining magnetic field effects.

The T-stations

Three tracking stations (T1-T3) are positioned downstream of the magnet. These stations are further divided into two parts, namely the Inner Tracker (IT) [98] and the Outer Tracker (OT) [99]. These T-stations consist of four layers arranged in a stereo configuration x-u-v-x, similar to the TT, and they cover an active area of approximately 30 m^2, as displayed in Figure 3.8. The Inner Tracker (IT) is a silicon micro-strip detector featuring strips with a pitch of 198 μm. As shown in Figure 3.9, the IT is positioned in a cross-shaped region at the center of the three tracking stations and it spans a width of 126 cm and a height of 41 cm. Although it covers only 2% of the LHCb acceptance, it encompasses around 20% of the tracks generated in pp collisions.

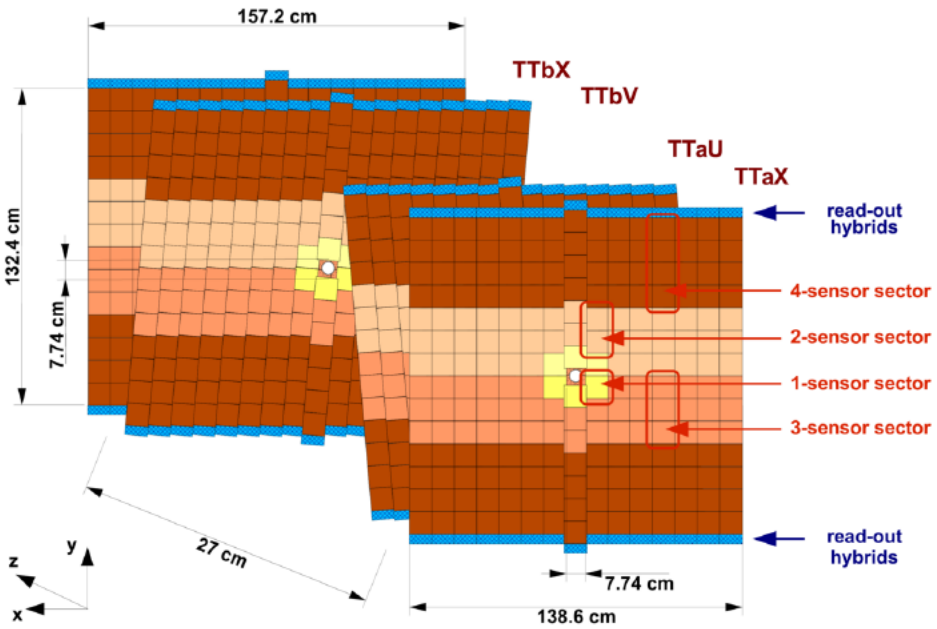


Figure 3.7: Scheme of the Tracker Turincensis subdetector. The 4 layers in x-u-v-x composition are shown. The different colours correspond to different readout sectors [97].

The Outer Tracker (OT) is a straw drift-tube detector comprising approximately 200 gas-tight modules with inner an inner diameter of 5.0 mm and a wall thickness of 75 μm. Charged particles need a minimum momentum of 2 GeV/c to reach the T-stations, forming long tracks. These tracks are crucial for physics analysis and are frequently utilised to assess the performance of the LHCb tracking system.

Track reconstruction and performance

The LHCb reconstruction software features five (standard) track categories, categorised by the individual sub-detectors and combinations of them involved in the track reconstruction process:

- **VELO tracks** register hits in both the $r-$ and $\phi-$sensors of the VELO and are not correlated with hits in any other tracking station. They are employed for the reconstruction of primary vertices;
- **Long tracks** leave hits in the VELO and the T-stations. Long tracks are the most relevant for physics analysis due to a higher precision in momentum determination;
- **Downstream tracks** are produced by charged, long living particles with a origin vertex considerably displaced from the interaction point. Typical examples are K_S^0 mesons or Λ baryons. They leave hits exclusively in the TT and the T-stations. The momentum resolution for these track types is less precise compared to long tracks because of the absence of hits in the VELO, resulting in a longer lever arm;
- **Upstream tracks** are created by low-momentum particles which are considerably deflected by the magnetic field, preventing them from reaching the T-stations. They are

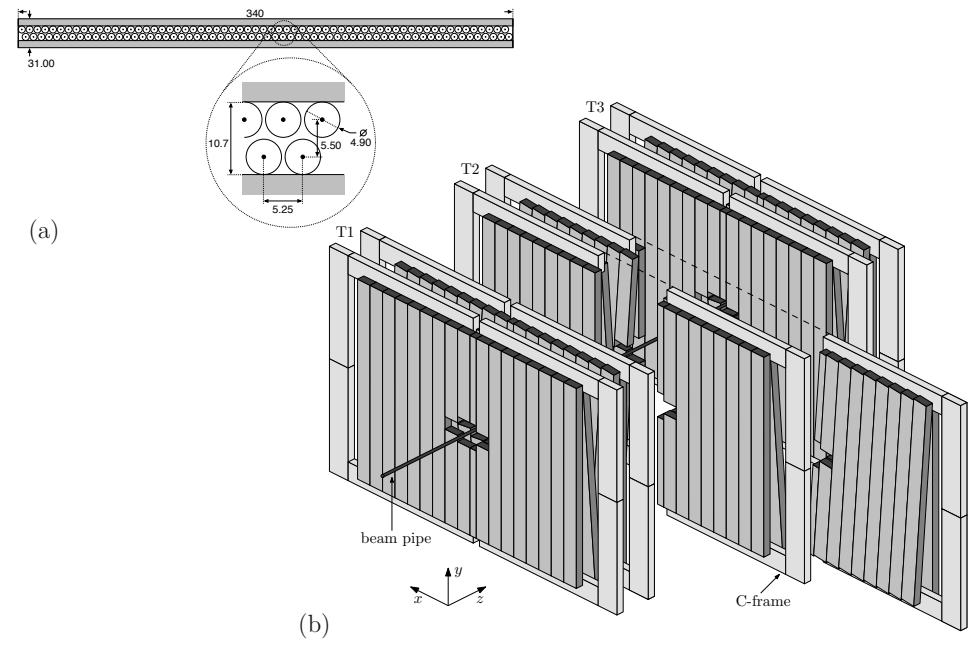
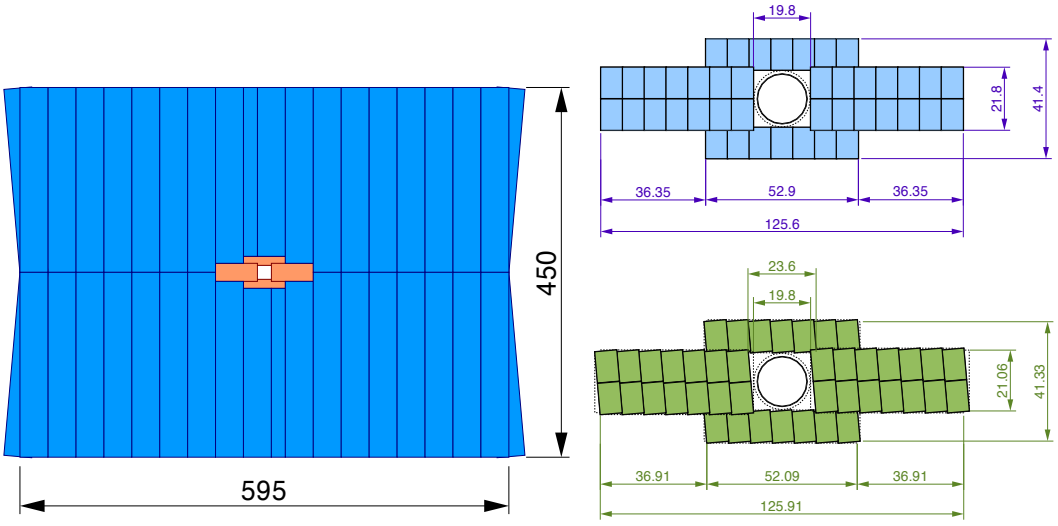


Figure 3.8: (a) OT module cross section. (b) Arrangement of OT straw-tube modules in layers and stations [100].



(a) (b)

Figure 3.9: (a) Front view of a tracking station. Dimensions are given in cm. (b) Layout of x-layer (top) and stereo layer (bottom) in T2. Dimensions are given in cm and refer to the sensitive surface covered by the Inner Tracker [98].

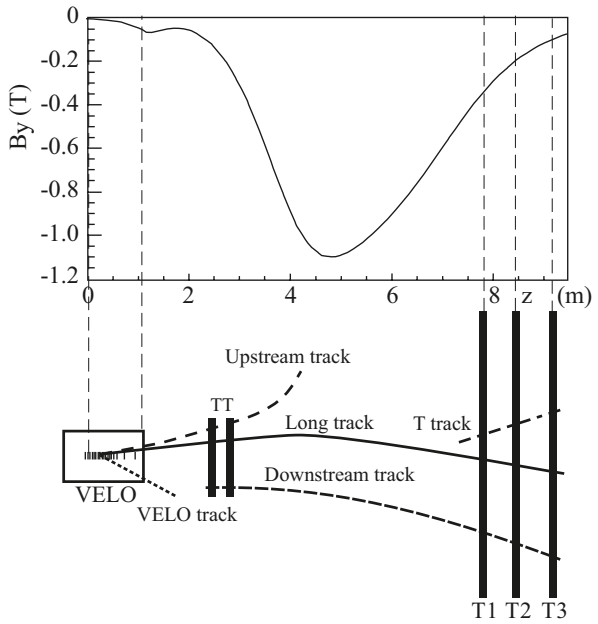


Figure 3.10: A schematic illustration of the various track types. For reference, the main B-field component (By) is plotted above as a function of the z coordinate [90].

employed in the reconstruction process for RICH1;

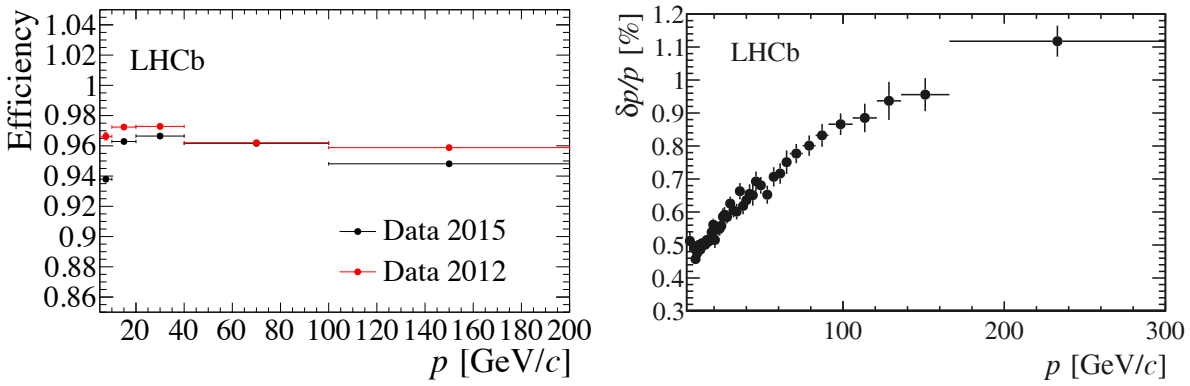
- **T tracks**: tracks reconstructed with hits only in the T stations. The corresponding particles may come from material interactions or decays of very long living particles.

LHCb employs two algorithms, namely "Forward Tracking" and "Track Matching," for identifying long tracks. Since long tracks hold particular significance for physics analyses, a brief discussion of these tracking methods will follow:

- **Forward tracking** [101]: the initial track seed is reconstructed within the modules of the VELO detector. Given the negligible magnetic field intensity at the VELO z coordinates, a straight-line model is adopted to estimate track positions and slopes. The reconstructed track segment is extended to the first modules of the TT, and corresponding energy deposits are sought. Utilising the available rough momentum estimation, the corresponding position in the T-stations is predicted, assuming that the magnetic field effect bends in the particle trajectory.

- **Track matching** [102]: track seeds are simultaneously reconstructed in the VELO detector and the T-stations, then extrapolated both forward and backward to the magnet plane. A compatibility criterion is established to match the two track segments, considering their distances in x and y at the magnet bending plane. Subsequently, tracks are fitted to determine the track parameters and particle momentum.

A Kalman filter model [103] is applied to address the effects of multiple scattering. The resulting χ^2_{ndf} from the track fit is utilised to assess the quality of track reconstruction. To eliminate



(a) (b)

Figure 3.11: (a) Long track reconstruction efficiency plotted against momentum for both Run1 and Run2 data [106]. (b) Momentum resolution as a function of momentum in Run 1 [107].

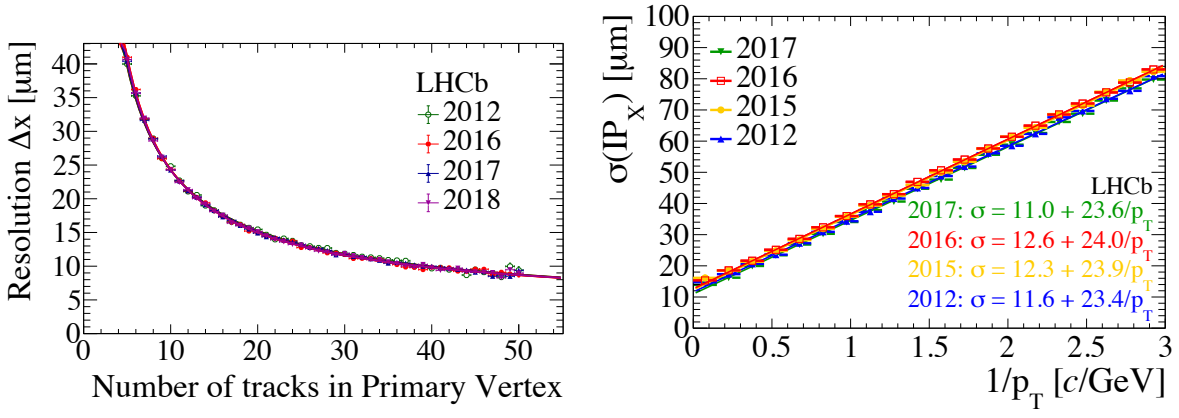
redundancy arising from the two track finding strategies, a clone killer algorithm [104] is executed, removing tracks reconstructed twice. Additionally, a dedicated neural network classifier is employed to determine the probability that a track is reconstructed from energy deposits left by different particles [105]. The output of this classifier is referred as GhostProb.

The tracking efficiency of long tracks is assessed through a tag-and-probe technique involving $J/\psi \to \mu^+\mu^-$ decays [106]. In this method, one muon serves as the "tag" and is fully reconstructed, while the other serves as the "probe" and is only partially reconstructed. Following the reconstruction of J/ψ candidates via an invariant mass fit, a search is conducted for a long track match for the partially reconstructed probe. The final efficiency is obtained with the ratio of efficient candidates over the total number of candidates. The efficiency of reconstructing long tracks as a function of momentum is shown in Figure 3.11, achieving a performance exceeding 94% efficiency. The figure also illustrates the distribution of momentum resolution as a function of momenta, showing values below 1% for a broad range of momenta.

Primary Vertex reconstruction and performance

Primary Vertices (PVs) reconstruction at LHCb relies on VELO tracks. The reconstruction algorithm analyses the distribution of tracks in an event to deduce the point of origin for most tracks. The reconstruction process consists of two main steps: seeding and fitting. In the seeding step, potential Primary Vertex (PV) candidates are identified by examining the spatial points where an accumulation of track trajectories is evident. The second step involves employing the weighted least square method to determine the final vertex position. Candidates are fitted by decreasing multiplicity, a strategy that helps prevent low-multiplicity secondary vertices from absorbing tracks from their corresponding PVs. Following the fitting process, the separation between the newly obtained PV and those already fitted is verified to prevent against the identification of fake PVs. The seeding and fitting steps are iterated until no new PVs are identified in the event.

The optimisation of PV reconstruction aims at achieving a high efficiency, defined as the ratio of reconstructed to generated vertices in simulations, and minimising the occurrence of fake



(a) (b)

Figure 3.12: (a) The primary vertex resolution along the x-axis during the Run1 and Run2 data collection periods, as a function of the number of tracks comprising the vertex. (b) The distribution of impact parameter resolution along the x-axis in relation to the inverse transverse momentum, measured during the Run1 and Run2 data acquisition phases [108].

PVs. Fake PVs may result from the random association of tracks in high-multiplicity events. Generally, reconstructing PVs in events with higher multiplicity, where more tracks are generated, is more straightforward, given that more tracks have higher probability to fall within the VELO acceptance. However, the challenge becomes more pronounced in events with multiple PVs.

Figure 5.9 illustrates the distribution of primary vertex and IP resolutions. The linear dependence of the IP resolution on $1/p_T$ is a consequence of multiple scattering and the geometry of the vertex detector. At high-p_T, the IP resolution is primarily influenced by PV resolution rather than track reconstruction, as tracking at high-p_T achieves very high precision while PV reconstruction is mainly contingent on the number of tracks associated with it.

3.2.2 The particle identification system

Particle identification (PID) involves categorising detected particles based on their mass and holds crucial importance for flavour physics studies. Within the LHCb experiment, particle identification is provided by three subdetectors: the Ring Imaging Cherenkov detectors (RICH), a calorimeter system, and the muon stations. Information from these subdetectors is typically integrated into a combined likelihood to optimise identification efficiency and minimise mis-identification rates. The efficiency, determined using calibration samples, varies between $90-100\%$ for different particle types.

The Ring Imaging Cherenkov system

The RICH detectors [109] play a crucial role in providing particle identification information for the LHCb experiment. Their primary function is to identify charged hadrons (π, K, p) within the momentum range of 2.6 to 100 GeV and also contribute significantly to the identification of charged leptons (e, μ), especially at low momentum. Identifying charged hadrons is essential

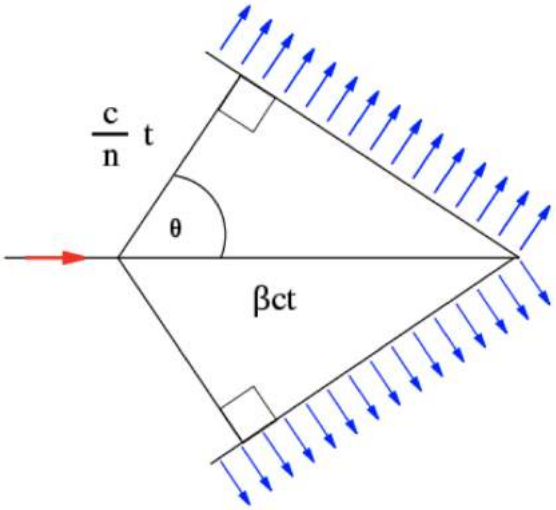


Figure 3.13: Diagram illustrating the process of Cherenkov radiation emission.

for the majority of the LHCb physics analyses. The RICH system enables the discrimination of final states with otherwise identical topologies, such as $B_0 \to \pi^+\pi^-, K^+\pi^-, K^+K^-$ decay modes. Furthermore, it reduces significantly the combinatorial background in decay modes involving hadrons in the final state, for instance, $B_s^0 \to \phi\phi$, where $\phi \to K^+K^-$. Without PID requirements, the combinatorial background in such cases would be unsustainably large.

The working principle of the RICH system relies on the Cherenkov effect: a charged particle traversing a dielectric medium of refractive index n with a velocity $\beta > 1/n$ emits Cherenkov photons with a characteristic Cherenkov angle (θ_c), as shown in Figure 3.13. The value of the (θ_c) angle is directly related to the velocity of the particle β through the formula

$$\cos(\theta_c) = \frac{1}{\beta n} . \tag{3.1}$$

The mass of the particle m is then evaluated from the momentum estimate provided by the tracking system, as $m = p/v$, with the velocity v retrieved by the θ_c measurement from the RICH detectors. The minimum particle velocity required for Cherenkov radiation to be emitted at $\theta_c = 0°$ is then given by

$$\beta_{\text{th}} = \frac{1}{n} . \tag{3.2}$$

Conversely, the maximum angle of emission occurs when the particle velocity approaches the speed of light ($\beta \to 1$), and it is given by

$$(\theta_c)_{\text{max}} = \arccos\left(\frac{1}{n}\right) . \tag{3.3}$$

Under this condition, different charged particles cannot be distinguished experimentally. These saturated tracks can be used to determine the Cherenkov angle resolution of a RICH detector, by comparing the measured Cherenkov angle with the expected Cherenkov angle obtained from the known refractive index. Cherenkov radiation occurs promptly in time. The timing of photons

hitting the photodetector planes distributed within a few hundred picoseconds, spread in a continuous frequency spectrum. The yield of Cherenkov photons emitted per unit length of the particle's path in the radiator and per unit of wavelength is determined by the formula

$$\frac{\mathrm{d}^2 N_\gamma}{\mathrm{d}x\mathrm{d}\lambda} = \frac{2\pi z^2 \alpha}{\lambda^2}\left(1 - \frac{1}{\beta^2 n^2}\right) \tag{3.4}$$

where z is the charge of the particle in units of the electron charge and α is the electromagnetic coupling constant [110]. This equation shows that the Cherenkov radiation is dominated by short wavelengths in the spectrum.

The RICH system comprises two detectors, RICH 1 and RICH 2. RICH 1 is situated between the VELO and the TT, covering the entire spectrometer angular acceptance of 25–300 mrad. It is optimised for low-momentum particles of $2-60$ GeV and employs C_4F_{10} as a radiator material with a refractive index of $n = 1.0014$. RICH 2 is positioned after the third tracking station and before the first muon station, covering the angular acceptance of $15-120$ mrad. It employs CF_4 as a radiator material ($n = 1.0005$). Approximately 5% of CO_2 has been added to CF_4 to quench scintillation in the gas. RICH 2 covers the high-momentum region of $15-100$ GeV. Both RICH detectors use spherical mirrors to focus the cones of Cherenkov light emitted from a charged particle into rings on an array of Hybrid Photon Detectors (HPDs). An intermediate flat mirror, positioned between the spherical mirror and the HPD array, allows to position the HPDs outside the spectrometer's acceptance. Figure 3.14 illustrates a schematic of the two RICH detectors.

The HPDs [112] are comprised of a photocathode deposited on the inner surface of the quartz window within a vacuum tube containing guiding fields. These fields guide the electrons generated at the photocathode (photoelectrons) to a silicon sensor, where they are registered as hits. Operating within a high-voltage range of 16 to 20 kV applied between the anode and cathode, the tube ensures that the high energy of the photoelectrons entering the silicon sensor effectively isolates the signal from the electronic noise pedestal. This results in a highly efficient and virtually noise-free photon detector. Collectively, the two RICH detectors feature approximately 500,000 active channels. Within a typical LHC event readout window of 25 ns, only 10 dark noise counts are observed. The trajectory of the photoelectron through the focusing optics of the HPD can be influenced by external electric and magnetic fields. Although most of these effects are minimised through extensive magnetic shielding around the HPDs, any residual distortions are rectified through software corrections.

The overall Cherenkov angle resolution plays a pivotal role in determining the quality of the RICH detectors, as it serves as the fundamental parameter for distinguishing between particle types. This distinction is particularly crucial at high particle momentum, where Cherenkov angles tend to saturate. Alongside chromatic dispersion, other factors contribute to limiting the Cherenkov angle resolution, including imperfect focusing of the optics, leading to uncertainty regarding the emission point of Cherenkov radiation, and the pixel size of the photodetectors. On average, the Cherenkov angle resolution during Run 2 is 1.65 mrad for RICH 1 and 0.67 mrad for RICH 2. Figure 3.16 shows the relation between the Cherenkov angle and momentum for various particle species in each of the radiators used in the LHCb RICH detectors. A summary

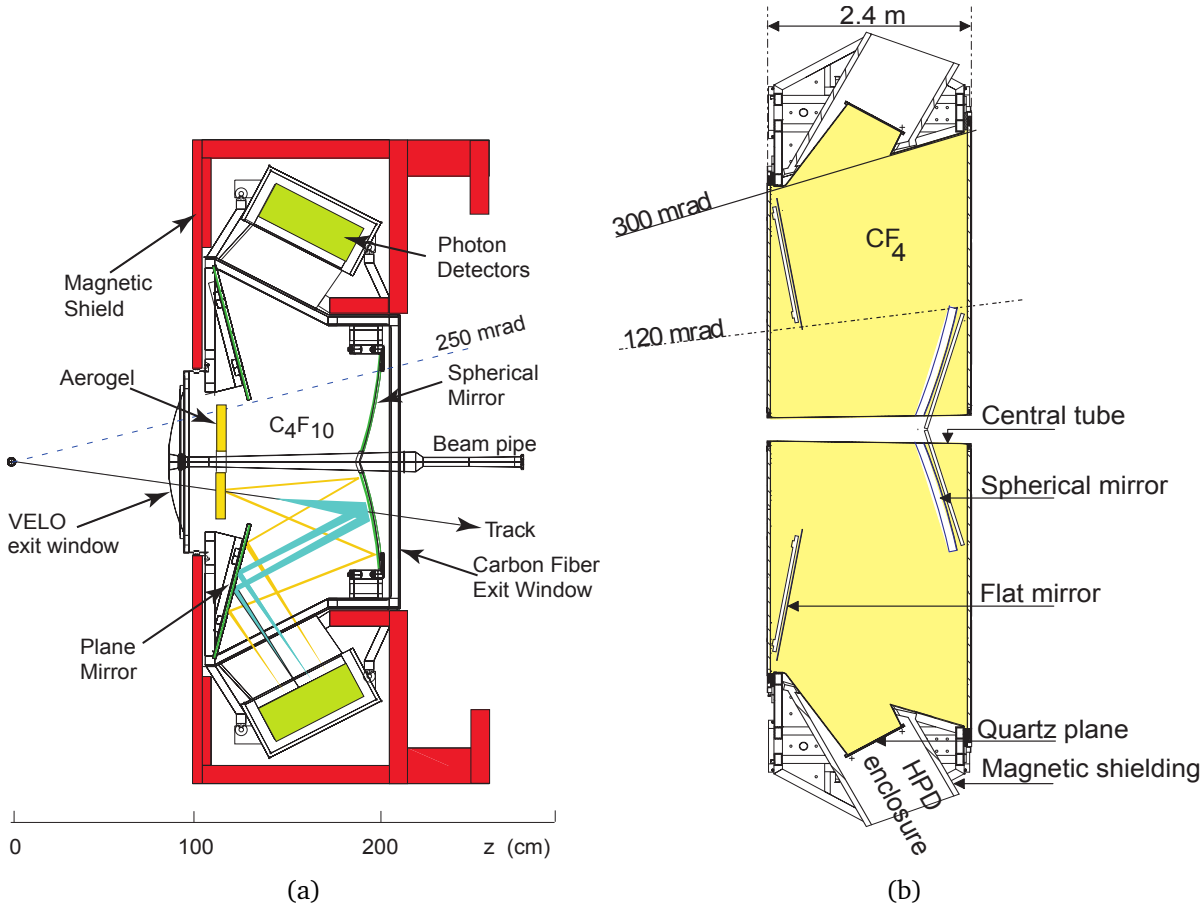


Figure 3.14: The optical system of (a) RICH1 and (b) RICH2 [111]

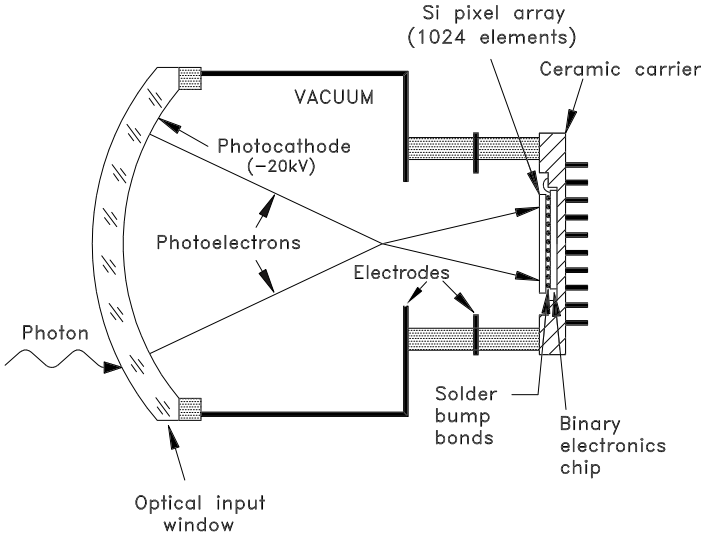
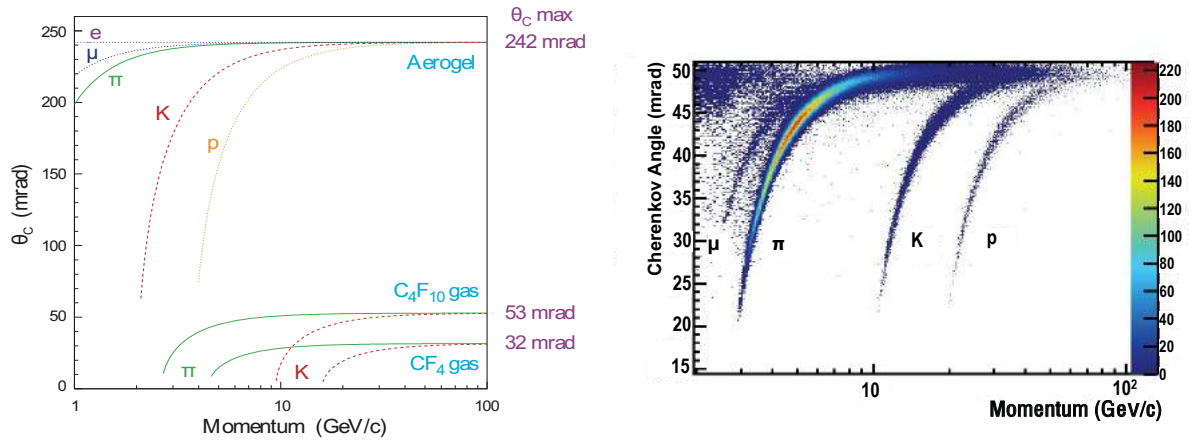


Figure 3.15: The focusing system of an HPD [111]. A photoelectron generated at the photocathode undergoes acceleration towards the silicon sensor chip.



(a) (b)

Figure 3.16: (a) Cherenkov angle as a function of particle momentum for the RICH radiators (including aerogel) for different particle masses [90]. (b) Cherenkov angle for isolated tracks, defined as tracks where the Cherenkov ring does not overlap with any other ring, in the C_4F_{10} radiator [113].

	RICH 1	RICH 2
Gas	C_4F_{10}	CF_4
Refractive index	1.0014	1.0005
Angular acceptance [mrad]	25 – 300	15 – 120
Length [cm]	86	196
Momentum range [GeV/c]	2 – 60	15 – 100
$(\theta_c)_{\max}$ [mrad]	53	32
$\pi^{\pm}_{\text{th}}$ [GeV/c]	2.6	4.4
$K^{\pm}_{\text{th}}$ [GeV/c]	9.3	15.6
$p^{\pm}_{\text{th}}$ [GeV/c]	17.7	29.7

Table 3.1: Summary of the main properties of the LHCb RICH detectors [114].

of main the properties of the RICH system is reported in Table 3.1.

The calorimeter system

The calorimeter system [115] is responsible for measuring the energies deposited by electrons, photons, and hadrons and positions of the relative clusters. It also identifies candidates with high transverse energy for the L0 trigger. Scintillation light generated throughout the calorimeter system is guided to Photo Multipliers Tubes (PMTs), which convert this light into an electric signal. The calorimeters comprise three components: a scintillating pad/preshower detector (SPD/PS), an electromagnetic calorimeter (ECAL), and a hadronic calorimeter (HCAL). A layout of the calorimeter system is displayed in Figure 3.17.

The SPD/PS consists of a 15 mm thick lead absorber sandwiched between plates of scintillation pads. These pads are read out through wavelength-shifting fibres guiding the light to photomultipliers. The first layer of scintillating pads records the deposited ionisation, enabling the rejection of π_0 with high transverse energy (E_T). The lead absorber initiates the showering,

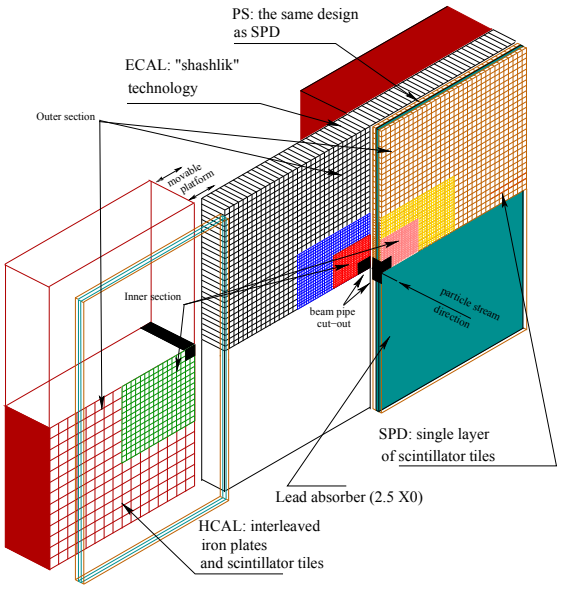


Figure 3.17: Layout of the calorimeter system [116].

detectable in the PS. It introduces longitudinal segmentation to the ECAL, enhancing discrimination between pions and electrons due to their distinct shower lengths. The dimensions of SPD and PS pads are approximately 4×4 cm^2, 6×6 cm^2, and 12×12 cm^2, corresponding to a radiation length of $X_0 \sim 2.5$, and a hadronic interaction length of $\lambda \sim 0.06$.

The ECAL, positioned after the SPD/PS, is a heterogeneous lead/scintillator sampling calorimeter designed for measuring the energy of photons and electrons. Separation between photons and electrons is accomplished by the SPD and PS preceding the ECAL. A lead layer's thickness between the SPD and PS is chosen so that photons initiate the electromagnetic shower within it. Additionally, electrons leave hits in the SPD, enabling their differentiation from photons. Once this separation is achieved, the ECAL determines the energies of these particles with an energy resolution of

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 1\% \tag{3.5}$$

with a stochastic term of 10 %, due to fluctuations related to the physical development of the shower, and a 1 % constant term, corresponding to mis-calibrations, non-linearities or leakages of the detector.

The HCAL is located next to the ECAL, and provides a measurement of the transverse energy of hadrons. This information is utilised both for the L0 trigger and for the offline PID. The detector is composed of alternating layers of scintillating tiles and iron absorbers. The measured energy resolution is:

$$\frac{\sigma_E}{E} = \frac{70\%}{\sqrt{E}} \oplus 10\% \tag{3.6}$$

with a stochastic term of 10 %, due to fluctuations related to the physical development of the shower, and a 1 % constant term, corresponding to mis-calibrations, non-linearities or leakages of the detector. Distinct segmentation schemes were devised for the calorimeter system to ac-

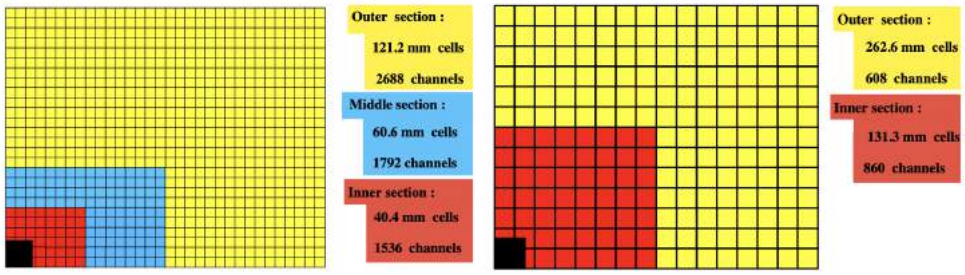


Figure 3.18: Segments with varying granularity for the ECAL (left) and HCAL (right) [115].

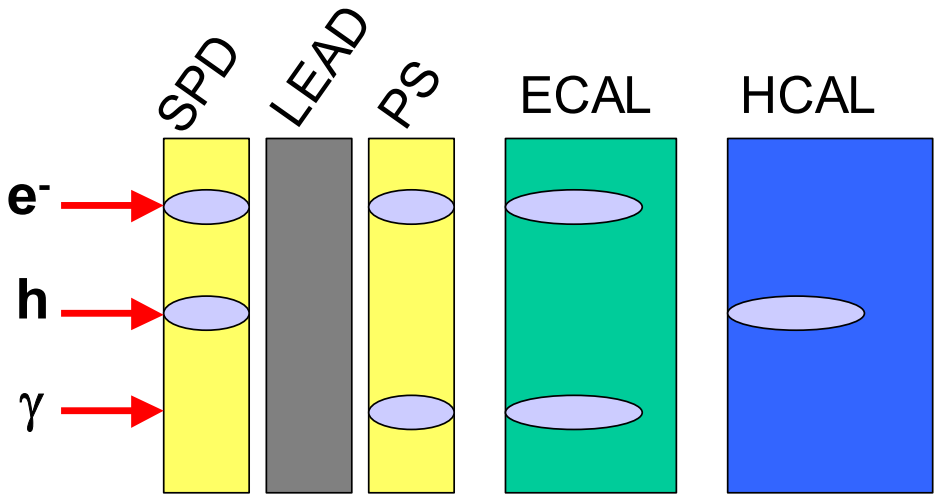


Figure 3.19: Illustration of the typical electromagnetic and hadronic showers and their interaction with the various detectors within the calorimeter system [117].

commodate variations in detector occupancy between the inner and outer regions. The SPD, PS, and ECAL detectors are subdivided into three regions, whereas the HCAL is divided into two regions, as illustrated in Figure 3.18.

The energy deposits in different detectors resulting from various particle types are displayed in Figure 3.19. Photons do not register hits in the SPD station but interact with the lead converter, initiating an electromagnetic shower in the PS and ECAL. Electrons exhibit similar behaviour to photons at the PS and ECAL levels but also leave hits in the SPD detector. Neutral hadrons, such as π_0, can be categorised as merged if reconstructed from a single cluster in ECAL or resolved when reconstructed from two photons. Generally, hadrons act as Minimum Ionising Particles (MIPs) in the SPD, PS, and ECAL (with the energy deposit in the latter being highly variable). Almost all the energy is released in the HCAL.

The muon system

The original muon system [118] comprises five stations, denoted as M1 to M5, aimed at the identification and triggering of muons. The detection of muons is a crucial part of LHCb operation, both for fast triggering at L0 and the identification and momentum measurement of muons. In particular, muons play a significant role as decay products in the study of quarkonia, rare de-

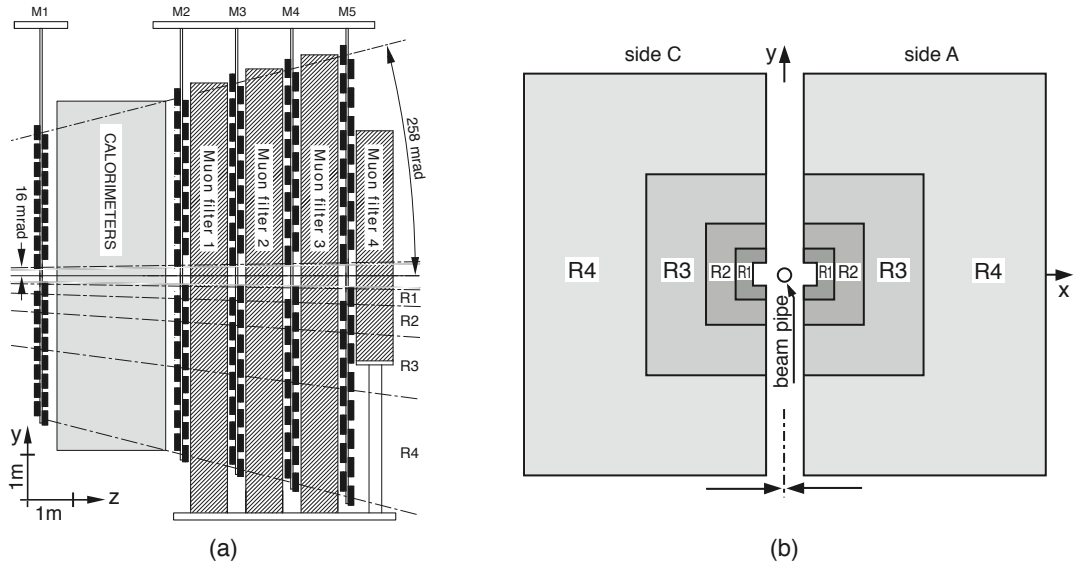


Figure 3.20: (a) Profile view of the LHCb Muon Detector. (b) Configuration of the stations featuring the four regions R1–R4 [118].

cays, and CP violation. The muon system layout is displayed in Figure 3.20. The first station is positioned before the calorimeter to enhance the transverse momentum resolution of the muon trigger. Stations M2 to M5 are situated after the HCAL and are interspersed with 80 cm thick iron absorbers. Consequently, only muons with a minimum momentum of 6 GeV/c can traverse the entire system. In the innermost M1 region, where MWPC sustenance becomes challenging due to high detector occupancy, Gas Electron Multipliers (GEM) chambers are employed. GEM chambers offer greater radiation resistance and higher granularity and spatial resolution to handle the large particle flux. The filling gas mixture for GEMs is $Ar/CO_2/CF_4$ (45:15:40). Chambers from from M2 to M5 are equipped with MultiWire Proportional Chambers (MWPC) detectors, featuring an $Ar/CO_2/CF_4$ (40/55/5) gas mixture. Each muon station is subdivided into four regions (R1-R4) with varying granularity, featuring a finer segmentation closer to the beam pipe and coarser segmentation in the outer regions.

Particle identification performance

All the information collected by the PID detectors are combined to perform particle identification of each track. In the LHCb experiment, two methods are used to achieve this task.

The first considers likelihood functions [109] for each particle type, computed on the basis of the information from RICH detectors. Charged particles are identified by constructing a log-likelihood value using the hits registered on the detector plane and the expected hit patterns calculated from all the tracks passing through the detector [119]. Each particle-type hypothesis is tried iteratively to find the configuration of particle hypotheses that minimises the global likelihood. Then, a delta-log-likelihood (ΔLL) value is computed for each track as the difference between the log-likelihoods when the track is given a specific hypothesis and that of the pion hypothesis. The performance of the RICH detectors is measured using control samples of $K^0_s \rightarrow$

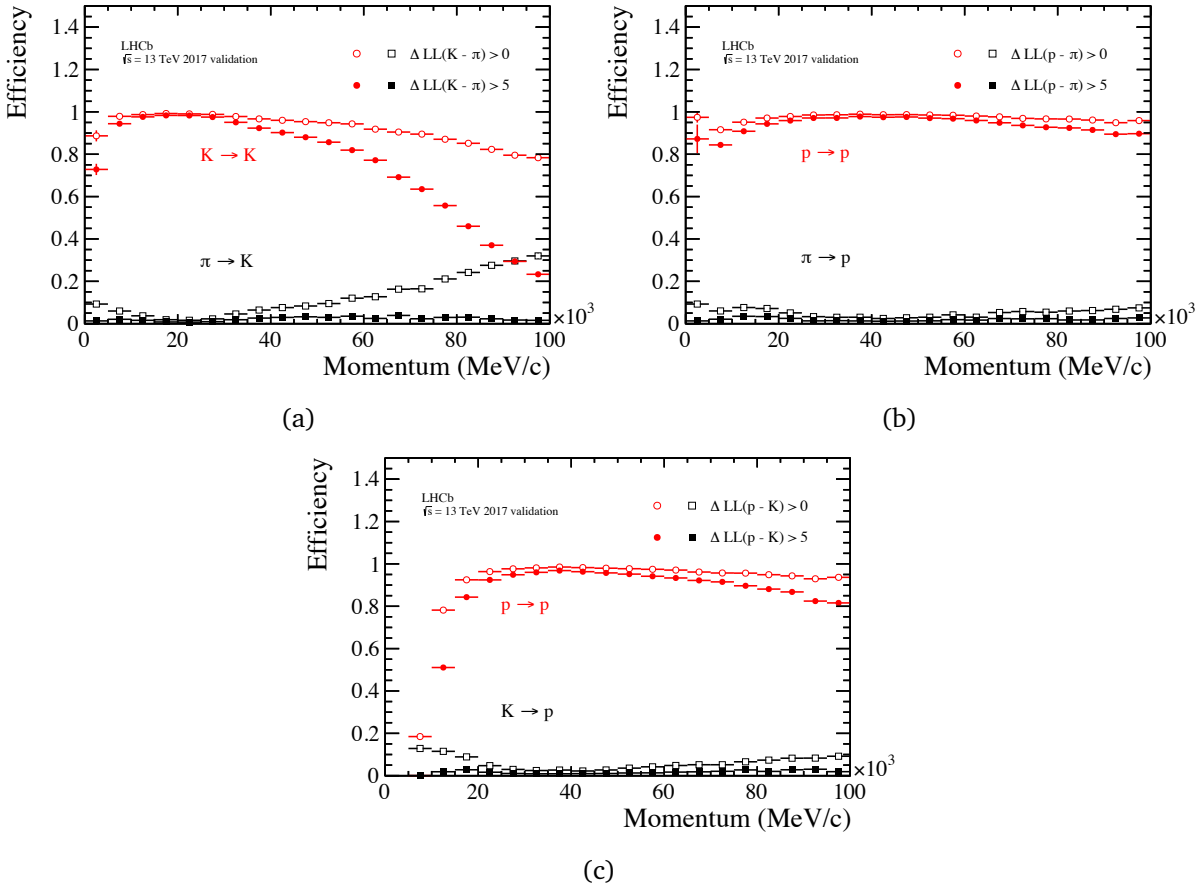
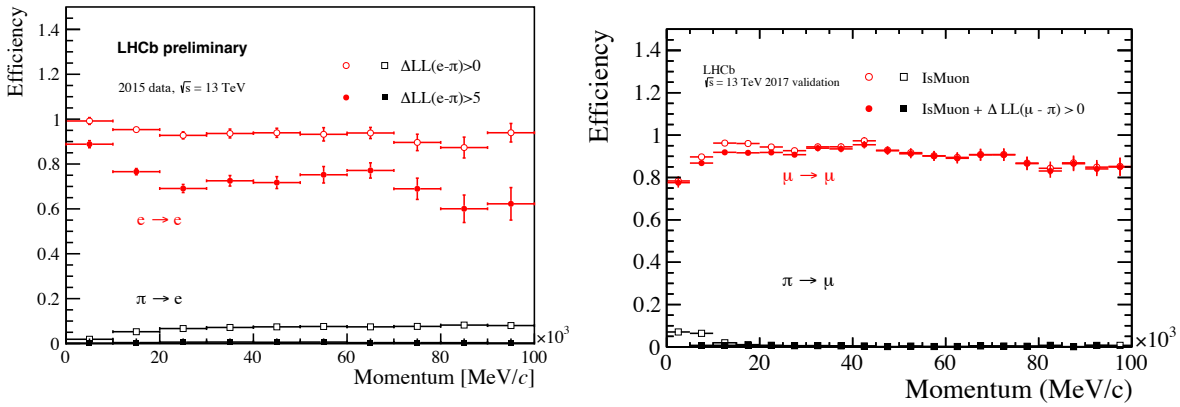


Figure 3.21: The efficiency of selecting kaons (a), protons (b and c), with the associate leakage from misidentifying pions (a and b) and kaons (c) as a function of momentum. Two selections are made, a loose selection (hollow circles) and a tight selection (solid circles) [109].

$\pi^+\pi^-, \Lambda \to p\pi^-$ and $D_*^+ \to D_0(\to K^-\pi^+)\pi^+$ decays, which can be selected by using kinematic requirements alone. The identification efficiencies and misidentification rates of pions, kaons and protons are calculated for a range of track momenta. An example is given in Figure 3.21, where the $K(p)$ identification efficiency and the $\pi \to K(K \to p)$ misidentification probabilities for two different PID requirements are shown as a function of the particle momenta. Electrons and photons are identified and distinguished using the SPD, the PS and the ECAL detectors. The muon chambers isolate muons from all the previously mentioned particles. Figure 3.22 shows the identification efficiency for $e(\mu)$ and the misidentification probabilities for $e \to \pi(\mu \to \pi)$ under two distinct PID requirements, presented as a function of particle momenta.

The second approach is based on neural networks [120] trained on simulated events and validated on calibration samples from data, employing the TMVA toolkit [121]. Information from the tracking system, calorimeter system, RICH detectors, and muon chambers is accounted to classify a given track as a $K, \pi, e, p,$ or μ particle. An advantage over standard ΔLL PID variables is that the ProbNN variables consider correlations among the various subdetectors and different ΔLL, resulting in improved performance. Figure 3.23 illustrates the performance of both ProbNNμ and ProbNNp compared to their respective ΔLL variables.



(a) (b)

Figure 3.22: (a) Electron PID performances during 2015 data taking. Electron efficiency ($e \to e$) and mis-identification ($\pi \to e$) for the two different PIDe cuts[122]. (b) Muon PID performances during 2017 data taking. Muon efficiency ($e \to e$) and mis-identification ($\pi \to e$) for the two different PIDμ cuts [122].

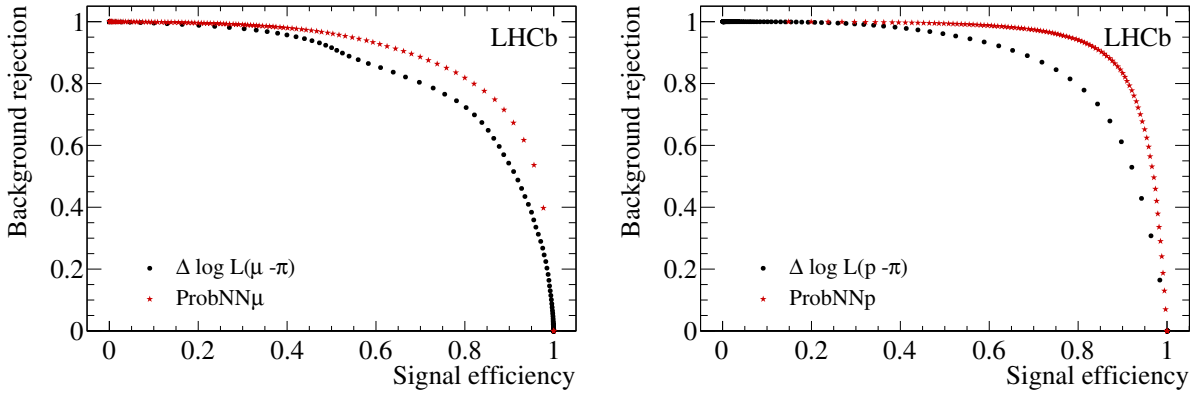


Figure 3.23: Rates of background rejection are presented as a function of muon (left) and proton (right) identification efficiency. The variables ΔL and ProbNN are represented in black and red, respectively, for comparison, utilising samples from the 2012 data-taking period.

3.2.3 The trigger system

The LHCb trigger system utilises information gathered by various subdetectors and reduce the event rate from the 40 MHz rate of bunch crossings at the LHC to a manageable 2-5 kHz for events to be recorded. The trigger system is organised into two levels: the Level-0 trigger (L0) and the High Level Trigger (HLT). L0 trigger is implemented in hardware and reduces the rate at which the entire detector is read out to 1 MHz, with the maximum rate dictated by the frontend electronics of the various subdetectors. The HLT is a software trigger that conducts a full event reconstruction in a computer farm comprising approximately 25,000 CPU cores across roughly 1600 physical nodes, each equipped with at least 1 TB of local storage space.

L0 hardware trigger

L0 is implemented through custom hardware that synchronises with the 40 MHz collision rate and reduces the event rate to 1 MHz. L0 trigger comprises two independent systems: the calorimeter trigger and the muon trigger. The L0-calorimeter system utilises information from the SPD, PS, ECAL, and HCAL detectors, making decisions based on the presence of high E_T electrons, photons, and hadrons. The L0-muon trigger looks for the two highest p_T muons in each quadrant. The muon system is able to perform standalone track reconstruction. Tracks are built from hits that form a straight line through all five muon stations and must be consistent with originating from the interaction point. By using information from the SPD/PS system and the muon system, it is possible to classify different types of L0 trigger streams:

- L0Hadron refers to the HCAL cluster with the highest E_T. If this cluster aligns with the highest E_T cluster in the ECAL, the E_T of the hadron candidates is determined by the sum of the HCAL and ECAL clusters.
- L0Electron is defined as the ECAL cluster with the highest E_T, featuring one or two hits in the SPD and corresponding PS cells.
- L0Photon is identified as the ECAL cluster with the highest E_T, possessing no hits in the SPD but having one or two hits in the corresponding PS cells.
- L0Muon trigger seeks the two highest p_T muons in each quadrant. An event is retained if either the highest p_T of any muon candidate in the event is above a certain threshold (L0Muon) or the product of the highest and second-highest p_T is above another threshold (L0DiMuon).

High level trigger

The HLT is the software trigger, operating asynchronously with the collision rate, on a processor farm. It reduces the event rate from 1 MHz to 2-5 kHz, and all such events are stored as raw data. An event accepted by L0, undergoes processing by the HLT, which merges data from different subdetectors. The HLT consists of two layers: HLT1 and HLT2. At HLT1, a partial event reconstruction is performed to confirm the L0 decision and identify objects of interest in the

event, as defined prior to data-taking in trigger lines. These trigger lines generally correspond to specific track configurations with defined p_T or geometric requirements. At HLT2, a full event reconstruction takes place, enabling a finer selection of the event topology. The trigger conditions for a specific data-taking run are specified with the trigger configuration key (TCK). A given TCK encodes the complete set of trigger lines, including the specific threshold values for trigger requirements and prescales that were active during data-taking. Each run has a unique TCK.

The offline software framework

The full LHCb software is employed for data generation and processing, ensuring the uniformity of LHCb tools and algorithms throughout the experiment and maintaining global consistency across all analyses. The key elements of the data processing are illustrated in Figure 3.24:

- **Event Generation and Simulation**: the Gauss package [123] handles the generation of simulated Monte Carlo (MC) events. Primary proton-proton interactions are simulated using PYTHIA [124], and decays of b- and c-hadrons are generated with the EvtGen package [125]. The Geant4 [126] toolkit is then employed to simulate the interaction of particles with the detector.
- **Digitisation**: the Boole application [127] emulates the LHCb detector response for the simulated event. The digitised output represents the MC-equivalent of the real data detector response.
- **Trigger Processing**: Both simulated and real data undergo processing through three stages of triggers (L0, HLT1, and HLT2), as explained in section 3.2.3, using the Moore framework [128].
- **Reconstruction**: the Brunel application [129] performs the full offline reconstruction, clustering signal hits in the detector. Particle properties, such as momentum and PID probability variables, are evaluated and stored in Data Storage Tape (DST) files.
- **Analysis**: the DaVinci software package [130] utilises the information stored in the DST files to assess various kinematic and topological variables while conducting the event selection. Quantities like the invariant masses of the decayed particles, their distance of flight, or the decay times are determined in this step. The application facilitates the production of output data files for subsequent physics analyses.

3.3 Fixed-target physics at LHCb

LHCb stands out as the only experiment at the LHC capable of acquiring data in both collider and fixed-target modes [132]. The fixed-target system at LHCb, known as SMOG (System for Measuring the Overlap with Gas) [133], was initially designed to enable accurate luminosity calibration for colliding proton beams. Through the SMOG device, noble gases such as helium,

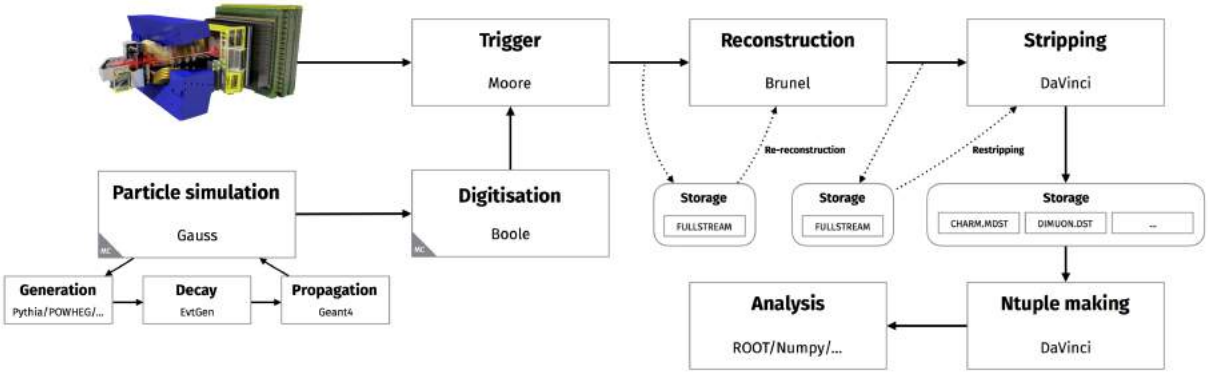


Figure 3.24: The data flow for LHCb Run 2, along with the corresponding applications for both real data and simulation [131].

neon, and argon can be introduced into the LHC vacuum pipe near the LHCb collision point at a nominal pressure of approximately 2×10^{-7} mbar. For a beam of 10^{14} protons, this corresponds to a luminosity of $6 \times 10^{29} cm^{-2} s^{-1}$ for collisions occurring within one meter of gas along the beam direction, roughly matching the acceptance of the LHCb vertex detector. The consequent rise in the beam-gas collision rate, increased by two orders of magnitude, facilitates a precise determination of the beam density profiles. This capability has been effectively utilised for highly accurate luminosity measurements employing the beam-gas imaging technique [134, 135].

The SMOG system, coupled with the excellent capabilities for particle reconstruction and identification in the forward direction of LHCb, introduces the unique opportunity to perform experiments in fixed-target mode. Fixed-target collisions, exploiting LHC proton and lead beams, allow for the study of particle production with a large momentum fraction of the target nucleon in the nucleon-nucleon centre-of-mass (c.m.) frame, at an energy scale of up to 115 GeV. Another advantage of the fixed-target configuration over the beam-beam setup is the broader selection of collision systems (Figure 3.25), offering new avenues for exploring the nucleon structure, along with measurements of significant interest to cosmic-ray physics and heavy-ion physics.

3.3.1 The SMOG system

The SMOG system [133] enables the controlled injection of specific gas species (He, Ne, and Ar), increasing the pressure in the interaction region from approximately 10^{-9} mbar to around $\sim 10^{-7}$ mbar. The key components of the SMOG gas injection system are summarised in Figure 3.26. The SMOG gas injection system, shown in Figure 3.27, was developed and commissioned with the purpose of significantly increasing the beam-gas collision rate in order to take full advantage of the beam-gas imaging capabilities for precision luminosity measurements. A turbo pump (TP 301) is connected via the gate valve GV302 to the VELO vessel. By keeping the valve closed, a gas with flow rate Q is injected into the pump, resulting in a pressure $p_0 = Q/S$, where S is the pumping speed of the TP 301, of about 500 l/s. When opening the valve, the VELO beam vacuum pressure evolves to the same value p_0. During the gas injection, the two

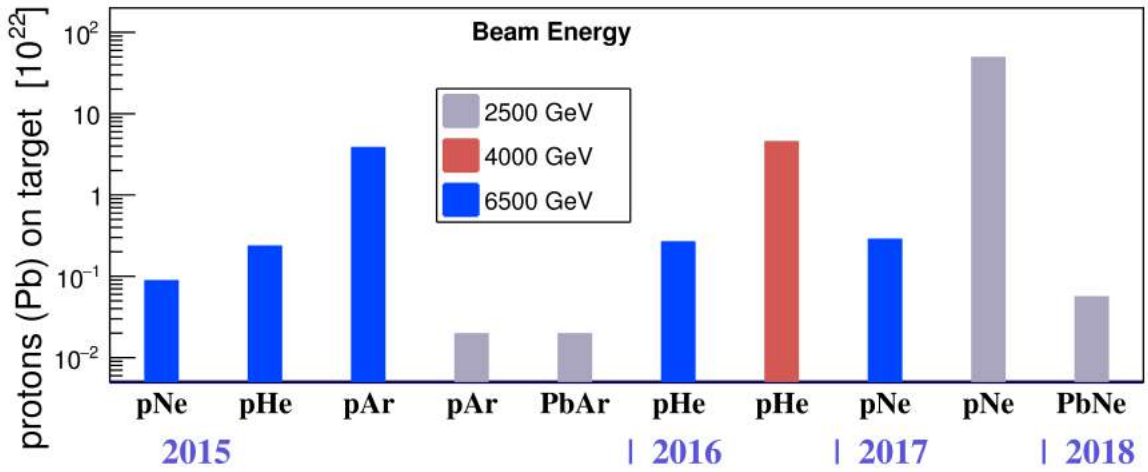


Figure 3.25: Dedicated SMOG runs collected since 2015. Beam-gas collisions have been recorded using different gas types (He, Ar, Ne) and beam energies [132].

VELO vacuum ion pumps located right under the interaction point are switched off and the VELO beam vacuum is opened to an additional volume pumped by a turbo-molecular pump previously stabilised at the nominal gas pressure. When gas injection is stopped and the ion pumps are switched back on, the nominal pressure of 10^{-9} mbar is recovered in a few minutes. The injected gas pressure can be monitored by four cold-cathode gauges (Penning type) and one hot-filament ionization gauge (Bayard-Alpert type) located at various positions around the VELO.

3.3.2 Physics opportunities and recent results

Fixed-target collisions involving LHC proton or lead beams with gas atoms open unique physics opportunities [136, 137]. In this setup, the beam-gas collision energy in the nucleon-nucleon centre-of-mass frame is given by

$$\sqrt{s_{NN}} = \sqrt{2E_N M_N c^2} \in [29, 115] \text{ GeV} \ , \ E_N \in [450, 7000] \text{ GeV}. \tag{3.7}$$

This represents an intermediate scale between experiments operating at the SPS and at the RHIC accelerators, providing an additional test bench to study the energy evolution in a region poorly constrained by experimental data. The corresponding rapidity of the centre-of-mass frame with respect to the laboratory ranges

$$y^* = \arcsin(\sqrt{E_N/2M_N c^2}) \in [3.8, 4.8] \ . \tag{3.8}$$

This implies that, given the LHCb instrumented region $\eta \in [2,5]$, $y^* \in [-2.8, 0.2]$ rapidity values in the centre-of-mass frame can be accessed at $E_N = 7$ TeV. As this is related to Feynman-x according

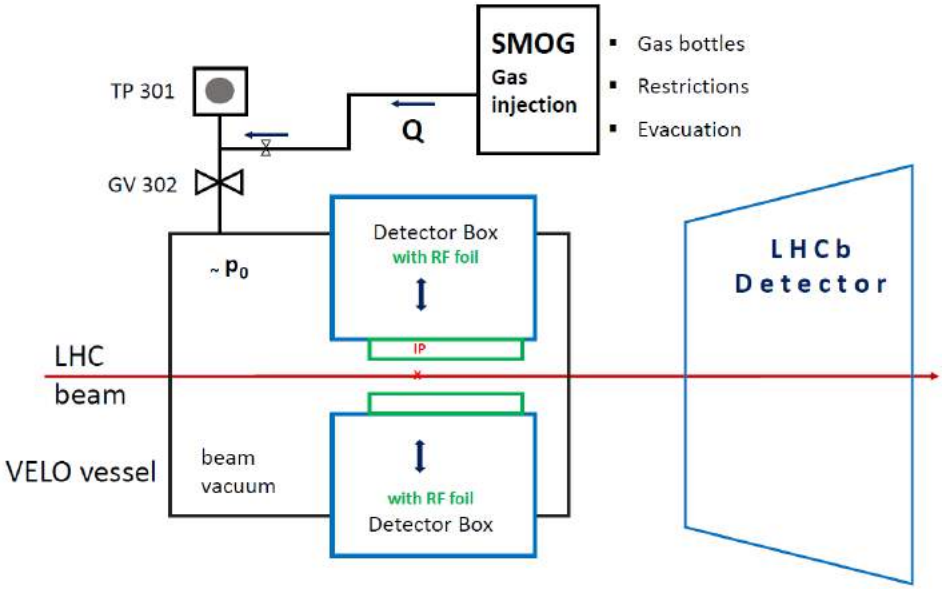


Figure 3.26: Scheme of the SMOG system setup: a gas feed system injects gas into a pump (TP 301). When the valve GV 302 is open, the VELO vessel is filled with gas at a low density, regulated by the injected flow rate [132]. Courtesy of E. Steffens.

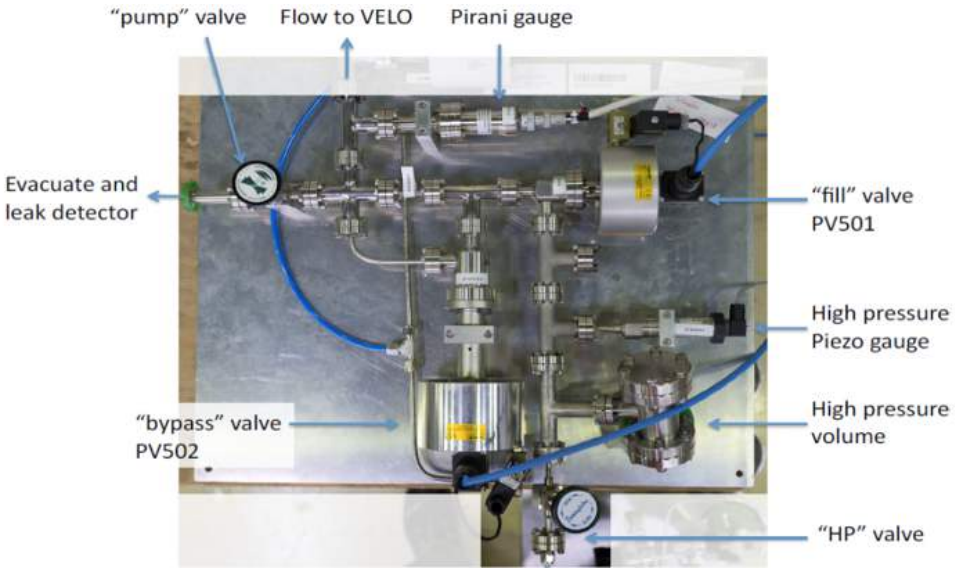


Figure 3.27: The SMOG gas feed system [132].

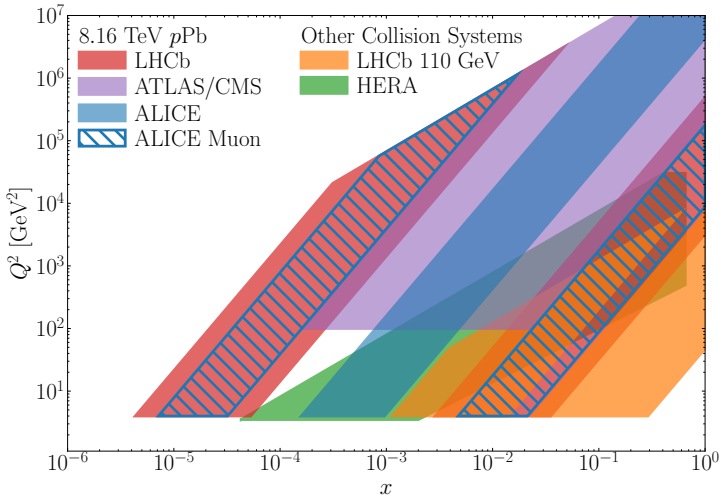


Figure 3.28: Accessible x-Bjorken regions for Q^2-target with LHCb in both its collider and fixed-target configurations, in comparison with other experiments [138].

to

$$x_F = \frac{2}{\sqrt{s_{NN}}} \sqrt{(Mc^2)^2 + (p_T c)^2}\ \sinh(y^*) \simeq x_1 - x_2\ , \tag{3.9}$$

SMOG allows the observation of particles originating from partons carrying a relatively large fraction of the target nucleon momentum (large Bjorken-x). The distinctive kinematic coverage of SMOG is evident in Figure 3.28, where it is compared with the fixed-target LHCb acceptance and those of other experiments.

Several results have been published from the analysis of fixed-target data collected during Run 2. LHCb previously released two papers focusing on the measurements of heavy-flavour and prompt antiproton production using SMOG [139, 140], establishing the feasibility of the project. Recently, three additional measurements based on SMOG data have been published. Charmonia production, involving $c\bar{c}$ bound states, as a crucial probe for studying various nuclear effects, including modifications arising from nPDFs, multiple scattering within the nucleus, nuclear absorption, and comover dissociation. The effects of cold nuclear matter (CNM) depend on collision energy, emphasising the importance of conducting measurements across a wide energy range to comprehend the underlying mechanisms. In a recent study utilising SMOG data recorded with a 2.5 TeV proton beam incident on neon nuclei at rest, LHCb investigated the production of charmonia and open-charm states, specifically J/ψ, $\psi(2S)$ and D^0 mesons, in pNe collisions at a centre-of-mass energy of $\sqrt{s_{NN}} = 68.5$ GeV [141]. In the context of nucleus–nucleus collisions, the production of heavy quarks offers insights into the transition from ordinary hadronic matter to the hot and dense QGP. Due to colour screening within the QGP, charmonium-bound states experience additional suppression in the presence of QGP compared to normal nuclear matter. Using a SMOG dataset from PbNe collisions, LHCb measured the production of heavy-flavour hadrons, including J/ψ and D_0 mesons, at $\sqrt{s_{NN}} = 68.5$ GeV. This marks the first measurement of nucleus–nucleus collisions in the fixed-target mode at LHCb [142]. The SMOG program also contributes valuable insights to cosmic-ray physics. A recent

measurement focused on the production of antiprotons originating from antihyperons, in proton–helium collisions at $\sqrt{s_{NN}} = 110$ GeV. This study extends the earlier SMOG measurement of prompt $\bar{p}$ production in the same collisions [143]. These measurements provide crucial information to reduce the uncertainties in the present model predictions of $\bar{p}$ flux in cosmic rays, addressing theoretical uncertainties in the production cross section of $\bar{p}$ in collisions between cosmic rays and the interstellar medium, ultimately enhancing the interpretation of cosmic ray data from space-borne experiments.

Chapter 4

Study of cold nuclear matter effects in PbNe and pNe collisions at $\sqrt{s_{NN}} = 69$ GeV

The study on cold nuclear matter effects in PbNe and pNe collisions at $\sqrt{s_{NN}} = 69$ GeV is presented in this chapter. The author led the entire data analysis process, covering the formulation of the analysis strategy (Section 4.1), global event and candidate selection (Section 4.3), evaluation of selection efficiencies from simulated samples (Section 4.5), and the development and implementation of a data-driven approach for modelling Particle Identification (PID) in fixed-target data using machine-learning techniques (Section 4.6), determination of $\pi^{\pm}, K^{\pm}, {}^{(-)}p$ yields through fitting collision data (Section 4.7), estimation of main systematic uncertainties (Section 4.8), and evaluation of single and double ratios of particle production with a subsequent comparison of these results with simulations (Section 4.9).

4.1 Analysis strategy

The objective of the analysis is to measure the inclusive production of prompt charged particles, specifically $\pi^{\pm}, K^{\pm}, {}^{(-)}p$, in fixed-target collisions involving proton-nucleus and nucleus-nucleus collisions at the LHCb experiment. This study is relevant because the comparison of light hadron yield distributions between light and heavy collision systems, with respect to pseudorapidity (η), transverse momentum (p_T), and collision centrality, may reveal modifications indicative of CNM effects.

Specifically, the analysis utilises the PbNe and pNe datasets collected at a centre-of-mass energy per nucleon pair of $\sqrt{s_{NN}} = 69$ GeV by the LHCb experiment with the SMOG system. The processes of interest are of the type:

$$\text{PbNe} \to h + X,$$
$$p\text{Ne} \to h + X,$$

where $h = \pi^{\pm}, K^{\pm}, \overset{(-)}{p}$, and X represents any other particle produced in the collision. At the present time, the luminosity has only been determined for the pNe sample and is unavailable for the PbNe dataset. This lack of information prevents the possibility to measure absolute cross-sections for the latter sample. To overcome this issue, the main goal of the analysis is to measure ratios of candidates yields, as the luminosity cancels out in the ratio. In particular, it is of interest to investigate particle production and various combinations of particle ratios as functions of η, p_T, and event multiplicity for both collision systems. Given the close correspondence between rapidity and pseudorapidity for light particles, especially notable for pions and kaons, it has been decided to utilise pseudorapidity for all particle types.

The relevant observables encompass:

- 1D distributions of π, K, p yields corrected by the efficiencies, for each of the two collision systems:

$$(N^{\pi}/\varepsilon^{\pi})_{\text{PbNe}}, \quad (N^{K}/\varepsilon^{K})_{\text{PbNe}}, \quad (N^{p}/\varepsilon^{p})_{\text{PbNe}} , \tag{4.1}$$

$$(N^{\pi}/\varepsilon^{\pi})_{\text{pNe}}, \quad (N^{K}/\varepsilon^{K})_{\text{pNe}}, \quad (N^{p}/\varepsilon^{p})_{\text{pNe}} ; \tag{4.2}$$

- 1D and 2D single ratios of particle production, corrected by the efficiencies, for each of the two collision systems:

$$\left(\frac{N^{p}/\varepsilon^{p}}{N^{\pi}/\varepsilon^{\pi}}\right)_{\text{PbNe}}, \quad \left(\frac{N^{K}/\varepsilon^{K}}{N^{\pi}/\varepsilon^{\pi}}\right)_{\text{PbNe}} , \tag{4.3}$$

$$\left(\frac{N^{p}/\varepsilon^{p}}{N^{\pi}/\varepsilon^{\pi}}\right)_{\text{pNe}}, \quad \left(\frac{N^{K}/\varepsilon^{K}}{N^{\pi}/\varepsilon^{\pi}}\right)_{\text{pNe}} ; \tag{4.4}$$

- double ratios of particle production in 1D and 2D, corrected by the efficiencies:

$$\frac{\left(\frac{N^{p}/\varepsilon^{p}}{N^{\pi}/\varepsilon^{\pi}}\right)_{\text{PbNe}}}{\left(\frac{N^{p}/\varepsilon^{p}}{N^{\pi}/\varepsilon^{\pi}}\right)_{\text{pNe}}}, \quad \frac{\left(\frac{N^{K}/\varepsilon^{K}}{N^{\pi}/\varepsilon^{\pi}}\right)_{\text{PbNe}}}{\left(\frac{N^{K}/\varepsilon^{K}}{N^{\pi}/\varepsilon^{\pi}}\right)_{\text{pNe}}} . \tag{4.5}$$

Section 4.2 delineates the datasets collected and analysed in this study. The analysis focuses on events triggered under Minimum Bias (MB) conditions in a beam-empty Bunch Crossing (BC) configuration, where the event is triggered by the bunch arriving from upstream of the detector colliding with the gas target injected into the beam pipe. Additionally, simulated samples are generated to assess selection efficiencies and examine the prompt condition selection. Section 4.3 details the requirements applied to both data and simulation for selecting charged tracks, resulting in a sample consisting of a mixture of pions, kaons, and protons. Following candidate selection, the simulation is validated with a data-driven procedure through a reweighing technique, particularly crucial for the multiplicity variables. This validation procedure is presented in Section 4.4.

After candidate selection, both data and simulation datasets are divided into intervals (bins), with binning strategies outlined in Section 4.1.1. Efficiency calculations on simulation are per-

formed after selection and reweighing, considering each bin and each hadron type independently. These calculations, reported in Section 4.5 account for potential variations between the different hadron species and incorporate effects of bin migration and smearing in the reconstruction process.

Particle Identification (PID) is a critical aspect of the analysis. At the selection level, no PID requirement is imposed on candidate tracks. Instead, the number of pions, kaons, and protons is determined through a bi-dimensional fit in the PID plane, utilising a fully data-driven method based on a Neural Network (NN) Gaussian Mixture Model discussed in Section 4.6. The shape of each template used for fitting the data is determined by the NN's prediction for pions, kaons, and protons. Additionally, a template modelling the ghost tracks is included, sourced from simulations. The signal yields of pions, kaons, and protons are obtained independently in each bin. Separate distributions are obtained for each collision system, considering both positive and negative tracks. Further details on this procedure are provided in Section 4.7. The results are presented in Section 4.9.

4.1.1 Binning definition

The determination of the binning for measuring candidates yields and correction factors involves careful consideration of various aspects. While a finer binning approach might be preferable for capturing small CNM effects, practical constraints must be considered. Key factors, such as detector resolution and the size of data and simulation samples, significantly impact the relative uncertainties associated with each bin measurement. In this case, the constraints on the bin size arise from the limited statistics of the simulation and the calibration samples used to produce the PID templates. Several binning schemes have been considered, including both 1D schemes on η, p_T and multiplicity, and some 2D combinations (η-p_T and p_T-multiplicity). The strategy for 1D binning is to define a set of bins with the same number of events. The 2D binning is defined simply by considering combinations between 1D intervals.

Regarding the binning variable for multiplicity, various options are possible. In theory, any observable that scales monotonically with the impact parameter could be used for a classification based on collision centrality. Centrality is related to the volume of the medium formed by the colliding nuclei, and measures the overlap region between the two nuclei in a collision. However, the range in centrality achievable with the LHCb tracking and particle identification systems is limited by their performance at high track multiplicities. The energy deposited in the ECAL is utilised for multiplicity determination in both PbPb and PbNe collisions. Centrality classes are defined as quantiles of the inelastic PbNe cross-section [144]. Figure 4.1 illustrates the categorisation of events in PbNe data based on specified centrality classes.

For this reason, yields from PbNe are computed as a function of the number of hits in the SPD detector (nSPDHits), allowing the establishment of a mapping between event multiplicity and collision centrality information. This mapping is crucial for studying the centrality dependence of the ratio of charged-particle production. The results of the mapping are summarised in Table 4.1, which presents the mean values and collision centrality percentiles along with the

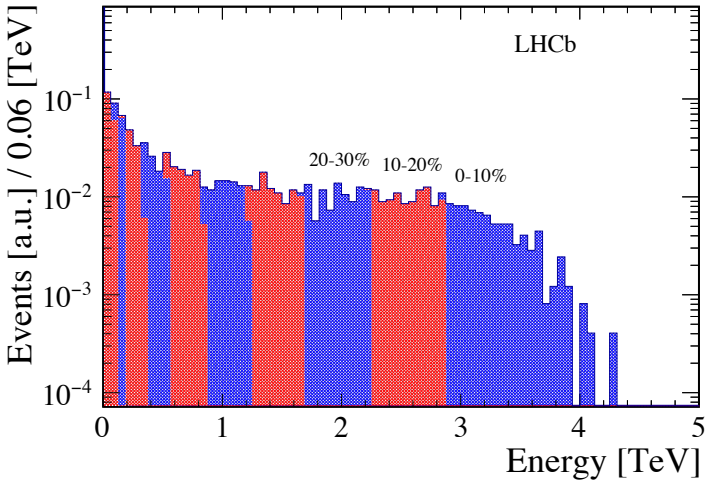


Figure 4.1: Distribution of the energy deposited in the ECAL which is utilised for multiplicity classification of events from PbNe data according to the defined collision centrality classes [144].

nSPDHits	Centrality (%)	Mean centrality (%)	rms centrality (%)
[0,300]	[88,71]	79.87	8.86
[300,446]	[72,62]	66.88	5.21
[446,715]	[61,51]	55.89	5.04
[715,1000]	[47,39]	42.90	4.25

Table 4.1: Centrality quantities corresponding to the nSPDHits bins used in the analysis for PbNe data. Lower centrality intervals are excluded to mitigate contamination from Ultra Peripheral Collisions (UPC), which predominate at low energies. By implementing a selection criterion for energy deposition in the ECAL exceeding 98.9 GeV, UPC contamination is reduced to less than 5% [144]. Higher centrality intervals, characterised by nSPDHits > 1000, are excluded due to sub-optimal performance in charged-hadron identification, as elaborated in Section 4.3.2.

corresponding RMS values, for intervals of nSPDHits.

The chosen range for the intervals is as follows:

- 4 bins in η: [2.4,3.7], [3.7,4.0], [4.0,4.2], [4.2,4.5];
- 8 bins in p_T: [600,676], [676,755], [755,839], [839,935], [935,1051], [1051,1206], [1206,1459], [1459,3000], expressed in MeV/c;
- 4 bins in nSPDHits: [0,300], [300,446], [446,715], [715,1000].

Given that CNM effects are prominently observed in the modification of the p_T distribution of the produced light hadrons, a finer binning is adopted for this variable. Moreover, the p_T binning is designed to accommodate the exponential decay of the distribution, with a slight increment in bin size as p_T increases. The first bin in nSPDHits covers the whole multiplicity range for the pNe dataset. The choice of the binning range for all three variables is detailed in Section 4.3.

4.2 Datasets

The analysis employs two primary datasets corresponding to PbNe and pNe, both gathered at a collision energy of $\sqrt{s_{NN}} = 69$ GeV. These datasets are described in Sections 4.2.1 and 4.2.2. The corresponding simulated samples are presented in Section 4.2.3.

4.2.1 PbNe dataset

The PbNe data sample corresponds to the data recorded throughout the entire 2018 heavy-ion data-taking period when gas was injected. This period amounts to 214 hours of recorded data. PbNe collisions occur at a centre-of-mass energy per nucleon of $\sqrt{s_{NN}} = 69$ GeV, where the Ne atoms act as a fixed target. To prevent contamination from simultaneously recorded PbPb collisions, only events from bunch crossings where a filled bunch from the incoming Pb beam crosses an empty bunch of the outgoing Pb beam are selected. Since there is some residual contamination from beam-beam collisions in the data sample, additional selections, as described in Section 4.3, are applied to enhance the purity of PbNe sample. An overview of fill numbers and the corresponding list of runs labelled with good data quality for PbNe data is presented in Table 4.2.

Given that the relevant event for the analysis is the inclusive production of charged hadrons, no particular trigger line is needed; therefore, a MB trigger is applied (Hlt1BEMicroBiasVelo). During data taking, no scaling of the HLT1 line was necessary, and no additional HLT2 selection is applied. Events are selected with the MB stripping line (StrippingMBMicroBias), which simply requires the trigger condition and at least one reconstructed primary vertex. In this case, a prescale factor of 0.05 is applied the stripping line.

4.2.2 pNe dataset

The pNe dataset corresponds to the data recorded in the "magnet down" configuration, at a centre-of-mass energy per colliding nucleon-nucleon pair of $\sqrt{s_{NN}} = 69$ GeV. Data collection for pNe occurred between November 11^{th} and November 22^{nd}, 2017, spanning approximately 170 hours. An overview of fill numbers and the corresponding list of runs labelled with good data quality for pNe data is provided in Table 4.3. It is noteworthy that standard proton-proton collisions at $\sqrt{s} = 5$ TeV were concurrently recorded during this data-taking period, and a dedicated study on the background originating from these collisions is presented in Section 4.3.

Similarly to PbNe data, a MB trigger is implied which requires at least one reconstructed VELO track (Hlt1BEMicroBiasVelo). Since the bandwidth is shared with pp collisions, and due to the high beam intensity, an unprescaled MB trigger is not feasible. Therefore, a prescale of 0.25 is applied to the line. Additionally, a random trigger (Lumi Trigger) is employed to control the L0 trigger efficiency. No HLT2 selection is applied.

During offline processing, the data is selected using a specific MB stripping line, denoted as StrippingMBMicroBiasVelo. Also in this case, the stripping line requires only the trigger condition and at least one reconstructed primary vertex. The stripping has a postscale of 0.50,

Fill	Good run list
7438	217949, 217951, 217952, 217953, 217954, 217955, 217956
7439	217971, 217973, 217974, 217976, 217977, 217978, 217979, 217980, 217983
7440	218007, 218009, 218010, 218011, 218012, 218013, 218014, 218016, 218017
7441	218032, 218033, 218035, 218036, 218037, 218038, 218039, 218040, 218041, 218043, 218044, 218045
7442	218067, 218068
7443	218098, 218099, 218109, 218111
7444	218130, 218131, 218132, 218133, 218134, 218135, 218136, 218137
7446	218166, 218173, 218174, 218175, 218176, 218177, 218178
7448	218204, 218205, 218206, 218207, 218208, 218209, 218210, 218211
7449	218230, 218231, 218232, 218233, 218234, 218235, 218236, 218237
7450	218251, 218253, 218254, 218255, 218256, 218257, 218258
7453	218311, 218312, 218313, 218314, 218315, 218316, 218317, 218318, 218319, 218320, 218325, 218326
7454	218343, 218344, 218349, 218350, 218351, 218352, 218353, 218354, 218355, 218357, 218359, 218371
7456	218402, 218403, 218407, 218408, 218409, 218410, 218411, 218412
7457	218428, 218429, 218430, 218431, 218432, 218433, 218434
7460	218472, 218473
7466	218558, 218559, 218560, 218561, 218562, 218563, 218564, 218565, 218567, 218568
7467	218584, 218585, 218586, 218587, 218588, 218590, 218591, 218592, 218596, 218598, 218599
7468	218620, 218621, 218622, 218623, 218624, 218625, 218626, 218627, 218628, 218629, 218631
7471	218660, 218661, 218664, 218665, 218666, 218667, 218668, 218669
7472	218685, 218686, 218687, 218688, 218689, 218690, 218691, 218692, 218693
7473	218704, 218705, 218706, 218707, 218708, 218709, 218710, 218711, 218712, 218714
7475	218775
7477	218791, 218792, 218793, 218794, 218795, 218796, 218799, 218800, 218801
7480	218843, 218844, 218845, 218846, 218847, 218848, 218849, 218850, 218851
7483	218905, 218906, 218907, 218908, 218909, 218910, 218913, 218914, 218915, 218917, 218918
7485	218938, 218939, 218940, 218941, 218942, 218943, 218944, 218945, 218947
7486	218962, 218965, 218966, 218967, 218968, 218969, 218971, 218974
7487	218986, 218987, 218989, 218990, 218991, 218992, 218993, 218994, 218996
7488	219019, 219021, 219022, 219023, 219024, 219025, 219028, 219029, 219030
7489	219042
7490	219055, 219056, 219058, 219059, 219060, 219062, 219063, 219064, 219065, 219066, 219067
7491	219081, 219083, 219084, 219085

Table 4.2: Summary of fill numbers and the corresponding list of runs marked with good data quality for PbNe data.

implying that only 50% of the events meeting the conditions are retained. The scaling factors of the trigger and stripping cancel out in the calculations of the ratio but need to be accounted for in the calculation of the cross-section for direct particle production comparison between different datasets.

4.2.3 Simulated samples

The analysis chain includes the computation of efficiencies by estimating samples with fully simulated events using the standard LHCb simulation software. These simulated events are then reconstructed and analysed using the same software tools used for the real data. Since the generator used to simulate pp data at LHCb, Pythia [145], does not support proton-nucleus interaction, EPOS-LHC [78], specialised in soft hadronic production and nuclear effects, is used instead.

The other simulation steps are common to pp data: the particle decays are described by EvtGen [146], and their interactions with the detector material are modelled using the Geant4 toolkit [125, 147]. The gas pressure evolution with the z-coordinate, and consequently the PbNe and pNe collision vertices, is assumed to be uniform in $z \in [-1000, 300]$ mm, which is a reasonable approximation considering that the gas is extracted by two pumps located at $z = \pm 20$ m. Two Monte Carlo (MC) simulation samples are used in this analysis:

- **MC PbNe**: A minimum-bias sample with 11 million PbNe simulated collisions is generated;
- **MC pNe**: A minimum-bias sample with 11 million pNe simulated collisions is generated.

Fill	Good run list
6384	202217, 202218, 202219, 202220, 202221, 202222, 202223
6385	202234, 202238, 202239, 202240, 202242, 202244, 202245, 202246, 202247, 202248, 202249, 202250, 202251, 202252, 202253, 202254, 202255, 202256, 202257, 202258, 202259, 202260, 202261, 202262, 202263, 202264, 202265, 202266, 202267, 202268
6386	202280, 202283, 202284, 202285, 202286, 202287, 202288, 202289, 202290, 202291, 202292
6389	202313, 202316, 202317, 202318, 202319, 202320, 202321, 202322, 202323
6390	202339, 202341, 202342, 202343, 202345, 202346, 202347
6392	202360, 202363, 202364, 202365, 202366, 202367, 202368, 202369, 202370, 202372, 202373, 202374
6396	202406, 202407, 202409, 202411, 202412, 202413, 202414, 202415, 202416, 202417, 202418, 202419, 202420, 202421, 202425, 202426, 202428, 202429, 202430, 202431, 202432, 202433, 202434, 202435, 202437, 202438, 202439, 202441, 202442, 202449, 202450, 202452, 202442, 202449, 202450, 202452, 202454, 202455, 202456, 202457, 202458, 202459, 202460, 202461, 202464, 202465, 202466, 202467, 202468, 202469, 202470, 202471, 202472, 202473, 202474
6397	202485, 202486, 202490, 202495, 202496, 202497, 202500, 202503, 202504, 202506, 202508, 202510, 202511, 202514, 202515, 202517, 202518, 202519, 202521, 202522, 202524, 202525, 202527, 202529, 202531, 202534, 202536, 202538, 202539, 202540, 202541, 202542, 202543, 202544, 202545, 202547, 202550
6398	202602, 202608, 202610, 202612, 202614, 202615, 202616, 202617, 202620, 202624, 202625, 202626, 202627, 202628, 202630
6399	202640, 202642, 202643, 202644, 202645, 202646, 202647, 202648, 202649, 202650, 202651, 202652, 202653, 202654, 202655, 202656, 202657, 202658, 202660

Table 4.3: Summary of fill numbers and the corresponding list of runs marked with good data quality for *p*Ne data.

4.3 Event and candidate selection

This section outlines the strategy used to select $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ candidates from fixed-target events while minimising background contamination. Initially, a global event selection is implemented to select fixed-target events and reject background contamination resulting from beam debunching, as detailed in Section 4.3.1. Following this, a selection on track quality and kinematic regions is applied to improve the purity and ensure to match the kinematic coverage of the RICH detectors, which is crucial to ensure good particle identification, as described in Section 4.3. Subsequently, a cut on the track χ^2_{IP} is employed to select prompt candidates. Finally, a summary of all the selections is presented in Section 4.3.6.

4.3.1 Global Event Selection

The global event selection is implemented to achieve several objectives:

- select fixed-target events by requiring the correct event BC type (BCTYPE=1);
- choose events with at least one reconstructed primary vertex in a specified region of z;
- remove background contamination arising from misreconstructed primary vertices. This is accomplished by setting conditions based on the number of tracks associated with the primary vertex. Specifically, for pNe events, the condition is PVNTRACKS[0] > 4, and for PbNe events, it is PVNTRACKS[0] > 13. An additional requirement to reject fake primary vertices consist in ensuring that the best primary vertex is the one with the highest track multiplicity of the event, i.e., PVNTRACKS[0] > PVNTRACKS[i] for $i > 0$.
- suppress background from beam debunching, a significant source in the analysis.

For the last point, considering that the PbNe and pNe datasets exhibit on average different event multiplicities, two distinct strategies are employed, as described in the following sections.

PbNe dataset

The 2018 PbNe data were collected concurrently with PbPb collisions using 2.5 TeV/nucleon Pb beams. SMOG data samples are collected under specific Beam Crossing Type Conditions (BCType=1), where fully-filled Beam 1 bunches cross empty Beam 2 bunches at the Interaction Point (IP8). In such conditions, it is not expected to have simultaneous $\sqrt{s_{NN}} = 5$ TeV PbPb collisions along with PbNe collisions. However, due to beam ghost-charge contamination resulting from Pb ions debunching, some PbPb collisions may occur even for BCType=1, thereby contaminating the SMOG sample.

The unique topology of the Pb-gas fixed-target events allows to examine and control the purity of the sample using "backward" information, such as nPUHits (the number of hits in the VELO Pile Up (PU) stations located between $z = -315$ and $z = -220$ mm) or nBackTracks (the number of VELO tracks going backward). Unlike "collider" events, SMOG events are expected to exhibit little or no backward activity (small nPUHits and small nBackTracks).

