UNIVERSITY OF SALERNO

DEPARTMENT OF PHYSICS "E. R. CAIANIELLO"

Ph.D. COURSE IN
"MATHEMATICS, PHYSICS AND APPLICATIONS"
XXXIV CYCLE
CURRICULUM: PHYSICS

Λ_c^+ production vs the event charged-particle multiplicity in pp collisions at 13 TeV with ALICE at the LHC

Candidate:
Luigi Dello Stritto

Supervisors:
Prof. S. De Pasquale
Prof.ssa A. De Caro
Dott. G.M. Innocenti

Ph.D. Coordinator:
Prof. C. Attanasio

Abstract

The Quantum Chromo-Dynamics (QCD) theory predicts, in condition of very high temperature and/or energy density, a phase transition from the ordinary nuclear matter to a colour-deconfined medium called Quark–Gluon Plasma (QGP). The ALICE experiment was designed and optimized for the investigation of this hot and dense medium, produced via heavy-ion collisions at the Large Hadron Collider (LHC). Due to the very short lifetime of the QGP, its properties cannot be directly revealed and its characterization is done through indirect signals, obtained from the observation of the ordinary particles that emerge from the interaction region. In particular, charm quarks are effective probes used for the investigation of QGP. Due to their large masses, they are produced in hard scattering processes on a timescale shorter than the QGP formation time and therefore experience the whole system evolution. The measurement of charm-baryon production, and in particular the baryon-to-meson ratios, provides unique information to characterize novel mechanisms of hadron formation beyond in-vacuum fragmentation, e.g. coalescence, which are expected to be significant in presence of a medium characterized by free colour charges. Measurements of charm-baryon production in pp collisions are essential to establish a baseline for p-A and A-A collisions. In addition, they provide critical tests of perturbative QCD (pQCD) calculations and models of charm hadronisation in hadronic collisions.

The aim of the studies carried out in this thesis is the measurement of Λ_c^+ charmed baryon yield, employed for the estimation of the baryon-to-meson ratio Λ_c^+/D^0. The first measurements of the Λ_c^+ production yields and of the Λ_c^+/D^0 baryon-to-meson ratios as a function of the charged-particle pseudorapidity density are presented. The study allows the characterization of the evolution of the Λ_c^+/D^0 baryon-to-meson ratio from very low to high charged particle density and provides new experimental constraints on the production mechanisms in pp collisions. The analysed sample is collected in pp collisions at the energy in the centre-of-mass system of $\sqrt{s} = 13$ TeV with the ALICE detector. The measurement is performed by reconstructing the hadronic decay channel $\Lambda_c^+ \to \mathrm{pK}_s^0 \to \mathrm{p}\pi^+\pi^-$, exploiting selections on its decay topology and on the particle identification (PID) of the decay products, extracting the signal via an invariant mass analysis and correcting for its selection and reconstruction efficiency and for the detector acceptance. A machine learning algorithm based on Boosted Decision Trees (BDT) has been developed and is used in order to improve the signal extraction by optimally combining topological and PID variables that allow discriminating signal candidates from the combinatorial background. The results are compared with a theoretical model that explains the multiplicity dependence by a canonical treatment of quantum charges in the statistical hadronisation approach and with predictions from `PYTHIA` event generators that implement colour reconnection mechanisms beyond the leading colour approximation to model the hadronisation process.

Contents

Introduction

Heavy quarks (charm and beauty) are extraordinary probes to investigate the Quark Gluon Plasma, a medium characterized by deconfined quarks and gluons, which is produced in heavy-ion collisions. Due to their large masses, they can only be produced in the initial stages of the collision in hard scattering processes on a timescale shorter than the QGP formation time. Since their masses are larger than the temperature of the medium, indeed, they cannot be produced thermally or via string fragmentation. So, the heavy quarks experience the full evolution of the system. The measurement of charm-baryon production, and in particular the baryon-to-meson ratio, allows the investigation of the propagation, thermalization and hadronisation mechanisms of heavy quarks in the medium. Moreover, measurements of charm-baryon production in pp collisions are essential to establish a baseline for p-A and A-A collisions. In addition, they provide critical tests of perturbative QCD (pQCD) calculations and models of charm hadronisation in hadronic collisions.

In this thesis work, the first Λ_c^+ corrected yield measurement versus the event charged-particle multiplicity is presented. The analysed sample is collected in pp collisions at the energy in the centre-of-mass system of $\sqrt{s} = 13$ TeV with the ALICE detector. The study allows one to characterize the evolution of the Λ_c^+/D^0 baryon-to-meson ratio from very low to high charged particle density, to explore if the hadronisation mechanisms get modified as a function of the event multiplicity and to provide new experimental constraints on the nature of these modifications in pp collisions.

The thesis is organized as follows. In Chapter 1 the hadron matter phase transition,

from ordinary matter to QGP, and the probes experimentally employed to investigate its properties are described. The state of the art of the heavy-flavour measurements is presented in Chapter 2. The Chapter 3 is dedicated to an in-depth description of the experimental apparatus employed to collect the data analysed in this work: the ALICE detector at the LHC. All its components are illustrated in detail, with a major focus on those that are crucial for the Λ_c^+ analysis, subject of this thesis. In Chapter 4, the analysis strategy for the reconstruction of the Λ_c^+ through the decay channel $\Lambda_c^+ \rightarrow pK_s^0 \rightarrow p\pi^+\pi^-$ is carefully described. In particular, an innovative machine learning approach has been used. In the final sections, the systematic uncertainty evaluation and the achieved results are reported. The measured Λ_c^+ corrected yields and the Λ_c^+/D^0 ratios in pp collisions at $\sqrt{s} = 13$ TeV versus the event charged-particle multiplicity and the comparison with the theoretical model is discussed. Finally, in Chapter 5, the prospects for the heavy-flavour analyses with the next-generation detector ALICE 3, proposed to replace ALICE in LHC Run 5 and 6, are presented. In particular, some Λ_c^+ performance studies in the decay channel $\Lambda_c^+ \rightarrow pK^-\pi^+$ employed to optimize the detector configuration and to provide a baseline for more complex decay topologies are shown.

Chapter 1

The Quark Gluon Plasma

In the Standard Model (SM) [1], the strong interaction is described by the Quantum Chromodynamics (QCD) field theory. The QCD theory predicts, in condition of very high temperature and/or energy density, a phase transition between the ordinary nuclear matter and a colour deconfined medium called Quark–Gluon Plasma (QGP). This state of the nuclear matter is created and studied in the laboratory via ultra-relativistic heavy-ion collisions. In this chapter, the QCD theory will be introduced with particular attention to the QGP phase transition and properties.

1.1 Introduction to the QCD

QCD is a non-Abelian gauge theory (i.e. in which the generators of the symmetry group do not commute) based on the symmetry group $SU(3)$ that describes the interactions between quarks by the exchanging of gluons. Quarks are elementary particles of spin 1/2 and electric charge +2/3 or -1/3. There are six types of quarks, up (u), down (d), charm (c), strange (s), bottom (b), top (t), known as quark *flavour*. Up and down are the lightest quarks. The heavier ones are not stable and rapidly decay into the two light quarks. Because of this, up and down quarks are the most common in the universe, whereas strange, charm, bottom and top quarks can only be produced in high energy collisions (such as those involving cosmic rays and in particle accelerators). Each quark has an additional quantum number called *colour*,

which plays the role of charge of the strong interaction and can assume three values conventionally indicated with red, green and blue. Coloured object cannot exist in isolation (see sec. 1.3), so, in nature, quarks are always in composite particles called *hadrons*, which have zero colour charge. Hadrons are classified into *mesons*, made up of a quark-antiquark pair, and *baryons*, composed by three quarks (or antiquarks). Recent experimental evidence has also pointed to the existence of more exotic forms of hadronic matter, namely tetraquarks and pentaquarks [2, 3]. Apart of the quarks that determine the quantum numbers of hadrons, called *valence* quarks, any hadron may contain an indefinite number of virtual *sea* quarks and antiquarks that do not influence its quantum numbers.

Gluons are massless particles with spin 1. They are the gauge bosons of the strong interaction and therefore mediate the interactions between quarks. Being responsible for the attraction between quarks in hadrons, such as protons and neutrons, they guarantee the stability of the atomic nucleus. The non-Abelian nature of QCD leads to the gluon carrying colour charge, meaning gluons can self-interact. This property of gluon self-interactions represents the major difference to Quantum Electrodynamics (QED), the (Abelian) quantum field theory describing the electromagnetic interaction, where the photon does not carry electric charge. Moreover, gluons, like quarks, cannot exist in isolation but only in zero colour charge groups called *glueball*.

1.2 The QCD Lagrangian

In this section, the QCD lagrangian, that controls the quarks and gluons dynamics, is derived.

Each flavour of quark comes in three copies of different colour, q_f^r, q_f^g and q_f^b. These tree fields are grouped in the so-called colour triplet:

$$q_f = \begin{pmatrix} q_f^r \\ q_f^g \\ q_f^b \end{pmatrix} \tag{1.1}$$

Since quarks are fermions, the free Lagrangian to be used is the Dirac one:

$$\mathcal{L}_0 = \sum_f \bar{q}_f (i\not{\partial} - m_f) q_f \tag{1.2}$$

with q_f e $\bar{q}_f$ representing the quark- and antiquark-colour triplets, respectively. The free Lagrangian (1.2) possesses the global symmetry $SU(3)_c$[1], i.e. it is invariant under global special unitary transformations of quark and antiquark field colours. This symmetry is introduced to describe the lack of direct observation of coloured particles. Therefore, the colour can be changed without altering the observable. The operation of a $SU(3)_c$ transformation on a colour triplet is the following:

$$q_f \to q'_f = U q_f \qquad U \in SU(3)_c \qquad UU^\dagger = I \tag{1.3}$$

$$\bar{q}_f \to \bar{q}'_f = U^\dagger \bar{q}_f \qquad U^\dagger \in SU(3)_c \qquad UU^\dagger = I \tag{1.4}$$

The U operator admits the exponential representation:

$$U = e^{i\omega_a T_a} \qquad T_a = \frac{\lambda_a}{2} \qquad [T_a, T_b] = i f^{abc} T_c \tag{1.5}$$

where the T_a (with a = 1,...,8) are the eight generators of the $SU(3)_c$ group. In their fundamental representation they correspond to the Gell-Mann matrices λ_a. The f^{abc} are the structure constants of the group.

The Lagrangian (1.2) is invariant only for global transformations (1.5), i.e. in which the parameters ω_a do not depend on the coordinates. For local symmetry transformations, due to the derivative term in the Lagrangian, the gauge invariance for $SU(3)_c$ transformations is broken. To make the Langrangian (1.2) invariant even under local gauge transformation, the ordinary derivative, ∂, is replaced with the covariant derivative, D, defined in such a way as to remove the excess terms produced

[1] The special unitary group of degree n, $SU(n)$, is the group of $n \times n$ unitary matrices with determinant 1. The $SU(n)$ group admits $(n^2 - 1)$ generators.

when the ordinary derivative acts. The covariant derivative D is defined as:

$$D_\mu = \partial_\mu + i\alpha_s A_\mu \qquad\qquad A_\mu = A_\mu^a T^a \tag{1.6}$$

where α_s is the strong coupling constant and A_μ are the eight gluon fields. In order to make the Lagrangian invariant under local gauge transformations $SU(3)_c$, the covariant derivative must behave as follows:

$$D'_\mu q'_f = U(x) D_\mu q_f \tag{1.7}$$

To satisfy this condition, the gauge fields must transform as in the following:

$$A'_\mu = U(x) A_\mu U(x)^{-1} + \frac{i}{\alpha_s}(\partial_\mu U(x)) U(x)^{-1} \tag{1.8}$$

Therefore, the Lagrangian for quark field, invariant under local $SU(3)_c$ gauge transformations, will be:

$$\mathcal{L}_0 = \sum_f \bar{q}_f (i\not{D} - m_f) q_f \tag{1.9}$$

To describe the dynamics of the gluon fields A_μ as well, another term must be added in the Lagrangian. Similarly to what is done in the QED theory, the tensor $F_{\mu\nu}$ has been defined:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + i\alpha_s [A_\mu, A_\nu] \tag{1.10}$$

Unlike QED, however, the generators of the group do not commute and this implies the presence of an additional term in the tensor (1.10). Physically, this term corresponds to the fact that gluons, differently from photons, carry colour charges and so, they can self-interact. Starting from the $F_{\mu\nu}$ tensor, the simplest gauge invariant scalar that can be inserted in the Lagrangian, again in analogy with quantum electrodynamics, is $F_{\mu\nu}F^{\mu\nu}$. Hence, the Lagrangian of the gauge bosons will be:

$$\mathcal{L}_{gluon} = -\frac{1}{4} \sum_{a=1}^{8} F^{(a)\,\mu\nu} F^{(a)}_{\mu\nu} \tag{1.11}$$

Therefore, the final Lagrangian of the QCD will be:

$$\mathcal{L}_{QCD} = -\frac{1}{4}\sum_{a=1}^{8} F^{(a)\,\mu\nu} F^{(a)}_{\mu\nu} + \sum_{f} \bar{q}_f (i\not{D} - m_f) q_f \tag{1.12}$$

The first term of the (1.12) describes the free propagation of the eight gluon fields A^{a_μ} while the second one describes the free propagation of the quark fields q_f and their interactions with the gluon-gauge fields.

1.3 Confinement and asymptotic freedom

One of the greatest triumphs of quantum chromodynamics is the discovery that the strong coupling constant α_s is not constant, but depends on the distance between the interacting partons (or, similarly, by the momentum Q transferred in the interaction). The so-called "running coupling" can be understood via the vacuum polarization. The vacuum, according to Quantum Field Theory (QFT), is not a vacuum in the sense of being completely empty but instead contains virtual particles. In QED, a charge in vacuum attracts virtual particles of opposite charge that lessen the effect of the charge seen at some distance r. As r increases, or if the charge is probed at smaller energies, there are more virtual particles screening the charge such that the effective charge is reduced. Conversely, at smaller r or if probed at higher energies, the screening effect reduces such that the effective charge increases. In QCD the colour charge is screened in a similar way by quark-antiquark (colour-anticolour) pairs, decreasing the overall colour charge. However, since gluons are self interacting, this effect also occurs with gluons where gluon loops contribute to the overall colour charge instead, leading to an anti-screening effect. The behaviour of the coupling constant at different quark distances will depend on the interplay of the screening and anti-screening mechanisms. The experimental measurements of the α_s at different transferred momentum Q, shown in Fig. 1.1, reproduce the expected scaling of the coupling constant with the energy.

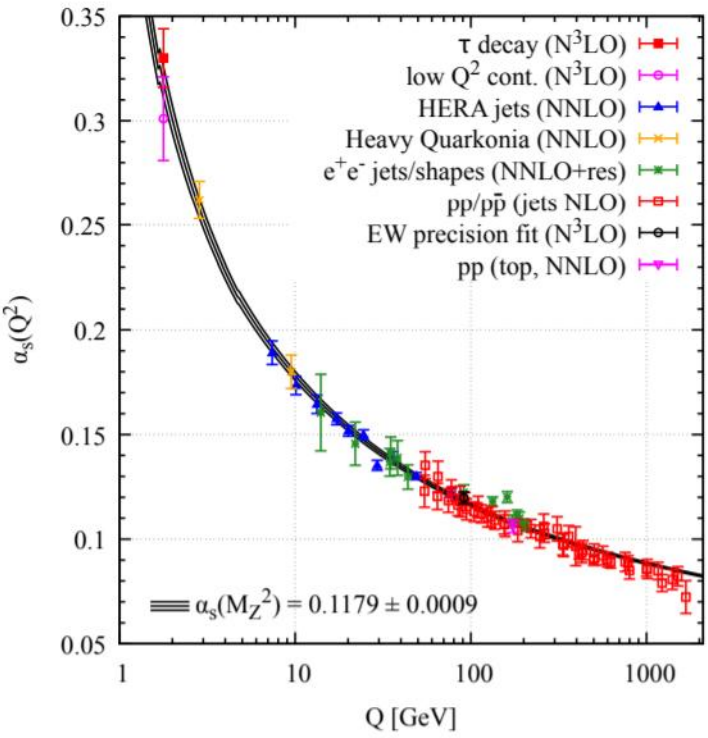

Figure 1.1: Summary of the strong coupling constant $\alpha_S(Q^2)$ measurements as a function of the four-momentum transfered Q^2. Measurements (shown as points) are compared with a parameterisation of α_s at the scale of the Z-boson mass. The brackets in the legend indicate the order to which perturbation theory is used (NLO:next-to-leading order; NNLO: next-to-next-to-leading order; NNLO+res.: NNLO matched to a resummed calculation; N3LO: next-to-NNLO) [4].

For low values of transferred momentum (and, therefore, for relatively large distances between quarks), the coupling constant grows indefinitely, making impossible the quarks separation. Beyond a critical distance, the potential becomes high enough such that a new quark-antiquark pair pops up from the vacuum. In this way, the original quark-antiquark pair splits in two pairs. So, quarks are always confined inside hadrons and we can never isolate them in QCD. This phenomenon is called *confinement*. The interaction between quarks and gluons is too strong and it cannot be treated perturbatively. Complex numerical methods are used for its resolution, such as the lattice QCD (lQCD) in which the QCD is formulated on a discrete Euclidean space time grid (lattice) [5].

Particles possessing a non-zero colour charge cannot be isolated and therefore observed individually. Quarks are bound together by the strong interaction in hadrons, so that the net colour charge is neutral. Conversely, for sufficiently high momentum values (and, therefore, small distances), the value of the coupling constant decreases and the quarks behave as free non-interacting particles. This phenomenon is known as *asymptotic freedom*. In this case, the coupling strength can be determined using perturbative expansion of Quantum Chromodynamics (pQCD). If we truncate this series to the first term only, the expression for α_s relative to some fixed scale M looks as follow:

$$\alpha_s(Q^2) = \frac{\alpha_s(M^2)}{1 + \left(\frac{3N - \frac{2}{3}n_f}{4\pi}\right) \ln\left(\frac{Q^2}{M^2}\right)} \tag{1.13}$$

where $N = 3$ is the number of colours and n_f is the number of quark flavours. As can be easily seen, $\alpha_s(Q^2) \to 0$ when $Q^2 \to \infty$, describing in this way the asymptotic freedom. This running of the coupling strength has been verified at a variety of energy scales using results from relativistic colliders, as shown in Fig. 1.1, where the α_s obtained from perturbative QCD calculation at the scale of the Z-boson mass, represented by black lines, is in agreement, within the uncertainties, with the experimental points. After all the considerations done, the QCD interaction potential between a quark-antiquark pair can be written as:

$$V_{QCD}(r) = -\frac{4}{3}\frac{\alpha_s(r)}{r} + kr \tag{1.14}$$

where k is a constant called *string tension* and r is the distance between the interacting quarks. The interaction potential (1.14) is able to match the aforementioned characteristics. Indeed $V_{QCD}(r) \to \infty$ when $r \to \infty$, reproducing the quark confinement. Conversely, for low distances between the interacting quarks ($r \to 0$), the first term dominates and, since the coupling constant $\alpha_s(r) \to 0$ when $r \to 0$, the asymptotic freedom of the quark is reached.

1.4 QGP Phase transition

The discovery of the asymptotic freedom has a very strong consequence: at very high temperature and/or density conditions, quarks and gluons interact so weakly that they form a colour-deconfined state of quarks and gluons called *Quark Gluon Plasma (QGP)*. Since the system passes from hadronic to partonic matter, by profoundly modifying its physical properties, there is a real phase transition. It is hypothesized that this was the state of matter of the universe in the initial moments of its existence ($10^{-10}s - 10^{-6}s$), right after the Big Bang. Also in neutron stars, given the very high densities reached in the core, the presence of a plasma of this type is hypothesized. The QCD phase diagram is shown in Fig. 1.2 as a function of the temperature of the system (y axis) and of the baryochemical potential (x axis), defined as the energy needed to increase by one unity the total baryon number. It is proportional to the net baryon density (the total baryon number per unit volume).

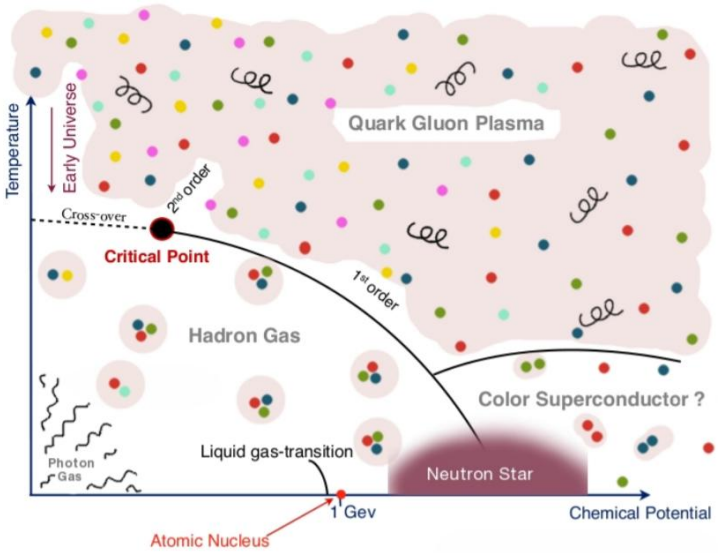

Figure 1.2: Phase diagram (temperature, baryonchemical potential) of the nuclear matter. [6].

In the bottom-left corner of the diagram, the confined ordinary matter can be found. In particular, the point at ($\mu \sim 1$ GeV, $T \sim 0$ MeV) corresponds to nuclear matter under standard conditions in the present Universe. Increasing the baryochemical potential and keeping low the temperature, a transition to the QGP is expected to

happen. These very high density conditions have been hypothesised to be present in the core of the neutron stars. The pressure is such that the separation between hadrons is smaller than their radius (∼1 fm), causing them to effectively overlap. At this point, each quark can no longer be meaningfully attributed to any given hadron. A further increase of μ_B should lead to the formation of Cooper pairs and therefore of a colour-superconductive state. This very high baryon density cannot be reproduced in laboratory, so, the region of the diagram explored by the colliders is the one at low baryochemical potential ($\mu_B \simeq 0$) and at very high temperatures. The phase transition is indeed expected to occur also in the limit of $\mu_B = 0$, at a temperature of ∼170 MeV, called critical temperature, corresponding to an energy density of the order of 1 GeV/fm^3. It is postulated that these conditions were those of the early Universe, a few microseconds after the Big Bang.

The presence of a phase transition above the critical temperature has been confirmed by lQCD calculation as well. In Fig.1.3, the trend of the strong coupling constant is shown as a function of the hadron radius for systems at different temperatures.

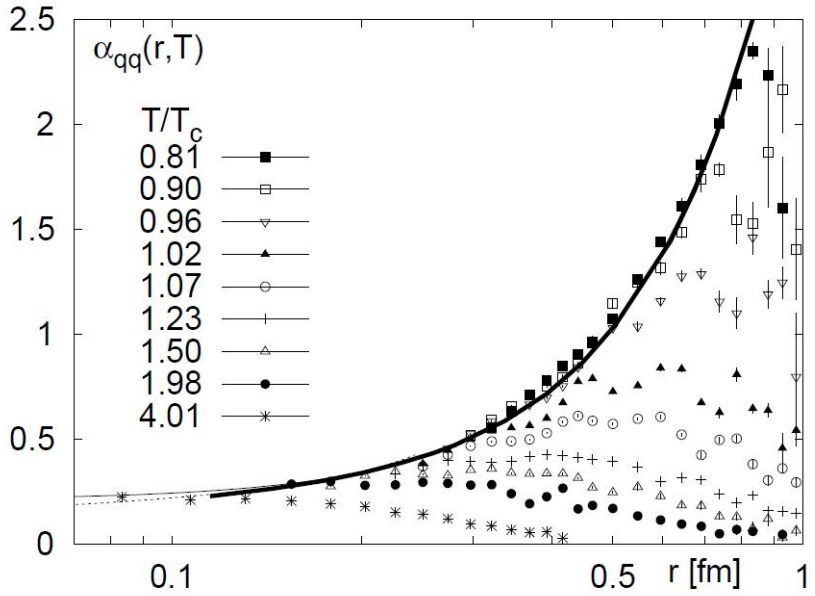

Figure 1.3: Trend of the strong coupling constant α_s as a function of the hadron radius for systems at different temperatures [7].

At low temperatures, the potential diverges for distances of the order of hadronic dimensions (confinement) but, increasing the temperature, the coupling tends to vanish allowing the existence of a deconfined state of quarks and gluons.
In the phase transition from hadronic to partonic matter, the number of degrees of freedom of the QGP system, n_{QGP}, increases and it can be calculated as:

$$n_{\mathrm{QGP}} = n_g + \frac{7}{8}(n_q + n_{\bar{q}}) \tag{1.15}$$

where n_g and n_q ($n_{\bar{q}}$) are the number of degree of freedom for gluons and quarks (antiquarks), respectively, and the factor 7/8 derives from the Fermi-Dirac statistics for fermions. Since there are 8 "type" of gluons and they can have 2 spin states, the gluonic degrees of freedom are 16. Concerning the quark number of degree of freedom, there are 3 different colours, 2 charge and 2 spin states to which the number of flavours involved has to be added. From the statistical mechanics of ideal quantum gases in thermal equilibrium [8], the energy density of fermions and boson in the massless limit[2] is proportional to the number of degree of freedom:

$$\epsilon = n_{\mathrm{QGP}} \frac{\pi^2}{30} T^4 \tag{1.16}$$

where T is the temperature of the system. So, what is expected is an increase in the energy density at the phase transition. The dependence of the energy density divided by T^4 (equivalent to the number of degree of freedom) as a function of the temperature, obtained from lQCD calculation, is reported for different hypothesis on the free quark masses in Fig. 1.4. A transition is observed at a temperature of about 173 MeV and energy density of 0.7 GeV/fm^3.

