

分类号_____

密级_____

U D C_____

编号_____

華中師範大學

博士学位论文

题目：LHC-ALICE 开粲强子产生
及非瞬时 D^0 介子椭圆流的实验研究

学位申请人姓名：张彪
申请学位学生类别：全日制博士
申请学位学科专业：粒子物理与原子核物理
指导教师姓名：王亚平 教授

博士学位论文

论文题目：LHC-ALICE 开粲强子产生及非
瞬时 D^0 介子椭圆流的实验研究

论文作者：张彪

指导教师：王亚平 教授

学科专业：粒子物理与原子核物理

研究方向：相对论重离子物理

华中师范大学物理科学与技术学院

2023 年 5 月

Dissertation

Measurements of open charm hadron production and non-prompt D^0 elliptic flow with the ALICE at the LHC

By

Biao Zhang

Supervisor: Yaping Wang

Specialty: Particle Physics and Nuclear Physics

Research Area: Heavy Ion Collisions

College of Physical Science and Technology

Central China Normal University

May. 2023



华中师范大学学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的科研成果。除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

作者签名：张彪

日期：2023年6月2日

学位论文版权使用授权书

学位论文作者完全了解华中师范大学有关保留、使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属华中师范大学。学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许学位论文被查阅和借阅；学校可以公布学位论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。（保密的学位论文在解密后遵守此规定）

保密论文严格按照《华中师范大学涉密研究生与涉密学位论文管理办法》进行管理。非保密论文不属于保密范围，适用本授权书。

因特殊情况需延期发布的论文，同意论文提交后延期☐半年；☐一年；☐二年发布。到期后适用本授权书。

作者签名：张彪

日期：2023年6月2日

导师签名：

日期： 年 月 日

摘 要

人类的梦想是不会止步的！随着科学技术和人类认知水平的提高，探索人类起源和世界构成的领域衍生出了许多不同的学科。宇宙大爆炸后约 10^{-5} s 会形成一种新的物质形态：夸克胶子等离子体（QGP）。研究这种物质有助于我们理解宇宙的起源和形成。然而有一个问题一直困扰着人类：如何在实验室中制造这种高温高密度的强相互作用物质？在 20 世纪 70 年代，李政道先生提出了利用接近光速的重离子碰撞来研究夸克胶子等离子体的想法。基于这个伟大的想法，重离子碰撞实验开始兴起！其中最著名的是美国布鲁克海文国家实验室（BNL）的相对论重离子对撞机（RHIC）上的螺旋管径迹探测器（STAR）和欧洲核子研究中心（CERN）的大型强子对撞机（LHC）上的大型重离子实验（ALICE）。

本论文的研究载体是 LHC 上的 ALICE 实验，旨在通过分析 ALICE Run2 实验数据来探究夸克胶子等离子体的性质。由于重离子碰撞的演化过程复杂且短暂，我们只能通过重建并分析碰撞末态的强子信息来反推 QGP 的性质。这些强子信息可分为软探针和硬探针两类，其中软探针包括粒子各向异性流、粒子产额比等，而硬探针则包括喷注（jet）、重味强子谱和高横动量粒子谱等。其中，重味强子指带有重味夸克（charm）和底（beauty）的强子。重味夸克被视为研究 QGP 的理想探针，因为它们主要产生于碰撞早期的硬散射过程，并经历了整个碰撞系统的演化过程。实验上主要通过研究重味强子的产额比、各向异性流以及核修正因子来研究强子化机制、QGP 的输运性质和部分子在 QGP 介质中的能损效应。

目前观测到的实验现象不足以证明小系统（如质子-质子和质子-重离子对撞）中也产生了 QGP。通常，小系统的实验测量可用于检验微扰量子色动力学（pQCD）并作为核核对撞实验的参考。重味强子的产额通常可以用基于微扰量子色动力学的因子化理论来描述。在因子化理论中，重味强子的产额可以由对撞入射质子的部分子分布函数、重味夸克的硬散射产生截面以及重味夸克的碎裂函数三部分卷积而成。其中，重味夸克碎裂成重味强子的碎裂函数和入射质子的部分子分布函数是非微扰过程，只能通过实验测量来约束，而硬散射截面可以通过 pQCD 计算得出。值得一提的是，长期以来碎裂函数一直被认为是普适的，即此过程不依赖于碰撞系统和能量。基于正负电子（ e^+e^- ）和电子-质子（ep）对撞实验提取的碎裂函数约束的次领头阶微扰计算可以成功描述 LHC 能区的粲介子产额。但最近 ALICE 实验上质心系能量为 5.02 和 7 TeV 的质子-质子（pp）对撞中测量的粲重子-介子比 Λ_c^+/D^0 在低横动量区间却显著高于基于 e^+e^- 碎裂函数参数化的事件产生器 PYTHIA 8 的模拟结果。类似的现象也在 ALICE 的中心快度区和 LHCb 的向前快度区的质子-铅核（p-Pb）对撞数据中被观察到。最后，结合最近 ALICE pp 对撞中测量的其它粲重子介子比（如 Ξ_c^+/D^0 ）的产额进一步表明碎裂函数并不是普适的。值得一提的是，以上 ALICE 实验结果使用的是无偏事件（Minimum Bias events），其带电粒子多重

数约为整体事件的均值。测量不同带电粒子多重数区间的 Λ_c^+/D^0 产额比可以系统地研究碎裂函数的非普适现象和重味重子的形成机制，有助于进一步约束理论模型。

目前实验上对于中心快度且低横动量区底夸克的测量一直十分缺乏。ALICE 实验组通过测量底强子衰变产物（如底夸克衰变产生的电子和缪子，以及非瞬时的 D 介子和夸克偶素）来间接研究底夸克的性质。重味夸克强子化模型可以通过测量非瞬时 D^0 介子产额随带电粒子多重数的依赖关系来进一步完善。而在重离子实验中，由于底夸克质量较大，其在 QGP 中是否会被热化一直是实验上的一个疑问。最近，ALICE 实验组观测到底强子衰变电子的椭圆流显著大于零，这表明底夸克也有很大可能参与了热化过程。然而，底强子衰变电子的横动量仅反映了较高横动量区间的底强子集体行为，无法充分反映底夸克的热化行为。基于目前的探测器空间分辨率和亮度，测量强子道的衰变产物即非瞬时粲介子（ D^0 ）的椭圆流可以更好地研究底夸克与 QGP 之间的相互作用。

本论文旨在通过测量 LHC 能区 ALICE 实验中粲强子和非瞬时 D^0 介子的产额及椭圆流，研究重味夸克强子化机制和重味夸克与 QGP 的相互作用。针对重味夸克强子化机制，利用 2016 至 2018 年采集的质心系能量为 13 TeV 的 pp 碰撞数据（约 18 亿次事件）测量了瞬时粲重子 Λ_c^+ 产额随带电多重数的依赖性关系，以及非瞬时 D^0 介子所占比例随带电多重数的依赖性和非瞬时 D^0 介子的产额。对于底夸克与 QGP 的相互作用的研究，利用 2018 年采集的质心系能量为 5.02 TeV 的 Pb – Pb 碰撞数据测量了非对心碰撞中的非瞬时 D^0 介子的椭圆流。

关键词：夸克胶子等离子体 (QGP)，重离子对撞实验，大型重离子对撞实验 (ALICE)，重味夸克，强子化，非瞬时 D^0 介子，椭圆流

Abstract

The human dream will not stop! With the development of science and technology and human cognitive level, exploration of the origin of human beings and the composition of the world has given rise to many different disciplines. The theory of the Big Bang indicates that the universe was born from a singularity explosion, and in the early stages of the explosion (0.1 fm - 10 fm), a new form of matter is formed: quark-gluon plasma (QGP). Therefore, studying this substance helps us to understand the origin and formation of the universe. However, there has always been a question that has perplexed humanity: how to produce this high-temperature and high-density substance with strong interactions in the laboratory? In the 1970s, Dr. Lee Tsung-Dao proposed the idea of colliding heavy ions approaching the speed of light to study quark-gluon plasma. Based on this great idea, heavy ion collision experiments began to rise! The most famous ones include the Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Laboratory (BNL) in the United States with the Solenoidal Tracker at RHIC (STAR) and the Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN) with a large heavy ion experiment (ALICE).

The subject of this thesis is the ALICE experiment at the LHC, aiming to explore the properties of quark-gluon plasma by analyzing the experimental data from ALICE Run2. Due to the complex and fleeting evolution process of heavy-ion collisions, we can only rely on reconstructing and analyzing the hadron information at the collision endpoint to infer the properties of QGP. These hadron information can be classified into two types of probes: soft probes and hard probes. Soft probes include particle anisotropic flow, particle yield ratios, and so on, while hard probes include jets, heavy-flavor hadrons transverse momentum (p_T) spectrum, and high p_T particle spectra. Among them, heavy-flavor hadrons refer to hadrons carrying heavy-flavor quarks charm and beauty, which are considered as ideal probes for studying QGP because they mainly originate from the hard scattering process in the early stage of collision and undergo the entire collision system evolution process. Experimentally, the mechanism of hadronization, transport properties of QGP, and energy loss effect of partons in the QGP medium are mainly studied by analyzing the yield ratios, anisotropic flow, and nuclear modification factor of heavy-flavor hadrons.

The currently observed experimental phenomena are not sufficient to prove the existence of quark-gluon plasma (QGP) in small systems, such as proton-proton and proton-nucleus collisions. Typically, experimental measurements in small systems can be used to test perturbative quantum chromodynamics (pQCD) and serve as a baseline for heavy-ion collision experiments. The production of heavy-flavored

hadrons can usually be described by factorization theory based on pQCD. In factorization theory, the production of heavy-flavored hadrons is the convolution of three parts: the parton distribution functions of the incident protons, the cross-section of hard scattering between partons and heavy quarks, and the fragmentation function that describes the fragmentation of heavy quarks into heavy-flavored hadrons. Among them, the fragmentation function and the parton distribution functions of the incident protons are non-perturbative processes and can only be constrained by experimental measurements, while the hard scattering cross-section can be calculated by perturbative QCD. It is worth mentioning that the quark fragmentation function has long been considered universal, meaning that this process does not depend on the collision system and energy. However, despite the successful description of the charmonium production at the LHC energy region by the next-to-leading order perturbative calculation of the fragmentation function extracted from e^+e^- and ep collision experiments, the production of the Λ_c^+/D^0 baryon measured in proton-proton (pp) collisions at center-of-mass energies of 5.02 and 7 TeV at the ALICE experiment is significantly underestimated by the PYTHIA 8 calculation, which tuned the fragmentation function from the e^+e^- and ep collision experiments. Similar phenomena have also been observed in the proton-lead collisions at the central rapidity region of ALICE and the forward rapidity region of LHCb. Other heavy-flavored hadrons such as $\Xi_c^{+,0}/D^0$ measured recently in pp collisions also confirmed that the fragmentation function is not universal for the collision system. It is worth mentioning that the ALICE experiment used the minimum bias events, with the charged particle multiplicity being the average of the entire event. Measurement of the Λ_c^+/D^0 yield ratio as a function as charged multiplicity can help us further understand the non-universal phenomena of fragmentation functions and the formation mechanism of open baryons in more detail, which will also better constrain the theoretical models.

In addition, there has been a lack of measurement for bottom quarks in the central rapidity and low transverse momentum region in current experiments. The ALICE collaboration indirectly studies the properties of bottom quarks by measuring the decay products of bottom hadrons, such as electrons and muons from bottom quark decays, as well as non-prompt D mesons and quarkonia. In proton-proton collisions, studying the multiplicity of charged particles produced with bottom hadrons can help improve hadronization models. However, in heavy-ion collisions, it remains an experimental question whether bottom quarks with their larger mass are thermalized in the quark-gluon plasma (QGP). Recently, the ALICE collaboration observed a significant positive elliptic flow of electrons from bottom hadron decays, indicating that bottom quarks could also participate in the thermalization process. However, the transverse momentum of these electrons only reflects the collective behavior of high transverse momentum bottom hadrons and cannot fully reflect the thermalization behavior of bottom quarks. Based on the

current detector spatial resolution and luminosity, measuring the elliptic flow of non-prompt D mesons can better study the interaction between bottom quarks and the QGP.

This thesis aims to study the hadronization mechanisms of heavy quarks and their interaction with the quark-gluon plasma (QGP) by measuring the production yields and elliptic flow of charmed hadrons and non-prompt D^0 mesons in the ALICE experiment at the LHC energy region. For the hadronization mechanisms of heavy quarks, the dependence of the prompt Λ_c^+ production on the charged particle multiplicity and the proportion of non-prompt D^0 mesons as a function of the charged particle multiplicity, as well as the non-prompt D^0 meson yield, were measured using proton-proton collision data collected from 2016 to 2018 at a center-of-mass energy of 13 TeV (about 1800 million events). For the study of the interaction between bottom quarks and QGP, the elliptic flow of non-prompt D^0 mesons in non-central collisions was measured using lead-lead collision data collected in 2018 at a center-of-mass energy of 5.02 TeV.

Keywords: Quark gluon plasma; Heavy-ion; ALICE, Heavy quark; Hadronization; Non-prompt D^0 -meson; Elliptic flow

Contents

摘 要	I
Abstract	III
List of Figures	IX
List of Tables	XVIII
1 General introduction	1
1.1 Quantum Chromodynamics	2
1.2 Quark Gluon Plasma	4
1.3 Heavy-ion collision	6
1.4 Experimental observables	8
1.4.1 Particle multiplicity	8
1.4.2 Identified-hadron abundances	9
1.4.3 Nuclear modification factor	12
1.4.4 Anisotropic flow	13
1.5 Open heavy-flavour physics	15
1.5.1 Open heavy-flavour production	16
1.5.2 Interactions with the cold-nuclear matter	20
1.5.3 Interactions with the hot-nuclear matter	22
1.5.4 Hadronisation mechanisms	28
1.6 Motivation and structure of thesis	36
2 ALICE: A Large Ion Collider Experiment	38
2.1 The ALICE detectors	38
2.1.1 Inner Tracking System	38
2.1.2 Time Projection Chamber	40
2.1.3 Time-of-Flight detector	41
2.1.4 V0 and T0 detector	42
2.2 Detector performance	43
2.2.1 tracking and vertexing reconstruction	43
2.2.2 Particle identification	44
2.2.3 Centrality determination	45

2.3 The ALICE offline framework	47
2.4 Data samples	48
3 Advanced analysis strategy	50
3.1 Machine-learning based candidate selection	51
3.1.1 Boosted Decision Trees and XGBoost	51
3.1.2 BDT training with variables	53
3.1.3 BDT application and signal extraction	58
3.2 Data-driven based fraction estimation	60
4 Open charm-hadron measurements in pp collisions at $\sqrt{s} = 13$ TeV	65
4.1 Prompt Λ_c^+ baryon production in pp collisions at $\sqrt{s} = 13$ TeV	65
4.1.1 Event selection and multiplicity definition	67
4.1.2 Analysis strategy	68
4.1.3 Candidate selection and raw yield extraction	69
4.1.4 Corrections	76
4.1.5 Systematic uncertainties	82
4.1.6 Results and discussion	89
4.2 Non-prompt D-meson fraction as a function of multiplicity in pp collisions at $\sqrt{s} = 13$ TeV	97
4.2.1 Event selection and multiplicity definition	97
4.2.2 Candidate selection and raw yield extraction	98
4.2.3 Non-prompt fraction extraction	104
4.2.4 Systematic uncertainties	105
4.2.5 Results and discussion	110
4.3 Non-prompt D^0 production in pp collisions at $\sqrt{s} = 13$ TeV	116
4.3.1 Raw yield extraction	117
4.3.2 Corrections	117
4.3.3 Systematic uncertainties	120
4.3.4 Results and discussion	121
5 Elliptic flow of non-prompt D^0 in Pb – Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV	126
5.1 Event selection and data samples	127
5.2 Candidate selection and v_2 extraction	127
5.3 Non-prompt D^0 v_2 extraction	134

5.3.1 Linear fit with correlated cut sets	134
5.3.2 Linear fit with uncorrelated cut sets	136
5.4 Systematic uncertainties	140
5.4.1 Fit procedure	140
5.4.2 $f_{\text{non-prompt}}$ calculation	140
5.4.3 MC p_T shape	141
5.4.4 Scalar product denominator	148
5.5 Statistic uncertainty estimate using Jackknife method	148
5.6 Results and discussion	153
6 Conclusions and Perspectives	156
References	159
Appendix 博士论文中文简介	181
A.1 论文的研究意义	181
A.2 论文的框架	182
A.2.1 理论背景介绍	182
A.2.2 实验装置	183
A.2.3 实验分析方法	184
A.2.4 重味强子的产生	185
A.2.5 非瞬时 D^0 介子椭圆流的测量	185
A.2.6 总结与展望	185
攻读学位期间发表的学术论文	187
致 谢	188

List of Figures

1.1	Standard Model in particle physics	1
1.2	Summary of α_s measurements as a function of the transferred momentum Q . The respective degree of QCD perturbation theory used in the extraction of α_s is indicated in parentheses.	3
1.3	The QCD phase diagram as a function of the temperature and the baryon-chemical potential.	5
1.4	The space-time diagram of a heavy-ion collision.	6
1.5	Left: The charged-particle pseudorapidity density ($\langle dN_{\text{ch}}/d\eta \rangle$) is plotted as a function of the centre-of-mass energy ($\sqrt{s_{\text{NN}}}$) for various types of collisions. Right: The centrality dependence of $\langle dN_{\text{ch}}/d\eta \rangle$ for Pb – Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV is compared to different model predictions. The division of $2/\langle N_{\text{part}} \rangle$ in both figures is introduced for comparing different colliding systems.	10
1.6	Thermal fit to the identified-particle yields measured by the ALICE Collaboration in the 10% most central Pb – Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV.	11
1.7	ALICE measured the nuclear modification factors for charged particles in central (0–5%) and peripheral (70–80%) Pb – Pb collisions, as well as in minimum-bias p – Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV.	14
1.8	The measurements of the elliptic flow coefficient v_2 for charged pions, charged and neutral kaons, protons, ϕ mesons, and hyperons as a function of p_{T} in several centrality classes of Pb – Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV by ALICE.	16
1.9	The measured p_{T} -differential production cross section of prompt and non-prompt D^0 mesons in pp collisions at $\sqrt{s} = 5$ TeV.	19
1.10	The measured p_{T} -differential production cross section of prompt and non-prompt D^0 mesons in pp collisions at $\sqrt{s} = 5$ TeV.	20
1.11	The nuclear modification factor R_{pPb} of prompt D mesons and Λ_c^+ baryons in p – Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV as measured.	23
1.12	The radiative and collisional energy loss for charm and bottom quarks in the Quark-Gluon Plasma (QGP).	24
1.13	R_{AA} of prompt non-strange D mesons and charged pions, charged particles, inclusive J/ψ	25
1.14	The p_{T} -dependent v_2 and v_3 coefficients for prompt D^0 , D^+ , and D^{*+} mesons in Pb – Pb collisions.	27
1.15	Average R_{AA} and elliptic flow v_2 of prompt D^0 , D^+ , and D^{*+} mesons.	29

1.16 (a) A flux tube extends between a quark and an antiquark. (b) The movement and dis- integration of a string system.	31
1.17 $f(c \rightarrow H_c)$ as measured in e^+e^- and ep collisions by different experiments.	32
1.18 Λ_c^+/D^0 ratio in pp collisions at $\sqrt{s} = 5.02$ TeV.	32
1.19 Illustration of a multi-parton state, with a leading-colour string topology (C) and with a junction-antijunction string topology (D).	34
1.20 The interplay between coalescence and fragmentation processes.	35
2.1 ALICE schematic layout	39
2.2 Structure of the Inner Tracking System.	40
2.3 Layout of the Time Projection Chamber.	41
2.4 Modular structure of the ALICE TOF detector.	42
2.5 The impact parameter resolution as a function of p_T in pp collisions at $\sqrt{s} = 7$ TeV, in p – Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV and in Pb – Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV.	44
2.6 PID from TPF and TPC in ALICE	46
2.7 Distribution of the V0 amplitudes measured in Pb – Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV.	48
3.1 A simple Decision Tree Structure	51
3.2 A sketch of the $D^0 \rightarrow K^- \pi^+$ decay, representing the primary, secondary vertices and some topological quantities	53
3.3 The $D^0 \rightarrow K^- \pi^+$ candidates distributions of the decay length related variables(left col- umn) and impact parameter related variables (right column) for three components, which extracted from the data (background) and MC production (prompt and feed-down) in pp collisions at $\sqrt{s} = 13$ TeV in the $4 < p_T < 6$ GeV/c.	55
3.4 The $D^0 \rightarrow K^- \pi^+$ candidates distributions of the cosine point angle (left) and the projec- tion in the transverse plane (right column) for three components, which extracted from the data (background) and MC production (prompt and feed-down) in pp collisions at $\sqrt{s} = 13$ TeV in the $4 < p_T < 6$ GeV/c.	56
3.5 The $D^0 \rightarrow K^- \pi^+$ candidates distributions of the $n\sigma_{\text{comb}}^\pi$ (left column) and $n\sigma_{\text{comb}}^K$ of each daughter tracks for three components, which extracted from the data (background) and MC production (prompt and feed-down) in pp collisions at $\sqrt{s} = 13$ TeV in the $4 < p_T < 6$ GeV/c.	57

3.6	Distributions of ML output score of Feed-down for combinatorial background candidates, prompt D0 mesons, and Feed-down D ⁰ in pp collisions at $\sqrt{s} = 13$ TeV in the $4 < p_T < 6$ GeV/c.	58
3.7	ROC curves for the models trained for D ⁰ in pp collisions at $\sqrt{s} = 13$ TeV in the $4 < p_T < 6$ GeV/c.	59
3.8	Invariant-mass distribution of D ⁰ candidates. The left (right) column corresponds to the prompt (non-prompt) D ⁰ meson candidates dominated sample.	60
3.9	((Acc $\times$ ϵ)) factors as a function of the cut set (top-left panel), correlation matrix (top-right panel), raw-yield as a function of the cut set with the prompt and non-prompt contributions obtained from the minimisation procedure (bottom-left panel), and the resulting prompt and non-prompt fractions as a function of the cut set (bottom-right panel) for the D ⁰ mesons in the $2 < p_T < 4$ GeV/c interval for the MB class of events.	64
4.1	The distributions of the total number of SPD tracklets and the V0M amplitude in pp collisions at $\sqrt{s} = 13$ TeV.	68
4.2	Distributions for signal and background of Λ_c^+ candidate variables in the $1 < p_T < 2$ GeV/c interval.	71
4.3	The plot shows the correlations between the selected features and the reconstructed invariant mass of the Λ_c^+ in the momentum range of 1 GeV/c to 2 GeV/c, using signal samples.	71
4.4	ROC curves for Λ_c^+ candidates in the p_T intervals $2 < p_T < 4$ GeV/c (left) and $8 < p_T < 12$ GeV/c (right).	72
4.5	Distributions for signal and background of Λ_c^+ candidate variables in the $1 < p_T < 2$ GeV/c interval.	74
4.6	Widths obtained in V0M multiplicity integrated bin for all p_T bins (left) and their ratio (right).	74
4.7	Invariant mass spectra for the Λ_c^+ baryon selected in the decay chain $\Lambda_c^+ \rightarrow pK_S^0$ for MB triggered data.	75
4.8	Invariant mass spectra for the Λ_c^+ baryon selected in the decay chain $\Lambda_c^+ \rightarrow pK_S^0$ for HMOV0M triggered data.	75
4.9	MC p_T shapes comparing Pythia8 with Monash tune, as used for the simulated samples, to the Mode2 tune.	77
4.10	Left top: $N_{\text{tracklets}}$ distribution in data (circles) and MC (triangles) for the three MB years and for the HM. Right: Multiplicity weights as a ratio of $N_{\text{tracklets}}$ distribution in data/MC.	79

4.11	Barrel multiplicity in selected events containing at least one $\Lambda_c^+ \rightarrow pK_S^0$ candidate within the mass window of $\pm 20 \text{ MeV}/c^2$ around the PDG value. All years of data taking from left to right, for MB triggered data in the top and for HVMZERO triggered data in the bottom row, respectively.	79
4.12	MC multiplicity weight for each year with the 4 multiplicity intervals	80
4.13	Prompt and feed-down acceptance-times-efficiency of V0M percentile multiplicity bins as a function of p_T for the Λ_c^+ baryon selected in the decay chain $\Lambda_c^+ \rightarrow pK_S^0$	80
4.14	Prompt and feed-down acceptance-times-efficiency of the integrated V0M percentile multiplicity bins as a function of p_T for the Λ_c^+ baryon selected in the decay chain $\Lambda_c^+ \rightarrow pK_S^0$	81
4.15	Left: Comparison of uncertainties using Nb and Nb $\times$ 2 convolution method (old FONLL predictions). Mid: $\Lambda_c^+ \rightarrow pK_S^0$ prompt fraction (multiplicity integrated) vs. p_T using the N_b method. Green = new FONLL predictions, blue = old FONLL predictions. Right: Ratio of new to old corrected yield.	82
4.16	The output of the MultiTrial fitter for the $\Lambda_c^+ \rightarrow pK_S^0$ decay in the [8-12] GeV/c p_T bin for the multiplicity integrated case. The figure displays the Gaussian mean (top left), Gaussian sigma (top middle), the reduced χ^2 of the fit (top right), and the extracted raw yield distributions using both the fit and bin counting methods (bottom row).	84
4.17	Ratios of corrected yield with different ML working points over the nominal values for the $\Lambda_c^+ \rightarrow pK_S^0$ V0M multiplicity bins [30 – 100]% (left), [0 – 0.1]% (right).	85
4.18	Total tracking systematic, containing the TPC track quality and ITS-TPC matching efficiency part, propagated to the kinematics of the selected hadrons for $\Lambda_c^+ \rightarrow pK_S^0$	85
4.19	Estimation of relative uncertainty from generated p_T shape for Λ_c^+ . The left plot displays the p_T shapes of Pythia8 Mode2 and from FONLL calculations. The middle plot shows the weights obtained by dividing the FONLL shape by the Pythia8 shape. The right plot shows the ratio of prompt efficiencies, one reweighted to Pythia8 Mode2 p_T shape (nominal shape used in this analysis) and the second reweighted to the FONLL shape. ...	86
4.20	Ratio of $f_{\text{non-prompt}}/f_{\text{non-prompt-MB}}$ versus relative $dN_{\text{ch}}/d\eta$ for charmed mesons (top left) and charmed baryons (top right) for different PYTHIA 8 tunes using softQCD.	88
4.21	Prompt Λ_c^+ yield per event as a function of candidate p_T and in intervals of V0M percentile p_{V0M} . Statistical and systematic uncertainties are indicated by vertical bars and boxes, respectively.	89

4.22	Transverse-momentum spectra of D^0 , D_s^+ , and Λ_c^+ hadrons measured in pp collisions at $\sqrt{s} = 13$ TeV for different multiplicity classes selected with the p_{V0M} estimator at forward rapidity. The corresponding ratios to $INEL > 0$ are shown in the bottom panels.	91
4.23	The D_s^+/D^0 (top) and Λ_c^+/D^0 (bottom) ratios measured in pp collisions at $\sqrt{s} = 13$ TeV for different multiplicity classes at mid- (left) and forward (right) rapidity.	91
4.24	The D_s^+/D^0 (top) and Λ_c^+/D^0 (bottom) ratios measured in pp collisions at $\sqrt{s} = 13$ TeV for the lowest (left) and highest (right) multiplicity classes at midrapidity.	93
4.25	The baryon-to-meson ratios in the light-flavour.	95
4.26	Ratios of p_T -integrated yields of Λ_c^+ and D^0 hadrons as a function of $\langle dN_{ch}/d\eta \rangle$ in pp collisions at $\sqrt{s} = 13$ TeV.	96
4.27	Distributions for prompt D^0 -meson, non-prompt D^0 -meson, and combinatorial-background candidate variables in the $1 < p_T < 2$ GeV/ c interval.	100
4.28	Linear correlations between the training variables, the invariant mass and the p_T , for prompt D^0 mesons (top-left panel) and non-prompt D^0 mesons (top-right panel) and combinatorial background (bottom panel) in the $1 < p_T < 2$ GeV/ c interval. The colours, from red (fully correlated) to blue (fully anti-correlated), indicate the correlation level.	100
4.29	Distributions of ML output scores for combinatorial background candidates, prompt D^0 mesons, and non-prompt D^0 mesons in the $2 < p_T < 4$ GeV/ c interval.	101
4.30	ROC curves for the models trained for D^0 mesons in the $2 < p_T < 4$ GeV/ c (left panel) and $8 < p_T < 12$ GeV/ c (right panel) intervals.	102
4.31	Invariant-mass distribution of D^0 candidates and their charge conjugates in selected p_T and multiplicity intervals.	103
4.32	Examples of raw-yield distribution as a function of the BDT-based selection employed in the χ^2 -minimisation procedure adopted for the determination of $f_{non-prompt}$ of D mesons. D^0 mesons in low multiplicity (left) and high multiplicity (right) classes.	105
4.33	Corrected $f_{non-prompt}$ as a function of p_T for each multiplicity interval for D^0 meson obtained varying the fitting configurations. The multiplicity classes are MB (top left), 0-0.1% (top right), 0.1-30% (bottom left) and 30-100% (bottom right). Each color corresponds to a different variation.	106
4.34	Sum in quadrature of RMS and shift with respect to the default value of the corrected f_{prompt} , $f_{non-prompt}$, and their ratio with respect to the MB obtained by varying the raw-yield extraction configurations for the D^0 meson.	107

4.35	Corrected $f_{\text{non-prompt}}$ as a function of p_T for each multiplicity interval for D^0 mesons obtained varying the sets of selections in the cut-variation method. The multiplicity classes are MB (top left), 0-0.1% (top right), 0.1-30% (bottom left) and 30-100% (bottom right). Each color corresponds to a different variation.	107
4.36	Sum in quadrature of RMS and shift with respect to the default value of the corrected f_{prompt} , $f_{\text{non-prompt}}$, and their ratio with respect to the MB obtained by varying the sets of selections in the cut-variation method for the D^0 meson.	108
4.37	Comparison of the $f_{\text{non-prompt}}$ (top panel) and $f_{\text{non-prompt}}^{30-100\%} / f_{\text{non-prompt}}^{\text{INEL}>0}$ (bottom panel) of D^0 mesons obtained with the different multiplicity weights in the 30-100% multiplicity class.	109
4.38	Comparison of $f_{\text{non-prompt}} / f_{\text{non-prompt}}^{\text{MB}}$ for D^0 mesons obtained with the different p_T weights in the 30-100% multiplicity class.	109
4.39	Fractions of non-prompt D^0 (left column) and D^+ (right column) mesons as a function of p_T for the $\text{INEL} > 0$ class and the three multiplicity classes of events in pp collisions at $\sqrt{s} = 13$ TeV.	111
4.40	Average fractions of non-prompt D^0 and D^+ mesons as a function of p_T for different multiplicity intervals normalised to the one measured in the $\text{INEL} > 0$ class of pp collisions at $\sqrt{s} = 13$ TeV.	113
4.41	Average fractions of non-prompt D^0 and D^+ mesons as a function of multiplicity, both normalised to the value corresponding to the $\text{INEL} > 0$ class, for pp collisions at $\sqrt{s} = 13$ TeV in different p_T intervals and integrated in $1 < p_T < 24$ GeV/c.	114
4.42	Fractions of non-prompt D^0 (first row) and D^+ (second row) mesons in $1 < p_T < 24$ GeV/c as a function of multiplicity for pp collisions at $\sqrt{s} = 13$ TeV compared with predictions obtained with the PYTHIA 8.	115
4.43	Fraction of charm and beauty quarks hadronising to baryons as a function of the charged particle multiplicity at midrapidity in PYTHIA 8.	115
4.44	Invariant mass distribution of D^0 candidates for the prompt enhanced set of selections in the MB sample.	118
4.45	Invariant mass distribution of D^0 candidates for the prompt enhanced set of selections in the MB sample.	118
4.46	Examples of raw-yield distribution as a function of the BDT-based selection employed in the χ^2 -minimisation procedure adopted for the determination of $f_{\text{non-prompt}}$ of D mesons. D^0 mesons be extracted in $2 < p_T < 4$ GeV/c MB triggered events.	119

4.47	p_T -differential production cross sections of non-prompt D^0 and Λ_c^+ hadrons in pp collisions at $\sqrt{s} = 13$ TeV compared with predictions obtained with FONLL calculations combined with PYTHIA 8 for the $H_b \rightarrow H_c + X$ decay kinematics.	122
4.48	Non-prompt Λ_c^+/D^0 , prompt Λ_c^+/D^0 , and Λ/K^0 production ratios measured in pp collisions at $\sqrt{s} = 13$ TeV compared with predictions.	125
5.1	Distributions of training variables, invariant mass, and p_T for prompt D^0 mesons, non-prompt D^0 meson, and combinatorial-background candidate variables in the $2 < p_T < 3$ GeV/ c interval.	128
5.2	ROC curves for the models trained for D^0 meson in the all p_T interval. Top three are(from left to right): $2 < p_T < 3$ GeV/ c , $3 < p_T < 4$ GeV/ c , $4 < p_T < 5$ GeV/ c . bottom three are(from left to right): $5 < p_T < 6$ GeV/ c , $6 < p_T < 8$ GeV/ c , $8 < p_T < 12$ GeV/ c	130
5.3	Full V0, V0A, V0C (left panels) and TPC $ \eta < 0.8$, TPC $0.0 < \eta < 0.8$, TPC $-0.8 < \eta < 0.0$ (left panels) event-plane distributions for the 2 nd harmonic in the 30–50% centrality classes.	131
5.4	Resolution terms for the SP methods as a function of centrality, in the 30–50% centrality class.	132
5.5	Simultaneous fits of the invariant-mass distribution and $v_2^{\text{tot}}(M_{D^0})$ of D^0 mesons in $3 < p_T < 4$ GeV/ c	133
5.6	Example of the raw-yield distribution as a function of the minimum non-prompt D^0 -meson BDT probability threshold to determine the non-prompt D^0 -meson fraction in $3 < p_T < 4$ GeV/ c	135
5.7	Right panel: the selection based on the correlated cut. Left panel: the selection based on the uncorrelated cut.	136
5.8	Non-prompt v_2 distribution with the dependence of the $f_{\text{non-prompt}}$ with the correlated cuts for the 6 p_T in the 30–50% centrality class.	137
5.9	Central value estimated of prompt and non-prompt v_2 based on the fit method in the 30–50% centrality classes(dummy error bar).	137
5.10	$f_{\text{non-prompt}}$ obtained with independent cut sets	138
5.11	v_2^{sig} as a function of $f_{\text{non-prompt}}$. The blue band represents the 1σ confidence interval obtained from the linear fit.	138

5.12	Prompt and non-prompt v_2 based on uncorrelated cuts with the fit method in the 30–50% centrality classes (left), right panel show the comparison result between uncorrelated cuts fit and correlated cuts fit.	139
5.13	Non-prompt $D^0 v_2$ as a function as multi-trials based on the different fit parameters	140
5.14	$f_{\text{non-prompt}}$ distribution as a function as p_T with the different configuration of the points (right for each panel) and the ratio between $f_{\text{non-prompt}}$ of trials and the default value(left for each panel).	141
5.15	The non-prompt $D^0 v_2$ distribution as a function as p_T with the different configuration (right) and the ratio between non-prompt $D^0 v_2$ of trials and the default value(left).	141
5.16	p_T distributions for the two models (right) and the weight for the prompt D^0	142
5.17	p_T distributions for the two models (right) and the weight for the prompt D^0	143
5.18	$f_{\text{non-prompt}}$ distribution as a function as p_T with the different p_T shape of the points.	143
5.19	The non-prompt $D^0 v_2$ distribution as a function as p_T with the different p_T shape (right) and the ratio between non-prompt $D^0 v_2$ of LIDO model and the default value(left).	144
5.20	Prompt D and B-meson v_2 as function as p_T predicted by TAMU (left) and LIDO (right) model.	144
5.21	The ratios between N_{IP} (N_{OOP}) and $N_{IP} + N_{OOP}$ with respect to charm and beauty hadron as function as p_T predicted by TAMU (top) and LIDO (bottom) model.	145
5.22	The p_T weights for in and out of plane for charm (left) and beauty (right) from TAMU (top) and LIDO (bottom) model.	146
5.23	The possible effect from the different p_T shapes of in (top) and out (bottom) plane is estimated from TAMU model.	147
5.24	The possible effect from the different p_T shapes of in (top) and out (bottom) plane is estimated from LIDO model.	147
5.25	Different sub-samples used to the delete-d jackknife test.	149
5.26	Full procedure of the delete-d jackknife test in non-prompt $D^0 v_2$ analysis.	150
5.27	Non-prompt $D^0 v_2$ distribution using $\binom{20}{17}$ jackknife samples based on the correlated cut fit.	150
5.28	Non-prompt $D^0 v_2$ distribution using $\binom{20}{17}$ jackknife samples based on the uncorrelated cut fit.	151
5.29	Non-prompt $D^0 v_2$ distribution as function as trials using $\binom{20}{17}$ jackknife samples based on the correlated cut fits.	151
5.30	Non-prompt $D^0 v_2$ distribution as function as trials using $\binom{20}{17}$ jackknife samples based on the uncorrelated cut fits.	152

5.31	Non-prompt D^0 v_2 distribution as function as trials using $\binom{20}{17}$ jackknife samples based on the uncorrelated cut fits.	152
5.32	Left: non-prompt D^0 -meson (blue points) and prompt average non-strange D-meson (red points) v_2 as a function of p_T in 30–50% Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. Statistical uncertainties are shown as vertical lines and systematic uncertainties as boxes. Right: non-prompt D^0 -meson v_2 compared with model calculations.	153
5.33	Non-prompt D^0 -meson (black points) and electrons from beauty hadron decay (cyan points) v_2 as a function of p_T in 30–50% Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV.	154
5.34	Non-prompt D^0 -meson v_2 and R_{AA} as a function of p_T in 30–50% Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV.	155

List of Tables

2.1	parameters of ITS layers.....	39
2.2	Summary of the characteristics of the used data samples for the analyses presented in this thesis. The numbers for the Pb – Pb data samples are after selecting a centrality interval of 30-50%.	49
2.3	Data taking periods and number of events for minimum bias (MB), high-multiplicity V0 (HMV0), and high-multiplicity SPD (HMSPD) triggers in different years.	49
3.1	The mean proper decay length of heavy hadrons.	50
4.1	The summary of the multiplicity event classes at midrapidity (Ntrkl) and forward rapidity (%V0M), which corresponds to the visible V0M cross section, is shown below. The average charged-particle densities $\langle dN_{ch}/d\eta \rangle$ at midrapidity are presented, along with the value corresponding to the INEL > 0 event class. Additionally, the trigger efficiency ϵ^{INEL} is reported for each multiplicity interval.	69
4.2	Pre-selection for Λ_c^+ candidates	72
4.3	p_T intervals for Λ_c^+	72
4.4	Training variables used for the Λ_c^+ baryons.	73
4.5	The (p_T -independent) XGBoost BDT hyperparameters used for candidate selection in $\Lambda_c^+ \rightarrow pK_S^0$	73
4.6	Variations used for $\Lambda_c^+ \rightarrow pK_S^0$ MultiTrial.	83
4.7	Systematic uncertainties for raw yield extraction of $\Lambda_c^+ \rightarrow pK_S^0$ were evaluated in both multiplicity integrated and multiplicity range cases.	83
4.8	Variations on $f_{non-prompt}$ used for the different multiplicity bins.	88
4.9	Systematic uncertainties due to the multiplicity dependence in the feed-down subtraction for $\Lambda_c^+ \rightarrow pK_S^0$ in the multiplicity ranges on spectra(for mult/MB ratio) in percentage. ...	88
4.10	Training variables used for the D^0 mesons.	99
4.11	Selection criteria applied for the D^0 meson in the p_T bins of the analysis.	117
4.12	Production cross sections at midrapidity per unit of rapidity $(d\sigma/dy) _{ y < 0.5}$ of non-prompt Λ_c^+ and D^0 hadrons and $b\bar{b}$ pairs $(d\sigma b\bar{b}/dy) _{ y < 0.5}$ compared to pQCD calculations.	121
4.13	p_T -integrated Λ_c^+/D^0 production ratio measured at midrapidity ($ y < 0.5$) in pp collisions at $\sqrt{s} = 13$ TeV and in e^+e^- collisions at LEP for prompt and non-prompt production.	124

5.1	BDT score cuts for the background and FD in the p_T intervals	139
5.2	Configuration of the fit procedure.....	140

1 General introduction

The Standard Model is a fundamental theory in particle physics, based on a local gauge symmetry of $SU(3) \times SU(2) \times U(1)$ and a relativistic quantum field theory. It provides a classification for all known elementary particles and explains their interactions through the strong, weak, and electromagnetic forces. However, it does not include the fourth known interaction, gravitation. Each force acts on specific charges and is mediated by gauge bosons, such as photons (γ) for the electromagnetic force, $W^\pm$ and Z^0 bosons for the weak force, and gluons (g) for the strong force, which acts on particles with color charge. The Standard Model consists of twelve elementary particles and their corresponding antiparticles, known as fermions, which are grouped into quarks and leptons as shown in Fig. 1.1. Quarks experience all three types of interaction, while leptons experience only the electromagnetic and/or weak interaction. A more detailed explanation of the Standard Model can be found in Ref. [1]. The strong interaction between quarks and gluons, the elementary constituents of the hadronic matter, is described by *Quantum Chromodynamics* (QCD). This theory predicts the existence of a phase transition for the strongly interacting matter from the hadronic phase to a deconfined one, called *quark-gluon plasma* (QGP), where quarks and gluons are not confined within hadrons. In this Chapter, a brief introduction to the QCD and QGP, and a selection of the main experimental observables in this field and their interpretation are presented.

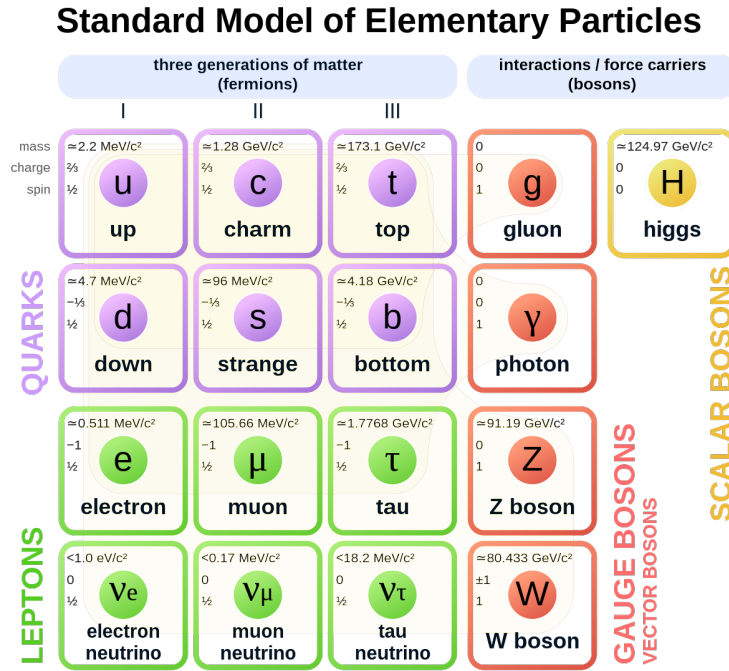


Figure 1.1: Standard Model in particle physics



1.1 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is a non-Abelian quantum gauge field theory. Its *Lagrangian* is invariant under local transformations of the $SU(3)$ group, as described in references [2, 3]. This symmetry of the Lagrangian introduces a conserved charge called "colour," which can take on three distinct values: red, blue, and green. Quarks in this theory are characterized by their colour charge, and their interactions are mediated by massless vector gauge bosons known as gluons. These gluons carry both colour and anticolour charges. The QCD Lagrangian density can be expressed as:

$$\mathcal{L}_{QCD} = \sum_f \bar{q}_f^k \left(i\gamma_\mu D_{kl}^\mu - m_f \delta_{kl} \right) q_f^l - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu}, \quad (1.1)$$

The sum in the equation runs over the six quark flavours (*up, down, strange, charm, beauty, and top*). Here, q_f^l represents the quark field with flavour f , colour charge l , and mass m_f . The covariant derivative D_{kl}^μ introduces the interaction between quarks and gluons, and it is defined by the equation:

$$D_{kl}^\mu = \partial^\mu \delta_{kl} - ig_s \frac{\lambda_{kl}^a}{2} A_a^\mu, \quad (1.2)$$

In this equation, A_a^μ represents the gluon fields, which correspond to the eight possible colour configurations. The Gell-Mann matrices are represented by λ_{kl}^a . The constant g_s quantifies the intensity of the interaction and is related to the strong coupling constant α_s by $g_s = \sqrt{4\pi\alpha_s}$. The second term of the equation describes the kinematics and dynamics of the gluons, and $G_{\mu\nu}^a$ can be expanded as follows:

$$G_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g_s f_{abc} A_\mu^b A_\nu^c \quad (1.3)$$

The structure constants of the $SU(3)$ -group Lie algebra are represented by f_{abc} . Since the $SU(3)$ group is non-Abelian, the structure constants are not null, and vertices among three and four gluons are possible, in addition to vertices between a quark, an antiquark, and a gluon. The presence of gluon self-interactions implies that the colour force is a short-range interaction, even though gluons are massless particles.

The unique characteristics of the strong interaction can be understood by analyzing the dependence of the strong coupling constant α_s on the energy scale of the interaction. The evolution of the α_s value as a function of the squared transferred momentum Q^2 is given by:

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu^2)}{1 + \frac{\alpha_s(\mu^2)}{12\pi} (33 - 2n_f) \ln(Q^2/\mu^2)} \quad (1.4)$$

where μ is the renormalisation scale and n_f is the number of quark flavours with masses smaller than μ . This dependence is a consequence of the renormalisation process used to eliminate divergences in perturbative QCD calculations, which includes quantum corrections like vertex corrections, self-energy, and vacuum polarisation in the value of the coupling constant.

Experimental measurements confirm the evolution of α_s described by Eq.1.4, as illustrated in Fig. 1.2. The value of α_s decreases as the momentum transfer increases, and two distinct regimes can be distinguished: asymptotic freedom and confinement. At high values of $|Q|$, α_s is small, and asymptotic freedom is attained, meaning that quarks and gluons inside hadrons are weakly interacting and can be considered free. Perturbative QCD calculations based on a perturbative expansion in α_s are applicable in this regime. For $|Q| < 1 \text{ GeV}/c$, the strong coupling constant diverges, and quarks and gluons are strongly confined inside hadrons, leading to the confinement regime. In this regime, the large value of α_s prevents the application of perturbative techniques, and QCD calculations can be performed using lattice QCD[4], a discretised space-time grid approach.

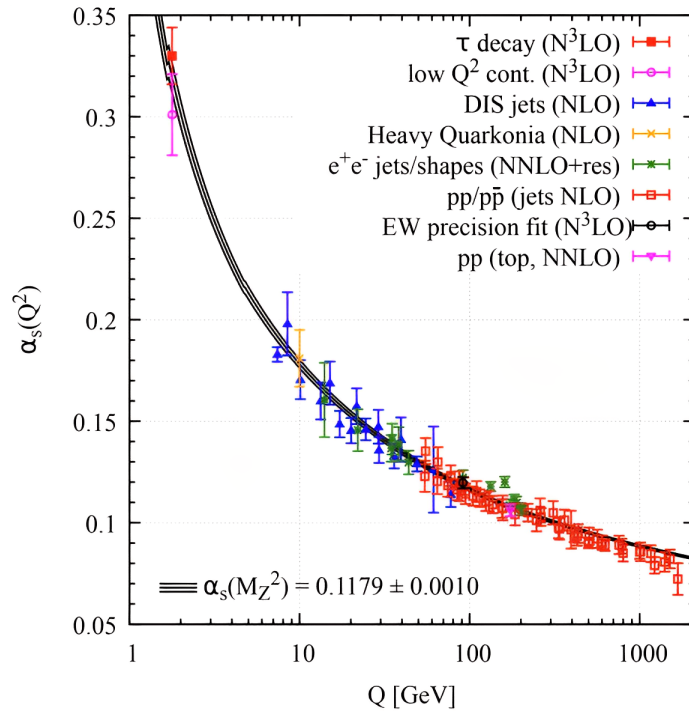


Figure 1.2: Summary of α_s measurements as a function of the transferred momentum Q . The respective degree of QCD perturbation theory used in the extraction of α_s is indicated in parentheses. Figure from Ref. [5]



1.2 Quark Gluon Plasma

The strong coupling constant exhibits a significant range of values, resulting in diverse properties of nuclear matter, especially at high temperatures and densities. Such extreme conditions are encountered in nature during the early moments of the universe after the Big Bang or within the core of neutron stars. Understanding the behavior of nuclear matter under these conditions is crucial, and investigating its phase diagram can provide valuable insights. High-energy collisions between heavy nuclei in the laboratory can replicate these conditions, and lattice QCD calculations can help constrain phenomenological models, enabling both theoretical and experimental approaches to the problem.

The phase diagram of QCD matter is shown in Fig. 1.3 as a function of T and μ_B . Ordinary nuclear matter is located in the region where $T \approx 0$ and $\mu_B \approx 1$ GeV. At higher values of the baryon-chemical potential, nuclear matter can be described as a degenerate gas of neutrons, similar in structure to the atomic nucleus. This matter is expected to make up the core of neutron stars. As the baryon-chemical potential increases, a low-temperature gas is predicted where quarks are no longer confined inside hadrons. This state is called a colour-superconductive state[6]. At high temperatures, nuclear matter transitions to a hadron gas, and above $T \approx 100 - 150$ GeV, it is expected to convert to the *quark-gluon plasma*. The QGP is a strongly-interacting state where the colour degrees of freedom are not confined in colour-singlet hadrons, allowing quarks and gluons to move more freely in a larger volume of space. The nature of the transition from hadronic matter to QGP can be studied using lattice QCD calculations. In the low- μ_B limit, a crossover between the two phases is expected due to the chiral-symmetry restoration at high energies [7]. At higher μ_B , a first-order phase transition is predicted, implying the existence of a critical point due to coexistence with the crossover. For $\mu_B \approx 0$, the QGP is predicted to form at the pseudo-critical temperature $T_{pc} \approx 154 - 158$ MeV, corresponding to an energy density ϵ_{pc} of about 0.42 GeV/fm^3 [7, 8]. These conditions are reached in the initial stages of ultrarelativistic heavy-ion collisions, and current measurements indicate that the QGP is a strongly-coupled liquid-like system with a remarkably low shear viscosity to entropy density ratio of $\eta/s \approx 1/4\pi$ [9].

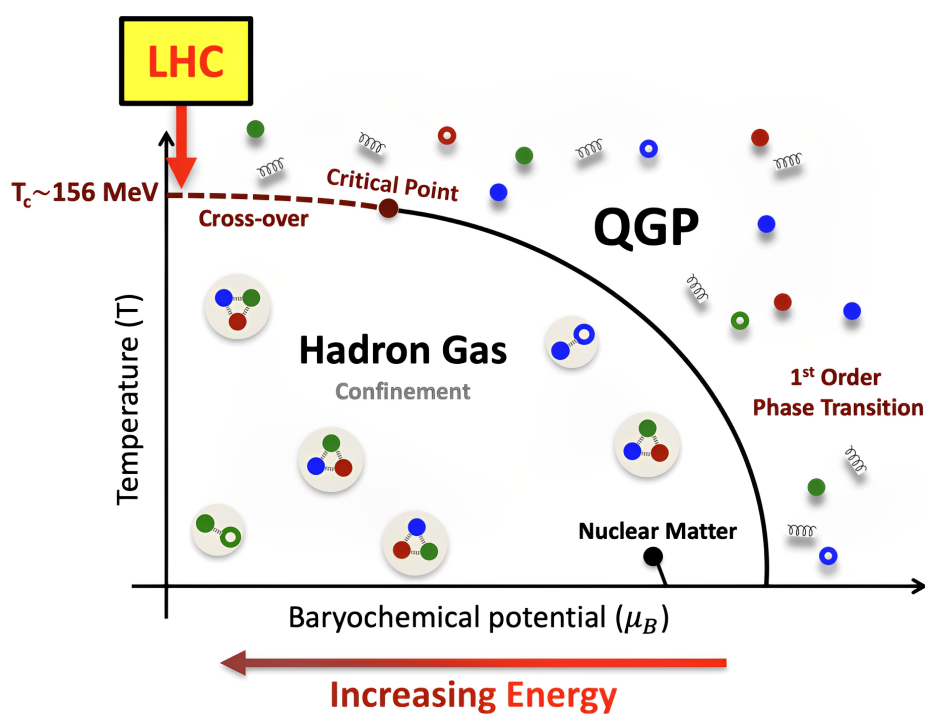


Figure 1.3: The QCD phase diagram as a function of the temperature and the baryon-chemical potential.

1.3 Heavy-ion collision

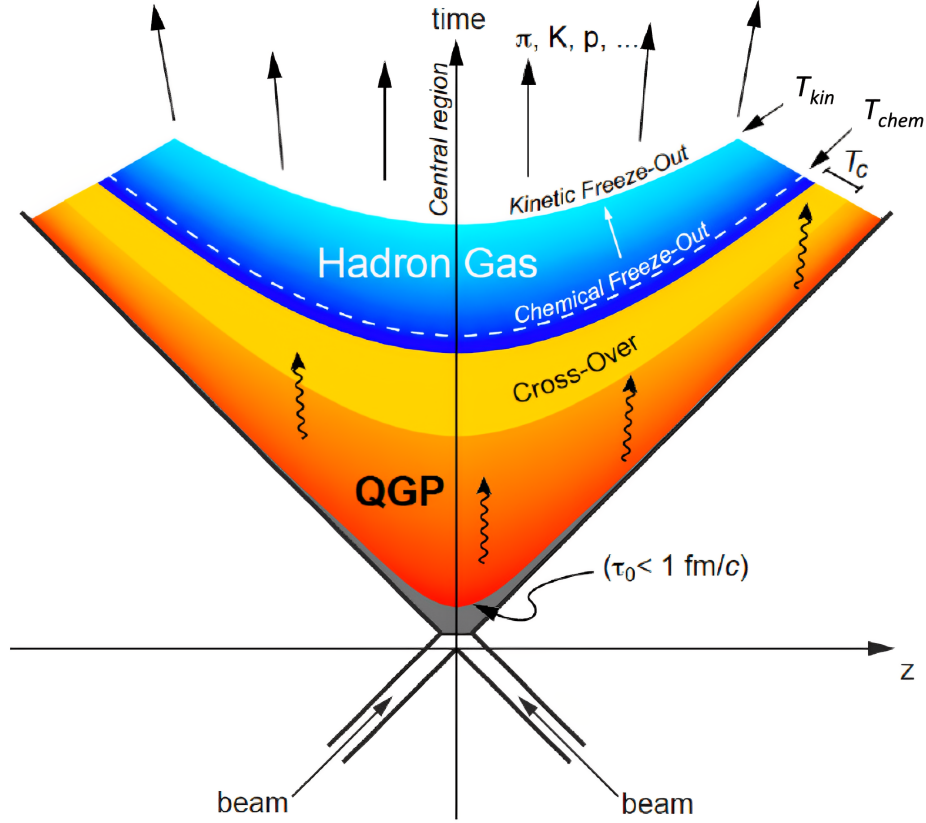


Figure 1.4: The space-time diagram of a heavy-ion collision between two nuclei at $t = 0$ and $z = 0$ in the longitudinal direction (transverse direction not shown). The evolution of the system starts with a pre-equilibrium phase, which leads to the formation of a Quark-Gluon Plasma, followed by a cross-over phase transition to a hadron gas. During this process, the hot-fireball formed in the collision emits various types of particles, indicated by the arrows. Figures from [10]

As discussed in the previous section, QGP is a state of matter that can be observed in nature during the early moments of the universe after the Big Bang or within the core of neutron stars. In laboratory settings, QGP is recreated via ultra-relativistic heavy ion collisions. The first experiments to explore heavy-ion collisions began in the second half of the 1980s at the Super Proton Synchrotron (SPS) at CERN and the Alternating Gradient Synchrotron (AGS) at Brookhaven National Laboratory (BNL). At these facilities, several fixed-target experiments operated at a centre-of-mass energy per nucleon-pair ($\sqrt{s_{NN}}$) in the range of 7.7-17.3 GeV, probing the baryon chemical potential (μ_B) interval of 200-500 MeV. At these energies, nucleons are stopped in the collision region, resulting in high net baryon density. This condition is known as the stopping regime. Subsequently, heavy-ion physics moved to colliders, first with the Relativistic Heavy Ion Collider (RHIC) at BNL in 2000 and then with the Large Hadron Collider (LHC) at CERN in 2010. The top centre-of-mass energies of the two accelerators are 200 GeV and 5.5 TeV for Au and Pb ions, respectively, allowing for exploration of the QCD phase diagram in the



region of μ_B close to zero. In this case, the collision between two heavy nuclei occurs under the *Bjorken regime* or *transparency regime* [11]. A heavy-ion collision is a complex space-time event. Assuming that the Bjorken condition is satisfied, the stages of this evolution can be described in terms of the light-cone variables t (time) and z (direction of the colliding beams), as shown in Fig. 1.4. The following phases of the system evolution can be defined:

- **Initial state:** Before colliding, the incident nuclei are Lorentz-contracted disks in the centre-of-mass frame, resulting from their ultra-relativistic velocities. These disks are primarily composed of gluons, which carry only a small fraction (x) of the longitudinal momenta of their parent nucleons, but have relatively large transverse momenta (e.g., $k_\perp \simeq 2 \text{ GeV}/c$ for $x = 10^{-4}$) and densities that rapidly increase with $1/x$. These high-density gluon systems contain a significant fraction of the colliding partons and therefore govern the dynamics of the early stages in heavy-ion collisions. To properly characterise the produced QCD medium, a detailed understanding of these small- x gluons in the nuclei wave function is essential. Bulk observables, used to extract the medium transport coefficients, are as sensitive to features of the initial stage as to QGP properties themselves [12, 13].
- **Pre-equilibrium:** Particles are produced from the energy released in the interaction region by the colliding nuclei after a formation time $\tau_0 > \tau_{\text{cross}}$. In the midrapidity region, the net baryon density is nearly zero, and the energy density of the system can be approximated as follows:

$$\epsilon = \frac{\langle E_T \rangle}{\pi R^2 \tau_0} \frac{dN}{dy} \bigg|_{y=0}, \quad (1.5)$$

In this formula, πR^2 represents the transverse area of the collision region, $\langle E_T \rangle$ is the average transverse energy of the measured particles, and dN/dy is the particle multiplicity per rapidity unit [11, 14]. Based on the measured particle multiplicities at the LHC and assuming a formation time of $\tau_0 \sim 1 \text{ fm}/c$, the energy densities achieved at this collider exceed $12 \text{ GeV}/\text{fm}^3$ [15]. This is significantly above the pseudo-critical value for QGP formation.

- **QGP:** After the pre-equilibrium stage, the QGP is formed and it evolves following the laws of relativistic viscous hydrodynamics [16]. The QGP droplet expands driven by significant pressure gradients, resulting in cooling down. The lifetime of QGP for the energy densities attained at the LHC is around $10 \text{ fm}/c$. Experimental quantification of transport coefficients has become a crucial aspect of heavy-ion physics [17].
- **Hadronization:** When the local temperature of the QGP droplet falls below the critical temperature T_c , a crossover transition occurs, and hadronisation takes place. This process, known as the chemical freeze out, confines the previously free quarks and gluons inside colour-singlet hadrons. The formed hadronic system is known as a *hot hadron gas*. This system is still relatively dense and



preserves local thermal equilibrium while it continues to expand. Eventually, the system dilutes, and the hadrons stop interacting with each other. This is known as the kinetic or thermal *freeze out*. After this stage, the particles freely stream until they reach the detector.

1.4 Experimental observables

In the investigation of heavy-ion collisions, a range of experimental probes are employed to explore the various phases of the collision, including the initial state, the QGP phase, and the final hadronic phase. Each of these probes has distinct sensitivities to the different phases involved [18]. This thesis will primarily focus on the experimental probes of the QGP, which can be broadly divided into two categories: measurements of particle yields and measurements of particle correlations.

Particle yields are often measured with respect to kinematic variables such as the transverse momentum (p_T) or pseudorapidity (η) of identified particles, or event characteristics such as the charged hadron multiplicity (N_{ch}). The bare yields provide information on the cross sections of the underlying production processes [19], while yield ratios can be more informative about specific physical mechanisms. For instance, ratios of yields between nucleus-nucleus and proton-proton collisions can reveal in-medium energy loss mechanisms [19], while ratios of yields between different hadron species can be sensitive to hadronisation processes [20].

Alternatively, particle correlations can also be studied, usually through the calculation of correlations between the kinematic variables of specific identified particles or all produced particles in an event. For instance, correlations in the azimuthal angle can provide information on anisotropic flow [21], while momentum correlations can be sensitive to the spatial evolution of the system or in-medium energy loss mechanisms [22]. In this section, we will discuss the experimental observables in greater detail.

1.4.1 Particle multiplicity

The total number of particles produced in a collision is a fundamental quantity that provides crucial information about the properties of the created medium, such as the collision geometry and energy density. While the Glauber models can give insight into the collision geometry, the Bjorken hydrodynamic picture can estimate the energy density for nucleus-nucleus collisions [11]. The Bjorken energy density (ϵ_{BJ}) can be estimated using the equation:

$$\epsilon_{BJ} = \frac{\langle m_T \rangle}{\tau_0 \cdot A} \frac{dN(\tau_0)}{dy} \Big|_{y=0}, \quad (1.6)$$



where A is the transverse overlap area of the collision region, τ_0 is the thermalisation time for the QGP (usually estimated as $1\text{fm}/c$), m_T is the transverse mass ($m_T = (m^2 + p_T^2)^{1/2}$), and dN/dy is the particle multiplicity per unit of rapidity. While this is a simplification given the complex sequence of stages preceding particle production, it provides a good estimate of the energy density of the medium using the measured charged-particle pseudorapidity density at midrapidity.

The left panel of Fig. 1.5 [23] presents a compilation of charged-particle multiplicity measurements in various collision systems. These measurements provide important information on the produced medium, such as its energy density, which can be estimated using Eq. 1.1. At the LHC, the estimated energy density is found to be approximately $\sim 14\text{ GeV}/\text{fm}^3$ [24], while at RHIC it is about three times smaller [25]. The averaged charged-particle pseudorapidity density, $\langle dN_{\text{ch}}/d\eta \rangle$, shows a smooth trend as a function of $\sqrt{s_{\text{NN}}}$ for different collision systems. For nucleus-nucleus collisions, the rise of $dN_{\text{ch}}/d\eta$ is observed to be steeper than for proton-proton and proton-nucleus collisions. Interestingly, the measurements in $p - A$ and $d - A$ systems can be described by the same power-law function as for pp systems, implying that the observed difference for nucleus-nucleus collisions is not solely related to the multiple interactions between the participating nucleons, since the proton in $p - A$ collisions also experiences multiple nucleons.

The centrality dependence of $\langle dN_{\text{ch}}/d\eta \rangle$ for $\text{Pb} - \text{Pb}$ collisions at $\sqrt{s_{\text{NN}}} = 5.02\text{ TeV}$ is shown in the right panel of Fig. 1.5, with N_{part} serving as the centrality measure as discussed earlier. The charged-particle pseudorapidity density exhibits a steep increase for peripheral collisions, but the slope decreases for $N_{\text{part}} > 50$. Model predictions [23] attempting to describe this observation can be broadly categorized into two groups: those that combine pQCD processes with soft interactions and those that implement different mechanisms to control the gluon density in the initial stage. The latter models do a better job of reproducing the measurement, suggesting the presence of a gluon saturation domain that limits the number of produced partons and hence, particles [12, 13]. These measurements have another implication, which is not yet fully understood - the energy and centrality dependence appear to be surprisingly independent of each other [23–25]. This may indicate that the increase in hard relative to soft processes, which would naively lead to an increase in particle multiplicity with energy, does not play a dominant role in determining the particle multiplicity.

1.4.2 Identified-hadron abundances

The temperature T_{ch} and volume V of a system at the moment of chemical freeze-out can be estimated by measuring the production yields of different hadron species and comparing them with predictions from the statistical hadronisation model (SHM) [26]. The SHM assumes that a system composed

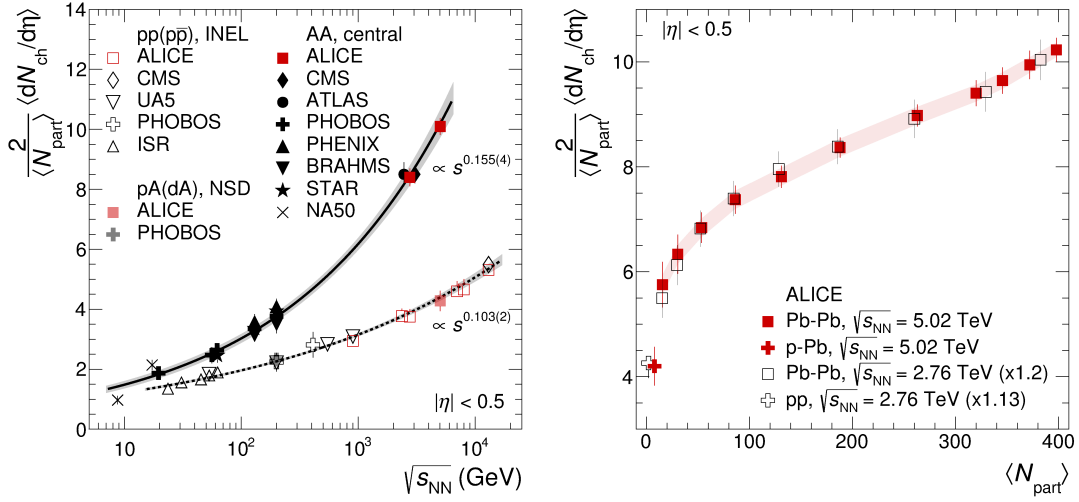


Figure 1.5: Left: The charged-particle pseudorapidity density ($\langle dN_{\text{ch}}/d\eta \rangle$) is plotted as a function of the centre-of-mass energy ($\sqrt{s_{\text{NN}}}$) for various types of collisions. Right: The centrality dependence of $\langle dN_{\text{ch}}/d\eta \rangle$ for Pb – Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02 \text{ TeV}$ is compared to different model predictions. The division of $2/\langle N_{\text{part}} \rangle$ in both figures is introduced for comparing different colliding systems. Figures taken from Ref. [23].

of strongly interacting hadrons is in thermal and chemical equilibrium at chemical freeze-out, forming an ensemble that can be treated using the grand-canonical ensemble formalism for A – A collisions at collider energies. Under these conditions, the thermodynamic properties of the system determine the hadron production. Specifically, the abundance N_i of a particle species i is given by:

$$N_i(T_{ch}, V, \mu_i) = \frac{g_i V}{2\pi^2} \int_0^{+\infty} \frac{p^2 dp}{\exp[(E - \mu_i)/T_{ch}] \pm 1} \quad (1.7)$$

The system volume is denoted by V , and p and E represent the particle momentum and energy, respectively. The particle spin degeneracy is quantified by g_i , and the chemical potential for the relevant species, μ_i , ensures the conservation of relevant quantum numbers on average. The distinction between bosons and fermions is indicated by the $\pm$ sign. In some cases, an additional parameter γ_s is introduced to account for non-perfect chemical equilibrium of strange quarks. This parameter scales thermal predictions and is less than one in the case of non-equilibrium, such as in peripheral A-A collisions or smaller p – A and pp collision systems. The SHARE model introduces another parameter, γ_q , which considers a non-equilibrium QGP undergoing sudden hadronization without further hadronic re-scatterings [27].

The ALICE Collaboration's measurements of total yields per rapidity unit for different particle species in the 10% most central Pb – Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02 \text{ TeV}$ are presented in Fig. 1.6. These measurements are compared to the values obtained by three different models: THERMUS 4 [28], GSI-

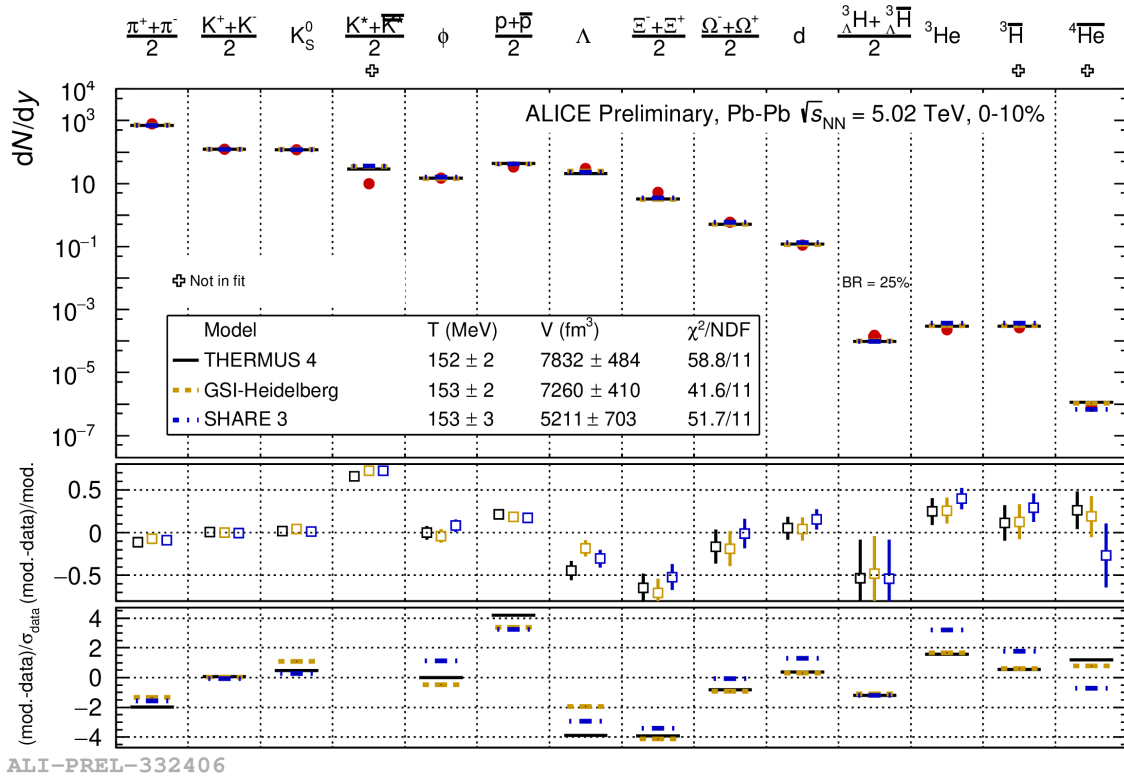


Figure 1.6: Thermal fit to the identified-particle yields measured by the ALICE Collaboration in the 10% most central Pb – Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. The results obtained with the THERMUS 4 [28], GSI-Heidelberg [29], and SHARE 3 [27] models are reported. Figure from ALICE figure repository



Heidelberg [29], and SHARE 3 [27]. These models simultaneously fit the particle yields to determine the T_{ch} and V parameters based on different assumptions for the system's equilibrium condition and conservation laws at chemical freeze-out. The models successfully describe the observed particle abundances over nine orders of magnitude, with a χ^2/ndf ranging from approximately 3.8 to 5.3. However, a significant discrepancy with the measurements is observed for the proton yield, while excited kaons are excluded from the fit due to their strong susceptibility to rescatterings and regeneration in the hadronic phase. The estimated chemical freeze-out temperature is about 153 MeV, and it is consistent among the different models. This value is close to the expected pseudo-critical temperature for the crossover transition from the hadronic matter to the QGP.

1.4.3 Nuclear modification factor

To investigate the hottest and densest phases of high-energy heavy-ion collisions, one can examine particles emitted with high transverse momentum. These high- p_T partons, such as quarks or gluons, are sensitive to various effects that can modify their distribution compared to pp collisions. Specifically, jet quenching, caused by energy loss from induced gluon radiation or elastic collisions, can be measured to study the thermodynamic or transport properties of the QGP.

The nuclear modification factor R_{AB} is a way to measure the nuclear modification of a given observable obs_{AB} , where A and B represent the respective colliding objects with an impact parameter b . It is measured as a function of various parameters such as collision energy $\sqrt{s_{NN}}$, transverse momentum p_T , or rapidity y , by comparing it to the same observable measured in pp collisions obs_{pp} . The expression for R_{AB} is given by:

$$R_{AB}(\sqrt{s_{NN}}, p_T, y; b) = \frac{\text{"hot/cold QCD medium"}}{\text{"QCD vacuum"}} \propto \frac{obs_{AB}(\sqrt{s_{NN}}, p_T, y; b)}{obs_{pp}(\sqrt{s_{NN}}, p_T, y)} \quad (1.8)$$

which can be more specifically defined, in p – Pb or A – A collisions, here as a function of p_T , as

$$R_{AA} = \frac{dN_{AA}/dp_T}{\langle N_{coll} \rangle dN_{pp}/dp_T} = \frac{dN_{AA}/dp_T}{\langle T_{AA} \rangle d\sigma_{pp}^{INEL}/dp_T}, \quad (1.9)$$

$$R_{pPb} = \frac{dN_{pPb}/dp_T}{\langle N_{coll} \rangle dN_{pp}/dp_T} = \frac{dN_{pPb}/dp_T}{\langle T_{pPb} \rangle d\sigma_{pp}^{INEL}/dp_T} \quad (1.10)$$

where $N_{AA/pPb/pp}$ is the observable's yield in A, p – Pb or pp collisions, respectively, σ_{pp}^{INEL} is the observable's inelastic cross section in pp collisions, $\langle N_{coll} \rangle$ is the average number of binary nucleon-nucleon



collisions for the considered centrality interval, and $\langle T_{AA/pPb} \rangle = N_{\text{coll}}/\sigma_{pp}$ is the nuclear overlap function. This function represents the effective integrated luminosity per collision. A nuclear modification factor less than unity indicates suppression, and a nuclear modification factor greater than unity indicates enhancement, relative to the same number of independent pp collisions.

Fig. 1.7 illustrates the nuclear modification factor of charged particles as a function of p_T , as measured by the ALICE Collaboration [30]. The measurements were taken in p – Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV and in central (0-5%) and peripheral (70-80%) Pb – Pb collisions. The nuclear modification factor for high- p_T particles ($(p_T \gtrsim 6 \text{ GeV}/c)$) in p – Pb collisions is close to unity, indicating no significant effects in the R_{AA} associated with the initial state. The maximum at intermediate p_T ($2 < p_T < 6 \text{ GeV}/c$) is attributed to the Cronin effect [31]. In central Pb – Pb collisions, the nuclear modification factor is significantly suppressed, while for peripheral collisions, the suppression is less pronounced and approaches unity for the highest p_T intervals. The measurement of the R_{AA} serves as a sensitive probe of medium-induced effects, and theoretical calculations can be compared with experimental data to extract the transport properties of the medium [32].