PVZ [mm]	nPUHits	nBackTracks
$[-700,-200]$	No cut	< 50
$[-200,-100]$	< 10	No cut
$[-100,100]$	< 10	No cut
$[100,200]$	< 10	No cut

Table 4.4: Fiducial global-event cuts applied in different PVZ regions for PbNe data.

Figure 4.2 illustrates the effects of the various applied cuts on relevant 2D distributions between the PVZ and other variables such the distance between the PV and the beam line, nPUHits and nBackTracks. When selecting SMOG events only with the BCTYPE=1 condition a large background is observed, primarily due to the presence of fake reconstructed primary vertices. This is expected due to the high track multiplicity in PbNe collisions. This background can be suppressed by requiring that the minimum number of tracks in the best reconstructed primary vertex is lower than 13 (PVNTRACKS[0] < 13) and that the best-reconstructed PV has the highest track multiplicity (PVNTRACKS[0] $>$ PVNTRACKS[i], $i > 0$). The fraction of fake PVs is significantly reduced after applying these requirements.

The remaining background originates from PbPb collisions occurring in the central PVZ region. This contribution can be suppressed by applying a selection on the backward activity in the detector, i.e., by imposing constraints on nPUHits and nBackTracks. For the PbNe dataset, it is not possible to perform an optimisation on the selection since a pure sample of fixed-target events is not available. Similarly to what has been done in a previous analysis [148], the strategy is based on fiducial cuts applied in different PVZ regions. Table 4.4 provides a summary of the fiducial global-event cuts.

pNe dataset

The 2017 pNe data were collected concurrently with high-intensity pp collisions using 2.5 TeV proton beams. SMOG data samples were collected under specific conditions with BCType=1, where fully-filled Beam 1 bunches intersected with empty Beam 2 bunches at the interaction point. Under these conditions, no 5 TeV proton-proton collisions were expected to occur simultaneously with pNe collisions. However, due to ghost charges, some collider events may occur, leading to contamination of the SMOG sample. Similar to the description in the previous section for PbNe, a selection based on nPUHits and nBackTracks has been employed to suppress this particular background. For pNe, a more refined study is conducted, benefiting from the availability of background and control samples.

A background sample from pp collisions is obtained from a dedicated data stream, selected by requiring the trigger lines corresponding to the signal Hlt1BBSMOGSingleTrack. The criteria for these lines are identical to those used for selecting SMOG candidates, with the distinction that the trigger is applied to beam-beam crossing type events. A signal sample, referred to as the "clean train," is derived as a subset of the pNe dataset. Depending on the filling scheme of the machine, specific bunch IDs can be selected to minimise ghost-charge contamination.

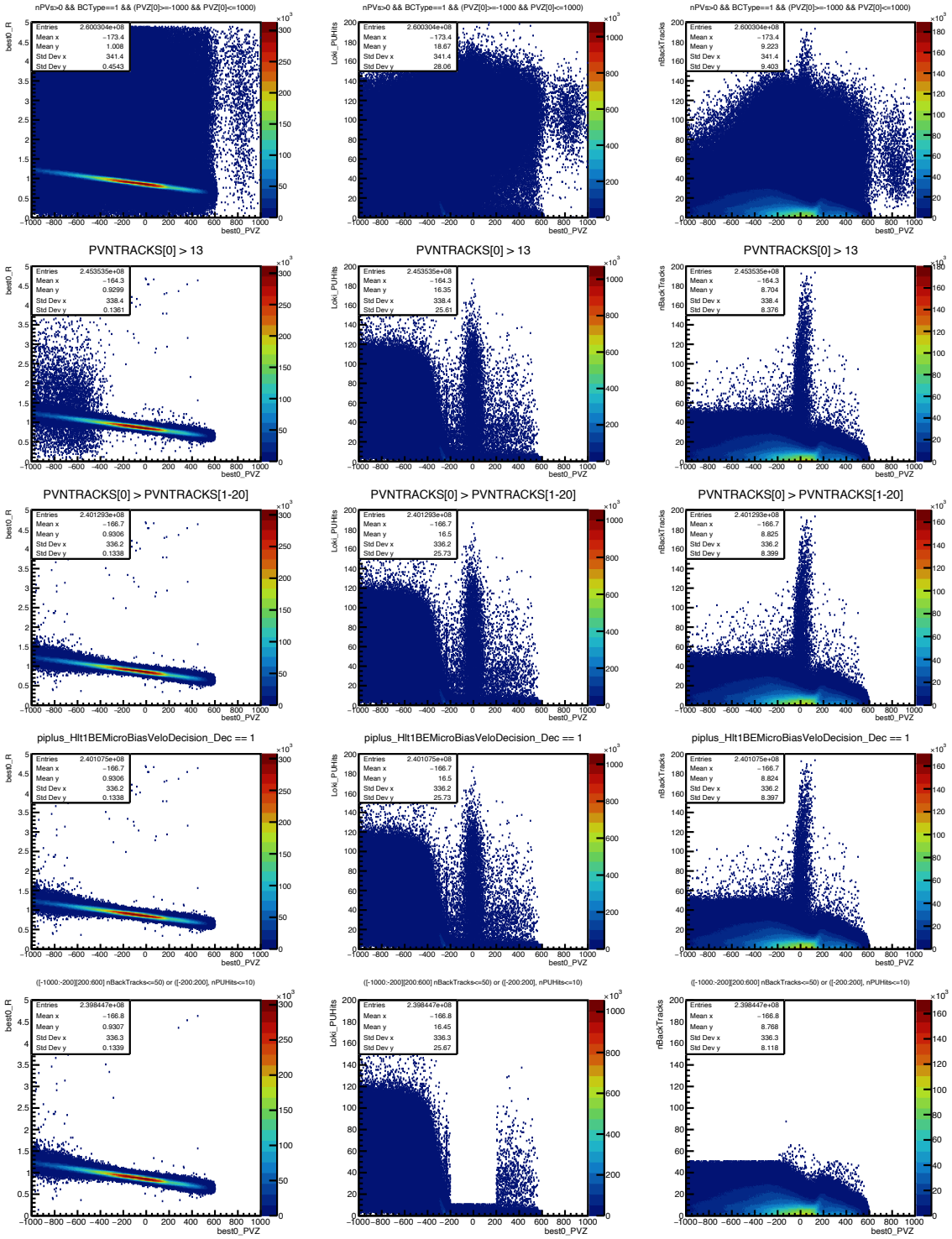


Figure 4.2: Effects of different global-event selection on different 2D distributions such as PVZ (cm) vs distance of the PV (com) from the beamline (left column), PVZ (cm) vs. nPUHits (middle column), and PVZ vs. nBackTracks (right column). Each row corresponds to different selections applied subsequently: the first row is obtained by requiring only BCTYPE=1; the second row requires PVNTRACKS[0] > 13; the third row requires that the best-reconstructed PV has the highest track multiplicity; the fourth row requires the MB trigger and stripping conditions; in the last row, the backward selection is applied to remove ghost-charge contamination.

In Figure 4.3, the fraction of reconstructed PVs within the luminous region $-100 < z < 100$ mm is presented based on the bunch ID. While a significant fraction of PVs is reconstructed within the luminous region when triggering on beam-beam crossing type, a small fraction of PVs is reconstructed when triggering on beam-empty crossing type. The selected bunch IDs, depicted as green points in Figure 4.3, are also listed in Table 4.5.

Figure 4.4 illustrates the distribution of the PVZ position for the entire SMOG dataset obtained by triggering on beam-empty events, the pp background-enriched sample acquired during beam-beam crossings triggering, and the signal-enriched sample (clean trains). The PVZ positions in the clean train sample span a broad range from -1000 to 600 mm, while the interaction points of pp background-enriched sample are concentrated within ± 100 mm. In the data, a distinct peak at ± 100 mm is evident, indicating significant contamination in this region. This peak is absent in the signal-enriched dataset of clean trains, where the contamination from pp collisions is minimal.

Figure 4.5 illustrates for the entire SMOG dataset, the pp background-enriched samples and the signal-enriched (clean trains) the 2D distributions of nPUHits and nBackTracks against PVZ. While, in the data, and especially in the signal-enriched sample, mostly one PV per event is reconstructed, the background-enriched sample exhibits multiple PVs per event. To align with the characteristics of the SMOG dataset, the number of reconstructed PVs is constrained to 1 for the pp background-enriched sample in the subsequent analysis steps. Two distinct populations are observed:

- **SMOG pNe fixed-target events**: Primary Vertices extend over a broad PVZ range. When events occur in the VELO area, outgoing particles are produced in the forward direction with respect to the nominal interaction point. Consequently, a small number of PU hits (nPUHits) and a small number of back tracks (nBackTracks) are expected for these events. It's important to note that events can also occur upstream of the VELO at large negative PVZ values. In such cases, the VELO is hit from the back, leading to larger nPUHits values.

- **Ghost pp collision events**: These events exhibit large nPUHits and nBackTracks, with the PV position located around PVZ~ 0. These occurrences are likely due to the interaction of Beam 1 protons with the Beam 2 ghost charges, resulting in spurious 5 TeV pp collisions. This background constitutes the predominant contamination of the fixed-target event sample.

Filling scheme	Bunch ID selection
1292b (492 n.c.)	$1 \leq \text{BunchID} \leq 25$
1548b (702 n.c.)	$769 \leq \text{BunchID} \leq 915$ or $3239 \leq \text{BunchID} \leq 3399$
1836b (740 n.c.)	$832 \leq \text{BunchID} \leq 902$ or $923 \leq \text{BunchID} \leq 977$
1836b (742 n.c.)	$824 \leq \text{BunchID} \leq 915$ or $938 \leq \text{BunchID} \leq 969$

Table 4.5: Selection criteria for clean trains based on specific filling schemes and associated Bunch ID ranges. The number of non-colliding bunches (n.c.) is indicated for each filling scheme.

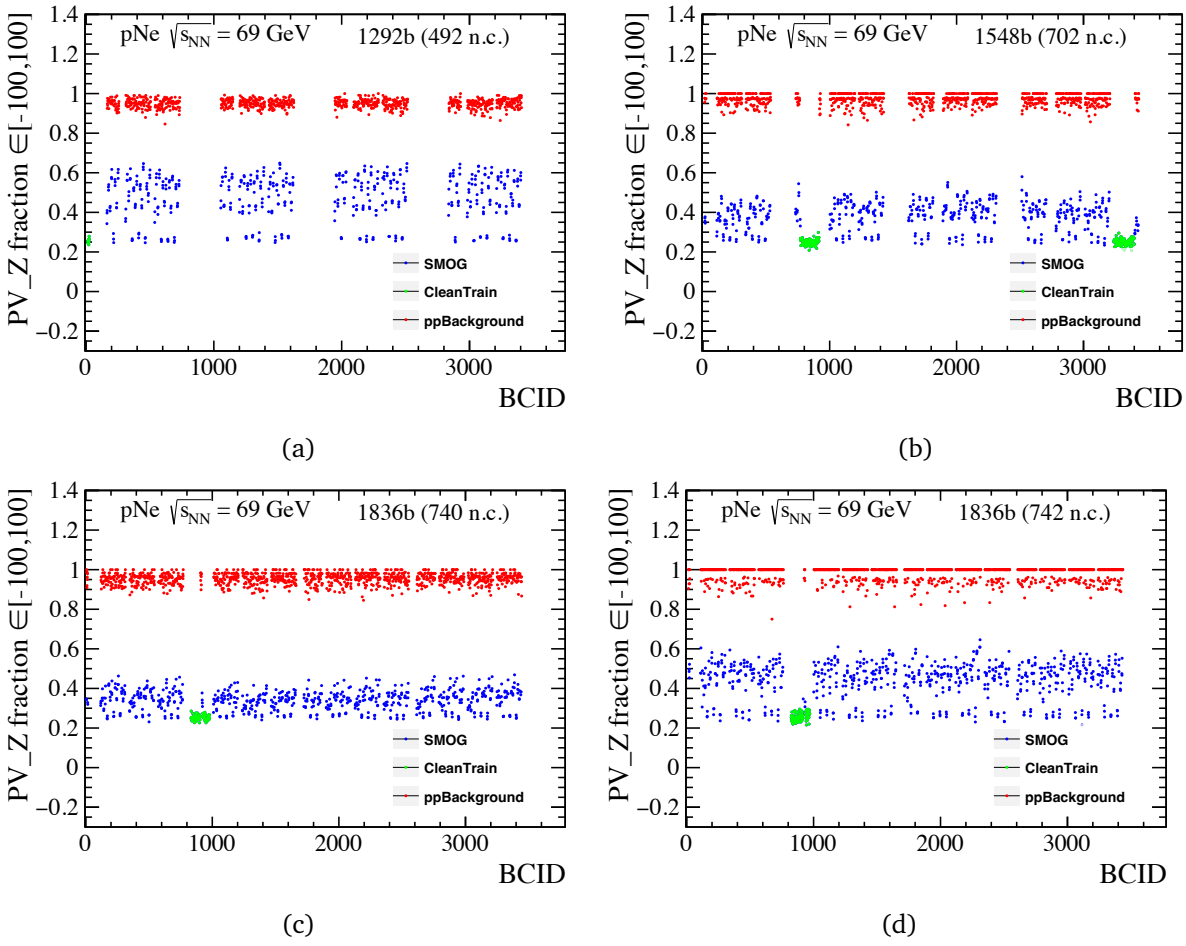


Figure 4.3: Fraction of reconstructed PVs inside the region $|z| < 100$ mm for (a) 1292b (492 n.c.) (b) 1548b (702 n.c.) (c) 1836b (740 n.c.) (d) 1836b (742 n.c.) filling schemes used during the data taking. The blue points represent the fraction when triggering on beam-empty crossing type, the green points represent the fraction for clean train events and the red points when triggering on beam-beam crossing type.

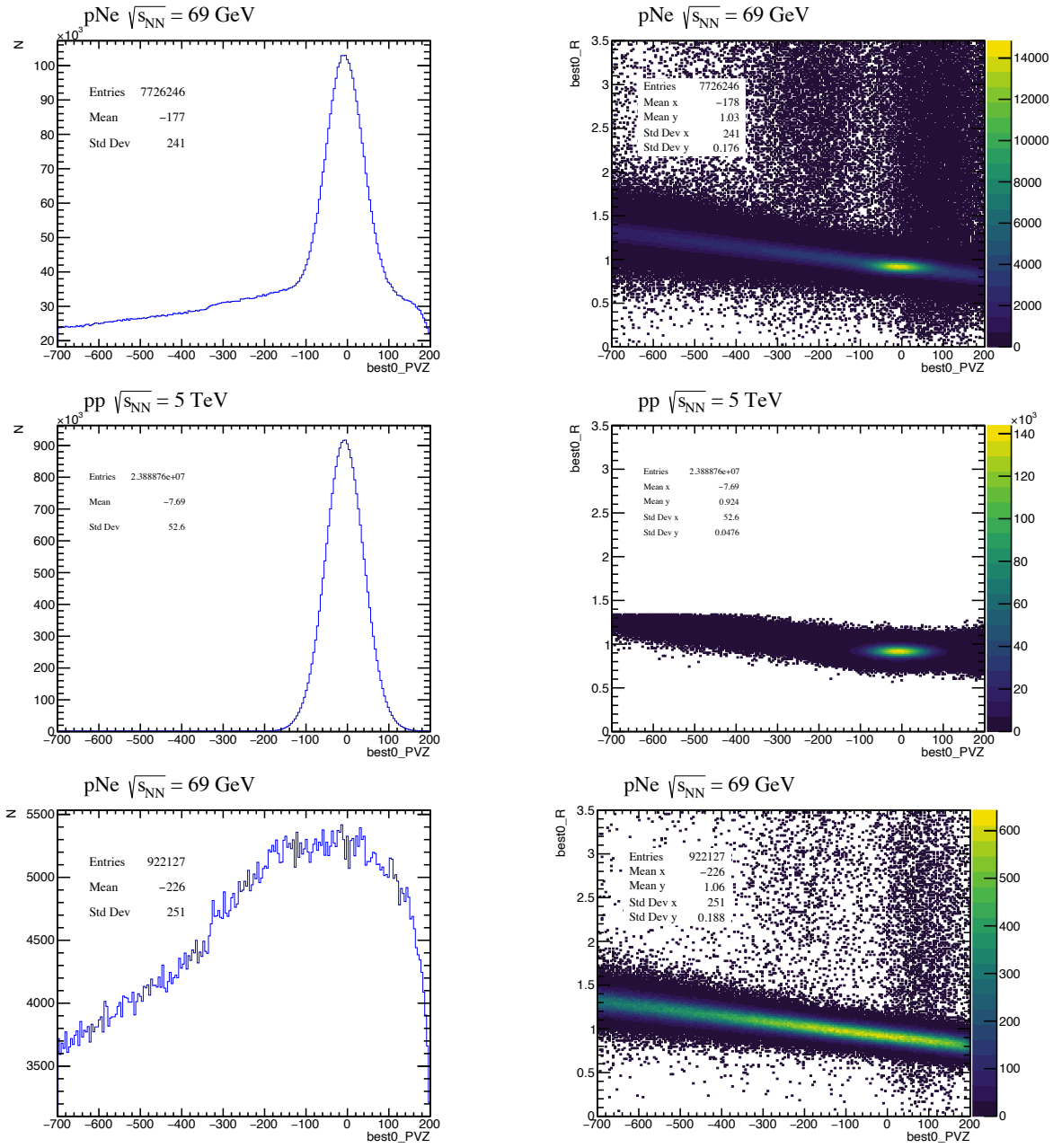


Figure 4.4: Distribution of the reconstructed PV (left) and correlation between radius and z positions of the reconstructed PV (right) for the entire SMOG dataset (top), the pp background-enriched sample (middle), and the clean train sample (bottom). The SMOG and clean train distributions are obtained by requiring the number of tracks used to reconstruct the PV to be greater than 4 and selecting the PV with the highest number of tracks among all the reconstructed PVs in the event.

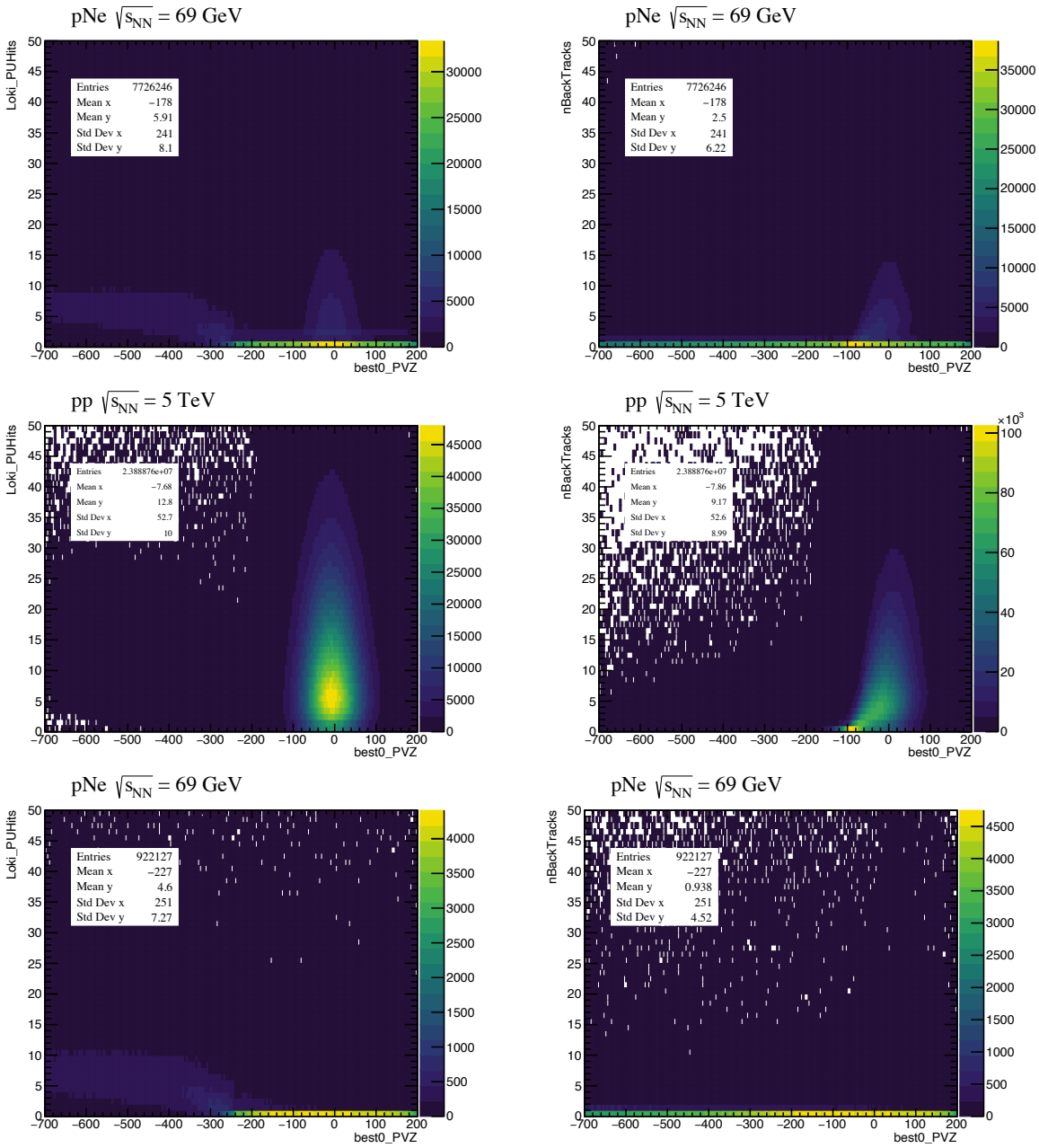


Figure 4.5: Distribution of number of hits in the Pile Up stations (nPUHits) versus PVZ (left) and distribution of the number of back tracks (nBackTracks) versus PVZ (right). The plots are reported for the data (top), pp background enriched (middle) and signal enriched clean trains (bottom) samples.

In order to define the optimal selection for the event variables nPUHits and nBackTracks, the following strategy is utilised. First, considering that the three datasets have different number of events, it is important to rescale the signal and background-enriched samples to data. In particular, two regions are identified:

- the signal region is defined as $100 < |PVz| < 200$ mm and nPUHits$= 0$ and nBackTracks$= 0$;
- the background region is defined as $|PVz| < 100$ mm and nPUHits> 10 and nBackTracks$>$ 10.

This allows to write the following relation

$$\begin{pmatrix} N_{CT}^{sig} & N_{pp}^{sig} \\ N_{CT}^{bkg} & N_{pp}^{bkg} \end{pmatrix} \begin{pmatrix} w_S \\ w_B \end{pmatrix} = \begin{pmatrix} N_{SMOG}^{sig} \\ N_{SMOG}^{bkg} \end{pmatrix} , \tag{4.6}$$

where w_S (w_B) is the scale factor for the signal (background) enriched sample, $N_{CT}^{sig}(N_{pp}^{sig})$ is the number of events for clean trains signal (pp background) enriched sample in the signal region, $N_{CT}^{bkg}(N_{pp}^{bkg})$ is the number of events for clean trains signal (pp background) enriched sample in the background region. Therefore it is possible to extract the corresponding weights by inverting the equation, as follows

$$\begin{pmatrix} w_S \\ w_B \end{pmatrix} = \frac{1}{N_{pp}^{bkg} N_{CT}^{sig} - N_{pp}^{sig} N_{CT}^{bkg}} \begin{pmatrix} N_{pp}^{bkg} & -N_{pp}^{sig} \\ N_{CT}^{bkg} & N_{CT}^{sig} \end{pmatrix} \begin{pmatrix} N_{SMOG}^{sig} \\ N_{SMOG}^{bkg} \end{pmatrix} \tag{4.7}$$

$$= \frac{1}{N_{pp}^{bkg} N_{CT}^{sig} - N_{pp}^{sig} N_{CT}^{bkg}} \begin{pmatrix} N_{pp}^{bkg} N_{SMOG}^{sig} - N_{pp}^{sig} N_{SMOG}^{sig} \\ N_{CT}^{bkg} N_{SMOG}^{sig} - N_{CT}^{sig} N_{SMOG}^{sig} \end{pmatrix} = \begin{pmatrix} 5.83 \\ 0.01 \end{pmatrix} , \tag{4.8}$$

where lower value of w_B is due to the higher statistics of the pp background enriched sample. To verify the calculation of the normalisation factors, the distributions of nPUHits and nBackTracks are plotted for both the pp background and clean train samples, each multiplied by their respective weights and compared with the entire SMOG dataset in Figure 4.6. The agreement between the two SMOG datasets and the weighed sum of clean and pp background distributions, demonstrates that the normalised sum of these components effectively captures the SMOG data, thereby validating the distribution shapes and affirming the precision of the calculated weights.

The yields are estimated for different nPUHits and nBackTracks thresholds i and j, where i and j are integer numbers, defined as

$$N_{sig}^{i,j} = w_S N_{CT}(\text{nPUHits} < i; \text{nBackTracks} < j) \tag{4.9}$$

$$N_{bkg}^{i,j} = w_B N_{pp}(\text{nPUHits} < i; \text{nBackTracks} < j) \, . \tag{4.10}$$

The Global Event Cuts (GEC) are optimised for different PVZ regions in terms of the significance, defined as

$$S(i,j) = \frac{N_{sig}^{i,j}}{\sqrt{N_{sig}^{i,j} + N_{bkg}^{i,j}}}, \tag{4.11}$$

PVZ[0]	nPUHits	nBackTracks	$\mathscr{S}$	$\varepsilon * \mathscr{P}(\%)$	$\varepsilon(\%)$	$\mathscr{P}(\%)$
$[-700,-200]$	No cut	No cut	1736.2	99.8	100.0	99.8
$[-200,-100]$	No cut	No cut	838.6	94.9	100.0	94.9
$[-100,100]$	<3	<2	1094.1	80.7	86.1	93.7
$[100,200]$	No cut	No cut	793.1	96.7	100.0	96.7

Table 4.6: Optimal GEC cuts for different PVZ regions for pNe data.

where N_{sig} and N_{bkg} represent the estimated number of SMOG and pp background events, respectively. Another Figure Of Merit (FOM), utilised to optimise the selection and to cross-check the results, is the product of the efficiency and the purity, defined as

$$\varepsilon_S \cdot P = \frac{N_{CT}^{i,j}}{N_{CT}^{tot}} \cdot \frac{N_{sig}^{i,j}}{N_{bkg}^{i,j}}. \tag{4.12}$$

In Figure 4.7, S and $\varepsilon_S \cdot P$ are displayed as a function of various cuts applied to different PVZ regions. The optimal cuts are determined by maximising both FOMs, and there is agreement between the two metrics regarding the optimal cut values. Notably, in the central region |PVZ| < 100 mm, a stringent selection on nPUHits and nBackTracks is necessary to mitigate pp contamination. Conversely, no cuts are imposed for the region |PVZ| > 100 mm, where the purity is already sufficiently high. The optimal cut values are detailed in Table 4.6.

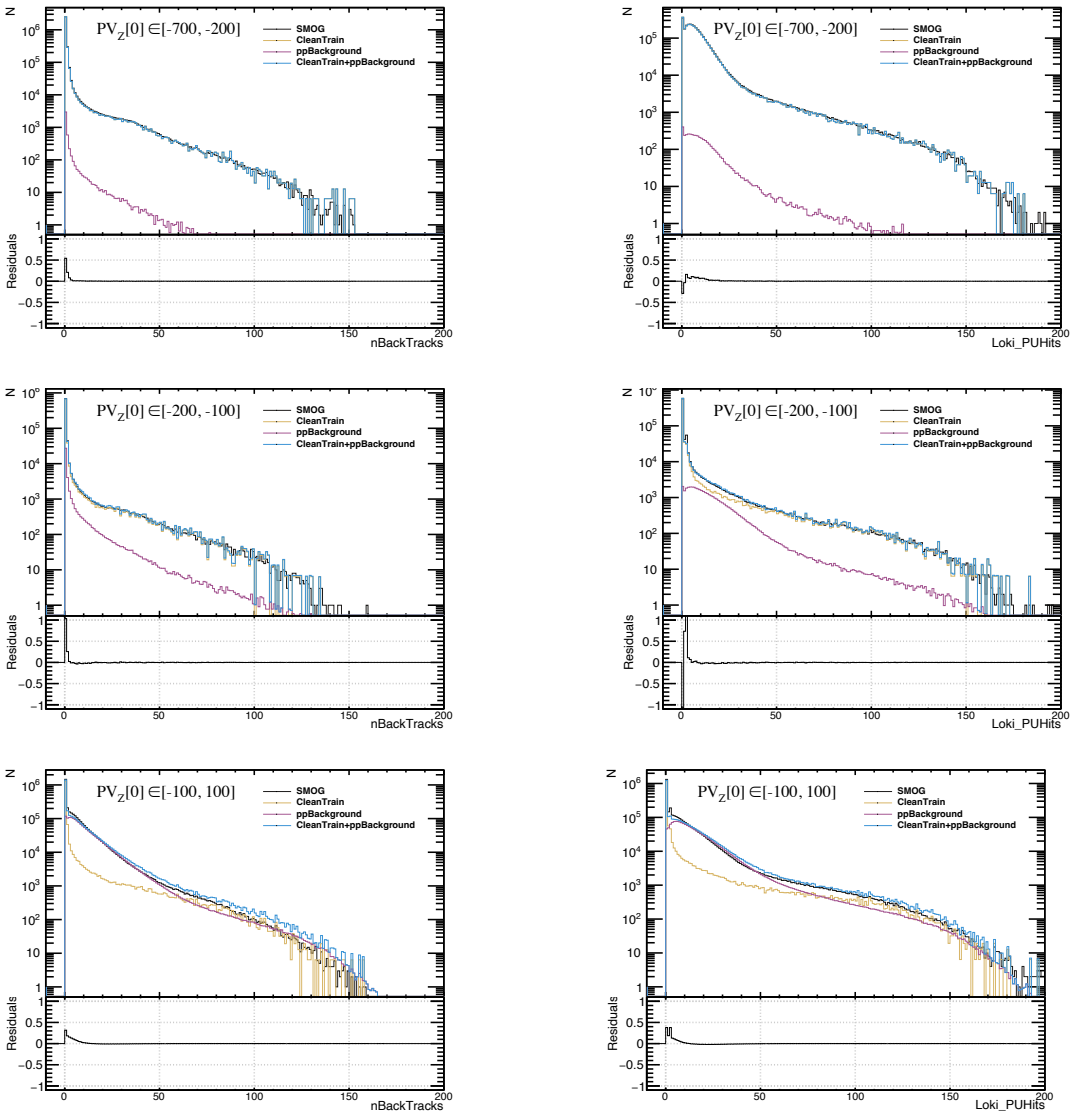


Figure 4.6: One-dimensional distributions of nPUHits and nBackTracks are presented for different z regions where the PV is reconstructed, considering the entire SMOG dataset and the weighted pp background and clean train sample. The residual are obtained by evaluating the differences between SMOG data and the sum of the signal-enriched and the background-enriched samples, divided by the SMOG data itself. The agreement between the entire SMOG dataset and the sum of the weighted background sample and the clean train sample validates the calculation of the normalisation factors. The discrepancies between the SMOG dataset and the combined signal-enriched and background-enriched samples are attributed to differences in the data-taking conditions.

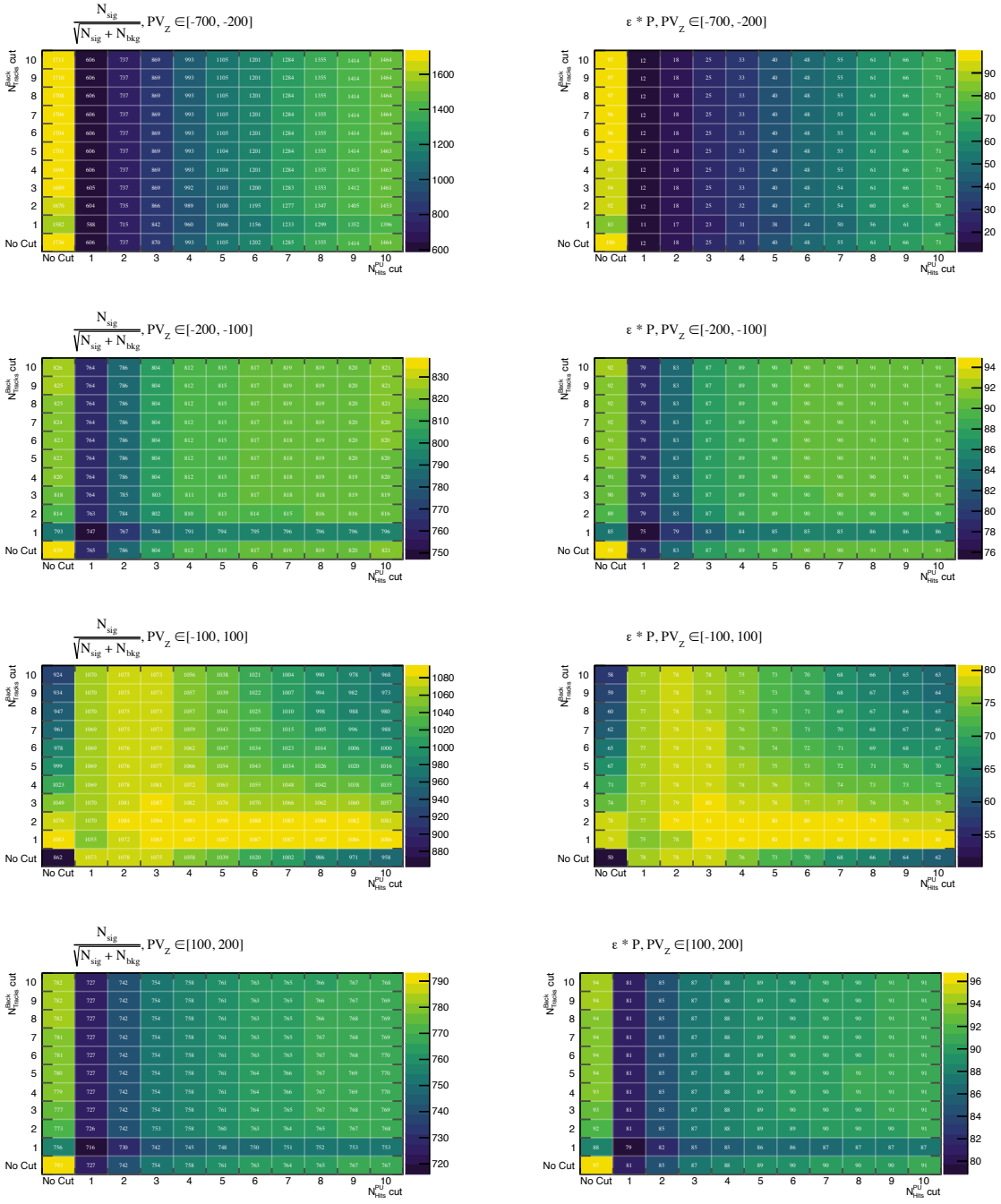


Figure 4.7: Distribution of significance (left) and $\varepsilon \times P$ (right) figures of merit for various thresholds on nPUHits and nBackTracks variables. The optimisation is conducted in four distinct PVZ regions, presented from top to bottom.

4.3.2 Track quality and kinematic selection

Following the event selection, the candidates are selected with loose criteria on kinematics and track quality. At this stage, all charged particles are assumed to be pions using the StdAllNoPIDsPions definition, which applies minimal quality requirements to charged tracks and assigns them the mass hypothesis of a pion. This assumption is made solely to obtain a particle object during reconstruction. The actual particle abundances for each hypothesis (π, K, p) are determined independently from this initial assumption. Then an additional selection is required, involving the following criteria:

- **Long track**: only tracks that leave hits in all tracking subdetectors are selected. Generally, the long tracks have the best quality and better momentum and position resolution;
- **Track quality**: a loose requirement is applied on the quality of the track fitting during the reconstruction. This cut corresponds to

$$\chi^2_{\text{track}}/\text{ndof} < 5; \tag{4.13}$$

- **RICH acceptance**: the kinematic range is chosen to match the RICH detector's acceptance. The following requirement are applied

$$\begin{aligned} &p \in [15, 80] \text{ GeV/c} \\ &\text{nSPDHits} \in [0, 1000] \\ &p_T \in [600, 3000] \text{ MeV/c} \\ &\eta \in [2.4, 4.5] \ . \end{aligned}$$

The momentum lower limit is chosen to be above the threshold for generating Cherenkov light in RICH 1 for kaons ($p > 10$ GeV/c), enabling active discrimination between pions and kaons. The momentum upper limit is determined to ensure reliable particle identification, as PID performance starts to degrade at higher momentum due to the saturation of the Cherenkov angle.

Based on the PID performance in Run 2, as presented in Figure 3.21, the momentum region has been constrained to $p \in [15, 80]$ GeV/c to guarantee at least 80% efficiency and less than 10% misidentification for all three particle types. The lower limit is defined by the separation between protons and kaons, while the upper limit is constrained by the separation of kaons and pions.

Similarly, due to considerations about PID performance, the centrality range for PbNe collisions is limited, corresponding to centrality up to approximately 40%. Above this centrality threshold, the RICH detector's performance is estimated to provide poor particle identification. The pNe dataset is not affected by this constraint, as it covers centrality values up to approximately 70%. No direct study on PID performance as a function of event multiplicity was conducted during Run 2, making it challenging to estimate efficiency and misidentification rates for different centrality ranges directly.

Both momentum and centrality ranges may potentially be extended in future analyses during Run 3, taking into account the expected improvement in PID performance discussed in Chapter 6. Additionally, the initial estimation of PID as a function of event centrality is presented, offering the opportunity to better define the fiducial region of the analysis.

The p_T selection is based on statistics and on the kinematic coverage of the PID calibration samples available for fixed-target collisions. A better description of these datasets is presented in Section 4.6. The η selection corresponds to the pseudorapidity range that remains after applying the selections on p and p_T, accounting for the correlation among these three variables.

4.3.3 Ghost tracks suppression

Among the candidates selected in the preceding section, there are tracks that do not correspond to actual particles and are instead reconstructed from random hits in the detector. These specific candidates are referred to as ghost tracks, primarily originating from random matches of unrelated track segments upstream and downstream of the magnet. To eliminate ghost tracks, a selection is applied with the following requirement:

$$\text{TRACK\_GHOST\_PROB} < 0.4 \ ,$$

where the variable GhostProb is the output of a dedicated neural network classifier described in Section 3.2.1. After implementing this cut, a residual fraction of ghost candidates persists. The behaviour of the ghost particle fraction relative to the total number of events has been investigated using simulation samples, and the results are depicted in Figures 4.8 and 4.9.

In both PbNe and pNe simulations, the rate of ghost tracks scales with the increase in nSPDHits, serving as a proxy for event multiplicity. This phenomenon is expected as a higher detector occupancy enhances the likelihood of random hits reconstructing a track. The overall fractions of ghosts surviving the selection for both datasets are summarised in Table 4.7. Notably, the total fraction of surviving ghosts is higher for PbNe, consistent with the higher average detector activity in PbNe events. Notably, the residual ghost fraction remains relatively small in both datasets. This factor will be considered when extracting yields, as detailed in Section 4.7. Specifically, the average ghost fraction obtained from simulations is utilised to constrain the initialisation and range of ghost track yields.

Dataset	Ghosts fraction (%)
PbNe	1.4
pNe	1.0

Table 4.7: Average fraction of ghost particles surviving the selection cut in PbNe and pNe simulations.

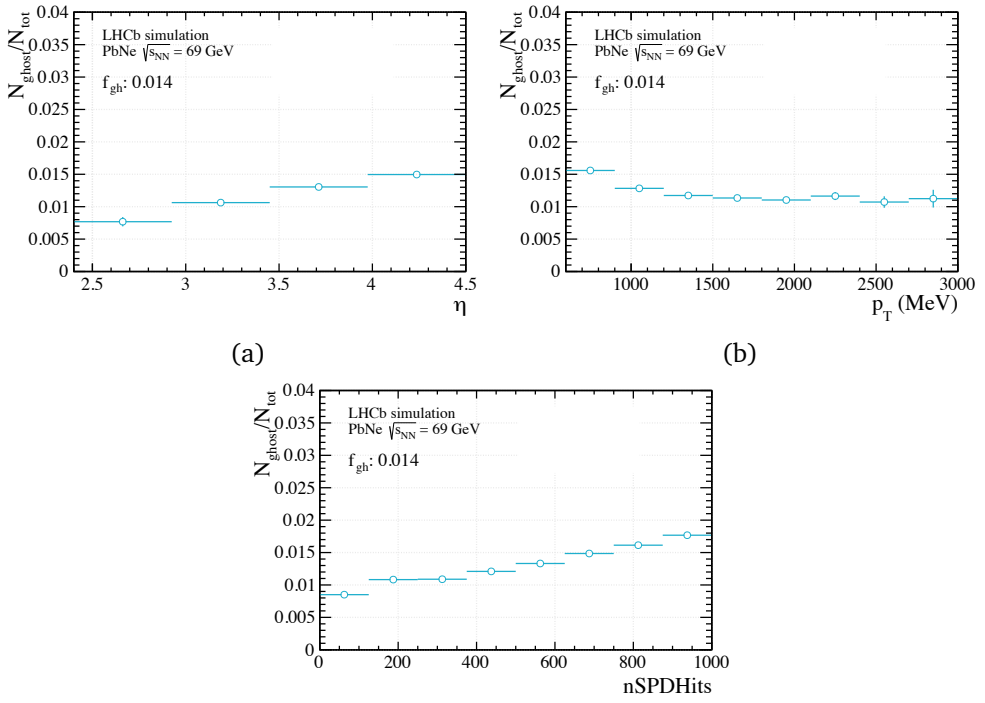


Figure 4.8: Fraction of ghost tracks with respect to the total number of tracks in PbNe simulation as a function of η (a) p_T (b) and nSPDHits (c).

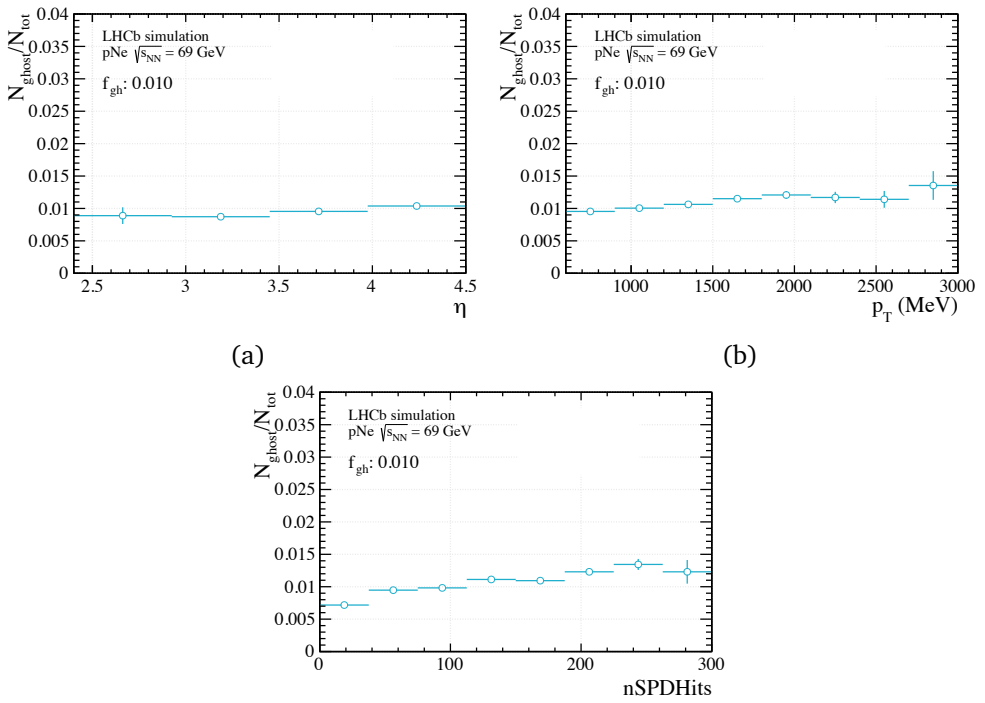


Figure 4.9: Fraction of ghost tracks with respect to the total number of tracks in pNe simulation as a function of η (a) p_T (b) and nSPDHits (c).

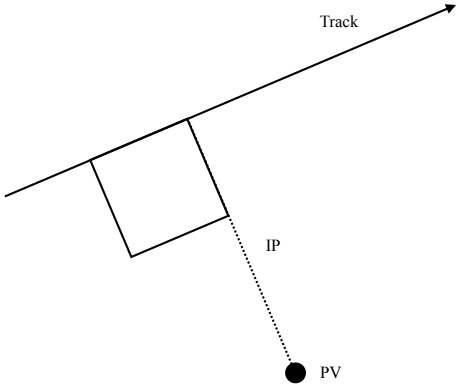


Figure 4.10: Illustration of the IP (length of the dotted line) of a track (solid line) in relation to the PV.

4.3.4 Prompt selection

The prompt selection aims to mitigate background contributions from $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ pairs originating from the decays of long-lived particles rather than being directly produced in the collision. A particle is considered long-lived if its lifetime exceeds 30 ps, and prompt classification is assigned to particles either directly produced in the primary interaction or lacking long-lived ancestors [149]. This category encompasses electrons, muons, pions, kaons, protons, as well as Σ^+, Σ^-, Ξ^-, and Ω^- baryons, and their corresponding antiparticles. A summary on the lifetime of different particle types with corresponding values of the mean proper lifetime is presented in Table 4.8.

To identify prompt particles, the Impact Parameter (IP) variable is crucial. The IP of a particle is defined as the distance of closest approach of its extrapolated trajectory to the primary vertex, as illustrated in Figure 4.10. The IP χ^2 is determined as the difference in the χ^2 of the primary vertex reconstructed with and without the considered particle, providing a measure of the track's displacement from the primary vertex. Particles originating from the primary vertex should have a small IP χ^2, while higher values indicate inconsistency with the primary vertex.

In both PbNe and pNe data, prompt candidates are selected by applying the condition:

$$\chi^2_{IP} < 12 \, .$$

To validate this selection, the prompt requirement is investigated in the simulation using the IS_PROMPT condition. This variable pertains to a track and designates it as a prompt particle if the lifetime of the parent particle is less than 10^{-7} ns. The distribution of charged tracks can be separated into prompt and non-prompt particles. The two contributions are depicted in Figures 4.11 and 4.12 for PbNe and pNe simulations respectively, for tracks that pass all the previously defined selections. Overall, the non-prompt contamination after the χ^2 cut is approximately 1-2%. In Section 4.7, the yields are corrected by the prompt purity factor. Among the three hadron species, the proton exhibits the highest non-prompt contamination that survives the track selection cuts. A potential explanation for this contamination is protons originating

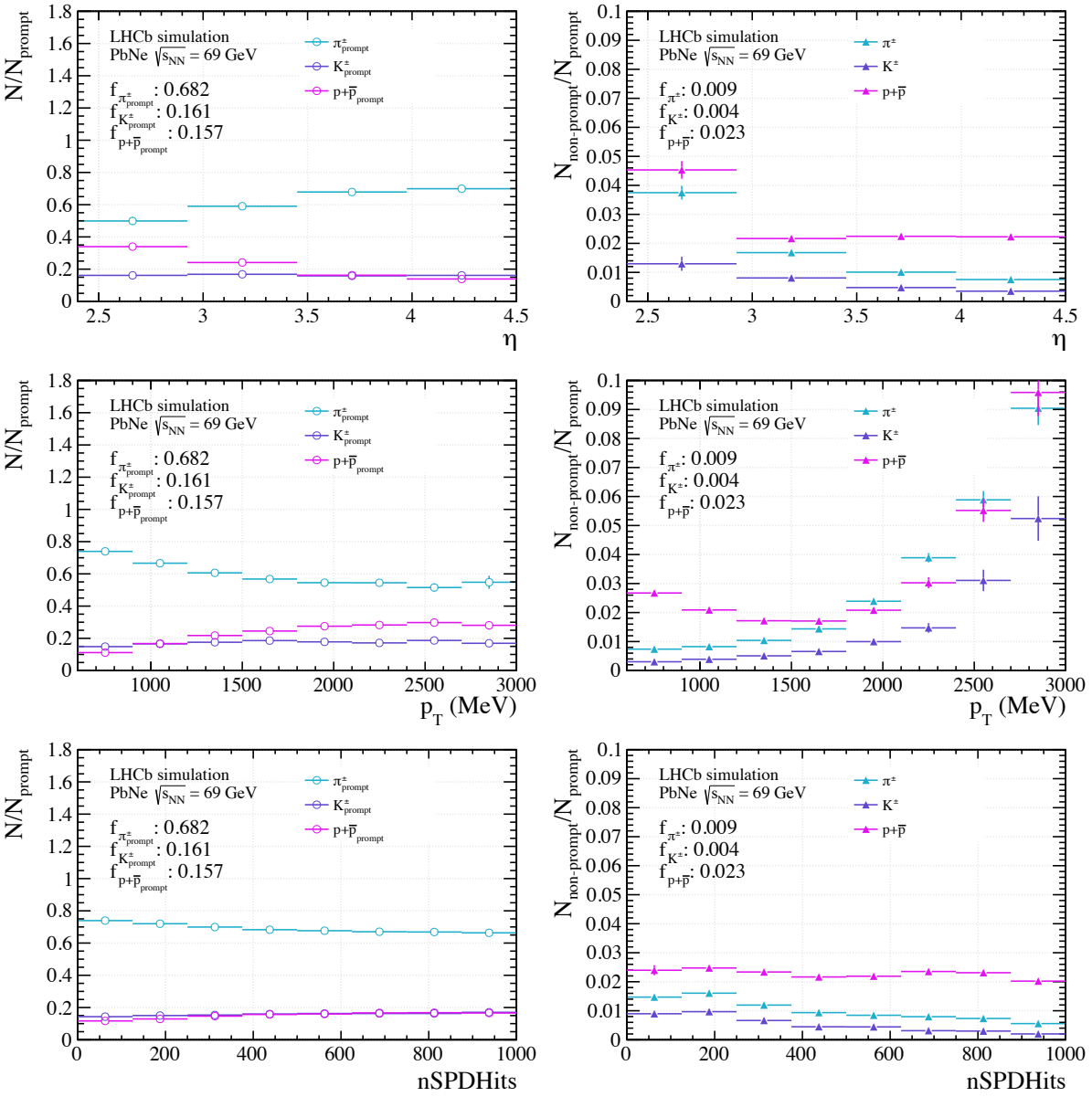


Figure 4.11: Fraction of prompt $\pi^\pm, K^\pm, \overset{(-)}{p}$ with respect to the total number of prompt charged tracks and the fraction of non-prompt $\pi^\pm, K^\pm, \overset{(-)}{p}$ with respect to the corresponding number of prompt particles of the same type, evaluated in the PbNe simulation. Both distributions are presented as functions of η (top), p_T (middle), and nSPDHits (bottom).

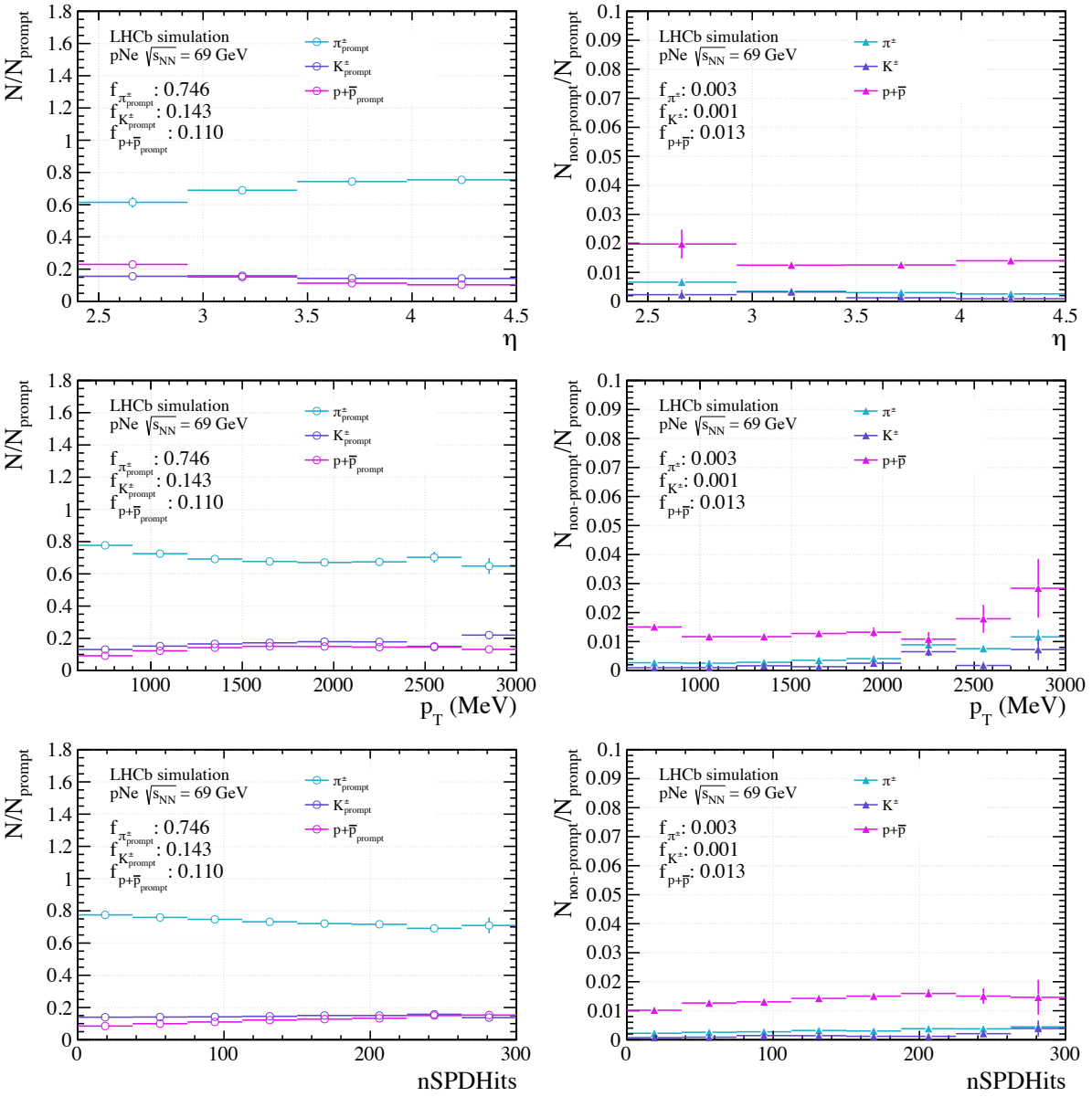


Figure 4.12: Fraction of prompt $\pi^\pm, K^\pm, \overset{(-)}{p}$ with respect to the total number of prompt charged tracks and the fraction of non-prompt $\pi^\pm, K^\pm, \overset{(-)}{p}$ with respect to the corresponding number of prompt particles of the same type, evaluated in the pNe simulation. Both distributions are presented as functions of η (top), p_T (middle), and nSPDHits (bottom).

Specie	Width Γ (GeV)	Mean proper lifetime τ (ps)	cm/c
p^+	$\ll 10^{-29}$	$\gg 10^{+15}$	$\gg 10^{+14}$
γ	$\ll 10^{-29}$	$\gg 10^{+15}$	$\gg 10^{+14}$
K^0	$\ll 10^{-29}$	$\gg 10^{+15}$	$\gg 10^{+14}$
e^-	$\ll 10^{-29}$	$\gg 10^{+15}$	$\gg 10^{+14}$
n	7.478×10^{-28}	$8.861 \times 10^{+14}$	$2.656 \times 10^{+13}$
μ^-	2.996×10^{-19}	$2.212 \times 10^{+06}$	$6.63 \times 10^{+04}$
K^0_L	1.287×10^{-17}	$5.148 \times 10^{+04}$	1543
π^+	2.528×10^{-17}	$2.621 \times 10^{+04}$	785.7
K^+	5.317×10^{-17}	$1.246 \times 10^{+04}$	373.6
Ξ^0	2.27×10^{-15}	291.9	8.751
Λ^-	2.501×10^{-15}	264.9	7.943
Ξ^-	4.02×10^{-15}	164.8	4.941
Σ^-	4.45×10^{-15}	148.9	4.464
K^0_S	7.351×10^{-15}	90.14	2.702
Ω^-	8.071×10^{-15}	82.1	2.461
Σ^+	8.209×10^{-15}	80.72	2.42
B^+	4.018×10^{-13}	1.649	0.04944
Ω^-_b	4.2×10^{-13}	1.578	0.0473
Ξ^-_b	4.22×10^{-13}	1.57	0.04707
B^0	4.33×10^{-13}	1.53	0.04588
B^0_s	4.359×10^{-13}	1.52	0.04557
Λ_b	4.49×10^{-13}	1.476	0.04424
Ξ^0_b	4.5×10^{-13}	1.472	0.04414
D^+	6.33×10^{-13}	1.047	0.03138
B^+_c	1.298×10^{-12}	0.5105	0.0153
D^+_s	1.317×10^{-12}	0.5031	0.01508
Ξ^+_c	1.49×10^{-12}	0.4447	0.01333
D^0	1.605×10^{-12}	0.4128	0.01238
τ^-	2.267×10^{-12}	0.2923	0.008762
Λ^+_c	3.3×10^{-12}	0.2008	0.00602
Ξ^0_c	5.9×10^{-12}	0.1123	0.003367
Ω^0_c	9.6×10^{-12}	0.06902	0.002069

Table 4.8: Width (Γ), and mean proper lifetime (τ) of various particles, sorted by descending lifetime [149].

from decaying Λ and other strange baryons (Σ, Ξ) which do not fall within the prompt definition due to their mean lifetime being longer than 10^{-6} ns.

Dataset	Candidate	Prompt / Total (%)	Non-Prompt / Prompt (%)
PbNe	$\pi^{\pm}$	68.2	0.9
	$K^{\pm}$	16.1	0.4
	$\overset{(-)}{p}$	15.7	2.3
pNe	$\pi^{\pm}$	74.6	0.3
	$K^{\pm}$	14.3	0.1
	$\overset{(-)}{p}$	11.0	1.3

Table 4.9: Prompt and non-prompt fractions for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ obtained from PbNe and pNe simulation.

4.3.5 Fraction of $e^{\pm}$ and $\mu^{\pm}$

As described in Section 4.1, the candidate selection is tailored to apply loose criteria for selecting charged tracks. Given that the primary focus of the analysis is on prompt $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$, any charged particle surviving the candidate selection is considered as background.

The predominant contribution may arise from $e^{\pm}$ and $\mu^{\pm}$ promptly produced in the collision. To estimate the fraction of these particles surviving the candidate selection, a dedicated study is conducted on simulated data. The results are presented in Figures 4.13 and 4.14, and the overall fractions of $e^{\pm}$ and $\mu^{\pm}$ that pass the selection for both datasets are summarised in Table 4.10.

The contribution from muons is smaller for both datasets. The electron component is more prominent at low p_T, and the distribution shape resembles that of pions, as illustrated, for instance, in Figure 4.12. Additionally, electrons exhibit a RICH PID response compatible with the pions hypothesis in the condition of saturated Cherenkov angle. Consequently, the correction for electron contamination is exclusively applied when extracting the experimental yield from pions, a step subsequently described in Section 4.7.

Dataset	$e^{\pm}$ fraction (%)	$\mu^{\pm}$ fraction (%)
PbNe	0.5	0.1
pNe	0.7	< 0.1

Table 4.10: Average fraction of $e^{\pm}$ and $\mu^{\pm}$ surviving the selection cut in PbNe and pNe simulations.

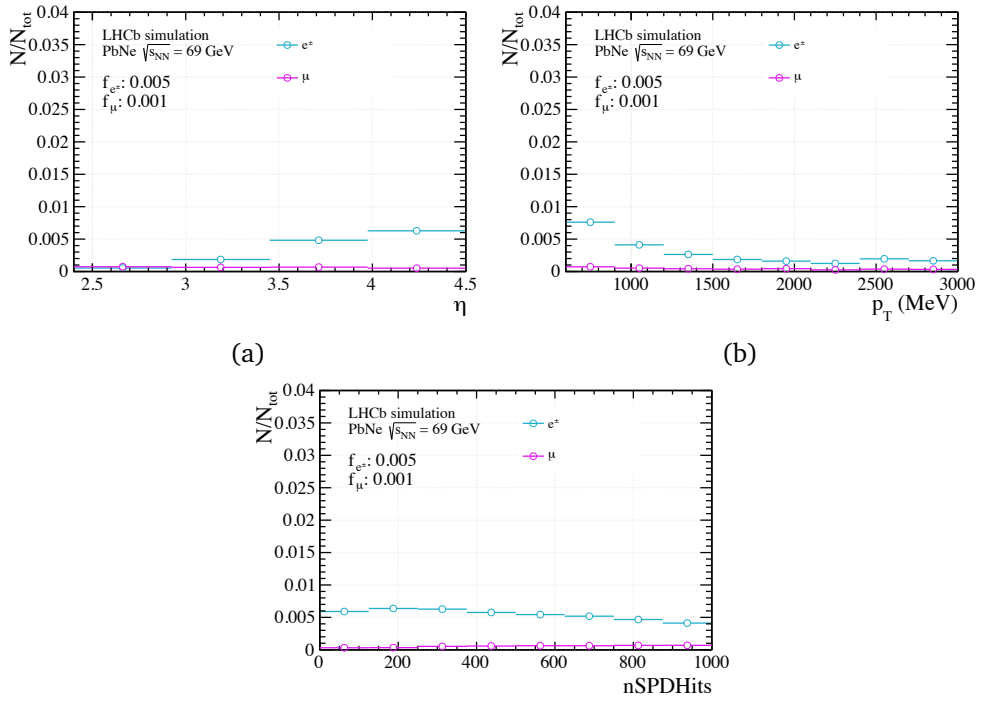


Figure 4.13: Fraction of $e^{\pm}$ and $\mu^{\pm}$ with respect to the total number of tracks in PbNe simulation as a function of η (a) p_T (b) and nSPDHits (c).

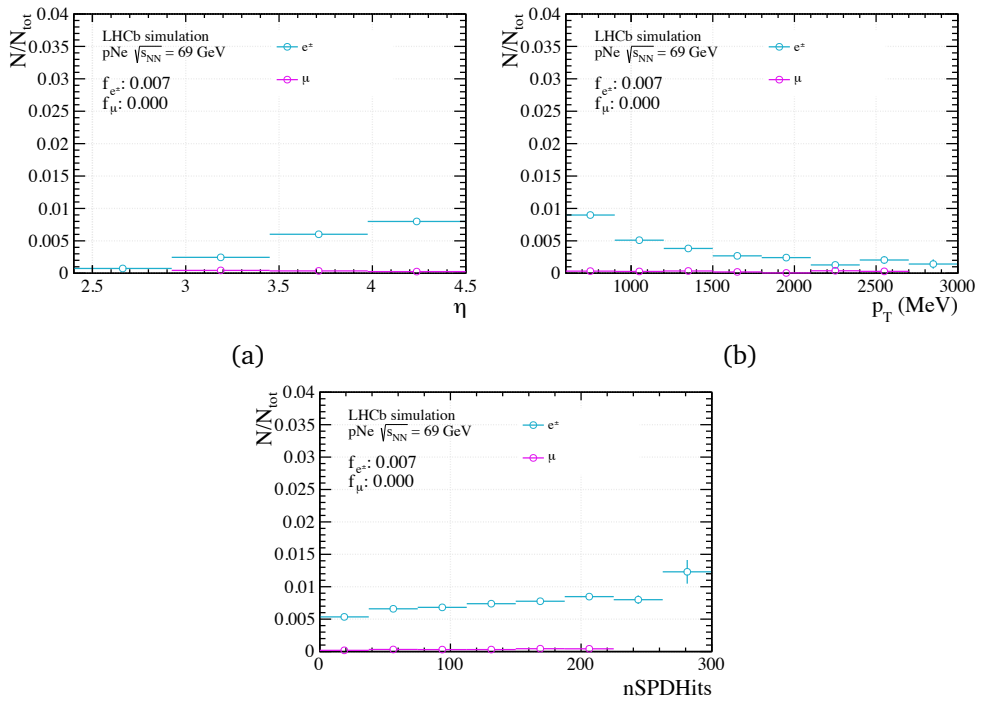


Figure 4.14: Fraction of $e^{\pm}$ and $\mu^{\pm}$ with respect to the total number of tracks in pNe simulation as a function of η (a) p_T (b) and nSPDHits (c).

4.3.6 Summary of the selection

A summary of both the global event and candidate selection is provided in Table 4.11. In order to address potential biases arising from differences between the two datasets, a common global event selection is implemented, taking into account the specific studies conducted for each dataset. To assess the overall quality of the candidate selection, the distributions of relevant variables after applying the candidate selection are compared between PbNe and pNe in Figure 4.15. The kinematic coverage in η and p_T is found to be consistent across both datasets. There are two notable differences between the PbNe and pNe datasets:

- **Multiplicity Distributions:** PbNe events cover the entire multiplicity range selected for the analysis, while pNe covers only a limited range, as expected.
- **PID Variables:** The PIDK and PIDp distributions for PbNe display a narrower spread, attributed to a general degradation in PID performance at high multiplicities. This results in a greater overlap of the PID distributions for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ and, consequently, a reduced spread in the PIDK and PIDp variables. This underscores the necessity to treat the modelling of PID templates independently for PbNe and pNe, as elaborated in Section 4.6.

The impact of the candidate selection criteria on simulations and the agreement between the simulation and the data are discussed in the Section 4.4.

Selection type	Cut PbNe (pNe)
Global Event Selection	BC_TYPE=1
	nPVs >0
	PVZ$\in[-700,200]$
	PVNTRACKS[0] >13 (4)
	PVNTRACKS[0] $>$ PVNTRACKS[i], $i>0$
	nBackTracks <50
	PVZ $\in[-200,-150]$, nPUHits <10
	PVZ $\in[-150,150]$, nPUHits <5 (nPUHits <3, nBackTracks <2)
	PVZ $\in[150,200]$, nPUHits <10
Track quality	IS_LONG=1
	TRACK_CHI2NDOF <5
	TRACK_GHOST_PROB <0.4
Kinematic	$\eta\in[2.4,4.5]$
	$p_T\in[600,3000]$ MeV/c
	$p\in[15,80]$ GeV/c
	nSPDHits $\in[0,1000]$, corresponding to centrality 88-39%
Prompt	OWNPV_IPCHI2 <12
Stripping	StrippingMBMicroBiasVelo >0
Trigger	Hlt1MBMicroBiasVelo >0

Table 4.11: Summary of the candidate selection for PbNe (pNe).

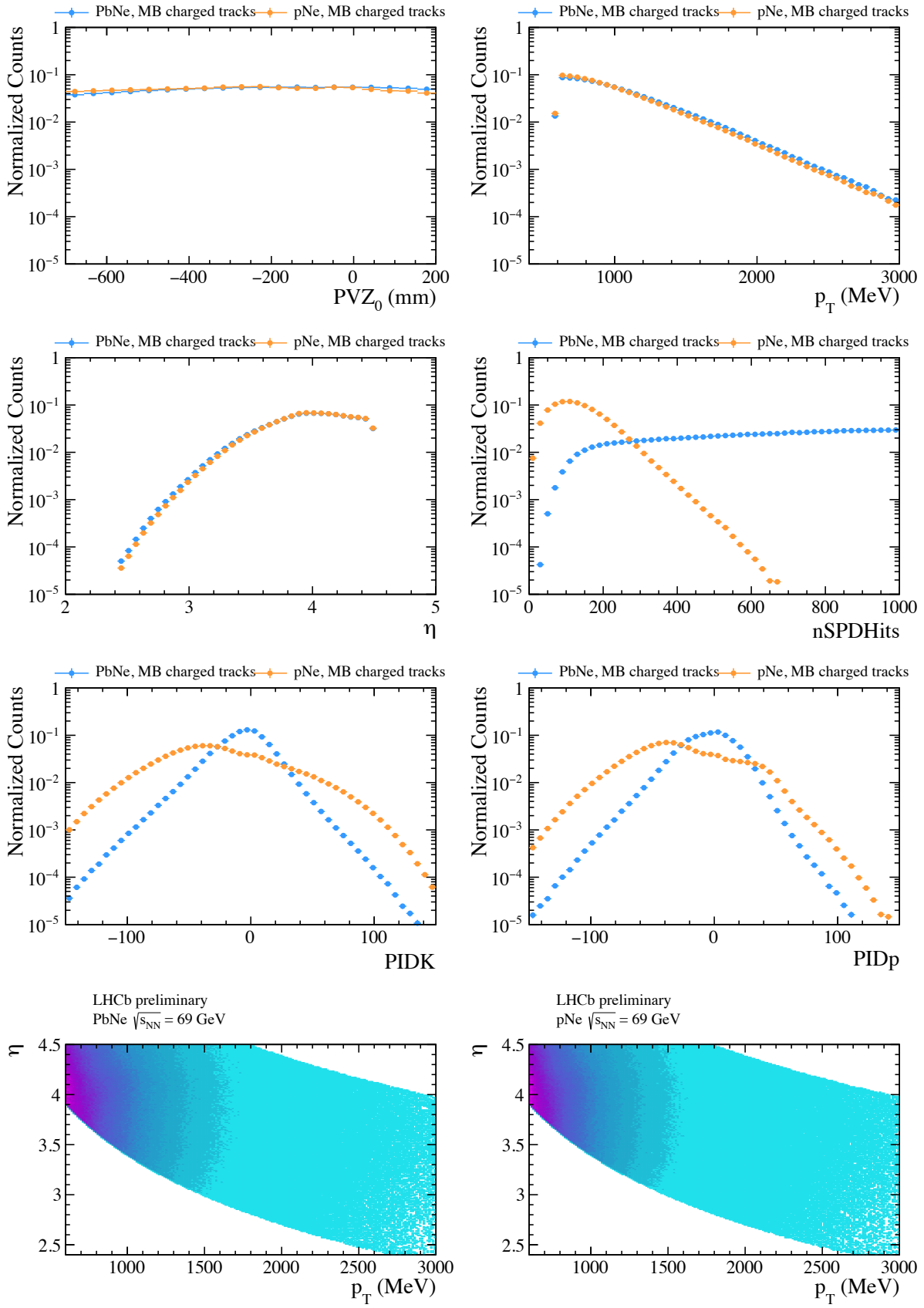


Figure 4.15: Comparison of PbNe and pNe datasets after applying the candidate selection, illustrating the dependence on relevant variables for the analysis. Additionally, the correlation between η and p_T is depicted in the two lower plots for both datasets.

4.4 Simulation reweighing and validation

The same candidate selection described in Section 4.3 is applied to simulated data. To address discrepancies between the data and simulation after the candidate selection, a reweighing procedure is applied to simulated events. The reweighing is executed using a multivariate method employing the Gradient Boost Reweighing (GBR) algorithm [150]. This approach allows for the adjustment of multidimensional distributions, taking into account correlations among the considered variables. The selected variables for reweighing the simulations include:

- **nSPDHits**: the selected multiplicity variable (based on the hits on the SPD detector), influencing the categorisation of events based on centrality. Additionally, this variable exhibits the largest disagreement between data and simulation;
- η, p_T: key kinematic variables for the analysis;
- **PVZ**: a distribution that is essential to align between the data and simulation for a correct evaluation of the PV reconstruction efficiency, detailed in Section 4.5.

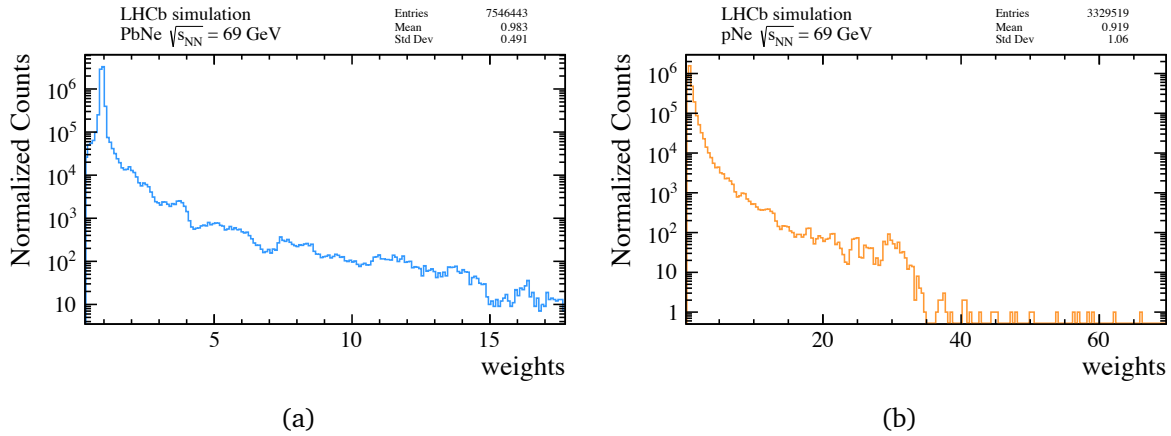
Different subsets of variables can be utilised as input for the model, and the difference in the results obtained with different sets is accounted as a systematic uncertainties associated with the weighing procedure, as detailed in Section 4.8.

The weighing procedure employs a cross-validation method for its implementation, chosen for its tendency to yield a less biased estimate of the model compared to other methods like a simple train/test split [151]. In this approach, the set of observations is randomly divided into k groups or folds of approximately equal size. The first fold serves as a validation set, and the method is fitted on the remaining $k-1$ folds. The choice of k is typically 5 or 10, and as k increases, the difference in size between the training set and the resampling subsets diminishes. This reduction in difference leads to a smaller bias in the technique. For this application, the number of folds is selected to be equal to 5.

A GBR model is trained independently for each dataset, optimising the model's hyperparameters, such as the number of iterations, learning rate, maximum depth, and minimum number of leaves, to prevent overfitting. Given the overall agreement between data and simulation and the satisfactory performance of the model, the hyperparameters are set to be equal for both PbNe and pNe datasets. The values of the hyperparameters suitable for this analysis are shown in Table 4.12. A preliminary validation of the model involves examining the distribution of weights.

Parameter	Value
N folds	5
N estimators	25
Learning rate	0.5
Maximum Depth	5
N minimum entries per leaf	100000

Table 4.12: Summary of the hyperparameters of the GBR models used for multivariate reweighing of PbNe and pNe simulations.



(a) (b)

Figure 4.16: Distributions of the weights derived from the GBR algorithms for PbNe (a) and *p*Ne (b). Subsequently, these weights are applied to reweight the simulation.

The resulting weight distributions for the GBR models trained on PbNe and *p*Ne are shown in Figure 4.16. In both cases, the weight distribution displays no irregularities or excessively elevated values, suggesting that the reweighing process maintains the quality of the simulation without introducing significant biases. In the case of the *p*Ne simulation, the weights' distribution reaches slightly higher values. This phenomenon may be linked to a non-trivial correlation between PVZ and the multiplicity variables following the global event selection. To mitigate overfitting, the minimum number of candidates in each leaf has been increased.

The reweighed simulation distributions are displayed in Figures 4.17 and 4.18, along with data and unweighted simulation distributions for PbNe and *p*Ne, respectively. The metric adopted to quantify the improvement related to the reweighing is the Kolmogorov-Smirnov (KS) distance between data and the weighted simulation. A notable enhancement is observed in the decrease of the KS distance for all variables included in the model. Particularly noteworthy is the improvement in the nSPDHits distribution for the *p*Ne simulation. Once validated, the classifier is ready to compute weights for the entire simulation sample.

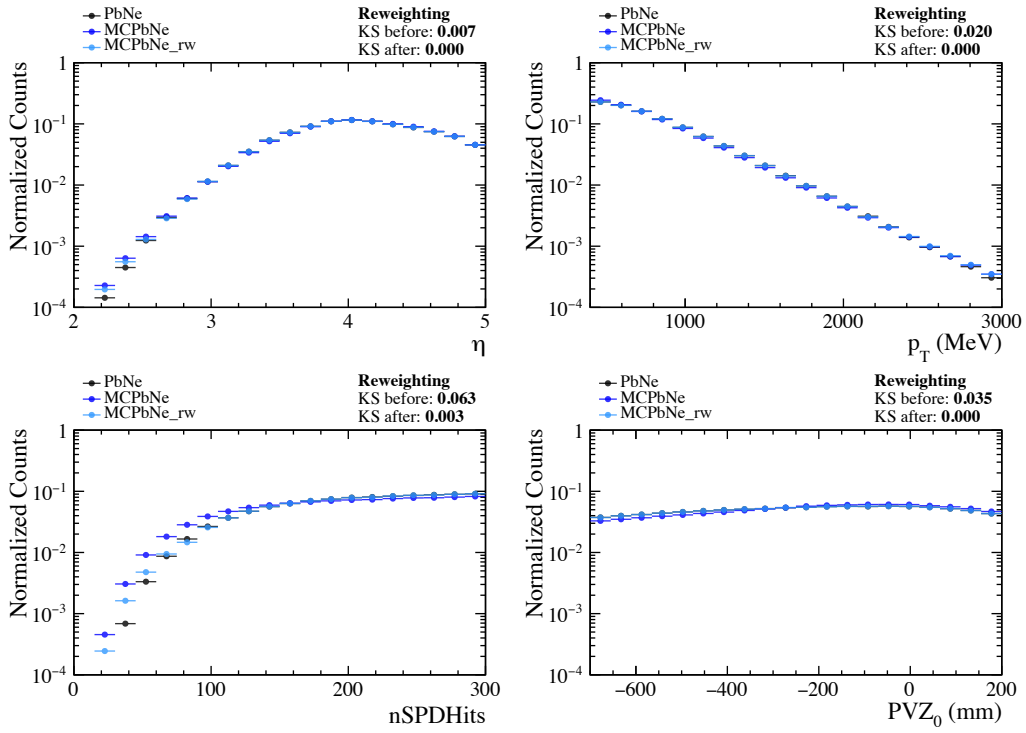


Figure 4.17: Comparison of the distributions for selected candidates in PbNe data and simulation before and after reweighing, including η, p_T, nSPDHits, and PVZ. The KS distance between the data and simulation distributions before and after reweighing is also provided.

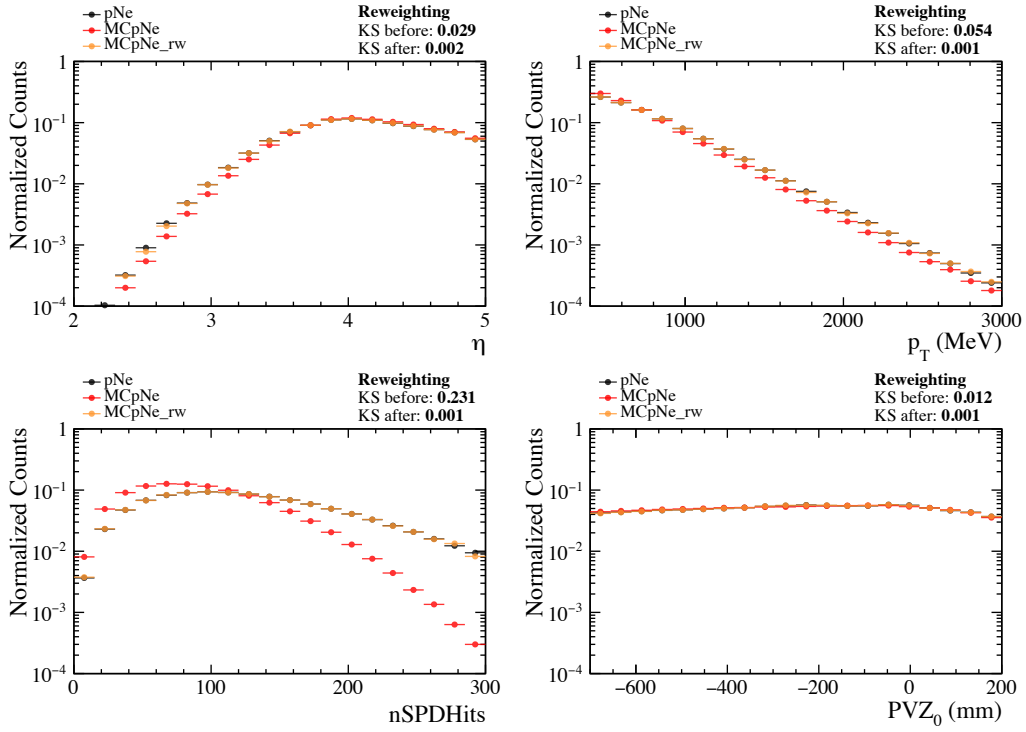


Figure 4.18: Comparison of the distributions for selected candidates in pNe data and simulation before and after reweighing, including η, p_T, nSPDHits, and PVZ. The KS distance between the data and simulation distributions before and after reweighing is also provided.

4.5 Efficiencies

The calculation of efficiencies is a crucial step in the analysis to take into account geometrical, tracking, and selection effects that influence the number of candidates observed experimentally. Efficiencies are computed from simulation, which has been appropriately reweighed to address discrepancies with data, as described in Section 4.4. The correction to the measured yields is expressed as follows:

$$N^h_{\text{corr}} = \frac{N^h}{\varepsilon_{\text{Tot}}}, \quad h = \pi^\pm, K^\pm, \overset{(-)}{p} \tag{4.14}$$

where N^h_{corr} represents the real particle yield, N^h is the measured yield, and ε_{Tot} is the total efficiency. In particular, the total efficiency can be decomposed as follows:

$$\varepsilon_{\text{Tot}} = \varepsilon_{\text{Acceptance}} \times \varepsilon_{\text{Tracking}} \times \varepsilon_{\text{PVreco}} \times \varepsilon_{\text{GEC,Sel,Trig}}. \tag{4.15}$$

The different terms in the expression are computed sequentially and are defined as follows:

- $\varepsilon_{\text{Acceptance}}$: geometrical acceptance;
- $\varepsilon_{\text{Tracking}}$: tracking reconstruction efficiency;
- $\varepsilon_{\text{PVreco}}$: primary vertex reconstruction efficiency;
- $\varepsilon_{\text{GEC,Sel,Trig}}$: GEC, selection and trigger efficiency.

This factorisation provides a detailed understanding of how the total efficiency varies with the relevant variables in the analysis. The calculations are conducted separately for $\pi^\pm, K^\pm, \overset{(-)}{p}$, utilising the information of the true particle identity (TRUE_ID) at the generator level in the simulation.

Efficiencies used to correct the yields are computed in simultaneous bins following the binning scheme described in Section 4.1.1. In the following sections, a more refined binning scheme is applied to study the dependence of the efficiencies more comprehensively. Additionally, the efficiency measurement encompasses the effects of bin migration of events between neighbouring kinematic intervals due to resolution effects in the reconstruction. Bin migration involves assigning reconstructed tracks to (η, p_T) bins that may not correspond to the real pseudorapidity and transverse momentum of the particle. This effect becomes more pronounced in regions of poor detector resolution and smaller binning sizes. The true values, denoted as $(\eta^{\text{TRUE}}, p_T^{\text{TRUE}})$, can be studied in the simulation sample from generator-level information.

A detailed discussion of the various steps involved in calculating the total efficiency is presented in the subsequent sections.

4.5.1 Acceptance

The geometrical acceptance ($\varepsilon_{\text{Acceptance}}$) is defined as the fraction of candidates geometrically intercepted by a sufficient number of components in the detector allowing the track to be subsequently reconstructed. To estimate this value, the "Reconstructible" [152] condition available at the generator level, is utilised. Specifically, a stable charged MC particle is considered to be reconstructible if any of the following criteria are met:

- **Long tracks**: the MC particle has 3 VELO clusters and 1 stereo clusters in each of the 3 seed stations;
- **Upstream tracks**: the MC particle has 1 stereo clusters in each of the 3 seed stations and 3 TT clusters;
- **VELOTT tracks**: the MC particle has 3 VELO clusters and 3 TT clusters.

The evaluation of the geometrical efficiency can be expressed as follows:

$$\varepsilon_{\text{Acceptance}} = \frac{N\ (\text{Reconstructible} == \text{True})}{N\ (\text{Generated in the region of interest})}, \tag{4.16}$$

where the total number of generated prompt charged particles in the simulation is determined based on all generated prompt particles in the kinematic region selected for the analysis.

The distributions of $\varepsilon_{\text{Acceptance}}$, evaluated independently for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ as functions of η, p_T, nSPDHits, and PVZ, are depicted in Figures 4.19 and 4.20 for PbNe and pNe, respectively. The overall geometrical acceptance in both datasets is approximately 98%. The acceptance efficiency for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ is very similar in both PbNe and pNe simulations, with a slight dependence observed concerning η. This dependency could be related to events near the detector's peripheral regions where simulated particles may not intercept enough detector components. No significant modulations are observed with p_T, PVZ or event multiplicity in either dataset.

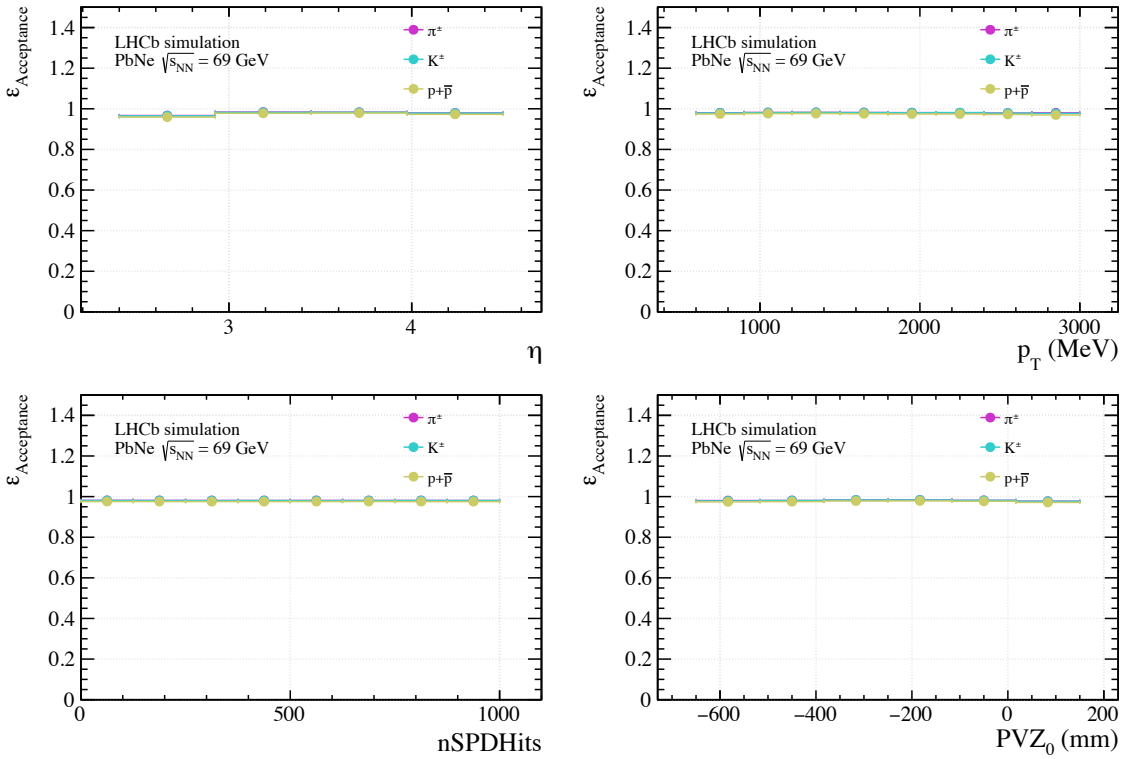


Figure 4.19: Acceptance efficiency calculated independently for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ from reweighed PbNe simulation as a function η, p_T, nSPDHits and PVZ.

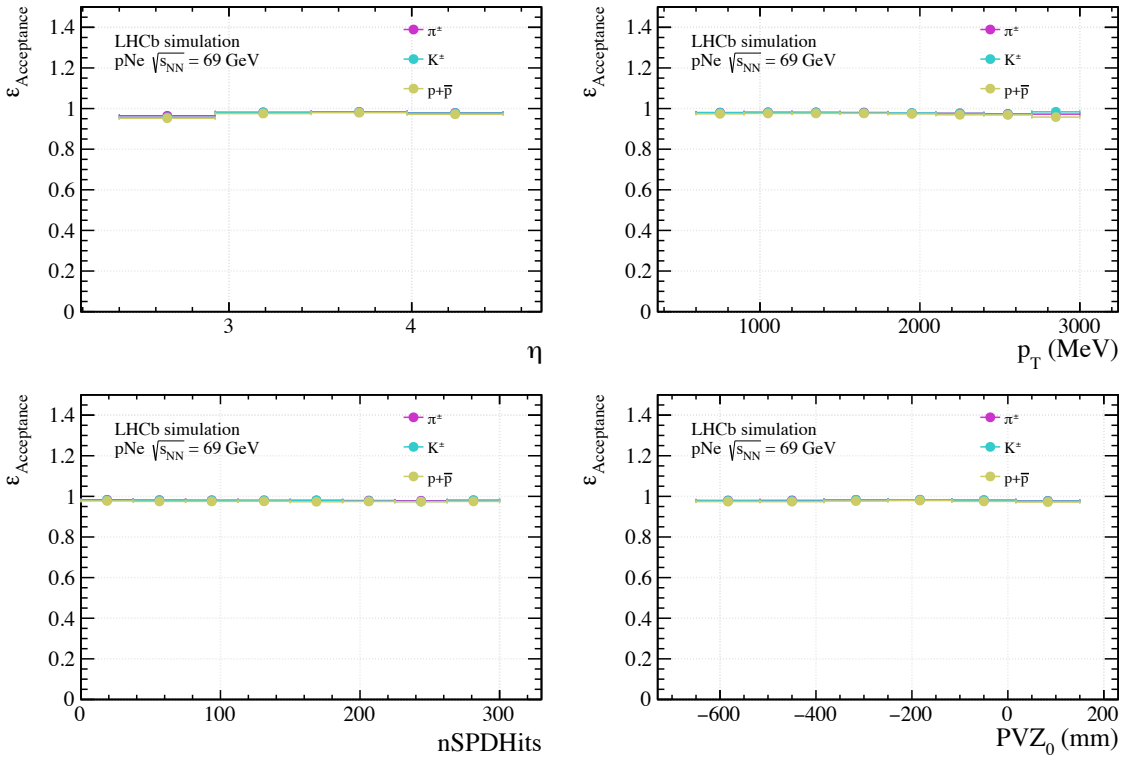


Figure 4.20: Acceptance efficiency calculated independently for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ from reweighed pNe simulation as a function η, p_T, nSPDHits and PVZ.

4.5.2 Tracking

The tracking efficiency is determined through simulation by calculating the ratio of reconstructed tracks matched to the number of candidates within the detector acceptance. To estimate this value, the "Reconstructed" condition [152] at the generator level is employed. In this scenario, a charged reconstructed particle is associated to a MC particle if:

- **Long tracks**: share at least 70% of the clusters in the VELO and 70% in the Seeding Stations;
- **Upstream tracks**: share at least 70% of the clusters in the Seeding Stations and have no more than 1 different cluster in TT out of those used;
- **VELOTT tracks**: share at least 70% of the clusters in the VELO and have no more than 1 different cluster in TT out of those used.

The evaluation of the tracking efficiency can be expressed as follows:

$$\varepsilon_{\text{Tracking}} = \frac{N\ (\text{Reconstructed} == \text{True})}{N\ (\text{Reconstructible} == \text{True})}\ . \tag{4.17}$$

The overall tracking efficiency for pions and kaons in the PbNe simulation is approximately 83%, while the tracking efficiency for protons is around 80%. Small variations among different particle types primarily arise from differences in their interactions with the detector material and the fact that pions and kaons may decay before leaving a long track, unlike protons. In the pNe simulation, the overall tracking efficiency for pions and kaons is close to 88%, and around 84% for protons. The lower tracking efficiency in PbNe can be attributed to a generally higher average detector occupancy, leading to a degradation in tracking performance.