[2]The up and down quarks mass inside the hadrons is ∼300 MeV. Valence quarks interact with gluons and sea quarks enhancing their effective mass. After the transition to the deconfined medium, the chiral symmetry [9] is expected to be restored and the up and down quarks mass can be approximated to zero.

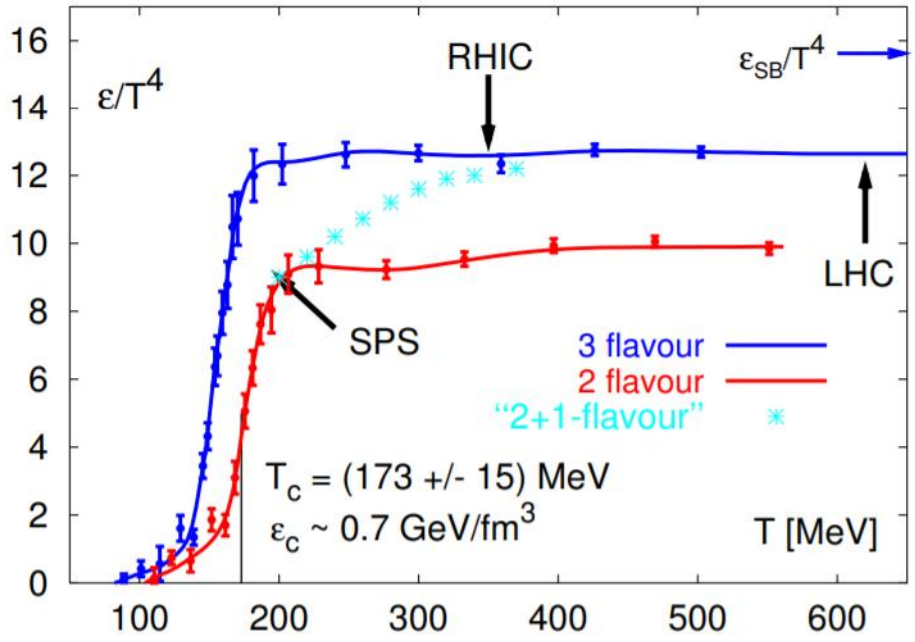

Figure 1.4: Dependence of the energy density as a function of the temperature of the hadronic matter at null baryonic potential given by lattice QCD calculations at finite temperature. The calculations are performed for two massless quarks (up and down), three massless quarks (up, down and strange) and two massless quark and one (strange) with its real mass [8].

To conclude this section, the lattice QCD can provide also some information about the order of the phase transition from the hadronic to the partonic matter. At the condition reproduced in the colliders, i.e. large T and small μ, on the basis of lattice calculations, the transition is expected to be a rapid crossover, whereas it is naturally expected that the phase transition along the μ-axis (with $T = 0$) is an actual first-order phase transition [10]. This implies that, at some point in the phase diagram 1.2, the first-order phase transition is converted into a rapid crossover. The point at which this occurs is commonly known as the QCD critical point and its determination is still under study.

1.5 Experimental study of the QGP

In the previous sections, the phase transition from the hadronic to the partonic matter, under extreme condition of temperature and density, has been introduced. Experimentally, the condition for the QGP production are reproduced via ultrarelativistic heavy-ion collisions. The first attempts were performed with fixed target experiment, by the Alternating Gradient Synchrotron (AGS) at the Brookhaven National Laboratory (BNL) with Au ions at $\sqrt{s_{\mathrm{NN}}} = 5$ GeV and the Super Proton Synchrotron (SPS) at European Organization for Nuclear Research (CERN) with Pb nuclei at $\sqrt{s_{\mathrm{NN}}} = 17$ GeV. The convincing experimental results presented by the WA97 and NA57 experiments at SPS brought to the official announcement by the CERN of the discovery of a new state of matter at the beginning of 2000. During the following years, important quantities that allowed infering some of the QGP medium properties, such as the collective anisotropy and the jet quenching in the medium, were measured by the STAR and PHENIX experiments at the Relativistic Heavy Ion Collider (RHIC) [11, 12] at BLN. Cu-Cu and Au-Au collisions at $\sqrt{s_{\mathrm{NN}}} = 130$ GeV and $\sqrt{s_{\mathrm{NN}}} = 200$ GeV, respectively, were employed since 2000 to achieve these crucial results. A totally new energy regime was reached with the beginning of the LHC operation in 2010. Pb-Pb collisions were performed at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV and $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. Both RHIC and LHC allow the exploration of the QCD phase diagram in the region of μ_B close to zero and very high temperature.

1.5.1 QGP system evolution

The collision between two heavy nuclei is an event with a complex space-time evolution that can be described with the 2D light-cone, reported in Fig 1.5. Here, the collision between the beams occurs at ($z = 0$, $t = 0$), where z is the direction of the colliding beams and t is the time.

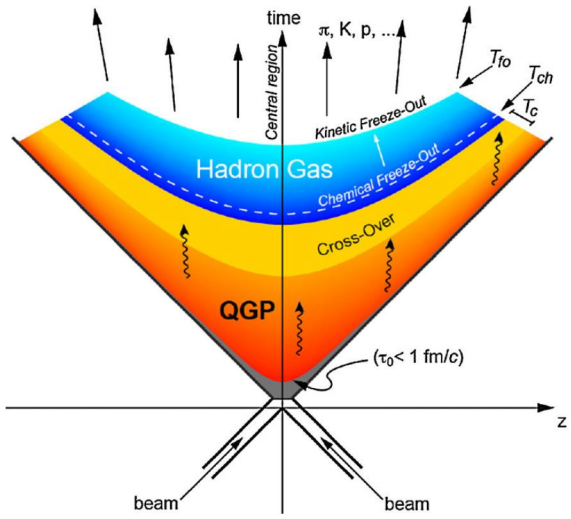

Figure 1.5: The space-time evolution of the QGP in a heavy-ion collision, as a 2D representation of the light-cone.

The system goes through several stages:

- Thermal equilibrium and QGP formation: in the very first moments of the collision, a very high partonic density is produced. Indeed, the frequent collisions between nucleons do not allow them to escape freely from the interaction zone, as happens instead in proton-proton collisions, but force them to continuously collide with each other. So, they can form a region with a high density of particles that reaches the thermal equilibrium very quickly ($\tau_0 \sim 1$ fm/c) thanks to elastic scattering, leading to the formation of QGP.

- Expansion and cooling: after reaching thermal equilibrium, the pressure exerted by the plasma on the surrounding vacuum causes the partonic system to rapidly expand and cool. The QGP behaves like a nearly perfect fluid [13]. So, the expansion occurs collectively due to the low viscosity of the medium, and it is generally described using relativistic hydrodynamics.

- Hadronisation: when the temperature decreases below the critical value T_c, the QGP cannot longer exist and its constituents hadronise. The medium becomes an interacting gas of hadrons.

- Chemical freeze-out: the hadron gas continues its expansion and cooling since, below a temperature threshold T_{ch}, the inelastic collisions cease. The hadron species abundances are fixed at this stage.

- Kinematical freeze-out: then, the elastic interactions cease (T_{fo}) and the momenta of the hadrons become fixed as well. The system continues its expansion reaching the detectors without further changes.

The overall lifetime of the system is strongly dependent on the energy density produced in the collision, i.e. by the collision energy. For instance, the QGP lifetime at RHIC in Au–Au collisions at $\sqrt{s_{\mathrm{NN}}} = 200$ GeV is $\tau_{QGP} = 6.2$ fm/c while at LHC in Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV is $\tau_{\mathrm{QGP}} = 14$ fm/c [14] at the critical temperature $T_c \approx 180$ MeV.

1.5.2 Proton-proton and proton-ion collision systems

In addition to the heavy-ion collisions, performed to reproduce and study the QGP in laboratory, also proton-proton and proton-ion collisions are carried out at colliders. pp collisions can be used not only to establish a baseline for measurements in different collision systems but also to understand important aspects of QCD and test the pQCD calculations. An important tool that allows the comparison of the results obtained in pp and A-A collisions is the *nuclear modification factor*. From the Glauber model [15], in which the collision between two nuclei is described as incoherent superposition of binary nucleon–nucleon interactions, is expected that the yield of particles produced in hard processes in heavy-ion collisions scales with the average number of binary nucleon–nucleon collisions $\langle N_{\mathrm{coll}} \rangle$. The Glauber model is valid until there are no medium effects, and the interactions are pp-collisions like. To test this hypothesis, the nuclear modification factor R_{AA} is defined as:

$$R_{AA}(p_{\mathrm{T}}, y) = \frac{1}{\langle N_{\mathrm{coll}} \rangle} \frac{\mathrm{d}^2 N_{AA}/\mathrm{d}p_{\mathrm{T}}\mathrm{d}y}{\mathrm{d}^2 N_{pp}/\mathrm{d}p_{\mathrm{T}}\mathrm{d}y} \tag{1.17}$$

where d^2N_{AA} and d^2N_{pp} are the p_T- and y-differential yields in A-A and pp collisions, respectively. The nuclear modification factor is the ratio between the signals obtained in A-A collisions normalized by the average number of collisions between nucleons and those obtained in pp collisions. It describes the modification of a given observable in heavy-ion collisions, relative to the pp one. So it is expected to be 1 in absence of medium effect and discrepancies from unity can hint to QGP formation. The nuclear modification factor can be defined analogously for proton-ion collisions as:

$$R_{pA}(p_\mathrm{T}, y) = \frac{1}{A} \frac{\mathrm{d}^2N_{pA}/\mathrm{d}p_\mathrm{T}\mathrm{d}y}{\mathrm{d}^2N_{pp}/\mathrm{d}p_\mathrm{T}\mathrm{d}y} \tag{1.18}$$

where A is the mass number of the colliding ion. The R_{pA} is useful to spot mechanism, called *cold nuclear matter effects* (CNM), that modify the output of the collision but that are not related to the QGP formation. The CNM effects could arise from the colliding nucleus internal structure that is different with respect to free protons (initial-state effects) or to the interaction of particles with remnants of the colliding nuclei, such as final-state energy loss. It is important to factorise these mechanisms from those induced by the creation of the QGP, in order to properly interpret the measurements in A–A collisions.

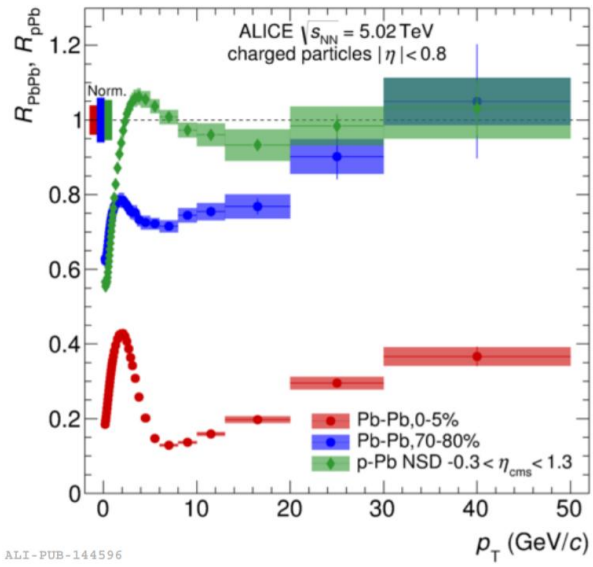

Figure 1.6: Nuclear modification factors of primary charged particles measured by ALICE in central (0-5%) and peripheral (70-80%) Pb–Pb collisions and in p–Pb collisions at $\sqrt{s_\mathrm{NN}} = 5$ TeV [16].

Examples of nuclear modification factors R_{AA} and R_{pA} of the primary charged particles measured by the ALICE experiment in Pb-Pb and p-Pb collisions at $\sqrt{s_{\rm NN}} = 5$ TeV are shown in Fig. 1.6. For low-$p_{\rm T}$ particles the nuclear modification factor is suppressed as a consequence of the *shadowing*, due to the destructive interference arising from multiple-scattering of the incoming particle (more detail in sec. 2.2), while at high-$p_{\rm T}$ ($p_{\rm T} > 5$ GeV/c) the nuclear modification factor is compatible with unity for p–Pb collisions and it is significantly suppressed for Pb–Pb collisions, as a consequence of the energy loss in the medium (see sec. 1.5.3).

1.5.3 Experimental signatures of QGP formation

Due to the very short lifetime of the QGP, it cannot be directly revealed. The experimental challenge is to determine whether the QGP has actually formed in the interaction zone. This can be done through indirect signals, obtained from the observation of the ordinary particles that emerge from the interaction region after the medium cooling:

- **Electromagnetic radiation:** information about the early stages of QGP can be obtained by studying photons and lepton pairs e^+e^- and $\mu^+\mu^-$. They are formed either from the interaction of quarks and gluons or by quark-antiquark annihilation. Indeed, they carry no colour charge and, once formed, they can escape the medium without further interaction. These particles are of crucial importance for the determination of some critical parameters, such as the temperature. However, it is not easy to discriminate between photons and leptons produced in the early stage from those deriving by hadron decays and then formed in the later stages of the collision.

- **Strangeness enhancement:** since no valence strange quarks are present in the colliding nuclei, they must be produced in the collision or in the QGP. In particular, a strangeness production enhancement is expected in the medium. The mass of a strange valence quark is ∼500 MeV/c^2 and its production in the collision is disfavoured. But, if the system reaches the critical temperature,

the phase transition occurs and the mass of the strange quark decreases to its bare mass value ∼100 MeV/c^2 due to the restoration of the chiral symmetry. Hence, strange quark-antiquark pairs are expected to be abundantly produced in the QGP, leading to an increase of the relative abundance of particles with strange-quark content. The enhanced production of strange hadrons was also experimentally observed at RHIC and LHC energies in heavy-ion collisions (Fig. 1.7).

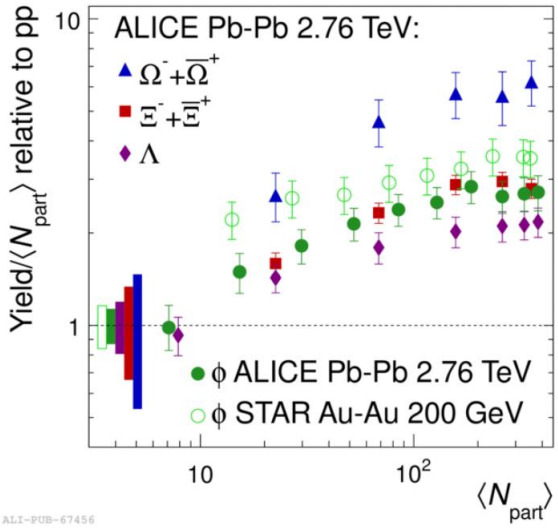

Figure 1.7: Strange hadrons enhancement (Λ, Ξ, Ω) in Pb-Pb collisions at $\sqrt{s_{\rm NN}}$ = 2.76 TeV at the LHC w.r.t. the pp reference yields. The ϕ mesons enhancement observed in Au-Au collissions at $\sqrt{s_{\rm NN}}$ = 200 GeV at RHIC is reported as well [17].

It is important to note that hierarchy of the enhancement with the strangeness content is observed (the Λ baryon has one strange quark, the ϕ meson and the Ξ baryon have two strange quarks while the Ω baryon has three strange quarks). Recently, the strangeness enhancement has been observed in high-multiplicity pp collisions as well by the ALICE Collaboration [18]. The particle composition evolves smoothly across collision systems, depending on charged particle multiplicity. Hence, the strangeness enhancement could have a common origin in all systems.

- **Quarkonia:** quarkonium is a meson whose constituents are a heavy quark

(charm or bottom) and its own antiquark. Quarkonia can exist bound in the medium up to temperatures similar to their binding energies at which point they can melt, suppressing the final quarkonia yield with respect to pp collisions. The ground states of charmonium (J/Ψ) and bottonomium (ϒ) binding energy are expected to make them able to survive in the QGP. However, the interaction potential between the quark pair could be strongly decreased because of the colour screening effects due to deconfined colour charge in the medium. Taking into account the colour screening and in the condition of small distances between quarks, the interaction potential (1.14), assumes the form:

$$V_{QCD}(r) = -\frac{4}{3}\frac{\alpha_s(r)}{r}e^{-\frac{r}{\lambda_D}} \tag{1.19}$$

where λ_D is the Debye screening length, which defines the range of the strong force between quarks and decreases as the charge density of the medium increases. Moreover, the higher excited quarkonium states are less tightly bound and have a different melting temperatures. So, a suppression of the quarkonia bound states, dependent by the medium temperature, is expected. The J/Ψ production yields measured in different collision systems at SPS are shown in Fig. 1.8.

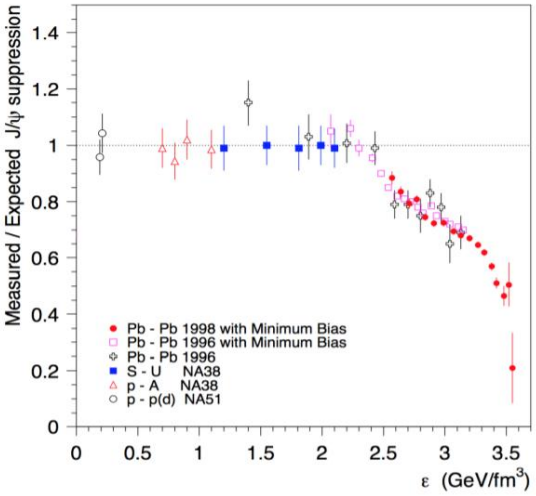

Figure 1.8: Measured J/Ψ production yields, normalised to the expected yield, as a function of the energy density reached in the several collision systems [19].

The experimental data show a clear evidence of the J/Ψ suppression when the energy density is high enough. However, for very high energy collisions, such as those delivered at the LHC, a regeneration of the J/Ψ is observed. Indeed, a greater number of charm quarks is produced at that energy scale. As a consequence, there is a secondary production of J/Ψ due to the statistical recombination of the c quarks during the hadronisation phase. The J/Ψ restoration is evident by comparing the results obtained at the RHIC with those at LHC, superimposed in Fig. 1.9, where the J/Ψ suppression is lower because the energies reached are higher.

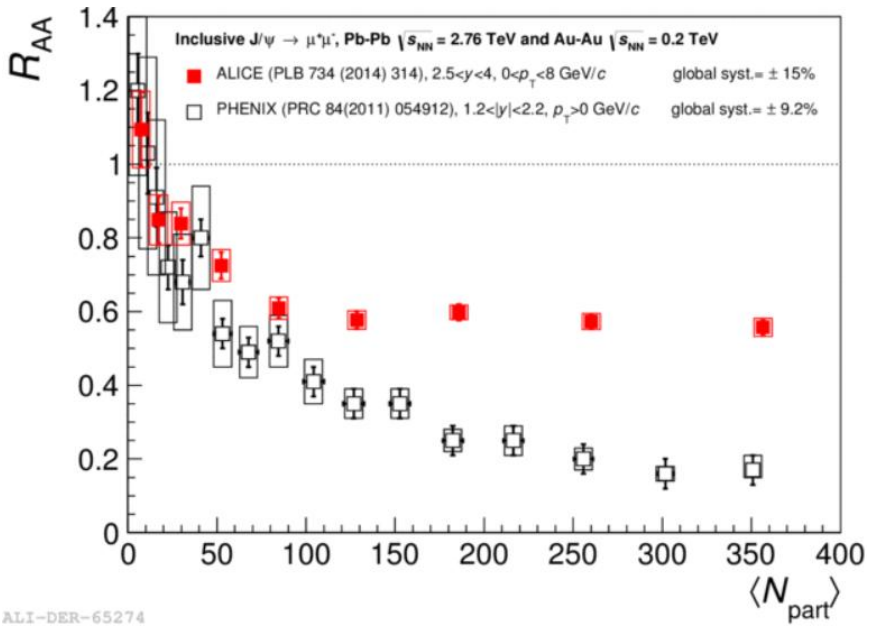

Figure 1.9: Centrality dependence (N_{part} is the number of partecipants) of the nuclear modification factor, R_{AA}, of inclusive J/Ψ production in Pb-Pb collisions at $\sqrt{s_{\text{NN}}} = 2.76$ TeV, measured at forward-rapidity compared to PHENIX results in Au-Au collisions at $\sqrt{s_{\text{NN}}} = 200$ GeV [20, 21].

- **Jet quenching:** high-p_{T} partons (quarks or gluons) are produced in hard-scattering processes in the early stage of the collision. In this hard scattering a single pair of partons from the incoming nuclei strike each other directly with such force that they scatter with high momentum away from the initial beam direction. These partons then fragment, giving rise to localized sprays of

energetic particles called jets. When a high-p_{T} parton propagates in the QGP, it strongly interacts with the extremely dense medium and loses its energy via medium-induced gluon radiation and elastic scattering. The final result is a softening of the p_{T} distributions of the hadrons arising from the jets. This energy loss is called *jet quenching*. The expected situation is schematized in Fig.1.10.

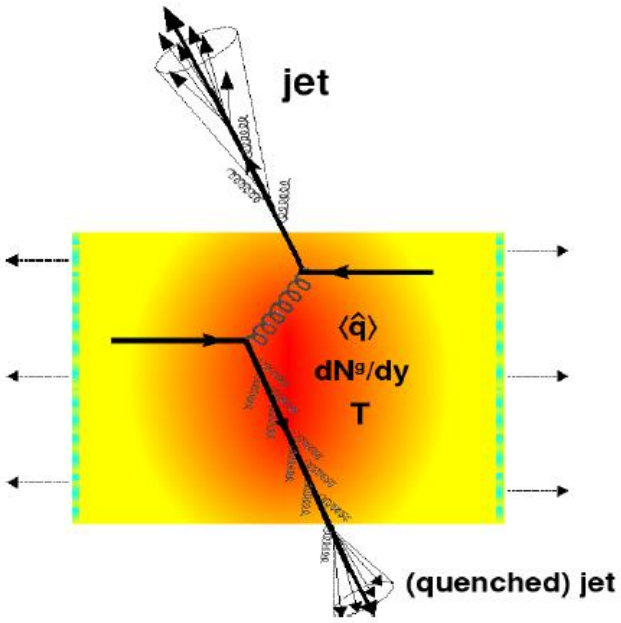

Figure 1.10: Jet quenching in a nucleus-nucleus collision. Two quarks suffer a hard scattering: one goes out directly to the vacuum, radiates a few gluons and hadronises, the other goes through the dense plasma formed in the collision, suffers energy loss due to medium-induced gluon radiation and elastic scattering and finally fragments outside into a (quenched) jet [22].

The quantity of energy lost by the partons travelling in the medium when they come out, can provide useful information about the medium density. In Fig. 1.11 the jet quenching observed at RHIC in Au-Au collisions at $\sqrt{s_{\mathrm{NN}}} = 200$ GeV is shown. The experiment measures a high momentum jet particle (the near-side trigger) and then looks at the correlation between the trigger and the softer particles in the opposite direction (away-side region), where it is assumed that the recoil jet passed through the medium, was suppressed. Strong modification

of the recoil jet indicates substantial partonic interaction and subsequent energy loss within the medium.

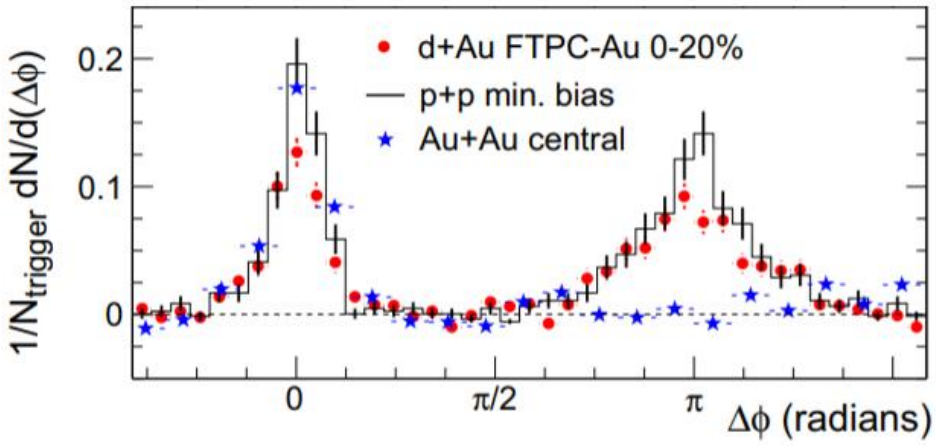

Figure 1.11: Two-particle azimuthal distributions for minimum bias pp, central d+Au and Au+Au collisions [23].

