

Single and double charm production in pp collisions at $\sqrt{s} = 13.6$ TeV with ALICE

*Mesure de la production de simple et double charme en collisions pp
à $\sqrt{s} = 13.6$ TeV avec ALICE*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 576 : Particules, hadrons, énergie et noyau: instrumentation, imagerie,
cosmos et simulation (PHENIICS)

Spécialité de doctorat: Physique nucléaire

Graduate School : Physique. Référent : Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche **IJCLab** (Université Paris-Saclay, CNRS), sous la direction
de **Zaida CONESA DEL VALLE**, Chargée de recherche CNRS, et le co-encadrement de **Bruno**
ESPAGNON, Professeur de l'Université Paris-Saclay

Thèse soutenue à Paris-Saclay, le 30 Septembre 2025, par

Andrea TAVIRA GARCÍA

Composition du jury

Membres du jury avec voix délibérative

Javier CASTILLO CASTELLANOS Directeur de recherche (HDR), Paris-Saclay U. (CEA)	Président
Francesco PRINO Chercheur sénior (équivalent HDR), INFN Torino	Rapporteur & Examineur
Cibrán SANTAMARINA RÍOS Professeur, Santiago de Compostela U. (IGFAE)	Rapporteur & Examineur
David D'ENTERRIA Chercheur sénior, CERN	Examineur
Silvia MASCIOCCHI Professeure, Heidelberg U. (GSI)	Examinatrice

Titre: Mesure de la production de simple et double charme en collisions pp à $\sqrt{s} = 13.6$ TeV avec ALICE

Mots clés: Charme, LHC, ALICE, plasma de quarks et de gluons, interactions multiples de partons, systèmes petits

Résumé: L'expérience ALICE au grand collisionneur de hadrons (LHC) est consacrée à l'étude du plasma quarks-gluons (QGP), un état de la matière dans laquelle les quarks et les gluons, les constituants fondamentaux de la matière nucléaire, sont déconfinés. La chromodynamique quantique (QCD) prédit une transition de phase entre la matière nucléaire ordinaire et le QGP à une densité d'énergie d'environ 1-2 GeV/fm³ et une température d'environ 200 MeV. On pense que l'Univers a connu une telle phase quelques microsecondes après le Big Bang. Les collisions ultra-relativistes d'ions lourds offrent une occasion unique de recréer en laboratoire les conditions extrêmes nécessaires à la formation du QGP. Le milieu qui en résulte se comporte comme un liquide quasi-parfait composé de partons fortement couplés.

Parmi les observables clés utilisées pour sonder le QGP figurent les sondes dures (angl. "hard probes"), qui proviennent des premières étapes de la collision et traversent tout le milieu, offrant ainsi un aperçu de son évolution. Cette thèse se concentre sur la mesure de la QCD perturbative (pQCD), qui sert de référence pour l'interprétation des résultats des collisions Pb-Pb. En particulier, on a étudié la production de mésons D⁰, qui ont été sélectionnés à l'aide de techniques d'apprentissage automatique et en utilisant les données du Run 3 de l'expérience ALICE. Cet ensemble de données est beaucoup plus grand et comprend des informations plus complètes sur les détecteurs par rapport aux runs précédents.

La thèse explore la production de quarks beaux dans les collisions hadroniques ultra-relativistes, ce qui constitue un test fondamental de la pQCD et des modèles d'hadronisation. Une approche pour étudier les quarks beaux consiste à mesurer les mésons D non-prompt, qui proviennent de la dés-

intégration des hadrons B. La fraction de mésons D⁰ non-prompt est mesurée dans des collisions pp à $\sqrt{s} = 13,6$ TeV. Les résultats préliminaires de la fraction de D non-prompt sont comparés aux prédictions théoriques, qui ne reproduisent pas complètement les données expérimentales.

L'impact des interactions de parties multiples (MPI), qui sont composés d'un quark charmé et d'un quark léger, sur la double production de mésons D⁰, qui sont composés d'un quark charmant et d'un quark léger, est aussi exploré. En mesurant à la fois leurs taux de production et leurs corrélations angulaires, il est possible de mieux comprendre les mécanismes de production sous-jacents. Les MPIs sont une explication possible de l'existence de signatures similaires au QGP qui ont été trouvées dans les collisions proton-proton et proton-noyau à haute multiplicité. Ces signatures comprennent un renforcement de la production des hadrons étranges avec la multiplicité et l'existence de coefficients d'anisotropie azimuthale non nuls. Notre analyse comprend l'étude et la comparaison de la production de deux mésons D⁰ de même signe (principalement produits via des MPI) et de deux mésons D⁰ de signe opposé (principalement produits dans une seule collision parton-parton). Les sections efficaces intégrées et différentielles des mésons D⁰ doubles sont mesurées pour la première fois en ALICE, et le rapport entre les mésons de même signe et de signe opposé est calculé.

En outre, la thèse présente des études préliminaires sur l'efficacité du suivi des muons dans les chambres à muons pour les données Pb-Pb collectées en 2023. Cette étude est essentielle pour comprendre les performances des chambres à muons dans le Run 3 et pour garantir l'exactitude des mesures impliquant des muons dans les collisions d'ions lourds.

Title: Single and double charm production in pp collisions at $\sqrt{s} = 13.6$ TeV with ALICE

Keywords: Charm, LHC, ALICE, quark-gluon plasma, multiple parton interactions, small systems

Abstract: The ALICE experiment at the Large Hadron Collider (LHC) is dedicated to the study of the Quark-Gluon Plasma (QGP), a state of matter in which quarks and gluons, the fundamental constituents of nuclear matter, are deconfined. Quantum Chromodynamics (QCD) predicts a phase transition from ordinary nuclear matter to the QGP at an energy density of approximately 1-2 GeV/fm³ and a temperature of around 200 MeV. It is believed that the Universe underwent such a phase a few microseconds after the Big Bang. Ultra-relativistic heavy-ion collisions provide a unique opportunity to recreate in the laboratory the extreme conditions necessary to form the QGP. The resulting medium behaves as a quasi-perfect liquid composed of strongly coupled partons.

Among the key observables used to probe the QGP are hard probes, which originate in the early stages of the collision and traverse the entire medium, thereby offering insight into its evolution. This thesis focuses on measuring fundamental perturbative QCD (pQCD) processes, which serve as a reference for interpreting results from Pb-Pb collisions. All D-meson candidates used in this thesis are selected using Machine Learning techniques and using Run 3 data from the ALICE experiment. This dataset is significantly larger and includes more comprehensive sub-detector information compared to previous runs.

The thesis explores beauty-hadron production in ultra-relativistic hadronic collisions, which is a fundamental test of pQCD and hadronisation models. One approach to studying beauty quarks

involves measuring non-prompt D mesons, which originate from the decays of B hadrons. The fraction of non-prompt D⁰ mesons is measured in pp collisions at $\sqrt{s} = 13.6$ TeV. The preliminary results of the non-prompt D⁰ fraction are compared to theoretical predictions, which do not fully reproduce the experimental data.

The impact of multiple parton interactions (MPI) on the production of double D⁰ mesons, composed of a charm and a light quark, is also investigated. MPIs are one possible explanation for the existence of QGP-like signatures that have been found in high-multiplicity proton-proton and proton-nucleus collisions. These signatures include an enhancement of the strange hadron production with multiplicity and the existence of non-zero azimuthal anisotropy coefficients. Our analysis includes the study and comparison of the production of two like-sign D⁰ mesons (mainly produced via MPIs) and two opposite-sign D⁰ mesons (mainly produced in a single parton-parton collision). The integrated and transverse-momentum-differential cross sections of the double D⁰ mesons are measured for the first time in ALICE, and the ratio of like-sign to opposite-sign D⁰ mesons is calculated.

In addition, the thesis presents preliminary studies on the muon tracking efficiency in the muon chambers for the Pb-Pb data collected in 2023. These studies are essential for understanding the performance of the muon chambers in Run 3 and for ensuring the accuracy of the measurements involving muons in heavy-ion collisions.

ACKNOWLEDGEMENTS

There are lots of people that have helped me and supported me throughout these three years. It has truly been a huge journey that I could not have done without you. I will try to mention all of you, but I apologise if I forgot someone.

First and foremost, I would like to thank my supervisor, Zaida. Thank you so much for your guidance, your support, and your help accross these three years. And thank you to you and to Bruno for reading and re-reading the thesis drafts a couple thousand times!

A special gratitude also goes to Francesco Prino, Cibrán Santamarina, Javier Castillo, David d'Enterria and Silvia Masciocchi for accepting to be a part of my thesis jury.

Thank you so much to the full ALICE group at IJCLab, who made me feel welcome all these three years: Christophe, Laure, Cynthia, Theraa, Corentin, Afnan, Dukhishyam, Lucía, and Olga. A special thanks goes to Sara, who became a great friend. This past year would have not been the same without you in the office. A big thank you goes to the people from the theory pole, especially Jean-Philippe and Allencris, for accepting to work on the theoretical input for my analysis.

A big thanks to all the people from ALICE who helped me throughout these years. Thanks to the HF convenors, Fabio Colamaria and Fabrizio Grosa, for your huge help. Thank you to Marianna Mazzilli and Mingyu Zhang for the times we worked together. A big thanks goes also to the D2H and HFCL people that have helped me improve my analyses and my physics knowledge in general. Thank you all!

To all the people I met in France, both at the lab and Cité. To Raoul, Allencris (again), Ragansu, Dimitris, Piera, Tomas, Sebastián, Daniel, Lucas, Shu. To Jenny, Guglielmo, Greta, and Marcello from the MDI, and to the people at the Colegio de España that made me feel a bit more at home from far away.

Thank you to the amazing people I met from all accross the world at CERN and at conferences. A special thanks to the IGFAE group who made me feel like one more of them from day one in every conference I have been to. Thanks to the ALICE junior group that made my stay at CERN a lot more fun, especially to Maxim, who also made me visit every single Pokemon card store in Japan.

To all my friends back in Madrid. A huge thanks to my friends from Sanma: Raquel, JP, Guille, Íñigo, Álvaro, Irene. Sois los mejores amigos que podría pedir, mil mil gracias por haberme acompañado en mis aventuras por Francia estos tres años. To Paula, my friend since we were 10. To my uni friends for all the fun times and the videocalls accross countries.

To Marquinho, muito muito obrigada pelo teu apoio e carinho. És o melhor. To my family, my grandparents and my parents. No podría haber hecho esto sin vosotros.

A todos vosotros, gracias. Merci. Thank you.

CONTENTS

Contents	vii
General Introduction	1
Synthese en français	3
1 Introduction	5
1.1 The Standard Model of Particle Physics	5
1.2 Quantum chromodynamics (QCD)	7
1.2.1 Confinement and asymptotic freedom	7
1.2.2 Chiral symmetry breaking	9
1.3 The quark-gluon plasma	9
1.3.1 The QCD phase diagram	10
1.3.2 Evolution of a heavy-ion collision in the Bjorken scenario	12
1.4 Collision and particle kinematic properties	13
1.5 Experimental probes to study heavy-ion collisions	14
1.5.1 Initial state probes	15
1.5.2 Final state probes	16
1.6 Studies in small systems	20
1.6.1 Using pp collisions as a baseline: the nuclear modification factor	20
1.6.2 QGP-like behaviours in small systems	21
2 Open heavy-flavour production	25
2.1 Charm and beauty production	25
2.1.1 Parton Distribution Functions	27
2.1.2 Parton scattering cross section	29
2.1.3 Fragmentation functions and hadronisation	30
2.1.4 PYTHIA	32
2.1.5 EPOS 4	34
2.1.6 State-of-the-art of the hadron production cross sections	34
2.2 Multiple parton interactions	37
2.3 Cold nuclear matter effects in heavy-ion collisions	42
2.4 Hot nuclear matter effects in heavy-ion collisions	44

2.4.1	Collisional processes	44
2.4.2	Radiative processes	46
2.4.3	Heavy-flavour hadronisation in Pb–Pb collisions	46
3	A Large Ion Collider Experiment – ALICE	49
3.1	The Large Hadron Collider	49
3.2	ALICE	51
3.2.1	Central barrel detectors	52
3.2.2	Forward detectors	57
3.2.3	Muon Spectrometer	59
3.3	Computing and data taking in Run 3	63
3.3.1	The First Level Processors (FLP) farm	64
3.3.2	The Event Processing Nodes (EPN) farm	65
3.3.3	Analysis framework: O^2 Physics	65
4	Muon tracking efficiency studies of Run 3 data	67
4.1	Acceptance and efficiency in the muon spectrometer	67
4.2	The muon reconstruction algorithm	68
4.3	Method to estimate the tracking efficiency	70
4.4	Tracking efficiency in Pb–Pb collisions	71
4.4.1	Muon tracking efficiency studies in data	73
4.4.2	Muon tracking efficiency studies in the initial Monte Carlo	74
4.4.3	Results using an anchored Monte Carlo	76
4.4.4	Comparison between data and MC	78
4.5	Evaluation of the tracking systematic	80
5	D^0-meson reconstruction	83
5.1	Real and simulated data samples	83
5.2	Event reconstruction and selection	84
5.2.1	Event selection	85
5.2.2	Primary vertex reconstruction	85
5.3	Track reconstruction and selection	86
5.3.1	Track selection	88
5.3.2	The track-tuner feature	88
5.4	Track-to-collision association	89
5.5	PID selections	90
5.6	The heavy-flavour workflow in O^2 Physics	90
5.7	A (very) brief introduction to Machine Learning	93
5.7.1	Decision tree classifiers (DTs)	93
5.7.2	Boosted Decision Trees (BDTs)	95
5.8	Reconstructing D^0 mesons using Machine Learning	96
5.8.1	D^0 mesons in ALICE	96
5.8.2	Main variables to reconstruct D^0 mesons	97

5.8.3	Training and test sets	103
5.8.4	Training results and model quality evaluation	104
5.8.5	Machine Learning inference	107
5.9	The Offline Trigger System (OTS)	107
5.10	Acceptance and efficiency corrections	109
6	Measurement of the non-prompt D^0 fraction at $\sqrt{s} = 13.6$ TeV	111
6.1	Data samples and MC used	111
6.2	Machine Learning based selection	112
6.2.1	Preselections	112
6.2.2	Machine Learning training	113
6.3	The cut-variation method	115
6.4	D^0 yield extraction	117
6.5	Acceptance-times-efficiency calculation	122
6.6	Extraction of the non-prompt fraction	123
6.7	Systematic uncertainties	124
6.7.1	Raw yield extraction	125
6.7.2	BDT selections	126
6.7.3	DCA resolution	126
6.7.4	p_T shape	127
6.7.5	Summary of the considered systematics	127
6.8	Results	128
7	Associated production of D^0 mesons in pp collisions at $\sqrt{s} = 13.6$ TeV	133
7.1	Data samples and MC used	133
7.2	Luminosity determination	134
7.3	Comparison of the triggered and minimum bias samples	135
7.4	Machine Learning based selection	135
7.4.1	Trigger-based preselection	136
7.4.2	Machine Learning training	136
7.4.3	Model application	139
7.5	Double D^0 yield extraction	139
7.6	Acceptance-times-efficiency corrections	143
7.7	Prompt fraction determination	147
7.8	Systematic uncertainties	150
7.8.1	Corrected yield extraction	150
7.8.2	ITS-TPC matching efficiency	150
7.8.3	Single-track selections	151
7.8.4	DCA resolution	151
7.8.5	BDT selections	152
7.8.6	Prompt fraction extraction	153
7.8.7	Luminosity and BR	153

7.8.8	Summary of the considered systematics	154
7.9	Cross section results	154
7.10	p_T -differential results	157
7.11	Towards DPS extraction	164
8	Conclusions	169
Appendices		
A	Comparison of the muon tracking efficiency in data and the anchored MC	190
B	List of runs included in 2024 dataset	195
C	Comparison of the double D^0 production results with the Preliminary results of March 2025	197
C.1	Updates in the ML models and selection criteria	197
C.2	Comparison of the integrated results	198
	List of Figures	201
	List of Tables	212

GENERAL INTRODUCTION

This thesis is dedicated to the study of open heavy-flavour production in small systems using data collected by the ALICE experiment at the Large Hadron Collider (LHC). ALICE is a general-purpose detector mainly designed to study the quark-gluon plasma (QGP), a state of matter composed of deconfined quarks and gluons. Even though this state of matter is expected to be produced in high-energy heavy-ion collisions, such as Pb–Pb, recent studies in small systems, namely proton-proton (pp) and p–Pb collisions, have shown that some QGP signatures are also present in these systems. One possible explanation for the presence of QGP-like signatures in small systems could be related to the presence of multiple parton interactions (MPI), which occur when two or more parton-parton collisions happen in a single hadronic collision. These MPIs can lead to a higher particle multiplicity and could potentially create conditions suitable for QGP formation, even in small systems.

The measurements presented in this thesis focus on the production of open charm hadrons, specifically D^0 mesons, in pp collisions at $\sqrt{s} = 13.6$ TeV. The main goal is to study the associated production of two D^0 mesons, which is expected to be enhanced by double parton scattering, the simplest form of MPI. Furthermore, the study of open charm hadrons in pp collisions is relevant for understanding the production mechanisms of heavy-flavour hadrons. In fact, the measurement of non-prompt D^0 mesons, which are produced from the decay of beauty hadrons, provides insights into the beauty production cross-section and serves as a test of perturbative quantum chromodynamics.

The structure of this manuscript is as follows:

Chapter 1 describes the physical context behind the measurements performed by ALICE. It introduces the concept of quark-gluon plasma (QGP) in the context of quantum chromodynamics (QCD). Lastly, it emphasises the relevance of measurements in small systems.

Chapter 2 is dedicated to open heavy-flavour production, remarking its importance as probes in proton-proton, proton-nucleus and nucleus-nucleus collisions. Special attention is paid to multiple parton interactions, which is the main focus of this thesis.

Chapter 3 introduces the ALICE experiment and the details of the new framework de-

veloped in Run 3, used for all analyses performed in this thesis. The main detectors introduced are the ones located in the central barrel, which are employed to reconstruct open heavy-flavour hadrons. The detectors in the muon spectrometer are also introduced, which are used to calculate the muon tracking efficiency in Chapter 4.

Chapter 5 presents the specifics of the D^0 -meson reconstruction in Run 3, from the event and track reconstruction to the Machine Learning tools used to select D^0 mesons minimising the combinatorial background. The strategies introduced in this Chapter are employed in Chapters 6 and 7, which present the measurements of the non-prompt D^0 fraction and the associated production of D^0 mesons in pp collisions at $\sqrt{s} = 13.6$ TeV, respectively.

SYNTHÈSE EN FRANÇAIS

Cette thèse est dédiée à l'étude de la production de saveurs lourdes ouvertes dans de petits systèmes, à partir de données collectées par l'expérience ALICE au Large Hadron Collider (LHC). ALICE est un détecteur polyvalent conçu pour étudier le plasma de quarks et de gluons (QGP), un état de la matière composé de quarks et de gluons déconfinés. Même s'il est attendu que cet état soit produit lors de collisions d'ions lourds, telles que Pb–Pb, des études récentes avec de petits systèmes (proton-proton et p–Pb) suggèrent que certaines signatures du QGP sont aussi présentes dans ces systèmes. Une possible explication est que ces signatures sont liées aux interactions multiples de partons (MPI), qui apparaissent quand deux ou plusieurs collisions de partons se produisent dans la même collision hadronique. Ces MPI peuvent mener à une multiplicité de particules plus élevée et pourraient potentiellement créer des conditions favorables à la formation du QGP, même dans de petits systèmes.

Les mesures présentées dans cette thèse se concentrent sur la production de hadrons charmés ouverts, en particulier les mésons D^0 , dans les collisions pp à $\sqrt{s} = 13,6$ TeV. L'objectif principal est d'étudier la production associée de deux mésons D^0 , qui devrait être renforcée par les doubles interactions de partons, la forme la plus simple d'interactions multiples de partons. De plus, l'étude des hadrons charmés ouverts dans les collisions pp est pertinente pour comprendre les mécanismes de production des hadrons de saveur lourde. En effet, la mesure des mésons D^0 non-prompt, qui sont produits par la désintégration des hadrons beaux, fournit des informations sur la section efficace de production de la beauté et sert de test à la chromodynamique quantique perturbative.

Voici comment le manuscrit est structuré :

Le premier chapitre décrit le contexte physique des mesures d'ALICE. Il introduit le concept de plasma de quarks et de gluons (QGP) dans le cadre de la chromodynamique quantique (QCD). Enfin, il souligne la pertinence des mesures avec des petits systèmes.

Le deuxième chapitre est consacré à la production de saveurs lourdes ouvertes, dont l'importance en tant que sondes dans des collisions proton-proton, proton-noyau et noyau-noyau est soulignée. Une attention particulière est portée aux interactions multiples de partons, qui constituent l'objet principal de cette thèse.

Le troisième chapitre introduit l'expérience ALICE et les détails du nouveau cadre développé pour le Run 3 et utilisé pour toutes les analyses effectuées dans cette thèse. Les principaux détecteurs présentés sont ceux situés dans la partie centrale, qui sont utilisés pour reconstruire les hadrons de saveur lourde ouverts. Les détecteurs du spectromètre à muons sont également introduits car ils sont utilisés pour calculer l'efficacité de reconstruction des muons dans le quatrième chapitre.

Le cinquième chapitre présente les spécificités de la reconstruction des mésons D^0 dans le Run 3, depuis la reconstruction des événements et des traces jusqu'aux outils d'apprentissage automatique utilisés pour sélectionner les mésons D^0 en minimisant le fond combinatoire. Les stratégies introduites dans ce chapitre sont employées dans les chapitres 6 et 7, qui présentent respectivement les mesures de la fraction de D^0 non-prompt et de la production associée de mésons D^0 dans les collisions pp à $\sqrt{s} = 13,6$ TeV.

INTRODUCTION

The objective of this Chapter is to introduce the context and motivations underlying the work presented in this thesis. It begins with a brief overview of the Standard Model, which describes the interactions among the fundamental building blocks of matter. Particular emphasis is placed on quantum chromodynamics (QCD), as it provides the framework necessary to understand the quark-gluon plasma (QGP), a state of matter that emerges under extreme conditions of temperature and density. Finally, the role of proton-proton (pp) collisions as a crucial reference in the study of the QGP is highlighted.

1.1 The Standard Model of Particle Physics

Out of the four fundamental forces present in nature, the Standard Model (SM) of particle physics can describe three of them [1]. It describes the electromagnetic, strong and weak forces as interactions between elementary particles. The elementary particles included in the SM can be consulted in Fig. 1.1. They are characterised by their masses and a set of quantum numbers, such as spin (S), electric charge (Q), baryonic number (B), leptonic number (L), etc. Depending on their spin, it is possible to separate particles into two groups: **fermions**, which have a half-integer spin, and **bosons**, whose spin is an integer.

Fermions are a set of particles whose half-integer spin makes them follow Fermi-Dirac statistics. They can be classified into two groups: **quarks** and **leptons**, depending on whether or not they undergo strong interactions. Each of these groups is composed of three generations of particles, following a mass hierarchy from lower to higher masses. The three generations of quarks are (u, d), (s, c) and (b, t). They all have a fractional electric charge of $-1/3$ or $2/3$, $B \neq 0$, $L = 0$, and colour charge. The leptonic family is composed of three negatively charged particles and three associated neutral left-handed neutrinos, organised as (e^-, ν_e) , (μ^-, ν_μ) and (τ^-, ν_τ) . They are characterised by having $B = 0$, $L \neq 0$, and no colour charge. Each fermion, lepton or quark, has an associated anti-particle of opposite charge and, in the case of neutrinos, helicity¹.

¹ The helicity is the projection of the spin onto the momentum direction.

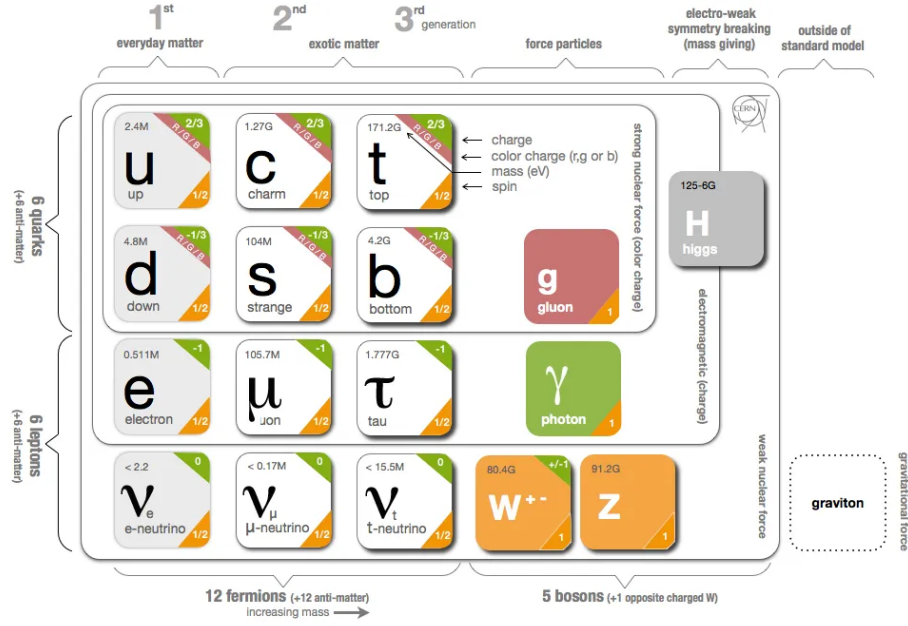


Figure 1.1: Particles described by the Standard Model. Image from the CERN Webfest 2012 [2].

Only the first generation of quarks and leptons participates in the generation of ordinary matter, as their higher-mass counterparts are highly unstable. Quarks tend to be bound in colourless hadrons of integer electric charge. These hadrons can be classified depending on the number of constituent (anti)quarks. If they are composed of a quark and an antiquark, they are called **mesons**, while hadrons that contain three quarks are called **baryons**. Hadrons composed of more than three quarks were recently discovered, such as **tetraquarks**, found for the first time at the Belle experiment [3], or **pentaquarks**, observed by the LHCb experiment [4].

Formally speaking, the SM can be represented as a Quantum Field Theory (QFT) based on a gauge symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$. The interaction forces related to the SM are the following:

- **The electromagnetic interaction:** it corresponds to the gauge group $U(1)_Y$, and describes the interaction between electrically charged particles. It is mediated by photons (γ), which are massless and neutral in charge. The associated QFT is Quantum Electrodynamics (QED).
- **The weak interaction:** it is mediated by three vector gauge bosons, two with electric charge ($W^\pm$, $M_W \sim 80 \text{ GeV}/c^2$), and one neutral (Z^0 , $M_Z \sim 90 \text{ GeV}/c^2$). The weak interaction generates changes in flavour. It is unified with the electromagnetic interaction to create the electroweak theory, which is based on a gauge symmetry $SU(2)_L \times U(1)_Y$.
- **The strong interaction:** it is mediated by gluons (g) and occurs between particles with colour charge. It is described by quantum chromodynamics using a non-Abelian $SU(3)_C$ group. In this theory, gluons also carry a combination of colour and anti-colour charge.

In addition to the interactions mentioned above, all massive particles are subject to the **gravitational interaction**, although the SM formalism does not describe it. In the SM, particles gain mass by interacting with the Higgs field [5]. The associated particle is the Higgs boson, H , which was discovered in 2012 at the LHC by the ATLAS [6] and CMS [7] collaborations.

1.2 Quantum chromodynamics (QCD)

Quantum chromodynamics describes the strong interactions of coloured quarks and gluons with an $SU(3)_C$ gauge symmetry. Its Lagrangian is given by

$$\mathcal{L} = \sum_q \bar{\psi}_{q,a} (i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t_{ab}^C \mathcal{A}_\mu^C - m_q \delta_{ab}) \psi_{q,b} - \frac{1}{4} F_{\mu\nu}^A F^{A\mu\nu}, \quad (1.1)$$

using Einstein notation of repeated indices. In this equation, γ^μ represent the Dirac γ -matrices, g_s is the QCD coupling constant, and $\psi_{q,a}$ are the quark-field spinors for a quark of flavour q and mass m_q , with a colour index a , running from $a = 1$ to the total number of colours, N_C . The total number of colours is set to three: *red*, *blue* and *green*. The $\mathcal{A}_\mu^C$ represent the gluon fields and correspond to the adjoint representation of $SU(3)$. In this case, C runs from one to $N_C^2 - 1 = 8$. The t_{ab} matrices are the $SU(3)$ group generators and are used to describe the rotation of the colour of the quarks as they interact with gluons. Lastly, the field tensor $F_{\mu\nu}^A$ is defined as:

$$F_{\mu\nu}^A = \partial_\mu \mathcal{A}_\nu^A - \partial_\nu \mathcal{A}_\mu^A - g_s f_{ABC} \mathcal{A}_\mu^B \mathcal{A}_\nu^C, \quad (1.2)$$

with

$$[t^A, t^B] = i f_{ABC} t^C,$$

and with f_{ABC} being the structure constants of the $SU(3)$ group. An extensive description of the theory and its formalism can be found, for instance, in [8]. Nonetheless, this text will mainly focus on some of its most relevant consequences, such as the confinement and asymptotic freedom or the breaking of the chiral symmetry at low energy.

1.2.1 Confinement and asymptotic freedom

The fact that QCD is a non-Abelian theory has implications for how charges are screened as they propagate. In QFT theories, particles can create loops of virtual pairs of particles and antiparticles in the vacuum. For instance, in QED, an electron can emit a photon, which can fluctuate into an electron-positron pair (see the left image of Fig. 1.2 for an equivalent case in QCD), and so on. Hence, this electron is not isolated, but rather it is surrounded by e^-e^+ pairs, and because of their charges, positrons will tend to be preferably closer to the electron. Thus, the negative charge of the electron is **screened** by the surrounding particles.

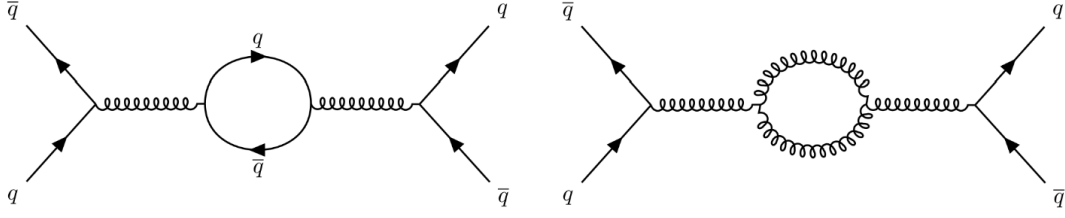


Figure 1.2: Vacuum polarisation Feynman diagrams. On the left, diagram contributing to the screening, and on the right, to anti-screening.

However, to perform the same thought experiment in QCD, it is necessary to consider that gluons also carry charge, which makes them capable of coupling not just with quarks but also with themselves (Fig. 1.2, right). As a consequence, the loops of virtual particles that get created can also enhance the effective charge of the initial particle. This phenomenon is called **anti-screening** and is, in fact, the dominant phenomenon in QCD.

The dominance of the anti-screening can be observed by analysing the QCD coupling constant, parametrised at leading order as:

$$\alpha_s = \frac{g_s^2}{4\pi} = \frac{\alpha_s(\mu^2)}{1 + \alpha_s(\mu^2) \left(\frac{33-2n_f}{12\pi} \right) \ln \left(\frac{Q^2}{\mu^2} \right)}. \quad (1.3)$$

This definition takes as parameters the number of flavours, n_f , the renormalisation mass scale, μ , and the momentum transfer squared, Q^2 .

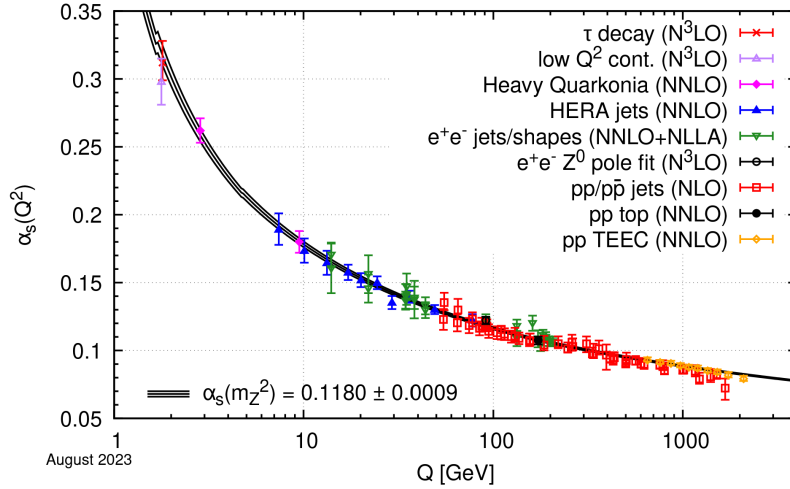


Figure 1.3: Summary of the measurements of the running coupling of α_s as a function of the energy scale, Q [9].

The behaviour of α_s as a function of the momentum transfer can be seen in Fig. 1.3. At short distances, *i.e.* at high energies, the strength of α_s decreases. In other words, quarks are able to move in a quasi-free state at short distances. This phenomenon is called **asymptotic freedom** and allows for the usage of a perturbative approach for

QCD calculations (pQCD). On the other hand, at large distances, that is, at low energies, quarks are tightly bonded, which explains why they are always found in nature forming neutral colour states. This is called **confinement**. At this point, pQCD is no longer applicable, and other methods have to be used, such as effective theories or lattice QCD (lQCD). Lattice QCD is based on discretising the QCD Lagrangian in a lattice of space-time points, where the Lagrangian can be solved. The crossover between the two regions is around $Q = \Lambda_{\text{QCD}} \sim 255 \text{ MeV}$ [10].

1.2.2 Chiral symmetry breaking

Another feature of QCD that we need to consider is its chiral symmetry breaking, which causes particles to have mass. If quarks had no mass, the QCD Lagrangian would be invariant under helicity transformations. Using Noether's theorem, this invariance implies that a **chiral symmetry** prevents quarks with different helicities from interacting. In bracket notation, $\langle \bar{\psi}\psi \rangle = \langle \bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L \rangle = 0$.

However, quarks possess mass in nature, which implies that $\langle \bar{\psi}\psi \rangle \neq 0$. Hence, the **chiral symmetry is explicitly broken**, since the masses introduce terms in the action which are not invariant under transformations. These quark masses originate from the Higgs mechanism in the electroweak sector of the Standard Model: through Yukawa interactions between quark fields and the Higgs doublet, quarks acquire mass once the Higgs field develops its non-zero vacuum expectation value, $\langle \phi \rangle = v \simeq 246 \text{ GeV}$.

On top of that, the chiral symmetry is also **spontaneously broken** in QCD, upon which the currents are conserved, but the ground state is not invariant under the action of said currents. This process was first formulated by Nambu [11, 12] and later improved by Goldstone [13, 14], who introduced the need for a set of **massless Goldstone bosons**. In QCD, this role is fulfilled by eight bosons: $\pi^0, \pi^\pm, K^0, K^\pm, \bar{K}^0$ and η_8 , whose masses are significantly lower than other hadrons.

The mass of the nucleon is generated predominantly by QCD dynamics, namely confinement and spontaneous chiral symmetry breaking, which together account for about 99 % of its mass. The Higgs mechanism contributes only the small explicit symmetry-breaking component through the light-quark masses. The symmetry is expected to be restored at high energies, where the light quark (u, d, and s) masses are expected to become null.

1.3 The quark-gluon plasma

There is now considerable evidence to believe that the Universe as we know it started when the Big Bang happened, which had extremely high temperatures and energy densities. Afterwards, it expanded, cooling down in the process. As the temperature lowered, the Universe went through different phases. In particular, when the temperatures were high enough, quarks and gluons were able to move as quasi-free particles

due to the asymptotic freedom of QCD. The resulting medium is called the **quark-gluon plasma (QGP)**. It was first predicted in 1975 when Collins and Perry suggested that a "superdense matter (found in neutron-star cores, exploding black holes, and the early Big Bang Universe) consists of quarks rather than of hadrons" [15]. The term "quark-gluon plasma" was attributed soon after in a 1978 paper by Shuryak [16].

Nowadays, it is possible to recreate the conditions needed to generate the QGP in the laboratory via **heavy-ion collisions**. Several experiments are dedicated to studying these collisions, mainly at Brookhaven National Laboratory (BNL) and CERN. Nevertheless, heavy-ion collisions can only reproduce these conditions for a very short time ($\sim 10^{-23}$ s). Thus, its properties have to be studied indirectly using probes affected by the QGP.

Heavy-ion experiments over the years

The first experiments that used ion beams at a few GeV took place in the late 1970s, starting at **BEVALAC** [17], at Lawrence Berkeley National Laboratory, USA. CERN joined the quest to study heavy-ion collisions in 1986 at the **Super Proton Synchrotron (SPS)**, where they collided relatively light nuclei such as oxygen and sulphur. At the same time, the **Alternating Gradient Synchrotron (AGS)** experiments at Brookhaven National Laboratory (BNL) began its relativistic heavy-ion programme. In 1994, the lead beam programme at the SPS started, colliding lead ions with a centre-of-mass energy of 160 GeV per nucleon. This programme allowed them to report the creation of a new state of matter compatible with the QGP in the year 2000 [18]. Soon after, the **Relativistic Heavy Ion Collider (RHIC)** began its operations at BNL, using the existing AGN as injector and increasing the available collision energy by one order of magnitude. RHIC studied Au–Au collisions and its discoveries greatly boosted the field in the following years. For instance, they discovered that the QGP medium behaves like a perfect liquid rather than a perfect gas [19].

In 2009, the **Large Hadron Collider (LHC)** began its operations at CERN (more details in Section 3.1), including a heavy-ion programme using Pb–Pb collisions at a nominal centre-of-mass energy of 5.5 TeV. The ALICE experiment was primarily designed to study the QGP, although all four main experiments currently include heavy-ion groups.

1.3.1 The QCD phase diagram

One reason to study high-energy heavy-ion collisions is the possibility to investigate the **phase diagram of QCD matter**. The different phases in which QCD matter can be found are represented in Fig. 1.4 as a function of the temperature, T , and the baryonic chemical potential, μ_B . In this context, μ_B represents the net baryon density, as it is proportional to the number of baryons.

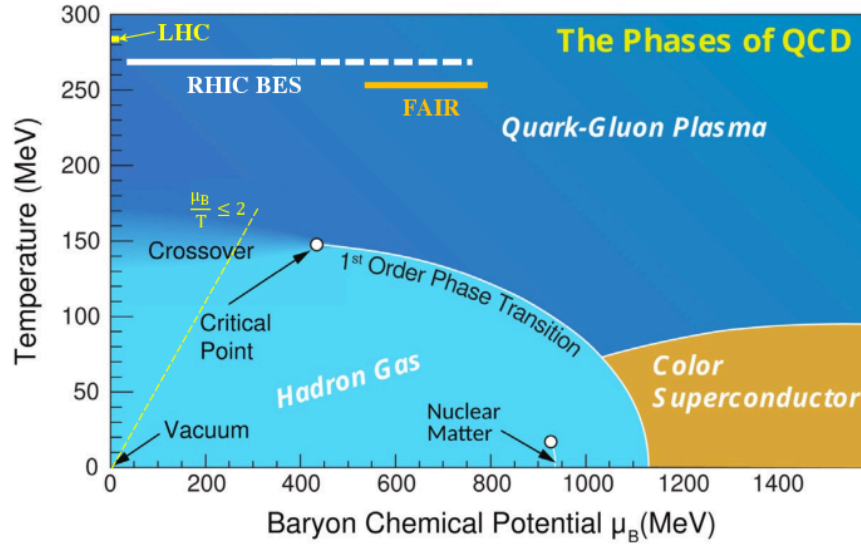


Figure 1.4: Schematic view of the QCD phase diagram. Figure extracted from [20].

Quarks and gluons are confined into hadrons at low temperatures and baryon densities, creating a **hadron gas**. In particular, nuclear matter in standard conditions is located at $\mu_B \sim 1$ GeV and at very low temperatures. However, if the baryon chemical potential rises but the temperatures remain low, QCD matter behaves as a degenerated gas of neutrons. This state of matter is believed to be found in the core of neutron stars [15]. If the baryonic density continues to grow to the point where $\mu_B \rightarrow \infty$, Cooper pairs of quarks are expected to be created, leading to a **colour superconducting phase** [21].

An equal amount of matter and antimatter were created in the very early Universe, providing a null baryon density at very high temperatures. These conditions favoured the production of QGP and can be reproduced in the Pb–Pb collisions that take place at the LHC. The transition temperature between the hadron gas and the QGP phases was calculated using different models, predicting a pseudocritical temperature, T_C , of $T_C \sim 155$ MeV at zero μ_B [22]. Also, lattice QCD (lQCD) predicted a **smooth crossover** between the two states rather than an abrupt transition at low μ_B . Nonetheless, the nature of this transition at higher values of μ_B is still under study, as lQCD is harder to apply in this regime. Overall, a **first order phase transition** is expected for higher values of μ_B [23]. First order phase transitions are abrupt transitions that exhibit a discontinuity in the first derivative of the free energy with respect to some thermodynamic parameters. A **pseudocritical point** could separate the two phase transition regimes. The current efforts to find the position of this point include the

studies at RHIC performed with different ion systems and collision energies. This search will continue in the future at FAIR, a heavy-ion facility currently being built in Darmstadt, Germany.

1.3.2 Evolution of a heavy-ion collision in the Bjorken scenario

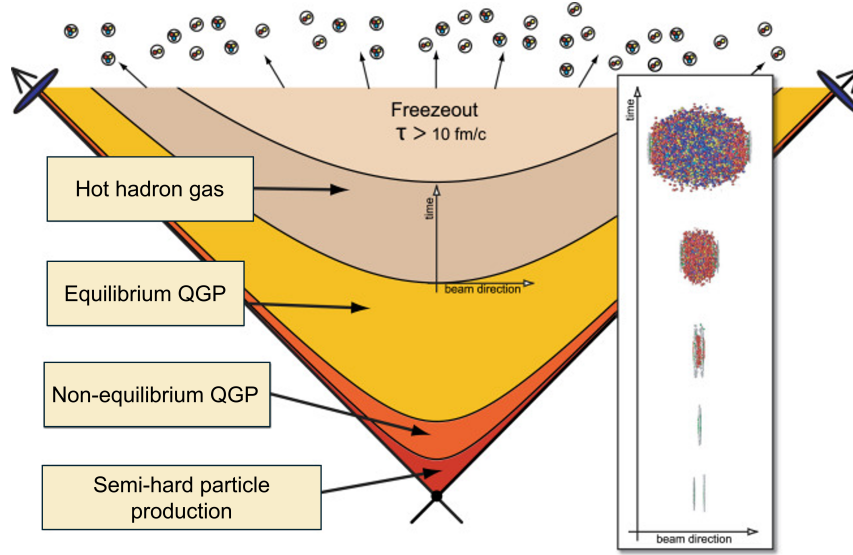


Figure 1.5: Space-time evolution of a heavy-ion collision. On the right, the evolution on the laboratory frame is displayed. Image from [24].

In order to study the QGP via heavy-ion collisions, it is necessary to understand how the system evolves in time. The most common model to describe this is the **Bjorken scenario** [25], first proposed in 1983. In this scenario, it is assumed that the initial system after a heavy-ion collision is isotropic, regardless of the rapidity. Starting from this premise, it is possible to study the hydrodynamic space-time evolution of the system, as represented in Fig. 1.5. If the collision took place at a proper time $\tau = 0$, it is possible to distinguish several phases, which will be introduced in the following. It is worth mentioning that the proper times mentioned are approximative and should be considered in the most central class of a heavy-ion collision at the LHC.

1. $0 < \tau < 0.3 \text{ fm}/c$: right after the collision, the system evolution is dominated by hard particle production processes. At high energies, such as the ones achieved at the LHC, this period can be described by the **colour glass condensate** (CGC) [26, 27] approach. For a more detailed description of this approach, see Sec. 2.1.1.
2. $0.3 < \tau < 1 \text{ fm}/c$: once the majority of the hard scatterings occur, the system enters a **pre-equilibrium** phase. Conditions are not stable enough to use linear viscous hydrodynamic equations to describe it yet [24].
3. $1 < \tau < 6 \text{ fm}/c$: as the system expands, it **thermalises**. At this point, the medium can be described with hydrodynamic equations. The energy density is around 20 times higher than that of hadrons, and temperatures can reach up to 300 MeV at the LHC [28]. Particles are strongly coupled and collectively behave

as a relativistic hydrodynamic fluid with low viscosity [29].

4. $6 < \tau < 10 \text{ fm}/c$: the medium cools down until it reaches the critical temperature, making the system go through a crossover phase transition. Thus, quarks and gluons begin to hadronise, forming a **hot hadron gas**. At a temperature of around 150 MeV [22], a **chemical freezeout** stops the inelastic interactions between the hadrons in the gas. The relative abundances are fixed to the values they had at the moment of the chemical freezeout.
5. $\tau > 10 \text{ fm}/c$: as a last step, a **kinetic freezeout** causes all elastic interactions to stop. Particles are now capable of moving freely, eventually reaching the detectors. The momentum distributions are fixed to their values at this point.

1.4 Collision and particle kinematic properties

The objective of this section is to introduce some terminology commonly used when talking about heavy-ion collisions. More specifically, we will introduce the concepts of collision centrality, particle (pseudo)rapidity and invariant mass.

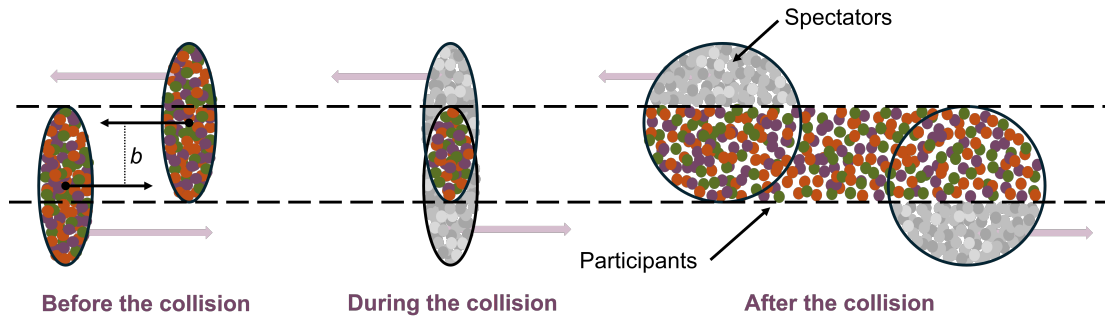


Figure 1.6: Visual representation of the collision of two heavy-ion nuclei.

Centrality

When two nuclei collide, not all nucleons will likely meet another nucleon. In other words, not all nucleons might take part in the collision. Hence, if we assume that nucleons propagate in straight, parallel trajectories, it is possible to distinguish between **participants** (N_{part}) and **spectators** (N_{spec}) of the collision (see Fig. 1.6).

The vector connecting the centres of the two colliding nuclei is called the *impact vector*, and its modulus is the **impact parameter**, b . In general, a coordinate system is chosen so that the z -axis is parallel to the beam and the impact vector points in the x -direction. The x and z axes generate the **reaction plane** of a given collision.

The number of participants in a collision can be related to the number of binary nucleon-nucleon collisions (N_{coll}) using the **Glauber model** [30]. This model assumes that heavy-ion collisions are composed of a superposition of individual nucleon-nucleon collisions. In practice, since neither N_{part} nor N_{coll} are directly measurable, centrality is estimated using the number of charged particles produced in the collision (N_{ch} , also

known as the **charged-particle multiplicity**) and N_{spec} . It is commonly expressed as a percentile, with lower values representing more head-on collisions. This way, we can separate collisions into centrality classes. The two nuclei collide almost head-on in **central collisions**, translating to large values of N_{ch} and a small impact parameter. On the other hand, in **peripheral collisions**, the impact parameter is larger. Thus, N_{ch} is smaller than in central collisions (and the percentage is larger).

Rapidity and pseudorapidity

The rapidity is a Lorentz-invariant quantity defined as:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) = \text{arctanh} \left(\frac{p_z}{E} \right), \quad (1.4)$$

with p_z representing the component of the momentum in the beam direction. E is the energy of a particle, defined as $E = \sqrt{m^2 + \mathbf{p}^2}$. Similarly, the pseudorapidity is defined as:

$$\eta = \frac{1}{2} \ln \left(\frac{|\mathbf{p}| + p_z}{|\mathbf{p}| - p_z} \right) = -\ln \left(\tan \frac{\theta}{2} \right), \quad (1.5)$$

where θ is the scattering angle. If the particle mass is small, $m \rightarrow 0$, the rapidity and the pseudorapidity are equal.

Invariant mass

The **invariant mass** is a Lorentz-invariant quantity which can be related to the reconstruction of the mass of a particle from its decay products, or prongs. Its expression depends on the energy and momentum as:

$$M_{\text{inv}}^2 = \left(\sum_{i=1}^{N_{\text{prongs}}} \sqrt{m_i^2 + \mathbf{p}_i^2} \right)^2 - \left(\sum_{i=1}^{N_{\text{prongs}}} \mathbf{p}_i \right)^2, \quad (1.6)$$

where m_i represents the mass hypothesis of each decay track and $\mathbf{p}_i$ is the spatial momentum vector of the same decay track. Lastly, N_{prongs} is the number of decay products. For example, in the $D^0 \rightarrow K^+ \pi^-$ channel m_i is either the pion (m_π) or the kaon (m_K) mass.

1.5 Experimental probes to study heavy-ion collisions

Since the QGP only lives for a few femtoseconds, it is not possible to make direct measurements. Instead, we must resort to using observables who change as they travel through the QGP medium, called probes. One can distinguish between several types of probes:

- **Soft probes:** these processes involve low momentum transfers and can be produced at all stages of the collision. Some examples include the anisotropic flow

or the strangeness enhancement.

- **Hard probes:** in these cases, high momentum transfers are involved. Hard probes are typically produced by hard scatterings at the early stages of the collision. Hence, they experience the full evolution of the system. For instance, some hard probes are the weak bosons ($W^\pm, Z^0$), heavy-flavour hadrons, and jets.
- **Electromagnetic probes:** photons and dileptons are produced at all stages of the collision, but they do not interact strongly with the QGP medium. Hence, they can provide information about the entire evolution of the system.

An exhaustive list of experimental QGP probes can be found in [29, 31–34]. In the following, we present a snippet of some of these probes, organised depending on whether they provide insight on the initial or the final state of the collision.

1.5.1 Initial state probes

We define the initial state probes as those that remain unaffected after they pass through the QGP medium. Two examples of these probes are the direct photons and the electroweak bosons.

Direct photon production

Photons and di-lepton pairs are less likely to interact with the deconfined medium since, unlike quarks, they do not carry colour charge. Instead, they interact via the electromagnetic force, which is significantly weaker than the strong interaction. For this reason, the photons and lepton pairs created at the beginning of the collision are hard probes which can navigate through the QGP medium almost undisturbed, providing insights on the status of the system at that time.

However, it is important to take into account that photons and di-leptons are created at all stages of the collision. As a consequence, it is necessary to distinguish between those created in the initial hard scattering and those created in the deconfined medium or in the decay of hadrons in the final state. Photons created in the initial hard scattering are characterised by their large transverse momentum, allowing them to be described with pQCD calculations. Lepton pairs created at the beginning of the collision, on the other hand, are produced via the Drell-Yan mechanism ($q\bar{q} \rightarrow l^+l^-$, with $l = e, \mu$).

Weak boson production

Electroweak bosons ($W^\pm$ and Z^0) have very large masses, so considerable momentum is needed to create them. At the same time, they are transparent to the QGP since they have no colour. Hence, they are produced at the beginning of the collision and remain unaltered as they pass through the medium, making them a good set of **standard candles** of heavy-ion collisions. Furthermore, they can be used to verify the Glauber

model since their production is expected to scale with the number of binary collisions. Additionally, they can be used to calibrate the jet energy loss in the QGP by measuring Z-jet processes [35].

1.5.2 Final state probes

The final state probes provide insight into the QGP evolution. They are affected by the QGP medium, inducing effects which are not reproduced in smaller systems. In this section, we will introduce a few examples of final state probes, such as the anisotropic flow, the strangeness enhancement, the jet quenching, or the quarkonia suppression and regeneration mechanisms. Open heavy-flavours will not be discussed since Chapter 2 is exclusively dedicated to them.

Anisotropic flow

In non-central heavy-ion collisions, the overlap region between nuclei (see Fig. 1.6) generates an **initial geometrical anisotropy**. At large enough energy densities, the particles in the overlap region of the collision reinteract, creating a pressure gradient. This gradient is stronger in the short axis, where less matter is in the way. Hence, most matter will move towards this axis and lose less momentum due to interactions with neighbouring particles. However, particles that try to exit the overlap region in the perpendicular direction will be slowed down by interactions with the medium, decreasing their momentum. This process creates an anisotropy in the momentum distribution of the final states, the anisotropic flow. A schematic representation of the process can be seen in Fig. 1.7.

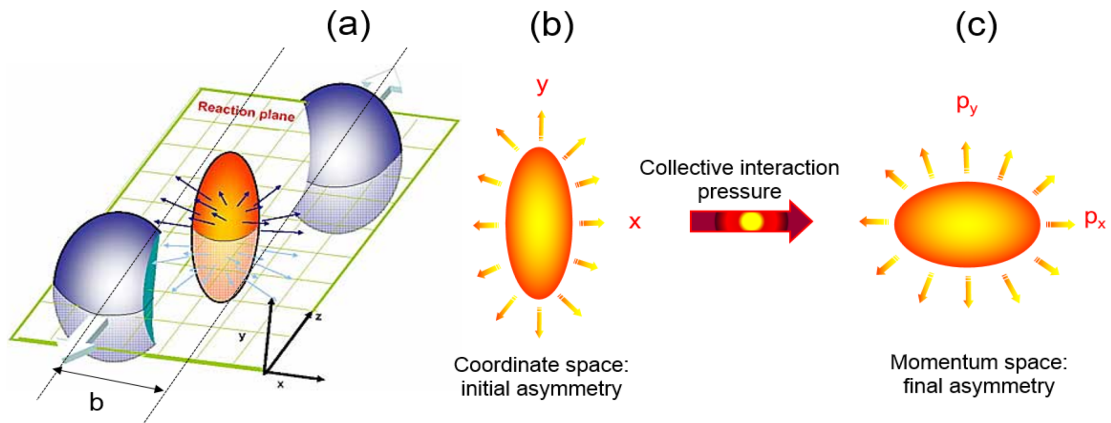


Figure 1.7: Description of the formation of the anisotropic flow [36].

The angular distribution of the final state particles can be described using a Fourier

expansion of the azimuthal distribution of the emitted particles:

$$\frac{dN}{d\varphi} \propto 1 + \sum_{n=1}^{\infty} 2v_n \cos(n\varphi - n\Psi_{\text{RP}}), \quad (1.7)$$

$$v_n = \langle \cos(n\varphi - n\Psi_{\text{RP}}) \rangle, \quad (1.8)$$

where v_n is the harmonic at the n -th order, φ is the azimuthal angle and Ψ_{RP} is the angle of the reaction plane. The first harmonic term, v_1 , is called the **directed flow**. It represents the difference between the emission of particles at 0 and 180 degrees. The second harmonic, v_2 , can be interpreted as the **elliptic flow**. Higher harmonics account for higher-order asymmetries in the initial state geometry. Overall, these harmonics can be used to extract hydrodynamic information about the medium, such as the shear viscosity [37].

Strangeness enhancement

The enhancement of strangeness production relative to that observed in pp collisions was one of the first proposed signatures of the quark-gluon plasma (QGP) [38]. Unlike up and down quarks, strange quarks are not present as valence quarks in the colliding nucleons, but they are still relatively common due to their low mass. In the QGP, several mechanisms contribute to the efficient production of strange quarks. First, the restoration of chiral symmetry effectively reduces the quark masses ($m_q \sim 0$), making the production of strange hadrons more energetically favourable than in a vacuum. Furthermore, the high gluon density in the QGP enhances the probability of processes such as $gg \rightarrow s\bar{s}$ with respect to the vacuum, further increasing strangeness production. As a result, an enhancement of strange particle yields is expected with increasing system size, from pp to p-A to A-A collisions. It was also predicted that there should be an **enhancement of multistrange (anti)baryons** [39], which should scale with their strangeness content, such that the relative yields with respect to pp, R , fulfil that

$$R_{\Lambda} < R_{\Xi} < R_{\Omega}.$$

Several collaborations have observed a strangeness enhancement experimentally at different collision energies [40–42]. It was initially considered as a strong evidence for the formation of a deconfined partonic medium in heavy-ion collisions, but the observation of strangeness enhancement in small systems (pp [43] and p-Pb [44]) challenges this interpretation. We will discuss these results in the next section.

Thermal photons and di-lepton production

In analogy to black body radiation, the QGP medium emits electromagnetic radiation as low-momentum photons and thermal di-leptons. In-medium-produced di-lepton pairs are usually located in the low-mass region ($M < 1$ GeV). This region is domi-

nated by hadronic interactions and decays produced after the freezeout. A contribution at intermediate mass ($1 < M < 2.5$ GeV) can also be expected, in this case coming from the thermalised QGP. The low mass spectrum is the superposition of these different contributions dominated by the two-body decays of several mesons, such as ρ^0 , ϕ or ω . Studying their widths and masses, it is possible to infer information on the hadron properties at the late stages of the collision process [45].

Jet quenching

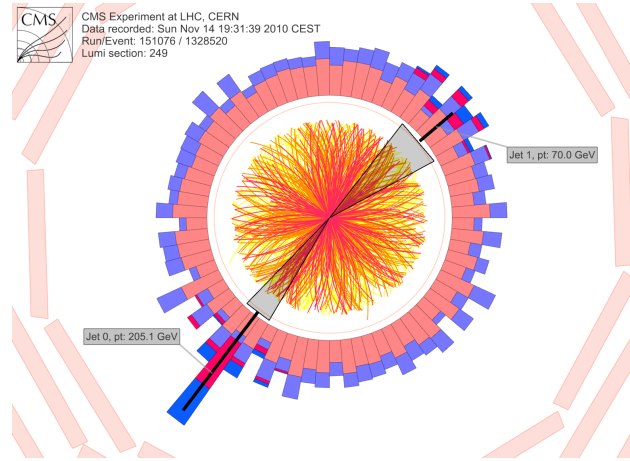


Figure 1.8: Event display of a CMS Pb-Pb event recorded in November 14 2010 showing the energy imbalance of a pair of jets caused by jet quenching [46].

Jets are collimated sprays of particles produced in a hard scattering. They are defined through a jet reconstruction algorithm, the most common being the anti- k_T algorithm [47]. They have been found to experience significant modifications as a consequence of their passing through the QGP medium, both at RHIC [32, 33] and LHC [48, 49] energies.

In pp collisions, jets produced in hard scatterings come in pairs of back-to-back jets as a consequence of momentum conservation. In events where the di-jet momentum balance appears to be violated, the imbalance is often attributed to the presence of a third hard jet or a high- p_T particle that balances the transverse momentum.

However, in heavy-ion collisions, this is no longer the case. Often, only one of the jets can be reconstructed. At the same time, its back-to-back partner appears to be either completely absorbed or significantly softened and broadened, making it difficult to identify as a distinct jet. This phenomenon is referred to as **jet quenching** [50], representing one of the key signatures of QGP formation. Jet quenching arises primarily due to energy loss of high-energy partons as they traverse the QGP. This energy loss occurs through a combination of radiative and collisional processes, where the parton interacts with the medium constituents. The amount and nature of this energy loss depends on several factors, including the path length traversed by the parton, the colour charge, and the local properties of the medium. Hence, by studying jet quench-

ing, it is possible to probe the medium at different resolution scales and learn about the medium response to the passing of the jet. These effects are investigated experimentally through a variety of observables, such as di-jet asymmetry [51], the nuclear modification factor [52], jet fragmentation functions [53], and jet-hadron correlations [54].

Quarkonia suppression and regeneration

Heavy-flavour (HF) quarks, namely charm (c) and beauty (b), are produced in the early stages of high-energy collisions, primarily through hard-scattering processes. As a result, they experience the entire evolution of the medium created in the collision. Due to their large masses, their production cross sections can be reliably calculated using pQCD over the full transverse momentum (p_T) range. After propagating through the QGP, heavy quarks hadronize by fragmenting into hadrons, which decay before detection. During hadronization, an HF quark may combine with a light quark to form an **open heavy-flavour** hadron, or with another HF quark to form a **quarkonium** state. A more detailed discussion of open heavy-flavours is presented in Chapter 2.

In the case of quarkonium production, measurements in Pb–Pb collisions at the SPS at $\sqrt{s_{NN}} = 17$ GeV [55] showed a yield suppression with respect to pp collisions. Afterwards, RHIC [56, 57] also found evidence of suppression at a higher collision energy. These results were associated with the predictions of [58], who anticipated a **sequential suppression** of the quarkonia states. Tightly bound states, such as the J/ψ , are more resilient to dissociation and less suppressed than weakly bound states like the $\psi(2S)$. This behaviour led to the proposal of quarkonia as **QGP “thermometers”** [59]. However, when the same measurements were performed at the LHC with a centre-of-mass energy per nucleon around 10 times larger, the **regeneration of quarkonium states** was found relevant at low p_T for central collisions. This regeneration process is caused by the recombination of independently produced heavy quarks [60]. It only appears to be significant at LHC energies, suggesting it requires a sufficiently high temperature or charm-quark density to become prominent. Studies of quarkonia production in the statistical hadronisation model and transport models suggest that the main factor driving the regeneration is the total charm cross section [61, 62], which is larger at the LHC.

Figure 1.9 aims to illustrate concepts of suppression and regeneration. It contains the results of the J/ψ nuclear modification factor (R_{AA}), which is proportional to the ratio of yields in pp and A–A collisions (see the next section for a description of this factor), as a function of the number of participants in the collision. The measurements of STAR [56] and PHENIX [57] are compatible with each other and show no signs of regeneration, whereas the R_{AA} measured in ALICE [63] at a larger $\sqrt{s_{NN}}$ tends to grow for $\langle N_{part} \rangle > 100$.

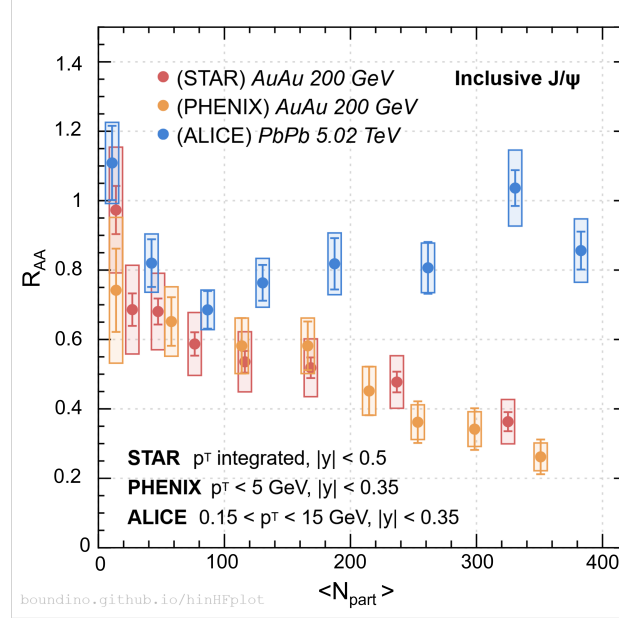


Figure 1.9: Nuclear modification factor of J/ψ mesons as a function of the number of participants in a heavy-ion collision. In red, results by the STAR collaboration [56], in orange, by PHENIX [57], both measured in Au–Au collisions at $\sqrt{s_{NN}} = 200$ GeV. In blue, results by ALICE [63] in Pb–Pb collisions at 5.02 TeV. Plot generated using [64].

1.6 Studies in small systems

To assess the effect of the QGP medium in heavy-ion collisions, it is necessary to establish a basis for those measurements in smaller collision systems, where no QGP is expected to be formed. Hence, to understand A–A collisions, the pp and p–A systems must also be studied. Moreover, results in these systems can be used to investigate QCD properties and predictions made with pQCD calculations.

1.6.1 Using pp collisions as a baseline: the nuclear modification factor

One variable that can be used to compare the behaviour in pp and A–A collisions is the **nuclear modification factor**, R_{AA} . In the context of the Glauber model, the nuclear modification factor can be considered as a test of the assumption that particle production in a heavy-ion collision is the sum of individual binary collisions. It is defined as

$$R_{AA} = \frac{1}{\langle N_{coll} \rangle} \frac{d^2 N_{AA}/dp_T dy}{d^2 N_{pp}/dp_T dy}. \quad (1.9)$$

Here $\langle N_{coll} \rangle$ represents the average number of binary collisions. $d^2 N_{AA}/dp_T dy$ and $d^2 N_{pp}/dp_T dy$ are the number of measured particles as a function of the transverse momentum and the rapidity in A–A and pp collisions, respectively. Deviations from unity of the R_{AA} imply modifications in the p_T and y distributions of the produced particles. These modifications can be caused by the interaction of particles with the QGP medium, or by the so-called **cold nuclear matter effects** (CNM). To distinguish

between the two cases, the nuclear modification factor can be calculated in p–A collisions, where no QGP is expected. A short review about the CNM effects can be read in [65], and the specific case for open heavy-flavours is discussed in Sec. 2.3.

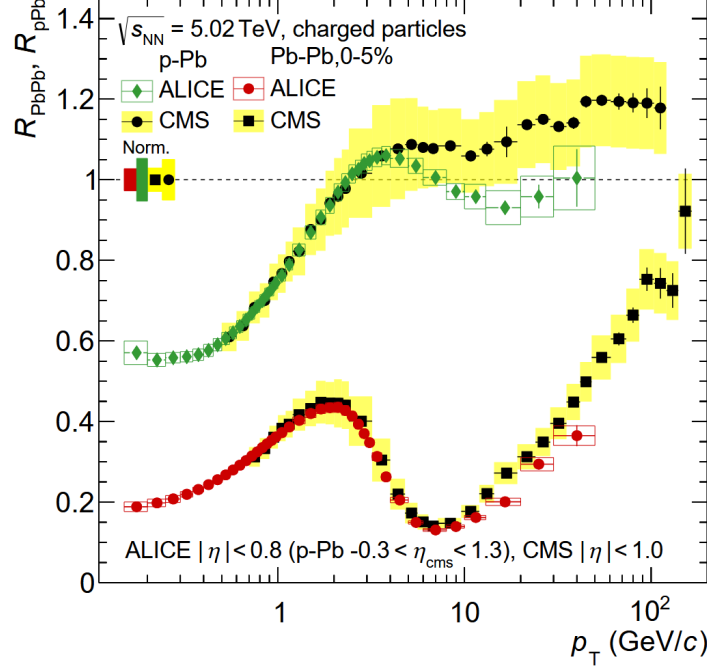


Figure 1.10: Nuclear modification factor for charged particles measured at $\sqrt{s_{NN}} = 5.02$ TeV by the ALICE (coloured) and CMS (black) collaborations in p–Pb and central Pb–Pb collisions [66].

The R_{AA} and the R_{pA} of charged particles as a function of p_T measured by ALICE [66] and CMS [67] in p–Pb and central (0–5%) Pb–Pb collisions is presented in Fig. 1.10. As can be seen, the p–Pb nuclear modification factor is larger than the Pb–Pb one at all p_T ranges for charged particles, suggesting that hadrons lose energy as they travel through the deconfined medium. However, both p–Pb and Pb–Pb collisions are susceptible to CNM effects, explaining why they both have regions where the nuclear modification factor is different from unity.

1.6.2 QGP-like behaviours in small systems

In small systems, namely pp and p–Pb collisions, the temperature and energy density conditions are not expected to be extreme enough to produce a quark-gluon plasma. For this reason, the studies in these systems have been traditionally considered as a baseline for QGP studies in heavy-ion collisions. However, recent measurements in high-multiplicity pp and p–Pb collisions have found similar behaviours to those found in the QGP [43, 68, 69]. In this section, a short review of some of these effects will be given, providing a motivation to study small systems as more than a necessary step to understand heavy-ion collisions.

Strangeness enhancement in small systems

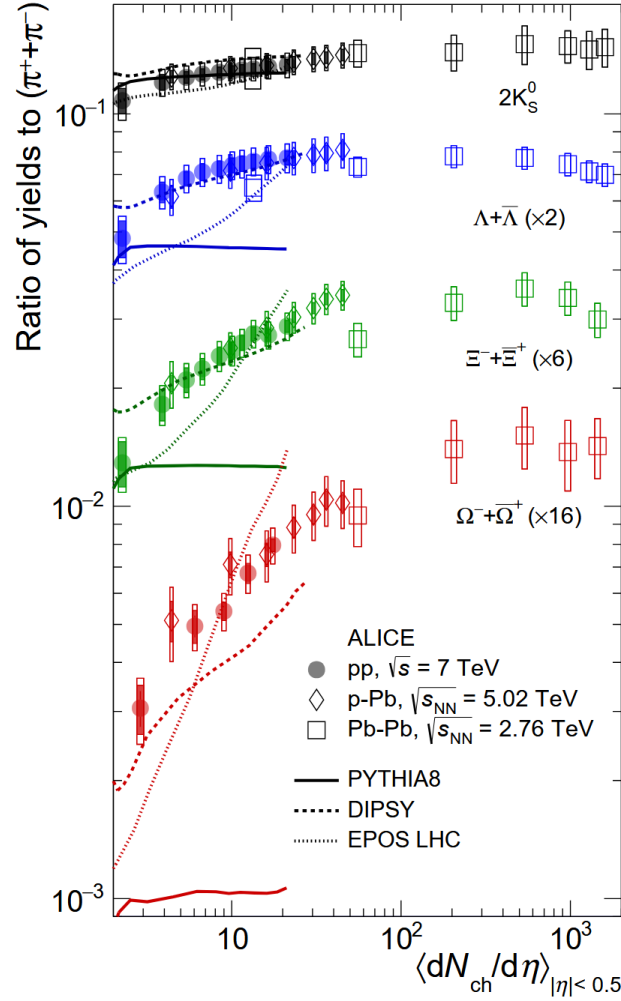


Figure 1.11: Ratio of the production yield of K_s^0 (in black), Λ (in blue), Ξ (in green), and Ω (in red) particles to pion yields measured in pp, p-Pb and Pb-Pb collisions as a function of charged-particle multiplicity [43]. pp, p-Pb and Pb-Pb measurements are represented with circles, diamonds, and squares, respectively. Results are compared to PYTHIA 8 [70], DIPSY [71] and EPOS LHC [72] predictions.

As discussed in section 1.5.2, one of the earliest proposed signatures of the formation of the QGP medium is the enhancement of the production of strange particles. However, an analysis performed in ALICE in 2016 [43] using pp data at $\sqrt{s} = 7$ TeV [43], p-Pb data at $\sqrt{s_{NN}} = 5.02$ TeV [44] and Pb-Pb data at $\sqrt{s_{NN}} = 2.76$ TeV [42] calculated the production yields of several strange hadrons (K_s^0 , Λ , Ξ , and Ω) as a function of charged-particle multiplicity. The results of this measurement can be found in Fig. 1.11. In general, a growing trend can be seen for all particles, as well as a smooth overlap between the different collision systems. Results are also compared to theoretical predictions calculated using MC generators, namely PYTHIA 8 [70], DIPSY [71] and EPOS LHC [72]. However, none of the considered models at the time were able to provide a description for all hadrons.

Elliptic flow in small systems

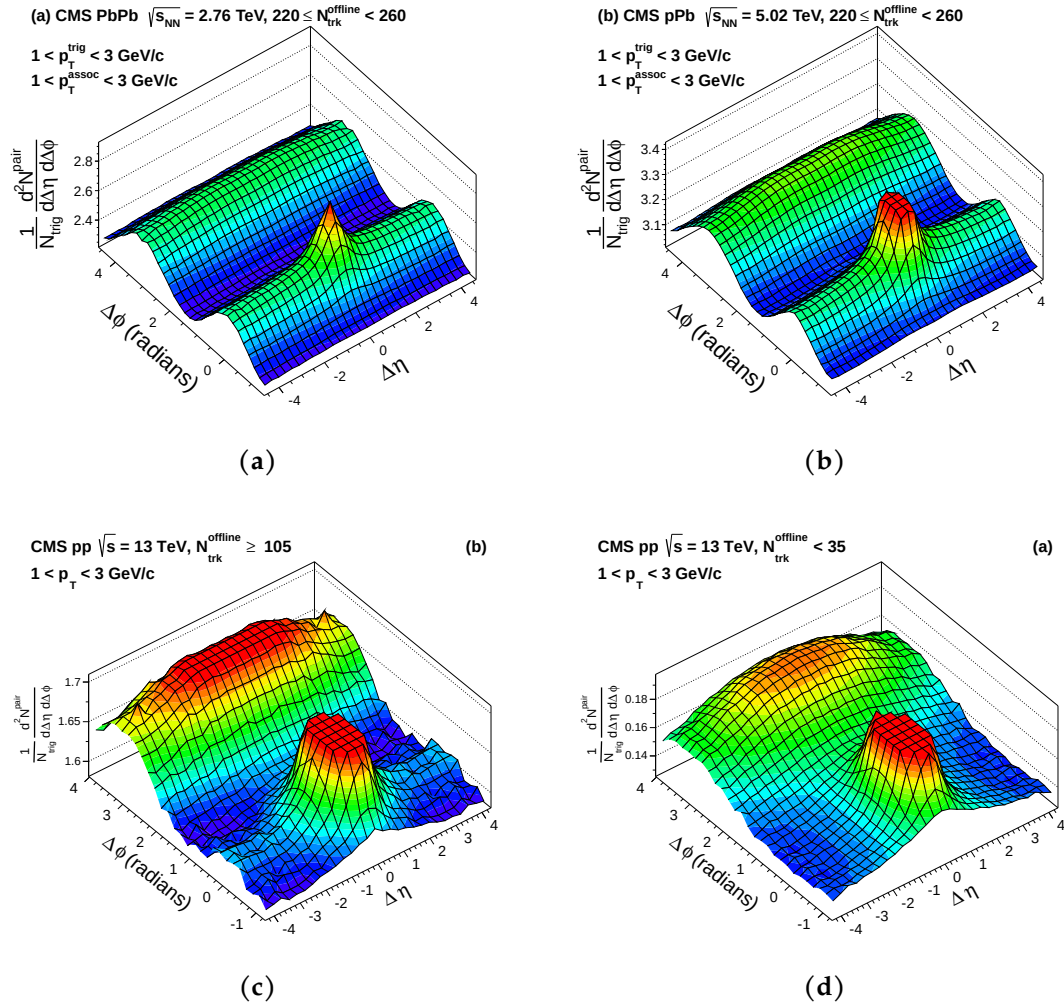


Figure 1.12: Measurements performed by CMS of the $(\Delta\eta, \Delta\phi)$ two-particle correlations in different collision systems: (a) Pb–Pb [69]; (b) p–Pb [69]; (c) pp with high multiplicity [68]; (d) pp with low multiplicity [68].

One way to study the elliptic flow is by analysing the $(\Delta\eta, \Delta\phi)$ two-particle correlations, as was performed in a CMS study at $\sqrt{s_{\text{NN}}} = 2.76$ TeV [69] (Fig. 1.12a). In heavy-ion collisions, the expected shape is composed of a near-side ($\Delta\phi = 0$) ridge covering a wide range of $\Delta\eta$ a clear peak at $\Delta\eta \sim 0$, generated by the short-range interactions that take place in jets. The ridge structure is created by long-range correlations, which can be linked with the anisotropic flow originating from the QGP medium. Although this behaviour is expected in heavy-ion collisions, several studies [68, 69] have found similar ridge-like structures in smaller systems, such as p–Pb and high-multiplicity pp collisions (Figs. 1.12b and 1.12c, respectively). Nonetheless, the ridge has not been observed for low-multiplicity pp collisions (1.12d).

Summary of the current observations in small systems

A summary table of the findings per collision system can be found in Tab. 1.1 for the probes introduced in section 1.5. As can be observed, not all probes are present in small systems. For example, jet quenching is only found in heavy-ion collisions. However, the fact that other observables seem to exist in smaller systems is not fully explained yet.

One possible interpretation is that small systems do not reach the conditions to create a full QGP, but are still extreme enough to generate some **QGP droplets** if the multiplicity is large enough. **Multiple Parton Interactions (MPIs)**, where two or more partons interact in a single hadron-hadron collision, could also play a role. Lastly, it is likely that p-Pb results are affected by **cold nuclear matter effects** (Sec. 2.3).

Table 1.1: Summary of the observations per collision system of the probes introduced in section 1.5 at LHC energies.

Probe	Pb-Pb	p-Pb (high mult.)	pp (high mult.)
Azimuthal anisotropy (from two part. correlations)	Yes [69]	Yes [69]	Yes [68]
Jet quenching	Yes [48, 49]	Not observed [73]	Not observed [74]
Quarkonia	Suppression + regeneration [77]	Suppressed [75]	Not suppressed [76]
Strangeness enhancement	Yes [42]	Yes [44]	Yes [43]
Thermal photons at low p_T	Yes [78]	Not excluded [79]	Not observed [79]

OPEN HEAVY-FLAVOUR PRODUCTION

Open heavy-flavour hadrons are characterised by containing at least one charm or beauty quark, but not their corresponding antiquark. They constitute powerful probes of Quantum Chromodynamics (QCD) and the properties of the matter created in high-energy collisions at the LHC. Produced primarily in the earliest, hardest scatterings, heavy quarks traverse the entire evolution of the collision system, making their study essential for understanding both initial and final state effects.

This Chapter introduces the main mechanisms of open heavy-flavour production in proton-proton (pp), proton-nucleus (p-A), and nucleus-nucleus (A-A) collisions. Proton-proton collisions provide a baseline for QCD studies, while p-A and A-A collisions reveal the influence of cold and hot nuclear matter, respectively. Despite being heavy-flavour quarks, top quarks are excluded from this discussion due to their inability to hadronise.

Special attention is given to multiple parton interactions (MPI), which can significantly affect heavy-flavour yields, and to key experimental measurements that underpin our current understanding of these processes.

2.1 Charm and beauty production

Heavy quarks are mostly produced in hard-scattering processes in the early stages of the collision. Processes involving heavy quarks are characterised by a large momentum transfer (Q) due to their large masses ($m_c \sim 1.3 \text{ GeV}/c^2$, $m_b \sim 4.2 \text{ GeV}/c^2$ [9]), since these imply that $Q^2 \gtrsim 4m_{c,b}^2$. This large momentum transfer causes α_s to be smaller than unity, allowing us to use perturbative QCD (pQCD) methods to calculate heavy-flavour production down to $p_T = 0$. Heavy-flavours are mainly produced in $c\bar{c}$ and $b\bar{b}$ pairs by either quark antiquark annihilation ($q\bar{q} \rightarrow Q\bar{Q}$) or gluon fusion ($gg \rightarrow Q\bar{Q}$) processes. They can later hadronise to form either quarkonia or open heavy-flavour hadrons. Due to the abundance of light quarks with respect to heavy quarks, open heavy-flavour hadrons are more likely to be produced than quarkonia states. In fact, only about 1% of the produced heavy quarks form quarkonia states [80].

