INTERNATIONAL DOCTORAL SCHOOL OF THE USC

Sara
Sellam

PhD Thesis

Measurement of the production and nuclear modification factor of prompt charged pions, charged kaons and protons in proton-proton and proton-lead collisions at the LHCb experiment.

Santiago de Compostela, 2025

Doctoral Programme in Nuclear and Particle Physics

ESCOLA DE DOUTORAMENTO
INTERNACIONAL DA USC

DOCTORAL THESIS

Measurement of the production and nuclear modification factor of prompt charged pions, charged kaons and protons in proton-proton and proton-lead collisions at the LHCb experiment.

Author

Sara Sellam

Supervisor/s: Cibrán Santamarina Ríos, Elena González Ferreiro

Tutor: Elena González Ferreiro

PHD PROGRAMME IN NUCLEAR AND PARTICLE PHYSICS

SANTIAGO DE COMPOSTELA – 2025

ESCOLA DE DOUTORAMENTO
INTERNACIONAL DA USC

DECLARACIÓN DE CONFLITO DE INTERESE

A autora desta tese declara non ter **ningún conflito de interese** real, potencial ou aparente en relación coa elaboración, contido, análise de datos ou conclusións presentadas no presente traballo.

Sara Sellam

Santiago de Compostela, Outubro de 2025

Firmado por SELLAM SARA - ****5272* el día
08/10/2025 con
un certificado emitido por AC FNMT Usuarios

Acknowledgements

This thesis is the result of not only my own work but also the generosity, guidance, and support of many extraordinary people.

First and foremost, I am deeply grateful to my supervisors, Cibrán and Elena, for their mentorship, insight, and encouragement. Their guidance was invaluable at every stage of this journey.

I thank Abraham, Juan, Antonio, and Sam for generously sharing their expertise, and Adlene, who first introduced me to the LHCb collaboration in Algeria, setting me on this path.

Warm thanks go to my colleagues and friends Óscar, Imanol, Praveen, Manpreet, Luigi, Lidia, Clara, Cris, Victoria, and Iván. Your collaboration, discussions, and shared moments made the daily work both productive and enjoyable. I am also grateful to the entire LHCb collaboration and to everyone at IGFAE, whose support and collaborative spirit enriched this experience both scientifically and personally.

I would also like to remember my colleague Arnau, who had promised to attend my thesis defense even if he could not be in Santiago; you will be deeply missed.

Beyond academia, I have been fortunate to find a second family in Spain and beyond. I thank my Palestinian family in Paris for their generosity and kindness, my friend Theraa for her warm hospitality, and Rahima, whose encouragement from afar bridged the distance between Brazil and Spain.

To my friends in Santiago, Djamila, Paula, Iria, Lucía, Liliana, and María, Alex and to all my friends outside physics, your laughter, support, and companionship turned cultural and language barriers into opportunities for connection and community.

To my parents, thank you for your unwavering love, encouragement, and sacrifices, which made my higher education possible. To my brothers, though I could only watch from afar, thank you for your humor and constant support.

This thesis reflects not only my own work, but also the efforts, kindness, and encouragement of many others. To all of you: thank you.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

And finally, to the girl who left Algeria in 2020 with nothing but dreams and determination, you made it against all odds.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Abstract

This thesis presents a measurement of identified charged pion ($\pi^{\pm}$), kaon ($K^{\pm}$), and proton ($p^{\pm}$) production rates in high-energy proton-proton (pp) and proton-lead (pPb) collisions at a centre-of-mass energy of $\sqrt{s_{NN}} = 5.02$ TeV, using data collected by the LHCb experiment at CERN. These measurements aim to provide insights that help isolate Cold Nuclear Matter (CNM) effects in pPb collisions from possible Quark-Gluon Plasma (QGP) signatures in lead-lead (PbPb) collisions. They also offer new constraints on the mass ordering of the Cronin enhancement and contribute to improving our understanding of gluon fragmentation function

Double-differential cross-sections of the identified particles are determined in both systems as functions of pseudorapidity in the nucleon-nucleon centre-of-mass frame (η_{cms}) and transverse momentum (p_{T}) with respect to the proton beam. Using these measured cross-sections, the nuclear modification factor is extracted for charged pions, kaons, and protons within the acceptance ranges $-4.8 < \eta_{\text{cms}} < -2.5$ (backward region) and $2.0 < \eta_{\text{cms}} < 4.8$ (forward region), and for transverse momenta in the range $0.5 < p_{\text{T}} < 5\,\text{GeV}/c$.

In the forward region, all three species exhibit a similar suppression relative to proton-proton collisions, consistent with initial-state effects such as shadowing or gluon saturation. In contrast, in the backward region, while pions and kaons show modest enhancement or values near unity, protons display a pronounced enhancement at higher p_{T}, pointing toward additional contributions from final-state effects such as baryon transport or radial flow. These results provide valuable constraints on nuclear parton distribution functions and offer sensitivity to gluon saturation at parton momentum fractions as low as $x \sim 10^{-6}$.

Key words: Quantum Chromodynamics (QCD), LHCb, Proton-proton collisions, Proton-lead collisions, Identified charged particles, Pions, Kaons, Protons, Cold Nuclear Matter (CNM) effects, Cronin effect, Gluon fragmentation functions, High Energy Physics.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Resumo

Esta tese presenta a medida das taxas de produción de pións ($\pi^{\pm}$), kaóns ($K^{\pm}$), e protóns ($p^{\pm}$) cargados e identificados en colisións protón-protón (pp) e protón-chumbo (pPb) de alta enerxía no experimento LHCb do CERN. Este traballo ten como obxectivo proporcionar novas restricións sobre a dependencia na masa do aumento de Cronin no factor de modificación nuclear e contribuír a unha mellor comprensión das funcións de fragmentación de gluóns. Asemade, estas medidas son cruciais para mellorar os modelos de Monte Carlo, comprender as interaccións brandas e illar os efectos da Materia Nuclear Fría (CNM) nas colisións pPb para poder distinguilos das sinaturas do Plasma de Quark-Gluón (QGP) nas colisións PbPb que se analicen en futuros estudos. A cobertura cinemática única cara adiante e cara atrás do LHCb permite sondar diversos efectos da materia nuclear fría nun amplo rango da x de Bjorken.

Palabras clave: Cromodinámica Cuántica (QCD), LHCb, Colisións protón-protón, Colisións protón-chumbo, Partículas cargadas identificadas, Pións, Kaóns, Protóns, Efectos da Materia Nuclear Fría (CNM), Efecto Cronin, Funcións de fragmentación de gluóns, Física de Altas Enerxías.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Index

Acknowledgements **vi**

Abstract **vii**

Resumo **1**

1 Resumo en galego da tese **1**
1.1 Motivación teórica e obxectivos da tese 1
1.2 O experimento LHCb 4
1.3 Partículas cargadas primarias 5
1.4 Mostras de datos 6
1.5 Medida do espectro de partículas cargadas primarias identificadas 7
1.6 Resultados e discusión 14

2 Research Objectives, Hypotheses, and Methodology **20**
2.1 Objectives 20
2.2 Hypotheses 21
2.3 Methodology 21

3 Introduction **24**

4 Theoretical Background **26**
4.1 The Standard Model 26
4.1.1 Quantum Chromodynamics 27
4.1.2 The QCD Phase Diagram 28
4.2 Kinematics of High-Energy Collisions 30
4.3 High-Energy Hadronic Collisions 32
4.3.1 Elastic and Inelastic Scattering 32
4.3.2 Hard Parton Scattering and Subsequent Processes 33
4.3.3 Classification of Interaction Processes 33
4.3.4 Factorization Theorem 34
4.3.5 Proton-Proton Collision Physics 36
4.4 Heavy-Ion Collisions and Quark-Gluon Plasma 39
4.4.1 Signatures of the QGP 41

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.5 Phenomenology of Proton-Nucleus Collisions . . . 42
4.5.1 Nuclear Parton Distribution Functions and Initial State Effects . . 42
4.5.2 Saturation Physics and the Color Glass Condensate . . . 44
4.5.3 The Cronin Effect . . . 45
4.6 Monte Carlo Event Generators . . . 46
4.6.1 PYTHIA . . . 46
4.6.2 EPOS-LHC . . . 48

5 The LHCb experiment at the LHC 49
5.1 The Large Hadron Collider . . . 49
5.2 The LHCb Detector . . . 51
5.2.1 Tracking System . . . 51
5.2.1.1 Vertex Locator . . . 52
5.2.1.2 Primary Vertex Reconstruction . . . 52
5.2.1.3 Magnet System . . . 53
5.2.1.4 Silicon Tracker . . . 53
5.2.1.5 Outer Tracker . . . 53
5.2.1.6 Track Reconstruction . . . 53
5.2.2 Particle Identification . . . 54
5.2.2.1 RICH Detectors . . . 55
5.2.2.2 Calorimeters . . . 56
5.2.2.3 Muon System . . . 57
5.3 Data flow at LHCb . . . 61
5.3.1 Trigger System . . . 61
5.3.1.1 Level-0 Trigger . . . 61
5.3.1.2 High-Level Trigger . . . 61
5.3.2 LHCb Software Framework . . . 62

6 Definition of the Observables and Physics Motivation 64
6.1 Definition of Prompt Particles . . . 64
6.2 Kinematic Shift in Proton-Lead Collisions . . . 65
6.3 Unique Coverage of LHCb . . . 66
6.4 Physics Motivation and Key Observables . . . 67

7 Description of the Datasets 71
7.1 Luminosity . . . 71
7.2 Datasets . . . 72
7.2.1 Proton-Lead Dataset . . . 72
7.2.2 Luminosity Determination in Proton-Lead Data . . . 74
7.2.3 Proton-Proton Dataset . . . 75
7.2.4 Luminosity Determination in Proton-Proton Data . . . 75

8 Preparation and Selection of the Data and Simulation 78

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8.1 Event Selection . . . 78
8.1.1 Luminous Region . . . 78
8.1.2 Primary Vertex Conditions . . . 79
8.2 Candidate Selection . . . 84
8.2.1 Kinematic Range . . . 84
8.3 Signal Candidates . . . 84
8.3.1 PID Selection . . . 85
8.3.2 Background Sources . . . 86
8.4 Detector Occupancy . . . 93

9 Measurement of the Identified Prompt Charged Particle Spectra 98
9.1 Definitions . . . 98
9.1.1 Selecting The Binning in The (η, p_{T}) plane . . . 99
9.2 Isolating the Prompt Contribution: Purity Calculation . . . 101
9.2.1 Fake Tracks . . . 101
9.2.2 Secondary Tracks . . . 108
9.2.3 Purity . . . 122
9.3 Separating Prompt Contributions . . . 124
9.3.1 Particle Identification Efficiency . . . 124
9.3.2 Reweighting the Calibration Samples . . . 125
9.3.3 Determination of the Efficiencies with the Tag & Probe Method . . . 126
9.3.4 Correcting the Total Momentum Dependency per Species with an Iterative Data-Driven Method . . . 127
9.4 Reconstruction Efficiency . . . 140
9.5 Selection Efficiency . . . 144
9.6 Truth Matching Efficiency . . . 148

10 Systematic Uncertainties 149
10.1 Purity Calculation . . . 149
10.1.1 Fake Tracks . . . 150
10.1.2 Secondary Tracks . . . 156
10.1.3 Purity . . . 159
10.2 Separating Prompt Contributions . . . 161
10.2.1 Fit Model . . . 161
10.2.2 PID Selection on the Tag Particle . . . 161
10.2.3 Reweighting of the Calibration Samples . . . 161
10.2.4 Iterative Data-Driven Method . . . 163
10.2.5 Total Uncertainty of the PID Matrix . . . 163
10.2.6 Propagating the PID Matrix Uncertainty and the *Selected* Yields Uncertainty to the *True* Yields . . . 163
10.3 Reconstruction Efficiency . . . 170
10.3.1 Tracking Correction Factor . . . 170

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

10.3.2 Occupancy Correction 171
10.3.3 Total Reconstruction Efficiency Systematic Uncertainty 172
10.4 Selection Efficiency 176

11 Results and Discussion **178**
11.1 Cross Section and Nuclear Modification Factor 178
11.2 Particle Ratios and Correlations 187

12 Conclusions **196**

A Calibration samples preparation **198**
A.1 Λ calibration sample 198
A.2 K^0_S calibration sample 201
A.3 ϕ calibration sample 203

B Additional figures **205**

List of figures **208**

List of tables **221**

Glossary **223**

References **225**

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

1

Resumo en galego da tese

1.1 Motivación teórica e obxectivos da tese

A Cromodinámica Cuántica (QCD, do inglés "Quantum Chromodynamics") é a teoría fundamental que describe a interacción forte, que rexe o comportamento de quarks e gluóns e o seu confinamento en partículas compostas coñecidas como hadróns (como protóns e neutróns) [1]. Mentres que a QCD proporciona un marco robusto para predicir a produción de partículas en interaccións "duras" (hard), caracterizadas por unha gran transferencia de momento, a comprensión das interaccións "brandas" (soft), que implican unha pequena transferencia de momento, segue a ser un desafío teórico significativo. Estes procesos brandos dominan moitos aspectos das colisións hadrónicas, polo que a súa descrición precisa resulta crucial para diversos ámbitos. Para investigar a intricada estrutura dos hadróns e explorar as propiedades da QCD, realízanse colisións ultra-relativistas en aceleradores de partículas, como o Gran Colisor de Hadróns (LHC, polas súas siglas en inglés Large Hadron Collider). O LHC, o acelerador de hadróns máis grande do mundo, facilita colisións protón-protón (pp), protón-chumbo (pPb) e chumbo-chumbo (PbPb) a enerxías sen precedentes. Situado no punto de interacción 8 do LHC, o experimento LHCb (do inglés "Large Hadron Collider beauty") [2] é un espectrómetro deseñado para detectar partículas cargadas producidas a ángulos pequenos respecto das traxectorias dos feixes en colisión. Nas colisións das partículas dos feixes as partículas cargadas, como pións ($\pi^{\pm}$), kaóns ($K^{\pm}$) e protóns ($p^{\pm}$) son algúns dos produtos primarios. A produción de partículas cargadas de baixo momento está en gran parte gobernada por interaccións brandas, o que fai que as predicións teóricas desde primeiros principios sexan un desafío. Porén a súa medición precisa é indispensable. Isto esténdese máis alá da física de partículas fundamental, impactando áreas como a física dos raios cósmicos, onde a modelaxe precisa das fervenzas hadrónicas é vital para interpretar as observacións de instalacións como o Observatorio Pierre Auger [3].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Nas colisións de ións pesados, as complexidades observadas nas colisións *pp* fanse aínda máis evidentes. En particular, nas colisións PbPb, os científicos predín a formación dun estado de materia denominado Plasma de Quarks e Gluóns (QGP,do inglés "Quark gluon plasma"). Neste estado, os quarks e gluóns deixan de estar confinados dentro de partículas individuais para existir nun estado desconfinado [4]. A formación do QGP inflúe significativamente na produción de partículas, levando a miúdo a unha diminución en relación ás colisións *pp*. Para interpretar con precisión as sinaturas experimentais das colisións PbPb e illar os efectos do QGP, as colisións *p*Pb serven como unha referencia crucial. Nas colisións *p*Pb, aínda que a formación a gran escala do QGP é xeralmente limitada, a produción de partículas cargadas vese afectada polas diferenzas intrínsecas na estrutura partónica do núcleo de chumbo en comparación coa dun protón libre [5]. Estas modificacións, coñecidas como efectos da Materia Nuclear Fría (CNM, do inglés "Cold Nuclear Matter"), inclúen fenómenos como o *shadowing*, *anti-shadowing* e o efecto Cronin [6]. Unha comprensión profunda destes efectos CNM é primordial, xa que se superpoñen ás sinaturas do QGP nas colisións PbPb.

A exploración do réxime de x pequenos, onde as incertezas teóricas son significativas, require de medicións experimentais. A relación entre x e a pseudorapidez no centro de masas ($\eta_{\rm cms}$) descríbese cuantitativamente pola ecuación:

$$x \approx \frac{p_{\rm T}}{\sqrt{s_{\rm NN}}} e^{-\eta_{\rm cms}}, \tag{1.1}$$

onde $p_{\rm T}$ representa o momento transverso das partículas producidas e $\sqrt{s_{\rm NN}}$ denota a enerxía do centro de masas. A cobertura de varios experimentos en x e o momento transferido na colisión ao cadrado, denotado como Q^2, en colisións protón-chumbo móstrase na figura 1.1.

Como se observa, o LHCb pode acceder a fraccións de momento tan baixas como 10^{-6} na **rexión cara adiante** (*p*Pb, co feixe de protóns apuntando cara ao espectrómetro LHCb) e ata 10^{-2} na **rexión cara atrás** (Pb*p*, co feixe de chumbo apuntando cara ao espectrómetro LHCb). Isto converte ao LHCb nun laboratorio único para o estudo das funcións de distribución de partóns nucleares (nPDFs, do inglés "Nuclear Parton Distribucion Functions"). Os efectos experimentais da materia nuclear fría cuantifícanse a miúdo usando o factor de modificación nuclear, que compara o número de partículas producido en colisións *p*Pb co de colisións *pp*, escalados polo número de masa atómica do núcleo. Para unha especie de hadrón dada h, defínese como:

$$R^h_{p\rm Pb} = \frac{1}{A} \frac{d\sigma^h_{p\rm Pb}/dp_{\rm T}d\eta_{\rm cms}}{d\sigma^h_{pp}/dp_{\rm T}d\eta_{\rm cms}}, \tag{1.2}$$

onde A é o número de masa atómica do núcleo (208 para o chumbo) e $d\sigma^h_{p\rm Pb/}{}_{pp}/dp_{\rm T}d\eta_{\rm cms}$ representa a sección eficaz diferencial para a produción de hadróns lixeiros (h) en colisións *p*Pb e *pp* en función do momento transverso e da pseudorapidez do centro de masas. As

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

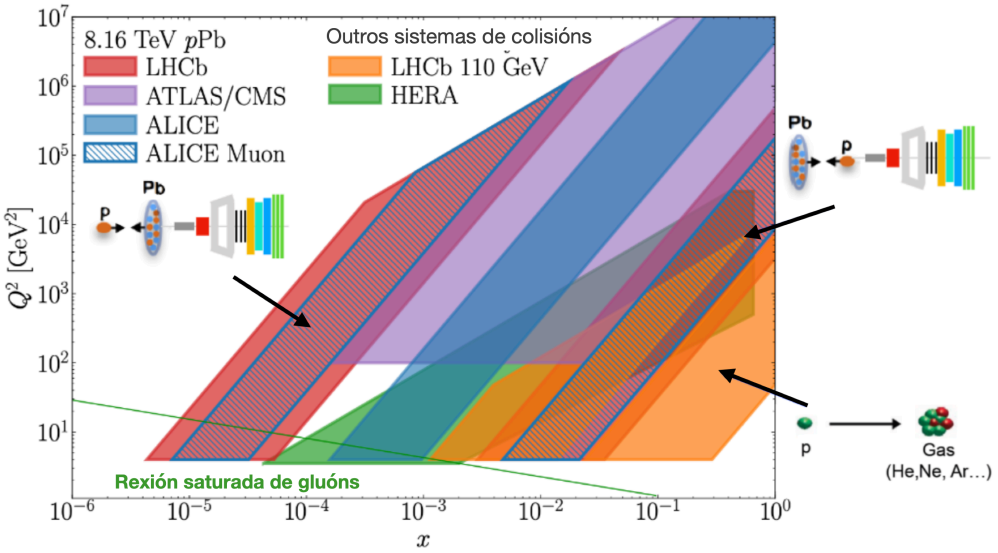


Figura 1.1: Cobertura cinemática de LHCb e outros experimentos para colisións pPb en termos do momento lonxitudinal fraccionario do nucleón no branco, x, e da masa invariante partón-partón ao cadrado, Q^2.

seccións eficaces diferenciais de produción dos hadróns cargados pódense estimar como:

$$\frac{\mathrm{d}^2\sigma^h_{pp,p\mathrm{Pb}}}{\mathrm{d}p_T\mathrm{d}\eta_{\mathrm{cms}}} = \frac{1}{\mathcal{L}}\frac{N^h(\eta_{\mathrm{cms}},p_T)}{\Delta p_T \Delta\eta_{\mathrm{cms}}}. \tag{1.3}$$

Aquí, N^h é o número de hadróns lixeiros no intervalo bidimensional $(\Delta\eta_{\mathrm{cms}},\Delta p_T)$, e $\mathcal{L}$ é a luminosidade integrada da mostra de datos analizada. Pode obterse información complementaria da relación de producción de dous hadróns, h_1 e h_2, definida como a razón dos seus números de hadróns medidos no mesmo intervalo cinemático:

$$R_{h_1/h_2} = \frac{N^{h_1}(\eta_{\mathrm{cms}},p_T)}{N^{h_2}(\eta_{\mathrm{cms}},p_T)}. \tag{1.4}$$

Este observable permite o estudo da producción relativa de pións, kaóns e protóns en colisións de alta enerxía. A relación protón-pión sonda a producción de barións en relación cos mesóns, mentres que a relación kaón-pión é sensible á producción de extrañeza. A relación kaón-protón proporciona información sobre o equilibrio entre o contido de quarks extraños e non extraños no estado final

Entre as contribucións máis recentes relacionadas con esta tese, a colaboración LHCb mediu o factor de modificación nuclear $R_{p\mathrm{Pb}}$ de pións neutros (π^0) en colisións pPb a

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

$\sqrt{s_{\mathrm{NN}}} = 8.16\,\mathrm{TeV}$, así como o de partículas cargadas primarias a $\sqrt{s_{\mathrm{NN}}} = 5\,\mathrm{TeV}$ [7,8]. Na rexión de pseudorrapidez cara adiante ambas medicións mostran unha supresión significativa da produción de partículas en relación coas colisións pp escaladas, en coherencia coas predicións de nPDFs. Os resultados para π^0 son compatibles cos observados para partículas cargadas. Pola contra, na rexión cara atrás, obsérvase un aumento da produción, tanto para os π^0 como para as partículas cargadas. Porén, a magnitude deste aumento é menor no caso dos pións neutros, en comparación co observado para partículas cargadas. Esta diferenza podería atribuírse á contribución adicional de mesóns máis pesados e barións na medición de partículas cargadas, o que suxire unha posible dependencia coa masa, consistente cun fluxo radial ou con mecanismos de recombinación na fase final. O estudo con hadróns identificados pode permitir discriminar entre estas hipóteses. Esta tese aborda os desafíos mencionados anteriormente mediante unha medición das taxas de produción de pións, kaóns e protóns cargados, identificados en colisións de alta enerxía pp e pPb no experimento LHCb.

1.2 O experimento LHCb

O experimento LHCb é un espectrómetro dun só brazo, optimizado para estudar partículas producidas a baixo ángulo respecto do feixe incidente, e descrito en detalle en [9]. O detector abrangue un intervalo de pseudorrapidez de $2 < \eta < 4.8$ e permite estudar colisións en condicións de momento lonxitudinal elevado.

Os subsistemas fundamentais para esta análise son: un detector de vértice de tiras de silicio (VELO, do inglés "Vertex LOcator"), situado arredor da rexión de interacción e encargado de determinar a posición do vértice primario (VP); un sistema de trazado composto polo Trigger Tracker (TT) e as estacións T, que proporciona a medida do momento das partículas cargadas grazas á presenza dun imán dipolar; e dous detectores de Cherenkov de imaxe en anel (RICH, do inglés "Ring Imaging Cherenkov Detector"), que permiten identificar distintas especies de partículas cargadas.

As traxectorias das partículas cargadas reconstrúense a partir da información rexistrada polos distintos subdetectores. Estas traxectorias, denominadas **trazas**, clasifícanse segundo os subdetectores implicados na súa reconstrución:

- **Trazas VELO**: baseadas unicamente en sinais no VELO.
- **Trazas Upstream**: combinan sinais no VELO e no TT.
- **Trazas T**: construídas a partir de sinais só nas estacións T.
- **Trazas Downstream**: baseadas en sinais no TT e nas estacións T.
- **Trazas Long**: combinan sinais no VELO, no TT e nas estacións T.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

As trazas Long e Downstream son as máis utilizadas nas análises de física, debido á súa calidade e á información completa que proporcionan. O LHC xera colisións a unha frecuencia de 40 MHz, unha taxa moi alta para o detector. Para seleccionar os eventos máis relevantes, LHCb emprega un sistema de **trigger** en tempo real que reduce significativamente esta taxa, permitindo almacenar e procesar os datos de interese.

O sistema de trigger divídese en dúas fases principais: unha primeira etapa en hardware, o Nivel 0, que realiza unha primeira selección para reducir a taxa de eventos, seguida dunha segunda etapa en software, o Trigger de Alto Nivel, que leva a cabo unha selección máis detallada mediante a reconstrución completa dos eventos.

O software offline de LHCb integra ferramentas para simulación, reconstrución, selección e análise de datos, garantindo a consistencia entre datos físicos e simulados. Os principais paquetes de software inclúen: Gauss para simulación, Boole para dixitalización, Moore para o trigger, Allen para procesamento acelerado de datos por GPU, e DaVinci para reconstrución e selección de eventos.

1.3 Partículas cargadas primarias

Na física experimental de altas enerxías, defínese unha partícula cargada primaria como calquera hadrón ou leptón con vida media propia superior a 30 ps, producido directamente na colisión ou a partir da desintegración de partículas con vidas máis curtas. Esta definición, recomendada polo grupo de traballo do *LHC Minimum Bias and Underlying Event Working Group* [10,11], permite incluír partículas cuxos produtos de desintegración non se poden distinguir do vértice primario. Pola contra, partículas de tempo de vida longo, como o K_S^0 ou a Λ, exclúense deste grupo. Ao longo desta tese, todas as referencias a **partículas cargadas** ou **partícula cargada primaria** seguirán estritamente esta definición. A Táboa 1.1 resume as vidas medias e anchuras de partículas relevantes para esta clasificación.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

		Vida media propia τ	
Especie	**Ancho Γ GeV**	**(ps)**	**(cm/c)**
p	0	∞	∞
γ	0	∞	∞
K^0	0	∞	∞
e^-	0	∞	∞
n	7.478×10^{-28}	8.861×10^{14}	2.656×10^{13}
μ^-	2.996×10^{-19}	2.212×10^{6}	6.63×10^{4}
K^0_{L}	1.287×10^{-17}	5.148×10^{4}	1543
π^+	2.528×10^{-17}	2.621×10^{4}	785.7
K^+	5.317×10^{-17}	1.246×10^{4}	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

Táboa 1.1: Anchura (Γ) e vida media propia (τ) das partículas de tempo de vida longo. As partículas con $\Gamma = 0$ considéranse estables para os efectos desta definición. A táboa está estruturada cunha listaxe inicial das partículas estables, seguida das partículas inestables ordenadas en orde descendente segundo τ. A entrada para K^0 con vida infinita mantense tal e como aparece na fonte, reflectindo un tratamento experimental específico dos kaóns neutros no contexto da definición de partículas primaria. Datos tomados de [11].

1.4 Mostras de datos

Nesta tese utilízanse datos de colisións pPb e Pbp recollidos en febreiro de 2013, cunha luminosidade integrada[1] de 42.7 ± 1.0 $\mu\mathrm{b}^{-1}$ e 38.7 ± 1.0 $\mu\mathrm{b}^{-1}$, respectivamente. As incertezas non están correlacionadas entre ambas mostras. Durante este período os sentidos dos feixes foron invertidos nas dúas configuracións para permitir a adquisición de datos con dúas polaridades de campo magnético: *MagDown*, cando o campo apunta cara ao chan, e *MagUp*, cando apunta cara ao teito. Deste xeito, a toma de datos abarca catro configuracións distintas: rexión cara adiante e rexión cara atrás, cada unha con dúas polaridades de campo. Os eventos selecciónanse mediante un sistema de trigger sen nesgos, que require unicamente a presenza dunha traza reconstruída no VELO. Ademais, só se consideran eventos cun único vértice primario reconstruído dentro de tres desviacións

[1] A luminosidade integrada representa a cantidade total de datos recollidos por un experimento e é proporcional ao número total de colisións observadas. Exprésase en unidades inversas de barn, onde 1 barn = 10^{-24} cm^2.

usc UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

estándar respecto á media da distribución dos vértices primarios da mostra completa. Esta zona denomínase rexión luminosa e corresponde á área onde se cruzan os feixes. Para a simulación, empréganse mostras oficiais de LHCb xeradas co xerador de eventos EPOS, axeitado para describir interaccións hadrónicas de mínimo nesgo en colisións pPb. A mostra de datos correspondente ás colisións pp foi recollida en novembro de 2015 e consta dunha luminosidade integrada de 3.49 ± 0.07 nb^{-1}. Os eventos selecciónanse mediante un sistema de trigger sen nesgo. Para esta tese só se consideran os eventos producidos nas colisións feixe-feixe. Esta estratexia permite evitar a contaminación procedente de interaccións en paquetes veciños do feixe, reducindo así os efectos de contaminación por interaccións tardías ou adiantadas (spill-over). No proceso de reconstrución, aplícanse criterios de calidade adicionais ás trazas, esixindo unha probabilidade de traza ghost (`GhostP`) inferior a 0.4 e unha calidade da traza caracterizada por $\chi^2/\text{ndf} < 4$. Estes requisitos non se aplican aos datos de protón-chumbo. Para a simulación das colisións pp, emprégase o xerador Pythia configurado para producir eventos de mínimo nesgo. Dado que os datos reais só están dispoñibles coa configuración *MagDown*, a simulación restrínxese a esta polaridade.

1.5 Medida do espectro de partículas cargadas primarias identificadas

A selección de eventos céntrase en colisións feixe-feixe empregando un sistema de trigger sen nesgos. Para reducir o fondo de interaccións de feixe-gas ou feixe-detector só se consideran eventos cun vértice primario (VP) ben reconstruído dentro dunha rexión luminosa. Nas colisións pPb, debido á alta multiplicidade, esíxese un único VP por evento, mentres que nas colisións pp esta restrición non se aplica para evitar introdución de nesgos.

A selección de candidatos baséase en trazas do tipo "long" que cumpren requisitos cinemáticos: $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}/c$, $p > 10\,\mathrm{GeV}/c$ e $2 < \eta < 4.8$. A identificación de partículas primarias utiliza o pseudo-parámetro de impacto (`pseudoIP`), que se constrúe como a distancia transversa entre o punto estimado de orixe da traza e a media gaussiana da distribució de VPs na mostra total. A vantaxe desta variable radica na súa utilidade cando o parámetro de impacto estándar resulta difícil de usar, debido á presenza de múltiples VPs ou á súa ausencia nas colisións pp. Ademais, a probabilidade de traza ghost (`GhostP`) utilízase para reducir as trazas falsas e, deste xeito, maximizar a pureza do sinal. Finalmente, a identificación de especies (PID, do inglés "Particle IDentification") realízase empregando a variable $\mathrm{DLL}_{\mathrm{X}\pi}$, que permite seleccionar mostras de pións, kaóns e protóns. Un resumo de todas as selección aplicadas nesta tese preséntase na Táboa 1.2. É esencial identificar e cuantificar as principais fontes de fondo coa fin de minimizar a súa influencia:

- **Trazas pantasma ou falsas ("Fake Tracks")**: artefactos da reconstrución que non se corresponden cunha partícula cargada. Reducidas pola selección en `GhostP`. O fondo remanente estimouse con simulación.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

- **Partículas secundarias ("Secondaries")**: provintes de desintegracións doutros estados (mesóns K^0_S, barións Λ, hiperóns, etc.) ou interaccións co material do detector, como conversións de fotóns ou interaccións hadrónicas. Caracterízanse por unha distribución cun maior `pseudoIP` e cuantifícanse mediante simulación segundo a súa orixe.
- **Partículas mal identificadas ("Mis-Identified Particle")**: resultado da separación imperfecta entre especies na selección de PID. As trazas mal identificadas cuantifícanse con mostras de calibración en datos para corrixir a pureza das mostras.
- **Trazas clonadas ("Clone Tracks")**: dúas trazas xeradas a partir dunha única partícula. A maioría elimínanse mediante algoritmos específicos e a contribución do fondo restante é desprezábel.

Como se definiu na sección 1.1, a sección eficaz diferencial de produción de hadróns lixeiros cargados nas colisións pp, pPb e Pbp conséguese a partir da ecuación (1.3), onde N^h representa o número de hadróns lixeiros cargados. Para obter o valor de N^h o número de candidatos, N^h_{cand}, é inicialmente extraído unha vez aplicados os criterios de selección mencionados na Táboa 1.2. Este N^h_{cand} debe ser entón corrixido en dous pasos principais. O primeiro paso consiste en eliminar a contribución de trazas ghost e secundarias, estimando a pureza, P^h, para obter só o número de partículas primarias, N^h_{Sel}. Isto faise mediante a seguinte expresión:

$$N^h_{Sel} = N^h_{cand} \cdot P^h, \tag{1.5}$$

onde a pureza P^h está definida como:

$$P^h = 1 - \left(\sum_{\alpha=\{ghost,sec\}} R^h_\alpha f^{sim}_\alpha(h) \right), \tag{1.6}$$

sendo f^{sim}_α a fracción de trazas ghost ($\alpha = ghost$) ou secundarias ($\alpha = sec$) na simulación e R_α é un factor que corrixe as diferenzas entre datos e a simulación. Os resultados correspondentes a esta pureza móstranse na figura 1.2. As seleccións de identificación de partículas non conseguen separar as especies de partículas de xeito absoluto. Por iso, o segundo paso da corrección consiste en estimar a fracción ben identificada e a contaminación entre especies presente en N^h_{Sel}. Para isto, utilízase unha matriz de resposta de identificación de partículas. Esta matriz relaciona N^h_{Sel} co número do partículas verdadeiras, libres de contaminación entre especies, N^h_{True}, segundo a seguinte expresión :

$$\begin{bmatrix} N^\pi_{True} \\ N^K_{True} \\ N^p_{True} \end{bmatrix} = \begin{bmatrix} \varepsilon^{\pi\to\pi} & \varepsilon^{K\to\pi} & \varepsilon^{p\to\pi} \\ \varepsilon^{\pi\to K} & \varepsilon^{K\to K} & \varepsilon^{p\to K} \\ \varepsilon^{\pi\to p} & \varepsilon^{K\to p} & \varepsilon^{p\to p} \end{bmatrix}^{-1} \begin{bmatrix} N^\pi_{Sel} \\ N^K_{Sel} \\ N^p_{Sel} \end{bmatrix}, \tag{1.7}$$

onde $\varepsilon^{h\to h}$ son as eficiencias de identificación da partícula h, e $\varepsilon^{h\to h'}$ son factores de falsa identificación da partícula h que foi mal identificada como h' (é dicir, partículas h que

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

		pp	pPb	Pbp
Selección de evento	Trigger	`Hlt1NoBiasLeadingCrossing`	`Hlt1MBMicroBiasVelo`	`Hlt1MBMicroBiasVelo`
	Bunch crossing type	bunch-bunch	bunch-bunch	bunch-bunch
	number of PV	-	1	1
	PV position	-	within luminous region	within luminous region
Selección de candidato	Track container	`/Event/Phys/StdAllNoPIDsPions/Particles`	`/Event/Phys/StdAllNoPIDsPions/Particles`	`/Event/Phys/StdAllNoPIDsPions/Particles`
	Track type	Long	Long	Long
	η	$2 < \eta < 4.8$	$2 < \eta < 4.8$	$2 < \eta < 4.8$
	p	$p > 10\,\mathrm{GeV}/c$	$p > 10\,\mathrm{GeV}/c$	$p > 10\,\mathrm{GeV}/c$
	p_T	$0.5 < p_\mathrm{T} < 4.619\,\mathrm{GeV}/c$	$0.5 < p_\mathrm{T} < 4.619\,\mathrm{GeV}/c$	$0.5 < p_\mathrm{T} < 4.619\,\mathrm{GeV}/c$
	`pseudoIP`	0.368 mm	0.348 mm	0.348 mm
	`GhostP`	0.078	0.103	0.109
Selección de PID	pions	$\mathrm{DLL}_{K\pi} < 0$ & $\mathrm{DLL}_{p\pi} < 0$ & hasRich >0.5& `isMuon`==0	$\mathrm{DLL}_{K\pi} < 0$ & $\mathrm{DLL}_{p\pi} < 10$ & hasRich >0.5 & `isMuon`==0	$\mathrm{DLL}_{K\pi} < 0$ & $\mathrm{DLL}_{p\pi} < 10$ & hasRich >0.5 & `isMuon`==0
	Kaons	$\mathrm{DLL}_{K\pi} > 5$ & ($\mathrm{DLL}_{K\pi}$-$\mathrm{DLL}_{p\pi}$) >5 & hasRich >0.5 & `isMuon`==0	$\mathrm{DLL}_{K\pi} > 5$ & ($\mathrm{DLL}_{K\pi}$-$\mathrm{DLL}_{p\pi}$) >0 & hasRich >0.5& `isMuon`==0	$\mathrm{DLL}_{K\pi} > 0$ & ($\mathrm{DLL}_{K\pi}$-$\mathrm{DLL}_{p\pi}$) >0 & hasRich >0.5& `isMuon`==0
	protons	$\mathrm{DLL}_{p\pi} > 5$ & ($\mathrm{DLL}_{p\pi}$-$\mathrm{DLL}_{K\pi}$) >10 & hasRich >0.5& `isMuon`==0	$\mathrm{DLL}_{p\pi} > 10$ & ($\mathrm{DLL}_{p\pi}$-$\mathrm{DLL}_{K\pi}$) > 10& hasRich >0.5& `isMuon`==0	$\mathrm{DLL}_{p\pi} > 10$ & ($\mathrm{DLL}_{p\pi}$-$\mathrm{DLL}_{K\pi}$) > 10& hasRich >0.5& `isMuon`==0

Táboa 1.2: Resumo de todas as seleccións consideradas para a tese.

Especia	**Baixo momento**	**Alto momento**
$\pi^{\pm}$	$K^0_S \to \pi^+\pi^-$	$D^{*+} \to D^0\pi^+$ with $D^0 \to K^-\pi^+$
$K^{\pm}$	$\phi \to K^+K^-$	$D^{*+} \to D^0\pi^+$ with $D^0 \to K^-\pi^+$
p, $\overline{p}$	$\Lambda \to p\pi^-$	$\Lambda \to p\pi^-$

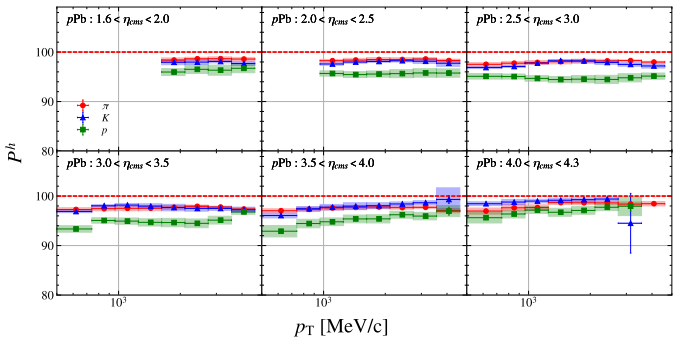
Táboa 1.3: Mostras de calibración consideradas nesta tese. As mostras proceden das canles de desintegración empregadas para calibrar as eficiencias de PID en LHCb [12].

pasan os criterios de PID de $h^{'}$). Estas eficiencias e factores de falsa identificación determínanse estudando modos de desintegración específicos para cada partícula, recollidos en mostras de calibración. Estas mostras están listadas na Táboa 1.3.

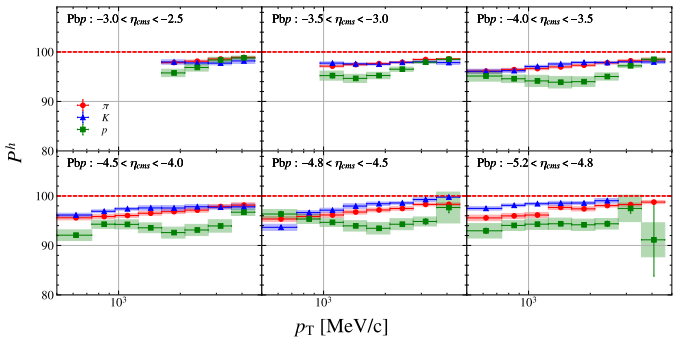
Finalmente, os valores de N^h_{True} corríxense polas eficiencias de reconstrución ε^h_{reco}, de selección ε^h_{sel}, e de asociación á traza xerada ε_{TM}. As eficiencias de reconstrución e selección están mostradas na figura 1.3 e na figura 1.4, respectivamente. A corrección faise segundo:

$$N^h = N^h_{True} \cdot \frac{1}{\varepsilon^h_{reco} \cdot \varepsilon^h_{sel} \cdot (1/\varepsilon_{TM})}. \tag{1.8}$$

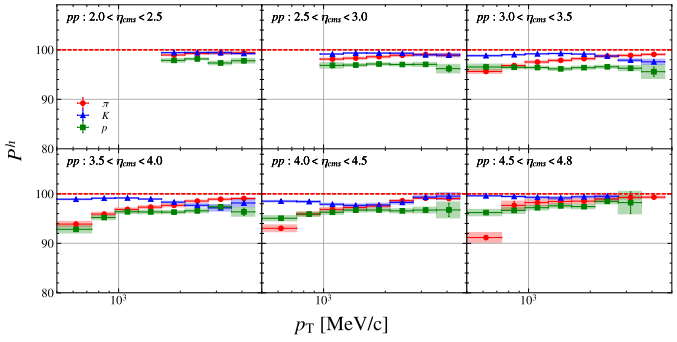
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Pureza nas colisións pPb.



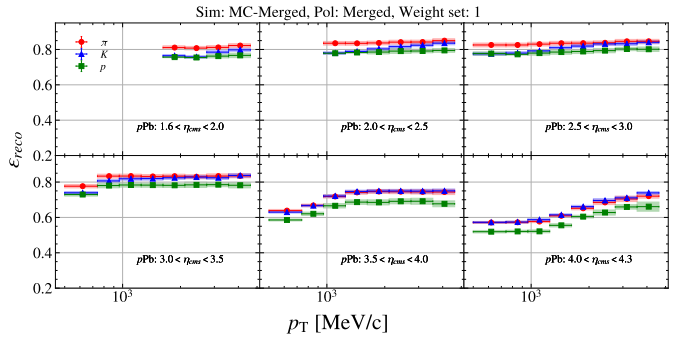
(b) Pureza nas colisións Pbp.



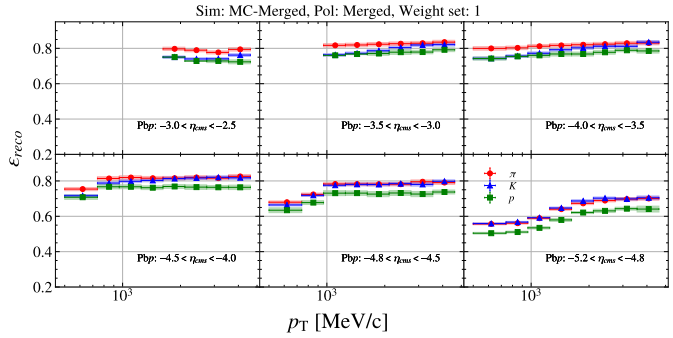
(c) Pureza nas colisións Pbp.

Figura 1.2: Porcentaxe de pureza da mostra de candidatos nas colisións pPb (parte superior), Pbp (parte central) e pp (parte inferior). Os resultados represéntanse para cada tipo de hadrón lixeiro: os pións en vermello, os kaóns en azul e os protóns en verde. As incertezas estatísticas indícanse mediante liñas verticais, mentres que as incertezas sistemáticas móstranse como caixas.

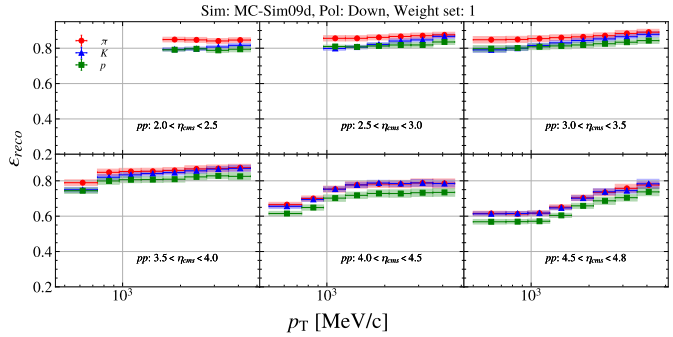
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Eficiencia de reconstrucción en pPb.

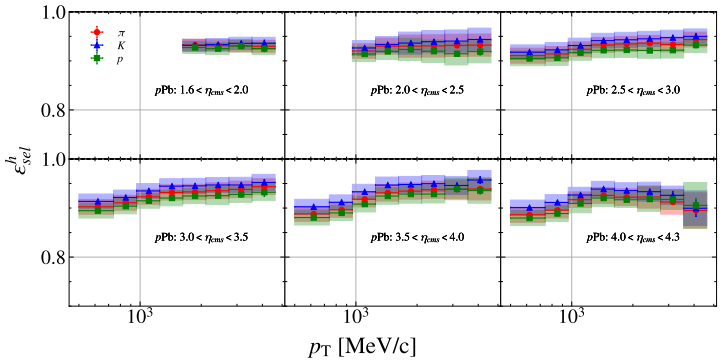


(b) Eficiencia de reconstrucción en Pbp.

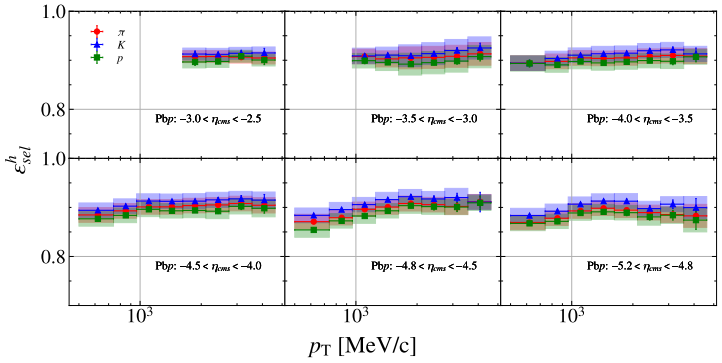


(c) Eficiencia de reconstrucción en pp.

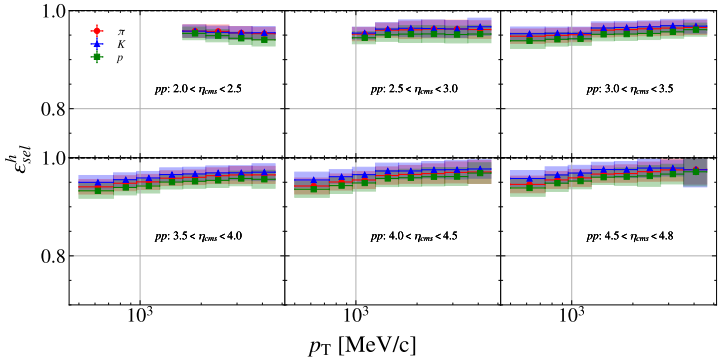
Figura 1.3: Eficiencia de reconstrucción nas colisións pPb (arriba), Pbp (no centro) e pp (abaixo). Os resultados están representados para cada tipo de hadrón lixeiro: os pións en vermello, os kaóns en azul e os protóns en verde. As incertezas estatísticas están indicadas con liñas verticais, mentres que as incertezas sistemáticas aparecen como caixas.



(a) Eficiencia de selección en pPb.



(b) Eficiencia de selección en Pbp.



(c) Eficiencia de selección en pp.

Figura 1.4: Eficiencia de selección nas colisión pPb (arriba), Pbp (no centro) e pp (abaixo). Os resultados están representados para cada tipo de hadrón lixeiro: os pións en vermello, os kaóns en azul e os protóns en verde. As incertezas estatísticas están indicadas con liñas verticais, mentres que as incertezas sistemáticas aparecen como caixas.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

1.6 Resultados e discusión

Seguindo a metodoloxía descrita na sección 1.5, obtéñense as seccións eficaces diferenciais para a produción de hadróns lixeiros primarios. As partículas analizadas son pións, kaóns e protóns, e os resultados preséntanse separadamente para colisións pp, pPb e Pbp. Os resultados obtidos compáranse con predicións teóricas baseadas en cálculos de QCD perturbativa a orde seguinte á principal (NLO), nas que se incorporan efectos nucleares a través da parametrización EPPS21. Os modelos inclúen as funcións de fragmentación de Kretzer e utilizan os conxuntos de PDFs CT18ANLO. As incertezas estatísticas represéntanse mediante barras verticais, mentres que as incertezas sistemáticas aparecen como caixas sombreadas. As predicións, representadas como bandas sombreadas, abranguen as incertezas asociadas ás escalas de renormalización e factorización, así como aos erros nas PDFs. Os resultados móstranse na figura 1.5. Os pións presentan a sección eficaz máis elevada, seguidos polos kaóns e finalmente polos protóns

Unha vez determinadas as seccións eficaces, calcúlase o factor de modificación nuclear. Esta cantidade compara a produción en colisións pPb coa produción en pp, tendo en conta o número de nucleóns do chumbo. A incerteza sistemática asociada á luminosidade cancela neste cociente ao estar completamente correlacionada entre as dúas mostras de datos. Os resultados móstranse separadamente para a rexión de pseudorapidez cara adiante e cara atrás fig. 1.6.

Os valores do factor de modificación nuclear mostran unha clara dependencia coa pseudorapidez. Tomando en conta a referencia de que a unidade indicaría que o núcleo é unha unión incoherente de nucleóns, podemos observar que na rexión cara adiante se observa unha supresión significativa, común a todas as especies de hadróns, o que suxire un mecanismo de perda de enerxía independente da masa. Pola contra, na rexión cara atrás detéctase un aumento, especialmente acentuado nos protóns, o que podería indicar efectos dependentes da masa, posiblemente relacionados coa fragmentación ou coas interaccións finais. Os resultados obtidos neste traballo compáranse con medidas anteriores da colaboración LHCb. Para a rexión cara adiante, os factores de modificación nuclear dos pións, kaóns e protóns son consistentes cos obtidos anteriormente para partículas cargadas primarias. Na rexión cara atrás, os resultados para pións e kaóns coinciden co comportamento observado para o π^0 a 8 TeV, mentres que os protóns presentan unha mellora comparable á observada en colisións a 5 TeV figura 1.7a.

Os resultados experimentais deste traballo tamén se comparan coas predicións teóricas baseadas en EPPS21. Estas predicións reproducen correctamente a supresión observada na rexión cara adiante. Porén, na rexión cara atrás, se ben os modelos explican o comportamento dos pións e kaóns, non logran describir a mellora observada nos protóns figura 1.7b.

Finalmente, como se introduciu na sección 1.1, a razón entre as producións de dúas especies de hadróns defínese como

$$R_{h_1/h_2} = \frac{N^{h_1}(\eta_{\mathrm{cms}}, p_{\mathrm{T}})}{N^{h_2}(\eta_{\mathrm{cms}}, p_{\mathrm{T}})}, \qquad \text{(eq. (1.4))}$$

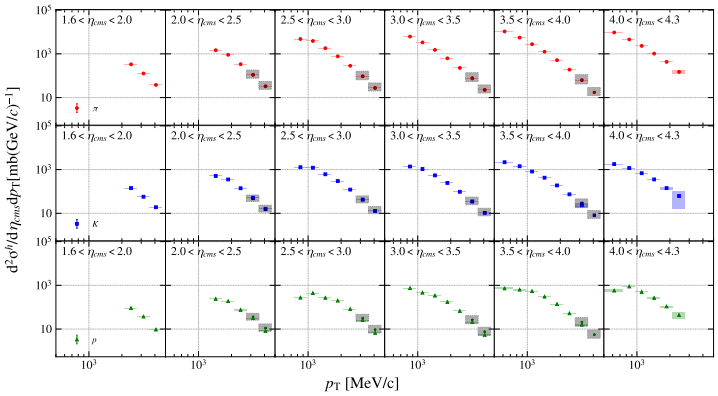
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

onde N^{h_i} representa o número verdadeiro de hadróns da especie i no intervalo $(\eta_{\mathrm{cms}}, p_{\mathrm{T}})$. Esta razón proporciona información sobre as taxas relativas de produción de pións, kaóns e protóns. As producións verdadeiras das partículas determínanse mediante a inversión da matriz de resposta de identificación de partículas, tal e como se define na ecuación 1.7. Non obstante, as probabilidades finitas de mala identificación introducen correlacións estatísticas entre as producións verdadeiras das diferentes especies de hadróns. Por exemplo, unha sobreestimación na produción dunha especie, debido a eventos mal identificados doutro tipo de hadrón, levará a unha subestimación correspondente na produción desa segunda especie unha vez aplicada a corrección. Estas correlacións son fundamentais para unha estimación adecuada da incerteza estatística nas razóns de produción, o que se reflicte na expresión para o cálculo da incerteza relativa propagada, dada por:

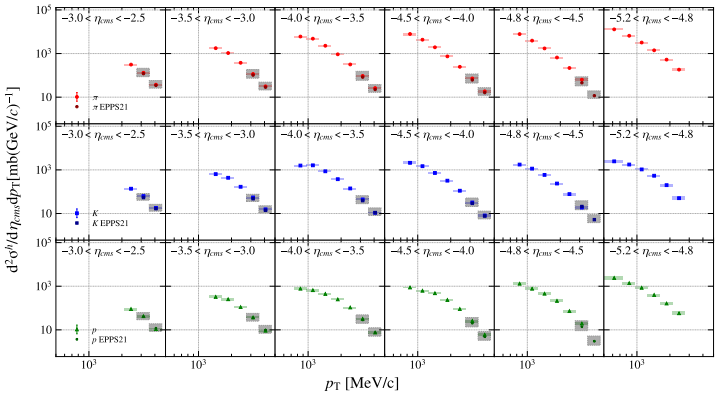
$$\left(\frac{\Delta R_{h_1/h_2}}{R_{h_1/h_2}}\right)^2 = \left(\frac{\Delta N^{h_1}}{N^{h_1}}\right)^2 + \left(\frac{\Delta N^{h_2}}{N^{h_2}}\right)^2 - 2\frac{\mathrm{Cov}(N^{h_1}, N^{h_2})}{N^{h_1} N^{h_2}}, \tag{1.9}$$

onde ΔN^{h_i} representa a incerteza na produción verdadeira da especie i e $\mathrm{Cov}(N^{h_1}, N^{h_2})$ é o termo de covarianza que cuantifica a correlación entre as incertezas das dúas producións. O resultado do cociente de kaóns sobre pións móstrase na figura 1.8.

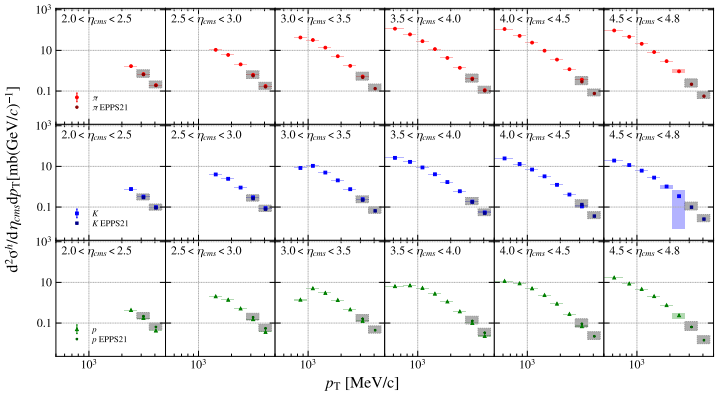
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Sección eficaz en pPb.



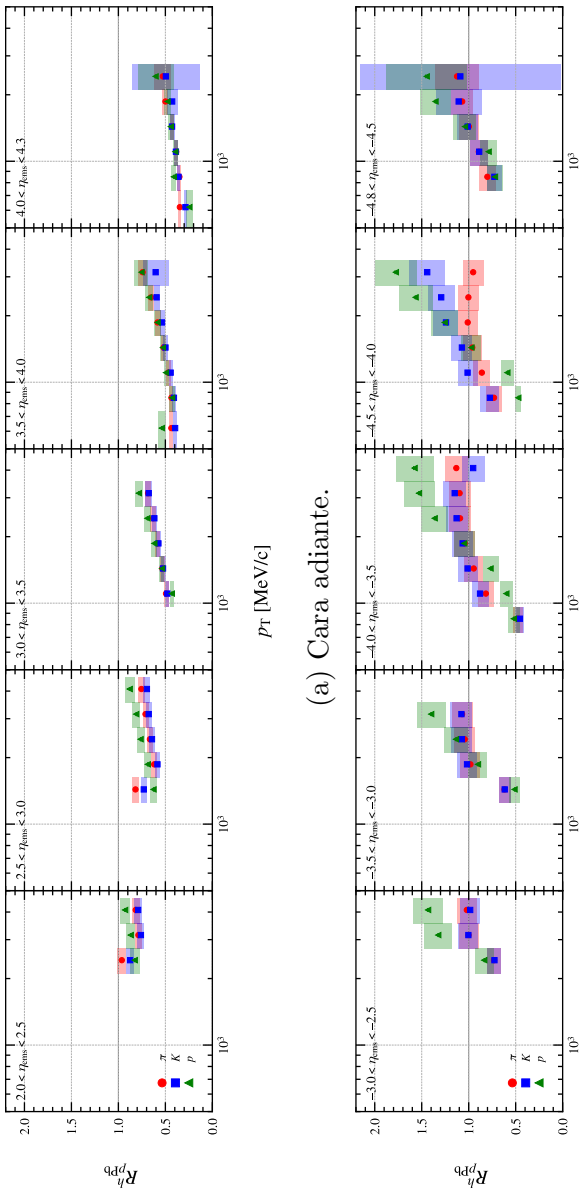
(b) Sección eficaz en Pbp.



(c) Sección eficaz en pp.

Figura 1.5: A sección eficaz de produción en colisións pPb (arriba), Pbp (centro) e pp (abaixo), comparada coas predicións do modelo EPPS21 (liñas negras discontinuas). Os resultados represéntanse para cada tipo de hadrón lixeiro: os pións móstranse en vermello, os kaóns en azul e os protóns en verde. As incertezas estatísticas indícanse mediante liñas verticais, mentres que as incertezas sistemáticas aparecen como caixas sombreadas.

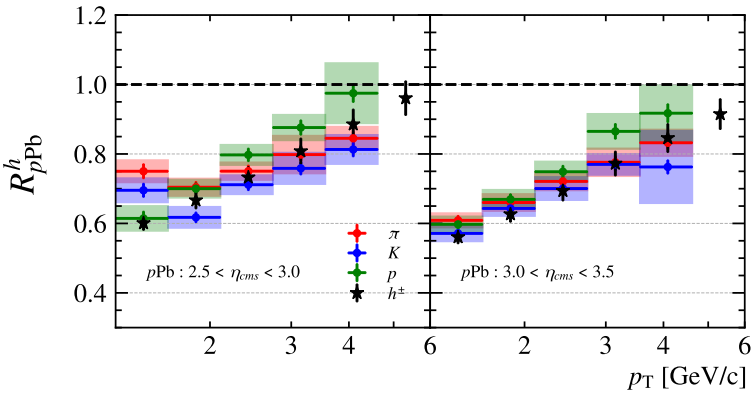
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



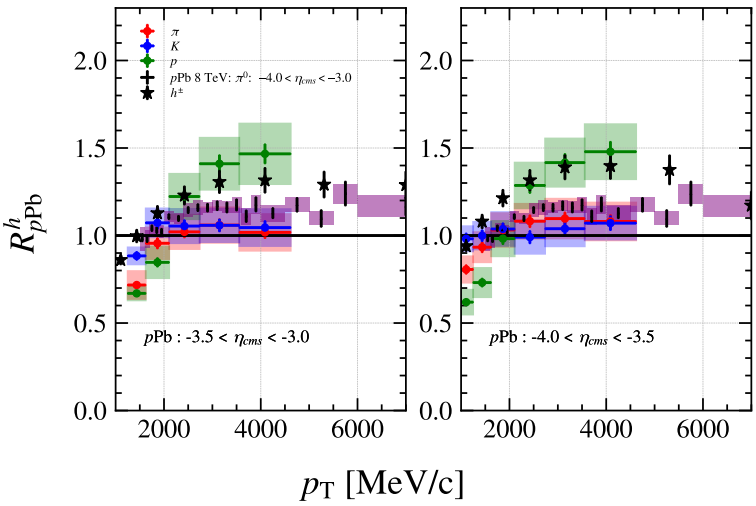
(a) Cara adiante.

(b) Cara atrás.

Figura 1.6: O factor de modificación nuclear na rexión cara adiante (arriba) e na rexión cara atrás (abaixo). Os resultados represéntanse para cada tipo de hadrón lixeiro: os pións móstranse en vermello, os kaóns en azul e os protóns en verde. As incertezas estatísticas indícanse mediante liñas verticais, mentres que as incertezas sistemáticas aparecen como caixas sombreadas.

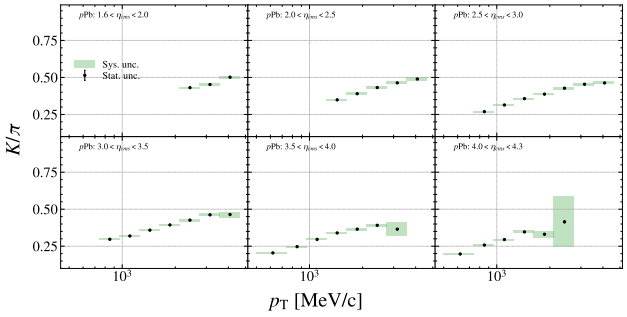


(a) Cara adiante.

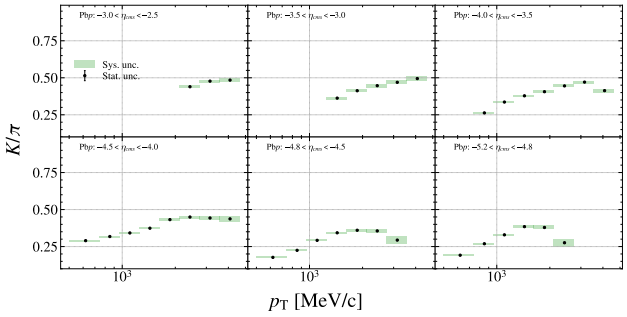


(b) Cara atrás.

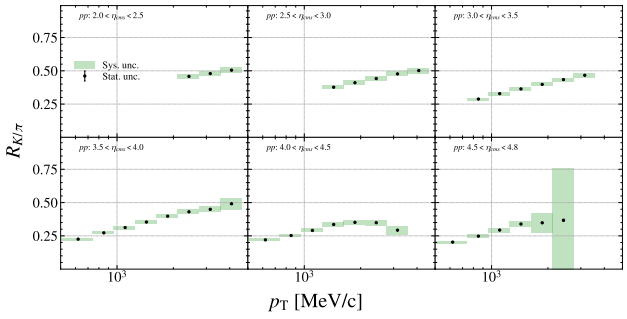
Figura 1.7: O factor de modificación nuclear na rexión cara adiante (arriba) e na rexión cara atrás (abaixo). Os resultados represéntanse para cada tipo de hadrón lixeiro: os pións móstranse en vermello, os kaóns en azul e os protóns en verde. As incertezas estatísticas indícanse mediante liñas verticais, mentres que as incertezas sistemáticas aparecen como caixas sombreadas. Os resultados na rexión cara adiante compáranse coa análise a 5 TeV de partículas cargadas primarias [8], mentres que na rexión cara atrás compáranse tanto cos resultados a 5 TeV como cos pións neutros a 8 TeV [7].



(a) pPb.



(b) Pbp.



(c) pp.

Figura 1.8: Cociente de produción de kaóns sobre pións en pPb (arriba), Pbp (medio), pp (abaixo). As incertezas estatísticas están indicadas con liñas verticais, mentres que as incertezas sistemáticas se amosan con caixas.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

2

Research Objectives, Hypotheses, and Methodology

2.1 Objectives

The primary objective of this thesis is to measure the production of identified prompt charged hadrons (π, K, p) in proton–lead (pPb), lead-proton (Pbp) and proton-proton (pp) collisions at a nucleon–nucleon centre-of-mass energy of $\sqrt{s_{NN}} = 5$ TeV using the LHCb detector. These measurements allow us to understand how the presence of a heavy nucleus, like lead, modifies the production of particles compared to the proton–proton case. In particular, they provide sensitivity to nuclear parton distribution functions. The LHCb detector plays a unique role here, as it can explore regions of momentum and angle (phase space) that are not accessible to other experiments at the LHC.

The specific objectives of this thesis are:

- **Cross-section measurements**: Determine the production rates of pions, kaons, and protons, expressed as double-differential cross-sections. These are measured as a function of the transverse momentum of the particle (p_{T}) and pseudorapidity (η_{cms}).

- **Nuclear modification factor**: Calculate the nuclear modification factor, $R^{h}_{p\mathrm{Pb}}$, for each hadron species. This observable compares particle production in pPb collisions to the expectation from pp collisions, and is measured separately in the forward and backward rapidity regions. Deviations from unity reveal the influence of nuclear effects.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

- **Particle production ratios**: Study ratios such as p/π, K/π, and K/p. These quantities are less sensitive to experimental uncertainties and provide insight into hadronisation as well as the abundance of strange quarks.

- **Input for theory**: Provide new experimental results that can be used to test theoretical descriptions of hadron production in cold nuclear matter, and to improve global fits of nuclear parton distribution functions.

2.2 Hypotheses

Building on the objectives described above, this thesis tests how the nuclear environment affects the production of charged hadrons. The following hypotheses summarize the main physical expectations that guide this work:

- **Modification of hadron production in nuclear collisions:** When a proton collides with a lead nucleus, the quarks and gluons inside the nucleus are distributed differently than in a single proton. As a result, the number of produced hadrons (pions, kaons, and protons) is expected to differ from that observed in proton–proton collisions. These changes can arise from nuclear effects such as the modification of parton densities (shadowing or anti-shadowing), multiple scattering, or energy loss of partons as they traverse the nucleus.

- **Dependence on hadron type:** The size of these nuclear effects is not necessarily the same for all hadron species. Pions, kaons, and protons are formed from quarks of different types and through different hadronisation processes. Therefore, the degree of modification may depend on the particle type, reflecting the role of strange quarks and baryon production mechanisms.

- **Ratios as clear probes of nuclear effects:** Ratios of particle yields, such as K/π, p/π, and K/p, provide a clean way to study differences between nuclear and proton–proton collisions. Because these ratios are less sensitive to many experimental uncertainties, any deviation from the pp reference would indicate an effect related to the nuclear environment or the hadronisation process.

2.3 Methodology

The analysis strategy is designed to extract reliable measurements of identified hadron production from the raw collision data, while controlling all relevant detector effects and uncertainties. It relies on the precise tracking and particle identification capabilities of the LHCb detector, combined with a series of calibration and correction procedures. The overall workflow is summarised below.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Data sample

The analysis uses pPb collision data collected by the LHCb experiment during the 2016 run of the Large Hadron Collider. Data were recorded in both beam configurations, with protons or lead nuclei travelling towards the detector. This allows access to both forward and backward rapidity regions.

Track and particle selection

Prompt charged hadrons are defined as particles produced either directly in the primary pp or pPb collision or through the decay of very short-lived resonances. Secondary particles originating from weak decays or from interactions with the detector material are excluded. Tracks are required to satisfy strict quality criteria and to fall within the geometrical acceptance of the LHCb detector.

Particle identification

Pions, kaons, and protons are separated using the Ring Imaging Cherenkov detectors. The particle identification (PID) performance is calibrated using clean samples from well-known decays:

- $K^0_S \to \pi^+\pi^-$ for pions,
- $\Lambda \to p\pi^-$ for protons,
- $\phi \to K^+K^-$ for kaons.

These samples are used to determine PID efficiencies and misidentification probabilities as a function of the kinematic variables relevant to this analysis.

Yield determination and corrections

Raw yields of pion, kaon, and proton candidates are measured in bins of transverse momentum and pseudorapidity. Corrections are applied to account for:

- Background contributions from secondary particles and misreconstructed tracks.
- Particle misidentification using the response matrix derived from calibration samples.
- Selection and reconstruction efficiencies, including detector effects.

The formal definitions and formulas for these corrections are provided in Chapter 9.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Systematic uncertainties

Each correction step introduces potential systematic uncertainties. The dominant sources are the integrated luminosity, tracking efficiency, particle identification calibration, and background subtraction. These are quantified individually and combined while accounting for correlations.

Final observables

After applying all corrections and evaluating uncertainties, the analysis yields the fully corrected measurements of:

- Production cross-sections of pions, kaons, and protons in bins of transverse momentum and pseudorapidity,
- Particle production ratios such as p/π, K/π, and K/p,
- Nuclear modification factors $R^{h}_{p\mathrm{Pb}}$ for the different hadron species.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

3

Introduction

Quantum Chromodynamics (QCD) is the fundamental theory describing the strong interaction, governing the behavior of quarks and gluons and their confinement within composite particles known as hadrons, such as protons and neutrons [1]. While QCD provides a robust framework for predicting particle production in "hard" interactions, characterized by large momentum transfer, the understanding of "soft" interactions, involving small momentum transfer, remains a significant theoretical challenge. These soft processes dominate many aspects of hadronic collisions, making their accurate description crucial for various fields.

To probe the intricate structure of hadrons and explore the properties of QCD, ultra-relativistic collisions are performed at particle accelerators like the Large Hadron Collider (LHC) at CERN. The LHC, the world's largest hadron accelerator, facilitates proton-proton (pp) and proton-lead (pPb) collisions at unprecedented energies. Situated at interaction point 8 of the LHC, the LHCb experiment [2] is a forward spectrometer uniquely designed to detect charged particles produced at low angles relative to the colliding beam trajectories. In these collisions, identified charged particles such as pions ($\pi^{\pm}$), kaons ($K^{\pm}$), and protons (p) are among the primary products. While the production of low-momentum charged particles is largely governed by soft interactions, making first-principles predictions challenging, their precise measurement is indispensable. This extends beyond fundamental particle physics, impacting areas like cosmic ray physics, where accurate modeling of hadronic showers is vital for interpreting observations from facilities such as the Pierre Auger Observatory [3]. Addressing these challenges often relies on refining Monte Carlo models with experimental data [13].

The complexity observed in pp collisions is further amplified in collisions involving heavy ions. In these heavy-ion collisions, particularly lead-lead (PbPb) collisions, a deconfined state of quarks and gluons, the Quark-Gluon Plasma (QGP), is believed to form [4]. The formation of QGP significantly influences particle production, often leading to a

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

suppression of yields relative to pp collisions. To precisely interpret the experimental signatures from PbPb collisions and isolate the effects of QGP, proton-lead (pPb) collisions serve as a crucial baseline. In pPb collisions, while large-scale QGP formation is generally limited, the production of charged particles is notably affected by the intrinsic differences in the partonic structure of the lead nucleus compared to that of a free proton [5]. These modifications, known as Cold Nuclear Matter (CNM) effects, include phenomena like shadowing, anti-shadowing, and the Cronin effect [6]. A deep understanding of CNM effects is paramount, as they are superimposed on QGP signatures in ion-ion (AA) collisions. Theoretical models often rely on the hard part of the charged particle spectra, which can be treated perturbatively, to characterize these nuclear effects in pA collisions, but require stringent validation with experimental data.

This thesis addresses the aforementioned challenges by presenting a precise measurement of identified charged pion, kaon, and proton production rates in high-energy pp and pPb) collisions at the LHCb experiment. Building on the insights from previous inclusive charged particle and neutral pion measurements at LHCb [8, 14], this work aims to provide new constraints on the mass ordering of the Cronin enhancement and contribute to a better understanding of gluon fragmentation functions. The unique forward and backward kinematic coverage of LHCb allows for probing diverse Cold Nuclear Matter (CNM) effects across a wide range of Bjorken x.

The thesis is structured as follows. In chapter 4, an overview of the theoretical background relevant to understanding pp and pA collisions is provided. This includes QCD, QGP phase diagram, kinematic concepts, and a discussion of nuclear effects such as nuclear parton distribution functions, gluon saturation, and the Cronin effect. Chapter 5 describes the Large Hadron Collider and the LHCb detector, detailing its sub-detectors and their roles in particle identification and tracking. Chapter 6 focuses on prompt identified light hadrons, defining prompt particles, discussing kinematic considerations, describing the kinematic coverage of LHCb and presenting previous LHCb measurements of prompt charged particles, alongside the motivation for studying identified pions, kaons, and protons. Chapter 7 covers the preparation and selection of the data and simulation samples, including event selection criteria, candidate selection, and background sources. Chapter 9 describes the analysis techniques used to measure the identified charged particle spectra, including particle identification, track reconstruction, and separating prompt contributions. Chapter 10 discusses the treatment of systematic uncertainties associated with the measurements. Chapter 11 presents the results, including cross-sections and nuclear modification factors, along with comparisons to theoretical predictions. Chapter 12 provides the concluding remarks, summarizing the findings and outlining future research directions.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4

Theoretical Background

4.1 The Standard Model

The Standard Model (SM) is a well-established theory in particle physics that explains most of the observed phenomena involving fundamental particles and their interactions, excluding gravity. It encompasses three of the four fundamental forces: electromagnetic, weak, and strong interactions. The SM categorizes elementary particles, summarized in fig. 4.1, into two groups: fermions and bosons. Fermions, which constitute the matter components, possess a spin of 1/2 and are further divided into leptons and quarks. Each fermion has a corresponding antifermion with opposite charge and quantum numbers, except for mass and spin, which remain unchanged. Fermions are organized into three generations, with each generation containing a pair of quarks and a pair of leptons. Quarks, bearing color and electroweak charges, combine to form hadrons: baryons (composed of three quarks or three antiquarks) and mesons (composed of a quark-antiquark pair). In addition to these conventional states, experimental observations have established the existence of exotic hadronic states such as tetraquarks (two quarks and two antiquarks) and pentaquarks (four quarks and one antiquark) [15, 16]. In contrast, leptons only carry electroweak charges. Bosons, which mediate the fundamental forces, possess a spin of 1. Gluons are responsible for the strong interaction between color-charged particles. Photons mediate electromagnetic interactions between electrically charged particles. The $W^{\pm}$ and Z^0 bosons govern weak interactions among particles with weak charge. The discovery of the scalar Higgs boson, a spin-0 particle, in 2012 by the Large Hadron Collider (LHC) provided evidence for the Higgs field, which imparts mass to other elementary particles [17–20].

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

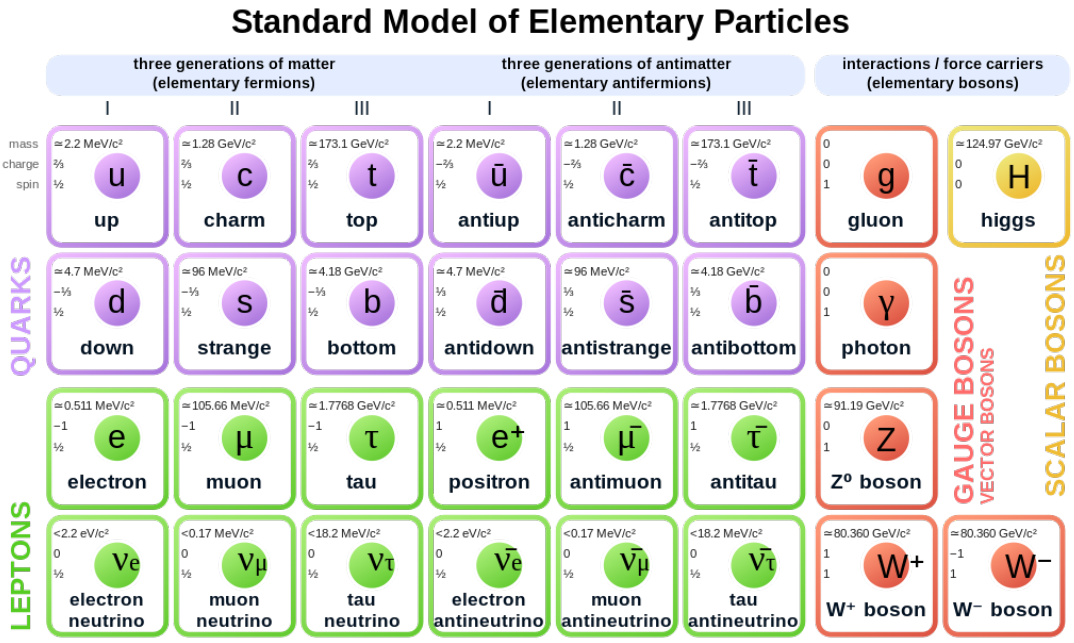


Figure 4.1: Elementary particles in the Standard Model. Figure taken from [21] under Creative Commons license.