- **Collective flow:** collectivity relates to the physics governing the expansion of the system. Once formed, the QGP starts to expand because of its own internal pressure, leading to a global hydronamic flow. This collective isotropic motion, called *radial flow*, is superimposed to the thermal motion, giving an additional overall boost in momentum to the produced hadrons. The radial flow provides information about the QCD interactions in the expanding fireball. In particular, a greater flow generally indicates that the matter comes closer to thermal equilibrium. Moreover, if the initial condition were not spherically symmetric, as for the peripheral heavy-ion collisions, the difference in pressure in the spatial directions will lead to additional anisotropic flow components. The anisotropic flow will be reflected in an anisotropy in the momentum distribution of the produced particles and therefore it can be measured through the azimuthal angle distributions of the particles in the final state. The overall collective flow can be studied by expanding the particle spectrum in terms of Fouries series of the azimuthal angle ϕ (angle between the particle and the

reaction plane):

$$\frac{\mathrm{d}N}{\mathrm{d}\phi} = \frac{N}{2\pi}(1 + 2\sum_{n=1}^{\infty} \nu_n \cos(n\phi)) \tag{1.20}$$

where ν_n are the Fourier coefficients of the n^{th} harmonic. The first coefficient, ν_1, characterises the radial flow. The second order armonic, ν_2, is the so-called *elliptic flow*. It originates from the shape of the interaction region and strongly depends on the centrality of the collisions. The comparison of the measured ν_n harmonics with viscous hydrodynamic models provides information about medium properties, such as equation of state and shear viscosity.

- **Hadron production:** a lot of information can be retrieved analysing hadron radiation dependence on transverse momentum, multiplicity, rapidity, centrality, beam energy, particle species. The hadron production studies are classified in light (hadrons constituted only by light quarks, i.e. up, down and strange) and heavy flavour, i.e. particles made of at least one heavy quark (charm or beauty). Light-flavour hadrons are mainly produced at later stages of the evolution and are able to describe the thermodynamic properties and the hydrodynamics of the QGP. In particular, the total light-hadron yields are used to study the chemical freeze-out characteristics, such as the temperature T_{ch}, the baryochemical potential, the volume of the system. Those physical quantities can be in fact evaluated by comparing the measured abundances of different hadron species with the predictions of the statistical hadronisation model (SHM) [24], which assumes that the system is in the thermal and chemical equilibrium at the chemical freeze-out. This is done by using a limited number of parameters (T, μ, V) that are fixed fitting the measured yield with the SHM. The results obtained by the ALICE experiment in central Pb-Pb collision at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV are shown in Fig. 1.5.3.

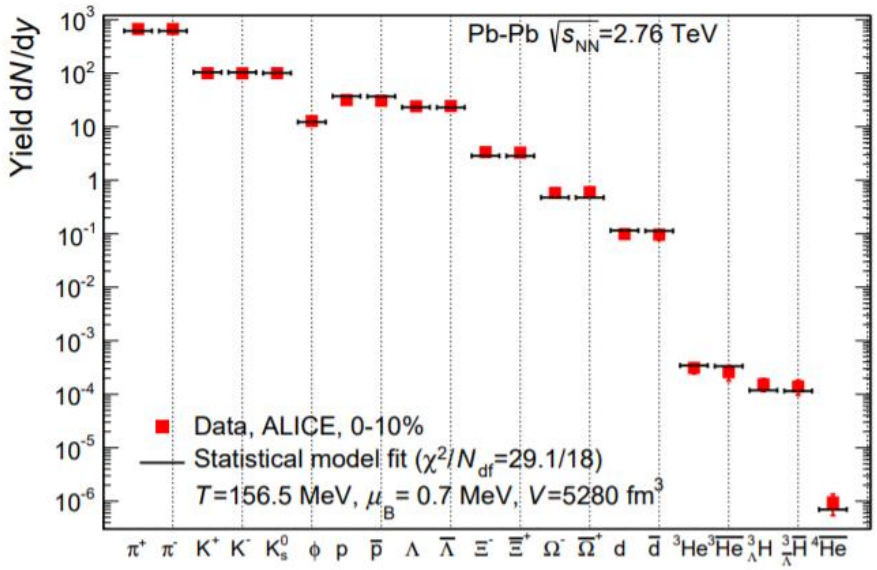

Figure 1.12: Hadron multiplicities in central (0-10%) Pb–Pb collisions at the LHC with the ALICE data and best fit obtained with the Standard Hadronisation Model. The values of fitted parameters (T, V, μ) are reported as well [25].

Conversely, due to their large mass, heavy quarks are produced in the initial stages of the collisions via hard partonic scattering processes and so they give better insight into the early stages of the collision system. Then, they interact with the medium constituents experiencing its whole evolution. The relative heavy-hadron yields can give insight into hadronisation. In particular, the relative abundances of baryons and mesons ("baryon-to-meson ratios") are sensitive to the hadronisation mechanisms in the QGP (see 2.4 for more details).

Chapter 2

Open heavy-flavour production

Heavy-flavour hadrons, i.e. those containing at least one charm or beauty quark[1], provide an important testing ground for perturbative QCD predictions and they are a unique probe into the properties of the medium formed in heavy-ion collisions and into the hadronisation mechanisms. In fact, heavy quarks are produced in hard-scattering processes occurring in the early stage of the collision and subsequently experience the full system evolution, interacting with the medium constituents via elastic and inelastic processes. Charm, being the lightest heavy quark, is the most commonly produced in ultrarelativistic hadronic collisions. Due to the strong force flavour conservation law, charm quarks produced in hard-scatterings are always produced in $c\bar{c}$ pairs, which may either hadronise into flavour-neutral charmonia or into open-charm baryons and mesons, i.e. particles with a non-zero charm quantum number (most commonly D mesons and Λ_c baryons). In this chapter, theoretical details of heavy-flavour production and hadronisation mechanism as well as the state of the art of the open-charm studies are discussed.

[1]The top quark is also considered to belong to the heavy-flavour quarks, but it decays before it can hadronise due to the large difference between the top and beauty masses; it therefore is not mentioned in the following.

2.1 Factorization theorem

As already pointed out in sec. 1.5.2, pp collisions are crucial to provide a baseline and to test pQCD calculations. Predictions for the pp p_{T}-differential production cross section of open heavy-flavour hadrons can be obtained with the QCD factorization theorem [26] as a convolution of three terms:

$$\mathrm{d}\sigma_{pp\to h} = \sum_{i,j=q,\bar{q},g} f_{i/A}(x_i, Q^2) \otimes f_{j/B}(x_j, Q^2) \otimes \mathrm{d}\sigma_{ij\to q\bar{q}}(x_i, x_j, Q^2) \otimes D_{q\to h}(z, Q^2) \quad (2.1)$$

where:

- $f_{i/A}(x_i, Q^2)$ and $f_{j/B}(x_j, Q^2)$ are the parton distribution functions (PDFs). In proton-proton collisions, the proton must be considered as a composite structure of quarks and gluons. An inelastic proton-proton collision is therefore a superposition of one or more inelastic parton-parton interactions. The PDFs represent the probability to find the parton species i (j) in the colliding proton labelled as A (B) as a function of the fraction $x_i = p_i/p_{\mathrm{proton}}$ ($x_j = p_j/p_{\mathrm{proton}}$) of the proton's momentum carried by the parton i (j) at the resolution scale Q^2. Because of the non-perturbative nature of partons, which cannot be observed as free particles, PDFs cannot be calculated using pQCD. They are usually obtained by fitting observables to deep-inelastic scattering experimental data [27, 28] or derived from lQCD calculations [29].

- $\mathrm{d}\sigma_{ij\to q\bar{q}}(x_i, x_j, Q^2)$ is the cross section of the partonic scatterings producing the heavy quarks. Because of their large masses, $\mathrm{m}_c \simeq 1.3$ GeV/c^2 and $\mathrm{m}_b \simeq 4.2$ GeV/c^2, heavy quarks are always produced in hard-scattering processes, with momentum transfer of the order of $Q^2 \geq 4\mathrm{m}_{b,c}^2$. The strong coupling constant is significantly smaller than unity in this Q^2 region (see sec. 1.3), so the heavy quark-antiquark pair cross-section production, $\sigma_{ij} \to q\bar{q}$, can be calculated with perturbative techniques. At the leading order (LO) the two processes that contributed to the heavy-quark production are the quark-antiquark annihilation, $q + \bar{q} \to Q + \bar{Q}$, and the gluon fusion, $g + g \to Q + \bar{Q}$. If

also the Next-to-Leading Order (NLO) is considered in the computation, both the virtual one-loop (top diagrams in Fig. 2.1) and the real ($q+\bar{q} \to Q+\bar{Q}+g$, $g+g \to Q+\bar{Q}+g$, $q+g \to Q+\bar{Q}+q$, bottom diagrams in of Fig. 2.1) corrections, must be added.

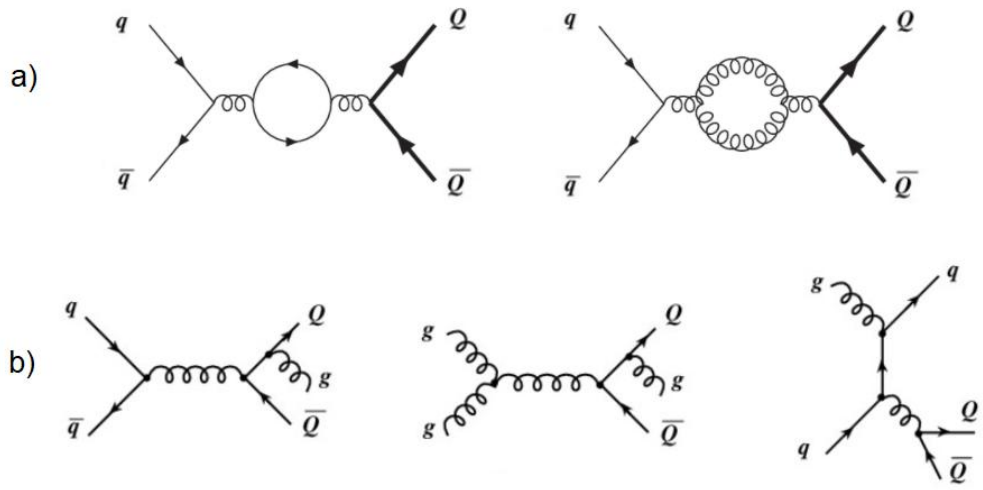

Figure 2.1: Next-to-leading order mechanisms for the heavy quark-antiquark pair production.

- $D_{q\to h}(z, Q^2)$ is the quark fragmentation function (FF). The FFs represent the probability that the outgoing heavy-quark q fragments in the hadron h carrying the fraction z of the initial parton momentum. Because they describe the heavy quark transition into a given hadron, a non-perturbative parametrisation tuned to experimental results must be used for its determination. They are assumed to be universal across different collision systems. So, the FFs are typically measured in e^+e^- collisions [30] because they provide the cleanest possible environment.

Both the PDFs and the FFs are determined experimentally at a given energy scale and then projected at the needed scale with the DGLAP (Dokshitzer – Gribov – Lipatov – Altarelli – Parisi) equations [31]. In this way, parton distribution functions and fragmentation functions determined from experimental data at lower energies can be scaled for higher energy regimes, such as those at the LHC.

2.2 Modification in large systems

As explained in sec. 1.5.2, already in p-A collisions, the presence of nuclei implies a variation in the production of open heavy-flavours also in absence of a colour-deconfined medium. The initial-state effects, introduced in sec. 1.5.2, derive from the modification of nuclear parton distribution functions with respect to simple superpositions of proton PDFs. The modified PDFs, called nuclear PDFs (nPDFs), $f_i^A(x_i, Q^2)$, assume the form:

$$f_i^A(x_i, Q^2) = R_i^A(x_i, Q^2) f_i(x_i, Q^2) \tag{2.2}$$

where $R_i^A(x_i, Q^2)$ is the nuclear modification to the free proton PDF $f_i(x_i, Q^2)$ for a parton of species i: a value less than 1 at a given fractional momentum x_i and momentum transfer Q^2 indicates a depletion of partons with respect to a free nucleon, while a value greater than 1 indicates an enhancement. The overall $R_i^A(x_i, Q^2)$ shape versus the fractional momentum x is shown in Fig.2.2.

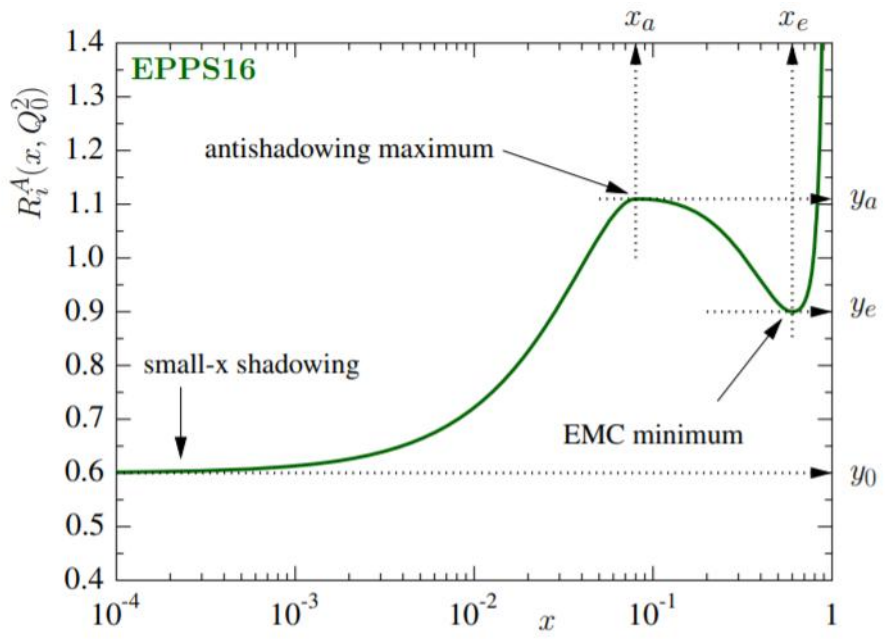

Figure 2.2: $R_i^A(x_i, Q^2)$ shape. Typical nuclear effect are indicated in the figure [32].

The magnitude of the $R_i^A(x_i, Q^2)$ depends on the phenomena experienced by partons

at different fractional momentum x. Moving from low to high x, the predominant effects are:

- **Shadowing**: for $x < 0.05$ a suppression of the PDFs is observed. This effect is due to the high phase-space density of small-x partons and it is induced by the destructive interference arising from multiple-scattering of the incoming particle with the overlapping small-x partons from different protons of the nuclei.

- **Anti-shadowing**: in the region $0.05 < x < 0.3$, constructive interferences due to multiple scatterings dominate and the nPDFs are enhanced.

- **EMC (European Muon Collaboration) effect**: in the high-x region ($0.3 < x < 0.8$) the $R_i^A(x_i, Q^2)$ is lower than unity, reaching a minimum around $x \simeq 0.7$. The dominant effect is the EMC effect. It is still not fully understood but it seems to be related to the modifications of the kinematic distributions of quarks and gluons in bound nucleons. In particular, partons in bound nucleons move through a larger confinement volume and, as the uncertainty principle implies, they carry less momentum than in free nucleons.

- **Fermi motion**: Finally, for $x \sim 1$, the $R_i^A(x_i, Q^2)$ increases above unity because of the nucleons Fermi motion.

After the introduction of the nPDF, the factorization theorem can be extended to proton-ion collisions as well, modifying the (2.1) as follows:

$$\mathrm{d}\sigma_{pA\to h} = A \cdot \sum_{i,j=q,\bar{q},g} f_i(x_i, Q^2) \otimes f_j^A(x_j, Q^2) \otimes \mathrm{d}\sigma_{ij\to q\bar{q}}(x_i, x_j, Q^2) \otimes D_{q\to h}(z, Q^2) \quad (2.3)$$

where A is the mass number of the nucleus.

A crucial CNM mechanism that contributes to the modification of the open heavy-flavour production in p–A collisions compared to that in pp collisions is the *Cronin effect* [33]. The partons in the projectile particles go through multiple, softer, elastic scatterings with the target nucleus constituent, before the hard-scattering. After

each scattering, the parton will acquire additional transverse momentum p_{T}, which grows with the number of elastic collisions. These elastic interactions transfer an initial transverse momentum to the partons, which is responsible of the p_{T}-spectrum shift for the particles produced in hard scatterings, such as heavy-quarks. This extra momentum given to the hard parton in its final state can lead to an enhancement of the R_{pPb} at intermediate p_{T}. As the p_{T} of the produced particle increases, the extra momentum supplied by the Cronin effect becomes a smaller fraction of the total observed p_{T}, so the Cronin enhancement is expected to disappear for $p_{\mathrm{T}} \to \infty$. Concerning A-A collisions, also the medium effect must be considered. In this case the factorization theorem can be written as:

$$\mathrm{d}\sigma_{AA\to h} = A^2 \cdot \sum_{i,j=q,\bar{q},g} f_i^A(x_i, Q^2) \otimes f_j^A(x_j, Q^2) \otimes \mathrm{d}\sigma_{ij\to q\bar{q}}(x_i, x_j, Q^2) \otimes R^A_{\mathrm{QGP}} \otimes D^A_{q\to h}(z, Q^2) \tag{2.4}$$

Here, R^A_{QGP} is the modification of the heavy-quark momentum due to energy loss in the medium and/or collective flow, and $D^A_{q\to h}(z, Q^2)$ is the medium-modified fragmentation function due to the different hadronisation mechanisms involved in the QGP (see sec 2.4).

2.3 pQCD models and MC generators

Several pQCD models were developed to estimate the heavy quark production cross section, $\mathrm{d}\sigma_{ij\to q\bar{q}}(x_j, x_j, Q^2)$. One of the most employed is `FONLL` (Fixed Order + Next-to-Leading Logarithms) [34]. In the `FONLL` framework, the fixed next-to-leading order heavy-quark production processes are matched with a resummation of the next-to-leading logarithms (NLL) in the limit of heavy-quark transverse momentum much larger than its mass (massless limit in the high p_{T} region), according to the formula:

$$\sigma_Q^{FONLL} = FO + (RS - FOM0)G(p_{\mathrm{T}}, m) \tag{2.5}$$

where p_{T} and m are the transverse momentum and the mass of the produced heavy quark Q and σ_Q^{FONLL} is the production cross section of the heavy quark. FO is the cross section estimated with the NLO calculation, performed with massive quarks. RS is the resummed calculation at the massless limit from which $FOM0$, that is the massless limit of the NLO calculation, is subtracted to remove the terms shared by both approaches. The factor of $G(p_{\mathrm{T}}, m)$ is a suppression term, which ensures that the massless limit is only considered at high p_{T}.

Analytical tools, as pQCD calculation, cannot represent the full complexity of a real event in enough detail. An alternative approach for the description of the heavy-flavour hadron production is represented by the Monte Carlo (MC) generators, as `PYTHIA` [35] and `HERWIG` [36], which allow for a more complete description of the final state. `PYTHIA` generator, in particular, includes hard and soft interactions, parton distributions, initial- and final-state parton showers, multiparton interactions, fragmentation and decay. However, it implements the hard processes only at the LO. Monte Carlo generators with NLO accuracy, such as `POWHEG` [37], are also available and can be matched to the parton shower of `PYTHIA` to provide predictions for more differential observables. To conclude, generators such as `HIJING` [38] extend `PYTHIA` to include effects present in heavy-ion collisions.

Hadronisation of heavy quarks (as well as light quarks) is modelled in MC generators using phenomenological, non perturbative models. The most widely used hadronisation model is the Lund string model [39]. In this model the potential between two colour-connected quarks is described by (1.14). When two colour-connected quarks (string) are moving apart, the linear term dominates. The potential energy stored in the string increases with the separation between the quark-antiquark pair until the string breaking into smaller strings via the production of a new $q\bar{q}$ pairs becomes more energetically favourable. In the Lund string model, only partons from the same scattering can be colour connected by a string (leading colour picture (LC)). Quarks are string endpoints while gluons become transverse kinks, connecting two string pieces. Since quark-antiquark pair are progressively generated, with this

mechanism the meson production is strongly favoured, while the baryon production is suppressed. To reproduce the results recently obtained at the hadron colliders [40, 41], an extension on the simple Lund string model based on the colour-reconnection mechanism (CR) [42] has been developed. Examples of string colour reconnection are shown in Fig. 2.3. Partons from two separate parton scatterings may become colour-connected, in such a way that the total string length becomes as short as possible. The strings must be causally connected and different space-time constraint may be applied to determine the allowed strings. Therefore, two partons produced from independent hard scattering can be colour connected and make a large transverse boost. In addition, colour reconnection allows for 'junctions'. Since this topology is closely related to baryons, the result is a baryon enhancement. For this reasons, the CR mechanism was found to mimic several collective-like effects, namely, enhanced production of baryon over meson at intermediate p_{T}, similar to those observed in heavy-ion collisions.

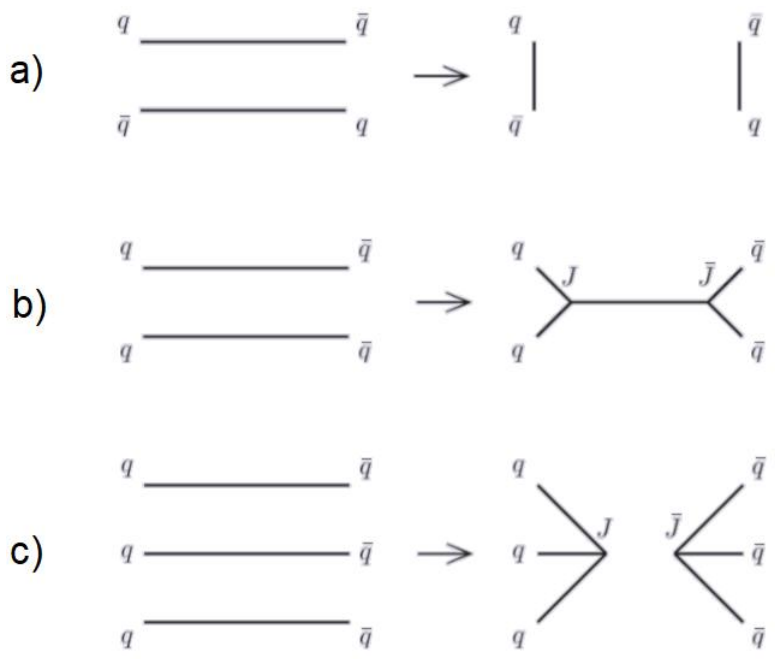

Figure 2.3: Examples of colour reconnection. a) Ordinary string reconnection, b) double junction reconnection, c) triple junction reconnection.

2.4 Hadronisation mechanisms in the medium

The quark hadronisation in small systems is foreseen to occur by parton fragmentation, as described in sec. 2.3. However, in heavy-ion collisions, modification of the hadronisation mechanisms are predicted in the medium. In particular, the interplay between two different mechanism is expected. Heavy quarks can hadronise via fragmentation into a jet of lower-momentum hadrons as in the vacuum; on the other hand, partons that are close in phase space (position and momentum) can recombine with other quarks in the medium, giving rise to a hadron with momentum higher than that of the initial quark. Recombination (or coalescence) [43] is expected to be the dominant hadronisation mechanism at low p_{T} (few GeV/c) while, for high-p_{T} heavy quarks, that leave the medium too fast, the fragmentation is the most likely hadronisation process. The coalescence naturally enhances the baryon production with respect to the meson one and it can explain the measured p_{T} differential baryon-to-meson ratios. In Fig 2.4, the antiproton to pion ratio measured at RHIC in Au-Au collisions at $\sqrt{s_{\mathrm{NN}}} = 200$ GeV and the Λ to K ratio measured at LHC in Pb+Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV are shown. Superimposed, the expected yield ratios from different model are also reported. A good description of the observed data is achieved with a coalescence plus fragmentation model.

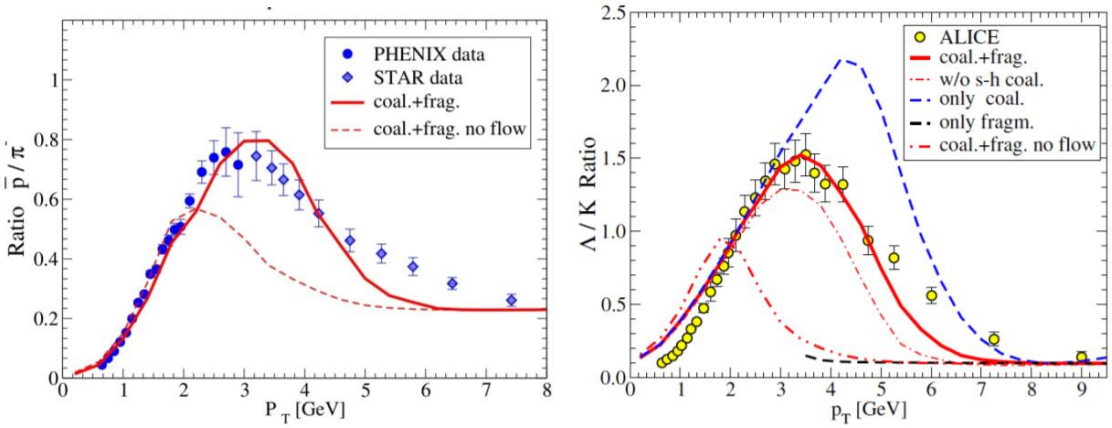

Figure 2.4: Left: Antiproton to positive pion ratio at RHIC from Au+Au collisions at $\sqrt{s_{\mathrm{NN}}} = 200$ GeV [44]. Right: Λ to K ratio in central (0-5%) Pb+Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV [45].

