Alma Mater Studiorum · Università di Bologna

Dottorato di Ricerca in Fisica
Ciclo XXXV

Settore Concorsuale: 02/A1 - Fisica Sperimentale Delle Interazioni Fondamentali

Settore Scientifico Disciplinare: FIS/01 - Fisica Sperimentale

TITOLO TESI

Study of charm hadronisation in pp collisions at $\sqrt{s} = 13$ TeV with the ALICE experiment

Presentata da:
Marco Giacalone

Supervisore:
Dr. Pietro Antonioli

Coordinatore Dottorato:
Prof. Michele Cicoli

Co-supervisore:
Prof. Andrea Alici

Esame finale anno 2023

To my family

Abstract

In high-energy hadron collisions, the production at parton level of heavy-flavour quarks (charm and bottom) is described by perturbative Quantum Chromo-dynamics (pQCD) calculations, given the hard scale set by the quark masses. However, in hadron-hadron collisions, the predictions of the heavy-flavour hadrons eventually produced entail the knowledge of the parton distribution functions, as well as an accurate description of the hadronisation process. The latter is taken into account via the fragmentation functions measured at e^+e^- colliders or in ep collisions, but several observations in LHC Run 1 and Run 2 data challenged this picture.
In this dissertation, I studied the charm hadronisation in proton-proton collision at $\sqrt{s}$ = 13 TeV with the ALICE experiment at the LHC, making use of a large statistic data sample collected during LHC Run 2.
The production of heavy-flavour in this collision system will be discussed, also describing various hadronisation models implemented in commonly used event generators, which try to reproduce experimental data, taking into account the unexpected results at LHC regarding the enhanced production of charmed baryons. The role of multiple parton interaction (MPI) will also be presented and how it affects the total charm production as a function of multiplicity.
The ALICE apparatus will be described before moving to the experimental results, which are related to the measurement of relative production rates of the charm hadrons $\Sigma_c^{0,++}$ and Λ_c^+, which allow us to study the hadronisation mechanisms of charm quarks and to give constraints to different hadronisation models. Furthermore, the analysis of D mesons (D^0, D^+ and D^{*+}) as a function of charged-particle multiplicity and spherocity will be shown, investigating the role of multi-parton interactions.
This research is relevant per se and in the wider context of the mission of the ALICE experiment at the LHC. ALICE is devoted to the study of Quark-Gluon Plasma, which is a state of matter predicted by the QCD theory where quarks and gluons are deconfined. This is done by reproducing energy densities and temperatures similar to the one existing few microseconds after the Big Bang, on a smaller scale via relativistic heavy-ion collisions. Charm and beauty quarks are a powerful probe of that state since they are produced by the hard scattering at the first stages of the collisions and they can interact with the medium during its whole lifetime. Therefore, studying heavy-ion collision events requires a precise baseline of the expected heavy-flavour hadron production, which is provided by proton–proton collisions.

Contents

Introduction

The ALICE experiment at the LHC is devoted to study the Quark Gluon Plasma (QGP), which is a deconfined state of matter created at very high temperature and energy densities via ultra-relativistic heavy-ion collisions. The entire experiment has been designed to detect the extreme amount of particle multiplicity developed in these collision systems and to withstand the impressive amount of data that the analysis of these processes require.
For the sake of studying the QGP, it is also important to have a reference in proton-proton and proton-nucleus collisions in order to be able to discern different production mechanisms that might exist between the different collision systems, for example, when particles interact with the newly generated plasma medium.
Analysing pp collisions also provides insight on the Quantum Chromo-dynamics (QCD) properties of the hadronic matter and gives the opportunity to test theoretical estimates of perturbative QCD (pQCD) calculations.
This dissertation is based on analyses performed in pp collisions at $\sqrt{s} = 13$ TeV with data collected during Run 2, which is the datataking period lasting from April 2015 to December 2018.
In the first chapter, the physics studied in ALICE and the QGP are introduced, describing the main properties of the strong nuclear interaction and of the extreme state of hadronic matter. The main focus of this section is to highlight the importance of pp collision studies as a tool to understand QCD and as an important prerequisite to analyse QGP observables.
In the second chapter, the open heavy-flavour production in pp collisions will be illustrated in its basic mechanisms, and a detailed report about different hadronisation models related to Monte Carlo generators will be provided. These generators will be used later for comparison with the obtained results.
The ALICE experiment subdetectors used for this dissertation together with the online and offline software framework that was used to perform simulations and analyses of the experimental data will be described in Chapter 3.
The last two chapters are dedicated to the analysis performed on $\Sigma_c^{0,+,++}$, Λ_c^+, and D^0 in the full multiplicity range and on the D mesons (D^+, D^0 and D^{*+}) as a function of multiplicity and spherocity. These observables are important to shed light on the contribution of Multiple Parton Interaction (MPI), which is predicted to be large in pp collisions at LHC energies, and also to study the jet-likelihood of events in different multiplicity con-

ditions.
The analysis of $\Sigma_c^{0,+,++}$ was published at the end of 2021 in Physics Review Letters, while the D mesons analysis as a function of multiplicity and spherocity was released as preliminary result during 2022 by the ALICE experiment and is planned to be published. Additional activities performed during the doctoral programme, both at Bologna and at CERN, will be described in the appendices of this thesis.

Chapter 1

pp-collision measurements as a baseline for QGP studies

Among the four fundamental interactions in nature, the standard model (SM) [1] is capable of describing three of them using the language of Quantum Field Theory (QFT) [2]. The interactions are associated to different gauge bosons carriers of the interaction charge: the electromagnetic force is carried out by photons which act between particles with electric charge, the weak interaction is carried out by $W^{\pm}$ and Z^0 bosons and affects all the particles of the SM (quarks and leptons), while the strong interaction is applied only to quarks, and its carriers are gluons. Using local gauge symmetries in different transformation groups, the SM can classify all known elementary particles (Fig. 1.1) and describe electromagnetic, weak, and strong interactions. The first two are unified by the electroweak interaction, which employs the symmetry group $SU(2)\times U(1)$, respectively to describe the weak and electromagnetic interactions, while the strong force is described by the symmetry group SU(3).

It is recognised that SM is not the definitive theory of physics because it cannot describe several physics phenomena, such as matter and antimatter asymmetry, and, overall, it does not include a description of gravitational interaction. Nevertheless, it is an extremely valuable tool for discussing hadron-hadron interactions in collider physics. In particular, the strong interaction between hadrons elementary particles (quarks and gluons) is described by the QFT of Quantum Chromo-Dynamics (QCD), whose predicted phase diagram via lattice computation suggests a deconfined state of matter, at extreme conditions of temperature and energy density (achievable in ultra-relativistic heavy-ion collisions), where quarks and gluons are not bound to each other and move freely in the medium: the Quark Gluon Plasma (QGP).

Given the physics goals of the ALICE experiment, this chapter provides an introduction to the QCD and the phase diagram of nuclear matter, focusing mostly on the reasons why it is important to study pp collisions in order to understand heavy-ion collisions data.

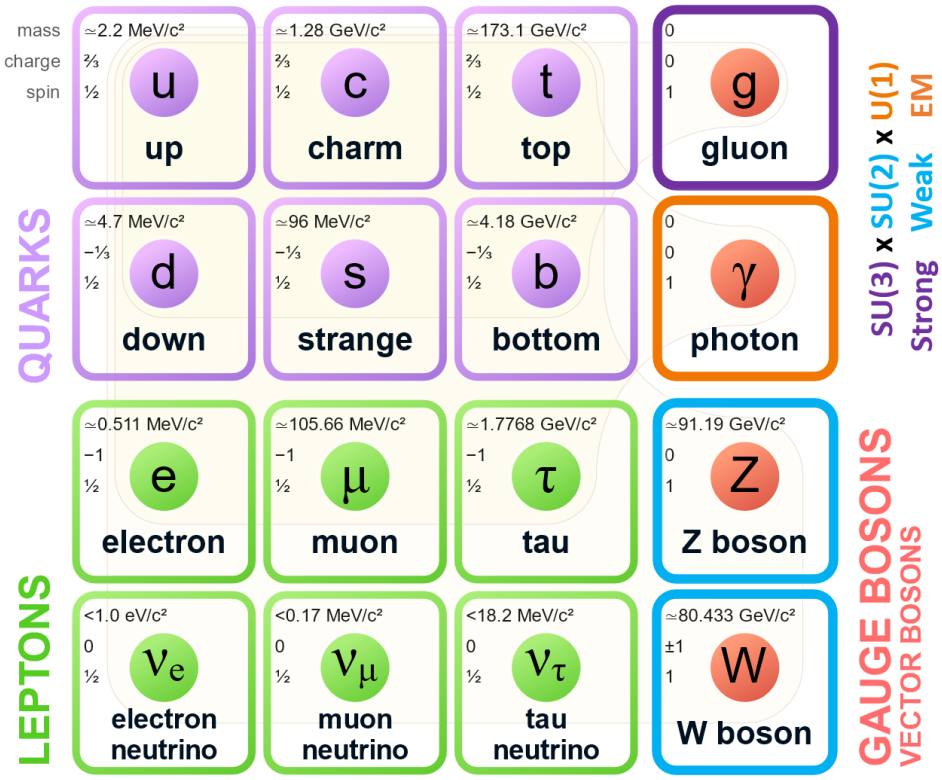


Figure 1.1: Particles described by the Standard Model and interactions connections between symmetries and particles.

1.1 Quantum Chromo-Dynamics (QCD)

Quantum Chromo-Dynamics is the SU(3) component of the Standard Model of particle physics that describes the strong interaction. This symmetry is chosen because the interaction needs three colour charges in order to be described, which are carried by quarks with six different flavours (as illustrated in Fig. 1.1) and by antiquarks with their antiflavours. In addition, eight massless gauge vector bosons, called gluons, containing a colour and anti-colour charge, carry the interaction with quarks and, differently from Quantum Electrodynamics (QED), they can self-interact with other gluons since they carry colour charge.
The Lagrangian density of QCD is expressed by:

$$\mathcal{L} = \sum_f \bar{q}_i^f (i\gamma_\mu D_{ij}^\mu - m_f \delta_{ij}) q_j^f - \frac{1}{4} G_a^{\mu\nu} G_{\mu\nu}^a, \tag{1.1}$$

in which the first term expresses the free propagation of the quark fields q_i^f with colour f and charge i and their interactions with the gluon-gauge fields, while the second one instead describes the free propagation of the 8 gluon fields G_a^μ, which are contained inside

the computation of the covariant derivative D^{μ}_{ij}:

$$D^{\mu}_{ij} = \partial^{\mu}\delta_{ij} - ig_s\left(\frac{\lambda^a_{ij}}{2}\right)G^{\mu}_a, \tag{1.2}$$

in which the strong coupling constant appears as g_s and the Gell-Mann matrices λ^a_{ij} [2] are included.
Avoiding to provide too many theoretical details, since it is not the target of this dissertation, it can be said that the gluon tensor $G^a_{\mu\nu}$ of the Lagrangian in Eq. 1.1 contains a non-Abelian term that describes self-interactions among gluon fields, leading to the anti-screening phenomenon in colour interaction.

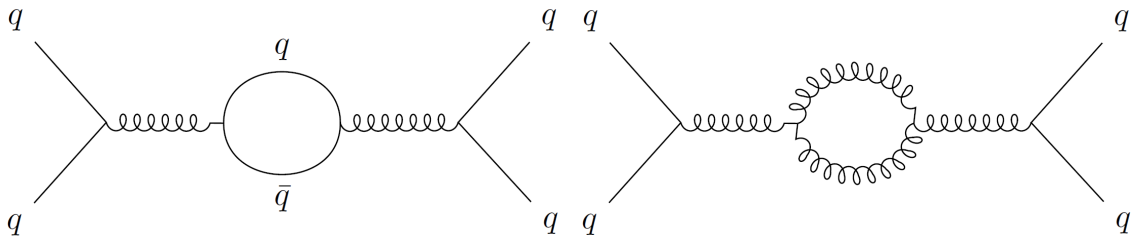


Figure 1.2: Feynman loop diagrams of strong interaction showing screening (a) and anti-screening (b) processes.

As in QED for electromagnetic interaction, the carrier can create loops of virtual particles-antiparticles in the QCD vacuum, as seen in Fig. 1.2a, effectively reducing the strong interaction strength of the initial colour charges (*screening*). However, the possibility of having loops with self-interacting gluons (Fig. 1.2b) increases the strength of strong interaction due to the fact that they carry colour charges that increase the effective charge of the process: *anti-screening*.
The latter is the dominant phenomenon in QCD and, as a result, the QCD coupling constant

$$\alpha_s = \frac{g_s^2}{4\pi} = \frac{\alpha_s(M^2)}{1 + \alpha_s(M^2)\frac{33-2n_f}{12\pi}\ln\frac{Q^2}{M^2}}, \tag{1.3}$$

calculated using the number of flavours n_f and the renormalisation mass scale M, increases when the squared transfer momentum Q^2 decreases, which means that α_s becomes small with short distances (higher Q^2), and vice versa. Many experimental results confirm the *running* trend of the coupling constant α_s as a function of momentum, as shown in Fig. 1.3.
This behaviour defines two QCD specific phenomena that occur, respectively, for low and large values of Q^2: *confinement*, in which the coupling constant diverges, making quarks strongly bound and inseparable, and *asymptotic freedom* [3], where elementary hadron particles can be considered free to move and weakly interacting due to a very

small α_s. It is worth noting that the behaviour of the running coupling constant also depends on the number of flavours in Eq. 1.3: a large n_f value could potentially cause the sign inversion of the second term in the denominator, preventing both confinement and asymptotic freedom from happening. For this to occur, the experimentally measured number of flavours should be $n_f > 33/2$, which, however, is not the case.

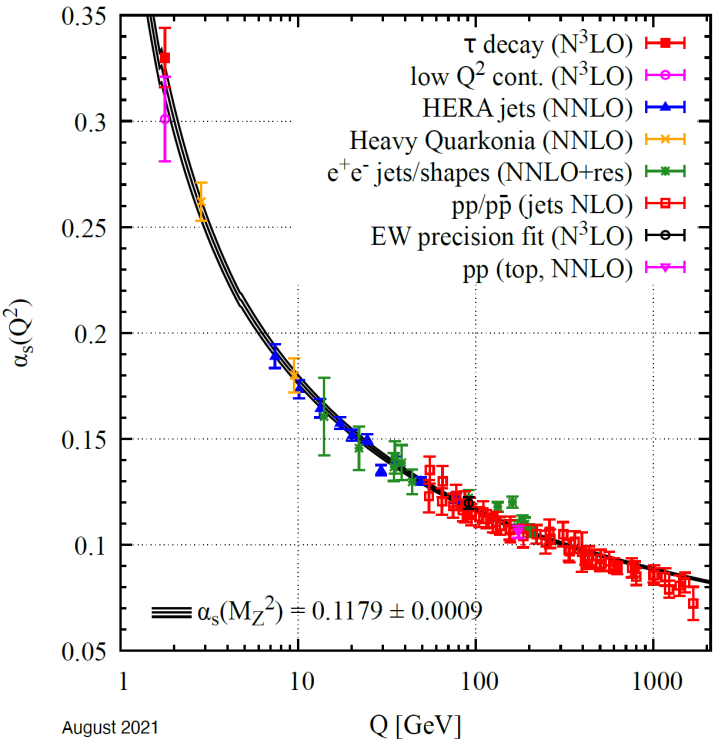


Figure 1.3: α_s results as a function of the energy scale Q. The calculations were performed using multiple degrees of QCD perturbation theory [4].

A perturbative approach (pQCD) can be used for the calculation of α_s with values of Q^2 higher than 1 GeV$^2/c^2$, but this approach cannot be applied with lower momentum due to the higher value of the coupling constant. In the latter case, the QCD Lagrangian can be evaluated on a discrete space-time grid (lattice QCD calculations).
Figure 1.3 shows different levels of perturbation theory applied to the calculation of the running coupling constant, going from next-to-leading (NLO) to next-to-next-to-NLO (N^{3}LO). The estimates (shown as points in the plot) are determined from various experimental measurements of peculiar processes. These values are compared with a parametrisation of α_s at the renormalisation mass scale of the weakly interacting Z^0 boson, shown as a 3-lines curve, and with the latest precise global average $\alpha_s = 0.1179 \pm 0.0009$ [4].

1.2 Phase diagram of QCD

The behaviour of the running coupling constant α_s leads to different properties of strongly interacting matter as a function of the physics conditions in which it is considered.
If we took into account extreme conditions of energy densities and temperature, we would experience a state of matter that behaves differently from the ordinary one (embedded in atomic nuclei), and these conditions were present few microseconds after the Big Bang. The predicted state of matter is the Quark-Gluon Plasma, which can be described as a mixture of quarks and gluons completely deconfined. Experimentally, the conditions under which we could create QGP are achievable in the laboratory, thanks to heavy-ion collisions at extremely high energies.

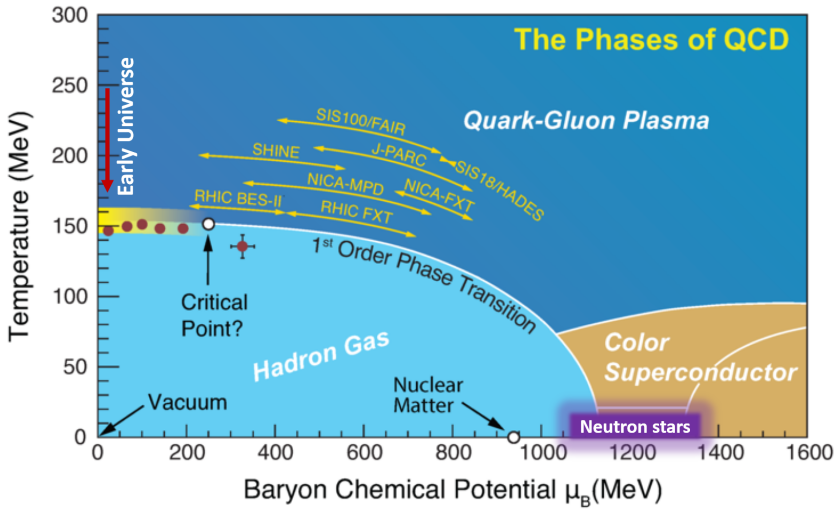
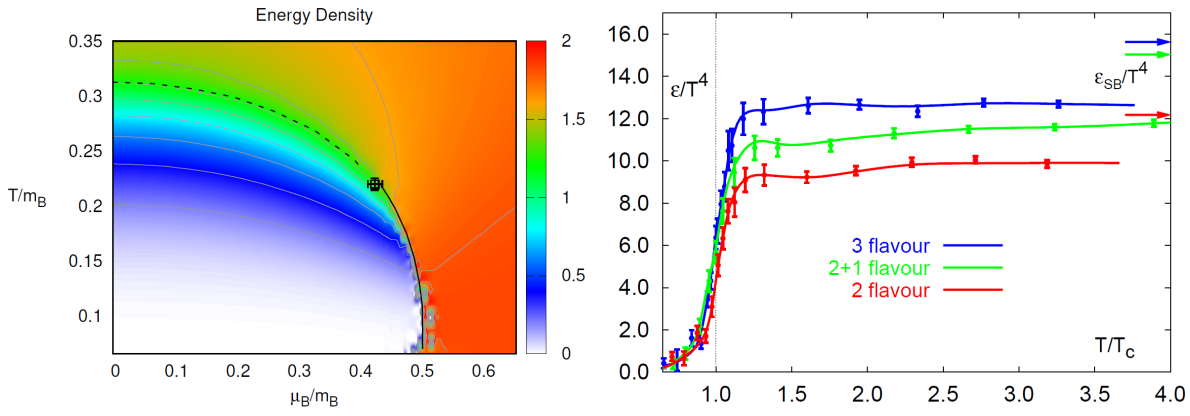


Figure 1.4: QCD phase diagram showing the application ranges of various experiments [5].

The behaviour of QCD under different physical conditions can be shown via a phase diagram, as it is done in Fig. 1.4 where the phase is expressed by the temperature as a function of the baryochemical potential μ_B, which is the energy needed to increase by one the total baryon number ($\propto$ total number of baryons per unit of volume - *net baryon density*).
Nuclear matter in standard conditions in the present Universe is located in the phase diagram around 1 GeV with a temperature close to zero (1 eV $\approx$ 11600 K)[6]. Different conditions could be met in order to have a phase transition, e.g. leaving the temperature close to zero and increasing the baryochemical potential from standard conditions will eventually lead to a QGP, conditions which could be reached eventually inside neutron stars. These specific stellar objects are supposed to have a temperature below 1 MeV, so their state of matter basically lies along the x-axis of Fig. 1.4. This phase is predicted to begin to have Cooper colour pairs (pairs of quantum-state bound coloured quarks), which result in colour superconductivity. The window phase of colour superconductivity

is quite large and is not experimentally achievable in a laboratory, so the target of experimental physics relies on lower values of μ_B.
Considering the limit of the vacuum state at $\mu_B = 0$ and increasing the temperature to a critical value ranging from $T_c \approx 145 - 165$ MeV (which corresponds to an energy density $\epsilon = 0.18 - 0.5$ GeV/fm^3), the conditions at the early stages of the Universe are theoretically reached (QGP after the Big Bang) and a transition should occur as a fast cross-over between hadronic matter and QGP. This behaviour should hold at low-μ_B until a certain point in which the region is replaced by a first-order phase transition. The relativistic heavy-ion experiments at accelerators try to probe these two phase regions. In particular, LHC is able to study the former ($\mu_B \approx 0$) thanks to the high centre of mass energies achieved, while multiple other facilities can only inspect the QCD phase space at lower temperatures and higher baryochemical potential values (as shown for example in Fig. 1.4 for RHIC experiments [5]).
The description of QCD thermodynamics and the estimate of the cross-over region and the critical point can be performed thanks to lattice QCD (lQCD) calculations, which can quite well describe the phase transition, as shown in Fig. 1.5a, where the energy density is expressed in the form of temperature as a function of the baryochemical density (both divided by the baryon mass). In the plot a clear distinction between cross-over region and first-order phase transition seems to happen [7]: when QCD deconfinement is involved, it is expected that new degrees of freedom are added to the system at a phase transition temperature, leading to unbound quarks and gluons now able to contribute to thermodynamic observables like the energy density.



(a) (b)

Figure 1.5: Simulation with lattice QCD of the energy density as a function of baryochemical density (a) [7] and lattice calculation of the energy density as a function of temperature (b) for different number of flavours. The (2+1)-flavour curve implements 2 light flavours plus a four-times heavier strange quark mass [8]

As shown in Fig. 1.5b, this last quantity reflects the relevant number of degrees of freedom, providing a significant contribution in the high-temperature limit, and this

drastic change in the behaviour of the energy density, when the temperature approaches the critical value T_c, indicates a QCD phase transition to a deconfining plasma state of matter. Assuming an infinite temperature, the observable is expected to reach approximately its free gas value, which is provided by the Stefan-Boltzmann law and it is illustrated on the right side of the plot [8]. However, the significant discrepancies obtained in the calculations suggest that QCD matter behaves more like a fluid at that temperature.

1.3 Quark Gluon Plasma in heavy-ion collisions

The analysis of QGP in experimental facilities such as the LHC is related to a baryochemical potential value close to 0 and temperatures well above T_c, which are achievable via ultra-relativistic heavy-ion collisions.

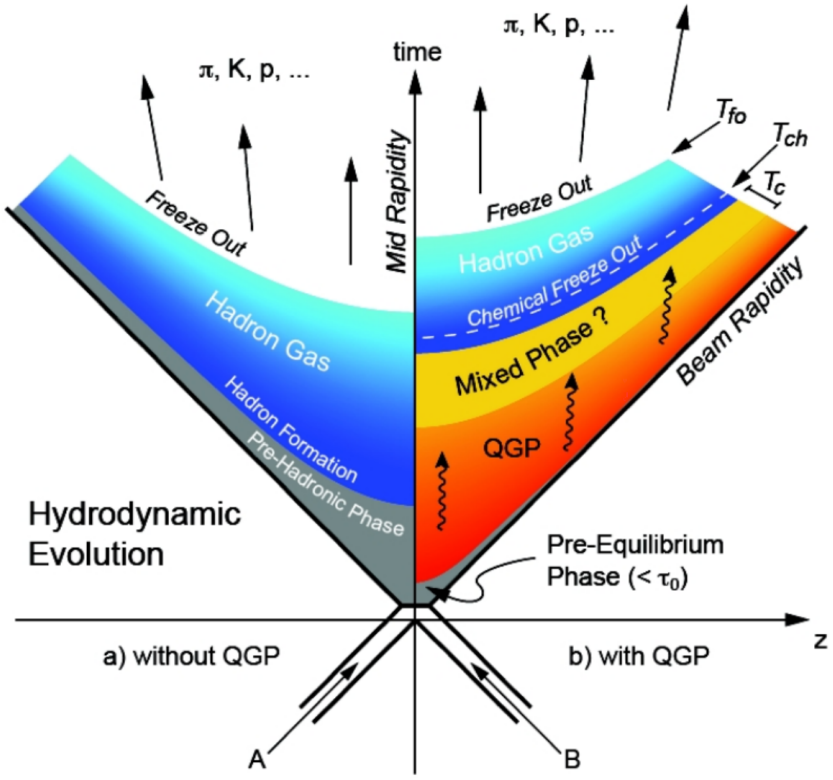


Figure 1.6: Space-time evolution of heavy-ion collisions without and with QGP.

Experimental activities in such a field began in particle accelerators in the 1980s, first with fixed target experiments at the Alternating Gradient Synchroron (AGS) at Brookhaven National Laboratory (BNL) using gold ions at $\sqrt{s_{\mathrm{NN}}} = 5$ GeV and at the Super Proton Synchrotron (SPS) at CERN using lead nuclei at $\sqrt{s_{\mathrm{NN}}} = 17$ GeV, and then with collider experiments until today, in which the colliding nuclei are not stopped at the collision

point and, due to the high energies, a net-baryon density close to zero is reached (Bjorken or transparency regime [9]).
The current top energies achieved by the colliders are 200 GeV with Au ions with the Relativistic Heavy Ion Collider (RHIC) at BNL and 5.5 TeV with Pb ions with LHC at CERN.
Fig. 1.6 illustrates the various phases of beam collisions with and without QGP using relativistic hydrodinamic models. On the left side of the illustration, the case without QGP formation in heavy-ion collisions is represented, so after the beams collision the hadron formation would follow directly the initial stage of the collision, originating in the end the particles that would be detected experimentally. On the other hand, on the right side of Fig. 1.6, the QGP evolution in heavy-ion collisions is shown, which is composed of various steps:

- **Pre-Equilibrium** → Two highly Lorentz-contracted nuclei collide, creating the maximum energy density, but the originating system is not in equilibrium. This phase is supposed to last less than ≈1 fm/c and thanks to the multiple scattering occurring among nuclei partons, the energy density values are well above the limit to generate QGP. In particular, using measurements obtained at RHIC and LHC, and the Bjorken formula, one can obtain the initial energy densities achieved at central rapidity[1] when two nuclei collide, which are respectively $\epsilon_{RHIC} = 5.4$ GeV/fm^3 and $\epsilon_{CERN} = 15$ GeV/fm^3. Assuming a time τ_0 after the collision, the Bjorken formula states that the energy density at central rapidity y with a $\mu_B \approx 0$ can be estimated as
$$\epsilon_{\mathrm{BJ}} = \frac{1}{\mathcal{A} \cdot \tau_0} \frac{\mathrm{d}\langle E_{\mathrm{T}}(\tau_0)\rangle}{\mathrm{d}y}\bigg|_{y=0}, \tag{1.4}$$
where E_{T} is the transverse energy and $\mathcal{A}$ is the transverse area of the collision region between nuclei [9, 10];

- **Quark Gluon Plasma** → At approximately 1 fm/c the system of generated particles reaches local thermal equilibrium while the partonic matter expands and starts to cool down. This phase can last up to 10 fm/c (in the LHC);

- **Hadronisation** → With the system cooling down there will be local regions in which the temperature will decrease below the critical value T_c and the system will move through the cross-over region between QGP and hadronic matter. At this point, the medium will start to become an interacting gas of hadrons;

- **Freeze-out** → This phase is split in two: the first process to take place is the chemical freeze-out in which the inelastic interactions between hadrons stop (below a certain temperature) and the relative abundances of the hadronic species

[1] Variable defined along the z-axis as
$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}.$$

are fixed; afterwards, the thermal (or kinetic) freeze-out will cause all the elastic interactions to stop, leading to the definition of the momentum distribution of the particles, which will now reach the various detectors of the experiment.

Overall, the lifetime of the QGP is highly dependent on the hydrodynamic scenarios in which the plasma is developing, including initial conditions and the type of expansion of the system. The faster the expansion of the system in the transverse plane, for example, the shorter the plasma lifetime will be. On the other hand, a higher initial energy density will lead to a longer lifetime: the plasma lives approximately 5-10 fm/c respectively in RHIC and LHC ultra-relativistic heavy-ion collisions, approximately $3 \cdot 10^{-23}$ s, but theoretical estimates suggest that the order of magnitude of the energy released by the Big Bang was able to make the QGP live for microseconds, a difference of 17 orders of magnitude compared to what scientists are able to achieve right now in the largest laboratories with the most powerful colliders in the world [11].

1.4 pp-collision measurements as a baseline for QGP studies

Considering the extremely short lifetime of QGP in heavy-ion colliders, the only way to study the existence of the plasma is via indirect probes. To assess the effect of the plasma medium on produced particles, it is necessary to provide a baseline for those measurements using more elementary collision systems. In LHC, in particular, studying Pb–Pb collisions requires both pp and p–Pb data beforehand as a reference.
Furthermore, results obtained from the former systems are extremely useful for studying the properties of QCD and to test predictions provided by perturbative QCD calculations.

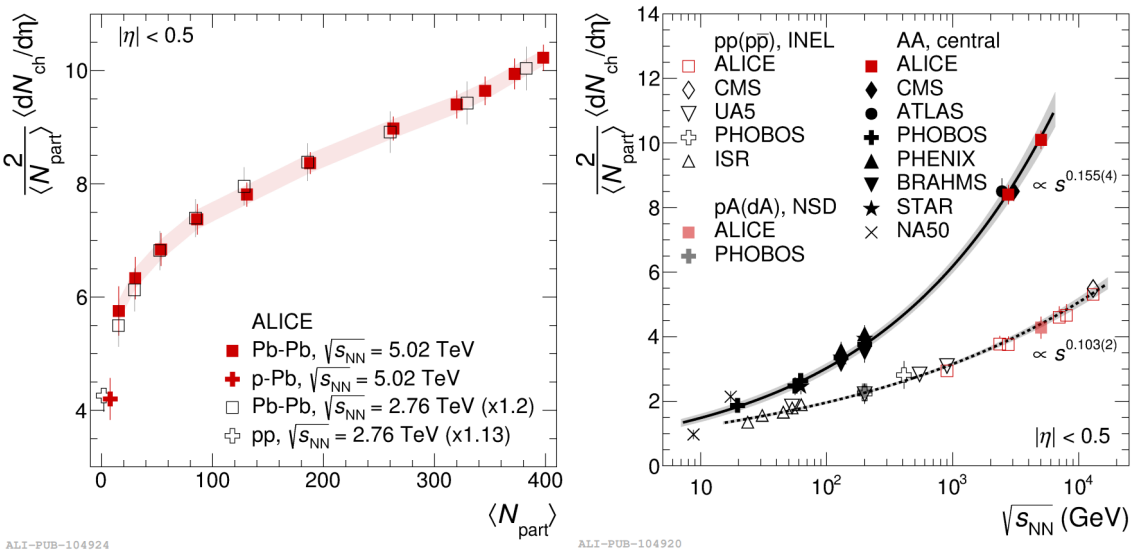
The simplest observable that provides important information about the properties of the medium in heavy-ion collisions is particle multiplicity. From the Bjorken formula discussed earlier, one can calculate the energy density via the charged-particle multiplicity measured per unit of rapidity in the experiment by replacing the transverse energy with the average transverse mass multiplied by the number of particles:

$$E_{\mathrm{T}}(\tau_0) \approx N(\tau_0) \cdot m_{\mathrm{T}} = N\sqrt{(m^2 + p_{\mathrm{T}}^2)}, \tag{1.5}$$

$$\epsilon_{\mathrm{BJ}} = \frac{\langle m_{\mathrm{T}} \rangle}{\mathcal{A} \cdot \tau_0} \frac{\mathrm{d}N(\tau_0)}{\mathrm{d}y}\bigg|_{y=0}, \tag{1.6}$$

considering a QGP thermalisation time $\tau_0 \approx 1$ fm/c. Hence, from the Bjorken model there is a strict correlation between energy density and the number of charged particles, which is usually expressed in experimental results as:

$$\text{Charged particle pseudorapidity density} = \frac{2\langle \mathrm{d}N_{\mathrm{ch}}/\mathrm{d}\eta \rangle}{\langle N_{\mathrm{part}} \rangle}, \tag{1.7}$$



(a) (b)

Figure 1.7: Charged-particle pseudorapidity density per participant-nucleon pair as a function of the number of participant nucleons in the ALICE experiment (a) and comparison with various measurements as a function of $\sqrt{s_{\rm NN}}$ (b) [12].

where the factor $2/\langle N_{\rm part}\rangle$ is introduced to compare different collision systems, and the pseudorapidity η can be used as a rapidity approximation when the transverse momentum is much higher than the mass value (as it happens in ultra-relativistic collisions)[2].
Measurements by the ALICE experiment in pp, p–Pb and Pb–Pb as a function of $\langle N_{\rm part}\rangle$ are illustrated in Fig. 1.7a. It is clear that the charged-particle pseudorapidity density increases as a function of the number of participant collision particles, but the rising slope changes substantially when $\langle N_{\rm part}\rangle \approx 50$, showing a much softer increase for higher values. A similar trend is visible for Pb–Pb collisions at both $\sqrt{s_{\rm NN}} = 5.02$ and 2.76 TeV [12].
On the other hand, Fig. 1.7b shows the multiplicity density of charged particles as a function of the centre of mass per pair of nuclei using measurements performed by different collaborations and in different collision systems. Increasing $\sqrt{s_{\rm NN}}$ causes a steeper rising trend for nuclei-nuclei collisions than the results obtained in p–nuclei, but it is interesting to notice that the p–A and d–A results lie on the curve of pp measurements, indicating that other mechanisms other than multiple interactions are involved in the production.
Furthermore, a simple calculation using the Bjorken formula (Eq. 1.6) with the charged-

[2]This quantity is experimentally easier to measure than rapidity and it is defined as

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

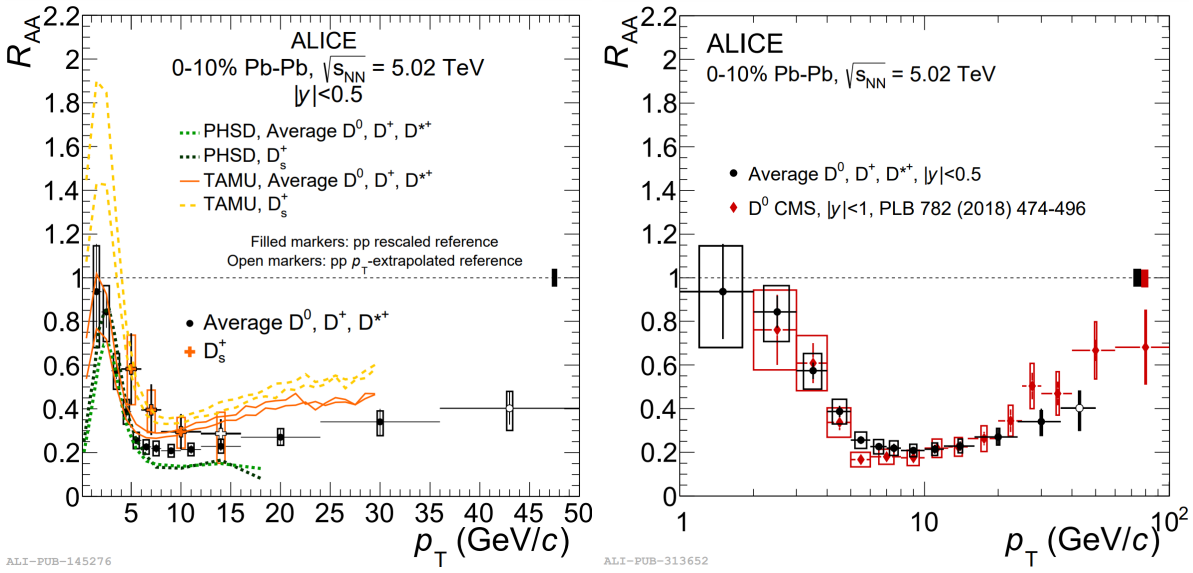
where θ is the angle between the momentum of the particles and the z axis.

particle density values obtained in heavy-ion collisions, leads to energy densities larger than the estimated one needed for the creation of a deconfined state [9, 12].
In addition to the energy density, the collision geometry can also be studied by comparing pp collisions with the Pb–Pb and p–Pb measurements.
An initial geometrical description can be provided by the Glauber model calculations [13], which help to estimate important geometrical parameters that are precluded from direct observation due to the femtoscopic length scales of the events. These are, for example, the impact parameter b, the number of participating nucleons N_{part}, and the number of binary nucleon-nucleon collisions N_{coll}.
However, the Glauber model describes the collision in a regime called the *optical limit*, i.e. the collision between two nuclei must be described as an incoherent superposition of binary nucleon-nucleon interactions (with point-like indistinguishable protons and neutrons) and, over a certain energy, the nucleons will carry enough momentum to be undeflected as the nuclei pass through each other [13].
The theoretical model is capable of showing that the cross section of hard processes depends on $\langle N_{\mathrm{coll}} \rangle$ and this hypothesis can actually be tested by analysing a physical quantity called the nuclear modification factor (R_{AA}) which is:

$$R_{\mathrm{AA}} = \frac{1}{\langle N_{\mathrm{coll}} \rangle} \frac{\mathrm{d}^2 N_{\mathrm{AA}}/(\mathrm{d}p_{\mathrm{T}}\mathrm{d}y)}{\mathrm{d}^2 N_{\mathrm{pp}}/(\mathrm{d}p_{\mathrm{T}}\mathrm{d}y)}, \tag{1.8}$$

where the double-differential yields ratio between the Pb–Pb and pp data is corrected for the average number of binary nucleon-nucleon collisions.

A deviation from unity of the R_{AA} value implies modifications of the p_{T} distribution of the produced particles due to interactions with the medium, that is the presence of QGP (hence, a reference of pp collision measurements is fundamental), or due to other nuclear processes called cold-nuclear matter effects (CNM) that are not related to the creation of a deconfined state. These initial and final-state effects, originating from the bounding of nucleons inside the nuclei, can be studied by analysing the nuclear modification factor in p–A collisions R_{pA}, which is calculated in a manner similar to Eq. 1.8 by considering the differential cross section of p–Pb, and they include phenomena such as nuclear shadowing, gluon saturation, and nuclear absorption [16].
Furthermore, the energy loss of high-p_{T} partons in heavy-ion collisions via inelastic interactions (gluon radiation or *gluonsstrahlung*) with the QGP causes a shift in the energy distribution, which leads to a softening of the final-state hadron p_{T} distribution. In general, the production of high-energy partons is related to hard-scattering processes in which two partons from the initial-state hadrons interact with a large momentum exchange producing high-energy back-to-back partons. These will generate a series of hadrons from the parton shower caused by high-energy parton fragmentation, the so-called *jets*. The energy loss in the QGP medium by the jets is called *jet-quenching* and it was first observed at RHIC in 2003 [17], by the STAR Collaboration comparing data in d–Au and Au–Au collisions. This phenomenon can also be estimated using the R_{AA}, by comparing the heavy-ion collision results with measurements obtained in small systems (i.e. pp and p–Pb collisions).



(a) (b)

Figure 1.8: Average non-strange D mesons (D^0, D^+ and D^{*+}) and D_s^+ nuclear modification factor as a function of p_T compared to models (a) and R_{AA} comparison of the average D mesons ALICE measurements with D^0 results obtained by the CMS experiment (b) [14, 15].

Some examples of nuclear modification factor measurements performed by the ALICE experiment are shown in Fig. 1.8a, comparing the R_{AA} values measured for D mesons [14] as a function of p_T with TAMU [18] and the Parton-Hadron-String Dynamics (PHSD)[19] models in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV with a centrality[3] class of 0-10%. The same results are compared with CMS D^0 measurements [15] in Fig. 1.8b. The models provide a good description of the general qualitative trend of the results and the measurements obtained by the two experiments are consistent within the uncertainties.
Another key signature of QGP formation is the enhancement of strangeness in heavy-ion collisions compared to smaller collision systems, as predicted by Rafelski and Muller [20]. The first observations of this phenomenon in Pb–Pb collisions were obtained at CERN by the WA97 and NA57 [21, 22] experiments, whose enhancements were directly proportional to the quantity of strange quarks contained in the analysed hadrons.
When the temperature is below the critical value T_c, strange production is disfavoured due to the effective mass of the quark being larger than the temperature of the system (≈ 0.5 GeV/c^2). However, when the critical temperature is reached by the system, the chiral symmetry restoration in QGP causes the effective mass of the quark to decrease up to a factor 10, increasing the abundance of strangeness production in heavy-ion collisions. In fact, pairs of $s\bar{s}$ are expected to be abundantly produced through gluon fusion

[3]Percentile of hadronic cross section in heavy-ion collisions. The centrality classes can be estimated by measuring the particles multiplicity and exploiting the Glauber model [13].

in the QGP, given the high density of gluons in the system, leading to an increase in the relative production of hadrons with strangeness.
A possible explanation for this behaviour can be linked to the suppression of strangeness in small systems via canonical suppression [23]. The grand-canonical formulation cannot be implemented in pp and p–A, for example, due to the small number of particles produced in the collisions. Considering the quantum number conservation, this means that in small systems the strangeness must be exactly conserved, whereas in a grand-canonical ensemble the quantum numbers are conserved on average. As a consequence, the production of a hadron with strangeness in the QGP fireball does not require the nearby creation of another hadron with the same antistrangeness content, but this can be generated anywhere in the medium. Hence the suppression of the strange hadron production in canonical ensembles, which require, instead, the production of compensating strange/anti-strange hadrons in the same space-time position.

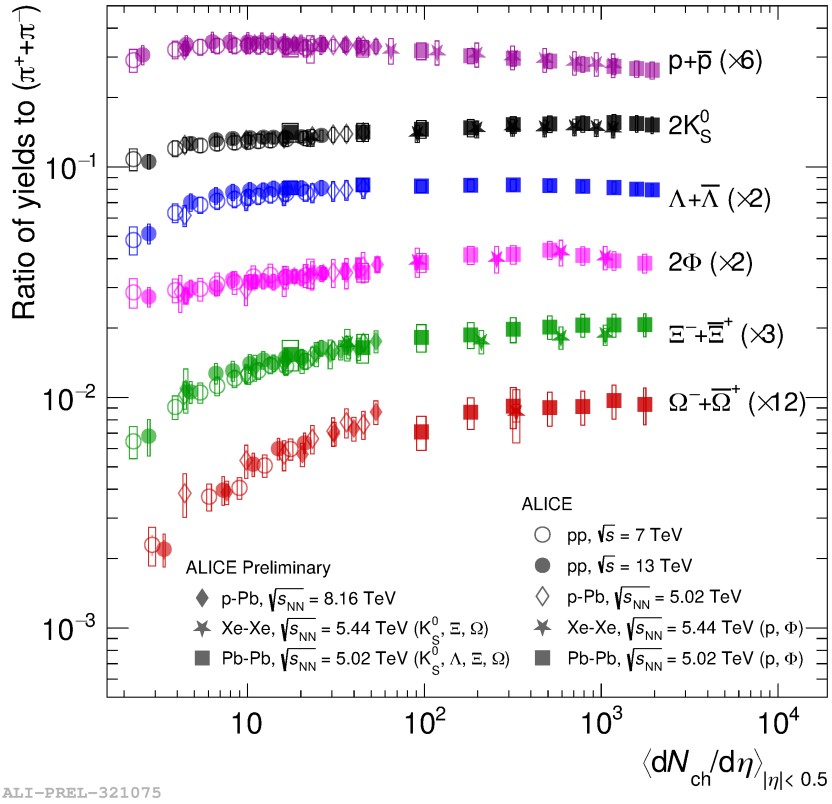


Figure 1.9: Relative yields to $\pi^+\pi^-$ of hadrons with strange quark content as a function of charged-particle multiplicity at mid-rapidity by the ALICE experiment in many collision systems.

Preliminary and published results by the ALICE experiment of the p_{T}-integrated ratio between several hadrons with strange content and the sum of π^+ and π^- in various collision systems are shown in Fig. 1.9 as a function of charged-particle multiplicity at

mid-rapidity. The measurements define a smooth increase from low multiplicity pp to Pb–Pb collisions for all the strange hadrons, defining a steeper increasing trend for the particles containing a higher strangeness content. These results confirm those obtained at RHIC in Cu–Cu and Au–Au collisions by the STAR Collaboration [24] and the previous ALICE results in Pb–Pb at $\sqrt{s_{\rm NN}} = 2.76$ TeV [25].
In 2016 the ALICE Collaboration first reported the surprising result [26] (also shown in Fig. 1.9) that strangeness production measurements in high multiplicity pp and p–Pb collisions are compatible to Pb–Pb collisions results at the lowest multiplicity. It seems therefore that the strangeness enhancement increases as a function of multiplicity independently of the collision system. Hence, currently this phenomenon is actively under study, but at the moment it is not yet understood if a similar underlying mechanism to heavy-ion collisions (e.g. formation of QGP) contributes to the increased production of strange quarks after a certain multiplicity threshold in small systems.

Chapter 2

Open heavy-flavour production in pp collisions

In this chapter, the heavy-flavour (charm and beauty) production and hadronisation mechanisms in pp collisions will be described in detail, given the original results presented in this dissertation. A discussion will be opened on the ways small systems are useful to test different perturbative QCD calculations, and on different hadronisation models.

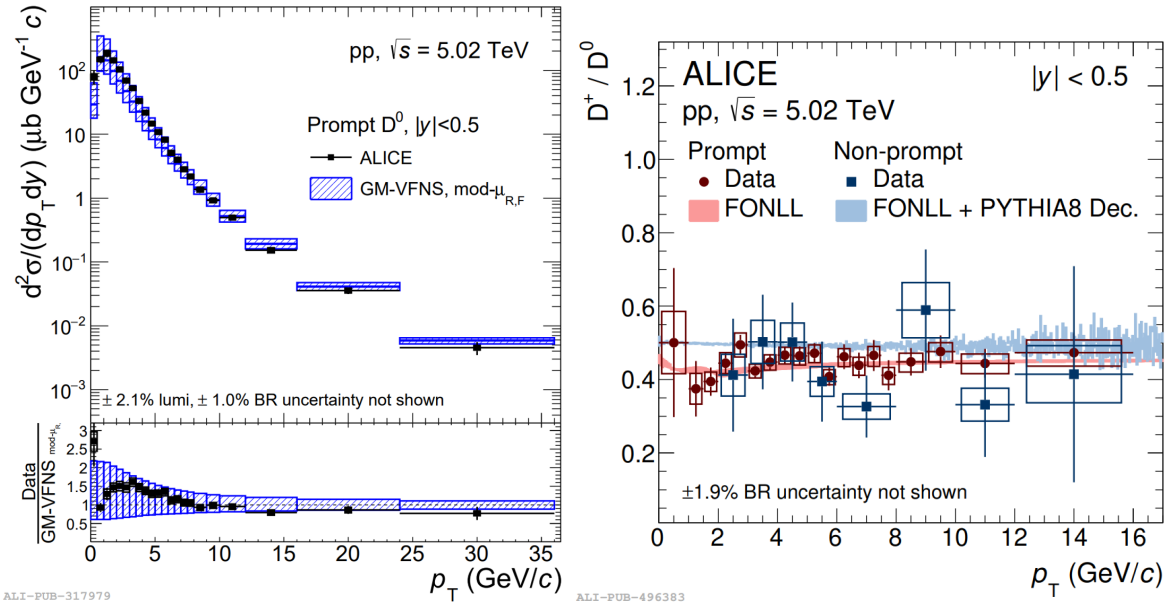
2.1 Charm and beauty production

Heavy quarks are produced in hard scattering processes at the early stages of the collision due to their large mass being of the GeV scale ($m_c \approx 1.3$ and $m_b \approx 4.2$ GeV/c^2 [4]). As a consequence of the strong interaction flavour conservation law, they are always produced in $b\bar{b}$ and $c\bar{c}$ pairs, which can later hadronise in bottomonia/charmonia states or into open heavy-flavour hadrons, which are particles with a non-zero beauty or charm quantum number.
These scattering processes require a large Q^2. The value of the coupling constant α_s is, therefore, lower than unity, so the estimation of the elementary cross section of quark-antiquark pairs production can be performed with perturbative calculations in which the main channels to heavy flavours production are via annihilation of quark-antiquark pairs $q\bar{q} \rightarrow Q\bar{Q}$ and gluon fusion $gg \rightarrow Q\bar{Q}$ [27]. Given its lower mass, charm is more abundant than beauty in ultra-relativistic collisions.
The production cross section of open heavy-flavour hadrons in pp collisions $\sigma_{\mathrm{pp}\rightarrow\mathrm{H_Q+X}}$, employing pQCD calculations, can be obtained by applying the factorisation approach [28] combining three different terms via the formula:

$$\sigma_{\mathrm{pp}\rightarrow\mathrm{H_Q+X}} = \sum_{i,j=q,\bar{q},g} f_1(x_i,\mu_f^2)\cdot f_2(x_j,\mu_f^2) \otimes \sigma_{\mathrm{ij}\rightarrow\mathrm{Q\bar{Q}}}(\alpha_s(\mu_R^2)) \otimes D_Q^{\mathrm{H_Q}}(z,\mu_f^2). \tag{2.1}$$

In detail, the first term is the product of the parton distribution functions (PDFs) $f_{1/2}$ of the colliding protons (labelled 1 and 2), which expresses the probability to locate

a specific parton (i, j) inside the colliding proton carrying a fraction $x_{i,j}$ of the total proton momentum (Bjorken-x); the second one is the production cross section of a quark-antiquark pair via parton interactions (hard scattering); while the third one is the fragmentation function (FF) component, which describes the probability of the hadronisation of a heavy quark Q in a hadron H_Q keeping a fraction z of the original quark momentum.
Both PDFs and FF depend on the energy scale of the process, so they are expressed as a function of the factorisation scale μ_F, which is related to the value of Q^2. The former are usually estimated from measurements of deep-inelastic scattering processes (for example, by the NNPDF Collaboration [29]), while the latter are obtained from hadronic production measurements performed in e^+e^- collisions, assuming the universality of fragmentation functions in different collision systems. Furthermore, these terms are assumed at a given energy scale Q^2, so the values at a specific energy are calculated employing the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [30].
The hard scattering cross section is calculated instead as a perturbative expansion on the strong nuclear coupling constant assumed at a renormalisation scale μ_R, which is usually chosen of the same order as the momentum transfer of the hard process. Both μ_R and μ_F can be approximated to the value $\sqrt{m_Q^2 + p_{T,Q}^2}$.



(a) (b)

Figure 2.1: Production cross section of prompt D^0 (a) [31], and D^+/D^0 production ratios (b) [32] as a function of p_T compared to pQCD calculations.

At the time of writing this dissertation, the most precise predictions of the production cross section for open heavy-flavours with next-to-leading order accuracy (NLO) are provided by FONLL [33] and GM-VFNS [34] groups, which both implement, in ad-

dition, next-to-leading logarithmic corrections (NLL). These are important because the computed cross sections contain mass-dependent logarithmic terms that will dominate and diverge at higher interaction scales ($p_{\mathrm{T}} \gg m_q$), so they must be resummed in the final NLO results [35].
ALICE measurements in pp collisions at $\sqrt{s} = 5.02$ TeV are compared with these pQCD calculations in Fig. 2.1. In particular the differential cross section of the prompt[1] D^0 meson is illustrated in Fig. 2.1a as a function of p_{T} compared to GM-VFNS calculations, while the mesons production ratio $\mathrm{D}^+/\mathrm{D}^0$ is shown in Fig. 2.1b both in the prompt and non-prompt cases compared respectively with FONLL, and FONLL including the PYTHIA8 decayer (which is a Monte Carlo event generator that will be discussed in Sec. 2.2).
For mesons production, the pQCD calculations describe well the measurements both in the prompt and non-prompt production, but it is not the same for the baryons production.

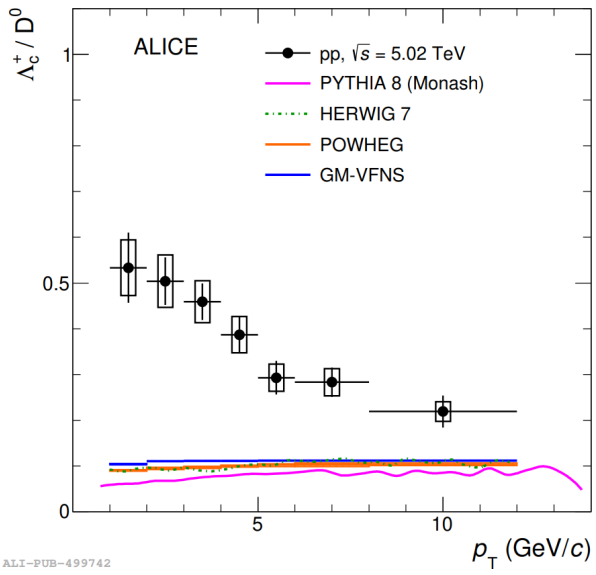


Figure 2.2: $\Lambda_{\mathrm{c}}^+/\mathrm{D}^0$ production ratio as a function of p_{T} compared to Monte Carlo models parametrised using $\mathrm{e}^+\mathrm{e}^-$ collisions results [36].

The baryon-to-meson ratio is a valuable physical observable sensitive to hadronisation mechanisms, so it is often used to probe possible differences in different collision systems and to analyse particle production. For the integrated-p_{T} case, average results obtained from experiments at LEP in e^+e^- collisions showed a $\Lambda_{\mathrm{c}}^+/\mathrm{D}^0$ ratio of $0.113 \pm 0.013 \pm 0.006$, which is calculated with fragmentation fractions[2] derived from measure-

[1] *Prompt* indicates the production of a specific hadron with a charm quark that is directly produced in the hard scattering, while *non-prompt* production refers to the decay of a beauty hadron in a charm hadron.

[2] Probability of a specific quark to fragment into a given hadron H.

ments obtained for both species [37]. This value is largely lower than the ALICE results obtained in pp and p–Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV, which are respectively $0.51 \pm 0.04 \pm 0.04$ and $0.42 \pm 0.03 \pm 0.06$ [36]. Moreover, ALICE measurements as a function of $p_{\rm T}$ in pp collisions at $\sqrt{s} = 5.02$ TeV in Fig. 2.2 are also inconsistent with models results that use e^+e^- FF measurements as reference: the former show a decreasing trend of the ratio with increasing $p_{\rm T}$, while e^+e^- based estimates have an overall flat trend in $p_{\rm T}$. Specifically, ALICE results are compared, in addition to pure GM-VFNS pQCD calculations, with estimates from the POWHEG pQCD framework [38] and with the commonly used event generators HERWIG7 [39] and PYTHIA8 [40] with the Monash tune (see Sec. 2.2), but none of them manages to adequately describe neither the trend nor the magnitude of the ALICE results.