Examples of open heavy-flavour hadrons include the D-meson family (D^0 , D^+ , D_s^+ , D^{*+}), which contain a charm and a light quark, the B mesons (B^0 , B^+), containing a beauty and a light quark, and several baryons, such as Λ_q , Σ_q , Ω_q , and Ξ_q . The full list of hadrons and their known decay channels is included in [9]. These decays can be categorised into one of the following types of reconstruction channels:

- **Hadronic channels:** In these channels, the particle we want to study decays into only hadrons, such as protons, kaons or pions. Thanks to the particle identification (PID) capabilities of ALICE, it is possible to obtain a complete reconstruction of open heavy-flavours in these channels. For instance, all measurements presented in this thesis include D^0 mesons reconstructed in the $D^0 \rightarrow K^- \pi^+$ channel.
- **Semi-inclusive channels:** In these cases, only a limited number of particles are detected and identified, while the rest of the contributions are summed over. Such is the case for all channels which include neutrinos, since they cannot be directly measured at the LHC.
- **Semi-leptonic channels:** These channels involve a partial hadron reconstruction using exclusively leptons in channels of the form $(D, B, \dots) \rightarrow \mu X, eX$, where X represents a hadron.

All of these reconstruction channels enable experimentalists to identify and measure open heavy-flavour hadrons in different final states, providing the necessary data to quantify their production rates in high-energy collisions. To interpret these measurements and compare them with theoretical predictions, it is essential to understand how the production cross section of heavy-flavour hadrons is calculated in pp collisions. This is typically done using the **factorisation approach** [81], which separates the total cross section into distinct components:

$$\frac{d\sigma^{\text{pp} \rightarrow H_Q}}{dp_T} = \int dx_i \int dx_j \underbrace{f_i(x_i, Q^2) f_j(x_j, Q^2)}_{\text{PDFs}} \otimes \underbrace{\frac{d\sigma^{ij \rightarrow Q}}{dp_T}}_{\text{Parton cross section}} \otimes \underbrace{D_{Q \rightarrow H_Q} \left(z_Q = \frac{p_{H_Q}}{p_Q}, Q^2 \right)}_{\text{Hadronisation}}. \quad (2.1)$$

The first component contains the information related to the probability distributions of the parton momentum fractions in the proton. This information is parametrised using the **parton distribution functions (PDFs)**. They depend on the Bjorken- x , which parametrises the momentum fraction of a parton at a given hard scale, μ_f , which is generally considered of the order of Q . In the two-body case, the Bjorken- x can be defined as

$$x = \frac{Q^2}{s + Q^2}, \quad (2.2)$$

where s is the centre-of-mass energy of the collision. The next component is the **cross section of the partons** involved in the process. As we anticipated, this factor can be calculated using pQCD in the case of heavy-flavours. Finally, the third factor of Eq. 2.1

includes the information related to the non-perturbative **hadronisation** of the heavy-flavour quarks to form baryons or mesons. In the following subsections, we will go through all of these steps individually.

2.1.1 Parton Distribution Functions

The first component of Eq. 2.1 refers to the parton distribution functions (PDFs), which describe the probability of finding a parton i with a momentum fraction x at a hard scale $Q \gg \Lambda_{\text{QCD}}$. They were first introduced in the context of **deep inelastic scattering** (DIS) experiments [82–84], where a target proton was bombarded with high-energy electrons that interacted via virtual photons. This experiment configuration allowed physicists to infer that protons are composite particles. In particular, Feynman suggested that protons were composed of point-like objects [85], which he called **partons**.

The discovery of partons led to the development of several parton models to describe the proton structure. At the beginning, protons were thought to be composed of three **valence quarks** (uud), which determined their quantum numbers [86, 87]. Nevertheless, the increasingly higher precision of the DIS experiments revealed the presence of additional components, namely **sea quarks** and **gluons** [88, 89]. Sea quarks are virtual quark-antiquark pairs produced by quantum fluctuations. Although each sea quark carries only a small fraction of the momentum of the proton, its contribution becomes increasingly significant as the resolution of the probe increases. A more extensive and quantitative introduction to these concepts can be found in [1, 8].

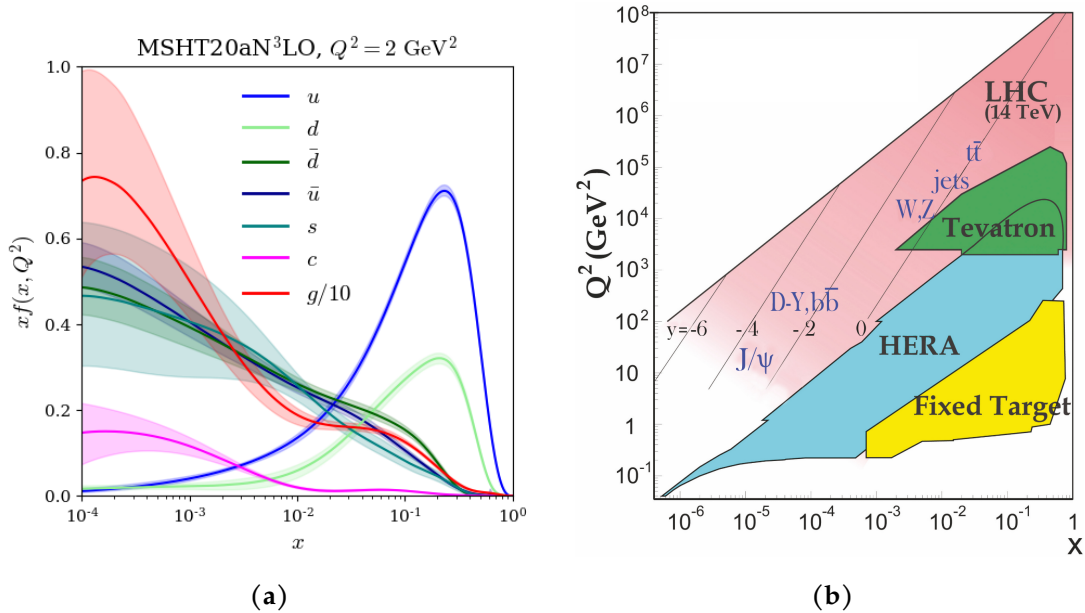


Figure 2.1: (a) PDFs at $Q^2 = 2 \text{ GeV}^2$ from the MSHT20aN³LO [90] set. (b) Kinematic coverage in the (x, Q^2) plane of key experiments contributing to PDF determinations. Figure from [9].

All the mentioned discoveries lead to the introduction of parton distribution functions (PDFs) to quantitatively describe the momentum distribution of partons inside

the proton. These functions are not perturbative and must be extracted from experimental data. Multiple global analyses are carried out by various groups (such as CT [91], MMHT [92], NNPDF [93], and others), each employing different methodologies, levels of theoretical precision, and datasets, including not just DIS but also hadron collider data. An example of the PDFs of the MSHT20aN³LO [90] set can be observed in Fig. 2.1a. As can be seen, the valence quark distributions dominate at large x , and the distribution of the u quark is twice as large as that of the d quark, which matches the idea that a proton is built as a (uud) triplet. However, for small x values (*i.e.* large energies), the contributions of the gluons and the sea quarks gain importance until the gluons dominate.

The dependence of PDFs on the energy scale is governed by the **DGLAP** [94–96] evolution equations, named after Dokshitzer, Gribov, Lipatov, Altarelli, and Parisi. These equations describe how PDFs evolve with changing momentum scales due to the emission of gluons and the splitting of partons. For instance, the distributions in Fig. 2.1a correspond to an energy scale of 2 GeV². The DGLAP evolution is derived within perturbative QCD. It allows one to extrapolate PDFs across different kinematic regimes, from low-energy fixed-target experiments to the high-energy collisions observed at the LHC and beyond. The phase space regions covered by the main experiments which study or have studied PDFs can be found in Fig. 2.1b.

The saturation region and the colour glass condensate

So far, we have discussed how the system evolves when we change the energy scale, Q^2 . The DGLAP equations describe this process. However, to see how it evolves with x , we need to use the BFKL [97–100] (Balitsky-Fadin-Kuraev-Lipatov) formalism. This formalism is based on resumming the logarithmic terms arising from soft gluon emissions. As x decreases, the transverse-momentum scale remains fixed, and the BFKL evolution leads to high parton densities [101]. When the Bjorken- x becomes small enough, the system is dominated by small- p_T gluons. At this point, the density is very high because large-momentum gluons tend to split into gluons with smaller momentum. The process continues until the system reaches saturation, when the density is so large that gluon-gluon recombination balances out the splitting and the linear evolution described by the DGLAP and BFKL formulations stops describing the medium. Instead, the system can be described with an effective field theory called the **colour glass condensate** (CGC) [102].

A schematic view of the QCD evolution as a function of Q^2 and x is presented in Fig. 2.2. The validity of the DGLAP and BFKL formalisms is represented, as well as the region where the system saturates and their validity is compromised.

The CGC framework describes the properties of saturated gluons in the Regge-Gribov limit, reached when $x \rightarrow 0$ for a fixed Q^2 . At this point, high- x partons contain colour sources which remain *frozen* in time due to the time dilation of the fast-moving nu-

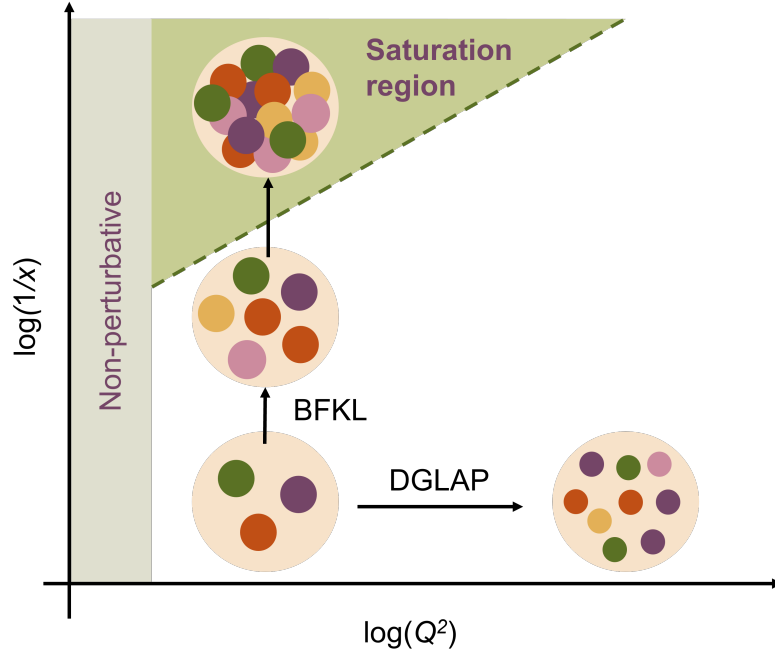


Figure 2.2: QCD evolution from DGLAP to BFKL and saturation.

cleon. These partons are randomly distributed across the transverse disk and produce gluons with small momentum fractions ($x \ll 1$) which can be treated as dynamic classical fields. The interplay of the time scale of the long-lived partons and the short-lived gluons, together with the stochastic nature of the gluon fields, explains the *glass* appellation. Lastly, the term *condensate* references the high density of the medium in this region.

An important characteristic of the CGC is that, given an initial non-perturbative distribution of colour sources at an initial scale x_0 , it is possible to compute the n -point gluon correlation functions and their evolution with x order by order in perturbation theory using the JIMWLK [103–105] renormalisation group equations. They indicate that a proton and a heavy nucleus become indistinguishable at a fixed impact parameter at high energy.

Overall, the behaviour of PDFs, including in the low- x regime where saturation effects and the CGC become relevant, plays a critical role in the factorisation approach to heavy-flavour production. With this foundation, we now turn to the next of the three key components of the factorisation framework: the partonic cross sections.

2.1.2 Parton scattering cross section

The parton scattering cross section can be calculated using perturbative QCD (pQCD), where the hard scattering matrix elements are computed as a power series in the strong coupling constant, α_s . Most modern calculations of hadronic cross sections, such as FONLL [106, 107] and GM-VFNS [108–113] (see Sec. 2.1.6), include pQCD contribu-

tions up to next-to-leading order (NLO), which significantly improves the precision over leading-order estimates. Recently, next-to-next-to-leading order (NNLO) calculations have become available [114], providing even greater accuracy and reducing theoretical uncertainties. In addition to fixed-order calculations, all-order resummations of large logarithmic corrections are often employed to refine predictions and ensure reliable results across a wide range of energies and momenta. Such is the case of FONLL and GM-VFNS, as will be discussed in Section 2.1.6.

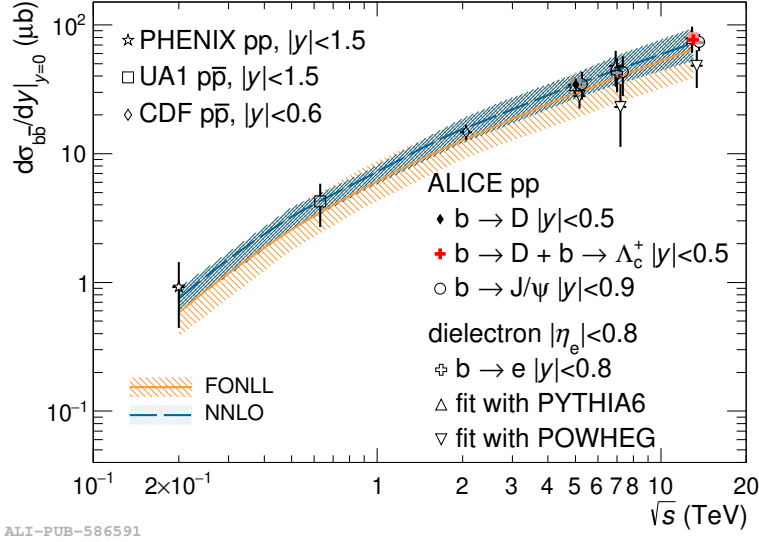


Figure 2.3: $b\bar{b}$ cross section per unit of rapidity measured in pp collisions at midrapidity as a function of the centre-of-mass energy. The figure is extracted from [115].

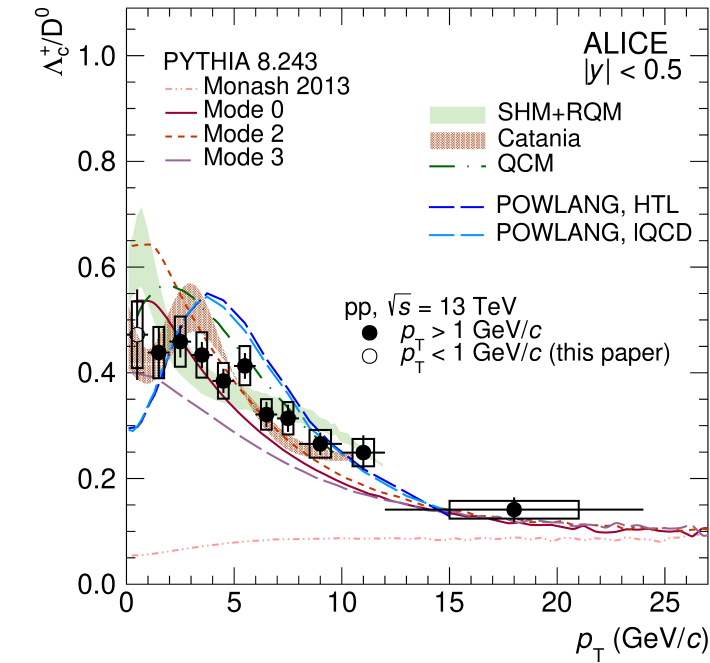
An example of the measurement of the $b\bar{b}$ cross section can be found in Fig. 2.3 as a function of the centre-of-mass energy. The results of PHENIX [116], UA1 [117], CDF, and ALICE [79, 115, 118–122] at midrapidity are compared to the theoretical predictions of FONLL and NNLO. In the case of ALICE, several values were extracted using different b -decay channels. Overall, both models are compatible with the data. As was expected, the NNLO calculations have a smaller uncertainty, and their central values are closer to the data points.

2.1.3 Fragmentation functions and hadronisation

The last ingredient of Eq. 2.1 is the hadronisation mechanism, by which quarks and gluons produced in the hard scattering form hadrons. Hadronisation constitutes a non-perturbative problem which must be studied through phenomenological models. It is parametrised using the **fragmentation functions**, $D_{Q \rightarrow H_Q}(z_Q, Q^2)$, which depend on the momentum fraction taken by the hadron H_Q , called z_Q , and the energy scale, Q^2 . Fragmentation functions are generally assumed to be universal across collision systems and centre-of-mass energies.

To study the universality of the fragmentation functions, we can resort to particle production ratios, where the effects of hadronisation can be isolated. In particular, studies

of meson-to-meson ratios in pp , e^+e^- and ep seem to confirm this universality [123, 124]. However, when the same test was performed using baryons, the results in pp showed a clear deviation from e^+e^- [125]. This measurement is presented in Fig. 2.4. As observed, the experimental results display an upward trend towards small- p_T , which PYTHIA Monash does not describe. This event generator tune uses e^+e^- measurements as a reference (more on PYTHIA in the next section). At present, multiple possible explanations for this behaviour are under discussion. For instance, the quark recombination (QCM) [126] model includes the possibility of having coalescence in pp collisions. Furthermore, the Catania [127] and the POWLANG transport model [128] expect the creation of a dense QGP-like medium even in pp collisions and model the hadronisation via quark coalescence, also in pp collisions. The statistical hadronisation model (SHM) [129] predicts the existence of a set of yet unobserved excited charm baryons which decay strongly and enrich the abundance of ground-state baryons. Their abundance is predicted by the relativistic quark model (RQM) [130] via their thermal densities, which depend on their mass and spin degeneracy.



ALI-PUB-567876

Figure 2.4: Ratio between the p_T -differential cross sections at midrapidity ($|y| < 0.5$) of prompt Λ_c baryons and D^0 mesons in pp collisions at $\sqrt{s} = 13$ TeV compared with the predictions from the PYTHIA Monash tune, the PYTHIA CR-BLC modes 0, 2 and 3, SHM+RQM, Catania, QCM, and POWLANG models in pp collisions at $\sqrt{s} = 13$ TeV [125].

Fragmentation is also considered in MC event generators, where it is the last stage of the parton shower. The most used event generator to model pp collisions is PYTHIA [70], which uses leading-order approximations. Another MC generator that is commonly used and which will be discussed throughout this thesis is EPOS, based on pomeron exchanges. In the following, we will introduce the basics of these two generators and the different parametrisations, known as tunes or modes, studied in this thesis.

2.1.4 PYTHIA

PYTHIA is a general-purpose event generator whose hadronisation mechanism is based on the Lund string model [131, 132]. According to this model, $q\bar{q}$ pairs are connected by a colour field represented by a string. Their interaction can be described using a Cornell potential, composed of a Coulomb-like term and a linear term, related to the string tension. This generates the following potential:

$$V(r) = \kappa r - \frac{\alpha}{r}. \quad (2.3)$$

When the distance between the quark and the antiquark increases, the Coulomb-like term becomes negligible, and the potential can be approximated as $V(r) \sim \kappa r$, with κ being a string tension of around 1 GeV/fm. Once this happens, the system tries to minimise its energy, creating a new $q\bar{q}$ pair that breaks the previous string into two smaller ones. This process to create new pairs is repeated at other points of the stretched colour string, generating pairs of particles close in phase space. At this point, the generated quarks can combine with other produced pairs to create hadrons.

Baryons are produced in the Lund string model by assuming that strings can break into quark-antiquark pairs in a $\bar{3} - 3$ colour state. A state gets created with two quarks or antiquarks. This state has a specific spin and flavour, and gets combined with a quark or an antiquark that neutralises it, creating a baryon. As a result, baryon production is strongly influenced by spin suppression factors and effective diquark masses, both of which contribute to its significant suppression.

The default PYTHIA tune is called **Monash** [133], and it includes a leading-colour (LC) approximation to trace the colour flow on an event-by-event basis. This means each partonic final state contains quarks that can only be connected to one other parton, whereas gluons can be connected to two partons via two colour strings. A schematic visualisation of the LC approximation can be found on the left panel of Fig. 2.5. In addition, Monash was tuned on e^+e^- data. Thus, this tune can describe collisions in this system, but it can differ when describing hadronic collisions at the LHC, where we could have multiple parton interactions (MPIs). In the LC approximation, each MPI evolves as a separate system, and no interactions between MPI systems are considered. The current Monash versions go one step further from a strict LC scenario, allowing different MPI initiators to be connected by colour chains. Nonetheless, cross-talk between the products of different MPIs via colour reconnection is still limited in the Monash tune. In any case, this is one of the possible reasons why this tune cannot explain the low- p_T excess of Λ_c^+ baryons over D^0 mesons in pp collisions that we introduced earlier.

The idea of having **colour reconnections (CRs) beyond the leading colour approximation** [134] was first studied in the context of the e^+e^- collisions of LEP [135, 136]. Back then, CRs were not found to have a substantial effect, since baryon-over-meson

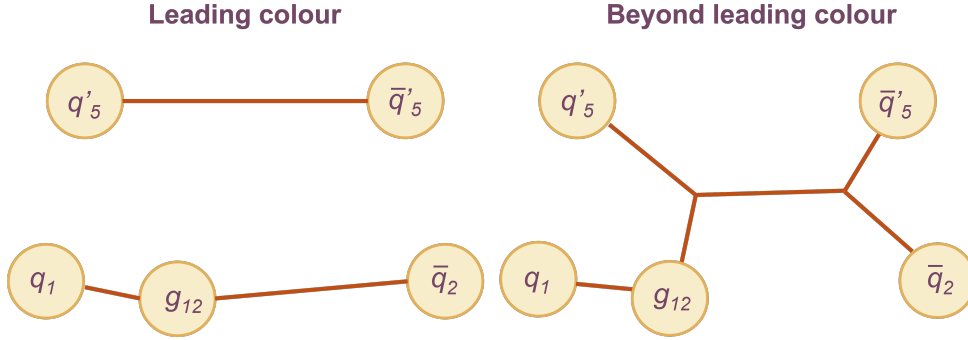


Figure 2.5: Colour strings at leading and beyond leading colour. Image adapted from [134].

ratios were expected to have a flat trend in p_T . However, their effect becomes non-negligible in hadronic collisions. In this framework, parton-parton combinations are encoded using an index ranging between 1 and 9, with the corresponding anti-index assigned for antipartons. (Anti)quarks are assigned one single colour (anti)index, while gluons have one colour index and one anti-index. This provides a total of 9 possible quark and 72 gluon states [134]. It is worth noting that these indices are not equal to the ordinary ones of $SU(3)$. Instead, they mark whether two partons form a colour singlet state, created if they are a quark and an antiquark whose index and anti-index have the same value. Otherwise, the pair of partons is in an octet state.

Confining potentials are created between any two partons containing matching colour and anticolour indices. All connected partons in the LC approximation remain connected in this framework, since they are forced to have matching colour indices. However, more combinations are now possible for the cases when two partons *accidentally* have matching indices and, for instance, belong to different MPIs. Furthermore, combinations of specific indices are also allowed, as the anticolour index is missing in the triplets $[1, 4, 7]$, $[2, 5, 8]$, and $[3, 6, 9]$. In other words, two partons with colours 2 and 5 can appear as one parton with anticolour 8. An example of the resulting topologies is shown in the right panel of Fig. 2.5.

Due to the newly available topologies, the CR model introduces new possible quark combinations or **junctions** [134]. The kind of junctions present in an event depends on the minimisation of the potential energy of the string. This expression contains parameters such as the time dilation caused by the relative particle boosts. The relevant modes in this thesis are **mode 0**, where no time-dilation constraints are included. In **mode 2**, the time dilation is calculated using the boost factor obtained from the final-state mass of the $q\bar{q}$ pairs. In this case, all pairs are required to be causally connected. Lastly, **mode 3** includes the same time-dilation constraints as mode 2, but only requires one single connection to be causally connected.

Besides colour reconnection mechanisms, we can also introduce the concept of **colour ropes** [71]. They are composed of multiple overlapping Lund strings. They can provide an alternative to hydrodynamics in small systems to explain some of the obser-

vations presented in Sec. 1.6.2.

2.1.5 EPOS 4

EPOS is a model based on the Parton-Based Gribov-Regge Theory (PBGRT) [137]. It merges the Parton Model and Gribov-Regge theory to describe both hard and soft processes using a single theoretical framework based on multiple parallel scatterings. The framework also carefully considers the energy-momentum conservation via energy sharing between the different scatterings, and includes the concepts of factorisation and saturation. These principle predictions are included in EPOS using an S-matrix approach. The resulting model is capable of handling pp, p-A and A-A collisions. A longer introduction can be seen in [138–141].

In EPOS 4, the basic interaction unit is a pomeron, which corresponds to a parton ladder in the QCD picture. When cut, a pomeron represents an inelastic interaction and is translated into a relativistic string in the hadronisation phase. Strings are $q\bar{q}$ pairs with the same flavour that are linked by a chain of gluons, forming a colour flux of the form

$$q - g - \dots - g - \bar{q}.$$

As in the case of PYTHIA, the momentum of the initial partons causes the string to get stretched as it expands, increasing its potential energy. Hence, it becomes more favourable to create a new $q\bar{q}$ pair at some point. It is worth noting that, in EPOS, the partons at the end of the chain can have any flavour, but those created at the breaking points can only be light quarks (u, d, or s). String segments can join to form colour-neutral objects called **prehadrons**, virtual particles whose evolution depends on the density of the region where they are created. In dense environments, string segments cannot hadronise. Instead, they thermalise, creating a hot, dense environment that can be described using viscous hydrodynamics. This region is called the **core**; it evolves as a fluid and hadronises via statistical hadronisation. In this model, the existence of this core with QGP-like properties is allowed even in small systems.

On the other hand, if the string density is low or the particles are highly energetic, strings can independently form hadrons. This is the case for particles located on the surface of the core, in a region called the **corona**. The existence of both the core and the corona in small systems implies that hadrons can be produced via statistical hadronisation as well as independent string fragmentation. This explains the excess of Λ_c^+ baryons over D^0 mesons with respect to the e^+e^- results, as shown in Fig. 2.4.

2.1.6 State-of-the-art of the hadron production cross sections

Several theoretical approaches are used to reproduce the heavy-flavour hadron production cross sections we see in data. A common feature is that they are performed in a fixed order of α_s . In these calculations, heavy-quark production is computed using

a factorisation formula with the following structure:

$$d\sigma^Q \simeq \sum_{a,b} f_a^A \otimes f_b^B \otimes d\tilde{\sigma}_{ab \rightarrow Q+X}, \quad (2.4)$$

where the sum is performed over all possible partonic subprocesses that do not include heavy quarks, f corresponds to the PDFs, and $d\tilde{\sigma}$ is the cross section of the process calculated with perturbation theory kept at a fixed order. Using these heavy-quark cross sections, the heavy-flavour hadronic cross section is calculated as:

$$d\sigma^H = d\sigma^Q \otimes D_Q^H(z), \quad (2.5)$$

with $D_Q^H(z)$ being a scale-independent fragmentation function which describes the hadronisation process of the heavy-quark.

Two of the state-of-the-art approaches are the FONLL (Fixed Order plus Next-to-Leading Logarithms) [106, 107] and GM-VFNS (General-Mass Variable Flavor Number Scheme) [108–113] methods. Both approaches perform their perturbative calculations at next-to-leading order (NLO) and perform a resummation of the terms at next-to-leading logarithm (NLL). FONLL is based on calculating the massive cross section at NLO, which has a high accuracy at low- p_T , plus a high- p_T term, calculated as the difference between the resummed term at NLL and the massless limit of the NLO result. This last term is introduced to avoid double-counting. Lastly, the hadronisation is treated using a fixed-order fragmentation function, which is typically extracted from e^+e^- data.

GM-VFNS calculations, on the other hand, incorporate the heavy quark mass directly into the coefficient functions and parton densities, allowing a consistent treatment of mass effects across the entire kinematic range. It uses a variable number of active quark flavours, smoothly transitioning between regimes where the heavy quark is produced in the final state or treated as an active parton in the initial state. The hadronisation in this case is performed via fits to e^+e^- data, which are then used to extract the fragmentation functions. These functions are then evolved to the scale of interest using the DGLAP evolution equations.

On the other hand, approaches such as MC@NLO [143] and POWHEG [144] combine NLO fixed-order matrix elements with parton showers, typically generated using Monte Carlo event generators like PYTHIA. These methods preserve the accuracy of the NLO calculation for inclusive observables while including the resummation of leading logarithmic (LL) soft and collinear emissions through the parton shower. They provide a fully exclusive and realistic simulation of the hadronic final state, including hadronisation and multiple parton correlations, making them especially useful for experimental analyses. However, since parton showers and hadronisation models rely on phenomenological parameters that must be tuned to data, these simulations are not purely first-principles predictions and may depend on the tuning.

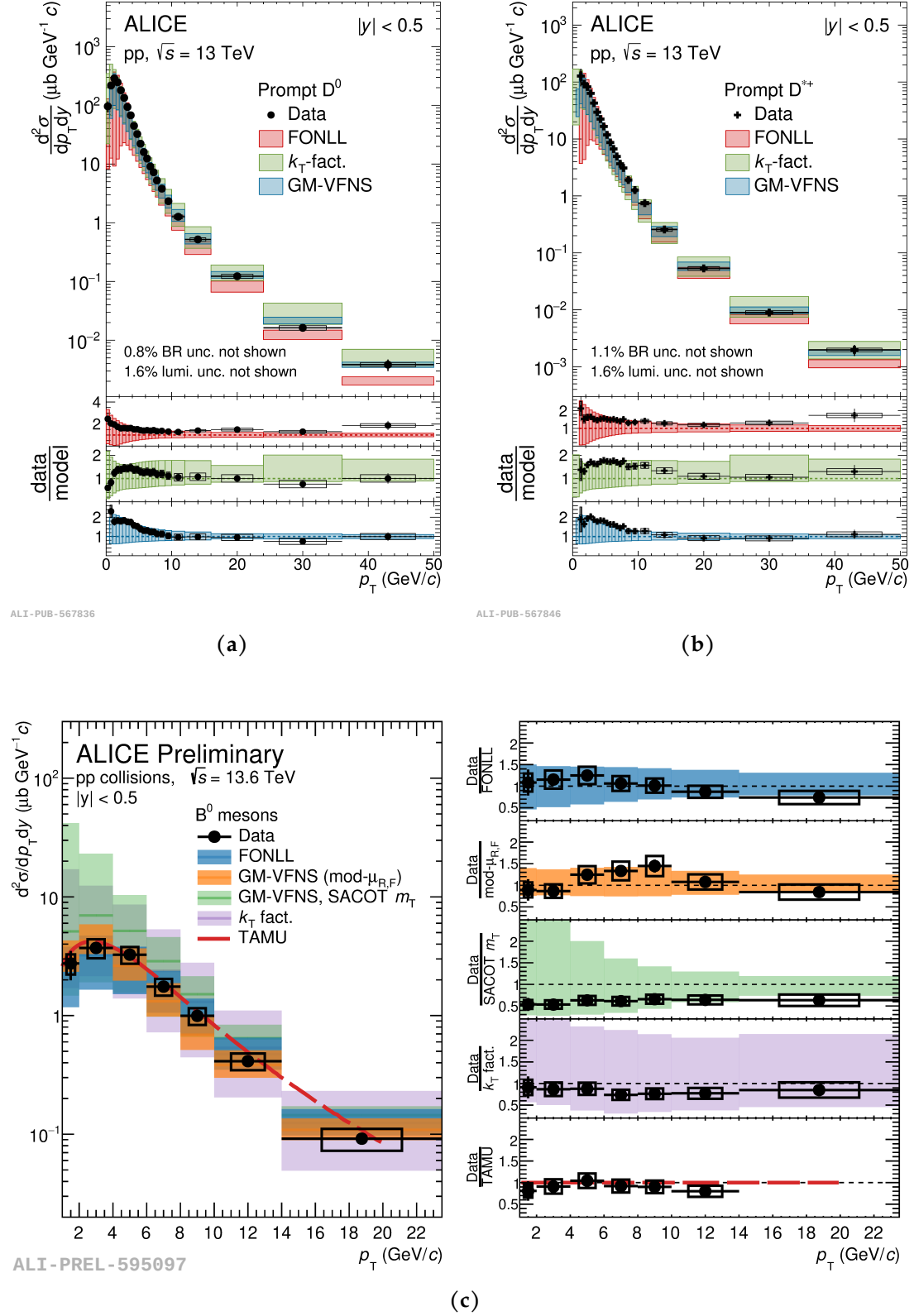


Figure 2.6: (a) D^0 cross section measured at midrapidity at $\sqrt{s} = 13$ TeV. (b) D^{*+} cross section measured at midrapidity at $\sqrt{s} = 13$ TeV [125]. (c) B^0 cross section measured at midrapidity at $\sqrt{s} = 13.6$ TeV [142]. Experimental results are compared with theoretical predictions.

For example, we present the results of the measurements of the D^0 and D^{*+} mesons at midrapidity in pp collisions at $\sqrt{s} = 13$ TeV in Figs. 2.6a and 2.6b [125]. The results are compared with several theoretical predictions, including FONLL and GM-VFNS. Both pQCD calculations agree with the data within their uncertainties for both particles at low and intermediate p_T , but FONLL tends to underestimate the data at high- p_T . Furthermore, the data points tend to lie on the upper limit of the GM-VFNS predictions at $p_T < 10$ GeV/c. The results are also compared with the predictions of the k_T factorisation method. This method implements an alternative to the collinear approach. These calculations are performed at the leading order and include higher-order corrections. The k_T predictions [145, 146] from the factorisation model are compatible with the data in all p_T ranges.

The results of the B^0 cross section at midrapidity in pp collisions at $\sqrt{s} = 13.6$ TeV are shown in Fig. 2.6c [142]. The data are compared with the predictions of FONLL, GM-VFNS, k_T factorisation and TAMU. Overall, the predictions of FONLL are more precise for B^0 mesons than for D mesons due to the larger mass of the b quark, which reinforces the use of pQCD. The GM-VFNS predictions were tested using two different sets of fragmentation functions, presented as GM-VFNS (mod- $\mu_{R,F}$) and GM-VFNS SACOT m_T . Although both sets are compatible with the data, the SACOT m_T set has a larger uncertainty at low- p_T and its central value is systematically higher than the data points. The k_T factorisation predictions are also affected by large uncertainties, and their central values are lower than the data points at all p_T ranges. Lastly, the TAMU predictions are compatible with the data at all p_T ranges. These were obtained using a $b\bar{b}$ production cross section at midrapidity of $85.3 \mu b$ and a b-quark p_T distribution extracted from FONLL calculations.

2.2 Multiple parton interactions

Multiple parton interactions, or MPIs, originate when two or more parton-parton collisions occur in a single hadron-hadron collision. MPIs were first studied when the parton model was developed, in [147, 148]. The simplest form of MPI is the case for two particles, known as **double parton scattering (DPS)**. They were found to be particularly relevant in processes whose final states involve four jets, four leptons, or a vector boson plus jets [149–152]. These channels were studied in several experiments at different collision energies, ranging from 63 GeV to LHC energies [153, 154]. Studying DPS can help probe the internal structure of the proton, since it can shed light on the correlations between the partons that compose it. In addition, MPIs constitute a necessary input for MC generators, such as PYTHIA, to adequately describe collisions at LHC energies.

In this thesis, we will focus on the study of double parton scattering through double charm production. These processes involve two $c\bar{c}$ pairs in the final state, whose production in one single parton-parton collision is suppressed. A schematic view of the

single parton scattering (SPS) and the DPS processes under consideration is located in Fig. 2.7.

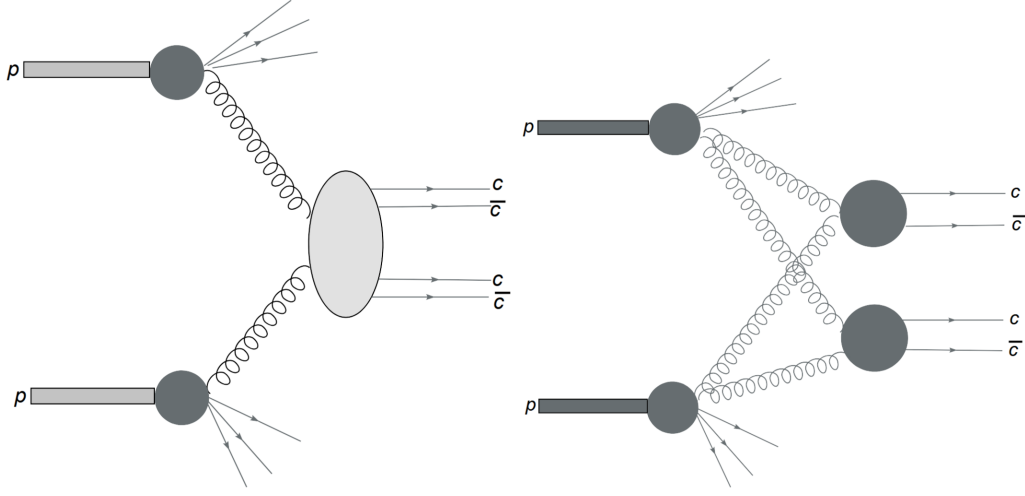


Figure 2.7: Example of a SPS (left) and a DPS (right) process involving two $c\bar{c}$ pairs. Figure taken from [155].

In the factorisation approach, we can describe double parton scattering processes involving two independent parton scatterings, A and B, as:

$$\sigma_{\text{DPS}}^{A,B} = \sum_{ijkl} \int dx_1 \int dx_2 \int d^2\mathbf{y} F_{ik}(x_1, x_2, \mathbf{y}) \otimes \hat{\sigma}_{ij}^A \hat{\sigma}_{kl}^B \otimes F_{jl}(x_1, x_2, \mathbf{y}), \quad (2.6)$$

where $\hat{\sigma}_{ij}$ is the partonic cross section and $F_{ik}(x_1, x_2, \mathbf{y})$ is the double parton density (DPD) [156]. Assuming that there are no correlations between partons, the DPDs can be defined in the context of generalised parton distributions (GPDs) as

$$F^{ij}(x_1, x_2, \mathbf{y}) \rightarrow \int d^2\mathbf{b} f^i(x_1, \mathbf{b}) f^j(x_2, \mathbf{b} + \mathbf{y}). \quad (2.7)$$

In this case, $\mathbf{b}$ represents the impact parameter, or the distance between the centre of the hadron and the parton, and $f^{i,j}$ represents the GPDs. Generalised parton distributions are characterised by having a longitudinal and a transverse component, as opposed to PDFs, in which the transverse component is integrated. We can write the GPDs as the product of a PDF and a function $G(\mathbf{b})$, which contains the information relative to the impact parameter, resulting in

$$f^i(x_1, \mathbf{b}) = f^i(x_1) G(\mathbf{b}). \quad (2.8)$$

Using these approximations, the DPS cross section can be expressed as

$$\sigma_{\text{DPS}}^{A,B} = \frac{m}{2} \frac{\sigma_A \sigma_B}{\sigma_{\text{eff}}}. \quad (2.9)$$

This equation is commonly called the **pocket formula**, and parametrises the transverse

overlap function between the partons of the proton. The factor m is a symmetry factor which equals one if $A = B$ and two otherwise. σ_{eff} is an effective cross section that was originally assumed to be universal across channels and collision systems. This parameter is related to the internal structure of partons within the proton, and a number of effects can invalidate the assumption of universality, as discussed in [157].

The size of the DPS cross section can be roughly estimated via a power counting, resulting in

$$\sigma_{\text{DPS}} \sim \sigma_{\text{SPS}} \frac{\Lambda_{\text{QCD}}^2}{Q^2}. \quad (2.10)$$

This expression demonstrates that, although the DPS cross section is generally suppressed with respect to the SPS one, it can compete if the SPS process is suppressed or if the energy of the collision is high, which grants access to lower Bjorken- x values. The evolution of the DPS cross section for the case of double charm production was calculated in [155] and found to grow faster than the SPS cross section as a function of the centre-of-mass energy. In fact, at LHC energies, the DPS cross section for $pp \rightarrow c\bar{c} c\bar{c}$ processes becomes larger than the SPS cross section. This behaviour can be observed in Fig. 2.8.

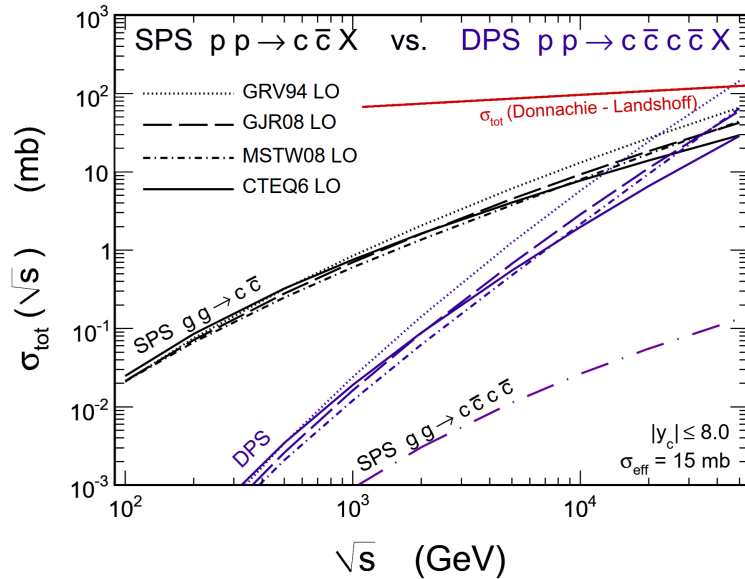


Figure 2.8: Total cross section of DPS events in the process $pp \rightarrow c\bar{c} c\bar{c}$ (in purple) compared to the cross section of the SPS processes corresponding to $pp \rightarrow c\bar{c}$ (in black) and $pp \rightarrow c\bar{c} c\bar{c}$ (in purple, dashed) as a function of $\sqrt{s}$ [155].

A summary of the measurements of σ_{eff} up to date can be observed in Fig. 2.9 [158]. Overall, the processes produced by gluon fusion, such as those involving charm mesons, tend to have a smaller effective cross section (~ 5 mb) than quark-mediated measurements (~ 20 mb), as the channels involving vector bosons or jets. This suggests that the effective cross section might not contain all the physical information. Some of the possible corrections that might be missing in the formulation of σ_{eff} are the non-

perturbative correlations or the correlations in spin and colour between partons. Furthermore, a study performed in [157] revisited the DPS measurements available at the time and extracted the σ_{DPS} using a modified pocket formula:

$$\sigma_{\text{DPS}} = \sum_{i,j;k',l'} \frac{\sigma_{ik'}(A)\sigma_{jl'}(A)}{\sigma_{k'l'}^{ij;\text{eff}}}. \quad (2.11)$$

This formulation allowed for the effective cross section to vary depending on whether the quarks involved are valence or sea quarks. Fitting these parameters to the existing data, they found that processes which only involve sea quarks tend to have a smaller effective cross section than those which include at least one valence quark. With this procedure, they were able to exclude the hypothesis of a universal σ_{eff} by 3.8σ .

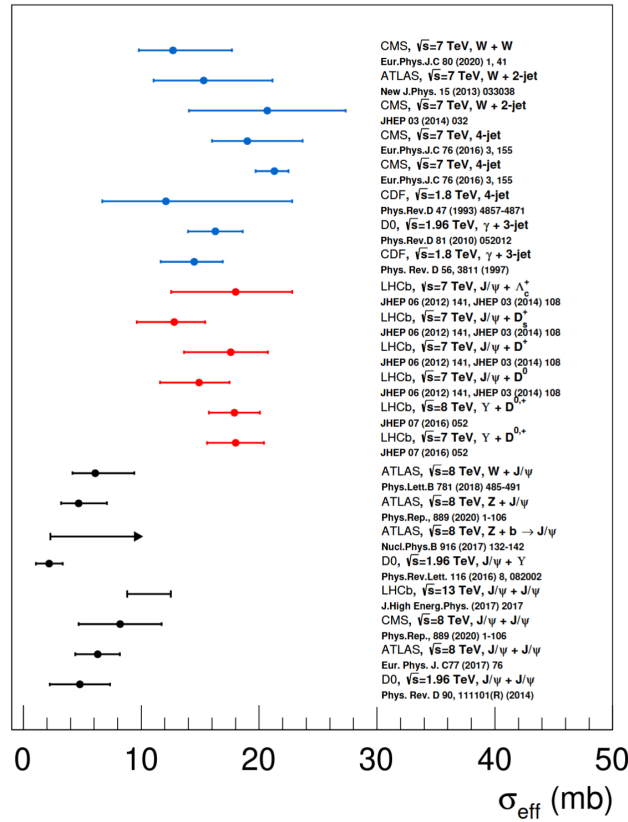


Figure 2.9: Summary of the measurements of σ_{eff} up to date. Figure from [158].

In the context of double charm production, LHCb measured the double production of J/ψ mesons alongside several D mesons and Λ_c^+ baryons in [159]. The results of the pair cross sections can be found in Fig. 2.10. They are separated into three categories: the first one corresponds to the cross sections of pairs which are mainly composed of SPS events, such as $D^0 \bar{D}^0$. These are pairs which can be produced from the hadronisation of a single $c\bar{c}$ pair. On the other hand, the second and third categories include pairs which are primarily produced via DPS processes. These involve the production of same-sign pairs, such as $D^0 D^0$, which require the production of two same-sign

charm quarks. At last, the production of J/ψ together with an open charm hadron is included. These pairs require the production of at least three charm or anticharm quarks, making them unlikely to originate in an SPS. They are compared with the theoretical predictions from Refs. [160–162], which tend to underestimate the data. Overall, the cross section of opposite-sign pairs is systematically higher than that of pairs which are mainly produced through DPS processes. As a last note on double

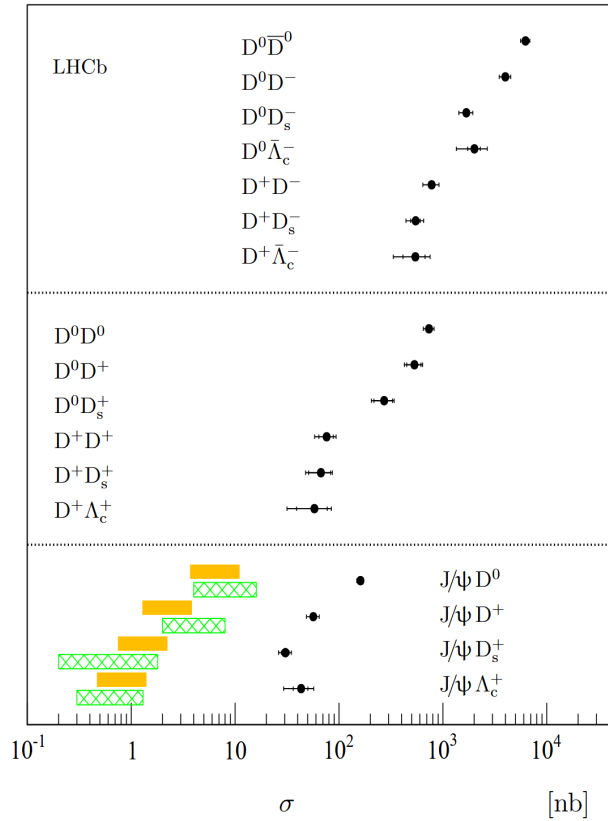


Figure 2.10: Summary of the measurements of the double cross sections measured by LHCb in [159]. On the upper pannel, the opposite-sign open charm hadron pairs are presented. In the middle pannel, the like-sign open charm hadron pairs are included, and the last pannel corresponds to the pairs composed of a J/ψ meson and an open charm hadron. These last pairs are compared to theoretical predictions.

parton scattering, it is worth mentioning that DPS events populate the phase space in a different way from SPS events. For instance, SPS events prefer smaller values of the pair $|\Delta y|$ than DPS ones, for which the distribution can be approximated by considering that there are no correlations between the two hadrons. DPS events display a flat $|\Delta\phi|$ distribution, contrary to SPS events, which peak around π . An example of these distributions measured by the LHCb collaboration at $\sqrt{s} = 13$ TeV for the case of double J/ψ production can be found in Fig. 2.11 [163]. As can be seen, the measured $|\Delta y|$ distribution cannot be described without invoking the DPS contribution. Similarly, the data of the $|\Delta\phi|$ distribution is compatible with a straight line.

The study of DPS in different rapidity regions can also be used to study production asymmetries, which could point to the existence of parton correlations between par-

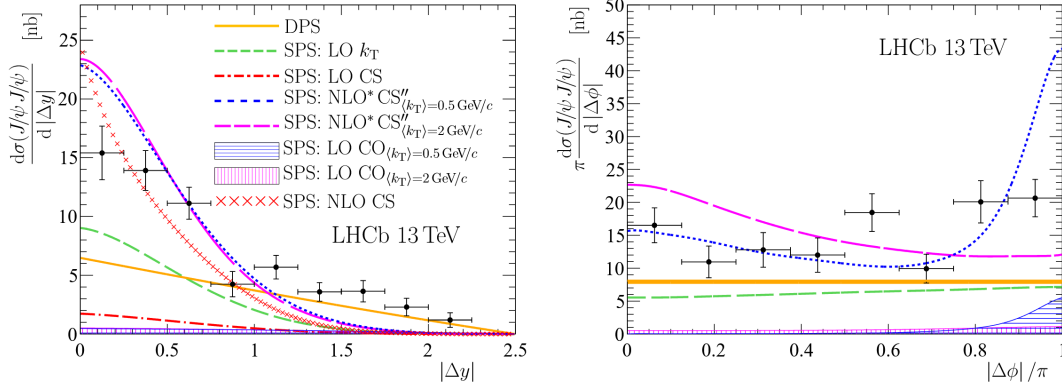


Figure 2.11: Left pannel: Double J/ψ cross section per unit of $|\Delta y|$ as a function of $|\Delta y|$. Right pannel: Double J/ψ cross section per unit of $|\Delta\phi|$ as a function of $|\Delta\phi|$. The measurements are compared to the expected SPS and DPS distributions. Plots extracted from [163].

tons from different partonic collisions. Thus, it would be possible to study double parton scattering processes beyond the pocket formula. For instance, asymmetries in the production between the case where both processes occur in the same direction versus opposite directions were studied for the case of double W production in [164, 165]. In this study, it was found that an asymmetry could be present in the case where spin correlations between the vector bosons are as large as possible.

2.3 Cold nuclear matter effects in heavy-ion collisions

Collision systems involving heavy ions exhibit cold nuclear matter effects, which are present in both p–A and A–A collisions, but are absent in pp interactions. In A–A collisions, the simultaneous presence of hot nuclear matter effects, arising from the formation of a QGP medium, and cold nuclear matter effects complicates the disentanglement of their respective contributions. In contrast, p–A collisions are of particular interest since the formation of a large QGP is not expected in these systems. However, recent experimental observations suggest that features typically attributed to QGP formation, such as collective behavior, may also emerge in high-multiplicity p–A collisions (see Sec. 1.6.2). Nevertheless, p–A systems remain a valuable reference for constraining cold nuclear matter effects, due to the reduced presence of hot QCD medium effects compared to A–A collisions. In this section, we will discuss these aspects in detail.

One of the features of heavy-ion collisions (p–A and A–A) is the modification of the partonic structure of nucleons within nuclei, which manifests in both the parton distribution functions (PDFs) and structure functions. A commonly used observable to quantify such nuclear effects is the ratio:

$$R(x, Q^2) = \frac{F_2^A(x, Q^2)}{AF_2^N(x, Q^2)}. \quad (2.12)$$

This ratio is obtained by dividing the structure function of a nucleus composed of A nucleons ($F_2^A(x, Q^2)$) and the nucleus structure function ($F_2^N(x, Q^2)$) multiplied by the number of nucleons. If there were no CNM effects, this ratio would be equal to unity. The resulting distributions are called **nuclear PDFs** (nPDFs), and we can distinguish several regions by comparing them with PDFs [166]. These are represented in Fig. 2.12 as a function of the Bjorken- x [167].

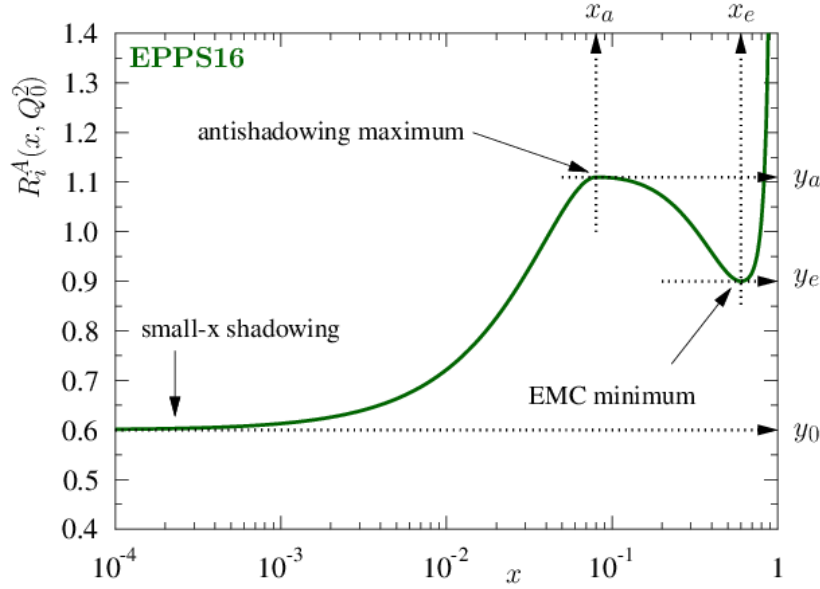


Figure 2.12: Sketch of the EPPS16 parametrisation of the nuclear PDFs as a function of the Bjorken- x [167].

1. **Shadowing:** it is located at $x \lesssim 0.05$ and is characterised by a reduction of the PDF values in nuclei compared to free nucleons ($R < 1$). In this region, the main contribution to the cross section comes from sea quarks and gluons. The shadowing is found to increase with decreasing x , though the behaviour is compatible with a saturation at the smallest available values of x [168]. Furthermore, the level of shadowing increases with the mass number of the nucleus and decreases with increasing momentum transfer squared. The presence of shadowing reflects the onset of nonlinear QCD dynamics, which, in the Colour Glass Condensate framework (see Sec. 2.1.1), arises from gluon saturation.
2. **Anti-shadowing:** $0.05 \lesssim x \lesssim 0.3$. The ratio shows a small increase of a few per cent above unity. This effect is linked to the necessity to conserve the momentum. The reduction of the gluon density caused by the shadowing at low- x requires an excess at intermediate p_T .
3. **EMC effect:** this effect was discovered by the European Muon Collaboration, [169]. A small suppression is found at $0.3 \lesssim x \lesssim 0.8$, with a local minimum at $x \sim 0.7$. However, as opposed to the shadowing region, the contribution of the sea quarks is almost negligible; most of the effect is due to the valence quarks.
4. **Fermi motion:** At $x \gtrsim 0.8$, the ratio increases rapidly with x due to the Fermi

motion of the bound nucleons in the nucleus. This effect is mainly kinematic, because the cross section vanishes at $x \rightarrow 1$.

Heavy-flavour hadrons in heavy-ion collisions at the LHC are mostly affected by the shadowing effect. At midrapidity, $b\bar{b}$ and $c\bar{c}$ pairs are produced in hard scatterings between partons with Bjorken- x values of 10^{-3} and 10^{-4} , respectively. Furthermore, the low- p_T region dominates their production, and its yield normalised to the number of nucleon-nucleon collisions is found to be suppressed with respect to the results in pp [170, 171]. The predictions of nuclear PDFs are parametrised using Drell-Yan and DIS measurements [172, 173].

Another CNM effect is the **Cronin enhancement**, which causes the nuclear modification factor to be slightly greater than one in p-A collisions at intermediate p_T . This effect is produced by a shift of the hadron p_T distribution to higher values compared to pp collisions. It has been interpreted as arising from the multiple scatterings of partons from the proton off partons from the nucleus [174]. As a result, the partons experience a transverse momentum kick, shifting the distribution towards higher values, and thus explaining the depletion and enhancement in the distribution.

2.4 Hot nuclear matter effects in heavy-ion collisions

Heavy-flavour quarks are produced at the beginning of the collision, before the creation of the QGP medium, and interact with it as they pass through. These interactions cause the HF quark propagation to be described as a Brownian motion characterised by small momentum kicks [175]. Those kicks are produced by elastic and inelastic scatterings, which are commonly referred to as **collisional** and **radiative** scatterings, respectively. The expected contribution of collisional and radiative scatterings to the total energy loss per heavy-quark is presented in Fig. 2.13 [176]. These predictions were made for central collisions (0-7.5%) in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. They show that collisional scatterings account for most of the energy loss at low p_T , while radiative energy losses dominate at high p_T . The cross-over between the two contributions takes place at $E \sim m_{c,b}^2 \text{ GeV}/c^2$ ($m_c \sim 1.5 \text{ GeV}/c^2$, $m_b \sim 4.2 \text{ GeV}/c^2$). Theoretical frameworks that incorporate at least one of these mechanisms and model the in-medium evolution of heavy quarks are commonly referred to as **transport models**.

The existence of the QGP also affects the hadronisation, which can still be produced via fragmentation, as introduced in Sec. 2.1.3, but also via recombination. In this section, these processes will be briefly introduced.

2.4.1 Collisional processes

The effect of the multiple elastic scatterings that occur as the heavy quark goes through the medium results in an energy loss ΔE_{coll} . If we assume that the energy of the quark is much larger than its mass, $E_Q \gg m_Q^2 T$, this energy loss is expected to evolve log-

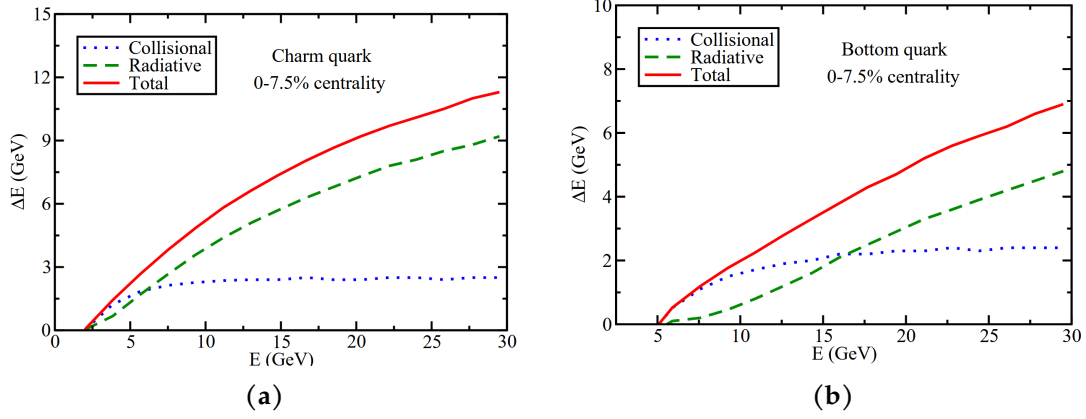


Figure 2.13: Expected contribution to the total energy loss (in red) produced by collisional (in blue) and radiative processes (in green) for (a) charm and (b) beauty quarks in central Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV [176].

arithmically with E_Q and linearly with the thickness of the medium [175]. At low energies, heavy quarks interact with the medium via scatterings, resulting in a random, Brownian motion-like trajectory, and their movement becomes influenced by the collective flow of the medium.

The dynamics and space-time evolution of heavy quarks can be described using the Boltzmann equation in the transport formalism:

$$\left[\frac{\partial}{\partial t} + \frac{\mathbf{p}}{E_p} \frac{\partial}{\partial \mathbf{x}} + \mathbf{F} \frac{\partial}{\partial \mathbf{p}} \right] f_Q(t, \mathbf{x}, \mathbf{p}) = C[f_Q], \quad (2.13)$$

where E_p represents the quark energy, $\mathbf{F}$ encodes the forces generated by the electromagnetic and strong fields, and f_Q provides the heavy-quark phase-space distribution function. Finally, $C[f_Q]$ is a collisional integral which contains the parton scattering amplitudes. Assuming that the momentum kicks are small enough to generate a Brownian motion, which corresponds to the limit where $m_Q \gg T$, the Boltzmann equation can be approximated by the Fokker-Planck equation:

$$\frac{\partial}{\partial t} f_Q(t, \mathbf{p}) = \frac{\partial}{\partial p_i} \left[A_i(\mathbf{p}) f_Q(t, \mathbf{p}) + \frac{\partial}{\partial p_j} (B_{ij}(\mathbf{p}) f_Q(t, \mathbf{p})) \right] \quad (2.14)$$

The Fokker-Planck equation is assumed to be valid for heavy quarks because their masses are significantly larger than $T_{QGP} \sim 300$ MeV [175]. It corresponds to a differential equation that depends on the drift and diffusion coefficients, $A_i(\mathbf{p})$ and $B_{ij}(\mathbf{p})$. In particular, the drift coefficient (A_i) can be interpreted as the friction of the quark as it moves through the medium, and the diffusion coefficient (B_{ij}) represents the average momentum broadening.

In the non-relativistic limit, the Fokker-Planck equation can be reduced to the Langevin

equation:

$$\frac{\partial}{\partial t} f_Q(t, \mathbf{p}) = \gamma \frac{\partial}{\partial p_i} [p_i f_Q(t, \mathbf{p})] + D_p \nabla_p f_Q(t, \mathbf{p}). \quad (2.15)$$

In this case, $\gamma = A_i(\mathbf{p})$ and D_p is the momentum diffusion coefficient. The γ factor is also the inverse of the quark thermalisation coefficient, τ_Q^{eq} . This coefficient is estimated to be around 5 fm/c for charm quarks and around 15 fm/c for beauty quarks, although the specific values can vary significantly between models, since it depends on the friction coefficient, $A_i(\mathbf{p})$. Thus, the HF thermalisation time is comparable to the QGP lifetime at the LHC.

2.4.2 Radiative processes

Energetic partons traversing the QGP mainly lose energy by emitting gluons. One possible description of this phenomenon is provided by the BDMPS (Baier-Dokshitzer-Mueller-Peigné-Schiff) formalism [177]. This formalism assumes that the parton, which propagates through the medium, interacts with the medium constituents, which remain static. In each interaction, the parton gains some momentum, k_T , until it obtains enough momentum to emit a gluon. The resulting energy loss is proportional to:

$$\Delta E_{\text{rad}} \propto \alpha_s C_R \hat{q} L^2, \quad (2.16)$$

where C_R is the Casimir factor, which is 4/3 for the quark-gluon coupling and 3 for the gluon-gluon one, $\hat{q} = \langle k_T^2 \rangle / L$ is the average of the momentum transferred to the parton per unit of distance travelled by the particle in the medium, L . As a consequence of the Casimir factors, if a gluon and a quark travel the same distance through the same medium, the gluon will lose $C_R^{(g)} / C_R^{(q)} = 9/4$ times more energy than the quark by radiation.

The radiative losses also depend on the quark mass as a consequence of the **dead cone effect** [178, 179]. This effect, which is caused by kinematic constraints, forbids gluons to be emitted in the angular region $|\theta| < m_Q / E_Q$. This implies that the radiative energy loss follows the following hierarchy:

$$\Delta E_g > \Delta E_{u,d,s} > \Delta E_c > \Delta E_b. \quad (2.17)$$

Similarly, a dependence on the quark mass is also expected for collisional losses, although it is less pronounced [180].

2.4.3 Heavy-flavour hadronisation in Pb–Pb collisions

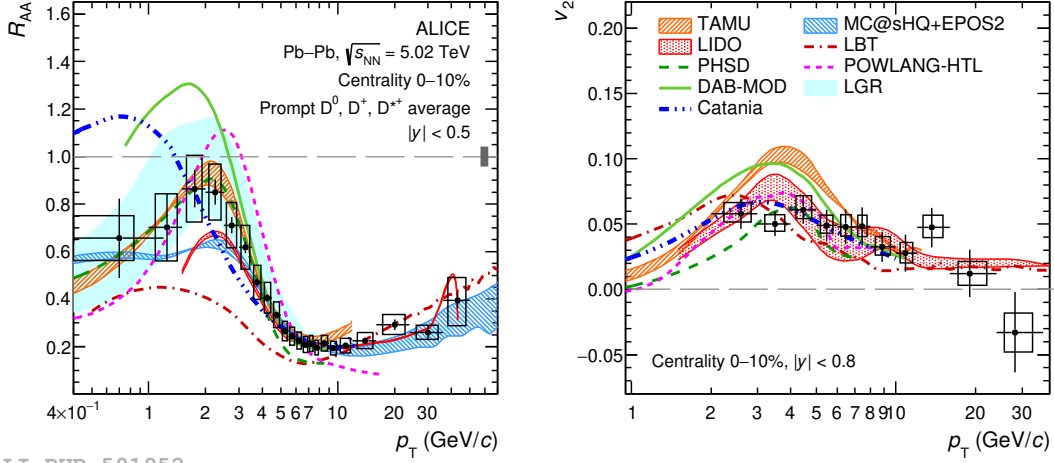
In heavy-ion collisions, the process of hadronisation (already defined in Sec. 2.1.3) is not only driven by the **fragmentation**, but also by the **recombination** (or **coalescence**) of hadrons. The increased density of colour charges in the QGP medium increases the probability for heavy quarks to combine with light quarks present in the medium,

which are close in phase space. This mechanism is relevant for low and intermediate- p_T hadrons, as opposed to fragmentation, which is dominant in high- p_T particle production. The relative abundance of baryons with respect to mesons is also expected to increase due to the recombination. Note that, as we discussed in Sec. 2.1.3, this effect is not exclusive to heavy-ion collisions, as it also present in pp collisions. This presence could be attributed to the formation of a QGP-like medium, as suggested by models such as Catania or POWLANG. However, other models, such as Herwig [181], which implements a cluster hadronisation mechanism, can describe the data without invoking the presence of a medium.

Different implementations of the recombination mechanism have been included in transport models:

- **Instantaneous coalescence:** This model was originally created to explain the production of light-nuclei from the coalescence of nucleons on a freeze-out hypersurface [182]. It was later expanded to describe meson and baryon formation from quark hadronisation in the QGP through $2 \rightarrow 1$ and $3 \rightarrow 1$ processes [183]. The model assumes that the phase space at the time of hadronisation is filled with partons. At that point, a very rapid freeze-out takes place on an infinitely thin hypersurface, where the recombination of nearby partons occurs. The resulting quarks are dressed, and the gluons are split into $q\bar{q}$ pairs. The recombination probability is calculated using the quark spatial and momentum distributions and a Wigner function. This formalism is used in models such as PHSD [184], LBT [185], and Catania [186, 187].
- **Resonance recombination model (RRM):** This model is based on the idea of treating the recombined mesons as resonances [188]. With this procedure, the energy is conserved at all stages of the recombination thanks to the finite resonance width. The formation and dissociation rates are estimated from the Boltzmann equation, and their transport properties are calculated using the Langevin equation. This model is implemented by models such as TAMU [189].
- **In-medium string formation:** In this case, the hadronisation process is explained as a fragmentation of strings. These strings have a similar structure to the one introduced in Sec. 2.1.4, and their endpoints are composed of a heavy (anti)quark, Q ($\bar{Q}$), and a thermal parton, $\bar{q}$ (q) from the medium [190]. This method is implemented in POWLANG.

The effects of recombination are relevant up to a higher p_T for more massive quarks. In other words, heavy-flavour quarks are affected by recombination in a larger p_T range than light quarks, and the same is true for beauty quarks with respect to charm. The effects of hadronisation in Pb–Pb collisions can be investigated by comparing model predictions of both the nuclear modification factor (R_{AA}) [171] and elliptic flow coefficient (v_2) [191] to experimental data, as illustrated in Fig. 2.14. By testing whether different hadronisation mechanisms can simultaneously describe both observables, one



ALI-PUB-501952

Figure 2.14: Average R_{AA} (left) and elliptic flow (right) of prompt D mesons in the 0-10% centrality class measured at $\sqrt{s_{NN}} = 5.02$ TeV [171]. The experimental data is compared with model predictions.

can assess the consistency and validity of their implementations within a given model. The average nuclear modification factor for D mesons measured in central Pb-Pb collisions is presented on the left, and their corresponding elliptic flow is displayed on the right. These results are compared to several transport models, including TAMU, PHSD, Catania and POWLANG. Even though their recombination mechanisms differ, as we have already introduced, they are all capable of describing the data of the elliptic flow and the R_{AA} at high- p_T . The main discrepancies are found on the R_{AA} at $p_T < 10$ GeV/c : the Catania and DAB-MOD [192] models overestimate the data at low- p_T , and the LBT model underestimates the data up to intermediate- p_T . In general, most models are unable to simultaneously describe the R_{AA} and the elliptic flow, showcasing a need for further studies of the hadronisation mechanisms of heavy quarks in the high-density QCD medium.

A LARGE ION COLLIDER EXPERIMENT – ALICE

This Chapter aims to explain the technicalities of the experimental apparatus used to obtain the data employed in this thesis. The Large Hadron Collider (LHC), ALICE, and its sub-detectors will be introduced. Special attention will be paid to the central barrel region, where most of the data analysed in this thesis was recorded, and the muon spectrometer. Finally, the current status of the ALICE analysis framework will be introduced, as well as the strategies followed in Run 3 for data taking.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [193] is a particle accelerator with two separate beam pipes located at CERN on the border between Switzerland and France. It is situated between 45 and 170 m underground, in the tunnel where the Large Electron Positron (LEP) collider was built. It has a circumference of 26.7 km, making it the largest particle accelerator up to date. It was mainly created to collide protons with a design centre-of-mass energy of $\sqrt{s} = 14$ TeV and a luminosity of $10^{34} \text{ cm}^{-2} \cdot \text{s}^{-1}$. It can also accelerate heavy ions. In particular, it has accelerated Pb ions to produce collisions which can reach a $\sqrt{s_{\text{NN}}} = 5.5$ TeV and a luminosity of $10^{27} \text{ cm}^{-2} \cdot \text{s}^{-1}$. It began its operations in September 2008. Since then, it has gone through three data-taking periods, called Runs, and two long shutdown (LS) periods, during which upgrades and maintenance work have occurred.

Before reaching the LHC, each particle must go through smaller accelerators, progressively gaining more energy. Proton production begins at LINAC4, where negatively ionised hydrogen atoms are accelerated until they reach energies of 160 MeV. Afterwards, they enter the Proton Synchrotron Booster, where the H-ions are stripped of their two electrons and reach an energy of 2 GeV. Then, they enter the Proton Synchrotron, which can make protons reach energies up to 26 GeV. They continue by entering the Super Proton Synchrotron (SPS), where they get to 450 GeV and, finally,

reach the LHC. In the case of heavy ions, production starts at LINAC3 and continues at the Low Energy Ion Ring (LEIR). Here, ion bunches are prepared and accelerated until they achieve energies between 4.2 MeV and 72 MeV. At this point, they pass onto the Proton Synchrotron and continue on the same sequence of accelerators as proton bunches. In each step, the atoms are stripped of electrons, leaving a final beam composed exclusively of ions.

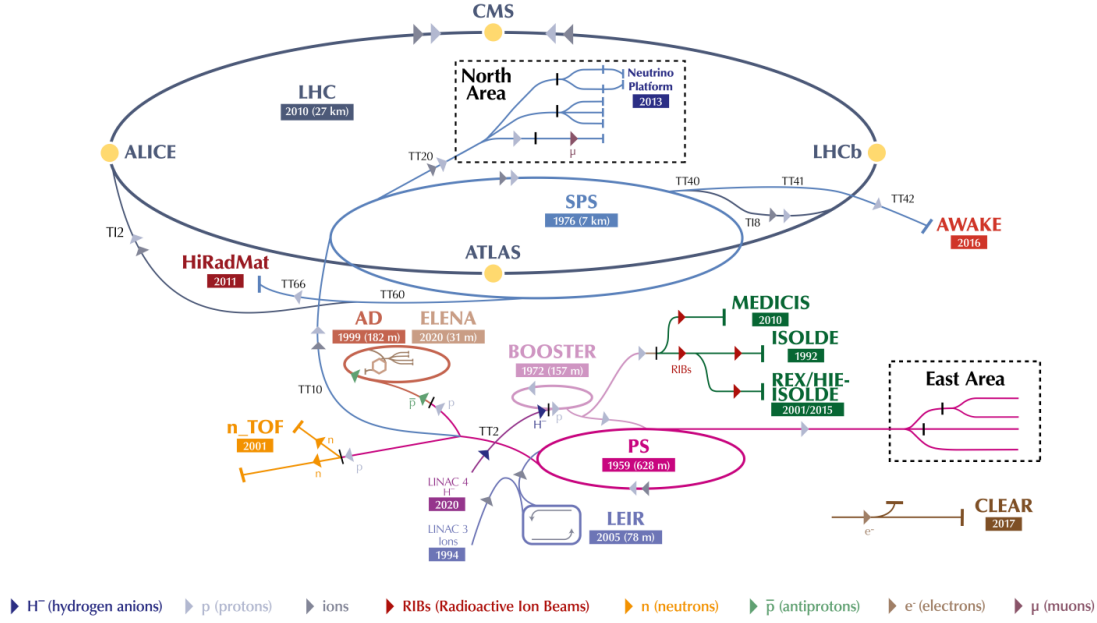


Figure 3.1: Schematic view of the LHC complex with its different accelerators and experiments.

The LHC is not a perfect circle. Instead, it is composed of eight 2.54 km-long arcs and eight 545 m-long straight sections. The arcs contain a set of superconducting electromagnets, which curve the beams. The LHC is structured in eight octants, which expand from the middle of an arc to the middle of the following arc, including a full straight section. Figure 3.1 shows a schematic view of the LHC. Four major experiments are located at the LHC in the four beam interaction points (IP). These are the following:

- **A Large Ion Collider Experiment (ALICE)** [194]: it is located at IP2, and its main purpose is to study heavy-ion collisions to obtain more insight into the QCD sector of the Standard Model. To be more precise, it studies the physics of strongly interacting matter, as well as quark-gluon plasma that can be reached via heavy-ion collisions. It will be explained in more detail in Section 3.2.
- **A Toroidal LHC ApparatuS (ATLAS)** [195] and **Compact Muon Solenoid (CMS)** [196]: they are general purpose experiments located at IP1 and IP5, respectively. Their main objective was to find the Higgs boson, which they found in 2012 [6, 7]. Since then, they have studied its properties. They are also searching for sources of new physics beyond the Standard Model.
- **Large Hadron Collider beauty (LHCb)** [197]: it is situated at IP8, and its main

purpose is to study CP violation and rare decays of beauty and charm hadrons. Its objective is to search for sources of CP violation outside of the Standard Model since they are required to explain the matter-antimatter asymmetry we see in the Universe nowadays.

In addition, other experiments are located in the LHC complex. These are TOTEM [198], LHCf [199], MoEDAL-MAPP [200], FASER [201], and SND@LHC [202].

3.2 ALICE

A Large Ion Collider Experiment, ALICE, is one of the four main experiments located at the LHC. It is a general-purpose experiment dedicated primarily to studying nuclear matter in extreme temperature and pressure conditions. For this reason, it was built to cope with large multiplicities of charged particles produced at midrapidity in Pb–Pb collisions, being optimised for a charged particle multiplicity density value of $dN/d\eta = 4000$. ALICE can also reconstruct particles down to very low transverse momentum values. For instance, it can reconstruct pions down to 80 MeV/ c and has suitable particle identification capabilities up to $p_T \sim 20$ GeV/ c . It has dimensions of $16 \times 16 \times 26$ m³ and an approximate weight of 10,000 t, primarily located at its magnet. Several upgrades were performed during the Long Shutdown 2 (LS2), which took place between 2019 and 2021, in order to exploit the full capabilities of the LHC in Runs 3 and 4. The main objectives of these upgrades involve improving tracking and vertexing, which can be achieved with the implementation of new tracking systems, such as the ITS2 (section 3.2.1) and the MFT (section 3.2.3). In addition, the upgrades aimed at handling Pb–Pb collisions with a collision rate up to 50 kHz in continuous readout and adding the possibility to use GPUs for online reconstruction. The resulting detector can be seen in Fig. 3.2, as well as the 18 subdetectors that compose it.

The coordinate system of ALICE is centred around the interaction point. Its z -axis is parallel to the beamline and is positive in the direction towards IP1. The x -axis points towards the centre of the LHC ring, and the y -axis points upwards.

ALICE is composed of 18 subdetectors, which can be separated into several groups. First, the **central barrel region** is introduced. This region is encapsulated by a solenoid magnet taken from the L3 experiment that provides a magnetic field of $B \sim 0.5$ T and covers the pseudorapidity region $-0.9 < \eta < 0.9$. They are mostly in charge of identifying charged particles, as well as photons. Detectors in this region include the Inner Tracking System (ITS), the Time Projection Chamber (TPC), the Transition Radiation Detector (TRD), the Time-of-Flight (TOF) detector, a set of calorimeters (DCal, PHOS, CPV, EMCal), and the High Momentum Particle Identification Detector (HMPID). Next, we have a set of detectors used for muon identification. They are included in the **muon spectrometer**, which covers the pseudorapidity window $-4 < \eta < -2.5$. The

muon spectrometer comprises a set of absorbers, a dipole magnet, the Muon Chambers (MCH), and the Muon IDentifier (MID) system. Next to the interaction point, the Muon Forward Tracker (MFT) is used for muon tracking and vertexing. Finally, we have some **forward detectors**, such as the Zero Degree Calorimeter (ZDC) and the Fast Interaction Trigger (FIT), which are used for triggering, multiplicity and centrality determination. The following will introduce the characteristics of the main detectors employed for the analyses performed in this thesis.

3.2.1 Central barrel detectors

Inner Tracking System

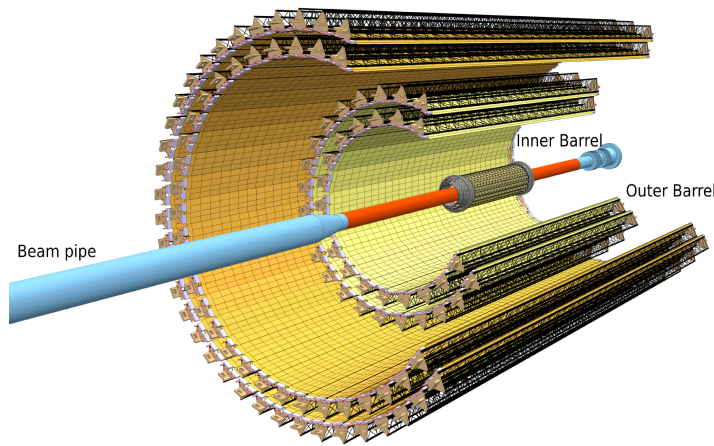


Figure 3.3: Layout of the ITS2 [203].

The Inner Tracking System (ITS) [204] is the innermost detector of ALICE and was a major focus of upgrades during the second LHC long shutdown [205]. It consists of seven concentric layers of Monolithic Active Pixel Sensors (MAPS), as illustrated in Fig. 3.3, covering a total active area of approximately 10 m². The upgraded ITS, or ITS2, is structured into three Inner Layers, two Middle Layers, and two Outer Layers, covering a radial extension from 22 mm to 430 mm with respect to the beamline. The first three layers form the Inner Barrel, while the remaining four constitute the Outer Barrel. The main improvements introduced in the ITS2 with respect to the older ITS (ITS1) are reported in table 3.1. The diameter of the beampipe was also reduced, allowing for the first layer of the ITS2 to be located closer to the interaction point.