1.4.4 Anisotropic flow

The geometry of a nucleus-nucleus collision can be described by the anisotropy of the overlap region in the transverse plane, which increases with the impact parameter. The spatial anisotropy of the initial state generates pressure gradients in the Quark-Gluon Plasma (QGP), leading to collective flow. This flow results in an azimuthal anisotropy of the momentum distribution of the produced hadrons. To study this anisotropy, particle production is measured as a function of the azimuthal angle φ , which can be decomposed into a Fourier expansion [33] as

$$E \frac{d^3N}{dp^3} = \frac{1}{2\pi} \frac{d^2N}{p_T dp_T dy} \left\{ 1 + 2 \sum_{n=1}^{\infty} v_n(p_T) \cos[n(\varphi - \Psi_n)] \right\}. \quad (1.11)$$

The equation above describes the Fourier decomposition of the azimuthal angle dependence of the particle production in heavy-ion collisions. The harmonic order n represents the degree of anisotropy, while $v_n(p_T)$ is the flow coefficient or the p_T -dependent Fourier coefficient. The angle Ψ_n denotes the symmetry-plane angle. The flow coefficient $v_n(p_T)$ is defined as

$$v_n(p_T) = \langle \cos[n(\varphi - \Psi_n)] \rangle, \quad (1.12)$$

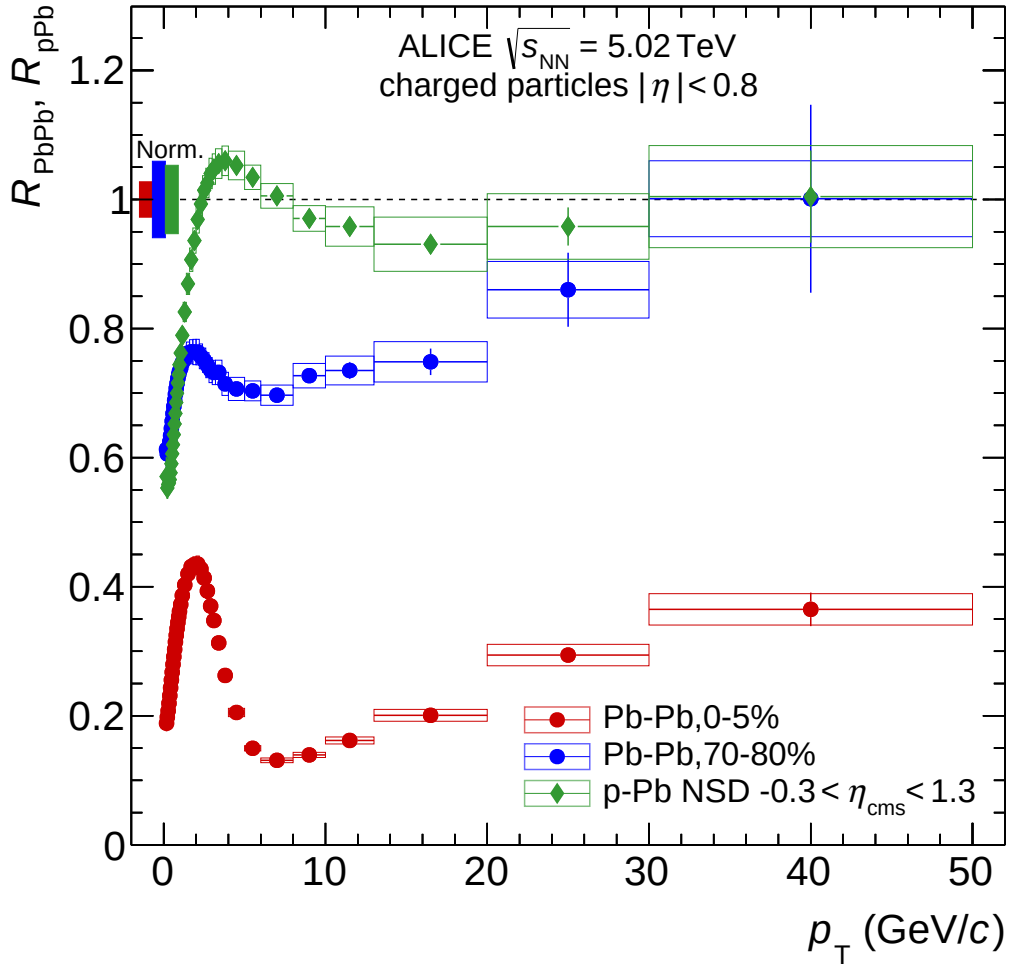


Figure 1.7: ALICE measured the nuclear modification factors for charged particles in central (0-5%) and peripheral (70-80%) Pb-Pb collisions, as well as in minimum-bias p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. Figure from [30]



In a given p_T interval and within a specific centrality interval, the average over all particles is denoted by brackets. The angle Ψ_n of the symmetry plane is determined by the spatial distribution of participating nucleons during the collision. For $n = 2$, and without considering fluctuations in the participant distribution, Ψ_n identifies the reaction plane, which is defined by the impact parameter vector and the direction of the colliding nuclei. The dominant term in the Fourier expansion during non-central collisions is the second-harmonic coefficient, v_2 , also known as *elliptic flow*, owing to the almond-shaped interaction region. The measurement of a significant elliptic flow directly indicates collective motion. Higher-order coefficients such as v_3 and v_4 primarily result from fluctuations in the initial participant distribution in the overlap region and the non-linear hydrodynamic response of the medium [34, 35]. Comparing the measured v_n coefficients with viscous hydrodynamic models of the QGP provides information about the system's equation of state and transport coefficients, such as the shear and bulk viscosity to entropy density ratios [36, 37].

Fig. 1.8 presents the p_T -differential elliptic-flow coefficient v_2 for several particle species, including charged pions, charged and neutral kaons, ϕ mesons, protons, and Λ hyperons, measured by the ALICE Collaboration in Pb – Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, across different centrality intervals [38]. The coefficient shows an increasing trend from the most central (0-1%) to semicentral (30-40%) collisions, attributable to the progressively higher spatial anisotropy of the overlap region between the colliding nuclei. Conversely, in more peripheral collisions, the coefficient decreases as the pressure gradients originating from the lower density of the produced medium become weaker. The coefficient exhibits a mass ordering in the low- p_T region, $p_T < 2 - 3$ GeV/ c , with lighter particles displaying a larger v_2 compared to heavier ones, due to the radial flow effect that imparts an equal velocity boost to all particles, pushing more massive hadrons to higher p_T . In the p_T region $3 < p_T < 8$ GeV/ c , the v_2 coefficients display a grouping by baryons and mesons, with the baryon values being larger than those of mesons. Interestingly, the ϕ meson, whose mass is similar to that of the proton, exhibits a v_2 value comparable to that of the other mesons in this p_T range, indicating a dependence on the number of constituent quarks rather than the hadron mass. This observation lends support to the notion of quark hadronization via coalescence, where the elliptic flow of each constituent quark contributes to the final-state hadron v_2 .

1.5 Open heavy-flavour physics

This thesis focuses on open heavy-flavour hadrons, namely mesons and baryons that carry one charm or bottom quark or antiquark. More than 95% of the produced heavy quarks in heavy-ion collisions at RHIC and the LHC are distributed into these hadrons [39]. These hadrons provide are ex-

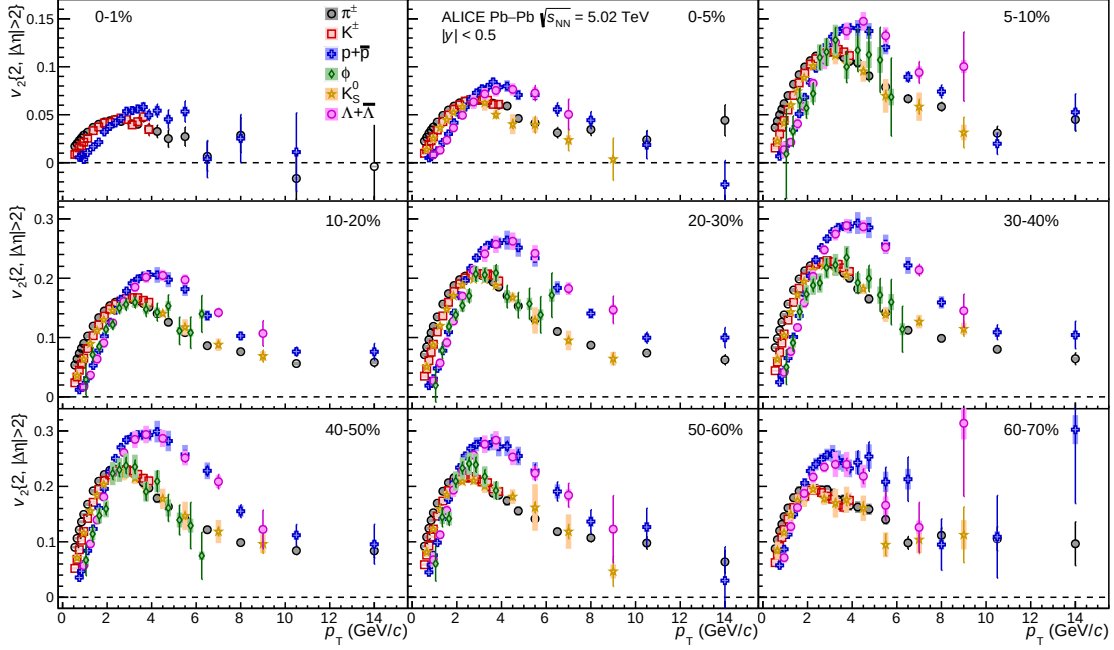


Figure 1.8: The measurements of the elliptic flow coefficient v_2 for charged pions, charged and neutral kaons, protons, ϕ mesons, and hyperons as a function of p_T in several centrality classes of Pb – Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV by ALICE. Figure from [38]

cellent probes of the properties of the Quark-Gluon Plasma. The heavy quarks, due to their large mass ($m_c \simeq 1.3 \text{ GeV}/c^2$ and $m_b \simeq 4.2 \text{ GeV}/c^2$), are produced in hard-scattering processes in the initial stages of the collision and subsequently interact with the QGP, allowing us to probe the properties of the QGP in greater detail. Furthermore, the study of open-heavy flavour physics also allows us to investigate the mechanisms of hadronisation, which is the process by which quarks and gluons combine to form hadrons. This process occurs in the later stages of the collision and is crucial for the formation of the final state hadrons that are detected by experimental detectors. Overall, the study of open-heavy flavour physics plays a crucial role in our understanding of the Quark-Gluon Plasma and the properties of hadronisation, providing important insights into the fundamental laws of nature.

This chapter provides an overview of open-heavy flavour hadrons and their evolution throughout the various stages of the collision, from production to interaction with the nuclear matters and eventual hadronisation.

1.5.1 Open heavy-flavour production

In pp collisions, the proton is composed of quarks and gluons, collectively referred to as *partons*. Inelastic pp collisions involve one or more inelastic parton-parton interactions. For high-energy parton



interactions, the QCD coupling α_S is small, allowing the cross section for the production of a parton c in a parton-parton collision $ab \rightarrow c + X$ to be computed using perturbative techniques. Assuming the parton-parton interaction occurs on much shorter timescales than any interactions that can occur among partons within the proton before the collision or with the outgoing parton after the collision, the production of a high- p_T or high-mass hadron h in a pp collision $AB \rightarrow h$ can be calculated using the QCD *factorisation theorem* [40]:

$$d\sigma_{AB \rightarrow h}^{\text{hard}} = f_{b/B}(x_1, Q^2) \otimes f_{a/A}(x_2, Q^2) \otimes d\sigma_{ab \rightarrow c}^{\text{hard}}(x_1, x_2, Q^2) \otimes D_{c \rightarrow h}(z, Q^2). \quad (1.13)$$

The idea is that interactions between partons within the colliding proton, or between outgoing partons and the hadronic final state, occur at longer distances and can be separated into three components: a perturbative parton-parton scattering cross section $d\sigma_{ab \rightarrow c}^{\text{hard}}$, and two non-perturbative terms. The parton distribution function (PDF) $f_{a/A}(x, Q^2)$ represents the probability of finding a parton with momentum fraction x and momentum transfer Q^2 in the proton A , while the fragmentation function (FF) $D_{c \rightarrow h}(z, Q^2)$ describes the probability of the outgoing parton c fragmenting into a hadron h with fractional momentum z . These functions cannot be calculated perturbatively and must be determined experimentally.

Both the PDFs and the FFs are crucial ingredients in calculating the production of hadrons in high-energy collisions, but they are inherently non-perturbative and cannot be calculated using pQCD. The PDFs describe the distribution of quarks and gluons within the proton and are typically parametrized using measurements of deep-inelastic scattering [41–43], while the FFs describe the probability that a parton fragments into a specific hadron and are determined from measurements in e^+e^- collisions [44]. These functions are defined at a specific energy scale Q^2 and must be evolved to the desired scale using the DGLAP equations [45]:

$$\mu_F \frac{\partial}{\partial \mu_F} f_{i/A}(x, \mu_F^2) = \frac{\alpha_S(\mu_F^2)}{2\pi} \int_x^1 \frac{dz}{z} P_{ij}(z, \alpha_S(\mu_F^2)) f_{j/A}\left(\frac{x}{z}, \mu_F^2\right), \quad (1.14)$$

$$\mu_R \frac{\partial}{\partial \mu_R} D_{c \rightarrow h}(z, \mu_R^2) = \frac{\alpha_S(\mu_R^2)}{2\pi} \int_z^1 \frac{d\xi}{\xi} P_{ck}(\xi, \alpha_S(\mu_R^2)) D_{k \rightarrow h}\left(\frac{z}{\xi}, \mu_R^2\right), \quad (1.15)$$

where P_{ij} and P_{ck} are the perturbative splitting functions for parton evolution. The factorisation scale μ_F is introduced to separate the perturbative and non-perturbative parts of the calculation, and it is usually chosen to be of the same order as the momentum transfer of the hard process. The renormalisation scale μ_R is used to remove divergences arising from virtual loop corrections in the pQCD calculation. Both



scales are typically chosen to be of the same order as the relevant momentum scale in the problem, such as the transverse momentum of the produced hadron in the hard process $\mu_R \sim \mu_F \sim \sqrt{m_Q^2 + p_{T,Q}^2}$ or $\sim \sqrt{4m_Q^2 + p_{T,Q}^2}$.

In addition to the factorization scale μ_F and the heavy quark mass m_Q , the cross section also depends on the square of the center-of-mass energy of the proton-proton collision $s = 4E_p^2 \simeq 4p_p^2$. Currently, perturbative calculations are performed up to next-to-leading order (NLO) $O(\alpha_S^3)$, with fixed-order next-to-leading-log resummation in GM-VFNS [46, 47] and FONLL [48, 49]. The next-to-leading-log corrections become more relevant at high p_T , where $p_T \gg m$ due to the quark mass m . Recently, next-to-next-to-leading-order (NNLO) calculations with QCD radiative corrections have been used to provide predictions for the $c\bar{c}$ and $b\bar{b}$ production cross sections [50–53].

Apart from the pQCD calculations, there is an alternative approach for describing the production of heavy-flavour hadrons that utilizes general purpose Monte Carlo (MC) generators, including PYTHIA [54] and HERWIG [55]. Unlike the pQCD approach, these generators offer a more complete representation of the final state, incorporating soft and multi-parton interactions, initial- and final-state parton showers, and a description of hadron decays. However, only leading-order processes are exactly implemented in these generators, and higher-order effects are considered with coarse-grained approaches. Furthermore, there are Monte Carlo event generators that provide next-to-leading order accuracy, such as MC@NLO [56] and POWHEG [57]. These generators not only predict differential observables and jets but can also be consistently matched to the parton showers generated by PYTHIA and HERWIG. Overall, the Monte Carlo generators offer an alternative tool for studying heavy-flavour hadron production and provide a complementary approach to the pQCD calculations.

Cutting-edge theoretical calculations of the production cross section of open heavy-flavour hadrons are furnished by two theoretical schemes: FONLL [48, 49, 58] and Gm-vfns [46, 47, 59–62]. Both methodologies utilize the collinear factorisation approach and carry out next-to-leading-order (NLO) pQCD calculations that incorporate the resummation of next-to-leading logarithms. However, they differ in their selection of the matching of the NLO and resummation calculations, the sets of parton distribution functions (PDFs) and fragmentation functions (FFs) used, and the values assigned to the masses of the charm and beauty quarks. In Fig. 1.9, the production cross section of prompt D^0 mesons, i.e., those originating from the hadronisation of charm quarks or the decays of excited charm-hadron states, in pp collisions at $\sqrt{s} = 5$ TeV [63] is compared with the FONLL and Gm-vfns calculations. The production cross section of non-prompt D^0 mesons, which arise from decays of beauty hadrons, is also shown. The pQCD calculations agree with the prompt D^0 -meson measurement within the uncertainties, demonstrat-

ing that the factorisation approach coupled with the FFs extracted from e^+e^- collisions is successful in the meson sector. The FONLL calculations provide a good description of the non-prompt D^0 production, whereas the Gm-vfns estimates underestimate the measured values. These disparities arise from the treatment of the fragmentation of beauty quarks into beauty hadrons and their subsequent decay to a D^0 meson. In particular, a better agreement is observed if the two processes are handled separately ($b \rightarrow H_b \rightarrow D$), as done to obtain the FONLL predictions, rather than being combined into a single step ($b \rightarrow D + X$) [62]. Additionally, predictions of the production of open heavy-flavour hadrons are made by the k_T -factorisation model [64], which considers an alternative factorisation scheme to the collinear approach. These estimates are based on leading-order calculations and incorporate higher-order corrections. The k_T -factorisation predictions reproduce the measurements at low and intermediate p_T but overestimate the data for transverse momenta above approximately 10 GeV/c.

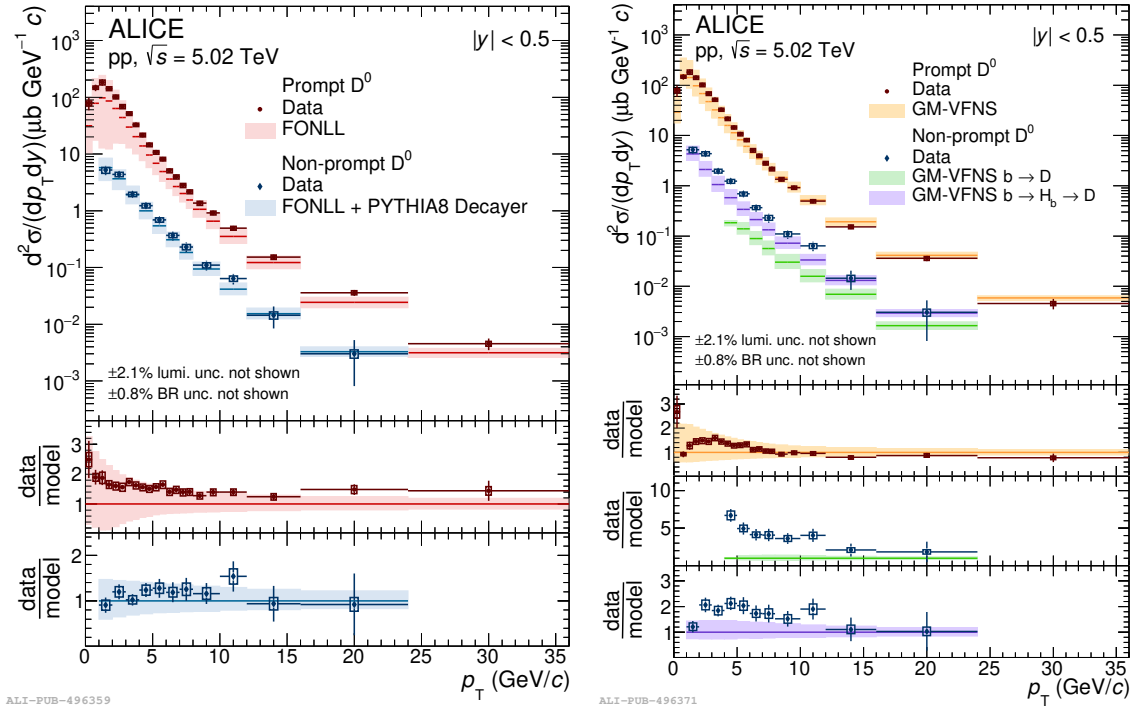


Figure 1.9: The measured p_T -differential production cross section of prompt and non-prompt D^0 mesons in pp collisions at $\sqrt{s} = 5$ TeV compare with the FONLL (left panel) and Gm-vfns (right panel) calculations. Figure from Ref. [63]

It is important to note that all the aforementioned statements regarding data-theory comparisons were specifically made for open heavy-flavour mesons. Nevertheless, the story is much different in the open heavy baryon sector, as evidenced by the first measurements of heavy-flavour baryon production at the LHC [65–68]. While theoretical predictions for the meson sector perform well, they significantly underestimate the baryon production measurements, as shown in Fig. 1.10. This figure compares the

prompt Λ_c^+ -baryon production cross section in pp collisions at $\sqrt{s} = 5$ TeV as measured by the ALICE Collaboration [65, 69] with Gm-vfns [59], PYTHIA 8 [54], and POWHEG [55] predictions, all of which significantly underestimate the data. More details about this puzzle will be discussed in Section 1.5.4.

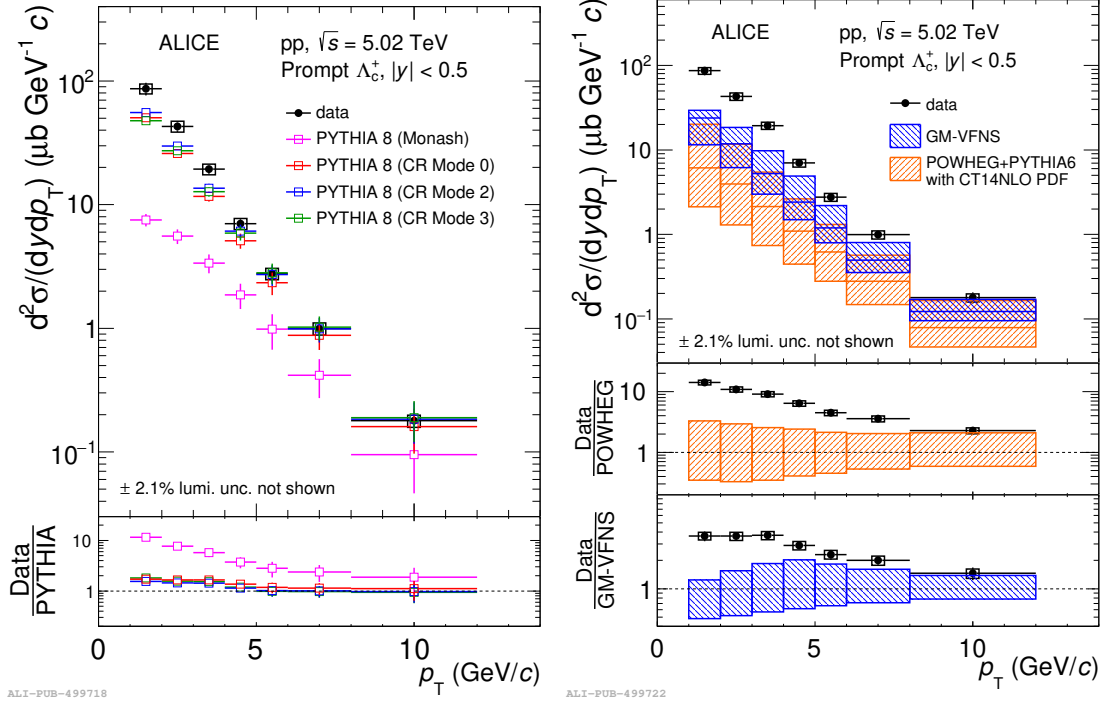


Figure 1.10: The measured p_T -differential production cross section of prompt and non-prompt D^0 mesons in pp collisions at $\sqrt{s} = 5$ TeV compare with the FONLL (left panel) and Gm-vfns (right panel) calculations. Figure from Ref. [63]

1.5.2 Interactions with the cold-nuclear matter

Before characterizing the properties of the medium produced in nucleus-nucleus collisions, it is crucial to quantitatively understand the cold-nuclear matter (CNM) effects that result from the presence of nuclei in the initial state. One way to achieve this understanding is by studying the production in proton-nucleus collisions. Open heavy-flavour production, in particular, is affected by two CNM effects: (i) a modification of the effective partonic luminosity due to the different dynamics of partons in nuclei and (ii) multiple scatterings of partons in the nucleus before and after the hard scattering.

At LHC energies, the first effect dominates for heavy flavours, and it is a consequence of the larger parton density in nuclei, which modifies the parton distribution function (PDF). The structure of a high-energy nucleus is much more complex than a superposition of non-interacting component nucleons, as discussed. The PDF for a bound proton in a nucleus with A nucleons, also known as the nuclear PDF (nPDF) $f^A(x; \mu_F^2)$, can be related to that of a free proton ($f^P(x; \mu_F^2)$) by the ratio:



$$R(x; \mu_F^2) = \frac{f^A(x; \mu_F^2)}{A \cdot f^P(x; \mu_F^2)}, \quad (1.16)$$

The modification of nuclear parton distribution functions (nPDFs) is believed to be caused by different mechanisms in different regions of x . This can be divided into four regions. At small $x \lesssim 10^{-2}$, a depletion ($R < 1$) known as *shadowing* [70] is observed, which is thought to be caused by phase-space saturation resulting from a high probability of gluon fusion in high-density gluon systems in the initial state. At intermediate values $10^{-2} \lesssim x \lesssim 10^{-1}$, a possible enhancement ($R > 1$) known as *anti-shadowing* is seen, which is generally explained as arising from the application of momentum conservation rules [71]. At large $x \gtrsim 10^{-1}$, there is the EMC effect (European Muon Collaboration effect), a depletion whose cause is still not fully understood [72]. Finally, an enhancement of the nPDF is observed at $x \gtrsim 0.8$ due to Fermi motion, which describes the movement of nucleons within the nucleus with larger transverse momentum than that of a free proton. There is no comprehensive theoretical calculation that explains the observed pattern across the entire x range, so nPDFs are determined through phenomenological parametrizations based on global fits to several experimental data sets [72, 73], similarly to proton PDFs. Heavy-flavour production at midrapidity at the LHC is predominantly influenced by the modification of the nPDFs in the shadowing region since it results in Bjorken $x \sim 10^{-4}$ for charm and $x \sim 10^{-3}$ for beauty [74].

The second CNM effect for heavy-flavour production is the multiple elastic scatterings of partons with the constituents of the target nucleus before and after the hard scattering. This process results in additional transverse momentum (k_T) being acquired by the parton after each scattering, which leads to a shift of the transverse momentum spectrum, commonly known as k_T -broadening or the Cronin effect [75]. The parton also experiences energy loss as it travels through cold-nuclear matter, which has a similar qualitative behavior to the final-state energy loss in the QGP, albeit with much smaller magnitude [76].

For a long time, it was believed that proton–nucleus collisions were only suitable for studying the CNM effects arising from the initial state. However, recent experimental data [77–79] have raised the question of whether the large number of particles produced in $p - A$ collisions can form a “medium” with some degree of collectivity. If this is the case, final-state effects will also affect the measurement of heavy-flavour production in such collisions, making it more difficult to quantitatively interpret the genuine CNM effects. In addition to in-medium energy loss mechanisms, which will be discussed in more detail in the next section, the acquisition of radial and elliptic flow may influence the p_T -spectrum



shape of open heavy-flavour hadrons. This was investigated in Refs. [80–82], which concluded that these flow effects, in a small and short-lived medium, do indeed leave signatures in the final hadronic observables of heavy-flavour hadrons, visible as an enhancement in the nuclear modification factor R_{pPb} at intermediate p_T .

Finally, it is important to note that extrapolating the CNM effects from $p - A$ collisions to A -collisions is not a straightforward task, given the possible interplay with final-state effects. However, studies have shown that such an extrapolation is possible for the relevant CNM effects in open heavy-flavour production. For nuclear PDF effects, as long as the factorisation theorem holds for the production cross section, the nuclear modification factors also factorise. Specifically, $R_{AA}^{CNM}(y) = R_{pA}(+y) \cdot R_{pA}(-y)$ [74]. Additionally, cold-nuclear matter effects due to multiple scatterings in A -collisions at midrapidity are roughly double the nuclear modification expected in $p - A$ collisions, owing to the symmetry in the $Q\bar{Q}$ production [83].

Fig. 1.11 shows the nuclear modification factor at $\sqrt{s_{NN}} = 5.02$ TeV in $p - Pb$ collisions for averaged non-strange D and Λ_c^+ hadrons [84, 85]. As previously discussed, obtaining a complete understanding of the cold-nuclear matter effects is challenging due to their interplay with other effects, including final-state effects. However, it can be concluded that the R_{pPb} for D mesons is consistent with unity, indicating that the effects are either insignificant or cancel each other out. Therefore, any suppression of D mesons in Pb – Pb collisions is expected to be a result of final-state interactions with a hot QCD medium. For Λ_c^+ baryons, an enhancement (depletion) in the R_{pPb} at intermediate (low) p_T is observed, albeit with large uncertainties. This suggests that Λ_c^+ production in Pb – Pb collisions should take these effects into account. Although theoretical calculations on the market cannot fully describe the R_{pPb} for Λ_c^+ , it is interesting to note the similarities with the light-flavour sector [79, 86, 87], where enhanced R_{pPb} values were also observed for baryons compared to mesons (and unity), believed to be due to radial flow.

1.5.3 Interactions with the hot-nuclear matter

Heavy quarks interact with the medium constituents losing energy via gluon radiation (inelastic) and collisional (elastic) processes [88]. The significant suppression observed at both RHIC [89–93] and LHC [94–102] energies of charm- and beauty-hadron production yields at intermediate and high transverse momentum ($p_T > 6$ GeV/ c) in heavy-ion collisions compared to appropriately scaled yields from proton-proton (pp) collisions, indicates a substantial energy loss of heavy quarks in the QGP. Heavy quarks in a medium experience momentum exchange with the heat bath, which is of the order of the medium temperature $Q^2 \approx T$, leading to several momentum kicks. Due to their typically low thermal

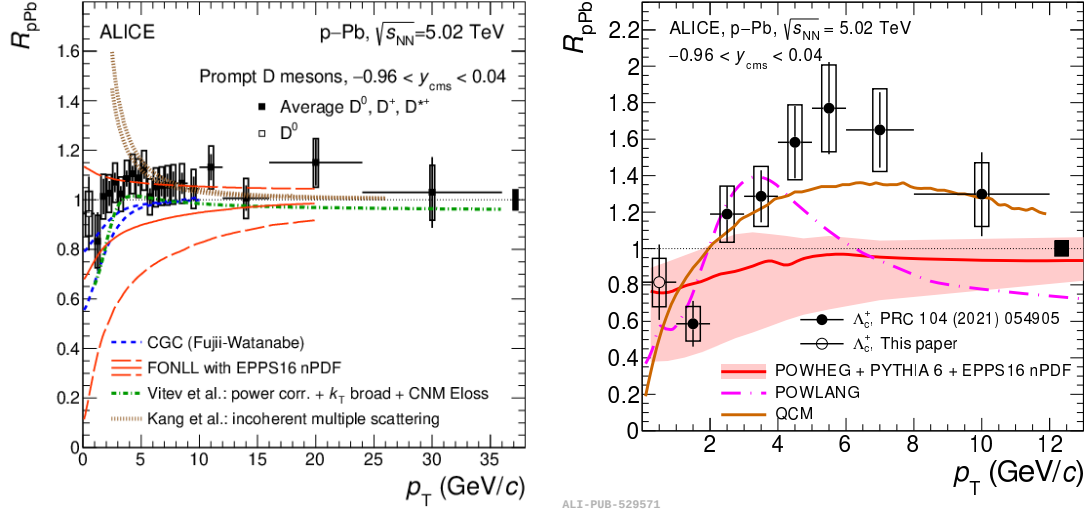


Figure 1.11: The nuclear modification factor R_{pPb} of prompt D mesons (an average of the non-strange D meson species [84] and Λ_c^+ baryons in p – Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV as measured by the ALICE Collaboration [85].

momentum, $p_Q^{\text{th}} = \sqrt{2m_Q T}$, they tend to execute a Brownian motion in the medium. The interactions with the constituents of the medium cause a loss of energy of the parton as it traverses the medium, but also provide a push to low- p_T quarks due to the radial flow of the bulk. Moreover, the heavy-quark thermalisation time is typically of the order of the medium lifetime or more [103], and therefore the momentum spectra of heavy-flavour hadrons carry the memory of the interaction history, making them excellent probes of the medium properties. For high momenta, the mass of heavy quarks becomes negligible ($m_Q \ll p$), and they behave like light particles, losing energy mainly via gluon radiation. Fig. 1.12 illustrates the energy loss via collisional and gluon radiation processes as a function of the quark energy for charm (left panel) and beauty (right panel) quarks, as computed in Ref. [104].

1.5.3.1 Energy loss

The mechanisms of collisional energy loss occur through scatterings of $q(g)Q \rightarrow q(g)Q$ and become significant for relativistic quarks. Let us first consider collisional energy loss in the ultra-relativistic limit ($p_Q \gg m_Q$), where similarities can be drawn with the scenario of a fast muon traversing a QED e^+e^- plasma. In this case, two partonic processes need to be considered, namely Coulomb collisions (with electrons and positrons) and Compton collisions (with photons). Ref. [105] presents a detailed review where it is demonstrated that the total energy loss due to elastic collisions will have a linear dependence on the in-medium path length and a logarithmic dependence on the initial parton energy.

The radiative energy loss in the medium is a result of $q(g)Q \rightarrow q(g)Qg$ processes, which are

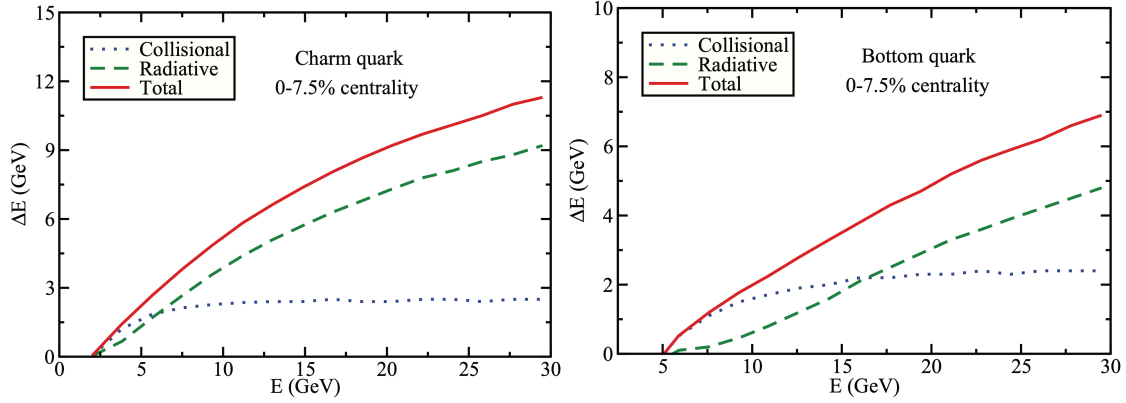


Figure 1.12: The left panel of the figure compares the radiative and collisional energy loss for charm quarks in the Quark-Gluon Plasma (QGP) created in central Pb – Pb collisions at a center-of-mass energy of $\sqrt{s_{NN}} = 2.76$ TeV. Similarly, the right panel of the figure compares the radiative and collisional energy loss for bottom quarks in the same system. Figures from [104]

similar to the electromagnetic bremsstrahlung process. The total energy loss due to radiation depends on the properties of the medium, particularly the length of the path in the medium, denoted by L . The long formation times of these inelastic scatterings lead to a coherent contribution, which means that the total energy loss $\Delta E = p_Q^{\text{in}} - p_Q^{\text{out}}$ can only be evaluated at the end of the in-medium path length. The quadratic dependence on L can be shown using the BDMPS-Z formalism [106].

$$\langle \Delta E_{\text{rad}} \rangle \propto \frac{\alpha_s C_R}{2} w_c = \frac{\alpha_s C_R}{4} \hat{q} L^2. \quad (1.17)$$

In this equation, w_c represents the characteristic energy of the emitted gluons, C_R are the Casimir factors associated with the QCD vertices, and $\hat{q}$ is the transport coefficient that characterizes the medium-induced k_T^2 transferred to the projectile per unit path length. The dead-cone effect, which suppresses gluon-bremsstrahlung in the collinear region for emitters with mass m and energy E , is also expected to introduce a quark-mass dependent correction [107]. This effect is particularly relevant for heavy quarks, while for the light-flavour sector, radiative energy loss remains the dominant mechanism even in the low- p_T regime. As a result, there is a flavour hierarchy in the radiative energy loss for relativistic energies, with $\Delta E(g) > \Delta E(q) > \Delta E(Q)$.

The nuclear modification factor (R_{AA}) is widely used to quantify the modifications of hadron spectra in A relative to pp collisions (discussed in sec.). As seen in Fig. 1.12, at high p_T , radiative energy loss is expected to dominate over collisional energy loss. Therefore, it is the primary mechanism causing the suppression of hadron yields at high p_T . By comparing the nuclear modification factor R_{AA} of

heavy- and light-flavour hadrons, it is possible to study the dependence of energy loss on quark mass and colour charge. Fig. 1.13 shows the average R_{AA} of prompt D^0 , D^+ , and D^{*+} mesons [108], charged pions [109], charged particles [30], and inclusive J/ψ mesons [110] measured by the ALICE Collaboration in central Pb – Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. The R_{AA} of prompt and non-prompt J/ψ mesons measured by the CMS Collaboration [111] is also shown. For $p_T \gtrsim 8$ GeV/ c , the R_{AA} of D mesons and charged pions are similar due to the interplay of larger energy loss expected for gluons and light-flavour quarks than for charm quarks and the harder charm-quark p_T distribution and fragmentation function. Additionally, the nuclear modification factor of non-prompt J/ψ mesons from beauty-hadron decays is higher than that of prompt D mesons, consistent with the quark-mass dependence of radiative energy loss. The dependence of radiative energy loss on the distance travelled by the parton in the medium is the origin of the positive elliptic flow observed at high p_T in non-central collisions. In this regime, the elliptic flow arises from path-length asymmetries resulting from the spatial anisotropy of the fireball in the first instants of the collision, leading to different energy losses depending on the heavy-flavour azimuthal direction [112, 113].

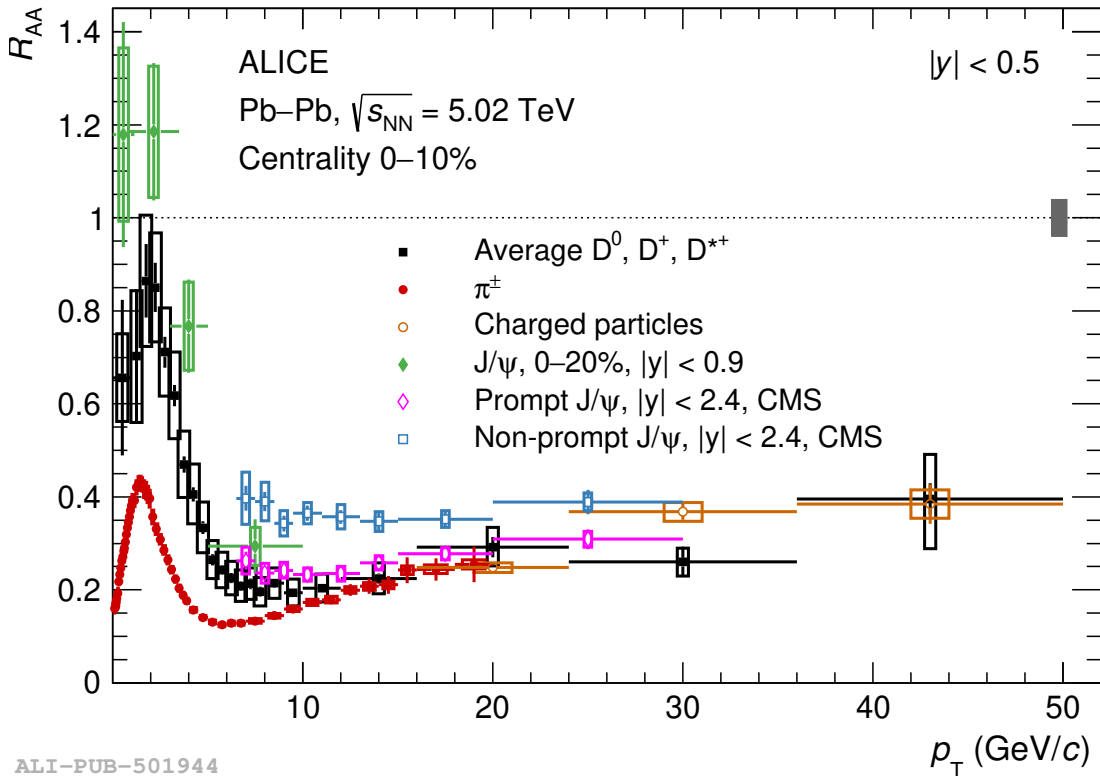


Figure 1.13: Average R_{AA} of prompt non-strange D mesons in the 0–10% centrality class compared to R_{AA} of charged pions, charged particles, inclusive J/ψ measured by ALICE, and of prompt and non-prompt J/ψ from CMS in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. Figure from [108]



1.5.3.2 Azimuthal anisotropy

In heavy-ion collision, the initial spatial eccentricity in the azimuthal plane is converted to particle momenta via the pressure gradients produced by the hot and de-confined medium. However, heavy quarks, which are produced in the initial hard scattering processes, have different dynamics from the light partons that make up the bulk of the medium. As a result, heavy quarks are not expected to be in equilibrium with the light partons at the formation time of the QGP, and therefore they only marginally participate in the collective motions of the bulk in the early stages of system evolution. Nevertheless, the multiple interactions of heavy quarks with the medium constituents transfer the bulk flow to heavy quarks, giving rise to non-vanishing elliptic and triangular flows in charm and beauty quarks [114, 115]. The anisotropic flow is quantified by the harmonic coefficients $v_n = \langle \cos[n(\varphi - \Psi_n)] \rangle$ of the Fourier expansion of the particle azimuthal angle (φ) relative to the collision symmetry planes with angles (Ψ_n) for the n^{th} harmonic. In non-central heavy-ion collisions, the second harmonic, called *elliptic flow* v_2 , is the largest coefficient. The heavy-flavour v_2 can be used to measure the extent to which charm and beauty quarks participate in the collective expansion of the medium at low p_T ($p_T < 6 \text{ GeV}/c$) [116], and the fraction of heavy quarks hadronizing via recombination with light quarks in the QGP medium in the intermediate p_T region ($6 < p_T < 10 \text{ GeV}/c$) [117, 118]. Conversely, the v_2 of heavy-flavour hadrons can be used to constrain the in-medium heavy-quark path-length dependence of energy loss at high p_T ($p_T > 10 \text{ GeV}/c$) [112, 113]. The coefficient v_3 , also known as the *triangular flow*, is primarily attributed to the event-by-event fluctuations in the initial distribution of nucleons and gluons within the overlap region [119].

Fig. 1.14 shows prompt D^0 , D^+ , and D^{*+} meson average v_2 and v_3 coefficients as a function of p_T in 0–10% and 30–50% centrality classes measured by ALICE collaboration [120]. The D-meson elliptic flow increases with centrality, while the triangular flow is similar in both classes. The v_2 coefficients of D mesons are lower than those of pions and protons for low p_T , consistent with the mass hierarchy. The v_2 coefficients for D mesons and J/ψ mesons in semi-central events follow the mass hierarchy, but no firm conclusions can be drawn for central events due to large uncertainties. For $p_T \gtrsim 4 \text{ GeV}/c$, the D-meson v_2 coefficients are similar to those of charged pions and lower than those of protons, supporting the quark coalescence hypothesis. In the 30–50% centrality class, the larger values of v_2 for D mesons compared to J/ψ mesons can be explained by hadronisation via coalescence and the fraction of J/ψ mesons coming from beauty-hadron decays. For $p_T \gtrsim 8 \text{ GeV}/c$, the measured v_2 coefficients for all hadron species are compatible within uncertainties.

To model the space-time evolution of heavy quarks undergoing multiple elastic scatterings in a QGP,

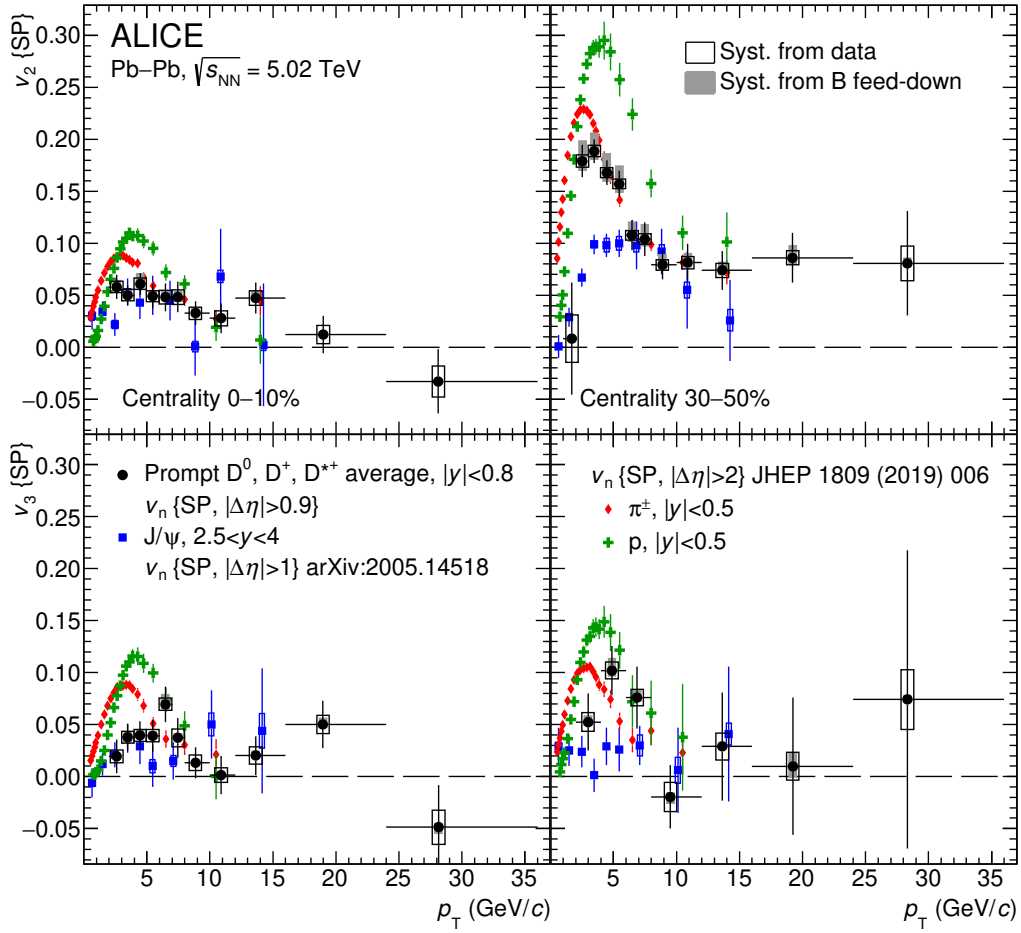


Figure 1.14: The p_T -dependent v_2 and v_3 coefficients for prompt D^0 , D^+ , and D^{*+} mesons in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV in the 0-10% (left panels) and 30-50% (right panels) centrality classes, along with the corresponding coefficients measured for π^+ , $p\bar{p}$ [121], and inclusive J/ψ mesons [122]. Figure from [120]



which is crucial as these scatterings transfer the radial and anisotropic flows of the QGP to the heavy quarks, transport theories based on the Boltzmann equation are employed [123]. As the mass of the heavy quark m_Q is much greater than the temperature T of the medium, the typical momentum exchange between the heavy quarks and the medium, $\vec{q}^2 \simeq T^2$, is small in comparison to the thermal momentum of the heavy quarks, $p_Q^{\text{th}} \simeq \sqrt{2m_Q T}$. This enables treating the heavy quark motion as Brownian motion. In this regime, the Boltzmann equation can be simplified to the Fokker-Planck equation [124], which can be stochastically solved using a Monte-Carlo simulation through the macroscopic Langevin equation [103]:

$$\frac{d\vec{p}}{dt} = -\eta_D(p)\vec{p} + \vec{\xi}, \quad (1.18)$$

The function $\eta_D(p)$ represents the frictional drag coefficient experienced by heavy quarks, while $\vec{\xi}$ accounts for random thermal momentum impulses. The drag coefficient is related to the momentum diffusion coefficient D_p (or the spatial diffusion coefficient D_s) through the Einstein fluctuation-dissipation relation, which states that $D_p = m_Q \eta_D T$ (or $D_s = T/(m_Q \eta_D)$). The spatial diffusion coefficient is closely linked to the heavy quark's thermalisation time $\tau_Q = m_Q/TD_s$. A higher value of D_s indicates a weaker interaction between the heavy quark and the medium (fewer interactions), resulting in a longer thermalisation time. Thus, heavy flavour production can be an effective way to explore the transport properties of the QGP.

Models attempting to describe the interactions between heavy quarks and the constituents of the QGP can be tested by comparing their predictions for both R_{AA} and v_2 with experimental measurements in Pb – Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, as illustrated in Fig.1.15. However, the majority of these models fail to simultaneously describe both observables in central and semicentral Pb – Pb collisions. By focusing on the few models that are in good agreement with both observables, it was possible to estimate the heavy-quark spatial diffusion coefficient to be within the range of $1.5 < 2\pi D_s T_c < 4.5$ at the pseudocritical temperature of $T_{pc} = 155\text{MeV}$ [108]. This interval is narrower than previous estimations based on D-meson measurements at LHC energies [120, 125]. Therefore, comparing experimental data with theoretical predictions for these observables has led to more precise constraints on the heavy-quark spatial diffusion coefficient and significant progress in understanding the interaction processes and hadronisation mechanisms of charm quarks in the high-density QCD medium.

1.5.4 Hadronisation mechanisms

The hadronization of quarks and gluons into color-neutral hadrons is a fundamental process in QCD that poses a significant challenge due to its nonperturbative nature. To address this issue, numerous mod-

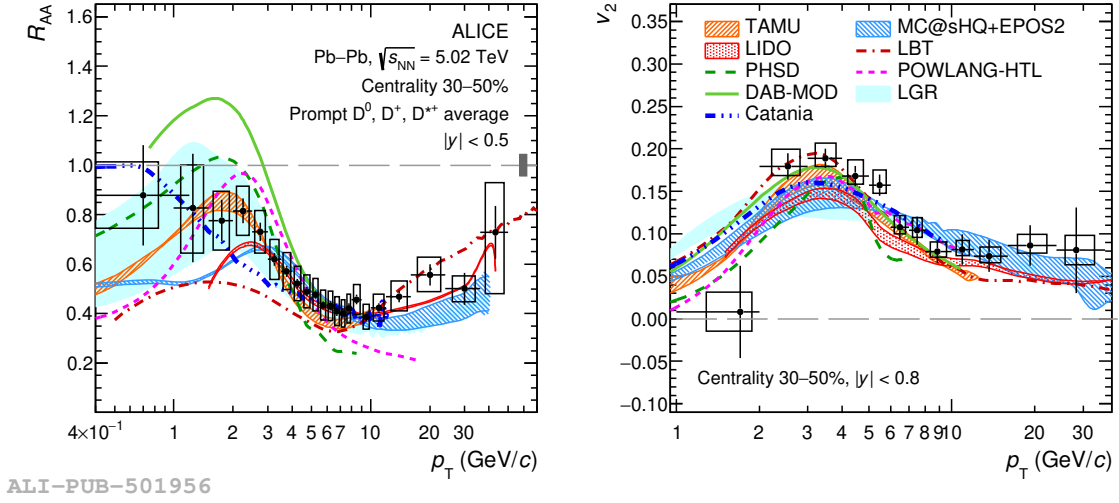


Figure 1.15: Average R_{AA} [108] (left) and elliptic flow v_2 [120] (right) of prompt D^0 , D^+ , and D^{*+} mesons in the 30–50% centrality classes compared with predictions of models implementing the charm-quark transport in a hydrodynamically expanding medium. Figure from [108]

els and parameterizations of hadronization processes have been developed through phenomenological studies of particle production in elementary collisions.

1.5.4.1 String fragmentation

The interest in the internal structure of hadrons arose in the 1970s when the first models on hadron formation were proposed, including the *Bag Model*. In this model, a hadron is considered as a “bag” of massless valence quarks surrounded by a perturbative QCD vacuum, which differs from the normal QCD vacuum outside the bag. The difference between them is characterized by the Bag constant (B), which quantifies the energy density difference between the perturbative and normal QCD vacuums. The bag constant value is approximately 200 MeV, obtained by minimizing the total energy of the bag as a function of its radius. With this approach, the critical temperature (T_c) for deconfinement is estimated to be approximately 150 MeV [126, 127].

In the same era, different approaches were used to treat quark formation and hadronization, with the most widely used being the *Lund string* model [128] based on color strings to account for $u\bar{u}$, $d\bar{d}$, and $s\bar{s}$ pair production [129, 130]. This model is currently implemented in Monash, the default tune of the PYTHIA event generator [131].

Default Pythia mode

In the Lund model, a $q\bar{q}$ pair in a meson is confined by the linear component of the Cornell potential. As the quarks move away from each other, the physical picture is that of a colour flux tube being stretched



between the q and the $\bar{q}$ as shown in Fig. 1.16 (a), the linear part becomes dominant and the Cornell potential is approximated as $V(r) \simeq \kappa r$, where κ is the string tension, typically assumed to be $\kappa \sim 1$ GeV/fm. For a massless $q\bar{q}$ pair flying apart at the speed of light, a colour string is formed between them with a constant tension until the stored potential in the string encodes all their kinetic energy. At that point, the colour string turns the quark and antiquark around, accelerating them towards each other until they meet back and repeat the sequence. In a high-energy e^+e^- collision, where the two leptons annihilate in a $q\bar{q}$ pair, this motion helps understand the structure of hadrons [132]. As shown in Fig. 1.16 (b), gradual distancing between the original quark and antiquark causes the colour string to stretch, and when it exceeds the amount required to produce an additional $q\bar{q}$ pair, the string can break into two $q\bar{q}$ pairs. The new $q\bar{q}$ pair created in a space-time point starts moving on a parallel direction with respect to the original light cone. Multiple string breaks may happen and produce new $q\bar{q}$ pairs, giving new $q\bar{q}$ systems, which are either hadrons or further fragments, until only hadrons remain. In this framework, the production of a meson with transverse mass m_T originating from a quark with flavour α that combines with an antiquark with flavour β is described by the fragmentation function:

$$f_{\alpha\beta}(z) = N_{\alpha\beta} \frac{1}{z} \cdot z^{a_\alpha} \left(\frac{1-z}{z} \right)^{a_\beta} e^{-bm_T^2/z}, \quad (1.19)$$

The parameter z represents the proportion of the initial energy-momentum carried by the generated meson, while $N_{\alpha\beta}$ is a normalization factor, b is a universal parameter for all vertices, and $a_{\alpha,\beta}$ are independent parameters associated with the quark flavors involved, often assumed to be identical.

A similar discussion can be done for massive quarks, $c\bar{c}$ and $b\bar{b}$ pairs, which are forced to move along a hyperbola due to their finite and non-negligible mass. The production of baryons can also be described within the framework of the Lund model, by considering the breakup of a color string into three quarks, with each quark carrying a fraction of the string's energy. The details of baryon production in the Lund model are more complex than for mesons, but the basic idea is the same: the string between the quarks stores energy as they move apart, and this energy is eventually released in the form of hadrons.

As discussed in Sec. 1.5.1, one successful approach involves the use of fragmentation functions with the Lund model in collinearly factorized production processes, which have been applied to describe a wide range of measured D-meson spectra, as depicted in Fig. 1.9. These functions are flavor-dependent but are generally assumed to be universal and can be constrained using e^+e^- or ep collision data and tested in hadronic collisions. Fig. 1.17 shows the fragmentation fractions ($f(c \rightarrow H_c)$) for charm hadrons obtained from a combined analysis of different experiments [134]. As seen in Fig. 1.10,

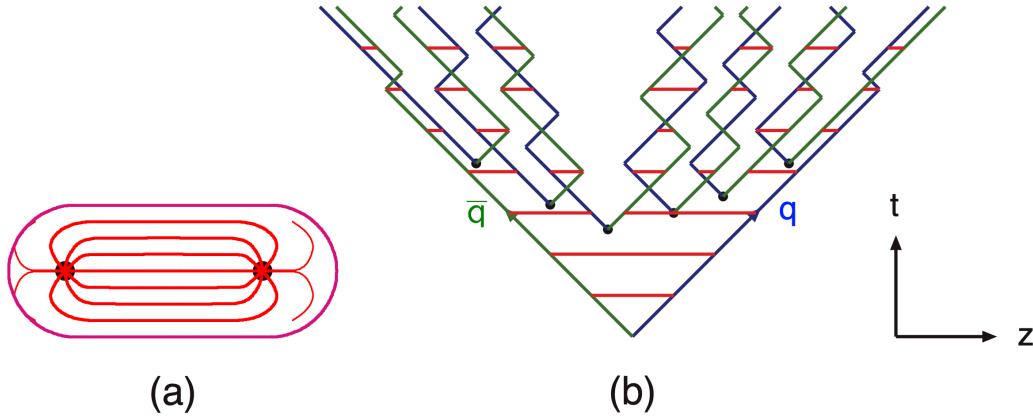


Figure 1.16: (a) A flux tube extends between a quark and an antiquark. (b) The movement and disintegration of a string system, with the two perpendicular degrees of freedom eliminated. The (anti)quarks are represented by diagonal lines, while snapshots of the string field are depicted by horizontal lines. Figure from [133]

this approach fails to describe baryon production, which is significantly larger than expected based on universal fragmentation functions constrained by e^+e^- and ep data. This behavior also be confirmed by the ALICE measurement of Ξ_c^0 in pp collisions at $\sqrt{s} = 5.02$ TeV as summarised in Fig. 1.17 [53]. Then It's worth to go deeper with the p_T -differential of Λ_c^+/D^0 ratio as shown in Fig. 1.18 in pp collisions at $\sqrt{s} = 5.02$ TeV. The experimental results have been found to be significantly underestimated by the predictions from Monash, which is a PYTHIA tune that implements conventional fragmentation. Specifically, at low p_T , the data exceeds the model predictions by a factor of approximately 5. To better comprehend this "puzzle", alternative theoretical approaches have been explored, and several models that deviate from the Lund model have been proposed.

Pythia colour reconnection modes

Monte Carlo generators often use the *leading-colour* (LC) approximation to trace the colour flow in an event [137]. This approach assumes that each quark is only connected to another coloured parton in the event, similar to a leading-colour QCD dipole, where the gluons carry a colour and an anticoulour charge, and are connected to two other partons. This can be interpreted as gluons forming transverse "kinks" on strings whose endpoints are a quark and an antiquark, as per the Lund model. However, this approach is not suitable for hadronic collision systems such as pp , where the initial state is sensitive to the strong force and coloured beam remnants remain after the collision. Furthermore, in such collision systems, *multiple parton interactions* (MPI) can occur, which are crucial to explain the underlying event and soft-inclusive physics, especially in high-multiplicity events. The possibility of phase-space overlaps among particles produced in different MPI events is non-negligible. Earlier works considered the possibility for different MPI systems to be colour-connected within a colour "chain", in order to minimize the total colour charge of the remnant and the number of strings stretched to it. However, the

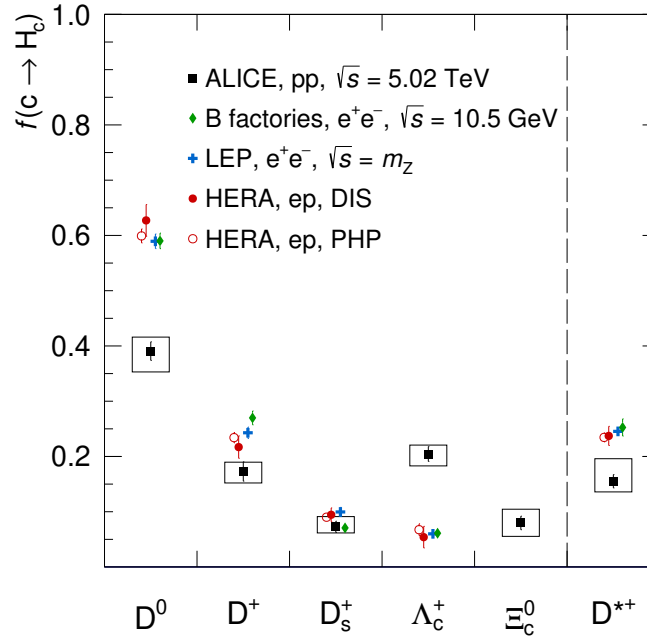


Figure 1.17: The fragmentation fractions of charm quarks into charm hadrons have been measured in pp collisions at a center-of-mass energy of $\sqrt{s} = 5.02$ TeV and compared to experimental results obtained from e^+e^- collisions at LEP and B factories, as well as from ep collisions at HERA. Figure taken from Ref. [53]

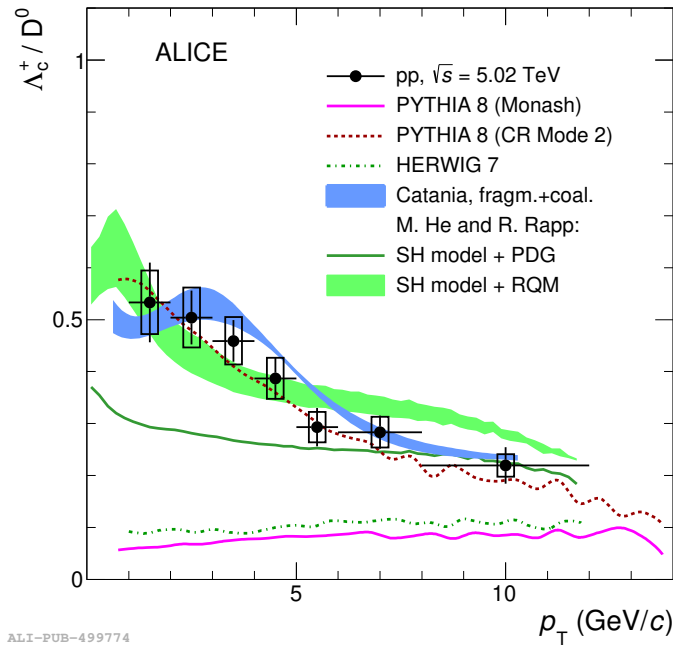


Figure 1.18: Λ_c^+/D^0 ratio in pp collisions at $\sqrt{s} = 5.02$ TeV compared with predictions from Pythia predictions with Monash and colour reconnection beyond leading colour approximation. As well as different theoretical calculations [135, 136]. Figures from from Ref. [53]



effect of this basic colour reconnection was estimated to be small, without significant changes in the per-particle spectra in high-multiplicity events. The average p_T of charged particles in such events is estimated by this approach to be flat as a function of particle multiplicities, but experimental data ruled out this possibility, highlighting the need to introduce non-trivial colour reconnection effects.

The pure SU(3) group structure of QCD provides some sub-leading combinations among different partons that can create colour singlets before single parton fragmentation. However, these additional combinations are insufficient to describe the measured baryon production in hadronic collision. Therefore, a revisited *colour reconnection beyond the leading colour approximation* (CR-BLC) has been implemented. This model assumes a “simplified QCD” framework, in which parton combinations are ruled by nine colour indices, leading to nine possible states for quarks and 72 for gluons, which allow recombinations among partons in different MPIs and from beam remnants. In this new framework, all possible combinations among partons in the whole event are considered, and this enriches the string topology that will undergo the hadronisation. The additional combinations introduced in this framework are the so-called *junctions* shown in the right panel of Fig. 1.19 and augment the set of possible string topologies that would be available with the LC structure. The possible topologies that can be realized in a specific event are dictated by the minimization of the λ *string-length measure*, which corresponds to the area spanned by the string before the hadronisation, namely the potential energy of the string.

A necessary condition for a colour string to be suitable for a reconnection is that its invariant mass must overcome a certain threshold m_0 , whose value influences the amount of possible reconnections. The parameter $C_j \equiv m_{0j}/m_0$, where a high m_{0j} corresponds to a low λ and $C_j > 1$ translates into an enhancement of junctions, finally governing the amount of reconnections. Another ingredient that influences the string topology is the causal connection among the strings themselves. If either string has already hadronised before the formation of the other, there is no space-time region available for the reconnection among them to occur. To make sure that two strings are able to reconnect, they must resolve each other between the string formation and its hadronisation, taking into account the time dilation occurring due to the relative boost. The different “modes” of this model depend on the requirement and strength of this causal contact criterion. In the Mode 0, no time-dilation constraint is applied, and the amount of colour reconnection is dictated only by the invariant-mass threshold m_{0j} . In Mode 2, the strictest condition is applied, where the causal contact is required to be maintained between the strings. Mode 3 relaxes the constraint by considering only the first reconnection between two dipoles, even if there are subsequent reconnections that occur later.

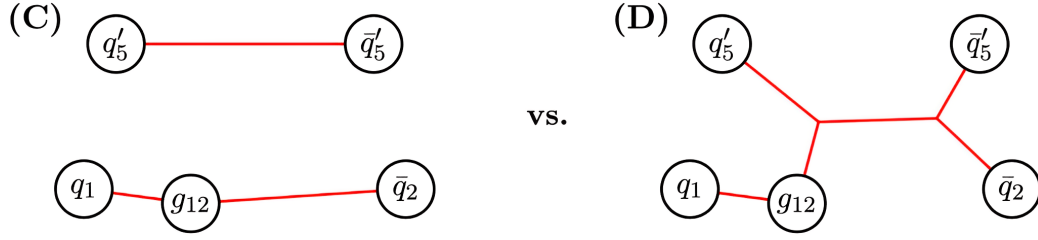


Figure 1.19: Illustration of a multi-parton state, with a leading-colour string topology (C) and with a junction-antijunction string topology (D), allowed by cyclically matching “2” and “5” indices. Figure from [137]

1.5.4.2 Coalescence or recombination

In heavy-ion collisions, the *recombination* or *coalescence* process occurs due to the presence of surrounding free color charges. This process allows heavy quarks to produce hadrons by associating with lighter quarks picked up from the deconfined medium instead of fragmenting. Fig. 1.20 displays the interplay between coalescence and fragmentation processes. The production of baryons with a given p_T (p_h) is preferred over mesons with the same transverse momentum, or at least, the baryon-to-meson ratio is increased at intermediate p_T , compared to a scenario with pure quark fragmentation. As the quark p_T distribution decreases rapidly, low- p_T quarks are more abundant. Thus, it is more likely to form a baryon by combining three low- p_T quarks, rather than creating a meson. This is because in the latter case, the two quarks require a larger p_T to reach the final p_h . An enhancement of the relative abundance of baryons with respect to mesons in heavy-ion collisions has been observed at RHIC [138] and the LHC [139] firstly for the light-flavour sector. A similar behaviour has been recently observed also for charm hadrons (Λ_c^+/D^0 ratio) both in RHIC [140] and LHC [141]. In pp collisions at the LHC as shown in Fig. 1.18, a similar baryon-to-meson enhancement as observed in AA collisions, which also good described by Catania model [135] with a hybrid hadronisation via fragmentation and coalescence is assumed for the charm quark.

1.5.4.3 Statistical Hadronization Model

Statistical hadronisation models (SHMs) [142] aim to describe a fully thermalised system statistically, based on its macroscopic properties, in contrast to macroscopic models acting on the parton level. Hadron abundances are predicted from the volume V , temperature T , and chemical potential μ of the thermalised system at the chemical freeze-out during heavy-ion collisions. Assuming an equilibrated system, its macroscopic properties can be derived from the grand canonical ensemble with the partition function:

$$Z(T, V, \mu) = \text{Tr} \left[\exp \left(-\beta \left(H - \sum_i \mu_{Q_i} Q_i \right) \right) \right], \quad (1.20)$$

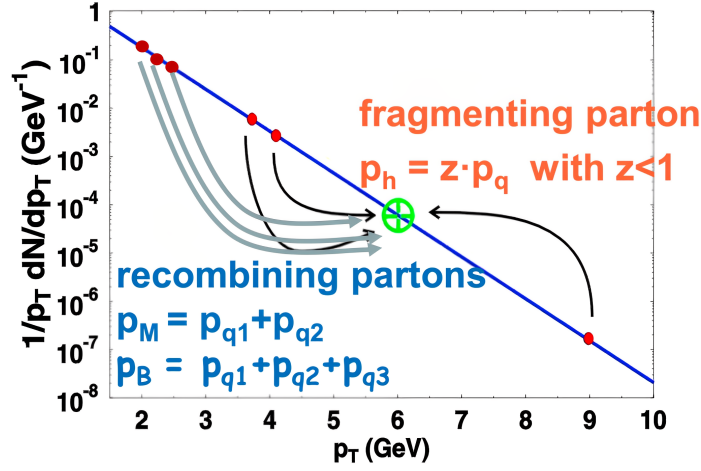


Figure 1.20: The interplay between coalescence and fragmentation processes.

Here, $\beta = 1/T$, μ_{Q_i} are the chemical potentials to conserve the charge Q_i on average in the system, and H is the Hamiltonian that describes the hadron resonance gas. The conserved quantities in such a system are the electrical charge, strangeness, and the baryon number with their chemical potentials denoted as μ_Q , μ_S , and μ_B , respectively.

According to [143], the formation of single hadrons is independent, and the overall partition function can be approximated as the product of the partition functions Z_h of each hadron species h :

$$Z \approx \prod_h Z_h. \quad (1.21)$$

The fraction of each species per volume V can then be written as:

$$n_h = \frac{N_h}{V} = \frac{T}{V} \left(\frac{\partial \ln Z_h}{\partial \mu_h} \right)_{V,T}, \quad (1.22)$$

where μ_h is the sum over the chemical potentials applicable to h . Strangeness is conserved to be 0 in the overall system, and the electrical charge is conserved and given from the studied collision system in a specific rapidity region [144]. In a modeling approach, T , μ , and V are free parameters that can be obtained from a fit to measured data. Once determined, SHMs can be used to predict the particle abundances at a given chemical freeze-out temperature T_{cfo} . Note that to obtain the correct production ratio of each hadron species h , contributions from higher mass resonant states decaying to h must also be taken into account.



The statistical hadronisation model has been successful in describing the production of light and strange hadrons in heavy-ion collisions as well as in smaller systems. In the heavy flavour sector, this model incorporates an underlying charm-baryon spectrum from the PDG, or alternatively, includes excited charm baryons predicted by the Relativistic Quark Model (RQM) but not yet observed [136]. These additional states decay strongly to Λ_c^+ baryons, which contribute to the prompt Λ_c^+ spectrum. However, the RQM predictions are subject to uncertainty related to the branching ratios of the excited baryon states into Λ_c^+ final states, which is estimated by varying the branching ratios between 50% and 100%. While the model underpredicts the data when using the PDG charm-baryon spectrum, it provides a good description of the pp data when including the additional baryon states, both in terms of the magnitude of the ratio and the decreasing trend with p_T [53].

1.6 Motivation and structure of thesis

Measurements of open charm- and beauty-meson production in proton–proton (pp) collisions are successfully described by quantum chromodynamics (QCD) calculations based on the factorization of the soft (non-perturbative) and hard (perturbative) processes [18, 63, 145–154]. Within this approach, the production cross sections of heavy-flavor hadrons are computed as the convolution of (i) the parton distribution functions (PDFs) of the incoming protons, (ii) the perturbative partonic cross section, and (iii) the fragmentation functions (FFs) describing the transition from the heavy quark to the hadron. The partonic cross section in these calculations is typically computed at next-to-leading order accuracy with all-order resummation of next-to-leading logarithms (e.g., FONLL [48, 49, 58] and Gm-vfns [46, 47, 59–62]), but recently calculations at next-to-next-to-leading order (NNLO) also became available [155]. The FFs are instead parametrized from measurements performed in e^+e^- or ep collisions [156], assuming that the hadronization process of charm and beauty quarks is independent of the collision system. However, the production of the Λ_c^+/D^0 baryon measured in proton-proton (pp) collisions at center-of-mass energies of 5.02 and 7 TeV at the ALICE experiment is significantly underestimated by the theoretical calculation. Similar phenomena have also been observed in the proton-lead collisions at the central rapidity region of ALICE and the forward rapidity region of LHCb. Other heavy-flavored hadrons such as Ξ_c^{+0} measured recently in pp collisions also confirmed that the fragmentation function is not universal for the collision system. In order to gain a more comprehensive understanding of fragmentation functions and the formation mechanism of heavy-flavored hadrons, a detailed examination of non-universal phenomena is essential. To this end, a study was conducted on the system size dependence of open charm baryons and non-prompt D mesons. This involved measuring the production of open-charm hadrons in various charged particle multiplicity ranges. Through this approach, a deeper insight into the behavior of heavy-flavored hadrons can be obtained, shedding light on the underlying mechanisms that govern



their formation.

However, in heavy-ion collisions, it remains an experimental question whether bottom quarks with their larger mass are thermalized in the quark-gluon plasma (QGP). Recently, the ALICE and ATLAS collaboration observed a significant positive elliptic flow of electrons and muons from bottom hadron decays, indicating that bottom quarks could also participate in the thermalization process [157, 158]. In heavy ion experiments, due to the large mass of the bottom quark, whether it will be fully thermalized in QGP has always been an experimental question. Based on the current detector spatial resolution and luminosity, measuring the elliptic flow of non-prompt D mesons can better study the interaction between bottom quarks and the QGP.

This thesis aims to study the hadronization mechanisms of heavy quarks and their interaction with the quark-gluon plasma (QGP) by measuring the production yields and elliptic flow of charmed hadrons and non-prompt D^0 mesons in the ALICE experiment at the LHC energy region. For the hadronization mechanisms of heavy quarks, the dependence of the prompt Λ_c^+ production on the charged particle multiplicity and the proportion of non-prompt D^0 mesons as a function of the charged particle multiplicity, as well as the non-prompt D^0 meson yield, were measured using proton-proton collision data collected from 2016 to 2018 at a center-of-mass energy of 13 TeV (about 1800 million events). For the study of the interaction between bottom quarks and QGP, the elliptic flow of non-prompt D^0 mesons in non-central collisions was measured using lead-lead collision data collected in 2018 at a center-of-mass energy of 5.02 TeV. Chapter 1 introduces the theoretical background of QCD, QGP, heavy-ion experimental evolution and the main observables, open charm and heavy flavor physics. Chapter 2 introduces the ALICE experiment's detectors, their functions, and the ALICE offline analysis framework and data samples used in this thesis. Chapter 3 mainly discusses the two advanced analysis method, including the reconstruction of candidate particles based on machine learning algorithms and the data-driven method used to estimate prompt and non-prompt fraction. Chapter 4 introduces the measurements of the open charm-hadron in pp collisions at 13 TeV, which include the production of prompt Λ_c^+ baryons as a function of charged particle multiplicity, the non-prompt D^0 mesons fraction as a function of charged particle multiplicity and its production cross section. Chapter 5 describes the first measurement of the elliptic flow of non-prompt D^0 mesons at a center-of-mass energy of 5.02 TeV in the ALICE experiment. Chapter 6 summarizes all the experimental measurements and gives prospects for future research.



2 ALICE: A Large Ion Collider Experiment

The ALICE experiment [159], one of four main experiments at the CERN Large Hadron Collider (LHC) [160] near Geneva, is designed to investigate heavy-ion collisions at ultra-relativistic center-of-mass energies. Its primary objective is to analyze the quark-gluon plasma created in Pb – Pb collisions, characterized by extreme temperature and energy density [161]. In addition, the physics program of ALICE includes the study of p – Pb and pp collisions to address various QCD topics. This chapter presents an introduction to the detectors utilized in the ALICE experiment, along with their performance specifications. Finally, the computing and software framework of the ALICE Collaboration is briefly discussed.

2.1 The ALICE detectors

The final setup of ALICE until LHC Run 2 is shown in Fig. 2.1. The experiment is divided into two main components: the central barrel and a muon arm, covering the mid-rapidity region ($|\eta| < 0.9$) and the forward rapidity range ($-4 < \eta < -2.5$), respectively [161]. The central barrel detectors are situated within the L3 magnet, a large solenoid providing a 0.5 T magnetic field along the beam axis, which is necessary for tracking and identifying charged particles. The central barrel comprises the Inner Tracking System (ITS), the Time Projection Chamber (TPC), the Transition Radiation Detector (TRD), the Time Of Flight (TOF) detector, the Ring Imaging Cherenkov (HMPID) detector, and the electromagnetic calorimeters PHOS, EMCal, and DCal. In addition to these, ALICE also include various other detectors for global event characterization and triggering (ZDC, PMD, FMD, T0, V0A, and V0C) located at forward and backward rapidity, as well as an array of scintillators (ACORDE) positioned atop the L3 magnet for triggering on cosmic rays. More detailed descriptions of all subdetectors are provided in subsequent sections.

2.1.1 Inner Tracking System

The Inner Tracking System (ITS) consists of a six-layer silicon detector and is positioned closest to the interaction point and is designed for tracking charged particles and determining the primary vertex, as shown in Fig. 2.2. Moreover, it plays a crucial role in reconstructing secondary vertices, i.e., the points of decay for short-lived hadrons, and measuring the impact parameter of tracks, defined as the distance of closest approach between the track trajectory and the primary vertex. The ITS uses three different technologies for its inner, middle, and outer layers.