2.5 Experimental results

The purpose of this section is to provide an overview of the open heavy-flavour experimental results, which serves as a starting point for the study performed in this thesis.

Due to their short lifetime, the production of open heavy-flavour particles is studied through their decay products. Different analysis strategies are employed:

- Full decay chain reconstruction of the heavy-flavour particles: this is the technique utilized in this work. It will be described in detail in chapter 4.
- Semi-leptonic decay reconstruction of heavy-flavour particles: the known background sources are subtracted from the measured inclusive electron or muon spectra [46, 47].
- heavy-flavour reconstruction inside c- and b-jets [48].

Measurement of D-mesons (D^0, D^+, D^{*+}, D_s^+) and heavy-flavours decay electron and muons were performed in pp collisions at different centre of mass energies by the ALICE experiment and are found to be compatible with pQCD predictions [49, 50]. The p_T-differential production cross sections of D^0 and D^+ mesons (prompt and non-prompt) measured in pp collisions at $\sqrt{s} = 5.02$ TeV and the p_T-differential production cross sections of electrons from heavy-flavour hadron decay in pp collisions at $\sqrt{s} = 2.76$ TeV are shown, respectively, in Fig. 2.5 and 2.6. The prompt D-meson spectra lie on the `FONLL` prediction upper edge while the non-prompt D-mesons are compatible with the central values. The `PYTHIA` 8 MC generator is used to describe the beauty-hadron decay $H_b \to D + X$. The electron spectra are found to be compatible within the uncertainties with pQCD including calculations from `FONLL`, `GM-VFNS` [51] and k_T-factorization [52].

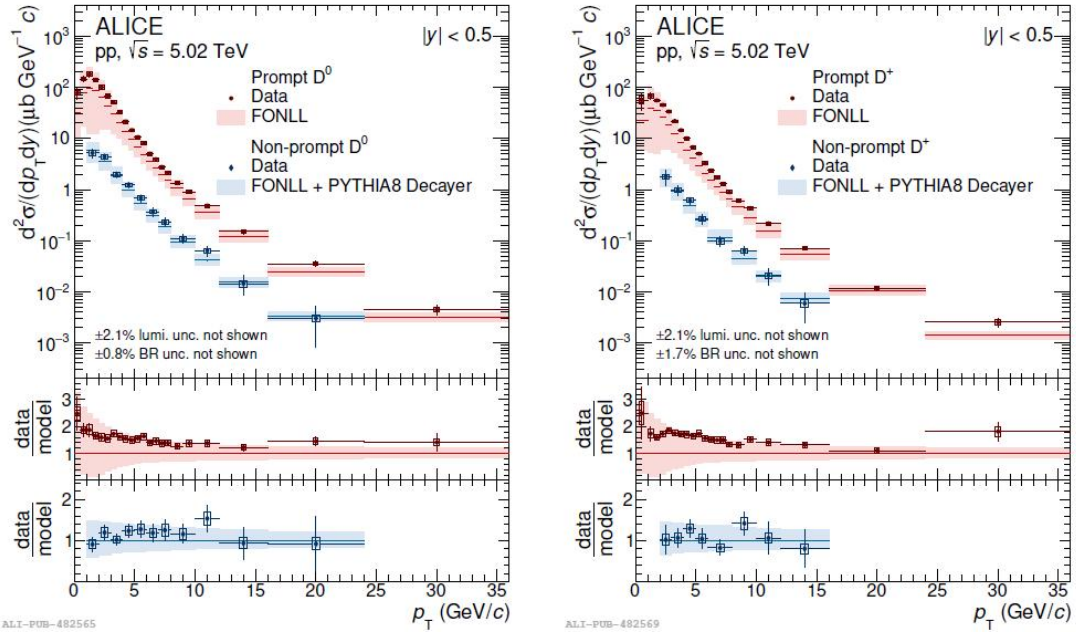

Figure 2.5: p_{T}-differential production cross sections of prompt and non-prompt D^0(left panel) and D^+ (right panel) mesons measured in pp collisions at $\sqrt{s} = 5.02$ TeV by the ALICE experiment. Superimposed `FONLL` predictions [49].

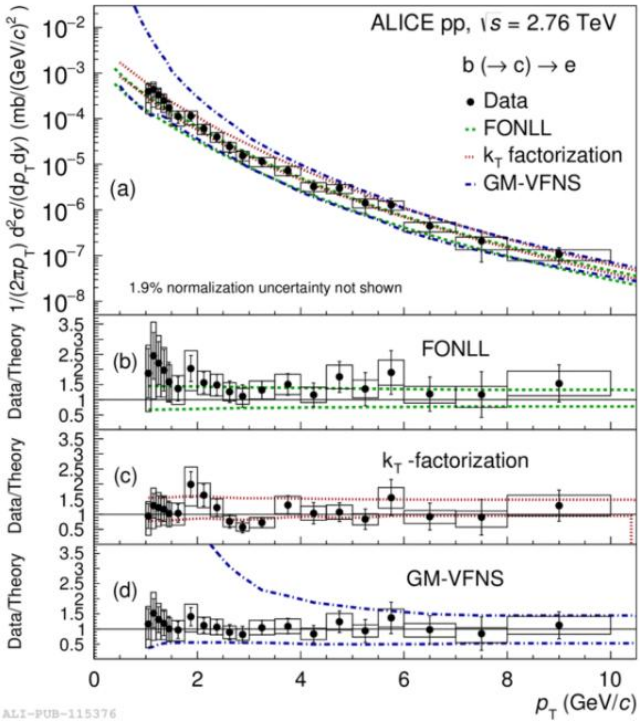

Figure 2.6: p_{T}-differential inclusive production cross sections of electrons from beauty-hadron decays compared with pQCD models (`FONLL`, `GM-VFNS`, k_T-factorization) [50].

The measurements of charm mesons and baryons (D^0, D^+, D_s^+ , Λ_c^+ and Ξ_c^0) production cross sections were employed to calculate the $c\bar{c}$ production cross section and the charm-quark fragmentation fractions at midrapidity ($|y| < 0.5$).

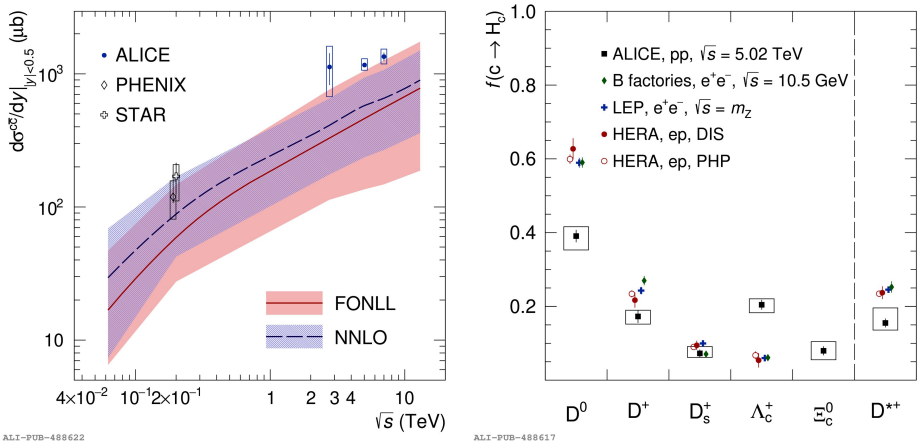

Figure 2.7: Left: $c\bar{c}$ production cross section at midrapidity per unit of rapidity as a function of the collision energy. The comparison with FONLL (red band) and NNLO (violet band) pQCD calculations is shown. Right: charm-quark fragmentation fractions into charm hadrons measured in pp collisions at $\sqrt{s} = 5.02$ TeV, compared to data from e^+e^- collisions at LEP and at B factories, and in ep collisions at HERA [53].

The behaviour of the prompt charm cross section as a function of the collision energy, measured by ALICE, STAR and PHENIX, is reported on the left panel of Fig. 2.7. The data sit on the upper edges of the NNLO and FONLL calculations. The hadron fragmentation fractions were estimated as the ratio of the hadron-production cross section over the sum of cross sections of all measured ground states of charm hadrons. The charm fragmentation fractions measured in pp collisions at $\sqrt{s} = 5.02$ TeV and the comparison with values derived from experimental measurements performed in e^+e^- collisions at LEP and B factories, as well as in ep collisions at HERA, at the DESY accelerator, are shown on the right panel of Fig. 2.7. The fragmentation fractions measured at midrapidity in pp collisions at the LHC are different from the ones measured in e^+e^- and ep collisions, providing significant evidence that the assumption of universality, i.e. independence from the collision system, of

parton-to-hadron fragmentation is not valid already in pp collisions.

Another crucial observable in the charm heavy-flavour sector, is the relative abundances of charm baryons and mesons ("baryon-to-meson ratio"). It is indeed sensitive to the charm hadronisation mechanisms. Recent observations [54] suggest that coalescence of the charm quarks with other light quarks in the hot medium is a process effective for charm hadron formation in heavy-ion collisions and concurrent with fragmentation. This is expected to produce an enhancement of the baryon-to-meson ratio at intermediate p_{T} in heavy-ion collisions with respect to measurements performed in pp collisions (see Fig. 2.8).

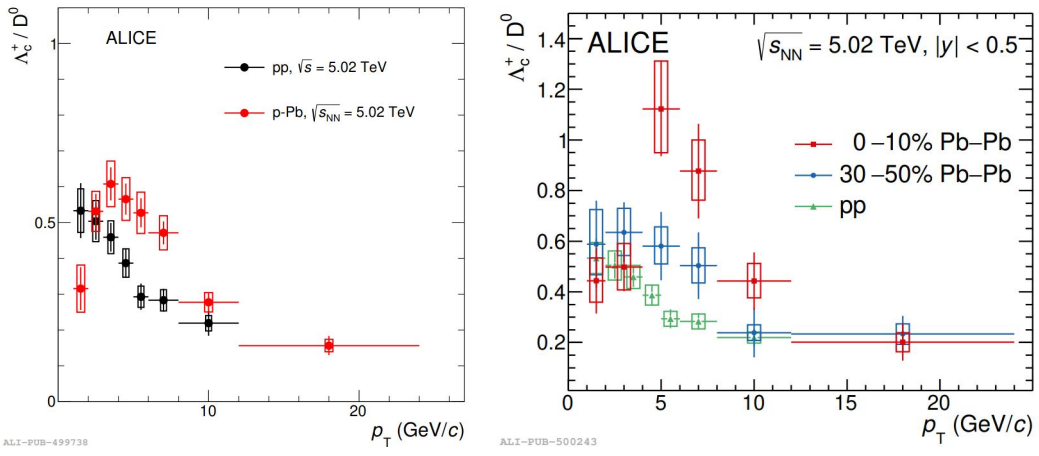

Figure 2.8: Left: Λ_c^+/D^0 baryon-to-meson ratio measured in pp collisions compared with the ratio measured in p-Pb (left) [41]. Right: Λ_c^+/D^0 baryon-to-meson ratio measured in Pb-Pb collisions in two centrality classes. The measurements were performed with the ALICE detector.

There is a hint of a higher Λ_c^+/D^0 ratio in Pb–Pb with respect to pp collisions at intermediate p_{T} (for both the 0-10% and 30-50% centrality classes). More precision is needed to investigate a trend between pp, p–Pb, and Pb–Pb collisions. But the surprising thing is that, also in pp collisions, the ratio shows significantly higher results than in e^+e^- collisions, confirming that the fragmentation of charm is not universal across different collision systems. Therefore, first of all, it is fundamental to improve the understanding of Λ_c^+/D^0 production in pp collisions. In Fig. 2.9 the baryon-to-meson ratio Λ_c^+/D^0 measured in pp collisions at $\sqrt{s} = 5.02$ TeV as a function of p_{T}

is shown and compared with predictions from several MC generators and models where different hadronisation processes are implemented. The PYTHIA 8 Monash tune and HERWIG 7.2 predictions are driven by the fragmentation fraction $f(c \to \Lambda_c)$ implemented in these generators, that is tuned on e^+e^- collisions. They predict a relatively constant Λ_c^+/D^0 ratio versus p_T of about 0.1, significantly underestimating the data at low p_T. A significant enhancement of the Λ_c^+/D^0 ratio is seen with colour reconnection beyond the leading-colour approximation (PYTHIA 8 CR Mode 2). This prediction is consistent with the measured ratio. The statistical hadronisation model, SH, uses either an underlying charm-baryon spectrum taken from the PDG [24], or includes additional excited charm baryons that have not yet been observed but are predicted by the Relativistic Quark Model (RQM) [55]. These additional states decay strongly to Λ_c^+ baryons, which contribute to the prompt Λ_c^+ spectrum. With the PDG charm-baryon spectrum the model underpredicts the data while with the additional baryon states the model gives a good description of the pp data. It is interesting to note that also the Catania model [56], where a colour-deconfined state of matter formation is assumed and hadronisation can occur via coalescence in addition to fragmentation, provides a reasonable description of the data, reproducing the shape and the magnitude of the data over the full p_T range.

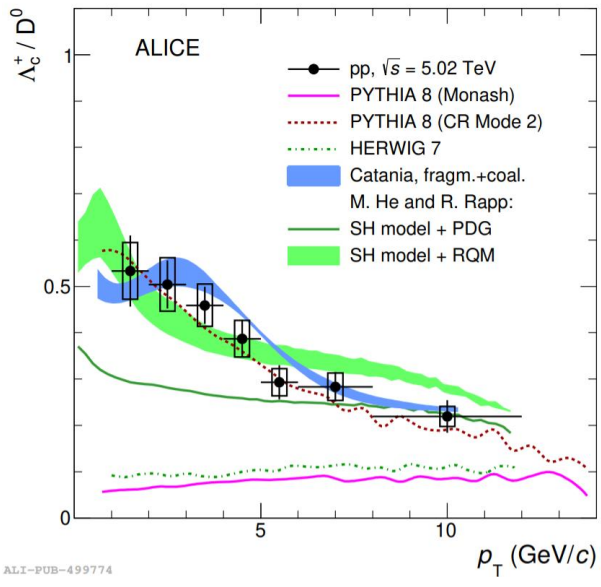

Figure 2.9: The Λ_c^+/D^0 ratio as a function of p_T measured in pp collisions at $\sqrt{s} =$ 5.02 TeV by the ALICE collaboration compared with theoretical predictions [40].

Finally, the Λ_c^+ R_{AA} is shown in Fig. 2.10. A suppression is observed for the Λ_c^+ in Pb–Pb collisions, with a hint of a larger suppression for central collisions.

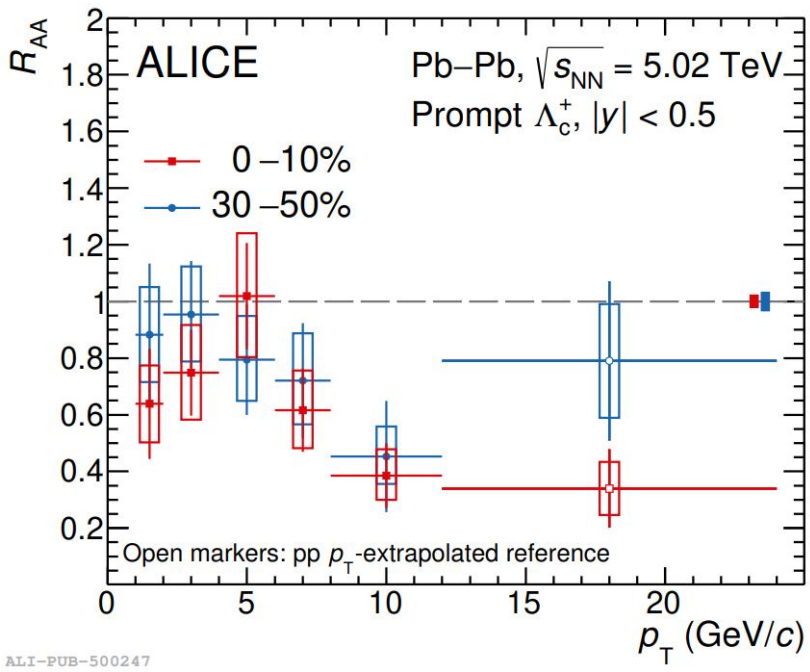

Figure 2.10: The R_{AA} of the Λ_c^+ baryon in the 0-10% and 30-50% centrality classes.

The results shown in this chapter suggest modified or additional hadronisation mechanisms in small hadronic collision systems with respect to fragmentation in vacuum that still have to be fully understood. The aim of this work is to characterise the evolution of the Λ_c^+/D^0 baryon-to-meson ratio from very low to moderate charged-particle density and provide new experimental constraints on the nature of these modifications in pp collisions. In particular, we would like to verify whether the observed trend versus p_T in small systems becomes more similar to those observed in p-A and A-A collisions when high multiplicity events, with a charged-particle density closer to the one produced in large systems, are considered. Possible common mechanisms can be identified and the effects not arising from the hot medium production can be isolated.

Chapter 3

The ALICE experiment at the LHC

A Large Ion Collider Experiment (ALICE) [57] is one of the four main experiments at the Large Hadron Collider (LHC) [58], the accelerator built to collide protons and ions at unprecedented energies (TeV scale). ALICE has been designed and optimised for the high-multiplicity environment of heavy-ion collisions, to study the deconfined state of matter known as the Quark-Gluon Plasma. This chapter will give a detailed description of the ALICE detector system employed to collect the data analysed in this work.

3.1 The Large Hadron Collider

The Large Hadron Collider, installed at the European Organisation for Nuclear Research (CERN) laboratories, is the largest and most powerful particle collider in the world. It is a circular accelerator 27 km long, made of two rings of superconductive magnets, located between 45 and 170 m underground (it is 1.3° inclined with respect to the horizontal plane). To keep the charged particles in a circular orbit, 1600 superconducting magnets in niobium and titanium alloy, operating at the temperature of 1.9 K and cooled with superfluid helium, are used. The magnetic field established is ≈8 T. Along its circumference, four main experiments are located in four different

interaction points: ATLAS (A Toroidal LHC ApparatuS), CMS (Compact Muon Solenoid), LHCb (Large Hadron Collider beauty) and ALICE. The ATLAS and CMS experiments are general purpose detectors, designed to confirm the existence of the Higgs boson, discovered in 2012, and to search for physics beyond the Standard Model (SM). The LHCb experiment is focused on the beauty physics and in particular on the study of the CP-symmetry violation.

The LHC was constructed to collide protons with a maximum centre-of-mass energy of $\sqrt{s} = 14$ TeV at a peak luminosity[1] of $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1} as well as Pb ions at a centre-of-mass energy per nucleon of $\sqrt{s_{\mathrm{NN}}} = 5$ TeV at a peak luminosity of $\mathcal{L} = 10^{27}$ cm^{-2}s^{-1}. To reach such high energies, a complex injection chain, in which the particles are pre-accelerated, is needed. Protons are first accelerated by a linear accelerator (LINAC 2), then are passed into the Proton Synchrotron Booster, and sequentially accelerated via the Proton Synchrotron (PS) and Super Proton Synchrotron (SPS). Only after all these steps, they are injected into the LHC at the energy of 450 GeV. Lead nuclei follow a similar path, but they are accelerated via the LINAC 3 and Low-Energy Ion Ring (LEIR) before being passed into the PS. Their injection energy at the LHC is 177 GeV per nucleon. Once injected, the beams are further accelerated around the LHC ring before being brought to collision in the interaction points (IP) of each experiment. The CERN accelerator complex is shown in Fig. 3.1.

[1] The instantaneous luminosity depends on the particle beam parameters and it can be defined as the ratio of the number of inelastic events in a certain period of time to the pp inelastic cross-section:

$$\mathcal{L} = \frac{1}{\sigma_{\mathrm{INEL}}} \frac{\mathrm{d}N_{\mathrm{ev}}}{\mathrm{d}t} \tag{3.1}$$

The integrated luminosity is the integral of the luminosity with respect to time:

$$\mathcal{L}_{\mathrm{int}} = \frac{N_{\mathrm{ev}}}{\sigma_{\mathrm{INEL}}} \tag{3.2}$$

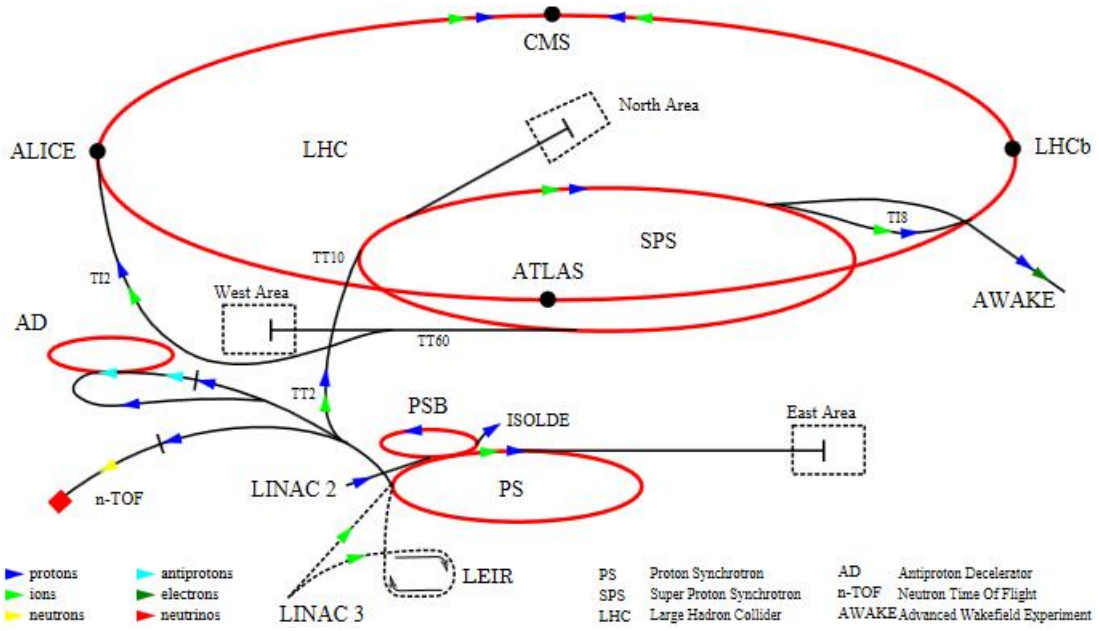

Figure 3.1: Schema of the CERN accelerator complex. The locations of the four main LHC experiments (ALICE, ATLAS, CMS and LHCb) around the LHC ring is also shown.

3.2 The ALICE Detector

ALICE is a general-purpose detector designed to study the hadronic matter under extreme conditions of temperature and energy density, as those reached fractions of seconds after the Big Bang, to verify the QCD predictions about the existence of a phase transition from the common hadronic matter to the QGP, with quarks and gluons not confined into hadrons. ALICE is also interested in pp collisions, needed as a reference case, and p-nucleus collisions to disentangle the effects related to the presence of nuclei in the colliding system that could mimic final-state medium-related effects (Cold Nuclear Matter effects). Performing measurements in heavy-ion collisions requires the capability to work in the high-track density environment, with a charged particle multiplicity at mid-rapidity in a central Pb–Pb event up to $dN/d\eta = 2500$. To face the high particle density, detectors with high granularity were adopted in order to reduce the detector occupancy. Moreover, an excellent tracking down to very low p_T and a particle identification (PID) in a wide momentum

range are crucial features for a detector with such an ambitious target. To fulfil these requirements, low material budget to reduce multiple scattering at low p_{T} and larger tracking distance to improve the resolution at high p_{T} were adopted and many different PID techniques were combined. The final ALICE setup until the LHC Run 2 is shown in Fig. 3.2. In particular, ALICE consists of two main parts: the central barrel and a muon arm, covering, respectively, the mid-rapidity region ($|\eta| < 0.9$) and the forward rapidity range -4 $< \eta <$ -2.5. The central-barrel detectors sit inside the L3 magnet, which is a large solenoid providing a magnetic field of 0.5 T along the beam axis, needed for the tracking and the identification of the charged particles. From the inside out, the central barrel is composed of the Inner Tracking System (ITS), the Time Projection Chamber (TPC), the Transition Radiation Detector (TRD), the Time Of Flight (TOF) detector, the Ring Imaging Cherenkov (HMPID) detector, and the electromagnetic calorimeters PHOS, EMCal and DCal. ALICE includes also several other detectors for global event characterisation and triggering (ZDC, PMD, FMD, T0, V0A and V0C) located at forward and backward rapidity and an array of scintillators (ACORDE) located on top of L3 magnet, used to trigger on cosmic rays. A more detailed description of all the subdetectors is provided in the next sections.

The ALICE coordinate system is schematized in Fig. 3.3. It is a right-handed Cartesian system. The z axis corresponds to the beam direction, pointing in the opposite site of the muon spectrometer, while the transverse plane is defined by the x and y coordinates. In particular, the x axis points toward the accelerator centre, while the y axis is oriented upward. The positive and negative side of the z axis are called side A and side C, respectively, and will be used to identify the subdetector position with respect to the interaction point. The polar angle θ is defined with respect to the z direction while the azimuthal angle ϕ increases counter-clockwise starting from the x axis towards the y one.