4.1.1 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the quantum field theory that describes the strong interaction, the force responsible for binding quarks into protons, neutrons, and other hadrons. Quarks possess an internal degree of freedom known as **color charge**, which comes in three types: red, green, and blue. These color charges are not related to actual color but represent labels for quantum states. Because color is a local property, meaning it can vary from point to point in spacetime, the theory must be invariant under local transformations among these three color states. To ensure this invariance, QCD requires symmetry transformations that preserve the probabilities of quantum states, i.e., unitary transformations. Additionally, to avoid introducing an unphysical overall phase, the transformations must have determinant equal to one. Together, these conditions define the mathematical structure known as the **special unitary group in three dimensions**, or SU(3). Since this symmetry acts on the color degrees of freedom of quarks, it is referred to as $SU(3)_{\text{color}}$.

This $SU(3)_{\text{color}}$ symmetry forms the foundation of QCD. It leads to the existence of eight massless gauge bosons, the gluons, each associated with one of the eight generators of the group. Unlike photons in electromagnetism, gluons carry color charge themselves and can interact with one another. This feature, arising from the nonabelian nature of SU(3), gives rise to essential QCD phenomena such as gluon self-interaction, color confinement, and asymptotic freedom.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The behavior of QCD is encoded in its Lagrangian, which is given by:

$$\mathcal{L}_{\text{QCD}} = \sum_{i=u,d,s,c,b,t} \bar{q}_i(i\gamma^\mu D_\mu - m_i)q_i - \frac{1}{4}G^a_{\mu\nu}G^{a\mu\nu}, \tag{4.1}$$

where $\bar{q}_i$ and q_i are the quark fields, D_μ is the covariant derivative:

$$D_\mu = \partial_\mu - ig_s T^a G^a_\mu, \tag{4.2}$$

with g_s representing the strong coupling constant, T^a (for $a = 1, \ldots, 8$) the generators of $\text{SU}(3)_{\text{color}}$, and G^a_μ the gluon fields. The term $\frac{1}{4}G^a_{\mu\nu}G^{a\mu\nu}$ describes the self-interactions of gluons, where $G^a_{\mu\nu}$ is the gluon field strength tensor:

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu, \tag{4.3}$$

and f^{abc} are the structure constants of the $\text{SU}(3)_{\text{color}}$ group [22–24].

A key property of QCD is asymptotic freedom, where quark-gluon interactions weaken at high energy scales due to the running of the strong coupling constant $\alpha_s(Q^2)$. This dimensionless coupling is related to g_s via:

$$\alpha_s(Q^2) = \frac{g_s^2(Q^2)}{4\pi}, \tag{4.4}$$

where Q represents the energy-momentum transfer scale (typically the momentum transfer in a scattering process). The energy dependence of α_s is governed by:

$$\alpha_s(Q^2) \sim \frac{1}{\beta_0 \ln\left(\frac{Q^2}{\Lambda^2}\right)}, \tag{4.5}$$

where $\beta_0 = 11 - \frac{2}{3}n_f$ (n_f: active quark flavors), and $\Lambda \sim 200 - 300$ MeV is the QCD scale parameter. The logarithmic decrease of α_s with Q is confirmed experimentally, as shown in fig. 4.2. The figure demonstrates excellent agreement between theory (including 5-loop calculations) and diverse experimental and computational results (e.g., lattice QCD, deep inelastic scattering, and heavy quarkonia decays). The current Particle Data Group (PDG) average $\alpha_s(m_Z^2) = 0.1180 \pm 0.0009$ serves as the normalization point for the running coupling [25].

4.1.2 The QCD Phase Diagram

The concept of asymptotic freedom in QCD, introduced in section 4.1.1, suggests the existence of the Quark-Gluon Plasma (QGP), a high-temperature phase in which quarks and gluons interact weakly [26, 27]. This state arises when the temperature exceeds a critical threshold T_c, and quarks and gluons behave as nearly free particles due to the weak coupling at high temperatures. The QCD phase diagram, shown in fig. 4.3, maps the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

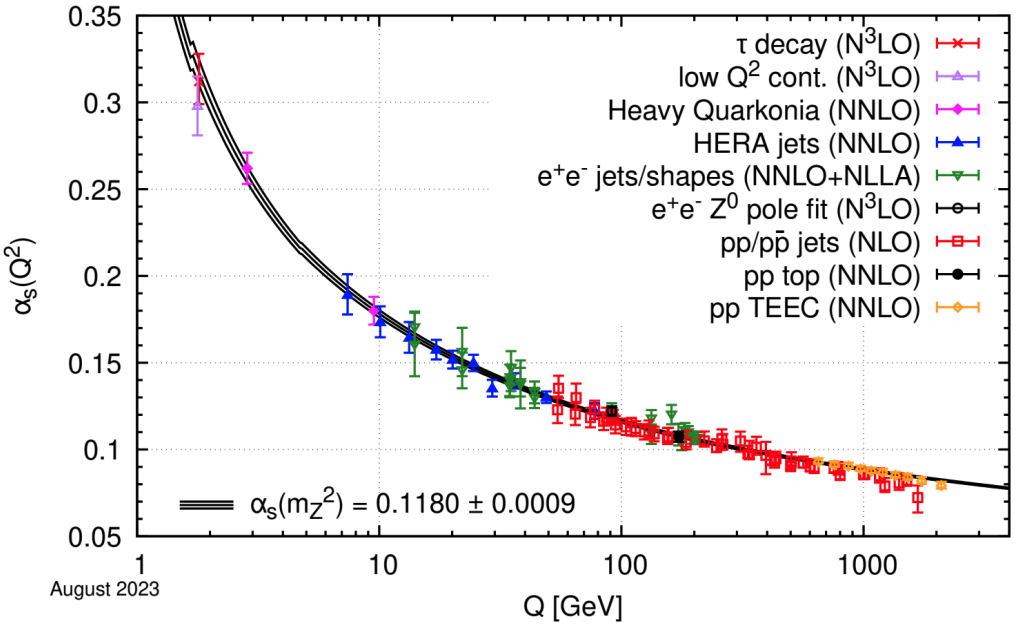


Figure 4.2: Measurements of α_s across energy scales compared to the 5-loop QCD prediction (solid curve), using $\alpha_s(m_Z^2) = 0.1180 \pm 0.0009$ from the Particle Data Group (PDG). Figure taken from [25] under Creative Commons license.

different phases of strongly interacting matter as a function of temperature T and baryon chemical potential μ_B [28]. In this diagram, higher T values correspond to more energetic particles and stronger interactions, while the baryon chemical potential μ_B correlates with the quark density in the system, higher values indicating increased quark concentrations.

At low temperatures and baryon chemical potentials, matter is predominant in the hadronic phase, where quarks are confined within hadrons. As the temperature increases and the baryon chemical potential remains low, matter transitions to the QGP [29], where quarks and gluons become deconfined and interact freely. At low temperatures and high baryon chemical potentials, matter exists in a high-density nuclear phase, which is important for understanding conditions inside neutron stars [30].

The QCD phase diagram includes boundaries that separate different phases, indicating where transitions occur. These transitions can be smooth crossovers, marked by a gradual change in properties, or sharp phase transitions, characterized by abrupt changes in behavior. In particular, studies suggest that at low values of μ_B, the transition from hadronic matter to QGP is a crossover, where the change is quick yet continuous, without a distinct phase separation. The QGP is believed to have existed naturally in the early universe, just microseconds after the Big Bang, when temperatures were sufficiently high to prevent hadron formation [31].

The existence and properties of the QGP are studied experimentally through ultra-

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

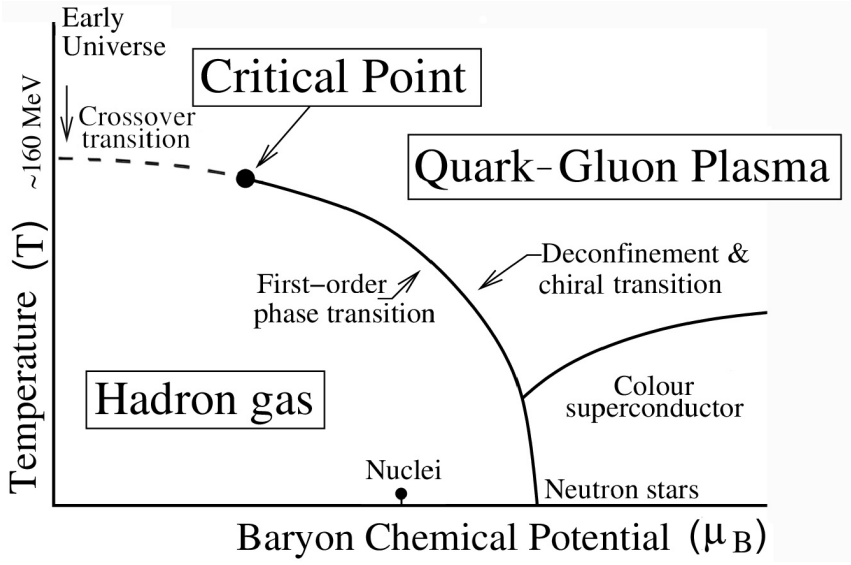


Figure 4.3: Schematic QCD phase diagram showing the phases of strongly interacting matter as functions of temperature T and baryon chemical potential μ_{B}. The dashed curve indicates the crossover transition at low μ_{B}. Figure taken from [32] under Creative Commons license.

relativistic heavy-ion collisions, where temperatures exceeding T_c can be achieved in the laboratory. Signatures of QGP formation, such as jet quenching, strangeness enhancement, and collective flow, are expected to become more pronounced at higher center-of-mass energies, as the system reaches higher temperatures and energy densities (as detailed in section 4.4.1). For instance, gold-gold (AuAu) collisions at the Relativistic Heavy Ion Collider (RHIC) reached a center-of-mass energy of 200 GeV per nucleon-nucleon pair [33], while more recent lead-lead (PbPb) collisions at the Large Hadron Collider achieved 5.02 TeV per nucleon-nucleon pair [34].

4.2 Kinematics of High-Energy Collisions

The energy E and momentum $\vec{p}$ of a particle can be described by the four-momentum vector p^μ that in natural units ($c = \hbar = 1$), is given by:

$$p = (E, \vec{p}) = (E, p_x, p_y, p_z), \tag{4.6}$$

satisfying the relativistic energy-momentum relation $p^\mu p_\mu = E^2 - |\vec{p}|^2 = m^2$. In high-energy collisions, momenta are described relative to the beam axis (z-axis):

$$\vec{p} = (\vec{p_{\mathrm{T}}}, p_z), \tag{4.7}$$

where the transverse momentum is:

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}. \tag{4.8}$$

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

The transverse mass, defined as:

$$m_{\mathrm{T}}^2 = E^2 - p_z^2 = m^2 + p_{\mathrm{T}}^2, \tag{4.9}$$

is a particularly useful quantity as it remains invariant under Lorentz boosts along the beam direction, making it ideal for analyzing transverse dynamics independent of longitudinal motion [35].

The longitudinal dynamics are characterized by the rapidity:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \tag{4.10}$$

which precisely measures the velocity of a particle along the beam axis [36]. However, the experimental determination of rapidity presents significant challenges. First, the energy requires knowledge of the mass of the particle. Second, while modern detectors can precisely measure momentum components (p_x, p_y, p_z), they typically cannot directly determine particle masses on an event-by-event basis. For these reasons, experiments predominantly use the pseudorapidity:

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right), \tag{4.11}$$

where θ is the polar angle relative to the beam axis. Pseudorapidity offers three principal advantages: it is directly measurable from track trajectories in the detector, it is independent of the particle mass and, for relativistic particles ($p \gg m$), it approximates the true rapidity ($\eta \sim y$). The relationship between rapidity and pseudorapidity becomes exact for massless particles. The laboratory frame (detector frame) and center-of-mass (cms) frame are related by a Lorentz boost along the beam axis, resulting in a transformation of pseudorapidity:

$$\eta_{\mathrm{cms}} = \eta + \delta, \tag{4.12}$$

where δ represents the rapidity shift between frames [24]. The pseudorapidity coverage spans from $\eta \to -\infty$ (backward direction, anti-parallel to the beam) to $\eta \to +\infty$ (forward direction), with $\eta = 0$ corresponding to particles produced perpendicular to the beam axis.

The key kinematic variables discussed are illustrated in fig. 4.4. The beam axis defines the z-direction, with the transverse plane (x-y plane) perpendicular to it. The interaction point (IP), where the colliding beams intersect, serves as the origin of this coordinate system. The momentum vector $\vec{p}$ of a produced particle is decomposed into transverse p_{T} and longitudinal p_z components, with θ representing the polar angle relative to the beam axis and ϕ being the azimuthal angle in the transverse plane. The pseudorapidity η, directly related to θ through eq. (4.11), provides a convenient measure of angular distribution that accounts for the cylindrical symmetry of collider experiments. The coordinate system that constitute η, p_T and ϕ, provides a complete description of particle emission patterns in the detector acceptance.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

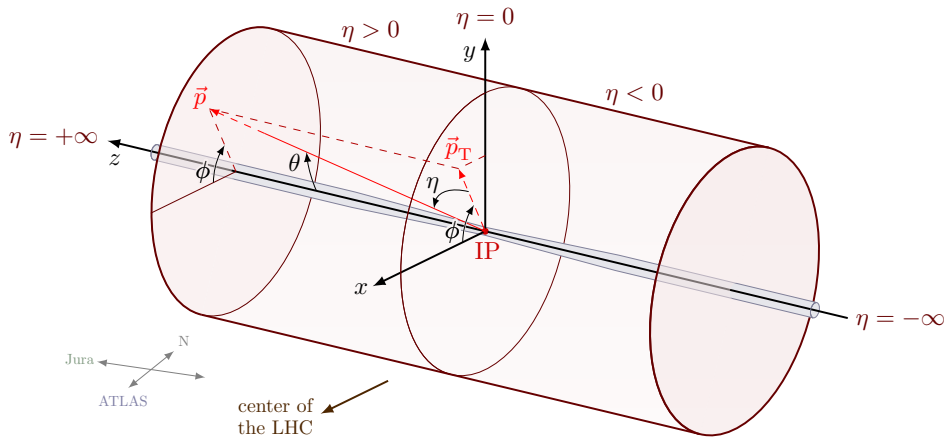


Figure 4.4: Illustration of kinematic variables in collider particle physics. The beamline defines the z-axis, with the transverse plane defined by the x-y axes. The momentum vector $\vec{p}$ is shown, along with its transverse component $\vec{p}_T$ and longitudinal component p_z. The angles θ and ϕ are the polar and azimuthal angles, respectively. Figure taken from [37] under Creative Commons license.

4.3 High-Energy Hadronic Collisions

High-energy hadronic collisions are a crucial systems for probing the fundamental structure of matter and the interactions between its elementary constituents. By accelerating hadrons to ultrarelativistic energies, such as the center-of-mass energies of $\sqrt{s_{\mathrm{NN}}} \sim 1$ TeV achieved at the LHC, these collisions provide access to partonic interactions across a broad range of energy scales.

4.3.1 Elastic and Inelastic Scattering

Hadronic collisions can be divided into two main types: elastic and inelastic scattering. In **elastic scattering**, the colliding hadrons maintain their identity and internal structure throughout the interaction, with no new particles produced. This process provides valuable insights into the fundamental forces that govern hadronic interactions [38]. On the other hand, **inelastic scattering** involves a more dynamic process where the energy from the collision is used to create new particles. This results in a more complex final state. Inelastic scattering provides a framework for studying particle production mechanisms, the formation of hadrons, and the underlying processes that govern the behavior of matter at extremely high energies.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.3.2 Hard Parton Scattering and Subsequent Processes

The primary interaction involves high-momentum partons (quarks/gluons,as introduced in section 4.1) undergoing hard scattering in inelastic collisions, potentially producing heavy particles like top-quark pairs [13]. This hard-scattering process is accompanied by other phenomena that shape the final-state:

- **Initial-state radiation (ISR)** occurs when the incoming partons emit radiation before the hard interaction takes place.
- **Final-state radiation (FSR)**, on the other hand, happens after the interaction, when the outgoing partons emit radiation.

Both ISR and FSR contribute significantly to the energy flow of the events and final-state particle multiplicity, complicating the reconstruction of the original hard-scattering kinematics.

Following the hard interaction, the scattered partons radiate gluons, initiating parton showers. These showers evolve through successive branchings (e.g., $g \to q\bar{q}$ or $q \to qg$) until the partons reach a low virtuality scale, typically $\mathcal{O}(1\ \text{GeV})$ [39]. At this stage, the partons undergo **hadronization**, transitioning into color-neutral hadrons via non-perturbative processes that can be understood with theories such as the Lund string fragmentation model [40]. The resulting hadrons dominate the final-state particles detected in experiments.

In addition to single hard scatterings, hadronic collisions may involve **multiple parton interactions (MPI)**. A notable example is **double parton scattering (DPS)**, where two independent hard interactions occur within the same collision [41]. DPS and other MPI processes enhance particle multiplicity and introduce correlations that should be described in event generators [42].

The **underlying event** (UE) encompasses all activity not directly attributable to the hard scattering, including contributions from ISR, FSR, and MPI. The UE is critical for modeling minimum-bias collisions and jet substructure observables. In high-multiplicity environments (e.g., heavy-ion collisions or high-energy proton-proton collisions), **final-state interactions** such as parton rescattering or hadronic reinteractions may further modify the observed particle spectra. Collectively, these processes, including hard scattering, parton showers, hadronization, MPI and final-state interactions, determine the properties of the final-state particles. Their interplay governs the total cross-section and the event topology in proton-proton collisions.

4.3.3 Classification of Interaction Processes

In hadronic collisions, the interaction processes can be broadly classified into diffractive and non-diffractive categories. This classification depends on the nature of the particle exchange and the characteristics of the final state.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Diffractive interactions: Diffractive processes involve collisions where the incoming hadrons either stay intact or break up into low-mass states. These interactions are mediated by a colorless object, commonly described as a Pomeron ($\mathcal{P}$) in Regge theory [43]. Because no color charge is exchanged, there is no transfer of quantum numbers between the particles. This leads to the formation of a rapidity gap, which is a detector region with no observed particle activity.

Diffractive interactions can be further classified into the following types:

- *Elastic scattering*: both hadrons remain intact and scatter without producing additional particles.
- *Single diffraction (SD)*: one hadron remains intact, while the other dissociates into a particle system.
- *Double diffraction (DD)*: both hadrons dissociate into distinct particle systems, typically separated by a large rapidity gap.
- *Central diffraction (CD)*: a central particle system is produced, with both hadrons remaining intact (also called double Pomeron exchange).

Non-diffractive interactions: In contrast to diffractive, non-diffractive processes involve the exchange of colored objects between the partons of the incoming hadrons. This results in color being transported across the interaction, leading to particle production throughout the detector with no large rapidity gaps. These processes dominate the inelastic cross-section and lead to more significant particle multiplicities.

4.3.4 Factorization Theorem

In high-energy physics, the factorization theorem provides a crucial framework for understanding hadronic collisions, allowing for the separation of short- and long-distance physics. At high energy scales, the strong interaction exhibits asymptotic freedom, as detailed in section 4.1.1, meaning the coupling constant α_s decreases as the energy increases. This makes perturbative Quantum Chromodynamics (pQCD) an effective tool for describing interactions at short distances, where the momentum transfer is large, and the coupling constant is small enough for perturbative techniques to apply. On the other hand, at low energy scales, where the momentum transfer is small, the strong coupling increases, and non-perturbative effects, such as confinement and hadronization (further discussed in section 4.3.2 regarding the formation of final-state particles), become important. These long-distance effects are essential for understanding the internal structure of hadrons, which are composed of quarks and gluons bound together by the strong interaction [44].

The factorization theorem allows for a systematic separation of these short- and long-distance effects in the computation of hadronic cross-sections. It asserts that the total

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

cross-section for the production of a particle in a hadronic collision can be factored into two distinct parts: one that describes the short-distance, high-energy dynamics of the partonic interaction, and another that accounts for the long-distance, low-energy physics governing the distribution of partons within hadrons [45].

To understand the production of particles in hadronic collisions, the cross-section must be broken down into its fundamental components. The inclusive production cross-section of a particle k in a collision between two hadrons, h_1 and h_2, is described by the following equation:

$$d\sigma^{h_1+h_2\to k+X}(\mu^2, Q^2) = \sum_{i,j,X'} f_{i/h_1}(x_1, Q^2) \otimes f_{j/h_2}(x_2, Q^2) \otimes \hat{d\sigma}^{ij\to k+X'}(\mu^2, Q^2). \quad (4.13)$$

Here, the probability of producing a particle k and any associated debris, $X^{'}$, in the collision, is expressed as a convolution of two components:

Parton distribution functions (PDFs): The PDFs, $f_{i/h_1}(x_1, Q^2)$ and $f_{j/h_2}(x_2, Q^2)$, describe the probability of finding a parton i within hadron h_1 and a parton j within hadron h_2. These functions depend on the momentum fractions x_1 and x_2, often referred to as Bjorken-x, which represent how the momentum of the hadron is shared among its constituents. Since PDFs are non-perturbative, they cannot be calculated directly from first principles but are instead extracted from experimental data through fits, as will be further explained in section 4.3.5. The factorization scale Q^2 determines the energy scale at which the parton distribution is evaluated.

Partonic cross-section: The partonic cross-section, $\hat{\sigma}^{ij\to k+X'}(\mu^2, Q^2)$, describes the hard scattering process between partons i and j, resulting in the production of $k + X'$. This quantity is computed using pQCD techniques, which are applicable due to the high energies involved. Calculations up to next-to-leading order (NLO) and next-to-next-to-leading order (NNLO) are considered. The partonic cross-section depends on the renormalization scale μ^2, which regulates divergences arising from high-energy interactions in QCD.

The total cross-section for producing a hadron h is obtained by integrating the contributions from all possible intermediate particles k, along with the contribution from the fragmentation process. This is described by:

$$d\sigma^{h_1+h_2\to h+X}(\mu^2, Q^2, Q_F^2) = \sum_k d\sigma^{h_1+h_2\to k+X}(\mu^2, Q^2, Q_F^2) \otimes D_{h/k}(z, Q_F^2). \quad (4.14)$$

Fragmentation function: The fragmentation function, $D_{h\to k}(z, Q_F^2)$, connects the partonic final state to the observed hadronic state. It defines the probability that a parton fragments into a hadron h with a momentum fraction z. This final-state process is non-perturbative and, like PDFs, must be determined experimentally through fits to data.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

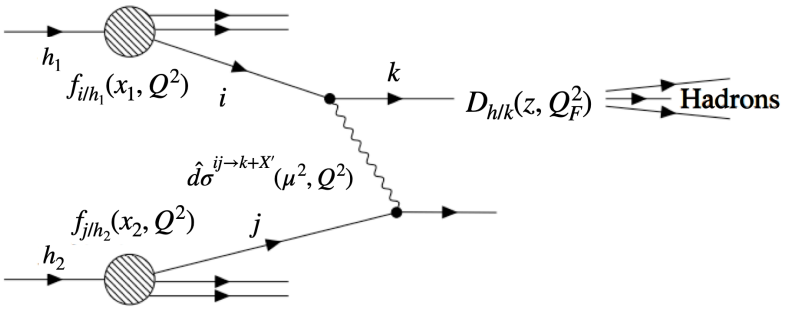


Figure 4.5: Illustration of the factorization theorem for a process $h_1 + h_2 \to h + X$. Figure was adapted from [46] under Creative Commons license.

The fragmentation function depends on the fragmentation scale Q_F^2, which is associated with the energy scale of the hadronization process.

To summarize, the partons inside the colliding hadrons interact. Then, the resulting partons transform into the observed particles, such as pions or kaons, through fragmentation. The process is illustrated in fig. 4.5.

4.3.5 Proton-Proton Collision Physics

Proton-proton (pp) interactions at high energies are governed by the dynamics of partons within the colliding protons. The PDFs describe the momentum distributions of these partons and are essential for predicting scattering cross-sections. As discussed in section 4.3.4, PDFs cannot be calculated directly from first principles in QCD. Instead, they are extracted through global analysis, fitting theoretical predictions to experimental data from processes like deep inelastic scattering and hadron-hadron collisions.

An important feature of PDFs is that they are process-independent. This means that a single set of PDFs can be used across a wide range of processes. This universality is crucial for experimental particle physics, as it enables predictions for various processes, such as Higgs boson production or top quark pair production, using the same set of parton distributions derived from deep inelastic scattering experiments [47]. The global analysis method relies on data from many different types of experiments, including deep inelastic scattering and hadron-hadron collisions. By using a broad range of data, the uncertainties in the PDFs are minimized, leading to more accurate predictions for high-energy processes.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

As the energy scale of the interactions changes, the PDFs also evolve. This evolution is described by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations. These equations model how partons, such as quarks and gluons, split into additional partons as the energy scale increases. This allows the prediction of how the internal structure of hadrons evolves with changes in the energy resolution of the interaction. The DGLAP equations are integral-differential equations, which are solved numerically to obtain PDFs at different energy scales. These numerical solutions account for parton emissions and absorptions, incorporating both theoretical and experimental uncertainties.

One well-known example of this analysis is the CT18 PDF set, developed by the CTEQ-CT18 collaboration [48]. The CT18 set provides PDFs for various partons at different energy scales, including quarks, antiquarks, and gluons, along with uncertainty estimates that reflect both experimental and theoretical sources of error. The CT18 PDFs are employed to analyze hadronic collisions at the LHC and other particle colliders [49]. For example, parton PDFs from the CT18 set can be used to predict cross-sections for particle production at different energy scales. These predictions are essential for understanding hadronic collisions at high-energy accelerators. Examples of parton distributions from the CT18 set are shown in fig. 4.6, with PDFs provided for two energy scales: $Q \sim 2$ GeV and $Q \sim 100$ GeV.

At the lower scale ($Q \sim 2$ GeV), which corresponds to relatively low-energy processes, the proton is primarily made up of valence quarks, specifically u and d quarks, especially at high values of x. At small x the gluon distribution dominates over all other partons. At the higher energy scale ($Q \sim 100$ GeV), gluons become even more dominant at small x and sea quarks start to contribute more significantly. This increase in the gluon contribution is due to the higher energy, which leads to more gluon splittings into quark-antiquark pairs and additional gluons. At large x valence quarks still dominate, but their relative contribution decreases. Another approach for obtaining PDFs is through lattice QCD calculations [50], although these methods result in substantial uncertainties.

The total cross-section for proton-proton collisions can be expressed as the sum of the contributions from all interaction types (detailed in section 4.3.3) as:

$$\sigma_{\text{total}} = \sigma_{\text{elastic}} + \sigma_{\text{inelastic}} = \sigma_{\text{elastic}} + \sigma_{\text{diff}} + \sigma_{\text{ND}} = \sigma_{\text{elastic}} + \sigma_{\text{SD}} + \sigma_{\text{DD}} + \sigma_{\text{CD}} + \sigma_{\text{ND}}. \quad (4.15)$$

An illustration of diffractive and non-diffractive processes in proton-proton collisions can be seen in fig. 4.7. Within this breakdown, non-diffractive interactions contribute dominantly to the overall particle production observed in high-energy experiments, especially for particles produced with high transverse momentum. While diffractive processes contribute significantly to the total inelastic cross-section, potentially up to 40% at LHC energies, their event characteristics typically involve lower momentum transfer and distinct rapidity gaps, resulting in fewer particles produced, particularly at high p_{T}. Conversely, non-diffractive interactions are characterized by more uniform energy deposition across rapidity and a higher multiplicity of produced particles.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

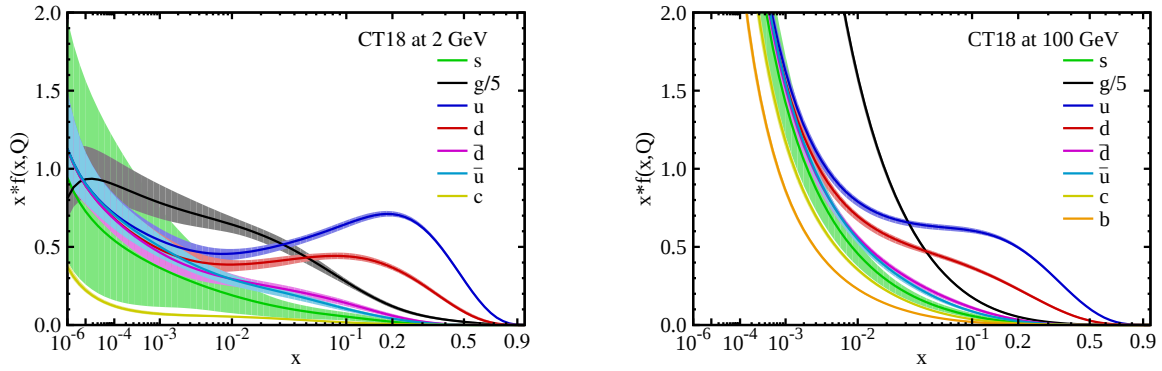


Figure 4.6: The CT18 parton distribution functions at $Q \sim 2\,\text{GeV}$ and $Q \sim 100\,\text{GeV}$ for u, $\bar{u}$, d, $\bar{d}$, s $= \bar{s}$, and g, as functions of the parton momentum fraction x. Figure was adapted from [48] under Creative Commons license.

Distinguishing between diffractive and non-diffractive processes experimentally poses significant challenges, particularly under the high-luminosity conditions of modern hadron colliders, like the LHC, which lead to substantial "pile-up" (multiple proton-proton interactions within a single beam crossing). The experimental signatures of these processes, especially minimum bias events, can overlap. Diffractive events, typically identified by the presence of large rapidity gaps (regions devoid of particle activity), can be obscured by particles originating from additional pile-up interactions. This ambiguity makes a precise and unambiguous separation of diffractive from non-diffractive contributions challenging in many analyses.

Therefore, for the analysis presented in this thesis, the focus will be on generic inelastic interactions, encompassing both non-diffractive processes and the more active diffractive components, where a significant amount of hadronic activity is observed. This approach is justified because non-diffractive interactions represent the primary source of hard scattering processes and the vast majority of particle production relevant to the phenomena under investigation.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

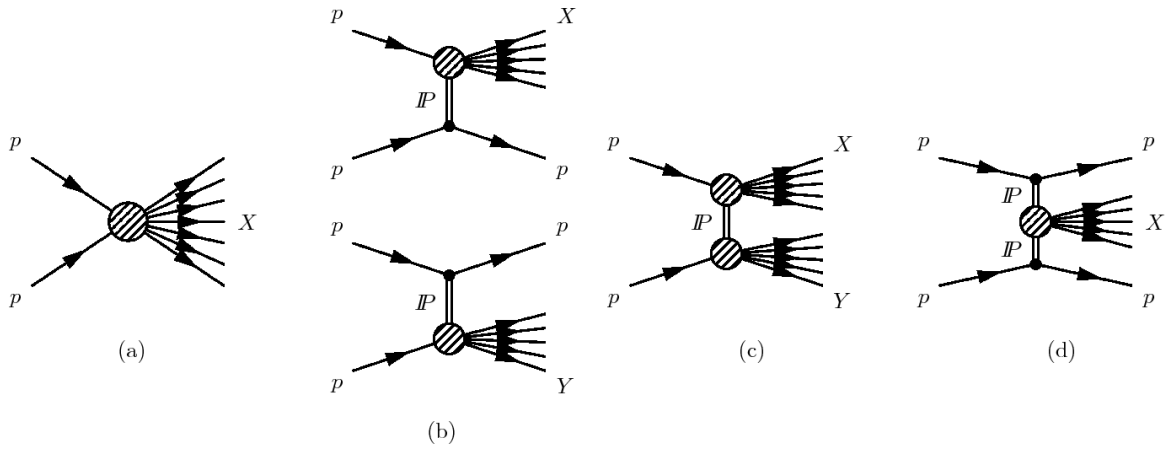


Figure 4.7: Illustration of diffractive and non-diffractive processes in proton-proton collisions. (a) Non-diffractive interaction, where the entire proton energy contributes to particle production. (b) Single diffractive dissociation, where one proton dissociates into a system X, while the other remains intact, mediated by the exchange of a color-singlet object (denoted as $\mathbb{P}$). (c) Double diffractive dissociation, where both protons dissociate into systems X and Y through color-singlet exchanges. (d) Central diffraction, where a central system X is produced via the exchange of two color-singlet objects ($\mathbb{P}$) while both protons remain intact. Figure was adapted from [51] under Creative Commons license.

4.4 Heavy-Ion Collisions and Quark-Gluon Plasma

Ultra-relativistic heavy-ion collisions serve as an indispensable tool for exploring QCD under extreme conditions of temperature and energy density. When two heavy nuclei, such as lead (Pb), collide at nearly the speed of light, they create a unique and transient environment where the fundamental properties of nuclear matter can be probed through a sequence of distinct temporal stages. This intricate evolution, spanning from the initial impact to the final detection of particles, is crucial for understanding the formation and characteristics of the Quark-Gluon Plasma, which was introduced in section 4.1.2 and is expected to be formed in nucleus-nucleus collisions at sufficiently high enough energy. The collision process can be broadly categorized into the following stages, each associated with a characteristic proper time (τ), as depicted in fig. 4.8:

1. **Initial State ($\tau \sim 0$ fm/c):** At the beginning of the collision, two lead nuclei approach each other at relativistic speeds. The immediate overlap of the two nuclei creates a highly compressed region, characterized by an extremely high energy density. This initial interaction sets the stage for the formation of the QGP.

2. **Initial Energy Density Peak and Deconfinement ($\tau \approx 1$ fm/c):** As the collision progresses, the energy density within the interaction zone reaches its maximal value. At this critical point, the constituent quarks and gluons, normally confined

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

within hadrons, may become deconfined due to the immense energy and temperature, initiating the formation of the QGP. In this phase, the system consists of free quarks and gluons rather than hadrons, marking a transition to a new state of matter.

3. **QGP Formation and Hydrodynamic Expansion:** Following deconfinement, the system enters the QGP phase, characterized by the collective, thermalized motion of deconfined quarks and gluons. The evolution of this hot and dense strongly-interacting medium is well-described by relativistic hydrodynamic models, which account for its rapid expansion and cooling over time. This collective behavior provides key insights into the transport properties of the QGP.

4. **Hadronization and Hadron Gas Formation ($\tau \approx 10$ fm/c):** As the QGP continues to expand and cool, its temperature eventually drops below the critical temperature for deconfinement. At this juncture, the system undergoes a phase transition known as hadronization, where the deconfined quarks and gluons recombine and condense to form color-neutral hadrons (e.g., pions, kaons, and protons). This signifies the transition from the QGP to a hadron gas phase, where the newly formed hadrons interact via conventional strong forces.

5. **Kinetic Freeze-out and Free Streaming:** After hadronization, the strong interactions among the hadrons rapidly diminish as the system expands further and becomes more dilute. The system reaches a state of kinetic freeze-out, at which point the elastic interactions between individual hadrons effectively cease, and their momentum distributions become "frozen". From this stage onward, these hadrons propagate freely towards the detector without further significant re-scattering.

6. **Particle Detection ($\tau \approx 10^{15}$ fm/c):** Finally, the freely streaming hadrons reach the detectors of the experimental apparatus. Their trajectories, momenta, and other properties are precisely measured. These data are used to reconstruct the properties of the QGP and the overall dynamics of the collision. This marks the last phase of the collision evolution, where the information gathered provides insight into the earlier stages of the collision.

Each phase in this process provides critical information about the behavior of the strongly interacting matter created in the collision. Understanding these phases is fundamental to study the properties of the QGP and the underlying mechanisms of high-energy nuclear interactions.

A crucial aspect of QGP research involves comparing different collision systems to isolate and understand specific phenomena. As previously detailed in section 4.3.5, pp collisions serve as a fundamental baseline, being largely devoid of both nuclear effects and QGP formation, thus providing a reference for elementary parton-parton interactions.

Proton-nucleus (pA) collisions, in contrast, are instrumental in isolating initial-state effects, such as modifications to nPDFs and early-time dynamics (e.g., cold nuclear matter

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

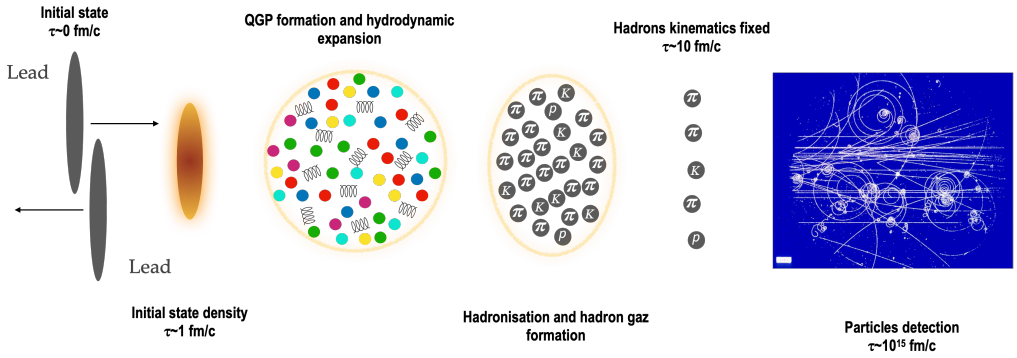


Figure 4.8: Evolution of a PbPb collision at the LHC, showing the transition from initial state to particle detection.

effects), as will be discussed in more detail in section 4.5. By systematically comparing nucleus-nucleus (AA) collisions to both *pp* and *p*A collisions, genuine QGP signatures can be effectively decoupled from initial-state modifications and other cold nuclear matter phenomena, allowing for a clearer understanding of the hot and dense medium properties.

4.4.1 Signatures of the QGP

The formation of the QGP leads to distinct experimental signatures that differentiate heavy-ion collisions from *pp* collisions. These signatures are primarily linked to the strong interactions between energetic partons and the deconfined medium. As a parton traverses the QGP, it experiences significant energy loss through two primary mechanisms:

- **Collisional energy loss**: occurs when the traversing parton exchanges momentum with quarks and gluons in the dense medium, resulting in its energy reduction.
- **Radiative energy loss**: arising from medium-induced gluon radiation, where the emitted gluons further interact with the QGP.

These energy loss mechanisms suppress hadron production at high transverse momentum in heavy-ion collisions relative to *pp* collisions, a phenomenon known as jet quenching. This effect was first observed by the PHENIX collaboration [52]. Furthermore,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

reconstructed jets reveal asymmetric energy distributions in dijet events, consistent with medium-induced gluon radiation and collisional energy loss. These observations establish the presence of a dense colored medium with high opacity.

Anomalous suppression of charmonium, observed by the NA50 collaboration in Pb-Pb collisions at SPS [53], provides further evidence of deconfinement. The phenomenon of quarkonium suppression reflects the effect of Debye color screening on heavy-quark bound states. Lattice QCD predicts sequential melting of charmonium and bottomonium states depending on their binding energies. The threshold-like drop in J/ψ production with increasing transverse energy reflects color screening within the QGP, preventing bound $c\bar{c}$ formation.

Another pivotal observable for probing QGP properties is elliptic flow, offering insights into the early-stage dynamics and anisotropic pressure gradients of the produced medium. Quantified by the second Fourier coefficient of the azimuthal particle distribution, elliptic flow arises from the initial spatial anisotropy of the collision zone in non-central heavy-ion collisions [54]. The collective expansion of the QGP converts this initial spatial anisotropy into a momentum anisotropy, imprinted on the detected hadron distributions. The measurement of large elliptic flow coefficients indicates strong collective motion driven by initial spatial anisotropies in the collision geometry. Hydrodynamic modeling of these flow harmonics suggests that the QGP exhibits an extremely low shear viscosity-to-entropy density ratio. This behavior implies that the QGP is a nearly perfect fluid.

Note that the above signals of QGP formation need to be corrected by any initial state effect, also present in proton-nucleus collisions. Thus, the study of these small-system collisions becomes crucial for the interpretation of QGP signatures.

4.5 Phenomenology of Proton-Nucleus Collisions

In the following the most relevant effects affecting proton-nucleus collisions are presented. Among them, the modification of the nuclear parton distribution functions and saturation effects are considered to be the most relevant initial-state effects. Also, the phenomenological Cronin effect, known to be present in proton-nucleus and also in nucleus-nucleus collisions is introduced.

4.5.1 Nuclear Parton Distribution Functions and Initial State Effects

In the study of heavy-ion collisions, understanding the partonic substructure of the colliding nuclei is essential. The majority of the entropy produced in such collisions originates from the initial stages of the interaction. This initial state plays a critical role in determining the conditions before the possible formation and hydrodynamic evolution of the QGP.

usc
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

For soft particles with transverse momentum $p_{\mathrm{T}} \leq 1\,\mathrm{GeV}/c$, which constitute the dominant contribution to particle yields, the properties of the initial state directly influence event observables, such as the particle multiplicity and the geometry of the collision. For hard probes, particles produced in scatterings with large momentum transfers, the production cross-section depends on PDFs of the colliding nuclei (a concept thoroughly introduced in section 4.3.5).

The assumption of a nucleus as a mere incoherent sum of free proton and neutron PDFs fails to describe experimental data accurately. Deep Inelastic Scattering (DIS) experiments reveal that nuclear effects must be considered. These modifications are encoded in nuclear PDFs (nPDFs), which adjust the free-proton PDF with a nuclear modification factor

$$f_i^{p/A}(x, Q^2) = R_i^A(x, Q^2) \cdot f_i^p(x, Q^2), \tag{4.16}$$

where $f_i^p(x, Q^2)$ is the free-proton PDF for parton species i, and $R_i^A(x, Q^2)$ is the nuclear modification factor encapsulating the nuclear effects. The nPDF for the entire nucleus can be obtained by combining the bound-proton and bound-neutron PDFs:

$$f_i^{\text{nucleus}}(x, Q^2) = \frac{Z}{A} f_i^{p/A}(x, Q^2) + \frac{A - Z}{A} f_i^{n/A}(x, Q^2), \tag{4.17}$$

where Z is the number of protons, and $A - Z$ represents the number of neutrons in the nucleus. The neutron PDF, $f_i^{n/A}(x, Q^2)$, is often derived from the proton PDF using isospin symmetry. The nuclear modification factor, $R_i^A(x, Q^2)$, varies across different regions of Bjorken-x and encapsulates distinct physical phenomena:

Shadowing and anti-shadowing ($x \lesssim 0.2$): At low x, a depletion in the parton densities is observed due to multiple scattering effects. In this regime, nucleons at the periphery of the nucleus absorb parts of the incoming flux, casting a "shadow" on the inner nucleons. This leads to destructive interference, named shadowing. In contrast, at slightly higher x values, constructive interference can result in an enhancement known as anti-shadowing.

EMC effect ($x \approx 0.2 - 0.8$): The discovery of the EMC (European Muon Collaboration) effect [55] revealed an enhancement in parton densities at intermediate x. Various models attribute this to increased pion contributions, modified quark confinement, or changes in nucleon structure within the nucleus.

Fermi motion ($x \gtrsim 0.8$): At high x, nucleons inside the nucleus are not stationary but possess Fermi motion. The convolution of parton momentum distributions with the nucleon motion leads to an observed rise in parton densities.

The modification factor of nPDFs in Pb ($R_i^{Pb}(x, Q^2)$) at a scale $Q \sim 10\,\mathrm{GeV}$ as a function of x for u and d quarks is shown fig. 4.9. This figure illustrates the regions of shadowing, anti-shadowing, the EMC effect, and the Fermi motion. The results are

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

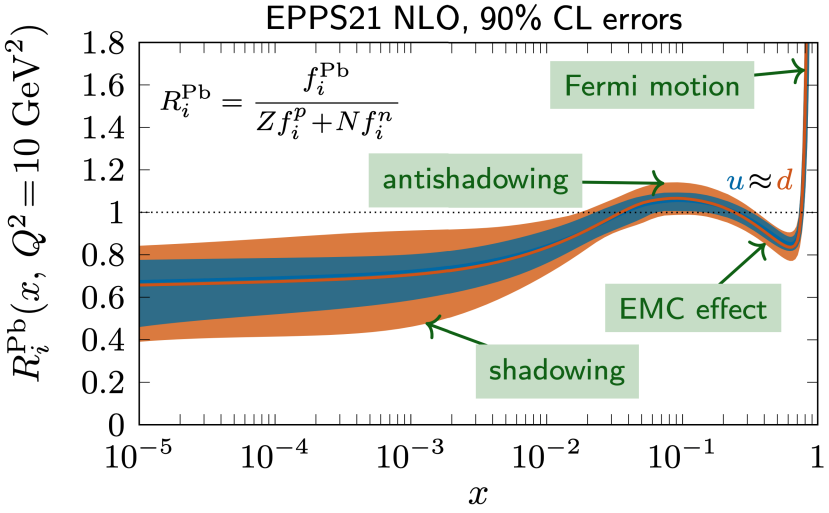


Figure 4.9: The nuclear modification factor $R_i^{\mathrm{Pb}}(x, Q^2 = 10\,\mathrm{GeV}^2)$ as a function of the Bjorken-x, calculated using the EPPS21 NLO nuclear parton distribution functions with 90% confidence level (CL) uncertainties. The definition of R_i^{Pb} is given in the equation, where f_i^{Pb} denotes the PDF for parton i in lead, and Z and N represent the atomic and neutron numbers, respectively. The figure illustrates different nuclear effects: shadowing at small x, anti-shadowing around $x \sim 10^{-1}$, the EMC effect at moderate x, and the Fermi motion at high x. Uncertainty bands for u and d quarks are shown in blue and orange, respectively. Figure taken from [56] under Creative Commons license.

derived from EPPS21 NLO nuclear parton distribution functions, with the associated uncertainties also shown.

4.5.2 Saturation Physics and the Color Glass Condensate

An alternative framework to introduce initial-state effects is the saturation within the Color Glass Condensate. At high energies, especially at small values of x, a new regime emerges where the density of partons in the nucleus becomes so high that non-linear effects become important. This regime is characterized by saturation, which refers to the phenomenon where the parton densities become limited or "saturated", preventing further growth as the energy increases. Saturation is a crucial feature of high-energy QCD that significantly modifies the dynamics of proton-nucleus collisions.

The theoretical framework that describes this high-density regime is known as the Colour Glass Condensate (CGC). The CGC is a state of matter in which the partons inside a nucleus are densely packed, and their interactions are described by non-linear equations, such as the JIMWLK equations [57]. These equations incorporate the effects of gluon saturation and the evolution of the system at high energies. The CGC formalism provides a description of the nucleus as a collection of high-energy gluons that undergo

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

color interactions, leading to a color-neutral configuration at large distances [58].

The saturation scale is a key parameter in this framework and sets the momentum scale at which saturation effects become significant. It increases with energy and decreases with the nuclear size, meaning that larger nuclei have a higher saturation scale, thus modifying the dynamics of parton interactions in proton-nucleus collisions [59].

4.5.3 The Cronin Effect

A key phenomenon observed in proton-nucleus and nucleus-nucleus collisions is the Cronin effect, which refers to the enhancement of particle production at intermediate transverse momenta. This effect is understood to originate from the multiple scatterings of partons as they traverse the nuclear medium before the hard interaction with a nucleon in the target nucleus [6]. These multiple elastic scatterings lead to a broadening of the parton's transverse momentum distribution within the nucleus, effectively increasing the probability of producing high-p_{T} particles compared to a naïve superposition of proton-proton collisions. The Cronin effect is typically observed at intermediate p_{T} (around a few GeV). It provides valuable insight into the dynamics of QCD in nuclear matter and is a distinct cold nuclear matter effect.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.6 Monte Carlo Event Generators

Monte Carlo (MC) event generators play a crucial role in high-energy physics experiments, providing simulated data that are essential for understanding detector responses, developing analysis strategies, and estimating background processes. These generators model both the hard scattering interactions and the subsequent parton showering, hadronization, and underlying event activity, allowing for a detailed comparison between theoretical predictions and experimental observations.

MC simulations are used to model both signal and background processes at hadron colliders such as the LHC. They are also employed in detector design optimization and sensitivity studies. Detector effects are incorporated by interfacing the generator-level output with simulation packages such as GEANT4 [60], which propagates particles through the detector geometry and simulates their interactions with the material.

Beyond collider physics, MC generators find extensive applications in cosmic ray studies. A prominent example is their use in simulating extensive air showers produced by ultra-high-energy cosmic rays, as conducted at the Pierre Auger Observatory. Persistent discrepancies between simulations and experimental data, particularly in muon production yields from air showers, highlight current limitations in modeling hadronic interactions, especially in the forward kinematic region [61].

The parameters of phenomenological models undergo a multi-stage tuning process to experimental data. Initial tuning focuses on hadronization and fragmentation parameters, followed by subsequent optimization of parton shower and multiple parton interaction components. This systematic approach is facilitated by comparison tools like Rivet [62], which enable direct confrontation of generator predictions with unfolded experimental data, allowing for rigorous model validation and iterative improvement.

Several MC generators are available for simulating particle collisions. Among them, Pythia [63] is commonly used for pp collisions, while EPOS [64] is employed in heavy-ion simulations. Both generators are used within the LHCb simulation framework and are employed in the analysis presented in this thesis.

4.6.1 PYTHIA

Pythia [63] is a general-purpose Monte Carlo event generator widely used for the simulation of pp collisions at the LHC. It is specifically designed to model the complex high-energy hadronic scattering processes and subsequent phenomena detailed in section 4.3. These processes involve several key components, which Pythia simulates, and are schematically depicted in fig. 4.10 for a representative $pp \to \bar{t}t$ event:

- **Hard Interaction**: depicted at the center, it represents the hard scattering of partons.
- **Initial- and Final-State Radiation**: shown as outward-directed gluon emissions from both incoming and outgoing partons.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

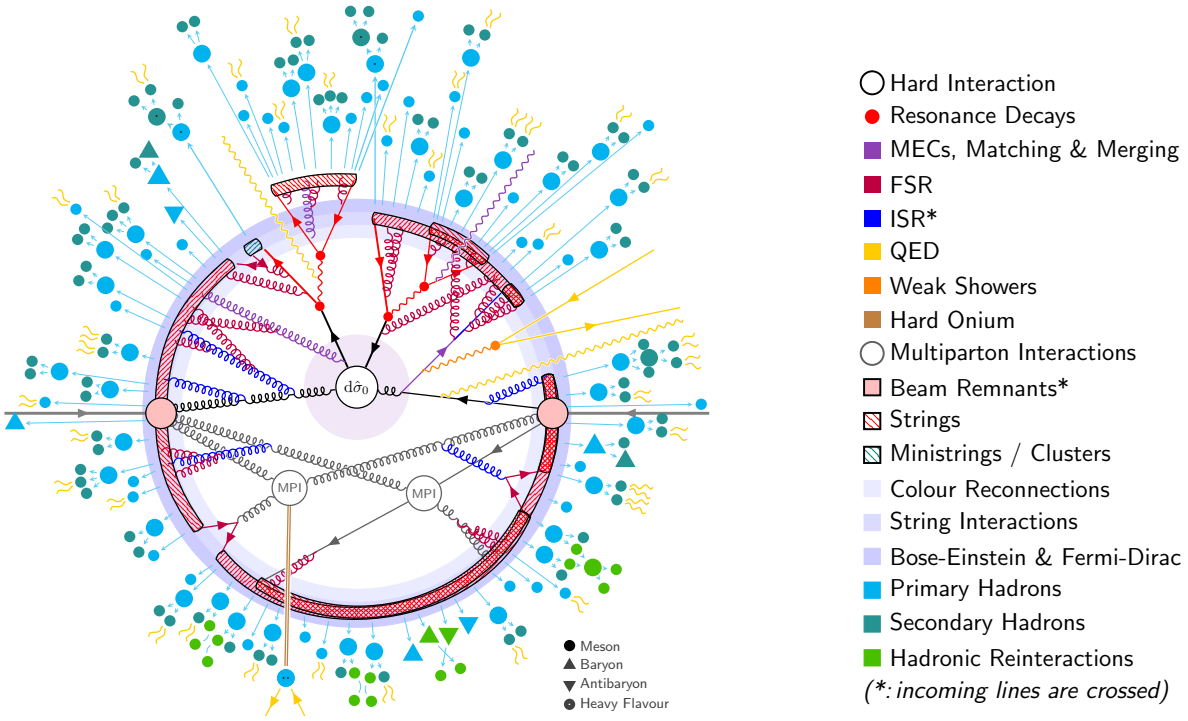


Figure 4.10: Schematic of the structure of a $pp \to \bar{t}t$ event, as modeled by PYTHIA. The figure illustrates hard interactions, parton showers, hadronization, multiple parton interactions, and the role of beam remnants. Figure taken from [65] under Creative Commons license.

- **Multiple Parton Interactions**: additional partonic scatterings, contributing to the overall event structure.
- **Hadronization**: the transformation of partons into observable hadrons, represented by the radiating hadronic lines.
- **Beam Remnants**: residual fragments of the colliding protons, located at the periphery of the figure.

Beyond these core components, PYTHIA specifically includes models for both non-diffractive and diffractive contributions, which are crucial for describing the total inelastic pp cross-section, as introduced in section 4.3.3. Diffractive processes are incorporated via the Ingelman–Schlein model, allowing PYTHIA to provide a comprehensive simulation of the various interaction types.

The phenomenological models within PYTHIA are subject to extensive parameter tuning to ensure agreement with experimental data from various facilities, including the LHC. For LHCb analyses, a dedicated tune is employed to account for the unique forward

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

geometry of the detector and its specific kinematic acceptance. This tuned configuration is crucial for improving the description of event properties relevant to the LHCb physics program, thereby enhancing the precision of simulated samples used in analyses.

4.6.2 EPOS-LHC

EPOS-LHC [66] is a Monte Carlo event generator designed to simulate proton-proton, proton-nucleus, and nucleus-nucleus collisions. While applicable to *pp* interactions, it is particularly suited for modeling dynamics observed in heavy-ion environments. Its design integrates key theoretical and phenomenological approaches, notably the Parton-Based Gribov Regge theory for initial-state effects and a core–corona hadronization approach for final-state collective phenomena.

The Parton-Based Gribov Regge theory describes hadronic interactions through the exchange of multiple parton ladders, forming a central part of the design of EPOS-LHC. This multi-ladder framework accounts for initial-state nuclear effects, such as modifications to nPDFs, including shadowing and anti-shadowing (as detailed in section 4.5.1). It is also formulated to describe phenomena related to gluon saturation, a high-density QCD regime elaborated in section 4.5.2. Each individual ladder incorporates a hard scattering process, calculated via perturbative QCD, supplemented by initial- and final-state parton radiation (introduced in section 4.3.2).

For hadronization, EPOS uses a core–corona approach. In this model, formed string segments are classified: those in the lower-density corona hadronize via string fragmentation, following principles similar to the Lund model discussed in section 4.3.2. Conversely, high-density overlapping string segments constitute a "core" that undergoes collective hadronization through a $(3 + 1)$-dimensional hydrodynamic expansion. This hydrodynamic treatment enables EPOS to simulate collective flow phenomena, notably elliptic flow, which serves as a key signature of the QGP (as discussed in section 4.4.1). This capability allows for probing the dynamics of dense systems formed in high-energy pA and AA collisions, and for simulating heavy-ion-like effects in smaller systems.

EPOS is extensively tuned to reproduce minimum-bias data across diverse collision systems and energies. Its framework, which integrates perturbative QCD calculations with models for collective phenomena, provides a tool for investigating initial-state effects like the Cronin effect (section 4.5.3) and final-state QGP signatures, such as jet quenching and charmonium suppression (section 4.4.1).

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5

The LHCb experiment at the LHC

5.1 The Large Hadron Collider

The Large Hadron Collider (LHC), located at the European Organization for Nuclear Research (CERN) near Geneva, Switzerland, is the largest particle accelerator in the world. CERN, established in 1954, is a multinational research facility dedicated to advancing the understanding of particle physics. The LHC is housed in a 27-kilometer tunnel, initially constructed for the Large Electron-Positron Collider (LEP), and is situated between 50 and 175 meters underground.

Protons and ions are accelerated through a series of pre-accelerators before reaching the LHC ring. Protons are first accelerated to 50 MeV in the Linear Accelerator 2 (LINAC2), then to 1.4 GeV in the Proton Synchrotron Booster (PSB), 25 GeV in the Proton Synchrotron (PS), and 450 GeV in the Super Proton Synchrotron (SPS). Finally, they are injected into the LHC, where each beam reaches energies of up to 7 TeV. The LHC employs 1232 superconducting niobium–titanium (NbTi) dipole magnets, cooled to 1.9 K by a liquid helium system, to bend the beams and generate a magnetic field of 8.33 T. An additional 392 quadrupole magnets are used to focus the beams, increasing the collision probability at the interaction points [67].

The proton beams are organized into bunches, each containing about 1.1×10^{11} protons. Bunches are separated by 25 ns, corresponding to a spacing of approximately 7 m, and travel in opposite directions inside two independent beam pipes [68]. When brought into collision, the counter-rotating beams achieve a center-of-mass energy of up to 14 TeV. In addition to proton–proton collisions, the LHC can also accelerate heavy ions, such as lead (^{208}Pb) and xenon (^{129}Xe). Heavy ions follow a similar acceleration chain, beginning with LINAC3 and the Low Energy Ion Ring (LEIR), before being injected into the PS and SPS. The design energy for lead ions corresponds to 2.76 TeV per nucleon.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

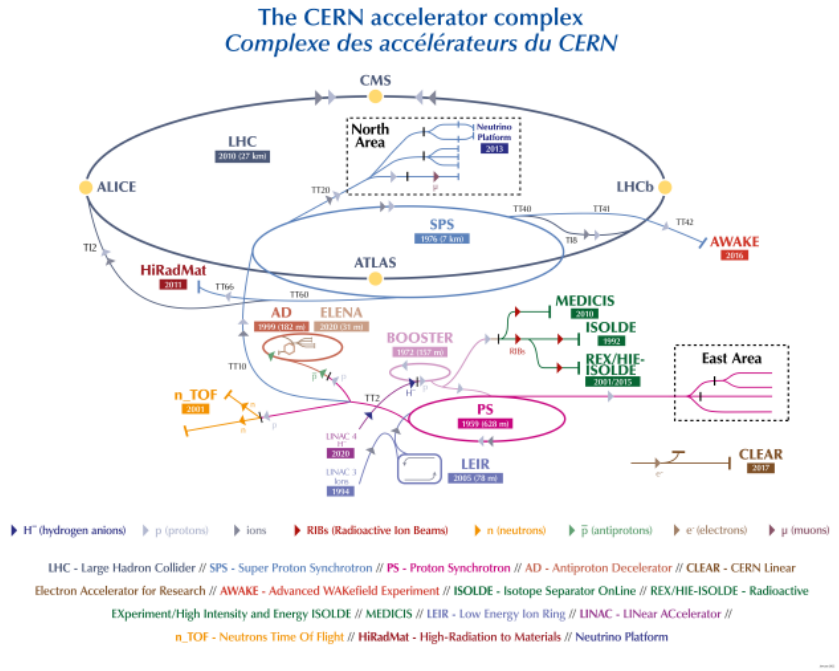


Figure 5.1: The diagram illustrates the layout of the CERN accelerator complex, highlighting the key components: the Large Hadron Collider (LHC), Super Proton Synchrotron (SPS), Proton Synchrotron (PS), and various other accelerators. It shows the flow of different particle beams, including protons, ions, and radioactive ion beams (RIBs), through the accelerator stages. Additionally, it highlights the main experimental areas such as ALICE, ATLAS, CMS, and LHCb. Figure taken from [69] under Creative Commons license.

The accelerator delivers collisions to four main experiments: ATLAS, CMS, ALICE, and LHCb [2, 70–72]. Each experiment is supported by large international collaborations of thousands of physicists, engineers, and computer scientists, who design, construct, and operate the detectors and analyze the resulting data. ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid) are general-purpose detectors designed to explore a wide range of physics, including the origin of mass, the unification of fundamental forces, and searches for dark matter and dark energy. Both experiments played key roles in the discovery of the Higgs boson. ALICE (A Large Ion Collider Experiment) is optimized for studying strongly interacting matter at extreme energy densities through heavy-ion collisions [73]. LHCb (Large Hadron Collider beauty) specializes in precision measurements of b-quark decays, CP violation, and rare processes in hadron physics [74], while also contributing to hadron spectroscopy and heavy-ion studies [75]. A more indepth description of the LHCb experiment is given in the following section 5.2.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5.2 The LHCb Detector

The LHCb detector, shown in fig. 5.2, was originally designed to study the production and decays of heavy-flavor hadrons, in particular those containing b and $\bar{b}$ quarks. Its primary physics program focused on CP violation, rare decays, and searches for physics beyond the Standard Model. Over time, the scope has broadened to include studies of heavy-ion collisions and other forward-physics topics [2].

The detector is located at Point 8 of the LHC and covers a forward angular region of 10–300 mrad horizontally and 10–250 mrad vertically, corresponding to a pseudorapidity range of $2 < \eta < 5$. This acceptance is optimized for the detection of b hadrons, which are predominantly produced at small polar angles in proton–proton collisions. Despite covering only about 4% of the full solid angle, LHCb is able to record roughly 27% of all b and $\bar{b}$ hadrons produced at LHC energies. This distinctive forward geometry sets LHCb apart from the other three main LHC experiments.

A central requirement of the experiment is the precise reconstruction of displaced vertices, crucial for studying B-hadron decays. This involves separating the the primary vertex, where the pp interaction occurs, from the secondary vertex, where the B-hadron decays. To achieve this, the VELO (described in section 5.2.1.1) is placed as close as 8 mm to the beam. Operating under these conditions requires reduced instantaneous luminosity, which quantifies the number of proton-proton collisions per unit area and time at the interaction point, compared to the nominal LHC design. LHCb typically runs at around 4×10^{32} cm^{-2}s^{-1}, maintained through a process called "luminosity leveling," which dynamically adjusts beam parameters to keep the interaction rate nearly constant. This strategy minimizes pile-up[1], thereby preserving the excellent vertex resolution required for B-physics.

The overall performance of LHCb relies on several subsystems, including high-resolution tracking detectors, particle-identification systems, and calorimeters, which together provide efficient event reconstruction. The detector has been continuously upgraded to handle increasing luminosities, with a major upgrade completed in 2023 that enhanced tracking, vertexing, and trigger capabilities [77–80]. Future upgrades will further extend its capacity to operate at higher interaction rates. The following sections provide a detailed description of the individual subsystems and their contributions to the overall performance of the LHCb detector.

5.2.1 Tracking System

The tracking system of LHCb is designed to reconstruct the trajectories of charged particles and measure their momentum. The system consists of several sub-detectors, including the VELO, the magnet, the silicon tracker, and the outer tracker. Each of these

[1] Pile-up refers to multiple proton–proton interactions occurring in a single LHC bunch crossing, leading to overlapping events in space and time.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

components contributes to the overall process of track reconstruction and momentum determination.

5.2.1.1 Vertex Locator

The Vertex Locator (VELO) is a critical component of the LHCb detector, surrounding the interaction point and designed to operate extremely close to the LHC beam. Its unique geometry allows it to reconstruct charged particle tracks emitted both forward and backward with respect to the LHCb acceptance, making it the only subdetector capable of backward track reconstruction. This feature is essential for the precise determination of primary vertices and secondary decay vertices of long-lived particles, such as b- and c-hadrons, which typically decay a few millimeters to centimeters from the PV. The fine spatial resolution of the VELO enables accurate measurements of decay lengths and impact parameters[2].

In addition to vertex reconstruction, the VELO provides seeding information for tracking in downstream subdetectors and serves as a luminosity monitor by using reconstructed tracks and vertices to estimate the instantaneous luminosity during data taking.

The VELO consists of 26 modules per detector half (for Upgrade I), each equipped with silicon microstrip sensors measuring the radial (R) and azimuthal (ϕ) coordinates. The modules are arranged along the beam line over roughly one meter, with a small central opening to allow the LHC beam to pass. The detector halves can retract during beam injection and close around the beam when stable conditions are reached. The sensors operate in a secondary vacuum separated from the primary LHC vacuum by a thin RF foil, minimizing multiple scattering effects [81].

The VELO achieves a hit resolution of 4 μm at optimal track angles, a signal-to-noise ratio of around 20, and a track reconstruction efficiency of about 98% under standard conditions. The transverse and longitudinal primary vertex resolutions are 13 μm and 71 μm, respectively, assuming 25 tracks. High-p_{T} tracks ($p_{\mathrm{T}} > 1$ GeV/c) have impact parameter resolutions below 35 μm, although high-occupancy events, such as proton-lead collisions, can reduce performance [9].

5.2.1.2 Primary Vertex Reconstruction

The reconstruction of primary vertices (PVs) is a key component of the event reconstruction at LHCb. Using the tracks reconstructed by the VELO detector, the algorithm identifies the spatial location where the majority of tracks in a given event originate. This process involves two stages: seeding and fitting. In the seeding stage, potential PV candidates are identified by searching for clusters of tracks in space. During the fitting stage, a weighted least squares method is employed to refine the vertex position. The fitting procedure prioritizes PVs with higher multiplicity[3], which helps to prevent secondary

[2]The impact parameter is the perpendicular distance between the trajectory of a charged particle and the primary vertex.

[3]Multiplicity refers to the number of charged particle produced in a collision.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

vertices from being falsely identified as primary vertices.

The reconstruction of PVs is an iterative process. Once a PV is fitted, its separation from previously identified vertices is checked to avoid the inclusion of false vertices. This process continues until no new PV candidates are found, ensuring high efficiency and accuracy in vertex reconstruction [82].

5.2.1.3 Magnet System

The LHCb detector incorporates a normal temperature dipole magnet capable of generating a magnetic field with an integrated field strength of 3.6 Tm. This magnetic field bends the paths of charged particles, enabling the determination of their momentum, p. The magnet system consists of a large yoke weighing 1500 tons, accompanied by two coils with a combined mass of 54 tons. These coils are positioned symmetrically within the yoke, designed to accommodate the angular acceptance of the detector in the pseudorapidity range of $2 < \eta < 5$. The systematic effects of the detector are controlled by periodically reversing the magnetic field during operation [83].

5.2.1.4 Silicon Tracker

The silicon tracker consists of two main sub-detectors: the Tracker Turicensis (TT) and the Inner Tracker (IT). The TT is located upstream of the dipole magnet and serves to link the tracks reconstructed in the VELO with those in downstream tracking stations. The Tracker Turicensis (TT) is a planar detector with dimensions of 150 cm width by 130 cm height, covering the full acceptance of the detector. The IT, located downstream of the magnet, consists of three subdetectors with silicon microstrip sensors arranged in a stereo layout. The sensors in the outer layers have vertical strips, while those in the inner layers are tilted at an angle of $\pm 5^\circ$ to improve tracking resolution [84].

5.2.1.5 Outer Tracker

The outer tracker (OT) covers the largest portion of the sensitive area in stations T1-T3. It is composed of straw-tube drift chambers arranged in two layers of drift tubes. Each straw has a diameter of 4.9 mm and is housed in modules that constitute the four layers of each tracking station. The OT serves as the primary tracking detector for particles passing through the region after the magnet, complementing the tracking information provided by the silicon detectors [85].

5.2.1.6 Track Reconstruction

In the LHCb experiment, the reconstruction of charged particle tracks involves associating hits from various tracking detectors with the magnetic field map to deduce the particle trajectories and momenta. This procedure is handled by the LHCb reconstruction software, Brunel. The tracks reconstructed in the spectrometer are categorized into

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

different types based on the subdetector used and the associated trajectory information. These types include VELO tracks, T tracks, upstream tracks, downstream tracks, and long tracks, each serving distinct roles in the overall tracking process [86].

VELO tracks are reconstructed using only the hits from the VELO and typically correspond to tracks with a large angle or tracks originating from particles emitted backwards to the LHCb detector. These tracks are essential for precise primary and secondary vertex reconstruction.

T tracks are formed exclusively from hits in the TT stations. These tracks are typically low-momentum and have poorer momentum resolution due to their limited detector coverage.

Upstream tracks traverse both the VELO and TT stations, generally corresponding to low-momentum tracks that are removed from the acceptance by the magnetic field. These tracks are useful in certain calibration and tagging tasks.

Downstream tracks are reconstructed using hits from the TT and downstream tracking stations, such as IT and OT. These tracks typically arise from particles produced in the decays of long-lived resonances, such as K^0_L or Λ, and have well-defined momentum measurements.

Long tracks are those that traverse the full tracking system, from the VELO to the TT and tracking stations downstream of the magnet. They provide the most accurate momentum determination and are the most commonly used tracks for physics analyses.

Track reconstruction begins by identifying seed tracks in the VELO and T stations, where the magnetic field strength is relatively low. These initial track candidates are then extended by incorporating hits from additional tracking stations. To refine the track trajectory, a Kalman filter[4] is applied. The quality of the track fit is assessed by evaluating the difference between the measured and predicted track positions with simulation, accounting for the uncertainties in the measurements. This difference is quantified using the χ^2 statistic, which is then normalized by the number of degrees of freedom (ndf), the number of independent data points minus the number of fitted parameters. After successful reconstruction, the tracks are stored in Transient Event Store (TES), ready for further analysis.

5.2.2 Particle Identification

In the LHCb experiment, particle identification is accomplished by combining data from multiple sub-detectors, including the Ring Imaging Cherenkov (RICH) detectors, calorimeters, and the muon detection system.

Each sub-detector is designed to identify specific types of particles. The RICH system consists of two detectors, which are responsible for distinguishing pions from kaons. The

[4]The Kalman filter is an algorithm that estimates the trajectory of a particle by combining measurements from different detector layers while considering effects like multiple scattering, energy loss due to ionization, and other detector-related uncertainties.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

calorimeters are primarily used for identifying electrons and photons, while the muon detection system identifies muons.

5.2.2.1 RICH Detectors

The Ring Imaging Cherenkov (RICH) detectors identify charged particles by exploiting Cherenkov radiation, which is emitted when a particle travels faster than the speed of light in a given medium. This radiation forms a cone at a specific angle, θ_c, relative to the particle's trajectory, which depends on the particle's velocity. By measuring θ_c, the particle's identity can be inferred. The Cherenkov angle is related to the particle velocity through:

$$\cos\theta_c = \frac{c}{nv}, \tag{5.1}$$

where c is the speed of light in vacuum, n is the refractive index of the medium, and v is the velocity of the particle. Combining this measurement with the momentum obtained from the tracking system allows the determination of the particle mass:

$$m = \sqrt{E^2 - p^2} = p\sqrt{(n\cos\theta_c)^2 - 1}, \tag{5.2}$$

where the velocity of a particle has been expressed as $v = p/E$ in natural units ($c = 1$). The precision of this particle identification depends on the accuracy of both the Cherenkov angle measurement and the momentum determination from the tracking system [88]. The measured performance of the LHCb RICH system, showing the reconstructed Cherenkov angle as a function of track momentum in the C_4F_{10} radiator, is presented in fig. 5.5. This figure illustrates how different particle species can be distinguished across a broad momentum range. The RICH detectors use optical systems composed of spherical and flat mirrors to focus Cherenkov light onto Hybrid Photon Detectors (HPDs), which detect photons across a wavelength range from 200 nm to 600 nm. The LHCb RICH system comprises two detectors [89]:

- RICH1, optimized for low-momentum particles (1 GeV/c to 60 GeV/c) using aerogel and C4F10 radiators.
- RICH2, covering higher momenta (15 GeV/c to over 100 GeV/c) with a CF4 radiator.

This configuration allows efficient particle identification over a wide momentum range, essential for flavor tagging and reconstruction of hadronic decays in the LHCb experiment.

In the LHCb experiment, particle identification in the RICH detectors is performed using a log-likelihood algorithm that evaluates all tracks within an event across both RICH1 and RICH2. This method is essential for accurately reconstructing events in high-occupancy environments, particularly when Cherenkov cones from different particles overlap. The algorithm begins by assuming that each track corresponds to a pion, as pions are the most abundant particles in proton-proton collisions.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The event likelihood is initially calculated based on the distribution of photon hits, track information, and their associated uncertainties. For each track, the likelihood is recomputed by systematically varying the particle mass hypothesis (electron, muon, kaon, or proton) while keeping all other tracks fixed. The mass hypothesis that maximizes the event likelihood for a given track is selected.

To improve efficiency, the algorithm prioritizes tracks that exhibit the largest likelihood changes. Once a mass hypothesis of a track stabilizes, it is excluded from further iterations. Typically, the process converges after two iterations, and the final particle identification is determined using log-likelihood differences denoted as $\mathrm{DLL}_{X\pi}$. Using the pion as a reference, since it is the most abundant species, $\mathrm{DLL}_{X\pi}$ is defined as the difference between the log-likelihoods of the particle X ($X \in K, p, e, \mu$) and the pion hypotheses:

$$\mathrm{DLL}_{X\pi} = \log \mathcal{L}_X - \log \mathcal{L}_\pi. \tag{5.3}$$

A positive $\mathrm{DLL}_{X\pi}$ value indicates that hypothesis X is more likely than the pion hypothesis for that track. Differences between other hypotheses can be constructed from these quantities. For example, the log-likelihood difference between the kaon and proton hypotheses can be written as:

$$\mathrm{DLL}_{Kp} = \mathrm{DLL}_{K\pi} - \mathrm{DLL}_{p\pi}. \tag{5.4}$$

In addition to the log-likelihood method, particle identification can be alternatively approached using ProbNN, a machine-learning-based method based on neural networks. ProbNN computes the probability that a track corresponds to a given particle hypothesis (electron, muon, kaon, proton, or pion) by combining information from the RICH detectors and other sub-detectors. The output is a set of probability values in the range [0, 1], one for each hypothesis:

$$\mathrm{ProbNN}(i) \in [0, 1], \quad \text{for each particle hypothesis } i. \tag{5.5}$$

This probabilistic approach provides a more flexible classification and is particularly effective when particle separation is ambiguous. ProbNN is trained on simulated pp data using neural networks, which learn to associate detector responses with particle types based on patterns in the input variables. When applied to heavy-ion collision data, the differences in event characteristics, such as higher particle multiplicities, may affect detector responses and particle identification performance.

5.2.2.2 Calorimeters

The calorimeter system in the LHCb experiment is designed to identify electrons, photons, and hadrons, and to measure their energy and position. The system consists of an electromagnetic calorimeter (ECAL) followed by a hadron calorimeter (HCAL), together with a preshower detector (PS), which helps to reject background from charged pions. The calorimeter system provides input to the hardware trigger, requiring fast readout to select interesting events in real time [90].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The ECAL is a sampling calorimeter composed of alternating layers of lead and scintillator, with a total thickness of 25 radiation lengths, sufficient to fully contain electromagnetic showers. Its energy resolution is given by:

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 1\%, \tag{5.6}$$

where E is the energy in GeV. The HCAL is a sampling calorimeter made of iron and scintillator tiles, optimized for measuring the energy of hadronic showers, with an energy resolution of::

$$\frac{\sigma_E}{E} = \frac{69\%}{\sqrt{E}} \oplus 9\%. \tag{5.7}$$

Both calorimeters are segmented into cells of varying sizes, with finer granularity in regions of higher hit density to maintain good spatial resolution. This segmentation is essential for accurate reconstruction of electromagnetic and hadronic showers, as well as for efficient particle identification and trigger performance.