The two innermost layers of the ITS are equipped with Silicon Pixel Detectors (SPD), which, owing

THE ALICE DETECTOR

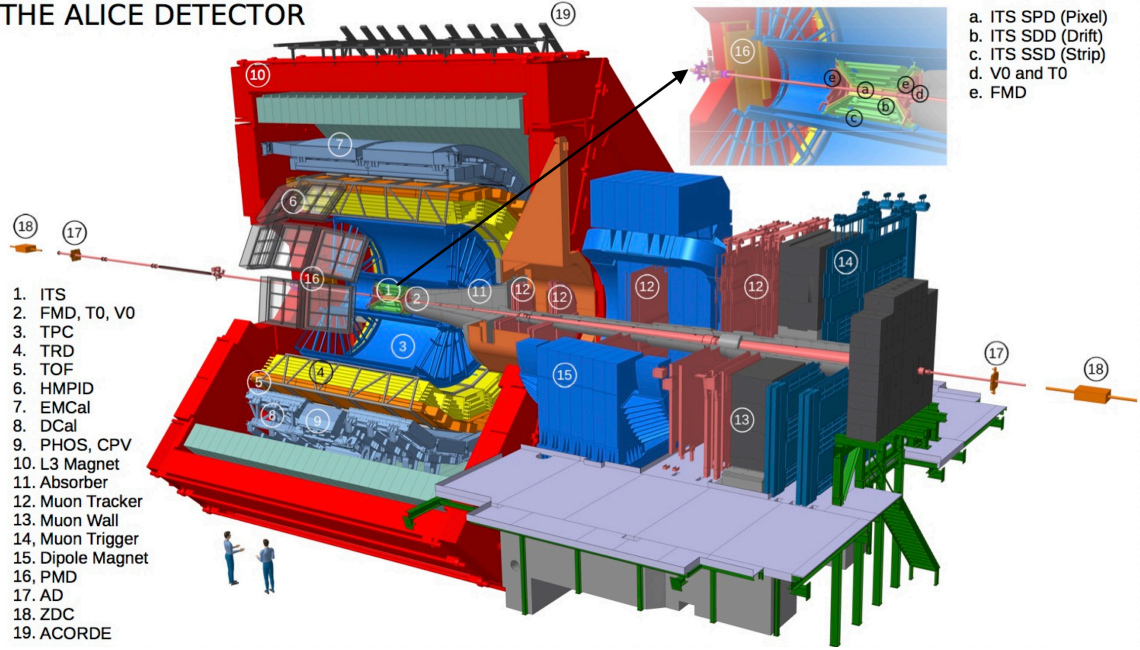


Figure 2.1: ALICE schematic layout

to their position and resolution, are fundamental for the reconstruction of weakly-decaying strange, charm, and beauty hadrons. The SPD layers also serve as a trigger due to their fast response and provide a midrapidity multiplicity estimator via tracklets, which are track segments built by associating pairs of hits in both layers. The third and fourth layers of the ITS are Silicon Drift Detectors (SDD), which provide excellent multitrack capability and the possibility of particle identification via dE/dx measurements. However, this latter feature is not used for this dissertation because other detectors possess excellent PID capabilities. The two outermost layers of the ITS are equipped with double-sided Silicon Strip Detectors (SSD), which are crucial for the prolongation of tracks from the main tracking detector of ALICE (i.e., the Time Projection Chamber) to the ITS. The SSD layers also provide dE/dx information. Table 2.1 summarizes the main differences between the three technologies.

Table 2.1: parameters of ITS layers [159].

Layer	r (cm)	Material budget ($\%X_0$)	Active area per module (mm^2)	Resolution (μm^2)
SPD	3.9	14.1	12.8×70.7	12×100
SPD	7.6	14.1	12.8×70.7	12×100
SDD	15.0	22.2	70.17×75.26	35×25
SDD	23.9	29.7	70.17×75.26	35×25
SSD	38.0	43.1	73×40	20×830
SSD	43.0	48.9	73×40	20×830

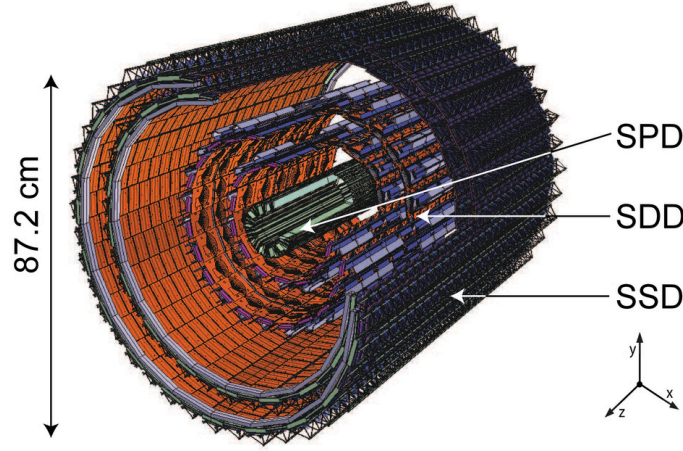


Figure 2.2: Structure of the Inner Tracking System, figure taken from [162]

2.1.2 Time Projection Chamber

The main tracking detector of the ALICE experiment is the Time Projection Chamber (TPC). Its main functions are to measure the momentum of charged particles, identify them, and reconstruct the interaction vertex along with the ITS. The TPC has a cylindrical shape with an internal radius of approximately 85 cm, an external radius of approximately 247 cm, and a length of around 510 cm, making it the largest TPC currently mounted on an experimental apparatus. Its total active volume is approximately 88 m³. The azimuthal acceptance is 360°, while the covered pseudorapidity range is $|\eta| < 0.9$ for particles that can reach the TOF detector. The pseudorapidity acceptance for slower particles is $|\eta| < 1.5$.

The TPC is divided into two sectors by a central cathode maintained at a high negative potential of $V \sim -100$ kV. This generates a constant electric field of $E \approx 400$ V/cm thanks to the action of the external field cage. The internal volume is divided into two sections, each approximately 2.5 m long, filled with a Ne/CO₂/N₂ mixture (90/10/5). This mixture allows for a maximum drift time for electrons of approximately 90 μ s. However, the disadvantage of this mixture is its strong velocity dependence on temperature, making it necessary for the TPC to maintain proper thermal stability ($\Delta T \leq 0.1$ K). Electrons produced in the gas ionization from charged particles passing through the TPC undergo a drift motion towards the cylinder basis. The cylinder basis is azimuthally segmented into 18 trapezoidal sections, divided into an inner and an outer region, each equipped with a Multi-Wire Proportional Chamber (MWPC), for a total of 36 shown in Fig. 2.3. The two MWPCs together count 159 readout pad rows, with the cathodes collecting the avalanche originated by the primary electrons. This signal is used to reconstruct the x - y projection of the particle trajectory, while the spatial information along the beam direction is provided by the drift time, measured thanks to a buffering of the collected charge with a fixed

frequency. The TPC detector provides a three-dimensional track reconstruction for charged particles and delivers a resolution of 1% for low-momentum particles ($p_T \sim 1 \text{ GeV}/c$) with a magnetic field of 0.5 T. The resolution grows to approximately 3.5% for $p_T \sim 100 \text{ GeV}/c$. Another important function of the TPC is to identify charged particles by measuring their energy loss per unit length as they pass through the gas, using the amplitude of the MWPC signal.

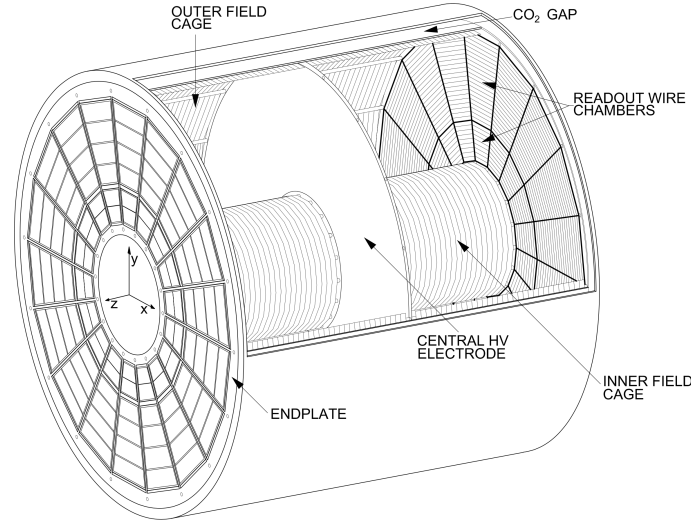


Figure 2.3: Layout of the Time Projection Chamber. Figure from [163]

2.1.3 Time-of-Flight detector

The cylindrical Time of Flight (TOF) detector has a 360° azimuthal acceptance and covers a pseudorapidity range of $|\eta| < 0.9$. The modular structure of the TOF detector is shown in Fig. 2.4. Its dimensions include an internal radius of approximately 370 cm and an external radius of around 399 cm. The detector consists of double-stacked Multi-gap Resistive-Plate Chambers (MRPCs), each with an active area of $120 \times 7.4 \text{ cm}^2$. The MRPCs have two cathodes and a central anode, and a stack of fine glass plates with a $250 \mu\text{m}$ gap between them. A uniform electric field of around kV/cm is applied throughout the detector's volume. When a charged particle interacts with the gas in the active volume, producing ionization charges, an electron avalanche is generated in every vacuum space between two plates. The plates themselves stop the formed avalanches, but the electrodes detect the fast signals induced by the ionization electrons' movement. The final signal detected on the electrodes is the sum of the signals coming from every vacuum space between two plates.

The TOF detector is made up of five modules, tilted to suppress oblique transversal paths that can cause signal sharing among adjacent pads. Each detecting module is made up of strips transversally oriented with respect to the beam direction. To cover the full cylinder along the z -direction, modules of

varying lengths are used. The central module is approximately 117 cm long, intermediate modules are around 137 cm, and the external modules are roughly 177 cm. These modules are combined to create one of the 18 supermodules that make up the barrel. The TOF detector is a dedicated system for Particle Identification (PID), which measures the time of flight of charged particles on a fixed length L and with a given momentum p measured by the TPC. More details on the PID are available in Sec. 3.7. During the LHC Run 2, the TOF detector achieved a time of flight resolution of about 56 ps [164].

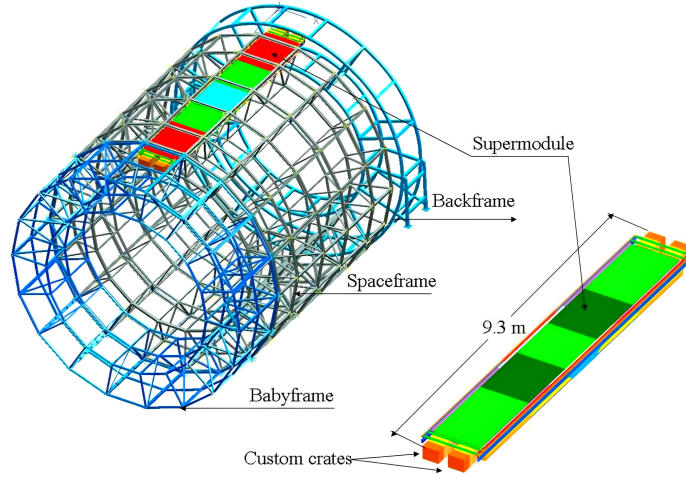


Figure 2.4: Modular structure of the ALICE TOF detector. Figure from [161]

2.1.4 V0 and T0 detector

The V0 detector consists of two arrays of scintillator counters, V0A and V0C, which are divided into 4 rings radially and 8 sectors azimuthally. V0A is located opposite to the muon spectrometer at a distance of 340 cm from the nominal interaction point along the beam axis, while V0C is positioned in front of the hadronic absorber at a distance of 90 cm from the interaction point. The detectors cover the full azimuthal range and the pseudorapidity ranges of $2.8 < \eta < 5.1$ for V0A and $-3.7 < \eta < -1.7$ for V0C. The V0 provides a Minimum Bias trigger (MB) by taking a logical AND between the signals from V0A and V0C. The timing information from V0 is utilized to reject background events generated by beam-gas interactions and pileup during the SPD readout time. Additionally, the V0 signals are employed to estimate the collision centrality and the reaction plane in Pb–Pb collisions [161].

The T0 detector comprises two arrays of Cherenkov counters, T0A and T0C, which cover the full azimuthal range and the pseudorapidity intervals of $4.61 < \eta < 4.92$ and $-3.28 < \eta < -2.97$, respectively. It is used to determine the start time for the TOF detector and as the primary online luminosity monitoring detector.



2.2 Detector performance

2.2.1 tracking and vertexing reconstruction

Particle track finding [161] is a crucial step in the ALICE experiment, and it is performed using the Inner Tracking System (ITS) and Time Projection Chamber (TPC). Before this procedure, the detector data undergoes a *clusterisation* process where the data is converted into clusters using information such as positions, signal amplitudes, and signal times with associated errors. The interaction vertex is then determined using *tracklets*, which are lines defined by pairs of clusters in each Silicon Pixel Detector (SPD) layer. The vertex is found when the maximum number of tracklets converges onto a single point.

The tracks are then reconstructed in three stages. In the first stage, the tracks are searched for starting at large radius at the edge of the TPC, where the track is built inwards. Track "seeds" are built first with two TPC clusters and the interaction vertex, then with three clusters and no vertex, and are propagated inwards. At each step, the nearest cluster is updated provided it fulfills a proximity cut. Tracks are accepted if they have a minimum number of clusters and do not miss more than 50% of expected clusters for the position of the track. Reconstruction of the same track is avoided by rejecting tracks which share a fraction of common clusters, where the worse of the two tracks is rejected. After the track has been propagated to the inner TPC edge, a preliminary particle identification is performed based on the specific energy loss, and this information is used in the following steps. The tracks are then propagated to the outermost ITS layers, becoming seeds for the ITS track finding, and the tracks are propagated inwards based on clusters in the ITS layers. At this point, any "path" layer-to-layer is saved provided it passes a proximity cut, such that each TPC track produces a tree of track hypotheses in the ITS. The highest quality candidate from each hypothesis tree is saved based on the χ^2 of each path, and each track is extrapolated to the point of closest approach to the interaction vertex.

The second stage involves propagation outwards. The tracks are refitted using a Kalman filter technique [165] in the outward direction using the same clusters from the first stage. The tracks are extended and matched to clusters in the Transition Radiation Detector (TRD) and Time-of-Flight (TOF) where a time-of-flight is assigned. Finally, in the third stage, all tracks are propagated inwards from the TPC to the ITS, where the tracks are refitted.

The impact parameter resolution is shown as a function of transverse momentum (p_T) in figure 2.5 for all charged particles in three collision systems: proton-proton (pp), proton-lead (p – Pb), and lead-lead (Pb – Pb) collisions. Due to the larger multiplicities in heavier systems, the interaction vertex can be determined more precisely, and as such, the impact parameter resolution is improved. Topological



selection on charmed weakly-decaying particles such as D mesons and Λ_c^+ baryons also becomes more effective at higher p_T due to both the improved impact parameter resolution and the longer decay length of the particle in the lab frame.

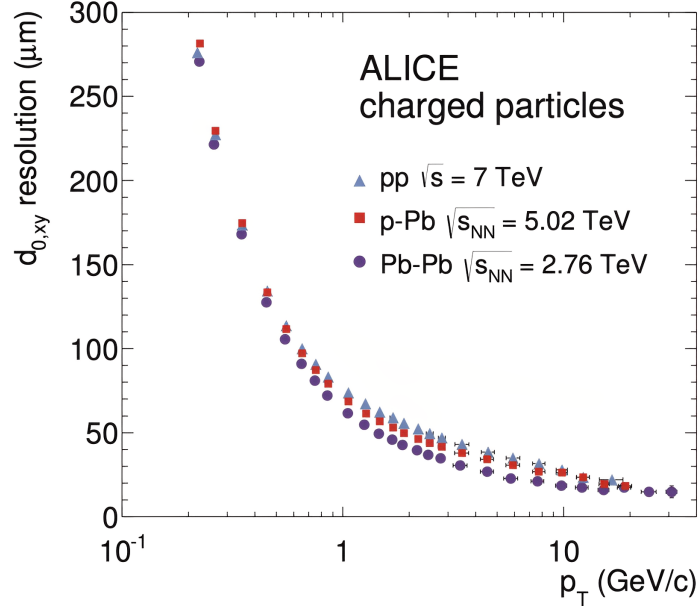


Figure 2.5: The impact parameter resolution as a function of p_T in pp collisions at $\sqrt{s} = 7$ TeV, in p – Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV and in Pb – Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Figure from [161]

2.2.2 Particle identification

The ALICE experiment is characterised by excellent particle-identification capabilities in a wide momentum range, which result from the different and complementary techniques used by the experiment detectors. Charged hadrons are identified by exploiting measurements of their specific energy loss in the ITS and TPC detectors, via particle time-of-flight measurements with the TOF detector, and through angle measurements of the Cherenkov light produced traversing the HMPID detector material. Electrons and positrons are identified via transition-radiation measurements with the TRD detector, and by comparing the energy deposited in the PHOS and EMCal calorimeters with their momentum. These latter two detectors can also reconstruct photons. The analyses presented in this thesis employ the PID information provided by the TPC and TOF detectors, whose particle-identification techniques and performances are described in more details in the following Sections.

TPC particle identification

Particle identification (PID) in the TPC relies on the measurement of the specific energy loss dE/dx , charge, and momentum of particles that pass through the detector gas. To calculate the dE/dx of a track, the 65% lowest-amplitude charge measurements from up to 159 pad rows are averaged, thereby eliminating secondary-ionisation contributions. Using this measured dE/dx and the expected value



for a given particle species at a specific momentum, PID is performed by comparing the two. The expected dE/dx is obtained from a data-driven parametrization based on either splines or the Bethe-Bloch formula, originally adopted by the ALEPH collaboration [166]. The formula is given

$$f(\beta\gamma) = \frac{P_1}{\beta^{P_4}} \left[P_2 - \beta^{P_4} - \ln \left(P_3 + (\beta\gamma)^{-P_5} \right) \right] \quad (2.1)$$

where β is the particle velocity, γ is the Lorentz-boost factor, and π are the fit parameters. The left panel of Fig. 2.6 shows the measured specific energy loss of different particle species as a function of momentum in Pb – Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. The expected mean energy losses for electrons, pions, kaons, protons, deuterons, tritons, and ^3He nuclei are also plotted. The TPC has a relative dE/dx resolution of about 5% in pp collisions, which degrades to 6% in heavy-ion collisions with high multiplicity [166]. While good pion-kaon and kaon-proton separations are achieved for transverse momenta below 0.8 GeV/ c and 1.6 GeV/ c , respectively. However, in the relativistic-rise region ($2 < p_T < 20$ GeV/ c), the relative abundances of particle species can be determined using a statistical approach.

TOF particle identification

The TOF detector is used to identify particle species by measuring their time of flight from the interaction point to the detector chambers. The start time for the TOF measurement is determined by either the signal from the T0 detector or the particle arrival times at the TOF detector. A combinatorial minimisation algorithm is used when both methods are available to account for all possible mass hypotheses. The distribution of particle velocities β as a function of momentum is displayed in the right panel of Fig. 2.6 for Pb – Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. The electron, pion, kaon, proton, and deuteron bands are clearly separated, while the background is caused by tracks that are incorrectly matched to TOF hits. The overall TOF resolution is approximately 56 ps for pions with a momentum of 1.5 GeV/ c in Pb – Pb collisions. This value includes the intrinsic detector resolution, the contribution from electronics and calibration, the uncertainty on the event start time, and the tracking and momentum resolution. The TOF detector provides a good pion-kaon separation for $p_T < 3$ GeV/ c and a satisfactory kaon-proton discrimination up to $p_T \approx 5$ GeV/ c . However, at $p_T < 1$ GeV/ c , the fraction of tracks matched to the TOF detector is small, so low- p_T tracks without TOF particle-identification information are typically identified using only the TPC.

2.2.3 Centrality determination

Lead beams accelerated at ultra-relativistic energies are used to create the QGP in the laboratory, reaching energy-density values exceeding 1 GeV/fm³ [161]. The topology of a single Pb – Pb collision

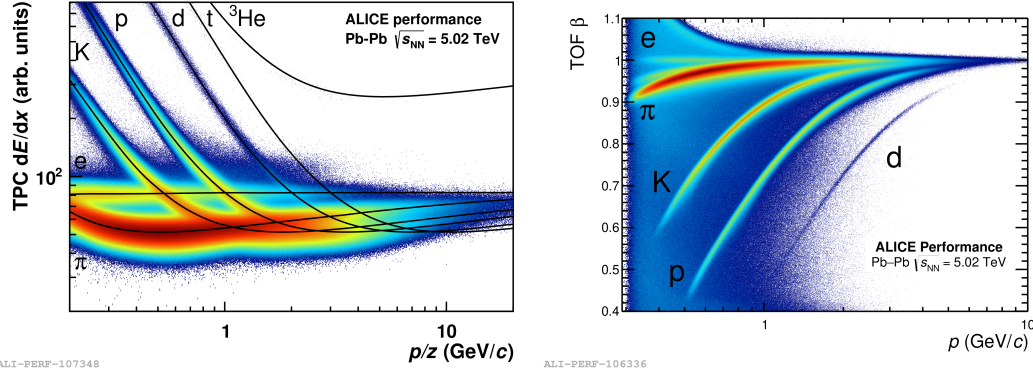


Figure 2.6: The left panel displays the specific energy loss (dE/dx) in the TPC as a function of the particle momentum in Pb – Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. The lines represent the expected mean energy loss based on parametrization. On the other hand, the right panel shows the velocity β of particles reaching the TOF detector as a function of momentum in Pb – Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. This figure clearly separates the electron, pion, kaon, proton, and deuteron bands. Figures from the ALICE figure repository

is characterized by the impact parameter b , defined as the distance between the centers of the two colliding ions on the transverse plane, and the $\hat{z}$ axis defining the reaction plane. The number of nucleons participating in the collision (N_{part}) and the number of binary interactions among nucleons (N_{coll}) depend on b , as shown in Fig. 2.7. b determines the centrality of the event, expressed as a fraction of the total hadronic cross section σ_{AA} :

$$c(b) = \frac{\int_0^b \frac{d\sigma}{db'} db'}{\int_0^{+\infty} \frac{d\sigma}{db'} db'} = \frac{1}{\sigma_{\text{AA}}} \int_0^b \frac{d\sigma}{db'} db' \quad (2.2)$$

The impact parameter cannot be measured directly, so collision centrality is estimated based on the event's particle activity, specifically the produced particle multiplicity (N_{ch}). It is assumed that the particle multiplicity increases monotonically with event centrality, allowing for experimental calculation of the centrality using the following formula:

$$c \approx \frac{1}{\sigma_{\text{AA}}} \int_{N_{\text{ch}}}^{+\infty} \frac{d\sigma}{dN'_{\text{ch}}} dN'_{\text{ch}} = \frac{1}{n_{\text{ev}}^{\text{tot}}} \int_{N_{\text{ch}}}^{+\infty} \frac{dn_{\text{ev}}}{dN'_{\text{ch}}} dN'_{\text{ch}}, \quad (2.3)$$

The total number of collisions is denoted by $n_{\text{ev}}^{\text{tot}}$, and the number of events within a certain event multiplicity dN'_{ch} is represented by dn_{ev} . The event multiplicity used to estimate the event centrality is typically determined using the V0 detectors. The distribution of the sum of signal amplitudes on the two V0 modules is utilized to determine the percentile of the hadronic cross section associated with each event, which is linked to a given value of V0 signal amplitudes. The centrality percentile is determined



with respect to an "anchor point," which identifies the $\sim 90\%$ of the total hadronic cross section and determines the absolute scale of the centrality. This choice is made because collisions with an event multiplicity lower than that of the "anchor point" are strongly contaminated by electromagnetic events, which are not related to the strong interaction between the colliding beams and therefore are not used for centrality determination. Moreover, low-multiplicity events are characterized by a rapidly decreasing trigger efficiency, resulting in a larger number of rejected events. Fig. 2.7 displays the distributions of the sum of V0 signal amplitudes for Pb – Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV [167]. The distribution of particle multiplicity produced in Pb – Pb collisions is described by the analytical function:

$$g(n) = P_{\mu,k}(n) \times [f \cdot N_{\text{part}} + (1 - f)N_{\text{coll}}] \quad (2.4)$$

where the particle multiplicity produced in a single nucleon-nucleon interaction is described by the Negative binomial Distribution (NBD):

$$P_{\mu,k}(n) = \frac{\Gamma(n+k)}{\Gamma(n+1)\Gamma(k)} \cdot \frac{(\mu/k)^n}{(\mu/k+1)^{n+k}}. \quad (2.5)$$

Eq. 2.5 gives the probability for a single ancestor to produce n particles, with μ as mean multiplicity and k as the distribution width parameter $D := \sqrt{\langle n^2 \rangle - \langle n \rangle} = \sqrt{\mu(1 + \mu/k)}$. To determine the multiplicity distribution probability in a Pb – Pb collision with N_{part} participant nucleons and N_{coll} binary collisions, Eq. 2.5 is multiplied by the number of ancestors described in the second factor of Eq. 2.4. This combination is based on two-component models [168, 169] that decompose a nucleus-nucleus collision into soft processes proportional to N_{part} and hard processes proportional to N_{coll} . N_{part} and N_{coll} are calculated using the Glauber model [170]. Centrality class limits are defined as percentiles of the distribution's total integral, with a centrality resolution ranging from 0.5% in central events to 2% in peripheral Pb – Pb collisions.

2.3 The ALICE offline framework

The offline framework used by ALICE is based on ROOT software [171], which is specifically designed for high-energy physics experiments. It is used for handling big data processing, statistical analyses, and data visualization and storage. The ALICE offline framework is employed for a variety of purposes, including simulations, detector alignment and calibration, event reconstruction, and analysis of the experimental data collected. Simulations are generated using Monte Carlo event generators. In particular, the PYTHIA [172, 173] generators are used for pp collisions, while the HIJING [174] gen-

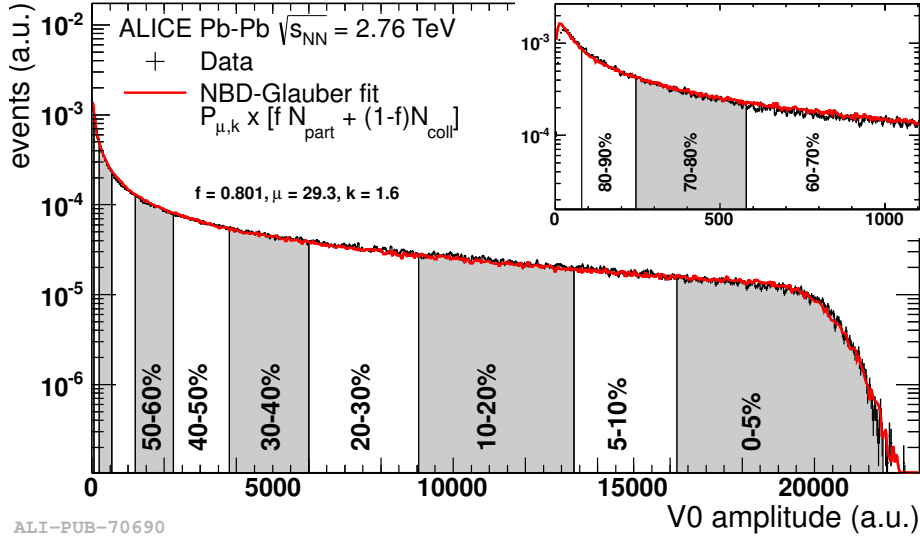


Figure 2.7: Distribution of the V0 amplitudes measured in Pb – Pb collisions at $\sqrt{s_{NN}} = 2.76\text{TeV}$. The red line reports the NBD-Glauber fit, while the bands show the different centrality classes. Figure from Ref. [167]

erator is used for p–Pb and Pb–Pb collisions. The simulations take into account a detailed description of the detector, and the transport packages like GEANT [175] are used to simulate the responses of the detectors to the passage of particles. The simulations are anchored to actual data-taking runs to ensure the same detector and data-taking conditions are used. The reconstructed data, both real and simulated, are stored in Event Summary Data (ESD) ROOT files, which contain all the physical information required for analyses and quality assurance for each subdetector. To reduce the output size of these files and therefore reduce the computational time required for analysis, Analysis Object Data (AOD) ROOT files are produced, containing only the information necessary for analyses.

All computational tasks, including generating simulations, reconstructing data, and performing analyses, are carried out on the worldwide LHC computing grid. This infrastructure connects several computing centers from more than 40 countries, enabling the processing of vast amounts of data in a timely manner.

2.4 Data samples

The data analysed in this thesis were collected by the ALICE detector in LHC Run 2, during the 2016, 2017, and 2018 pp runs at $\sqrt{s} = 13\text{TeV}$ and the 2018 Pb – Pb run at $\sqrt{s_{NN}} = 5.02\text{TeV}$. All data samples were collected with the minimum-bias trigger, although enhanced with the two high-multiplicity triggers and a centrality triggers for pp collisions at $\sqrt{s} = 13\text{TeV}$ and Pb – Pb at $\sqrt{s_{NN}} = 5.02\text{TeV}$, respectively. The characteristics of the data samples are summarised in Tab. 2.2. In particular for pp collisions, more



detailed with the periods related to year-by-year are shown in Tab. 2.3.

Table 2.2: Summary of the characteristics of the used data samples for the analyses presented in this thesis. The numbers for the Pb – Pb data samples are after selecting a centrality interval of 30-50%.

System	$\sqrt{s_{NN}}$	Trigger	N_{events}	$\mathcal{L}_{\text{int}}(\mu\text{b}^{-1})$	Year
pp	13	MB	1800×10^6	32000 ± 1600	2016-2018
	13	HMV0	1000×10^6	n.a.	2016-2018
	13	HMSPD	300×10^6	n.a.	2018
Pb – Pb	5.02	MB or mid-central	85×10^6	55.5 ± 0.2	2018

Table 2.3: Data taking periods and number of events for minimum bias (MB), high-multiplicity V0 (HMV0), and high-multiplicity SPD (HMSPD) triggers in different years.

Year	Trigger	Periods	Number of events
2016	MB	d, e, g, h, j, k, l, o, p	424×10^6
	HMV0	k, l, o, p	
2017	MB	e, f, h, i, j, k, l, m, o, r	580×10^6
	HMV0	e, f, h, i, k, l, m, o, r	
2018	MB	b, d, e, f, g, h, i, j, k, l, m, n, o, p	706×10^6
	HMV0	b, d, e, f, h, j, k, l, m, o, p	
	HMSPD	f, h, j, k, l, m, o, p	

For pp collisions, monte Carlo simulations were utilised in the analysis mainly for the machine-learning training, and to obtain the correction factors for the limited detector acceptance as well as the reconstruction and selection efficiencies. They were obtained by simulating pp collisions with the PYTHIA 8.243 event generator [172, 173] (Monash-13 tune [131]). In order to enrich the simulated data samples of prompt and non-prompt D mesons, either a $c\bar{c}$ or $b\bar{b}$ quark pair was required in each simulated PYTHIA pp event and D mesons were forced to decay into the hadronic channels of interest for the analysis. Then the simulated Pb–Pb events were obtained with the HIJING v1.383 generator [174]. Similarly, in each simulated event, additional $c\bar{c}$ and $b\bar{b}$ -quark pairs were injected using the event generator [172, 173] (Monash-13 tune [131]) and D mesons were forced to decay into the hadronic channel of interest for the analysis. The generated particles in both case were propagated through the detector using the GEANT3 transport package [175]. The conditions of all the ALICE detectors in terms of active channels, gain, noise level, and alignment, and their evolution with time during the data taking period, were taken into account in the simulations.



3 Advanced analysis strategy

Heavy hadrons, due to weak decays with mean proper decay length being at most a few hundred micrometers (shown in Tab. 3.1), can not be detected directly as they decay before reaching the first detector layer. They can be detected via indirect measurements via semileptonic decays and b-hadrons decay products like non-prompt J/ψ and D mesons. Meanwhile, full reconstruction of exclusive hadronic decay channels: $D^0 \rightarrow K^- \pi^+$, $D^+ \rightarrow K^- \pi^+ \pi^+$, $D_s^+ \rightarrow \phi \pi^+$, $\Lambda_c^+ \rightarrow p K_S^0$ and $\Lambda_c^+ \rightarrow p K^- \pi^+$ would provide more information related to HF parents. To suppress the large combinatorial background, stringent criteria based on the properties of these decay channels are necessary. In addition to utilizing the particle identification abilities of ALICE, the detector's spatial resolution enables the reconstruction of the secondary decay vertex, providing a crucial selection opportunity. Most physics analyses aim to extract a physics "signal" with high precision. This is often achieved by reducing the "background" through selection criteria based on certain features of the signal. These features can include topological and kinematic properties of the signal or detector responses, and exploit our knowledge of the physics behind the signal. While tighter selections reduce the background, they can also lead to a reduction in the signal. Therefore, optimizing the signal/background levels is crucial. In high-energy physics, multivariate techniques are commonly employed to perform this task.

Recently, machine learning (ML) algorithms have gained popularity in high-energy physics for a variety of tasks, including classification and regression problems, as well as simulation and event reconstruction [176, 177]. The algorithms utilize samples of known signal/background make-up (often simulated) to respond to samples of unknown make-up, thus optimizing the signal/background separation in these samples. This chapter will present a machine learning (ML) technique that was employed to improve the extraction of the prompt charm hadron signal and separate it from the non-prompt charm hadron background in this thesis. Additionally, an advanced data-driven method that was used to estimate the prompt and non-prompt fraction in this thesis will also be introduced in this chapter.

Table 3.1: The mean proper decay length of heavy hadrons. Taken from PDG [178]

Particle	$c\tau$ [μm]
D^0	123.01
D^+	309.8
D_s^+	151.2
Λ_c^+	60.4

3.1 Machine-learning based candidate selection

3.1.1 Boosted Decision Trees and XGBoost

Boosted Decision Trees (BDTs) is a popular machine learning algorithm used in high-energy physics. They are based on *Decision Trees* (DT), which are a type of supervised learning algorithm that can be used for both classification and regression tasks. Fig. 3.1 shows a simple decision tree (DT) example for a binary classification problem between two classes, S and B, using numerical features (labelled as var_i). The thresholds for each node of the tree are labelled as c_i . The logic rules used by DT algorithms are relatively easy to interpret, making it advantageous in terms of comprehensibility compared to other machine-learning algorithms. At each node of the decision tree, the feature and its corresponding value that best separates the classes is chosen, and the training data is divided accordingly. The quality of the partition can be evaluated using different criteria, such as the Gini index or the cross-entropy, and the feature that results in the greatest decrease in impurity is chosen for each split. This process is repeated recursively until all instances in a node belong to the same class, or a stopping condition set by the user is met. At this point, the node becomes a leaf of the tree.

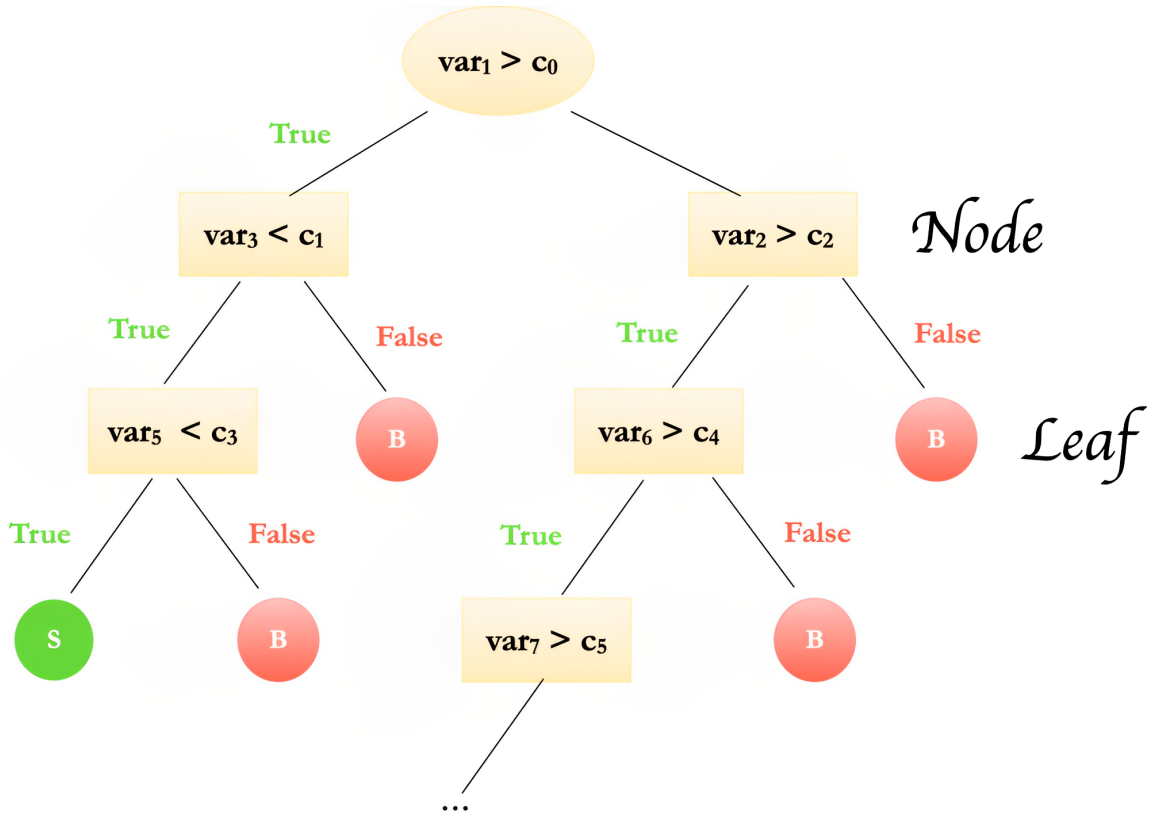


Figure 3.1: A simple Decision Tree Structure

The boosted part of BDTs comes from the fact that multiple decision trees are combined in an iter-



ative process, with each subsequent tree trained to correct the errors of the previous ones. This process effectively boosts the performance of the individual decision trees, resulting in a more powerful and accurate classification or regression algorithm. BDTs have several advantages in high-energy physics analyses, including their ability to handle a large number of input features, their robustness to overtraining, and their ability to handle both continuous and categorical variables. They have been used for a wide range of applications, including particle identification, event classification, and parameter estimation.

Large decision trees with many nodes can classify the training set well, but they tend to have poor performance on unknown data, which is why small trees with few nodes are preferred. Although small trees have a lower discrimination power, their performance on the training and test set is similar. As a result, small trees can learn general patterns from the data rather than the data itself, as large trees do. To create a model with good classification power, numerous decision trees are combined, usually in the order of hundreds. This is accomplished through a technique called boosting, where trees are constructed sequentially, compensating for the shortcomings of one tree with the others. Boosting can be achieved through the AdaBoost algorithm [179], where during the training of a DT, higher importance is given to the correct classification of samples that were misclassified by the preceding tree. After all the decision trees have been built, their outputs are combined to obtain a prediction. The ensemble of these decision trees forms the BDT model. The AdaBoost algorithm is a special case of the gradient boosting technique [180].

The analyses presented in this thesis use the implementation of BDT provided by the XGBoost [181] library, an open-source package that provides an efficient implementation of gradient boosting algorithms. It was developed by Tianqi Chen and his team at the University of Washington and has been widely adopted in industry and academia for its accuracy and scalability [182]. XGBoost supports both classification and regression tasks, and is especially useful in handling large datasets with high dimensionality. The key features of XGBoost include its ability to handle missing values, regularization techniques to prevent overfitting, and parallel processing capabilities. It has won several machine learning competitions on Kaggle and is known for its high predictive accuracy and speed. In multi-class classification, XGBoost can be used to train an ensemble of decision trees that can classify data into multiple classes. XGBoost supports multiple objective functions for multi-class classification, including multi-softprob, multi-softmax function. The multi-softprob objective function outputs a probability distribution for each class, while the multi-softmax objective function outputs the class with the highest probability.

3.1.2 BDT training with variables

A sketch of the $D^0 \rightarrow K^- \pi^+$ decay is shown in Fig.3.2. Weakly decaying charmed hadrons have proper decay lengths $c\tau$ greater than or of the order of the impact parameter resolution of the ALICE tracking detectors. The reconstruction of weakly-decaying particles in ALICE is therefore based on exploiting the displacement of the particles from the primary vertex. The variables used as BDT input include kinematic variables (variables relating to the kinematics of the final state), topological variables (variables relating to the decay topology), and PID variables (variables relating to particle identification information). typically the variables are used in this analysis:

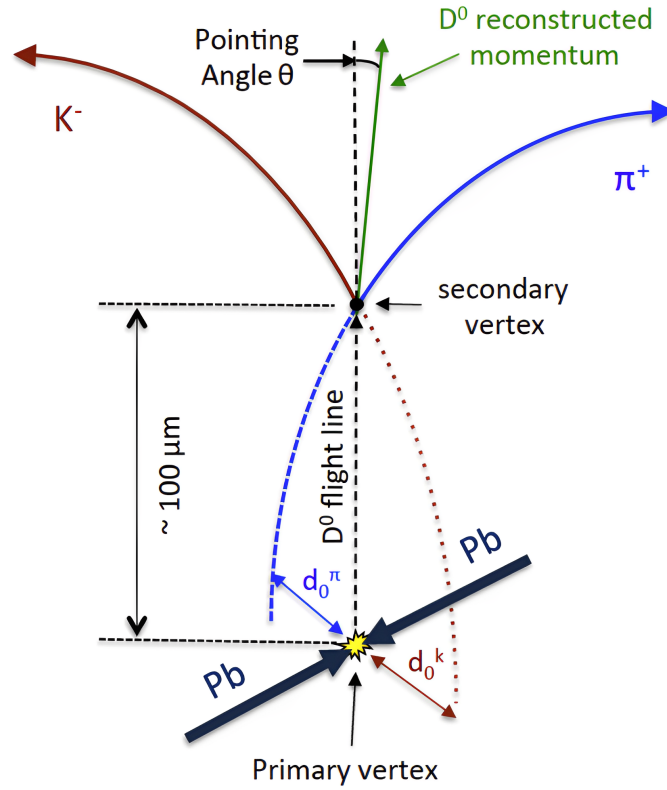


Figure 3.2: A sketch of the $D^0 \rightarrow K^- \pi^+$ decay, representing the primary, secondary vertices and some topological quantities

- **Decay length related:** Decay length (d_{len}) is constructed as the distance between the reconstructed primary and secondary vertex (three dimensional). A cut on the decay length separates the signal from background sources originating from the primary vertex. Addtinal, The resolution in the x-y plane (the plane perpendicular to the beam line) is known to be better than that in the z plane, and so projecting a variable in this plane can add discriminating power (d_{len_xy}). Finally, The decay length in the x-y plane normalised by its uncertainty is included as an input variable to



the BDT ($d_{len_xy}/\sigma_{d_{len_xy}}$). The distributions of the decay length related variables for three components are shown in the left column of Figs. 3.3, which extracted from the data (background) and MC production (prompt and Feed-down) in pp collisions at $\sqrt{s} = 13$ TeV in the $4 < p_T < 6$ GeV/c.

- **Impact parameter related:** The impact parameter (imp_par_xy) of the reconstructed heavy-flavour hadron in the transverse plane is defined as the distance of closest approach between the reconstructed flight line (or measured track in case one considers a decay daughter) and the primary vertex position, projected in the transverse plane. And also a important variable is the product of the impact parameters of their daughter tracks (imp_par_prod). Then an effective way to reject feed-down and combinatorial background candidates is adopted the maximum normalised difference between the measured and the expected impact parameters in the transverse plane $max_norm_d0d0exp$. The selection is applied for decay daughter i to the maximum value of $(d_{0,i}^{reco} - d_{0,i}^{exp})/\sigma_{\Delta}$. The distributions of the Impact parameter related variables for three components are shown in the right column of Figs. 3.3.
- **Cosine of the pointing angle related:** The pointing angle θ (cos_p) is defined as the angle between the momentum vector of the reconstructed particle, and the decay path from the primary vertex to the secondary vertex. To better exploit the x- and y-coordinate resolutions, a projection in the transverse plane (cos_p_xy) is used as well. The distributions of the Cosine of the pointing angle related variables for three components are shown in the right column of Figs. 3.4.

The particle-identification information is also provided to the machine-learning models used to perform the candidate selection. For each one of the D^0 -meson decay tracks, its $n\sigma$ difference with respect to the kaon and pion hypotheses, independently for the TPC and TOF detectors, is available. Therefore, a total of twelve variables related to the candidate PID can be employed. The combine the TPC and TOF information as follow as follows:

$$n\sigma_{comb}^{\pi,K} = \begin{cases} |n\sigma_{TPC}^{\pi,K}| & \text{tracks only in TPC} \\ |n\sigma_{TOF}^{\pi,K}| & \text{tracks only in TOF} \\ \frac{1}{\sqrt{2}} \sqrt{(n\sigma_{TPC}^{\pi,K})^2 + (n\sigma_{TOF}^{\pi,K})^2} & \text{tracks in TPC and TOF} \end{cases} \quad (3.1)$$

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 low the number of variables used 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, which contain the information on the compatibility of the track with the

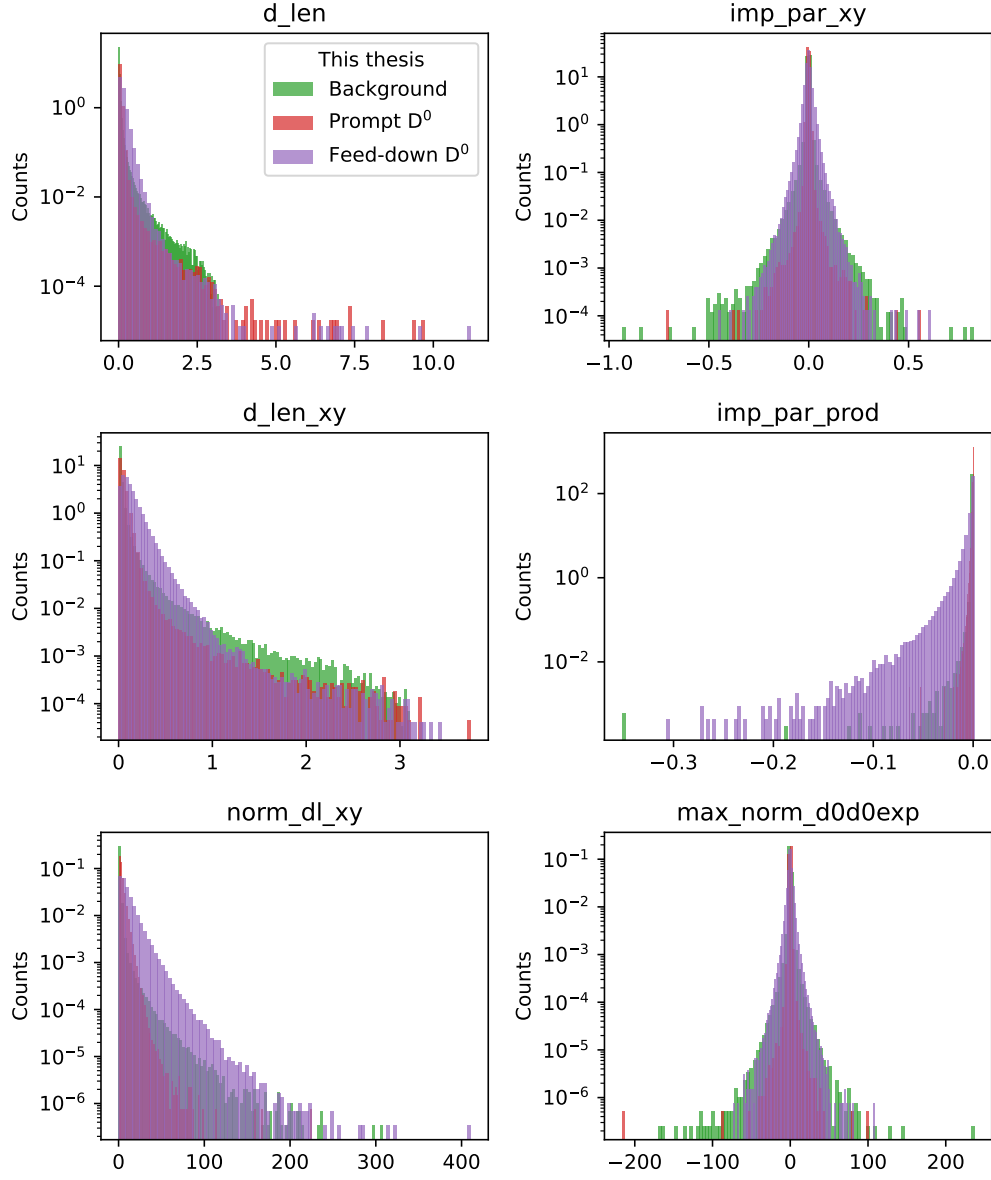


Figure 3.3: The $D^0 \rightarrow K^- \pi^+$ candidates distributions of the decay length related variables(left column) and impact parameter related variables (right column) for three components, which extracted from the data (background) and MC production (prompt and feed-down) in pp collisions at $\sqrt{s} = 13$ TeV in the $4 < p_T < 6$ GeV/ c .

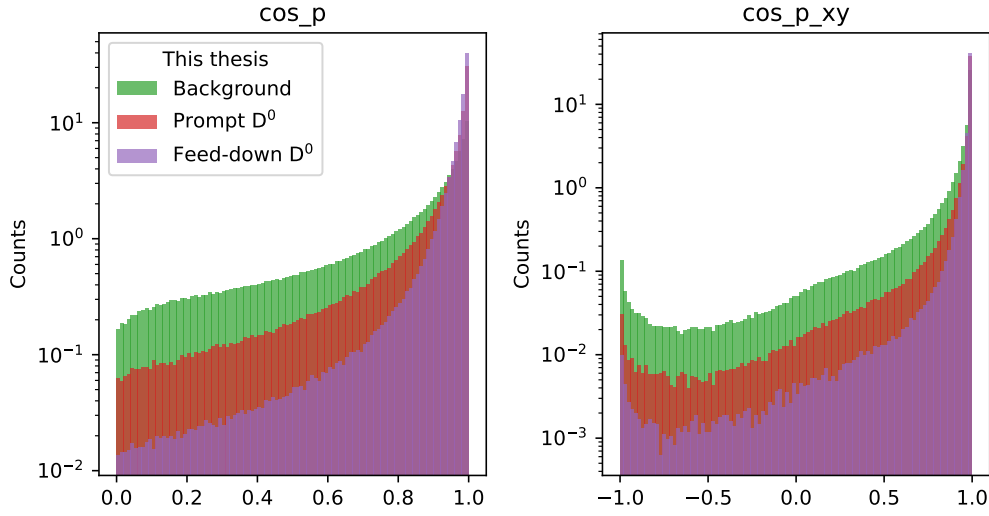


Figure 3.4: The $D^0 \rightarrow K^- \pi^+$ candidates distributions of the cosine point angle (left) and the projection in the transverse plane (right column) for three components, which extracted from the data (background) and MC production (prompt and feed-down) in pp collisions at $\sqrt{s} = 13$ TeV in the $4 < p_T < 6$ GeV/c.

pion and kaon hypotheses. The $D^0 \rightarrow K^- \pi^+$ candidates distributions of the $n\sigma_{\text{comb}}^\pi$ and $n\sigma_{\text{comb}}^K$ of each daughter tracks for three components are shown in Figs. 3.5.

In order to obtain a 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. In this thesis, the training and test sets are composed by three classes of candidates: prompt and non-prompt charm hadrons are taken from the MC sample, while combinatorial-background candidates are taken from real data. The MC productions, enhanced in charm hadrons, are used for the prompt and non-prompt charm-hadron components. The examples for the combinatorial background, instead, are extracted from the data sample of real candidates, selecting charm hadron candidates in the sidebands of the the invariant-mass distribution, away from the charm hadrons peaks. The idea behind this decision is to use real data, where feasible, to avoid a loss in the model predictive power caused by possible shortcomings of simulations in describing the real combinatorial background. The prompt and non-prompt charm-hadron candidates used for the training and test sets are all those available in the MC production. The number of background candidates for the low p_T region is chosen to be the same number as the signal, while at high p_T less background candidates are used to maintain small the fraction of real data employed in the training with respect to the total amount available.

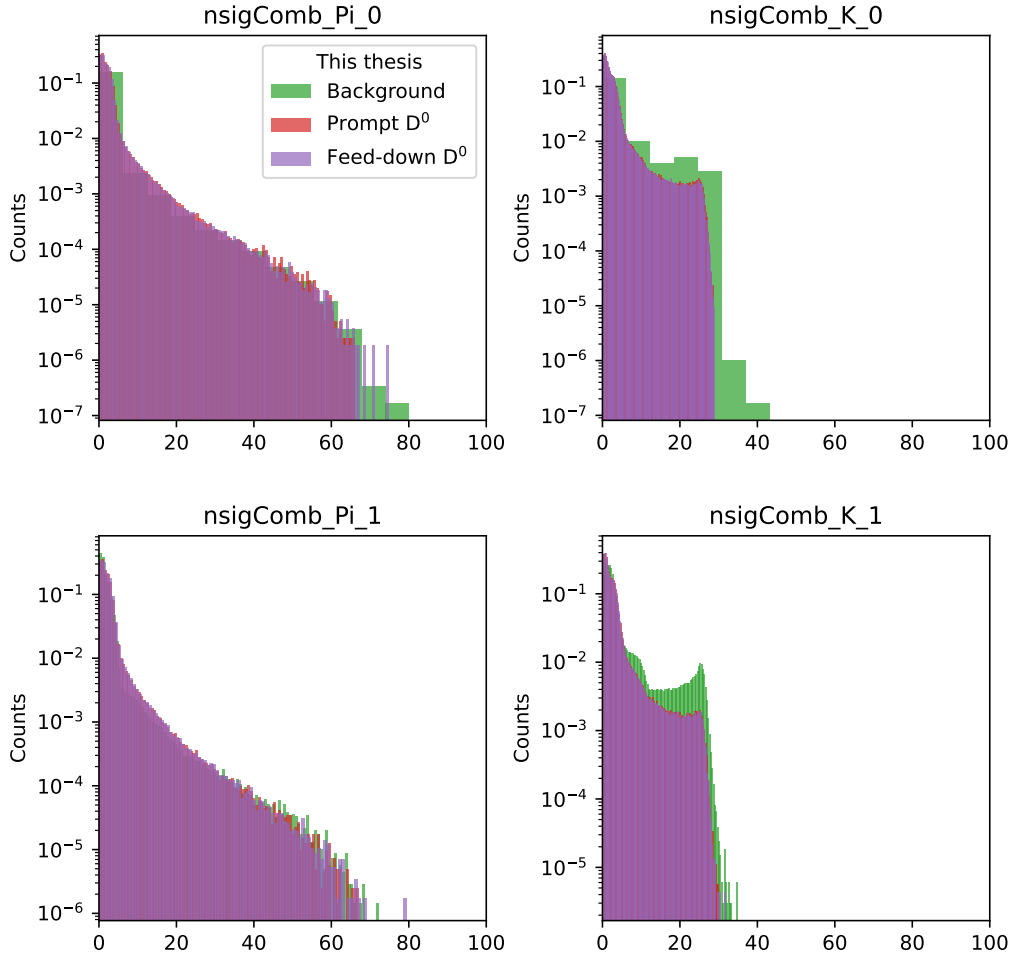


Figure 3.5: The $D^0 \rightarrow K^- \pi^+$ candidates distributions of the $n\sigma_{\text{comb}}^{\pi}$ (left column) and $n\sigma_{\text{comb}}^K$ of each daughter tracks for three components, which extracted from the data (background) and MC production (prompt and feed-down) in pp collisions at $\sqrt{s} = 13$ TeV in the $4 < p_T < 6$ GeV/c.



The hyper-parameter configuration used for the training, in the different p_T intervals was chosen with a Bayesian optimisation procedure [183] with a 5-fold cross-validation [184] implemented in the hipe4ml package [185]. It has been observed that this approach gives a good tuning of the model complexity without the need of an extensive and time consuming brute-force search over all the hyper-parameters. The complete list of model hyper-parameters, together with their default values and a description of their meaning can be found in the XGBoost documentation [181].

3.1.3 BDT application and signal extraction

After the training, the model is applied to the test and training sets to visualise the distributions of the ML output scores obtained and to check their agreement between the two sets. In particular, in case of classification with three classes of candidates, three scores are provided by the ML algorithm, which represent the probability of belonging to each of the considered classes (prompt D mesons, non-prompt D mesons, combinatorial background). By construction, for each candidate, the probabilities of belonging to the three classes sum to one. Fig. 3.6 show the example of probabilities be a non-prompt D^0 distributions in the $4 < p_T < 6$ GeV/ c intervals. For each probability, the distribution corresponding to the respective true class is close to unity, while the other two distributions are shifted towards zero. Moreover, it is possible to observe that the distributions of Feed-down D^0 candidates are typically more separated from those of prompt and non-prompt D mesons, indicating that the separation between Feed-down D^0 and orther classes is efficient.

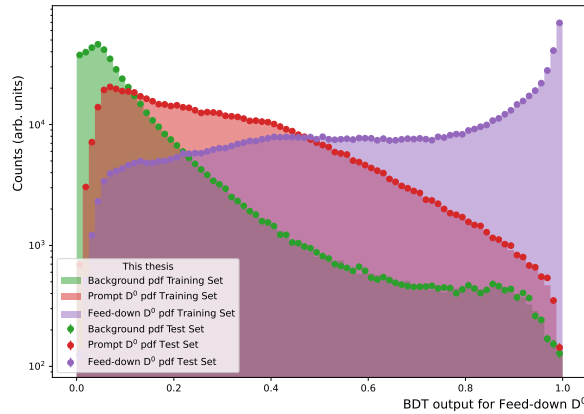


Figure 3.6: Distributions of ML output score of Feed-down for combinatorial background candidates, prompt D^0 mesons, and Feed-down D^0 in pp collisions at $\sqrt{s} = 13$ TeV in the $4 < p_T < 6$ GeV/ c .

The performance of a model can be evaluated using the Receiver Operating Characteristic (ROC) curve, which plots the True Positive Rate (signal selection efficiency) against the False Positive Rate

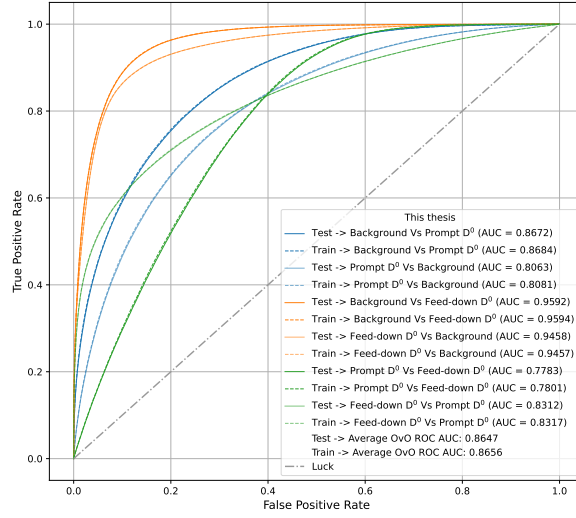


Figure 3.7: ROC curves for the models trained for D^0 in pp collisions at $\sqrt{s} = 13$ TeV in the $4 < p_T < 6$ GeV/c.

(background selection efficiency) for various threshold settings. The area under the ROC curve, known as ROC AUC, is used as a measure of the model's performance. The ROC AUC ranges between 0.5 and 1, where 0.5 indicates random classification and 1 indicates perfect classification. The ROC AUC provides a global estimation of the model's performance, independent of the chosen threshold and the class distribution in the dataset. However, for multi-class classification problems, the ROC curve cannot be directly defined. Instead, the "One-vs-One" approach is used, where a separate ROC curve is defined for each pair of classes, treating one class as signal and the other as background, and vice versa. To summarize the performance of a multi-class classification model, the macro average of the individual ROC AUCs is computed. The macro average is the average of the ROC AUCs for each individual ROC curve, which corresponds to a pair of classes. The equation for macro average is given as

$$\text{macro average} = \frac{1}{6} \sum_i^6 \text{ROCAUC}_i \quad (3.2)$$

where i runs over all the single ROC curves. The ROC curves and the values of ROC AUC, obtained on test set, for the models trained for D^0 meson in the $4 < p_T < 6$ GeV/c intervals is reported in Fig. 3.7.

The use of machine learning (ML) in candidate selection can be influenced by the response scores assigned to each candidate when applying BDT models to the data. An illustration of this is seen in Fig. 3.8, which displays examples of invariant-mass fits for beauty-hadron decays in proton-proton col-



lisions at $\sqrt{s} = 13\text{TeV}$ within the $2 < p_T < 4 \text{ GeV}/c$ range of minimum-bias triggered data. Selections based on the BDT output scores were used to identify samples dominated by non-prompt or prompt candidates. Low probabilities for a candidate to be combinatorial background and high (low) probabilities to be non-prompt were required to select samples dominated by non-prompt (prompt) candidates. The corresponding selection applied to the BDT output score related to the probability of a non-prompt D meson is shown in each right (left) panel, displaying the invariant-mass fits from non-prompt (prompt) enhanced samples.

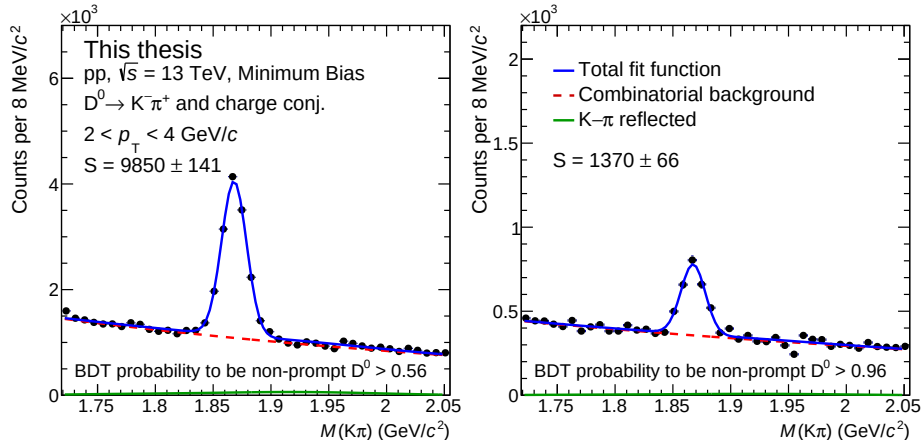


Figure 3.8: Invariant-mass distribution of D^0 candidates and their charge conjugates in selected in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ in the $2 < p_T < 4 \text{ GeV}/c$ interval for the minimum-bias triggered. The blue solid curves show the total fit function and the red dashed curves show the combinatorial-background contribution. The green solid lines represent the reflection contribution. The raw-yield (S) values are reported together with their statistical uncertainties resulting from the fit. The corresponding BDT probability minimum threshold for the candidate selection is reported. The left (right) column corresponds to the prompt (non-prompt) D^0 meson candidates dominated sample.

3.2 Data-driven based fraction estimation

The heavy hadrons reconstructed in this thesis originate not only from prompt heavy hadron production, but also from decays of beauty hadrons by the weakly decay $H_b \rightarrow H_c + X$. The final prompt and non-prompt heavy hadron production cross section is corrected by the fraction of the reconstructed heavy hadron yield, which namely f_{prompt} and $f_{\text{non-prompt}}$, respectively. There are two method used in the previous HF analysis in ALICE. One is pure theory-driven method which called 'fc'. The idea of fc method is entirely based on FONLL prediction to calculated cross sections for prompt and non-prompt production of heavy hadrons:

$$f_{\text{prompt}} = \left(1 + \frac{(\text{Acc} \times \epsilon)_{\text{feed-down}}}{(\text{Acc} \times \epsilon)_{\text{prompt}}} \right)^{-1} \frac{\left. \frac{d\sigma_{\text{FONLL}}^{D \text{ from } B}}{dp_T} \right|_{|y| < 0.5}}{\left. \frac{d\sigma_{\text{FONLL}}^D}{dp_T} \right|_{|y| < 0.5}} \quad (3.3)$$



The idea of another method is to combine the theory model prediction and raw yield measured ($N^{\text{D raw}}$) from the analysis, which namely ' N_b ' method. As show in Eq. 3.4, the feed-down raw yield $N^{\text{D feed-down}}$ was estimated from the production cross section of FONLL prediction where the fragmentation fraction of beauty quarks into charm hadrons was taken as the combined average from e^+e^- collisions. f_{prompt} was then calculated applying the correction factors for acceptance and reconstruction efficiency $\text{Acc} \times \epsilon)_{\text{feed-down}}$, the branching ratio BR of $H_b \rightarrow H_c + X$ and the total events N_{evt} :

$$\begin{aligned} f_{\text{prompt}} &= 1 - (N^{\text{D feed-down raw}} / N^{\text{D raw}}) = \\ &= 1 - \left(\frac{d^2\sigma}{dy dp_T} \right)_{\text{feed-down}}^{\text{FONLL+PYTHIA8}} \cdot \frac{(\text{Acc} \times \epsilon)_{\text{feed-down}} \cdot \Delta y \Delta p_T \cdot \text{BR} \cdot N_{\text{evt}}}{N^{\text{D raw}}/2}, \end{aligned} \quad (3.4)$$

The disadvantage of these two methods is that they can lead to bias if the theoretical calculation is not good agreement with experimental data. Thus in this section, a new data-driven will be introduced to estimate the f_{prompt} and $f_{\text{non-prompt}}$ in the analysis related to this thesis, which was already adopted for several measurements of non-prompt D mesons. The method is based on the definition of several set of selections. For each set of selections i it is possible to define an equation, which relates the extracted raw yield Y_i and the acceptance-times-efficiency factors of prompt D ($(\text{Acc} \times \epsilon)_i^{\text{prompt}}$) and non-prompt D ($(\text{Acc} \times \epsilon)_i^{\text{non-prompt}}$):

$$(\text{Acc} \times \epsilon)_i^{\text{prompt}} \cdot N_p + (\text{Acc} \times \epsilon)_i^{\text{non-prompt}} \cdot N_{\text{np}} = Y_i, \quad (3.5)$$

where N_p and N_{np} are the corrected yields of prompt and non-prompt D mesons, respectively, which we want to determine in order to compute their production cross sections. However, it is not possible to obtain the corrected yields from a single equation (i.e. a single set of cuts): at least two are needed. The basic idea of the cut-variation method is therefore to define n different sets of topological cuts in order to obtain a system of n equations

$$\left\{ \begin{array}{l} (\text{Acc} \times \epsilon)_1^{\text{prompt}} \cdot N_p + (\text{Acc} \times \epsilon)_1^{\text{non-prompt}} \cdot N_{\text{np}} = Y_1 \\ \dots \\ (\text{Acc} \times \epsilon)_n^{\text{prompt}} \cdot N_p + (\text{Acc} \times \epsilon)_n^{\text{non-prompt}} \cdot N_{\text{np}} = Y_n \end{array} \right. \quad (3.6)$$

and determine N_p and N_{np} . The system can be exactly solved in the case of two equations, otherwise it is likely to be overdetermined. Furthermore, since the extracted raw yields and the efficiencies are affected by statistical and systematic uncertainties, a number of sets of topological cuts larger than two



would lead to a better estimation of the corrected yields. For this reason, a minimisation procedure is needed in order to obtain an approximated solution of the system.

The data-driven method is used to get an approximated solution of the system in Eq. 3.6 is based on the minimisation of a χ^2 . In order to give a general description of this method, the system of Eq. 3.6, can be rewritten in a more compact way with a the matrix formalism for n generic sets of topological cuts, corresponding to n equations involving the measured raw yields and the efficiencies from simulations:

$$\epsilon \times N - Y = \mathbf{0}, \quad (3.7)$$

where $\mathbf{0}$ is the n -dimensional null vector. As already mentioned, for $n > 2$ the system is likely to be overdetermined, and therefore an additional quantity δ is needed. In particular, δ is the residual vector, which coincides with the null vector $\mathbf{0}$ in case of a determined system, while is different otherwise. Therefore, the latter equation becomes

$$\begin{pmatrix} (\text{Acc} \times \epsilon)_1^{\text{prompt}} & (\text{Acc} \times \epsilon)_1^{\text{non-prompt}} \\ \vdots & \vdots \\ (\text{Acc} \times \epsilon)_n^{\text{prompt}} & (\text{Acc} \times \epsilon)_n^{\text{non-prompt}} \end{pmatrix} \times \begin{pmatrix} N_p \\ N_{np} \end{pmatrix} - \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \delta_2 \\ \vdots \\ \delta_n \end{pmatrix}. \quad (3.8)$$

The best estimation of the corrected yields is obtained by minimising the χ^2 , defined as

$$\chi^2 = \delta^T C^{-1} \delta, \quad (3.9)$$

where δ^T is the row vector of residuals and C the covariance matrix

$$C = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \dots & & \\ \sigma_{2,1} & \sigma_2^2 & & & \\ \vdots & & \ddots & & \vdots \\ & & & \sigma_{n-1}^2 & \sigma_{n-1,n} \\ \dots & \sigma_{n,n-1} & & \sigma_n^2 & \end{pmatrix}, \quad (3.10)$$

which is defined from the uncertainties inherent to each equation. The uncertainties σ_i^2 are calculated starting from the statistical errors on the raw yields and the efficiencies, according to the following relation:

$$\sigma_i^2 = \sigma_{Y_i}^2 + N_p \cdot \sigma_{(\text{Acc} \times \epsilon)_i^{\text{prompt}}}^2 + N_{np} \cdot \sigma_{(\text{Acc} \times \epsilon)_i^{\text{non-prompt}}}^2. \quad (3.11)$$

Since the corrected yields are not yet determined at this stage, the weights are evaluated iteratively: in the first step N_p and N_{np} are set to zero and therefore only the uncertainty on the raw yields is taken



into account, while from the second iteration the weights are computed using the value of the corrected yields obtained in the previous step. The covariances $\sigma_{i,j}$ are instead computed considering the sets as fully correlated, and therefore with a correlation coefficient:

$$\rho_{i,j} = \frac{\sigma_i}{\sigma_j}, \text{ with } i \subset j. \quad (3.12)$$

This is motivated by the choice of the cut sets, as will be described in the following. Only few iterations are needed to obtain a stable result. Then, the minimisation of the χ^2 with respect to the corrected yields leads to

$$\mathbf{N} = \mathbf{Cov}(\mathbf{N})\boldsymbol{\epsilon}^T\mathbf{C}^{-1}\mathbf{Y}, \quad (3.13)$$

where $\mathbf{Cov}(\mathbf{N}) = (\boldsymbol{\epsilon}^T\mathbf{C}^{-1}\boldsymbol{\epsilon})^{-1}$ is the covariance matrix which contains the statistical uncertainties on the corrected yields and their correlation. Finally, from N_p and N_{np} is possible to evaluate the corrected fraction of non-prompt D mesons:

$$f_{np} = \frac{N_{np}}{N_{np} + N_p}. \quad (3.14)$$

In addition, it is possible to evaluate a fraction of non-prompt D mesons for any set of selections i :

$$f_{np}^i = \frac{(\text{Acc} \times \epsilon)_i^{\text{non-prompt}} \cdot N_{np}}{(\text{Acc} \times \epsilon)_i^{\text{non-prompt}} \cdot N_{np} + (\text{Acc} \times \epsilon)_i^{\text{prompt}} \cdot N_p}. \quad (3.15)$$

Eq. 3.14 is used for the measurement of the fraction of non-prompt D mesons as a function of multiplicity, while Eq. 3.15 for the measurement of the non-prompt D^0 cross section in the MB sample. In fact, with Eq. 3.14 we obtain a cancellation of the systematic uncertainties which are in common between prompt and non-prompt D mesons, while using Eq. 3.14 is convenient for the cross section measurement, as done in previous analyses, such as Ref. [?] and [186].

An example of the minimization procedure is shown in Fig.3.9 provides and its outcomes for the $2 < p_T < 4$ GeV/ c transverse-momentum region. The top left panel shows the acceptance-times-efficiency factors for prompt and non-prompt D mesons as a function of the BDT-based selection. The loosest selection on the BDT output related to the candidate probability of being a non-prompt D meson is at the leftmost data point, while the strictest selection, which is expected to select non-prompt D mesons preferentially, is at the rightmost. In the top right panel, the correlation coefficients ρ , defined in Eq. 3.12, between the samples obtained with different selections are displayed, with values ranging from 0.5 to 1. The raw-yield distribution as a function of the BDT-based selection is shown in the bottom left panel of Fig. 3.9, compared to the sum of the prompt and non-prompt components obtained from the minimization procedure, represented by the stacked red and blue histograms, respectively. The prompt

and non-prompt components are obtained as $(\text{Acc} \times \epsilon)_i^{\text{prompt}} \cdot N_p$ and $(\text{Acc} \times \epsilon)_i^{\text{non-prompt}} \cdot N_{np}$, where i represents prompt or non-prompt. Finally, the estimated fraction of prompt and non-prompt D mesons for the different selections is reported in the bottom right panel. As expected, stricter selections result in a higher fraction of non-prompt candidates in the extracted raw yields.

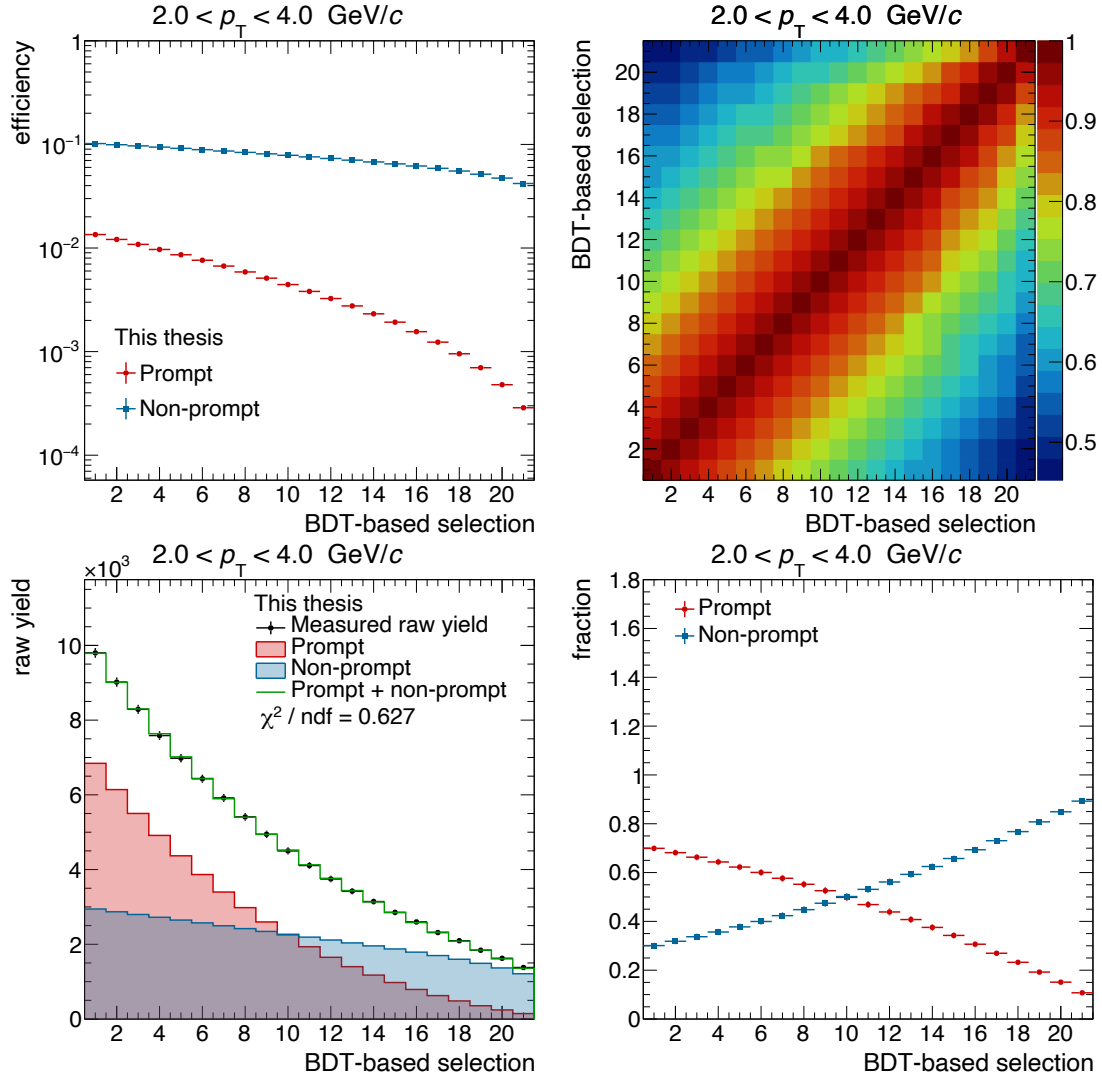


Figure 3.9: $((\text{Acc} \times \epsilon))$ factors as a function of the cut set (top-left panel), correlation matrix (top-right panel), raw-yield as a function of the cut set with the prompt and non-prompt contributions obtained from the minimisation procedure (bottom-left panel), and the resulting prompt and non-prompt fractions as a function of the cut set (bottom-right panel) for the D^0 mesons in the $2 < p_T < 4 \text{ GeV}/c$ interval for the MB class of events.