The main objectives of the ITS are:

- To determine the position of the collision (primary vertex) with a resolution better than 5 μm .
- To reconstruct secondary vertices coming from the decays of B and D mesons.
- To track and identify particles with a momentum lower than 200 MeV/c.
- To improve the momentum and angle resolution of the particles reconstructed with the TPC.

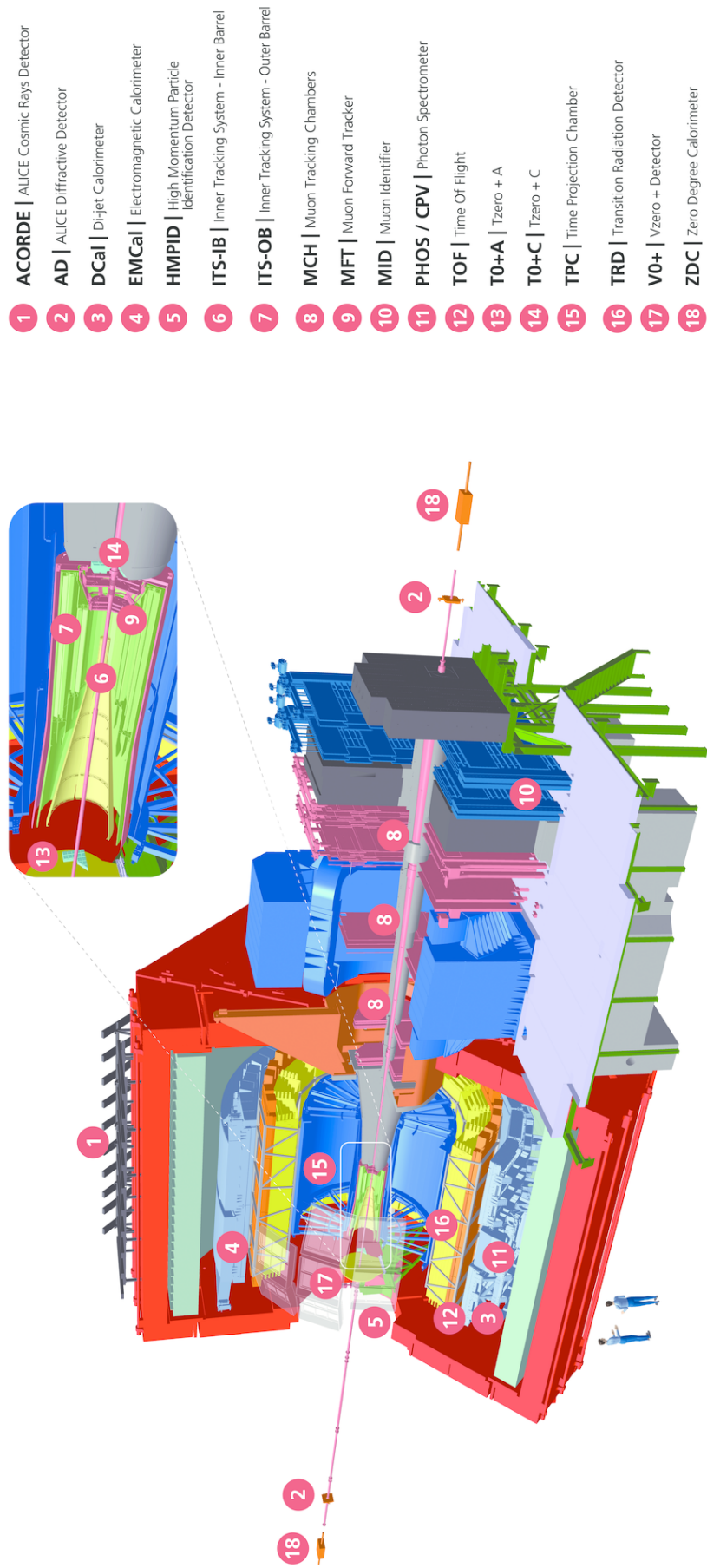


Figure 3.2: Overview of ALICE in Run 3. Image taken from the ALICE image repository.

Table 3.1: Table showing the main upgrades of ITS2 compared to ITS1 [203].

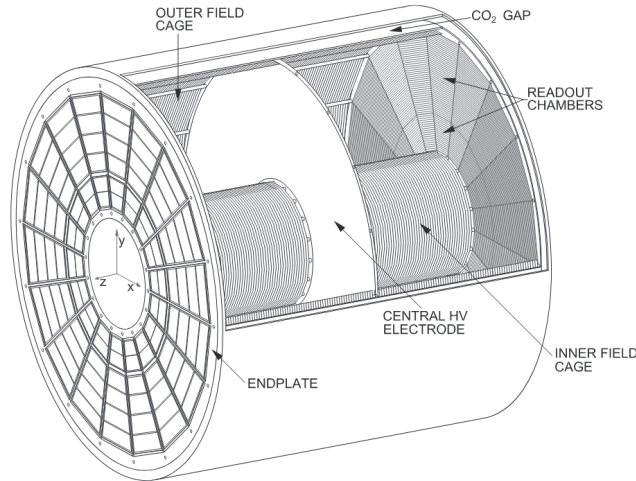
	ITS1	ITS2
Readout rate	up to 1 kHz	> 100 kHz (Pb–Pb) > 400 kHz (pp)
Material budget per layer	$1.1\%X_0$	$0.36\%X_0$ (Inner Barrel) $1.1\%X_0$ (Outer Barrel)
Pixel size	$50 \times 425 \mu\text{m}^2$	$27 \times 29 \mu\text{m}^2$
Pointing resolution ($p_T = 500 \text{ MeV}/c$)	$\sim 120 \mu\text{m}$ ($r\varphi$) $\sim 240 \mu\text{m}$ (z)	$\sim 50 \mu\text{m}$ ($r\varphi$) $\sim 50 \mu\text{m}$ (z)
Standalone tracking efficiency ($p_T = 200 \text{ MeV}/c$)	$\sim 60\%$	$\sim 90\%$
Radius of innermost layer	3.9 cm	2.5 cm

- To reconstruct particles that go through dead regions of the TPC located between sections.

Time Projection Chamber

The Time Projection Chamber (TPC) is the main detector in the central barrel for tracking and particle identification [206] and one of the upgraded detectors for Run 3 [207]. It consists of a hollow cylinder, as shown in Fig. 3.4, located between the ITS and the TRD detectors. Its active radial range goes from around 85 cm to 250 cm, with an overall length in the beam axis of 500 cm. It covers the following phase space:

- a pseudo-rapidity range of $|\eta| < 0.9$ for tracks that can be matched with ITS, TRD and TOF, and up to $|\eta| = 1.5$ for unmatched tracks;
- a full azimuthal coverage, except for dead zones between sectors; and
- the p_T region between $0.1 \text{ GeV}/c$ and around $100 \text{ GeV}/c$.

**Figure 3.4:** Layout of the TPC [194].

The TPC is divided in two by a central high-voltage electrode, resulting in two symmetric drift zones. It is filled with a gas mixture of Ne-CO₂-N₂ at atmospheric pres-

sure, which ionises when charged particles pass through. The electrons generated by this process are then affected by an electric field, which makes them drift towards the detector endplates. Finally, their arrival points are measured using Gas Electron Multiplier (GEM) detectors, which allow for a continuous readout up to 50 kHz in Pb–Pb collisions. This way, the trajectories of charged particles can be measured with a p_T ranging from 100 MeV/ c (in the case of pions) to 100 GeV/ c . Particles can be identified by measuring the energy loss, dE/dx , which is parametrised via the Bethe-Bloch formula:

$$f(\beta\gamma) = \frac{P_1}{\beta^{P_4}} [P_2 - \beta^{P_4} - \ln P_3 + (\beta\gamma)^{-P_5}], \quad (3.1)$$

where β is the velocity of the particle, γ is the Lorentz boost factor, and P_i are fit parameters.

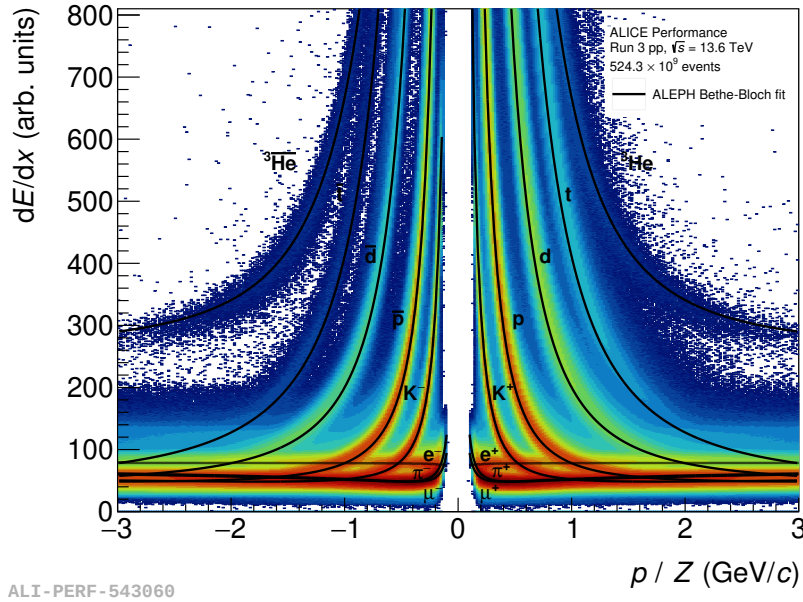


Figure 3.5: Particle identification using the TPC of the pp data collected in 2022 at $\sqrt{s} = 13.6$ TeV. Particles are identified by fitting to a Bethe-Bloch distribution. Figure taken from the ALICE Figure repository.

An example of the TPC performance in pp collisions at $\sqrt{s} = 13.6$ TeV can be seen in Fig. 3.5. It shows the energy loss distributions as a function of p/Z for several charged particles, which are identified with a fit to a Bethe-Bloch distribution.

Transition Radiation Detector

The Transition Radiation Detector (TRD) [208] is primarily used to identify electrons and trigger electrons and jets. It covers the pseudorapidity range $|\eta| < 0.84$, and is located between the TPC and TOF detectors. It is composed of 522 chambers, arranged in 6 layers located between 2.90 and 3.68 m from the beam axis. Each chamber contains a foam/fibre radiator, followed by a Xe-CO₂-filled multi-wire proportional chamber (MWPC). MWPCs are separated by a 3 cm drift region. This drift region is used to

reconstruct local track segments, to be matched with ITS-TPC tracks.

The TRD is based on the transition radiation process, which occurs when particles go through the boundaries of two media with different dielectric constants. For highly relativistic particles, the emitted radiation reaches the X-ray domain, and peaks forward in the particle direction. Hence, by absorbing the emitted X-ray photons using high-Z gases, it is possible to detect the traversing particle by measuring the energy deposition. This procedure can be used to separate electrons from hadrons, as the former produce X-ray photons and have a higher dE/dx due to the relativistic rise of the ionisation energy loss. The TRD has also been proven succesful in reconstructing high- p_T electrons and jets. Lastly, the TRD also plays a key role in correcting the distortions in the TPC tracks. A schematic view of the principle of detection can be seen in Fig. 3.6.

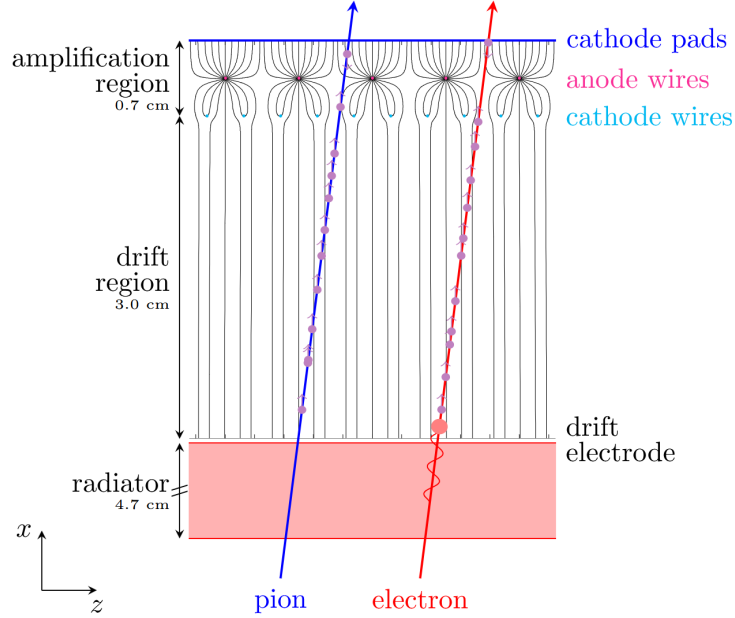


Figure 3.6: Layout of one of the modules of the TRD [208].

Time-of-Flight detector

The Time-of-Flight (TOF) detector [209] was designed to improve the particle identification (PID) capabilities in regions not fully covered by other detectors. It is the fourth detector with respect to the interaction point, after the ITS, TPC and TRD. It covers the pseudorapidity range $|\eta| < 0.9$, and the full azimuthal range. It measures the time it takes for a particle to travel between the interaction point and the detector itself with a precision better than 100 ps. This is particularly useful to measure and distinguish pions, kaons, and protons at intermediate p_T (between 0.5 and 2.5 GeV/c).

The TOF is composed of 1593 Multi Wire Resistive Plate Chambers (MRPC), separated into 18 supermodules. MRPCs are gaseous detectors composed of two stacks, each

containing five gas gaps. The gas ionises when a charged particle passes through the detector, creating a signal. The sum of the signals gives the final signal in every gap. A schematic view of the TOF detector can be seen in Fig. 3.7. During the Long Shutdown 2, the readout rate was improved from 40 kHz to > 200 kHz in Pb–Pb collisions [210].

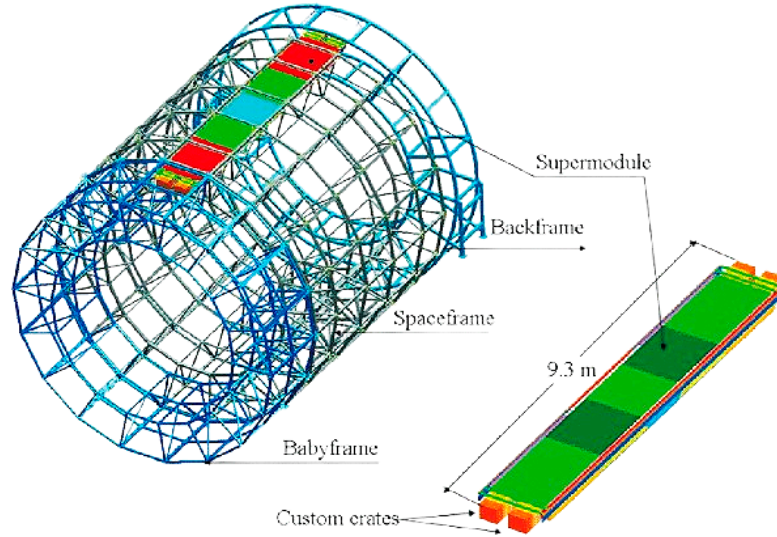


Figure 3.7: Layout of the TOF detector with a close-up in one of its supermodules [210].

3.2.2 Forward detectors

Zero Degree Calorimeter

The Zero Degree Calorimeter (ZDC) [211] is a detector whose main purpose is to exclude background events coming from beam-gas interactions and beam-halo particles. In addition, it contributes in the determination of the centrality of the collision through the measurement of the energy deposited by spectator nucleons. It consists of two hadronic calorimeters situated on both sides of the IP at a distance of 115.5 m, where the LHC beams circulate in two different pipes. Each set of detectors contains a neutron (ZN) and a proton (ZP) calorimeter. ZN calorimeters are located at a zero-degree angle with respect to the LHC axis. At the same time, ZPs are placed externally to the outgoing pipe since, at that point, spectator protons are deflected from the ion beams. The ZDC system is completed with two forward electromagnetic calorimeters (ZEM) located 7.35 m from the IP.

The ZDC is used to measure the forward energy carried by non-interacting nucleons, directly measuring the number of participants in heavy-ion collisions. Its operation is based on detecting Cherenkov light produced by showers of particles created in the calorimeters as charged particles pass through. This method is very fast, which makes this detector suitable for triggering purposes.

Fast Interaction Trigger

In order to cope with the larger luminosity in Run 3 with respect to previous runs, the triggering system needed to be upgraded. For this reason, three Run 2 detectors, the time zero detector (T0), the VZERO system (V0), and the Forward Multiplicity Detector (FMD), were replaced during the LS2 with the Fast Interaction Trigger (FIT) [212], composed of three subdetectors: FV0, FT0 and FDD. These subdetectors are located on both sides of the IP, as shown in Fig. 3.8, and they are used for triggering, online luminosity determination, vertexing, and as forward multiplicity counters. The FIT electronics are prepared to operate in both triggered and continuous modes and can provide a precise collision time that is used for the TOF PID.

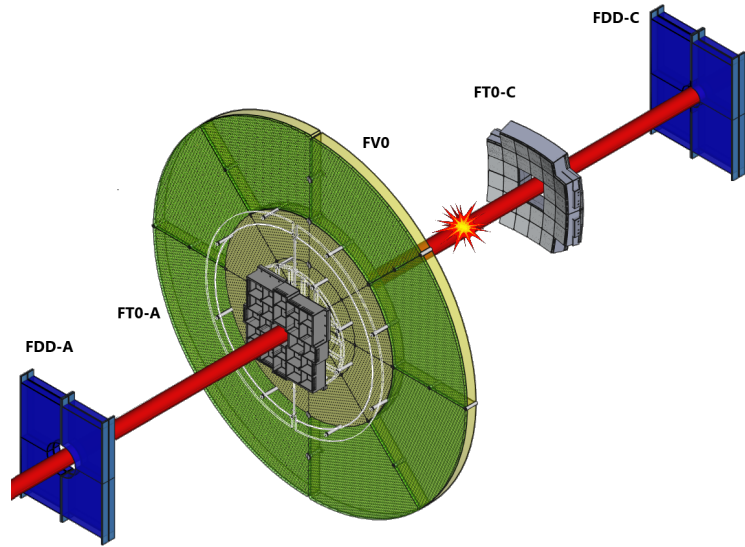


Figure 3.8: Layout of the FIT detector. Figure taken from the ALICE Figure repository.

- **FT0:** the FT0 is composed of two subdetectors, called FT0-A and FT0-C, located at opposite sides of the IP. They are quartz Cherenkov radiators coupled to microchannel plate based (MCP) photomultipliers, which provide a time resolution for high-multiplicity heavy-ion collisions of 7 ps. This level of precision is key for online vertex determination and identifying charged particles using the TOF.
- **FV0:** the FV0 is a scintillator disk with an outer diameter of 144 cm and an inner diameter of 8 cm. It is divided into 48 sectors, each providing an independent readout channel. This way, it can yield the required centrality and event plane resolution, as well as information on the LHC background conditions and luminosity.
- **FDD:** the Fit Forward Detector (FDD) contains two nearly identical arrays situated at opposite ends of the IP, FDD-A and FDD-C. Each array contains eight scintillator pads distributed around four sectors. Pads are connected to independent FDD channels on each IP side. The FDD detector contributes to measuring diffractive cross-sections and studying ultra-peripheral collisions.

3.2.3 Muon Spectrometer

The muon spectrometer [213, 214] was mainly designed to detect the spectrum of quarkonia states, as well as the muon decay products of light mesons, heavy-flavour hadrons and electroweak bosons. The spectrometer is able to reconstruct quarkonium candidates via their decay to $\mu^+\mu^-$ pairs in the rapidity range $-4 < y < -2.5$ with a resolution of $70 \text{ MeV}/c^2$ for charmonium and $100 \text{ MeV}/c^2$ for bottomonium. A schematic view of the muon spectrometer can be seen in Fig. 3.9, and its main components include:

- A silicon-based detector used to reconstruct secondary vertices,
- A set of absorbers,
- A dipole magnet that generates a field of 3 Tm,
- Ten muon tracking chambers of high granularity, and
- Four planes composed of Resistive Plate Chambers to identify muon candidates.

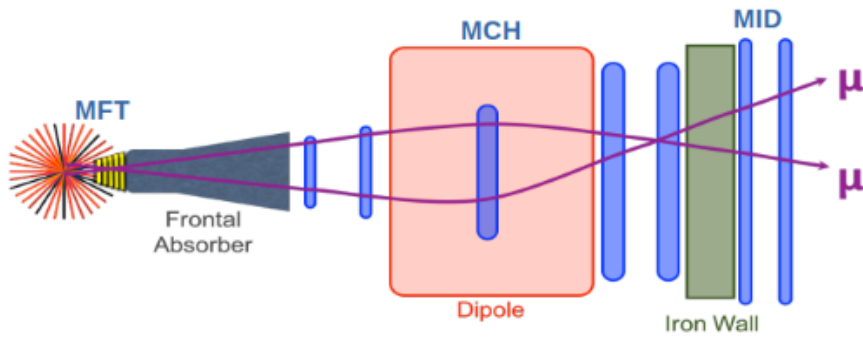


Figure 3.9: Layout of the muon spectrometer in Run 3.

In the following, each of the components of the MS will be briefly introduced.

Muon Forward Tracker

The Muon Forward Tracker (MFT) [215] is an all-silicon detector installed during the LS2. It was designed to improve the secondary vertex reconstruction of muons, which can give access to new physics measurements. It is located before the absorbers, inside the ITS outer barrel, and covers the pseudorapidity range $-3.6 < \eta < -2.5$.

In Run 1 and Run 2, prompt (i.e. produced in the collision) and non-prompt (i.e. coming from the decay of B-hadrons) J/ψ production could not be separated due to the multiple scatterings inside the front absorber. The MFT aims to disentangle the two, employing the same technology as the ITS detector based on MAPS monolithic pixel sensors. This way, the uncertainties in matching muon tracks to the primary vertex will also be reduced. The geometry of the MFT detector can be seen in Fig. 3.10. It is composed of two half-cones containing five detection half-disks each.

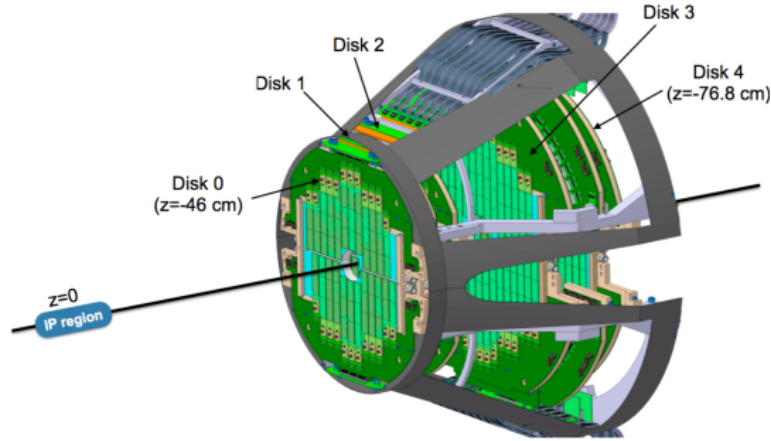


Figure 3.10: Layout of the MFT. Image taken from the ALICE figure repository.

Absorbers

Before reaching the forward detectors located in the muon arm, particles must go through the **front absorber** (Fig. 3.11), whose task is to reduce the background of muons coming from the decay of pions and kaons. At the same time, the deterioration of the invariant-mass resolution due to multi-scattering should be minimised. For this reason, the front absorber was built using low-Z materials, such as carbon and concrete, in the region closest to the interaction point. However, the rear end is made of high-Z materials since their purpose is to absorb low-energy photons and neutrons as well as the particles generated by the front absorber itself. It is located inside the L3 magnet, at a distance of 90 cm from the interaction point, and has a total length of 4.13 m. Finally, a **beam shield**, made out of tungsten and Pb, is situated around the beam pipe to diminish the low energy background generated by secondary interactions along the pipe.

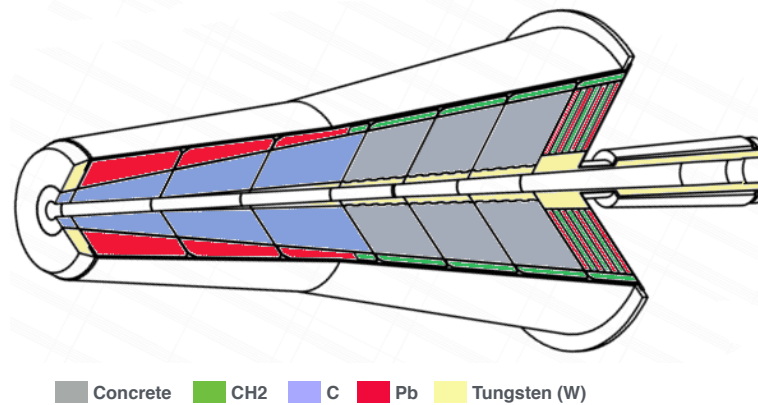


Figure 3.11: Internal structure of the front absorber and the beam pipe shield. Figure extracted from [216].

A 1.2 m **iron wall** can be found between the muon tracking and identification cham-

bers to avoid contamination from particles other than muons in the muon chambers. At last, a **second iron wall** acts as a rear absorber and protects the identification chambers from beam-gas interactions.

Dipole magnet

The MS counts with a **dipole magnet** [217] made with resistive coils used to measure the charge and momenta of muons. This measurement can be done by deflecting particles using a magnetic field and registering their curvature. The dipole magnet provides a magnetic field toward the LHC radius of 3 Tm. It is located 7 m from the IP and is 5 m long, 6.6 m wide and 8.6 m tall.

Muon Chambers

The MCH [214] is composed of five stations of two planes each, consisting of a total of 156 Multi-Wire Proportional Chambers (MWPC) of 4.2 and 5 mm. The first two stations are located before the dipole magnet, the third station is inside the magnet, and the last two are situated after. Station sizes vary greatly, ranging from a few square meters in station 1 to around 30 m² in station 5. The muon spectrometer angular acceptance limits the size of the first two stations, while the last three stations are relatively larger to detect particles deviated by the dipole magnet. Hence, all stations were constructed employing Cathode Strip Chambers (CSC). Depending on the station, different geometries were adopted by varying parameters as the pad sizes and the distances between wires. A schematic representation of two chambers can be seen in Fig. 3.12.

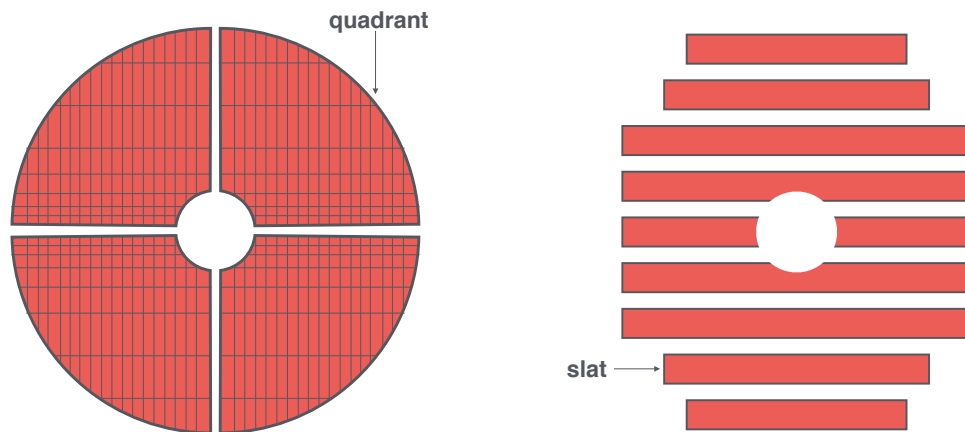


Figure 3.12: On the left, the quadrant structure of stations 1 and 2 of the MCH. On the right, the slat structure of stations 3, 4, and 5 [216].

MWPCs are characterised by the presence of a central anode plane containing a set of

parallel, equally spaced wires situated between two cathode planes. As seen in Fig. 3.13, this configuration creates an electric field with a maximum in the wire surface. The space between the anode and cathode planes is filled with a gas composed of 80% Argon and 20% CO_2 . The gas ionises when a particle passes through the chamber, liberating electrons. The electric field moves these electrons towards the closest anodes, where they generate avalanches of secondary electrons, causing a redistribution of the charges in the nearest cathode pads. The primary ionisation follows the path of the track. This provides information on the x and y hit positions, while the z position can be inferred by the station number that was hit.

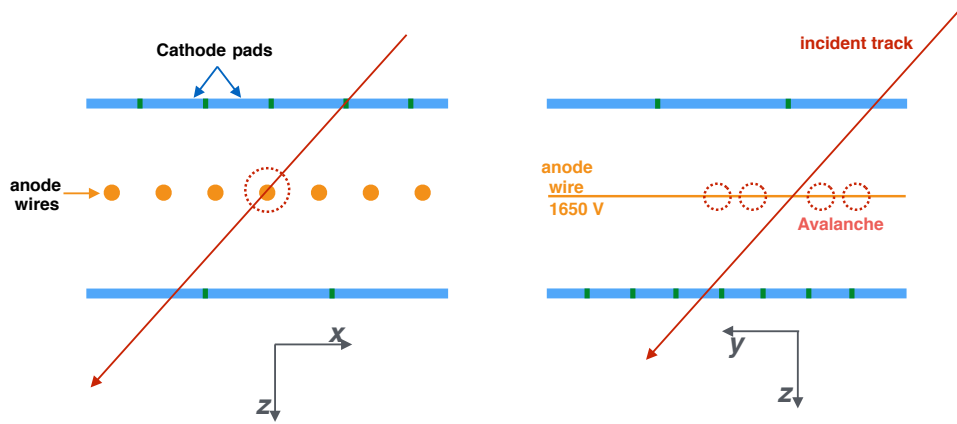


Figure 3.13: Diagram of the principle behind MWPCs [216].

Cathode pads are finely granulated in order to keep the occupancy at a maximum of around 5% in the most central Pb–Pb collisions. The size of these pads depends on their location since higher multiplicities are expected in the regions closest to the beam. For this reason, pads are larger at larger radii. The readout electronics are implemented on the surface of stations 1 and 2 and on the side of stations 3, 4, and 5. In all cases, electronics were upgraded for Run 3 [218] to cope with the expected higher collision rate. A new front-end electronics (FEE) and readout chain were implemented, which can be used in both triggered and continuous modes. The signal of 40 FEEs is gathered by the GigaBite Transceiver (GBT) readout card, which directs it to the Common Readout Unit (CRU), connecting it to the O^2 computing system (see section 3.3).

Muon Identifier

The last detector that can be found in the MS, located after the first iron wall, is a muon trigger system [219]. This system was upgraded Run 3 to cope with the continuous

readout scheme, resulting in the Muon IDentifier detector or MID. Its main purpose is to select events suitable for quarkonia analyses, i.e. pairs of unlike-sign muons. This is needed due to the relatively low cross-section of these kinds of events. To do so, a fast signal and a large covered area were needed. Hence, Resistive Plate Chambers (RPCs) were the detectors chosen to build the MID. Finally, to select the desired muon candidates, a trigger decision is made based on the p_T of the measured particle. Low- p_T muons are rejected since the decays of pions and kaons predominantly produce them.

The MID consists of two stations, each containing two planes of 18 RPCs. RPCs comprise two electrodes separated by a 2 mm gap filled with gas. One of the electrodes is connected to a high voltage, and the other is grounded, creating an electric field. When a charged particle goes through, it ionises the gas and generates an avalanche of secondary electrons, which move towards the electrode plates and are registered as signals. The readout electronics were upgraded during the LS2 in a similar fashion to the MCH.

3.3 Computing and data taking in Run 3

Detecting and processing Pb–Pb events at 50 kHz in a continuous readout is a great computational challenge. Over the course of Run 3, roughly 100 times more heavy-ion data is expected to be collected than in Run 1. Moreover, an estimated interaction rate of 500 kHz is expected for pp collisions. Considering that pp data is taken during roughly seven months (around 210 days) per year, and assuming a 35% LHC efficiency, we can expect to have an integrated luminosity of around 40 pb^{-1} per year. This luminosity corresponds to 3.2×10^{12} collisions per year, constituting about 79 PB of total raw output size [220]. Hence, an upgrade was needed in the computing system to minimise its cost and requirements for processing and data storage. This has led to the upgrade of the ALICE Online and Offline Computing to a new common system called O^2 [221].

The new system is designed to reduce the data volume from the detectors as early as possible during the data flow. This reduction is performed by processing data on the fly rather than rejecting full events using high-level triggers. Since data can be taken continuously, it is not delimited by events but by Time Frames (TF), which constitute the basic processing unit in ALICE in Run 3. Each TF corresponds to 32 LHC orbits (around 2.85 ms) of continuous data.

The functional flow of the O^2 system can be seen in Fig. 3.14, from the raw data gathered from the detectors to the final archival of the reconstructed events. Data is collected in either triggered or continuous mode and replaced by the results of a first local processing performed with a high level of parallelism. Afterwards, a step of data aggregation is done to assemble the information of all detectors and a global calibration

and reconstruction is performed synchronously with the data taking. Results are then temporarily stored and recalibrated asynchronously before being stored permanently.

3.3.1 The First Level Processors (FLP) farm

The FLP farm [222] performs the traditional Data Acquisition (DAQ) from the detectors and reduces the data rate from around 3.5 TB/s to 600 GB/s. Most of this reduction comes from performing zero-suppression of the TPC data since, out of all detectors, the one that generates the highest amount of data is by far the TPC, constituting around 90 %.

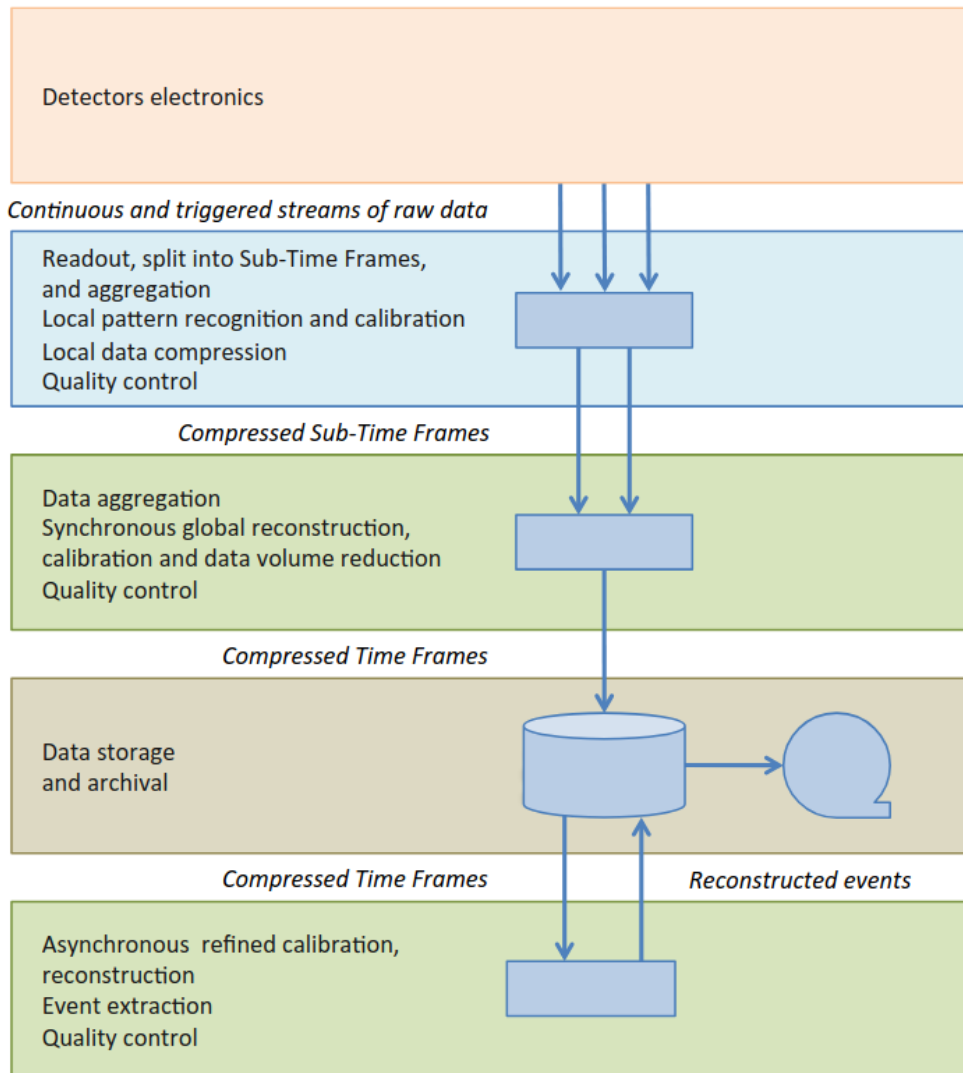


Figure 3.14: Functional flow of O^2 , from the raw data collected in the detectors to its final storage. Image taken from [223].

It is located at CR1, in the experimental site at P2, and it contains a total of 42 racks, out of which 24 are currently in use and contain 260 servers. To readout the input from the detectors, roughly 9000 fibres are used to connect the front-end electronics of the

detectors to the computing room. Two types of readout cards can be used: CRUs or the legacy Common Readout Receiver Card (C-RORC). Each FLP only sees data from up to 3 CRUs or 4 C-RORCs, i.e. parts of individual detectors. They run a multi-process topology that reads the detector data, performs an initial processing and/or quality control and sends the data to the EPN farm.

3.3.2 The Event Processing Nodes (EPN) farm

The EPN farm was designed to sustain the highest maximum expected data rate possible during the synchronous processing step, during which Compressed Time Frames (CTF) are created and calibrated for the first time. Hence, EPNs are optimised to provide the fastest TPC track reconstruction possible in order to compress it further. This compression is done in several steps, starting with clustering algorithms to reconstruct hits. Then, hits that will not be used for physics are discarded, and the remaining are further compressed in several entropy-reducing steps. This allows a data rate reduction from 600 GB/s to around 100 GB/s. The final samples are stored in the disk buffers as CTFs.

EPNs also participate in asynchronous processing alongside the GRID, a set of computing centres distributed in more than 40 countries. All the information stored in the CTFs from all detectors is used during this step to refine calibrations before storing the final datasets as Analysis Object Data (AOD) in ROOT [224] Tree format.

3.3.3 Analysis framework: O^2 Physics

The analysis software employed in Run 3 is called O^2 Physics [225, 226]. It is a collection of flat tables interconnected via indices. This choice was made because, in Run 3, vertex-to-track association, for instance, is no longer unambiguously defined. Thus, collisions and tracks must be defined in separate arrow tables to be connected via indices later, greatly increasing processing speed. Tables are stored in ROOT format as TTrees located in TFiles, and each row of the table contains a physical entity, such as collisions, tracks, etc. This columnar memory layout is optimised for bulk operations, contributing to the boost in computing speed needed to deal with the larger statistics of Run 3. Each table can be linked to others which have a compatible structure, avoiding possible data repetitions. In the same way, users can define their own derived tables, which can be connected to all other existing tables.

A simplified diagram of how tables are interconnected is shown in Fig. 3.15. In this image, grey tables contain the information on disk coming from the data, green tables can be calculated from the grey ones, and orange tables contain MC-simulated information. For instance, the most general tables in data are the ones which contain the information of the bunch crossings. These bunch crossings are referred to by collisions, which contain properties such as the time stamps, the event selection criteria, the multiplicities, or the centralities. Other objects can also refer to the bunch crossings, such

MUON TRACKING EFFICIENCY STUDIES OF RUN 3 DATA

Before being able to perform any kind of analysis, the data we collected must be cross-checked to ensure that it is suitable for use. However, even after removing the runs that are not good for physics analyses, there may still be detector effects and imperfections that need to be simulated in the Monte Carlos. For this reason, we will focus on the study of the **muon tracking efficiency of the muon chambers (MCH)** and calculate a systematic uncertainty that encompasses for the differences between the data and the simulations available. A detailed description of the MCH can be found in Section 3.2.3.

In this Chapter, we will start by describing how muons are reconstructed in the muon spectrometer and the procedure to calculate the muon tracking efficiency. Afterwards, we will present the results of this calculation corresponding to the Pb–Pb data collected in November of 2023 at $\sqrt{s_{\text{NN}}} = 5.36$ TeV. In particular, we obtained the efficiencies of a set of runs labelled as the *golden runs* of this Pb–Pb period. These are:

544095, 544098, 544116, 544121, 544122, 544123, 544124.

4.1 Acceptance and efficiency in the muon spectrometer

In the muon spectrometer, the acceptance is defined as the ratio of muons produced within the angular coverage of the detector to the total number of produced muons. This definition can be used to extract the number of single muons which can be reconstructed. In the case of di-muon pairs, such as the ones used to reconstruct quarkonia or weak bosons, one of the muons may get detected, while the other one falls outside of the production window. How likely it is to happen depends on parameters such as the particle rapidity, transverse momentum or polarisation. For instance, the muon pair produced by a high- p_{T} particle will have a smaller relative angle thanks to the Lorentz boost. Thus, it is more likely for both muons to be inside the detection window. For instance, the J/ψ acceptance is defined as the ratio of the number of J/ψ

produced within the angular coverage of the detector to the total number of di-muon pairs detected in that angular coverage.

An example of the J/ψ acceptance in the muon spectrometer calculated with PYTHIA 6.2 is presented in Fig. 4.1 [227]. As can be observed, at fixed rapidity, the acceptance reaches a minimum value at $p_T \sim 3.1$ GeV/c, and increases with p_T from there on. At fixed momentum, the acceptance decreases at the borders of the rapidity region, because it is harder to reconstruct both muons in which the J/ψ decays in these areas.

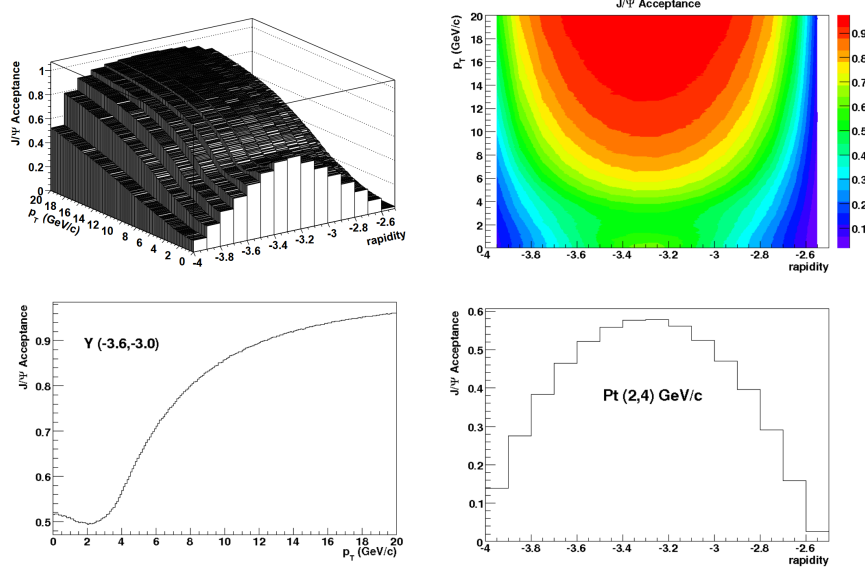


Figure 4.1: J/ψ acceptance. Figures from [227].

On the other hand, the efficiency of the muon spectrometer is computed by considering the tracking and triggering efficiencies. Our purpose in this Chapter is to obtain an estimation of the tracking efficiency systematic uncertainty by comparing the results in data and Monte Carlo using the same methodology. This systematic uncertainty is present in all measurements involving muons due to the impossibility to reconstruct all tracks that pass through the detector. For example, a track that passes through two dead zones, one in each chamber of a station, will not be reconstructed in the data, but we will be able to know it existed in MC thanks to the MC-truth.

4.2 The muon reconstruction algorithm

To study the muon tracking efficiency, we must first describe the tracking algorithm employed by the muon spectrometer. Similarly to the central barrel tracks, whose reconstruction is explained in Sec. 5.3.1, this procedure is based on the **Kalman filter algorithm** [228, 229]. The reconstruction steps can be found in [230, 231] and are the following:

1. The tracking algorithm begins in station 5. Track segments are built by connecting the cluster information of the two chambers that compose the station. These

segments are then extrapolated to the primary vertex and reconstructed with the ITS. At this stage, the local positions, directions and impact parameters are computed, alongside their respective uncertainties. The extrapolation of the segments is done taking into account the effect of the magnetic field, allowing us to estimate their momentum thanks to the Lorentz law.

2. A selection is made only to filter the muons within the acceptance. In addition, muon candidates with $p_T < 0.5 \text{ GeV}/c$ are rejected.
3. The steps 1 and 2 are repeated in station 4.
4. The track candidates reconstructed in station 5 are extrapolated to station 4, and vice versa. Candidates are required to have at least one compatible cluster in the other station. Track candidates are duplicated if there is more than one compatible cluster per station.
5. Tracks sharing the same clusters are removed. Further selections are made to remove candidates with an impact parameter and/or momentum outside of the given limits.
6. The track candidates are propagated to stations 3, 2, and 1, in that order, taking into account the effect of the magnetic field. In all cases, at least one compatible cluster is required per station. The track parameters are recalculated at each step, and selections are made by cutting on the impact parameter and transverse momentum. Tracks are duplicated if there are multiple compatible clusters in the same chamber.
7. The candidates sharing clusters in the first three stations are removed. Further cuts are made to reject the tracks with the lowest number of clusters. If there are several tracks with the same number of clusters, the ones with the lowest χ^2 are kept. The tracks that reach this step and pass these selections are called **MCH standalone tracks**.
8. The surviving tracks are matched with the MID track segments. These segments are required to belong to the same period as the MCH tracks within four bunch crossings. MCH-MID tracks are extrapolated to the primary vertex reconstructed with the ITS. As in the previous steps, only tracks within the detector acceptance that have a transverse momentum above the minimum required are kept. The tracks that reach this stage are called **muon standalone tracks**.
9. The muon standalone tracks are matched with the tracks reconstructed using the MFT, known as **MFT standalone tracks**. Only the tracks that pass the selection cuts in the transverse momentum and impact parameter are kept. The tracks also need to be within the MFT acceptance. The tracks that could not be matched with the MFT receive the name of **MCH-MID tracks**. In all cases, the track parameters are updated.
10. The remaining tracks, called **global muon tracks**, are extrapolated to the primary vertex reconstructed with the ITS.
11. For all track types, the track parameters are recalculated and corrected for the energy loss and the multiple scattering in the absorber.

We introduced the names of several types of muon tracks depending on the stage they reached of the muon reconstruction algorithm. In summary, the different tracks we can have in the muon spectrometer are (see their representation in Fig. 4.2):

- **Global muon track:** These tracks are created by matching clusters in the MFT, MCH and MID detectors.
- **Global forward track:** The tracks are created by matching clusters in the MFT and MCH detectors.
- **Muon standalone track:** These tracks are created by matching clusters in the MCH and MID detectors.
- **MCH standalone track:** These tracks are reconstructed using exclusively the MCH clusters.
- **MFT standalone track:** These tracks are reconstructed using exclusively the MFT clusters.

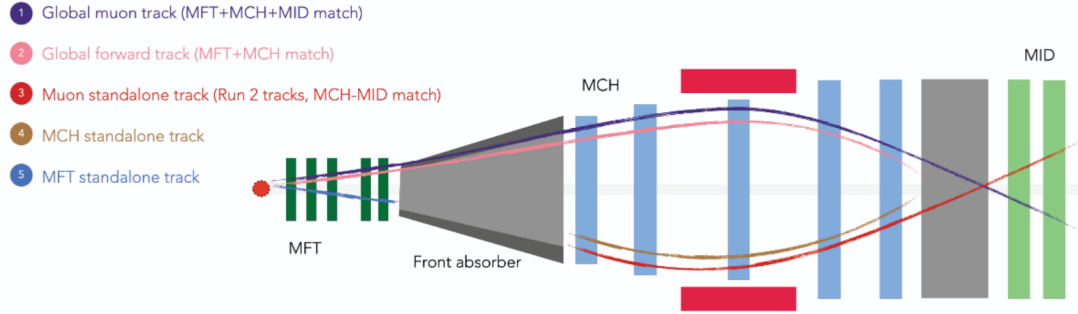


Figure 4.2: Muon track types. Figure extracted from [232].

4.3 Method to estimate the tracking efficiency

The strategy to extract the tracking efficiency of the MCH is based on the existing chamber redundancy per station. Tracks can leave a hit in both chambers of a station (N_{ij}), in just one chamber (N_{0j} , N_{i0}), or in none (N_{00}), in which case the detector is unable to reconstruct them. Figure 4.3 contains a representation of the MCH and the different types of tracks.

Therefore, the total number of tracks passing through a station composed of chambers i and j can be defined as:

$$N_{\text{tot}} = N_{ij} + N_{0j} + N_{i0} + N_{00}. \quad (4.1)$$

Assuming that the efficiencies of chambers i and j are independent, N_{ij} can be expressed as:

$$N_{ij} = \epsilon_i \epsilon_j N_{\text{tot}}. \quad (4.2)$$

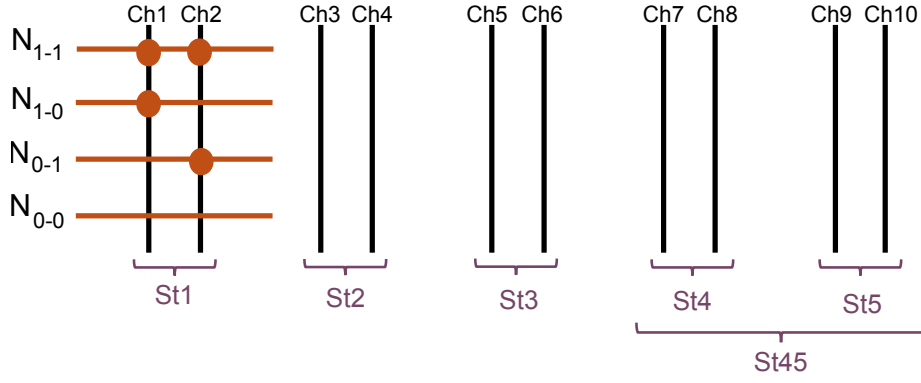


Figure 4.3: Schematic representation of the MCH showing the different responses of the detector to the passage of a track through Station 1. Adapted from [230].

Similarly, for N_{0j} and N_{i0} :

$$N_{i0} = \epsilon_i(1 - \epsilon_j)N_{\text{tot}}, \quad (4.3)$$

$$N_{0j} = (1 - \epsilon_i)\epsilon_j N_{\text{tot}}. \quad (4.4)$$

Manipulating these equations, one can extract the definitions of the chamber efficiencies as a function of the number of hits per chamber:

$$\epsilon_i = \frac{N_{ij}}{N_{ij} + N_{0j}} \quad (4.5)$$

$$\epsilon_j = \frac{N_{ij}}{N_{ij} + N_{i0}}. \quad (4.6)$$

Furthermore, the efficiency per station can be extracted by combining the efficiencies per chamber. In the case of the first three stations, this can be done as:

$$\epsilon_{\text{St}(1)(2)(3)} = 1 - (1 - \epsilon_i)(1 - \epsilon_j), \quad (4.7)$$

where i and j represent the two chambers of the station. On the other hand, the last two stations need to be considered together, due to the restrictions of the tracking algorithm. The efficiency requires that at least three out of the four chambers detect the muon:

$$\epsilon_{\text{St}(45)} = \prod_{i=7}^{i=10} \epsilon_i + \sum_{i=7}^{i=10} \left((1 - \epsilon_i) \prod_{j=7, j \neq i}^{j=10} \epsilon_j \right). \quad (4.8)$$

The final efficiency is computed as the product of the efficiencies of all stations:

$$\epsilon_{\text{trk}} = \epsilon_{\text{St}(1)} \epsilon_{\text{St}(2)} \epsilon_{\text{St}(3)} \epsilon_{\text{St}(45)}. \quad (4.9)$$

4.4 Tracking efficiency in Pb–Pb collisions

Using the methods explained in the previous sections, we calculated the muon tracking efficiency of the *golden runs* of Pb–Pb data collected in November of 2023. In this

section, we report our findings on the data and an initial anchored Monte Carlo simulation and compare the two.

As a first cross-check, we compared the cluster distributions that we observed with the 2D representation of the fired pads from the Quality Control (QC). Those plots were generated online at the time the data was collected to ensure data quality. The QC plots corresponding to the cluster distributions in run 544122 can be observed in Fig. 4.4. More specifically, these plots contain the cluster distributions recorded on October 6 2023, between 4:11 and 5:04. The most notable features of the cluster distributions are the holes present on the bottom-left corner of chamber 2 and the top-left region of chamber 6. The hole in chamber 2 corresponds to a known issue that prevented the sector from receiving any voltage, caused by a broken or very loose wire. This issue was present in all Pb–Pb runs considered in this chapter, and is planned to be fixed during the long shutdown 3.

On the other hand, the hole in chamber 6 is only present in runs 544098, 544116, 544121, 544122, 544123 and 544124. Its origin can be traced down to a low voltage channel that was not powered in the SOLAR crate number 1. This crate powers the boards that ensure the readout of the detection elements without clusters, so we have an exact matching.

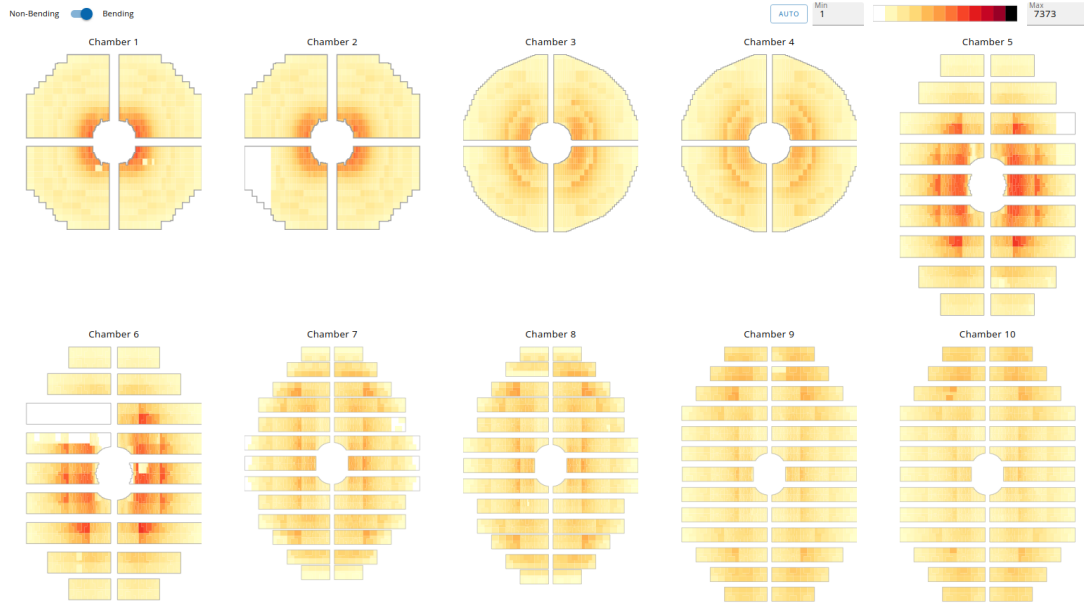


Figure 4.4: Online cluster distributions of the MCH chambers when run 544122 was collected.

If we compare the online cluster distributions in Fig. 4.4 with the ones we obtain offline using the code developed in O^2 Physics, we would expect to find holes in the same regions of chambers 2 and 6. As can be observed in Fig. 4.5, they display minima in the same regions as the ones seen in the QC plots. From the cluster map, in Fig. 4.4 no other sizeable areas with missing signal could be seen in the tracking chamber.

In all cases, it is worth noting the differences in binning between the online and of-

fine reconstructions, which can make some finer structures appear different. This is observed, for instance, in the cluster distributions of chambers 8 and 9 shown in the bottom plots of Fig. 4.5. These chambers do not present any missing signal in the on-line reconstruction, a feature that is also reproduced offline. Overall, we may consider this test successful, as the O^2 Physics code manages to reproduce, at least in broad terms, the real-time data taking conditions of the detector.

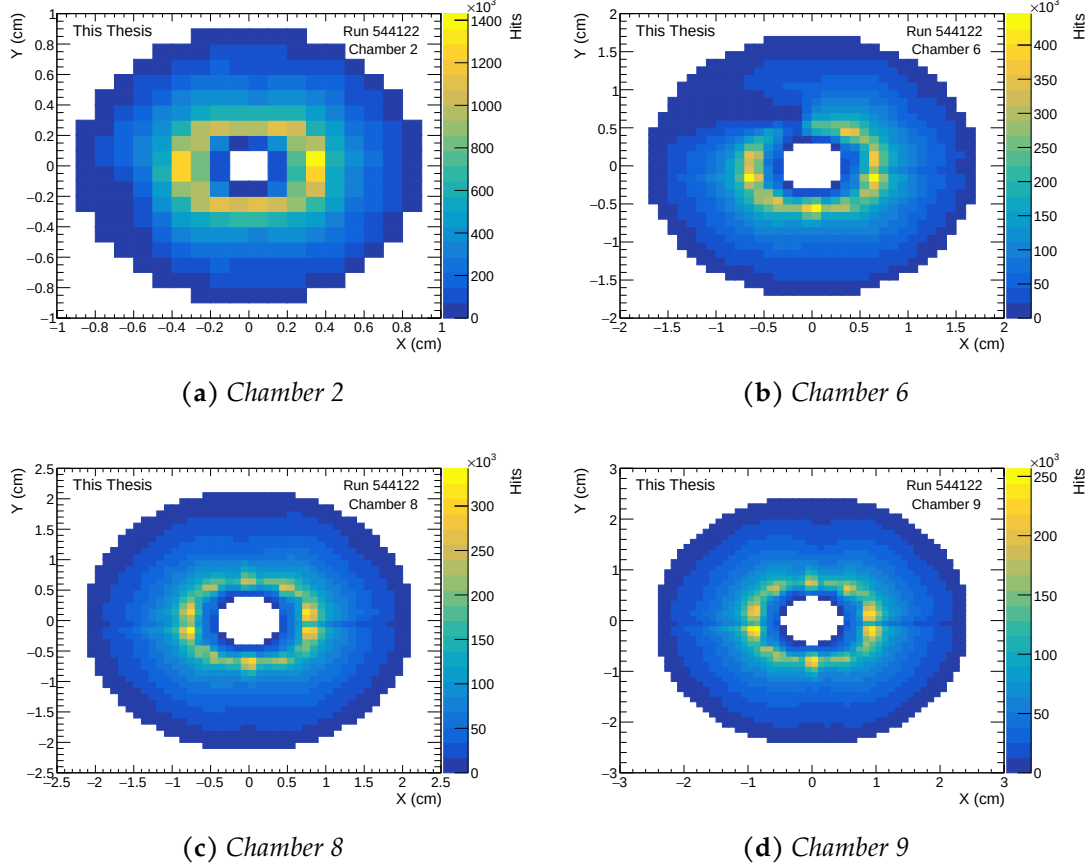


Figure 4.5: Offline results obtained with O^2 Physics for the clusters in chambers 2, 6, 8 and 9.

4.4.1 Muon tracking efficiency studies in data

After proving that the cluster distributions we reconstruct offline correspond to the ones visible online at the time of the run, we estimated the muon tracking efficiency of the MCH in each chamber using Eq. 4.5. The resulting efficiencies are presented as a function of η and φ in Figs. 4.6 and 4.7.

The hole in the cluster distribution of chamber 2 translates to a reduced efficiency in the bottom-left corner of Fig. 4.6b, corresponding to $\eta \sim (2.5, 2.9)$ and $\varphi \sim (-\pi, 2.4)$ rad. The efficiency in this area reaches a minimum of 20%. Furthermore, the hole in chamber 6 is also visible in the upper half of the $\eta - \varphi$ distributions.

Other chambers, such as chambers 5, 7, or 8, have small holes in their $\eta - \varphi$ distributions caused by faulty front-end electronic (FEE) boards. For instance, it is possible to see

that there is a small sector in the upper-right corner of chamber 5 where no clusters were detected (Fig. 4.4), which is present in the top-right corner of Fig. 4.7c. Lastly, two straight lines are visible in chamber 10, located at $\varphi \sim \pm 1.6$. These lines could be due to an improper closing of the chamber half-planes.

4.4.2 Muon tracking efficiency studies in the initial Monte Carlo

Our chosen method to calculate a systematic uncertainty related to the muon tracking efficiency requires comparing the efficiencies found in data with those in MC¹. At the same time, all Monte Carlo simulations are required to recreate the passage of particles through the detectors, but only those **anchored** to a specific production include the status of the detectors at the time the data were taken (more on simulated data samples in 5.1). MC simulations which do not include these conditions are called **ideal**. In our case, we begun by studying the efficiencies using an anchored MC which contained information on the alignment and noise channels.

The $\eta - \varphi$ distributions of the muon tracking efficiency in all chambers are presented in Figs. 4.8 and 4.9. As expected, most chambers display an efficiency close to 100%. The hole that we saw in the data in chamber 2 is not present in the MC, while the distribution in chamber 6 contains a hole in the $\eta \sim (-2.5, 3.1), \varphi \sim (-2.5, 1.6)$ rad region. In addition, a hole is present on the left side of the distribution of chamber 5. This hole is similarly located to the one present in the data in the same chamber, but there is no sign of the efficiency drop present in the upper-right corner of the data.

The discrepancies between the data and the initial MC simulation in Chamber 2 may be due to missing high-voltage information in the MC sample. In Chambers 5 and 6, the absence of certain holes observed in the data could also result from incomplete input information, such as the low-voltage information. By contrast, the holes that

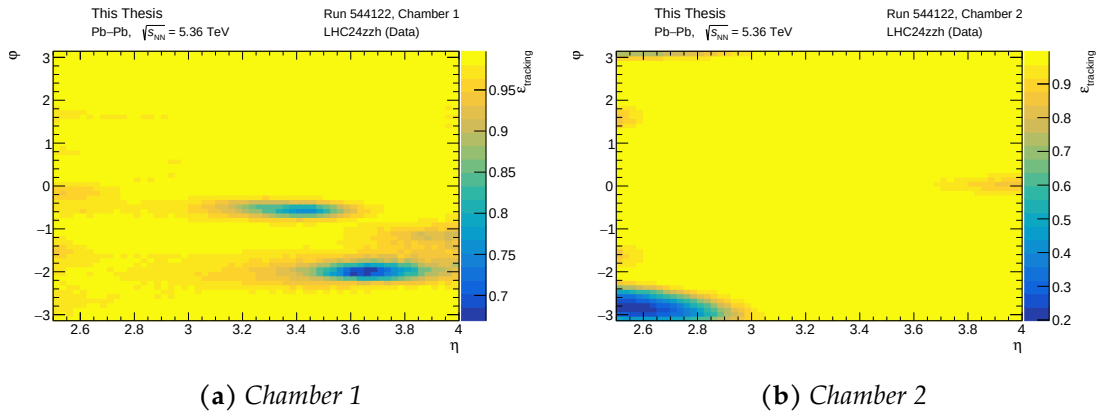


Figure 4.6: MCH tracking efficiency of the data corresponding to run 544122 as a function of η and φ (part 1).

¹ This is true for the cases where the efficiency is computed from MC simulations. In some cases, it is possible to estimate the efficiencies using a data-driven approach, such as the *tag-and-probe* method, which do not suffer from this systematic uncertainty.

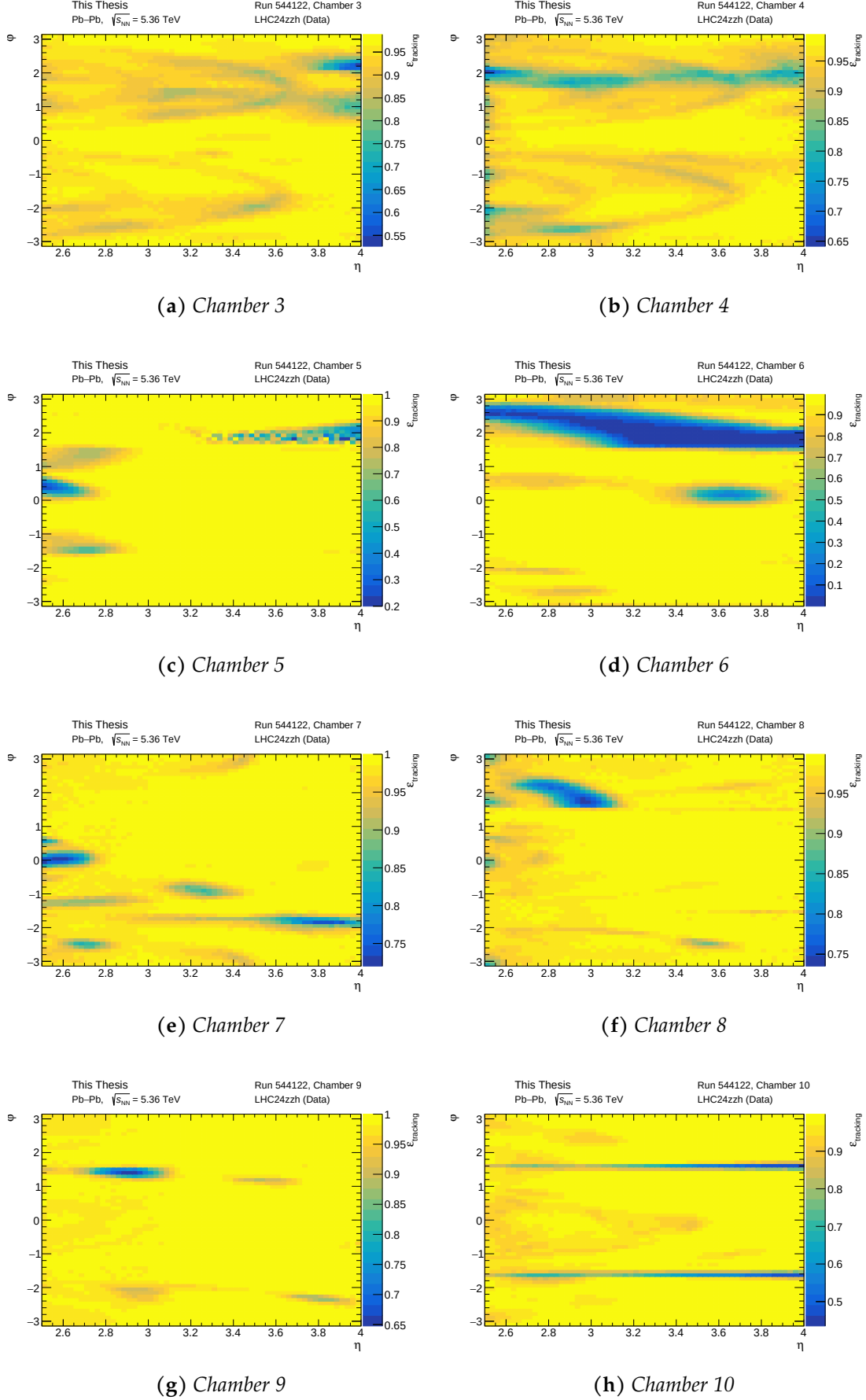


Figure 4.7: MCH tracking efficiency of the data corresponding to run 544122 as a function of η and φ (part 2).

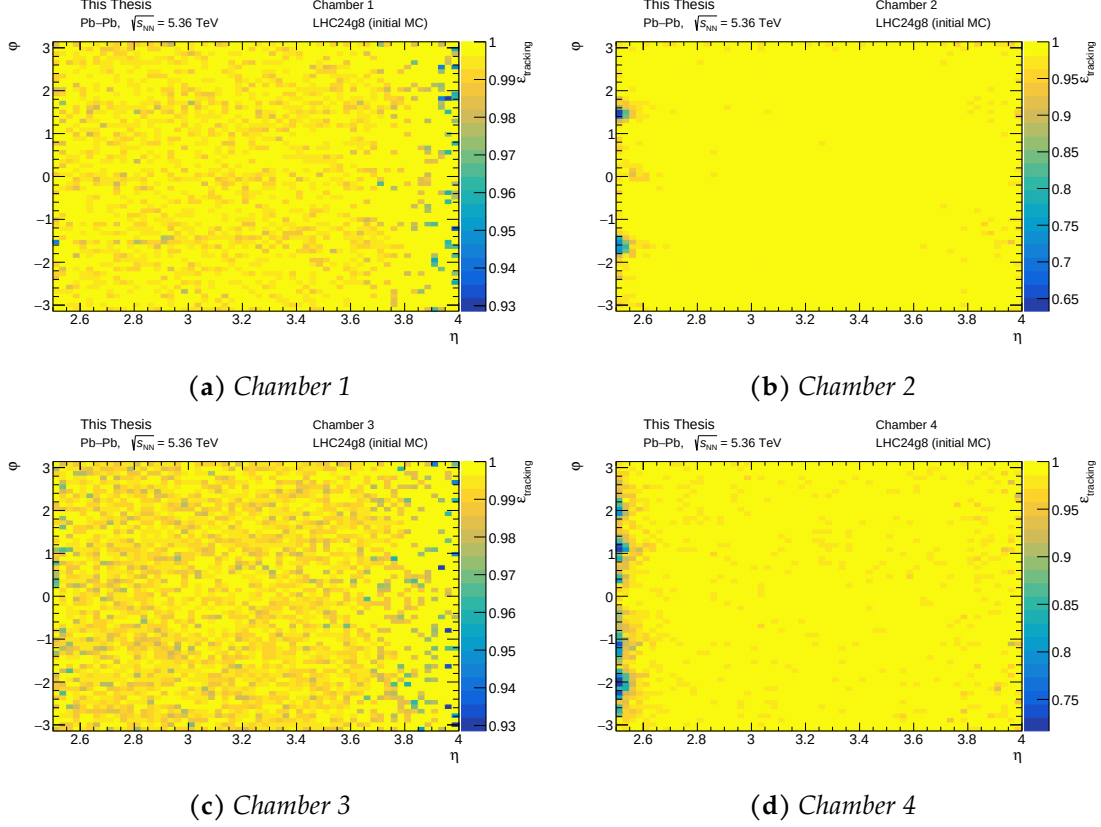


Figure 4.8: MCH tracking efficiency calculated using the initial MC corresponding to run 544122 as a function of η and ϕ (part 1).

do appear in the MC but are misaligned with those in the data may be caused by the simulation using an incorrect cluster map. Lastly, the straight lines seen in Chamber 10 data are reproduced in this simulation.

4.4.3 Results using an anchored Monte Carlo

In order to study the muon tracking efficiency in a more realistic scenario, we used an updated **anchored** Monte Carlo simulation. This simulation includes the high and low voltage information and dead zones per run on top of what is present on the previous MC at the time of the run. Furthermore, the bug in the alignment was fixed. The resulting $\eta - \phi$ distributions of the muon tracking efficiency in all chambers are presented in Fig. 4.10 for chambers 2 and 6. These chambers were selected because they contained the most significant differences in the initial MC with respect to the data. As we can see, the updated anchored MC is capable of reproducing the holes in the $\eta - \phi$ distributions of both chambers. In particular, the hole in chamber 2 is now present in the bottom-left corner, and the one in chamber 6 occupies most of the upper half of the distribution. Other features, such as the straight lines found in chamber 10, are also present in this MC. The comparison between the data and this MC for chambers 2 and 6 is located in the Appendix A. Hence, we expect the anchored MC to be able to reproduce the muon tracking efficiencies found in the data.

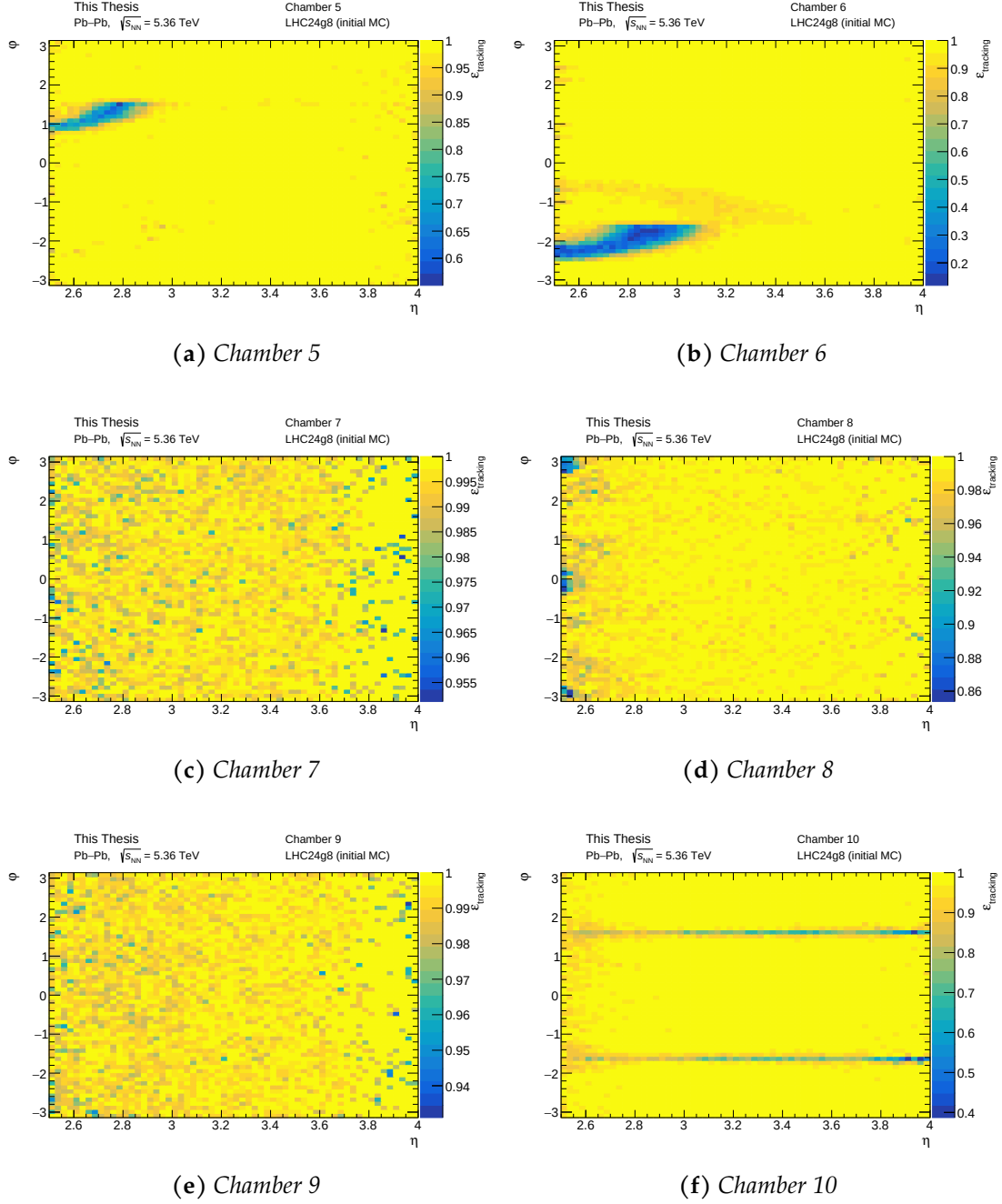


Figure 4.9: MCH tracking efficiency calculated using the initial MC corresponding to run 544122 as a function of η and ϕ (part 2).

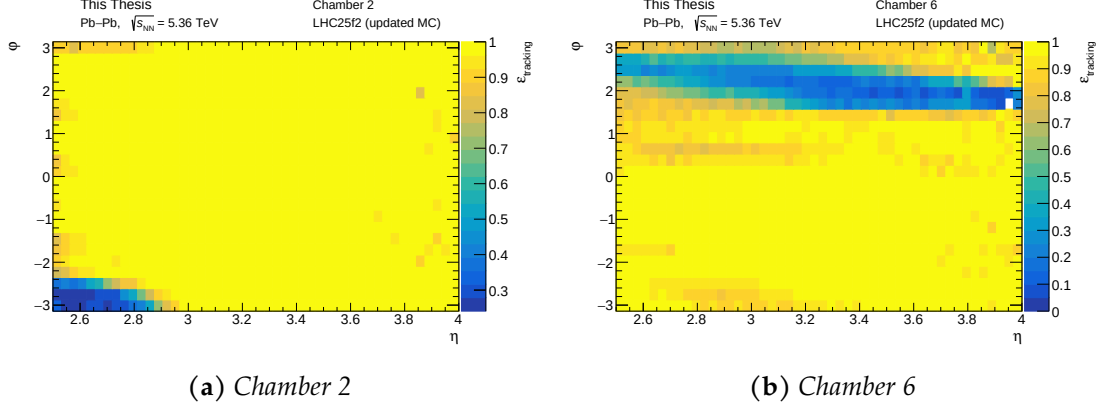


Figure 4.10: MCH tracking efficiency calculated using an anchored MC corresponding to run 544122 as a function of η and φ for chambers 2 and 6.

4.4.4 Comparison between data and MC

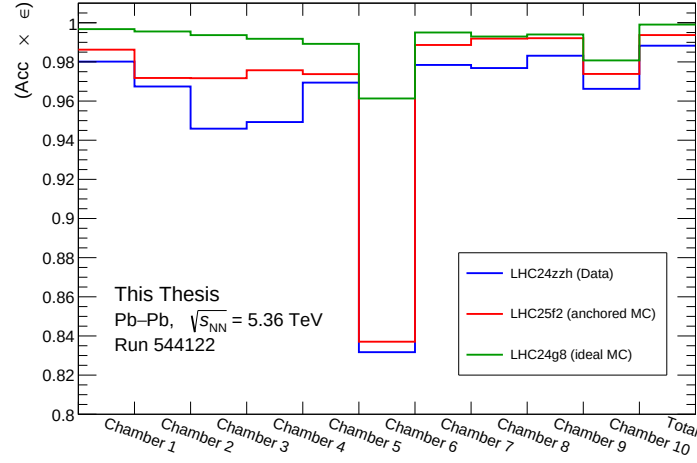


Figure 4.11: Muon tracking efficiency per chamber for run 544122. The initial MC is represented in green, the updated anchored MC in red, and the data in blue.

After studying the $\eta - \varphi$ distributions in data and MC, we present the results of the muon tracking efficiency per chamber in Fig. 4.11 for run 544122. As we anticipated, the efficiencies of the initial MC are very close to unity for all chambers except chamber 6, due to the limited per-run information available. Nevertheless, the dip in the efficiency in chamber 6 is not as significant as the one in the data, where it drops to 0.84. On the other hand, the results using the updated MC are much closer to the data, as expected. The situation is similar when we study the efficiency per station, calculated using the expressions in Eqs. 4.7, 4.8 and 4.9. In this case, which is presented in Fig. 4.12, we can see that the initial MC does not reproduce the drops in the efficiencies per station, making the total efficiency significantly closer to unity than what we find in data. Overall, we can conclude that the initial MC does not reproduce the trends present in the data, as we expected, and that the updated MC is capable of following the trends marked by the data.

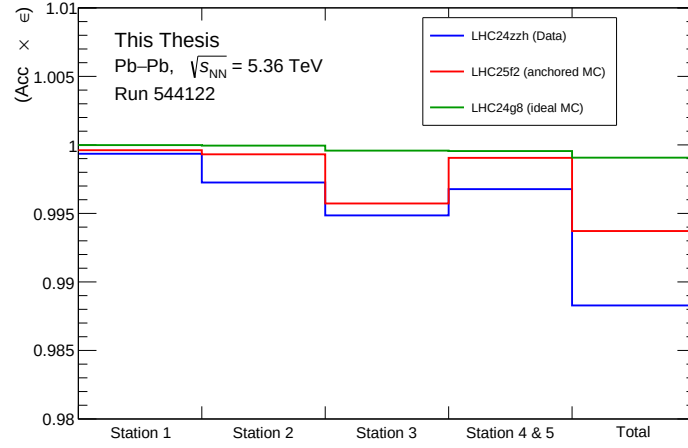


Figure 4.12: Muon tracking efficiency per station for run 544122. The initial MC is represented in green, the updated MC in red, and the data in blue.