The distributions of $\varepsilon_{\text{Tracking}}$ evaluated independently for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ as functions of η, p_T, nSPDHits, and PVZ are depicted in Figures 4.21 and 4.22 for PbNe and pNe, respectively. The efficiency exhibits a slight decrease with rising detector occupancy, as anticipated. Additionally, a dependence on PVZ and especially η is observed. These two effects may be correlated and can be attributed to the fact that as collisions approach the end of the VELO, there are fewer modules available to reconstruct the track, leading to a drop in the track reconstruction efficiency.

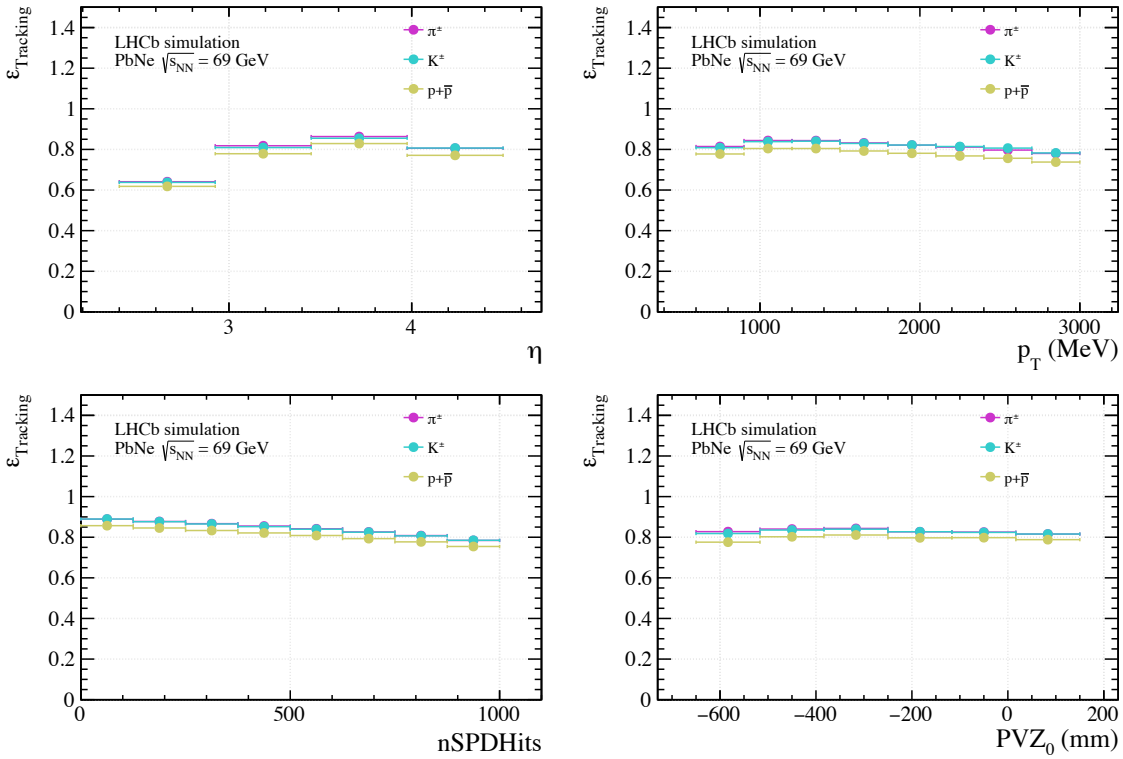


Figure 4.21: Tracking efficiency calculated independently for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ from reweighed PbNe simulation as a function η, p_T, nSPDHits and PVZ.

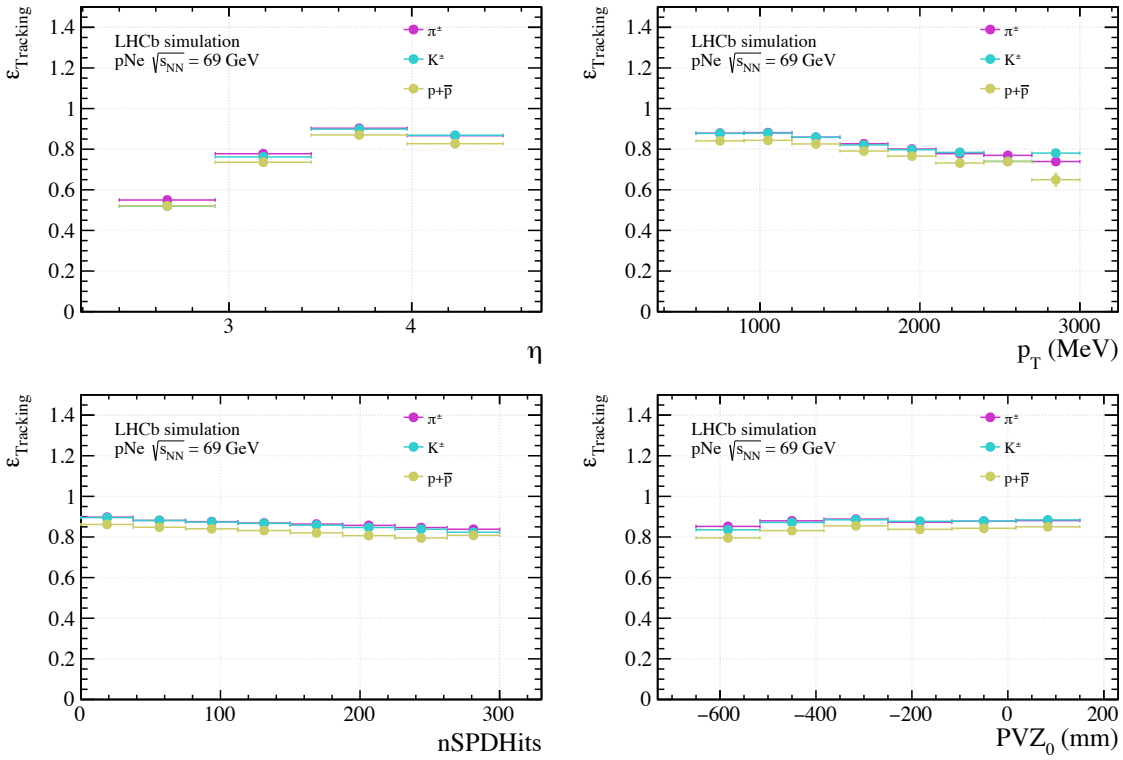


Figure 4.22: Tracking efficiency calculated independently for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ from reweighed pNe simulation as a function η, p_T, nSPDHits and PVZ.

4.5.3 PV reconstruction

The efficiency of reconstructing the primary vertex is assessed through the reweighed simulation. This step is especially important for SMOG fixed-target collisions due to the extended region along the z-direction. The efficiency is calculated by requiring that at least one primary vertex is reconstructed in the event. Given the substantial number of tracks originating from the collision, an additional requirement for true matching is applied. More precisely, the truth matching condition deems a primary vertex as reconstructed if its position is determined within 20 mm in z and 2 mm in x and y relative to the truth-simulated position. These specific values are chosen based on the average resolution of the reconstructed primary vertex position along the different directions.

The primary vertex reconstruction efficiency is computed as follows:

$$\varepsilon_{\text{PVreco}} = \frac{N\ (\text{nPVs}{>}0 \text{ and PV true-matching})}{N\ (\text{Reconstructed} == \text{True})}\ . \tag{4.18}$$

The overall PV reconstruction efficiency in the PbNe and pNe simulations is approximately 88% and 83%, respectively. This difference is attributed to the higher average number of tracks in PbNe collisions, resulting in a higher efficiency in PV reconstruction. This observation aligns with the findings in the global event studies presented in Section 4.3.1.

The distributions of $\varepsilon_{\text{PVreco}}$, depicted as functions of η, p_T, nSPDHits, and PVZ, are shown in Figures 4.23 and 4.24 for PbNe and pNe, respectively. The PV efficiency exhibits a slight decrease in the negative PVZ region, mainly due to the degradation of VELO resolution when moving away from the nominal interaction point ($z = 0$). Additionally, the efficiency is lower at small values of nSPDHits, reflecting the correlation between nSPDHits and other multiplicity variables such as the number of tracks in the event. A low number of tracks leads to a degradation in PV reconstruction performance. In the PbNe simulation, a small decrease at high nSPDHits is observed, attributed to the effect of the true matching requirement. No significant η and p_T dependencies are observed.

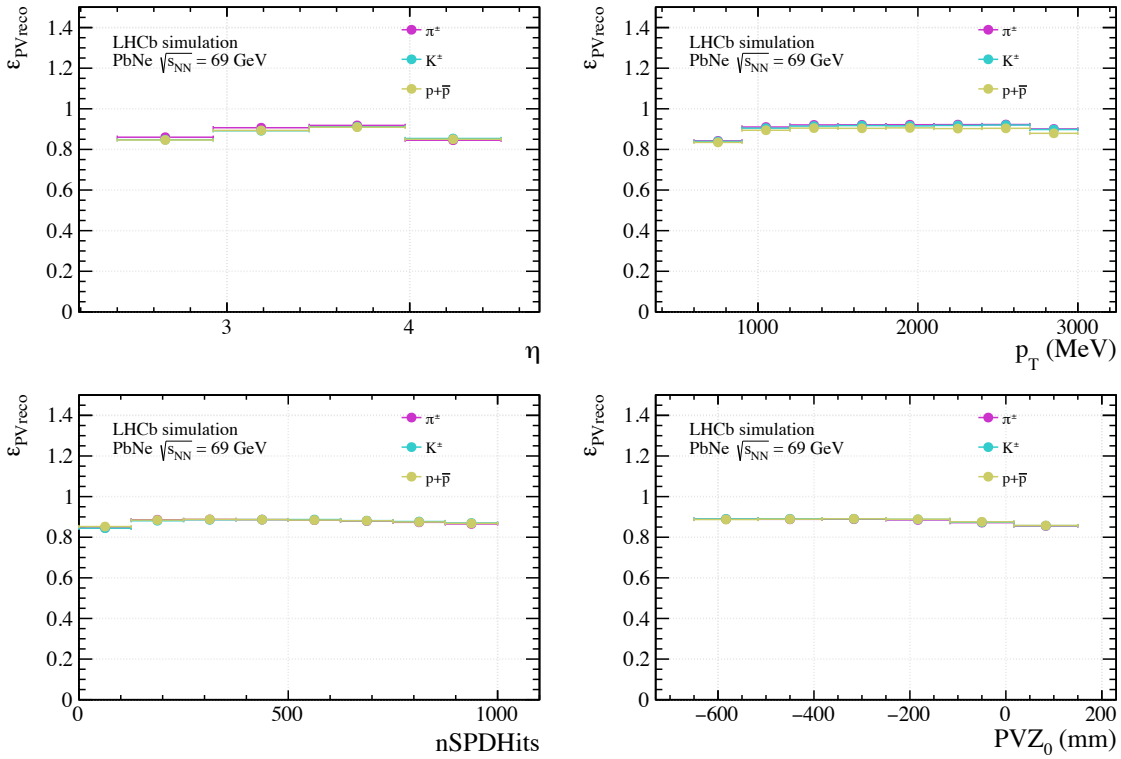


Figure 4.23: Primary vertex reconstruction efficiency calculated independently for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ from reweighed PbNe simulation as a function η, p_T, nSPDHits and PVZ.

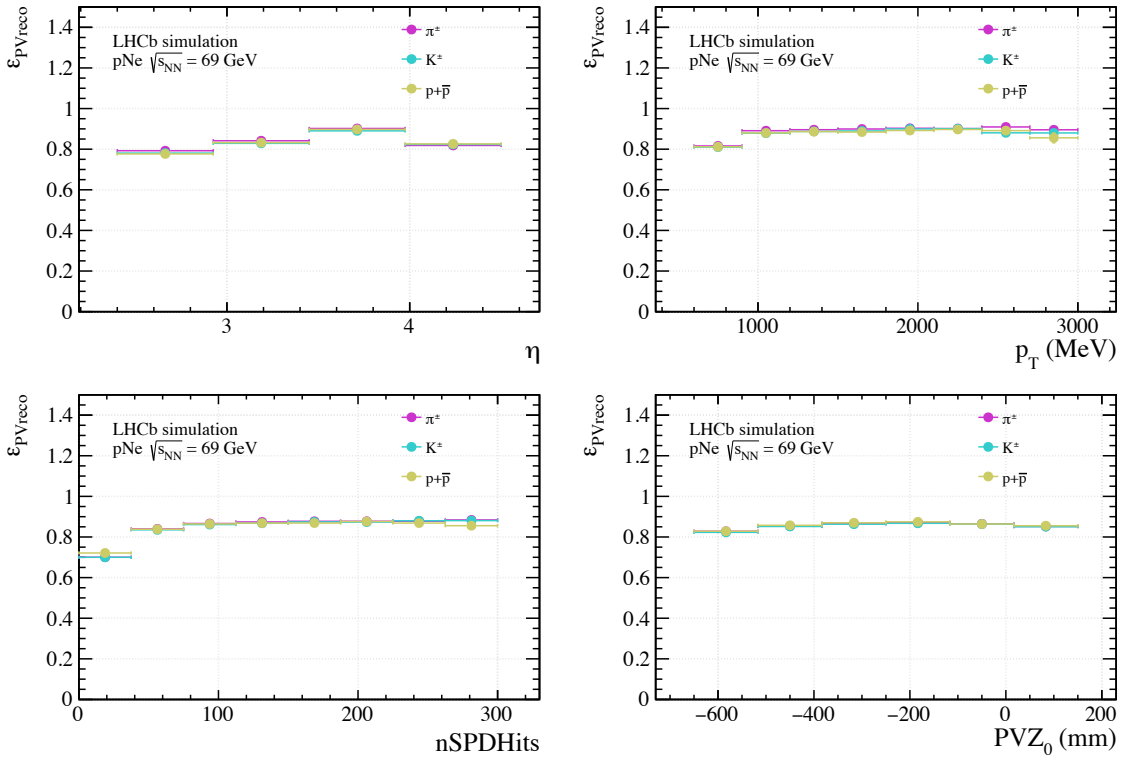


Figure 4.24: Primary vertex reconstruction efficiency calculated independently for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ from reweighed pNe simulation as a function η, p_T, nSPDHits and PVZ.

4.5.4 Selection

The selection efficiency is determined from simulation by computing the ratio of matched tracks that satisfy the selection criteria to the total number of tracks eligible to originate from a reconstructed primary vertex. This process includes the global event selection, stripping, and trigger efficiencies.

The expression for the evaluation of the selection efficiency is:

$$\varepsilon_{\text{GEC, Sel, Trig}} = \frac{N \text{ (Passing candidate selection)}}{N \text{ (nPVs>0 \& PV true-matching)}} \,. \tag{4.19}$$

The overall selection in both PbNe and pNe simulations is approximately 80% and 40%, respectively. Notably, the selection efficiency in pNe is almost half of that in PbNe. This discrepancy is primarily due to the 0.5 post-scaling factor applied to the stripping selection in pNe, as outlined in Section 4.3. Moreover, the simulation suggests a fully efficient minimum-bias trigger. To validate this, the efficiency is cross-checked using data. For the L0 requirement (nSPDHits > 0), selected candidates triggered at the L0 level by both the sequencer trigger and the NoBias trigger exhibited consistent inefficiency values. NoBias triggers passing the L0 requirement in the first fill confirmed that the HLT requirement is fully efficient. In conclusion, MB trigger inefficiency is considered negligible.

The distributions of $\varepsilon_{\text{GEC, Sel, Trig}}$, assessed independently for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ as functions of η, p_T, nSPDHits, and PVZ, are depicted in Figures 4.25 and 4.26 for PbNe and pNe, respectively. The dependence on PVZ and multiplicity is mainly related to the GEC applied to remove ghost-charge contamination, while the dependence on η and p_T can be attributed to the effect of the cut on the track momentum.

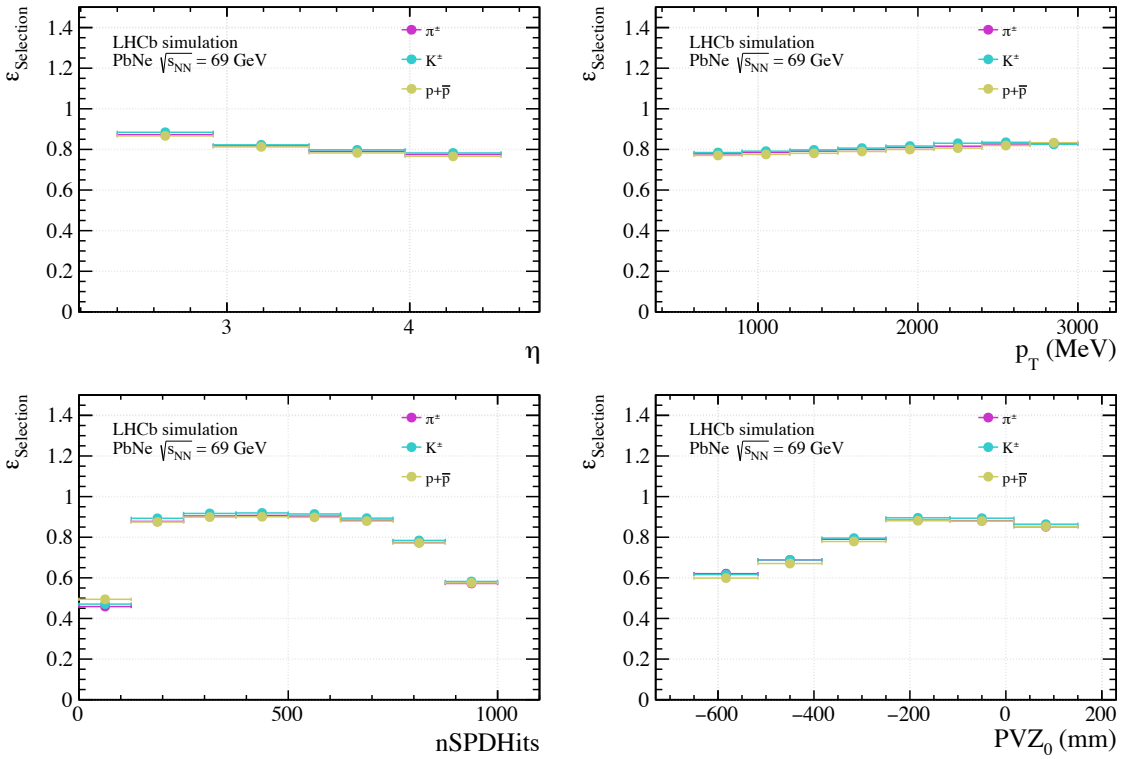


Figure 4.25: Selection efficiency calculated independently for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ from reweighed PbNe simulation as a function η, p_T, nSPDHits and PVZ.

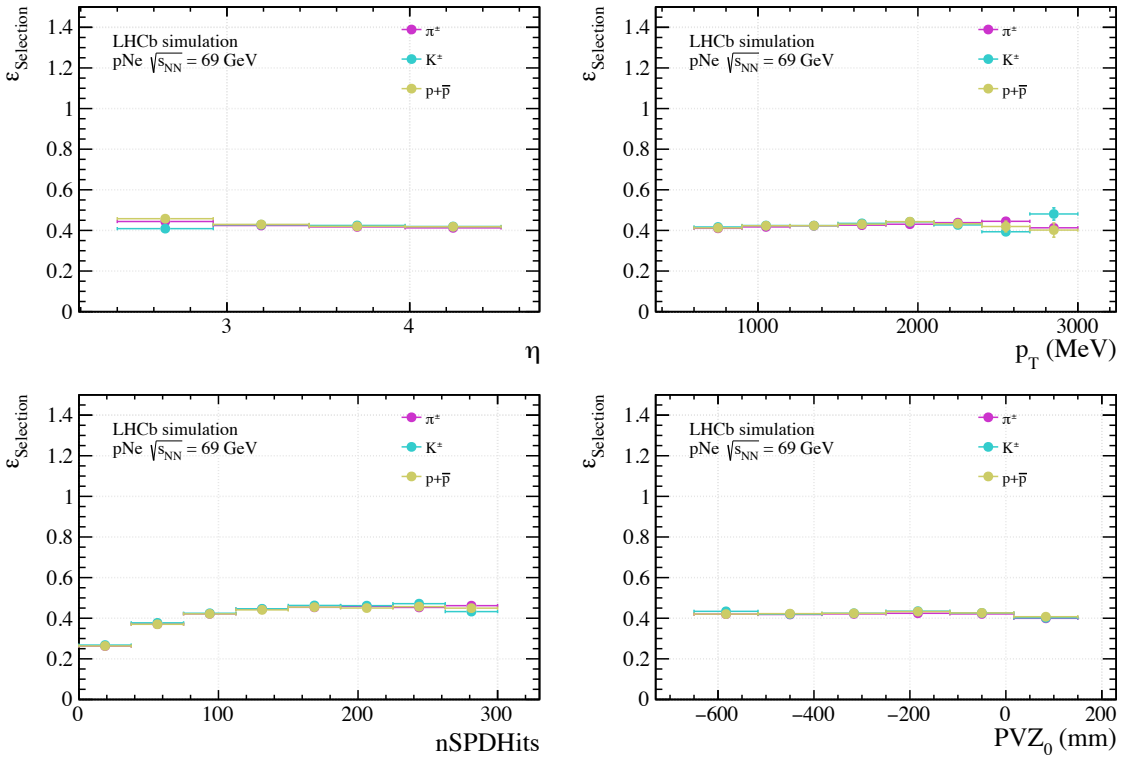


Figure 4.26: Selection efficiency calculated independently for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ from reweighed pNe simulation as a function η, p_T, nSPDHits and PVZ.

4.5.5 Total efficiency

The total efficiency is determined by combining the acceptance, tracking, PV reconstruction, and selection efficiencies outlined in the previous sections using Equation 4.15. The distributions of ε_{Tot}, evaluated independently for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ as functions of η, p_T, nSPDHits, and PVZ, are shown in Figures 4.27 and 4.28 for PbNe and pNe, respectively. The overall values of the total efficiencies for each step are summarised in Tables 4.13 and 4.14, for PbNe and pNe simulations, respectively.

Label	$\pi^{\pm}$	$K^{\pm}$	$\overset{(-)}{p}$
$\varepsilon_{\text{Acceptance}}$	98.09 ± 0.07	98.03 ± 0.12	97.55 ± 0.10
$\varepsilon_{\text{Tracking}}$	83.11 ± 0.18	82.80 ± 0.33	79.57 ± 0.25
$\varepsilon_{\text{PVreco}}$	87.61 ± 0.16	87.74 ± 0.31	87.71 ± 0.24
$\varepsilon_{\text{GEC,Sel}}$	79.75 ± 0.17	80.81 ± 0.30	79.55 ± 0.25
ε_{Tot}	56.96 ± 0.18	57.55 ± 0.33	54.16 ± 0.25

Table 4.13: Integrated efficiency values for various event selection cuts for simulated PbNe collisions, showcasing the percentage and the corresponding statistical errors of identified pions, kaons, and protons at different stages of the analysis pipeline.

Label	$\pi^{\pm}$	$K^{\pm}$	$\overset{(-)}{p}$
$\varepsilon_{\text{Acceptance}}$	98.13 ± 0.28	98.00 ± 0.89	97.50 ± 0.53
$\varepsilon_{\text{Tracking}}$	88.00 ± 0.75	87.83 ± 1.58	84.25 ± 1.53
$\varepsilon_{\text{PVreco}}$	82.74 ± 0.79	82.57 ± 1.63	83.49 ± 1.78
$\varepsilon_{\text{GEC,Sel}}$	39.00 ± 1.05	39.70 ± 1.98	39.62 ± 2.16
ε_{Tot}	27.85 ± 0.57	28.25 ± 0.98	27.17 ± 1.11

Table 4.14: Integrated efficiency values for various event selection cuts for simulated pNe collisions, showcasing the percentage and the corresponding statistical uncertanties of identified pions, kaons, and protons at different stages of the analysis pipeline.

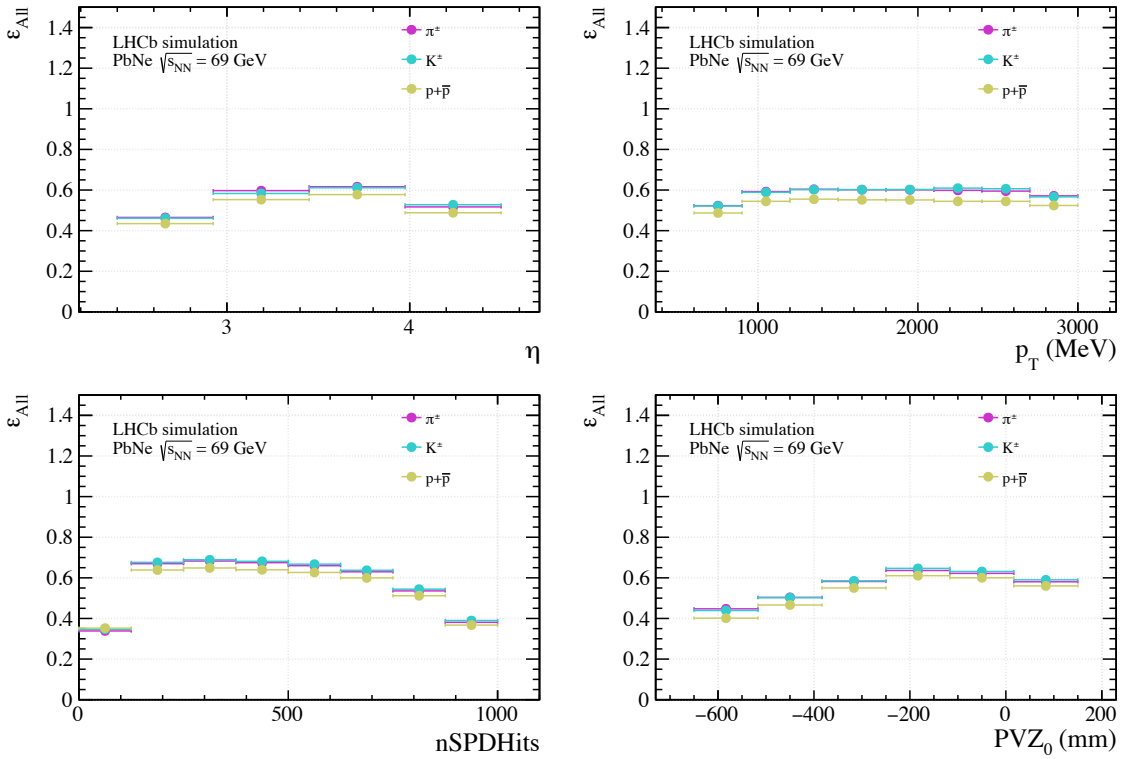


Figure 4.27: Total efficiency calculated independently for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ from reweighed PbNe simulation as a function η, p_T, nSPDHits and PVZ.

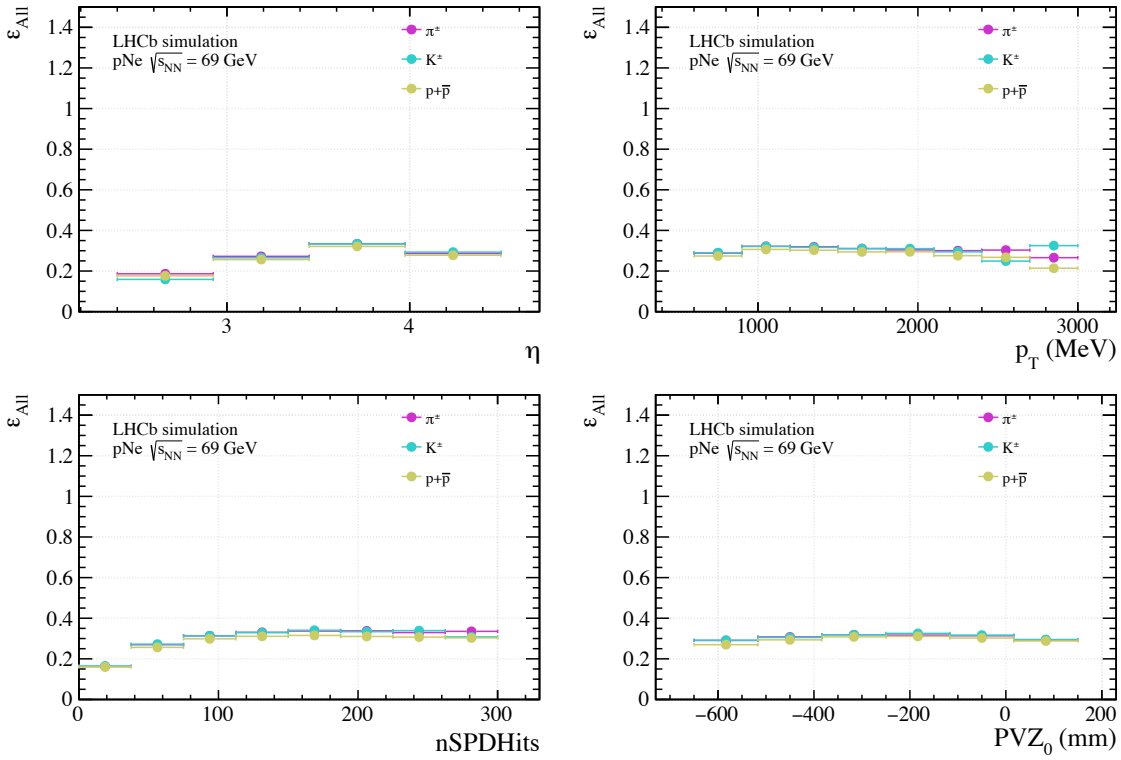


Figure 4.28: Total efficiency calculated independently for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ from reweighed PbNe simulation as a function η, p_T, nSPDHits and PVZ.

4.6 Particle identification for fixed-target events at LHCb

One of the crucial aspects in the analyses of fixed-target samples at the LHCb experiment is particle identification. In the case of pp data, high-statistics calibration channels are typically employed to validate the PID detector response. However, for fixed-target samples, which are mostly collected during short dedicated periods, their statistics are insufficient for such calibration. Additionally, the use of pp calibration channels for SMOG data is not suitable due to the poor overlap in the distributions of the quantities influencing the PID detector responses.

4.6.1 Limitations of SMOG particle identification

Particle identification in high-energy physics experiments commonly depends on a set of dedicated detectors with sufficient redundancy to distinguish different particle species. Describing the probability density function (pdf) of the PID classifier as a function of the parameters influencing the response of these detectors is essential for physics analyses. In the case of charged hadrons at the LHCb experiment, the primary reliance is on the RICH system, whose response is influenced by:

- the particle momentum, determining the Cherenkov photon emission angle according to Equation 3.1;
- the particle trajectory in the RICH gas radiator, which is a function of its origin vertex, its direction, and the quality of its reconstruction;
- the total number of hits in the RICH detectors and of the reconstructed tracks.

To illustrate the PID performance on fixed-target data, the approach used for measuring the antiproton production cross-section in pHe collisions [143] is described in the following. For this analysis, the selected sample of negatively-charged particles consists of a mixture of pions, kaons, and antiprotons, with abundances of all other species considered negligible. Instead of applying a selection, antiprotons are distinguished from other negatively-charged particles by considering the DLL(p,π) - DLL(p,K) distribution in pHe data, illustrated in Figure 4.29. To measure the antiproton abundance, a fit incorporating one template for each particle species and their relative abundances as free parameters is performed. The template distributions are drawn from pHe simulation or from the $\Lambda \to p\pi^-$, $K^0_S \to \pi^+\pi^-$, and $\phi \to K^+K^-$calibration channels, reconstructed and selected in data with no PID requirements. A fourth category related to ghost tracks, which accounts for 1.6% of the total selected particles, is added.

The simulation provides the same experimental conditions as the data and adequate statistics, but biases in the results could arise due to inevitable imperfections in the description of the detector response. This effect is mitigated by reweighing the simulation. The right plot of Figure 4.29 shows a projection of a fit in one kinematic interval with simulated templates onto a combination of the DLL variables where the three particle species can be distinguished. Imperfections in the description of the pHe data can be observed as a consequence of the small size of the calibration samples and the resulting large statistical uncertainty affecting the evaluated

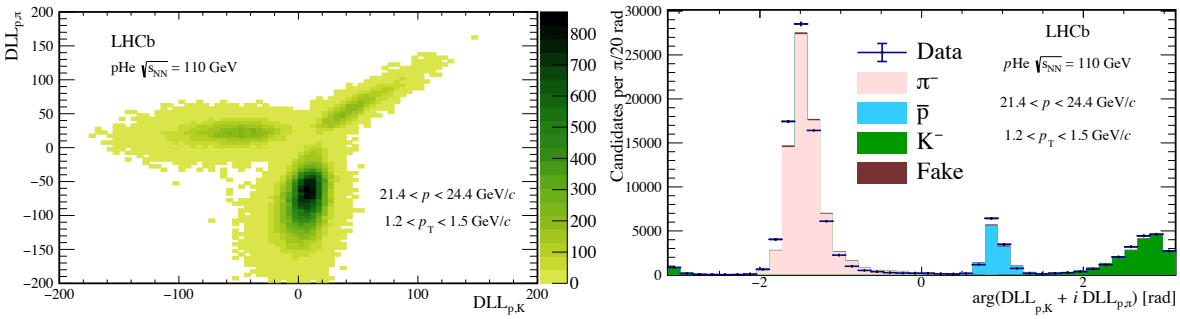


Figure 4.29: Illustration of the PID performance on fixed-target data. The left plot displays the DLL(p,π) vs DLL(p,K) distribution for all negatively-charged prompt particles selected in the pHe data. To discriminate antiprotons from kaons and pions, a fit to the pHe data is conducted, incorporating template distributions for all particle species drawn from simulation or calibration channels reconstructed and selected in the pHe data, with their relative abundances treated as free parameters. The right plot provides an example of the fit projected onto arg(DLL(p,K) + DLL(p,π)). Imperfections in the description of data are evident, arising from the limited statistics of the calibration channels.

weights. For $p > 30$ GeV/c, calibration channels reconstructed and selected in pp collisions have been considered. These channels are abundantly produced, but, due to their different distribution in the quantities affecting the response of the RICH detectors, need to be weighted as well. In particular, for the calibration channels reconstructed from pp data, the occupancy in the detector is much larger than in pHe, and there is no statistics for $|z| > 100$ mm. Hence, their application to the pHe data was found to be poorly efficient.

4.6.2 Neural Network for particle identification

The proposed solution to overcome the limitations for particle identification in fixed-target collisions, described in the previous section, is based on a data-driven Neural Network (NN) approach [153]. The fundamental concept behind the proposed model is to acquire knowledge from a data sample with high statistics regarding the marginal probability density function (pdf) of a PID classifier, denoted as x, with respect to certain pertinent experimental features represented by $\underline{\theta}$. By considering the distributions of these variables in smaller samples, a PID template distribution can be constructed, thereby addressing the constraint of inadequate calibration statistics. During the training, the x pdf in a given calibration channel is modelled as a Gaussian Mixture Model (GMM), expressed by the sum of N_g normal distributions $\mathcal{G}$ as follows

$$x_p \sim \sum_{j=1}^{N_{g,p}} \alpha_{j,p}(\underline{\theta}) \mathcal{G}(x, \mu_{j,p}(\underline{\theta}), \sigma_{j,p}(\underline{\theta})) \ , \tag{4.20}$$

where p represent the specific particle species under consideration. The selection of the normal distribution guarantees a rapid convergence of the model, and for a sufficiently large N_g, it can accurately describe any smooth profile featuring exponential-like tails, such as those observed in PID classifiers constructed at LHCb. To extend the model to a bi-dimensional PID distribu-

tion, as discussed in the previous section, the Gaussian Mixture Model (GMM) representation can be generalised to a vector target $\underline{x}$ by replacing the normal distributions with multinormal distributions:

$$\underline{x}_p \sim \sum_{j=1}^{N_{g,p}} \frac{\exp(-\frac{1}{2}(\underline{x}-\underline{\mu}_{j,p}(\underline{\theta}))^T \Sigma_{j,p}^{-1}(\underline{\theta})(\underline{x}-\underline{\mu}_{j,p}(\underline{\theta})))}{2\pi\sqrt{\det(\Sigma_{j,p}(\underline{\theta}))}} \tag{4.21}$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} . \tag{4.22}$$

For every multinormal distribution in the Gaussian Mixture Model (GMM), the free parameters comprising central values μ, standard deviations (σ_1, σ_2), correlation coefficient ρ, and the relative weight in the combination α, demonstrate a non-trivial evolution with the features $\underline{\theta}$. This dynamic behaviour is the learning objective for the machine-learning model. Each parameter is determined through a series of Multi-Layer Perceptron (MLP) NN algorithms, which are fed with the feature values $\underline{\theta}$. The internal parameters of the NN are iteratively determined by training on a calibration data set n_p. A schematic idea of the architecture of the model is presented in Figure 4.30.

The loss function, a distinctive aspect of the described approach, is defined as the negative of the log-likelihood

$$\mathscr{L} = -\sum_{i=1}^{n_p} w_i \log\left[\alpha_{j,p}(\underline{\theta}_i)\mathscr{G}(\underline{x}_i, \mu_{j,p}(\underline{\theta}_i), \sigma_{j,p}(\underline{\theta}_i))\right] , \tag{4.23}$$

where w_i are additional weights that can be obtained through the sPlot technique [154] to remove potential background contributions that can affect the considered calibration channel. In this case, the resulting weights for the signal hypothesis (w_i) act as multiplicative factors in the loss function [155, 156, 157]. The GMM parameters adapt to the n_p entries in calibration data and $x_p(\underline{\theta})$ is learned through the minimisation of $\mathscr{L}$ over multiple processing iterations (epochs).

Prior to minimising the loss function, x and $\underline{\theta}$ undergo preprocessing using the scikit-learn [158] MinMaxScaler and QuantileTransformer algorithms, respectively. Given that the model aims to predict the x pdf, the MinMaxScaler applies scaling to the [0, 1) range to make it more suitable for the NN. The QuantileTransformer is a non-linear transformation that maps the $\underline{\theta}$ distribution in the calibration data to a Gaussian. Equalising the range and functional form of the pdf for the features facilitates the numerical evaluation of derivatives in the loss minimisation, resulting in a significant speed-up of the training phase. With the processed variables, a simple NN model with a few layers and a low number of nodes, activated with a tanh function, is considered.

The minimisation occurs in two steps using mini-batch gradient descent. Initially, the x pdf is assumed to be independent of $\underline{\theta}$, and the NN parameters are determined with a low number of epochs. Subsequently, the obtained results are used as input for the second step, where $x(\theta)$ is learned with a larger number of epochs.

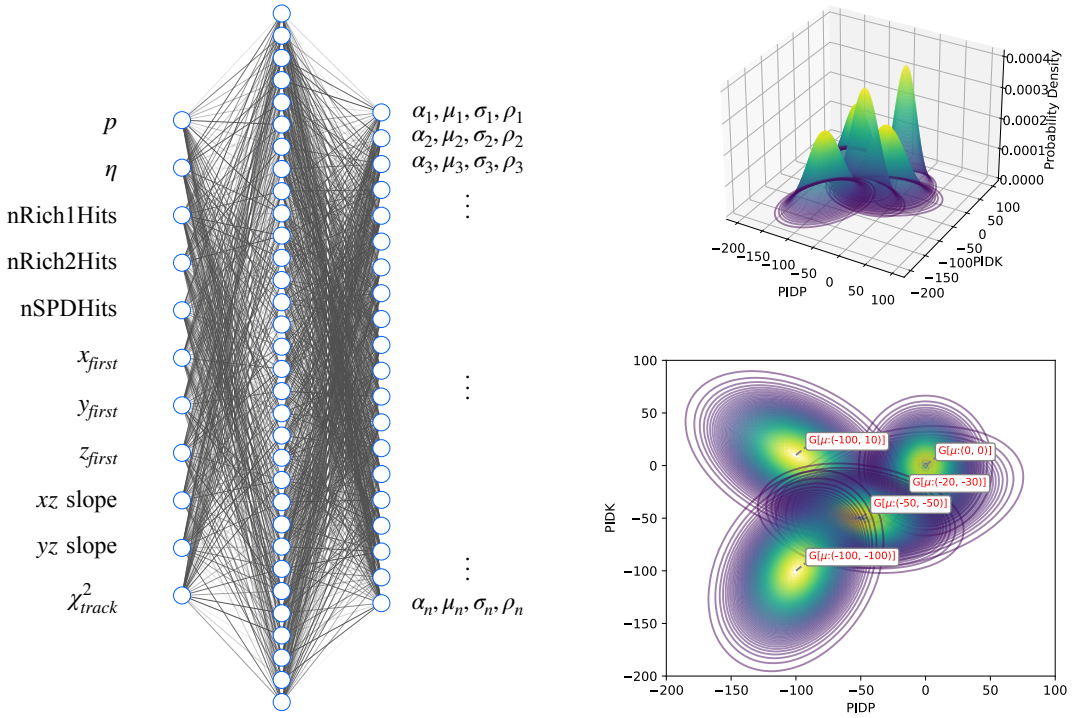


Figure 4.30: Schematic representation of the proposed GMM model used to replicate the bi-dimensional marginal probability density function (pdf) for modelling the PID distribution in fixed-target collisions. (Left) Schematic view of the architecture of the model: the NN comprises a set of fully connected layers with one hidden layer, taking a few relevant input features for particle identification and optimising the parameters $(\alpha_n, \mu_n, \sigma_n, \rho_n)$ of a set of bi-dimensional Gaussians. (Right) Illustration of the model's output, featuring 5 Gaussians with their corresponding projection on the PID plane.

Considering that ideally the model learns the dependence of the PID response on specific variables, in principle only one GMM model trained for each particle type is required to reproduce the PID response for any given dataset. However, in practice, the detector response may vary between different data-taking periods and under different detector conditions, such as high pile-up or high detector occupancy. Considering the substantial differences in event topology between the PbNe and pNe datasets collected in different years, the strategy for this analysis involves training two sets of three $(\pi^{\pm}, K^{\pm}, \overset{(-)}{p})$ GMM models each, with each set dedicated to reproducing the PID response for a specific dataset.

4.6.3 Calibration samples for fixed-target collisions

The GMM models are trained using dedicated fixed-target calibration samples obtained solely with kinematics cuts, without applying any PID selection to avoid biasing the results. Calibration samples for the PID response of $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ are reconstructed and selected in the both PbNe and pNe samples, collected respectively in 2017 and 2018 at $\sqrt{s_{NN}} = 69$ GeV. As explained in Section 4.6.2, two sets of calibration samples are needed due to the fact that PbNe dataset

has a larger coverage at high multiplicity, as shown in Figure 4.15. The PID responses for pions and protons are calibrated using $\Lambda \to p\pi^-$and $K^0_S \to \pi^+\pi^-$decays. In the case of kaons, the $D^0 \to K^-\pi^+$ decay is usually considered for pp PID calibration. However, due to low statistic for charm production in fixed-target collisions, the $\phi \to K^+K^-$process is utilised instead. For all calibration channels, the selection is performed in two steps, as summarised in Tables 4.15 and 4.16.

Particles reconstructed by combining information from all LHCb tracking detectors and with a track fit $\chi^2/\text{ndf} < 5$ are associated with final-state hypotheses and accepted if they meet loose kinematic thresholds. For the decays involving Λ and K^0_S, final-state particles are required to be well detached from the collision vertex, with the condition $\chi^2_{\text{IP}} > 25$. To reduce the combinatorial background, a selection is made within an interval around the nominal mass of the parent particle. The decaying vertex is also required to have a low χ^2 and be located upstream of the LHCb magnet, reducing contamination from long-lived particles that decay downstream of the magnet and thus have low-quality tracking and PID performance.

Contamination between the $K^0_S \to \pi^+\pi^-$and $\Lambda \to p\pi^-$calibration channels, arising from misidentifying pions as protons and vice versa, is addressed. The invariant mass of two-track combinations is re-evaluated under the hypothesis of the correct mass for the negatively-charged final-state particle, and a region around the nominal mass values of K^0 or Λ is vetoed.

For the $\phi \to K^+K^-$calibration channel, where there is a larger background contamination, one of the two kaons is selected with tighter criteria to enhance the fraction of signal candidates. The tagged kaon is required to have a low probability of being reconstructed from spurious energy deposits (GhostProb) and a high probability of being recognized as a kaon (PROBNNK > 0.75). Considering that for PbNe a larger combinatorial background is expected due to the larger event multiplicity, an even tighter selection (PROBNNK > 0.95) is applied to increase the purity. Thanks to this selection, the purity of the PbNe sample becomes similar to the one in pNe. However, the limiting factor for PbNe is the lower statistics.

A more detailed description of the decays is then performed to reduce their background contamination. The same global event selection applied for candidates is applied in order to mitigate the ghost-charge contamination. Considering that the presented approach learns the PID dependence on all variables and takes into account their correlations, the residual background is not expected to impact the results but rather to reduce the overlap between the calibration and application phase-spaces.

The mass windows and kinematic thresholds are tightened compared to the first selection, with the minimum momentum and transverse momentum set to 12 GeV/c and 400 MeV/c, respectively. All probe particles are required to be associated with signals in RICH system. For the $K^0_S \to \pi^+\pi^-$and $\Lambda \to p\pi^-$calibration channels, discrimination between signal and background is achieved through the analysis of the Armenteros-Podolanski plot [159]. For example, for the

Decay	Selection
$K^0_S \to \pi^+\pi^-$	π^+,π^- : $p > 2$ GeV/c, track χ^2/ndf $< 5, \chi^2_{\text{IP}} > 25$ comb. : $M < 1$ GeV/c^2, DV$\chi^2 < 16$ K^0_S : $\left\|M - M\left(K^0_S\right)\right\| < 50$ MeV/c^2, DV$_z < 2200$ mm, Λ veto
$\phi \to K^+K^-$	$p_T > 380$ MeV/c, track χ^2/ndf < 5 GhostProb < 0.025, K probability > 0.75 (pNe) > 0.95 (PbNe) $\|M - M(\phi)\| < 40$ MeV/c^2, DV$\chi^2 < 16$ $\|M - M(\phi)\| < 20$ MeV/c^2, DV$_z < 2200$ mm
$\Lambda \to p\pi^-$	p,π^- : $p > 2$ GeV/c, track χ^2/ndf $< 5, \chi^2_{\text{IP}} > 25$ comb. : $M < 1.5$ GeV/c^2, DV$\chi^2 < 16$ Λ : $\|M - M(\Lambda)\| < 25$ MeV/c^2, DV$_z < 2200$ mm, K^0_S veto

Table 4.15: PID calibration channels first selection stage is applied to both PbNe and pNe calibrations samples.

Decay	Selection
Event	Global event selection and kinematic ranges
$K^0_S \to \pi^+\pi^-$	π^- : $p > 12$ GeV/c, $p_T > 400$ MeV/c, RICH signals π^- : beam POCA $z \in [-700, 100]$ mm Arm.: $\left\|\left(\frac{p^{\pi}_{Transv}}{206\ \text{MeV}/c}\right)^2 + \left(\frac{\alpha}{0.83}\right)^2 - 1\right\| < 0.0275$ K^0_S : $M \in [450, 540]$ MeV/c^2
$\phi \to K^+K^-$	K^- : $p > 12$ GeV/c, $p_T > 400$ MeV/c, RICH signals K^- : beam POCA $z \in [-700, 100]$ mm ϕ : $M \in [1010, 1028]$ MeV/c^2
$\Lambda \to p\pi^-$	p : $p \in [15, 80]$ GeV/c, $p_T > 600$ MeV/c, RICH signals p : beam POCA $z \in [-700, 200]$ mm Arm.: $\left\|\left(\frac{p^{p}_{\text{Transv}}}{101\ \text{MeV}/c}\right)^2 + \left(\frac{\alpha - 0.69}{0.18}\right)^2 - 1\right\| < 0.0275$ Λ : $M \in [1100, 1150]$ MeV/c^2

Table 4.16: PID calibration channels second selection stage is applied to both PbNe and pNe calibrations samples

$\Lambda \to p\pi^-$channel the selection is defined as

$$\begin{aligned} x-\text{axis}: \alpha &= \frac{p_{\pi^-} \cdot \cos\left(\theta_{\pi^-,\Lambda}\right) - p_p \cdot \cos\left(\theta_{p,\Lambda}\right)}{p_{\pi^-} \cdot \cos\left(\theta_{\pi^+,\Lambda}\right) + p_p \cdot \cos\left(\theta_{p,\Lambda}\right)} \ , \\ y-\text{axis}: \text{pTrans} &= p_{\pi^-} \cdot \sin\left(\theta_{\pi^-,\Lambda}\right) \ , \end{aligned} \tag{4.24}$$

being $p_{\pi^-}(p)$ the pion (proton) momentum and $\theta_{\pi^-,(p),\Lambda}$ the respective angles with respect to the Λ flight direction. Depending on the particle masses involved in the decay, a unique ellipse is drawn in the Armenteros-Podolanski plot, and a high selection purity is reached.

The results of various cuts on the Armenteros-Podolanski plot for the $K^0_S \to \pi^+\pi^-$and $\Lambda \to p\pi^-$decays are presented in Figures 4.31 and 4.32 for PbNe, and in Figures 4.33 and 4.34 for pNe. This selection mainly reduces the sidebands of the invariant mass of the parent particle while keeping the signal peak essentially unchanged, allowing reaching very high purity in the calibration sample, which is a critical aspect for the training of the model. The distributions of the PID variables for different cuts are also presented. No relevant biases are observed in the PID distributions. This ensures that the selection improves the purity of the calibration sample without biasing the PID response, which is crucial to avoid introducing bias during the training of the Deep Neural Network.

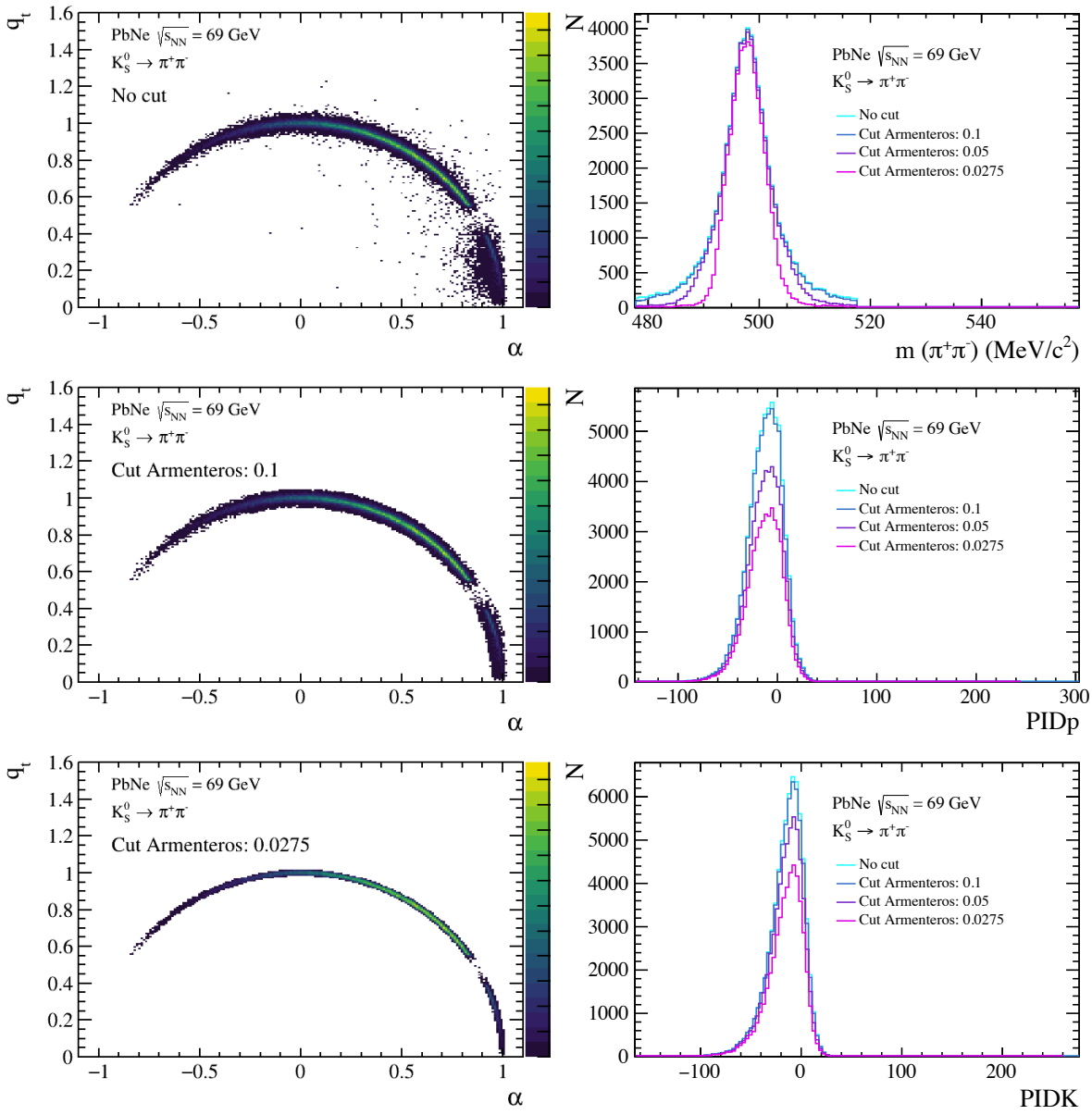


Figure 4.31: Impact of various cuts on the Armenteros-Podolanski plot for the $K^0_S \to \pi^+\pi^-$ channel in PbNe collisions: (left) distribution of the Armenteros-Podolanski variables; (right) distributions of the invariant mass (top), PIDp (middle), and PIDK (bottom) variables.

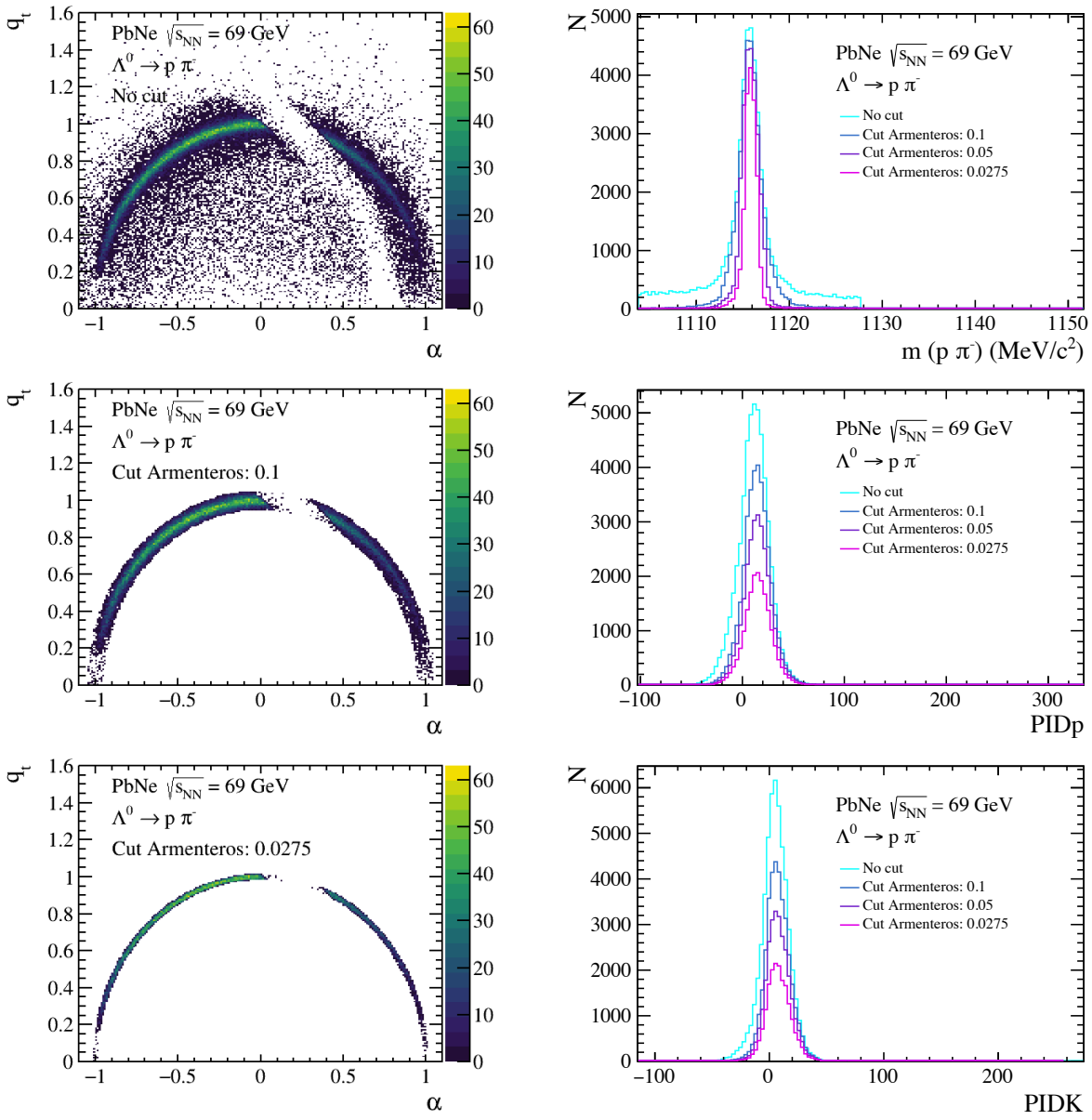


Figure 4.32: Impact of various cuts on the Armenteros-Podolanski plot for the $\Lambda \rightarrow p\pi^-$channel in PbNe collisions: (left) distribution of the Armenteros-Podolanski variables; (right) distributions of the invariant mass (top), PIDp (middle), and PIDK (bottom) variables.

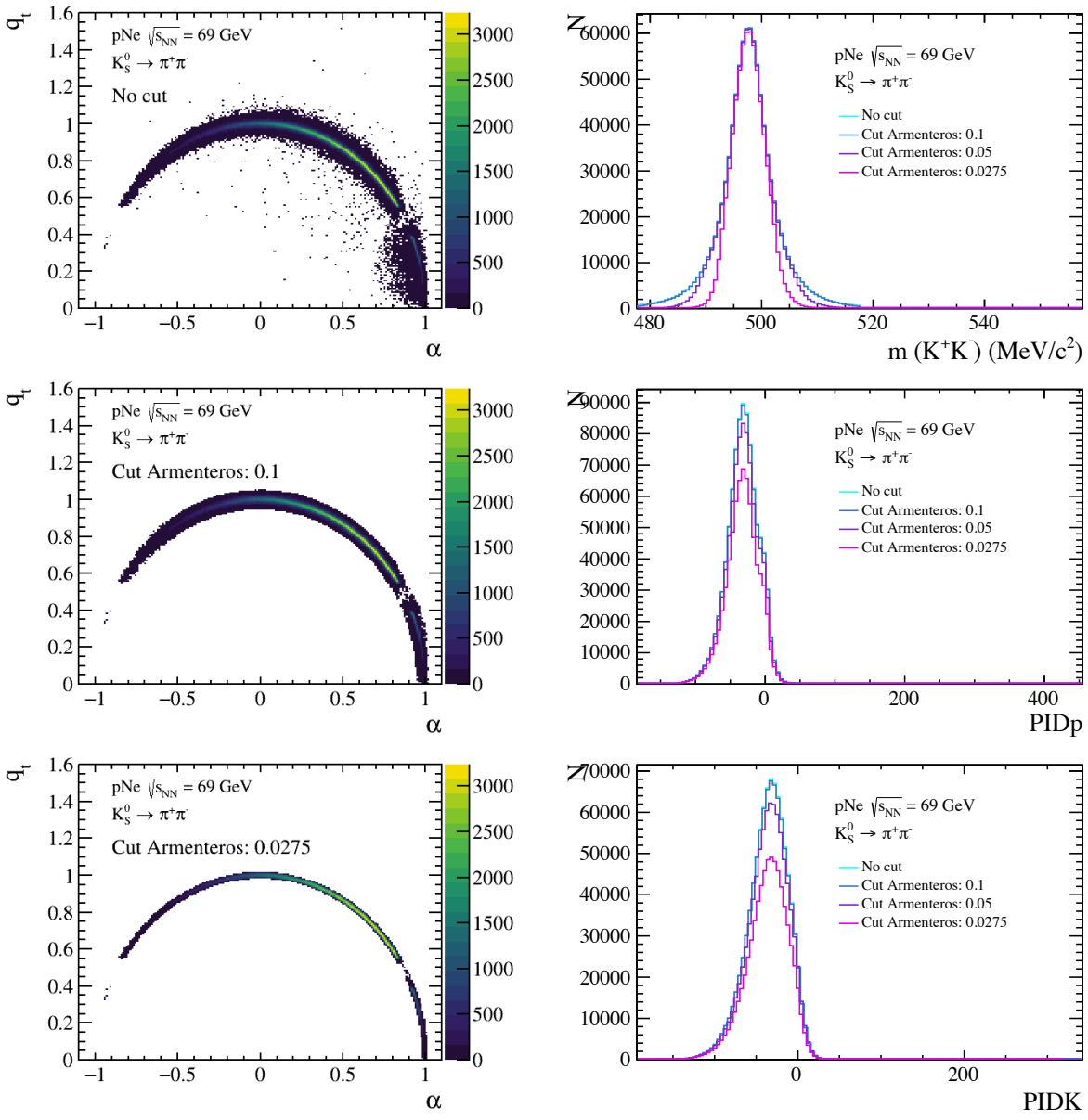


Figure 4.33: Impact of various cuts on the Armenteros-Podolanski plot for the $K^0_S \to \pi^+\pi^-$ channel in pNe collisions: (left) Distribution of the Armenteros-Podolanski variables; (right) distributions of the invariant mass (top), PIDp (middle), and PIDK (bottom) variables.

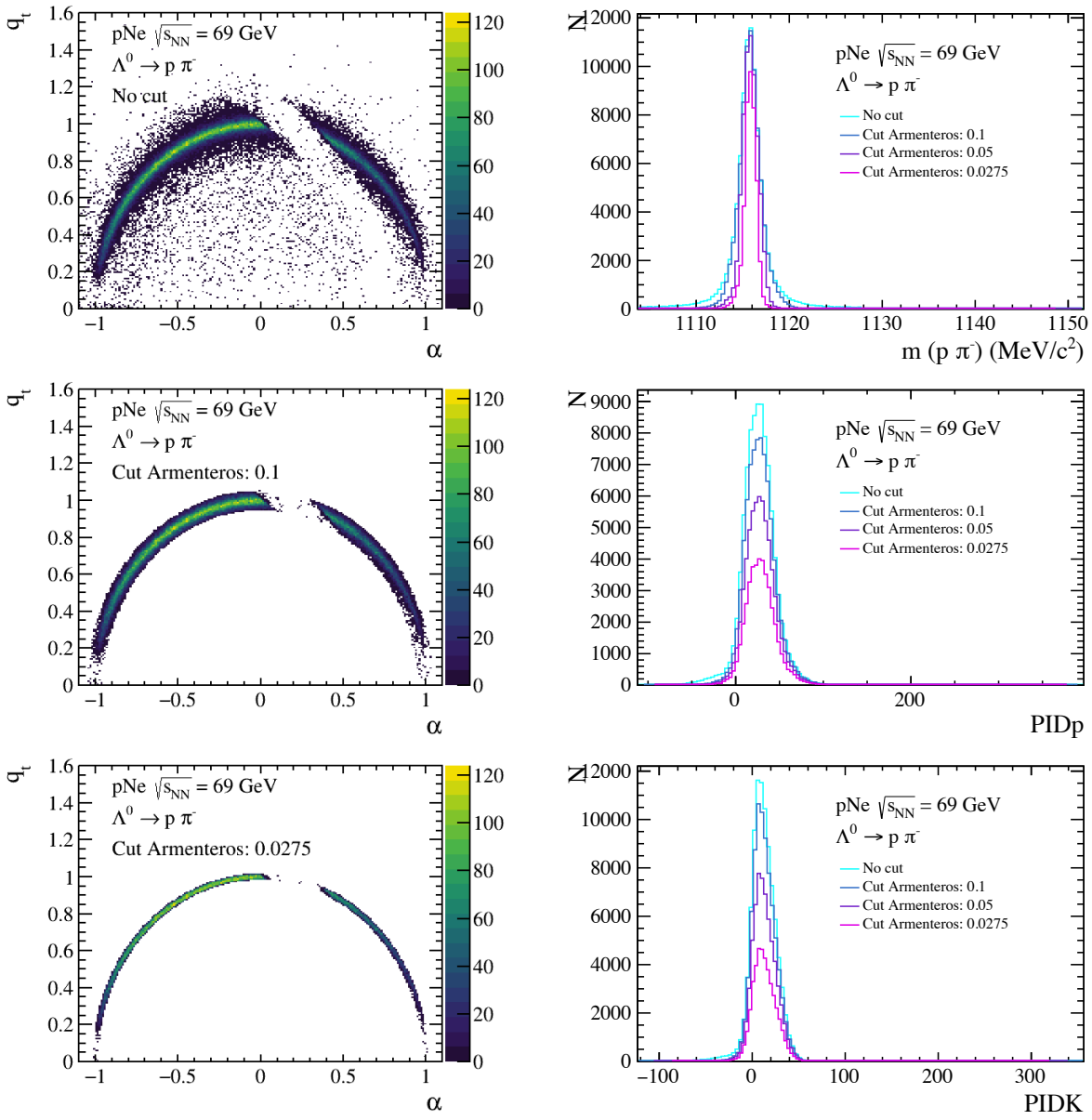


Figure 4.34: Impact of various cuts on the Armenteros-Podolanski plot for the $\Lambda \rightarrow p\pi^-$channel in pNe collisions: (left) Distribution of the Armenteros-Podolanski variables; (right) distributions of the invariant mass (top), PIDp (middle), and PIDK (bottom) variables.

For the $\phi \to K^+K^-$decay, the background is statistically subtracted using the sPlot technique. The invariant mass distribution for the ϕ candidates is modelled with the sum of a Voigtian distribution (the convolution of a Gaussian and a Breit-Wigner) for the signal and a polynomial function for the background, as depicted Figures 4.35 and 4.36 for PbNe and pNe calibratrion samples, respectively.

The validation of the extracted sWeights is conducted using two different techniques:

- **Closure test:** this test validates the sPlot formalism by checking if the number of events matches the sum of the signal and background components within the uncertainties;
- **PID discriminating variable:** the weights assigned to each event belonging to the signal or background category are validated by comparing the normalized distributions of the variable arg(PIDp + i PIDK). This variable allows for clear differentiation between the contributions of pions, kaons, and protons. The sPlot is considered validated if the signal contribution is compatible with the kaon.

Both datasets successfully pass the closure test. The test on the PID discriminating variable is also validated for both datasets, as illustrated in Figures 4.35 and 4.36. The contributions from pions and protons are clearly subtracted by applying the sPlot technique. Additionally, the weighted PID distributions for the signal are presented to demonstrate the direct effect of applying the weights to these variables. In both datasets, the dominant left-most peak, corresponding to pion background contamination, and the small peak at high values of PIDp, are suppressed, serving as an additional cross-check of the precedure.

Another crucial step for validating the data used to train the GMM models is to ensure that the kinematic coverage of the calibration sample post-selection matches that of the minimum bias data under study. Figures 4.37 and 4.38 display the distributions for charged tracks and the PID calibration channels after applying all the selections. Notably, all calibration samples exhibit a kinematic coverage that spans the entire range observed in the minimum bias data. This ensures that the model can effectively learn across the full spectrum of the analysis and accurately reproduce the PID response for the complete range.

In the $\Lambda \to p\pi^-$channel for pNe, a distinctive peak is observed in the p_T distribution at around 1 GeV/c. This occurrence is attributed to the application of scaling factors on the $\Lambda \to p\pi^-$channel to bias the selection of the decay in different p_T regions, enabling better coverage of the calibration sample. Given that the analysis involves simultaneous binning over multiple variables, it is conceivable that some regions may be inadequately covered, particularly at high p_T. This consideration contributes to limit the analysis to the $p_T < 3000$ MeV/c range.

Additionally, upon examining the PID distributions in both datasets, it is evident that the combination of the three components corresponding to samples of pure pions, kaons, and protons covers the entire range of charged tracks. This observation serves as further validation, confirming that the selected calibration samples are indeed suitable candidates for accurately reproducing the shape of each particle type in the PID plane.

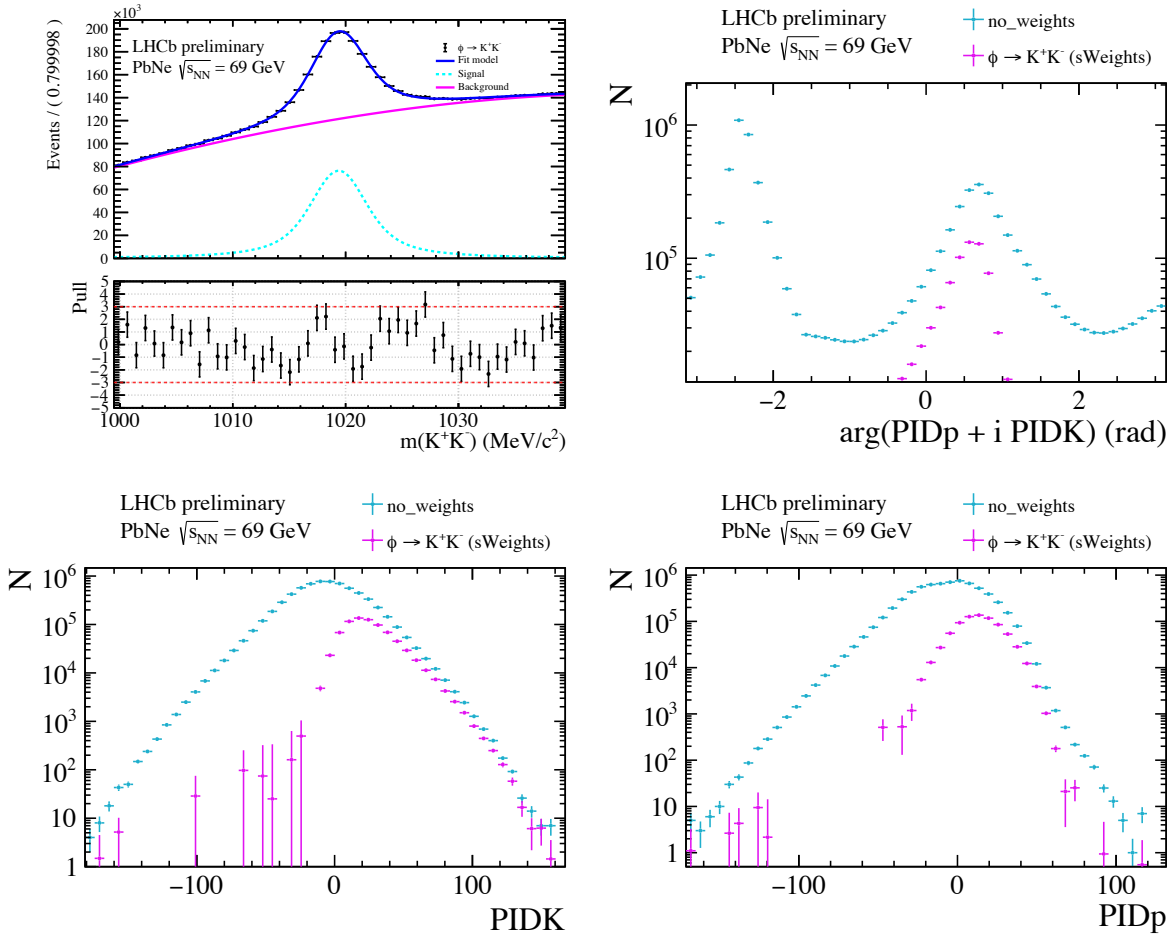


Figure 4.35: Application of the sPlot technique on the $\phi \to K^+K^-$ decays from the PbNe sample: (top-left) invariant mass fit of ϕ candidates with a combination of a Voigtian for the signal and a polynomial function for the background; (top-right) Normalised distributions of arg(PIDp + i PIDK) before (in blue) and after (in magenta) their application; (bottom) effect on PID variables of the application of the sPlot technique for $\phi \to K^+K^-$ events.

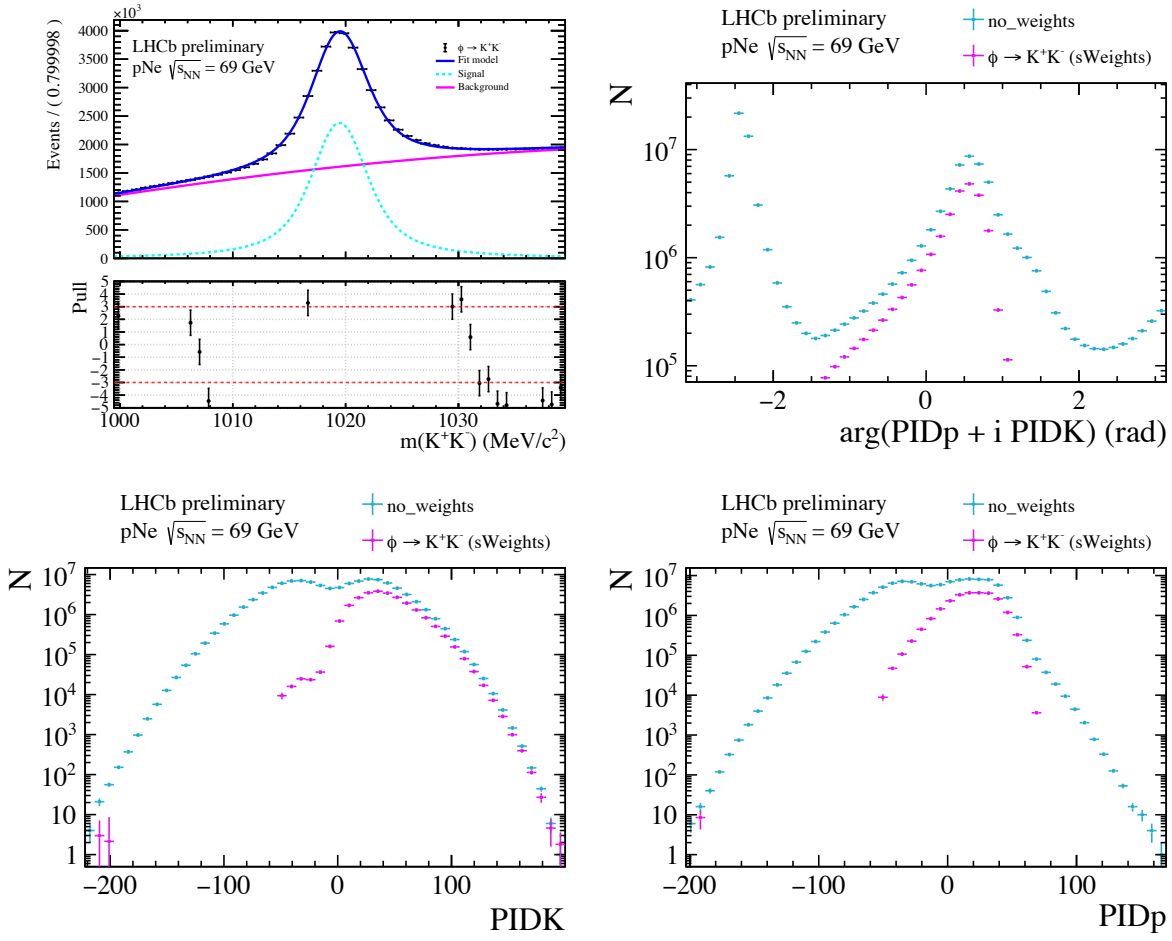


Figure 4.36: Application of the sPlot technique on the $\phi \to K^+K^-$decays from the pNe sample: (top-left) invariant mass fit of ϕ candidates with a combination of a Voigtian for the signal and a polynomial function for the background; (top-right) Normalised distributions of arg(PIDp + i PIDK) before (in blue) and after (in magenta) their application; (bottom) effect on PID variables of the application of the sPlot technique for $\phi \to K^+K^-$events.

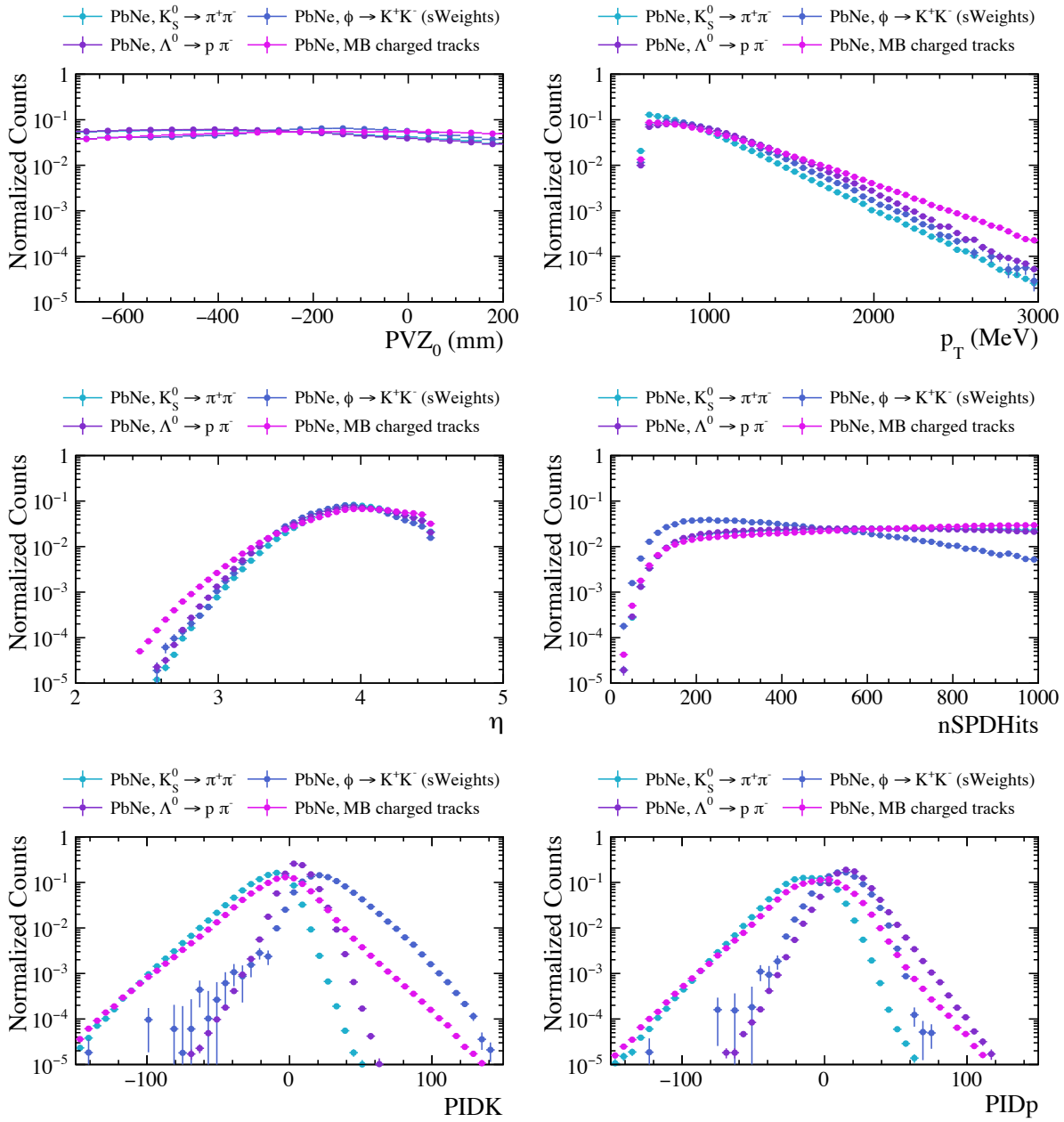


Figure 4.37: Comparison of η, p_T, nSPDHits, PVZ, PIDp and PIDK distributions for $K^0_S \to \pi^+\pi^-$, $\phi \to K^+K^-$, $\Lambda \to p\pi^-$ and MB prompt charged tracks in PbNe data.

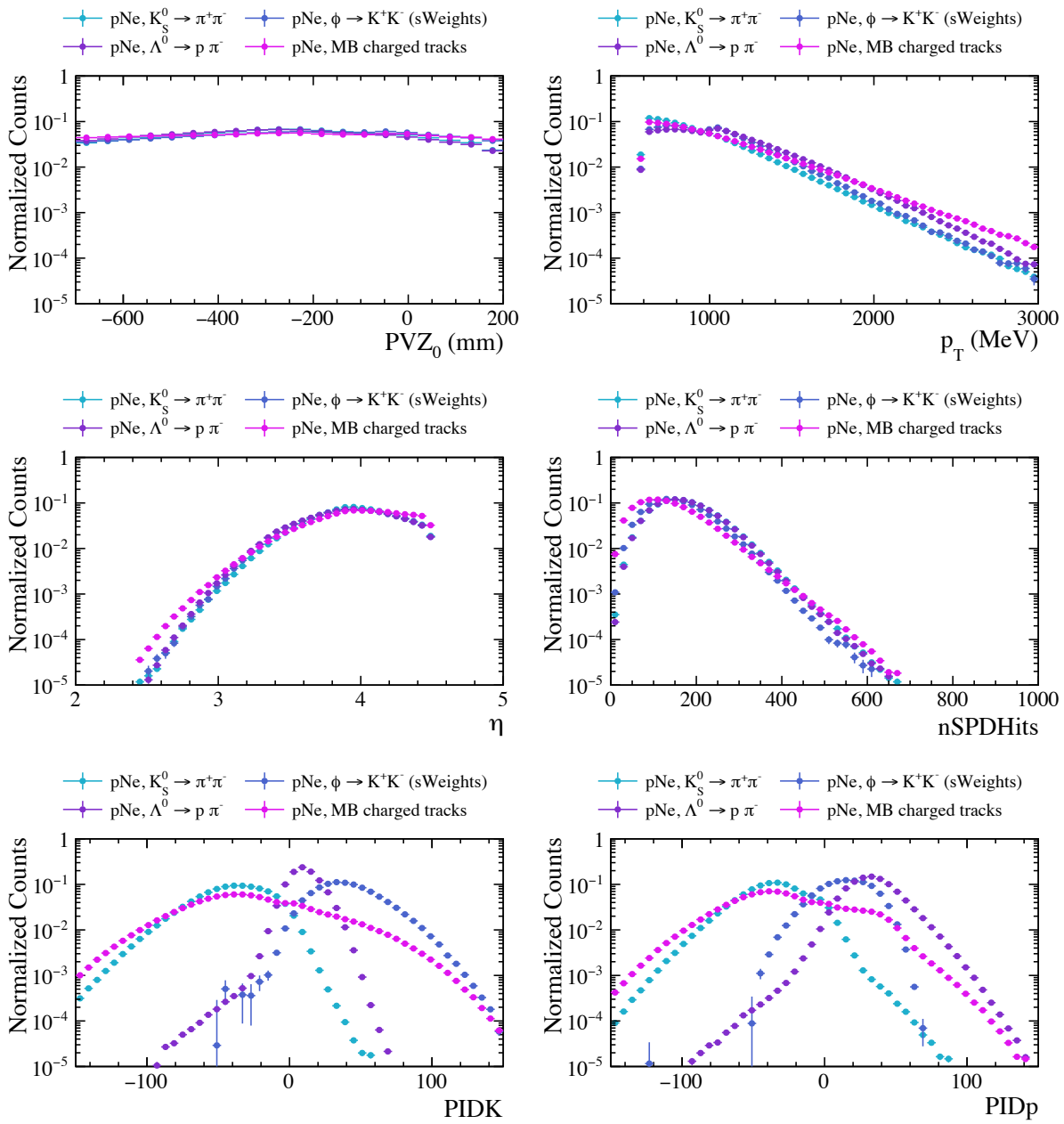


Figure 4.38: Comparison of η, p_T, nSPDHits, PVZ, PIDp and PIDK distributions for $K^0_S \to \pi^+\pi^-$, $\phi \to K^+K^-$, $\Lambda \to p\pi^-$ and MB prompt charged tracks in pNe data.

4.6.4 Feature selection

The identification of key features influencing the response of the RICH detectors is crucial for this approach. It is essential to consider an adequate number of variables to ensure a comprehensive description of the RICH response while maintaining a low model complexity for efficient hyperparameter tuning.

To prioritise variables, a method based on the Kolmogorov-Smirnov distance is employed. A comparison is made for each potential feature using the PID variables for pions in the pHe data. The ordering parameter is determined by the maximum KS distance measured between all possible histogram pairs, signifying a greater importance of the feature candidate in determining the PID variables. Table 4.17 lists the KS distances in decreasing order by iterating over a set of potential feature candidates obtained on a study on pHe calibration samples [153]. The leading features reflect:

- Particle kinematics, such as momentum (p), its longitudinal (p_z) and transverse (p_T) components, and pseudorapidity (η);
- Occupancy in the detector, expressed as the number of hits in the SPD, RICH1, and RICH2 detectors or as the number of reconstructed tracks;
- Quality of track reconstruction, exemplified by χ^2/ndf or the number of degrees of freedom (ndf) of its fit;
- Geometry of the event, such as track slopes with respect to the z-axis.

Variables listed in Table 4.17, which are not significantly correlated with others, are included in the model. For example, despite a lower KS value compared to nSPDHits, the number of hits in the RICH detectors is considered due to its connection with PID. The p_T is discarded, considering its correlation with η and p which exhibit stronger dependence. Variables related to different geometries between training and application particles are added. While accounting for different distributions of the Origin Vertex (OV) coordinates, tracks associated with training particles are extrapolated to the beam, and the coordinates of the intersection (beam POCA) are added to the set of features.

The models are applied to prompt $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$. While for prompt particles the distribution of the z coordinate of the OV coincides by definition with the selected range PVZ$\in [-700, 200]$mm, the candidates used for the training originate from a decay. Considering that PID depends on the trajectory and the kinematic of the track, this can lead non negligible differences in the PID response. To mitigate this effect, the OV coordinates for particles originating from decays are extrapolated and the coordinates of the intersection are added to the set of features.

4.6.5 Preprocessing and model training

Once the set of features influencing the RICH detector responses is identified, a separate model is trained for each calibration sample. As mentioned above, an initial preprocessing of

Variable	Max KS	Variable	Max KS	Variable	Max KS
p	0.64	p_z	0.64	η	0.54
p_T	0.51	yz slope	0.38	track ndf	0.34
xz slope	0.34	nTracks	0.34	nRICH2Hits	0.33
nSPDHits	0.32	nRICH1Hits	0.28	track χ^2/ ndf	0.26

Table 4.17: Maximum Kolmogorov-Smirnov distances for bidimensional PID distribution comparisons, listed in decreasing order of relevance. A higher indicator value signifies greater importance of the feature to the PID [153].

Training dataset	Input parameter	$K^0_S \to \pi^-\pi^+$	$\phi \to K^-K^+$	$\Lambda \to p\pi^-$
PbNe	Number of Gaussians	64	32	32
	Number of NN nodes	64	64	64
	Starting learning rate	$5\cdot10^{-4}$	$5\cdot10^{-4}$	10^{-4}
	Batch size [events]	20000	40000	20000
pNe	Number of Gaussians	64	64	20
	Number of NN nodes	64	64	64
	Starting learning rate	10^{-3}	$5\cdot10^{-6}$	10^{-4}
	Batch size [events]	10000	20000	10000

Table 4.18: Values of the input parameters for the model training of the three calibration channels

all variables is conducted. The results for the η, p and nSPDHits for reconstructed and selected $K^0_S \to \pi^+\pi^-$ decays are illustrated in Figure 4.39. Despite having different input functional forms, all features are transformed to follow normal distributions.

The GMM is then constructed using the input parameters for the three calibration channels as indicated in Table 4.18. The same architecture is used for both PbNe and pNe. A higher level of complexity is required for the K^0 and ϕ decays due to the threshold effects involved in the classifier definition, leading to multimodal PID distributions. This led to the choice of increasing the number of gaussians in these two calibration samples. In the case of the ϕ calibration channel, the application of sPlot weights (w_i) implies that only a fraction of the events contribute to the minimisation of the loss function. To counterbalance this limitation, a larger batch size is set.

The parameters of the GMM are randomly initialised according to uniform distributions within the ranges:

- $\left[\langle \underline{x}\rangle - \frac{1}{2}\sqrt{\langle \underline{x}^2\rangle - \langle \underline{x}\rangle^2}, \langle \underline{x}\rangle + \frac{1}{2}\sqrt{\langle \underline{x}^2\rangle - \langle \underline{x}\rangle^2}\right]$ for the mean values $\underline{\mu}$, indicating by $\langle\cdot\rangle$ the average. For the ϕ calibration channel, the background-subtracted distribution is considered;
- $\left[\frac{1}{10}\sqrt{\langle \underline{x}^2\rangle - \langle \underline{x}\rangle^2}, \sqrt{\langle \underline{x}^2\rangle - \langle \underline{x}\rangle^2}\right]$ for the widths $\sigma_{1,2}$;
- $[0, 2\pi]$ for the correlation angle, defined as $\omega = \arcsin(\rho)$;
- $[0, 1]$ for the component weights α_j.

The initial training stage is conducted with this initialisation only. Subsequently, for each parameter of the Gaussian Mixture Model (GMM), the output of a Neural Network (NN), fed with the

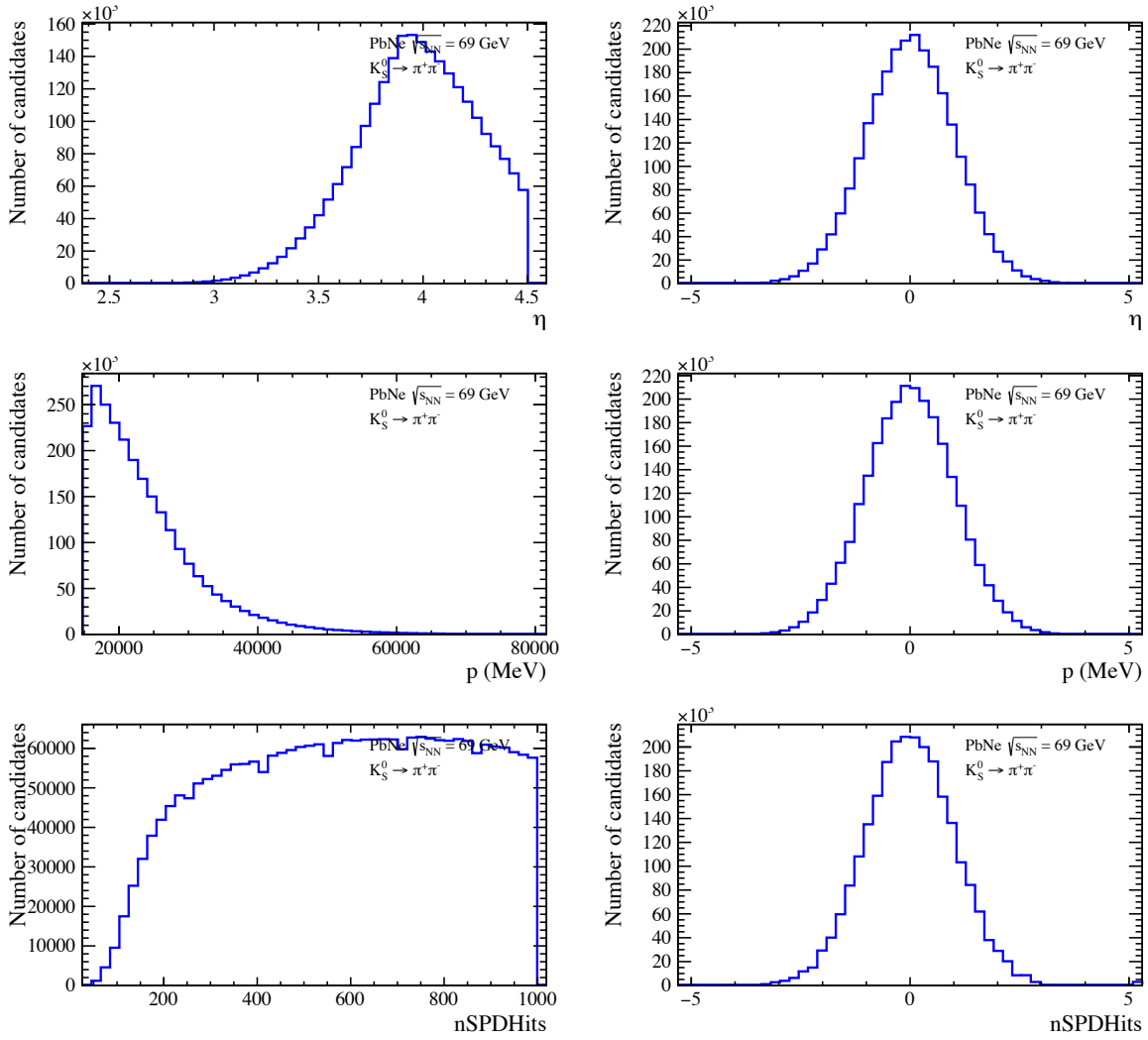


Figure 4.39: Pseudorapidity, momentum and nSDPHits distributions before (left) and after (right) the application of the QuantileTransformer preprocessing algorithm for the negative pions in reconstructed and selected $K^0_S \to \pi^+\pi^-$ decays in the 2018 PbNe data.