5.2.2.3 Muon System

The muon system in the LHCb experiment is designed to identify muons and provide input to the trigger system. It consists of five detector stations (M1–M5) covering a total area of 435 m^2 and containing 1380 individual chambers. Most of the stations use multi-wire proportional chambers (MWPCs), while the inner region of station M1 employs triple-GEM detectors to cope with the higher radiation levels near the beam line [91].

The system layout ensures that muons with momenta greater than 6 GeV/c can traverse all five stations. By measuring the hits in multiple stations, the muon system provides information on the trajectory of muons, which can be used to estimate their transverse momentum.

Muon identification is primarily performed using a binary criterion labeled `IsMuon`. The requirement for a track to satisfy `IsMuon` depends on its momentum: higher-momentum particles must traverse more stations to be classified as muons. The system achieves high identification efficiency, though the misidentification probability can reach approximately 10% for lower-momentum particles.

In addition to the `IsMuon` decision, a log-likelihood difference variable, $\mathrm{DLL}_{\mu\pi}$, is used to further improve muon identification by discriminating muons from pions based on the consistency of hits in the muon stations with the muon hypothesis [92].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

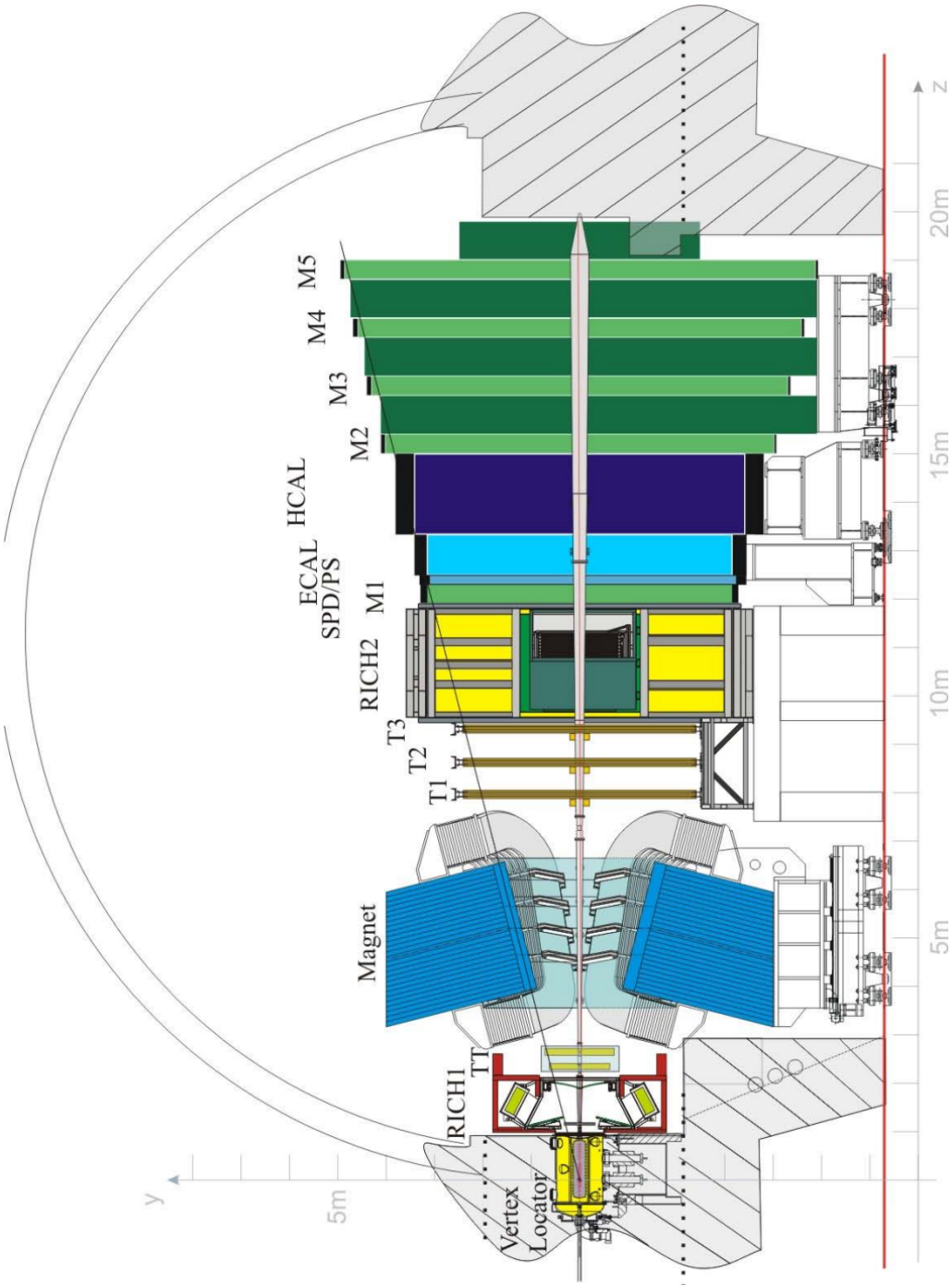


Figure 5.2: Schematic side view of the LHCb detector, with the proton-proton interaction point located inside the Vertex Locator, shown on the left side of the image. In this context, "downstream" refers to detector components located after the dipole magnet along the beam direction (to the right in the image), while "upstream" refers to those located before the magnet. Figure taken from [76] under Creative Commons license.

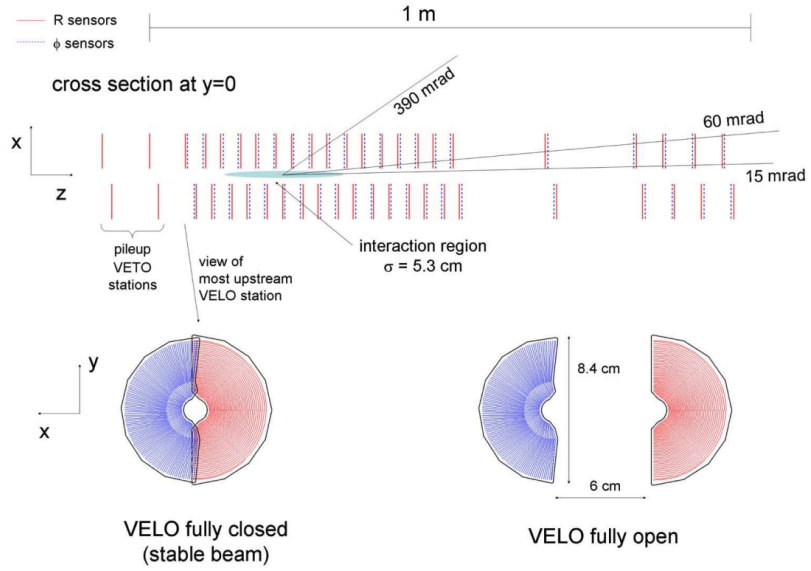


Figure 5.3: Cross-sectional view of the VELO silicon sensors in both fully closed and open positions, with the front face of the first modules and the upstream pile-up veto stations also shown. Figure taken from [2] under Creative Commons license.

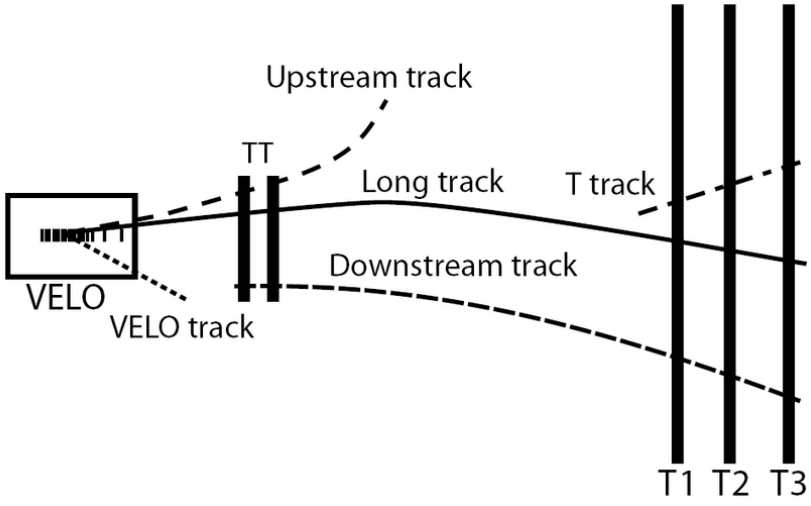


Figure 5.4: Diagram illustrating the five different track types reconstructed in LHCb tracking. Figure taken from [87] under Creative Commons license.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

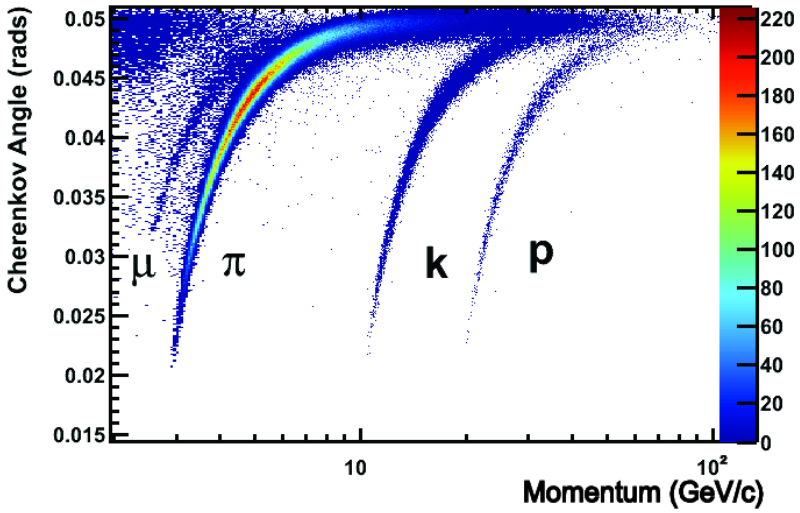


Figure 5.5: Reconstructed Cherenkov angle as a function of track momentum in the C_4F_{10} radiator. Figure taken from [89] under Creative Commons license.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5.3 Data flow at LHCb

The LHC delivers proton-proton collisions at a 40 MHz bunch-crossing rate, which far exceeds the readout and storage capabilities of the LHCb detector. To address this, LHCb uses a real-time trigger system that reduces the event rate to $\sim$ 12 kHz while preserving events of scientific interest. Simulated data is processed through the same reconstruction and selection chains as real data to ensure consistency in physics analyses.

5.3.1 Trigger System

The LHCb trigger system reduces the 40 MHz collision rate to a manageable 3-5 kHz for storage and offline analysis, while retaining events with high physics potential. This is achieved via two stages: the hardware-based Level-0 (L0) trigger and the software-based High-Level Trigger (HLT) [93].

5.3.1.1 Level-0 Trigger

The Level-0 (L0) trigger is implemented in hardware and represents as the initial stage in event selection, reducing the event rate from 40 MHz to approximately 1 MHz. It uses inputs from various sub-detectors, including the calorimeters, the muon system, and the pile-up system in the VELO detector. The L0 trigger decision is based on predefined thresholds applied to calorimeter clusters and muon tracks, allowing the system to filter out background events while retaining those with high physics potential.

The VELO pile-up system additionally helps mitigate combinatorial issues by suppressing high-multiplicity events. The L0 decision is made within a latency of 4 microseconds, with only 2 microseconds available for actual decision-making after accounting for signal propagation and electronics delays. A small fraction of NoBias triggers are also generated for offline studies, providing insight into the bunch filling pattern of the LHC.

5.3.1.2 High-Level Trigger

Following the L0 decision, the High-Level Trigger (HLT) further reduces the event rate through a detailed, software-based selection process. This stage operates on the Event Filter Farm (EFF), a large computing infrastructure. The HLT is highly flexible and can be adjusted to accommodate evolving physics priorities and data characteristics. It is divided into two stages: HLT1 and HLT2.

In HLT1, a partial event reconstruction is performed to validate L0 decisions and identify potential objects of interest. T. These objects are typically selected using loose criteria on track parameters, such as momentum or geometric constraints. HLT1 reduces the event rate from 1 MHz to around 100 kHz, providing a manageable number of events for further processing.

HLT2 performs full event reconstruction, equivalent to the offline reconstruction, and refines the event selection based on detailed tracking and interaction topology. This stage

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

reduces the rate to a few kHz, ensuring that only the most relevant events are recorded for offline analysis.

Each data-taking run is associated with a unique Trigger Configuration Key (TCK), which encodes the specific trigger conditions, including thresholds, prescales, and active trigger lines. The TCK ensures reproducibility and allows offline analyses to reference the exact trigger settings used during data collection.

The combination of L0, HLT1, and HLT2 enables LHCb to handle the high collision rate efficiently while maximizing the physics potential of recorded events, particularly for heavy-flavor decay studies.

5.3.2 LHCb Software Framework

The LHCb offline software framework is a modular system designed to handle the full processing chain from Monte Carlo (MC) event generation to final analysis. The framework has undergone significant updates to incorporate real-time reconstruction within the trigger system, unifying the software used for online and offline processing. The core framework is written primarily in modern C++, with Python employed for configuration and job control.

The simulation workflow begins with event generation using Pythia, POWHEG, or similar tools, followed by decay simulation with EvtGen. Detector interactions are simulated with Geant4, and the digitization step, mimicking the electronic response of the detector, is performed by the Boole package. These components are integrated within the Gauss framework [94].

For real-time and offline reconstruction, the system uses several specialized applications:

- Moore: Configures and runs the HLT, including hit decoding and track reconstruction.
- Allen: Provides GPU-accelerated processing for the first trigger stage, enabling real-time event selection [95].
- DaVinci: Performs offline analysis tasks, including track and particle reconstruction, event selection (stripping), and final physics analysis.

Events are categorized into dedicated streams, allowing efficient handling of different types of data (e.g., physics, calibration, monitoring). The entire framework is subject to rigorous code reviews, continuous integration, and automated testing to ensure stability and reproducibility.

A schematic overview of the complete software workflow, from event generation to analysis, is shown in fig. 5.6.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

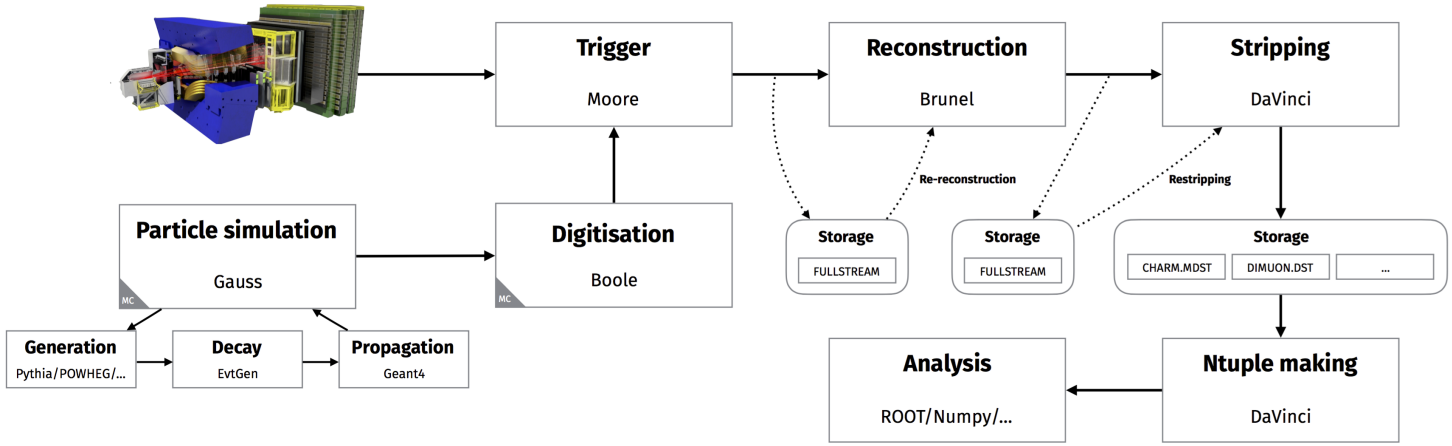


Figure 5.6: Schematic overview of the LHCb offline software framework, illustrating the data flow from event generation (e.g., Pythia, EvtGen) through reconstruction (e.g., DaVinci, Moore) to storage and analysis (e.g., ROOT). Key components include Geant4 for particle propagation, Boole for digitization, and Allen for GPU-accelerated processing. Figure taken from [96] under Creative Commons license.

6

Definition of the Observables and Physics Motivation

6.1 Definition of Prompt Particles

In experimental high-energy physics, particularly in the context of event characterization and track reconstruction, it is important to establish a consistent definition of what constitutes a "prompt particle." Following the recommendations of the LHC Minimum Bias and Underlying Event Working Group [10, 11], a **prompt charged particle** is defined as:

> Any hadron or lepton, with a mean proper lifetime $\tau > 30$ ps, produced directly in the collision or from decays of shorter-lifetime particles.

This definition ensures that particles whose decay products are not distinguishable from those originating directly from the primary interaction vertex are included in analyses focusing on primary production. Conversely, particles with longer lifetimes, which might traverse significant distances before decaying (e.g., K^0_S, Λ), are typically excluded from analyses of "prompt" production and are often referred to as "long-lived" particles. For clarity, throughout this thesis, any subsequent mention of "charged particle" or "prompt charged particle" will strictly adhere to this definition. The widths (Γ) and mean proper lifetimes (τ) of various long-lived particles relevant to this definition are presented in table 6.1.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

		Mean proper lifetime τ	
Specie	**Width Γ GeV**	**(ps)**	**(cm/c)**
p	0	∞	∞
γ	0	∞	∞
K^0	0	∞	∞
e^-	0	∞	∞
n	7.478×10^{-28}	8.861×10^{14}	2.656×10^{13}
μ^-	2.996×10^{-19}	2.212×10^{6}	6.63×10^{4}
K^0_{L}	1.287×10^{-17}	5.148×10^{4}	1543
π^+	2.528×10^{-17}	2.621×10^{4}	785.7
K^+	5.317×10^{-17}	1.246×10^{4}	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

Table 6.1: Width (Γ) and mean proper lifetime (τ) of long-lived particles. Particles with $\Gamma = 0$ are considered stable for the purposes of this definition. The table is structured with truly stable particles listed first, followed by unstable particles sorted by descending τ. The K^0 entry with infinite lifetime is as presented in the source, representing a specific experimental treatment of neutral kaons in the context of prompt particle definition. Data taken from [11].

6.2 Kinematic Shift in Proton-Lead Collisions

In asymmetric collision systems, such as proton-ion (pA) collisions at colliders, the center-of-mass frame does not coincide with the laboratory (LAB) frame. This kinematic asymmetry arises because, while the beam momentum-to-charge ratio (p/Z) is typically matched for both colliding species (e.g., at the LHC), their mass-to-charge ratios (A/Z) differ significantly. This results in the total energy of the ion beam being approximately Z times that of the proton beam ($E_A \approx ZE_p$). Consequently, the CMS system is boosted relative to the laboratory frame. As discussed in section 4.2, the relationship between the laboratory frame and the cms frame is described by a Lorentz boost along the beam axis, which for pseudorapidity results in a constant shift. This boost manifests as a shift (δ) between the laboratory pseudorapidity (η) and the cms pseudorapidity (η_{cms}):

$$\eta_{\mathrm{cms}} = \eta + \delta. \tag{6.1}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

In the ultra-relativistic regime, the pseudorapidity shift δ for a proton-nucleus collision is approximately given by:

$$\delta \approx \frac{1}{2} \ln \left(\frac{A}{Z} \right), \qquad \text{(eq. (4.12))}$$

where A and Z refer to the mass and atomic numbers of the nucleus, respectively.

For proton-lead (pPb) collisions, using $A = 208$ and $Z = 82$ for ^{208}Pb, the calculated shift is $\delta \approx \frac{1}{2} \ln \left(\frac{208}{82} \right) \approx 0.465$. This positive value indicates that the CMS frame is boosted by 0.465 units of pseudorapidity in the direction of the proton beam (conventionally defined as the positive z-direction). Conversely, for lead-proton (Pbp) collisions (where the lead beam defines the positive z-direction), the boost is in the opposite direction, resulting in a shift of -0.465 [97].

6.3 Unique Coverage of LHCb

Understanding the evolution of nPDFs across various kinematic regions necessitates experimental measurements, particularly within the small x regime, where theoretical predictions face significant uncertainties. The explored range of x depends on the experimental acceptance, especially the pseudorapidity in the centre of mass frame. This relationship is quantitatively described by the equation:

$$x \approx \frac{p_{\mathrm{T}}}{\sqrt{s_{\mathrm{NN}}}} e^{-\eta_{\mathrm{cms}}}, \tag{6.2}$$

where p_{T} represents the transverse momentum of the produced particles and $\sqrt{s_{\mathrm{NN}}}$ denotes the centre-of-mass energy. As η_{cms} increases, the corresponding Bjorken x decreases. Conversely, for small or negative values of η_{cms}, the Bjorken x increases. The coverage of various experiments in x and the momentum transfer in the collision squared, denoted as Q^2 in proton-Lead collisions, is shown in fig. 6.1. As depicted, LHCb can access momentum fractions as low as 10^{-6} in the forward region, or pPb, the proton beam pointing towards the LHCb spectrometer, and up to 10^{-2} in the backward region or, Pbp, the lead beam pointing towards the LHCb spectrometer. This makes LHCb a unique laboratory to study the nPDFs.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

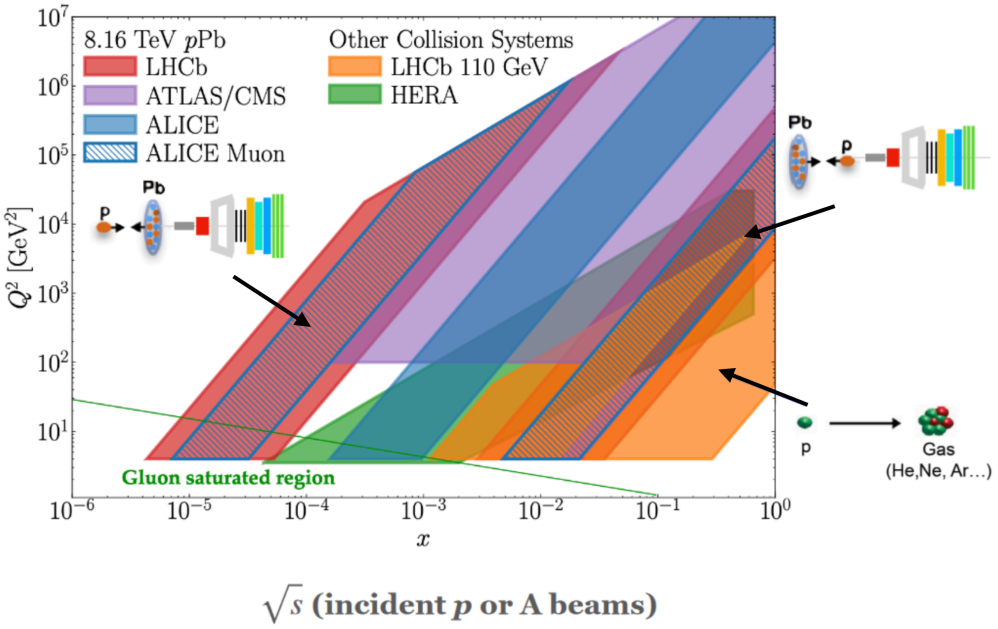


Figure 6.1: Kinematic coverage of LHCb and other experiments for pPb collisions in terms of the fractional longitudinal momentum, x, of the nucleon in the target and the squared parton-parton invariant mass, Q^2. Figure taken from [98] under Creative Commons license.

6.4 Physics Motivation and Key Observables

Measurements of light hadron production in pPb collisions at the LHC provide essential constraints on the mechanisms governing particle production in the presence of nuclear matter. Since the energy density is lower than the one in AA collisions, these studies are crucial for disentangling cold nuclear matter effects from those related to the formation of a QGP observed in ion-ion collisions. In this context, the pPb system serves as a control experiment, isolating nuclear effects that are not related to the hot, dense QGP medium.

Cold nuclear matter effects are commonly quantified using the nuclear modification factor, which compares particle yields in pPb collisions to those in pp collisions, scaled by the atomic mass number of the nucleus. For a hadron species h, it is defined as:

$$R^h_{p\mathrm{Pb}} = \frac{1}{A}\frac{d\sigma^h_{p\mathrm{Pb}}/dp_\mathrm{T}d\eta_\mathrm{cms}}{d\sigma^h_{pp}/dp_\mathrm{T}d\eta_\mathrm{cms}}, \tag{6.3}$$

where A is the atomic mass number of the nucleus (208 for lead), and $d\sigma^h_{p\mathrm{Pb}/pp}/dp_\mathrm{T}d\eta_\mathrm{cms}$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

is the differential cross-section for light hadron production in pPb and pp collisions as a function of transverse momentum and center-of-mass pseudorapidity. The differential production cross-sections can be estimated as:

$$\frac{\mathrm{d}^2\sigma^h_{pp,p\mathrm{Pb}}}{\mathrm{d}p_T\mathrm{d}\eta_{\mathrm{cms}}} = \frac{1}{\mathcal{L}}\frac{N^h(\eta_{\mathrm{cms}}, p_T)}{\Delta p_T \Delta\eta_{\mathrm{cms}}}, \tag{6.4}$$

where N^h is the number of light hadrons in the selected $\Delta\eta_{\mathrm{cms}}$, Δp_T interval, and $\mathcal{L}$ is the integrated luminosity of the analyzed sample (see section 7.1 for a detailed definition).

Complementary information can be obtained from the production ratio of two hadrons, h_1 and h_2:

$$R_{h_1/h_2} = \frac{N^{h_1}(\eta_{\mathrm{cms}}, p_T)}{N^{h_2}(\eta_{\mathrm{cms}}, p_T)}, \tag{6.5}$$

which probes the relative production of pions, kaons, and protons. Specifically, the proton-to-pion ratio reflects baryon production relative to mesons, the kaon-to-pion ratio is sensitive to strangeness production, and the kaon-to-proton ratio informs on the balance between strange and non-strange quarks in the final state.

Experimental Context and Previous Measurements

The study of CNM effects in pPb collisions has been a central objective of the LHC heavy-ion program. The LHCb, ALICE, and CMS experiments have each provided complementary results covering distinct rapidity regions and kinematic ranges.

At midrapidity, ALICE has performed measurements of identified hadron production in pPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV [99]. The nuclear modification factors for pions, kaons, and protons show a characteristic pattern: $R_{p\mathrm{Pb}}$ is close to unity at high p_{T} (> 8 GeV/c), indicating the absence of large nuclear effects, while a moderate enhancement is observed at intermediate p_{T} ($2 < p_{\mathrm{T}} < 6$ GeV/c), especially for protons. This enhancement suggests transverse momentum broadening or collective radial flow even in small collision systems. Moreover, the p/π and K/π ratios measured by ALICE increase at intermediate p_{T}, consistent with observations in PbPb collisions.

At forward and backward rapidities the LHCb collaboration have provided detailed insight into these nuclear effects. The nuclear modification factor has been measured for neutral pions (π^0) at $\sqrt{s_{\mathrm{NN}}} = 8.16$ TeV in forward ($2.5 < \eta_{\mathrm{cms}} < 3.5$) and backward ($-4.0 < \eta_{\mathrm{cms}} < -3.0$) rapidity regions [7], covering $1.5 < p_T < 10.0$ GeV/c. Complementary measurements of inclusive charged hadrons ($h^{\pm}$) at $\sqrt{s_{\mathrm{NN}}} = 5$ TeV were performed in similar rapidity intervals [8], allowing a direct comparison between particle species.

At forward rapidity, both π^0 mesons and charged hadrons exhibit significant suppression, consistent with shadowing or gluon saturation effects in the low-x region of the nucleus. The similar suppression patterns suggest that these low-x effects affect light and heavier hadrons alike.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

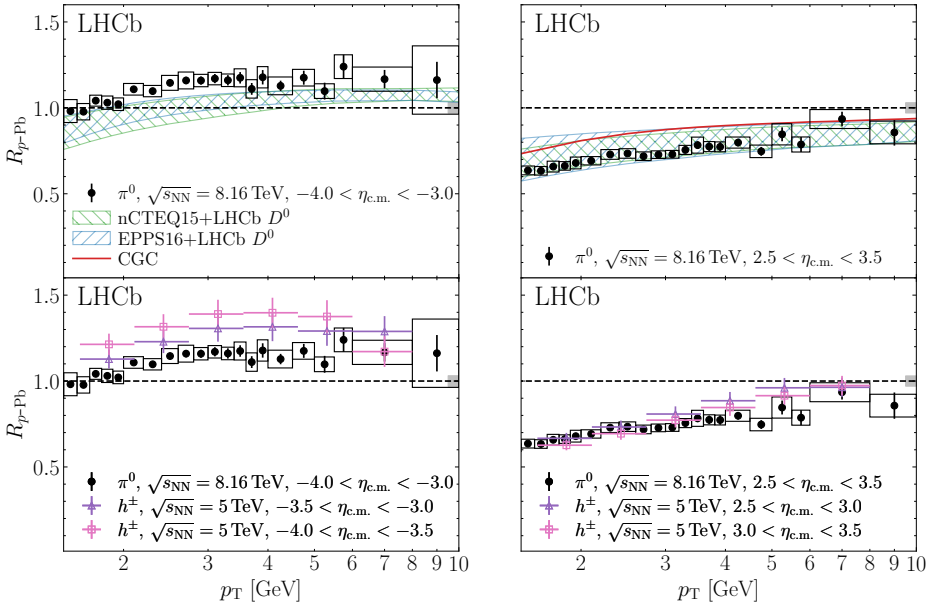


Figure 6.2: Nuclear modification factor ($R_{p\text{Pb}}$) for neutral pions (π^0) and inclusive charged hadrons ($h^\pm$) in pPb collisions at LHCb. The top row (left and right panels) shows π^0 results at $\sqrt{s_{\text{NN}}} = 8.16$ TeV in the backward ($-4.0 < \eta_{c.m.} < -3.0$) and forward ($2.5 < \eta_{c.m.} < 3.5$) [7] rapidity regions, respectively. Theoretical predictions from nCTEQ15+LHCb D^0 [49], EPPS16+LHCb D^0 [100], and CGC [101] models are shown for π^0. The bottom row (left and right panels) compares the π^0 results with inclusive charged hadron ($h^\pm$) measurements at $\sqrt{s_{\text{NN}}} = 5$ TeV in similar backward and forward [8] rapidity regions. The x-axis represents transverse momentum (p_{T}) in GeV. The y-axis represents $R_{p\text{Pb}}$.

At backward rapidity, an enhancement is observed for both particle species, although it is stronger for charged hadrons than for π^0 mesons. This backward enhancement is commonly interpreted as a Cronin-like effect, resulting from transverse momentum broadening due to multiple scattering within the nucleus. The stronger effect for charged hadrons may also reflect mass-dependent effects, such as radial flow or baryon enhancement from recombination. Overall, these measurements probe parton momentum fractions from $x \sim 10^{-6}$ to $x \sim 10^{-1}$, covering both the gluon saturation regime at very low x and the antishadowing region at intermediate x.

The comparison of neutral-pion and charged-hadron $R_{p\text{Pb}}$ values, illustrated in fig. 6.2, highlights both species-dependent differences and distinct forward/backward behavior. Understanding these patterns is essential for building a complete picture of cold nuclear matter effects. Future studies of identified hadrons, including protons and heavier unflavored mesons such as η and η', will help disentangle the different mechanisms contributing to the observed nuclear modifications.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Building on these measurements, this thesis focuses on identified charged hadrons ($\pi^{\pm}$, $K^{\pm}$, and p) to probe nuclear effects beyond inclusive measurements. Studying these hadrons allows the investigation of:

- **Mass-dependent effects (radial flow):** Different hadron masses ($m_{\pi} \approx 0.14$, $m_K \approx 0.49$, $m_p \approx 0.94$ GeV/c^2) produce characteristic ordering in $R_{p\mathrm{Pb}}$.
- **Hadronization mechanisms:** Ratios such as p/π and K/π reveal the relative contributions of recombination/coalescence versus fragmentation.
- **Flavor-sensitive nuclear effects:** Kaons probe strange-quark modifications, while pions and protons probe light-quark dynamics.
- **Initial state effects:** Forward η_{cms} accesses low-x partons sensitive to shadowing or saturation, while backward η_{cms} probes higher-x regions affected by multiple scattering or antishadowing.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

7

Description of the Datasets

7.1 Luminosity

In collider physics, the integrated luminosity ($\mathcal{L}$) is a fundamental measure of the total number of collisions collected in an experiment. It relates to the number of recorded events through:

$$N_{\text{recorded}} = \mathcal{L} \cdot \sigma, \tag{7.1}$$

where N_{recorded} is the number of observed collision events for a given process, and σ is the probability of that process happening, known as the cross-section. Knowing the integrated luminosity accurately is crucial for determining cross-sections with precision.

At the LHCb experiment, the Luminosity Working Group (LWG) is responsible for measuring the integrated luminosity for all data-taking periods. Each run requires independent calibration because beam parameters, such as intensity, size, and position, can vary over time. The LWG carefully accounts for these variations.

A primary method for absolute luminosity calibration at LHCb is the Van der Meer (VDM) scan. This technique measures the effective overlap between the two colliding beams, which directly affects the collision rate. Instead of attempting to measure the complex beam density profiles directly, the VDM scan uses the observable collision rate. During a VDM scan, the beams are systematically displaced relative to each other in both horizontal and vertical directions. By precisely measuring how the collision rate changes with beam displacement, the effective overlap and the transverse beam sizes can be inferred. Larger overlap corresponds to higher collision rates. This method allows a model-independent calculation of luminosity, as it does not require prior knowledge of the exact bunch density functions, making it a robust calibration technique.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The transverse beam sizes (Σ_x, Σ_y) are obtained by performing Gaussian fits to the measured interaction profiles. The luminosity is then given by:

$$\mathcal{L}_{\text{VDM}} = \frac{N_1 N_2 f}{2\pi \Sigma_x \Sigma_y}, \tag{7.2}$$

where N_1 and N_2 are the number of particles per bunch in each beam, and f is the revolution frequency of the bunches. This method provides an absolute, model-independent calibration that is consistent across all LHC experiments.

Even after the absolute calibration, the luminosity can vary over time due to changes in beam conditions. To monitor these variations continuously, **relative luminosity detectors** are employed, such as beam-gas counters and scintillator arrays. These detectors track collision rates during data-taking, and correction factors are applied to account for:

- **Beam current decay:** gradual loss of particles from the beam.
- **Beam spreading:** the beam grows wider over time, reducing collision rates.
- **Detector drifts:** small changes in the response of the monitor over time.

These corrections maintain the luminosity measurement accurate to better than 1% between absolute calibration scans.

Systematic uncertainties in luminosity measurements at LHCb originate from several key factors. These include variations between the calibration scans and real physics runs, detector drifts, and interactions between the beams. Each source of uncertainty is evaluated separately and combined to obtain the total uncertainty. The LHCb experiment always reports these uncertainties alongside the measured luminosity, ensuring the precision of the dataset is well understood [102].

7.2 Datasets

The analysis presented in this thesis uses data from proton-lead collisions collected in 2013 (run I) and a reference proton-proton dataset from 2015 (run II), detailed in section 7.2.1 and section 7.2.3, respectively. Corresponding simulation samples are also considered. The following subsections detail the characteristics, data-taking conditions, and integrated luminosities for each of these datasets.

7.2.1 Proton-Lead Dataset

The proton-lead collision data analyzed in this thesis were collected by the LHCb experiment during the 2013 LHC run, spanning from January 20th to February 10th. The total integrated luminosity accumulated for this period was $\mathcal{L} = 1.7$ nb^{-1} at a center-of-mass energy per nucleon pair of $\sqrt{s_{\text{NN}}} = 5.02\,\text{TeV}$. To account for potential

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

		Stripping Line	ODIN RunNumber	Prescale
Pbp	*MagDown*	`StrippingProtonIonMagDownMinBiasLine`	(136799 – 137045)	0.025
	MagUp	`StrippingProtonIonMagUpMinBiasLine`	(136417 – 136758)	0.110
pPb	*MagDown*	`StrippingIonProtonMagDownMinBiasLine`	(135576 – 136212)	0.080
	MagUp	`StrippingIonProtonMagUpMinBiasLine`	(136237 – 136341)	0.090

Table 7.1: Minimum bias lines from stripping v20r3 which are considered in the analysis. The information was extracted from [106].

detector asymmetries and nuclear effects, the directions of the proton and lead beams were swapped during data-taking. This resulted in two distinct configurations: pPb, where the proton beam was directed towards the LHCb spectrometer (forward rapidity for proton-going particles), and Pbp, where the lead beam was directed towards the spectrometer (backward rapidity for proton-going particles). Furthermore, in both the pPb and Pbp configurations, the polarity of the LHCb magnetic field was reversed, collecting data with the field pointing down (*MagDown*) and up (*MagUp*). Consequently, four different data configurations were acquired: pPb with *MagDown*, pPb with *MagUp*, Pbp with *MagDown*, and Pbp with *MagUp*.

The considered trigger configuration employed for the pPb dataset is a minimum bias, designed to select events with no specific kinematic or topological pre-selection. At the HLT2 level, this trigger was configured in pass-through mode, meaning no event rejection occurred. The L0 trigger acted as an interaction trigger, rejecting empty bunch crossings. The HLT1 stage considered the `Hlt1MBMicroBiasVelo` line, which required the presence of at least one reconstructed track in the VELO detector. This trigger selected every bunch-crossing where a proton-lead interaction occurred without any preferred event feature. Therefore, the trigger efficiency is close to 100%.

Following data acquisition, raw data undergo a stripping process to reconstruct particles and filter events into specific streams. For this analysis, minimum bias stripping lines were applied, with the specific lines and their corresponding prescale factors detailed in table 7.1. These prescale factors, which reduce the number of events written to disk for specific configurations, are crucial and will be explicitly accounted for in the cross-section normalization.

Different Trigger Configuration Keys (TCK) were assigned to the minimum bias trigger in the considered proton-lead data samples. The data collected with TCK 0x6e0049 have a high threshold for the sum of the transverse energy in the HCAL (SumEt), and were not considered. In the data with TCK 0x6f004a the `MBMicroBiasVelo` line is prescaled by a factor of 1/20. Since the other TCKs are not prescaled, they are neither considered. In addition, runs with muon inefficiency problems, according to the online monitoring (some with TCK 0x6a1710), are also excluded from the analysis [103–105]. Details on excluded and included TCKs and runs for each beam configuration and magnet polarity are summarised in table 7.2.

For simulation, the LHCb official samples were generated using the EPOS event

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

	pPb		Pbp		pp
	MagDown	*MagUp*	*MagDown*	*MagUp*	*MagDown*
included TCK	0x006e004a 0x006d0048 0x006e004a	0x006e004a	0x006e004a	0x006e004a	0x0115014e
excluded TCK	0x006a1710 0x006f004a	0x006e0049	0x006f004a		
excluded runs	135710, 135711, 135900, 135901			136729 to 136736	168072, 168128, 168183, 168182, 168143,

Table 7.2: Summary of the included and excluded TCK and runs for every dataset.

generator [64], which is well-suited for producing minimum bias hadronic interactions in pp, pA and AA collisions, where A is a heavy ion nucleus. Following the analysis of prompt charged particles, the simulation version `Sim09e` will be used here. In this MC, events are created with the minimum-bias configuration, and the number of proton-lead interactions in each event is fixed to one. To increase the size of the sample, the `Sim09e` data are merged with a sample produced with a newer simulation version `Sim09k-Multiplicity`. This version was developed to study prompt D^0 meson production and considered enlarged pile-up adding a $D^+ \to K^-\pi^+\pi^-$ decay per event. This sample is used in this analysis eliminating all the tracks originating from the D^+ decay in the event. The merged sample is named `MC-Merged`. The bookkeeping[1] path of each sample is mentioned in table 7.4.

7.2.2 Luminosity Determination in Proton-Lead Data

For the proton-lead datasets, the luminosity is determined using the same general methods described in section 7.1, with adjustments to account for the asymmetric beam configuration. The LWG measures the luminosity on a run-by-run basis, ensuring that each period of data taking is properly calibrated. The recommended detector condition database (CondDB) tag `cond-20141002` and data quality flags (DQFLAGS) `dq-20140822` are used to retrieve the correct luminosity values from DaVinci [102, 103]. The uncertainties, about 2%, include both statistical and systematic contributions. The luminosity values and the minimum bias stripping line prescale factors are shown in table 7.3.

[1] In LHCb, bookkeeping refers to the central data management system that records datasets, storage locations, and processing history.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Beam Magnet	Dataset	Luminosity (nb^{-1})	Uncertainty (%)	Prescale
pPb	*MagDown*	0.849 ± 0.019	2.29	0.080
pPb	*MagUp*	0.1955 ± 0.0045	2.29	0.090
Pbp	*MagDown*	0.2564 ± 0.0064	2.5	0.025
Pbp	*MagUp*	0.2023 ± 0.0051	2.5	0.110

Table 7.3: Luminosity measurements for proton-lead datasets. Uncertainties include both statistical and systematic components.

7.2.3 Proton-Proton Dataset

The proton-proton data collected in November 2015 at $\sqrt{s_{\mathrm{NN}}} = 5\,\mathrm{TeV}$ are used as a reference sample for this analysis. This dataset is characterised by a bunch spacing of 25 ns, ensuring it is free from spill-over effects. No inversion of the magnet polarity was made in the data-taking and the sample was collected with the *MagDown* configuration. NoBias[2] from leading bunch interactions events were selected with the `Hlt1NoBiasLeadingCrossing` line and were assigned the TCK key `0x0115014e`. The reconstruction algorithm applied a selection requiring the tracks to have a ghost probability (`GhostP`) less than 0.4 and a track quality of χ^2/ndf less than 4. These criteria were not applied to the proton-lead data.

For simulation, the LHCb official samples were produced using the Pythia [108] event generator. The events were created with minimum bias[3] conditions with the LHCb official tune `sim09d`. Given that the data sample is only available in the *MagDown* configuration, only *MagDown* events were generated. The paths for the datasets in bookkeeping are summarised in table 7.4.

7.2.4 Luminosity Determination in Proton-Proton Data

For proton-proton datasets, the luminosity is determined by the LHCb Luminosity Working Group using the log-zero method [109]. This method infers the average number of interactions per bunch crossing from the fraction of empty events in a random trigger stream, and normalises the result using the absolute calibration from van der Meer scans, as described in section 7.1. For the 2015 5.02 TeV dataset, only leading bunch crossings were recorded with the `Hlt1/2NoBiasLeadingCrossing` trigger lines. The luminosity is therefore restricted to these crossings and calculated as:

[2]NoBias events are selected with a pass-through approach, meaning no selection requirements are applied at L0, HLT1, and HLT2 trigger lines. Usually taken with a large prescale such that the output rate of HLT2 is low [107].

[3]Minimum bias events are required to pass minimal selections, for example requiring at least one track to be reconstructed in the VELO [107].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 7.4: Bookkeeping path to the data-set used in this analysis for both Data and MC

Data Type	Configuration	Magnet Polarity	Path
Data	Pbp	*MagDown*	/LHCb/Ionproton13/Beam4000GeV-VeloClosed-*MagDown*/Real Data/Reco14r1/Stripping20r3/90000000/ALL.DST
		MagUp	/LHCb/Ionproton13/Beam4000GeV-VeloClosed-*MagUp*/Real Data/Reco14r1/Stripping20r3/90000000/ALL.DST
	*p*Pb	*MagDown*	/LHCb/Protonion13/Beam4000GeV-VeloClosed-*MagDown*/Real Data/Reco14r1/Stripping20r3/90000000/ALL.DST
		MagUp	/LHCb/Protonion13/Beam4000GeV-VeloClosed-*MagUp*/Real Data/Reco14r1/Stripping20r3/90000000/ALL.DST
	pp	*MagDown*	/LHCb/Collision15/Beam2510GeV-VeloClosed-*MagDown*/Real Data/Reco15a/96000000/FULL.DST
MC-Sim09e	Pbp	*MagDown*	/MC/2013/pPb-Beam4000GeV-1580GeV-2013-*MagDown*-Fix1-Epos/Sim09e/L0Trig0x1710/Trig0x406a1710/Reco14r1/Stripping20r3p1/30000000/ALLSTREAMS.DST
		MagUp	/MC/2013/pPb-Beam4000GeV-1580GeV-2013-*MagUp*-Fix1-Epos/Sim09e/L0Trig0x1710/Trig0x406a1710/Reco14r1/Stripping20r3p1/30000000/ALLSTREAMS.DST
	*p*Pb	*MagDown*	/MC/2013/Pbp-Beam4000GeV-1580GeV-2013-*MagDown*-Fix1-Epos/Sim09e/L0Trig0x1710/Trig0x406a1710/Reco14r1/Stripping20r3p1/30000000/ALLSTREAMS.DST
		MagUp	/MC/2013/Pbp-Beam4000GeV-1580GeV-2013-*MagUp*-Fix1-Epos/Sim09e/L0Trig0x1710/Trig0x406a1710/Reco14r1/Stripping20r3p1/30000000/ALLSTREAMS.DST
MC-Sim09d	*pp*	*MagDown*	/MC/2015/Beam2510GeV-2015-*MagDown*-Nu1.5-25ns-Pythia8/Sim09d/Trig0x411501fe/Reco15a/Turbo01aEM/Stripping22NoPrescalingFlagged/30000000/ALLSTREAMS.DST
Sim09k-Multiplicity	Pbp	*MagDown*	/MC/2013/Pbp-Beam4000GeV-1580GeV-2013-*MagDown*-Fix1-Epos/Sim09k-Multiplicity/L0Trig0x1710/Trig0x406a1710/Reco14r1/Stripping20r3p1/21263010/ALLSTREAMS.DST
		MagUp	/MC/2013/Pbp-Beam4000GeV-1580GeV-2013-*MagUp*-Fix1-Epos/Sim09k-Multiplicity/L0Trig0x1710/Trig0x406a1710/Reco14r1/Stripping20r3p1/21263010/ALLSTREAMS.DST
	*p*Pb	*MagDown*	/MC/2013/pPb-Beam4000GeV-1580GeV-2013-*MagDown*-Fix1-Epos/Sim09k-Multiplicity/L0Trig0x1710/Trig0x406a1710/Reco14r1/Stripping20r3p1/21263010/ALLSTREAMS.DST
		MagUp	/MC/2013/pPb-Beam4000GeV-1580GeV-2013-*MagUp*-Fix1-Epos/Sim09k-Multiplicity/L0Trig0x1710/Trig0x406a1710/Reco14r1/Stripping20r3p1/21263010/ALLSTREAMS.DST

$$\mathcal{L} = \frac{\sum_{\mathrm{BX}} \langle \mu_{\mathrm{BX}} \rangle \cdot N_{\mathrm{BX,NoBias}}}{\sigma_{\mathrm{VDM}}}, \tag{7.3}$$

where $\langle \mu_{\mathrm{BX}} \rangle$ is the average number of interactions per bunch crossing and per run, $N_{\mathrm{BX,NoBias}}$ is the number of events selected by the no-bias leading bunch crossing triggers, and $\sigma_{\mathrm{VDM}} = (50.91 \pm 1.0)\,\mathrm{mb}$ is the reference cross-section from van der Meer scans. The full validation of this determination, including cross-checks with alternative datasets and stability studies, is documented in [110]. The total integrated luminosity for the 2015 5.02 TeV dataset is:

$$\mathcal{L}_{\mathrm{tot}} = 3.490 \pm 0.070\ \mathrm{nb}^{-1}, \tag{7.4}$$

with a 2% uncertainty dominated by the VDM calibration. This value is approximately 9% lower than the luminosity previously computed using DaVinci, as that calculation does not properly account for the selection of only leading bunch crossings, whereas the method applied here does.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8

Preparation and Selection of the Data and Simulation

8.1 Event Selection

In the LHCb experiment, there are different bunch-crossing types, depending on whether each bunch is filled with a beam or not. In the analysis, only events produced in bunch-bunch crossings, selected with `BunchCrossingType==3`, are considered. These events are also required to pass the minimum bias trigger, as detailed in section 7.2. However, these requirements do not remove all the background in proton-proton and proton-lead collisions. Other interactions can be wrongly identified as bunch-bunch interactions and be accepted by the trigger system. These include beam interactions with gas molecules within the beam pipe (beam-gas interactions) or with the detector material (beam-splash interactions), both of which have lower energy particles and occur displaced from the beams interaction region. Additionally, spill-over interactions between a nominal bunch and a preceding bunch can appear in proton-lead events. Such effects are less likely in *pp* data due to the longer intervals between bunch crossings, as explained in section 7.2. The reduction of these background events in the analysis is addressed in the two following subsections.

8.1.1 Luminous Region

Beam-gas and beam-splash interactions are spread over the entire VELO vessel. These background events are reduced by considering only events with a reconstructed primary vertex (PV) within a fiducial luminous region. The size of this region is defined analysing the distribution of the PV coordinates $(x_{\mathrm{PV}}, y_{\mathrm{PV}}, z_{\mathrm{PV}})$, where z is parallel to the beam

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

	pPb					
	MagDown			*MagUp*		
	x_{PV}	y_{PV}	z_{PV}	x_{PV}	y_{PV}	z_{PV}
μ[mm]	0.677	0.211	-2.7	0.698	0.208	3.83
σ[mm]	0.030	0.025	40.6	0.024	0.023	61.08

	Pbp					
	MagDown			*MagUp*		
	x_{PV}	y_{PV}	z_{PV}	x_{PV}	y_{PV}	z_{PV}
μ[mm]	0.694	0.192	-1.17	0.687	0.201	2.66
σ[mm]	0.025	0.024	38.5	0.025	0.024	60.40

Table 8.1: Result of the means (μ) and standard deviations (σ) of the Gaussian fits used to define the luminous region for the forward (top) and backward (bottom) configuration for both magnet polarities. Reproduced from [103].

pipe and x and y are the transverse coordinates. Each distribution features a Gaussian core centred around the beam crossing location, accompanied by a broader background contribution. The mean (μ_x, μ_y, μ_z) and standard deviation $(\sigma_x, \sigma_y, \sigma_z)$ of these distributions are determined with Gaussian fits for each coordinate projection. Consequently, the luminous region is defined within a range of three standard deviations from these mean values, as follows:

$$|x_{\mathrm{PV}} - \mu_x| < 3\sigma_x, \quad |y_{\mathrm{PV}} - \mu_y| < 3\sigma_y, \quad |z_{\mathrm{PV}} - \mu_z| < 3\sigma_z. \tag{8.1}$$

The results of the Gaussian fits, presented in table 8.1, are reproduced from [103].

For pp data the impact of background events is less significant. For this reason, no restrictions on the position of the primary vertex are considered. A detailed discussion can be found in [103].

8.1.2 Primary Vertex Conditions

In proton-lead collisions multiple PVs can be reconstructed per event due to the high multiplicity. A detailed study was performed in [103], demonstrating that events with more than one reconstructed PV suffer from more background, defined in section 8.3.2. Therefore, in this analysis, pPb collisions are required to have one single PV per event.

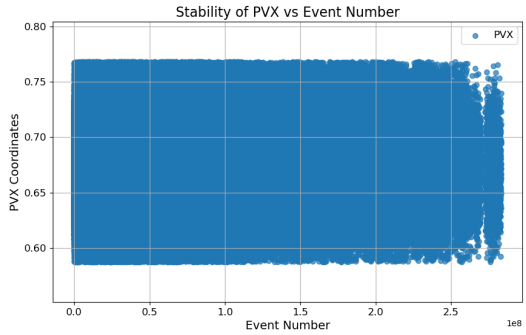
The instantaneous luminosity in the proton-proton data-taking was much higher than for the proton-lead runs. Therefore, the average number of reconstructed PVs per-bunch crossing was as high as 0.6 for the pp events and below 0.045 in the pPb events. This means that, on average, inelastic collisions are much more probable for the pp sample. However, despite that there is a non-negligible chance of having more than one pp interaction per event, because of the lower multiplicity in pp collisions, imposing a requirement on the number of PVs could introduce biases in the measurements. Therefore, no requirement

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

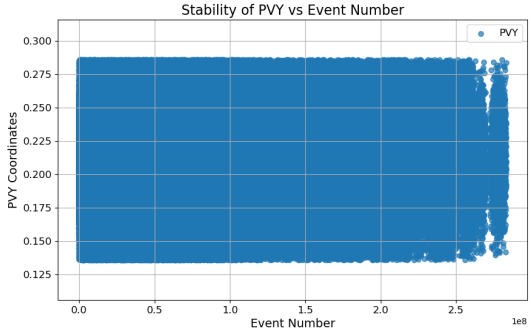
on the PV was made for the pp dataset. As stated above, a very detailed study of the data samples can be consulted in [103]. The selections described in this section for both the pp and pPb events can be found in table 8.4.

The stability of the PV distribution is evaluated by studying the PV projections x_{PV}, y_{PV}, and z_{PV} as a function of the number of events. The distributions of x_{PV}, y_{PV}, and z_{PV} are shown in fig. 8.1, fig. 8.2, and fig. 8.3 for the pPb, Pbp, and pp datasets, respectively. In the case of pp collisions, where multiple PVs may be reconstructed, the mean position of the PVs is taken for events with more than one PV.

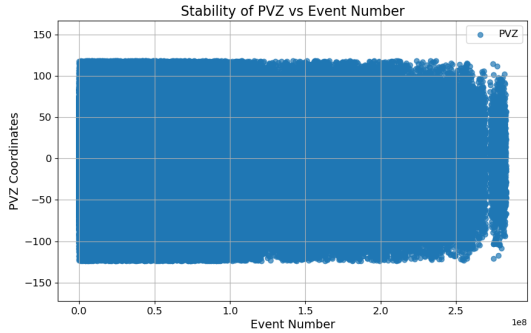
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) x_{PV} projection for the pPb dataset



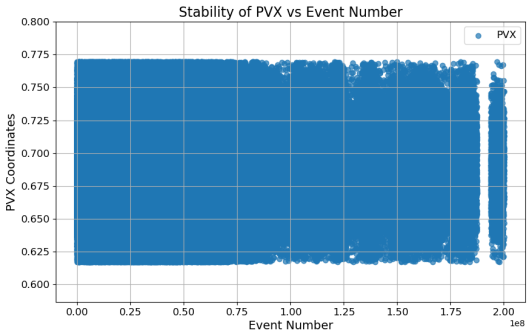
(b) y_{PV} projection for the pPb dataset



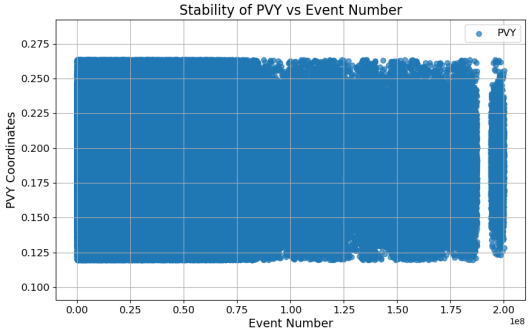
(c) z_{PV} projection for the pPb dataset

Figure 8.1: Primary vertex position projected along the x (top), y (middle), and z (bottom) axes as a function of the event number for the pPb dataset. The quantities x_{PV}, y_{PV}, and z_{PV} correspond to the PV position projections onto the respective axes. These distributions are used to monitor the stability of the vertex reconstruction and the detector alignment over time.

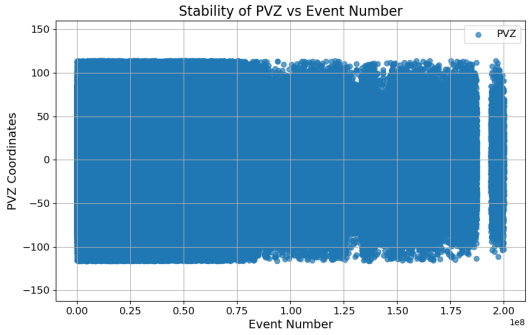
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) x_{PV} projection for the Pbp dataset

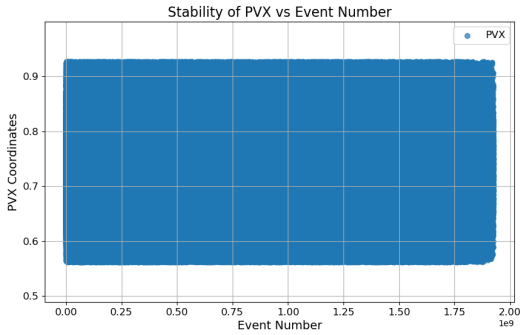


(b) y_{PV} projection for the Pbp dataset

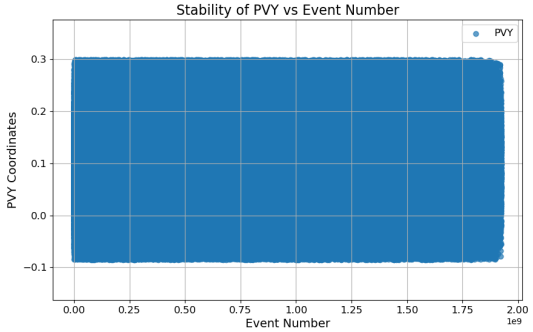


(c) z_{PV} projection for the Pbp dataset

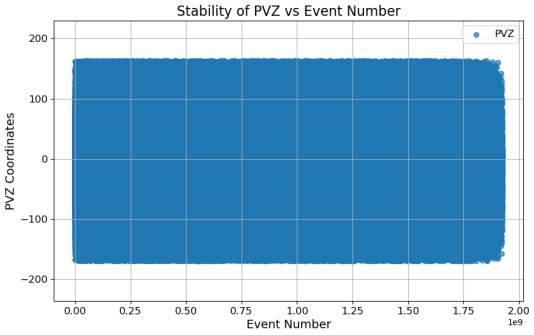
Figure 8.2: Primary vertex position projected along the x (top), y (middle), and z (bottom) axes as a function of the event number for the Pbp dataset. The quantities x_{PV}, y_{PV}, and z_{PV} correspond to the PV position projections onto the respective axes. These distributions are used to monitor the stability of the vertex reconstruction and the detector alignment over time.



(a) x_{PV} projection for the pp dataset



(b) y_{PV} projection for the pp dataset



(c) z_{PV} projection for the pp dataset

Figure 8.3: Primary vertex position projected along the x (top), y (middle), and z (bottom) axes as a function of the event number for the pp dataset. The quantities x_{PV}, y_{PV}, and z_{PV} correspond to the PV position projections onto the respective axes. These distributions are used to monitor the stability of the vertex reconstruction and the detector alignment over time.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8.2 Candidate Selection

This section discusses the selections focused on the track and its quality. It also describes the background sources and how to reduce and treat their presence in the analysed sample. In the rest of the document, the long-tracks passing the selections summarised in table 8.4, are referred to as **pion candidates**, **kaon candidates** and **proton candidates** depending on the PID cut applied. The term **candidates** (or **signal candidates**) refers generically to these categories, with the species explicitly indicated as $h \equiv \pi, K, p$ where needed.

8.2.1 Kinematic Range

In this analysis, long tracks are used as proxies for charged particle tracks. These tracks provide the best momentum and spatial resolution compared to other track types, as detailed in section 5.2. The tracks are taken from the `Phys/StdAllNoPIDsPions/Particles` container. As the name `NoPIDs` suggests, they do not have any PID requirements applied. Within this container tracks are assigned the pion mass only to calculate the energy. However, the energy is not used in this analysis. Charged particle tracks are required to fulfil kinematic requirements to reach the main tracking stations: $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}/c$, $p > 10\,\mathrm{GeV}/c$ and $2 < \eta < 4.8$. Tracks with $p_{\mathrm{T}} > 4.619\,\mathrm{GeV}/c$ [1] are excluded, since only the low and intermediate p_{T} ranges are of interest for this work. The kinematic coverage is changed from that used in the prompt charged particle production in pp and pPb analysis at 5 TeV [103] to align with the kinematic coverage of the calibration sample used to compute the particle identification efficiencies, described in section 9.3.

8.3 Signal Candidates

Prompt charged particles are those that meet the criteria in section 6.1. In particular, they are produced in the PV. Often, the impact parameter (IP), defined as the perpendicular distance from a particle trajectory to the collision point, is used to select particles produced near the collision point. However, in the proton-proton collisions analysed here, not every event has a single reconstructed PV, as discussed in section 8.1. Some events have no vertices[2] and events with multiple PVs may present ambiguities in matching reconstructed tracks with the appropriate PV. To address these two challenges, the pseudo impact parameter is used instead. For the i-th track the `pseudoIP` is defined as

$$\texttt{pseudoIP}^i = \sqrt{(x^i - \bar{x}_{PV})^2 + (y^i - \bar{y}_{PV})^2}, \tag{8.2}$$

[1] this analysis follows the binning of prompt charged particles [103] and is limited to p_{T} values up to 4.619 GeV/c due to the coverage of the calibration samples.

[2] A minimum of three tracks is required to reconstruct a PV.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Particle	True ID
Pion	211
Kaon	321
Proton	2212

Table 8.2: True ID values for pions, kaons, and protons in the MC simulation

where (x^i, y^i) are the transverse coordinates of the closest point of the track to the barycenter of PVs ($\bar{x}_{PV}$,$\bar{y}_{PV}$,$\bar{z}_{PV}$), which is calculated with fits of their projected densities to one-dimensional Gaussian distributions, see the results in fig. 8.4. A detailed explanation of this procedured can be found in [103].

In simulation samples a tool was developed to determine whether a track corresponds to a prompt-charged particle or not, based on the algorithm explained in [11]. The information is saved in a boolean variable called `isPrompt`. Pions, kaons and protons are tagged using their Monte Carlo True ID, shown in table 8.2.

The LHCb tracking system requires only a small number of hits to reconstruct a track. However, a fraction of these reconstructed tracks are derived from a combination of hits that do not correspond to the trajectory of a charged particle. In LHCb, such tracks are referred to as ghost, or fake, tracks. The algorithm used to determine the probability that a particle track is a ghost is described in [111], and this probability is denoted as `GhostP`.

A simultaneous selection on `GhostP` and `pseudoIP` was studied in detail in [103] by maximising the result of the product of significance and purity in simulation events,

$$\mathcal{SP}(g,p) = \frac{\mathcal{S}(g,p)}{\sqrt{\mathcal{S}(g,p)+\mathcal{B}(g,p)}} \cdot \frac{\mathcal{S}(g,p)}{\mathcal{S}(g,p)+\mathcal{B}(g,p)}, \tag{8.3}$$

where $\mathcal{S}(g,p)$ is the amount of prompt charged particles and $\mathcal{B}(g,p)$ the amount of background of fake tracks, clone tracks and secondary particles, with `GhostP` $< g$ and `pseudoIP` $< p$. This analysis reproduces the same cut values of that reference, summarised in table 8.4.

8.3.1 PID Selection

At LHCb the information from calorimeters, muon and RICH subsystems are combined in two types of PID variables:

ProbNN: this variable is derived from a neural network that classifies particles based on the data obtained from detector subsystems. The ProbNN was tuned using pp simulated events. Since the detector occupancy in proton-lead is higher than for proton-proton collisions, this variable will not be used in this analysis.

DLL$_{\mathbf{X}\pi}$: Using pion as reference, since it is the most abundant species, this variable provides the logarithm of the likelihood difference between the particle X (X $\equiv K, p, e, \mu$)

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

and the π hypotheses,

$$\mathrm{DLL}_{\mathrm{X}\pi} = \Delta\log\mathcal{L}(\mathrm{X}-\pi) \equiv \frac{\log\mathcal{L}(\mathrm{X})}{\log\mathcal{L}(\pi)}. \tag{8.4}$$

A likelihood difference between two arbitrary species (X and Y) can be obtained as:

$$\mathrm{DLL}_{\mathrm{XY}} = \mathrm{DLL}_{\mathrm{X}\pi} - \mathrm{DLL}_{\mathrm{Y}\pi} \equiv \frac{\log\mathcal{L}(\mathrm{X})}{\log\mathcal{L}(\pi)}\frac{\log\mathcal{L}(\pi)}{\log\mathcal{L}(\mathrm{Y})} = \frac{\log\mathcal{L}(\mathrm{X})}{\log\mathcal{L}(\mathrm{Y})}. \tag{8.5}$$

This variable does not rely directly on simulation for its definition, as it is constructed from the likelihood ratios based on the detector response to different particles. However, the calibration and validation of the likelihood functions may involve simulation. In this analysis, this variable will be used for PID to select enriched pion, kaon and proton samples. The pion sample is selected by applying a cut on $\mathrm{DLL}_{\mathrm{K}\pi}$ and $\mathrm{DLL}_{\mathrm{p}\pi}$ to remove proton and kaon contributions. Similarly, kaons are selected by cutting on $\mathrm{DLL}_{\mathrm{K}\pi}$ and $\mathrm{DLL}_{\mathrm{Kp}}$ and protons on $\mathrm{DLL}_{\mathrm{p}\pi}$ and $\mathrm{DLL}_{\mathrm{pK}}$.

To choose the optimal value of $\mathrm{DLL}_{\mathrm{K}\pi}$ and $\mathrm{DLL}_{\mathrm{p}\pi}$ for pion selection, the pion purity is defined as,

$$\mathrm{Purity} = \frac{N^{\pi}_{selected}}{N^{total}_{selected}}, \tag{8.6}$$

where $N^{\pi}_{selected}$ is the number of truth-matched candidates to a pion and $N^{total}_{selected}$ is the total number of candidates, both after the $\mathrm{DLL}_{\mathrm{K}\pi}$ and $\mathrm{DLL}_{\mathrm{p}\pi}$ cuts. The purity is computed across a range of 2-dimensional $\mathrm{DLL}_{\mathrm{K}\pi}$ and $\mathrm{DLL}_{\mathrm{p}\pi}$ cuts. The final cuts are then decided within regions where the purity exceeds 70%. The procedure is repeated for kaons and protons. The results of the purity are shown in fig. 8.6, fig. B.1 and fig. B.2 for pPb, Pbp and pp, respectively, The chosen pion, kaon and proton PID selection cuts are listed in table 8.4. The sample is divided into candidates for protons, kaons, and pions after the PID cut.

The tracks are also required to be registered by the `RICH` subsystem. This is accomplished by applying the criterion `hasRICH==1`. Additionally, to exclude muon-like tracks, tracks are required to pass the `isMuon==0` selection.

For the pPb, Pbp, and pp collision systems, the PID selection criteria for pion, kaon, and proton candidates were optimized to yield mutually exclusive samples. The specific selection requirements, provided in table 8.4, were validated through a Venn diagram analysis, which confirmed the absence of overlap between the samples, ensuring that they are disjoint.

8.3.2 Background Sources

Candidates that do not represent prompt-charged particles are considered background for the analysis. Understanding and quantifying these background sources is essential for minimising their influence. The main sources of background are detailed below.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Track Type	Selection Criteria
decay	abs(TRUEID)!=13 & ORIGINVERTEX_TYPE==2 & TRACK_Type==3
$X \to \mu$	ORIGINVERTEX_TYPE==2 & abs(TRUEID)==13 & TRACK_Type==3
$\gamma \to e$	abs(TRUEID)==11 & abs(MOTHER_ID)==22 &ORIGINVERTEX_TYPE==102 & TRACK_Type==3
had_int	ORIGINVERTEX_TYPE==100 & TRACK_Type==3
all sec	abs(TRUEID)!=0

Table 8.3: Selection criteria for secondary track sources in simulation: "all sec" encompasses every secondary track, "decay" specifies decay products of prompt and non-prompt particles, "$X \to \mu$" identifies particles decaying into muons, "$\gamma \to e$" denotes photon conversions and "had_int" refer to secondaries from hadronic origins.

Fake Tracks: or ghost tracks, are reconstruction artefacts that do not correspond to a trajectory of a charged particle. This background is reduced with the `GhostP` selection introduced in section 8.3. The remaining background tracks are then estimated with simulation, selecting tracks without a corresponding particle (`True_ID==0`). This is discussed in detail in section 9.2.1.

Secondary particles: originating from decays of other particles with lifetime $\tau > 30\ p$s, or from the interaction with the detector material. The following are the main sources of secondary particles:

- Particle decay product (decay):
  - Decays of K^0_S mesons, principally into pions.
  - Decays of Λ baryons, principally into pions and protons.
  - Decays of other hyperons, mainly Σ^+, Σ^-, Ξ^0, Ξ^-, Ω^-.
  - Pion and kaon decays into leptons.
- Particle decays into muons ($X \to \mu$).
- Converted photons ($\gamma \to e$).
- Interactions of hadrons with the detector material producing showers (had_int).

The secondary tracks are separated based on their origin into four main contributions: decay products of prompt and secondary particles, muons from particle decays, converted photons and hadronic interactions. The fraction of these backgrounds in the candidate sample is estimated with simulation by selecting tracks matched to a particle with `IsPrompt==False` and the selection criteria summarised in table 8.3. In data, secondary tracks originating from decays away from the PV have a larger `pseudoIP`, defined in section 8.3. The `pseudoIP` distribution for each contribution to the secondary tracks is illustrated in fig. 8.5. The secondary tracks background is further discussed in section 9.2.2, where use of the fraction of its different contributions is required.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Particle Mis-identification: The PID selection, detailed in section 8.3.1, cannot completely separate between pions, kaons, and protons. Consequently, a fraction of kaons and protons can be mis-identified as pions and pass the pion candidate selections. Similarly, the kaon candidates may include pions and protons, while proton candidates are contaminated with pions and kaons. The mis-identification (mis-ID) rates affect the purity of the selected particle sample. The PID efficiency and mis-ID rates are estimated with PID calibration samples. The approach used to separate between the species is detailed in section 9.3.

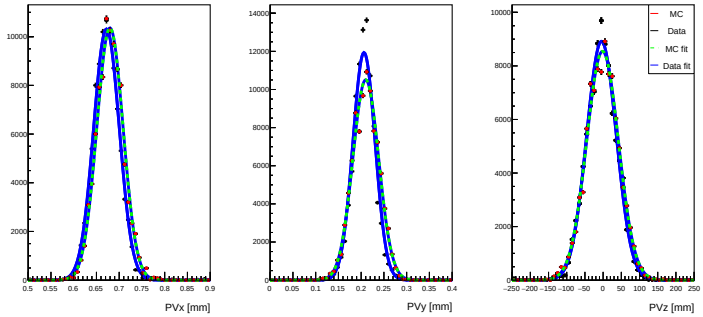
Clone Tracks: Another source of background arises when two reconstructed tracks originate from a single charged particle. These are called clone tracks. A large fraction of these tracks is removed from the `Phys/StdAllNoPidPions/Particles` container by the Clone Killer algorithm [112]. The contribution of this background was studied and its effect was found to be small enough to be neglected [113].

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

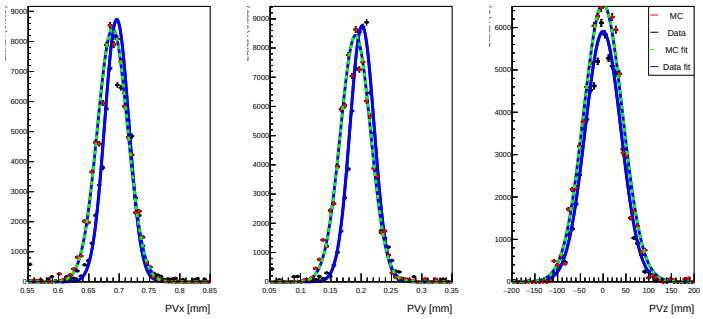
		pp	pPb	Pbp
Event selection	Trigger	`Hlt1NoBiasLeadingCrossing`	`Hlt1MBMicroBiasVelo`	`Hlt1MBMicroBiasVelo`
	Bunch crossing type	bunch-bunch	bunch-bunch	bunch-bunch
	number of PV	-	1	1
	PV position	-	within luminous region	within luminous region
Candidate selection	Track container	`/Event/Phys/StdAllNoPIDsPions/Particles`	`/Event/Phys/StdAllNoPIDsPions/Particles`	`/Event/Phys/StdAllNoPIDsPions/Particles`
	Track type	Long	Long	Long
	η	$2 < \eta < 4.8$	$2 < \eta < 4.8$	$2 < \eta < 4.8$
	p	$p > 10\,\text{GeV}/c$	$p > 10\,\text{GeV}/c$	$p > 10\,\text{GeV}/c$
	p_{T}	$0.5 < p_{\text{T}} < 4.619\,\text{GeV}/c$	$0.5 < p_{\text{T}} < 4.619\,\text{GeV}/c$	$0.5 < p_{\text{T}} < 4.619\,\text{GeV}/c$
	`pseudoIP`	0.368 mm	0.348 mm	0.348 mm
	`GhostP`	0.078	0.103	0.109
PID selection	pions	$\text{DLL}_{K\pi} < 0$ & $\text{DLL}_{p\pi} < 0$ & hasRich >0.5& `isMuon`==0	$\text{DLL}_{K\pi} < 0$ & $\text{DLL}_{p\pi} < 10$ & hasRich >0.5 & `isMuon`==0	$\text{DLL}_{K\pi} < 0$ & $\text{DLL}_{p\pi} < 10$ & hasRich >0.5 & `isMuon`==0
	Kaons	$\text{DLL}_{K\pi} > 5$ & ($\text{DLL}_{K\pi}$-$\text{DLL}_{p\pi}$) >5 & hasRich >0.5 & `isMuon`==0	$\text{DLL}_{K\pi} > 5$ & ($\text{DLL}_{K\pi}$-$\text{DLL}_{p\pi}$) >0 & hasRich >0.5& `isMuon`==0	$\text{DLL}_{K\pi} > 0$ & ($\text{DLL}_{K\pi}$-$\text{DLL}_{p\pi}$) >0 & hasRich >0.5& `isMuon`==0
	protons	$\text{DLL}_{p\pi} > 5$ & ($\text{DLL}_{p\pi}$-$\text{DLL}_{K\pi}$) >10 & hasRich >0.5& `isMuon`==0	$\text{DLL}_{p\pi} > 10$ & ($\text{DLL}_{p\pi}$-$\text{DLL}_{K\pi}$) > 10& hasRich >0.5& `isMuon`==0	$\text{DLL}_{p\pi} > 10$ & ($\text{DLL}_{p\pi}$-$\text{DLL}_{K\pi}$) > 10& hasRich >0.5& `isMuon`==0

Table 8.4: Summary of all the selections considered for the analysis.