As a last note, the integrated efficiency was calculated for all considered runs in the data and MC. The result is presented in Fig. 4.13. The initial MC has an almost constant efficiency with respect to the run number. In contrast, the efficiencies in data show a clear drop between runs 544095 and 544098. The apparition of the hole in chamber 6 causes this drop. Lastly, the updated MC is situated between the ideal MC and the data. The drop in efficiency is not as significant as in the data, but it is still present. This is expected, since the updated MC is designed to reproduce the conditions of the detector at the time of the run, but it does not include all the information that we have in the data. For instance, the current version of the anchored MC does not include the structures seen in chambers 3 and 4 in the data. All in all, the updated MC differs from the data by less than 1% in the integrated efficiency for all runs considered.

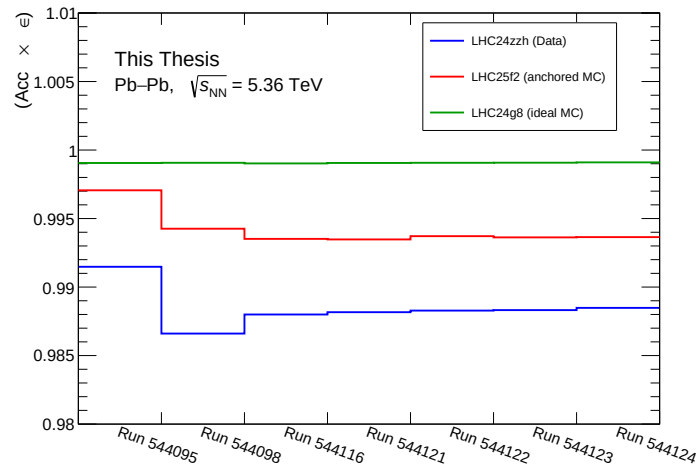


Figure 4.13: Integrated muon tracking efficiency for all runs considered. The initial MC is represented in green, the updated MC in red, and the data in blue.

4.5 Evaluation of the tracking systematic

The comparison of the muon tracking efficiency calculated from the data and from the updated anchored MC allows us to estimate its systematic uncertainty. In the previous section, we presented the muon tracking efficiency per run, chamber, and station, and we observed that the efficiencies in the anchored MC differed by around 1% from the data. To validate this difference, we studied the muon tracking efficiency as a function of p_T , η , and φ in the data and the anchored MC. The results are presented in Figs. 4.14a, 4.14b, and 4.14c.

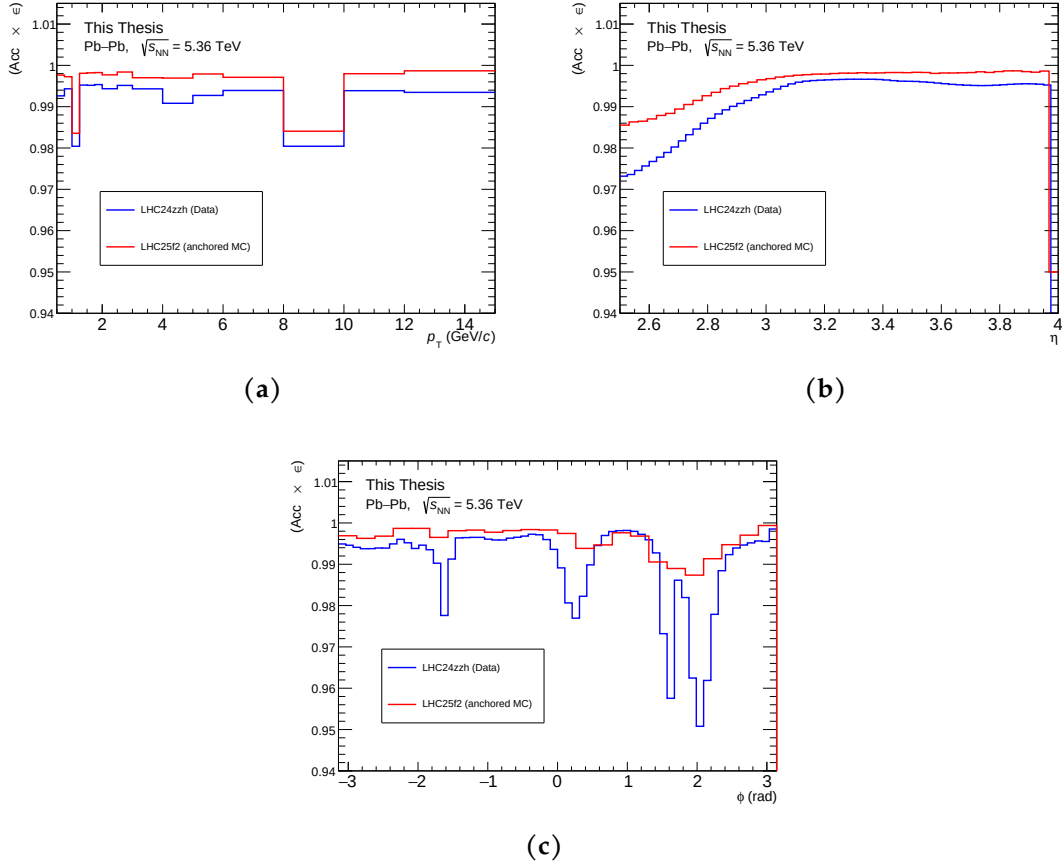


Figure 4.14: Integrated muon tracking efficiency for all Pb–Pb runs recorded in 2023 as a function of (a) p_T ; (b) η and (c) φ . The data is represented in blue and the anchored MC in red.

The distribution of the efficiency as a function of p_T shows that the behaviour of the data and the anchored MC is similar. Furthermore, there is a drop in the efficiency at $p_T \sim 10$ GeV/c that is present in both distributions. Overall, the discrepancies remain within 1%.

The efficiency as a function of η is also very similar in both distributions. In this case, we can see that the differences in the efficiencies are more pronounced between $\eta \sim 2.5$ and $\eta \sim 3.1$, reaching a maximum difference of 1%. At higher rapidity values, both distributions show differences of around 0.2%. Finally, at $\eta \sim 4$, the efficiencies drop for data and MC, possibly due to border effects.

Finally, the distributions of the efficiency as a function of φ display how the MC is capable of reproducing the overall trend of the data. However, in this case, the differences are more pronounced, especially in regions such as $\varphi \sim 2$ rad which show a difference of around 4%.

From the work in this thesis, we can assign a preliminary systematic uncertainty to the muon tracking efficiency, σ_{trk}^μ , of 1% for single muons with p_T above 0.5 GeV/ c measured in Pb–Pb collisions at $\sqrt{s_{\text{NN}}} = 5.36$ TeV taken in 2023. To account for the larger differences observed as a function of p_T , η and ϕ , we propose to use a global uncertainty of 1, 3% and 5%, respectively. These values were obtained as the maximum difference between the data and the anchored MC.

D⁰-MESON RECONSTRUCTION

This Chapter introduces the strategy followed to perform all D⁰ analyses included in this thesis. The Chapter starts by describing the criteria for selecting events and tracks and explains how tracks are associated with collisions in continuous-readout mode. The particle identification (PID) requirements are also introduced. Then, the specific workflow to reconstruct heavy-flavour particles in the framework employed in Run 3, O² Physics, will be introduced. The procedure to select D⁰ mesons using Machine Learning algorithms is explained, starting with a description of the algorithms themselves. The variables used to train the models, the training procedure, and their inference onto the O² Physics workflow are indicated. Lastly, a few words are dedicated to the ALICE Offline Trigger System to prefilter rare events in large datasets, such as those with double charm mesons.

In short, the objective of this Chapter is to go through all the steps described in Fig. 5.1. These go from the detection (or simulation) of the particles to the final physical analysis, which requires a prior particle reconstruction.

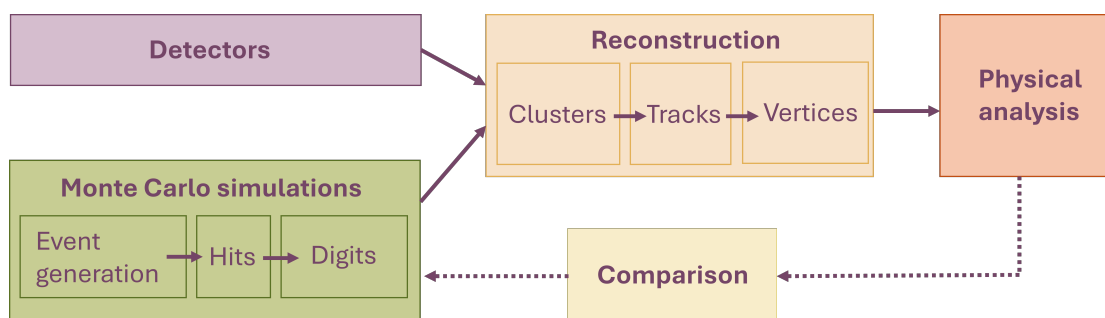


Figure 5.1: General structure of a high-energy physics analysis. This sketch is a modified version of the one from [233].

5.1 Real and simulated data samples

Experimental physics is based on the study of the *real* data collected with detectors in order to extract physical meaning out of it, which can only be done by comparing what

we measure with what our models expect. However, we need to take into account the fact that our detectors are also imperfect, either because of geometrical constraints or because of faulty detection elements. In addition, the efficiency of our selections to maximise our desired signal over all possible background sources also needs to be evaluated. In these situations, **simulating data** is of its best use.

Simulated data can be defined as fictitious data that is created to mimic the properties of real data. Simulations are commonly generated using **Monte Carlo (MC) methods**¹ in the context of high energy physics.

Monte Carlo simulations are mainly used to simulate the detector response and to estimate the acceptance-efficiency corrections (more on Sec. 5.10). The acceptance of a detector is defined as the geometrical range of the kinematic properties of a particle in which it can be detected. All simulations in this thesis are created using PYTHIA 8 [70]. The interactions of particles passing through the detectors are simulated using GEANT4 [234]. Furthermore, all MC simulations used in this thesis are heavy-flavour enriched to overcome the low probability of finding these kinds of events in nature. To do so, a $c\bar{c}$ or a $b\bar{b}$ event was injected in the $|y| < 1$ region every five events, and the generated D^0 , D^+ and D_s^+ mesons were forced to decay into their channels of interest.

Depending on the level of detail of the simulation of the detector conditions, it is possible to distinguish two types of MC:

- **Ideal Monte Carlo:** The nominal geometry of the detectors is recreated, as well as the materials that compose them.
- **Anchored Monte Carlo:** On top of the ideal simulations, the conditions of the detectors at the time when a specific dataset was taken are recreated, such as faulty detection units, dead zones, or calibrations. We say that an MC is *anchored* to a specific production.

5.2 Event reconstruction and selection

In particle physics experiments, particles are detected indirectly through elements like energy depositions, timestamps, hit positions, etc. These elements are collected in raw data samples, which are not directly usable for physics analyses. A reconstruction process is needed to build up collisions from these depositions and obtain information such as the fitted particle momenta, coordinates of interaction and decay vertices, or normalised PID probabilities. Each of these detected collisions is called an **event**.

Event reconstruction generally begins by creating **clusters**, groups of neighbouring detection units (such as pixels in the ITS), which get flagged simultaneously. Each cluster corresponds to a **hit** in a detector, the basic building block of track and vertex reconstruction [235]. A **vertex** refers to a point in the collision where a particle

¹ Monte Carlo generators are a set of computational algorithms which use a controlled random sampling to obtain numerical results.

is created. We can distinguish between **primary** (PV) and **secondary** (SV) vertices. The primary vertex corresponds to the collision point. Secondary vertices are those reconstructed from decay particles, which are displaced from the PV by some mm up to a few cm.

5.2.1 Event selection

Out of all the reconstructed events, only those that pass a set of criteria are considered for physics analyses: rejecting pileup, interactions of the beam with the gas, and poor-quality events. This process constitutes the main challenge of ALICE in Run 3 due to the continuous data-taking mode. The drift time of the ionisation electrons produced by the charged particles crossing the TPC is large (around $100\ \mu\text{s}$), leading to an ambiguity in the z -time component of tracks reconstructed solely with the TPC. Also, the ITS records data in integrated time-frames of $5\ \mu\text{s}$ for pp collisions and $15\ \mu\text{s}$ for Pb–Pb collisions, limiting the timestamp determination precision. Thus, the most probable **bunch crossing** (BC) for a certain collision will have a limited time precision.

To overcome the limitations mentioned above and be able to select physics events, a matching of the information gathered with the FIT detector is required. In particular, **minimum bias** (MB) events² are selected using the so-called **TVX selections**, which require a coincidence of the FT0-A and FT0-C signals. However, TVX events might be shifted with respect to the BCs found using the FIT, especially in collisions with a high **interaction rate** (IR). To mitigate this effect, the FT0 information of the neighbouring BCs is used to find the best possible match.

All selected events in this thesis were required to satisfy the `se18` condition, a standard selection based on the timing information of the FT0 detector. This ensures that collisions are temporally close enough to be associated with a bunch crossing. At the analysis level, each reconstructed vertex is associated with the most likely bunch crossing by replaying this timing-based association. Additional cuts were applied to exclude TPC and ITS events reconstructed near the time frame border. Only collisions passing these criteria were considered in further steps.

5.2.2 Primary vertex reconstruction

Primary vertices are fundamental ingredients to reconstruct events. Their position is estimated in two steps using different algorithms. First, only the hits in the three innermost layers of the ITS are used. A set of **tracklets** is created by connecting the hits in two consecutive layers with straight lines. The magnetic field is ignored at this stage.

In the continuous read-out mode of Run 3, tracks cannot be directly associated with collisions. Instead, detectors have associated time windows, or time frames (TF), di-

² Minimum bias events are selected using a loose trigger, with the purpose of capturing the maximum amount of potentially interesting interactions in a certain collision.

vided into smaller Read-Out frames (ROF) of $5 \mu\text{s}$ for pp and $15 \mu\text{s}$ for Pb–Pb data. To construct vertices, tracklets measured in the same Read-Out Frame are required to be close to each other so that their distance of closest approach (DCA) is smaller than $400 \mu\text{m}$. The vertex is computed as the centroid of these tracklets and is recalculated each time a new tracklet is found within this distance. Only vertices associated with at least 16 hits and close to the interaction point in the transverse plane are kept.

The first reconstruction step gives a fast estimation of the primary vertex position, but it is not very precise since it does not take into account the curvature of the tracks due to the magnetic field. For this reason, a second step of the algorithm is employed, exploiting the full track information once tracks are reconstructed (see section 5.3.1 to learn more about the track selection and section 5.4 for the track-to-collision association). The procedure followed is equivalent to the first vertex reconstruction, but the track characteristics and uncertainties are more precise at this stage.

5.3 Track reconstruction and selection

This section explains the process followed to reconstruct and select charged particle tracks. The full track reconstruction is based on the **Kalman filter algorithm** [228, 229], which combines the information of the detectors located in the central barrel, namely the ITS, TPC, TRD and TOF detectors. It consists of the steps shown in Fig. 5.2.

Tracking starts at the outermost part of the TPC, where a **track seed** is created from a cluster of energy depositions. Then, the seed is propagated inwards, through the rest of the TPC and down to the ITS, where it gets matched to the clusters reconstructed using this detector. The resulting ITS-TPC tracks are propagated outwards, from the innermost ITS layer to the TRD and TOF detectors. Lastly, bad-quality and fake tracks are removed by applying a set of selections. The remaining tracks are the ones used for vertexing following the procedure introduced in the previous section. The details of each of these steps are presented below.

TPC reconstruction

Each sector of the TPC is composed of 158 **pad rows**, whose tasks are to detect the signal left by a particle as it passes through the detector. By detecting the hits per pad row, clusters can be created, which are used to compose TPC tracks. TPC tracks lack precise information along the z -axis due to the unknown **interaction time** (t_0). This is a direct consequence of the continuous readout mode established in ALICE in Run 3. The data are taken in time windows of the order of 10 ms, whereas the average drift time of a particle inside the TPC is around a hundred times smaller ($\sim 100 \mu\text{s}$) [236]. For this reason, it is not possible to distinguish when a particle was produced.

A plausible t_0 hypothesis must be assigned to accurately evaluate cluster corrections

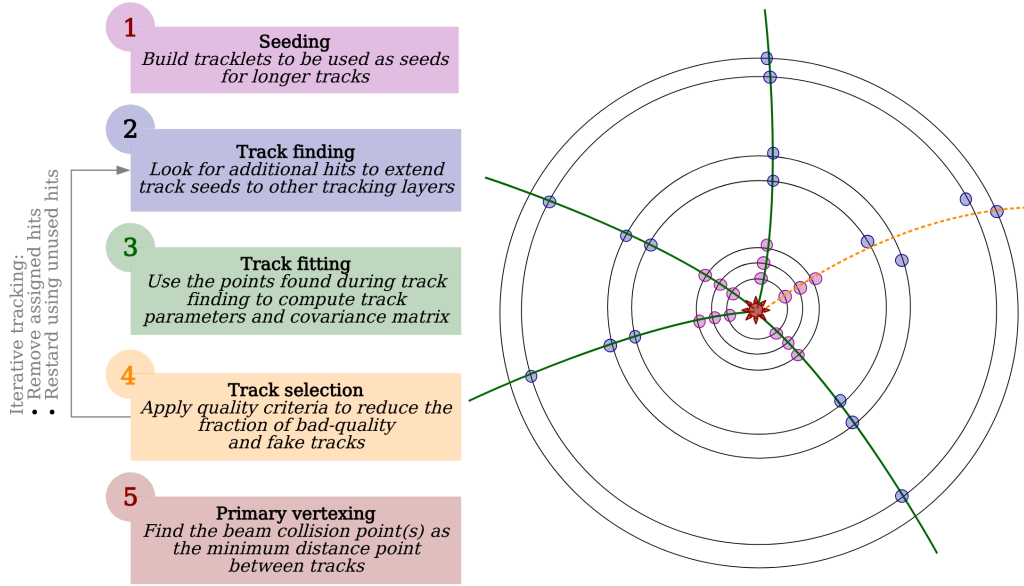


Figure 5.2: Description of the tracking process. The figure contains a schematic representation of the ITS, with hits represented as circles and traces as green lines. The dashed orange line corresponds to a fake track. Image taken from [233].

and fit errors. These corrections account for slow changes in the luminosity, ambient conditions and variations in the collision frequency and centrality [237]. When a track can be propagated to the beamline, it is reconstructed under the assumption that $z = 0$ at the beamline. Otherwise, it is assumed that the track originates from the decay of a primary particle located at $|\eta| = 0.5$. A **time bracket** is then defined around the estimated t_0 , ensuring that under any time or z assignment, the edge clusters remain within the TPC volume. If the track has clusters on both sides of the TPC, its z -position is well constrained, and a timing uncertainty of $1 \mu\text{s}$ is assigned.

ITS reconstruction – ITS-TPC tracks

ITS tracks are characterised by their precise positions, simplifying the cluster unfolding process. Despite this precision, the clustering process is still a challenge due to the proximity of the ITS to the beamline, which greatly increases the number of hits detected. Furthermore, ITS tracks have an uncertainty on the time within the ITS read-out frame. In Run 3, standalone ITS tracks are created and later matched to TPC tracks, whose position z position has large uncertainties due to the unknown t_0 . The matching is performed by creating pairs of ITS-TPC tracks in overlapping time brackets. The TPC track is moved along the z -axis to match the position of the ITS track, and the χ^2 of each pair is computed. Multiple matches are possible at this stage for each ITS and TPC track. Later, the best ITS-TPC track is chosen as the one with the minimum χ^2 . At last, the remaining TPC tracks are matched with unused ITS tracks.

Propagation of tracks to the TRD and TOF detectors

ITS-TPC tracks, which have a well-defined z position, get propagated to the TRD. In a similar fashion to how ITS and TPC tracks were matched, a hypothesis tree is built per candidate, and the “best winner” per ITS-TPC track is chosen as the one with the minimum χ^2 . The time of the TRD trigger is then selected as the track timestamp. The same procedure is repeated for the TOF detector.

Lastly, matched tracks which had a contribution in the innermost layer of the ITS are propagated back to this detector to generate the primary vertex following the procedure described in Section 5.2.2. These tracks are called **primary vertex contributors**. Every other track is associated with a PV as long as they are time-compatible. A single non-contributor track can be associated with several PVs.

5.3.1 Track selection

Tracks are required to pass some requirements to be considered in further analysis steps. These include loose kinematic cuts on the distance of closest approach (DCA), p_T , and η , and quality requirements on the tracks that traverse the ITS and TPC. Tracks that pass all of these selections are called **global tracks** and were employed in all analyses in this thesis. The specific values for each cut can be found in Table 5.1.

Table 5.1: List of track selection conditions for global tracks [225].

Selection cuts	
Min. of TPC crossed rows	70
Min. ratio of crossed rows over findable TPC clusters	0.8
Max. χ^2 per TPC cluster	4.0
Max. χ^2 per ITS cluster	36.0
Require TPC refit	True
Require ITS refit	True
Max. DCA to vertex in the beam axis (cm)	2.0
Max. DCA to vertex in the transverse plane (cm)	$0.0105 + 0.035/p_T^{1.1}$
Cluster requirement ITS	At least one hit in the 3 innermost layers
p_T range (GeV/ c)	(0.1, 10^{10})
η range	(−0.8, 0.8)

5.3.2 The track-tuner feature

In Run 3, a discrepancy is observed between the reconstructed DCA resolution in data and in MC simulations. These differences arise from several sources, including imperfect knowledge of the detector alignment, inaccuracies in the description of the detector material budget, and residual calibration effects that are not fully reproduced in the MC. To diminish the difference, the so-called *track-tuner* feature was introduced to smear the DCA resolution in MC and make it closer to the values found in the data. It also worsens the MC p_T resolution by a factor of 1.5. The track-tuner feature was used in all MCs involved in this thesis.

5.4 Track-to-collision association

As introduced in 3.3, the continuous read-out strategy followed in Run 3 is handled by separating data into time frames. However, different detectors have different default track time windows or Read-Out Frames (ROF). Moreover, due to the increased interaction rate, bunch crossings corresponding to different collisions can fall into the same TF. Tracks need to be reassigned to the collision that produced them, which can be done using the approximate time each track crossed a specific detector and the time of each collision.

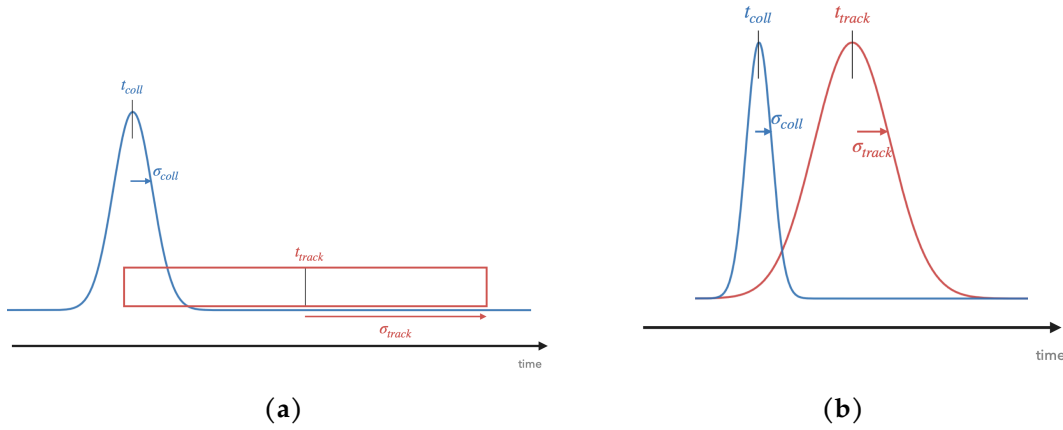


Figure 5.3: Track-to-collision association mechanism for detectors that take data in a set time frame (a) and for detectors with a gaussian-like resolution (b). In red, detected track time, and in blue, detected collision time [225].

The procedure to associate tracks and collisions depends on how each detector records the time information. Some detectors, such as the ITS, collect track information in allocated time windows. In these cases, track time, t_{track} , is recorded as the centre of the time window. The uncertainty (σ_{track}) is represented by a rectangular distribution covering the whole duration of the window (Fig. 5.3a). In other cases, the track time distribution is Gaussian-like, as in the TPC. Therefore, t_{track} is recorded as the peak of the Gaussian, with an uncertainty corresponding to its width (Fig. 5.3b). The same is done for the collision time t_{coll} , which always has a Gaussian distribution. To associate t_{track} and t_{coll} , the difference between the two is required to be smaller than a **threshold time**, which depends on the type of detector and an external parameter that provides the maximum number of standard deviations allowed ($n\sigma_{max}$), set to 4. This threshold time, τ , is defined as

$$\tau = \sigma_{track} + n\sigma_{max} \cdot \sigma_{coll} + \text{timeMargin} \quad (5.1)$$

for tracks coming from detectors that collect information in rectangular time windows. The parameters σ_{track} and σ_{coll} represent the uncertainties of t_{track} and t_{coll} , respectively, and the time margin is an optional parameter added to account for the possibility of not having the TPC fully calibrated. For tracks with a Gaussian-like error, such as TPC

tracks, the threshold time is computed as:

$$\tau = n\sigma_{\max} \cdot \sqrt{\sigma_{\text{track}}^2 + \sigma_{\text{coll}}^2} + \text{timeMargin}. \quad (5.2)$$

5.5 PID selections

Particle identification of charged particles, such as pions or kaons, is accessed using the information of the TPC and TOF detectors. PID information is usually accessed in units of the **experimental resolution**, σ , defined as:

$$n\sigma_{\text{TPC, TOF}}^{\pi, K} = \frac{|S_{\text{measured, TPC, TOF}} - S_{\text{expected, TPC, TOF}}|}{\sigma_{\text{TPC, TOF}}}. \quad (5.3)$$

For a given mass hypothesis (electron, muon, pion, kaon, proton, deuteron, etc.), the experimental resolution expresses the difference between the measured signal, S_{measured} and the expected one, S_{expected} , in the TPC and TOF. In the case of the TPC, particle identification is based on the measurement of the deposited energy (dE/dx) in the detector. This energy loss is parametrised with the Bethe-Bloch formula, which is further introduced in Section 3.2.1. The detected energy is combined into clusters, which are used to reconstruct tracks. The cluster finder used for all analyses in this thesis employs a neural network (NN) algorithm instead of the default clusterizer used in previous Runs. The switch to an ML-based approach improves the resolution of around 30% [238]. NNs are used to decide whether a hit belongs to a cluster and to extract the properties of the mentioned cluster. The resulting $n\sigma$ distributions corresponding to the pion and kaon hypotheses as a function of momentum can be observed in Fig. 5.4a and 5.4b, respectively.

The PID of the TPC in MC is corrected using the *tune-on data* feature. This feature is used to parametrise the MC distributions using the data. A Gaussian distribution centred at the average dE/dx is extracted from the data and used to correct the $n\sigma$ MC distributions as a function of the track momentum.

The TOF detector can distinguish particles by measuring the time it takes them to fly from the vertex to the detector. An introduction to this detector and its detection mechanism can be found in Section 3.2.1. It is worth mentioning that a bug was found in the T0 digitalisation of the MC simulations, causing post-calibrations to be necessary to generate the $n\sigma$ TOF distributions. Figs. 5.4c and 5.4d contain the $n\sigma$ distributions for the pion and kaon hypotheses.

5.6 The heavy-flavour workflow in O^2 Physics

The heavy-flavour (HF) workflow starts by gathering the information presented in the previous sections as input, namely the event selection, collisions, tracks and PID. These tables are introduced in the `track-index-skims-creator`, which creates a set of

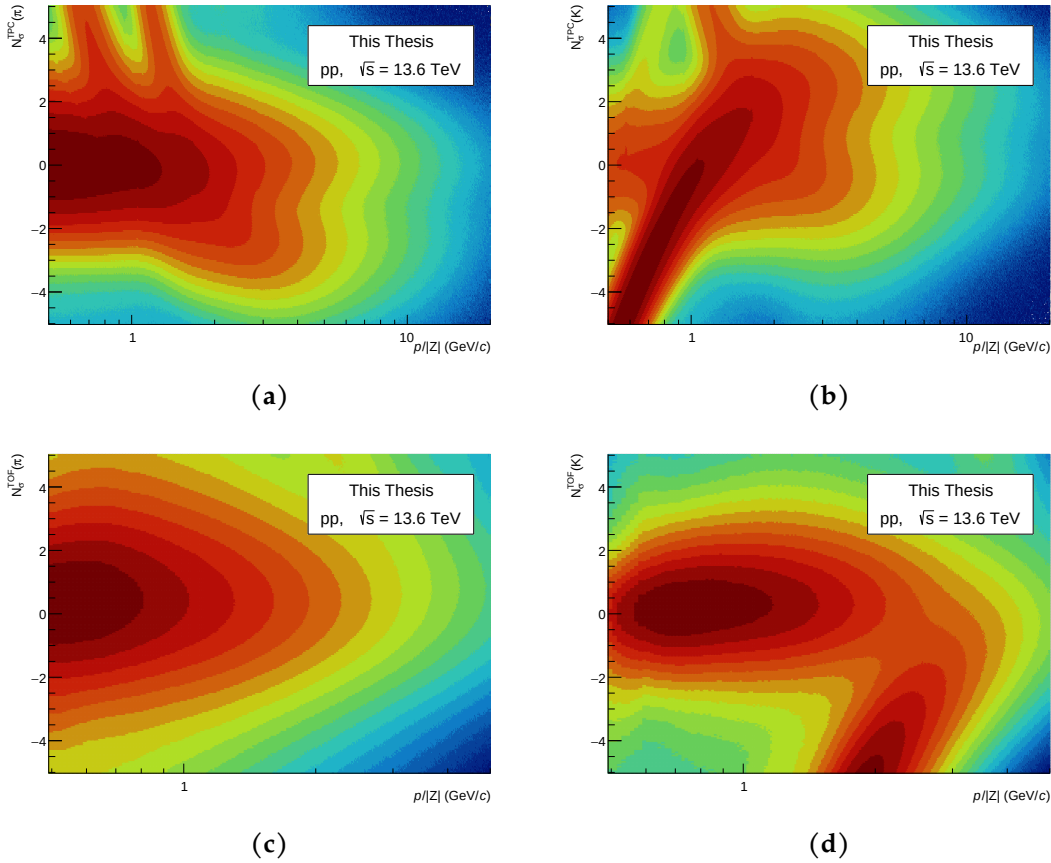


Figure 5.4: Upper row: $n\sigma$ TPC distributions for (a) pions and (b) kaons corrected using a NN as a function of the momentum. Bottom row: $n\sigma$ TOF distributions for (c) pions and (d) kaons as a function of the momentum.

two and three-prong *protocandidates* using very loose selection cuts, with the purpose of **skimming**³ the data that will enter the rest of the workflow. Afterwards, *protocandidates* enter the candidate-creator, where they **get fully reconstructed as two and three-prong decay candidates**. When running on an MC sample, it is at this stage that particles get matched to their generated counterparts.

All created candidates go to the next step, where they get selected as specific HF hadrons. For instance, to reconstruct D^0 meson candidates, one should use the helper task named `candidate-selector-d0-to-k-pi`, where selections based on kinematics and PID are made to keep D^0 candidates reconstructed in the $D^0 \rightarrow K^+\pi^-$ channel. These selections will be explained in the following sections. At last, the final output is obtained in the analysis tasks, such as `task-d0`, which creates histograms that can be used for further local analysis. At this stage, other tasks can also be made to, for instance, study correlations between particles.

The complete heavy-flavour workflow can be observed in Fig. 5.5 [225]. It is structured as a combination of different interconnected boxes: in yellow, input AOD tables; in

³ Skimming refers to getting a first selection of potentially interesting candidates to reduce the data size.

green, processes; in grey, derived tables produced by the corresponding processes; and in blue, final outputs. HF processes are put together in dedicated tasks, the scope of which is marked by orange boxes.

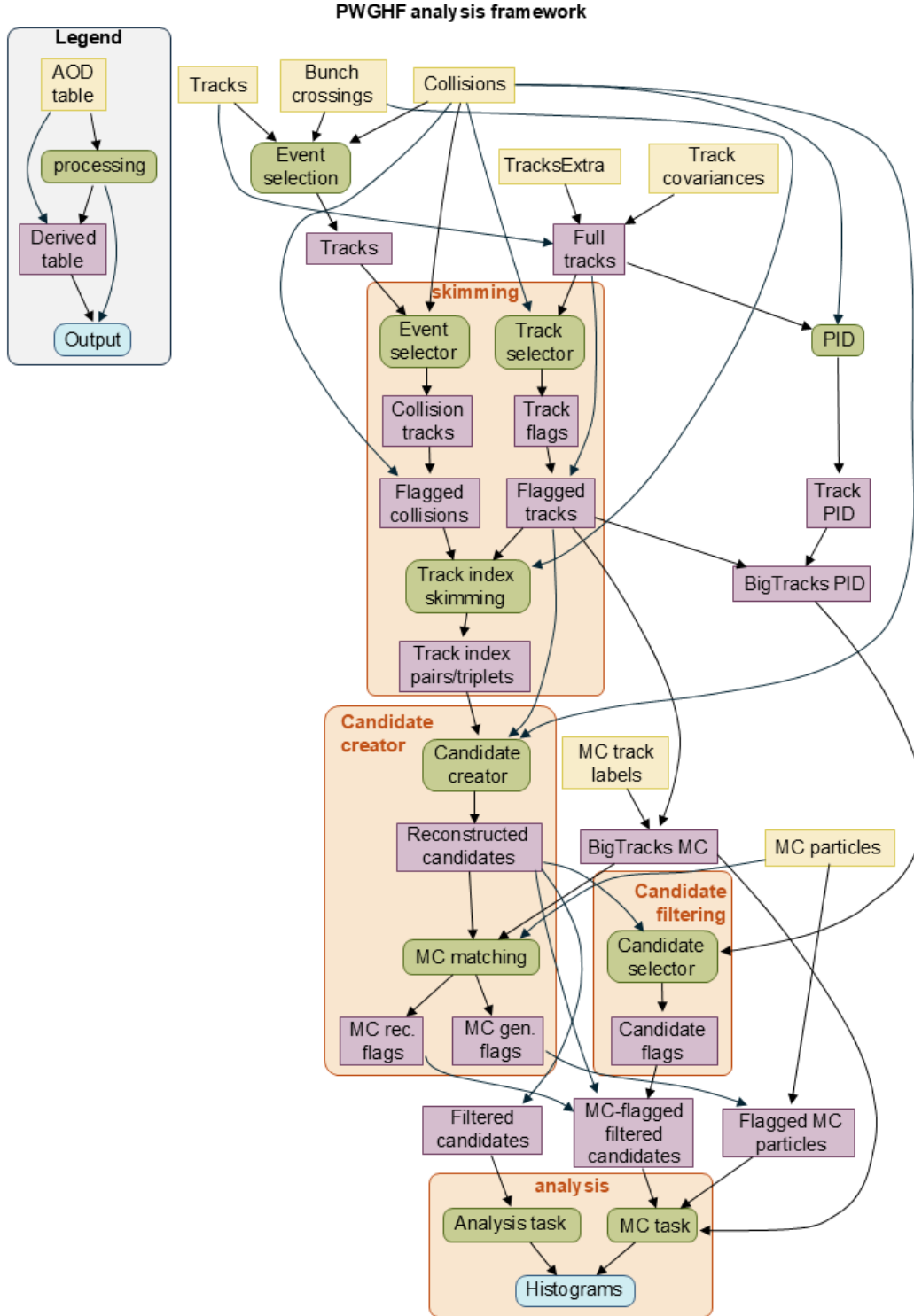


Figure 5.5: Schematic view of the main components of the HF workflow, from the initial skimming to the tasks developed by analysers. Image adapted from [225].

5.7 A (very) brief introduction to Machine Learning

Machine Learning (ML) [239, 240] is a statistical learning and optimisation process based on data exposure and pattern identification, mimicking how humans naturally acquire new information. Most ML algorithms start by establishing a **decision process**. Then, they evaluate the quality of the resulting model using an **error function**, which compares the guesses of the model with known examples. As a last step, ML algorithms can review the wrong guesses and update their decision process in order to **optimise the results**. This process is iterative and automatic and can be performed until a certain **accuracy threshold** is met. Hence, ML is a very powerful tool that can be used to classify objects into groups (**classification**) as well as to predict numerical values (**regression**).

Depending on how models learn new information, ML algorithms can be classified as:

- **Supervised learning algorithms:** They are characterised by using labelled data to train the models.
- **Unsupervised learning algorithms:** They use unlabelled data and identify patterns without help from the user.
- **Semi-supervised learning algorithms:** They use labelled and unlabelled data. It is useful when it becomes too costly to label all data.
- **Reinforcement learning algorithms:** They are based on a "reward/punishment" system. It does not use sample data for training but instead learns as it goes from trial and error.

In the context of this thesis, a set of supervised ML algorithms called **Boosted Decision Trees (BDTs)** are used to separate between classes, namely signal (prompt and/or non-prompt D^0 -meson candidates) and background. The purpose of this section is to introduce BDTs and how they are used to select D^0 mesons. Special attention is paid to determining the variables for the training and the model quality evaluation. To illustrate each introduced concept, an example BDT is used, which was developed for one of the analyses that will be described in this thesis.

5.7.1 Decision tree classifiers (DTs)

Decision Trees (DTs) are a non-parametric supervised learning method for classification and regression. They are composed of a collection of questions and have a hierarchical structure reminiscent of a tree. The first question, or node, is called the **root node**, out of which two branches are created. These branches can contain either a prediction, in which case the branch is called a **leaf node**, or another condition that will subsequently create more branches with more conditions or leaf nodes. Fig. 5.6 shows a simple visualisation of how this method works. In this case, all decision nodes contain questions that provide only two options, "yes" or "no". Trees of this sort are

called **binary decision trees**, as opposed to **non-binary decision trees**, which have more than two possible outcomes.

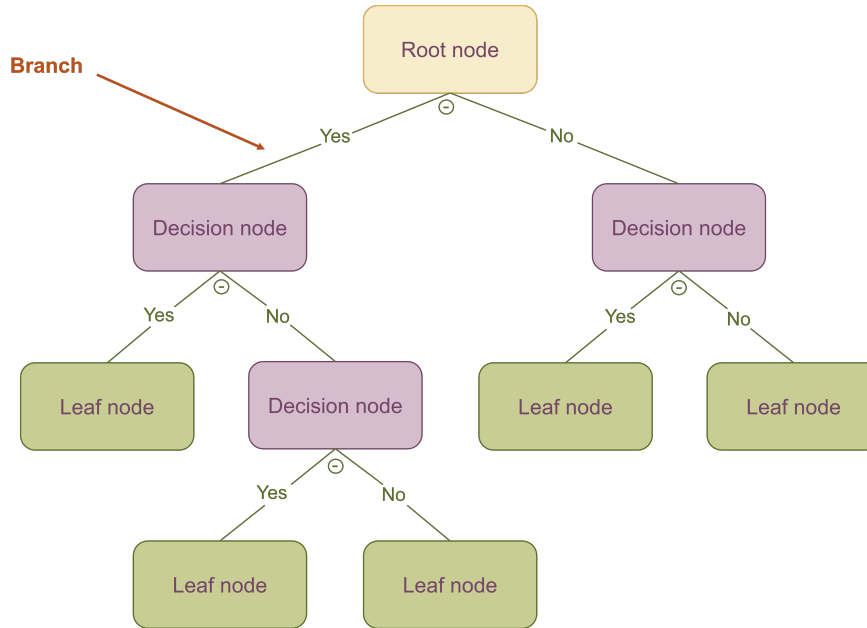


Figure 5.6: Structure of a binary decision tree.

Decision trees learn by applying a *divide-and-conquer* strategy to search for the optimal split points within a tree. The process of splitting is repeated until most (ideally all) candidates are classified. As a result, decision trees separate the input data into rectangular subsets and label each rectangle with their corresponding class.

Advantages of using decision trees include:

- **Easy to visualise and interpret:** Since they are based on Boolean logic, it is possible to visualise the whole tree structure that lead the model to a certain prediction. Also, it is possible to study the relative importance of each feature.
- **Little data preparation required:** As opposed to other techniques, decision trees can handle several data types and missing values.
- **Flexible:** decision trees can handle problems with several possible outputs. They are also insensitive to correlations between features; if two features are highly correlated, the algorithm will only perform a split using one of them.

Nonetheless, there are also possible disadvantages that need to be considered, mainly concerning the model stability. First of all, trees can have significant variations when small changes are made in the input data. In addition, decision trees become more **prone to overfitting** as the model complexity grows. If a tree is too large, some subtrees may contain too few data points, making the resulting model more prone to learning from random variations in the data rather than capturing generalisable patterns. To minimise this problem, trees can be **pruned** so tree growth is stopped when data are insufficient before (pre-pruning) or after (post-pruning) tree reconstruction.

5.7.2 Boosted Decision Trees (BDTs)

Combining multiple models is one method to overcome the difficulties DTs pose without losing their advantages. Amongst the possible ensemble methods, one of the most used, and the one that was used for this thesis, is boosting, which creates Boosted decision tree (BDT) models.

A simple representation of how a BDT works is presented in Fig. 5.7. The method involves comparing two models: the weak learner, which is a single decision tree (DT), and the strong learner, which is an ensemble of multiple weak learners combined. The process begins by training a weak learner on the dataset, making predictions, and identifying misclassified points. These misclassified points are then assigned higher weights, which increases their importance in the next iteration. A second weak learner is trained on this reweighted dataset, improving the performance of the model.

This process is repeated iteratively, with each new weak learner focusing on the errors of the previous ensemble. The final strong learner is formed by combining all weak learners, typically using a weighted voting or averaging scheme. The iterations continue until a predefined stopping criterion is met, such as reaching a maximum number of iterations or detecting signs of overfitting, such as performance stagnation or degradation on validation data.

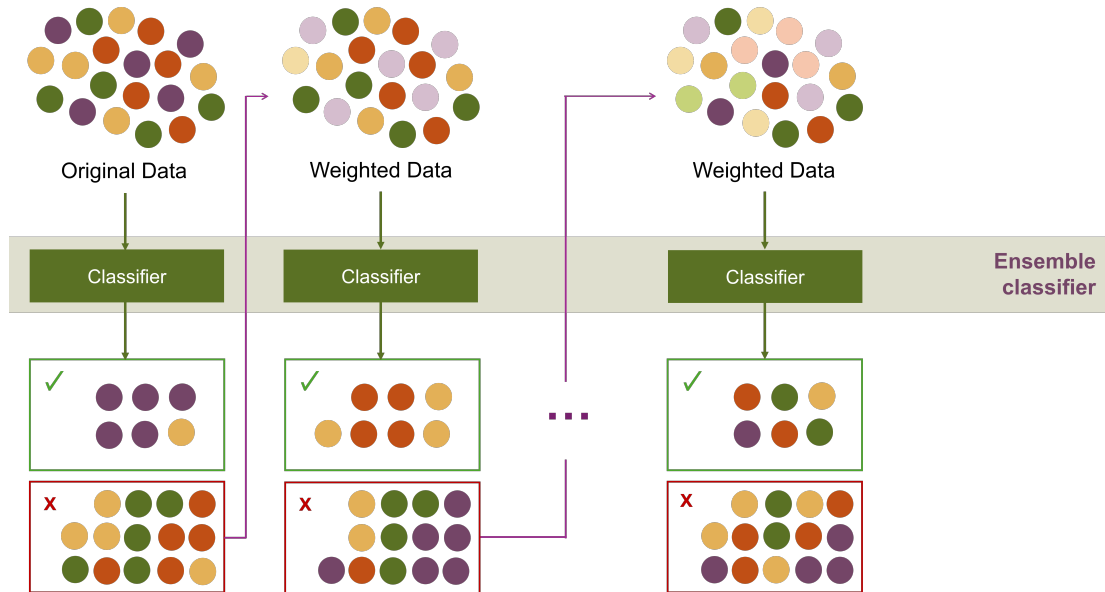


Figure 5.7: Structure of a BDT.

In the context of this thesis, the algorithms mentioned above were used through the `hipe4ML` [241] library. It was developed in ALICE to perform ML-based physics analyses. It is optimised to convert ROOT TTree objects into data frames that can be manipulated using the Python library Pandas [242, 243] and to wrap models from different ML libraries to be able to use them in one single format. At present, it supports LightGBM [244], XGBoost [245], and sklearn [246] models.

All models presented in this thesis were trained using XGBoost. The name of the library comes from *Extreme Gradient Boosting*, and its algorithm is based on the CART (Classification and Regression Tree) algorithm [247], allowing it to deal with both classification and regression problems. The CART algorithm is a modified version of a simple DT known for its flexibility and scalability. XGBoost is also a notoriously fast algorithm, thanks to its capacity to be executed in CPUs and GPU parallel. Thus, it has become a state-of-the-art algorithm for data analysis studies that use Machine Learning.

5.8 Reconstructing D^0 mesons using Machine Learning

This thesis will present a measurement of the non-prompt D^0 fraction and the double production of D^0 mesons. Even though both measurements require selecting D^0 mesons, and this selection was made using an ML approach, there are differences in the preselections before this step. Thus, two different types of BDT models were trained: a three-class model for the non-prompt D^0 fraction and a binomial classification model for the double D^0 production measurement. Nonetheless, the final goal in both cases is to optimise the D^0 signal extraction over the background. Hence, a lot of the characteristics of these models are shared, such as the main variables used for the ML training or the study of the model quality. We aim to explain these common features in this section.

5.8.1 D^0 mesons in ALICE

All D^0 mesons (and their charge conjugates) included in this thesis were measured in pp collisions at $\sqrt{s} = 13.6$ TeV and reconstructed in the $D^0 \rightarrow K^- \pi^+$ channel. The reason why this channel was chosen is that its **branching ratio (BR)** is the largest among all hadronic decay channels [9]. Compared to their possible decays, such as the semi-inclusive or leptonic decays, hadronic channels do not have to account for missing energy, allowing for a complete reconstruction of the D meson. The decay vertex position is determined by minimising the distance between the two decay tracks. A summary table of our D meson of interest properties can be found in 5.2. This particle was chosen due to its abundance: approximately 40% of all charm hadrons produced in the initial hard-scattering hadronise into D^0 mesons. We call these types of particles **prompt** D^0 . These mesons can also provide indirect input in the study of open beauty production, as around 70% of B mesons decay to D mesons. In this case, D^0 particles, which are not produced in the hadronisation of charm quarks but rather due to a beauty decay, are called **non-prompt** D^0 .

Table 5.2: General characteristics of D^0 mesons. Information obtained from [9].

Particle	Mass (GeV/ c^2)	Mean life (fs)	Decay channel	Branching ratio (%)
D^0 ($c\bar{u}$)	1864.84 ± 0.05	420.3 ± 1.0	$D^0 \rightarrow K^- \pi^+$	3.947 ± 0.030

The signal yield is extracted once all selections are made by performing an invariant mass fit. In this fit, it is necessary to consider the D^0 signal, the combinatorial background, and the $\bar{D}^0$ candidates, which are misidentified as D^0 mesons.

Say we have a D^0 , which we detect by searching for a negative kaon and a positive pion. It can happen that the prong candidates can be simultaneously assigned to a kaon and a pion. In this case, the reconstruction algorithm will reconstruct two candidates, one with a positively-charged pion and a negatively-charged kaon and another one with the opposite charges. Thus, in addition to the D^0 that we have in nature, we will have reconstructed an extra $\bar{D}^0$ meson. The resulting $\bar{D}^0$ is called a **reflection** and is characterised by having the wrong mass hypothesis. Reflections constitute a background of all D^0 analyses, which can be minimised by tightening the selection cuts on the PID variables. Nonetheless, other D mesons, such as the D^+ , do not have this issue, since one can distinguish them by looking at the charge of the reconstructed particle.

The details of how the signal is extracted and how the reflections are treated in the two analyses of this thesis will be introduced in their respective dedicated Chapters (6 and 7). For now, we will focus on explaining how D^0 mesons are selected using ML.

5.8.2 Main variables to reconstruct D^0 mesons

The chosen single track selections, introduced in 5.3.1, combined with the geometrical limitations of the detectors in the central barrel, result in a reduction of the **detector acceptance**. In particular, we can define a **fiducial rapidity region**, y_{fid} , to describe the acceptance in the central barrel:

$$y_{\text{fid}}(p_T) = \begin{cases} 0.5 + \frac{1.9}{15}p_T - \frac{0.2}{15}p_T^2 & \text{for } p_T < 5 \text{ GeV}/c \\ 0.8 & \text{for } p_T > 5 \text{ GeV}/c \end{cases} \quad (5.4)$$

To avoid introducing border effects at low- p_T (*i.e.* $p_T < 5 \text{ GeV}/c$), all analyses in this thesis were computed in the $|y| < 0.5$ region.

Topological and kinematic variables are used together with the PID information of each candidate daughter to reconstruct D^0 mesons. The topological variables are typically used in other analyses of prompt D mesons performed without ML techniques [248]. Figure 5.8 contains a sketch of the main observables for prompt and non-prompt candidates, which are introduced in the following.

- **Decay length (d_0, d_0^{xy})**

The decay length, d_0 , is defined as the distance a particle travels (in this case, the D^0 meson) before decaying into its daughters. In other words, the distance between the primary vertex (PV), where the particle is created, and the secondary

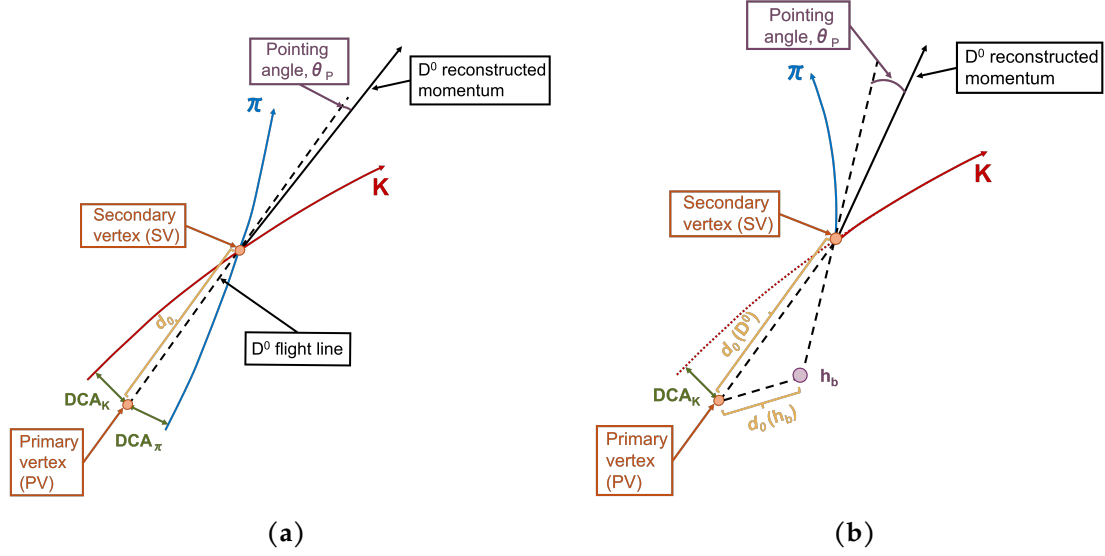


Figure 5.8: Representation of the decay of a (a) prompt and (b) non-prompt D^0 meson.

vertex (SV), where it decays. This three-dimensional variable and its transverse projection (d_0^{xy}) are particularly useful to distinguish between prompt and non-prompt particles.

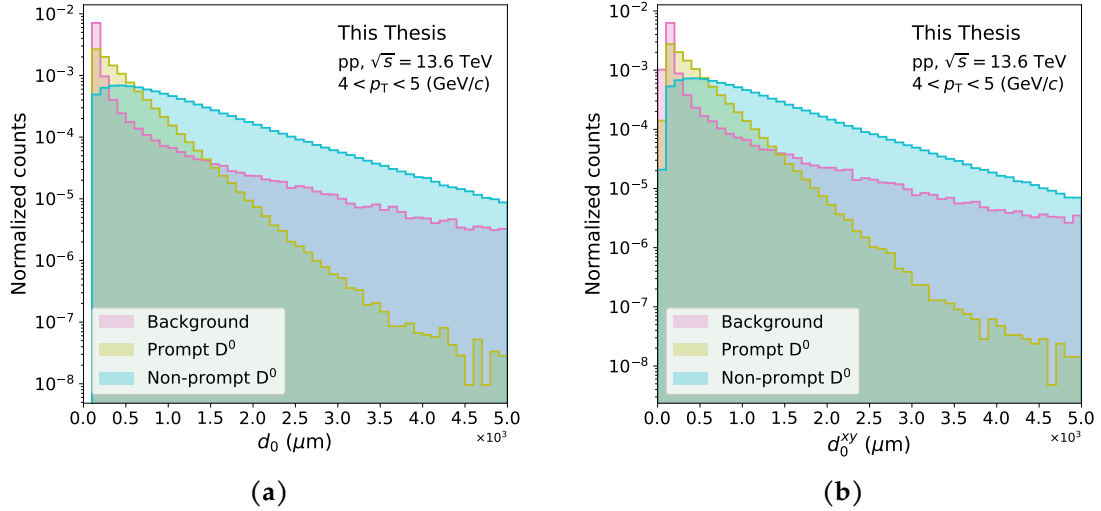


Figure 5.9: Decay length distributions of background, prompt D^0 and non-prompt D^0 candidates in the $4 < p_T < 5$ GeV/c range.

Figures 5.9a and 5.9b present the decay lengths of D^0 -meson candidates in pp collisions at $\sqrt{s} = 13.6$ TeV in the $4 < p_T < 5$ GeV/c region. The prompt and non-prompt D^0 distributions were extracted from MC simulations, while the background was extracted from data. Looking at the background distributions, we can observe that they have a peak around zero. This is because most background candidates come from pion and kaon tracks located around the PV, which do not belong to a real D^0 , making their decay lengths close to zero. Nonetheless, their

distribution is widened by resolution effects caused by the detectors. On the other hand, signal D^0 candidates (prompt or non-prompt) are expected to travel for some distance before decaying ($d_0^{\text{prompt}} \sim 300\mu\text{m}$, $d_0^{\text{non-prompt}} \sim 500\mu\text{m}$). On average, the decay length of non-prompt D^0 mesons is larger than that of prompt D^0 candidates since their decay length also includes the distance travelled by the parent beauty hadron. Finally, the decay length of signal candidates is expected to increase with the transverse momentum because of the Lorentz boost.

- **Distance of closest approach (DCA)**

The distance of closest approach (DCA) describes the minimum distance between the PV and the decay particles. As a general rule, the DCA of the decay prongs of a prompt D meson (a pion and a kaon, in the case of the D^0) will tend to be smaller than that of a non-prompt D meson. This is due to the displacement of the decay vertex of the D in the latter case. In this thesis, the chosen variable to compactify this information is the product of the DCA of the two decay prongs, written as $\text{DCA}_\pi \times \text{DCA}_K$, which is represented in Fig. 5.10. The figure contains the DCA product measured in pp collisions at $\sqrt{s} = 13.6$ TeV in the p_T range between 4 and 5 GeV/c. As can be observed, the DCA product of prompt D^0 candidates is found to be wider than that of background candidates because of the displacement of the decay vertex. Similarly, the DCA product of non-prompt D^0 candidates is wider than that of prompt candidates, as their decay vertex is even more displaced as a consequence of the presence of a beauty hadron in the decay chain.

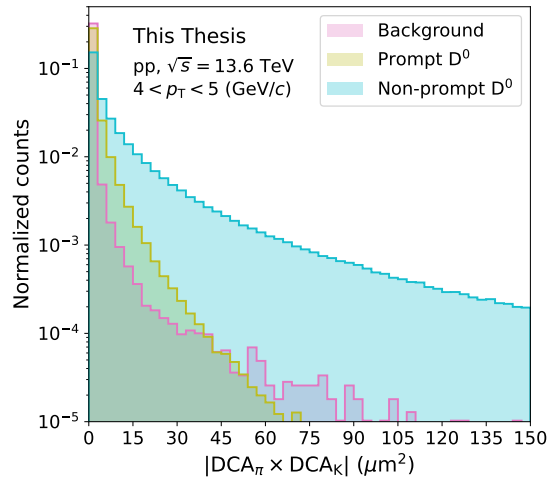


Figure 5.10: Product of the distance of closest approach distributions of the two decay particles for the background, prompt D^0 and non-prompt D^0 candidates in the $4 < p_T < 5$ GeV/c range.

- **Cosine of pointing angle ($\cos(\theta_p), \cos(\theta_p^{xy})$)**

The pointing angle, θ_p , is defined as the angle between the mother-particle flight line and its reconstructed momentum. Since we can determine the position with a higher precision in the xy plane, its projection is also employed in the training. The distributions of the cosine of pointing angle and its transverse projection in

pp collisions at $\sqrt{s} = 13.6$ TeV in the p_T range between 4 and 5 GeV/c can be found in Fig. 5.11a and 5.11b. Excluding resolution effects, the cosine of pointing angle of prompt D^0 mesons is expected to equal to unity. However, that is not true for non-prompt D^0 candidates since they originate from the beauty-hadron vertex. Still, the D^0 meson inherits a large portion of the beauty-hadron momentum when it decays, making the average pointing angle for non-prompt candidates larger than for prompt but smaller than for background candidates. As a consequence, the $\cos(\theta_p)$ distribution is more peaked towards unity for prompt D^0 candidates, while the non-prompt D^0 and background candidates have wider distributions.

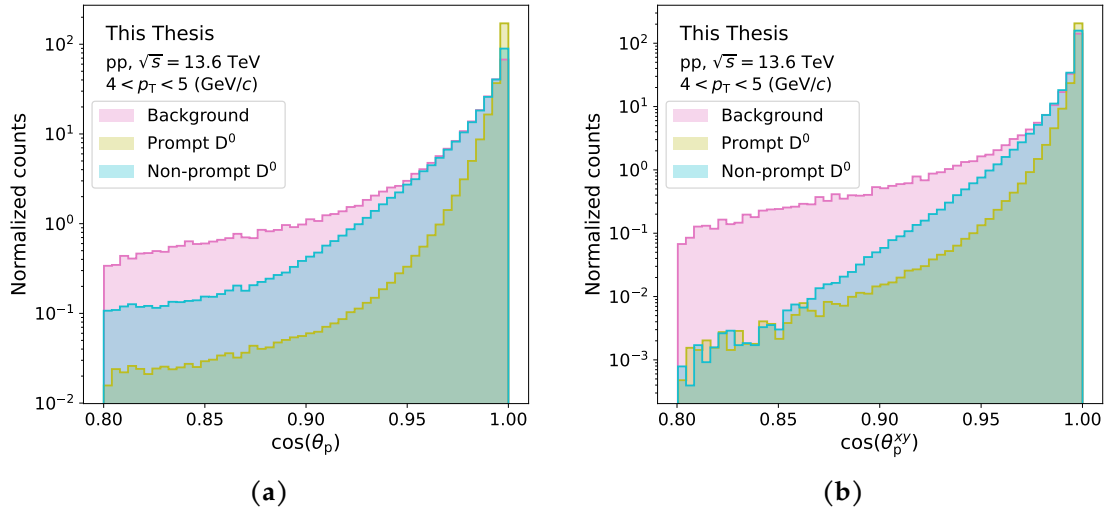


Figure 5.11: Cosine of pointing angle distributions of background, prompt D^0 and non-prompt D^0 candidates in the $4 < p_T < 5$ GeV/c range.

- **PID variables**

Adding particle identification information in the model can help distinguish between signal (prompt or non-prompt) and background by increasing confidence in the identity of the decay prongs.

In the analyses presented in this thesis, the PID information is applied in two steps. First, loose preselections are used to reduce the number of background candidates before the ML training. In particular, rectangular selections are applied to the $n\sigma$ of the TPC and TOF. These selections are computed by assigning a -1, 0, or 1 to each mass hypothesis. If $n\sigma > 3$, a response value of -1 is assigned. A value of 0 is assigned if $n\sigma < 3$ and $p_T > 1.5$ GeV/c, and +1 otherwise. If a track does not have TOF PID information, its response value is set to 0, but it is still considered compatible with all the different mass hypotheses. To identify a particle as a pion or a kaon, the sum of the TPC and TOF scores must be larger or equal to 0.

As a second step, the PID variables are used to perform the training. The chosen

PID variables combine the TPC and TOF information as follows:

$$n\sigma_{\text{comb}}^{\pi,K} = \begin{cases} |n\sigma_{\text{TPC}}^{\pi,K}| & \text{tracks only in TPC} \\ \frac{1}{\sqrt{2}} \sqrt{(n\sigma_{\text{TPC}}^{\pi,K})^2 + (n\sigma_{\text{TOF}}^{\pi,K})^2} & \text{tracks in TPC and TOF} \end{cases} \quad (5.5)$$

where $n\sigma^{\pi,K}$ is computed from the specific energy loss dE/dx (for the TPC) or the time-of-flight (for the TOF) measured for a track. This was done to maintain a low number of used variables in the training and to account for tracks that do not have hits in the TOF detector.

In the end, four PID variables are obtained, two for each track ($n\sigma_{\text{comb}}^{\pi}$ and $n\sigma_{\text{comb}}^K$ for each of the two prongs), which contain the information on the compatibility of the track with the pion and kaon hypotheses.

Evaluation of the chosen variables

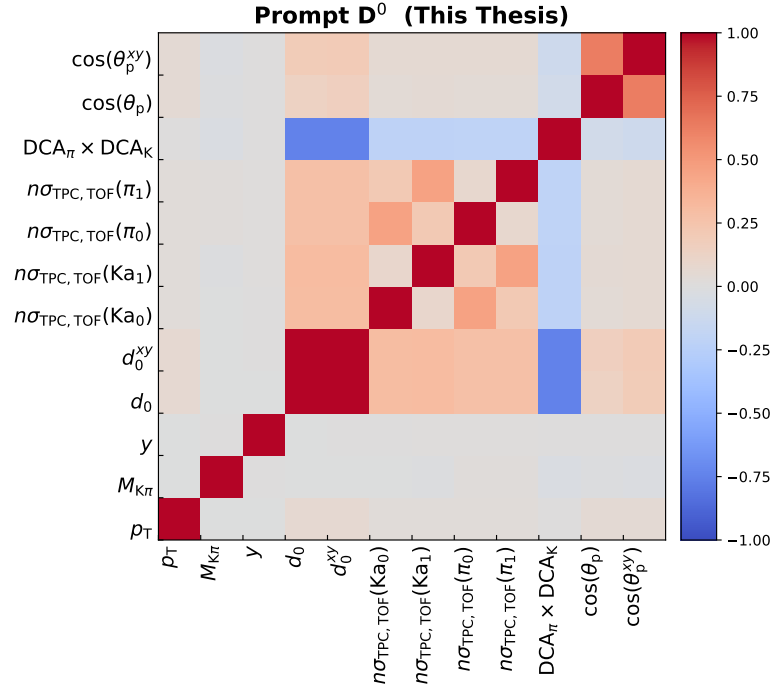
One practical method to analyse whether a variable is suitable for ML training is to examine the **correlation matrices** of the variables per class. They are created to evaluate the relationship between variables in a dataset one by one and constitute a powerful tool for identifying patterns between features. Each cell contains a correlation value that ranges between 1 (fully correlated) and -1 (fully anti-correlated). The diagonal of the matrix must be of equal unity since all variables are fully correlated with themselves. For example, the correlation matrices of the prompt D^0 and background classes in the $4 < p_T < 5$ GeV/ c range can be found in Fig. 5.12. The correlation matrices of more p_T ranges for the two analyses presented in this thesis can be consulted in their respective Analysis Notes (ANs)⁴⁵.

Correlation matrices provide a first test to check the validity of a feature for training models. As a general rule, the chosen variables must not be correlated with the variables that will be used for the primary analysis, nor should they contain redundant information. Highly redundant or correlated features might slow down the training process by increasing the model complexity without adding new information. In the example shown in Fig. 5.12, the invariant mass ($M_{K\pi}$), the transverse momentum (p_T), and the rapidity (y) were plotted in the leftmost region of each correlation matrix but were not used in the training. All variables correlated with these three were removed from the training as well. Finally, correlations should be different between classes so that the model can use them to separate between classes.

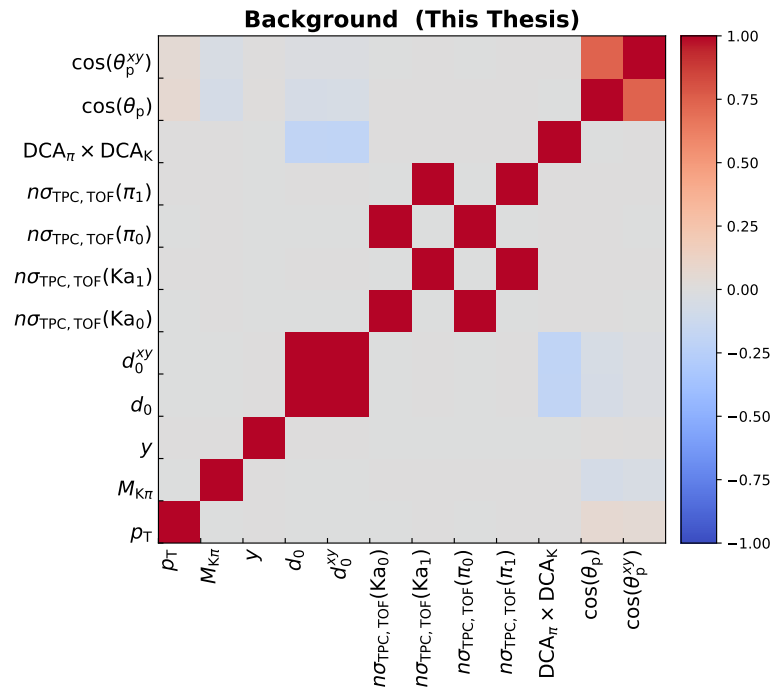
In Fig. 5.12, some variables have a high correlation level between them. For instance, the decay length and its xy projection fully correlate for prompt and background candidates. This is expected since they both refer to the same kinematic variable. At the

⁴ AN of the non-prompt D^0 fraction

⁵ AN of the double D^0 cross section



(a)



(b)

Figure 5.12: Linear correlations between the training variables, the invariant mass and the p_T , for prompt D^0 mesons (a) and combinatorial background (b). The colours, from red (fully correlated) to blue (fully anti-correlated), indicate the degree of correlation.

same time, the decay length is almost entirely anti-correlated to the product of DCAs for the prompt case but is less anti-correlated for background candidates. This is also expected, as the prongs that create the background candidates do not come from a displaced decay vertex.

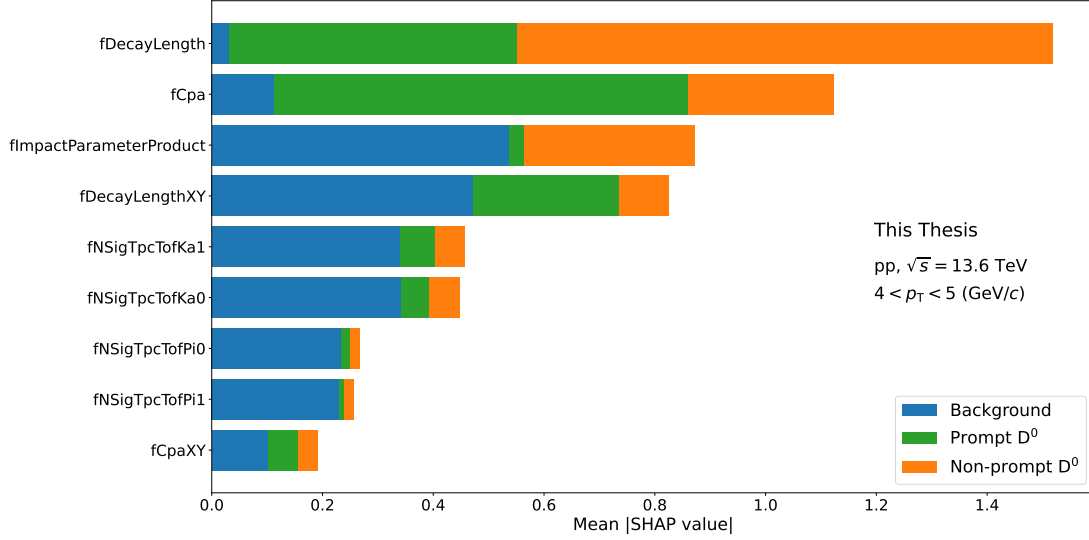


Figure 5.13: Feature importances per variable to separate between prompt D^0 (green), non-prompt D^0 (orange) and background (blue) measured in the $4 < p_T < 5$ GeV/c region.

Model inputs can also be tested with the *SHapley Additive exPlanations* SHAP [249] value of each variable per class, effectively providing their relative importance. This approach is based on the usage of **Shapley values** [250], which were developed in cooperative game theory to distribute the outcome of a game among all players fairly. In our case, SHAP values are based on the assumption that the ML output score can be expressed as a linear combination of the feature values. They are computed iteratively by individually calculating the average output score after adding the feature values. SHAP values show how each feature affects the final prediction and can be used to compare their relative importance. For example, the influence of each variable after training the three-class D^0 model in the $4 < p_T < 5$ GeV/c range can be found in Fig. 5.13. Different colours represent different classes. Features are plotted in descending order of importance, showing that the decay length is overall the most crucial variable to distinguish between the tree classes, followed by the cosine of pointing angle ($fCpa$ in the Figure). The decay length is the main feature to distinguish the non-prompt class, whereas the product of the DCAs is the most important to separate D^0 mesons and combinatorial background.

5.8.3 Training and test sets

To obtain an ML algorithm able to make predictions, it is necessary to build the data sets on which the model training is performed and its performance is evaluated. These are called **training** and **test** sets and are constructed from MC and real data samples.

These two sets must contain data with the same characteristics but be independent. For this reason, it is common to separate the data sample in two: most candidates (traditionally between 70 and 80%) are used to perform the training, and the remaining candidates are used to test the model. Then, the training and testing outputs have to be compared to each other. A well-trained model should have similar results for both sets.

To study the non-prompt D^0 fraction, a three-class model was developed to separate between prompt D^0 , non-prompt D^0 and combinatorial background. Prompt and non-prompt D^0 candidates are extracted from MC simulations. However, background candidates are taken from sidebands in the invariant mass distribution of the real data sample. These sidebands are created by taking the data points between 5 and 7 σ away from the D^0 mass peak on both sides to ensure that they are mainly composed of background candidates. The purpose behind this choice of samples is to use real data where possible to avoid biases produced by imperfections in the MC simulations, which could affect the predictive power of the model.

A similar strategy was followed to study double charm production. However, due to the specific preselections of this analysis, most of the considered candidates are prompt D^0 mesons. Hence, a binomial classification is enough to separate between signal and combinatorial background.

5.8.4 Training results and model quality evaluation

To perform the training, a set of **hyper-parameters**⁶ present in XGBoost were tuned to maximise the performance of the models. These are setting parameters that control the learning process but are not taken from data. Over the course of this thesis, all presented models have the same values for all hyper-parameters, namely:

- **n_estimators**: it accounts for the number of trees to train. In this thesis, this value is 1000.
- **learning_rate**: it is the size of the step at each boosting iteration, which is set to 0.01.
- **max_depth**: it specifies the maximum amount of levels (depth) of a tree, allowing to control the complexity of each DT. It is set to 4.
- **min_child_weight**: it accounts for the minimum number of weighted events needed per leaf node. It is set to 5.

Once the training is completed, the model is applied to the test and training sets. The distributions of the **ML output scores** for the example model can be found in Fig. 5.14. In particular, for a case where the model has to distinguish between three classes of candidates, the algorithm provides three ML output scores, each representing the probability of belonging to a specific class. The different subfigures represent the pro-

⁶ Hyper-parameters are a set of weights or values which determine the learning process of an algorithm.

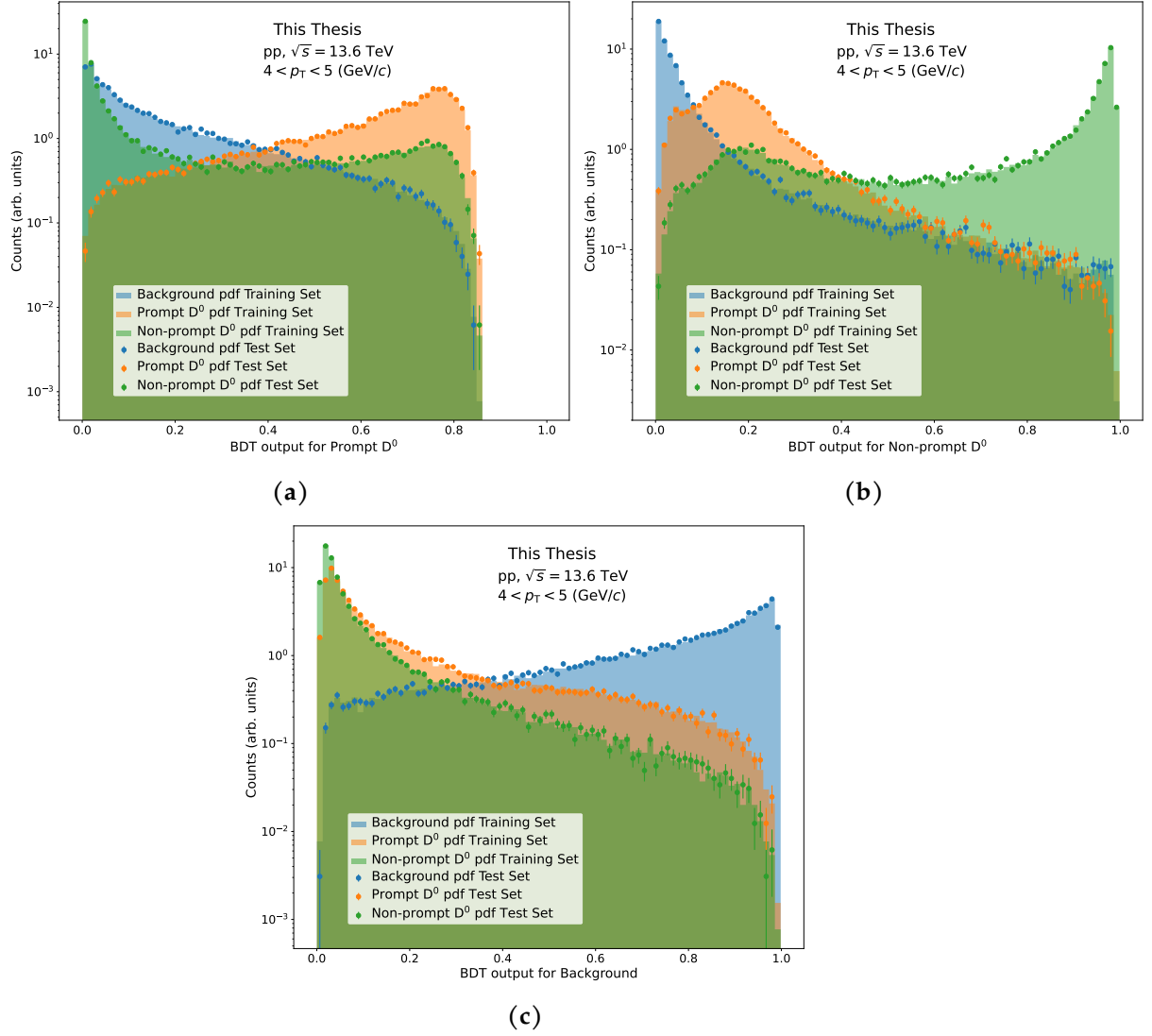


Figure 5.14: Distributions of the ML output scores for the prompt (a), non-prompt (b) and combinatorial background (c) templates measured in the $4 < p_T < 5$ GeV/c region.

jections of the different ML output scores, and the prompt, non-prompt and combinatorial background candidate distributions are depicted in orange, green and blue, respectively.

By construction, the sum of all output scores adds up to unity for each candidate. For the non-prompt and the background scores, it can be observed that the distributions of the respective true classes peak close to unity (BDT output = 1), while the other background score distribution peaks towards zero and the prompt score distribution around 0.2. This means the model successfully separated between the three classes. However, this cannot be observed in the prompt case in Fig. 5.14a. Instead, all distributions are abruptly stopped before reaching unity due to how XGBoost handles multi-class classifications. Nonetheless, the prompt distribution still peaks above all others at the highest possible prompt score values, indicating that the model can also separate this class.

The last thing that can be studied from Fig. 5.14 is the comparison between the results obtained with the training (represented with templates) and the testing (represented with points) sets. To be precise, one should check that both distributions are similar for all classes, as is the case for the example model. Significant differences might imply the existence of overfitting.

Although visualising the BDT output scores of the training and test sets gives an idea of the training quality, more quantitative tests can be performed. A standard tool to assess the quality of a specific BDT model is the **receiver-operating characteristic (ROC) curve**. It is created by plotting the True Positive Rate (TPR) over the False Positive Rate (FPR) for various threshold settings on the model output. In the cases that concern this thesis, the TPR corresponds to the signal selection efficiency, whereas the FPR is calculated using the background selection efficiency. An ideal classifier would correspond to one with a threshold with a TPR of 1.0 and an FPR of 0.0. However, real cases resemble the example shown in Fig. 5.15, where we see several ROC curves obtained from the example model. To estimate the quality of these curves, the **area under the curve (AUC)** is computed. It gives a global estimation of the model performance unrelated to the chosen threshold value. Moreover, it is independent of the relative abundances of the class instances in the dataset used for the evaluation. For a given positive and negative example, the AUC represents the probability that the model will rank the positive higher than the negative. The ROC AUC ranges from 0.5, corresponding to a random classifier, to 1.0, which would be a perfect classifier.

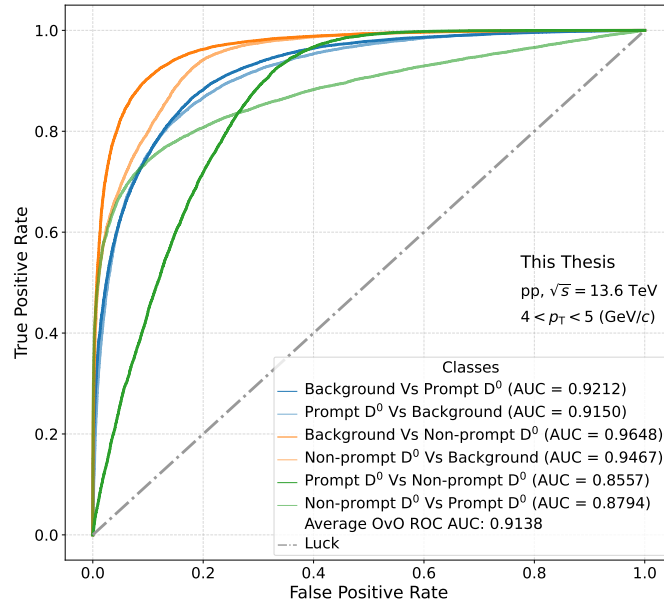


Figure 5.15: Example of a set of ROC curves for a three-class model trained using candidates in the $4 < p_T < 5$ GeV/c region. In blue, curves involving the background and non-prompt distributions, in orange, the background and prompt distributions, and in green the prompt and non-prompt distributions.