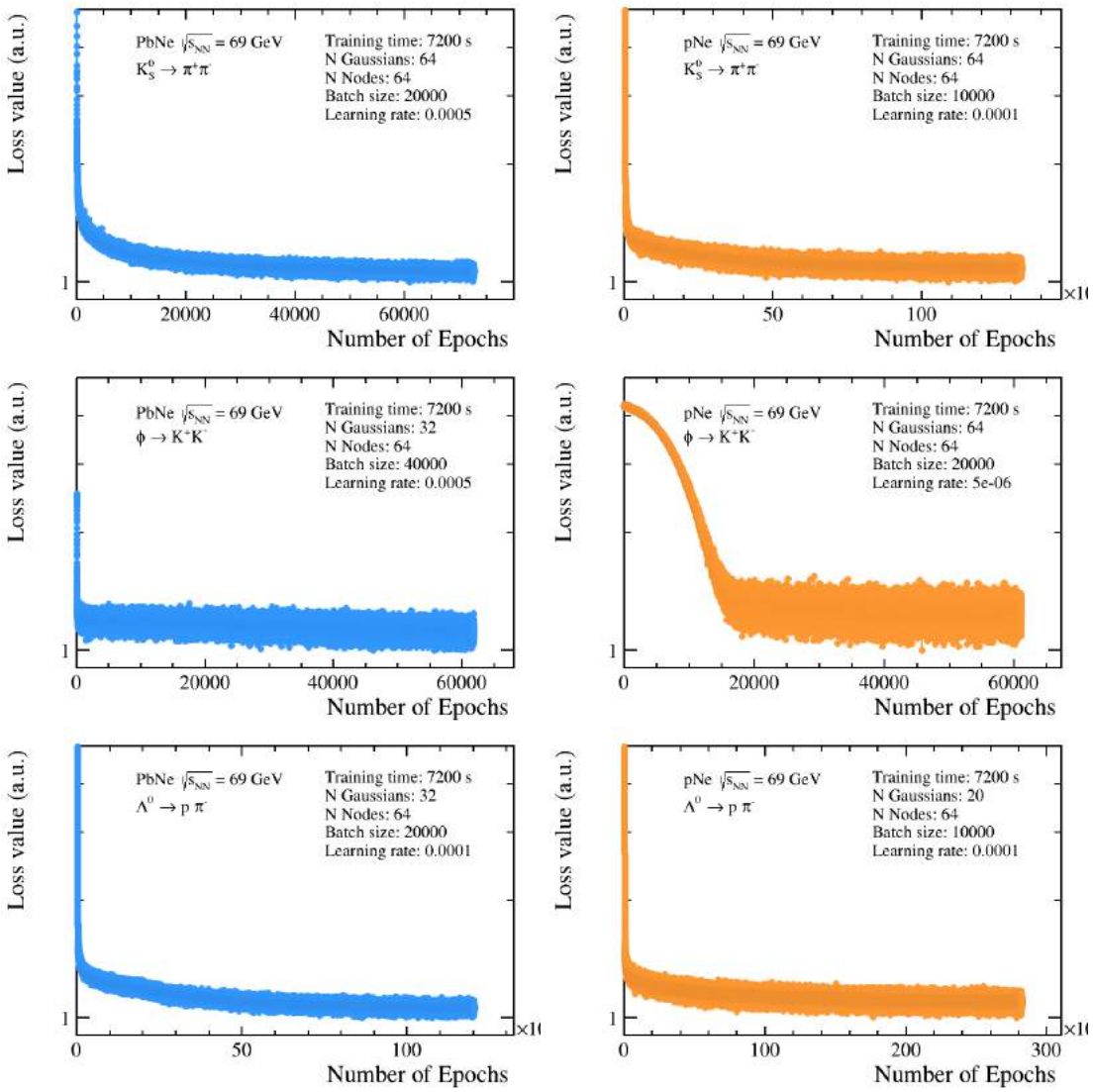


Figure 4.40: Evolution of the loss function with the number of epoch for GMM models trained on $K^0_S \to \pi^+\pi^-$, $\phi \to K^+K^-$and $\Lambda \to p\pi^-$calibration samples for PbNe (left) and pNe (right) data.

feature values and constructed using the parameters in Table 4.18, is incorporated. The number of training parameters for the three calibration decays is approximately O(10^5), depending on the chosen complexity for the NN and the GMM. Even with different input configurations, the loss minimisation procedure converges in a few hours. The evolution of the loss function with the number of epochs is illustrated in Figure 4.40 for GMM models trained on $K^0_S \to \pi^+\pi^-$, $\phi \to K^+K^-$and $\Lambda \to p\pi^-$calibration samples on both PbNe and pNe data. For all calibration channels, the loss function exhibits a steep decrease initially, followed by a smoother decline, and then a gentle oscillation towards the minimum. The learning rate choice is validated by the absence of abrupt variations of the curves.

4.6.6 Overtraining and validation

The subsequent steps after the training include several checks to validate the accuracy and reliability of the model. In contrast to other machine-learning problems, notably the classification ones, the overtraining is not a key aspect to monitor in this case considering that the model is always applied to other samples than the training one. Nevertheless, in corner regions of the hyper-space, statistical fluctuations could induce fast variations of the model parameters to reproduce single data entries or clusters. To exclude this effect, the evolution of the model parameters as a function of the feature variables is controlled.

As an example, the distributions of the Gaussians μ_n and σ_n as functions of p and nSPDHits are shown in Figures 4.41, 4.42, and 4.43 for GMM models trained on $K^0_S \to \pi^+\pi^-$, $\phi \to K^+K^-$, and $\Lambda \to p\pi^-$in PbNe data. Similar distributions are presented for pNe in Figures 4.44, 4.45, and 4.46. These distributions offer relevant insights into the training of the model and the correct choice of hyperparameters. First of all, the smooth evolution of all the curves, corresponding to the multinormal components in the GMM, indicates the regularity of the training. Another important aspect is that the dependence of the parameters on the input features is significantly different between models trained on different datasets for the same calibration line. This fact underlines the importance of having two distinct sets of GMMs to model the PID of PbNe and pNe, rather than a single set of models trained on a single dataset.

The model can undergo a direct validation by comparing its predictions with the expected PID distributions of the calibration sample. Specifically, the model must successfully learn to reproduce the PIDp and PIDK variables across the entire considered multi-dimensional feature hyper-space. To assess this, each model undergoes separate validation in equally-populated bins of all pairs of the considered features. In each interval, calibration events are randomly selected, and a bi-dimensional PID template is generated based on their feature distribution. The model passes validation if the two distributions agree in all regions of the feature hyper-space.

Examples of such comparisons for the $K^0_S \to \pi^+\pi^-$, $\phi \to K^+K^-$and $\Lambda \to p\pi^-$calibration channels with different feature pairs are presented in Figures 4.47 and 4.48 for models trained on PbNe and pNe data, respectively. In all figures, bi-dimensional comparisons and projections onto the two axis are displayed. A good match between the data and the distributions generated by the trained models is consistently observed. The agreement between the two distributions also in low statistics regions underscores the model's ability to create a smooth template based on the available information, overcoming limitations posed by scarce calibration data, and without significant overtraining.

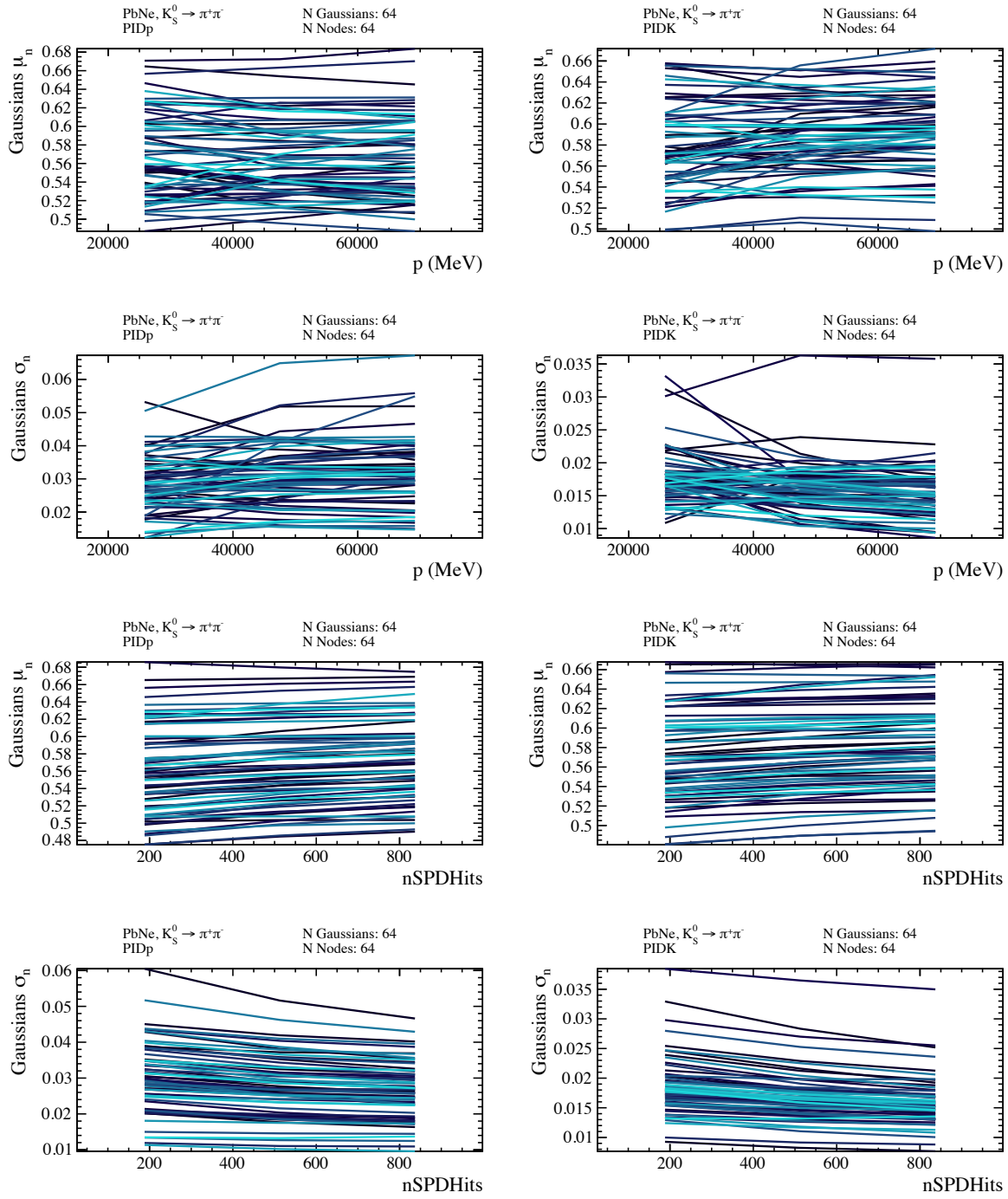


Figure 4.41: Check against a possible overtraining effect of the model trained on the $K^0_S \to \pi^+\pi^-$calibration channel for PbNe. The evolution of the Gaussian mean values and widths for the PIDp (left) and the PIDK (right) variables are presented as functions of p (top) and nSPDHits (bottom). The different colors represent the Gaussian components included in the model.

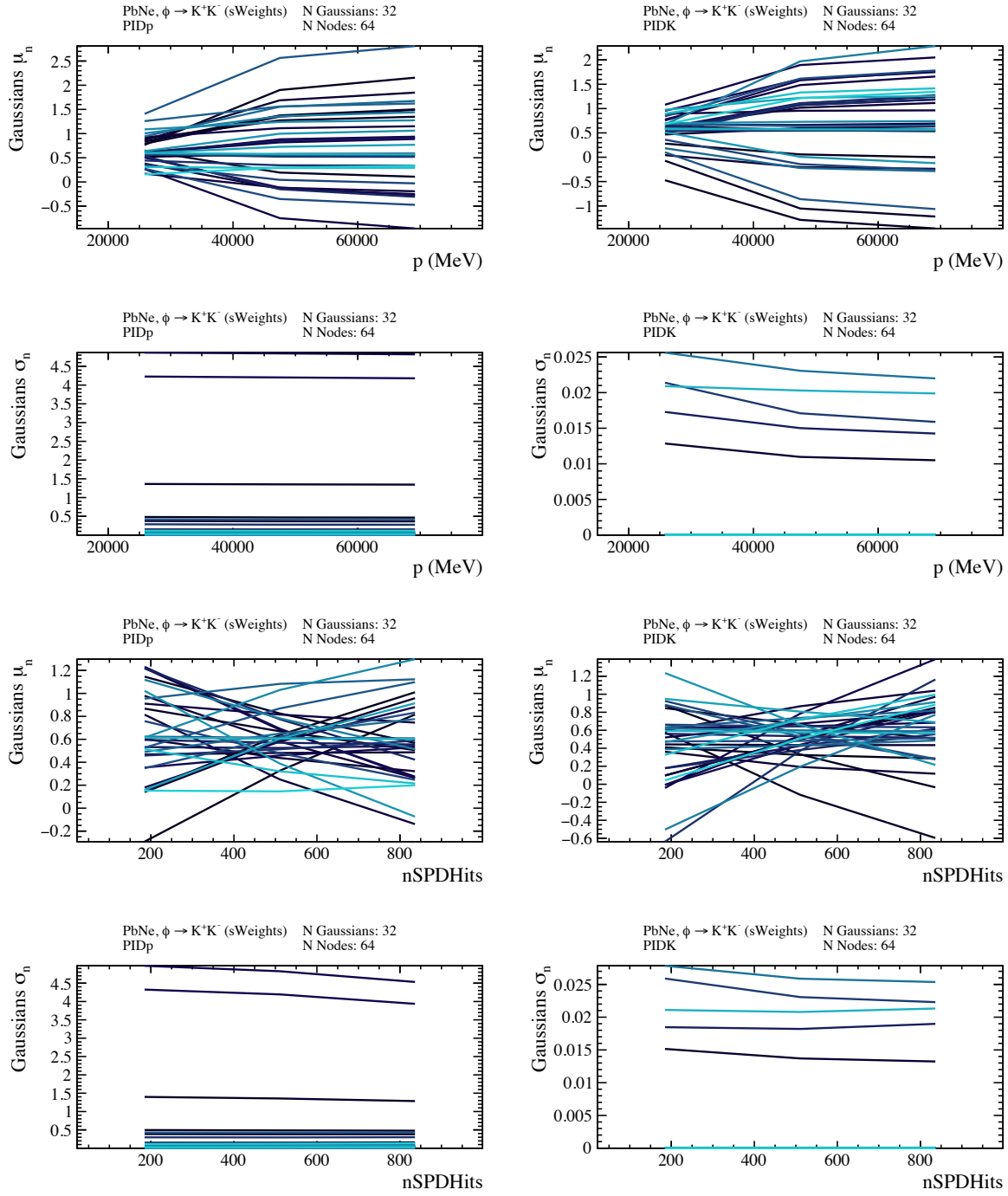


Figure 4.42: Check against a possible overtraining effect of the model trained on the $\phi \to K^+K^-$calibration channel for PbNe. The evolution of the Gaussian mean values and widths for the PIDp (left) and the PIDK (right) variables are presented as functions of p (top) and nSPDHits (bottom). The different colors represent the Gaussian components included in the model.

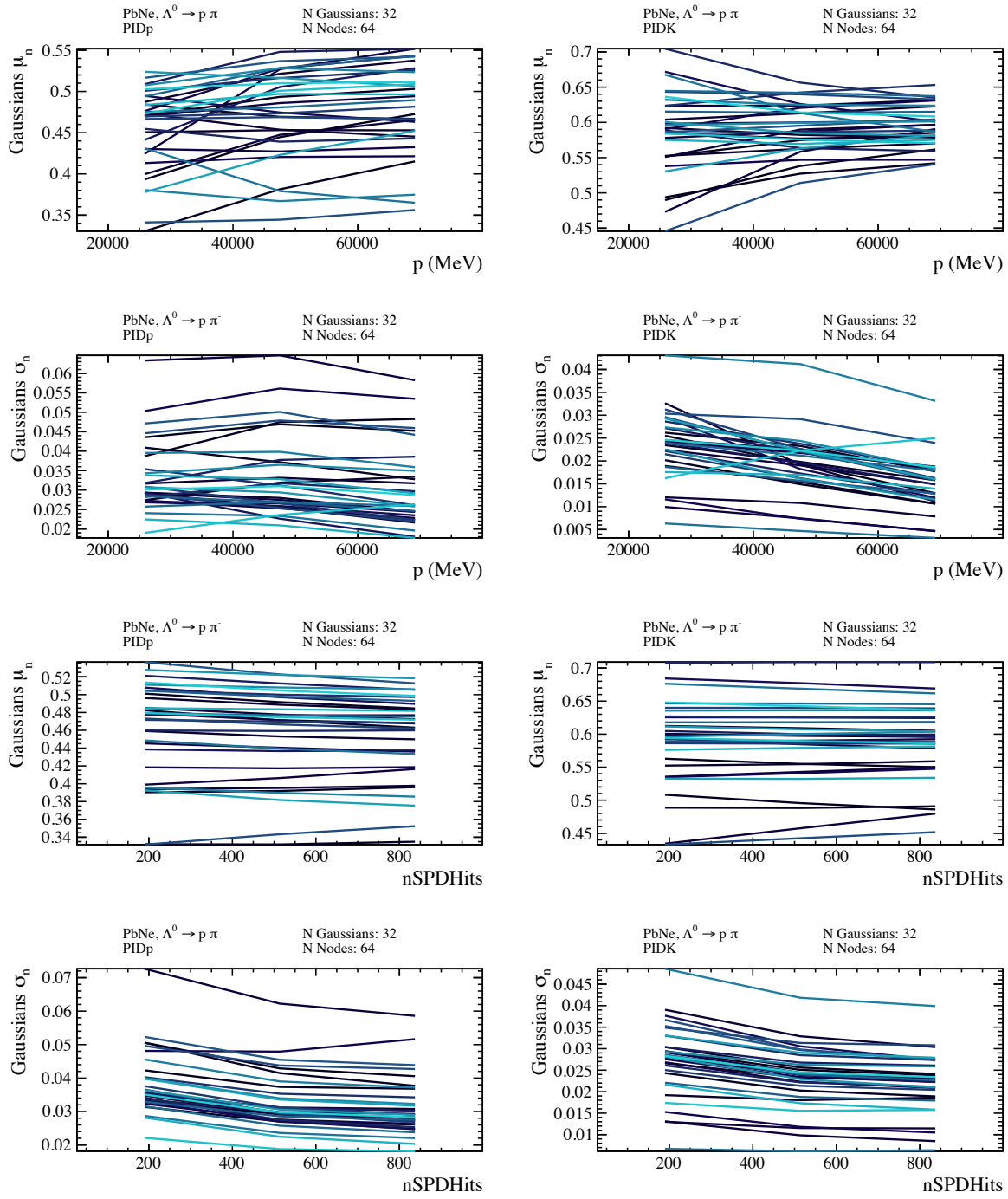


Figure 4.43: Check against a possible overtraining effect of the model trained on the $\Lambda \to p\pi^-$ calibration channel for PbNe. The evolution of the Gaussian mean values and widths for the PIDp (left) and the PIDK (right) variables are presented as functions of p (top) and nSPDHits (bottom). The different colors represent the Gaussian components included in the model.

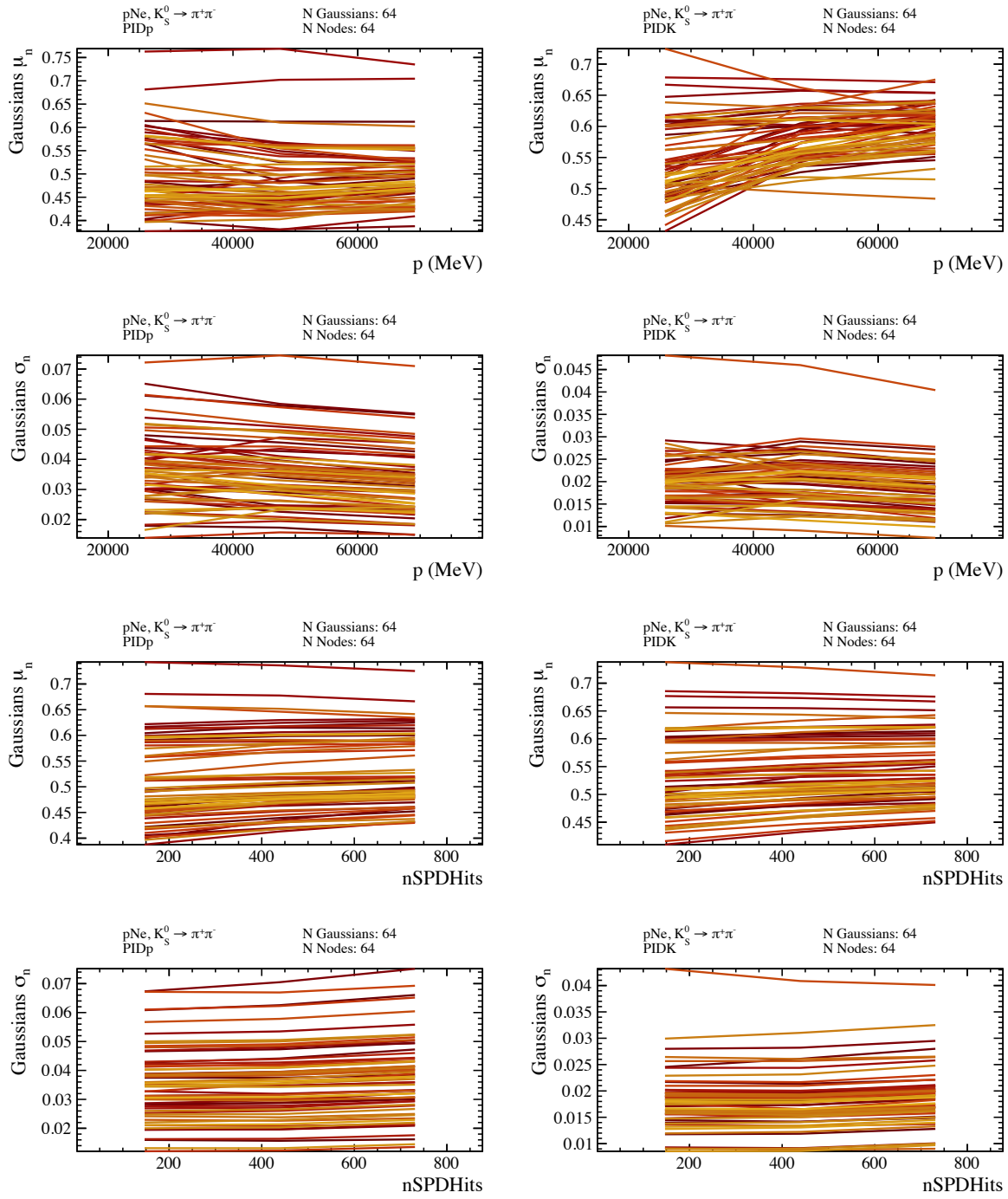


Figure 4.44: Check against a possible overtraining effect of the model trained on the $K^0_S \to \pi^+\pi^-$calibration channel for pNe. The evolution of the Gaussian mean values and widths for the PIDp (left) and the PIDK (right) variables are presented as functions of p (top) and nSPDHits (bottom). The different colors represent the Gaussian components included in the model.

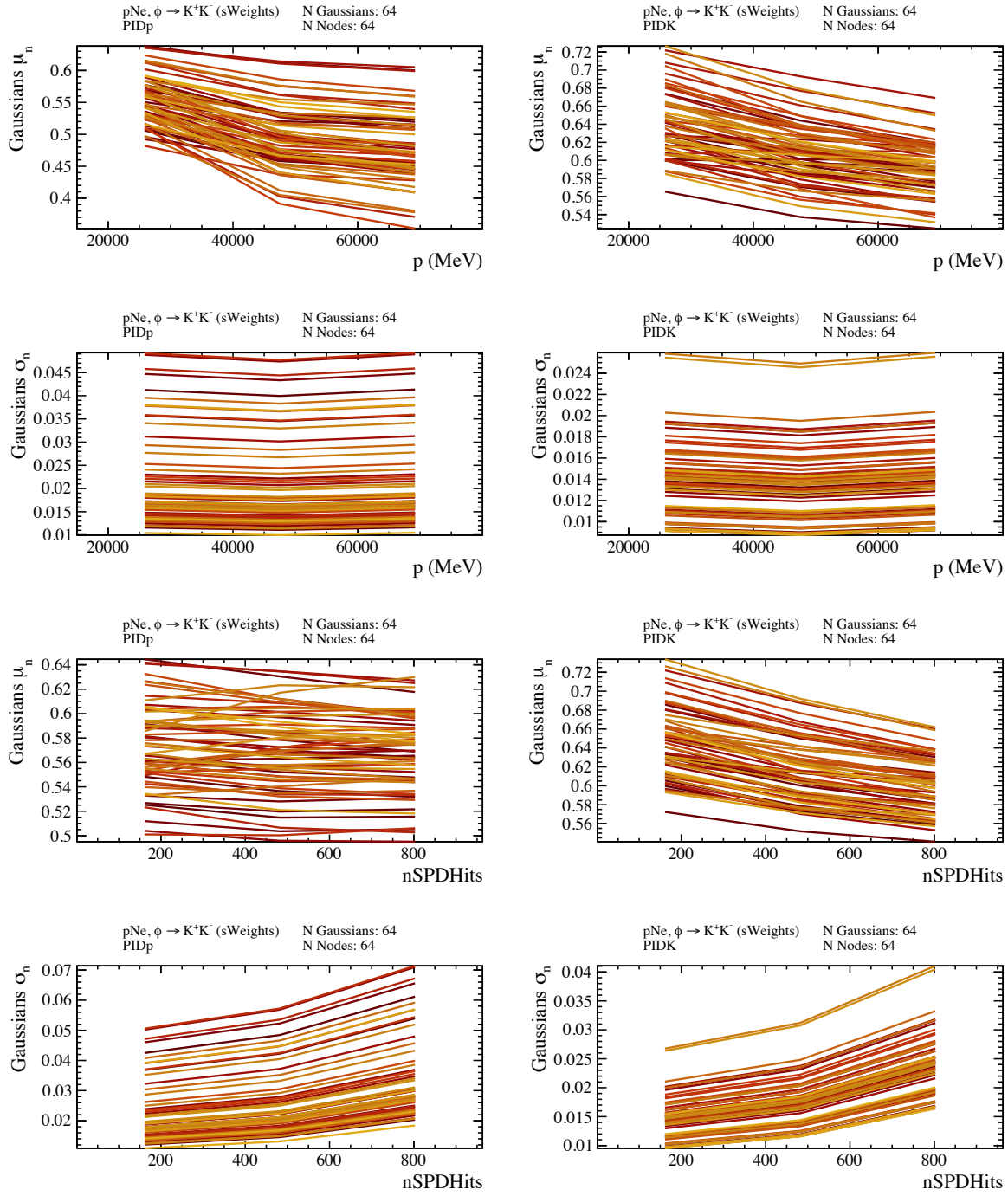


Figure 4.45: Check against a possible overtraining effect of the model trained on the $\phi \to K^+K^-$calibration channel for pNe. The evolution of the Gaussian mean values and widths for the PIDp (left) and the PIDK (right) variables are presented as functions of p (top) and nSPDHits (bottom). The different colors represent the Gaussian components included in the model.

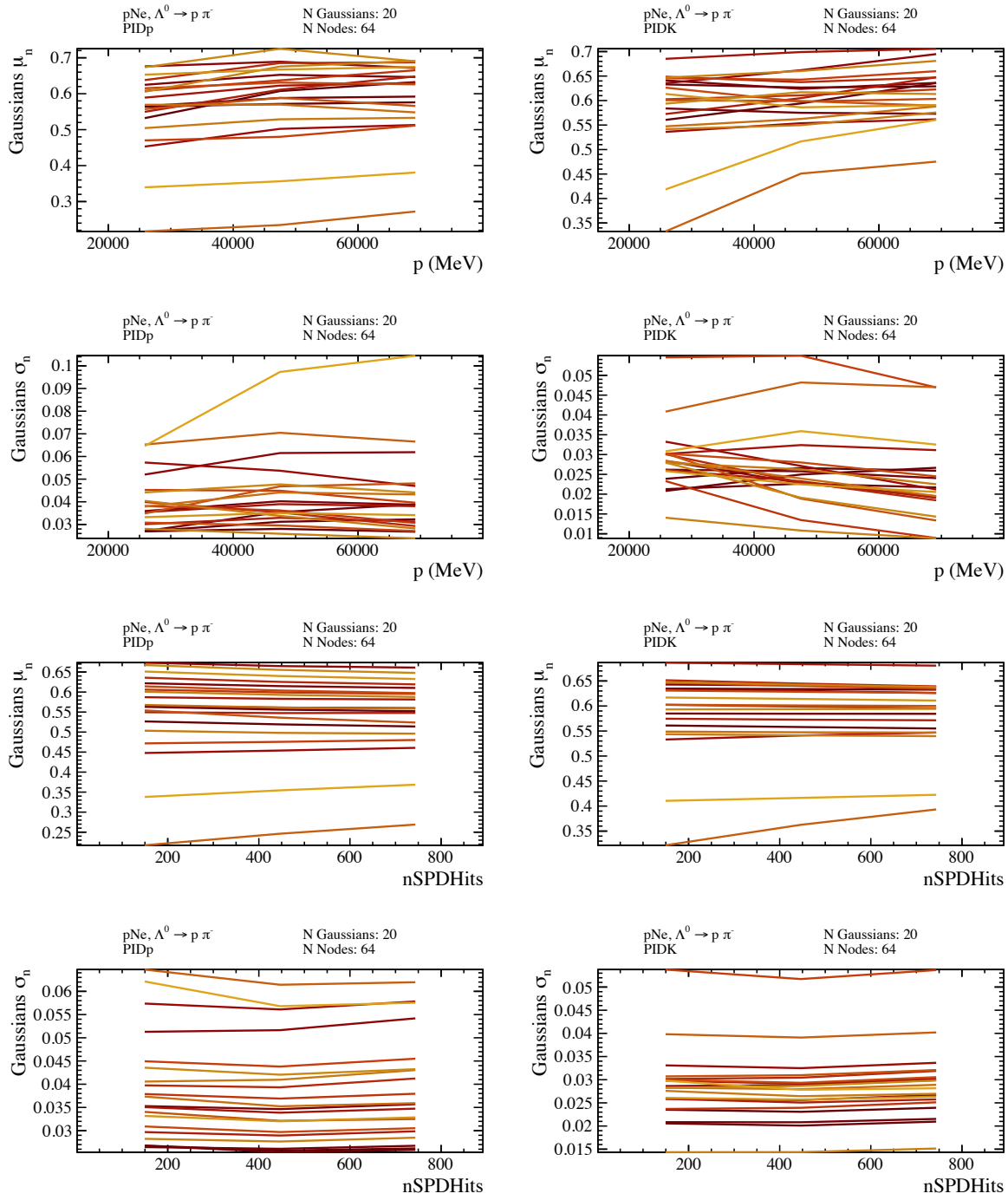


Figure 4.46: Check against a possible overtraining effect of the model trained on the $\Lambda \to p\pi^-$ calibration channel for pNe. The evolution of the Gaussian mean values and widths for the PIDp (left) and the PIDK (right) variables are presented as functions of p (top) and nSPDHits (bottom). The different colors represent the Gaussian components included in the model.

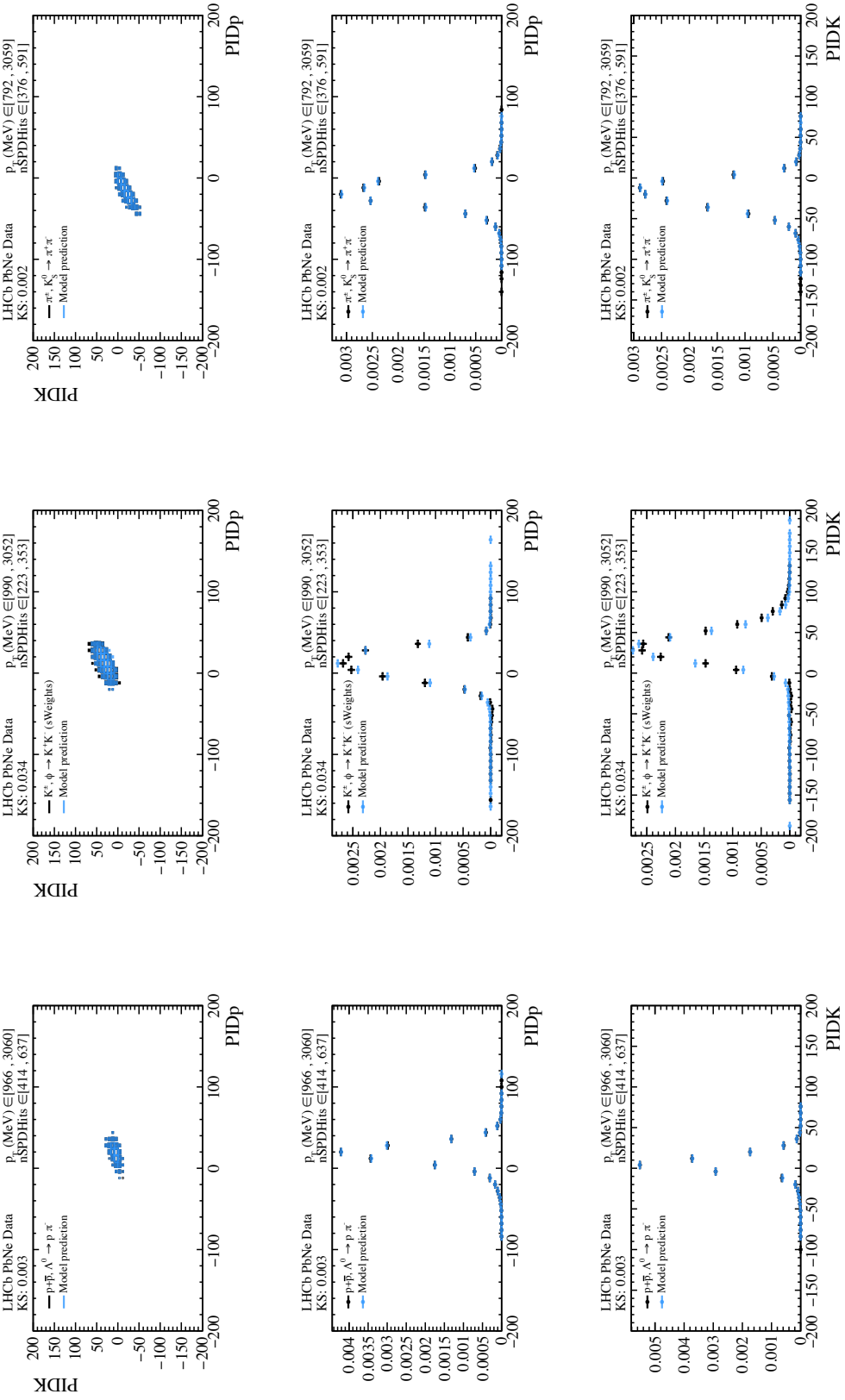


Figure 4.47: Comparison between the bi-dimensional PID response in PbNe data and the GMM model prediction in $K^0_S \to \pi^+\pi^-$(top), $\phi \to K^+K^-$(middle) and $\Lambda \to p\pi^-$(bottom) calibration samples. The projections along the two axis are also reported.

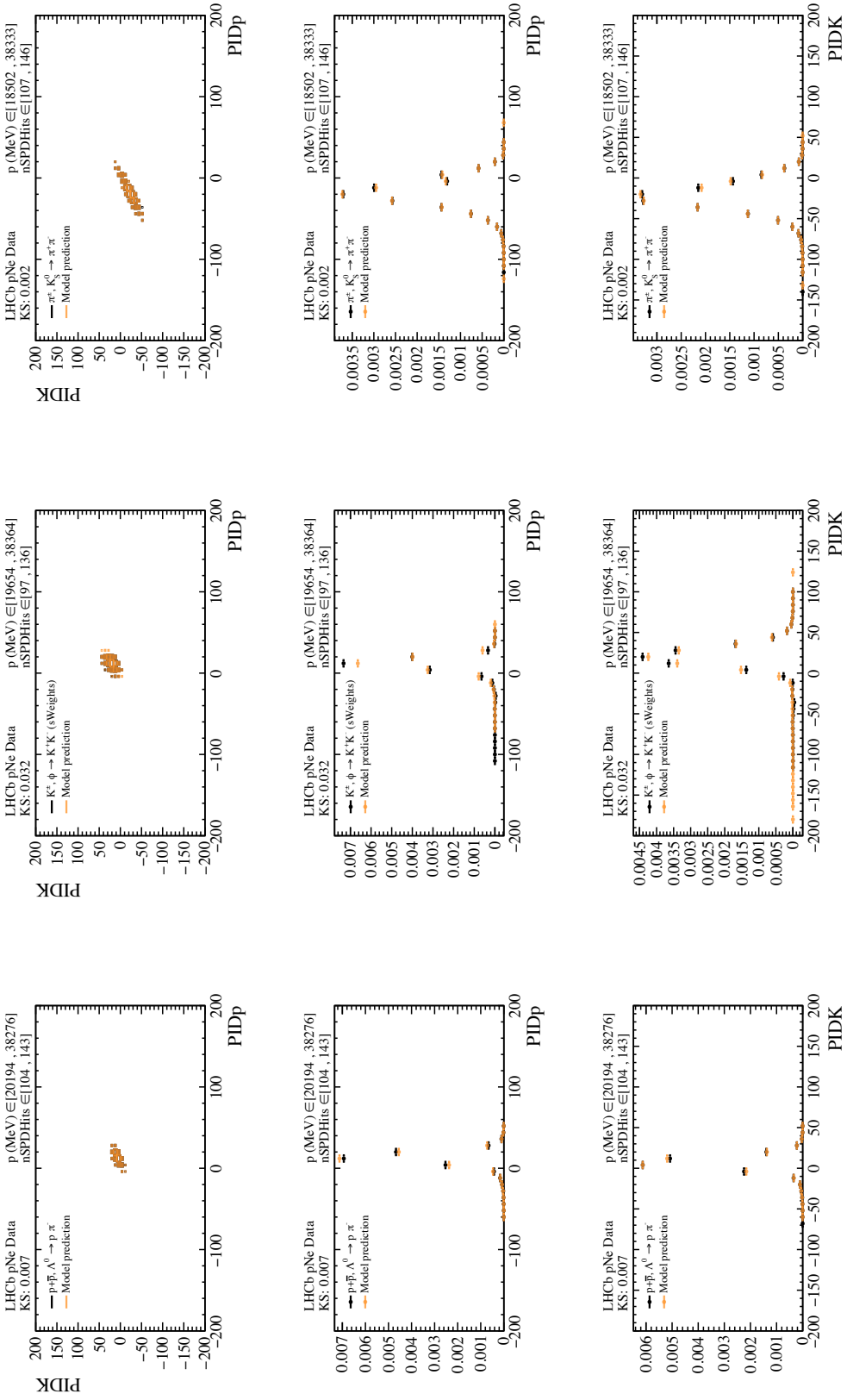


Figure 4.48: Comparison between the bi-dimensional PID response in pNe data and the GMM model prediction in $K^0_S \to \pi^+\pi^-$(top), $\phi \to K^+K^-$(middle) and $\Lambda \to p\pi^-$(bottom) calibration samples. The projections along the two axis are also reported.

4.7 Signal yields

The extraction of the signal yields for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ is obtained in different steps. The strategy is to perform a bi-dimensional fit over the distribution of prompt charged tracks in the PIDp-PIDK plane with a joint 2D pdf, described by the following expression

$$f(\text{PIDp, PIDK}) = \sum^{h} N_h f_h(\text{PIDp, PIDK}) \tag{4.25}$$

$$= N_\pi f_\pi + N_K f_K + N_p f_p + N_{gh} f_{gh} \ , \tag{4.26}$$

where N_π, N_K, N_p, N_{gh} are, respectively, the yields for pions, kaons, protons and ghost candidates, and f_π, f_K, f_p, f_{gh} are the PID templates for pions, kaons, protons and ghost candidates, respectively. The templates for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ are generated from the trained NN networks, described in Section 4.6, whereas the ghost template is derived from the reweighed simulation. Given the anticipated small contamination of ghost tracks (around 1-2%), studied in Section 4.3.3, no specific data-driven model is deemed necessary, with the reweighing considered sufficient to address any potential discrepancy with the data.

In addition, electrons and muons contributions are not considered in the modelling of the total pdf. As described in Section 4.3.5, the contamination of electrons is estimated from simulation to be less than 1%, whereas muons have even smaller contamination. Considering the similar PID response between electrons and pions, the residual contamination of electrons is applied as a correction factor for the experimental yield of pions.

4.7.1 Fit templates generation

After the training process outlined in Section 4.6.5, the different GMM models available for both PbNe and pNe datasets and for each hadron type, are utilised to predict the PID bi-dimensional distribution. Specifically, the NN network trained on the $K^0_S \to \pi^+\pi^-$, $\phi \to K^+K^-$and $\Lambda \to p\pi^-$decays predicts the PID response for pions, kaons and protons, respectively. This procedure is repeated independently on both PbNe and pNe datasets.

The template extraction involves applying the binning scheme, described in Section 4.1.1. For each bin, the following steps are repeated:

- select the data within the given bin interval;
- preprocess the candidates using the Quantile transformer for the features. The Scikit-learn transformation objects are initialised in the preprocessing step before the network training, defined in Section 4.6.5;
- feed the candidates' feature variables to the network. For each candidate, predict the probability density function in the bi-dimensional PID plane;
- obtain the PID template by sampling the pdf and applying the inverse MinMaxScaler transformation to the sampled points in the bi-dimensional PID plane. The number of samples is chosen to be high enough to sample also marginal regions of the pdf;

- normalise the resulting distribution;
- save the templates into ROOT files. This step facilitates the usage and the access of the templates for the fit.

The outcomes of this process for each hadron type for PbNe data is presented in Figure 4.49 for a particular bin interval. The binned bi-dimensional distributions obtained are smooth and represent the expected PID response in the 2D plane.

4.7.2 Fit in the PID plane

The templates are used to fit the PID distributions observed in binned data. The strategy for the fit is based on an extended likelihood method which allows Poisson fluctuations in the extraction of the normalisation constants [160]. The starting value of the yields, introduced in Equation 4.26, are initialised randomly to avoid potential biases in the convergence of the fit. The fit is performed for all possible bins for the different binning schemes of the analysis. The results of the 2D fits considering the binning over p_T are presented in Figures 4.51 and 4.52 for PbNe and pNe, respectively. The corresponding projection over the PID axis are presented as well for each bin.

To check the goodness-of-fit, the Kolmogorov-Smirnov (KS) [161] statistical test is employed. Formally the KS test is used to decide if a sample comes from a population with a specific distribution and the exact test only applies when comparing empirical data to continuous distributions. In case of 2D histograms the KS test uses the distance between the pseudo Cumulative Distribution Functions (CDFs) obtained from the histogram. Since in 2D the order for generating the pseudo-CDF is arbitrary, two pairs of pseudo-CDF are used, one starting from the x axis, the other from the y axis, and the maximum distance is the average of the two maximum distances obtained [162]. The sensitivity of the KS test is set to $\alpha = 0.05$. Bin intervals in which the of the KS test on the fit is higher than this value are rejected. A summary for the result of the KS test in different bins is reported in Figures 4.50. The majority of the bins pass the requirement, except few bins at the edge of the phase space. In addition, at least 100 candidates are required to be in each bin and the relative error on the efficiency has to be lower than 0.1.

The bin intervals that pass these criteria are accepted and the extracted yields for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ are corrected by the corresponding efficiencies. The value of the corrected yields with the corresponding efficiencies is presented in 2D in Figures 4.53 and 4.54 along with the corresponding projections presented in Figures 4.55 and 4.56 for PbNe and pNe, respectively. In both datasets, the corrected yields exhibit a mild dependence as a function of p_T for all three particle species. A dependence on η is also observed, which is attributed to the kinematic relation between η and p_T and the p region chosen for the analysis convoluted with the detector acceptance. It is important to note that it is not possible to compare directly the extracted yields for PbNe and pNe, since these results have not been normalised by the luminosity. Nevertheless, a smooth dependence is observed on both datasets which have been treated independently until this point.

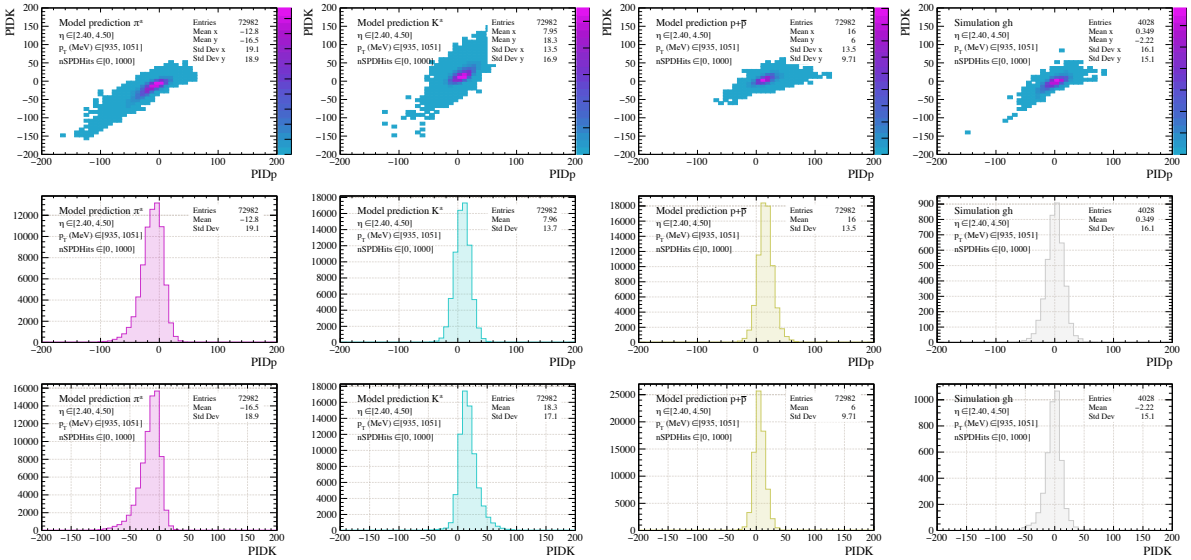
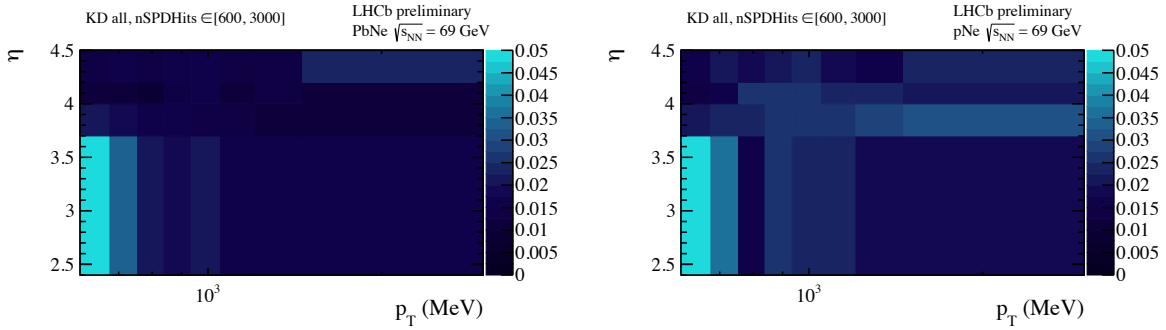


Figure 4.49: Illustration of the 2D PID templates for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$, and ghost in the PbNe dataset within the bin interval associated with $p_T \in [935, 1051]$ MeV/c. The predictions for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ are generated by their respective GMM models. Ghost templates are extracted from the PbNe simulation. Corresponding projections along the two axes are presented below each template.



(a) (b)

Figure 4.50: Distribution of the KS distance, used as an estimator for the goodness-of-fit, for PbNe (a) and pNe (b) data and the corresponding fit models in different bins of η and p_T. At the edges of the covered phase space, the fit quality is worse.

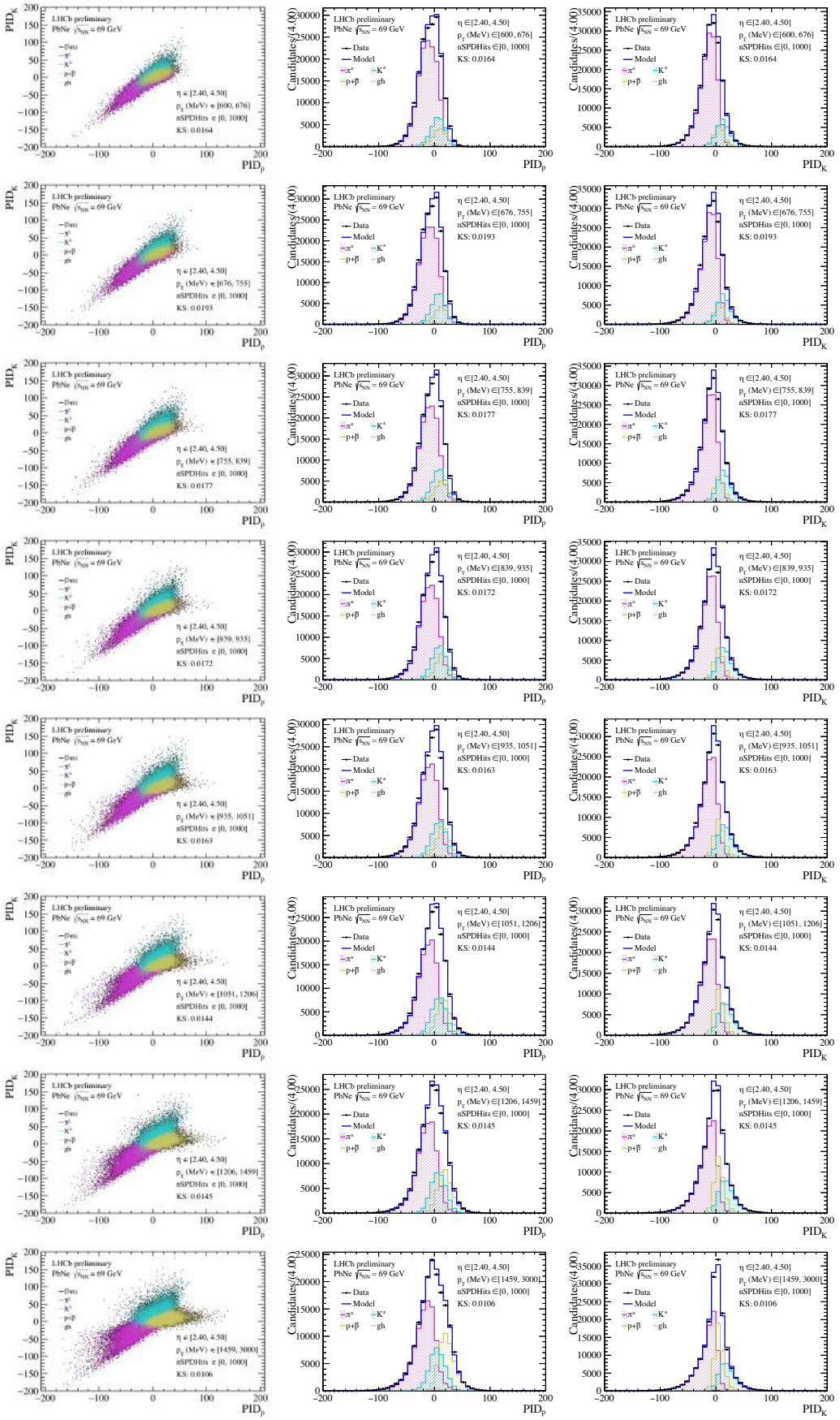


Figure 4.51: Results of the 2D fit in the PID plane and the corresponding projections onto the PID axes to PbNe data in p_T intervals.

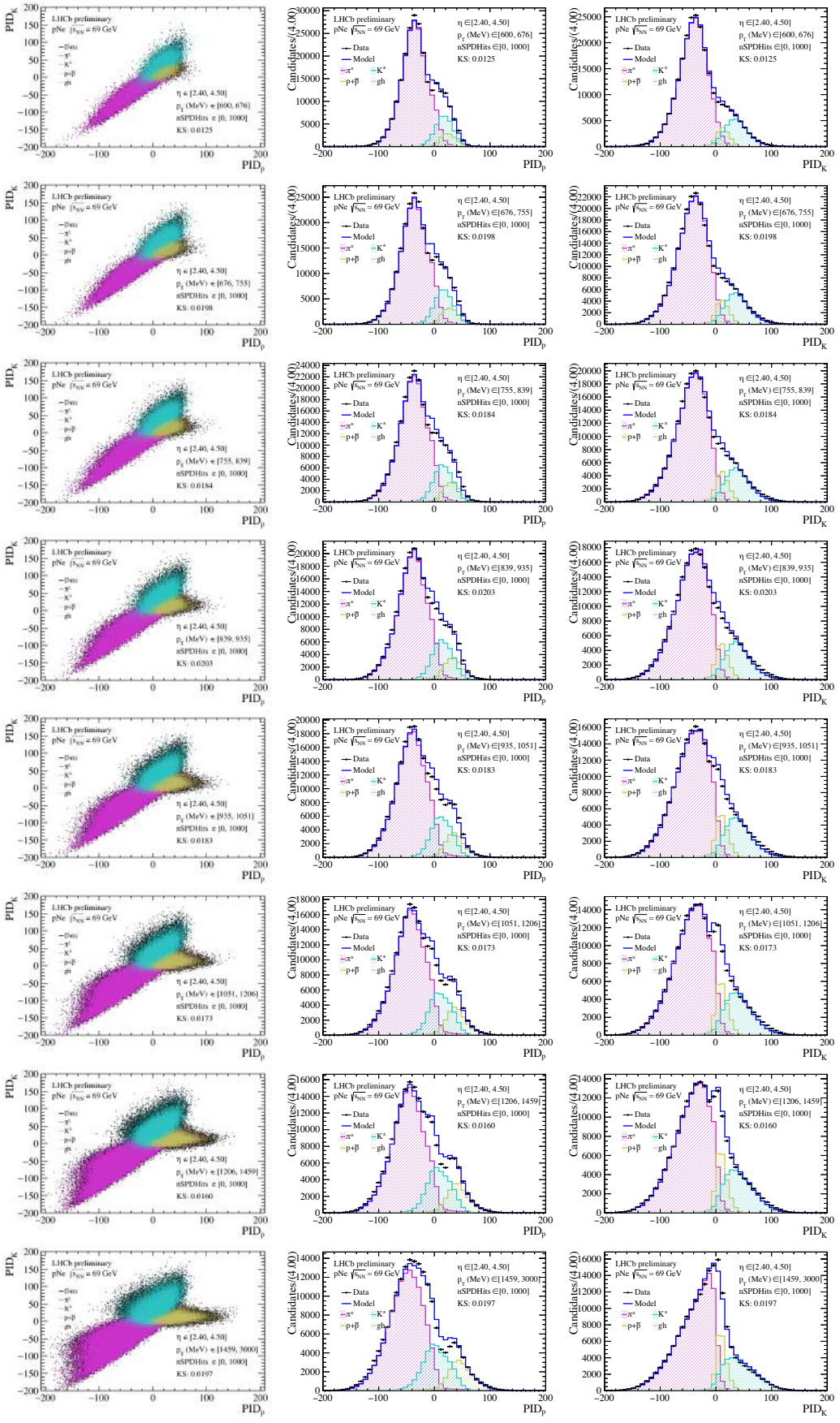


Figure 4.52: Results of the 2D fit in the PID plane and the corresponding projections onto the PID axes to pNe data in p_T intervals.

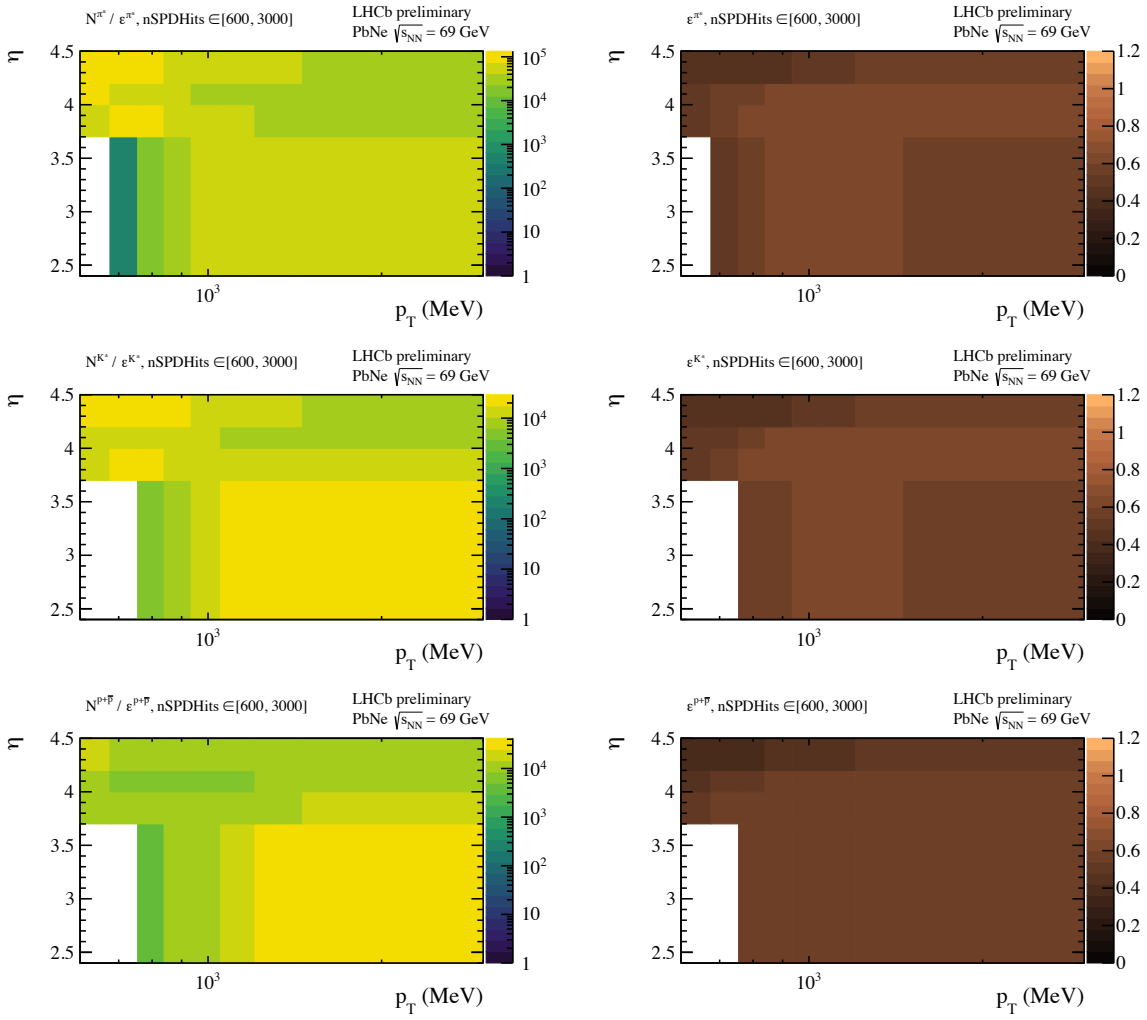


Figure 4.53: Distributions of the efficiency-corrected yields (left) and corresponding values of the total efficiencies (right) for pions (top), kaons (middle) and protons (bottom) for PbNe data. The results are displayed in different intervals in the $\eta - p_T$ plane.

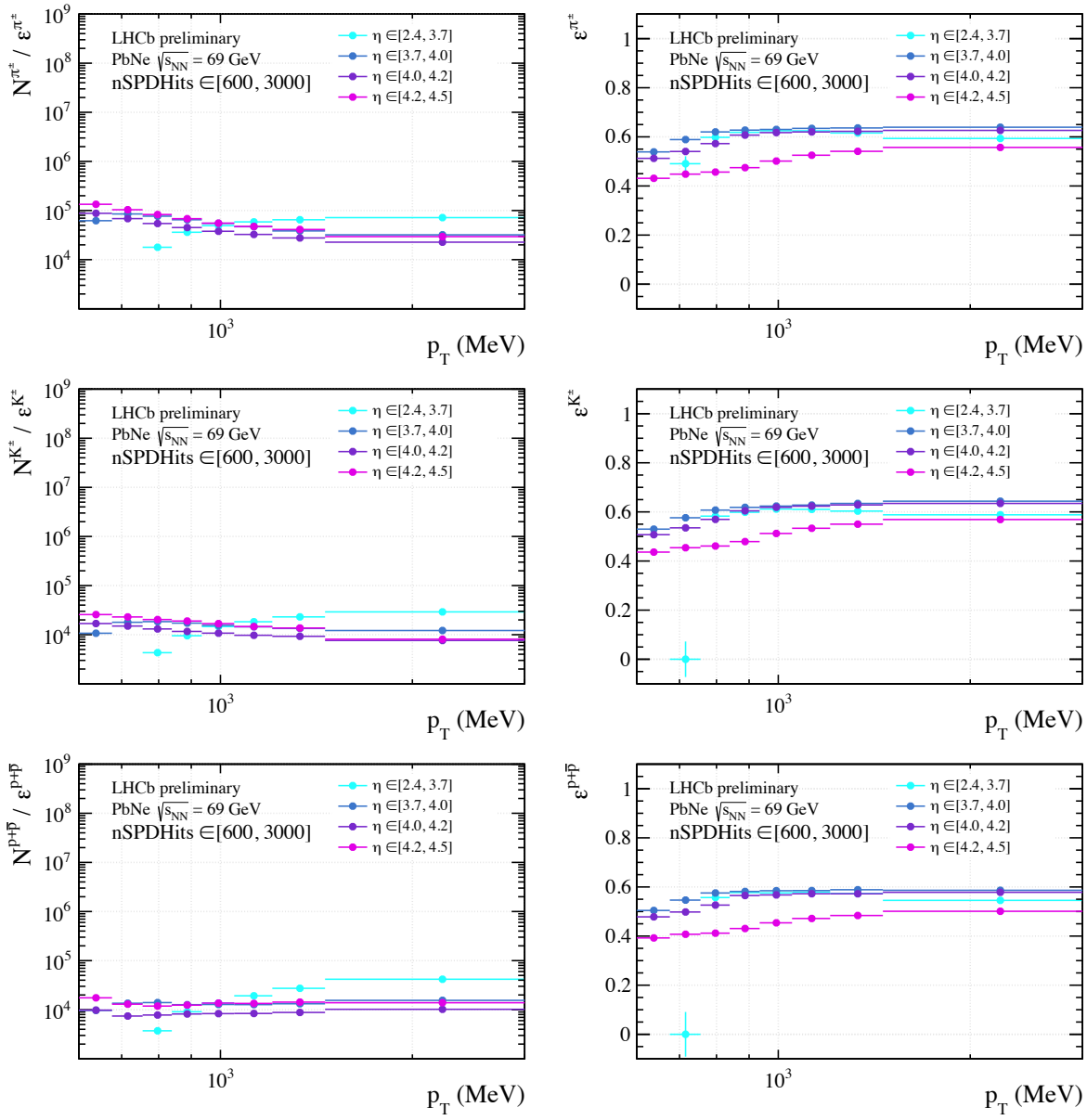


Figure 4.54: Distributions of the efficiency-corrected yields (left) and corresponding values of the efficiencies (right) for pions (top), kaons (middle) and protons (bottom) for PbNe data. The results are displayed as a function of p_T for different bins in η.

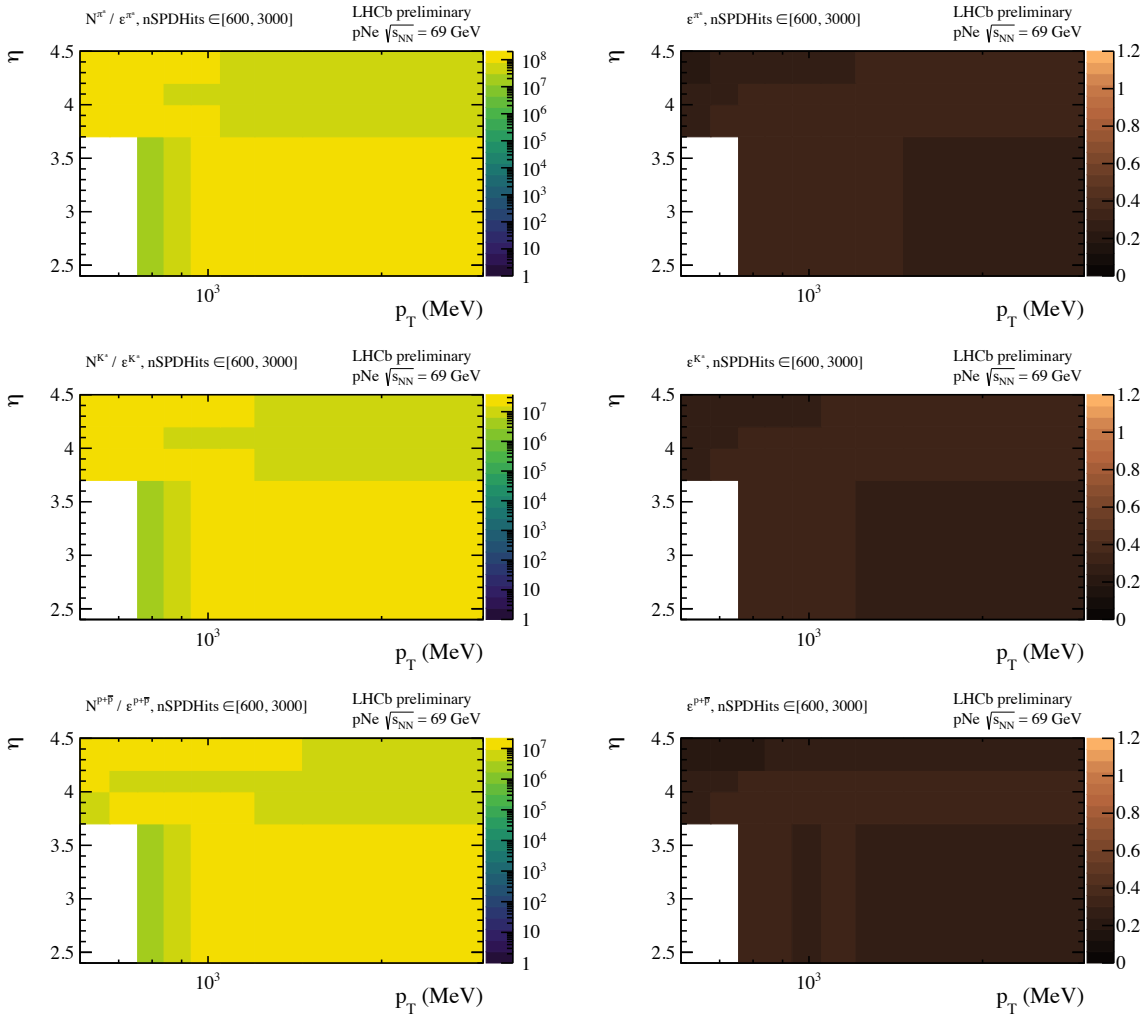


Figure 4.55: Distributions of the efficiency-corrected yields (left) and corresponding values of the total efficiencies (right) for pions (top), kaons (middle) and protons (bottom) for pNe data. The results are displayed in different intervals in the $\eta - p_T$ plane.

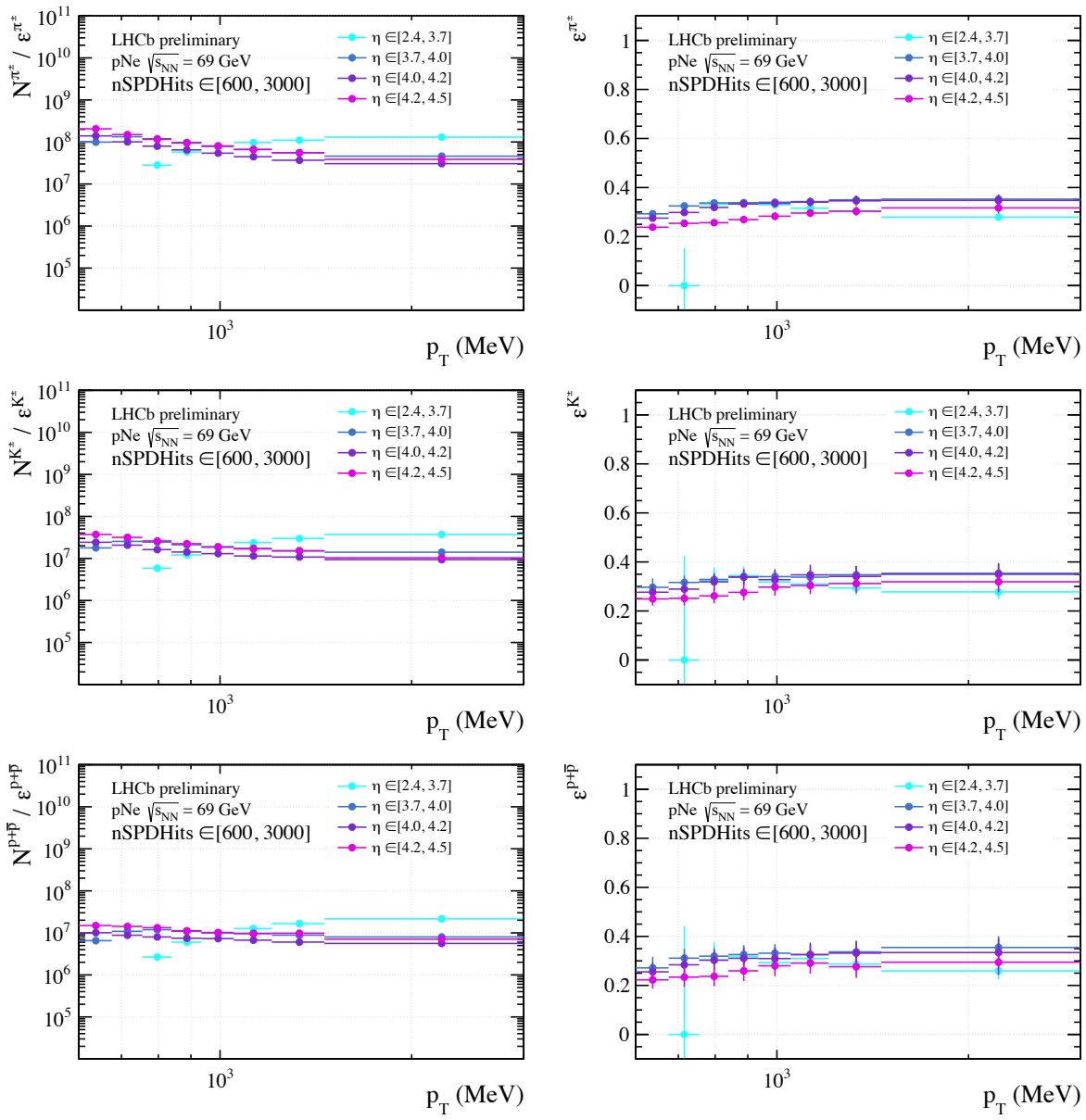


Figure 4.56: Distributions of the efficiency-corrected yields (left) and corresponding values of the total efficiencies (right) for pions (top), kaons (middle) and protons (bottom) for pNe data. The results are displayed as a function of p_T for different bins in η.

4.8 Systematic uncertainties

Systematic uncertainties affecting the ratios between two hadron types within the same dataset must be carefully estimated. In this section, the systematic uncertainties influencing efficiency calculations and signal extraction, are identified and calculated. The primary sources of uncertainties may arise from:

- **Ghost charge contamination:** As explained in Section 4.3, residual contamination from spurious PbPb and pp collisions occurs during fixed-target data taking due to Pb nuclei and protons migrating to a nominally empty bunch in beam 2. Notably, this uncertainty cancels out in the measurements of particle ratios for both datasets. Therefore, no specific systematic uncertainty is assigned for the ratio measurements.

- **Gas purity:** Outgassing from the beam-pipes or components of the SMOG device, primarily composed of hydrogen, has the potential to dilute the Ne gas, affecting its purity and, consequently, the data sample's purity. Similar to the previous point, this contribution cancels out in particle ratios; thus, no systematic uncertainty is assigned.

- **NN template estimation:** It is crucial to assess the reliability and stability of the neural network response. The shape of the generated templates influences the fit in the bi-dimensional PID plane and the yields obtained in each bin. Additionally, the yield variation related to the NN template estimate is correlated between hadron types. For instance, a modification in the p template not only affects the relative yield of p but also that of K and π.

- **Simulation statistics:** This systematic uncertainty is accounted by propagating the statistical uncertainty on the evaluation of the efficiencies.

- **Simulation reweighing:** Efficiencies change for each hadron type, and this variation needs to be controlled in the ratio. Default efficiencies are computed after reweighing the simulation, as outlined in Section 4.4. Hence a systematic uncertainty related to the MC reweighing procedure must be considered.

The most relevant systematic uncertainties which do not cancel out in single and double ratios are those related to the NN template estimation and the simulation reweighing procedure and will be further addressed in the following sections.

4.8.1 NN template estimation

The extraction of the relative yields from the template fit relies almost entirely on the parametrisation of the pdf in the 2D PID plane of the NN trained on both PbNe and pNe calibration lines. To evaluate this uncertainty, different NNs have been trained after varying the network parameters with respect to a reference configuration defined by the settings reported in Table 4.18 for all the 3 networks trained to predict $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ templates.

Label	Feature changes	Targets	nNodes changes	nGaussians changes
nn1	/	PIDp, PIDK	+20	/
nn2	/	PIDp, PIDK	-20	/
nn3	/	PIDp, PIDK	/	+10
nn4	/	PIDp, PIDK	/	-10
nn5	+ PT	PIDp, PIDK	/	/

Table 4.19: Various GMM models configurations were employed to assess the fluctuations in the yield estimation following the fitting process. The $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ template prediction networks were all trained with identical parameters for both PbNe and pNe datasets.

The GMM models were altered from the aforementioned configuration as outlined in Table 4.19. These variations in NNs were selected to assess performance across different network complexities by adjusting the number of nodes in hidden layers and the number of Gaussians. Consequently, the total number of NN parameters optimised during the loss function minimisation in the training process was affected. Additionally, variations in input features were introduced, incorporating p_T as additional variable.

The experimental yields for the Gaussian Mixture Model (GMM) models listed in Table 4.19 are depicted in Figures 4.57 and 4.58 for PbNe and pNe, respectively. The variation in yield within a specific bin is influenced by both the NN structure and the parameter space covered by that bin. No discernible preference for a particular NN structure was observed, and thus all structures were retained for evaluating the systematic error associated with PID. The associated systematic PID error is represented by the standard deviation of the distribution of these yields. The distributions of systematic uncertainties associated with the choice of the GMM model as a function of p_T are presented in Figures 4.59 and 4.60 for PbNe and pNe, respectively.

Future improvements of the systematic uncertainties related to template generation can be estimated by varying target particle identification (PID) variables, encompassing various combinations of log-likelihood hypotheses.

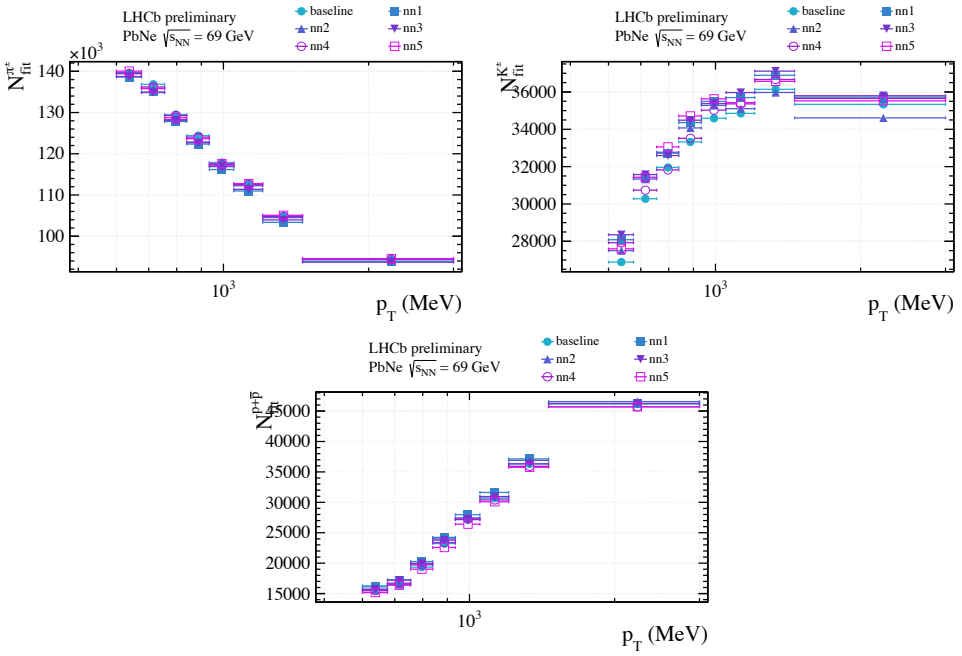


Figure 4.57: Experimental $\pi^\pm, K^\pm, \overset{(-)}{p}$ yields extracted in PbNe data for different GMM models as a function of p_T, integrated in η and event multiplicity. The discrepancies between the different models is accounted as a systematic uncertainty.

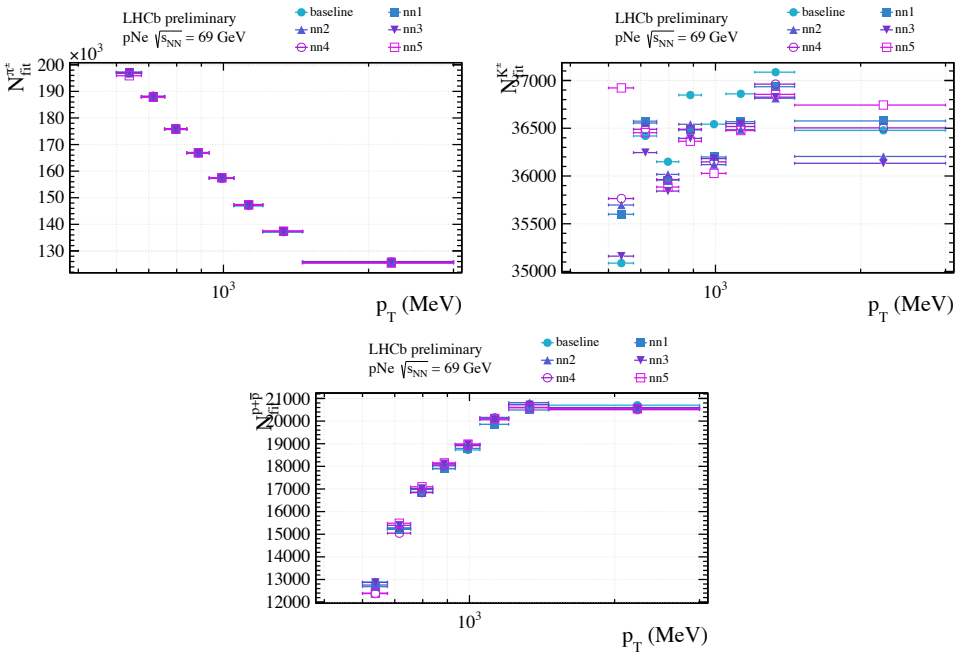


Figure 4.58: Experimental yields of in pNe data for different GMM models as a function of p_T, integrated in η and event multiplicity. The discrepancies between the different models is accounted as a systematic uncertainty.

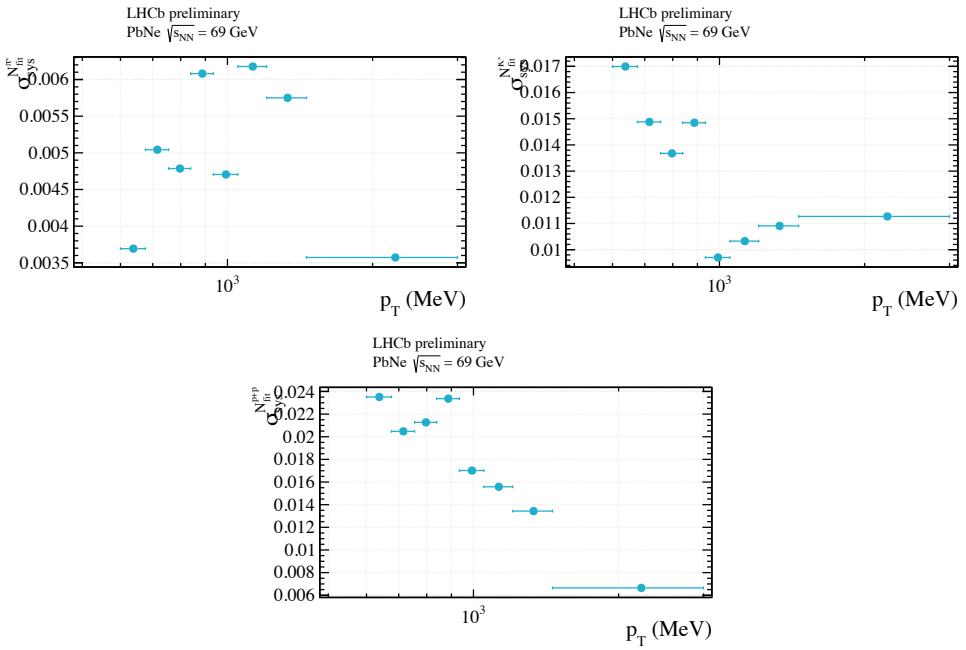


Figure 4.59: Systematic uncertainty for the experimental $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ yields for PbNe data associated to the choice of the GMM model, as a function of p_T, integrated over η and event multiplicity.

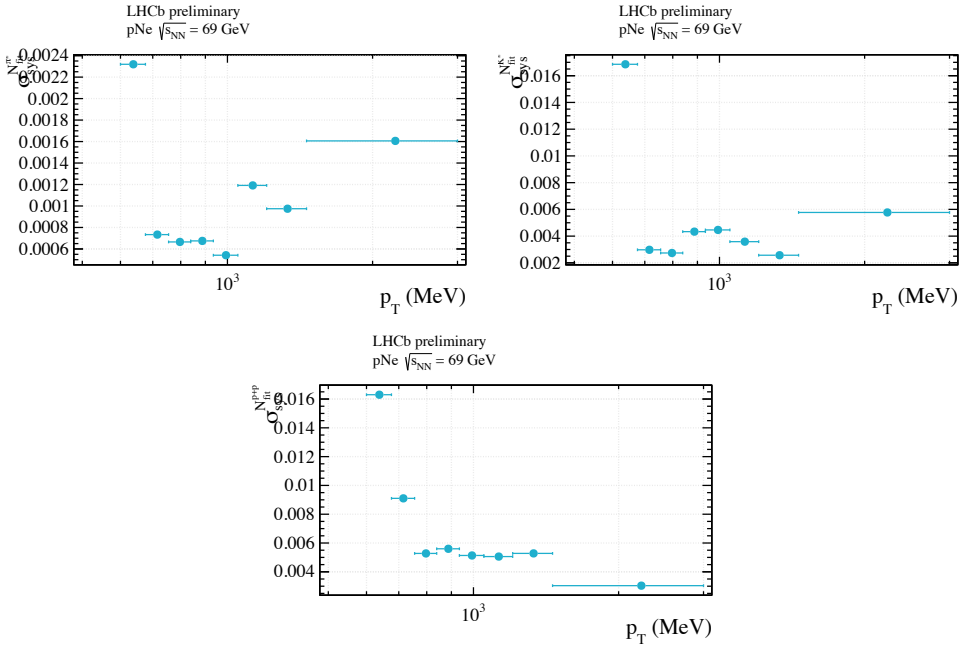


Figure 4.60: Systematic uncertainty for the experimental $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ yields for pNe data associated to the choice of the GMM model, as a function of p_T, integrated over η and event multiplicity.

4.8.2 Simulation reweighing

The choice on the model for the reweighing of the MC sample can affect the final value of the efficiencies. To take into account this systematic uncertainty, alternative reweighing models have been trained and utilised to quantify the impact of the GBR input variables. The nominal reweighing exploits η, p_T and nSPDHits. Four alternative reweighing methods were tested, summarised in Table 4.20. Notably, the nSPDHits has to be always reweighed considering the significant differences with respect to data.

The distributions of the efficiencies evaluated for $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$ for different simulation reweighings are presented as a function of p_T in Figures 4.61 and 4.62 for PbNe and pNe, respectively. No particular preferences are observed for a specific model for reweighing, and thus values evaluated for all the presented models are retained for evaluating the systematic error. The associated systematic uncertainty is represented by the standard deviation of the distribution of these yields.

The distributions of systematic uncertainties associated with the simulation reweighing are presented as a function of p_T in Figures 4.63 and 4.64 for PbNe and pNe, respectively. The systematic uncertainty associated to the simulation reweighing reaches values up to approximately 0.3% for PbNe and 1.0% for pNe. This can be due to the reduced ability of the GBR reweighing model at higher values of p_T considering the lower statistics in the simulation in this region. In any case, the effect is expected to be small which shows that the choice of the reweighing model does not significantly affect the final results.

Label	Model input variables
baseline	η, p_T, nSPDHits
rw1	p_T, nSPDHits
rw2	η, nSPDHits
rw3	nSPDHits
rw4	p, nSPDHits

Table 4.20: Input variables used to train different GBR model used to for reweighing both PbNe and pNe simulations.

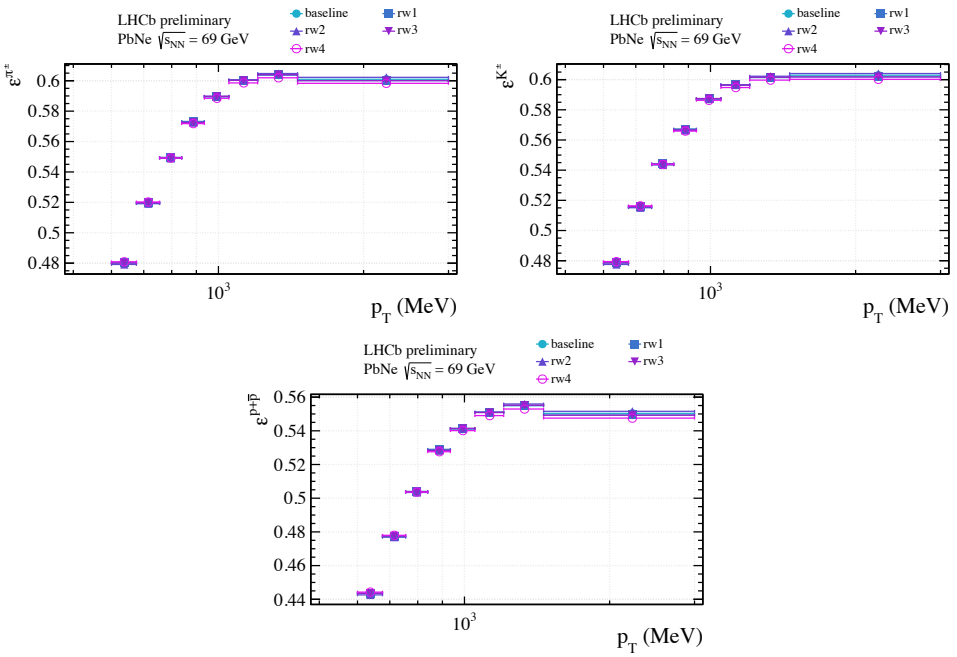


Figure 4.61: Distributions of the efficiencies of $\pi^\pm, K^\pm, \overset{(-)}{p}$ evaluated in PbNe data for different simulation reweighing models as a function of p_T, integrated in η and event multiplicity. The discrepancies between the different models is accounted as a systematic uncertainty.

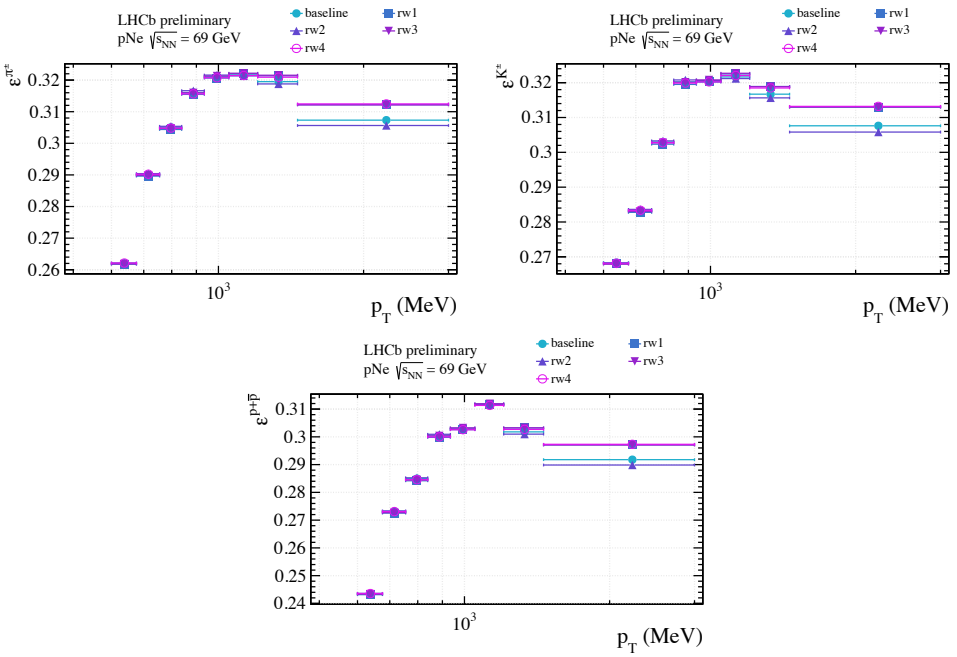


Figure 4.62: Distributions of the efficiencies of $\pi^\pm, K^\pm, \overset{(-)}{p}$ evaluated in pNe data for different simulation reweighing models as a function of p_T, integrated in η and event multiplicity. The discrepancies between the different models is accounted as a systematic uncertainty.