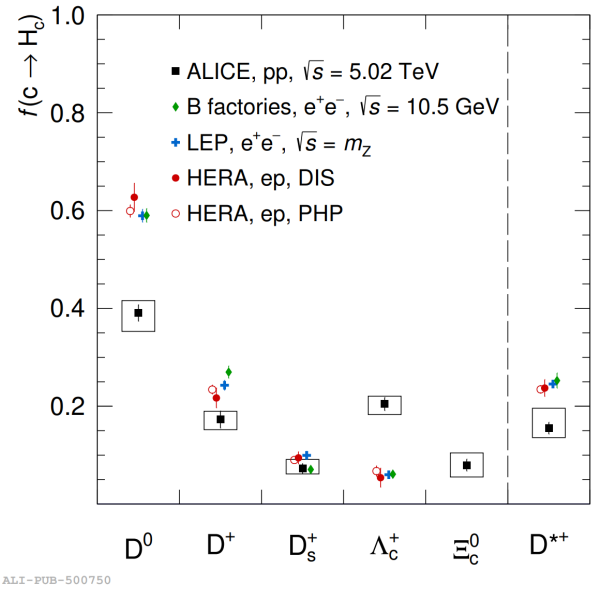


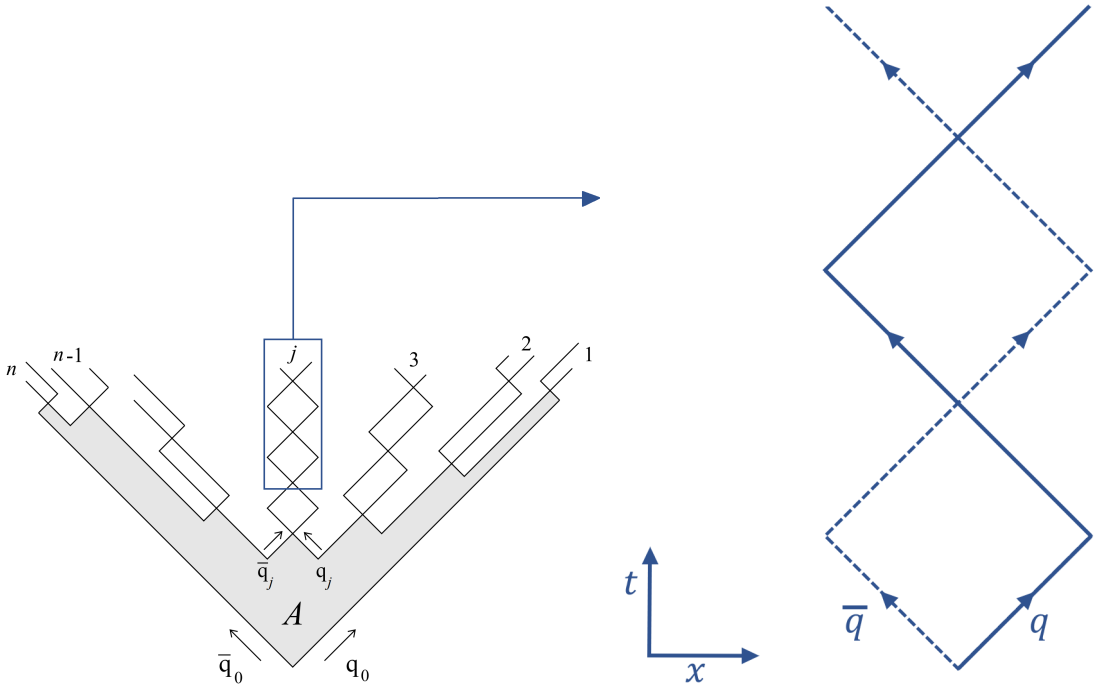
Figure 2.3: Comparison of the fragmentation fractions of charm quarks in charm hadrons obtained in pp collisions at $\sqrt{s} = 5.02$ TeV at the ALICE experiment and in other experiments in e^+e^- and ep collisions [41].

Hence, pQCD models seem to properly describe only mesons production, while baryon production results hint at the possibility of non-universal fragmentation functions or different mechanisms playing a role during hadronisation which specifically increase the baryons production in pp collisions compared to e^+e^-.
This is clear when studying the charm fragmentation fractions, whose results in pp by the ALICE experiment and by other collaborations in e^+e^-/ep collision systems are shown in Fig. 2.3.
In particular, in the latter collision systems, the measured fragmentation fractions were consistent within the uncertainties, supporting the theory of the universality of the fragmentation fractions.
In the plot, D^0 and Λ_c^+ are the ones with the largest discrepancy, and they differ, respec-

tively, by a factor ≈ 0.8 and ≈ 3.3 with a significance greater than 5σ, so more studies need to be conducted on the topic.
Various theoretical models were proposed trying to reproduce ALICE measurements and to describe the correct mechanisms behind hadronisation in hadronic collisions. They are reported in the following sections.

2.2 Hadronisation mechanisms and models

Studying hadronisation of heavy quarks in small systems (pp and p–Pb collisions) means mostly analysing parton fragmentation, which is supposed to be the main channel of hadron creation after the collision. On the other hand, in heavy-ion collisions heavy-flavour hadrons can also be created by recombination (coalescence) with other quarks or antiquarks in the medium, making them a potentially powerful probe to study the QGP. However, as discussed in the previous section (Fig. 2.2), latest experimental measurements showed an enhanced production of Λ_c^+ baryons over D^0 mesons in pp collisions at the LHC energies compared to e^+e^- and ep collisions [36] suggesting that coalescence processes or other mechanisms could also be present in smaller collision systems.



(a) (b)

Figure 2.4: Illustration of the fragmentation of Lund strings (a) [42] and zoomed detail of a single *yoyo* process (b) in the x-t plane (based on the scheme found in [43]).

The most used event generators differ in the way they model the hadronisation of par-

tons using phenomenological non-perturbative calculations. They generally employ only leading-order approximations and can provide different measurements as a function of the parameters considered while running them.
One of the most widely used hadronisation mechanisms is the Lund phenomenological string model [44], which deserves a brief introduction, since it is the basic hadronisation mechanism in many Monte Carlo generators.
The interaction between two partons in the Lund model occurs through coloured fields that are represented by strings. The interaction of a $q\bar{q}$ pair can be described by the Cornell potential [45]:

$$V(r) = \kappa r - \alpha r^{-1}, \tag{2.2}$$

which is a simple way to account mathematically for quarks confinement, and contains a string tension term κ and a Coulombian-like constant α. When the distance between the parton increases the second term of the Eq. 2.2 becomes negligible and the potential can be approximated with $V(r) \approx \kappa r$ with a typical string tension of 1 GeV/fm. This term is similar to the Hooke's law describing the elastic mechanical force exerted by a loaded string.
Increasing the distance between the original quark-antiquark pair causes the colour string to stretch due to the linear relation of the energy stored in the strings, and at a certain time the creation of a new $q_j\bar{q}_j$ pair will become more favourable than further extending the colour string. As illustrated in Fig. 2.4a the strings are then broken into two, and the newly created pair moves on a parallel direction with respect to the original light cone that generated it.
Both the original pair and the newly generated ones are accelerated and forced together by the linear potential when the distance increases in a motion called *yo-yo* (due to the visual similarity of the process to the famous toy).
An example of yo-yo motion is illustrated in Fig. 2.4b, in which the process is shown in space-time coordinates. Starting from a t = 0, each intersection of the quark-antiquark pair is characterised by the quarks inverting their momentum, in a *yo-yo* way [44].
When more pairs are produced at other points of the stretched colour strings, the quarks originating can either combine with other produced pairs to form hadrons or they can proceed with their yo-yo modes until they fragment and only hadrons remain.
The Lund model is able to describe the hadronisation following hard parton scatterings, and it is often employed in Monte Carlo generators, which provide complete particle simulations by including other mechanisms such as initial and final-state parton showers, hadron decays, multiple parton interactions, etc., to improve their predictions. Moreover, event generators often offer only leading-colour (LC) processes, reducing their accuracy, but some generators with NLO accuracy in the hard scattering, such as POWHEG [38], are already available.
In the next sections, a collection of generators used in the final results of this dissertation, and for other activities related to the PhD (see App. A.1), will be described.

2.2.1 PYTHIA8

As anticipated in the previous section, the Lund string model is not able to describe the production of some heavy-flavour hadrons in pp collisions at the LHC, especially in the baryon sector.
The basic fragmentation Lund model is fully implemented in the PYTHIA8 Monash 2013 tune which, as seen in Fig. 2.2, greatly underestimates the production of the Λ_c^+ baryon over D^0.
Other theoretical approaches are introduced to better describe the enhancement of the baryon-to-meson ratio; the one mainly used in this dissertation is the introduction of enhanced colour reconnection mechanisms [46]. This model was also developed taking into account the enhanced baryon production in pp collisions in the light sector: the baryon-to-meson ratio is tuned to the measurements of the Λ/K_s^0 ratio (obtained by the CMS Collaboration in 2011 [47]).
PYTHIA uses a leading-colour approximation to trace the colour flow for each event, and each partonic final state contains quarks that are connected uniquely to a single other parton in the event. On the other hand, gluons carry both a colour and an anticolour charge, and are hence connected to two partons by two colour strings. The leading-colour approximation is represented in Fig. 2.5a.
This description implemented in the Monash tune is quite effective in describing e^+e^- collisions, but the situation could change significantly in pp collisions where, due to the composite nature of hadrons, events with several distinct parton interactions can occur simultaneously during a collision: multiple parton interactions (MPI) [48]. All the partonic systems generated are to be considered in the description of the collision, and there is a non-negligible possibility of overlapping between their phase spaces.

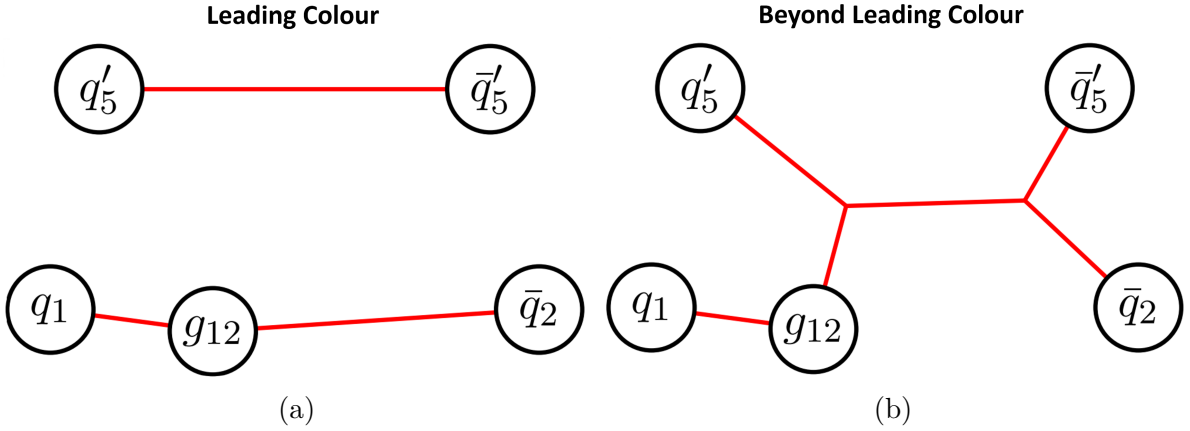


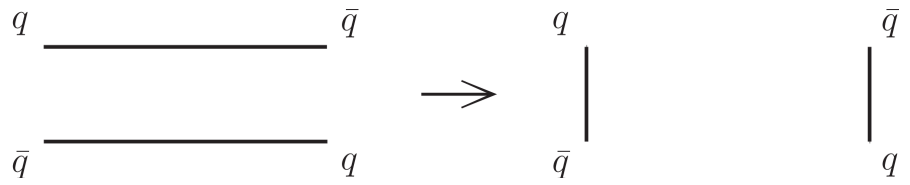
Figure 2.5: Quarks scheme of a leading-colour string topology (a) and of a junction beyond-leading-colour topology (b) [46] implemented in PYTHIA8.

Early studies on the possible colour connection between multiple systems were performed [49], but the effect was expected to be small even in high-multiplicity events. As in LEP results, the production as a function of p_T was expected to have a flat trend [37],

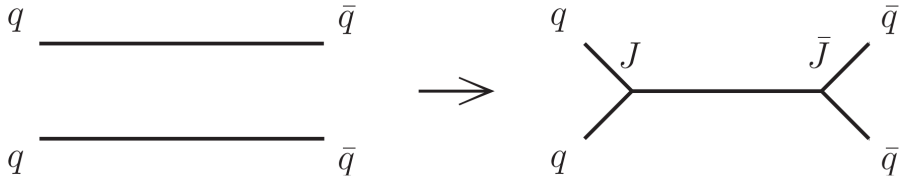
which is not the case in recent experimental measurements [36], meaning that non-trivial colour reconnection mechanisms could be introduced.
In the enhanced colour reconnection model in PYTHIA, the QCD framework is simplified by encoding the main parton-parton combination possibilities in a single colour index from 1 to 9 (and anti-index for anti-colours). Quarks have a single colour index as antiquarks have a single anticolour one, while gluons are characterised by one of each, allowing the model to have 9 different quark colours and 72 gluon states. These indices are not the ordinary SU(3) colours of the SM, but they rather indicate the possible colour states of the combinations of two-parton systems: a colour singlet is expressed if the colour index of a quark equals the anticolour index of its coupled antiquark, otherwise they are in an octet state (9 colour index values in total).

Confining potentials between quark-antiquark couples can now be created between any two partons that have opposite and matching indices, making the original leading colour string topology still possible, but increasing the possibilities of accidental recombination between quarks. Moreover, specific indices can also combine in pairs in order to appear in the system as the anticolour index missing in the triplets [1,4,7], [2,5,8] and [3,6,9], e.g. two partons with colours 1 and 4 can appear as one parton with anticolour index 7.
Recombination of quarks is allowed both among partons arising from MPIs and from beam remnants, increasing with these mechanisms the string topologies that will hadronise, as shown in Fig. 2.5b.
This model enriches the topologies available with the default leading colour string structures through the so-called junctions, which can be of different types and are classified into 4 groups related to their kind of reconnection topology (Fig. 2.6), which starts as an ordinary dipole-style reconnection with Type I and becomes more complicated with Type IV in which connected junction-antijuction pairs allow gluons to be added between junctions [46]. The other two kinds of junctions, Type II and Type III, implement the formation of a junction-antijunction pair that is connected in the former, and not directly connected in the latter.

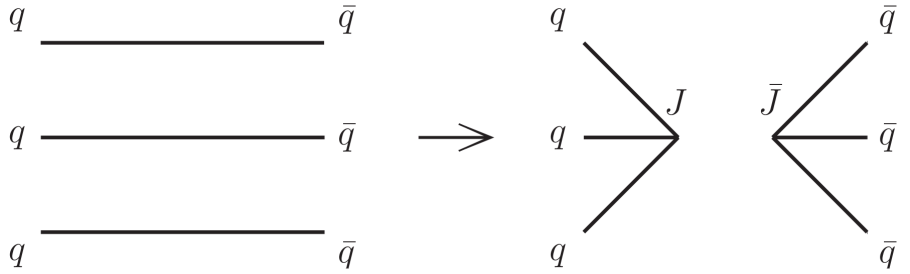
This model predicts the enhancement, in particular, of charmed baryons, and a collection of their relative yields using this method is shown in Tab. 2.1. The standard colour reconnection model production estimates are also shown in the table (old CR model). The comparison between the new full production and that without the CR mechanism (corresponding to the column labelled as "string") shows up to a factor ≈20 of increase on heavier charmed baryons (as Σ_c^0). To obtain these results ten million events were simulated in pp collisions at $\sqrt{s} = 7$ TeV and hadron decays were turned off to focus on the primary particles production.
The achievable topologies in an event are ruled by the minimisation of the potential energy of the string, and different modes of the model are possible thanks to the tuning of various parameters which are shown in Tab. 2.2. They take into account the requirement that, in order for reconnection to be feasible between two string strands, they must be able to resolve each other between the time of formation and hadronisation (including time-dilation effects caused by their relative boosts [50]).



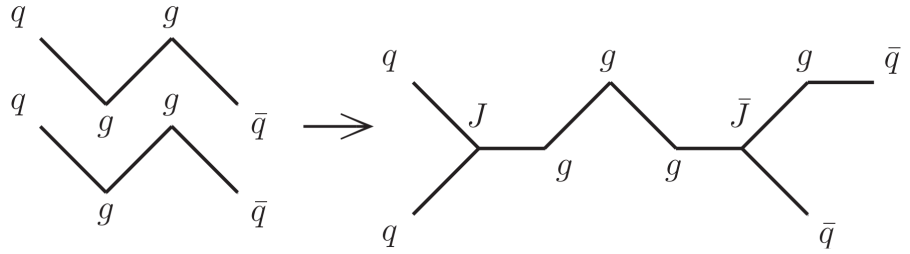
(a) Type I: dipole-style reconnection



(b) Type II: junction-style reconnection



(c) Type III: baryon-style junction reconnection



(d) Type IV: zipper-style junction reconnection

Figure 2.6: Allowed reconnection junctions in the enhanced colour reconnection model [46].

In particular in mode 0 there is no time dilation at all and the value of the constant m_0[3] controls the amount of colour reconnection; in mode 2 the time dilation is performed by using the boost factor obtained from the final-state mass of the dipoles, and it is required that all the dipoles involved in a reconnection are causally connected; mode 3, instead, is very similar to the previous one, but requires only a single connection to be causally connected.

[3]This is usually of the order of the Λ_{QCD} scale, with energy dimension.

Particle	New CR model			Old CR model
	string	junction	all	(all)
D^+	$5.3 \cdot 10^{-2}$	0	$5.3 \cdot 10^{-2}$	$6.5 \cdot 10^{-2}$
Λ_c^+	$4.0 \cdot 10^{-3}$	$7.9 \cdot 10^{-3}$	$1.2 \cdot 10^{-2}$	$6.6 \cdot 10^{-3}$
Σ_c^{++}	$2.7 \cdot 10^{-4}$	$1.3 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$	$5.4 \cdot 10^{-4}$
Σ_c^{+}	$2.5 \cdot 10^{-4}$	$1.5 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$5.2 \cdot 10^{-4}$
Σ_c^{0}	$2.5 \cdot 10^{-4}$	$1.3 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$	$5.1 \cdot 10^{-4}$
Σ_c^{*++}	$5.1 \cdot 10^{-4}$	$1.7 \cdot 10^{-3}$	$2.2 \cdot 10^{-3}$	$9.5 \cdot 10^{-4}$
Σ_c^{*+}	$4.9 \cdot 10^{-4}$	$1.9 \cdot 10^{-3}$	$2.4 \cdot 10^{-3}$	$9.4 \cdot 10^{-4}$
Σ_c^{*0}	$4.8 \cdot 10^{-4}$	$1.7 \cdot 10^{-3}$	$2.2 \cdot 10^{-3}$	$9.1 \cdot 10^{-4}$

Table 2.1: Relative primary particle production of charmed hadrons (N_{par} /N_{events}) in pp collisions at $\sqrt{s} = 7$ TeV over the full y and p_{T} range, summed over particles and antiparticles [46].

Parameter	Monash	Mode 0	Mode 2	Mode 3
StringPT:sigma	0.335	0.335	0.335	0.335
StringZ:aLund	0.68	0.36	0.36	0.36
StringZ:bLund	0.98	0.56	0.56	0.56
StringFlav:probQQtoQ	0.081	0.078	0.078	0.078
StringFlav:ProbStoUD	0.217	0.2	0.2	0.2
StringFlav:probQQ1toQQ0join	0.5, 0.7, 0.9, 1.0	0.0275, 0.0275, 0.0275, 0.0275	0.0275, 0.0275, 0.0275, 0.0275	0.0275, 0.0275, 0.0275, 0.0275
MultiPartonInteractions:pT0Ref	2.28	2.12	2.15	2.05
BeamRemnants:remnantMode	0	1	1	1
BeamRemnants:saturation	–	5	5	5
ColourReconnection:mode	0	1	1	1
ColourReconnection:allowDoubleJunRem	on	off	off	off
ColourReconnection:m0	–	2.9	0.3	0.3
ColourReconnection:allowJunctions	–	on	on	on
ColourReconnection:junctionCorrection	–	1.43	1.20	1.15
ColourReconnection:timeDilationMode	–	0	2	3
ColourReconnection:timeDilationPar	–	–	0.18	0.073

Table 2.2: List of all the parameters among the considered PYTHIA8 tunes [46].

2.2.2 Hadronisation via coalescence: the Catania model

The measurements obtained by the ALICE experiment in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV have shown large productions of the Λ_c^+ baryon [51], in agreement with previous results of heavy-ion collisions obtained at RHIC by the STAR collaboration in Au–Au

collisions at 200 GeV [52]. The Λ_c^+/D^0 ratio obtained in Pb–Pb collisions is similar to the results obtained at RHIC, but shows a small reduction as estimated by theoretical calculations, providing hadronisation via a combined coalescence plus fragmentation approach. This mechanism (see [53] for a review) is the basis of the Catania model, which was initially developed to be used only in nucleus-nucleus collisions. However, a thorough study has been performed on the production of charm hadrons in pp collisions assuming a coalescence plus fragmentation approach in a bulk matter following hydrodynamic simulations, included to study the spectra and collectivity phenomena in the system [54].

The Catania model assumes the presence of a thermalised system of gluons and light flavour quarks (u, d, s) and anti-quarks, already in pp, with a lifetime $\tau \approx 2.5$ fm/c, a radius of $R = 2$ fm and a temperature of the bulk matter of $T = 165$ MeV for pp collisions at $\sqrt{s} = 13$ TeV.

The implemented coalescence model is based on the Wigner formalism and predicts a momentum spectrum of the hadrons formed by quark coalescence as:

$$\begin{aligned}\frac{dN_H}{dyd^2P_{\mathrm{T}}} = g_H \int \prod_{i=1}^{N_q} \frac{d^3p_i}{(2\pi)^3 E_i} p_i \cdot d\sigma_i \; f_{q_i}(x_i, p_i) \\ \times f_H(x_1...x_{N_q}, p_1...p_{N_q})\, \delta^{(2)}\left(P_{\mathrm{T}} - \sum_{i=1}^{n} p_{\mathrm{T},i}\right),\end{aligned} \tag{2.3}$$

where g_H is the statistical factor to form a colourless hadron H from quarks and antiquarks with spin 1/2, corresponding to 1/36 and 1/108, respectively, for mesons and baryons; $d\sigma_i$ denotes an element of space-like surface; f_{q_i} are the quark (anti-quark) phase-space distribution functions for the i-th quark (anti-quark), while $f_H(x_1...x_{N_q}, p_1...p_{N_q})$ is the Wigner function describing the spatial and momentum distribution of quarks in a hadron, which is assumed to follow a Gaussian shape; finally, N_q is the number of quarks that form a hadron (2 for mesons and 3 for baryons).

The Wigner distribution[4] contains a normalisation constant that is fixed to guarantee that in the momentum limit $p \to 0$ all charm quarks hadronise by coalescence in a heavy hadron, requiring that the total coalescence probability P_{coal}^{tot} satisfies $\lim_{p\to 0} P_{coal}^{tot} = 1$.

On the other hand, the momentum spectrum of a hadron produced by charm parton fragmentation can be defined as:

$$\frac{dN_{had}}{d^2p_{\mathrm{T}}\, dy} = \sum \int dz \frac{dN_{fragm}}{d^2p_{\mathrm{T}}\, dy} \frac{D_{had/c}(z, Q^2)}{z^2}, \tag{2.4}$$

in which $D_{had/c}(z, Q^2)$ is the fragmentation function dependent on the momentum fraction z of the heavy quarks transferred to the final heavy hadron and on $Q^2 = (p_{had}/2z)^2$

[4]It was originally introduced in non-relativistic quantum mechanics and the expectation value of any physical operator can be obtained by averaging over a Wigner distribution. De facto, it is a quasi-distribution (due to the Heisenberg uncertainty principle in quantum theories) that describes the non-positive definite density of finding a particle in average values of position and momentum. This function has been applied in various areas, including heavy-ion collisions and hadron physics [55].

which is the momentum scale for the fragmentation process. The model assumes the commonly used Peterson fragmentation function for heavy quarks [56].

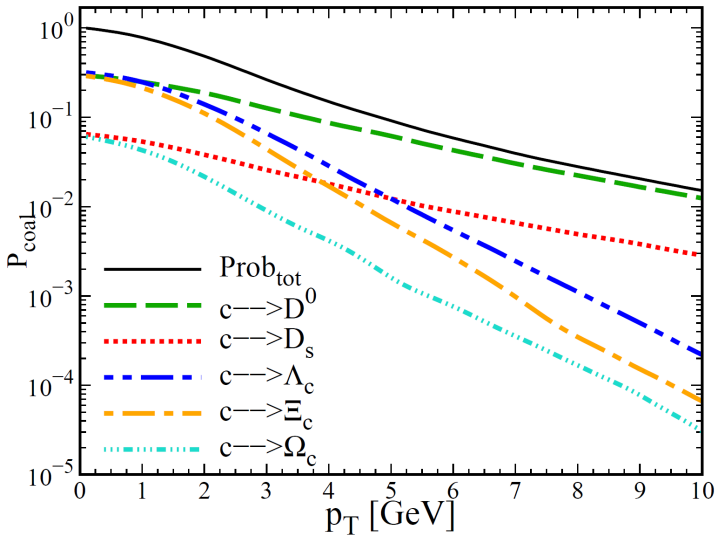


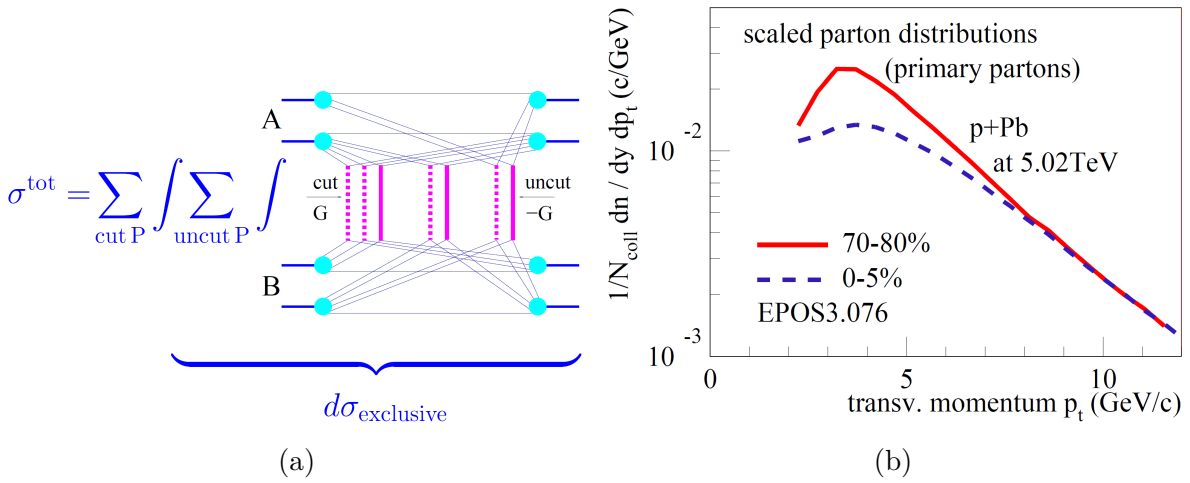
Figure 2.7: Probability of charm quark coalescence as a function of p_{T} of the charm quark for pp collisions at LHC [57]. The solid black line is the total coalescence probability of all the species.

The probability of coalescence has a decreasing trend as a function of p_{T}, as shown in Fig. 2.7, which means that at low momentum charm has a higher probability of hadronising through coalescence with light partons from the thermalised medium. As required in the Wigner distribution assumed in the model, at $p_{\mathrm{T}} = 0$ all charm will hadronise only by coalescence, while at higher momentum heavy quarks that cannot coalesce will hadronise via fragmentation with a probability given by $P_{frag} = 1 - P_{coal}$. This implies that, at high p_{T}, fragmentation becomes the main mechanism of hadronisation in the Catania model and a charm quark will hadronise following specific fragmentation fractions defined in [58].
The coalescence probabilities in Fig. 2.7, for different charmed hadrons, show that the Λ_{c}^{+} and Ξ_{c} trends are compatible with the D^0 one for $p_{\mathrm{T}} < 1$ GeV/c, which is a feature of the coalescence mechanism and causes an enhancement of the $\Lambda_{\mathrm{c}}^{+}/\mathrm{D}^0$ and $\Xi_{\mathrm{c}}/\mathrm{D}^0$ ratio.

2.2.3 EPOS3

The EPOS model is a sophisticated multiple scattering approach based on partons and parton ladders, which are referred in the model as pomerons and define individual scattering. The acronym stands for **E**nergy conserving quantum mechanical multiple scattering approach, based on **P**artons (parton ladders), **O**ff-shell remnants, and **S**plitting of parton ladders, mentioning, in its entirety, the basic mechanisms on which the model is based. The generator provides a hydrodynamical evolution in pp, p–A and heavy-ion collisions

through parton ladders which show up as flux tubes (also called strings in the model). In EPOS3 the nonlinear effects in the parton evolution are treated by considering per pomeron individual saturation scales and a 3+1 dimension viscous hydrodynamical evolution of the system.
The initial hydrodynamic conditions are provided by the Gribov-Regge multiple scattering framework, while the parton ladders are composed of a pQCD hard process combined with initial and final state linear parton emissions [59]. These conditions are applicable to many collision systems, including pp, p–A, A–A and lepton-proton collisions.
The complex theoretical calculations will not be discussed, but the main features of the model are introduced.
The formalism is capable of computing exclusive cross sections for subprocesses using partial summations of cut and uncut pomerons (defined in [59]), as illustrated in Fig. 2.8, such as the cross section for triple scattering (exchange of three pomerons) in pp collisions.



(a) (b)

Figure 2.8: Illustration of EPOS3 total cross section in terms of cut (uncut) pomerons using dashed (solid) lines for pp, p–A and A–A collisions (a) and parton distribution divided by the number of binary collisions as a function of p_{T} in p–Pb collisions in two different centrality intervals at $\sqrt{s_{\text{NN}}} = 5.02$ TeV (b) [60].

The single pomeron exchange amplitude, which is contained in the cross section calculation, is defined as the sum of different contributions [59, 60] that are related to the interaction between *sea* or *valence* quarks, in the combinations sea-sea, val-val, sea-val and val-sea, respectively, on the projectile and target side, starting the parton ladder. The saturation scale was introduced in EPOS3 to account for nonlinear effects, replacing the previous description in EPOS2, and it is implemented through a value Q_s representing the virtuality scale below which non-linear effects (like gluon-gluon fusion) are important. This variable is adapted to the EPOS3 formalism, implemented in the hard scattering

cross section calculation, and it is assumed to be

$$Q_s^2 = B_{sat} \frac{N_{part}}{(1/\hat{s})^{\lambda}}, \tag{2.5}$$

taking into account the centre-of-mass energy $\hat{s}$ of the pomeron, the number of participants N_{part}, and the proportionality constant B_{sat} that is chosen to ensure binary scaling of the parton multiplicity and number of collisions in p-A and A-A at high p_{T}. λ is assumed to be constant and equal to 0.25. The number of participants is estimated via a theoretical approach, which defines it as the maximum value of the calculated prediction of the number of target and projectile participants.
The saturation scale affects the value of parton multiplicity at low and high p_{T} values, as shown in Fig. 2.8b for p–Pb collisions, in which the primary parton distribution results, scaled by the number of binary collisions, are expressed as a function of p_{T} in two different centrality intervals. The peripheral and central collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV show different trends that become compatible only at high p_{T}. In particular, the low p_{T} region in central collisions is significantly suppressed compared to the peripheral case.

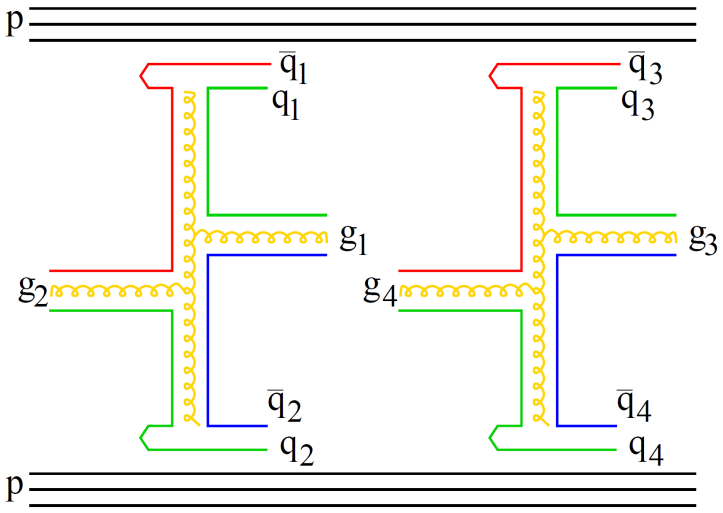


Figure 2.9: Simplified scattering, with double-pomeron exhange, illustrating the colour flow [60]

For the generation of events in EPOS, real physical events are used, without introducing "test particles" as performed in other generators. The generation is performed in two steps: the first one generates the multiple scattering configuration accordingly to the total cross section described earlier, via the number of cut pomerons per possible nucleon–nucleon pair and the momentum fractions of the light cone pomeron ends; while the second step creates, for a given configuration, the partons associated with each pomeron. These last partons make the parton ladders of the models which are identified with flux tubes.
A simple example of scattering is shown in Fig. 2.9 using the colour flow of two cut

pomerons in *sea-sea* with a simple $gg \rightarrow gg$ process. The interaction concerns the quarks and antiquarks q_i of the first ($i = 1, 2$) and second pomerons ($i = 3, 4$). Following the colour flows, one can identify the kinky strings, e.g. q_1 - g_1 - $\bar{q_2}$ and q_2 - g_2 - $\bar{q_1}$ for the first pomeron. So in Fig. 2.9 four kinky strings are found, which are essentially one-dimensional colour flux tubes, with a possible very small finite transverse additional dimension. High-p_{T} partons will show as transversely moving string pieces, but they are rare in the TeV energy range, and most kinks have a p_{T} of only few GeV/c.
Quark-antiquarks pairs are generated when the strings break due to elementary reactions and the originating segments are identified as hadrons and resonances. This mechanism is also used to generate jets through the segmentation of high-p_{T} hard partons.
However, in heavy-ion collisions and in high-multiplicity events in pp and p–Pb scatterings at high energy, this mechanism is not sufficient to describe particle production, since the density of strings is so high that they cannot decay independently only by segmentation. Thus, even in high-multiplicity pp collisions, the flux tubes will constitute thermalising bulk matter that expands collectively, called *core*. The strings segments that are close to the surface of the bulk matter or with a large p_{T} will leave the core and hadronise, defining the region of the *corona*.
The model allows strings segments to constitute bulk matter or escape according to their p_{T} and the local string density ρ. Particles from the corona will then appear as hadrons, while the core ones will provide the initial condition of a hydrodynamical evolution and they will later hadronise at a hadronisation temperature T_{H} following the freeze-out of the medium. After this process, a hadronic cascade procedure will allow the hadrons to continue interacting with each other [61]. The treatment of high-multiplicity events in the core-corona description improves the agreement between data and simulations.

2.2.4 Statistical Hadronisation Model with enhanced set of charm-baryon states

A revisited version of the Statistical Hadronisation Model (SHM [62]) was developed to explain the results in the charm sector regarding the enhancement of heavy-flavour baryon production [63] by including an enhanced set of charm-baryon states predicted by the Relativistic Quark Model (RQM [64]).
The SHM replaces the complexity of the hadronisation process by using thermostatistical weights that are ruled by the masses of the available hadron states at a hadronisation temperature T_{H}. The model has been successful in describing the production of light and strange hadrons in both heavy-ion collisions and smaller systems, by adding in the latter case and in peripheral heavy-ion collisions a strangeness suppression factor $\gamma_S < 1$.
But the model was unable to describe the $\Lambda_{\mathrm{c}}^{+}/\mathrm{D}^{0}$ ratio results obtained by the ALICE experiment [65] in pp collisions at $\sqrt{s} = 7$ TeV. In fact, the value predicted by SHM for the ratio (≈ 0.22 [66]) was based on charm-hadron states from PDG [4], but the measurements showed a ratio that was more than double (≈ 0.54). Hence, it seems that the use of PDG states only underpredicts the results, while a good description can be

achieved by employing the RQM predictions.
The RQM predicts additional charm-baryon states which amount to 18 Λ_c^+, 42 $\Sigma_c^{0,+,++}$, 62 $\Xi_c^{0,+}$ and 34 Ω_c excited states up to a mass of 3.5 GeV/c^2. Thermal hadron densities n_i, as described by SHM, depend strictly on the mass of the hadrons m_i and their spin degeneracies $d_i = 2J + 1$ estimated at T_H:

$$n_i = \frac{d_i}{2\pi^2} m_i^2 T_{\mathrm{H}} K_2 \left(\frac{m_i}{T_H}\right), \tag{2.6}$$

which uses the second-order Bessel function K_2 and an upper estimate of the hadronisation temperature $T_{\mathrm{H}} = 170$ MeV [67]. The thermal densities of the ground-state charm hadrons calculated using this method are listed in Tab. 2.3.

n_i $(\cdot 10^{-4}\ fm^{-3})$	D^0	D^+	D^{*+}	D_s^+	Λ_c^+	$\Xi_c^{0,+}$	Ω_c
PDG (170 MeV)	1.161	0.5098	0.5010	0.3165	0.3310	0.0874	0.0064
RQM (170 MeV)	1.161	0.5098	0.5010	0.3165	0.6613	0.1173	0.0144

Table 2.3: Thermal densities of prompt ground-state charmed hadrons for $T_{\mathrm{H}} = 170$ MeV [63].

The additional excited states of the particles mentioned in the table are forced to decay into one of the ground states governed by the branching ratios (BR) from the PDG, while those without listed BRs are considered equally weighted. For example, this means that for the excited Ξ_c state the PDG declares a decay channel of Λ_c^+ + K without specifying the branching ratio, so in the model it is assumed that 50% of excited Ξ_c decay to Λ_c^+ and the other half goes to the ground state of Ξ_c. This leads to a specific enhancement of the ground states of charmed baryons without increasing the production of D mesons.

2.2.5 Quark Combination Mechanism framework

The Quark (re-)Combination Mechanism (QCM) framework [68] implements as hadronisation mechanism the recombination of a charm quark c with a co-moving light antiquark $\bar{l}$ or two co-moving light quarks ll to form a single hadron H (respectively, a meson or a baryon), which will assume as momentum $p_{\mathrm{H}} = p_c + p_{\bar{l},ll}$. This mechanism influences the momentum spectra of charm hadrons and, in particular, the baryon-to-meson ratio. QCM is similar to the coalescence mechanism described above, but it does not employ the Wigner distributions.
The momentum distributions of the single charmed meson $M_{c\bar{l}}$ and baryon $B_{cll'}$ in the model can be expressed as

$$f_{M_{c\bar{l}}}(p) = \int dp_1 dp_2 f_{c\bar{l}}(p_1, p_2)\, \mathcal{R}_{M_{c\bar{l}}}(p_1, p_2; p), \tag{2.7}$$

$$f_{B_{cll'}}(p) = \int dp_1 dp_2 dp_3 f_{cll'}(p_1, p_2, p_3)\, \mathcal{R}_{B_{cll'}}(p_1, p_2, p_3; p)., \tag{2.8}$$

where $f_{c\bar{l}}(p_1, p_2)$ is the joint momentum distribution for the charm quark and the light antiquark ($\bar{l}$), $\mathcal{R}_{M_{c\bar{l}}}(p_1, p_2; p)$ is the probability density of $c\bar{l}$ with momenta p_1, p_2 to combine into a meson $M_{c\bar{l}}$ with momentum p (combination function). Similar variables are used for the charmed baryon production in Eq. 2.8.
The distributions of quarks of different flavours are considered independent, neglecting the correlations, and the combination functions are assumed to be the product of Dirac delta functions

$$\mathcal{R}_{M_{c\bar{l}}}(p_1, p_2; p) = \kappa_{M_{c\bar{l}}} \prod_{i=1}^{2} \delta(p_i - x_i p), \tag{2.9}$$

$$\mathcal{R}_{B_{cll'}}(p_1, p_2, p_3; p) = \kappa_{B_{cll'}} \prod_{i=1}^{3} \delta(p_i - x_i p). \tag{2.10}$$

The values $\kappa_{M_{c\bar{l}}}$ and $\kappa_{B_{cll'}}$ in the combination functions are momentum independent constants that depend on other elements, such as the number of quarks.
The momentum fraction of the i-th quark (x_i) is defined by assuming the co-moving approximation, which implies that recombination can take place only if the charm quark is moving at the same velocity of the other light antiquark (or two light quarks) to form the charm hadron. Under this assumption, the fraction can be expressed as

$$x_i = m_i / \sum_j m_j, \tag{2.11}$$

considering in the quark model the constituent fixed masses $m_u = m_d = 0.33$ GeV/c^2, $m_s = 0.5$ GeV/c^2, and $m_c = 1.5$ GeV/c^2.
Therefore, the distribution functions of the charm hadrons can be rewritten

$$f_{M_{c\bar{l}}}(p) = N_{M_{c\bar{l}}} f^{(n)}_{M_{c\bar{l}}}(p), \tag{2.12}$$

$$f_{B_{cll'}}(p) = N_{B_{cll'}} f^{(n)}_{B_{cll'}}(p), \tag{2.13}$$

by substituting all the terms defined above and introducing the normalised distribution function $f^{(n)}_{M_{c\bar{l}}}(p)$ and $f^{(n)}_{B_{cll'}}(p)$, whose integration on momentum return unity, and the momentum integrated yields $N_{M_{c\bar{l}(ll')}}$ which can be calculated as

$$N_{M_{c\bar{l}}} = N_c N_{\bar{l}} \frac{\kappa_{M_{c\bar{l}}}}{A_{M_{c\bar{l}}}} = N_c N_{\bar{l}} \mathcal{R}_{c\bar{l}\to M_{c\bar{l}}}, \tag{2.14}$$

$$N_{B_{cll'}} = N_c N_l N_{l'} \frac{\kappa_{B_{cll'}}}{A_{B_{cll'}}} = N_c N_l N_{l'} \mathcal{R}_{cll'\to B_{cll'}}, \tag{2.15}$$

where $A^{-1}_{M_{c\bar{l}}} = \int dp\, f^{(n)}_c(x_1 p)\, f^{(n)}_{\bar{l}}(x_2 p)$ and $A^{-1}_{B_{cll'}} = \int \mathrm{d}p \prod_{i=1}^{3} f^{(n)}_{q_i}(x_i p)$ with the normalised charm and light quark distribution $\int dp\, f^{(n)}_{c,l}(p) = 1$. The quantity $\mathcal{R}_{c\bar{l}\to M_{c\bar{l}}} \equiv \kappa_{M_{c\bar{l}}}/A_{M_{c\bar{l}}}$ is the momentum-integrated combination probability of a charm quark and a

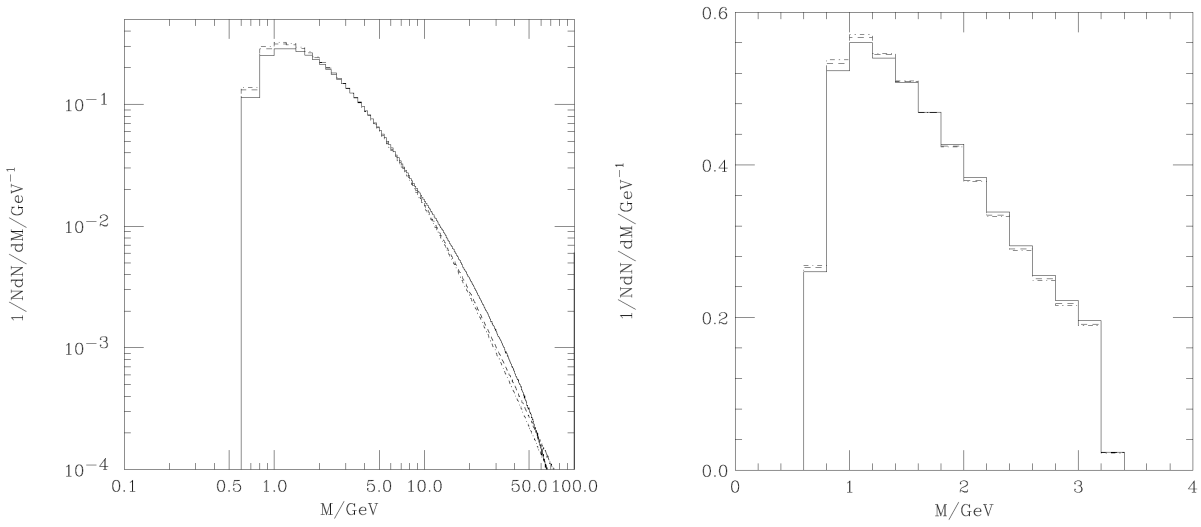
light antiquark to produce a meson, while $\mathcal{R}_{cll' \to B_{cll'}}$ is similar, but related to the production of a baryon.
Finally, the integrated momentum yield of a charm meson $M_{i,c\bar{l}}$ can be expressed as

$$N_{M_{i,c\bar{l}}} = C_{M_{i,c\bar{l}}} P_{\bar{l}} N_{M_c}, \tag{2.16}$$

by introducing the probability $P_{\bar{l}} = N_{\bar{l}}/N_{\bar{q}}$ of an antiquark $\bar{q}$ to be of flavour $\bar{l}$, and the probability $C_{M_{i,c\bar{l}}}$ of creating a particular spin state i, given that a specific $c\bar{l}$ combination can form different J^P states. A similar formula to Eq. 2.16 is achievable for the charm baryon yield [68].
These formulas are applied to the one-dimensional p_{T} space, and the p_{T} distributions of the quarks at hadronisation are used as inputs to the model. In particular, the light-quark spectra were obtained from the light-flavour hadron production in pp collisions at $\sqrt{s} = 7$ TeV [69], while the charm quark spectrum in QCM is extracted from the data of D^{*+} results in pp and p–Pb collisions, respectively at $\sqrt{s} = 7$ and $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV [70].
Via the inputs, the model is able to provide estimates for the production of many single charm hadrons, including all the mesons and baryons that contribute to the results presented in this dissertation.

2.2.6 HERWIG7

Hadron Emission Reactions With Interfering Gluons (HERWIG) is a general purpose event generator capable of providing simulations at next-to-leading order in QCD keeping its capabilities of coherent parton showers, Beyond Standard Model (BSM) events description, cluster hadronisation model, etc. [39].
The generator can assemble NLO QCD corrections for virtually all SM processes, by calculating the amplitudes with external libraries that are integrated in the software, and uses a complex mechanism to develop parton showering via the coherent branching algorithm [71]. This mechanism is based on angular ordering (used in previous versions of the generator) [72], but introduces a series of improvements, such as the covariant formulation of the showering algorithm (invariant under boosts along the jet axis) and the introduction of heavy-quark fragmentation via mass-dependent splitting functions and kinematics [73].
HERWIG7 hadronisation employs a cluster model which is executed after the parton shower in order for quarks and gluons to combine in the observed hadrons. The basis of the mechanism is the colour preconfinement property of the angular-ordered parton shower.
First, the remaining gluons after the parton shower are split non-perturbatively into quark-antiquark pairs (with any flavour). The isotropic decays create, in the event, only colour connected parton pairs, which are colour singlets defined in the model with clusters, which have a momentum given by the sum of their constituents momenta. These clusters have a mass distribution that is independent of the hard scattering process and its centre-of-mass energy, due to the principle of colour-preconfinement.



(a) (b)

Figure 2.10: Mass spectrum of the primary clusters (a) and of the clusters after fission (b). The plot shows only clusters containing light quarks. Solid, dashed, and dot-dashed lines show the clusters produced in hadronisation of $e^+e^- \to d\bar{d}$ events at a centre-of-mass energy of 100 GeV, 1 TeV, and 10 TeV, respectively [73].

Due to this property and to the fact that the cluster mass distribution is peaked at low values (Fig. 2.10a), the clusters can be considered as highly excited hadron resonances that decay into observed hadrons according to the phase space.
However, a small portion of the generated clusters will be split into lighter ones before they decay, due to their heavy mass. The mass limit for the cluster fission to occur is defined by

$$M^{\mathrm{Cl}_{pow}} \geq \mathrm{Cl}_{max}{}^{\mathrm{Cl}_{pow}} + (m_1 + m_2)^{\mathrm{Cl}_{pow}}, \tag{2.17}$$

where M is the mass of the cluster, Cl_{max} and Cl_{pow} are parameters of the model (flavour dependent), and $m_{1,2}$ are the masses of the constituent partons of the cluster.
During the fission, a pair $q\bar{q}$ is selected from the vacuum and the cluster is then broken down into two new clusters, with one of the original partons in each one, with mass distributions

$$M_{1,2,} = m_{1,2} + (M - m_{1,2} - m_q)\mathcal{R}_{1,2}^{1/P}, \tag{2.18}$$

which contain the mass m_q of the parton taken from the vacuum. $M_{1,2}$ indicate the masses of the new clusters obtained after the fission (Fig. 2.10b), that are required to be smaller than the mass of the original cluster, but greater than the sum of the mass of their constituent partons.
If a beam remnant is contained in the original cluster, a soft distribution is used for the masses of the daughter clusters produced in the process. A series of parameters are employed to control (and tune) the mechanism during the simulation.
Finally, clusters are decayed into a pair of hadrons by combining their quarks with a

pair of quark-antiquark or diquark-antidiquark extracted from the vacuum. The species are selected from all the possible alternatives with the appropriate flavour based on the available phase space, spin, and flavour of the hadrons.
As done in other Monte Carlo models, HERWIG uses weights for hadron production through cluster decays, but its peculiar feature is to use a single weight for each combination of decay products to account for the suppression of same flavour hadrons production in multiple weights approaches [74] (employed for example in HERWIG5.9).
Moreover, since many more mesons than baryons are included in the model, in order to increase baryon production, if the mass of a cluster is large enough to decay into the lightest baryon-antibaryon pair, then a parameter is used to select whether the cluster will perform a mesonic or baryonic decay [73].
Additionally, three colour reconnection models are implemented (and tunable) in HERWIG: the default plain reconnection model, a more complicated statistical model, and the baryonic reconnection model which allows the formation of baryonic clusters. The reconnection happens before cluster fission takes place.
The plain reconnection model (PCR) is composed of a series of steps which are followed exactly once for every quark created in random order in the event. A specific quark will be part of a cluster A and it can reconnect with all other clusters existing at that time, so its potential reconnection partner will be called B. The newly possible generated clusters C and D, after reconnection, are limited by conditions which make them lighter compared to the parent clusters and they cannot be colour octets. If at least one possible reconnection is found, then the one with the lowest cluster mass sum is selected and the reconnection is accepted following a tunable reconnection probability p_{reco}. After this last step, the reconnection cycle continues for another quark [75].
The statistical reconnection model (SCR) aims at finding a cluster configuration with a preferably small colour length obtained as

$$\lambda = \sum_{i=1}^{N_{cl}} m_i^2, \tag{2.19}$$

considering the number of clusters N_{cl} and the invariant mass m_i of the i-th cluster. However, finding the global minimum of λ is a very difficult task, given that the possible cluster configurations scale as the factorial of the parton pairs (100 parton pairs $\rightarrow 100! \approx 10^{158}$ cluster to be tested), so the Simulated Annealing algorithm is used to optimise the process.
The SCR model selects random pairs of clusters and proposes them for the colour reconnection, with the same conditions for daughter clusters as PCR. A reconnection step that reduces the colour length is always accepted, while if the value is raised, then the reconnection is accepted with the probability

$$p = \exp\left\{-\frac{\lambda_f - \lambda_i}{T}\right\}, \tag{2.20}$$

where T is the temperature and $\lambda_{i,f}$ are the initial and final colour lengths, respectively. T is a parameter that is gradually decreased during the reconnection procedure, reducing

the probability of accepting colour length and increasing reconnection steps. The transition of temperature is dictated by the annealing schedule which depends on the number of clusters. In HERWIG7 this model introduced a rule that forbids a reconnection step that would lead to a gluon produced in any stage of the parton shower evolution becoming a colour singlet after hadronisation [76].
Both SCR and PCR deal only with mesonic clusters, so an extension to the colour reconnection mechanism in the Monte Carlo generator was needed to produce baryonic clusters via reconnection. The baryonic reconnection model (BCR) is an algorithm that provides these results following a series of steps.
Due to the large invariant mass of baryon clusters, the baryonic reconnection does not employ the reduction of the cluster masses, but it uses instead a geometric picture of nearest neighbours in which the search of quarks is performed approximately in the population of the same phase space region based on the rapidity y. The model defines the formation of two baryonic clusters out of three mesonic ones.

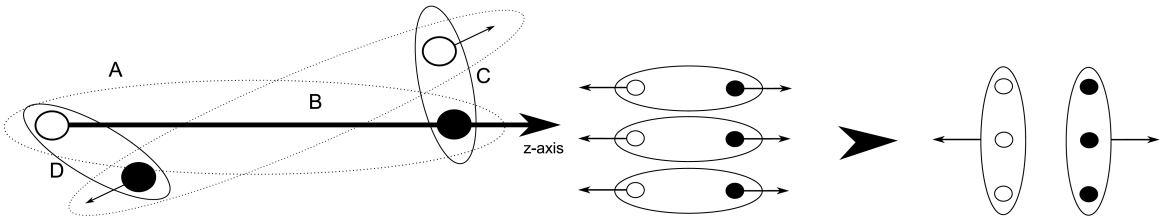


Figure 2.11: Illustration of mesonic (a) and baryonic (b) clusters colour reconnection in HERWIG7, using the BCR model [77].

After the creation of a list of colour-connected quarks and anti-quark clusters, following the parton shower and the gluons splitting, the order in which the clusters appear is shuffled to remove a possible bias for the reconnection. A cluster A is selected now from the shuffled list, and the rest frame of that cluster is considered, so that the two constituents (namely $q_A, \bar{q}_A$) are moving back-to-back, with the antiquark direction defining the positive z-direction of the quark axis. A loop is then performed on all the other clusters (B), calculating the rapidity of their constituents with respect to the quark axis in the rest frame of the cluster A. Therefore, the constituents of B ($q_B, \bar{q}_B$) will be listed in two categories: mesonic if $y(q_B) > 0 > y(\bar{q}_B)$ and baryonic if $y(\bar{q}_B) > 0 > y(q_B)$. In any other case, the cluster will not be considered for reconnection.
The category and the sum $|y(q_B)| + |y(\bar{q}_B)|$ for the clusters with the two largest sums are saved. They will be assumed as B and C clusters, and they will proceed to the reconnection step with cluster A depending on the category they fall into: if mesonic the normal reconnection with a probability p_R will be performed, otherwise the baryonic reconnection will occur and the considered clusters will be removed from the list and will not be considered for further reconnection. Fig. 2.11 shows the rapidity-based reconnection for both mesonic (Fig. 2.11a) and baryonic (Fig. 2.11b) cases. In particular,

the latter sketch shows how in the model the reconnection can happen only if all quarks (antiquarks) move in the same direction. Subsequently, the procedure is repeated for the next cluster in the list [77].

2.3 Multiple Parton Interactions in event generators

In the description provided by the factorisation approach (see Eq. 2.1) it is implicit that in the collision between two hadrons only two partons, each belonging to one of the hadrons, are actually taking part in the collision (SPS: Single Parton Scattering). However, given the nature of hadrons as composite objects (valence quarks and gluons, and quarks belonging to the sea), there is also a non-negligible probability that other partons take part in the collision. These interactions, generally at a lower scale with respect to the hard scattering described by the factorisation approach, are called Multiple Parton Interactions and their role is key in describing successfully the so-called soft underlying events. It is difficult to treat the partons (and in turn the final states) produced by the MPI with non-perturbative QCD calculations. Having said that, at fixed final state invariant masses, the cross sections for MPI will tend to increase with collision energy. This is because, generally speaking, partons with successively lower momentum fraction x are being probed. Therefore, an increase in MPI was expected moving to LHC energies (multi-TeV scale).
At these energies, some of the "soft" multiple parton interactions can reach a sufficient energy scale to produce heavy quarks, in particular charm (few GeV are enough). Traditionally, a second "hard" scattering is described by the so-called Double Parton Scattering (DPS), where pQCD techniques can be applied. The description of MPI (and the possibility of heavy quark production) through them is therefore relevant for heavy-quark production studies at the LHC.
This phenomenon should not be confused with pile-up, which is related, instead, to several simultaneous hadron-hadron collisions recorded at the same time in the detectors.
In the following paragraphs, a description of how commonly used event generators implement MPI will be provided.