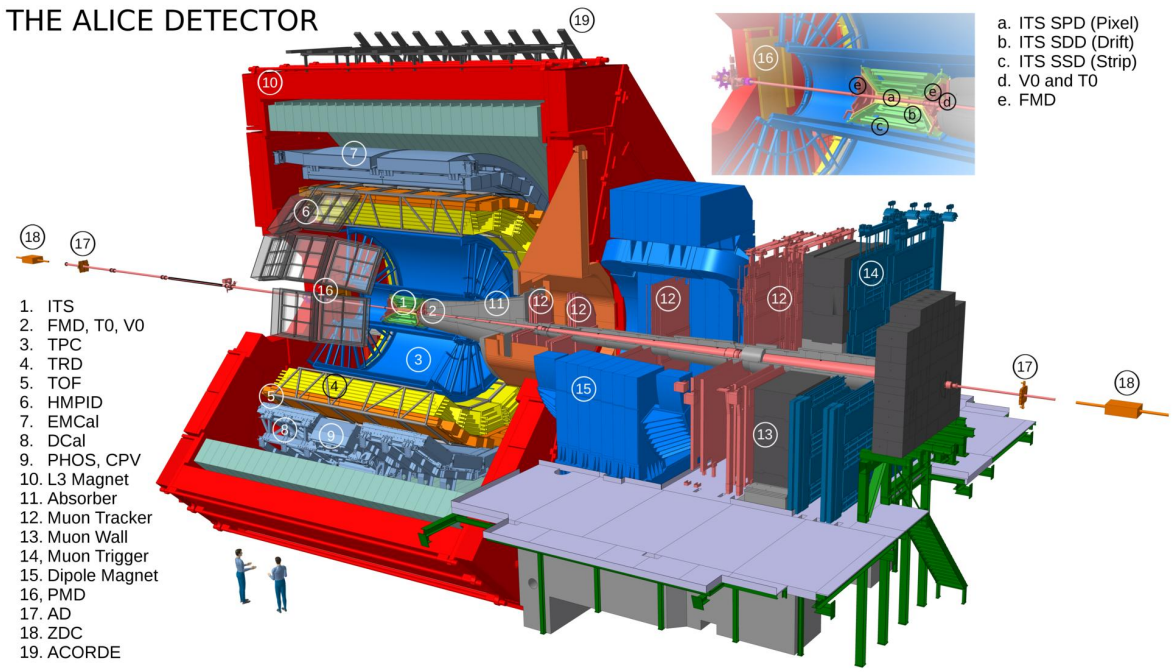

Figure 3.2: The ALICE experimental apparatus with individual sub-detectors labelled. The inlay at the top-right shows a zoomed-in view of the beam line (pink) and of the Inner Tracking System.

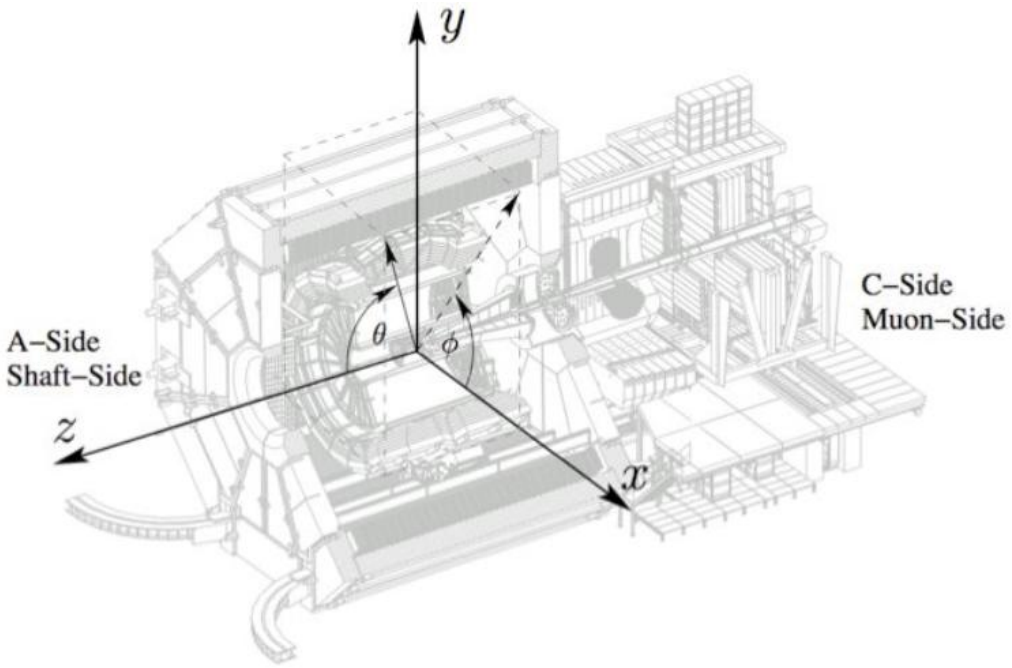

Figure 3.3: Schema of the ALICE coordinate system.

3.2.1 Inner Tracking System

The Inner Tracking System (ITS) is the detector closest to the IP and it surrounds the beam pipe. The main purposes of the ITS detector are:

- determination of the position of the primary vertex (position of the point of interaction between the two colliding beam particles) and measurement of the impact parameter (distance of closest approach between the track trajectory and the primary vertex);
- improve the reconstruction of the secondary (decay) vertices, i.e., originating from the decay of strange, charm and beauty particles;
- tracking and low-momentum particle identification ($p_{\mathrm{T}} < 200$ MeV/c);
- measurement of the event charged-particle multiplicity.

The ITS is a cylindrical silicon pixel detector made by six concentric layers built by exploiting three different technologies: the two innermost layers are silicon pixel detectors (SPD), the third and the forth ones are silicon drift detectors (SDD), and the two outermost layers are silicon strip detectors (SSD). The ITS layout is shown in Fig. 3.4. The SPD layers are fundamental for the primary- and decay-vertices reconstruction, as well as for the measurement of the impact parameter, which is crucial for the measurement of secondary tracks originating from the weak decays of strange, charm and beauty particles. The high SPDs granularity allows for a spatial precision of 12 μm in the $r\phi$ direction and 100 μm in the z direction. The SDDs and the SSDs purposes are the tracking and the particle identification via specific energy loss ($\mathrm{d}E/\mathrm{d}x$) measurements, described by the Bethe-Bloch formula. The ITS dimensions are chosen in order to optimize the reconstruction of the tracks. To guarantee the best possible resolution on the impact parameter, the first SPD layer must be as close as possible to the IP. The minimum value allowed is determined by the presence of the beam pipe, that has a radius of $\sim$3 cm. For this reason, it is not possible to reduce the SPD dimension to a radius that is less than 4 cm. This

configuration allows one to achieve, for transverse momentum higher than 1 GeV/c, a resolution on the impact parameter lower than 60 μm in the $r\phi$ direction. Concerning the ITS external radius, it is ∼44 cm and it is determined by the need to combine the tracks reconstructed by the ITS with those in the TPC. The ITS covers the pseudorapidity[2] range $|\eta| < 0.9$. The innermost layer has a wider acceptance of $|\eta| < 1.98$, that allows, together with the FMD, having a full pseudorapidity coverage and measuring the event charged-particles multiplicity in the full range. More details about the multiplicity measurement with the SPD are given in section 4.2. The ITS detector, like all semiconductor detectors, undergoes some degradation due to the radiation passing through them. These effects have been studied in detail and are discussed in sec. 4.2.1.

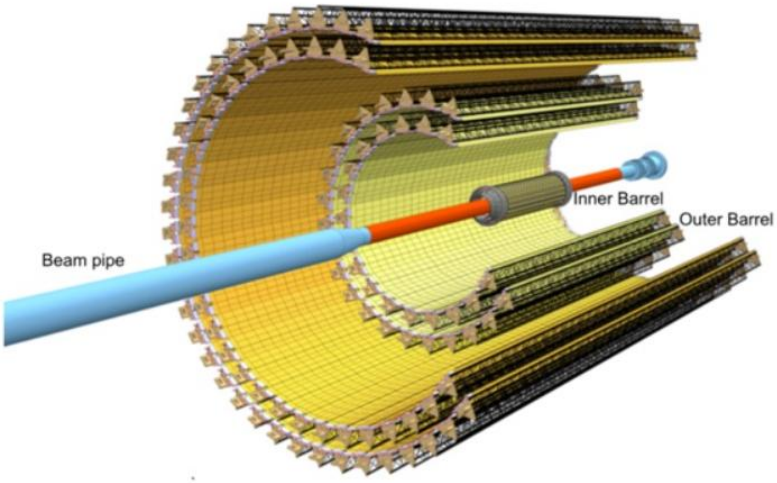

Figure 3.4: Layout of the ALICE ITS detector.

[2] The rapidity y of a particle with mass m and energy E is defined in natural units as:

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right) \tag{3.3}$$

where p_z is the longitudinal momentum. At very high energy or in the massless limit, the rapidity reduces to the pseudorapidity η, defined as:

$$\eta = \frac{1}{2}\ln\left(\frac{|\vec{p}|+p_z}{|\vec{p}|-p_z}\right) = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{3.4}$$

where θ is the particle-emission angle relative to the beam axis.

3.2.2 Time Projection Chamber

The Time Projection Chamber (TPC) is the main tracking detector of the ALICE barrel. It is a cylindrical detector with the active volume spanning a radial position of $85 < r < 247$ cm and a total length of 5 m. It covers the pseudorapidity interval of $|\eta| < 0.9$ and the full azimuth. The active volume of $\sim$90 m^3 is filled with a gas mixture chosen to optimize the drift speed, to keep the diffusion low avoiding multiple scattering, to have a low radiation length and to be stable (it must not change its chemical composition). When a charged particle generated in the collision passes through the chamber, it ionizes the gas releasing electrons. In the middle of the chamber there is a 30 μm thick electrode that divides the chamber into two parts. A uniform electric field, which will collect electrons towards the readout planes on the two edges, is established. Each end plate is divided in 18 trapezoidal sectors, where multi-wire proportional chambers (MWPC) are mounted. The freed electrons cause an electron avalanche in the MWPC, which reaches the cathode pad readout. The hit location on the cathode gives the two-dimensional track position in $r\phi$, and the time taken for the electrons to drift to the end plates gives the track position in z. The maximum drift time of the electrons is $\sim$88 μs, setting in this way a limit on the sustainable event rate of the TPC. At high interaction rates, pile up becomes relevant. This effect can be rejected considering that tracks from pileup point to a different primary vertex.

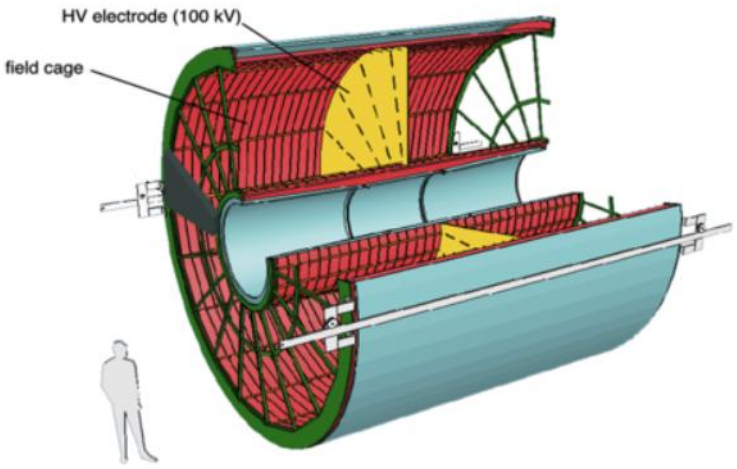

Figure 3.5: Layout of the ALICE TPC detector.

With this configuration, shown in Fig. 3.5, the detector provides a very high tracking efficiency, $> 90\%$, for tracks of p_{T} between 100 MeV/c and 100 GeV/c and guarantee a good two-tracks separation with a momentum resolution of few (< 5) MeV/c. Measuring the tracks deflection in the magnetic field, the TPC is able to determine the momentum of charged particles with a resolution better than 1% at low p_{T} and better than 20% for p_{T} ~100 GeV/c.

PID with the TPC

The TPC is also crucial for the PID of charged particles, performed via a measurement of the specific energy loss $\mathrm{d}E/\mathrm{d}x$, which is directly related to the number of electrons ionised by the charged particle propagating through the TPC. The energy loss distributions in the TPC for protons, pions and kaons are well separated for momentum lower than 1 GeV/c while the deuterons until 2 GeV/c, as can be seen in Fig. 3.6.

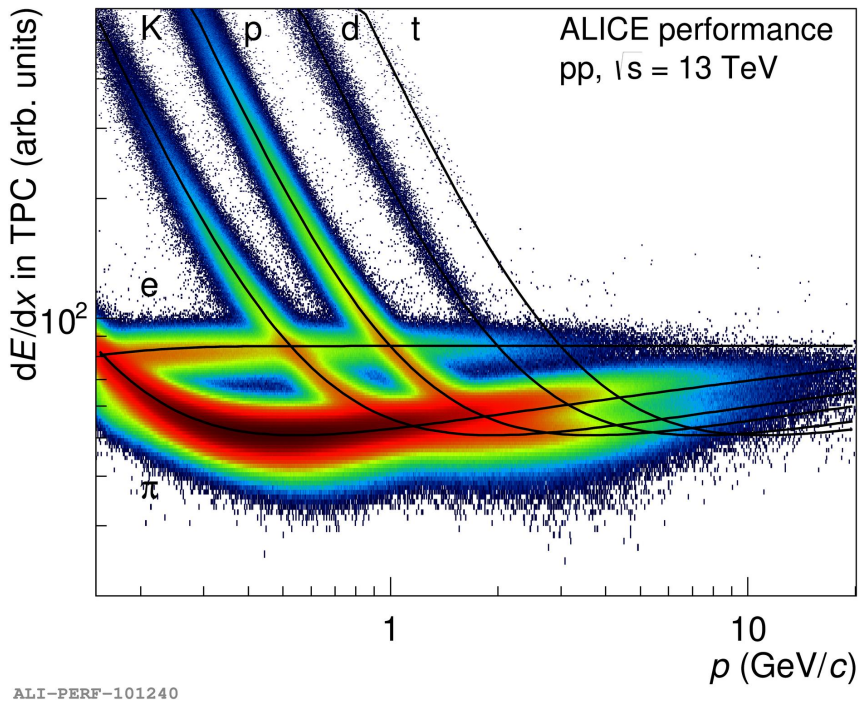

Figure 3.6: Distribution of $\mathrm{d}E/\mathrm{d}x$ for charged particles as a function of particle momentum, measured by the TPC in pp collisions at $\sqrt{s} = 13$ TeV. The overlapped black lines are the $\mathrm{d}E/\mathrm{d}x$ parametrizations, based on the Bethe-Bloch formula.

In most of the ALICE analysis (and in this analysis as well), to verify if a track is compatible with a certain species hypothesis and to quantify how the agreement is good, the "number of sigmas" variable (nσ) is defined. The nσ is unitless and describes the distance of a signal from the expectation value in units of the detector resolution:

$$\mathrm{n}\sigma_{\mathrm{TPC}}(i) = \frac{(\mathrm{d}E/\mathrm{d}x)_{\mathrm{exp}} - (\mathrm{d}E/\mathrm{d}x)_{\mathrm{th}}(i)}{\sigma_{\mathrm{TPC}}} \tag{3.5}$$

where σ_{TPC} represents the detector energy loss resolution, $\mathrm{d}E/\mathrm{d}x_{\mathrm{exp}}$ is the measured energy loss and $\mathrm{d}E/\mathrm{d}x_{\mathrm{th}}$ is the theoretical expectation for a species i, based on the Bethe-Bloch formula. It is possible to make simple selections based on the nσ variable; typically this will involve a cut of $|n\sigma| <$ n, in which all particles lying within the given region around the expectation are accepted as being compatible with that hypothesis. In regions where the energy loss distributions are well separated for different species, this leads to a clear identity being assigned to each track; in other cases, a particle may be selected as being compatible with multiple mass hypotheses.

3.2.3 Transition Radiation Detector

The Transition Radiation Detector (TRD) is a cylindrical detector placed between the TPC and the TOF. The inner radius is 2.90 m while the outer one is 3.68 m. The detector covers the pseudorapidity region $|\eta| < 0.9$ and the full azimuth. It has a modular structure: it is divided in 18 sectors (called supermodules) arranged along the longitudinal direction and 7.8 m long. Each sector is divided in 5 modules and each of them is in turn made up of 6 layers along the radial direction (TRD chambers). So, in total, the TRD includes 540 detectors and its structure is reported in Fig. 3.7. Each detector element consists of a radiator of 4.8 cm thickness, a drift section of 30 mm thickness and a multiwire proportional chamber (MWPC) with pad readout. The main purpose of the TRD is to provide electron identification for $p_{\mathrm{T}} > 1$ GeV/c. The principle of operation is based on the transition radiation (TR) emitted by particles passing at very high speed ($\gamma \sim$1000) through a material with discontinuity in the refractive index. The TRD can detect transition radiation in

order to discriminate electrons from other particles passing through the detector that do not generate any TR. The high-p_{T} electron identification is crucial for the study of D- and B-mesons through their semi-leptonic decays and the production of vector meson resonances (like J/Ψ, Υ, Υ') through their leptonic decay channel $e^{+}e^{-}$.

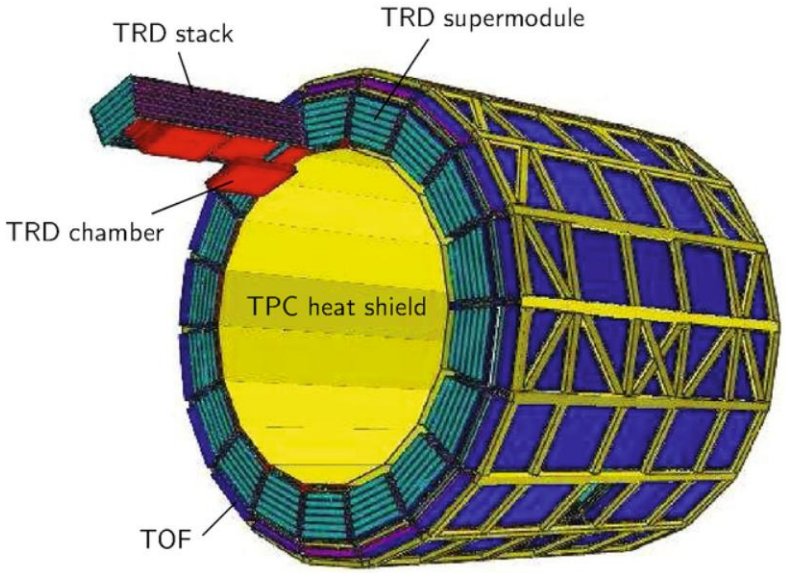

Figure 3.7: ALICE TRD modular structure.

3.2.4 Time Of Flight

The Time Of Flight (TOF) detector is a large cylindrical array of Multi-gap Resistive Plate Chambers (MRPCs) placed around the TRD, at a radial distance of 370–399 cm from the beam line. Its main purpose is the particle identification in the intermediate momentum range. It also provides the trigger for cosmic rays analyses, taking back-to-back coincidence between two hits in two different supermodules and for the ultra-peripheral collision, characterized by the presence of only a few tracks coming from resonance decays (J/Ψ, ρ, ϕ). It covers the rapidity region of $|\eta| < 0.9$ and the full azimuth, except for the region $260 < \phi < 320$ and $|\eta| < 0.12$ in which the TOF modules have not been installed to allow a good photons measurement with the PHOS detector (3.2.6) reducing the material budget in front of it. The active length of the TOF is 7.41 m corresponding to a total active surface of 141 m^2. The TOF detector has a modular structure. It is made up of 18 linear sectors called

supermodules, positioned on the surface of the cylinder, along the beam direction. Each supermodule is composed by 5 modules in which the MRPC detectors are contained. To guarantee a low detector occupancy, needed to face the high particle rate, each MRPC strip is segmented into two rows of 48 pickup pads, with a total of more than 10^5 independent readout channels. On the edge of each supermodule, two crates, in which the read-out electronic is placed, are installed. The modular structure of the TOF detector is shown in Fig. 3.8.

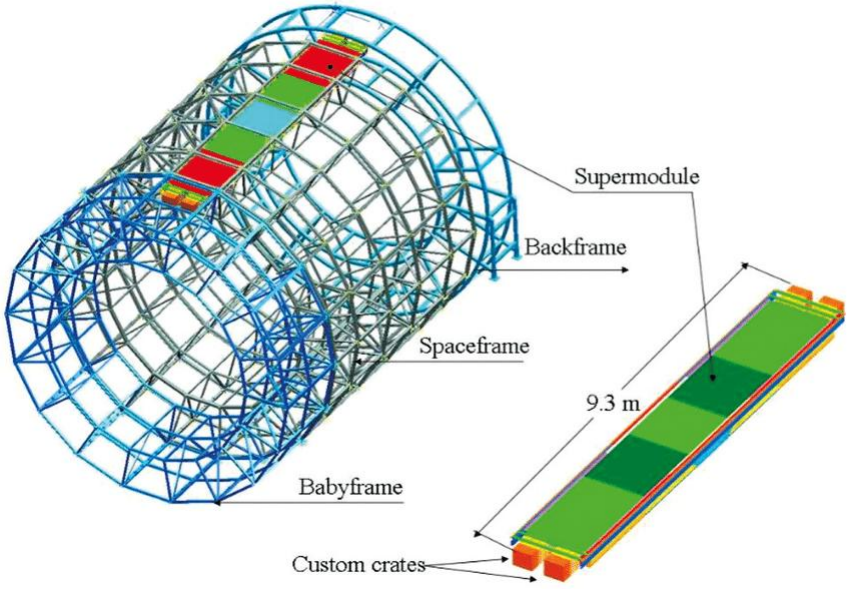

Figure 3.8: Modular structure of the ALICE TOF detector.

Multi-gap Resistive Plate Chamber

The MRPCs are gaseous parallel detectors derived from the Resistive Plate Chamber (RPC). The RPCs consist of two electrodes separated by a single gas-filled gap, across which a uniform electric field is established. When a charged particle passes through the gas, it ionizes it and the electrons produced are collected by the electric field. While they move towards the anode, they accelerate due to the potential difference, acquiring enough energy to cause secondary ionizations. In this way, an avalanche intense enough to produce a detectable signal is formed. The RPCs have some limits that do not make them suitable to be used in the ALICE TOF detector. In particular, to reach the required timing resolution, a high electric field needs to be

applied between the electrodes, enhancing the chance to have sparks in the chamber with the risk of damaging the pick-up electrodes. Moreover, with this configuration, the RPCs cannot manage the huge amount of charges produced in heavy-ion collision that would collect at the electrodes reducing the intensity of the electric field and badly affecting the efficiency. This effect can be limited reducing the electric field applied (and so the avalanche) but it would lead to a worsening of the time resolution that is not compatible with the experiment requirements. To cope with the high rate and to match the timing resolution requirements, the MRPC were employed. The RPC initial gap is divided in many smaller and identical gaps, through resistive plates, parallel to the two outer electrodes and made of the same material. These plates reach a certain voltage value due to the electrostatic field and the potential difference across all the sub-gaps is the same. In the MRPC, the dimension of the avalanche and the amount of charges collected are limited by the small dimensions of the gaps. The resistive internal gaps are transparent to the fast signals generated by the avalanches inside each gap. So the induced signal on the external electrodes is the sum of the signals induced by the single avalanches in the single gaps. Many gaps allow for the achievement of high efficiency, the small size of each gap is aimed to achieved a good time resolution. A schematic representation of the RPC and MRPC operation is shown in Fig. 3.9.

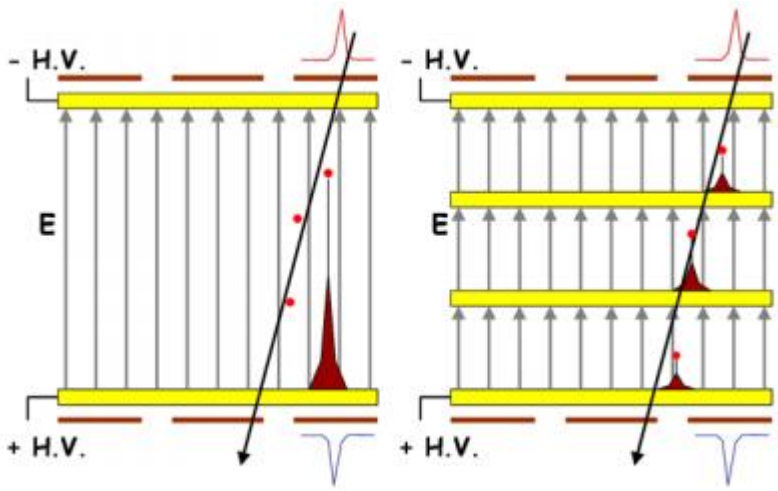

Figure 3.9: Schematic representation of the RPC and MRPC operation.

The TOF MRPC is based on a double-stack design: it is made of two stacks of five gas gaps of 250 μm, built on each side of the anode plane (Fig. 3.10). What it is measured hence is the sum of the signals from the two stacks. In this way, the collected signal is the same with respect to a single-stack MRPC, while the applied voltage is half. The performance achieved with this configuration are an efficiency close to 100% and an intrinsic time resolution better then 50 ps (which includes also the contribution from the full electronic chain).