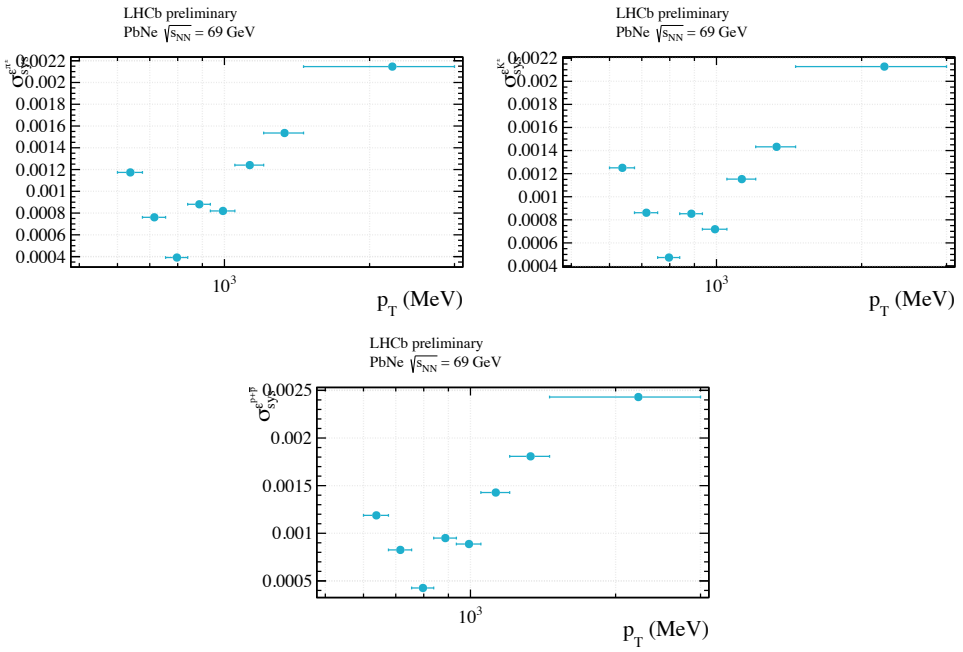


Figure 4.63: Systematic uncertainty for the efficiencies of $\pi^\pm, K^\pm, \overset{(-)}{p}$ associated to the simulation reweighing models for PbNe data, as a function of p_T, integrated over η and event multiplicity.

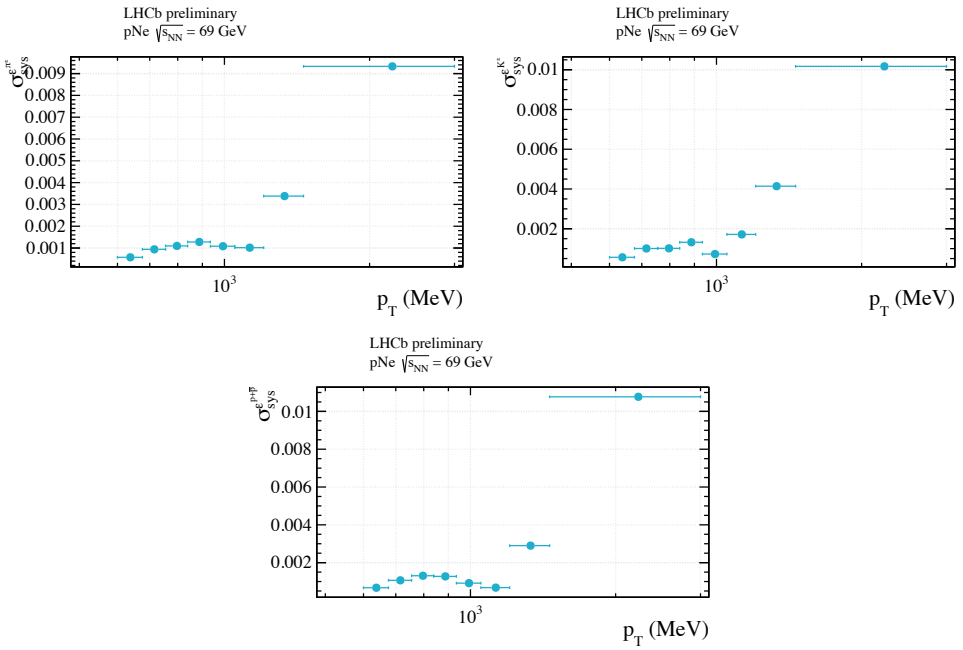


Figure 4.64: Systematic uncertainty for the efficiencies of $\pi^\pm, K^\pm, \overset{(-)}{p}$ associated to the simulation reweighing models for pNe data, as a function of p_T, integrated over η and event multiplicity.

4.8.3 Total systematic uncertainties

The total systematic uncertainty comprises both the uncertainty associated to the yield extraction procedure and the one related to the calculation of the efficiencies, since both affect the value of the corrected signal yields utilised to calculate the single and double ratios.

The distributions of the corrected yields evaluated for $\pi^{\pm}, K^{\pm}, {}^{(-)}_{p}$ for different GMM models and for different simulation reweighing models are presented as a function of p_T in Figures 4.65 and 4.66 for PbNe and pNe, respectively. The distributions of total systematic uncertainties are presented as a function of p_T in Figures 4.67 and 4.68 for PbNe and pNe, respectively. For PbNe the systematic uncertainty is dominated by the choice on the NN model to produce the templates for all the p_T bins. For pNe the systematic uncertainty at high p_T is dominated by the simulation reweighing.

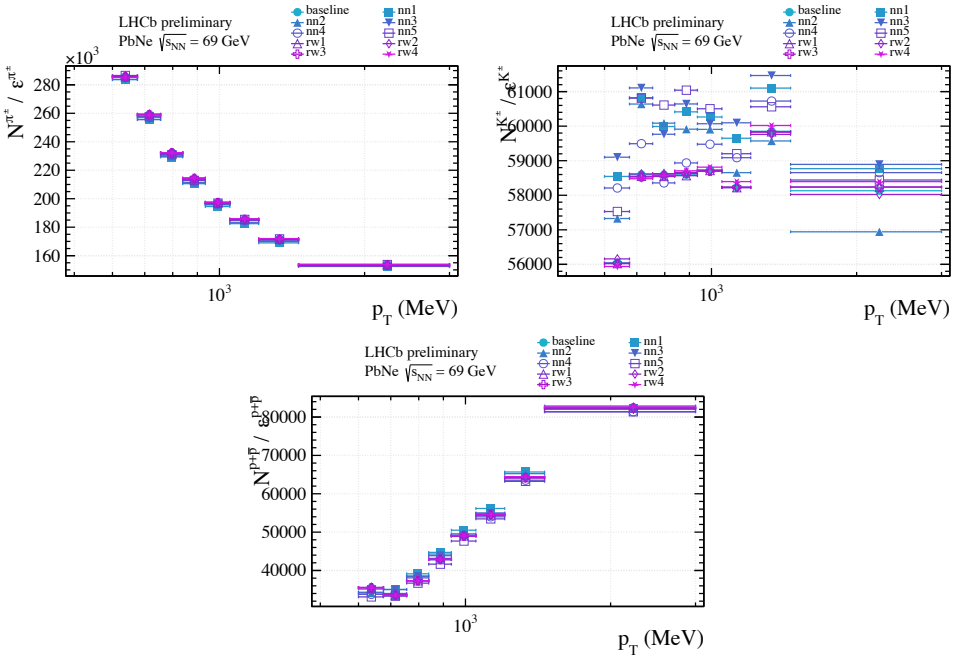


Figure 4.65: Distributions of the corrected yields of $\pi^\pm, K^\pm, \overset{(-)}{p}$ in PbNe data for different analysis methods as a function of p_T, integrated in η and event multiplicity. The discrepancies between the different models is accounted as a systematic uncertainty.

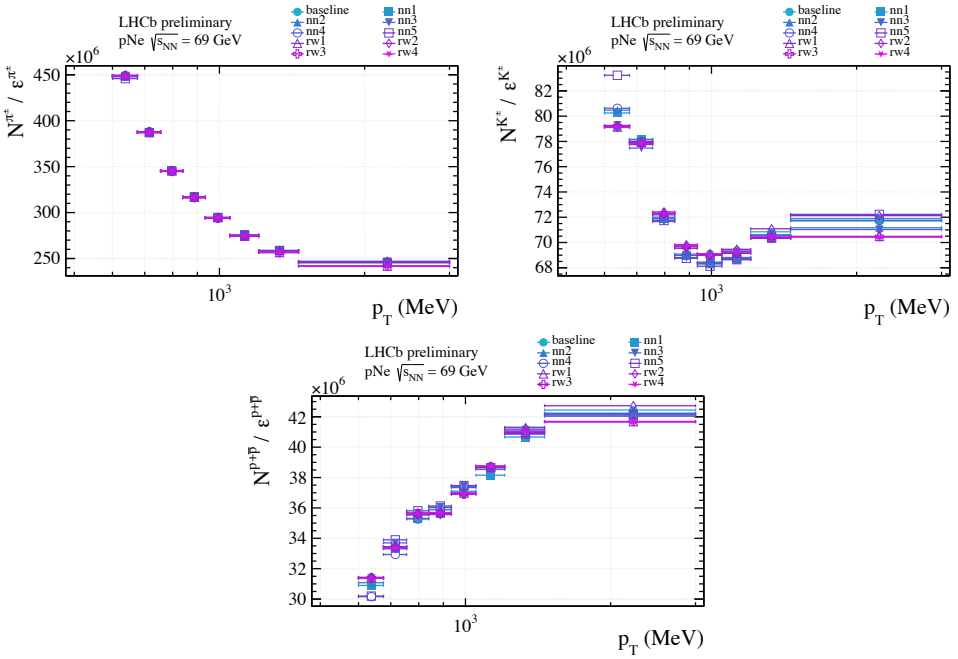


Figure 4.66: Distributions of the corrected yields of $\pi^\pm, K^\pm, \overset{(-)}{p}$ in pNe data for different analysis methods as a function of p_T, integrated in η and event multiplicity. The discrepancies between the different models is accounted as a systematic uncertainty.

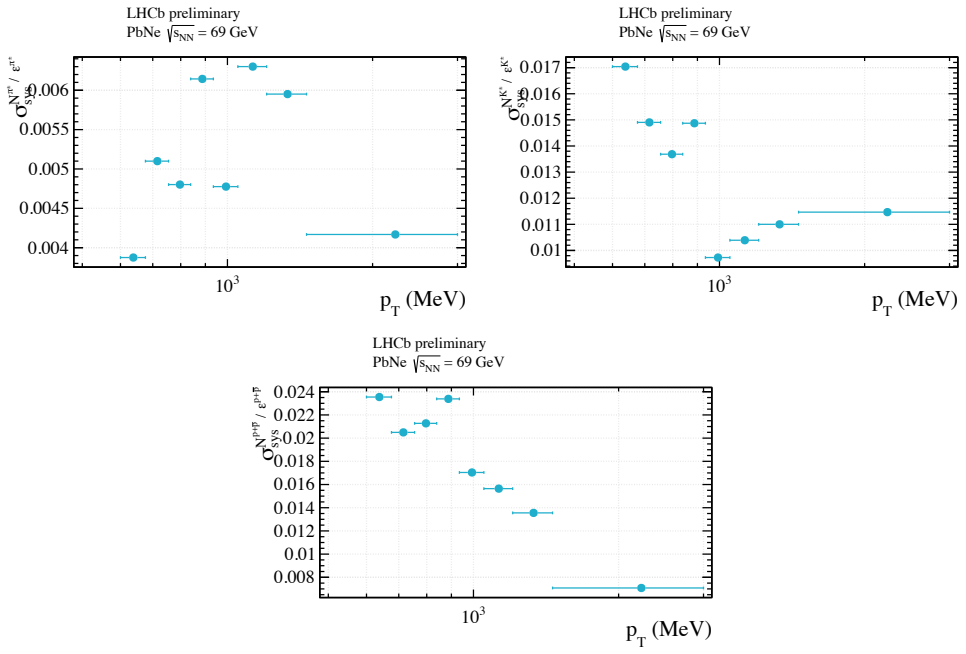


Figure 4.67: Total systematic uncertainty for the corrected yields of $\pi^\pm, K^\pm, \overset{(-)}{p}$ for PbNe data, as a function of p_T, integrated over η and event multiplicity.

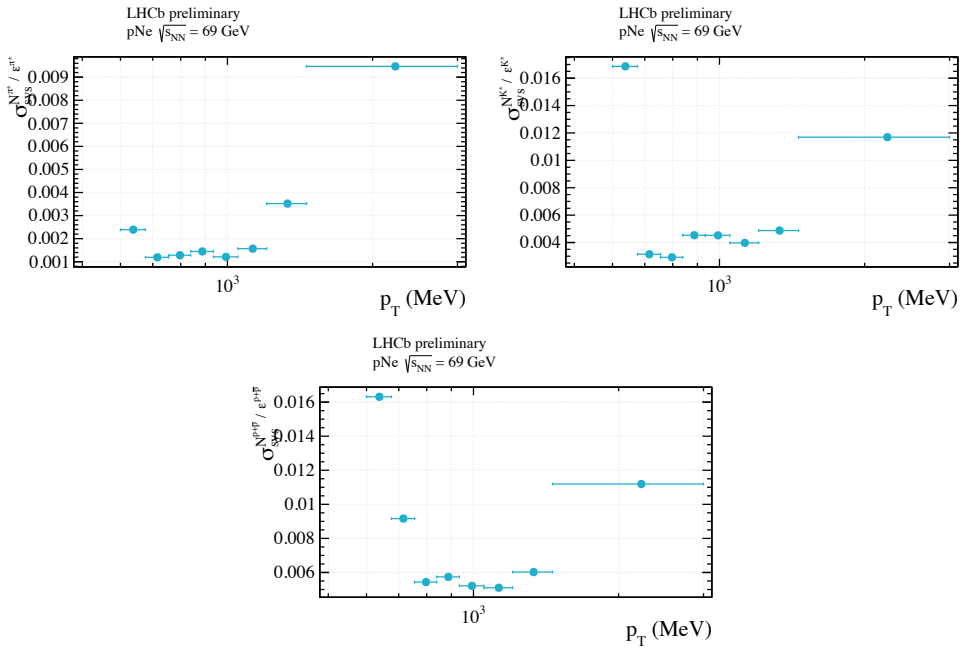


Figure 4.68: Total systematic uncertainty for the corrected yields of $\pi^\pm, K^\pm, \overset{(-)}{p}$ for pNe data, as a function of p_T, integrated over η and event multiplicity.

4.9 Results

The outcomes of the analysis on light hadron production are organised according to various binning schemes. Yields and ratios are computed for each particle type, considering all tracks, as well as separately for positive and negative tracks. The focus primarily lies on presenting results against transverse momentum (p_T), given its high sensitivity to CNM effects. It is also interesting to look at the ratios as a function of the multiplicity. One-dimensional results are compared for different hadron species and collision types. The key yield ratios of interest include:

- **Single ratios** for a specific collision system: p/π, K/π, with the heavier hadron placed in the numerator are presented in Section 4.9.1;
- **Double ratios** between PbNe and pNe collisions to facilitate a direct comparison between the two collision systems are presented in Section 4.9.2;
- **Validation of the results**: a set of cross-check measurement is conducted to validate the results and the analysis strategy the results presented in Section 4.9.3;
- **Comparison with the simulation**: the results are compared with simulation in Section 4.9.4. Any discrepancy can lead to improvement of the simulation.

All the results provided in the following sections comprehends both systematic and statistical uncertainties.

4.9.1 Single ratios

The single K/π , p/π , and p/K ratios, as defined in Equation 4.4, offer valuable insights into particle production for each collision system. The 1D comparisons of the $\pi^{\pm}, K^{\pm}, {}^{(-)}p$ single ratio in PbNe and pNe are presented as functions of η, p_T, and nSPDHits. Figure 4.69 displays the distributions as functions of η, integrated over p_T and collision centrality. All three hadron ratios exhibit enhancements in PbNe data compared to pNe data. Particularly, a stronger dependence on η is observed in PbNe data, with a larger enhancement of the p/π and p/K ratios for $\eta < 3.7$.

A more intriguing feature emerges when examining the same ratios as functions of p_T, presented in Figure 4.70, integrated over η and collision centrality. Notably, when focusing on baryon-meson ratios (p/π and p/K), the contrast between the two collision systems becomes more apparent in the high-p_T region. Specifically, for $p_T > 1$ GeV/c, the single ratio in PbNe data doubles that in pNe data. This observation suggests a more significant modification due to, e.g., the Cronin effect in the $p_T > 1$ GeV/c region. The less pronounced deviation in the K/π ratio, compared to the deviations in the p/π and p/K ratios, may be associated with the baryon-meson anomaly discussed in Chapter 2. In the p/K ratio for PbNe data, a minimum is observed at $p_T \sim 700$ MeV/c. Further investigation in a 2D binning over pseudorapidity can provide insights into this effect.

The distribution as a function of nSPDHits is presented in Figure 4.71. As anticipated in Section 4.3, the pNe dataset covers only a limited range in multiplicity, while the PbNe dataset

extends to higher values. This result indicates that the modification of the single ratios is observed even in the low multiplicity region (nSPDHits < 300), corresponding to centrality values up to approximately 70%. A comparison of these results with smaller collision systems (pHe, pAr) collected by LHCb in Run 2 can verify if these effects are observed at even smaller event multiplicities, providing a more comprehensive understanding of these effects.

A more detailed examination of these effects can be conducted by analysing the single ratios in simultaneous bins of η and p_T. Figures 4.72 and 4.73 illustrate the single K/π , p/π , and p/K ratios as function of p_T for different bins of η, integrated over collision centrality. In both datasets, the distribution of the single ratio consistently increases as a function of p_T in all η bins. Particularly for PbNe data, a clear simultaneous dependence in η is observed. In the $p_T < 1$ GeV/c region, an increase in the p/π and p/K ratios is observed for lower values of η. This result is consistent with the 1D distribution as a function of η, which shows a strong modification for $\eta < 3.7$. Additionally, this provides more information on the minimum observed in the 1D distribution as a function of p_T. This effect is not observed in pNe data, where the trend is consistent for all η bins. These observations could suggest the presence of additional effects in larger collision systems that modify baryon-meson ratios in the backward region of the centre-of-mass frame.

In this context, it is interesting to study the p_T dependence as a function of collision centrality. Since pNe data covers only one centrality bin, this investigation is conducted only for the PbNe dataset. Figures 4.74 present K/π , p/π , and p/K as functions of p_T in different centrality bins. A noticeable trend, especially in K/π and p/K ratios, is observed as a function of collision centrality. This effect is more pronounced in the $p_T < 1$ GeV/c region, aligning with previous findings from 2D binning in η and p_T.

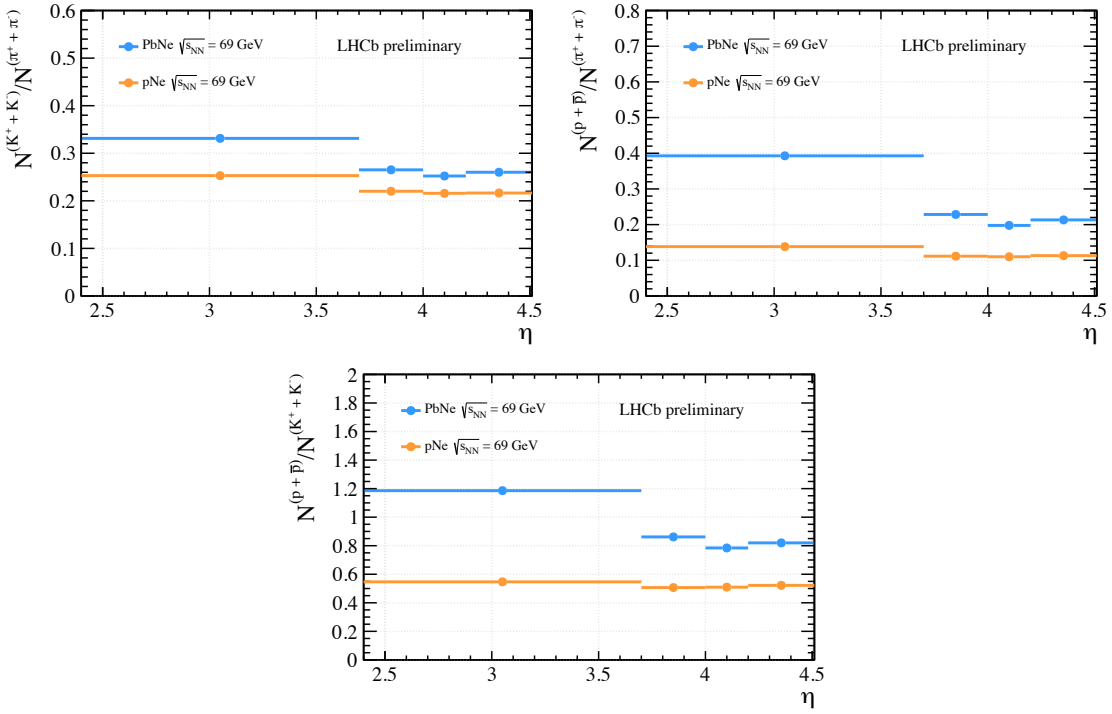


Figure 4.69: Single ratio of prompt K/π , p/π and p/K in PbNe and pNe data collected at $\sqrt{s_{NN}} = 69$ GeV, as a function of η, integrated over p_T and for collision centrality in the range of 88-39%. The error bars indicate the total uncertainty (statistical and systematic).

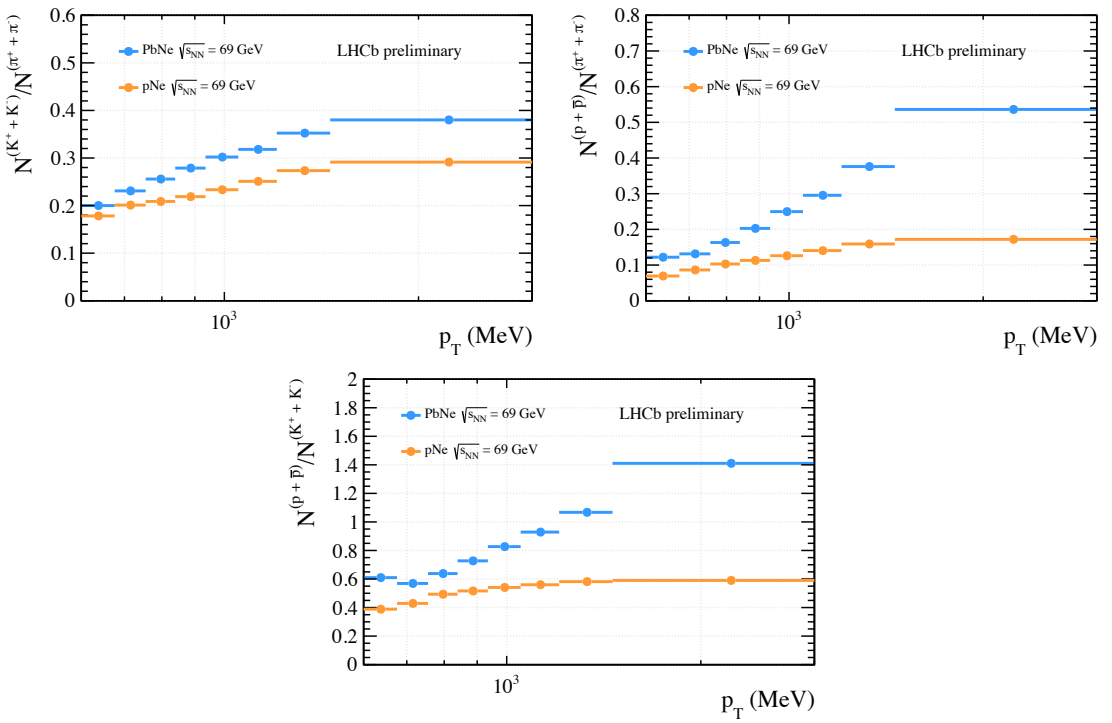


Figure 4.70: Single ratio of prompt K/π , p/π and p/K in PbNe and pNe data collected at $\sqrt{s_{NN}} = 69$ GeV, as a function of p_T, integrated over η and for collision centrality in the range of 88-39%. The error bars indicate the total uncertainty (statistical and systematic).

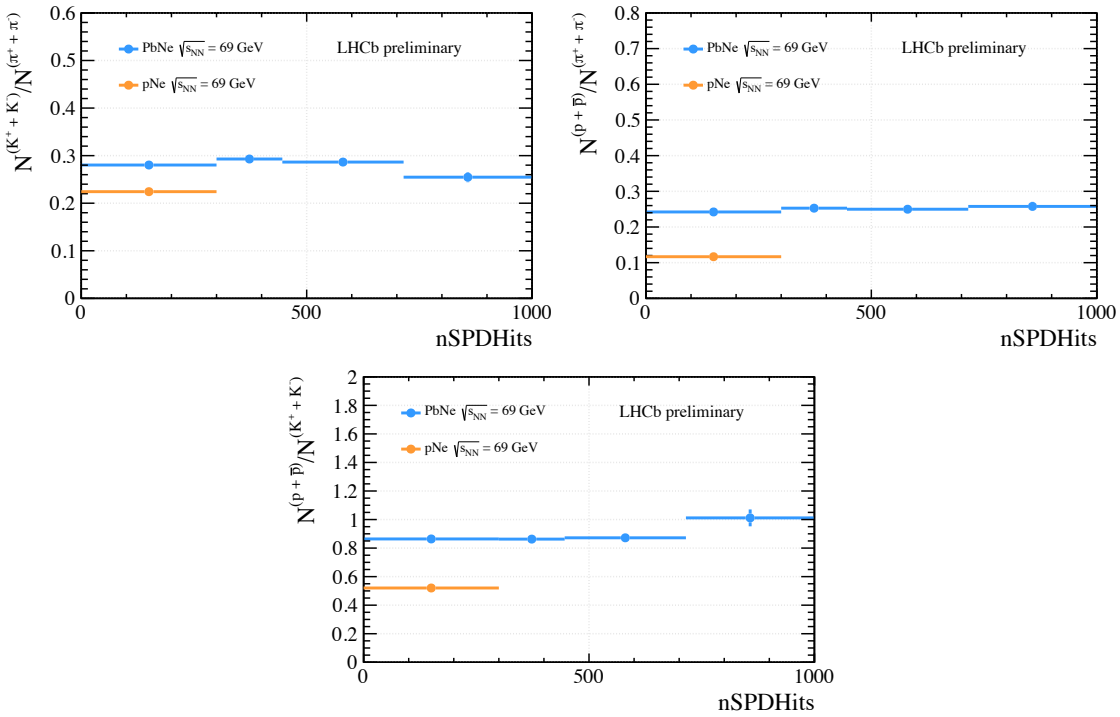


Figure 4.71: Single ratio of prompt K/π , p/π and p/K in PbNe and pNe data collected at $\sqrt{s_{NN}} = 69$ GeV, as a function of nSPDHits, integrated over η and p_T. The error bars indicate the total uncertainty (statistical and systematic).

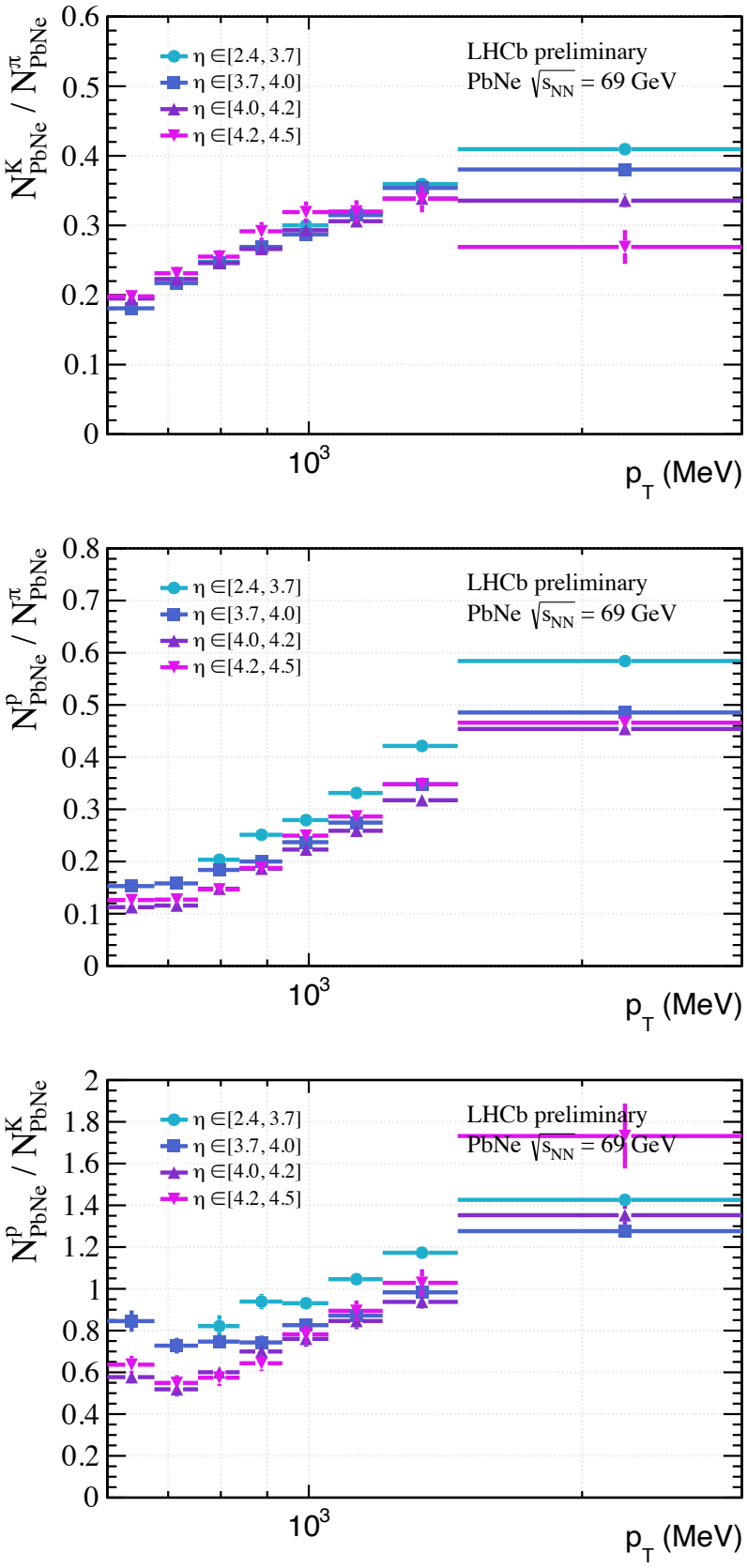


Figure 4.72: Single ratio of prompt K/π , p/π and p/K in PbNe data collected at $\sqrt{s_{NN}} = 69$ GeV, as a function of p_T for different bins of η, and for collision centrality in the range of 88-39%. The error bars indicate the total uncertainty (statistical and systematic).

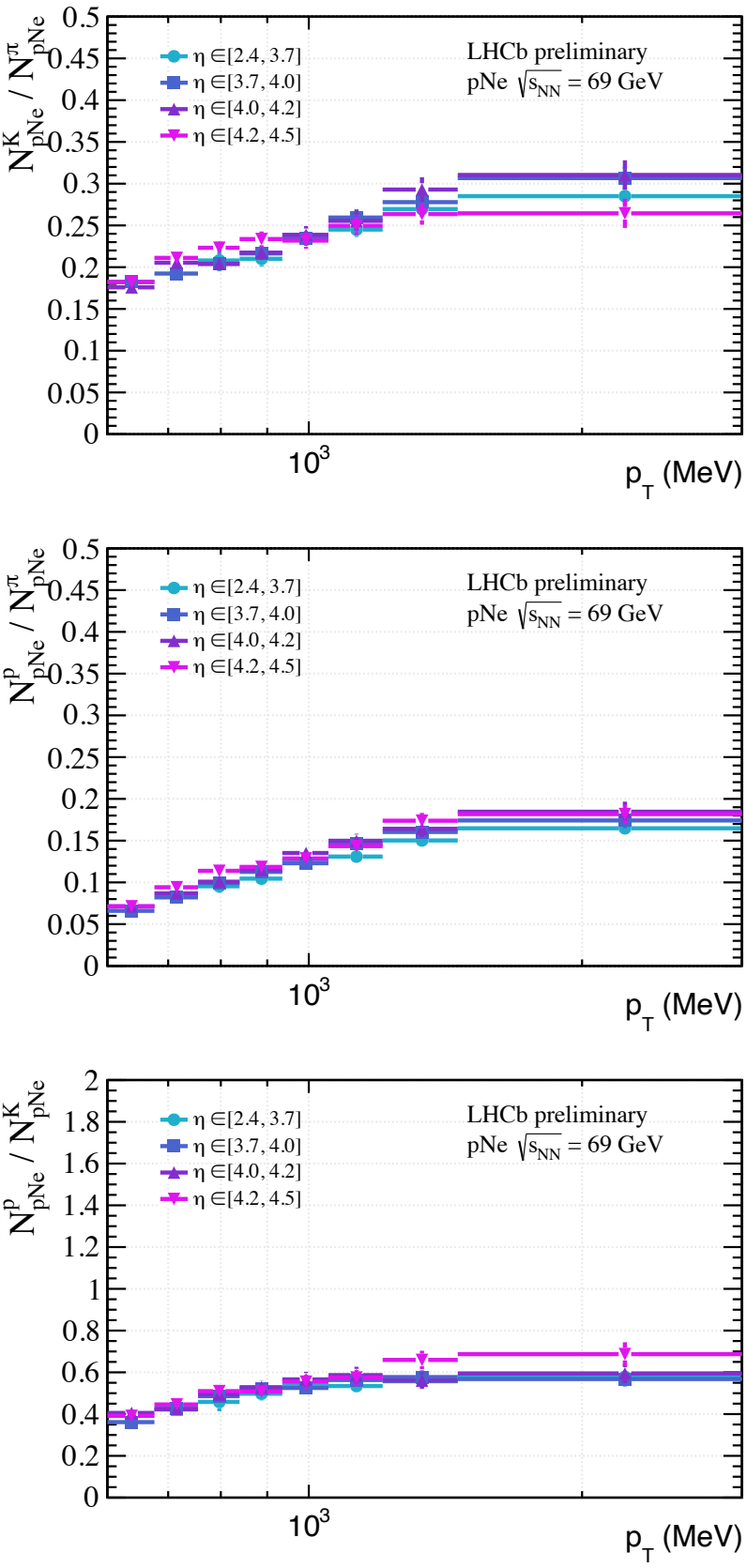


Figure 4.73: Single ratio of prompt K/π , p/π and p/K in pNe data collected at $\sqrt{s_{NN}} = 69$ GeV, as a function of p_T for different bins of η, and for collision centrality in the range of 88-39%. The error bars indicate the total uncertainty (statistical and systematic).

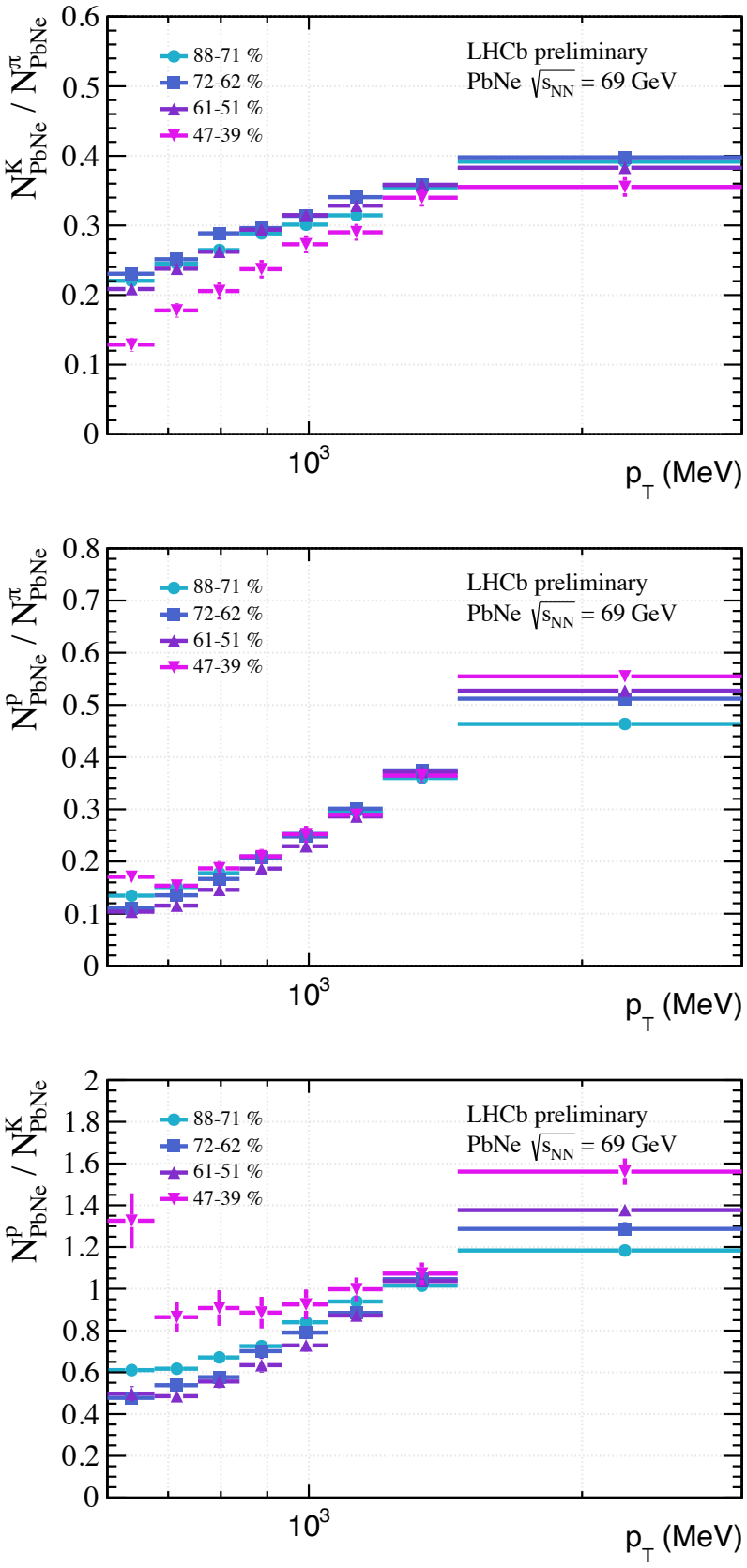


Figure 4.74: Single ratio of prompt K/π , p/π and p/K in PbNe data collected at $\sqrt{s_{NN}} = 69$ GeV, as a function of p_T for different bins of collision centrality, integrated over η. The error bars indicate the total uncertainty (statistical and systematic).

4.9.2 Double ratios

The double ratios, defined in Equation 4.5, allow for a direct assessment of modifications in particle production between a heavier system (PbNe) and a lighter one (pNe) in a controlled manner. As explained in Section 4.1, since a direct comparison of production cross-sections in PbNe and pNe collisions is unfeasible due to the absence of luminosity information in the PbNe dataset, the double ratios offer an alternative means of comparing the two datasets. Additionally, akin to the calculation of single ratios, computing double ratios helps to mitigate various unknowns and systematic uncertainties arising from both different data-taking conditions and the analysis strategy.

The double ratios of prompt K/π , p/π , and p/K between PbNe and pNe data collected at $\sqrt{s_{NN}} = 69$ GeV are presented in Figure 4.75 as functions of p_T, integrated over η and collision centrality. The results exhibit clear deviations from unity, indicating larger contributions from CNM effects in PbNe than in pNe due to the substantial size of the collision system. The deviation from unity observed in all three cases, expected due to the considerably larger volume of the PbNe system compared to the pNe one, exhibits a strong p_T dependence. Particularly, the p/π and p/K double ratios demonstrate a noticeable increase with p_T, starting at approximately 0.7 GeV/c, reaching a modification factor greater than 1.5 for $p_T > 1$ GeV/c. Conversely, a smaller increase is observed in the K/π double ratio as a function of p_T, reaching a maximum value between $1.3 - 1.4$. The double ratios as functions of η, integrated over p_T and collision centrality, are presented in Figure 4.76. A deviation from unity is observed in all three cases, showing a noticeable η dependence.

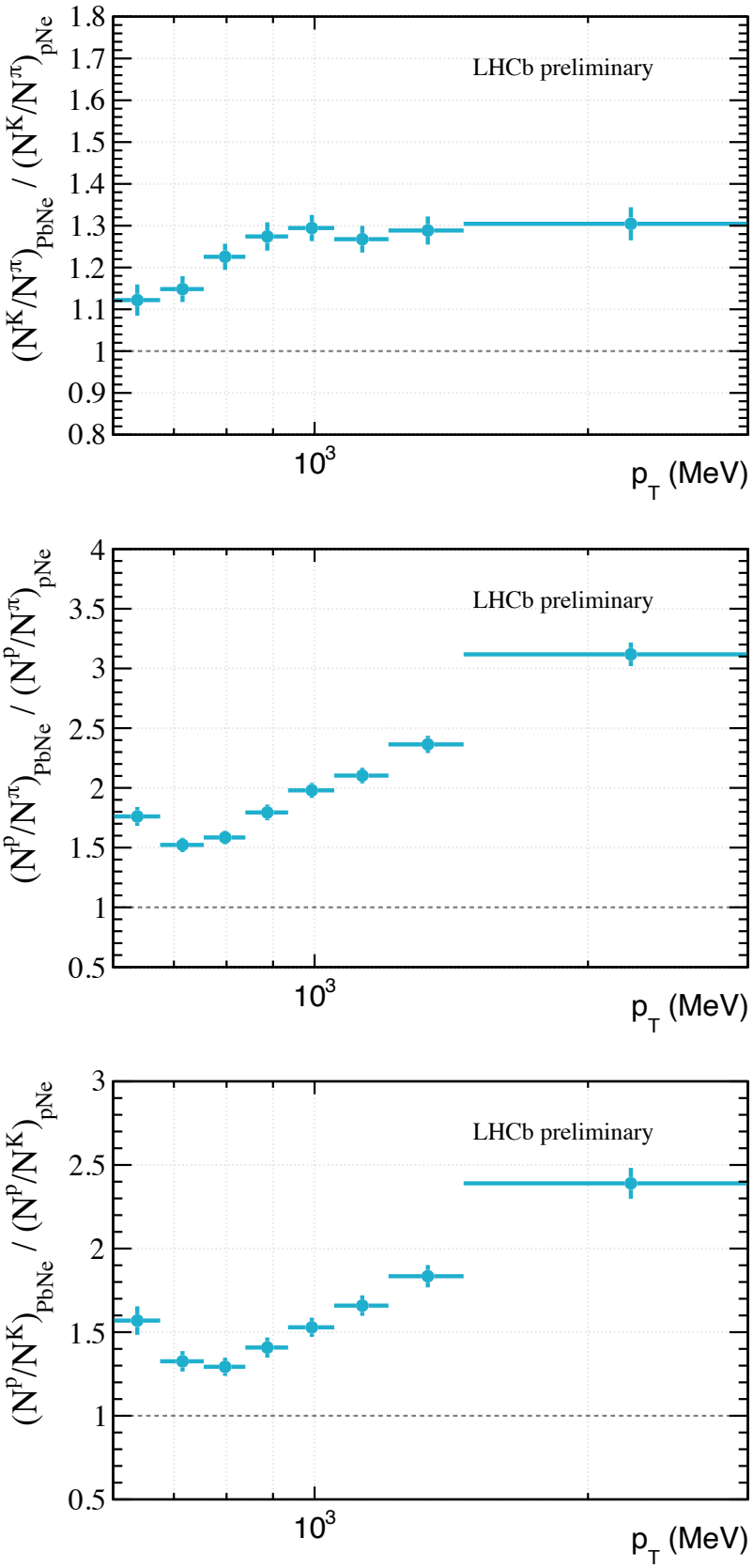


Figure 4.75: Double ratio of prompt K/π , p/π and p/K between PbNe and pNe data collected at $\sqrt{s_{NN}} = 69$ GeV as a function of p_T, integrated over η and for collision centrality in the range of 88-39%. A dotted line is added for better comparison with the double ratio consistent to unity. The error bars indicate the total uncertainty (statistical and systematic).

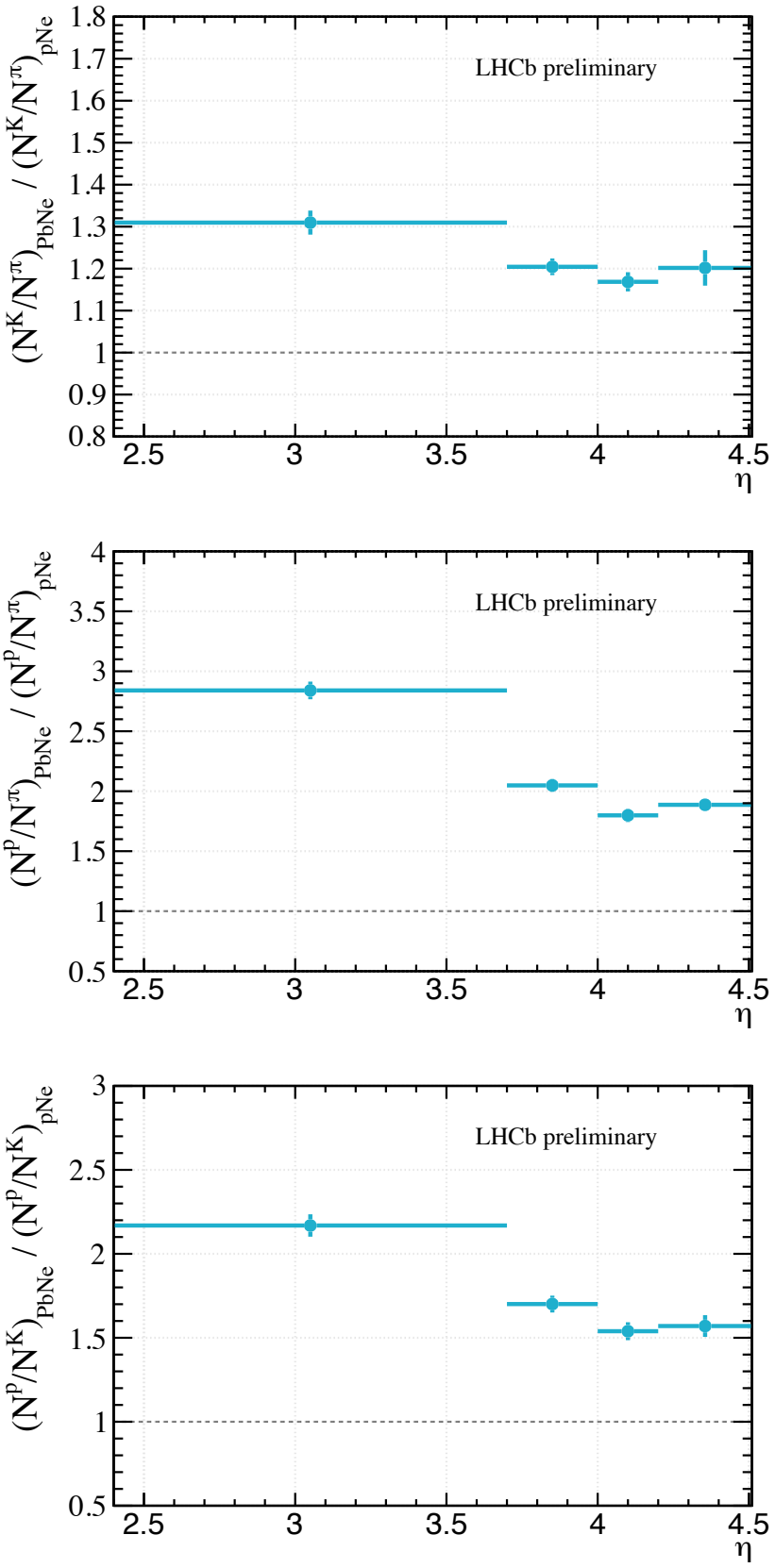


Figure 4.76: Double ratios of prompt K/π , p/π and p/K between PbNe and pNe data collected at $\sqrt{s_{NN}} = 69$ GeV as a function of η, integrated over p_T and for collision centrality in the range of 88-39%. A dotted line is added for better comparison with the double ratio consistent to unity. The error bars indicate the total uncertainty (statistical and systematic).

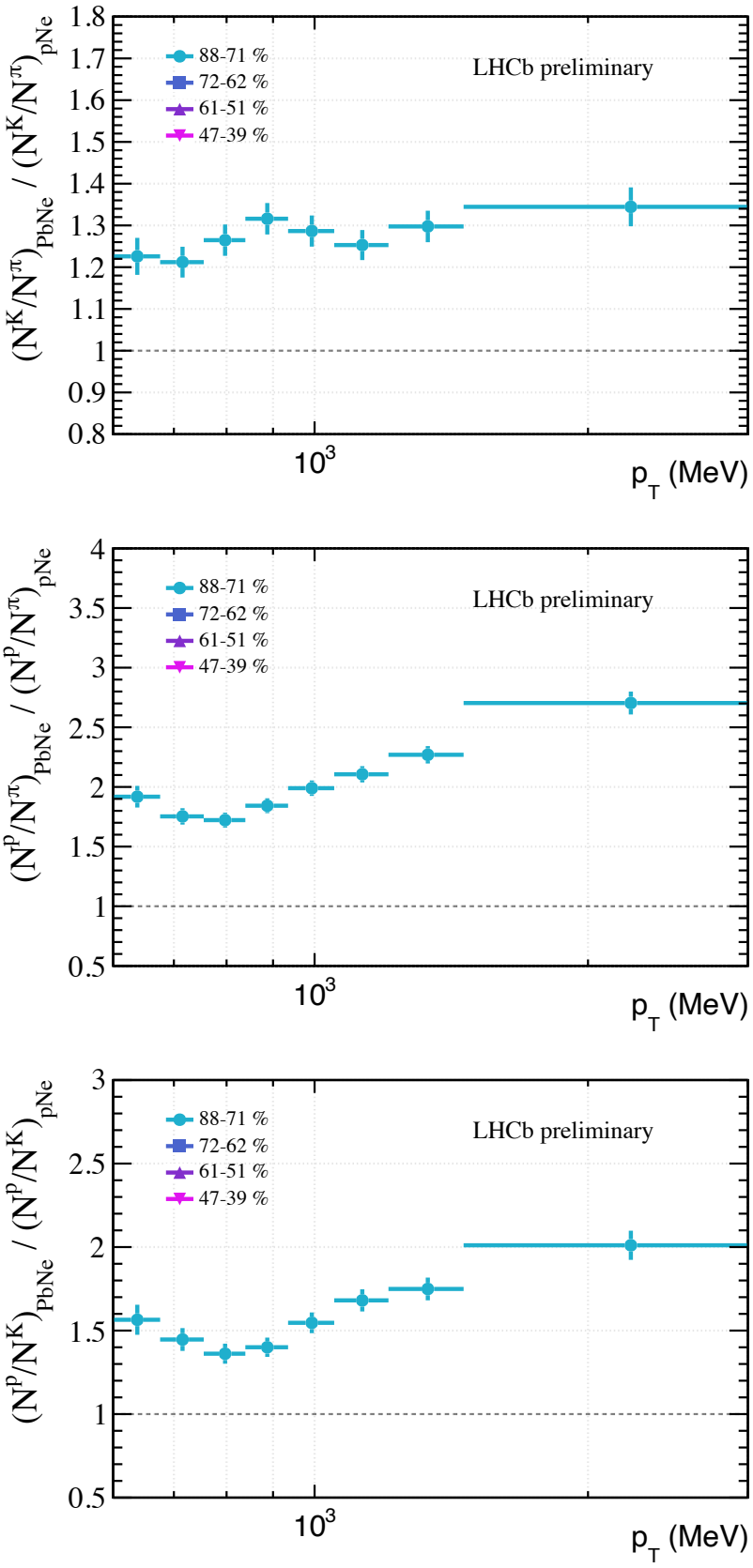


Figure 4.77: Double ratio of prompt K/π , p/π and p/K between PbNe and pNe data at $\sqrt{s_{NN}}$ = 69 GeV as a function of p_T for the collision centrality in the range of 88-71%, integrated over η. A dotted line is added for better comparison with the double ratio consistent to unity. The error bars indicate the total uncertainty (statistical and systematic).

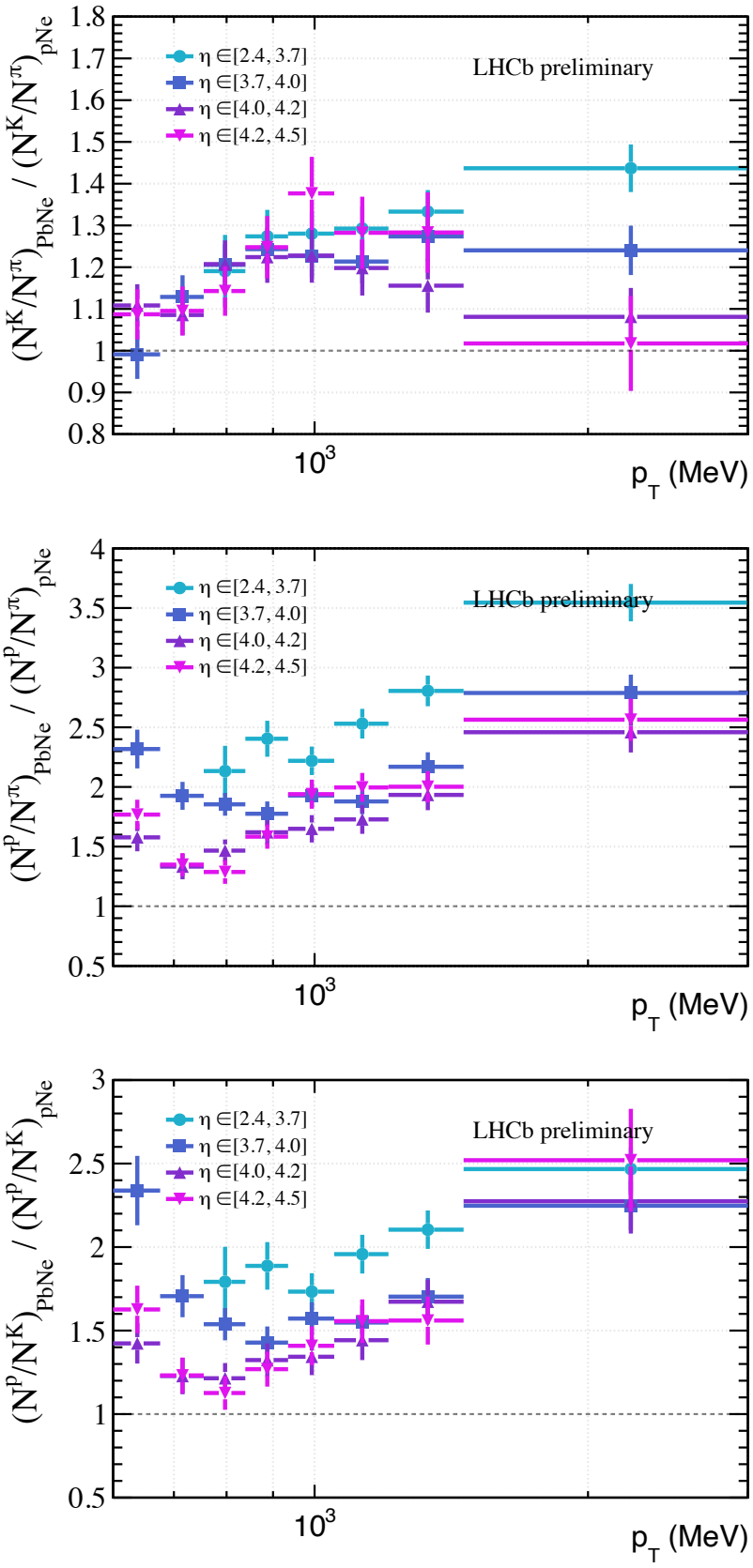


Figure 4.78: Double ratio of prompt K/π , p/π and p/K between PbNe and pNe data collected at $\sqrt{s_{NN}} = 69$ GeV as a function of p_T for different bins of η, and for collision centrality in the range of 88-39%. A dotted line is added for better comparison with the double ratio consistent to unity. The error bars indicate the total uncertainty (statistical and systematic).

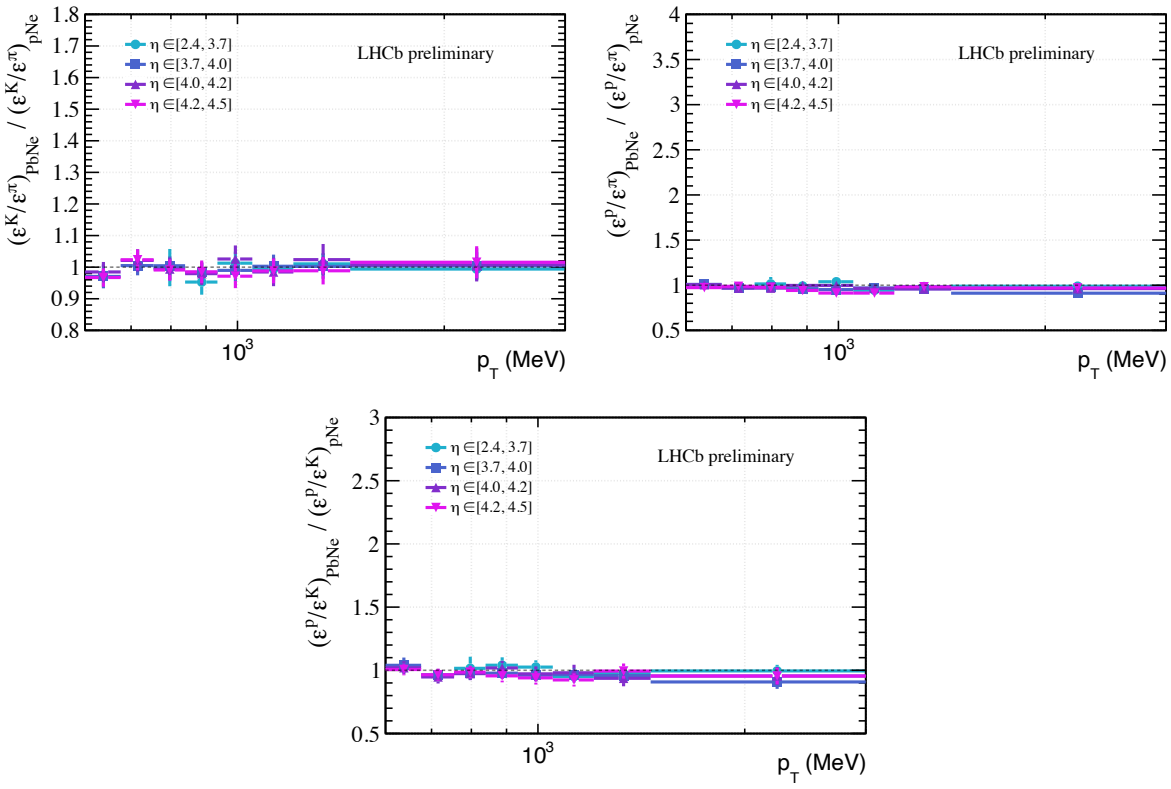


Figure 4.79: Double ratio of efficiencies in K/π , p/π and p/K between PbNe and pNe simulation at $\sqrt{s_{NN}} = 69$ GeV as a function of p_T for different bins of η, integrated over collision centrality. A dotted line is added for better comparison with the double ratio consistent to unity. The error bars indicate the total uncertainty (statistical and systematic).

4.9.3 Validation of the results

A crucial aspect of validating the outcomes for both single and double ratios involves assessing the impact of the simulation. This step is essential to ensure a meaningful comparison between the results derived from the data and to leverage the data for refining the simulation. Figure 4.79 illustrates the double ratios of efficiencies in the K/π , p/π , and p/K ratios, consistently hovering around unity for each η bin. This serves as a key cross-check, affirming that the impact of the simulation on the observed modifications in the double ratios, as depicted in Figures 4.75 and 4.78, is almost negligible. This observations ensures independence from the simulation, facilitating a direct and reliable comparison between experimental data and simulation.

4.9.4 Comparison with simulated data

After extracting the results for both single and double ratios, a compelling aspect is to compare these outcomes against simulation data. This comparative analysis, involving simulation generators, can yield additional insights into the observed modification effects.

The comparison between PbNe data and simulation for the single ratios K/π , p/π , and p/K is depicted in Figure 4.80 as a function of p_T, integrated over η and collision centrality. The K/π ratio in simulation consistently appears lower than in the data for all p_T bins. The p/π ratio seems well-described by the simulation, except for the high p_T region where a slight underestimation is evident. Conversely, the p/K ratio is consistently overestimated by the simulation, with no observed minimum in the distribution. Similar considerations can be made by comparing data and simulation as a function of η, as shown in Figure 4.81. The comparison in PbNe data is also performed as a function of nSPDHits, presented in Figure 4.82. The main discrepancies between data and simulation are observed at low multiplicity, especially for K/π and p/K single ratios. In contrast, the simulation appears to describe the data well in the highest multiplicity bin.

The comparison between pNe data and simulation for the single ratios K/π , p/π , and p/K is presented in Figure 4.83 as a function of p_T, integrated over η and collision centrality. The K/π ratio in simulation is slightly lower than in the data for all p_T bins. The p/π ratio shows a significant overestimation by the simulation, which becomes more pronounced with increasing p_T. Similarly, the p/K ratio is consistently overestimated by the simulation. Despite the absolute values, data and simulations exhibit similar trends as a function of p_T. Comparable considerations can be made by comparing data and simulation as a function of η, as demonstrated in Figure 4.84. In addition to the direct comparison of single ratios in data and simulation, double ratios are also compared between the two. This offers valuable insight into whether the simulation accurately captures CNM effects for these collision systems. To facilitate this comparison, the single ratios evaluated in the simulation of PbNe and pNe collisions are presented in Figure 4.85 as a function of p_T, integrated over η and multiplicity.

The comparison between the data and simulation for the double ratios K/π , p/π , and p/K is displayed in Figure 4.86 as a function of p_T, integrated over η and collision centrality. The K/π double ratio is consistent with the simulation at low p_T, whereas it is underestimated by the simulation for $p_T > 1$ GeV/c. The p/π and p/K double ratios are consistently underestimated by the simulation, with discrepancies increasing at higher p_T. In both data and simulation, a minimum is observed in the p/π and p/K double ratios at $p_T \sim 700$ MeV/c. Despite absolute values, data and simulations show similar trends as a function of p_T. Similar considerations can be made by comparing data and simulation as a function of η, as shown in Figure 4.87. The region where data and simulation show the largest discrepancy corresponds to $\eta < 3.7$.

Interestingly, the data-MC comparison for K/π and p/π ratios, shown in Figures 4.80 and 4.83, exhibits a qualitative agreement with the one shown in Figure 2.16, where similar ratios measured by CMS in the same p_T range are compared with the EPOS 1.99 model with and without core formation. The results presented in this section underscore that this analysis can

provide relevant inputs for the tuning of MC generator models, specifically for the EPOS 1.99 model, described in Section 2.5.1.

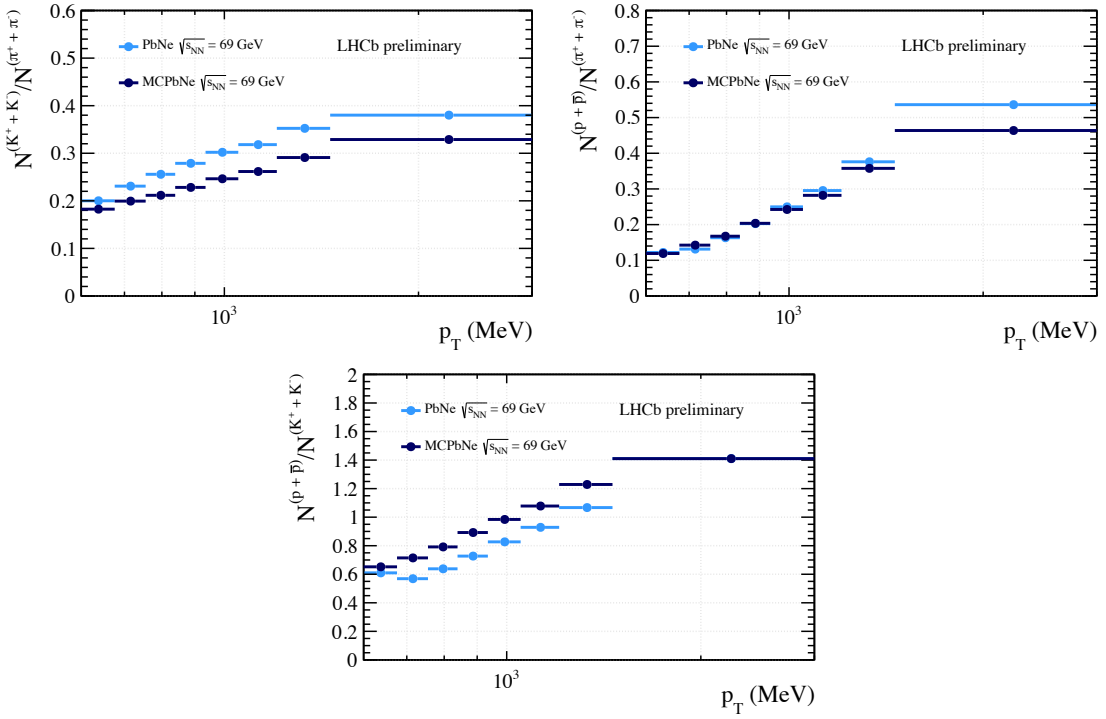


Figure 4.80: Single ratio of prompt K/π , p/π and p/K in PbNe data and simulation at $\sqrt{s_{NN}} = 69$ GeV, as a function of p_T, integrated over η and for collision centrality in the range of 88-39%. The error bars in real data indicate the total uncertainty (statistical and systematic).

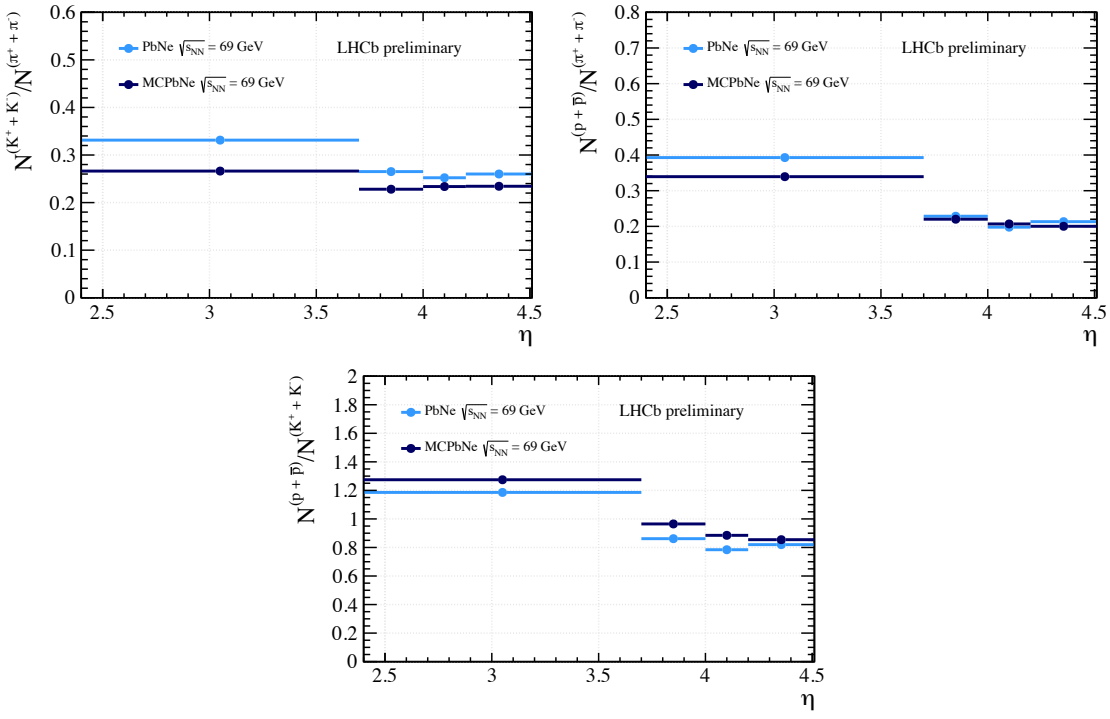


Figure 4.81: Single ratio of prompt K/π , p/π and p/K in PbNe data and simulation at $\sqrt{s_{NN}} = 69$ GeV, as a function of η, integrated over p_T and for collision centrality in the range of 88-39%. The error bars indicate the total uncertainty (statistical and systematic).

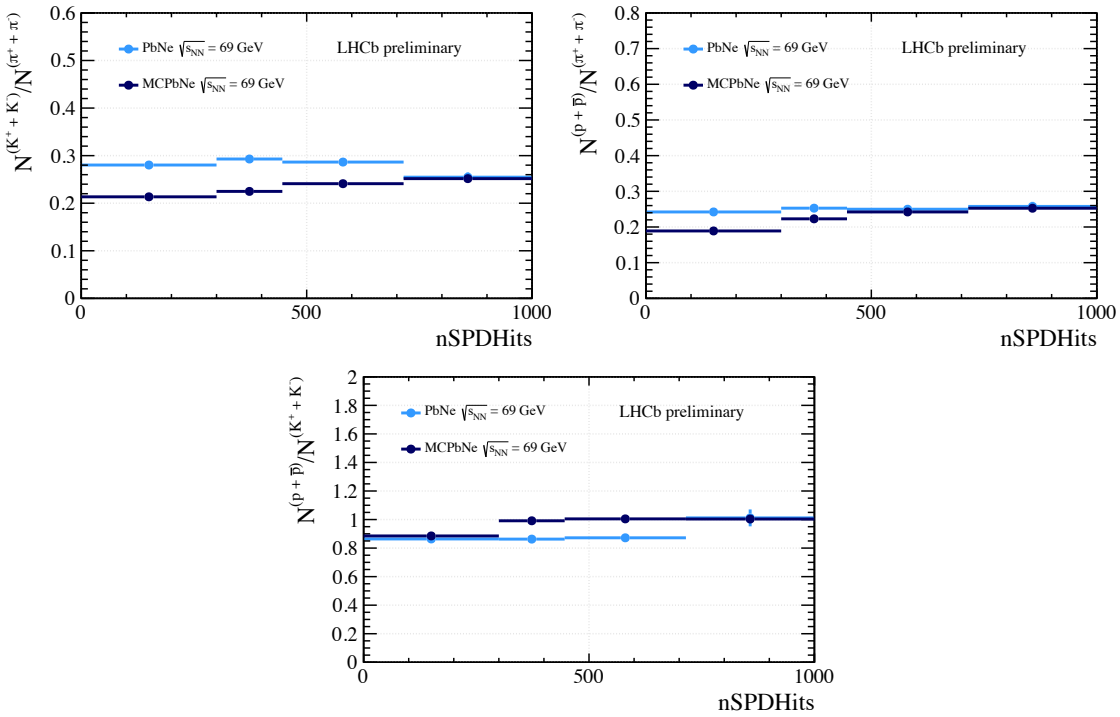


Figure 4.82: Single ratio of prompt K/π , p/π and p/K in PbNe data and simulation at $\sqrt{s_{NN}} = 69$ GeV, as a function of nSPDHits, integrated over η and p_T. The error bars indicate the total uncertainty (statistical and systematic).

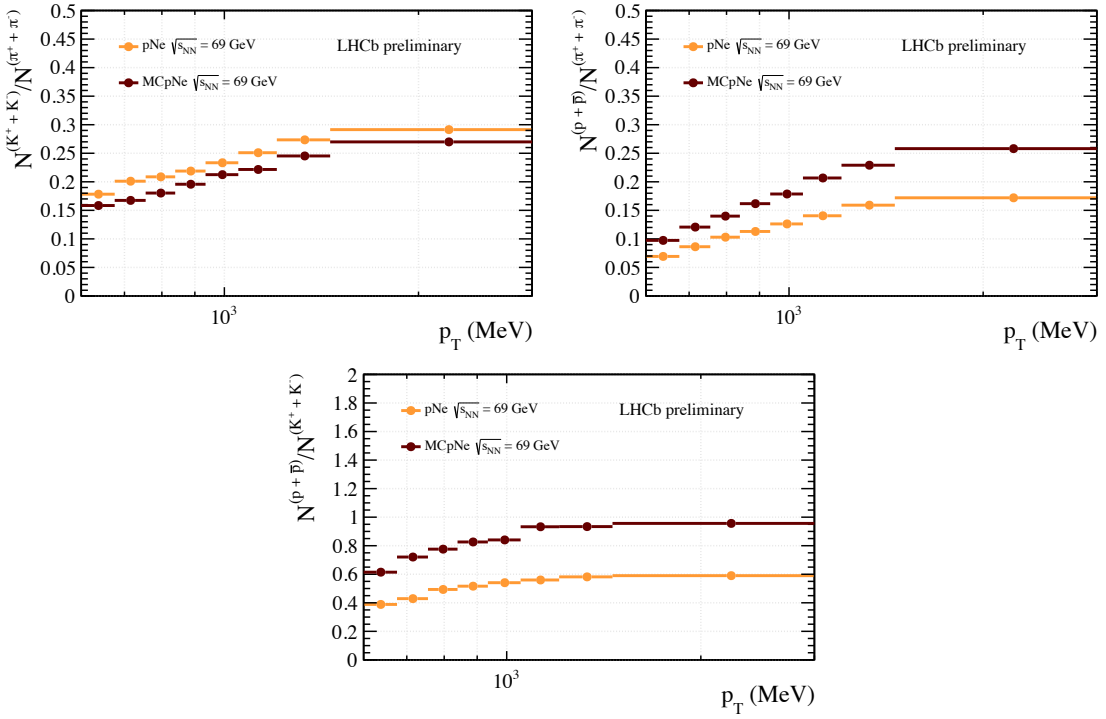


Figure 4.83: Single ratio of prompt K/π , p/π and p/K in pNe data and simulation at $\sqrt{s_{NN}} = 69$ GeV, as a function of p_T, integrated over η and for collision centrality in the range of 88-39%. The error bars indicate the total uncertainty (statistical and systematic).

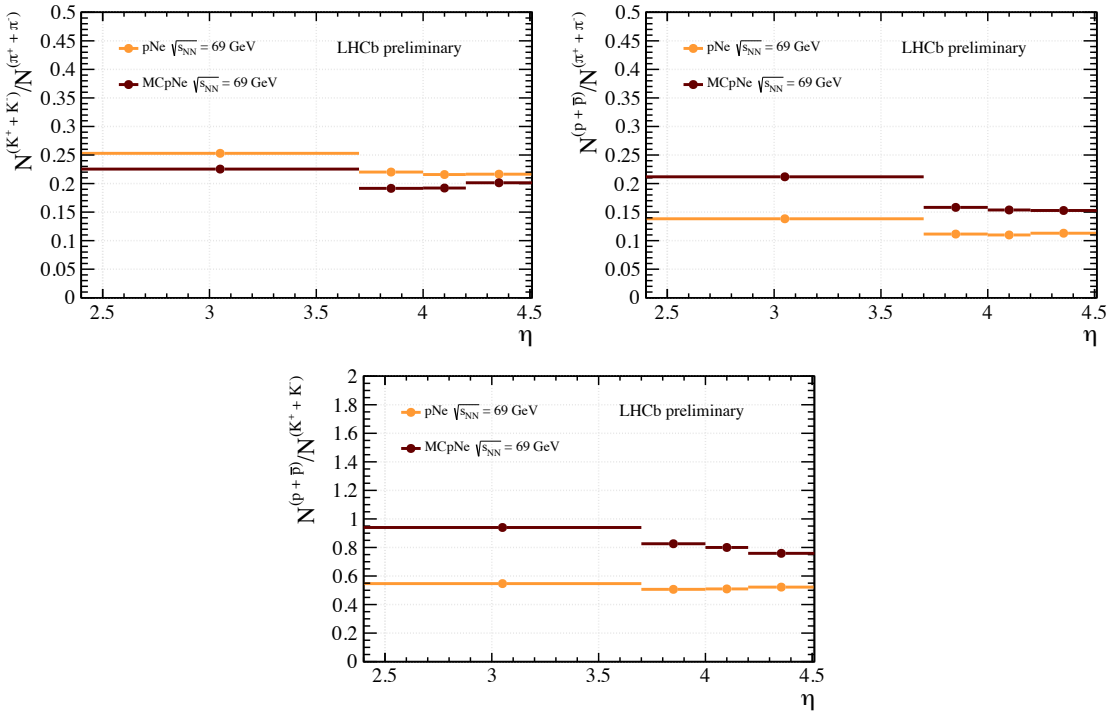


Figure 4.84: Single ratio of prompt K/π , p/π and p/K in pNe data and simulation at $\sqrt{s_{NN}} = 69$ GeV, as a function of η, integrated over p_T and for collision centrality in the range of 88-39%. The error bars indicate the total uncertainty (statistical and systematic).

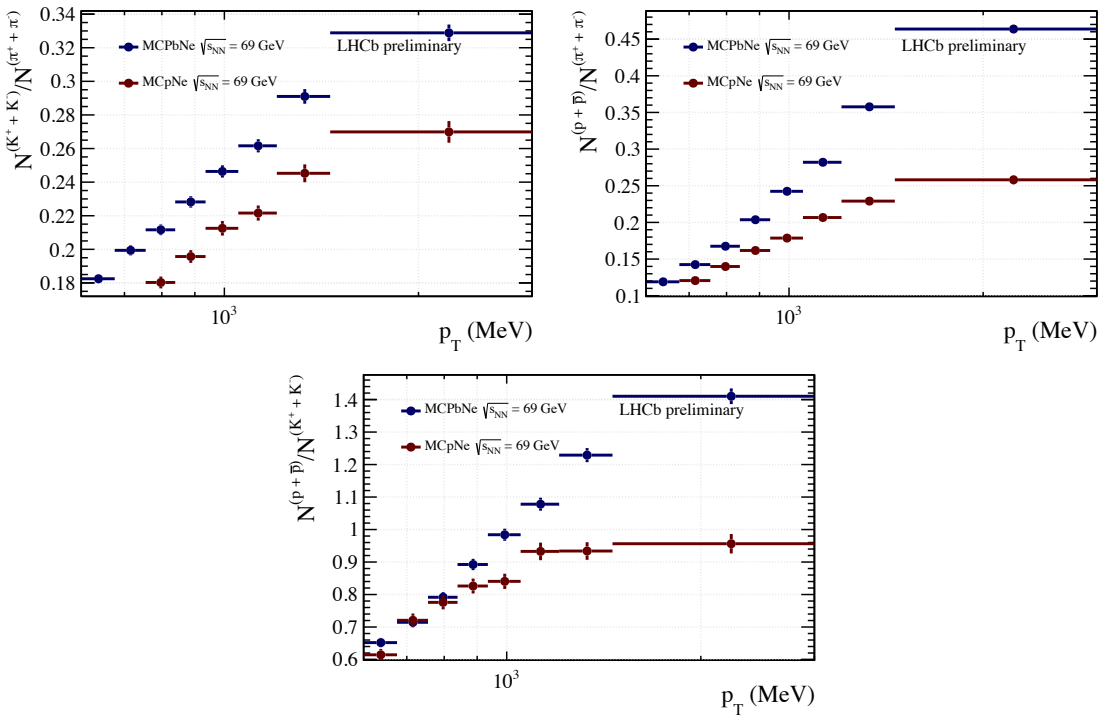


Figure 4.85: Single ratio of prompt K/π , p/π and p/K in pNe data and simulation at $\sqrt{s_{NN}} = 69$ GeV, as a function of p_T, integrated over η and for collision centrality in the range of 88-39%. The error bars indicate the total uncertainty (statistical and systematic).

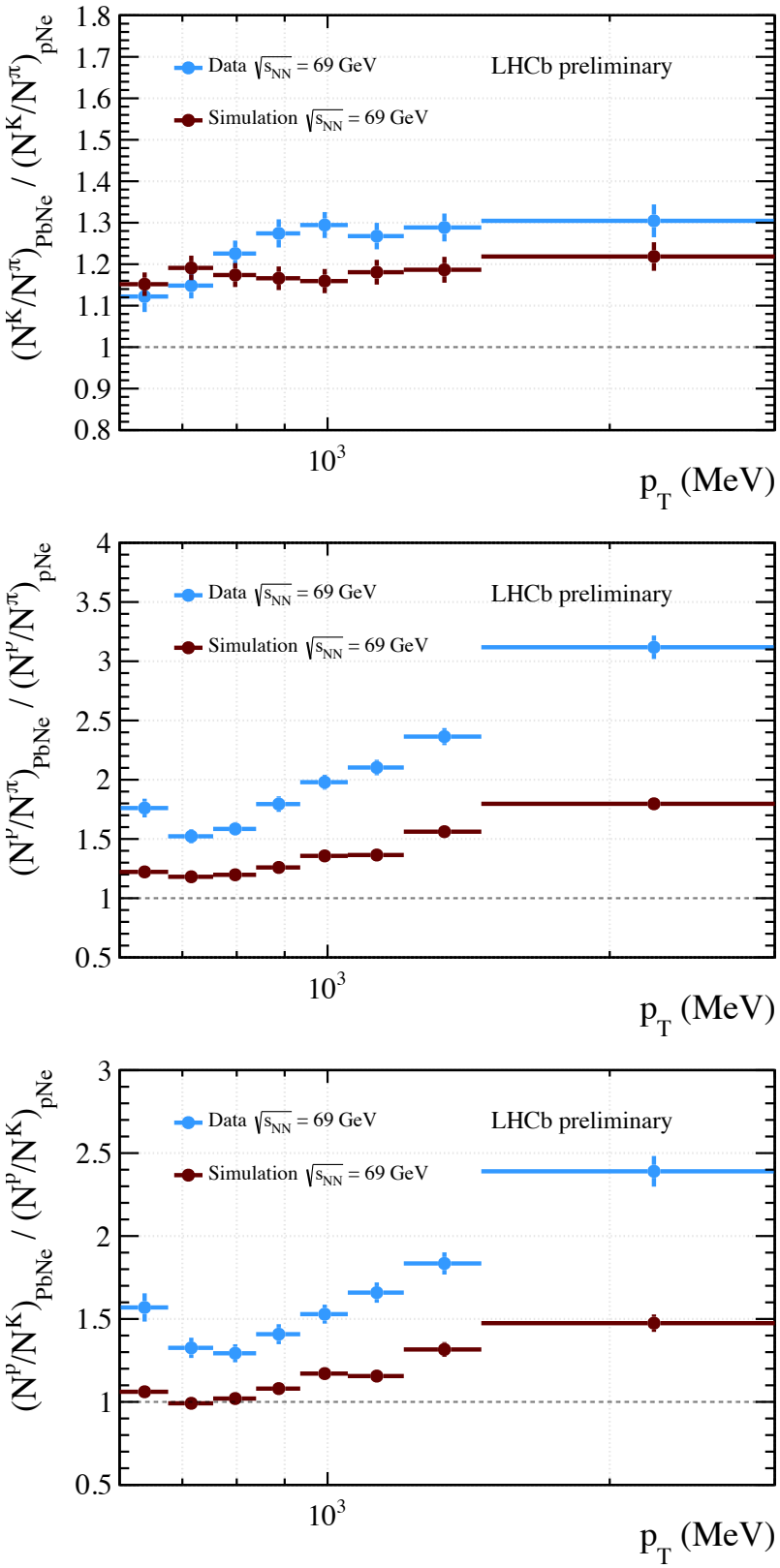


Figure 4.86: Double ratio of prompt K/π , p/π and p/K between PbNe and pNe data and simulation at $\sqrt{s_{NN}} = 69$ GeV as a function of p_T, integrated over η and for collision centrality in the range of 88-39%. A dotted line is added for better comparison with the double ratio consistent to unity. The error bars indicate the total uncertainty (statistical and systematic).

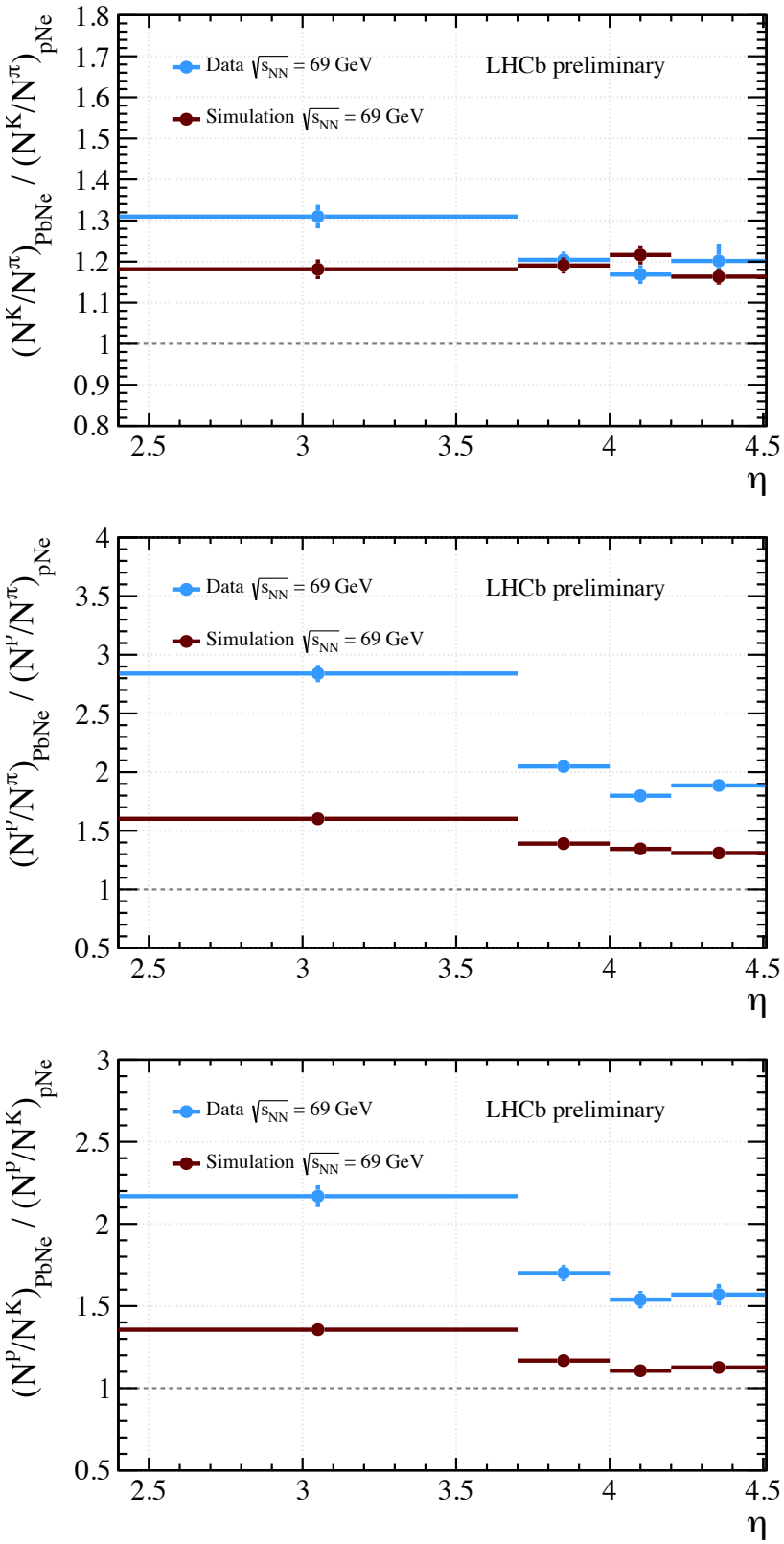


Figure 4.87: Double ratio of prompt K/π , p/π and p/K between PbNe and pNe data and simulation at $\sqrt{s_{NN}} = 69$ GeV as a function of η, integrated over p_T and for collision centrality in the range of 88-39%. A dotted line is added for better comparison with the double ratio consistent to unity. The error bars indicate the total uncertainty (statistical and systematic).

Chapter 5

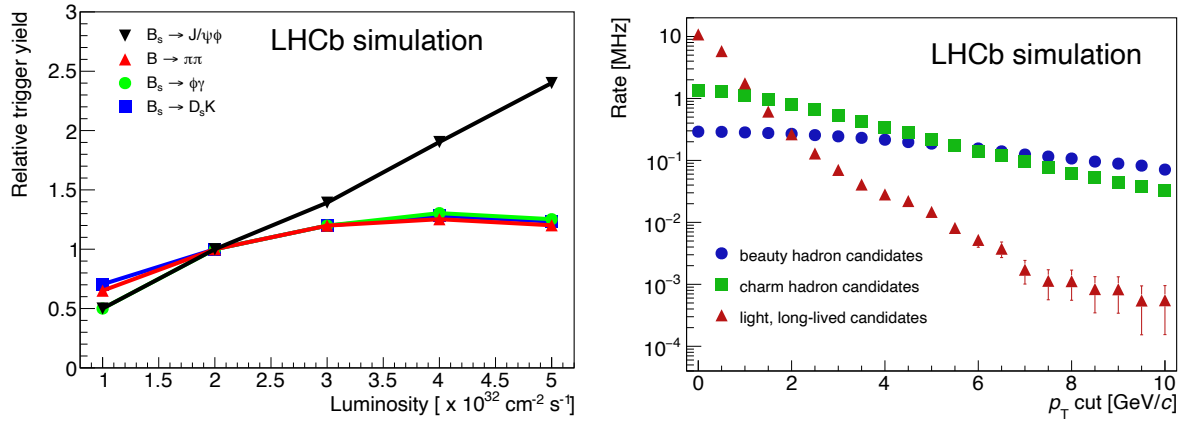
The LHCb Upgrade I

In this chapter, an extensive overview of the LHCb upgrade is provided. A general description of the upgrade is presented in Section 5.1, followed by specific discussions on the upgraded tracking system in Section 5.2 and the enhanced particle identification systems in Section 5.3. Section 5.4 details the replacement of the hardware trigger with a real-time fully-software data processing strategy. Lastly, Section 5.5 presents the upgrade of the fixed-target program in Run 3.

5.1 Motivations for the Run 3 upgrade

LHCb demonstrated outstanding performance throughout the LHC Run 1 (2010–2012) and Run 2 (2015–2018) data-taking periods, successfully operating from 2010 to 2018 [93]. The experiment collected a total of 9 fb^{-1} of pp data, approximately 30 nb^{-1} of PbPb and pPb collisions, and around 200 nb^{-1} of fixed-target data. Despite the substantial dataset, the precision of several crucial flavour-physics observables investigated and measured by LHCb is still limited by statistical constraints [163]. Consequently, achieving the necessary sensitivity to detect potential new-physics effects and probing the Standard Model at the level of precision attained by theoretical calculations requires substantially larger datasets.

The design of the LHCb Run 1-2 system posed a challenge in increasing statistics, especially for fully hadronic final state decays. The main bottleneck arises from the maximum allowable output rate of the L0 trigger. The simplistic inclusive selection criteria implemented at the L0 trigger stage, predominantly based on particle transverse momentum requirements, leads to an efficiency loss with rising luminosity. This phenomenon is particularly pronounced for the most common processes featuring hadrons in the final state, resulting in the saturation of the event yield, as shown in the left panel of Figure 5.1. In addition, inclusive flavour physics signals have relatively large cross sections and, at the upgrade instantaneous luminosity of $L = 2 \times 10^{33}$ cm^{-2} s^{-1}, every event in the LHCb acceptance will contain on average two long-lived hadrons not containing heavy quarks [164, 165]. Consequently, straightforward cuts relying on displaced vertices or p_T would either be ineffective in background rejection or, once achieving sufficient purity, result in downscaling the signal, as illustrated in the right panel of



(a) (b)

Figure 5.1: (a) Normalised relative trigger yields plotted against instantaneous luminosity, with normalisation to $L = 2 \times 10^{33}$ cm^{-2} s^{-1}. (b) The rate of reconstructed decays within the LHCb acceptance plotted against the cut in p_T of the decaying particle, considering a decay time $\tau > 0.2$ ps [164].