2.3.1 PYTHIA8

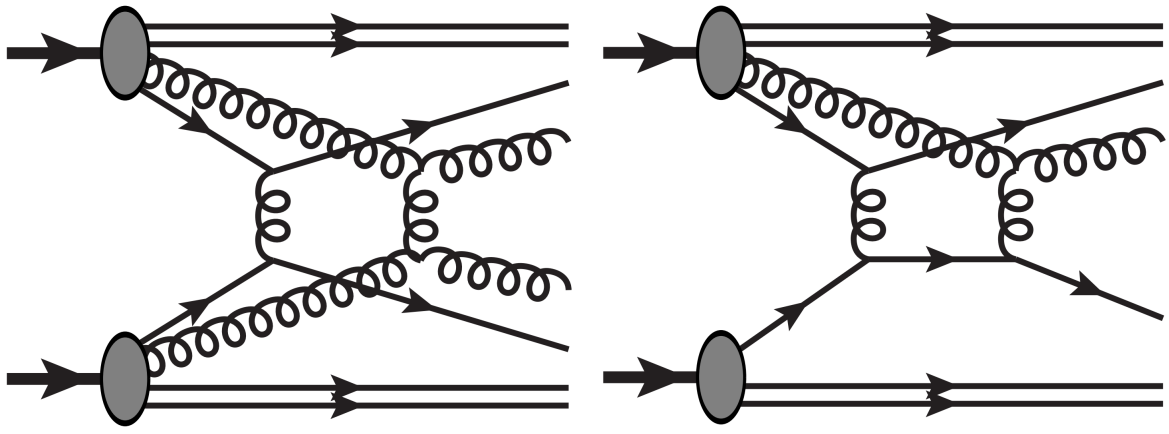
The PYTHIA8 approach on handling MPIs is to merge the DPS mode with the Gribov-Regge theory, which allows events with multiple cut pomerons, meaning that there are several strings crossing from one rapidity end of the event to the other [78], extending both of them. In particular, semi-perturbative MPIs create several minijets that contribute to the p_{T} flow and multiple colour connections between beam remnants, leading to higher multiplicity events.
The first step is to consider a $2 \to 2$ perturbative QCD process (with cross section σ) in a hadron-hadron collision[5], shown in Fig. 2.12a, in which the probability of having

[5]Hadrons are considered as *bunches* of incoming partons.

a MPI at a given p_{T} in a non-diffractive event is $\mathcal{P} = (1/\sigma_{\mathrm{nd}})\mathrm{d}\sigma/\mathrm{d}p_{\mathrm{T}}$, where σ_{nd} is the non-diffractive cross section [40]. If the interactions occur independently, then the number of MPIs is distributed in a Poissonian manner with a suppression of the zero.
An inspired parton shower paradigm approach is used by the model, that generates consecutive MPIs as a downward progression in p_{T}, resulting in a sequence of n interactions with $\sqrt{s}/2 > p_{\mathrm{T}_1} > p_{\mathrm{T}_2} > \cdots > p_{\mathrm{T}_n} > 0$. The evolution in p_{T} is handled by using the veto algorithm, as done in parton shower, and if no MPIs are created, a sequence is rejected, and the process will restart.
The calculation leads to a probability distribution for the p_{T} of the i-th parton equal to

$$\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}p_{\mathrm{T}_i}} = \frac{\widetilde{\mathcal{O}}(b)}{\langle\widetilde{\mathcal{O}}\rangle}\frac{1}{\sigma_{\mathrm{nd}}}\frac{\mathrm{d}\sigma}{\mathrm{d}p_{\mathrm{T}_i}}\exp\left(-\frac{\widetilde{\mathcal{O}}(b)}{\langle\widetilde{\mathcal{O}}\rangle}\int_{p_{\mathrm{T}_i}}^{p_{\mathrm{T}_{i-1}}}\frac{1}{\sigma_{\mathrm{nd}}}\frac{\mathrm{d}\sigma}{\mathrm{d}p'_{\mathrm{T}}}\,\mathrm{d}p'_{\mathrm{T}}\right)\,, \tag{2.21}$$

after selecting the impact parameter b with a similar doubly differential expression using $i = 1$. In the formula, $\widetilde{\mathcal{O}}(b)$ is the overlap between the distributions of two partons in the transverse plane, averaged in $\langle\widetilde{\mathcal{O}}\rangle$.
Even though the probability distribution in Eq. 2.21 provides an interpolation between hard and soft processes, only those in which the hardest interactions are selected by the MPI mechanism (non-diffractive ones) employ full correlation. In PYTHIA, if standalone hard processes are considered, the selection on the processes kinematics is performed without taking into account the MPIs, but, in other cases, the selection of the impact parameter, depending on the p_{T} scale, is done only after the MPI machinery is invoked, also considering Sudakov-like form factors[6]. A similar formulation can be extended from $2 \to 2$ processes to other QCD ones in the PYTHIA MPI framework.



(a) (b)

Figure 2.12: Two independent $2 \to 2$ scatterings (a) [79] and $2 \to 2$ scattering followed by a rescattering (b) [48].

In addition, another process can influence the collective effect of MPI, which is when an outgoing state from one scattering is allowed to become the incoming state of another

[6]Survival probabilities that a state remains untouched after a change in the evolution scale.

(*rescattering*), as schematically illustrated in Fig. 2.12b. This process is expected to have a small effect compared to independent $2 \to 2$ processes, but it is implemented in the event generator [48].
In MPI processes, the PYTHIA generator performs modifications to PDFs to take into account the conservation of flavour and momentum. The standard values are used for the first emissions, while the subsequent ones are treated with gradually modified PDFs to take into account the effects of the previous emissions. This is performed by using a rescaling factor which considers the flavour counting, by assuming the quark distributions can be split into a valence and a sea part.
The former quarks are limited by the reduction in quark content by the previous MPIs and by constraints from momentum conservation, while the sea quarks must produce a corresponding antisea parton in the beam remnant when it is extracted from a hadron, for the flavour conservation, which can later participate in another interaction. The addition of these quarks breaks the momentum sum rule, but a compensation is performed by scaling down the gluons and the non-companion sea-quark PDFs.
This is the basic handling of MPIs in PYTHIA, but further effects are introduced in the generator, such as the association with processes with initial and final-state radiation showers, which are not discussed here (see reference [40] for a description). It is worth mentioning that the magnitude of these effects can be tuned using a series of editable parameters in the generator.
A comparison of the PYTHIA8 results with and without multiple parton interactions for the D mesons as a function of charged-particle multiplicity and spherocity is shown in Chapter 5, where the experimental results of this dissertation are discussed.

2.3.2 HERWIG7

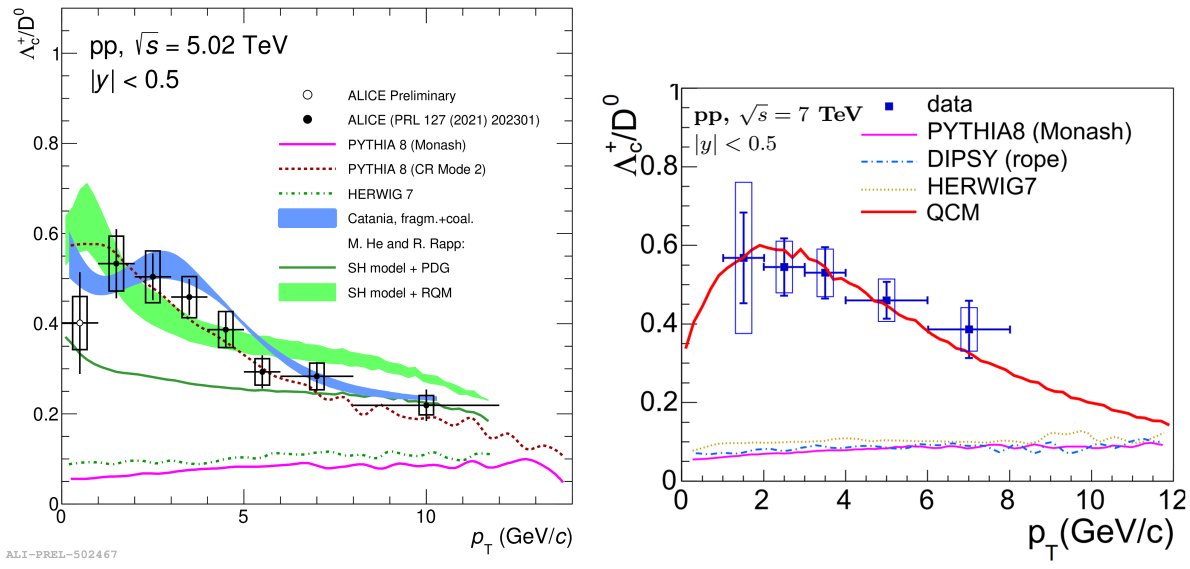
In HERWIG7 the underlying events are based on the eikonal multiple parton–parton scattering model [80], which describes them as additional semi-hard and soft partonic interactions. This mechanism derives from the assumption that at a fixed impact parameter b individual scatterings are independent and that the distribution of partons in hadrons factorises with respect to its dependence on b and on the momentum fraction x. The MPIs are tuned via the various parameters of the MPIHandler of the generator, which employs a phase-space sampling for the additional scatterings (via the MPISampler).
In a generic hard, high-momentum transfer process in HERWIG the coloured particles, in the elementary hard subprocesses, follow a perturbative evolution from the hard scale of the collisions to an infrared cutoff region. For large centre-of-mass energies, multiple scattering is dominant in the perturbative regime above the cutoff and creates additional parton showers, while non-perturbative partonic scatters are included below this region, making it possible to simulate both minimum bias events and underlying events in hard scattering processes.
This section will mainly discuss the implementation in the generator of additional semi-hard and soft scatters in the Monte Carlo simulation, as the theoretical calculation of the

multiplicities of these events is not the purpose of this work [73]. Nevertheless, from the theory, two main parameters are exploited for the description of MPI in the generator: the minimal transverse momentum p_{T}^{min} of the additional hard scatters, and the typical inverse proton radius squared (μ^2), appearing in the spatial transverse overlap of the incoming hadrons.
The semi-hard implementation was recovered from HERWIG++ [80] and starts from the first single hard scattering evolving backward to end on a valence parton, causing the partons to be (anti)quarks or (anti)diquarks. However, this process does not allow subsequent scatters, given that it saturates the valence structure of the hadron by extracting a valence parton and forming the hadron remnant with the rest. Hence, in the next interactions, the backward evolution is modified so that they terminate on a gluon. During the evolution the used PDFs are the standard ones, but with a valence quark less, hence not obeying the momentum sum rule, which does not cause issues since the algorithm is only sensitive to ratios of PDFs. When the process reaches the gluons, its colour connections are then inserted into those of the previous remnant, creating a structure that can be iterated an arbitrary number of times.
The soft scatters are generated after the perturbative evolution has ended, since the parameter p_{T}^{min} is typically of the order of the parton shower cutoff scale. Instead, the non-perturbative remnant decays produce diquarks that emit soft gluons scattering off each other. These gluons carry colour charge and in the MPI process of the generator their colour connections to the diquarks are severed so that two outgoing gluons from each soft scattering are colour connected to each other.
After the kinematics has been generated in the centre-of-mass frame, it is boosted back to the lab frame, and the diquark momenta are reshuffled so that they remain on their original mass shell. Afterwards, the available energy for the next soft interaction can be determined and the process is iterated until the requested multiplicity is reached or all the available diquarks energy has been depleted.
These are the basic mechanism of MPIs at the parton level, but in order to provide a full description of the event at the hadrons level, HERWIG also connects the multi-parton scattering to the parton shower and the hadronisation models by extending the introduced concepts [73].

2.4 Comparisons to ALICE results

The comparison of the Λ_{c}^{+} over D^0 ratio in pp collisions at $\sqrt{s} = 5.02$ TeV measured by the ALICE experiment and various Monte Carlo models is shown in Fig. 2.13a including, specifically, models that implement enhanced baryon production mechanisms. Compared to the default Monash 2013 tune in PYTHIA, the enhanced colour reconnection mode 2 describes much better the magnitude and the decreasing trend of the results as a function of p_{T}, being consistent with the experimental values within the uncertainties even at intermediate-p_{T} intervals. Mode 2 in particular is the standard tune when using this specific PYTHIA model, as suggested by its developers, though more comparisons

will be shown including all of the different enhanced colour reconnection modes in the next chapters, and in the RIVET appendix at the end of this dissertation.



(a) (b)

Figure 2.13: Preliminary ALICE results of Λ_c^+ over D^0 ratio down to $p_T \approx 0$ GeV/c in pp collisions at $\sqrt{s} = 5.02$ TeV compared to different models, and Monte Carlo event generators predictions (a) and comparison of the Λ_c^+/D^0 in pp collisions at $\sqrt{s} = 7$ TeV obtained by ALICE [65] with several models including QCM (b). This last plot was derived from [68].

A consistent description of the experimental results is also provided by the Catania and SHM+RQM models. The latter calculations show how the additional baryon states predicted by the RQM model significantly improve the results if compared to the SHM model using the PDG estimates.
On the other hand, a default run with HERWIG7, without including colour reconnection mechanisms, provides underestimated results that are aligned with the predictions of PYTHIA with Monash.
Similar conclusions are reached when comparing the models with the Λ_c^+/D^0 ratio at a lower centre-of-mass energy $\sqrt{s} = 7$ TeV. As a matter of fact, it is noticeable from Fig. 2.13b that all the models not using colour recombination mechanisms (PYTHIA8 with Monash and HERWIG7 with the default tune) are coherently underestimating the results. The DIPSY rope hadronisation model [81] introduces a kind of colour reconnection[7], but the obtained results are still underestimating the baryon-to-meson ratio.
On the other hand, the predictions obtained by the QCM framework describe properly

[7]It is based on the assumption that if several parton pairs are next to each other in geometric space, they can act together coherently to form a colour rope, but if an overlap on strings is found then they will either colour reconnect or they will end up forming a new colour rope [82].

the experimental results.

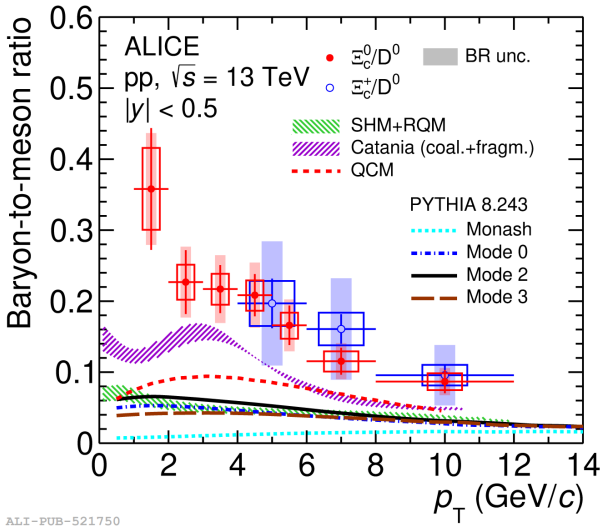


Figure 2.14: Prompt Ξ_c/D^0 baryon-to-meson ratio as a function of p_T in pp collisions at $\sqrt{s} = 13$ TeV compared to different models [83].

Additional information can be obtained from the results of the ratio between the charmed strange baryon Ξ_c and the D^0 meson yields shown in Fig. 2.14 [83]. The p_T dependence of the experimental results is similar to what was measured for the Λ_c^+/D^0 ratio shown in Fig. 2.13, but the measurements are underestimated by all the models considered, including the configurations with enhanced baryon production. However, among them, the Catania model is capable of better describing the shape of the measurement as a function of p_T.
This could lead to the conclusion that additional assumptions and different mechanisms must be further introduced to describe the production of charmed strange baryons.
The contribution to the charm production of MPI can be studied experimentally, for example, via the analysis of D mesons production as a function of multiplicity. Separating the production processes of charm in PYTHIA leads to different trends in the relative yields of the mesons as shown in Fig. 2.15a in pp collisions at $\sqrt{s} = 7$ TeV as a function of multiplicity: including first hard processes leads to a shape that is slower than linear, while the quickest production as a function of multiplicity is obtained through MPI processes. The inclusive production predictions in four different p_T intervals are presented in Fig. 2.15b, where the fastest relative yield trend with multiplicity is observed for higher transverse momentum values, while the $1 < p_T < 2$ GeV/c interval shows a linear trend. ALICE results in pp collisions at $\sqrt{s} = 7$ TeV as a function of multiplicity (Fig. 2.15c) have a trend that is faster than linear for all the five p_T intervals considered, and it is similar to PYTHIA predictions including all the production contributions, although the event generator underestimates data. This acknowledges the important role of multiple parton interactions in the description of pp collisions, which have the biggest contribution to the increase of the relative yields.

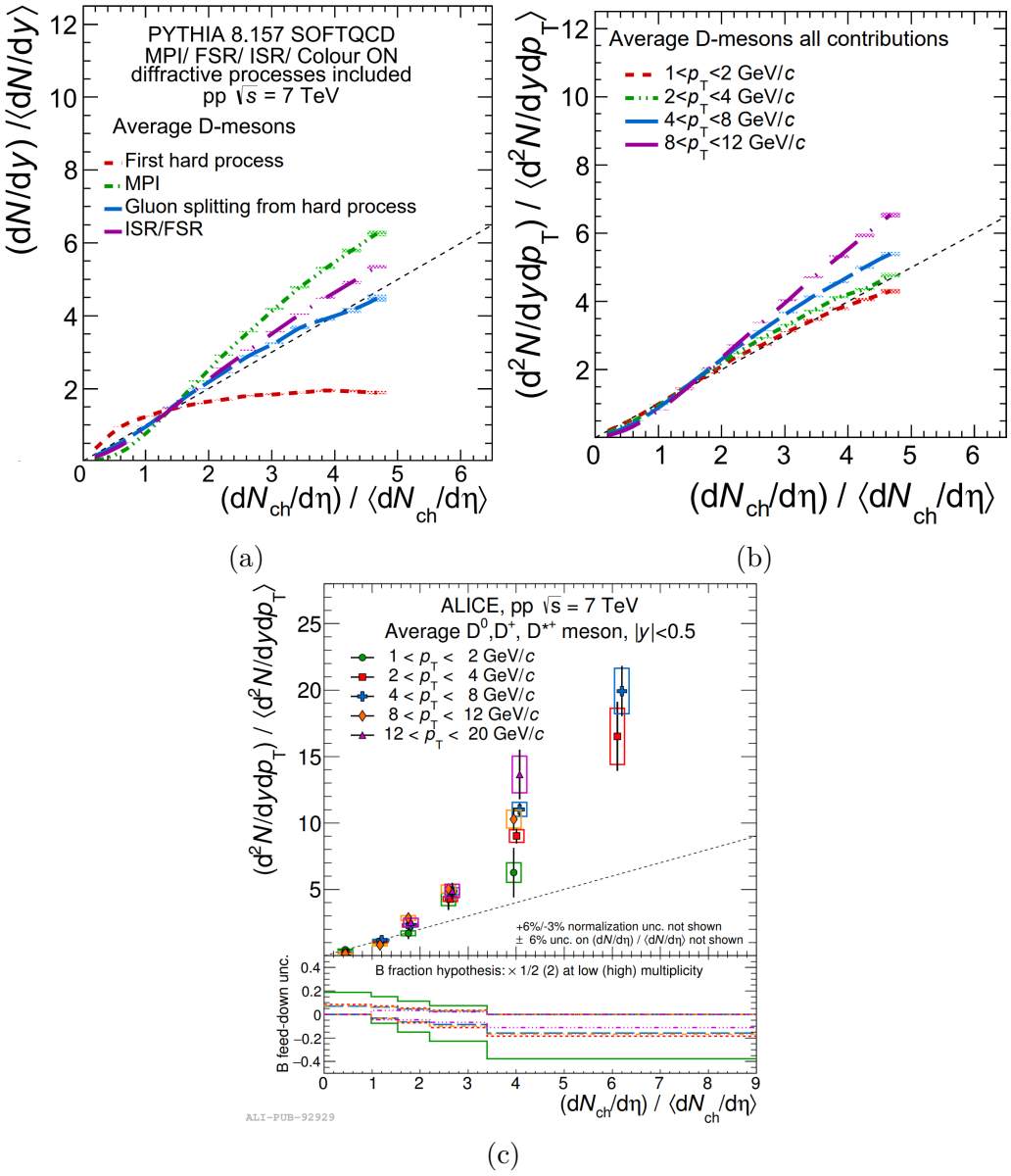


Figure 2.15: (a) Average of D-meson relative yields as a function of charged-particle multiplicity at central rapidity separating charm production processes in PYTHIA for integrated p_{T} and (b) in p_{T} intervals for all production mechanisms, derived from ALI-PUB-92978; (c) measurements in pp collisions at $\sqrt{s} = 7$ TeV as a function of multiplicity in p_{T} intervals by ALICE [84].

To conclude, at the LHC energy scale, hadronisation mechanisms can be analysed in pp collisions via the creation of a high-density parton state, in which the interaction of colour charges, carried by partons, allows us to study colour reconnection and MPI mechanisms that are not accessible at lower energies.
In this dissertation two specific channels are studied: the production of charmed $\Sigma_{\mathrm{c}}^{0,+,++}$

and Λ_c^+ baryons, and of charmed D mesons, comparing the results with the models described in this chapter. These analyses are presented, respectively, in Ch. 4 and 5.

Chapter 3

A Large Ion Collider Experiment: ALICE

The ALICE detector was built to study the Quark Gluon Plasma (QGP) thanks to a design that allows particle detection to be performed even with very high particle multiplicities originating from heavy-ion collisions at the TeV energy scale.
In this chapter, the main features of the experiment will be discussed, starting from a brief introduction to the LHC apparatus and concluding with an illustration of the ALICE analysis framework.

3.1 Introduction to the Large Hadron Collider

The Large Hadron Collider (LHC) is the biggest and most energetic hadron accelerator and collider in the world with a circonference of 26.7 km, and the possibility to run collisions with a centre-of-mass energy (per nucleons pair) up to $\sqrt{s} = 14$ TeV in proton-proton (pp) collisions and up to $\sqrt{s_{\mathrm{NN}}} = 5.5$ TeV in Pb–Pb systems as per design [85, 86, 87, 88].
The collider was installed inside the previously existing tunnel dedicated to the Large Electron-Positron Collider (LEP) at CERN in Geneva (Switzerland) and benefits from multiple acceleration stages performed by older smaller accelerators which are all illustrated in Fig. 3.1.
Regarding proton beams, all the following steps are performed in order to inject particles in the LHC: protons are extracted from hydrogen tanks which are then sequentially accelerated in the LINAC2 (a linear accelerator), the Proton Synchrotron Booster (BOOSTER), Proton Synchrotron (PS) and lastly in the Super Proton Synchrotron (SPS), after which they will experience the last acceleration process inside LHC, using Radio-Frequency cavities (RF), stabilising the beams with its ≈ 1600 superconducting magnets. The LHC magnets reach a 8 T magnetic field to maintain a circular orbit of the beams and are kept to a temperature of around 1.9 K thanks to a cooling system based on liquid helium.

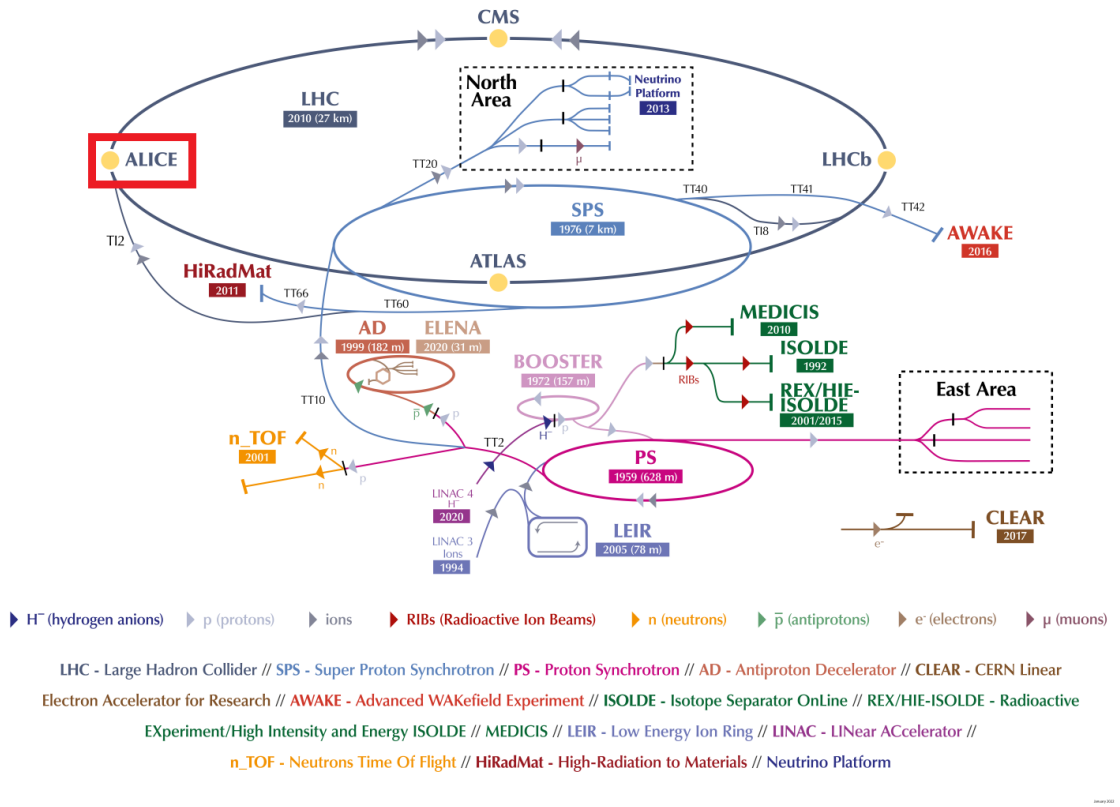


Figure 3.1: Scheme of the accelerator system ensemble at CERN that ultimately leads to the LHC. All the accelerators and main experiments are illustrated, highlighting the ALICE experiment.

The main differences between accelerating protons and ions are in the first two stages of acceleration, which for the latter include an initial ionisation and acceleration process in a linear accelerator (LINAC 3) right after the extraction of the atoms which are then accumulated in a dedicated ion ring (LEIR), before being injected inside the PS accelerator.
After all the acceleration stages, protons are inserted inside the LHC with an energy of ≈ 450 GeV, while $Pb(82)^+$ ions are injected at ≈ 177 GeV for the last acceleration that will lead to the collision of the beams at specific interaction points (IPs) of the tunnel, in which the four main detectors of CERN are installed: A Toroidal LHC ApparatuS (ATLAS), Compact Muon Solenoid (CMS), Large Hadron Collider Beauty (LHCb), and A Large Ion Collider Experiment (ALICE).
Although all detectors perform a wide range of measurements, they were optimised for specific purposes which include the study of beauty quark physics for LHCb and in particular the study of CP-violating processes, the study of the Higgs boson and its interactions by CMS and ATLAS, and the Quark Gluon Plasma formation in heavy-ion collisions by the ALICE experiment.

The first collisions at LHC were recorded in 2009, which started Run 1 and continued until 2013. After the first long shutdown of the machine, data taking was restarted in April 2015 up to December 2018 (Run 2). For the purpose of the analyses discussed in this dissertation, only data from Run 2 were considered, since pp collisions at $\sqrt{s} = 13$ TeV were achieved only in that specific period, so the further description of the ALICE detector is limited to its status during Run 2 and does not include the latest upgrades to the experiment.

3.2 The ALICE detector

The experiment is required to track charged particles with extreme precision down to very low momenta (which reach 80 MeV/c for pions) in an environment dominated by a very large amount of charged particles created in Pb–Pb collisions. The originally estimated charged-particle density at mid-rapidity, between $2000 < \mathrm{d}N/\mathrm{d}y < 8000$ [89], has driven the design of the experiment, so a series of high granularity and low material budget detectors were developed.

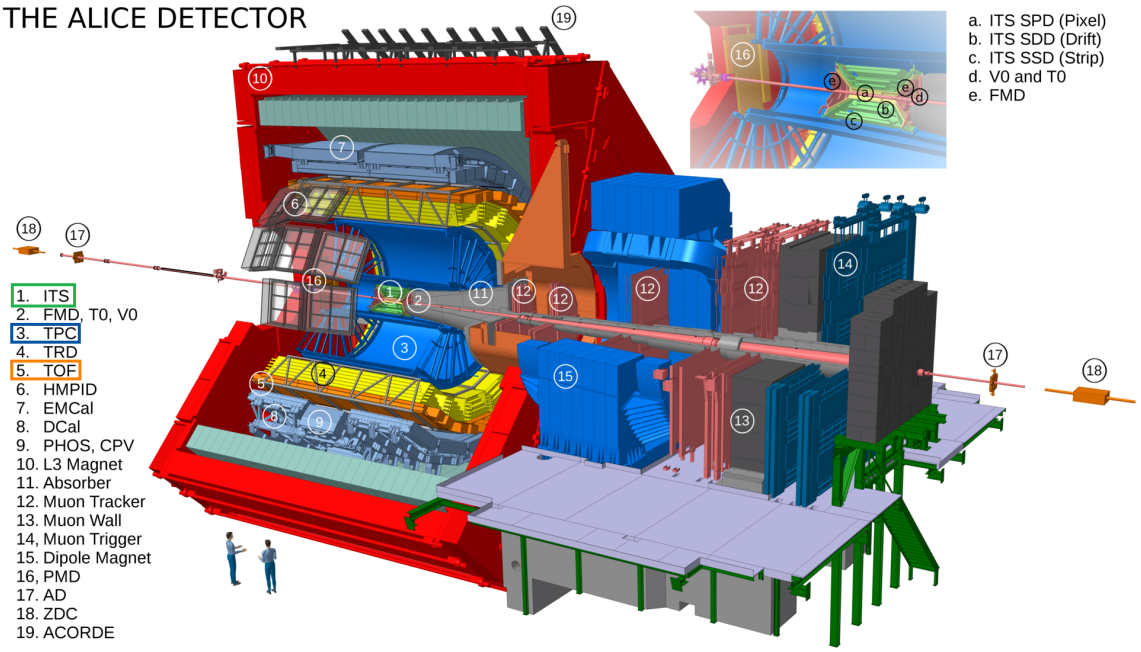


Figure 3.2: Schematic illustration of the ALICE experiment during Run 2. Labels indicating all the subdetectors are shown and the main parts used for the analysis are highlighted. Figure taken from ALICE repository ©.

In order to properly analyse various physics signals from particle collisions, an important feature of the experiment is also its excellent particle identification (PID) capabilities.
The experiment is 26 m long, 16 m tall, and 16 m wide, weighting more than 10 ktons

(Fig. 3.2). It can be divided into two main parts based on the pseudorapidity coverage (η): detectors inside and around the L3 magnet covering the range $|\eta| < 0.9$ (central barrel) and the muon spectrometer, covering the interval $-4 < \eta < -2.5$ and the full azimuthal angle.
Apart from the array of scintillators used for triggering cosmic rays (ACORDE)[90], useful for calibration and alignment purposes and located on top of the L3 magnet, Fig. 3.3 shows a transversal view of all the central barrel detectors used during Run 2 data taking.

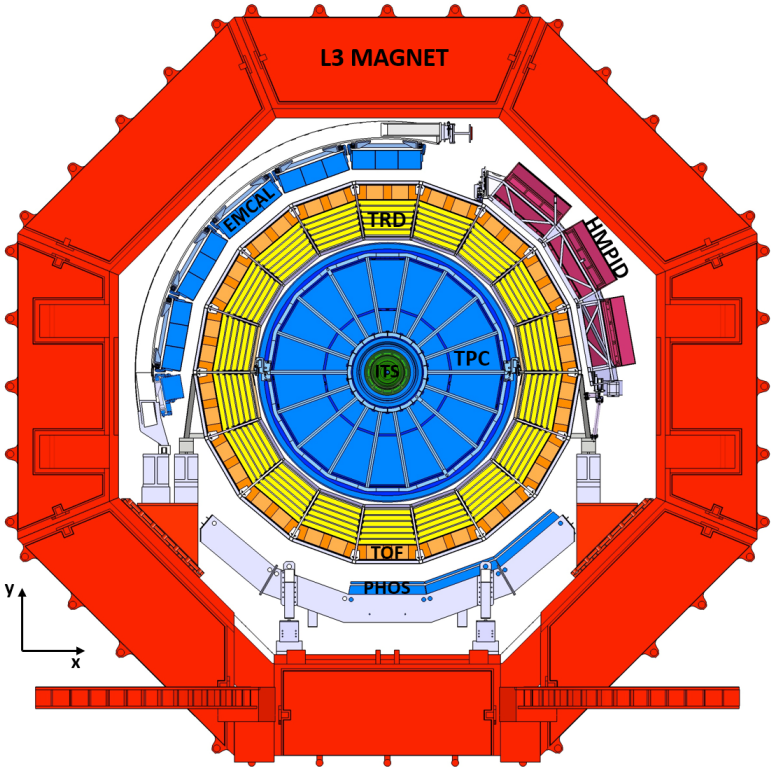


Figure 3.3: Transverse cross section of the central barrel of the ALICE experiment, showing inside the L3 magnet with labels indicating the positions of the main detectors. Base figure taken from ALICE repository ©.

Ordering them from the inner to the outer shell of the detector, they are: 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 two electromagnetic calorimeters (PHOS and EMCAL). The last three detectors are the only ones that do not cover the full azimuthal angle.
The forward region muon spectrometer includes a series of absorbers, a large dipole magnet (able to reach a magnetic field of 0.67 T), and 14 Resistive plate chambers (RPC) used for tracking and triggering. Furthermore, Fig. 3.2 also shows a series of smaller detectors used in the experiment for global event characterisation and triggering, which are:

the Zero Degree Calorimeter (ZDC) [89], the Photon Multiplicity Detector (PMD)[91], the Forward Multiplicity Detector (FMD), two arrays of photo-multipliers equipped with Cherenkov radiatiors (T0) and two arrays of scintillator counters (V0A and V0C)[92]. Out of all the many detectors of the experiment, in the next paragraphs the ones used for the analyses performed in this dissertation will be described in detail, which are specifically the ITS, TPC and TOF detectors.

3.2.1 The Inner Tracking System (ITS)

ITS is the closest detector to the interaction point of the ALICE experiment as illustrated in Fig. 3.2. It is made up of six layers of silicon trackers, coupled in series of two, as shown in Fig. 3.4, employing different technologies. In particular, the ones that will be thoroughly used in this dissertation are the two innermost layers of silicon pixel detectors (SPD), followed by two layers of silicon drift detectors (SDD), and finally the two outermost double-sided silicon strip detectors (SSD) layers [93, 94].

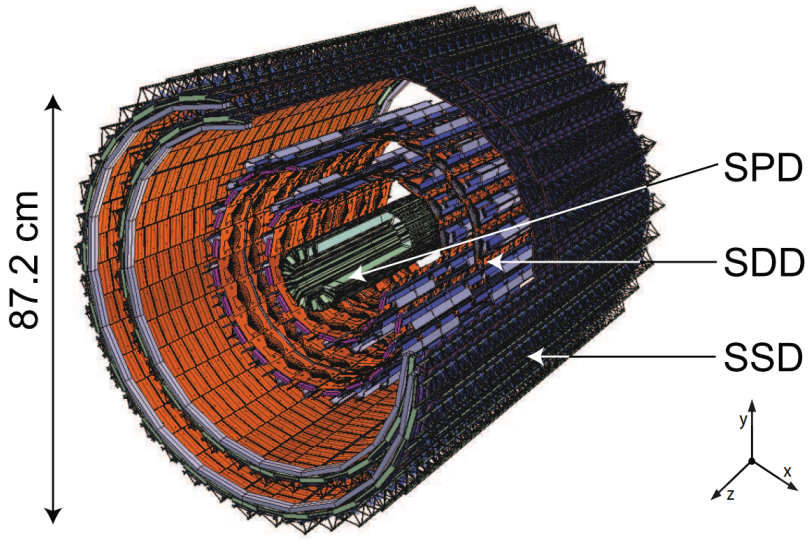
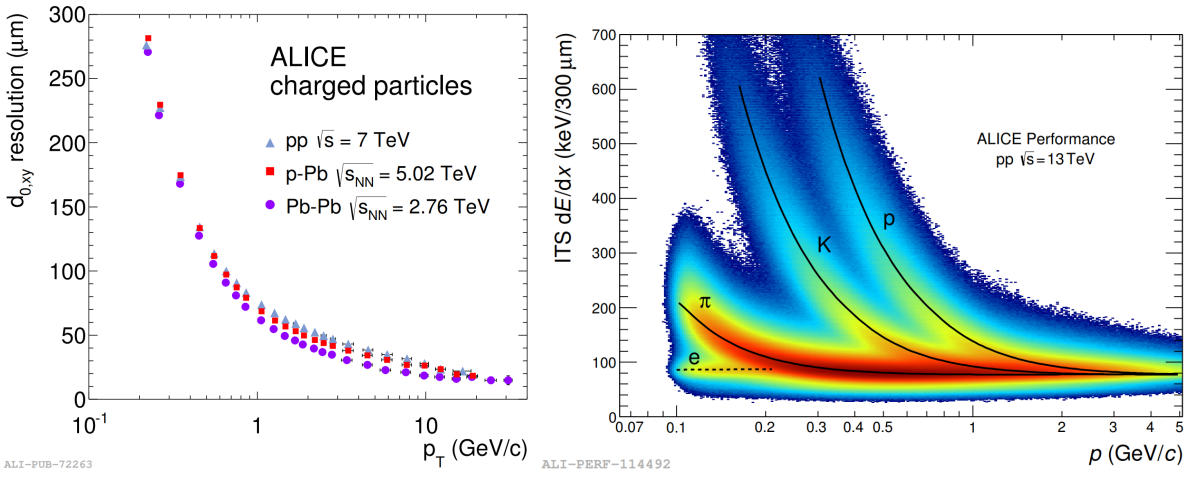


Figure 3.4: 3D illustration of the ITS detector with its 6 layers [94].

ITS was built primarily for the tracking of charged particles, the determination of the interaction point (primary vertex), and for the reconstruction of secondary vertices, which is useful for the analysis of short-lived hadrons and for measuring the impact parameter of the tracks (b)[1]. The resolution of b defines the tracking precision of the system and it is shown in the x-y plane in Fig. 3.5a as a function of the transverse momentum for identified ITS-TPC charged particle tracks in different collision systems. The detector has a resolution better than 100 μm for the primary vertex reconstruction and it can track and identify particles with a momentum below 200 MeV/c, as shown in the performance example in Fig. 3.5b of the energy loss distribution obtained by the

[1]Distance of closest approach between the track trajectory and the primary vertex.

detector in 2016 in pp collisions at $\sqrt{s} = 13$ TeV. In the last plot, the average specific energy loss $\mathrm{d}E/\mathrm{d}x$ of the charged particles is shown as a function of their momentum in a pure standalone ITS reconstruction. A fit based on a Bethe-Bloch parametrised function was also applied to the distribution (black lines). A similar plot in Pb–Pb at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV can be found in [95].



(a) (b)

Figure 3.5: ITS impact parameter resolution with Run 2 data for all charged tracks in the x-y plane as a function of p_{T} in pp collisions at $\sqrt{s} = 7$ TeV, p–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV and in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV (a), and PID performance plot in pp collisions at $\sqrt{s} = 13$ TeV (b) [95].

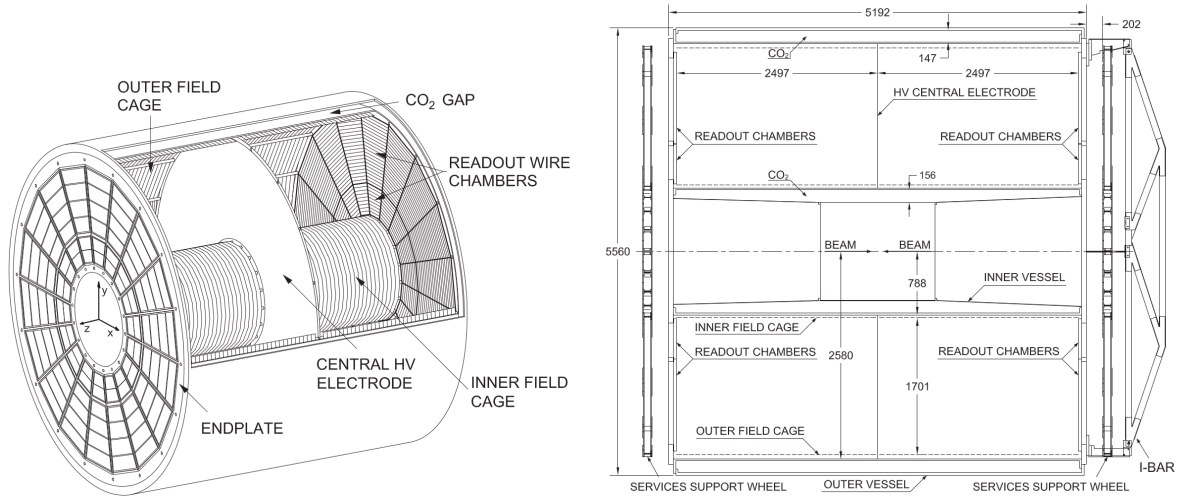
The *primary vertex* position is determined directly through the innermost SPD layers of the detector, which are also used to locate secondary vertices and measure the impact parameter. These layers also provide a fast trigger signal thanks to their fast response. The remaining outer SDD and SSD layers are also used for PID via $\mathrm{d}E/\mathrm{d}x$ and the latter are also fundamental for the prolongation of the TPC tracks, covering dead regions of the detector and improving its momentum and angle resolution. The main parameters and characteristics of the three layer types are listed in Tab. 3.1.

3.2.2 The Time Projection Chamber (TPC)

The TPC is the main tracking and PID detector of charged particles of the ALICE experiment in the central barrel region. The detector consists of a hollow cylinder aligned with the LHC beampipe and parallel to the solenoid magnetic field [97]. It contains a high-voltage conductive electrode (Fig. 3.6a) that is charged up to 100 kV providing, together with a voltage dividing network on the surface of the outer and inner cylinder, an axial electric field of 400 V/m. The active volume of approximately 90 m^2 (exact

Parameter		Silicon Pixel	Silicon Drift	Silicon Strip
Spatial precision rφ	(μm)	12	35	20
Spatial precision z	(μm)	100	25	830
Two track resolution rφ	(μm)	100	200	300
Two track resolution z	(μm)	850	600	2400
Cell size	(μm^2)	50×425	202×294	95×40000
Active area per module	(mm^2)	12.8×69.6	72.5×75.3	73×40
Readout channels per module		40960	2×256	2×768
Total number of modules		240	260	1698
Total number of readout channels	(k)	9835	133	2608

Table 3.1: Characteristics and parameters of the different layers of the ITS detector [89].



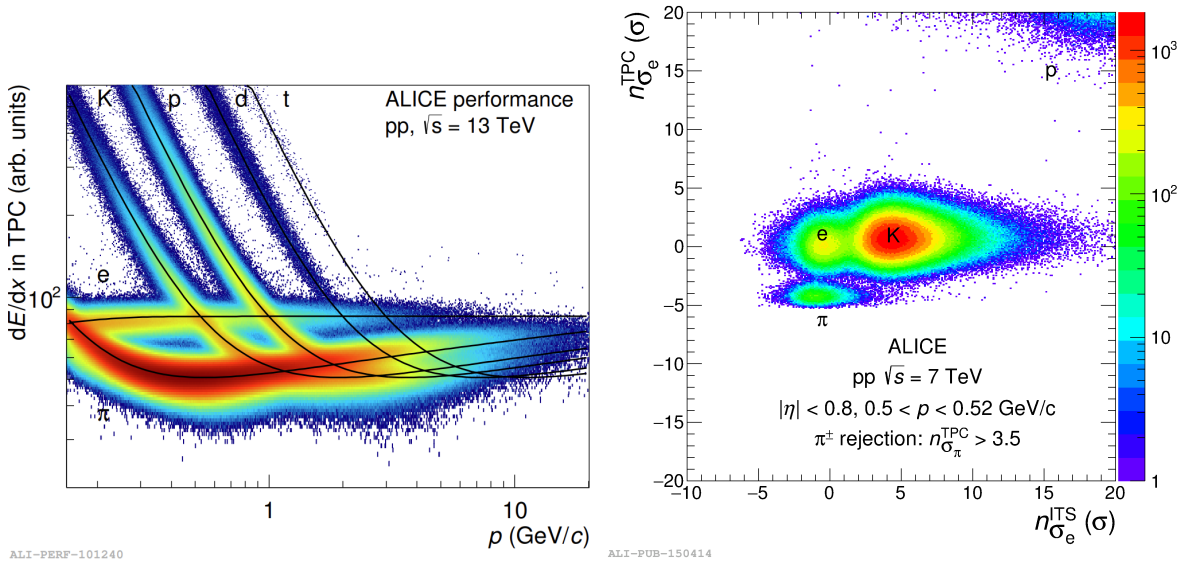
(a) (b)

Figure 3.6: 3D illustration of the TPC field cage (a) and longitudinal cross section with chamber dimensions in millimetres (b) [96].

values are reported in Fig. 3.6b) is filled with a mixture of Ne-CO_2-N_2[2] at atmospheric pressure that ionises when charged particles interact with it.
The electrons generated by the ionisation processes drift toward the detector endplates, thanks to the electric field, where their arrival points in the cylinder plane are measured through proportional multi-wire chambers with cathode pads readout installed in the 18 trapezoidal sectors of TPC. The full trajectory of the charged particles in the active volume of the detector is then precisely determined by combining the charge deposited on the cathode pad and accurate measurement of the drift time of the electrons [96].

The detector allows the reconstruction of charged particle tracks with a p_{T} as low as

[2]For the first two years of Run 2 Neon was replaced by Argon, but the replacement was taken back in 2017 due to large phase-charge distortions caused by the gas mixture.



(a) (b)

Figure 3.7: Performance plot of the TPC detector in pp collisions at $\sqrt{s} = 13$ TeV using the average energy loss $\mathrm{d}E/\mathrm{d}x$ of charged particles as a function of their momentum (a), and (b) TPC PID signal expressed as $\mathrm{n}\sigma$(TPC) of electrons as a function of the ITS PID signal for the same particles using selected tracks with $0.5 < p < 0.52$ GeV/c [98].

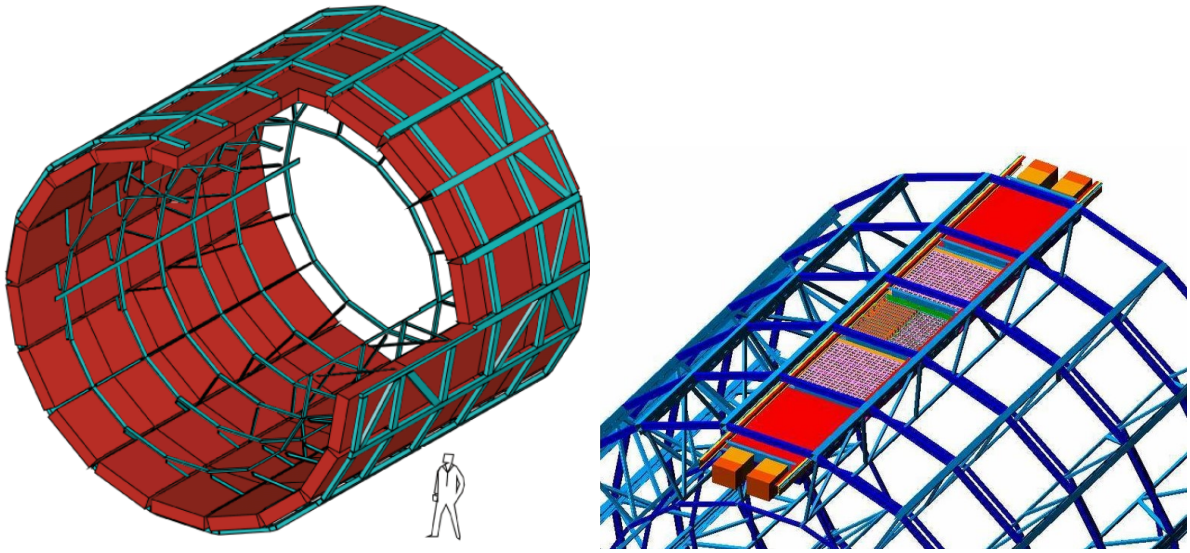
about 100 MeV/c (for pions), and up to 100 GeV/c. PID is performed by measuring the specific energy loss (parametrised by the Bethe-Bloch formula [4]) and the identification is based on the comparison of the measured $\mathrm{d}E/\mathrm{d}x$ with the expectation for a specific charged particle as a function of p_{T}.
An example of the specific energy loss measurement in pp collisions at $\sqrt{s} = 13$ TeV is shown in Fig. 3.7a as a function of the transverse momentum, while Fig. 3.7b illustrates the combined PID capabilities of TPC and ITS in pp collisions at $\sqrt{s} = 7$ TeV using the number of sigma for electrons for TPC as a function of the ITS PID σ after applying the pion rejection in the TPC with a $n_\sigma > 3.5$ [98]. For a detector with a Gaussian response, the PID information is usually referred as the deviation of the measured signal S_α from the one expected for a specific particle H_i using the expected average signal $\hat{S}(H_i)_\alpha$ of the α=(ITS,TPC,...) detector, normalised by the detector resolution σ^i_α:

$$n_{\sigma^i_\alpha} = \frac{S_\alpha - \hat{S}(H_i)_\alpha}{\sigma^i_\alpha}. \tag{3.1}$$

This approach implies the identification of a track if the $n_{\sigma^i_\alpha}$ of that species falls within a certain range around the estimate, which is generally around 2 or 3 σ [99].

3.2.3 The Time-of-Flight detector (TOF)

The TOF detector is the fourth in order of distance from the interaction point, located at 3.7 m from the beam axis (after ITS, TPC and TRD), and covers the midrapidity region ($|\eta| < 0.9$) and the full azimuthal angle with a large cylindrical active area array of $\approx 141\text{m}^2$ that provides particle identification and trigger capabilities for the ALICE experiment [100, 101].

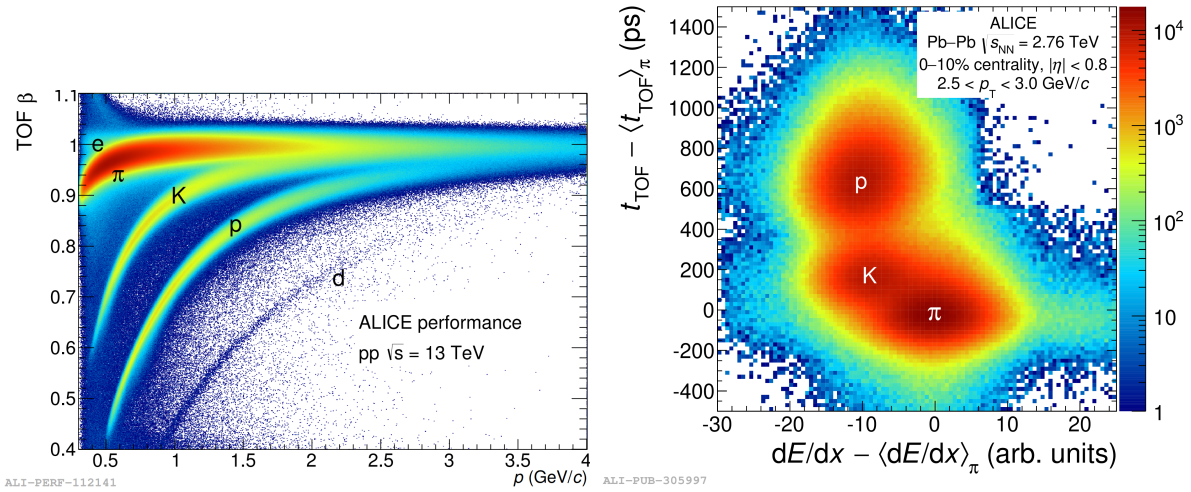


(a) (b)

Figure 3.8: Illustration of the TOF detector MRPCs (a) and close schematic of the components of a single supermodule of the detector (b).

The system is composed of 1593 Multi-Gap Resistive Plate Chambers (MRPC) strip detectors and, like the TPC, it is divided into 18 sectors (shown in Fig. 3.8a). Each chamber is made of 10-gaps double-stack strip detectors, reaching a time resolution of 40 ps. Each MRPC strip is also segmented into two rows of 48 readout pickup pads of approximately 2.5×3.5 cm^2 for a total of 96 pads per strip and 152928 active readout channels in the entire detector [102].
The analog signal collected from the MRPCs is pre-amplified and discriminated in Front-End Analogue cards, sitting close to the MRPC. The discriminated signals are then read and digitised by TDC cards housed, for each supermodule, in 4 crates located at the outer edges of each sector, as in Fig. 3.8b. The digitised data are read-out and sent to the central DAQ of the experiment.
Specifically, the time-of-flight of each charged particle is measured by determining the event collision time (provided by the T0 detector or by the TOF itself through a combinatorial algorithm based on a χ^2 minimisation between all mass hypotheses) and the time in which the particle hits the detector [103].
The excellent MRPC resolution and the optimised electronics allow the detector to iden-

tify, with a precision greater than 3σ, protons from kaons up to a momentum of ≈ 4 GeV/c and kaons from pions up to about 2.5 GeV/c. The identification of particles is based on the distribution of the particles measured velocity (β) as a function of the momentum measured by the TPC detector, as can be seen in pp collisions at $\sqrt{s} = 13$ TeV in Fig. 3.9a.



(a) (b)

Figure 3.9: Performance plot of the TOF detector in pp collisions $\sqrt{s} = 13$ TeV with β expressed as a function of particle momentum (a) and combined particle identification in TPC and TOF in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV (b) [99].

Taking into account tracks with a momentum of 1 GeV/c, it was shown that TOF reaches an overall resolution of 56 ps in Pb–Pb and 84 ps in pp collisions [104], where more uncertainties arise due to the poorer resolution in the determination of the event collision time and greater uncertainties in the calibration. The separation of hadron species can be further improved by combining signals from different detectors, increasing the momentum range for the identified particle measurements, as shown in Fig. 3.9b, where the difference between the measured and expected PID signals for the TPC and TOF detectors is illustrated in a two-dimensional scatter plot using data from Run 1 in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV [99].

3.2.4 V0 and T0 detectors

The two detectors V0 and T0 are important, respectively, for trigger, and for particle identification and reconstruction performed via ITS, TPC, and TOF [92].

V0A and V0C

The V0 detector is made up of two arrays of 32 scintillators along the beamline on both sides of the ALICE interaction point, called V0A and V0C. They are segmented into

4 rings along the radial direction and 8 sectors throughout the azimuth, and they are located, respectively, 340 cm and 90 cm away from the nominal interaction point, the former on the opposite side of the muon spectrometer, while the latter is in front of the hadronic absorber.
They cover, respectively, the pseudorapidity intervals $2.8 < \eta < 5.1$ and $-3.1 < \eta < -1.7$ in the full azimuthal angle and their purpose is mainly to provide trigger to the experiment and define event characteristics such as centrality in heavy-ion collisions and multiplicity, which is obtainable thanks to the proportionality of the scintillation light signal amplitude to the number of particles interacting with the detector.
A logical AND between the arrays of the V0 detector is used to define both the Minimum Bias and the High Multiplicity triggers that will be used for the analyses performed in this dissertation.

T0

Similarly to V0, T0 is composed of two arrays of Cherenkov detectors, T0A and T0C, covering the full azimuthal angle and, respectively, the pseudorapidity intervals $4.61 < \eta < 4.92$ and $-3.28 < \eta < -2.97$. The former array is placed at a distance of 375 cm from the nominal interaction point, while the latter is on the opposite side at a distance of 73 cm from the IP.
The detector is capable of achieving a time resolution of 20-25 ps in Pb–Pb collisions and $\approx$ 40 ps in pp collisions, providing a very precise measurement of the event collision time for the TOF detector and online luminosity monitoring.

3.3 Trigger

The Central Trigger Processor (CTP)[105] is responsible for the selection of the collision events, which will be stored later. It receives input from trigger detectors and generates a central trigger distributed to detector sub-systems if a series of requirements are met. The evaluation occurs at each clock cycle ($\approx$ 25 ns), but due to the different latency of the various detectors, the trigger procedure is divided into three different layers. The fastest trigger is **L0** which is produced with a latency of 0.9 μs (from the collision of the beams) and is based on "fast" trigger inputs (from SPD, V0 and T0). The L0 accepted events are then considered as candidates for the second level **L1** trigger algorithm whose results are obtained 6.5 μs after the interaction takes place, due to slower trigger detectors (longer computing time of TRD and EMCal, for example, depending on the event size). Finally, the **L2** trigger level is performed after 100 μs, which is approximately the maximum drift time of electrons in the TPC detector.
In case all triggers are fulfilled, the event fragments from the different sub-systems are sent to the Data Acquisition System (DAQ)[106] which coordinates the data flow from electronics to the tape archiving procedure and includes a High-Level Trigger (HLT) that rejects events not satisfying a fast track reconstruction and online analysis, builds

events, and compresses data. If the events are labelled *good* after the last step, they are stored on tape.

Detector	Function	Level
SPD	hit-multiplicity and hit-topology based triggers	L0
TRD	electron, high-p_{T} particle and charged-jet triggers	L1
TOF	multiplicity, topological (back-to-back) and cosmic-ray triggers	L0/L1
PHOS	photon trigger	L0
EMCal	photon and neutral-jet triggers	L0/L1
ACORDE	cosmic-ray trigger (single and multiple hits)	L0
V0	coincidence based MB interaction and centrality triggers	L0
T0	event-vertex selection and interaction triggers	L0
ZDC	MB interaction and electromagnetic-dissociation triggers in Pb-Pb	L1
MTR	single-muon and dimuon triggers	L0

Table 3.2: Trigger capabilities of the ALICE detectors [95].