ROC curves cannot be easily defined for a multi-class classification problem, such as the one used to create Fig. 5.15. To overcome this issue, an alternative approach must

be adopted. The main ones are:

- **"One-vs-all"**: compares each class against all others, which are assumed as one.
- **"One-vs-one"**: compares all possible pairs of classes, considering one as the signal and the other as the background, and vice versa.

The quality of all models trained for this thesis was estimated using the "one-vs-one" approach. In the example shown in Fig. 5.15, this method resulted in six ROC curves, two for each pair of classes. For instance, the blue curves represent the ROC curves obtained when comparing prompt vs. background with the class labels reversed. Similarly, the orange curves compare non-prompt vs. background, and the green curves compare prompt vs. non-prompt. The curves for each pair of classes are nearly identical, except for the prompt-non-prompt case where a visible difference can be observed. However, this difference arises because these two curves are reflections of each other with respect to the diagonal; they represent the same binary classifier with reversed positive and negative class assignments.

At last, the overall quality of the model can be studied by computing the **macro average AUC**. It is defined as:

$$\text{Macro average} = \frac{1}{2n} \sum_i^{2n} \text{ROC AUC}_i, \quad (5.6)$$

where n is the number of classes considered in the model. As shown in the Figure, all AUCs have values close to unity, providing a macro average AUC of 0.91. As a result, we can confirm that the model can separate the three classes. In particular, the larger AUC was found for the separation of the non-prompt and background classes, while the lowest corresponds to the prompt-non-prompt case.

5.8.5 Machine Learning inference

After the models are trained and optimised locally, they are applied to the full datasets available on the GRID. This uses the `ONNXRuntime` package [251], allowing cross-platform model inference. Models are uploaded to the Condition and Calibration Data Base (CCDB), and their outputs are saved as variables that can be used to perform selection cuts following the same procedure as any other kinematic or PID variable.

5.9 The Offline Trigger System (OTS)

As discussed in section 3.3, the increased statistics collected during Run 3, compared to previous runs, introduced new challenges associated with processing and managing substantially larger datasets [220]. For this reason, the **Offline Trigger System (OTS)** was created. It comprises a series of analysis tasks that select potentially interesting events. These tasks are applied in O^2 Physics after full reconstruction and are used to create derived data samples that serve as preselections for subsequent analyses. An

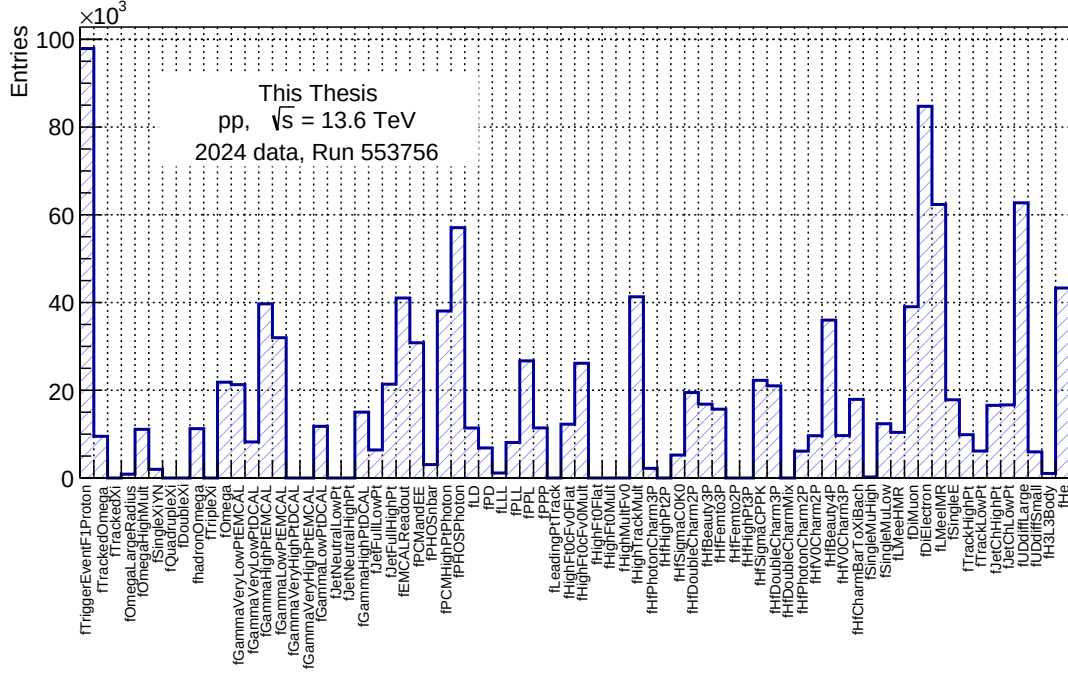


Figure 5.16: Output trigger table containing all selected events per trigger corresponding to run 553756.

example of a table which contains the summary information for all triggers in one run can be seen in Fig. 5.16. Hence, by choosing the desired **trigger masks**, it is possible to reduce the analysed data sample via a skimming process before its analysis.

In this thesis, an OTS was used to trigger double charm events, using the so-called `kDoubleCharm2P` trigger mask. This trigger was created by selecting events with at least two D^0 mesons. D^0 mesons, as well as all other charm hadron species, were reconstructed in their main hadronic channel and selected using a BDT. In this case, the chosen channel is $D^0 \rightarrow K^- \pi^+$. The corresponding BDTs⁷ were trained using the p_T and DCA distributions of the daughter tracks. The specific kinematic and PID selections can be found in Table 5.3. For the `kDoubleCharm2P` trigger, an extra requirement was made to remove non-prompt D^0 candidates. In particular, the prompt D^0 BDT score was required to be larger than 0.5. Also, global tracks were used, and a 4σ time compatibility plus 500 ns time margin was required to perform the track-to-collision association.

Table 5.3: List of trigger preselections for D^0 mesons.

p_T (GeV/c)	M (GeV/ c^2)	$\cos \theta_p$	Bkg. BDT score	PID
(1, 5)	(1.70, 2.10)	> 0.95	< 0.06	$[N_{\sigma}^{\text{TPC}}(\pi) < 3 \text{ and } N_{\sigma}^{\text{TOF}}(\pi) < 3]$ $[N_{\sigma}^{\text{TPC}}(K) < 3 \text{ and } N_{\sigma}^{\text{TOF}}(K) < 3]$
(5, 1000)	(1.70, 2.15)	> 0.98	< 0.06	$[N_{\sigma}^{\text{TPC}}(\pi) < 3 \text{ and } N_{\sigma}^{\text{TOF}}(\pi) < 3]$ $[N_{\sigma}^{\text{TPC}}(K) < 3 \text{ and } N_{\sigma}^{\text{TOF}}(K) < 3]$

⁷ See the full settings in this [CodiMD](#).

5.10 Acceptance and efficiency corrections

As we already anticipated at the beginning of this Chapter, detectors can only reconstruct particles in a certain geometrical region, called the detector **acceptance** (Acc). In addition, we need to account for the selection cuts we have performed to maximise the signal over the background. To do so, the **efficiency** (ϵ) is computed from MC simulations as:

$$\epsilon(p_T) = \frac{N^{\text{reco}}(p_T)|_{y < |y_{\text{fid}}(p_T)|}}{N^{\text{acc}}_{\text{gen}}(p_T)|_{y < |y_{\text{fid}}(p_T)|}}, \quad (5.7)$$

where N^{reco} is the number of reconstructed particles after applying selections and $N^{\text{acc}}_{\text{gen}}$, is the number of MC generated events which decay into our channel of interest. Generated events must also originate from a primary vertex with $|z| < 10$ cm, and prongs must have a $p_T > 0.3$ GeV/ c . In our analyses, these corrections need to be taken into account at **midrapidity** ($|y| < 0.5$). They also depend on p_T due to the track reconstruction efficiency and the selection to reject background.

The acceptance is computed as:

$$\text{Acc}(p_T) = \frac{N^{\text{acc}}_{\text{gen}}(p_T)|_{y < |y_{\text{fid}}(p_T)|}}{N^{\text{gen}}(p_T)}, \quad (5.8)$$

where the numerator is the number of generated D^0 mesons within $|y| < 0.5$, which decay into a kaon and a pion and satisfy that $|z| < 10$ cm.

Lastly, a **correction factor** $c_{\Delta y}$ was applied to account for the differences between Eq. 5.4 and the considered rapidity region, $|y| < 0.5$. It is defined as the fraction of the generated D-meson yield in the rapidity region $\Delta y = 2y_{\text{fid}}$ to that in $|y| < 0.5$. This factor was determined using PYTHIA and FONLL pQCD calculations [106, 107] in pp collisions and found to be $c_{\Delta y} \sim y_{\text{fid}}/0.5$.

The product of the acceptance, the efficiency and the correction factor $c_{\Delta y}$ will be used to correct the yield extractions in the analyses introduced in Chapters 6 and 7.

MEASUREMENT OF THE NON-PROMPT D^0 FRACTION AT $\sqrt{s} = 13.6$ TeV

This Chapter presents the measurement results of the non-prompt D^0 fraction, originating from beauty-hadron decays, at midrapidity in pp collisions at $\sqrt{s} = 13.6$ TeV. We start by introducing the data samples and the selections we used, which rely heavily on the methodologies introduced in the previous Chapter. Later, we describe the algorithm employed to evaluate the fraction, called the **cut-variation** method, for which we need to extract a set of raw yields and $(\text{Acc} \times \epsilon)$ factors. Lastly, the systematic uncertainties are estimated, allowing us to obtain the result discussed in the last section. Similarly to what was done during Run 2 [252], Machine Learning algorithms were employed to extract the prompt and non-prompt signals. This way, it is possible to overcome the relatively low natural fraction of non-prompt D^0 in the total production (about 5-10%).

The results presented in this Chapter were obtained by collaborating with another PhD. student from Central China University called Mingyu Zhang. Two parallel analyses were performed, in order to compare the results and methodology, until these were found to be compatible with each other. At that point, both analyses were merged, and the systematic uncertainties were calculated together, dividing the work between both analysers.

6.1 Data samples and MC used

The data sample chosen for this analysis corresponds to the minimum bias pp data collected in 2022 at $\sqrt{s} = 13.6$ TeV. A total of 5.7×10^{10} events were considered. It consists of 13 runs deemed good for physics analyses, according to the quality assurance selections. These runs are:

526641, 526964, 527041, 527057, 527109, 527240, 527850, 527871, 527895, 527899, 528292, 528461, 528531.

A set of MC simulations created using the PYTHIA 8 event generator was used to train the Machine Learning models (see Sec. 5.8.3 for a description of how ML is used) and to calculate the efficiencies (procedure explained in Sec. 5.10). The events used for the training must be different to the ones used to estimate the efficiencies to avoid introducing biases. For this reason, an ideal MC was used to train while a MC anchored to the 2022 pp data sample was employed to calculate the efficiencies. As a test, a set of models were also trained using a subset of the anchored MC, and the training results showed no significant improvements with respect to the ideal MC. For this reason, and to preserve statistics, the ideal MC was kept for the training. Both MC simulations were enriched with heavy-flavour particles following a gap-triggered approach. In other words, a heavy-flavour hadron is generated and forced to decay into its hadronic decay channel every five minimum bias events. The ideal MC contains around 100 M pp events at $\sqrt{s} = 13.6$ TeV, while the anchored MC contains approximately 400 M events.

6.2 Machine Learning based selection

As introduced in Sec. 5.8, Machine Learning algorithms, particularly BDTs, were employed to select D^0 mesons. The example plots shown in that section were extracted from the analysis presented in this Chapter. For this reason, in this Chapter, we will go through the specifics that were not introduced before, providing references to the parts already explained. Therefore, we will skip most of the explanations on the BDT trainings themselves.

6.2.1 Preselections

Table 6.1: Rectangular preselection cuts on the number of σ of the TOF, the TPC and the combination of both detectors, the cosine of pointing angle, and its projection in the xy plane.

Variable	Selection	p_T range (GeV/c)
$\cos(\theta_p)$	> 0.8	(0, 5)
	> 0.85	(5, 24)
$\cos(\theta_p^{xy})$	> 0.8	(0, 5)
	> 0.85	(5, 24)
$n\sigma_{\pi,K}^{\text{TOF}}$	< 3	(0, 5)
$n\sigma_{\pi,K}^{\text{TPC}}$	< 3	(0, 5)
$n\sigma_{\pi,K}^{\text{comb}}$	< 5	(0, 5)

Before the training, **loose rectangular preselection cuts** were made in some topological and kinematical variables to reduce the amount of background. They are applied atop the track selections introduced in Sec. 5.3.1. The full list of preselection cuts is reported in table 6.1, including the PID requirements. The reconstruction efficiencies of prompt and non-prompt D^0 meson candidates are shown in Fig. 6.1. The nature of the topological cuts causes the upward trend in the efficiencies with increasing p_T .

Furthermore, PID selections are not present at high p_T , making their selection looser. On the other hand, non-prompt candidates are expected to have a better resolution on the pointing angle than prompt candidates due to their larger displacement. Furthermore, the non-prompt D^0 candidate tends to inherit most of the momentum of the beauty hadron, making the reconstructed momentum point close to the PV. Thus, selection cuts on this variable tend to be looser for non-prompt rather than prompt candidates, explaining why its efficiencies are higher for the former than the latter.

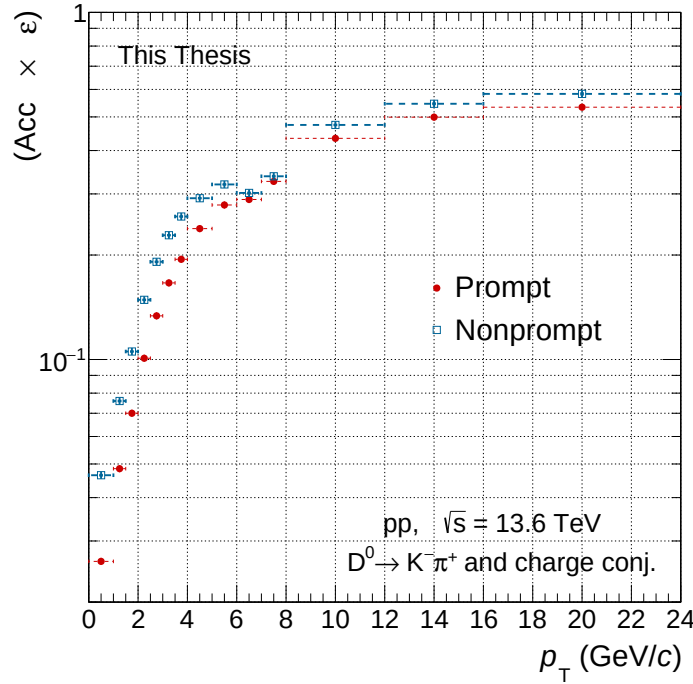


Figure 6.1: Reconstruction efficiencies as a function of p_T after the preselections for prompt (red) and non-prompt (blue) D^0 mesons in the $0 < p_T < 24$ GeV/ c range.

6.2.2 Machine Learning training

Instead of rectangular selection cuts, we used a set of ten three-class BDTs to separate between prompt D^0 , non-prompt D^0 and combinatorial background candidates. Each model was trained using candidates in a different p_T range, going from 0 to 24 GeV/ c . This way, it is possible to accommodate for the differences in the considered kinematic and topological variables as they evolve in p_T . Nevertheless, all models were trained using the same variables, which are introduced in Table 6.2. The description of each variable and the definition of $n\sigma^{\text{comb}}$ can be found in Sec. 5.8.2.

Due to the limited statistics of the MC sample, the ML models were trained in larger p_T ranges than the ones used for the analysis. In particular, the analysis was performed in the following ranges, expressed in GeV/ c :

$$(0, 1), (1, 1.5), (1.5, 2), (2, 2.5), (2.5, 3), (3, 3.5), (3.5, 4), (4, 5), (5, 6), (6, 7), (7, 8), (8, 12), (12, 16), (16, 24),$$

Table 6.2: Chosen variables to train the models to separate prompt D^0 , non-prompt D^0 and combinatorial background candidates.

Variables
Cosine of pointing angle, $\cos(\theta_p)$
xy -projection of the cosine of pointing angle, $\cos(\theta_p^{xy})$
Decay length, d_0
xy -projection of the decay length, d_0^{xy}
$DCA_0 \times DCA_1$
$n\sigma_{\pi,i}^{\text{comb}}, i = 0, 1$
$n\sigma_{K,i}^{\text{comb}}, i = 0, 1$

while the models were trained in the ranges:

$$(0, 1), (1, 2), (2, 3), (3, 4), (4, 5), (5, 6), (6, 8), (8, 12), (12, 16), (16, 24),$$

also in GeV/ c . All prompt and non-prompt D^0 candidates were extracted from an HF-enriched Monte Carlo production with an equal number of generated $b\bar{b}$ and $c\bar{c}$ pairs. The combinatorial background sample is extracted from data by collecting all candidates whose invariant mass is in the ranges $1.68 < m_{K\pi} < 1.77$ GeV/ c^2 or $1.95 < m_{K\pi} < 2.04$ GeV/ c^2 . The number of available candidates per class is reported in Tab. 6.3. However, the number of candidates per class was chosen to be equal across classes to avoid introducing biases in the training. In Tab. 6.3, the final value of considered candidates is marked in bold and corresponds to the smallest number of candidates of all three classes. Lastly, it is worth mentioning that 80% of the considered candidates were used for training, and the remaining 20% was used to test the models.

Table 6.3: Approximate number of available candidates in each class to perform the BDT training per p_T range. The final values used for the training and testing are marked in bold.

p_T (GeV/ c)	Prompt ($\times 10^4$)	Non-prompt ($\times 10^4$)	Background ($\times 10^4$)
(0, 1)	50	38	351
(1, 2)	173	125	554
(3, 4)	242	177	335
(4, 5)	215	155	122
(5, 6)	146	106	38
(6, 8)	89	67	159
(8, 12)	86	70	86
(12, 16)	12	12	9
(16, 24)	5	5	3

The evaluation of the models trained for this analysis can be seen in Sec. 5.8, which was used to illustrate the features of all BDTs included in this thesis. The study of the feature importances is also included in that section. Overall, the models can distinguish between the three classes and do not show signs of overfitting (see Fig. 5.14). The average ROC AUC values (Sec. 5.8 introduces the concept and shows a distribution using these models) per p_T range are presented in Tab. 6.4. All values are close

to unity and tend to increase with p_T . This trend is expected, as variables such as θ_P have a higher discrimination power at high p_T .

Table 6.4: Values of the average ROC AUC of all trained models.

p_T (GeV/c)	(0, 1)	(1, 2)	(2, 3)	(3, 4)	(4, 5)	(5, 6)	(6, 8)	(8, 12)	(12, 16)	(16, 24)
ROC AUC	0.880	0.900	0.918	0.929	0.935	0.940	0.941	0.943	0.945	0.951

6.3 The cut-variation method

Our objective in this analysis is to estimate the fraction of non-prompt D^0 mesons present in nature. This physical observable is defined as

$$f_{\text{non-prompt}} = \frac{N_{\text{non-prompt}}}{N_{\text{prompt}} + N_{\text{non-prompt}}}, \quad (6.1)$$

where N_{prompt} and $N_{\text{non-prompt}}$ are the number of prompt and non-prompt D^0 mesons. A data-driven procedure called the **cut-variation** method was employed to do this. This method was already used in previous Run 2 measurements in ALICE, such as the one in [115]. In this section, we will explain the details of the algorithm following the description provided in [253].

In an analysis, the raw yields of prompt and non-prompt D^0 mesons are extracted from the invariant mass distributions of the candidates. However, these raw yields are not directly related to the physical observables we want to measure, as they are affected by the acceptance and efficiency of the detector, which differ for prompt and non-prompt D^0 candidates. Therefore, it is necessary to correct them using the acceptance-times-efficiency factors, $(\text{Acc} \times \epsilon)$. This necessity motivates the use of the cut-variation method, which starts by defining a set of selections, $\{i\}$. For each selection, it is possible to define an equation which relates the extracted raw yield, Y_i , and the acceptance-times-efficiency factors of prompt and non-prompt D^0 mesons as:

$$(\text{Acc} \times \epsilon)_i^{\text{prompt}} \cdot N_{\text{prompt}} + (\text{Acc} \times \epsilon)_i^{\text{non-prompt}} \cdot N_{\text{non-prompt}} = Y_i, \quad (6.2)$$

where N_{prompt} and $N_{\text{non-prompt}}$ are the corrected yields of prompt and non-prompt D^0 mesons, respectively. Hence, it is possible to create a set of n equations using different selection cuts, leading to the following system of equations:

$$\begin{cases} (\text{Acc} \times \epsilon)_1^{\text{prompt}} \cdot N_{\text{prompt}} + (\text{Acc} \times \epsilon)_1^{\text{non-prompt}} \cdot N_{\text{non-prompt}} = Y_1 \\ \dots \\ (\text{Acc} \times \epsilon)_n^{\text{prompt}} \cdot N_{\text{prompt}} + (\text{Acc} \times \epsilon)_n^{\text{non-prompt}} \cdot N_{\text{non-prompt}} = Y_n \end{cases} \quad (6.3)$$

This system can be solved exactly if $n = 2$. Otherwise, it is likely to be overdetermined. In our case, however, since the extracted raw yields and the efficiencies are affected by

statistical and systematic uncertainties, introducing more than two equations to solve the system leads to a better estimation of the corrected yields. For this reason, we opted to create between ten and twenty sets of cuts. In this case, the problem can only be solved approximately through a minimisation procedure.

The method used in this analysis to get an approximated solution of the system in Eq. 6.3 is based on a χ^2 minimisation¹. To explain how this is done, we can start by rewriting the system of Eq. 6.3 in a more compact way using a matrix formalism for n generic sets of topological cuts. This leads us to the following formulation of the previous system of equations:

$$(\mathbf{Acc} \times \epsilon) \times \mathbf{N} - \mathbf{Y} = \mathbf{0}, \quad (6.4)$$

where $\mathbf{0}$ is the n -dimensional null vector. As mentioned, the system will likely be overdetermined for $n > 2$. An additional quantity δ is needed to account for this. The δ represents a vector of residuals, which coincides with the null vector $\mathbf{0}$ in the case of a determined system but is different otherwise. Therefore, by introducing the residual term and expanding the matrices, we reach the following:

$$\begin{pmatrix} (\mathbf{Acc} \times \epsilon)_1^{\text{prompt}} & (\mathbf{Acc} \times \epsilon)_1^{\text{non-prompt}} \\ \vdots & \vdots \\ (\mathbf{Acc} \times \epsilon)_n^{\text{prompt}} & (\mathbf{Acc} \times \epsilon)_n^{\text{non-prompt}} \end{pmatrix} \times \begin{pmatrix} N_{\text{prompt}} \\ N_{\text{non-prompt}} \end{pmatrix} - \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_n \end{pmatrix}. \quad (6.5)$$

To extract N_{prompt} and $N_{\text{non-prompt}}$ using a χ^2 minimisation, we first need to define the parameter to use. In this case, it is defined as

$$\chi^2 = \delta^T \mathbf{C}^{-1} \delta, \quad (6.6)$$

where δ^T is the row vector of residuals and $\mathbf{C}$ the covariance matrix

$$\mathbf{C} = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & & \\ \sigma_{2,1} & \sigma_2^2 & & & \\ \vdots & & \ddots & & \vdots \\ & & & \sigma_{n-1}^2 & \sigma_{n-1,n} \\ \cdots & \sigma_{n,n-1} & & \sigma_n^2 & \end{pmatrix}, \quad (6.7)$$

which is defined by the uncertainties inherent to each equation. The uncertainties σ_i^2 are calculated using the statistical errors on the raw yields and the efficiencies accord-

¹ The χ^2 is a statistical tool used to compare the observed and the expected results as:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{\sigma_i^2},$$

where O_i is the observed value, E_i is the expected value, and σ_i^2 is the variance.

ing to the following relation:

$$\sigma_i^2 = \sigma_{Y_i}^2 + N_{\text{prompt}} \cdot \sigma_{(\text{Acc} \times \epsilon)_i^{\text{prompt}}}^2 + N_{\text{non-prompt}} \cdot \sigma_{(\text{Acc} \times \epsilon)_i^{\text{non-prompt}}}^2. \quad (6.8)$$

Since the corrected yields are not yet determined at this stage, the weights must be evaluated iteratively. In the first step, N_{prompt} and $N_{\text{non-prompt}}$ are set to zero; therefore, only the uncertainty on the raw yields is considered. From the second iteration onwards, the weights are computed using the value of the corrected yields obtained in the previous step. The covariances $\sigma_{i,j} = \sigma_i^2$ are instead computed considering the sets as fully correlated, and therefore with a Pearson correlation coefficient, defined as:

$$\rho_{i,j} = \frac{\sigma_i}{\sigma_j}, \text{ with } i \subset j. \quad (6.9)$$

Only a few iterations are needed to obtain a stable result. Then, the minimisation of the χ^2 with respect to the corrected yields leads to a final value of the form:

$$N = \text{Cov}(N)(\text{Acc} \times \epsilon)^T C^{-1} Y, \quad (6.10)$$

where $\text{Cov}(N) = [(\text{Acc} \times \epsilon)^T C^{-1} (\text{Acc} \times \epsilon)]^{-1}$ is the covariance matrix which contains the statistical uncertainties on the corrected yields and their correlation. Finally, once N_{prompt} and $N_{\text{non-prompt}}$ are extracted, it is possible to evaluate a fraction of non-prompt D^0 mesons for any set of selections i as:

$$f_{\text{non-prompt}}^i = \frac{(\text{Acc} \times \epsilon)_i^{\text{non-prompt}} \cdot N_{\text{non-prompt}}}{(\text{Acc} \times \epsilon)_i^{\text{non-prompt}} \cdot N_{\text{non-prompt}} + (\text{Acc} \times \epsilon)_i^{\text{prompt}} \cdot N_{\text{prompt}}}. \quad (6.11)$$

6.4 D^0 yield extraction

Before defining the prompt and non-prompt enhanced cuts we will use to extract the non-prompt fraction, we need to maximise the D^0 signal (prompt and non-prompt) over the background. We scanned the background BDT score to search for a selection near the **maximum statistical significance**. This parameter is defined as $S/\sqrt{S+B}$, where S symbolises the number of signals and B the number of background candidates within a particular region, commonly 3σ around the invariant mass signal peak. An example of the background scan employed to find the optimal background selections in the $2 < p_T < 2.5 \text{ GeV}/c$ region can be seen in Fig. 6.2. In this case, the 3σ significance was plotted alongside the peak width, as well as the prompt and non-prompt efficiencies. The significance and width distributions tend to peak for the tightest BDT scores and show a clear decreasing trend up to 0.20, where both parameters stabilise. The peak width tends to decrease when tightening the cuts, which could be related to the fact that the remaining tracks have a higher quality, which provides a better resolution. On the other hand, the prompt and non-prompt efficiencies show an increasing

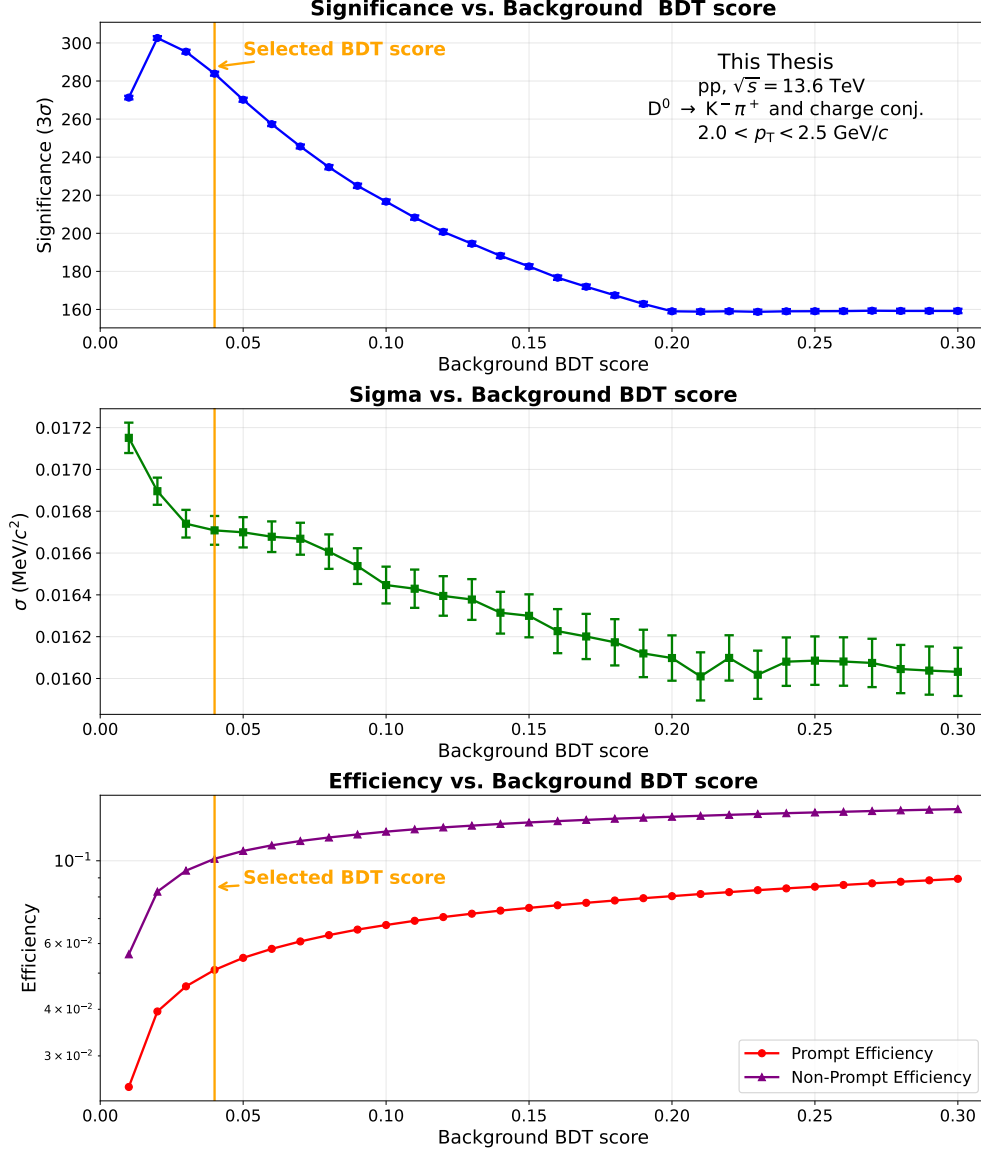


Figure 6.2: Significances in the 3σ region, invariant mass widths, and efficiencies as a function of the background BDT score for the $2 < p_T < 2.5$ GeV/c range. The selected score is marked with a vertical orange line.

trend. This is expected, since cutting on larger BDT scores implies loosening the cuts. The optimal cut was chosen where the significance was close to its maximum, but the mass width was located in a plateau, which assures better stability.

The resulting signal-enriched samples were employed as a basis to extract the non-prompt fraction. The raw yields were estimated by performing binned maximised-likelihood fits on the invariant mass distributions ($m_{K\pi}$) of the candidates. A total of fourteen p_T ranges were considered between 0 and 24 GeV/c. A Gaussian function was used to model the signal contribution in all cases. Reflections are fitted to a double Gaussian whose parameters are extracted from MC simulations. Lastly, the combinatorial background is estimated using a third-degree polynomial.

The resulting **invariant mass distributions** can be consulted in Figs. 6.3, 6.4 and 6.5 for all considered p_T ranges, alongside their respective fitting functions. The raw yields, widths, means, and statistical significance were also included. As can be observed, the significance varies widely across p_T ranges, obtaining values between 29 and 220. The widths also evolve with the transverse-momentum. In particular, we find a better resolution at low- p_T , with widths between 15 and 20 MeV/c^2 . At high- p_T , however, this resolution worsens, reaching up to 40 MeV/c^2 . This effect is expected due to the p_T resolution of the decay tracks.

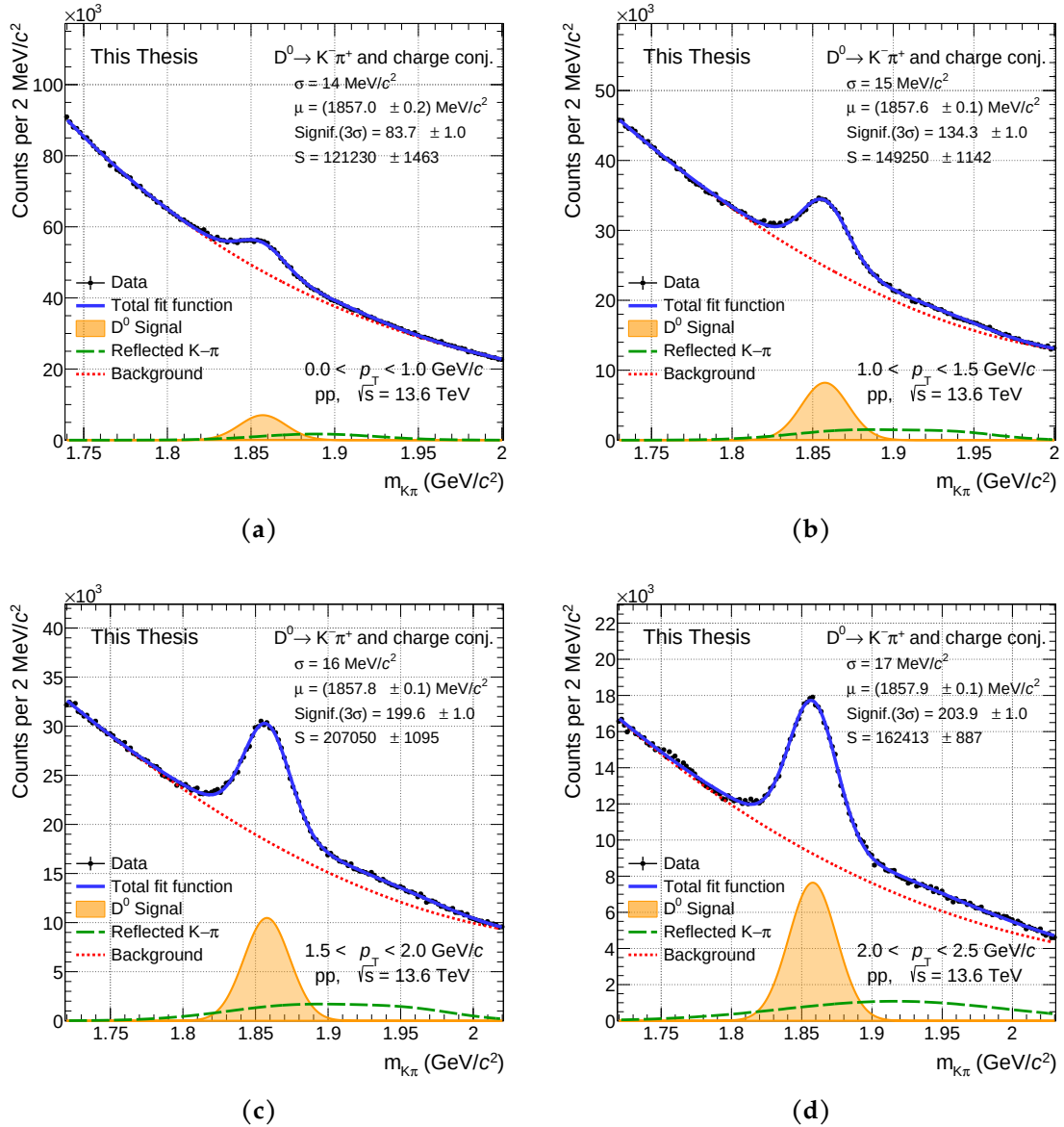


Figure 6.3: Fits to the invariant mass distributions in the $0 < p_T < 24 \text{ GeV}/c$ range (part 1). The black points represent the data, the blue line represents the total fit function, the red line is the background, the green line is used for the reflected D^0 candidates, and the yellow figure represents the D^0 signal.

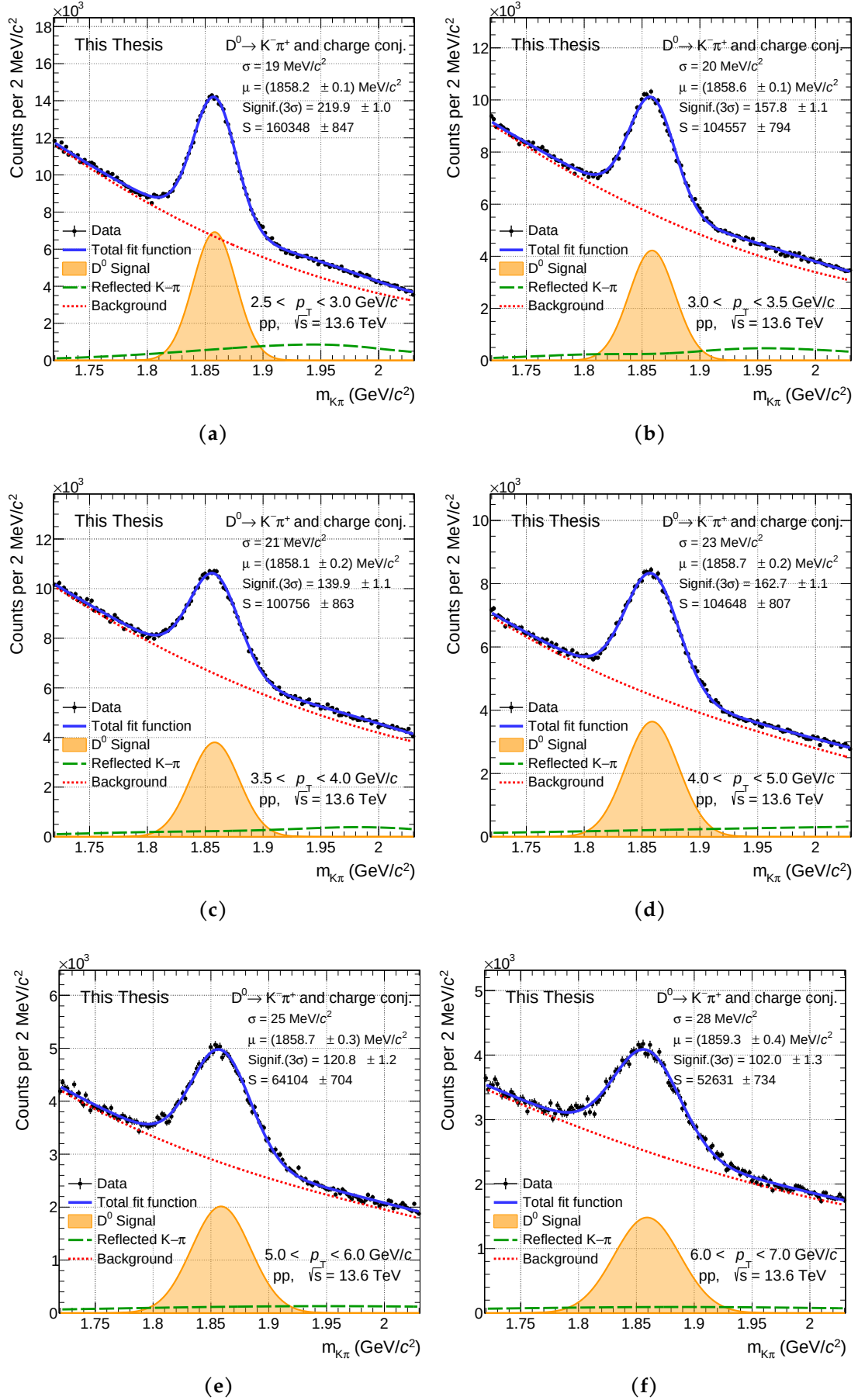


Figure 6.4: Fits to the invariant mass distributions in the $0 < p_T < 24 \text{ GeV}/c$ range (part 2). The black points represent the data, the blue line represents the total fit function, the red line is the background, the green line is used for the reflected D^0 candidates, and the yellow figure represents the D^0 signal.

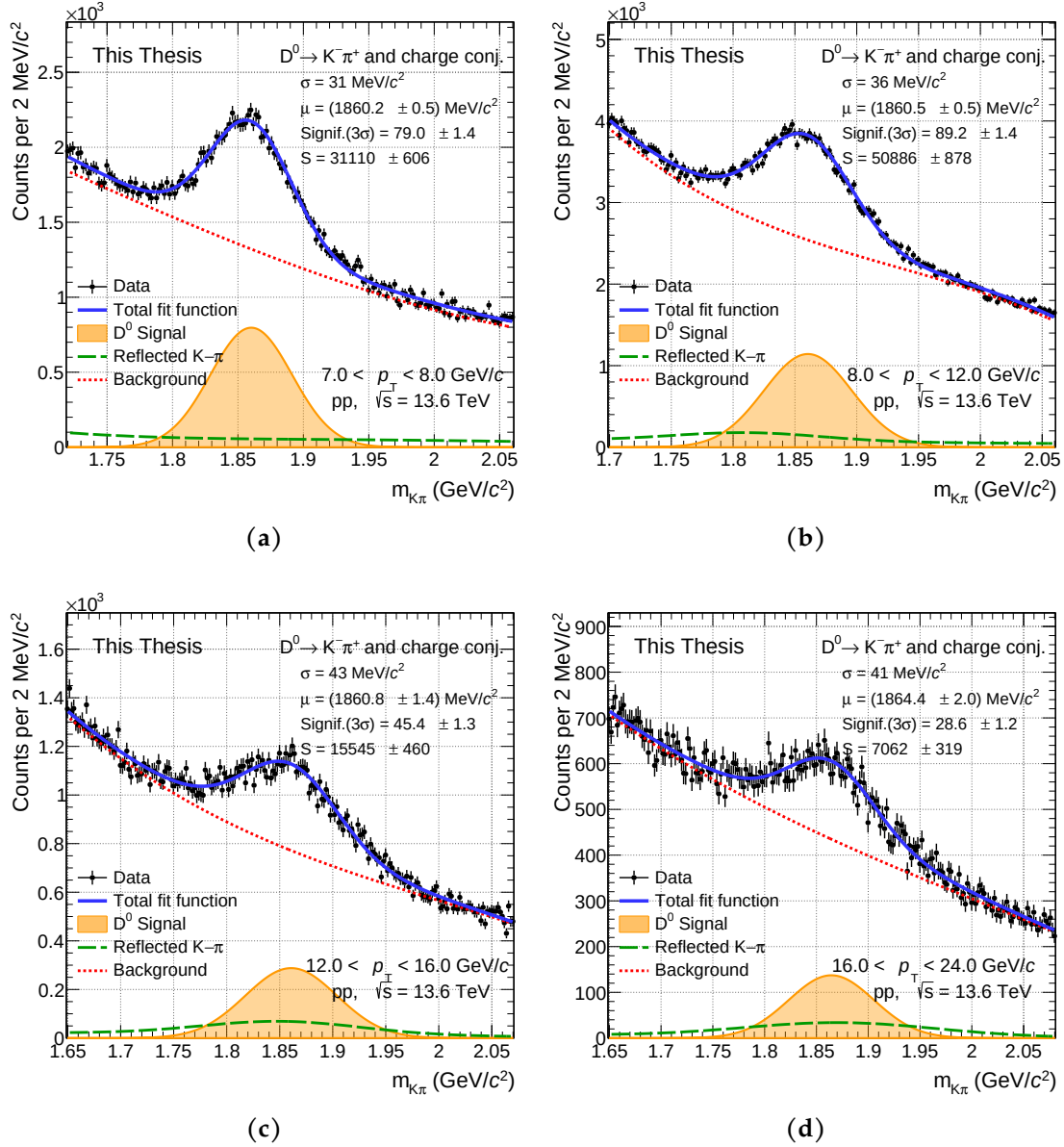


Figure 6.5: Fits to the invariant mass distributions in the $0 < p_T < 24 \text{ GeV}/c$ range (part 3). The black points represent the data, the blue line represents the total fit function, the red line is the background, the green line is used for the reflected D^0 candidates, and the yellow figure represents the D^0 signal.

After maximising the signal over the background, a set of cuts were performed on the non-prompt BDT score to obtain the raw yields needed for the cut-variation procedure. For all cuts and p_T ranges, we performed maximum-likelihood fits on the invariant mass distributions to extract these yields, using the same functions as in the signal optimisation case, *i.e.* a Gaussian for the signal, a double Gaussian for the reflected D^0 candidates and a third-degree polynomial for the background.

As we have already explained, we expect a lower quantity of non-prompt D^0 mesons than of prompt D^0 mesons in nature. Thus, tightening the selections on the non-prompt BDT score will cause the signal to diminish, making its resolution more susceptible to fluctuations. To avoid these fluctuation from having an effect on the raw

yield extraction of the non-prompt enhanced distributions, the width of the signal Gaussians per p_T range were fixed to the values of Fig. 6.3. However, this decision did not have a strong effect on the final signal yield extraction, as can be observed in Fig. 6.6. In this example, a cut was made on the non-prompt BDT score, and the results of the means and the widths are presented for all considered p_T ranges of this analysis. We found that the values obtained are compatible with both methods, and the largest difference between the two is located in the high- p_T region, where the statistics is lower.

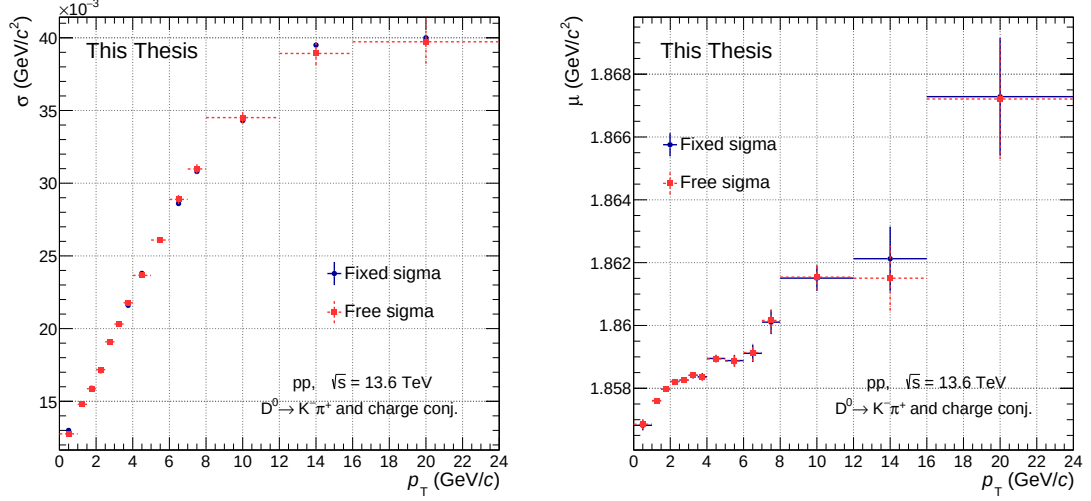


Figure 6.6: Comparison between the widths (left) and means (right) obtained fixing the width to the case without applying selection cuts on the non-prompt BDT score and leaving the parameter free for a non-prompt BDT score of 0.45.

6.5 Acceptance-times-efficiency calculation

Alongside the raw yields, we need to obtain the prompt and non-prompt **acceptance-times-efficiency** distributions for all considered p_T ranges in order to solve Eq. 6.5. These are calculated following the procedure explained in Sec. 5.10. To avoid introducing biases, we employed different MC simulations to train the models and to extract the acceptance-times-efficiency. Whereas the training was performed using an ideal MC, the efficiencies required using a MC anchored to the 2022 data sample.

The acceptance-times-efficiency distribution as a function of p_T after optimising the signal can be found in Fig. 6.7. Overall, the non-prompt efficiency is higher than the one for prompt for the same reasons as in Fig. 6.1. When introducing the cuts on the non-prompt BDT score, this situation gets amplified, and the difference between the prompt and non-prompt efficiencies grows larger. The reason behind this behaviour is that we are tightening the selections to keep the non-prompt candidates and remove the prompt ones, lowering the prompt efficiency but maintaining the non-prompt one between cuts.

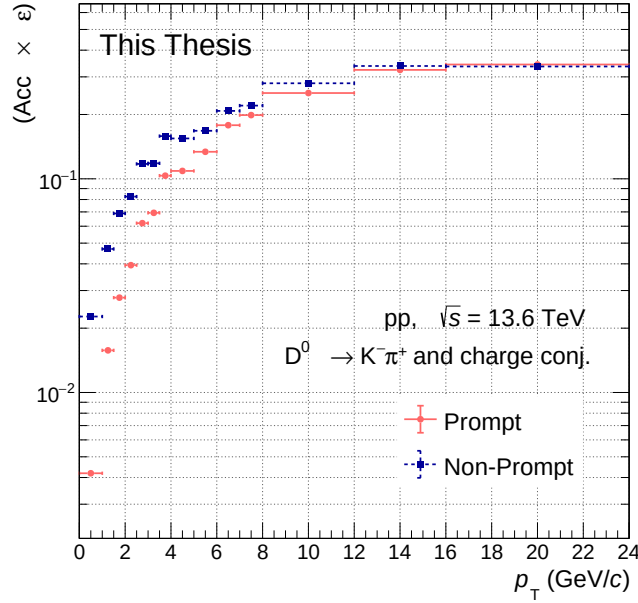


Figure 6.7: Prompt (red) and non-prompt (blue) acceptances multiplied by the efficiencies per p_T bin after selecting the working point.

6.6 Extraction of the non-prompt fraction

Seventeen selection cuts were made on the non-prompt BDT score to define the system of equations in Eq. 6.5. These were applied on top of the selections on the background BDT scores that maximised the statistical significance. With this strategy, each cut set is a subsample of the previous one. Therefore, it is possible to compute the correlation among the cut sets as reported in Eq. 6.9. In particular, the loosest selection required a non-prompt score larger than 0.2, while the tightest one required a non-prompt score of at least 0.85.

The minimisation results of the system of equations are reported in Fig. 6.8 for the $0 < p_T < 1$ GeV/c range. The raw yield distributions as a function of the cut set are presented on the top-left panel. These raw yields are fitted to the prompt and non-prompt contributions obtained through the minimisation procedure. As can be observed, the raw yield diminishes as the cuts tighten. The prompt contribution displays a similar trend to the data, while the non-prompt contribution is approximately flat. The same behaviour can be seen in the top-right plot, which contains the prompt and non-prompt efficiencies as a function of the cut set. Again, the non-prompt efficiency is almost flat, while the prompt efficiency decreases as the cut set number increases.

The bottom-left plot displays the prompt and non-prompt fraction as a function of the cut set. In the first cut sets, the proportion of prompt candidates is significantly larger than that of non-prompt candidates. However, around the tenth cut, the number of non-prompt candidates exceeds the number of prompt candidates. Around 80% of all

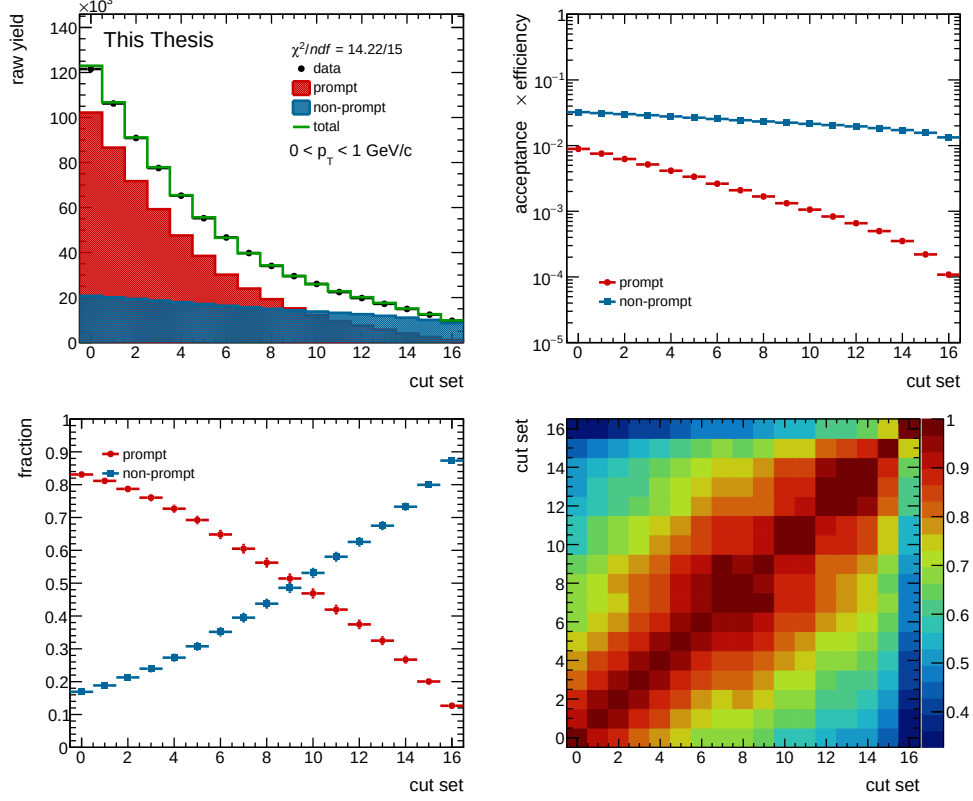


Figure 6.8: Results of applying the cut-variation procedure. Top-left: raw yields as a function of the cut set. They are fitted to extract the prompt and non-prompt contributions. Top-right and bottom-left: acceptance-times efficiency and fraction as a function of the cut set, respectively. In both plots, the prompt contribution is plotted in red and the non-prompt contribution is in blue. Bottom-right: covariance matrix.

candidates are found to be non-prompt with the tightest cut set. This large change in the proportion of prompt and non-prompt candidates helps stabilise the algorithm.

Lastly, the correlation matrix as a function of the cut set is presented in the bottom-right panel. As expected, all cut sets are fully correlated with themselves and show strong signs of correlation with the adjacent cut sets. However, the level of correlation decreases as the cut sets deviate from each other. For instance, the 14th cut set is almost entirely correlated with the 13th cut set but has a correlation level close to 0.5 with the second one.

6.7 Systematic uncertainties

We considered several sources of systematic uncertainties to the non-prompt D^0 fraction extraction. In particular, these are (i) the signal extraction, (ii) the BDT selections, (iii) the DCA resolution, and (iv) the p_T weighting. In this section, we aim to explain how these uncertainties were estimated. Other possible sources of uncertainty, such as the single track reconstruction efficiency, are the same for prompt and non-prompt candidates. Thus, they were not considered because they are expected to cancel out in

the ratio.

6.7.1 Raw yield extraction

The signal extraction constitutes a possible source of systematic uncertainties, which can be evaluated by varying the invariant mass ranges and fitting functions used to extract the raw yield. More precisely, we used second-degree polynomials to estimate the background distribution, apart from the default third-degree polynomials, rebinned the histograms to 4 and 6 MeV/ c^2 instead of the default 2 MeV/ c^2 , and extended the mass range by 0.02, 0.04 and 0.06 GeV/ c^2 . In addition, combinations of these variations were also considered. Nevertheless, trials whose χ^2/ndf was larger than three were excluded from evaluating the systematic uncertainty.

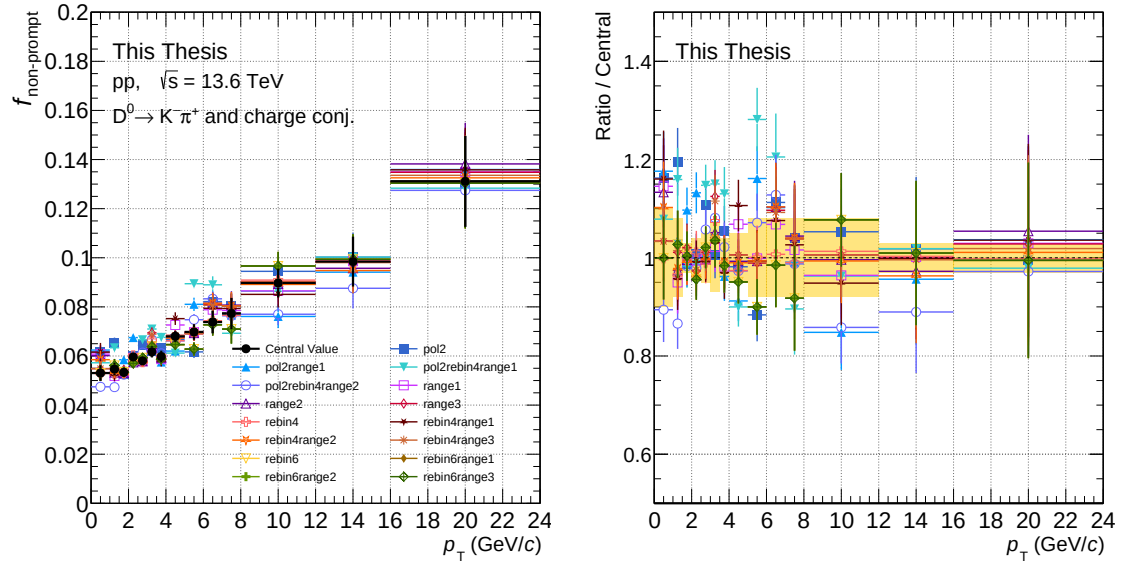


Figure 6.9: Raw yield systematic uncertainty and its ratio w.r.t. the central value.

The resulting non-prompt D^0 fractions can be found in the left panel of Fig. 6.9. On the right, the ratio of these fractions with respect to the default value as a function of p_T is presented. Overall, all trials are located within 20% of the central value for all p_T ranges. The largest deviations, specially at low- p_T are caused by the trials where the changes are the most significant, such as the light blue points, where the background fitting function, binning and invariant mass range were changed simultaneously.

The assigned systematic per p_T range was calculated as

$$\sqrt{\text{RMS}^2 + \text{shift}^2},$$

where RMS represents the root-mean square error and the shift is the average deviation of the trials from the central value. These can be found in the right panel of Fig. 6.9, and vary between 3 and 10% depending on the transverse-momentum range.

6.7.2 BDT selections

One possible source of uncertainties are the choices of selection cuts to perform the cut-variation procedure. Possible imperfections on the description of the detectors and the decay parameters can affect the efficiencies calculated in Sec. 6.5, which in turn can have an impact on the measured fraction. For this reason, the cut-variation method was recalculated using variations of the cuts employed to obtain the central value. More specifically, the procedure was repeated using both wider and finer steps, and making cuts tighter and looser than the default case. The resulting fractions are presented in the left panel of Fig. 6.10, and the ratio with respect to the default value can be observed on the right. The largest deviations are produced by the trials which include extra cuts which are looser than the ones considered in the default calculation.

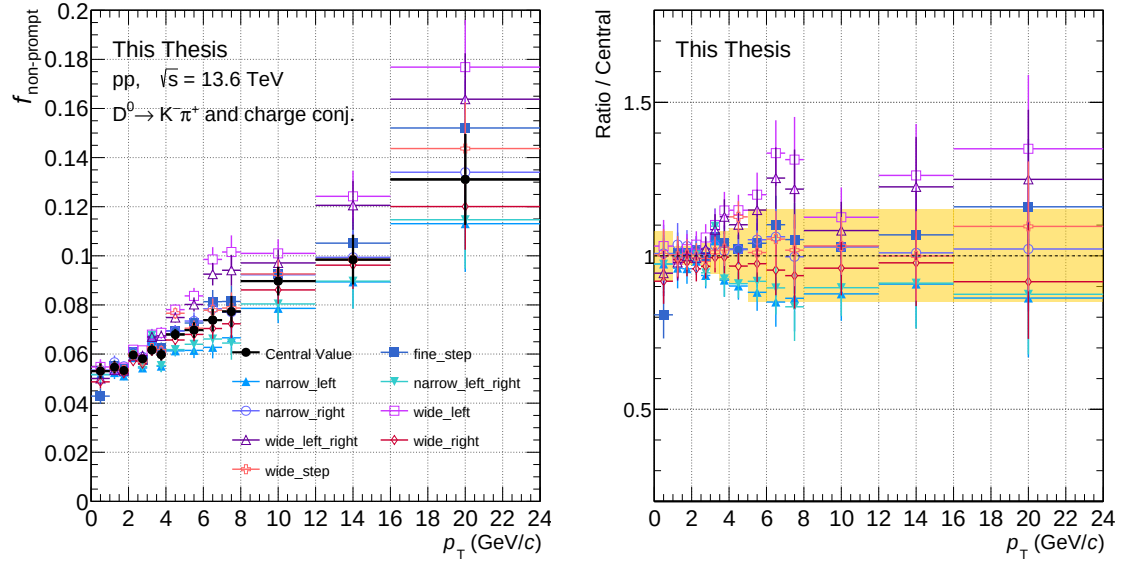


Figure 6.10: BDT selection systematic uncertainty and its ratio w.r.t. the central value.

The final systematic is computed by considering the RMS and the shift from the central value. The resulting values were also plotted in the right panel of Fig. 6.10. They range between 3 and 15% depending on the p_T interval.

6.7.3 DCA resolution

Another possible source of uncertainties is the possibility of having discrepancies between the track impact parameter distributions in MC and the data. This is related to the description of the track parameters in the vicinity of the primary vertex and the performance of separating displaced tracks and displaced decay vertex topologies. In particular, we employed the track-tuner feature (see Sec. 5.3.2) to worsen the DCA resolution in an attempt to mimic the one we see in data. However, it is necessary to study its effect, which was done by recalculating the fraction applying different track-tuner configurations: one where the q/p_T distribution is smeared by a factor of 2, and another one where the DCA resolution is a 15% worse than in the default case. The

resulting fractions and the ratio with respect to the default case as a function of p_T are presented in 6.11. As before, the systematic uncertainty was obtained by considering the RMS and the shift with respect to the central value, and is plotted in the right panel alongside the ratio. The assigned systematic varies between 4 and 9% depending on the p_T range.

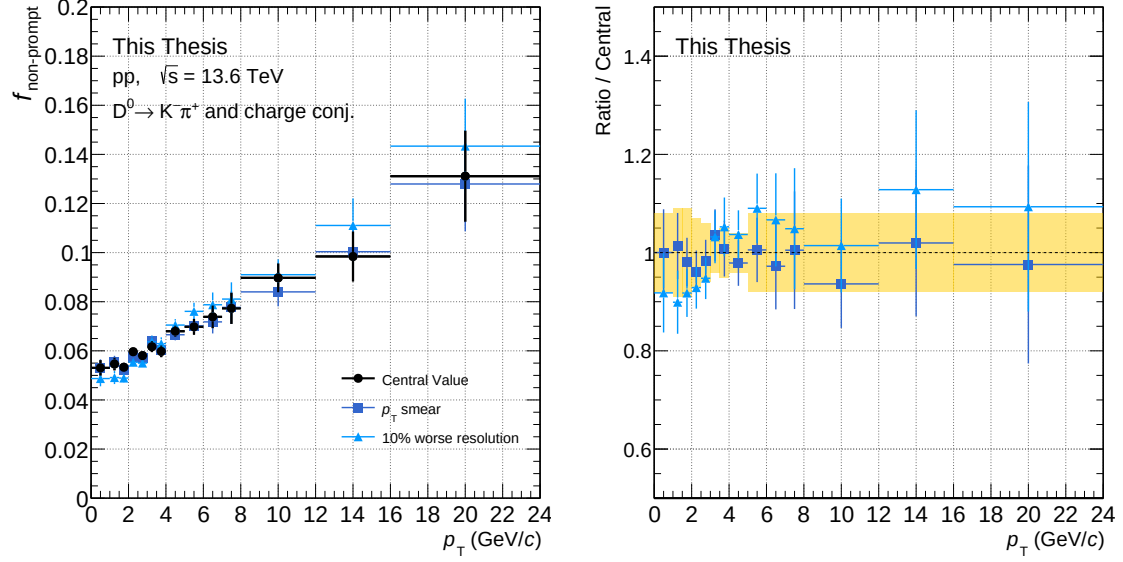


Figure 6.11: DCA resolution systematic uncertainty and its ratio w.r.t. the central value.

6.7.4 p_T shape

The last systematic uncertainty that we need to consider is the one linked to the possible difference between the real and the simulated D-meson p_T distributions. To account for this, the MC distributions obtained with PYTHIA were compared with the central value of the FONLL predictions, which are closer to the distributions found in data. Afterwards, the p_T distributions of PYTHIA were reweighted to account for this difference, and the efficiencies were recalculated. The weights were applied to the p_T distributions of prompt D mesons and to the parent beauty-hadron p_T distributions in case of non-prompt D mesons. The final uncertainty was found to range between 4% at low- p_T and 1% at high- p_T . In all ranges, it constitutes a minor source of uncertainty.

6.7.5 Summary of the considered systematics

A summary table containing all systematic uncertainties per transverse-momentum interval can be observed in Tab. 6.5. The full systematic is calculated by summing all contributions in quadrature and ranges between 9 and 19%.

The main source of uncertainty at low- p_T is the raw yield extraction, followed by the DCA resolution. However, as the p_T increases, so does the BDT selection systematic, until it reaches a stable value of 15% for $p_T > 5$ GeV/c, becoming the largest source

of uncertainty. The uncertainty on the p_T shape decreases with p_T , as is expected, although in all cases it is the smallest source of uncertainty of the ones considered.

6.8 Results

The final results of the non-prompt D^0 meson measurement at midrapidity in pp collisions at $\sqrt{s} = 13.6$ TeV can be found in Fig. 6.12. They are compared with the previous results by the ALICE collaboration measured at $\sqrt{s} = 13$ TeV, and the predictions of EPOS 4 and several PYTHIA 8 tunes. The objective of this section is to summarise our findings and comment on these results.

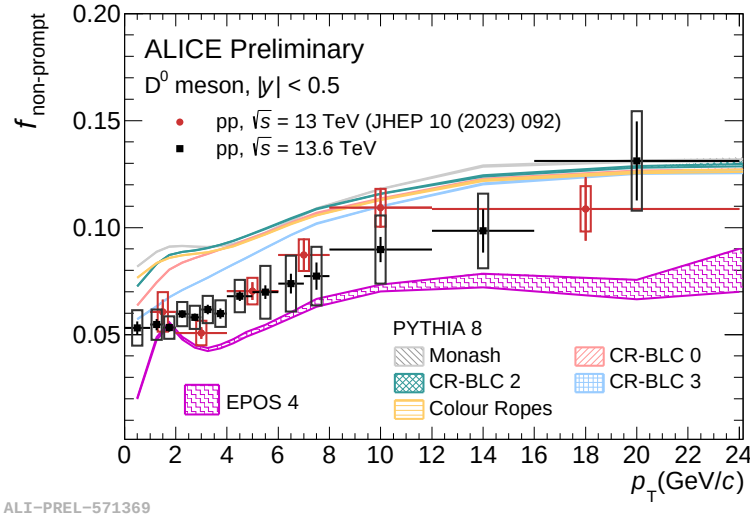


Figure 6.12: Results of the measurement of the non-prompt D^0 fraction at $\sqrt{s} = 13.6$ TeV (in black) and at $\sqrt{s} = 13$ TeV (in red) [252]. The data is compared to the theoretical predictions by PYTHIA [70] and EPOS [72].

The Run 3 measurement, in black, shows a **higher granularity** and an increased p_T coverage compared to the Run 2 result [252], which is displayed in red. These features are a consequence of the larger statistics collected with the continuous read-out mode of Run 3. This way, the measurement is extended to $p_T = 0$ and the statistical uncertainties are more precise than before. At the same time, both results are compatible with each other in all p_T ranges, indicating no significant variation arising from the higher collision energy.

Even though the statistical uncertainties are smaller in the Run 3 measurement compared to the Run 2 one, the systematic uncertainties are larger at all p_T ranges. As we indicated in the previous section, the main source of systematic uncertainty lies in the BDT selections, adding up to 15% in the highest p_T ranges. This uncertainty depends greatly on the Monte Carlo simulations used to train the models and to extract the efficiencies. Hence, we can conclude that the level of systematic uncertainty can be improved by further tuning the MCs to mimic the data, which would increase the confidence in the efficiency distributions that we use to extract the fraction.

Table 6.5: Summary of the systematics vs. p_T (transposed).

	(0,1)	(1,1.5)	(1.5,2)	(2,2.5)	(2.5,3)	(3,3.5)	(3.5,4)	(4,5)	(5,6)	(6,7)	(7,8)	(8,12)	(12,16)	(16,24)
BDT selection	8%	3%	3%	3%	4%	6%	8%	9%	15%	15%	15%	15%	15%	15%
DCA resolution	8%	9%	8%	7%	6%	4%	5%	4%	8%	8%	8%	8%	8%	8%
p_T shape	4%	4%	4%	4%	3%	2%	1%	1%	1%	1%	1%	1%	1%	1%
Raw yield	10%	8%	3%	4%	5%	7%	4%	5%	8%	8%	8%	8%	3%	3%
Total	16%	13%	10%	9%	9%	10%	10%	11%	19%	19%	19%	19%	17%	17%

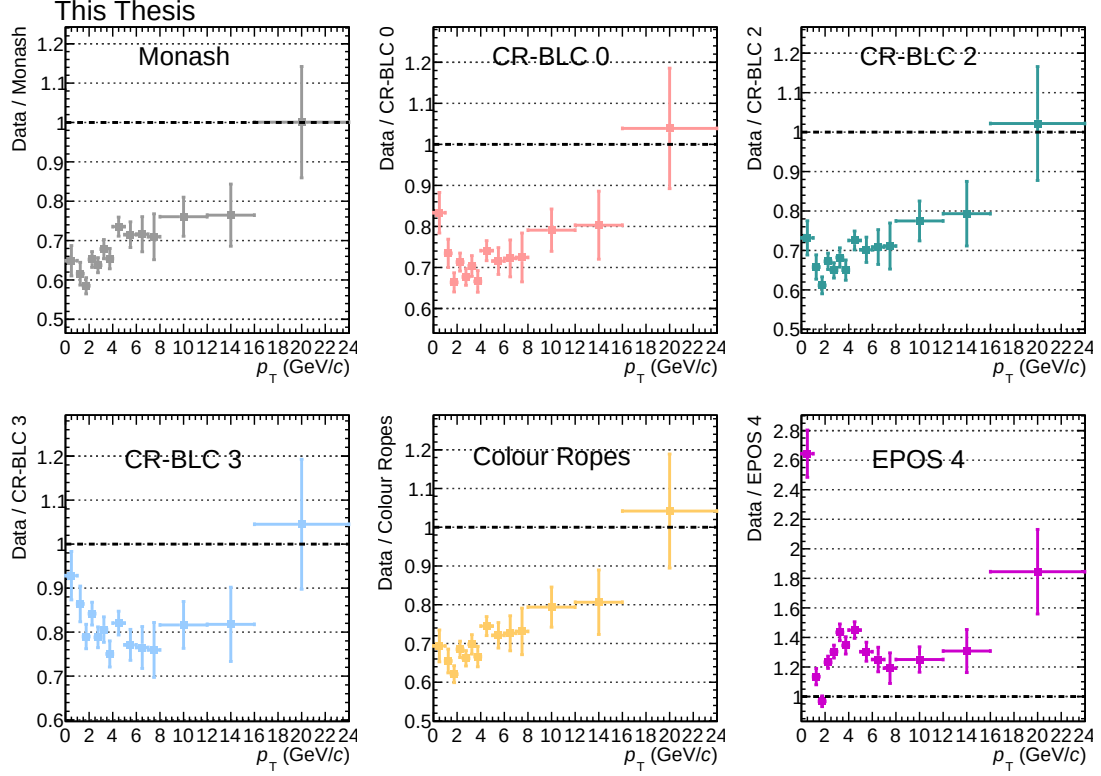


Figure 6.13: Ratio of the data over the considered PYTHIA tune and EPOS predictions as a function of p_T .

The experimental results are also compared with theoretical models. In particular, we compared our results with **EPOS 4** (Sec. 2.1.5), which relies on a hydrodynamical description of the evolution of the system even in pp collisions. This model tends to underestimate the data but remains compatible with the results within uncertainties.

In contrast, PYTHIA-based models (Sec. 2.1.4) tend to overestimate the data, as shown in Fig. 6.13. In these models, the system is not described using hydrodynamics, but rather MPIs, parton showers and hadronisation mechanisms. A total of four tunes were used. The **Monash** tune, which is the default configuration, relies on a leading-colour approximation and standard MPI-based colour reconnection. The different **colour-reconnection beyond-leading colour** (CR BLC) modes include new possible junctions between partons, especially between partons originating from different MPIs. Lastly, the **colour rope** model incorporates a rising effective string tension when multiple colour strings overlap, potentially enhancing baryon and strangeness production. Overall, all PYTHIA tunes provide similar predictions, which tend to overestimate the data.

From a theoretical standpoint, our results can be compared to **FONLL predictions**, as introduced in Sec. 2.1.6. The comparison as a function of p_T is shown in Fig. 6.14. These predictions were obtained assuming a fragmentation fraction of beauty quarks into D^0 mesons of 0.405 [254] and a D^0 branching ratio of 0.38 [125] at $\sqrt{s} = 13$ TeV. Across

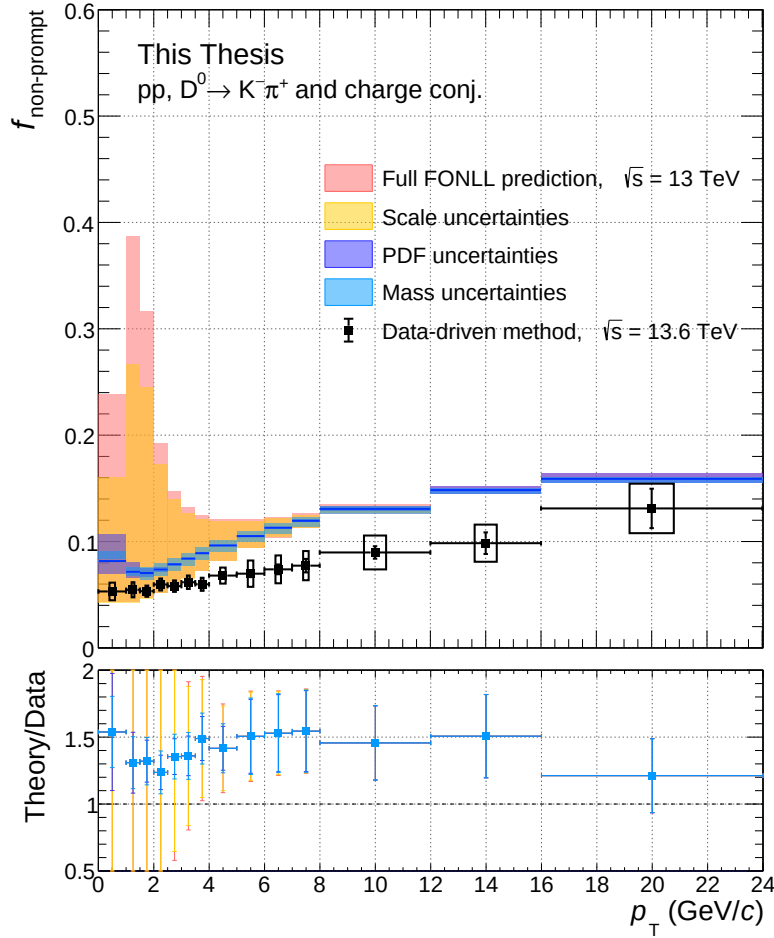


Figure 6.14: Results of the non-prompt fraction compared to FONLL [106] predictions as a function of the transverse-momentum.

all p_T ranges considered, the FONLL predictions overestimate the measured values by a factor of approximately 1.5. This behavior is expected: the FONLL D-meson cross sections generally underestimated the data, while the B^0 cross section was found to be in agreement with the data. These features, discussed in detail in Sec. 2.1.6, can be attributed to the larger mass of the b quark, which improves the reliability of pQCD calculations for beauty relative to charm. Therefore, the FONLL prediction of the non-prompt fraction, which depends on the ratio of beauty to charm cross sections, tends to be higher than the experimental data.

The different components of the uncertainties of the FONLL calculations are also presented in Fig. 6.14. As can be observed, the main sources of uncertainties come from the variation of the renormalisation and factorisation scales and is taken as the envelope of the considered variations. Other sources include the PDFs and the variation of the heavy-quark masses, which are smaller than the previous ones. The total uncertainty is calculated by summing all contributions in quadrature, and it is represented in red in the figure.

Overall, this measurement showcases the improved capabilities of ALICE in Run 3, particularly in the heavy-flavour sector, when compared to Run 2. The main current limitation is the level of systematic uncertainties, but these are expected to decrease as the understanding of the Run 3 data evolves and the Monte Carlo tuning increases. Moreover, the methodology developed here lays a solid foundation for the upcoming double D^0 measurement, to be presented in the next Chapter.

ASSOCIATED PRODUCTION OF D^0 MESONS IN PP COLLISIONS AT $\sqrt{s} = 13.6$ TeV

This Chapter presents the analysis of the production of D^0 mesons in pairs, measured in pp collisions at $\sqrt{s} = 13.6$ TeV. We begin by describing the data and MC samples employed in the analysis. These datasets were created using the Offline Trigger System (OTS), presented in Sec. 5.9. Next, we present the machine learning (ML) based selection of the D^0 candidates, which is similar to the one used in the previous Chapter, but adapted to the new trigger requirements. The raw yields of the D^0 pairs are extracted using a two-dimensional unbinned likelihood fit to the invariant mass distribution of the D^0 candidates. The efficiencies are calculated on a candidate-by-candidate basis, and the prompt fraction is estimated using the cut-variation method described in the previous Chapter. Finally, we present the results of the analysis, including the cross section and differential cross section measurements.

The results presented in this Chapter constitute an update with respect to those approved as Preliminary in March of 2025. A comparison between the two can be found in the Appendix C.

7.1 Data samples and MC used

Different datasets were used to train the BDT models and to perform the analysis in order to avoid possible biases. To train the BDTs, we used the same dataset as the previous analysis. It corresponds to the minimum bias pp data collected in the year 2022 at $\sqrt{s} = 13.6$ TeV, adding up to 1.7×10^{10} events. The full sample consists of the following runs considered suitable for physics analyses:

526641, 526964, 527041, 527057, 527109, 527240, 527850, 527871, 527895, 527899, 528292, 528461, 528531.

On the other hand, a triggered dataset was employed to perform the primary analysis. This sample contains a total of 4.4×10^{10} events extracted from the full 2024 pp sample

at $\sqrt{s} = 13.6$ TeV. It contains 465 runs, specified in the Appendix B, which correspond to the total number of good runs recorded in 2024. These events were required to pass the fHfDoubleCharm trigger, which selects events with at least two charm hadron candidates. The description of the offline trigger system (OTS) used to generate all triggered samples, as well as the common features for all triggers involving charm quarks, is introduced in Sec. 5.9.

Monte Carlo samples are required for two main tasks: to train and test the machine learning (ML) models and to calculate efficiencies. Depending on the purpose, we used two different MC samples. To train and test our models, we used a heavy-flavour enriched MC anchored to the 2022 data sample with the same calibration pass as the data used to train the models. This MC contains about 100M pp events at 13.6 TeV. It was created using PYTHIA Colour Reconnection (CR) mode 2 with a gap-triggered approach, meaning that one charm decay is forced every five minimum bias events. Particles are propagated through the detector using the GEANT4 transport packages. The run numbers are the same as the ones in the data sample.