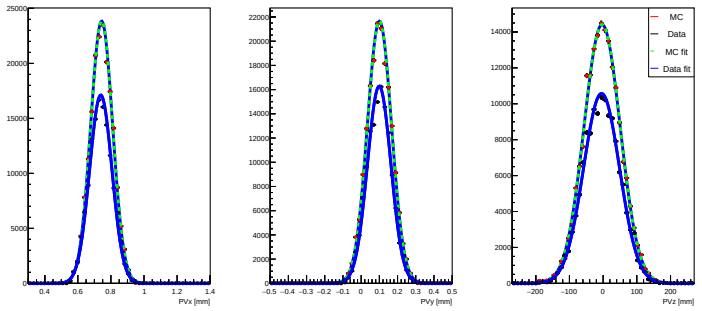
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) pPb

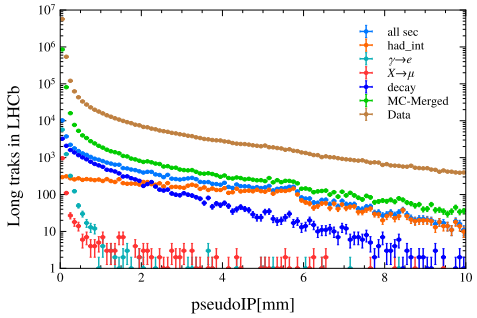


(b) Pbp

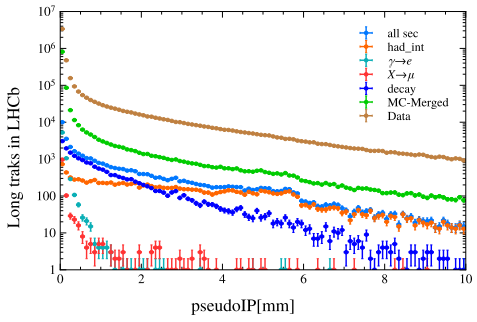


(c) pp

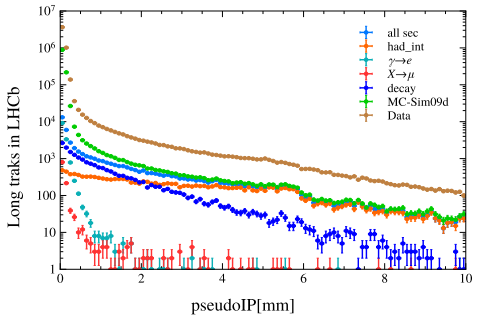
Figure 8.4: Projections of the Primary Vertex distribution and the fit to a three-dimensional Gaussian function that is used to calculate the average transverse PV position entering in the `pseudoIP` calculation of eq. (8.2). These Figures correspond to *MagDown* samples in three distinct beam configurations: Pbp (upper), pPb (middle), and pp (lower). The black and red dots represent data and simulation, respectively. The Gaussian function from the fit to the simulation dataset is presented as a green dashed line. The Gaussian function from the fit to data is represented with a blue line.



(a) `pseudoIP` pPb



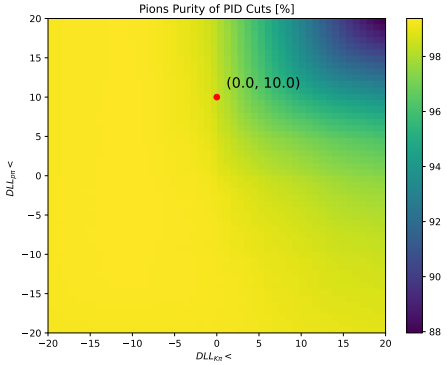
(b) `pseudoIP` Pbp



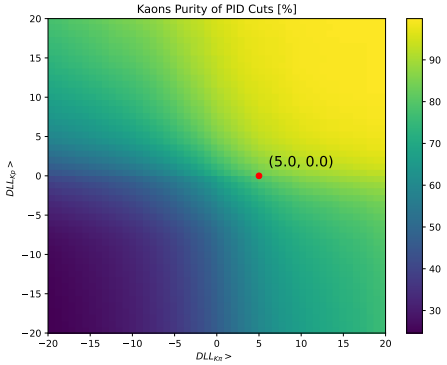
(c) `pseudoIP` pp

Figure 8.5: Distribution of the `pseudoIP` variable in pPb, Pbp and pp collisions. Data are shown in brown and compared to the simulation (green). In the simulation, contributions from different sources of secondary backgrounds are distinguished: hadronic interactions (orange), particle decays (dark blue), photon conversions $\gamma \to e$ (cyan), and decays producing muons $X \to \mu$ (red). The total of all secondary tracks is labelled "all sec" (light blue). The definitions of the track categories are provided in table 8.3. These distributions are used to validate the modelling of secondary backgrounds in the simulation relative to data.

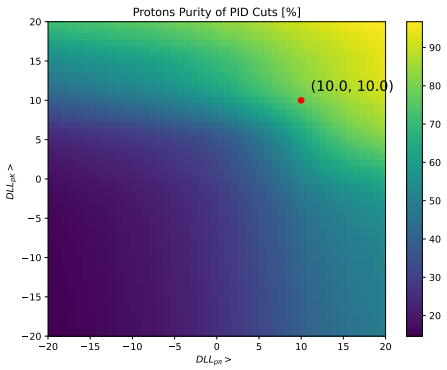
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Purity of pions



(b) Purity of kaons



(c) Purity of protons

Figure 8.6: Purity of pion (top), kaon (middle), and proton (bottom) candidates after applying PID cuts in pPb simulated events. The purity is defined as in eq. (8.6), using truth-matched tracks after cuts on $\mathrm{DLL}_{\mathrm{K}\pi}$ and $\mathrm{DLL}_{\mathrm{pK}}$. The scans across the two-dimensional DLL space allow the determination of optimal PID working points, with final cuts chosen in regions where the purity exceeds 70%.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8.4 Detector Occupancy

The spectrometer occupancy is quantified with variables such as the number of clusters, tracks, or hits in each subdetector. These quantities are correlated, within an event, to the charged particles produced during the collision. For instance, the multiplicity in the tracking stations (VELO, TT, IT, or OT) or the calorimeters (SPD). The considered variables in this analysis are the cluster multiplicity in the VELO, TT, IT, and OT detectors, the number of hits in the SPD detector and the number of reconstructed tracks (`nTracks`).

The occupancy in simulation does not properly reproduce that of data, particularly in proton-lead events. The disagreement between data and simulation is corrected for by assigning weights. This correction is accomplished using the `GBReweighter` tool [114], which is based on the "gradient boosting" machine learning technique [115]. Gradient boosting is one of the most powerful techniques for building predictive models.

The gradient boosting model is trained to weight the data to simultaneously tune distributions of the occupancy variables described above. Multiple combinations of these variables were tested, some of them only consider upstream VELO clusters or upstream tracks, others downstream tracks exclusively, and some sets consider hits across upstream and downstream stations simultaneously. The combinations are listed in table 8.5, each one is assigned a unique identifier for easy reference. All the events passing the minimum bias trigger were considered for the reweighting. The dataset was divided into training and testing subsets. The training set is used to build the model, while the testing set is used to evaluate its performance. The `GBReweighter` hyperparameters used to train proton-lead and proton-proton samples are shown in table 8.6. Once computed, the weights are applied to the entire dataset for each beam configuration.

Out of the initially proposed 14 sets, three sets, considering occupancy variables across the spectrometer, upstream and downstream the magnet, were considered:

- set 1: `nVeloClusters`, `nTTClusters`, `nITClusters`, `nOTClusters`.
- set 2: `nVeloClusters`, `nSPDHits`.
- set 7: `nPVs`, `nTracks`.

The remaining sets were either equivalent to these ones or did not correctly reproduce all of the occupancy variables.

The results of the reweighting for the three considered weight sets are shown in fig. 8.7, fig. 8.8 and fig. 8.9 for the pPb, Pbp and pp datasets, respectively. In every set, the occupancy distributions are better reproduced for the considered variables as compared to those that are not in the set. Set 1 was chosen as baseline for the analysis and deviations when employing set 2 and 7 are considered to estimate systematic uncertainties evaluating the model sensitivity to different aspects of track reconstruction and event characteristics, see section 10.3.2.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

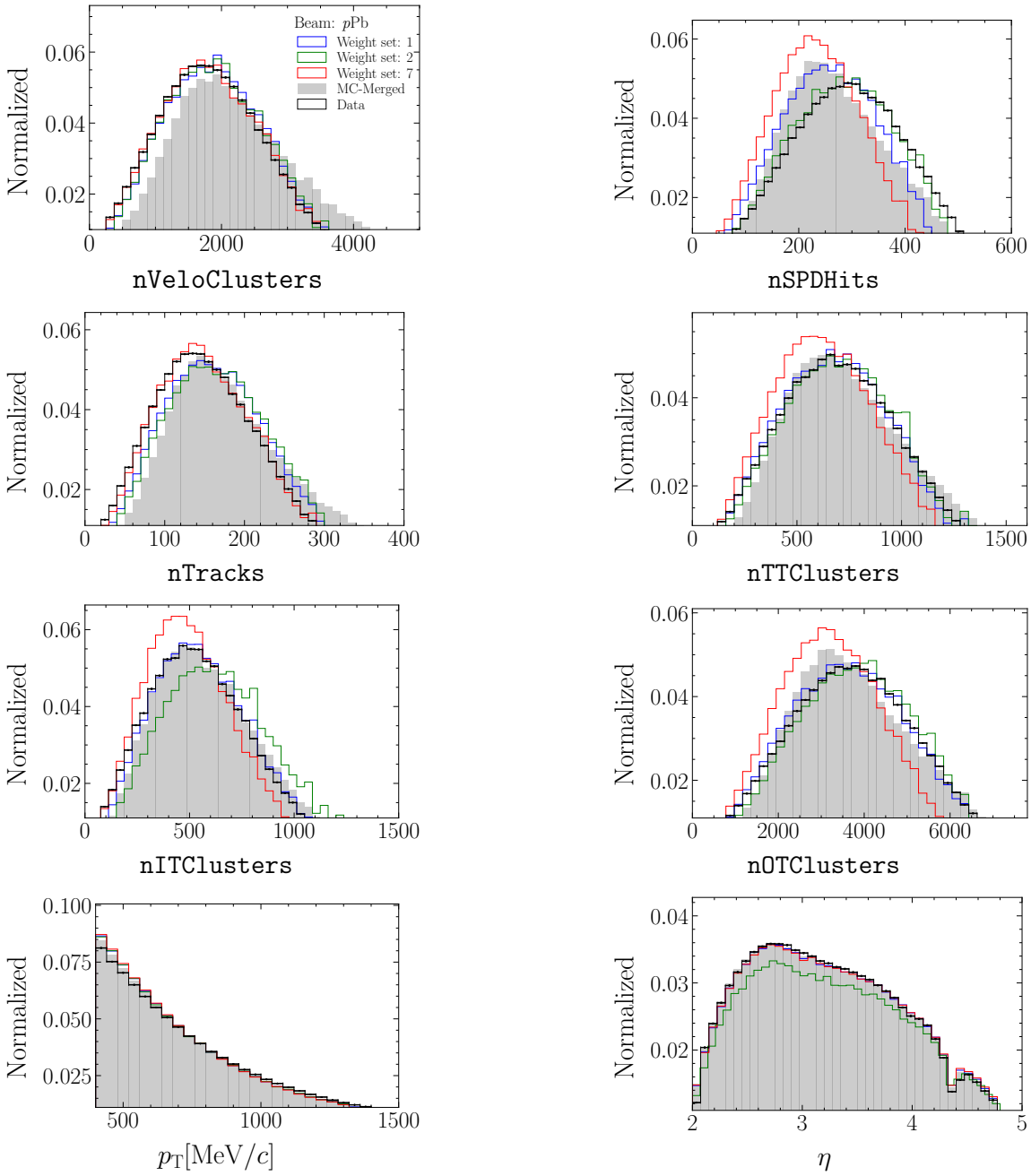


Figure 8.7: Comparison of occupancy-related and kinematic variables in data and in the `MC-Merged` pPb simulation sample. The simulation is shown after applying different sets of event weights: set 1 (blue), set 2 (green), and set 7 (red). The unweighted simulation is shown in grey, while the data are shown in black. These weighting strategies are based on different combinations of detector occupancy variables.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

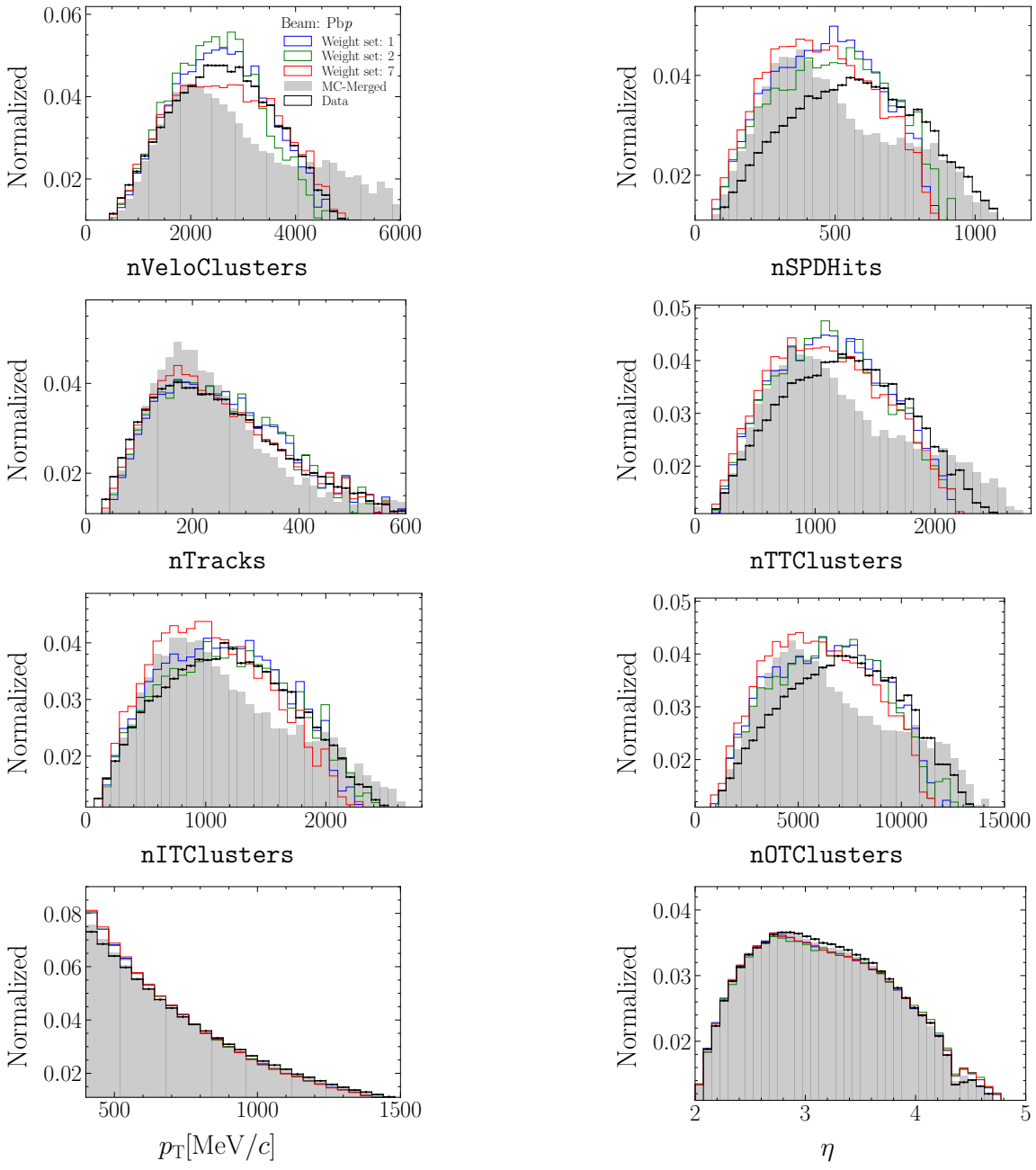


Figure 8.8: Comparison of occupancy-related and kinematic variables in data and in the `MC-Merged` Pbp simulation sample. The simulation is shown after applying different sets of event weights: set 1 (blue), set 2 (green), and set 7 (red). The unweighted simulation is shown in grey, while the data are shown in black. These weighting strategies are based on different combinations of detector occupancy variables.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

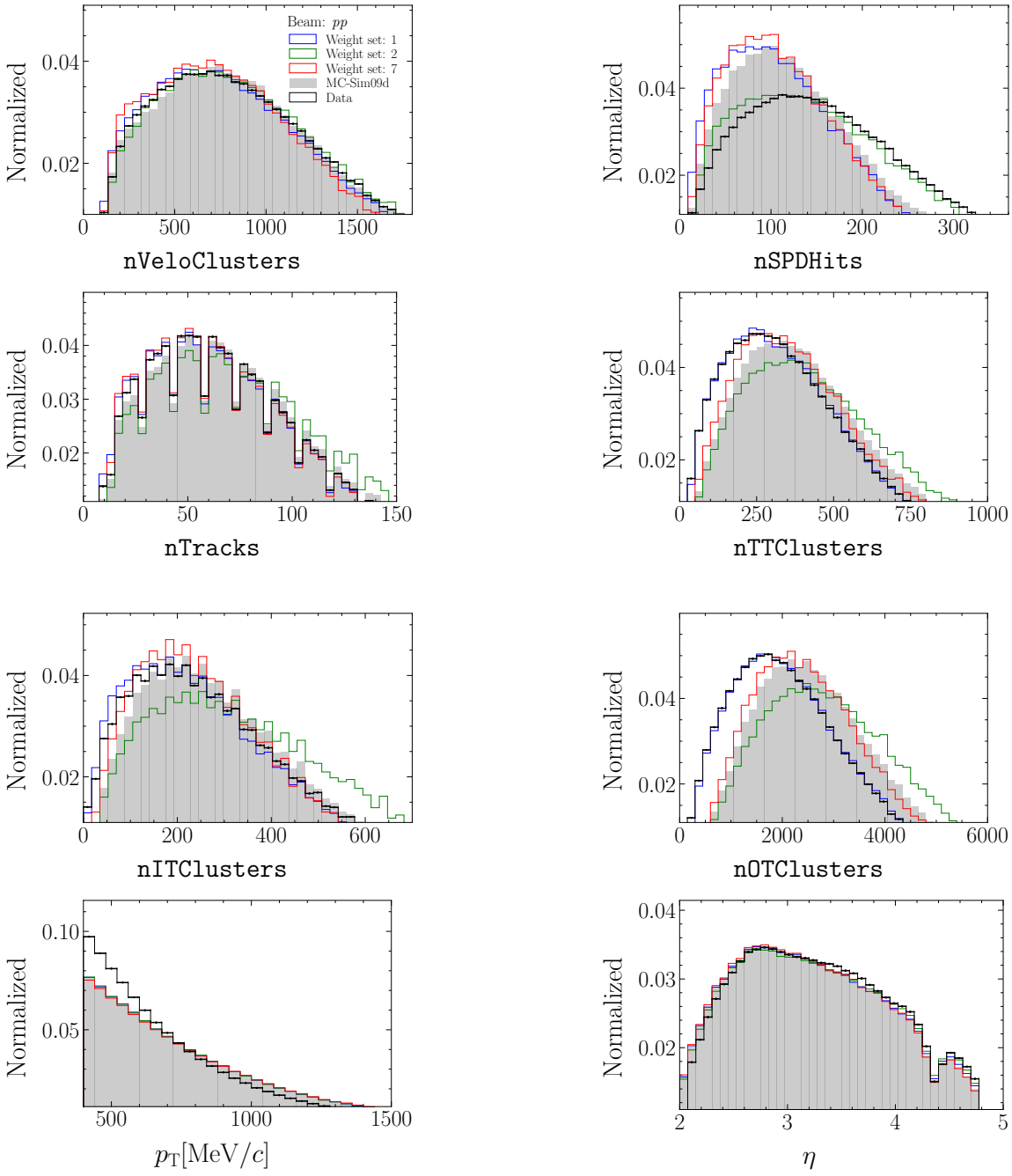


Figure 8.9: Comparison of occupancy-related and kinematic variables in data and in the `MC-Sim09d` pPb simulation sample. The simulation is shown after applying different sets of event weights: set 1 (blue), set 2 (green), and set 7 (red). The unweighted simulation is shown in grey, while the data are shown in black. These weighting strategies are based on different combinations of detector occupancy variables.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Set number	Detector Components
1	nVeloClusters, nTTClusters, nITClusters, nOTClusters
2	nVeloClusters, nSPDHits
3	nVeloClusters, nTTClusters, nTracks
4	nVeloClusters
5	nVeloClusters, nITClusters, nOTClusters, nTTClusters, nSPDHits
6	nVeloClusters, nITClusters, nOTClusters, nSPDHits, nTracks
7	nPVs, nTracks
8	nTracks
9	nVeloClusters, nPVs
10	nVeloClusters, nITClusters, nOTClusters, nTTClusters, nPVs
11	nVeloClusters, nITClusters, nOTClusters, nTracks
12	nTTClusters, nITClusters, nOTClusters
13	nPVs, nTTClusters, nITClusters, nOTClusters
14	nVeloClusters, nTracks

Table 8.5: List of Occupancy Variables Sets used for training the GBReweighter tool.

Parameter	Value	Description
n_estimators	50	Number of trees.
learning_rate	0.08	Learning rate.
max_depth	3	Maximum depth of each tree.
min_sample_leaf	1000	Minimum number of samples required to be at a leaf node.
gb_args	'subsample': 0.4	Dictionary of additional arguments. With the key of subsample set to a value of 0.4 each iteration will use 40% of the training samples randomly selected from the entire dataset.

Table 8.6: GBReweighter hyperparameters used in the analysis.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

9

Measurement of the Identified Prompt Charged Particle Spectra

9.1 Definitions

The goal of this analysis is to measure the production cross-sections of light charged hadrons (pions, kaons, and protons) in pPb, Pbp, and pp collisions. As introduced in section 6.4, this differential cross-section is defined as:

$$\frac{\mathrm{d}^2\sigma^h_{pp,p\mathrm{Pb}}}{\mathrm{d}p_\mathrm{T}\mathrm{d}\eta_\mathrm{cms}} = \frac{1}{\mathcal{L}}\frac{N^h(\eta_\mathrm{cms}, p_\mathrm{T})}{\Delta\eta_\mathrm{cms}\Delta p_\mathrm{T}}, \qquad \text{(eq. (6.4))}$$

where $\mathcal{L}$ is the integrated luminosity of the dataset, and $N^h(\eta_\mathrm{cms}, p_\mathrm{T})$ is the amount of charged light hadrons of type h in the kinematic bin $\Delta\eta_\mathrm{cms} \times \Delta p_\mathrm{T}$. The binning scheme is described in section 9.1.1.

The amount of charged light hadrons is determined in several steps:

1. Starting from the observed hadron yield, N^h_{cand}, after applying the selections listed in table 8.4.

2. Correcting this yield for purity to obtain the number of selected hadrons,

$$N^h_{Sel} = N^h_{cand} \cdot P^h, \tag{9.1}$$

where the purity is

$$P^h = 1 - \sum_\alpha R^h_\alpha f^{sim}_\alpha(h). \tag{9.2}$$

Here, f^{sim}_α is the fraction of background (ghosts or secondaries) in simulation, and R^h_α corrects for data–simulation discrepancies. Details on their determination are given in section 9.2.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Given the presence of non-zero misidentification rates, the three samples contaminate each other. To obtain the *True*[1] yields without inter-species contamination, N^h_{True}, the selected yields are corrected using the PID response matrix:

$$\begin{bmatrix} N^{\pi}_{True} \\ N^{K}_{True} \\ N^{p}_{True} \end{bmatrix} = \begin{bmatrix} \varepsilon^{\pi\to\pi} & \varepsilon^{K\to\pi} & \varepsilon^{p\to\pi} \\ \varepsilon^{\pi\to K} & \varepsilon^{K\to K} & \varepsilon^{p\to K} \\ \varepsilon^{\pi\to p} & \varepsilon^{K\to p} & \varepsilon^{p\to p} \end{bmatrix}^{-1} \begin{bmatrix} N^{\pi}_{Sel} \\ N^{K}_{Sel} \\ N^{p}_{Sel} \end{bmatrix}. \tag{9.3}$$

The diagonal terms correspond to PID efficiencies, while the off-diagonal terms represent misidentification probabilities. The determination of these efficiencies and mis-ID rates is described in section 9.3.

Finally, the *True* yields are corrected for detector and selection effects to obtain the fully corrected hadron yields:

$$N^h = N^h_{True} \cdot \frac{1}{\varepsilon^h_{reco}\,\varepsilon^h_{sel}\,(1/\varepsilon_{TM})}, \tag{9.4}$$

where ε^h_{reco}, ε^h_{sel}, and ε_{TM} are the reconstruction, selection, and truth-matching efficiencies, respectively. The determination of these efficiencies is described in detail in sections 9.4 to 9.6.

9.1.1 Selecting The Binning in The (η, p_{T}) plane

The binning in $(\eta_{\mathrm{cms}}, p_{\mathrm{T}})$ is a crucial step in the analysis, as it affects the statistical precision, detector resolution, and the physical interpretation of the results. The choice is guided by three main considerations:

1. **Detector acceptance:** The LHCb detector covers a finite range in laboratory pseudorapidity, approximately $2.0 < \eta < 4.8$. Bins must be fully contained within this acceptance.

2. **Detector resolution:** The resolution in η and p_{T} sets a lower limit on the bin size.

3. **Statistical precision:** Each bin must contain a sufficient number of candidates.

As discussed in section 6.2, using η_{cms} provides a common frame of reference and allows a direct comparison between the three beam configurations: pp, pPb, and Pbp. The rapidity in the cms frame is shifted by:

$$|\delta| \approx 0.465 \approx 0.5, \tag{9.5}$$

Then depending on the beam configuration, the relation between η and η_{cms} is approximately:

[1] The term *True* does not refer to Monte Carlo truth information. It is used here in a mathematical sense to represent the inferred particle numbers after correcting for misidentifications.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

$\eta_{\rm cms}$ (pp)	$\eta_{\rm cms}$ (pPb)	$\lvert\eta_{\rm cms}\rvert$(Pb$p$)
-	1.6–2.0	-
2.0–2.5	2.0–2.5	-
2.5–3.0	2.5–3.0	2.5–3.0
3.0–3.5	3.0–3.5	3.0–3.5
3.5–4.0	3.5–4.0	3.5–4.0
4.0–4.3	4.0–4.3	4.0–4.3
4.0–4.5	-	4.0–4.5
4.5–4.8	-	4.5–4.8
-	-	4.8–5.3

Table 9.1: $\eta_{\rm cms}$ bins used in this analysis for pp, pPb, and Pbp configurations.

- pPb **(proton forward):** $\eta_{\rm cms} = \eta - 0.5$.
- Pbp **(proton backward):** $\eta_{\rm cms} = \eta + 0.5$
- pp **collisions:** $\eta_{\rm cms} = \eta$ (no boost)

Using these relations, the detector coverage in $\eta_{\rm cms}$ becomes:

- pp**:** $2.0 < \eta_{\rm cms} < 4.8$
- pPb**:** $1.6 < \eta_{\rm cms} < 4.3$
- Pbp**:** $-5.3 < \eta_{\rm cms} < -2.5$ (or in absolute value $2.5 < |\eta_{\rm cms}| < 5.3$)

Binning in $\eta_{\rm cms}$: A uniform bin width of 0.5 in $\eta_{\rm cms}$ is chosen to match the natural resolution of the rapidity shift. The $\eta_{\rm cms}$ binning scheme for each beam configuration is summarized in table 9.1. Empty entries (-) indicate that this region lies outside the LHCb acceptance for that configuration. For pp and Pbp samples, the bin $[4.0, 4.3]$ is an additional bin introduced to ensure overlap with the highest $\eta_{\rm cms}$ region accessible in pPb collisions. This overlap allows us to measure the $R^h_{p\rm Pb}$ and the forward to backward ratios in the same kinematic region.

Binning in $p_{\rm T}$: Since the multiplicity decreases exponentially with the transverse momentum, an increasing $p_{\rm T}$-bin size is suggested. Considering all of the above, the proposed binning scheme is summarized in table 9.2.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

η	p_{T} bins [GeV/c]
$(2.0, 2.5)$	$[2.107, 2.737, 3.556, 4.619]$
$(2.5, 3.0)$	$[1.249, 1.622, 2.107, 2.737, 3.556, 4.619]$
$(3.0, 3.5)$	$[0.740, 0.961, 1.249, 1.622, 2.107, 2.737, 3.556, 4.619]$
$(3.5, 4.0)$	$[0.500, 0.740, 0.961, 1.249, 1.622, 2.107, 2.737, 3.556, 4.619]$
$(4.0, 4.5)$	$[0.500, 0.740, 0.961, 1.249, 1.622, 2.107, 2.737, 3.556, 4.619]$
$(4.5, 4.8)$	$[0.500, 0.740, 0.961, 1.249, 1.622, 2.107, 2.737, 3.556, 4.619]$

Table 9.2: Proposed binning scheme in the (p_{T}, η) plane.

9.2 Isolating the Prompt Contribution: Purity Calculation

In this section, the contributions to the background from fake tracks and secondary particles are discussed. Both simulation and data-driven techniques are used. A detailed understanding of these background contributions is necessary for effective background subtraction. This process enhances the purity of the measured signal, enabling a more accurate estimation of the prompt contribution.

9.2.1 Fake Tracks

The corrected fraction of fake tracks is calculated as:

$$f^h_{fake} = f^{sim}_{fake}(h) R^h_{fake}, \tag{9.6}$$

where $f^{sim}_{fake}(h)$ is the fraction of fake tracks in simulation and R^h_{fake} is a factor correcting for discrepancies between simulation and data.

The fraction of fake tracks in the simulation sample is obtained as:

$$f^{sim}_{fake}(h) = \frac{\sum_{i}^{N^h_{fake_cand}} \varepsilon_{TM}(\eta^i_{\mathrm{cms}}, p^i_{\mathrm{T}}) w_i}{\sum_{i}^{N^h_{cand}} w_i}, \tag{9.7}$$

where N^h_{cand} represents the total number of candidates that pass the selections outlined in table 8.4. The term $N^h_{fake_cand}$ refers to N^h_{cand} that do not match any MC particle (`TRUEID` equal to 0). The occupancy weight of the event, introduced in section 8.4, is denoted as w_i, and ε_{TM} is the efficiency of the truth matching algorithm. This efficiency accounts in eq. (9.7) for the truth-matching algorithm failures when the fake track is attributed to candidates not matched to their corresponding MC particles. The procedure to estimate this efficiency is discussed in section 9.6.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The cut on `GhostP` applied in the analysis removes a large fraction of fake tracks. The remaining background is studied across different occupancy weights to check for deviations for each light hadron. The obtained fractions are shown in fig. 9.1. Across all three collision types, proton-lead, lead-proton and proton-proton, a consistent trend is observed. The proton sample (green) exhibits the highest fake track rate, followed by the kaon sample (blue), with the pion sample (red) having the lowest rate.

Calculating the fraction of fake tracks with simulation is not sufficient for an accurate description of the data. Discrepancies arise from limitations in the detector response description and from event characteristics, which lead to different fake track fractions between simulated events and data. To estimate the discrepancies, proxy data and simulation samples dominated by fake tracks are considered using the selection criteria described in table 9.3. The purity of the proxy sample is estimated with the simulation sample as:

$$\Pi^{h,sim}_{fake} = \frac{N^{h,sim\text{-}TM}_{proxy_fake}}{N^{h,sim}_{proxy_fake}}, \tag{9.8}$$

where $N^{h,sim}_{proxy_fake}$ is the total number of candidates in the proxy sample and $N^{h,sim\text{-}TM}_{proxy,fake}$ corresponds to the number of truth matched (TM) fake tracks (`TRUEID` $== 0$) within $N^{h,sim}_{proxy_fake}$. The results of the purity of the proxy sample for pPb, Pbp and pp events are shown in fig. 9.2a, fig. 9.2b and fig. 9.2c, respectively. The fact that these are not 100% pure samples contributes to the systematic uncertainty of the fake fractions.

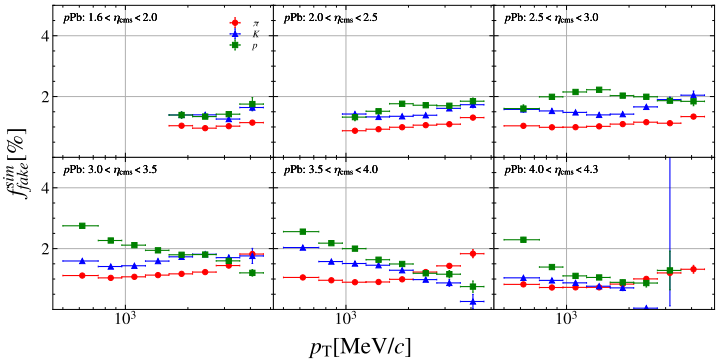
The fake track yield ratio of the proxy data sample, as compared to simulation, is referred to as the proxy factor and is calculated for each light hadron as:

$$R^{h}_{fake} = \frac{\frac{1}{N^{h,data}_{cand}} N^{h,data}_{proxy_fake}}{\frac{1}{\sum_{i=1}^{N^{h,sim}_{cand}} w_i} \sum_{i=1}^{N^{h,sim}_{proxy_fake}} w_i}. \tag{9.9}$$

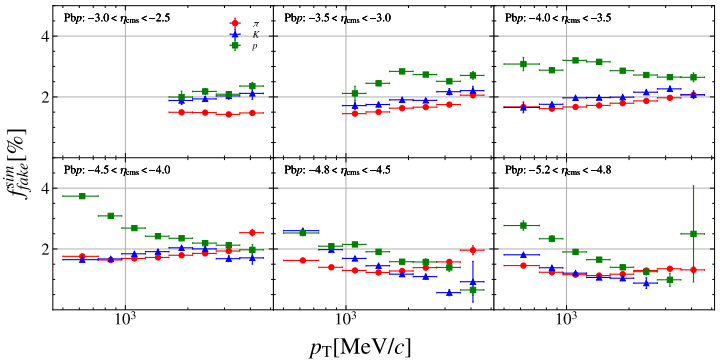
Here, $N^{h,data}_{proxy_fake}$ represents the number of tracks in the data that pass the selection criteria for the proxy sample of fake tracks, while $N^{h,sim}_{proxy_fake}$ represents the number of proxy fake track candidates in simulation, also used in eq. (9.12). The total number of candidates in data and simulation are represented by $N^{h,data}_{cand}$ and $N^{h,sim}_{cand}$, respectively. Finally, w_i is again the occupancy weight for each candidate in simulation (introduced in section 8.4). The results for R^{h}_{fake} are shown in fig. 9.3a for pPb, fig. 9.3b for Pbp and fig. 9.3c for pp events, respectively.

Finally, the estimated values of the fake track fractions for pPb, Pbp and pp events are shown in fig. 9.4a, fig. 9.4b and fig. 9.4c, respectively. The systematic uncertainty in the estimation of these fractions is discussed in detail in section 10.1.1.

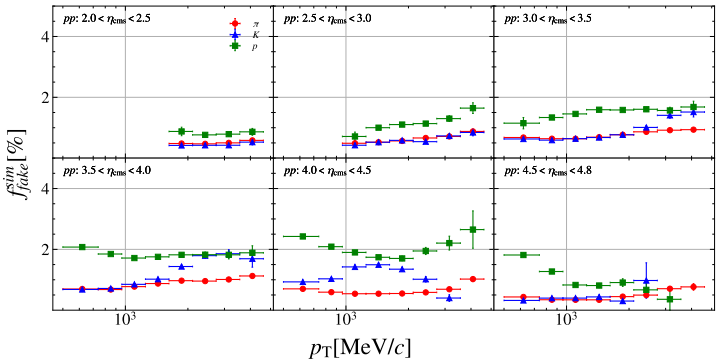
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Fake tracks in pPb.



(b) Fake tracks in Pbp.



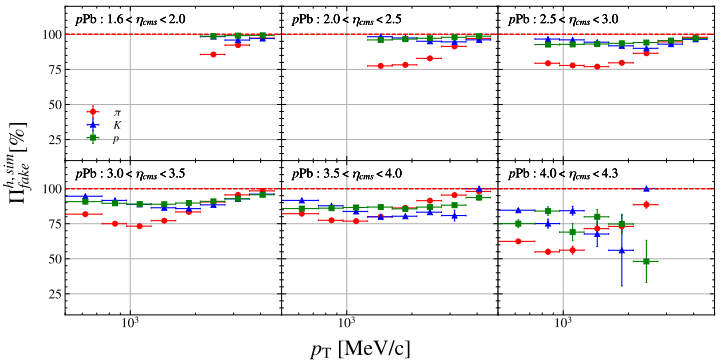
(c) Fake tracks in pp.

Figure 9.1: Fraction of fake tracks in simulation, $f^{sim}_{fake}(h)$, for pPb (top), Pbp (middle), and pp (bottom) collisions. The fractions are shown separately for pions (red), kaons (blue), and protons (green). Only statistical uncertainties are displayed. These distributions illustrate the expected rate of fake tracks in the simulation before applying the data-driven correction factor R^h_{fake} (see eq. (9.6)).

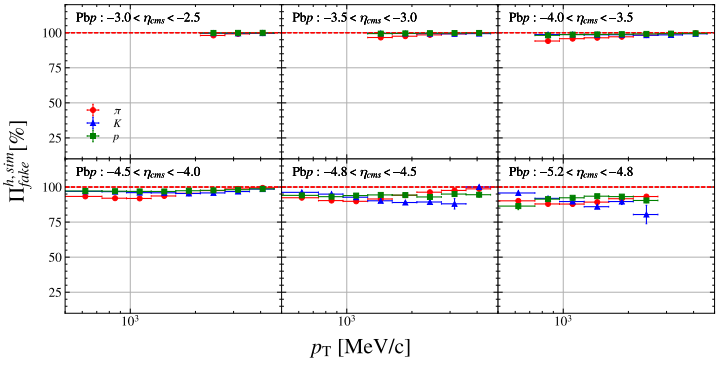
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

	pp	pPb	Pbp
Track type	Long	Long	Long
η	$2 < \eta < 4.8$	$2 < \eta < 4.8$	$2 < \eta < 4.8$
p_{T}	$0.5 < p_{\mathrm{T}} < 5\,\mathrm{GeV}/c$	$0.5 < p_{\mathrm{T}} < 5\,\mathrm{GeV}/c$	$0.5 < p_{\mathrm{T}} < 5\,\mathrm{GeV}/c$
p	$p > 10\,\mathrm{GeV}/c$	$p > 10\,\mathrm{GeV}/c$	$p > 10\,\mathrm{GeV}/c$
pseudoIP	$< 0.368\,\mathrm{mm}$	$< 0.348\,\mathrm{mm}$	$< 0.348\,\mathrm{mm}$
GhostP	$0.1 <$ GhostP < 0.4	$0.5 <$ GhostP < 0.9	$0.5 <$ GhostP < 0.9
	PID selection		
pions	$\mathrm{DLL}_{K\pi} <$ -2 & $\mathrm{DLL}_{p\pi} <$-2 & hasRich >0.5& isMuon==0	$\mathrm{DLL}_{K\pi} < 0$ & $\mathrm{DLL}_{p\pi} <$10 & hasRich >0.5 & isMuon==0	$\mathrm{DLL}_{K\pi} < 0$ & $\mathrm{DLL}_{p\pi} <$10 & hasRich >0.5 & isMuon==0
Kaons	$\mathrm{DLL}_{K\pi} > 5$ & $(\mathrm{DLL}_{K\pi}$-$\mathrm{DLL}_{p\pi}) > 5$ & hasRich >0.5 & isMuon==0	$\mathrm{DLL}_{K\pi} > 5$ & $(\mathrm{DLL}_{K\pi}$-$\mathrm{DLL}_{p\pi}) > 0$ & hasRich >0.5& isMuon==0	$\mathrm{DLL}_{K\pi} > 5$ & $(\mathrm{DLL}_{K\pi}$-$\mathrm{DLL}_{p\pi}) > 0$ & hasRich >0.5& isMuon==0
protons	$\mathrm{DLL}_{p\pi} > 5$ & $(\mathrm{DLL}_{p\pi}$-$\mathrm{DLL}_{K\pi}) > 10$ & hasRich >0.5& isMuon==0	$\mathrm{DLL}_{p\pi} > 5$ & $(\mathrm{DLL}_{p\pi}$-$\mathrm{DLL}_{K\pi}) > 15$& hasRich >0.5& isMuon==0	$\mathrm{DLL}_{p\pi} > 5$ & $(\mathrm{DLL}_{p\pi}$-$\mathrm{DLL}_{K\pi}) > 15$& hasRich >0.5& isMuon==0

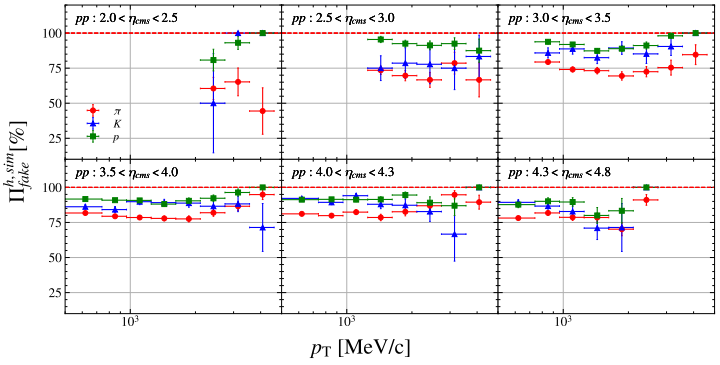
Table 9.3: List of requirements for the fake track proxy in each configuration.



(a) Purity in pPb.



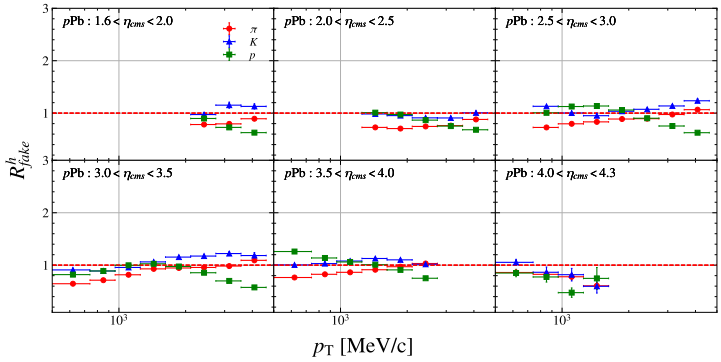
(b) Purity in Pbp.



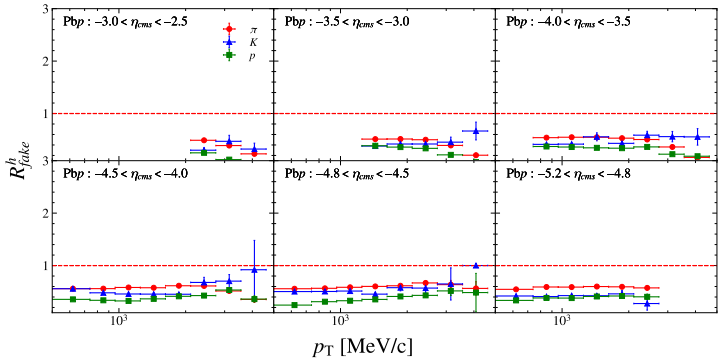
(c) Purity in pp.

Figure 9.2: Purity of the proxy sample used to estimate fake tracks in pPb (top), Pbp (middle), and pp (bottom) collisions. Fractions are shown for each hadron type: pions (red), kaons (blue), and protons (green). Only statistical uncertainties are displayed.

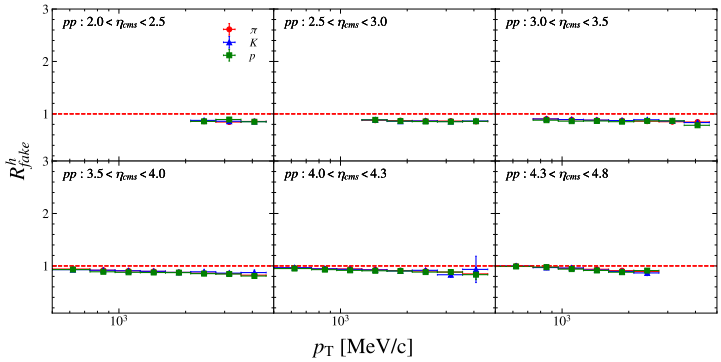
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Proxy factor in pPb.

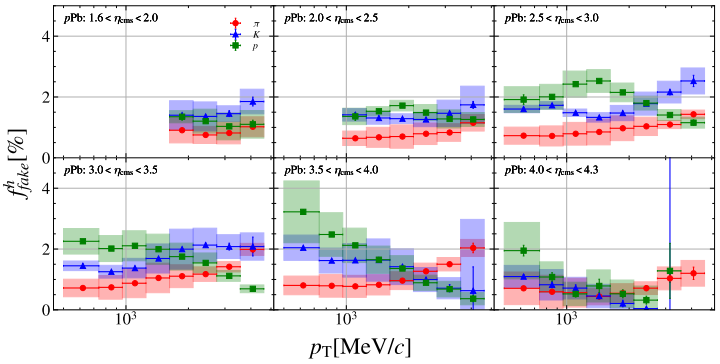


(b) Proxy factor in Pbp.

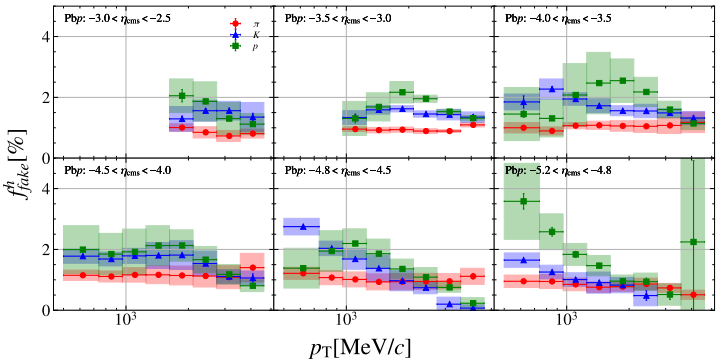


(c) Proxy tracks in pp.

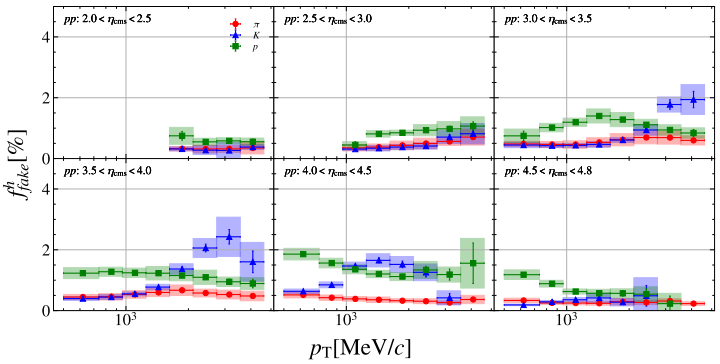
Figure 9.3: Data-to-simulation correction factors R^h for the proxy sample of fake tracks in pPb (top), Pbp (middle), and pp (bottom) collisions. Fractions are shown for each hadron type: pions (red), kaons (blue), and protons (green). Only statistical uncertainties are displayed. These factors quantify the discrepancy between simulation and data for fake-track rates.



(a) Fake tracks in pPb.



(b) Fake tracks in Pbp.



(c) Fake tracks in pp.

Figure 9.4: Fraction of fake tracks in pPb (top), Pbp (middle), and pp (bottom) collisions for each type of light hadron: pions (red), kaons (blue), and protons (green). Statistical uncertainties are shown as vertical lines, and systematic uncertainties are represented by boxes. Only the *MagDown* polarity is displayed. These distributions illustrate the expected fake-track contributions in the dataset.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

9.2.2 Secondary Tracks

The fraction of secondary tracks is obtained as:

$$f^h_{sec} = f^{sim}_{\text{had_decay}}(h) R^h_{\text{had_decay}} + f^{sim}_{\gamma\to e}(h) R^h_{\gamma\to e} + f^{sim}_{X\to\mu}(h), \tag{9.10}$$

where $f^{sim}_j(h)$ are the secondary track fractions obtained in simulation and R^h_j are factors correcting for discrepancies between data and simulation obtained with proxy samples, to be discussed below, similarly to what was done for fake tracks in section 9.2.1. The subscripts $j = \{\text{had_decay}, X\to\mu, \gamma\to e\}$ refer to the origin of the secondary tracks, see section 8.3.2, and they stand for:

- had_decay: this category accounts for tracks from hadronic interactions (had_int) and tracks different than muons originating from decays (decay). These two contributions were originally separated but are merged because the contribution of muons from decays is negligible due to the PID requirement ISMUON==0. However, in some of the figures the two separate contributions are displayed.
- $X\to\mu$: muons originating from decays.
- $\gamma\to e$: tracks resulting from photon conversions.

Each fraction is computed with simulation as

$$f^{sim}_j(h) = \frac{\sum_i^{N^h_j} w_i}{\sum_i^{N^h_{cand}} w_i}, \tag{9.11}$$

where N^h_{cand} is the amount of candidates after the selections of table 8.4, N^h_j is the subset of these candidates that fulfil the secondary tracks requirements based on its origin, summarised in table 9.4, and w_i is the occupancy weight per event introduced in section 8.4. The results of the fractions are shown for pPb, Pbp and pp in fig. 9.5, fig. 9.6 and fig. 9.7, where the subcategories had_int and decay are shown disaggregated from the had_decay category. The had_int category is indicated in red, μ in brown, γ in green, and decay in blue. The different secondary sources are stacked to compare the aggregated result with all secondary tracks in black.

Similar to the strategy for fake tracks discussed in section 9.2.1, proxy samples with the highest purity in the different category for secondary tracks are defined to address data-simulation discrepancies. The selection criteria defining the different proxy samples are shown in table 9.5. The purity of the j-th proxy sample is estimated in simulation as the fraction of truth-matched (TM) candidates in the sample,

$$\Pi^{h,sim}_j = \frac{N^{h,sim\text{-}TM}_{proxy,j}}{N^{h,sim}_{proxy,j}}. \tag{9.12}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Category	Selection Criteria
had_int	`isPrompt==0 && ORIGINVERTEX_TYPE==100 && TRACK_Type==3`
$\gamma \to e$	`isPrompt==0 && abs(TRUEID)==11 && abs(MOTHER_ID)==22 && ORIGINVERTEX_TYPE==102 && TRACK_Type==3`
$X \to \mu$	`isPrompt==0 && ORIGINVERTEX_TYPE==2 && abs(TRUEID)==13 && TRACK_Type==3`
decay	`isPrompt==0 && abs(TRUEID)!=13 && ORIGINVERTEX_TYPE==2 && TRACK_Type==3`
all	`isPrompt==0 && abs(TRUEID)!=0`

Table 9.4: Selection criteria for Secondary Tracks in simulation

The purity of the different proxy samples is used when estimating the systematic uncertainty originated in the correction of the different secondary particle fractions, as discussed in section 10.1.2.

The correction factor for secondary tracks, based on their origin, is defined as:

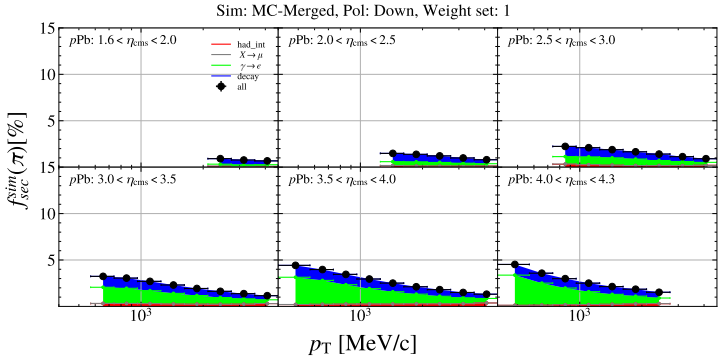
$$R_j^h = \frac{\frac{1}{N_{cand}^{h,data}} N_{proxy,j}^{h,data}}{\frac{1}{\sum_{i=1}^{N_{cand}^{h,sim}} w_i} \sum_{i=1}^{N_{proxy,j}^{h,sim}} w_i}. \tag{9.13}$$

This approach follows the strategy of [103].

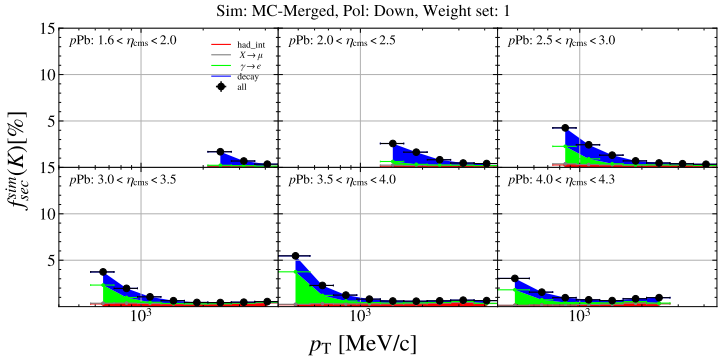
In eq. (9.10) no correction factor is applied to the fraction for secondary tracks arising from $X \to \mu$ decays, since constructing a dedicated proxy sample was not possible due to limited statistics. Consequently, the fraction in simulation is not scaled and a large uncertainty of $\sigma(f_{X\to\mu}^{sec}) = 0.5 \cdot f_{X\to\mu}^{sec}$ is considered for it in section 10.1.2.

The purity of the had_decay and $\gamma \to e$ proxy samples for the three beam configurations is shown in fig. 9.8 and fig. 9.9, respectively. The R_j^h results are shown in fig. 9.10 and fig. 9.11, respectively. The final fraction of secondaries based on its source for the three species is shown in fig. 9.12 and fig. 9.13 for the pPb and Pbp samples, respectively. The total fraction of secondaries is shown in fig. 9.15.

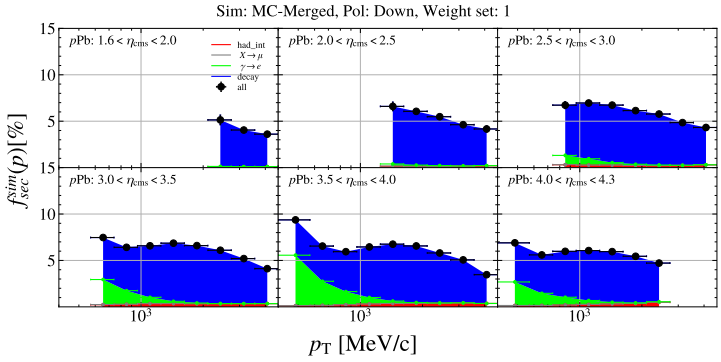
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Secondary tracks fraction in the pion sample.



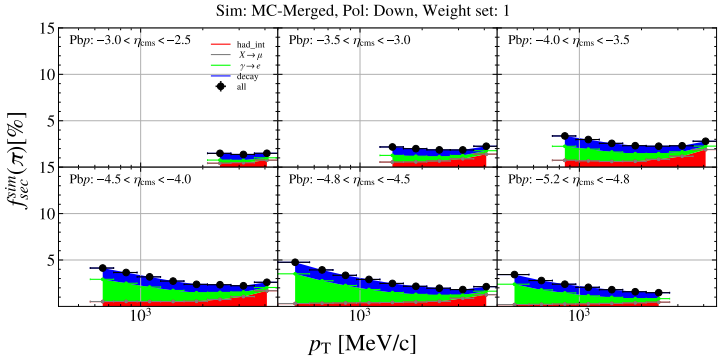
(b) Secondary tracks fraction in the kaon sample.



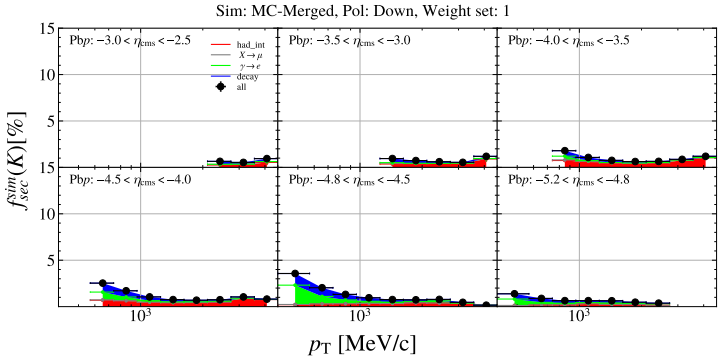
(c) Secondary tracks fraction in the proton sample.

Figure 9.5: Fraction of secondary tracks in simulation, $f_{sec}^{sim}(h)$, contaminating the candidate sample in pPb events for each hadron type: pions (top), kaons (middle), and protons (bottom). The stacked contributions from individual secondary sources are shown with the same labels as in the figure: had_int (red), $X \to \mu$ (brown), $\gamma \to e$ (green), and decay (blue). The black points labeled "all" correspond to the total fraction of secondary tracks. The occupancy weight set 1 is applied, only *MagDown* polarity is displayed, and only statistical uncertainties are shown.

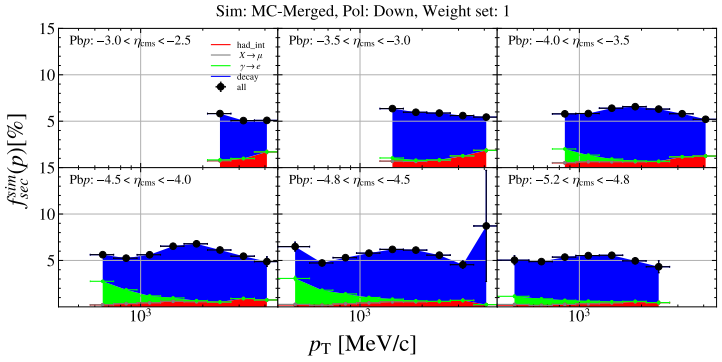
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Secondary tracks fraction in the pion sample.



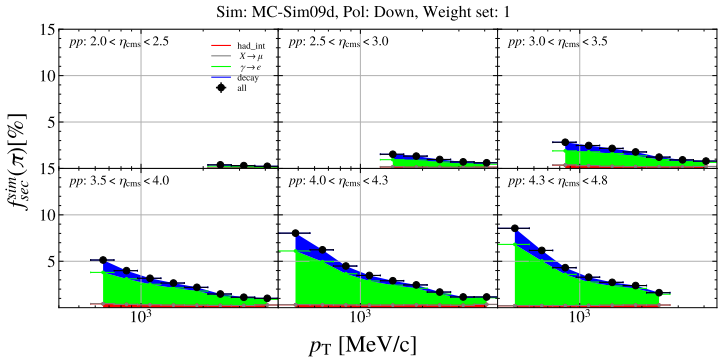
(b) Secondary tracks fraction in the kaon sample.



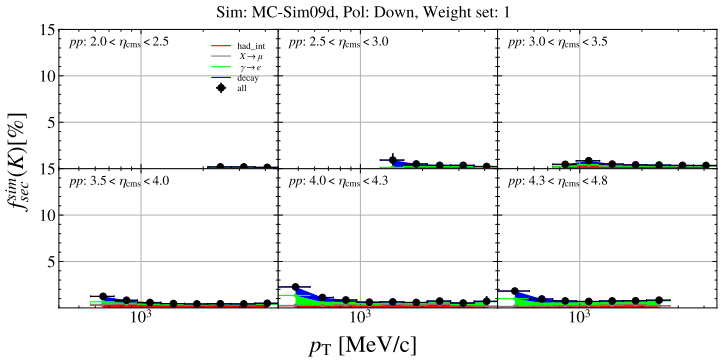
(c) Secondary tracks fraction in the proton sample.

Figure 9.6: Fraction of secondary tracks in simulation, $f_{sec}^{sim}(h)$, contaminating the candidate sample in Pbp events for each hadron type: pions (top), kaons (middle), and protons (bottom). The stacked contributions from individual secondary sources are shown with the same labels as in the figure: had_int (red), $X \to \mu$ (brown), $\gamma \to e$ (green), and decay (blue). The black points labeled "all" correspond to the total fraction of secondary tracks. The occupancy weight set 1 is applied, only *MagDown* polarity is displayed, and only statistical uncertainties are shown.

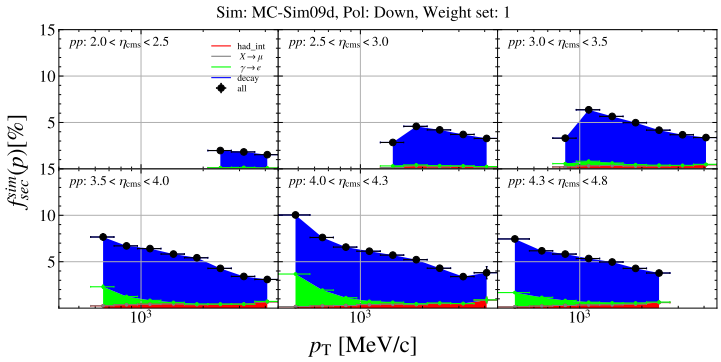
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Secondary tracks fraction in the pion sample.



(b) Secondary tracks fraction in the kaon sample.



(c) Secondary tracks fraction in the proton sample.

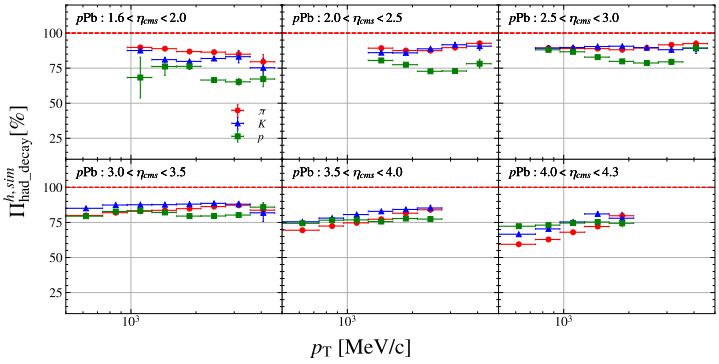
Figure 9.7: Fraction of secondary tracks in simulation, $f_{sec}^{sim}(h)$, contaminating the candidate sample in pp events for each hadron type: pions (top), kaons (middle), and protons (bottom). The stacked contributions from individual secondary sources are shown with the same labels as in the figure: had_int (red), $X \to \mu$ (brown), $\gamma \to e$ (green), and decay (blue). The black points labeled "all" correspond to the total fraction of secondary tracks. The occupancy weight set 1 is applied, and only statistical uncertainties are shown.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

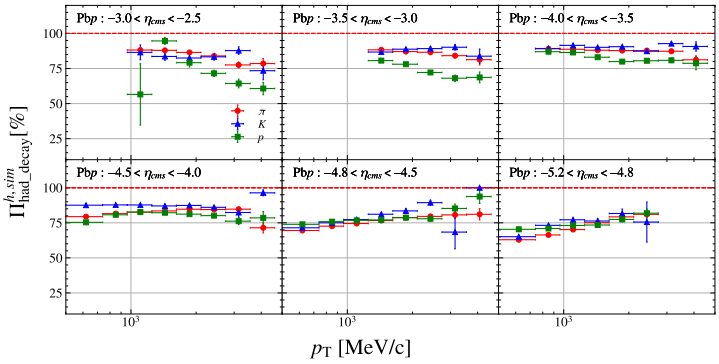
	pp	pPb	Pbp
Track type	Long	Long	Long
η	$2 < \eta < 4.8$	$2 < \eta < 4.8$	$2 < \eta < 4.8$
p_{T}	$0.5 < p_{\mathrm{T}} < 5\,\mathrm{GeV}/c$	$0.5 < p_{\mathrm{T}} < 5\,\mathrm{GeV}/c$	$0.5 < p_{\mathrm{T}} < 5\,\mathrm{GeV}/c$
p	$p > 10\,\mathrm{GeV}/c$	$p > 10\,\mathrm{GeV}/c$	$p > 10\,\mathrm{GeV}/c$
`GhostP`	$< 0.078\,\mathrm{mm}$	$< 0.103\,\mathrm{mm}$	$< 0.109\,\mathrm{mm}$
	PID selection		
pions	$\mathrm{DLL}_{K\pi} < -2$ & $\mathrm{DLL}_{p\pi} < -2$ & hasRich >0.5& `isMuon`==0	$\mathrm{DLL}_{K\pi} < 0$ & $\mathrm{DLL}_{p\pi} <10$ & hasRich >0.5 & `isMuon`==0	$\mathrm{DLL}_{K\pi} < 0$ & $\mathrm{DLL}_{p\pi} <10$ & hasRich >0.5 & `isMuon`==0
Kaons	$\mathrm{DLL}_{K\pi} > 5$ & ($\mathrm{DLL}_{K\pi}$-$\mathrm{DLL}_{p\pi}$) >5 & hasRich >0.5 & `isMuon`==0	$\mathrm{DLL}_{K\pi} > 5$ & ($\mathrm{DLL}_{K\pi}$-$\mathrm{DLL}_{p\pi}$) >0 & hasRich >0.5& `isMuon`==0	$\mathrm{DLL}_{K\pi} > 5$ & ($\mathrm{DLL}_{K\pi}$-$\mathrm{DLL}_{p\pi}$) >0 & hasRich >0.5& `isMuon`==0
protons	$\mathrm{DLL}_{p\pi} > 5$ & ($\mathrm{DLL}_{p\pi}$-$\mathrm{DLL}_{K\pi}$) > 10 & hasRich >0.5& `isMuon`==0	$\mathrm{DLL}_{p\pi} > 5$ & ($\mathrm{DLL}_{p\pi}$-$\mathrm{DLL}_{K\pi}$) > 15& hasRich >0.5& `isMuon`==0	$\mathrm{DLL}_{p\pi} > 5$ & ($\mathrm{DLL}_{p\pi}$-$\mathrm{DLL}_{K\pi}$) > 15& hasRich >0.5& `isMuon`==0
	PID selection		
sec. except. $\gamma \to e$	`pseudoIP` > 0.7	`pseudoIP` > 0.7	`pseudoIP` > 0.7
$\gamma \to e$	$\mathrm{DLL}_{e\pi} > 10$	$\mathrm{DLL}_{e\pi} > 10$	$\mathrm{DLL}_{e\pi} > 10$

Table 9.5: List of requirements for the secondary track proxy in each configuration.

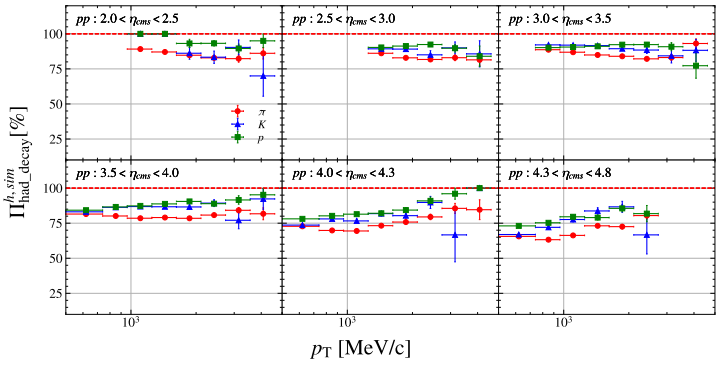
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Purity in pPb.

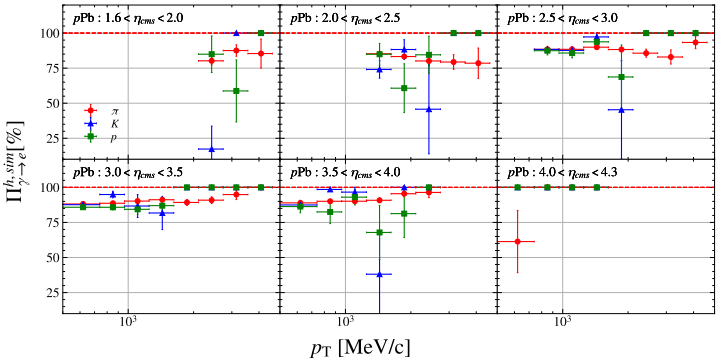


(b) Purity in Pbp.

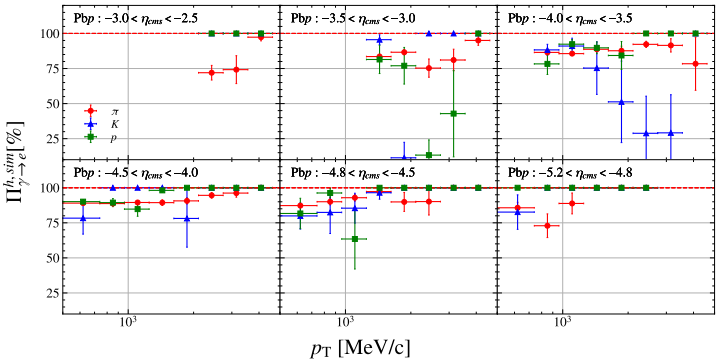


(c) Purity in pp.

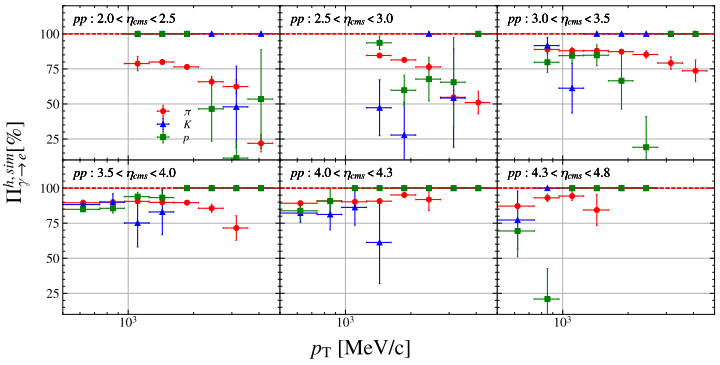
Figure 9.8: Purity of the proxy sample used to estimate the had_decay component of secondary tracks in pPb (top), Pbp (middle), and pp (bottom) collisions. Fractions are shown for each hadron type: pions (red), kaons (blue), and protons (green). Only statistical uncertainties are displayed.



(a) Purity in pPb.



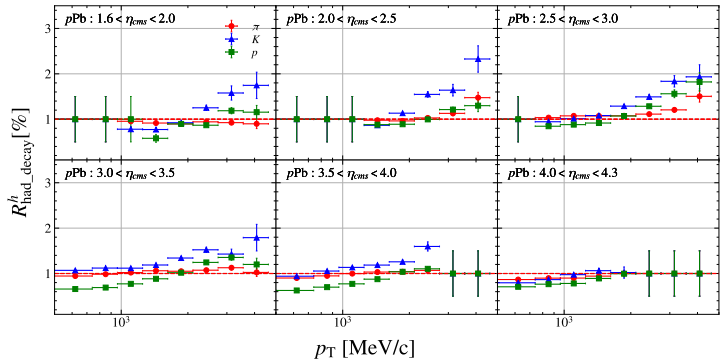
(b) Purity in Pbp.



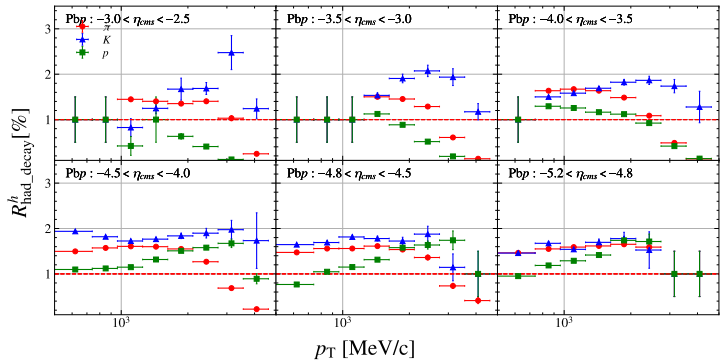
(c) Purity in pp.

Figure 9.9: Purity of the proxy sample used to estimate the $\gamma \rightarrow e$ component of secondary tracks in pPb (top), Pbp (middle), and pp (bottom) collisions. Fractions are shown for each hadron type: pions (red), kaons (blue), and protons (green). Only statistical uncertainties are displayed.

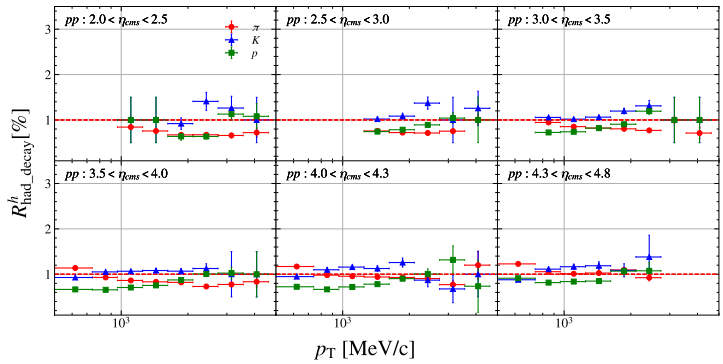
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Proxy tracks in pPb.



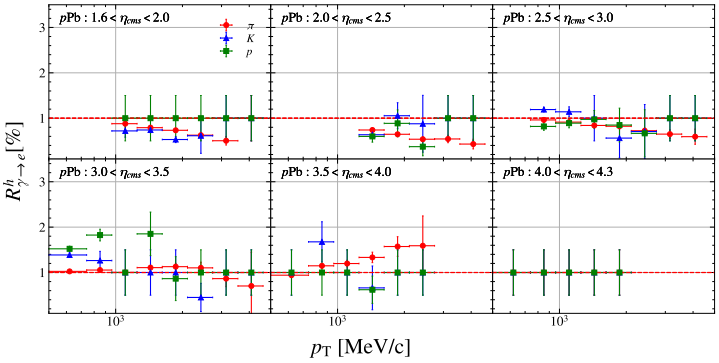
(b) Proxy tracks in Pbp.



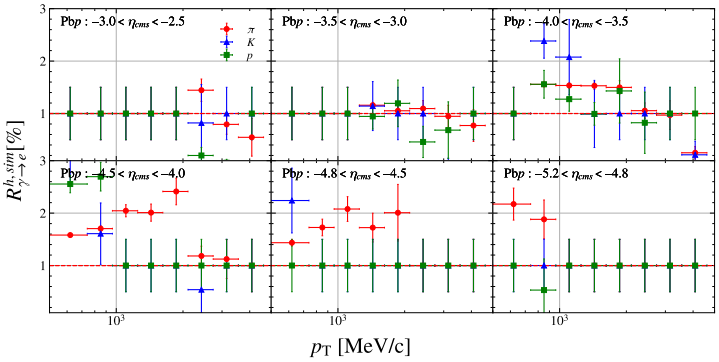
(c) Proxy tracks in pp.

Figure 9.10: Data-to-simulation correction factors R^h for the proxy sample of had_int tracks in pPb (top), Pbp (middle), and pp (bottom) collisions. Fractions are shown for each hadron type: pions (red), kaons (blue), and protons (green). Only statistical uncertainties are displayed. These factors quantify the discrepancy between simulation and data for hadronic-interaction secondary tracks.

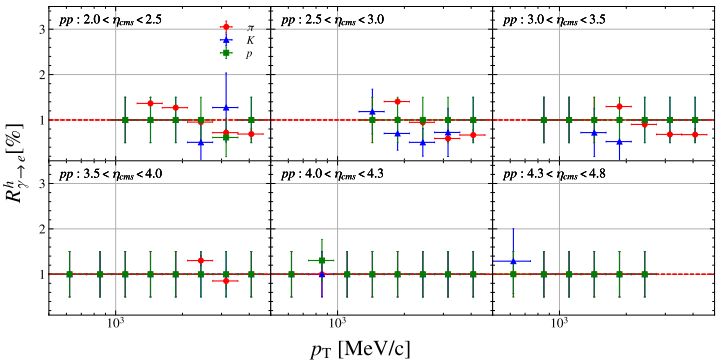
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Proxy tracks in pPb.



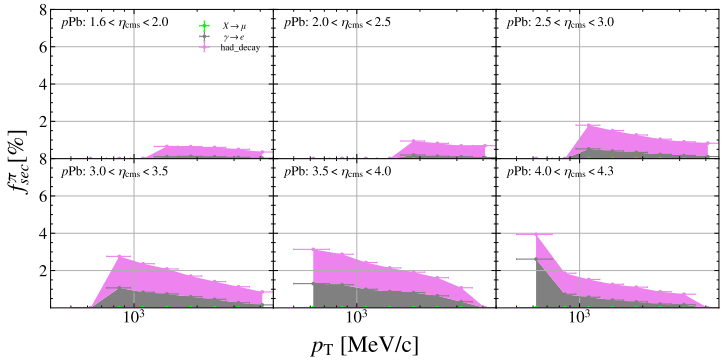
(b) Proxy tracks in Pbp.



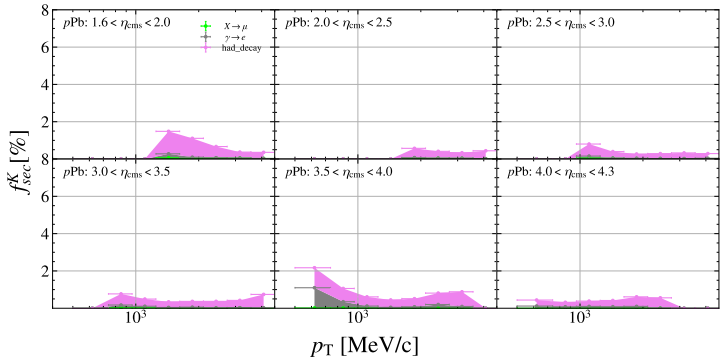
(c) Proxy tracks in pp.