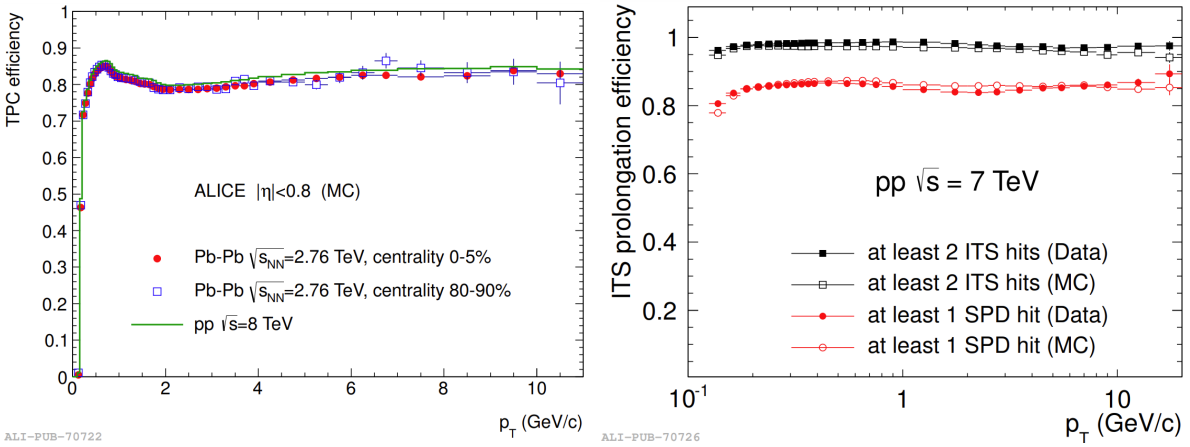
In general, trigger decisions are made using a total of 60 trigger inputs divided into different trigger levels, and starting from L0 (which receives 24 of them). The CTP is also responsible for sending the BUSY signal to all detectors in an affected cluster[3]. The detectors that provide input to trigger decisions are listed in Tab. 3.2.
In the analyses presented in this dissertation, the trigger schemes used are the minimum bias (MB) and the high-multiplicity trigger (HMSPD). The former events are selected by considering the conditions with less bias during data acquisition, with measurements at low transverse momenta, in which very demanding trigger selections cannot be applied, and they require the coincidence of the V0A and V0C signals with the arrival time of the proton bunches. On the other hand, the high-multiplicity trigger requires a coincidence of the signal between the SPD and V0 detectors. About 100 million events were collected in 2018 by the experiment using this selection and were used to perform the analysis of D mesons at high multiplicity intervals.

3.4 ALICE data reconstruction

In the ALICE experiment, tracking and vertex reconstruction are procedures performed offline in sequence. Track finding in the central barrel begins with the conversion of the data from each detector to *clusters*, which are characterised by a series of parameters such as positions, signal amplitudes, etc., and their associated errors [95]. Then a preliminary determination of the interaction vertex is performed using the two layers of the SPD

[3]Group of detectors sending/receiving specific triggers from the CTP. During a run it is possible to operate different clusters of detectors, typically the barrel detectors and the forward muon spectrometer, for example.

detector via the convergence in space of the maximum number of tracklets[4]. Due to the pile-up in pp collisions, this algorithm is performed several times, discarding at each repetition the clusters that contributed to already reconstructed vertices. Following this method, the first vertex found, which has the largest number of tracklets, is considered as the primary one.
In low multiplicity events (i.e. in pp collisions), in case no coincidence points are detected, the algorithm performs a one-dimensional minimisation of the interaction point along the beamline axis (z) employing the distance of closest approach (DCA) of the tracklets to the average beam position in the transverse plane.



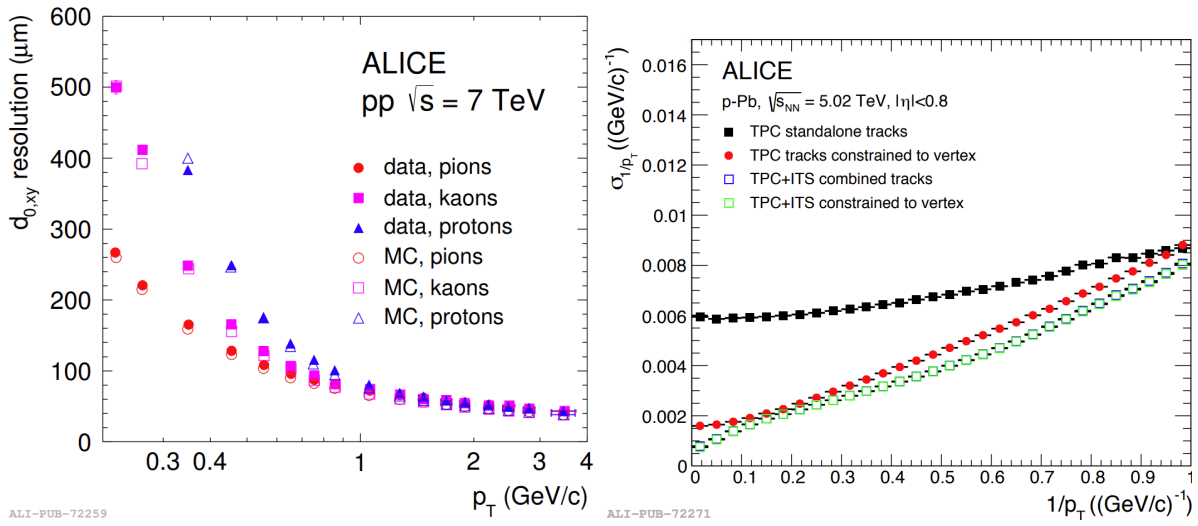
(a) (b)

Figure 3.10: (a) TPC track finding efficiency for primary particles in pp and Pb–Pb simulations and (b) ITS-TPC matching efficiency as a function of p_{T} in pp collisions at $\sqrt{s} = 7$ TeV [95].

The next step is the reconstruction of the track, which is divided in three stages. The first one starts by reconstructing tracks in the TPC at large radius: track seeds are built initially using two TPC clusters and the vertex point, which is then removed as a constraint in favour of three detector clusters. At each step of the process, the seeds are propagated inward the TPC with the closest cluster satisfying a proximity cut.
An algorithm makes sure that the same physical tracks are not reconstructed multiple times, and the computed results obtained are propagated to the inner radius of the TPC. Tracks are accepted only if they have at least 20 clusters of the 159 available and they miss no more than 50% of the estimated clusters for a given track position. In Fig. 3.10a the detector tracking efficiency is shown in pp at $\sqrt{s} = 8$ TeV and Pb–Pb at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV at different centrality intervals [95].
The reconstructed tracks from the TPC are then propagated to the outer SSD layer, becoming the seeds for track reconstruction in the ITS. Similarly to what was done

[4]Segments defined by pairs of clusters in the SPD. They will be discussed later to determine the multiplicity of the events.

for the TPC, the seeds are extended inward the detector and updated at each layer using proximity cuts that take into account positions and errors. The highest quality track candidate (selected through reduced χ^2 and checked for cluster sharing with other similar ones), from each *tree* of track hypothesis from the TPC, is eventually added to the reconstructed event of the TPC. The combined efficiency obtained from the TPC track prolongation to the ITS detector in pp and Pb–Pb collisions as a function of p_{T} is shown in Fig. 3.10b, where the results for both experimental data and Monte Carlo simulations are illustrated considering different contributions of the ITS layers. In particular, a higher efficiency is obtained considering 2 hits in the ITS, rather than a single one on the SPD layers.
The ITS detector is used in addition as a standalone tracker for the clusters belonging to the cutoff regions of the TPC detector at low p_{T} (due to energy loss and multiple scattering in the detector material), which are below 200 MeV/c for pions and 400 MeV/c for protons. In this region, a drop in reconstruction efficiency can be observed in Fig. 3.10a.



(a) (b)

Figure 3.11: Distance to the primary vertex resolution for ITS–TPC tracks (a) and p_{T} resolution for standalone TPC and combined TPC+ITS matched tracks with and without vertex constraint (b) [95].

After the ITS reconstruction is complete, the tracks of TPC and ITS detectors are extrapolated to their point of closest approach to the preliminary interaction vertex and they are propagated in the outward direction via the Kalman filter algorithm [107] to match the clusters in the TRD, TOF and the signals from the EMCal, PHOS and HMPID detectors. The filter is used again for refitting the tracks with the clusters found in the previous stages.
Finally, the tracks are propagated inward from the outer radius of the TPC and, by refitting the previously found clusters, the characteristics of the tracks are determined.

Tracks from the primary vertex follow this procedure, whereas secondary tracks, coming from decays and further interactions with the detector materials, can be suppressed using cuts on the longitudinal and transverse distances of closest approach to the primary vertex.
The achievable resolution of the transverse distance from the primary vertex for tracks reconstructed by ITS and TPC combined in pp collisions at $\sqrt{s} = 7$ TeV for different hadrons with simulated and experimental data is shown in Fig. 3.11a, while the transverse momentum resolution for both TPC and ITS-TPC combined tracks is shown in Fig. 3.11b in p–Pb at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. The latter plot is expressed as a function of inverse-p_{T} because the variable can be extracted directly from the covariance matrix of the Kalman filter and it is connected to the relative resolution via the formula

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} = p_{\mathrm{T}}\, \sigma_{1/p_{\mathrm{T}}} \,. \tag{3.2}$$

The effect of applying the constraint to the primary vertex of the reconstructed tracks is also shown in the plot, both in the TPC only and TPC+ITS combined reconstruction [95].

3.5 ALICE software framework

The ALICE experiment withstands a very large amount of data that need to be processed, compressed in size, and made accessible to analysers, who will use a series of tools that require specific data formatting. The basic component of a vast majority of analyses performed by the Collaboration is the `ROOT` software developed by CERN [108], which was designed to perform statistical analyses, data visualisation, and storage purposes for high-energy physics experiment. It is almost entirely based on C++, but it also contains some integration with other languages such as `Python` and `R`.
The `AliRoot` framework [109] was developed from the ROOT software by the ALICE Collaboration and includes detailed detector geometry specifications with each subdetector described in independent modules for better simulation and reconstruction. In fact, in addition to the analysis of experimental data, `AliRoot` is also used for simulation, calibration, alignment, and reconstruction. In pp collisions, simulated events are generated from Monte Carlo (MC) event generators such as PYTHIA8 [40] and HERWIG7 [39] using various tuning parameters.
Within the framework, all subdetector modules are fully described, including material budget, support structures, and the response to particle interaction is simulated using `GEANT3` [110], `GEANT4` [111] and `FLUKA` [112]. In order to obtain a simulation close to the real experimental conditions, the generated data are anchored to experimental data-taking periods, simulating the status of the detectors when data were obtained (for example considering channels that were off for some malfunctioning).
The simulated MC events and experimentally obtained events are then stored in Event Summary Data (ESD) ROOT files, which include all the physics information required to

perform the analysis and quality assurance data for each subdetector. From these files containing the entirety of the information stored after the trigger procedures, smaller Analysis Object Data (AOD) ROOT files are produced containing only selected parts needed for specific analyses, making the data easier to handle and reducing the computing time to process them.
All computing tasks (simulations, reconstruction, and analysis) are performed through the Worldwide LHC Computing Grid (WLCG), which is a geographically distributed infrastructure connecting a large number of computing centres in 42 countries [113, 114]. WLCG centres are divided in hierarchical order based on three *tiers*: Tier 0 centres comprehend the CERN computing centre and the Wigner Research Centre for Physics in Budapest, they hold the main copy of all the experimental data and perform the first reconstruction on the raw events; Tier 1 structures hold a second copy of the raw data and they are involved in the re-processing and in part of the reconstruction; Tier 2, finally, are the centres dedicated to the MC simulations and to the processing of the analysis.
The ALICE Environment (AliEn) interface [115], provided with the `AliRoot` package, allows ALICE users to easily access the data on the WLCG, performing analysis tasks and simulations (via command-line or the web-based MonALISA interface [116]) and monitoring their status.
In addition, the offline framework of ALICE includes also the AliPhysics software which contains multiple physics analysis tasks designed to process all the data processed by `AliRoot` and stored in the grid, providing the analysers a user-developed series of useful algorithms and macros.

Chapter 4

Measurement of $\Sigma_{\rm c}^{0,++}$ and $\Lambda_{\rm c}^{+}$ production cross section in pp collisions at $\sqrt{s} = 13$ TeV

As discussed in the previous chapters, the analysis of the heavy-flavour production cross section is an important tool to validate pQCD calculations. Furthermore, recent measurements at the LHC show charm baryon-to-meson ratios with a significant dependence on $p_{\rm T}$ in pp collisions, indicating an enhancement in the charmed baryons production compared to models results (based on $\rm e^+e^-$ and ep $p_{\rm T}$ integrated measurements) which, on the other hand, have an overall flat trend. In particular, in this dissertation the ratios $\Lambda_{\rm c}^+/{\rm D}^0$ and $\Xi_{\rm c}/{\rm D}^0$ have been discussed in pp collisions and compared to different MC models (see Chap. 2).
Another charmed heavy baryon that was not yet measured in hadronic collisions at CERN is the $\Sigma_{\rm c}(2455)$, which is identified by the isospin triplet I = 1 baryons $\Sigma_{\rm c}^0$, $\Sigma_{\rm c}^+$ and $\Sigma_{\rm c}^{++}$, partner of the $\Lambda_{\rm c}$ baryon (isospin singlet).
The enhanced colour reconnection model implemented in PYTHIA8 [46], described in Chap. 2, predicts an increase in the production measurement of $\Sigma_{\rm c}$ states, which might, consequently, increase the production of $\Lambda_{\rm c}^+$ when decaying. The enhancement of $\Lambda_{\rm c}^+$ production was observed already in ALICE results with Run 1 data [65] (Fig. 2.13b). One of the hypotheses discussed to explain this result is an increased feed-down from $\Sigma_{\rm c}$ states, which was originally proposed in [117]. Hence, this was one of the motivations of this study. This analysis reports, for the first time, the production of the $\Sigma_{\rm c}$ baryon in hadronic collisions using the large statistics of Run 2 and data from the ALICE experiment. Using the same data, the production of $\Lambda_{\rm c}^+$ and the feed-down contribution of $\Lambda_{\rm c}^+$ from $\Sigma_{\rm c}^{0,++}$ baryons have been measured as well.
In ALICE, $\Lambda_{\rm c}^+$ is studied via the two decay channels $\Lambda_{\rm c}^+ \rightarrow {\rm pK}^+\pi^-$ and $\Lambda_{\rm c}^+ \rightarrow {\rm pK}_{\rm s}^0$, which have branching ractios (BR) respectively of (6.28 $\pm$ 0.32)% and (1.59 $\pm$ 0.08)% [4]. The $\Lambda_{\rm c}^+$ baryon (udc) has a mass $m = 2286.46 \pm 0.14$ MeV/c^2 and a decay length of $c\tau = 60.7$ μm, while the masses of $\Sigma_{\rm c}^0$(ddc) and $\Sigma_{\rm c}^{++}$(uuc) are $m_{\Sigma_{\rm c}^0} = 2453.75 \pm 0.14$ MeV/c^2 and $m_{\Sigma_{\rm c}^{++}} = 2453.97 \pm 0.14$ MeV/c^2 with a resonance width respectively of

$\Gamma \approx 1.83$ MeV/c^2 and $\Gamma \approx 1.89$ MeV/c^2 [4].
In this chapter, I report my analysis and measurement of the production cross section of prompt $\Sigma_c^{0,++}$, Λ_c^+ and $\Lambda_c^+ \leftarrow \Sigma_c^{0,++}$ baryons at mid-rapidity in pp collisions at $\sqrt{s} = 13$ TeV by studying the $\Lambda_c^+ \rightarrow p+K_s^0$ decay channel with $K_s^0 \rightarrow \pi^+ + \pi^-$, that has a BR of $(69.20 \pm 0.05)\%$. The obtained data will be merged in the end with the measurements obtained from the $\Lambda_c^+ \rightarrow pK^+\pi^-$ decay performed by another analysis group. These results have been published in the [118] paper, together with the D^0 meson cross section which is fundamental to calculate baryon-to-meson ratios.
The contributions from each $\Sigma_c(2455)$ isospin state are combined in the analysis, whose production is studied for the first time in pp collisions at $\sqrt{s} = 13$ TeV at $|y| < 0.5$. Experimentally, the study of the baryon was performed using only the $\Sigma_c^{0,++}$ states, due to their decays having a charged pion, while the Σ_c^+ (udc) was added indirectly to the final results, assuming isospin symmetry.
For the purpose of clarity, from now on I will use the Σ_c symbol to refer to the sum of all the $\Sigma_c(2455)$ isospin states.

4.1 Data samples and event selection

The differential cross section of $\Sigma_c^{0,++}$, Λ_c^+ and Λ_c^+ feed-down from $\Sigma_c^{0,++}$ in pp collisions at $\sqrt{s} = 13$ TeV is measured using data collected by the ALICE Collaboration during Run 2. The Minimum Bias trigger was enabled when selecting the events, requiring matching signals in the V0 detector. Moreover, the same MC productions, anchored to the data-taking periods and dedicated to $\Lambda_c^+ \rightarrow pK_s^0$ decays, are used both to calculate the efficiency corrections and for the training procedure of the machine learning algorithm.
The primary vertices, reconstructed only by TPC+ITS tracks, are required to be $\pm$ 10 cm from the nominal interaction point and the pile-up contribution to the measurement is suppressed by rejecting triggered events with more than one reconstructed primary vertex.
The total number of selected events for the analysis is 1.82×10^9 that, considering an inelastic cross section in pp collisions at $\sqrt{s} = 13$ TeV of $\sigma_{\rm MB} \approx 57.8$ mb (from the measurements of the visible cross section by a set of detectors obtained in Van Der Meer scans of 2018 [119]), correspond to an integrated luminosity of $L_{\rm int} \approx 32$ nb^{-1}.

4.2 Analysis Strategy

Machine learning (ML) techniques are algorithms developed to *learn* to perform a task without being explicitly programmed. They found applications in multiple fields (such as medicine, speech recognition, etc.) and in the past years they became increasingly popular in high-energy physics, in particular for classification and regression problems [120].
A ML technique was used, in this thesis, to select $\Lambda_c^+ \rightarrow p + K_s^0$ decay candidates. The

baryons are reconstructed by combining a proton candidate track with a K^0_s identified via its characteristic V-shaped neutral decay topology (called V^0).

Variable	**Conditions**
$\|\eta\|$	< 0.8
p_T (GeV/c)	> 0.3
ITS refit	yes
TPC refit	yes
Number of TPC crossed rows	≥ 70
TPC clusters found / findable	≥ 0.8
Variable	**Conditions**
$\|\eta\|$	< 0.8
p_T (GeV/c)	> 0.3
ITS refit	yes
$q_{T,Arm-Pod}/\alpha_{Arm-Pod}$	> 0.15
Number of TPC crossed rows	≥ 70
TPC clusters found / findable	≥ 0.8

Table 4.1: Selection requirements of bachelor proton (top) and V^0 (bottom) tracks.

A series of selection requirements are applied to the tracks in order to build the Λ_c^+ decay candidates (Tab. 4.1) before running the ML classification, described in detail in the next section.
For the selection of K^0_s, a cut in the Armenteros-Podolanski[1] space is applied to remove contributions from Λ decays. The condition applied on this dissertation analysis is based on the ratio of $q_{T,Arm-Pod}$, which is the transverse momentum component of the positive decay particle relative to the direction of the V^0, and the α parameter.
Finally, the Λ_c^+ candidates are filtered using a 3σ selection both on the TPC compatibility and, if available, on the TOF PID responses for the bachelor track with the expected values for a proton.
Candidates of $\Sigma_c^{0,++}$ are reconstructed by combining the Λ_c^+ decays just selected, with invariant mass close to the nominal Λ_c^+ mass value, to negative and positive tracks (for Σ_c^0 and Σ_c^{++} respectively) selected as *soft-pion* candidates (reconstructing their charge

[1]Considering a two-body decay $M \to m_1 + m_2$, this method represents the transverse momentum as a function of the longitudinal momentum asymmetry α of the decay products, obtained by considering the following relation between the longitudinal momenta p_L of daughter particles

$$\alpha = \frac{p_{L_1} - p_{L_2}}{p_{L_1} + p_{L_2}}.$$

Hence, the decay is described with semi-ellipses, whose parameters provide information on the masses of the parent and child particles [121]. It was developed in 1954 specifically for V^0 decays, and it was introduced to study the separation of $K^0_s \to \pi^+\pi^-$ and the $\Lambda^0 \to p + \pi^-$ decay channels, from which it was possible to retrieve the masses of parents and daughter decay species by fitting the ellipses observed in data, without a pre-existent assumption on final state masses [122].

conjugates as well), thanks to the $\approx 100\%$ branching ratio of the $\Sigma_c^{0,++} \to \Lambda_c^+ + \pi^{-,+}$ decay. The conditions applied for the selection of the pions are shown in Tab. 4.2.

Variable	**Conditions**
$\lvert\eta\rvert$	< 0.9
p_{T} (GeV/c)	> 0.05
ITS clusters	≥ 3
Track-to-vertex distance in the xy-plane	< 0.065 cm
Track-to-vertex distance in the z-plane	< 0.15 cm

Table 4.2: Selection requirements for soft $\pi^{\pm}$ particles.

4.2.1 Boosted Decision Trees (BDT) configuration

Several machine learning techniques were developed for the analysis of high-energy physics collision events, among which Boosted Decision Trees (BDT) were adopted for this analysis, through the TMVA library of ROOT [123]. They belong to the family of supervised models that learn how to classify data starting from a subset of examples of which the true classification is already known: the training sample [124].
In particular, MC signal datasets are used in this case for the training and testing, by randomly sampling them with an equal amount of events for the two processes. The background of Λ_c^+ was estimated by considering samples of real candidates sitting in the two side-bands of the baryon invariant mass peak[2]. Our specific configuration parameters are described in Tab. 4.3, in which the number of trees and maximum depth of the decision tree are defined before the splitting is stopped. They correspond to the default TMVA settings, which should deliver very good performance without modifications. Adaptive boosting (AdaBoost) is used for the BDT classification, which means that signal events (or background) from the training sample ending up in the opposite class node are given a larger weight compared to events classified correctly, causing the creation of a re-weighed training event sample from which a new decision tree is developed. The boosting process is repeated several times until a set of new decision trees are generated (a forest). The weights w are calculated via

$$w = \left(\frac{1 - err}{err}\right)^{\beta}, \tag{4.1}$$

where err is the fraction of misclassified events in the tree, and β is a free parameter modifying the boosting [123].

Furthermore, the Gini index is used as a separation criterion in the node splitting [108], which is a number that defines the purity of a sample: the larger it is (maximum

[2]A Gaussian fit (χ^2) is performed to the true Λ_c^+ of the enriched Monte Carlo sample for each p_{T} interval and after the pre-selection cuts are applied. The side-bands contain candidates within $\pm 3\sigma$ from the obtained Λ_c^+ invariant mass peak.

Parameter	**Value**
Number of trees	850
Maximum depth	3
Boosting	AdaBoost with $\beta = 0.5$
Separation	Gini index

Table 4.3: Parameters configuration of the BDT algorithm.

value 0.5) the more the sample is well mixed between signal and background, while smaller values imply a good separation. Taking into account only the signal (s) and background (b) classes, the Gini index (G) of a sample can be obtained with:

$$\mathrm{G(sample)} = \frac{2sb}{(s+b)^2}. \tag{4.2}$$

Multiple input variables were considered for the classification:

- invariant mass, impact parameter with respect to primary vertex ($d_{0,\mathrm{V^0}}$), cosine of the pointing angle[3] and $c\tau$ of a $\mathrm{V^0}$ structure ($\mathrm{K^0_s}$ decay);
- impact parameter of the bachelor track with respect to the primary vertex ($d_{0,\mathrm{p}}$);
- cosine of the proton emission angle θ^* in the Λ_c^+ centre-of-mass system with respect to the baryon momentum direction;
- $n_{\sigma,\mathrm{TOF}}$(bachelor)[4] for the proton, pion and kaon mass hypotheses;
- $n_{\sigma,\mathrm{TPC}}$(bachelor)[4] for the proton, pion and kaon mass hypotheses.

Independent BDTs were trained for each of the p_T intervals in the analysis that was carried out in two different p_T width configurations: wide and fine intervals. The former was used for all the baryons and it was performed in the four p_T intervals $2 \leq p_\mathrm{T} < 4$ GeV/c, $4 \leq p_\mathrm{T} < 6$ GeV/c, $6 \leq p_\mathrm{T} < 8$ GeV/c and $8 \leq p_\mathrm{T} \leq 12$ GeV/c; on the other hand, the latter was used only for the inclusive Λ_c^+ analysis and it is the same employed for the measurements performed in the $\mathrm{pK^-\pi^+}$ channel, namely 10 p_T intervals going from 1 to 24 GeV/c [125]. The following tables will show only the wide configuration results, but a separate training was performed for the finer p_T intervals using the same procedures described further (same task for tree selection, pre-selection conditions, BDT variables, etc.).
To give an idea on the statistics needed for the BDT procedure, the numbers of training candidates for the signal and background characterisation of the Λ_c^+ baryons are illustrated in Tab. 4.4.

[3]It is defined as the angle between the reconstructed Λ_c^+ momentum vector and the line connecting the primary vertex to the Λ_c^+ decay point.

[4]Defined in Eq. 3.1.

p_T interval (GeV/c)	Training signal candidates	Training bkg candidates
2–4	412364	500000
4–6	276740	500000
6–8	123328	131259
8–12	77798	36914

Table 4.4: Number of signal and background (bkg) Λ_c^+ candidates processed for the training in different Λ_c^+ p_T intervals.

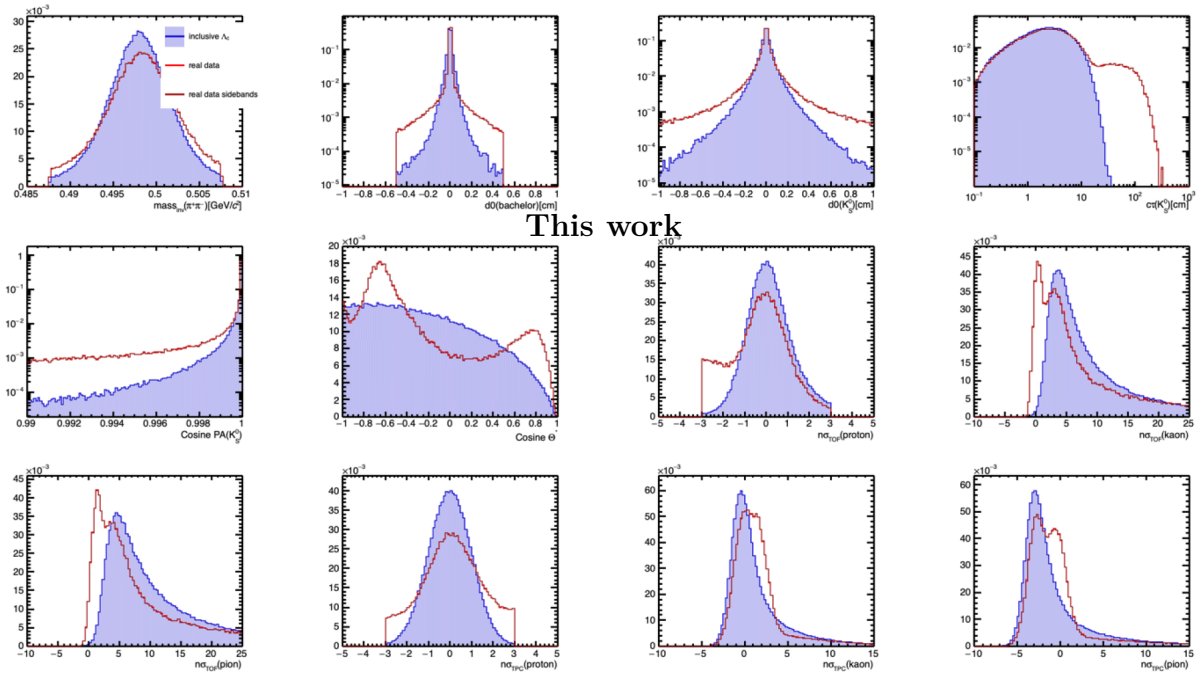
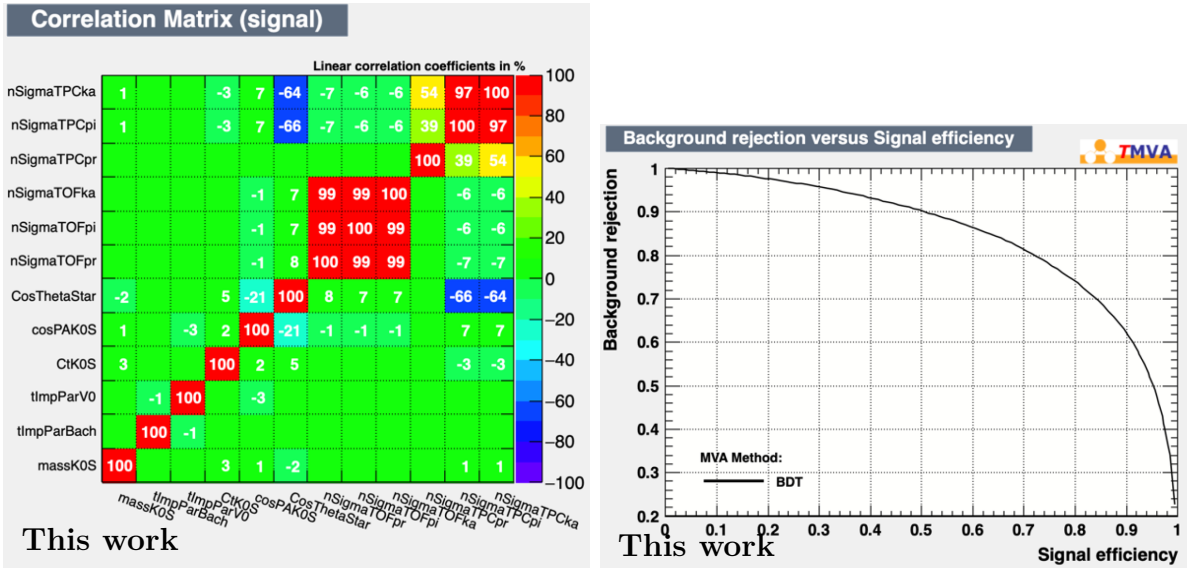


Figure 4.1: Distributions of BDT input variables in the $4 \leq p_T$ (Σ_c) < 6 GeV/c interval for MC signal (blue) and real data (red for the full Λ_c^+ invariant mass range and brown for candidates sitting in the two side bands around the Λ_c^+ invariant mass peaks. Top: V^0 invariant mass, bachelor impact parameter, V^0 impact parameter, ct of the V^0. Middle: cosine of pointing angle of V^0, $\cos\theta^*$, $n_{\sigma,\mathrm{TOF}}$(proton), $n_{\sigma,\mathrm{TOF}}$(kaon). Bottom: $n_{\sigma,\mathrm{TOF}}$(pion), $n_{\sigma,\mathrm{TPC}}$(proton), $n_{\sigma,\mathrm{TPC}}$(kaon), $n_{\sigma,\mathrm{TPC}}$(pion).

Figure 4.1 shows an example of the input variables distributions for the $4 \leq p_T < 6$ GeV/c interval used in the BDT analysis for simulated inclusive Λ_c^+ for the signal, and real data for the background characterisation.
A study was also performed on the input variables in order to avoid any possible linear correlation between each one of them and the Λ_c^+ invariant mass (improving in this way the ML algorithm efficiency). An example of the obtained results is shown in Fig. 4.2a for the signal obtained in the $6 \leq p_T < 8$ GeV/c interval (the background was studied separately with the same method). A high correlation is clear among the PID variables, which are also slightly correlated, as expected, to the $\cos\theta^*$, since these quantities are

both partially correlated on the p_{T} of the Λ_{c}^{+}. However, this has been demonstrated not to be an issue because BDT analyses and performances are not affected by the correlation between input variables.



(a) (b)

Figure 4.2: Linear correlation coefficients of the BDT input variables for signal (a) and ROC curve (b) for Λ_{c}^{+} candidates in the 6–8 GeV/c p_{T} range.

One of the first output of the BDT algorithm is the comparison between the training and the testing distributions of the classifier, since the input events are copied internally and split into one training and one testing ROOT trees. From the obtained results we concluded that the specific choice of a certain training sample does not introduce a bias in the output, given that the same output is obtained if the classification is applied to the testing sample.
The performances of the BDT classifier are evaluated via the Receiver Operating Characteristic (ROC) curves, obtained for all p_{T} intervals, which are expressed as the relative amount of background rejection as a function of the signal efficiency and provided as an output by the algorithm, as shown in the example of Fig. 4.2b.
Finally, the ranking of the BDT input variables is shown in Tab. 4.5 for two p_{T} intervals, where the scores are calculated by counting the number of times the variables are used to split the nodes of the decision tree (divided by the number of events), thus giving an indication on the importance of each parameter. For all the considered p_{T} intervals, the PID variables related to the bachelor track are in the top rankings (first 5 positions).
Eventually, the choice of the selections on the BDT output scores is performed by es-

timating the expected significance[5] on the training sample, and it is tuned in each p_T interval in order to maximise the expected statistical significance. This last quantity is based on the signal-over-background ratios and total number of candidates in the data for each p_T interval.

Rank	**Variable**	**Importance**	**Rank**	**Variable**	**Importance**
1	$n_{\sigma,\mathrm{TOF}}(\mathrm{p})$	0.1100	1	$n_{\sigma,\mathrm{TPC}}(\mathrm{p})$	0.1215
2	$n_{\sigma,\mathrm{TPC}}(\mathrm{p})$	0.1096	2	$n_{\sigma,\mathrm{TOF}}(\mathrm{p})$	0.1131
3	$d_{0,\mathrm{p}}$	0.1094	3	$d_{0,\mathrm{p}}$	0.1107
4	$\cos \mathrm{PA}_{\mathrm{V}^0}$	0.1070	4	$\cos\theta^*$	0.1067
5	$\cos\theta^*$	0.1050	5	$n_{\sigma,\mathrm{TPC}}(\pi)$	0.09925
6	$n_{\sigma,\mathrm{TPC}}(\mathrm{K})$	0.09452	6	$\cos \mathrm{PA}_{\mathrm{V}^0}$	0.09422
7	$n_{\sigma,\mathrm{TPC}}(\pi)$	0.08401	7	$n_{\sigma,\mathrm{TPC}}(\mathrm{K})$	0.08681
8	$c\tau(\mathrm{V}^0)$	0.07352	8	$n_{\sigma,\mathrm{TOF}}(\mathrm{K})$	0.06514
9	$n_{\sigma,\mathrm{TOF}}(\mathrm{K})$	0.06456	9	mass K_s^0	0.06286
10	$n_{\sigma,\mathrm{TOF}}(\pi)$	0.05950	10	$n_{\sigma,\mathrm{TOF}}(\pi)$	0.05075
11	d_{0,V^0}	0.04526	11	d_{0,V^0}	0.04948
12	mass K_s^0	0.03762	12	$c\tau(\mathrm{V}^0)$	0.03942

Table 4.5: Ranking of input variables after BDT classification for $4 \leq p_T < 6$ GeV/c (left) and $6 \leq p_T < 8$ GeV/c (right) intervals.

p_T (GeV/c)	BDT cut	$\lvert m(\Lambda_c) - m(\Lambda_c^{\mathrm{PDG}})\rvert$ (GeV/c^2)
2–4	0.02	≥ 0.025
4–6	0.03	≥ 0.018
6–8	0.04 (0.03)	≥ 0.018
8–12	0.04	≥ 0.030

Table 4.6: Offline cuts applied before fitting the invariant mass distributions for the $\Sigma_c^{0,++}$ (Λ_c^+) analysis.

The maximum value of significance is linked to a particular working point on the efficiency curve which can be easily derived by TMVA after the cut optimisation scan has converged. Both the signal efficiency and the background rejection curves as a function of the classifier output for each Λ_c^+ p_T interval are shown in Fig. 4.3.
Using this procedure, working points were selected as well as cuts on the Λ_c^+ invariant mass for each p_T interval, which are reported in Tab. 4.6 for the analysis of both $\Sigma_c^{0,++}$ and Λ_c^+ baryons[6].

[5]It is calculated by considering the amount of signal (S) and background (B):

$$\text{Significance} = \frac{S}{\sqrt{S+B}}$$

[6]The same cuts are applied both for the inclusive and for the $\Sigma_c^{0,++}$ feed-down productions.

The BDT algorithm is trained in the same way for both the analyses, but the cuts have some slight differences because in some cases they were shifted from the optimal working point, provided by the TMVA. This was done to avoid having a central value in the tail of the cut variation distributions due to statistical fluctuations (more on Sec. 4.5).

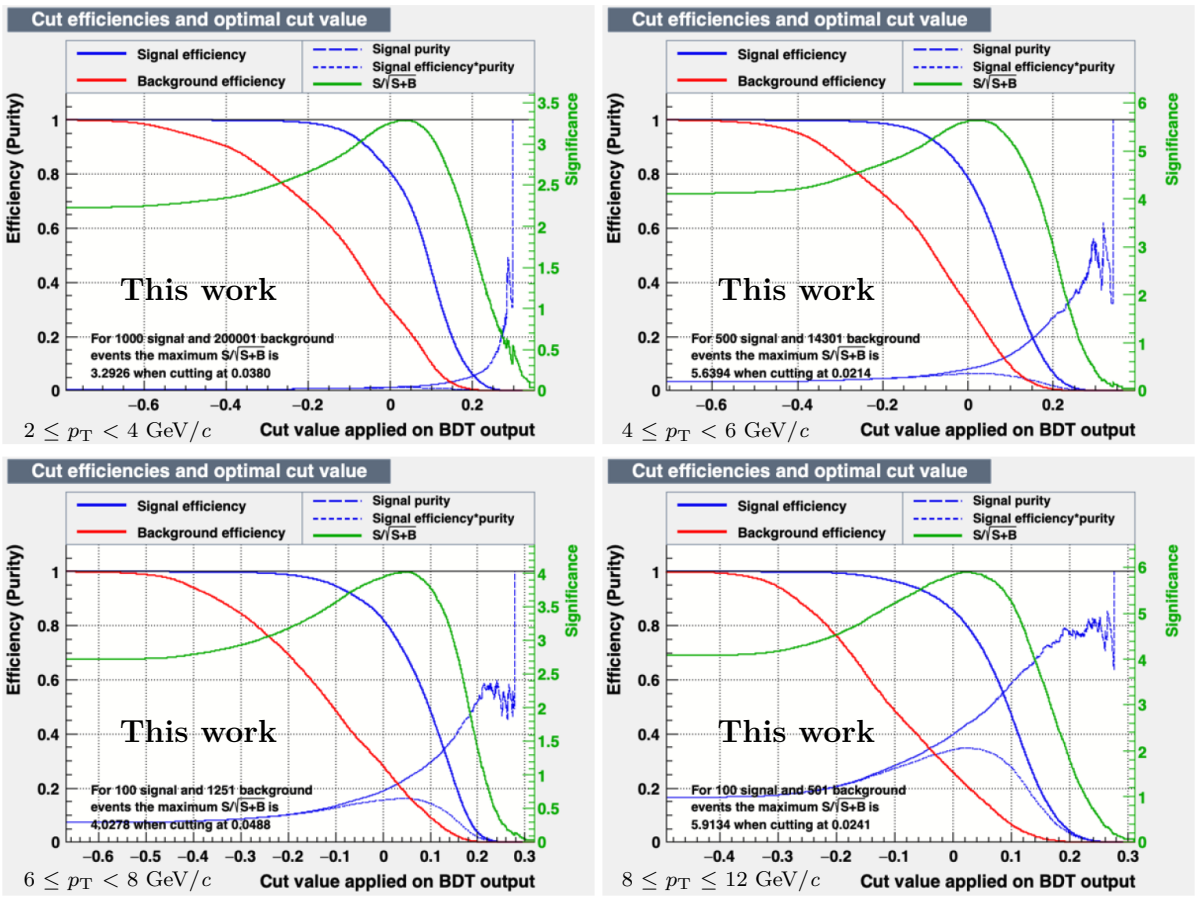


Figure 4.3: Signal efficiency, background rejection, purity and statistical significance as a function of the classifier output for each p_T interval of Λ_c^+ candidates.

4.3 Signal extraction

From the mass values reported by the PDG [4] and showed earlier, the difference in mass between the Σ_c and Λ_c baryons is ≈ 167 MeV/c^2, so both the $\Sigma_c^{0,++}$ and Λ_c^+ from $\Sigma_c^{0,++}$ decays are measured by counting the entries corresponding to the mass difference distribution $\Delta m = m(\Sigma_c) - m(\Lambda_c) = m(\mathrm{pK_s^0}\pi) - m(\mathrm{pK_s^0})$ as a function of the transverse momentum of the analysed baryon. Moreover, since the considered isospin states $\Sigma_c^{0,++}$ differ in mass for just $\delta m = 0.22$ MeV/c^2, the expected signal peak centered at ≈ 167 MeV/c^2 benefits from both of them. Taking this into account, the fit function chosen to

describe the invariant mass peak is

$$f(x) = \frac{C}{2}\left(\text{Voigt}\left(x-\mu,\sigma,\Gamma_{\text{BW}}\right) + \text{Voigt}\left(x-\mu+\delta m,\sigma,\Gamma_{\text{BW}}\right)\right) + f_{\text{bkg}}(x), \tag{4.3}$$

where two Voigt functions are used to implement the Σ_c^0 and Σ_c^{++} mass peaks. These are defined as a convolution of a Breit-Wigner, implementing the physics of the strong decay with a $\Gamma_{\text{BW}} = 1.89$ MeV, and a Gaussian, deriving from the finite resolution on the reconstructed daughter momenta.
On the other hand, the background was modelled through two functions:

- a product between a power law and an exponential function $\rightarrow f_{\text{bkg}}(x) = a \cdot (x - m_\pi)^b \cdot e^{-c(x-m_\pi)}$, for the [2,4] and [4,6] GeV/c p_{T} intervals;

- a third order polynomial function, otherwise.

The stability of the background fit function was tested by repeating the fitting procedure using different values of the fit range.
Using the values from the PDG, the Breit-Wigner Γ_{BW} and δm were fixed, as well as the width of the Gaussian function which is fixed to 1 MeV/c^2. All the other parameters are left free, together with the parameters of the background fit function.
The statistical significance of the signal is calculated as the inverse of the relative signal uncertainty (S/σ(S)), since the usual 3σ interval around the peak cannot be considered because the Breit-Wigner inside the Voigt function does not have a finite variance. This variable is larger than 3 in all the analysed p_{T} intervals for the invariant mass extraction of both the $\Sigma_c^{0,++}$ and $\Lambda_c^+ \leftarrow \Sigma_c^{0,++}$, whose measurements are shown in Fig. 4.4–4.5. The second column of plots in the figure contains the residuals, which are data subtracted by the fitted background.
In the plots, the signal-over-background ratio (S/B) is also indicated and it was estimated in the interval $\pm$ the full width at half maximum (FWHM) of the Voigt profile calculated as $f_{\text{FWHM}} \approx 0.5346 \cdot f_{\text{L}} + \sqrt{0.2166 \cdot f_{\text{L}}^2 + f_{\text{G}}^2}$ [126], where f_{G} is the FWHM of the Gaussian function $\left(f_{\text{G}} = 2\sigma\sqrt{2\ln(2)}\right)$, while $f_{\text{L}} = 2\Gamma_{\text{BW}}$.
On the other hand, for the Λ_c^+ inclusive analysis the invariant mass was fitted using a Gaussian distribution for the signal, while the background employed a 2$^{\text{nd}}$ order polynomial function for all the p_{T} bins except the last one in which a linear function was applied. The width of the Gaussian distribution is fixed to values obtained from Monte Carlo simulations to improve the stability of the fit and reduce statistical uncertainties. The invariant mass spectra in smaller p_{T} intervals (thanks to the higher yield available for the Λ_c^+) are shown in Fig. 4.6.

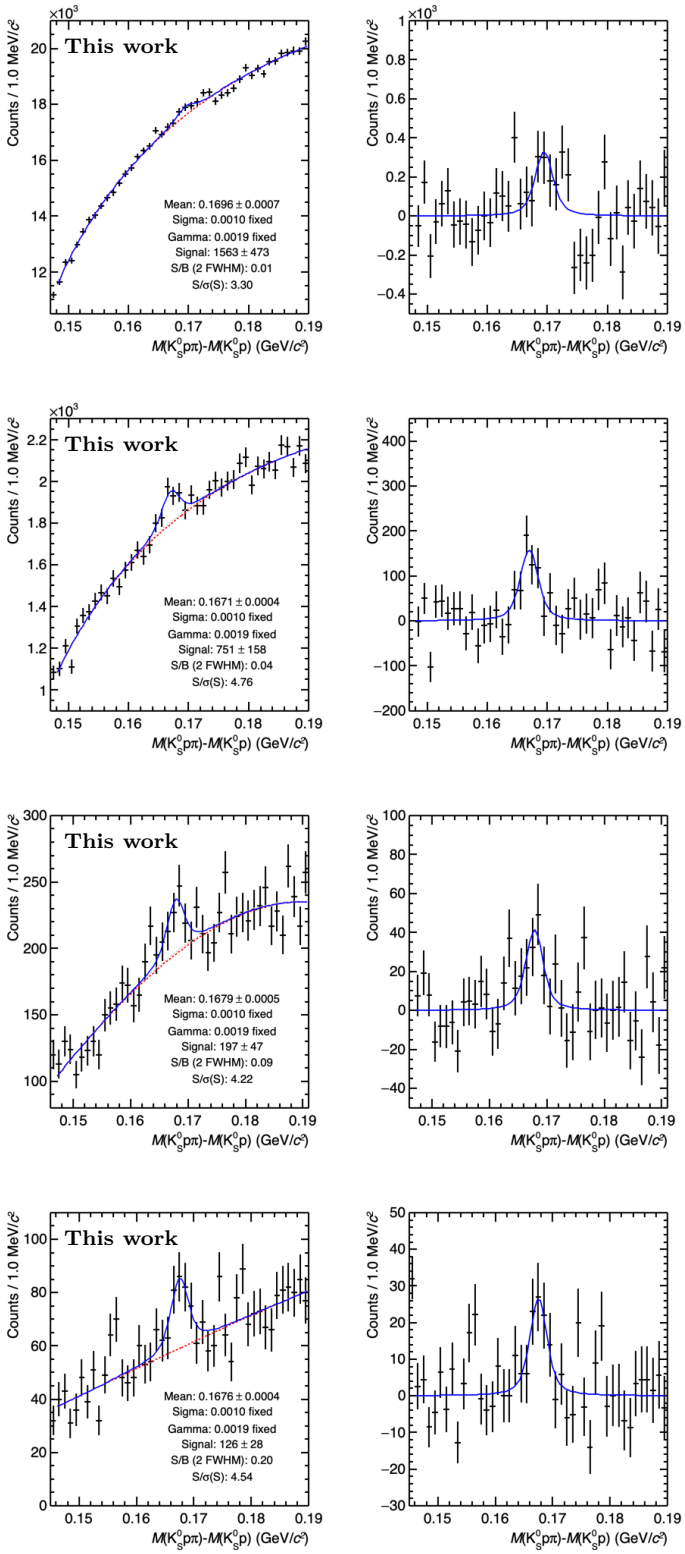


Figure 4.4: Invariant mass distributions of the $\Sigma_c^{0,++}$ baryon, showing the $m(\mathrm{pK_s^0}\pi) - m(\mathrm{pK_s^0})$ with the fitting function superimposed (left) and the residuals distribution (right) for all the considered p_T intervals.

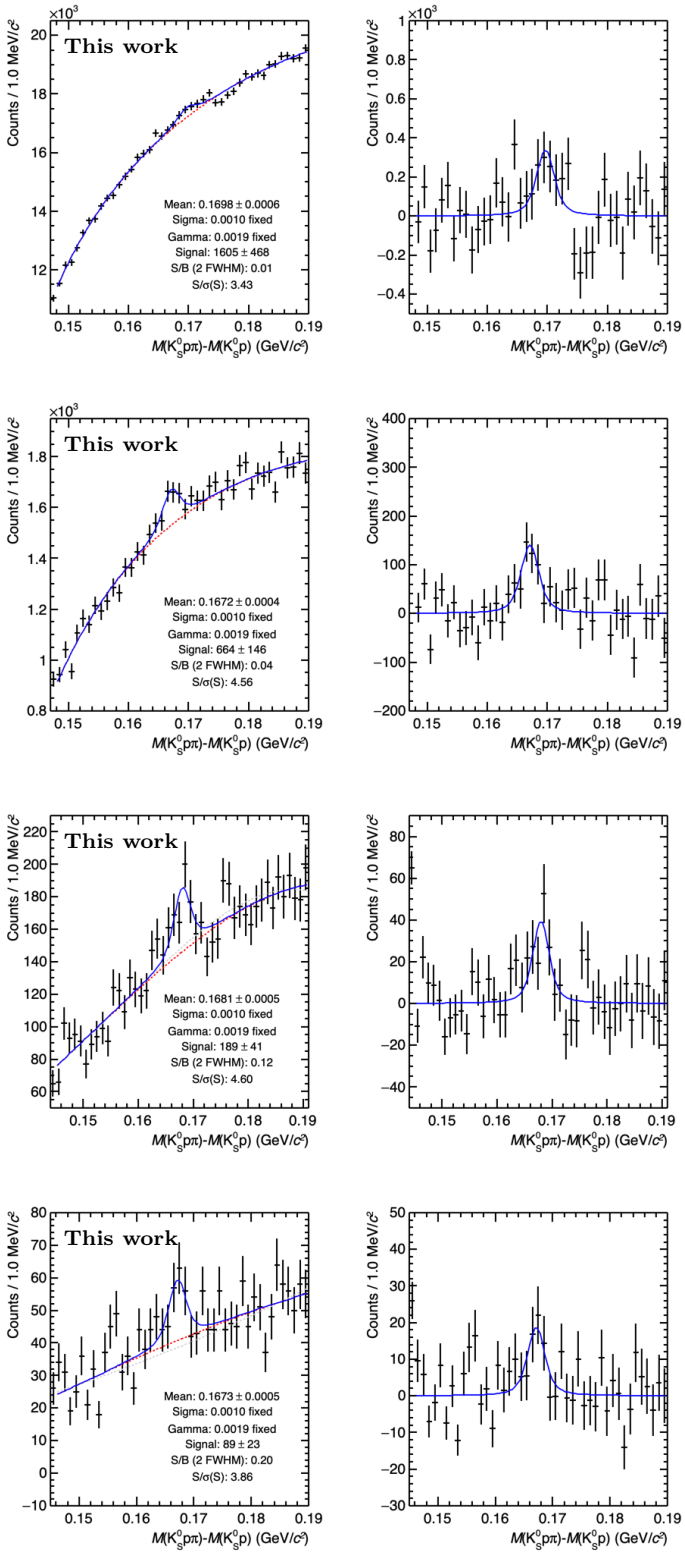


Figure 4.5: Invariant mass distributions of the $\Lambda_c^+ \leftarrow \Sigma_c^{0,++}$ baryon, showing the $m(\mathrm{pK_s^0}\pi) - m(\mathrm{pK_s^0})$ with the fitting function superimposed (left) and the residuals distribution (right) for all the considered p_T intervals.

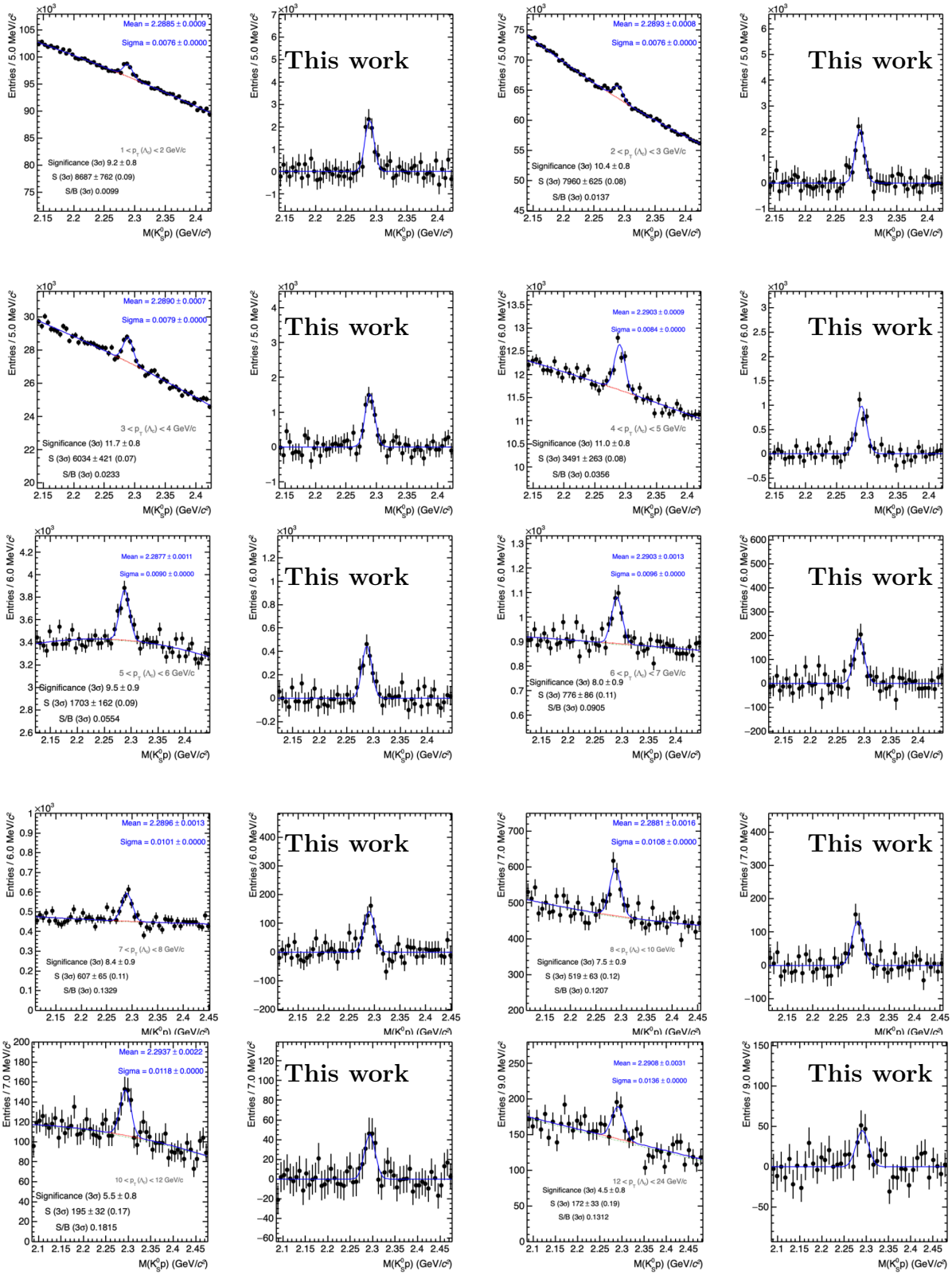


Figure 4.6: Λ_c invariant mass distributions and residuals with fitting functions superimposed in ten p_T intervals.

4.4 Corrections

Calculating the differential production cross section of the baryons requires taking into account a series of corrections dependent on the detector acceptance and reconstruction efficiency of the considered particles. Furthermore, since we are interested in extracting the prompt production, the contribution from beauty decays has to be subtracted.