To calculate the efficiencies, however, we used another heavy-flavour enriched MC. In this case, it is anchored to the minimum bias 2024 data sample for pp collisions measured at $\sqrt{s} = 13.6$ TeV. As in the previous case, the simulated signals are obtained using PYTHIA8 CR mode 2, with a gap trigger of 5. This MC contains 750M events, and its run numbers are the same as the ones in the 2024 data sample used to perform the analysis.

7.2 Luminosity determination

The usage of a triggered data sample makes the luminosity extraction non-trivial. A specific system titled ZORRO was employed. This system is tasked with the calculation of the number of events which were triggered per run. To extract the integrated luminosity, we need to include the information of the total number of detected events, providing the following final expression:

$$\mathcal{L}_{\text{int}}^{\text{DoubleCharm}} = \frac{N_{\text{inspected TVX}}}{\sigma_{\text{TVX}}} \cdot \frac{N_{\text{DoubleCharm}}^{\text{Analysed}}}{N_{\text{DoubleCharm}}^{\text{Total}}}, \quad (7.1)$$

where $N_{\text{DoubleCharm}}^{\text{Analysed}}$ is the number of triggered events that were analysed, $N_{\text{DoubleCharm}}^{\text{Total}}$ is the total number of triggered events in the data sample, and $N_{\text{inspected TVX}}$ is the number of TVX events in the sample. The resulting number of events is divided by $\sigma_{\text{TVX}} = 59.4$ mb, which is the TVX cross section, in order to obtain a luminosity. At last, the value is corrected for the BC existing cuts, reaching an integrated luminosity of 37.8 nb^{-1} .

7.3 Comparison of the triggered and minimum bias samples

Before proceeding with the analysis, we must evaluate the efficiency of the trigger in selecting D^0 mesons. To do this, we compared the normalised raw yields obtained with a triggered sample and a minimum bias (MB) sample, both corresponding to the same runs. To maximise the statistics, we looked at the invariant mass distributions of single D^0 candidates in both samples. Since we are only interested in studying how the trigger captures D^0 mesons, we used the `fHfSingleCharm` trigger rather than the `fHfDoubleCharm` one, because this way we can have more stats in a smaller dataset. The same selections were applied for both datasets, which correspond to the ones for D^0 mesons used to create the trigger. They can be consulted in Sec. 5.9.

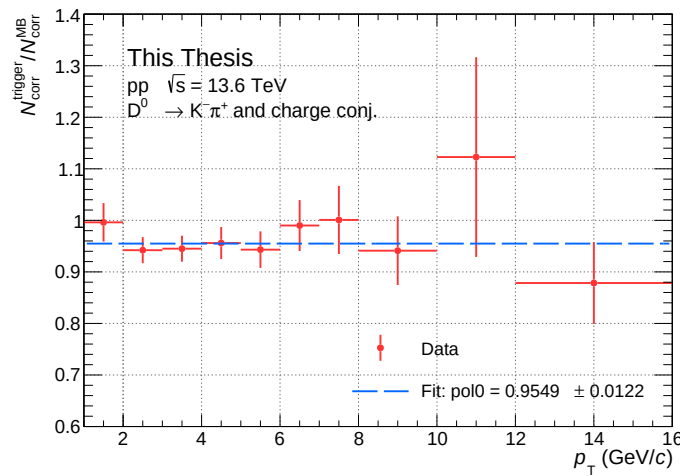


Figure 7.1: Acceptance-times-efficiency of the preselection cuts imposed by the OTS.

Dividing the normalised raw yields of the triggered and the MB sample in several p_T ranges, we obtained the plot shown in Fig. 7.1. The results were fitted to a constant, which was found to be around 0.95, meaning that the trigger is capable of selecting 95% of the D^0 mesons in the MB sample. Hence, a 5% systematic error must be accounted for in the final results.

7.4 Machine Learning based selection

The main difference between the selection of the D^0 mesons presented in this Chapter and the one in the previous Chapter is that here we employed the Offline Trigger System (OTS) to preselect D^0 pairs in the full 2024 data sample. This way, we could significantly reduce the amount of statistics to be processed. On top of the OTS preselections, we used a set of Boosted Decision Trees (BDTs) to select the D^0 mesons further. Our objective in this section is to present the preselections and the details of the BDT training and the resulting models, which will be used to select the D^0 candidates in the analysis.

7.4.1 Trigger-based preselection

The OTS uses three-class BDTs to preselect the D^0 candidates. These BDTs are trained to separate between prompt D^0 , non-prompt D^0 and combinatorial background using the DCAs and the p_T of the prongs. They are applied on top of very loose preselection cuts on the cosine of pointing angle and the PID variables, which can be consulted in Sec. 5.9. The resulting models are applied to all D^0 candidates, and selection cuts are performed in order to minimise the background and keep exclusively prompt mesons.

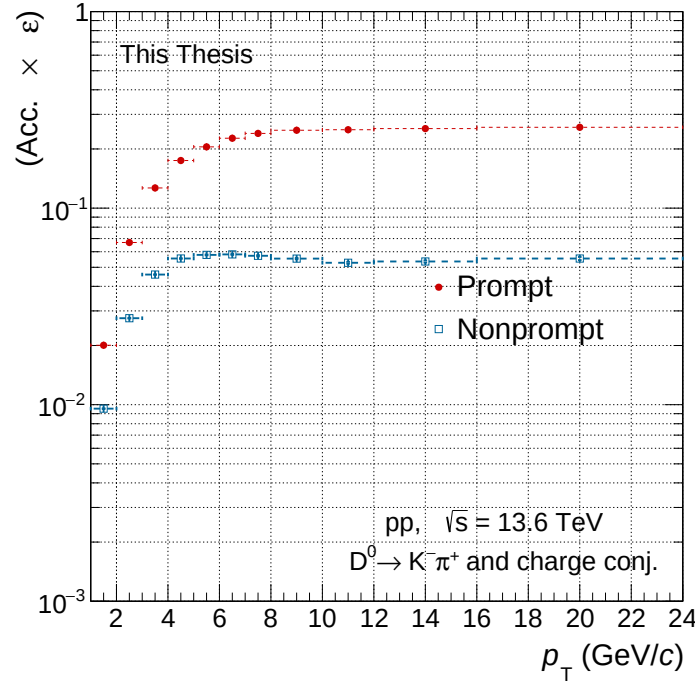


Figure 7.2: Acceptance-times-efficiency of the preselection cuts imposed by the OTS. Prompt candidates are shown in red, non-prompt candidates in blue, and all candidates in black.

To study the effectiveness of these BDTs, the acceptance-times-efficiency of the trigger selections is estimated in Fig. 7.2. As can be observed, the efficiency is higher for prompt than non-prompt D^0 candidates, which is expected because the trigger is designed to select prompt D^0 candidates. The efficiency is also higher for D^0 candidates with larger p_T as a consequence of the smaller background at high- p_T , favouring the signal detection.

7.4.2 Machine Learning training

Although the trigger preselections are designed to select prompt D^0 candidates, they are not sufficient to guarantee a good signal-to-background ratio. Hence, we trained a set of BDTs to further separate the D^0 candidates from the combinatorial background. The training was performed using the same procedure as in the previous Chapter, but adapted to the new trigger requirements and the new dataset. Specifically, we performed a binomial classification between signal D^0 candidates, composed of prompt

the residual non-prompt D^0 candidates, and background. A set of BDT models was trained in the following p_T ranges, measured in GeV/c :

$$(1, 2), (2, 3), (3, 5), (5, 7), (7, 10), (10, 24).$$

These regions were chosen to maximise the available statistics per range and to account for the ranges where the trigger BDTs were trained. The selected variables to train the models are presented in Tab. 7.1. Most variables are the same as in the previous

Table 7.1: Chosen variables to train the models to separate signal D^0 and combinatorial background candidates.

Variables
Cosine of pointing angle, θ_p
Decay length, d_0
xy -projection of the decay length, d_0^{xy}
$\text{DCA}_0 \times \text{DCA}_1$
$n\sigma_{i,j}^{\text{TPC},n}, i, j = \pi, K$
$n\sigma_{i,j}^{\text{TOF},n}, i, j = \pi, K$

chapter, except for the PID variables. First, using the separated TPC and TOF variables instead of the combined ones was found to produce a better ROC AUC. In addition, the PID distributions were reorganised as a function of whether the hypothesis was that of a pion or a kaon, rather than by the prong number. These variables are called *swapped*, as their implementation requires swapping the prongs order. This update was included to minimise the effect of an issue found in the TOF reconstruction, which reduced the efficiency of the particle identification with this detector. Even though the ML model is not trained to separate between D^0 and $\bar{D}^0$ candidates, this analysis is based on the distinction between D^0 and $\bar{D}^0$ candidates, so having the best possible PID reconstruction at all stages is imperative to guarantee its success.

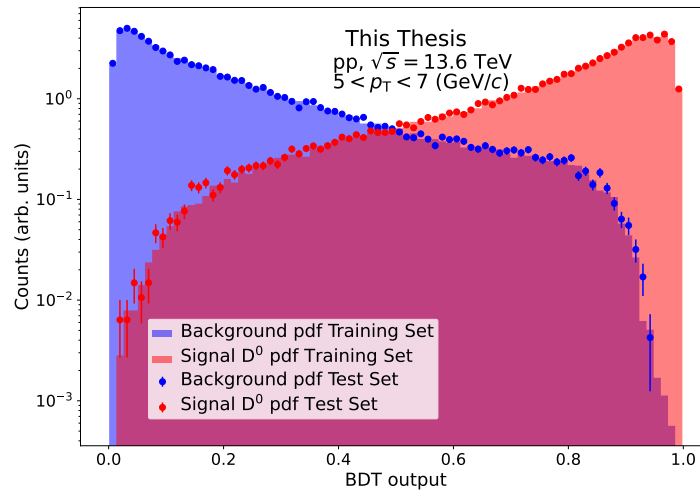


Figure 7.3: BDT output distribution for the model trained in the $5 < p_T < 7 \text{ GeV}/c$ region.

The total number of candidates available per p_T region is included in Tab. 7.2. As in

the previous analysis, we used the same number of candidates from both classes to train the models. In this case, this value corresponds to the number of signal candidates in all p_T ranges, since it is the lowest of the two. An example of the resulting models is presented in Fig. 7.3, which corresponds to the $5 < p_T < 7$ GeV/ c region. As opposed to the models presented in Sec. 5.8, only two classes are provided: signal D^0 and background. This is because of the trigger preselections, which included a cut to remove most non-prompt candidates. All candidates, prompt and non-prompt, were considered for the training, although we can expect this signal to be mainly composed of prompt D^0 candidates. A precise estimation of the prompt fraction present in the data will be presented in Sec. 7.7.

Table 7.2: Approximate number of available candidates in each class to perform the BDT training per p_T range. The final values used for the training and testing are marked in bold.

p_T (GeV/ c)	Signal ($\times 10^4$)	Background ($\times 10^4$)
(1, 2)	10	149
(2, 3)	22	258
(3, 5)	40	268
(5, 7)	18	53
(7, 10)	10	22
(10, 24)	5	7

The model performance can be analysed using the ROC AUC, as explained in Sec. 5.8.4. For instance, we present the ROC curve for the $5 < p_T < 7$ GeV/ c region in Fig. 7.4.

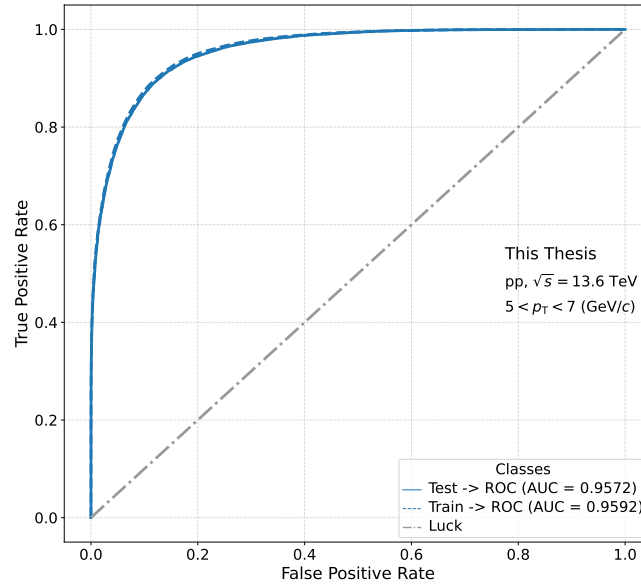


Figure 7.4: ROC distribution in the $5 < p_T < 7$ GeV/ c region for the training and testing sets. The areas under the curves (AUCs) are also shown.

As can be observed, the ROC AUC is almost equal to unity, and the training and testing sets provide almost identical ROC curves. Hence, our model is capable of distinguishing between the two classes with almost no sign of overfitting. The rest of the ROC

AUCs are included in Tab. 7.3. The training and testing sets provide very similar results in all cases, and the AUCs are very close to unity. The performance peaks in the (5, 7) GeV/ c region, where both the statistics and the p_T are most favourable. At higher p_T , the model benefits from more boosted particles, which makes learning easier. However, the available statistics are lower in that region, which makes training more challenging.

Table 7.3: Values of the average ROC AUC of all trained models.

p_T (GeV/ c)	(1, 2)	(2, 3)	(3, 5)	(5, 7)	(8, 10)	(10, 24)
ROC AUC (Train)	0.882	0.919	0.951	0.959	0.945	0.933
ROC AUC (Test)	0.877	0.918	0.950	0.957	0.943	0.932

7.4.3 Model application

The models were applied to all D^0 candidates before selecting the D^0 pairs. The reason is that, otherwise, we might risk biasing the distributions of the pair candidates. However, this means that selecting the cuts to optimise the signal is not as direct as in the non-prompt D^0 fraction analysis. The smaller statistics of the pairs compared to the single D^0 mesons might play against us, making the optimal cuts too strong to work with the pairs. Hence, we opted for a compromise, choosing looser cuts than we would otherwise. Furthermore, we excluded the D^0 candidates which were detected twice, once as a real D^0 and a second time as a reflection, for all candidates with $p_T < 5$ GeV/ c . This is because we really need to distinguish correctly between D^0 and $\bar{D}^0$ candidates to perform this analysis, so this provides a greater purity. The cut is only applied up to 5 GeV/ c because from then on, the PID is not available anymore, causing the efficiency to drop significantly if this selection is retained.

Applying these cuts, we obtained the invariant mass distributions in the $1 < p_T < 5$ GeV/ c and the $5 < p_T < 24$ GeV/ c regions 7.5. These distributions contain all D^0 candidates, not only those forming pairs. They serve as a cross-check of the cut quality, since the selections were applied to single D^0 candidates rather than exclusively to pairs. In both p_T ranges, the signal was fitted using a Gaussian function, and the background was fitted with an exponential. The reflected candidates present at $p_T > 5$ GeV/ c were fitted using a double Gaussian distribution. As expected, the D^0 peak width increases with p_T , and both the widths and peak positions are consistent with the values obtained in the previous Chapter.

7.5 Double D^0 yield extraction

Using the same cuts as in the previous section, we extracted the raw yields for the D^0 D^0 (Fig. 7.6a) and $D^0 \bar{D}^0$ (Fig. 7.6b) pairs. In this case, the yields were obtained by performing a two-dimensional unbinned likelihood fit to the invariant mass distribution

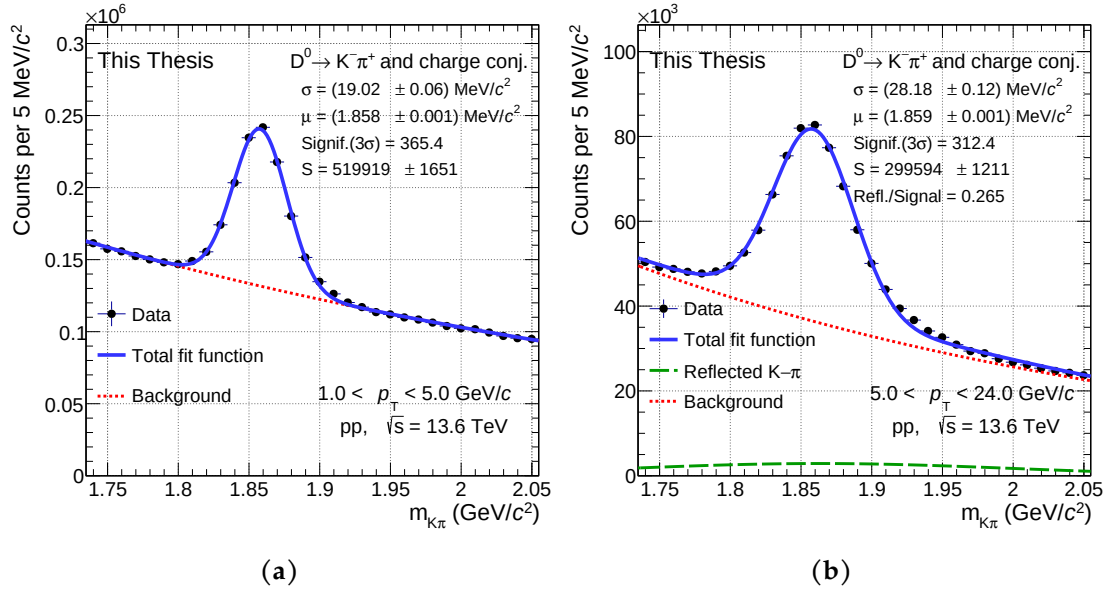


Figure 7.5: Invariant mass fit of the D^0 raw yield distribution in the (a) $1 < p_T < 5$ GeV/c (b) $5 < p_T < 24$ GeV/c regions. The total fit function is marked in blue, the background in red, and the reflected D^0 distribution in green. The reflected D^0 candidates were not included in the $5 < p_T < 24$ GeV/c range.

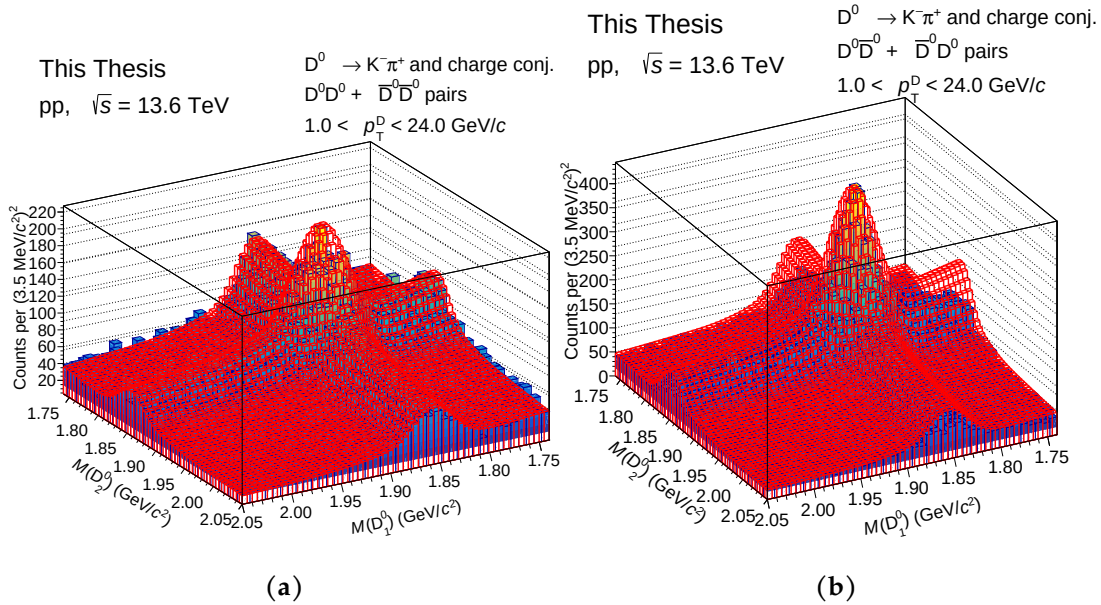


Figure 7.6: Two-dimensional plots of the invariant mass distributions of the D^0 pair candidates in the $1 < p_T < 24$ GeV/c range for (a) Like-sign pairs (b) Opposite-sign pairs. The fit to the invariant mass is represented with a red mesh.

of the D^0 pairs. To construct the two-dimensional functions, we multiplied the one-dimensional functions for each of the D^0 candidates. We used Gaussian distributions to describe the signal and an exponential function for the background. Other functions were also tested, such as a bifurcated Crystal Ball for the signal, or several modified exponential and polynomial functions for the background. They were considered for the systematic evaluation. At last, the reflected D^0 candidates were included in the fit using a double Gaussian. The reflected-over-signal fraction was extracted from the MC distributions of the D^0 candidates in the $1 < p_T < 24$ GeV/ c . The size of the reflected component for each candidate was required to follow the expression:

$$N_{\text{reflected}} = \alpha \cdot N_{\text{signal}}, \quad (7.2)$$

with

$$\alpha = \frac{N_{\text{reflected}}^{\text{MC}}}{N_{\text{signal}}^{\text{MC}}}. \quad (7.3)$$

This way, the number of fit parameters could be reduced. In this analysis, α was found to be around 0.09. This value was obtained after removing all reflected candidates in the $1 < p_T < 5$ GeV region, as explained in the previous section. It is used throughout the entire p_T -integrated analysis.

To increase the stability of the fits, we started by fitting all D^0 candidates separately, using the same procedure as in the previous section. The same background functions were chosen for the one and two-dimensional fits. The peak width and position of the signal and the background parameters were then extracted and used as initial values for the two-dimensional fit. The normalisation coefficients were left free. The fits were performed using the RooFit package [255], which is part of the ROOT [224] framework. The resulting invariant mass distributions are presented in Figs. 7.6, which show the mass correlations between the two D^0 candidates in the opposite-sign (OS) and like-sign (LS) cases, respectively. The projections of these distributions are shown in Figs. 7.7a and 7.7b. The different components that compose the total fit are presented in the projections, including the signal-signal component in green. We can observe that the signal-signal component is significantly smaller for the LS case compared to the OS. This is expected, since the combinatorics is smaller for like-sign pair candidates.

The raw yields were extracted from the fits, and the results are presented in Tab. 7.4.

Table 7.4: Raw yields of the D^0 pairs in the $1 < p_T < 24$ GeV/ c range. The first row corresponds to the like-sign pairs and the second row to the opposite-sign pairs.

Pair type	Raw yield
Like-sign pairs	1708 ± 46
Opposite-sign pairs	4003 ± 69

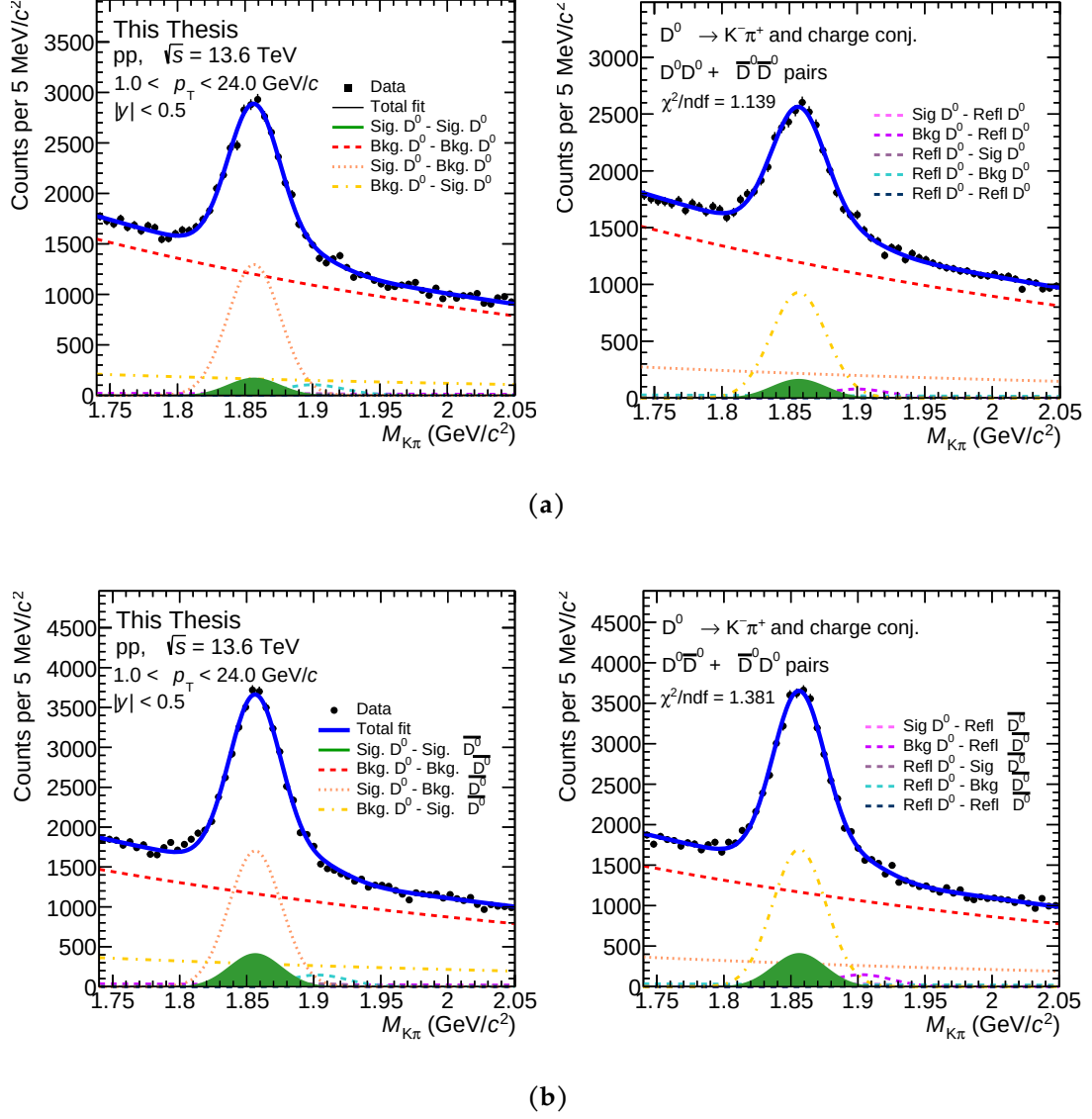


Figure 7.7: Projections of the two-dimensional plots of the invariant mass distributions of the D^0 pair candidates in the $1 < p_T < 24$ GeV/c range for (a) Like-sign pairs (b) Opposite-sign pairs. The figures on the left represent the first D^0 candidate, and the ones on the right the second.

7.6 Acceptance-times-efficiency corrections

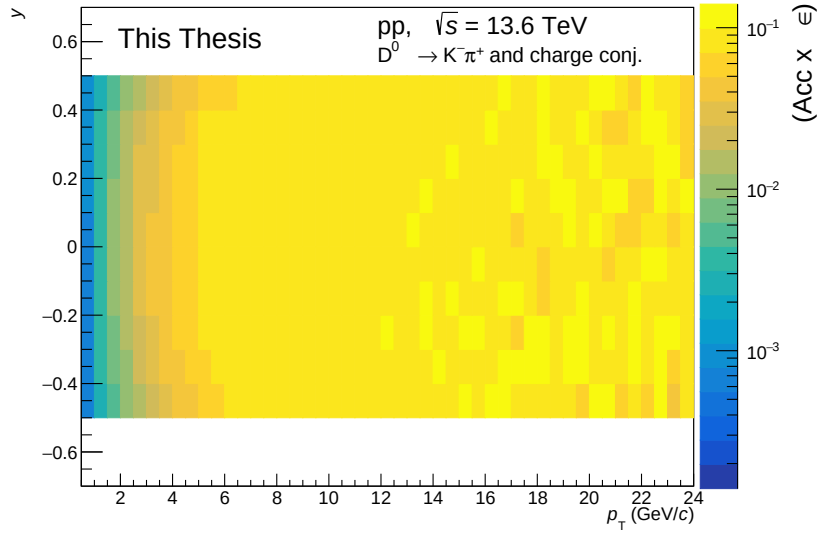


Figure 7.8: Efficiency map of the D^0 candidates as a function of the rapidity and the transverse momentum.

The efficiency corrections were calculated on a candidate-by-candidate basis. To do this, an efficiency map was created as a function of p_T and y for all D^0 candidates. This efficiency represents both the trigger and the analysis selections. Afterwards, each candidate was assigned a weight corresponding to the inverse of its efficiency. The efficiency map was created using the MC sample anchored to the 2024 data sample, as explained in Sec. 7.1. The efficiency map is presented in Fig. 7.8.

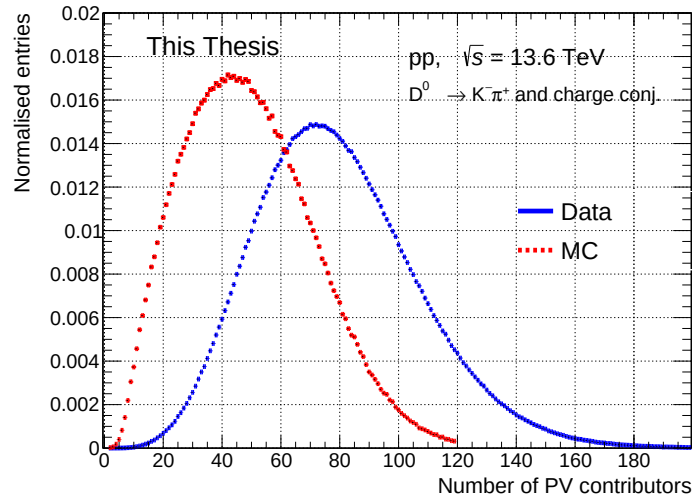


Figure 7.9: Distributions of the primary vertex (PV) contributors in data (blue) and MC (red).

It should be noted that the distributions of the number of primary vertex (PV) contributors as a function of p_T were found to differ between data and MC, as shown in Fig. 7.9. This difference was taken into account when calculating the efficiency corrections by reweighting the PV contributor distributions. Hence, the final expression

used to calculate the efficiencies was:

$$(\text{Acc} \times \epsilon)(p_T, y) = \frac{N^{\text{reco}}(p_T, y)|_{\text{PV contrib.}}}{N^{\text{gen}}(p_T, y)|_{\text{PV contrib.}}}, \quad (7.4)$$

where N^{reco} and N^{gen} are the number of reconstructed and generated candidates, respectively. Before extracting N^{reco} and N^{gen} , the MC distribution of the PV contributors was reweighted to match that of the data.

To study the acceptance-times-efficiency distributions, we also present the $(\text{Acc} \times \epsilon)$ of all candidates that enter the analysis in the $|y| < 0.5$ region as a function of p_T in Fig. 7.10. The prompt and non-prompt efficiencies are shown. The efficiencies are also higher at larger p_T , which is expected because there is less background at high- p_T , which favours signal detection. In addition, the Lorentz boost at high p_T causes the D^0 decay vertices more displaced from the PV, which also helps to distinguish from the background.

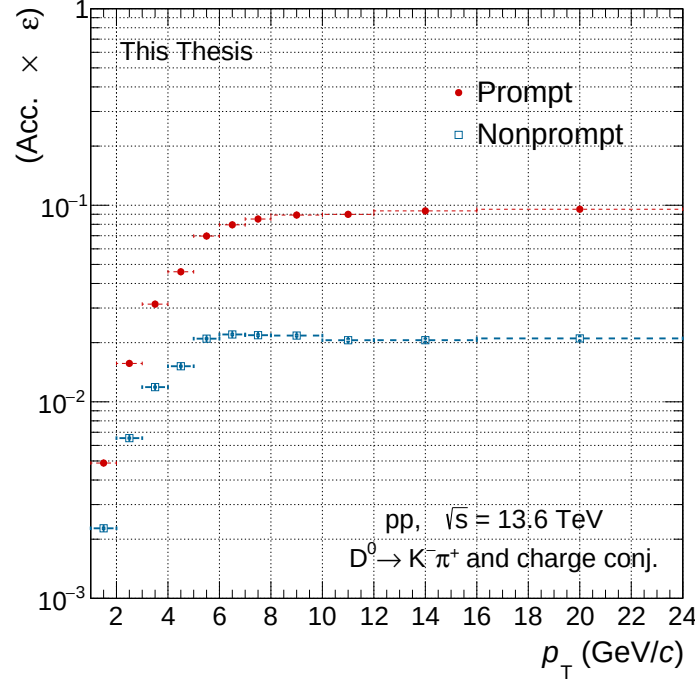


Figure 7.10: Acceptance-times-efficiency distributions of the prompt (red), non-prompt (blue) after applying the BDT selections on top of the OTS.

The final corrected yields were obtained by fitting the efficiency-corrected invariant mass distributions. These are presented in Fig. 7.11 for like-sign (upper plot) and opposite-sign pairs (bottom plot). They were fitted using the same functions and procedures as in the previous section. Thus, a Gaussian was used to model the signal, a double Gaussian was used for the reflections, and an exponential was used for the background. The normalisation coefficients were left free, while other fitting parameters, such as the signal width and position or the background parameters, were extracted from the raw yield fits and used as initial parameters.

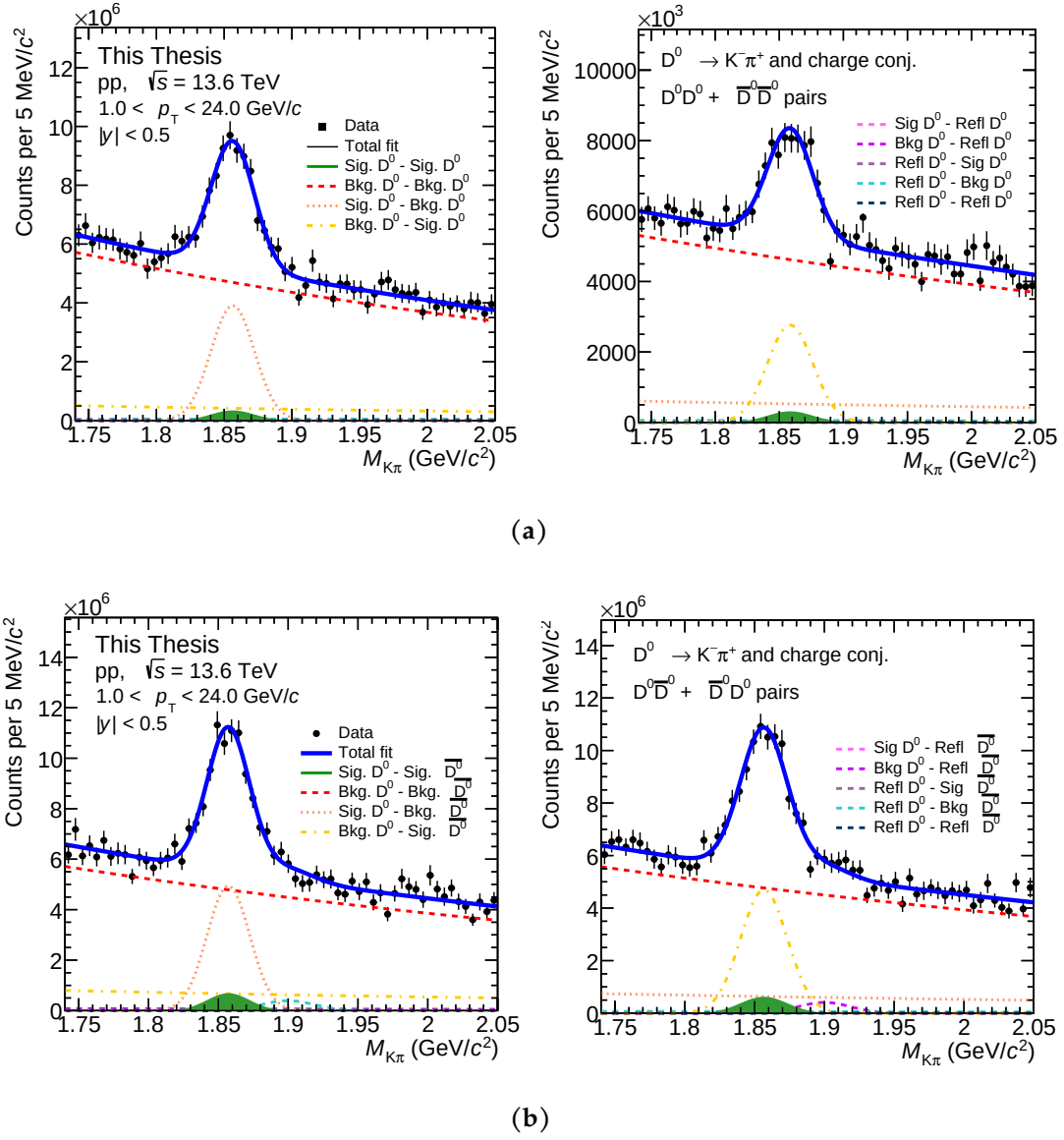


Figure 7.11: Weighted projections of the two-dimensional plots of the invariant mass distributions of the D^0 pair candidates in the $1 < p_T < 24$ GeV/c range for (a) Like-sign pairs (b) Opposite-sign pairs. The figures on the left represent the first D^0 candidate, and the ones on the right the second.

The corrected yields computed by applying the efficiencies candidate-by-candidate are presented in the first column of Tab. 7.5. The second column shows the yields computed with the integrated efficiency, calculated as the average of the efficiencies in the $1 < p_T < 24$ GeV/ c and $|y| < 0.5$ region. We should expect that:

$$Y^{\text{corr}} \sim \frac{Y^{\text{raw}}}{(\text{Acc} \times \epsilon)_{\text{int}}^2}, \quad (7.5)$$

where Y^{corr} represents the yield corrected candidate-by-candidate, Y^{raw} is the raw yield extracted from the invariant mass fit, and $(\text{Acc} \times \epsilon)_{\text{int}}$ is the integrated efficiency. If the efficiencies are correctly computed, the two corrected yields should be similar, although small differences are expected due to the mathematical differences between the two procedures: the candidate-by-candidate procedure starts by performing a division of the candidate over the efficiency, and then sums these values, while the integrated efficiency procedure begins by summing all candidates and then divides by the efficiency. Furthermore, it is also possible to have small discrepancies due to variations in the simulated y or p_T shape and data. Comparing the results of Tab. 7.5, we can observe that the yields obtained with both methods are consistent within uncertainties, which serves as a cross-check for the candidate-by-candidate efficiency-correction procedure.

Table 7.5: Check on the signal-signal yield computed with the efficiencies calculated candidate-by-candidate vs. using the integrated efficiency values.

	Candidate-by-candidate ($\text{Acc} \times \epsilon$)	Integrated ($\text{Acc} \times \epsilon$)
Opposite-sign yield	$(5.55 \pm 0.35) \cdot 10^6$	$(5.30 \pm 0.09) \cdot 10^6$
Like-sign yield	$(2.59 \pm 0.25) \cdot 10^6$	$(2.30 \pm 0.06) \cdot 10^6$

Another cross-check performed to increase the confidence in the application of the acceptance-times-efficiencies to the data is the comparison of the raw and corrected invariant mass distributions. This was done by computing the ratio of the normalised invariant mass distributions before and after correcting for efficiencies. This ratio is reported in Fig. 7.12 as a function of the invariant mass of each of the two D^0 candidates. The peak region of the efficiency-corrected distribution is marked with dashed red lines. As can be observed, the ratio is compatible with unity in most of the mass space for both LS and OS pairs. The raw and the efficiency-corrected invariant mass distributions differ the most in the high-mass region. This behaviour is expected, since this region is characterised by a smaller statistics primarily composed of background candidates. Hence, we can conclude that this cross-check was successful.

As a follow-up of the previous check, we present the peak widths and means of the raw and efficiency-corrected yield fits. These are presented in Tab. 7.6. The mean positions are found to be compatible, while the widths differ by less than 2σ .

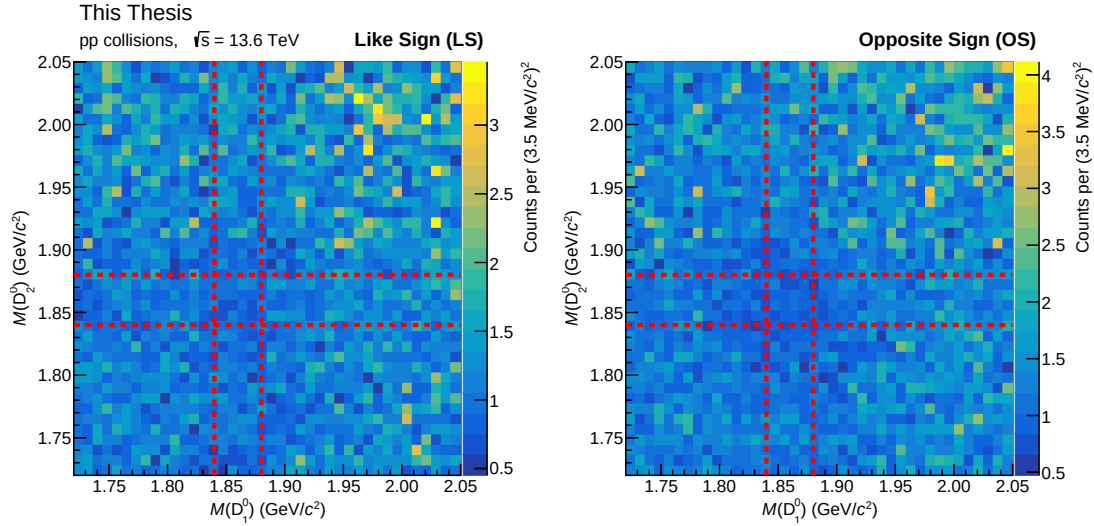


Figure 7.12: Ratio of the normalised invariant mass distributions before and after correcting by efficiencies for like-sign pairs (left) and opposite-sign pairs (right).

Table 7.6: Comparison of the widths and means for the efficiency-corrected and raw yields.

Pair type	Efficiency-corrected yield		Raw yield	
	Width (MeV/c ²)	Mean (MeV/c ²)	Width (MeV/c ²)	Mean (MeV/c ²)
Like-sign	17.65 ± 1.22	1858.40 ± 0.91	19.50 ± 0.38	1858.00 ± 0.36
Opposite-sign	17.55 ± 0.78	1857.90 ± 0.81	19.85 ± 0.28	1857.70 ± 0.26

7.7 Prompt fraction determination

To estimate the prompt fraction, we started by cross-checking whether the 2024 data sample provided the same non-prompt fraction as the 2022 sample, which was presented in the previous Chapter. As we mentioned in previous sections, it was not possible to use the BDT models developed for the D^0 pair analysis, since the trigger is heavily biased towards prompt D^0 candidates. Thus, we cannot extract enough non-prompt candidates to perform a training with three classes. Instead, we used an updated version of the model used in the previous Chapter¹ and applied it to the 2024 sample without employing any triggers. The preselections are the same as in the previous Chapter, and the models used the same variables as the ones used in the previous Chapter, except for the PID variables, which were updated to use the *swapped* variables described in Sec. 7.4.2. The models were trained using the 2022 data sample, and then applied to the 2024 data sample in the following p_T ranges:

$$(1, 2), (2, 3), (3, 4), (4, 5), (5, 6), (6, 8), (8, 12), (12, 16), (16, 24).$$

The results of applying the cut-variation method are presented in Fig. 7.13 for the $1 < p_T < 2$ GeV/c range. As in the previous Chapter, the top-left plot shows the yields of the D^0 candidates as a function of the cut on the non-prompt BDT output, as well as

¹ The details of these models are located in this [Analysis Note](#).

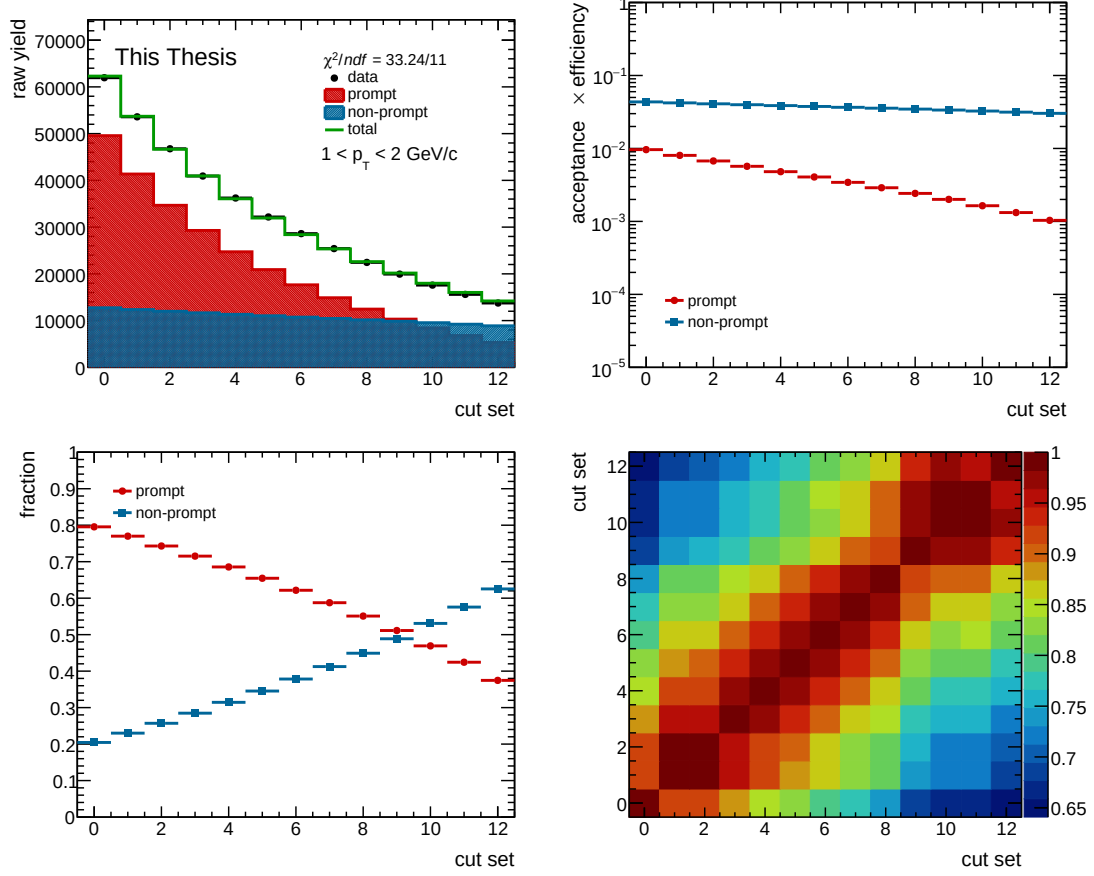


Figure 7.13: Results of the cut-variation method in the $1 < p_T < 2$ GeV/c range.

the resulting N_{prompt} and $N_{\text{non-prompt}}$ values. The top-right plot contains the $(\text{Acc} \times \epsilon)$ values as a function of the cut set, and the bottom-left plot shows the resulting prompt and non-prompt fractions as a function of the cut set. Lastly, the bottom-right plot shows the correlation matrix of the cuts used in the cut-variation algorithm.

As in the previous Chapter, we can extract the non-prompt fraction present in nature by obtaining the N_{prompt} and $N_{\text{non-prompt}}$ values using a set of yields and acceptance-times-efficiencies. Doing so should provide the same results as the ones obtained in the previous Chapter, since the 2022 and 2024 data samples refer to the same collision energy. The results of the non-prompt fractions obtained with both datasets are compared in Fig. 7.14. As can be observed, the non-prompt fractions obtained with both datasets are consistent within uncertainties. The results are also compared with the Run 2 results measured at $\sqrt{s} = 13$ TeV, which were presented in the previous Chapter. The results are also consistent with the Run 2 results. Thus, we can conclude that the 2024 data sample provides a compatible non-prompt fraction as the 2022 sample, which serves to validate the cut-variation procedure.

However, in this case, we are interested in calculating the prompt fraction of the D^0 candidates which pass the selections introduced in previous sections. To do this, we can use the results of the cut-variation method corrected by the efficiencies of the se-

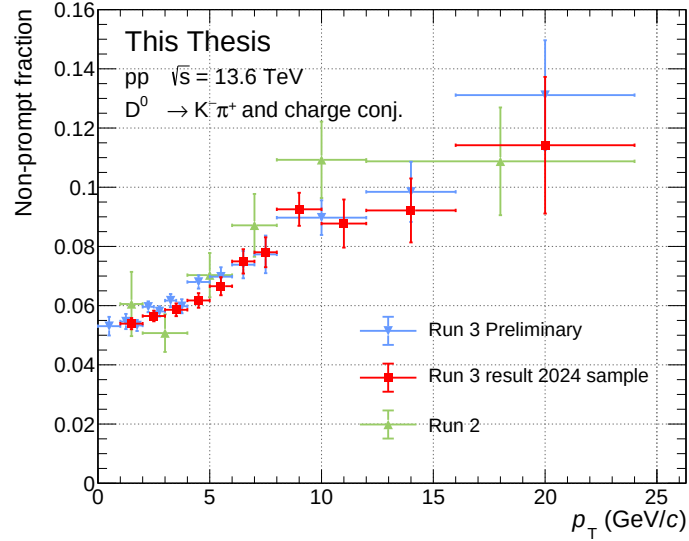


Figure 7.14: Comparison of the non-prompt D^0 fractions obtained using the 2022 and 2024 datasets. Results are also compared with the Run 2 results measured at $\sqrt{s} = 13$ TeV.

lections of this Chapter. That is:

$$f_{\text{prompt}}^{D^0 D^0 \text{ cuts}} = \frac{N_{\text{prompt}} \cdot (\text{Acc} \times \epsilon)_{\text{prompt}}^{D^0 D^0 \text{ cuts}}}{N_{\text{prompt}} \cdot (\text{Acc} \times \epsilon)_{\text{prompt}}^{D^0 D^0 \text{ cuts}} + N_{\text{non-prompt}} \cdot (\text{Acc} \times \epsilon)_{\text{non-prompt}}^{D^0 D^0 \text{ cuts}}}. \quad (7.6)$$

Plotting the resulting prompt fraction as a function of p_T , we obtain the results presented in Fig. 7.15. The prompt fraction is seen to be nearly constant across all p_T ranges. Hence, we opted to include the prompt fraction as a constant value in the analysis, rather than a p_T -dependent value. The chosen prompt fraction was obtained by fitting a polynomial of order 0 to the prompt distribution, which is also plotted in Fig. 7.15. The resulting prompt fraction is 0.9780 ± 0.0003 .

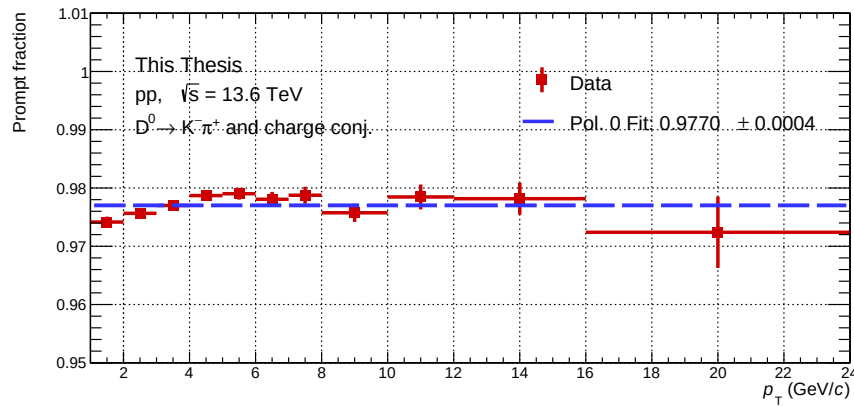


Figure 7.15: Prompt fraction obtained with the acceptance-times-efficiencies corresponding to the double D^0 analysis fitted to a polynomial of order 0.

7.8 Systematic uncertainties

A number of systematic uncertainties were considered in this analysis. These are (i) the corrected yield extraction, (ii) the ITS-TPC matching, (iii) the single track selections, (iv) the BDT selections, (v) the DCA resolution, evaluated via the track-tuner, (vi) the prompt fraction estimate, and (vii) the luminosity and BR. The systematic uncertainties were then added in quadrature to obtain the final uncertainty on the results for both OS and LS pairs.

7.8.1 Corrected yield extraction

The systematic uncertainties on the corrected yield extraction were estimated by varying the parameters of the invariant mass fits. The variations were performed by changing the invariant mass ranges and the background functions. In particular, we used first-degree polynomials and exponential functions of the form

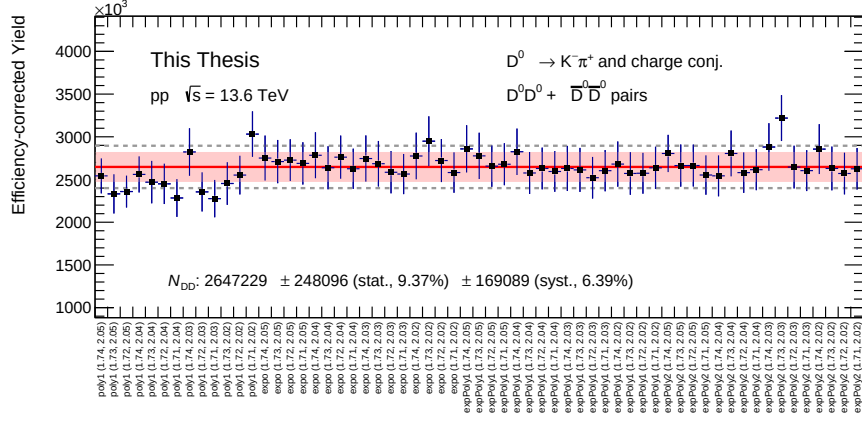
$$f(x) = e^{\alpha x + \beta}; \quad g(x) = e^{\alpha x^2 + \beta x + \gamma}$$

as background functions. The different invariant mass ranges were created by varying the lower limit between 1.71 and 1.74 GeV/ c^2 and the upper limit between 2.02 and 2.05 GeV/ c^2 . Only those trials whose χ^2/ndf was below 2 were considered valid.

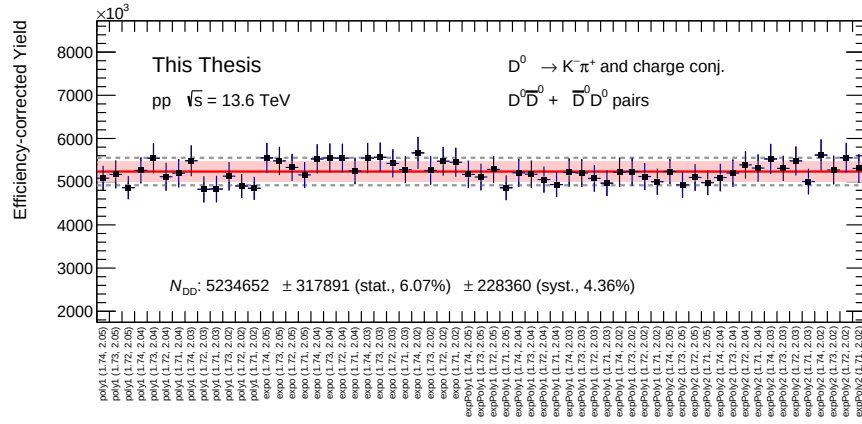
The resulting variations were then used to estimate the systematic uncertainties on the yields as $\sqrt{\text{RMS}^2 + \Delta^2}$, where Δ is the average of the trials with respect to the default case. The results of the variations are presented in Figs. 7.16a and 7.16b for the LS and OS pairs, respectively. They provide a systematic uncertainty of 6% for the LS pairs and 4% for the OS pairs. However, since we expect the yield extraction to be as stable for both types of pairs, we opted to use the average of the two values as the final systematic uncertainty on the yield extraction, corresponding to 5%.

7.8.2 ITS-TPC matching efficiency

The ITS-TPC matching efficiency was studied in [256] by measuring the discrepancies between the data and the MC. A *tag-and-probe* method was employed to reconstruct D^+ candidates using different ITS and TPC criteria. This method is based on *tagging* the two pion prongs by requiring they pass tight criteria and *probing* the kaon prong with the specific selections to be studied. At the end, an efficiency is computed as the ratio of probes that pass the criterion over the total number of probes. Using this method, they calculated a systematic uncertainty of 2% per prong, so we assigned an 8% for the D^0 pairs.



(a)



(b)

Figure 7.16: Corrected yields for the different trials varying the background fitting functions and invariant mass ranges in the $1 < p_T < 24$ GeV/c range for (a) Like-sign pairs (b) Opposite-sign pairs.

7.8.3 Single-track selections

The single-track selections were extracted from the analysis of the B^0 cross section in pp collisions at $\sqrt{s} = 13.6$ TeV [142]. In this analysis, they studied the effect of changing the minimum number of ITS clusters, the number of TPC rows and the maximum TPC χ^2/ndf per number of clusters allowed. They found a variation of 7% for B^0 candidates in the $1 < p_T < 2$ GeV/c range and 5% in all other p_T ranges up to 24 GeV/c. Since the B^0 mesons are reconstructed using four single-tracks, as in our case ($D^0 D^0 \rightarrow K^- \pi^+ K^- \pi^+$), we can assign the same systematic uncertainties for the selected p_T ranges.

7.8.4 DCA resolution

The systematic uncertainties on the DCA resolution were estimated by varying the parameters of the track-tuner (introduced in Sec. 5.3.2) and studying how they affect on the extraction of the efficiencies. The variations were performed by changing the parameters of the Gaussian smearing applied to the DCA distributions to worsen the resolution by 10% and 15% with respect to the resolution used throughout the analysis.

The ratio of the efficiencies with respect to the efficiency presented in Fig. 7.10 is located in Fig. 7.17. As can be observed, both variations provide efficiencies within 1% in all p_T ranges, which allows us to assign a global systematic uncertainty of 1% for each D^0 meson. Hence, 2% was assigned for the pairs.

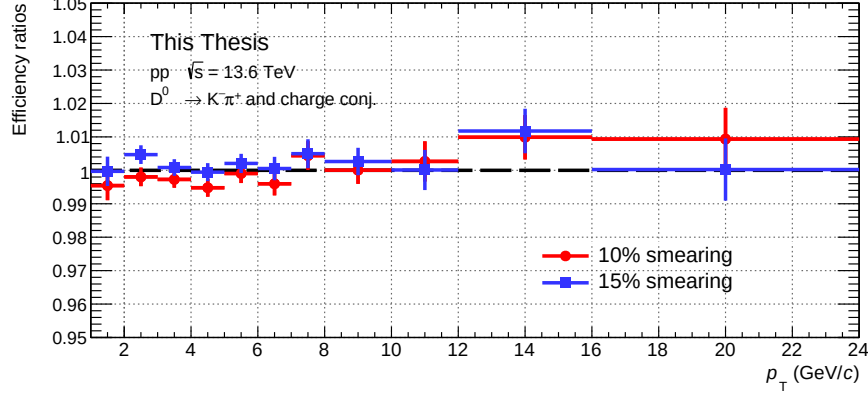


Figure 7.17: Ratio of the acceptance-times-efficiencies using worse resolution configurations of the track-tuner over the default configuration.

7.8.5 BDT selections

A systematic uncertainty was assigned to account for possible variations caused by the BDT selections. These variations were measured by recalculating the raw yields of the single D^0 candidates changing the BDT score cuts. A set of cuts was chosen to allow for a 20% change rate with respect to the default value used across this analysis. We considered the raw yields in the following p_T ranges, in GeV/c :

$$(1, 2), (2, 3), (3, 5), (5, 7), (7, 10), (10, 24),$$

which correspond to the ranges where the different BDT models were trained.

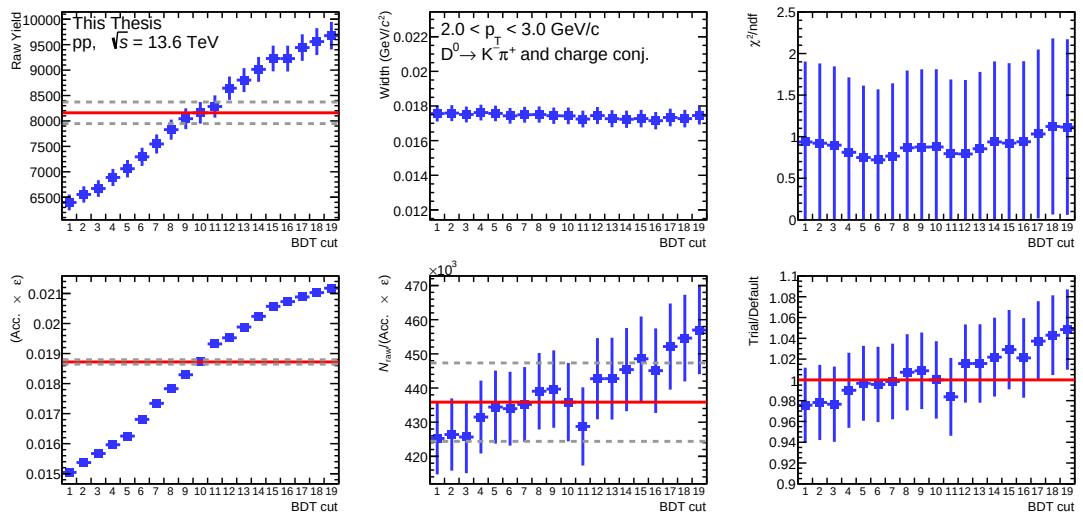


Figure 7.18: Calculation of the systematic uncertainty tied to the BDT selection.

An example of the systematic study can be found in Fig. 7.18, corresponding to the

$2 < p_T < 3$ GeV/ c range. The top-left plot contains the raw yield values as a function of the BDT cut, with the red line representing the raw yield obtained with the cut used for the analysis. To ensure the stability of the fits, we plotted the width of the D^0 peak and the corresponding χ^2/ndf per BDT cut in the top-centre and top-right plots, respectively. The bottom-left plot contains the acceptance-times-efficiencies per cut. These were used to calculate the corrected yields as the fraction of the raw yields by their associated efficiency, shown in the bottom-centre plot. They were later compared with the corrected yield obtained with the default cuts used throughout the analysis. The result of this comparison is presented on the right plot of the bottom half for the $2 < p_T < 3$ GeV/ c range.

Table 7.7: BDT selection systematic uncertainty as a function of p_T .

p_T range (GeV/ c)	(1, 2)	(2, 3)	(3, 5)	(5, 7)	(7, 10)	(10, 24)
BDT sel. systematic	5%	2%	2.5%	4%	6%	12%

Lastly, we present Tab. 7.7, containing the systematic uncertainty due to the BDT selections as a function of p_T . This uncertainty is calculated as the $\sqrt{\text{RMS}^2 + \Delta^2}$ of the efficiency-corrected yields per p_T range, and displays an increasing trend with p_T that ranges from 2% to 12% for $2 < p_T < 24$ GeV/ c , while the (1, 2) GeV/ c range has an uncertainty of 5%. Thus, to obtain a systematic uncertainty for the integrated cross section measurement, we used the candidate- p_T distributions and calculated the weighted average of the systematic uncertainty. This provided a final value for this systematic of 4% for the single D^0 candidates, since most of the statistics are located at low- p_T . Hence, we can assign an 8% uncertainty for the pairs. As a final remark, it is worth noting that the differences in the prompt fraction when varying the BDT cuts were assumed to be negligible. The rather conservative systematic uncertainty assigned to the prompt fraction, which will be presented in the next subsection, is expected to cover any possible variations.

7.8.6 Prompt fraction extraction

The prompt fraction was extracted in Sec. 7.7. We assigned a 1% systematic uncertainty for the extraction of the prompt fraction to all p_T ranges. Given that the prompt fraction itself is small, contributing approximately 2%, this level of uncertainty is considered sufficient to account for its impact.

7.8.7 Luminosity and BR

An uncertainty on the luminosity determination of 10% is assigned to all p_T ranges. The luminosity experts suggested this value for all measurements. On the other hand, a relative uncertainty on the BR of 0.76% was extracted from [9].

7.8.8 Summary of the considered systematics

The summary of all systematics included in this analysis is presented in Tab. 7.8 for single and double D^0 candidates. It is worth noting that the yield extraction and the single-track selection systematics were computed directly for the D^0 pair candidates (or for the four prongs of the B^0 meson in the case of the single-track selection). Thus, there is no systematic assignment specifically for the single D^0 case. The dominant source of uncertainty is the luminosity, providing a 10%. However, this systematic is expected to be reduced with more refined studies in the near future. The next uncertainty by level of dominance is the ITS-TPC matching efficiency, which adds up to 8% when considering all four prong tracks and the BDT selection. The total systematic uncertainty is computed as the sum in quadrature of all contributions and has a value of 18% for D^0 pairs.

Table 7.8: Summary of the systematic uncertainties for single and double D^0 candidates.

Systematic source	Final sys. single D^0	Final sys. double D^0
DCA resolution	1%	2%
BDT selections	4%	8%
Yield extraction	–	5%
Prompt fraction	1%	2%
ITS-TPC matching (from [256])	4%	8%
Trigger vs. MB check	5%	5%
Luminosity	10%	10%
BR	0.76%	1.52%
Single-track selections (from [142])	–	5%
Total systematic		18%

7.9 Cross section results

The integrated cross section of prompt ($D^0 D^0 + \bar{D}^0 \bar{D}^0$) and $D^0 \bar{D}^0$ pairs is computed as:

$$\sigma_{DD}^{\text{Prompt}} = \frac{N_{\text{corr}} \cdot f_{\text{prompt}}^2}{\mathcal{L}_{\text{int}} \cdot \text{BR}^2(D^0 \rightarrow K^- \pi^+)} \quad (7.7)$$

In this equation, f_{prompt} is the prompt fraction present in the data after performing the selections introduced in this Chapter. $\mathcal{L}_{\text{int}}$ corresponds to the integrated luminosity of $37.8 \pm 3.8 \text{ nb}^{-1}$, BR represents the branching ratio of the $D^0 \rightarrow K^+ \pi^-$ channel, and N_{corr} is the efficiency-corrected yield obtained in Sec. 7.6. The cross sections of both types of pairs at midrapidity computed using D^0 candidates in the $1 < p_T < 24 \text{ GeV}/c$ range are plotted in Fig. 7.19. They correspond to the values reported in the first column of Tab. 7.9.

The OS cross section is approximately twice as large as the LS one. More specifically,

Table 7.9: Integrated cross sections of the associated production of D^0 -meson pairs at $\sqrt{s} = 13.6$ TeV in the rapidity region $|y| < 0.5$.

	Cross section (μb)	$\mathcal{R}_{DD}$ (mb)
Opposite-sign pairs	$86.8 \pm 5.1(\text{stat.}) \pm 14.6(\text{sys.})$	$1.41 \pm 0.18(\text{stat.}) \pm 0.32(\text{sys.})$
Like-sign pairs	$42.7 \pm 3.9(\text{stat.}) \pm 7.2(\text{sys.})$	$2.88 \pm 0.46(\text{stat.}) \pm 0.66(\text{sys.})$

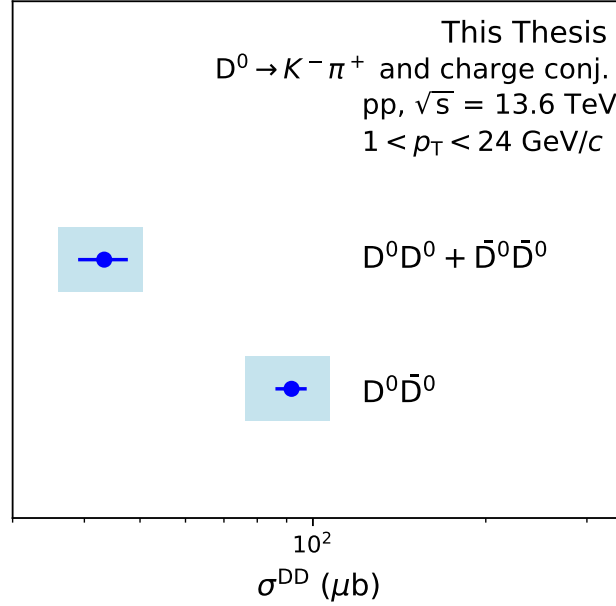


Figure 7.19: Cross section of the production of prompt $D^0 D^0$ (upper point) and $D^0 \bar{D}^0$ (bottom point) pairs measured at midrapidity at $\sqrt{s} = 13.6$ TeV.

it amounts to

$$\sigma_{OS}/\sigma_{LS} = 2.12 \pm 0.25(\text{stat.}) \pm 0.50(\text{sys.}).$$

This contrasts with the results found by the LHCb collaboration measured at forward rapidity at $\sqrt{s} = 7$ TeV [159], where they report a ratio of 10.9. However, PYTHIA 8 simulations show that this difference is caused by the transverse-momentum and rapidity regions explored by each analysis. The ratio of OS over LS pair cross sections as a function of the transverse momentum of the single D^0 candidate and the rapidity can be found in Fig. 7.20. The LHCb measurement covered the rapidity region $2 < y < 4$ at $3 < p_T < 12$ GeV/c. Our measurement, on the other hand, covers the rapidity region $|y| < 0.5$ at $1 < p_T < 24$ GeV/c. Hence, the average p_T of our measurement is significantly lower than that of the LHCb measurement, making the ratio smaller.

We can also study the value of the following ratio:

$$\mathcal{R}_{DD} = \alpha \cdot \frac{\sigma_{D^0} \sigma_{D^0}}{\sigma_{DD}}. \quad (7.8)$$

where $\alpha = 1/4$ accounts for the two particles being identical and non-self conjugate, and the single D^0 cross section is extracted from the ALICE measurement at $\sqrt{s} =$

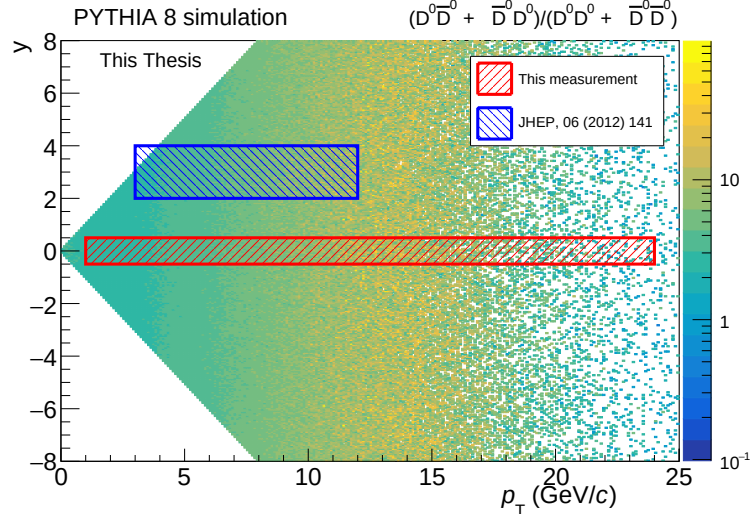


Figure 7.20: PYTHIA 8 simulation of the ratio of OS over LS pair production as a function of p_T and y . The coverage of the LHCb measurement [159] is marked in blue, while our coverage is indicated in red.

13 TeV [125]. This ratio is defined similarly to σ_{eff} (see Sec. 2.2), but it is not the same value. For this ratio to be equal to σ_{eff} , it is necessary to subtract the SPS contribution from the double cross section and be left exclusively with the DPS component. The measured results of $\mathcal{R}_{DD}$ in the $1 < p_T < 24$ GeV/c range at midrapidity are presented in Fig. 7.21 and reported in the second column of Tab. 7.9.

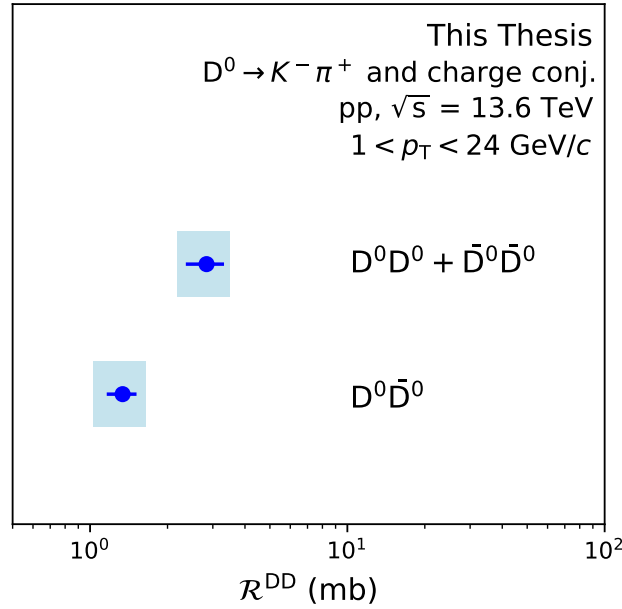


Figure 7.21: Ratio of the single-over-double D^0 production measured at midrapidity for like-sign (upper point) and opposite-sign (bottom point) D^0 pairs.

Even though this ratio is not equal to the effective cross section that is commonly measured in DPS studies, as we stated, we can study its order of magnitude to cross-check

the double D^0 cross section measurement. In particular, we expect the LS ratio, $\mathcal{R}_{DD}$, to be slightly smaller than the effective cross section we would obtain after separating the DPS and SPS components. Comparing our LS ratio with the effective cross section measurements of Fig. 2.9, we can conclude that this ratio is lower than these measurements, but lies in the same order of magnitude as the measurements that involve charm quarks.

7.10 p_T -differential results

The analysis was repeated separating the data sample in five pair- p_T ranges, p_T^{DD} . The pair- p_T is defined as the transverse momentum of the vectorial sum of the Lorentz vectors of the two D^0 candidates that form the pair. The chosen ranges were:

$$(0, 2), (2, 4), (4, 6), (6, 12), \text{ and } (12, 24) \text{ GeV}/c.$$

As in the p_T -integrated case, we used an unbinned-likelihood fit on the invariant mass distributions to extract the yields, which were corrected by the $(\text{Acc} \times \epsilon)$ on a candidate-by-candidate basis using the efficiency map in Fig. 7.8. In all considered p_T^{DD} ranges, a Gaussian was used to model the signal, a double Gaussian was used for the reflections, and an exponential was used for the background.

To study the reflected-over-signal value to add in each p_T^{DD} region, we began by studying the reflected-over-signal fraction for the single- p_T candidates. The results are located in Fig. 7.22. To understand them, it is worth reminding that we removed from the analysis all candidates simultaneously passing the D^0 and $\bar{D}^0$ hypotheses whose p_T was smaller than 5 GeV/ c . The PID selections are only applied up to 5 GeV/ c as well, which explains why the values rise with p_T and reach a maximum of around 0.6.

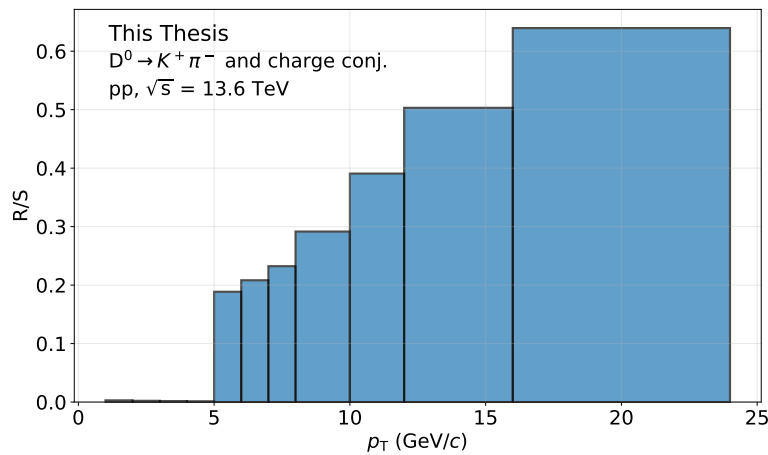


Figure 7.22: Value of the reflected-over-signal fraction for single D^0 candidates. The reflections were removed for all candidates with $p_T < 5$ GeV/ c .

Afterwards, we took the single- p_T distribution of all candidates per pair- p_T bin and

assigned them a reflection-over-signal value based on the previous results. This way, we were able to compute the average reflection-over-signal fraction per pair- p_T bin. The resulting fractions are presented in Tab. 7.10. As can be observed, the fractions tend to increase with the pair- p_T , since high- p_T pairs are more likely to be composed of high- p_T candidates.

Table 7.10: *Reflected-over-signal fractions per pair- p_T range.*

p_T^{DD} GeV/c	(0, 2)	(2, 4)	(4, 6)	(6, 12)	(12, 24)
R/S	0.003	0.004	0.012	0.06	0.11

The resulting corrected yield extractions are located in Figs. 7.23, 7.24, 7.25, and 7.26. The first two figures correspond to OS and the last two to LS pairs.

We considered the same systematic uncertainties as the ones presented in Sec. 7.8. The yield extraction systematic was recomputed for each p_T^{DD} interval. We employed the same procedure as in the integrated case, varying the background functions and mass ranges per range. The same functions and mass ranges were considered, and only the trials whose χ^2/ndf was below 2 were kept. The final uncertainty was computed as the $\sqrt{\text{RMS}^2 + \Delta^2}$. We also recomputed the weighted average of the BDT selection systematic with the single- p_T distributions obtained for each pair- p_T range. All other systematics were assigned the same value as in the integrated case. The systematic uncertainties per pair- p_T range are located in Tab. 7.11.

Table 7.11: *Systematic uncertainties on the yield extraction as a function of the p_T of the pair.*

p_T^{DD} GeV/c	(0, 2)	(2, 4)	(4, 6)	(6, 12)	(12, 24)
Yield extraction sys.	5%	6%	5%	7%	11%
BDT selection sys.	8%	8%	6%	8%	15%
Total sys.	18%	18%	17%	18%	24%

As can be observed, the systematic uncertainties tend to increase at high- p_T^{DD} . This is caused by the fact that these pairs contain a larger proportion of candidates with a higher p_T , which have a worse resolution.