Figure 9.11: Data-to-simulation correction factors R^h for the proxy sample of $\gamma \to e$ tracks in pPb (top), Pbp (middle), and pp (bottom) collisions. Fractions are shown for each hadron type: pions (red), kaons (blue), and protons (green). Only statistical uncertainties are displayed. These factors quantify the discrepancy between simulation and data for photon-conversion secondary tracks.

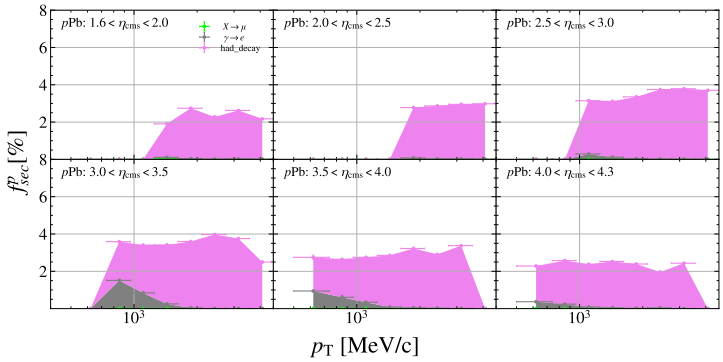
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Secondary tracks fraction in the pion sample.

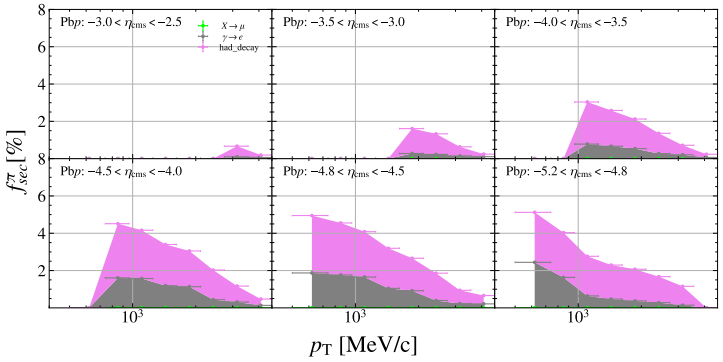


(b) Secondary tracks fraction in the kaon sample.

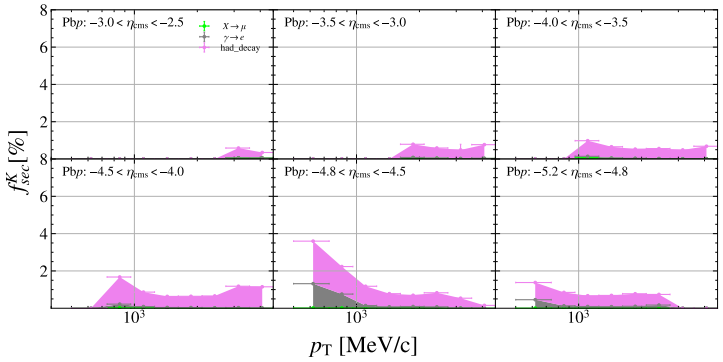


(c) Secondary tracks fraction in the proton sample.

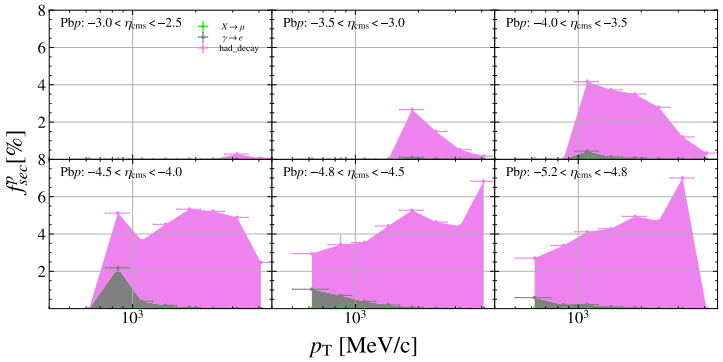
Figure 9.12: Fraction of secondary tracks in pPb collisions for each type of light hadron: pions (red), kaons (blue), and protons (green). The contributions are separated into three categories: muons from decays ($X \to \mu$, green), photon conversions ($\gamma \to e$, brown), and had_decay (pink). Statistical uncertainties are shown as vertical lines. These distributions illustrate the expected secondary-track contributions in the dataset.



(a) Secondary tracks fraction in the pion sample.



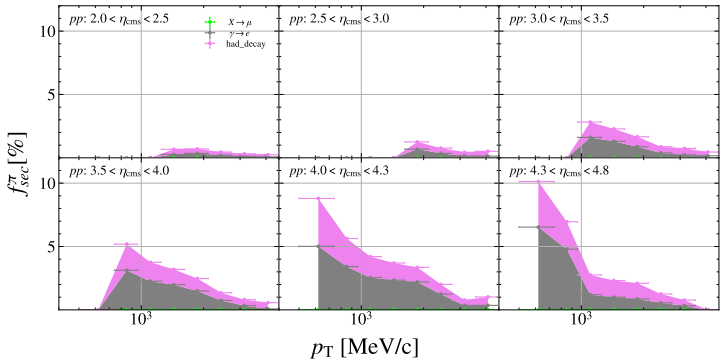
(b) Secondary tracks fraction in the kaon sample.



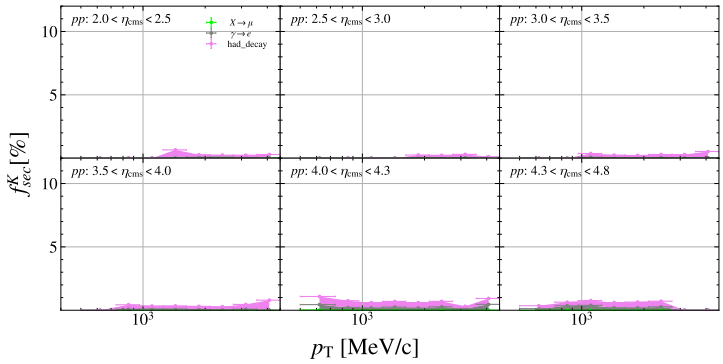
(c) Secondary tracks fraction in the proton sample.

Figure 9.13: Fraction of secondary tracks in Pbp collisions for each type of light hadron: pions (red), kaons (blue), and protons (green). The contributions are separated into three categories: muons from decays ($X \to \mu$, green), photon conversions ($\gamma \to e$, brown), and had_decay (pink). Statistical uncertainties are shown as vertical lines. These distributions illustrate the expected secondary-track contributions in the dataset.

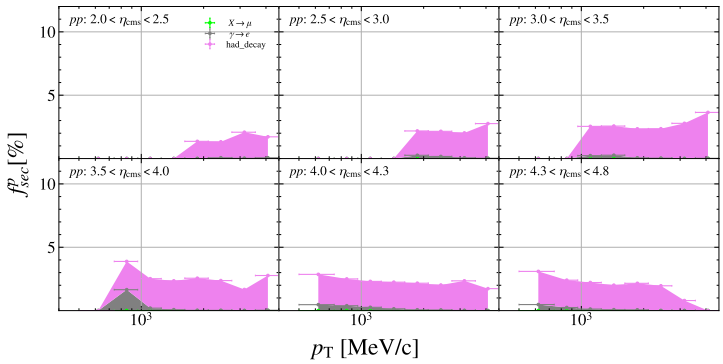
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Secondary tracks fraction in the pion sample.

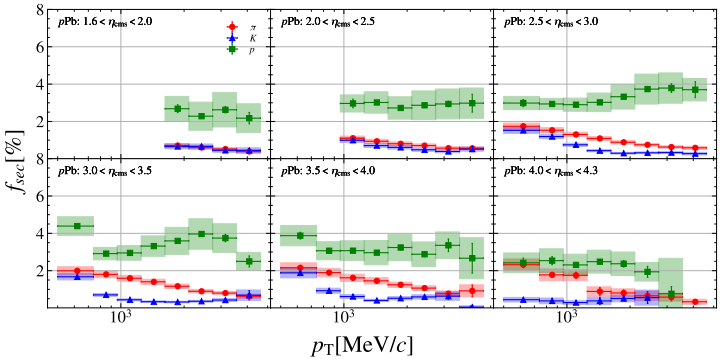


(b) Secondary tracks fraction in the kaon sample.

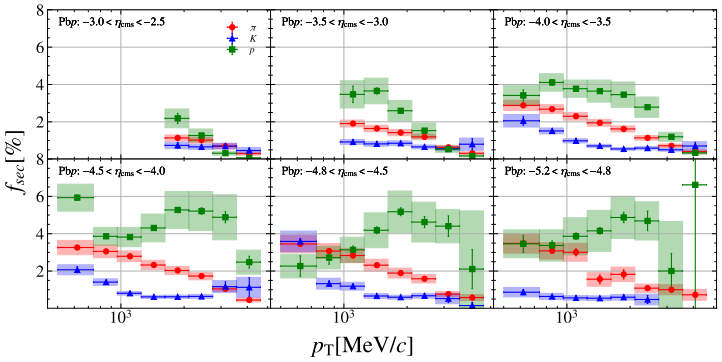


(c) Secondary tracks fraction in the proton sample.

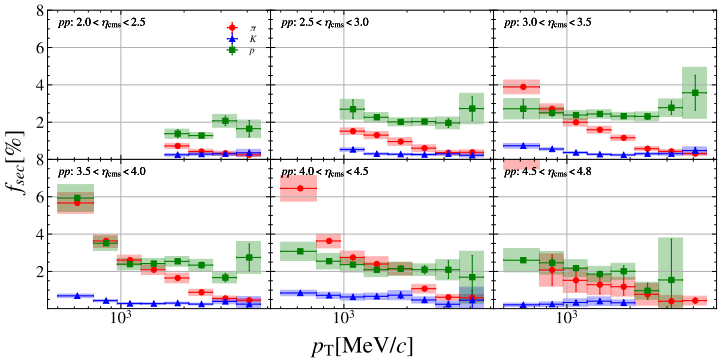
Figure 9.14: Fraction of secondary tracks in *pp* collisions for each type of light hadron: pions (red), kaons (blue), and protons (green). The contributions are separated into three categories: muons from decays ($X \to \mu$, green), photon conversions ($\gamma \to e$, brown), and had_decay (pink). Statistical uncertainties are shown as vertical lines. These distributions illustrate the expected secondary-track contributions in the dataset.



(a) Secondary tracks fraction in pPb.



(b) Secondary tracks fraction in Pbp.



(c) Secondary tracks fraction in pp.

Figure 9.15: Fraction of secondary tracks in pPb (top), Pbp (middle), and pp (bottom) collisions for each type of light hadron: pions (red), kaons (blue), and protons (green). Statistical uncertainties are shown as vertical lines, and systematic uncertainties are represented by boxes. These distributions illustrate the expected secondary-track contributions in the dataset.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

9.2.3 Purity

The purity, P^h, quantifies the fraction of true primary light hadrons in the sample after accounting for background contributions as discussed in section 9.2.2 and section 9.2.1. It is defined as:

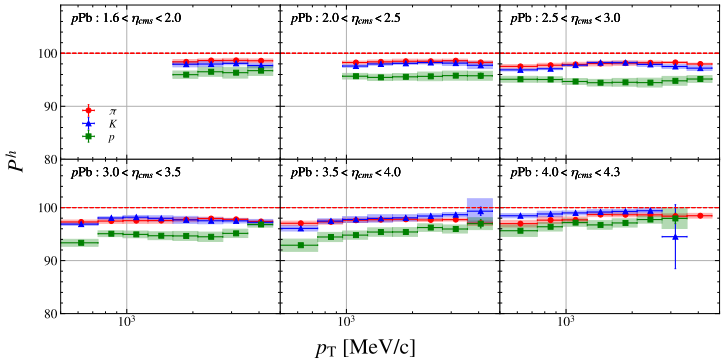
$$P^h = 1 - \sum R^h_\alpha f^{sim}_\alpha(h), \qquad \text{(eq. (9.2))}$$

the resulting purity values for the pPb, Pbp, and pp samples are presented in fig. 9.16. This purity is essential for correcting the initial charged light hadron yield N^h_{cand} to get the prompt contribution of the selected candidates defined as :

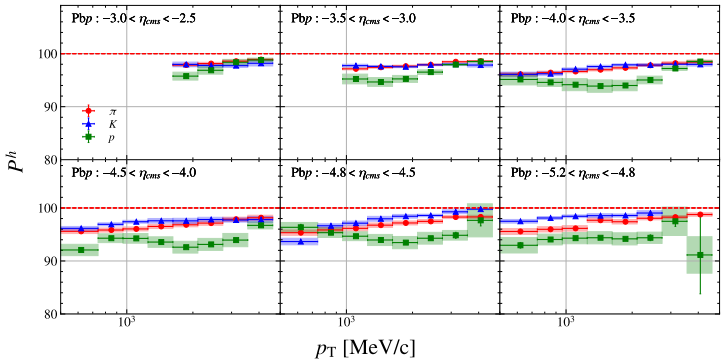
$$N^h_{Sel} = N^h_{cand} \cdot P^h. \qquad \text{(eq. (9.1))}$$

After applying this correction, only the PID efficiency correction remains to be applied to N^h_{Sel} to derive the final yield of the three species. This will be addressed in the subsequent section. The systematic uncertainties on the purity are discussed in section 10.1.3.

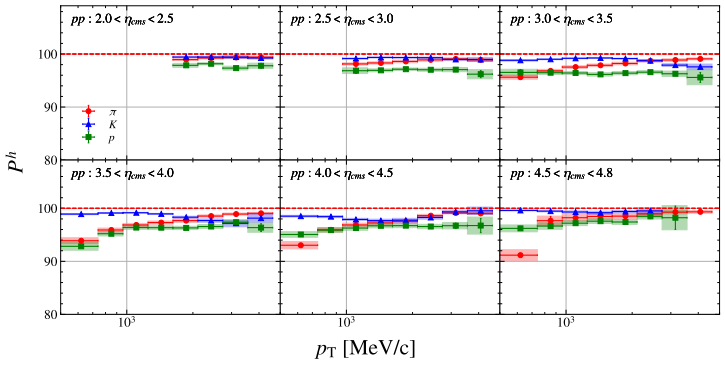
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Purity in pPb.



(b) Purity in Pbp.



(c) Purity in Pbp.

Figure 9.16: Percentage of the purity in pPb (top), Pbp (middle), and pp (bottom) collisions. The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green. Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

9.3 Separating Prompt Contributions

9.3.1 Particle Identification Efficiency

The PID efficiency for each particle type (π, K, p, μ, or e) is estimated using abundant samples of particle decays containing the corresponding species in the final state. The reconstruction of the parent particle invariant mass permits an independent identification of the particle. The efficiency is obtained using the tag-and-probe method: one particle is selected based on stringent PID criteria (the *tag*), and the other particle (the *probe*) is used to assess the PID performance. This approach isolates the probe particle, providing an unbiased measurement of the efficiency. Identification with the Cherenkov effect is very dependent of the total momentum of the particles [116]. Also, the algorithms for pattern recognition of rings in RICH detectors are very sensitive to the multiplicity of charged particles in the event. The LHCb collaboration prepares dedicated particle identification calibration samples for the PIDCalib tool [117]. The complete list of these samples and the particle species they are used for is shown in table 9.6. In this analysis we have considered a subset of these calibration samples, listed in table 9.7. The calibration samples and the kinematic regions of the analysed tracks were chosen to ensure that the former cover the p_{T} and p ranges of the latter. Consequently, the selections in this analysis differ from [103].

For pions an abundant calibration sample is obtained from $K^0_{\mathrm{S}} \to \pi^+\pi^-$ decays. This sample covers the low range of the p_{T} spectrum. This range is extended with the use of $D^{*+} \to D^0\pi^+$ decays, with $D^0 \to K^-\pi^+$. The soft pion produced in the strong decay of the D^{*+} hadron allows to tag the flavour of the D^0 and, consequently, to identify K^- and π^+ mesons. This channel is also used for kaons in the high p_{T} range, while the $\phi \to K^+K^-$ decay is considered to cover the low p_{T} region. The sample purity is further enhanced by requiring the ϕ meson to originate from $D_s^+ \to \phi\pi^+$ decays. Finally, for protons $\Lambda \to p\pi^-$ and $\Lambda_c^+ \to pK^-\pi^+$ decays are used for low and high p_{T} values, respectively. Since Λ decays are very abundant in LHCb, the kinematic coverage of the sample is extended by collecting datasets with different cuts on the proton p_{T} (or p) and prescaling the sub-samples[2] [118].

Using calibration samples collected during the same run as the analysed data is crucial for ensuring an accurate PID efficiency determination. This guarantees that the calibration precisely reflects the detector conditions at the time of data collection, accounting for any variations in the detector performance. For pp data collected during run II, all the samples specified in table 9.6 were available for PIDCalib. In contrast, for proton-lead collisions collected during run I, there were no calibration samples available. For this run period the samples are limited to the pp collisions high momentum calibration samples listed in table 9.6. These data are not ideal for pPb and Pbp collisions due to differences in kinematic coverage and multiplicity. Therefore, the calibration samples summarised in table 9.7 were specifically produced for pPb and Pbp data.

[2] In this analysis, these sub-samples are weighted by the inverse of the prescale factor and then merged. This is discussed in detail in appendix A.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Species	Low momentum	High momentum
$e^{\pm}$	$B^+ \to J/\psi K^+$ with $J/\psi \to e^+e^-$	
$\mu^{\pm}$	$B^+ \to J/\psi K^+$ with $J/\psi \to \mu^+\mu^-$	$J/\psi \to \mu^+\mu^-$
$\pi^{\pm}$	$K^0_S \to \pi^+\pi^-$	$D^{*+} \to D^0\pi^+$ with $D^0 \to K^-\pi^+$
$K^{\pm}$	$D^+_s \to \phi\pi^+$ with $\phi \to K^+K^-$	$D^{*+} \to D^0\pi^+$ with $D^0 \to K^-\pi^+$
p, $\overline{p}$	$\Lambda \to p\pi^-$	$\Lambda \to p\pi^-$ $\Lambda^+_c \to pK^-\pi^+$

Table 9.6: Decay channels used to calibrate PID efficiencies in LHCb [12].

Species	Low momentum	High momentum
$\pi^{\pm}$	$K^0_S \to \pi^+\pi^-$	$D^{*+} \to D^0\pi^+$ with $D^0 \to K^-\pi^+$
$K^{\pm}$	$\phi \to K^+K^-$	$D^{*+} \to D^0\pi^+$ with $D^0 \to K^-\pi^+$
p, $\overline{p}$	$\Lambda \to p\pi^-$	$\Lambda \to p\pi^-$

Table 9.7: Calibration samples considered in this analysis. The samples are from the decay channels used to calibrate PID efficiencies in LHCb [12].

9.3.2 Reweighting the Calibration Samples

In this analysis the efficiency is segmented in η and p_{T} bins, since these are the relevant variables for the final measurement. However, within these η and p_{T} bins, the total momentum distribution of the probe particles of the calibration samples may differ from that of the prompt particles under study. Also, the multiplicity distribution in calibration events can be different to that of the analysed data.

To minimise these discrepancies the calibration samples are reweighted. This is achieved with the GBRreweighter tool [114], also considered to improve the agreement between data and simulation, as discussed in section 8.4. Kinematic variables of the probe, $p_{\mathrm{T}}^{\mathrm{probe}}$, p^{probe}, η^{probe}, as well as event multiplicity proxies are used to reweight the calibration samples to match the target distributions in the minimum bias data sample, where PID requirements are not applied. The specific combinations of kinematic and occupancy variables used for reweighting are outlined in table 9.9. As previously explained, the performance of the RICH sub-detector relies on the total momentum of the particles, p^{probe}, as well as on the detector occupancy. For this reason, both variables are considered in the three proposed weight sets. Set 2 considers only these two dependencies. In addition, Set 1, which is used as baseline for the measurements, and Set 7 also consider the variables in which the efficiency is binned, η^{probe} and $p_{\mathrm{T}}^{\mathrm{probe}}$. These two sets differ in the multiplicity proxy and the discrepancies in the obtained results will be used to

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

asses systematic uncertainties. The result differences between Set 1 and Set 2 are also employed to estimate systematic effects originated by introducing the η and p_{T} binning. Other potential sets of variables were considered but ultimately discarded as they were found to give equivalent results to the ones already discussed.

9.3.3 Determination of the Efficiencies with the Tag & Probe Method

The PID efficiencies are determined with the tag-and-probe method, which involves fitting the mass distribution of the parent particle before and after applying the PID cuts. This technique allows for the extraction of efficiencies by comparing the yields of events with specific selections on the *tag* and *probe* particles. Initially, the mass distribution is fitted after applying a PID cut on the *tag* particle only. Then, a subsequent fit is performed with the PID cut applied to both the *tag* and *probe* particles. The ratio of the yields from these two fits is used to estimate the efficiency of the PID cut on the probe particle. For different decay modes, specific models are used to describe both the signal and background distributions. In the case of $\Lambda \to p\pi^-$ and $K^0_{\mathrm{S}} \to \pi^+\pi^-$ decays the signal is modelled using a double Gaussian function describing the detector resolution. For the $\phi \to K^+K^-$ decay a Voigtian function is used to model the signal, as it accounts for both the detector resolution and the natural width of the resonance. For the $D^{*+} \to D^0\pi^+$ decay, the fit is performed in the mass difference $\Delta M = M(D^{*+}) - M(D^0)$ instead of the individual masses, providing better resolution. The signal for this decay is modelled with a Gaussian function. For the decays $\Lambda \to p\pi^-$, $K^0_{\mathrm{S}} \to \pi^+\pi^-$, and $\phi \to K^+K^-$, the background is modeled using a Chebyshev polynomial. This polynomial effectively captures the smooth variations of the background beneath the signal peaks in these channels. In contrast, the background for the ΔM distribution is represented by a special PDF in RooFit, referred to as the `DstD0BKG` PDF. This PDF is specifically designed to accurately depict the unique shape of the background in the ΔM distribution. The tag particle selection criteria are summarised in table 9.8, and the mass fits of the parent particle for various collision types are shown in fig. 9.17, fig. 9.18, and fig. 9.19.

Taking into consideration the definition of the tag-and-probe method and the weights obtained after the reweighting procedure discussed in the previous sub-section, the PID efficiency ($h^{'} = h$) and the misidentification rates ($h^{'} \neq h$) for the calibration sample can be obtained as:

$$\varepsilon^{h \to h^{'}}_{tag\&probe,\alpha} = \frac{N^{h^{'}}_{\alpha}(tag\&probe)}{N_{\alpha}(tag)}, \tag{9.14}$$

where $N^{h^{'}}_{\alpha}(tag\&probe)$ is the yield from the fit to the mass distribution after applying the PID cuts on both the tag and probe, and $N_{\alpha}(tag)$ is the yield from the fit to the mass distribution of the mother of h after applying the PID cut only on the tag.[3] The $\alpha = \{\mathrm{Set\ 1}, \mathrm{Set\ 2}, \mathrm{Set\ 7}\}$ subscript indicates that the candidates have been reweighted

[3] The tag particle is not necessarily of the same h species as the probe particle.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Decay Samples	Probe Particle	Tag Particle	PID cut on the tag
$K^0_S \to \pi^+\pi^-$	π^+	π^-	$\mathrm{DLL}_{K\pi} < 0$
$D^{*+} \to D^0\pi^+$ with $D^0 \to K^-\pi^+$	π^+	K^-	$\mathrm{DLL}_{K\pi} < 0$
$\phi \to K^+K^-$	K^+	K^-	$\mathrm{DLL}_{K\pi} > 0$
$D^{*+} \to D^0\pi^+$ with $D^0 \to K^-\pi^+$	K^-	π^+	$\mathrm{DLL}_{K\pi} < 0$
$\Lambda \to p\pi^-$	p	π^-	$\mathrm{DLL}_{p\pi} < 0$

Table 9.8: Summary of PID cuts applied to tag particles for each calibration sample, including the decay process and associated probe particles.

Set Number	Variables
1	p^{probe}, p_{T}^{probe}, η^{probe}, `nVeloClusters`
2	p^{probe}, p_{T}^{probe}, η^{probe}, `nTracks`
3	p^{probe}, `nVeloClusters`

Table 9.9: Summary of variables used to reweight the calibration sample.

with the corresponding weight set prior to the invariant mass fits. In the following the α subscript will be omitted and the use of the reweight with the baseline Set 1 assumed implicitly. The reweighting with Sets 2 and 7 will be revisited when discussing the estimation of systematic uncertainties in section 10.2.

The choice of binning for measuring efficiencies depends on the calibration sample statistics. For PID efficiencies a finer binning is used for precise measurements, while a coarser binning is employed for mis-ID rates due to lower statistics. To ensure consistency with the binning of the analysis, the PID efficiencies are averaged as:

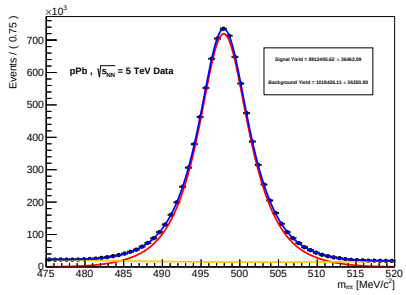
$$\varepsilon^{h\to h'}(\eta, p_{\mathrm{T}}) = \frac{\sum_{i=1}^{N^h_{cand}} \varepsilon^{h\to h'}_{tag\&probe}}{N^h_{cand}}, \tag{9.15}$$

where the sum runs over candidates within the computed bin. This strategy can be also applied to any given subsample of candidates. For instance, it will be used to compute efficiencies as a function of the total momentum in eq. (9.16).

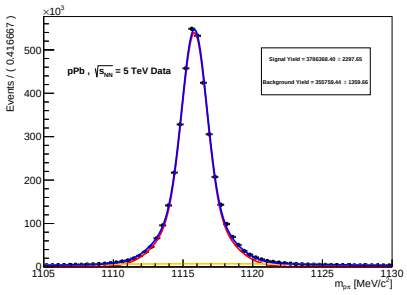
9.3.4 Correcting the Total Momentum Dependency per Species with an Iterative Data-Driven Method

As previously discussed, the performance of the RICH detector is primarily influenced by the particle total momentum and event multiplicity. While the event multiplicity is independent of the particle species, the total momentum spectrum of kinematic variables

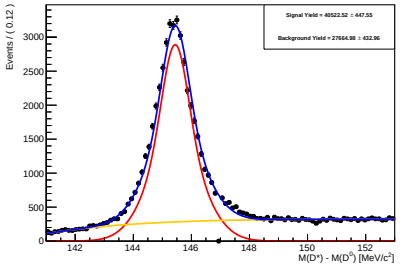
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



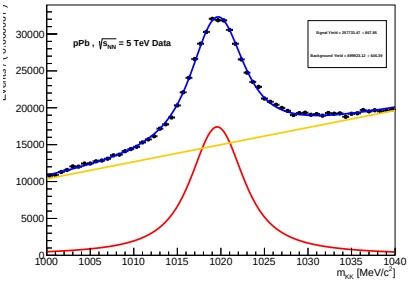
(a) $K^0_S \to \pi^+\pi^-$



(b) $\Lambda \to p\pi^-$



(c) $M(D^{*+}) - M(D^0)$



(d) $\phi \to K^+K^-$

Figure 9.17: Mass fits for different particles in pPb collisions: (a) $K^0_S \to \pi^+\pi^-$, (b) $\Lambda \to p\pi^-$, (c) and (d) $\phi \to K^+K^-$.

is specific to protons, kaons and pions. The sets introduced in section 9.3.2 consider the total momentum as one of the weighting variables. However, these weights make that the p^{probe} distributions of the calibration samples match the total momentum distribution of the minimum bias data ($p^{\text{min-bias}}$). The differences between the minimum bias and the particle species total momentum spectrum is corrected for with the following data-driven approach:

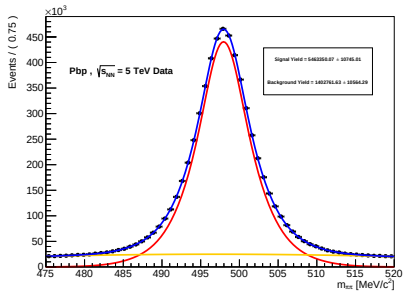
1) Considering candidate samples binned in the total momentum, N^h_{cand}, total momentum (p) dependent PID efficiencies are obtained as

$$\varepsilon^{h\to h'} = \frac{1}{N^h_{cand}} \sum_{i=1}^{N^h_{cand}} \varepsilon^{h\to h'}_{\text{tag\&probe}}(\eta_i, p_{\text{T},i}). \tag{9.16}$$

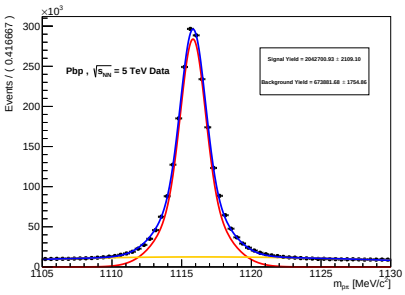
2) These efficiencies are applied to separate the particle species with the matrix strategy introduced in chapter 9. In particular, an analogous of eq. (9.3), binned in the total momentum, is employed:

$$\begin{bmatrix} N^\pi_{True} \\ N^K_{True} \\ N^p_{True} \end{bmatrix} = \begin{bmatrix} \varepsilon^{\pi\to\pi} & \varepsilon^{K\to\pi} & \varepsilon^{p\to\pi} \\ \varepsilon^{\pi\to K} & \varepsilon^{K\to K} & \varepsilon^{p\to K} \\ \varepsilon^{\pi\to p} & \varepsilon^{K\to p} & \varepsilon^{p\to p} \end{bmatrix}^{-1} \begin{bmatrix} N^\pi_{cand} \\ N^K_{cand} \\ N^p_{cand} \end{bmatrix}. \tag{9.17}$$

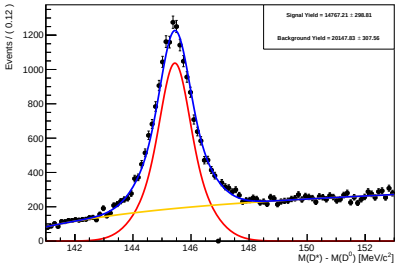
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



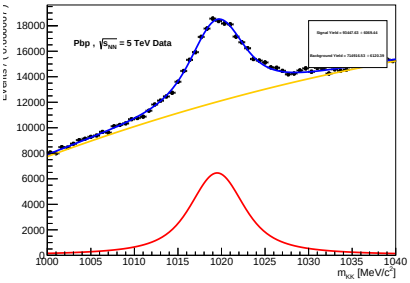
(a) $K^0_S \to \pi^+\pi^-$



(b) $\Lambda \to p\pi^-$



(c) $M(D^{*+}) - M(D^0)$



(d) $\phi \to K^+K^-$

Figure 9.18: Mass fits for different particles in Pbp collisions: (a) $K^0_S \to \pi^+\pi^-$, (b) $\Lambda \to p\pi^-$, (c) $D^0 \to K^-\pi^+$, and (d) $\phi \to K^+K^-$.

3) For a given total momentum bin the fraction of each particle species is calculated as:

$$\omega^h = \frac{N^h_{True}}{N_{cand}}. \tag{9.18}$$

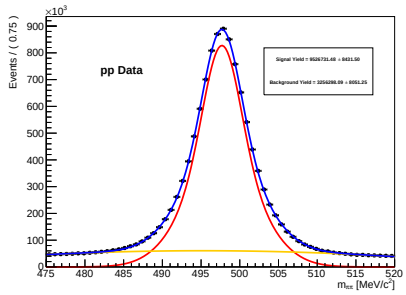
4) The reweighting, described in section 9.3.2, is repeated including ω^h as a total momentum dependent weight to the minimum-bias data, used as target, according to the species of the calibration sample. This procedure allows for the accurate matching of the p^{probe} distribution to the $p^{\text{min-bias}}$ distribution for each of the h species.

5) The PID efficiencies are again calculated, as described in section 9.3.3, with the updated weights for every α set according to eq. (9.14), which is now species dependent.

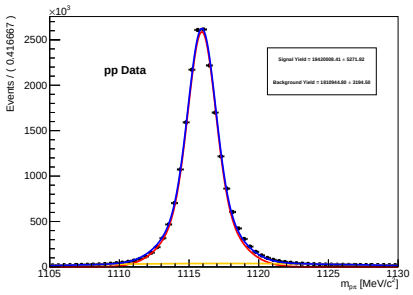
6) The process is iterated from 1) until convergence is reached. The differences between the last two iterations are considered as a systematic uncertainty as described in section 10.2.4. The iterative process is illustrated in fig. 9.20.

This procedure ensures that the momentum distribution of the probe particles in the calibration sample progressively aligns with the momentum distribution of the different

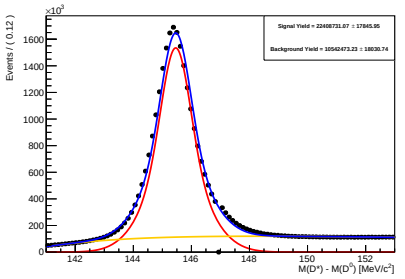
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



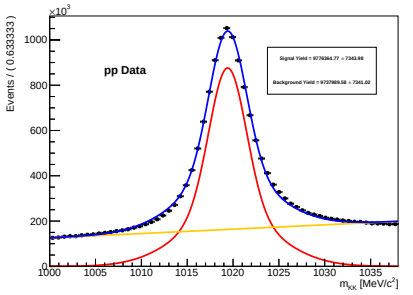
(a) $K^0_S \to \pi^+\pi^-$



(b) $\Lambda \to p\pi^-$



(c) $M(D^{*+}) - M(D^0)$



(d) $\phi \to K^+K^-$

Figure 9.19: Mass fits for different particles in pp collisions: (a) $K^0_S \to \pi^+\pi^-$, (b) $\Lambda \to p\pi^-$, (c) $D^0 \to K^-\pi^+$ and (d) $\phi \to K^+K^-$.

prompt particle species. Convergence is considered to be achieved when subsequent iterations show negligible changes in the PID efficiency matrix, indicating that the momentum distributions and efficiencies are consistent. The iterative reweighting method improves the precision of PID efficiency measurement and provides a more accurate characterisation of particle momentum distributions in the minimum bias data.

In this analysis, convergence was reached after five iterations, as no significant differences were observed between the third, fourth, and fifth iterations. Consequently, the efficiencies from the fifth iteration were chosen as the baseline for computing the yields of pions, kaons, and protons. The systematic uncertainty related to the PID efficiency estimation is discussed in detail in section 10.2.

The distributions of the kinematic variables from the final iteration are shown in fig. 9.21, fig. 9.22, fig. 9.23, fig. 9.24, and fig. 9.25. These figures compare the reweighted pPb data, shown in the green-shaded region, with the unweighted pPb data (grey-shaded region) and the decay particles from various resonances (red-hatched distributions). Each plot displays the kinematic variables for different particle species and decay channels after applying the final set of weights from the fifth iteration.

The weights obtained after the fifth iteration are used to compute the PID efficiency according to eq. (9.14). The binning is selected based on the statistical significance of the sample, with finer binning applied to the PID efficiencies and coarser binning for mis-ID

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

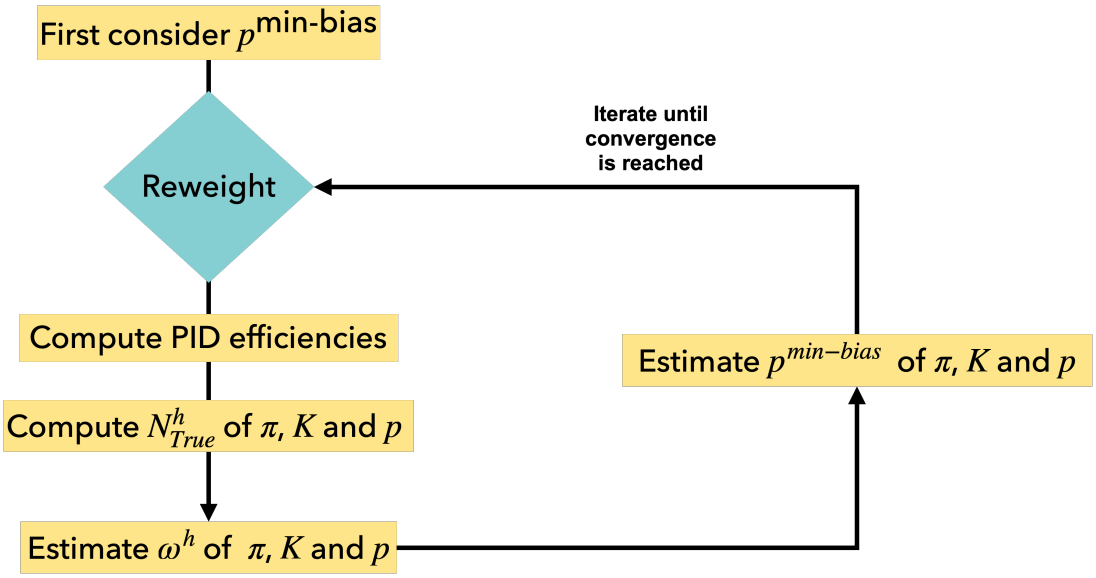


Figure 9.20: Flowchart illustrating the iterative reweighting process for PID efficiency calculations. The process begins with initial reweighting of probe momentum to the total momentum of the minimum bias data, followed by estimating the momentum distributions of pions, kaons and protons in the minimum bias data and iterative refinement of weights until convergence.

rates. For each (η, p_{T}) bin, the convergence of the fit is evaluated by calculating the χ^2/ndf and checking the fit status. If the fit fails to converge, it is repeated by adjusting the initial parameters until convergence is achieved for each (η, p_{T}) bin.

For the misidentification efficiency for kaon-to-pion and kaon-to-proton the resolution of the $\phi(1020)$ mass peak is insufficient for reliable particle identification, even when the sample is divided into (η, p_{T}) bins. This is the reason why the PIDCalib relies on the ϕ meson from D_s decay channels as a reference for improving the identification performance. However, in the pPb and Pbp minimum bias data, the sample of ϕ mesons from D_s decays is very limited, making it unsuitable for precise calibration. The only fit with sufficiently high resolution that could be reconstructed was the mass difference $\Delta M = M(D_s^*) - M(D^0)$. Nevertheless, due to the limited statistics, this fit could only be performed for a single (η, p_{T}) bin. The obtained efficiency is used as a correction factor to the mis-ID rates obtained from simulation instead of using one single bin of

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

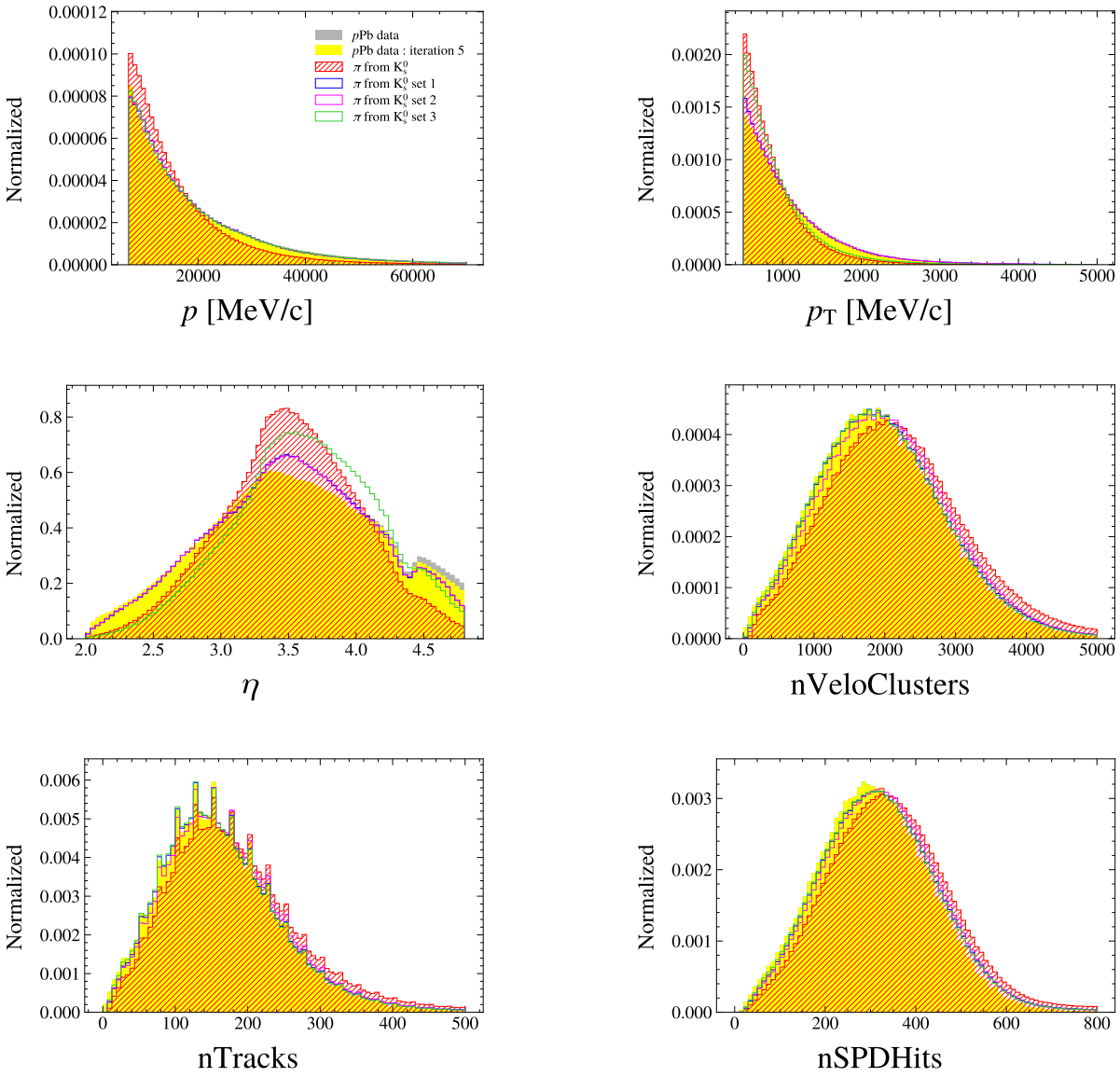


Figure 9.21: Kinematic distributions of charged-pion candidates in pPb data. The grey shaded region shows the unweighted minimum-bias sample (all charged tracks). The yellow shaded region corresponds to the same sample after applying pion-specific weights derived from the iterative procedure (see section 9.3.4). Pions from $K^0_S \to \pi^+\pi^-$ decays are shown as a red hatched distribution, while the same $K^0_S \to \pi^+\pi^-$ sample reweighted with sets 1, 2, and 3 (see section 9.3.2) are shown in blue, pink, and green, respectively.

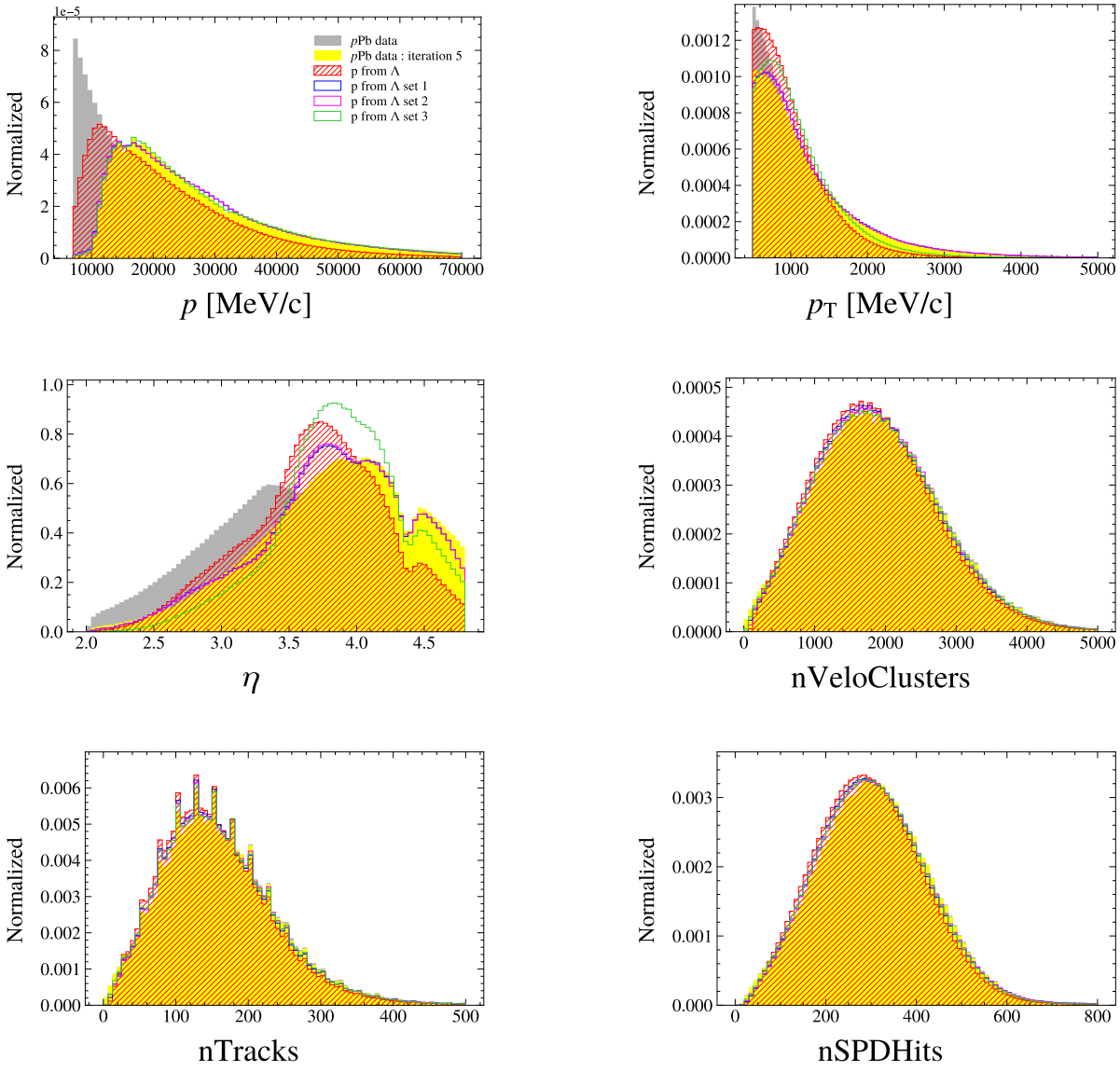


Figure 9.22: Kinematic distributions of proton candidates in pPb data. The grey shaded region shows the unweighted minimum-bias sample (all charged tracks). The green shaded region with stars corresponds to the same sample after applying the proton-specific weights from the fifth iteration of the iterative procedure (see section 9.3.4). Protons from $\Lambda \to p\pi^-$ decays are shown as a red hatched distribution, while the same $\Lambda \to p\pi^-$ sample reweighted with sets 1, 2, and 3 (see section 9.3.2) are shown in blue, magenta, and green, respectively.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

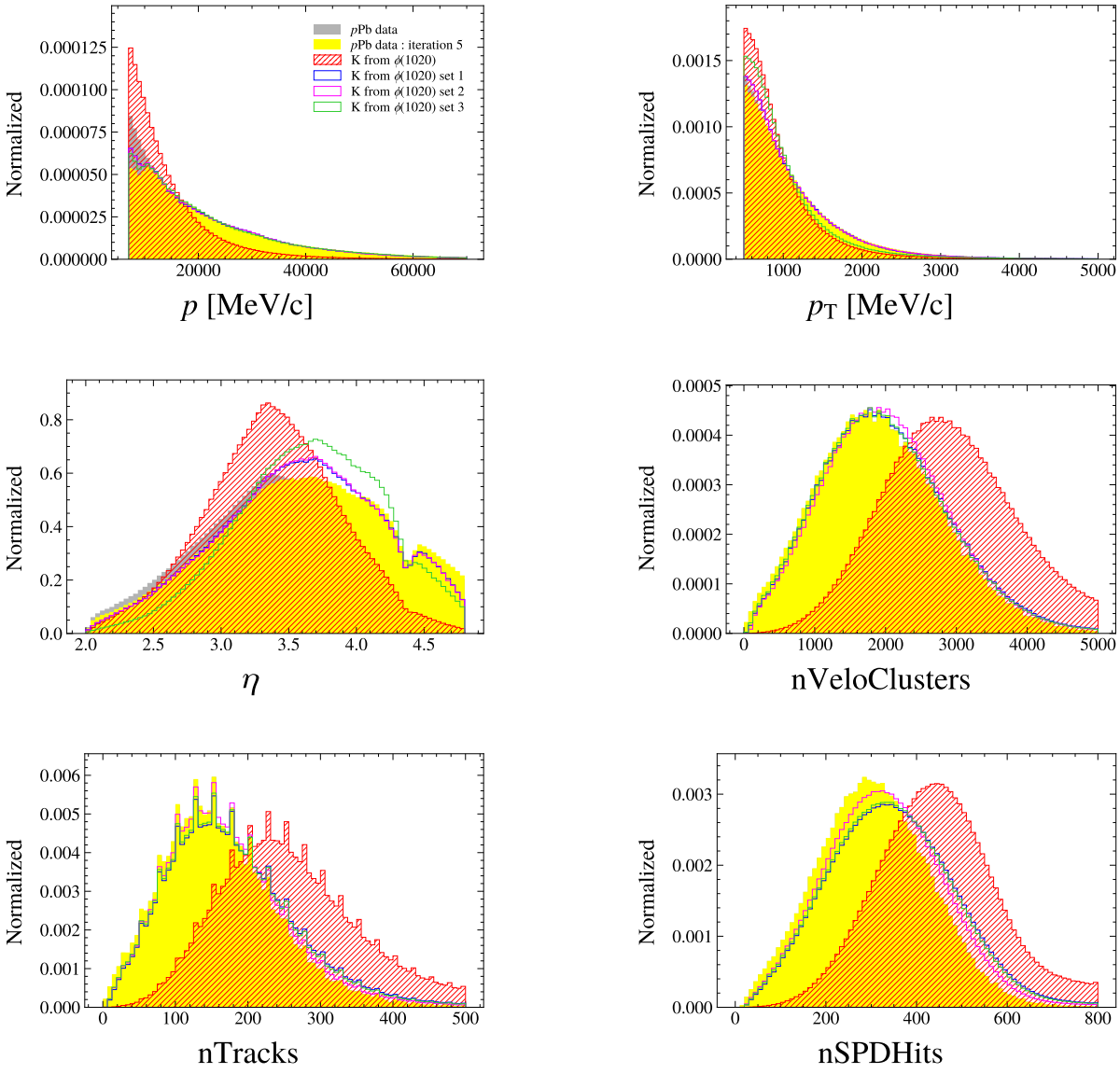


Figure 9.23: Kinematic distributions of kaon candidates in pPb data. The grey shaded region shows the unweighted minimum-bias sample (all charged tracks). The green shaded region with stars corresponds to the same sample after applying kaon-specific weights from the fifth iteration of the iterative procedure (see section 9.3.4). Kaons from $\phi \to K^+K^-$ decays are shown as a red hatched distribution, while the same $\phi \to K^+K^-$ sample reweighted with sets 1, 2, and 3 (see section 9.3.2) are shown in blue, magenta, and green, respectively.

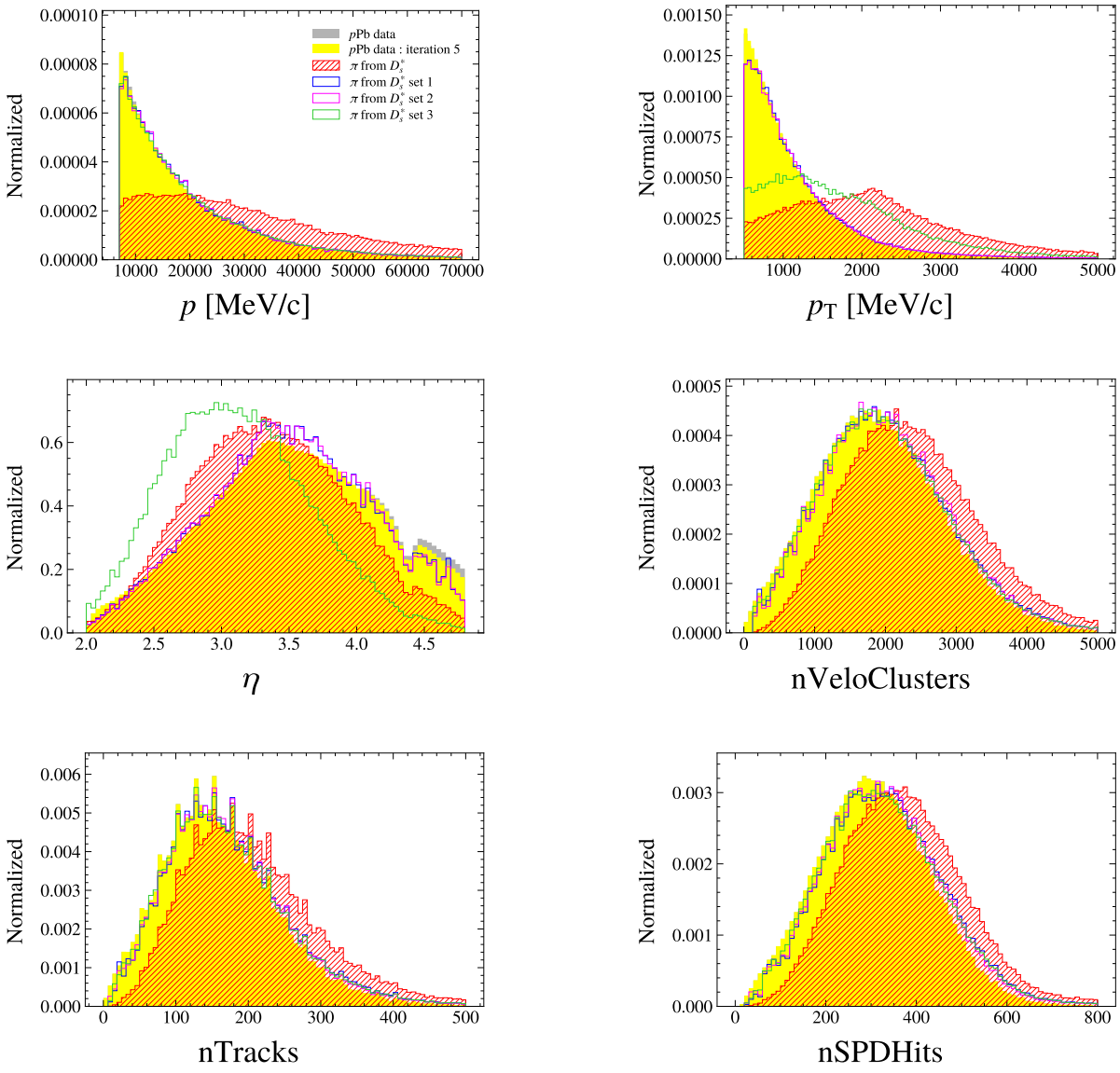


Figure 9.24: Kinematic distributions of charged-pion candidates in pPb data. The grey shaded region shows the unweighted minimum-bias sample (all charged tracks). The green shaded region with stars corresponds to the same sample after applying pion-specific weights from the fifth iteration of the iterative procedure (see section 9.3.4). Pions from $D^{*+} \to D^0\pi^+$ decays are shown as a red hatched distribution, while the same $D^{*+} \to D^0\pi^+$ sample reweighted with sets 1, 2, and 3 (see section 9.3.2) are shown in blue, magenta, and green, respectively.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

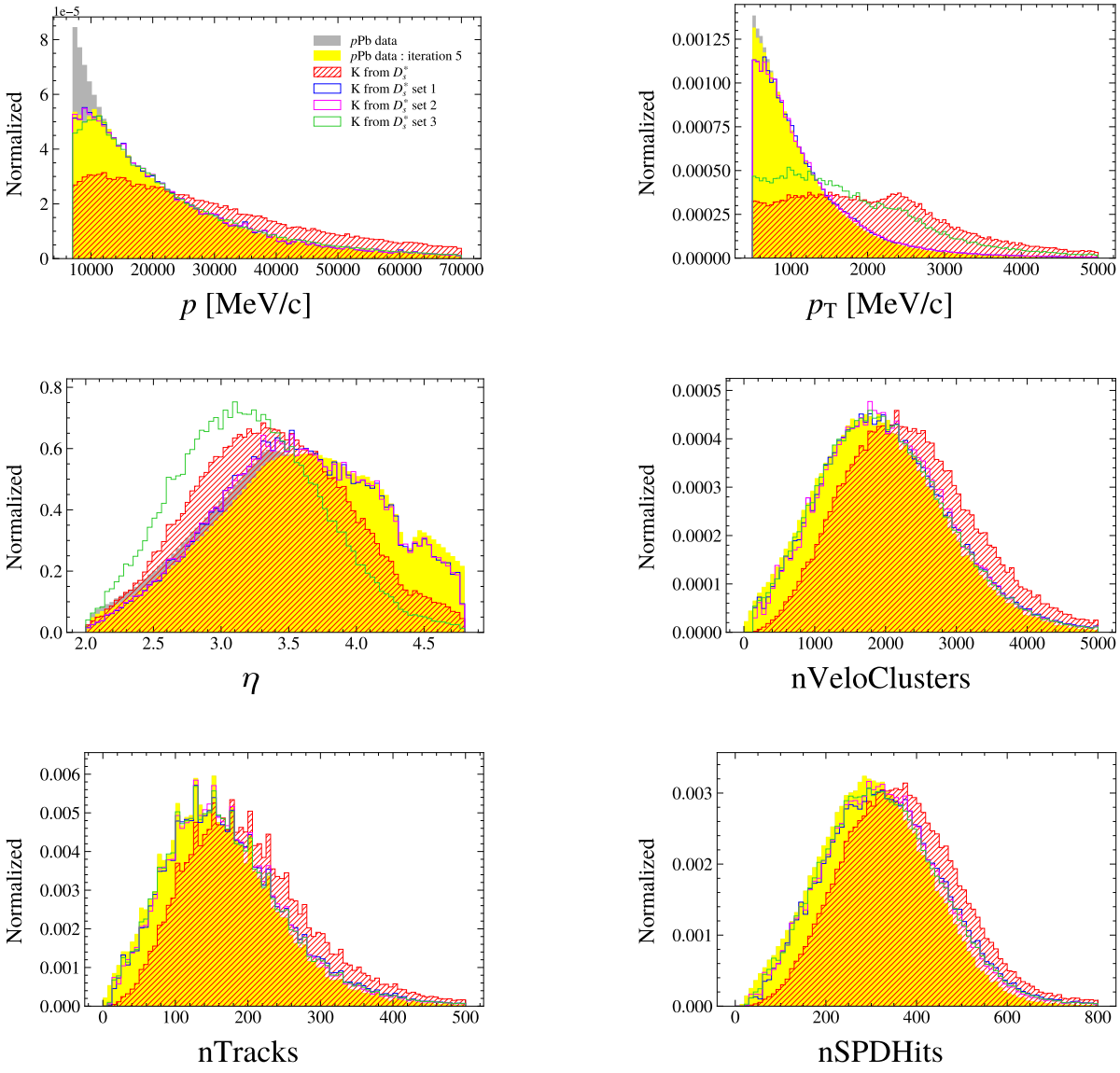


Figure 9.25: Kinematic distributions of charged-kaon candidates in pPb data. The grey shaded region shows the unweighted minimum-bias sample (all charged tracks). The green shaded region with stars corresponds to the same sample after applying pion-specific weights from the fifth iteration of the iterative procedure (see section 9.3.4). Kaons from $D^{*+} \to D^0\pi^+$ decays are shown as a red hatched distribution, while the same $D^{*+} \to D^0\pi^+$ sample reweighted with sets 1, 2, and 3 (see section 9.3.2) are shown in blue, magenta, and green, respectively.

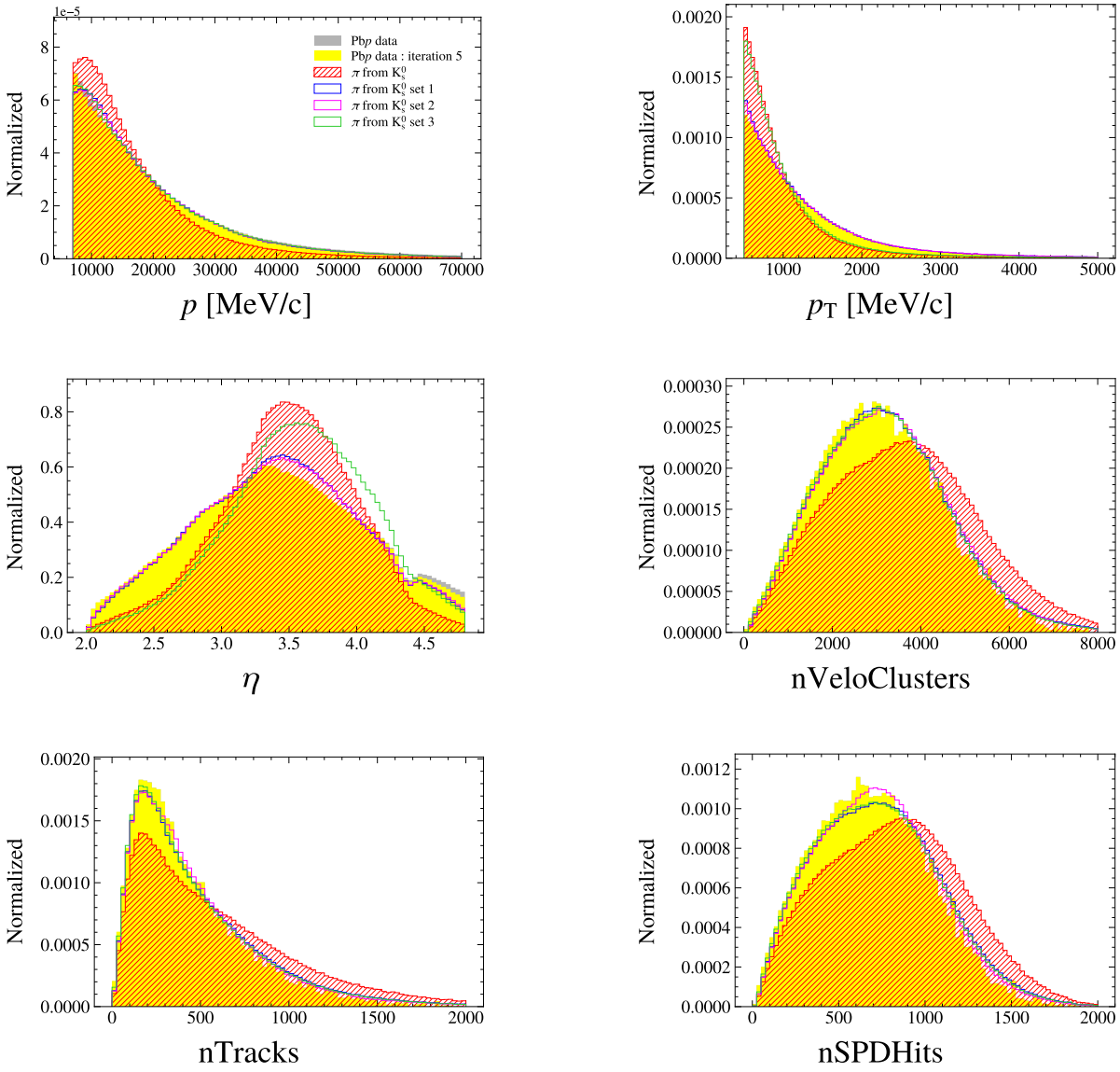


Figure 9.26: Kinematic distributions of charged-pion candidates in Pbp data. The grey shaded region shows the unweighted minimum-bias sample (all charged tracks). The yellow shaded region corresponds to the same sample after applying pion-specific weights derived from the iterative procedure (see section 9.3.4). Pions from $K^0_S \to \pi^+\pi^-$ decays are shown as a red hatched distribution, while the same $K^0_S \to \pi^+\pi^-$ sample reweighted with sets 1, 2, and 3 (see section 9.3.2) are shown in blue, pink, and green, respectively.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

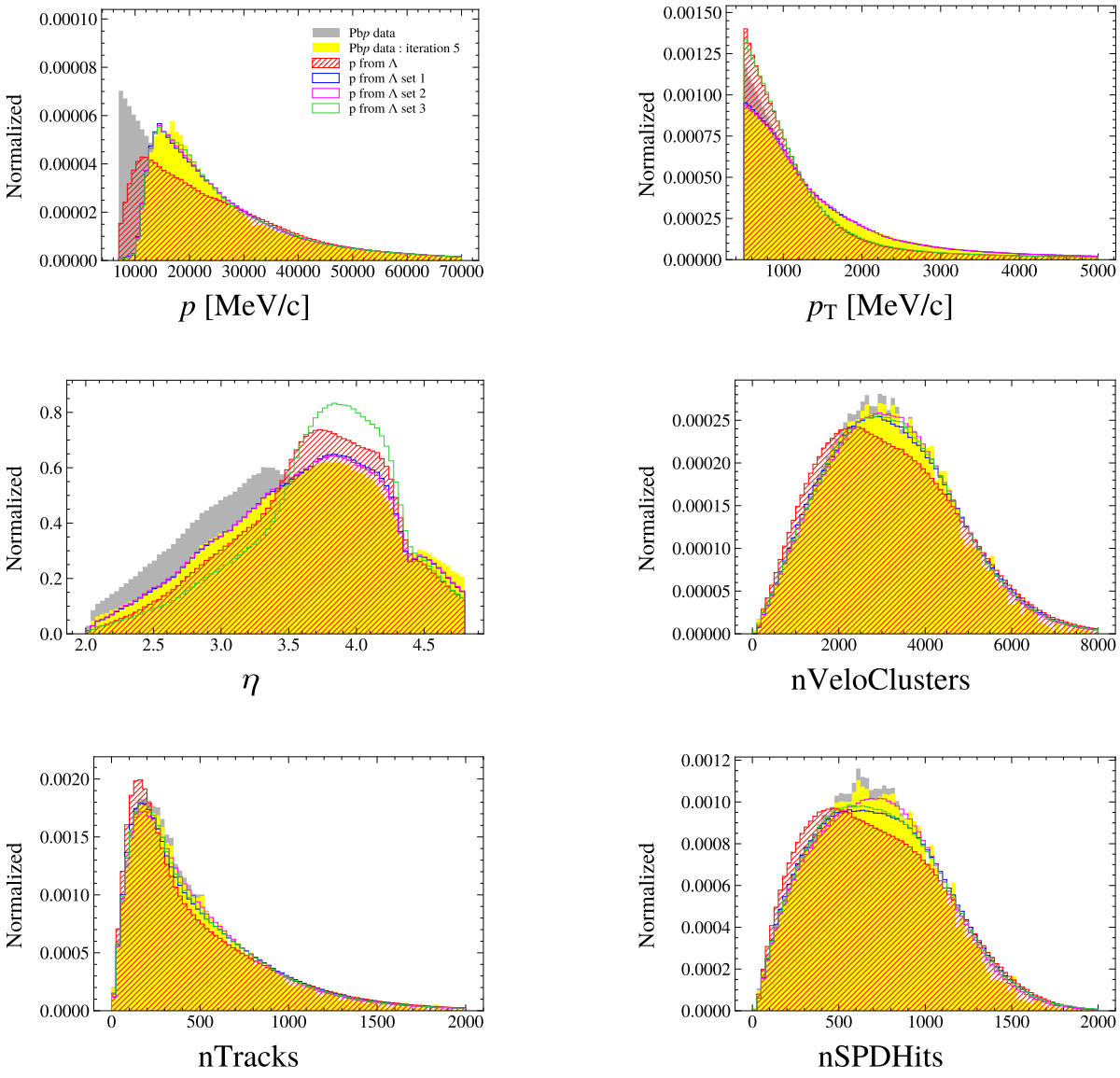


Figure 9.27: Kinematic distributions of proton candidates in Pbp data. The grey shaded region shows the unweighted minimum-bias sample (all charged tracks). The green shaded region with stars corresponds to the same sample after applying the proton-specific weights from the fifth iteration of the iterative procedure (see section 9.3.4). Protons from $\Lambda \to p\pi^-$ decays are shown as a red hatched distribution, while the same $\Lambda \to p\pi^-$ sample reweighted with sets 1, 2, and 3 (see section 9.3.2) are shown in blue, magenta, and green, respectively.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

misidentification. This misID rate is calculated using the following equation:

$$\varepsilon^{K\to(\pi,p)} = \frac{\sum_i^{N^K_{cand,sim}} \varepsilon^{K\to(\pi,p)}_{tag\&probe}(\eta_i, p_{\mathrm{T},i})}{N^K_{cand,sim}}. \tag{9.19}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

9.4 Reconstruction Efficiency

The reconstruction efficiency is computed from simulation as the ratio of the number of reconstructed prompt-charged tracks to the number of generated prompt-charged particles for each light hadron. Both the reconstructed tracks and the generated particles are required to be within the LHCb spectrometer acceptance, discussed in section 8.2.1. Since the reconstruction efficiency is computed with simulated events, any discrepancies between simulation and data need to be accounted for. This is done by considering a correction factor,

$$C_{tracking} = \varepsilon^{\text{data}}_{tracking}/\varepsilon^{\text{MC}}_{tracking}, \tag{9.20}$$

where $\varepsilon^{\text{data}}_{tracking}$ and $\varepsilon^{\text{MC}}_{tracking}$ are the tracking efficiency in data and MC simulation, respectively. This efficiency is calculated in both datasets using $J/\psi \to \mu^+\mu^-$ decays,[4] due to their abundance and wide momentum range. In the event, the J/ψ is reconstructed from one long track with muon identification (*tag*) and another that is partially reconstructed (*probe*). The reconstruction efficiency is then estimated as the fraction of reconstructed J/ψ decays where the probe track can be matched to a long track. The $C_{tracking}$ factor is provided as a table by the TrackCalib package [120] in bins of η and total momentum for different collisions. These tabulated values are considered for each simulation candidate in addition to the event occupancy weights w of set 1,[5] introduced in section 8.4 to compute the reconstruction efficiency as:

$$\varepsilon^{h}_{reco} = \frac{\sum_i^{total\,reco\,h} C_{tracking}(p_i, \eta_i) \cdot w_i}{\sum_i^{total\,gen\,h} w_i}, \tag{9.21}$$

where, for a given species h, the sum in the numerator is over the long tracks matched with a generated particle and in the denominator over the total number of generated particles.

No $C_{tracking}$ table was produced for the 2013 proton-lead data. The one from the 2012 proton-proton collisions is used instead. This is justified as both datasets are collected during run I and the detector condition is the same in both samples. The $C_{tracking}$ tracking factors for this sample, with statistical uncertainties, are shown in fig. 9.28 in bins of η and p in GeV/c. The systematic uncertainty for the values in this table is 0.4% [121].

For proton-proton events the tracking calibration table from 2015 proton-proton collisions is used. The $C_{tracking}$ correction factor is computed with $B^+ \to J/\psi K^+$ decays, where a cut of `GhostP` < 0.3 was applied to the tracks. In the present analysis reconstructed tracks that do not satisfy this cut are considered without calibration and are assigned a systematic uncertainty, see section 10.3. The values of $C_{tracking}$ with statistical uncertainties are shown in fig. 9.29 in bins of η and p in MeV/c. The systematic uncertainty of the table is 0.8%. [122].

[4] In run I, the efficiency was computed using a clean sample of $J/\psi \to \mu^+\mu^-$ decays [119]. In run II, the purity of the sample is enhanced with J/ψ mesons from b-hadron decays [120].

[5] The other two weight sets, sets 2 and 3, are used to estimate systematic uncertainties originated in the occupancy weighting in section 10.3.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

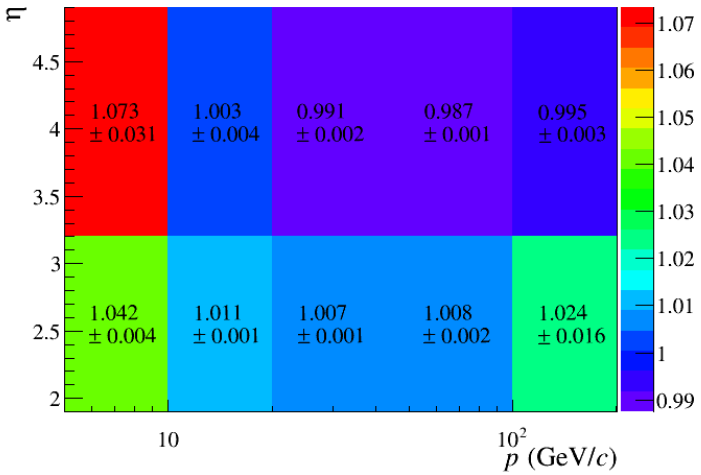


Figure 9.28: Ratio between data and simulation tracking efficiency $C_{tracking} = \varepsilon^{\text{data}}_{tracking}/\varepsilon^{\text{MC}}_{tracking}$ in 2012 pp calibration data.

For proton-lead, the reconstruction efficiency is computed separately for each magnet polarity and MC samples and the discrepancies are found to be negligible.

The reconstruction efficiency results for pions, kaons and protons are shown in bins of η_{cms} and p_{T} in MeV/c, for pPb in fig. 9.30a, Pbp in fig. 9.30b, and pp in fig. 9.30c, with the name of the MC sample, magnet polarity configuration and weight set used to produce the plots as a title.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

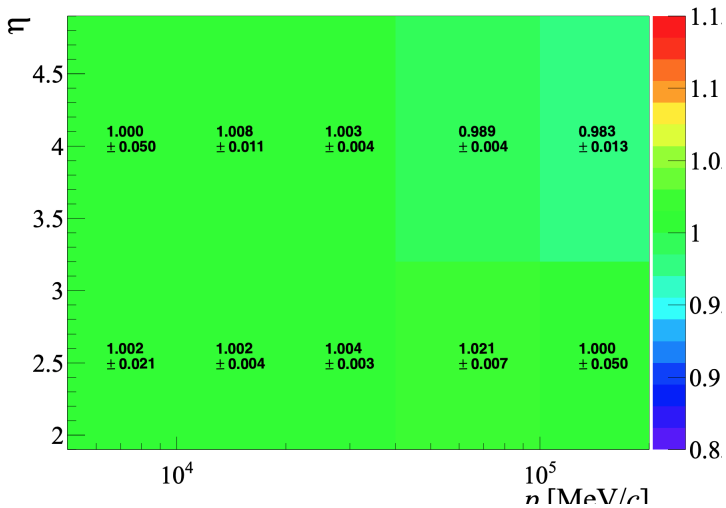
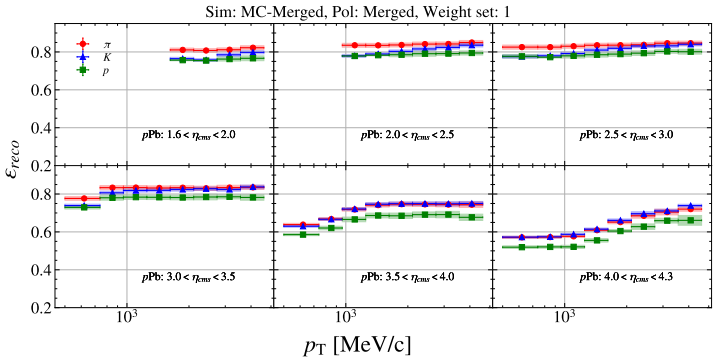
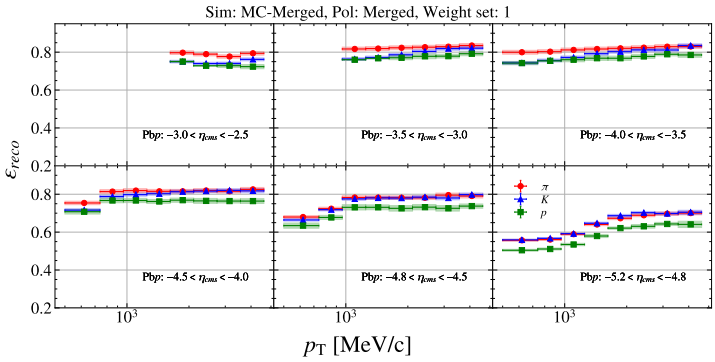


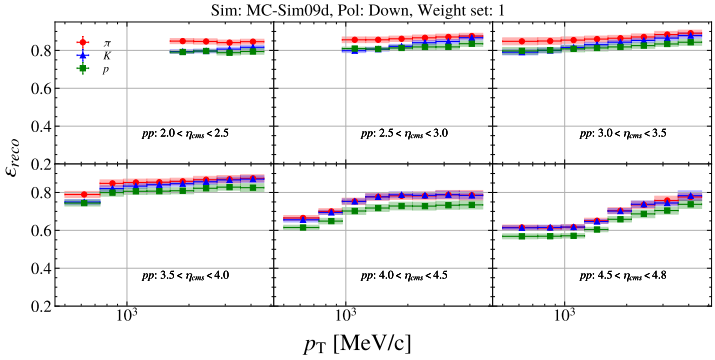
Figure 9.29: Ratio between data and simulation tracking efficiency $C_{tracking} = \varepsilon^{\text{data}}_{tracking}/\varepsilon^{\text{MC}}_{tracking}$ in 2015 pp calibration data.