4.4.1 Efficiency and detector acceptance

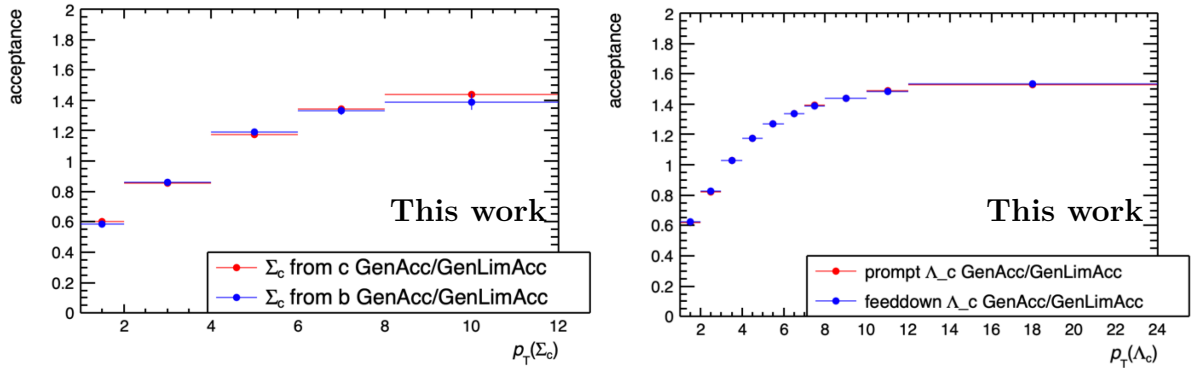
The prompt production yields are obtained by correcting the raw measured yields for the product of detector acceptance and reconstruction efficiency (Acc$\times\varepsilon$) for prompt Λ_c^+ and for the feed-down contribution from Λ_b decays. These corrections are obtained using enriched Monte Carlo simulations, in which the events were generated with PYTHIA8 (using Monash 2013 tune) simulations requiring the presence of at least one $c\bar{c}$ pair. The Monte Carlo datasets analysed are required to contain only events with at least one Λ_c^+ decaying in the p + K_s^0 hadronic decay channel. Furthermore, after the generation, the particles are propagated through the ALICE experiment via the GEANT3 package [110], implementing a detailed description of the apparatus layout and of the detector response. The simulation was configured to reproduce the conditions of the luminous region and of all the ALICE subsystems, in terms of active electronic channels, calibration level, and their evolution within the pp data-taking period.
Λ_c^+ or $\Sigma_c^{0,++}$ candidates are selected by applying a fiducial acceptance cut $|y| < y_{\rm fid}(p_{\rm T})$, with $y_{\rm fid}$ smoothly increasing from 0.5 to 0.8 in the interval $0 < p_{\rm T} < 5$ GeV/c and set to 0.8 above 5 GeV/c, assuming that the rapidity distribution of Λ_c^+ or $\Sigma_c^{0,++}$ is uniform in the considered rapidity ranges.
Two contributions take part in the calculation of the total efficiency: the efficiency of the pre–selection cuts, which takes into account all steps prior to the BDT application, and the offline efficiency, that accounts for the cut on the BDT output score, the required refitting of the soft-pion tracks in the ITS (ITS-refit) and the cut on the Λ_c invariant mass.
The acceptance is defined as the ratio between the number of generated candidates within the fiducial rapidity and with the daughters of the baryon decay in the acceptance of the central barrel of ALICE ($|\eta| < 0.9$), and the number of generated particles decaying in the desired channel within a $|y| < 0.5$ rapidity range, using the formula

$$\mathrm{Acc} = \frac{N_{\rm gen.}^{\rm daug.\ in\ acc.}\Big|_{|y|<y_{\rm fid}}(p_{\rm T}^{\rm gen.})}{N_{\rm gen.}\Big|_{|y|<0.5}(p_{\rm T}^{\rm gen.})}. \tag{4.4}$$

The efficiency ε is calculated as

$$\varepsilon = \frac{N_{\rm reco.}\Big|_{|y|<y_{\rm fid}}(p_{\rm T}^{\rm reco.})}{N_{\rm gen.}^{\rm daug.\ in\ acc.}\Big|_{|y|<y_{\rm fid}}(p_{\rm T}^{\rm gen.})}, \tag{4.5}$$

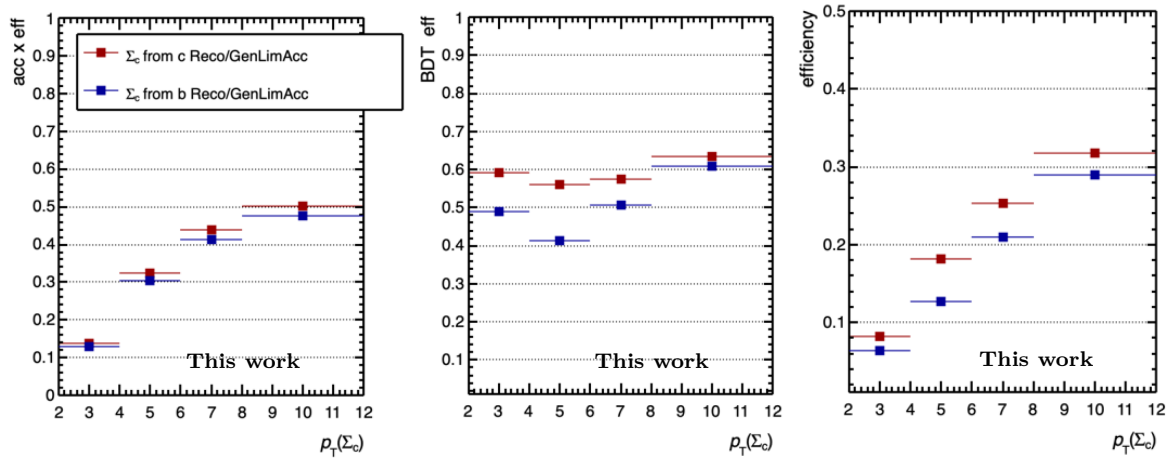
through the ratio between the number of reconstructed baryons within the fiducial acceptance and the generated baryons within the same rapidity interval and with the decay products in the $|\eta| < 0.9$ acceptance interval.



(a) (b)

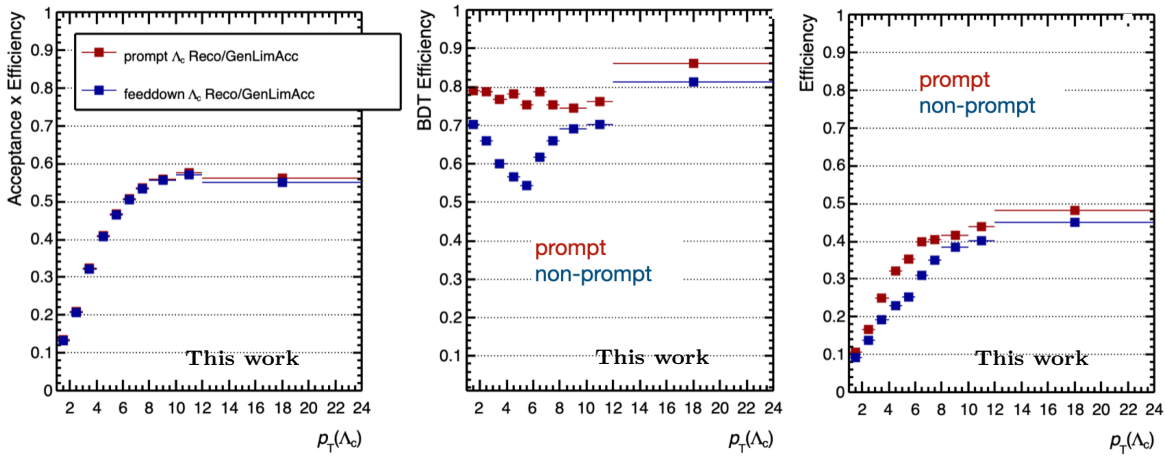
Figure 4.7: Geometrical acceptance of prompt and non-prompt $\Sigma_c^{0,++}$ (a) and Λ_c^+ with finer p_T intervals (b).

The obtained corrections for the $\Sigma_c^{0,++}$ selection are shown in Fig. 4.7a, illustrating the acceptance of the baryon from prompt and non-prompt decays, visually compared to the results for the inclusive Λ_c^+ in Fig. 4.7b. While, in Fig. 4.8 efficiencies calculated in the two steps and the combined final results are expressed as a function of p_T for both prompt and non-prompt $\Sigma_c^{0,++}$ production. The same variables are displayed for the Λ_c^+ using finer p_T intervals in Fig. 4.9.



(a) (b) (c)

Figure 4.8: Prompt and non-prompt $\Sigma_c^{0,++}$ pre-selection (a), offline (b) and total efficiencies (c).



(a) (b) (c)

Figure 4.9: Prompt and non-prompt Λ_c^+ pre-selection (a), offline (b) and total efficiencies (c) in ten p_T intervals.

4.4.2 Feed-down from b quarks

The raw yields extracted for the analysis of the baryons have contributions from both prompt ($c \to \Lambda_c$) and non-prompt ($\Lambda_b \to \Lambda_c$) candidates. Hence, this second contribution needs to be removed by considering only the Λ_c prompt fraction from charm quark hadronisation, which can be estimated by computing the feed-down fraction from beauty-hadron decays ($f_{\rm prompt} = 1 - f_{\rm non-prompt}$).
For the Λ_c^+, this term is calculated via the production cross section of feed-down charm baryons from the decay of beauty hadrons $\left(\frac{d\sigma^2}{dydp_T}\right)^{\rm theory}_{\rm feed\text{-}down}$, which is obtained through: the beauty p_T-shape from FONLL calculations, the beauty fragmentation functions to Λ_b^0 hadrons[7], and the $\Lambda_b^0 \to \Lambda_c + X$ kinematics modelled through PYTHIA8 simulations. The result is then normalised by the branching ratio of the $\Lambda_b^0 \to \Lambda_c + X$ decay channel, which is ≈82%, as in PYTHIA8.
Hence, the prompt fraction $f_{\rm prompt}$ is calculated as:

$$f_{\rm prompt} = 1 - \left(\frac{N^{\Lambda_c,\text{feed-down,raw}}}{N^{\Lambda_c,\text{raw}}}\right) \tag{4.6}$$

$$= 1 - \left(\frac{d\sigma^2}{dydp_T}\right)^{\rm theory}_{\rm feed\text{-}down} \frac{(\text{Acc}\times\varepsilon)_{\rm feed\text{-}down}\,\Delta y \Delta p_T \cdot \text{BR} \cdot \text{N}_{\rm evt}}{\text{N}^{\Lambda_c}_{\rm raw}/2} \tag{4.7}$$

whose result is used to obtain the differential cross-section of the analysed baryons. In the formula, the branching ratio (BR) of the decay is considered, as well as the number of events ($\text{N}_{\rm evt}$) and of the raw number of selected Λ_c^+ baryons ($\text{N}^{\Lambda_c}_{\rm raw}$), which is divided

[7] This is evaluated from b-hadron fraction measurements by LHCb in pp collisions at $\sqrt{s} = 13$ TeV [127].

by two to account for the selection of both particles and antiparticles. This procedure is commonly referred as 'N_b' method. The product of acceptance and efficiency for non-prompt baryons is also considered as a correction factor.
The only beauty hadron with a significant branching fraction to the Σ_c is Λ_b, which has a predicted probability for $\Lambda_b \to \Sigma_c$ decay that is 3% of the corresponding beauty feed-down Λ_c^+ decay. Hence, due to the lack of direct measurements of the non-prompt Σ_c baryons, the prompt fraction for the latter baryon was estimated using these assumptions.

4.5 Systematic uncertainties

All the sources of systematic uncertainties will be described in the following paragraphs, and a summary of the uncertainties applied to the final results will be shown.

4.5.1 Raw yield extraction

The systematic error on the yield extraction for $\Sigma_c^{0,++}$ and $\Lambda_c^+ \leftarrow \Sigma_c^{0,++}$ was estimated by repeating the fitting procedure of the invariant mass distribution under different conditions. Multiple variations to different parameters of the fit were considered, in particular:

- different functions to describe the background, specifically:
  - 3^{rd} and 4^{th} order polynomial functions for the [2,4] and [4,6] GeV/c p_T intervals;
  - 2^{nd}, 3^{rd} and 4^{th} order polynomial functions and a power law combined with an exponential function for the [6,8] and [8,12] GeV/c p_T intervals;
- five different values for the lower and upper limits of the fitted mass range;
- two different bin widths of the invariant mass $\to$ 0.5 and 1 MeV/c^2.

These different fitting variations were performed one at a time using:

- three values for the width of the Gaussian component in the Voigt function, namely 0.9, 1 and 1.1 MeV/c^2;
- three values for the Γ_{BW} in the Voigt function (1.71, 1.89 and 1.98 MeV/c^2);
- the bin-counting method;

making a total of 1050 trials, whose measurements are considered only if they have a significance > 2.5 and the reduced χ^2 is lower than two.

An example of results for one p_T interval is shown in the output plots of Fig. 4.10. This shows various histograms, which (from left to right) contain the raw yield for different trials (with central value specified with a red solid line and two dashed lines

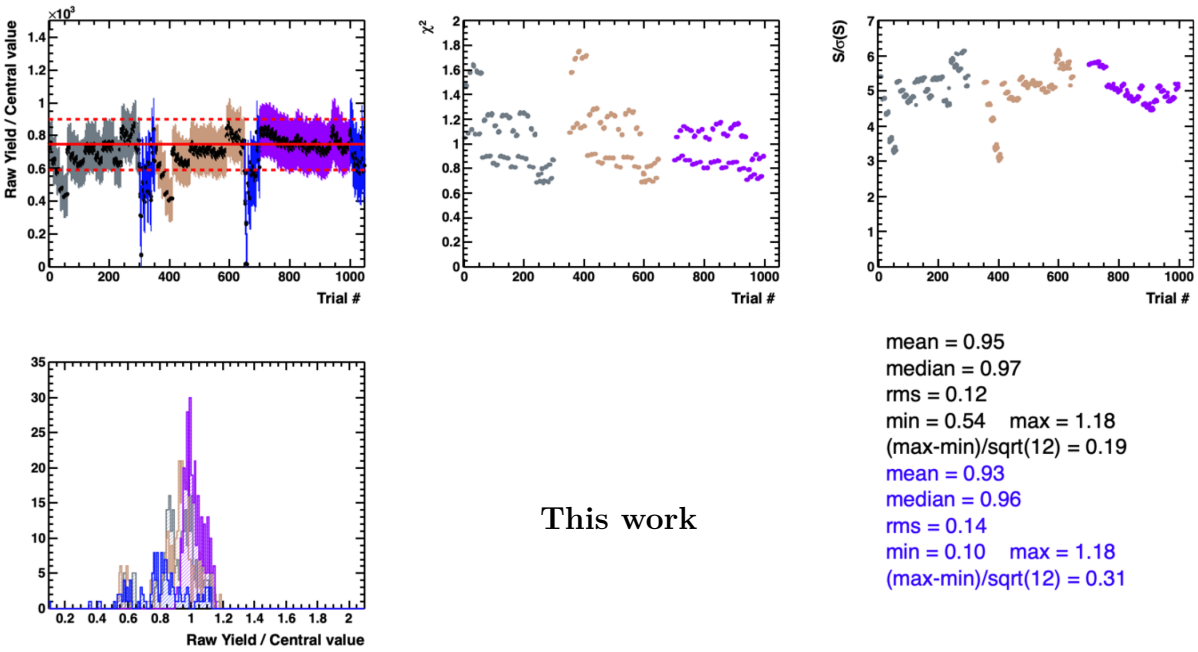


Figure 4.10: $\Sigma_c^{0,++}$ raw yield, reduced χ^2, significance and ratio of the extracted raw yields to the central value for the multitrial fitter in the p_T interval [4,6] GeV/c.

indicating the statistical uncertainty), the reduced χ^2, the significance of the obtained signal, and the distribution of the ratio of the extracted raw yields with the central reference value.
The root mean square (RMS) of the latter (without considering the bin counting that is only used as a cross-check[8]) is used to estimate the systematic uncertainty for all the p_T intervals, except the first one in which a 15% uncertainty was considered due to the values of the mean and the median being significantly shifted with respect to unity.
On the other hand, different conditions are applied to the systematic raw yield estimate of the Λ_c^+ inclusive analysis, even though the procedure is maintained the same. The fitting variables changed in this case are:

- three different functions to describe the background which are: 2^{nd}, 3^{rd} and 4^{th} order polynomial functions;
- five values for the lower/upper limit of the fitted mass range;
- three different bin widths of the invariant mass.

Each variation was tested six times, changing the fixed to MC (or unfixed) combinations of peak width and position, for a total of 1350 trials.

[8]It is referred to the blue points, histograms and statistics shown in Fig 4.10.

4.5.2 Cut variation

The systematic uncertainty on the efficiency correction due to imperfections in the MC simulation, for the description of the input variables used in the training of the BDT algorithm, was calculated by considering multiple variations of the selection on the output classifier. In addition, for a given cut on the BDT response, five different cuts on the Λ_c^+ invariant mass were applied: the central one and $\pm$ 10%, and 15%. The range of the cuts was chosen in order to scan a reasonably wide range in the BDT efficiencies. The systematic uncertainty is evaluated as the RMS of the distribution of the ratios of the BDT–corrected yields with the central value.

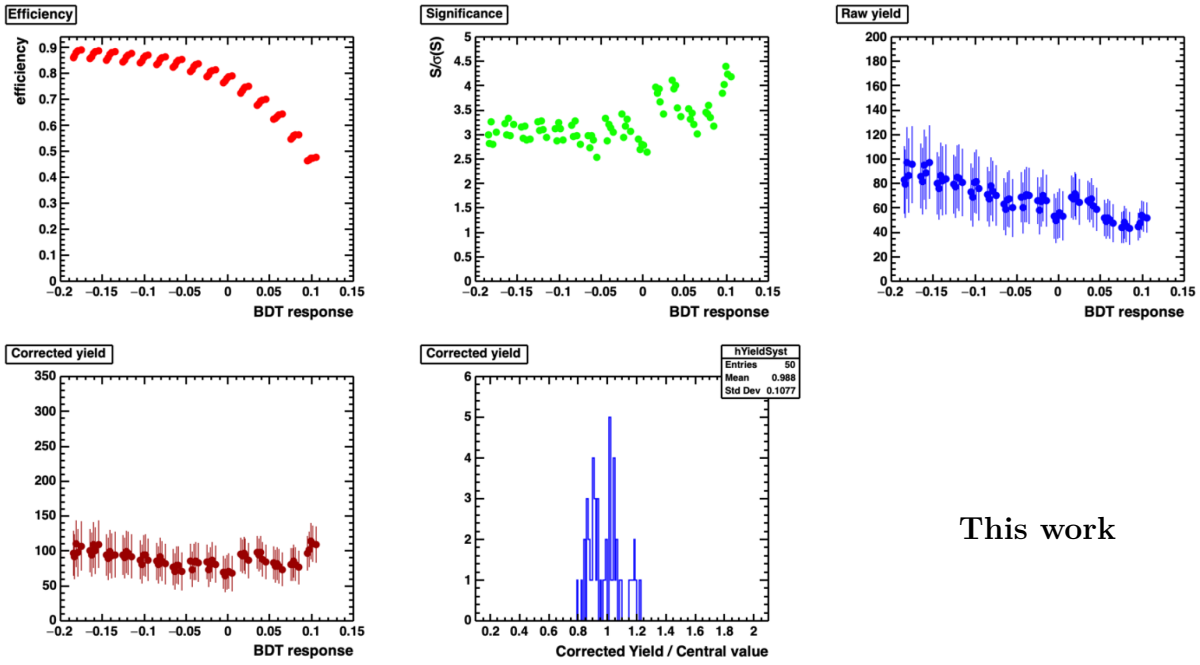


Figure 4.11: $\Lambda_c^+ \leftarrow \Sigma_c^{0,++}$ BDT efficiency, significance, raw yield, corrected yield and ratio of the corrected yields with the central one for different cuts on the BDT output score in the [8,12] GeV/c p_T interval.

The output provided by this multitrial cut variation procedure includes five histograms that are shown as an example in Fig. 4.11 for the $\Lambda_c^+ \leftarrow \Sigma_c^{0,++}$ in the highest p_T interval. They indicate (in order) the BDT efficiency, the significance, the raw yield, the corrected yield as a function of the BDT cut and the ratio between the corrected yield and the reference central value used for the analysis.
The signal extraction is very sensitive to statistical fluctuations, due to the low signal-to-background ratio, influencing the corrected yields distributions. Furthermore, the selection on the BDT output score selects the Λ_c candidates before looping on the soft-pion, thus, these uncertainties can be evaluated with a higher precision looking at the corrected yield distributions of inclusive Λ_c candidates where the signal extraction is less prone to statistical fluctuations thanks to the higher statistics.

4.5.3 ITS-TPC matching efficiency and TPC track quality

When analysing tracking efficiency, the effects arising from residual mismatches between data and Monte Carlo simulations in the track finding in the TPC, track propagation from TPC to ITS, and from quality selections must be taken into account and assigned as systematic uncertainties. They are estimated by comparing the corrected yields obtained with different TPC track selection cuts on the bachelor proton and the two charged pions, and by comparing the TPC-ITS track matching efficiency in data and simulations.
These uncertainties were evaluated in the study performed by the ALICE Collaboration on the multiplicity dependence in the p_{T}-differential charm baryon-to-meson ratios in proton-proton collisions at $\sqrt{s} = 13$ TeV paper [128]. Meanwhile, the corresponding uncertainty for the soft pions has been evaluated as not larger than $\approx 2\%$ for each p_{T} interval, and it is linearly summed to the previous systematic uncertainties.

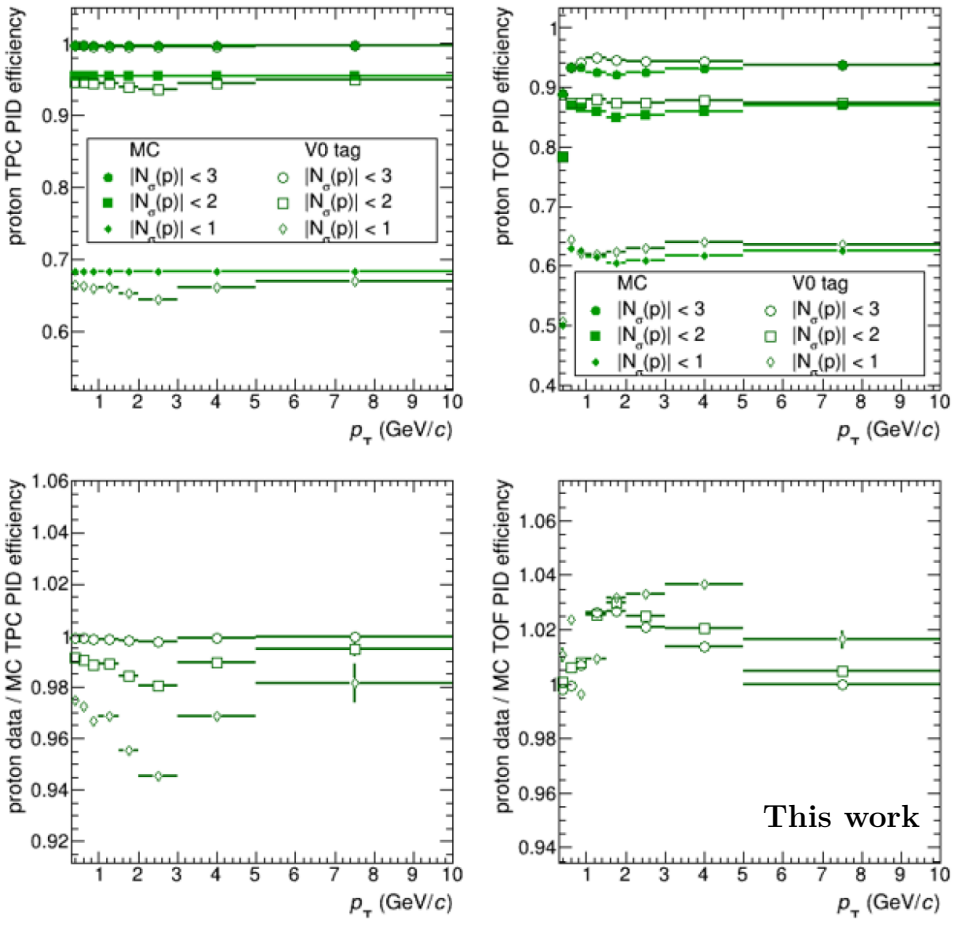


Figure 4.12: Proton PID efficiency using TPC and TOF detector for data and MC for three different $n\sigma$ cuts (top) and ratio between data and MC obtained efficiencies (bottom).

4.5.4 PID strategy in the pre-selection

The particle identification efficiency for protons is obtained by comparing MC estimates to experimental results with different σ cuts up to 3σ (using the V0 tagging for the experimental data).
A reference proton sample is obtained selecting the $\Lambda \rightarrow p\pi$ decay. Both TPC and TOF PID selection uncertainties are studied and the comparisons in Fig. 4.12 show how the proton σ cut is fairly described by simulations in both detectors. The bottom plots contain the ratio between data and Monte Carlo estimates and they are in agreement within $\approx 2\%$ using the 3σ selection, with Monte Carlo slightly overestimating the efficiency in TPC and slightly underestimating it in the TOF detector.
The obtained uncertainties are then propagated to Λ_c using the MC trees developed for the machine learning training, and the PID efficiencies obtained for the proton systematic calculation.

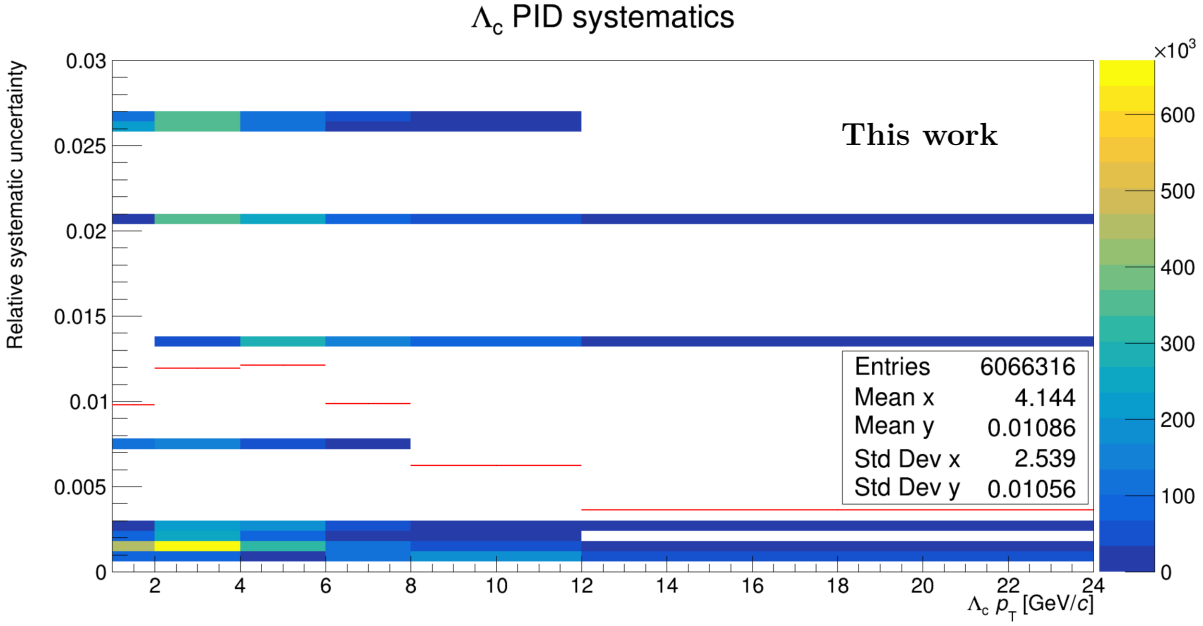


Figure 4.13: Inclusive Λ_c PID systematic uncertainty (relative error) as a function of p_T. The red lines are the final values obtained from the average of each bin.

All the protons coming from Λ_c^+ are taken into account and the 2D histogram shown in Fig. 4.13 is filled considering the efficiencies for each proton and linking them to the original baryons with their own p_T.
The uncertainties of events containing both TPC and TOF information are combined for the final results contained in the plot via the formula:

$$\sqrt{\left(\frac{|1-(\varepsilon_{\text{data}}^{\text{TPC}}/\varepsilon_{\text{MC}}^{\text{TPC}})|}{\varepsilon_{\text{data}}^{\text{TPC}}}\right)^2+\left(\frac{|1-(\varepsilon_{\text{data}}^{\text{TOF}}/\varepsilon_{\text{MC}}^{\text{TOF}})|}{\varepsilon_{\text{data}}^{\text{TOF}}}\right)^2}\varepsilon_{\text{data}}^{\text{TPC}}\varepsilon_{\text{data}}^{\text{TOF}} \tag{4.8}$$

where $\varepsilon^{\mathrm{TPC(TOF)}}_{\mathrm{data(MC)}}$ indicates the proton efficiency for TPC (TOF) using data (MC).
The combined systematic uncertainties for the 3σ cut are around 1% for all the p_{T} bins in all the performed analyses.

4.5.5 Generated Λ_c p_{T} shape

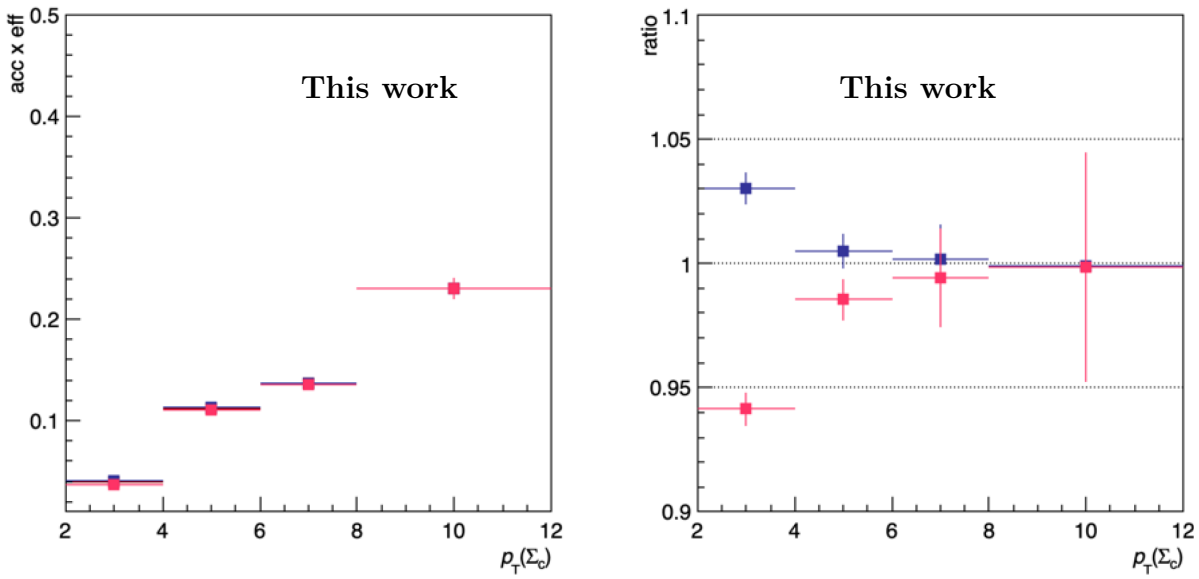


Figure 4.14: (left) $\Sigma_c^{0,++}$ efficiencies obtained with weights from PYTHIA8 Monash tune (blue) and enhanced colour reconnection Mode 2 (pink), and (right) ratio between the obtained reweighted efficiencies and the unweighted values as a function of p_{T}.

The efficiency in each p_{T} interval is sensitive to the generated MC p_{T} distribution which is used for computing the corrections, due to the p_{T} dependence of the Acc$\times\varepsilon$ and the finite p_{T} bin size of the analysis. This systematic effect is evaluated through the shift in the efficiencies after re-weighting the PYTHIA8 generated $\Sigma_c^{0,++}$ (or the Λ_c^+) p_{T} shape. The applied weights are defined as the ratio of the p_{T} distribution provided by simulations run with two different modes of the generator (Monash 2013 and enhanced colour reconnection Mode 2 [46]). Hence, the evaluation of these systematics depends on two different sets of weights, deriving from the usage of PYTHIA with the two considered tunes.
Finally, the uncertainties were computed as the ratio of the corrected efficiencies with the uncorrected ones (Fig. 4.14 for $\Sigma_c^{0,++}$), assigning the uncertainty to the maximum spread in each p_{T} interval with respect to unity.
The same procedure was applied to the inclusive Λ_c^+ analysis with finer p_{T} intervals, whose efficiencies comparison with and without reweighting and their ratio with the default obtained values are reported in Fig. 4.15.

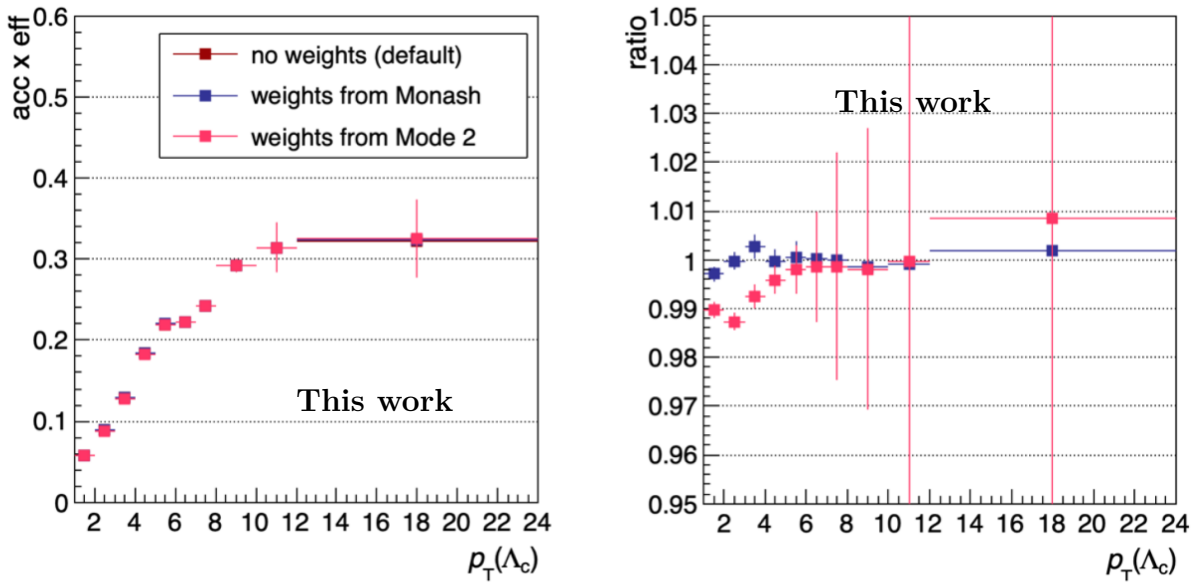


Figure 4.15: (left) Λ_c^+ efficiencies obtained with weights from PYTHIA8 Monash tune (blue) and enhanced colour reconnection Mode 2 (pink), and (right) ratio between the obtained reweighted efficiencies and the unweighted values as a function of p_T.

4.5.6 Feed-down subtraction

The systematic uncertainty on the prompt fraction is estimated by varying independently: the production cross section of beauty quarks within the theoretical uncertainties in FONLL, and the function describing the fragmentation fraction $f_{\Lambda_b^0}$, with the additional variation from 3% to 6% of the ratio of Λ_c feed-down calculated from PYTHIA8, as described in Sec. 4.4.2.

4.5.7 Systematic uncertainties: summary

The systematic uncertainties assigned to the $\Sigma_c^{0,++}$ and $\Lambda_c^+ \leftarrow \Sigma_c^{0,++}$ analyses in pp collisions at $\sqrt{s} = 13$ TeV as a function of p_T are summarised in Tabs. 4.7-4.8, which indicate the contributions coming from raw yield extraction, tracking efficiency and cut variation as the main sources of systematic uncertainty.
All the sources of systematic uncertainties contribute in a similar way to the results of both Λ_c^+ (inclusive and from $\Sigma_c^{0,++}$ decay) and $\Sigma_c^{0,++}$.
In addition, Tab. 4.9 collects a summary of all the systematic uncertainty contributions calculated for the inclusive Λ_c^+ measurements in finer p_T intervals, that will be combined later with the $pK^-\pi^+$ decay channel for the final differential cross section results.

p_T interval (GeV/c)	[2, 4] (%)	[4, 6] (%)	[6, 8] (%)	[8, 12] (%)
cut variation	7	7	10	10
raw yield extraction	15	13	13	13
tracking efficiency	8	8	9	9
PID	1	1	1	1
MC- p_T shape	6	2	1	1
feed-down (upper)	0.3	0.3	0.6	0.6
feed-down (lower)	0.9	1.2	2.5	2.5

Table 4.7: Summary of systematic uncertainties for the $\Sigma_c^{0,++}$ analysis.

p_T interval (GeV/c)	[2, 4] (%)	[4, 6] (%)	[6, 8] (%)	[8, 12] (%)
cut variation	7	7	10	10
raw yield extraction	15	10	10	10
tracking efficiency	8	8	9	9
PID	1	1	1	1
MC- p_T shape	5	2	1	1
feed-down (upper)	0.3	0.3	0.6	0.6
feed-down (lower)	0.9	1.2	2.5	2.5

Table 4.8: Summary of systematic uncertainties for the $\Lambda_c^+ \leftarrow \Sigma_c^{0,++}$ analysis.

p_T range (GeV/c)	[1, 2] (%)	[2, 3] (%)	[3, 4] (%)	[4, 5] (%)	[5, 6] (%)	[6, 7] (%)	[7, 8] (%)	[8, 10] (%)	[10, 12] (%)	[12, 24] (%)
cut var.	5	5	5	5	5	7	7	7	7	7
raw yield	8	8	8	8	8	8	8	8	11	14
tracking eff.	5	6	6	6	6	7	7	7	7	7
PID	1	1	1	1	1	1	1	1	1	1
MC-p_T shape	1	1	1	1	1	1	1	1	1	1
feed-down (+)	2	2	2	2	3	3	3	5	5	7
feed-down (-)	3	3	3	3	4	4	5	7	7	10

Table 4.9: Summary of systematic uncertainties for the Λ_c^+ inclusive analysis.

4.6 Results

The raw yields measured for the baryons, using the fitting procedure described earlier, are used for the calculation of the p_{T}-differential cross sections, according to the formula:

$$\left.\frac{\mathrm{d}\sigma^{\mathrm{H}}}{\mathrm{d}y\mathrm{d}p_{\mathrm{T}}}\right|_{|y|<0.5} = \frac{1}{2}\frac{1}{\Delta p_{\mathrm{T}}} \cdot \frac{f_{\mathrm{prompt}} N^{\mathrm{H}}_{|y|<y_{\mathrm{fid}}}}{c_{\Delta y}(\mathrm{Acc}\times\varepsilon)_{\mathrm{prompt}}} \cdot \frac{1}{\mathrm{BR}} \cdot \frac{1}{\mathcal{L}_{\mathrm{int.}}}, \tag{4.9}$$

where the 1/2 factor is introduced because baryons and antibaryons are not separated in the analyses, $N^{\mathrm{H}}_{|y|<y_{\mathrm{fid}}}$ is the signal measured with the fit functions (see Sec. 4.3) in the fiducial rapidity acceptance and different p_{T} intervals, with Δp_{T} width, and $c_{\Delta y}$ is a correction factor for the rapidity coverage [129].
The obtained cross sections are shown in Fig. 4.16, in particular for the prompt Λ_{c}^{+} in Fig. 4.16a, $\Sigma_{\mathrm{c}}^{0,++}$ in Fig. 4.16b and $\Lambda_{\mathrm{c}}^{+} \leftarrow \Sigma_{\mathrm{c}}^{0,++}$ in Fig. 4.16c. The measurements in the $\mathrm{pK}^{0}_{\mathrm{s}}$ decay channel are compared to the ones performed in the $\mathrm{pK}^{-}\pi^{+}$ channel [125] and they show compatible values, as visible in the zoomed ratios in the bottom parts of the plots. Both $\Sigma_{\mathrm{c}}^{0,++}$ and $\Lambda_{\mathrm{c}}^{+} \leftarrow \Sigma_{\mathrm{c}}^{0,++}$ measurements are multiplied by a factor 3/2 to account for the contribution of the Σ_{c}^{+}, assuming a symmetrical production of the three isospin states of Σ_{c}.
The final results [118] are the average of the two decay channels, increasing the precision of the measurements. The procedure uses the uncorrelated uncertainties between the two measurements to define the weigths used for the average calculation. Specifically, these are the statistical uncertainties, the uncertainty of the Acc$\times\varepsilon$ term, and among the systematic uncertainties those related to: the raw yield extraction, the selection efficiency correction, and the PID. The BR uncertainties are considered partially correlated, due to a linear correlation of 54% between the two decay channels of Λ_{c}^{+}.
The weights are calculated using the 2×2 covariance matrix between the two measurements

$$V_x\left(p_{\mathrm{T}}\right) = \begin{pmatrix} \mathrm{syst}^2_{\mathrm{pK}^-\pi^+}\left(p_{\mathrm{T}}\right) + \mathrm{stat}^2_{\mathrm{pK}^-\pi^+}\left(p_{\mathrm{T}}\right) & r \cdot \sigma_{\mathrm{pK}^-\pi^+}\left(p_{\mathrm{T}}\right)\sigma_{\mathrm{pK}^0_{\mathrm{s}}}\left(p_{\mathrm{T}}\right) \\ r \cdot \sigma_{\mathrm{pK}^-\pi^+}\left(p_{\mathrm{T}}\right)\sigma_{\mathrm{pK}^0_{\mathrm{s}}}\left(p_{\mathrm{T}}\right) & \mathrm{syst}^2_{\mathrm{pK}^0_{\mathrm{s}}}\left(p_{\mathrm{T}}\right) + \mathrm{stat}^2_{\mathrm{pK}^0_{\mathrm{s}}}\left(p_{\mathrm{T}}\right) \end{pmatrix} \tag{4.10}$$

where the statistical and uncorrelated systematic uncertainties are addressed, together with the covariance between the two cross sections in the out-of-diagonal terms via the correlation (r) of the BR uncertainties ($\sigma_{\mathrm{pK}^-\pi^+,\mathrm{pK}^0_{\mathrm{s}}}$). The weights are then calculated using the formula

$$w_i\left(p_{\mathrm{T}}\right) = \frac{\sum_k \left(V_x^{-1}\right)_{ik}}{\sum_{jk}\left(V_x^{-1}\right)_{jk}} \tag{4.11}$$

whose values appear in the average cross section calculation of the H baryon, namely

$$\left\langle \sigma_{\mathrm{H}}\left(p_{\mathrm{T}}\right)\right\rangle = \sum_i w_i\left(p_{\mathrm{T}}\right)\sigma_{\mathrm{H},i}\left(p_{\mathrm{T}}\right). \tag{4.12}$$

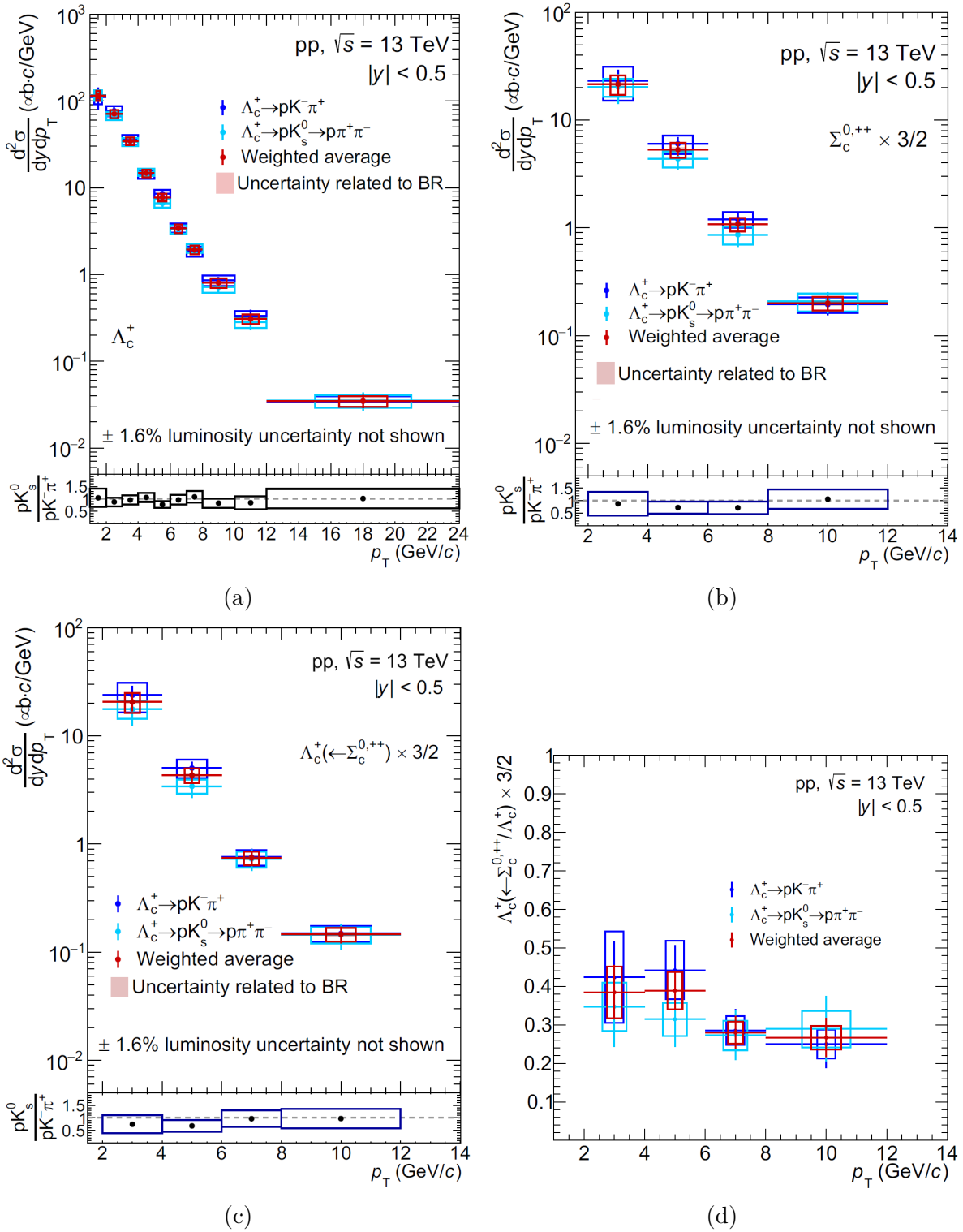


Figure 4.16: Differential production cross section of prompt Λ_c^+ (a), $\Sigma_c^{0,++}$ (b), $\Lambda_c^+ \leftarrow \Sigma_c^{0,++}$ (c), and ratio between Λ_c^+ ($\leftarrow \Sigma_c^{0,++}$) and Λ_c^+ results (d) [125]. The measurements in (b), (c) and (d) are scaled by a 3/2 factor to account for the three isospin states of Σ_c. The p+K_s^0 channel results of this dissertation are compared to the $pK^-\pi^+$ decay channel results and combined in the [118] publication.

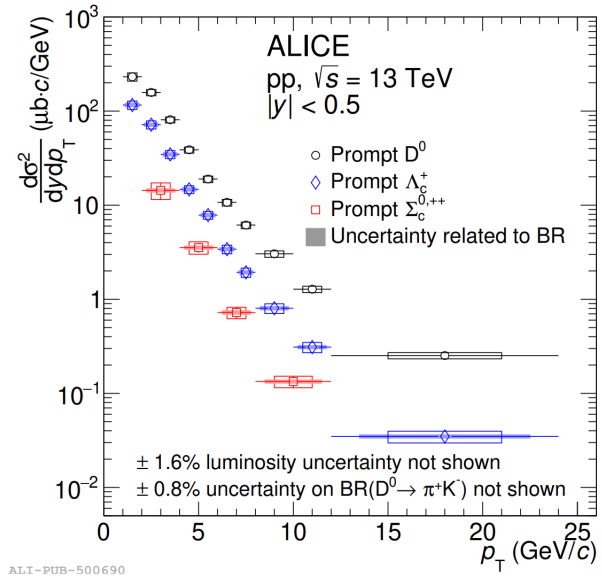


Figure 4.17: Cross section of prompt $\Sigma_c^{0,++}$, Λ_c^+ and D^0 in pp collisions at $\sqrt{s} = 13$ TeV [118].

The same procedure was followed to calculate the weighted average of the Λ_c^+ ($\leftarrow \Sigma_c^{0,++}$) to Λ_c^+ ratio shown in Fig. 4.16d in red (as for the baryons cross sections). Performing the ratio separately for the $\Lambda_c^+ \rightarrow pK^-\pi^+$ and pK_s^0 decay channels allows us to cancel the uncertainties related to the sources assumed as correlated, hence this procedure was done before calculating the average.
The systematic uncertainties are calculated accordingly to the correlation of the systematic source between the two decay channels [125] and the final systematic uncertainties contributions to the average results are calculated through the quadrature sum of the different sources values, which are considered uncorrelated by definition. The BR uncertainty is shown aside in the plots of Fig. 4.16 as a shaded box.
The final cross section measurements are compared with the D^0 cross section obtained in the same collision system [118] in Fig. 4.17. These last values are useful to calculate the baryon-to-meson ratios in Fig. 4.18, showing the comparison between Λ_c^+/D^0 (left), Σ_c/D^0 (centre) and $\Lambda_c^+ \leftarrow \Sigma_c$ $/\Lambda_c^+$ (right) experimental results and various MC models described in Chap. 2.
Systematic uncertainties are determined in different ways depending on their correlation degree between the two species considered in the ratios (most of the uncertainty sources are considered uncorrelated). In particular, the tracking and luminosity ones are considered as correlated and they cancel partly and completely, respectively. The feed-down uncertainty is propagated as partially correlated, depending on the contribution source: FONLL uncertainties are cancelled in the ratio due to a partial correlation, while the uncertainty on the Λ_c^+ feed-down fraction for the calculation of $f_{\rm prompt}$ of $\Sigma_c^{0,+,++}$ is propagated as uncorrelated in the ratio.
All the calculated baryon-to-meson ratios show a decreasing trend with p_T, differently

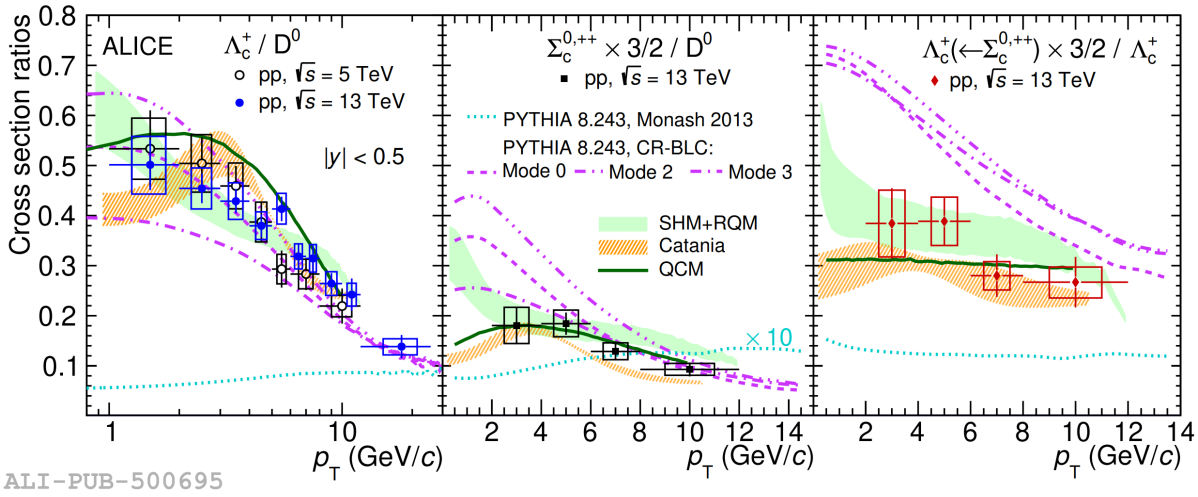


Figure 4.18: Comparison of Λ_c^+/D^0 (left), $\Sigma_c^{0,+,++}/D^0$ (centre) and $\Lambda_c^+ \leftarrow \Sigma_c$ $/\Lambda_c^+$ (right) with several MC models, including PYTHIA8 with various tunes [118].

from the results obtained from MC event generators based on measurements in e^+e^- [130] and ep [131] collisions, which are much flatter in p_T. Furthermore, all the ratios in pp collisions at $\sqrt{s} = 13$ TeV are significantly larger than the e^+e^- and ep results.
From the $\Lambda_c^+ \leftarrow \Sigma_c^{0,+,++}/\Lambda_c^+$ ratio, the Λ_c^+ feed-down from $\Sigma_c^{0,+,++}$ is quantified in the $2 < p_T < 12$ GeV/c interval to be 0.38 ± 0.06(stat.) ± 0.06(syst.). This is an important result because it confirms that $\Sigma_c^{0,+,++}$ has a significant influence on the enhancement observed in the Λ_c^+ production. Analysing each ratio in detail, the Λ_c^+/D^0 ratio measurements are in agreement within uncertainties with Catania, SHM+RQM and PYTHIA8 Mode 2 and 3 predictions, similarly to results obtained in pp collisions at 5 TeV [132].
Besides, PYTHIA8 with Monash largely underestimates ALICE results, but is able to properly describe measurements from e^+e^-, ep and B-factories, which suggest a Λ_c^+/D^0 ratio closer to 0.12 [130, 131], that is up to 5 times smaller than in hadronic collisions for low p_T.
The enhancement in hadronic collisions is even larger (up to a factor 10) when the $\Sigma_c^{0,+,++}/D^0$ ratio is considered, taking in consideration a value of the ratio ≈0.02 in e^+e^- collisions at $\sqrt{s} = 10.52$ GeV[9], on which Monash was tuned.
All the enhanced baryons production models are able to describe well the results, except PYTHIA8 mode 3 that overestimates data.
On the other hand, PYTHIA with enhanced colour reconnection is significantly overestimating the Λ_c^+ feed-down from $\Sigma_c^{0,+,++}$, as shown in Fig. 4.18 (right), while the other models (except Monash tune) are overall in agreement with the data within the uncertainties.
The final published results were also translated into a RIVET plugin (see App. A.1) that

[9]It is calculated by multiplying the Λ_c^+/D^0 result in e^+e^- collisions to the $\Sigma_c^{0,+,++}/\Lambda_c^+$ ratio evaluated from BELLE measurements [133].

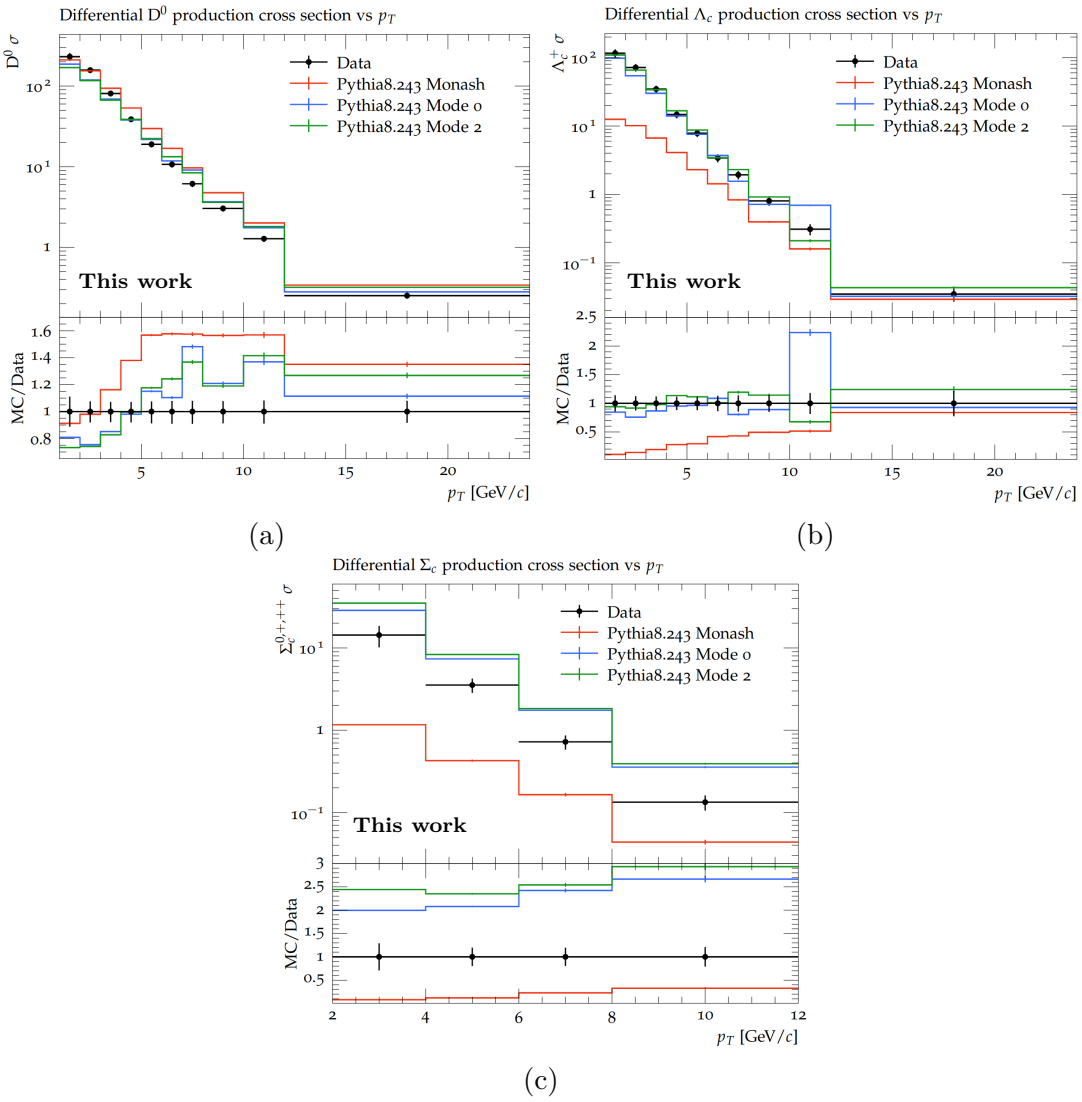


Figure 4.19: Comparison of experimental D^0 (a), Λ_c^+ (b) and $\Sigma_c^{0,+,++}$ (c) cross sections with PYTHIA 8 using different tunes.

was validated by the ALICE Collaboration MC analysis group. This was used to perform some further comparison of the experimental data with various tunes of PYTHIA8 and HERWIG. The cross section measurements shown earlier, are compared with PYTHIA8 estimates in Fig. 4.19. It is possible to see the different behaviours of the tunes when a meson (D^0) or a baryon (Λ_c^+, $\Sigma_c^{0,+,++}$) is considered. In particular, PYTHIA8 with Monash is coherent with the enhanced colour reconnection tunes for the description of the D^0 cross section (Fig. 4.19a), while it largely underestimates those of Λ_c^+ and $\Sigma_c^{0,+,++}$. The predictions of the enhanced colour reconnection modes in PYTHIA are consistent with the Λ_c^+ cross section, but they overestimate the production of Σ_c, even though the $\Sigma_c^{0,+,++}/D^0$ ratio is properly addressed as shown in Fig. 4.18 (centre).