4 Open charm-hadron measurements in pp collisions at $\sqrt{s} = 13$ TeV

4.1 Prompt Λ_c^+ baryon production in pp collisions at $\sqrt{s} = 13$ TeV

Theoretical predictions based on pQCD calculations with the collinear factorisation approach at next-to-leading order, such as the general-mass variable-flavour-number scheme (GM-VFNS) or at fixed order with next-to-leading-log resummation (FONLL [48, 49]), describe the inclusive production cross sections of charm mesons measured in pp collisions at the LHC at both central and forward rapidity, within theoretical uncertainties [187, 188]. These predictions rely on the factorisation theorem, which states that the p_T -differential cross sections are obtained by convolving three factors: the parton distribution functions (PDFs) of the incoming (anti)protons, the partonic cross section computed as a perturbative series in the strong coupling constant α_s , and the fragmentation functions that describe the transition from charm quarks to charm hadrons [156]. The fragmentation functions used in these calculations are typically parametrised from measurements in e^+e^- or ep collisions. However, These predictions are not available for charm baryons, due to the lack of knowledge of the fragmentation function of charm quarks into baryons. Fig. 1.17 shown the charm-quark fragmentation fractions into charm hadrons measured in pp collisions at $\sqrt{s} = 5.02$ TeV in comparison with experimental measurements performed in e^+e^- collisions at LEP and at B factories, and in ep collisions at HERA. For charm baryon Λ_c^+ and $\Xi_c^{+,0}$ fragmentation fractions in pp collisions are significantly different than the measurement in e^+e^- or ep collisions [53]. Possible explanations are non-universality of the fragmentation functions, need of additional colour reconnection models, unobserved charm-baryon states, or hot nuclear matter produced in pp and p – Pb collisions.

In small collision systems, several differential studies can be performed, such as studying the evolution of identified hadrons with multiplicity. One such study is the measurement of heavy-flavour production in pp collisions as a function of the charged-particle multiplicity produced in the collision, which could provide insight into the processes occurring in the collision at the partonic level and the interplay between the hard and soft mechanisms in particle production. The NA27 Collaboration observed in 1988 that the average charged-particle multiplicity in events with open charm production was higher by about 20% than in events without charm production [189]. A softening of the momentum spectra of hadrons produced in association with charm was also observed. This result was interpreted as a consequence of the more central nature of collisions leading to charm production.



At LHC energies, two additional contributions to charm production and its relation to multiplicity have to be considered. The first effect is the likely larger amount of gluon radiation associated with short distance production processes at larger energies and particle transverse momenta. The second is the contribution of Multiple-Parton Interactions (MPI) [190–192], i.e. several hard partonic interactions occurring in a single pp collision.

Recently, the study of heavy-flavour production as a function of the multiplicity of charged particles produced in the collision has attracted growing interest. The measurements of prompt D mesons, inclusive and non-prompt J/ψ in pp collisions at $\sqrt{s} = 7$ TeV [193, 194], and of the three Υ states in pp collisions at $\sqrt{s} = 2.76$ TeV [195], provide evidence for a similar increase of open and hidden heavy-flavour yields as a function of charged-particle multiplicity. These results suggest that the enhancement probably originates in short-distance production processes and is not influenced by hadronisation mechanisms. The enhancement was found to be stronger-than-linear and quantitatively described by calculations including MPI contributions. However, as argued in Refs. [54, 196], this non-linear increase is due to the fact that the charged-particle multiplicity and the heavy-flavour yield are not independent quantities; the latter influences the former. It was shown in Ref. [54] that colour reconnection cannot be responsible for a stronger-than-linear increase. To disentangle the auto-correlation effects and true correlation effects between the hard probe and the underlying event, the multiplicity should be measured in the transverse region with respect to the HF candidate. For the J/ψ measurement, this was recently done by the ALICE Collaboration; see Ref. [197] for a description of this preliminary measurement.

Also in the light-flavour sector, there is interest in the evolution of identified hadrons with multiplicity. The ALICE and CMS Collaborations observed an enhancement of baryon-to-meson ratios (p/π) at intermediate p_T in high multiplicity pp and p–Pb collisions, similar to what is observed in heavy-ion collisions [198, 199]. The ALICE Collaboration recently published several papers related to strangeness production, considering primary hadrons such as K_S^0 and Λ , as well as multi-strange hadrons such as Ξ , Ω , and ϕ , versus multiplicity [69, 200, 201] and sphericity [202]. The mass and multiplicity dependences are reminiscent of the patterns seen in p – Pb and Pb – Pb collisions at the LHC. Furthermore, it has been demonstrated that the evolution of the baryon-to-meson ratios as a function of $\langle dN_{ch}/d\eta \rangle$ exhibits a universal pattern for all collision systems. This behavior might indicate a common mechanism at work that depends solely on final-state multiplicity density. Another observation is that the observed enhancement in the particle ratios increases with strangeness content rather than with mass or baryon number of the hadron. The evolution of this strangeness enhancement steadily increases from low multiplicity pp to high multiplicity p – Pb and reaches the values observed in Pb – Pb collisions.



This may point towards a common underlying physics mechanism that gradually compensates for the strangeness suppression in fragmentation. As discussed before, the "fragmentation bias" [54, 203] is taken into account in these measurements by measuring the particle's p_T and event multiplicity in different pseudorapidity intervals.

In this section, I presented the first measurement of the production yield of prompt Λ_c^+ and the corresponding ratio Λ_c^+/D^0 as a function of multiplicity in pp collisions at $\sqrt{s} = 13$ TeV. The study was conducted by selecting events based on the charged-particle density at mid- and forward rapidities to account for possible biases in the determination of multiplicity. By examining the dependence of production yield on charged-particle multiplicity, we aim to gain insights into the hadronization mechanism of charm baryons.

4.1.1 Event selection and multiplicity definition

A standard physics selection was applied to reject background events stemming from the interaction of beam particles with the beam pipe materials or beam-gas interactions. Only events with interaction vertex within the range $|z| < 10$ cm were used for the analysis. In case where the primary vertex could be reconstructed with both SPD and tracks, the two z-positions were requested to be compatible. A dedicated algorithm to detect multiple interaction vertices based on the tracklets (i.e. pairs of reconstructed hits in the SPD) was used to further reduce the pile-up contribution. An event is rejected from the analysed data sample if a second interaction vertex is found, it has at least 5 associated tracklets. Further selections are applied on the χ^2 of the SPD pile-up vertex and on the weighted distance between the primary and the pile up vertex. When the two vertices are found to be too close in z-direction, no pile up rejection is applied since it becomes very difficult to distinguish between a real pile-up event and a fake one. Furthermore, events were selected based on the presence of at least one charged particle within the pseudorapidity range $|\eta| < 1$, known as the INEL>0 class. This class of events reduces diffractive corrections, enabling a more accurate comparison with theoretical calculations.

In addition the so-called offline trigger decision is used to select events offline that triggered the corresponding decision and which fall at least in one of the classes MB, HMSPD or HVMZERO. Where MB (minimum-bias) required signals in both V0A and V0C in coincidence. HMSPD is a new trigger strategy, only adopted in 2018 pp data taking period, was developed to increase the number of events with large number of SPD tracklets, and more generally with large multiplicity inside the central barrel. The trigger algorithm was based on the number of outer fast-OR chips fired online. The latter quantity is strongly correlated with the offline SPD tracklet multiplicity. HVMZERO is based on the energy deposited in the two V0 scintillators is proportional to the number of primary charged particles produced



in the pseudorapidity ranges covered by the detectors, namely $3.7 < \eta < 1.7$ and $2.8 < \eta < 5.1$. The high multiplicity V0 triggers are activated when the V0 amplitude reaches a threshold value of approximately ≥ 2700 arbitrary units.

The analyses split the events in different multiplicity intervals. There are two multiplicity determination methods based on the charged-particle density at mid- and forward rapidity. At midrapidity ($|\eta| < 1$), the event multiplicity was estimated using the number of tracklets (N_{trkl}), which are defined as track segments constructed by associating pairs of hits in the two SPD layers. On the other hand, the event multiplicity in the forward rapidity region was estimated using the percentile distribution (p_{VOM}) of the V0M amplitude, which is the sum of signal amplitudes in the V0A and V0C scintillators. Fig. 4.1 shows the normalised distributions for the two multiplicity estimators. This analysis data was collected in the years 2016, 2017, and 2018 as shown in Table 2.3. Then three or four independent multiplicity intervals was considered to be used to extract the raw yield signals. The corresponding observables N_{trkl} and p_{VOM} for the multiplicity intervals are presented in Table 4.1.

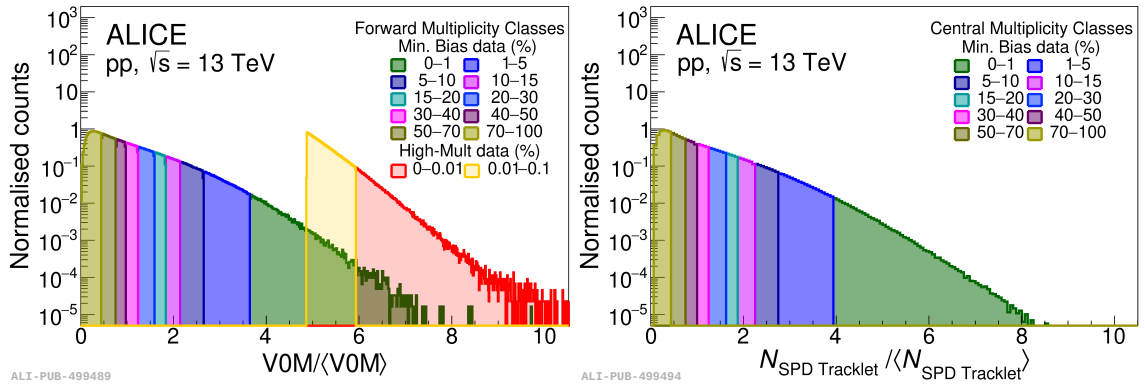


Figure 4.1: The distributions of the total number of SPD tracklets and the V0M amplitude in pp collisions at $\sqrt{s} = 13$ TeV. The right panel displays the distribution of the total number of SPD tracklets within $|\eta| < 2$ scaled by its average value, while the left panel shows the distribution of the V0M amplitude scaled by its average value. The figures are taken from Ref. [204]

4.1.2 Analysis strategy

The final observable for analysis is the corrected yield per event in each multiplicity bin for heavy-flavour hadrons. After selecting candidates in all multiplicity and p_T bins, a fit is performed on their invariant mass distributions to estimate the *raw* signal and background yields. The corrected per-event yields were computed for each p_T and multiplicity interval as

$$\frac{1}{N_{\text{mult}}^{\text{ev}}} \frac{d^2 N_{\text{mult}}^{\text{hadron}}}{dy dp_T} = \frac{\epsilon_{\text{mult}}^{\text{INEL}}}{N_{\text{mult}}^{\text{ev}}} \frac{1}{c_{\Delta y}(p_T) \times \Delta p_T} \frac{1}{\text{BR}} \frac{f_{\text{prompt}}(p_T) \times \frac{1}{2} \times N_{\text{mult}}^{\text{hadron, raw}}(p_T) \Big|_{|y| < y_{\text{fid}}(p_T)}}{(\text{Acc} \times \epsilon)_{\text{prompt, mult}}(p_T)}, \quad (4.1)$$



Table 4.1: The summary of the multiplicity event classes at midrapidity (Ntrkl) and forward rapidity (%V0M), which corresponds to the visible V0M cross section, is shown below. The average charged-particle densities $\langle dN_{\text{ch}}/d\eta \rangle$ at midrapidity are presented, along with the value corresponding to the INEL > 0 event class. Additionally, the trigger efficiency ϵ^{INEL} is reported for each multiplicity interval, as estimated in Ref. [204].

Mult. estimator	Mult. interval	$\langle dN_{\text{ch}}/d\eta \rangle_{ \eta < 0.5}$	ϵ^{INEL}
N_{trkl}	[1, 9]	3.10 ± 0.02	0.862 ± 0.015
	[10, 29]	10.54 ± 0.01	0.997 ± 0.002
	[30, 59]	22.56 ± 0.07	1 (negl. unc.)
	[60, 99]	37.83 ± 0.06	1 (negl. unc.)
$p_{\text{V0M}} [\%]$	[30, 100]	4.41 ± 0.05	0.897 ± 0.013
	[0.1, 30]	13.81 ± 0.14	0.997 ± 0.001
	[0, 0.1]	31.53 ± 0.38	1 (negl. unc.)
INEL > 0		6.93 ± 0.09	0.920 ± 0.003

where $N_{\text{mult}}^{\text{hadron,raw}}$ is the raw yield (sum of particles and antiparticles) extracted in a given p_T and multiplicity interval. It is multiplied by the prompt fraction f_{prompt} in order to correct for the corresponding beauty-hadron decay contribution, and divided by the multiplicity-dependent prompt acceptance-times-efficiency, $(\text{Acc} \times \epsilon)_{\text{prompt,mult}}$. It is further divided by a factor of two to obtain the charge-averaged yield, by the BR of the decay channel, the p_T -interval width (Δp_T), and the correction factor for the rapidity coverage $c_{\Delta y}$, computed as the ratio between the generated heavy-flavour hadron yield in $\Delta y = 2y_{\text{fid}}$ and that in $|y| < 0.5$. The factor $N_{\text{mult}}^{\text{ev}}$ denotes the number of recorded events in each multiplicity class, which is then corrected for the fraction of INEL > 0 events that were not selected by the trigger, $\epsilon_{\text{mult}}^{\text{INEL}}$, whose values are reported in Table 4.1. The ingredients of the Eq. 4.1 will be discussed in the below sections.

4.1.3 Candidate selection and raw yield extraction

The Λ_c^+ hadrons and their anti-particles were reconstructed in the central rapidity region using their charged hadronic decay channels $\Lambda_c^+ \rightarrow pK^-\pi^+$ (BR = $(6.28 \pm 0.32)\%$) and $\Lambda_c^+ \rightarrow pK_S^0$ (BR = $(1.10 \pm 0.06)\%$). Standard selection criteria were applied to the tracks (or V0 candidates) used to build the Λ_c^+ candidates. The criteria included ITS and TPC refit (only TPC refit was required for the K_S^0 daughters), at least 70 (out of a maximum of 159) associated crossed rows in the TPC, a ratio of crossed rows over findable clusters in the TPC larger than 0.8, $\chi^2/\text{ndf} < 2$ in the TPC, and at least two (out of six) hits in the ITS, out of which at least one had to be in either of the two SPD layers (this was only required for the proton in the $\Lambda_c^+ \rightarrow pK_S^0$ channel). The tracks were also required to be in the kinematic range $|\eta| < 0.8$ and $p_T > 0.3 \text{ GeV}/c$ (the K_S^0 prongs were selected in the $|\eta| < 1.1$ and $p_T > 0.1 \text{ GeV}/c$ range). The track selection criteria reduced the acceptance for the Λ_c^+ mesons, which dropped steeply to zero for $|y_{\text{lab}}| > 0.5$ at low p_T and for $|y_{\text{lab}}| > 0.8$ at $p_T > 5 \text{ GeV}/c$. Therefore, a p_T -dependent fiducial acceptance region was defined as $|y_{\text{lab}}| < y_{\text{fid}}(p_T)$, with $y_{\text{fid}}(p_T)$ increasing from 0.5 to 0.8 in the transverse momentum



range $0 < p_T < 5 \text{ GeV}/c$ according to a second-order polynomial function, and $y_{\text{fid}} = 0.8$ for $p_T > 5 \text{ GeV}/c$. For the $\Lambda_c^+ \rightarrow pK_S^0$ channel, where the secondary vertex cannot be reconstructed due to the poor resolution on the resonant state, the selection strategy was based only on kinematical and particle identification (PID) criteria on the daughter tracks. The PID selection strategy used the information on specific energy deposit in the Time Projection Chamber (TPC) and on the time of flight measured by the Time of Flight (TOF) detector. The TOF detector allowed for a reliable identification of kaons and pions in a wide momentum range up to $2 \text{ GeV}/c$.

The Λ_c^+ candidates were selected in the decay chain $\Lambda_c^+ \rightarrow pK_S^0$ with the loose rectangular cut based selection followed by a machine learning (ML) based approach using Boosted Decision Trees (BDTs). In particular the BDT implemented by the XGBoost library [181], which provides the possibility to perform a multi-class classification, is employed. This library is an open-source package and its algorithms achieved state-of-the-art results on many ML challenges, such as those hosted on Kaggle [205]. The machine learning models be trained with the filter data and MC based on the preselection shown in Table 4.2. There are 6 p_T intervals were considered to extract the raw yield of Λ_c^+ baryon in this analysis (Table 4.3). Topological and kinematic variables are used in the model training together with the PID information of each candidate daughter. The variables adopted are shown in Table 4.4 and their distributions for signal and background are shown in Figs. 4.2. The linear correlations between the variables used in the training as well as the invariant mass of Λ_c^+ candidate and its transverse momentum, which are not employed in the training, are reported for the three classes of candidates in Figs. 4.3. To obtain optimised values for the BDT's hyperparameters, a grid search was performed using the Area-Under-Curve (AUC) to judge on the performance of the different configurations^①. The final hyperparameter values are reported in 4.5. The Receiver-Operating-Curves (ROC) for a low and a high p_T bin as an example shown in Fig 4.4.

After applying the machine learning models to the candidates, a binary decision-based BDT selection is performed, resulting in probability values within the range $[0, 1]$. A candidate is considered signal if its assigned probability exceeds a chosen working point (WP) and is classified as background if the assigned probability value falls below the WP. To account for the changing topology and kinematics across different p_T bins, a dedicated model is constructed for each analysis bin using the multiplicity-integrated MB MC and data. It is assumed that the topology and kinematics do not vary significantly with multiplicity. Due to insufficient statistics when splitting the training samples into multiplicity bins, the model is constructed using the multiplicity-integrated samples.

① A grid search scans regions of different hyperparameters and comparing the scores. The nominal classifier is fixed to the one which performs best during the grid search.

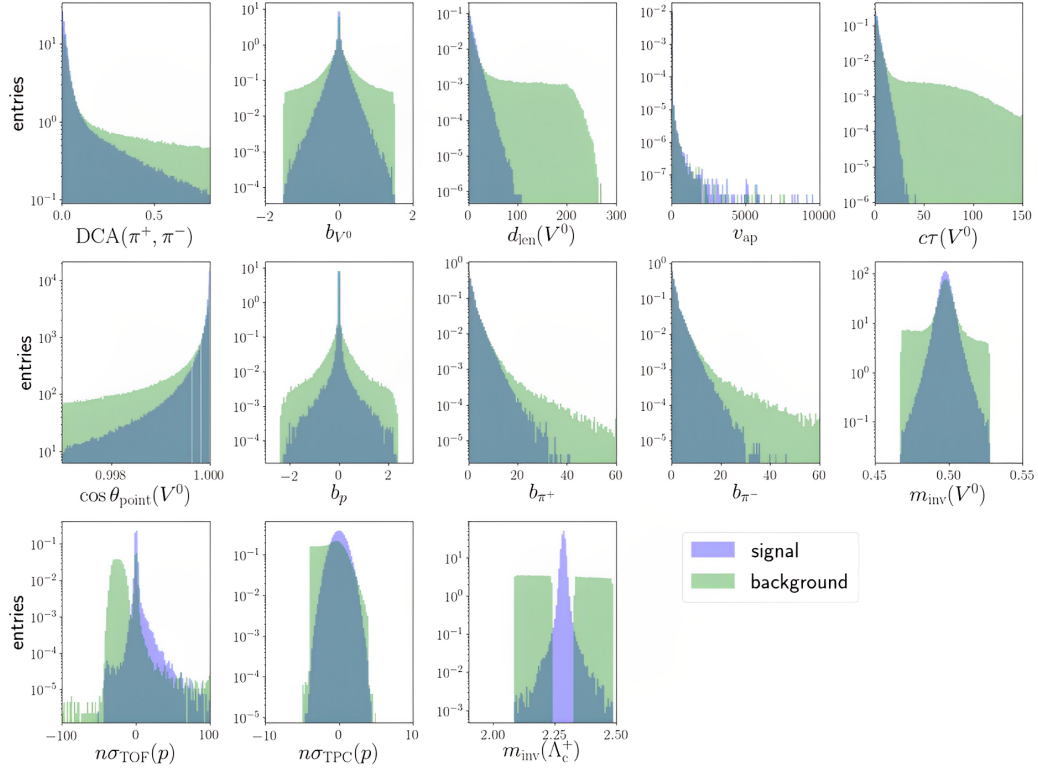


Figure 4.2: Distributions for signal and background of Λ_c^+ candidate variables in the $1 < p_T < 2$ GeV/c interval.

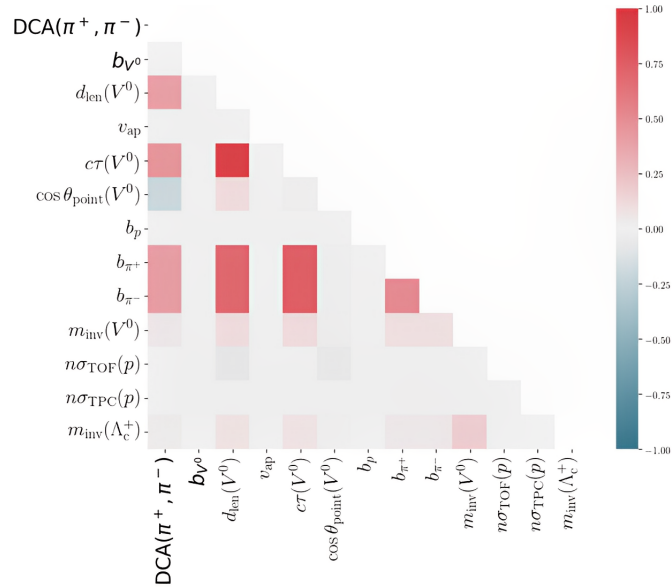


Figure 4.3: The plot shows the correlations between the selected features and the reconstructed invariant mass of the Λ_c^+ in the momentum range of 1 GeV/c to 2 GeV/c, using signal samples.



Table 4.2: Pre-selection for Λ_c^+ candidates

variables	selection
$\Delta M_{\Lambda_c^+}$	$< 0.2 \text{ GeV}/c^2$
$\Delta M_{K_S^0}$	$< 0.03 \text{ GeV}/c^2$
DCA(V^0)	$< 10 \text{ cm}$
$n\sigma(\text{DCA}(V^0))$	$< 0.8 \text{ cm}$
impact parameter(V^0)	$< 1.5 \text{ cm}$
impact parameter(proton)	$< 3 \text{ GeV}/c^2$
$\cos\theta_p$	> 0.997
$N\sigma(\text{TPC})$	< 4
$N\sigma(\text{TOF})$	< 4

Table 4.3: p_T intervals for Λ_c^+

$p_T [\text{GeV}/c]$	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
----------------------	--------	--------	--------	--------	---------	----------

To maximize the expected significance $S/\sqrt{S+B}$, the optimal WP is chosen based on FONLL predictions for the $\Lambda_c^+ \rightarrow pK_S^0$ baryon, which are the same as those used for the feed-down subtraction. The expected signal is estimated using the formula:

$$S_{\text{exp}} = 2 \left(\frac{d\sigma}{dp_T} \right) \text{prompt}^{\text{FONLL}} \cdot (\text{Acc} \times \varepsilon)_{\text{prompt}} \cdot c_{\Delta y} \cdot \Delta p_T \text{BR} \cdot \mathcal{L}_{\text{int}} / (f_{\text{prompt}}), \quad (4.2)$$

where the factor of 2 accounts for the particle-antiparticle average, Δp_T is the width of the p_T interval, $c_{\Delta y}$ is the correction factor for the rapidity coverage, BR is the branching ratio, and $\mathcal{L}_{\text{int}} = \sigma_{\text{MB}}/N_{\text{ev}}$ is the integrated luminosity. The f_{prompt} fraction is assumed to be constant (0.9) with p_T . The expected background under the signal peak is estimated by fitting the sidebands of the invariant-mass distribution using a fraction of the data, which is then scaled to match the background estimate in the full data sample.

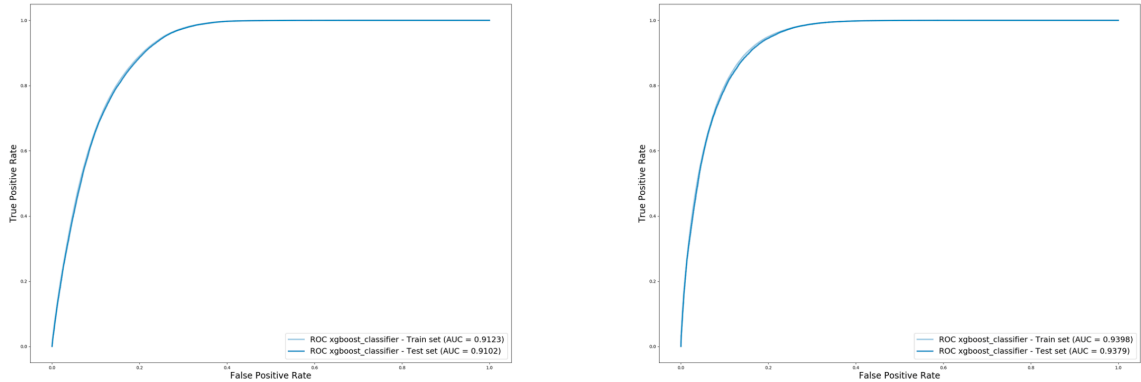


Figure 4.4: ROC curves for Λ_c^+ candidates in the p_T intervals $2 < p_T < 4 \text{ GeV}/c$ (left) and $8 < p_T < 12 \text{ GeV}/c$ (right).



Table 4.4: Training variables used for the Λ_c^+ baryons.

Variable
Decay length (V^0)
$\cos \theta^*$
DCA(V^0)
impact parameter(V^0)
impact parameter(daughter 1)
impact parameter(daughter 2)
impact parameter(proton)
$c\tau(V^0)$
$\cos \theta_p(V^0)$
Inv.Mass(V^0)
$n\sigma_{\text{TPC}}(\text{proton})$
$n\sigma_{\text{TOF}}(\text{proton})$

Table 4.5: The (p_T -independent) XGBoost BDT hyperparameters used for candidate selection in $\Lambda_c^+ \rightarrow pK_s^0$.

XGBoost hyperparameters	
min-child-weight	3
max-depth	3
gamma	0
subsample	0.8
colsample-bytree	0.8
learning-rate	0.1
n-estimators	850

The optimal probability cut for each analysis was chosen to balance high significance and selection efficiency. Low efficiencies were avoided as they could lead to large systematics resulting from significant differences in the distribution of cut variables between data and MC. The results of the scans for the low and high p_T intervals are shown in Fig. 4.5, and the optimal working point of 0.7 was chosen for all p_T bins.

After pre-filtering the candidates, their invariant mass distributions are modelled using a single Gaussian function for the signal and a 2nd order polynomial function for the background across all multiplicity and p_T bins. The comparison of the widths obtained in the integrated multiplicity bin for data and simulation is shown in Fig. 4.6 on the left and their ratio on the right. During the fit, Gaussian means and widths are left free. However, to avoid large fluctuations of the width across multiplicity bins and since the fit in the highest p_T bin does not converge in data, all widths are fixed to the values per p_T bin obtained in MC for the integrated multiplicity. Finally, Fig. 4.7 and Fig. 4.8 shown the mass spectra for the INEL>0 class minimum bias triggered sample and high V0M triggered sample, respectively.

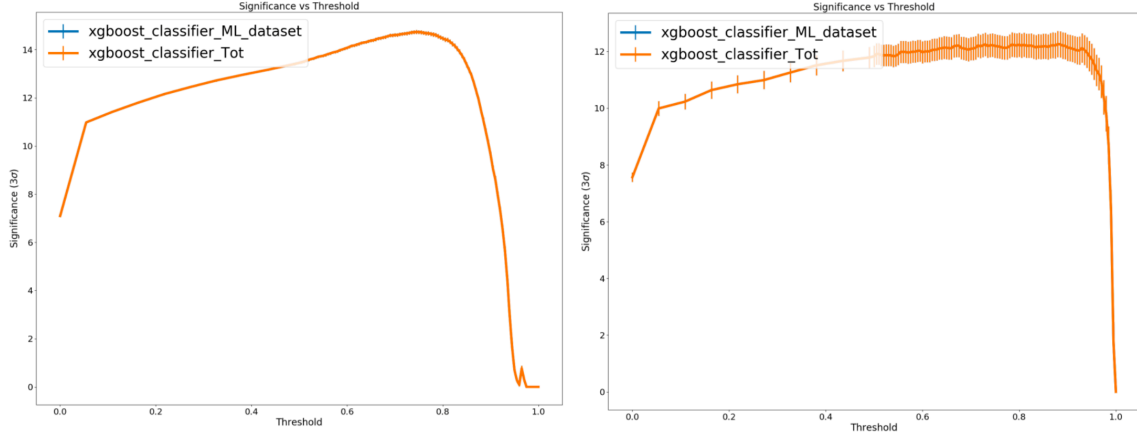


Figure 4.5: Distributions for signal and background of Λ_c^+ candidate variables in the $1 < p_T < 2$ GeV/c interval.

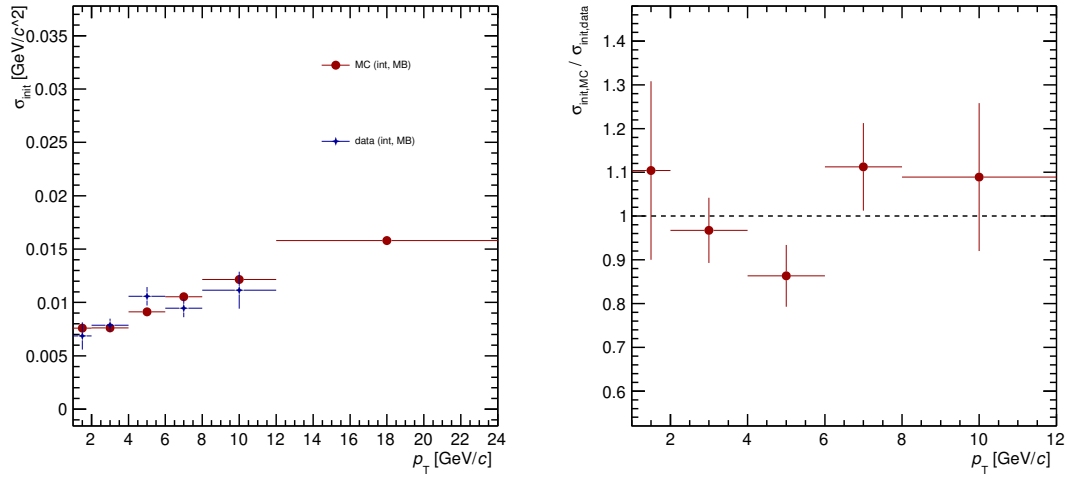


Figure 4.6: Widths obtained in V0M multiplicity integrated bin for all p_T bins (left) and their ratio (right).

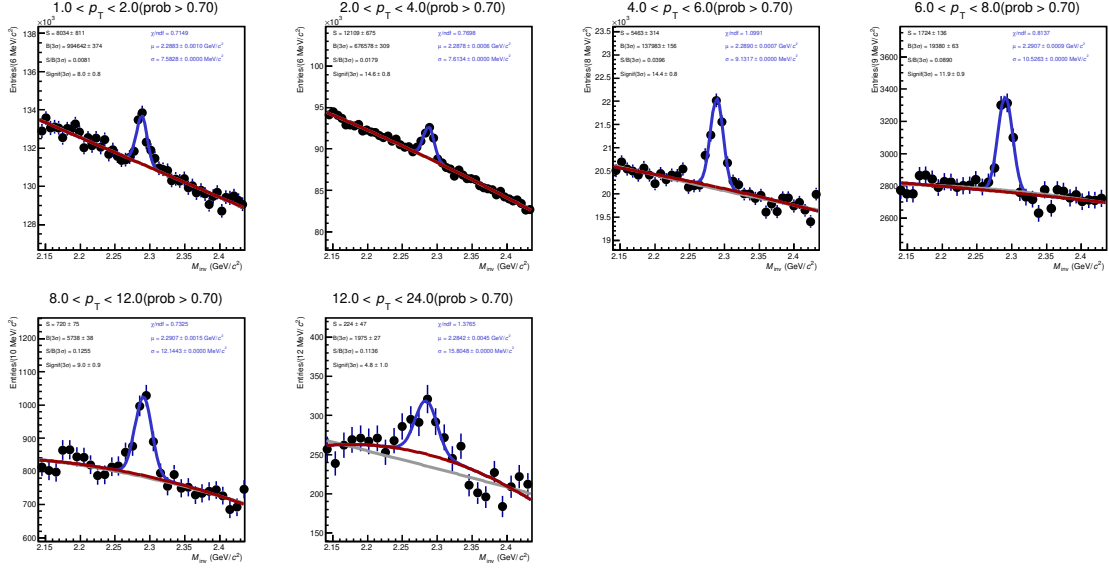


Figure 4.7: Invariant mass spectra for the Λ_c^+ baryon selected in the decay chain $\Lambda_c^+ \rightarrow pK_S^0$ for MB triggered data.

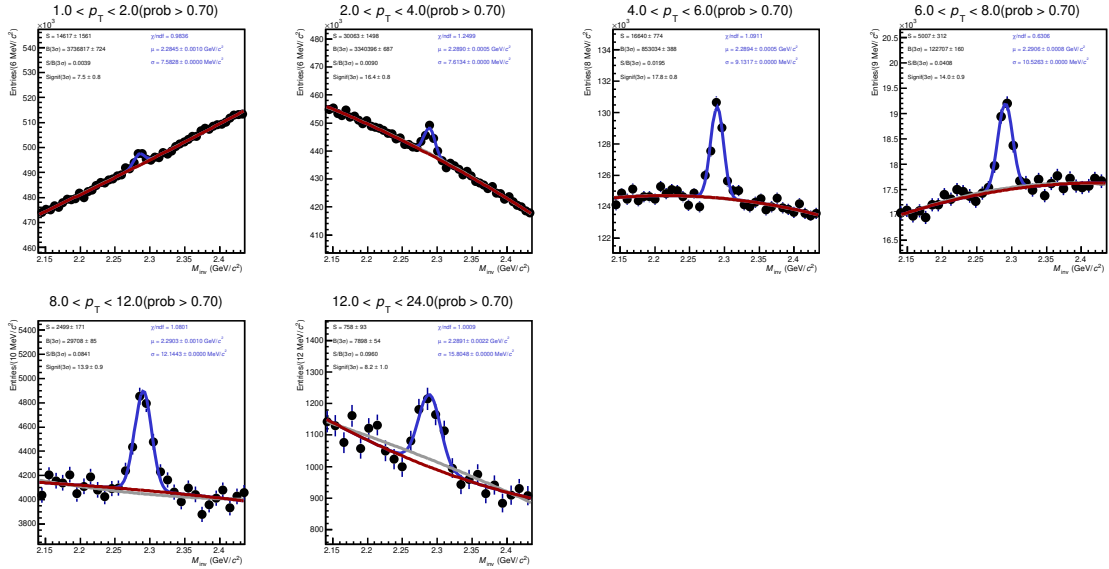


Figure 4.8: Invariant mass spectra for the Λ_c^+ baryon selected in the decay chain $\Lambda_c^+ \rightarrow pK_S^0$ for HMV0M triggered data.



4.1.4 Corrections

There are two corrections are considered in this analysis: the acceptance-times-efficiency correction and feed-down subtraction.

4.1.4.1 Acceptance and efficiency corrections

In this analysis, the raw counts of Λ_c^+ baryons in each p_T and multiplicity bin were corrected for the reconstruction and selection efficiency, as well as the acceptance. Monte Carlo simulations were used to obtain the correction factor for each species. The reconstruction and selection efficiency depend on the multiplicity of charged particles produced in the collision, which affects the primary vertex resolution and the resolution of the topological selection variables. As a result, the selection efficiency has a multiplicity dependence, and the MC multiplicity distributions were reweighted.

The acceptance and efficiency correction factors ($\text{Acc} \times \epsilon$), were determined using Monte Carlo simulations of pp collisions generated with the PYTHIA 8.243 event generator with the Monash 2013 tune. The particles were propagated through the ALICE apparatus using the GEANT3 transport codes, including the luminous region distribution and the conditions (active channels, gain, noise level, and alignment) of all the detectors, taking into account their evolution over the 2016, 2017, and 2018 LHC data taking period. Only events containing a $c\bar{c}$ or a $b\bar{b}$ pair were transported through the apparatus and reconstructed. Λ_c^+ baryons were forced to decay into the hadronic channels considered in this analysis. The efficiency was determined separately for prompt charmed particles and particles from B-meson decays (feed-down).

The nominal simulation samples used Pythia8 with the Monash tune. However, in order to better reproduce the p_T shape of the Λ_c^+ , its p_T shape was re-weighted according to the shape obtained from Pythia8 with the Mode2 tune, which includes colour reconnection effects. The left plot in Fig. 4.9 shows the p_T shapes of the Monte Carlo samples generated with Pythia8 Monash per year, compared to the p_T shape obtained using Mode2. The re-weighting distribution as function as p_T shown in right plot.

Multiplicity binning and corrections

Table 4.1 provides a summary of the multiplicity bins used in the analysis, along with the corresponding mean charged particle multiplicity $\langle dN_{\text{ch}}/d\eta \rangle$, the trigger used to collect events of the respective multiplicity class, the corresponding trigger efficiency, and whether a trigger correction is needed due to the trigger turn on curve.

The interval limits of the p_{VOM} multiplicity bins are defined based on the signal-extraction performance, with the requirement that the edges are in common with the intervals used for light-flavour multiplicity measurements. In addition to the multiplicity binning, the measured particle spectra are

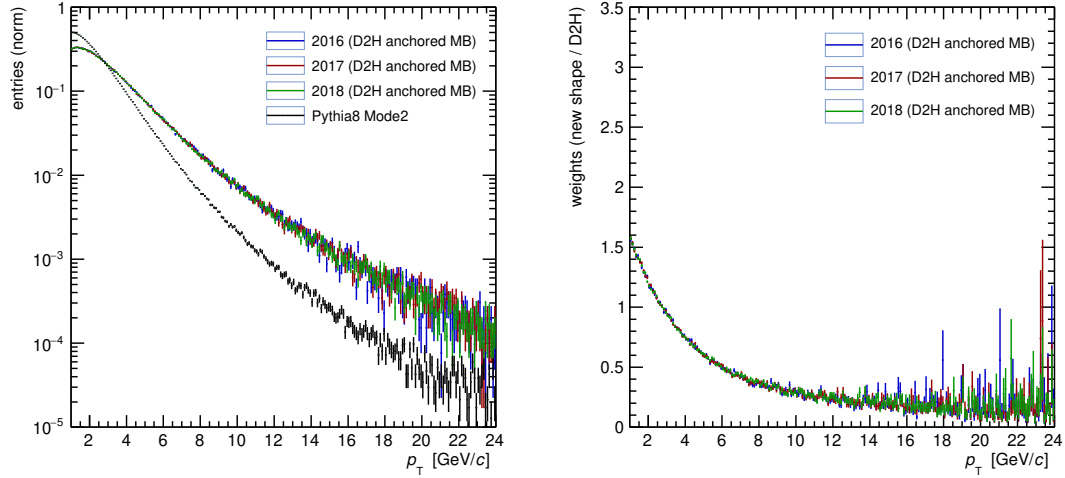


Figure 4.9: MC p_T shapes comparing Pythia8 with Monash tune, as used for the simulated samples, to the Mode2 tune.

normalised to the "accepted" $INEL > 0$ events, which require at least one SPD reconstructed tracklet in the pseudorapidity interval $|\eta| < 1$. This is especially important for the V0M percentile multiplicity estimator, which includes this additional selection. For the SPD tracklet multiplicity estimator, this is taken into account by starting to count from 1 tracklet instead of zero.

When normalising to the $INEL > 0$ class, it is important to correct for event detection efficiency and signal detection efficiency. The first factor does not depend on the particle under study and can be provided by $dN_{ch}/d\eta$ analyses, and these values can be taken from the PWGMM ReferenceMult Twiki page, with the details of their analyses explained in Ref.[?]. The quoted trigger efficiencies in Tab.4.1 are the average of the three available periods, and a weighted average of the V0M percentile multiplicity bins is performed where needed to merge the finer bins used by PWGMM. For the SPD multiplicity estimator in Tab. 4.1, the same table is used, taking the values from V0M percentile multiplicity bins with matching $\langle dN_{ch}/d\eta \rangle$ values. The quoted uncertainty is the envelope of the weighted average for the three available periods separately. The second factor is usually more delicate, but in the charm sector, this correction will be negligible since no signal should be lost after requiring at least one tracklet due to the SPD requirements of the heavy-flavour prongs. In fact, the distributions of n_{tkl} shift towards larger values when at least one D-meson candidate is required in the event.

MC event re-weighting

The selection and reconstruction efficiencies of heavy-flavour hadrons, particularly those with a reconstructed secondary vertex, depend on the multiplicity of charged particles produced during collisions. This is because the resolution of the primary vertex improves at high multiplicity, leading to an



improvement in the resolution of selection variables that use the primary vertex position (such as track impact parameter, particle decay length, and pointing angle). As a result, the selection efficiency exhibits a dependence on multiplicity. It is crucial that the multiplicity distributions in Monte Carlo (MC) samples accurately reproduce the multiplicity distribution in data.

However, the barrel multiplicity is not adequately reproduced in our simulations, and thus the N_{trkl} distributions in each MC are re-weighted to match the distribution from data. Additionally, since the p_{VOM} multiplicity estimator is not calibrated in MC simulations, events in MC are re-weighted based on the barrel multiplicity for analyses using this estimator. Typically, the MC weights are obtained per year through the following steps (with additional steps for p_{VOM} analysis):

Apply full event selection on MC and data (without selecting on the multiplicity estimator). Select only events in MC and data with at least one candidate that meets all candidate selection criteria and requires the candidate invariant mass to be within $\pm 20 \text{ MeV}/c^2$ of the PDG value. Extract the normalised N_{trkl} distribution for both MC and data. Divide the obtained distribution from data by the MC distribution to obtain event weights ω_i for each barrel multiplicity bin i . For HMSPD-triggered data, the corresponding high-multiplicity MC is used. However, since the VOM percentile is not calibrated in simulation, the approach differs slightly. First, events in data are sliced into the desired percentile bins, while no slicing is applied to MC. Then, for each slice, the above procedure is applied, resulting in one set of weights per multiplicity bin. For HVMZERO-triggered data, the corresponding high-multiplicity MC is used. This results in an efficiency correction given by

$$(\text{Acc} \times \varepsilon)_{\text{corr}} = \frac{\sum_i N_{\text{sel},i} \omega_i}{\sum_i N_{\text{gen},i} \omega_i} \quad (4.3)$$

for barrel multiplicity bins i , where $N_{\text{sel},i}$ and $N_{\text{gen},i}$ are the reconstructed and selected, and generated number of candidates in bin i in MC, respectively.

The weights for the $\Lambda_c^+ \rightarrow pK_S^0$ were computed selecting Monte Carlo and data events with at least one $\Lambda_c^+ \rightarrow pK_S^0$ candidate in the $\Lambda_c^+ \rightarrow pK_S^0$ mass range (within a 20 MeV window). The weights have been computed separately for each year. The input distributions and weights for all three years are shown in Fig. 4.10. Then Fig. 4.11 shows the barrel multiplicities in data and MC, MB triggered in the upper and HVMZERO triggered in the lower row for all three years of data taking. The data is sliced into the corresponding p_{VOM} multiplicity bins as explained above and for HM triggered data the corresponding high-multiplicity MC is used. The derived weights for all three years are then summarised in Fig. 4.12.

Fig. 4.13 shows the efficiencies when the MC is reweighted according to the weights as derived and shown in Fig. 4.12. Given the flat trend of the efficiency as a function of N_{trkl} , the efficiency of the

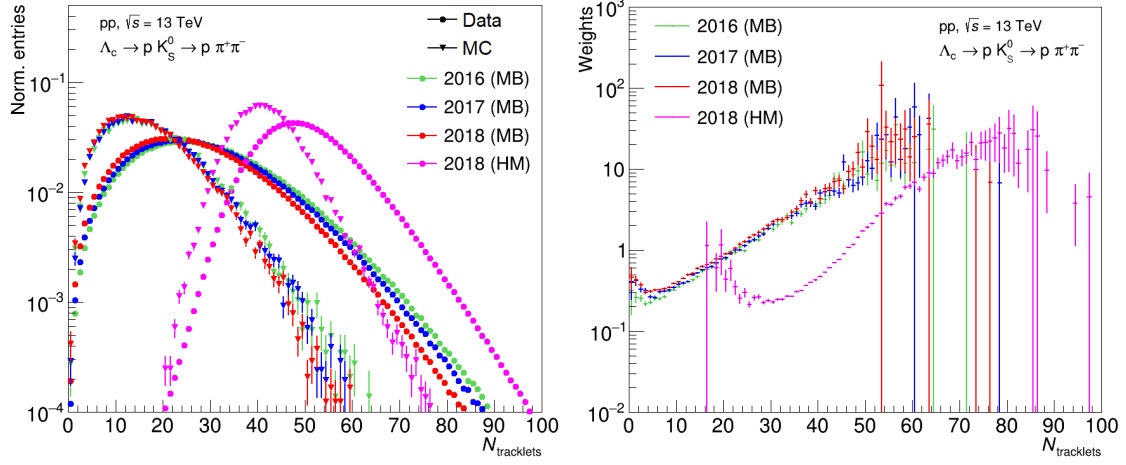


Figure 4.10: Left top: $N_{\text{tracklets}}$ distribution in data (circles) and MC (triangles) for the three MB years and for the HM. Right: Multiplicity weights as a ratio of $N_{\text{tracklets}}$ distribution in data/MC.

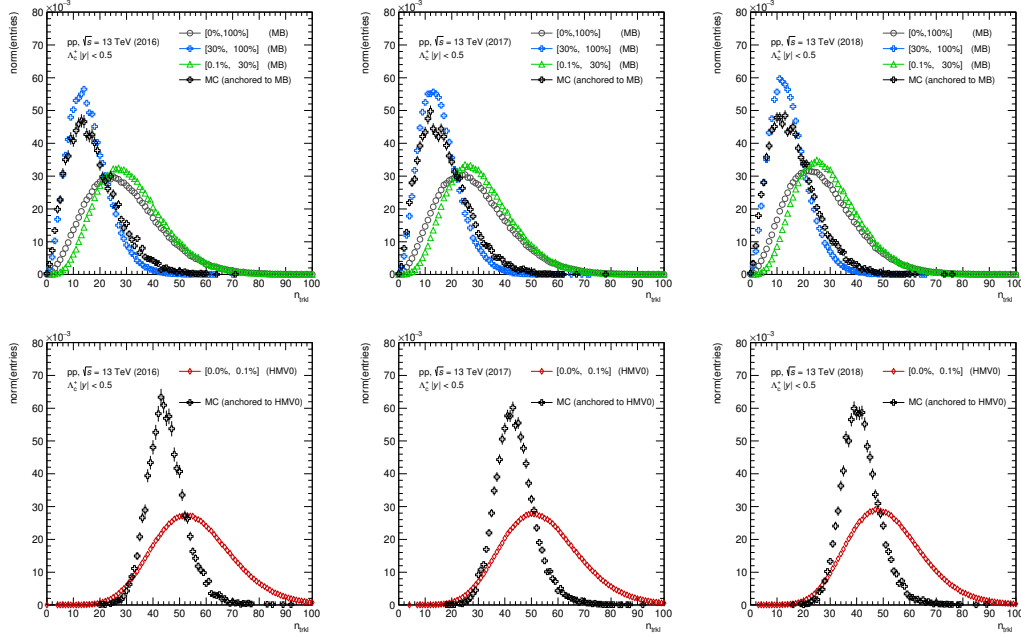


Figure 4.11: Barrel multiplicity in selected events containing at least one $\Lambda_c^+ \rightarrow p K_S^0$ candidate within the mass window of $\pm 20 \text{ MeV}/c^2$ around the PDG value. All years of data taking from left to right, for MB triggered data in the top and for HMVZERO triggered data in the bottom row, respectively.

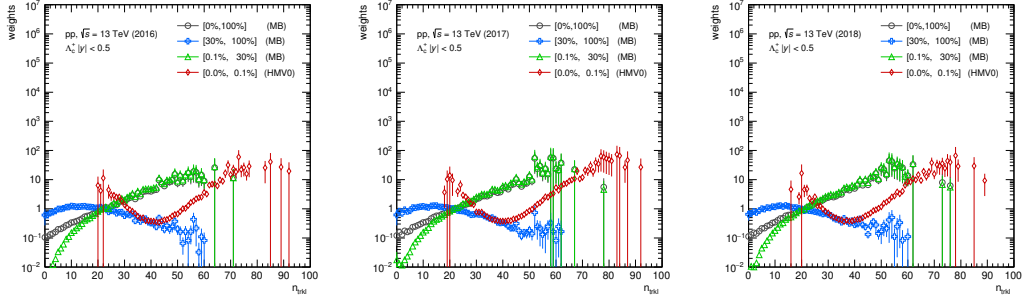


Figure 4.12: MC multiplicity weight for each year with the 4 multiplicity intervals

integrated multiplicity is used for all differential bins, again shown in Fig. 4.14.

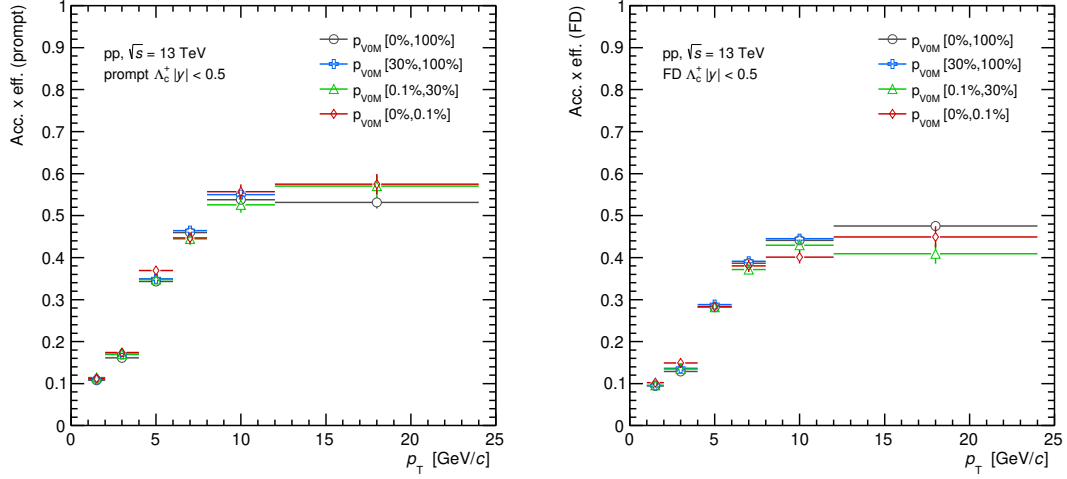


Figure 4.13: Prompt and feed-down acceptance-times-efficiency of VOM percentile multiplicity bins as a function of p_T for the Λ_c^+ baryon selected in the decay chain $\Lambda_c^+ \rightarrow pK_S^0$.

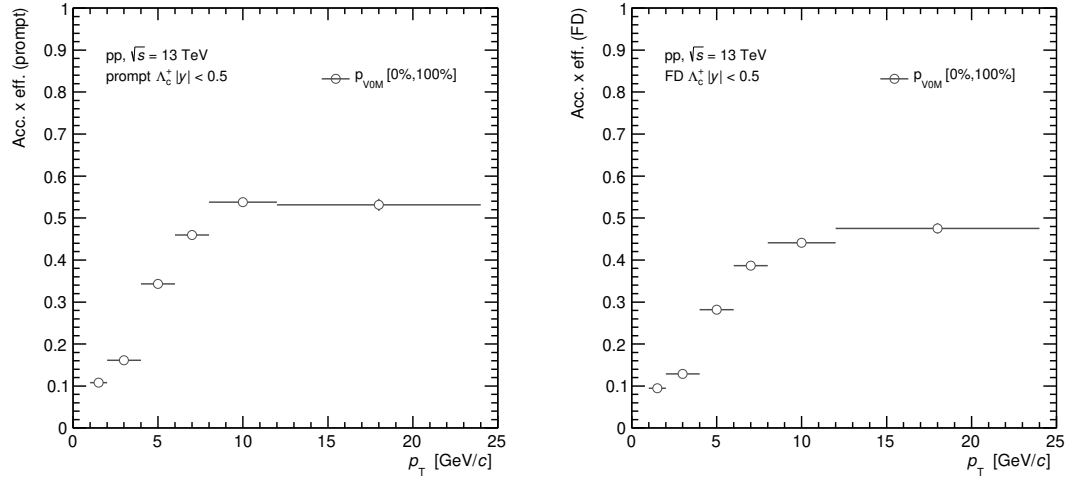


Figure 4.14: Prompt and feed-down acceptance-times-efficiency of the integrated VOM percentile multiplicity bins as a function of p_T for the Λ_c^+ baryon selected in the decay chain $\Lambda_c^+ \rightarrow pK_S^0$.



4.1.4.2 feed-down subtraction

The prompt Λ_c^+ production yields normalised by the number of events in pp collisions were obtained by subtracting the contribution of Λ_c^+ from B decays. The detail as discussed in Sec. 3.2 shown in Eq. 3.4, to calculate the fraction of prompt charm hadrons in pp collisions, the contribution of charm hadrons from beauty hadron decays was subtracted using the FONLL calculation for the beauty production cross section, the PYTHIA8 decayer for $H_B \rightarrow D + X$ decay kinematics, and the Monte Carlo efficiencies for feed-down prompt charm hadrons. The acceptance-times-efficiency for feed-down charm hadrons and the variation of parameters used for the FONLL B predictions were taken into account to evaluate part of the systematic uncertainties related to feed-down subtraction. The Nb method used for this calculation is not multiplicity-dependent, so the f_{prompt} fraction for multiplicity-dependent analyses is taken from the corresponding multiplicity-integrated one. The beauty fragmentation functions are taken from e^+e^- measurements for D mesons and from LHCb measurements for Λ_c^+ baryons. The $\text{Nb} \times 2$ method previously used for Λ_c^+ analyses was replaced by this new strategy, which significantly reduces the systematic uncertainty on the f_{prompt} estimation for Λ_c^+ .

In the middle plot of Fig. 4.15, the fraction obtained using the new FONLL predictions (shown in green) is compared to the old predictions (shown in blue). The right panel suggests that the downward differences between the fractions may be due to the different p_T shapes observed with PYTHIA8 below 5 GeV/c as compared to EvtGen. However, the f_{prompt} fractions are still within compatible ranges of uncertainty. In the left plot of Fig. 4.15, one can see that the removal of the $\text{Nb} \times 2$ procedure for estimating the feed-down FONLL systematic has resulted in significant improvement for the Λ_c^+ case.

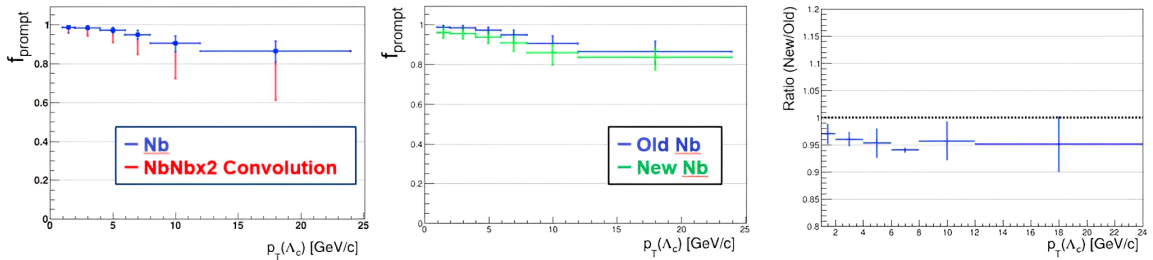


Figure 4.15: Left: Comparison of uncertainties using Nb and $\text{Nb} \times 2$ convolution method (old FONLL predictions). Mid: $\Lambda_c^+ \rightarrow pK_S^0$ prompt fraction (multiplicity integrated) vs. p_T using the N_b method. Green = new FONLL predictions, blue = old FONLL predictions. Right: Ratio of new to old corrected yield.

4.1.5 Systematic uncertainties

The values of systematic uncertainty on the Λ_c^+ corrected yields in different multiplicity intervals were estimated with these procedures: (i) the raw-yield extraction from the invariant-mass distributions; (ii) the selection efficiency estimation; (iii) the track-reconstruction efficiency; (iv) the MC p_T shape in



the simulation; (v) the feed-down subtraction from beauty-hadron decays.

4.1.5.1 Raw-yield extraction

To evaluate the raw yield extraction systematic uncertainty for the Λ_c^+ baryon, a *MultiTrial* approach was used by varying the background fit function, range of the invariant mass spectra, binning, and mass peak sigma. Four variations in the fit strategy were considered for each case, including constraining both the mean and sigma of the signal fit function, constraining only one of the two parameters, or leaving both parameters free. Two methods of bin counting were also considered. Only fits with a reduced χ^2 below two were considered for variation. Tab. 4.6 summarizes the parameter spaces for $\Lambda_c^+ \rightarrow pK_S^0$. The systematic uncertainties on yield extraction were evaluated for the multiplicity integrated case, lower and higher multiplicity intervals, and HM triggered data for each p_T bin.

Table 4.6: Variations used for $\Lambda_c^+ \rightarrow pK_S^0$ MultiTrial.

Fix/BC cases	fixed sigma & free mean fixed width & fixed mean 1.15· width & free mean 0.85· width & free mean
lower bound fit range	{2.136, 2.116, 2.126, 2.1146, 2.156}
upper bound fit range	{2.446, 2.426, 2.436, 2.456, 2.466}
rebin values	{0}
background functions	{ Exponential and first, second, third order polynomial }

Figure 4.16 shows an example of the MultiTrial fitter output for $\Lambda_c^+ \rightarrow pK_S^0$, including the Gaussian mean, Gaussian sigma, reduced χ^2 , and raw yield distributions. The systematic uncertainty was estimated by considering the RMS of the fit distribution and the mean of the fit bincounting distributions, compared to the central value. The final values, which include an additional p_T smoothening, are presented in Table 4.7.

4.1.5.2 Selection efficiency

To ensure accurate measurements, detector acceptances and efficiencies are determined through Monte Carlo simulations and used to correct the data. However, discrepancies between data and simulation in the selection criteria could introduce biases in the extracted yields. To quantify this effect and estimate a corresponding systematic uncertainty, we compared corrected yields obtained using different

Table 4.7: Systematic uncertainties for raw yield extraction of $\Lambda_c^+ \rightarrow pK_S^0$ were evaluated in both multiplicity integrated and multiplicity range cases.

p_T intervals [GeV/c]	1 - 2	2 - 4	4 - 6	6 - 8	8 - 12	12 - 24
VOM mult int (%)	6	8	8	6	7	10
[30 - 100]%	6	8	7	6	6	-
[0.1 - 30]%	7	9	8	6	8	8
[0 - 0.1]%	7	9	8	6	7	8

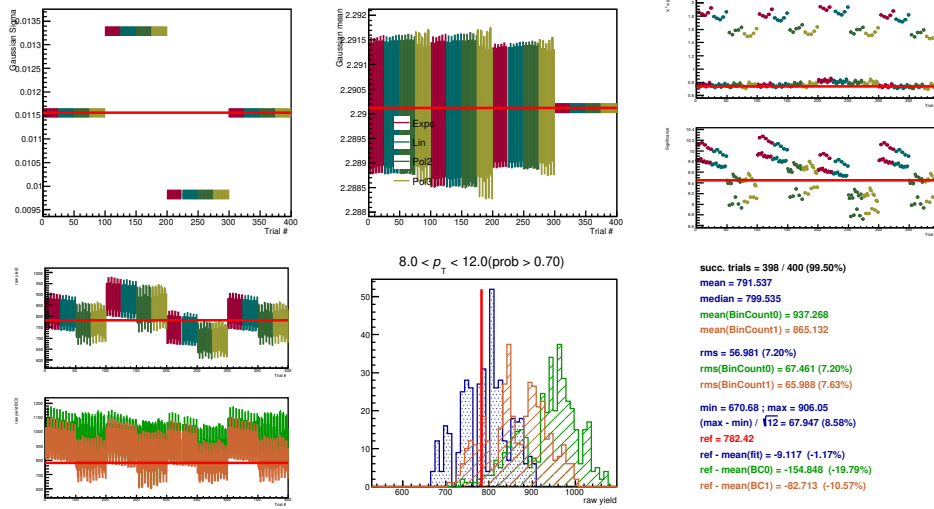


Figure 4.16: The output of the MultiTrial fitter for the $\Lambda_c^+ \rightarrow pK_S^0$ decay in the $[8-12]$ GeV/ c p_T bin for the multiplicity integrated case. The figure displays the Gaussian mean (top left), Gaussian sigma (top middle), the reduced χ^2 of the fit (top right), and the extracted raw yield distributions using both the fit and bin counting methods (bottom row).

machine learning (ML) working points. For the decay $\Lambda_c^+ \rightarrow pK_S^0$, the ML working point was varied while keeping the efficiency within a $\pm 25\%$ window around the nominal efficiency, choosing 20 higher and 20 lower working points equidistantly around the nominal value. The full analysis was then conducted for each of the selected working points, with ratios computed relative to the nominal values. The resulting variations in corrected yield were used to estimate systematic uncertainties for the analyzed V0M multiplicity bins $[30 - 100]\%$, and $[0 - 0.1]\%$, as shown in Fig. 4.17.

4.1.5.3 Track-reconstruction efficiency

The systematic uncertainty related to the tracking efficiency is determined by considering the impact of track finding in the TPC, track propagation from TPC to ITS, and track quality selections. Two tests were conducted to estimate this uncertainty: comparing the corrected yields obtained with different TPC track selection cuts and comparing the TPC-ITS track matching efficiency in data and simulations. To avoid dependence on the multiplicity estimator and intervals, this uncertainty is only estimated for the MB case. The track selection cuts was applied to $\Lambda_c^+ \rightarrow pK_S^0$. The assigned uncertainties per daughter track range from 1-1.5% for pions, while the matching efficiencies can be found in [206]. The systematic uncertainties for different periods are averaged into one final value: 1.8% for $[1-2]$ GeV/ c , 2.7% for $[2-4]$ GeV/ c , 2.3% for $[4-6]$ GeV/ c , 2.4% for $[6-8]$ GeV/ c , 3.0% for $[8-12]$ GeV/ c , and 2.9% for $[12-24]$ GeV/ c . To propagate the uncertainty at the track level to the Λ_c^+ baryon level, a MC simulation was used, taking into account the daughter's kinematics in the HF hadron p_T range of our analysis. The final assigned uncertainties for $\Lambda_c^+ \rightarrow pK_S^0$ can be found in Fig. 4.18.

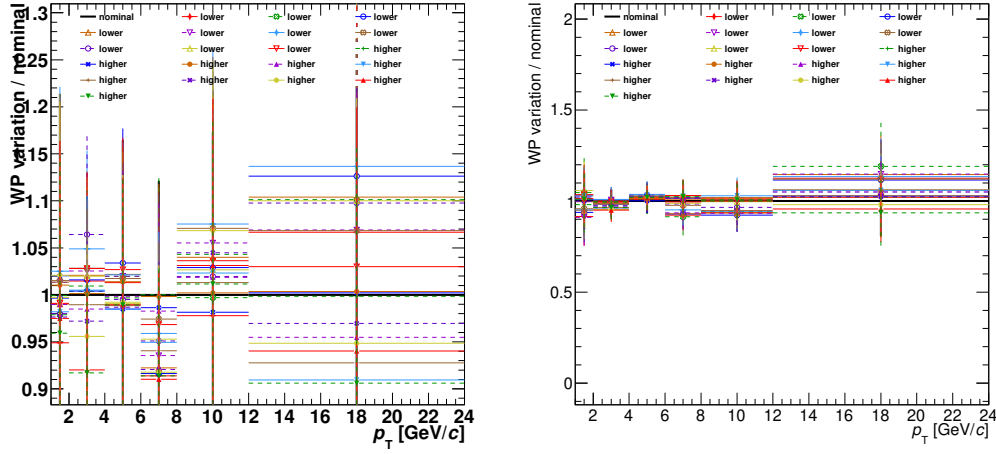


Figure 4.17: Ratios of corrected yield with different ML working points over the nominal values for the $\Lambda_c^+ \rightarrow pK_S^0$ VOM multiplicity bins [30 – 100]% (left), [0 – 0.1]% (right).

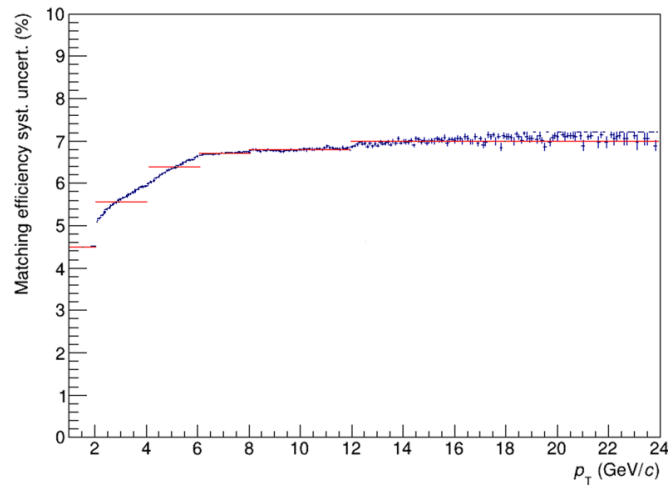


Figure 4.18: Total tracking systematic, containing the TPC track quality and ITS-TPC matching efficiency part, propagated to the kinematics of the selected hadrons for $\Lambda_c^+ \rightarrow pK_S^0$.



4.1.5.4 MC generated p_T shape

To evaluate the systematic effect on the efficiency derived from simulations due to the generated p_T shape of Λ_c^+ baryons, we calculated the efficiency with and without using p_T -dependent weights to correct for the difference between the p_T spectrum of generated Λ_c^+ baryons in the PYTHIA8 MC production used in our analyses and the prediction from FONLL. In Sec. 4.1.4.1, we reweighted the Λ_c^+ p_T shape to the simulated shape from Pythia8 with tune Mode2. To estimate the systematic effect of the generated p_T shape on the efficiency, we alternatively reweighted the simulated samples to the calculated FONLL shape. The two p_T shapes obtained from the Pythia8 Mode2 simulation and FONLL calculation are shown on the left of Fig. 4.19. The effective weights were calculated by dividing the FONLL by the Pythia8 shape, which is shown in the middle. The impact on the prompt efficiency is then shown on the right, after dividing the two efficiencies obtained with the different reweighting. The ratio is taken as the relative systematic uncertainty, which is marginal and reaches a maximum value of 2% in the highest p_T bin.

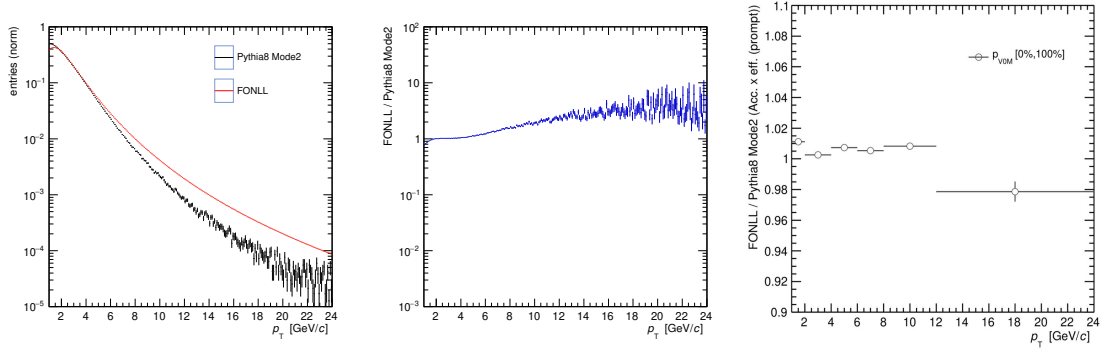


Figure 4.19: Estimation of relative uncertainty from generated p_T shape for Λ_c^+ . The left plot displays the p_T shapes of Pythia8 Mode2 and from FONLL calculations. The middle plot shows the weights obtained by dividing the FONLL shape by the Pythia8 shape. The right plot shows the ratio of prompt efficiencies, one reweighted to Pythia8 Mode2 p_T shape (nominal shape used in this analysis) and the second reweighted to the FONLL shape.

4.1.5.5 Feed-down subtraction

As discussed in Sec. 4.1.4.2, the f_{prompt} factor is computed from FONLL predictions. In addition, we made the assumption that the relative contribution of beauty hadron decays to the Λ_c^+ yields in the different multiplicity bins is constant and equal to that of the multiplicity integrated sample. We therefore quote, in constrast to the usual HF studies, two different feed-down subtraction systematics. Depending on the observable, one or both may cancel. This will be discussed in more detail later, in this section we will discuss the strategy and estimation of both contributions to the systematic. The systematic on the feed-down estimation is performed by varying the parameters used for the FONLL B predictions using the Nb method, see Eq. 3.4. In particular, the b quark mass and the factorisation and renormalisation scales in the FONLL calculations were varied. For the Λ_c^+ baryons, the Nb $\times$ 2 method is no longer applied



to estimate the systematic. Instead, a similar procedure as for the D mesons is performed, exploiting a recent Λ_b^0 measurement of LHCb [207]. For more details, see Sec. 4.1.4.2.

We assume that the contribution of beauty hadron decays to the yields of D or Λ_c^+ hadrons in different multiplicity bins remains constant and is equal to that of the multiplicity-integrated sample. However, the dependence of B hadrons and prompt D mesons on the charged particle density may vary. Therefore, we evaluate the systematic uncertainty due to this effect. We used PYTHIA 8.2.34 tunes [173] to define multiplicity-dependent variations to be applied on $f_{\text{non-prompt}}$, following a similar strategy as the one used for the previous HF versus multiplicity analysis in pp collisions at 7 TeV [208], but with updated variations.

We investigated different PYTHIA8 tunes, including standard ones like the Monash and 4Cx tune [131], as well as the so-called colour-reconnection tunes [137]. The top two panel of Fig. 4.20 shows a schematic diagram of the multiplicity trends for B and D hadrons, based on the ratio $f_{\text{non-prompt}}/f_{\text{non-prompt-MB}}$, for most of the tunes. A clear dependency is visible, with a higher f_{prompt} fraction at low $dN_{\text{ch}}/d\eta$ values, and a higher $f_{\text{non-prompt}}$ fraction at high $dN_{\text{ch}}/d\eta$, as expected. No significant difference is observed for charmed mesons and charmed baryons, so we decided to use uniform variations for D mesons and the Λ_c^+ baryons. The variations are asymmetric and defined by the deviation with one (the MB $f_{\text{non-prompt}}$ fraction). The bottom panel of Fig. 4.20 shows the defined variations for the SPD multiplicity estimator case, and Tab. 4.8 summarizes them.

We utilized the multiplicity integrated prompt fraction, as discussed in Sec. 4.1.4.2, for the charm hadron species, and varied it with respect to multiplicity assuming the B/D yields as described earlier. We first computed the fraction of yield coming from B feed-down, denoted as f_B , for each hadron species and p_T bin using the relation $f_B = 1 - f_{\text{prompt}}$. Then, we applied a variation to f_B in each multiplicity bin using the factors listed in Tab. 4.8. This resulted in a range of f_B values, which were converted into a range of f_{prompt} using $f_{\text{prompt}} = 1 - f_B$. Finally, we obtained a relative uncertainty on our measurement by computing the ratio $f_{\text{prompt}}^{\text{mult.,bin}}/f_{\text{prompt}}^{\text{mult.,integrated}}$, where the f_{prompt} range was derived from the multiplicity bins and integrated values.

We performed the systematic variation independently for two cases: single spectra and ratio with multiplicity integrated. For the single spectra case, we took the envelope after applying the variations on the upper, central, and lower f_{prompt} values. We also included the upper and lower values of the uncertainties on f_{prompt} , which were obtained from the variation of the pQCD parameters in the FONLL calculation. Since the same set of parameters had to be considered for both the multiplicity integrated and



multiplicity bins, we considered these uncertainties as well. For the ratio with multiplicity integrated, we only applied the variations on the central f_{prompt} value. We assumed that the values of pQCD scales and quark mass did not depend on multiplicity, and thus did not consider the systematic uncertainty coming from the FONLL theoretical uncertainties in the variation.

The FONLL part essentially "cancels" between the multiplicity interval and minimum bias. Hence, a smaller systematic uncertainty was assigned in the ratio with multiplicity integrated than in the corrected yield spectra. The final uncertainties ranged between 0 and 15% and are presented in Tab. 4.9.

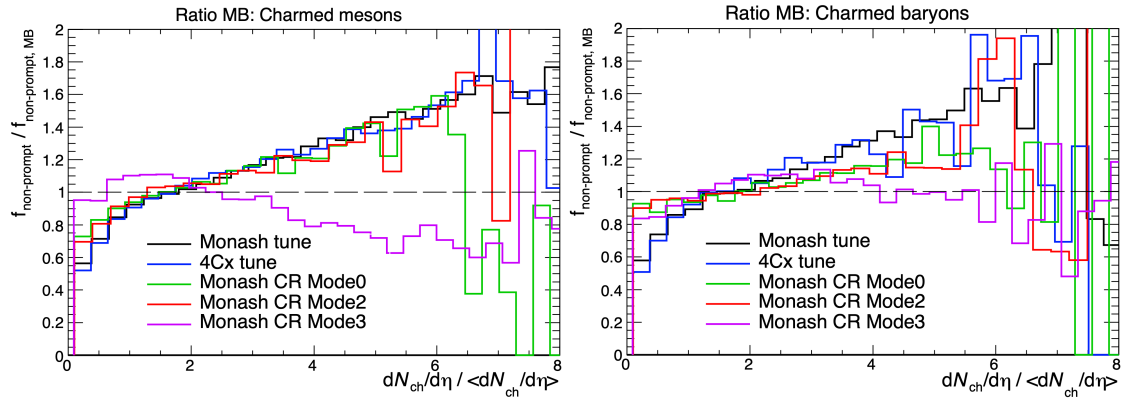


Figure 4.20: Ratio of $f_{\text{non-prompt}}/f_{\text{non-prompt-MB}}$ versus relative $dN_{\text{ch}}/d\eta$ for charmed mesons (top left) and charmed baryons (top right) for different PYTHIA 8 tunes [131, 137] using softQCD. This input is used to define the systematic variations is reported in Tab. 4.8.

Table 4.8: Variations on $f_{\text{non-prompt}}$ used for the different multiplicity bins.

Multiplicity interval	$f_{\text{non-prompt}}$ variation
[30 – 100]%	[0.7 - 1.0]
[0.1 – 30]%	[0.9 - 1.2]
[0 – 0.1]%	[1.0 - 1.4]

Table 4.9: Systematic uncertainties due to the multiplicity dependence in the feed-down subtraction for $\Lambda_c^+ \rightarrow pK_S^0$ in the multiplicity ranges on spectra(for mult/MB ratio) in percentage.

p_T intervals [GeV/c]	1 - 2	2 - 4	4 - 6	6 - 8	8 - 12	12 - 24
[30-100]%	+1 (+1) -0 (-0)	+2 (+1) -0 (-0)	+2 (+1) -0 (-0)	+3 (+2) -0 (-0)	+4 (+3) -0 (-0)	/
[0.1-30]%	+0 (+0) -1 (-1)	+1 (+0) -1 (-1)	+1 (+0) -1 (-1)	+1 (+1) -2 (-1)	+1 (+1) -3 (-2)	+2 (+1) -4 (-2)
[0-0.1]%	+0 (+0) -2 (-1)	+0 (+0) -2 (-1)	+0 (+0) -3 (-2)	+0 (+0) -4 (-2)	+0 (+0) -6 (-4)	+0 (+0) -7 (-5)



4.1.6 Results and discussion

The p_T -differential corrected yield of the Λ_c^+ baryon was obtained in the different event-multiplicity classes via the V0M percentile p_{V0M} , averaging the results from the two decay channels $\Lambda_c^+ \rightarrow pK^-\pi^+$ and $\Lambda_c^+ \rightarrow pK_S^0$ to obtain a more precise measurement, for which the inverse of the quadratic sum of the relative statistical and uncorrelated systematic uncertainties were used as weights. The statistical and systematic uncertainties, as estimated in the previous Section 4.1.5, have been taken into account and are represented by vertical bars and boxes, respectively. The results are depicted in Fig. 4.21, which shows that the Λ_c^+ yield per event increases with multiplicity. The measured values of the [0.1-30]% and [0-0.1]% intervals lie above the multiplicity-integrated yield, while the data for the low-multiplicity interval of [30-100]% lie below. This observation is consistent with the expectation that the abundance of all particle species should scale with multiplicity by the same amount in pp collisions, such that a higher (lower) number of particles is produced with increasing (decreasing) multiplicity.