(a) Reconstruction efficiency in pPb events.



(b) Reconstruction efficiency in Pbp events.



(c) Reconstruction efficiency in pp events.

Figure 9.30: Reconstruction efficiency in pPb (top), Pbp (middle), and pp (bottom) events. The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green. Statistical uncertainties are indicated with vertical lines, while systematic uncertainties are displayed as boxes. This efficiency quantifies the probability that a charged hadron produced in the collision is successfully reconstructed as a track by the detector.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

9.5 Selection Efficiency

For each light hadron, the selection efficiency is calculated with simulated events as a weighted ratio of matched reconstructed tracks that pass the candidate selection criteria, summarised in table 8.4 (*sel tracks*), to the total number of matched long tracks within the LHCb detector (*reco tracks*):

$$\varepsilon^{h}_{\text{sel}} = \frac{\sum_{sel\ tracks} C_{\text{sel}}(\eta^{i}_{\text{cms}}, p^{i}_{\text{T}})}{N_{reco\ tracks}}. \tag{9.22}$$

The C_{sel} weights account for simulation and data discrepancies estimating the ratio

$$C_{\text{sel}} \sim \frac{\varepsilon^{\text{data}}_{\text{sel}}}{\varepsilon^{\text{sim}}_{\text{sel}}}, \tag{9.23}$$

where $\varepsilon^{\text{data}}_{\text{sel}}$ and $\varepsilon^{\text{sim}}_{\text{sel}}$ are the data and simulation selection efficiencies, respectively, in bins of $(\eta_{\text{cms}}, p_{\text{T}})$.

In this analysis C_{sel} is determined using tag-and-probe kaons from $\phi \to K^+K^-$ decays in data and simulation samples. The yield of ϕ candidates in these samples is obtained fitting the K^+K^- mass spectrum with a Probability Density Function (PDF) combining a Voigtian lineshape for the signal and a background function defined as:

$$f(M) = q \cdot m_b^{a_1} \cdot e^{-a_2 \cdot m_b}, \tag{9.24}$$

where q is the momentum of a daughter in the resonance rest frame, a_1 and a_2 are free parameters in the fit, and $m_b = M(K^+K^-) - 2m_K$. The yields of signal and background are left free in the fit. Tag and probe kaons from ϕ decays are selected applying the cuts in table 9.10 and C_{sel} is defined as

$$C_{\text{sel}} = \frac{\varepsilon^{\phi\text{-data}}_{\text{sel}}}{\varepsilon^{\phi\text{-sim}}_{\text{sel}}}, \tag{9.25}$$

where

$$\varepsilon^{\phi\text{-data(-sim)}}_{\text{sel}} = \frac{\text{Yield of } \phi \text{ mesons in data (simulation) after tag and probe kaon selection}}{\text{Yield of } \phi \text{ mesons in data (simulation) after tag kaon selection}}. \tag{9.26}$$

The selection efficiency, ε_{sel}, is used to correct N^{h}_{True} in eq. (9.4) to account for prompt charged light hadrons reconstructed by the spectrometer that do not meet the selection criteria for `GhostP` and `pseudoIP`. The systematic uncertainty in its determination is estimated by computing C_{sel} with a different PDF model, as discussed in section 10.4. The selection efficiency coefficient is illustrated in fig. 9.31. The results, provided for different (η, p_{T}) bins, include both the statistical uncertainty from the signal yields of the fit and the systematic uncertainty. The coefficient is shown for pPb collisions in fig. 9.31a,

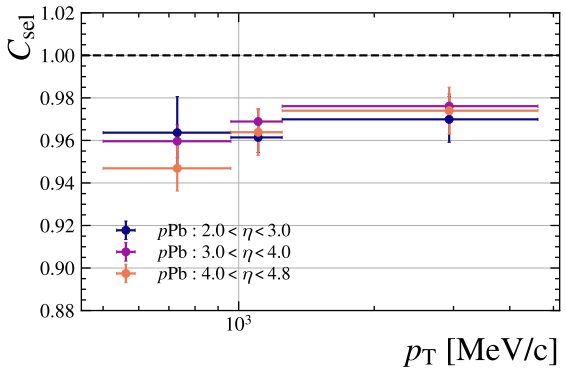
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

$K^{\pm}$ (tag)	$K^{\mp}$ (probe)	$\phi(1020)$ (combination cuts)
`StdAllNoPIDsKaons` probNNK > 0.5	`StdAllNoPIDsKaons` Long tracks In LHCb acceptance	$M(K^{\pm}K^{\mp}) < 1040$ MeV/c $M(K^{\pm}K^{\mp}) > 1000$ MeV/c DOCA$(K^{\pm}, K^{\mp}) < 0.2$

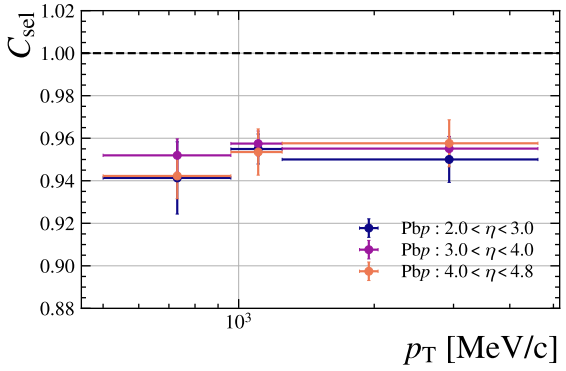
Table 9.10: Selection criteria for $K^{\pm}$ (tag), $K^{\mp}$ (probe), and $\phi(1020)$ (combination cuts).

for Pbp collisions in fig. 9.31b and for pp collisions in fig. 9.31c, respectively. The binning in p_{T} and η is finer in pPb and Pbp collisions, with three bins in rapidity. In contrast, due to the limited statistics, for the pp collisions only two rapidity bins are used. The values of C_{sel} are generally close to unity, ranging between 0.95 and 0.98. In pPb collisions, the selection efficiency varies slightly across the three rapidity bins, with $2.0 < \eta < 3.0$ exhibiting slightly higher coefficient, especially at low p_{T}. As p_{T} increases, the values of C_{sel} tend to converge across the different rapidity bins. A similar pattern is observed in Pbp collisions with consistent behaviour across different p_{T} ranges. In pp collisions, the values of the selection coefficient are close to 0.98 across the p_{T} range, and the difference between the two rapidity bins is minimal. Overall, C_{sel} remains consistently close to unity across all collision systems. The selection efficiencies for pions, kaons, and protons in pPb, Pbp, and pp collisions are presented in fig. 9.32. The systematic uncertainties are discussed in section 10.4.

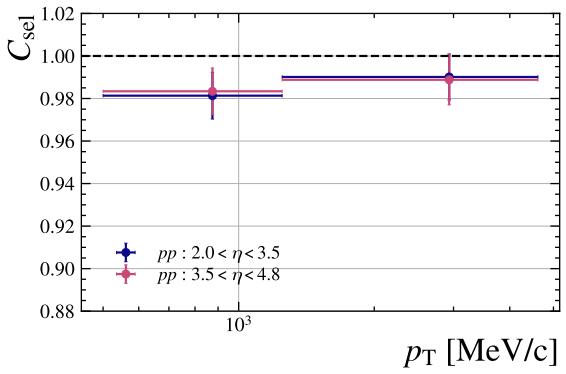
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) C_{sel} for pPb events in (η, p_{T}) bins.



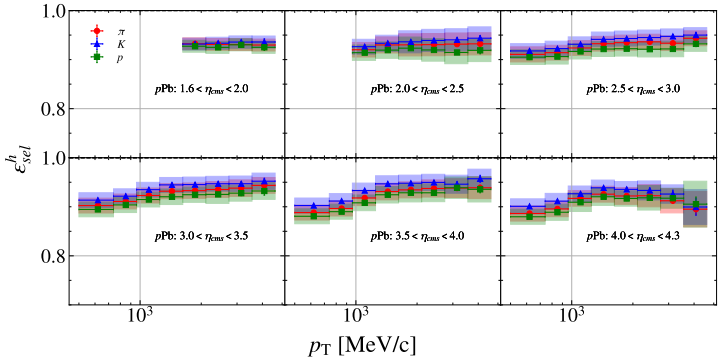
(b) C_{sel} for Pbp events in (η, p_{T}) bins.



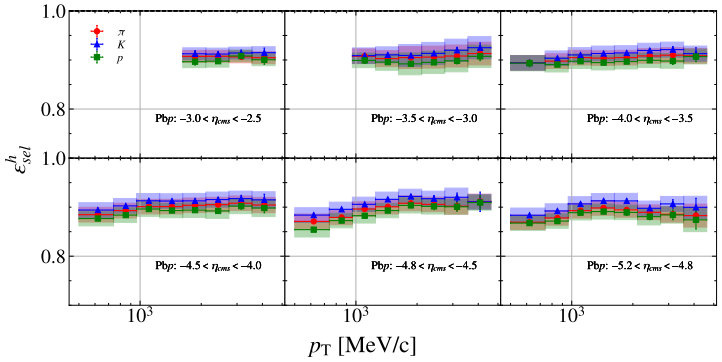
(c) C_{sel} for pp events in (η, p_{T}) bins.

Figure 9.31: Selection efficiency coefficients (C_{sel}) are shown as a function in $(\eta_{\mathrm{cms}}, p_{\mathrm{T}})$ bins for different collision systems: (a) pPb (top), (b) Pbp (middle), and (c) pp (bottom).

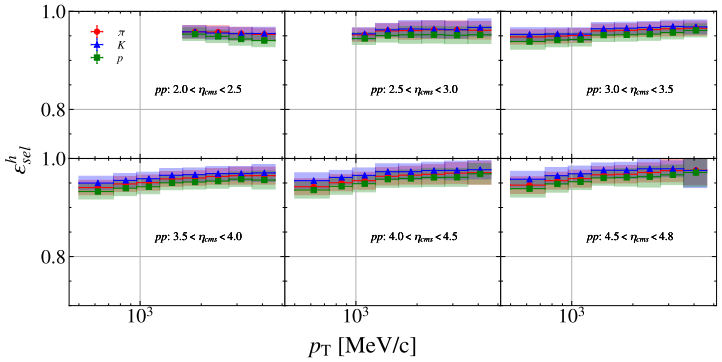
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Selection efficiency in pPb events.



(b) Selection efficiency in Pbp events.



(c) Selection efficiency in pp events.

Figure 9.32: Selection efficiencies are shown as a function of $(\eta_{\text{cms}}, p_{\text{T}})$ bins for different collision systems: (a) pPb (top), (b) Pbp (middle), and (c) pp (bottom). The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green. Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes. Only *MagDown* polarity is shown in the figure.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

9.6 Truth Matching Efficiency

Following the prompt charged tracks analysis [103], the truth-matching efficiency evaluates the performance of the truth-matching algorithm in establishing the correspondence between generator-level particles and reconstructed tracks. A reconstructed track that fails to match its corresponding `MC` particle is labelled as fake. This inefficiency can be estimated using the tag-and-probe method applied to simulated $\phi \to K^+K^-$ events, similarly to the method used for determining selection efficiency in section 9.5. In this approach, the tag kaon is truth-matched to its corresponding `MC` particle (`TrueID==321`), while the probe kaon is considered as fake (`TrueID==0`). The proportion of non-matched candidates is given by:

$$\varepsilon_{\text{non-matched}} = \frac{\text{Yield of } \phi \text{ meson in simulation after tag and probe kaon selection}}{\text{Yield of } \phi \text{ meson in simulation after tag kaon selection}}. \quad (9.27)$$

The truth-matching efficiency is then calculated as:

$$\varepsilon_{TM} = 1 - \varepsilon_{\text{non-matched}}. \quad (9.28)$$

This efficiency is included in the evaluation of N^h_{True} in eq. (9.4). The calculation of ε_{TM} is performed across different kinematic bins of $(\eta_{\text{cms}}, p_{\text{T}})$ for the probe kaon. Due to the low percentage of mismatched tracks, only four kinematic bins are considered. The truth-matching efficiency, as shown in table 9.11, varies across different experimental configurations, with the highest efficiency observed in the pp configuration, followed by pPb, and then Pbp configurations. Similar to the method described in section 9.5, the systematic uncertainty is estimated by varying the PDF used to model the ϕ meson.

p_{T} [GeV/c]	η	$\varepsilon_{TM;pp}$ (`MCsim09d`)	$\varepsilon_{TM,p\text{Pb}}$(`MCsim09e`)	$\varepsilon_{TM;\text{Pb}p}$ (`MCsim09e`)
$0.2 < p_{\text{T}} < 1.249$	$2 < \eta < 3.2$	0.9944 ± 0.0015	0.99029 ± 0.00056	0.98942 ± 0.00070
$0.2 < p_{\text{T}} < 1.249$	$3.2 < \eta < 4.8$	0.9913 ± 0.0012	0.98873 ± 0.00059	0.98785 ± 0.00062
$1.249 < p_{\text{T}} < 8$	$2 < \eta < 3.2$	0.9903 ± 0.0017	0.98708 ± 0.00014	0.98358 ± 0.00082
$1.249 < p_{\text{T}} < 8$	$3.2 < \eta < 4.8$	0.9870 ± 0.0021	0.985514 ± 0.000082	0.98267 ± 0.00012

Table 9.11: Values of ε_{TM} for the pPb, Pbp, and pp simulation samples. Results taken from [103].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

10

Systematic Uncertainties

10.1 Purity Calculation

The determination of the number of selected hadrons in eq. (9.1) relies on the estimation of the sample purity, obtained after subtracting the contributions from fake tracks and secondary tracks, as described in section 9.2.

The purity is defined in eq. (9.2) as:

$$P^h = 1 - \sum R^h_\alpha f^{sim}_\alpha(h),$$

where $\alpha = \{fake, sec\}$. Here, $f^{sim}_\alpha(h)$ is the fraction of background component α obtained from simulation, and R^h_α is a correction factor accounting for discrepancies between data and simulation.

Equivalently, by defining the corrected background fraction as

$$f^h_\alpha = R^h_\alpha f^{sim}_\alpha(h), \tag{10.1}$$

the purity can be expressed more simply as:

$$P^h = 1 - \sum_\alpha f^h_\alpha. \tag{10.2}$$

In this analysis, the systematic uncertainty on the purity arises from the uncertainties in the estimation of the background fractions f^h_α. The uncertainty in the fake-track fraction is discussed in section 10.1.1, while the uncertainty in the secondary-track fraction is detailed in section 10.1.2. Finally, the impact of these uncertainties on the purity itself is evaluated in section 10.1.3.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

10.1.1 Fake Tracks

As explained in section 9.2.1, the determination of the fake track fraction is obtained as:

$$f^h_{fake} = f^{sim}_{fake}(h) R^h_{fake}. \tag{9.6}$$

where $f^{sim}_{fake}(h)$ is the fraction of fake tracks obtained from simulation and R^h_{fake} is a data-to-simulation correction factor. Uncertainties on f^h_{fake} arise from both terms in eq. (9.6):

- $f^{sim}_{fake}(h)$ is affected only by statistical uncertainty, due to the limited size of the simulation samples ($\sigma_{stat, f^{sim}_{fake}(h)}$).
- R^h_{fake} has a statistical component ($\sigma_{stat, R^h_{fake}}$), but also suffers from systematic effects related to the proxy-sample method.

The systematic uncertainties on R^h_{fake} include the following contributions:

- **Purity of the proxy sample.** The proxy sample is not composed entirely of fake tracks; its purity is $\Pi^h_{fake} < 1$. Consequently, a fraction $1 - \Pi^h_{fake}$ of tracks in the proxy sample are not fakes, which contaminates the estimate of R^h_{fake}. Since the correction factor scales the simulated fraction to match data, the uncertainty due to impurity propagates proportionally to R^h_{fake}:

  $$\sigma^h_{purity} = (1 - \Pi^h_{fake})\, R^h_{fake}. \tag{10.3}$$

  The purity in data cannot be measured directly, but it can be estimated in simulation as :

  $$\Pi^{h,sim}_{fake} = \frac{N^{h,sim\text{-}TM}_{proxy_fake}}{N^{h,sim}_{proxy_fake}}. \tag{9.8}$$

  where $N^{h,sim}_{proxy_fake}$ is the number of tracks in the simulated proxy sample and $N^{h,sim\text{-}TM}_{proxy_fake}$ is the subset matched to fake tracks. Since the purity in data is expected to be higher than in simulation, the following inequality holds:

  $$(1 - \Pi^h_{fake}) R^h_{fake} < (1 - \Pi^{h,sim}_{fake}) R^h_{fake}. \tag{10.4}$$

  A conservative estimate of the uncertainty is therefore:

  $$\sigma^h_{purity} = (1 - \Pi^{h,sim}_{fake})\, R^h_{fake}. \tag{10.5}$$

- **Choice of the proxy sample.** The selection of the proxy sample also introduces uncertainties. To estimate the sensitivity of R^h_{fake} to the proxy sample selection, alternative definitions of the proxy sample are explored, as detailed in table 10.1. These alternative proxy samples provide an independent assessment of the fake track fraction. The purity of the fake tracks sample with the alternative selections is

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

presented in fig. 10.1a, fig. 10.1b and fig. 10.1c for the pPb Pbp and pp configurations, respectively. The corresponding R^h_{fake} of the alternative proxy samples is presented in fig. 10.2a, fig. 10.2b and fig. 10.2c. The uncertainty arising from the proxy sample selection, $\sigma^h_{proxy_sys}$, is defined as the standard deviation between the nominal and the alternative values of R^h_{fake}.

- **Limited size of the proxy and simulation samples.** The limited size of the proxy data samples induces a statistical uncertainty in the R^h_{fake} factors: $\sigma_{stat,R^h_{fake}}$, whereas the limited size of the simulation samples induces a statistical uncertainty in the f^{sim}_{fake} fractions: $\sigma_{stat,f^{sim}_{fake}(h)}$.

Total R^h_{fake} uncertainty. The overall uncertainty on R^h_{fake}, $\sigma(R^h_{fake})$, is calculated by adding in quadrature the uncertainties from the purity and the proxy sample selection with its statistical uncertainty:

$$\sigma(R^h_{fake}) = \sqrt{\left(\sigma^h_{purity}\right)^2 + \left(\sigma^h_{proxy_sys}\right)^2 + \left(\sigma_{stat,R^h_{fake}}\right)^2}. \tag{10.6}$$

Finally, the total uncertainty on f^h_{fake} is obtained by combining the uncertainties on R^h_{fake} and $f^{sim}_{fake}(h)$ in quadrature:

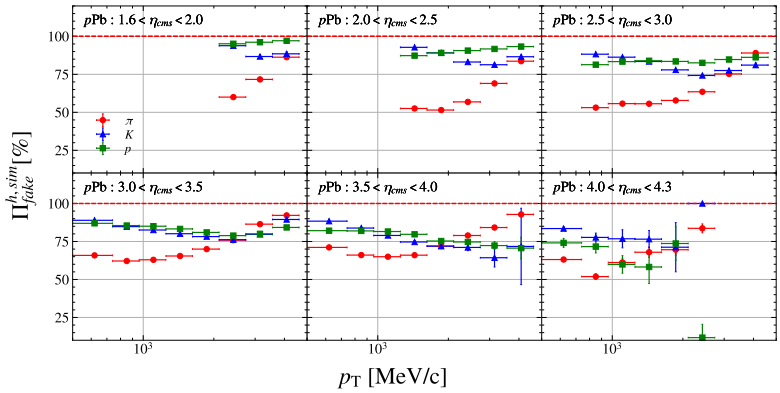
$$\sigma(f^h_{fake}) = \sqrt{\left(\sigma(R^h_{fake}) \cdot f^{sim}_{fake}(h)\right)^2 + \left(\sigma_{stat,f^{sim}_{fake}(h)} \cdot R^h_{fake}\right)^2}. \tag{10.7}$$

The obtained results for the total uncertainty on f^h_{fake} are shown in fig. 10.3.

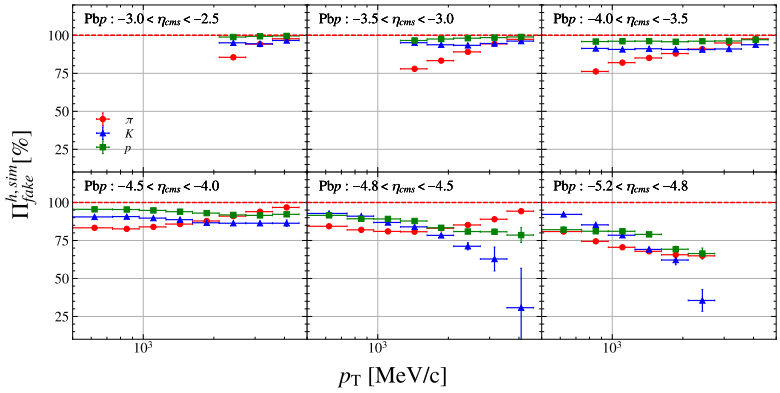
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

	pp	pPb	Pbp
Track type	Long	Long	Long
η	$2 < \eta < 4.8$	$2 < \eta < 4.8$	$2 < \eta < 4.8$
p_{T}	$0.5 < p_{\mathrm{T}} < 8\,\mathrm{GeV}/c$	$0.5 < p_{\mathrm{T}} < 8\,\mathrm{GeV}/c$	$0.5 < p_{\mathrm{T}} < 8\,\mathrm{GeV}/c$
p	$p > 10\,\mathrm{GeV}/c$	$p > 10\,\mathrm{GeV}/c$	$p > 10\,\mathrm{GeV}/c$
`pseudoIP`	$< 0.368\,\mathrm{mm}$	$< 0.348\,\mathrm{mm}$	$< 0.348\,\mathrm{mm}$
`GhostP`	$0.3 <$ `GhostP` < 0.4	$0.3 <$ `GhostP` < 0.4	$0.3 <$ `GhostP` < 0.4
	PID selection		
pions	$\mathrm{DLL}_{K\pi} <$ -2 & $\mathrm{DLL}_{p\pi} <$-2 & hasRich >0.5& `isMuon`==0	$\mathrm{DLL}_{K\pi} < 0$ & $\mathrm{DLL}_{p\pi} <$10 & hasRich >0.5 & `isMuon`==0	$\mathrm{DLL}_{K\pi} < 0$ & $\mathrm{DLL}_{p\pi} <$10 & hasRich >0.5 & `isMuon`==0
Kaons	$\mathrm{DLL}_{K\pi} > 5$ & ($\mathrm{DLL}_{K\pi}$-$\mathrm{DLL}_{p\pi}$) >5 & hasRich >0.5 & `isMuon`==0	$\mathrm{DLL}_{K\pi} > 5$ & ($\mathrm{DLL}_{K\pi}$-$\mathrm{DLL}_{p\pi}$) >0 & hasRich >0.5& `isMuon`==0	$\mathrm{DLL}_{K\pi} > 5$ & ($\mathrm{DLL}_{K\pi}$-$\mathrm{DLL}_{p\pi}$) >0 & hasRich >0.5& `isMuon`==0
protons	$\mathrm{DLL}_{p\pi} > 5$ & ($\mathrm{DLL}_{p\pi}$-$\mathrm{DLL}_{K\pi}$) > 10 & hasRich >0.5& `isMuon`==0	$\mathrm{DLL}_{p\pi} > 5$ & ($\mathrm{DLL}_{p\pi}$-$\mathrm{DLL}_{K\pi}$) > 15& hasRich >0.5& `isMuon`==0	$\mathrm{DLL}_{p\pi} > 5$ & ($\mathrm{DLL}_{p\pi}$-$\mathrm{DLL}_{K\pi}$) > 15& hasRich >0.5& `isMuon`==0

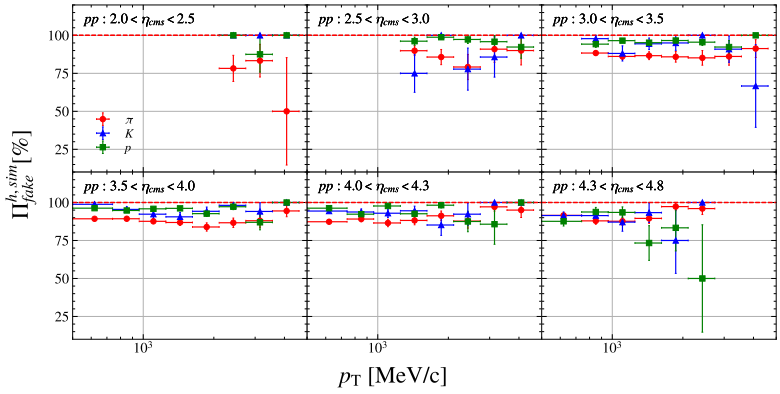
Table 10.1: Selection criteria for the alternative proxy samples used to assess systematic uncertainties in the determination of the fake-track correction factors R^h_{fake} in each collision configuration.



(a) pPb sample.



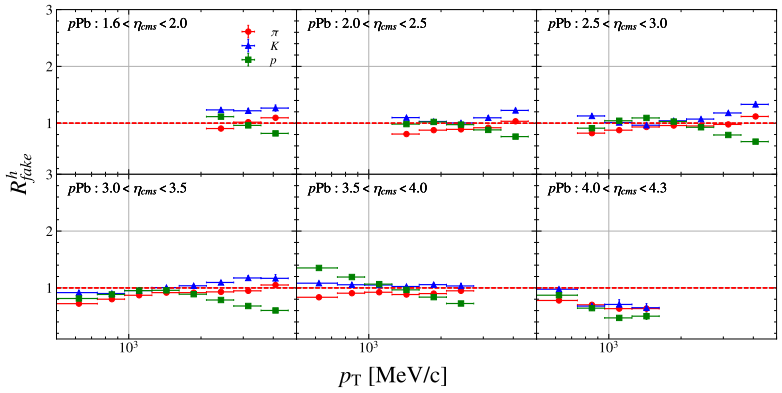
(b) Pbp sample.



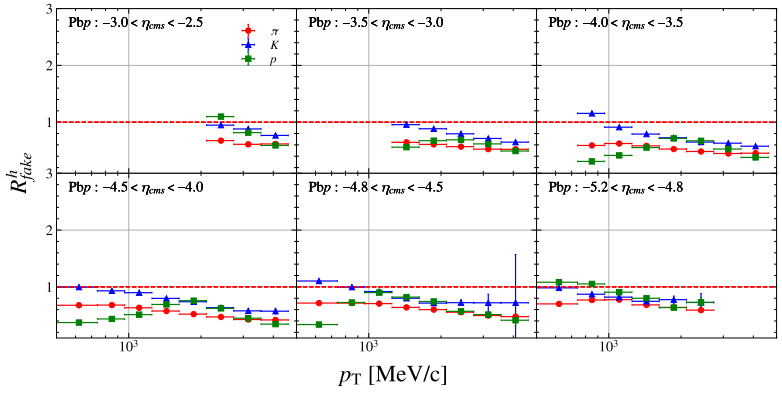
(c) pp sample.

Figure 10.1: Purity of the alternative proxy samples used to assess systematic uncertainties in the determination of R^h_{fake} in pPb (top), Pbp (middle), and pp (bottom) collisions. Fractions are shown for each hadron type: pions (red), kaons (blue), and protons (green).

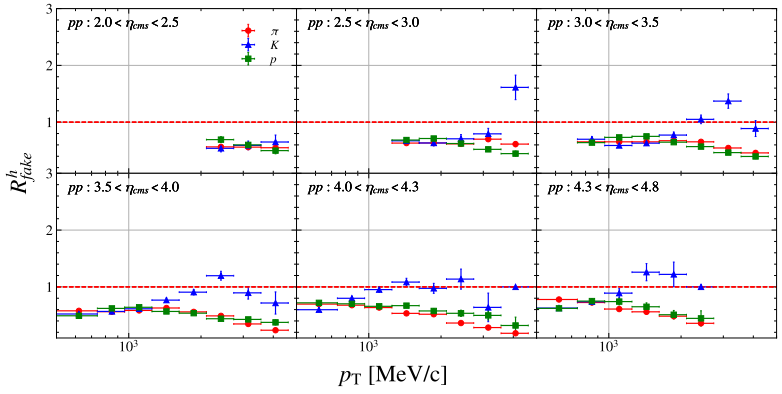
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) pPb sample.



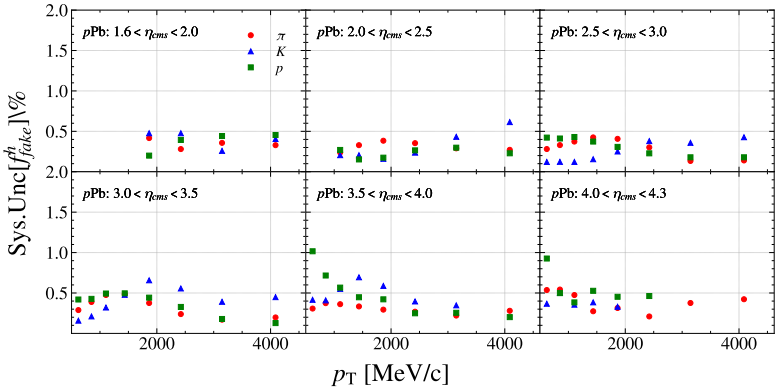
(b) Pbp sample.



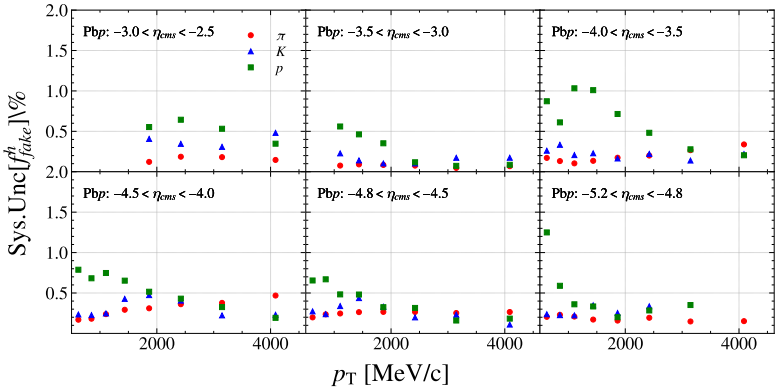
(c) pp sample.

Figure 10.2: Data-to-simulation correction factors R^h for the alternative proxy sample of fake tracks in pPb (top), Pbp (middle), and pp (bottom) collisions. Fractions are shown for each hadron type: pions (red), kaons (blue), and protons (green). These factors quantify the discrepancy between simulation and data for fake-track rates.

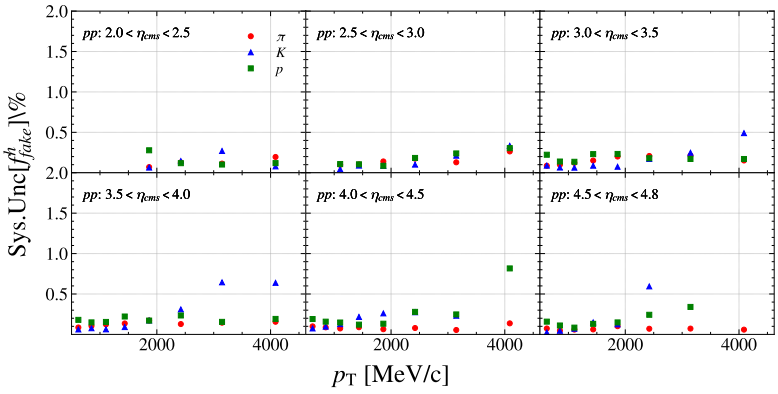
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) pPb sample.



(b) Pbp sample.



(c) pp sample.

Figure 10.3: Total systematic uncertainty of f^h_{fake} in pPb (top), Pbp (middle), and pp (bottom) collisions. The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

10.1.2 Secondary Tracks

As explained in section 9.2.2, the determination of the secondary track fraction is factorised as:

$$f^h_{sec} = f^{sim}_{\text{had_decay}}(h) R^h_{\text{had_decay}} + f^{sim}_{\gamma\to e}(h) R^h_{\gamma\to e} + f^{sim}_{X\to\mu}(h). \qquad \text{(eq. (9.10))}$$

This fraction is subject to the following effects that originate an uncertainty in its determination:

- **Purity of the proxy sample.** The R^h_j factors, with $j = \{\text{had_decay}, \gamma \to e\}$, account for discrepancies between simulation and data and are computed with proxy samples enriched in the corresponding secondaries category. However, these samples are not 100% pure. Similarly to what was discussed in section 10.1.1 for fake tracks, an uncertainty $\sigma^h_{purity,j}$ for the R^h_j factors is derived from the purity of the proxy sample:
$$\sigma^h_{purity,j} = (1 - \Pi^{h,sim}_j) R^h_j, \qquad (10.8)$$
where the purity of each of the secondary track proxy samples in simulation is:
$$\Pi^{h,sim}_j = \frac{N^{h,sim\text{-}TM}_{proxy,j}}{N^{h,sim}_{proxy,j}}. \qquad \text{(eq. (9.12))}$$

- **Limited size of the proxy and simulation samples.** For $j = \{\text{had_decay}, \gamma \to e\}$ the limited size of the proxy data samples induces a statistical uncertainty in the R^h_j factors: σ_{stat,R^h_j}, whereas the limited size of the simulation samples induces a statistical uncertainty in the f^{sim}_j fractions: $\sigma_{stat,f^{sim}_j(h)}$.

- **No correction of the data-simulation discrepancies for the decays into muons.** As discussed in section 9.2.2, for the decays into muons, the lack of statistics prevents from defining a proxy sample to calculate a correction factor accounting for data-simulation discrepancies. Therefore, the fraction for this secondaries contribution is assigned a large 50% total uncertainty: $\sigma(f^h_{X\to\mu}) = 0.5 \cdot f^h_{X\to\mu}$.

Contrary to what was done in the fake tracks contribution, no alternative definitions for the proxy data samples were considered. The proxy data samples of the secondaries background are smaller than the ones that are defined for the fake track contribution. The consequent larger statistical uncertainty is considered to cover the absence of alternative proxy data samples.

Combining the different uncertainty sources to estimate the uncertainty on f^h_{sec}: The overall uncertainty on R^h_{sec}, $\sigma(R^h_j)$, is calculated by combining the uncertainties from the purity of the proxy sample and the statistical uncertainty in quadrature:

$$\sigma(R^h_j) = \sqrt{\left(\sigma^h_{purity,j}\right)^2 + \left(\sigma_{stat,R^h_j}\right)^2}. \qquad (10.9)$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Subsequently, for $j = \{\text{had_decay}, \gamma \to e\}$, using error propagation, the uncertainty on the secondary track fraction ($f_j^h = f_j^{sim}(h) R_j^h$) is:

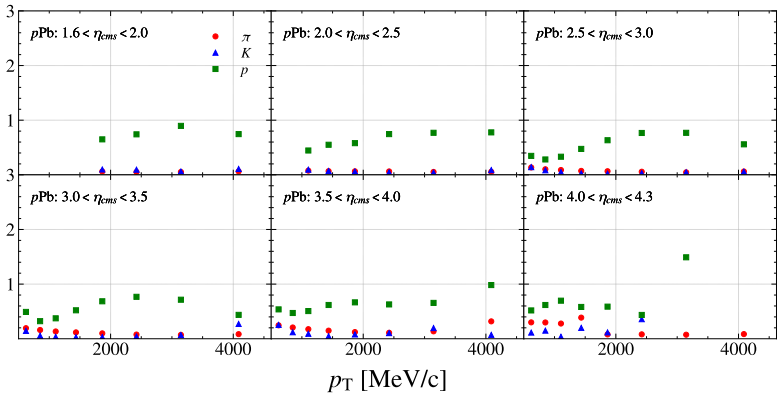
$$\sigma(f_j^h) = \sqrt{\left(\sigma(R_j^h) \cdot f_j^{sim}(h)\right)^2 + \left(\sigma_{stat, f_j^{sim}(h)} \cdot R_j^h\right)^2}. \tag{10.10}$$

Finally, the total uncertainty on f_{sec}^h is simply computed as:

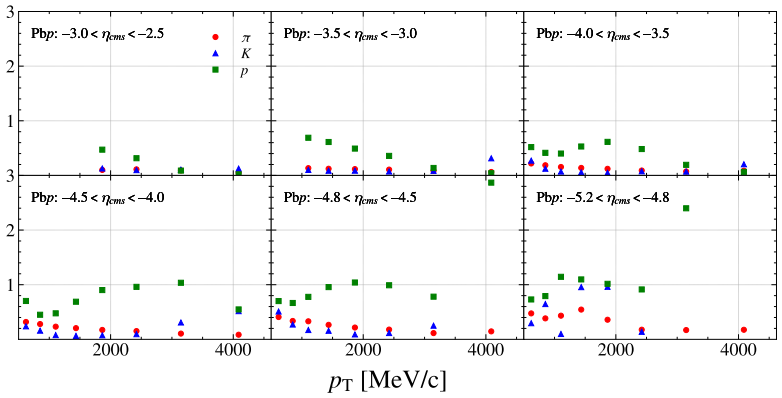
$$\sigma(f_{sec}^h) = \sqrt{\left(\sigma(f_{\text{had_decay}}^h)\right)^2 + \left(\sigma(f_{\gamma \to e}^h)\right)^2 + \left(\sigma(f_{X \to \mu}^h)\right)^2}. \tag{10.11}$$

The results for $\sigma(f_{sec}^h)$ are shown in fig. 10.4.

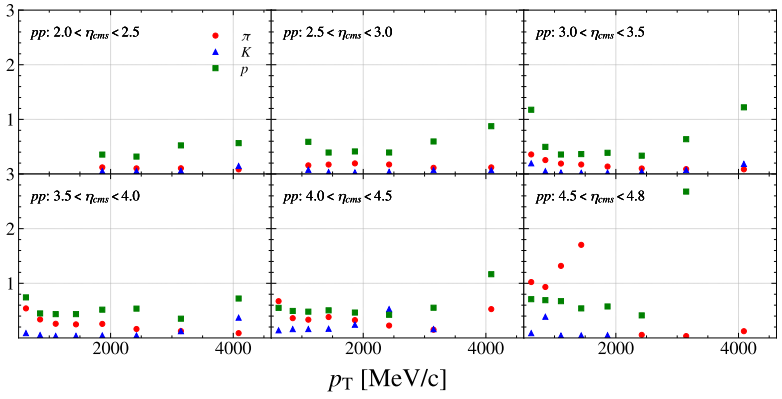
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) pPb sample.



(b) Pbp sample.



(c) pp sample.

Figure 10.4: Total systematic uncertainty of f^{h}_{sec} in pPb (top), Pbp (middle), and pp (bottom) collisions. The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

10.1.3 Purity

With the results of section 10.1.1 and section 10.1.2, the uncertainty on the purity, $\sigma(P^h)$, is calculated using error propagation in eq. (10.2). Assuming the uncertainties on f_α^h are uncorrelated, the uncertainty of P^h is given by:

$$\sigma(P^h) = \sqrt{\sum \left(\sigma(f_\alpha^h)\right)^2}. \tag{10.12}$$

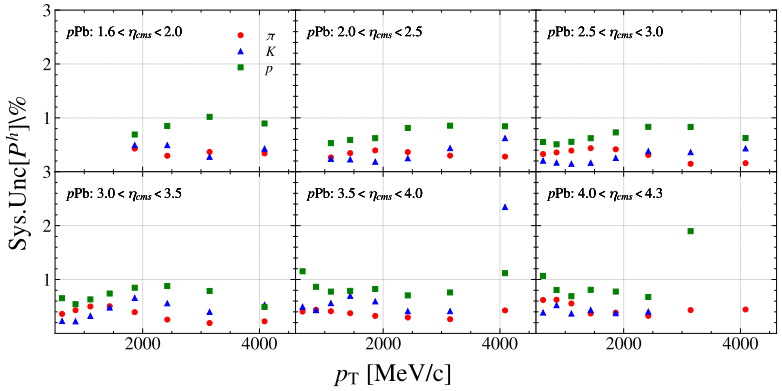
The results for the total uncertainty on P^h are shown in fig. 10.5. This uncertainty is propagated as a systematic uncertainty on the number of selected hadrons of each type, N_{Sel}^h, which was defined as:

$$N_{Sel}^h = N_{cand}^h \cdot P^h. \tag{eq. (9.1)}$$

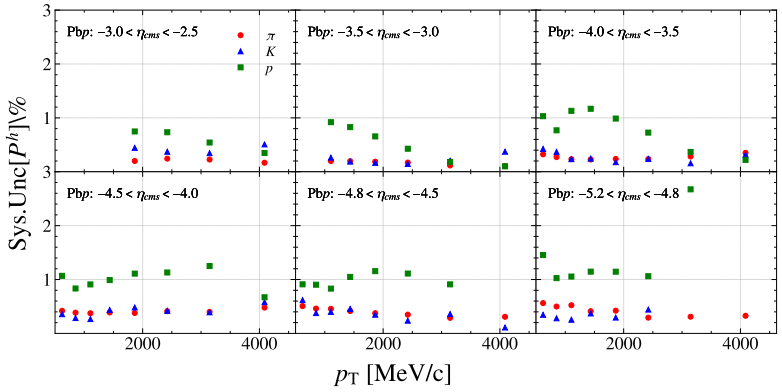
The error propagation is trivial and gives:

$$\sigma(N_{Sel}^h) = N_{cand}^h \cdot \sigma(P^h). \tag{10.13}$$

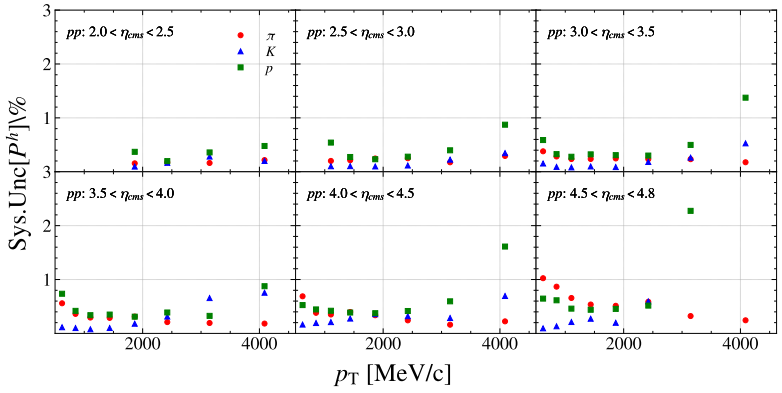
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) pPb sample.



(b) Pbp sample.



(c) pp sample.

Figure 10.5: Total systematic uncertainty in the purity (P^h) determination for pPb (top), Pbp (middle), and pp (bottom) collisions. The results are represented for each type of light hadron: pions are shown in red, kaons in blue and protons in green.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

10.2 Separating Prompt Contributions

A first classification of candidates is made with PID cuts which, after the purity correction, results in the per species number of *Selected* hadrons. The estimation of $True$ yields from *Selected* candidates is given by eq. (9.3), which relies in the determination method of PID efficiencies and misidentification factors, described in section 9.3. In this section the uncertainties in the determination of the PID matrix elements, are addressed. First, the fit model of the mass spectra of calibration samples is discussed in section 10.2.1 followed by the PID selection of tag particles in the considered decays in section 10.2.2. Then, the reweighting of the calibration samples to match the occupancy of analysed data is discussed in section 10.2.3 and, finally, the effect of the iterative method to correct for the total momentum dependency is addressed in section 10.2.4.

10.2.1 Fit Model

As explained in section 9.3.3, the PID matrix is obtained with fits to the invariant mass spectrum of resonances. Fit models were proposed as baseline but alternatives could produce different results assessing how sensitive the PID matrix elements are to the choice of the resonance and background line-shapes. The uncertainty in the PID matrix elements from the fit model, $\sigma_{\text{fit}}^{h\to h'}$, is estimated, considering an alternative resonance line-shape, as the difference of the alternative result with respect to the one obtained with the baseline model. The baseline and alternative models used for each calibration sample are listed in table 10.2.

10.2.2 PID Selection on the Tag Particle

The PID matrix uncertainty related to the PID selection of the tag particle in the calibration samples is assessed by proposing alternative selection criteria. Aside from the baseline selections, listed in table 9.8, two other sets of PID cuts, loose and tight, are applied to the tag particle, and the PID matrix is recalculated for both. The standard deviation of the results with the baseline and the two alternative selections is taken as a contribution to the PID-matrix uncertainty, denoted as $\sigma_{\text{tag}}^{h\to h'}$.

10.2.3 Reweighting of the Calibration Samples

The uncertainty in the PID matrix originated from the reweighting procedure to correct for discrepancies in occupancy variables in simulation with respect to data, explained in section 8.4, is evaluated by comparing the PID matrix obtained using three distinct weight sets. The standard deviation in the PID matrix elements across these weight sets quantifies this contribution to the PID matrix uncertainty, denoted as $\sigma_{\text{reweighting}}^{h\to h'}$.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Decay Mode	Signal Model	Signal Systematics Model	Background Model	Background Systematics Model
Λ	2 Gaussians	VG	first-degree Chebyshev polynomial	polynomial
K_s	2 Gaussians	VG	first-degree Chebyshev polynomial	polynomial
$\phi(1020)$	Voigtian	VG	Cheby1	Cheby1

Table 10.2: Models employed as baseline and alternative for the invariant mass spectrum of the calibration samples.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

10.2.4 Iterative Data-Driven Method

An additional contribution to the PID matrix uncertainty originates in the iterative method to correct the total momentum spectrum per species, see section 9.3.4. As discussed, the method is considered to converge after five iterations. The difference in the matrix elements between the fourth and fifth iterations of the reweighting process are taken as this contribution to the PID matrix uncertainty, denoted as $\sigma^{h\to h'}_{\text{iteration}}$.

10.2.5 Total Uncertainty of the PID Matrix

The total uncertainty on the PID matrix elements, $\sigma(\varepsilon^{h\to h'})$, is calculated as the quadratic sum of the previously discussed contributions. The statistical uncertainty, originated in the limited size of the calibration samples, is also accounted for, and is denoted as $\sigma^{h\to h'}_{\text{stat}}$. The result is then given by:

$$\sigma(\varepsilon^{h\to h'}) = \sqrt{\left(\sigma^{h\to h'}_{\text{fit}}\right)^2 + \left(\sigma^{h\to h'}_{\text{tag}}\right)^2 + \left(\sigma^{h\to h'}_{\text{reweighting}}\right)^2 + \left(\sigma^{h\to h'}_{\text{iteration}}\right)^2 + \left(\sigma^{h\to h'}_{\text{stat}}\right)^2} \tag{10.14}$$

10.2.6 Propagating the PID Matrix Uncertainty and the *Selected* Yields Uncertainty to the *True* Yields

As discussed in section 10.1, the yields without inter-species contamination, N^h_{True}, are estimated with the PID matrix:

$$\begin{bmatrix} N^{\pi}_{True} \\ N^{K}_{True} \\ N^{p}_{True} \end{bmatrix} = \begin{bmatrix} \varepsilon^{\pi\to\pi} & \varepsilon^{K\to\pi} & \varepsilon^{p\to\pi} \\ \varepsilon^{\pi\to K} & \varepsilon^{K\to K} & \varepsilon^{p\to K} \\ \varepsilon^{\pi\to p} & \varepsilon^{K\to p} & \varepsilon^{p\to p} \end{bmatrix}^{-1} \begin{bmatrix} N^{\pi}_{Sel} \\ N^{K}_{Sel} \\ N^{p}_{Sel} \end{bmatrix}. \qquad \text{(eq. (9.3))}$$

The systematic uncertainty on N^h_{True} is estimated by considering two contributions: the uncertainty on the selected yield N^h_{Sel}, denoted $\sigma_{\text{ToyMC-Nsel}}$, and the uncertainty from the PID efficiency matrix, denoted $\sigma_{\text{ToyMC-PID}}$.

The contribution $\sigma_{\text{ToyMC-Nsel}}$ originates from the uncertainty on the purity, as discussed in section 10.1.3. The contribution $\sigma_{\text{ToyMC-PID}}$ arises from uncertainties in the particle identification efficiency matrix, described in section 10.2. These uncertainties are propagated using a toy Monte Carlo approach:

- **PID Matrix Uncertainty Propagation:**
  Each matrix element is fluctuated according to a Gaussian distribution:

$$\varepsilon^{h\to h'}_{\text{toy}} \sim \mathcal{N}(\mu = \varepsilon^{h\to h'}, \sigma = \sigma(\varepsilon^{h\to h'})), \tag{10.15}$$

  where $\sigma(\varepsilon^{h\to h'})$ is the systematic uncertainty on the efficiency. For each of the 1000 toy experiments, a new fluctuated matrix is generated, inverted, and used to compute N^h_{True}. The standard deviation of the resulting values defines $\sigma^h_{\text{ToyMC-PID}}$.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

- **Selection Purity Uncertainty Propagation:**
  The selected yield N^h_{Sel} is varied within its uncertainty, also assumed to follow a Gaussian distribution. The resulting variation in N^h_{True} defines $\sigma^h_{\text{ToyMC-Nsel}}$.

The total systematic uncertainty on N^h_{True} is given by the quadrature sum:

$$\sigma(N^h_{True}) = \sqrt{(\sigma^h_{\text{ToyMC-PID}})^2 + (\sigma^h_{\text{ToyMC-Nsel}})^2}. \tag{10.16}$$

To illustrate the stability of the procedure, two representative bins in η_{cms} and p_T from the pPb dataset are shown in fig. 10.6 and fig. 10.7. The figures display the distributions of the N^h_{True} values for pions, kaons, and protons obtained with the toy procedure. The relative contributions of the different systematic sources to N^h_{True} are summarized in fig. 10.8, fig. 10.9 and fig. 10.10 for the pPb, Pbp, and pp samples, respectively.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

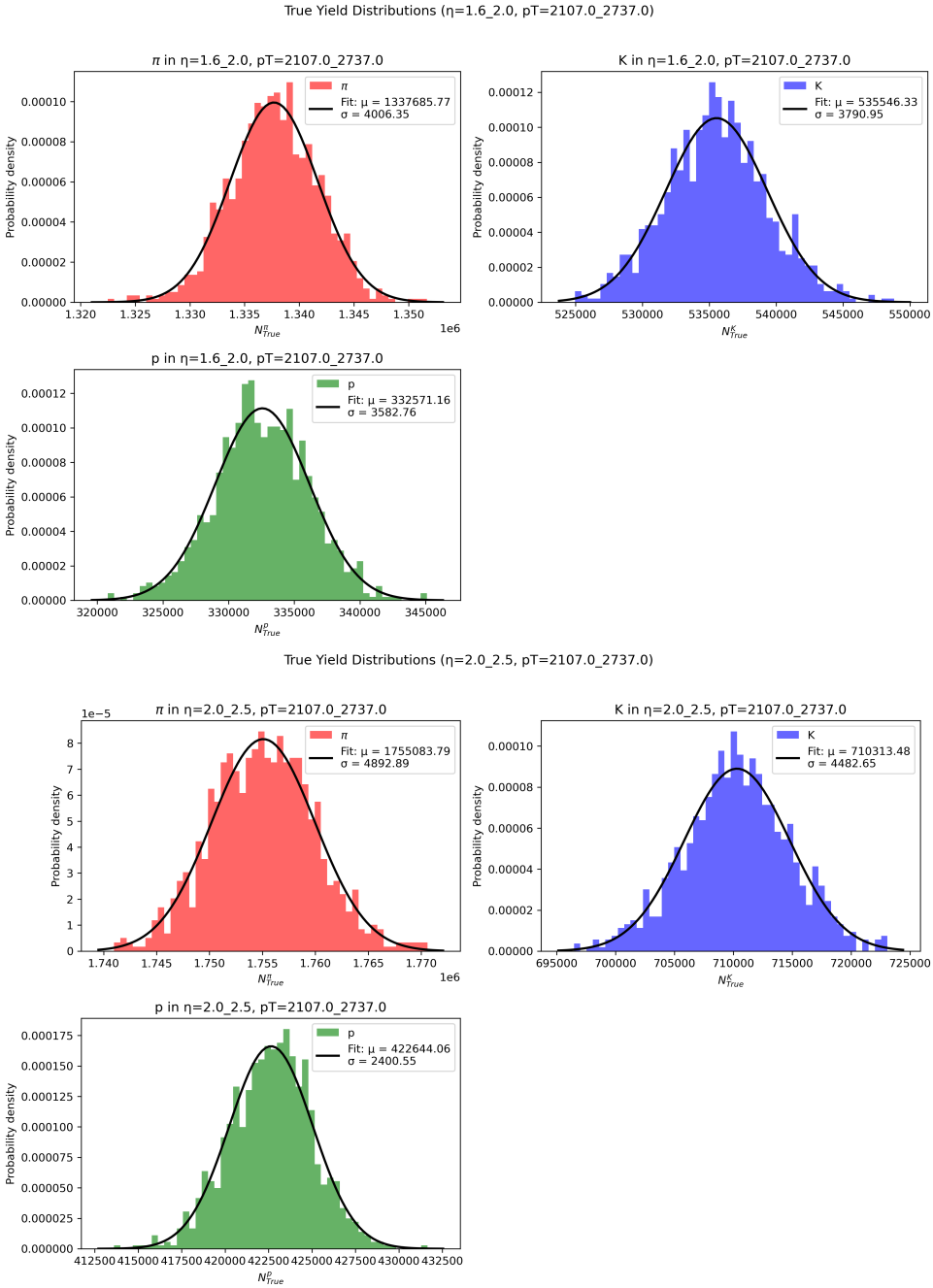


Figure 10.6: Distributions of the N_{True} values for pions, kaons, and protons obtained with the toy procedure in pPb collisions, for $p_{\mathrm{T}} \in [2107, 2737]$ MeV/c and $\eta_{\mathrm{cms}} \in [1.6, 2.0, 2.5]$.

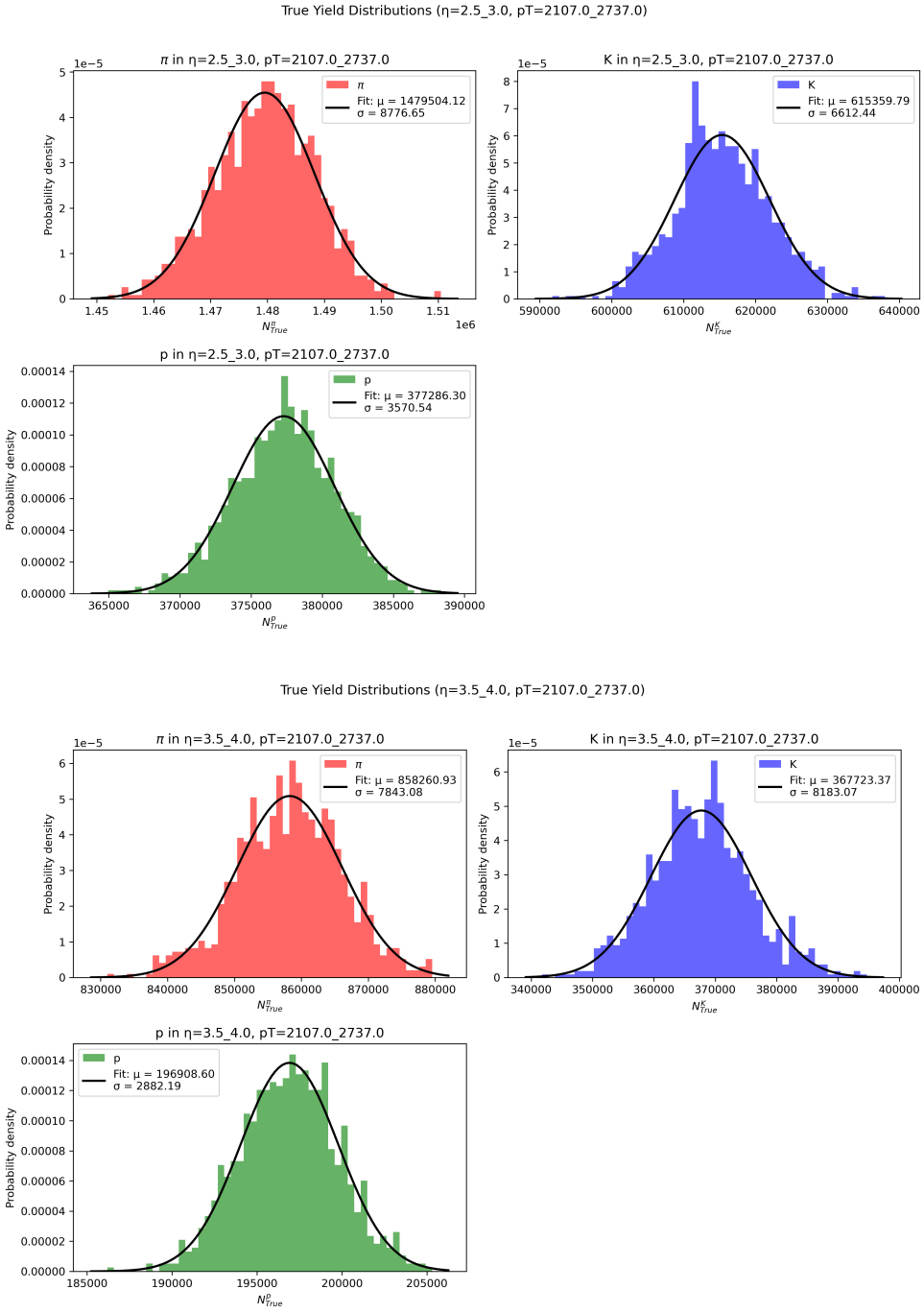
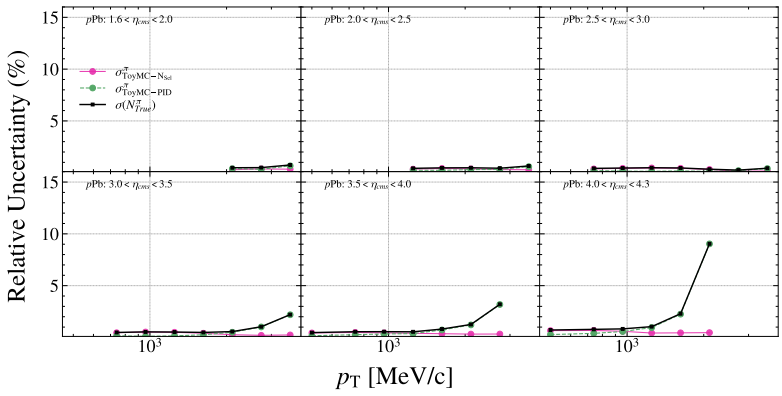
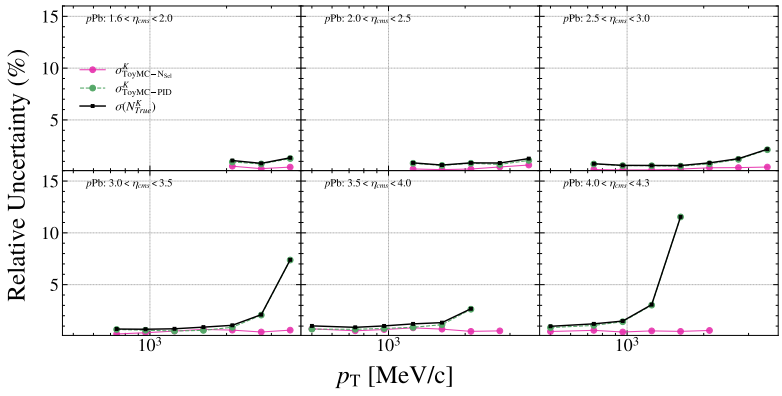


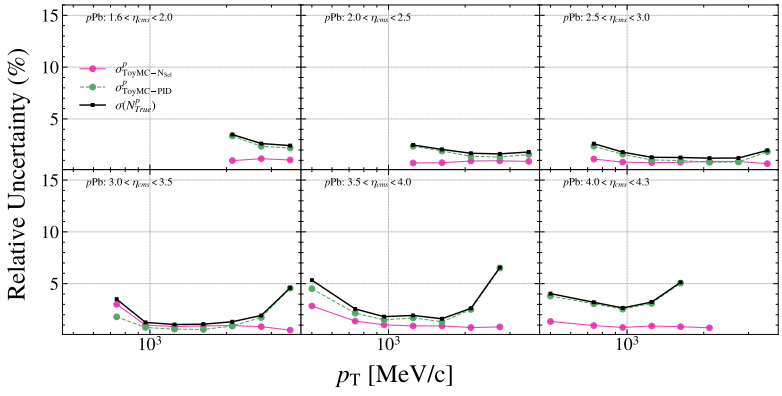
Figure 10.7: Distributions of the N_{True} values for pions, kaons, and protons obtained with the toy procedure in pPb collisions, for $p_{\mathrm{T}} \in [2107, 2737]$ MeV/c and $\eta_{\mathrm{cms}} \in [2.5, 3.0, 3.5]$.



(a) pions.



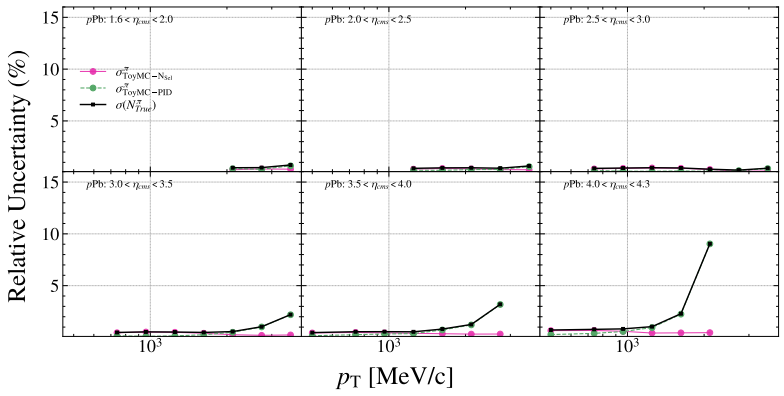
(b) kaons



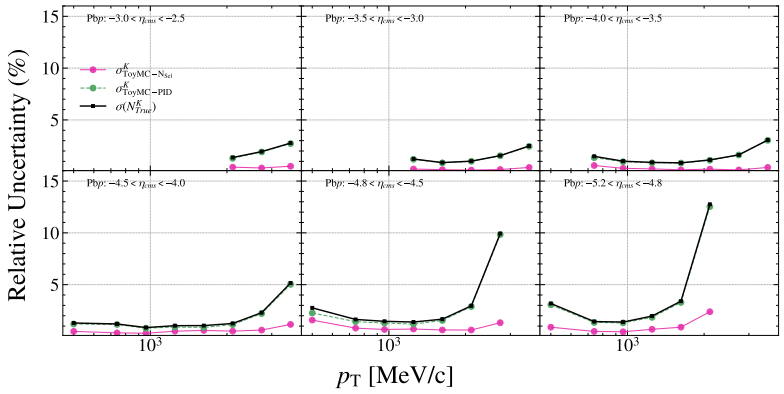
(c) protons.

Figure 10.8: Summary of the relative systematic uncertainties on N^h_{True} in pPb collisions for (a) pions, (b) kaons, and (c) protons. The contributions from the purity uncertainty ($\sigma^h_{\text{ToyMC-Nsel}}$, magenta), from the PID efficiency matrix uncertainty ($\sigma^h_{\text{ToyMC-PID}}$, green), and their quadrature sum ($\sigma(N^h_{True})$, black) are shown. The uncertainties are evaluated using the toy Monte Carlo procedure described in the text.

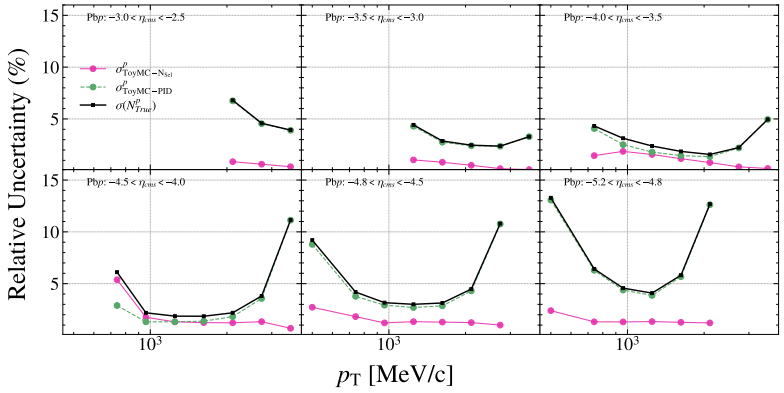
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) pions.



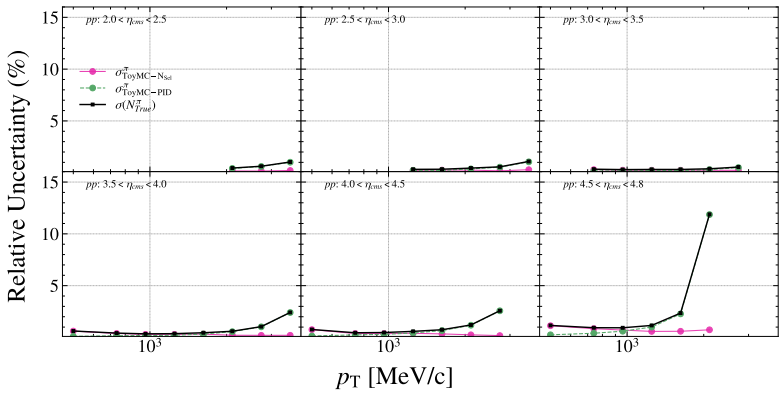
(b) kaons



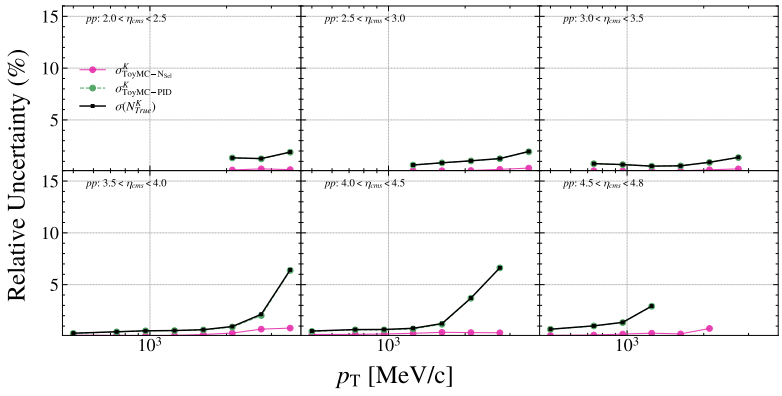
(c) protons.

Figure 10.9: Summary of the relative systematic uncertainties on N^h_{True} in Pbp collisions for (a) pions, (b) kaons, and (c) protons. The contributions from the purity uncertainty ($\sigma^h_{\text{ToyMC-Nsel}}$, magenta), from the PID efficiency matrix uncertainty ($\sigma^h_{\text{ToyMC-PID}}$, green), and their quadrature sum ($\sigma(N^h_{True})$, black) are shown. The uncertainties are evaluated using the toy Monte Carlo procedure described in the text.

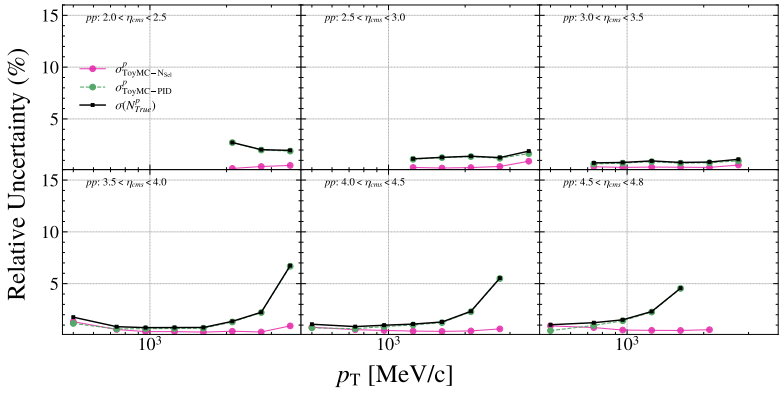
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) pions.



(b) kaons



(c) protons.

Figure 10.10: Summary of the relative systematic uncertainties on N^h_{True} in pp collisions for (a) pions, (b) kaons, and (c) protons. The contributions from the purity uncertainty ($\sigma^h_{\text{ToyMC-Nsel}}$, magenta), from the PID efficiency matrix uncertainty ($\sigma^h_{\text{ToyMC-PID}}$, green), and their quadrature sum ($\sigma(N^h_{True})$, black) are shown. The uncertainties are evaluated using the toy Monte Carlo procedure described in the text.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

10.3 Reconstruction Efficiency

The systematic uncertainty in the reconstruction efficiency, denoted as $\sigma(\varepsilon^h_{reco})$, arises from the determination of the $C_{tracking}$ tracking correction factor, introduced in section 9.4 to account for discrepancies between data and simulation in the tracking reconstruction efficiency, and from the reweighting occupancy correction discussed in Section 8.4. The contribution to $\sigma(\varepsilon^h_{reco})$ originated in uncertainties of $C_{tracking}$ is addressed in section 10.3.1 and the one originated in the reweighting procedure in section 10.3.2.

10.3.1 Tracking Correction Factor

The procedure to determine $C_{tracking}$ suffers from the following uncertainties:

Global systematic uncertainty. A per-track basal systematic uncertainty of 0.4% in 2012 pp data, which is the sample used to compute the reconstruction efficiency in pPb and Pbp events, and 0.8% in 2015 pp data is estimated for $C_{tracking}$ [120]. This uncertainty originates in three sources: the weight of simulation samples to reproduce data multiplicity, the discrepancies in the subdetector choice for the probe track and the signal model used to fit the J/ψ invariant mass. This uncertainty translates into a contribution to the systematic uncertainty of ε^h_{reco} as :

$$\sigma^h_{\text{global}} = \begin{cases} 0.004 \cdot \varepsilon^h_{reco} & (p\text{Pb/Pb}p), \\ 0.008 \cdot \varepsilon^h_{reco} & (pp). \end{cases} \tag{10.17}$$

Muon-based-corrections for hadrons. The $C_{tracking}$ factor is calculated with muon tracks. An uncertainty, arising from different effects in the detector material interactions of hadrons as compared to muons, should be considered when applying this factor to light hadrons. The uncertainty related to applying a muon-based correction to hadrons is estimated using simulated events and is taken as 1.4% for pions ($\sigma^\pi_{\text{mat}} = 0.014 \cdot \varepsilon^\pi_{reco}$) and 1.1% for kaons ($\sigma^K_{\text{mat}} = 0.014 \cdot \varepsilon^K_{reco}$), following [120]. However, for protons no value for this uncertainty was provided. This uncertainty can be estimated, following [113], from the difference in efficiency between protons and muons. Since protons are stable and muons interact very minimally with the detector material, any difference in efficiency is attributed to material interactions. The proton efficiency after material interactions is consequently factorised as:

$$\varepsilon^p_{reco} \approx \varepsilon^p_{mat} \cdot \varepsilon^\mu_{reco}, \tag{10.18}$$

where ε^μ_{reco} is the muon efficiency, assumed to be unaffected by material interactions. The lost proton fraction due to material interactions is:

$$f_{p,\text{mat}} = 1 - \varepsilon^p_{\text{mat}}, \tag{10.19}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

that can be alternatively expressed from eq. (10.18) as:

$$f_{p,\mathrm{mat}} = \frac{\varepsilon^{\mu}_{reco} - \varepsilon^{p}_{reco}}{\varepsilon^{\mu}_{reco}}. \tag{10.20}$$

The uncertainty in the loss fraction is proportional to the material budget μ, with a relative uncertainty of 10%:

$$\frac{\sigma(f_{p,\mathrm{mat}})}{f_{p,\mathrm{mat}}} = 0.1. \tag{10.21}$$

The uncertainty in the proton efficiency due to material interactions can be derived via error propagation as:

$$\sigma_{\mathrm{mat}}(\varepsilon^{p}_{reco}) = \frac{\partial \varepsilon^{p}_{reco}}{\partial f_{p,\mathrm{mat}}} \cdot \sigma(f_{p,\mathrm{mat}}). \tag{10.22}$$

The relation:

$$\frac{\partial \varepsilon^{p}_{reco}}{\partial f_{p,\mathrm{mat}}} = -\varepsilon^{\mu}_{reco}, \tag{10.23}$$

combined with eq. (10.20) and eq. (10.21), leads to:

$$\sigma^{p}_{\mathrm{mat}} = 0.1 \cdot \left(\varepsilon^{\mu}_{reco} - \varepsilon^{p}_{reco}\right). \tag{10.24}$$

The results of $\sigma^{p}_{\mathrm{mat}}$ obtained for pp data are shown in fig. 10.11.

Tracking bin uncertainty. The tracking correction factor is provided for each (p, η) bin with its corresponding uncertainty σ_{C_i}. The resulting contribution to the uncertainty on ε^{h}_{reco} due to tracking corrections is computed for the analysis (p_{T}, η) bins as:

$$\sigma^{h}_{\mathrm{bin}} = \sqrt{\frac{\sum_i \left(w_i \cdot \sigma_{C_i}\right)^2}{\left(\sum_i w_i\right)^2}}, \tag{10.25}$$

where i runs over the candidates within a given (p_{T}, η) bin.