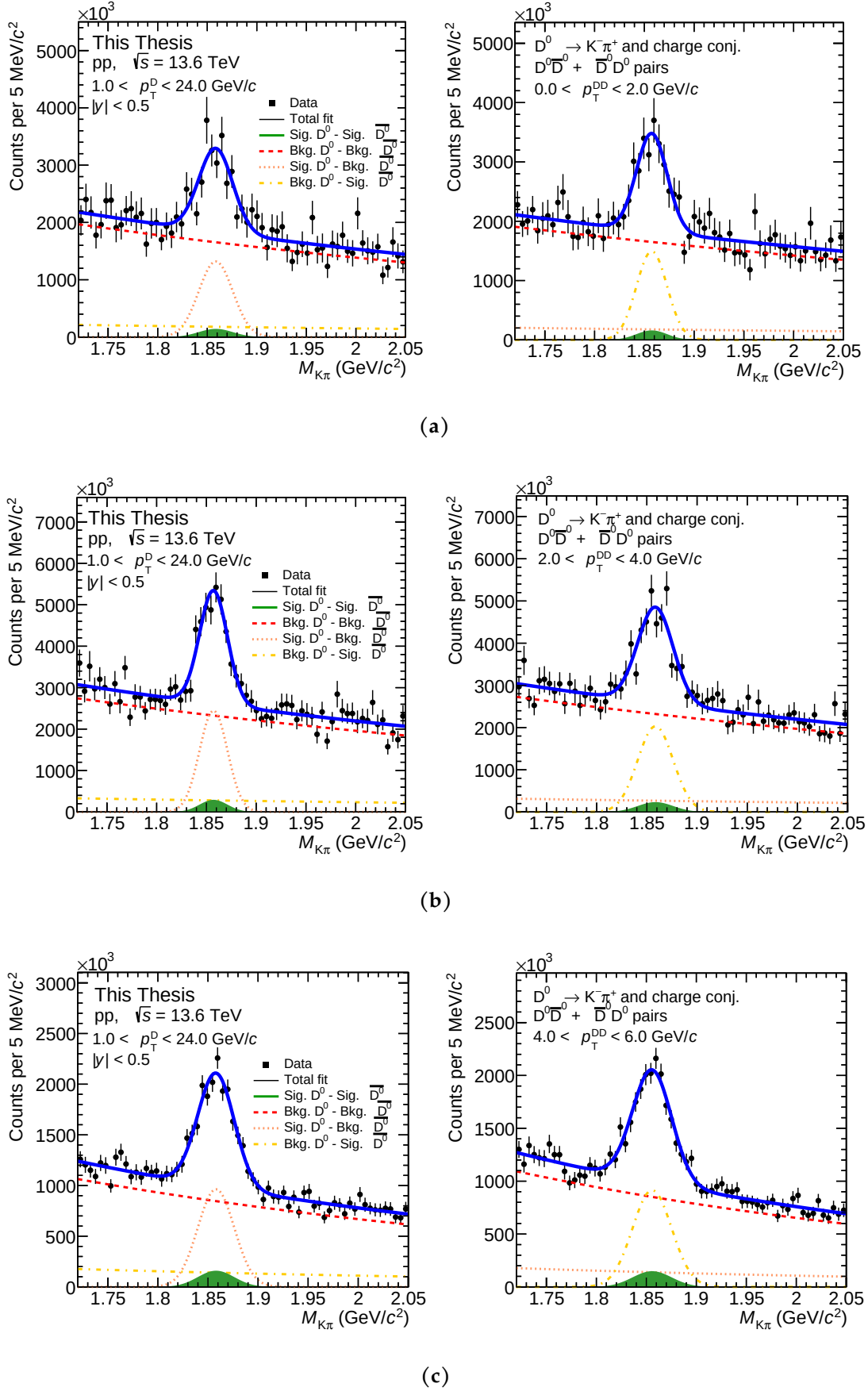


Figure 7.23: Projections of the two-dimensional plots of the invariant mass distributions of the D^0 pair candidates in the (a) $0 < p_T^{D\bar{D}} < 2$, (b) $2 < p_T^{D\bar{D}} < 4$, and (c) $4 < p_T^{D\bar{D}} < 6$ GeV/c range for opposite-sign pairs. In all cases, the left and right panels represent the projection on the invariant mass of the two D^0 candidates that form the pair.

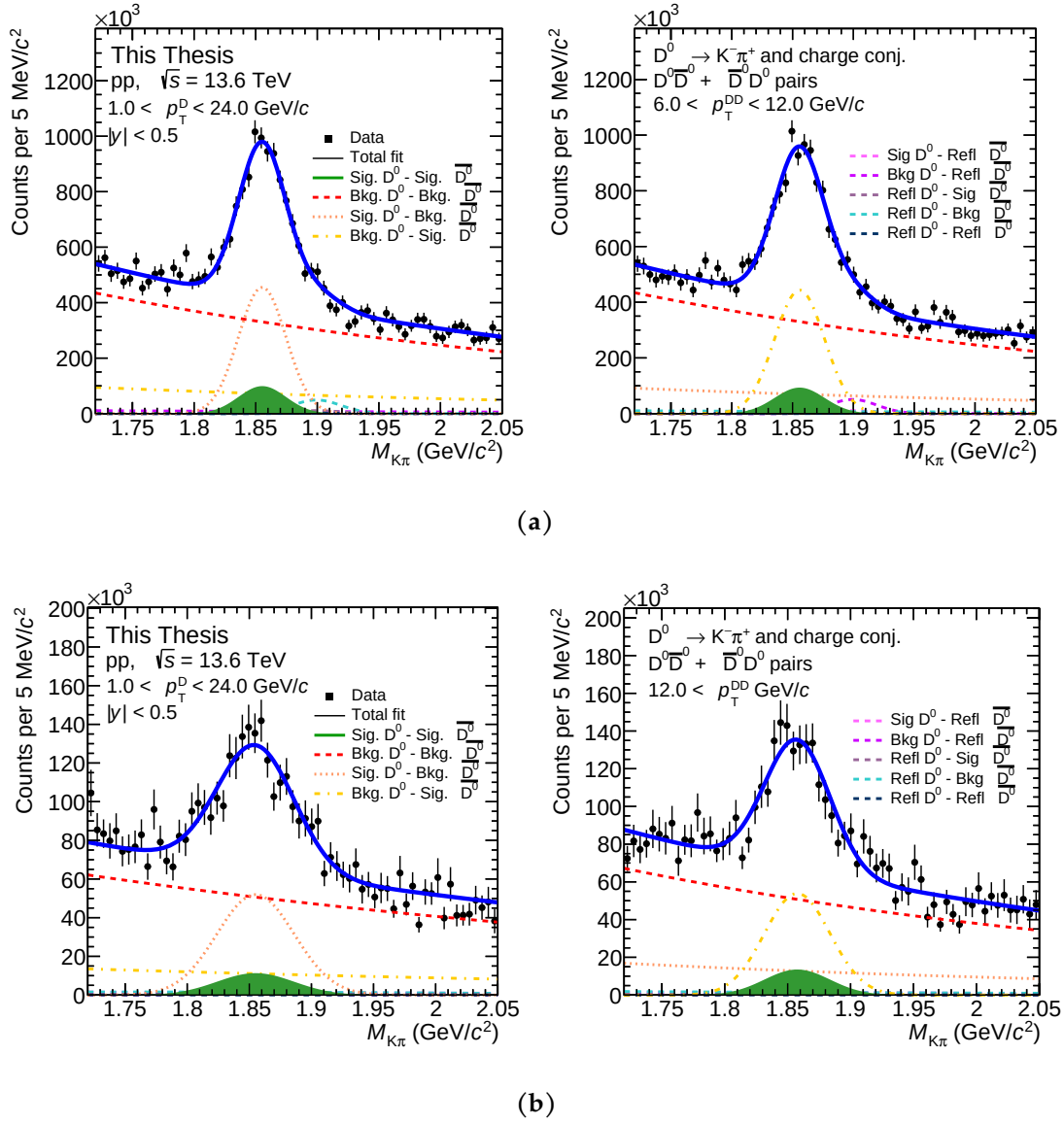


Figure 7.24: Projections of the two-dimensional plots of the invariant mass distributions of the D^0 pair candidates in the (a) $6 < p_T^{DD} < 12$ and (b) $p_T^{DD} > 12$ GeV/c range for opposite-sign pairs. In all cases, the left and right panels represent the projection on the invariant mass of the two D^0 candidates that form the pair.

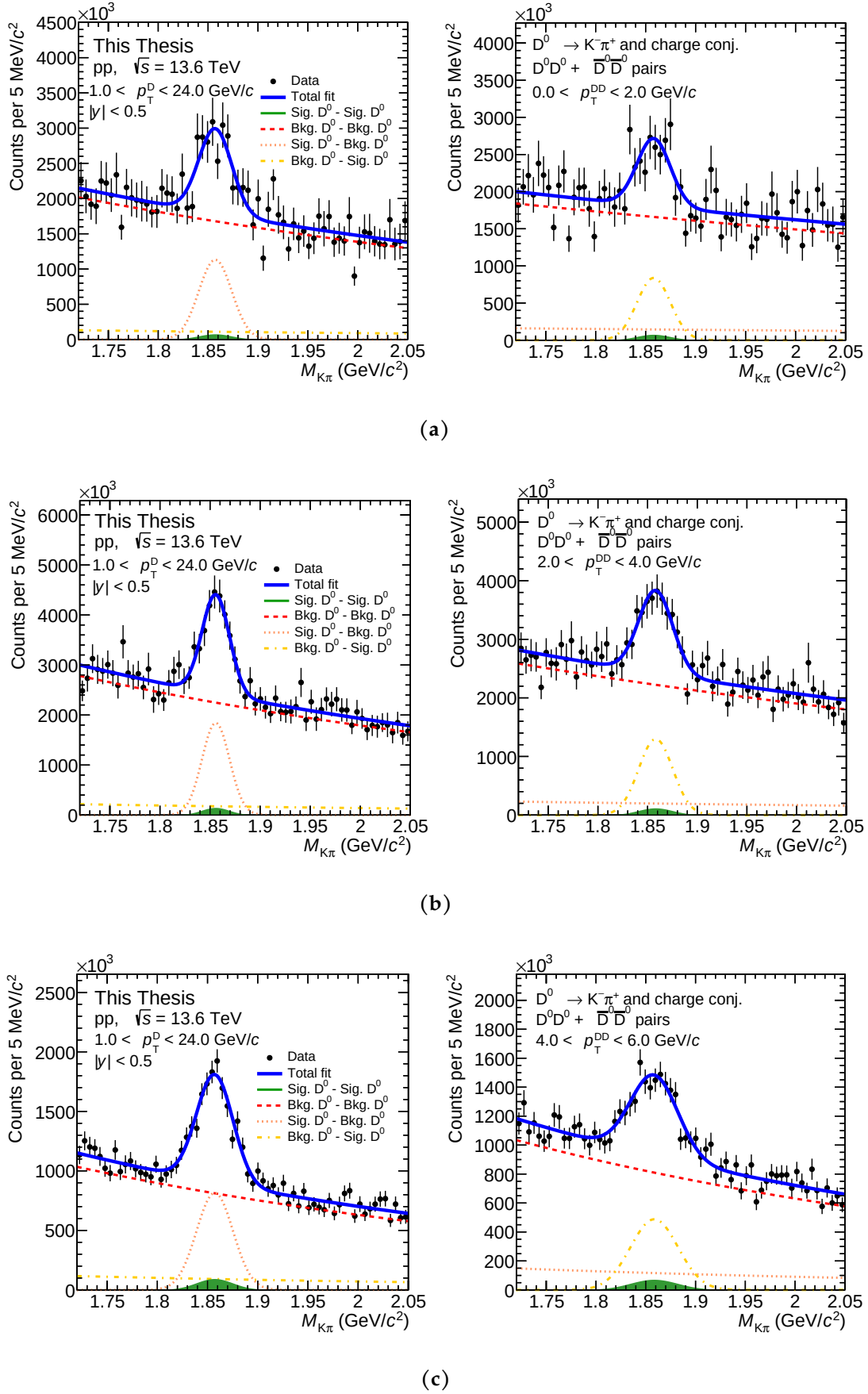


Figure 7.25: Projections of the two-dimensional plots of the invariant mass distributions of the D^0 pair candidates in the (a) $0 < p_T^{DD} < 2$, (b) $2 < p_T^{DD} < 4$, and (c) $4 < p_T^{DD} < 6$ GeV/c range for Like-sign pairs. In all cases, the left and right panels represent the projection on the invariant mass of the two D^0 candidates that form the pair.

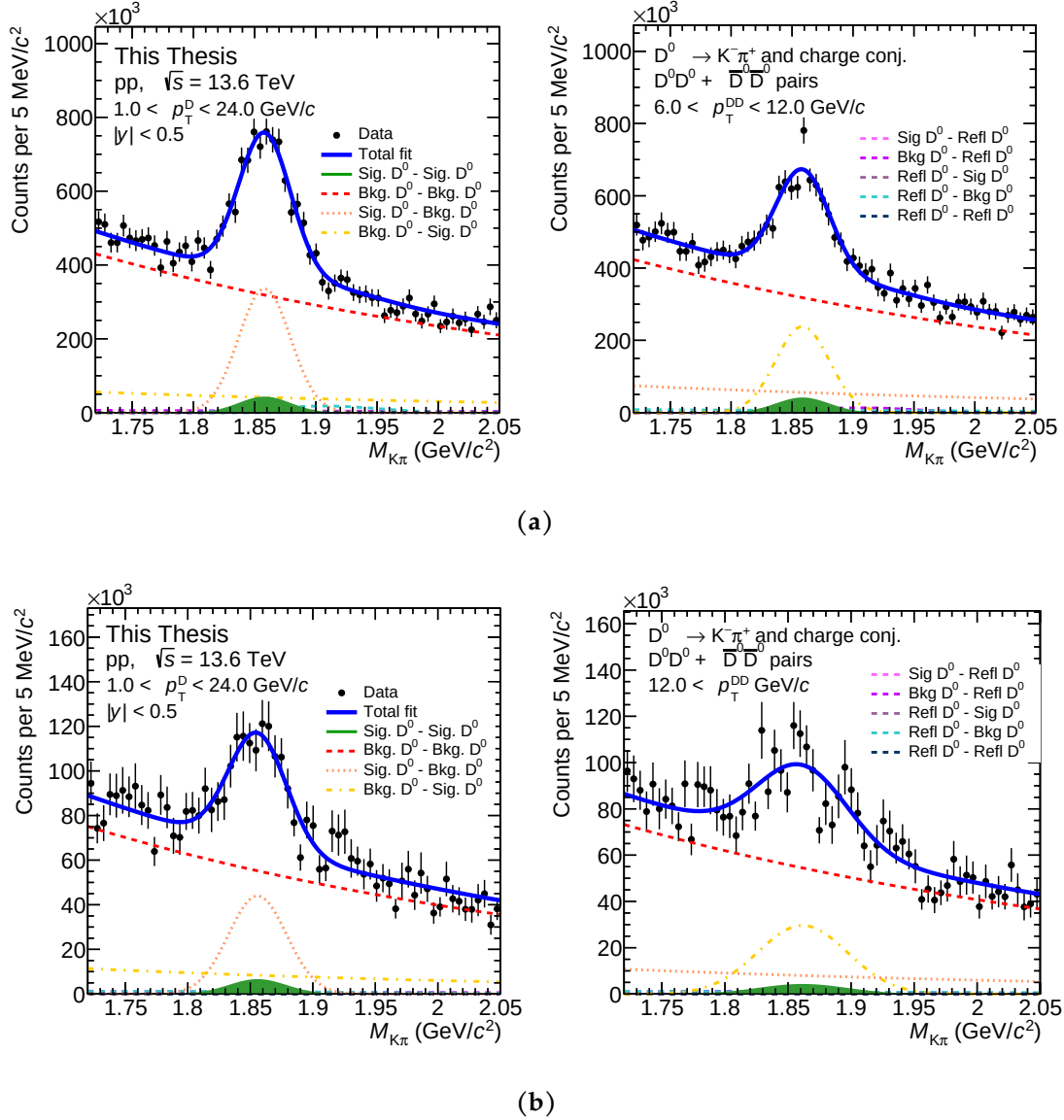


Figure 7.26: Projections of the two-dimensional plots of the invariant mass distributions of the D^0 pair candidates in the (a) $6 < p_T^{DD} < 12$ and (b) $p_T^{DD} > 12$ GeV/c range for Like-sign pairs. In all cases, the left and right panels represent the projection on the invariant mass of the two D^0 candidates that form the pair.

The p_T^{DD} -differential cross sections measured in pp collisions at $\sqrt{s} = 13.6$ TeV at midrapidity are presented in Figs. 7.27 and 7.28 for OS and LS pairs, respectively. They were computed with the following expression:

$$\frac{d\sigma_{\text{prompt}}^{DD}}{dp_T^{DD}} = \frac{N_{\text{corr}} \cdot f_{\text{prompt}}^2}{\mathcal{L}_{\text{int}} \cdot \text{BR}^2(D^0 \rightarrow K^- \pi^+) \cdot \Delta p_T^{DD}}, \quad (7.9)$$

where the prompt fraction, f_{prompt} , is the one obtained in Sec. 7.7. N_{corr} is the corrected yield computed from the invariant mass plots presented above per p_T^{DD} range, $\mathcal{L}_{\text{int}} = 37.8 \pm 3.8 \text{ nb}^{-1}$ is the integrated luminosity, and BR is the branching ratio. Lastly, Δp_T^{DD} accounts for the chosen pair- p_T bin size.

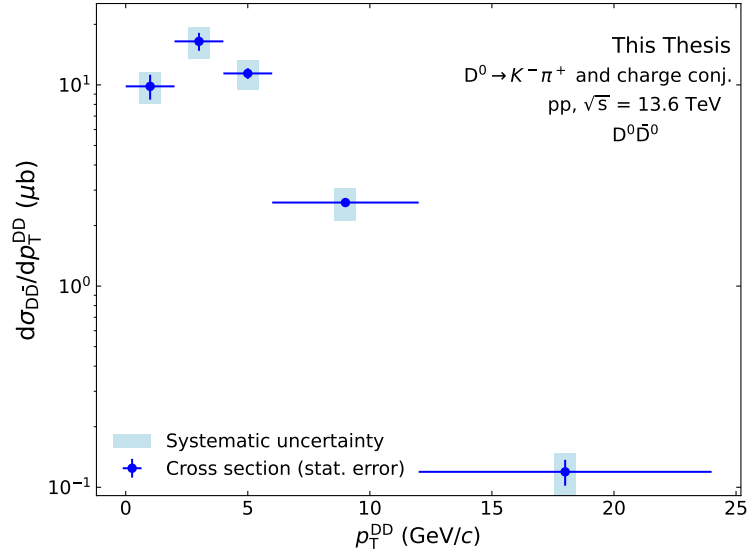


Figure 7.27: Differential cross section for opposite-sign pairs measured at midrapidity in pp collisions at $\sqrt{s} = 13.6$ TeV.

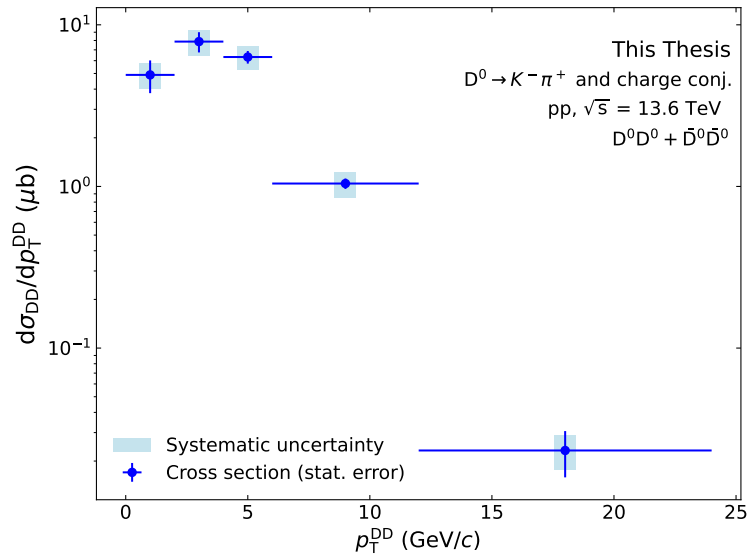


Figure 7.28: Differential cross section for like-sign pairs measured at midrapidity in pp collisions at $\sqrt{s} = 13.6$ TeV.

Lastly, we computed the OS-over-LS ratio of the differential cross sections, which we present in 7.29. As we can see, the ratios are approximately constant at low- p_T^{DD} , where the OS pairs are approximately twice as abundant as the LS pairs. However, at high- p_T^{DD} , the ratio tends to increase, reaching a value of around 5 at $p_T^{\text{DD}} > 12$ GeV/ c . This is consistent with the PYTHIA 8 simulation presented in Fig. 7.20, since, as we discussed before, this range includes pairs with a higher amount of high- p_T candidates, where this ratio is larger. Nonetheless, it is worth noting that the uncertainties are also larger in this region, making the ratio compatible with a less pronounced increase.

As a sanity check, we integrated the differential cross sections and compared the re-

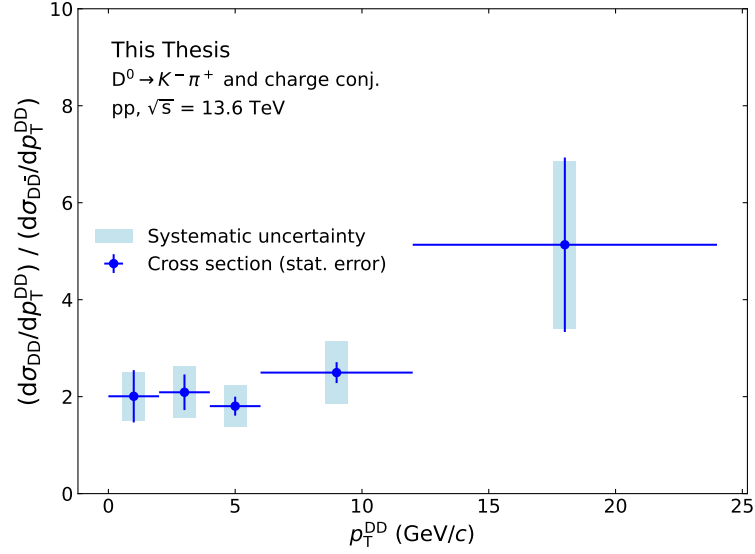


Figure 7.29: Differential cross section ratios of opposite-sign over like-sign pairs measured at midrapidity in pp collisions at $\sqrt{s} = 13.6$ TeV.

sults with the full cross sections measured in Sec. 7.9 and obtained the following results in Tab. 7.12. As can be seen, the two methods provide compatible results for both types of pairs.

Table 7.12: Comparison of the cross sections of the associated production of D^0 -meson pairs computed in one pair- p_T bin vs. multiple bins at $\sqrt{s} = 13.6$ TeV in the rapidity region $|y| < 0.5$.

	Full cross section (μb)	Diff. cross sections summed (μb)
Opposite-sign pairs	$86.8 \pm 5.1(\text{stat.}) \pm 14.6(\text{sys.})$	$92.6 \pm 8.4(\text{stat.}) \pm 14.5(\text{sys.})$
Like-sign pairs	$42.7 \pm 3.9(\text{stat.}) \pm 7.2(\text{sys.})$	$44.8 \pm 6.1(\text{stat.}) \pm 7.0(\text{sys.})$

7.11 Towards DPS extraction

The like-sign D^0 pair production is expected to contain a large number of events coming from double parton scattering (DPS) processes. As we introduced in Sec. 2.2, it is possible to separate between single and double scattering processes by studying kinematic variables such as $\Delta\varphi$ and Δy . Due to the limited rapidity range covered in this analysis ($|\Delta y| < 1$), it is very challenging to separate SPS from DPS using this variable, since DPS events are only expected to dominate at large Δy values. In addition, SPS production peaks at midrapidity, further complicating DPS extraction in our kinematic region. The $|\Delta y|$ distribution of LS pairs in the invariant mass peak region is plotted in Fig. 7.30a, and is found to have a very similar distribution to that of OS pairs, which is located in Fig. 7.30b.

We can also look at the $|\Delta\varphi|$ distributions for LS and OS pairs in Figs. 7.31a and 7.31b, respectively. Both distributions include peaks at $|\Delta\varphi| = 0$ and $|\Delta\varphi| = \pi$, but the peak at 0 is less pronounced for the LS case, while both distributions reach a minimum at

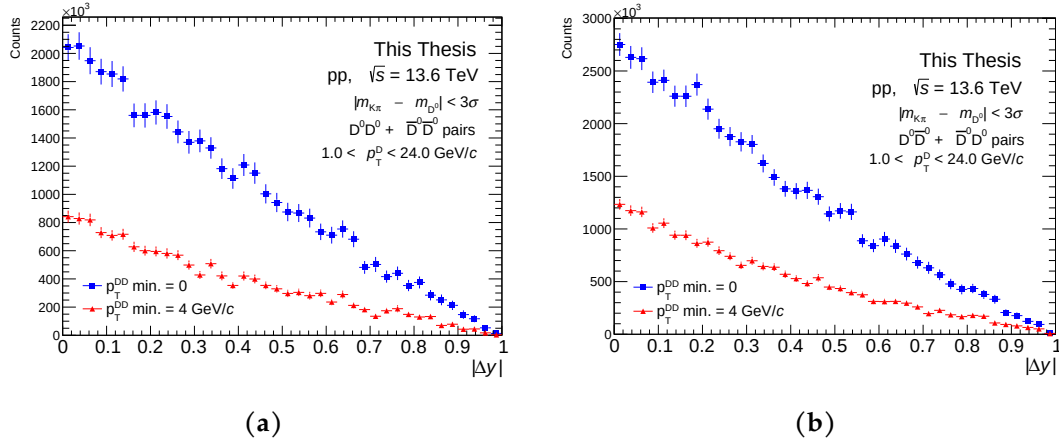


Figure 7.30: Distribution of $|\Delta y|$ for (a) like-sign pairs, (b) opposite-sign pairs.

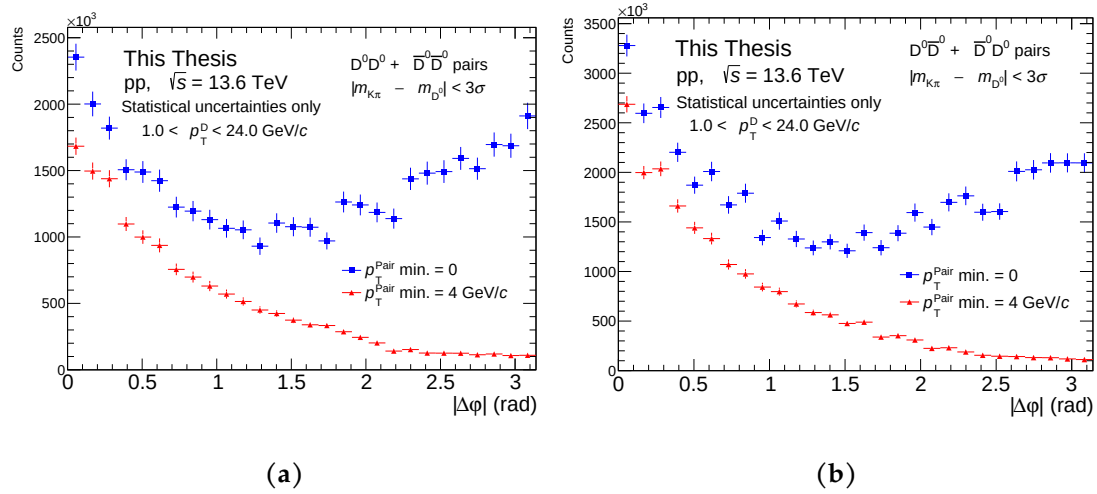


Figure 7.31: Distribution of $|\Delta\phi|$ for (a) like-sign pairs, (b) opposite-sign pairs.

$|\Delta\phi| = 1.5$, containing around 1500 counts. Thus, we can conclude that LS pairs have a flatter distribution than OS pairs. This is expected, since DPS events are naively expected to be more prominent in the LS case, and DPS events are expected to have a flat $|\Delta\phi|$ distribution. However, the differences between the two distributions disappear when we look at pairs of events whose $p_T^{\text{DD}} > 4$ GeV/c.

To better understand the differences when adding a cut on the p_T of the pair, we can look at a toy DPS MC created using PYTHIA 8. In this MC, two independent hard scatterings producing a D^0 meson each were generated. The resulting $|\Delta\phi|$ distribution is presented in Fig. 7.32. As can be observed, the distribution is flat when considering all pairs, but it develops a peak at $|\Delta\phi| = 0$ when we only consider pairs with $p_T^{\text{DD}} > 4$ GeV/c. This is caused by the fact that, to create a high- p_T pair, both D^0 candidates must have a similar direction in the transverse plane. This effect could explain why the differences between OS and LS pairs disappear when we add a cut on the pair- p_T .

All in all, it could be possible to extract the DPS component from the measurement of

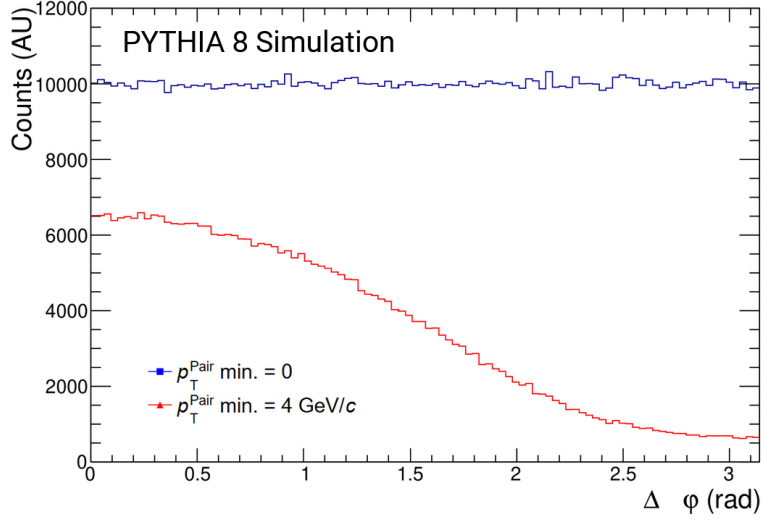


Figure 7.32: $|\Delta\varphi|$ distribution computed with a toy DPS MC.

LS D^0 pairs at midrapidity by exploiting the $|\Delta\varphi|$ distribution. This could be done by performing a templated fit, assuming that DPS pairs have a flat $|\Delta\varphi|$ distribution. The SPS distribution, on the other hand, could be extracted from theoretical inputs. This separation could allow us to obtain a value of the effective cross section, σ_{eff} , for the first time in ALICE and compare it with previous measurements (Fig. 2.9). However, at the time of writing this thesis, the theoretical inputs were not yet available, explaining why this separation was not performed.

Summary of this chapter

In this chapter, we presented the first measurement of the associated production of D^0 -meson pairs at midrapidity in pp collisions at $\sqrt{s} = 13.6$ TeV, separating the like-sign (LS) and opposite-sign (OS) cases. The analysis was performed using a data sample collected with a minimum-bias trigger during the 2024 LHC run, on top of which an offline selection trigger was applied. The final integrated luminosity is $37.8 \pm 3.8 \text{ nb}^{-1}$. The candidates were selected using a machine-learning algorithm based on boosted decision trees, which allowed us to improve the signal-to-background ratio significantly. The candidates were reconstructed via their hadronic decay channel $D^0 \rightarrow K^- \pi^+$ and its charge conjugate. The candidates were selected in the $1 < p_T < 24 \text{ GeV}/c$ range and $|y| < 0.5$. The integrated cross sections of prompt $D^0 D^0 + \bar{D}^0 \bar{D}^0$ and $D^0 \bar{D}^0$ pairs were measured to be $42.7 \pm 3.9(\text{stat.}) \pm 7.2(\text{sys.}) \mu\text{b}$ and $86.8 \pm 5.1(\text{stat.}) \pm 14.6(\text{sys.}) \mu\text{b}$, respectively. The ratio of OS over LS cross sections was found to be $2.12 \pm 0.25(\text{stat.}) \pm 0.50(\text{sys.})$, which is significantly smaller than the value of 10.9 reported by LHCb at forward rapidity at $\sqrt{s} = 7 \text{ TeV}$ [159]. This difference is explained by the different kinematic regions covered by both analyses, as shown by PYTHIA 8 simulations. In addition, the measurement was repeated in five pair- p_T ranges, and the differential cross sections were presented. The OS-over-LS ra-

tio of the differential cross sections was also computed, showing that the ratio tends to increase with the pair- p_T . Lastly, we discussed the possibility of extracting the DPS component from the LS measurement by exploiting the $|\Delta\varphi|$ distribution, which could allow us to obtain a value of the effective cross section, σ_{eff} . This extraction remains as a next step of the analysis. Model calculations are required to interpret data, and are missing to date. These results will encourage theorists to provide their inputs.

CONCLUSIONS

ALICE is an experiment located at the Large Hadron Collider which specialises in the study of quark-gluon plasma (QGP) through ultra-relativistic heavy-ion collisions. In these collisions, the QGP medium is only produced for a few femtoseconds, making direct measurements impossible. Instead, we must resort to its study via probes, such as heavy-flavour hadrons. Heavy-flavours are produced at the beginning of the collision, so they experience the full evolution of the medium. To understand how heavy-flavour production is altered by their passage through the QGP, it is necessary to compare the measurements in heavy-ion collisions with smaller systems, such as proton-nucleus (p-A) or proton-proton (pp) collisions. In addition, recent studies in smaller systems suggest that some signatures commonly attributed to the QGP are also present in high-multiplicity p-A and pp collisions. For instance, a strangeness enhancement was observed in high-multiplicity pp and p-Pb collisions, as well as a non-zero elliptic flow of charged particles. One possible explanation for these behaviours could be related to multiple parton interactions, or MPIs. These occur when more than one parton collides with another parton in a single hadron-hadron collision.

All measurements presented in this thesis were performed using ALICE Run 3 data. In this data-taking period, ALICE is using a continuous data-taking mode, which allows for a much larger statistics than previous Runs. The collected data must be understood before being able to perform physics analyses with them. For this reason, I have contributed in the study of the muon tracking efficiency of the muon chambers (MCH). In particular, I have studied the efficiencies as a function of η and φ of the Pb-Pb 2023 period with several Monte Carlo (MC) samples. These studies were then used to propose a systematic uncertainty that will need to be included in all analyses that involve the MCH with this data period.

Afterwards, I have focused on the study of D^0 mesons in pp collisions at $\sqrt{s} = 13.6$ TeV. These particles are produced in hard-scattering processes, and their large masses allow us to calculate their production with perturbative quantum chromodynamics (pQCD). They were selected using Boosted Decision Trees, a Machine Learning technique commonly used to separate signal and background candidates in experimental high energy

physics. I begun by studying the fraction of non-prompt D^0 mesons in pp collisions at $|y| < 0.5$, *i.e.* the number of D^0 mesons produced in the decay of beauty hadrons. I extracted the fraction using a data-driven method called the cut-variation method. The results presented in this thesis are compatible with previous measurements, and were compared with EPOS 4, several PYTHIA 8 tunes and FONLL predictions. Overall, EPOS tends to underestimate the data, while PYTHIA and FONLL tend to overestimate the data. These results were expected, given the data-to-model comparison of the prompt and non-prompt D-meson cross sections shown in Ref. [115, 125].

At last, I have studied the double production of D^0 mesons with the purpose of gaining insights on the double parton scattering (DPS). It is the simplest form of MPI, as it requires at least two parton-parton collisions in one hadron-hadron collision. It is expected that like-sign (LS) D^0 pairs, *i.e.* $D^0 D^0$ and $\bar{D}^0 \bar{D}^0$ pairs, have a large DPS contribution, since they require the production of at least two $c\bar{c}$ pairs. I have also studied the production of opposite-sign (OS) pairs, $D^0 \bar{D}^0$, and provided their p_T -integrated and differential cross sections. The ratio of opposite-sign to like-sign production is also calculated. This study constitutes the first measurement of D-meson pair production in ALICE, and was only possible thanks to the higher statistics of Run 3. Boosted Decision Trees were also crucial to separate between signal and background D^0 candidates.

As a further step, I aim to compare our results of the double D^0 production with theoretical predictions. Furthermore, I intend to use their $\Delta\varphi$ distribution to separate the single and double parton scattering components. This way, it will be possible to provide a measurement of the effective cross section for the first time in ALICE.

BIBLIOGRAPHY

- [1] F. Halzen and A. D. Martin. *Quarks and leptons: An introductory course in modern Particle Physics*. 1984. ISBN: 978-0-471-88741-6.
- [2] A. Purcell. “Go on a particle quest at the first CERN webfest. Le premier webfest du CERN se lance à la conquête des particules”. In: 35/2012 (2012), p. 10.
- [3] **Belle** Collaboration. “Observation of a Narrow Charmonium-like State in Exclusive $B^\pm \rightarrow K^\pm \pi^+ \pi^- J/\psi$ Decays”. In: *Phys. Rev. Lett.* 91.26 (Dec. 2003), p. 262001. DOI: 10.1103/physrevlett.91.262001.
- [4] **LHCb** Collaboration. “Observation of J/ψ p Resonances Consistent with Pentaquark States in $\Lambda_b^0 \rightarrow J/\psi K^- p$ Decays”. In: *Phys. Rev. Lett.* 115.7 (Aug. 2015), p. 072001. ISSN: 1079-7114. DOI: 10.1103/physrevlett.115.072001.
- [5] P. W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.
- [6] **ATLAS** Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett. B* 716.1 (July 31, 2012), pp. 1–29. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].
- [7] **CMS** Collaboration. “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].
- [8] M. E. Peskin and D. V. Schroeder. *An Introduction to Quantum Field Theory*. Reading, USA: Addison-Wesley 842 p. Westview Press, 1995.
- [9] S. Navas et al. “Review of particle physics”. In: *Phys. Rev. D* 110.3 (2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001.
- [10] J. -L. Kneur and A. Neveu. “ $\Lambda_{\overline{MS}}^{\text{QCD}}$ from Renormalization Group Optimized Perturbation”. In: *Phys. Rev. D* 85.1 (Jan. 2011), p. 014005. ISSN: 1550-2368. DOI: 10.1103/physrevd.85.014005. arXiv: 1108.3501 [hep-ph].
- [11] Y. Nambu. “Quasi-Particles and Gauge Invariance in the Theory of Superconductivity”. In: *Phys. Rev.* 117.3 (Feb. 1960), pp. 648–663. ISSN: 0031-899X. DOI: 10.1103/physrev.117.648.
- [12] Y. Nambu and G. Jona-Lasinio. “Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. I”. In: *Phys. Rev.* 122.1 (Apr. 1961), pp. 345–358. ISSN: 0031-899X. DOI: 10.1103/physrev.122.345.
- [13] J. Goldstone. “Field theories with « Superconductor » solutions”. In: *Il Nuovo Cimento* 19.1 (Jan. 1961), pp. 154–164. ISSN: 1827-6121. DOI: 10.1007/bf02812722.
- [14] J. Goldstone, A. Salam, and S. Weinberg. “Broken Symmetries”. In: *Phys. Rev.* 127.3 (Aug. 1962), pp. 965–970. ISSN: 0031-899X. DOI: 10.1103/physrev.127.965.

- [15] J. C. Collins and M. J. Perry. “Superdense Matter: Neutrons or Asymptotically Free Quarks?” In: *Phys. Rev. Lett.* 34.21 (May 1975), pp. 1353–1356. ISSN: 0031-9007. DOI: 10.1103/physrevlett.34.1353.
- [16] E.V. Shuryak. “Quark-gluon plasma and hadronic production of leptons, photons and pions”. In: *Phys. Lett. B* 78.1 (Sept. 1978), pp. 150–153. ISSN: 0370-2693. DOI: 10.1016/0370-2693(78)90370-2.
- [17] R. Stock et al. “Compression Effects in Relativistic Nucleus-Nucleus Collisions”. In: *Phys. Rev. Lett.* 49.17 (Oct. 1982), pp. 1236–1239. ISSN: 0031-9007. DOI: 10.1103/physrevlett.49.1236.
- [18] U. W. Heinz and M. Jacob. “Evidence for a new state of matter: An Assessment of the results from the CERN lead beam program”. In: (Jan. 2000). arXiv: nucl-th/0002042.
- [19] E. Shuryak. “Why does the Quark-Gluon Plasma at RHIC behave as a nearly ideal fluid?” In: *Prog. Part. Nucl. Phys.* 53.1 (July 2003), pp. 273–303. DOI: 10.1016/j.ppnp.2004.02.025. arXiv: hep-ph/0312227 [hep-ph].
- [20] A. Aprahamian. “Reaching for the horizon: The 2015 long range plan for nuclear science”. In: (Oct. 2015).
- [21] M. G. Alford et al. “Color superconductivity in dense quark matter”. In: *Rev. Mod. Phys.* 80.4 (Nov. 2008), pp. 1455–1515. ISSN: 1539-0756. DOI: 10.1103/revmodphys.80.1455.
- [22] A. Andronic et al. “Decoding the phase structure of QCD via particle production at high energy”. In: *Nat. Phys.* 561.7723 (Sept. 2017), pp. 321–330. DOI: 10.1038/s41586-018-0491-6. arXiv: 1710.09425 [nucl-th].
- [23] W. Busza, K. Rajagopal, and W. van der Schee. “Heavy Ion Collisions: The Big Picture, and the Big Questions”. In: *Annual Review of Nuclear and Particle Science* 68.1 (Oct. 2018), pp. 339–376. ISSN: 1545-4134. DOI: 10.1146/annurev-nucl-101917-020852. arXiv: 1802.04801 [hep-ph].
- [24] J. Jankowski and M. Spaliński. “Hydrodynamic attractors in ultrarelativistic nuclear collisions”. In: *Prog. Part. Nucl. Phys.* 132 (Sept. 2023), p. 104048. ISSN: 0146-6410. DOI: 10.1016/j.ppnp.2023.104048.
- [25] J. D. Bjorken. “Highly relativistic nucleus-nucleus collisions: The central rapidity region”. In: *Phys. Rev. C* 27.1 (Jan. 1983), pp. 140–151. ISSN: 0556-2821. DOI: 10.1103/physrevd.27.140.
- [26] E. Iancu and R. Venugopalan. “The Color Glass Condensate and High Energy Scattering in QCD”. In: (Mar. 24, 2003), pp. 249–363. DOI: 10.1142/9789812795533_0005. arXiv: hep-ph/0303204 [hep-ph].
- [27] H. Weigert. “A compact introduction to evolution at small x and the Color Glass Condensate”. In: *Nucl. Phys. A* 783.1–4 (Feb. 2007), pp. 165–172. ISSN: 0375-9474. DOI: 10.1016/j.nuclphysa.2006.11.032.
- [28] ALICE Collaboration. “Direct photon production in Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV”. In: *Phys. Lett. B* 754 (Mar. 2015), pp. 235–248. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2016.01.020. arXiv: 1509.07324 [nucl-ex].
- [29] ALICE Collaboration. “The ALICE experiment: a journey through QCD”. In: *Eur. Phys. J. C* 84.8 (Nov. 8, 2022). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-024-12935-y. arXiv: 2211.04384 [nucl-ex].
- [30] M. L. Miller et al. “Glauber Modeling in High Energy Nuclear Collisions”. In: *Ann. Rev. Nucl. Part. Sci.* 57.1 (Jan. 17, 2007), pp. 205–243. ISSN: 1545-4134. DOI: 10.1146/annurev.nucl.57.090506.123020. arXiv: nucl-ex/0701025 [nucl-ex].

- [31] **BRAHMS** Collaboration. “Quark Gluon Plasma an Color Glass Condensate at RHIC? The perspective from the BRAHMS experiment”. In: *Nucl. Phys. A* 757.1–2 (Aug. 2004), pp. 1–27. doi: 10.1016/j.nuclphysa.2005.02.130. arXiv: nucl-ex/0410020 [nucl-ex].
- [32] **PHENIX** Collaboration. “Formation of dense partonic matter in relativistic nucleus-nucleus collisions at RHIC: Experimental evaluation by the PHENIX collaboration”. In: *Nucl. Phys. A* 757.1–2 (Aug. 2004), pp. 184–283. doi: 10.1016/j.nuclphysa.2005.03.086. arXiv: nucl-ex/0410003 [nucl-ex].
- [33] **STAR** Collaboration. “Experimental and theoretical challenges in the search for the quark–gluon plasma: The STAR Collaboration’s critical assessment of the evidence from RHIC collisions”. In: *Nucl. Phys. A* 757.1–2 (Aug. 2005), pp. 102–183. issn: 0375-9474. doi: 10.1016/j.nuclphysa.2005.03.085.
- [34] **PHOBOS** Collaboration. “The PHOBOS Perspective on Discoveries at RHIC”. In: *Nucl. Phys. A* 757.1–2 (Aug. 2004), pp. 28–101. doi: 10.1016/j.nuclphysa.2005.03.084. arXiv: nucl-ex/0410022 [nucl-ex].
- [35] **CMS** Collaboration. “Study of Jet Quenching with Z+jet Correlations in Pb-Pb and pp Collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV”. In: *Phys. Rev. Lett.* 119.8 (Aug. 2017), p. 082301. issn: 1079-7114. doi: 10.1103/physrevlett.119.082301.
- [36] R. Pasechnik and M. Šumbera. “Phenomenological Review on Quark–Gluon Plasma: Concepts vs. Observations”. In: *Universe* 3.1 (Jan. 2017), p. 7. issn: 2218-1997. doi: 10.3390/universe3010007.
- [37] **ALICE** Collaboration. “Elliptic Flow of Charged Particles in Pb–Pb Collisions at $\sqrt{s_{\text{NN}}} = 2.76$ TeV”. In: *Phys. Rev. Lett.* 105.25 (Dec. 2010), p. 252302. issn: 1079-7114. doi: 10.1103/physrevlett.105.252302.
- [38] J. Rafelski and B. Müller. “Strangeness Production in the Quark-Gluon Plasma”. In: *Phys. Rev. Lett.* 48.16 (Apr. 1982), pp. 1066–1069. issn: 0031-9007. doi: 10.1103/physrevlett.48.1066.
- [39] A. Tounsi and K. Redlich. “Strangeness Enhancement and Canonical Suppression”. In: (Nov. 2001). doi: 10.48550/ARXIV.HEP-PH/0111159. arXiv: hep-ph/0111159 [hep-ph].
- [40] **WA97** Collaboration. “Strangeness enhancement at mid-rapidity in Pb–Pb collisions at 158 A GeV/c”. In: *Phys. Lett. B* 449.3–4 (Mar. 1999), pp. 401–406. issn: 0370-2693. doi: 10.1016/s0370-2693(99)00140-9.
- [41] **STAR** Collaboration. “Strangeness Enhancement in Cu–Cu and Au–Au Collisions at $\sqrt{s_{\text{NN}}} = 200$ GeV”. In: *Phys. Rev. Lett.* 108.7 (Feb. 2012), p. 072301. issn: 1079-7114. doi: 10.1103/physrevlett.108.072301.
- [42] **ALICE** Collaboration. “Multi-strange baryon production at mid-rapidity in Pb–Pb collisions at $\sqrt{s_{\text{NN}}} = 2.76$ TeV”. In: *Phys. Lett. B* 728 (July 21, 2013), pp. 216–227. issn: 0370-2693. doi: 10.1016/j.physletb.2013.11.048. arXiv: 1307.5543 [nucl-ex].
- [43] **ALICE** Collaboration. “Enhanced production of multi-strange hadrons in high multiplicity proton-proton collisions”. In: *Nat. Phys.* 13.6 (Apr. 2016), pp. 535–539. issn: 1745-2481. doi: 10.1038/nphys4111. arXiv: 1606.07424 [nucl-ex].
- [44] **ALICE** Collaboration. “Multiplicity Dependence of Pion, Kaon, Proton and Lambda Production in p–Pb Collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV”. In: *Phys. Lett. B* 728 (July 25, 2013), pp. 25–38. issn: 0370-2693. doi: 10.1016/j.physletb.2013.11.020. arXiv: 1307.6796 [nucl-ex].
- [45] W. Florkowski. *Phenomenology of Ultra-Relativistic Heavy-Ion Collisions*. Mar. 2010. isbn: 978-981-4280-66-2.

- [46] **CMS** Collaboration. <https://www.int.washington.edu/programs-and-workshops/21r-2b>.
- [47] M. Cacciari, G. P. Salam, and G. Soyez. “The anti- k_T jet clustering algorithm”. In: *J. High Energy Phys.* 2008.04 (Apr. 2008), p. 063. doi: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].
- [48] **ALICE** Collaboration. “Suppression of charged particle production at large transverse momentum in central Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV”. In: *Phys. Lett. B* 696.1–2 (Jan. 2010), pp. 30–39. doi: 10.1016/j.physletb.2010.12.020. arXiv: 1012.1004 [nucl-ex].
- [49] **CMS** Collaboration. “Study of high- p_T charged particle suppression in PbPb compared to pp collisions at $\sqrt{s_{NN}} = 2.76$ TeV”. In: *Eur. Phys. J. C* 72.3 (Feb. 12, 2012). doi: 10.1140/epjc/s10052-012-1945-x. arXiv: 1202.2554 [nucl-ex].
- [50] J. G. Milhano. “Theoretical overview of jet quenching”. In: *Nucl. Phys. A* 904–905 (May 2013), pp. 122c–129c. doi: 10.1016/j.nuclphysa.2013.01.053.
- [51] **ATLAS** Collaboration. “Observation of a Centrality-Dependent Dijet Asymmetry in Lead-Lead Collisions at $\sqrt{s_{NN}} = 2.76$ TeV with the ATLAS Detector at the LHC”. In: *Phys. Rev. Lett.* 105.25 (Dec. 2010), p. 252303. issn: 1079-7114. doi: 10.1103/physrevlett.105.252303.
- [52] **ALICE** Collaboration. “Measurement of the radius dependence of charged-particle jet suppression in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *Phys. Lett. B* 849 (Mar. 1, 2023), p. 138412. issn: 0370-2693. doi: 10.1016/j.physletb.2023.138412. arXiv: 2303.00592 [nucl-ex].
- [53] **ALICE** Collaboration. “Charged jet cross section and fragmentation in proton-proton collisions at $\sqrt{s} = 7$ TeV”. In: *Phys. Rev. D* 99.1 (Jan. 2018), p. 012016. issn: 2470-0029. doi: 10.1103/physrevd.99.012016. arXiv: 1809.03232 [nucl-ex].
- [54] **ALICE** Collaboration. “Measurements of jet quenching using semi-inclusive hadron+jet distributions in pp and central Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *Phys. Rev. C* 110.1 (July 2024), p. 014906. issn: 2469-9993. doi: 10.1103/physrevc.110.014906.
- [55] **NA50** Collaboration. “Charmonium production in Pb–Pb interactions at 158 GeV/c per nucleon”. In: *Nucl. Phys. A* 638.1–2 (Aug. 1998), pp. 261c–278c. issn: 0375-9474. doi: 10.1016/s0375-9474(98)00360-1.
- [56] **STAR** Collaboration. “Measurement of inclusive J/ψ suppression in Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV through the dimuon channel at STAR”. In: *Phys. Lett. B* 797 (Oct. 2019), p. 134917. issn: 0370-2693. doi: 10.1016/j.physletb.2019.134917.
- [57] **PHENIX** Collaboration. “ J/ψ Production vs Centrality, Transverse Momentum, and Rapidity in Au+Au Collisions at $\sqrt{s_{NN}} = 200$ GeV”. In: *Phys. Rev. Lett.* 98.232301, 2007 98.23 (June 2006), p. 232301. issn: 1079-7114. doi: 10.1103/physrevlett.98.232301. arXiv: nucl-ex/0611020 [nucl-ex].
- [58] T. Matsui and H. Satz. “ J/ψ suppression by quark-gluon plasma formation”. In: *Phys. Lett. B* 178.4 (Oct. 1986), pp. 416–422. issn: 0370-2693. doi: 10.1016/0370-2693(86)91404-8.
- [59] A. Mocsy. “Potential Models for Quarkonia”. In: *Eur. Phys. J. C* 61.4 (Jan. 2008), pp. 705–710. issn: 1434-6052. doi: 10.1140/epjc/s10052-008-0847-4. arXiv: 0811.0337 [hep-ph].
- [60] **ALICE** Collaboration. “Prompt and non-prompt J/ψ production at midrapidity in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *J. High Energy Phys.* 2024.2 (Feb. 2024). doi: 10.1007/jhep02(2024)066.

- [61] A. Andronic et al. “Transverse momentum distributions of charmonium states with the statistical hadronization model”. In: *Phys. Rev. Lett. B* 797 (Jan. 26, 2019), p. 134836. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2019.134836. arXiv: 1901.09200 [nucl-th].
- [62] K. Zhou et al. “Medium Effects on Charmonium Production at ultrarelativistic energies available at the CERN Large Haron Collider”. In: *Phys.Rev.C* 89.5 (Jan. 23, 2014), p. 054911. ISSN: 1089-490X. DOI: 10.1103/physrevc.89.054911. arXiv: 1401.5845 [nucl-th].
- [63] ALICE Collaboration. “Measurements of inclusive J/ψ production at midrapidity and forward rapidity in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *Phys. Lett. B* 849 (Feb. 2023), p. 138451. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2024.138451. arXiv: 2303.13361 [nucl-ex].
- [64] J. Wang. *boundino/hinHFplot*. Version latest. Apr. 2025. <https://github.com/boundino/hinHFplot>.
- [65] N. Armesto. “Small collision systems: Theory overview on cold nuclear matter effects”. In: *EPJ Web Conf.* 171 (2018). Ed. by A. Mischke and P. Kuijter, p. 11001. ISSN: 2100-014X. DOI: 10.1051/epjconf/201817111001.
- [66] ALICE Collaboration. “Transverse momentum spectra and nuclear modification factors of charged particles in pp, p–Pb and Pb–Pb collisions at the LHC”. In: *J. High Energy Phys.* 2018.11 (Nov. 2018). ISSN: 1029-8479. DOI: 10.1007/jhep11(2018)013. arXiv: 1802.09145 [nucl-ex].
- [67] CMS Collaboration. “Charged-particle nuclear modification factors in PbPb and p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *J. High Energy Phys.* 2017.4 (Apr. 2017). ISSN: 1029-8479. DOI: 10.1007/jhep04(2017)039.
- [68] CMS Collaboration. “Observation of Long-Range Near-Side Angular Correlations in Proton-Proton Collisions at the LHC”. In: *J. High Energy Phys.* 2010.9 (Sept. 21, 2010). ISSN: 1029-8479. DOI: 10.1007/jhep09(2010)091. arXiv: 1009.4122 [hep-ex].
- [69] CMS Collaboration. “Multiplicity and transverse-momentum dependence of two- and four-particle correlations in p–Pb and Pb–Pb collisions”. In: *Phys. Lett. B* 724.4–5 (May 3, 2013), pp. 213–240. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2013.06.028. arXiv: 1305.0609 [nucl-ex].
- [70] T. Sjöstrand et al. “An Introduction to PYTHIA 8.2”. In: *Computer Physics Communications* 191 (Oct. 11, 2014), pp. 159–177. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].
- [71] C. Bierlich and J. R. Christiansen. “Effects of Colour Reconnection on Hadron Flavour Observables”. In: *Phys. Rev. D* 92.9 (Nov. 2015), p. 094010. ISSN: 1550-2368. DOI: 10.1103/physrevd.92.094010. arXiv: 1507.02091 [hep-ph].
- [72] T. Pierog et al. “EPOS LHC : test of collective hadronization with LHC data”. In: *Phys. Rev. C* 92.3 (Sept. 2013), p. 034906. ISSN: 1089-490X. DOI: 10.1103/physrevc.92.034906. arXiv: 1306.0121 [hep-ph].
- [73] ALICE Collaboration. “Constraints on jet quenching in p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV measured by the event-activity dependence of semi-inclusive hadron-jet distributions”. In: *Phys. Lett. B* 783 (Dec. 15, 2017), pp. 95–113. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2018.05.059. arXiv: 1712.05603 [nucl-ex].
- [74] ALICE Collaboration. “Search for jet quenching effects in high-multiplicity pp collisions at $\sqrt{s} = 13$ TeV via di-jet acoplanarity”. In: *J. High Energy Phys.* 2024.5 (Aug. 30, 2023). ISSN: 1029-8479. DOI: 10.1007/jhep05(2024)229. arXiv: 2309.03788 [hep-ex].

- [75] **ALICE** Collaboration. “Measurement of nuclear effects on $\psi(2S)$ production in p–Pb collisions at $\sqrt{s_{NN}} = 8.16$ TeV”. In: *J. High Energy Phys.* 2020.7 (July 2020). issn: 1029-8479. doi: 10.1007/jhep07(2020)237.
- [76] **ALICE** Collaboration. “Forward rapidity J/ψ production as a function of charged-particle multiplicity in pp collisions at $\sqrt{s} = 5.02$ and 13 TeV”. In: *J. High Energy Phys.* 2022.6 (Dec. 17, 2021). issn: 1029-8479. doi: 10.1007/jhep06(2022)015. arXiv: 2112.09433 [nucl-ex].
- [77] **ALICE** Collaboration. “Studies of J/ψ production at forward rapidity in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *J. High Energy Phys.* 2020.2 (Sept. 7, 2019). issn: 1029-8479. doi: 10.1007/jhep02(2020)041. arXiv: 1909.03158 [nucl-ex].
- [78] **ALICE** Collaboration. “Dielectron production in central Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: (Aug. 2023). arXiv: 2308.16704 [nucl-ex].
- [79] **ALICE** Collaboration. “Dielectron production in proton-proton and proton-lead collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *Phys. Rev. C* 102.5 (May 25, 2020), p. 055204. issn: 2469-9993. doi: 10.1103/physrevc.102.055204. 2005.11995.
- [80] N. Brambilla, S. Eidelman, B. K. Heltsley, et al. “Heavy Quarkonium: Progress, Puzzles, and Opportunities”. In: *Eur. Phys. J. C* 71 (2011), p. 1534. doi: 10.1140/epjc/s10052-010-1534-9. arXiv: 1010.5827 [hep-ph].
- [81] J. C. Collins, D. E. Soper, and G. Sterman. “Factorization of Hard Processes in QCD”. In: *Adv. Ser. Direct. High Energy Phys.* 5:1-91, 1988 (Sept. 2004). doi: 10.48550/ARXIV.HEP-PH/0409313. arXiv: hep-ph/0409313 [hep-ph].
- [82] J. D. Bjorken and E. A. Paschos. “Inelastic Electron-Proton and γ -Proton Scattering and the Structure of the Nucleon”. In: *Phys. Rev.* 185.5 (Sept. 1969), pp. 1975–1982. issn: 0031-899X. doi: 10.1103/physrev.185.1975.
- [83] J. I. Friedman and H. W. Kendall. “Deep Inelastic Electron Scattering”. In: *Annual Review of Nuclear Science* 22.1 (Dec. 1972), pp. 203–254. issn: 0066-4243. doi: 10.1146/annurev.ns.22.120172.001223.
- [84] J. F. Owens and W. Tung. “Parton Distribution Functions of Hadrons”. In: *Annual Review of Nuclear and Particle Science* 42.1 (Dec. 1992), pp. 291–332. issn: 1545-4134. doi: 10.1146/annurev.ns.42.120192.001451.
- [85] R. P. Feynman. “Very High-Energy Collisions of Hadrons”. In: *Phys. Rev. Lett.* 23.24 (Dec. 1969), pp. 1415–1417. issn: 0031-9007. doi: 10.1103/physrevlett.23.1415.
- [86] M. Gell-Mann. “A schematic model of baryons and mesons”. In: *Phys. Lett.* 8.3 (Feb. 1964), pp. 214–215. issn: 0031-9163. doi: 10.1016/s0031-9163(64)92001-3.
- [87] G. Zweig. *An SU_3 model for strong interaction symmetry and its breaking*. en. 2002. doi: 10.17181/CERN-TH-412.
- [88] D. J. Gross and F. Wilczek. “Ultraviolet Behavior of Non-Abelian Gauge Theories”. In: *Phys. Rev. Lett.* 30.26 (June 1973), pp. 1343–1346. issn: 0031-9007. doi: 10.1103/physrevlett.30.1343.
- [89] H. D. Politzer. “Reliable Perturbative Results for Strong Interactions?” In: *Phys. Rev. Lett.* 30.26 (June 1973), pp. 1346–1349. issn: 0031-9007. doi: 10.1103/physrevlett.30.1346.
- [90] J. McGowan et al. “Approximate N^3 LO Parton Distribution Functions with Theoretical Uncertainties: MSHT20a N^3 LO PDFs”. In: *Eur. Phys. J. C* 83.3 (July 11, 2022). issn: 1434-6052. doi: 10.1140/epjc/s10052-023-11236-0. arXiv: 2207.04739 [hep-ph].

- [91] S. Dulat et al. "New parton distribution functions from a global analysis of quantum chromodynamics". In: *Phys. Rev. D* 93.3 (June 24, 2015), p. 033006. ISSN: 2470-0029. DOI: 10.1103/physrevd.93.033006. arXiv: 1506.07443 [hep-ph].
- [92] L. A. Harland-Lang et al. "Parton distributions in the LHC era: MMHT 2014 PDFs". In: *Eur. Phys. J. C* 75.5 (Dec. 12, 2014). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3397-6. arXiv: 1412.3989 [hep-ph].
- [93] R. D. Ball et al. "Parton distributions for the LHC run II". In: *J. High Energy Phys.* 2015.4 (Apr. 2015). ISSN: 1029-8479. DOI: 10.1007/jhep04(2015)040.
- [94] Y. L. Dokshitzer. "Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics." In: *Sov. Phys. JETP* 46 (1977), pp. 641–653.
- [95] V. N. Gribov and L. N. Lipatov. "Deep inelastic $e p$ scattering in perturbation theory". In: *Sov. J. Nucl. Phys.* 15 (1972), pp. 438–450.
- [96] G. Altarelli and G. Parisi. "Asymptotic freedom in parton language". In: *Nucl. Phys. B* 126.2 (Aug. 1977), pp. 298–318. ISSN: 0550-3213. DOI: 10.1016/0550-3213(77)90384-4.
- [97] L. N. Lipatov. "Reggeization of the Vector Meson and the Vacuum Singularity in Non-abelian Gauge Theories". In: *Sov. J. Nucl. Phys.* 23 (1976), pp. 338–345.
- [98] V.S. Fadin, E.A. Kuraev, and L.N. Lipatov. "On the pomeranchuk singularity in asymptotically free theories". In: *Phys. Lett. B* 60.1 (Dec. 1975), pp. 50–52. ISSN: 0370-2693. DOI: 10.1016/0370-2693(75)90524-9.
- [99] E. A. Kuraev, L. N. Lipatov, and Victor S. Fadin. "Multi - Reggeon Processes in the Yang-Mills Theory". In: *Sov. Phys. JETP* 44 (1976), pp. 443–450.
- [100] I. I. Balitsky and L. N. Lipatov. "The Pomeranchuk Singularity in Quantum Chromodynamics". In: *Sov. J. Nucl. Phys.* 28 (1978), pp. 822–829.
- [101] G. Chachamis. "BFKL phenomenology". In: *New Trends in High-Energy Physics and QCD*. 2016, 2016, pp. 4–24. arXiv: 1512.04430 [hep-ph].
- [102] F. Gelis et al. "The Color Glass Condensate". In: *Ann.Rev.Nucl.Part.Sci.* 60:463-489,2010 60.1 (Feb. 1, 2010), pp. 463–489. ISSN: 1545-4134. DOI: 10.1146/annurev.nucl.010909.083629. arXiv: 1002.0333 [hep-ph].
- [103] H. Weigert. "Unitarity at small Bjorken x ". In: *Nucl. Phys. A* 703.3–4 (Apr. 5, 2000), pp. 823–860. ISSN: 0375-9474. DOI: 10.1016/S0375-9474(01)01668-2. arXiv: hep-ph/0004044 [hep-ph].
- [104] J. Jalilian-Marian et al. "Wilson renormalization group for low x physics: Towards the high density regime". In: *Phys. Rev. D* 59.1 (Nov. 1998), p. 014014. ISSN: 1089-4918. DOI: 10.1103/physrevd.59.014014.
- [105] E. Ferreiro et al. "Nonlinear Gluon Evolution in the Color Glass Condensate: II". In: *Nucl. Phys. A* 703.1–2 (Sept. 13, 2001), pp. 489–538. ISSN: 0375-9474. DOI: 10.1016/S0375-9474(01)01329-X. arXiv: hep-ph/0109115 [hep-ph].
- [106] M. Cacciari, S. Frixione, and P. Nason. "The p_T Spectrum in Heavy-Flavour Photoproduction". In: *J. High Energy Phys.* 2001.03 (Feb. 12, 2001), pp. 006–006. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2001/03/006. arXiv: hep-ph/0102134 [hep-ph].
- [107] M. Cacciari, M. Greco, and P. Nason. "The p_T Spectrum in Heavy-Flavour Hadroproduction". In: *J. High Energy Phys.* 1998.05 (Mar. 20, 1998), pp. 007–007. ISSN: 1029-8479. DOI: 10.1088/1126-6708/1998/05/007. arXiv: hep-ph/9803400 [hep-ph].
- [108] M. Benzke et al. "Prompt neutrinos from atmospheric charm in the general-mass variable-flavor-number scheme". In: *J. High Energy Phys.* 2017.12 (May 29, 2017). ISSN: 1029-8479. DOI: 10.1007/jhep12(2017)021. arXiv: 1705.10386 [hep-ph].

- [109] G. Kramer and H. Spiesberger. “Study of heavy meson production in p–Pb collisions at $\sqrt{s}=5.02$ TeV in the general-mass variable-flavour-number scheme”. In: *Nucl. Phys. B* 925 (2017), pp. 415–430. doi: 10.1016/j.nuclphysb.2017.10.016. arXiv: 1703.04754 [hep-ph].
- [110] B. A. Kniehl et al. “Inclusive D^{*+-} Production in $p\bar{p}$ Collisions with Massive Charm Quarks”. In: *Phys. Rev. D* 71.1 (Oct. 21, 2004), p. 014018. issn: 1550-2368. doi: 10.1103/physrevd.71.014018. arXiv: hep-ph/0410289 [hep-ph].
- [111] B. A. Kniehl et al. “Collinear Subtractions in Hadroproduction of Heavy Quarks”. In: *Eur. Phys. J. C* 41.2 (Feb. 21, 2005), pp. 199–212. issn: 1434-6052. doi: 10.1140/epjc/s2005-02200-7. arXiv: hep-ph/0502194 [hep-ph].
- [112] B. A. Kniehl et al. “Inclusive Charmed-Meson Production at the CERN LHC”. In: *Eur. Phys. J. C* 72.7 (Feb. 2, 2012). issn: 1434-6052. doi: 10.1140/epjc/s10052-012-2082-2. arXiv: 1202.0439 [hep-ph].
- [113] I. Helenius and H. Paukkunen. “Revisiting the D-meson hadroproduction in general-mass variable flavour number scheme”. In: *J. High Energy Phys.* 2018.5 (Apr. 10, 2018). issn: 1029-8479. doi: 10.1007/jhep05(2018)196. arXiv: 1804.03557 [hep-ph].
- [114] S. Catani et al. “Bottom-quark production at hadron colliders: fully differential predictions in NNLO QCD”. In: *J. High Energy Phys.* 2021.3 (Oct. 22, 2020). issn: 1029-8479. doi: 10.1007/jhep03(2021)029. arXiv: 2010.11906 [hep-ph].
- [115] ALICE Collaboration. “Measurement of beauty-quark production in pp collisions at $\sqrt{s} = 13$ TeV via non-prompt D mesons”. In: *J. High Energy Phys.* 2024.10 (Feb. 26, 2024). issn: 1029-8479. doi: 10.1007/jhep10(2024)110. arXiv: 2402.16417 [hep-ex].
- [116] PHENIX Collaboration. “Measurement of Bottom versus Charm as a Function of Transverse Momentum with Electron-Hadron Correlations in p+p Collisions at $\sqrt{s} = 200$ GeV”. In: *Phys.Rev.Lett.* 103:082002,2009 103.8 (Mar. 27, 2009), p. 082002. issn: 1079-7114. doi: 10.1103/physrevlett.103.082002. arXiv: 0903.4851 [hep-ex].
- [117] UA1 Collaboration. “Beauty production at the CERN $p\bar{p}$ collider”. In: *Phys. Lett. B* 256.1 (Feb. 1991), pp. 121–128. issn: 0370-2693. doi: 10.1016/0370-2693(91)90228-i.
- [118] ALICE Collaboration. “Measurement of prompt J/ψ and beauty hadron production cross sections at mid-rapidity in pp collisions at $\sqrt{s} = 7$ TeV”. In: *J. High Energy Phys.* 2012.11 (May 26, 2012). issn: 1029-8479. doi: 10.1007/jhep11(2012)065. arXiv: 1205.5880 [hep-ex].
- [119] ALICE Collaboration. “Dielectron production in proton-proton collisions at $\sqrt{s} = 7$ TeV”. In: *J. High Energy Phys.* 2018.9 (May 11, 2018). issn: 1029-8479. doi: 10.1007/jhep09(2018)064. arXiv: 1805.04391 [hep-ex].
- [120] ALICE Collaboration. “Measurement of electrons from beauty hadron decays in pp collisions at $\sqrt{s} = 7$ TeV”. In: *Phys. Lett. B* 721.1–3 (Aug. 9, 2012), pp. 13–23. issn: 0370-2693. doi: 10.1016/j.physletb.2013.01.069. arXiv: 1208.1902 [hep-ex].
- [121] ALICE Collaboration. “Prompt and non-prompt J/ψ production cross sections at midrapidity in proton-proton collisions at $\sqrt{s} = 5.02$ and 13 TeV”. In: *J. High Energy Phys.* 2022.3 (Aug. 5, 2021). issn: 1029-8479. doi: 10.1007/jhep03(2022)190. arXiv: 2108.02523 [nucl-ex].
- [122] ALICE Collaboration. “Dielectron and heavy-quark production in inelastic and high-multiplicity proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Phys. Lett. B* 788 (May 11, 2018), pp. 505–518. issn: 0370-2693. doi: 10.1016/j.physletb.2018.11.009. arXiv: 1805.04407 [hep-ex].

- [123] T. Kneesch et al. “Charmed-Meson Fragmentation Functions with Finite-Mass Corrections”. In: *Nucl. Phys. B* 799.1–2 (Dec. 4, 2007), pp. 34–59. ISSN: 0550-3213. DOI: 10.1016/j.nuclphysb.2008.02.015. arXiv: 0712.0481 [hep-ph].
- [124] B. A. Kniehl and G. Kramer. “Charmed-Hadron Fragmentation Functions from CERN LEP1 Revisited”. In: *Phys. Rev. D* 74.3 (July 27, 2006), p. 037502. ISSN: 1550-2368. DOI: 10.1103/physrevd.74.037502. arXiv: hep-ph/0607306 [hep-ph].
- [125] ALICE Collaboration. “Charm production and fragmentation fractions at midrapidity in pp collisions at $\sqrt{s} = 13$ TeV”. In: *J. High Energy Phys.* 2023.12 (Aug. 9, 2023). ISSN: 1029-8479. DOI: 10.1007/jhep12(2023)086. arXiv: 2308.04877 [hep-ex].
- [126] Belle Collaboration. “Production cross sections of hyperons and charmed baryons from e^+e^- annihilation near $\sqrt{s} = 10.52$ GeV”. In: *Phys. Rev. D* 97.7 (June 21, 2017), p. 072005. ISSN: 2470-0029. DOI: 10.1103/physrevd.97.072005. arXiv: 1706.06791 [hep-ex].
- [127] V. Minissale, S. Plumari, and V. Greco. “Charm Hadrons in pp collisions at LHC energy within a Coalescence plus Fragmentation approach”. In: *Phys. Lett. B* 821 (Dec. 22, 2020), p. 136622. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2021.136622. arXiv: 2012.12001 [hep-ph].
- [128] A. Beraudo et al. “Heavy-flavor transport and hadronization in pp collisions”. In: *Phys. Rev. D* 109.1 (Jan. 2024), p. 1011501. ISSN: 2470-0029. DOI: 10.1103/physrevd.109.1011501.
- [129] P. Braun-Munzinger, K. Redlich, and J. Stachel. “Particle Production in Heavy Ion Collisions”. In: (Apr. 3, 2003), pp. 491–599. DOI: 10.1142/9789812795533_0008. arXiv: nucl-th/0304013 [nucl-th].
- [130] D. Ebert, R. N. Faustov, and V. O. Galkin. “Spectroscopy and Regge trajectories of heavy baryons in the relativistic quark-diquark picture”. In: *Phys. Rev. D* 84.1 (May 3, 2011), p. 014025. ISSN: 1550-2368. DOI: 10.1103/physrevd.84.014025. arXiv: 1105.0583 [hep-ph].
- [131] B. Andersson. *The Lund Model*. Cambridge University Press, Oct. 1998. ISBN: 978-0-511-52436-3. DOI: 10.1017/cbo9780511524363.
- [132] S. Ferreres-Solé and T. Sjöstrand. “The Space-Time Structure of Hadronization in the Lund Model”. In: *Eur. Phys. J. C* 78.11 (Aug. 14, 2018). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-018-6459-8. arXiv: 1808.04619 [hep-ph].
- [133] P. Skands, S. Carrazza, and J. Rojo. “Tuning PYTHIA 8.1: the Monash 2013 Tune”. In: *Eur. Phys. J. C* 74.8 (Apr. 22, 2014). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-014-3024-y. arXiv: 1404.5630 [hep-ph].
- [134] J. R. Christiansen and P. Skands. “String formation beyond leading colour”. In: *J. High Energy Phys.* 2015.8 (Aug. 2015). ISSN: 1029-8479. DOI: 10.1007/jhep08(2015)003.
- [135] L3 Collaboration. “Search for Colour Reconnection Effects in $e^+e^- \rightarrow W^+W^- \rightarrow$ hadrons through Particle-Flow Studies at LEP”. In: *Phys. Lett. B* 561.3–4 (Mar. 27, 2003), pp. 202–212. ISSN: 0370-2693. DOI: 10.1016/s0370-2693(03)00490-8. arXiv: hep-ex/0303042 [hep-ex].
- [136] OPAL Collaboration. “Tests of model of color reconnection and a search for glueballs using gluon jets with a rapidity gap”. In: *Eur. Phys. J. C* 35.3 (June 6, 2003), pp. 293–312. ISSN: 1434-6052. DOI: 10.1140/epjc/s2004-01809-2. arXiv: hep-ex/0306021 [hep-ex].
- [137] H. J. Drescher et al. “Parton-Based Gribov-Regge Theory”. In: *Phys. Rept.* 350.2–4 (July 18, 2000), pp. 93–289. ISSN: 0370-1573. DOI: 10.1016/s0370-1573(00)00122-8. arXiv: hep-ph/0007198 [hep-ph].

- [138] K. Werner and B. Guiot. “Perturbative QCD concerning light and heavy flavor in the EPOS4 framework”. In: *Phys. Rev. C* 108.3 (2023), p. 034904. doi: 10.1103/PhysRevC.108.034904. arXiv: 2306.02396 [hep-ph].
- [139] K. Werner. “Revealing a deep connection between factorization and saturation: New insight into modeling high-energy proton-proton and nucleus-nucleus scattering in the EPOS4 framework”. In: *Phys. Rev. C* 108.6 (2023), p. 064903. doi: 10.1103/PhysRevC.108.064903. arXiv: 2301.12517 [hep-ph].
- [140] K. Werner. “Parallel scattering, saturation, and generalized Abramovskii-Gribov-Kancheli (AGK) theorem in the EPOS4 framework, with applications for heavy-ion collisions at sNN of 5.02 TeV and 200 GeV”. In: *Phys. Rev. C* 109.3 (2024), p. 034918. doi: 10.1103/PhysRevC.109.034918. arXiv: 2310.09380 [hep-ph].
- [141] K. Werner. “Core-corona procedure and microcanonical hadronization to understand strangeness enhancement in proton-proton and heavy ion collisions in the EPOS4 framework”. In: *Phys. Rev. C* 109.1 (June 17, 2023), p. 014910. issn: 2469-9993. doi: 10.1103/physrevc.109.014910. arXiv: 2306.10277 [hep-ph].
- [142] ALICE Collaboration. “Preliminary Physics Summary: measurement of B^0 -meson production cross section in proton-proton collisions at $\sqrt{s} = 13.6$ TeV”. In: (2025).
- [143] S. Frixione et al. “The MC@NLO 4.0 Event Generator”. In: (Oct. 2010). arXiv: 1010.0819 [hep-ph].
- [144] A. Banfi et al. “A POWHEG generator for deep inelastic scattering”. In: *J. High Energy Phys.* 2024.2 (Sept. 5, 2023). issn: 1029-8479. doi: 10.1007/jhep02(2024)023. arXiv: 2309.02127 [hep-ph].
- [145] R. Maciula and A. Szczurek. “Production of Λ_c baryons at the LHC within the k_T -factorization approach and independent parton fragmentation picture”. In: *Phys. Rev. D* 98.1 (Mar. 15, 2018), p. 014016. issn: 2470-0029. doi: 10.1103/physrevd.98.014016. arXiv: 1803.05807 [hep-ph].
- [146] R. Maciula and A. Szczurek. “Consistent treatment of charm production in higher-orders at tree-level within k_T -factorization approach”. In: *Phys. Rev. D* 100.5 (May 16, 2019), p. 054001. issn: 2470-0029. doi: 10.1103/physrevd.100.054001. arXiv: 1905.06697 [hep-ph].
- [147] F. Takagi. “Multiple Production of Quark Jets off Nuclei”. In: *Phys. Rev. Lett.* 43.18 (Oct. 1979), pp. 1296–1299. issn: 0031-9007. doi: 10.1103/physrevlett.43.1296.
- [148] C. Goebel, D. M. Scott, and F. Halzen. “Double Drell-Yan annihilations in hadron collisions: Novel tests of the constituent picture”. In: *Phys. Rev. D* 22.11 (Dec. 1980), pp. 2789–2792. issn: 0556-2821. doi: 10.1103/physrevd.22.2789.
- [149] N. Paver and D. Treleani. “Multi-quark scattering and large- p_T jet production in hadronic collisions”. In: *Il Nuovo Cimento A* 70.3 (Aug. 1982), pp. 215–228. issn: 1826-9869. doi: 10.1007/bf02814035.
- [150] N. Paver and D. Treleani. “Multiple parton interactions and multijet events at collider and Tevatron energies”. In: *Phys. Lett. B* 146.3–4 (Oct. 1984), pp. 252–256. issn: 0370-2693. doi: 10.1016/0370-2693(84)91029-3.
- [151] CMS Collaboration. “Measurement of double-parton scattering in inclusive production of four jets with low transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *J. High Energy Phys.* 2022.1 (Sept. 28, 2021). issn: 1029-8479. doi: 10.1007/jhep01(2022)177. arXiv: 2109.13822 [hep-ex].
- [152] ATLAS Collaboration. “Measurement of hard double-parton interactions in $W \rightarrow l\nu^+$ 2 jet events at $\sqrt{s} = 7$ TeV with the ATLAS detector”. In: *New J. Phys.* 15.3 (Jan. 29,

- 2013), p. 033038. ISSN: 1367-2630. DOI: 10 . 1088 / 1367 - 2630 / 15 / 3 / 033038. arXiv: 1301 . 6872 [hep-ex].
- [153] T. Åkesson et al. "Double parton scattering in pp collisions at $\sqrt{s} = 63$ GeV". In: *Zeitschrift für Physik C Particles and Fields* 34.2 (June 1987), pp. 163–174. ISSN: 1434-6052. DOI: 10 . 1007/bf01566757.
- [154] D0 Collaboration. "Double parton interactions in photon+3 jet events in $p\bar{p}$ collisions $\sqrt{s} = 1.96$ TeV". In: *Phys. Rev. D* 81.5 (Dec. 27, 2009), p. 052012. ISSN: 1550-2368. DOI: 10.1103/physrevd.81.052012. arXiv: 0912.5104 [hep-ex].
- [155] A. Szczurek and R. Maciula. "Production of one and two $c\bar{c}$ pairs at LHC". In: (Dec. 2012), pp. 255–259. ISSN: 0094-243X. DOI: 10 . 1063 / 1 . 4802161. arXiv: 1212 . 5427 [hep-ph].
- [156] M. Diehl, D. Ostermeier, and A. Schafer. "Elements of a theory for multiparton interactions in QCD". In: *J. High Energy Phys.* 2012.3 (Nov. 3, 2011). ISSN: 1029-8479. DOI: 10.1007/jhep03(2012)089. arXiv: 1111.0910 [hep-ph].
- [157] E. Huayra, J. V. C. Lovato, and E. G. de Oliveira. "Valence and sea parton correlations in double parton scattering from data". In: *J. High Energy Phys.* 09 (2023), p. 177. DOI: 10.1007/JHEP09(2023)177. arXiv: 2305.11106 [hep-ph].
- [158] Z. Conesa del Valle. "Heavy flavours in small systems". Accreditation to supervise research. Université Paris Saclay, Apr. 2023. <https://theses.hal.science/tel-04923473>.
- [159] LHCb Collaboration. "Observation of double charm production involving open charm in pp collisions at $\sqrt{s}=7$ TeV". In: *J. High Energy Phys.* 2012.6 (May 4, 2012). ISSN: 1029-8479. DOI: 10.1007/jhep06(2012)141. arXiv: 1205.0975 [hep-ex].
- [160] A. V. Berezhnuy et al. "Doubly Charmed Baryon Production in Hadronic Experiments". In: *Phys. Rev. D* 57.7 (Oct. 14, 1997), pp. 4385–4392. ISSN: 1089-4918. DOI: 10 . 1103 / physrevd.57.4385. arXiv: hep-ph/9710339 [hep-ph].
- [161] S. P. Baranov. "Topics in associated $J/\psi + c + \bar{c}$ production at modern colliders". In: *Phys. Rev. D* 73.7 (Apr. 2006), p. 074021. ISSN: 1550-2368. DOI: 10.1103/physrevd.73.074021.
- [162] J. P. Lansberg. "On the mechanisms of heavy-quarkonium hadroproduction". In: *Eur. Phys. J. C* 61.4 (Nov. 25, 2008), pp. 693–703. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-008-0826-9. arXiv: 0811.4005 [hep-ph].
- [163] LHCb Collaboration. "Measurement of the J/ψ pair production cross-section in pp collisions at $\sqrt{s} = 13$ TeV". In: *J. High Energy Phys.* 2017.6 (June 2017). ISSN: 1029-8479. DOI: 10.1007/jhep06(2017)047.
- [164] S. Cotogno, T. Kasemets, and M. Myska. "A spin on same-sign W-boson pair production". In: *Phys. Rev. D* 100.1 (Sept. 24, 2018), p. 011503. ISSN: 2470-0029. DOI: 10.1103/physrevd.100.011503. arXiv: 1809.09024 [hep-ph].
- [165] S. Cotogno, T. Kasemets, and M. Myska. "Confronting same-sign W-boson production with parton correlations". In: *J. High Energy Phys.* 2020.10 (Oct. 2020). ISSN: 1029-8479. DOI: 10.1007/jhep10(2020)214.
- [166] N. Armesto. "Nuclear shadowing". In: *J. High Energy Phys.* 32.11 (Apr. 12, 2006), R367–R393. ISSN: 1361-6471. DOI: 10 . 1088 / 0954 - 3899 / 32 / 11 / r01. arXiv: hep-ph/0604108 [hep-ph].
- [167] P. Paakkinen et al. "EPPS16 – Bringing nuclear PDFs to the LHC era". In: *Proceedings of 12th International Workshop on High- p_T Physics in the RHIC/LHC Era*. High- p_T 2017. Sissa Medialab, June 2019, p. 003. DOI: 10.22323/1.320.0003.
- [168] E665 Collaboration. "Saturation of shadowing at very low Bjorkenx". In: *Phys. Rev. Lett.* 68.22 (June 1992), pp. 3266–3269. ISSN: 0031-9007. DOI: 10.1103/physrevlett.68.3266.