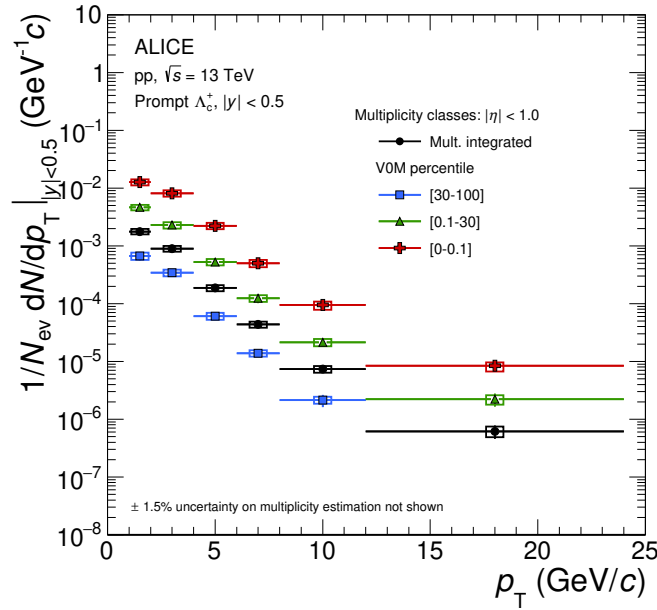


Figure 4.21: Prompt Λ_c^+ yield per event as a function of candidate p_T and in intervals of V0M percentile p_{V0M} . Statistical and systematic uncertainties are indicated by vertical bars and boxes, respectively.

The p_T -differential spectra of D^0 and D_s^+ measured in $|y| < 0.5$ are also shown in Fig. 4.22 for the INEL > 0 class and the four multiplicity classes selected using the V0M percentile p_{V0M} estimator at forward rapidity. The statistical and total systematic uncertainties are shown by vertical error bars and boxes, respectively, as for all the figures in this section. The bottom panels of Fig. 4.22 present the ratios to the INEL > 0 class, for which the multiplicity-dependent systematic sources were considered as un-



correlated among different multiplicity classes and the contributions of the tracking and PID efficiency, the shape of MC p_T spectra and z_{vtx} distribution, the beauty feed-down, and the branching ratio as correlated. The statistical uncertainties and the systematic uncertainties related to the selection efficiency and to the raw-yield extraction were considered partially correlated with respect to the measurement performed in the $\text{INEL} > 0$ class.

The measured p_T -differential yields increase from the lowest to the highest multiplicity class for the three hadron species. Their ratios to $\text{INEL} > 0$ increase (decrease) with increasing p_T for the highest (lowest) multiplicity class, suggesting a plateau towards $p_T > 10 \text{ GeV}/c$, as recently observed for the light-flavour hadrons in Refs. [200, 201, 209], where it was explained as a hardening of the measured p_T spectra with increasing $\langle dN_{\text{ch}}/d\eta \rangle$. Different MPI models were able to describe such effects [210], and it was shown to be more pronounced for protons than for kaons and pions, while similar for strange baryons and mesons.

In order to investigate potential differences in the $\langle dN_{\text{ch}}/d\eta \rangle$ dependence of the D^0 -meson production with respect to the D_s^+ meson and Λ_c^+ baryon, the D_s^+/D^0 and Λ_c^+/D^0 yield ratios are compared in different multiplicity event classes in Fig. 4.23, considering both forward and midrapidity multiplicity estimators. The sources of uncertainty assumed to be uncorrelated between different charm-hadron species included the raw-yield extraction, the selection efficiency, the shape of the MC p_T spectra, the z_{vtx} distribution, and the branching ratio. The systematic uncertainty deriving from the variation of the multiplicity-interval limits was propagated as partially correlated, while the other systematic uncertainties were assumed to be fully correlated.

Within the current experimental uncertainties, the D_s^+/D^0 ratios are independent of p_T in the measured p_T range. They are compatible with measurements performed in pp collisions at $\sqrt{s} = 5.02$ and 7 TeV [211], and also with the average of the p_T -integrated results from experiments at e^+e^- and ep colliders, 0.17 ± 0.03 [134, 212]. A dependence of these ratios with multiplicity, as seen for the ratio of (multi-)strange hadrons to $\pi^\pm$ [69, 201], is not observed within the uncertainties.

The p_T -differential Λ_c^+/D^0 ratios show an evident dependence on multiplicity, and a hierarchy is observed going from the lowest to the highest multiplicity intervals, for both the N_{trk} and p_{VOM} estimators, for all but the first p_T bin. The increasing trend with $\langle dN_{\text{ch}}/d\eta \rangle$ for the Λ_c^+/D^0 ratio is consistent among the measurements done with the two multiplicity estimators, indicating that the enhancement between low and high multiplicity intervals is not a consequence of a possible bias arising from the coinciding rapidity regions between the multiplicity estimator and the measurement of interest at midra-

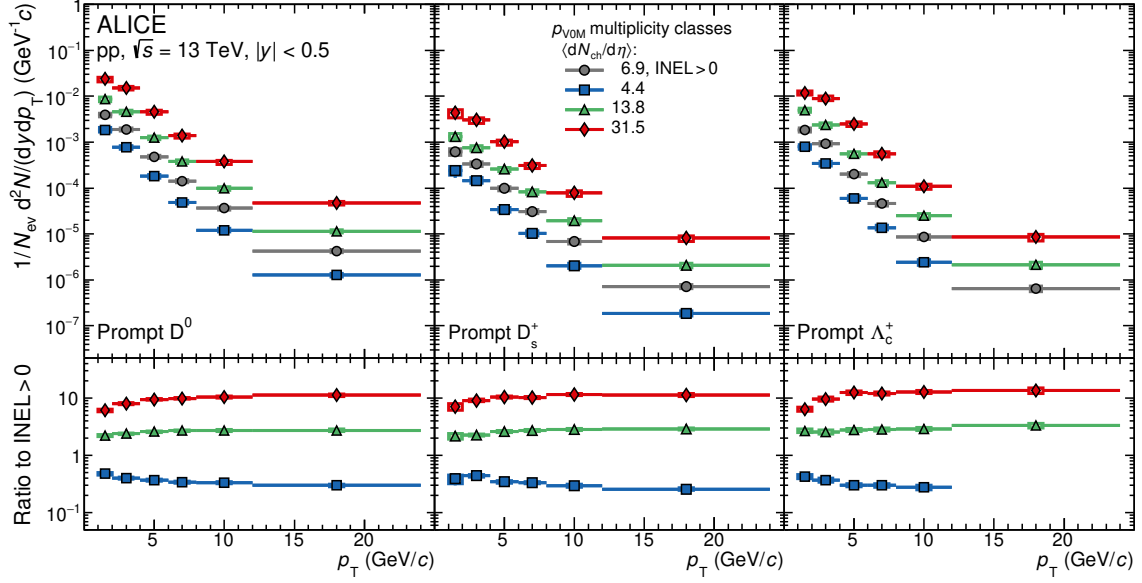


Figure 4.22: Transverse-momentum spectra of D^0 , D_s^+ , and Λ_c^+ hadrons measured in pp collisions at $\sqrt{s} = 13$ TeV for different multiplicity classes selected with the p_{V0M} estimator at forward rapidity. The corresponding ratios to INEL > 0 are shown in the bottom panels.

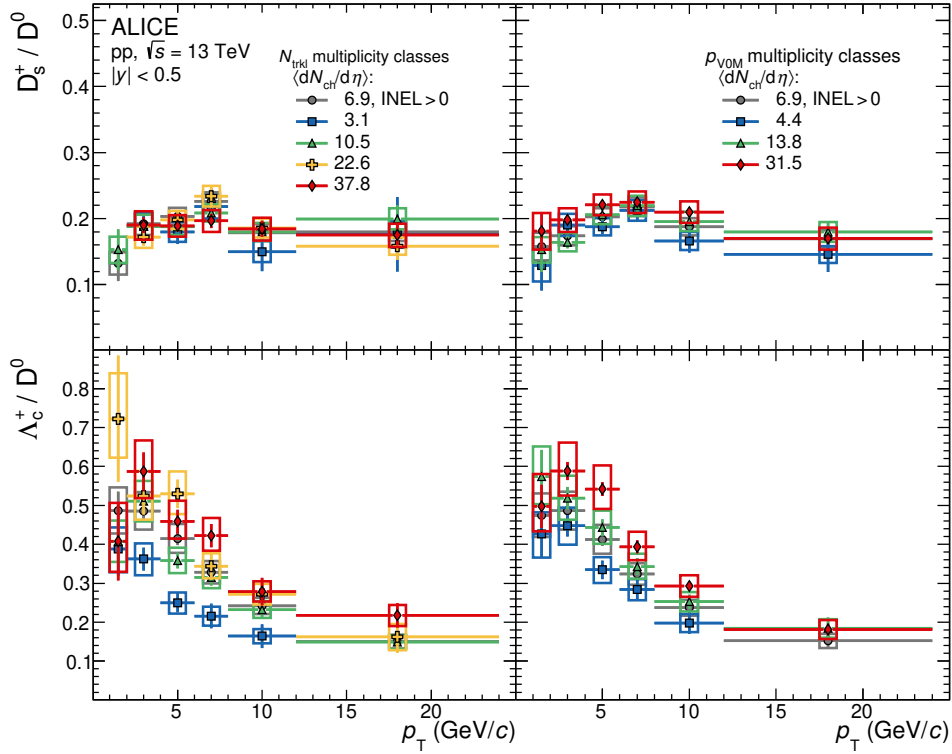


Figure 4.23: The D_s^+/D^0 (top) and Λ_c^+/D^0 (bottom) ratios measured in pp collisions at $\sqrt{s} = 13$ TeV for different multiplicity classes at mid- (left) and forward (right) rapidity.



pidity. It is worth noticing that the measured Λ_c^+/D^0 ratio in the lowest multiplicity class is still higher, in the measured p_T range, than the average of corresponding ratios measured in e^+e^- collisions at LEP, which was found to be $0.113 \pm 0.013(\text{stat}) \pm 0.006(\text{syst})$ [212, 213].

In order to estimate a significance level for the difference observed in the two extreme multiplicity classes at midrapidity, the highest multiplicity (HM) over the lowest multiplicity (LM) Λ_c^+/D^0 ratio was computed. The probability of the measured double-ratio $\text{DR} = (\Lambda_c^+/D^0)_{\text{HM}}/(\Lambda_c^+/D^0)_{\text{LM}} > 1$, corresponds to a significance of 5.3σ in the $1 < p_T < 12$ GeV/ c interval, considering as null hypothesis $\text{DR} = 1$. This estimate was performed taking into account statistical and systematic uncertainties, for which the raw-yield extraction, the selection efficiency, the shape of the MC p_T spectra, and the z_{vtx} distribution sources were considered as uncorrelated, the multiplicity-interval limits as partially correlated, while the other sources cancelled out in the double ratio. With the aim of investigating the least favourable case, the measured values in all p_T intervals were shifted down by one standard deviation, by considering the sources of systematic uncertainties correlated with p_T that do not cancel out in the double ratio, i.e. those arising from the selection efficiency and the generated p_T spectra.

The measured charm-hadron ratios for the lowest and highest multiplicity class for the N_{trkl} multiplicity estimator are compared to model predictions from MC generators and a statistical hadronisation model in Fig. 4.24. The simulations with the PYTHIA event generator were performed with the Monash and the CR-BLC tunes. For the latter, three modes are suggested by the authors, applying different constraints on the allowed reconnections among colour sources, in particular concerning the causality connection among strings involved in a reconnection, and time dilation caused by relative boosts of the strings [137]. The simulations are shown in intervals of primary particle multiplicities selected at midrapidity, evaluated by studying the correlation between N_{trkl} intervals and N_{ch} values. The estimated intervals are $1 \leq N_{\text{ch}} \leq 12$ and $N_{\text{ch}} > 75$ for the lowest and highest multiplicity event classes, respectively. The measured D_s^+/D^0 ratios at low and high multiplicity are compatible with PYTHIA with Monash and CR-BLC tunes. The Monash tune, however, does not reproduce the Λ_c^+/D^0 ratio, and furthermore it does not show a multiplicity dependence. By contrast, the CR-BLC tunes describe the Λ_c^+/D^0 decreasing trend with p_T , and are closer to the overall magnitude, as also observed for minimum-bias pp collisions at $\sqrt{s} = 5.02$ and 13 TeV [67, 213]. The CR-BLC tunes show a clear dependence with multiplicity, qualitatively reproducing the trend observed in data.

The measurements in Fig. 4.24 are also compared with the predictions of a canonical-ensemble statistical hadronisation (CE-SH) model [214], where the authors generalise the grand-canonical statistical hadronisation model (SHM) [136] of charm-hadron production to the case of canonical SHM, and

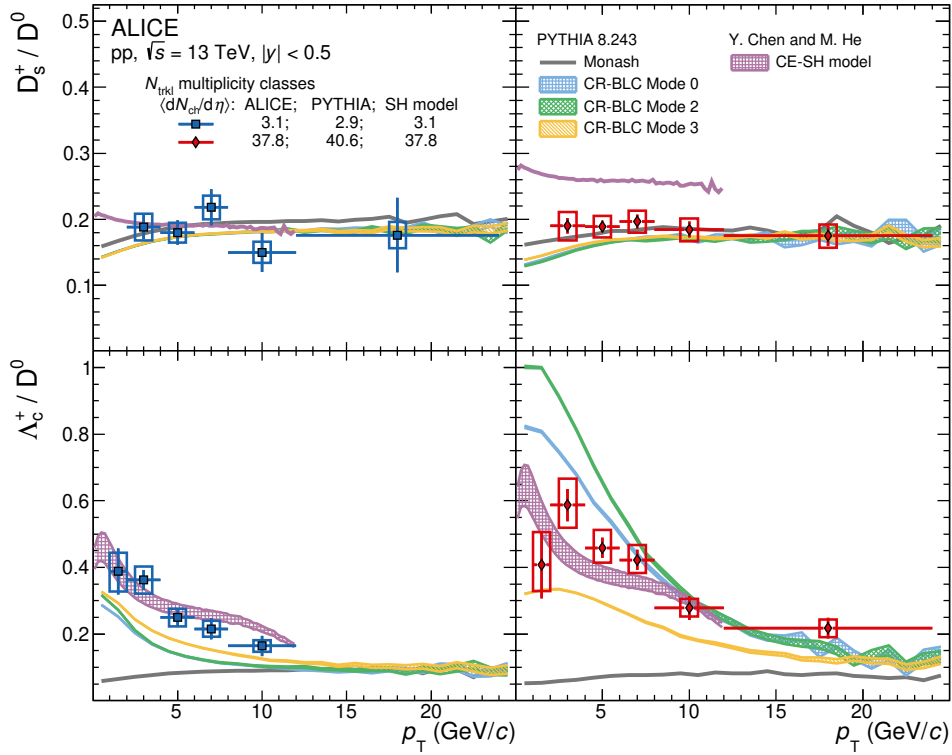


Figure 4.24: The D_s^+/D^0 (top) and Λ_c^+/D^0 (bottom) ratios measured in pp collisions at $\sqrt{s} = 13$ TeV for the lowest (left) and highest (right) multiplicity classes at midrapidity. The measurements are compared to PYTHIA predictions with the Monash [131] and the CR-BLC tunes [137], and the CE-SH model [214], estimated in similar multiplicity classes. The uncertainty bands for the PYTHIA predictions are the statistical uncertainties on the simulations, while for the CE-SH model they refer to the variation of the branching ratios of the additional charm-baryon states from RQM [215].



explore the multiplicity dependence of charm-hadron particle ratios. The version of the SHM model based on the measured charm-baryon spectrum reported by the PDG [216] was observed to strongly underestimate the Λ_c^+/D^0 measurements in minimum-bias pp collisions [213]. For this reason, for the Λ_c^+/D^0 case, the underlying charm-baryon spectrum in the calculations is augmented to include additional excited baryon states predicted by the Relativistic Quark Model (RQM) [215]. For the D_s^+/D^0 predictions, only the PDG case is shown, since the RQM does not modify the D-meson yields with respect to the PDG set. The model calculations describe the Λ_c^+/D^0 ratios, reproducing the multiplicity dependence. The D_s^+/D^0 prediction is compatible with the measurement for the low multiplicity class, while it overestimates the data in the highest multiplicity interval. The CE-SH model explains the multiplicity dependence as deriving from the reduced volume size of the formalism towards smaller multiplicity, where a decrease of the Λ_c^+/D^0 ratio is a consequence of the strict baryon-number conservation. Such behaviour is also predicted for charm-strange mesons relative to charm mesons, based instead on strangeness-number conservation.

Fig. 4.25 shows the comparison of the Λ_c^+/D^0 and the Λ/K^0 [201] baryon-to-meson ratios as a function of p_T in pp collisions at $\sqrt{s} = 13$ TeV, in similar low and high N_{trkl} and p_{VOM} multiplicity classes. In the vacuum-fragmentation scenario, the light-flavour hadron production has a significant contribution from gluon fragmentation, whereas heavy-flavour hadrons are primarily produced through the fragmentation of a charm quark, which is in turn produced in the initial hard scattering. In addition, at low p_T , light-flavour hadrons originate mainly from small-momentum soft scattering processes. Despite these differences, the light- and heavy-flavour baryon-to-meson ratios, Λ_c^+/D^0 and Λ/K^0 , show a remarkably similar trend as a function of $\langle dN_{\text{ch}}/d\eta \rangle$. The measurements also suggest a similar shift of the baryon-to-meson ratio peaks towards higher momenta, with increasing multiplicity. These similarities, observed as well in minimum-bias pp and p – Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV both in terms of shape and magnitude [217], hint at a potential common mechanism for light- and charm-baryon formation in hadronic collisions at LHC energies.

The p_T -integrated yields of Λ_c^+ and D^0 were computed by integrating the p_T -differential spectra in their measured range and extrapolating them down to $p_T = 0$ in each multiplicity interval. In the integration, the systematic uncertainties were propagated considering the uncertainty due to the raw-yield extraction and the statistical uncertainty on the efficiency as fully uncorrelated and all the other sources as fully correlated among p_T intervals. The PYTHIA predictions with CR-BLC Mode 2 were used for the extrapolation in each multiplicity interval, for both Λ_c^+ and D^0 , following a similar procedure as the one described in Ref. [213]. The extrapolation factor was computed as the ratio of the PYTHIA

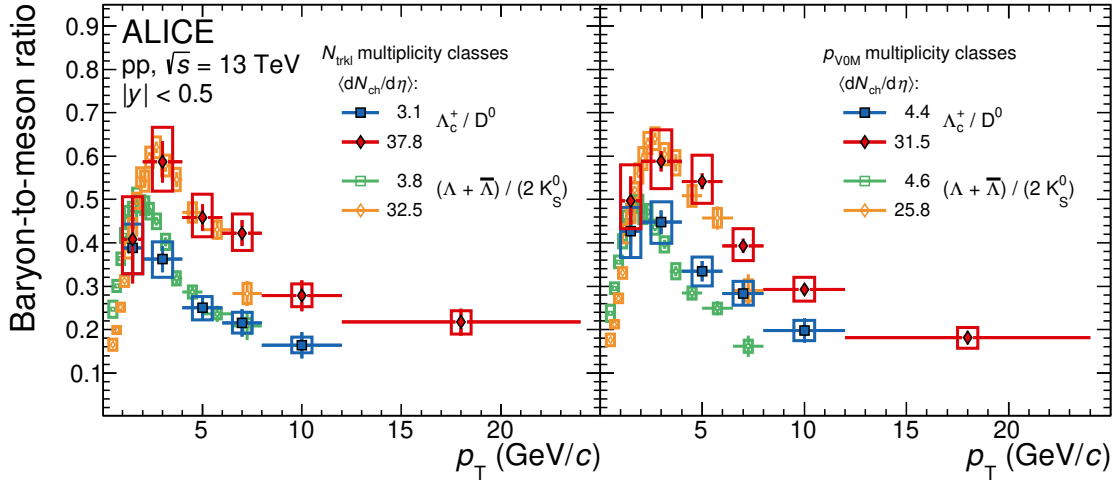


Figure 4.25: The baryon-to-meson ratios in the light-flavour, based on measurements from Ref. [201] and charm sector measured in pp collisions at $\sqrt{s} = 13$ TeV for similar low- and high-multiplicity classes at mid- (left) and forward (right) rapidity.

spectrum integrated in the full p_T range to the integral in the visible p_T range. The Λ_c^+ and D^0 yields in the full p_T range were obtained by integrating the yield in the visible p_T interval and scaling by the extrapolation factor. The fraction of extrapolated yield from the lowest to the highest multiplicity interval is about 39% (31%), 28% (22%), 20% (16%), and 15% (13%) for Λ_c^+ (D^0). The procedure was repeated considering also the CR-BLC Mode 0 and Mode 3 as well as two different functions fitted to the spectra (a Tsallis-Lévy [218] and a power-law function). The fits were performed considering the statistical and p_T -uncorrelated sources of systematic uncertainties, and also shifting up and down the data by one sigma of the p_T -correlated systematic uncertainties. The envelope of the extrapolation factors obtained with all the trials was assigned as the extrapolation uncertainty on Λ_c^+ and D^0 , and it was propagated to the Λ_c^+/D^0 ratio, resulting in a value that ranges from 2% to 21% depending on multiplicity. The same procedure was used to estimate the p_T -integrated D_s^+ yields and D_s^+/D^0 yield ratios in the different multiplicity intervals, reported in Ref. [219]. The Λ_c^+ and D^0 p_T -integrated yields are also reported in Ref. [219], together with the p_T -integrated Λ_c^+/D^0 yield ratios in the visible p_T range, and the tables with the numerical values of the p_T -integrated ratios. The p_T -integrated Λ_c^+/D^0 yield ratio as a function of $\langle dN_{ch}/d\eta \rangle$ is shown in Fig. 4.26, where the systematic uncertainties from the extrapolation (shaded boxes, assumed to be uncorrelated among multiplicity intervals) are drawn separately from the other sources of systematic uncertainties (empty boxes). The sources related to the raw-yield extraction, the multiplicity-interval limits, the high-multiplicity triggers, the multiplicity-independent prompt fraction assumption, and the statistical uncertainties on the efficiencies are also considered uncorrelated with multiplicity. The other systematic uncertainties are assumed to be correlated. The measurements performed in pp and p-Pb

collisions at $\sqrt{s} = 5.02$ TeV [213] are also shown. The result does not favour an increase of the yield ratios with multiplicity and the trend is compatible with a constant function. The same trend was also observed for the Λ/K_S^0 ratio in Ref. [201]. This suggests that the increasing trend observed for the $1 < p_T < 24$ GeV/ c range comes from a re-distribution of p_T that acts differently for baryons and mesons, while this is not observed in the meson-to-meson ratios, as shown in Fig. 4.24 for D_s^+/D^0 and in Ref. [200] for K/π . The results are compared to the p_T -integrated PYTHIA predictions. The measurements exclude the Monash prediction in the whole multiplicity range, and tend to be significantly below the CR-BLC Mode 2 for the three highest multiplicity intervals.

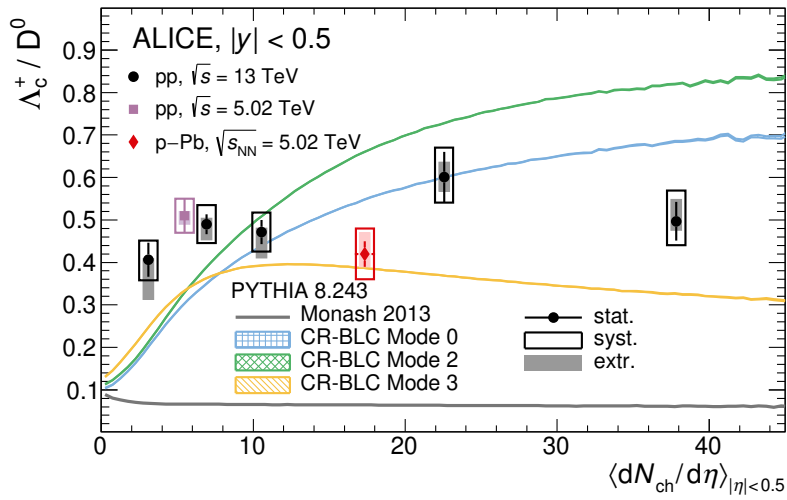


Figure 4.26: Ratios of p_T -integrated yields of Λ_c^+ and D^0 hadrons as a function of $\langle dN_{ch}/d\eta \rangle$ in pp collisions at $\sqrt{s} = 13$ TeV. Measurements performed in pp and p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV from Ref. [213] are also shown. Statistical and systematic uncertainties are shown by error bars and empty boxes, respectively. Shaded boxes represent the extrapolation uncertainties. The corresponding PYTHIA predictions [131, 137] are also shown.



4.2 Non-prompt D-meson fraction as a function of multiplicity in pp collisions at $\sqrt{s} = 13$ TeV

In Sec. 4.1, a multiplicity dependence of the transverse momentum (p_T) differential Λ_c^+/D^0 yield ratio be observed in pp collisions at $\sqrt{s} = 13$ TeV. The same quantity with p_T integrated was found not sensitive to the charged-particle multiplicity as shown in Fig. 4.26. No modification of the D_s^+/D^0 yield ratio as increasing multiplicity was observed in pp collisions [141, 220]. Conversely, in the beauty sector, the LHCb Collaboration has observed that the production ratio of B_s^0 to B^0 increases with the charged-particle multiplicity, when estimated using tracks in the same pseudorapidity interval as the B mesons [221]. However, there are no measurements available for beauty-baryon production as a function of charged-particle multiplicity. Finally, the fraction of $\chi_{c1}(3872)$ and $\psi(2S)$ states promptly produced at the collision vertex was found by the LHCb Collaboration to decrease as charged-particle multiplicity increases [222]. This suppression is interpreted as a consequence of the heavy-quark breakup via interactions with comoving hadrons [223, 224].

In this section, the first measurement of the fraction of D^0 and D^+ mesons originating from beauty-hadron decays ($f_{\text{non-prompt}}$) at midrapidity ($|\eta| < 0.5$) is reported as a function of the charged-particle multiplicity in pp collisions at $\sqrt{s} = 13$ TeV. In addition, the ratio between the fraction measured in different multiplicity classes divided by the one measured in the multiplicity-integrated sample is presented. Event selection and multiplicity definition is presented in in Section 4.2.1. The candidate selection and raw yield extraction are described in Section 4.2.2. The measurement of $f_{\text{non-prompt}}$ in six transverse momentum intervals and integrated in $1 < p_T < 24$ GeV/ c is described in Section 4.2.3, while the evaluation of the systematic uncertainties is discussed in Section 4.2.4. Finally, the results are presented and compared to model calculations in Section 4.2.5.

4.2.1 Event selection and multiplicity definition

The data used for this analysis are the same as previous analysis described in Sec. 4.1. Similarly, a minimum-bias (MB) trigger was used, based on coincident signals in V0A and V0C. To enrich the data sample in the highest multiplicity regions, a high-multiplicity trigger based on a minimum threshold for the V0 amplitudes (HMOV0) was used as well. The data sample collected with such a trigger corresponds to the 0.17% highest-multiplicity events out of all inelastic collisions with at least one charged particle in the pseudorapidity range $|\eta| < 1$ (denoted as INEL > 0). Offline selections were applied to remove background from beam–gas collisions, as described in Ref. [204]. Events with multiple reconstructed primary vertices were rejected. The remaining pile-up events were at a percent level and, therefore, did not affect the present analysis. Only the events with a primary vertex reconstructed within $|z_{\text{vtx}}| <$



10 cm from the nominal interaction point along the beam-line direction were selected for the analysis. To select events in the $\text{INEL} > 0$ class, at least one track segment reconstructed with the first two ITS layers (denoted as tracklet) within the pseudorapidity region $|\eta| < 1$ was required. After these selections, the integrated luminosities are about $\mathcal{L}_{\text{int}} \approx 32 \text{ nb}^{-1}$ for the MB triggered events, and $\mathcal{L}_{\text{int}} \approx 7.7 \text{ pb}^{-1}$ for the HVM0 triggered events [220]. The event multiplicity was determined in the forward rapidity region, exploiting the sum of signal amplitudes in the V0A and V0C scintillators, V0M, and defining its percentile distribution, p_{V0M} . Low p_{V0M} values represent high-multiplicity events. The definition of the mean multiplicity density ($\langle dN_{\text{ch}}/d\eta \rangle_{|\eta| < 0.5}$) of charged-primary particles at midrapidity is given in Ref. [225]. It was obtained by converting the measured event multiplicities as described in Ref. [204]. Tab. 4.1 summarises the multiplicity event classes at forward rapidity used in this analysis ($p_{\text{V0M}}(\%)$) and the corresponding values for $\langle dN_{\text{ch}}/d\eta \rangle_{|\eta| < 0.5}$, together with the corresponding value for the multiplicity-integrated class [204].

4.2.2 Candidate selection and raw yield extraction

The D^0 mesons and their charge conjugates were reconstructed via the hadronic decay channels $D^0 \rightarrow K^- \pi^+$, with branching ratio $\text{BR} = (3.947 \pm 0.030)\%$ [178]. D-meson candidates were built by combining pairs or triplets of tracks with the proper charge signs, each track with $p_T > 0.3 \text{ GeV}/c$, $|y| > 0.8$, at least 70 out of a maximum of 159 crossed TPC pad rows, a minimum number of two hits (out of six) in the ITS, with at least one in either of the two innermost layers, and a track fit quality $\chi^2/\text{ndf} < 2$ in the TPC. These track-selection criteria reduce the D-meson acceptance in rapidity, which falls steeply to zero for $|y| > 0.5$ at low p_T and for $|y| > 0.8$ at $p_T > 5 \text{ GeV}/c$. Thus, a fiducial acceptance selection $|y| < y_{\text{fid}}(p_T)$, was applied to grant a uniform acceptance inside the rapidity range considered. The $y_{\text{fid}}(p_T)$ value was defined as a second-order polynomial function, increasing from 0.5 to 0.8 in the transverse-momentum range $0 < p_T < 5 \text{ GeV}/c$, and as a constant term, $y_{\text{fid}} = 0.8$, for $p_T > 5 \text{ GeV}/c$ [63].

As discussed in Sec. 3.1, to suppress the large combinatorial background and to separate at the same time the contribution of prompt and non-prompt D mesons, a machine-learning approach with multi-class classification, based on Boosted Decision Trees (BDT) provided by the XGBoost [182, 226] library was adopted. Signal samples of prompt and non-prompt D mesons for the BDT training were obtained from PYTHIA 8 simulations as described in Sec. ???. The background samples were obtained from candidates in the sideband region in the data, i.e. in the interval $5\sigma < |\Delta M| < 9\sigma$ of the invariant mass distribution, where ΔM is the deviation between the invariant mass of the candidate and the mean of a Gaussian function describing the signal peak and σ is the Gaussian width. The training procedures



are the same as reported in Ref. [63]. Before the training, loose kinematic and topological selections were applied to the D-meson candidates. The D-meson candidate information used for training the BDT models was mainly based on the displacement of the tracks from the primary vertex, the impact parameter of the D-meson daughter tracks, the distance between the D-meson decay vertex and the primary vertex, the cosine of the pointing angle between the D-meson candidate line of flight (the vector connecting the primary and secondary vertices) and its reconstructed momentum vector, and the PID information of the decay tracks. The training variables used are reported in Tab. 4.10 and their distributions for signal and background are shown in Fig. 4.27 for the D^0 mesons in the $1 < p_T < 2$ GeV/ c interval.

Table 4.10: Training variables used for the D^0 mesons.

Variable	D^0
Decay length	x
Decay length xy	✓
Normalized decay length xy	✓
$\cos \theta_p$	x
$\cos \theta_p^{xy}$	✓
σ_{vertex}	x
$\max d_0 - d_0^{\text{exp}} ^{\text{prong}}(n\sigma)$	✓
$n\sigma_{\text{comb}}^{\pi}(0)$	✓
$n\sigma_{\text{comb}}^{\pi}(1)$	✓
$n\sigma_{\text{comb}}^{\pi}(2)$	x
$n\sigma_{\text{comb}}^K(0)$	✓
$n\sigma_{\text{comb}}^K(1)$	✓
$n\sigma_{\text{comb}}^K(2)$	x

The linear correlations between the variables used in the training as well as the invariant mass of the D-meson candidate and its transverse momentum, which are not employed in the training, are reported for the three classes of candidates in Fig. 4.28 for the D^0 mesons.

After the training, the model is applied to the test and training sets to visualise the distributions of the ML output scores obtained and to check their agreement between the two sets. In particular, in case of classification with three classes of candidates, three scores are provided by the ML algorithm, which represent the probability of belonging to each of the considered classes (prompt D mesons, non-prompt D mesons, combinatorial background). By construction, for each candidate, the probabilities of belonging to the three classes sum to one. Fig. 4.29 shows an example of distributions for D^0 meson in the $2 < p_T < 4$ GeV/ c intervals. For each probability, the distribution corresponding to the respective true

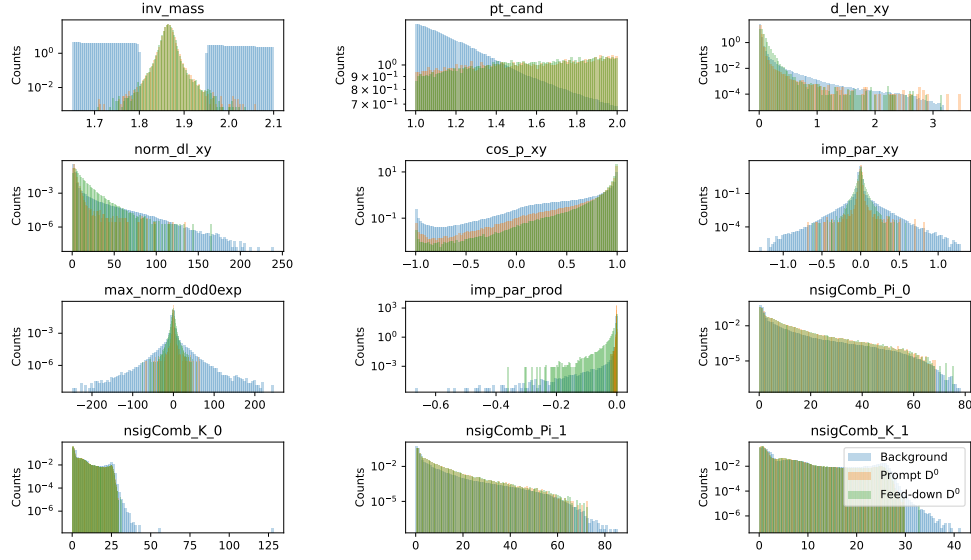


Figure 4.27: Distributions for prompt D^0 -meson, non-prompt D^0 -meson, and combinatorial-background candidate variables in the $1 < p_T < 2 \text{ GeV}/c$ interval.

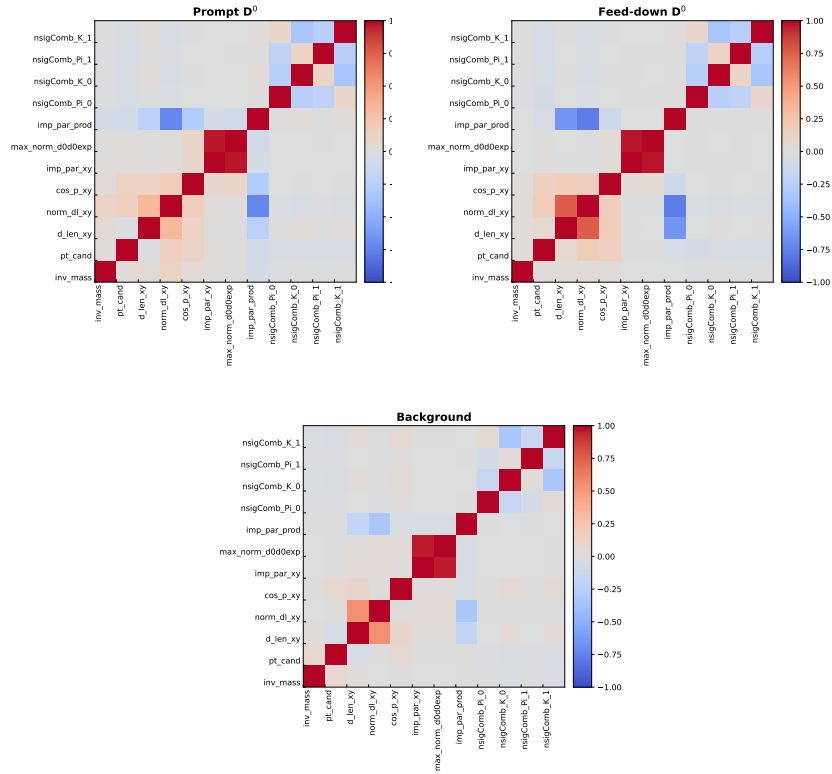


Figure 4.28: Linear correlations between the training variables, the invariant mass and the p_T , for prompt D^0 mesons (top-left panel) and non-prompt D^0 mesons (top-right panel) and combinatorial background (bottom panel) in the $1 < p_T < 2 \text{ GeV}/c$ interval. The colours, from red (fully correlated) to blue (fully anti-correlated), indicate the correlation level.

class is close to unity, while the other two distributions are shifted towards zero. Moreover, it is possible to observe that the distributions of combinatorial background candidates are typically more separated from those of prompt and non-prompt D mesons, indicating that the separation between combinatorial background and signal is more efficient than the separation between prompt and non-prompt D mesons, as expected.

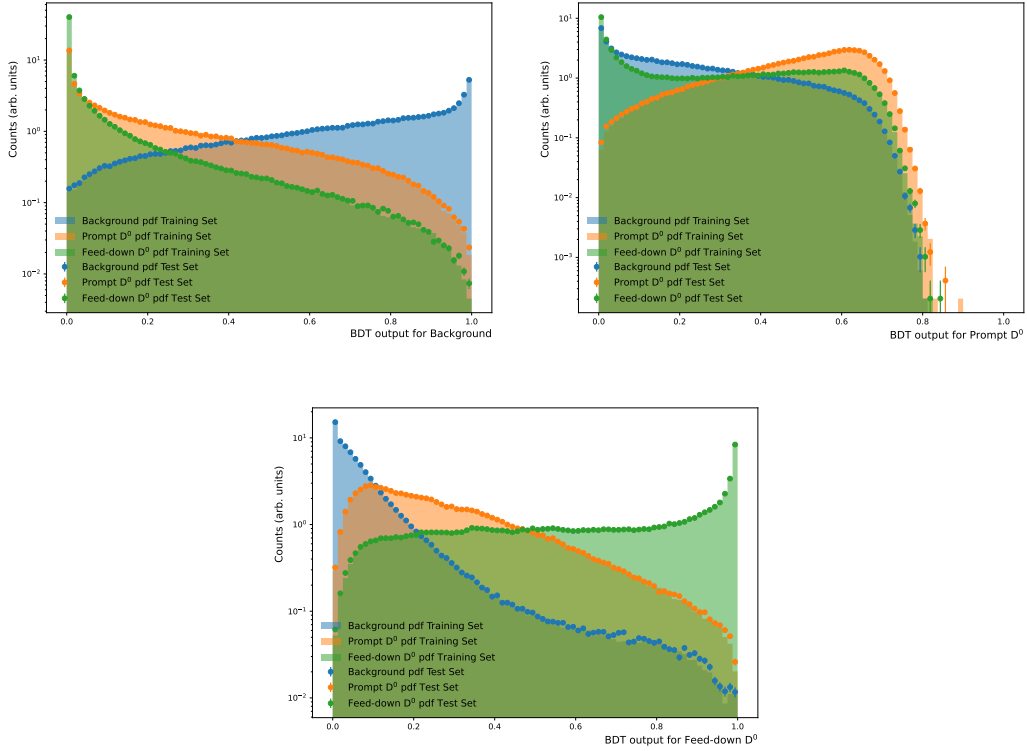


Figure 4.29: Distributions of ML output scores for combinatorial background candidates, prompt D^0 mesons, and non-prompt D^0 mesons in the $2 < p_T < 4$ GeV/ c interval.

To assess the performance of the model, the area under the Receiver Operating Characteristic curve (ROC AUC) is considered as discussed in Sec. 3.1.3. The ROC curves and the values of ROC AUC, obtained on test set, for the models trained for D^0 mesons in the $2 < p_T < 4$ GeV/ c and $8 < p_T < 12$ GeV/ c intervals are reported in Fig. 4.30. It is possible to observe that for both D^0 meson, the ROC AUC scores are higher at high p_T , as expected from the larger Lorentz Boost of high p_T particles, which enlarges the displacement of prompt and non-prompt D mesons, amplifying the differences among the classes of candidates. Independent BDTs were trained for each D -meson species and p_T interval in the multiplicity-integrated sample. Subsequently, the BDTs were applied to the real data sample in which the type of candidate is unknown. The BDT outputs are related to the candidate probability to be a



non-prompt D meson or combinatorial background. Selections on the BDT outputs were optimised to obtain a high non-prompt D-meson fraction while maintaining a reliable signal extraction (with statistical significance larger than 5).

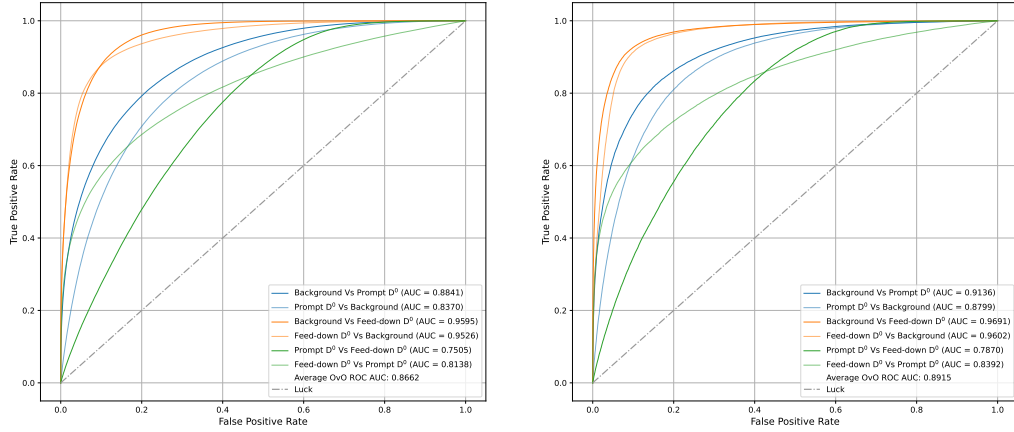


Figure 4.30: ROC curves for the models trained for D^0 mesons in the $2 < p_T < 4$ GeV/c (left panel) and $8 < p_T < 12$ GeV/c (right panel) intervals.

The signal extraction was performed in each p_T and multiplicity interval via a binned maximum-likelihood fit to the candidate invariant-mass distribution. The raw yields could be extracted in the transverse momentum interval $1 < p_T < 24$ GeV/c and in six subranges, for both D^0 and D^+ mesons. A Gaussian function and an exponential function were used to describe the signal peak and the background distribution, respectively. To improve the stability of the fits, the widths of the D-meson signal peaks were fixed to the values extracted from data samples dominated by prompt candidates, given the naturally higher abundance of prompt compared to non-prompt D mesons. In addition, for the D^0 meson, the contribution of signal candidates to the invariant-mass distribution with the wrong decay-particle mass assignment (reflections) was included in the fit. It was parameterised by fitting the invariant-mass distribution of reflections with a double Gaussian function, and normalised according to the reflection-to-signal ratio from the PYTHIA 8 simulations. The contribution of reflections to the raw yield is about 0.5%–4%, increasing with increasing p_T . Examples of invariant-mass fits with different contribution of signal from beauty-hadron decays in the $2 < p_T < 4$ GeV/c interval for the lowest multiplicity class and in the $1 < p_T < 24$ GeV/c interval for the highest multiplicity class are shown in Fig. 4.31 for D^0 . Based on the selections on the BDT outputs, samples dominated by non-prompt (prompt) candidates were selected by requiring low probability for a candidate to be combinatorial background and a

high (low) probability to be non-prompt. The invariant-mass fits from non-prompt (prompt) enhanced samples are shown in each right (left) panel, indicating the corresponding selection applied on the BDT output score related to the probability to be a non-prompt D meson.

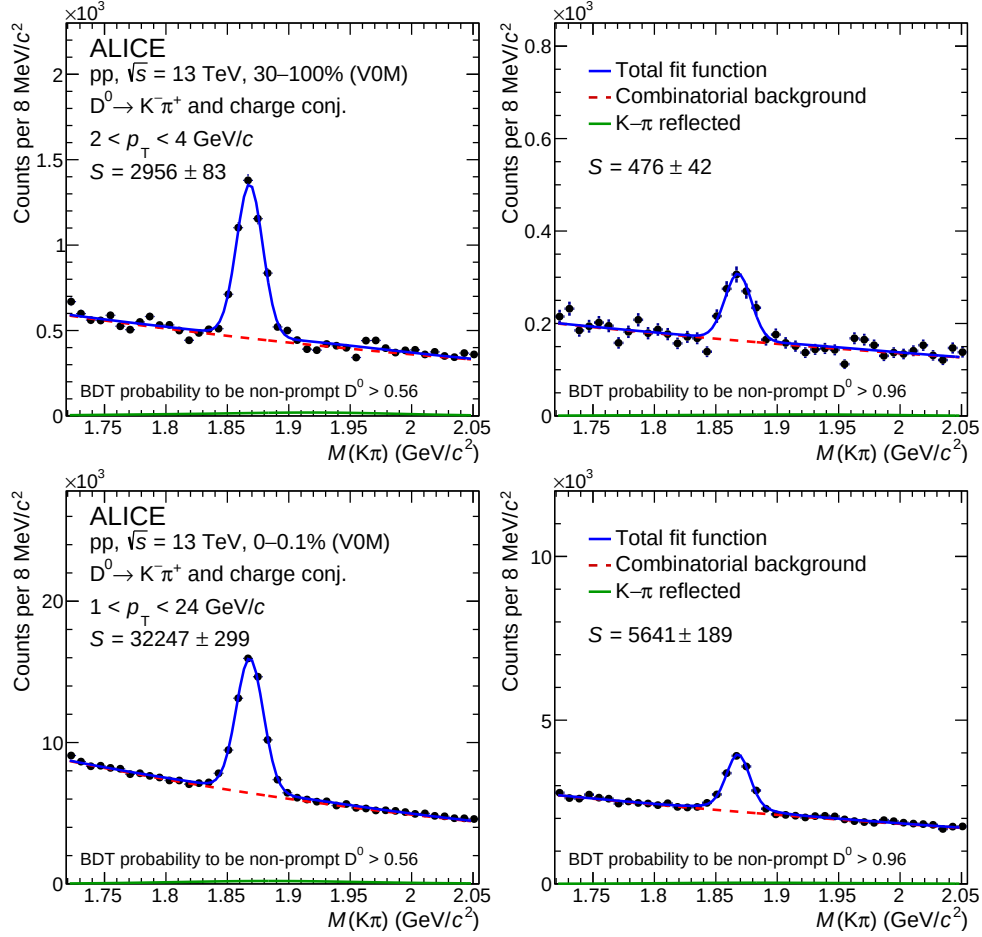


Figure 4.31: Invariant-mass distribution of D^0 candidates and their charge conjugates in selected p_T and multiplicity intervals. The blue solid curves show the total fit function and the red dashed curves show the combinatorial-background contribution. The green solid lines represent the reflection contribution. The raw-yield (S) values are reported together with their statistical uncertainties resulting from the fit. Top row: D^0 mesons in the $2 < p_T < 4$ GeV/ c interval for the low multiplicity class. Bottom row: D^0 mesons in the $1 < p_T < 24$ GeV/ c interval for the high multiplicity class. The corresponding BDT probability minimum threshold for the candidate selection is reported. The left (right) column corresponds to the prompt (non-prompt) D^0 meson candidates dominated sample.



4.2.3 Non-prompt fraction extraction

As discussed in Sec. 3.2, In each p_T and multiplicity interval, the fraction of non-prompt D mesons, $f_{\text{non-prompt}}$, was estimated by sampling the raw yield with different BDT selections related to the candidate probability of being a non-prompt D meson. In this way, a set of raw yields Y_i (index i refers to a given selection on the BDT output) with different contributions from prompt and non-prompt D mesons was obtained. These raw yields are related to the corrected yields of prompt (N_{prompt}) and non-prompt ($N_{\text{non-prompt}}$) D mesons via the acceptance-times-efficiency ($\text{Acc} \times \epsilon$) factors as

$$(\text{Acc} \times \epsilon)_i^{\text{prompt}} \times N_{\text{prompt}} + (\text{Acc} \times \epsilon)_i^{\text{non-prompt}} \times N_{\text{non-prompt}} - Y_i = \delta_i. \quad (4.4)$$

In the above equation, the δ_i term represents a residual originating from the uncertainties on Y_i , $(\text{Acc} \times \epsilon)_i^{\text{non-prompt}}$, and $(\text{Acc} \times \epsilon)_i^{\text{prompt}}$. The $\text{Acc} \times \epsilon$ factors were obtained from MC simulations as described in Sec. 4.1.4.1. They are different for prompt and non-prompt D mesons due to the different decay topology. Since the resolution of the reconstructed primary vertex depends on the multiplicity, the simulated events were weighted to reproduce the charged-particle multiplicity distribution measured in data for events that contain D-meson candidates having an invariant mass compatible with the one of the signal. After that, the $\text{Acc} \times \epsilon$ factors were computed for each BDT selection for prompt and non-prompt D mesons within the fiducial acceptance region. In the case of the number of sets of selections $n \geq 2$, a χ^2 function can be defined based on the set of equations of Eq. 4.4, which can be minimised to obtain N_{prompt} and $N_{\text{non-prompt}}$. More details can be found in Ref. [63]. The $N_{\text{non-prompt}}$ and N_{prompt} values can be used to calculate the corrected fraction of non-prompt D mesons as follows

$$f_{\text{non-prompt}} = \frac{N_{\text{non-prompt}}}{N_{\text{non-prompt}} + N_{\text{prompt}}}. \quad (4.5)$$

In addition, the ratio between the fraction of non-prompt D mesons measured in each multiplicity interval and the one measured in the $\text{INEL} > 0$ class of events, $f_{\text{non-prompt}}^{\text{mult}}/f_{\text{non-prompt}}^{\text{INEL}>0}$, was computed in multiplicity and p_T intervals in order to investigate the modification of the non-prompt fraction with respect to the one measured in the multiplicity-integrated sample.

Fig. 4.32 shows an example of the raw-yield distribution as a function of the BDT-based selection used in the χ^2 -minimisation procedure for D^0 meson in the transverse-momentum intervals $2 < p_T < 4 \text{ GeV}/c$ and $1 < p_T < 24 \text{ GeV}/c$ for the low-multiplicity and high-multiplicity classes of events, respectively. The raw yield decreases with increasing minimum threshold for the probability to be a non-prompt D meson, corresponding to an increasing non-prompt D fraction. Note also that the raw yields used in this procedure are largely correlated among each other, implying that adjacent data points

are expected to fluctuate in the same direction. The prompt and non-prompt components of the raw yields for each BDT-based selection obtained from the χ^2 -minimisation procedure as $(\text{Acc} \times \epsilon)_i^{\text{prompt}} \times N_{\text{prompt}}$ and $(\text{Acc} \times \epsilon)_i^{\text{non-prompt}} \times N_{\text{non-prompt}}$, are reported as the red and blue distributions, and their sum is represented by the green histogram.

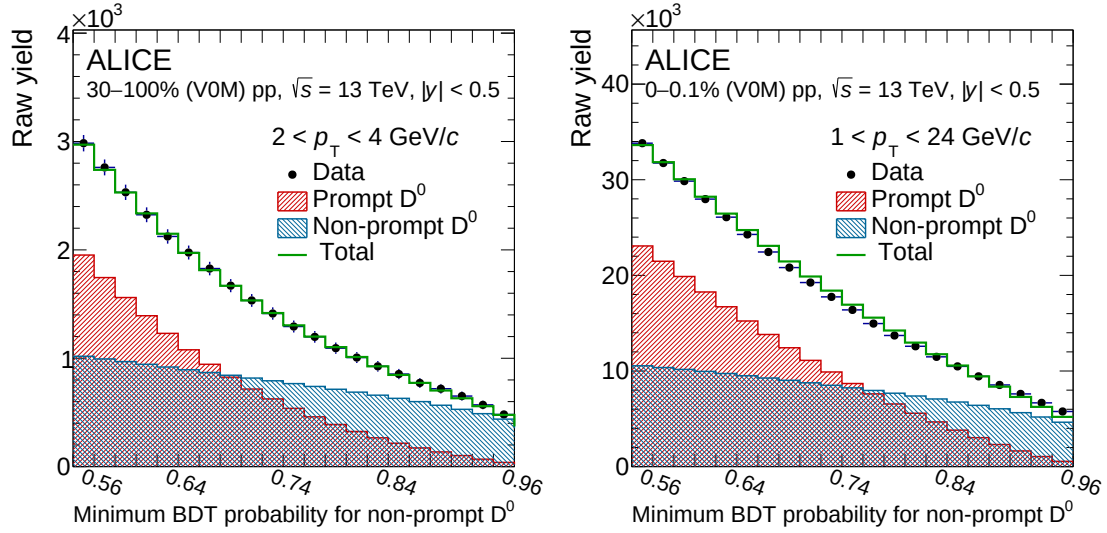


Figure 4.32: Examples of raw-yield distribution as a function of the BDT-based selection employed in the χ^2 -minimisation procedure adopted for the determination of $f_{\text{non-prompt}}$ of D mesons. D^0 mesons in low multiplicity (left) and high multiplicity (right) classes.

4.2.4 Systematic uncertainties

The values of systematic uncertainty on the non-prompt D -meson fraction were estimated with procedures similar to previous analysis in Refs. [63, 220]. They include the uncertainties on (i) the raw-yield extraction from the invariant-mass distributions; (ii) the selection efficiency estimation; (iii) the dependency of the efficiency on the charged-particle multiplicity; and (iv) the D -meson p_T shape in the simulation.

4.2.4.1 Raw-yield extraction

The systematic uncertainty of the raw-yield extraction was evaluated by repeating the fits to the invariant-mass distribution varying the fit range and the functional form of the background and signal fit functions. To further test the sensitivity to the line shape of the signal, a bin-counting method, in which the signal yield was obtained by integrating the background-subtracted invariant-mass distribution within the $\pm 3\sigma$ region relative to the peak position, was used. In the case of D^0 mesons, an additional contribution due to signal reflections in the invariant-mass distribution was estimated by varying the normalisation and the shape of the templates used for the reflections in the invariant-mass fits. The outcome of this procedure is shown for each multiplicity interval of the analysis in Figs. 4.33 for the D^0



meson. Then the final systematic uncertainty on fraction obtained from distribution of trials (considering sum in quadrature of RMS and shift with respect to the default value) show in Figs. 4.34. The sum in quadrature of RMS and shift for f_{prompt} in each multiplicity bin is shown in the upper left panel as a function of p_T , $f_{\text{non-prompt}}$ in the upper right panel. In the bottom left panel, the ratio between the f_{prompt} of multiplicity intervals and integrated multiplicity bin is shown as a function of the p_T are shown. The $f_{\text{non-prompt}}$ case is shown in the bottom right panel.

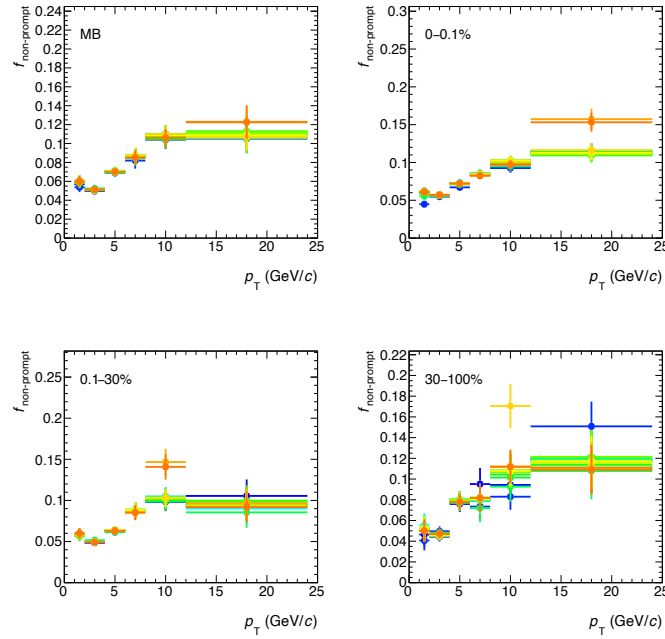


Figure 4.33: Corrected $f_{\text{non-prompt}}$ as a function of p_T for each multiplicity interval for D^0 meson obtained varying the fitting configurations. The multiplicity classes are MB (top left), 0-0.1% (top right), 0.1-30% (bottom left) and 30-100% (bottom right). Each color corresponds to a different variation.

4.2.4.2 Selection efficiency estimation

The systematic uncertainty of the selection-efficiency determination, arising from possible imperfections of the description of the decay topologies or the detector resolution in the simulation, was estimated by using alternative sets of BDT-output selections for the procedure described in Section 4.2.3. In particular, stricter and looser selections were tested, as well as different combinations of selections adopted to define the system of equations described in Eq. 4.4. The f_{prompt} and $f_{\text{non-prompt}}$ values obtained for all the variations for each multiplicity interval are reported in Figs. ?? for the D^0 meson. The systematic uncertainty has been assigned considering the RMS and shift of the trials with respect to the central value similarly to what was done for the systematic uncertainty on the raw-yield extraction described previously. The results are reported in Fig. 4.36 for the D^0 meson.

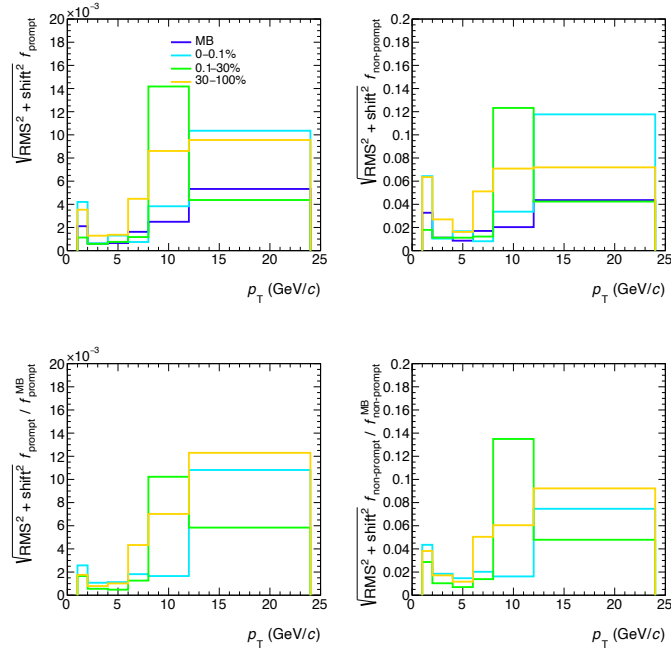


Figure 4.34: Sum in quadrature of RMS and shift with respect to the default value of the corrected f_{prompt} , $f_{\text{non-prompt}}$, and their ratio with respect to the MB obtained by varying the raw-yield extraction configurations for the D^0 meson.

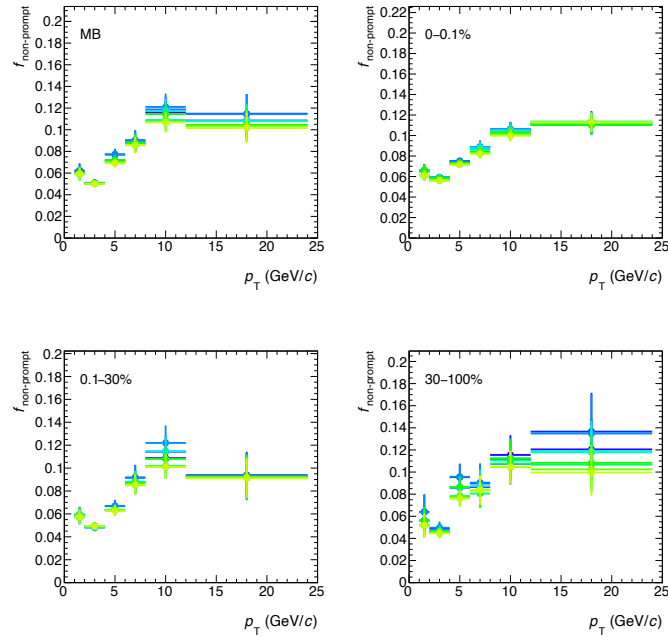


Figure 4.35: Corrected $f_{\text{non-prompt}}$ as a function of p_T for each multiplicity interval for D^0 mesons obtained varying the sets of selections in the cut-variation method. The multiplicity classes are MB (top left), 0-0.1% (top right), 0.1-30% (bottom left) and 30-100% (bottom right). Each color corresponds to a different variation.

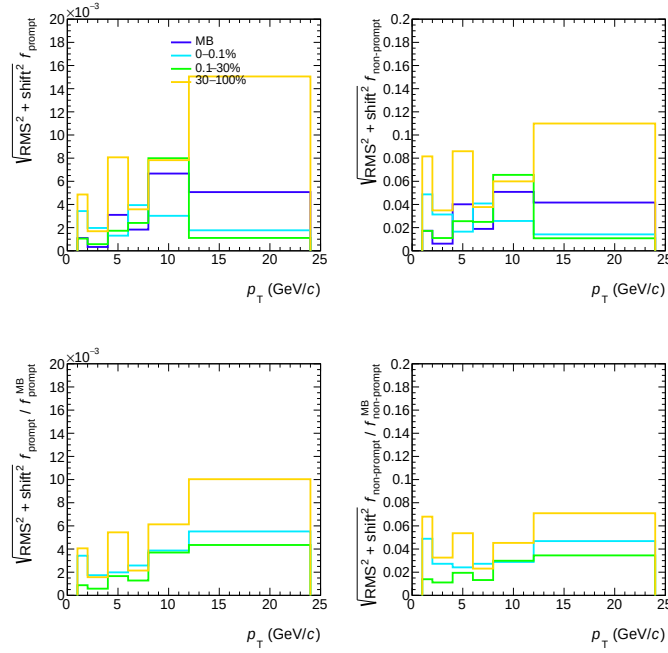


Figure 4.36: Sum in quadrature of RMS and shift with respect to the default value of the corrected f_{prompt} , $f_{\text{non-prompt}}$, and their ratio with respect to the MB obtained by varying the sets of selections in the cut-variation method for the D^0 meson.

4.2.4.3 Multiplicity distribution in the MC simulation

To estimate the systematic uncertainty on the sensitivity of the efficiency on the charged-particle multiplicity, due to the multiplicity dependence of the primary-vertex reconstruction resolution, the distribution of the number of tracklets in the MC simulation for each V0M class of events was weighted using the one obtained in the real data considering events containing a D-meson candidate, without requiring the invariant-mass region selection. As an example with lowest multiplicity, the effect on the $f_{\text{non-prompt}}$ and $f_{\text{non-prompt}}^{30-100\%} / f_{\text{non-prompt}}^{\text{INEL}>0}$ of D^0 meson are shown in Figs. 4.37.

4.2.4.4 MC generated p_T shape

The systematic uncertainty on the efficiency calculation due to a possible difference between the real and simulated D-meson transverse-momentum distributions was estimated by evaluating the efficiency after reweighting the p_T shape from the PYTHIA 8 generator to match the one from FONLL calculations, in addition to the reweighting of the multiplicity distribution mentioned above. 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. Figs. 4.38 shows an example for the effect on the $f_{\text{non-prompt}}$ and $f_{\text{non-prompt}}^{30-100\%} / f_{\text{non-prompt}}^{\text{INEL}>0}$ of D^0 meson in lowest multiplicity bin.

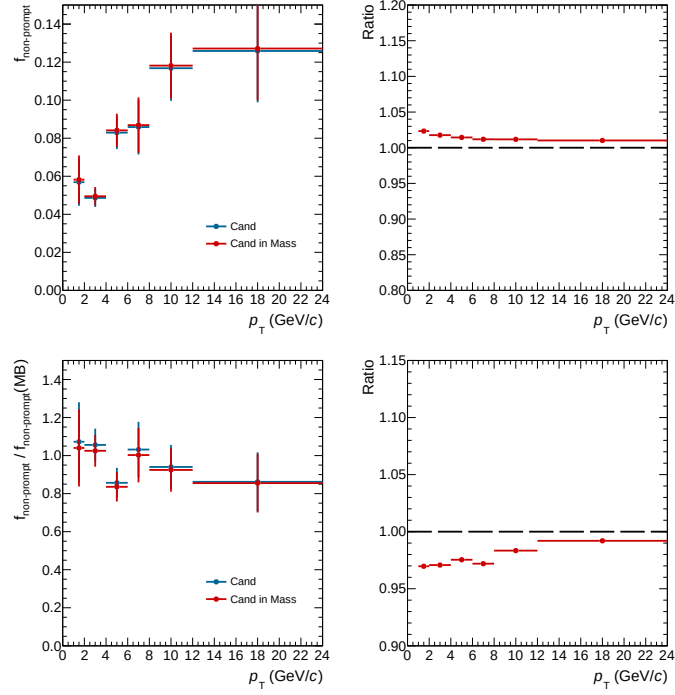


Figure 4.37: Comparison of the $f_{\text{non-prompt}}$ (top panel) and $f_{\text{non-prompt}}^{30-100\%} / f_{\text{non-prompt}}^{\text{INEL}>0}$ (bottom panel) of D^0 mesons obtained with the different multiplicity weights in the 30-100% multiplicity class.

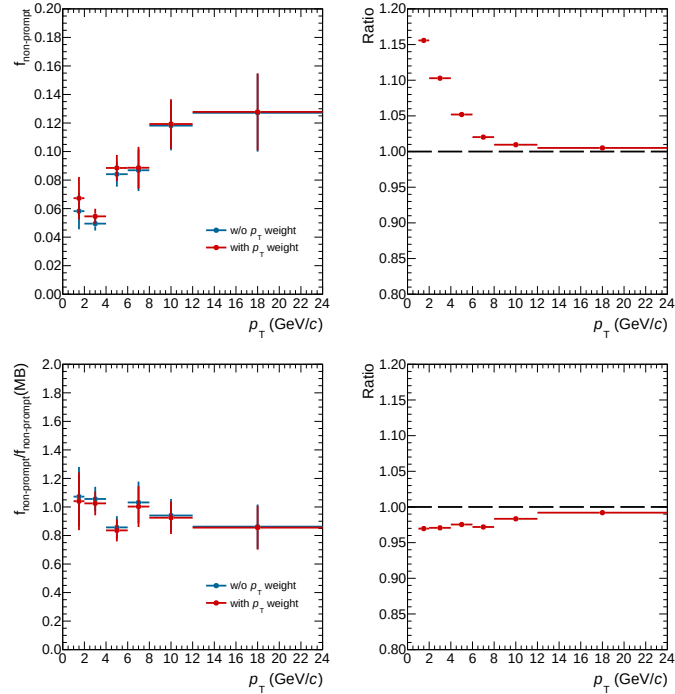


Figure 4.38: Comparison of $f_{\text{non-prompt}} / f_{\text{non-prompt}}^{\text{MB}}$ for D^0 mesons obtained with the different p_T weights in the 30-100% multiplicity class.



4.2.5 Results and discussion

The measured fractions of D-mesons originating from beauty-hadron decays, $f_{\text{non-prompt}}$, in pp collisions at $\sqrt{s} = 13$ TeV are shown in Fig. 4.39 as a function of p_T . The results are reported in different panels for D^0 (left) and D^+ (right) mesons and for the $\text{INEL} > 0$ class (top panels) and the three multiplicity classes of events (lower panels). The statistical and total systematic uncertainties are shown by vertical error bars and boxes, respectively. In all the event classes and for both D^0 and D^+ mesons, $f_{\text{non-prompt}}$ increases with p_T from 5%–7% to about 10%. This increase is motivated by the harder p_T distribution of beauty hadrons compared to the charm ones, which is only partly compensated by the $b \rightarrow D + X$ decay kinematics [48, 63]. The fraction of non-prompt D^0 mesons is slightly larger than that of D^+ mesons, as a consequence of the different branching ratios of B mesons with a D^0 or D^+ meson in the final state, and of the different charm-quark fragmentation fractions for the prompt D-meson production. This increasing trend is expected from pQCD calculations, as shown in Ref. [62]. Measurements are compared to predictions from the PYTHIA 8 [173, 228] and EPOS 3 [227, 229] event generators. PYTHIA 8 simulations were obtained using the standard Monash 2013 tune [131] as well as with colour reconnection settings beyond-leading-colour approximation (CR-BLC Mode 0, 2, and 3) [137], and with colour ropes [228] using PYTHIA version 8.307. Both the version 3 and 4.0.0 of the EPOS 3 MC generator were tested. In EPOS 4, parallel partonic scatterings based on the S -matrix theory are implemented, leading to the factorisation of the hard and soft scales, particularly important for heavy quarks. This factorisation allows the computation of the PDFs within the EPOS framework itself. In both EPOS 3 versions, the hadronisation is implemented via a microcanonical ensemble, consisting in the decay of plasma droplets which conserves energy, momentum, and flavour. The EPOS 3 predictions presented in this paper do not include a hydrodynamic expansion of the system. However, the results were found not to significantly change if the latter is included. Following what was done for data, all PYTHIA 8 and EPOS 3 simulations were selected according to percentiles of the $\text{INEL} > 0$ cross section based on the charged-particle multiplicity counts in the ALICE V0A and V0C acceptance. While all models qualitatively reproduce the increase of $f_{\text{non-prompt}}$ with increasing p_T , EPOS 3 significantly underpredicts $f_{\text{non-prompt}}$ of D^+ mesons and D^0 mesons in the $\text{INEL} > 0$ and in the lowest multiplicity classes of events, by up to a factor of two. Moreover, EPOS 3 predicts a slightly stronger multiplicity dependence compared to EPOS 4. On the other hand, PYTHIA 8 is generally closer to data but overpredicts $f_{\text{non-prompt}}$ by approximately 20–30%. No significant difference in the various PYTHIA 8 settings tested in this work is observed, with the exception of the ‘CR-BLC Mode 3’ setting (see Ref. [137] for a detailed description), which predicts a lower D^+ and D^0 non-prompt fraction especially in the two highest multiplicity intervals, providing a better description of the data.

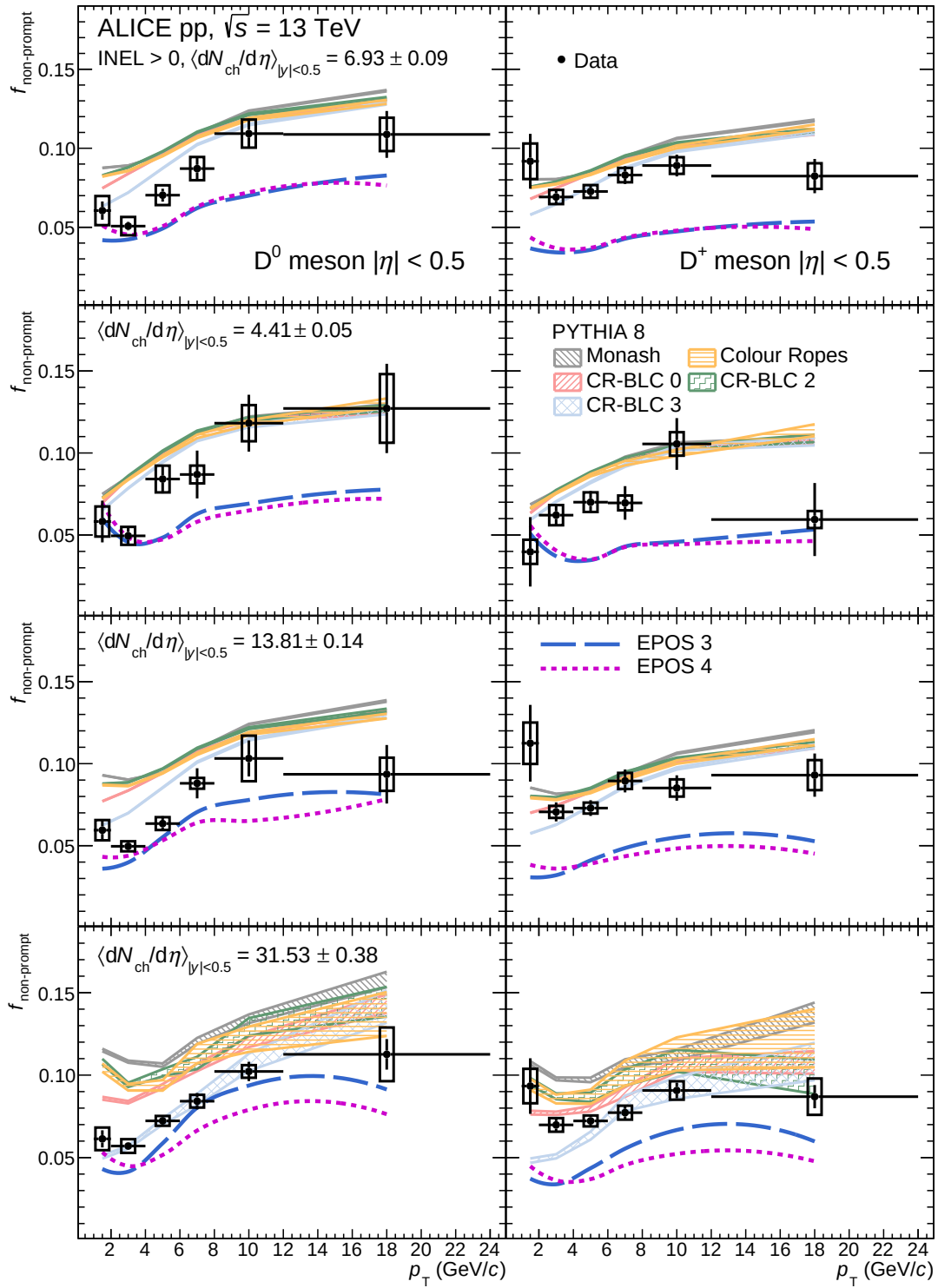


Figure 4.39: Fractions of non-prompt D^0 (left column) and D^+ (right column) mesons as a function of p_T for the $\text{INEL} > 0$ class and the three multiplicity classes of events in pp collisions at $\sqrt{s} = 13$ TeV. The measurements are compared with the predictions obtained with PYTHIA 8 [137] and EPOS 3 [227] event generators.



The ratio of the D-meson non-prompt fractions in the multiplicity classes relative to that in the $\text{INEL} > 0$ class, $f_{\text{non-prompt}}^{\text{mult}}/f_{\text{non-prompt}}^{\text{INEL}>0}$, is shown in Fig. 4.40 as a function of transverse momentum for the three multiplicity classes. This double ratio isolates the relative variation of $f_{\text{non-prompt}}$ as a function of the charged particle multiplicity from absolute scaling factors. The double ratio of D^0 and D^+ was found to be compatible for all the multiplicity classes as expected. In order to improve the statistical precision, the average $f_{\text{non-prompt}}^{\text{mult}}/f_{\text{non-prompt}}^{\text{INEL}>0}$ of D^0 and D^+ was computed. The average was computed using the inverse of the quadratic sum of the relative statistical and uncorrelated systematic uncertainties as weights. The systematic uncertainties were propagated through the averaging procedure considering the contributions from the raw-yield extraction and the selection efficiency as uncorrelated, while the other sources as fully correlated between the two D-meson species. In all multiplicity classes, the measured ratio is compatible with unity within uncertainties. This finding suggests similar production mechanisms of charm and beauty quarks as a function of multiplicity. The expectation obtained with EPOS 3 shows a modification of the p_T spectrum different for charm and beauty hadrons due to their different mass, which is not supported by the data. A qualitatively similar but weaker behaviour is obtained with EPOS 4, which is more in agreement with the data. All the PYTHIA 8 configurations reproduce the measurements within the uncertainties, indicating a small influence of the hadronisation in the multiplicity dependence, except for the CR-BLC Mode 3 setting, which underestimates the data at low p_T in the high-multiplicity class of events. The data points are further compared to a CGC model that includes the three-pomeron exchange mechanism [230]. In this model, the transition from the beauty quark to the charm hadron is modelled in a single step using $f(b \rightarrow H_c)$ fragmentation functions measured in e^+e^- collisions [231]. Even though these fragmentation functions were shown to be unable to reproduce the measured cross sections of non-prompt D mesons in previous studies [63], the CGC predictions indicate a very weak multiplicity dependence of the double ratio and are consistent with the data within uncertainties.