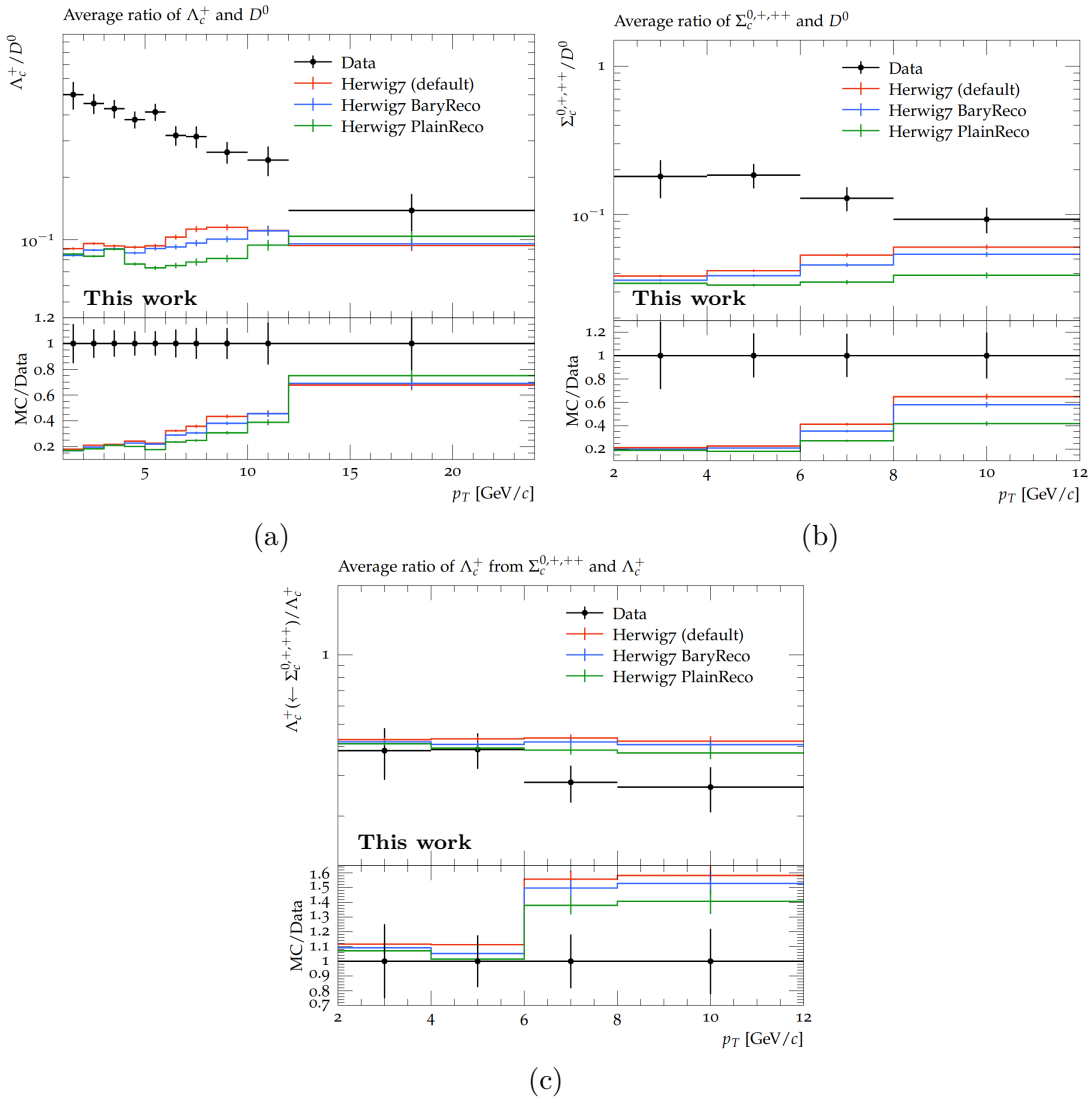


Figure 4.20: Comparison of Λ_c^+/D^0 (a), $\Sigma_c^{0,+,++}/D^0$ (b) and $\Lambda_c^+ \leftarrow \Sigma_c^{0,+,++}/\Lambda_c^+$ (c) with HERWIG7 using the default tune, baryonic, and plain reconnection models.

To conclude, a similar comparison was performed with HERWIG7, whose results are shown in Fig. 4.20. The baryon-to-meson ratios are underestimated by all considered tunes of the generator (for both Λ_c^+ and $\Sigma_c^{0,+,++}$), but predictions seem to be compatible instead with the ratio of Λ_c^+ from $\Sigma_c^{0,+,++}$ decays with the prompt Λ_c^+ production.

4.7 Summary

The measurement of the $\Sigma_c^{0,++}$ production cross section in pp collisions at $\sqrt{s} = 13$ TeV was obtained (first result in hadronic collisions), together with the measurement of the Λ_c^+ and $\Lambda_c^+ \leftarrow \Sigma_c^{0,++}$ using the $\Lambda_c^+ \rightarrow p+K_s^0$ decay channel. The results were combined with the measurements performed in the channel $\Lambda_c^+ \rightarrow pK^-\pi^+$, that is the other decay in which the baryon is studied in the ALICE Collaboration. In addition, the Σ_c^+ baryon was included in the measurement by assuming that all isospin states of Σ_c contribute in the same way to the production, hence a factor 3/2 was included in the calculation.
The $\Sigma_c^{0,+,++}/D^0$ and Λ_c^+/D^0 baryon-to-meson ratios, together with the Λ_c^+ feed-down from $\Sigma_c^{0,+,++}$ measured via the ratio $\Lambda_c^+ \leftarrow \Sigma_c^{0,+,++}/\Lambda_c^+$, were found to be larger than measurements obtained in e^+e^- and ep collision systems.
The comparison of the experimental results with Monte Carlo estimates puts important constraints on various models, which, via additional hadronisation mechanisms, aim at describing the enhancement of the heavy-flavour baryons production confirmed by the measurements discussed in this chapter. At the moment several event generators were consistent with data within the measured experimental uncertainties, explaining the observed increase of charm baryons in a parton-rich environment through enhanced colour reconnection (PYTHIA8), coalescence mechanisms with lighter particles (Catania and QCM), or by assuming decays from higher-mass baryon states that are yet to be observed (SHM+RQM). However, event generators tuned only on e^+e^- data are clearly disfavoured, especially at low p_T.

Chapter 5

Study of D mesons production as a function of multiplicity and spherocity

The analysis of charmed mesons as a function of multiplicity is a valuable tool to understand the role of MPI in pp collisions, since a higher production of particles can be connected to the interaction of multiple partons in pp collisions. This underlying mechanism can also be studied as a function of spherocity, a variable that allows the characterisation of the event in terms of its "jet-likelihood" as it will be discussed, letting us tag the events where a hard scattering (generating jets) took place. Both observables give insights on the QCD processes happening during the collisions and leading to the production of hadrons.
In this chapter, the average prompt production of D mesons as a function of multiplicity and spherocity in different p_{T} intervals in pp collisions at $\sqrt{s} = 13$ TeV will be studied, discussing, in particular, the production of D^{*+} that I personally analysed. All the procedures listed in this chapter are similarly followed for the D^{+} and D^{0} mesons, which have been analysed by other members of the ALICE Collaboration.

5.1 Data samples and events selection

The Run 2 datasets were used to measure the D mesons production, as done in the analysis described in the previous chapter. So a total of approximately 1.82×10^{9} minimum bias events were analysed. For the multiplicity only (spherocity integrated) case, an additional set of events was considered, which was selected using the High Multiplicity trigger provided by the SPD detector (HMSPD), for a total of ≈ 300 millions events.
Furthermore, for the calculation of corrections several Monte Carlo datasets anchored to the Run 2 data-taking period are used, implementing an enriched heavy-flavour production. This means that a pair of $\mathrm{c\bar{c}}$ or $\mathrm{b\bar{b}}$ is specifically required in each pp collision event during the simulation. In addition, three general purpose Monte Carlo samples are

exploited to determine the correlation between the number of tracklets and the number of charged particles (as it will be explained in Sec. 5.2.1).

5.2 Definition of the observables

The procedures followed to compute the observables used to classify the production of D mesons will be introduced in this section. The selections used and the discussion on how they were obtained are important to understand the results that will be shown later.

5.2.1 Multiplicity

The multiplicity estimator used in these analyses is related to the number of reconstructed SPD tracklets within a pseudorapidity range of $|\eta| < 1$. Tracklets are segments built from the reconstructed position of the primary vertex and two subsequent hits on the two layers of the SPD detector as shown in Fig. 5.1, so the estimation is limited by the efficiency of the detector itself [134].

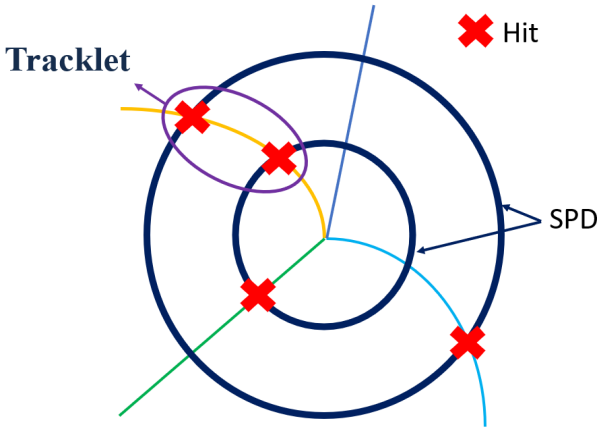


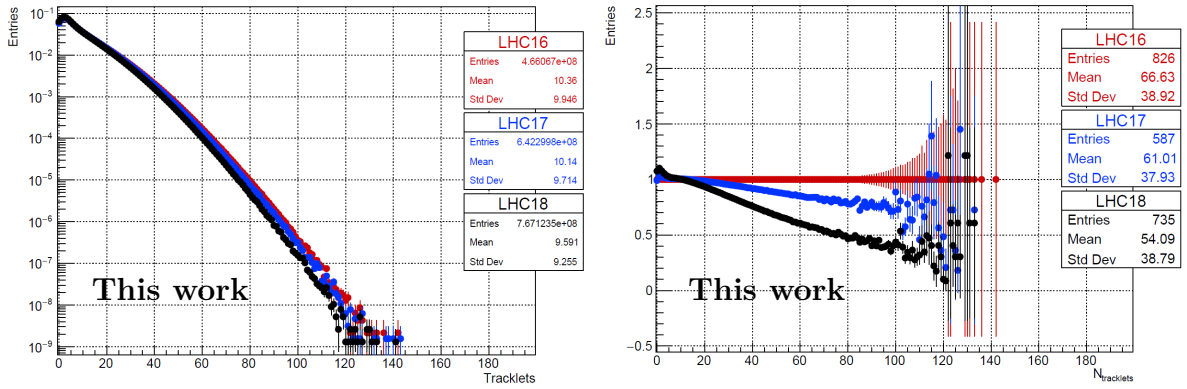
Figure 5.1: Cartoon explaining the definition of tracklet used in the multiplicity estimation.

Various multiplicity intervals were studied in the analysis (Tab. 5.1), starting from a minimum of one tracklet in order to consider only inelastic collision events with at least one charged particle (INEL>0 event class). However, before using the multiplicity values obtained from the detector, the number of SPD tracklets (N_{trk}) needs to be corrected due to the acceptance times efficiency having a significant z-vertex dependence, taking into account the different status (number of pixels actually turned "on") of the SPD pixels in the different years of Run 2 data-taking.
The acceptance of the detector has decreased slowly over time, with the aging of the pixels, and this is observed on the raw distributions of the N_{trk} in the three years of Run 2 data-taking, shown in Fig. 5.2, especially in the ratio of data divided by the 2016 trend, which is the one with the highest efficiency.

SPD tracklets	Trigger
1-8	MB
9-13	MB
14-19	MB
20-30	MB
31-59	MB
60-99	HMSPD

SPD tracklets	Trigger
20-30	MB
31-81	MB
20-81	MB

Table 5.1: Multiplicity intervals as number of SPD tracklets for measurements with spherocity integrated (left) and for the analysis as a function of spherocity (right).



(a) (b)

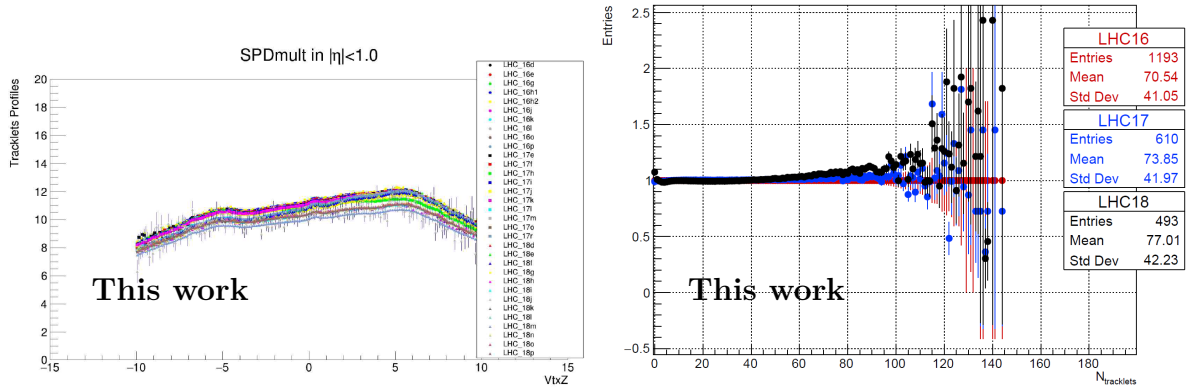
Figure 5.2: N_{trk} distributions per year of data-taking (a) and ratios of these distribution with the 2016 curve (b).

The raw N_{trk} profiles can also be plotted as a function of the z-vertex position, from which a clear dependence is observed (Fig. 5.3a). The profiles are used on event by event basis to correct the measured multiplicities in each data-taking period (identified by the year and a letter).
The raw distribution of the number of tracklets (N_{raw}) is corrected event by event to equalise the number of tracklets among the three different periods and to correct the dependence on the z-vertex using the formula:

$$N^{\mathrm{corr}} = N_{\mathrm{raw}} \frac{N_{\mathrm{ref}}}{\langle N_{\mathrm{period}} \rangle}, \tag{5.1}$$

where the numerator of the fraction is the average amount of tracklets in a z-vertex position in the LHC16h period, which is considered as a reference because it has the largest SPD acceptance, and it is equal to 12.25 and 55.02 respectively for MB and HMSPD triggered events, while the denominator is the average number of tracklets in the various data-taking periods.
The correction procedure exploits Poissonian statistics to obtain an integer value of N^{corr} starting from the value obtained in Eq. 5.1.

The corrections are shown in Fig. 5.3b, in which the ratio of the distributions with respect to the 2016 dataset is illustrated. Comparing this plot with the previous distributions shown in Fig. 5.2b one can clearly notice the trends for the three years of data-taking are now compatible within the errors after the correction procedure, especially below 60 N_{trk}, which is the interval used for the analysis with MB triggered data. A similar result is obtained for the HMSPD triggered events.



(a) (b)

Figure 5.3: N_{trk} distributions per year of data-taking as a function of z-vertex position (a) and corrected distribution ratios with 2016 trend as a reference (b).

The distributions of the number of tracklets (normalised for the number of events) for MB and HMSPD triggered events with D mesons candidates for the 2018 datasets are compared in Fig. 5.4, where the shift of the distribution peak to higher values of N_{trk} is noticeable using the high-multiplicity triggered dataset.

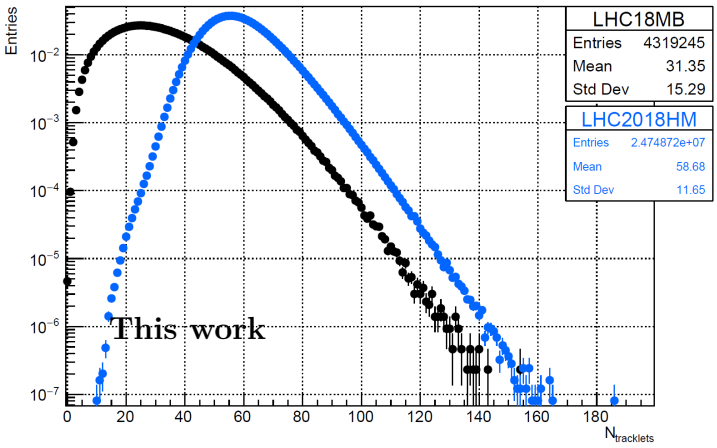


Figure 5.4: Comparison of HMSPD- (blue) and MB-triggered (black) number of tracklets distributions for 2018 data.

However, the SPD multiplicity estimator is not a physical observable and it is detector dependent, being linked to the functionality of the SPD detector. Thus, a conversion

procedure is mandatory to obtain a physical observable that can be used outside of the ALICE experiment: the number of primary charged particles produced in the event N_{ch}. The conversion is performed via the general purpose Monte Carlo simulations after the z-vertex correction of the tracklets distributions, applied in the same way as experimental data. A 2D plot is then created with the number of tracklets within $|\eta| < 1$ as a function of the amount of charged particles, within the same pseudo-rapidity interval, and it is displayed in Fig. 5.5, in which a linear correlation is visible.

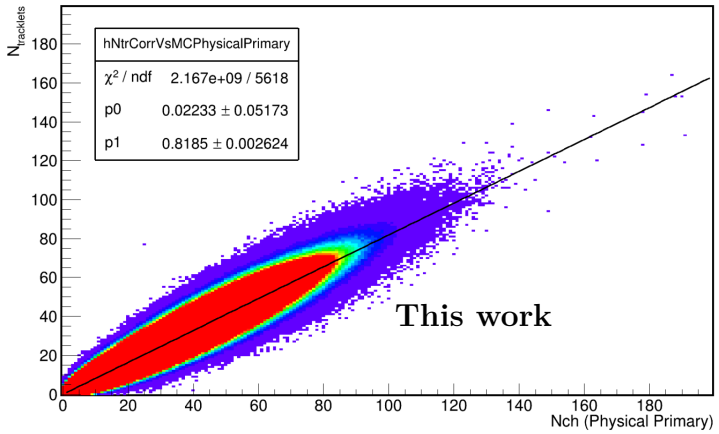


Figure 5.5: Two-dimensional plot of the number of tracklets as a function of the number of charged particles. A linear fit is applied.

The obtained correlation plot allows one to obtain the root mean square (RMS), maximum and minimum value of the $\mathrm{d}N_{\mathrm{ch}}/\mathrm{d}\eta$ distributions for each N_{trk} interval. To do so, the 2D histogram shown in Fig. 5.5 is projected in each multiplicity interval in order to obtain the distributions shown in Fig. 5.6.

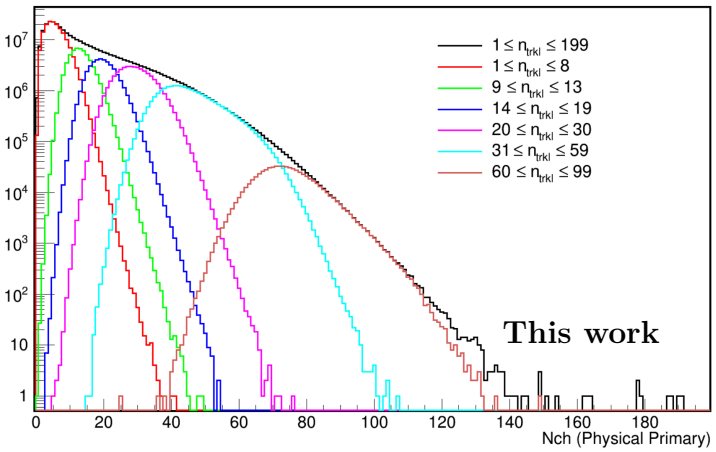


Figure 5.6: Projection in N_{trk} intervals of the 2D histogram in Fig. 5.5.

The minimum and maximum values (used for estimating one of the feed-down systematic uncertainties, as discussed in detail in Sec. 5.6.4) are extracted from a global

first-order polynomial fit to a profile histogram weighted by the RMS, whose offset and slope are then used to convert the $N_{\rm trk}$ to $N_{\rm ch}$.
The final conversion values are in the form of the ratio $\langle \mathrm{d}N_{ch}/\mathrm{d}\eta\rangle_{\rm trk}/\langle \mathrm{d}N_{ch}/\mathrm{d}\eta\rangle_{\rm MB}$ (charged-particle density) and are shown in Tab. 5.2, in which MB indicates the 1-199 $N_{\rm trk}$ interval (integrated minimum bias).
The conversion has not yet been performed for the analysis as a function of spherocity, but it will be introduced in the final published results.

Trackl.	$\langle N_{ch}/\mathrm{d}\eta\rangle$	RMS	[min,max]	$\langle \mathrm{d}N_{ch}/\mathrm{d}\eta\rangle_{\rm trk}/\langle \mathrm{d}N_{ch}/\mathrm{d}\eta\rangle_{\rm MB}$
$[1-199]$	7.42 ± 0.02	6.49 ± 0.00	-	Val $\pm$ stat $\pm$ syst
$[1-8]$	2.87 ± 0.02	1.46 ± 0.02	$[0.55, 4.86]$	$0.39 \pm 0.003 \pm 0.03$
$[9-13]$	6.59 ± 0.04	1.68 ± 0.06	$[5.47, 7.93]$	$0.89 \pm 0.006 \pm 0.01$
$[14-19]$	9.88 ± 0.06	2.04 ± 0.07	$[8.55, 11.62]$	$1.33 \pm 0.009 \pm 0.01$
$[20-30]$	14.59 ± 0.1	2.78 ± 0.09	$[12.23, 18.38]$	$1.97 \pm 0.014 \pm 0.02$
$[31-59]$	22.64 ± 0.17	4.45 ± 0.06	$[19.23, 36.21]$	$3.05 \pm 0.024 \pm 0.03$
$[60-99]$	37.77 ± 0.47	4.55 ± 0.12	$[36.83, 60.8]$	$5.10 \pm 0.065 \pm 0.03$

Table 5.2: Conversion to charged particle density values of number of tracklets multiplicity intervals.

5.2.2 Event shape: spherocity

In hadronic collisions, event shape observables (aiming to separate "soft" and "hard" events, and, within events, a larger or softer underlying event activity) measure the deviation of energy flow of events from jetty-like to isotropic structures, defined in terms of the $p_{\rm T}$ geometrical distribution of charged hadrons in the final state.
In this dissertation, the D mesons were studied using the transverse spherocity $S_{\rm O}$, which is generally formulated as

$$S_{\rm O} = \frac{\pi^2}{4} \min_{\vec{n}=(n_x,n_y,0)} \left(\frac{\sum_i |\vec{p}_{\rm T_i} \times \hat{n}|}{\sum_i p_{\rm T_i}} \right)^2 . \tag{5.2}$$

For each event, the spherocity is obtained by finding the minimum of the ratio in the equation by looping over steps of the transverse unit vector $\hat{n}$ (360 steps in the ϕ angle). In particular, during each iteration of the loop, the numerator is obtained by doing the sum of the vector product between $\hat{n}$, in that specific step, and the $p_{\rm T}$ of single charged tracks selected within $|\eta| < 0.8$ and $p_{\rm T} > 0.15$ GeV/c ($p_{\rm T_i}$) [135].
Spherocity has been a valuable tool for discriminating between jet-like and isotropic events associated with underlying event activity, which is either suppressed or enhanced [134, 136]. Thanks to the normalisation factor $\pi^2/4$, transverse spherocity value runs from 0 to 1, implying

$$S_{\rm O} = \begin{cases} 0 & \textit{pencil}\text{-like jet limit (hard events)}, \\ 1 & \text{isotropic limit (soft events)}. \end{cases} \tag{5.3}$$

To better understand this concept, an illustration of two events with opposite extreme values of spherocity is shown in Fig. 5.7.

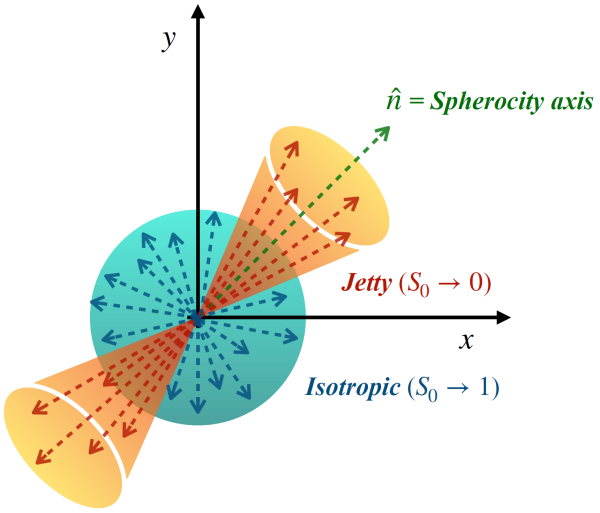


Figure 5.7: Cartoon explaining the definition of spherocity illustrating the jet-like and isotropic cases [137].

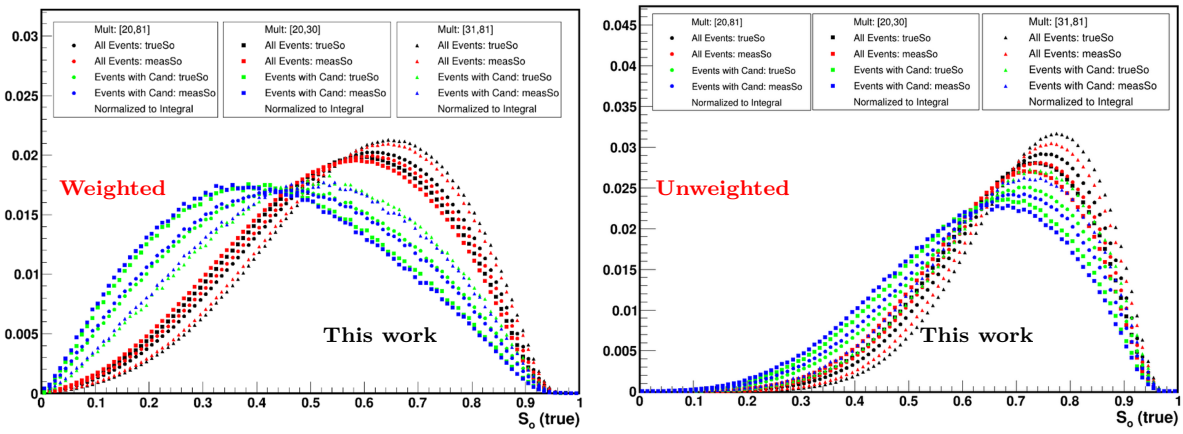
However, this term may be subject to p_{T} bias of the particle tracks, so an unweighted-p_{T} definition of spherocity is introduced and used for the D mesons analysis.
The observable is computed similarly to Eq. 5.2, but considering only the angular component of the transverse momenta of the tracks, namely

$$S_{\mathrm{O}}^{p_{\mathrm{T}}=1.0} = \frac{\pi^2}{4} \min_{\vec{n}=(n_x,n_y,0)} \left(\frac{\sum_i |\hat{p}_{\mathrm{T_i}}| \times \hat{n}|}{N_{\mathrm{tracks}}} \right)^2 . \tag{5.4}$$

The distributions of the two definitions of spherocity, normalised to their integral values, are shown in Fig. 5.8. The three multiplicity intervals considered for the analysis are included in the plots, which show also the measurements obtained with the true simulated spherocity and the reconstructed values on the Monte Carlo dataset. In addition, the distributions were also obtained both from considering all the events and only the ones containing D^+ candidates (it is assumed that the distributions are very similar between the three D mesons). From the measurements, one can notice that the unweighted spherocity distribution peak is shifted more towards isotropic events compared to the weighted definition, but also that, using the former, the reconstructed and true spherocity values for the selected D meson events are in general more compatible with the inclusive measurements. As it will be shown later, moving to this kind of observable improved the reconstruction efficiency for the analysis.
Due to the shape of the unweighted spherocity distribution, the lower spherocity intervals are wider than the higher ones, in order to have enough statistics to be able to perform the analyses. In particular, the four spherocity[1] intervals considered are: [0,0.5], [0.5,0.65],

[1]From now on *spherocity* or S_{O} will indicate the unweighted definition, unless specified otherwise.

[0.65,0.8], [0.8,1].



(a) (b)

Figure 5.8: Comparison between measured and true spherocity of events using Monte Carlo simulations applying the weighted (a) and unweighted (b) definition of the variable in different multiplicity intervals.

5.3 Analysis strategy

The D^{*+} (and charge conjugate) charmed meson ($c\bar{d}$) has a mass of (2010.26 $\pm$ 0.05) MeV/c^2 and it is reconstructed via the $D^{*+} \rightarrow D^0\ \pi^+ \rightarrow K^-\pi^+\pi^+$ decay channel which has a combined branching ratio of (2.67 $\pm$ 0.02)% [4]. Similarly to the $\Lambda_c^+\ \pi^-$ decay channel, the pion from the D^{*+} decay is referred as *soft* because it has a limited phase space available in the decay. Due to the fact that the meson decays strongly, its decay length is only of few picometers ($c\tau$ = 2.37 pm), hence it is not possible to separate its decay vertex from the primary one. For this reason, the selections are mainly based on the topology of the D^0 decay, which is simpler to reconstruct due to a better separation of the decay and primary vertices (few hundreds micrometers).
Before proceeding to the kinematic reconstruction of the specific decay channel, general track quality selections are applied to all the tracks to be considered as daughters of the studied decay (similarly to the cuts applied before the BDT selection on the $\Sigma_c^{0,++}$) and they are shown in Tab. 5.3.
All the variables used for the topological selection of the D^{*+} meson are shown in a subset of p_T intervals in Tab. 5.4. This selection is spherocity and multiplicity independent, but it is applied as a function of the reconstructed transverse momentum. The p_T-dependence comes from the fact that at low transverse momentum the contribution from combinatorial background is large, while at high p_T the statistical significance of the D^{*+} signal is relatively high before applying topological cuts. The distance of closest approach (DCA) between D^0 and the reconstructed secondary vertices, the cosine

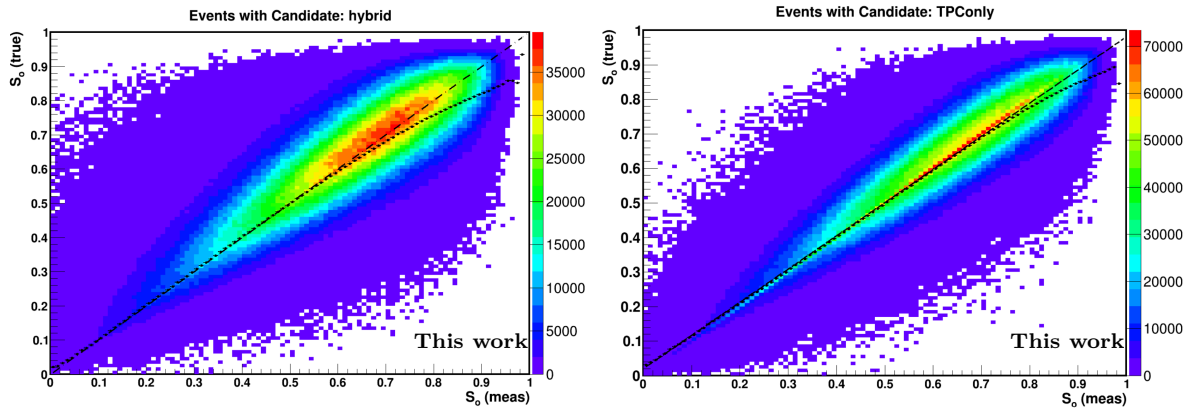
Variable	**Condition**
$\|\eta\|$	< 0.8
p_{T} (GeV/c)	> 0.3
ITS refit	yes
TPC refit	yes
Number of TPC crossed rows	≥ 70
TPC clusters found / findable	≥ 0.8
ITS minimum layers (max 6)	≥ 2

Table 5.3: General track quality selections.

$\mathrm{p_T}$ interval (GeV/c)	1-2	4-6	8-12	12-24
Inv. mass $\mathrm{D^0}$ - $\mathrm{M_{D^0}^{PDG}}$ [GeV/c^2]	0.03	0.036	0.05	0.7
DCA [cm]	0.0315	0.05	0.105	0.15
$\cos\theta^*$	0.8	1.0	1.0	1.0
p_{T}K [GeV/c]	0.5	1.0	1.0	0.3
$p_{\mathrm{T}}\pi$ [GeV/c]	0.5	1.0	1.0	0.3
Imp. par. $\mathrm{d_0^K}$ [cm]	0.1	0.09	0.1	0.5
Imp. par. $\mathrm{d_0^\pi}$ [cm]	0.1	0.09	0.1	0.5
$\mathrm{d_0^K} \times \mathrm{d_0^\pi}$ [cm^2]	-33e-5	55e-6	0.01	0.01
$\cos\theta_{point}$	0.865	0.79	0.68	0.6
Inv. mass half-width $\mathrm{D^{*+}}$ [GeV/c^2]	0.3	0.3	0.3	0.3
Half-width $\mathrm{M}_{D^{*+}} - \mathrm{M}_{D^0}$ [GeV/c^2]	0.15	0.1	0.1	0.3
p_{T} max soft π [GeV/c]	0.5	100	100	100
p_{T} min soft π [GeV/c]	0.05	0.05	0.05	0.05

Table 5.4: Topological selection criteria for $\mathrm{D^{*+}}$ in a selection of p_{T} intervals. The values are all upper limits unless specified.

of the emission angle of $\mathrm{D^0}$ in the $\mathrm{D^{*+}}$ centre-of-mass system with respect to the $\mathrm{D^{*+}}$ momentum direction ($\cos\theta^*$), impact parameters of kaon and pions from the $\mathrm{D^0}$ decay, and the invariant mass of $\mathrm{D^0}$ and $\mathrm{D^{*+}}$ appear on the variables list on which thresholds are applied for the selection procedure.
For the spherocity dependent analysis, two track selection cuts were considered: a hybrid approach, which considers tracks from both TPC and ITS, and a TPC only selection requiring the refitting of the tracks in the same detector (most commonly used method). These methods were tested by studying the correlation between generated and measured spherocity using Monte Carlo simulations. Hence, the correlation plot of the true spherocity as a function of the measured S_{O} with the hybrid track cuts (Fig. 5.9a) is compared with the TPC only case, shown in Fig. 5.9b.
The obtained values are related to D mesons track candidates and the TPC only choice provides a narrower correlation between the measured and generated spherocity. Although both profiles show a drop at large values of generated spherocity, the latter is



(a) (b)

Figure 5.9: D-meson candidates true spherocity as a function of the measured one for hybrid (a) and TPC only with TPC refit (b) track cuts obtained from a Monte Carlo simulation.

the case in which the most probable value remains closer to the diagonally fitted line. Consequently, the standard TPC cuts with the TPC refitting procedure were used in the analysis discussed in this dissertation.
A common PID selection strategy was implemented for the analyses, applying a $N_\sigma < 3$ compatibility selection for both TPC and TOF signals on the masses of pions and kaons daughter candidates of the D mesons (D^0 decay tracks for the D^{*+}).

5.4 Signal extraction

Similarly to the $\Sigma_c^{0,++}$ analysis, the difference in the invariant mass between $K^-\pi^+\pi^+$ (D^{*+}) and $K^-\pi^+$ (D^0) is measured when analysing D^{*+} production. This corresponds to $\Delta m = m_{\rm inv}(K^-\pi^+\pi^+) - m_{\rm inv}(K^-\pi^+)$ which for the true D^{*+} candidates has a nominal value of ≈ 145.42 MeV/c^2 from PDG values [4].
The production measurement is performed as a function of $p_{\rm T}$ in the range from 1 to 24 GeV/c, for the spherocity integrated analysis, and starting from 2 GeV/c for the multiplicity and spherocity dependent analysis (due to the smaller number of events in this double-differential analysis).
The invariant mass peak is described by a Gaussian fit, while the background is described using a power function convoluted with an exponential

$$F_{bkg} = \alpha\sqrt{x - m_\pi}\, e^{\beta(x-m_\pi)}, \tag{5.5}$$

where α and β are free parameters.
An example of the Δm distributions with superimposed the best fit functions is shown in the plots in Fig. 5.10-5.11 for the multiplicity dependent only, and multiplicity and spherocity dependent analyses, respectively.

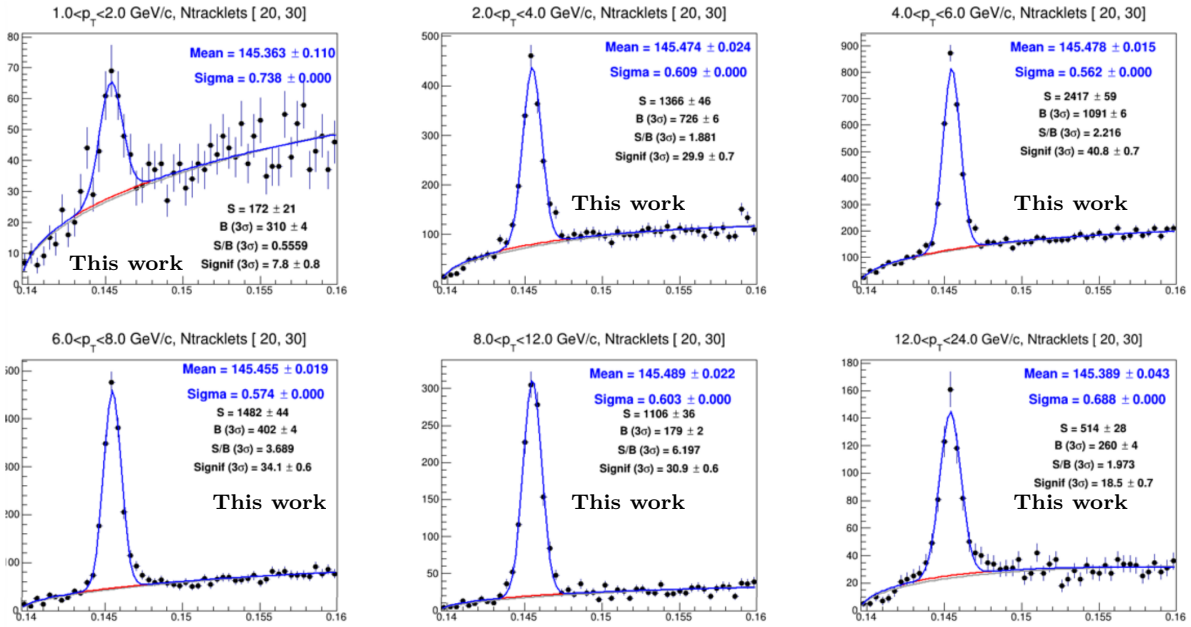


Figure 5.10: Invariant mass spectra for different p_{T} intervals for D^{*+} in the [20,30] multiplicity interval with integrated spherocity.

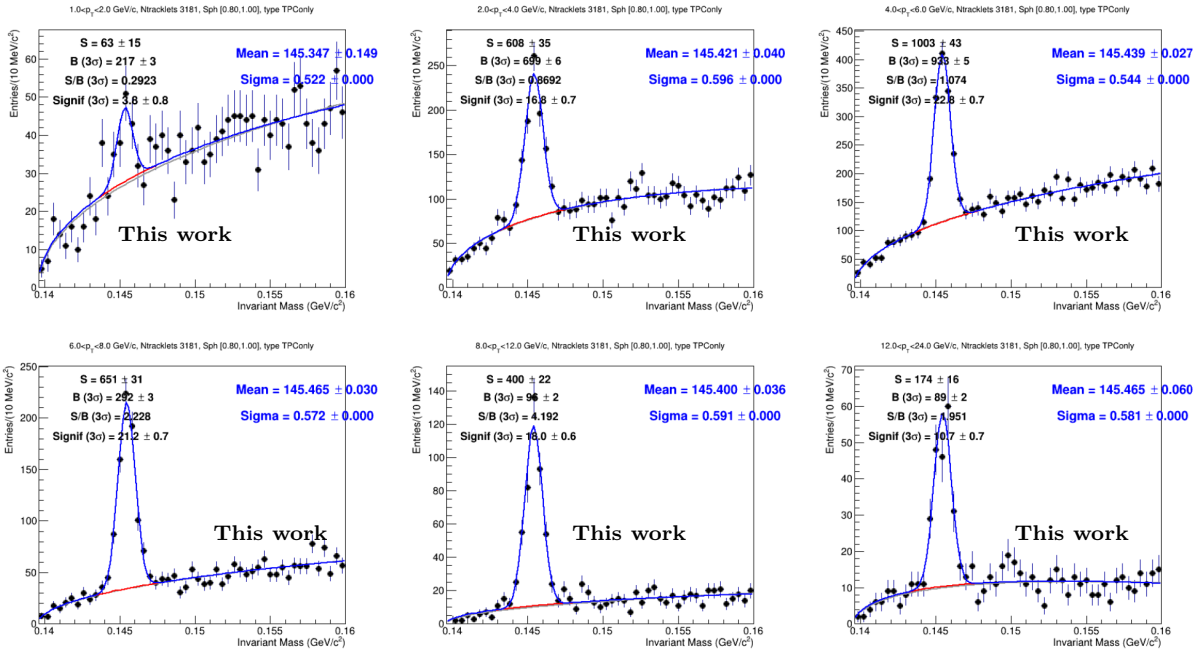


Figure 5.11: Invariant mass spectra for different p_{T} intervals for D^{*+} in the [31,81] multiplicity and [0.8,1] spherocity intervals.

For these results, the widths of the invariant mass were fixed in the fitting procedure to the MB values in order to reduce statistical fluctuations in the results.

5.4.1 Self-normalised yield

The invariant mass spectra of D mesons will be used for the calculation of the self-normalised yield, which is a commonly used observable in multiplicity dependent analyses since several systematic uncertainties cancel out when considering this quantity. All the final results shown in this chapter will be expressed using this variable, which, for the multiplicity dependent only analysis, is calculated using the formula:

$$\mathrm{Y}_{\mathrm{corr}}^{\mathrm{mult}} = \left(\frac{\mathrm{Y}^{\mathrm{mult}}}{\left((\mathrm{Acc} \times \varepsilon)^{\mathrm{mult}} \times N_{\mathrm{event}}^{\mathrm{mult}} \right) / \varepsilon_{\mathrm{mult}}^{\mathrm{trg}}} \right) \Bigg/ \left(\frac{\mathrm{Y}_{\mathrm{MB}}^{\mathrm{mult}}}{\left((\mathrm{Acc} \times \varepsilon)^{\mathrm{MB}} \times N_{\mathrm{event}}^{\mathrm{MB}} \right) / \varepsilon_{\mathrm{MB}}^{\mathrm{trg}}} \right), \quad (5.6)$$

where $\mathrm{Y}_{\mathrm{corr}}^{\mathrm{mult}}$ is the self-normalised yield in a multiplicity interval, Y are the yield in multiplicity (mult) intervals or integrated multiplicity (MB) and the same notation is applied for the acceptance times efficiency factor and the number of events. An additional factor is introduced in the formula which is the trigger efficiency, implemented to account for trigger imperfections in the event discrimination. This factor is multiplicity dependent and it is estimated, via Monte Carlo simulations, to be 0.92 for the minimum bias interval, 0.87 for 1–8 tracklets, and 0.98 in the 9–13 case [138]. All other multiplicity intervals have a predicted trigger efficiency of 1.
A similar normalisation is used for the spherocity dependent (in multiplicity intervals) analysis, in which the normalisation is done with the spherocity integrated case, and not multiplicity integrated. Therefore the trigger efficiency factor cancels out and we obtain the formula

$$\mathrm{Y}_{\mathrm{corr}}^{S_{\mathrm{O}},\,\mathrm{mult}} = \left(\frac{\mathrm{Y}_{S_{\mathrm{O}}}^{\mathrm{mult}}}{(\mathrm{Acc} \times \varepsilon)^{S_{\mathrm{O}},\,\mathrm{mult}} \times N_{\mathrm{event}}^{S_{\mathrm{O}},\,\mathrm{mult}}} \right) \Bigg/ \left(\frac{\mathrm{Y}^{\mathrm{mult}}}{(\mathrm{Acc} \times \varepsilon)^{\mathrm{mult}} \times N_{\mathrm{event}}^{\mathrm{mult}}} \right), \quad (5.7)$$

in which "S_{O}, mult" indicates the spherocity and multiplicity dependent case, while the "mult" only label refers to the spherocity integrated values in the considered multiplicity class.

5.5 Corrections

Before obtaining the final production results of the prompt D mesons, the raw yields need to be corrected for the acceptance and selection efficiency (according to Eqs. 5.6-5.7), similarly to what was done in Chapter 4, which are obtained using Monte Carlo simulations.
The reconstruction and selection efficiency depends on the multiplicity of charged particles produced in the collision, since the resolution of the primary vertex and of the topological selection variables improve at high multiplicity. In addition, the reconstruction efficiency is affected also by the spherocity interval, given that more jet-like (low) spherocity values are associated to better efficiency compared to isotropic events.
The efficiencies in each multiplicity interval considered are evaluated using Monte Carlo

samples that reproduce the multiplicity distribution observed in data, by using tracklet weights that are computed by doing a ratio of the $N_{\rm trk}$ distribution of data over the one in Monte Carlo simulations (selecting events with at least one D meson candidate). These weights are computed separately for each data-taking year and it has been verified that they are compatible within the uncertainties. They are applied to the year-wise efficiencies, whose values are merged and multiplied by the acceptance, which are respectively calculated using the Eq. 4.5 and 4.4 (used already for the $\Sigma_{\rm c}^{0,++}/\Lambda_{\rm c}^{+}$ corrections).

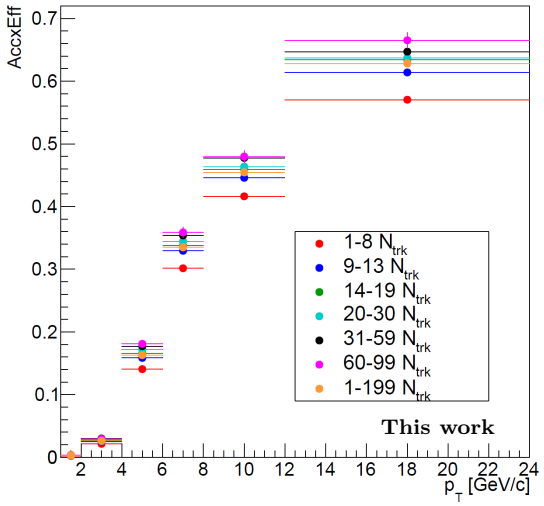
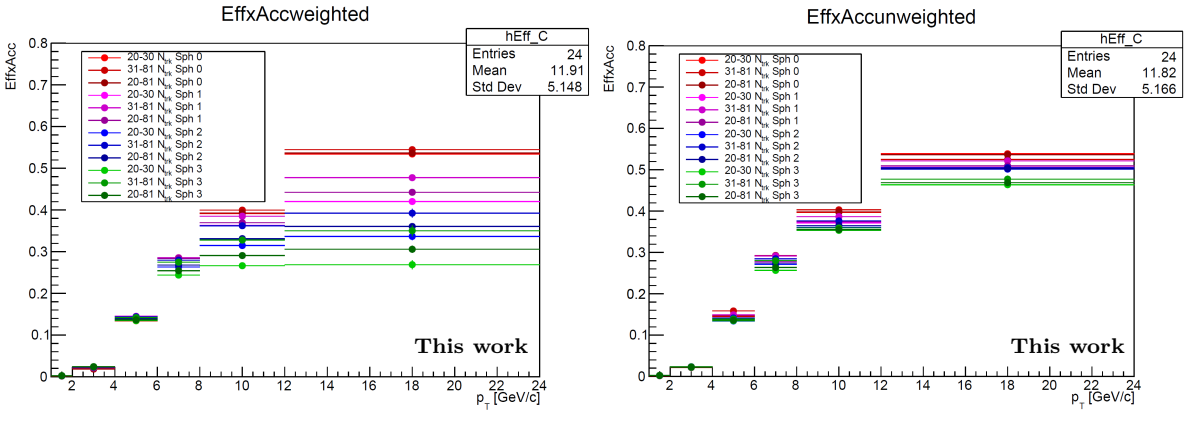


Figure 5.12: Product of acceptance and efficiency in multiplicity intervals (spherocity integrated) for prompt $\rm D^{*+}$ reconstruction.



(a) (b)

Figure 5.13: Product of acceptance and efficiency in multiplicity and spherocity intervals for prompt $\rm D^{*+}$ reconstruction using the weighted (a) and unweighted (b) definition of $S_{\rm O}$.

Year-wise weighted efficiency values are merged before being combined with the acceptance and, other than being calculated as a function of p_{T}, they are uniquely computed for each multiplicity and spherocity interval.
Specifically, the Acc $\times\, \varepsilon$ values for the spherocity integrated analysis in multiplicity intervals are shown as a function of p_{T} in Fig. 5.12, from which one can observe that higher efficiency is achieved at higher p_{T} and with higher multiplicity intervals. Furthermore, as expected, the MB correction is somewhat the average of the multiplicity-dependent values.

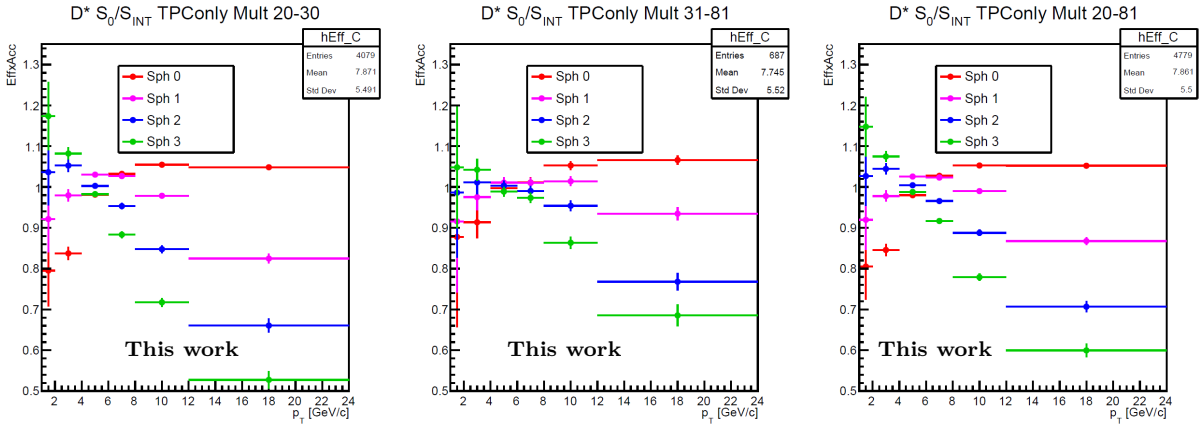


Figure 5.14: Ratio of efficiency correction in multiplicity and spherocity intervals with the spherocity integrated case for prompt D^{*+} reconstruction using the weighted definition of S_{O}.

For the spherocity dependent analysis, both the weighted and unweighted S_{O} were considered in order to study which definition could lead to a better reconstruction efficiency. The obtained results in Fig. 5.13 for both cases show a peculiar spread in the efficiency values among the spherocity intervals in a specific p_{T}. However, this phenomenon is significantly less pronounced in the unweighted spherocity results in Fig. 5.13b.
To quantify the spread, the efficiency ratios as a function of p_{T} and multiplicity in spherocity intervals are computed with the spherocity integrated case and shown in Fig. 5.14-5.15. The former contains results of the weighted spherocity definition and a difference up to almost 50% compared to the spherocity integrated case is observed in the lowest multiplicity interval at the highest p_{T} measured, while the spread of the unweighted spherocity efficiencies is limited within uncertainties to about 10%. It must be noted that the plots are also showing the 1-2 GeV/c p_{T} interval, in which a preliminary analysis was performed. Due to the current amount of data, this interval has a lower efficiency compared to the others and was not considered for the final results (given also the difficulties in obtaining an invariant mass spectrum in specific spherocity and multiplicity intervals).

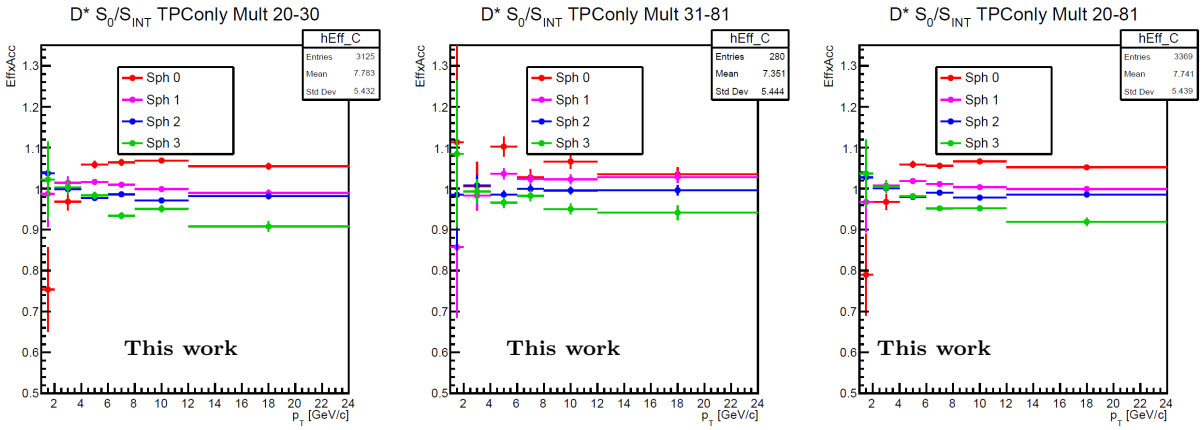


Figure 5.15: Ratio of efficiency correction in multiplicity and spherocity intervals with the spherocity integrated case for prompt D^{*+} reconstruction using the unweighted definition of S_O.

5.6 Systematic uncertainties

The contributions of systematic uncertainties considered for the D mesons analysis will be described in the following paragraphs. Some of them are similar to those discussed in Sec. 4.5, but the extraction procedure might be different.
Due to large statistical fluctuations, the estimate of the systematic uncertainties for the spherocity dependent analysis are carried out in the wide multiplicity interval 20–81 tracklets and the same uncertainties are assigned to the two narrower multiplicity classes.

5.6.1 Raw yield extraction

The multitrial approach for the calculation of the raw yield extraction systematics described in Sec. 4.5.1 is implemented in the D mesons analysis. For the D^{*+}, the variations on the fit are presented in Tab. 5.5 and the fit procedure is executed for each p_T, multiplicity and spherocity interval. A power law was considered for the background fitting,

$m_{\rm inv}$ **low (GeV/c^2)**	(0.1396, 0.1397, 0.1398, 0.1399, 0.14, 0.1401, 0.1402...0.1406)
$m_{\rm inv}$ **up (GeV/c^2)**	(0.1592, 0.1593, 0.1594, 0.1595, 0.1596, 0.1597, 0.1598...0.1602)
Background fit	PowerLaw, PowerLawExpo
Sigma	fixed sigma and free sigma

Table 5.5: Variations used for D^{*+} multitrial fitting. The invariant mass values ($m_{\rm inv}$), low and up, define combinations of fitting intervals.

in addition to the default function, and the sigma value of the invariant mass was considered both fixed to the value obtained for the multiplicity and spherocity integrated case, and kept as a free parameter. Moreover, a cut on the reduced χ^2 is applied at 2.5 for the D^{*+}, while the other mesons (with more statistics) have a tighter cut at 2.