- [169] **EMC** Collaboration. “The ratio of the nucleon structure functions F_2^N for iron and deuterium”. In: *Phys. Lett. B* 123.3 (1983), pp. 275–278. issn: 0370-2693. doi: [https://doi.org/10.1016/0370-2693\(83\)90437-9](https://doi.org/10.1016/0370-2693(83)90437-9).
- [170] Betty Abelev et al. “Suppression of high transverse momentum D mesons in central Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV”. In: *JHEP* 09 (2012), p. 112. doi: 10.1007/JHEP09(2012)112. arXiv: 1203.2160 [nucl-ex].
- [171] **ALICE** Collaboration. “Prompt D^0 , D^+ , and D^{*+} production in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *J. High Energy Phys.* 2022.1 (Jan. 2022). issn: 1029-8479. doi: 10.1007/jhep01(2022)174.
- [172] Kari J. Eskola et al. “EPPS21: a global QCD analysis of nuclear PDFs”. In: *Eur. Phys. J. C* 82.5 (2022), p. 413. doi: 10.1140/epjc/s10052-022-10359-0. arXiv: 2112.12462 [hep-ph].
- [173] K. Kovarik et al. “nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework”. In: *Phys. Rev. D* 93.8 (2016), p. 085037. doi: 10.1103/PhysRevD.93.085037. arXiv: 1509.00792 [hep-ph].
- [174] A. Krzywicki et al. “Does a nucleus act like a gluon filter?” In: *Phys. Lett. B* 85.4 (Aug. 1979), pp. 407–412. issn: 0370-2693. doi: 10.1016/0370-2693(79)91284-x.
- [175] R. Prino F. and Rapp. “Open heavy flavor in QCD matter and in nuclear collisions”. In: *J. Phys. G: Nucl. Part. Phys.* 43.9 (Aug. 2016), p. 093002. issn: 1361-6471. doi: 10.1088/0954-3899/43/9/093002.
- [176] S. Cao, G. Qin, and S. A. Bass. “Heavy quark dynamics and hadronization in ultra-relativistic heavy-ion collisions: collisional versus radiative energy loss”. In: *Phys. Rev. C* 88.4 (Aug. 2, 2013), p. 044907. issn: 1089-490X. doi: 10.1103/physrevc.88.044907. arXiv: 1308.0617 [nucl-th].
- [177] R. Baier et al. “Radiative energy loss of high energy quarks and gluons in a finite volume quark-gluon plasma”. In: *Nucl. Phys. A* 483.1–2 (July 17, 1996), pp. 291–320. issn: 0550-3213. doi: 10.1016/s0550-3213(96)00553-6. arXiv: hep-ph/9607355 [hep-ph].
- [178] Y. L. Dokshitzer and D. E. Kharzeev. “Heavy quark colorimetry of QCD matter”. In: *Phys. Lett. B* 519.3–4 (June 18, 2001), pp. 199–206. issn: 0370-2693. doi: 10.1016/s0370-2693(01)01130-3. arXiv: hep-ph/0106202 [hep-ph].
- [179] **ALICE** Collaboration. “Direct observation of the dead-cone effect in QCD”. In: *Nat. Phys.* 605.7910 (June 10, 2021), 440–446. issn: 1476-4687. doi: 10.1038/s41586-022-04572-w. arXiv: 2106.05713 [nucl-ex].
- [180] W. M. Alberico et al. “Heavy-flavour spectra in high-energy nucleus–nucleus collisions”. In: *Eur. Phys. J. C* 71.6 (June 2011). issn: 1434-6052. doi: 10.1140/epjc/s10052-011-1666-6.
- [181] J. Bellm et al. “Herwig 7.2 Release Note”. In: *Eur. Phys. J. C* 80.5 (Dec. 12, 2019). issn: 1434-6052. doi: 10.1140/epjc/s10052-020-8011-x. arXiv: 1912.06509 [hep-ph].
- [182] R. Scheibl and U. Heinz. “Coalescence and flow in ultrarelativistic heavy ion collisions”. In: *Phys. Rev. C* 59.3 (Mar. 1999), pp. 1585–1602. issn: 1089-490X. doi: 10.1103/physrevc.59.1585.
- [183] Y. Oh et al. “Ratios of heavy baryons to heavy mesons in relativistic nucleus-nucleus collisions”. In: *Phys. Rev. C* 79.4 (Apr. 2009), p. 044905. issn: 1089-490X. doi: 10.1103/physrevc.79.044905.
- [184] T. Song et al. “Tomography of the Quark-Gluon-Plasma by Charm Quarks”. In: *Phys. Rev. C* 92.1 (Mar. 10, 2015), p. 014910. issn: 1089-490X. doi: 10.1103/physrevc.92.014910. arXiv: 1503.03039 [nucl-th].

- [185] S. Cao et al. “Linearized Boltzmann transport model for jet propagation in the quark-gluon plasma: Heavy quark evolution”. In: *Phys. Rev. C* 94.1 (2016), p. 014909. DOI: 10.1103/PhysRevC.94.014909. arXiv: 1605.06447 [nucl-th].
- [186] F. Scardina et al. “Estimating the charm quark diffusion coefficient and thermalization time from D meson spectra at energies available at the BNL Relativistic Heavy Ion Collider and the CERN Large Hadron Collider”. In: *Phys. Rev. C* 96.4 (2017), p. 044905. DOI: 10.1103/PhysRevC.96.044905. arXiv: 1707.05452 [nucl-th].
- [187] S. Plumari et al. “Heavy-light flavour correlations of anisotropic flows at LHC energies within event-by-event transport approach”. In: *Phys. Lett. B* 805 (Dec. 19, 2019), p. 135460. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2020.135460. arXiv: 1912.09350 [hep-ph].
- [188] L. Ravagli and R. Rapp. “Quark coalescence based on a transport equation”. In: *Phys. Lett. B* 655.3–4 (Nov. 2007), pp. 126–131. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2007.07.043.
- [189] M. He and R. Rapp. “Hadronization and Charm-Hadron Ratios in Heavy-Ion Collisions”. In: *Phys. Rev. Lett.* 124, 042301 (2020) 124.4 (May 22, 2019), p. 042301. ISSN: 1079-7114. DOI: 10.1103/physrevlett.124.042301. arXiv: 1905.09216 [nucl-th].
- [190] A. Beraudo et al. “Heavy flavors in heavy-ion collisions: quenching, flow and correlations”. In: *Eur. Phys. J. C* 75.3 (Mar. 2015). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3336-6.
- [191] ALICE Collaboration. “Transverse-momentum and event-shape dependence of D-meson flow harmonics in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *Phys. Lett. B* 813 (May 22, 2020), p. 136054. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2020.136054. arXiv: 2005.11131 [nucl-ex].
- [192] R. Katz et al. “DAB-MOD sensitivity study of heavy flavor R_{AA} and azimuthal anisotropies based on beam energy, initial conditions, hadronization, and suppression mechanisms”. In: *Phys. Rev. C* 102.2 (June 25, 2019), p. 024906. ISSN: 2469-9993. DOI: 10.1103/physrevc.102.024906. arXiv: 1906.10768 [nucl-th].
- [193] L. Evans and P. Bryant. “LHC Machine”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08001–S08001. ISSN: 1748-0221. DOI: 10.1088/1748-0221/3/08/s08001.
- [194] ALICE Collaboration. “The ALICE experiment at the CERN LHC”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08002–S08002. ISSN: 1748-0221. DOI: 10.1088/1748-0221/3/08/s08002.
- [195] ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08003–S08003. ISSN: 1748-0221. DOI: 10.1088/1748-0221/3/08/s08003. arXiv:2305.16623.
- [196] CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08004–S08004. ISSN: 1748-0221. DOI: 10.1088/1748-0221/3/08/s08004.
- [197] LHCb Collaboration. “The LHCb Detector at the LHC”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08005–S08005. ISSN: 1748-0221. DOI: 10.1088/1748-0221/3/08/s08005.
- [198] TOTEM Collaboration. “CMS-TOTEM Precision Proton Spectrometer”. In: (Sept. 2014).
- [199] LHCf Collaboration. “Technical design report of the LHCf experiment: Measurement of photons and neutral pions in the very forward region of LHC”. In: (Feb. 2006).
- [200] MoEDAL-MAPP Collaboration. “MoEDAL-MAPP, an LHC Dedicated Detector Search Facility”. In: *Snowmass 2021*. Sept. 2022. arXiv: 2209.03988 [hep-ph].

- [201] J. L. Feng et al. “ForwArd Search ExpeRiment at the LHC”. In: *Phys. Rev. D* 97.3 (2018), p. 035001. doi: 10.1103/PhysRevD.97.035001. arXiv: 1708.09389 [hep-ph].
- [202] **SND@LHC** Collaboration. “SND@LHC: the scattering and neutrino detector at the LHC”. In: *Journal of Instrumentation* 19.05 (2024), P05067. doi: 10.1088/1748-0221/19/05/P05067. arXiv: 2210.02784 [hep-ex].
- [203] I. Ravasenga. “Commissioning and Performance of the New ALICE Inner Tracking System in the First Phase of LHC Run 3”. In: *JPS Conf. Proc.* 42 (2024), p. 011002. doi: 10.7566/JSPCP.42.011002. arXiv: 2302.00432 [physics.ins-det].
- [204] **ALICE** Collaboration. *ALICE Inner Tracking System (ITS) : Technical Design Report*. Tech. rep. 1999. cds.cern.ch/record/391175.
- [205] **ALICE** Collaboration. *Technical Design Report for the Upgrade of the ALICE Inner Tracking System*. Tech. rep. 2014, p. 087002. doi: 10.1088/0954-3899/41/8/087002.
- [206] **ALICE** Collaboration. *ALICE: Technical design report of the time projection chamber*. Tech. rep. Jan. 2000.
- [207] C. Lippmann. *Upgrade of the ALICE Time Projection Chamber*. Tech. rep. Mar. 2014.
- [208] **ALICE** Collaboration. “The ALICE Transition Radiation Detector: construction, operation, and performance”. In: *Nucl. Instrum. Meth. A* (Sept. 2017). doi: 10.1016/j.nima.2017.09.028. arXiv: 1709.02743 [physics.ins-det].
- [209] **ALICE** Collaboration. *ALICE Time-Of-Flight system (TOF) : Technical Design Report*. Tech. rep. 2000. <https://cds.cern.ch/record/430132>.
- [210] L. Terlizzi. “An upgraded muon identification system for ALICE at the LHC: from detector construction to testing and re-commissioning”. Presented 26 Apr 2023. Torino U., 2023. <https://cds.cern.ch/record/2858769>.
- [211] M. Gallio et al. *ALICE Zero-Degree Calorimeter (ZDC): Technical Design Report*. Technical design report. ALICE. Geneva: CERN, 1999. <https://cds.cern.ch/record/381433>.
- [212] A. Maevskaya. “ALICE FIT Data Processing and Performance during LHC Run 3”. In: *Phys. Atom. Nucl.* 84.4 (2021), pp. 579–584. doi: 10.1134/S1063778821040189. arXiv: 2012.02760 [physics.ins-det].
- [213] **ALICE** Collaboration. *ALICE dimuon forward spectrometer: Technical Design Report*. Technical design report. ALICE. Geneva: CERN, 1999. <https://cds.cern.ch/record/401974>.
- [214] **ALICE** Collaboration. *ALICE dimuon forward spectrometer: addendum to the Technical Design Report*. Technical design report. ALICE. Geneva: CERN, 2000. <https://cds.cern.ch/record/494265>.
- [215] **ALICE** Collaboration. *Technical Design Report for the Muon Forward Tracker*. Tech. rep. 2015. <https://cds.cern.ch/record/1981898>.
- [216] M. Tarhini. “Measurement of Z-boson and J/ψ Production in p-Pb and Pb-Pb Collisions at $\sqrt{s_{NN}} = 5.02$ TeV with ALICE at the LHC”. Theses. Universite Paris Saclay (COMUE), June 2017. <https://theses.hal.science/tel-01691725>.
- [217] **ALICE** Collaboration et al. *The ALICE Dipole Magnet Calculation*. Tech. rep. Geneva: CERN, 1997. <https://cds.cern.ch/record/689184>.
- [218] S. Siddhanta. “The ALICE Muon Tracking Chambers Upgrade”. In: *2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. IEEE, Oct. 2020, pp. 1–3. doi: 10.1109/nss/mic42677.2020.9507849.
- [219] F. Bossù. “The trigger system of the muon spectrometer of the ALICE experiment at the LHC”. In: *Nucl. Phys. B Proc. Suppl.* 215.1 (June 2011), pp. 204–207. ISSN: 0920-5632. doi: 10.1016/j.nuclphysbps.2011.04.009.
- [220] **ALICE** Collaboration. *Future high pp programme with ALICE*. Tech. rep. An update version will be posted on 9th September 2020. Geneva: CERN, 2020. <https://cds.cern.ch/record/2729936>.

- [221] P. Buncic, M. Krzewicki, and P. Vande Vyvre. *Technical Design Report for the Upgrade of the Online-Offline Computing System*. Tech. rep. 2015. <http://cds.cern.ch/record/2011297>.
- [222] V. Barroso et al. "The new ALICE data acquisition system (O2/FLP) for LHC Run 3". In: *EPJ Web Conf.* 295 (2024). Ed. by R. De Vita et al., p. 02029. ISSN: 2100-014X. DOI: 10.1051/epjconf/202429502029.
- [223] A. Alkin et al. "ALICE Run 3 Analysis Framework". In: *EPJ Web Conf.* 251 (2021), p. 03063. DOI: 10.1051/epjconf/202125103063. <https://cds.cern.ch/record/2814355>.
- [224] R. Brun and F. Rademakers. "ROOT: An object oriented data analysis framework". In: *Nucl. Instrum. Meth. A* 389 (1997). Ed. by M. Werlen and D. Perret-Gallix, pp. 81–86. DOI: 10.1016/S0168-9002(97)00048-X.
- [225] AliceO2Group. *O2 Analysis Framework Documentation*. 2024. <https://aliceo2group.github.io/analysis-framework/>.
- [226] AliceO2Group. *O2Physics*. <https://github.com/AliceO2Group/O2Physics>. 2025.
- [227] Z. Conesa del Valle. "Performance of the ALICE muon spectrometer.
Weak boson production and measurement in heavy-ion collisions at LHC." Theses. Université de Nantes ; Universitat Autònoma de Barcelona, July 2007. <https://theses.hal.science/tel-00198703>.
- [228] M. Ivanov et al. "TPC tracking and particle identification in high-density environment". In: (June 2003). DOI: 10.48550/ARXIV.PHYSICS/0306108. arXiv: physics/0306108 [physics.data-an].
- [229] M. Arslanodok et al. "Track Reconstruction in a High-Density Environment with ALICE". In: *Particles* 5.1 (Mar. 19, 2022), pp. 84–95. ISSN: 2571-712X. DOI: 10.3390/particles5010008. arXiv: 2203.10325 [physics.ins-det].
- [230] J. Martin Blanco. "Study of J/ψ production dependence with the charged particle multiplicity in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV and pp collisions at $\sqrt{s} = 8$ TeV with the ALICE experiment at the LHC". PhD thesis. Nantes U., Jan. 2016.
- [231] T. Tork. "Charmonium production as a function of charged-particle multiplicity in p-Pb collisions at $\sqrt{s_{NN}} = 8.16$ TeV with ALICE at the LHC". Theses. Université Paris-Saclay, Oct. 2023. <https://theses.hal.science/tel-04313420>.
- [232] R. Sadek Finot. "MFT commissioning and preparation for Run 3 data analysis with ALICE (LHC, CERN)". Theses. Ecole nationale supérieure Mines-Telecom Atlantique, Oct. 2022. <https://theses.hal.science/tel-04008085>.
- [233] A. Bigot. "Étude de la production de beauté ouverte avec le détecteur ALICE au LHC". Theses. Université de Strasbourg, Sept. 2024. <https://theses.hal.science/tel-04839692>.
- [234] S. Agostinelli et al. "GEANT4: A simulation toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.
- [235] I. Belikov. "Event reconstruction and particle identification in the ALICE experiment at the LHC". In: *EPJ Web Conf.* 70 (2014). Ed. by Larissa Bravina, Yiota Foka, and Sonia Kabana, p. 00029. ISSN: 2100-014X. DOI: 10.1051/epjconf/20147000029.
- [236] D. Rohr et al. "Track Reconstruction in the ALICE TPC using GPUs for LHC Run 3". In: *4th International Workshop Connecting The Dots 2018*. Nov. 2018. arXiv: 1811.11481 [physics.ins-det].
- [237] M. Ljunggren. "Performance simulation studies for the ALICE TPC GEM upgrade". In: *Nucl. Instrum. Methods Phys. Res., Sect. A* 824 (July 2016), pp. 548–550. ISSN: 0168-9002. DOI: 10.1016/j.nima.2015.12.059.

- [238] K. Deja, T. Trzciniński, and Ł. Graczykowski. “Generative models for fast cluster simulations in the TPC for the ALICE experiment”. In: *EPJ Web Conf.* 214 (2019). Ed. by A. Forti et al., p. 06003. ISSN: 2100-014X. DOI: 10.1051/epjconf/201921406003.
- [239] UC Berkeley School of Information. *Data Science Program*. UCB-UMT. Accessed: 2025-04-10. May 2025. <https://ischoolonline.berkeley.edu/data-science/>.
- [240] Inc. O’Reilly Media, ed. *Introduction to Machine Learning with Python: A Guide for Data Scientists*. 2016. ISBN: 978-1-4493-6989-7.
- [241] L. Barioglio et al. *hipe4ml/hipe4ml (v0.0.14)*. <https://hipe4ml.github.io/>. 2022.
- [242] W. McKinney. “Data Structures for Statistical Computing in Python”. In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van der Walt and Jarrod Millman. 2010, pp. 56–61. DOI: 10.25080/Majora-92bf1922-00a.
- [243] The pandas development team. *pandas-dev/pandas: Pandas*. Version latest. Feb. 2020. DOI: 10.5281/zenodo.3509134. <https://doi.org/10.5281/zenodo.3509134>.
- [244] Y. Shi et al. *lightgbm: Light Gradient Boosting Machine*. R package version 4.6.0.99. 2025. <https://github.com/Microsoft/LightGBM>.
- [245] T. Chen and C. Guestrin. “XGBoost: A Scalable Tree Boosting System”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ’16. ACM, Aug. 2016, pp. 785–794. DOI: 10.1145/2939672.2939785.
- [246] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *J. Mach. Learn. Res.* 12 (2011), pp. 2825–2830.
- [247] L. Breiman et al. *Classification And Regression Trees*. Routledge, Oct. 1984. ISBN: 978-1-315-13947-0. DOI: 10.1201/9781315139470.
- [248] ALICE Collaboration. “Measurement of beauty and charm production in pp collisions at $\sqrt{s} = 5.02$ TeV via non-prompt and prompt D mesons”. In: *J. High Energy Phys.* 2021.5 (May 2021). ISSN: 1029-8479. DOI: 10.1007/jhep05(2021)220. arXiv: 2102.13601.
- [249] S. Lundberg and S. Lee. “A Unified Approach to Interpreting Model Predictions”. In: (May 2017). DOI: 10.48550/ARXIV.1705.07874. arXiv: 1705.07874 [cs.AI].
- [250] A. E. Roth. *The Shapley Value: Essays in Honor of Lloyd S. Shapley*. Cambridge University Press, Oct. 1988. ISBN: 978-0-511-52844-6. DOI: 10.1017/cbo9780511528446.
- [251] ONNX Runtime developers. *ONNX Runtime, (1.21.0)*. <https://onnxruntime.ai/>. 2021.
- [252] ALICE Collaboration. “Measurement of the non-prompt D-meson fraction as a function of multiplicity in proton–proton collisions at $\sqrt{s} = 13$ TeV”. In: *J. High Energy Phys.* 2023.10 (Feb. 15, 2023). ISSN: 1029-8479. DOI: 10.1007/jhep10(2023)092. arXiv: 2302.07783 [nucl-ex].
- [253] F. Catalano. “Strange D-meson measurements in pp and Pb–Pb collisions exploiting machine learning with ALICE at the LHC”. PhD thesis. Polytech. Turin, 2022.
- [254] Y. Sara Amhis, S. Banerjee, et al. “Averages of b-hadron, c-hadron, and τ -lepton properties as of 2018”. In: *Eur. Phys. J. C* 81.3 (2021), p. 226. DOI: 10.1140/epjc/s10052-020-8156-7. arXiv: 1909.12524 [hep-ex].
- [255] Wouter V. and David K. *The RooFit toolkit for data modeling*. 2003. arXiv: physics/0306116 [physics.data-an]. <https://arxiv.org/abs/physics/0306116>.
- [256] V. Di Bella. *Data Driven studies of tracking efficiency with charm-hadron tag-and-probe method*. Sept. 2024.

APPENDICES

COMPARISON OF THE MUON TRACKING EFFICIENCY IN DATA AND THE ANCHORED MC

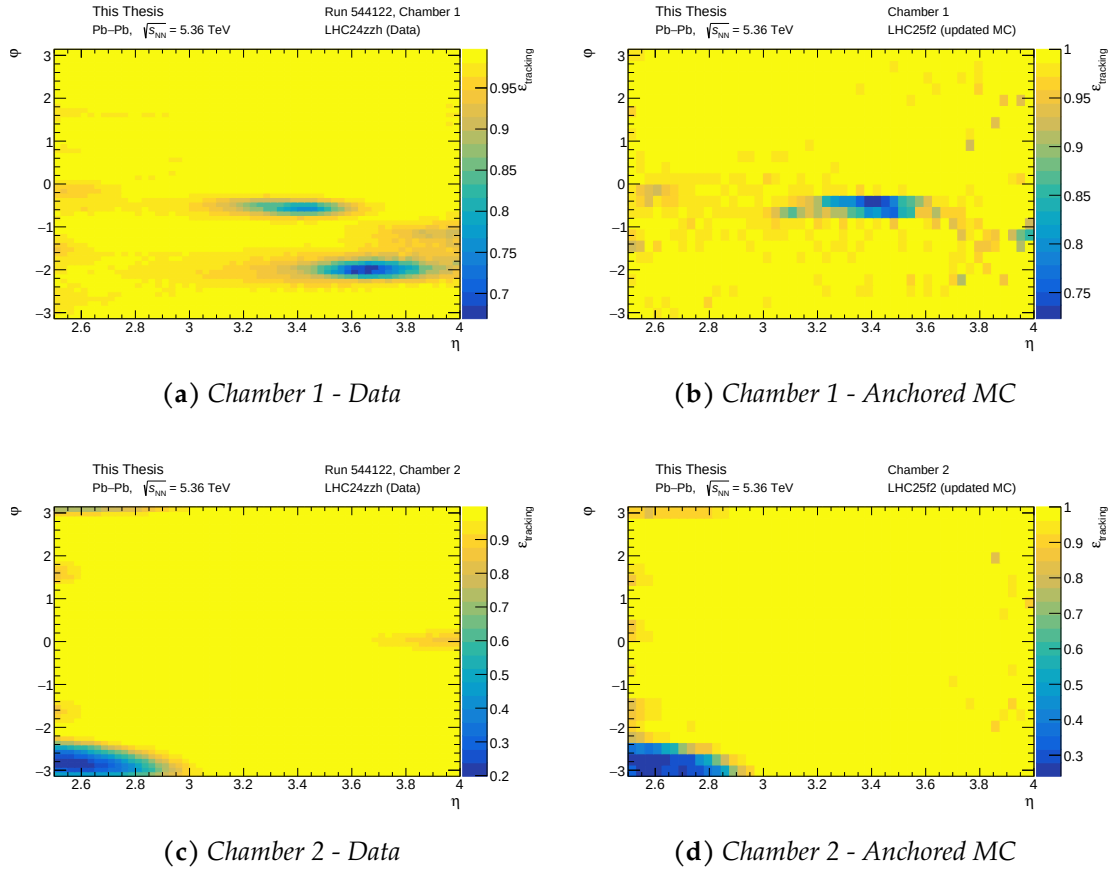


Figure A.1: Left figures: MCH tracking efficiency of the data. Right figures: MCH tracking efficiency of the anchored MC for the same chambers. All plots correspond to run 544122 and are measured as a function of η and ϕ (part 1).

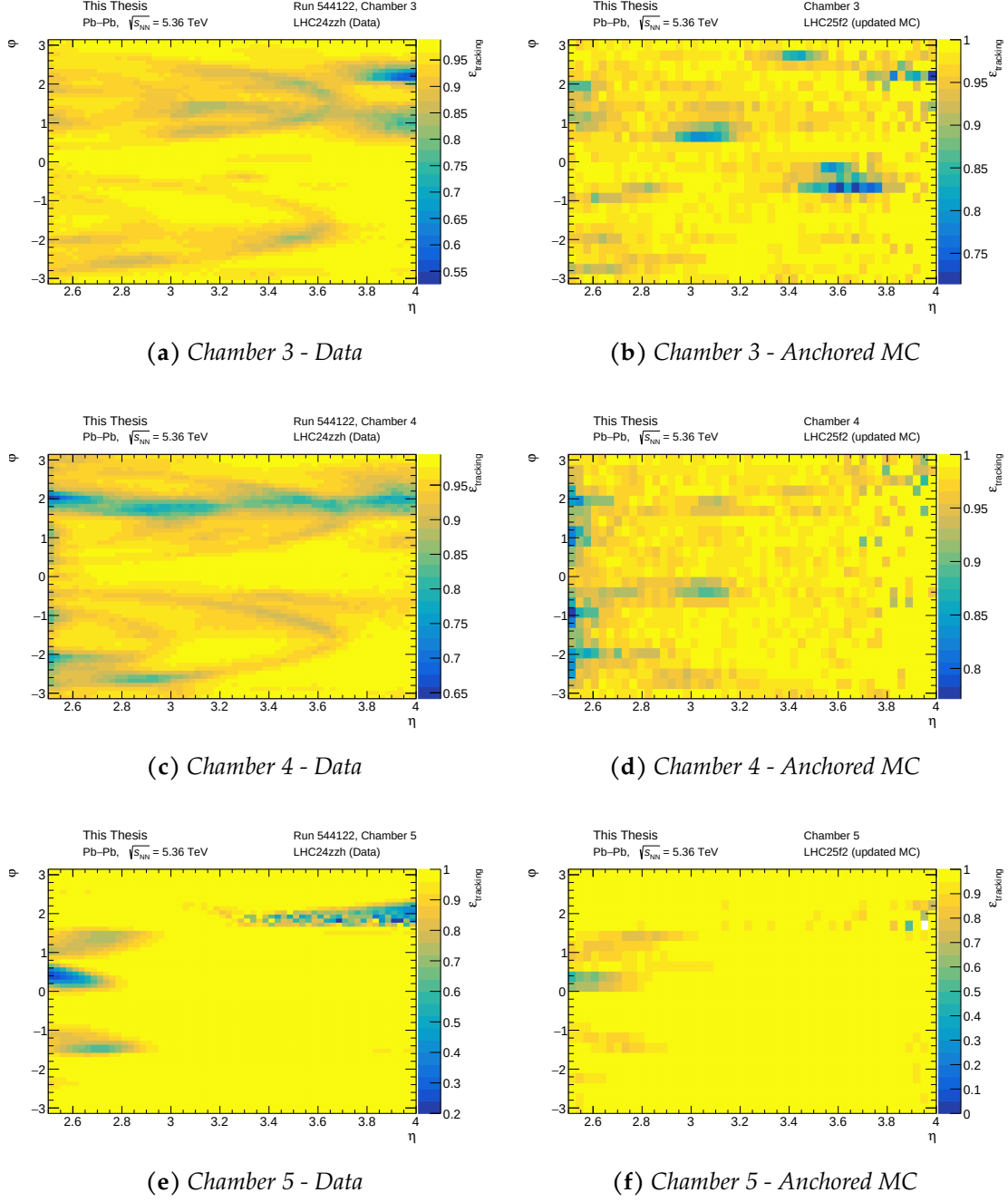


Figure A.2: Left figures: MCH tracking efficiency of the data. Right figures: MCH tracking efficiency of the anchored MC for the same chambers. All plots correspond to run 544122 and are measured as a function of η and ϕ (part 2).

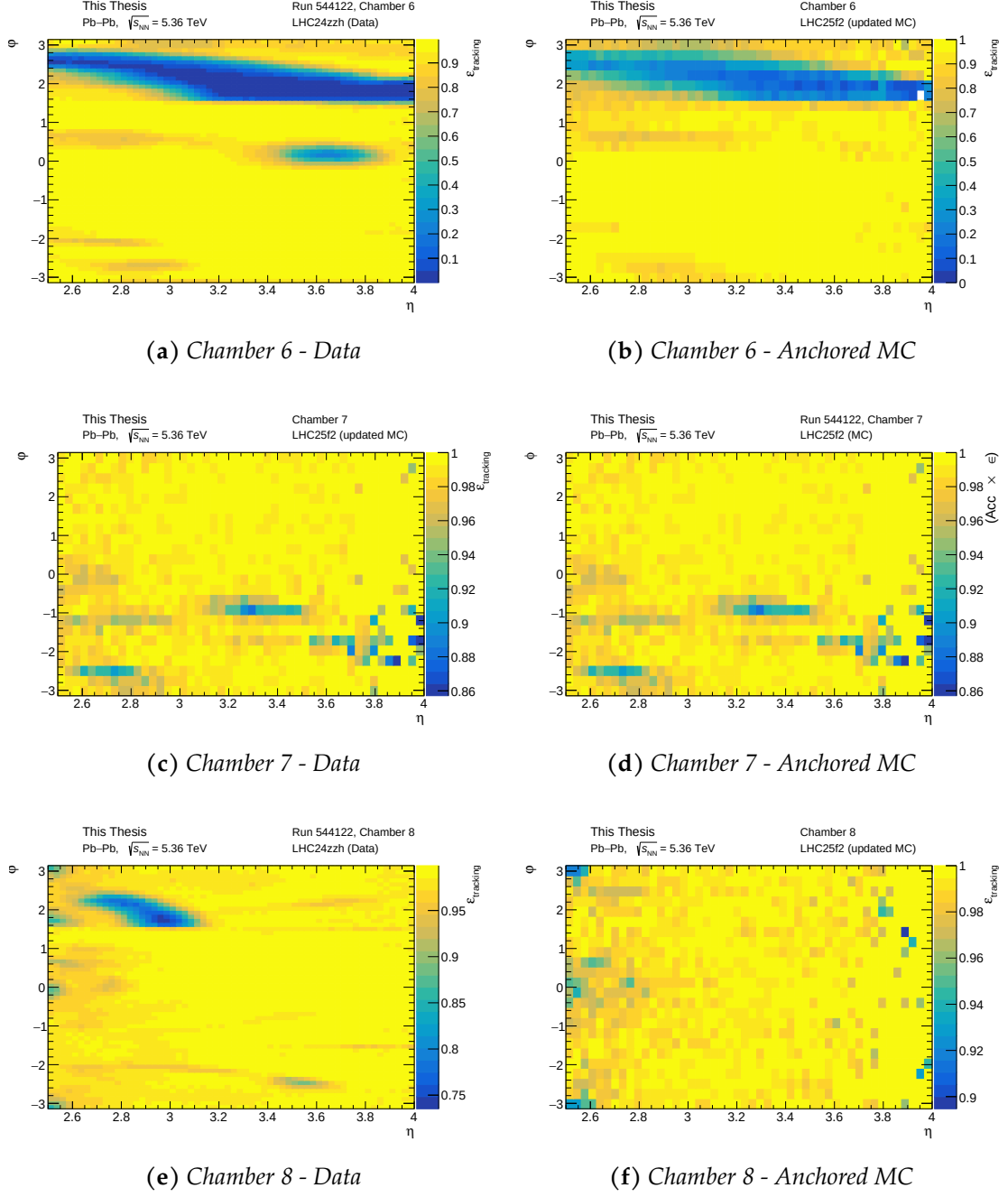


Figure A.3: Left figures: MCH tracking efficiency of the data. Right figures: MCH tracking efficiency of the anchored MC for the same chambers. All plots correspond to run 544122 and are measured as a function of η and ϕ (part 3).

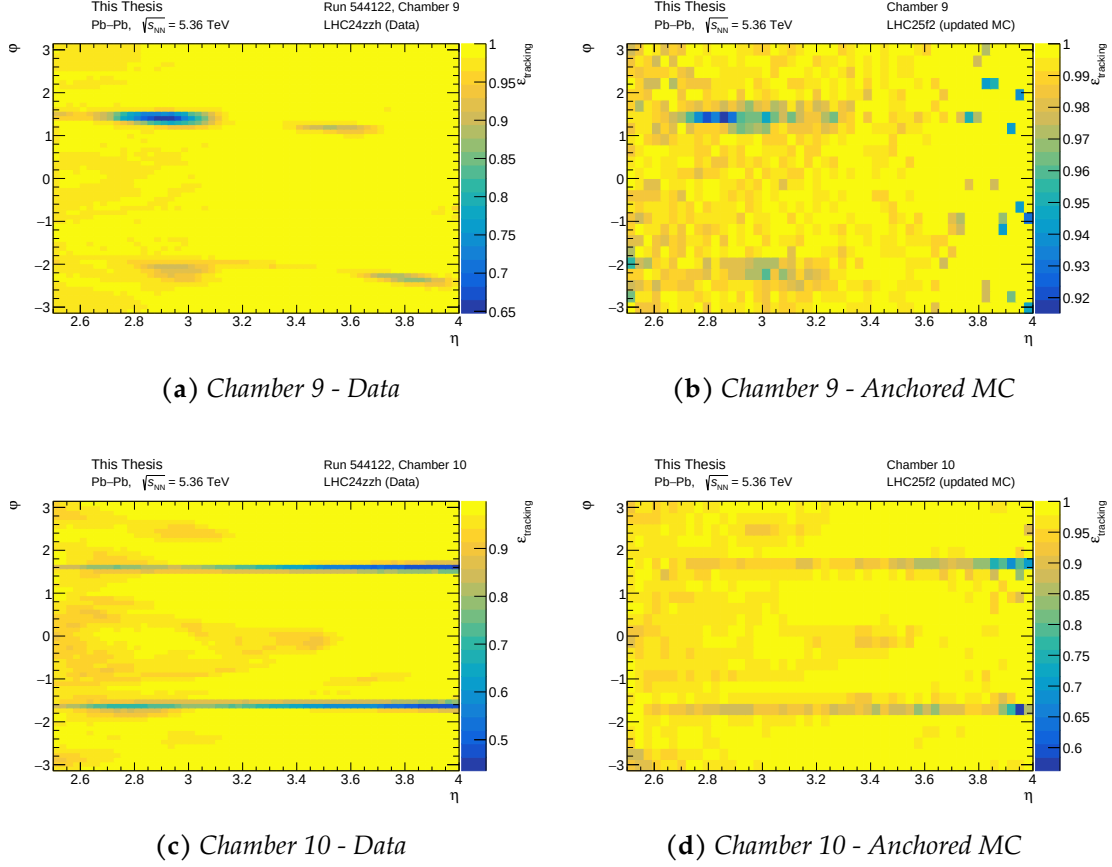


Figure A.4: Left figures: MCH tracking efficiency of the data. Right figures: MCH tracking efficiency of the anchored MC for the same chambers. All plots correspond to run 544122 and are measured as a function of η and ϕ (part 4).

LIST OF RUNS INCLUDED IN 2024 DATASET

- **LHC24af_pass1_MinBias**
550367, 550369, 550375, 550412, 550417, 550421, 550424, 550425, 550439, 550630,
550632, 550634, 550653, 550654, 550690, 550691, 550707, 550711, 550728, 550731,
550742, 550756, 550760, 550774, 550778, 550781, 550784, 550819, 550824, 550843,
550848, 550852, 550858, 550889, 550916.
- **LHC24ao_pass1_MinBias**
558122, 558124, 558126, 558150, 558153, 558155, 558179, 558182, 558215, 558217,
558221, 558244, 558247, 558273, 558275, 558284, 558288, 558291, 558327, 558329,
558330, 558354, 558369, 558383, 558387, 558390, 558406, 558409, 558410, 558422,
558433, 558437, 558449, 558482, 558535, 558551, 558602, 558604, 558606, 558615,
558627, 558633, 558656, 558676, 558685, 558726, 558744, 558750, 558752, 558757,
558783, 558801.
- **LHC24aj_pass1_MinBias**
551759, 551760, 551761, 551780, 551843, 551856, 551874, 551875, 551877, 551889,
551890, 551894, 551921, 551922, 551923, 551924, 551925, 551926, 551931, 551943,
551958, 551977, 551979, 551982, 551983, 551989, 551992, 551993, 551997, 552005,
552029, 552080, 552102, 552103, 552138, 552139, 552140, 552141, 552142, 552156,
552176, 552177, 552178, 552179, 552197, 552198, 552200, 552201, 552203, 552204,
552205, 552206, 552283, 552285, 552304, 552340, 552341, 552353, 552369, 552381,
552382, 552383, 552384, 552400, 552401, 552402, 552403.
- **LHC24ag_pass1_MinBias**
550997, 551005, 551007, 551008, 551013, 551023, 551027, 551066, 551083, 551105,
551107, 551122, 551127, 551149, 551219, 551221, 551229, 551230, 551232, 551255,
551257, 551260, 551272, 551290, 551296, 551365, 551387, 551389, 551391, 551392,
551394, 551398, 551418, 551427, 551463, 551468, 551498, 551504.
- **LHC24am_pass1_MB**
555344, 555345, 555370, 555374, 555398, 555401, 555408, 555411, 555431, 555435,
555443, 555451, 555476, 555478, 555482, 555504, 555540, 555543, 555546, 555575,
555591, 555596, 555612, 555649, 555651, 555676, 555693, 555695, 555705, 555707,
555722, 555723, 555740, 555742, 555759, 555761, 555763, 555789, 555790, 555798,

555801, 555850, 555853, 555860, 555881, 555883, 555900, 555917, 555933, 555958, 555960, 555965, 555967, 555976.

- **LHC24an_pass1_MB**

556152, 556160, 556164, 556182, 556210, 556218, 556237, 556248, 556269, 556284, 556370, 556372, 556412, 556437, 556454, 556461, 556482, 556485, 556491, 556497, 556517, 556542, 556562, 556639, 556640, 556641, 556662, 556664, 556680, 556716, 556734, 556741, 556767, 556816, 556834, 556872, 556889, 556907, 556909, 556913, 556923, 556939, 556954, 556958, 556979, 556981, 556997, 557012, 557019, 557021, 557026, 557074, 557104, 557112, 557118, 557119, 557138, 557149, 557226, 557233, 557251, 557271, 557291, 557299, 557321, 557336, 557339, 557350, 557374, 557415, 557425, 557444, 557481, 557482, 557509, 557531, 557547, 557613, 557659, 557681, 557691, 557717, 557723, 557726, 557744, 557749, 557862, 557876, 557897, 557913, 557926.

- **LHC24al_pass1_MinBias**

553185, 553187, 553188, 553189, 553193, 553219, 553225, 553250, 553253, 553255, 553274, 553294, 553297, 553299, 553305, 553486, 553512, 553530, 553536, 553555, 553588, 553590, 553610, 553633, 553655, 553660, 553663, 553700, 553702, 553739, 553756, 553785, 553807, 553816, 553819, 553821, 553824, 553825, 553844, 553862, 553880, 553903, 554092, 554094, 554095, 554097, 554098, 554194, 554198, 554201, 554203, 554207, 554208, 554223, 554247, 554261, 554293, 554295, 554316, 554322, 554323, 554354, 554394, 554404, 554408, 554413, 554427, 554462, 554471, 554494, 554495, 554504, 554507, 554524, 554526, 554538, 554558, 554564, 554569, 554588, 554603, 554613, 554615, 554633, 554701, 554703, 554714, 554728, 554732, 554736, 554752, 554768, 554772, 554774, 554791, 554808, 554835, 554837, 554873, 554880, 554898, 554920, 554968, 554970, 554973, 554998, 555020, 555022, 555023, 555047, 555071, 555073, 555121, 555122, 555124, 555150, 555152, 555156, 555160, 555166, 555172, 555187, 555202, 555208, 555226, 555232, 555254, 555259, 555267, 555270, 555308.

COMPARISON OF THE DOUBLE D^0 PRODUCTION RESULTS WITH THE PRELIMINARY RESULTS OF MARCH 2025

The results presented in this thesis correspond to an updated version of the results approved as ALICE Preliminary in March 2025. In this Appendix, we present those results and compare them with the ones in Chapter 7.

C.1 Updates in the ML models and selection criteria

The main differences in the BDT models used in this thesis and the ones used in the Preliminary results concern the PID variables. In the Preliminary results, the BDT models were trained using the same PID variables as in Chapter 6, while in this thesis, the BDT models were trained using the *swapped* PID variables described in Section 7.4. This resulted in a better performance of the BDT models which can be seen by comparing the ROC AUC values per p_T range (Tab. C.1).

Table C.1: Values of the average ROC AUC of all trained models in the Preliminary results vs. this thesis.

p_T (GeV/ c)	(1, 2)	(2, 3)	(3, 5)	(5, 7)	(8, 10)	(10, 24)
ROC AUC (This thesis)	0.882	0.919	0.951	0.959	0.945	0.933
ROC AUC (Preliminary)	0.846	0.870	0.887	0.880	0.881	0.890

The increased performance of the models allowed for an expansion of the p_T range used for the analysis. In the Preliminary results, the p_T range was limited to the (2, 24) GeV/ c region, while in this thesis, the p_T range was extended to (1, 24) GeV/ c . This allowed for a better coverage of the p_T spectrum and a more precise measurement of the associated production of double D^0 mesons.

Lastly, in the Preliminary results, candidates passing the selections both as D^0 and $\bar{D}^0$ were removed from the analysis in all p_T ranges, while in this thesis we only removed

the reflected candidates in the $(1, 5)$ GeV/ c range. This change was done to increase the statistics in the high- p_T region, where the PID is not applied anymore, causing a drop of the efficiencies. The efficiencies used in the Preliminary measurement are shown in Fig. C.1.

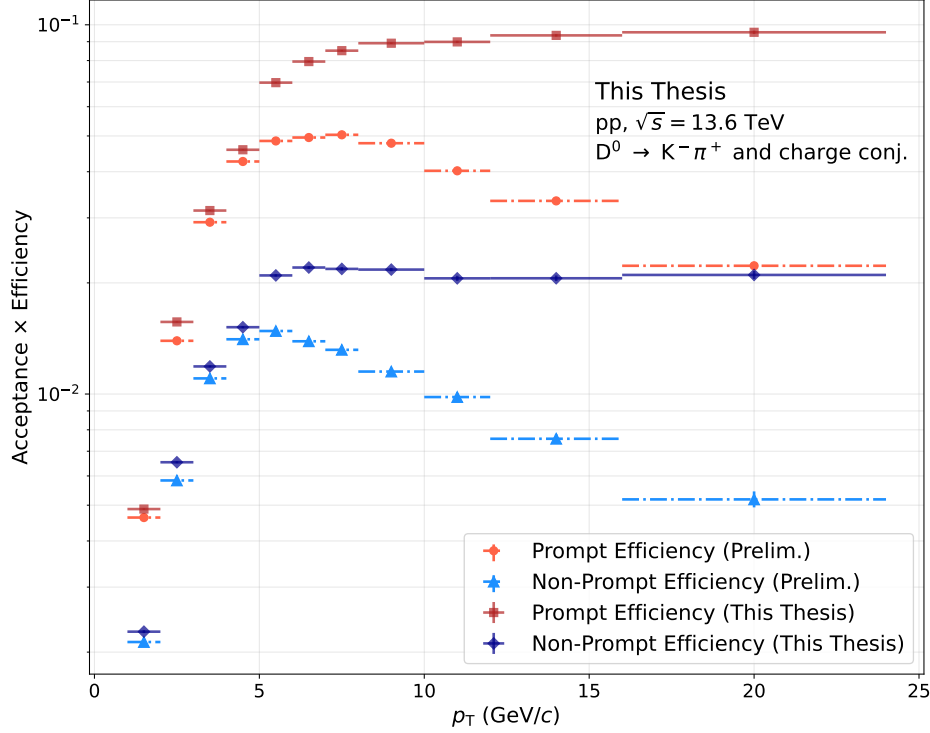


Figure C.1: Prompt and non-prompt efficiencies computed with the selections that were approved as Preliminary compared to the ones presented in this thesis.

C.2 Comparison of the integrated results

The Preliminary efficiency-corrected yields of the associated production of double D^0 pairs are presented in Fig. C.2. These fits were performed using the same method as in Chapter 7. A Gaussian distribution was used to fit the signal candidates and an exponential for the background. No reflected functions were used in the fits, as the reflected candidates were removed from the analysis in all p_T ranges.

The resulting p_T -integrated cross sections for LS and OS pairs are shown in Fig. C.3, and the corresponding ratios, $\mathcal{R}_{DD}$, can be found in Fig. C.4. We can compare these values with the ones we obtain when we use the selections of Chapter 7. This is presented in Tab. C.2. As can be observed, all measurements are compatible with each other within the uncertainties. The differences in the cross sections and ratios can be attributed to the different selections and BDT models used in the two analyses, as well as the removal of reflected candidates in all p_T ranges in the Preliminary results.

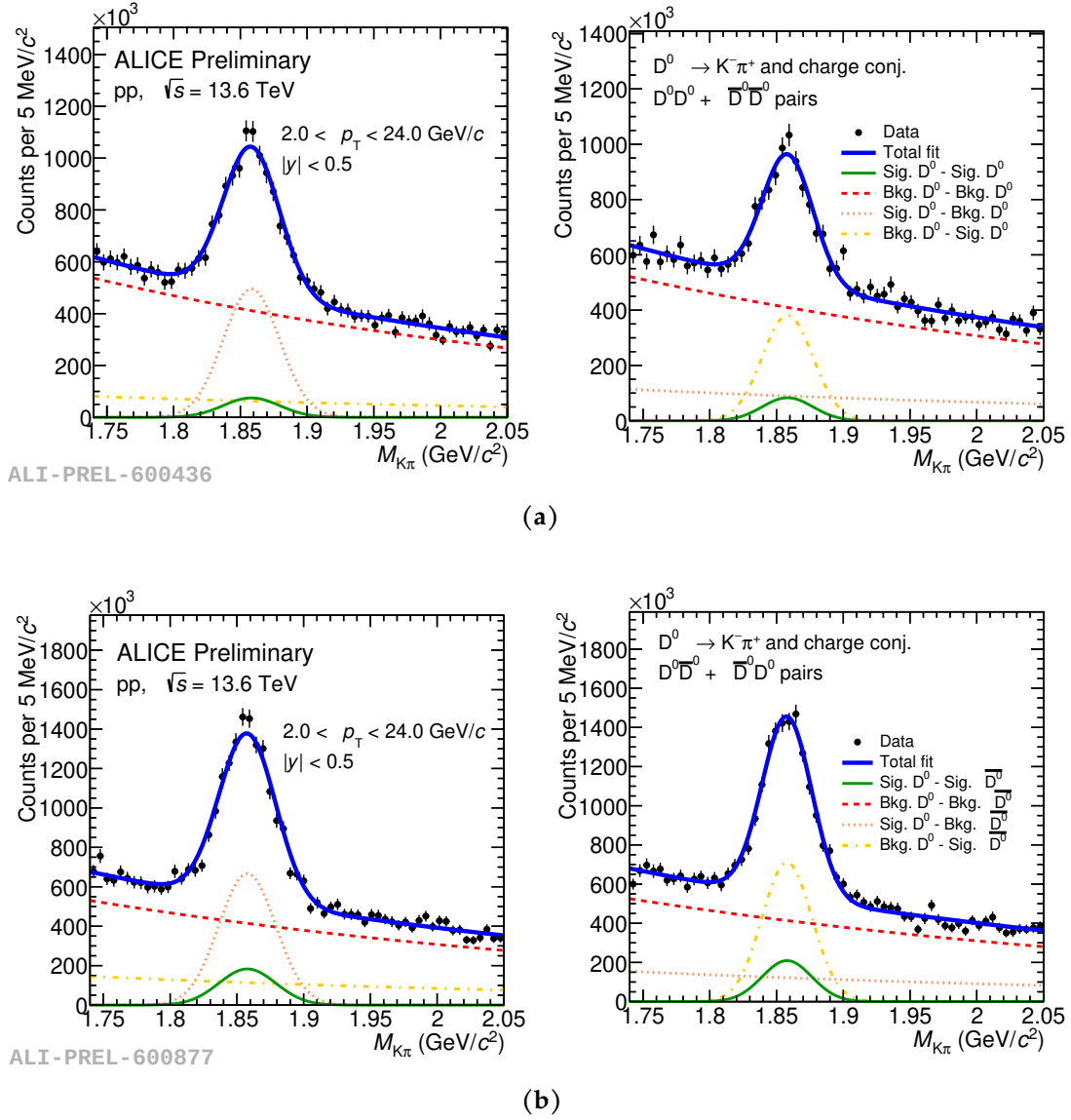


Figure C.2: Projections of the two-dimensional plots of the invariant mass distributions of the D^0 pair candidates in the $2 < p_T < 24$ GeV/c range for (a) Like-sign pairs (b) Opposite-sign pairs. The figures on the left represent the first D^0 candidate, and the ones on the right the second.

Table C.2: Comparison of the integrated cross sections and $\mathcal{R}_{DD}$ in the $2 < p_T < 24$ GeV/c using the selections of this thesis vs. the Preliminary results for OS and LS pairs. In all cases, the first uncertainty is statistical and the second is systematic.

	LS cross section (μb)	OS cross section (μb)	LS $\mathcal{R}_{DD}$ (mb)	OS $\mathcal{R}_{DD}$ (mb)
This thesis	$15.2 \pm 0.7 \pm 2.5$	$35.5 \pm 1.1 \pm 6.0$	$3.31 \pm 0.28 \pm 0.76$	$1.44 \pm 0.09 \pm 0.32$
Preliminary	$12.5 \pm 0.5 \pm 2.2$	$29.8 \pm 0.9 \pm 5.3$	$4.03 \pm 0.21 \pm 0.76$	$1.69 \pm 0.08 \pm 0.32$

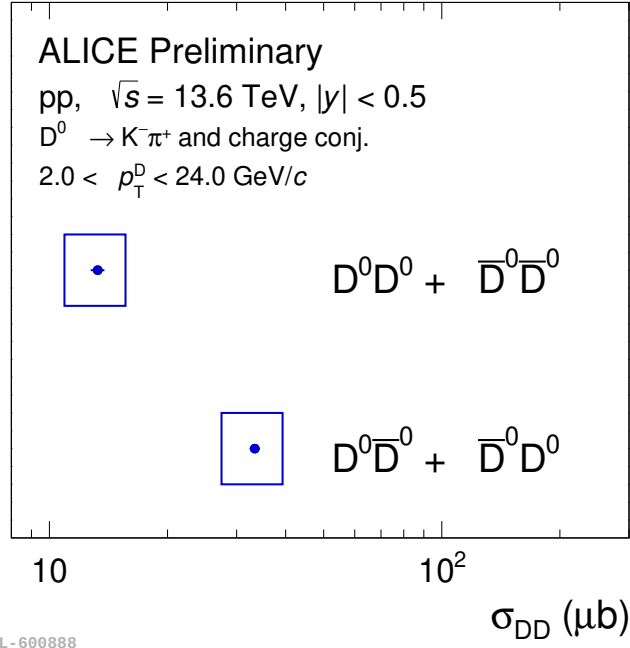


Figure C.3: Integrated cross sections of like-sign and opposite-sign pair D^0 mesons measured in the $2 < p_T < 24$ GeV/c range that were approved as Preliminary.

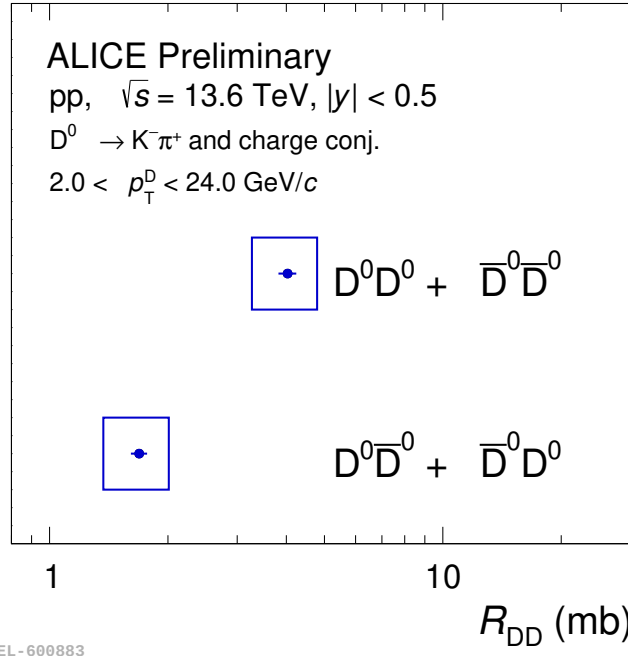


Figure C.4: Ratio of the single and double of like-sign and opposite-sign pair D^0 mesons measured in the $2 < p_T < 24$ GeV/c range that were approved as Preliminary.

LIST OF FIGURES

1.1	Particles described by the Standard Model. Image from the CERN Webfest 2012 [2].	6
1.2	Vacuum polarisation Feynman diagrams. On the left, diagram contributing to the screening, and on the right, to anti-screening.	8
1.3	Summary of the measurements of the running coupling of α_s as a function of the energy scale, Q [9].	8
1.4	Schematic view of the QCD phase diagram. Figure extracted from [20].	11
1.5	Space-time evolution of a heavy-ion collision. On the right, the evolution on the laboratory frame is displayed. Image from [24].	12
1.6	Visual representation of the collision of two heavy-ion nuclei.	13
1.7	Description of the formation of the anisotropic flow [36].	16
1.8	Event display of a CMS Pb–Pb event recorded in November 14 2010 showing the energy imbalance of a pair of jets caused by jet quenching [46].	18
1.9	Nuclear modification factor of J/ψ mesons as a function of the number of participants in a heavy-ion collision. In red, results by the STAR collaboration [56], in orange, by PHENIX [57], both measured in Au–Au collisions at $\sqrt{s_{NN}} = 200$ GeV. In blue, results by ALICE [63] in Pb–Pb collisions at 5.02 TeV. Plot generated using [64].	20
1.10	Nuclear modification factor for charged particles measured at $\sqrt{s_{NN}} = 5.02$ TeV by the ALICE (coloured) and CMS (black) collaborations in p–Pb and central Pb–Pb collisions [66].	21
1.11	Ratio of the production yield of K_s^0 (in black), Λ (in blue), Ξ (in green), and Ω (in red) particles to pion yields measured in pp, p–Pb and Pb–Pb collisions as a function of charged-particle multiplicity [43]. pp, p–Pb and Pb–Pb measurements are represented with circles, diamonds, and squares, respectively. Results are compared to PYTHIA 8 [70], DIPSY [71] and EPOS LHC [72] predictions.	22
1.12	Measurements performed by CMS of the $(\Delta\eta, \Delta\varphi)$ two-particle correlations in different collision systems: (a) Pb–Pb [69]; (b) p–Pb [69]; (c) pp with high multiplicity [68]; (d) pp with low multiplicity [68].	23

2.1	(a) PDFs at $Q^2 = 2 \text{ GeV}^2$ from the MSHT20aN ³ LO [90] set. (b) Kinematic coverage in the (x, Q^2) plane of key experiments contributing to PDF determinations. Figure from [9].	27
2.2	QCD evolution from DGLAP to BFKL and saturation.	29
2.3	$b\bar{b}$ cross section per unit of rapidity measured in pp collisions at midrapidity as a function of the centre-of-mass energy. The figure is extracted from [115].	30
2.4	Ratio between the p_T -differential cross sections at midrapidity ($ y < 0.5$) of prompt Λ_c baryons and D^0 mesons in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ compared with the predictions from the PYTHIA Monash tune, the PYTHIA CR-BLC modes 0, 2 and 3, SHM+RQM, Catania, QCM, and POWLANG models in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ [125].	31
2.5	Colour strings at leading and beyond leading colour. Image adapted from [134].	33
2.6	(a) D^0 cross section measured at midrapidity at $\sqrt{s} = 13 \text{ TeV}$. (b) D^{*+} cross section measured at midrapidity at $\sqrt{s} = 13 \text{ TeV}$ [125]. (c) B^0 cross section measured at midrapidity at $\sqrt{s} = 13.6 \text{ TeV}$ [142]. Experimental results are compared with theoretical predictions.	36
2.7	Example of a SPS (left) and a DPS (right) process involving two $c\bar{c}$ pairs. Figure taken from [155].	38
2.8	Total cross section of DPS events in the process $pp \rightarrow c\bar{c} c\bar{c}$ (in purple) compared to the cross section of the SPS processes corresponding to $pp \rightarrow c\bar{c}$ (in black) and $pp \rightarrow c\bar{c} c\bar{c}$ (in purple, dashed) as a function of $\sqrt{s}$ [155].	39
2.9	Summary of the measurements of σ_{eff} up to date. Figure from [158]. . .	40
2.10	Summary of the measurements of the double cross sections measured by LHCb in [159]. On the upper pannel, the opposite-sign open charm hadron pairs are presented. In the middle pannel, the like-sign open charm hadron pairs are included, and the last pannel corresponds to the pairs composed of a J/ψ meson and an open charm hadron. These last pairs are compared to theoretical predictions.	41
2.11	Left pannel: Double J/ψ cross section per unit of $ \Delta y $ as a function of $ \Delta y $. Right pannel: Double J/ψ cross section per unit of $ \Delta\phi $ as a function of $ \Delta\phi $. The measurements are compared to the expected SPS and DPS distributions. Plots extracted from [163].	42
2.12	Sketch of the EPPS16 parametrisation of the nuclear PDFs as a function of the Bjorken- x [167].	43
2.13	Expected contribution to the total energy loss (in red) produced by collisional (in blue) and radiative processes (in green) for (a) charm and (b) beauty quarks in central Pb–Pb collisions at $\sqrt{s_{\text{NN}}} = 2.76 \text{ TeV}$ [176].	45

2.14	Average R_{AA} (left) and elliptic flow (right) of prompt D mesons in the 0-10% centrality class measured at $\sqrt{s_{NN}} = 5.02$ TeV [171]. The experimental data is compared with model predictions.	48
3.1	Schematic view of the LHC complex with its different accelerators and experiments.	50
3.3	Layout of the ITS2 [203].	52
3.2	Overview of ALICE in Run 3. Image taken from the ALICE image repository.	53
3.4	Layout of the TPC [194].	54
3.5	Particle identification using the TPC of the pp data collected in 2022 at $\sqrt{s} = 13.6$ TeV. Particles are identified by fitting to a Bethe-Bloch distribution. Figure taken from the ALICE Figure repository.	55
3.6	Layout of one of the modules of the TRD [208].	56
3.7	Layout of the TOF detector with a close-up in one of its supermodules [210].	57
3.8	Layout of the FIT detector. Figure taken from the ALICE Figure repository.	58
3.9	Layout of the muon spectrometer in Run 3.	59
3.10	Layout of the MFT. Image taken from the ALICE figure repository.	60
3.11	Internal structure of the front absorber and the beam pipe shield. Figure extracted from [216].	60
3.12	On the left, the quadrant structure of stations 1 and 2 of the MCH. On the right, the slat structure of stations 3, 4, and 5 [216].	61
3.13	Diagram of the principle behind MWPCs [216].	62
3.14	Functional flow of O^2 , from the raw data collected in the detectors to its final storage. Image taken from [223].	64
3.15	Schematic view of connections between different AOD tables. Image taken from [223].	66
4.1	J/ψ acceptance. Figures from [227].	68
4.2	Muon track types. Figure extracted from [232].	70
4.3	Schematic representation of the MCH showing the different responses of the detector to the passage of a track through Station 1. Adapted from [230].	71
4.4	Online cluster distributions of the MCH chambers when run 544122 was collected.	72
4.5	Offline results obtained with O^2 Physics for the clusters in chambers 2, 6, 8 and 9.	73
4.6	MCH tracking efficiency of the data corresponding to run 544122 as a function of η and φ (part 1).	74
4.7	MCH tracking efficiency of the data corresponding to run 544122 as a function of η and φ (part 2).	75

4.8	MCH tracking efficiency calculated using the initial MC corresponding to run 544122 as a function of η and φ (part 1).	76
4.9	MCH tracking efficiency calculated using the initial MC corresponding to run 544122 as a function of η and φ (part 2).	77
4.10	MCH tracking efficiency calculated using an anchored MC corresponding to run 544122 as a function of η and φ for chambers 2 and 6.	78
4.11	Muon tracking efficiency per chamber for run 544122. The initial MC is represented in green, the updated anchored MC in red, and the data in blue.	78
4.12	Muon tracking efficiency per station for run 544122. The initial MC is represented in green, the updated MC in red, and the data in blue. . . .	79
4.13	Integrated muon tracking efficiency for all runs considered. The initial MC is represented in green, the updated MC in red, and the data in blue. . . .	79
4.14	Integrated muon tracking efficiency for all Pb–Pb runs recorded in 2023 as a function of (a) p_T ; (b) η and (c) φ . The data is represented in blue and the anchored MC in red.	80
5.1	General structure of a high-energy physics analysis. This sketch is a modified version of the one from [233].	83
5.2	Description of the tracking process. The figure contains a schematic representation of the ITS, with hits represented as circles and traces as green lines. The dashed orange line corresponds to a fake track. Image taken from [233].	87
5.3	Track-to-collision association mechanism for detectors that take data in a set time frame (a) and for detectors with a gaussian-like resolution (b). In red, detected track time, and in blue, detected collision time [225].	89
5.4	Upper row: $n\sigma$ TPC distributions for (a) pions and (b) kaons corrected using a NN as a function of the momentum. Bottom row: $n\sigma$ TOF distributions for (c) pions and (d) kaons as a function of the momentum.	91
5.5	Schematic view of the main components of the HF workflow, from the initial skimming to the tasks developed by analysers. Image adapted from [225].	92
5.6	Structure of a binary decision tree.	94
5.7	Structure of a BDT.	95
5.8	Representation of the decay of a (a) prompt and (b) non-prompt D^0 meson.	98
5.9	Decay length distributions of background, prompt D^0 and non-prompt D^0 candidates in the $4 < p_T < 5$ GeV/ c range.	98

5.10	Product of the distance of closest approach distributions of the two decay particles for the background, prompt D^0 and non-prompt D^0 candidates in the $4 < p_T < 5$ GeV/ c range.	99
5.11	Cosine of pointing angle distributions of background, prompt D^0 and non-prompt D^0 candidates in the $4 < p_T < 5$ GeV/ c range.	100
5.12	Linear correlations between the training variables, the invariant mass and the p_T , for prompt D^0 mesons (a) and combinatorial background (b). The colours, from red (fully correlated) to blue (fully anti-correlated), indicate the degree of correlation.	102
5.13	Feature importances per variable to separate between prompt D^0 (green), non-prompt D^0 (orange) and background (blue) measured in the $4 < p_T < 5$ GeV/ c region.	103
5.14	Distributions of the ML output scores for the prompt (a), non-prompt (b) and combinatorial background (c) templates measured in the $4 < p_T < 5$ GeV/ c region.	105
5.15	Example of a set of ROC curves for a three-class model trained using candidates in the $4 < p_T < 5$ GeV/ c region. In blue, curves involving the background and non-prompt distributions, in orange, the background and prompt distributions, and in green the prompt and non-prompt distributions.	106
5.16	Output trigger table containing all selected events per trigger corresponding to run 553756.	108
6.1	Reconstruction efficiencies as a function of p_T after the preselections for prompt (red) and non-prompt (blue) D^0 mesons in the $0 < p_T < 24$ GeV/ c range.	113
6.2	Significances in the 3σ region, invariant mass widths, and efficiencies as a function of the background BDT score for the $2 < p_T < 2.5$ GeV/ c range. The selected score is marked with a vertical orange line.	118
6.3	Fits to the invariant mass distributions in the $0 < p_T < 24$ GeV/ c range (part 1). The black points represent the data, the blue line represents the total fit function, the red line is the background, the green line is used for the reflected D^0 candidates, and the yellow figure represents the D^0 signal.	119
6.4	Fits to the invariant mass distributions in the $0 < p_T < 24$ GeV/ c range (part 2). The black points represent the data, the blue line represents the total fit function, the red line is the background, the green line is used for the reflected D^0 candidates, and the yellow figure represents the D^0 signal.	120

6.5	Fits to the invariant mass distributions in the $0 < p_T < 24 \text{ GeV}/c$ range (part 3). The black points represent the data, the blue line represents the total fit function, the red line is the background, the green line is used for the reflected D^0 candidates, and the yellow figure represents the D^0 signal.	121
6.6	Comparison between the widths (left) and means (right) obtained fixing the width to the case without applying selection cuts on the non-prompt BDT score and leaving the parameter free for a non-prompt BDT score of 0.45.	122
6.7	Prompt (red) and non-prompt (blue) acceptances multiplied by the efficiencies per p_T bin after selecting the working point.	123
6.8	Results of applying the cut-variation procedure. Top-left: raw yields as a function of the cut set. They are fitted to extract the prompt and non-prompt contributions. Top-right and bottom-left: acceptance-times efficiency and fraction as a function of the cut set, respectively. In both plots, the prompt contribution is plotted in red and the non-prompt contribution is in blue. Bottom-right: covariance matrix.	124
6.9	Raw yield systematic uncertainty and its ratio w.r.t. the central value.	125
6.10	BDT selection systematic uncertainty and its ratio w.r.t. the central value.	126
6.11	DCA resolution systematic uncertainty and its ratio w.r.t. the central value.	127
6.12	Results of the measurement of the non-prompt D^0 fraction at $\sqrt{s} = 13.6 \text{ TeV}$ (in black) and at $\sqrt{s} = 13 \text{ TeV}$ (in red) [252]. The data is compared to the theoretical predictions by PYTHIA [70] and EPOS [72].	128
6.13	Ratio of the data over the considered PYTHIA tune and EPOS predictions as a function of p_T	130
6.14	Results of the non-prompt fraction compared to FONLL [106] predictions as a function of the transverse-momentum.	131
7.1	Acceptance-times-efficiency of the preselection cuts imposed by the OTS.	135
7.2	Acceptance-times-efficiency of the preselection cuts imposed by the OTS. Prompt candidates are shown in red, non-prompt candidates in blue, and all candidates in black.	136
7.3	BDT output distribution for the model trained in the $5 < p_T < 7 \text{ GeV}/c$ region.	137
7.4	ROC distribution in the $5 < p_T < 7 \text{ GeV}/c$ region for the training and testing sets. The areas under the curves (AUCs) are also shown.	138
7.5	Invariant mass fit of the D^0 raw yield distribution in the (a) $1 < p_T < 5 \text{ GeV}/c$ (b) $5 < p_T < 24 \text{ GeV}/c$ regions. The total fit function is marked in blue, the background in red, and the reflected D^0 distribution in green. The reflected D^0 candidates were not included in the $5 < p_T < 24 \text{ GeV}/c$ range.	140

7.6	Two-dimensional plots of the invariant mass distributions of the D^0 pair candidates in the $1 < p_T < 24$ GeV/ c range for (a) Like-sign pairs (b) Opposite-sign pairs. The fit to the invariant mass is represented with a red mesh.	140
7.7	Projections of the two-dimensional plots of the invariant mass distributions of the D^0 pair candidates in the $1 < p_T < 24$ GeV/ c range for (a) Like-sign pairs (b) Opposite-sign pairs. The figures on the left represent the first D^0 candidate, and the ones on the right the second.	142
7.8	Efficiency map of the D^0 candidates as a function of the rapidity and the transverse momentum.	143
7.9	Distributions of the primary vertex (PV) contributors in data (blue) and MC (red).	143
7.10	Acceptance-times-efficiency distributions of the prompt (red), non-prompt (blue) after applying the BDT selections on top of the OTS.	144
7.11	Weighted projections of the two-dimensional plots of the invariant mass distributions of the D^0 pair candidates in the $1 < p_T < 24$ GeV/ c range for (a) Like-sign pairs (b) Opposite-sign pairs. The figures on the left represent the first D^0 candidate, and the ones on the right the second.	145
7.12	Ratio of the normalised invariant mass distributions before and after correcting by efficiencies for like-sign pairs (left) and opposite-sign pairs (right).	147
7.13	Results of the cut-variation method in the $1 < p_T < 2$ GeV/ c range.	148
7.14	Comparison of the non-prompt D^0 fractions obtained using the 2022 and 2024 datasets. Results are also compared with the Run 2 results measured at $\sqrt{s} = 13$ TeV.	149
7.15	Prompt fraction obtained with the acceptance-times-efficiencies corresponding to the double D^0 analysis fitted to a polynomial of order 0.	149
7.16	Corrected yields for the different trials varying the background fitting functions and invariant mass ranges in the $1 < p_T < 24$ GeV/ c range for (a) Like-sign pairs (b) Opposite-sign pairs.	151
7.17	Ratio of the acceptance-times-efficiencies using worse resolution configurations of the track-tuner over the default configuration.	152
7.18	Calculation of the systematic uncertainty tied to the BDT selection.	152
7.19	Cross section of the production of prompt $D^0 D^0$ (upper point) and $D^0 \bar{D}^0$ (bottom point) pairs measured at midrapidity at $\sqrt{s} = 13.6$ TeV.	155
7.20	PYTHIA 8 simulation of the ratio of OS over LS pair production as a function of p_T and y . The coverage of the LHCb measurement [159] is marked in blue, while our coverage is indicated in red.	156
7.21	Ratio of the single-over-double D^0 production measured at midrapidity for like-sign (upper point) and opposite-sign (bottom point) D^0 pairs.	156
7.22	Value of the reflected-over-signal fraction for single D^0 candidates. The reflections were removed for all candidates with $p_T < 5$ GeV/ c	157

7.23	Projections of the two-dimensional plots of the invariant mass distributions of the D^0 pair candidates in the (a) $0 < p_T^{DD} < 2$, (b) $2 < p_T^{DD} < 4$, and (c) $4 < p_T^{DD} < 6$ GeV/ c range for opposite-sign pairs. In all cases, the left and right panels represent the projection on the invariant mass of the two D^0 candidates that form the pair.	159
7.24	Projections of the two-dimensional plots of the invariant mass distributions of the D^0 pair candidates in the (a) $6 < p_T^{DD} < 12$ and (b) $p_T^{DD} > 12$ GeV/ c range for opposite-sign pairs. In all cases, the left and right panels represent the projection on the invariant mass of the two D^0 candidates that form the pair.	160
7.25	Projections of the two-dimensional plots of the invariant mass distributions of the D^0 pair candidates in the (a) $0 < p_T^{DD} < 2$, (b) $2 < p_T^{DD} < 4$, and (c) $4 < p_T^{DD} < 6$ GeV/ c range for Like-sign pairs. In all cases, the left and right panels represent the projection on the invariant mass of the two D^0 candidates that form the pair.	161
7.26	Projections of the two-dimensional plots of the invariant mass distributions of the D^0 pair candidates in the (a) $6 < p_T^{DD} < 12$ and (b) $p_T^{DD} > 12$ GeV/ c range for Like-sign pairs. In all cases, the left and right panels represent the projection on the invariant mass of the two D^0 candidates that form the pair.	162
7.27	Differential cross section for opposite-sign pairs measured at midrapidity in pp collisions at $\sqrt{s} = 13.6$ TeV.	163
7.28	Differential cross section for like-sign pairs measured at midrapidity in pp collisions at $\sqrt{s} = 13.6$ TeV.	163
7.29	Differential cross section ratios of opposite-sign over like-sign pairs measured at midrapidity in pp collisions at $\sqrt{s} = 13.6$ TeV.	164
7.30	Distribution of $ \Delta y $ for (a) like-sign pairs, (b) opposite-sign pairs. . .	165
7.31	Distribution of $ \Delta \varphi $ for (a) like-sign pairs, (b) opposite-sign pairs. . .	165
7.32	$ \Delta \varphi $ distribution computed with a toy DPS MC.	166
A.1	Left figures: MCH tracking efficiency of the data. Right figures: MCH tracking efficiency of the anchored MC for the same chambers. All plots correspond to run 544122 and are measured as a function of η and φ (part 1).	190
A.2	Left figures: MCH tracking efficiency of the data. Right figures: MCH tracking efficiency of the anchored MC for the same chambers. All plots correspond to run 544122 and are measured as a function of η and φ (part 2).	191
A.3	Left figures: MCH tracking efficiency of the data. Right figures: MCH tracking efficiency of the anchored MC for the same chambers. All plots correspond to run 544122 and are measured as a function of η and φ (part 3).	192

A.4	Left figures: MCH tracking efficiency of the data. Right figures: MCH tracking efficiency of the anchored MC for the same chambers. All plots correspond to run 544122 and are measured as a function of η and φ (part 4).	193
C.1	Prompt and non-prompt efficiencies computed with the selections that were approved as Preliminary compared to the ones presented in this thesis.	198
C.2	Projections of the two-dimensional plots of the invariant mass distributions of the D^0 pair candidates in the $2 < p_T < 24$ GeV/ c range for (a) Like-sign pairs (b) Opposite-sign pairs. The figures on the left represent the first D^0 candidate, and the ones on the right the second.	199
C.3	Integrated cross sections of like-sign and opposite-sign pair D^0 mesons measured in the $2 < p_T < 24$ GeV/ c range that were approved as Preliminary.	200
C.4	Ratio of the single and double of like-sign and opposite-sign pair D^0 mesons measured in the $2 < p_T < 24$ GeV/ c range that were approved as Preliminary.	200

LIST OF TABLES

1.1	Summary of the observations per collision system of the probes introduced in section 1.5 at LHC energies.	24
3.1	Table showing the main upgrades of ITS2 compared to ITS1 [203]. . . .	54
5.1	List of track selection conditions for global tracks [225].	88
5.2	General characteristics of D^0 mesons. Information obtained from [9]. .	96
5.3	List of trigger preselections for D^0 mesons.	108
6.1	Rectangular preselection cuts on the number of σ of the TOF, the TPC and the combination of both detectors, the cosine of pointing angle, and its projection in the xy plane.	112
6.2	Chosen variables to train the models to separate prompt D^0 , non-prompt D^0 and combinatorial background candidates.	114
6.3	Approximate number of available candidates in each class to perform the BDT training per p_T range. The final values used for the training and testing are marked in bold.	114
6.4	Values of the average ROC AUC of all trained models.	115
6.5	Summary of the systematics vs. p_T (transposed).	129
7.1	Chosen variables to train the models to separate signal D^0 and combinatorial background candidates.	137
7.2	Approximate number of available candidates in each class to perform the BDT training per p_T range. The final values used for the training and testing are marked in bold.	138
7.3	Values of the average ROC AUC of all trained models.	139
7.4	Raw yields of the D^0 pairs in the $1 < p_T < 24$ GeV/ c range. The first row corresponds to the like-sign pairs and the second row to the opposite-sign pairs.	141
7.5	Check on the signal-signal yield computed with the efficiencies calculated candidate-by-candidate vs. using the integrated efficiency values.	146
7.6	Comparison of the widths and means for the efficiency-corrected and raw yields.	147
7.7	BDT selection systematic uncertainty as a function of p_T	153

7.8	Summary of the systematic uncertainties for single and double D^0 candidates.	154
7.9	Integrated cross sections of the associated production of D^0 -meson pairs at $\sqrt{s} = 13.6$ TeV in the rapidity region $ y < 0.5$	155
7.10	Reflected-over-signal fractions per pair- p_T range.	158
7.11	Systematic uncertainties on the yield extraction as a function of the p_T of the pair.	158
7.12	Comparison of the cross sections of the associated production of D^0 -meson pairs computed in one pair- p_T bin vs. multiple bins at $\sqrt{s} = 13.6$ TeV in the rapidity region $ y < 0.5$	164
C.1	Values of the average ROC AUC of all trained models in the Preliminary results vs. this thesis.	197
C.2	Comparison of the integrated cross sections and $\mathcal{R}_{DD}$ in the $2 < p_T < 24$ GeV/ c using the selections of this thesis vs. the Preliminary results for OS and LS pairs. In all cases, the first uncertainty is statistical and the second is systematic.	199