The specific multiplicity dependence of $f_{\text{non-prompt}}^{\text{mult}}/f_{\text{non-prompt}}^{\text{INEL}>0}$ can be studied in more detail by plotting the values obtained in each individual transverse momentum interval as a function of the charged-particle multiplicity density normalised to the value corresponding to the $\text{INEL} > 0$ class of events, as shown in Fig. 4.41. In all p_T intervals, the average D^0 and D^+ $f_{\text{non-prompt}}^{\text{mult}}/f_{\text{non-prompt}}^{\text{INEL}>0}$ ratio is found to be compatible with unity, indicating a weak (if any) dependence of $f_{\text{non-prompt}}$ with the charged-particle multiplicity. Comparisons with models reveal that the EPOS 3 event generator predicts a multiplicity dependence at intermediate transverse momentum ($4 < p_T < 6 \text{ GeV}/c$) which is ruled out by the data. At lower and higher p_T , instead, EPOS 3 predicts a milder charged-particle multiplicity dependence and is hence closer to the data. Moreover, the multiplicity-independence of CGC predictions is

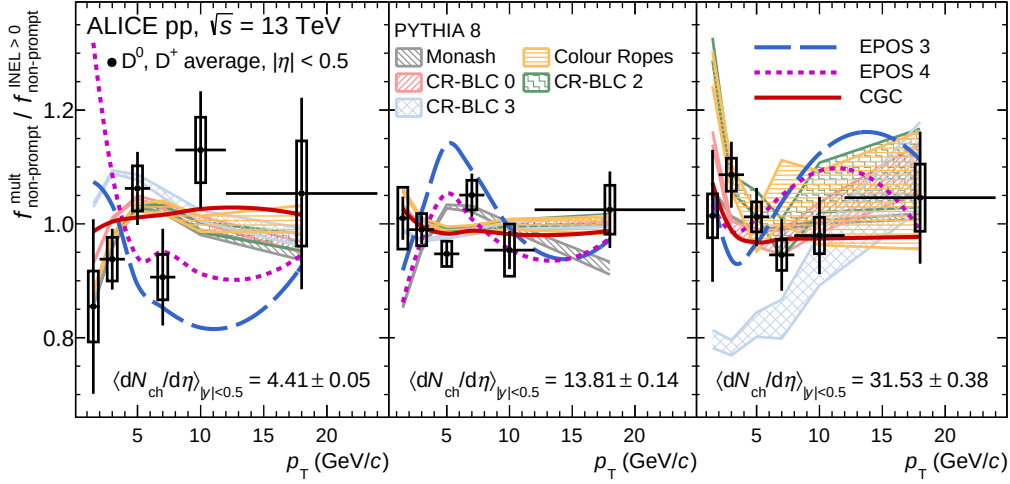


Figure 4.40: Average fractions of non-prompt D^0 and D^+ mesons as a function of p_T for different multiplicity intervals normalised to the one measured in the $\text{INEL} > 0$ class of pp collisions at $\sqrt{s} = 13$ TeV. The measurements are compared with the predictions obtained with PYTHIA 8 [137] and EPOS 3 [227] event generators and the CGC model.

also consistent with the data. Finally, most PYTHIA 8 predictions are consistent with the data, with the notable exception of CR-BLC Mode 3 results, in which the double ratio is shown to decrease with multiplicity. This behaviour can be understood by isolating the double ratio for D mesons originating from beauty-meson and beauty-baryon decays in each of the specific PYTHIA 8 configurations being used, as represented in Fig. 4.42. While in all cases the $f_{\text{non-prompt}}^{\text{mult}} / f_{\text{non-prompt}}^{\text{INEL}>0}$ ratio from beauty baryons increases systematically with multiplicity, the Mode 3 setting results in a decrease of this double ratio for D mesons originating from B-meson decays. More specifically, a clean MC-only test can be performed with the beyond-leading-colour tunes by calculating the ratio of baryons and mesons at hadronisation time in PYTHIA 8 as a function of multiplicity in each model, as depicted in Fig. 4.43. Notably, CR-BLC Mode 3 differs from other PYTHIA 8 predictions due to the fact that, in that case, beauty quarks produce significantly more baryons, and charm quarks produce fewer baryons than in other cases. Consequently, the fraction of non-prompt D mesons decreases with the multiplicity as a combination of two effects. On the one side, charm quarks hadronise more to D mesons, increasing the prompt contribution to the D-meson production and, on the other side, beauty quarks will tend towards being contained in baryons, which in turn will feed preferentially into charm baryons such as the Λ_c^+ baryon. This strong preference towards beauty baryons is not favoured by current ALICE data, which essentially rules out the CR-BLC Mode 3 dynamics in favour of models in which $f_{\text{non-prompt}}$ tends to either remain constant or increase slightly with multiplicity. Future studies of meson and baryon production in the beauty sector as a function of the charged-particle multiplicity will allow for firmer conclusions.

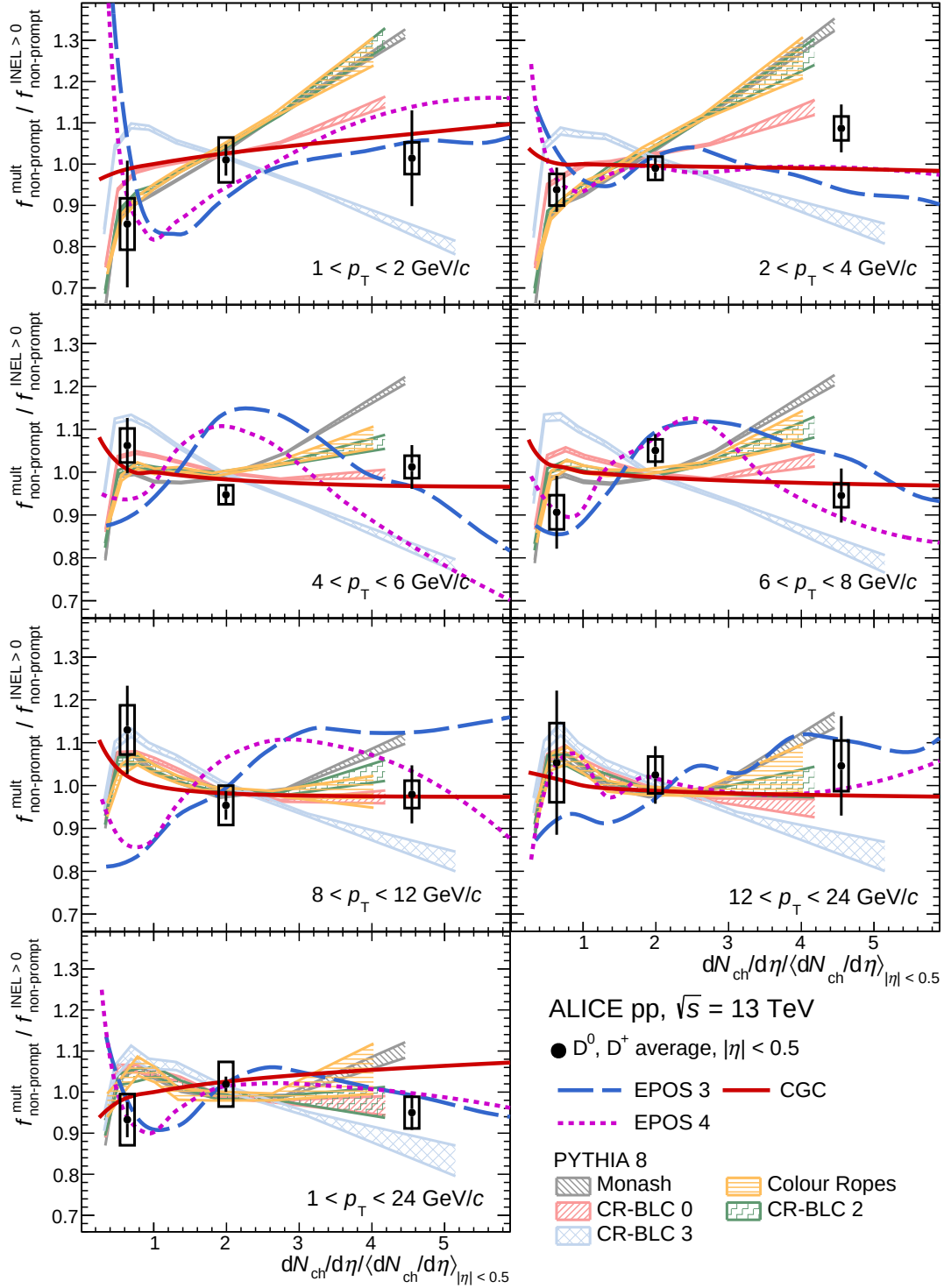


Figure 4.41: Average fractions of non-prompt D^0 and D^+ mesons as a function of multiplicity, both normalised to the value corresponding to the $\text{INEL} > 0$ class, for pp collisions at $\sqrt{s} = 13$ TeV in different p_T intervals and integrated in $1 < p_T < 24$ GeV/c. The measurements are compared with predictions obtained with the PYTHIA 8 [137] and EPOS 3 [227] event generators and the CGC model [230].

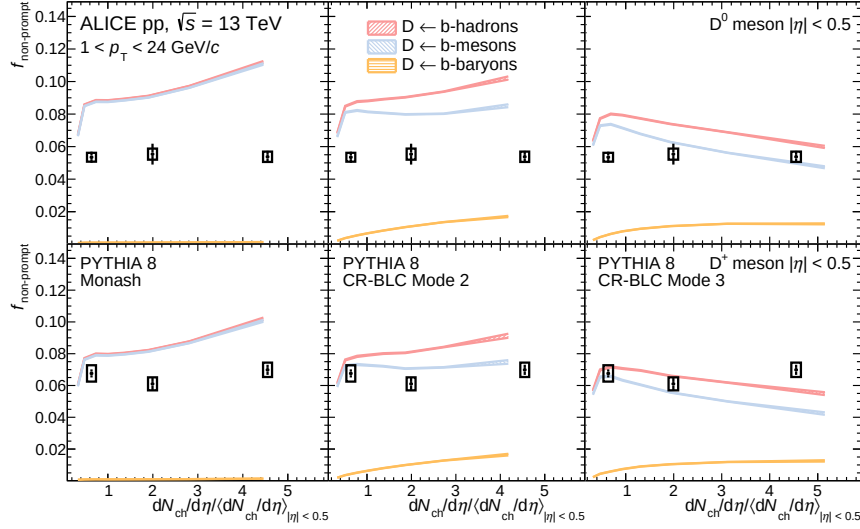


Figure 4.42: Fractions of non-prompt D^0 (first row) and D^+ (second row) mesons in $1 < p_T < 24 \text{ GeV}/c$ as a function of multiplicity for pp collisions at $\sqrt{s} = 13 \text{ TeV}$ compared with predictions obtained with the PYTHIA 8 [137] event generator. The contributions from beauty meson and baryon decays in PYTHIA 8 are displayed separately.

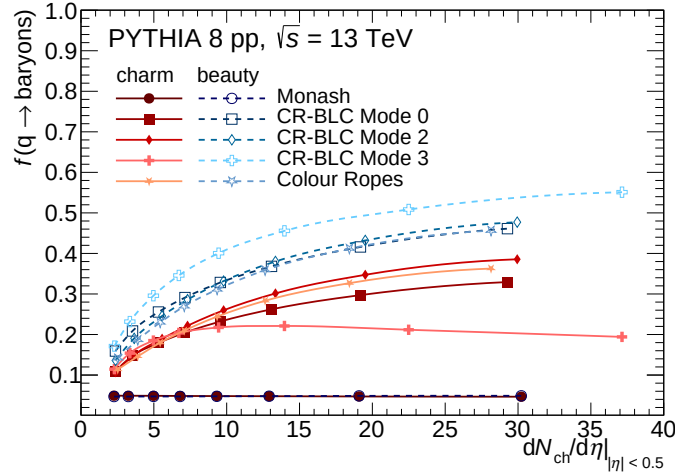


Figure 4.43: Fraction of charm and beauty quarks hadronising to baryons as a function of the charged particle multiplicity at midrapidity in PYTHIA 8 [137] simulations with different tunes.



4.3 Non-prompt D^0 production in pp collisions at $\sqrt{s} = 13$ TeV

In the charm sector, recent measurements of the production ratios of charm baryons to mesons and fragmentation fractions at midrapidity in pp collisions at the LHC have shown significant deviations from the values measured at electron-positron and electron-proton colliders [53, 66, 85, 232–237]. These deviations demonstrate that the assumption of a common hadronization process across collision systems needs to be reconsidered. The observed baryon-to-meson enhancement also leads to an increase in the derived $c\bar{c}$ production cross section at midrapidity compared to previous measurements, as reported in Ref. [53]. Several models have been proposed to explain these differences. Models based on coalescence hadronization, which involves the recombination of heavy and light quarks that are close in momentum and space, are able to reproduce the baryon enhancement and the p_T dependence of the measured Λ_c^+/D^0 yield ratio [135, 238]. Statistical hadronization models can also explain the enhancement if an augmented set of yet unobserved charm-baryon states predicted by the relativistic-quark model [215] and lattice QCD [136] is considered. String fragmentation models that include color reconnection mechanisms beyond the leading-color approximation also introduce new topologies that enhance the contribution of ”junctions” that fragment into baryons, thus providing augmented baryon production [173].

In the beauty sector, recent measurements at the LHC have shown that the assumption of a common hadronization process across different collision systems needs to be reconsidered, as there were significant deviations from the values measured at other colliders. The measurements of Λ_b^0 -baryon production relative to that of B mesons at forward rapidity show a modification of the fragmentation fractions among collision systems similar to that observed for the charm quark [239, 240]. However, the rapidity-dependent baryon-to-meson enhancement is still not fully understood, as different enhancements have been observed for the charm baryon-to-meson ratio between midrapidity and forward rapidity for $p_T < 8$ GeV/c [235, 241]. Beauty measurements at midrapidity are available only at high transverse momentum ($p_T > 7$ GeV/c) [101, 102, 148, 242–246], hence a firm conclusion cannot be drawn.

In this section, the first measurements of the p_T -differential and p_T -integrated production cross sections of D^0 mesons at midrapidity ($|y| < 0.5$) originating from beauty-hadron decays (denoted as non-prompt) in pp collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV with the ALICE experiment are reported. The differential measurements performed in $1 < p_T < 24$ GeV/c for D^0 mesons, are extrapolated down to $p_T = 0$ to compute the p_T -integrated production cross sections and measure the $b\bar{b}$ production cross section at midrapidity. Finally, the non-prompt Λ_c^+/D^0 yield ratio is compared with the yield ratio of prompt Λ_c^+ and D^0 , with the Λ/K^0 yield ratio, and with predictions based on the PYTHIA 8



Monte Carlo (MC) generator to study the flavor dependence of the baryons production relative to that of the mesons.

4.3.1 Raw yield extraction

The data and MC simulations used for this analysis are only used MB events as shown in Tab 2.3. The selected events correspond to an integrated luminosity of $\mathcal{L}_{\text{int}} = 31.9 \pm 0.5 \text{ nb}^{-1}$ [225]. Only the event based on minimum-bias trigger are used to extracted signals. The similar event and candidate selection are choosed for this analysis, as well as the same models trained from integrated multiplicity bin (MB) be applied on the MB triggered data. More details about the ML model performance be-described in Sec. 4.2.2. To enhance the contribution of non-prompt D^0 mesons in the MB sample while maintaining sufficient statistical significance, a set of threshold values were chosen as selection criteria. These values are presented in Tab. 4.11 and were determined by fixing the selection on the probability of the combinatorial background and varying the selection on the non-prompt D^0 probability. The selection on the background probability was chosen to ensure high statistical significance across the entire range of non-prompt D^0 probabilities used in the analysis. The specific values for the background probability selection for D^0 mesons are also provided in Tab. 4.11.

Table 4.11: Selection criteria applied for the D^0 meson in the p_T bins of the analysis.

p_T interval (GeV/c)	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
probability to be background <	0.006	0.008	0.012	0.015	0.05	0.10
probability to be non-prompt >	0.88	0.88	0.88	0.88	0.82	0.80

The selected candidates are divided into six bins based on their candidate p_T . This binning is relatively coarse, but was chosen to ensure sufficient statistics for the signal extraction for all heavy-flavour hadrons in all multiplicity ranges. Fig. 4.44 displays the prompt-enhanced invariant mass distribution for the six p_T bins of D^0 , which are necessary to determine the Gaussian width required to fix the corresponding parameter in the fits for the non-prompt enhanced sample. The fit function is composed of a Gaussian function for the signal and an exponential term for the combinatorial background. Fig. 4.45 shows a high level of significance, even for the non-prompt enhanced sample, which serves as the central working point for the cross section measurement.

4.3.2 Corrections

The non-prompt D^0 p_T -differential cross sections were obtained in the rapidity interval $|y| < 0.5$ as

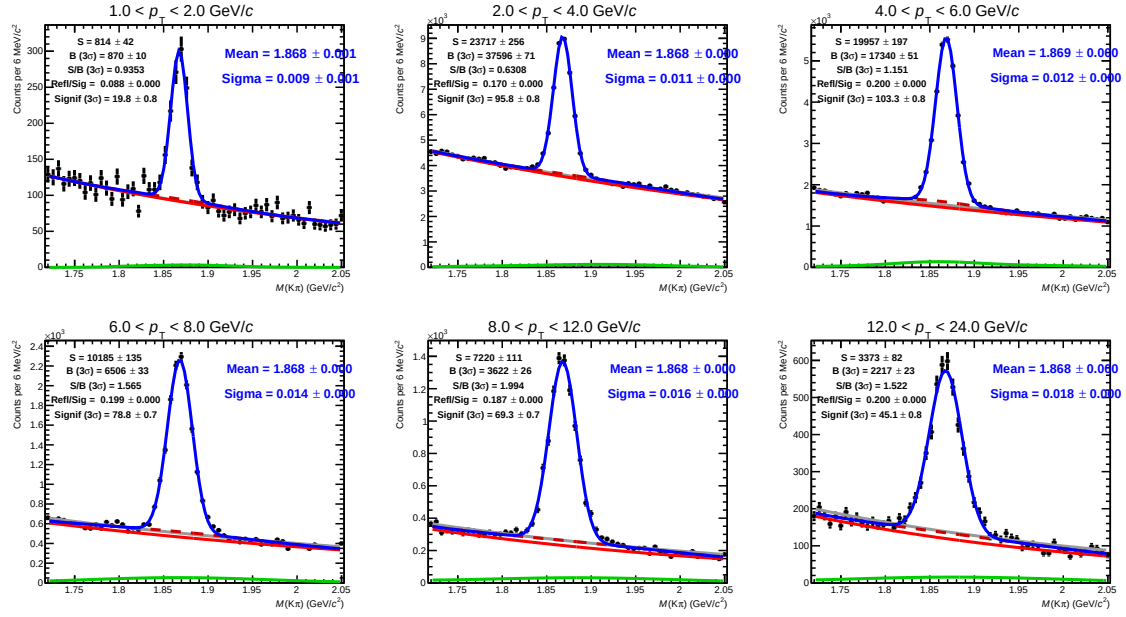


Figure 4.44: Invariant mass distribution of D^0 candidates for the prompt enhanced set of selections in the MB sample.

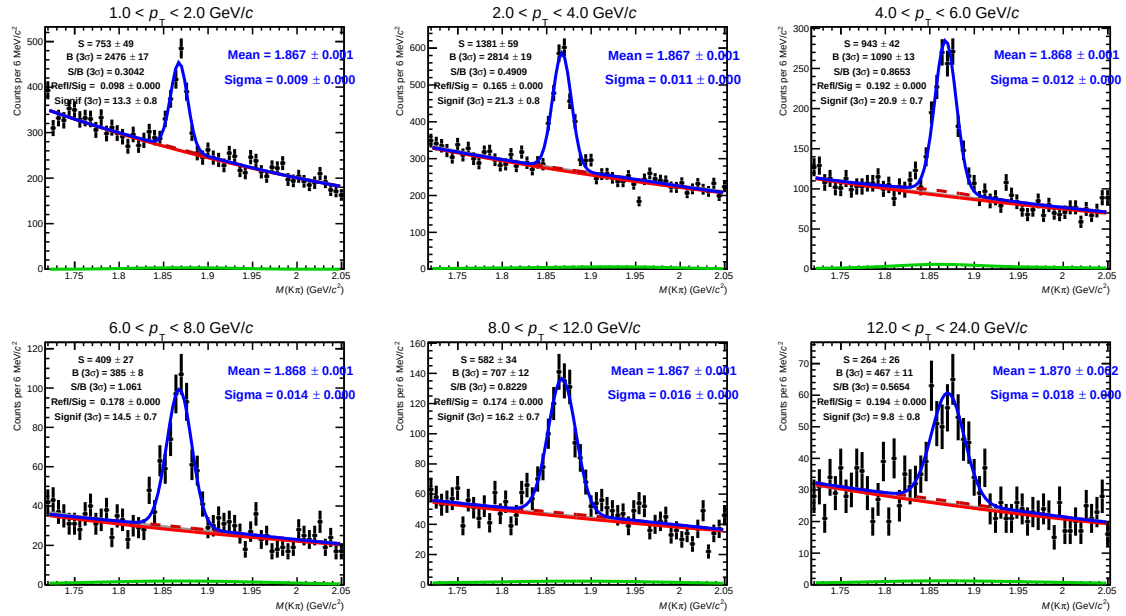


Figure 4.45: Invariant mass distribution of D^0 candidates for the prompt enhanced set of selections in the MB sample.

$$\frac{d\sigma^{D^0}}{dp_T dy} = \frac{1}{BR} \times \frac{1}{2c_{\Delta y}(p_T)\Delta p_T} \times \frac{f_{\text{non-prompt}} \times N_{\text{raw}}}{(\text{Acc} \times \epsilon)_{\text{non-prompt}}} \times \frac{1}{\mathcal{L}_{\text{int}}}, \quad (4.6)$$

where N_{raw} is the sum of particles and antiparticles, $c_{\Delta y}(p_T)$ and Δp_T denote the width of the rapidity and transverse momentum intervals respectively, BR is the branching ratio of the decay mode of $D^0 \rightarrow K^- \pi^+$ ($3.947 \pm 0.030\%$ [178]), the factor 2 accounts for both particles and antiparticles, $f_{\text{non-prompt}}$ is the fraction of non-prompt hadrons in the raw yield, and $(\text{Acc} \times \epsilon)_{\text{non-prompt}}$ is the product of the geometrical acceptance and the reconstruction and selection efficiency for non-prompt hadrons. The value of $(\text{Acc} \times \epsilon)_{\text{non-prompt}}$ increases with p_T and ranges from 5% to 40% for D^0 , depending on the BDT selections and decay channel. The integrated luminosity is denoted by $\mathcal{L}_{\text{int}}$. The detector acceptance and the signal reconstruction and selection efficiency correction factors $(\text{Acc} \times \epsilon)$ were determined using MC simulations.

The same data-driven procedure introduced in Sec. 3.2 was used to estimate the fraction $f_{\text{non-prompt}}$ of non-prompt D^0 mesons. Fig. 4.46 shown the examples of raw-yield distribution as a function of the BDT-based selection employed in the χ^2 -minimisation procedure adopted for the determination of $f_{\text{non-prompt}}$ of D mesons. The D^0 mesons be extracted in $2 < p_T < 4 \text{ GeV}/c$ MB triggered events.

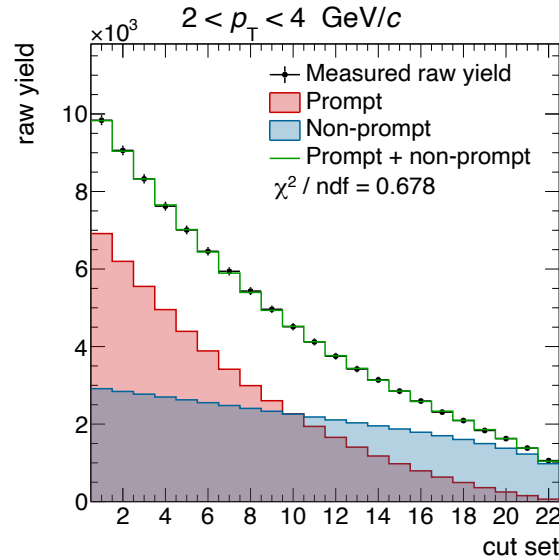


Figure 4.46: Examples of raw-yield distribution as a function of the BDT-based selection employed in the χ^2 -minimisation procedure adopted for the determination of $f_{\text{non-prompt}}$ of D mesons. D^0 mesons be extracted in $2 < p_T < 4 \text{ GeV}/c$ MB triggered events.



4.3.3 Systematic uncertainties

In this analysis, seven sources of systematic uncertainties are considered as the systematic uncertainties for non-prompt D^0 production cross section in MB case. (i) raw yield extraction; (ii) track reconstruction efficiency; (iii) ML selection efficiency; (iv) non-prompt fraction; (v) particle identification; (vi) MC p_T shape; (vii) material budget.

Since both systematic uncertainties are common sources with HF analysis and have been discussed in previous sections, here I will briefly report their values for the non-prompt D^0 production cross section. The systematic uncertainties of the non-prompt D^0 cross sections were evaluated as a function of p_T for different decay channels. The contributions from raw-yield extraction were evaluated by repeating the fits, varying the fit interval, and the functional form of the background fit function. The relative uncertainty from this contribution ranges from 1–3%. The uncertainties of the track reconstruction efficiency were estimated by summing in quadrature the uncertainty due to track quality selection and the uncertainty due to the TPC-ITS matching efficiency, and it ranges from 3.5% to 5%.

The systematic uncertainties of the non-prompt fraction were evaluated by varying the configuration and the number of BDT selections employed in the data-driven method, resulting in an uncertainty of 2–3%. The selection efficiency uncertainties, ranging from 2% to 4%, were studied by repeating the analyses using different BDT selections. The systematic uncertainties of the PID selection efficiency were found to be negligible, similar to those observed for prompt charm hadrons. The systematic effects due to a possible difference between the real and simulated p_T spectra were evaluated by reweighting the p_T shape from the PYTHIA 8.243 generator as the default shape to match the one from FONLL calculations. The weights were applied to the p_T distributions for prompt hadrons and to the mother beauty-hadron particles in the case of non-prompt hadrons. The systematic uncertainty from this contribution varies in the range 1–4%.

Additionally, the imperfect apparatus material budget description in the MC simulation, particularly for the effects of the absorption of protons, might result in a bias in the estimation of the efficiencies. It was evaluated by comparing the corrected yields of charged pions, kaons, and protons using a standard MC production and one with the material budget increased by 10%. The assigned systematic uncertainty is 2%. All the uncertainty sources described above were assumed to be uncorrelated. The total uncertainties for D^0 , 5–8%, were calculated as the quadratic sum of the contributions of different sources.



4.3.4 Results and discussion

The p_T -differential production cross sections of non-prompt Λ_c^+ baryons and D^0 mesons are shown in Fig. 4.47. The non-prompt Λ_c^+ cross section was obtained by computing a weighted average of the results from the analyses of the $\Lambda_c^+ \rightarrow pK_S^0$ and $\Lambda_c^+ \rightarrow pK^-\pi^+$ decay channels, using the inverse of the quadratic sum of the relative statistical and uncorrelated systematic uncertainties as weights. The systematic uncertainties related to the tracking, luminosity, and generated p_T spectrum in the MC simulations are treated as correlated, the uncertainty of the branching ratios as partially correlated as described in Ref. [247], while all the other sources of systematic uncertainties are considered uncorrelated. The data points are compared with theoretical models based on the B-meson cross section predicted by FONLL calculations. In these predictions, the beauty-quark fragmentation fraction to B mesons, $f(b \rightarrow B)$, was taken from e^+e^- collisions [247]. For the Λ_b^0 baryon, $f(b \rightarrow B)_{e^+e^-} \times (f(b \rightarrow \Lambda_b^0)/f(b \rightarrow B))_{\text{LHCb}}$ was used, where the ratio was measured by the LHCb Collaboration in pp collisions [240]. The resulting beauty-hadron cross section was then folded with the $H_b \rightarrow H_c + X$ decay kinematics obtained with PYTHIA 8 in order to obtain the non-prompt cross section. The uncertainties in the model are those of the FONLL calculation, which arise from the choice of the normalization and factorization scales, the mass of the beauty quark, combined with uncertainties of the CTEQ6.6 PDFs.

The measured visible cross sections of non-prompt Λ_c^+ and D^0 hadrons were computed by integrating the measured p_T -differential cross sections in the measured p_T range. All the systematic uncertainties were propagated as fully correlated among the measured p_T intervals, except for the raw-yield extraction uncertainty. The visible cross section was then extrapolated to the full p_T range, using an extrapolation factor computed as the ratio of the p_T -integrated cross sections predicted by FONLL + PYTHIA 8 integrated over $p_T > 0$ and that in the measured p_T interval. The systematic uncertainty of the extrapolation factors was computed considering (i) the FONLL uncertainties, (ii) the $f(b \rightarrow H_b)$ fragmentation fractions uncertainties, and (iii) the branching ratios uncertainties of the $H_b \rightarrow H_c + X$ decays. The second source was estimated by using different sets of beauty fragmentation fractions (from e^+e^- , $p\bar{p}$ collisions [247], or those measured by LHCb [240]), while for the third one the branching ratios implemented in PYTHIA 8 were reweighted in order to reproduce the measured values reported in

Table 4.12: Production cross sections at midrapidity per unit of rapidity $(d\sigma/dy)|_{|y| < 0.5}$ of non-prompt Λ_c^+ and D^0 hadrons and $b\bar{b}$ pairs $(d\sigma b\bar{b}/dy)|_{|y| < 0.5}$ compared to pQCD calculations [48, 49, 58, 155].

	Measurement (μb)	FONLL (μb)	NNLO (μb)
D^0	39.2 ± 1.4 (stat) ± 2.8 (syst) $^{+3.1}_{-4.3}$ (extrap)	34 ± 14	—
Λ_c^+	19.3 ± 2.3 (stat) ± 1.7 (syst) $^{+1.0}_{-1.5}$ (extrap)	11^{+7}_{-5}	—
$b\bar{b}$	85.0 ± 3.4 (stat) ± 3.4 (syst) $^{+8.6}_{-1.7}$ (extrap)	65^{+28}_{-26}	72^{+22}_{-20}

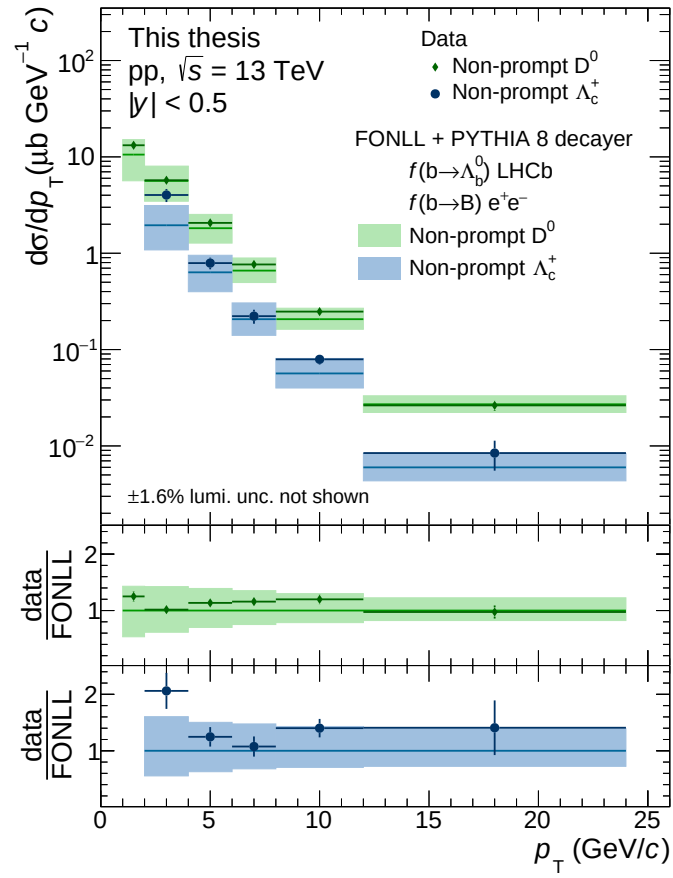


Figure 4.47: p_T -differential production cross sections of non-prompt D^0 and Λ_c^+ hadrons in pp collisions at $\sqrt{s} = 13$ TeV compared with predictions obtained with FONLL calculations [48, 49, 58] combined with PYTHIA 8 [131, 173] for the $H_b \rightarrow H_c + X$ decay kinematics.



Ref [247]. The resulting extrapolation factors are $\alpha_{\text{extrap}}^{D^0} = 1.241_{-0.047}^{+0.009}$ and $\alpha_{\text{extrap}}^{\Lambda_c^+} = 1.847_{-0.152}^{+0.108}$ for D^0 mesons and Λ_c^+ baryons, respectively. The p_T -integrated cross sections are reported in Table 4.12 and compared to FONLL + PYTHIA 8 calculations, which describe the measurements within the uncertainties.

Similarly, the $b\bar{b}$ production cross section per unit of rapidity at midrapidity was obtained using an extrapolation factor computed as the ratio of the beauty cross section and the sum of the visible cross sections of non-prompt D^0 and Λ_c^+ hadrons, estimated with FONLL + PYTHIA 8 as follows

$$\alpha_{\text{extrap}}^{b\bar{b}} = \frac{d\sigma_{b\bar{b}}^{\text{FONLL}}/dy|_{|y|<0.5}}{\sum_{H_c} d\sigma_{H_c \leftarrow b}^{\text{FONLL+PYTHIA 8}}/dy|_{|y|<0.5}(p_T^{\min H_c} < p_T < p_T^{\max H_c})} = 1.91_{-0.03}^{+0.19}. \quad (4.7)$$

The systematic uncertainty of the extrapolation factor includes the same sources considered for the extrapolation of the single-hadron production cross sections. In addition, a correction due to the difference between the rapidity distributions of beauty quarks and beauty hadrons, and between the $b\bar{b}$ pairs and beauty quarks was applied. The first factor was evaluated to be unity in the relevant rapidity range based on FONLL calculations with 1% uncertainty evaluated from the difference between FONLL and PYTHIA 8. The second correction factor is the ratio $(d\sigma_{b\bar{b}}/dy)/(d\sigma_b/dy) = 1.06 \pm 0.01$, which was estimated from POWHEG simulations [57]. The uncertainty was assigned by varying the factorization and renormalisation scales in the POWHEG calculation and using the CT10NLO [248] and CT14NLO [42] PDFs. The resulting $b\bar{b}$ production cross section per unit of rapidity at midrapidity is compatible with the predictions from FONLL and NNLO calculations, as reported in Tab. 4.12. The measurement is also compatible with previous estimates based on dielectrons [249–251] and non-prompt J/ψ [252] measurements.

The p_T -differential ratio of the production cross sections of non-prompt Λ_c^+ and D^0 hadrons is shown in Fig. 4.48. The data are compared, in the top three panels, with the prompt Λ_c^+/D^0 [235] and Λ/K_S^0 [253] ratios, and with PYTHIA 8 predictions obtained with different tunes. Interestingly, all the measurements for beauty, charm, and strange hadrons show a similar behavior as a function of p_T and are compatible within the uncertainties. The prompt and non-prompt Λ_c^+/D^0 ratios are significantly underestimated by PYTHIA 8 simulations carried out with the Monash 2013 tune [131], which implements a fragmentation process tuned to reproduce the measurements in e^+e^- collisions, as illustrated in the top panel of Fig. 4.48. A better agreement is instead obtained with the CLR-BLC tunes [137]. The Mode 2 setting, reported in the second panel of Fig. 4.48, predicts a similar baryon-to-meson ratio for all three flavors for $p_T > 2 \text{ GeV}/c$, while a significantly higher ratio for heavy-flavor hadrons than strange hadrons



Table 4.13: p_T -integrated Λ_c^+/D^0 production ratio measured at midrapidity ($|y| < 0.5$) in pp collisions at $\sqrt{s} = 13$ TeV and in e^+e^- collisions at LEP [212] for prompt and non-prompt production.

	ALICE	LEP average [212]
prompt Λ_c^+/D^0 [220]	$0.49 \pm 0.02(\text{stat})^{+0.05}_{-0.06}(\text{syst})$	0.105 ± 0.013
non-prompt Λ_c^+/D^0	$0.49 \pm 0.06(\text{stat})^{+0.05}_{-0.06}(\text{syst})$	0.124 ± 0.016

for lower p_T (e.g., a factor three is predicted at $p_T \approx 400$ MeV/c). A similar behavior is obtained with the Mode 0 configuration. Instead, the Mode 3 (third panel of Fig. 4.48) underestimates the ratio for charm hadrons for $p_T < 12$ GeV/c and overestimates that of beauty hadrons in the same p_T interval. In the lower panel of Fig. 4.48, the prompt and non-prompt Λ_c^+/D^0 ratios are compared with predictions from the statistical hadronization model TAMU [254]. In this model, the branching fractions of charm and beauty quarks to the different hadron species are assumed to follow the relative thermal densities calculated with the statistical hadronisation model and an enriched set of beauty-hadron states is obtained from the relativistic-quark model [215]. In particular, the non-prompt ratio prediction is obtained by folding the beauty-hadron cross section with the decay kinematics obtained from PYTHIA 8. The non-prompt Λ_c^+/D^0 ratio is well described for $p_T > 4$ GeV/c, while it is underestimated for $2 < p_T < 4$ GeV/c. The prompt ratio is in agreement within the uncertainties for the whole measured p_T interval. The features, observed in all model comparisons, indicate that more precise measurements of the baryon-to-meson ratios, especially at low p_T (e.g., $p_T < 2$ GeV/c), for heavy-flavor hadrons are crucial for tuning the model parameters involving the reconnection of quarks via junction topologies and to possibly validate this as the mechanism of the baryon enhancement observed in hadron collisions compared to e^+e^- collisions.

The p_T -integrated non-prompt Λ_c^+/D^0 yield ratio was computed by dividing the p_T -integrated cross sections reported in Tab. 4.12. All the systematic uncertainties, except for those related to the tracking efficiency, were propagated as uncorrelated in the ratio. The resulting value is compatible with the one measured for promptly produced particles and significantly higher than that measured in e^+e^- collisions at LEP [212]. All the values are reported in Table 4.13.

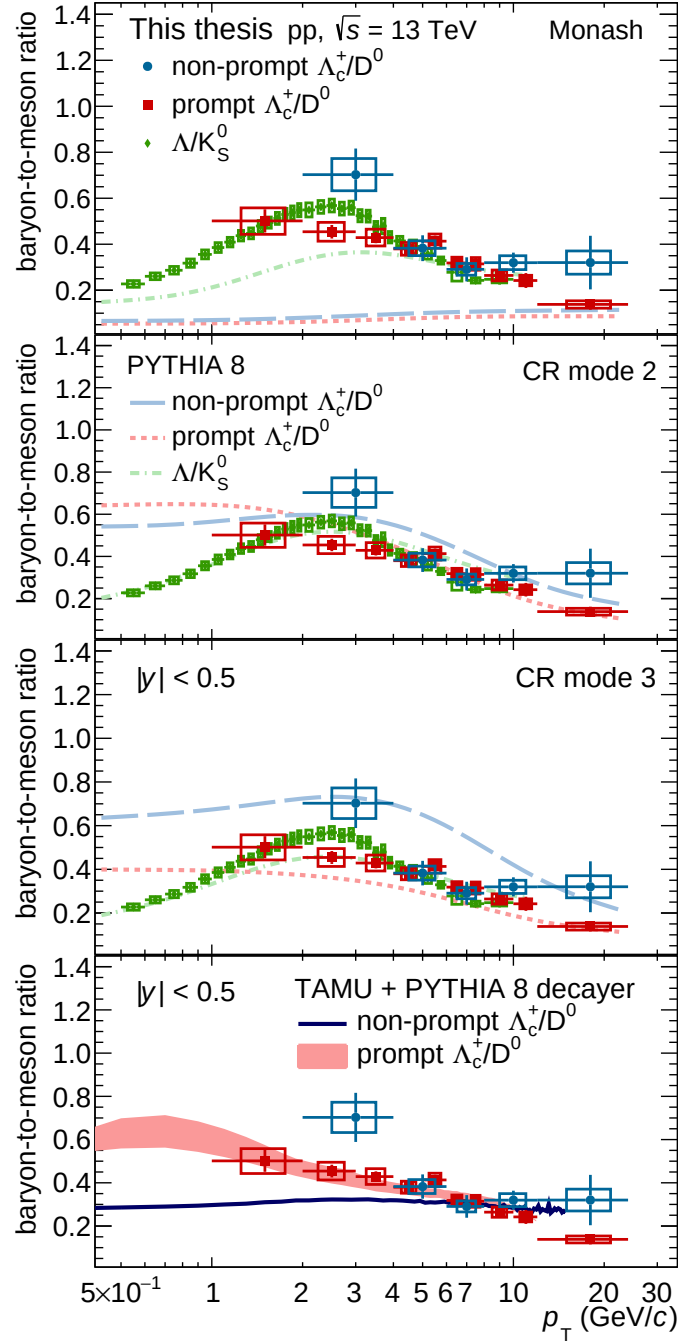


Figure 4.48: Non-prompt Λ_c^+/D^0 , prompt Λ_c^+/D^0 [235], and Λ/K^0 [253] production ratios measured in pp collisions at $\sqrt{s} = 13$ TeV compared with predictions obtained with the PYTHIA 8 MC generator, tuned on three different sets of parameters [131, 173], and with the TAMU model [254].



5 Elliptic flow of non-prompt D^0 in Pb – Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV

The azimuthal anisotropy in the momentum distribution of final state hadrons acts as an additional effective observable to probe the properties of the QGP. In non-central nucleus–nucleus collisions the spatial anisotropy in the initial matter distribution due to the asymmetry of the nuclear-overlap region is transferred to the final momentum distribution via multiple collisions, a phenomenon referred to as anisotropic flow [21, 33]. The anisotropic flow is quantified by the harmonic coefficients $v_n = \langle \cos[n(\varphi - \Psi_n)] \rangle$ of the Fourier expansion of the particle azimuthal angle (φ) relative to the collision symmetry planes with angles (Ψ_n) for the n^{th} harmonic. The second harmonic, called elliptic flow v_2 , is the largest coefficient in non-central heavy-ion collisions. The heavy-flavour v_2 can help to quantify the extent to which charm and beauty quarks participate in the collective expansion of the medium at low p_T ($p_T < 6$ GeV/ c) [116], and the fraction of heavy quarks hadronising via recombination with light quarks in the QGP medium in the intermediate p_T region ($6 < p_T < 10$ GeV/ c) [117, 118]. Instead, the v_2 of heavy-flavour hadrons can constrain the in-medium heavy-quark path-length dependence of energy loss at high p_T ($p_T > 10$ GeV/ c) [112, 113].

D mesons and charm-hadron decay leptons show a positive v_2 in nucleon–nucleon collisions at both RHIC [89, 255–257] and LHC [120, 125, 258–265] energies. The comparison of data measurements with theoretical models indicates that charm quarks participate in the collective expansion of the medium and both collisional processes and the hadronisation of charm quarks via coalescence with light quarks are important to describe the observed elliptic flow [104, 115, 266–272]. In particular, the D-meson v_2 has a similar magnitude as the v_2 of charged pions for $3 < p_T < 6$ GeV/ c , suggesting that low- p_T charm quarks have a relaxation time compatible to the QGP lifetime [103]. Due to their higher mass, beauty quarks are unlikely to reach thermalisation in the medium, therefore their azimuthal anisotropy can give further insight into the interactions of heavy quarks with the medium [120, 273–276]. The experimental information is still poor for low-momentum beauty-hadron v_2 . The elliptic flow of J/ψ mesons originating from beauty-hadron decays (non-prompt) measured by the CMS and ATLAS collaborations is consistent with zero within large uncertainties for $p_T > 3$ GeV/ c [277, 278]. The v_2 of leptons from beauty-hadron decays measured by ALICE and ATLAS is found to be positive [157, 158]. However, due to the small lepton masses, the correlation of p_T and direction between the beauty hadrons and the decay leptons are broad. This is improved when choosing a decay into a heavier particle.



In this chapter, the measurement of non-prompt D^0 -meson v_2 at midrapidity ($|y| < 0.8$) in Pb–Pb collisions at a centre-of-mass energy per nucleon pair $\sqrt{s_{NN}} = 5.02$ TeV with the ALICE detector is reported. The D^0 -meson v_2 is measured with the Scalar Product (SP) method in the 30–50% centrality class. The non-prompt D^0 -meson v_2 is extracted and compared with previous measurements of non-strange D-meson v_2 and v_2 of electrons from beauty hadron decays, as well as with theoretical models based on beauty-quark transport in the QGP.

5.1 Event selection and data samples

The events used in the analysis were recorded with a minimum bias (MB) trigger. An online selection based on the V0 signal amplitudes was also applied in order to enhance the sample of mid-central collisions as an additional trigger class. Background events from beam–gas interactions were rejected offline using the timing information provided by the V0 and the neutron Zero-Degree Calorimeter (ZDC) [279]. Events used in the analysis were required to have a primary vertex reconstructed within ± 10 cm from the centre of the nominal interaction point along the beam axis. Centrality intervals for events were defined in terms of percentiles of the hadronic Pb–Pb cross section based on the signal amplitude of the V0 detectors [23]. After the aforementioned selections, a sample of about 85×10^6 events in the 30–50% centrality class was utilised for further analysis, corresponding to an integrated luminosity of $\mathcal{L}_{\text{int}} \approx 56 \mu\text{b}^{-1}$ [280]. The Monte Carlo samples utilised in the analysis were obtained simulating Pb – Pb collisions with the HIJING v1.383 generator [174] as discussed in Sec 2.4.

5.2 Candidate selection and v_2 extraction

The D^0 mesons and their corresponding charge conjugates were reconstructed using the hadronic decay channel $D^0 \rightarrow K^- \pi^+$ (with a branching ratio of $\text{BR} = (3.947 \pm 0.030)\%$) [178]. To select the D^0 -meson candidates, pairs of tracks were combined with appropriate charge signs, each having $p_T > 0.3$ GeV/ c and $|\eta| < 0.8$, with at least 70 (out of 159) associated space points in the TPC and a minimum of two hits (out of six) in the ITS, with at least one hit in either of the two innermost layers. The track selection criteria decrease the D^0 -meson acceptance in rapidity, which drops steeply to zero for $|y| > 0.5$ at low p_T and for $|y| > 0.8$ for $p_T > 5$ GeV/ c . Therefore, a fiducial acceptance selection $|y| < y_{\text{fid}}(p_T)$ was applied to ensure a uniform acceptance within the considered rapidity range. The $y_{\text{fid}}(p_T)$ value was defined as a second-order polynomial function, increasing from 0.5 to 0.8 for $0 < p_T < 5$ GeV/ c , and as a constant term, $y_{\text{fid}} = 0.8$, for $p_T > 5$ GeV/ c .

The multi-class classification based on Boosted Decision Trees (BDT) was adopted to simultaneously suppress the large combinatorial background and to separate the contributions of prompt and

non-prompt D^0 mesons. Samples of prompt and non-prompt D^0 mesons for the BDT training were obtained from Monte Carlo (MC) samples. Samples for the combinatorial background were obtained from candidates in the sideband region in the data, i.e. $5\sigma < |\Delta M| < 9\sigma$ of invariant mass distribution, where ΔM is the difference between the invariant mass and the mean of signal distribution, σ is the invariant-mass resolution. Before the training, loose selections on kinematic and topological variables were applied to the D^0 -meson candidates to reduce the computation time. The distribution of training variables from signal and background are shown in Fig. 5.1.

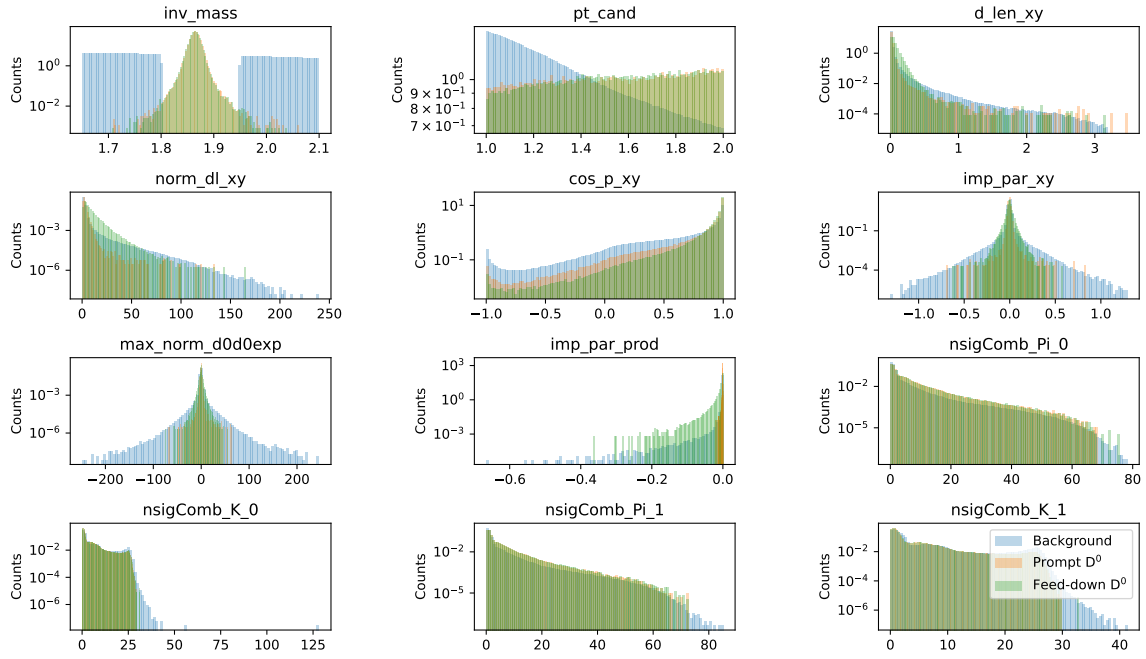


Figure 5.1: Distributions of training variables, invariant mass, and p_T for prompt D^0 mesons, non-prompt D^0 meson, and combinatorial-background candidate variables in the $2 < p_T < 3 \text{ GeV}/c$ interval.

The provided variables to the BDTs were mainly based on the displacement of the tracks from the primary vertex. This includes, the impact parameter of the D^0 -meson daughter tracks, the distance between the D^0 -meson decay vertex and the primary vertex, and the cosine of the pointing angle between the D^0 -meson candidate line of flight (the vector connecting the primary and secondary vertices) and its reconstructed momentum vector, as well as the PID information of the decay tracks.

The ROC curves and the values of ROC AUC, obtained on test set, for the models trained for D^0 meson in the all p_T interval is reported in Fig. 5.2. Independent BDTs were trained in the different p_T intervals of the analysis. Subsequently, the BDTs were applied to the real data sample to obtain the BDT output scores related to the candidate probability to be a non-prompt D^0 -meson or a combinatorial



background. Selections were applied on the scores to reduce the large combinatorial background and to obtain different fractions of non-prompt D^0 candidates.

The D^0 -meson v_2 coefficient was measured with the Scalar Product (SP) method [281–283],

$$v_2\{\text{SP}\} = \langle \langle \mathbf{u}_2 \cdot \frac{\mathbf{Q}_2^{A*}}{M^A} \rangle \rangle / \sqrt{\frac{\langle \frac{\mathbf{Q}_2^A}{M^A} \cdot \frac{\mathbf{Q}_2^{B*}}{M^B} \rangle \langle \frac{\mathbf{Q}_2^A}{M^A} \cdot \frac{\mathbf{Q}_2^{C*}}{M^C} \rangle}{\langle \frac{\mathbf{Q}_2^B}{M^B} \cdot \frac{\mathbf{Q}_2^{C*}}{M^C} \rangle}}, \quad (5.1)$$

where $\mathbf{u}_2 = e^{i2\varphi_{D^0}}$ is the unit flow vector of the D^0 -meson candidate with azimuthal angle φ_{D^0} . $\mathbf{Q}_2^k$ and M^k are the subevent 2nd harmonic flow vector and multiplicity for the subevent k , respectively. The denominator, called resolution (R_2), is calculated with the formula introduced in Ref. [281], where the three subevents, indicated as A, B, and C, are defined by the particles measured in the V0C, V0A, and TPC detectors, respectively. For the TPC detector, the azimuthal angles of charged tracks reconstructed with $|\eta| < 0.8$ and the number of measured tracks were used to calculate the $\mathbf{Q}_2$ vectors and M , respectively. For the V0A and V0C detectors, the $\mathbf{Q}_2$ vectors were calculated from the azimuthal distribution of the energy deposition in the detector sectors and M is the sum of the amplitudes measured in each channel [120]. The $\mathbf{Q}_2$ vectors are recalibrated using a recentering procedure [284] to correct effects of non-uniform acceptance. To test the calibration of the V0, the event-plane angle distribution was checked to be flat within 2%–3%, while the TPC event-plane angle distribution within less than 0.1%, as shown in Figs. 5.3. The single bracket $\langle \rangle$ in Eq. 5.1 refers to an average over all the events, while the double brackets $\langle \langle \rangle \rangle$ denote the average over all particles in the considered p_T interval and all events. The R_2 is extracted as a function of the collision centrality for the second harmonics in the 30–50% centrality classes is shown Fig. 5.4.

The D^0 -meson v_2 cannot be measured directly using Eq. 5.1 since D^0 mesons cannot be identified on a particle-by-particle basis. Therefore, a simultaneous fit to the invariant-mass spectrum and the v_2 distribution as a function of the invariant mass (M_{D^0}) was performed in each p_T interval for D^0 candidates, in order to measure the raw yields and the v_2 coefficients. The measured total elliptic flow coefficient, v_2^{tot} , can be written as a weighted sum of the v_2 of the D^0 -meson candidates v_2^{sig} , and that of background v_2^{bkg} [285] as

$$v_2^{\text{tot}}(M_{D^0}) = v_2^{\text{sig}} \frac{N^{\text{sig}}}{N^{\text{sig}} + N^{\text{bkg}}}(M_{D^0}) + v_2^{\text{bkg}}(M_{D^0}) \frac{N^{\text{bkg}}}{N^{\text{sig}} + N^{\text{bkg}}}(M_{D^0}), \quad (5.2)$$

where N^{sig} and N^{bkg} are the raw signals and background yields, respectively. The fit function for the D^0 -candidate invariant-mass distribution was composed of a Gaussian term to describe the signal and

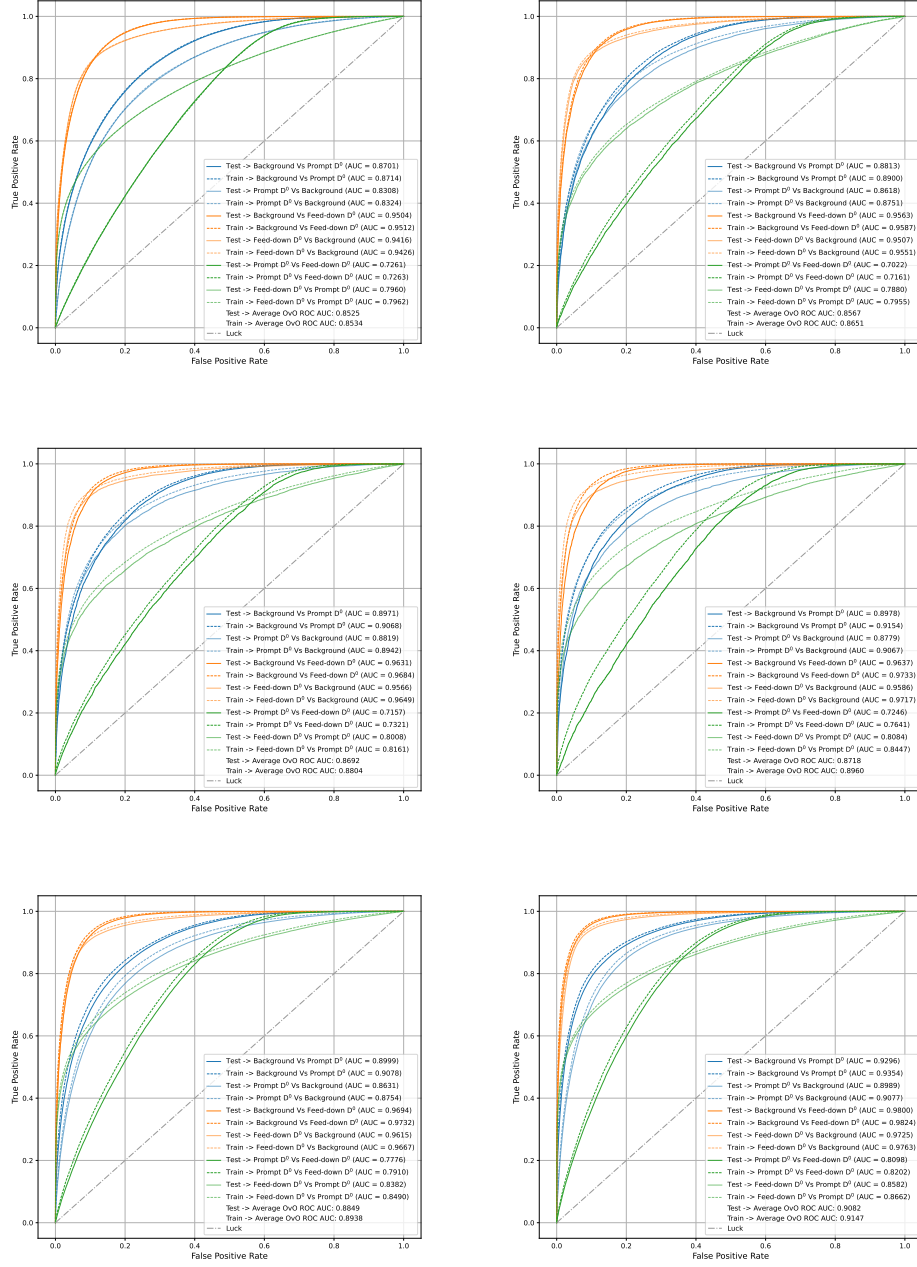


Figure 5.2: ROC curves for the models trained for D^0 meson in the all p_T interval. Top three are (from left to right): $2 < p_T < 3 \text{ GeV}/c$, $3 < p_T < 4 \text{ GeV}/c$, $4 < p_T < 5 \text{ GeV}/c$. bottom three are (from left to right): $5 < p_T < 6 \text{ GeV}/c$, $6 < p_T < 8 \text{ GeV}/c$, $8 < p_T < 12 \text{ GeV}/c$.

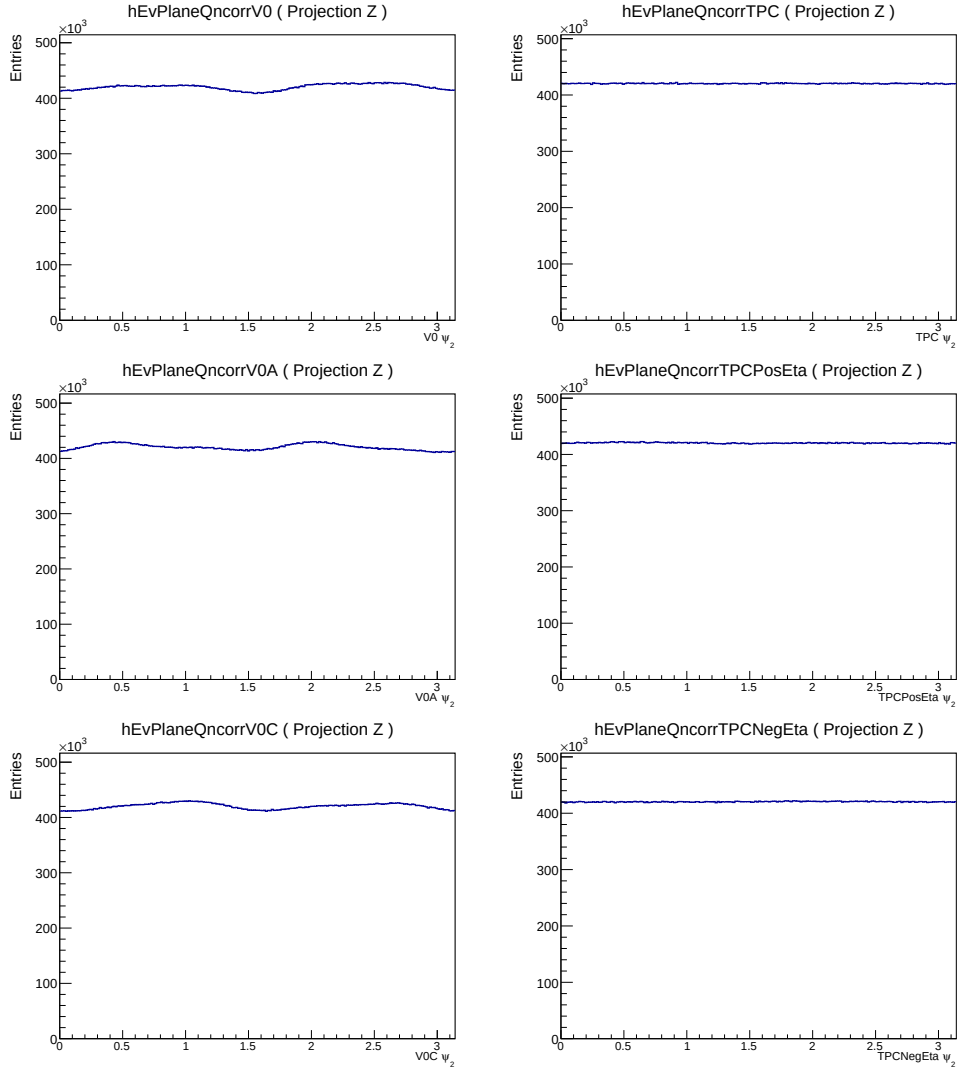


Figure 5.3: Full V0, V0A, V0C (left panels) and TPC $|\eta| < 0.8$, TPC $0.0 < \eta < 0.8$, TPC $-0.8 < \eta < 0.0$ (left panels) event-plane distributions for the 2nd harmonic in the 30–50% centrality classes.

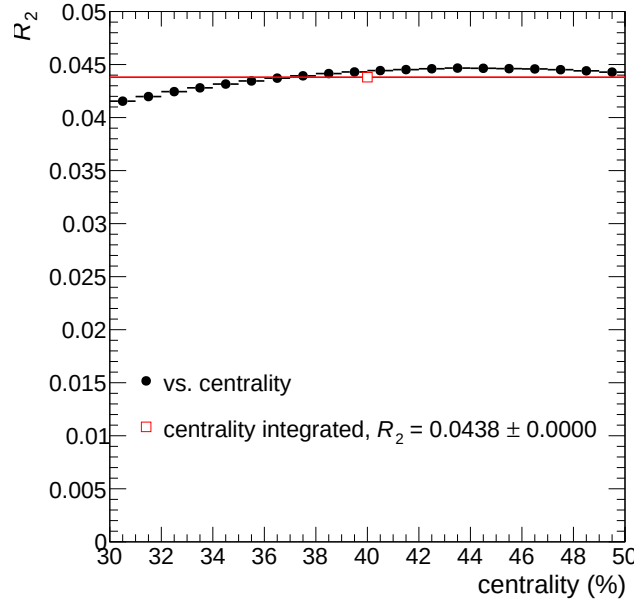


Figure 5.4: Resolution terms for the SP methods as a function of centrality, in the 30–50% centrality class.

an exponential distribution for the background. The contribution of signal candidates with the reflected $K-\pi$ mass assignment was taken into account with an additional term. It was parameterised by fitting the simulated invariant-mass distribution with a double Gaussian function. In the simultaneous fit, the v_2 parameter for the candidates with wrong $K-\pi$ mass assignment was set to be equal to v_2^{sig} , provided that the origin of these candidates are real D^0 mesons. The v_2^{sig} is measured from the fit to the v_2^{tot} distribution with the function of Eq. 5.2, where v_2^{bkg} is a linear function for $p_T > 3 \text{ GeV}/c$. For $p_T < 3 \text{ GeV}/c$, a second-order polynomial function was used to parametrise v_2^{bkg} . Different selections on the BDT output scores were applied in order to obtain samples with different non-prompt D^0 -meson fractions ($f_{\text{non-prompt}}$). Figure 5.5 shows an example of the simultaneous fit to the invariant-mass spectrum and $v_2^{\text{tot}}(M_{D^0})$ with low (left panel) and high (right panel) non-prompt D^0 -meson candidate probability selections in $3 < p_T < 4 \text{ GeV}/c$ in the 30–50% centrality class.

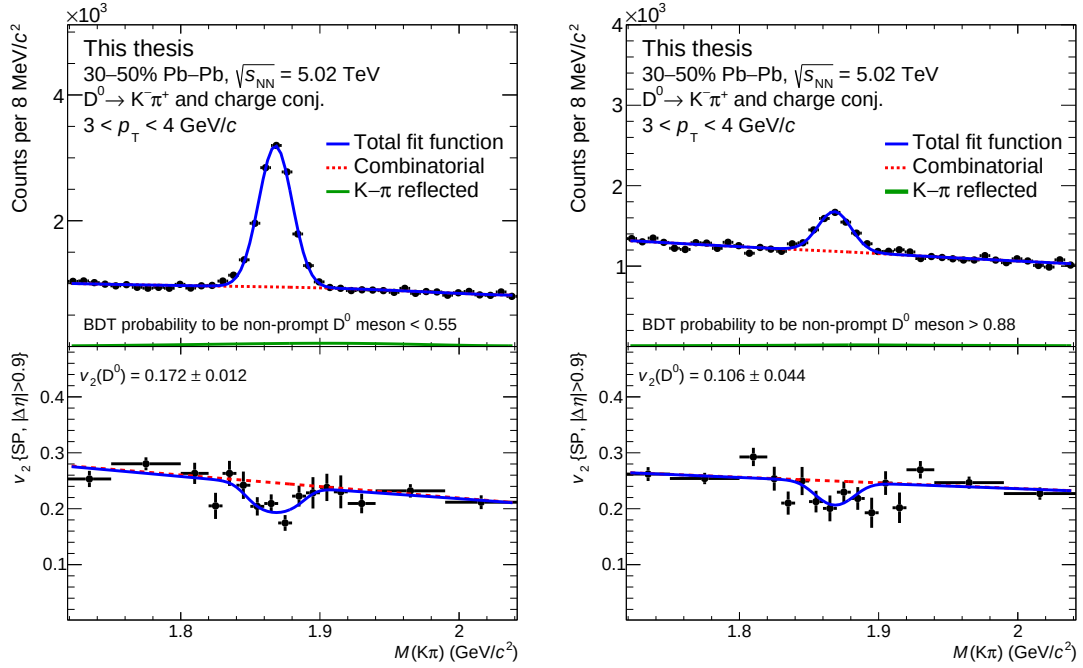


Figure 5.5: Simultaneous fits of the invariant-mass distribution and $v_2^{\text{tot}}(M_{D^0})$ of D^0 mesons in $3 < p_T < 4 \text{ GeV}/c$. Left panel: fits using D^0 meson candidates with low probability to be a non-prompt D^0 -meson. Right panel: fits using D^0 meson candidates with high probability to be a non-prompt D^0 -meson. The corresponding BDT score selection for the measured raw yield is reported. The blue lines, the dotted red curves, and the green solid lines represent the total, the combinatorial-background fit functions, and the contribution of the reflected signal, respectively.



5.3 Non-prompt D^0 v_2 extraction

The reconstructed D^0 -meson signals are a mixture of prompt and non-prompt D^0 mesons. The v_2^{sig} is therefore a linear combination of prompt (v_2^{prompt}) and non-prompt ($v_2^{\text{non-prompt}}$) contributions, which can be expressed as

$$v_2^{\text{sig}} = (1 - f_{\text{non-prompt}})v_2^{\text{prompt}} + f_{\text{non-prompt}}v_2^{\text{non-prompt}}, \quad (5.3)$$

where $f_{\text{non-prompt}}$ is estimated as a function of p_T with a data-driven method, which is based on the construction of data samples with different abundances of prompt and non-prompt candidates. A set of raw yields Y_i (index i refers to a given selection on the BDT output) can be obtained by varying the selection on the BDT output, which is related to the candidate probability to be a non-prompt D^0 meson. These raw yields are sensitive to the corresponding acceptance times efficiency ($\text{Acc} \times \epsilon$) of prompt and non-prompt D^0 mesons as follows

$$(\text{Acc} \times \epsilon)_i^{\text{prompt}} \cdot N_{\text{prompt}} + (\text{Acc} \times \epsilon)_i^{\text{non-prompt}} \cdot N_{\text{non-prompt}} - Y_i = \delta_i, \quad (5.4)$$

where δ_i represents a residual that accounts for the equation not holding exactly due to the uncertainties on Y_i , $(\text{Acc} \times \epsilon)_i^{\text{non-prompt}}$, and $(\text{Acc} \times \epsilon)_i^{\text{prompt}}$. In the case of $n \geq 2$ sets of selections, the corrected yields of prompt (N_{prompt}) and non-prompt ($N_{\text{non-prompt}}$) D^0 mesons can be obtained from Eq. 5.4 via a χ^2 minimisation. More details can be found in Ref. [63]. The Fig. 5.6 shows an example of the raw-yield distributions as a function of the minimum non-prompt D^0 -meson BDT probability threshold used in such a χ^2 -minimisation procedure in $3 < p_T < 4$ GeV/ c for the 30–50% centrality class. The raw yield decreases with the increasing minimum threshold for the probability to be a non-prompt D^0 meson, corresponding to an increasing non-prompt D^0 -meson fraction. The prompt and non-prompt components of the raw yields for each BDT-based selection obtained from the χ^2 -minimisation approach, $(\text{Acc} \times \epsilon)_i^{\text{prompt}} \times N_{\text{prompt}}$ and $(\text{Acc} \times \epsilon)_i^{\text{non-prompt}} \times N_{\text{non-prompt}}$, are shown in the histograms with red and blue colour, respectively, and their sum is reported by the green line. $N_{\text{non-prompt}}$ and N_{prompt} can be used to estimate the non-prompt D^0 -meson fraction in the raw yield for any set of selections i as follows

$$f_{\text{non-prompt}}^i = \frac{(\text{Acc} \times \epsilon)_i^{\text{non-prompt}} \cdot N_{\text{non-prompt}}}{(\text{Acc} \times \epsilon)_i^{\text{non-prompt}} \cdot N_{\text{non-prompt}} + (\text{Acc} \times \epsilon)_i^{\text{prompt}} \cdot N_{\text{prompt}}}, \quad (5.5)$$

5.3.1 Linear fit with correlated cut sets

As discussed in Section 5.2, XGBoost's multi-class classification algorithm can separate the three contributions (prompt D , non-prompt D , and combinatorial background). Therefore, the candidate

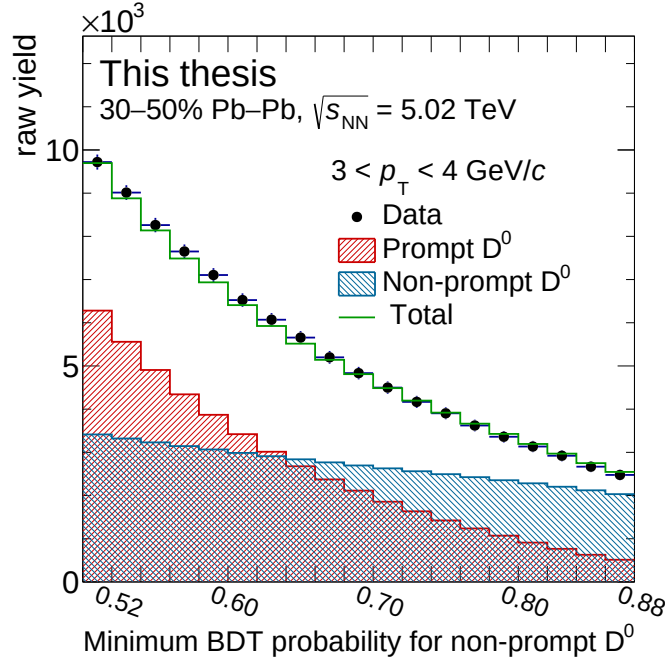


Figure 5.6: Example of the raw-yield distribution as a function of the minimum non-prompt D^0 -meson BDT probability threshold to determine the non-prompt D^0 -meson fraction in $3 < p_T < 4 \text{ GeV}/c$.

samples can be classified as prompt-enhanced and non-prompt-enhanced based on their output scores. Fig. 5.7 shows example plots for different BDT selection strategies. On the left side, the sub-samples selected based on the non-prompt output score are correlated, i.e., each selection is a sub-sample of the previous selection from left to right. This strategy is used to estimate the non-prompt fraction in the cut variation method. However, the statistical error propagation is incorrect with a linear fit when using this correlated cut method to estimate the final non-prompt D^0 v_2 . Thus, this method is only used to cross-check the central value for the non-prompt D^0 v_2 . On the right side, independent samples are selected based on the non-prompt output score using an uncorrelated cut. The v_2^{sig} is calculated using three or four independent samples along with their associated non-prompt D^0 fractions. The final results are extrapolated to $f_{\text{non-prompt}} = 0$ and $f_{\text{non-prompt}} = 1$ using a linear fit according to Eq. 5.5, in order to obtain the v_2 values for prompt and non-prompt D^0 mesons, respectively. The cut variation method can extract f_{prompt} and $f_{\text{non-prompt}}$ for each sample.

To cross-check the central value for the non-prompt D^0 v_2 , we utilized a linear fit with correlated cut sets. Fig. 5.8 displays the six bins of correlated cuts, and we fitted these cut sets using the linear fit and extrapolated to 1 to obtain the non-prompt D^0 v_2 . The linear fit was found to perform very well with the correlated cuts. Fig. 5.9 displays the v_2 values for prompt and non-prompt D^0 mesons in different

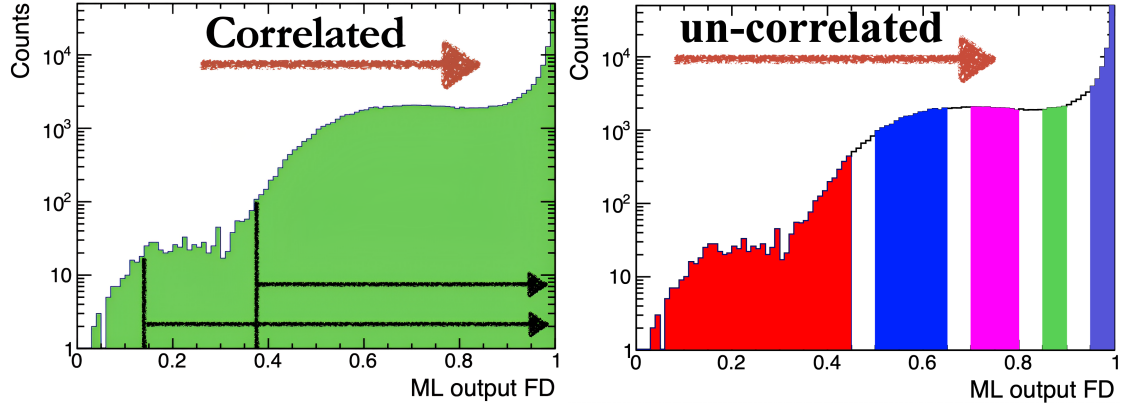


Figure 5.7: Right panel: the selection based on the correlated cut. Left panel: the selection based on the uncorrelated cut.

p_T intervals.