10.3.2 Occupancy Correction

The reweighting procedure, discussed in Section 8.4, introduces a source of uncertainty in ε^{h}_{reco}. This uncertainty is quantified by calculating the weighted standard deviation of the reconstruction efficiencies for the sets of weights $\{1, 2, 7\}$:

$$\sigma^{h}_{\mathrm{occ}} = \sqrt{\frac{\sum_{i\in\{1,2,7\}} r_i \cdot \left(\varepsilon^{h}_{reco,i}(\eta_{\mathrm{cms}}, p_{\mathrm{T}}) - \bar{\varepsilon}^{h}_{reco}(\eta_{\mathrm{cms}}, p_{\mathrm{T}})\right)^2}{\sum_{i\in\{1,2,7\}} r_i}}. \tag{10.26}$$

The variable $\varepsilon^{h}_{reco,i}(\eta_{\mathrm{cms}}, p_{\mathrm{T}})$ denotes the reconstruction efficiency in the i-th set for given η_{cms} and p_{T}, and $\bar{\varepsilon}^{h}_{reco}(\eta_{\mathrm{cms}}, p_{\mathrm{T}})$ is the weighted mean of these efficiencies across the considered sets. The weight r_i is defined as $r_i = 1/(\sigma^{h}_{\mathrm{stat},i})^2$, where $\sigma^{h}_{\mathrm{stat},i}$ is the statistical uncertainty on $\varepsilon^{h}_{reco,i}$, originated in the finite size of the simulation sample.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

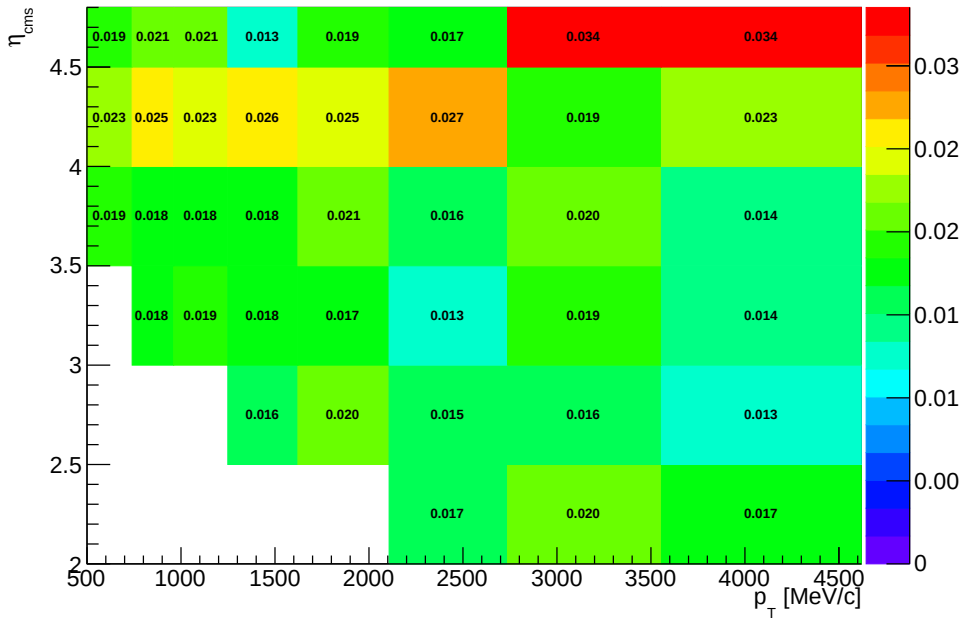


Figure 10.11: Uncertainty in proton reconstruction efficiency from material interactions, $\sigma_{\text{mat}}(\varepsilon^{p}_{reco})$, for pp collisions.

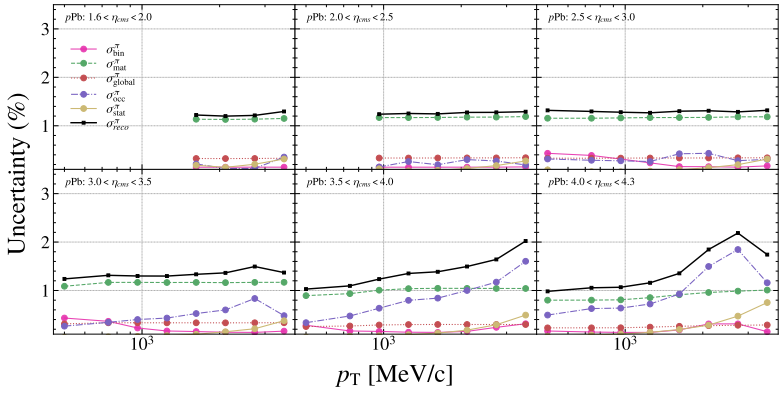
10.3.3 Total Reconstruction Efficiency Systematic Uncertainty

The total systematic uncertainty in the reconstruction efficiency is calculated as the sum in quadrature of the tracking and occupancy correction contributions:

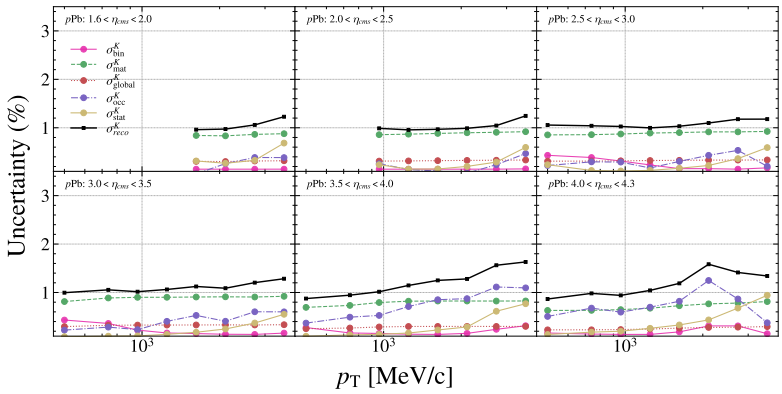
$$\sigma(\varepsilon^{h}_{reco}) = \sqrt{\left(\sigma^{h}_{\text{global}}\right)^2 + \left(\sigma^{h}_{\text{mat}}\right)^2 + \left(\sigma^{h}_{\text{bin}}\right)^2 + \left(\sigma^{h}_{\text{occ}}\right)^2 + \left(\sigma^{h}_{\text{stat}}\right)^2}. \tag{10.27}$$

The results of the systematic uncertainties on the reconstruction efficiency are shown in fig. 10.12, fig. 10.13, and fig. 10.14, for the pPb, Pbp, and pp samples, respectively.

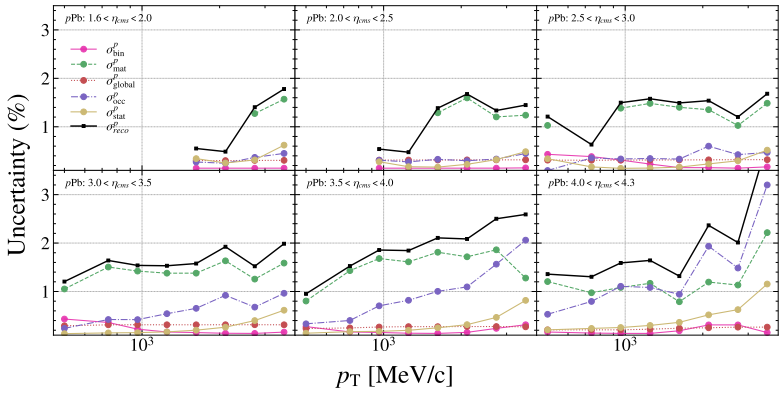
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) pions.



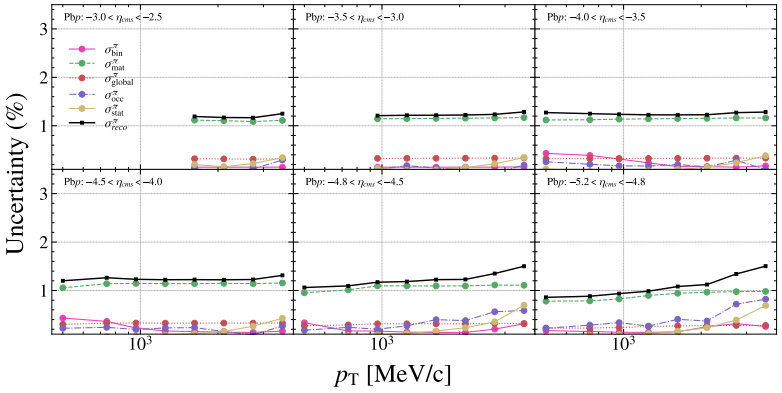
(b) kaons.



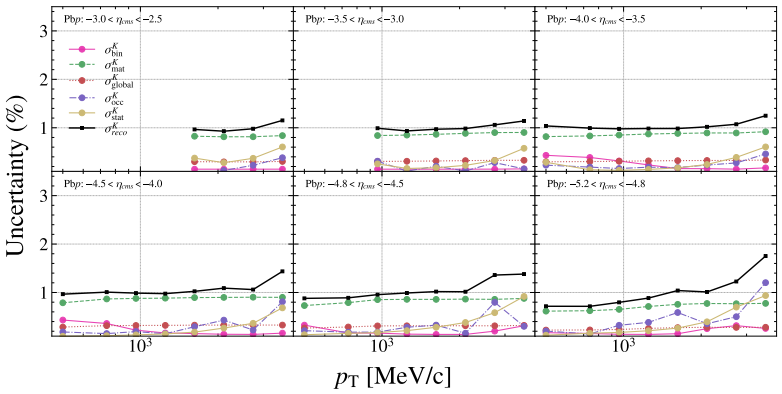
(c) protons

Figure 10.12: Systematic uncertainties on the reconstruction efficiency in the pPb sample for each particle species and source: σ^{h}_{bin} (tracking correction uncertainty), σ^{h}_{mat} (material differences), $\sigma^{h}_{\text{global}}$ (global effects), σ^{h}_{occ} (detector occupancy), and σ^{h}_{stat} (statistical uncertainty).

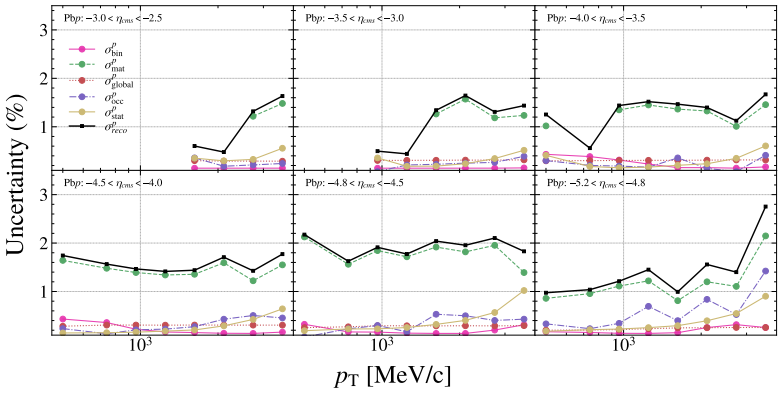
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) pions.



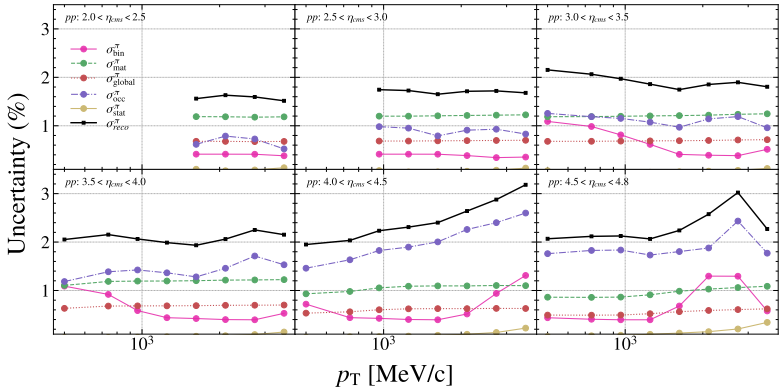
(b) kaons



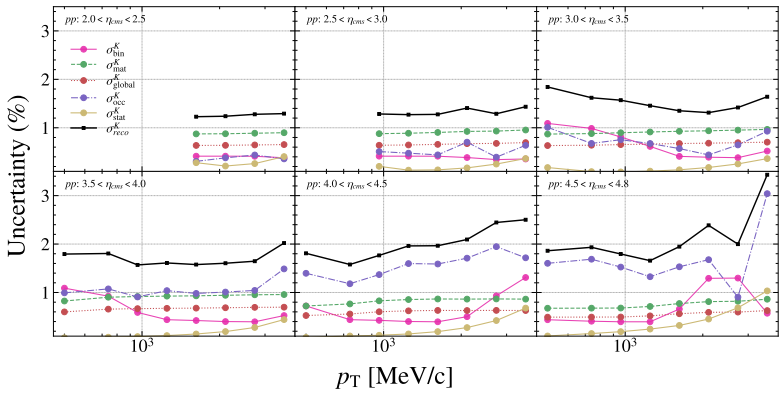
(c) protons.

Figure 10.13: Systematic uncertainties on the reconstruction efficiency in the Pbp sample for each particle species and source: σ^{h}_{bin} (tracking correction uncertainty), σ^{h}_{mat} (material differences), $\sigma^{h}_{\text{global}}$ (global effects), σ^{h}_{occ} (detector occupancy), and σ^{h}_{stat} (statistical uncertainty).

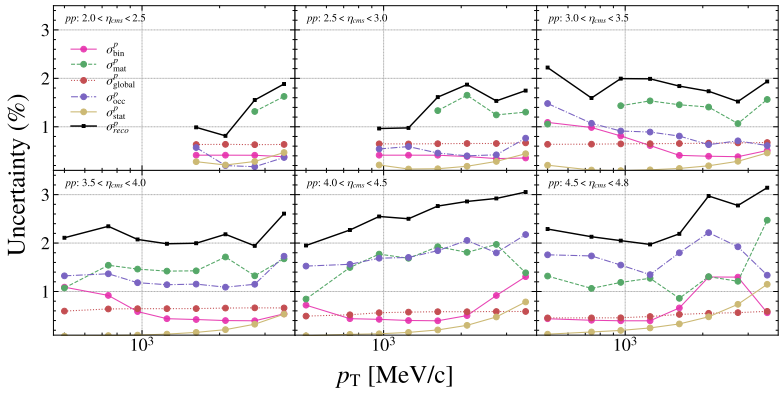
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) pions.



(b) kaons.



(c) protons.

Figure 10.14: Systematic uncertainties on the reconstruction efficiency in the *pp* sample for each particle species and source: σ^h_{bin} (tracking correction uncertainty), σ^h_{mat} (material differences), σ^h_{global} (global effects), σ^h_{occ} (detector occupancy), and σ^h_{stat} (statistical uncertainty).

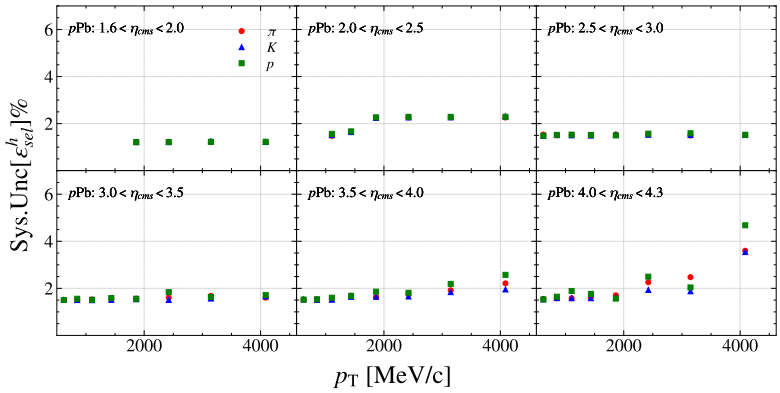
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

10.4 Selection Efficiency

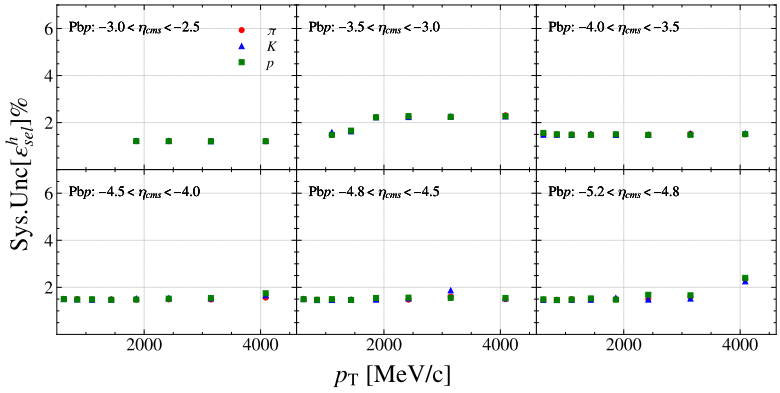
The selection efficiency was defined in section 9.5. The considered source of systematic uncertainty on the selection efficiency, $\sigma(\varepsilon_{sel})$, arises from the C_{sel} factor, defined in eq. (9.25) to account for the difference between data and simulated events.

The C_{sel} factor is obtained as an efficiency ratio from fitting the K^+K^- mass spectrum around the ϕ meson pole in data and simulation events. The systematic uncertainty is evaluated by performing the fits with a different PDF, replacing the background model of eq. (9.24) with a `Chebychev` polynomial. The alternative result for C_{sel} is then used to obtain a different ε_{sel}, and the uncertainty is estimated as the shift with respect to the baseline result. This difference corresponds to 0.8% of the value of C_{sel}, which is therefore taken as the systematic uncertainty estimate. The selection efficiency is then calculated by varying C_{sel} by $\pm\sigma$, and half the difference between the two results is taken as the systematic uncertainty. The $\sigma(\varepsilon_{sel})$ systematic uncertainties are shown for pPb, Pbp, and pp events in fig. 10.15

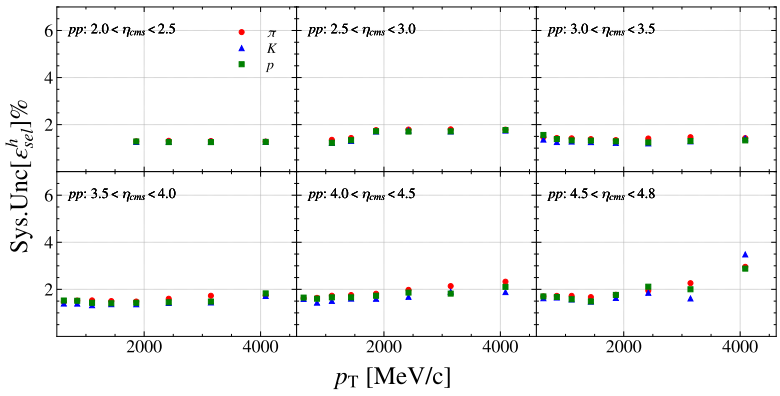
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) pPb.



(b) Pbp



(c) pp

Figure 10.15: Total systematic uncertainty of the selection efficiency in pPb (top), Pbp (middle), pp (bottom) events. The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

11

Results and Discussion

11.1 Cross Section and Nuclear Modification Factor

Following the methodology established in chapter 9, the double-differential cross-sections of prompt light hadrons are obtained as:

$$\frac{\mathrm{d}^2\sigma^h_{pp,p\mathrm{Pb}}}{\mathrm{d}p_\mathrm{T}\mathrm{d}\eta_\mathrm{cms}} = \frac{1}{\mathcal{L}}\frac{N^h(\eta_\mathrm{cms}, p_\mathrm{T})}{\Delta\eta_\mathrm{cms}\Delta p_\mathrm{T}}, \qquad \text{(eq. (6.4))}$$

where N^h is the corrected yield of prompt pions, kaons, and protons. The results for pPb, Pbp and pp events are shown in fig. 11.1. The pions, shown in red, have the highest cross section, followed by kaons in blue, then protons in green. Statistical uncertainties are represented by vertical lines, while systematic uncertainties are displayed as shaded boxes.

The measured cross-sections are compared to theoretical predictions obtained from next-to-leading order (NLO) perturbative QCD calculations, as shown in fig. 11.2. The predictions are displayed as shaded bands: red for pions, blue for kaons, and green for protons. The width of each band represents the uncertainty envelope resulting from the variation of renormalization and factorization scales as well as the PDF error sets, with both contributions added in quadrature. These predictions were computed using the INCNLO framework [123], modified as described in [124], and use the CT18ANLO proton PDFs [125]. For nuclear collisions, nuclear effects are implemented through the EPPS21 parametrization [126]. Hadronization is modeled using the Kretzer fragmentation functions [127]. The theoretical results correspond to the sum of positively and negatively charged pions, kaons, and protons. The theoretical calculations employ the same momentum cut ($p > 10\,\mathrm{GeV}/c$) as the analysis.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

After obtaining the cross sections, the nuclear modification factor is calculated as:

$$\mathrm{R}^{h}_{p\mathrm{Pb}} = \frac{1}{208}\frac{d\sigma^{h}_{p\mathrm{Pb}}/dp_T d\eta_{\mathrm{cms}}}{d\sigma^{h}_{pp}/dp_T d\eta_{\mathrm{cms}}}. \qquad \text{(eq. (6.3))}$$

The luminosity systematic uncertainty, which is fully correlated between the pPb and pp data samples, cancels in the ratio. The results are shown in fig. 11.3a for the forward region and fig. 11.3b for the backward region.

The nuclear modification factor shows a clear pseudorapidity dependence, characterized by a strong suppression in the forward region and a Cronin-like enhancement in the backward region. In forward pseudorapidity, the suppression is observed to be nearly identical for pions, kaons, and protons, indicating that the energy loss mechanisms in this region affects all particle species uniformly, regardless of their mass. In contrast, at backward pseudorapidity, the enhancement is significantly stronger for protons compared to pions and kaons. This mass-dependent behavior in the backward region may arise from differences in hadronisation processes or the interplay between initial-state nuclear modifications and final-state interactions.

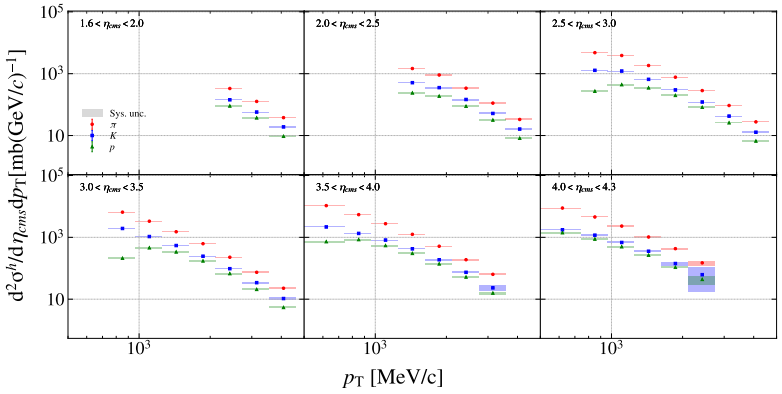
The results are first compared with previous LHCb measurements of prompt charged particles, as shown in section 11.1. In the forward region, the nuclear modification factors of pions, kaons, and protons are consistent with those of prompt charged particles. In the backward region, the proton nuclear modification factor shows an enhancement comparable to that observed for prompt charged particles at 5 TeV, whereas the pion and kaon nuclear modification factors do not display such a strong rise.

A separate comparison with the π^0 measurements is presented in fig. 11.5. In the forward region, the nuclear modification factors of the three hadron species are consistent with the corresponding π^0 results. In the backward region, the pion and kaon nuclear modification factors follow the trend observed for π^0 production at 8 TeV [7] in the overlapping kinematic range, while the proton nuclear modification factor clearly deviates from this behaviour, exhibiting an enhancement.

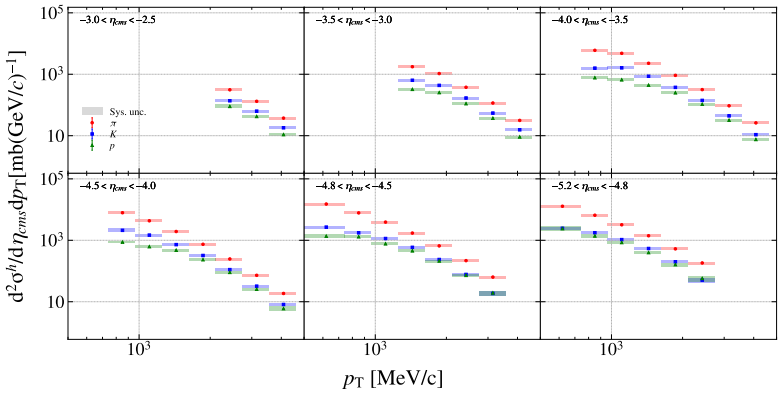
The comparison with EPPS21 theoretical predictions, shown in fig. 11.6, confirms that the forward-region suppression is well reproduced for all three hadron species. However, in the backward region, while the predictions capture the pion and kaon behaviour, they do not account for the enhancement observed for protons.

Finally, fig. 11.7 presents a comparison with ALICE results. Across all three hadron species, the nuclear modification factors measured here are consistent with those reported by ALICE.

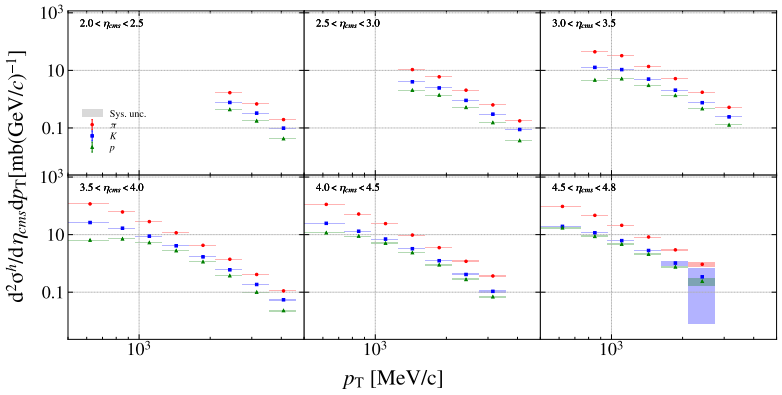
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) pPb sample



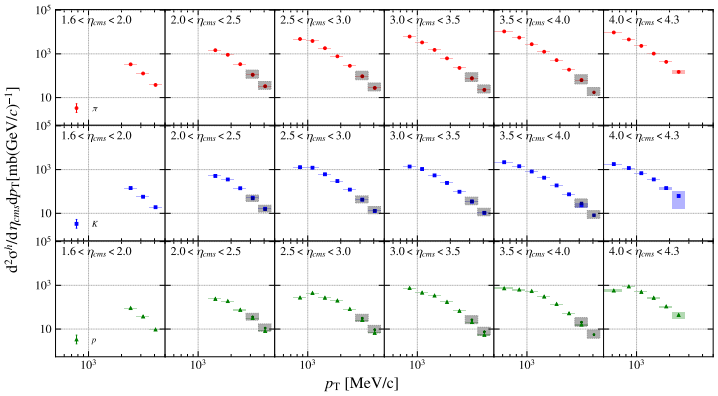
(b) Pbp sample



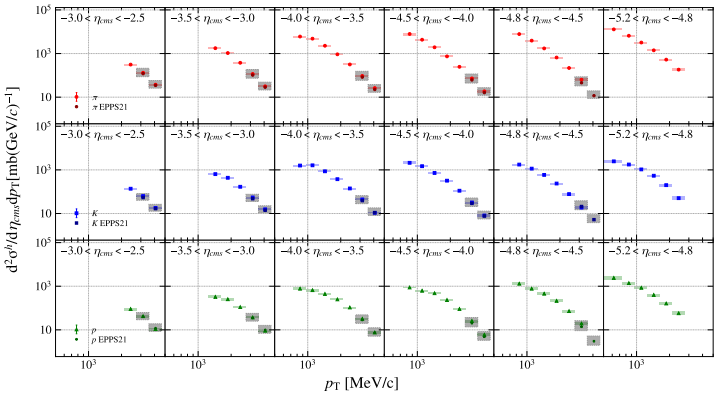
(c) pp sample

Figure 11.1: Production cross sections in pPb (top), Pbp (middle) and pp (bottom) events. The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green. Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes.

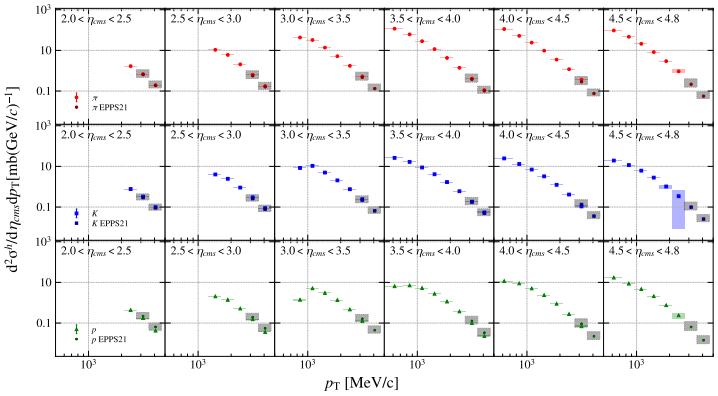
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) pPb sample.



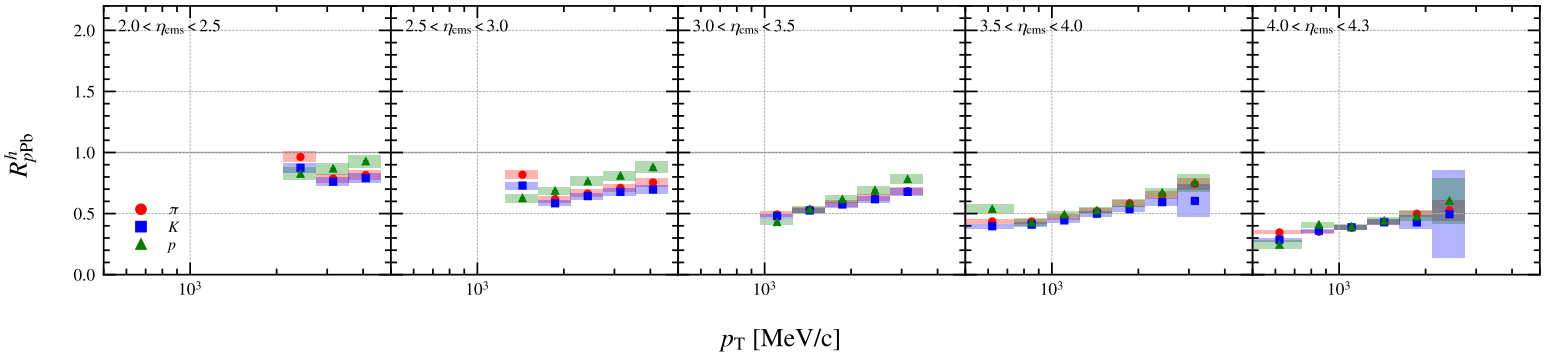
(b) Pbp sample.



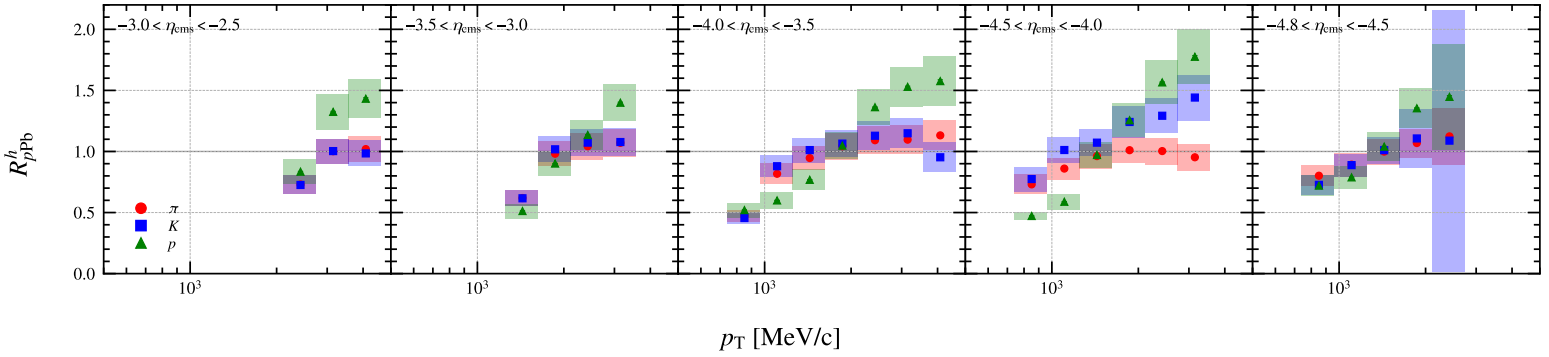
(c) pp sample.

Figure 11.2: Production cross sections in pPb (top), Pbp (middle) and pp (bottom) events compared to EPPS21 model predictions (dashed black lines). The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green. Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

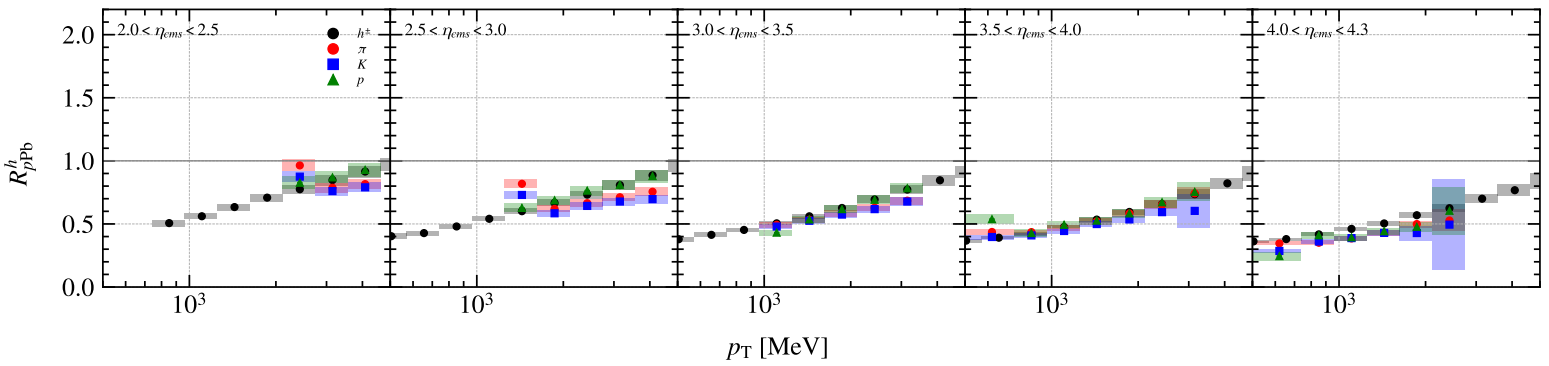


(a) Forward.

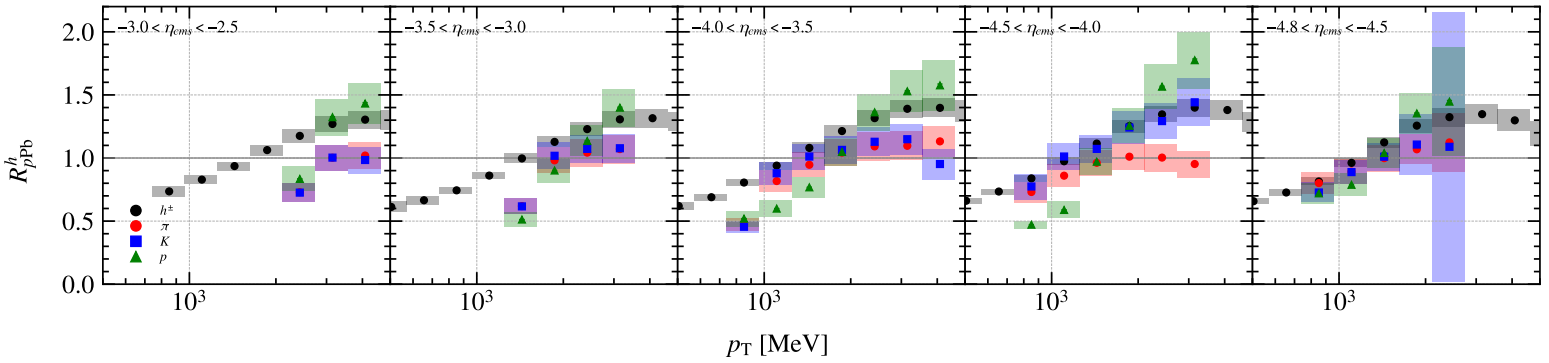


(b) Backward.

Figure 11.3: The Nuclear modification factor in the forward region(top) and the backward region (bottom). The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green. Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes.

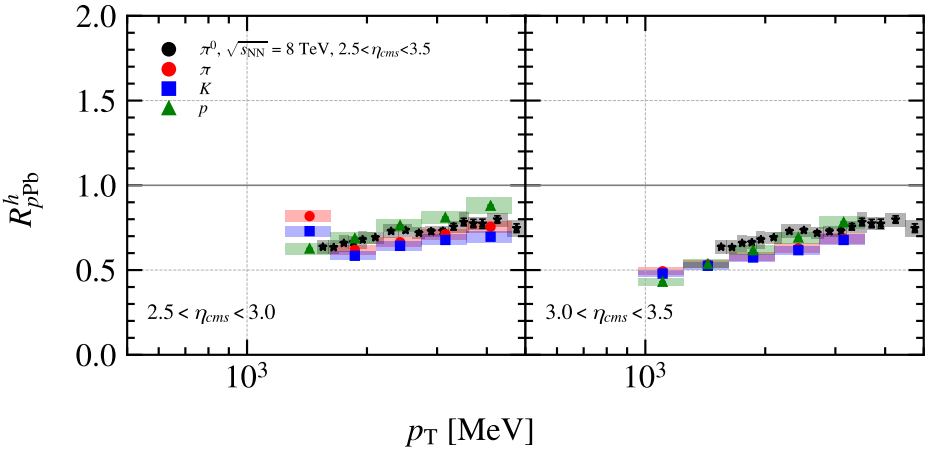


(a) Forward.

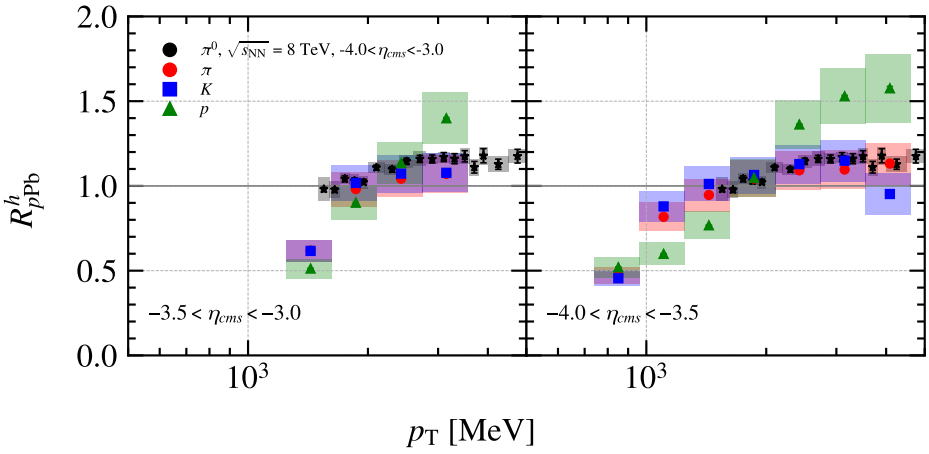


(b) Backward.

Figure 11.4: The Nuclear modification factor in the forward region(top) and the backward region (bottom) compared with prompt charged particles ($h^{\pm}$) [8]. The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green. Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes. The discrepancy observed at low p_{T} is due to the momentum cut applied in this analysis.



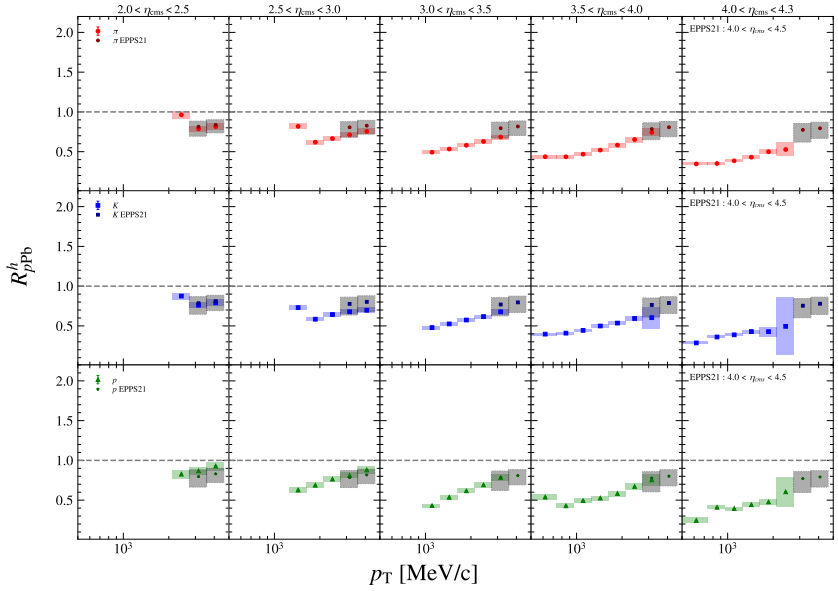
(a) forward.



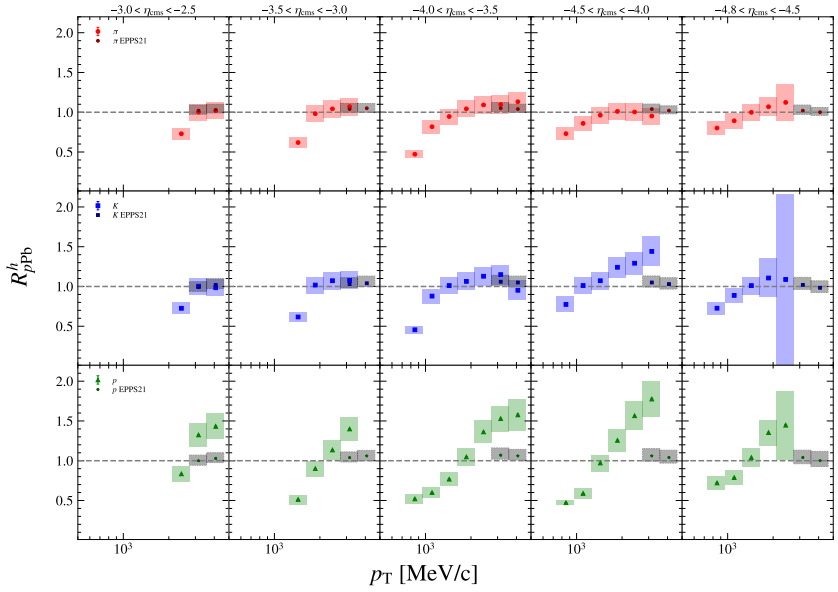
(b) backward.

Figure 11.5: Nuclear modification factors in the forward (top) and backward (bottom) regions. Results from this work are shown for pions (red), kaons (blue), and protons (green). Statistical uncertainties are indicated by vertical error bars, and systematic uncertainties by boxes. The result of this work is compared with previous LHCb measurement π^0 at 8 TeV [7] following the styles indicated in the legend. The discrepancy observed at low p_{T} is due to the momentum cut applied in this analysis.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Forward.



(b) Backward.

Figure 11.6: The nuclear modification factor in the forward (top) and the backward region (bottom) compared to EPPS21 model predictions. The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green. Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

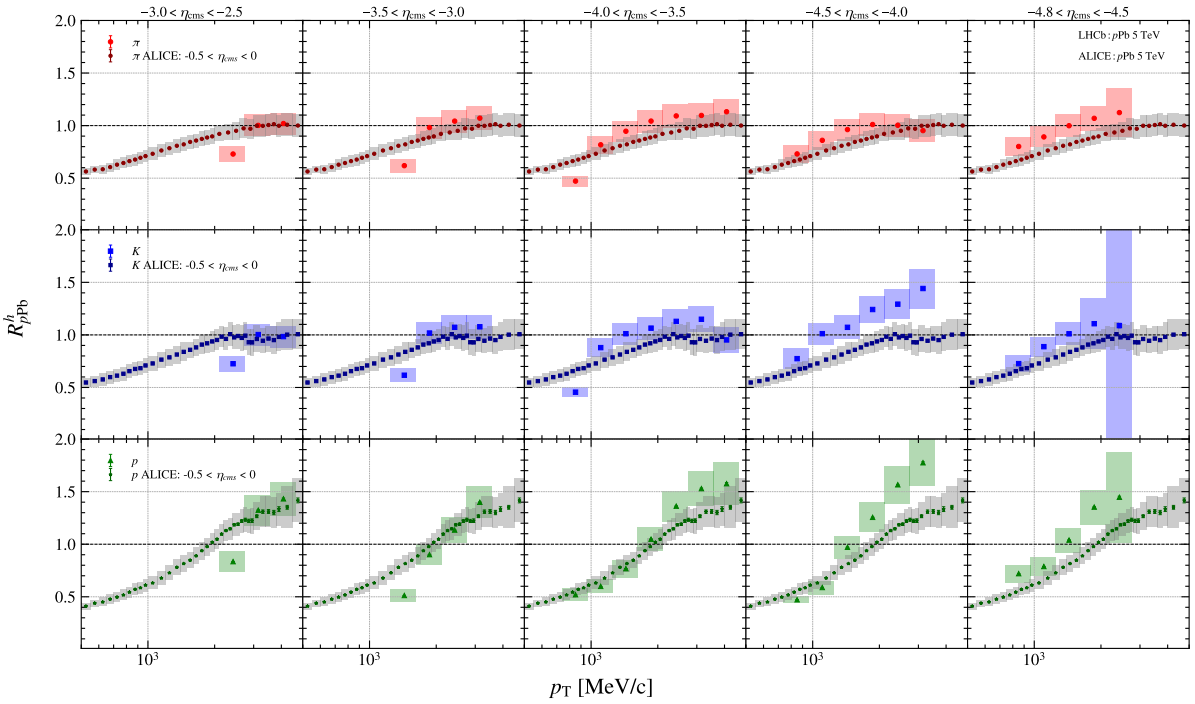


Figure 11.7: The nuclear modification factor of pions (red), kaons (blue), and protons (green). Statistical uncertainties are indicated by vertical error bars, and systematic uncertainties by boxes. The result of this work is compared with ALICE measurement at 5 TeV [99] following the styles indicated in the legend. The discrepancy observed at low p_{T} is due to the momentum cut applied in this analysis.

11.2 Particle Ratios and Correlations

The ratio of the yields of two hadron species is defined as:

$$R_{h_1/h_2} = \frac{N^{h_1}(\eta_{\mathrm{cms}}, p_{\mathrm{T}})}{N^{h_2}(\eta_{\mathrm{cms}}.p_{\mathrm{T}})}, \qquad \text{(eq. (6.5))}$$

and provides information on the relative production rates of pions, kaons, and protons. The yields of each particle species are obtained by inverting the PID matrix defined in eq. (9.3). This procedure introduces statistical correlations between the particle true yields due to the finite misidentification probabilities. For instance, an overestimate in the pion yield caused by misidentified kaons will result in a corresponding deficit in the kaon yield once the unfolding is applied. These interdependencies must be taken into account when evaluating the uncertainties on the particle ratios. To quantify the correlations between yields, a Monte Carlo approach is used following the next steps:

1. **Fluctuate the PID matrix**: For each $(\eta_{\mathrm{cms}}, p_{\mathrm{T}})$ bin, the elements of the PID matrix are varied within their uncertainties over $N = 1000$ iterations.

2. **Propagate fluctuations**: For each iteration, the inverted PID matrix is used to obtain a fluctuated estimate of the true yields.

3. **Evaluation of correlations**: The Pearson correlation coefficient is computed for each pair of hadron species to assess the linear dependence between their unfolded yields.

The obtained correlation coefficients range between -1 and 1, with positive (negative) values indicating a direct (inverse) linear relationship between the corresponding hadron yields. These coefficients are presented for pPb, Pbp, and pp in fig. 11.8, fig. 11.9, and fig. 11.10, respectively.

These correlations are propagated to the statistical uncertainty on the particle ratio via the covariance term. The covariance between two hadron species, h_1 and h_2, is obtained from the Pearson correlation coefficient $\rho(h_1, h_2)$ as:

$$\mathrm{Cov}(N^{h_1}, N^{h_2}) = \rho(h_1, h_2) \cdot \sigma(N^{h_1}) \cdot \sigma(N^{h_2}), \qquad (11.1)$$

where ΔN^{h_i} is the standard deviation of the unfolded yield N^{h_i} across the ensemble of Monte Carlo toys. The uncertainty on the particle ratio R_{h_1/h_2} is then calculated as:

$$\left(\frac{\sigma(R_{h_1/h_2})}{R_{h_1/h_2}}\right)^2 = \left(\frac{\sigma(N^{h_1})}{N^{h_1}}\right)^2 + \left(\frac{\sigma(N^{h_2})}{N^{h_2}}\right)^2 - 2\frac{\mathrm{Cov}(N^{h_1}, N^{h_2})}{N^{h_1} N^{h_2}}, \qquad (11.2)$$

where $\sigma(N^{h_i})$ is the uncertainty of the hadron yield i and $\mathrm{Cov}(N^{h_1}, N^{h_2})$ is the covariance between yields.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

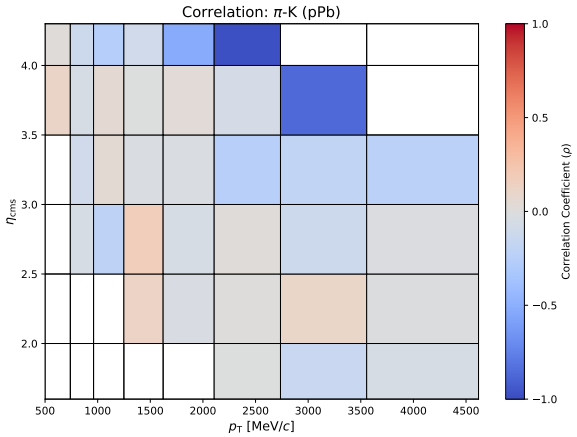
The proton-to-pion ratio, shown in fig. 11.12, for pPb, Pbp, and pp collisions, exhibits similar trends across the systems. The ratio is enhanced at intermediate transverse momenta, particularly at $p_{\mathrm{T}} \sim 2$–3 GeV/c. The kaon-to-pion ratio, shown in fig. 11.11, for pPb, Pbp, and pp collisions, exhibits similar trends across the systems. The ratio shows a smooth rise with p_{T}.

In these ratios, since the yields are obtained as:

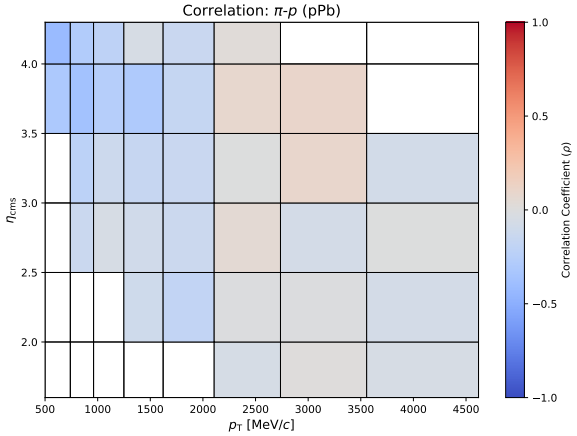
$$N^h = N^h_{True} \frac{1}{\varepsilon^h_{reco} \varepsilon_{sel} (1/\varepsilon_{TM})}, \qquad \text{(eq. (9.4))}$$

many systematic uncertainties cancel due to the similar reconstruction and selection conditions applied to both hadron species. In particular, uncertainties related to the event selection efficiency, tracking matching efficiency. However, species-dependent effects do not cancel and must be accounted for. These include the reconstruction efficiency, which differs due to particle-specific interactions with the detector material and acceptance, and the particle identification efficiency, which depends on the PID response and associated calibration. Additionally, differences in the contamination from secondary particles or fake tracks may contribute to the systematic uncertainty on the ratio. The measured kaon-to-pion and proton-to-pion ratios are further compared with published ALICE collaboration results for both Pbp and pp collisions [99, 128], as shown in fig. 11.13 and fig. 11.14, respectively.

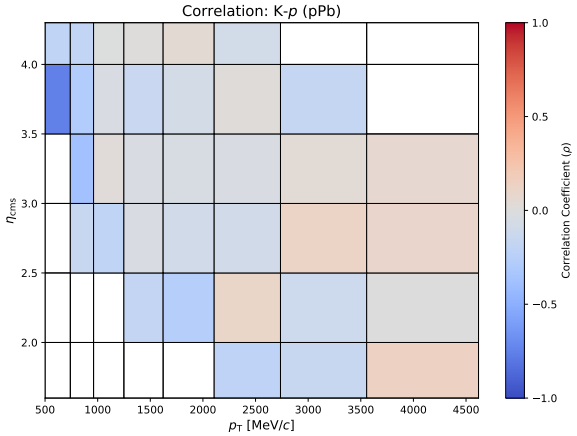
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) pion-kaon correlations



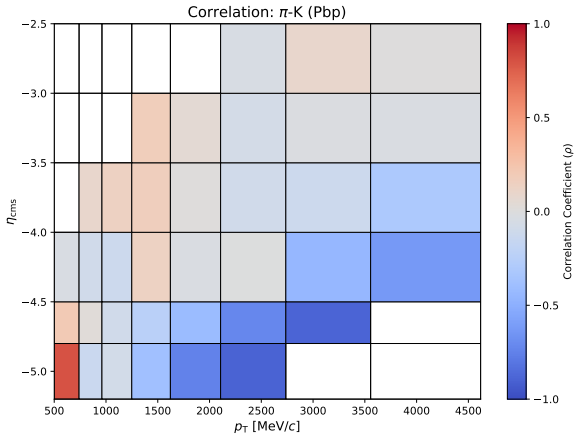
(b) pion-proton correlation



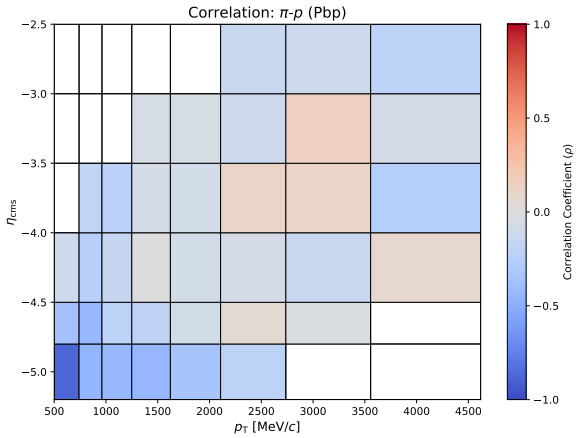
(c) proton-kaon correlation

Figure 11.8: Two-particle yield correlation heatmaps for identified hadron pairs in pPb collisions: (a) pion-kaon, (b) pion-proton, and (c) kaon-proton correlations.

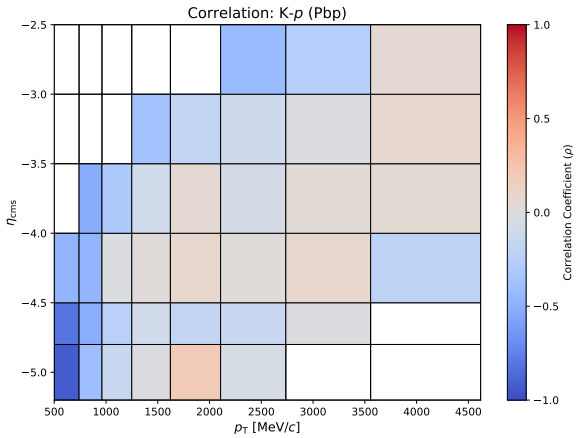
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) pion-kaon correlations



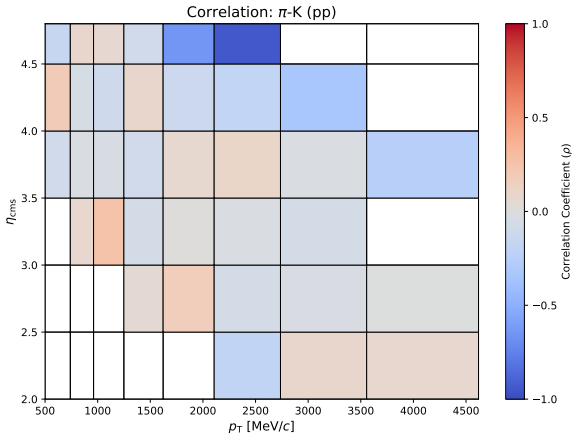
(b) pion-proton correlation



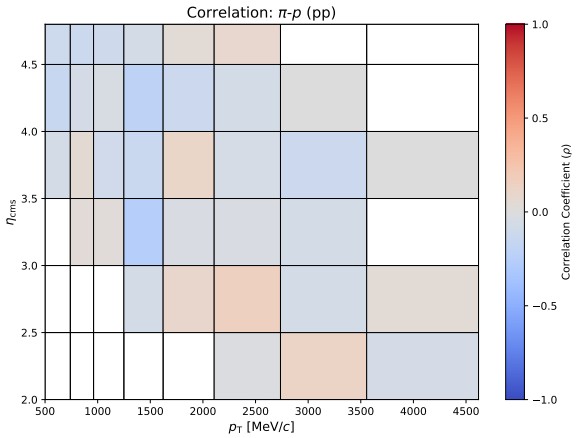
(c) proton-kaon correlation

Figure 11.9: Two-particle yield correlation heatmaps for identified hadron pairs in Pbp collisions: (a) pion-kaon, (b) pion-proton, and (c) kaon-proton correlations.

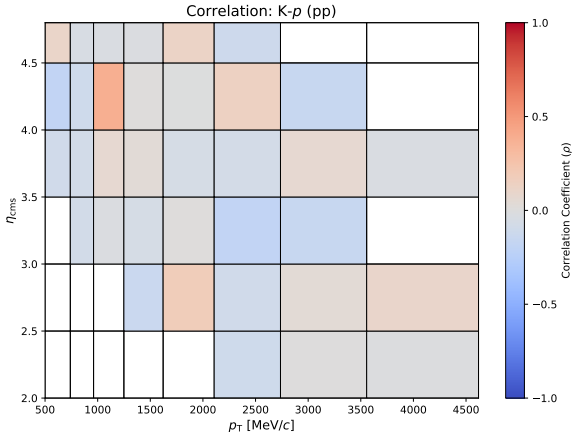
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) pion-kaon correlations



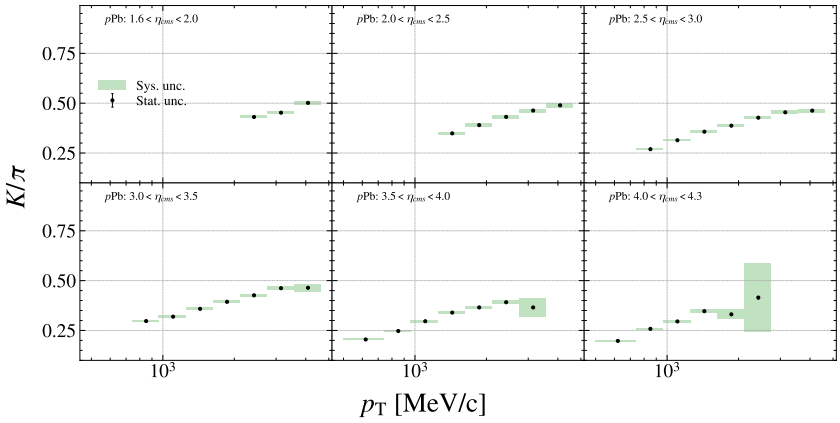
(b) pion-proton correlation



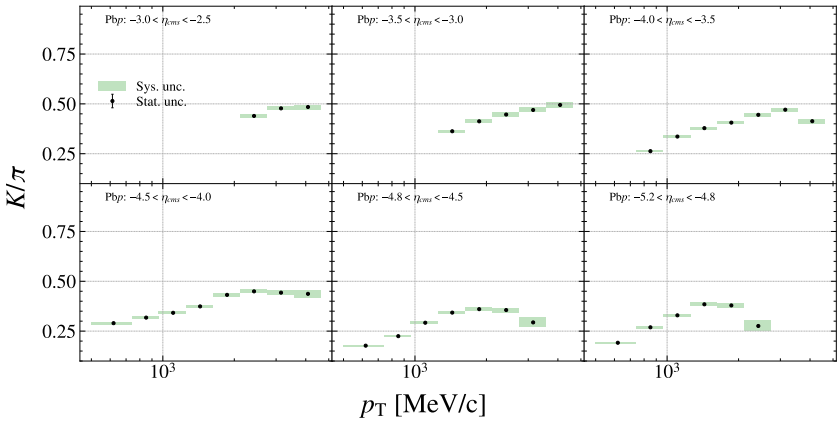
(c) proton-kaon correlation

Figure 11.10: Two-particle yield correlation heatmaps for identified hadron pairs in *pp* collisions: (a) pion-kaon, (b) pion-proton, and (c) kaon-proton correlations.

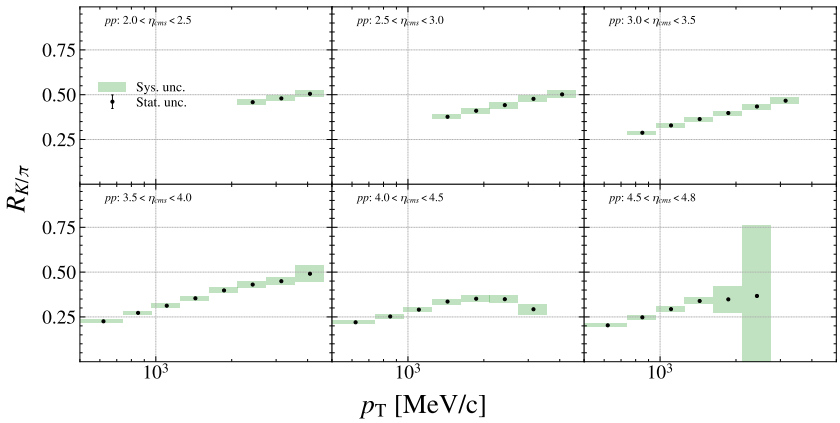
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) pPb.



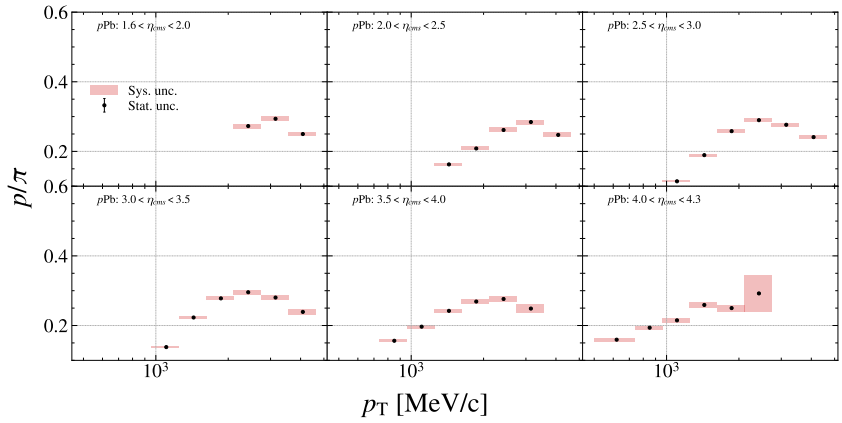
(b) Pbp.



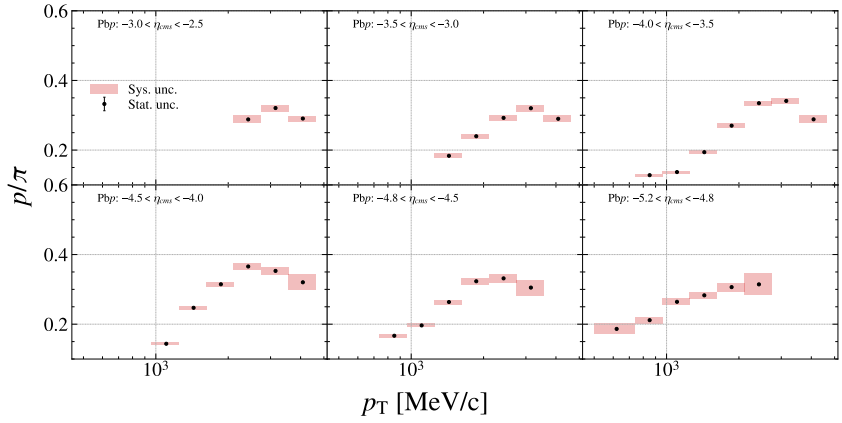
(c) pp.

Figure 11.11: Kaon over pion yield ratio in pPb (top), Pbp (middle), pp (bottom). Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes.

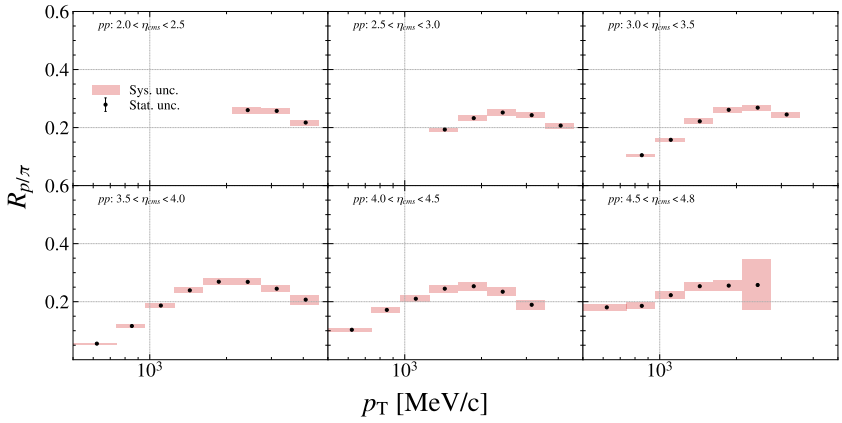
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) pPb.



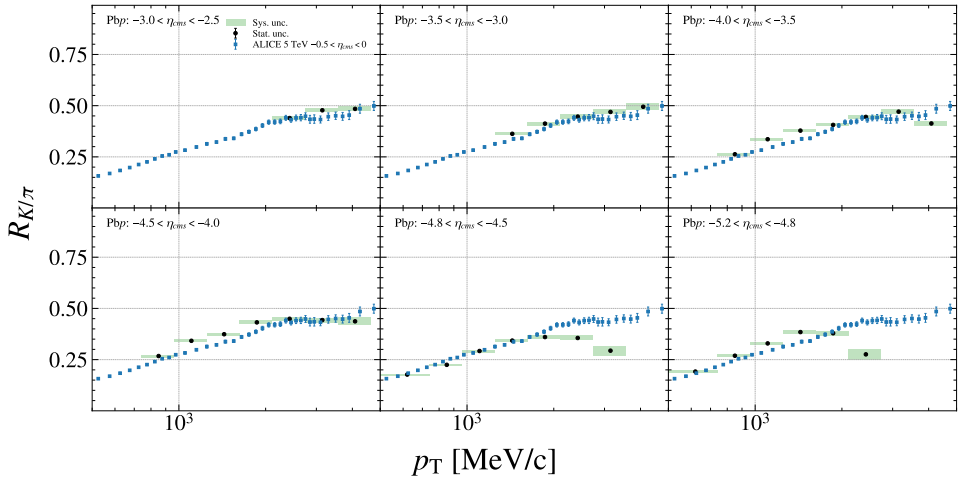
(b) Pbp.



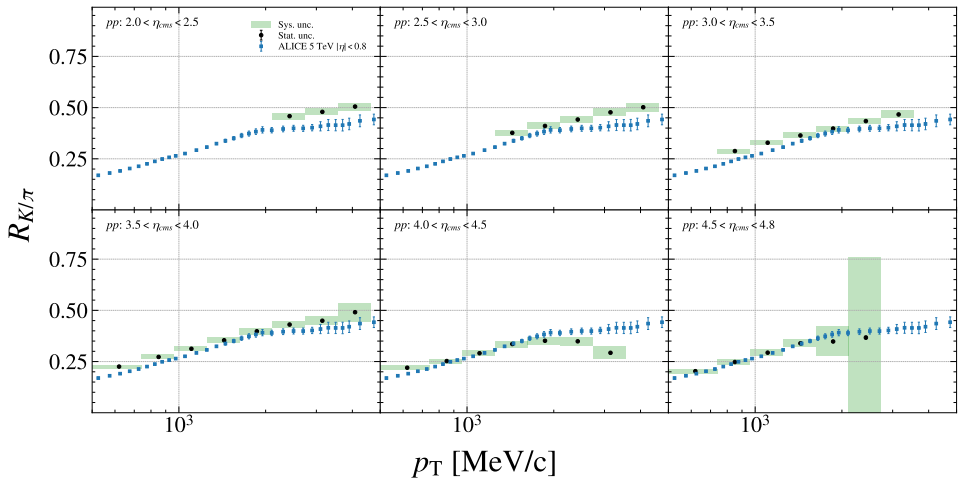
(c) pp.

Figure 11.12: proton over pion ratio in pPb (top), Pbp (middle), pp (bottom). Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

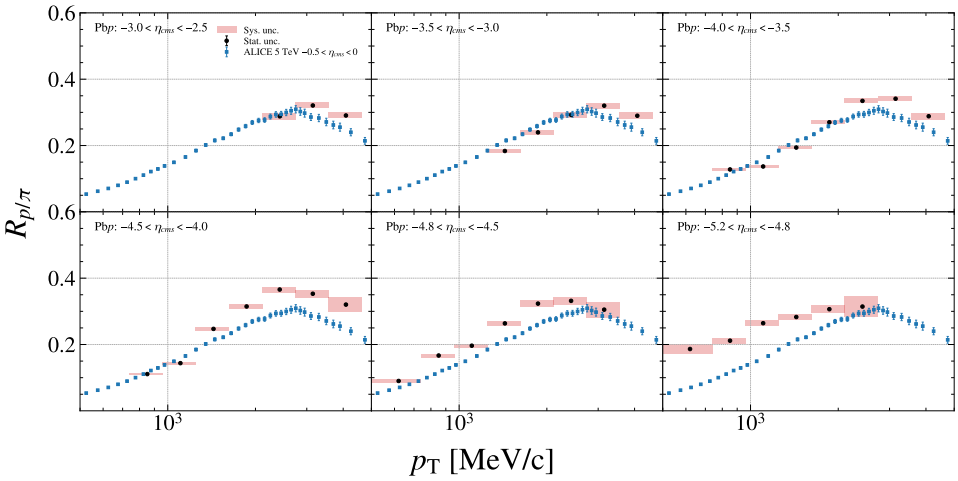


(a) Pbp.

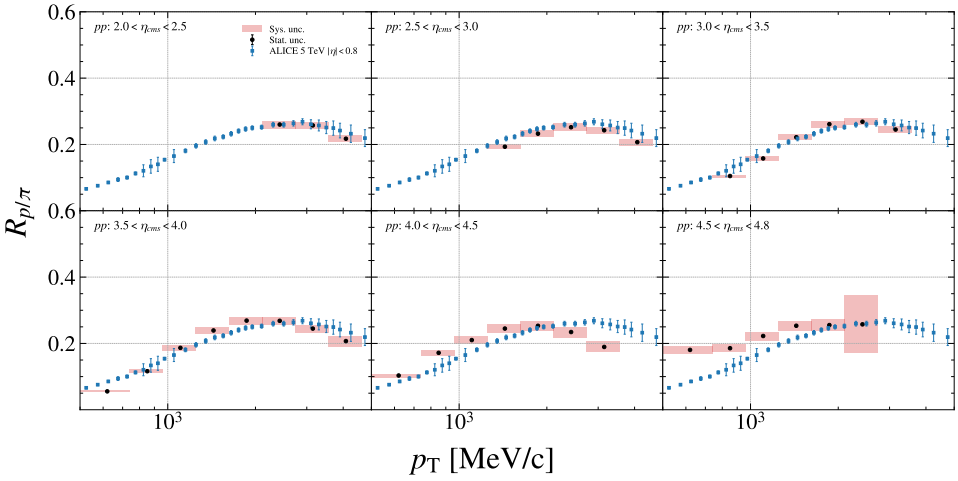


(b) pp.

Figure 11.13: kaon over pion ratio in Pbp (top) and pp (bottom) Compared with ALICE [99, 128]. Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes.



(a) Pbp.



(b) pp.

Figure 11.14: kaon over pion ratio in Pbp (top) and pp (bottom) Compared with ALICE [99, 128]. Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

12

Conclusions

This thesis presents the first measurements of identified charged pion, kaon, and proton production cross-sections in proton-proton and proton-lead collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV at the LHCb experiment. These measurements cover a broad kinematic range in pseudorapidity in the center of mass frame and transverse momentum. The study covers prompt charged particles with $p > 10\,\mathrm{GeV}/c$, $500 < p_{\mathrm{T}} < 4619\,\mathrm{MeV}/c$ and η_{cms} ranges of $-5.3 < \eta_{\mathrm{cms}} < -2.5$ forPbp data, $1.6 < \eta_{\mathrm{cms}} < 4.3$ for pPb data and $2 < \eta_{\mathrm{cms}} < 4.8$ for pp data.

The analysis uses data from pp collisions taken in 2015 and pPb collisions from 2013. identified charged particle yields were extracted from long tracks reconstructed by the LHCb tracking system, using a selection strategy to enhance signal purity. Validation of simulation samples for pPb, Pbp, and pp data was performed to determine and apply necessary corrections, including those for reconstruction efficiency, selection efficiency, and purity. Data-driven techniques were used to correct the simulation.

The differential production cross-sections for identified pions, kaons, and protons have been determined for the first time in pPb and pp collisions in both the forward and backward regions at the LHCb. These results, which are sensitive to different aspects of QCD dynamics, provide experimental constraints for Monte Carlo generators of hadronic collisions. Such generators are vital for a wide array of applications, from fundamental particle physics to cosmic ray physics, where their accuracy directly impacts the interpretation of observed phenomena.

A key output of this work is the determination of the nuclear modification factor, $R_{p\mathrm{Pb}}$. In the forward pseudorapidity region, a clear and strong suppression of identified charged particle production is observed, which appears to be largely independent of particle species (pion, kaon, or proton). In contrast, the backward region reveals a mass-dependent enhancement of particle production, particularly pronounced for protons compared to pions and kaons.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Comparisons with theoretical predictions using Next-to-Leading Order (NLO) perturbative QCD calculations, incorporating nuclear parton distribution functions (nPDFs) via the EPPS21 parametrization, show that current models successfully describe the observed suppression in the forward region. However, these models struggle to fully reproduce the observed mass-dependent enhancement in the backward region, particularly for protons. This discrepancy highlights the need for additional mechanisms beyond standard nPDFs to comprehensively describe particle production in nuclear collisions at LHC energies. The observed mass ordering in the backward region further provides unique constraints on gluon fragmentation functions and potential hadronization mechanisms in the nuclear environment.

The results gained from this thesis open several promising avenues for future research. A natural extension would involve a more detailed study of the hadron-dependent production mechanisms, potentially through comparisons with a wider range of theoretical models that explicitly account for mass-dependent effects or different hadronization schemes. From an experimental perspective, a key area for improvement lies in reducing the dominant systematic uncertainty, which currently stems from the size of the calibration samples used in this analysis. Future measurements with larger datasets, which would provide more statistics for the calibration procedures, would directly lead to a reduction in this uncertainty and a further increase in the precision of the presented results. Additionally, these measurements could be repeated for proton-oxygen collisions. This would offer a unique opportunity to constrain gluon nPDFs within the oxygen nucleus, providing valuable complementary information to the current proton-lead results and extending our understanding of nuclear effects across different nuclear sizes. Exploring these identified particle spectra at different collision energies or in other collision systems at LHCb could provide a richer dataset to constrain theoretical models and deepen our understanding of strong interaction dynamics.

To conclude, the original results presented in this thesis significantly advance our understanding of Quantum Chromodynamics in complex hadronic environments. The broad interest of these measurements spans various research areas, from refining our understanding of cold nuclear matter effects and gluon dynamics to improving the precision of Monte Carlo generators crucial for cosmic ray physics.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

A

Calibration samples preparation

A.1 Λ calibration sample

The decay channel $\Lambda \to p\pi^-$ is used to estimate proton PID efficiency. Events are selected using two stripping lines:
`StrippingLam0LLLine1V0ForPID` and `StrippingLam0LLLine2V0ForPID`.
The `StrippingLam0LLLine1V0ForPID` line is prescaled by a factor of 0.180, while the `StrippingLam0LLLine2V0ForPID` line is unprescaled and requires the p candidate total momentum to exceed 40 GeV/c. The events from the `StrippingLam0LLLine1V0ForPID` line are weighted by 1/0.180 to account for the prescaling before merging the two samples.