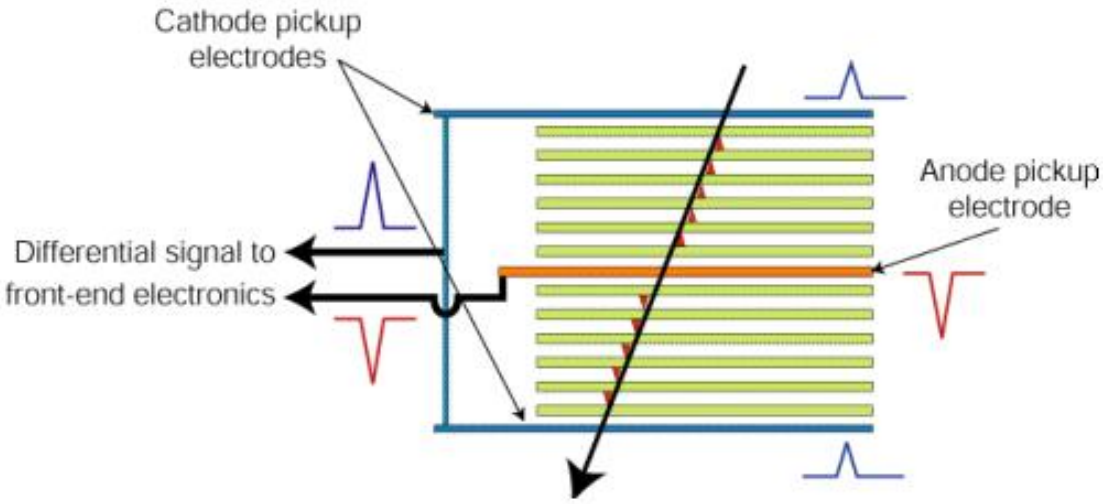

Figure 3.10: Schematic representation of a double-stack MRCP.

PID with the TOF

The TOF detector measures the time t needed for a particles coming from the interaction point to reach the sensitive surface of the detector. To estimate the particle travelling time, a start time, corresponding to the time of the event, has to be provided to the TOF. The start time, t_{Event}, can be provided by the T0 detector (section 3.2.10). However, the T0 has a limited acceptance and, especially in pp collisions, the information could be available only for a fraction of the events [59]. In order to overcome this limitation, an alternative method for the t_{Event} determination was developed by using the TOF itself when at least three tracks have an associate TOF signal [59]. A combinatorial algorithm compares the measured TOF times and the expected times, assuming a common t_{Event}. The t_{Event} is then obtained from a

χ-minimization procedure where all the possible mass combinations for the matched tracks are considered. If both the measurements are available, the final $t_{\rm Event}$ is calculated as a weighted average between T0 and TOF algorithms. Combining the time of flight information with that of the particle momentum p, provided by the tracking detectors (ITS and TPC), it is possible to determine the particle mass m as:

$$\begin{aligned} m &= \frac{p}{c\beta\gamma} \\ &= \frac{p}{c}\sqrt{\frac{c^2t^2}{L^2}-1} \end{aligned} \tag{3.6}$$

where L is the measured track length. The TOF detector allows one to distinguish two particles with the same track length, same momentum but different masses m_1 and m_2 measuring the different travelling time employed:

$$\begin{aligned} \Delta t &= \frac{L}{c}\left(\frac{1}{\beta_1}-\frac{1}{\beta_2}\right) \\ &= \frac{L}{pc^2}(E_1-E_2) \\ &\simeq \frac{Lc(m_1^2-m_2^2)}{2p^2} \end{aligned} \tag{3.7}$$

If the quantity on the right is comparable with the TOF time resolution, the two particles will not be clearly identified. Therefore, the ability of the TOF to distinguish two particles is quantified by the number $\mathrm{n}\sigma_{1,2}$ defined as follows:

$$\mathrm{n}\sigma_{1,2} = \frac{t_1-t_2}{\sigma_{\rm TOT}} = \frac{Lc}{2p^2}\frac{(m_1^2-m_2^2)}{\sigma_{\rm TOT}} \tag{3.8}$$

where $\sigma_{\rm TOT}$ is the overall time resolution. So, the TOF performance for the particle identification depends on the reached time resolution. The total time resolution, $\sigma_{\rm TOT}$, is given by the sum of several contributes:

$$\begin{aligned}(\sigma_{\mathrm{TOT}})^2 &= (\sigma_{\mathrm{TOF}})^2 + (\sigma_{\mathrm{Trk}})^2 + (\sigma_{\mathrm{Event}})^2 \\ &= (\sigma_{\mathrm{MRPC}})^2 + (\sigma_{\mathrm{TDC}})^2 + (\sigma_{\mathrm{Clock}})^2 + (\sigma_{\mathrm{FEE}})^2 + (\sigma_{\mathrm{Cal}})^2 + (\sigma_{\mathrm{Trk}})^2 + (\sigma_{\mathrm{Event}})^2\end{aligned} \tag{3.9}$$

where σ_{TOF} is the TOF time resolution, σ_{Trk} is the tracking particle jitter[3], negligible for tracks with a momentum higher than 1 GeV/c and σ_{Event} is the resolution on the event time[4]. The TOF time resolution is given by several contributions: σ_{MRPC} ($\sim$ 30 ps) and σ_{TDC} ($\sim$ 20 ps) are, respectively, the intrinsic resolution of the MRPCs and of the Time to Digital converter (TDC) cards that digitalize the signal; σ_{Clock} ($\sim$ 15 ps) is the resolution related to the LHC clock signal, σ_{FEE} ($\sim$ 10 ps) is the front-end electronic time resolution and σ_{Cal} summarises the uncertainties on the calibration parameters (cable lengths, paths on the readout boards, ecc.) and it depends on the calibration procedure.

As already explained in section 3.2.2, a nσ is defined for the TOF detector as well:

$$\mathrm{n}\sigma_{\mathrm{TOF}}(i) = \frac{(t_{\mathrm{TOF}} - t_{\mathrm{Event}}) - t_{\mathrm{exp}}(i)}{\sigma_{\mathrm{TOT}}} \tag{3.11}$$

where t_{TOF} is the time measured by the TOF detector, t_{Event} is the event time and $t_{\mathrm{exp}}(i)$ is the time of flight expected for a species i. As can be seen in Fig. 3.11, the TOF can provide a K/π separation up to 2.5 GeV/c and a p/K separation up to 4 GeV/c, extending the PID coverage provided by the TPC detector.

[3] After the particle is reconstructed by other ALICE detectors, it must be extrapolated to the active area of TOF; here a matching window of 10 cm for pp collisions (3 cm for Pb-Pb collisions) is opened around the extrapolation point. The matching algorithm then searches for at least one signal and the closest one is combined to the track. So, a time jitter is caused by fluctuations in the position of the initial clusters of primary ionization.

[4] The event time resolution σ_{Event} depends on TOF track multiplicity, i.e. the number of tracks matched with a hit on the TOF detector. The σ_{Event} improves with higher track multiplicity as:

$$\sigma_{\mathrm{Event}} = \frac{\sigma_{\mathrm{TOT}}}{\sqrt{\mathrm{TOF\ track\ multiplicity}}} \tag{3.10}$$

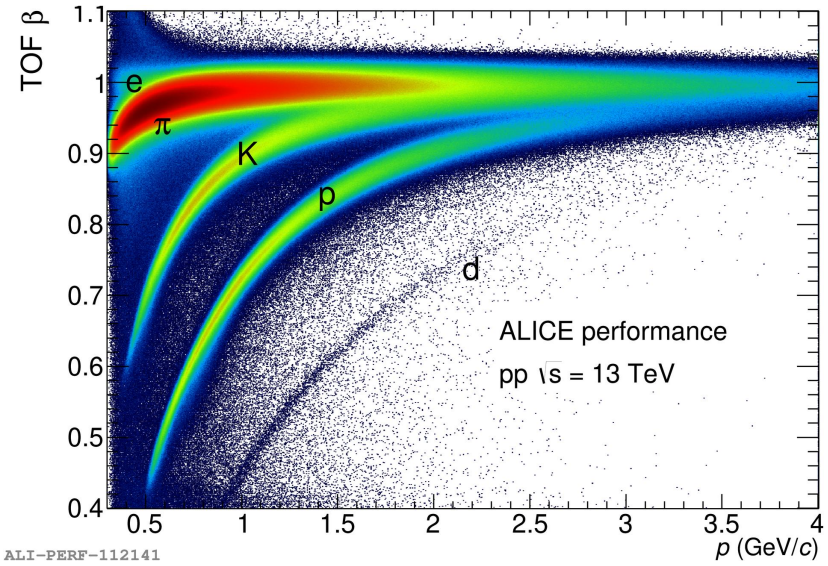

Figure 3.11: TOF-measured particle β as a function of particle momentum, measured by the TOF in pp collisions at $\sqrt{s} = 13$ TeV.

TOF Data Quality Assurance offline

The performance of all the ALICE detectors is monitored by online and offline algorithms to assess the quality of the recorded data and simulations and spot possible issues in the data taking conditions of the detector or in the reconstruction and calibration. This task is done by the TOF offline Quality Assurance (QA) expert and I personally covered this role for the whole 2021. The TOF QA algorithms validate the recorded data by checking the information on PID, track matching, calibration and event time determination for all reconstructed runs. Some of the QA checks usually performed are described in this section.

The difference between the time measured by TOF and the expected time, calculated during reconstruction for a given particle species, $(t_{\mathrm{TOF}} - t_{\mathrm{exp}})$, should be centered at zero. To test if the detector is properly calibrated, the $(t_{\mathrm{TOF}} - t_{\mathrm{exp}})$ distribution is checked by the offline QA. For each run, the mean value from the gaussian fit of the $(t_{\mathrm{TOF}} - t_{\mathrm{exp}})$ distribution is calculated and only the runs with an offset within 10 ps are flagged as good ones. In Fig. 3.12, the $(t_{\mathrm{TOF}} - t_{\mathrm{exp}})$ alignment for a set of 2018 pp runs at $\sqrt{s} = 13$ TeV is shown. The average difference stays close to 0, with an offset per run lower than 10 ps.

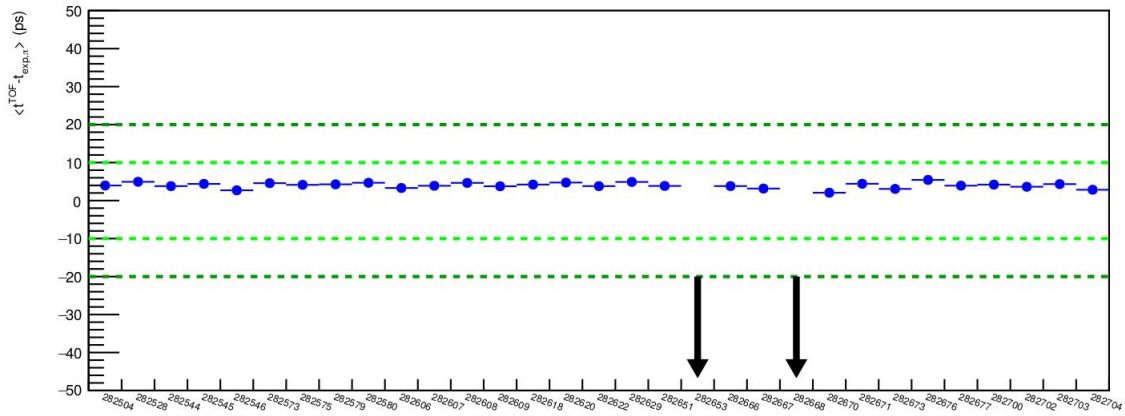

Figure 3.12: Trending distribution of the $(t_{\mathrm{TOF}} - t_{\mathrm{exp}})$ alignment for a set of 2018 pp runs at $\sqrt{s} = 13$ TeV. Each value per run is the mean value from the gaussian fit of the $(t_{\mathrm{TOF}} - t_{\mathrm{exp}})$ distribution. The black arrows point to runs in which the TOF was absent.

The offline QA checks the quality of the tracking process as well. The TOF track matching procedure has the purpose to assign a TOF cluster to each track propagated outside the TPC. A research region is defined around the position in which the extrapolated track would impinge on the TOF active area. TOF hits in this matching window (3 cm for high multiplicity events like those in Pb-Pb collisions and 10 cm for low multiplicity one like pp, pPb and peripheral Pb-Pb collisions) are selected as matchable candidates. The track is then matched to the nearest available cluster. The geometrical distance between the position of the extrapolated track and the matched TOF cluster (called residual distance) along the z direction is studied for every MRPC strip (Fig. 3.13). The strip index identifies the position in the TOF array along z (or η): the strip index 45 corresponds to z=0 ($\eta = 0$) while the strip indices 0 and 90 match the pseudorapidity η = -0.9 and η = 0.9, respectively. As expected, the residuals for the strips are distributed around zero and show no asymmetry in the two sides of the TOF detector. Deviations from zero could hint at the presence of tracking issues or a not perfect alignment between TOF and TPC detectors.

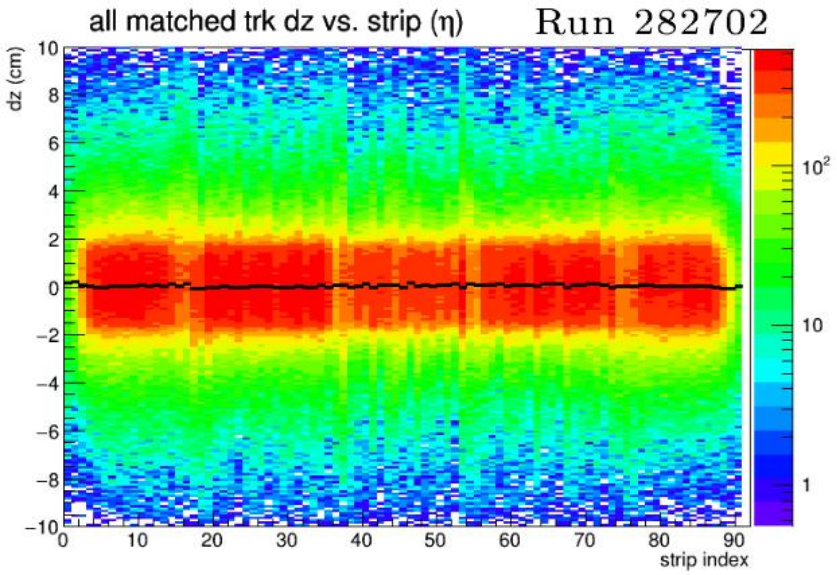

Figure 3.13: TOF matching residuals vs the strip number for a 2017 pp run at $\sqrt{s} = 13$ TeV.

Another crucial quantity that is checked by the offline QA is the TOF matching efficiency. Its trend is checked run by run and compared with the one observed in the MC anchored production (Fig. 3.14).

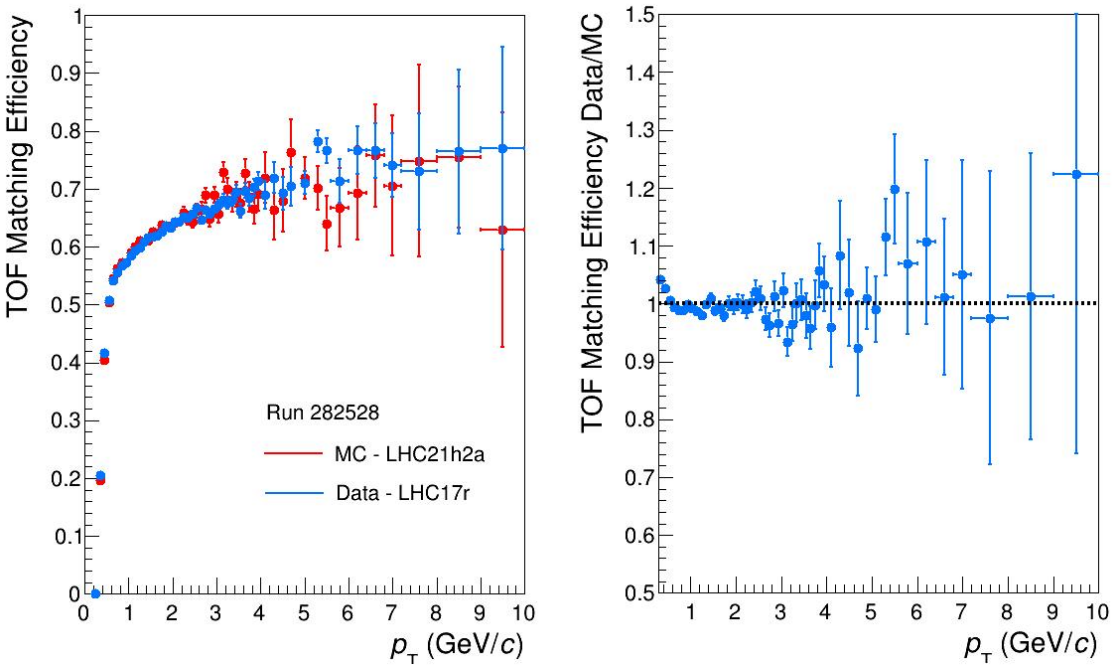

Figure 3.14: Left: TOF matching efficiency versus the track reconstructed p_T for tracks within the pseudorapidity range $|\eta| < 0.8$ in data and in the anchored MC production for a pp 2018 run at $\sqrt{s} = 13$ TeV. Right: Ratio between the TOF matching efficiency in data and in the anchored MC production.

It is defined as the ratio between the number of reconstructed tracks with a matchable hit on TOF and the total number of tracks reconstructed using the ALICE tracking detectors, i.e. ITS, TPC and TRD. Tracks with $p_{\mathrm{T}} < 0.3$ GeV/c do not reach the detector, due to the curvature of the trajectory in the magnetic field. For tracks with $p_{\mathrm{T}} > 0.3$ GeV/c, the matching efficiency rapidly increases and it reaches a more or less constant value for $p_{\mathrm{T}} > 1$ GeV/c. In order to fairly compare the matching efficiency in different runs, the efficiency must be normalized by the fraction of TOF active channels in that run. The trending plots of the average matching efficiency from tracks with $p_{\mathrm{T}} > 1$ GeV/c, for the different set of runs, before and after the normalization, are reported in Fig. 3.15. The normalization minimizes the differences between the runs. Any discrepancy from the roughly flat trend observed can be due to several issues and must be investigated by experts.

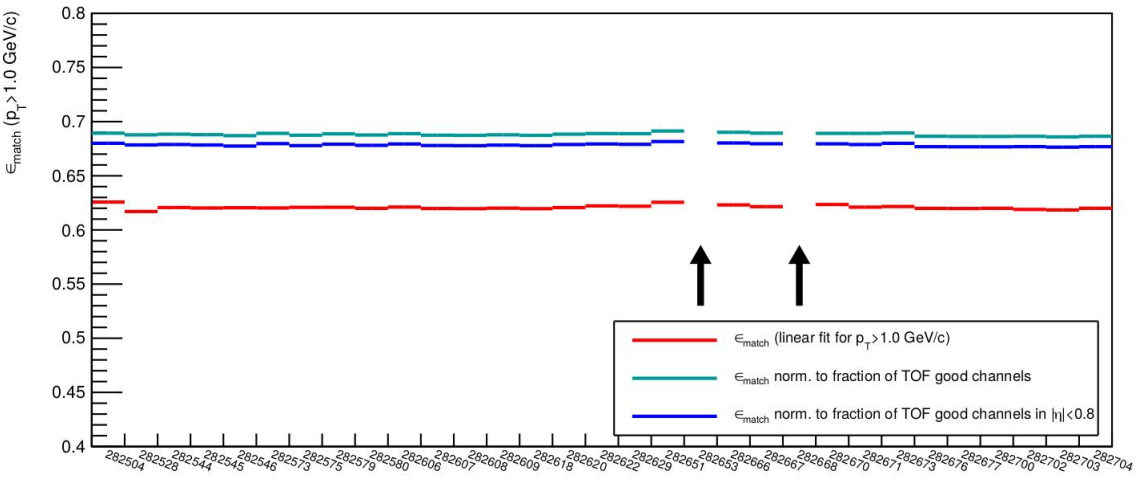

Figure 3.15: Trending distribution of the TOF matching efficiency versus the run number for a set of 2017 pp runs at $\sqrt{s} = 13$ TeV. The black arrows point to runs in which the TOF was turned off.

The PID capabilities are checked by computing the nσ distributions of pions, kaons and protons (eq. 3.11). These distributions are shown in Fig. 3.16 as a function of the track momentum. Fits with a Gaussian and a Gaussian with exponential tail are used to compute the mean and the width of the distributions.

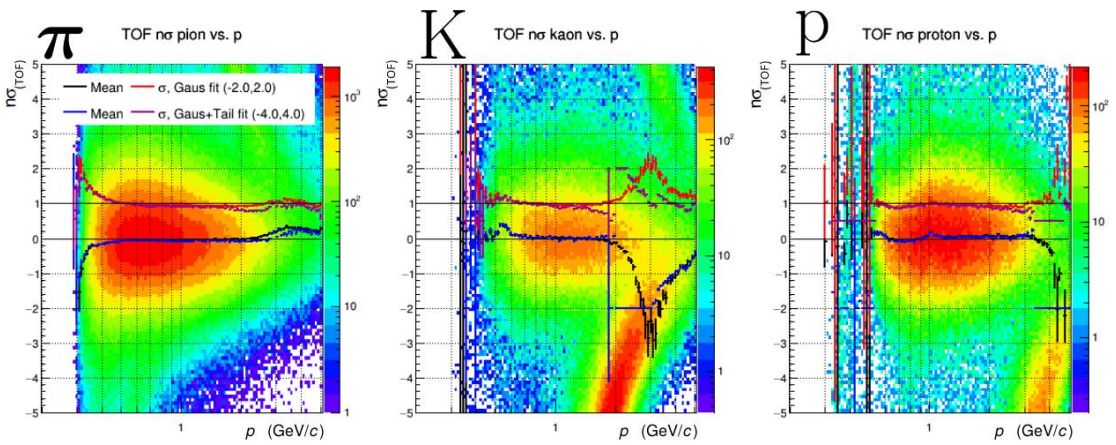

Figure 3.16: TOF nσ distribution for pions (left), kaons (middle) and protons (right). The mean and width parameters of the Gaussian (Gaussian with exponential tail) fitting function, are reported in red and black (blue and magenta) respectively.

As expected, the mean value of the distribution is aligned to 0 and the σ to 1, as long as the particle momenta are sufficiently high so that a good tracking resolution is achieved and the signals of different particle species are well separated. Discrepancies of the mean from 0 can point to detector miscalibration and must be investigated in detail by the QA expert.

3.2.5 High Momentum Particle Identification Detector

The High Momentum Particle Identification Detector (HMPID) is a Ring Imaging Cherenkov (RICH) detector mounted inside the solenoid. It has a very limited acceptance, covering the pseudorapidity range $|\eta| < 0.6$. The HMPID includes 7 independent modules (Fig. 3.17) formed by two main parts: a radiator medium, where the Cherenkov light is produced and a photon detector, consisting of a thin layer of CsI, deposited on the cathode pads of a proportional multi-wire chamber. The main purpose of this detector is to identify charged hadrons at $p_{\mathrm{T}} > 1$ GeV/c, enhancing the PID capabilities of ALICE beyond the momentum range allowed by the energy loss measurements in the TPC and by the time of flight measurements in the TOF, pushing the identification of π/K and K/p up to 3 GeV/c and 5 GeV/c, respectively.

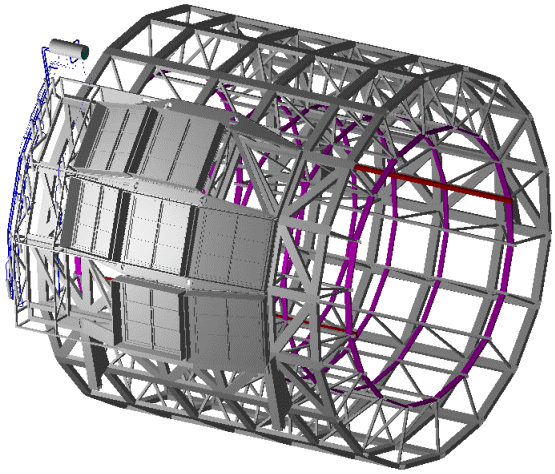

Figure 3.17: Layout of the HMPID modules on the support cradle as mounted on the ALICE space frame.

3.2.6 Photon Spectrometer

The Photon Spectrometer (PHOS) is a high-resolution electromagnetic spectrometer. It is positioned on the bottom of the ALICE set-up, at a distance 460 cm from the interaction point. The PHOS has a very limited acceptance, with a pseudorapidity and azimuth angle coverage of $|\eta| < 0.12$ and $260 < \phi < 320$, respectively. The PHOS is made by 3 independent modules, each consisting of an electromagnetic calorimeter and of the Charged Particle Veto detector (CPV) that allows for photon identification via suppression of charged-particle contamination. The PHOS is dedicated to the search of electromagnetic radiation from the hot strongly-interacting matter in nucleus-nucleus interactions at high energies, as well as for measurements of hadron spectra via their radiative decays and jet-quenching.