5.3.2 Linear fit with uncorrelated cut sets

The total samples can be divided into several independent samples based on the uncorrelated cut, these subsamples can be used for both $f_{\text{non-prompt}}$ calculation and prompt and non-prompt $D^0 v_2$ separation. For each p_T interval, the full data sample is divided into independent ones by selecting different ranges of Feed-down BDT score, which allows the variation for $f_{\text{non-prompt}}$ to be big enough for each sub-sample. Currently, only three or four sub-samples can be obtained to guarantee the χ^2/ndf is good enough for the non-prompt $D^0 v_2$ extraction. The Tab. 5.1 show the BDT selections for the independent samples used to do the linear fit. The advantage of using independent samples is that we will not suffer from the problem of error propagation. But as the numbers of subsamples are limited, the $f_{\text{non-prompt}}$ calculated with these subsamples fluctuated, as shown in Fig 5.10. Thus the $f_{\text{non-prompt}}$ values are still extracted with corrected cut sets. One can use the first-order polynomial fit the v_2^{obs} as a function as $f_{\text{non-prompt}}$ distribution. 1σ confidence interval obtained from the linear fit, which is considered as the statistical uncertainties of the v_2^{sig} . Figs. 5.11 show the $D^0 v_2^{\text{obs}}$ distribution with the dependence of the non-prompt fraction in the 6 p_T bins of uncorrelated cuts. Figs. 5.12 are shown the final non-prompt $D^0 v_2$ from the fit with $f_{\text{non-prompt}} = 1$, and also the prompt $D^0 v_2$ from the fit with $f_{\text{prompt}} = 0$ (left panel). And the comparison results between uncorrelated cuts fit and correlated cuts fit are also shown in right panel, the central values between these two methods are consistent.

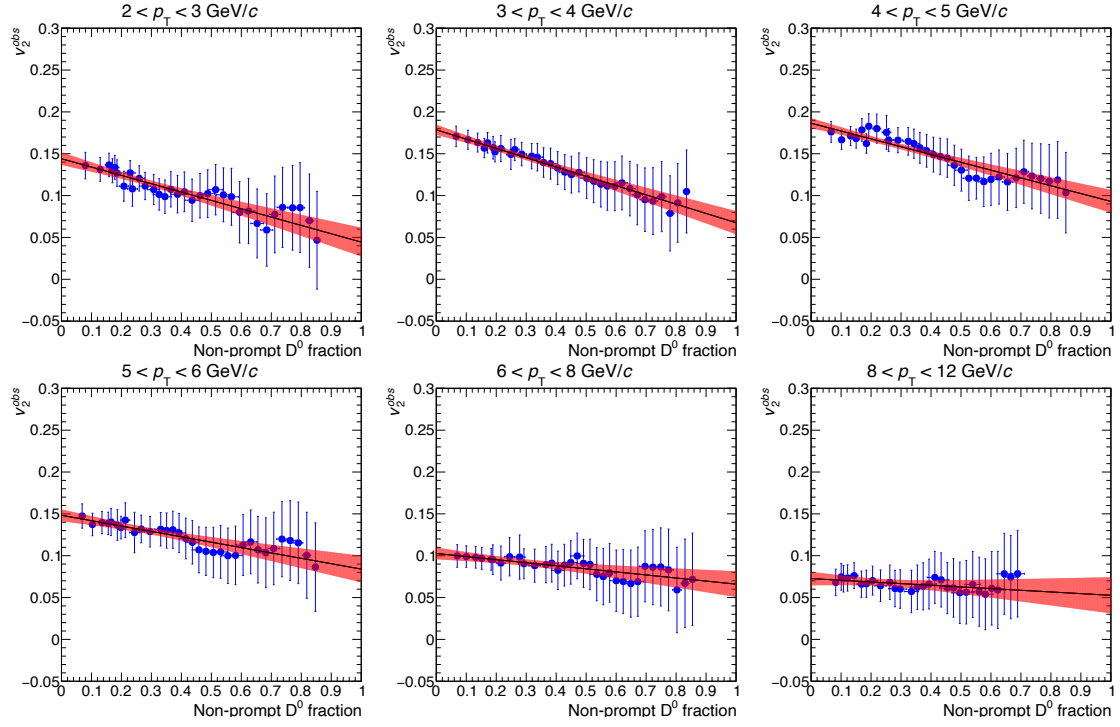


Figure 5.8: Non-prompt v_2 distribution with the dependence of the $f_{non-prompt}$ with the correlated cuts for the 6 p_T in the 30–50% centrality class.

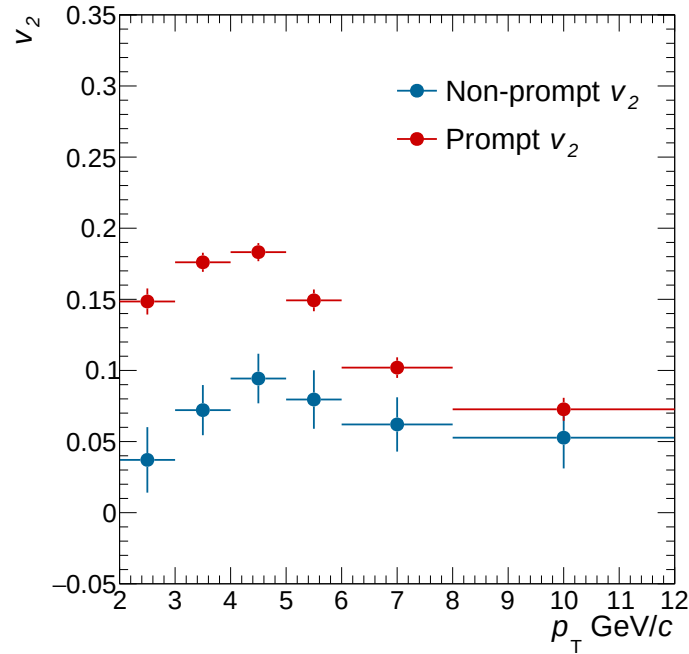


Figure 5.9: Central value estimated of prompt and non-prompt v_2 based on the fit method in the 30–50% centrality classes(dummy error bar).

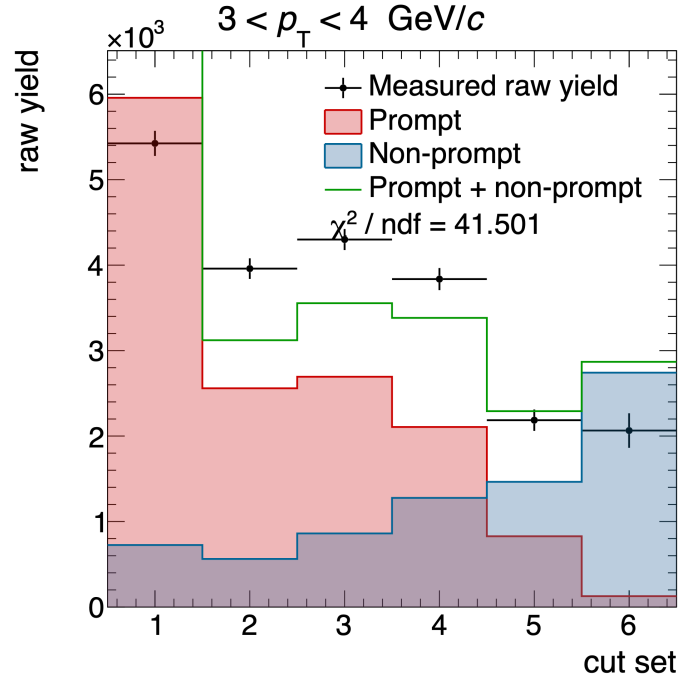


Figure 5.10: $f_{\text{non-prompt}}$ obtained with independent cut sets

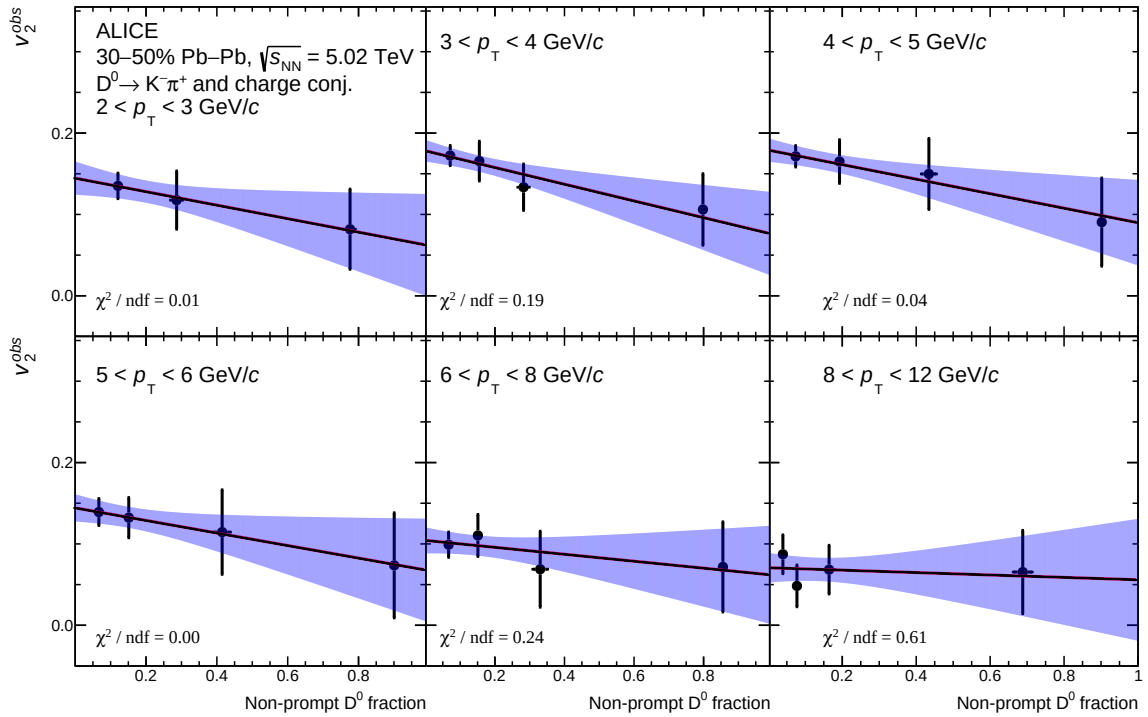


Figure 5.11: v_2^{sig} as a function of $f_{\text{non-prompt}}$. The blue band represents the 1σ confidence interval obtained from the linear fit.



Table 5.1: BDT score cuts for the background and FD in the p_T intervals

p_T interval (GeV/c)	[2, 3]	[3, 4]	[4, 5]	[5, 6]	[6, 8]	[8, 12]
Score (bkg) <	0.001	0.002	0.003	0.004	0.02	0.5
Score (FD) interval 1	[0, 0.65]	[0, 0.55]	[0, 0.60]	[0, 0.52]	[0, 0.50]	[0, 0.30]
Score (FD) interval 2	[0.65, 0.85]	[0.55, 0.70]	[0.60, 0.80]	[0.52, 0.75]	[0.5, 0.70]	[0.30, 0.45]
Score (FD) interval 3	[0.85, 1.00]	[0.70, 0.88]	[0.80, 0.94]	[0.75, 0.94]	[0.70, 0.90]	[0.45, 0.76]
Score (FD) interval 4	[0.85, 1.00]	[0.88, 1.00]	[0.94, 1.00]	[0.94, 1.00]	[0.90, 1.00]	[0.76, 1.00]

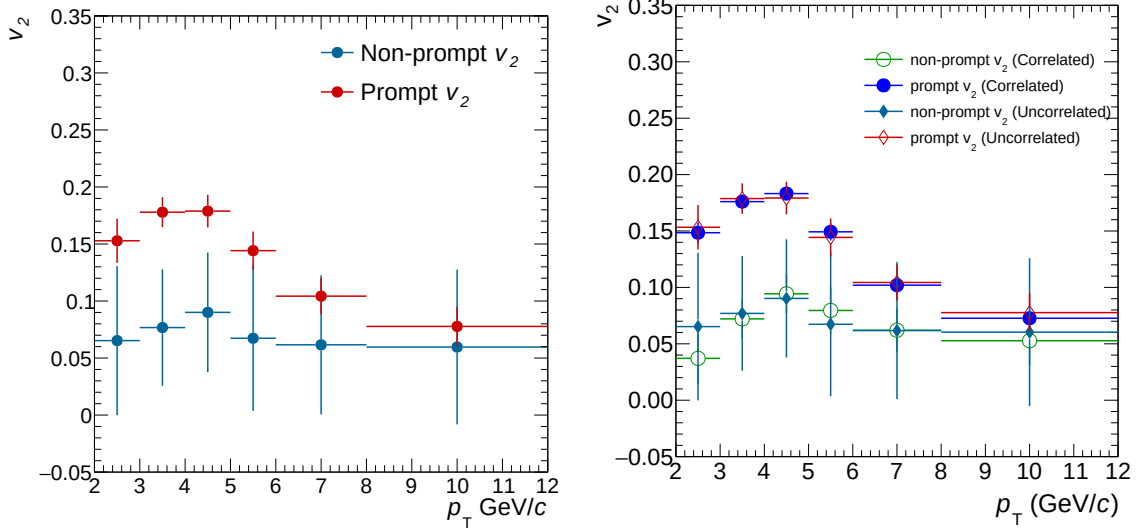


Figure 5.12: Prompt and non-prompt v_2 based on uncorrelated cuts with the fit method in the 30–50% centrality classes (left), right panel show the comparison result between uncorrelated cuts fit and correlated cuts fit.



5.4 Systematic uncertainties

Four major sources of systematic uncertainties were considered for the measurement of the non-prompt D^0 -meson v_2 : (i) the signal extraction from the invariant-mass and v_2^{tot} distributions; (ii) the non-prompt fraction estimation; (iii) the D -meson p_T shape in the simulation; (iv) the centrality dependence of the SP denominator.

5.4.1 Fit procedure

The systematic uncertainty of the signal extraction from the invariant-mass and v_2^{tot} distributions was evaluated by repeating the simultaneous fit with different configurations. In particular, the fit range, signal width and background fit functions used for the invariant-mass and v_2^{tot} distribution were varied, the details shown in Tab. 5.2. The systematic uncertainty was defined as the RMS of the distribution of the resulting $v_2^{\text{non-prompt}}$ obtained from all these variations as presented in Figs. 5.13.

Table 5.2: Configuration of the fit procedure.

D^0 meson	Configuration
Rebin	[1, 2, 3, 4, 5]
Inv.Mass min [GeV/c^2]	[1.70, 1.71, 1.72, 1.73, 1.74]
Inv.Mass mas [GeV/c^2]	[2.02, 2.03, 2.05, 2.06, 2.07]
Bkg function of mass	[Expo, Lin, Pol2]
Bkg function of v_2	[Lin, Pol2]

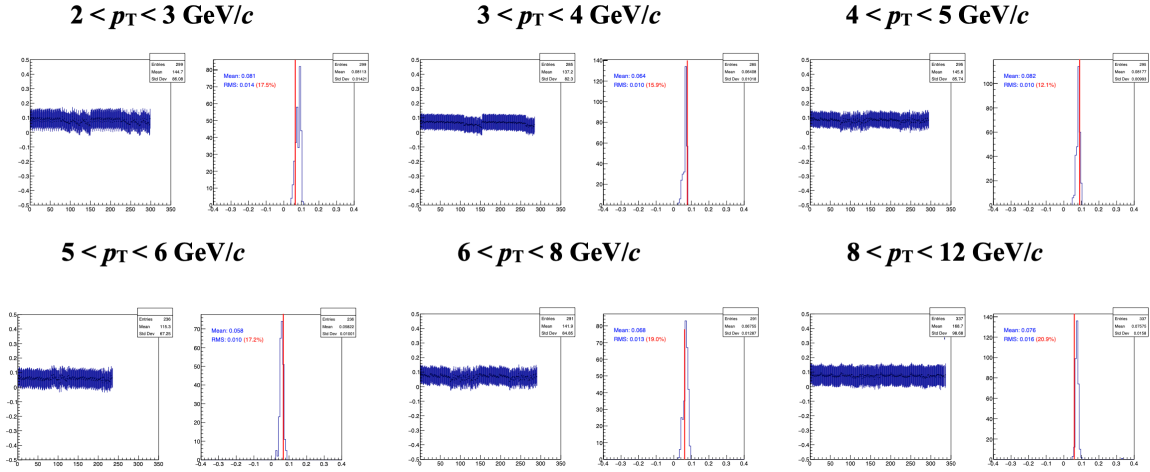


Figure 5.13: Non-prompt D^0 v_2 as a function as multi-trials based on the different fit parameters

5.4.2 $f_{\text{non-prompt}}$ calculation

To estimate the systematic uncertainty due to the non-prompt fraction estimation, we varied the cut sets used in the cut variation method. We considered wider and narrower scan intervals on prob(non-

prompt) and smaller and larger steps between the chosen cut sets. The values of $f_{\text{non-prompt}}$ obtained for each variation in each p_T interval are presented in Figs. 5.14 for the independent samples. We assigned the systematic uncertainty by considering the RMS and shift of the trials relative to the central value of the non-prompt $D^0 v_2$ estimated by the linear fit. The results are shown in Fig. 5.15.

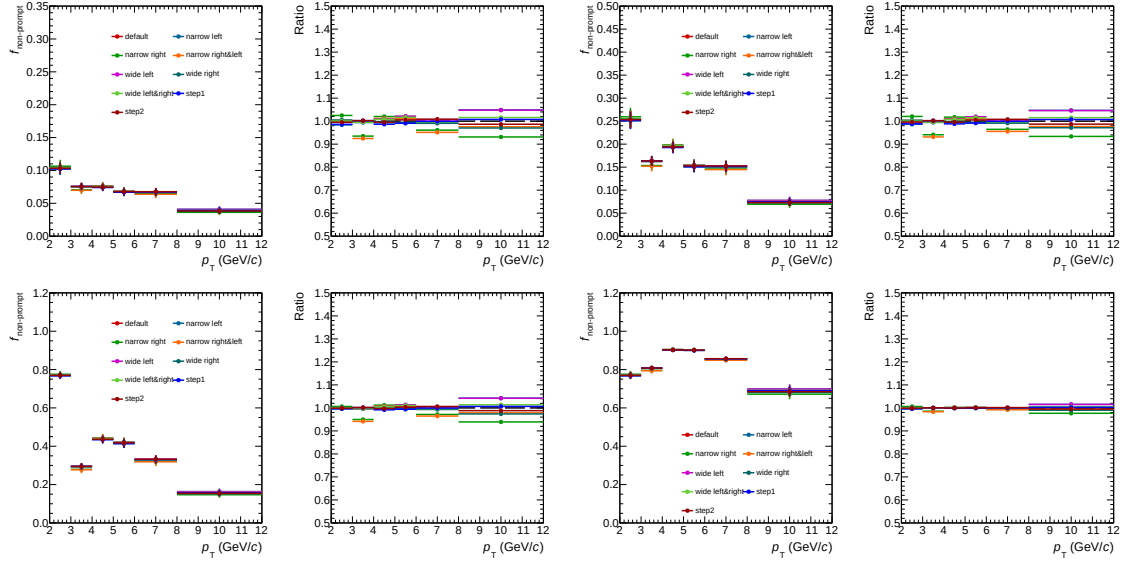


Figure 5.14: $f_{\text{non-prompt}}$ distribution as a function as p_T with the different configuration of the points (right for each panel) and the ratio between $f_{\text{non-prompt}}$ of trials and the default value(left for each panel).

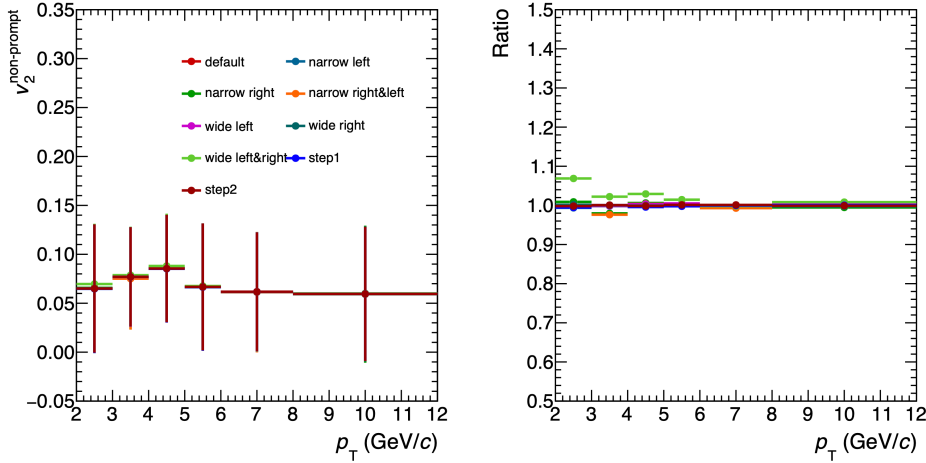


Figure 5.15: The non-prompt $D^0 v_2$ distribution as a function as p_T with the different configuration (right) and the ratio between non-prompt $D^0 v_2$ of trials and the default value(left).

5.4.3 MC p_T shape

The systematic effects due to a possible difference between the real and simulated p_T spectra, were estimated by applying different weights to the p_T distributions of prompt D^0 mesons and of mother



beauty hadrons in case of non-prompt D^0 mesons. For the nominal selection criterion, the weights were defined to match the shape given by FONLL in pp collisions [49, 58] multiplied by the nuclear modification factor (R_{AA}) prediction from the TAMU model [267]. The FONLL multiplied by R_{AA} from LIDO model [286] was used as an alternative shape for the systematic evaluation. Meanwhile, the effect due to beauty-hadron v_2 resulting the modification of the mother p_T spectra, was found to be negligible with respect to the assigned p_T shape systematic uncertainty. The p_T shape for the two models and the weight for charm and beauty are shown in Fig. 5.16 and Fig. 5.17, respectively. The $f_{\text{non-prompt}}$ distribution as function as p_T with the different p_T shape is reported in Fig. 5.18. The effect of p_T re-weight are reported in Fig. 5.19.

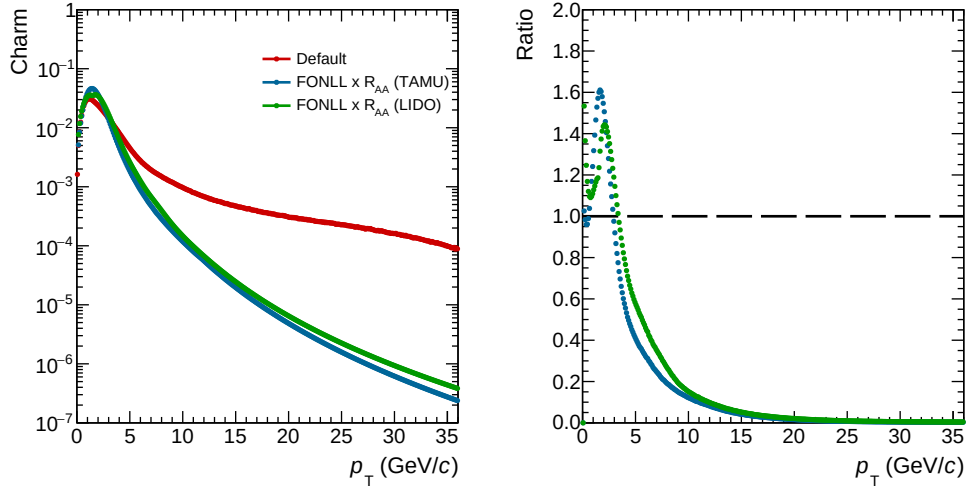


Figure 5.16: p_T distributions for the two models (right) and the weight for the prompt D^0 .

5.4.3.1 p_T shapes in plane and out of plane

Charm and beauty v_2 will lead the number of particle asymmetric between in and out plane, which is not simulated in MC. It will affect the p_T shape. So this effect will be evaluated by using the different p_T shapes from in and out plane. There is an alternative method to calculate the v_2 by extracting the signal in two $\Delta\varphi$ intervals, namely in-plane and out-of-plane. The v_2 can be given by

$$v_2 = \frac{1}{R_2} \frac{\pi}{4} \frac{N_{IP} - N_{OOP}}{N_{IP} + N_{OOP}} \quad (5.6)$$

Therefore the ratio between N_{IP} (N_{OOP}) and $N_{IP} + N_{OOP}$ can express by

$$\frac{N_{IP}}{N_{IP} + N_{OOP}} = \frac{1}{2} + \frac{2v_2}{\pi} \quad (5.7)$$

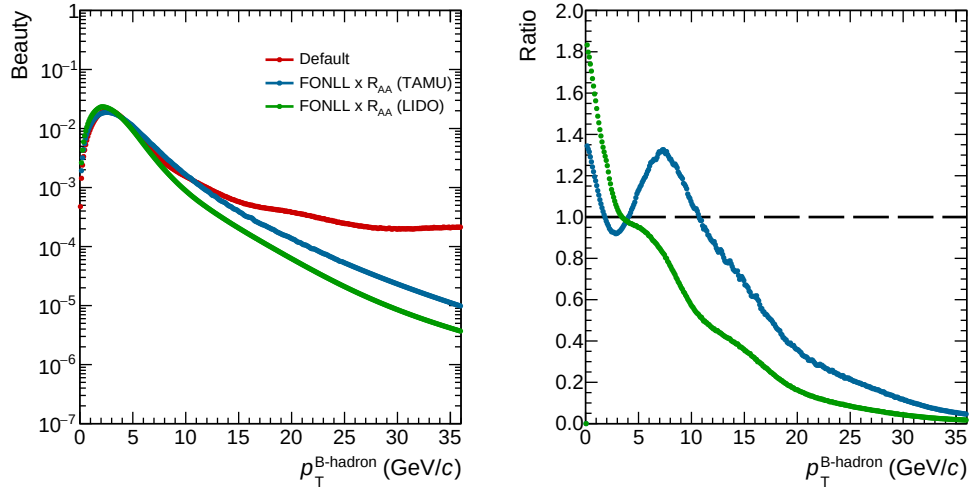


Figure 5.17: p_T distributions for the two models (right) and the weight for the prompt D^0 .

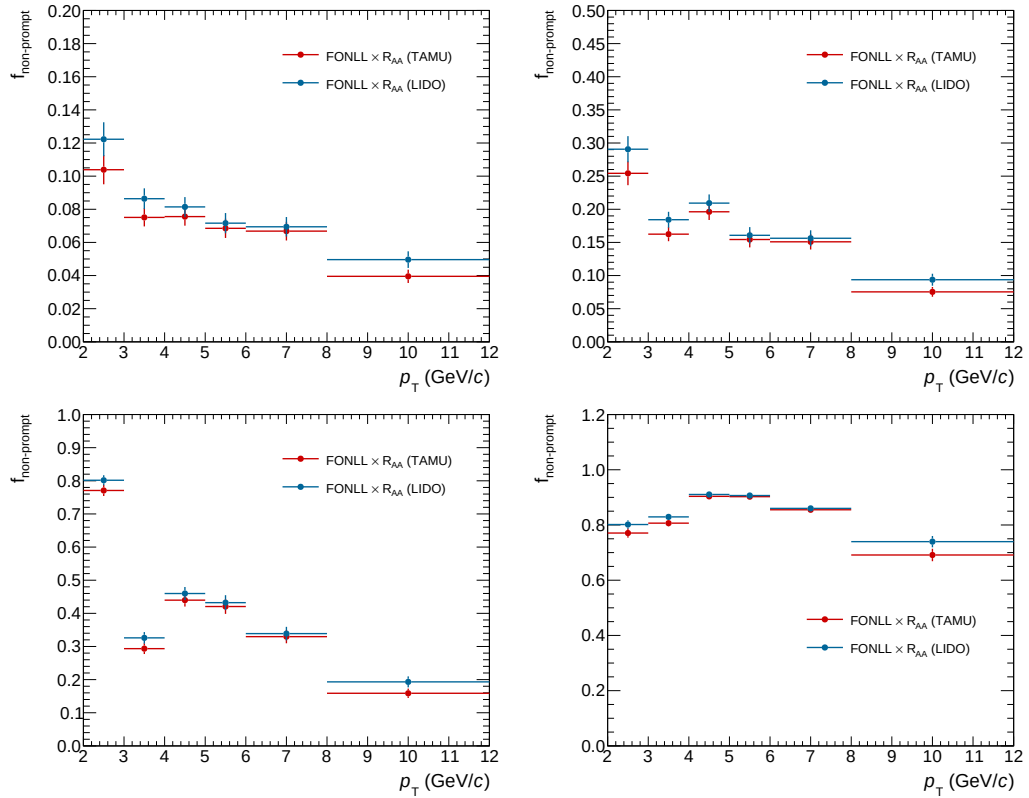


Figure 5.18: $f_{\text{non-prompt}}$ distribution as a function as p_T with the different p_T shape of the points.

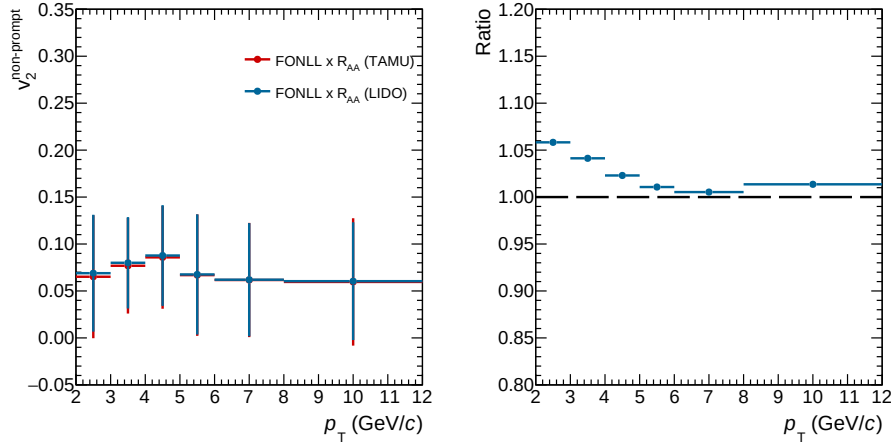


Figure 5.19: The non-prompt D^0 v_2 distribution as a function as p_T with the different p_T shape (right) and the ratio between non-prompt D^0 v_2 of LIDO model and the default value(left).

$$\frac{N_{OOP}}{N_{IP} + N_{OOP}} = \frac{1}{2} - \frac{2v_2}{\pi} \quad (5.8)$$

Consequently, we can obtain different ratios between N_{IP} (N_{OOP}) and $N_{IP} + N_{OOP}$ from charm and beauty hadron with different theoretical models. Figs. 5.20 shows the prompt D and B-meson v_2 as function as p_T predicted by TAMU (left) and LIDO (right) model. Figs. 5.21 shows the ratios between N_{IP} (N_{OOP}) and $N_{IP} + N_{OOP}$ with respect to charm and beauty hadron. The p_T shapes in plane and out of plane can be obtained from the default p_T shapes scaling with the ratios, the weights from the models shown in Figs. 5.22. Then the possible effect from the different p_T shapes of in and out plane is shown in Figs. 5.23 estimated by TAMU model and Figs. 5.24 estimated by LIDO model. The result indicates that the effect is minor (less than 1 %).

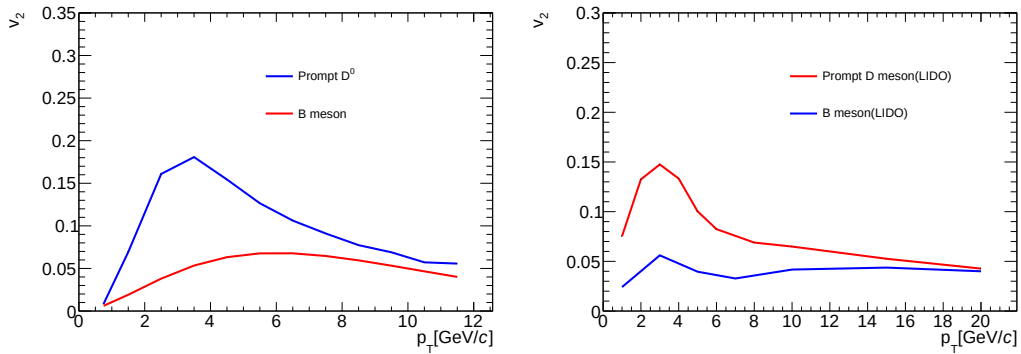


Figure 5.20: Prompt D and B-meson v_2 as function as p_T predicted by TAMU (left) and LIDO (right) model.

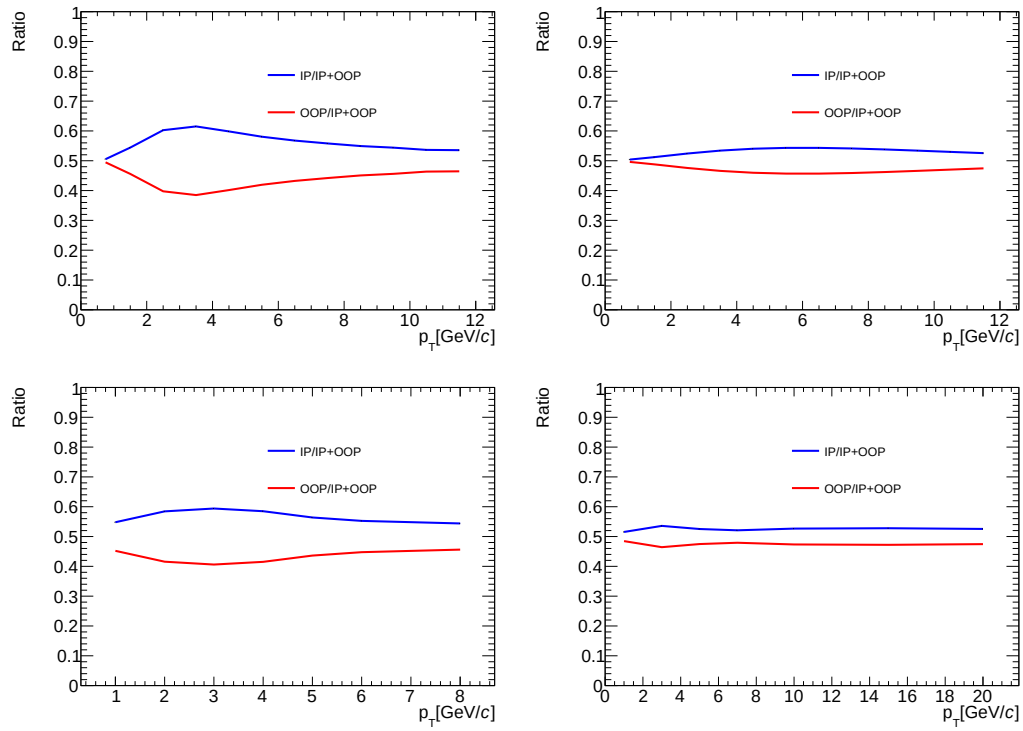


Figure 5.21: The ratios between N_{IP} (N_{OOP}) and $N_{IP} + N_{OOP}$ with respect to charm and beauty hadron as function as p_T predicted by TAMU (top) and LIDO (bottom) model.

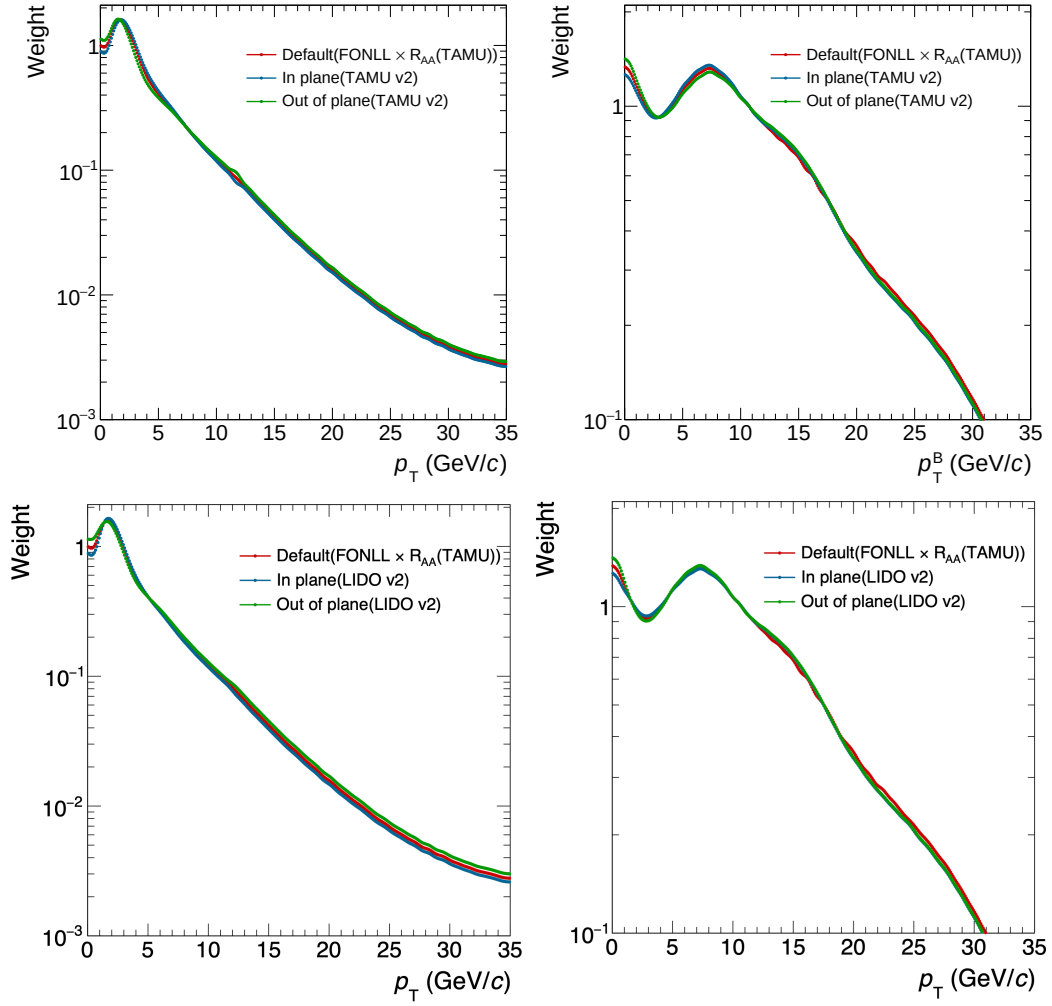


Figure 5.22: The p_T weights for in and out of plane for charm (left) and beauty (right) from TAMU (top) and LIDO (bottom) model.

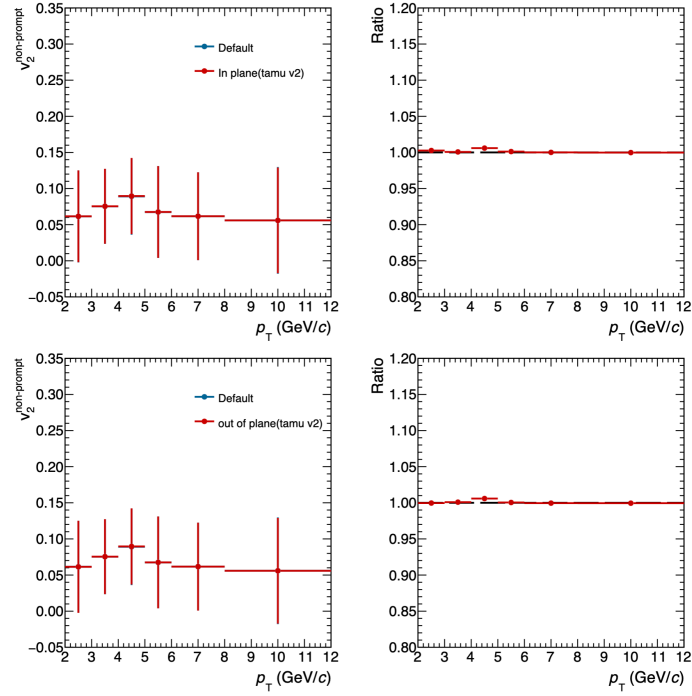


Figure 5.23: The possible effect from the different p_T shapes of in (top) and out (bottom) plane is estimated from TAMU model.

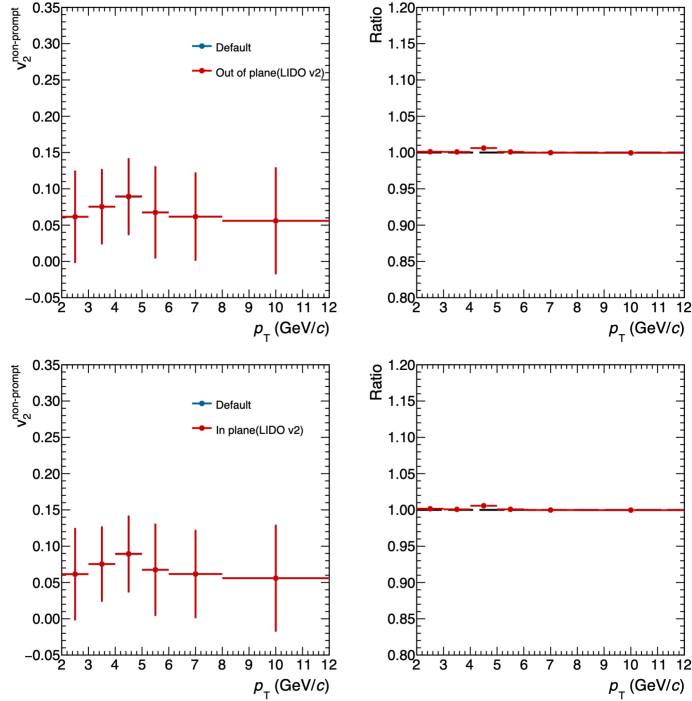


Figure 5.24: The possible effect from the different p_T shapes of in (top) and out (bottom) plane is estimated from LIDO model.



5.4.4 Scalar product denominator

The contribution of the SP denominator R_2 to the systematic uncertainty is due to the centrality dependence. It was evaluated as the difference of the centrality integrated R_2 values with those obtained from weighted average R_2 values in narrow centrality intervals using the D^0 -meson yields as weights [120].

5.5 Statistic uncertainty estimate using Jackknife method

Jackknife method [287] is applied to reduce the bias and to prove that interested estimation parameter is a true unbiased estimator of the population parameter samely interested. If there is a sample $x = (x_1, x_2, x_3, \dots, x_n)$ and an estimator $\hat{\theta} = s(x)$. The basic idea of the jackknife method is to calculate the estimator by sequentially deleting a single observation from the sample:

$$x_i = (x_1, x_2, x_3, \dots, x_{i-1}, x_{i+1}, \dots, x_n), \quad (5.9)$$

where $i = 1, 2, 3, \dots, n$, represent the jackknife samples. The i th jackknife sample consists of the data set with the i th observation removed. Let $\hat{\theta}_i = s(x_i)$ be the i th jackknife replication of $\hat{\theta}$. The jackknife estimate of standard error defined by

$$\sigma = \left[\frac{n-1}{n} \sum (\hat{\theta}_{(i)} - \hat{\theta}_{(.)})^2 \right]^{1/2} \quad (5.10)$$

where $\hat{\theta}_{(.)}$ is the mean of estimator. If the estimator is not smooth, the delete-1 jackknife method may lead to an inconsistent variance estimator [287]. Therefore, the delete-d jackknife method is developed to fix this problem. So the size of a delete-d jackknife sample is $(n-d)$, and there are $\binom{n}{d}$ jackknife samples. The standard error of delete-d jackknife given by

$$\sigma = \left[\frac{n-d}{d \binom{n}{d}} \sum (\hat{\theta}_{(i)} - \hat{\theta}_{(.)})^2 \right]^{1/2} \quad (5.11)$$

where $\hat{\theta}_{(.)}$ is the mean of estimator in all subsets, $\frac{n-d}{d \binom{n}{d}}$ is the scale factor of the RMS.

In this analysis, we re-sample the data with sub-samples randomly based on the FD score. The number of sub-samples varied with 12, 15, 20 and 40 shown in Figs. 5.25. Then we can repeat the analysis with the $\binom{n}{d}$ jackknife samples. Finally, the σ can be estimated by RMS of the non-prompt D^0 v_2 distribution of the new samples scaling with $\frac{n-d}{d \binom{n}{d}}$. The Full procedure is shown in Figs. 5.26.



More details can be found in Figs. 5.29 and Figs. 5.30 with respect to the non-prompt D^0 v_2 distribution using $\binom{20}{17}$ jackknife samples for correlated cut fits and the uncorrelated cut fits. Then to avoid the bias of RMS estimation, a Pol0 function was used to fit the non-prompt D^0 v_2 distribution as a function as trials (shown in Figs. 5.27 and Figs. 5.28). Therefore the RMS values were extracted from this fit.

The delete-d jackknife test is performed to check the statistic uncertainties in this analysis. We varied the deleted numbers and the size of sub-samples to estimate the statistic uncertainty for each p_T bin. Figs. 5.31 shows the statistic uncertainty as a function as p_T for each configuration. The black solid line represents the default statistic uncertainty from the uncorrelated cut fit. Although there are some fluctuations, We find the default values are not only consistent with the uncorrelated cut from the jackknife method but also the correlated cut from the jackknife method among all p_T intervals.

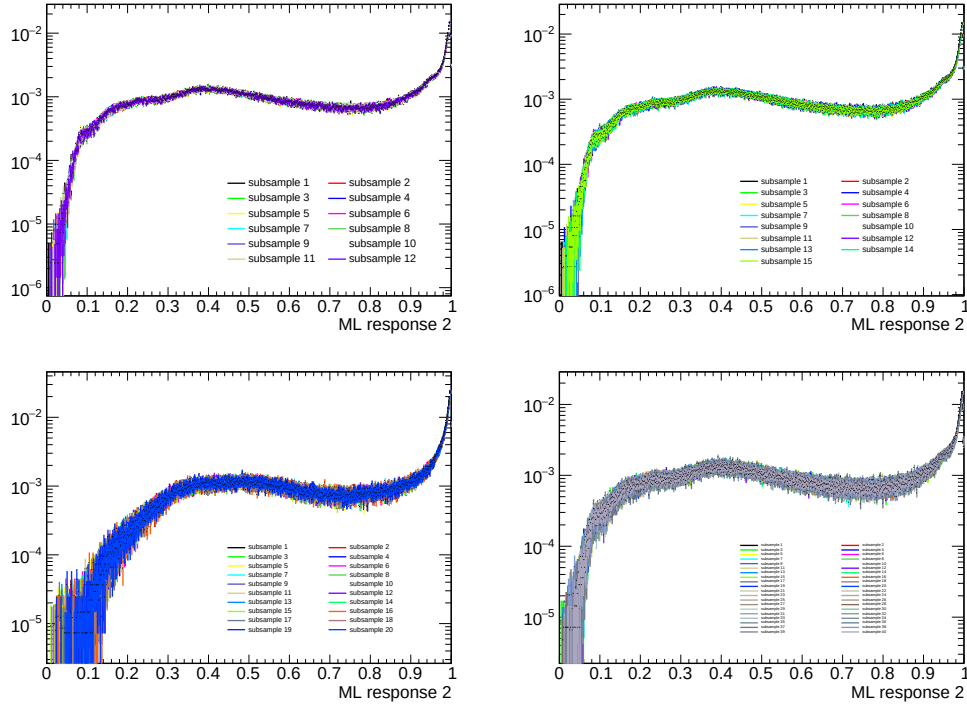


Figure 5.25: Different sub-samples used to the delete-d jackknife test.

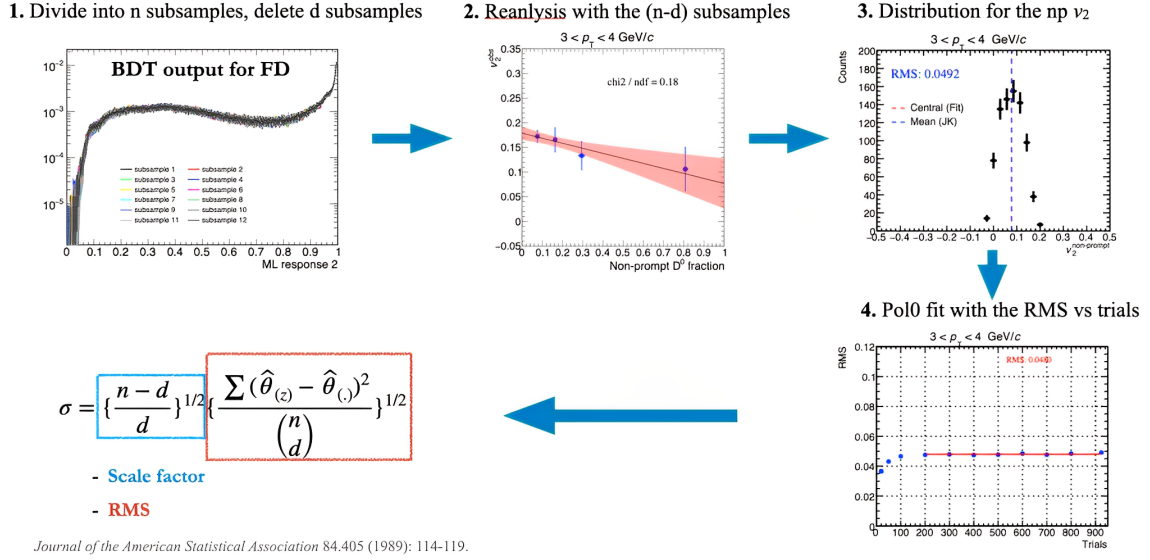


Figure 5.26: Full procedure of the delete-d jackknife test in non-prompt $D^0 v_2$ analysis.

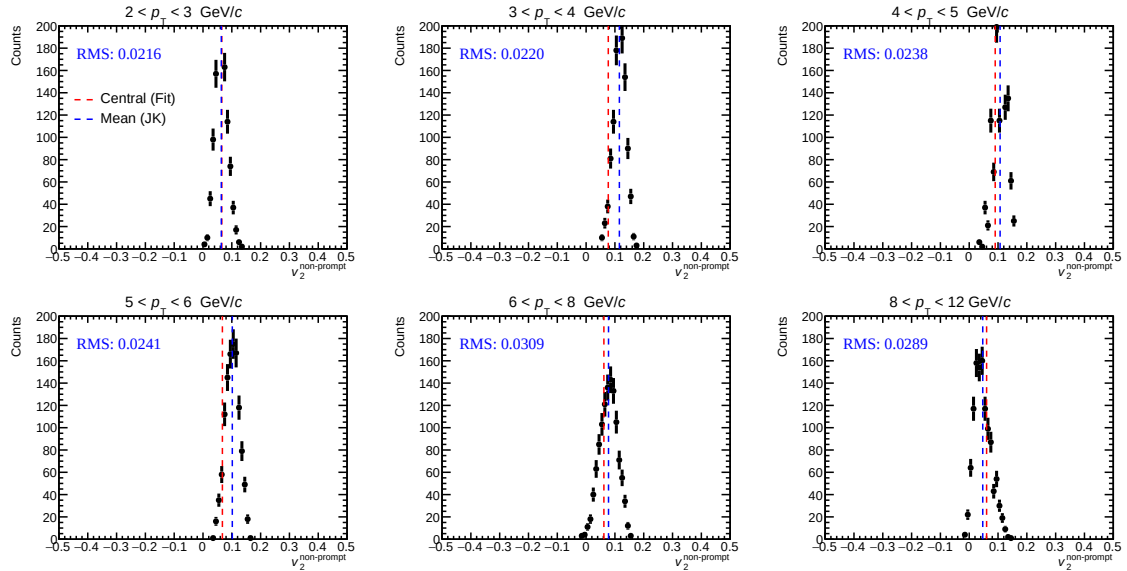


Figure 5.27: Non-prompt $D^0 v_2$ distribution using $\binom{20}{17}$ jackknife samples based on the correlated cut fit.

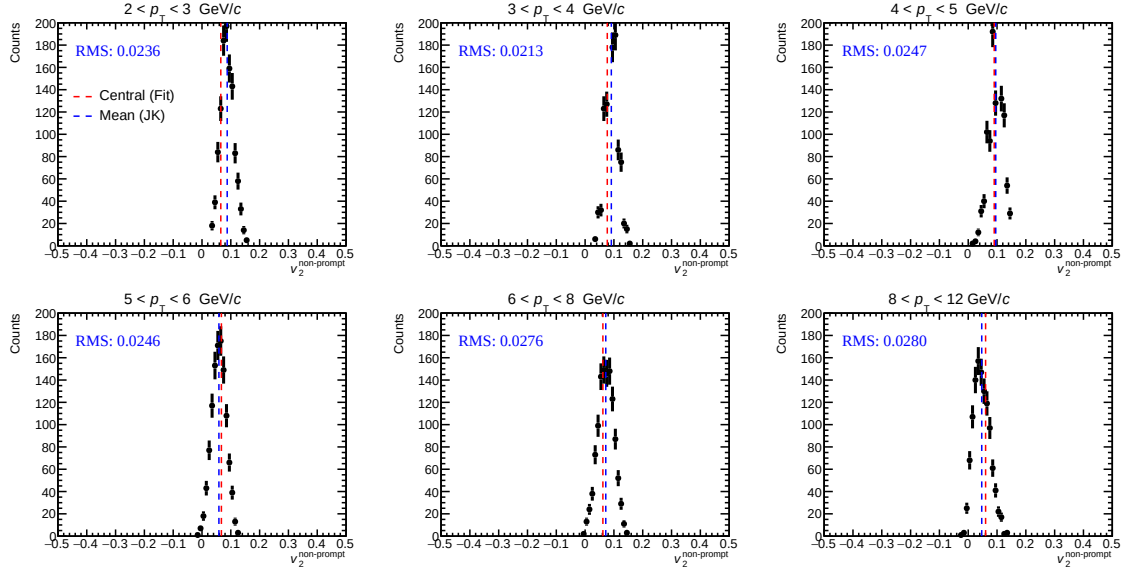


Figure 5.28: Non-prompt $D^0 v_2$ distribution using $\left(\frac{20}{17}\right)$ jackknife samples based on the uncorrelated cut fit.

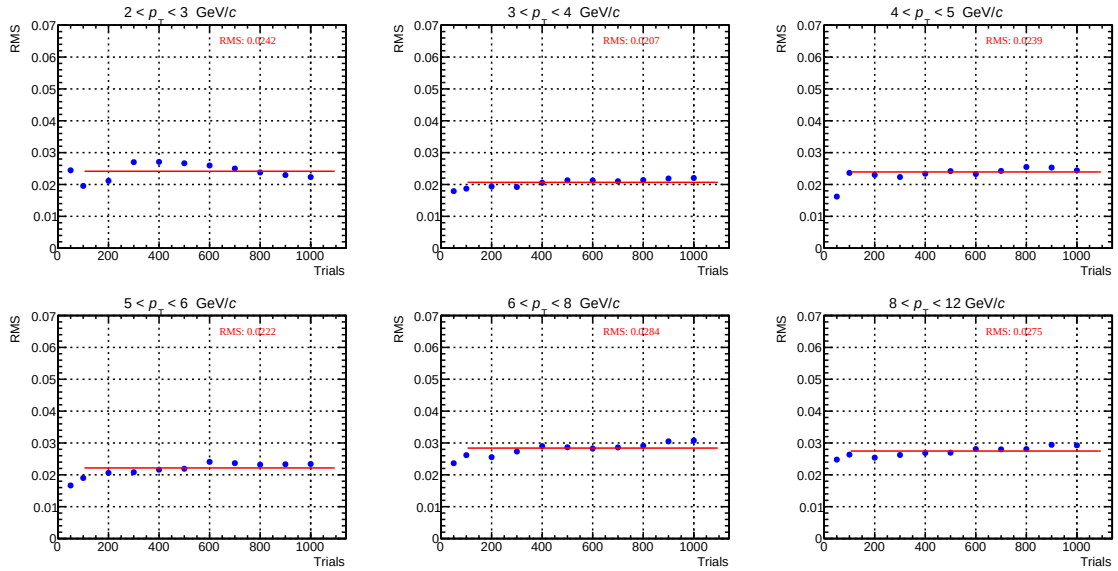


Figure 5.29: Non-prompt $D^0 v_2$ distribution as function as trials using $\left(\frac{20}{17}\right)$ jackknife samples based on the correlated cut fits.

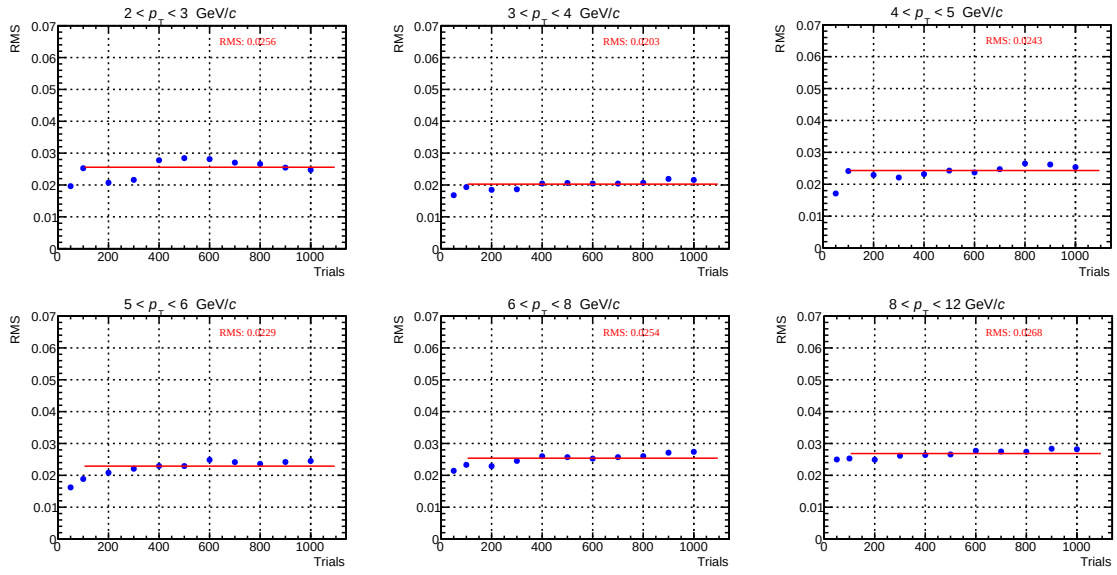


Figure 5.30: Non-prompt $D^0 v_2$ distribution as function as trials using $\binom{20}{17}$ jackknife samples based on the uncorrelated cut fits.

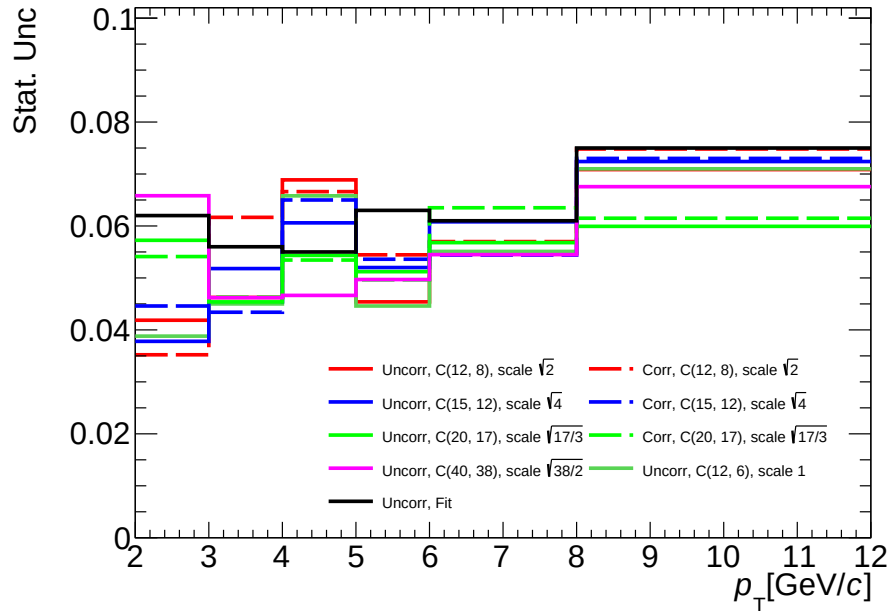


Figure 5.31: Non-prompt $D^0 v_2$ distribution as function as trials using $\binom{20}{17}$ jackknife samples based on the uncorrelated cut fits.



5.6 Results and discussion

The measurement of the non-prompt D^0 -meson elliptic flow is shown in Fig. 5.32 as a function of p_T at midrapidity ($|\eta| < 0.8$) in the 30–50% centrality class. the weighted mean of non-prompt D^0 -meson v_2 in the measured p_T range ($2 < p_T < 12$ GeV/c) is 2.7σ above 0. No significant p_T dependence of the v_2 is observed considering the uncertainties. In the left panel of Fig. 5.32, the non-prompt D^0 -meson v_2 is compared with the average of prompt non-strange D-meson v_2 [120]. The significance of the difference between these two measurements is 3.2σ in $2 < p_T < 8$ GeV/c, indicating a different degree of participation to the medium collective motion between charm and beauty quarks.

The measured v_2 of non-prompt D^0 mesons is compared with several theoretical models including beauty-quark transport in a hydrodynamically expanding medium [272, 286, 288–293] in the right panel of Fig. 5.32. All calculations include collisional interactions between beauty quarks and medium constituents. In addition, the LBT [272, 289], LIDO [286, 293], LGR [290], and Langevin [291, 292] models also include radiative processes. Beauty-quark hadronisation via coalescence is considered for all models, in addition to the fragmentation mechanism. Although the models are implemented with different assumptions on the interactions in the QGP and hadronic phases and on the medium expansion, all of them provide a fair description of the measurement within uncertainties.

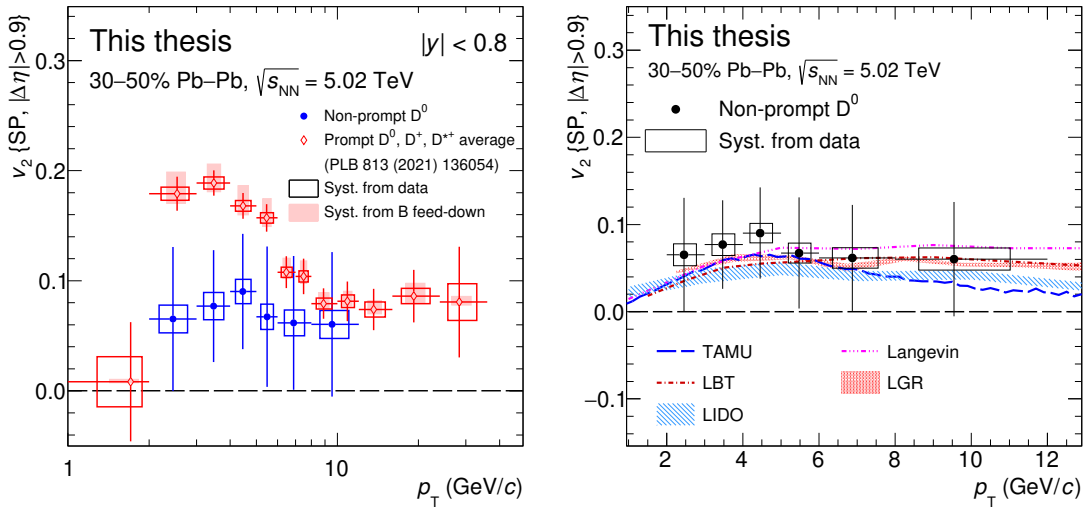


Figure 5.32: Left: non-prompt D^0 -meson (blue points) and prompt average non-strange D-meson (red points) v_2 as a function of p_T in 30–50% Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. Statistical uncertainties are shown as vertical lines and systematic uncertainties as boxes. Right: non-prompt D^0 -meson v_2 compared with model calculations [272, 286, 288–293].

Fig. 5.33 shows the comparison between the v_2 of electrons from beauty hadron decays and the non-prompt D^0 -meson v_2 measurements. They are compatible in the common p_T interval within un-

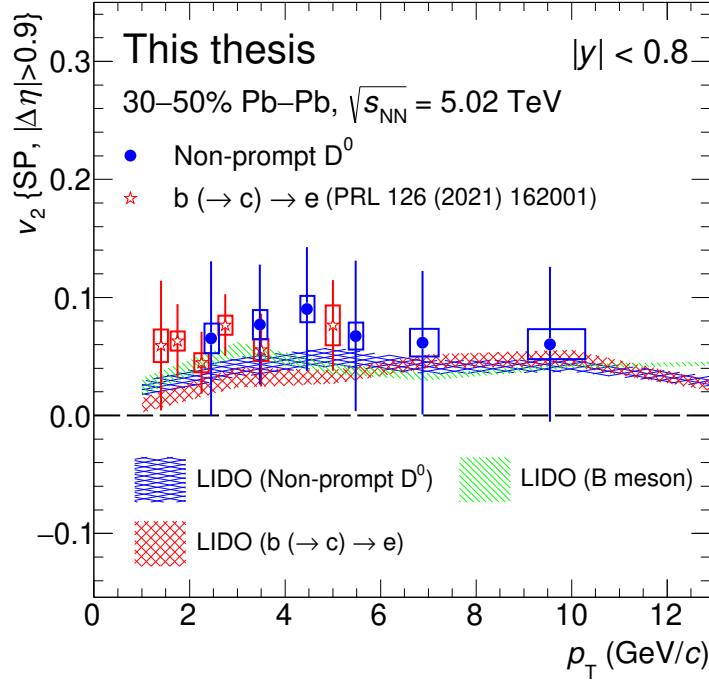


Figure 5.33: Non-prompt D^0 -meson (black points) and electrons from beauty hadron decay (cyan points) v_2 as a function of p_T in 30–50% Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, compared with the LIDO model predictions [286, 293].

certainties. The LIDO model provides reasonable descriptions for these measurements, reproducing the weak p_T dependence. The good agreement with the B-meson v_2 prediction from LIDO indicates that the decay kinematic effect is small for beauty-hadron v_2 measurements.

The simultaneous study of v_2 and R_{AA} can provide valuable insights into the interaction of beauty quarks with the QGP. The spatial diffusion coefficient (D_s) for heavy flavor, which characterizes their motion through a stochastic Langevin process [39], can be varied in different Langevin models to describe the data shown in Figs. 5.34. This allows for a more comprehensive investigation into the behavior of heavy quarks in the QGP. However the limited statistics from Run 2 data can't allow for firmer conclusions. The measurement is important for the understanding of the degree of thermalisation of beauty quarks in the QGP. Future data samples to be collected with the upgraded ALICE detector in Run 3, will allow for higher precision non-prompt D^0 -meson v_2 , together with the non-prompt D^0 -meson R_{AA} measurements [294].

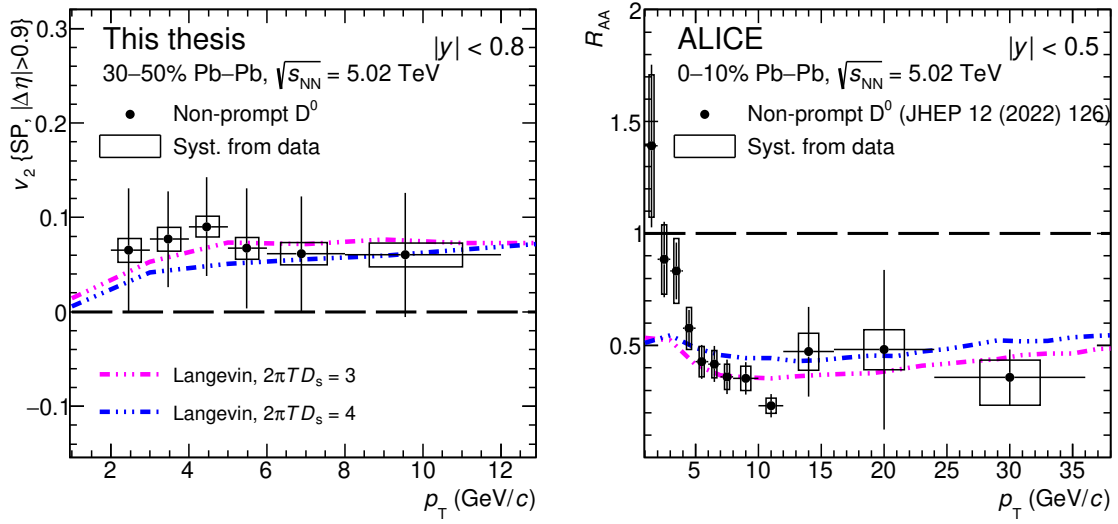


Figure 5.34: Non-prompt D^0 -meson v_2 and R_{AA} as a function of p_T in 30–50% Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, compared with the Langevin model predictions [291, 292].



6 Conclusions and Perspectives

The four data analyses presented in this thesis contribute to the investigation of the charm and beauty quark hadronization mechanism and dynamics in hadronic collisions at the LHC. The measurements related to open charm hadrons performed with the data in pp collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV and in Pb – Pb collisions at a centre-of-mass energy of $\sqrt{s_{NN}} = 5.02$ TeV. In the case of pp collisions, the size dependence of open heavy hadron production is studied through the measurement of the prompt Λ_c^+ baryon and non-prompt D-meson fraction as a function of multiplicity. The Λ_c^+ (D^0) candidates are reconstructed through the harmonic decay channels $\Lambda_c^+ \rightarrow pK_S^0$ ($D^0 \rightarrow K^-\pi^+$) in various multiplicity bins, which are estimated using the forward multiplicity estimators of the V0M detector. Additionally, the non-prompt D^0 production cross-section is also measured using minimum bias triggered events. For Pb – Pb collisions, the thesis presents the first measurement of non-prompt D^0 elliptic flow in ALICE. In all the measurements, the Machine-learning approach Boost Decision Tree (BDT) is adopted to suppress the combinatorial background and separate the prompt and non-prompt with the multi-classification algorithm provided by the XGBoost library for non-prompt D^0 meson analysis.

Together with the D_s^+ and D^0 , the first measurement of D_s^+/D^0 and Λ_c^+/D^0 ratios as a function of charged-particle multiplicity in pp collisions at $\sqrt{s} = 13$ TeV was presented. The p_T -differential D_s^+/D^0 yield ratio does not show a dependence on multiplicity, within uncertainties. In contrast, the charm baryon-to-meson ratio, Λ_c^+/D^0 , measured as a function of p_T , shows a significant increase (5.3σ) when comparing the measurements performed in the lowest and highest multiplicity intervals in $1 < p_T < 12$ GeV/ c . In addition, the Λ_c^+/D^0 ratio measured in the lowest multiplicity interval ($\langle dN_{ch}/d\eta \rangle = 3.1$) is higher, at low and intermediate p_T , than the values measured at e^+e^- colliders at lower centre-of-mass energies. These observations imply a modification of the hadronisation mechanisms that is collision-system and multiplicity dependent, further confirming the limited validity of the assumption of universality of the fragmentation functions. The measurements are compared with two calculations. Those based on PYTHIA with CR-BLC describe the D_s^+/D^0 measurements and capture the trend of the Λ_c^+/D^0 ratio, qualitatively describing the increasing magnitude of the baryon-to-meson ratios with multiplicity. Calculations based on a statistical hadronisation model, with the multiplicity dependence originating from the canonical treatment of quantum-charge conservation, describe the Λ_c^+/D^0 measurements in the lowest and highest multiplicity intervals. The prediction is also in agreement with the D_s^+/D^0 ratio for the low multiplicity interval, while it overestimates the data in the high-multiplicity class. The baryon-to-meson ratios in the charm sector, Λ_c^+/D^0 , are also compared to those in the light-flavour sector, Λ/K^0 ,



in similar multiplicity classes, showing a remarkably similar trend as a function of p_T and similar enhancement with $\langle dN_{ch}/d\eta \rangle$. These similarities hint at a possible common mechanism for light- and charm-baryon formation in pp collisions at LHC energies. The p_T -integrated Λ_c^+/D^0 ratios, extrapolated to $p_T = 0$ based on spectral shapes from PYTHIA with CR-BLC, show no significant dependence on multiplicity, suggesting that the increase in the baryon-to-meson yield ratio observed in the measured p_T range is due to a different redistribution of p_T between baryons and mesons, rather than to an enhancement in the overall baryon yield.

The fractions of the D^0 and D^+ mesons originating from beauty-hadron decays, $f_{\text{non-prompt}}$, were measured at midrapidity ($|y| < 0.5$) in pp collisions at $\sqrt{s} = 13$ TeV in events with at least a charged particle at midrapidity (INEL > 0 class of events) and as a function of charged-particle multiplicity and transverse momentum. Events with different charged-particle multiplicities were selected as percentiles of the INEL > 0 cross section based on the charged-particle multiplicity counts in the ALICE V0A and V0C at forward and backward rapidity. The D^+ and D^0 $f_{\text{non-prompt}}$ were observed to slightly increase from about 5%–7% for $1 < p_T < 3$ GeV/ c to about 10% for $8 < p_T < 24$ GeV/ c . The ratios $f_{\text{non-prompt}}^{\text{mult}}/f_{\text{non-prompt}}^{\text{INEL}>0}$ are compatible with unity both as a function of p_T and charged-particle multiplicity, suggesting either no or only a mild multiplicity dependence. This finding suggests a similar production mechanism of charm and beauty quarks as a function of multiplicity. The measured $f_{\text{non-prompt}}$ values are compared to predictions obtained with different MC generators. The EPOS 3 and EPOS 4 generators tend to underestimate the measurements, while PYTHIA 8 with different tunes, including the colour reconnection mechanism beyond leading colour approximation and colour ropes, slightly overestimates the data. The variation of $f_{\text{non-prompt}}$ with multiplicity is satisfactorily described by the MC simulations except for the $4 < p_T < 6$ GeV/ c interval, where the EPOS 3 generator predicts a significant increase. In all the considered p_T intervals, the CR-BLC Mode 3 tune of PYTHIA 8 foresees a decrease at high multiplicity. In that tune, this decrease with increasing multiplicity is motivated by an interplay between an increased fraction of charm quarks hadronising into mesons and an increased fraction of beauty quarks hadronising into baryons and is not favoured by data. The ratio $f_{\text{non-prompt}}^{\text{mult}}/f_{\text{non-prompt}}^{\text{INEL}>0}$ is also described well by the CGC model.

The p_T -differential and p_T -integrated production cross sections of non-prompt Λ_c^+ and D^0 hadrons were measured for the first time at midrapidity in pp collisions at $\sqrt{s} = 13$ TeV. The results are compatible with the theoretical models based on FONLL calculations and the PYTHIA 8 decayer. For the Λ_c^+ baryon, the beauty-quark fragmentation fraction measured by LHCb was used for the model prediction, suggesting a similar beauty-baryon enhancement at forward and midrapidity in pp collisions. The



extrapolated $b\bar{b}$ production cross section at midrapidity per unit of rapidity is found to be compatible with pQCD calculations with FONLL and NNLO accuracy. The measured baryon-to-meson ratios of strange, charm, and beauty hadrons are found to be similar, and in all cases significantly higher than the values measured in e^+e^- collisions. The p_T -differential baryon-to-meson ratios have been compared to predictions obtained with TAMU statistical hadronization model predictions and PYTHIA 8 simulations, that include the color-reconnection mechanism in the string fragmentation and indicate that all the flavors have to be considered simultaneously in order to obtain the best tuning of the model parameters involving the reconnection of quarks via junction topologies.

The non-prompt D^0 -meson v_2 in 30–50% Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV is measured in the transverse momentum interval $2 < p_T < 12$ GeV/ c . The non-prompt D^0 -meson v_2 is found to be positive with a significance of 2.7σ , and it is lower by 3.2σ than the non-strange D-meson v_2 in the range $2 < p_T < 8$ GeV/ c . The measurement is important for the understanding of the degree of thermalisation of beauty quarks in the QGP.

Despite the interesting insights provided by the results discussed in this thesis, a complete understanding of the charm and beauty quark hadronization mechanism and dynamics in hadronic collisions at the LHC requires further measurements. More precise measurements with the data sample collected in Run 3 of the LHC will allow us to further investigate with these prospects:

- Extend the multiplicity reach to lower and higher multiplicity intervals for the shape of the p_T -integrated baryon-to-meson ratios versus multiplicity.
- Further study of direct beauty-hadron production in pp collisions as a function of the charged-particle multiplicity.
- Direct measurement of beauty hadrons production cross section in the low p_T and mid-rapidity range.
- Beauty-meson and charm baryon elliptic flow in heavy-ion collision.



References

- [1] F. Halzen and Alan D. Martin. *QUARKS AND LEPTONS: AN INTRODUCTORY COURSE IN MODERN PARTICLE PHYSICS*. 1984.
- [2] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.
- [3] K. Melnikov. Lectures on QCD for hadron colliders. *CERN Yellow Rep. School Proc.*, 3:37–82, 2018.
- [4] Kenneth G. Wilson. Confinement of Quarks. *Phys. Rev. D*, 10:2445–2459, 1974.
- [5] P. A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020.
- [6] Helen Caines. The Search for Critical Behavior and Other Features of the QCD Phase Diagram – Current Status and Future Prospects. *Nucl. Phys. A*, 967:121–128, 2017.
- [7] A. Bazavov et al. Chiral crossover in QCD at zero and non-zero chemical potentials. *Phys. Lett. B*, 795:15–21, 2019.
- [8] Szabolcs Borsanyi, Zoltan Fodor, Jana N. Guenther, Ruben Kara, Sandor D. Katz, Paolo Parotto, Attila Pasztor, Claudia Ratti, and Kalman K. Szabo. QCD Crossover at Finite Chemical Potential from Lattice Simulations. *Phys. Rev. Lett.*, 125(5):052001, 2020.
- [9] Wit Busza, Krishna Rajagopal, and Wilke van der Schee. Heavy Ion Collisions: The Big Picture, and the Big Questions. *Ann. Rev. Nucl. Part. Sci.*, 68:339–376, 2018.
- [10] Peter Braun-Munzinger and Benjamin Dönigus. Loosely-bound objects produced in nuclear collisions at the LHC. *Nucl. Phys. A*, 987:144–201, 2019.
- [11] J. D. Bjorken. Highly Relativistic Nucleus-Nucleus Collisions: The Central Rapidity Region. *Phys. Rev. D*, 27:140–151, 1983.
- [12] Jean-Paul Blaizot. High gluon densities in heavy ion collisions. *Rept. Prog. Phys.*, 80(3):032301, 2017.
- [13] F. Gelis. Color Glass Condensate and Glasma. *Int. J. Mod. Phys. A*, 28:1330001, 2013.
- [14] Reinhard Stock. Relativistic Nucleus-Nucleus Collisions and the QCD Matter Phase Diagram. 7 2008.