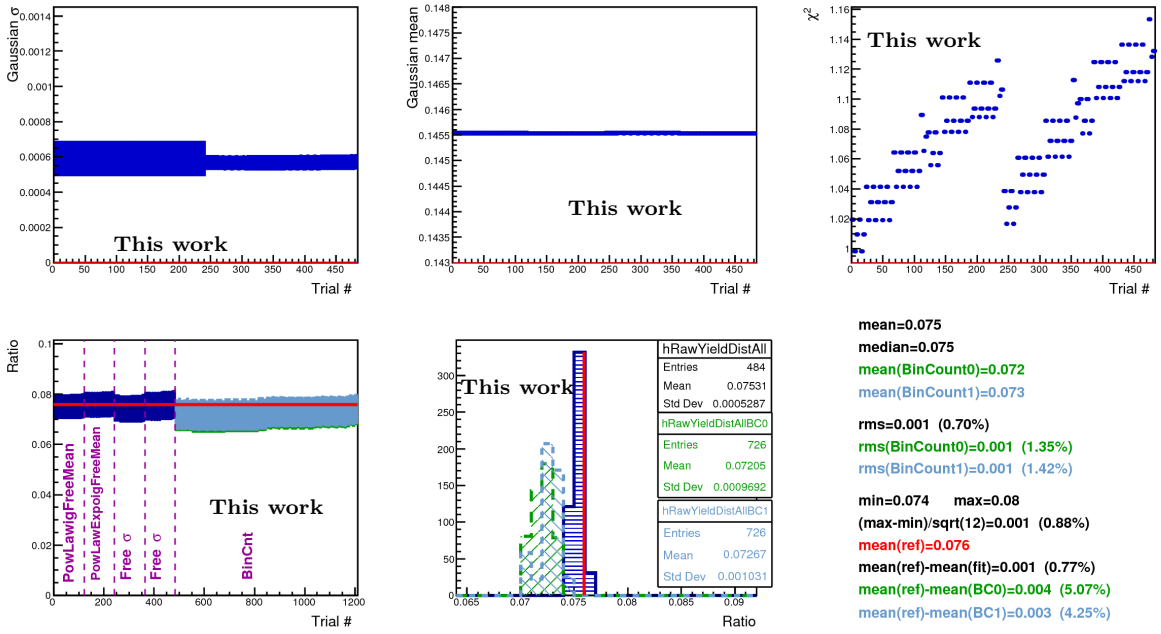


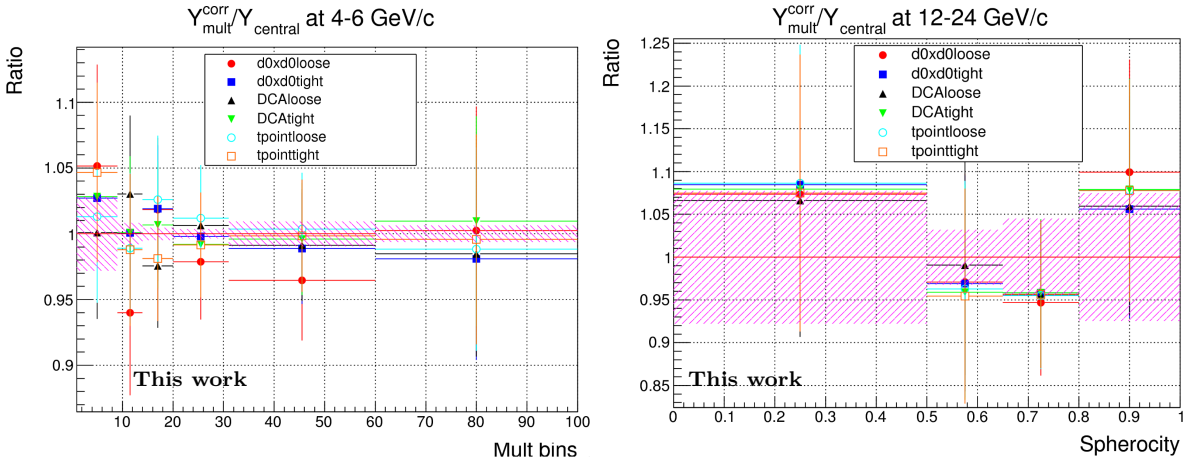
Figure 5.16: Output of the multitrial fitter for the D^{*+} meson for p_T interval [4-6] GeV/c, spherocity [0-0.5] and multiplicity [20-81]. It shows the Gaussian sigma (top left), Gaussian mean (top middle), the reduced χ^2 (top right), and the ratio of raw yields in multiplicity and spherocity intervals with the spherocity integrated one extracted from the fit. Bin counting, in green and light blue, is considered as a test (bottom row).

The main result shown in the output example of Fig. 5.16 is the ratio of the extracted yield in multiplicity and spherocity intervals with the spherocity integrated one. In case of the multiplicity dependent only analysis, the ratio is obtained by considering the minimum bias multiplicity interval as denominator. The obtained relative RMS is taken into account as systematic uncertainty.

5.6.2 Selection systematics

The obtained results from data need to be corrected for detector acceptances and efficiencies which are determined from Monte Carlo simulations, so biases could be introduced due to imperfections in the simulation of the variables used in the candidates selection. To quantify this effect, as done for the $\Sigma_c^{0,++}/\Lambda_c^+$ analysis, the self-normalised yields obtained using different selection criteria are compared, and the RMS of the difference of the ratio of all the cut variations to the central value with unity is applied as systematic uncertainty.
For the D^{*+}, six different cut variations were considered, making tighter and looser the cuts of specific selection variables. The output results of the procedure are expressed using the ratio of the self-normalised yield obtained in these conditions with the final

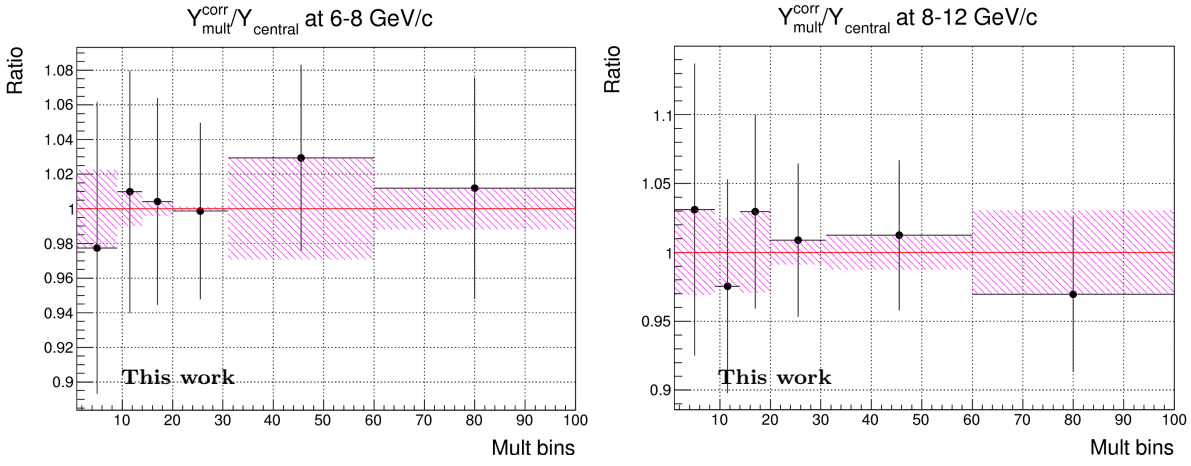
central results as a function of multiplicity and spherocity intervals in Fig. 5.17, where the pink band shows the estimated systematic uncertainty.



(a) (b)

Figure 5.17: Example of self-normalised yields ratio using different cuts with the central final results of D^{*+} as a function multiplicity (a) and spherocity (b), respectively, for 4 $\leq p_{\mathrm{T}} <$ 6 GeV/c and 12 $\leq p_{\mathrm{T}} <$ 24 GeV/c in the 20-81 multiplicity interval.

5.6.3 Particle identification



(a) (b)

Figure 5.18: Ratios of yields obtained with and without PID selection including the difference with unity (pink boxes) calculated in two p_{T} intervals for the D^{*+} meson with spherocity integrated.

Particle identification techniques using TOF and TPC signals are applied also in the D mesons analysis, therefore, studying the uncertainty contribution that this procedure has on the final results is mandatory.
For this purpose, the self-normalised yields calculated with the previously described formulas are compared with and without PID in order to check for differences in the analysis results.
For all D mesons, not just the D^{*+}, it was concluded that the contribution of the PID uncertainty to the total systematics is negligible, since the ratios obtained between the two sets of results are compatible with unity in almost all the multiplicity and p_T intervals considered within the large uncertainties, as shown in the examples in Fig. 5.18.

5.6.4 Feed-down

This analysis assumes that corrections and subtraction from feed-down beauty decays are multiplicity and spherocity independent, leading to the fact they are cancelled when calculating the self-normalised yields. However, this assumption contributes significantly to the final results as a source of systematic uncertainties, which are calculated in two different ways for the multiplicity and spherocity dependent analyses.

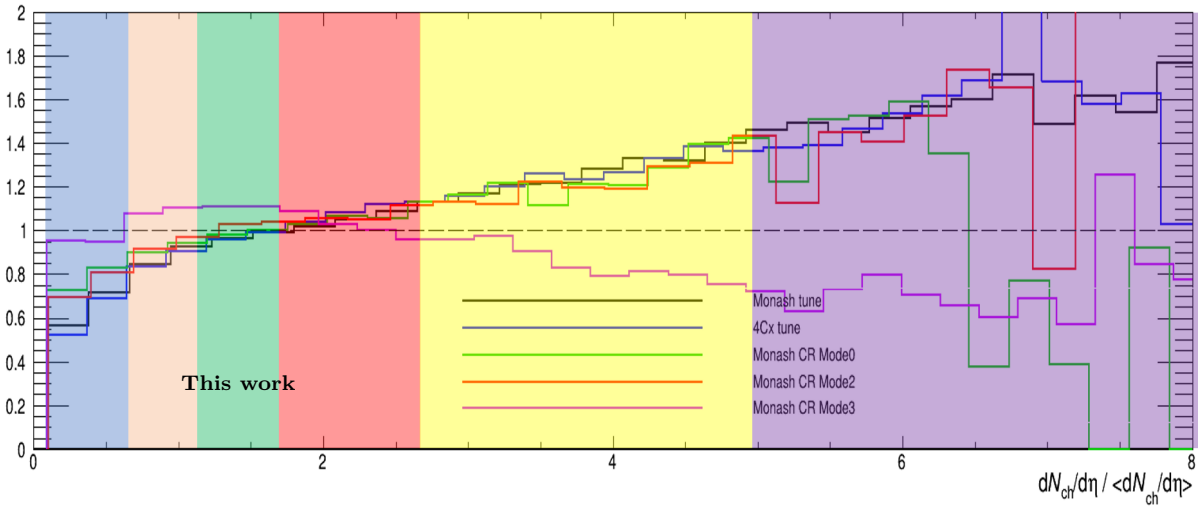


Figure 5.19: $f^{\text{mult}}_{\text{non-prompt}}/f^{\text{MB}}_{\text{non-prompt}}$ ratios as a function of charged-particle density obtained with various tunes of PYTHIA8, with the colours highligthing the multiplicity intervals of the D mesons analysis.

In the former (spherocity integrated) it is considered that the systematic uncertainty contribution of beauty hadron decays to the D meson yields is constant in p_T intervals and varies only with multiplicity. The systematics are obtained by taking into account PYTHIA8 simulations, run with different tunes (Monash, enhanced CR mode 0/2/3 and 4Cx[2] [140]), as a function of charged particle densities which are shown in Fig. 5.19. In

[2]This configuration is a variation of the past default 4C tune of PYTHIA8 (which was based on early

the plot, a series of coloured bands highlights the regions of multiplicities described by the conversions of tracklets to charged particles (Sec. 5.1).
The results are expressed as the ratio of the fraction of non-prompt decays in a multiplicity interval with the multiplicity integrated case $f^{\rm mult}_{\rm non-prompt}/f^{\rm MB}_{\rm non-prompt}$.
The fine multiplicity intervals obtained in the simulations were rebinned following the coverage bands of the D mesons multiplicity analysis, by doing a simple average of the values, and the results are illustrated in Fig. 5.20.

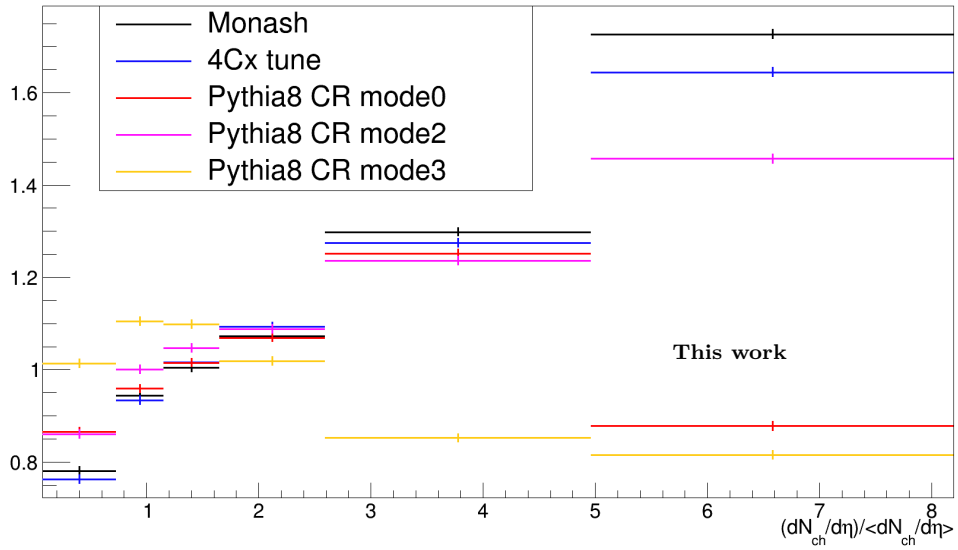


Figure 5.20: $f^{\rm mult}_{\rm non-prompt}/f^{\rm MB}_{\rm non-prompt}$ ratio obtained with various tunes of PYTHIA8, rebinned to the multiplicity intervals of the D mesons analysis.

Thus, the minimum and maximum values among the PYTHIA8 tunes in each multiplicity interval are identified and their shifts from unity are selected as lower and upper relative uncertainties of the non-prompt fraction.
After calculating $f_{\rm prompt}$ using the $\rm N_b$ method described in Sec. 4.4.2, which uses FONLL cross section predictions, the relative lower and upper feed-down systematic uncertainties $\sigma(f_{\rm prompt})^{\rm REL}$ are obtained via the equation

$$\sigma\left(f_{\rm prompt}\right)^{\rm REL}_{\rm low/up} = \frac{(1-f_{\rm prompt})\cdot\sigma\left(f_{\rm non-prompt}\right)^{\rm REL}_{\rm low/up}}{f_{\rm prompt}}. \tag{5.8}$$

On the other hand, in the spherocity dependent analysis the systematic contribution of feed-down subtraction is based on the assumption that $f_{\rm non-prompt}$ has a maximum variation of 50%. Hence, using the $f_{\rm prompt}$ calculated in the MB multiplicity, spherocity integrated case (central), the relative uncertainty is assigned in each $p_{\rm T}$ interval by using the formula

$$\sigma\left(f_{\rm prompt}\right)^{\rm REL}_{\rm low/up} = 1 - \frac{f^{\rm central}_{\rm prompt}}{f^{\rm max/min\ var}_{\rm prompt}}, \tag{5.9}$$

LHC data) [139], including a Gaussian matter profile with an x-dependent width.

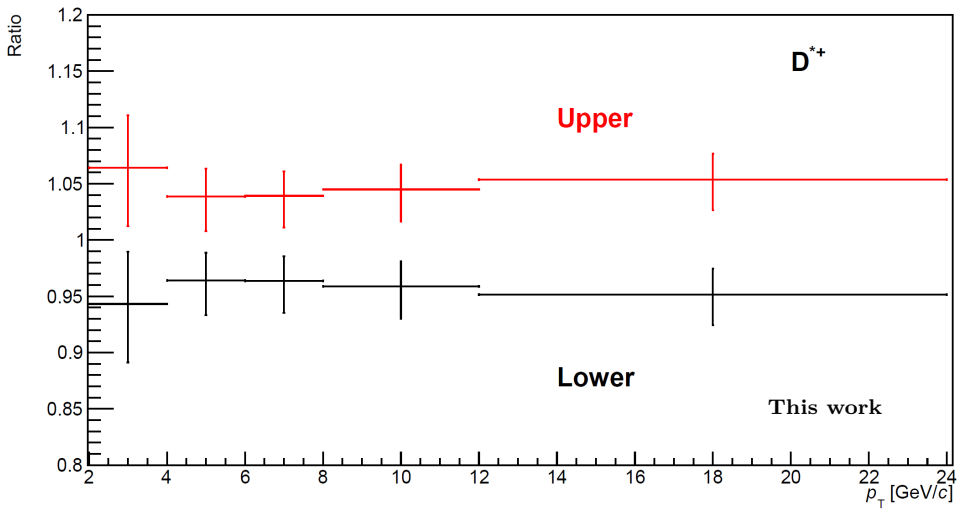


Figure 5.21: Upper and lower feed-down uncertainties for the D^{*+} meson in the spherocity dependent analysis.

where the variations of f_{prompt} are calculated, respectively, by adding and subtracting to $f_{\text{non-prompt}}$ half of its value. The obtained relative uncertainties are presented for the D^{*+} meson in Fig. 5.21.

5.6.5 D mesons p_{T} shape

Consistently to what discussed in Sec. 4.5.5, the efficiency calculations obtained from Monte Carlo simulations are sensitive to the p_{T} distribution of D mesons. These are obtained from PYTHIA6 and they are used as weights during data generation to account for the difference of the p_{T} trends with FONLL calculations.
The effect is considered constant in multiplicity and spherocity intervals, so the systematic uncertainty is evaluated in the largest multiplicity intervals (1-199 and 20-81 tracklets) and spherocity integrated case as

$$\sigma^{MC}_{p_T\ \text{shape}} = 1 - \frac{\varepsilon^{\text{FONLL}}}{\varepsilon^{p_{\text{T}}\ \text{weights}}}, \tag{5.10}$$

where the ratio of the efficiencies without (FONLL) and with p_{T} weights is computed and the values are illustrated in Fig. 5.22.

5.6.6 Multiplicity weights

While calculating the efficiencies, the distribution of number of tracklets were weighted using PYTHIA8 obtained distributions in order to better simulate the trends of experimental results in each year of data-taking, hence a source of uncertainty is introduced. Considering the nature of this correction, the spherocity dependent analysis is exempt

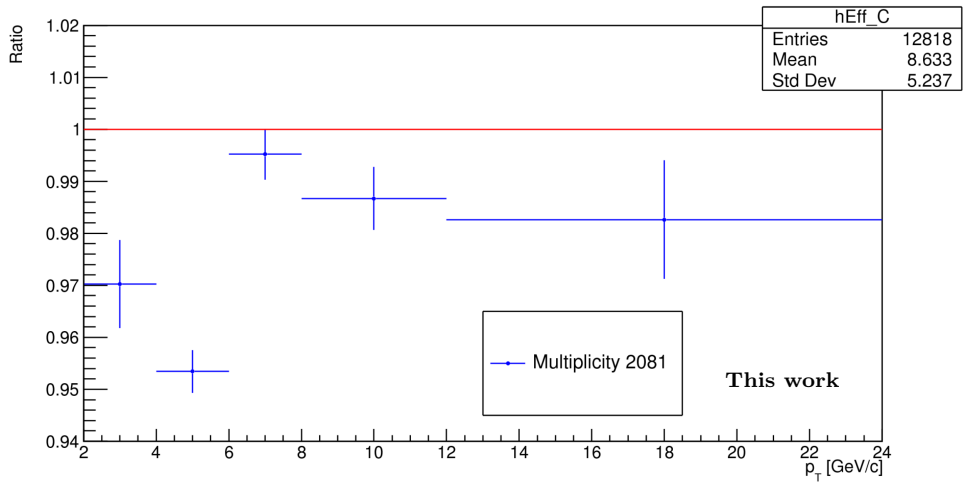


Figure 5.22: Efficiency ratios using FONLL calculations and p_{T} weights for the D^{*+} analysis.

from this uncertainty contribution because the weights are cancelled in the calculation of the self-normalised yield.
The related uncertainty is estimated by calculating the ratio of efficiencies to which two sets of weights are applied: one considers only events with D mesons candidates (also used to estimate the central value of the self-normalised yields), while the other includes all the events with at least a particle candidate (without minding its mass).
From the obtained results, one can obtain the systematic uncertainty with the deviation of the ratio from unity (in different multiplicity intervals). In the D^{*+} meson analysis, this uncertainty is below 0.5% for each considered interval, so the contribution to the final results is considered negligible.

5.6.7 Closure test on spherocity intervals

In the calculation of the efficiencies, it is expected that the values as a function of p_{T} should not change significantly in spherocity intervals. However, as shown in Fig. 5.13b, a difference up to $\approx 10\%$ is observed, meaning that the obtained efficiency might not consider properly the spherocity intervals (folded results).
A study was then performed to check if the phenomenon is occurring, by analysing the ratio

$$\frac{Y_{\mathrm{sph}}^{\mathrm{MCgen}}}{Y_{\mathrm{sph}}^{\mathrm{MCrec}}/\varepsilon_{\mathrm{sph}}}, \tag{5.11}$$

where $Y_{\mathrm{sph}}^{\mathrm{MCrec}}$ is the spherocity differential D-meson p_{T} distribution with reconstructed tracks, corrected with multiplicity integrated efficiency $\varepsilon_{\mathrm{sph}}$, triggered on reconstructed spherocity selection, and $Y_{\mathrm{sph}}^{\mathrm{MCgen}}$ is the spherocity differential D-meson p_{T} spectra with generated tracks, triggered on generated spherocity selection. This closure test implies that no folding is occurring if the ratio is close to unity, but the results in Fig. 5.23 for

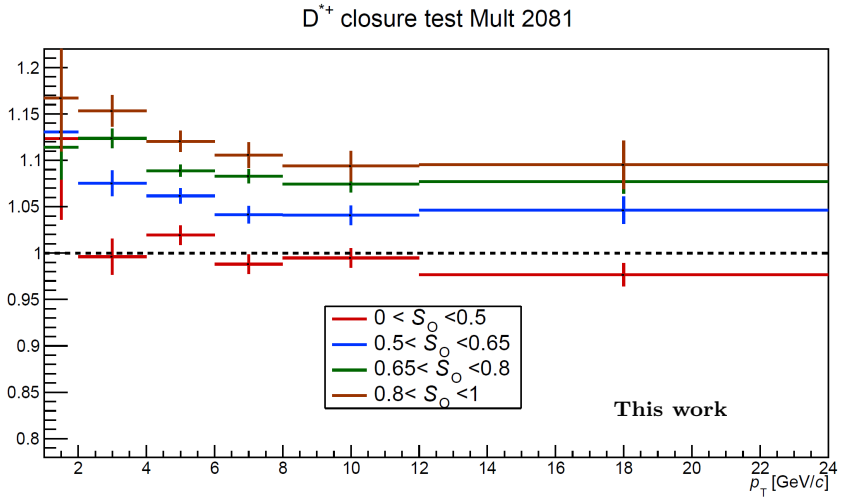


Figure 5.23: Closure test in the 20-81 multiplicity interval for the D^{*+} meson using the ratio in Eq. 5.11.

the D^{*+} meson (consistent with the other species) are showing a deviation up to 15% in the highest spherocity interval.
As a matter of fact, it is observed that the ratio increases with spherocity, but, as the ratio is calculated, ones would expect that the average of the obtained values in each p_{T} interval should be around unity (assuming the spherocity is reconstructed properly and it is compatible with the generated one).
Thus, the responsibility of the shift in the closure test average results is assigned to the difference in the distributions of reconstructed and generated spherocity, given that the applied cuts on the variables lead to different spherocity percentages. For this reason a 10% systematic uncertainty was estimated in all the p_{T} and spherocity intervals and a new closure test will be performed in the future using spherocity quantiles.

5.6.8 Summary of D^{*+} systematic uncertainties

A summary of the systematic uncertainties obtained for the D^{*+} meson is shown in Tab. 5.6 for the spherocity integrated case in the 14-19 N_{trk} multiplicity class. The uncertainties were refined and smoothed in multiplicity and p_{T} intervals, given that large variations in specific intervals are likely caused by statistical fluctuations.
A similar summary with the intermediate 0.5-0.65 spherocity interval is presented in Tab. 5.7 for the 20-81 multiplicity class.

5.7 Results

Measurements of self-normalised yields of the D^{*+} meson are combined with those of D^{+} and D^{0} in the preliminary results approved by the ALICE Collaboration in pp collisions

p_{T} [GeV/c]	$1-2$	$2-4$	$4-6$	$6-8$	$8-12$	$12-24$
RawYield	5%	1%	0%	0%	0%	1%
Cut Variation	15%	2%	0%	1%	5%	6%
PID	0%	0%	0%	0%	0%	0%
Feed-Down (lower)	0%	0%	0%	0%	0%	0%
Feed-Down (upper)	2%	1%	1%	1%	1%	1%
Monte Carlo p_{T} shape	4%	0%	0%	0%	0%	0%
Mult weights	0%	0%	0%	0%	0%	0%

Table 5.6: Summary of systematic uncertainties in the 14-19 multiplicity class as a function of p_{T} for the D^{*+} meson.

p_{T} [GeV/c]	$2-4$	$4-6$	$6-8$	$8-12$	$12-24$
RawYield	1%	1%	0%	0%	1%
Cut Variation	8%	0%	0%	2%	3%
Feed-Down (lower)	6%	4%	4%	4%	5%
Feed-Down (upper)	6%	4%	4%	4%	5%
Monte Carlo p_{T} shape	3%	5%	0%	1%	1%
Closure test	10%	10%	10%	10%	10%

Table 5.7: Summary of systematic uncertainties in the 20-81 multiplicity class and 0.5-0.65 spherocity interval as a function of p_{T} for the D^{*+} meson.

at $\sqrt{s} = 13$ TeV in midrapidity.
A weighted average was performed via the formula

$$\mathrm{Y}_{\mathrm{mean}} = \frac{\sum(w_i \mathrm{Y}_i)}{\sum(w_i)}, \text{ with weights} \tag{5.12}$$

$$w_i = \left(\frac{\mathrm{Y}_i}{\sigma_i}\right)^2, \tag{5.13}$$

in which the index i indicates the three D mesons, while σ is the statistical uncertainty of a D meson self-normalised yield in a specific p_{T}, spherocity and multiplicity interval. The average uncertainty of the combined measurement is obtained in a similar way, via

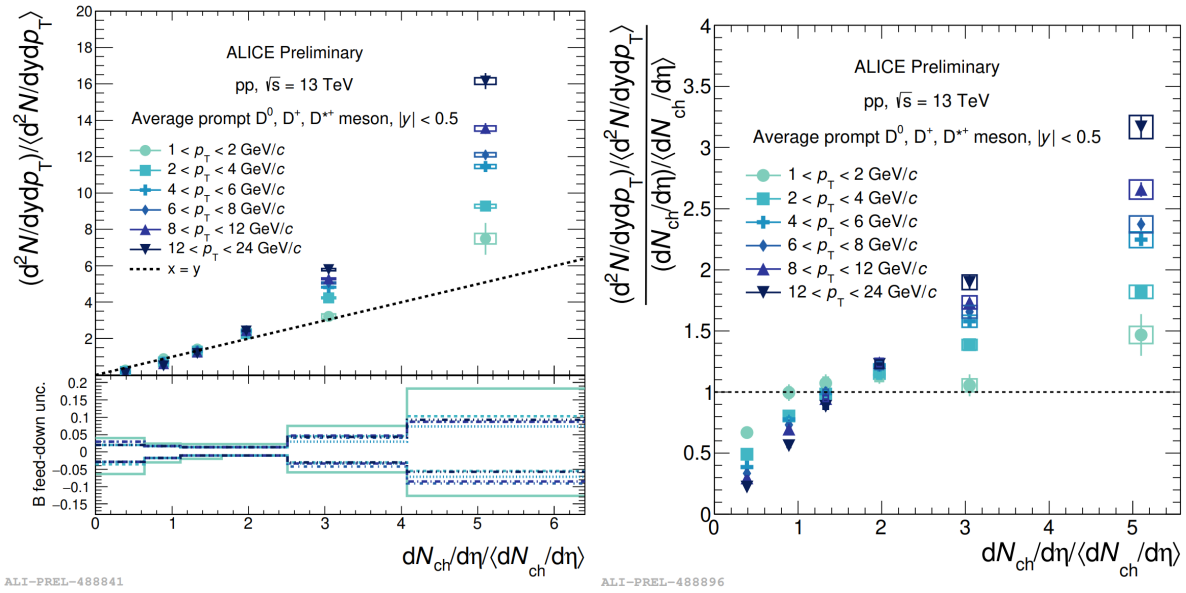
$$\sigma_{\mathrm{mean}} = \frac{\sqrt{\sum(w_i\sigma_i)^2}}{\sum(w_i)}. \tag{5.14}$$

The systematic uncertainties contributions are combined using error propagation via the equation

$$\sigma_i^{\mathrm{sys}} = \mathrm{Y}_i \sqrt{\sum (\sigma_j^{\mathrm{rel}})^2}, \tag{5.15}$$

where the total absolute systematic uncertainty σ_i^{sys} for a single D meson species is obtained from the squared sum of the relative systematic contributions (indicated with the

j index). Finally, the three final uncertainties are combined and assigned to the average D mesons self-normalised yields by using the approach described by Eq. 5.14, in which the weights are now calculated using the results from Eq. 5.15.
The spherocity integrated results in multiplicity intervals as a function of p_{T} in pp collisions at $\sqrt{s}$ = 13 TeV will be discussed first and the average self-normalised yields are shown in Fig. 5.24a.
The plot shows the results as a function of charged-particle density in p_{T} intervals, indicating the systematic uncertainties with boxes around the points and the feed-down uncertainties on the bottom panel. A faster-than-linear increasing trend is noticeable from the results, which make clear that the higher the p_{T}, the higher is the production of D mesons in high multiplicity intervals. This is even better observed on the ratio of self-normalised yield with the charged-particle density in Fig. 5.24b (double ratio).
It is observed that the results obtained for this dissertation are more precise and they have a higher multiplicity reach compared to past results obtained in pp collisions at $\sqrt{s}$ = 7 TeV by the ALICE Collaboration [84], which are fully illustrated in Fig. 2.15c.



(a) (b)

Figure 5.24: Average D mesons self-normalised yields (a) and ratio of yields divided by charged-particle densities (b) as a function of charged-particle densities in p_{T} intervals.

These results are compatible with the self-normalised yields found in this analysis, as shown in Fig. 5.25, in which the comparison between the measurements is performed in three p_{T} intervals. A full comparison was not possible due to different values of multiplicity and p_{T} intervals achieved.
The compatibility to past measurements is also noticeable from the comparison with J/ψ [141] and electrons from heavy-flavour decays as a function of multiplicity illustrated for a low p_{T} interval in Fig. 5.26a and for a higher p_{T} interval in Fig. 5.26b.

The plots show also the double ratio of the self-normalised yield in the lower panels to better visualise the compatibility among the values.
Moreover, a comprehensive study was performed with Monte Carlo generators, first with EPOS3 and Colour Glass Condensate (CGC[3]) models (results preliminarly approved by the ALICE Collaboration), and afterwards with different tunes of PYTHIA8.
As observed in Fig. 5.27, the EPOS3 generator with the hydrodynamical evolution enabled describes the data reasonably well, as opposed to the estimates without hydrodynamics, which underestimate results especially at high multiplicity. On the other hand, the CGC framework with the 3-pomeron fusion mechanism [143] overestimates significantly the experimental measurements in all the p_{T} intervals considered. Hence, the high multiplicity reach obtained in all p_{T} intervals is fundamental to constrain models and provide them with valuable inputs for their development.

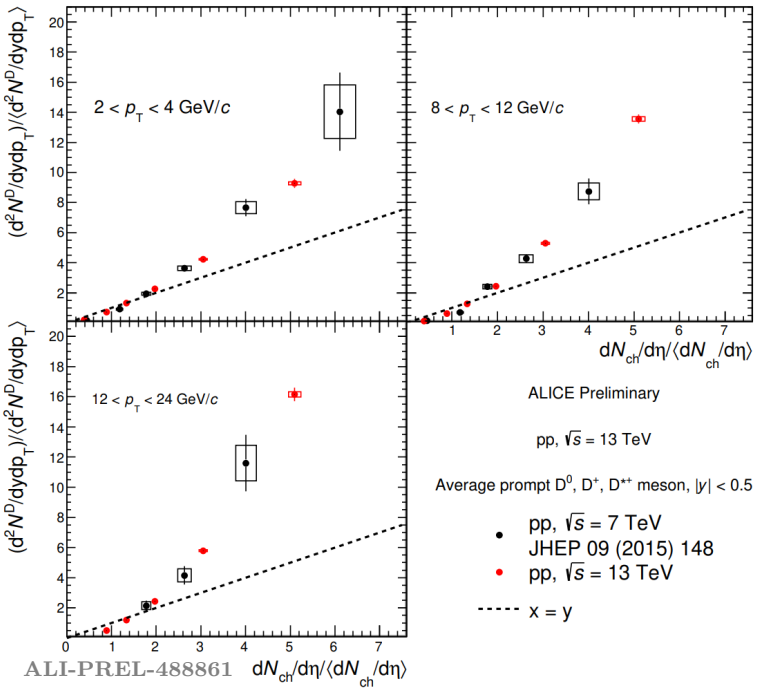
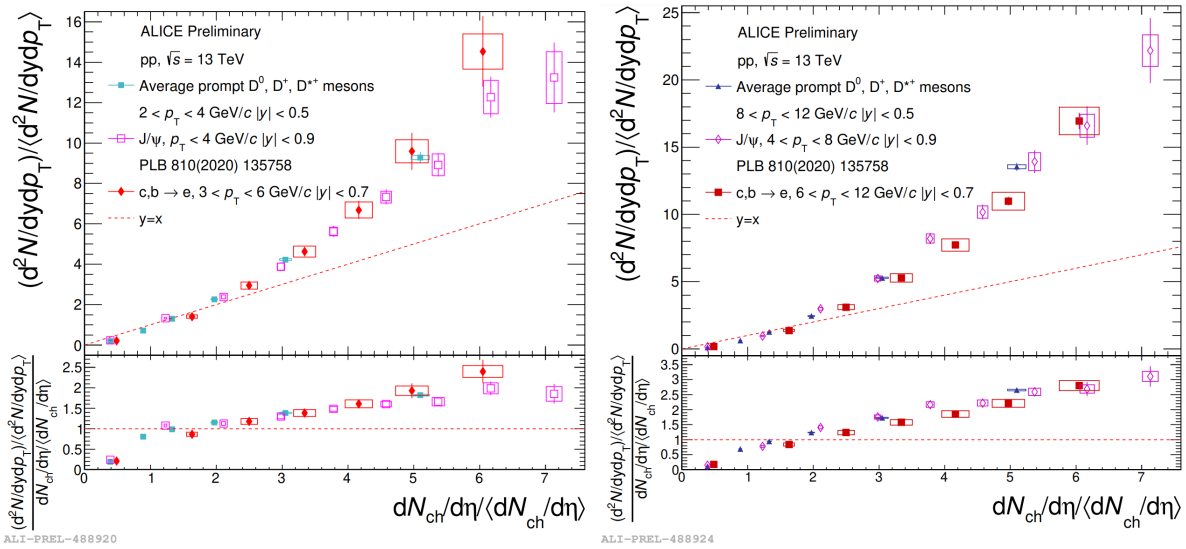


Figure 5.25: Comparison of average D mesons self-normalised yields in pp collisions at $\sqrt{s} = 13$ TeV with results obtained at $\sqrt{s} = 7$ TeV in three p_{T} intervals as a function of multiplicity.

[3]An effective field theory predicting gluon density saturation in the very low Bjorken-x region [142].



(a) (b)

Figure 5.26: Comparison of self-normalised yields of average D mesons, J/ψ and electrons from heavy-flavour decays in low p_{T} (a) and high p_{T} (b) intervals as a function of multiplicity.

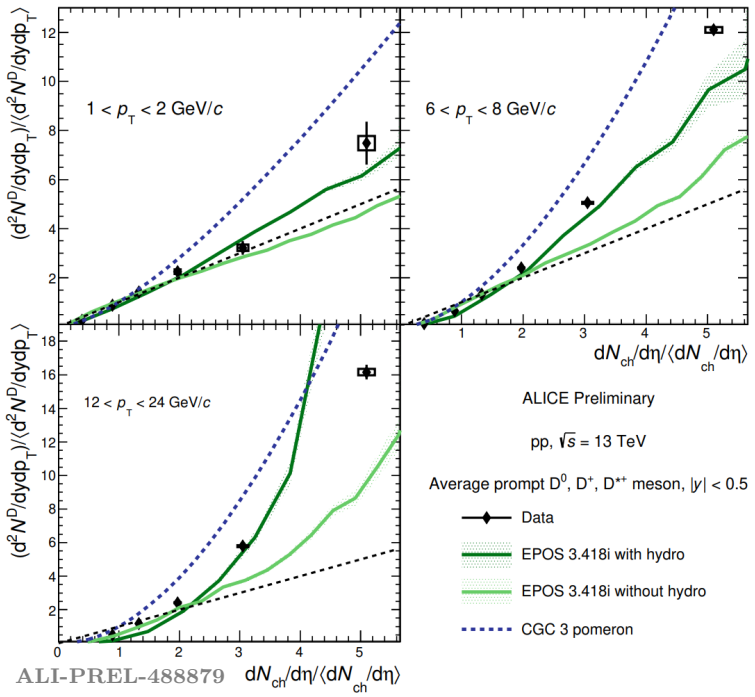


Figure 5.27: Comparison of average D mesons self-normalised yields in pp collisions at $\sqrt{s}$ = 13 TeV with EPOS3 (with and without hydrodynamical model) and CGC simulations in three p_{T} intervals as a function of multiplicity.

All the enhanced colour reconnection modes of PYTHIA8 are compared with data and with Monash tune in all the p_{T} intervals as a function of multiplicity in Fig. 5.28. The latter (Monash tune) describes well the data in the full p_{T} range and it is compatible with the enhanced colour reconnection Mode 0 and Mode 2 tunes. The Mode 3 configuration shows a significant production reduction compared to the others at high multiplicity.
Generally speaking, indeed, the different tunes show different predictions only at high multiplicity.

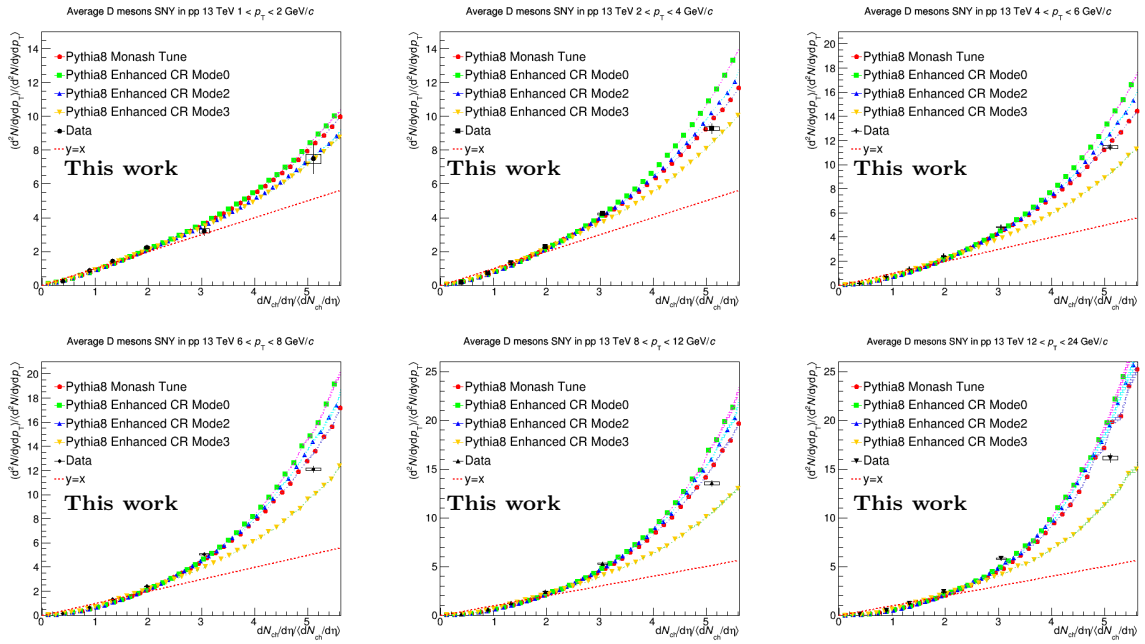


Figure 5.28: PYTHIA8 comparisons in p_{T} and multiplicity intervals of the average D mesons results using different tunes.

The contribution of multiple parton interactions and basic colour reconnection in PYTHIA was also studied by comparing the estimates of the generator with different Monash simulations forcely enabling or disabling the two mechanisms (one at a time). The predictions are also compared with the experimental results in Fig. 5.29 in all the p_{T} intervals.
The lack of statistics in the simulation without MPIs causes the results to have less multiplicity intervals and to be very different from the other simulations, drastically increasing the rapidity of the trend with p_{T}. The estimates were obtained by considering a maximum of 50 charged particles, which is half the value reached for the other simulations. On the other hand, disabling the colour reconnection yields to results that are compatible at low multiplicity with Monash tune, but they significantly underestimate data at higher charged particle density values.
This is an important point because it highlights the importance of MPI in a multiplicity dependent analysis, without which a powerful multi-purpose generator such as PYTHIA8

fails to describe properly the results.

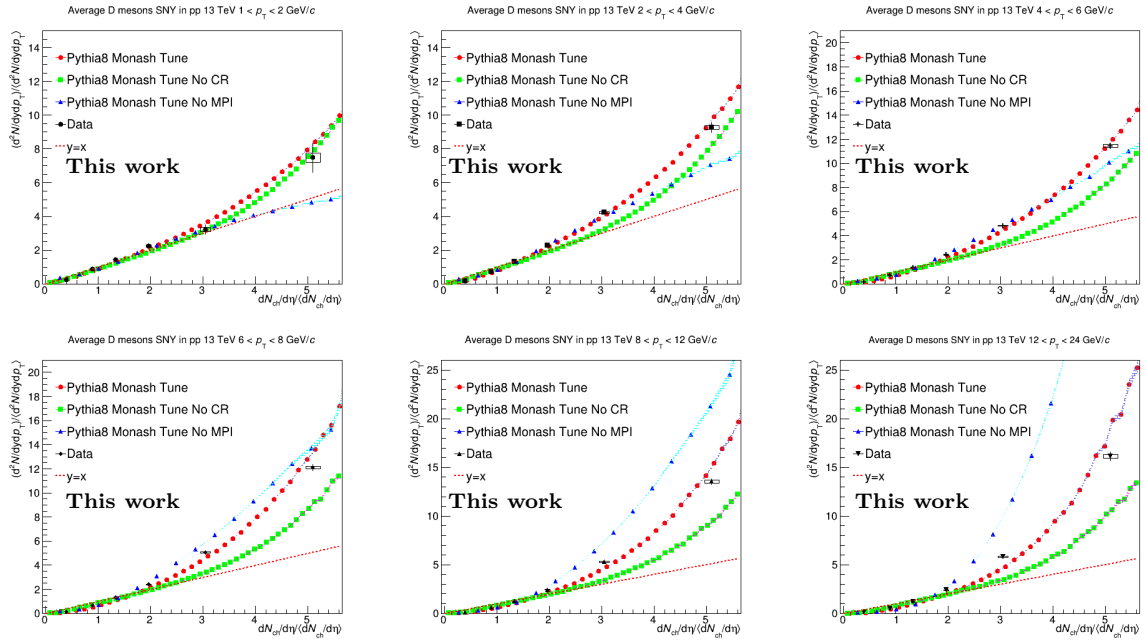


Figure 5.29: PYTHIA8 comparisons in p_{T} and multiplicity intervals of the average D mesons results using Monash with and without MPI, and colour reconnection.

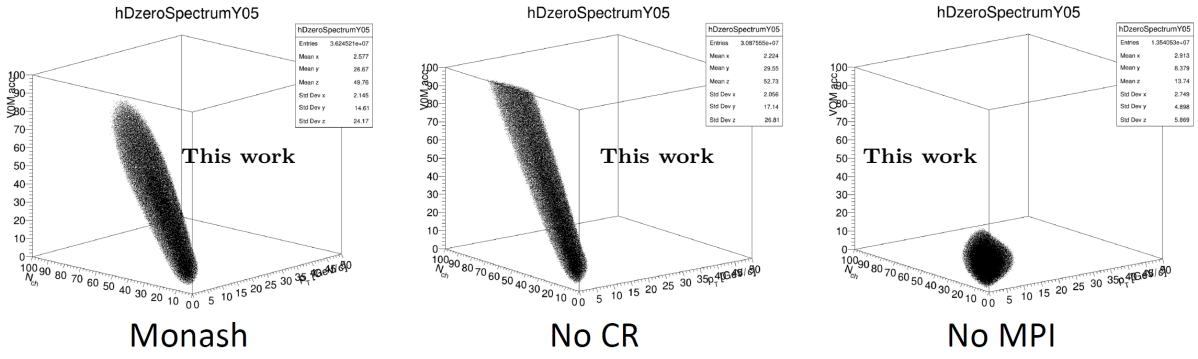


Figure 5.30: D^0 spectrum simulated with Monash enabling or not MPI and colour reconnection mechanisms.

The D^0 candidates as a function of charged particles, p_{T} and V0M acceptance (considered integrated in our study), illustrated in Fig. 5.30 for the D^0, are distributed in a similar way using Monash with and without colour reconnection (less events in the latter), but disabling MPIs causes the shape to converge in a much narrower bulk area.
Nevertheless, removing the MPI processes in PYTHIA causes inconsistencies with other predictions and data only in the highest multiplicity intervals, which are the most affected by multiple parton interactions. As a conclusion, from this results it emerges that

MPI description is a necessary ingredient to describe data and disabling them brings to inconsistent results. The enhanced colour reconnection models (plus MPI) are able to describe the measurements, with the exception of Mode 3 at high multiplicity.
For the spherocity dependent analysis, the same averaging procedure of the D mesons results was performed, and the obtained results in the three multiplicity intervals considered are shown in Fig. 5.31.

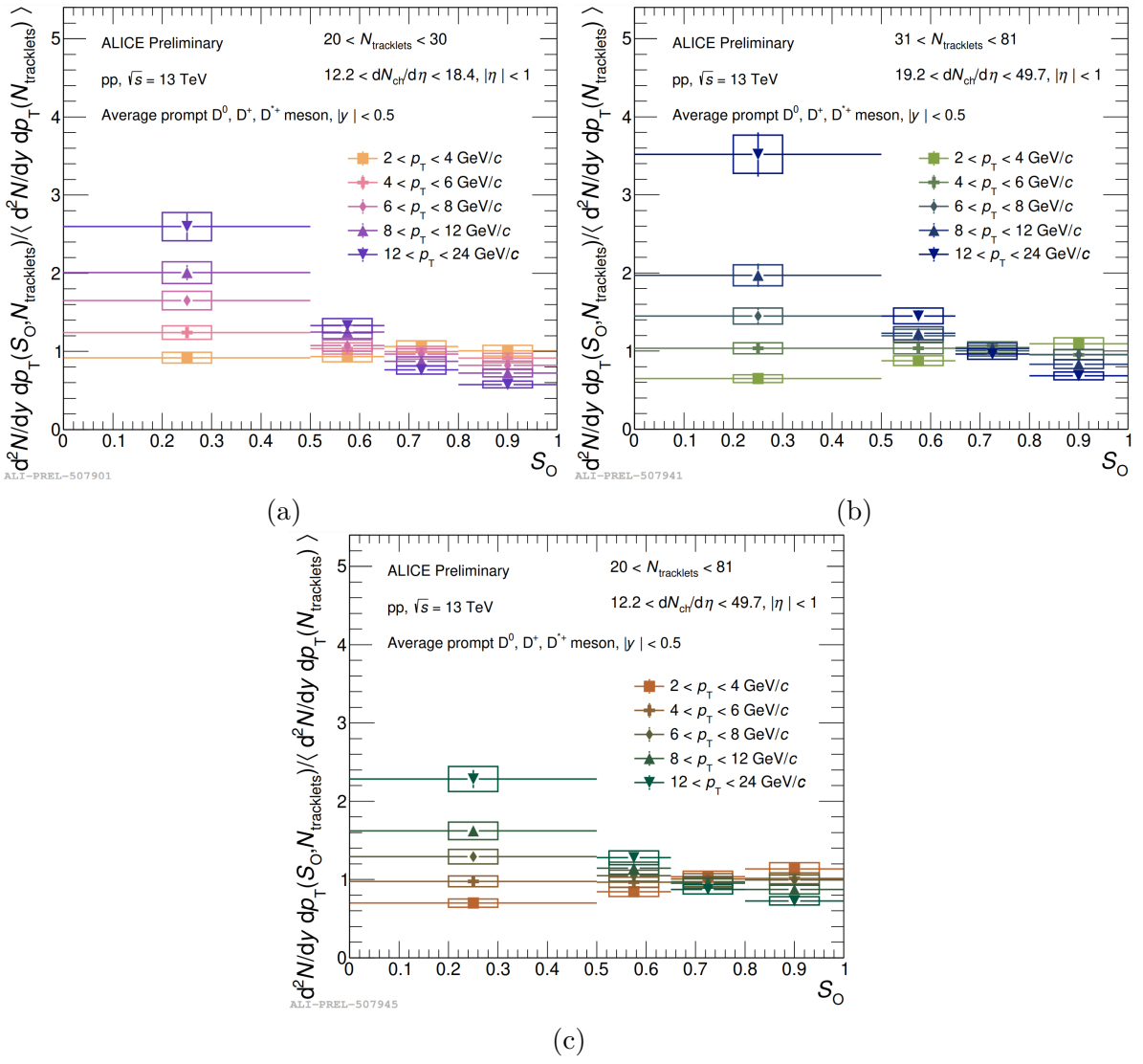


Figure 5.31: Self-normalised yields of average D mesons as a function of spherocity in different p_{T} intervals and for tracklets intervals of 20-30 (top left), 31-81 (top right) and 20-81 (bottom).

At low p_{T} and high multiplicity a slowly increasing trend is noticeable as a function of spherocity, which is flat instead at low multiplicity. In the other cases the self-normalised yields have a decreasing shape which becomes faster with higher p_{T} intervals, for all

the multiplicity classes considered. So, from this preliminary study it seems that the contribution of isotropic events in the production of prompt D mesons is very similar to jet-like ones for very low p_{T} intervals, while the latter are dominant for $p_{\mathrm{T}} > 6$ GeV/c. In particular, taking into account the integrated multiplicity interval, which has smaller statistical uncertainty, a factor larger than 2 is observed between the self-normalised yield values obtained at the highest p_{T} between the lowest and highest spherocity intervals (jet-like and isotropic limits).

A preliminary Monte Carlo study was performed on the analysis using PYTHIA8 with Monash and the enhanced colour reconnection tunes Mode 0/2/3. The results for the 20-81 multiplicity interval are shown in Fig. 5.32, illustrating a good compatibility among the simulation estimates and the experimental data within the uncertainties in the whole p_{T} range.

In addition, the measurements were also compared with the Monash tune disabling colour reconnection and MPI, as done for the spherocity integrated analysis (Fig. 5.33).

Disabling multiple parton interaction causes the model predictions to deviate significantly from the experimental data and Monash results, while, overall, the results without colour reconnection enabled (but with MPI) are showing trends and values which are compatible to data.

From this study we can conclude that enabling MPI makes a significant difference leading to a rough understanding of the real physics underlying processes, while the basic colour reconnection mechanisms are important to perform a fine tuning of the predictions to experimental measurements.

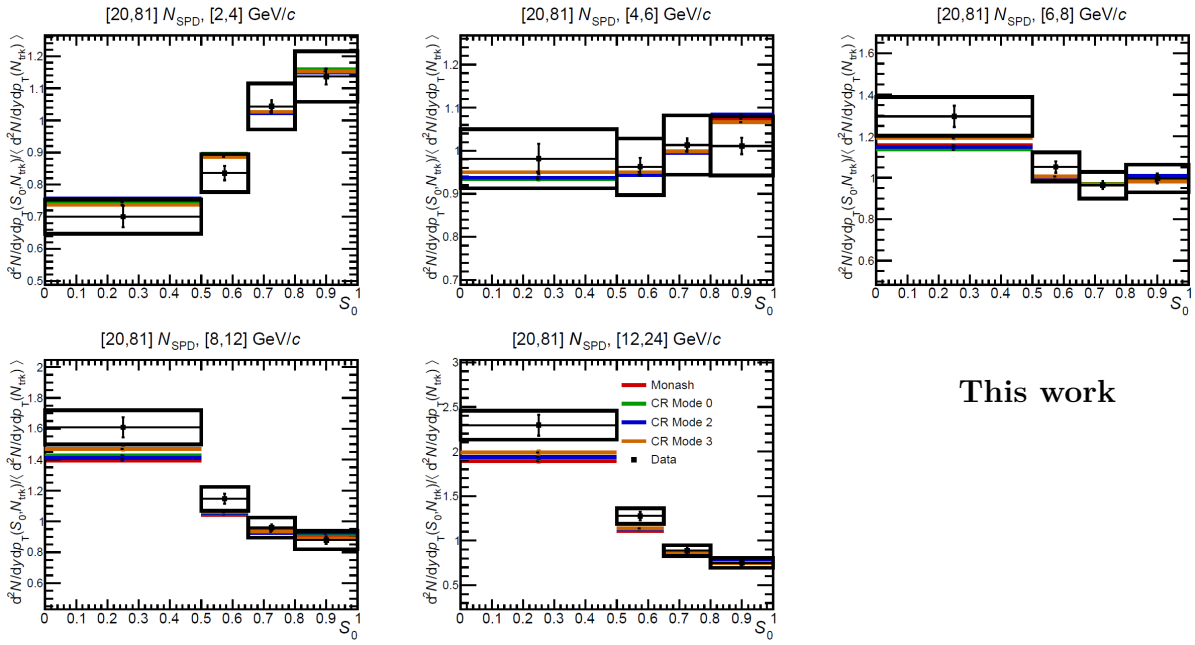


Figure 5.32: Preliminary comparison of average D mesons self-normalised yields as a function of spherocity in the 20-81 multiplicity interval with various tune of PYTHIA8.

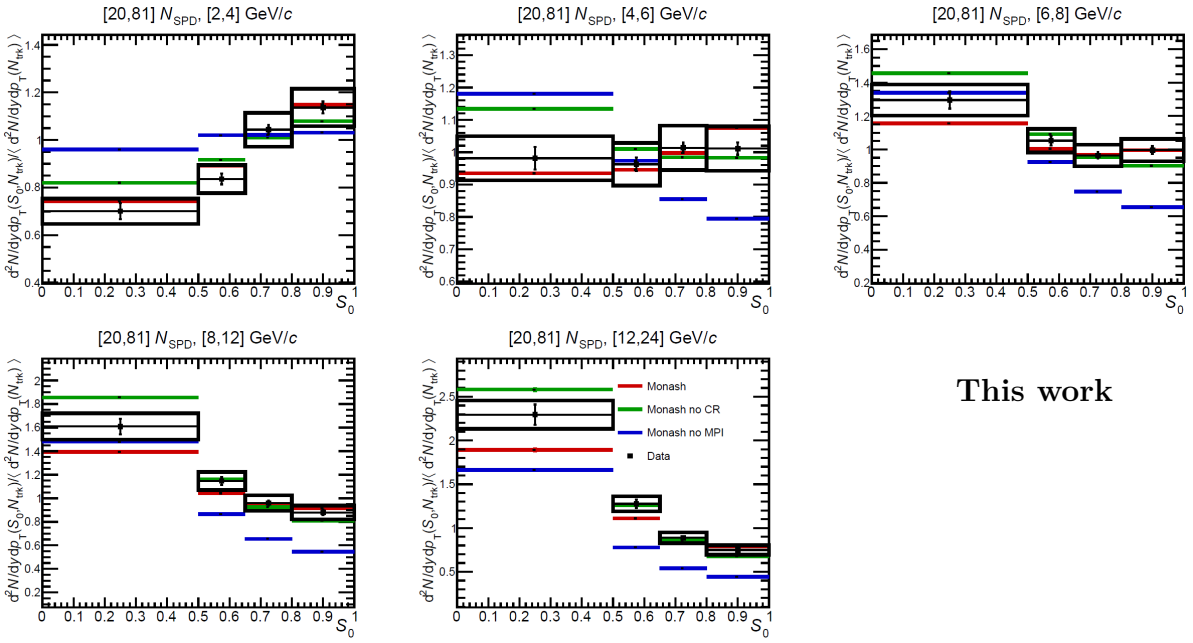


Figure 5.33: Preliminary comparison of average D mesons self-normalised yields as a function of spherocity in the 20-81 multiplicity interval with PYTHIA8 enabling and disabling colour reconnection and MPI from Monash 2013 tune.