3.2.7 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (EMCal) is a lead-scintillator sampling calorimeter located between the TOF detector and the L3 magnet. The EMCal covers the pseudorapidity region $|\eta| < 0.7$ and the azimuth angle of $80 < \phi < 187$. The EMCal enhances the capabilities of ALICE to measure highly energetic photons, electrons,

neutral pions and jets of particles, and the correlations between them, extending the p_{T} range reached by PHOS and providing a larger acceptance. To further extend the EMCal acceptance, the DCal (Di-Jet Calorimeter), a calorimeter with the same structure of EMCal but collocated in the opposite position, was installed during the Long Shutdown I of LHC (2013-2015). It allows measurements of back-to-back correlations of jets and hadrons.

3.2.8 ALICE Cosmic Ray Detector

The ALICE Cosmic Ray Detector (ACORDE) is an array of plastic scintillator modules located on top of the L3 magnet. It includes 60 modules, each consisting of 2 scintillators placed one on top of the other in coincidence. ACORDE is used during cosmic ray data taking sessions, providing a fast trigger signal useful for the commissioning operations, like calibration and alignment of the central barrel detectors. It also allows performing physics studies with cosmic rays.

3.2.9 Muon spectrometer

The muon spectrometer is placed outside of the central barrel, on the side C, and it covers the pseudorapidity region -4 $< \eta <$ -2.5. It allows the study of the complete spectrum of heavy quarkonia (J/Ψ, Ψ', Υ, Υ', Υ", ϕ) via their decay in the $\mu^+\mu^-$ channel. The structure of the forward muon arm is summarized below.

- The conical front absorber suppresses the hadron and electron background coming from the interaction vertex. It is 4.13 m thick and it is located 90 cm away from the interaction vertex, still inside the L3 magnet. It is made of carbon and concrete, to limit the multiple scattering and the energy loss of the muons, and of tungsten, lead and stainless steel to minimize the background arising from primary particles emitted in the collision and from their showers produced in the beam pipe and in the shield itself.

- The dipole magnet, positioned at $\sim$7 m from the interaction vertex, is used

for the determination of the muon momentum. It generates a magnetic field orthogonal to the direction of the beam of 0.7 T and an integrated magnetic field along the axis of the beam of 3 Tm. The magnetic field value is defined by the requirements on the mass resolution.

- The muon filter consists in a thick iron wall of 1.2 m. Its purpose is to reject the low energy background.

- The tracking system includes 10 muon tracking chambers, arranged in 5 stations (2 chamber each): two of them are located between the absorber and the dipole magnet, one inside the magnet and the remaining two between the magnet and the muon filter. The tracking system allows for the muon tracks reconstruction with a spatial resolution better than 100 mm, and provides a measurement of the muon momentum.

- The trigger system selects muons above a given p_{T} threshold. The 4 planes of RPCs, arranged in 2 stations and positioned behind the muon filter, provide a rough estimate of the transverse momentum of each μ.

The schematic representation of the muon spectrometer is reported in Fig. 3.18.

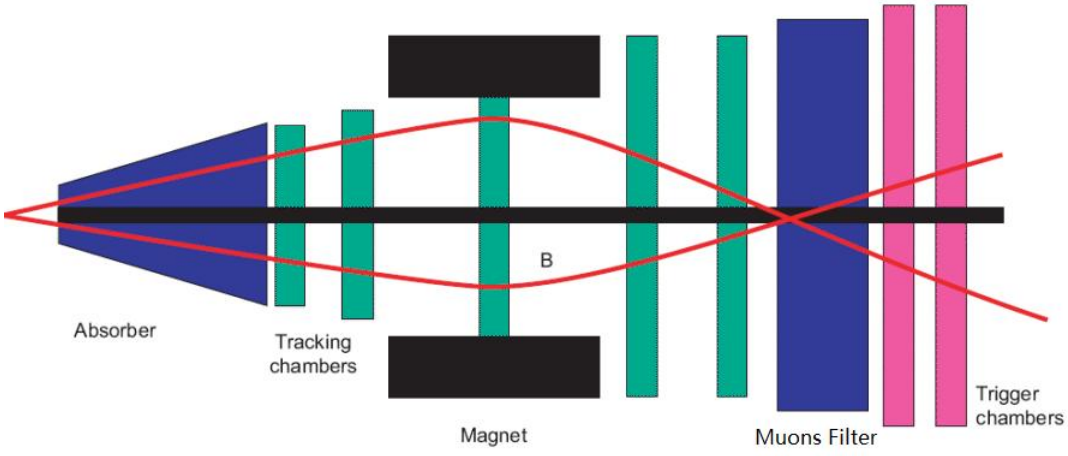

Figure 3.18: Schematic representation of the ALICE muon spectrometer: it consists of an absorber to filter the background, a set of tracking chambers, before, inside and after the magnet, and a set of trigger chambers.

3.2.10 Forward rapidity detectors

Outside of the ALICE barrel, in the high pseudorapidity region, several smaller detectors are installed. Some of the tasks that they perform are:

- Global event characterisation (centrality, multiplicity, time measurements).
- Provide triggers.
- Extend the charged particle measurements at forward rapidity.

These detectors are briefly described below.

Zero Degree Calorimeter

The Zero Degree Calorimeter (ZDC) detectors are used to measure the centrality in AA collision. The ZDC is indeed able to detect the number of nucleons that does not interact in the collision (spectators) measuring the energy that they deposit in the detector, which decreases with increasing centrality. The spectators keep their trajectory in the forward direction, emerging from the collisions at 0°, with a transverse momentum close to 0. Once the number of spectators is known, the number of nucleons that participates to the collision can be simply estimated by subtraction. There are two ZDC, placed at 115 m away from the IP on both sides, along the beam line. Each ZDC includes two hadronic calorimeters, the ZN that detects the spectator neutrons and the ZP for the spectator protons. The ZDC detector is completed by two forward electromagnetic calorimeters (ZEM) placed 7 m away from the IP. The ZEM calorimeters help to distinguish between central and ultra peripheral (UP) ion-ion collisions. Indeed, the behaviour of the energy deposited by nuclear fragments, produced in UP collisions, is no more monotonic, since some nuclear fragments can be deviated by the LHC magnets outside the acceptance of the ZDC leading to a energy signature similar to that of central collisions. This ambiguity can be solved by correlating the energy collected by the ZDC with the energy deposited in the ZEM, that measure the energy deposited by π^0 decays and photons produced at forward rapidity [60].

V0 Detector

The V0 is a small angle detector composed by two arrays of scintillator counters, V0-A and V0-C, placed at the opposite sides at 330 cm and 90 cm from the IP and covering the pseudorapidity range $2.8 < \eta < 5.1$ and $-3.7 < \eta < -1.7$, respectively. Scintillation light produced by particles traversing the detector's volume is transported to photomultiplier tubes. The signal produced is proportional to the number of particles hitting the detector, and there is a monotone dependence between the number of impinging particles on the V0 arrays and the number of primary particles emitted. For this reason, the V0 system is able to provide a measurement of the charged-particle multiplicity of the event, allowing also a centrality determination via interpolation with the Glauber model [61]. Moreover, the V0 detector is used to reject beam-induced background, due to the particles of the beam interacting with residual particles in the beam-pipe.

T0 Detector

The T0 detector includes two arrays of Cherenkov counters, T0A and T0C, placed at the opposite sides of the IP, at 370 cm and 70 cm. They cover the pseudorapidity intervals $4.61 < \eta < 4.92$ and $-3.28 < \eta < -2.97$, respectively. Thanks to its very good time resolution, $\sim$50 ps in pp collisions and reaches $\sim$25 ps at higher multiplicities, the T0's main purpose is to determinate the start time for the TOF detector.

Photon Multiplicity Detector

The Photon Multiplicity Detector (PMD) is a particle shower detector that measures the spatial distribution and the multiplicity of the forward rapidity photons, in the pseudorapidity interval $2.3 < \eta < 3.7$. The PMD is made of two plane of gas proportional counters preceded by two converter plates. The first plane in front of the converter is used to reject charged particles. Photons passing through a converter initiate an electromagnetic shower and produce large signals on several cells of the sensitive volume of the detector.

Forward Multiplicity Detector

The Forward Multiplicity Detector (FMD) measures the charged-particle multiplicity in the pseudorapidity range -3.4 $< \eta <$ -1.7 and 1.7 $< \eta <$ 5.0. Combined with the ITS, it provides a full pseudorapidity coverage. It is made by 5 ring silicon strip detectors, arranged in 3 sub-detectors located at three different position along the beam pipe: FMD1 is located at 320 cm from the IP on the A side while FMD2 and FMD3 are symmetrically placed 75 cm away from the IP.

3.3 The ALICE offline framework

The framework employed by the ALICE experiment for simulations, reconstruction and analysis of the data collected is AliRoot, a scientific software based on ROOT toolkit [62]. It is mainly written in C++, but it is integrated with other languages such as Python and R. The simulation of physics events is extremely important for any analysis, in particular for evaluating the efficiency of the algorithms used to reconstruct and analyse the data. Simulated events are produced with Monte Carlo generators, such as `PYTHIA` [35] for pp collisions, and HIJING [38] for proton–nucleus and nucleus–nucleus collisions. The generated particles are then propagated through the experimental apparatus, where they lose energy, decay and interact with the detectors. AliRoot includes a detailed description of the detector geometry, with the simulation and the reconstruction performed independently for each subdetector system. To simulate the detector response to the passage of the particles, AliRoot makes use of different transport packages like `GEANT3` [63], `GEANT4` [64], and `FLUKA` [65]. In these packages the detector material budget is simulated in detail, including support structures and the beam pipe. The reconstructed data, both real and simulated, are stored in files, called ESD (Event Summary Data), that contain all the physical information about the event and the tracks needed for the analysis, and related to the quality of the reconstruction for each sub-detector. In order to reduce the size of the files and the computational time needed for the analysis, the AOD (Analysis

Object Data) files, which contain only the information needed for the analysis, are produced. The huge and unprecedented amount of collected data at LHC require processing and storage space that cannot be concentrated in a single computing center. For this reason, the Worldwide LHC Computing Grid (WLCG) was created. It is a geographically distributed computing system made by the facilities of the institutes and universities participating in the experiment. The ALICE users can access the grid via the ALICE Environment (AliEn) [66], provided by AliRoot. In particular, the AliEn user interface allows accessing the data available everywhere on the grid, send analysis tasks and simulations, and monitor their status.

3.3.1 Tracking reconstruction algorithm

In ALICE the track-reconstruction procedure is performed offline. A charged particle passing through the detectors leaves a number of discrete hits in the corresponding positions in the space. First of all, the raw data of each detector are converted in clusters, defined as groups of adjacent detector cells firing. Each cluster corresponds to a hit (energy deposition) produced by a crossing particle and it is characterised by several quantities, such as position, signal amplitude and signal time, with their associated errors. Joining the pairs of clusters in the two innermost layers of the ITS, the SPD, the tracklets are obtained. The tracklets are crucial for the reconstruction because they are employed for a first evaluation of the primary vertex position, given by the space point that minimises the distance among the tracklets. The clusters are then passed to the track reconstruction algorithms, based on the Kalman filter [67] and divided in three steps (inward–outward–inward). The track finding algorithm starts from the reconstruction in the TPC, building the track seeds from the clusters in the two outermost pad rows of the TPC and the primary vertex estimated with the SPD tracklets. Then, the seeds are propagated inward, choosing, at each step, the cluster closest to the track, found by Kalman Filter algorithm according to a cut on the proximity. The procedure is repeated until the inner radius of the TPC is reached. The tracks reconstructed in the TPC are matched to the hits in the

outermost SSD layer and become the seed for the track finding in the ITS. Similarly to the previous step, the seed is propagated inward attaching the closest clusters within a proximity cut to the seed. Therefore, for each TPC track, a corresponding tree of track hypotheses in the ITS is produced. The track candidates are then selected according to their χ^2. In order to increase the efficiency for tracks with low transverse momentum ($p_{\mathrm{T}} < 200$ MeV/c) and to reconstruct tracks of particles traversing dead zones of the TPC or which decay before entering in the TPC, the hits in the ITS not attached to tracks propagated from the TPC are used to perform an ITS stand-alone reconstruction. In the second iteration, the tracks obtained with clusters in both the ITS and TPC detectors are prolonged in the outward direction to match the reconstructed points in the outer detectors, i.e. TRD, TOF, EMCal, DCal, PHOS and HMPID. Finally, the tracks are re-fitted inward using all the previously found clusters and propagated to their distance of closest approach to the SPD vertex. After the full tracking procedure, if at least two tracks are reconstructed, the position of the primary vertex is recomputed using global tracks (tracks with both ITS and TPC reconstruction). In order to remove contributions from secondary tracks originating from interactions with detector material or particle decays, a cut on the distance of closest approach between the tracks and the primary vertex is applied.

Chapter 4

Λ_c analysis via its $p\,K^0_S$ decay reconstruction

In this chapter, the measurement of the Λ_c^+ baryons and their antiparticles production versus the event charged-particle multiplicity, through the reconstruction of the decay channel $\Lambda_c^+ \rightarrow pK_s^0 \rightarrow p\pi^+\pi^-$ (BR $= 1.10 \pm 0.06\%$) [4], is presented. The measurement was performed at mid-rapidity ($|y| < 0.5$) using the data collected between 2016 and 2018 with the ALICE detector in pp collisions at $\sqrt{s} = 13$ TeV. The aim of the analysis is to extract the Λ_c^+ production yields and to estimate the baryon-over-meson ratio Λ_c^+/D^0 as a function of the transverse momentum p_T and the event charged-particle multiplicity.

The chapter is organized as follows. The data sample and the multiplicity corrections are first introduced. Successively, the analysis strategy, with the Λ_c candidate reconstruction, the innovative machine learning algorithm employed to reject the combinatorial background, the raw yield extraction and the efficiency times acceptance and feed-down corrections, is carefully described. The systematic uncertainties computation is then presented in details. In the last part of the chapter, the achieved results are finally shown. The measured Λ_c^+ corrected yield and Λ_c^+/D^0 ratio in pp collisions at $\sqrt{s} = 13$ TeV versus the charged particle multiplicity and their comparison with the theoretical model are discussed in detail.

4.1 Data samples and event selection

The analysis presented in this thesis was performed on the pp data sample at the collision energy in the centre-of-mass system of $\sqrt{s} = 13$ TeV collected by the ALICE detector during the RUN2 data taking period of the LHC. Two trigger selections were employed:

- The minimum-bias (MB) trigger requires signals in both V0A and V0C in coincidence with the proton bunch arrival time.
- To enrich the data sample size in the highest multiplicity regions, high-multiplicity trigger, based on a requirement of minimum number of hits in the outer SPD (HMSPD) layer, was used (available only for the 2018 data taking periods). In particular, the chosen threshold is of 77 chips fired. The HMSPD trigger will collect more HM events than the MB one due to the lower prescale[1].

Offline selection criteria were applied in order to remove the event coming from the interaction of beam particles with the beam pipe materials and from beam-gas interactions. Only events with a reconstructed primary vertex within 10 cm from the nominal IP were considered ($\mid z_{vtx} \mid < 10$ cm) to ensure a uniform acceptance. To reduce the superposition of more than one collision within the same colliding bunches (pile-up), events with multiple reconstructed primary vertices were rejected. In particular, an event is removed from the analysed data sample if a second interaction vertex with at least 5 associated tracklets is found. In addition, events were required to have at least one charged particle within the pseudorapidity region $\mid \eta \mid < 1$ (INEL > 0). This class of events minimises diffractive corrections and has a high trigger efficiency. It corresponds to about 75% of the total inelastic cross section [68, 69]. After the aforementioned selections, the data sample consist in $\sim$1.7 $\times$ 10^9 MB events and in $\sim$0.3 billion of HMSPD triggered events, corresponding,

[1] A trigger prescale factor is an integer $N > 1$ such that only one in every N events fulfilling the trigger requirements was actually sent to the output data stream. Prescaled triggers are used to perform measurements that do not require the full data sample, in cases where the unprescaled path would have prohibitively high rate.

respectively, to an integrated luminosity of 32 nb^{-1} and 0.8 pb^{-1}. Details on the analysed data are summarised in Tab 4.1.

Year	Subperiods	Trigger	N. events
LHC16	d,e,g,h,j,o,p,k,l	MB	$424{\cdot}10^6$
LHC17	e,f,h,i,j,k,l,m,o,r	MB	$580{\cdot}10^6$
LHC18	b,d,e,f,g,h,i,j,k,l,m,n,o,p	MB	$706{\cdot}10^6$
LHC18	b,f,h,j,k,l,m,o,p	HMSPD	$319{\cdot}10^6$

Table 4.1: Summary of the data samples used in the analysis.

4.2 Multiplicity definition and corrections

The charged-particle multiplicity was estimated in the mid-rapidity region, i.e. in the pseudorapidity range $|\eta| < 1$, employing the number of tracklets in the SPD detector, $N_{\text{tracklets}}$. A SPD tracklet is obtained by joining a pair of hits in the SPD layers and aligned with the reconstructed primary vertex. The measured raw tracklet distributions in 2016, 2017 and 2018 are shown in Fig. 4.1.

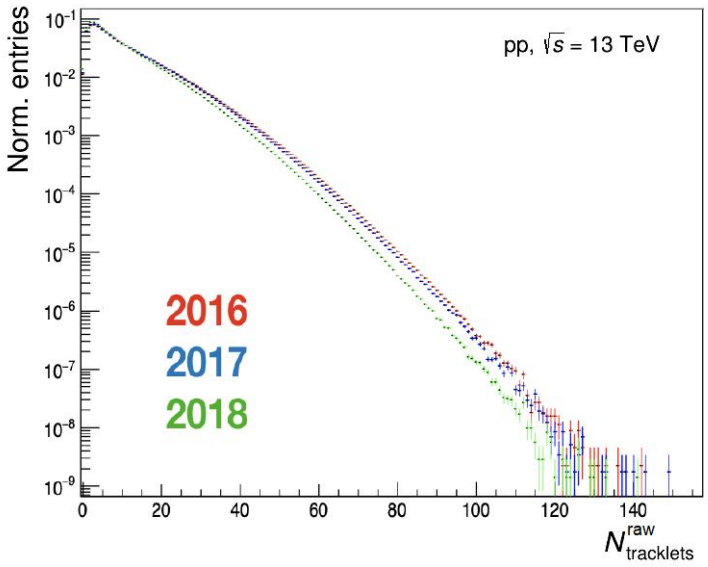

Figure 4.1: Measured raw tracklet distributions in 2016, 2017 and 2018 for minimum bias trigger.

The analysis was performed in four multiplicity intervals, between 1 and 99, and in the multiplicity integrated bin. The chosen bins are reported in Tab. 4.2 along with the trigger employed to collect the data. There are several corrections applied on the final number of tracklets, discussed in detail below.

$N_{\text{tracklets}}$	Trigger
INEL>0 ($[1, \infty]$)	MB
$[1, 9]$	MB
$[10, 29]$	MB
$[30, 59]$	MB
$[60, 99]$	HMSPD

Table 4.2: SPD tracklet multiplicity bins with the trigger used to collect the data in the given interval.

Selected candidates are further split into six bins of candidate transverse momentum p_{T}, which are summarised in Tab. 4.3. The p_{T} and multiplicity binning were chosen to ensure a large enough data sample size that allows for the signal extraction in all the multiplicity ranges.

Candidate p_{T} (GeV/c)
$[1, 2)$
$[2, 4)$
$[4, 6)$
$[6, 8)$
$[8, 12)$
$[12, 24)$

Table 4.3: Analysed bins of candidate p_{T}.

4.2.1 z-vertex correction

The first correction applied to the measured number of tracklets $N^{\text{raw}}_{\text{tracklets}}$ is meant to account for the z-vertex dependence of their profiles, which was observed to be significantly modified in the different data taking period. As the geometry and evolution of the collision system do not depend on its location within the detector, the distribution of multiplicity should ideally be independent of the position of the primary vertex. However, due to hardware detector effects (such as the presence of dead SPD pixels), the actual measured multiplicity does not follow this expectation. The measured tracklet profiles, i.e. the mean number of tracklets at each vertex position, $\langle N_{\text{period}}(z)\rangle$, for all the data taking periods in 2016, 2017 and 2018 are shown in Fig. 4.2.

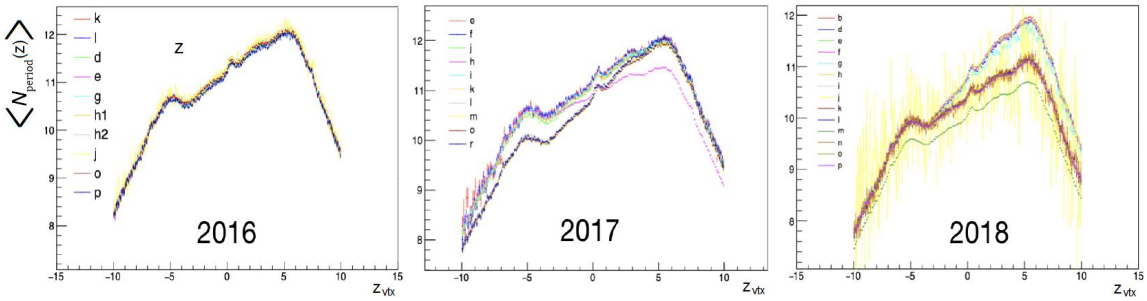

Figure 4.2: Measured raw tracklet profiles along the z vertex position, for each period of data taking during 2016, 2017, 2018 years, for minimum bias trigger.

A drop of the SPD acceptance is observed with the position of the vertex and it is even more pronounced for vertices with $z_{vtx} < 0$. Furthermore, as the data sample was taken over the course of multiple runs, the SPD performance declines over the time (aging effects). A reduction in the SPD performance over the time is clearly visible in Fig. 4.1. The distributions must therefore be corrected, in order to remove any possible time or position dependences that may affect the true multiplicity distribution. The raw profiles shown in Fig. 4.2 were used on event by event basis to correct the measured multiplicities in each period. The period with the highest acceptance was taken as a reference to correct the other ones. The maximum value in the average number of tracklets as a function of z_{vtx} was found in the LHC16

periods, and it is 12.25 ($\langle N_{\rm ref}\rangle$) at $z_{\rm vtx} = 5.55$ cm. The corrected mean number of tracklets for each position z, $N^{\rm corr}_{\rm tracklets}(z)$, was then obtained with equation (4.1):

$$N^{\rm corr}_{\rm tracklets}(z) = \frac{\langle N_{\rm ref}\rangle}{\langle N_{\rm period}(z)\rangle} N^{\rm raw}_{\rm tracklets}(z) \tag{4.1}$$

The mean $N_{\rm tracklets}$ values for all the three data periods without and with the z-vertex correction are reported in Tab. 4.4.

Period	$\langle N^{\rm raw}_{\rm tracklets}\rangle$	$\langle N^{\rm corr}_{\rm tracklets}\rangle$
LHC16	11.16	12.25
LHC17	10.91	12.25
LHC18	9.95	12.25

Table 4.4: Comparison between average multiplicities before $\langle N^{\rm raw}_{\rm tracklets}\rangle$ and after $\langle N^{\rm corr}_{\rm tracklets}\rangle$ the correction, for the minimum bias triggered sample.

For the HMSPD triggered sample, the z-vertex correction was done using the same profiles as for the MB analyses, after having checked that the average $N_{\rm tracklets}$ distribution vs the $z_{\rm vtx}$ was compatible in the two different samples.

4.2.2 Removal of daughter tracklets

A second correction is required because the charged-particle multiplicity is estimated in the same pseudorapidity region as the reconstructed charm hadrons. In this case, indeed, also the Λ_c-baryon decay products are included in the multiplicity evaluation. To reduce these effects of auto-correlation, the Λ_c decay tracks (also called prongs in the following) were excluded from the $N_{\rm traklets}$ estimation. The tracklets coming from the Λ_c baryons were subtracted from the number of $N_{\rm tracklets}$ if their prongs have hits on the first and second ITS layers. For the particular decay under study, i.e. $\Lambda_c^+ \to {\rm pK}^0_{\rm s} \to {\rm p}\pi^+\pi^-$, only the proton track crosses the SPD and it was then removed from the counts.

4.2.3 Trigger correction

Not all the events that fulfil the INEL > 0 requirement are selected by the MB triggered events with at least one charged particle within the pseudorapidity region $|\eta| < 1$ (MB$_{>0}$). For this reason, a trigger correction, ϵ^{INEL}, needs to be applied. This factor, usually called trigger efficiency, is estimated using `PYTHIA` 8 Monash 2013 tune simulations [35] as:

$$\epsilon^{\mathrm{INEL}} = \frac{N_{\mathrm{INEL>0}}}{N_{\mathrm{MB_{>0}}}} \tag{4.2}$$

where $N_{\mathrm{INEL>0}}$ and $N_{\mathrm{MB_{>0}}}$ are, respectively, the events that fulfil the INEL > 0 and the MB$_{>0}$ trigger conditions. The trigger efficiency does not depend on the particle under study. Is is estimated year by year in each multiplicity bin and the final values, reported in Tab. 4.5, are the average of the three available years [68].