-
- [15] Jaroslav Adam et al. Measurement of transverse energy at midrapidity in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Rev. C*, 94(3):034903, 2016.
- [16] Paul Romatschke and Ulrike Romatschke. *Relativistic Fluid Dynamics In and Out of Equilibrium*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 5 2019.
- [17] K. Aamodt et al. Two-pion Bose-Einstein correlations in central Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Lett. B*, 696:328–337, 2011.
- [18] The ALICE experiment – A journey through QCD. 11 2022.
- [19] K. Adcox et al. Suppression of hadrons with large transverse momentum in central Au+Au collisions at $\sqrt{s_{NN}} = 130$ -GeV. *Phys. Rev. Lett.*, 88:022301, 2002.
- [20] R. J. Fries, Berndt Muller, C. Nonaka, and S. A. Bass. Hadron production in heavy ion collisions: Fragmentation and recombination from a dense parton phase. *Phys. Rev. C*, 68:044902, 2003.
- [21] Jean-Yves Ollitrault. Anisotropy as a signature of transverse collective flow. *Phys. Rev.*, D46:229–245, 1992.
- [22] Michael Annan Lisa, Scott Pratt, Ron Soltz, and Urs Wiedemann. Femtoscopy in relativistic heavy ion collisions. *Ann. Rev. Nucl. Part. Sci.*, 55:357–402, 2005.
- [23] Jaroslav Adam et al. Centrality dependence of the charged-particle multiplicity density at midrapidity in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Rev. Lett.*, 116(22):222302, 2016.
- [24] Serguei Chatrchyan et al. Measurement of the pseudorapidity and centrality dependence of the transverse energy density in PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Rev. Lett.*, 109:152303, 2012.
- [25] S. S. Adler et al. Systematic studies of the centrality and $s(NN)^{1/2}$ dependence of the $d E(T) / d \eta$ and $d(N(ch) / d \eta$ in heavy ion collisions at mid-rapidity. *Phys. Rev. C*, 71:034908, 2005. [Erratum: *Phys.Rev.C* 71, 049901 (2005)].
- [26] Ulrich W. Heinz. Concepts of heavy ion physics. In *2nd CERN-CLAF School of High Energy Physics*, pages 165–238, 7 2004.
- [27] Giorgio Torrieri, S. Steinke, Wojciech Broniowski, Wojciech Florkowski, Jean Letessier, and Johann Rafelski. SHARE: Statistical hadronization with resonances. *Comput. Phys. Commun.*, 167:229–251, 2005.



- [28] S. Wheaton and J. Cleymans. THERMUS: A Thermal model package for ROOT. *Comput. Phys. Commun.*, 180:84–106, 2009.
- [29] A. Andronic, P. Braun-Munzinger, and J. Stachel. Hadron production in central nucleus-nucleus collisions at chemical freeze-out. *Nucl. Phys. A*, 772:167–199, 2006.
- [30] S. Acharya et al. Transverse momentum spectra and nuclear modification factors of charged particles in pp, p-Pb and Pb-Pb collisions at the LHC. *JHEP*, 11:013, 2018.
- [31] J. W. Cronin, Henry J. Frisch, M. J. Shochet, J. P. Boymond, R. Mermod, P. A. Piroue, and Richard L. Sumner. Production of hadrons with large transverse momentum at 200, 300, and 400 GeV. *Phys. Rev. D*, 11:3105–3123, 1975.
- [32] Man Xie, Shu-Yi Wei, Guang-You Qin, and Han-Zhong Zhang. Extracting jet transport coefficient via single hadron and dihadron productions in high-energy heavy-ion collisions. *Eur. Phys. J. C*, 79(7):589, 2019.
- [33] S. Voloshin and Y. Zhang. Flow study in relativistic nuclear collisions by Fourier expansion of Azimuthal particle distributions. *Z. Phys.*, C70:665–672, 1996.
- [34] Burak Han Alver, Clement Gombeaud, Matthew Luzum, and Jean-Yves Ollitrault. Triangular flow in hydrodynamics and transport theory. *Phys. Rev. C*, 82:034913, 2010.
- [35] Derek Teaney and Li Yan. Non linearities in the harmonic spectrum of heavy ion collisions with ideal and viscous hydrodynamics. *Phys. Rev. C*, 86:044908, 2012.
- [36] Scott McDonald, Chun Shen, Francois Fillion-Gourdeau, Sangyong Jeon, and Charles Gale. Hydrodynamic predictions for Pb+Pb collisions at 5.02 TeV. *Phys. Rev. C*, 95(6):064913, 2017.
- [37] Wenbin Zhao, Hao-jie Xu, and Huichao Song. Collective flow in 2.76 A TeV and 5.02 A TeV Pb+Pb collisions. *Eur. Phys. J. C*, 77(9):645, 2017.
- [38] S. Acharya et al. Anisotropic flow of identified particles in Pb-Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV. *JHEP*, 09:006, 2018.
- [39] Xin Dong, Yen-Jie Lee, and Ralf Rapp. Open Heavy-Flavor Production in Heavy-Ion Collisions. *Ann. Rev. Nucl. Part. Sci.*, 69:417–445, 2019.
- [40] John C. Collins, Davison E. Soper, and George F. Sterman. Factorization of Hard Processes in QCD. *Adv. Ser. Direct. High Energy Phys.*, 5:1–91, 1989.



-
- [41] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, Pavel M. Nadolsky, and W. K. Tung. New generation of parton distributions with uncertainties from global QCD analysis. *JHEP*, 07:012, 2002.
- [42] Sayipjamal Dulat, Tie-Jiun Hou, Jun Gao, Marco Guzzi, Joey Huston, Pavel Nadolsky, Jon Pumplin, Carl Schmidt, Daniel Stump, and C. P. Yuan. New parton distribution functions from a global analysis of quantum chromodynamics. *Phys. Rev. D*, 93(3):033006, 2016.
- [43] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J. C*, 77(10):663, 2017.
- [44] V.G. Kartvelishvili, A.K. Likhoded, and V.A. Petrov. On the fragmentation functions of heavy quarks into hadrons. *Physics Letters B*, 78(5):615–617, 1978.
- [45] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys. B*, 126:298–318, 1977.
- [46] M. Benzke, M.V. Garzelli, B. Kniehl, G. Kramer, S. Moch, and G. Sigl. Prompt neutrinos from atmospheric charm in the general-mass variable-flavor-number scheme. *JHEP*, 12:021, 2017.
- [47] 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. *Nucl. Phys. B*, 925:415–430, 2017.
- [48] Matteo Cacciari, Stefano Frixione, Nicolas Houdeau, Michelangelo L. Mangano, Paolo Nason, and Giovanni Ridolfi. Theoretical predictions for charm and bottom production at the LHC. *JHEP*, 10:137, 2012.
- [49] Matteo Cacciari, Mario Greco, and Paolo Nason. The P(T) spectrum in heavy flavor hadroproduction. *JHEP*, 05:007, 1998.
- [50] David d’Enterria and Alexander M. Snigirev. Triple parton scatterings in high-energy proton-proton collisions. *Phys. Rev. Lett.*, 118(12):122001, 2017.
- [51] David d’Enterria and Alexander M. Snigirev. Triple-parton scatterings in proton–nucleus collisions at high energies. *Eur. Phys. J. C*, 78(5):359, 2018.
- [52] S. Katuwal and R. P. Woodard. Gauge independent quantum gravitational corrections to Maxwell’s equation. *JHEP*, 21:029, 2020.
- [53] Shreyasi Acharya et al. Charm-quark fragmentation fractions and production cross section at midrapidity in pp collisions at the LHC. *Phys. Rev. D*, 105(1):L011103, 2022.



-
- [54] S. G. Weber, A. Dubla, A. Andronic, and A. Morsch. Elucidating the multiplicity dependence of J/ψ production in proton–proton collisions with PYTHIA8. *Eur. Phys. J.*, C79(1):36, 2019.
- [55] M. Bahr et al. Herwig++ Physics and Manual. *Eur. Phys. J.*, C58:639–707, 2008.
- [56] Stefano Frixione and Bryan R. Webber. Matching NLO QCD computations and parton shower simulations. *JHEP*, 06:029, 2002.
- [57] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.
- [58] Matteo Cacciari, Stefano Frixione, and Paolo Nason. The p_T spectrum in heavy flavor photoproduction. *JHEP*, 03:006, 2001.
- [59] B.A. Kniehl, G. Kramer, I. Schienbein, and H. Spiesberger. Inclusive $D^{*\pm}$ production in $p\bar{p}$ collisions with massive charm quarks. *Phys. Rev. D*, 71:014018, 2005.
- [60] B.A. Kniehl, G. Kramer, I. Schienbein, and H. Spiesberger. Inclusive Charmed-Meson Production at the CERN LHC. *Eur. Phys. J. C*, 72:2082, 2012.
- [61] Ilkka Helenius and Hannu Paukkunen. Revisiting the D-meson hadroproduction in general-mass variable flavour number scheme. *JHEP*, 05:196, 2018.
- [62] Paolo Bolzoni and Gustav Kramer. Inclusive charmed-meson production from bottom hadron decays at the LHC. *J. Phys. G*, 41:075006, 2014.
- [63] Shreyasi Acharya et al. Measurement of beauty and charm production in pp collisions at $\sqrt{s} = 5.02$ TeV via non-prompt and prompt D mesons. *JHEP*, 05:220, 2021.
- [64] Rafał Maciula and Antoni Szczurek. Production of Λ_c baryons at the LHC within the k_T -factorization approach and independent parton fragmentation picture. *Phys. Rev.*, D98(1):014016, 2018.
- [65] Shreyasi Acharya et al. Λ_c^+ production in pp collisions at $\sqrt{s} = 7$ TeV and in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *JHEP*, 04:108, 2018.
- [66] Albert M Sirunyan et al. Production of Λ_c^+ baryons in proton-proton and lead-lead collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Lett. B*, 803:135328, 2020.
- [67] Shreyasi Acharya et al. Measurement of Prompt D^0 , Λ_c^+ , and $\Sigma_c^{0,++}(2455)$ Production in Proton–Proton Collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 128(1):012001, 2022.



- [68] Roel Aaij et al. Prompt Λ_c^+ production in p Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *JHEP*, 02:102, 2019.
- [69] Jaroslav Adam et al. Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions. *Nature Phys.*, 13:535–539, 2017.
- [70] Nestor Armesto. Nuclear shadowing. *J. Phys. G*, 32:R367–R394, 2006.
- [71] Michele Arneodo. Nuclear effects in structure functions. *Phys. Rept.*, 240:301–393, 1994.
- [72] Donald F. Geesaman, K. Saito, and Anthony William Thomas. The nuclear EMC effect. *Ann. Rev. Nucl. Part. Sci.*, 45:337–390, 1995.
- [73] K. J. Eskola, H. Paukkunen, and C. A. Salgado. EPS09: A New Generation of NLO and LO Nuclear Parton Distribution Functions. *JHEP*, 04:065, 2009.
- [74] A. Andronic et al. Heavy-flavour and quarkonium production in the LHC era: from proton–proton to heavy-ion collisions. *Eur. Phys. J.*, C76(3):107, 2016.
- [75] Alberto Accardi. Cronin effect in proton nucleus collisions: A Survey of theoretical models. 12 2002.
- [76] Ivan Vitev. Non-Abelian energy loss in cold nuclear matter. *Phys. Rev. C*, 75:064906, 2007.
- [77] Serguei Chatrchyan et al. Observation of Long-Range Near-Side Angular Correlations in Proton-Lead Collisions at the LHC. *Phys. Lett. B*, 718:795–814, 2013.
- [78] Vardan Khachatryan et al. Evidence for Collective Multiparticle Correlations in p-Pb Collisions. *Phys. Rev. Lett.*, 115(1):012301, 2015.
- [79] Betty Bezverkhny Abelev et al. Multiplicity Dependence of Pion, Kaon, Proton and Lambda Production in p-Pb Collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Lett. B*, 728:25–38, 2014.
- [80] Anne M. Sickles. Possible Evidence for Radial Flow of Heavy Mesons in d+Au Collisions. *Phys. Lett. B*, 731:51–56, 2014.
- [81] Yingru Xu, Shanshan Cao, Guang-You Qin, Weiyao Ke, Marlene Nahrgang, Jussi Auvinen, and Steffen A. Bass. Heavy-flavor dynamics in relativistic p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Nucl. Part. Phys. Proc.*, 276-278:225–228, 2016.
- [82] Andrea Beraudo, Arturo De Pace, Marco Monteno, Marzia Nardi, and Francesco Prino. Heavy-flavour production in high-energy d-Au and p-Pb collisions. *JHEP*, 03:123, 2016.



- [83] Yang-Ting Chien, Alexander Emerman, Zhong-Bo Kang, Grigory Ovanesyan, and Ivan Vitev. Jet Quenching from QCD Evolution. *Phys. Rev. D*, 93(7):074030, 2016.
- [84] Shreyasi Acharya et al. Measurement of prompt D^0 , D^+ , D^{*+} , and D_s^+ production in p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *JHEP*, 12:092, 2019.
- [85] First measurement of Λ_c^+ production down to $p_T = 0$ in pp and p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. 11 2022.
- [86] J. Adams et al. Identified hadron spectra at large transverse momentum in p+p and d+Au collisions at $s(NN)^{1/2} = 200$ -GeV. *Phys. Lett. B*, 637:161–169, 2006.
- [87] Jaroslav Adam et al. Multiplicity dependence of charged pion, kaon, and (anti)proton production at large transverse momentum in p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Lett. B*, 760:720–735, 2016.
- [88] Eric Braaten and Markus H. Thoma. Energy loss of a heavy quark in the quark–gluon plasma. *Phys. Rev.*, D44(9):R2625, 1991.
- [89] A. Adare et al. Heavy Quark Production in $p + p$ and Energy Loss and Flow of Heavy Quarks in Au+Au Collisions at $\sqrt{s_{NN}} = 200$ GeV. *Phys. Rev.*, C84:044905, 2011.
- [90] B.I. Abelev et al. Transverse momentum and centrality dependence of high- p_T non-photonic electron suppression in Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV. *Phys. Rev. Lett.*, 98:192301, 2007. [Erratum: *Phys.Rev.Lett.* 106, 159902 (2011)].
- [91] L. Adamczyk et al. Observation of D^0 Meson Nuclear Modifications in Au+Au Collisions at $\sqrt{s_{NN}} = 200$ GeV. *Phys. Rev. Lett.*, 113(14):142301, 2014. [Erratum: *Phys.Rev.Lett.* 121, 229901 (2018)].
- [92] S. S. Adler et al. Nuclear modification of electron spectra and implications for heavy quark energy loss in Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV. *Phys. Rev. Lett.*, 96:032301, 2006.
- [93] A. Adare et al. Single electron yields from semileptonic charm and bottom hadron decays in Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV. *Phys. Rev.*, C93(3):034904, 2016.
- [94] Jaroslav Adam et al. Transverse momentum dependence of D-meson production in Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *JHEP*, 03:081, 2016.
- [95] Betty Abelev et al. Production of muons from heavy flavour decays at forward rapidity in pp and Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Rev. Lett.*, 109:112301, 2012.



-
- [96] Jaroslav Adam et al. Measurement of the production of high- p_T electrons from heavy-flavour hadron decays in Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Lett.*, B771:467–481, 2017.
- [97] Jaroslav Adam et al. Measurement of electrons from beauty-hadron decays in p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV and Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *JHEP*, 07:052, 2017.
- [98] Vardan Khachatryan et al. Suppression and azimuthal anisotropy of prompt and nonprompt J/ψ production in PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Eur. Phys. J.*, C77(4):252, 2017.
- [99] Albert M Sirunyan et al. Nuclear modification factor of D^0 mesons in PbPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Lett.*, B782:474–496, 2018.
- [100] S. Acharya et al. Measurement of D^0 , D^+ , D^{*+} and D_s^+ production in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *JHEP*, 10:174, 2018.
- [101] Albert M Sirunyan et al. Measurement of B_s^0 meson production in pp and PbPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Lett. B*, 796:168–190, 2019.
- [102] Albert M Sirunyan et al. Measurement of the B^+ Meson Nuclear Modification Factor in Pb–Pb Collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Rev. Lett.*, 119(15):152301, 2017.
- [103] Guy D. Moore and Derek Teaney. How much do heavy quarks thermalize in a heavy ion collision? *Phys. Rev.*, C71:064904, 2005.
- [104] Shanshan Cao, Guang-You Qin, and Steffen A. Bass. Heavy-quark dynamics and hadronization in ultrarelativistic heavy-ion collisions: Collisional versus radiative energy loss. *Phys. Rev.*, C88:044907, 2013.
- [105] S. Peigne and A. V. Smilga. Energy losses in a hot plasma revisited. *Phys. Usp.*, 52:659–685, 2009.
- [106] Carlos A. Salgado and Urs Achim Wiedemann. Calculating quenching weights. *Phys. Rev. D*, 68:014008, 2003.
- [107] Yuri L. Dokshitzer and D. E. Kharzeev. Heavy quark colorimetry of QCD matter. *Phys. Lett. B*, 519:199–206, 2001.
- [108] Shreyasi Acharya et al. Prompt D^0 , D^+ , and D^{*+} production in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *JHEP*, 01:174, 2022.
- [109] Shreyasi Acharya et al. Production of charged pions, kaons, and (anti-)protons in Pb–Pb and inelastic pp collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Rev. C*, 101(4):044907, 2020.



- [110] Shreyasi Acharya et al. Centrality and transverse momentum dependence of inclusive J/ψ production at midrapidity in Pb–Pb collisions at $\sqrt{s_{NN}}=5.02$ TeV. *Phys. Lett. B*, 805:135434, 2020.
- [111] Albert M. Sirunyan et al. Measurement of prompt and nonprompt charmonium suppression in PbPb collisions at 5.02 TeV. *Eur. Phys. J. C*, 78(6):509, 2018.
- [112] M. Gyulassy, I. Vitev, and X. N. Wang. High $p(T)$ azimuthal asymmetry in noncentral A+A at RHIC. *Phys. Rev. Lett.*, 86:2537–2540, 2001.
- [113] E. V. Shuryak. The Azimuthal asymmetry at large p_T seem to be too large for a ‘jet quenching’. *Phys. Rev.*, C66:027902, 2002.
- [114] Andrea Beraudo, Arturo De Pace, Marco Monteno, Marzia Nardi, and Francesco Prino. Development of heavy-flavour flow-harmonics in high-energy nuclear collisions. *JHEP*, 02:043, 2018.
- [115] Marlene Nahrgang, Joerg Aichelin, Pol Bernard Gossiaux, and Klaus Werner. Influence of hadronic bound states above T_c on heavy-quark observables in Pb + Pb collisions at the CERN Large Hadron Collider. *Phys. Rev.*, C89(1):014905, 2014.
- [116] S. Batsouli, S. Kelly, M. Gyulassy, and J. L. Nagle. Does the charm flow at RHIC? *Phys. Lett.*, B557:26–32, 2003.
- [117] Denes Molnar. Charm elliptic flow from quark coalescence dynamics. *J. Phys.*, G31:S421–S428, 2005.
- [118] V. Greco, C. M. Ko, and R. Rapp. Quark coalescence for charmed mesons in ultrarelativistic heavy ion collisions. *Phys. Lett.*, B595:202–208, 2004.
- [119] Marlene Nahrgang, Jörg Aichelin, Steffen Bass, Pol Bernard Gossiaux, and Klaus Werner. Elliptic and triangular flow of heavy flavor in heavy-ion collisions. *Phys. Rev.*, C91(1):014904, 2015.
- [120] Shreyasi Acharya et al. Transverse-momentum and event-shape dependence of D-meson flow harmonics in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Lett. B*, 813:136054, 2021.
- [121] S. Acharya et al. Anisotropic flow of identified particles in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *JHEP*, 09:006, 2018.
- [122] Shreyasi Acharya et al. J/ψ elliptic and triangular flow in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. 5 2020.
- [123] Francesco Prino and Ralf Rapp. Open Heavy Flavor in QCD Matter and in Nuclear Collisions. *J. Phys.*, G43(9):093002, 2016.



- [124] Ralf Rapp and Hendrik van Hees. Heavy Quarks in the Quark-Gluon Plasma. pages 111–206, 2010.
- [125] Shreyasi Acharya et al. D-meson azimuthal anisotropy in midcentral Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Rev. Lett.*, 120(10):102301, 2018.
- [126] A. Chodos, R. L. Jaffe, K. Johnson, and Charles B. Thorn. Baryon Structure in the Bag Theory. *Phys. Rev. D*, 10:2599, 1974.
- [127] Thomas A. DeGrand, R. L. Jaffe, K. Johnson, and J. E. Kiskis. Masses and Other Parameters of the Light Hadrons. *Phys. Rev. D*, 12:2060, 1975.
- [128] X. Artru and G. Mennessier. String model and multiproduction. *Nucl. Phys. B*, 70:93–115, 1974.
- [129] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton Fragmentation and String Dynamics. *Phys. Rept.*, 97:31–145, 1983.
- [130] Silvia Ferreres-Solé and Torbjörn Sjöstrand. The space–time structure of hadronization in the Lund model. *Eur. Phys. J. C*, 78(11):983, 2018.
- [131] Peter Skands, Stefano Carrazza, and Juan Rojo. Tuning PYTHIA 8.1: the Monash 2013 Tune. *Eur. Phys. J. C*, 74(8):3024, 2014.
- [132] Michael H. Seymour and Marilyn Marx. Monte Carlo Event Generators. In *69th Scottish Universities Summer School in Physics: LHC Physics*, pages 287–319, 4 2013.
- [133] Andy Buckley et al. General-purpose event generators for LHC physics. *Phys. Rept.*, 504:145–233, 2011.
- [134] Mykhailo Lisovyi, Andrii Verbytskyi, and Oleksandr Zenaiev. Combined analysis of charm-quark fragmentation-fraction measurements. *Eur. Phys. J. C*, 76(7):397, 2016.
- [135] Vincenzo Minissale, Salvatore Plumari, and Vincenzo Greco. Charm hadrons in pp collisions at LHC energy within a coalescence plus fragmentation approach. *Phys. Lett. B*, 821:136622, 2021.
- [136] Min He and Ralf Rapp. Charm-Baryon Production in Proton-Proton Collisions. *Phys. Lett. B*, 795:117–121, 2019.
- [137] Jesper R. Christiansen and Peter Z. Skands. String Formation Beyond Leading Colour. *JHEP*, 08:003, 2015.



- [138] G. Agakishiev et al. Strangeness Enhancement in Cu+Cu and Au+Au Collisions at $\sqrt{s_{NN}} = 200$ GeV. *Phys. Rev. Lett.*, 108:072301, 2012.
- [139] Betty Bezverkhny Abelev et al. K_S^0 and Λ production in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Rev. Lett.*, 111:222301, 2013.
- [140] Jaroslav Adam et al. First measurement of Λ_c baryon production in Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV. *Phys. Rev. Lett.*, 124(17):172301, 2020.
- [141] Shreyasi Acharya et al. Constraining hadronization mechanisms with Λ_c^+/D^0 production ratios in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. 12 2021.
- [142] Peter Braun-Munzinger, Krzysztof Redlich, and Johanna Stachel. Particle production in heavy ion collisions. pages 491–599, 4 2003.
- [143] Anton Andronic, Peter Braun-Munzinger, Krzysztof Redlich, and Johanna Stachel. Decoding the phase structure of QCD via particle production at high energy. *Nature*, 561(7723):321–330, 2018.
- [144] A. Andronic, P. Braun-Munzinger, J. Stachel, and M. Winn. Interacting hadron resonance gas meets lattice QCD. *Phys. Lett. B*, 718:80–85, 2012.
- [145] Shreyasi Acharya et al. Measurement of D^0 , D^+ , D^{*+} and D_s^+ production in pp collisions at $\sqrt{s} = 5.02$ TeV with ALICE. *Eur. Phys. J. C*, 79(5):388, 2019.
- [146] Shreyasi Acharya et al. Prompt and non-prompt J/ψ production cross sections at midrapidity in proton-proton collisions at $\sqrt{s} = 5.02$ and 13 TeV. *JHEP*, 03:190, 2022.
- [147] Shreyasi Acharya et al. Measurement of electrons from semileptonic heavy-flavour hadron decays at midrapidity in pp and Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Lett. B*, 804:135377, 2020.
- [148] Vardan Khachatryan et al. Measurement of the total and differential inclusive B^+ hadron cross sections in pp collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 771:435–456, 2017.
- [149] Albert M Sirunyan et al. Nuclear modification factor of D^0 mesons in PbPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Lett. B*, 782:474–496, 2018.
- [150] Albert M Sirunyan et al. Studies of Beauty Suppression via Nonprompt D^0 Mesons in Pb-Pb Collisions at $Q^2 = 4$ GeV². *Phys. Rev. Lett.*, 123(2):022001, 2019.
- [151] Roel Aaij et al. Measurements of prompt charm production cross-sections in pp collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 03:159, 2016. [Erratum: JHEP 09, 013 (2016), Erratum: JHEP 05, 074 (2017)].



- [152] Roel Aaij et al. Measurements of prompt charm production cross-sections in pp collisions at $\sqrt{s} = 5$ TeV. *JHEP*, 06:147, 2017.
- [153] Roel Aaij et al. Measurement of the $B^\pm$ production cross-section in pp collisions at $\sqrt{s} = 7$ and 13 TeV. *JHEP*, 12:026, 2017.
- [154] Georges Aad et al. Measurement of $D^{*\pm}$, $D^\pm$ and $D_s^\pm$ meson production cross sections in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Nucl. Phys. B*, 907:717–763, 2016.
- [155] Stefano Catani, Simone Devoto, Massimiliano Grazzini, Stefan Kallweit, and Javier Mazzitelli. Bottom-quark production at hadron colliders: fully differential predictions in NNLO QCD. *JHEP*, 03:029, 2021.
- [156] Eric Braaten, King-man Cheung, Sean Fleming, and Tzu Chiang Yuan. Perturbative QCD fragmentation functions as a model for heavy quark fragmentation. *Phys. Rev. D*, 51:4819–4829, 1995.
- [157] Shreyasi Acharya et al. Elliptic Flow of Electrons from Beauty-Hadron Decays in Pb–Pb Collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Rev. Lett.*, 126(16):162001, 2021.
- [158] Georges Aad et al. Measurement of azimuthal anisotropy of muons from charm and bottom hadrons in Pb+Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV with the ATLAS detector. *Phys. Lett. B*, 807:135595, 2020.
- [159] K. Aamodt et al. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.
- [160] P. J. Clark. The ATLAS detector simulation. *Nucl. Phys. B Proc. Suppl.*, 215:85–88, 2011.
- [161] Betty Bezverkhny Abelev et al. Performance of the ALICE Experiment at the CERN LHC. *Int. J. Mod. Phys. A*, 29:1430044, 2014.
- [162] K Aamodt et al. Alignment of the ALICE Inner Tracking System with cosmic-ray tracks. *JINST*, 5:P03003, 2010.
- [163] J. Alme et al. The ALICE TPC, a large 3-dimensional tracking device with fast readout for ultra-high multiplicity events. *Nucl. Instrum. Meth.*, A622:316–367, 2010.
- [164] Francesca Carnesecchi. Performance of the ALICE Time-Of-Flight detector at the LHC. *JINST*, 14(06):C06023, 2019.
- [165] R. Fruhwirth. Application of Kalman filtering to track and vertex fitting. *Nucl. Instrum. Meth. A*, 262:444–450, 1987.



- [166] Christian Lippmann. Performance of the ALICE Time Projection Chamber. *Phys. Procedia*, 37:434–441, 2012.
- [167] Betty Abelev et al. Centrality determination of Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV with ALICE. *Phys. Rev. C*, 88(4):044909, 2013.
- [168] Dmitri Kharzeev, Eugene Levin, and Marzia Nardi. Color glass condensate at the LHC: Hadron multiplicities in pp, pA and AA collisions. *Nucl. Phys. A*, 747:609–629, 2005.
- [169] Wei-Tian Deng, Xin-Nian Wang, and Rong Xu. Hadron production in p+p, p+Pb, and Pb+Pb collisions with the HIJING 2.0 model at energies available at the CERN Large Hadron Collider. *Phys. Rev. C*, 83:014915, 2011.
- [170] Michael L. Miller, Klaus Reygers, Stephen J. Sanders, and Peter Steinberg. Glauber modeling in high energy nuclear collisions. *Ann. Rev. Nucl. Part. Sci.*, 57:205–243, 2007.
- [171] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth. A*, 389:81–86, 1997.
- [172] T. Sjostrand, S. Mrenna, and P. Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 0605:026, 2006.
- [173] T. Sjöstrand et al. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.
- [174] Xin-Nian Wang and Miklos Gyulassy. HIJING: A Monte Carlo model for multiple jet production in pp, pA and AA collisions. *Phys. Rev.*, D44:3501–3516, 1991.
- [175] R. Brun et al. *GEANT: Detector Description and Simulation Tool; Oct 1994*. CERN Program Library. CERN, Geneva, 1993. Long Writeup W5013.
- [176] Alexander Radovic, Mike Williams, David Rousseau, Michael Kagan, Daniele Bonacorsi, Alexander Himmel, Adam Aurisano, Kazuhiro Terao, and Taritree Wongjirad. Machine learning at the energy and intensity frontiers of particle physics. *Nature*, 560(7716):41–48, 2018.
- [177] Kim Albertsson et al. Machine Learning in High Energy Physics Community White Paper. *J. Phys. Conf. Ser.*, 1085(2):022008, 2018.
- [178] R. L. Workman et al. Review of Particle Physics. *PTEP*, 2022:083C01, 2022.
- [179] Yoav Freund and Robert E. Schapire. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.*, 55(1):119–139, 1997.



- [180] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *The Annals of Statistics*, 28(2):337 – 407, 2000.
- [181] xgboost. <https://xgboost.readthedocs.io/en/latest/>.
- [182] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 785–794, 2016.
- [183] Fernando Nogueira. Bayesian Optimization: Open source constrained global optimization tool for Python, 2014–.
- [184] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.
- [185] Luca Barioglio, Fabio Catalano, Matteo Concas, Pietro Fecchio, Fabrizio Grosa, Francesco Mazzaschi, and Maximiliano Puccio. hipe4ml/hipe4ml, July 2021.
- [186] Shreyasi Acharya et al. Measurement of beauty production via non-prompt D^0 mesons in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *JHEP*, 12:126, 2022.
- [187] Rafal Maciula and Antoni Szczurek. Open charm production at the LHC - k_t -factorization approach. *Phys. Rev.*, D87(9):094022, 2013.
- [188] Marta Luszczak, Rafal Maciula, and Antoni Szczurek. Nonphotonic electrons at RHIC within $k(t)$ -factorization approach and with experimental semileptonic decay functions. *Phys. Rev.*, D79:034009, 2009.
- [189] M. Aguilar-Benitez et al. Comparative Properties of 400-GeV/c Proton - Proton Interactions With and Without Charm Production. *Z. Phys.*, C41:191, 1988.
- [190] Paolo Bartalini and Livio Fano, editors. *Proceedings, 1st International Workshop on Multiple Partonic Interactions at the LHC (MPI08)*, Hamburg, 2009. DESY, DESY.
- [191] Torbjorn Sjostrand and Maria van Zijl. A Multiple Interaction Model for the Event Structure in Hadron Collisions. *Phys. Rev. D*, 36:2019, 1987.
- [192] S. Porteboeuf and R. Granier de Cassagnac. J/Ψ yield vs. multiplicity in proton-proton collisions at the LHC. *Nucl. Phys. Proc. Suppl.*, 214:181–184, 2011.



- [193] B. Abelev et al. J/ψ Production as a Function of Charged Particle Multiplicity in pp Collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett.*, B712:165–175, 2012.
- [194] Jaroslav Adam et al. Measurement of charm and beauty production at central rapidity versus charged-particle multiplicity in proton-proton collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 09:148, 2015.
- [195] Serguei Chatrchyan et al. Event activity dependence of $Y(nS)$ production in $\sqrt{s_{NN}}=5.02$ TeV pPb and $\sqrt{s}=2.76$ TeV pp collisions. *JHEP*, 04:103, 2014.
- [196] K. Werner, B. Guiot, Iu. Karpenko, T. Pierog, and G. Sophys. Charm production in high multiplicity pp events. In *Proceedings, 7th International Workshop on Multiple Partonic Interactions at the LHC (MPI@LHC 2015): Miramare, Trieste, Italy, November 23-27, 2015*, pages 66–70, 2016.
- [197] Anisa Khatun. J/ψ production as a function of charged-particle multiplicity in pp collisions at $\sqrt{s} = 5.02$ TeV with ALICE. In *23rd DAE-BRNS High Energy Physics Symposium 2018 Chennai, India, December 10-14, 2018*, 2019.
- [198] Jaroslav Adam et al. Multiplicity dependence of charged pion, kaon, and (anti)proton production at large transverse momentum in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Lett.*, B760:720–735, 2016.
- [199] Albert M Sirunyan et al. Measurement of charged pion, kaon, and proton production in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev.*, D96(11):112003, 2017.
- [200] Shreyasi Acharya et al. Multiplicity dependence of light-flavor hadron production in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Rev.*, C99(2):024906, 2019.
- [201] Shreyasi Acharya et al. Multiplicity dependence of (multi-)strange hadron production in proton-proton collisions at $\sqrt{s} = 13$ TeV. *submitted to EPJC*, 2019.
- [202] Shreyasi Acharya et al. Charged-particle production as a function of multiplicity and transverse sphericity in pp collisions at $\sqrt{s} = 5.02$ and 13 TeV. *Eur. Phys. J.*, C79(10):857, 2019.
- [203] Betty Abelev et al. Multiplicity dependence of two-particle azimuthal correlations in pp collisions at the LHC. *JHEP*, 09:049, 2013.
- [204] Shreyasi Acharya et al. Pseudorapidity distributions of charged particles as a function of mid- and forward rapidity multiplicities in pp collisions at $\sqrt{s} = 5.02, 7$ and 13 TeV. *Eur. Phys. J. C*, 81(7):630, 2021.



-
- [205] Kaggle Inc. Kaggle home page, 2018.
- [206] ALICE DPG. [DPG - Analysis Object Tools - Systematic uncertainties on TPC-ITS matching in various Run-2 data samples](#). *ALICE Twiki*, 2019.
- [207] Roel Aaij et al. Measurement of b hadron fractions in 13 TeV pp collisions. *Phys. Rev. D*, 100(3):031102, 2019.
- [208] Renu Bala et al. [Measurement of charm production multiplicity dependence at central rapidity in proton-proton collisions at \$\sqrt{s} = 7\$ TeV](#). *ALICE Analysis Note*, 2013.
- [209] Shreyasi Acharya et al. Multiplicity dependence of π , K, and p production in pp collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 80(8):693, 2020.
- [210] Shreyasi Acharya et al. Charged-particle production as a function of multiplicity and transverse sphericity in pp collisions at $\sqrt{s} = 5.02$ and 13 TeV. *Eur. Phys. J. C*, 79(10):857, 2019.
- [211] Shreyasi Acharya et al. Measurement of D^0 , D^+ , D^{*+} and D_s^+ production in pp collisions at $\sqrt{s} = 5.02$ TeV with ALICE. *Eur. Phys. J.*, C79(5):388, 2019.
- [212] Leonid Gladilin. Fragmentation fractions of c and b quarks into charmed hadrons at LEP. *Eur. Phys. J. C*, 75(1):19, 2015.
- [213] Shreyasi Acharya et al. Λ_c^+ production in pp and in p – Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Rev. C*, 104:054905, 2021.
- [214] Yu Chen and Min He. Charged-particle multiplicity dependence of charm-baryon-to-meson ratio in high-energy proton–proton collisions. *Phys. Lett. B*, 815:136144, 2021.
- [215] D. Ebert, R.N. Faustov, and V.O. Galkin. Spectroscopy and Regge trajectories of heavy baryons in the relativistic quark-diquark picture. *Phys. Rev. D*, 84:014025, 2011.
- [216] P.A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020.
- [217] Shreyasi Acharya et al. Λ_c^+ production and Baryon-to-Meson Ratios in pp and p – Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV at the LHC. *Phys. Rev. Lett.*, 127:202301, 2021.
- [218] D. Prato and C. Tsallis. Nonextensive foundation of Levy distributions. *Phys. Rev. E*, 60:2398, 1999.
- [219] Shreyasi Acharya et al. Supplemental material: Observation of a multiplicity dependence in the p_T -differential charm baryon-to-meson ratios in proton–proton collisions at $\sqrt{s} = 13$ TeV.



- [220] Shreyasi Acharya et al. Observation of a multiplicity dependence in the p_T -differential charm baryon-to-meson ratios in proton-proton collisions at $\sqrt{s} = 13$ TeV. 11 2021.
- [221] Evidence for modification of b quark hadronization in high-multiplicity pp collisions at $\sqrt{s} = 13$ TeV. 4 2022.
- [222] Roel Aaij et al. Observation of Multiplicity Dependent Prompt $\chi_{c1}(3872)$ and $\psi(2S)$ Production in pp Collisions. *Phys. Rev. Lett.*, 126(9):092001, 2021.
- [223] Angelo Esposito, Elena G. Ferreira, Alessandro Pilloni, Antonio D. Polosa, and Carlos A. Salgado. The nature of $X(3872)$ from high-multiplicity pp collisions. *Eur. Phys. J. C*, 81(7):669, 2021.
- [224] Eric Braaten, Li-Ping He, Kevin Ingles, and Jun Jiang. Production of $X(3872)$ at High Multiplicity. *Phys. Rev. D*, 103(7):L071901, 2021.
- [225] Shreyasi Acharya et al. ALICE 2016-2017-2018 luminosity determination for pp collisions at $\sqrt{s} = 13$ TeV. (ALICE-PUBLIC-2021-005), 2021.
- [226] Luca Barioglio, Fabio Catalano, Matteo Concas, Pietro Fecchio, Fabrizio Grosa, Francesco Mazzaschi, and Maximiliano Puccio. hipe4ml/hipe4ml, July 2021.
- [227] K. Werner, B. Guiot, Iu. Karpenko, and T. Pierog. Analysing radial flow features in p-Pb and p-p collisions at several TeV by studying identified particle production in EPOS3. *Phys. Rev. C*, 89(6):064903, 2014.
- [228] Christian Bierlich et al. A comprehensive guide to the physics and usage of PYTHIA 8.3. 3 2022.
- [229] Klaus Werner. On a deep connection between factorization and saturation: new insight into modeling high-energy proton-proton and nucleus-nucleus scattering in the EPOS4 framework. 1 2023.
- [230] Ivan Schmidt and Marat Siddikov. Production mechanisms of open-heavy flavor mesons. *Phys. Rev. D*, 101(9):094020, 2020.
- [231] T. Kneesch, B. A. Kniehl, G. Kramer, and I. Schienbein. Charmed-meson fragmentation functions with finite-mass corrections. *Nucl. Phys. B*, 799:34–59, 2008.
- [232] Shreyasi Acharya et al. Measurement of the Cross Sections of Ξ_c^0 and Ξ_c^+ Baryons and of the Branching-Fraction Ratio $\text{BR}(\Xi_c^0 \rightarrow \Xi^- e^+ \nu_e)/\text{BR}(\Xi_c^0 \rightarrow \Xi^- \pi^+)$ in pp collisions at 13 TeV. *Phys. Rev. Lett.*, 127(27):272001, 2021.



- [233] Shreyasi Acharya et al. Λ_c^+ Production and Baryon-to-Meson Ratios in pp and p-Pb Collisions at $\sqrt{s_{NN}}=5.02$ TeV at the LHC. *Phys. Rev. Lett.*, 127(20):202301, 2021.
- [234] Shreyasi Acharya et al. Λ_c^+ production in pp and in p-Pb collisions at $\sqrt{s_{NN}}=5.02$ TeV. *Phys. Rev. C*, 104(5):054905, 2021.
- [235] Shreyasi Acharya et al. Measurement of Prompt D^0 , Λ_c^+ , and $\Sigma_c^{0,++}(2455)$ Production in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 128(1):012001, 2022.
- [236] First measurement of Ω_c^0 production in pp collisions at $\sqrt{s} = 13$ TeV. 5 2022.
- [237] Shreyasi Acharya et al. Measurement of the production cross section of prompt Ξ_c^0 baryons at midrapidity in pp collisions at $\sqrt{s} = 5.02$ TeV. *JHEP*, 10:159, 2021.
- [238] Jun Song, Hai-hong Li, and Feng-lan Shao. New feature of low p_T charm quark hadronization in pp collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J.*, C78(4):344, 2018.
- [239] R. Aaij et al. Study of the production of Λ_b^0 and $\bar{B}^0$ hadrons in pp collisions and first measurement of the $\Lambda_b^0 \rightarrow J/\psi p K^-$ branching fraction. *Chin. Phys. C*, 40(1):011001, 2016.
- [240] Roel Aaij et al. Measurement of b hadron fractions in 13 TeV pp collisions. *Phys. Rev. D*, 100(3):031102, 2019.
- [241] R Aaij et al. Prompt charm production in pp collisions at sqrt(s)=7 TeV. *Nucl. Phys. B*, 871:1–20, 2013.
- [242] Georges Aad et al. Measurement of the differential cross-section of B^+ meson production in pp collisions at $\sqrt{s} = 7$ TeV at ATLAS. *JHEP*, 10:042, 2013.
- [243] Vardan Khachatryan et al. Measurement of the B^+ Production Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV. *Phys. Rev. Lett.*, 106:112001, 2011.
- [244] Serguei Chatrchyan et al. Measurement of the Strange B Meson Production Cross Section with $J/\Psi \phi$ Decays in pp Collisions at $\sqrt{s} = 7$ TeV. *Phys. Rev. D*, 84:052008, 2011.
- [245] Serguei Chatrchyan et al. Measurement of the B^0 production cross section in pp Collisions at $\sqrt{s} = 7$ TeV. *Phys. Rev. Lett.*, 106:252001, 2011.
- [246] Serguei Chatrchyan et al. Measurement of the Λ_b cross section and the $\bar{\Lambda}_b$ to Λ_b ratio with $J/\Psi \Lambda$ decays in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett. B*, 714:136–157, 2012.
- [247] R. L. Workman and Others. Review of Particle Physics. *PTEP*, 2022:083C01, 2022.



- [248] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, Jon Pumplin, and C. P. Yuan. New parton distributions for collider physics. *Phys. Rev. D*, 82:074024, 2010.
- [249] Shreyasi Acharya et al. Dielectron production in proton-proton collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 09:064, 2018.
- [250] Shreyasi Acharya et al. Dielectron and heavy-quark production in inelastic and high-multiplicity proton-proton collisions at $\sqrt{s_{NN}} = 13$ TeV. *Phys. Lett. B*, 788:505–518, 2019.
- [251] Betty Abelev et al. Measurement of electrons from beauty hadron decays in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett. B*, 721:13–23, 2013. [Erratum: *Phys. Lett. B* **763** (2016) 507–509].
- [252] Betty Abelev et al. Measurement of prompt J/ψ and beauty hadron production cross sections at mid-rapidity in pp collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 11:065, 2012.
- [253] Shreyasi Acharya et al. Production of light-flavor hadrons in pp collisions at $\sqrt{s} = 7$ and $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 81(3):256, 2021.
- [254] Min He and Ralf Rapp. Bottom hadro-chemistry in high-energy hadronic collisions. 9 2022.
- [255] A. Adare et al. Energy Loss and Flow of Heavy Quarks in Au+Au Collisions at $\sqrt{s_{NN}} = 200$ GeV. *Phys. Rev. Lett.*, 98:172301, 2007.
- [256] L. Adamczyk et al. Elliptic flow of electrons from heavy-flavor hadron decays in Au+Au collisions at $\sqrt{s_{NN}} = 200, 62.4, \text{ and } 39$ GeV. *Phys. Rev.*, C95(3):034907, 2017.
- [257] L. Adamczyk et al. Measurement of D^0 Azimuthal Anisotropy at Midrapidity in Au+Au Collisions at $\sqrt{s_{NN}} = 200$ GeV. *Phys. Rev. Lett.*, 118(21):212301, 2017.
- [258] B. Abelev et al. D meson elliptic flow in non-central Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Rev. Lett.*, 111:102301, 2013.
- [259] Betty Abelev et al. Azimuthal anisotropy of D meson production in Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Rev.*, C90(3):034904, 2014.
- [260] Jaroslav Adam et al. Elliptic flow of electrons from heavy-flavour hadron decays at mid-rapidity in Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *JHEP*, 09:028, 2016.
- [261] Jaroslav Adam et al. Elliptic flow of muons from heavy-flavour hadron decays at forward rapidity in Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Lett.*, B753:41–56, 2016.



- [262] Albert M Sirunyan et al. Measurement of prompt D^0 meson azimuthal anisotropy in Pb–Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV. *Phys. Rev. Lett.*, 120(20):202301, 2018.
- [263] Shreyasi Acharya et al. Event-shape engineering for the D-meson elliptic flow in mid-central Pb–Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV. *JHEP*, 02:150, 2019.
- [264] Albert M Sirunyan et al. Measurement of prompt D^0 meson azimuthal anisotropy in Pb–Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV. *Phys. Rev. Lett.*, 120(20):202301, 2018.
- [265] Albert M Sirunyan et al. Measurement of prompt D^0 and $\bar{D}^0$ meson azimuthal anisotropy and search for strong electric fields in PbPb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV. *Phys. Lett. B*, 816:136253, 2021.
- [266] Jan Uphoff, Oliver Fochler, Zhe Xu, and Carsten Greiner. Open Heavy Flavor in Pb+Pb Collisions at $\sqrt{s} = 2.76$ TeV within a Transport Model. *Phys. Lett.*, B717:430–435, 2012.
- [267] Min He, Rainer J. Fries, and Ralf Rapp. Heavy Flavor at the Large Hadron Collider in a Strong Coupling Approach. *Phys. Lett.*, B735:445–450, 2014.
- [268] M. Monteno, W. M. Alberico, A. Beraudo, A. De Pace, A. Molinari, M. Nardi, and F. Prino. Heavy-flavor dynamics in nucleus-nucleus collisions: from RHIC to LHC. *J. Phys.*, G38:124144, 2011.
- [269] Taesoo Song, Hamza Berrehrah, Daniel Cabrera, Wolfgang Cassing, and Elena Bratkovskaya. Charm production in Pb + Pb collisions at energies available at the CERN Large Hadron Collider. *Phys. Rev.*, C93(3):034906, 2016.
- [270] Jan Uphoff, Oliver Fochler, Zhe Xu, and Carsten Greiner. Elastic and radiative heavy quark interactions in ultra-relativistic heavy-ion collisions. *J. Phys.*, G42(11):115106, 2015.
- [271] A. Beraudo, A. De Pace, M. Monteno, M. Nardi, and F. Prino. Heavy flavors in heavy-ion collisions: quenching, flow and correlations. *Eur. Phys. J.*, C75(3):121, 2015.
- [272] Shanshan Cao, Tan Luo, Guang-You Qin, and Xin-Nian Wang. Heavy and light flavor jet quenching at RHIC and LHC energies. *Phys. Lett.*, B777:255–259, 2018.
- [273] Ralf Rapp. Heavy-Flavor Transport and Hadronization in the QGP. *PoS*, BEAUTY2020:033, 2021.
- [274] A. Francis, O. Kaczmarek, M. Laine, T. Neuhaus, and H. Ohno. Nonperturbative estimate of the heavy quark momentum diffusion coefficient. *Phys. Rev. D*, 92(11):116003, 2015.



- [275] Shuai Y. F. Liu and Ralf Rapp. Spectral and transport properties of quark–gluon plasma in a nonperturbative approach. *Eur. Phys. J. A*, 56(2):44, 2020.
- [276] F. Riek and R. Rapp. Quarkonia and Heavy-Quark Relaxation Times in the Quark-Gluon Plasma. *Phys. Rev. C*, 82:035201, 2010.
- [277] Vardan Khachatryan et al. Suppression and azimuthal anisotropy of prompt and nonprompt J/ψ production in PbPb collisions at $\sqrt{s_{\text{NN}}} = 2.76$ TeV. *Eur. Phys. J. C*, 77(4):252, 2017.
- [278] Morad Aaboud et al. Prompt and non-prompt J/ψ elliptic flow in Pb+Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 78(9):784, 2018.
- [279] R. Arnaldi et al. The Zero degree calorimeters for the ALICE experiment. *Nucl. Instrum. Meth. A*, 581:397–401, 2007. [Erratum: Nucl.Instrum.Meth.A 604, 765 (2009)].
- [280] Centrality determination in heavy ion collisions. Technical Report ALICE-PUBLIC-2018-011, Aug 2018.
- [281] Sergei A. Voloshin, Arthur M. Poskanzer, and Raimond Snellings. Collective phenomena in non-central nuclear collisions. *Landolt-Bornstein*, 23:293–333, 2010.
- [282] Matthew Luzum and Jean-Yves Ollitrault. Eliminating experimental bias in anisotropic-flow measurements of high-energy nuclear collisions. *Phys. Rev.*, C87(4):044907, 2013.
- [283] C. Adler et al. Elliptic flow from two and four particle correlations in Au+Au collisions at $\sqrt{s_{\text{NN}}} = 130$ GeV. *Phys. Rev. C*, 66:034904, 2002.
- [284] Ilya Selyuzhenkov and Sergei Voloshin. Effects of non-uniform acceptance in anisotropic flow measurement. *Phys.Rev.C*, 77:034904, 2008.
- [285] N. Borghini and J. Y. Ollitrault. Azimuthally sensitive correlations in nucleus-nucleus collisions. *Phys. Rev.*, C70:064905, 2004.
- [286] Weiyao Ke, Yingru Xu, and Steffen A. Bass. Linearized Boltzmann-Langevin model for heavy quark transport in hot and dense QCD matter. *Phys. Rev.*, C98(6):064901, 2018.
- [287] Jun Shao and C. F. J. Wu. A General Theory for Jackknife Variance Estimation. *The Annals of Statistics*, 17(3):1176–1197, 2012.
- [288] Min He and Ralf Rapp. Hadronization and Charm-Hadron Ratios in Heavy-Ion Collisions. *Phys. Rev. Lett.*, 124(4):042301, 2020.



- [289] Shanshan Cao, Tan Luo, Guang-You Qin, and Xin-Nian Wang. Linearized Boltzmann transport model for jet propagation in the quark-gluon plasma: Heavy quark evolution. *Phys. Rev.*, C94(1):014909, 2016.
- [290] Shuang Li, Wei Xiong, and Renzhuo Wan. Relativistic Langevin dynamics: charm versus beauty. *Eur. Phys. J. C*, 80(12):1113, 2020.
- [291] Shu-Qing Li, Wen-Jing Xing, Feng-Lei Liu, Shanshan Cao, and Guang-You Qin. Heavy flavor quenching and flow: the roles of initial condition, pre-equilibrium evolution, and in-medium interaction. *Chin. Phys. C*, 44(11):114101, 2020.
- [292] Shu-Qing Li, Wen-Jing Xing, Xiang-Yu Wu, Shanshan Cao, and Guang-You Qin. Scaling behaviors of heavy flavor meson suppression and flow in different nuclear collision systems at the LHC. *Eur. Phys. J. C*, 81(11):1035, 2021.
- [293] Weiyao Ke, Yingru Xu, and Steffen A. Bass. Modified Boltzmann approach for modeling the splitting vertices induced by the hot QCD medium in the deep Landau-Pomeranchuk-Migdal region. *Phys. Rev. C*, 100(6):064911, 2019.
- [294] Z. Citron et al. Report from Working Group 5: Future physics opportunities for high-density QCD at the LHC with heavy-ion and proton beams. *CERN Yellow Rep. Monogr.*, 7:1159–1410, 2019.



Appendix 博士论文中文简介

A.1 论文的研究意义

随着科学技术和人类认知水平的提高,探索人类起源和世界构成的领域衍生出了许多不同的学科。宇宙大爆炸后约 10^{-5}s 会形成一种新的物质形态:夸克胶子等离子体 (QGP)。该物质形态也被量子色动力学 (QCD) 所描述,该理论描述了构成强子物质的基本组成部分夸克和胶子之间的强相互作用以及预言了强子相到 QGP 相可能存在的相变临界点。研究这种强相互作用物质也有助于我们理解宇宙的起源和形成。然而,有一个问题一直困扰着人类:如何在实验室中制造这种高温高密的强相互作用物质?在 20 世纪 70 年代,李政道先生提出了利用接近光速的重离子碰撞来研究夸克胶子等离子体的想法。基于这个伟大的想法,重离子碰撞实验开始兴起!其中最著名的是美国布鲁克海文国家实验室 (BNL) 的相对论重离子对撞机 (RHIC) 上的螺旋管径迹探测器 (STAR) 和欧洲核子研究中心 (CERN) 的大型强子对撞机 (LHC) 上的大型重离子实验 (ALICE)。本论文的研究载体是 LHC 上的 ALICE 实验,旨在通过分析 ALICE Run2 实验数据来探究夸克胶子等离子体的性质。

由于重离子碰撞的演化过程复杂且短暂,主要分为:包括初始状态、预平衡阶段、QGP 阶段和强子化阶段。实验上只能通过重建并分析碰撞末态的强子信息来探索碰撞的各个阶段,不同的实验观测量又可称为探针。不同探针针对不同重离子碰撞阶段阶段有不同的敏感性。这些探针也可分为软探针和硬探针两类,其中软探针包括粒子各向异性流、粒子产额比等,而硬探针则包括喷注 (jet)、重味强子谱和高横动量粒子谱等。其中,重味强子指带有重味夸克粲 (charm) 和底 (beauty) 的强子,重味夸克被视为研究 QGP 的完美探针,因为它们主要产生于碰撞早期的硬散射过程,并经历了整个碰撞系统的演化过程。实验上主要通过研究重味强子的产额比、各向异性流以及核修正因子来研究强子化机制、QGP 的输运性质和部分子在 QGP 介质中的能损效应。

目前观测到的实验现象不足以证明小系统 (如质子-质子和质子-重离子对撞) 中也产生了 QGP。通常,小系统的实验测量可用于检验微扰量子色动力学 (pQCD) 并作为核核对撞实验的参考。重味强子的产额通常可以用基于微扰量子色动力学的因子化理论来描述。在因子化理论中,重味强子的产额可以由对撞入射质子的部分子分布函数、部分子硬散射的重味夸克的产生截面以及描述重味夸克碎裂函数三部分卷积而成。其中,重味夸克碎裂成重味强子的碎裂函数和入射质子的部分子分布函数是非微扰过程,只能通过实验测量来约束,而硬散射截面可以通过 pQCD 计算得出。值得一提的是,长期以来夸克碎裂函数一直被认为是普适的,即此过程不依赖于碰撞系统和能量。然而,尽管基于正负电子 (e^+e^-) 和电子-质子 (ep) 对撞实验提取的



碎裂函数约束的次领头阶微扰计算可以成功描述 LHC 能区的粲介子产额。但最近 ALICE 实验上质心系能量为 5.02 和 7 TeV 的质子-质子 (pp) 对撞中测量的粲重子-介子比 Λ_c^+/D^0 在低横动量区间却显著高于基于 e^+e^- 碎裂函数参数化的事件产生器 PYTHIA 8 的模拟结果。类似的现象也在 ALICE 的中心快度区和 LHCb 的向前快度区的质子-铅核 (p-Pb) 对撞数据中被观察到。最后, 结合最近 ALICE pp 对撞中测量的其它粲重子介子比 (如 Ξ_c^{+0}/D^0) 的产额进一步表明碎裂函数并不是普适的。值得一提的是, 以上 ALICE 实验结果使用的是无偏事件 (Minimum Bias events), 其带电粒子多重数为整体事件的均值。测量不同带电粒子多重数区间的 Λ_c^+/D^0 产额比可以系统地研究碎裂函数的非普适现象和重味重子的形成机制, 将可以进一步约束理论模型。

目前实验上对于中心快度且低横动量区底夸克的测量一直十分缺乏。ALICE 实验组通过测量底强子衰变产物 (如底夸克衰变产生的电子和缪子, 以及非瞬时的 D 介子和夸克偶素) 来间接研究底夸克的性质。重味夸克强子化模型可以通过测量非瞬时 D^0 介子产额随带电粒子多重数的依赖关系来进一步完善。而在重离子实验中, 由于底夸克质量较大, 其在 QGP 中是否会被热化一直是实验上的一个疑问。最近, ALICE 实验组观测到底强子衰变电子的椭圆流显著大于零, 这表明底夸克也有很大可能参与了热化过程。然而, 底强子衰变电子的横动量仅反映了较高横动量区间的底强子集体行为, 无法充分反映底夸克的热化行为。基于目前的探测器空间分辨率和亮度, 测量强子道的衰变产物即非瞬时粲介子 (D^0) 的椭圆流可以更好地研究底夸克与 QGP 之间的相互作用。

本论文旨在通过测量 LHC 能区 ALICE 实验中粲强子和非瞬时 D^0 介子的产额及椭圆流, 研究重味夸克强子化机制和重味夸克与 QGP 的相互作用。针对重味夸克强子化机制, 利用 2016 至 2018 年采集的质心系能量为 13 TeV 的 pp 碰撞数据 (约 1.8 亿次事件) 首次在实验上测量了瞬时粲重子 Λ_c^+ 产额随带电多重数的依赖性关系, 以及非瞬时 D^0 介子所占比例随带电多重数的依赖性和非瞬时 D^0 介子的产额。对于底夸克与 QGP 的相互作用的研究, 利用 2018 年采集的质心系能量为 5.02 TeV 的 Pb-Pb 碰撞数据首次在 ALICE 实验上测量了非对心碰撞中的非瞬时 D^0 介子的椭圆流。

A.2 论文的框架

A.2.1 理论背景介绍

论文的第一章介绍了 QCD、QGP、重离子时空演化过程及主要的观测量和本文的主要研究对象开粲重味物理。

标准模型是描述基本粒子和它们之间相互作用的理论模型。它由三个相互作用基本粒子组成: 强相互作用粒子 (夸克和胶子)、电弱相互作用粒子 (带电粒子和中微子) 以及引力相互作用



用粒子（引力子）。这些粒子被认为是构成宏观物质的基本组成部分。量子色动力学（QCD）是描述强相互作用的理论，它研究的是夸克和胶子的相互作用。夸克是构成质子、中子等核子的基本粒子，而胶子则是负责夸克之间的相互作用的粒子。在 QCD 中，夸克和胶子被描述为带有色荷的粒子，它们之间的相互作用由强相互作用力所描述。自然界中夸克和胶子被禁闭在强子里面，QCD 理论预言在极高温高密的环境下夸克和胶子可能解禁闭形成一种新的物质形态，即夸克胶子等离子体（QGP）。研究 QGP 不仅有助于了解强相互作用力的本质，也可以为探索宇宙起源提供重要线索。人类通过将接近光速的重离子核对撞来产生并研究 QGP 的性质，重离子对撞的时空演化一般分为：包括初始状态、预平衡阶段、QGP 阶段和强子化阶段。然而其演化过程及其短暂，实验上只能通过探测器重建末态粒子来反推其过程，其主要的实验观测量有：带电粒子多重数、不同强子的产额、核修正因子和各向异性流。这些观测量还可分为软探针和硬探针两类，其中软探针包括粒子各向异性流、粒子产额比等，而硬探针则包括喷注 (jet)、重味强子谱和高横动量粒子谱等。

论文主要研究开放重味强子 (open heavy hadron)，即携带一个夸克或底夸克或反夸克的介子和重子。在 RHIC 和 LHC 的重离子碰撞中，95% 以上产生的重夸克分布在这些强子中。重夸克由于其大质量，在碰撞初始阶段的硬散射过程中产生，随后与 QGP 发生的相互作用。因此，重味夸克是研究 QGP 性质的良好探针。对于开发重味物理的研究主要可以通过三方面展开：重味物理产额测量、冷核相互作用和热核相互作用。此外，强子化过程由于其非微扰特性使得很难从第一性原理去理解其机制，目前只能通过诸多唯象模型来进行描述，目前实验上常用的模型有碎裂强子化、夸克重组以及统计强子化模型。重味重子介子产额比对于这些过程及其敏感，实验上的测量可以用来约束理论模型，加深对强子化机制的理解。

A.2.2 实验装置

第二章节介绍了 ALICE 实验的探测器、探测器性能、ALICE 离线分析框架以及本论文用到的实验数据。

ALICE (A Large Ion Collider Experiment) 是欧洲核子研究组织 (CERN) 位于瑞士日内瓦的一项重要实验项目。该实验利用 CERN 的 LHC 加速器加速重离子，探测宇宙大爆炸后形成的夸克胶子等离子体的性质和行为。ALICE 探测器中央桶探测器位于 L3 磁铁内，该大型螺线管提供沿着束流轴线方向的 0.5 T 磁场，这对于跟踪和识别带电粒子是必要的。中央筒包括内部跟踪系统 (ITS)、时间投影室 (TPC)、过渡辐射探测器 (TRD)、飞行时间探测器 (TOF)、环形成像切伦科夫探测器 (HMPID)、光子探测器 PHOS、EMCal 和 DCal。除此之外，ALICE 还包括用于全局事件特征描述和触发的各种其他探测器 (ZDC、PMD、FMD、T0、V0A 和 V0C)，这些探测器位于前向和后向区域，以及一个用于触发宇宙射线的闪烁体阵列 (ACORDE)，位于



L3 磁铁的顶部。

ALICE 探测器具有良好的径迹和顶点重建能力，主要是依靠中心快度区的 ITS 和 TPC。与此同时，ALICE 结合 TPC 和 TOF 探测器在很宽的动量范围内对于带电强子具有良好的鉴别能力，其原理是分别是根据带电粒子在气体中的电离能损和飞行时间的不同来鉴别粒子的种类。

实验得到的数据需要进行高度复杂的计算和分析，因此需要使用先进的计算机技术和高性能计算集群来处理和分析实验数据。实验离线分析软件是基于 CERN ROOT 开发的，称为 AliRoot 和 AliPhysics。对于质子质子对撞事件的模拟采用的是 PYTHIA 事件产生器，质子铅核和铅核铅核碰撞模拟则采用 HIJING 产生器。这些产生器产生的粒子随后都会通过 GEANT 模拟的探测器进行输运得到最后的模拟数据，用以物理分析的结果修正和验证。本论文用到的数据为 2016 至 2018 年采集的质心系能量为 13 TeV 的 pp 碰撞数据（约 18 亿次事件）以及 2018 年采集的质心系能量为 5.02 TeV 的 Pb – Pb 碰撞数据（约 8.5 亿次事件）。

A.2.3 实验分析方法

第三章主要讨论了论文实验数据分析所用到的实验方法，其中包括基于机器学习算法的备选粒子重建、重味强子衰变比例的实验数据修正方法等。

由于重味强子平均衰变长度最多只有几百微米，因此不能被探测到即在到达第一个探测器层之前就已经衰变了。因此由于其本身较低的产额和大量的本底，重味强子的重建一直具有挑战性。为了抑制大的组合本底，需要基于这些衰变道的特性严格的选择标准。除了利用 ALICE 的粒子鉴别能力外，探测器的空间分辨率还可以重建次级衰变顶点，提供关键的信息。大多数物理分析旨在提取具有高精度的物理“信号”。这通常是通过减少“背景”来实现的，这些减少背景的选择标准可以基于信号的某些特征。这些特征可以包括信号的拓扑和动力学特性或探测器响应，并利用我们对信号背后的物理学知识。虽然更严格的选择可以减少背景，但也可能导致信号减少。因此，优化信号和背景水平非常重要。显然这是一个多元变量分析任务。近年来，机器学习算法在高能物理中越来越受欢迎，用于各种任务，包括分类和回归问题，以及模拟和事件重构。该算法利用已知信号和背景组成的样本（通常是模拟的）来响应未知组成的样本，从而优化这些样本中的信号和背景分离。其中决策树 (BDT) 由于其简单的机制和高效的模型解释性被广泛用于稀有探针的寻找，我们同样也利用了基于 BDT 的梯度决策树算法 XGBoost 来提取重味强子信号，并且还进一步区分瞬时和非瞬时粲强子。

实验上由于对于底强子产额测量的匮乏，使得目前对于粲强子在结果修正时很难估计其中底强子所占比例。ALICE 之前的测量都是基于 FONLL 模型给出的横动量谱或者结合其他实验组测量的横动量谱来估计底强子所占比例。然而这些方法显然有很大的局限性。本论文采用了



一种先进的纯数据驱动方法来估计重味强子测量中粲强子和底强子所占的比例分数。该方法现在也广泛用于 ALICE 重味强子测量中。

A.2.4 重味强子的产生

第四章介绍了围绕 ALICE 质心系能量 13 TeV 质子质子对撞数据展开的实验分析。首先是对粲重子 Λ_c^+ 相对产额随带电多重数的测量, Λ_c^+ 备选粒子主要是通过强子道 $\Lambda_c^+ \rightarrow pK_S^0$ 在不同的带电粒子多重数区间进行重建。该结果首次在不同碰撞系统尺寸下对粲重子 Λ_c^+ 的产额进行测量, 结果表明粲重子介子比 Λ_c^+/D^0 的横动量分布随着带电粒子多重数的增加而抬高, 其中最低多重数和最高多重数之间有 5.3σ 的区分度。然而对结果进行横动量积分后发现积分的 Λ_c^+/D^0 并没有明显的多重数依赖, 这意味着之前的行为似乎表明粲重子和粲介子随着多重数的变化在横动量上进行了重新分配。在此实验分析的基础上, 对非瞬时 D^0 粲介子所占比例分数和带电多重数的依赖关系进行了测量。结果表明, 非瞬时 D 介子的比例分数几乎没有带电粒子多重数依赖, 这表明粲夸克和底夸克的产生机制在不同带电粒子多重数下似乎是一致的。接着对同样能量下非瞬时 D^0 介子产额进行了测量, 并结合非瞬时 Λ_c^+ 的产额测量。结果显示, 非瞬时的 Λ_c^+/D^0 和瞬时的 Λ_c^+/D^0 以及 Λ/K_S^0 随横动量一致且都显著大于 e^+e^- 和 ep 对撞的测量结果, 暗示 LHC 能区重子似乎有共同的产生机制。

A.2.5 非瞬时 D^0 介子椭圆流的测量

第五章介绍了在 ALICE 实验中首次测量质心系能量为 5.02 TeV 的非瞬时 D^0 介子椭圆流。 D^0 介子椭圆流的测量采用了标量积 (scalar-product) 方法, 备选粒子信号重建过程中也采用了 BDT 的方法。基于标量积方法得到的 D^0 介子椭圆流中包含了瞬时和非瞬时 D^0 介子的贡献。为了提取出我们最后的非瞬时 D^0 介子椭圆流, 我们采用了一个线性拟合的方法。主要是将实验数据基于 BDT 输出概率分为 3-4 个部分, 这些部分互相独立且可以用之前介绍到的方法计算其非瞬时 D^0 介子的比例。显然这些部分是线性分布且可以通过拟合来外推的到概率为 100% 的非瞬时 D^0 介子椭圆流。实验结果暗示非瞬时 D^0 介子椭圆流可能大于 0 (显著性为 2.7σ) 且与非奇异 D 介子椭圆流有着 3.2σ 的显著区分度。结果在误差范围内和理论模型一致, 表明底夸克在 QGP 中可能部分热化了。

A.2.6 总结与展望

第六章总结了所有的实验测量工作, 并给出了展望。本论文的旨在通过分析 ALICE Run2 数据来研究强子化机制和 QGP 的性质。本论文采用机器学习方法决策树 (BDT) 来抑制组合背景, 并使用 XGBoost 算法对实验数据进行多分类, 特别是对非瞬时和瞬时 D^0 介子的分离。还采用了一种纯数据驱动的方法来修正粲强子产额。

在小系统质心系能量为 $\sqrt{s} = 13$ TeV pp 碰撞中的开粲强子测量分为三部分。首先是粲重子



Λ_c^+ 随带电粒子多重数的测量, 基于向前区的多重数分类器 V0M 提取出不同带电粒子多重数区间的 Λ_c^+ 产额, 结果表明 Λ_c^+/D^0 的产额比有明显的多重数依赖。其结果可以被 PYTHIA 8 事件产生器的色重联机制以及统计强子化模型描述。然而对结果进行横动量积分后发现积分的 Λ_c^+/D^0 并没有明显的多重数依赖, 这意味着之前的行为似乎表明粲重子和粲介子随着多重数的变化只是在横动量上进行了重新分配。该实验结果对理论模型提供了很好的约束, 对于强子化机制的理解具有重大意义。然后由于实验上一直缺乏对底强子产额随带电粒子多重数依赖的测量, 于是在同样的能量和带电粒子多重数区间类测量了非瞬时 D 介子的比例。通过这一间接测量来研究底夸克随系统尺寸的演化过程, 结果表明非瞬时 D 介子的比例分数几乎没有带电粒子多重数依赖, 这表明粲夸克和底夸克的强子化机制在不同带电粒子多重数下似乎是一致的。最后对同样能量下非瞬时 D^0 介子产额进行了测量, 并结合非瞬时 Λ_c^+ 的产额测量。结果显示, 非瞬时的 Λ_c^+/D^0 和瞬时的 Λ_c^+/D^0 以及 Λ/K_S^0 随横动量一致且都显著大于 e^+e^- 和 ep 对撞的测量结果, 暗示 LHC 能区重子似乎有共同的产生机制。

在重离子实验中, 由于底夸克质量较大, 其在 QGP 中是否会被热化一直是实验上的一个疑问。最近, ALICE 实验组观测到底强子衰变电子的椭圆流显著大于零, 这表明底夸克也有很大可能参与了热化过程。然而, 底强子衰变电子的横动量仅反映了较高横动量区间的底强子集体行为, 无法充分反映底夸克的热化行为。基于目前的探测器空间分辨率和亮度, 测量强子道的衰变产物即非瞬时粲介子 (D^0) 的椭圆流可以更好地研究底夸克与 QGP 之间的相互作用。我们首次测量质心系能量为 5.02 TeV 的非瞬时 D^0 介子椭圆流。实验结果暗示非瞬时 D^0 介子椭圆流可能大于 0 (显著性为 2.7σ) 且与非奇异 D 介子椭圆流有着 3.2σ 的显著区分度。结果在误差范围内和理论模型一致, 表明底夸克在 QGP 中可能部分热化了。该实验结果对于 ALICE Run3 底强子椭圆流的测量提供了参考。

尽管本文的测量结果加深了重味夸克强子化机制和底夸克与 QGP 相互作用的理解, 并且给理论模型提供了很好的约束。但要完全理解 LHC 强子碰撞中的粲和底夸克强子化过程和 QGP 相互作用机制, 还需要进一步更为精细的测量。基于全新升级内部径迹探测器的 ALICE 第三次运行收集的数据样本目前正在矫正中, 更精确的测量将围绕以下几个方面展开研究:

- 测量极低和极高的带电粒子重数区间的粲重子介子比。
- 直接测量底强子产额的带电多重数依赖。
- 将在更低横动量中心快度区测量底强子的产生截面。
- 直接在铅铅碰撞中测量底强子椭圆流以及粲重子的椭圆流。



攻读学位期间发表的学术论文

- **Observation of a multiplicity dependence in the p_T -differential charm baryon-to-meson ratios in proton–proton collisions at $\sqrt{s} = 13$ TeV**
ALICE Collaboration, PC members: B. Zhang, A. Gal, B. Volkel, C. Terrevoli, and L. Dello Stritto, G. Michele Innocenti
Phys. Lett. B 829 (2022) 137065
- **Measurement of non-prompt D-meson fraction as a function of multiplicity in pp collisions at $\sqrt{s} = 13$ TeV**
ALICE Collaboration, PC members: B. Zhang, F. Grosa, X. Peng, X. Zhang, D. Dobrigkeit Chinellato
arXiv:2302.07783; Accepted by *Journal of High Energy Physics*
- **Study of flavour dependence of baryon-to-meson ratio in proton-proton collisions at $\sqrt{s} = 13$ TeV**
ALICE Collaboration, PC members: B. Zhang, F. Grosa, X. Peng, D. Battistini, S. Trogolo with ALICE internal review, aiming for *Phys. Rev. Lett*
- **Measurement of non-prompt D^0 elliptic flow in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV**
ALICE Collaboration, PC members: B. Zhang, F. Grosa, X. Peng, A. Rossi, X. Zhang with ALICE internal review, aiming for *European Physical Journal C*
- **Analysis of heavy-flavour particles in ALICE with the O² analysis framework**
ALICE Collaboration, B. Zhang, V. Kučera, G. Michele Innocenti, F. Prino et al
EPJ Web of Conferences 251, 03064 (2021)
- **ALICE 实验内径迹系统探测器升级**
张彪、王亚平等, 原子核物理评论, 2020, 037(003) - 734~741
- **Event plane determination from Zero Degree Calorimeter at the Cooling-Storage-Ring External-target Experiment**
PC members: L. Liu, B. Zhang, H. Pei, Y. Wang, N. Xu, S. Shi
arXiv: 2302.11759; Accepted by *Nuclear Science and Techniques*
- **Classifier for centrality determination from Zero Degree Calorimeter at the Cooling-Storage-Ring External-target Experiment**
PC members: B. Zhang, L. Liu, H. Pei, S. Shi, N. Xu, Y. Wang
Submitted to *Nuclear Science and Techniques*



致 谢

噫吁！不知不觉已过而立之年！疫情三年贯穿整个博士求学生涯，但求学之路本就不可能一帆风顺，迷茫、痛苦、挫折总是伴随整个旅途。即便如此，依然有大量可爱且真诚的人陪我渡过这段人生轨迹。

首先感谢我的导师王亚平教授，您丰富且深邃的学识令我倾佩。您积极推动我在 ALICE 合作组中的重味物理分析，并且提供机会让我去 CERN 进行访问，虽然最终由于疫情的原因没有去成，但是我依旧感激您给予的机会。您每次组会上给我的建议总是能够为我指明科研方向。

其次感谢 ALICE 中国组的同事们。感谢和彭欣烨师兄在学术上的大量讨论，没有这些努力的话，我们是不可能取得如今的成果。同时感谢张晓明老师，不仅在实验分析上给出有用的建议，而且还特别地传授我作报告的技巧。特别感谢殷老师和周老师一路以来的鼓励和支持。感谢毛老师以及裴老师在组会上给出的各种建议。感谢朱剑辉师兄在重味物理方向上的讨论以及无私分享自己对于找工作的见解，这让现阶段的我感触颇深。

感谢其他 ALICE PWGHF 组的同事们，特别是 Fabrizio Grosa 博士和 Gian Michele Innocenti 博士。和他们一起工作的日子让我收获颇丰且成长巨大，让我深刻明白了“只有和优秀的人在一起工作，你才可能变得更加优秀”这个道理。

感谢所有的小伙伴们对我科研和生活上的帮助：侯永珍、丁艳春、张明宇、范凤、崔捧瑶、朱亚、徐冉、徐浪、张茂林、张震、方涛、柳东海、程甜甜、张定伟、吴锦、傅川、刘利珂、张李昂、韩君仪、李鸿灿、顾元涛、安静、薄文云、文良涛、陈耀忠。感谢理论组的吴祥宇、汪磊、邢文静，七楼硅像素实验室的李自立、张伟，赵聪和游必辉，和你们的讨论极大丰富了我的知识面。也特别感谢我的 6 年室友兼好友郭文达一直以来的包容和关心。

最后感谢我的妻子周颖洁一路以来的鼓励和陪伴，你聪明、勤奋且好学，一直是我心中的榜样。你取得的科研成就也一直激励着我努力向你追赶。感谢我的爸妈和姐姐姐夫们对我求学路上物质和精神上的支持，往后余生我会用行动来回报你们的期待和支持。

张彪

2023 年 5 月 20 日于桂子山