5.8 Summary

The production of prompt D^{*+} meson as a function of multiplicity and spherocity in pp collisions was studied in this chapter, discussing corrections and systematic uncertainties applied to the measurements. The results were combined with D^+ and D^0 ones, obtained by other analysis groups, averaging them via a weighted procedure that was explained in Eq. 5.12.

Our measurements show an increase in the average self-normalised yields of D mesons with multiplicity in the spherocity integrated case, with a trend faster than linear. The enhancement is emphasised in higher p_T intervals, as shown in Fig. 5.24a, with a faster trend compared to lower p_T. Experimental data are compatible both with previous published results in pp collisions at $\sqrt{s} = 7$ TeV, and with other studies of heavy-flavour production, namely J/ψ and electrons from heavy-flavour decays.

EPOS3 [60] estimates, enabling the hydrodynamical evolution of the system, are able to properly describe the trend of results in three different p_T intervals, while disabling hydrodynamics leads to underestimating data, especially at high multiplicity. In addition, the Colour Glass Condensate [142, 143] model overestimates results in all the considered p_T intervals.

A full study was conducted with PYTHIA8, simulating the results with the default tune (Monash 2013), Modes 0/2/3 of the enhanced colour reconnection tune of the event generator [46], and two tunes in which MPI and colour reconnection were removed from

Monash. Both Monash and the enhanced colour reconnection modes describe well the results and are consistent with each other, although Mode 3 tends to underestimate data. The comparison with colour reconnection disabled shows a trend that is coherent with our measurement, but the values are underestimated, even compared to Monash, showing the importance of colour reconnection. Disabling MPI, instead, generates predictions whose trends become increasingly faster (higher slope) with multiplicity and p_{T}. The events generated with this configuration have a completely different shape in the distribution of the tracks compared to Monash with and without CR (Fig. 5.30). However the results are mainly incompatible with data and other predictions at high multiplicity, in which MPI has the highest contribution. All contributions considered, these results clearly indicate the main role of MPI in describing the observed trend with multiplicity.
The self-normalised yields as a function of multiplicity and spherocity in p_{T} intervals show a flat trend (slowly increasing) at low p_{T} and low (high) multiplicity. On the other hand, at higher p_{T} intervals, they have decreasing trends as a function of spherocity, as shown by Fig. 5.31. This behaviour is expressed in all multiplicity classes, but it is more dominant at the highest one, with a faster decreasing shape for the highest p_{T} interval. From this study, jet-like events seem to be predominant at high p_{T} in all multiplicity classes, as opposed to isotropic ones. On the other hand, at low p_{T} the uncertainties of the measurements do not allow us to draw a firm conclusion on the dominant shape of the production. They might tend to favour isotropic events, as shown in high multiplicity, or have a flat distribution as a function of spherocity.
The obtained results were compared with PYTHIA8 with Monash and enhanced colour reconnection Mode 0/2/3, in the widest multiplicity interval (20-81 N_{trk}), showing consistent trends and values, within uncertainties, with our results (Fig. 5.32).
The comparison with PYTHIA8 with Monash disabling MPI and CR highlights again the important role of multiple parton interactions in the proper description of the events. As a matter of fact results without this mechanism are not able to describe even the shape of data in some p_{T} intervals, while removing CR (with MPI enabled) seems to affect only predictions at low spherocity values. Combining both in the Monash 2013 tune provides values and shapes that are coherent with experimental results.
The results of self-normalised yields as a function of spherocity were approved only recently by the ALICE Collaboration, so a deeper study with other Monte Carlo event generators was not feasible at the time of this report.
From both the analyses, it is interesting to note that the baryon enhancement mechanisms included in the PYTHIA8 enhanced colour reconnection tunes (discussed in Sec. 2.2.1) do not influence significantly the production of D mesons, maintaining the compatibility of the estimates as for the Monash 2013 tune.
Finally, a deeper study with more data is needed to deepen our knowledge on the subject, improving the uncertainties and possibly selecting finer spherocity and multiplicity intervals, which could be performed with Run 3 data of LHC.

Conclusions

The main target of this thesis was to study QCD processes that occur in pp collisions at the LHC. In the ALICE experiment pp collisions are important because they provide a reference to heavy-ion collision measurements in which a QGP might form, but they are fundamental also for testing perturbative QCD calculations, hadronisation models, and the role of multi-parton interaction in the high-density parton state formed in the collisions.
The main phenomenological models and event generators used to describe these processes were presented, in order to better understand their differences in the comparison with the experimental results.
Furthermore, the main components of the ALICE detector used for the analyses were described, together with the software tools that allowed me to reconstruct the events based on the Run 2 datasets.
The first measurement of the $\Sigma_c^{0,++}$ baryon in hadronic collisions was an important part of my PhD work. It was performed together with Λ_c^+ in the $\Lambda_c^+ \to pK_s^0$ decay channel and combined with results obtained in the $\Lambda_c^+ \to pK^+\pi^-$ channel by another analysis group.
The final results refer to the sum of the three isospin states including also the Σ_c^+ by adding a 3/2 factor in the cross section formula, assuming a symmetric contribution of the isospin states to the production of the Σ_c baryon.
Apart from the cross section of Σ_c and Λ_c^+, the published results [118] also contain the cross section of the D^0 meson in pp collisions at $\sqrt{s} = 13$ TeV, with which the baryon-to-meson ratios were calculated. These were found to be larger than those obtained in e^+e^- and ep collisions, confirming the non-universality of fragmentation fractions in different collision systems.
The presented results provide important constraints on models aiming to explain the observed increase of charm baryons in a parton-rich environment, either increasing baryon-formation probability via enhanced colour reconnection or coalescence mechanisms, or assuming feed-down from yet-unobserved higher-mass baryon states.
Furthermore, the D meson self-normalised yields as a function of multiplicity and spherocity in pp collisions at $\sqrt{s} = 13$ TeV were studied in this dissertation. My contribution was mainly on the D^{*+} meson, whose results were later averaged with D^+ and D^0 results obtained by other members of the Collaboration.
The spherocity integrated results show an enhancement of the self-normalised yields with

multiplicity, with trends faster than linear which increase their rapidity in higher p_{T} intervals. The measurements are compatible with those obtained at $\sqrt{s} = 7$ TeV with D mesons, and with J/ψ and electrons from heavy-flavour decays results in pp collisions at $\sqrt{s} = 13$ TeV.
The comparison with Monte Carlo generators shows that EPOS3 with the hydrodynamical model is capable of properly describing the results, as well as PYTHIA8 with Monash and enhanced colour reconnection Mode 0 and 2. On the other hand, CGC, EPOS3 without hydrodynamics, and PYTHIA8 Mode 3 provide worse estimates compared to the other models.
Disabling separately MPI and CR in the Monash tune provides insight on the importance of these mechanisms for the description of the events: results without MPI fail to predict the trends of our result, while disabling the basic colour reconnection mechanism of PYTHIA8 allows the model to provide a good reproduction of the shape of the measurement, even if the overall description of data worsens. The results without MPI must be further analysed, since the disagreement between data and models is mainly observed at high multiplicity, in which a smaller amount of data is available, due to the shape of the generated tracks.
The analysis as a function of spherocity was performed using wider multiplicity intervals in order to achieve a higher statistics. The measurements show that jet-like events are dominant at high-p_{T} in all the multiplicity classes showing a decreasing trend of the self-normalised yields as a function of spherocity, while in lower p_{T} intervals the trends seem to be slowly increasing in high multiplicity and flattening in the lower class. The integrated 20-81 N_{trk} class also shows a slowly increasing trend, but more events are needed to obtain a clear trend of the results with a better precision of the measurements.
PYTHIA8 with Monash and the enhanced colour reconnection tune is able to properly describe our measurements with all the configurations used. However, the comparison with PYTHIA8 with and without MPI, and CR shows that both mechanisms are important to provide a good estimate of data, but, as in the spherocity integrated results, it seems that MPI has the strongest influence on the predictions. As a matter of fact, disabling the MPI causes the estimates to be unable to describe even the trends of the experimental results.
The work performed during my PhD will be expanded with future analyses of the larger statistics provided by Run 3 datasets, which will allow the results to reach a higher precision, especially in the intervals in which only few events were recorded (as in the D^{*+} analysis). In addition, the improved tracking resolution and the larger accessible samples of events, provided by the upgraded experiment after the LS2, will be exploited to improve the accuracy of current results on charmed baryons, including the $\Sigma_{\mathrm{c}}^{0,+,++}$.

Appendix A

Other research activities

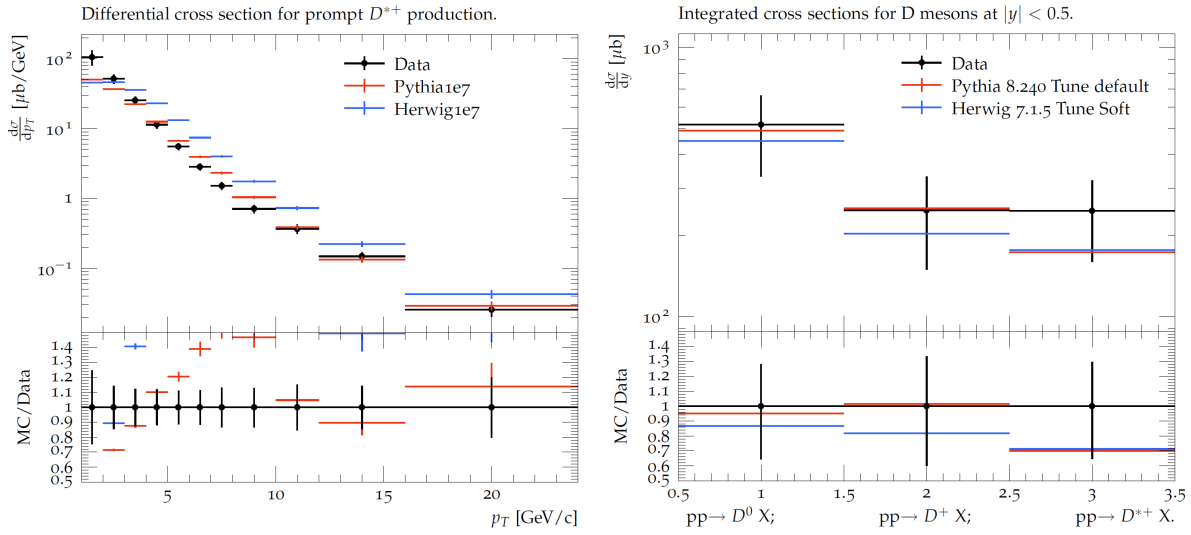
In this appendix additional research activities, in which I had a leading role, are described. They were carried out during the PhD enrolment, and they are not directly related to the main topic of this dissertation.

A.1 RIVET

Robust Indipendent Validation of Experiment and Theory (RIVET) [144] is a framework based on C++ and Python, which implements a series of valuable functions to make simple comparisons between experimental published data and predictions from Monte Carlo event generators. The published results are obtained through datasets from the Durham High-Energy Physics Database (HEPData), which are uploaded by the analysers in YAML [145] format at the time a paper is published on an international journal.
One of the main features of the RIVET framework is the possibility to obtain comparisons of Monte Carlo predictions with data in a short time, without the need to create different analyses macros for each generator. With RIVET the user creates a single C++ macro, independently on the event generator used. The macro analyses events written in HEPMC format[1] (already available for the most commonly used Monte Carlo generators, such as PYTHIA8 and HERWIG7). It is responsibility of the user to implement in the macro the same exact kinematical cut applied in the published analyses, in order to ensure a proper comparison is performed.
It is a standardised way to compare the results of different experiments with the Monte Carlo predictions and, from a theoretical point of view, allows users to validate their own models and further develop Monte Carlo generators.
During the PhD, I led the preparation and development of the specialised RIVET code needed for several published ALICE results: D mesons [147, 129], Λ_c^+ [65] and the $\Sigma_c^{0,+,++}$ [118] production (discussed in Chap. 4). The macros were also published inside the

[1] It is a simplified object oriented event record written in C++ for High Energy Physics MC event generators [146].

RIVET website.

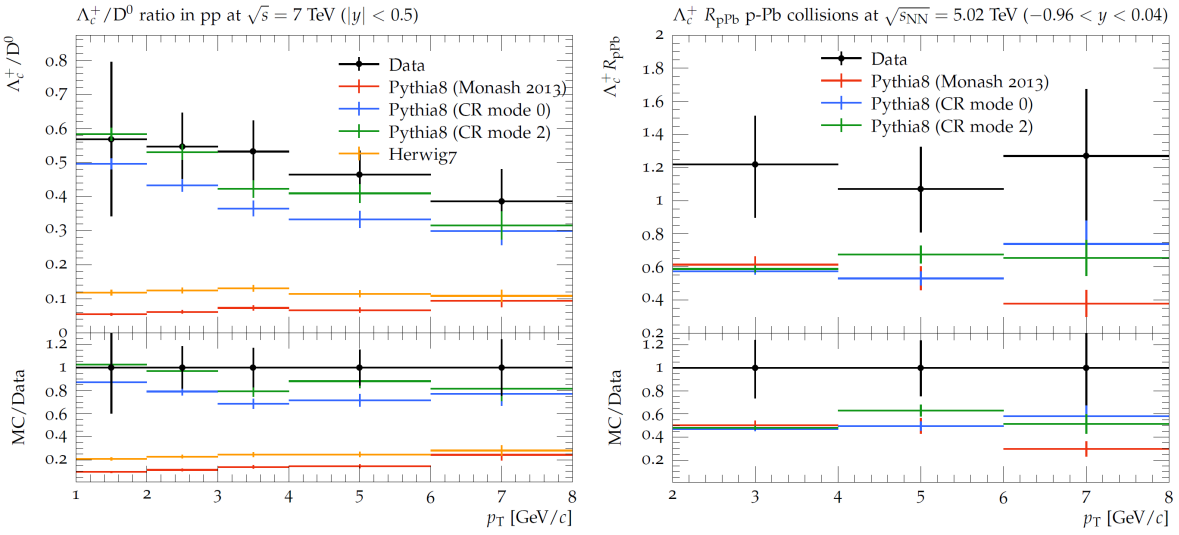


(a) (b)

Figure A.1: Differential cross section of prompt D^{*+} (a) and integrated cross section of D^0, D^+ and D^{*+} (b) mesons in pp collisions at $\sqrt{s} = 7$ TeV [129] compared to PYTHIA8 with Monash 2013 tune and HERWIG7 with the default tune using RIVET.

Furthermore, the experience on this software was useful for being then the supervisor of two Summer Students, in the Summer of 2022 at CERN, who developed in two months two standalone macros of ALICE published analyses.
An example of the RIVET outputs regarding two D mesons papers is shown in Fig. A.1a and A.1b, which illustrate, respectively, the differential production cross section of D^{*+} as a function of p_T and the integrated cross section of the D^0, D^+ and D^{*+} mesons in pp collisions at $\sqrt{s} = 7$ TeV compared to predictions from PYTHIA8 with Monash 2013 tune, and HERWIG7 using the default tune, which includes all the soft and hard QCD processes. In order to obtain the results, 10 million events were generated, and PYTHIA8 seems to be closer to the experimental data. However, even though the integrated cross sections are described within the uncertainties, both models are not consistent with the published ALICE results for the D^{*+} differential cross section (and a similar behaviour is shown also for D^0 and D^+).
The framework also allows us to perform ratios of different observables and particle species, and ratios of the production in different collision systems (by analysing multiple simulations and combining the outputs), as shown in the results obtained for the Λ_c^+ baryon.
Figure A.2a contains the experimental results of Λ_c^+/D^0 ratio obtained by the ALICE Collaboration in pp collisions at $\sqrt{s} = 7$ TeV [65], compared already to various Monte Carlo models in Fig. 2.13b. The RIVET computed plot shows both the PYTHIA8 and HERWIG default tunes, and, in addition, it includes the results obtained with

PYTHIA8 and two modes of the enhanced colour reconnection tune [46]. Both default tunes underestimate the ratio, while the colour reconnection tunes predict values similar to the experimental results.



(a) (b)

Figure A.2: Λ_c^+/D^0 ratio in pp collisions at $\sqrt{s} = 7$ TeV (a) and Λ_c^+ R_{pPb} in pp and p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV (b) [65] compared to different models by RIVET.

Estimates from PYTHIA8 predictions (both with Monash, and colour reconnection tunes) are also shown in Fig. A.2b, where the R_{pPb} of the Λ_c^+ baryon in pp and p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV at midrapidity is expressed as a function of p_T. The obtained predictions, in wider p_T bins compared to the analysis at $\sqrt{s} = 7$ TeV, are compatible with each other, but they are not coherent with the experimental data, showing a difference up to a factor 2. This inconsistency might depend on the Angantyr model, which is used in PYTHIA8 simulations for the description of heavy-ion collisions [148].

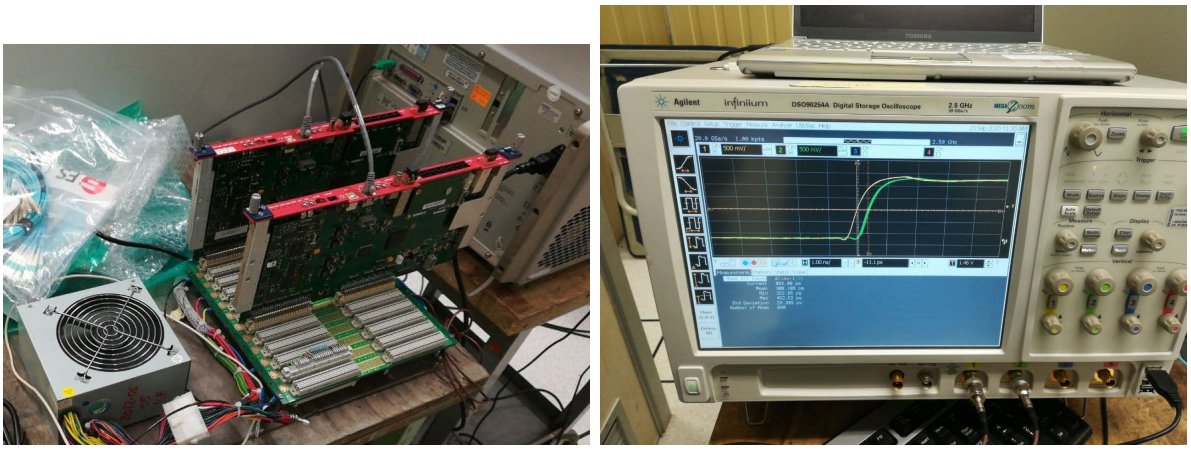
A.2 ALICE-TOF Clock Alignment procedures

During LS2 the whole trigger and clock distribution system was redesigned and moved from the previous standard Timing, Trigger and Control (TTC) system to transmission of the clock and trigger over GigaBit Transceiver (GBT) links. In addition, the TOF detector, since LHC Run 1, has a dedicated network for the distribution of the clock. When the LHC clock is received by the ALICE TTC Machine Interface it is routed both to the central CTP and to the TOF crates, without any additional encoding to minimise the clock jitter. From the CTP the clock via the TTC-PON network is then routed to all Common Readout Units (CRUs) and, there, encoded over GBT links down to the detectors. The DRM2 cards therefore receive two (equal) clocks: the *pure* LHC clock

and the GBT clock. The former is then used by the DRM2 to distribute the clock to the TDCs, while the latter, via the GBT link, is used for all synchronization signals and as clock for the data handshake with the DAQ. It is therefore of utmost importance to ensure locally in the crate a proper phase alignment between the two clocks, and a minimised absolute delay among the different TOF crates.
Even though such synchronisation was achieved during Run 1 commissioning, the re-cabling of the clock and trigger network during LS2 implied the need of a campaign of measurements to re-align the signals.
As part of the commissioning of the TOF detector during LS2 and before the data taking of Run 3 started, all the readout crates of TOF needed to have an aligned clock below a ns precision in order to achieve a raw synchronisation during data taking (O(10) ps alignment is then achieved with online and offline calibration working on data).
Furthermore, a thorough study was performed on the clock signals coming from different channels of the CRUs of the detector, in order to analyse the delays among them and to validate the timing values provided by the producer of the modules installed in our two First Level Processor (FLP) units.

A.2.1 CRU delay study

The procedures were first developed in Bologna and tested locally in a single-crate setup at the INFN laboratory. After they were validated, measurements were performed on the detector at CERN.



(a) (b)

Figure A.3: Test setup developed to study the delay between different channels of CRUs (a) and clock signals shown in the oscilloscope used for our study (b).

In order to obtain the clock signal from the CRUs, two detector readout boards (DRM2) were modified using a custom firmware that allowed them to work in the same VME backplane, which was powered by a simple computer power supply (Fig. A.3a).

These boards were fundamental in receiving the clock signal from the different transceiving optical ends of the CRUs channels and showing them as an output to be read, via LEMO cables, through a 2.5 GHz oscilloscope. This was used to measure the difference in the time measurements of the rising edge of the clock signals using its Δt function (Fig. A.3b). In addition, a laptop was setup with an NFS server and connected via a private network to the DRM2 boards to be able to access all the scripts needed for the tests. DRM2 can be controlled via a CAEN CONET2 link or through an ARM CPU running Linux, housed in a piggyback board and needing an NFS server to work. The ARM CPU was used during this lab test, at Bologna and at CERN.
CRUs provide the clock signal to the DRM2 boards via GBT optical link. There are 4 different CRUs in TOF, divided into 2 sections, each with 12 channels.
Three steps were performed to check the delays in the readout units:

1. delay between different channels of the same section of a CRU;
2. delay between different sections of the same CRU by considering one channel as reference;
3. delay between one output channel of one reference CRU with all the channels of the other CRUs.

The first tests were performed using the local clock of the readout units because the LHC clock was not ready yet (the procedures were performed at the end of September 2020, in the middle of LS2). When the clock from the machine was ready, the third test in the list was repeated.
Two breakout cables from Bologna were used and attached to the sections of the CRUs in order not to mess up the pre-existing cable management. These were connected to the sections to test and to the DRM2 boards in order to output the clock signals.
Each study considered the first channel of the breakout cables (labelled accordingly) as a reference.
The analysis led to the conclusion that the measurements performed at CERN were compatible with the preliminary results obtained in Bologna (on a smaller subset) and with the characteristics declared by the manufacturer. In this dissertation a full list of measurements will not be shown, but some ranges will be provided in Tab. A.1.

Test	**Delays at CERN**	**Delays in Bologna**
1	50–400 ps	100–250 ps
2	40–400 ps	400–550 ps
3 (int. clock)	10–400 ps	
3 (LHC clock)	10–400 ps	

Table A.1: CRUs delay ranges for different tests at CERN and in Bologna test setup.

A measure of the jitter of the LHC clock was also performed using one of the functions of the oscilloscope, obtaining an RMS value of ≈7–9 ps, in agreement with the expected 10 ps.

A.2.2 TOF detector delay measurements and alignment

The clock alignment procedure was performed in the ALICE cavern, analysing the clock signals from the GBT link and the LHC clock, without using the previous test setup. For this measurement, all LEMO outputs from DRM2 boards were used, since signals from both clock sources needed to be transmitted to the oscilloscope by the same board. The GBT clock signal is used as a backup for the LHC clock (with a compatible jitter) and it is also used by the GBTx ASIC chip to send and receive data to the CRUs (while the LHC clock is used for the VME and the HPTDC readout), so it is critical having their phase shift minimised.
The first test performed in the detector allowed us to find the local phase shift between LHC and the GBT clock signals using the oscilloscope and the same function used for the previously described tests. The GBT clock is transmitted to the board via optical links from the CRUs, while the LHC clock is sent to the DRM2 from the accelerator facilities via a dedicated input.
The firmware of the readout boards allows us to shift the phase of the GBT clock in order to align the clock signals, so a database was created listing all the phase alignment coefficients for each crate, which are loaded automatically at the DRM2 boot.
After this procedure, most of the crates reached a phase shift between the two clocks below 100 ps (starting from few ns), and only 7 on 72 readout boards remained with a higher delay (ranging from $\approx$ 200 to $\approx$ 700 ps, therefore still significantly lower than 1 ns). Any shift parameters (fine and coarse) applied to the GBT clock signal of these boards led to the worsening of the phase shift, so it was decided to keep the original default values.
The last measurements and alignment were performed on the absolute delay among the 72 TOF crates using the bunch reset signal, which is received every 89.1 μs by each crate and indicates the beginning of a new LHC orbit.
A reference crate was considered, and the delay of each readout board was calculated through the oscilloscope. The firmware allows us to shift the bunch reset using 25 ns steps, so the delay values were collected and the coefficients needed to align all the crates were computed offline. Subsequently, these constants were implemented in the database mentioned above. After the alignment procedure, all crates are now automatically configured to have a $\approx$1 ns absolute delay, with the exception of a single crate configured with a delay close to 3 ns. In addition, as explained above, the LHC clock and the GBT clock phase is aligned with a precision of O(10 ps) in each crate.

Bibliography

[1] F. Halzen and A. D. Martin, *Quarks and Leptons: An introductory course in Modern Particle Physics*. Wiley, 1984.

[2] I. J. R. Aitchison and A. J. G. Hey, *Gauge theories in particle physics: a practical introduction; 4th ed.* Boca Raton, FL: CRC Press, 2013.

[3] S. Bethke, "Experimental tests of asymptotic freedom", *Prog. Part. Nucl. Phys.*, vol. 58, pp. 351–386, 2007. arXiv: `hep-ex/0606035`.

[4] R. L. Workman *et al.* for the Particle Data Group Collaboration, "Review of Particle Physics", *PTEP*, vol. 2022, p. 083C01, 2022.

[5] H. Caines, "The search for critical behavior and other features of the QCD phase diagram – current status and future prospects", *Nuclear Physics A*, vol. 967, pp. 121–128, 2017.

[6] S. Borsanyi, "Thermodynamics of the QCD transition from lattice", *Nucl. Phys. A*, vol. 904-905, T. Ullrich, B. Wyslouch, and J. W. Harris, Eds., pp. 270c–277c, 2013. arXiv: `1210.6901 [hep-lat]`.

[7] M. Klegrewe and W. Unger, "Strong Coupling Lattice QCD in the Continuous Time Limit", *Phys. Rev. D*, vol. 102, no. 3, p. 034 505, 2020. arXiv: `2005.10813 [hep-lat]`.

[8] F. Karsch, "Lattice QCD at high temperature and density", *Lect. Notes Phys.*, vol. 583, W. Plessas and L. Mathelitsch, Eds., pp. 209–249, 2002. arXiv: `hep-lat/0106019`.

[9] J. D. Bjorken, "Highly relativistic nucleus-nucleus collisions: The central rapidity region", *Phys. Rev. D*, vol. 27, pp. 140–151, 1 Jan. 1983.

[10] R. Stock, "Relativistic Nucleus-Nucleus Collisions and the QCD Matter Phase Diagram", Jul. 2008. arXiv: `0807.1610 [nucl-ex]`.

[11] C. Markert, R. Bellwied, and I. Vitev, "Formation and decay of hadronic resonances in the QGP", *Physics Letters B*, vol. 669, no. 1, pp. 92–97, 2008.

[12] J. Adam *et al.* for the ALICE Collaboration, "Centrality dependence of the charged-particle multiplicity density at midrapidity in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV", *Phys. Rev. Lett.*, vol. 116, no. 22, p. 222 302, 2016. arXiv: `1512.06104 [nucl-ex]`.

[13] M. L. Miller *et al.*, "Glauber modeling in high energy nuclear collisions", *Ann. Rev. Nucl. Part. Sci.*, vol. 57, pp. 205–243, 2007. arXiv: nucl-ex/0701025.

[14] S. Acharya *et al.* for the ALICE Collaboration, "Measurement of D^0, D^+, D^{*+} and D_s^+ production in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV", *JHEP*, vol. 10, p. 174, 2018. arXiv: 1804.09083 [nucl-ex].

[15] A. M. Sirunyan *et al.* for the CMS Collaboration, "Nuclear modification factor of D^0 mesons in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV", *Phys. Lett. B*, vol. 782, pp. 474–496, 2018. arXiv: 1708.04962 [nucl-ex].

[16] K. Dusling, W. Li, and B. Schenke, "Novel collective phenomena in high-energy proton–proton and proton–nucleus collisions", *Int. J. Mod. Phys. E*, vol. 25, no. 01, p. 1 630 002, 2016. arXiv: 1509.07939 [nucl-ex].

[17] J. Adams *et al.* for the STAR Collaboration, "Evidence from d + Au measurements for final state suppression of high p_T hadrons in Au+Au collisions at RHIC", *Phys. Rev. Lett.*, vol. 91, p. 072 304, 2003. arXiv: nucl-ex/0306024.

[18] M. He and R. Rapp, "Hadronization and Charm-Hadron Ratios in Heavy-Ion Collisions", *Phys. Rev. Lett.*, vol. 124, no. 4, p. 042 301, 2020. arXiv: 1905.09216 [nucl-th].

[19] T. Song *et al.*, "Charm production in Pb + Pb collisions at energies available at the CERN Large Hadron Collider", *Phys. Rev. C*, vol. 93, no. 3, p. 034 906, 2016. arXiv: 1512.00891 [nucl-th].

[20] J. Rafelski and B. Müller, "Strangeness production in the quark-gluon plasma", *Phys. Rev. Lett.*, vol. 48, pp. 1066–1069, 16 Apr. 1982.

[21] E. Andersen *et al.* for the WA97 Collaboration, "Strangeness enhancement at mid-rapidity in Pb–Pb collisions at 158-A-GeV/c", *Phys. Lett. B*, vol. 449, pp. 401–406, 1999.

[22] F. Antinori *et al.* for the NA57 Collaboration, "Study of the production of strange and multi-strange particles in lead lead interactions at the CERN SPS: The NA57 experiment", *Nucl. Phys. A*, vol. 681, pp. 165–173, 2001.

[23] A. Tounsi and K. Redlich, "Strangeness enhancement and canonical suppression", Nov. 2001. arXiv: hep-ph/0111159.

[24] G. Agakishiev *et al.* for the STAR Collaboration, "Strangeness Enhancement in Cu+Cu and Au+Au Collisions at $\sqrt{s_{NN}} = 200$ GeV", *Phys. Rev. Lett.*, vol. 108, p. 072 301, 2012. arXiv: 1107.2955 [nucl-ex].

[25] B. B. Abelev *et al.* for the ALICE Collaboration, "Multi-strange baryon production at mid-rapidity in Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV", *Phys. Lett. B*, vol. 728, pp. 216–227, 2014. arXiv: 1307.5543 [nucl-ex].

[26] J. Adam *et al.* for the ALICE Collaboration, "Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions", *Nature Phys.*, vol. 13, pp. 535–539, 2017. arXiv: 1606.07424 [nucl-ex].

[27] M. L. Mangano, "Two lectures on heavy quark production in hadronic collisions", *Proc. Int. Sch. Phys. Fermi*, vol. 137, pp. 95–137, 1998. arXiv: hep-ph/9711337.

[28] J. C. Collins, D. E. Soper, and G. Sterman, "Factorization of hard processes in QCD", in *Perturbative QCD*. 1989, pp. 1–91.

[29] R. D. Ball *et al.* for the NNPDF Collaboration, "Parton distributions from high-precision collider data", *Eur. Phys. J. C*, vol. 77, no. 10, p. 663, 2017. arXiv: 1706.00428 [hep-ph].

[30] G. Altarelli and G. Parisi, "Asymptotic freedom in parton language", *Nuclear Physics B*, vol. 126, no. 2, pp. 298–318, 1977.

[31] S. Acharya *et al.* for the ALICE Collaboration, "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*, vol. 79, no. 5, p. 388, 2019. arXiv: 1901.07979 [nucl-ex].

[32] S. Acharya *et al.* for the ALICE Collaboration, "Measurement of beauty and charm production in pp collisions at $\sqrt{s} = 5.02$ TeV via non-prompt and prompt D mesons", *JHEP*, vol. 05, p. 220, 2021. arXiv: 2102.13601 [nucl-ex].

[33] M. Cacciari *et al.*, "Theoretical predictions for charm and bottom production at the LHC", *JHEP*, vol. 10, p. 137, 2012. arXiv: 1205.6344 [hep-ph].

[34] I. Helenius and H. Paukkunen, "Revisiting the D-meson hadroproduction in general-mass variable flavour number scheme", *JHEP*, vol. 05, p. 196, 2018. arXiv: 1804.03557 [hep-ph].

[35] M. Cacciari and M. Greco, "Large p_T hadroproduction of heavy quarks", *Nucl. Phys. B*, vol. 421, pp. 530–544, 1994. arXiv: hep-ph/9311260.

[36] S. Acharya *et al.* for the ALICE Collaboration, "Λ_c^+ production in pp and in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV", *Phys. Rev. C*, vol. 104, no. 5, p. 054 905, 2021. arXiv: 2011.06079 [nucl-ex].

[37] L. Gladilin, "Fragmentation fractions of c and b quarks into charmed hadrons at LEP", *Eur. Phys. J. C*, vol. 75, no. 1, p. 19, 2015. arXiv: 1404.3888 [hep-ex].

[38] S. Alioli *et al.*, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP*, vol. 06, p. 043, 2010. arXiv: 1002.2581 [hep-ph].

[39] J. Bellm *et al.*, "Herwig 7.0/Herwig++ 3.0 release note", *Eur. Phys. J. C*, vol. 76, no. 4, p. 196, 2016. arXiv: 1512.01178 [hep-ph].

[40] C. Bierlich *et al.*, "A comprehensive guide to the physics and usage of PYTHIA 8.3", Mar. 2022. arXiv: 2203.11601 [hep-ph].

[41] S. Acharya *et al.* for the ALICE Collaboration, "Charm-quark fragmentation fractions and production cross section at midrapidity in pp collisions at the LHC", *Phys. Rev. D*, vol. 105, no. 1, p. L011103, 2022. arXiv: 2105.06335 [nucl-ex].

[42] B. Andersson *et al.*, "Is there screwiness at the end of the QCD cascades?", *JHEP*, vol. 09, p. 014, 1998. arXiv: hep-ph/9807541.

[43] C. Bierlich, "Rope Hadronization, Geometry and Particle Production in pp and pA Collisions", Ph.D. dissertation, Lund University, 2017.

[44] B. Andersson *et al.*, "Parton fragmentation and string dynamics", *Physics Reports*, vol. 97, no. 2, pp. 31–145, 1983.

[45] T. Matsui and H. Satz, "J/ψ suppression by quark-gluon plasma formation", *Physics Letters B*, vol. 178, no. 4, pp. 416–422, 1986.

[46] J. R. Christiansen and P. Z. Skands, "String Formation Beyond Leading Colour", *JHEP*, vol. 08, p. 003, 2015. arXiv: 1505.01681 [hep-ph].

[47] V. Khachatryan *et al.* for the CMS Collaboration, "Strange Particle Production in pp Collisions at $\sqrt{s} = 0.9$ and 7 TeV", *JHEP*, vol. 05, p. 064, 2011. arXiv: 1102.4282 [hep-ex].

[48] P. Bartalini and J. R. Gaunt, Eds., *Multiple Parton Interactions at the LHC*. WSP, 2019, vol. 29.

[49] T. Sjöstrand and M. van Zijl, "A multiple-interaction model for the event structure in hadron collisions", *Phys. Rev. D*, vol. 36, pp. 2019–2041, 7 Oct. 1987.

[50] A. Ortiz Velasquez *et al.*, "Color Reconnection and Flowlike Patterns in pp Collisions", *Phys. Rev. Lett.*, vol. 111, no. 4, p. 042 001, 2013. arXiv: 1303.6326 [hep-ph].

[51] S. Acharya *et al.* for the ALICE Collaboration, "Λ_c^+ production in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV", *Phys. Lett. B*, vol. 793, pp. 212–223, 2019. arXiv: 1809.10922 [nucl-ex].

[52] J. Adam *et al.* for the STAR Collaboration, "First measurement of Λ_c baryon production in Au+Au collisions at $\sqrt{s_{\mathrm{NN}}} = 200$ GeV", *Phys. Rev. Lett.*, vol. 124, no. 17, p. 172 301, 2020. arXiv: 1910.14628 [nucl-ex].

[53] Y. Oh *et al.*, "Heavy baryon/meson ratios in relativistic heavy ion collisions", *Phys. Rev. C*, vol. 79, p. 044 905, 2009. arXiv: 0901.1382 [nucl-th].

[54] W. Zhao *et al.*, "Hydrodynamic collectivity in proton–proton collisions at 13 TeV", *Phys. Lett. B*, vol. 780, pp. 495–500, 2018. arXiv: 1801.00271 [nucl-th].

[55] Y. Han, T. Liu, and B.-Q. Ma, "Six-dimensional light-front Wigner distribution of hadrons", *Phys. Lett. B*, vol. 830, p. 137 127, 2022. arXiv: 2202.10359 [hep-ph].

[56] C. Peterson *et al.*, "Scaling Violations in Inclusive e^+e^- Annihilation Spectra", *Phys. Rev. D*, vol. 27, p. 105, 1983.

[57] V. Minissale, S. Plumari, and V. Greco, "Charm hadrons in pp collisions at LHC energy within a coalescence plus fragmentation approach", *Phys. Lett. B*, vol. 821, p. 136 622, 2021. arXiv: 2012.12001 [hep-ph].

[58] M. Lisovyi, A. Verbytskyi, and O. Zenaiev, "Combined analysis of charm-quark fragmentation-fraction measurements", *Eur. Phys. J. C*, vol. 76, no. 7, p. 397, 2016. arXiv: 1509.01061 [hep-ex].

[59] H. J. Drescher *et al.*, "Parton based Gribov-Regge theory", *Phys. Rept.*, vol. 350, pp. 93–289, 2001. arXiv: hep-ph/0007198.

[60] K. Werner *et al.*, "Analysing radial flow features in p-Pb and p-p collisions at several TeV by studying identified particle production in EPOS3", *Phys. Rev. C*, vol. 89, no. 6, p. 064 903, 2014. arXiv: 1312.1233 [nucl-th].

[61] K. Werner *et al.*, "Event-by-Event Simulation of the Three-Dimensional Hydrodynamic Evolution from Flux Tube Initial Conditions in Ultrarelativistic Heavy Ion Collisions", *Phys. Rev. C*, vol. 82, p. 044 904, 2010. arXiv: 1004.0805 [nucl-th].

[62] P. Braun-Munzinger, K. Redlich, and J. Stachel, "Particle production in heavy ion collisions", R. C. Hwa and X.-N. Wang, Eds., pp. 491–599, Apr. 2003. arXiv: nucl-th/0304013.

[63] M. He and R. Rapp, "Charm-Baryon Production in Proton-Proton Collisions", *Phys. Lett. B*, vol. 795, pp. 117–121, 2019. arXiv: 1902.08889 [nucl-th].

[64] 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*, vol. 84, p. 014 025, 2011. arXiv: 1105.0583 [hep-ph].

[65] S. Acharya *et al.* for the ALICE Collaboration, "Λ_c^+ production in pp collisions at $\sqrt{s} = 7$ TeV and in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV", *JHEP*, vol. 04, p. 108, 2018. arXiv: 1712.09581 [nucl-ex].

[66] A. Andronic *et al.*, "Charmonium and open charm production in nuclear collisions at SPS/FAIR energies and the possible influence of a hot hadronic medium", *Phys. Lett. B*, vol. 659, pp. 149–155, 2008. arXiv: 0708.1488 [nucl-th].

[67] A. Bazavov *et al.*, "The melting and abundance of open charm hadrons", *Phys. Lett. B*, vol. 737, pp. 210–215, 2014. arXiv: 1404.4043 [hep-lat].

[68] J. Song, H.-h. Li, and F.-l. Shao, "New feature of low p_T charm quark hadronization in pp collisions at $\sqrt{s} = 7$ TeV", *Eur. Phys. J. C*, vol. 78, no. 4, p. 344, 2018. arXiv: 1801.09402 [hep-ph].

[69] X. Gou *et al.*, "New insights into hadron production mechanism from p_T spectra in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. D*, vol. 96, no. 9, p. 094 010, 2017. arXiv: 1707.06906 [hep-ph].

[70] J. Adam *et al.* for the ALICE Collaboration, "D-meson production in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV and in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. C*, vol. 94, no. 5, p. 054 908, 2016. arXiv: 1605.07569 [nucl-ex].

[71] S. Gieseke, P. Stephens, and B. Webber, "New formalism for QCD parton showers", *JHEP*, vol. 12, p. 045, 2003. arXiv: hep-ph/0310083.

[72] G. Dissertori, I. G. Knowles, and M. Schmelling, *Quantum Chromodynamics: High Energy Experiments and Theory*. Oxford University Press, May 2009, ISBN: 9780199566419.

[73] M. Bahr *et al.*, "Herwig++ Physics and Manual", *Eur. Phys. J. C*, vol. 58, pp. 639–707, 2008. arXiv: 0803.0883 [hep-ph].

[74] A. Kupco, "Cluster hadronization in HERWIG 5.9", *Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting)*, pp. 292–300, Apr. 1998. arXiv: hep-ph/9906412.

[75] S. Gieseke, C. Rohr, and A. Siodmok, "Colour reconnections in Herwig++", *Eur. Phys. J. C*, vol. 72, p. 2225, 2012. arXiv: 1206.0041 [hep-ph].

[76] D. Reichelt, P. Richardson, and A. Siodmok, "Improving the Simulation of Quark and Gluon Jets with Herwig 7", *Eur. Phys. J. C*, vol. 77, no. 12, p. 876, 2017. arXiv: 1708.01491 [hep-ph].

[77] S. Gieseke, P. Kirchgaeßer, and S. Plätzer, "Baryon production from cluster hadronisation", *Eur. Phys. J. C*, vol. 78, no. 2, p. 99, 2018. arXiv: 1710.10906 [hep-ph].

[78] V. A. Abramovsky, V. N. Gribov, and O. V. Kancheli, "Character of Inclusive Spectra and Fluctuations Produced in Inelastic Processes by Multi - Pomeron Exchange", *Yad. Fiz.*, vol. 18, pp. 595–616, 1973.

[79] T. Sjostrand and P. Z. Skands, "Multiple interactions and the structure of beam remnants", *JHEP*, vol. 03, p. 053, 2004. arXiv: hep-ph/0402078.

[80] M. Bahr, S. Gieseke, and M. H. Seymour, "Simulation of multiple partonic interactions in Herwig++", *JHEP*, vol. 07, p. 076, 2008. arXiv: 0803.3633 [hep-ph].

[81] C. Bierlich and J. R. Christiansen, "Effects of color reconnection on hadron flavor observables", *Phys. Rev. D*, vol. 92, no. 9, p. 094 010, 2015. arXiv: 1507.02091 [hep-ph].

[82] C. Bierlich *et al.*, "Effects of Overlapping Strings in pp Collisions", *JHEP*, vol. 03, p. 148, 2015. arXiv: 1412.6259 [hep-ph].

[83] S. Acharya *et al.* for the ALICE Collaboration, "Measurement of the Cross Sections of Ξ_c^0 and Ξ_c^+ Baryons and of the Branching-Fraction Ratio BR($\Xi_c^0 \to \Xi^- e^+ \nu_e$)/BR($\Xi_c^0 \to \Xi^- \pi^+$) in pp collisions at 13 TeV", *Phys. Rev. Lett.*, vol. 127, no. 27, p. 272 001, 2021. arXiv: 2105.05187 [nucl-ex].

[84] J. Adam *et al.* for the ALICE Collaboration, "Measurement of charm and beauty production at central rapidity versus charged-particle multiplicity in proton-proton collisions at $\sqrt{s} = 7$ TeV", *JHEP*, vol. 09, p. 148, 2015. arXiv: 1505.00664 [nucl-ex].

[85] O. S. Brüning *et al.*, *LHC Design Report*. Geneva, Switzerland: CERN, 2004, vol. 1, ISBN: 9789290832249.

[86] O. S. Brüning *et al.*, *LHC Design Report*. Geneva, Switzerland: CERN, 2004, vol. 2, ISBN: 9789290832362.

[87] M. Benedikt *et al.*, *LHC Design Report*. Geneva, Switzerland: CERN, 2004, vol. 3, ISBN: 9789290832393.

[88] L. Evans and P. Bryant, "LHC machine", *JINST*, vol. 3, no. 08, S08001, 2008.

[89] K. Aamodt *et al.* for the ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST*, vol. 3, S08002, 2008.

[90] A. Fernández *et al.* for the ALICE Collaboration, "ACORDE a Cosmic Ray Detector for ALICE", *Nucl. Instrum. Methods Phys. Res., A*, vol. 572, 102–103. 2 p, Jun. 2006.

[91] G. Dellacasa *et al.* for the ALICE Collaboration, *ALICE Photon Multiplicity Detector (PMD): Technical Design Report* (Technical design report. ALICE). Geneva: CERN, 1999, https://cds.cern.ch/record/451099.

[92] P. Cortese *et al.* for the ALICE Collaboration, *ALICE forward detectors: FMD, TO and VO: Technical Design Report* (Technical design report. ALICE). Geneva: CERN, 2004, https://cds.cern.ch/record/781854.

[93] G. Dellacasa *et al.* for the ALICE Collaboration, *ALICE Inner Tracking System (ITS): Technical Design Report* (Technical design report. ALICE). Geneva: CERN, 1999, https://cds.cern.ch/record/391175.

[94] K. Aamodt *et al.* for the ALICE Collaboration, "Alignment of the ALICE Inner Tracking System with cosmic-ray tracks", *JINST*, vol. 5, P03003, 2010. arXiv: 1001.0502 [physics.ins-det].

[95] B. B. Abelev *et al.* for the ALICE Collaboration, "Performance of the ALICE Experiment at the CERN LHC", *Int. J. Mod. Phys. A*, vol. 29, p. 1 430 044, 2014. arXiv: 1402.4476 [nucl-ex].

[96] J. Alme *et al.*, "The ALICE TPC, a large 3-dimensional tracking device with fast readout for ultra-high multiplicity events", *Nucl. Instrum. Meth. A*, vol. 622, pp. 316–367, 2010, https://cds.cern.ch/record/451098. arXiv: 1001.1950 [physics.ins-det].

[97] G. Dellacasa *et al.* for the ALICE Collaboration, "ALICE: Technical design report of the time projection chamber", Jan. 2000.

[98] S. Acharya *et al.* for the ALICE Collaboration, "Dielectron production in proton-proton collisions at $\sqrt{s} = 7$ TeV", *JHEP*, vol. 09, p. 064, 2018. arXiv: 1805.04391 [hep-ex].

[99] J. Adam *et al.* for the ALICE Collaboration, "Particle identification in ALICE: a Bayesian approach", *Eur. Phys. J. Plus*, vol. 131, no. 5, p. 168, 2016. arXiv: 1602.01392 [physics.data-an].

[100] G. Dellacasa *et al.* for the ALICE Collaboration, "ALICE technical design report of the time-of-flight system (TOF)", Feb. 2000, https://cds.cern.ch/record/430132.

[101] P. Cortese *et al.* for the ALICE Collaboration, "ALICE: Addendum to the technical design report of the time of flight system (TOF)", Apr. 2002, https://cds.cern.ch/record/545834/.

[102] A. Akindinov *et al.*, "Performance of the ALICE Time-Of-Flight detector at the LHC", *Eur. Phys. J. Plus*, vol. 128, p. 44, 2013.

[103] J. Adam *et al.* for the ALICE Collaboration, "Determination of the event collision time with the ALICE detector at the LHC", *Eur. Phys. J. Plus*, vol. 132, no. 2, p. 99, 2017. arXiv: 1610.03055 [physics.ins-det].

[104] F. Carnesecchi for the ALICE Collaboration, "Performance of the ALICE Time-Of-Flight detector at the LHC", *JINST*, vol. 14, no. 06, p. C06023, 2019. arXiv: 1806.03825 [physics.ins-det].

[105] C. W. Fabjan *et al.* for the ALICE Collaboration, *ALICE trigger data-acquisition high-level trigger and control system: Technical Design Report* (Technical design report. ALICE). Geneva: CERN, 2004, https://cds.cern.ch/record/684651.

[106] F. Carena *et al.*, "The ALICE data acquisition system", *Nucl. Instrum. Meth. A*, vol. 741, pp. 130–162, 2014.

[107] R. Frühwirth, "Application of Kalman filtering to track and vertex fitting", *Nucl. Instrum. Meth. A*, vol. 262, no. 2, pp. 444–450, 1987.

[108] "ROOT", http://root.cern.

[109] "AliRoot", http://alice-offline.web.cern.ch.

[110] R. Brun *et al.*, "GEANT Detector Description and Simulation Tool", *CERN Program Library*, vol. CERN-W5013, Oct. 1994.

[111] S. Agostinelli *et al.*, "Geant4 — a simulation toolkit", *Nucl. Instrum. Meth. A*, vol. 506, no. 3, pp. 250–303, 2003.

[112] A. Ferrari *et al.*, "FLUKA: A multi-particle transport code (Program version 2005)", Oct. 2005, http://cds.cern.ch/record/898301.

[113] "LHC computing Grid. Technical design report", I. Bird *et al.*, Eds., Jun. 2005, http://cds.cern.ch/record/840543.

[114] "The Worldwide LHC Computing Grid ", "https://wlcg.web.cern.ch/.

[115] P. Saiz *et al.*, "AliEn — ALICE environment on the grid", *Nucl. Instrum. Meth. A*, vol. 502, no. 2, pp. 437–440, 2003.

[116] "Monalisa documentation", https://alien.web.cern.ch/content/alice-grid-monitoring-monalisa.

[117] R. Maciuła and A. Szczurek, "Production of Λ_c baryons at the LHC within the k_{T}-factorization approach and independent parton fragmentation picture", *Phys. Rev. D*, vol. 98, no. 1, p. 014 016, 2018. arXiv: 1803.05807.

[118] S. Acharya *et al.* for the ALICE Collaboration, "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.*, vol. 128, no. 1, p. 012 001, 2022. arXiv: 2106.08278 [hep-ex].

[119] S. Acharya *et al.* for the ALICE Collaboration, "ALICE 2016-2017-2018 luminosity determination for pp collisions at $\sqrt{s} = 13$ TeV", 2021, https://cds.cern.ch/record/2776672/.

[120] K. Albertsson *et al.*, "Machine Learning in High Energy Physics Community White Paper", *J. Phys. Conf. Ser.*, vol. 1085, no. 2, p. 022 008, 2018. arXiv: 1807.02876 [physics.comp-ph].

[121] P. B. Rodríguez *et al.*, "Calibration of the momentum scale of a particle physics detector using the Armenteros-Podolanski plot", *JINST*, vol. 16, no. 06, P06036, 2021. arXiv: 2012.03620 [physics.ins-det].

[122] J. Podolanski and R. Armenteros, "III. analysis of v-events", *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 45, no. 360, pp. 13–30, 1954.

[123] A. Hocker *et al.*, "TMVA - Toolkit for Multivariate Data Analysis", Mar. 2007. arXiv: physics/0703039.

[124] Y. Coadou, "Boosted decision trees", Mar. 2022. arXiv: 2206.09645.

[125] M. Faggin, "Measurement of heavy-flavour decay electrons and heavy-flavour baryon production with ALICE experiment at LHC", *PhD Thesis*, https://cds.cern.ch/record/2798337, 2021.

[126] J. Olivero and R. Longbothum, "Empirical fits to the Voigt line width: A brief review", *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 17, no. 2, pp. 233–236, 1977.

[127] R. Aaij *et al.* for the LHCb Collaboration, "Measurement of b hadron fractions in 13 TeV pp collisions", *Phys. Rev. D*, vol. 100, no. 3, p. 031 102, 2019. arXiv: 1902.06794 [hep-ex].

[128] S. Acharya *et al.* for the ALICE Collaboration, "Observation of a multiplicity dependence in the p_{T}-differential charm baryon-to-meson ratios in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B*, vol. 829, p. 137 065, 2022. arXiv: 2111.11948 [nucl-ex].

[129] S. Acharya *et al.* for the ALICE Collaboration, "Measurement of D-meson production at mid-rapidity in pp collisions at $\sqrt{s} = 7$ TeV", *Eur. Phys. J. C*, vol. 77, no. 8, p. 550, 2017. arXiv: 1702.00766 [hep-ex].

[130] H. Abramowicz *et al.* for the ZEUS Collaboration, "Measurement of charm fragmentation fractions in photoproduction at HERA", *JHEP*, vol. 09, p. 058, 2013. arXiv: 1306.4862 [hep-ex].

[131] L. Gladilin, "Fragmentation fractions of c and b quarks into charmed hadrons at LEP", *Eur. Phys. J. C*, vol. 75, no. 1, p. 19, 2015. arXiv: 1404.3888 [hep-ex].

[132] S. Acharya *et al.* for the ALICE Collaboration, "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.*, vol. 127, no. 20, p. 202 301, 2021. arXiv: 2011.06078 [nucl-ex].

[133] M. Niiyama *et al.* for the Belle Collaboration, "Production cross sections of hyperons and charmed baryons from e^+e^- annihilation near $\sqrt{s} = 10.52$~GeV", *Phys. Rev. D*, vol. 97, no. 7, p. 072 005, 2018. arXiv: 1706.06791 [hep-ex].

[134] S. Acharya *et al.* for the ALICE Collaboration, "Charged-particle production as a function of multiplicity and transverse spherocity in pp collisions at $\sqrt{s} = 5.02$ and 13 TeV", *Eur. Phys. J. C*, vol. 79, no. 10, p. 857, 2019. arXiv: 1905.07208 [nucl-ex].

[135] A. Ortiz, G. Paić, and E. Cuautle, "Mid-rapidity charged hadron transverse spherocity in pp collisions simulated with Pythia", *Nucl. Phys. A*, vol. 941, pp. 78–86, 2015. arXiv: 1503.03129 [hep-ph].

[136] A. Ortiz, "Experimental results on event shapes at hadron colliders", *Adv. Ser. Direct. High Energy Phys.*, vol. 29, pp. 343–357, 2018. arXiv: 1705.02056 [hep-ex].

[137] S. Prasad *et al.*, "Event topology and global observables in heavy-ion collisions at the Large Hadron Collider", *Sci. Rep.*, vol. 12, no. 1, p. 3917, 2022. arXiv: 2112.03892 [hep-ph].

[138] S. Acharya *et al.* for the ALICE Collaboration, "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*, vol. 81, no. 7, p. 630, 2021. arXiv: 2009.09434 [nucl-ex].

[139] R. Corke and T. Sjostrand, "Interleaved Parton Showers and Tuning Prospects", *JHEP*, vol. 03, p. 032, 2011. arXiv: 1011.1759 [hep-ph].

[140] R. Corke and T. Sjostrand, "Multiparton Interactions with an x-dependent Proton Size", *JHEP*, vol. 05, p. 009, 2011. arXiv: 1101.5953 [hep-ph].

[141] S. Acharya *et al.* for the ALICE Collaboration, "Multiplicity dependence of J/ψ production at midrapidity in pp collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B*, vol. 810, p. 135 758, 2020. arXiv: 2005.11123 [nucl-ex].

[142] F. Gelis *et al.*, "The Color Glass Condensate", *Ann. Rev. Nucl. Part. Sci.*, vol. 60, pp. 463–489, 2010. arXiv: 1002.0333 [hep-ph].

[143] I. Schmidt and M. Siddikov, "Production mechanisms of open-heavy flavor mesons", *Phys. Rev. D*, vol. 101, no. 9, p. 094 020, 2020. arXiv: 2003.13768 [hep-ph].

[144] C. Bierlich *et al.*, "Robust Independent Validation of Experiment and Theory: Rivet version 3", *SciPost Phys.*, vol. 8, p. 026, 2020. arXiv: 1912.05451 [hep-ph].

[145] "YAML Ain't Markup Language (YAML™) version 1.2", https://yaml.org/spec/1.2.2/.

[146] A. Buckley *et al.*, "The HepMC3 event record library for Monte Carlo event generators", *Comput. Phys. Commun.*, vol. 260, p. 107 310, 2021. arXiv: 1912.08005 [hep-ph].

[147] B. Abelev *et al.* for the ALICE Collaboration, "Measurement of charm production at central rapidity in proton-proton collisions at $\sqrt{s} = 7$ TeV", *JHEP*, vol. 01, p. 128, 2012. arXiv: 1111.1553 [hep-ex].

[148] C. Bierlich *et al.*, "The Angantyr model for Heavy-Ion Collisions in PYTHIA8", *JHEP*, vol. 10, p. 134, 2018. arXiv: 1806.10820 [hep-ph].