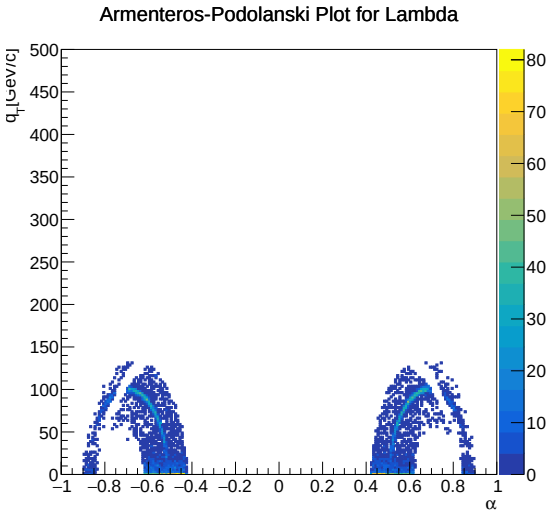
The purity of the sample is assessed using the Armenteros plot protect [129]. In two-body decays, this plot analyzes the kinematic properties of the decay products. The x-axis represents the transverse momentum q_T of one decay product relative to the direction of the parent particle, while the y-axis shows the longitudinal momentum asymmetry, defined as:

$$\alpha = \frac{(p_L)_1 - (p_L)_2}{(p_L)_1 + (p_L)_2}, \tag{A.1}$$

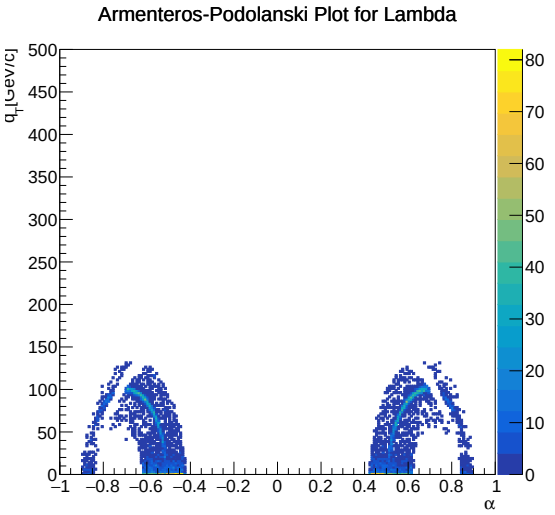
where $(p_L)_1$ and $(p_L)_2$ are the longitudinal momenta of the decay products. In the Armenteros plot, Λ clusters in the positive region of the y-axis, while $\bar{\Lambda}$ clusters in the negative region.

As illustrated in fig. A.1, the Armenteros plots for Λ baryons in both pPb and Pbp collisions demonstrate the high purity of the sample after the stripping line selections. The missing data from the characteristic arc of the Λ distribution is attributed to K^0_S mesons that have been successfully removed from our sample. A final check of the purity is also achieved by calculating the wrong mass assignment for p as π to check the presence of the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Armenteros plot for Λ in proton-lead (pPb) collisions

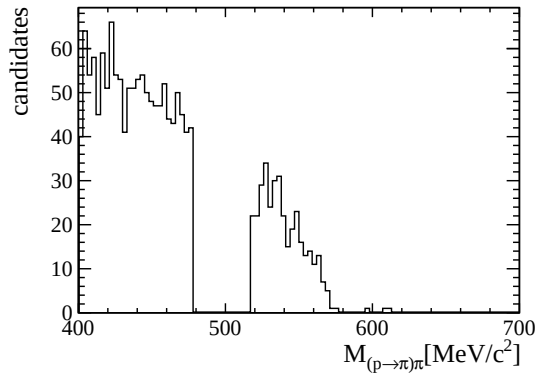


(b) Armenteros plot for Λ in lead-proton (Pbp) collisions

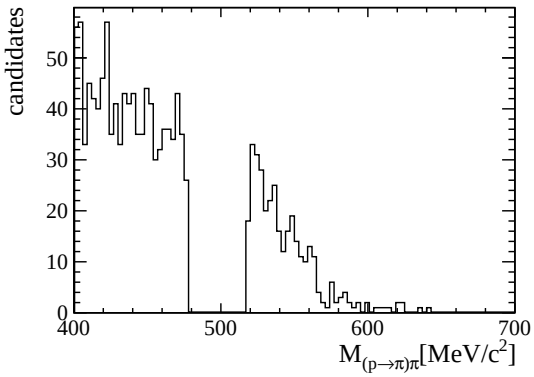
Figure A.1: Comparison of Armenteros plots for Λ baryons in different collision systems. The plots illustrate the transverse momentum (q_T) versus longitudinal momentum asymmetry (α) for the decay products of Λ baryons.

K^0_S mesons, the invariant mass with the $\pi\pi$ is therefore reconstructed and the result is shown in fig. A.2 To study the Λ signal, a dataset consisting of pPb and Pbp was generated using `MC-Sim09k-Multiplicity` with a magnet down polarity. The mass distribution of the Λ is then fitted to extract the fit parameters. These parameters are used for the fit of the mass yields when computing the PID and mis-ID efficiency.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

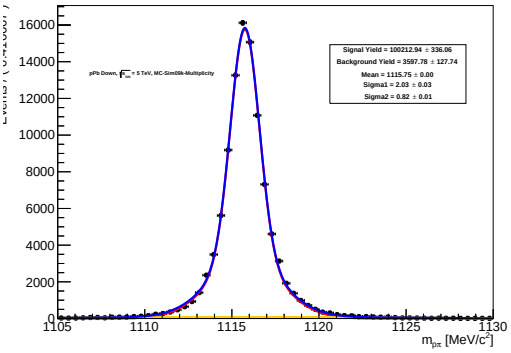


(a) Invariant mass in proton-lead (pPb) collisions

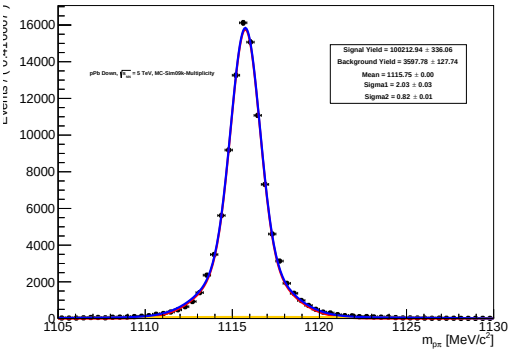


(b) invariant mass in lead-proton (Pbp) collisions

Figure A.2: The plots illustrate the invariant mass of the $\pi\pi$ final state assuming the p from Λ is a π



(a) Λ mass in proton-lead (pPb) collisions



(b) Λ mass in proton-lead (Pbp) collisions

Figure A.3: Results of the fit of Λ mass signal in pPb (left) and Pbp (right) with the fit parameters.

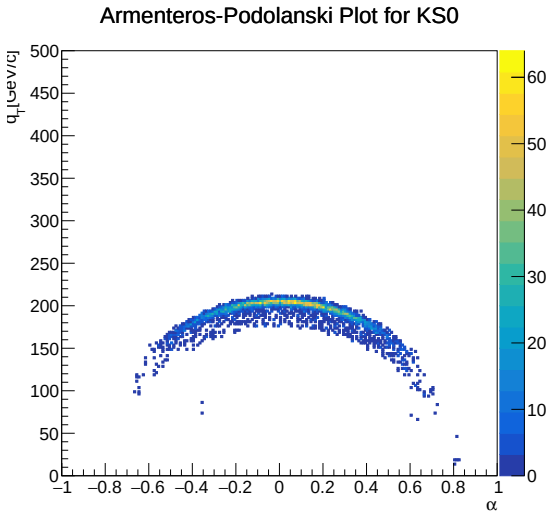
A.2 K^0_S calibration sample

The decay channel $K^0_S \to \pi^+\pi^-$ is used to estimate the PID efficiency of the pions. Similarly to the Λ sample explained in appendix A.1, there is a dedicated stripping line for the K^0_S sample `StrippingK0SLLLineV0ForPID` with a prescaling of 0.022. This stripping line is not used and will be replaced by the minimum bias with higher prescaling for more statistics.

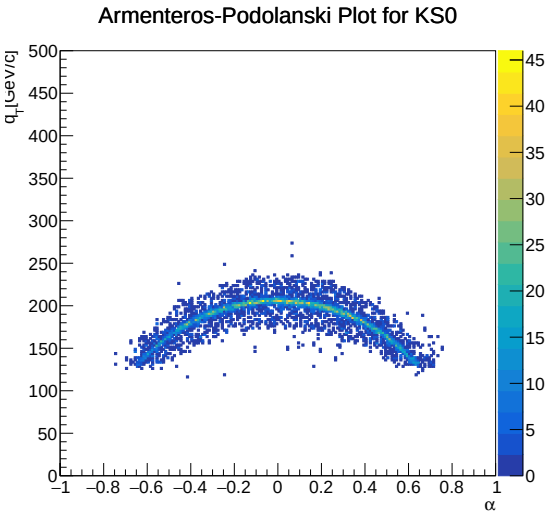
The q_T is required to be higher than 100 GeV/c to ensure the removal of any contamination from the Λ

To study the K^0_S signal, a dataset consisting of pPb and Pbp was generated using `MC-Sim09k-Multiplicity` with a magnet down polarity. The mass distribution of the K^0_S is then fitted to extract the fit parameters. These parameters are used for the fit of the mass yields when computing the PID and mis-ID efficiency.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

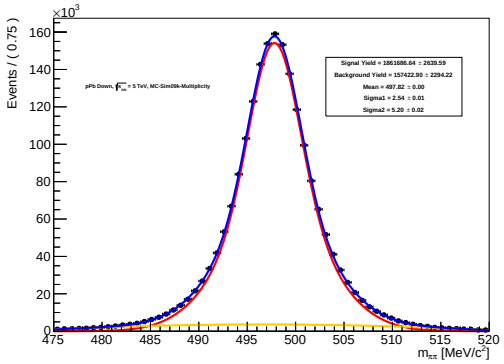


(a) Armenteros plot for K^0_S in proton-lead (pPb) collisions

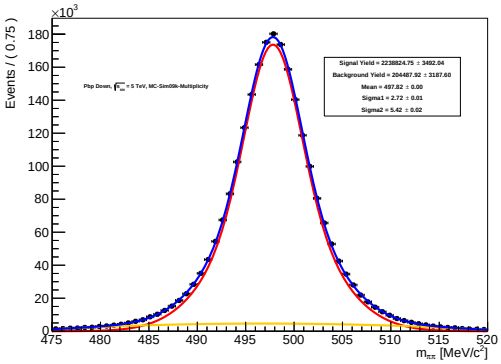


(b) Armenteros plot for K^0_S in lead-proton (Pbp) collisions

Figure A.4: Comparison of Armenteros plots for K^0_S baryons in different collision systems. The plots illustrate the transverse momentum (q_T) versus longitudinal momentum asymmetry (α) for the decay products of K^0_S baryons.



(a) K^0_S mass in proton-lead (pPb) collisions



(b) K^0_S mass in proton-lead (Pbp) collisions

Figure A.5: Results of the fit of K^0_S mass signal in pPb (left) and Pbp (right) with the fit parameters.

A.3 ϕ calibration sample

The decay channel $\phi(1020) \to K^+K^-$ is used. The decay $D_s \to \phi(1020)\pi$ is typically used to reduce the background contamination, this sample will not be used due to the statistical limitation. For this sample, the minimum bias stripping line is also used, in addition to the selections in table A.1.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Selection Component	**Criteria**
Daughters Cuts	K-: ALL
Combination Cut	(ADAMASS('phi(1020)') < 40 MeV) & (ACHI2DOCA(1,2) < 20)
Mother Cut	(ADMASS('phi(1020)') < 30 MeV) & (VFASPF(VCHI2/VDOF) < 25) & ((CHILDCUT((PROBNNk > 0.25),1)) | (CHILDCUT((PROBNNk > 0.25),2)))

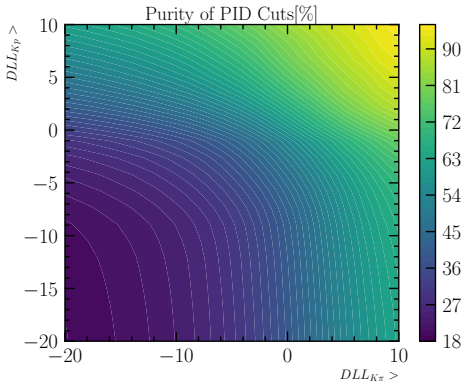
Table A.1: Selection criteria for $\phi(1020)$ decay into K^+K^-

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

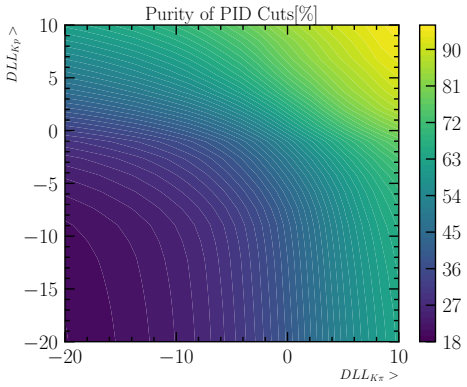
B

Additional figures

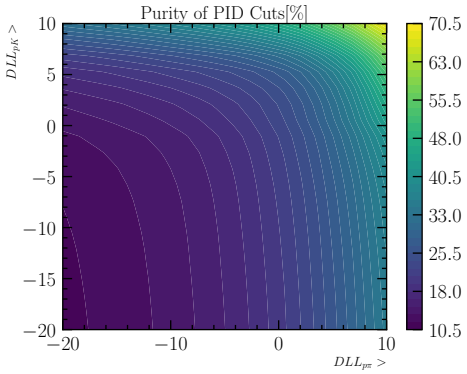
USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA



(a) Purity of pions



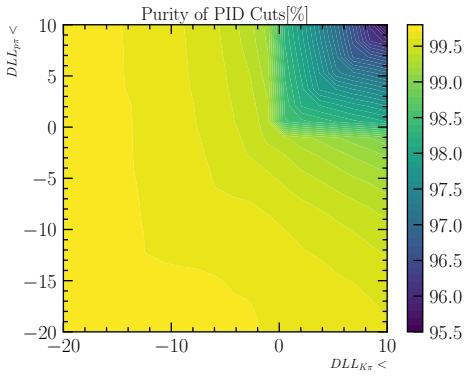
(b) Purity of kaons



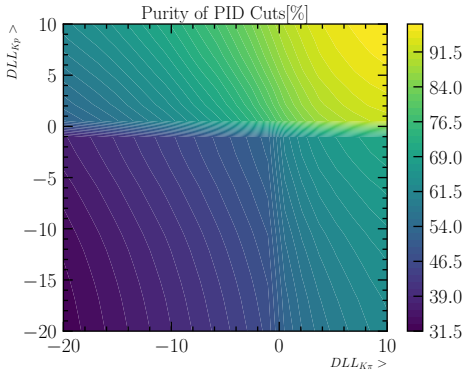
(c) Purity of protons

Figure B.1: Purity of pion (top), kaon (middle), and proton (bottom) candidates after applying PID cuts in Pbp simulated events. The purity is defined as in eq. (8.6), using truth-matched tracks after cuts on $\mathrm{DLL_{K\pi}}$ and $\mathrm{DLL_{pK}}$. The scans across the two-dimensional DLL space allow the determination of optimal PID working points, with final cuts chosen in regions where the purity exceeds 70%.

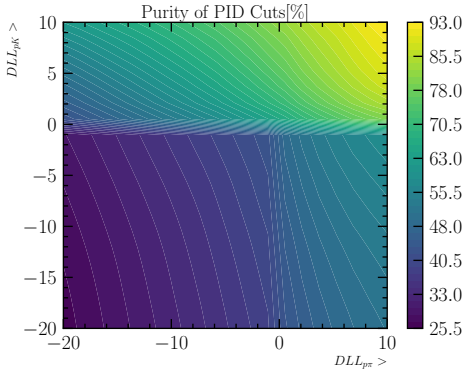
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Purity of pions



(b) Purity of kaons



(c) Purity of protons

Figure B.2: Purity of pion (top), kaon (middle), and proton (bottom) candidates after applying PID cuts in *pp* simulated events. The purity is defined as in eq. (8.6), using truth-matched tracks after cuts on $\mathrm{DLL_{K\pi}}$ and $\mathrm{DLL_{pK}}$. The scans across the two-dimensional DLL space allow the determination of optimal PID working points, with final cuts chosen in regions where the purity exceeds 70%.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

List of Figures

1.1 Cobertura cinemática de LHCb e outros experimentos para colisións pPb en termos do momento lonxitudinal fraccionario do nucleón no branco, x, e da masa invariante partón-partón ao cadrado, Q^2. 3

1.2 Porcentaxe de pureza da mostra de candidatos nas colisións pPb (parte superior), Pbp (parte central) e pp (parte inferior). Os resultados represéntanse para cada tipo de hadrón lixeiro: os pións en vermello, os kaóns en azul e os protóns en verde. As incertezas estatísticas indícanse mediante liñas verticais, mentres que as incertezas sistemáticas móstranse como caixas. . . 11

1.3 Eficiencia de reconstrucción nas colisións pPb (arriba), Pbp (no centro) e pp (abaixo). Os resultados están representados para cada tipo de hadrón lixeiro: os pións en vermello, os kaóns en azul e os protóns en verde. As incertezas estatísticas están indicadas con liñas verticais, mentres que as incertezas sistemáticas aparecen como caixas. 12

1.4 Eficiencia de selección nas colisión pPb (arriba), Pbp (no centro) e pp (abaixo). Os resultados están representados para cada tipo de hadrón lixeiro: os pións en vermello, os kaóns en azul e os protóns en verde. As incertezas estatísticas están indicadas con liñas verticais, mentres que as incertezas sistemáticas aparecen como caixas. 13

1.5 A sección eficaz de produción en colisións pPb (arriba), Pbp (centro) e pp (abaixo), comparada coas predicións do modelo EPPS21 (liñas negras discontinuas). Os resultados represéntanse para cada tipo de hadrón lixeiro: os pións móstranse en vermello, os kaóns en azul e os protóns en verde. As incertezas estatísticas indícanse mediante liñas verticais, mentres que as incertezas sistemáticas aparecen como caixas sombreadas. 16

1.6 O factor de modificación nuclear na rexión cara adiante (arriba) e na rexión cara atrás (abaixo). Os resultados represéntanse para cada tipo de hadrón lixeiro: os pións móstranse en vermello, os kaóns en azul e os protóns en verde. As incertezas estatísticas indícanse mediante liñas verticais, mentres que as incertezas sistemáticas aparecen como caixas sombreadas. 17

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

1.7 O factor de modificación nuclear na rexión cara adiante (arriba) e na rexión cara atrás (abaixo). Os resultados represéntanse para cada tipo de hadrón lixeiro: os pións móstranse en vermello, os kaóns en azul e os protóns en verde. As incertezas estatísticas indícanse mediante liñas verticais, mentres que as incertezas sistemáticas aparecen como caixas sombreadas. Os resultados na rexión cara adiante compáranse coa análise a 5 TeV de partículas cargadas primarias [8], mentres que na rexión cara atrás compáranse tanto cos resultados a 5 TeV como cos pións neutros a 8 TeV [7]. 18

1.8 Cociente de produción de kaóns sobre pións en pPb (arriba), Pbp (medio), pp (abaixo). As incertezas estatísticas están indicadas con liñas verticais, mentres que as incertezas sistemáticas se amosan con caixas. 19

4.1 Elementary particles in the Standard Model. Figure taken from [21] under Creative Commons license. 27

4.2 Measurements of α_s across energy scales compared to the 5-loop QCD prediction (solid curve), using $\alpha_s(m_Z^2) = 0.1180 \pm 0.0009$ from the Particle Data Group (PDG). Figure taken from [25] under Creative Commons license. 29

4.3 Schematic QCD phase diagram showing the phases of strongly interacting matter as functions of temperature T and baryon chemical potential μ_{B}. The dashed curve indicates the crossover transition at low μ_{B}. Figure taken from [32] under Creative Commons license. 30

4.4 Illustration of kinematic variables in collider particle physics. The beamline defines the z-axis, with the transverse plane defined by the x-y axes. The momentum vector $\vec{p}$ is shown, along with its transverse component $\vec{p}_T$ and longitudinal component p_z. The angles θ and ϕ are the polar and azimuthal angles, respectively. Figure taken from [37] under Creative Commons license. 32

4.5 Illustration of the factorization theorem for a process $h_1 + h_2 \to h + X$. Figure was adapted from [46] under Creative Commons license. 36

4.6 The CT18 parton distribution functions at $Q \sim 2\,\mathrm{GeV}$ and $Q \sim 100\,\mathrm{GeV}$ for u, $\bar{u}$, d, $\bar{d}$, s = $\bar{s}$, and g, as functions of the parton momentum fraction x. Figure was adapted from [48] under Creative Commons license. 38

4.7 Illustration of diffractive and non-diffractive processes in proton-proton collisions. (a) Non-diffractive interaction, where the entire proton energy contributes to particle production. (b) Single diffractive dissociation, where one proton dissociates into a system X, while the other remains intact, mediated by the exchange of a color-singlet object (denoted as $\mathbb{P}$). (c) Double diffractive dissociation, where both protons dissociate into systems X and Y through color-singlet exchanges. (d) Central diffraction, where a central system X is produced via the exchange of two color-singlet objects ($\mathbb{P}$) while both protons remain intact. Figure was adapted from [51] under Creative Commons license. 39

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.8 Evolution of a PbPb collision at the LHC, showing the transition from initial state to particle detection. 41

4.9 The nuclear modification factor $R_i^{\mathrm{Pb}}(x, Q^2 = 10\,\mathrm{GeV}^2)$ as a function of the Bjorken-x, calculated using the EPPS21 NLO nuclear parton distribution functions with 90% confidence level (CL) uncertainties. The definition of R_i^{Pb} is given in the equation, where f_i^{Pb} denotes the PDF for parton i in lead, and Z and N represent the atomic and neutron numbers, respectively. The figure illustrates different nuclear effects: shadowing at small x, anti-shadowing around $x \sim 10^{-1}$, the EMC effect at moderate x, and the Fermi motion at high x. Uncertainty bands for u and d quarks are shown in blue and orange, respectively. Figure taken from [56] under Creative Commons license. 44

4.10 Schematic of the structure of a $pp \to \bar{t}t$ event, as modeled by PYTHIA. The figure illustrates hard interactions, parton showers, hadronization, multiple parton interactions, and the role of beam remnants. Figure taken from [65] under Creative Commons license. 47

5.1 The diagram illustrates the layout of the CERN accelerator complex, highlighting the key components: the Large Hadron Collider (LHC), Super Proton Synchrotron (SPS), Proton Synchrotron (PS), and various other accelerators. It shows the flow of different particle beams, including protons, ions, and radioactive ion beams (RIBs), through the accelerator stages. Additionally, it highlights the main experimental areas such as ALICE, ATLAS, CMS, and LHCb. Figure taken from [69] under Creative Commons license. 50

5.2 Schematic side view of the LHCb detector, with the proton-proton interaction point located inside the Vertex Locator, shown on the left side of the image. In this context, "downstream" refers to detector components located after the dipole magnet along the beam direction (to the right in the image), while "upstream" refers to those located before the magnet. Figure taken from [76] under Creative Commons license. 58

5.3 Cross-sectional view of the VELO silicon sensors in both fully closed and open positions, with the front face of the first modules and the upstream pile-up veto stations also shown. Figure taken from [2] under Creative Commons license. 59

5.4 Diagram illustrating the five different track types reconstructed in LHCb tracking. Figure taken from [87] under Creative Commons license. 59

5.5 Reconstructed Cherenkov angle as a function of track momentum in the C_4F_{10} radiator. Figure taken from [89] under Creative Commons license. . 60

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5.6 Schematic overview of the LHCb offline software framework, illustrating the data flow from event generation (e.g., Pythia, EvtGen) through reconstruction (e.g., DaVinci, Moore) to storage and analysis (e.g., ROOT). Key components include Geant4 for particle propagation, Boole for digitization, and Allen for GPU-accelerated processing. Figure taken from [96] under Creative Commons license. 63

6.1 Kinematic coverage of LHCb and other experiments for pPb collisions in terms of the fractional longitudinal momentum, x, of the nucleon in the target and the squared parton-parton invariant mass, Q^2. Figure taken from [98] under Creative Commons license. 67

6.2 Nuclear modification factor ($R_{p\mathrm{Pb}}$) for neutral pions (π^0) and inclusive charged hadrons ($h^{\pm}$) in pPb collisions at LHCb. The top row (left and right panels) shows π^0 results at $\sqrt{s_{\mathrm{NN}}} = 8.16$ TeV in the backward ($-4.0 < \eta_{c.m.} < -3.0$) and forward ($2.5 < \eta_{c.m.} < 3.5$) [7] rapidity regions, respectively. Theoretical predictions from nCTEQ15+LHCb D^0 [49], EPPS16+LHCb D^0 [100], and CGC [101] models are shown for π^0. The bottom row (left and right panels) compares the π^0 results with inclusive charged hadron ($h^{\pm}$) measurements at $\sqrt{s_{\mathrm{NN}}} = 5$ TeV in similar backward and forward [8] rapidity regions. The x-axis represents transverse momentum (p_{T}) in GeV. The y-axis represents $R_{p\mathrm{Pb}}$. 69

8.1 Primary vertex position projected along the x (top), y (middle), and z (bottom) axes as a function of the event number for the pPb dataset. The quantities x_{PV}, y_{PV}, and z_{PV} correspond to the PV position projections onto the respective axes. These distributions are used to monitor the stability of the vertex reconstruction and the detector alignment over time. 81

8.2 Primary vertex position projected along the x (top), y (middle), and z (bottom) axes as a function of the event number for the Pbp dataset. The quantities x_{PV}, y_{PV}, and z_{PV} correspond to the PV position projections onto the respective axes. These distributions are used to monitor the stability of the vertex reconstruction and the detector alignment over time. 82

8.3 Primary vertex position projected along the x (top), y (middle), and z (bottom) axes as a function of the event number for the pp dataset. The quantities x_{PV}, y_{PV}, and z_{PV} correspond to the PV position projections onto the respective axes. These distributions are used to monitor the stability of the vertex reconstruction and the detector alignment over time. 83

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8.4 Projections of the Primary Vertex distribution and the fit to a three-dimensional Gaussian function that is used to calculate the average transverse PV position entering in the `pseudoIP` calculation of eq. (8.2). These Figures correspond to *MagDown* samples in three distinct beam configurations: Pbp (upper), pPb (middle), and pp (lower). The black and red dots represent data and simulation, respectively. The Gaussian function from the fit to the simulation dataset is presented as a green dashed line. The Gaussian function from the fit to data is represented with a blue line. . . . 90

8.5 Distribution of the `pseudoIP` variable in pPb, Pbp and pp collisions. Data are shown in brown and compared to the simulation (green). In the simulation, contributions from different sources of secondary backgrounds are distinguished: hadronic interactions (orange), particle decays (dark blue), photon conversions $\gamma \to e$ (cyan), and decays producing muons $X \to \mu$ (red). The total of all secondary tracks is labelled "all sec" (light blue). The definitions of the track categories are provided in table 8.3. These distributions are used to validate the modelling of secondary backgrounds in the simulation relative to data. 91

8.6 Purity of pion (top), kaon (middle), and proton (bottom) candidates after applying PID cuts in pPb simulated events. The purity is defined as in eq. (8.6), using truth-matched tracks after cuts on $\mathrm{DLL}_{\mathrm{K}\pi}$ and $\mathrm{DLL}_{\mathrm{pK}}$. The scans across the two-dimensional DLL space allow the determination of optimal PID working points, with final cuts chosen in regions where the purity exceeds 70%. 92

8.7 Comparison of occupancy-related and kinematic variables in data and in the `MC-Merged` pPb simulation sample. The simulation is shown after applying different sets of event weights: set 1 (blue), set 2 (green), and set 7 (red). The unweighted simulation is shown in grey, while the data are shown in black. These weighting strategies are based on different combinations of detector occupancy variables. 94

8.8 Comparison of occupancy-related and kinematic variables in data and in the `MC-Merged` Pbp simulation sample. The simulation is shown after applying different sets of event weights: set 1 (blue), set 2 (green), and set 7 (red). The unweighted simulation is shown in grey, while the data are shown in black. These weighting strategies are based on different combinations of detector occupancy variables. 95

8.9 Comparison of occupancy-related and kinematic variables in data and in the `MC-Sim09d` pPb simulation sample. The simulation is shown after applying different sets of event weights: set 1 (blue), set 2 (green), and set 7 (red). The unweighted simulation is shown in grey, while the data are shown in black. These weighting strategies are based on different combinations of detector occupancy variables. 96

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

9.1 Fraction of fake tracks in simulation, $f_{fake}^{sim}(h)$, for pPb (top), Pbp (middle), and pp (bottom) collisions. The fractions are shown separately for pions (red), kaons (blue), and protons (green). Only statistical uncertainties are displayed. These distributions illustrate the expected rate of fake tracks in the simulation before applying the data-driven correction factor R_{fake}^{h} (see eq. (9.6)). 103

9.2 Purity of the proxy sample used to estimate fake tracks in pPb (top), Pbp (middle), and pp (bottom) collisions. Fractions are shown for each hadron type: pions (red), kaons (blue), and protons (green). Only statistical uncertainties are displayed. 105

9.3 Data-to-simulation correction factors R^h for the proxy sample of fake tracks in pPb (top), Pbp (middle), and pp (bottom) collisions. Fractions are shown for each hadron type: pions (red), kaons (blue), and protons (green). Only statistical uncertainties are displayed. These factors quantify the discrepancy between simulation and data for fake-track rates. 106

9.4 Fraction of fake tracks in pPb (top), Pbp (middle), and pp (bottom) collisions for each type of light hadron: pions (red), kaons (blue), and protons (green). Statistical uncertainties are shown as vertical lines, and systematic uncertainties are represented by boxes. Only the *MagDown* polarity is displayed. These distributions illustrate the expected fake-track contributions in the dataset. 107

9.5 Fraction of secondary tracks in simulation, $f_{sec}^{sim}(h)$, contaminating the candidate sample in pPb events for each hadron type: pions (top), kaons (middle), and protons (bottom). The stacked contributions from individual secondary sources are shown with the same labels as in the figure: had_int (red), $X \to \mu$ (brown), $\gamma \to e$ (green), and decay (blue). The black points labeled "all" correspond to the total fraction of secondary tracks. The occupancy weight set 1 is applied, only *MagDown* polarity is displayed, and only statistical uncertainties are shown. 110

9.6 Fraction of secondary tracks in simulation, $f_{sec}^{sim}(h)$, contaminating the candidate sample in Pbp events for each hadron type: pions (top), kaons (middle), and protons (bottom). The stacked contributions from individual secondary sources are shown with the same labels as in the figure: had_int (red), $X \to \mu$ (brown), $\gamma \to e$ (green), and decay (blue). The black points labeled "all" correspond to the total fraction of secondary tracks. The occupancy weight set 1 is applied, only *MagDown* polarity is displayed, and only statistical uncertainties are shown. 111

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

9.7 Fraction of secondary tracks in simulation, $f_{sec}^{sim}(h)$, contaminating the candidate sample in pp events for each hadron type: pions (top), kaons (middle), and protons (bottom). The stacked contributions from individual secondary sources are shown with the same labels as in the figure: had_int (red), $X \to \mu$ (brown), $\gamma \to e$ (green), and decay (blue). The black points labeled "all" correspond to the total fraction of secondary tracks. The occupancy weight set 1 is applied, and only statistical uncertainties are shown. 112

9.8 Purity of the proxy sample used to estimate the had_decay component of secondary tracks in pPb (top), Pbp (middle), and pp (bottom) collisions. Fractions are shown for each hadron type: pions (red), kaons (blue), and protons (green). Only statistical uncertainties are displayed. 114

9.9 Purity of the proxy sample used to estimate the $\gamma \to e$ component of secondary tracks in pPb (top), Pbp (middle), and pp (bottom) collisions. Fractions are shown for each hadron type: pions (red), kaons (blue), and protons (green). Only statistical uncertainties are displayed. 115

9.10 Data-to-simulation correction factors R^h for the proxy sample of had_int tracks in pPb (top), Pbp (middle), and pp (bottom) collisions. Fractions are shown for each hadron type: pions (red), kaons (blue), and protons (green). Only statistical uncertainties are displayed. These factors quantify the discrepancy between simulation and data for hadronic-interaction secondary tracks. 116

9.11 Data-to-simulation correction factors R^h for the proxy sample of $\gamma \to e$ tracks in pPb (top), Pbp (middle), and pp (bottom) collisions. Fractions are shown for each hadron type: pions (red), kaons (blue), and protons (green). Only statistical uncertainties are displayed. These factors quantify the discrepancy between simulation and data for photon-conversion secondary tracks. 117

9.12 Fraction of secondary tracks in pPb collisions for each type of light hadron: pions (red), kaons (blue), and protons (green). The contributions are separated into three categories: muons from decays ($X \to \mu$, green), photon conversions ($\gamma \to e$, brown), and had_decay (pink). Statistical uncertainties are shown as vertical lines. These distributions illustrate the expected secondary-track contributions in the dataset. 118

9.13 Fraction of secondary tracks in Pbp collisions for each type of light hadron: pions (red), kaons (blue), and protons (green). The contributions are separated into three categories: muons from decays ($X \to \mu$, green), photon conversions ($\gamma \to e$, brown), and had_decay (pink). Statistical uncertainties are shown as vertical lines. These distributions illustrate the expected secondary-track contributions in the dataset. 119

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

9.14 Fraction of secondary tracks in pp collisions for each type of light hadron: pions (red), kaons (blue), and protons (green). The contributions are separated into three categories: muons from decays ($X \to \mu$, green), photon conversions ($\gamma \to e$, brown), and had_decay (pink). Statistical uncertainties are shown as vertical lines. These distributions illustrate the expected secondary-track contributions in the dataset. 120
9.15 Fraction of secondary tracks in pPb (top), Pbp (middle), and pp (bottom) collisions for each type of light hadron: pions (red), kaons (blue), and protons (green). Statistical uncertainties are shown as vertical lines, and systematic uncertainties are represented by boxes. These distributions illustrate the expected secondary-track contributions in the dataset. 121
9.16 Percentage of the purity in pPb (top), Pbp (middle), and pp (bottom) collisions. The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green. Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes. 123
9.17 Mass fits for different particles in pPb collisions: (a) $K^0_S \to \pi^+\pi^-$, (b) $\Lambda \to p\pi^-$, (c) and (d) $\phi \to K^+K^-$. 128
9.18 Mass fits for different particles in Pbp collisions: (a) $K^0_S \to \pi^+\pi^-$, (b) $\Lambda \to p\pi^-$, (c) $D^0 \to K^-\pi^+$, and (d) $\phi \to K^+K^-$. 129
9.19 Mass fits for different particles in pp collisions: (a) $K^0_S \to \pi^+\pi^-$, (b) $\Lambda \to p\pi^-$, (c) $D^0 \to K^-\pi^+$ and (d) $\phi \to K^+K^-$. 130
9.20 Flowchart illustrating the iterative reweighting process for PID efficiency calculations. The process begins with initial reweighting of probe momentum to the total momentum of the minimum bias data, followed by estimating the momentum distributions of pions, kaons and protons in the minimum bias data and iterative refinement of weights until convergence. . 131
9.21 Kinematic distributions of charged-pion candidates in pPb data. The grey shaded region shows the unweighted minimum-bias sample (all charged tracks). The yellow shaded region corresponds to the same sample after applying pion-specific weights derived from the iterative procedure (see section 9.3.4). Pions from $K^0_S \to \pi^+\pi^-$ decays are shown as a red hatched distribution, while the same $K^0_S \to \pi^+\pi^-$ sample reweighted with sets 1, 2, and 3 (see section 9.3.2) are shown in blue, pink, and green, respectively. . 132
9.22 Kinematic distributions of proton candidates in pPb data. The grey shaded region shows the unweighted minimum-bias sample (all charged tracks). The green shaded region with stars corresponds to the same sample after applying the proton-specific weights from the fifth iteration of the iterative procedure (see section 9.3.4). Protons from $\Lambda \to p\pi^-$ decays are shown as a red hatched distribution, while the same $\Lambda \to p\pi^-$ sample reweighted with sets 1, 2, and 3 (see section 9.3.2) are shown in blue, magenta, and green, respectively. 133

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

9.23 Kinematic distributions of kaon candidates in pPb data. The grey shaded region shows the unweighted minimum-bias sample (all charged tracks). The green shaded region with stars corresponds to the same sample after applying kaon-specific weights from the fifth iteration of the iterative procedure (see section 9.3.4). Kaons from $\phi \to K^+K^-$ decays are shown as a red hatched distribution, while the same $\phi \to K^+K^-$ sample reweighted with sets 1, 2, and 3 (see section 9.3.2) are shown in blue, magenta, and green, respectively. 134

9.24 Kinematic distributions of charged-pion candidates in pPb data. The grey shaded region shows the unweighted minimum-bias sample (all charged tracks). The green shaded region with stars corresponds to the same sample after applying pion-specific weights from the fifth iteration of the iterative procedure (see section 9.3.4). Pions from $D^{*+} \to D^0\pi^+$ decays are shown as a red hatched distribution, while the same $D^{*+} \to D^0\pi^+$ sample reweighted with sets 1, 2, and 3 (see section 9.3.2) are shown in blue, magenta, and green, respectively. 135

9.25 Kinematic distributions of charged-kaon candidates in pPb data. The grey shaded region shows the unweighted minimum-bias sample (all charged tracks). The green shaded region with stars corresponds to the same sample after applying pion-specific weights from the fifth iteration of the iterative procedure (see section 9.3.4). Kaons from $D^{*+} \to D^0\pi^+$ decays are shown as a red hatched distribution, while the same $D^{*+} \to D^0\pi^+$ sample reweighted with sets 1, 2, and 3 (see section 9.3.2) are shown in blue, magenta, and green, respectively. 136

9.26 Kinematic distributions of charged-pion candidates in Pbp data. The grey shaded region shows the unweighted minimum-bias sample (all charged tracks). The yellow shaded region corresponds to the same sample after applying pion-specific weights derived from the iterative procedure (see section 9.3.4). Pions from $K^0_S \to \pi^+\pi^-$ decays are shown as a red hatched distribution, while the same $K^0_S \to \pi^+\pi^-$ sample reweighted with sets 1, 2, and 3 (see section 9.3.2) are shown in blue, pink, and green, respectively. . 137

9.27 Kinematic distributions of proton candidates in Pbp data. The grey shaded region shows the unweighted minimum-bias sample (all charged tracks). The green shaded region with stars corresponds to the same sample after applying the proton-specific weights from the fifth iteration of the iterative procedure (see section 9.3.4). Protons from $\Lambda \to p\pi^-$ decays are shown as a red hatched distribution, while the same $\Lambda \to p\pi^-$ sample reweighted with sets 1, 2, and 3 (see section 9.3.2) are shown in blue, magenta, and green, respectively. 138

9.28 Ratio between data and simulation tracking efficiency $C_{tracking} = \varepsilon^{\text{data}}_{tracking}/\varepsilon^{\text{MC}}_{tracking}$ in 2012 pp calibration data. 141

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

9.29 Ratio between data and simulation tracking efficiency $C_{tracking} = \varepsilon^{\text{data}}_{tracking}/\varepsilon^{\text{MC}}_{tracking}$ in 2015 pp calibration data. 142
9.30 Reconstruction efficiency in pPb (top), Pbp (middle), and pp (bottom) events. The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green. Statistical uncertainties are indicated with vertical lines, while systematic uncertainties are displayed as boxes. This efficiency quantifies the probability that a charged hadron produced in the collision is successfully reconstructed as a track by the detector. 143
9.31 Selection efficiency coefficients (C_{sel}) are shown as a function in $(\eta_{\text{cms}}, p_{\text{T}})$ bins for different collision systems: (a) pPb (top), (b) Pbp (middle), and (c) pp (bottom). 146
9.32 Selection efficiencies are shown as a function of $(\eta_{\text{cms}}, p_{\text{T}})$ bins for different collision systems: (a) pPb (top), (b) Pbp (middle), and (c) pp (bottom). The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green. Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes. Only *MagDown* polarity is shown in the figure. 147

10.1 Purity of the alternative proxy samples used to assess systematic uncertainties in the determination of R^h_{fake} in pPb (top), Pbp (middle), and pp (bottom) collisions. Fractions are shown for each hadron type: pions (red), kaons (blue), and protons (green). 153
10.2 Data-to-simulation correction factors R^h for the alternative proxy sample of fake tracks in pPb (top), Pbp (middle), and pp (bottom) collisions. Fractions are shown for each hadron type: pions (red), kaons (blue), and protons (green). These factors quantify the discrepancy between simulation and data for fake-track rates. 154
10.3 Total systematic uncertainty of f^h_{fake} in pPb (top), Pbp (middle), and pp (bottom) collisions. The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green. 155
10.4 Total systematic uncertainty of f^h_{sec} in pPb (top), Pbp (middle), and pp (bottom) collisions. The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green. 158
10.5 Total systematic uncertainty in the purity (P^h) determination for pPb (top), Pbp (middle), and pp (bottom) collisions. The results are represented for each type of light hadron: pions are shown in red, kaons in blue and protons in green. 160
10.6 Distributions of the N_{True} values for pions, kaons, and protons obtained with the toy procedure in pPb collisions, for $p_{\text{T}} \in [2107, 2737]$ MeV/c and $\eta_{\text{cms}} \in [1.6, 2.0, 2.5]$. 165

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

10.7 Distributions of the N_{True} values for pions, kaons, and protons obtained with the toy procedure in pPb collisions, for $p_{\mathrm{T}} \in [2107, 2737]$ MeV/c and $\eta_{\mathrm{cms}} \in [2.5, 3.0, 3.5]$. 166

10.8 Summary of the relative systematic uncertainties on N^h_{True} in pPb collisions for (a) pions, (b) kaons, and (c) protons. The contributions from the purity uncertainty ($\sigma^h_{\text{ToyMC-Nsel}}$, magenta), from the PID efficiency matrix uncertainty ($\sigma^h_{\text{ToyMC-PID}}$, green), and their quadrature sum ($\sigma(N^h_{True})$, black) are shown. The uncertainties are evaluated using the toy Monte Carlo procedure described in the text. 167

10.9 Summary of the relative systematic uncertainties on N^h_{True} in Pbp collisions for (a) pions, (b) kaons, and (c) protons. The contributions from the purity uncertainty ($\sigma^h_{\text{ToyMC-Nsel}}$, magenta), from the PID efficiency matrix uncertainty ($\sigma^h_{\text{ToyMC-PID}}$, green), and their quadrature sum ($\sigma(N^h_{True})$, black) are shown. The uncertainties are evaluated using the toy Monte Carlo procedure described in the text. 168

10.10 Summary of the relative systematic uncertainties on N^h_{True} in pp collisions for (a) pions, (b) kaons, and (c) protons. The contributions from the purity uncertainty ($\sigma^h_{\text{ToyMC-Nsel}}$, magenta), from the PID efficiency matrix uncertainty ($\sigma^h_{\text{ToyMC-PID}}$, green), and their quadrature sum ($\sigma(N^h_{True})$, black) are shown. The uncertainties are evaluated using the toy Monte Carlo procedure described in the text. 169

10.11 Uncertainty in proton reconstruction efficiency from material interactions, $\sigma_{\text{mat}}(\varepsilon^p_{reco})$, for pp collisions. 172

10.12 Systematic uncertainties on the reconstruction efficiency in the pPb sample for each particle species and source: σ^h_{bin} (tracking correction uncertainty), σ^h_{mat} (material differences), σ^h_{global} (global effects), σ^h_{occ} (detector occupancy), and σ^h_{stat} (statistical uncertainty). 173

10.13 Systematic uncertainties on the reconstruction efficiency in the Pbp sample for each particle species and source: σ^h_{bin} (tracking correction uncertainty), σ^h_{mat} (material differences), σ^h_{global} (global effects), σ^h_{occ} (detector occupancy), and σ^h_{stat} (statistical uncertainty). 174

10.14 Systematic uncertainties on the reconstruction efficiency in the pp sample for each particle species and source: σ^h_{bin} (tracking correction uncertainty), σ^h_{mat} (material differences), σ^h_{global} (global effects), σ^h_{occ} (detector occupancy), and σ^h_{stat} (statistical uncertainty). 175

10.15 Total systematic uncertainty of the selection efficiency in pPb (top), Pbp (middle), pp (bottom) events. The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green . 177

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

11.1 Production cross sections in pPb (top), Pbp (middle) and pp (bottom) events. The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green. Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes. 180
11.2 Production cross sections in pPb (top), Pbp (middle) and pp (bottom) events compared to EPPS21 model predictions (dashed black lines). The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green. Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes. . . . 181
11.3 The Nuclear modification factor in the forward region(top) and the backward region (bottom). The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green. Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes. 182
11.4 The Nuclear modification factor in the forward region(top) and the backward region (bottom) compared with prompt charged particles ($h^{\pm}$) [8]. The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green. Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes. The discrepancy observed at low p_{T} is due to the momentum cut applied in this analysis. 183
11.5 Nuclear modification factors in the forward (top) and backward (bottom) regions. Results from this work are shown for pions (red), kaons (blue), and protons (green). Statistical uncertainties are indicated by vertical error bars, and systematic uncertainties by boxes. The result of this work is compared with previous LHCb measurement π^0 at 8 TeV [7] following the styles indicated in the legend. The discrepancy observed at low p_{T} is due to the momentum cut applied in this analysis. 184
11.6 The nuclear modification factor in the forward (top) and the backward region (bottom) compared to EPPS21 model predictions. The results are represented for each type of light hadron: pions are shown in red, kaons in blue, and protons in green. Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes. 185
11.7 The nuclear modification factor of pions (red), kaons (blue), and protons (green). Statistical uncertainties are indicated by vertical error bars, and systematic uncertainties by boxes. The result of this work is compared with ALICE measurement at 5 TeV [99] following the styles indicated in the legend. The discrepancy observed at low p_{T} is due to the momentum cut applied in this analysis. 186
11.8 Two-particle yield correlation heatmaps for identified hadron pairs in pPb collisions: (a) pion-kaon, (b) pion-proton, and (c) kaon-proton correlations. 189

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

11.9 Two-particle yield correlation heatmaps for identified hadron pairs in Pbp collisions: (a) pion-kaon, (b) pion-proton, and (c) kaon-proton correlations. 190
11.10Two-particle yield correlation heatmaps for identified hadron pairs in pp collisions: (a) pion-kaon, (b) pion-proton, and (c) kaon-proton correlations. 191
11.11Kaon over pion yield ratio in pPb (top), Pbp (middle), pp (bottom). Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes. 192
11.12proton over pion ratio in pPb (top), Pbp (middle), pp (bottom). Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes. 193
11.13kaon over pion ratio in Pbp (top) and pp (bottom) Compared with ALICE [99,128]. Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes. 194
11.14kaon over pion ratio in Pbp (top) and pp (bottom) Compared with ALICE [99,128]. Statistical uncertainties are indicated by vertical lines, while systematic uncertainties are displayed as boxes. 195

A.1 Comparison of Armenteros plots for Λ baryons in different collision systems. The plots illustrate the transverse momentum (q_T) versus longitudinal momentum asymmetry (α) for the decay products of Λ baryons. 199
A.2 The plots illustrate the invariant mass of the $\pi\pi$ final state assuming the p from Λ is a π 200
A.3 Results of the fit of Λ mass signal in pPb (left) and Pbp (right) with the fit parameters. 200
A.4 Comparison of Armenteros plots for K^0_S baryons in different collision systems. The plots illustrate the transverse momentum (q_T) versus longitudinal momentum asymmetry (α) for the decay products of K^0_S baryons. . . . 202
A.5 Results of the fit of K^0_S mass signal in pPb (left) and Pbp (right) with the fit parameters. 202

B.1 Purity of pion (top), kaon (middle), and proton (bottom) candidates after applying PID cuts in Pbp simulated events. The purity is defined as in eq. (8.6), using truth-matched tracks after cuts on $\mathrm{DLL}_{\mathrm{K}\pi}$ and $\mathrm{DLL}_{\mathrm{pK}}$. The scans across the two-dimensional DLL space allow the determination of optimal PID working points, with final cuts chosen in regions where the purity exceeds 70%. 206
B.2 Purity of pion (top), kaon (middle), and proton (bottom) candidates after applying PID cuts in pp simulated events. The purity is defined as in eq. (8.6), using truth-matched tracks after cuts on $\mathrm{DLL}_{\mathrm{K}\pi}$ and $\mathrm{DLL}_{\mathrm{pK}}$. The scans across the two-dimensional DLL space allow the determination of optimal PID working points, with final cuts chosen in regions where the purity exceeds 70%. 207

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

List of Tables

1.1 Anchura (Γ) e vida media propia (τ) das partículas de tempo de vida longo. As partículas con $\Gamma = 0$ considéranse estables para os efectos desta definición. A táboa está estruturada cunha listaxe inicial das partículas estables, seguida das partículas inestables ordenadas en orde descendente segundo τ. A entrada para K^0 con vida infinita mantense tal e como aparece na fonte, reflectindo un tratamento experimental específico dos kaóns neutros no contexto da definición de partículas primaria. Datos tomados de [11]. 6

1.2 Resumo de todas as seleccións consideradas para a tese. 9

1.3 Mostras de calibración consideradas nesta tese. As mostras proceden das canles de desintegración empregadas para calibrar as eficiencias de PID en LHCb [12]. 10

6.1 Width (Γ) and mean proper lifetime (τ) of long-lived particles. Particles with $\Gamma = 0$ are considered stable for the purposes of this definition. The table is structured with truly stable particles listed first, followed by unstable particles sorted by descending τ. The K^0 entry with infinite lifetime is as presented in the source, representing a specific experimental treatment of neutral kaons in the context of prompt particle definition. Data taken from [11]. 65

7.1 Minimum bias lines from stripping v20r3 which are considered in the analysis. The information was extracted from [106]. 73

7.2 Summary of the included and excluded TCK and runs for every dataset. . 74

7.3 Luminosity measurements for proton-lead datasets. Uncertainties include both statistical and systematic components. 75

7.4 Bookkeeping path to the data-set used in this analysis for both Data and MC 76

8.1 Result of the means (μ) and standard deviations (σ) of the Gaussian fits used to define the luminous region for the forward (top) and backward (bottom) configuration for both magnet polarities. Reproduced from [103]. 79

8.2 True ID values for pions, kaons, and protons in the MC simulation 85

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8.3 Selection criteria for secondary track sources in simulation: "all sec" encompasses every secondary track, "decay" specifies decay products of prompt and non-prompt particles, "$X \to \mu$" identifies particles decaying into muons, "$\gamma \to e$" denotes photon conversions and "had_int" refer to secondaries from hadronic origins. 87
8.4 Summary of all the selections considered for the analysis. 89
8.5 List of Occupancy Variables Sets used for training the `GBReweighter` tool. 97
8.6 `GBReweighter` hyperparameters used in the analysis. 97

9.1 η_{cms} bins used in this analysis for pp, pPb, and Pbp configurations. 100
9.2 Proposed binning scheme in the (p_{T}, η) plane. 101
9.3 List of requirements for the fake track proxy in each configuration. 104
9.4 Selection criteria for Secondary Tracks in simulation 109
9.5 List of requirements for the secondary track proxy in each configuration. . 113
9.6 Decay channels used to calibrate PID efficiencies in LHCb [12]. 125
9.7 Calibration samples considered in this analysis. The samples are from the decay channels used to calibrate PID efficiencies in LHCb [12]. 125
9.8 Summary of PID cuts applied to tag particles for each calibration sample, including the decay process and associated probe particles. 127
9.9 Summary of variables used to reweight the calibration sample. 127
9.10 Selection criteria for $K^{\pm}$ (tag), $K^{\mp}$ (probe), and $\phi(1020)$ (combination cuts). 145
9.11 Values of ε_{TM} for the pPb, Pbp, and pp simulation samples. Results taken from [103]. 148

10.1 Selection criteria for the alternative proxy samples used to assess systematic uncertainties in the determination of the fake-track correction factors R^{h}_{fake} in each collision configuration. 152
10.2 Models employed as baseline and alternative for the invariant mass spectrum of the calibration samples. 162

A.1 Selection criteria for $\phi(1020)$ decay into $K^{+}K^{-}$ 204

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Glossary

All acronyms and abbreviations used in this thesis are defined here. Some of these are re-defined also in the text for completeness.

SM Standard Model

LHC Large Hadron collider

QCD Quantum Chromodynamics

PDG Particle Data Group

QGP Quark-Gluon Plasma

RHIC Relativistic Heavy Ion Collider

AuAu Gold-Gold collisions

PbPb Lead-Lead

TeV Tera-electronvolt

GeV Giga-electronvolt

cms Center-of-Mass system (or frame)

IP Interaction Point

p_{T} Transverse momentum

η pseudorapidity

$\sqrt{s_{\mathrm{NN}}}$ Center-of-Mass energy

ISR Initial-State Radiation

FSR Final-State Radiation

MPI Multiple parton interactions

DPS Double Parton Scattering

UE Underlying event

SD Single Diffraction

DD Double Diffraction

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

CD Central Diffraction

pQCD Perturbative Quantum Chromodynamics

PDFs Parton Distribution Functions

NLO Next-to-Leading Order

NNLO Next-to-Next-to-Leading Order

DGLAP Dokshitzer-Gribov-Lipatov-Altarelli-Parisi

FFs Fragmentation Functions

pp Proton-proton collisions

MC Monte-Carlo

CGC Colour Glass Condensate

CNM Cold Nuclear Matter

TRUEID Monte Carlo Particle Data Group ID of the particle (e.g., |TRUEID| = 13 for a muon).

`MOTHER_ID` Monte Carlo Particle Data Group ID of the parent particle that decayed to produce the track (e.g., |`MOTHER_ID`| = 22 for a photon).

`ORIGINVERTEX_TYPE` A numerical code specifying the production process type (e.g., decay, heavy quark decay, hadronic interaction) as defined in the simulation framework.

`TRACK_Type` A numerical code classifying the reconstructed track quality (e.g., 3 for a Long Track)

w.r.t. With respect to

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

References

[1] D. J. Schwarz, *The first second of the universe*, Annalen der Physik **515** (2003) 220.

[2] LHCb collaboration, A. A. Alves, Jr. *et al.*, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[3] Pierre Auger collaboration, A. Aab *et al.*, *The Pierre Auger Cosmic Ray Observatory*, Nucl. Instrum. Meth. A **798** (2015) 172, arXiv:1502.01323.

[4] E. V. Shuryak, *Quantum Chromodynamics and the Theory of Superdense Matter*, Phys. Rept. **61** (1980) 71.

[5] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, *EPPS16: nuclear parton distributions with LHC data*, The European Physical Journal C **77** (2017) .

[6] J. W. Cronin *et al.*, *Production of hadrons with large transverse momentum at 200, 300, and 400 GeV*, Phys. Rev. D **11** (1975) 3105.

[7] LHCb collaboration, R. Aaij *et al.*, *Nuclear Modification Factor of Neutral Pions in the Forward and Backward Regions in p-Pb Collisions*, Phys. Rev. Lett. **131** (2023) 042302, arXiv:2204.10608.

[8] LHCb collaboration, R. Aaij *et al.*, *Measurement of the Nuclear Modification Factor and Prompt Charged Particle Production in $p-Pb$ and pp Collisions at $\sqrt{s_{NN}}$=5 TeV*, Phys. Rev. Lett. **128** (2022) 142004, arXiv:2108.13115.

[9] LHCb collaboration, R. Aaij *et al.*, *LHCb Detector Performance*, Int. J. Mod. Phys. A **30** (2015) 1530022, arXiv:1412.6352.

[10] *LHC Minimum Bias and Underlying Event Working GroupConveners' meeting*, http://lpcc.web.cern.ch/sites/lpcc.web.cern.ch/files/wg_docs/minutes-16-03-2015-final.pdf.

[11] ALICE Collaboration, S. Acharya *et al.*, *The ALICE definition of primary particles*, ALICE-PUBLIC-2017-005, CERN, 2017.

[12] O. Lupton, L. Anderlini, B. Sciascia, and V. Gligorov, *Calibration samples for particle identification at LHCb in Run 2* , CERN, Geneva, 2016. https://cds.cern.ch/record/2134057.

[13] T. Sjöstrand *et al.*, *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv:1410.3012.

[14] LHCb collaboration, *Nuclear modification factor of neutral pions in the forward and backward regions in pPb collisions*, arXiv:2204.10608.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[15] Belle collaboration, S. K. Choi *et al.*, *Observation of a narrow charmonium-like state in exclusive $B^{\pm} \to K^{\pm}\pi^{+}\pi^{-}J/\psi$ decays*, Phys. Rev. Lett. **91** (2003) 262001, arXiv:hep-ex/0309032.

[16] LHCb collaboration, R. Aaij *et al.*, *Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^{-} p$ Decays*, Phys. Rev. Lett. **115** (2015) 072001, arXiv:1507.03414.

[17] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. **13** (1964) 321.

[18] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132.

[19] ATLAS collaboration, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, arXiv:1207.7214.

[20] CMS collaboration, S. Chatrchyan *et al.*, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv:1207.7235.

[21] MissMJ and Pokey5190, *Standard model of elementary particles (anti-particle version)*, https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles_Anti.svg. Licensed under Creative Commons Attribution-Share Alike 3.0 Unported (CC BY-SA 3.0). Original by MissMJ, SVG by Pokey5190. Uploaded: 2006-11-14. Accessed: 2025-07-14.

[22] M. Gell-Mann, *Symmetries of baryons and mesons*, Phys. Rev. **125** (1962) 1067.

[23] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, *Advantages of the Color Octet Gluon Picture*, Phys. Lett. B **47** (1973) 365.

[24] F. Halzen and A. D. Martin, *QUARKS AND LEPTONS: AN INTRODUCTORY COURSE IN MODERN PARTICLE PHYSICS*, John Wiley and Sons, 1984.

[25] Particle Data Group, *Quantum Chromodynamics*, Review of Particle Physics (2024), Accessed: 15 October 2024.

[26] H. Satz, *The Quark-Gluon Plasma: A Short Introduction*, Nucl. Phys. A **862-863** (2011) 4, arXiv:1101.3937.

[27] K. Yagi, T. Hatsuda, and Y. Miake, *Quark-gluon plasma: From big bang to little bang*, vol. 23, Cambridge University Press, 2005.

[28] D. H. Rischke, *The Quark gluon plasma in equilibrium*, Prog. Part. Nucl. Phys. **52** (2004) 197, arXiv:nucl-th/0305030.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[29] P. Braun-Munzinger and J. Stachel, *The quest for the quark-gluon plasma*, Nature **448** (2007) 302.

[30] F. Weber, *Strange quark matter and compact stars*, Prog. Part. Nucl. Phys. **54** (2005) 193, arXiv:astro-ph/0407155.

[31] G. Baym, *THERMAL EQUILIBRATION IN ULTRARELATIVISTIC HEAVY ION COLLISIONS*, Phys. Lett. B **138** (1984) 18.

[32] S. Sumowidagdo, *Experimental study of matter under the influence non-abelian gauge theory*, J. Phys. Conf. Ser. **856** (2017) 012002.

[33] M. Harrison, T. Ludlam, and S. Ozaki, *RHIC project overview*, Nucl. Instrum. Meth. A **499** (2003) 235.

[34] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[35] M. Thomson, *Modern Particle Physics*, Cambridge University Press, 2013.

[36] V. Barger *et al.*, *Physics of the Top Quark and its Interactions*, World Scientific, 2012.

[37] TikZ, *3d axis for cms*, 2025. https://tikz.net/axis3d_cms/.

[38] S. Donnachie, H. G. Dosch, O. Nachtmann, and P. Landshoff, *Pomeron physics and QCD*, vol. 19, Cambridge University Press, 2004.

[39] Y. L. Dokshitzer, V. A. Khoze, A. H. Mueller, and S. I. Troian, *Basics of perturbative QCD*, Editions Frontières, 1991.

[40] B. Andersson, G. Gustafson, and B. Soderberg, *A General Model for Jet Fragmentation*, Z. Phys. C **20** (1983) 317.

[41] M. Diehl and A. Schafer, *Theoretical considerations on multiparton interactions in QCD*, Phys. Lett. B **698** (2011) 389, arXiv:1102.3081.

[42] T. Sjostrand and P. Z. Skands, *Multiple interactions and the structure of beam remnants*, JHEP **03** (2004) 053, arXiv:hep-ph/0402078.

[43] L. Schoeffel, *Diffractive physics*, Nucl. Instrum. Meth. A **423** (1999) 439.

[44] J. Collins, *Foundations of perturbative QCD*, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **32** (2011).

[45] J. C. Collins, D. E. Soper, and G. Sterman, *Factorization of Hard Processes in QCD*, Adv. Ser. Direct. High Energy Phys. **5** (1989) 1, [Frontiers in Physics: Volume 5, Perturbative Quantum Chromodynamics].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[46] F. Jonas, *Probing the initial state of heavy-ion collisions with isolated prompt photons*, PhD thesis, Westfälische Wilhelms-Universität Münster, 2023, Presented 11 Aug 2023.

[47] J. Gao, L. Harland-Lang, and J. Rojo, *The Structure of the Proton in the LHC Precision Era*, Phys. Rept. **742** (2018) 1, arXiv:1709.04922.

[48] T.-J. Hou *et al.*, *New cteq global analysis of quantum chromodynamics with high-precision data from the lhc*, Phys. Rev. D **103** (2021) 014013.

[49] I. Helenius, K. J. Eskola, and H. Paukkunen, *Probing the small-x nuclear gluon distributions with isolated photons at forward rapidities in p+Pb collisions at the LHC*, JHEP **09** (2014) 138, arXiv:1406.1689.

[50] M. Constantinou *et al.*, *Lattice QCD Calculations of Parton Physics*, arXiv:2202.07193.

[51] CMS collaboration, V. Khachatryan *et al.*, *Measurement of diffraction dissociation cross sections in pp collisions at $\sqrt{s}$ = 7 TeV*, Phys. Rev. D **92** (2015) 012003, arXiv:1503.08689.

[52] PHENIX collaboration, K. Adcox *et al.*, *Suppression of hadrons with large transverse momentum in central Au+Au collisions at $\sqrt{s_{NN}}$ = 130-GeV*, Phys. Rev. Lett. **88** (2002) 022301, arXiv:nucl-ex/0109003.

[53] NA50 collaboration, M. C. Abreu *et al.*, *Observation of a threshold effect in the anomalous J/ psi suppression*, Phys. Lett. B **450** (1999) 456.

[54] STAR collaboration, K. H. Ackermann *et al.*, *Elliptic flow in Au + Au collisions at (S(NN))**(1/2) = 130 GeV*, Phys. Rev. Lett. **86** (2001) 402, arXiv:nucl-ex/0009011.

[55] I. Sick and D. Day, *The EMC effect of nuclear matter*, Phys. Lett. B **274** (1992) 16.

[56] P. Paakkinen, *Pdfs in nuclear collisions - overview of recent results*, Presentation at the 10th Annual Large Hadron Collider Physics (LHCP) Conference, 2022. Accessed: 2025-01-24.

[57] J. Jalilian-Marian, A. Kovner, A. Leonidov, and H. Weigert, *The Wilson renormalization group for low x physics: Towards the high density regime*, Phys. Rev. D **59** (1998) 014014, arXiv:hep-ph/9706377.

[58] F. Gelis, E. Iancu, J. Jalilian-Marian, and R. Venugopalan, *The Color Glass Condensate*, Ann. Rev. Nucl. Part. Sci. **60** (2010) 463, arXiv:1002.0333.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[59] J. Jalilian-Marian, A. Kovner, L. McLerran, and H. Weigert, *The Intricate structure of color glass condensate at small x*, Phys. Rev. D **55** (1997) 5414, arXiv:hep-ph/9606337.

[60] GEANT4 collaboration, S. Agostinelli *et al.*, *GEANT4–a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.

[61] Pierre Auger collaboration, A. Aab *et al.*, *Measurement of the Fluctuations in the Number of Muons in Extensive Air Showers with the Pierre Auger Observatory*, Phys. Rev. Lett. **126** (2021) 152002, arXiv:2102.07797.

[62] A. Buckley *et al.*, *Rivet user manual*, Comput. Phys. Commun. **184** (2013) 2803, arXiv:1003.0694.

[63] S. Biderman *et al.*, *Pythia: A suite for analyzing large language models across training and scaling*, 2023. https://arxiv.org/abs/2304.01373.

[64] T. Pierog *et al.*, *EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider*, Phys. Rev. **C92** (2015) 034906, arXiv:1306.0121.

[65] C. Bierlich *et al.*, *A comprehensive guide to the physics and usage of PYTHIA 8.3*, SciPost Phys. Codeb. **2022** (2022) 8, arXiv:2203.11601.

[66] K. Werner, *The hadronic interaction model EPOS*, Nucl. Phys. B Proc. Suppl. **175-176** (2008) 81.

[67] L. Rossi, *The LHC main dipoles and quadrupoles toward series production*, in *Applied Superconductivity Conference*, 2002.

[68] E. Koukovini Platia *et al.*, *Protons Beyond the LHC Injectors Upgrade Project*, .

[69] E. Lopienska, *The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022*, , https://cds.cern.ch/record/2800984.

[70] ATLAS collaboration, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[71] CMS collaboration, S. Chatrchyan *et al.*, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004.

[72] ALICE collaboration, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[73] ATLAS, CMS collaboration, A. G. Clark, *Experimental and Physics Prospects at ATLAS and CMS - 2011 and Beyond*, in *21st Hadron Collider Physics Symposium*, 2010, arXiv:1011.6572.

[74] LHCb collaboration, O. Steinkamp, *LHCb physics, performance, prospects*, in *7th Patras Workshop on Axions, WIMPs and WISPs*, 169–173, 2011.

[75] LHCb collaboration, G. Manca, *Prospects for Heavy Ion Physics with LHCb*, Nucl. Phys. A **956** (2016) 240.

[76] R. Lindner, *LHCb layout . LHCb schema 2*, LHCb Collection., 2008.

[77] LHCb collaboration, V. Vagnoni, *Heavy ion physics at LHCb Upgrade II*, arXiv:2503.23093.

[78] LHCb collaboration, V. Vagnoni and C. Marin Benito, *Computing and software for LHCb Upgrade II*, arXiv:2503.24106.

[79] LHCb collaboration, V. Vagnoni, *Technology developments for LHCb Upgrade II*, arXiv:2504.03088.

[80] LHCb collaboration, V. Vagnoni, *Discovery potential of LHCb Upgrade II*, arXiv:2503.23087.

[81] LHCb, *LHCb VELO TDR: Vertex locator. Technical design report*, .

[82] M. Kucharczyk, P. Morawski, and M. Witek, *Primary Vertex Reconstruction at LHCb* , CERN, Geneva, 2014. https://cds.cern.ch/record/1756296.

[83] LHCb collaboration, S. Amato *et al.*, *LHCb magnet*, Technical design report. LHCb, CERN, Geneva, 2000.

[84] LHCb collaboration, P. R. Barbosa-Marinho *et al.*, *LHCb inner tracker*, Technical design report. LHCb, CERN, Geneva, 2002. https://cds.cern.ch/record/582793.

[85] LHCb collaboration, P. R. Barbosa-Marinho *et al.*, *LHCb outer tracker*, Technical design report. LHCb, CERN, Geneva, 2001. https://cds.cern.ch/record/519146.

[86] M. Needham, *Performance of the LHCb Track Reconstruction Software* , CERN, Geneva, 2008. https://cds.cern.ch/record/1080556.

[87] O. Callot and S. Hansmann-Menzemer, *The Forward Tracking* , CERN, Geneva, 2007. https://cds.cern.ch/record/1033584.

[88] A. Powell, *Particle Identification at LHCb. Particle ID in LHCb*, , https://cds.cern.ch/record/1322666.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[89] LHCbRICHgroup, M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. C **73** (2013) 2431, arXiv:1211.6759, https://cds.cern.ch/record/1495721.

[90] LHCb, *LHCb calorimeters: Technical design report*, , https://cds.cern.ch/record/494264,.

[91] LHCb, *LHCb muon system technical design report*, , https://cds.cern.ch/record/504326.

[92] F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, JINST **8** (2013) P10020, arXiv:1306.0249.

[93] T. Head, *The LHCb trigger system*, JINST **9** (2014) C09015.

[94] I. Belyaev *et al.*, *Simulation application for the LHCb experiment*, eConf **C0303241** (2003) TUMT003, arXiv:physics/0306035.

[95] R. Aaij *et al.*, *Allen: A high level trigger on GPUs for LHCb*, Comput. Softw. Big Sci. **4** (2020) 7, arXiv:1912.09161.

[96] LHCb Starterkit Collaboration, *First analysis steps: Data flow*, https://lhcb.github.io/starterkit-lessons/first-analysis-steps/dataflow.html. LHCb Starterkit Lessons. Accessed: 2025-07-14.

[97] M. Schmelling, J. He, Z. Yang, and F. Jing, *First Look at the pPb pilot run*, , https://cds.cern.ch/record/1496927.

[98] LHCb collaboration, T. Li, *Probing the valence quark region of nucleons with Z bosons at LHCb*, Acta Phys. Pol. B Proc. Suppl **16** (2023) 126, arXiv:2210.00783, 6 pages, 7 figures, Quark Matter 2022 Proceedings, https://indi.to/5XBNH.

[99] ALICE collaboration, J. Adam *et al.*, *Multiplicity dependence of charged pion, kaon, and (anti)proton production at large transverse momentum in p-Pb collisions at* $\sqrt{s_{\rm NN}}$ *= 5.02 TeV*, Phys. Lett. B **760** (2016) 720, arXiv:1601.03658.

[100] K. J. Eskola, I. Helenius, P. Paakkinen, and H. Paukkunen, *A QCD analysis of LHCb D-meson data in p+Pb collisions*, JHEP **05** (2020) 037, arXiv:1906.02512.

[101] T. Lappi and H. Mäntysaari, *Single inclusive particle production at high energy from HERA data to proton-nucleus collisions*, Phys. Rev. D **88** (2013) 114020, arXiv:1309.6963.

[102] LHCb collaboration, R. Aaij *et al.*, *Precision luminosity measurements at LHCb*, JINST **9** (2014) P12005, arXiv:1410.0149.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[103] O. B. G. et al. *Measurement of the nuclear modification factor and prompt charged particle production in pPb, Pbp and pp collisions at* $\sqrt{s_{NN}} = 5\ TeV$, PhD thesis, Universidade de Santiago de Compostela, 2020, https://minerva.usc.gal/entities/publication/1c15f4c2-ffc4-4f7b-a83b-846fb42290a9.

[104] Y. Gao, *Study of* Υ *production and cold nuclear matter effects in pPb collisions at* $\sqrt{s}_{NN} = 5TeV$, LHCb-ANA-2013-074, 2014.

[105] Y. Gao, *Study of the* $\psi(2S)$ *production and cold nuclear matter effects in pPb collisions at* $\sqrt{s}_{NN} = 5TeV$, LHCb-ANA-2015-047, 2016.

[106] *Stripping project webpage*, http://lhcbdoc.web.cern.ch/lhcbdoc/stripping/. Accessed: 2018-12-20.

[107] LHCb Collaboration, *Discussion on minimum bias events definition*, https://github.com/lhcb/glossary/issues/30, 2019.

[108] C. Bierlich, *Microscopic collectivity: The ridge and strangeness enhancement from string–string interactions*, Nucl. Phys. A **982** (2019) 499, arXiv:1807.05271.

[109] *Lhcb luminosity twiki*, https://twiki.cern.ch/twiki/bin/viewauth/LHCbPhysics/LHCbLuminosity.

[110] O. Boente Garcia, *Analysis of charged particle production in proton-nucleus and proton-proton collisions at the LHCb experiment*, phd thesis, University of Santiago de Compostela, 2021.

[111] M. Needham, *Identification of Ghost Tracks using a Likelihood Method* , CERN, Geneva, 2008. https://cds.cern.ch/record/1107564.

[112] P. I. H. Adrian Perieanu, *A Fast Algorithm to Identify and Remove Clone Tracks*, LHCb-2008-020, 2008.

[113] LHCb collaboration, R. Aaij *et al.*, *Measurement of prompt charged-particle production in pp collisions at* $\sqrt{\text{s}}$ *= 13 TeV*, JHEP **01** (2022) 166, arXiv:2107.10090.

[114] A. Rogozhnikov, *Reweighting with Boosted Decision Trees*, J. Phys. Conf. Ser. **762** (2016) 012036, arXiv:1608.05806.

[115] A. Rogozhnikov, *hep_ml python library*, https://arogozhnikov.github.io/hep_ml/reweight.html.

[116] R. Calabrese *et al.*, *Performance of the LHCb RICH detectors during LHC Run 2*, JINST **17** (2022) P07013, arXiv:2205.13400.

[117] L. Anderlini *et al.*, *The PIDCalib package* , CERN, Geneva, 2016. https://cds.cern.ch/record/2202412.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[118] R. Aaij *et al.*, *Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2*, EPJ Tech. Instrum. **6** (2019) 1, arXiv:1803.00824.

[119] M. De Cian *et al.*, *Measurement of the track finding efficiency* , CERN, Geneva, 2012. https://cds.cern.ch/record/1402577.

[120] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015) P02007, arXiv:1408.1251.

[121] TrackCalib, *Trackingeffstatus*2012*s*20, 2013. https://twiki.cern.ch/twiki/bin/view/LHCbInternal/TrackingEffStatus2012S20.

[122] TrackCalib, *Trackingeffstatus*20155*tev_sim*09*b*, 2017. https://twiki.cern.ch/twiki/bin/view/LHCbInternal/TrackingEffStatus20155TeV_Sim09b.

[123] P. Aurenche *et al.*, *Large p(T) inclusive pi0 cross-sections and next-to-leading-order QCD predictions*, Eur. Phys. J. C **13** (2000) 347, arXiv:hep-ph/9910252.

[124] D. d'Enterria, K. J. Eskola, I. Helenius, and H. Paukkunen, *Confronting current NLO parton fragmentation functions with inclusive charged-particle spectra at hadron colliders*, Nucl. Phys. B **883** (2014) 615, arXiv:1311.1415.

[125] T.-J. Hou *et al.*, *New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC*, Phys. Rev. D **103** (2021) 014013, arXiv:1912.10053.

[126] K. J. Eskola *et al.*, *EPPS21: a global QCD analysis of nuclear PDFs*, Eur. Phys. J. C **82** (2022) 413, arXiv:2112.12462.

[127] S. Kretzer, *Fragmentation functions from flavor-inclusive and flavor-tagged e+ e- annihilations*, Phys. Rev. D **62** (2000) 054001, arXiv:hep-ph/0003177.

[128] ALICE collaboration, J. Adam *et al.*, *Centrality dependence of the nuclear modification factor of charged pions, kaons, and protons in Pb-Pb collisions at* $\sqrt{s_{\rm NN}} = 2.76$ *TeV*, Phys. Rev. C **93** (2016) 034913, arXiv:1506.07287.

[129] J. Podolanski and R. Armenteros, *Iii. analysis of v-events*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **45** (1954) 13, arXiv:https://doi.org/10.1080/14786440108520416.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA





This thesis presents measurements of the production and nuclear modification factor of charged pions, kaons, and protons in high-energy proton–proton and proton–lead collisions with the LHCb experiment at CERN. The results provide new constraints on the mass dependence of the Cronin enhancement in the nuclear modification factor and improve the understanding of gluon fragmentation functions. These measurements are also essential for refining Monte Carlo models, studying soft interactions, and isolating Cold Nuclear Matter (CNM) effects in pPb collisions, allowing them to be distinguished from Quark–Gluon Plasma signatures in PbPb collisions. The unique forward and backward kinematic coverage of LHCb enables the study of CNM effects across a broad range of Bjorken-x.