Figure 5.1. Therefore, profiting on a higher luminosity to collect significantly more data is only feasible by eliminating the L0 trigger stage and introducing selections that are more discriminating than simple inclusive criteria. In particular, an entirely software-based trigger, capable of discriminating signal channels based on full event reconstruction, has been deemed necessary for this approach.

The revised trigger strategy, higher luminosity, and the resulting higher pile-up demanded an upgrade of the LHCb detectors and readout electronics [166]. Figure 5.2 illustrates the configuration of the upgraded detector. A comprehensive revision of the software, data processing, and computing strategy was also essential to handle the anticipated significant increase in data volume. Considering these factors, the LHCb upgrade has been designed to operate at a nominal instantaneous luminosity of $L = 2 \times 10^{33}$ cm^{-2} s^{-1} (a factor of five higher than in Run 2) and to capture events at the LHC crossing rate of 40 MHz. The expected total integrated luminosity, including Run 1 and Run 2, is approximately 50 fb^{-1} by the conclusion of Run 4 of the LHC. The projected distribution of the number of vertices per event visible in LHCb is displayed in Figure 5.3. At the upgrade conditions, where the average number of visible interactions per bunch crossing (μ) is approximately 5.2, the number of empty events is nearly eliminated, in contrast to the data-taking conditions in Run 2 ($\mu \sim 1.7$). The upgrade of the detector has been performed during the Long Shutdown 2 of the LHC (2019-2021) and the commissioning completed during the first phase of Run 3.

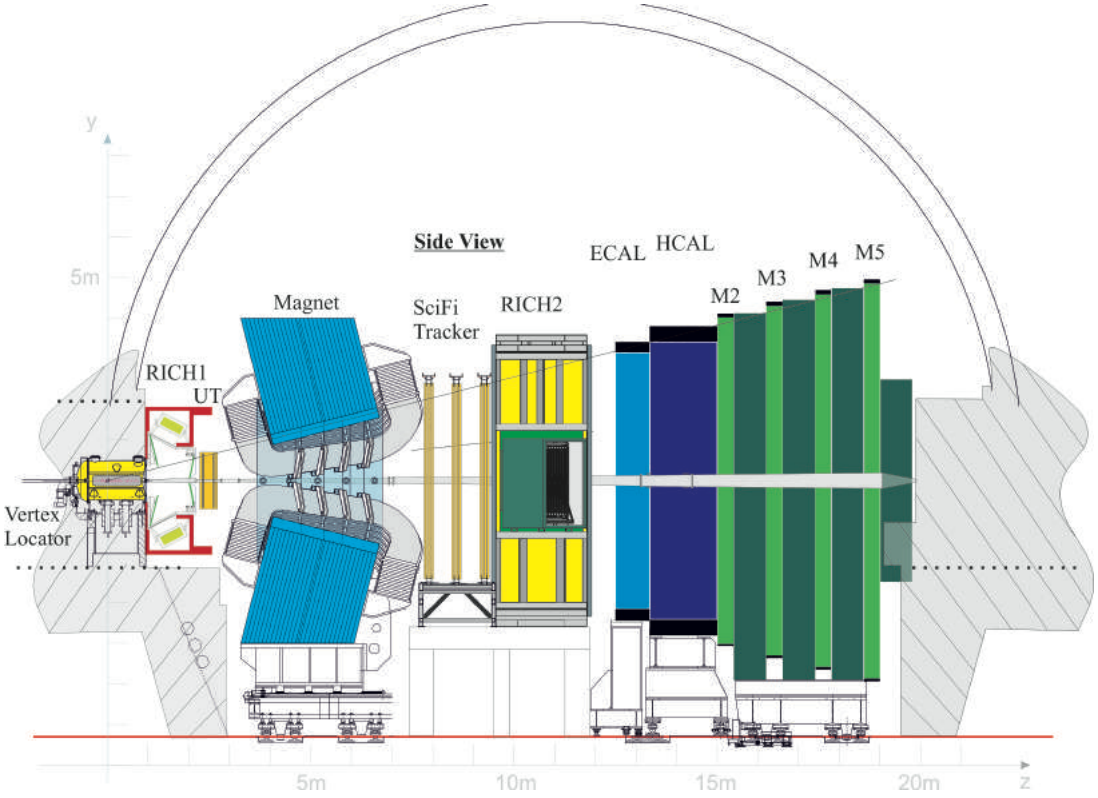


Figure 5.2: Layout of the upgraded LHCb detector [166].

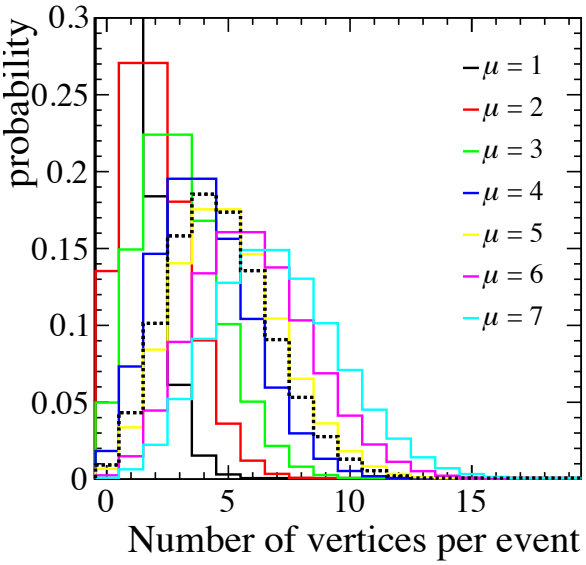


Figure 5.3: Number of vertices per event for operation at different values of μ. The default simulation value, corresponding to 2×10^{33} cm^{-2} s^{-1}, is marked by the dotted line [167].

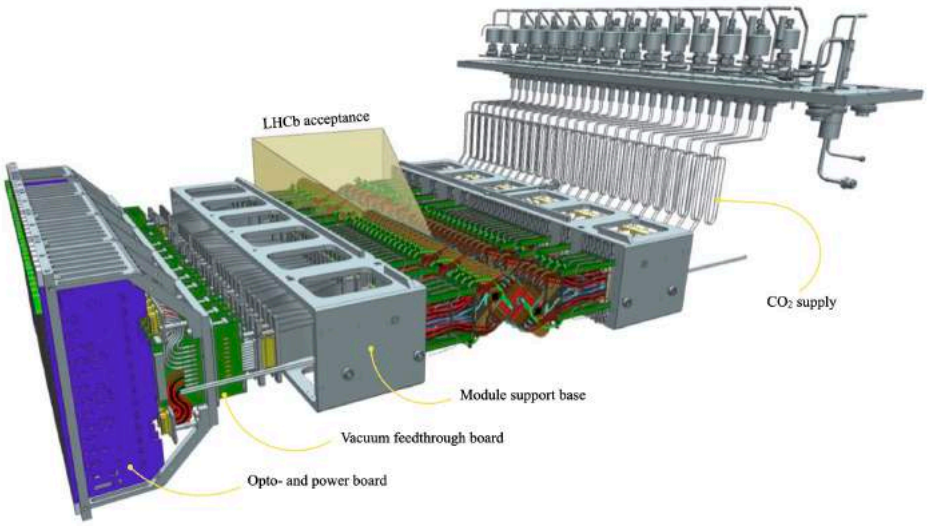


Figure 5.4: Visualisation of the VELO halves displaying modules mounted on the support bases and the LHCb acceptance depicted as a transparent pyramid.

5.2 The tracking system upgrade

The upgraded LHCb tracking system maintains the fundamental configuration of the previous tracking system. The individual detectors have been replaced: the upgraded VELO detector features a silicon pixel detector; the TT has been replaced by the Upstream Tracker (UT); a SciFi Tracker has been installed in place of the T stations.

5.2.1 VELO

The VELO has been redesigned [167] to be compatible with the increased luminosity and the trigger-less 40 MHz readout requirement of the upgraded experiment. The system must continue to deliver pattern recognition within an acceptable CPU budget while maintaining the highest track-finding efficiency. With the LHCb acceptance remaining unchanged, the optimised layout of the VELO resembles its predecessor, as illustrated in Figure 5.4. Active elements and their services are organised into a series of identical modules, equipped with pixelated ASICs, arranged perpendicular to the beamline. The distribution of the modules must span the full pseudorapidity acceptance of LHCb and ensure that the majority of tracks from the interaction region traverse at least four pixel sensors in the entire azimuthal range [168]. With the chosen sensor arrangement shown in Figure 5.5, 52 modules are required to meet these criteria, including the modules placed upstream of the interaction region, designed to enhance the unbiased measurement of primary vertices.

The core technology of the new VELO is pixelated hybrid silicon detectors, which are arranged into modules and cooled by a silicon microchannel cooler. Of the mechanical structures, only the principal vacuum vessel and motion services remain from the version that was in operation until 2018. In particular, the RF boxes, the enclosures that interface the detector to the LHC beams, were entirely redesigned reducing both material budget and the inner radius of

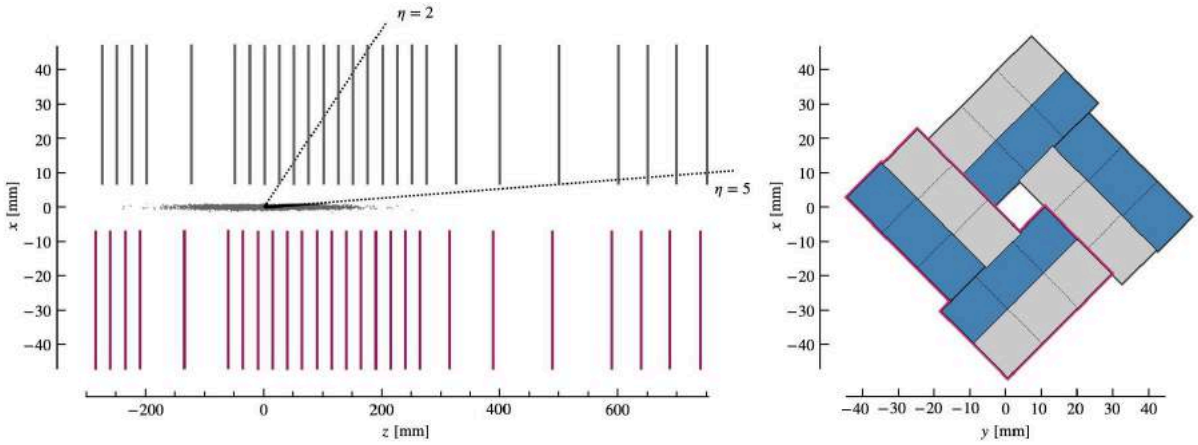


Figure 5.5: (Left) Schematic top view of the $z-x$ plane at $y=0$ illustrating the z-extent of the luminous region and the defined LHCb pseudorapidity acceptance, $2<\eta<5$. (Right) Layout of the standard arrangement of the ASICs around the z-axis in the closed VELO configuration. Half of the ASICs are positioned on the upstream module face (grey), and the remaining half on the downstream face (blue). The modules on Side C are highlighted in purple on both sketches.

the VELO along the beam line. The VELO module brings together the silicon detectors, their cooling, powering, readout and mechanical support into a single, repeating unit, as illustrated in Figure 5.6.

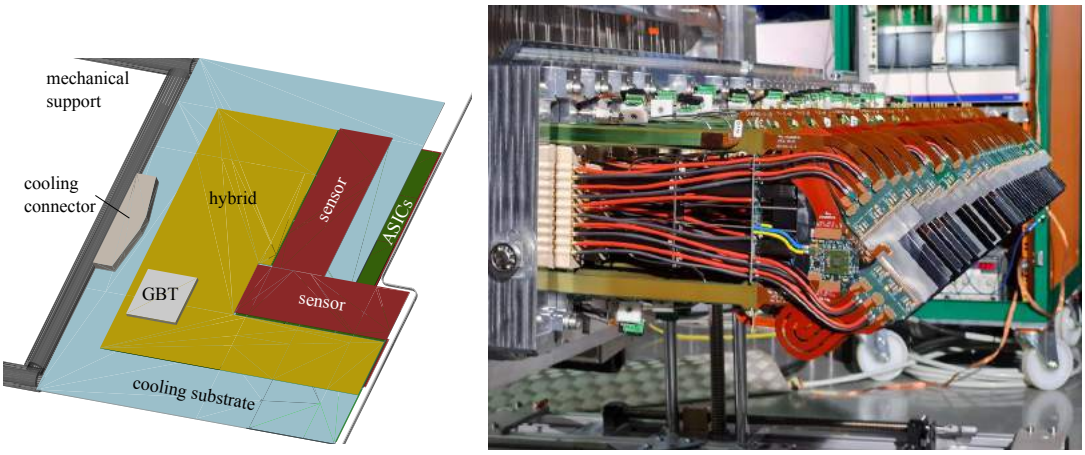
5.2.2 Upstream Tracker

The UT [169] is positioned between the RICH 1 detector and the dipole magnet, serving as a charged-particle tracking device with the same objectives and structure as the TT stations used in Run 1 and Run 2. The UT detector, illustrated in Figure 5.7, consists of four planes of silicon detectors organised into two stations, with a hole in the middle to accommodate the beam pipe. The silicon strip pitches and lengths are matched to the expected occupancy and are arranged in vertical units known as staves, as depicted in Figure 5.7. A UT stave serves as the mechanical support for the sensors and front-end electronics, and it incorporates active cooling. The sensors are positioned on both sides of the staves to ensure complete coverage in the vertical direction. Likewise, a z-staggered arrangement of the staves with horizontal overlaps enables full horizontal coverage. Additionally, special sensors are employed in the innermost area to maximise the active area near the beam pipe. These improvements, compared to the TT, significantly reduce the gaps in the acceptance.

The UT plays a crucial role in the initial processing algorithm of the software trigger, enabling the preliminary determination of track momentum with a precision of approximately 15% for tracks with $p_T > 0.2$ GeV/c.

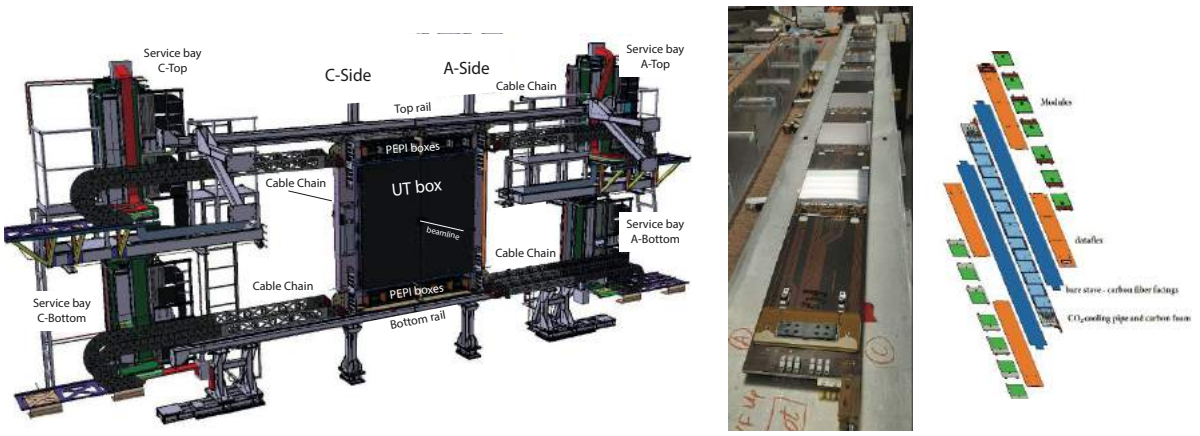
5.2.3 Scintillating Fibre Tracker

The SciFi Tracker [169] is positioned downstream of the LHCb dipole magnet and is responsible with charged-particle tracking and momentum measurement. The SciFi acceptance spans



(a) (b)

Figure 5.6: (a) Arrangement of a VELO module, as incorporated in the LHCb simulation framework, displaying the locations of key components, including a cross-section of the RF foil at the module's z-position [167] (b) The Side C half, with 26 modules, ready for the installation into the vacuum vessel [166].



(a) (b)

Figure 5.7: (a) 3D view of the UT system. (b) Fully assembled stave, featuring the visible end of the cooling tube with high voltage and signal connections at the bottom [169].

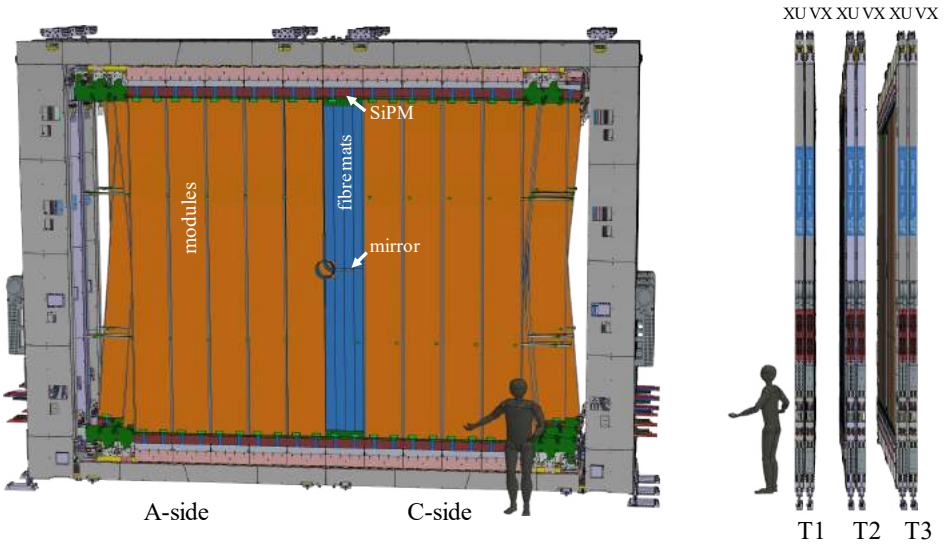
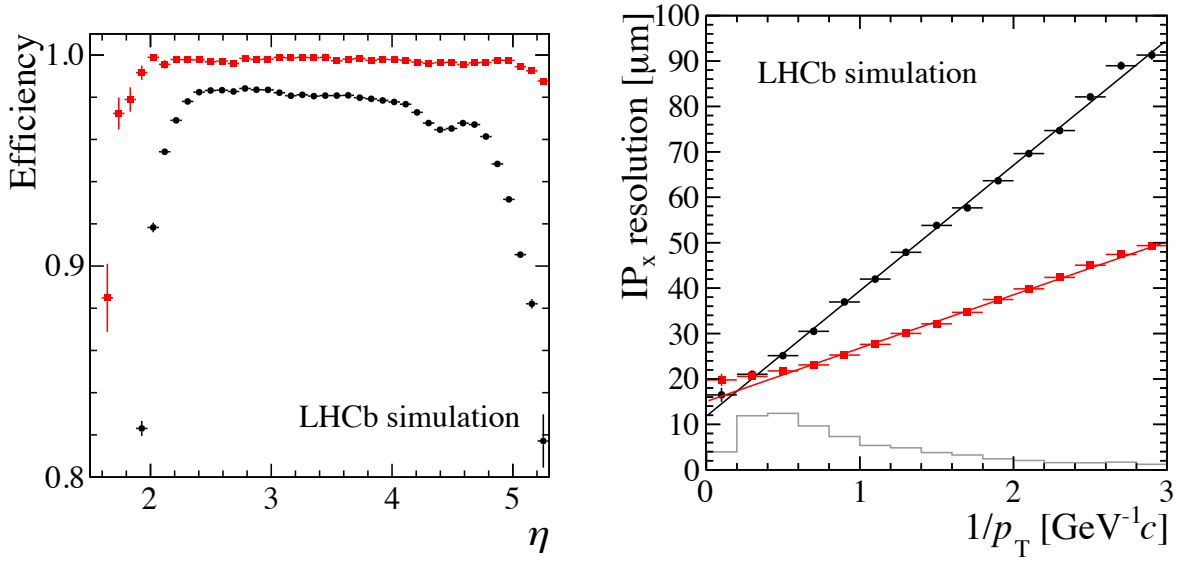


Figure 5.8: 3D representations depicting the SciFi Tracker detector from both frontal and lateral perspectives [166].

from approximately 20 mm near the edge of the beam pipe to distances of ±3186 mm and ±2425 mm in the horizontal and vertical directions, respectively. The detection system utilises a single technology employing 250 μm diameter plastic scintillating fibres organised in multi-layered fibre mats. There are a total of 12 detection planes distributed across 3 stations (T1, T2, T3), each comprising 4 layers arranged in an $xuvx$ configuration, as illustrated in Figure 5.8.

The detector modules are constructed in a honeycomb and carbon-fibre sandwich configuration, housing eight SciFi mats that are approximately 2.4 m long and 13 cm wide, composed of six staggered layers of fibres. A thin mirror is attached to the fibre end to reflect additional light back to the readout side [170]. In the experiment, these mirrors are positioned near the y = 0 plane. Four fibre mats point upward, and four point downward, covering a total height of almost 5 m. Special modules are used near x = 0 to account for the presence of the beam pipe. The optical signal from the scintillating fibres is detected by 128-channel arrays of SiPMs with a channel pitch of 250 μm. At the readout end of each module, 16 SiPM arrays are bonded to a 3D-printed titanium alloy cooling bar, aligned to the four fibre mats and housed in a cold-box.

Blue-green emitting double-clad plastic scintillating fibres with a diameter of 250 μm have been selected for the LHCb SciFi Tracker. The production of fibre mats involves winding six layers of fibres on a threaded winding-wheel with a diameter of approximately 82 cm [171]. This winding process arranges the fibres in a regular hexagonal matrix with a horizontal pitch of 275 μm. The tracker is equipped with over 500000 SiPM channels organised into 4096 arrays of 128 channels each, responsible for detecting the light emitted by the scintillating fibres. Each channel consists of 4×26 pixels connected in parallel, with each pixel measuring 57.5 μm × 62.5 μm. Consequently, a single channel has dimensions of 230 μm × 1625 μm, featuring a channel pitch of 250 μm.



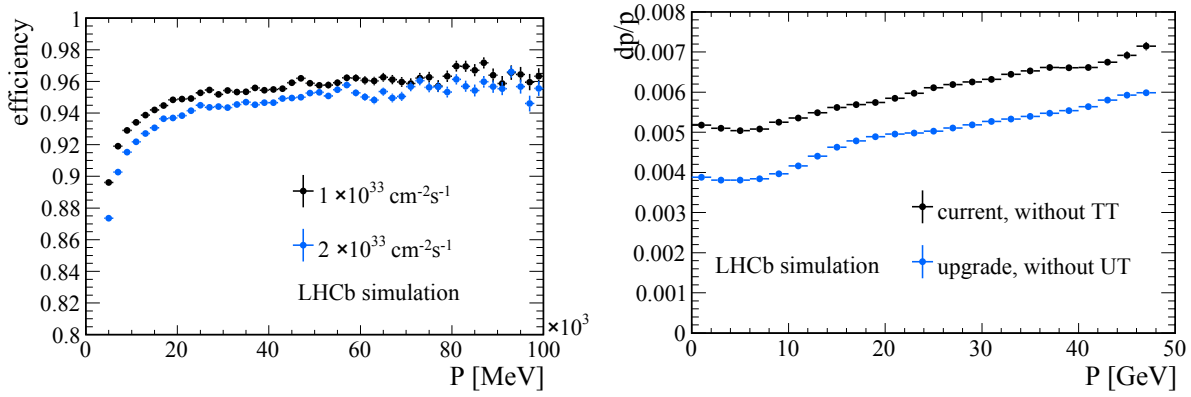
(a) (b)

Figure 5.9: Two performance indicators for the enhanced VELO detector (red) are assessed in comparison to Run 2 (black) under identical conditions. The left plot illustrates the tracking reconstruction efficiency concerning pseudorapidity for particles registering a single hit in at least three VELO modules. Meanwhile, the right figure portrays the resolution on the impact parameter x-coordinate for reconstructed VELO segments within $2 < \eta < 5$, showcased as a function of the inverse of the transverse momentum. [167].

5.2.4 Expected tracking performance

The performance of the enhanced VELO is studied through simulations and compared with the Run 2 detector performance under identical luminosity conditions [167]. Figure 5.9 illustrates two examples: the VELO tracking efficiency plotted against pseudorapidity (left) and the resolution on the x coordinate for the IP measurement as a function of the inverse of the transverse momentum (right). In both scenarios, a significant improvement is achieved.

The reconstruction of long tracks, crucial for physics analysis, primarily relies on the Forward tracking algorithm. The efficiency to reconstruct such tracks exhibits a $2-4\%$ decrease for the upgraded detector in events generated under upgrade conditions, compared to the efficiency observed with the current detector under 2011 conditions [169]. This difference is more pronounced for low-momentum tracks. However, the forward tracking algorithm ensures a considerable improvement for the upgraded detector in events simulated with the same running conditions, outperforming the Run 2 detector. These improvements are expected to restore the Run 2 efficiency and ghost rate in the upgrade environment. The momentum resolution of the upgraded tracking system is approximately $10-20\%$ better to that of the previous system, as shown in Figure 5.10. This improvement is attributed to reduced material budget and, consequently, decreased multiple scattering in the VELO, UT, and T-stations. The addition of the UT facilitates measurements for particles that undergo decay after passing through the VELO, such as long-lived K_S^0 and Λ particles. The expected performance in ghost rate reduction as a function



(a) (b)

Figure 5.10: (a) Forward tracking efficiency for long tracks with $p > 5$ GeV/c in bins of momentum and number of primary vertices for samples of simulated $B_s \to \phi\phi$ events. (b) Momentum resolution of long tracks fitted with the Kalman fit [169].

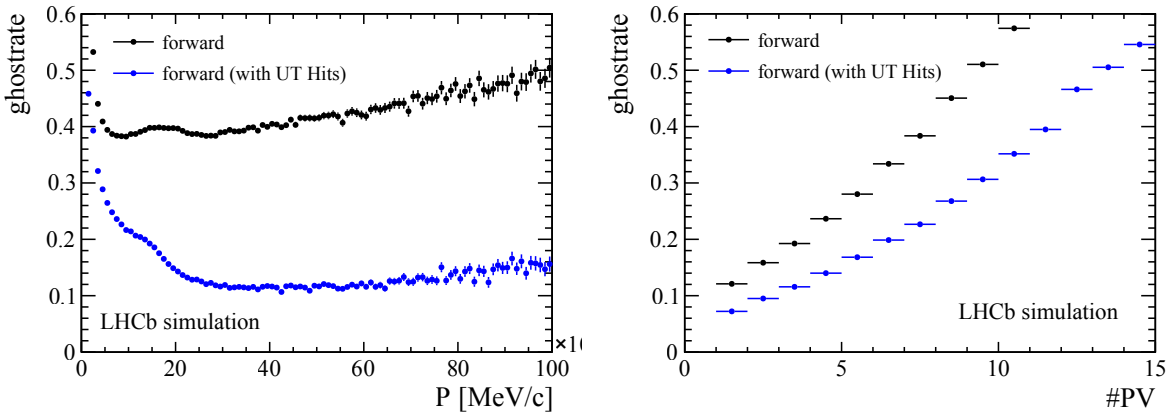


Figure 5.11: Ghost rate of long tracks reconstructed by the Forward algorithm with and without the requirement of at least three UT hits as a function of momentum and number of primary vertices for a sample of simulated $B_s \to \phi\phi$ events[169].

of momentum and number of primary vertices is shown in Figure 5.11.

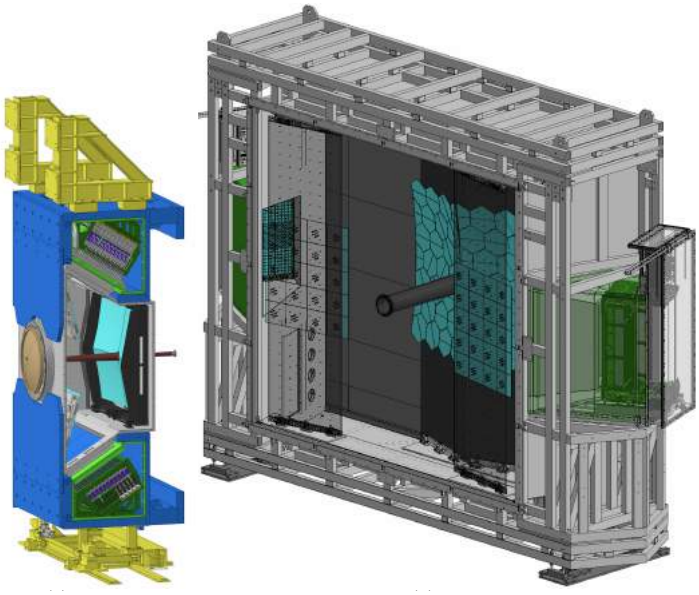


Figure 5.12: Schematic view of the upgraded RICH 1 (left) and RICH 2 (right) detectors [166].

5.3 The particle identification system upgrade

5.3.1 RICH

The overall layout and concept of the RICH system have been preserved from Run 1-2 LHCb. However, essential modifications were necessary to enable the system to operate at the higher design luminosity while maintaining performance levels comparable to Run 1 and Run 2 [113, 109]. To accommodate a 40 MHz readout rate, a complete overhaul of the photon detection chain was implemented in both RICH1 and RICH2 detectors. This was necessary because the previous hybrid photon detector (HPD) [112] featured embedded front-end electronics limited to a 1 MHz output rate. The layout of the upgraded RICH system is presented in Figure 5.12.

RICH 1 optical and mechanical systems

With a five-fold increase in instantaneous luminosity, a redesign of RICH 1 optics became necessary to reduce peak occupancy, defined as the fraction of detected photons per photon detector over the total number of channels [172]. This was done to spread the photon distribution over a wider area. Based on experience from Run 1 and Run 2 operations, optimal performance of the pattern recognition algorithm is achieved when the occupancy does not exceed 30%. To meet this criterion, the radius of the spherical mirrors was increased to 3.7 m, compared to the previous value of 2.7 m. As shown in Figure 5.13, the focal plane has been moved backward, resulting in an enlargement of the Cherenkov ring size. In this configuration, the photodetectors are placed deeper inside the magnetic shielding.

The upgraded photon detection chain

The upgraded photon detection chain has been optimised to handle the highly non-uniform

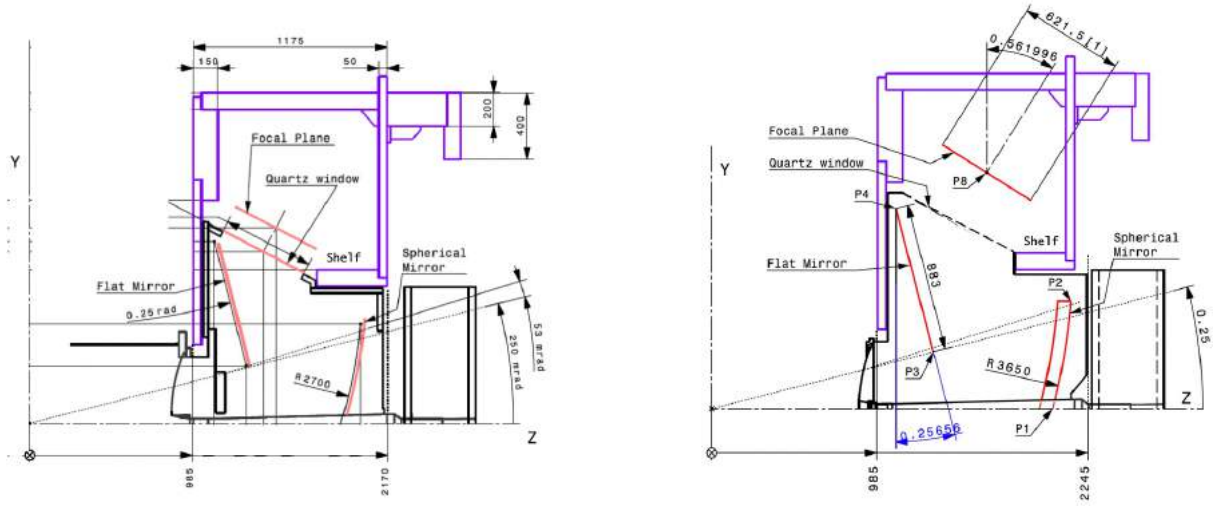


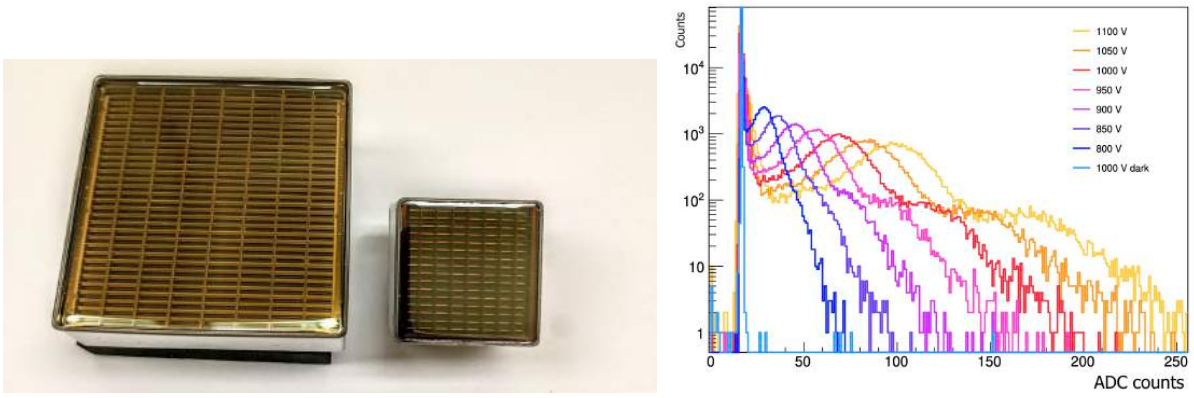
Figure 5.13: The optical geometries of (left) the original and (right) the upgraded RICH1 [166].

occupancy expected in the RICH system. This occupancy ranges from about 30% in the central region of RICH1 down to 5% in the peripheral region of RICH2, assuming the previous photon detectors remain unchanged. The highest hit rates correspond to Cherenkov photons associated with the large number of tracks produced at high pseudorapidity. Considering the observed occupancy distribution, the detector geometry and channel granularity have been optimised, taking into account existing mechanical constraints and the number of readout channels, which significantly impact the costs.

The former hybrid photon detectors (HPDs) have been replaced with multi-anode photomultiplier tubes (MaPMTs) equipped with new front-end (FE) electronics. The selected MaPMT models consist of an 8×8 anode matrix. RICH1 and the central region of RICH2 are outfitted with 1-inch MaPMT modules with a pixel size of 2.88×2.88 mm^2, ideal for the high-occupancy areas of the RICH system. In the outer region of RICH2, a 2-inch device with a pixel size of 6×6 mm^2 has been employed. The decision to use detectors with a coarser granularity in the peripheral regions of RICH2 led to a significant reduction in the number of MaPMT units and readout channels, with a negligible impact on the overall RICH performance. The MaPMTs installed in the upgraded RICH detectors, along with a typical signal amplitude spectra for a pixel as a function of the HV value, are shown in Figure 5.14.

The front-end (FE) boards are equipped with custom read-out ASICs known as CLARO, as displayed in Figure 5.15. CLARO is an 8-channel chip featuring an analog pulse shaping amplifier and a binary discriminator designed for single-photon detection. The signal undergoes amplification and digitisation, with the flexibility to set the threshold individually for each channel [173]. Subsequently, the CLARO output is directed to an FPGA, which synchronises the data with the LHC clock and transmits the events via a GigaBit Transceiver (GBT) link to the High Level Trigger (HLT) stage [174].

The photon detectors and CLARO chips are structured into compact units known as elementary cells (ECs), with two variations designed for the different MaPMT models: the R-type ele-



(a) (b)

Figure 5.14: (a) The MaPMTs selected for the upgraded RICH detectors with the 2-inch model on the left and the 1-inch model on the right. (b) Typical signal amplitude spectra for a pixel as a function of the HV value [166].

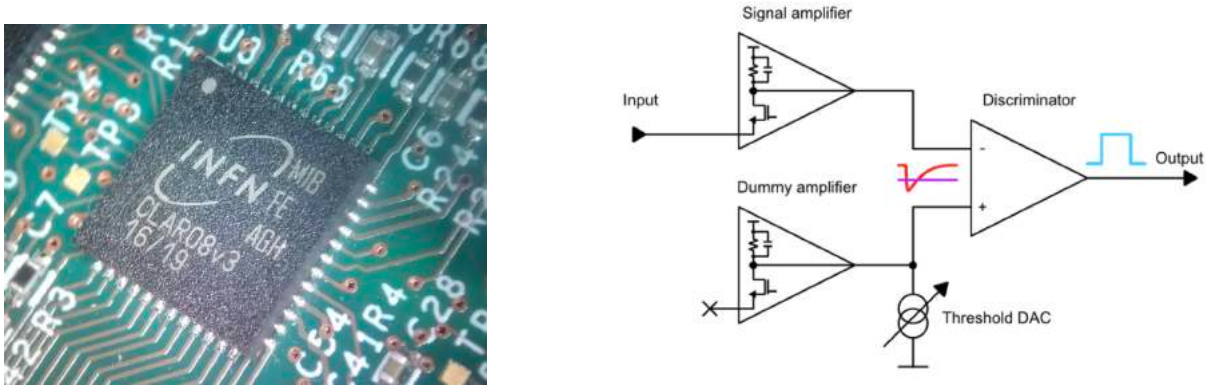
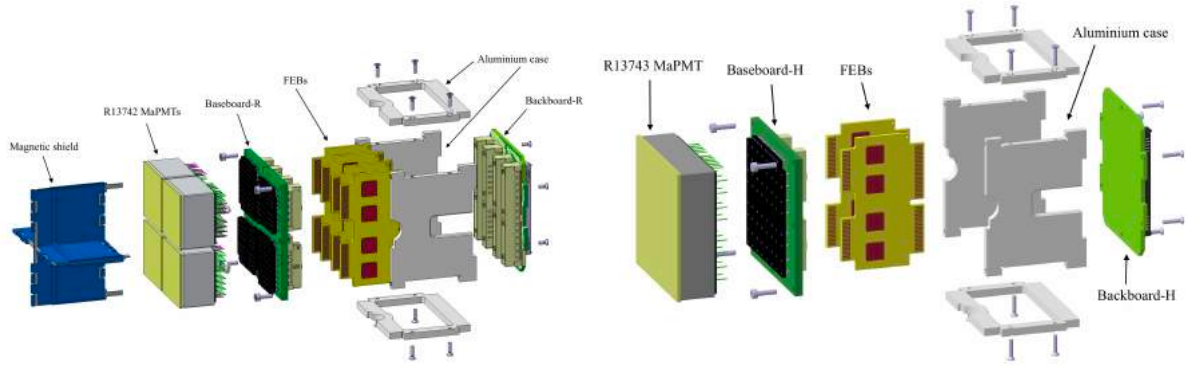


Figure 5.15: (Left) CLARO ASIC with its packaging. (Right) Block schematic of a CLARO channel. The purpose of the dummy amplifier is to give each channel a differential structure, improving the power supply rejection ratio and allowing DC-coupled input to the discriminator [166].



(a) (b)

Figure 5.16: Exploded view of R-type elementary cell (a) and H-type elementary cell (b)[166].

mentary cell (EC-R) and the H-type elementary cell (EC-H). Figure 5.16 illustrates a schematic view of both EC types. The EC-R reads out four 1-inch MaPMTs, covering a total of 256 pixels within an approximately 2×2 square-inch area. The MaPMTs are integrated into a baseboard, which accommodates four 3 MΩ resistive dividers in parallel to bias the dynodes of each MaPMT. In regions with high occupancy, where the drawn current is elevated, dedicated supply lines power the last two dynodes of each chain to mitigate nonlinear effects in MaPMT gain. A magnetic shield is placed in front of the MaPMTs in the RICH1 EC-Rs. Despite being within the magnetic shield, the stray magnetic field from the LHCb magnet reaches up to approximately 2 mT. The shield redirects the field lines, diminishing the magnetic field reaching the MaPMT by a factor of around 20, attenuating its impact on MaPMT performance.

The baseboard transfers the anode signals to four front-end boards (FEBs), each hosting eight CLARO ASICs (four on each face of the board). These FEBs are linked via a backboard that directs the output signals to the photon detector module digital boards (PDMDBs) through two high-density connectors. The CLARO power supply and control signals are generated on the PDMDBs and conveyed through the backboard. A 3.0 mm thick and 40.5 mm long aluminum case serves as a mechanical support structure for the electronic components, facilitating thermal transfer by conduction. Copper layers inside the baseboard enhance heat dissipation from the voltage dividers. Temperature monitoring is ensured by temperature probes. RICH1 incorporates 472 EC-Rs, while 192 are used in the central region of RICH2. The EC-H is responsible for reading out a single 2-inch MaPMT. It consists of a single 2.5 MΩ voltage divider and two FEBs, with half the CLARO channels disabled. The peripheral region of RICH2 contains a total of 384 EC-Hs.

Four ECs, accompanied by two Digital Boards, collectively form the Photo Detector Modules (PDM), serving as the essential components of the RICH columns. Each PDM interfaces with the new LHCb readout through the Photo Detector Module Digital Board (PDMDB), shown in Figure 5.17. The PDMDB plays a crucial role in conveying the digitised photon detector signals away from the high-radiation zone of the detector, ensuring seamless integration with the LHCb Experiment Control System (ECS) and avoiding dead time. Employing an FPGA-based approach, the PDMDB offers flexibility in capturing and formatting data while facilitating the interface

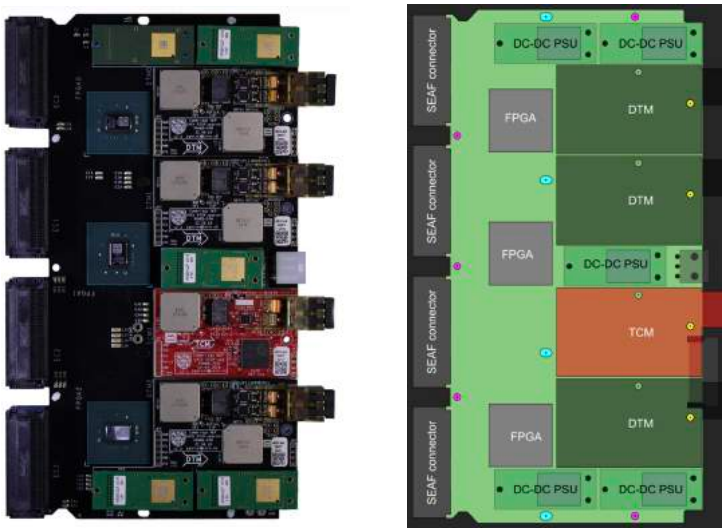


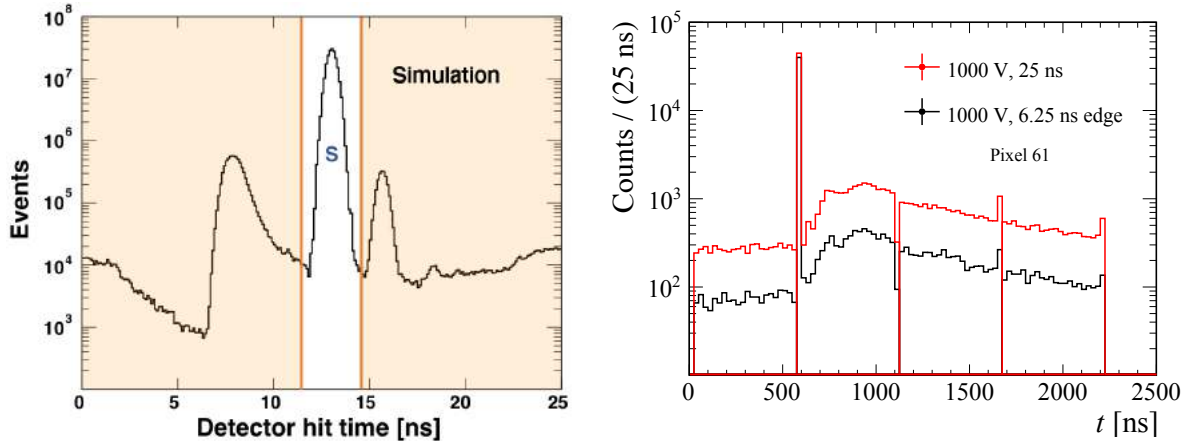
Figure 5.17: (Left) Picture of a fully populated PDMDB-R board. (Right) Schematic view of the PDMDB-R board main components. The PDMDB-H differs by having one less FPGA and DTM with respect to the PDMDB-R [166].

between the various electrical signaling standards of the front-end ASICs and the GBT chipset. Two variants of the PDMDB are utilised, corresponding to the distinct granularity of the photon detector planes. A pair of back-to-back PDMDB-Rs is coupled with a group of four EC-Rs, while a single PDMDB-H is linked to a group of four EC-Hs. The assembly of four ECs and one or two PDMDBs collectively constitutes a photon detector module (PDM). Each PDMDB accommodates one timing and control module (TCM) and supports up to three data transmission modules (DTMs).

Timing for background rejection

The prompt Cherenkov radiation and focusing mirror optics lead to the nearly simultaneous time-of-arrival to the detection plane of photons from a track in the RICH detector. This unique feature allows the application of a time gate at the FE electronics to filter out out-of-time background hits from the output data while admitting photon signals within a narrow time interval. In Figure 5.18, the distribution of photon hit times in RICH1 from a simulation is presented, with the signal peak spanning approximately 2 ns due to the spread of primary interactions in LHCb, defining the minimal width for the FE time gate. In practical terms, considering CLARO time walk, channel-to-channel variations, MaPMT transit time spread, and the digital sampling rate at the FE electronics, a time gate with a width set to 3.125 ns is required, which can be doubled to 6.250 ns if necessary. Beyond excluding background from beam interactions, the time gate proves effective in eliminating sensor noise, including MaPMT cross-talk and afterpulses [175], as shown in Figure 5.18. The achieved reduction in background significantly enhances the performance of particle identification (PID) through the RICH pattern recognition algorithms.

The generation of gates and the time alignment procedure are embedded in the Photo Detector Module Digital Board (PDMDB) FPGA firmware. The time-gating logic utilises the dese-



(a) (b)

Figure 5.18: (a) RICH1 simulated photon detector hit time distribution showing the signal peak (S) and a possible time gate in the front-end electronics [166]. (b) Photon hits accumulated on one MaPMT pixel during tests with a particle beam. When a 6.25 ns gate is applied the continuous background is reduced by a factor of three to four while the signal peak remains unaffected [174].

rialiser present in every input-output logic block of the FPGA, operating at gigabit rates [174]. The deserialiser samples CLARO signals using both edges of the 160 MHz clock, shifting the sampled data at 320 Mbit/s into an 8-bit shift register, as shown in Figure 5.19. This byte is cross-referenced against specific signal patterns using a lookup table, a memory resource readily available with a small logic footprint in the general-purpose logic of the FPGA. Upon a match with one of the configured lookup table patterns, a hit is registered on the 40 MHz system clock edge. The programmable lookup table allows flexibility for different data-taking modes, such as varying time gate widths, edge detection, and basic spillover checks. The time gate is implemented at a fixed latency with respect to the LHCb clock. The FPGA receives the 40 MHz system clock and 160 MHz sampling clock from the GBT, with clock phases adjustable over the 25 ns range in fine steps of 49 ps. This adjustment capability enables fine-tuning of the position of the time gate concerning the signal time-of-arrival in the RICH detector.

5.3.2 ECAL and HCAL

In response to the updated LHCb readout scheme, both the front-end (FE) and readout electronics for the electromagnetic and hadronic calorimeters have undergone a comprehensive redesign and replacement. Additionally, two subdetectors from the previous calorimeter system, namely the SPD and the PS, have been removed due to their diminished roles in the new LHCb full-software trigger. The overall layout of the electromagnetic calorimeter (ECAL) and hadronic calorimeter (HCAL) remains consistent with the upgrade. To minimise necessary modifications, the ECAL and HCAL calorimeter modules, along with their photomultiplier tubes, and coaxial cables, have been retained without alterations. In order to maintain a consistent average anode current at the higher luminosity, the photomultiplier high voltage has been reduced resulting in

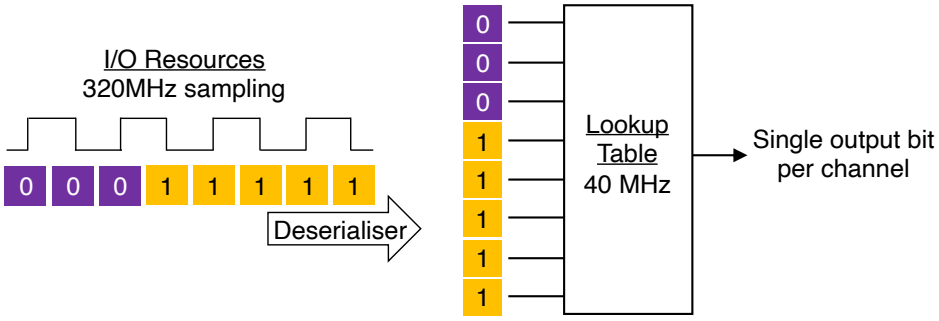


Figure 5.19: Diagram illustrating the timing gate implementation. The CLARO ASIC's output (not depicted) is sampled at 320 MHz, capturing the signal's arrival within a 3.125 ns time window and generating eight samples every 25 ns. The 8-bit byte is cross-referenced against a configurable lookup table for validity before being outputted by the FPGA [174].

an increased gain for the amplifier-integrator in the FE cards.

The FE electronics boards have been entirely redesigned to align with the 40MHz readout frequency. The format have been selected to ensure compatibility with the existing crates and racks. This decision to retain the calorimeter modules, assumes their capability to operate within the expected radiation levels corresponding to the foreseen integrated luminosity.

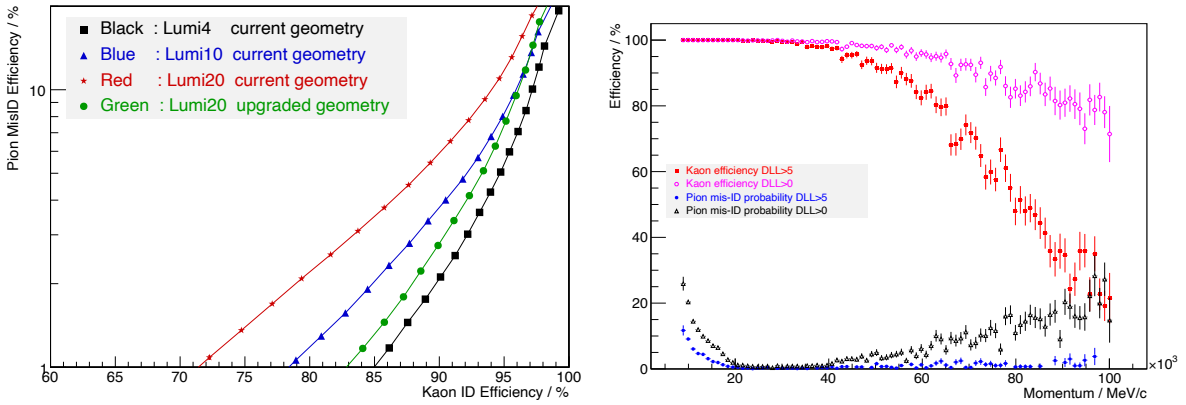
5.3.3 MUON

The muon system in LHCb Run 1-2 featured an additional station, M1, positioned upstream of the calorimeters. M1 played a role in the hardware L0 trigger, but its relevance diminishes in the upgraded setup. The present LHCb muon system comprises four stations, namely M2 to M5, equipped with MWPCs and interleaved with 80 cm thick iron absorbers to filter low-energy particles. The front-end (FE) electronics house an amplifier-shaper-discriminator stage implemented in a dedicated ASIC, along with a digital section enabling time alignment of signals and logical combinations of readout channels into "logical channels."

The FE electronics, designed to withstand radiation up to 100 kGy, remains unchanged as it is expected to be suitable for the new running conditions. However, a comprehensive overhaul of the readout electronics has been conducted to align with the updated LHCb readout scheme, representing the primary upgrade for the muon system. The monitoring and control electronics have also undergone a complete redesign to align with the new 40 MHz readout rate and the updated DAQ and control systems of the experiment. Despite these significant changes, the new electronics have been designed to maintain backward compatibility with the original architecture, minimising costs and allowing the reuse of original crates, cabling, and power supplies.

5.3.4 Expected particle identification performance

The simulations for the RICH upgrade [176] utilise the standard LHCb software framework. PID performance is characterised by the efficiency in correctly identifying a true kaon as a kaon and the misidentification probability of a true pion as a kaon or a heavier particle. These metrics



(a) (b)

Figure 5.20: (a) The PID performance of the current and the upgraded geometrical layouts for different luminosity conditions: "Lumi4" refers to a luminosity of $L = 3.9 \times 10^{32}$ cm^{-2} s^{-1} (the nominal 2012 running condition), "Lumi10" refers to $L = 10^{33}$ cm^{-2} s^{-1} and "Lumi20" refers to $L = 2 \times 10^{33}$ cm^{-2} s^{-1}. (b) The kaon identification efficiency (magenta and red) and pion misidentification probability (grey and blue) as a function of track momentum for the upgraded geometry at Lumi20 (with DLL cuts of 0 and 5, respectively) [176].

are plotted against each other for various cut values on the delta-log-likelihood (DLL) between kaon and pion hypotheses, which is derived from the PID algorithm. The results are compared between the current and upgraded RICH geometries for different luminosities. These comparisons are obtained by executing the full simulation and reconstruction chain in $B_s^0 \to \phi\phi$ events, as explained earlier. In Figure 5.20, the PID performance is illustrated for the current geometry at three luminosities and compared to the upgrade geometry. With the current geometry, there is a gradual decline in PID performance as luminosity increases from the present to the upgraded running conditions, as expected. The performance is regained when adopting the new RICH 1 geometry, indicating an overall improvement compared to the previous geometry. Figure 5.20 shows the kaon identification efficiency and pion misidentification probability as functions of track momentum for the upgraded geometry. As the DLL cut increases from 0 to 5, both the kaon identification efficiency and the pion misidentification probability decrease, demonstrating the excellent performance of the upgraded PID system.

The discriminatory power of the ECAL variables in distinguishing electrons from other charged particles is illustrated in Figure 5.21. The distribution shows the ratio between ECAL energy and track momentum, where EcalE is the sum of energies of the ECAL cells intersecting the track extrapolation and those compatible with potential bremsstrahlung emissions. The determination of these bremsstrahlung emissions involves projecting the track direction before bending in the magnetic field to the energy deposited in the ECAL. These plots are generated using $B^0 \to K^{*0} e^+ e^-$ simulation samples [176].

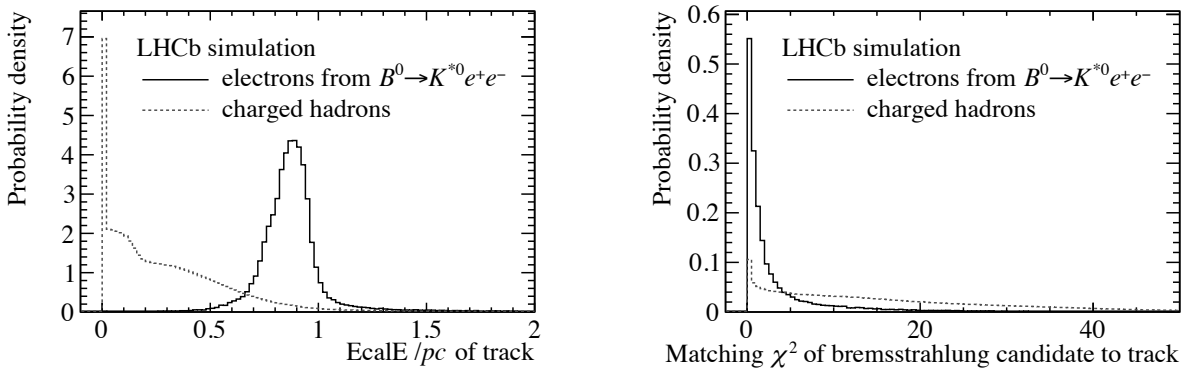


Figure 5.21: Main electron PID variables for the ECAL: distributions for signal and background separately for the variables (left) EcalE/p and (right) matching χ^2 of a bremsstrahlung cluster candidate to a track. The distributions of the bremsstrahlung matching χ^2 are conditional on having a cluster candidate in a 3×3 cell grid around the bremsstrahlung track extrapolation [176].

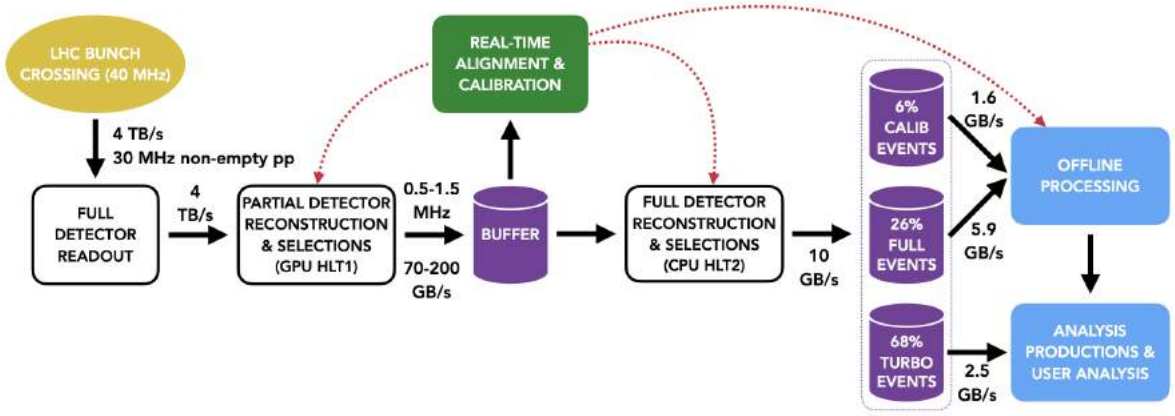


Figure 5.22: LHCb upgrade dataflow, from the LHC bunch crossing of 40 MHz down to the offline processing and analysis production [177].

5.4 The Run 3 dataflow

The removal of the L0 trigger requires a significant transformation in the data acquisition process at LHCb. The whole detector needs to be read out at the 40 MHz LHC rate, and event selection is carried out through a fully software-based trigger. The upgraded dataflow scheme is displayed in Figure 5.22. On average, only 30 MHz of the input rate corresponds to the non-empty proton-proton collisions that are relevant when making trigger decisions at LHCb. The 4 TB/s of input data is managed by the event builder (EB) CPU farm, which consolidates data from each subdetector and transfers it to the high-level trigger.

5.4.1 The online system

The LHCb data acquisition system, illustrated in Figure 5.23, comprises a farm of event builder (EB) servers housing back-end receiver boards (TELL40 boards) and graphics processing units (GPUs) responsible for running the HLT1 application. The data processed by the EB and HLT1 is subsequently transmitted to HLT2 for additional processing and final storage. This system features a farm of 162 EB servers hosting back-end receiver FPGA boards (TELL40) and GPUs executing the first high-level trigger stage. Each TELL40 board receives data from the various front-end electronics of the subdetectors, forming Multi-Fragment Packets (MFP). The primary objective of the EB servers is to gather the MFPs from all subdetectors, consolidating information from the same packet of events into Multi-Event Packets (MEP) containing 1000 events. The MEPs are then internally transferred within the EB server to the GPUs, where an initial event reconstruction and selection process takes place. This internal connection eliminates the need for further transfer to an external server, minimising overhead costs and significantly accelerating data processing.

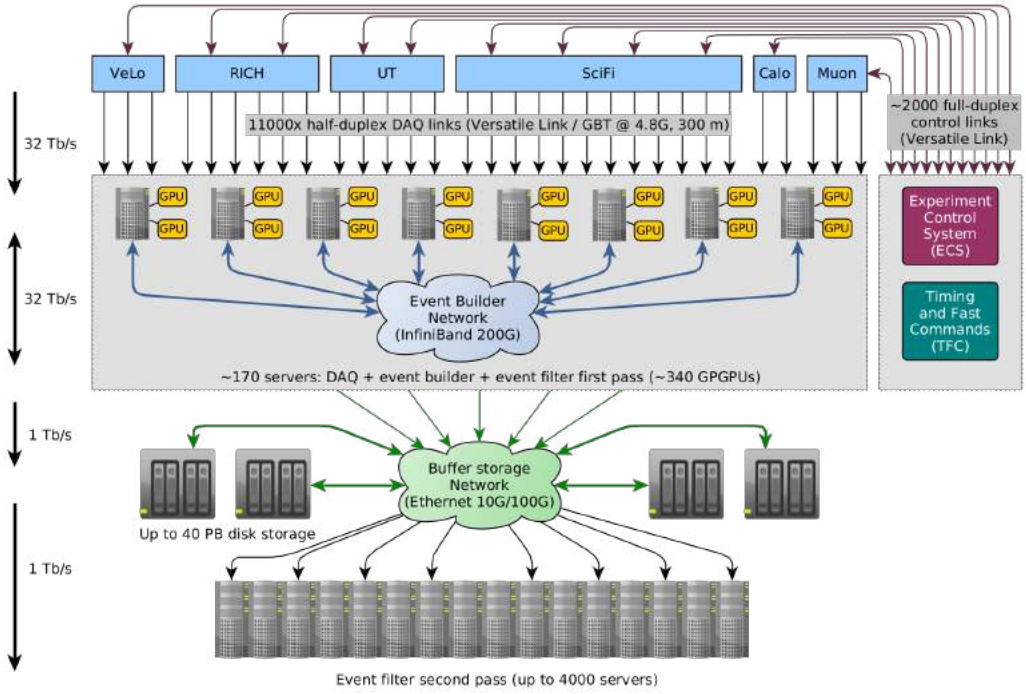


Figure 5.23: Upgraded LHCb online system. All system components are connected to the ECS shown on the right, although these connections are not shown in the figure for clarity [166].

5.4.2 The trigger system

The HLT1 trigger is entirely software-based and optimised for GPU. HLT1 conducts a partial event reconstruction to select interesting events and reduce the data rate by a factor of 30, resulting in a 1 MHz output rate. The selected events are temporarily stored in a buffer while a real-time alignment and calibration of the detector take place. Subsequently, the data is handed over to the second level trigger stage, HLT2. Here, a full event reconstruction occurs, leveraging the most accurate real-time detector parameters. HLT2 selects events using information from the entire LHCb detector, and its optimised reconstruction algorithms run on a CPU farm. Finally, the events selected by HLT2 are stored to tape and undergo further offline processing by the LHCb software to generate the final data samples for analysis.

HLT1

The primary objective of the HLT1 trigger is to reduce the rate of non-empty pp collisions from the LHC's 30 MHz to approximately 1 MHz. HLT1 achieves this by utilising information from the tracking detectors, electromagnetic calorimeter, and muon chambers to conduct a partial event reconstruction. This partial reconstruction imposes minimum momentum requirements in the tracking algorithms, aligning with the stringent throughput constraints of HLT1. Notably, HLT1 does not employ information from the RICH for particle identification, which is utilised at the HLT2 level. The reconstruction process involves finding primary and secondary vertices, tracking, and particle identification algorithms utilising data from the ECAL and muon stations.

A specialized sequence of algorithms optimised for GPUs was developed to handle event reconstruction in a higher multiplicity environment. The HLT1 reconstruction sequence encompasses real-time partial event reconstruction and selection, executed on the input data using the Allen framework [178].

GPUs are particularly well-suited for HLT1 purposes, as each physics event is independent and can be parallelised within the GPU threads, maintaining high throughput performance. The Allen algorithms leverage the parallelisation power of GPUs even when reconstructing individual tracks within an event. GPUs, characterised by a high number of floating-point operations per second (TFLOPS), align well with the substantial computational load of HLT1. They can efficiently process thousands of events, corresponding to approximately 100 kB of raw data each, fitting within the GPU's available memory of around 10 GB. Additionally, GPUs exhibit lower latency than CPUs, which proves beneficial during full detector readouts. The choice of GPU technology offers considerable advantages for future data-taking and upgrades. Each EB server has three slots for placing graphic cards, and a second GPU per server has been added for the 2023 data-taking period. In the future, a third GPU per server could be incorporated to enhance HLT1 performance, potentially extending the physics reach of HLT1 selections.

HLT2

After undergoing selection by the HLT1 trigger, events are directed to an intermediate buffer system before proceeding to the second stage of the high-level trigger, HLT2. The 40 PB buffer ensures sufficient storage time for the calculation of calibration and alignment constants, crucial for optimal event selection. Events are sampled from stored data to assess the required constants within minutes, facilitating the prompt update of HLT1 parameters, which are propagated to HLT2, accordingly. HLT2 executes a full event reconstruction, reconstructing all tracks and incorporating information from the RICH detectors. In addition, both inclusive and exclusive selections are performed at HLT2 level, reducing the event rate to around 100 kHz.

With over a thousand HLT2 lines, tailored for specific signal topologies and/or physics analyses, interesting events are selected using reconstruction information. Unlike HLT1, where a limited number of inclusive lines are sufficient to select numerous interesting events, HLT2 employs both inclusive and exclusive selections, exploiting the increased available bandwidth. The Turbo method [179] is employed to increase the number of recorded events, allowing flexibility in the amount of event information stored. In the Turbo persistence model, once a candidate decay is selected by the HLT2, as a bare minimum only the objects involved in the trigger decision, plus all the primary vertices are persisted in the Turbo stream. The event selection can be further customised (selective persistence) to save additional objects, like e.g. other tracks coming from a primary vertex or objects contained in a cone around the candidate, up to the full event, including possibly some raw data banks. A special use case of the last option is represented by the calibration TurCal stream where events selected for detector alignment and calibration are persisted. Depending on the specific physics channel being studied, event information is stored, ranging from the minimum of two tracks and their vertex to the complete event.

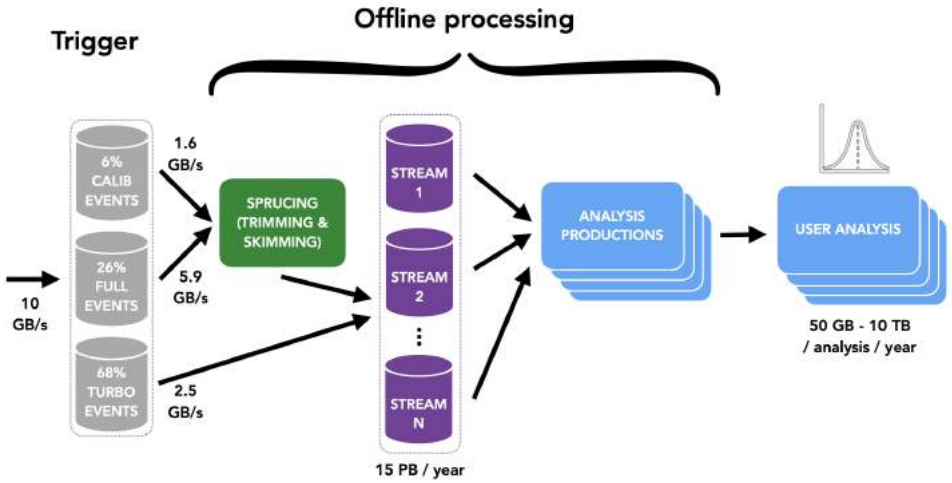


Figure 5.24: Scheme of the offline data flow. Events selected by HLT2 are processed through the sprucing model and the analysis production framework prepares the output files ready to be analysed [177].

5.4.3 The offline software framework

The offline data processing involves a "skimming" and "slimming" of the data, referred to as "sprucing". Sprucing can further apply selection criteria to events passed in the FULL stream (skimming), which constitutes approximately 30% of the total output rate from HLT2. The remaining 70% of the data undergoes processing through the Turbo mechanism, which reduces the data size by persisting only information related to particle signal candidates (slimming). Data saved in the FULL stream can be reprocessed in yearly re-sprucing campaigns, as the full raw data is stored to tape and accessible. This reprocessing capability is not available for data in the Turbo stream.

Typical examples of events saved in the FULL stream include those selected by inclusive topological lines, targeting, for instance, two or three-body decays with specific requirements on vertex quality and particle momenta. In contrast, events selected by exclusive lines, which request specific decay topologies and final-state particles, are usually processed in the Turbo model. Both FULL and Turbo data are then distributed into optimised physics streams, allowing analysts to access reduced data sets categorised by physics topics. The data undergoes further processing by the Analysis Productions framework, which generates the final output files ready for analysis. Analysis Productions is a centralised and automated framework responsible for handling the monitoring and archiving of the final output data files. The offline processing scheme is illustrated in Figure 5.24 and also includes calibration events processed by the sprucing model.

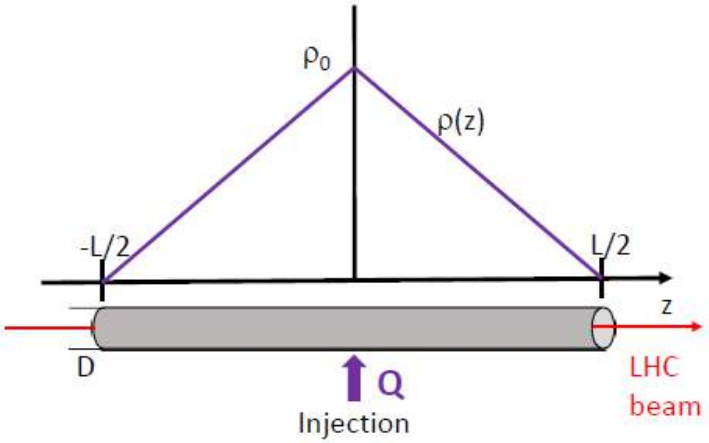


Figure 5.25: Sketch of the injected gas density as a function of z inside the SMOG2 cell [181].

5.5 The SMOG 2 project

Fixed-target physics with LHC beams and an internal gas target marked a significant achievement in the LHCb experiment during Run 2. The SMOG system, initially designed for precise colliding-beams luminosity calibrations, was utilised to inject light noble gas into the VELO vacuum vessel. This process created a temporary local pressure increase, peaking at approximately 10^{-7} mbar along the vessel's length (around 1 m) and gradually returning to the LHC background level ($\sim 10^{-9}$ mbar) over the 20 m LHCb beam pipe sections on each side of the interaction point. SMOG not only enabled accurate luminosity calibration but also provided a unique opportunity to operate the LHCb experiment in fixed-target mode. Gaseous targets with different nuclear sizes (He, Ne, and Ar) were utilised in combination with proton and lead beams, achieving centre-of-mass energies per nucleon of up to 115 GeV, with negligible impact on LHC operation. Encouraged by initial results and future prospects, an upgrade to SMOG, known as SMOG2, was proposed and implemented [132].

The fundamental concept of SMOG2 involves injecting gas directly into a storage cell coaxial to the LHC beams. This allows to significantly increase the areal density of the target with respect to SMOG while maintaining the same injected flux [180]. The schematic representation of this principle is illustrated in Figure 5.25. The open-ended cylindrical tube, with an inner diameter D and a length L, receives gas injection through a capillary at the centre of the storage cell, with a flow rate ϕ from a Gas Feed System (GFS). This results in an approximately triangular density distribution $\rho(z)$ with a maximum $\rho_0 = \phi / C_{tot}$ at the centre ($z = 0$). Here, C_{tot} represents the total flow conductance of the tube from the centre outwards and is determined by the conductance of two parallel tubular conductances of length $L/2$ in the molecular flow regime [180]. The areal density experienced by the beam is $\theta = \rho_0 L/2$.

The SMOG upgrade introduces several notable enhancements. Firstly, the determination of the target density (and beam-gas luminosity) becomes significantly more precise due to the confinement of the target within the storage cell. The conductance of the storage cell is well-

known and can be combined with an accurate measurement of the injected gas flow rate from the GFS. Secondly, the upgrade allows for the selection of various gas species without manual intervention, including non-noble gases such as H_2, D_2, O_2, etc. Finally, the beam-gas interaction region is distinctly defined and well-separated from the beam-beam collision region, offering the possibility of simultaneous beam-gas and beam-beam collisions.

SMOG2 comprises two principal systems: the storage cell assembly, situated inside the beam vacuum, and the GFS, located on the "balcony," a platform near the detector within the experimental cavern. Due to its proximity to the LHC beams, the design of the storage cell assembly must meet several requirements derived from aperture considerations, RF or impedance-related aspects, and dynamic vacuum phenomena.

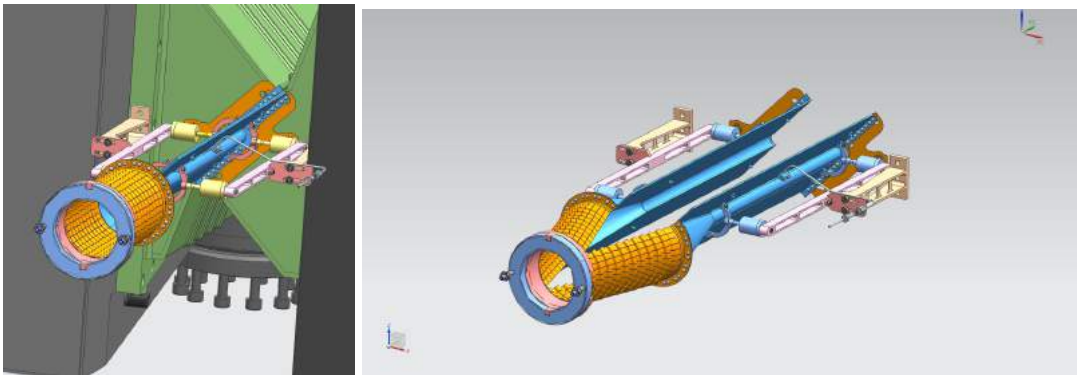
5.5.1 The storage cell

The configuration of the storage cell within the VELO vessel is illustrated in Figure 5.26. This assembly is specifically designed to fit within the limited space available inside the existing VELO vessel, positioned upstream of the VELO detector. To ensure sufficient beam aperture for various beam operations (such as injection, energy ramp, squeeze, etc.), the assembly is divided into two opposing halves, each connected to its respective VELO RF box.

This assembly includes the following components: a flexible wakefield suppressor, divided into two halves; two opposing storage cell shapes featuring a half cone, a half tube, and side wings; a short wakefield suppressor connecting to the VELO detector box; and two arms supporting the storage cell halves from the VELO RF box flanges. The conical shape allows for a smooth transition from the 56 mm diameter of the upstream beam pipe to the 10 mm diameter of the storage cell tube, which is 20 cm long. Gas injection from the Gas Feed System (GFS) occurs into the center of the tube through a flexible line terminated with a 0.8 mm inner diameter stainless steel capillary pressed into a hole in the Side C half of the storage cell. All components are sufficiently light to minimize beam-induced background from materials near the beams to a negligible level.

The synchronisation of the storage cell with the VELO detector boxes is achieved by attaching it with two cantilevers to the flanges of the VELO RF boxes. The VELO design allows for a slight retraction of the detector (approximately 0.1 mm) from the nominal closed position. The Side C half of the storage cell is securely attached to its cantilever, while the Side A half is connected through a spring system. This system ensures the reaching of the final closed position, guided by the rigid half, even if the VELO halves are not entirely closed, within a range of up to 1 mm. The minimum allowed aperture along the length of the storage cell, dictated by the Van Der Meer scan configuration, is 3 mm [182], well below the chosen radius of the storage cell.

Surfaces exposed to the beam are constructed from electrically conductive materials to shield the chamber from beam RF fields, prevent RF mode excitation, and maintain electrical continuity for any position of the VELO halves. The cell structure is crafted from a 99.5% pure aluminum block with a milling accuracy of approximately 20 μm. The storage cell is coated with amorphous carbon to present a surface to the beams with a secondary electron yield (SEY)



(a) (b)

Figure 5.26: (a) View of the storage cell (blue) supported from the VELO RF box flanges (in green) in the closed VELO position. Two flexible wakefield suppressors (orange) provide the electrical continuity. (b) Storage cell in the open position (without showing VELO elements).

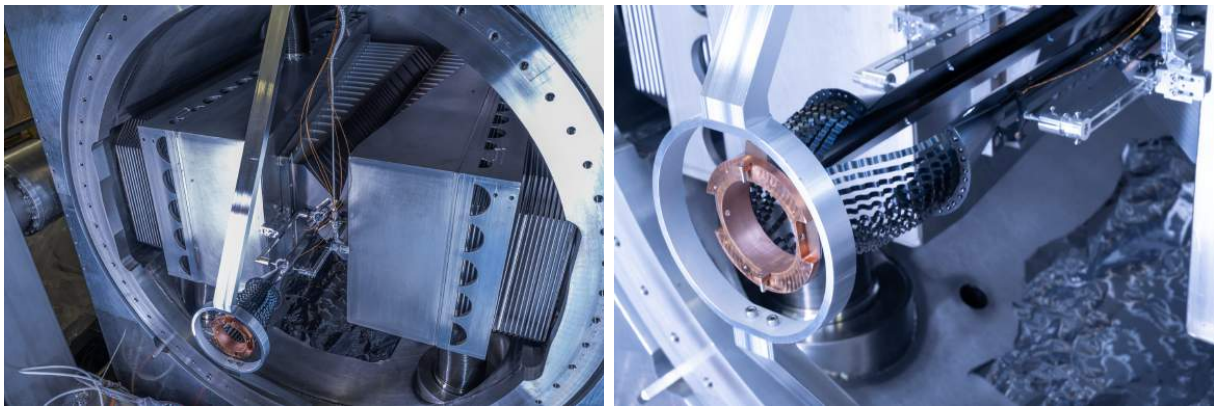


Figure 5.27: Pictures of the installed SMOG2 cell in its closed (left) and open (right) positions [184].

of approximately 1.0 [183]. This precaution prevents the formation of beam-induced electron clouds and potential beam instabilities. The coating, applied through sputtering, includes a 50 nm thick Ti adhesion layer followed by a 1–10 nm thick layer of amorphous carbon. Simulation studies confirm that this SEY is sufficient to prevent electron cloud buildup, even considering the potential impact of higher residual gas pressure.

Equipped with five 0.34 mm outer diameter K-type thermocouples (precision of about 0.1 K), the storage cell is insulated with a nickel-based super alloy and terminated with a ceramic connector suitable for ultrahigh vacuum use. Temperature measurements are essential for determining the areal density θ (temperature dependent) and monitoring potential temperature increases due to beam-induced effects. The installation of the storage cell into the VELO vessel was successfully completed in the summer of 2020, as shown in Figure 5.27. A detailed alignment survey confirmed no misalignment exceeding 0.25 mm.

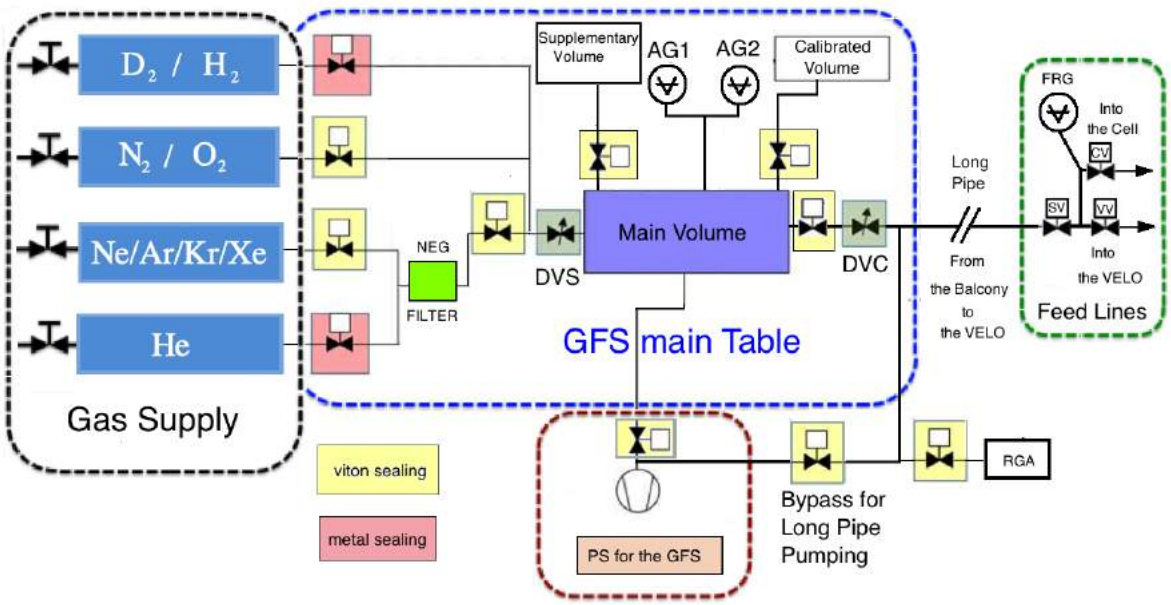


Figure 5.28: Sketch of the SMOG2 gas feed system [132].

5.5.2 The new gas feed system

The SMOG2 GFS offers the flexibility to select the gas type for injection from four reservoirs. The precise control and measurement of the injected gas volume allows for accurate computation of target densities based on the storage cell geometry and temperature. The GFS comprises four assembly groups, illustrated in Figure 5.28, utilising absolute thermo-stabilised gauges covering four decades of pressure readings. These gauges play a crucial role in monitoring and determining the stability of the injected flow. The flow is established by setting a nominal pressure and maintaining constancy through a thermo-regulated valve (DVS). Following the stabilisation of the injection pressure, another thermo-regulated valve (DVC) is adjusted to the appropriate value based on the chosen gas type and flow rate (typical values ranging from 0.5 to 8×10^{-5} mbar l/s). A full-range gauge (FRG) monitors the pressure just upstream of these two valves. Additionally, a residual gas analyser (RGA) is employed to analyse the composition of the injected gas in the main volume.

5.5.3 SMOG2 physics opportunities

During LHC Run 2, the LHCb collaboration demonstrated the ability to utilise the LHC beams and exploit LHCb-specific detector capabilities in a fixed-target configuration. Production measurements were conducted with various collision systems at unprecedented energies for fixed-target experiments, offering new insights into nuclear and cosmic-ray physics. Clean samples of charmed hadrons and light charged particles (positively identified by the PID subdetectors) were obtained, and absolute cross-sections were measured with a relative precision better than 10% [185].

The fixed-target configuration, now equipped with the SMOG2 storage cell, is poised to significantly advance the investigations previously carried out with SMOG. SMOG2 offers a broader selection of usable gas species, an improved control on the target gas pressure, and a significant increase in integrated luminosity for fixed-target samples.

Charm measurements, crucial for exploring cold nuclear matter effects and heavy-ion physics, stand to benefit from the increased statistics and improved luminosity determination achievable with SMOG2. Additionally, the possibility of measuring beauty production in pA collisions can be considered. Furthermore, investigations into QGP-related flow observables and correlations can a broader centrality range in PbA collisions, as explained in more details in Section 6.7.

Studies on cosmic rays will greatly benefit from the ability to inject hydrogen and deuterium, extending the exploration of antiproton production throughout the entire available centre-of-mass range. Measurements with pH will be possible. A deuterium target also enables the testing of isospin symmetry effects. The fixed-target mode presents an opportunity to probe the energy scale most relevant for the production of light anti-nuclei, prompting investigations into the potential detection of light anti-nuclei in LHCb.

In the realm of nucleon structure studies, SMOG2, with its increased statistics, will enable the exploration of quark and gluon PDFs in nucleons and nuclei, particularly at high-x and moderately-high Q^2. Processes like Drell-Yan are well-suited for constraining theoretical models and reducing uncertainties associated with light quark and anti-quark PDFs. Additionally, heavy-flavour production, dominated by gluon-gluon interactions in high-energy hadronic collisions, provides a means to access the poorly known gluon distribution functions.

Chapter 6

Charged-hadron identification performance in Run 3

Charged-hadron identification is crucial in particle physics experiments, tackling challenges such as mitigating combinatorial background in complex final states and discerning identical topologies. Achieving high performance in charged-particle identification becomes a vital component of the LHCb flavour physics program in Run 3. Consequently, the development of calibration tools will play a central role in Run 3, optimising the performance of the upgraded LHCb detector to maximise the physics outcomes.

The author has led the development and implementation of a software framework for assessing particle identification performance in various collision system collected during early Run 3 data-taking periods, considering various detector configurations and kinematic variables. The description of the strategy and a detailed study of the PID calibration samples collected in 2022 and 2023 is presented in Section 6.3. The performance for pp with early Run 3 data collected in 2022 and 2023, is presented in Sections 6.4 and 6.6, respectively. The study includes also the first assessment of PID performance for pA collisions in fixed-target mode and heavy-ion collisions, presented in Sections 6.5 and 6.7, respectively. In preparation for the 2024 data taking, this tool will be integrated into the LHCb software for monitoring high-level performance in real time, as described in Section 6.8. The results have been published in Reference [186].

6.1 Requirements for particle identification

One of the primary goals for charged-hadron identification in a flavour-physics experiment is the mitigation of combinatorial background. Numerous interesting decay modes of b- and c-flavoured hadrons involve complex hadronic multibody final states. In hadron colliders like the LHC, the most prevalent charged particle produced is the pion. The heavy flavour decays of interest typically involve kaons, pions, and protons. Consequently, when reconstructing the invariant mass of the decaying particle, it is crucial to select the charged hadrons of interest to decrease the combinatorial background.

The second relevant application of particle identification information is to differentiate final

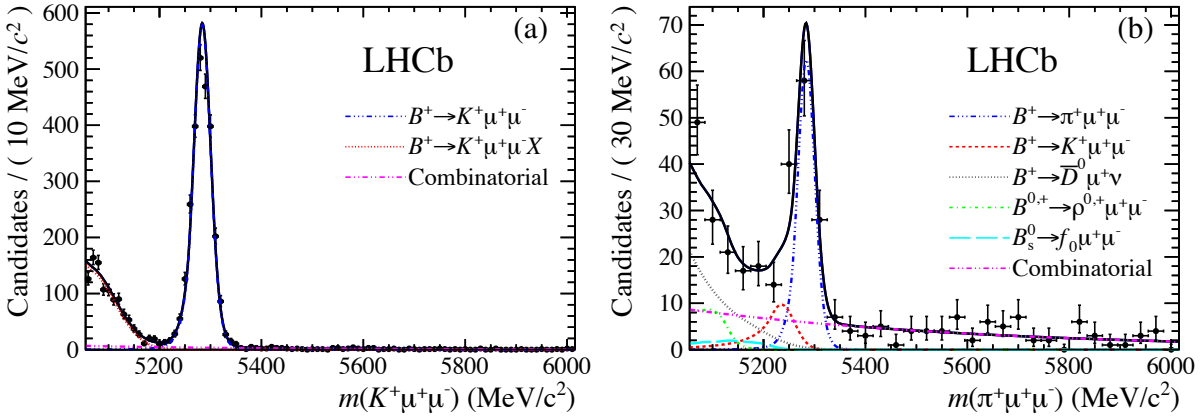


Figure 6.1: The effect of applying hadron PID selections to suppress the dominant kaon mode (a) to measure the CKM suppressed pion mode (b), in rare B decays. Here 82% of pion events are retained, suppressing the kaons by a factor of 80 relative to the pions, leaving a clear peak from the pion mode [189].

states with otherwise identical topology. An illustrative example is the two-body hadronic decays of B mesons, $B \to h^+h^-$, where h denotes a charged hadron [187]. Extracting a signal using only kinematic and vertex-related cuts involves a summation over all decay modes of this type, each generally exhibiting a distinct CP asymmetry. For a precise study of CP-violating effects, it is crucial to distinguish the various components (e.g. $B^0 \to \pi^+\pi^-$, $B_s^0 \to K^+K^-$, etc.). This is accomplished by leveraging the high efficiency of the RICH particle identification.

Another application of charged-hadron identification is for efficient flavour tagging [188]. Heavy-flavoured particles are predominantly generated in pairs. When studying CP asymmetries or neutral meson-antimeson oscillations, a precise knowledge of the production state of the heavy-flavoured particles becomes essential. This knowledge can be attained by tagging the particle/antiparticle state of the accompanying hadron. One of the most powerful methods of tagging the production state involves identifying charged kaons produced in the $b \to c \to s$ cascade decay of the associated particle. Such tagged kaons (as well as kaons from the b fragmentation when a B_s^0 is created) exhibit a soft momentum distribution, with an average around 10 GeV/c. Particle identification down to a few GeV/c can significantly enhance the tagging capability of the experiment.

The benefits of the application of PID on particular physics observables are dependent on the kinematics of the signal channels, and the nature of the background channels. An example of hadron PID used in rare decay analyses is shown in Figure 6.1 [189]. Separate sample of $B^+ \to K^+\mu^+\mu^-$ and $B^+ \to \pi^+\mu^+\mu^-$ decay candidates are obtained, where the pion mode is suppressed by the ratio of the CKM elements $\left|\frac{V_{td}}{V_{ts}}\right|^2$, approximately a factor of 25. The hadronic PID selection suppresses the kaon mode relative to the pion mode by a factor of 80, while retaining 82% of pion candidates. A clean peak from the pion mode is seen, which would be impossible to obtain without effective hadron PID.

The momenta distribution of the decay products in two-body b decays peaks at approx-

imately 50 GeV/c. Ensuring high efficiency for the reconstruction of these decays necessitates particle identification capabilities up to at least 100 GeV/c. The lower momentum limit of around 2 GeV/c arises from the need to identify decay products from high multiplicity B decays and the fact that particles below this momentum threshold will not traverse the dipole magnetic field (4 Tm) of the LHCb spectrometer.

Finally, particle identification has been crucial for the fixed-target and heavy-ions physics program at LHCb. Two physics analyses [141, 142] regarding charm production in fixed-target collisions utilised tight requirements of the kaon and pion particle identification information to reconstruct the $D^0 \to K^+\pi^-$ and $\bar{D}^0 \to K^-\pi^+$ decays. Given the high detector occupancy observed in pPb collisions, the study of prompt D^0 meson production in pPb collisions at $\sqrt{s_{NN}} = 5$ GeV [190] required particle identification criteria tighter than in pp collisions to increase the signal-over-background ratio. In addition, physics analysis regarding prompt and non-prompt antiproton production require high performance in particle identification to be able to efficiently identify antiprotons and distinguish them from the other charged hadrons.

6.2 Performance of the RICH detectors in Run 2

The overall performance of the RICH system can be determined in terms of the ability to distinguish between different species of charged hadrons and the resulting impact of these selections on the physics output of the experiment. The efficiency of the system for discriminating between different species of hadrons can be determined from control samples of well-identified particles obtained purely through kinematic selections. These calibration samples are selected without using information from the RICH detectors in order to not bias the results. High-purity samples of charged kaons and pions are obtained from the decay products of the D^0 meson identified in the decay chain $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ and the charge-conjugated decays. Samples of protons are obtained from $\Lambda^0 \to p\pi^-$ decays and the charge-conjugate decay.

The performance of the RICH detectors was investigated during LHC Run 1 (2011–2012) [113]. Subsequently, in Run 2 (2015–2018), the LHCb experiment collected data with a reduced proton-proton bunch collision spacing of 25 ns, halving that of Run 1. Additionally, the centre-of-mass energy was increased to $\sqrt{s} = 13$ TeV, resulting in a challenging environment characterised by elevated track multiplicities and higher detector occupancy. To improve event retention between HLT2 and offline data, a system for automated real-time alignment and calibration measurements of the LHCb detector was implemented in Run 2. This advancement marked the first-time utilisation of hadron identification in the software trigger. Details regarding the selection of calibration samples are provided in Reference [191].

Trigger selections vary across different data-taking periods, and the kinematics of the D^0 and Λ^0 decay products differ significantly. As RICH performance is a function of the track momentum, and pseudorapidity, as well as the activity in the detectors, a weighting procedure is employed to equalise the distributions of these variables between different species of hadrons and across data-taking periods. The momentum range is chosen such that, under at least one of the particle hypotheses compared, the particle would be above the velocity threshold for

generating Cherenkov radiation. Pseudorapidity ranges are selected to match the RICH detectors' acceptances. The weighting in the number of tracks cancels most of the differences in data-taking conditions and trigger thresholds.

Performance curves for separating kaons and pions, protons and pions, and protons and kaons, split by charge, are depicted in Figure 6.3 for the 2018 data-taking year and integrated over the two magnet polarities. Each data-point corresponds to a ΔLL value, with tighter selections moving towards the bottom-left corner of each plot. The agreement between the performance curves of positive and negative hadrons, particularly interesting for CP violation measurements, is excellent. The same behaviour is observed for all data-taking years. The same performance curves, but split by magnet polarity, are shown in Figure 6.3 for the 2018 data-taking year and integrated over the hadron charges. Again, the agreement is excellent, and the same behaviour is observed for all data-taking years. Curves demonstrating the stability of the PID performances across Run 2 are shown for the years 2015–2018 in Figure 6.2. Within the corresponding kinematics ranges, the probability to correctly identify each charged hadron consistently exceeds 90% for a 5% misidentification probability.

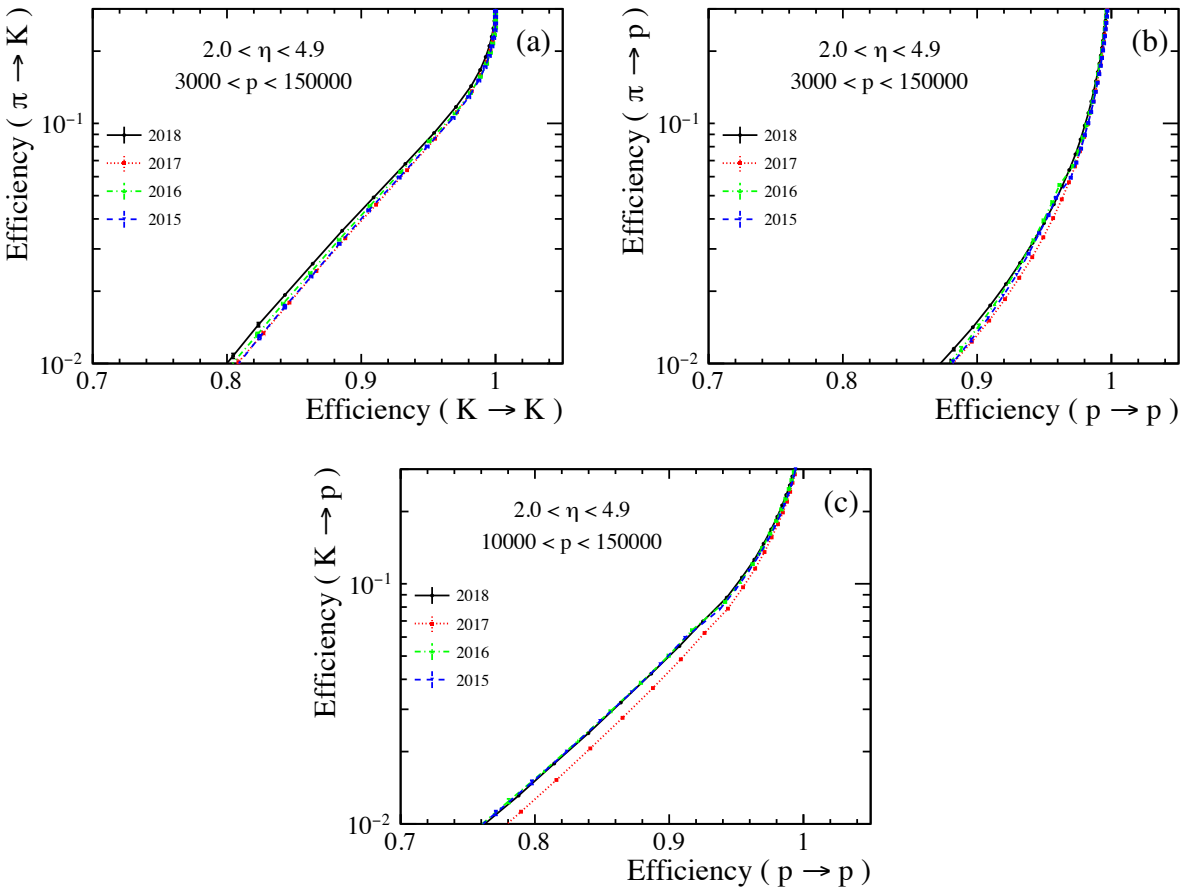


Figure 6.2: The efficiency of selecting kaons (a), protons (b and c), with the associate leakage from misidentifying pions (a and b) and kaons (c). The efficiency curves are shown for 2015 (blue, dashed), 2016 (green, dash-dotted), 2017 (red, dotted), 2018 (black solid). Uncertainties are statistical only, and are highly correlated between points on the same curve [109].

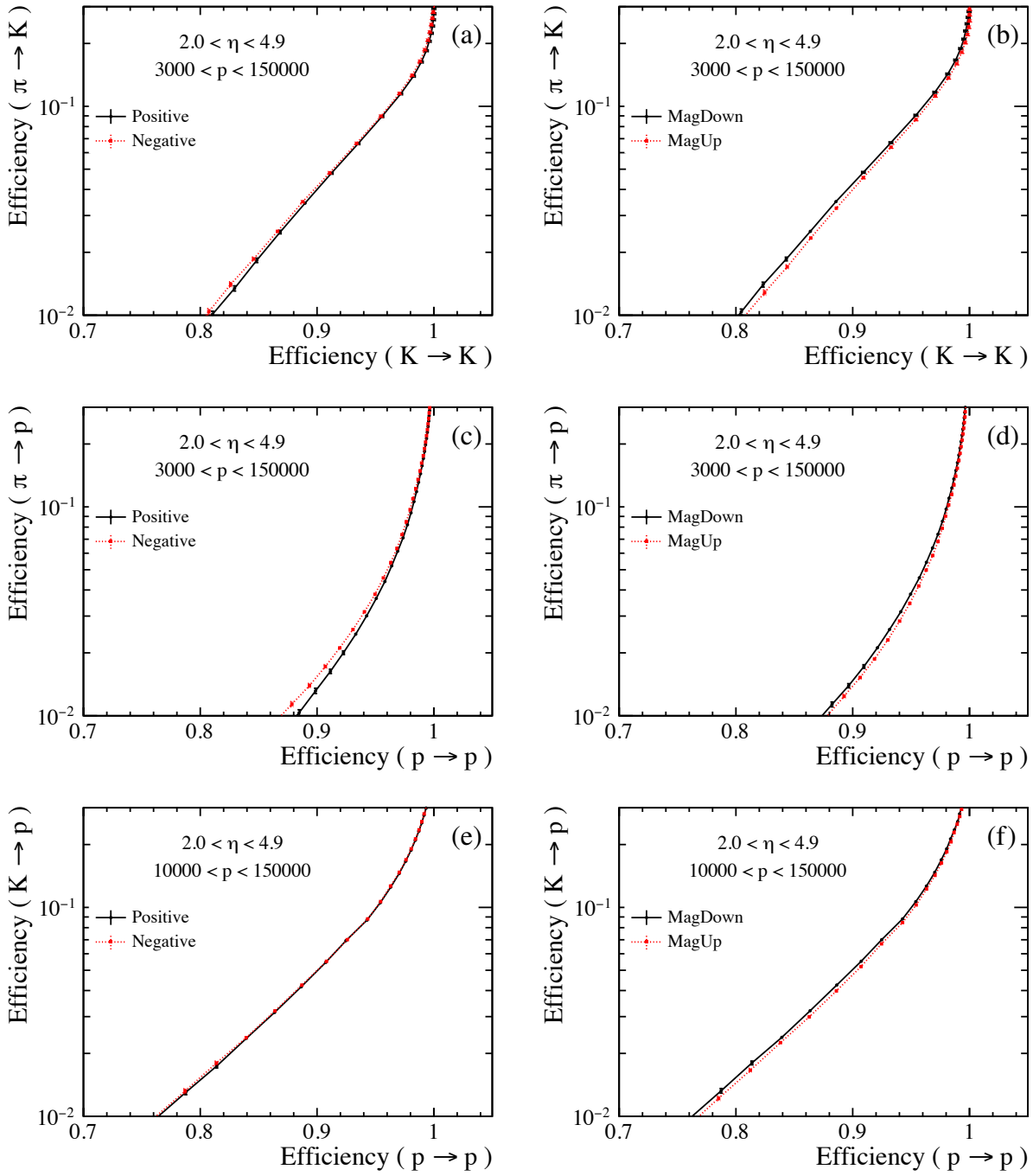


Figure 6.3: The efficiency of selecting kaons (a and b) and protons (c, d, e, and f), along with the associated misidentification from pions (a, b, c, d) and kaons (e and f) in data collected during 2018 is illustrated. For positively (black, solid) and negatively (red, dotted) charged particles, the efficiency curves are shown in panels (a, c, and e). In panels (b, d, and f), the efficiency curves are presented for Up (red, dotted) and Down (black, solid) magnetic field polarities. Statistical uncertainties are indicated and are highly correlated between points on the same curve [109].

6.3 Early performance of the RICH detectors in Run 3

Identification of charged particles is a critical aspect of the LHCb physics program in Run 3. As outlined in Section 5.3.1, the RICH system has undergone a significant upgrade, introducing several modifications aimed at enhancing particle identification capabilities. Notably, to withstand a five-fold increase in instantaneous luminosity compared to Run 1 and 2 the optical system of RICH1 has been redesigned. Additionally, both RICH1 and RICH2 now feature a redesigned opto-electronics chain for detector readout at 40 MHz, enabling single-photon counting capabilities at repetition rates up to 100 MHz/cm^2. This new configuration is anticipated to deliver discrimination between various species of long-lived charged hadrons with momenta ranging from 2.6 GeV/c to over 100 GeV/c, even in events with average multiplicities that are five times larger, thereby restoring the excellent performance achieved in Run 1 and 2.

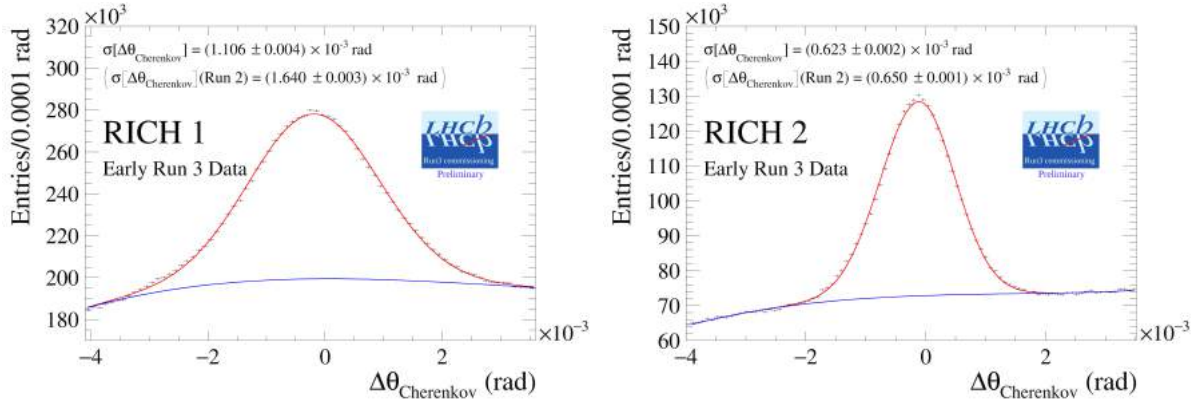
A crucial metric for studying the performance of the upgraded RICH detectors is the Cherenkov angle resolution. The total Cherenkov angle resolution for a track ($\Delta\theta_C$) is expressed as:

$$\Delta\theta_C = \frac{\sigma_\theta}{\sqrt{N_{ph}}} \oplus C_{tracking}, \tag{6.1}$$

where σ_θ is the average angle resolution per single photon in a Cherenkov ring, N_{ph} is the number of detected photons per tracks (photon yield), $C_{tracking}$ is the tracking resolution, and the $\oplus$ symbol denotes the quadratic sum of the two terms. As expressed in Equation 6.1, the particle identification performance is influenced by the photon yield. The average detected photon yield in RICH1 (RICH2) during the period 2016–2018 is reported as 30 ± 2 (18.5 ± 1.2) [109]. Although a precise determination of the photon yield in Run 3 is pending, an improvement is anticipated, primarily due to the enhanced quantum efficiency of the MaPMTs, which is nearly double that of the previously used HPDs. Preliminary estimates of the Cherenkov angle resolution in Run 3 are obtained by utilising data collected at the end of 2022. Figure 6.4 and Table 6.1 present the Cherenkov angle resolutions for reconstructed photons observed by RICH1 and RICH2 in the early stages of Run 3 data [192]. This preliminary outcome suggests improved performance in both RICH detectors compared to Run 2.

To reach the ultimate performance of the upgraded RICH detectors, a thorough assessment of particle identification performance is crucial. Moreover, as PID contains information from tracking and trigger systems, it can be utilised as an additional tool for monitoring detector stability and data quality. There are several points of interest to study:

- compare the performance between Run 3 at nominal pile-up and Run 2, to assess whether the RICH system's design goal has been achieved;
- study the PID as a function of different variables of interest, such as track pseudorapidity, momentum, and detector occupancy, to understand their impact on the detector performance;
- explore the performance under various hardware configurations (HV, threshold, time-of-arrival acceptance window) to determine the optimal working point for the RICH system;



(a) (b)

Figure 6.4: Cherenkov angle resolution for reconstructed photons detected by the (a) RICH 1, (b) RICH 2 detector using early Run 3 data. The red line represents an asymmetric Gaussian fit to the signal and the blue line represents a polynomial fit to the background [192].

	Run 3 $\sigma(\Delta\theta_C)$ (10^{-3} rad)	Run 2 $\sigma(\Delta\theta_C)$ (10^{-3} rad)
RICH 1	1.106 ± 0.004	1.640 ± 0.003
RICH 2	0.623 ± 0.002	0.650 ± 0.001

Table 6.1: Cherenkov angle resolution comparison between Run 3 and Run 2 for RICH 1 and RICH 2.

- conduct the performance assessment dedicated to fixed-target and heavy-ion collisions, and compare the results with those from pp collisions;
- implement online monitoring of the PID performance as an additional high-level metric to monitor the stability of the RICH system and the quality of the data-taking process.

The PID performance has undergone comprehensive examination using early Run 3 data across diverse collision systems at LHCb. The results for pp collisions are detailed in Sections 6.4 and 6.6, corresponding to the years 2022 and 2023, respectively. Section 6.5 presents a comparative analysis of PID performance between pp and pAr collisions. Furthermore, Section 6.7 delves into the early PID performance for PbPb and PbAr collisions, specifically examining its dependence on centrality. In all of these different configurations, the RICH detectors are time-aligned within a window of 6.25 ns around the arrival time of Cherenkov photons.

6.3.1 Efficiency and purity evaluation

The typical RICH performance figure of merit is represented by a misidentification (misID) rate versus an identification (ID) rate of different charged hadrons. The efficiency of a specific PID selection on a probe hadron (pion, kaon, or proton) is evaluated using a binned maximum likelihood fit to the invariant mass distributions of the parent particle. The fitting procedure allows to separate the signal candidates from the remaining background contributions. The

fit is performed with and without the PID selection to the particle of interest. The efficiency and the misidentification are determined by computing the ratio of the yields of candidate decays in the two fits. For example, the efficiency of correctly identifying kaons ($\varepsilon^i_{K\to K}$) and the misidentification of pions as kaons ($\varepsilon^i_{\pi\to K}$) for a specific PID cut (n^i) can be expressed as follows

$$\varepsilon^i_{K\to K} = \frac{N_{\text{fit}}(PID_K(K) > n^i)}{N_{\text{tot}}}, \tag{6.2}$$

$$\varepsilon^i_{\pi\to K} = \frac{N_{\text{fit}}(PID_K(\pi) > n^i)}{N_{\text{tot}}}, \tag{6.3}$$

where N_{fit} and N_{tot} represent the signal yields with and without the PID cut, respectively. The separation between kaons and pions, protons and pions, and protons and kaons is determined using information from the PID_K, PID_P, and PID_{PK} distributions, respectively. The PID selection is systematically varied from loose to tight, resulting in a profile of selection efficiencies versus misidentified particle leakage. In the next sections, this profile will be referred to as the PID performance curve.

6.3.2 Strategy and workflow

The objective of this study is to assess the efficiencies in identification and misidentification for various PID selections across different variables of interest, aiming to derive the profile of PID performance curves. For a comprehensive evaluation of PID performance, it is important to analyse and compare multiple detector and reconstruction configurations. Considering that the PID performance is particularly sensitive to the kinematic coverage of the calibration sample, systematic control of detector and trigger configurations is essential to validate the quality of the data-taking process.

To meet these requirements, an end-to-end software framework has been developed in pyROOT [193]. This tool facilitates the essential steps for the study, including data cleaning, exploratory data analysis, fitting and counting, efficiency evaluation, and plotting and comparing resulting PID performance curves. To ensure scalability and traceability, all information specific to a given analysis is stored in a dedicated configuration file. This allows for a straightforward execution of a new analysis by creating a new configuration file, ensuring prompt and efficient evaluation when new data becomes available. Additionally, general information shared across all analyses, such as global selections and display options, is stored in a global configuration file, ensuring uniformity and consistency across different processes. Furthermore, an automatic HTML report is generated for each step in the pipeline using the ROOT report package [194]. This feature enhances the visualisation and accessibility of the analysis results.

The algorithm is designed as a sequential pipeline, where each step can be independently executed for efficiency in time and computing resources. An automatic job scheduler has been implemented to simultaneously submit various jobs, further accelerating the overall process. A schematic view of the workflow is reported in Figure 6.5. The main steps in the pipeline include:

- **Ntuple Production**: This is the first step of the pipeline in which the tuples are centrally

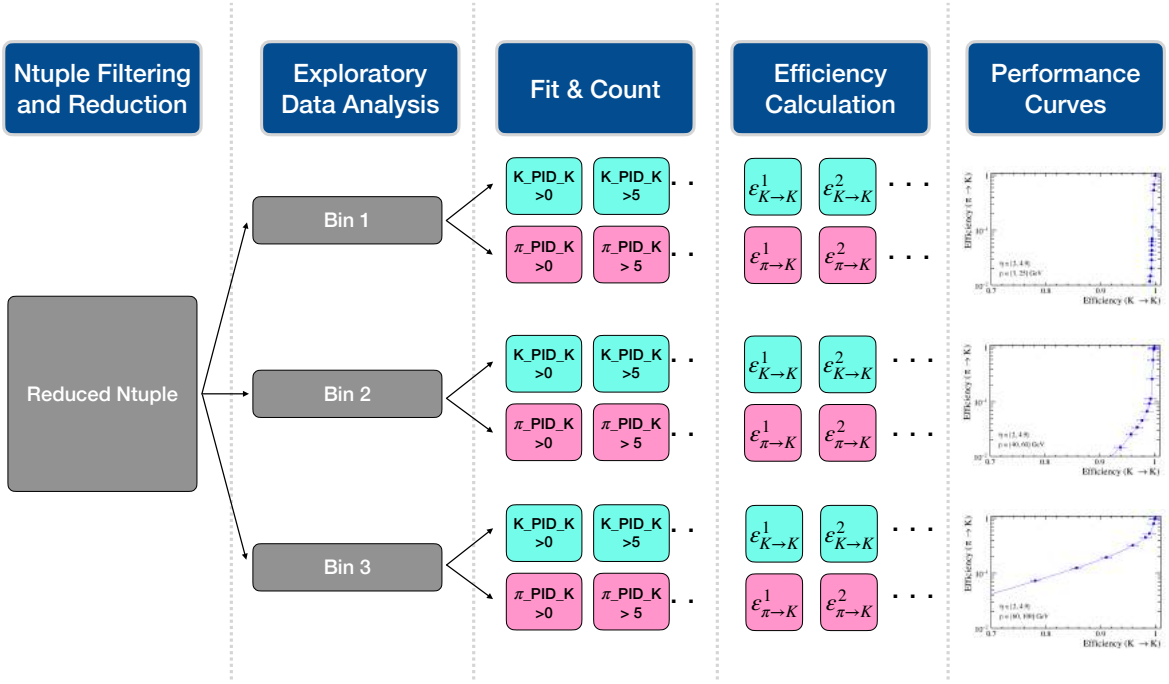


Figure 6.5: Schematic view of the pipeline for assessing particle identification performance. In this example, the dataset is divided into three subsets (bins), and the subsequent workflow is executed independently for each bin. This process results in the production of a performance curve for each subset. This modular approach allows for an efficient and parallelised analysis of different subsets.

produced with the LHCb Analysis Production [195] framework. The goal of the Analysis Productions is to optimise and automate the ntuple creation process, while maintaining a comprehensive record of dataset production. In Run 3, this is the preferred approach for generating ntuples in any type of analysis, including calibration and performance studies. The Analysis Production workflow involves analysts creating a merge request in a dedicated repository, with options files and associated metadata. Upon acceptance and merging of the request, the production process is submitted to LHCbDIRAC. Productions are executed using the DIRAC transformation system, and monitoring of the production is accessible on the Analysis Production webpage. Following the completion of transformations, the output data is replicated to CERN EOS. The AP framework is able to automatically update the existing ntuples when new data arrive.

- **Ntuple Filtering and Reduction**: This stage is designed to reduce data size by generating smaller, more manageable files. The data is directly loaded from the output of the Analysis Production, ensuring seamless integration with the standard dataflow of the nutple production for the calibration samples. The process involves selecting and saving only the relevant variables crucial for the performance study and applying offline candidate selection. When necessary, variable renaming is performed to maintain consistency with subsequent analysis steps, and new variables, such as PID_{PK} or ϕ, may be created. The RDataFrame interface [196] is employed to speed up the data processing through multi-

threading optimisation. The reduced data is then stored in a new file.

- **Exploratory Data Analysis**: This stage is crucial for evaluating the kinematic coverage of the calibration sample and ensuring the quality of data acquisition. The performance of PID detectors in detecting traversing particles is influenced by various factors, including the particle kinematics, detector occupancy, which may vary between events and for different particle production mechanisms, and experimental conditions such as spatial alignments. These conditions can modify the detectors response across runs [197]. Recognising any potential anomaly in the data-taking process is essential to validate the outcomes of the performance study. All pertinent variables used in the study are visualised, and relevant statistical information is presented. Two-dimensional distributions are plotted to check correlations. Custom selections on top of those defined in the preceding step can be employed, facilitating debugging or more in-depth investigations.

- **Fit & Count**: In this stage, a binned maximum likelihood fit is performed on the invariant mass distributions of the parent particle to extract signal yields, subsequently used for efficiency evaluation. The fit components are initialised and stored in a dedicated file, separating the model initialisation from the optimisation step. This division is useful when executing only one of the two steps, saving time accordingly. The sequence begins with fitting the invariant mass distribution with a very loose PID cut. Subsequently, the fit parameters are set to constant to fix the shape of the model, except from the yields and the background slope. Following this, a second fit is conducted to obtain the yields and background slope. This two-step fitting strategy serves to minimise uncertainties on the yields and facilitate the convergence of the fit. The fit status is checked iteratively, with automatic adjustments to initial parameters if convergence is not achieved. The procedure is repeated until the fit converges. There is also an option to perform the fit again based on the value of the χ^2/ndf. Once the fit converges, the PID cut is changed, and the previous routine is repeated, extracting again the signal yields for the new PID cut. Extracted signal yields and corresponding errors for both particle types are saved into separate arrays. These values are then utilised in the subsequent analysis step for computing PID efficiency.

- **Efficiency Calculation**: In this step, PID efficiencies and misidentification rates are determined by computing the ratio of the yields of candidate decays with and without PID selection, according to Equations 6.2 and 6.3. The uncertainties on the ratio are obtained through simple error propagation from the errors on the fit parameters. Plans for the future include implementing a more precise estimation of the error, such as binomial errors. The computed efficiencies, along with their corresponding errors, are saved into arrays.

- **Performance Curves**: In this step, the previously evaluated efficiency and misidentification are plotted to create the PID performance curve. Each point on the curve corresponds to a specific PID cut. If efficiency values are already available from previous results (e.g., Run 2 data), it is possible to generate the PID performance curves by loading data from simple text files, without the need to rerun the entire analysis. It is important to note the

uncertainties on the PID performance curves are purely statistical and highly correlated between points on the same curve. The process also involves the generation of tables containing relevant features, facilitating direct verification of computed values and their errors. Furthermore, this step enables the comparison of performance curves between different configurations. Finally, projections of the curves over relevant variables, such as pseudorapidity, momentum, and multiplicity, can be produced.

The analysis can be executed either over the entire dataset (integrated) or on specific data subsets (bins) obtained by dividing the data into intervals. Kinematic binning is achieved by ensuring that both particles under study fall within the same kinematic bin. The prospect of having binned performance curves is particularly valuable as it enables the mapping of RICH performance in specific phase-space regions.

Examining performance in relation to pseudorapidity (η) can be particularly insightful for identifying possible spatial dependencies. Additionally, studying performance as a function of particle momentum (p) allows for an assessment in different momentum regimes, with a critical focus on high momenta where Cherenkov angle saturation may impact PID performance, especially in the separation of kaons and pions. Another interesting study regards the performance dependence on the event multiplicity. The RICH reconstruction algorithm is expected to exhibit reduced performance with increasing event multiplicity due to a higher number of overlapping Cherenkov rings, leading to an overall degradation in PID performance.

Various multiplicity variables, including the number of reconstructed primary vertices (NPVS), the number of reconstructed tracks (NTRACKS), and the number of clusters in the fiber tracker (NFTCLUSTERS), are available. It is important to note that variables corresponding to the number of registered hits in the RICH detectors (nRich1Hits, nRich2Hits) are not accessible at the time of implementation. All the multiplicity variables exhibit correlations among themselves. As an example, Figure 6.6 provides the distributions of NPVS and NTRACKS, along with their correlation. In this study, NPVS is utilised as a proxy for multiplicity, specifically to emphasise the performance variation at different pile-up values.

Different binning schemes, whether combined or integrated, can be implemented. The choice of binning is presently limited by statistical constraints. More refined binning schemes will be introduced as additional statistics become available in the calibration sample. Table 6.2 presents several binning schemes utilised in the study of PID for pp data from 2022. A summary of the variables necessary for the analysis, along with their corresponding syntax names, is outlined in Table 6.3.

Binned variable	η	p (GeV/c)	NPVS
Integrated	[2.0, 4.9]	[3, 150]	[1, 20]
η	[2.0, 3.0] [3.0, 3.5] [3.5, 4.0] [4.0, 4.9]	[3, 150]	[1, 20[
p	[2.0, 4.9]	[3, 25] [25, 40] [40, 60] [60, 100]	[1, 20[
NPVS	[2.0, 4.9]	[3, 150]	[1, 2[[2, 3[[3, 5[[5, 10[

Table 6.2: Binning schemes used for PID analysis for pp data collected in 2022, with ranges specified for pseudorapidity (η), momentum (p), and number of reconstructed primary vertices (NPVS). The binning over a variable is obtained by requiring that both particles under study fall within the same interval of that variable.

Variable type	Syntax name
General	FILLID, RUNNUMBER
Invariant Mass	L0_M, D0_M, Dst_M
Track quality	TCHI2DOF, GHOSTPROB
Kinematic	ETA, P, PX, PY, PZ, PT, E_ECAL
Multiplicity	NPVS, NTRACKS
PID	PIDK, PIDP

Table 6.3: Summary of the variables necessary for the analysis showing the corresponding syntax names.

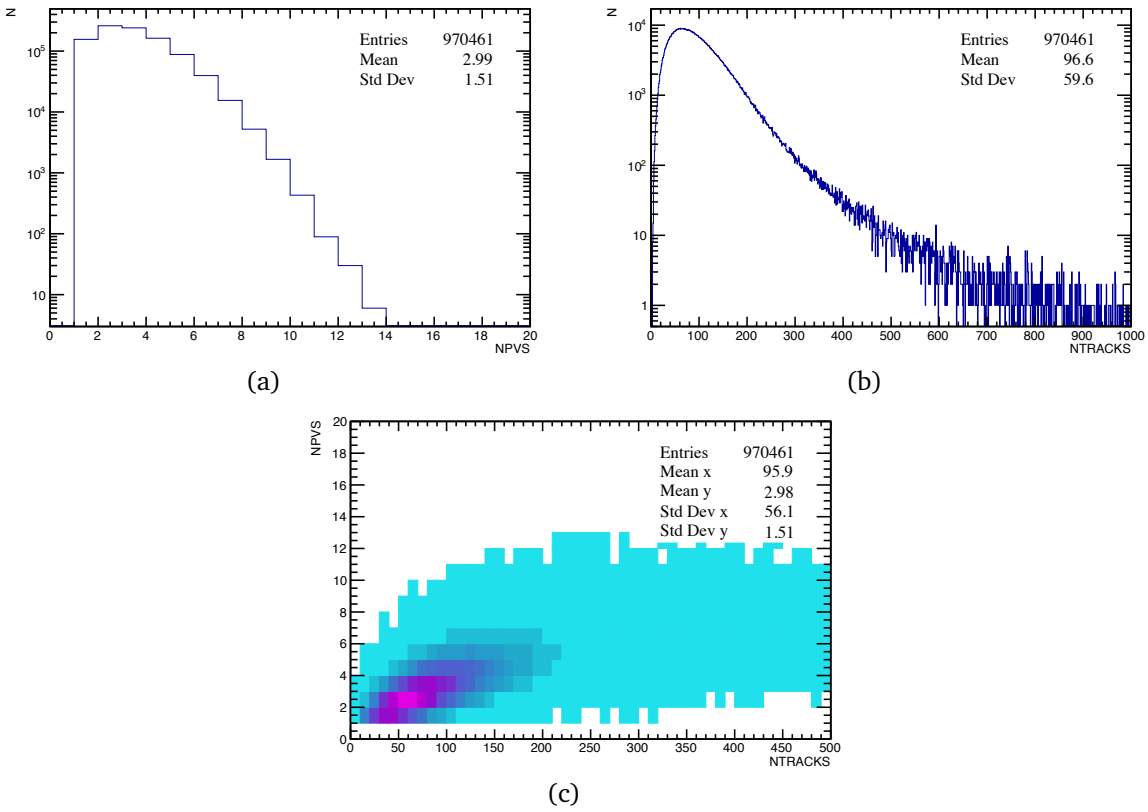


Figure 6.6: Distributions (a) number of reconstructed primary vertices (NPVS) and (b) the number of reconstructed tracks (NTRACKS) in pp data collected in 2022 for the $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ decay channel. (c) Correlation between the two distributions.

6.4 Early performance with 2022 pp data

The first preliminary study of the RICH performance for pp collisions in Run 3, is evaluated with data collected by the LHCb experiment at the end of the 2022 commissioning year. Data corresponding to the LHC fills 8489 and 8491 are considered, with run numbers between 255949 and 256170, corresponding to a total stored luminosity of approximately $L \sim 23$ pb^{-1}. Additional details about the data are summarised in Table 6.4. The detector conditions during this data-taking period are the following:

- the magnet polarity is down;
- the VELO is in closed position (1 mm gap);
- the SciFi tracker is in warm conditions at a temperature of about +20°C [199];
- the UT is not installed;

The calibration samples are selected in the HLT2 passthrough stream. Then the available data is reprocessed with the best subdetectors alignment and calibration conditions. Samples of $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ and $\Lambda^0 \to p\pi^-$ decays are selected through dedicated HLT2 lines. Due to the high production cross section of visible Λ^0 in LHCb, only a a small portion of the data is used for calibration. Certain categorisations can be applied to selectively bias the recorded candidates, enhancing the coverage of the sample. To achieve this, three HLT2 lines (L0ToPPi_LL, L0ToPPi_LL_HighPt, L0ToPPi_LL_VeryHighPt), which involve variations in the requirement on the proton transverse momentum (p_T), are implemented to target the $\Lambda^0 \to p\pi^-$ decay mode. The selection implemented in the HLT2 lines are reported in Table 6.5 and Table 6.6. The tuples are centrally produced with the LHCb Analysis Production framework.

Additionally, to enhance the purity of the calibration samples, further selections based on track quality (χ^2/ndf) are applied. A requirement on the probability of the tracks being ghosts (GHOSTPROB) is also applied, and all tracks are required to have $p_T > 250$ MeV. The momentum and pseudorapidity ranges are chosen to match with the acceptances of the RICH detectors. Both D^0 and Λ^0 are required to be consistent with originating from the primary vertex by applying a cut on the χ^2_{BPVIP}. For the $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ decay sample, additional criteria are imposed to enhance the purity of the D^{*+} invariant mass distribution, including requiring the D^0 meson to have $p_T > 2$ GeV. These requirements are summarised in Tables 6.7 and 6.8.

fill ID	Duration	Delivered Luminosity (pb^{-1})	Stored Luminosity (pb^{-1})	Average μ
8491	08:56:39	22.9	12.2	1.9
8489	15:57:36	23.9	10.8	1.0
Total	24:54:15	46.8	23.0	/

Table 6.4: Summary of the pp data collected by the LHCb experiment used to study the preliminary performance in 2022 [198].

Particle	Requirements
K^-, π^+	$p_{\mathrm{T}} > 250$ MeV $p > 2$ GeV $\chi^2_{IP} > 16$
$K^-\pi^+$	$\lvert m - m_{D0}\rvert < 95$ MeV/c^2, $p_T > 1.5$ Gev/c, DOCA < 0.1 mm
D^0	$\lvert m - m_{D0}\rvert < 75$ MeV/c^2, χ^2_{VTX}/ndf<10, χ^2_{VS}<49 DIRA$_{PV}$ > 0.9999 $\lvert m(K^- \to \pi^-, \pi^+ \to K^+) - m_{D0}\rvert > 25$ MeV/c^2 $\lvert m(\pi^+ \to K^+) - m_{D0}\rvert < 25$ MeV/c^2 NumChildren($p_T > 1$ GeV/c) ≥ 1
π^+	$p_{\mathrm{T}} > 0.1$ GeV/c, $p > 1$ GeV, χ^2_{trk}/ndf < 3
$D^0\pi^+$	$\lvert m - m_{D0}\rvert < 112.5$ MeV/c^2, $135 < m - m^{cand}_{D0} < 175$ MeV/c^2
D^{*+}	$\lvert m - m_{D^*}\rvert < 95$ MeV/c^2, χ^2_{VTX}/ndf < 10

Table 6.5: Definition of the HLT2 line DstToD0Pi_D0ToKPi used for selecting candidates from the $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ decay channel.

Particle	Requirements
p	$p_{\mathrm{T}} > 0/3/6$ GeV/c $p > 2$ GeV/c $\chi^2_{IP} > 36$
π^-	χ^2_{trk}/ndf < 4, $\chi^2_{IP} > 36$
$p\pi^-$	$\lvert m - m_{\Lambda^0}\rvert < 50$ MeV/c^2
Λ^0	$\tau > 2$ps, χ^2_{vtx}/ndf < 30, $\chi^2_{IP} > 50$ $\lvert m - m_{\Lambda^0}\rvert < 20$ MeV/c^2, $\lvert m(p \to \pi^+) - m_{\Lambda^0}\rvert < 20$ MeV/c^2

Table 6.6: Definition of the HLT2 line L0ToPPi_LL, L0ToPPi_LL_HighPt, L0ToPPi_LL_VeryHighPt used for selecting candidates from the $\Lambda^0 \to p\pi^-$ decay channel.

Particle	Requirements
K^-, π^+	$\chi^2/\text{ndf} < 3$
	GHOSTPROB < 0.4
	$p_\text{T} > 250$ MeV
	$3 < p < 150$ GeV
	$2 < \eta < 4.9$
π_s^+	$\chi^2/\text{ndf} < 3$
	GHOSTPROB < 0.4
D^0	$\chi^2_\text{BPVIP} < 20$
	$p_\text{T} > 2$ GeV

Table 6.7: Selection requirements on top of the HLT2 line output for $D^{*+} \to D^0(\to K^-\pi^+)\pi_\text{s}^+$ decays. The accompanying soft pion is labelled as π_s^+.

Particle	Requirements
p	$\chi^2/\text{ndf} < 4$
	GHOSTPROB < 0.4
	$p_\text{T} > 250$ MeV
	$3 < p < 150$ GeV
	$2 < \eta < 4.9$
π	$\chi^2/\text{ndf} < 4$
	GHOSTPROB < 0.4
Λ	$\chi^2_\text{BPVIP} < 20$

Table 6.8: Selection requirements on top of the HLT2 line output for $\Lambda^0 \to p\pi^-$ decays.

The data that pass the previous criteria undergoes further analysis to validate the quality of the calibration samples. The kinematic coverage of 2022 pp events passing the HLT2 and the offline selection is displayed in Figure 6.7. Kaons and pions exhibit similar distributions in pseudorapidity versus momentum, as they both originate from the same decay channel. In the pseudorapidity versus momentum distribution of protons, the three contributions in p_T originating from the three distinct HLT2 lines can be distinguished. The kinematic coverage is overall comparable across all three particles. This suggests that any variation in PID performance induced by differences in kinematic coverage could be considered negligible, at least for these preliminary studies. Consequently, no kinematic reweighting is applied in this analysis. The distribution of NPVS is also reported, showing that the average number of NPVS for events that pass the selection is higher than the average μ of the data-taking. This bias effect arises from the reconstruction and the event selection processes.

The efficiency and the misID fraction for kaons and pions are determined from the $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ channel with a 2D fit on the $D^0 \to K^-\pi^+$ invariant mass distribution, in the range $1800 < m_{D^0} < 1928\,\text{MeV}$, and on the $\Delta m = m_{D^{*+}} - m_{D^0}$ distribution, in the range $140 < \Delta m < 155\,\text{MeV}$, as shown in Figure 6.8. This approach aims to better constrain the residual background, mainly combinatorial and from random pions combined with a D^0 candidate. Efficiency and misID fractions for protons are evaluated from the $\Lambda^0 \to p\pi^-$ channel by performing a fit to the Λ^0 invariant mass distribution, in the range $1107 < m_{\Lambda^0} < 1125\,\text{MeV}$, as illustrated in Figure 6.9.

The first set of results focuses on the overall PID performance across the entire dataset. Figure 6.10 presents integrated performance curves, illustrating the ID fraction for kaons and protons, with the corresponding misID fraction involving pions and kaons. A comparison between Run 3 and Run 2 is provided. The momentum range is chosen to ensure that at least one particle hypothesis is above the velocity threshold for Cherenkov radiation generation. Notably, there is an evident improvement in Run 3 performance compared to Run 2. This enhancement is attributed to a better single-photon resolution [192] and increased photon yield. The enhancements are even more pronounced considering the higher average LHCb detector occupancy in the 2022 dataset, with an average number of reconstructed primary vertices around $N_{PV} \sim 3.0$, compared to $N_{PV} \sim 1.8$ in Run 2. Table 6.9 provides efficiency and misidentification rates for both Run 3 and Run 2 data.

Charged-particle identification has also been studied as a function of pseudorapidity. Figure 6.11 depicts the identification efficiency (ε) of kaons and protons, compared to the misID fraction with pions and kaons, as a function of the track pseudorapidity. The separation between protons and pions is nearly ideal in the range $\eta \in [2.0, 4.0]$. However, all three scenarios exhibit a degradation in performance at high pseudorapidity ($\eta > 4.0$).

Another significant analysis involves the PID curve binned in momentum intervals. Figure 6.12 present the identification efficiency of kaons and protons, with the corresponding misID fraction of pions and kaons, as a function of track momentum. The observed trend in momentum is associated with different velocity thresholds for generating Cherenkov radiation in RICH1 and RICH2. Specifically, the separation between kaons and pions is optimal for $p < 40$ GeV/c, fol-

lowed by a degradation attributed to the saturation of the Cherenkov angle at high momentum. This effect is more pronounced for particles with similar mass, therefore particularly evident in the separation of kaons and pions and less prominent in other particle combinations. The separation between protons and kaons exhibits a slightly different trend, showing sub-optimal performance for $p < 25$ GeV/c. Positive proton-kaon separation in both RICH1 and RICH2 is possible above approximately 30 GeV/c. The reduced performance in the momentum bin $[10, 25]$ GeV/c can be explained by the momentum thresholds for generating Cherenkov light by kaons and protons. For instance, below 15.6 GeV/c, protons and kaons cannot be distinguished in RICH2, while between 15.6 GeV/c and 29.7 GeV/c they can be distinguished in RICH2 in veto-mode only. Consequently, at low momentum, the discrimination of protons and kaons relies solely on RICH1, resulting in an overall degradation of the PID performance. This observation underscores the significant contribution of RICH2 to PID performance for high-momentum tracks.

The final set of results pertains to the study of PID performance as a function of the activity in the detector. Figures 6.13 illustrate the identification efficiency of kaons and protons, compared to the misID fraction with pions and kaons, as a function of the number of reconstructed primary vertices. All three plots demonstrate a degradation in PID performance with increased interaction multiplicity, consistent with expectations due to a higher number of overlapping Cherenkov rings. The separation of kaons and pions appears sub-optimal even at very low multiplicity ($NPVS \in [1, 2]$), whereas the separation of protons and pions and protons and kaons shows substantial improvements in the same interval. To investigate this behaviour further, Figure 6.14 compares two sets of curves in different momentum intervals. The plot corresponding to the momentum region $p \in [3, 80]$ GeV/c exhibits a sharp improvement, indicating that the sub-optimal performance at low multiplicity is dominated by high-momentum particles.

Furthermore, it is relevant to highlight that the last multiplicity interval, corresponding to $NPVS \in [5, 10]$, includes the largest event occupancies expected in nominal Run 3 instantaneous luminosity. Consequently, it is interesting to compare the PID curves in this specific subset with the nominal performance in Run 2. This comparison is presented in Figure 6.15. The performance in the highest N_{PV} bin closely resembles that of Run 2, confirming that the upgraded RICH system can achieve the same high-quality charged-hadron identification as in Run 2, despite the increase in the average number of visible interactions. Table 6.10 provides efficiency and misidentification rates for Run 2 and Run 3 with $NPVS \in [5, 10]$ for three different PID cuts. This preliminary result demonstrates the achievement of the RICH upgrade design goals.

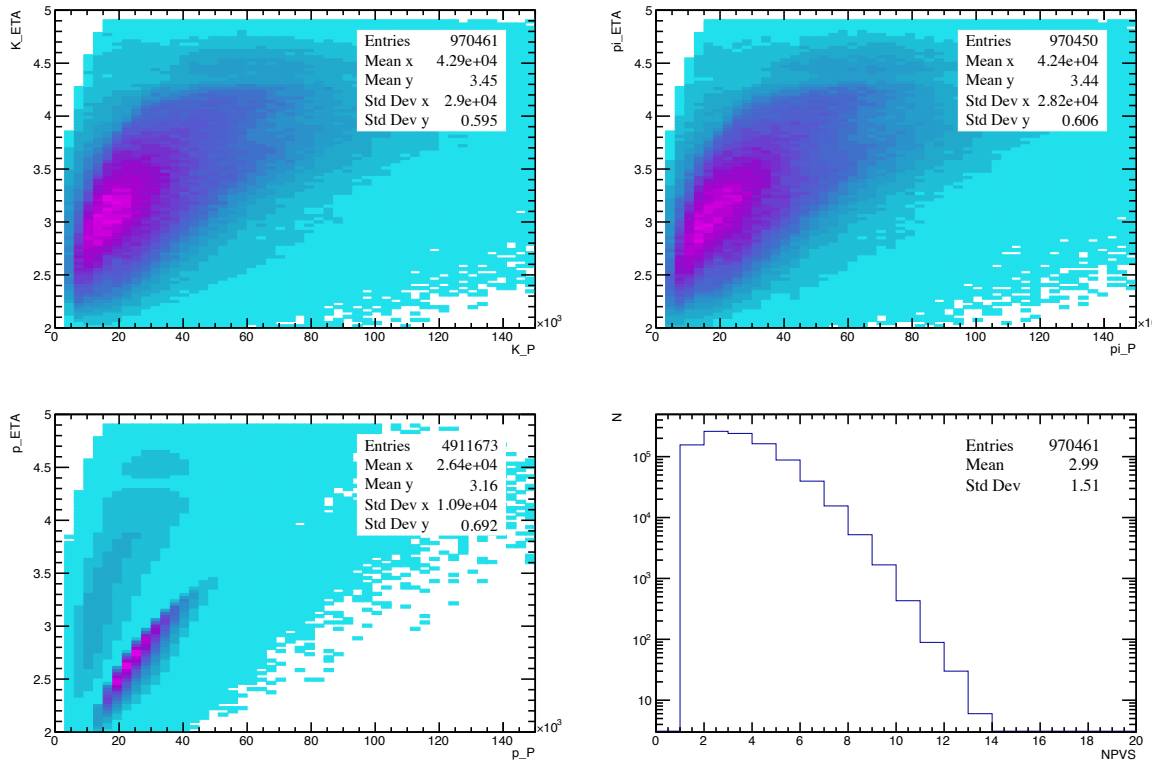
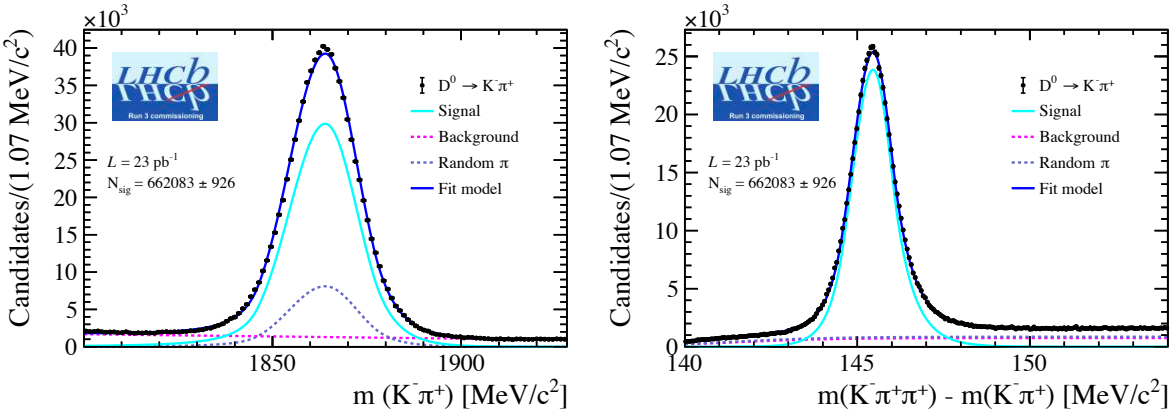


Figure 6.7: Kinematic coverage of pp data collected in 2022 after the HLT2 and the offline candidate selections. The pseudorapidity versus momentum distribution is plotted for protons, kaons and pions. The distribution of NPVS is also reported, as a proxy for the multiplicity.



(a) (b)

Figure 6.8: Invariant mass spectrum of the $K^-\pi^+$ final-state (a) and distribution of the mass difference $\Delta m = m_{D^{*+}} - m_{D^0}$ distribution (b) for pp data collected in 2022. The data is overlaid with the result of a 2D fit on the two distributions. In the $K^-\pi^+$ invariant mass spectrum (a), the $D^0 \to K^-\pi^+$ signal is modelled with a double-sided crystal ball function, the random$-\pi$ component is modelled with a Gaussian function and the background is modelled with a polynomial function. In the Δm distribution (b), the $D^0 \to K^-\pi^+$ signal is modelled with a double-sided crystal ball function and the random$-\pi$ component and the background are modelled with a Fermi function.

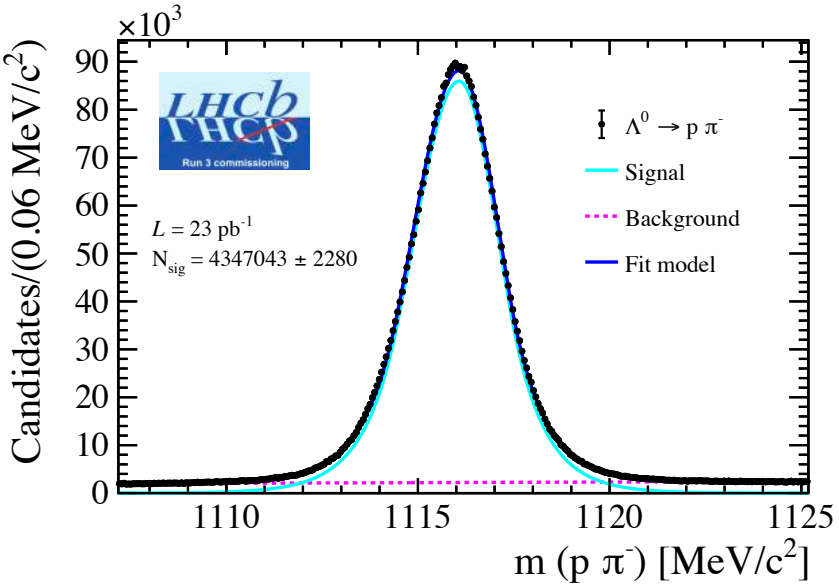


Figure 6.9: Invariant mass spectrum of the $p\pi^-$ final-state for pp data collected in 2022. The data is overlaid with the result of a fit, where the $\Lambda^0 \to p\pi^-$ signal is modelled with a double-sided crystal ball function and the background background is modelled with a polynomial function.

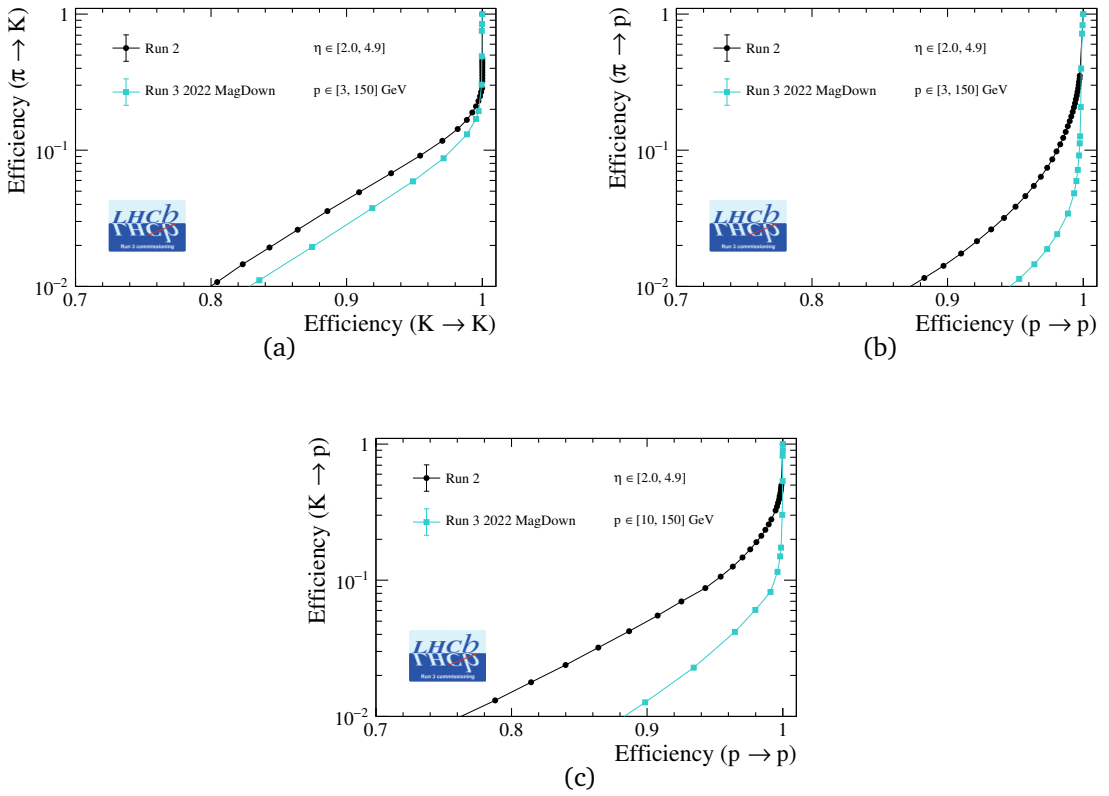


Figure 6.10: The efficiency of selecting (a) kaons, (b) protons, (c) protons, with the associate leakage from misidentifying (a) pions, (b) pions, (c) kaons. The momentum range is $3 < p < 150$ GeV for (a) and (b) and $10 < p < 150$ GeV for (c).

	Run 3 2022		Run 2	
PID_K cut	$\varepsilon\ (K \to K)$	misID $(\pi \to K)$	$\varepsilon\ (K \to K)$	misID $(\pi \to K)$
> 0	94.9 ± 0.2	5.73 ± 0.04	95.3	9.4
> 5	87.5 ± 0.2	1.87 ± 0.02	84.3	2.4
> 10	81.2 ± 0.2	0.74 ± 0.02	74.9	0.8
	Run 3 2022		Run 2	
PID_P cut	$\varepsilon\ (p \to p)$	misID $(\pi \to p)$	$\varepsilon\ (p \to p)$	misID $(\pi \to p)$
> 0	99.49 ± 0.07	5.94 ± 0.04	96.9	6.4
> 5	99.32 ± 0.07	3.42 ± 0.03	93.2	2.6
> 10	97.33 ± 0.07	1.88 ± 0.02	86.7	1.0
	Run 3 2022		Run 2	
PID_{PK} cut	$\varepsilon\ (p \to p)$	misID $(K \to p)$	$\varepsilon\ (p \to p)$	misID $(K \to p)$
> 0	97.98 ± 0.07	6.06 ± 0.04	95.3	9.4
> 5	93.44 ± 0.07	2.28 ± 0.03	84.3	2.4
> 10	87.28 ± 0.06	0.86 ± 0.02	74.9	0.8

Table 6.9: Efficiency (ε) and misidentification rates for pp data collected in 2022 and Run 2 for three different PID cuts (>0, >5, >10) for separations of kaons vs. pions (PID_K), protons vs. pions (PID_P), and protons vs. kaons (PID_{PK}).

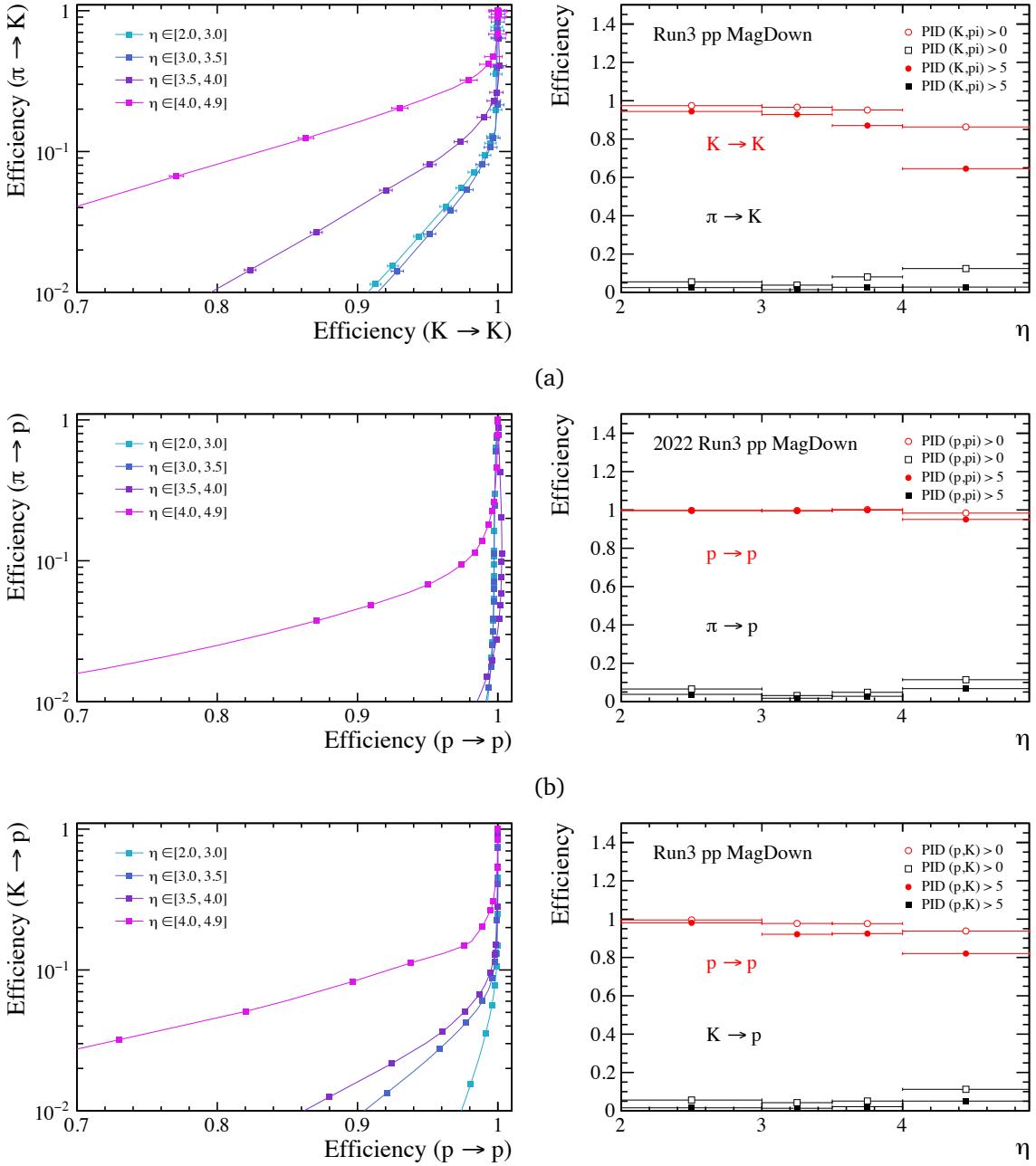
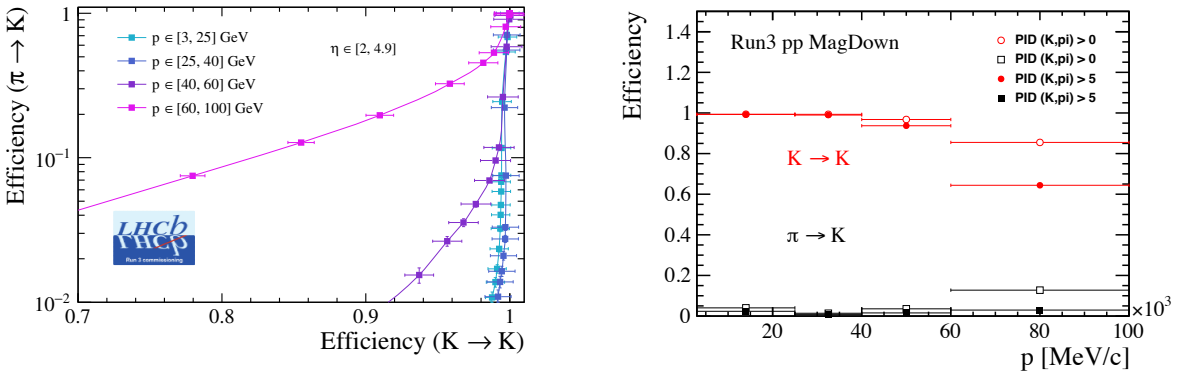
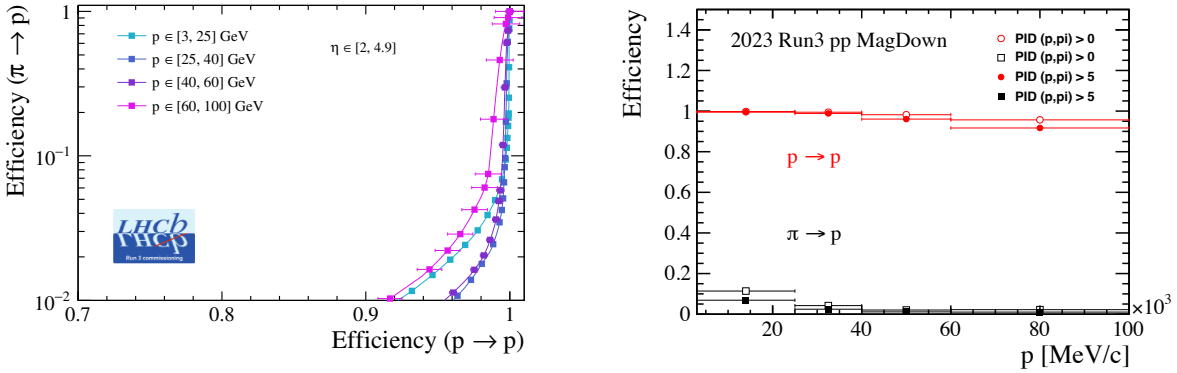


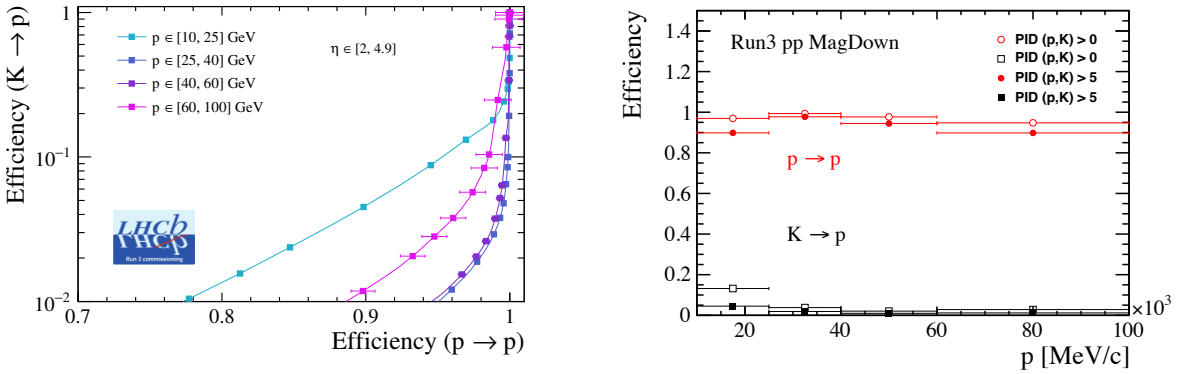
Figure 6.11: The efficiency of selecting (a) kaons, (b) protons, (c) protons, with the associate leakage from misidentifying (a) pions, (b) pions, (c) kaons as a function of the track pseudorapidity for pp data collected in 2022.



(a)

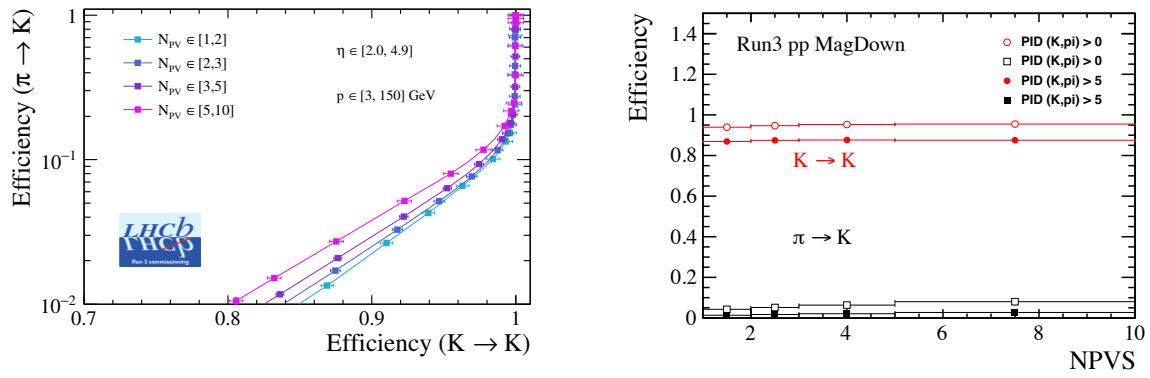


(b)

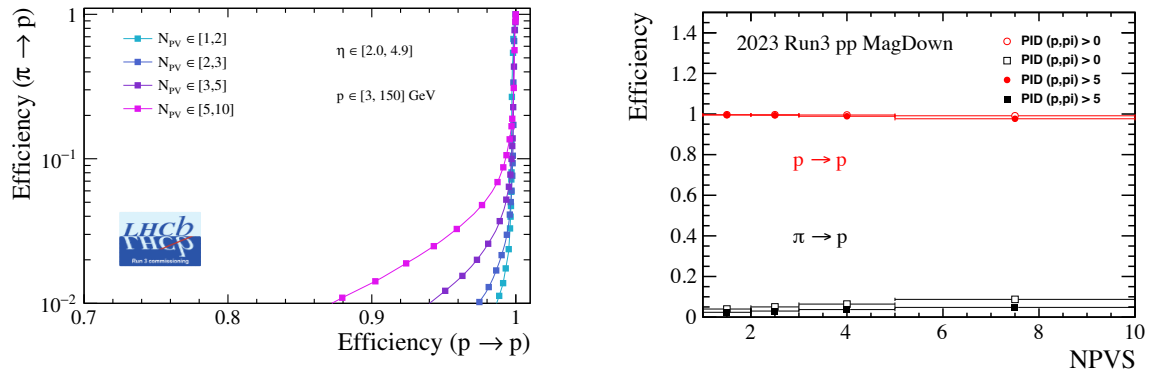


(c)

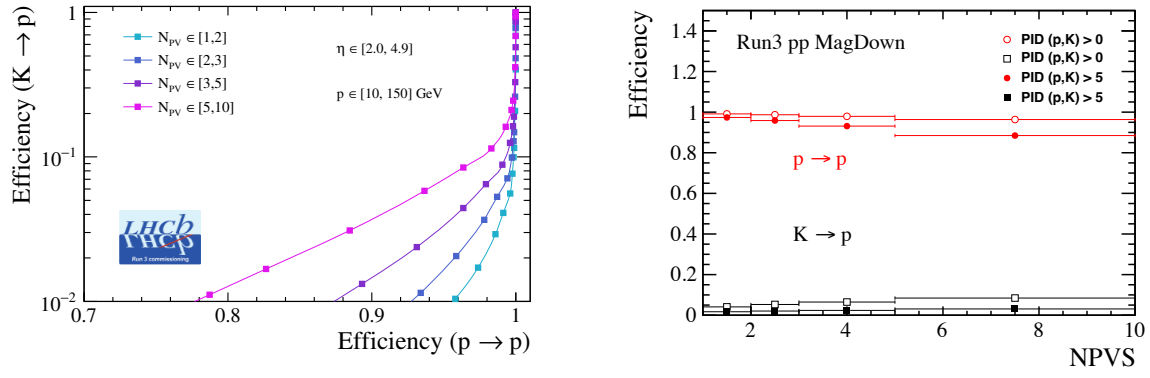
Figure 6.12: The efficiency of selecting (a) kaons, (b) protons, (c) protons, with the associate leakage from misidentifying (a) pions, (b) pions, (c) kaons as a function of the track momentum for pp data collected in 2022.



(a)

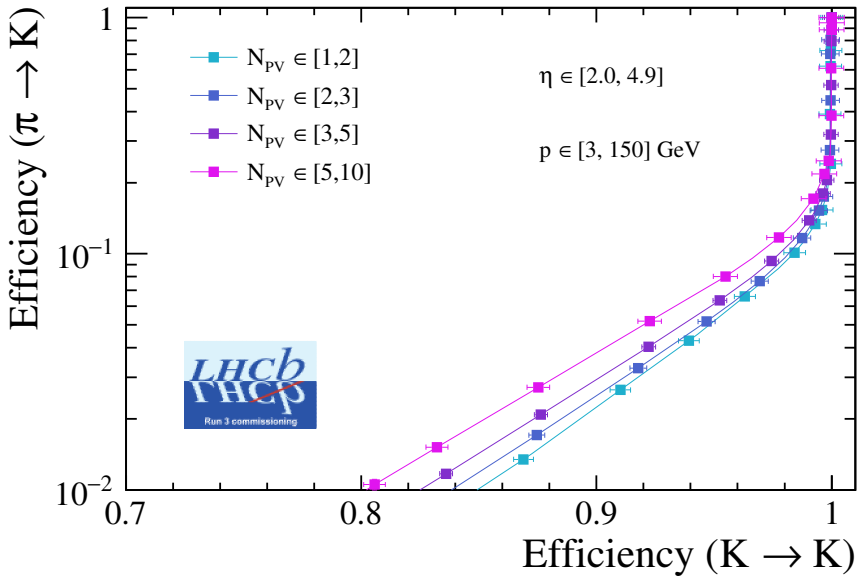


(b)

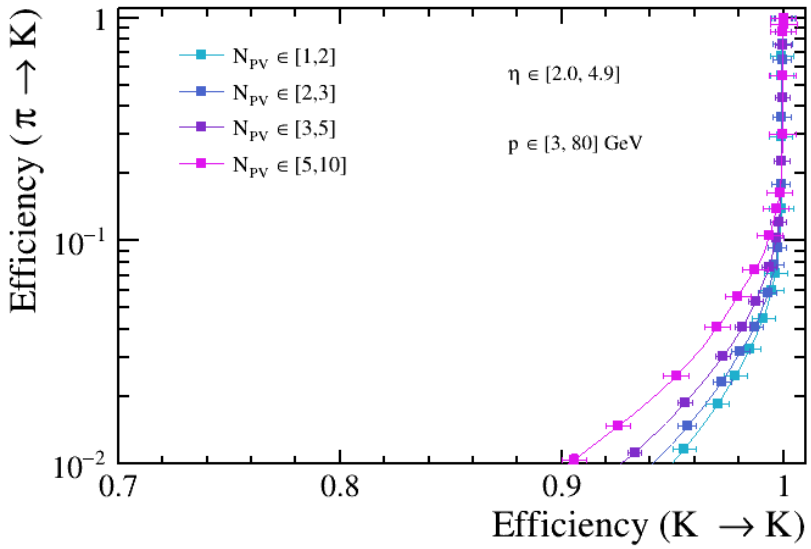


(c)

Figure 6.13: The efficiency of selecting (a) kaons, (b) protons, (c) protons, with the associate leakage from misidentifying (a) pions, (b) pions, (c) kaons as a function of the number of reconstructed primary vertices for pp data collected in 2022. The momentum range is $3 < p < 150$ GeV for (a) and (b) and $10 < p < 150$ GeV for (c)



(a)



(b)

Figure 6.14: The efficiency of selecting kaons with the associate leakage from misidentifying pions as a function of the number of reconstructed primary vertices in the momentum ranges (a) $3 < p < 150$ GeV and (b) $3 < p < 80$ GeV for pp data collected in 2022. The plot corresponding to the momentum region $p \in [3,80]$ GeV/c exhibits a notable enhancement, suggesting that the suboptimal performance at low multiplicity is primarily influenced by high-momentum particles.

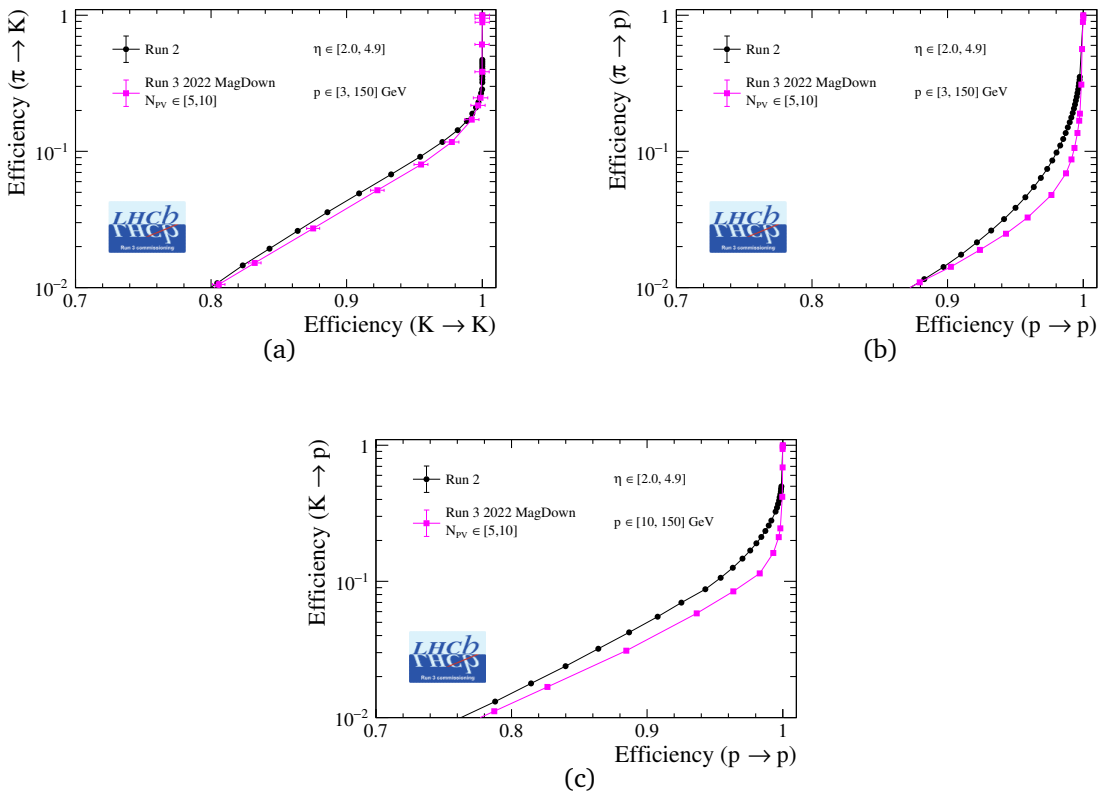


Figure 6.15: The efficiency of selecting (a) kaons, (b) protons, (c) protons, with the associate leakage from misidentifying (a) pions, (b) pions, (c) kaons for pp data collected in 2022 in a high multiplicity bin (magenta) and in Run 2 (black). The momentum range is $3 < p < 150$ GeV for (a) and (b) and $10 < p < 150$ GeV for (c).

	Run 3 2022, $NPVS \in [5,10]$		Run 2	
PID_K cut	ε $(K \to K)$	misID $(\pi \to K)$	ε $(K \to K)$	misID $(\pi \to K)$
> 0	96.7 ± 0.3	8.6 ± 0.1	95.3	9.4
> 5	90.1 ± 0.3	3.4 ± 0.1	84.3	2.4
> 10	83.1 ± 0.3	1.4 ± 0.1	74.9	0.8
	Run 3 2022, $NPVS \in [5,10]$		Run 2	
PID_P cut	ε $(p \to p)$	misID $(\pi \to p)$	ε $(p \to p)$	misID $(\pi \to p)$
> 0	99.1 ± 0.2	8.7 ± 0.1	96.9	6.4
> 5	97.7 ± 0.2	4.8 ± 0.1	93.2	2.6
> 10	94.3 ± 0.2	2.5 ± 0.1	86.7	1.0
	Run 3 2022, $NPVS \in [5,10]$		Run 2	
PID_{PK} cut	ε $(p \to p)$	misID $(K \to p)$	ε $(p \to p)$	misID $(K \to p)$
> 0	96.4 ± 0.2	8.5 ± 0.1	95.3	9.4
> 5	88.5 ± 0.2	3.1 ± 0.1	84.3	2.4
> 10	78.7 ± 0.2	1.1 ± 0.1	74.9	0.8

Table 6.10: Efficiency (ε) and misidentification rates for pp data collected in 2022 with $NPVS \in [5,10]$ and Run 2 data for three different PID cuts (>0, >5, >10) for separations of kaons vs. pions (PID_K), protons vs. pions (PID_P), and protons vs. kaons (PID_{PK}).

6.5 Early performance with 2022 pAr data

As detailed in Section 5.5, the LHCb fixed-target system underwent an upgrade with the installation of the SMOG2 storage cell. A notable difference from Run 2 is the separation of the fixed-target interaction point from the nominal pp interaction point. Specifically, the fixed-target interaction point is located approximately ~ 50 cm upstream along the z axis, as illustrated in Figure 6.16. Given that LHCb was initially designed and optimised for the nominal pp interaction point, it is crucial for the fixed-target physics program to evaluate the impact of the displacement in the z direction on the PID performance. This section presents preliminary studies comparing the PID performance in Run 3 between fixed-target and collider modes.

This comparison is based on data collected in November 2022 during the commissioning of the SMOG2 system, which involved simultaneous pp and pAr collisions. Specifically, for pAr collisions, data from run numbers 255622 and 255623 in LHC fill 8484 are considered. The detector conditions during this data-taking period are as follows:

- the magnet polarity is down;
- the VELO is in closed position;
- the SciFi tracker is in warm conditions at a temperature of about +20°C [199];
- the UT is not installed;

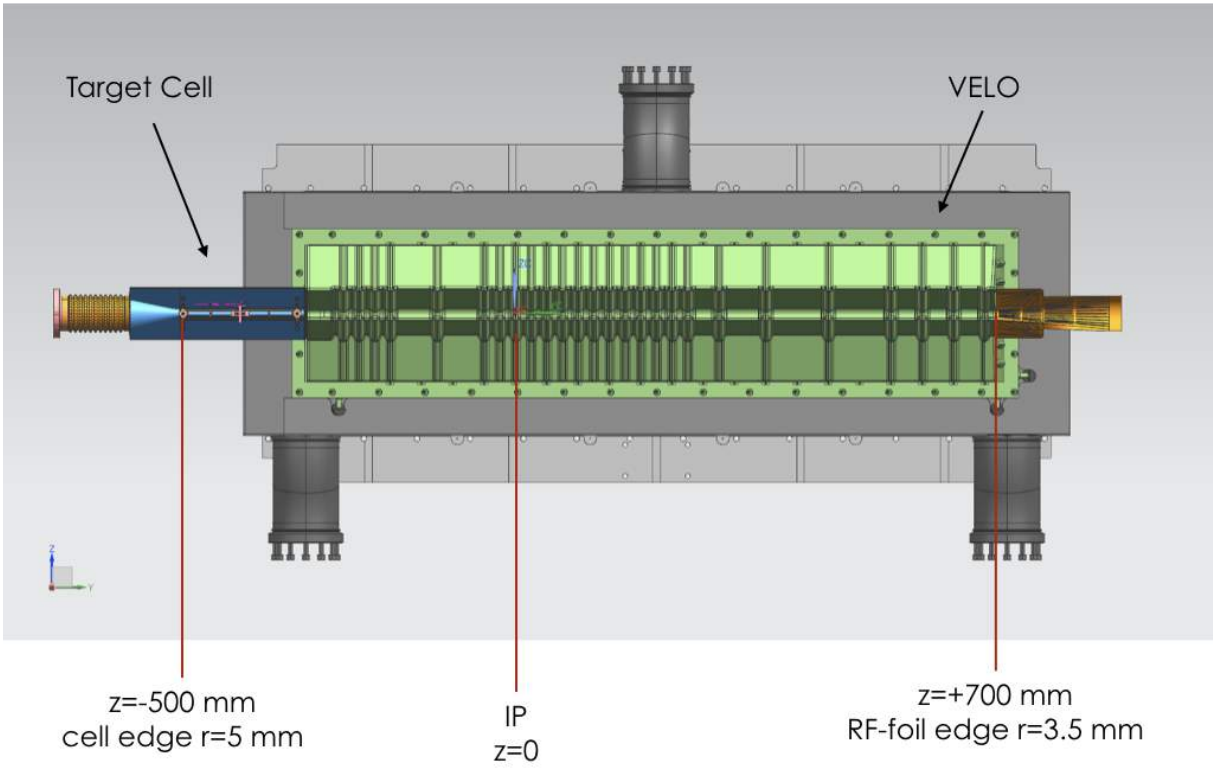


Figure 6.16: Side view of the SMOG2+VELO system. The coordinate with respect to the IP and the radius of the extreme apertures are reported [132].

- Argon gas is injected into the SMOG2 cell over a period of approximately 18 minutes. Throughout the injection process, the pressure inside the SMOG2 cell remains stable, maintaining an average value of around 9.6×10^{-8} mbar.

The Analysis Production framework presently does not automate the processing of data collected in the fixed-target configuration, therefore a private data production is performed. The available data from the HLT2 SMOG2 data stream is processed with Moore (v54r5). Tuples are obtained by running DaVinci (v63r5) on the Moore output, with the alignment version employed for pp collisions, presented in Section 6.4, ensuring consistency in detector calibration. Samples of identified protons originating from $\Lambda^0 \to p\pi^-$ decays are obtained through a dedicated HLT2 line. Further details regarding candidate selection are available in Reference [200]. In addition to baseline selection in the trigger lines, offline selection is applied to enhance signal purity, as detailed in Table 6.8. Identified kaon samples for fixed-target collisions are currently unavailable due to the absence of an unbiased calibration sample for D^0 in pAr data. Consequently, only the results concerning the separation of protons and pions are presented.

The kinematic coverage of protons and pions from the $\Lambda^0 \to p\pi^-$ calibration sample for pAr collisions collected in 2022 is illustrated in Figure 6.17. Due to the kinematics of the two-body decay, protons originating from the Λ^0 particle always carry a higher momentum fraction than pions. Consequently, protons are produced with momenta up to 150 GeV/c, while pions cover a smaller momentum region, up to 40 GeV/c. To address this imbalance, the separation of protons and pions is studied in a smaller momentum region common to both particles ($p \in [3, 20]$ GeV/c), reducing overall statistics but mitigating the influence of the lack of kinematic coverage on PID performance. Another aspect of fixed-target kinematics is the pseudorapidity coverage, with particles produced in fixed-target collisions boosted in the forward region in the laboratory frame. Considering this substantial difference from collider data, the comparison between pp and pAr is conducted within a specific subinterval of the entire RICH acceptance. This selected interval corresponds to $\eta \in [3.6, 4.2]$, where the majority of tracks in pAr data are reconstructed, as displayed in Figure 6.17.

Efficiency and misID fractions for protons in pp and pAr data are determined through a 1D fit on the corresponding Λ^0 invariant mass distribution, presented in Figure 6.18, within the range $1107 < m_{\Lambda^0} < 1125 \text{MeV}/c^2$. For a proper comparison, proton and pion candidates from pp data are selected from $\Lambda^0 \to p\pi^-$ decays using the same momentum and pseudorapidity intervals applied for pAr. Figure 6.19 compares proton selection efficiency to the misID fraction with pions in pp and pAr data collected in 2022, demonstrating overall agreement in PID performance. Although statistically limited due to low running time and tight selection to compensate for the lack of kinematic coverage, this preliminary result reveals that PID for fixed-target and collider data is similar, indicating that the position of the interaction point has a negligible impact on overall PID performance. Table 6.11 provides values for efficiency and misID rates for pp and pAr data.

A direct comparison between PID performance in fixed-target collisions in Run 2 and Run 3 is not feasible, as no dedicated study for PID performance in fixed-target collisions was con-

ducted in Run 2. However, considering the significant improvement in PID performance for pp collisions in Run 3 (Figure 6.10), it can be concluded that a similar improvement is expected to hold also for fixed-target collisions in Run 3. These enhancements are anticipated to have a positive impact on physics analyses, especially those focused on heavy-flavour and charged-hadron production. In particular, in an analysis similar to the one detailed in Chapter 4, a more accurate estimate of fit templates can be achieved in Run 3, thus leading to a reduction of the associated systematic uncertainty.

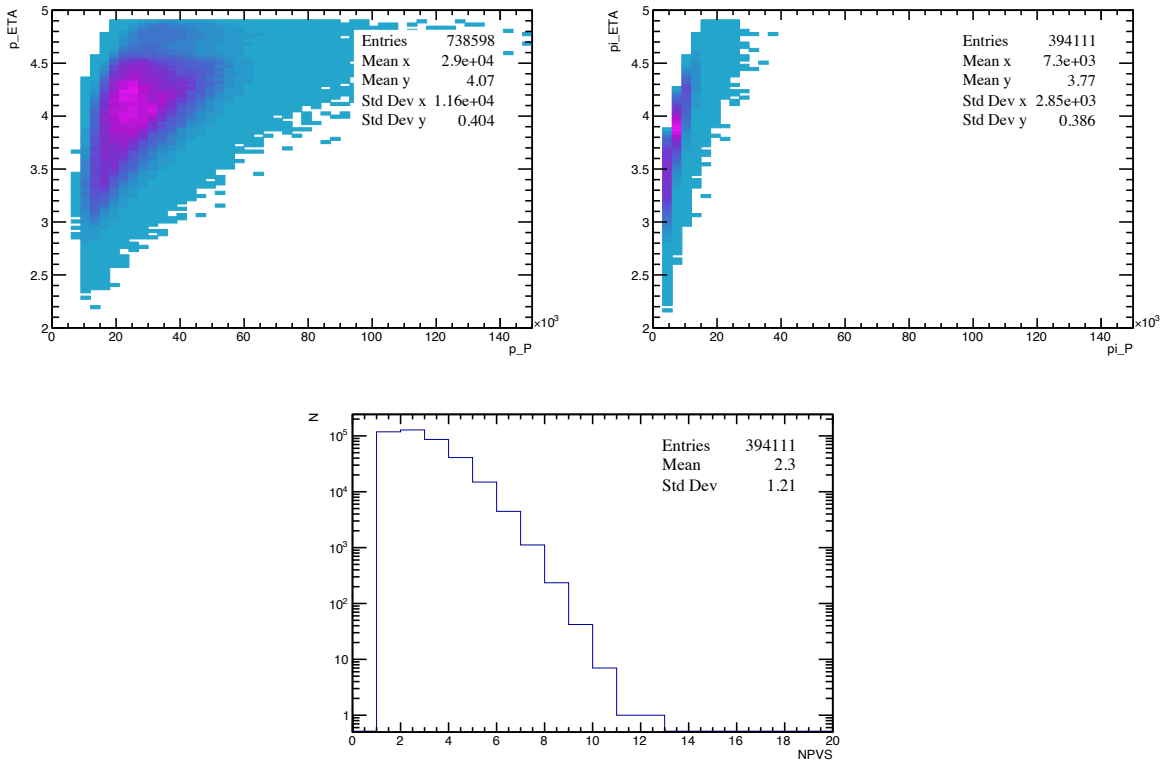


Figure 6.17: Kinematic coverage of *p*Ar data collected in 2022 after the HLT2 and the offline candidate selections. The pseudorapidity versus momentum distribution is plotted for protons and pions. The distribution of NPVS is also reported, as a proxy for the multiplicity.

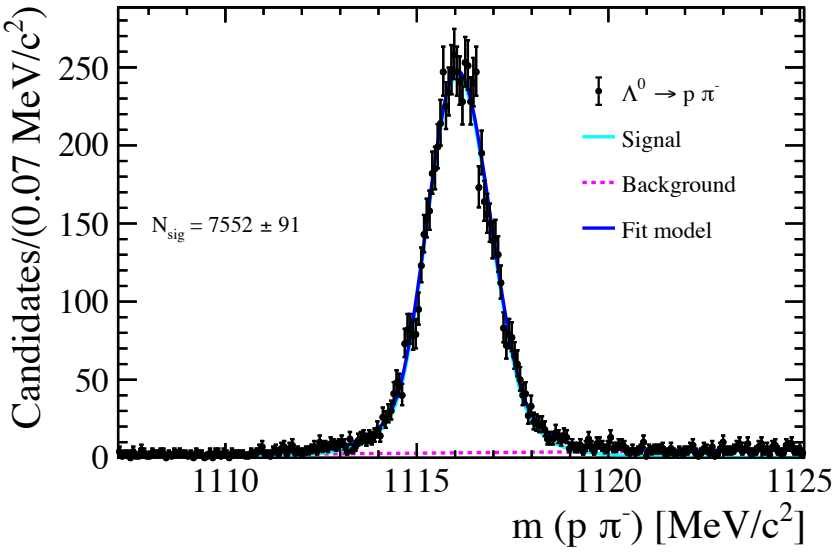


Figure 6.18: Invariant mass spectrum of the $p\pi^-$ final-state of *p*Ar data collected in 2022. The data is overlaid with the result of a fit, where the $\Lambda^0 \to p\pi^-$ signal is modelled with a double-sided crystal ball function and the background background is modelled with a polynomial function.

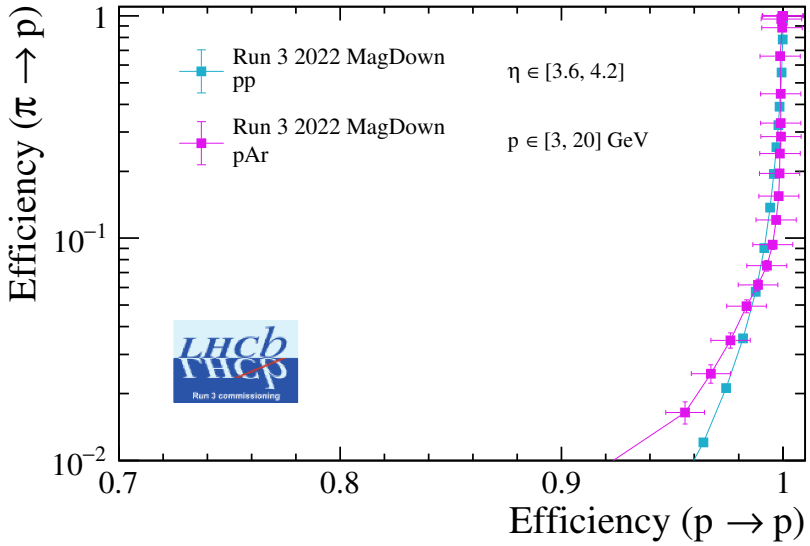


Figure 6.19: The efficiency of selecting protons with the associate leakage from misidentifying pions in the pseudorapidity range $3.6 < \eta < 4.2$ and in the momentum range $3 < p < 20$ GeV for pp and pAr data both collected in 2022. The ranges in pseudorapidity and momentum are chosen to compensate for the lack of kinematic coverage in pAr data.

	Run 3 pAr		Run 3 pp	
PID_P cut	ε $(p \to p)$	misID $(\pi \to p)$	ε $(p \to p)$	misID $(\pi \to p)$
> 0	99 ± 1	20 ± 1	99.7 ± 0.2	25.7 ± 0.2
> 5	99 ± 1	10 ± 1	99.2 ± 0.2	9.0 ± 0.2
> 10	99 ± 1	5 ± 1	98.2 ± 0.2	3.5 ± 0.1

Table 6.11: Efficiency (ε) and misidentification rates for pAr and pp data both collected in 2022 for three different PID cuts (>0, >5, >10), for separations of protons vs. pions (PID_P).

6.6 Early performance with 2023 pp data

In 2023, due to the LHC vacuum incident in the VELO volume, the commissioning of the LHCb detector faced several challenges, including variations in trigger and tracking performance, modifications in the detector acceptance, insufficient coverage in certain sub-detector regions. Despite these challenges, notable achievements were realised, including successful tests of automated constant propagation and the implementation of automatic monitoring with quality checks. These enhancements significantly contributed to improve the tracking, alignment, and calibration processes. Given the substantial impact of these improvements on PID, a detailed study of PID performance for 2023 pp data becomes relevant.

The data samples utilised in this supplementary analysis were collected by the LHCb experiment over a span of approximately 3 days in July 2023. These data correspond to LHC fills 9043 and 9045, with run numbers spanning from 269370 to 269542. The total stored luminosity for this dataset is estimated to be approximately 29 pb^{-1}. Table 6.12 presents a summary of the information pertaining to the data-taking. The detector conditions during this data-taking period are the following:

- the magnet polarity is down;
- the VELO is in open position (49 mm gap);
- the UT is installed but not included in the data taking;

The data from the TURCAL stream undergoes processing with the online conditions version. Similar to the approach outlined in Section 6.4, dedicated HLT2 lines are employed to select samples of $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ and $\Lambda^0 \to p\pi^-$ decays. The selections, detailed in Table 6.5 and Table 6.6, remain consistent. The resulting tuples are generated centrally using the LHCb Analysis Production framework, and the same offline candidate selection, as presented in Tables 6.7 and 6.8, is applied to enhance the quality of the calibration samples.

Considering the differences in the overall detector status from 2022, a careful analysis of the kinematic coverage for 2023, illustrated in Figure 6.20, is necessary. An important distinction in 2023 is the reduced pseudorapidity range for all three charged hadrons, primarily stemming from the reduced acceptance of the VELO in the open position. Additionally, a significant difference pertains to the average event multiplicity, with $< NPVS > \sim 2.9$ in 2022, to be compared

fill ID	Duration	Delivered Luminosity (pb^{-1})	Stored Luminosity (pb^{-1})	Average μ
9045	13:28:08	9.5	5.9	0.57
9044	13:35:12	17.1	12.7	1.07
9043	14:41:54	18.8	10.0	1.07
Total	41:45:15	45.4	28.6	/

Table 6.12: Summary of the pp data collected by the LHCb experiment used to study the preliminary performance in 2022 [198].

with $< NPVS > \sim 1.7$ in 2023. The kinematic coverage remains generally comparable across all three particles. Therefore, no kinematic reweighing is applied in this case as well.

Figure 6.23 shows the efficiency of selecting kaons and protons compared to the misID fraction with pions and kaons, for both 2022 and 2023 data. Notably, a marked improvement is observed in the 2023 performance across all three cases. These enhancements can be attributed to various factors. First, the limited pseudorapidity range for long tracks ($\eta \in [2.0, 4.0]$) in 2023, as opposed to the broader range in 2022 ($\eta \in [2.0, 4.9]$), plays a significant role. Specifically, the analysis of PID as a function of rapidity, presented in Figure 6.11, indicates a degradation in performance within the interval $\eta \in [4.0, 4.9]$. Therefore, the exclusion of this interval due to reduced acceptance contributes to the overall PID performance improvement. Another factor is the lower multiplicity ($N_{PV} \sim 1.7$) in 2023. As demonstrated in Figure 6.13, PID degrades with increasing multiplicity, thus a lower activity in the detector implies an improved performance. Lastly, additional improvements stem from an enhanced alignment and calibration of the overall LHCb detector, including better RICH mirror alignment and improved tracking performance. Table 6.13 provides efficiency and misidentification rates for both 2023 and 2022 data.

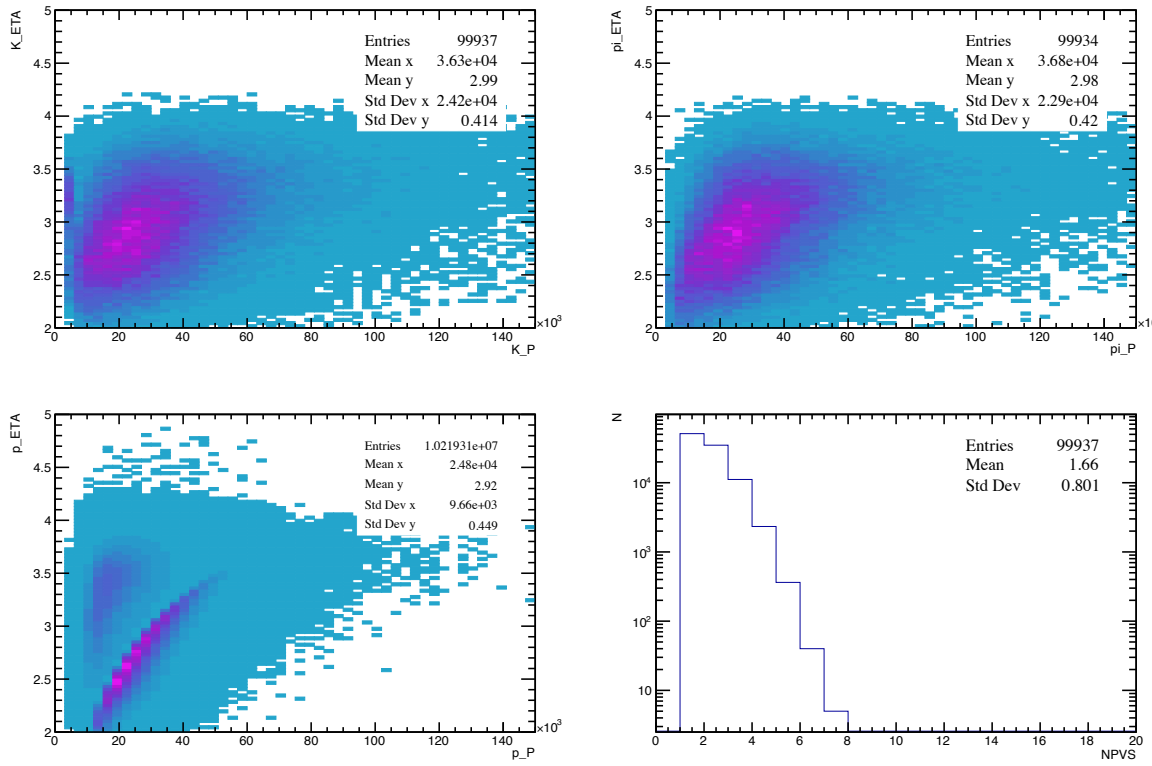
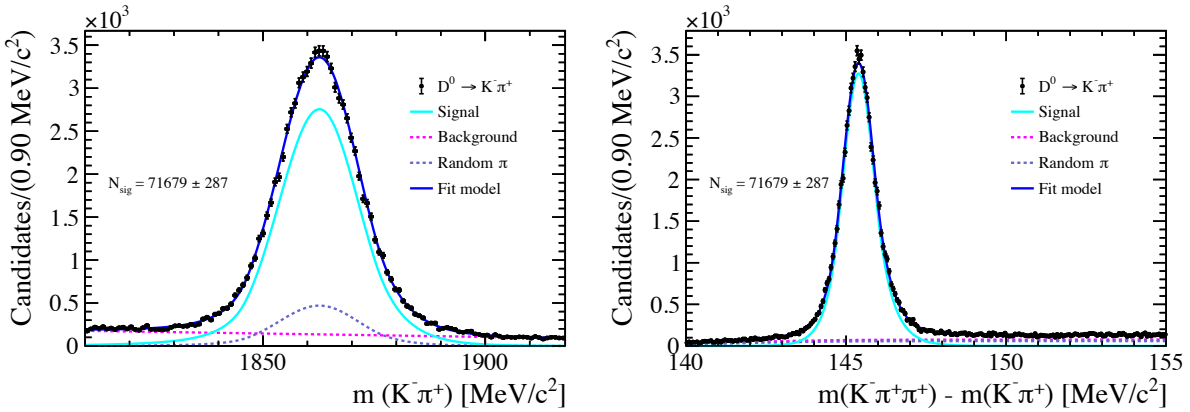


Figure 6.20: Kinematic coverage of pp data colleted in 2023 after the HLT2 and the offline candidate selections. The pseudorapidity versus momentum distribution is plotted for protons, kaons and pions. The distribution of NPVS is also reported, as a proxy for the multiplicity.



(a) (b)

Figure 6.21: Invariant mass spectrum of the $K^-\pi^+$ final-state (a) and distribution of the mass difference $\Delta m = m_{D^{*+}} - m_{D^0}$ distribution (b) for pp data colleted in 2023. The data is overlaid with the result of a 2D fit on the two distributions. In the $K^-\pi^+$ invariant mass spectrum (a), the $D^0 \to K^-\pi^+$ signal is modelled with a double-sided crystal ball function, the random$-\pi$ component is modelled with a Gaussian function and the background is modelled with a polynomial function. In the Δm distribution (b), the $D^0 \to K^-\pi^+$ signal is modelled with a double-sided crystal ball function and the random$-\pi$ component and the background are modelled with a Fermi function.

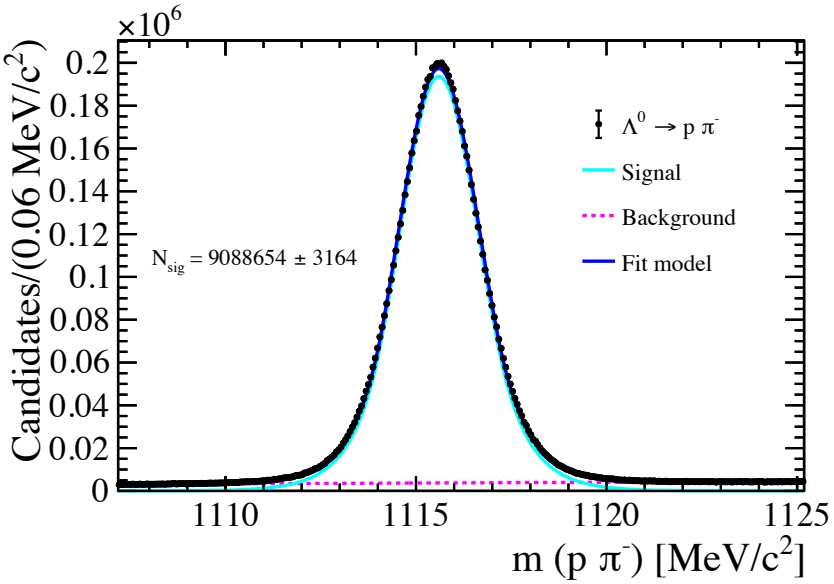


Figure 6.22: Invariant mass spectrum of the $p\pi^-$ final-state for pp data collected in 2023. The data is overlaid with the result of a fit, where the $\Lambda^0 \to p\pi^-$ signal is modelled with a double-sided crystal ball function and the background background is modelled with a polynomial function.

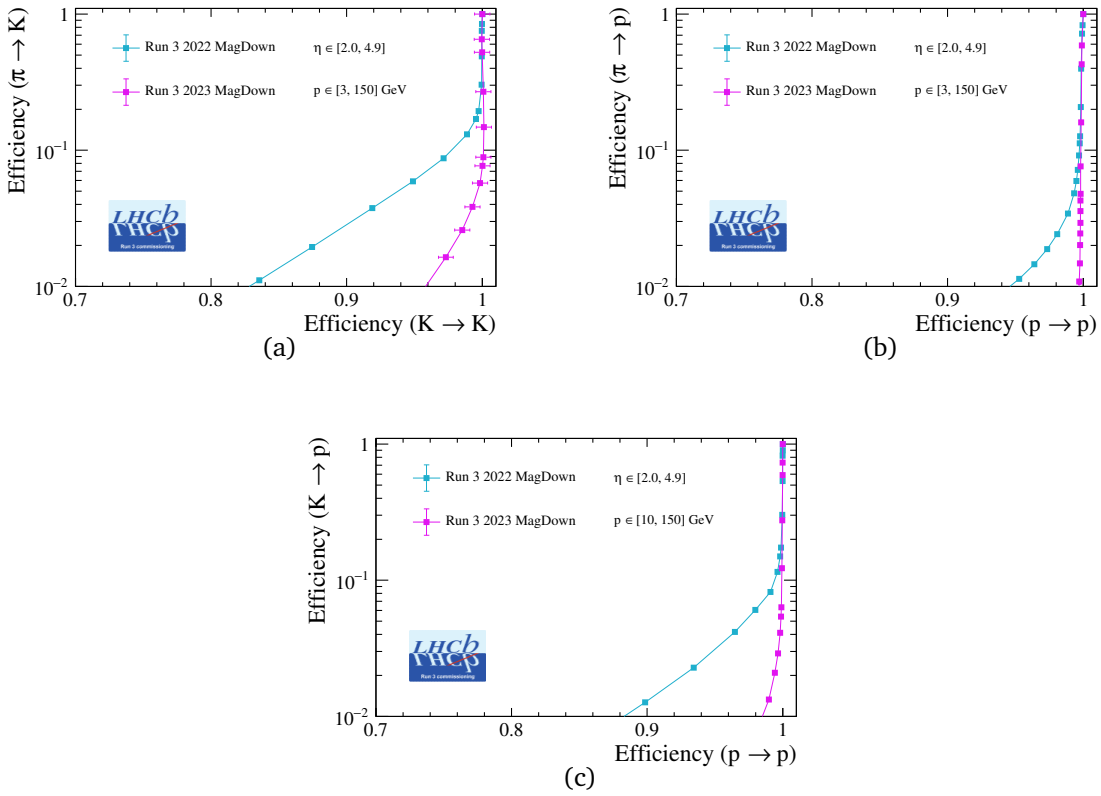


Figure 6.23: The efficiency of selecting (a) kaons, (b) protons, (c) protons, with the associate leakage from misidentifying (a) pions, (b) pions, (c) kaons for pp data collected in 2022 and 2023. The momentum range is $3 < p < 150$ GeV for (a) and (b) and $10 < p < 150$ GeV for (c).

	Run 3 2023		Run 3 2022	
PID_K cut	ε $(K \to K)$	misID $(\pi \to K)$	ε $(K \to K)$	misID $(\pi \to K)$
> 0	98.5 ± 0.1	2.6 ± 0.1	94.9 ± 0.2	5.73 ± 0.04
> 5	95.3 ± 0.1	0.9 ± 0.1	87.5 ± 0.2	1.87 ± 0.02
> 10	92.2 ± 0.1	0.3 ± 0.1	81.2 ± 0.2	0.74 ± 0.02
	Run 3 2023		Run 3 2022	
PID_P cut	ε $(p \to p)$	misID $(\pi \to p)$	ε $(p \to p)$	misID $(\pi \to p)$
> 0	99.7 ± 0.1	2.5 ± 0.1	99.49 ± 0.07	5.94 ± 0.04
> 5	99.7 ± 0.1	1.5 ± 0.1	99.32 ± 0.07	3.42 ± 0.03
> 10	99.7 ± 0.1	1.0 ± 0.1	97.33 ± 0.07	1.88 ± 0.02
	Run 3 2023		Run 3 2022	
PID_{PK} cut	ε $(p \to p)$	misID $(K \to p)$	ε $(p \to p)$	misID $(K \to p)$
> 0	99.4 ± 0.1	2.1 ± 0.1	97.98 ± 0.07	6.06 ± 0.04
> 5	97.9 ± 0.1	0.7 ± 0.1	93.44 ± 0.07	2.28 ± 0.03
> 10	94.9 ± 0.1	0.3 ± 0.1	87.28 ± 0.06	0.86 ± 0.02

Table 6.13: Efficiency (ε) and misidentification rates for pp data collected in 2022 and 2023 for three different PID cuts (>0, >5, >10) for separations of kaons vs. pions (PID_K), protons vs. pions (PID_P), and protons vs. kaons (PID_{PK}).

6.7 Early performance with 2023 PbPb and PbAr data

The heavy-ion physics program also benefits from improved particle identification performance, enabling physics measurements at lower centrality values and ensuring reliable particle identification for more central collisions. Given the elevated particle production in central heavy-ion collisions, it becomes essential to investigate the detector performance under these extreme multiplicity conditions. Figure 6.24 illustrates RICH hitmaps for various centrality values, demonstrating a substantial increase in the occupancy of both RICH1 and RICH2 with more central heavy-ion collisions. As centrality increases, the overlap between Cherenkov rings of different particles intensifies, making pattern recognition challenging in high occupancy regions. In addition, as the occupancy increases, the probability to have multiple hits in the same pixels is larger, causing saturation effects in photon detection. To better understand this effect, it is valuable to generate PID curves as a function of event centrality.

Towards the conclusion of the 2023 operations, LHCb collected PbPb data at $\sqrt{s_{NN}} = 5.76$ TeV and PbAr data at $\sqrt{s_{NN}} = 70.9$ GeV in September and October. A 10% subsample of the PbPb/PbAr dataset has been validated, and a preliminary study of PID performance for this specific data-taking period has been conducted. The detector conditions during this particular data-taking are outlined below:

- the magnet polarity is down;
- the VELO is in open position (49 mm gap);
- the SciFi tracker is in warm conditions at a temperature of about +20°C [199];
- the UT is installed but not included in the data-taking;
- Argon gas is injected "à la SMOG", which means directly into the VELO vessel.

The tuples are generated centrally using the LHCb Analysis Production framework. Samples of identified protons, originating from $\Lambda^0 \to p\pi^-$ decays, are obtained through a dedicated HLT2 line. The same offline candidate selection, outlined in Table 6.8, is applied to enhance the quality of the calibration samples. Similar to the scenario described in Section 6.5, identified kaon samples are not available in this case due to the absence of an unbiased calibration sample for D^0 in PbPb data. Consequently, only results pertaining to the separation of protons and pions are presented.

The distribution of energy deposited in the electromagnetic calorimeter (E_ECAL) is presented in Figure 6.25. This variable is employed to define centrality intervals for producing binned PID curves. Table 6.14 outlines these centrality intervals alongside their corresponding E_ECAL ranges, derived by categorising the distribution into centrality classes. The kinematic coverage of the $\Lambda^0 \to p\pi^-$ sample in PbPb data collected in 2023 is displayed in Figure 6.26. Also in this case, to mitigate the momentum imbalance arising from the two-body decay kinematics, the separation of protons and pions is examined in a reduced momentum range common to both particles, specifically $p \in [3, 40]$ GeV/c.

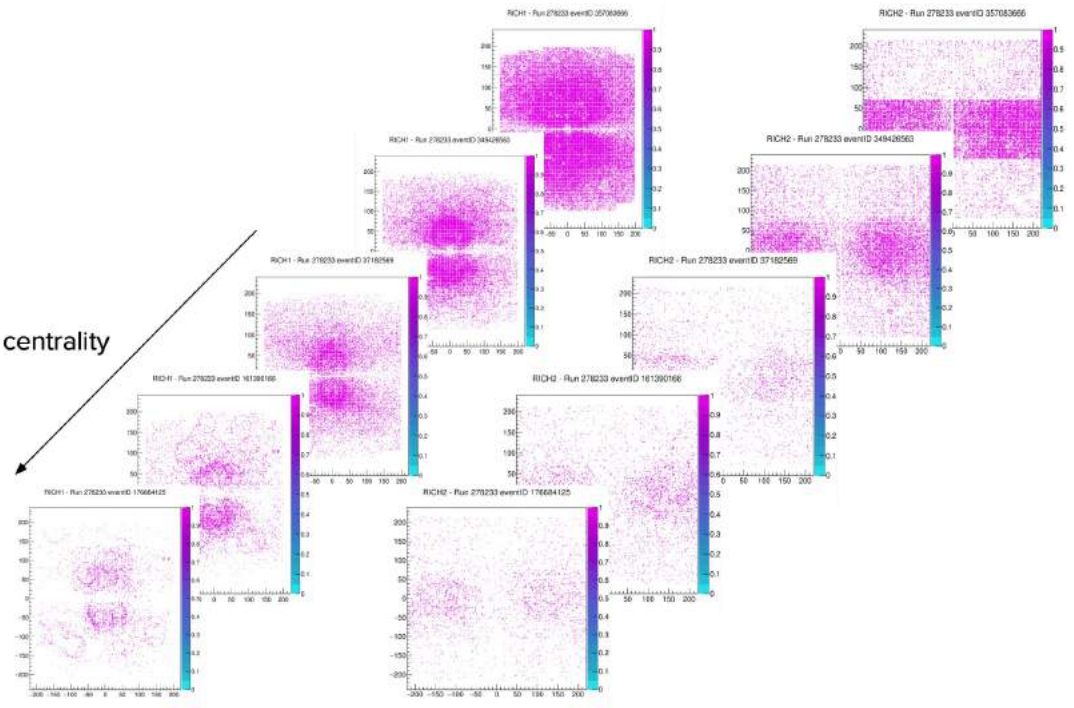


Figure 6.24: Single events is RICH1 and RICH2 in PbPb collisions collected on the 04/10/2023, corresponding to the LHC fill 9223.

Efficiency and misID fractions for protons and pions are determined through a 1D fit on the corresponding Λ^0 invariant mass distribution, presented in Figure 6.27, within the range $1107 < m_{\Lambda^0} < 1125$ MeV/c^2. Figure 6.28 illustrates the efficiency of proton identification with corresponding the misidentification fraction of pions, as a function of the centrality the collision. Remarkably, PID performance is excellent for centrality values ranging from 100% to 70%, indicating the ability of the RICH system to deliver robust particle identification for peripheral PbPb events. This observation is coherent with the results presented in Figure 6.23 for pp data in 2023. However, as collision centrality increases, the performance of PID degrades due to higher occupancy in the RICH detectors. In the case of 50-30% centrality, the performance is observed to be suboptimal.

All these preliminary results are limited by statistics. A more comprehensive analysis will be conducted upon processing the entire PbPb dataset, providing a clearer understanding of the RICH system capability in reconstructing central PbPb events. Despite this limitation, the initial results are promising and suggests that the heavy-ion physics program will benefit from the enhanced PID performance in Run 3 as compared to Run 2.

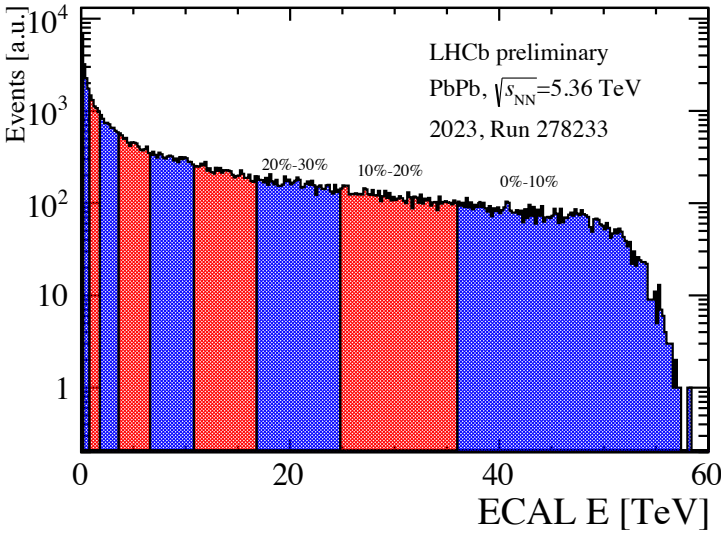


Figure 6.25: Distribution of the total energy collected in the ECAL for PbPb collision events collected in 2023 divided in event activity classes [201].

Centrality	Energy deposited in ECAL (TeV)
100%-70%	[0, 1.8]
70%-60%	[1.8, 3.6]
60%-50%	[3.6, 6.6]
50%-30%	[6.6, 16.8]

Table 6.14: Centrality bins and corresponding energy deposition intervals in the energy deposited in the ECAL for the 2023 PbPb collision data.

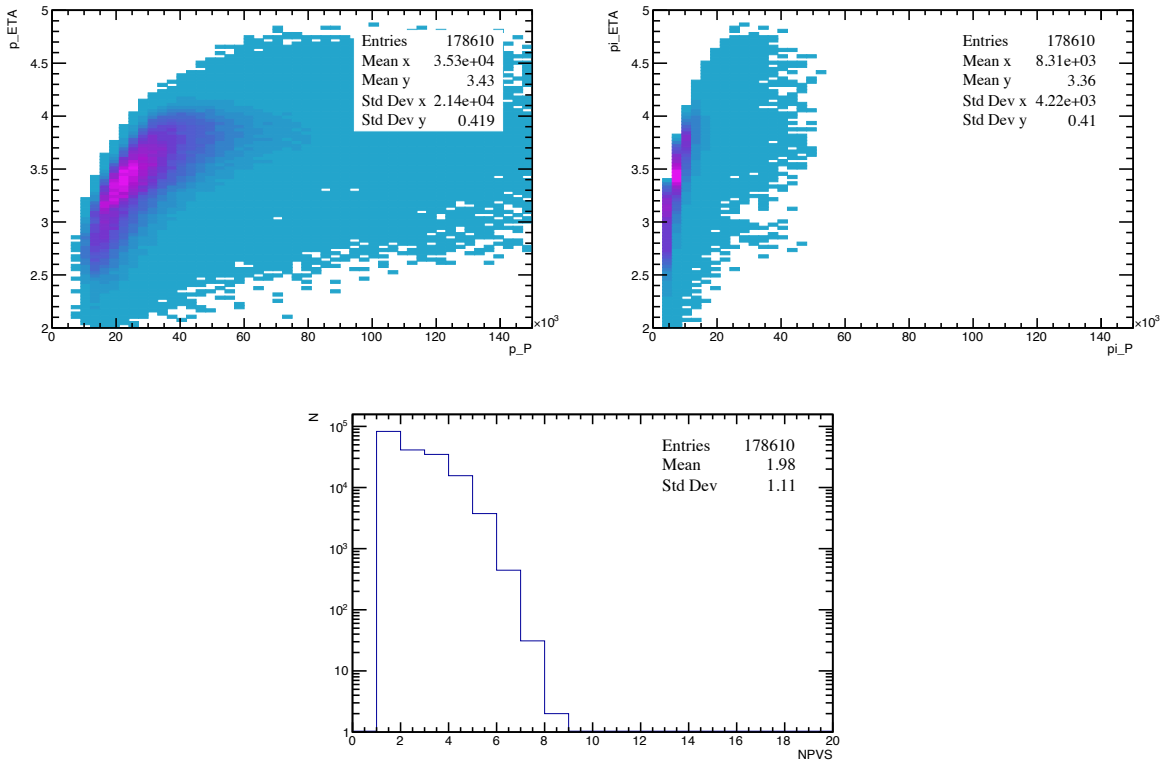


Figure 6.26: Kinematic coverage of PbPb data collected in 2023 after the HLT2 and the offline candidate selections. The pseudorapidity versus momentum distribution is plotted for protons and pions. The distribution of NPVS is also reported, as a proxy for the multiplicity.

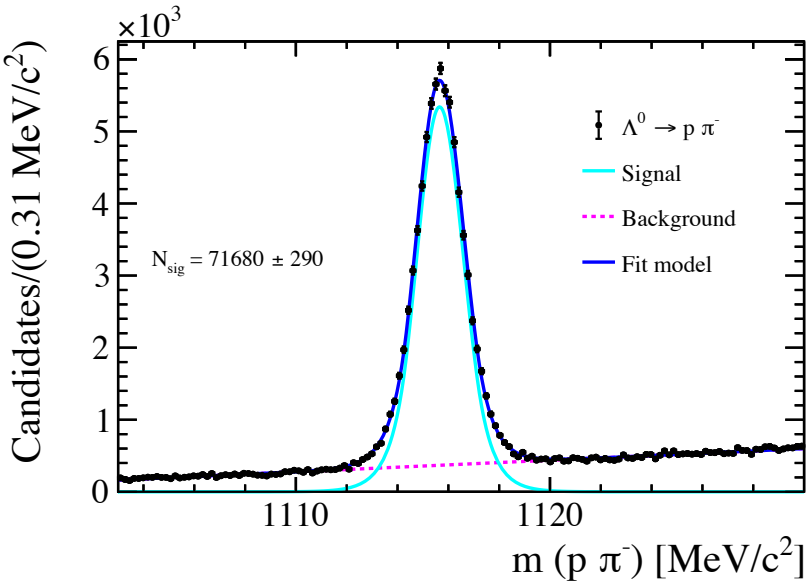


Figure 6.27: Invariant mass spectrum of the $p\pi^-$ final-state for PbPb and PbAr data collected in 2023. The data is overlaid with the result of a fit, where the $\Lambda^0 \to p\pi^-$ signal is modelled with a double-sided crystal ball function and the background background is modelled with a polynomial function.

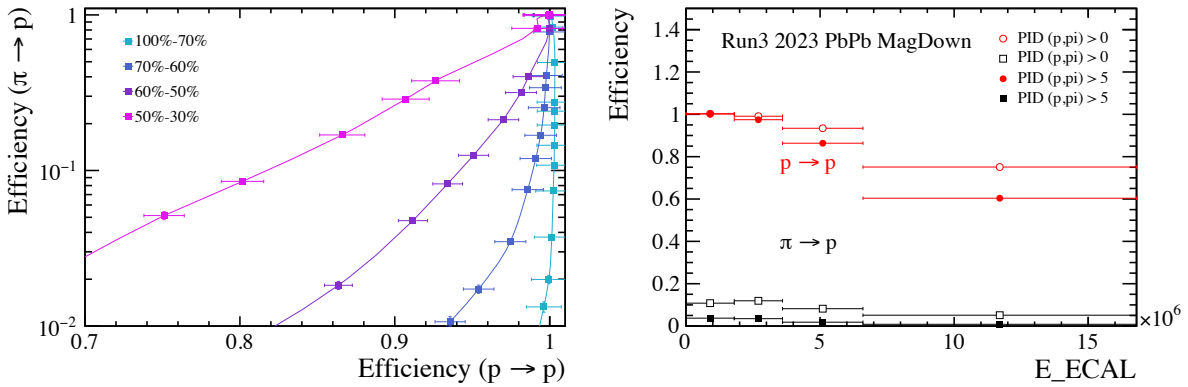


Figure 6.28: The efficiency of selecting protons with the associate leakage from misidentifying pions in the momentum range $3 < p < 40$ GeV/c for PbPb and PbAr data collected in 2023.

	Run 3 PbPb, centrality 80-70%	
PID_P cut	ε $(p \to p)$	misID $(\pi \to p)$
> 0	100 ± 2	12.1 ± 0.4
> 5	100 ± 2	8.2 ± 0.3
> 10	99 ± 2	1.4 ± 0.1
	Run 3 PbPb, centrality 70-60%	
PID_P cut	ε $(p \to p)$	misID $(\pi \to p)$
> 0	99 ± 1	11.9 ± 0.3
> 5	97 ± 1	3.5 ± 0.2
> 10	94 ± 1	1.1 ± 0.1
	Run 3 PbPb, centrality 60-50%	
PID_P cut	ε $(p \to p)$	misID $(\pi \to p)$
> 0	93 ± 1	8.2 ± 0.2
> 5	86 ± 1	1.8 ± 0.2
> 10	74 ± 1	0.3 ± 0.1
	Run 3 PbPb, centrality 50-30%	
PID_P cut	ε $(p \to p)$	misID $(\pi \to p)$
> 0	71.6 ± 2.2	8.2 ± 0.1
> 5	55.9 ± 1.8	1.3 ± 0.2
> 10	38.8 ± 1.4	0.1 ± 0.1

Table 6.15: Efficiency (ε) and misidentification rates for PbPb data collected in 2023 in different centrality intervals, for three different PID cuts (>0, >5, >10), for the separation of protons vs. pions (PID_P)

6.8 Monitoring of the PID performance

The framework introduced in this chapter serves as a valuable tool for an extensive study of the PID performance across various collision systems at LHCb, under different detector conditions and within a wide kinematic range. Although the framework is designed for offline use with high statistics to ensure stability and flexibility, the importance of PID in the LHCb physics programme drives the development of automated and interactive monitoring tools. These tools are crucial for ensuring the proper functioning of the experimental apparatus during data collection and the quality of recorded data. At LHCb, the monitoring is implemented through the Monet web-based application [202].

Currently, there are several tools available to data-taking shifters and experts for monitoring the RICH detectors, including hitmaps, distributions of the number of hits, and the distribution of the Cherenkov angle resolution. While these are the primary tools for experts to identify problems, they do not provide information on the overall PID performance. Hence, it is not trivial to assess differences in PID performance for different data-taking conditions and detector configurations.

In this scenario, online information about PID performance curves can serve as a figure-of-merit to optimise and monitor the performance of the RICH detectors across different configurations. Despite the excellent early performance presented in the previous sections, several parameters (HV, thresholds, time-of-arrival acceptance window of Cherenkov photons) can still be optimised to achieve the ultimate PID performance.

6.8.1 Data-taking time projections

A crucial aspect in the development of monitoring tools involves making educated guesses regarding the precision of the monitored quantity and determining the required computing resources for the monitoring task. Leveraging the production cross-section data of D^* [203] and adopting a highly conservative estimate for the reconstruction efficiency of the $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ channel set at 1%, it is possible to anticipate approximately 6 million events per LHC fill, with a conservative fill duration of 8 hours (expecting around 12 hours). In addition, a substantially larger number of $\Lambda^0 \to p\pi^-$ events can be expected due to the higher cross-section, assuming nominal luminosity. More precise estimates will become available following the restart of data-taking in 2024 with optimal LHCb detector conditions.

To efficiently explore the parameter space, the implementation of PID in online monitoring is crucial for identifying variations between different configurations. Specifically, the plan to test different latching windows in 2023 was not possible due to the LHC incident. This study is now rescheduled for early 2024 and is crucial for determining the absolute scale of improvements for Run 3 (and beyond) simulation tuning. This underscores the significance of developing and incorporating PID performance in online monitoring.

6.8.2 Requirements for PID monitoring

The current strategy for implementing PID in online monitoring is as follows:

- select HLT1 lines specifically for the monitoring task;
- set a routing bit for the collection of the sample;
- once the Alignment finishes running, run the HLT2 reconstruction and selections on the filtered sample;
- run Gaudi Algorithm in the HLT2 control flow, or use Python-based scripts to determine performance;
- display results in Monet.

Understanding the optimal set of HLT1 lines for PID monitoring is crucial for optimising the allocation of machines for each task and for making precise estimates of signal rates. Figure 6.30 illustrates the most active HLT1 lines for the $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ and $\Lambda^0 \to p\pi^-$ processes used in PID studies with 2022 pp data. The corresponding rates are provided in Table 6.16. Not all lines will be used for the monitoring task. Therefore, it is essential to verify whether using only a subset of the HLT1 lines introduces any bias in the PID response. To address this, the PID study was replicated using only the four most frequently triggered lines. The comparison between this configuration and the results presented in Section 6.4 for the separation of pions and kaons is shown in Figure 6.29. No biases are observed when using only these lines, ensuring the reliability of the study with a subset of lines.

Another critical consideration is understanding the optimal minimum input rate for monitoring to ensure sufficient precision. To address this concern, a dedicated study has been conducted. This study focuses on evaluating the precision of efficiency and purity at various signal input rates. The projected precision for two distinct PID cuts $(>0, >5)$ is investigated across different values of the signal event rate. The results for the $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ and $\Lambda^0 \to p\pi^-$ processes are presented in Figure 6.31. Rate computations are based on 1-hour runs, and data points are obtained by executing the complete analysis, including fit and efficiency calculations, on subsets of 2022 data with increasing statistics (e.g., N=10000, 100000, 500000, etc.). It is important to note that no offline selection is applied in this analysis, enabling a better comparison of signal purity directly at the output of HLT1.

Given the various constraints, the currently adopted approach for monitoring PID involves determining only the integrated PID curve. This strategy offers several advantages. Firstly, it simplifies the data storage process by reducing the number of histograms that need to be saved. This streamlined approach enhances the overall robustness of the fit and procedure, contributing to increased stability in the analysis. Another notable benefit is the reduction in input rate, as fewer statistics are required to achieve the desired precision. An investigation into the expected precision reveals that, with a signal rate ranging between 50-60 Hz, a precision level of 0.5-1.0% is expected.

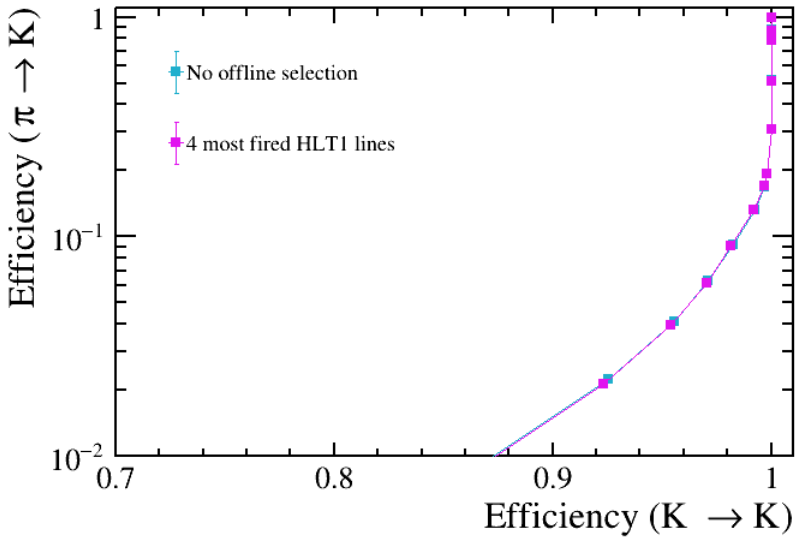


Figure 6.29: The efficiency of selecting kaons, along with the associated leakage from misidentifying pions, is evaluated for the 2022 pp data using both a subset of the most active HLT1 lines and the full dataset.

Calibration Line	Trigger Line	Fraction of Events
$D^{*+} \to D^0(\to K^-\pi^+)\pi^+$	Hlt1TwoTrackMVADecision	0.53
	Hlt1TrackMVADecision	0.29
	Hlt1D2KPiDecision	0.27
	Hlt1SingleHighPtMuonNoMuID	0.16
$\Lambda^0 \to p\pi^-$	Hlt1TwoTrackMVA	0.26
	Hlt1TrackMVA	0.19
	Hlt1SingleHighPtMuonNoMuID	0.17
	Hlt1DiMuonLowMass	0.16
$\Lambda^0 \to p\pi^-$ ($p_T > 3$ GeV/c)	Hlt1TwoTrackMVA	0.74
	Hlt1TrackMVA	0.19
	Hlt1SingleHighPtMuonNoMuID	0.076
	Hlt1DiMuonLowMass	0.057
$\Lambda^0 \to p\pi^-$ ($p_T > 6$ GeV/c)	Hlt1TwoTrackMVA	0.63
	Hlt1TrackMVA	0.21
	Hlt1SingleHighPtMuonNoMuID	0.19
	Hlt1Bs2GammaGamma	0.081

Table 6.16: Most active HLT1 trigger lines used to reproduce the PID study with 2022 data. The corresponding fraction of events of each line is presented.

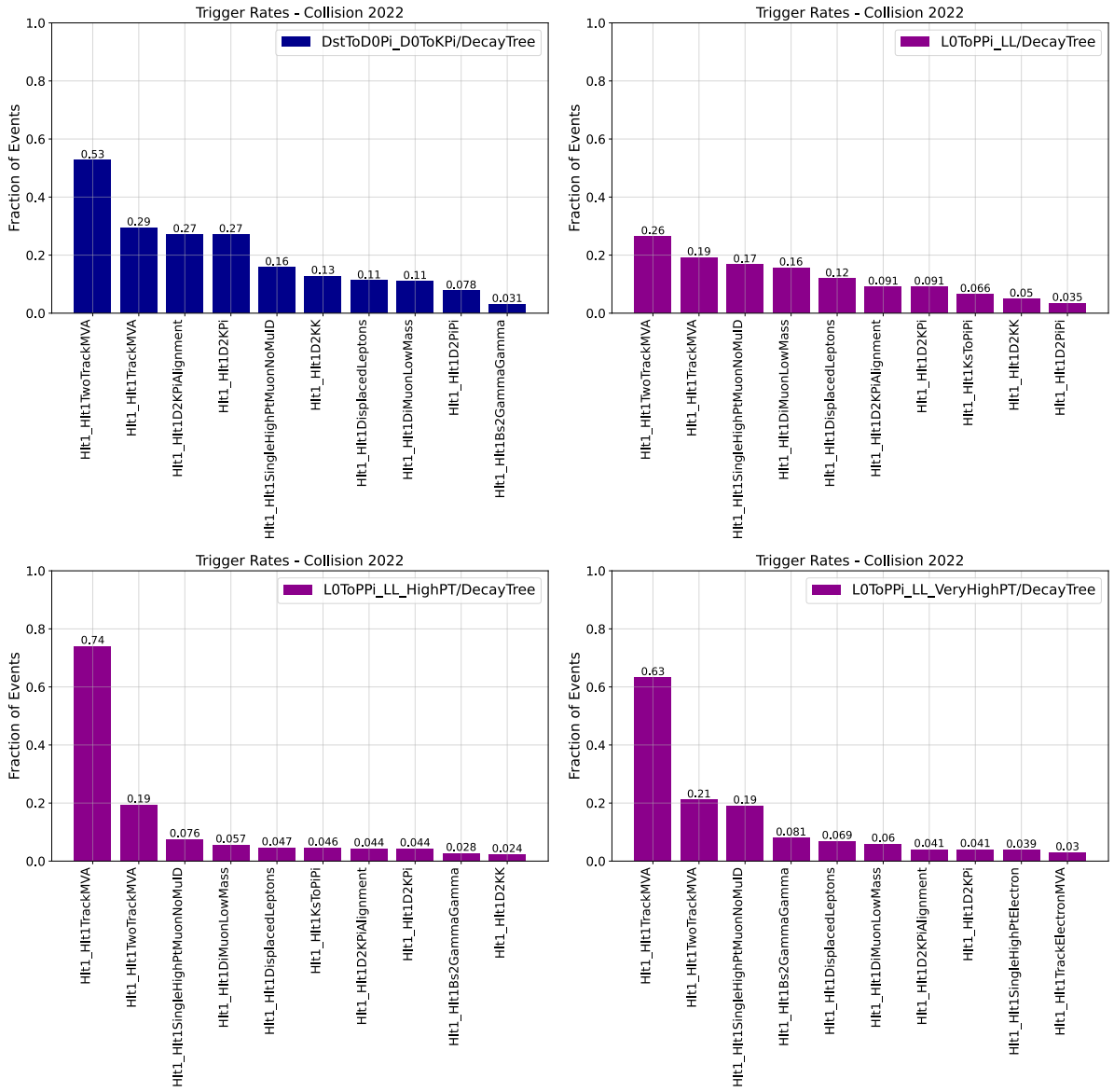
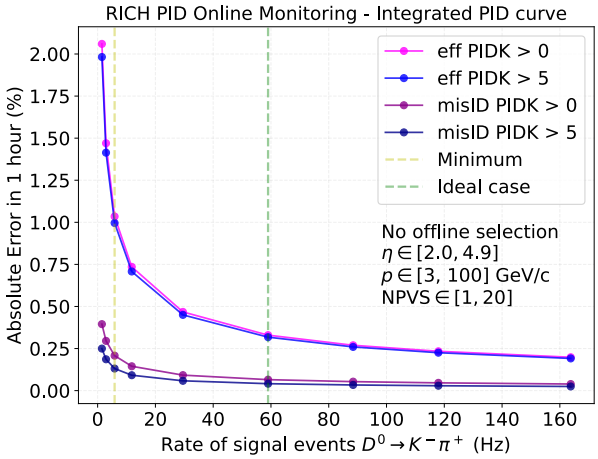
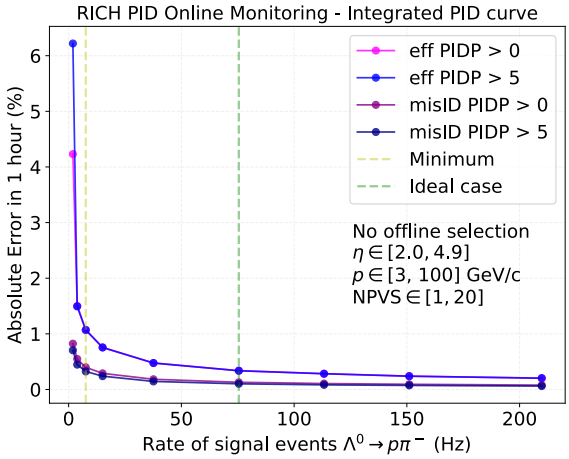


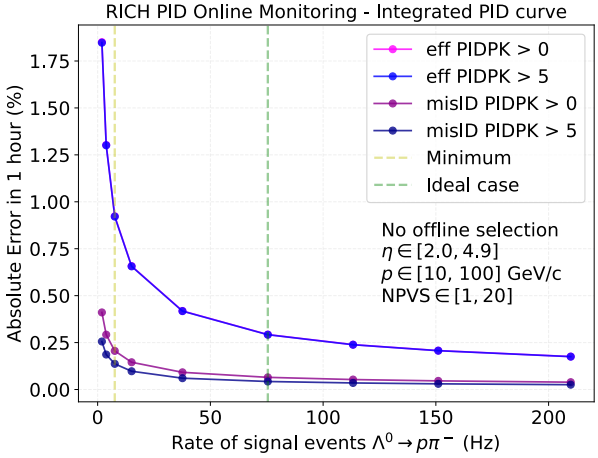
Figure 6.30: Comparison of most active HLT1 lines for $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ and $\Lambda^0 \to p\pi^-$ line. For the Λ^0 line the three lines with different p_T scaling factors are presented.



(a)



(b)



(c)

Figure 6.31: Estimated precision on the efficiency and the misID for the integrated PID curve for two different PID cuts(>0, >5) as a function of the signal rate of $D^0(\to K^-\pi^+)$ for PIDK (a) and of $\Lambda^0 \to p\pi^-$ for PIDP (b) and PIDPK (c). Vertical dashed lines indicate the signal rates to achieve 1% precision (minimum requirement) and less than 0.5% precision (ideal case) on the efficiency. In plot (c) the curve corresponding to PIDPK > 0 is not visible since it lies below the curve corresponding to PIDPK > 5.

Chapter 7

Conclusions

This thesis presents a comprehensive study of Cold Nuclear Matter (CNM) effects in PbNe and pNe collisions, utilising fixed-target data from the LHCb experiment at a center-of-mass energy of $\sqrt{s_{NN}} = 69$ GeV. Understanding charged-particle production in hadronic collisions is crucial for studying Quantum Chromodynamics (QCD), the quantum field theory of the strong interactions. Analysing changes in the production rate of charged particles in proton-nucleus (pA) collisions compared to pp collisions helps model various CNM effects. Experimental data is essential, especially as charged particles from soft interactions dominate, influencing models for both hadron collider and cosmic ray physics.

The analysis aims to detect signals of CNM effects by comparing the production of prompt light charged hadrons in two collision systems, a heavier one (PbNe) and a lighter one (pNe). A selection is applied to choose prompt charged tracks, suppressing background contributions, mainly originating from ghost-charge contamination. Simulation samples validate the candidate selection, and adjustments are made to mitigate discrepancies between simulation and real data. The primary observables consist of single and double ratios of $\pi^{\pm}, K^{\pm}, \overset{(-)}{p}$, studied as functions of transverse momentum, pseudorapidity, and collision centrality. The analysis of single ratios (K/π , p/π , and p/K) in PbNe and pNe collisions reveals intriguing insights into CNM effects. Enhancements are observed in all three ratios in PbNe collisions, with a notable factor of two enhancement in baryon-meson ratios (p/π and p/K) for $p_T > 1$ GeV/c, possibly indicating the onset of the Cronin Effect. A significant dependence on η has been observed in the 2D binning over η and p_T. The analysis includes measurements of double ratios, providing direct insights into particle production differences between PbNe and pNe collisions. All three double ratios show significant deviations from unity, exhibiting a strong p_T dependence. Additionally, the results are compared with corresponding simulations. The single ratios K/π , p/π , and p/K in PbNe data reveal discrepancies with simulation, particularly in the K/π ratio in all p_T bins, slight underestimation in p/π at high p_T, and overestimation in p/K . Similar trends are observed in pNe data, indicating overestimation in p/π and p/K by the simulation. Despite discrepancies, the similar trends in data and simulation validate the analysis. Overall, these observations can provide crucial input for simulation tuning, enhancing the current understanding of CNM effects and related QCD phenomena. One of the primary

challenges of the analysis is related to charged-hadron identification performance. In particular, the kinematic and centrality ranges are limited by the low statistics in PID calibration samples for fixed-target collisions and particle identification performance in Run 2. The LHCb upgrade features an upgraded RICH system that will provide better charged PID performance for pp, fixed-target, and heavy-ion collisions, opening new possibilities for measurements at LHCb in more central events and broader kinematic regions.

Part of this thesis is dedicated to an extensive study of charged-hadron identification performance with early Run 3 data for pp, pAr, PbPb and PbAr collisions collected by LHCb in 2022 and 2023. A software framework, designed to seamlessly integrate into the LHCb data workflow, has been developed to comprehensively study the PID performance. The algorithm is designed as a sequential pipeline, encompassing steps such as data cleaning, exploratory data analysis, fitting and counting, efficiency evaluation, and plotting and comparing resulting PID performance curves. The analysis of 2022 pp data at LHCb indicates promising results regarding the PID performance in Run 3, showing improved PID performance compared to Run 2. The study explores PID dependencies on pseudorapidity, momentum, and event multiplicity, providing insights into the upgraded RICH system's capabilities. Preliminary results suggest that the RICH upgrade achieves its design goals, remarkably enhancing the LHCb charged-hadron identification performance in Run 3. The 2022 pAr collision data shows comparable PID performance with respect to pp collisions for protons-pions separation. Despite statistical limitations, this preliminary analysis suggests that PID behaviour in fixed-target and collider data is similar, indicating minimal impact from the position of the interaction point. While a direct Run 2 comparison is unavailable, the observed improvement in PID for pp collisions in Run 3 suggests a positive impact on fixed-target collisions as well, promising improvements for analyses like the one presented in this thesis. The proton identification efficiency and misidentification fraction of pions have been studied for PbPb and PbAr data, demonstrating robust PID performance in 70-100% centrality class, consistent with trends observed in 2023 pp data. However, performance degrades with increasing centrality, notably in the 30-50% range. Despite limited statistics, these preliminary results are promising, indicating potential benefits for the heavy-ion physics program with enhanced PID performance in Run 3.

Despite the excellent early performance of the RICH systems presented in this thesis, there is room for optimisation to attain the ultimate PID performance. In this regard, online monitoring of the PID performance curves serves as a crucial metric for fine-tuning RICH detectors across various settings. A dedicated section of this work focuses on projections for the implementation of online monitoring for PID performance. Preliminary estimates suggest that 12 hours of data-taking can provide sufficient statistics for the calibration samples, allowing to repeat the PID performance study for different configurations. A study on the expected precision of the efficiency and misID rates at varying signal input rates indicates that a signal rate of 50-60 Hz can achieve a precision level of 0.5-1.0%. Further investigations across various detector configurations can uncover the optimal operational point for the RICH system. Subsequent examinations focused on dedicated physics channels, such as $B \to hh$ and $B \to Dh$ Cabibbo favored and suppressed modes, will allow to validate the absolute scales of PID efficiencies.

Bibliography

[1] S. L. Glashow. "Partial Symmetries of Weak Interactions". In: *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

[2] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (21 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264. URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[3] Abdus Salam. "Weak and Electromagnetic Interactions". In: *Conf. Proc. C* 680519 (1968), pp. 367–377. DOI: 10.1142/9789812795915_0034.

[4] Serguei Chatrchyan et al. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[5] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[6] Roel Aaij et al. "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays". In: *Phys. Rev. Lett.* 115 (2015), p. 072001. DOI: 10.1103/PhysRevLett.115.072001. arXiv: 1507.03414 [hep-ex].

[7] H. Fritzsch, Murray Gell-Mann, and H. Leutwyler. "Advantages of the Color Octet Gluon Picture". In: *Phys. Lett. B* 47 (1973), pp. 365–368. DOI: 10.1016/0370-2693(73)90625-4.

[8] P. A. Zyla et al. "Review of Particle Physics". In: *PTEP* 2020.8 (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104.

[9] O. Kaczmarek et al. "Static quark-antiquark free energy and the running coupling at finite temperature". In: *Phys. Rev. D* 70 (7 2004), p. 074505. DOI: 10.1103/PhysRevD.70.074505. URL: https://link.aps.org/doi/10.1103/PhysRevD.70.074505.

[10] John C. Collins, Davison E. Soper, and George F. Sterman. "Factorization of Hard Processes in QCD". In: *Adv. Ser. Direct. High Energy Phys.* 5 (1989), pp. 1–91. DOI: 10.1142/9789814503266_0001. arXiv: hep-ph/0409313.

[11] V. N. Gribov and L. N. Lipatov. "Deep inelastic e p scattering in perturbation theory". In: *Sov. J. Nucl. Phys.* 15 (1972), pp. 438–450.

[12] Guido Altarelli and G. Parisi. “Asymptotic Freedom in Parton Language”. In: *Nucl. Phys. B* 126 (1977), pp. 298–318. DOI: 10.1016/0550-3213(77)90384-4.

[13] Jun Gao, Lucian Harland-Lang, and Juan Rojo. “The Structure of the Proton in the LHC Precision Era”. In: *Phys. Rept.* 742 (2018), pp. 1–121. DOI: 10.1016/j.physrep.2018.03.002. arXiv: 1709.04922 [hep-ph].

[14] Tie-Jiun Hou et al. “New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC”. In: *Phys. Rev. D* 103.1 (2021), p. 014013. DOI: 10.1103/PhysRevD.103.014013. arXiv: 1912.10053 [hep-ph].

[15] S. M. Berman, J. D. Bjorken, and John B. Kogut. “Inclusive Processes at High Transverse Momentum”. In: *Phys. Rev. D* 4 (1971), p. 3388. DOI: 10.1103/PhysRevD.4.3388.

[16] Andreas Metz and Anselm Vossen. “Parton Fragmentation Functions”. In: *Prog. Part. Nucl. Phys.* 91 (2016), pp. 136–202. DOI: 10.1016/j.ppnp.2016.08.003. arXiv: 1607.02521 [hep-ex].

[17] Peter Braun-Munzinger et al. “Properties of hot and dense matter from relativistic heavy ion collisions”. In: *Phys. Rept.* 621 (2016), pp. 76–126. DOI: 10.1016/j.physrep.2015.12.003. arXiv: 1510.00442 [nucl-th].

[18] Bencédi Gyula. “Study of charged pion, kaon, and (anti)proton production at high transverse momenta in pp and p–Pb collisions with the ALICE experiment at the CERN LHC”. PhD thesis. Eotvos U., 2020.

[19] David d’Enterria. “Jet quenching”. In: *Landolt-Bornstein* 23 (2010). Ed. by R. Stock, p. 471. DOI: 10.1007/978-3-642-01539-7_16. arXiv: 0902.2011 [nucl-ex].

[20] Guang-You Qin and Xin-Nian Wang. “Jet quenching in high-energy heavy-ion collisions”. In: *Int. J. Mod. Phys. E* 24.11 (2015). Ed. by Xin-Nian Wang, p. 1530014. DOI: 10.1142/S0218301315300143. arXiv: 1511.00790 [hep-ph].

[21] David G. d’Enterria. “Hard scattering cross-sections at LHC in the Glauber approach: From pp to pA and AA collisions”. In: (Feb. 2003). arXiv: nucl-ex/0302016.

[22] Georges Aad et al. “Observation of a Centrality-Dependent Dijet Asymmetry in Lead-Lead Collisions at $\sqrt{s_{NN}} = 2.77$ TeV with the ATLAS Detector at the LHC”. In: *Phys. Rev. Lett.* 105 (2010), p. 252303. DOI: 10.1103/PhysRevLett.105.252303. arXiv: 1011.6182 [hep-ex].

[23] K. Aamodt et al. “Suppression of Charged Particle Production at Large Transverse Momentum in Central Pb-Pb Collisions at $\sqrt{s_{NN}} = 2.76$ TeV”. In: *Phys. Lett. B* 696 (2011), pp. 30–39. DOI: 10.1016/j.physletb.2010.12.020. arXiv: 1012.1004 [nucl-ex].

[24] Serguei Chatrchyan et al. “Observation and studies of jet quenching in PbPb collisions at nucleon-nucleon center-of-mass energy = 2.76 TeV”. In: *Phys. Rev. C* 84 (2011), p. 024906. DOI: 10.1103/PhysRevC.84.024906. arXiv: 1102.1957 [nucl-ex].

[25] Betty Abelev et al. "Centrality Dependence of Charged Particle Production at Large Transverse Momentum in Pb–Pb Collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV". In: *Phys. Lett. B* 720 (2013), pp. 52–62. DOI: 10.1016/j.physletb.2013.01.051. arXiv: 1208.2711 [hep-ex].

[26] Jaroslav Adam et al. "Centrality dependence of the nuclear modification factor of charged pions, kaons, and protons in Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV". In: *Phys. Rev. C* 93.3 (2016), p. 034913. DOI: 10.1103/PhysRevC.93.034913. arXiv: 1506.07287 [nucl-ex].

[27] K. Adcox et al. "Suppression of hadrons with large transverse momentum in central Au+Au collisions at $\sqrt{s_{NN}}$ = 130-GeV". In: *Phys. Rev. Lett.* 88 (2002), p. 022301. DOI: 10.1103/PhysRevLett.88.022301. arXiv: nucl-ex/0109003.

[28] S. Acharya et al. "Transverse momentum spectra and nuclear modification factors of charged particles in pp, p-Pb and Pb-Pb collisions at the LHC". In: *JHEP* 11 (2018), p. 013. DOI: 10.1007/JHEP11(2018)013. arXiv: 1802.09145 [nucl-ex].

[29] C. A. Salgado et al. "Proton-Nucleus Collisions at the LHC: Scientific Opportunities and Requirements". In: *J. Phys. G* 39 (2012), p. 015010. DOI: 10.1088/0954-3899/39/1/015010. arXiv: 1105.3919 [hep-ph].

[30] Nestor Armesto. "Nuclear shadowing". In: *J. Phys. G* 32 (2006), R367–R394. DOI: 10.1088/0954-3899/32/11/R01. arXiv: hep-ph/0604108.

[31] Betty Abelev et al. "Transverse momentum distribution and nuclear modification factor of charged particles in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV". In: *Phys. Rev. Lett.* 110.8 (2013), p. 082302. DOI: 10.1103/PhysRevLett.110.082302. arXiv: 1210.4520 [nucl-ex].

[32] Vardan Khachatryan et al. "Charged-particle nuclear modification factors in PbPb and pPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV". In: *JHEP* 04 (2017), p. 039. DOI: 10.1007/JHEP04(2017)039. arXiv: 1611.01664 [nucl-ex].

[33] Vardan Khachatryan et al. "Nuclear Effects on the Transverse Momentum Spectra of Charged Particles in pPb Collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV". In: *Eur. Phys. J. C* 75.5 (2015), p. 237. DOI: 10.1140/epjc/s10052-015-3435-4. arXiv: 1502.05387 [nucl-ex].

[34] Georges Aad et al. "Transverse momentum, rapidity, and centrality dependence of inclusive charged-particle production in $\sqrt{s_{NN}} = 5.02$ TeV p + Pb collisions measured by the ATLAS experiment". In: *Phys. Lett. B* 763 (2016), pp. 313–336. DOI: 10.1016/j.physletb.2016.10.053. arXiv: 1605.06436 [hep-ex].

[35] I. Arsene et al. "On the evolution of the nuclear modification factors with rapidity and centrality in d + Au collisions at s(NN)**(1/2) = 200-GeV". In: *Phys. Rev. Lett.* 93 (2004), p. 242303. DOI: 10.1103/PhysRevLett.93.242303. arXiv: nucl-ex/0403005.

[36] C. Aidala et al. "Nuclear-modification factor of charged hadrons at forward and backward rapidity in p+Al and p+Au collisions at $\sqrt{s_{NN}} = 200$ GeV". In: *Phys. Rev. C* 101.3 (2020), p. 034910. DOI: 10.1103/PhysRevC.101.034910. arXiv: 1906.09928 [hep-ex].

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

[38] 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.

[39] Kari J. Eskola et al. “EPPS16: Nuclear parton distributions with LHC data”. In: *Eur. Phys. J. C* 77.3 (2017), p. 163. DOI: 10.1140/epjc/s10052-017-4725-9. arXiv: 1612.05741 [hep-ph].

[40] Sayipjamal Dulat et al. “New parton distribution functions from a global analysis of quantum chromodynamics”. In: *Phys. Rev. D* 93.3 (2016), p. 033006. DOI: 10.1103/PhysRevD.93.033006. arXiv: 1506.07443 [hep-ph].

[41] 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.

[42] M. Hirai, S. Kumano, and T. H. Nagai. “Nuclear parton distribution functions and their uncertainties”. In: *Phys. Rev. C* 70 (2004), p. 044905. DOI: 10.1103/PhysRevC.70.044905. arXiv: hep-ph/0404093.

[43] D. de Florian and R. Sassot. “Nuclear parton distributions at next-to-leading order”. In: *Phys. Rev. D* 69 (2004), p. 074028. DOI: 10.1103/PhysRevD.69.074028. arXiv: hep-ph/0311227.

[44] Daniel de Florian et al. “Global Analysis of Nuclear Parton Distributions”. In: *Phys. Rev. D* 85 (2012), p. 074028. DOI: 10.1103/PhysRevD.85.074028. arXiv: 1112.6324 [hep-ph].

[45] K. J. Eskola, V. J. Kolhinen, and P. V. Ruuskanen. “Scale evolution of nuclear parton distributions”. In: *Nucl. Phys. B* 535 (1998), pp. 351–371. DOI: 10.1016/S0550-3213(98)00589-6. arXiv: hep-ph/9802350.

[46] K. J. Eskola, V. J. Kolhinen, and C. A. Salgado. “The Scale dependent nuclear effects in parton distributions for practical applications”. In: *Eur. Phys. J. C* 9 (1999), pp. 61–68. DOI: 10.1007/s100520050513. arXiv: hep-ph/9807297.

[47] Kari J Eskola et al. “A global reanalysis of nuclear parton distribution functions”. In: *Journal of High Energy Physics* 2007.05 (2007), p. 002.

[48] M. Hirai, S. Kumano, and T. H. Nagai. “Determination of nuclear parton distribution functions and their uncertainties in next-to-leading order”. In: *Phys. Rev. C* 76 (2007), p. 065207. DOI: 10.1103/PhysRevC.76.065207. arXiv: 0709.3038 [hep-ph].

[49] L. N. Lipatov. “Reggeization of the Vector Meson and the Vacuum Singularity in Nonabelian Gauge Theories”. In: *Sov. J. Nucl. Phys.* 23 (1976), pp. 338–345.

[50] E. A. Kuraev, L. N. Lipatov, and Victor S. Fadin. “The Pomeranchuk Singularity in Nonabelian Gauge Theories”. In: *Sov. Phys. JETP* 45 (1977), pp. 199–204.

[51] I. I. Balitsky and L. N. Lipatov. “The Pomeranchuk Singularity in Quantum Chromodynamics”. In: *Sov. J. Nucl. Phys.* 28 (1978), pp. 822–829.

[52] F. D. Aaron et al. “A Precision Measurement of the Inclusive ep Scattering Cross Section at HERA”. In: *Eur. Phys. J. C* 64 (2009), pp. 561–587. DOI: 10.1140/epjc/s10052-009-1169-x. arXiv: 0904.3513 [hep-ex].

[53] L. V. Gribov, E. M. Levin, and M. G. Ryskin. “Semihard Processes in QCD”. In: *Phys. Rept.* 100 (1983), pp. 1–150. DOI: 10.1016/0370-1573(83)90022-4.

[54] Alfred H. Mueller and Jian-wei Qiu. “Gluon Recombination and Shadowing at Small Values of x”. In: *Nucl. Phys. B* 268 (1986), pp. 427–452. DOI: 10.1016/0550-3213(86)90164-1.

[55] Larry D. McLerran and Raju Venugopalan. “Gluon distribution functions for very large nuclei at small transverse momentum”. In: *Phys. Rev. D* 49 (1994), pp. 3352–3355. DOI: 10.1103/PhysRevD.49.3352. arXiv: hep-ph/9311205.

[56] F. Gelis. “Color Glass Condensate and Glasma”. In: *Int. J. Mod. Phys. A* 28 (2013), p. 1330001. DOI: 10.1142/S0217751X13300019. arXiv: 1211.3327 [hep-ph].

[57] Amir H. Rezaeian. “CGC predictions for p+A collisions at the LHC and signature of QCD saturation”. In: *Phys. Lett. B* 718 (2013), pp. 1058–1069. DOI: 10.1016/j.physletb.2012.11.066. arXiv: 1210.2385 [hep-ph].

[58] Javier L. Albacete et al. “CGC predictions for p + Pb collisions at the LHC”. In: *Nucl. Phys. A* 897 (2013), pp. 1–27. DOI: 10.1016/j.nuclphysa.2012.09.012. arXiv: 1209.2001 [hep-ph].

[59] T. Lappi and H. Mäntysaari. “Single inclusive particle production at high energy from HERA data to proton-nucleus collisions”. In: *Phys. Rev. D* 88 (2013), p. 114020. DOI: 10.1103/PhysRevD.88.114020. arXiv: 1309.6963 [hep-ph].

[60] Francois Gelis et al. “The Color Glass Condensate”. In: *Ann. Rev. Nucl. Part. Sci.* 60 (2010), pp. 463–489. DOI: 10.1146/annurev.nucl.010909.083629. arXiv: 1002.0333 [hep-ph].

[61] D. Antreasyan et al. “Production of Hadrons at Large Transverse Momentum in 200-GeV, 300-GeV and 400-GeV p p and p n Collisions”. In: *Phys. Rev. D* 19 (1979), pp. 764–778. DOI: 10.1103/PhysRevD.19.764.

[62] S. S. Adler et al. “Absence of suppression in particle production at large transverse momentum in S(NN)**(1/2) = 200-GeV d + Au collisions”. In: *Phys. Rev. Lett.* 91 (2003), p. 072303. DOI: 10.1103/PhysRevLett.91.072303. arXiv: nucl-ex/0306021.

[63] B. B. Back et al. “Pseudorapidity dependence of charged hadron transverse momentum spectra in d+Au collisions at s(NN)**(1/2) = 200 GeV”. In: *Phys. Rev. C* 70 (2004), p. 061901. DOI: 10.1103/PhysRevC.70.061901. arXiv: nucl-ex/0406017.

[64] Jaroslav Adam et al. “Multiplicity dependence of charged pion, kaon, and (anti)proton production at large transverse momentum in p-Pb collisions at $\sqrt{s_{\mathrm{NN}}}$ = 5.02 TeV”. In: *Phys. Lett. B* 760 (2016), pp. 720–735. DOI: 10.1016/j.physletb.2016.07.050. arXiv: 1601.03658 [nucl-ex].

[65] A. Adare et al. “Spectra and ratios of identified particles in Au+Au and d+Au collisions at $\sqrt{s_{NN}} = 200$ GeV”. In: *Phys. Rev. C* 88.2 (2013), p. 024906. DOI: 10.1103/PhysRevC.88.024906. arXiv: 1304.3410 [nucl-ex].

[66] Zhong-Bo Kang, Ivan Vitev, and Hongxi Xing. “Multiple scattering effects on inclusive particle production in the large-x regime”. In: *Phys. Rev. D* 88 (2013), p. 054010. DOI: 10.1103/PhysRevD.88.054010. arXiv: 1307.3557 [hep-ph].

[67] Alberto Accardi. “Cronin effect in proton nucleus collisions: A Survey of theoretical models”. In: (Dec. 2002). arXiv: hep-ph/0212148.

[68] Zhong-Bo Kang et al. “Multiple scattering effects on heavy meson production in p+A collisions at backward rapidity”. In: *Phys. Lett. B* 740 (2015), pp. 23–29. DOI: 10.1016/j.physletb.2014.11.024. arXiv: 1409.2494 [hep-ph].

[69] Shreyasi Acharya et al. “Production of muons from heavy-flavour hadron decays in p-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = \mathbf{5.02}$ TeV”. In: *Phys. Lett. B* 770 (2017), pp. 459–472. DOI: 10.1016/j.physletb.2017.03.049. arXiv: 1702.01479 [nucl-ex].

[70] A. Adare et al. “Cold-Nuclear-Matter Effects on Heavy-Quark Production at Forward and Backward Rapidity in d+Au Collisions at $\sqrt{s_{NN}} = 200$ GeV”. In: *Phys. Rev. Lett.* 112.25 (2014), p. 252301. DOI: 10.1103/PhysRevLett.112.252301. arXiv: 1310.1005 [nucl-ex].

[71] Rudolph C. Hwa and C. B. Yang. “Final state interaction as the origin of the Cronin effect”. In: *Phys. Rev. Lett.* 93 (2004), p. 082302. DOI: 10.1103/PhysRevLett.93.082302. arXiv: nucl-th/0403001.

[72] Andy Buckley et al. “General-purpose event generators for LHC physics”. In: *Phys. Rept.* 504 (2011), pp. 145–233. DOI: 10.1016/j.physrep.2011.03.005. arXiv: 1101.2599 [hep-ph].

[73] Ralph Engel, Dieter Heck, and Tanguy Pierog. “Extensive air showers and hadronic interactions at high energy”. In: *Ann. Rev. Nucl. Part. Sci.* 61 (2011), pp. 467–489. DOI: 10.1146/annurev.nucl.012809.104544.

[74] Alexander Aab et al. “The Pierre Auger Cosmic Ray Observatory”. In: *Nucl. Instrum. Meth. A* 798 (2015), pp. 172–213. DOI: 10.1016/j.nima.2015.06.058. arXiv: 1502.01323 [astro-ph.IM].

[75] T. Abu-Zayyad et al. “Evidence for Changing of Cosmic Ray Composition between 10**17-eV and 10**18-eV from Multicomponent Measurements”. In: *Phys. Rev. Lett.* 84 (2000), pp. 4276–4279. DOI: 10.1103/PhysRevLett.84.4276. arXiv: astro-ph/9911144.

[76] T. Antoni et al. “A Nonparametric approach to infer the energy spectrum and the mass composition of cosmic rays”. In: *Astropart. Phys.* 16 (2002), pp. 245–263. DOI: 10.1016/S0927-6505(01)00111-6. arXiv: astro-ph/0102443.

[77] Johannes Albrecht et al. “The Muon Puzzle in cosmic-ray induced air showers and its connection to the Large Hadron Collider”. In: *Astrophys. Space Sci.* 367.3 (2022), p. 27. DOI: 10.1007/s10509-022-04054-5. arXiv: 2105.06148 [astro-ph.HE].

[78] T. Pierog et al. “EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider”. In: *Phys. Rev. C* 92.3 (2015), p. 034906. DOI: 10.1103/PhysRevC.92.034906. arXiv: 1306.0121 [hep-ph].

[79] T. Pierog and K. Werner. “EPOS Model and Ultra High Energy Cosmic Rays”. In: *Nucl. Phys. B Proc. Suppl.* 196 (2009). Ed. by Jean-Noël Capdevielle, Ralph Engel, and Bryan Pattison, pp. 102–105. DOI: 10.1016/j.nuclphysbps.2009.09.017. arXiv: 0905.1198 [hep-ph].

[80] 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.

[81] Klaus Werner, Fu-Ming Liu, and Tanguy Pierog. “Parton ladder splitting and the rapidity dependence of transverse momentum spectra in deuteron-gold collisions at RHIC”. In: *Phys. Rev. C* 74 (2006), p. 044902. DOI: 10.1103/PhysRevC.74.044902. arXiv: hep-ph/0506232.

[82] K. Werner et al. “Event-by-Event Simulation of the Three-Dimensional Hydrodynamic Evolution from Flux Tube Initial Conditions in Ultrarelativistic Heavy Ion Collisions”. In: *Phys. Rev. C* 82 (2010), p. 044904. DOI: 10.1103/PhysRevC.82.044904. arXiv: 1004.0805 [nucl-th].

[83] Klaus Werner. “Core-corona separation in ultra-relativistic heavy ion collisions”. In: *Phys. Rev. Lett.* 98 (2007), p. 152301. DOI: 10.1103/PhysRevLett.98.152301. arXiv: 0704.1270 [nucl-th].

[84] G. Aad et al. “Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC”. In: *New J. Phys.* 13 (2011), p. 053033. DOI: 10.1088/1367-2630/13/5/053033. arXiv: 1012.5104 [hep-ex].

[85] Serguei Chatrchyan et al. “Study of the Inclusive Production of Charged Pions, Kaons, and Protons in pp Collisions at $\sqrt{s} = 0.9$, 2.76, and 7 TeV”. In: *Eur. Phys. J. C* 72 (2012), p. 2164. DOI: 10.1140/epjc/s10052-012-2164-1. arXiv: 1207.4724 [hep-ex].

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

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

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

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

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

[91] Giorgio Apollinari et al. “High luminosity large hadron collider HL-LHC”. In: *arXiv preprint arXiv:1705.08830* (2017).

[92] Christian Elsasser. *$\overline{b}b$ production angle plots*. URL: {\href{https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html}{https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html}}.

[93] Roel Aaij et al. “LHCb Detector Performance”. In: *Int. J. Mod. Phys. A* 30.07 (2015), p. 1530022. DOI: 10.1142/S0217751X15300227. arXiv: 1412.6352 [hep-ex].

[94] “LHCb VELO TDR: Vertex locator. Technical design report”. In: (May 2001).

[95] S. Borghi et al. “First spatial alignment of the LHCb VELO and analysis of beam absorber collision data”. In: *Nucl. Instrum. Meth. A* 618 (2010), pp. 108–120. DOI: 10.1016/j.nima.2010.02.109.

[96] VELO LHCb. “Performance of the lhcb vertex locator”. In: *arXiv preprint arXiv:1405.7808* (2014).

[97] C. Abellan Beteta et al. “Monitoring radiation damage in the LHCb Tracker Turicensis”. In: *JINST* 15.08 (2020), P08016. DOI: 10.1088/1748-0221/15/08/P08016. arXiv: 1809.05063 [physics.ins-det].

[98] “LHCb: Inner tracker technical design report”. In: (Nov. 2002).

[99] “LHCb: Outer tracker technical design report”. In: (Sept. 2001).

[100] R Arink et al. “Performance of the LHCb Outer Tracker”. In: *JINST* 9.01 (2014), P01002. DOI: 10.1088/1748-0221/9/01/P01002. arXiv: 1311.3893 [physics.ins-det].

[101] O Callot and S Hansmann-Menzemer. *The Forward Tracking: Algorithm and Performance Studies*. Tech. rep. Geneva: CERN, 2007. URL: https://cds.cern.ch/record/1033584.

[102] M Needham. *Performance of the Track Matching*. Tech. rep. Geneva: CERN, 2007. URL: https://cds.cern.ch/record/1060807.

[103] Rudolph Emil Kalman. “A new approach to linear filtering and prediction problems”. In: (1960).

[104] E Rodrigues. *Dealing with clones in the tracking*. Tech. rep. Geneva: CERN, 2006. URL: https://cds.cern.ch/record/1000723.

[105] Michel De Cian et al. *Fast neural-net based fake track rejection in the LHCb reconstruction*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2255039.

[106] LHCb Collaboration et al. “Measurement of the track reconstruction efficiency at LHCb”. In: *Journal of Instrumentation* 10.02 (2015), P02007.

[107] LHCb Collaboration et al. “LHCb detector performance”. In: *International Journal of Modern Physics A, 2015, Vol. 30, No. 7* (2015).

[108] Roel Aaij et al. “Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC”. In: *JINST* 14.04 (2019), P04013. DOI: 10.1088/1748-0221/14/04/P04013. arXiv: 1812.10790 [hep-ex].

[109] R Calabrese et al. “Performance of the LHCb RICH detectors during LHC Run 2”. In: *Journal of Instrumentation* 17.07 (2022), P07013.

[110] IETIM Frank and Ig Tamm. “Coherent visible radiation of fast electrons passing through matter”. In: *Selected Papers*. Springer, 1991, pp. 29–35.

[111] R Antunes-Nobrega et al. *LHCb reoptimized detector design and performance: Technical Design Report*. Technical design report. LHCb. Geneva: CERN, 2003. URL: https://cds.cern.ch/record/630827.

[112] M. Alemi et al. “First operation of a hybrid photon detector prototype with electrostatic cross-focussing and integrated silicon pixel readout”. In: *Nucl. Instrum. Meth. A* 449 (2000), pp. 48–59. DOI: 10.1016/S0168-9002(99)01448-5.

[113] M. Adinolfi et al. “Performance of the LHCb RICH detector at the LHC”. In: *Eur. Phys. J. C* 73 (2013), p. 2431. DOI: 10.1140/epjc/s10052-013-2431-9. arXiv: 1211.6759 [physics.ins-det].

[114] Giovanni Cavallero. “Evidence for an $\eta_c(1S)\pi^-$ exotic resonance at LHCb and development of the Experiment Control System for the RICH upgrade”. Presented 26 Mar 2018. Genoa U., 2018. URL: https://cds.cern.ch/record/2647311.

[115] R Aaij et al. “Performance of the LHCb calorimeters, LHCb-DP-2013-004”. In: *preparation.[Cited on page 44.]* ().

[116] C Abellán Beteta et al. “Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC”. In: *arXiv preprint arXiv:2008.11556* (2020).

[117] Eduardo Picatoste Olloqui and (on behalf ofthe LHCb Collaboration). “LHCb Preshower(PS) and Scintillating Pad Detector (SPD): Commissioning, calibration, and monitoring”. In: *Journal of Physics: Conference Series* 160.1 (2009), p. 012046. DOI: 10.1088/1742-6596/160/1/012046. URL: https://dx.doi.org/10.1088/1742-6596/160/1/012046.

[118] A. A. Alves Jr. et al. “Performance of the LHCb muon system”. In: *JINST* 8 (2013), P02022. DOI: 10.1088/1748-0221/8/02/P02022. arXiv: 1211.1346 [physics.ins-det].

[119] R. W. Forty and O. Schneider. “RICH pattern recognition”. In: (Apr. 1998).

[120] Denis Derkach et al. “Machine-Learning-based global particle-identification algorithms at the LHCb experiment”. In: *Journal of Physics: Conference Series*. Vol. 1085. 4. IOP Publishing. 2018, p. 042038.

[121] Andreas Hoecker et al. “TMVA-toolkit for multivariate data analysis”. In: *arXiv preprint physics/0703039* (2007).

[122] Lhcb Collaboration. “PID Plots for Conference”. In: (). URL: https://twiki.cern.ch/twiki/bin/view/LHCb/PIDConferencePlots.

[123] M. Clemencic et al. “The LHCb simulation application, Gauss: Design, evolution and experience”. In: *J. Phys. Conf. Ser.* 331 (2011). Ed. by Simon C. Lin, p. 032023. DOI: 10.1088/1742-6596/331/3/032023.

[124] I. Belyaev et al. “Handling of the generation of primary events in Gauss, the LHCb simulation framework”. In: *J. Phys. Conf. Ser.* 331 (2011). Ed. by Simon C. Lin, p. 032047. DOI: 10.1088/1742-6596/331/3/032047.

[125] D. J. Lange. “The EvtGen particle decay simulation package”. In: *Nucl. Instrum. Meth. A* 462 (2001). Ed. by S. Erhan, P. Schlein, and Y. Rozen, pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4.

[126] John Allison et al. “Geant4 developments and applications”. In: *IEEE Trans. Nucl. Sci.* 53 (2006), p. 270. DOI: 10.1109/TNS.2006.869826.

[127] “Boole”. In: (). URL: http://lhcbdoc.web.cern.ch/lhcbdoc/boole/.

[128] “Moore”. In: (). URL: http://lhcbdoc.web.cern.ch/lhcbdoc/moore/.

[129] “Brunel”. In: (). URL: http://lhcbdoc.web.cern.ch/lhcbdoc/brunel/.

[130] “DaVinci”. In: (). URL: http://lhcbdoc.web.cern.ch/lhcbdoc/davinci/.

[131] Lhcb Collaboration. “LHCb starterkit”. In: (). URL: https://lhcb.github.io/starterkit-lessons/first-analysis-steps/dataflow.html.

[132] “LHCb SMOG Upgrade”. In: (). DOI: 10.17181/CERN.SAQC.EOWH.

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

[134] Roel Aaij et al. “Precision luminosity measurements at LHCb”. In: *JINST* 9.12 (2014), P12005. DOI: 10.1088/1748-0221/9/12/P12005. arXiv: 1410.0149 [hep-ex].

[135] M. Ferro-Luzzi. “Proposal for an absolute luminosity determination in colliding beam experiments using vertex detection of beam-gas interactions”. In: *Nucl. Instrum. Meth. A* 553 (2005), pp. 388–399. DOI: 10.1016/j.nima.2005.07.010.

[136] Albert Bursche et al. *Physics opportunities with the fixed-target program of the LHCb experiment using an unpolarized gas target*. Tech. rep. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2649878.

[137] C Hadjidakis et al. “A fixed-target programme at the LHC: physics case and projected performances for heavy-ion, hadron, spin and astroparticle studies”. In: *Physics Reports* 911 (2021), pp. 1–83.

[138] Giacomo Graziani. “Results on heavy ion physics at LHCb”. In: *J. Phys. Conf. Ser.* 1271.1 (2019), p. 012008. DOI: 10.1088/1742-6596/1271/1/012008. arXiv: 1904.04130 [hep-ex].

[139] Roel Aaij et al. “Measurement of Antiproton Production in pHe Collisions at $\sqrt{s_{NN}} = 110$ GeV”. In: *Phys. Rev. Lett.* 121.22 (2018), p. 222001. DOI: 10.1103/PhysRevLett.121.222001. arXiv: 1808.06127 [hep-ex].

[140] Roel Aaij et al. "First Measurement of Charm Production in its Fixed-Target Configuration at the LHC". In: *Phys. Rev. Lett.* 122.13 (2019), p. 132002. DOI: 10.1103/PhysRevLett.122.132002. arXiv: 1810.07907 [hep-ex].

[141] Roel Aaij et al. "Open charm production and asymmetry in pNe collisions at $\sqrt{s_{\rm NN}}$ = 68.5 GeV". In: *Eur. Phys. J. C* 83.6 (2023), p. 541. DOI: 10.1140/epjc/s10052-023-11641-5. arXiv: 2211.11633 [hep-ex].

[142] R. Aaij et al. "J/ψ and D^0 production in $\sqrt{s_{\rm NN}}$ = 68.5 GeV PbNe collisions". In: *Eur. Phys. J. C* 83.7 (2023), p. 658. DOI: 10.1140/epjc/s10052-023-11674-w. arXiv: 2211.11652 [hep-ex].

[143] R. Aaij et al. "Measurement of antiproton production from antihyperon decays in pHe collisions at $\sqrt{s_{NN}}$ = 110 GeV". In: *Eur. Phys. J. C* 83.6 (2023), p. 543. DOI: 10.1140/epjc/s10052-023-11673-x. arXiv: 2205.09009 [hep-ex].

[144] Roel Aaij et al. "Centrality determination in heavy-ion collisions with the LHCb detector". In: *JINST* 17.05 (2022), P05009. DOI: 10.1088/1748-0221/17/05/P05009. arXiv: 2111.01607 [nucl-ex].

[145] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "A Brief Introduction to PYTHIA 8.1". In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[146] Giulio Dujany and Barbara Storaci. "Real-time alignment and calibration of the LHCb Detector in Run II". In: *J. Phys. Conf. Ser.* 664.8 (2015), p. 082010. DOI: 10.1088/1742-6596/664/8/082010.

[147] S. Agostinelli et al. "GEANT4–a simulation toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[148] Frederic Fleuret et al. *Global event cuts for 2018 PbNe SMOG data*. Tech. rep. Geneva: CERN, 2021. URL: https://cds.cern.ch/record/2765569.

[149] "The ALICE definition of primary particles". In: (2017). URL: https://cds.cern.ch/record/2270008.

[150] A. Rogozhnikov. "Reweighting with Boosted Decision Trees". In: *J. Phys. Conf. Ser.* 762.1 (2016). Ed. by Luis Salinas and Claudio Torres, p. 012036. DOI: 10.1088/1742-6596/762/1/012036. arXiv: 1608.05806 [physics.data-an].

[151] Daniela Witten and Gareth James. *An introduction to statistical learning with applications in R*. springer publication, 2013.

[152] "Definition of Reconstructible and Reconstructed". In: (). URL: https://lhcb-comp.web.cern.ch/analysis/davinci/v8/recrecdefinition.htm.

[153] 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.

[154] Muriel Pivk and Francois R. Le Diberder. "SPlot: A Statistical tool to unfold data distributions". In: *Nucl. Instrum. Meth. A* 555 (2005), pp. 356–369. DOI: 10.1016/j.nima.2005.08.106. arXiv: physics/0402083.

[155] Maxim Borisyak and Nikita Kazeev. "Machine Learning on data with sPlot background subtraction". In: *JINST* 14.08 (2019), P08020. DOI: 10.1088/1748-0221/14/08/P08020. arXiv: 1905.11719 [cs.LG].

[156] Christoph Langenbruch. "Parameter uncertainties in weighted unbinned maximum likelihood fits". In: *Eur. Phys. J. C* 82.5 (2022), p. 393. DOI: 10.1140/epjc/s10052-022-10254-8. arXiv: 1911.01303 [physics.data-an].

[157] Yuehong Xie. "sFit: a method for background subtraction in maximum likelihood fit". In: *arXiv preprint arXiv:0905.0724* (2009).

[158] Fabian Pedregosa et al. "Scikit-learn: Machine Learn-ing in Python, Journal of Machine Learning Re-search, 12". In: (2011).

[159] R. W. Thompson et al. "The Disintegration of V^0 Particles". In: *Phys. Rev.* 90 (2 1953), pp. 329–330. DOI: 10.1103/PhysRev.90.329. URL: https://link.aps.org/doi/10.1103/PhysRev.90.329.

[160] Glen Cowan. *Statistical data analysis*. Oxford university press, 1998.

[161] "Kolmogorov-Smirnov Goodness-of-Fit Test". In: (). URL: https://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm.

[162] "ROOT TH2 Class Reference". In: (). URL: https://root.cern.ch/doc/master/classTH2.html#a14e58513e24f48726edda3010017fd5a.

[163] R. Aaij et al. "Implications of LHCb measurements and future prospects". In: *Eur. Phys. J. C* 73.4 (2013), p. 2373. DOI: 10.1140/epjc/s10052-013-2373-2. arXiv: 1208.3355 [hep-ex].

[164] C. Fitzpatrick and V. V. Gligorov. "Anatomy of an upgrade event in the upgrade era, and implications for the LHCb trigger". In: (2014).

[165] "LHCb Trigger and Online Upgrade Technical Design Report". In: (May 2014).

[166] Roel Aaij et al. "The LHCb upgrade I". In: (May 2023). arXiv: 2305.10515 [hep-ex].

[167] I. Bediaga. "LHCb VELO Upgrade Technical Design Report". In: (Nov. 2013).

[168] Thomas Bird. *Flavour studies with LHCb: b-meson mixing, lepton-flavour violation and the velo upgrade*. The University of Manchester (United Kingdom), 2016.

[169] "LHCb Tracker Upgrade Technical Design Report". In: (Feb. 2014).

[170] Ana Barbara Rodriguez Cavalcante et al. "Irradiation test of mirror samples for the LHCb SciFi tracker". In: (2016).

[171] C. Alfieri et al. "An experimental set-up to measure Light Yield of Scintillating Fibres". In: (2015).

[172] C. D'Ambrosio et al. "The Future of RICH Detectors through the Light of the LHCb RICH". In: *Nucl. Instrum. Meth. A* 876 (2017). Ed. by P. Krizan et al., pp. 194–197. DOI: 10.1016/j.nima.2017.02.076. arXiv: 1703.09927 [physics.ins-det].

[173] M. K. Baszczyk et al. "Test of the photon detection system for the LHCb RICH Upgrade in a charged particle beam". In: *JINST* 12.01 (2017), P01012. DOI: 10.1088/1748-0221/12/01/P01012. arXiv: 1610.02879 [physics.ins-det].

[174] A. Papanestis, F. Keizer, and S. A. Wotton. "The upgrade of the LHCb RICH system for the LHC Run 3". In: *JINST* 15.09 (2020), p. C09022. DOI: 10.1088/1748-0221/15/09/C09022.

[175] Mirco Andreotti et al. "Characterisation of signal-induced noise in Hamamatsu R11265 Multianode Photomultiplier Tubes". In: *JINST* 16.11 (2021), P11030. DOI: 10.1088/1748-0221/16/11/P11030. arXiv: 2110.00831 [physics.ins-det].

[176] "LHCb PID Upgrade Technical Design Report". In: (Nov. 2013).

[177] "RTA and DPA dataflow diagrams for Run 1, Run 2, and the upgraded LHCb detector ". In: (2020). URL: https://cds.cern.ch/record/2730181.

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

[179] *Computing Model of the Upgrade LHCb experiment*. Tech. rep. Geneva: CERN, 2018. DOI: 10.17181/CERN.Q0P4.57ON. URL: https://cds.cern.ch/record/2319756.

[180] Erhard Steffens and Willy Haeberli. "Polarized gas targets". In: *Reports on Progress in Physics* 66.11 (2003), R02.

[181] Caterina Boscolo Meneguolo et al. "Calculation of the allowed aperture for a gas storage cell in IP8". In: (2018). URL: https://cds.cern.ch/record/2651289.

[182] Branko Kosta Popovic and Christine Vollinger. "Measurement and Simulation of the Longitudinal Impedance of the LHCb VELO". In: (2019). URL: https://cds.cern.ch/record/2701362.

[183] C. Yin Vallgren et al. "Amorphous carbon coatings for the mitigation of electron cloud in the CERN Super Proton Synchrotron". In: *Phys. Rev. ST Accel. Beams* 14 (2011), p. 071001. DOI: 10.1103/PhysRevSTAB.14.071001.

[184] Noemi Caraban Gonzalez. "SMOG2 installation in the LHCb cavern". In: (2020). General Photo. URL: https://cds.cern.ch/record/2727007.

[185] Albert Bursche et al. "Physics opportunities with the fixed-target program of the LHCb experiment using an unpolarized gas target". In: (2018).

[186] "Charged hadron identification performance with early Run 3 data". In: (2023). URL: https://cds.cern.ch/record/2868904.

[187] R Aaij et al. "Measurement of b-hadron branching fractions for two-body decays into charmless charged hadrons". In: *JHEP* 10 (2012), p. 037. DOI: 10.1007/JHEP10(2012)037. arXiv: 1206.2794 [hep-ex].

[188] R. Aaij et al. "Opposite-side flavour tagging of B mesons at the LHCb experiment". In: *Eur. Phys. J. C* 72 (2012), p. 2022. DOI: 10.1140/epjc/s10052-012-2022-1. arXiv: 1202.4979 [hep-ex].

[189] Roel Aaij et al. "First measurement of the differential branching fraction and *CP* asymmetry of the $B^{\pm} \to \pi^{\pm}\mu^{+}\mu^{-}$ decay". In: *JHEP* 10 (2015), p. 034. DOI: 10.1007/JHEP10(2015)034. arXiv: 1509.00414 [hep-ex].

[190] Roel Aaij et al. "Study of prompt D^0 meson production in pPb collisions at $\sqrt{s_{NN}} = 5$ TeV". In: *JHEP* 10 (2017), p. 090. DOI: 10.1007/JHEP10(2017)090. arXiv: 1707.02750 [hep-ex].

[191] 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 Tech. Instrum.* 6.1 (2019), p. 1. DOI: 10.1140/epjti/s40485-019-0050-z. arXiv: 1803.00824 [hep-ex].

[192] "RICH Performance Plots - Cherenkov Angle Resolutions for RICH 1 and RICH 2". In: (2023). URL: https://cds.cern.ch/record/2859118.

[193] *Python interface: PyROOT*. URL: https://root.cern/manual/python/.

[194] *Root Report Class Reference*. URL: https://pypi.org/project/root-report/.

[195] *LHCb Analysis Productions Documentation*. URL: https://lhcb-ap.docs.cern.ch.

[196] *RDataFrame Class Reference*. URL: https://root.cern/doc/master/classROOT_1_1RDataFrame.html.

[197] 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 Tech. Instrum.* 6.1 (2019), p. 1. DOI: 10.1140/epjti/s40485-019-0050-z. arXiv: 1803.00824. URL: https://cds.cern.ch/record/2308409.

[198] *Run Database*. URL: https://lbrundb.cern.ch.

[199] "Dark count rate measurement for SciFi - LHCb commissioning 2022". In: (2023). URL: https://cds.cern.ch/record/2856208.

[200] "Invariant mass spectra from SMOG2 pAr and pH collisions from 2022 data". In: (2023). URL: https://cds.cern.ch/record/2859158.

[201] " First invariant mass spectra and performance figures of the 2023 ion run ". In: (2023). URL: https://cds.cern.ch/record/2883088.

[202] M. Adinolfi et al. "LHCb data quality monitoring". In: *J. Phys. Conf. Ser.* 898.9 (2017). Ed. by Richard Mount and Craig Tull, p. 092027. DOI: 10.1088/1742-6596/898/9/092027.

[203] Roel Aaij et al. "Measurements of prompt charm production cross-sections in pp collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 03 (2016). [Erratum: JHEP 09, 013 (2016), Erratum: JHEP 05, 074 (2017)], p. 159. DOI: 10.1007/JHEP03(2016)159. arXiv: 1510.01707 [hep-ex].