Concerning the HMSPD triggered data sample, the trigger is not fully efficient in the [60-100] $N_{\mathrm{tracklets}}$ interval. As discussed in section 4.1, the trigger is based on the hits in the outer SPD layer while the multiplicity class is estimated counting the number of tracklets. The first quantity is strongly correlated with the SPD tracklet multiplicity but there is not a one-to-one correspondence. To quantify the $N_{\mathrm{tracklets}}$ that are not detected by the HMSPD trigger, a comparison with the MB triggered events has been employed. The $N_{\mathrm{tracklets}}$ distributions in MB and HMSPD triggered events are shown in the left panel of Fig. 4.3. The ratio between the MB $N_{\mathrm{tracklets}}$ distribution and the HMSPD one, called trigger turn-on-curve, is reported on the right panel of Fig. 4.3 and it does not show a sharp increase. The turn-on-curve goes smoothly from 0 to a constant value in the $N_{\mathrm{tracklets}}$ range 40 to 70, indicating that the HMSPD trigger is not fully efficient in the $60 < N_{\mathrm{tracklets}} < 70$ interval. As expected, the turn-on-curve does not saturate at 1, due to the effect of the different prescales, which are lower for the HMSPD triggered events. Accordingly, the high multiplicity trigger provides an enhancement of the data sample size at high multiplicity range by a factor $\sim$100, as it can be seen by the value of the ratio in the plateau of the

turn-on-curve.

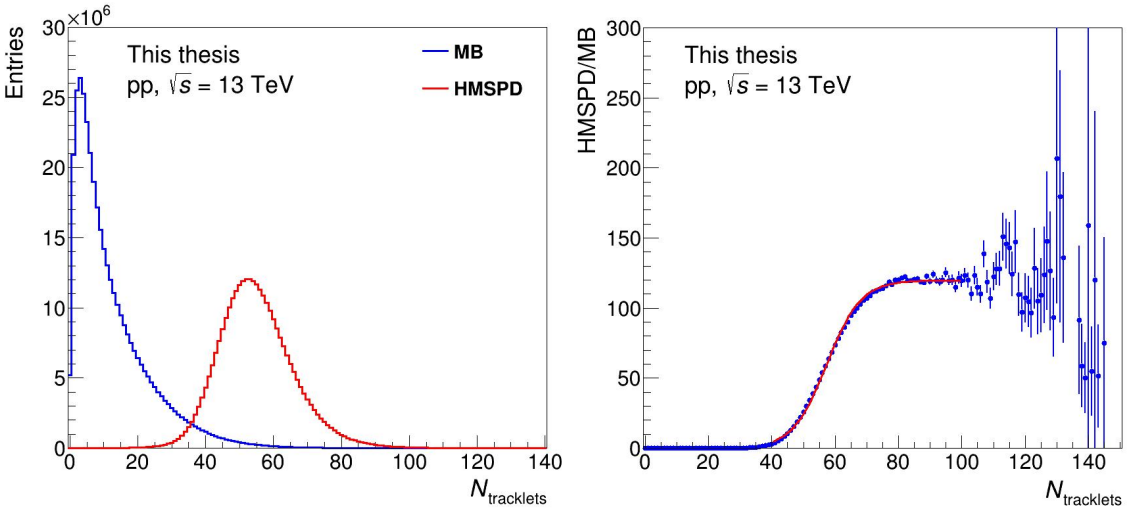

Figure 4.3: Left: $N_{\text{tracklets}}$ distributions for MB and HMSPD triggered sample. Right: HMSPD over MB $N_{\text{tracklets}}$ distribution ratio.

To account for the observed HMSPD trigger inefficiency, a correction was applied with a data-driven reweighting procedure. The ratio of the $N_{\text{tracklets}}$ distribution in the two trigger samples was normalized to 1 in the high multiplicity region by dividing the distribution by the observed ratio at $N_{\text{tracklets}} = 70$. Different ways of normalising were studied. A detailed discussion can be found in section 4.8.6. All of these normalisation methods give the same corrected results within a few per mille and the variation has been assumed as systematic uncertainty. For each value of $N_{\text{tracklets}}$, having defined R as the ratio HMSPD/MB for that particular value of $N_{\text{tracklets}}$, the weight 1/R was used when filling the invariant mass histograms. The number of events was also corrected for the HMSPD trigger efficiency with the same technique, i.e. by applying a weight to the number of measured events for each value of $N_{\text{tracklets}}$. The normalized turn-on-curve and the weights are shown, respectively, on the left and right panel of Fig. 4.4.

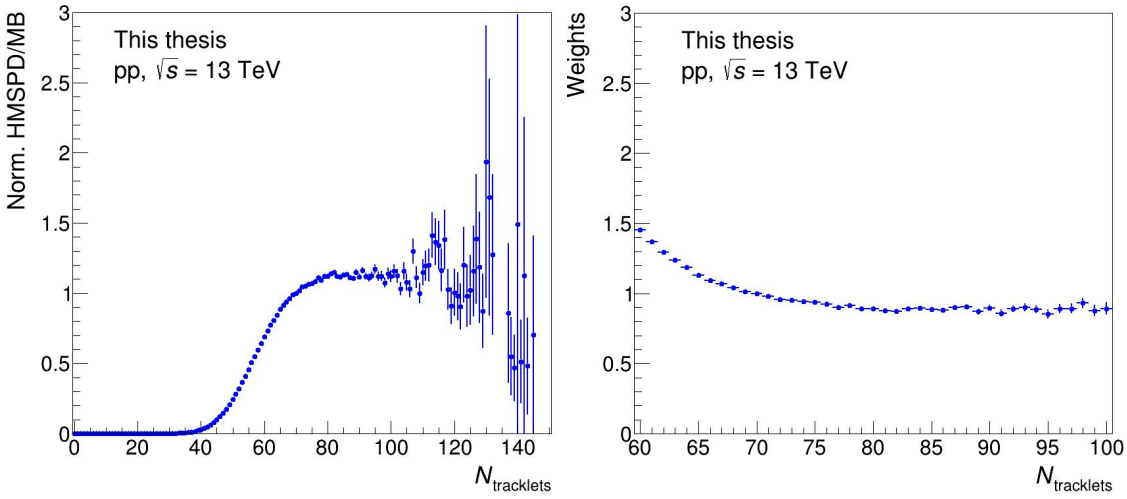

Figure 4.4: Left: Normalised ratio of the $N_{\text{tracklets}}$ distributions for MB and HM triggers. Right: The extracted weights used to correct for the HM trigger efficiency.

The used trigger, the trigger efficiency, and if a trigger correction is needed due to the trigger turn-on-curve or not are reported, for each multiplicity class, in Tab. 4.5.

$N_{\text{tracklets}}$	Trigger	Trigger efficiency	Trigger correction
INEL>0 ($[1, \infty]$)	MB	0.92 ± 0.003	No
$[1, 9]$	MB	0.862 ± 0.015	No
$[10, 29]$	MB	0.997 ± 0.002	No
$[30, 59]$	MB	1.0 ± 0.0	No
$[60, 99]$	HMSPD	1.0 ± 0.0	Yes

Table 4.5: SPD tracklet multiplicity bins, the used trigger, the trigger efficiency, and if a trigger correction is needed due to the trigger turn-on-curve.

4.2.4 Conversion of SPD tracklets to $\mathrm{d}N_{\text{ch}}/\mathrm{d}\eta$

The number of SPD tracklets is not a physical primary observable. It is simply the counts of the number of tracks seen by the detector, so it is detector-dependent and may change over time when the SPD ages. The actual physical quantity of interest is N_{ch}, the number of primary charged particles produced in the interaction [70]. Then, a strategy to convert the $N_{\text{tracklets}}$ values to a value of N_{ch} was developed.

Minimum-bias Monte Carlo simulations were employed to perform the conversion. The number of physical primary charged particles $N_{\rm ch}$ and the $N_{\rm tracklets}$ distribution corrected for the $z_{\rm vtx}$ profile dependence (discussed in 4.2.1), generated in the simulation within $|\eta| < 1$, were compared and shown via a two dimensional plot (Fig 4.5).

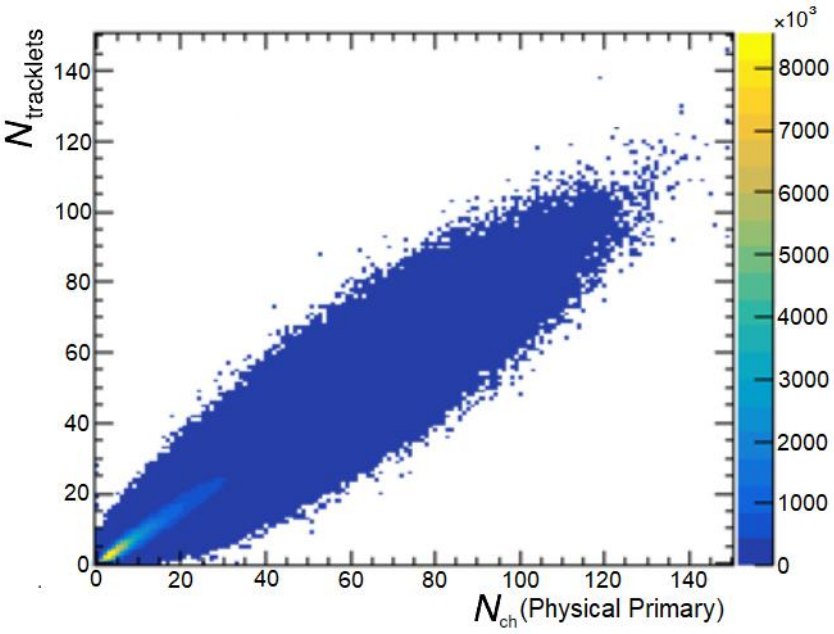

Figure 4.5: The correlation between the physical primary $N_{\rm ch}$ and the detector-dependent $N_{\rm tracklets}$ within $|\eta| < 1$ with the period-dependent $z_{\rm vtx}$ correction.

A clear linear correlation is visible between $N_{\rm ch}$ and $N_{\rm tracklets}$. A sharp cut on the reconstructed multiplicity does not correspond to a sharp cut on the true multiplicity estimator. Therefore, the mean and RMS of the $\mathrm{d}N_{\rm ch}/\mathrm{d}\eta$ distributions for each $N_{\rm tracklets}$ interval were estimated plotting the $N_{\rm ch}$ distributions in bins of $N_{\rm tracklets}$ as shown in Fig 4.6. This figure shows even better the clear overlap in $N_{\rm ch}$ for the different $N_{\rm tracklets}$ bins. The mean was simply taken from these distributions, merging all the analysed periods. Since $N_{\rm ch}$ is estimated considering the particles within $|\eta| < 1$, the computed values were divided by a factor 2 to get the mean number of primary charged particles produced per unity of pseudorapidity $\langle \mathrm{d}N_{\rm ch}/\mathrm{d}\eta \rangle$.

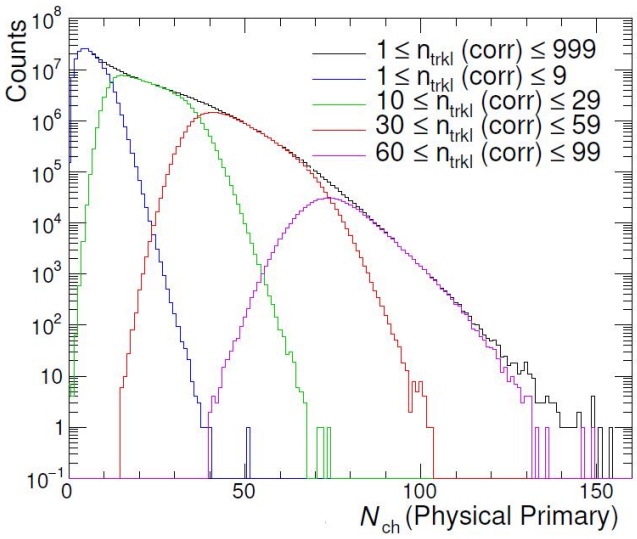

Figure 4.6: Distributions of the physical primary N_{ch} for the different used $N_{\mathrm{tracklets}}$ multiplicity intervals, applying the period-dependent z_{vtx} correction.

To estimate the uncertainty on these values, the procedure was repeated for each period separately. The quoted uncertainty is the RMS of the distributions for each period, reported in Fig 4.7.

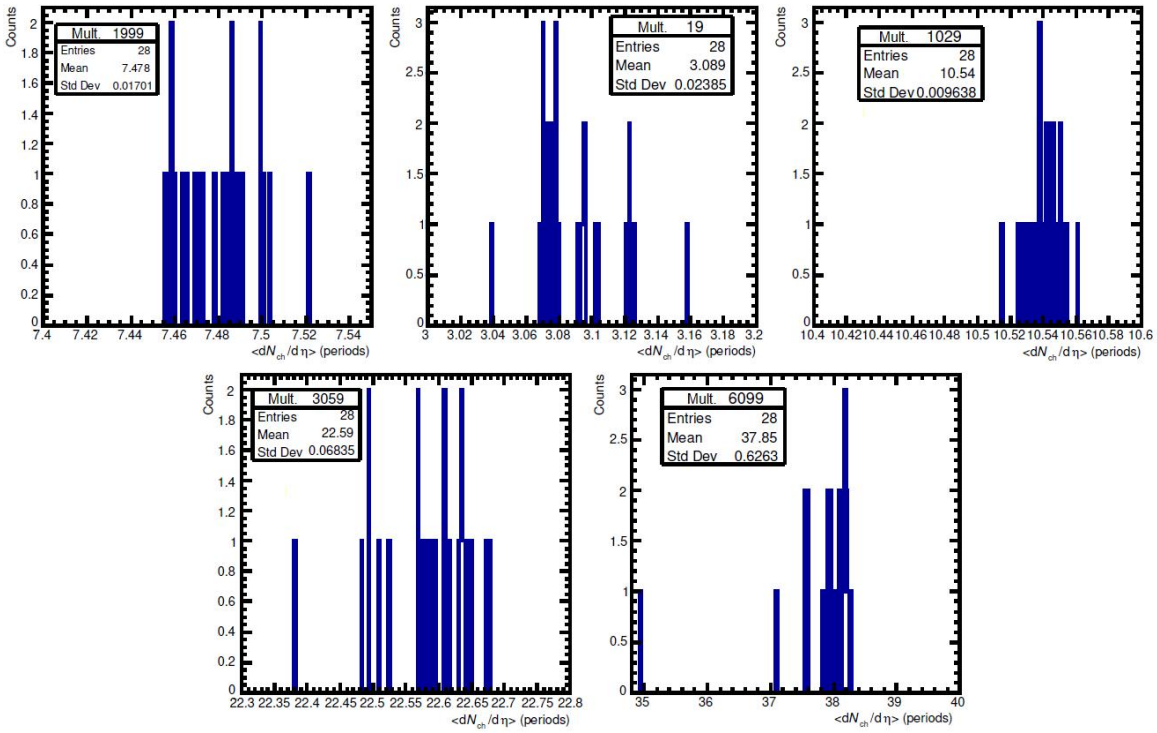

Figure 4.7: Extracted $\mathrm{d}N_{\mathrm{ch}}/\mathrm{d}\eta$ values for each period separately for the different $N_{\mathrm{tracklets}}$ intervals ([1-9999], [1-9], [10-29], [30-59], and [60-99]). The RMS values of these distributions are quoted as uncertainty.

The conversion between the $N_{\text{tracklets}}$ intervals to the $\mathrm{d}N_{\text{ch}}/\mathrm{d}\eta$ mean values and the related uncertainties are summarized in Tab. 4.6.

$N_{\text{tracklets}}$	$\mathrm{d}N_{\text{charged}}/\mathrm{d}\eta$	
Interval	Mean	RMS
$[1-9999]$	7.48 ± 0.02	6.49 ± 0.00
$[1-9]$	3.10 ± 0.02	1.64 ± 0.02
$[10-29]$	10.54 ± 0.01	3.77 ± 0.04
$[30-59]$	22.59 ± 0.07	4.51 ± 0.06
$[60-99]$	37.9 ± 0.6	4.51 ± 0.12

Table 4.6: Conversion of SPD $N_{\text{tracklets}}$ intervals to the physical primary N_{ch} observables for $|\eta| < 1$. For the bin [1-9999] the values estimated by the Luminosity, Multiplicity, Rivet and Monte Carlo generators ALICE working group (PWG-MM) is quoted.

4.3 Candidate reconstruction

The Λ_c^+ baryons cannot be directly revealed because of their mean proper decay length around 60 μm that prevents them to reach the detector. Hence, they are measured via the exclusive reconstruction of their hadronic decays. In particular, the measurement of the $\Lambda_c^+ \to \mathrm{pK}_s^0 \to \mathrm{p}\pi^+\pi^-$ decay chain relied on reconstructing the V-shaped decay of the K_S^0 meson into two pions with opposite charge, which was then combined with a proton-candidate track (bachelor). Each bachelor-V^0 candidate has therefore three prongs in total. A sketch of the decay chain is reported in Fig. 4.9.

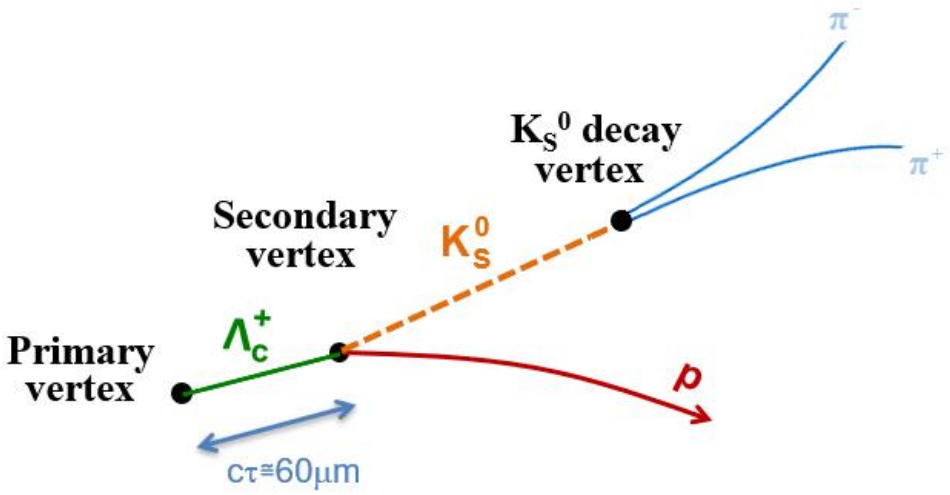

Figure 4.8: $\Lambda_c^+ \to \mathrm{pK}_s^0 \to \mathrm{p}\pi^+\pi^-$ decay chain.

The invariant mass of the Λ_c^+-baryon candidates is computed using the energy and the momentum of the decay tracks:

$$M_{\rm inv}(\Lambda_c^+) = \sqrt{(\mathrm{E_p} + \mathrm{E}_{\pi^+} + \mathrm{E}_{\pi^-})^2 - (\mathbf{p}_{\rm p} + \mathbf{p}_{\pi^+} + \mathbf{p}_{\pi^-})^2} \tag{4.3}$$

The large combinatorial background obtained with this combination procedure is then reduced applying quality selection on the single tracks and on the topological and particle identification properties of the bachelor and V^0 candidates.

4.3.1 Track quality selection

The daughter tracks were selected according to the following quality selection criteria:

- ITS refit[2] is required for bachelor tracks.
- both ITS and TPC refit are required for V^0 daughter tracks.
- at least 70 (out of a maximum of 159) associated crossed rows in the TPC.
- ratio of crossed rows (total number of hit TPC pad rows) over findable clusters (pad rows which, based on the geometry of the track, are possible clusters) in the TPC larger than 0.8.
- $\chi^2/ndf < 2$ of the momentum fit in the TPC (where ndf is the number of degrees of freedom involved in the tracking procedure[3]).
- for the bachelor tracks, at least two (out of six) hits in the ITS, out of which at least one has to be in either of the two SPD layers.

In addition, the tracks are selected in the kinematic range:

- $|\,\eta\,| < 0.8$ and $p_{\rm T} > 0.3$ GeV/c for bachelor tracks.
- $|\,\eta\,| < 1.1$ and $p_{\rm T} > 0.1$ GeV/c for $\mathrm{K}^0_{\rm S}$ daughter tracks.

[2] The tracks are required to successfully fit the detector hits.

[3] In this case *ndf* is the number of clusters available for the reconstruction. Since clusters can be identified by two coordinates, the *ndf* correspond to twice the number of clusters associated to a track.

The cut on the minimum p_T were applied because the TPC cannot reconstruct with high precision the low-momentum tracks (< 0.1 GeV/c). As a consequence of these track selection criteria, the detector acceptance for Λ_c baryons varies as a function of rapidity, falling steeply to zero for $| y | > 0.5$ at low p_T and for $| y | > 0.8$ at $p_T > 5$ GeV/c. For this reason, a fiducial acceptance selection was applied on the rapidity of the candidates, $| y | < y_{fid}(p_T)$, where the factor $y_{fid}(p_T)$ was defined as a second-order polynomial function, increasing from 0.5 to 0.8 in the transverse momentum range $0 < p_T < 5$ GeV/c, and a constant term, $y_{fid} = 0.8$, for $p_T \geq 5$ GeV/c:

$$y_{fid}(p_T) = \begin{cases} 0.5 + \dfrac{1.9}{15} p_T - \dfrac{0.2}{15} p_T^2 & \text{if } p_T < 5 \text{ GeV}/c \\ 0.8 & \text{if } p_T \geq 5 \text{ GeV}/c \end{cases} \tag{4.4}$$

4.3.2 Candidate selection

Further topological and particle identification (PID) selections on the candidates were exploited to reduce the combinatorial background. The short Λ_c decay length (60μm) combined with the particular decay topology of the $\Lambda_c^+ \to pK_S^0$ channel do not allow the separation of the Λ_c decay vertex from the interaction vertex with the current ITS resolution. So, no topological variables on the Λ_c candidates can be used to reduce the combinatorial background. The selection strategy is based only on kinematical and PID criteria on the daughter tracks. The selections are applied in two steps. First, loose prefilter cuts were applied, rejecting mainly background in regions where there is minimal signal. In a second steps, a binary classification (signal, background) was applied via a machine learning algorithm.

Prefilter selection

The prefilter selections are aimed to reduce the combinatorial background but keeping the signal efficiency as high as possible. The prefilter cuts applied are summarized in Tab. 4.7.

The particular V-shaped decay topology of the K^0_S, represented in Fig. 4.9, allows using the K^0_S topological variables to reduce the combinatorial background. An upper cut is applied on the V^0 daughters distance of closest approach (DCA) and on the V^0 impact parameter with respect to the primary vertex. The angle between the line connecting the primary and secondary vertices and the direction of the reconstructed V^0 (pointing angle) is expected to be very close to 0 because the smaller the pointing angle, the better the reconstruction line of the mother track and the vertex line match. So the cosine of the pointing angle (CPA) is required to be very close to 1. In addition, only the V^0 candidates that decay inside a fiducial volume are taken into account: the radial distance ρ between the secondary vertex and the beam line is required to be smaller than 1 m.

Concerning the bachelor candidates, also in this case the impact parameter with respect to the primary vertex is required to be small and, consequentially, the cosine of the angle between the Λ_c and the bachelor flying direction very close to 1. The measured TPC $\mathrm{d}E/\mathrm{d}x$ and TOF time-of-flight of the bachelor track must have a 4σ compatibility with the expected value for the proton hypothesis using OR-logic between the two detectors. In case of missing TOF information, the tracks were selected relying only on the TPC. Finally, the compatibility of the computed invariant mass from the decay products and the expected mother value from the PDG was checked for the Λ_c and the K^0_S candidates. A selection cut on the absolute value of their difference was applied:

$$|M(p,\pi^+,\pi^-) - M(\Lambda_c)| < 0.2 \text{ GeV}/c^2 \tag{4.5}$$

$$|M(\pi^+,\pi^-) - M(K^0_S)| < 0.03 \text{ GeV}/c^2 \tag{4.6}$$

In addition, the K^0_S misidentification cases with a Λ^0 were removed asking for the fulfilling of the condition:

$$|M(\pi^+,\pi^-) - M(\Lambda^0)| > 0.05 \text{ GeV}/c^2 \tag{4.7}$$

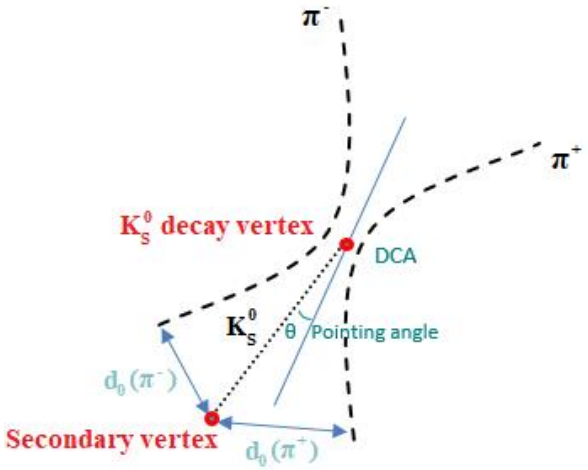

Figure 4.9: K^0_S decay in two opposite charged pions.

Variable	**Cut**
$\|d_0\|$ (V^0 daughters)	< 1.5 cm
DCA between V^0-daughters	< 0.8 cm
V^0 CPA	> 0.997
V^0 ρ (fiducial volume radius)	[0 m; 1 m]
$\|d_0\|$ (bachelor)	< 3 cm
$\mid cos(\Lambda_c, p) \mid$	> 0.9
TOF nσ(p) or TPC nσ(p)	< 4
$M(p, K^0_S) - M(\Lambda_c)$	< 0.2 GeV/c^2
$M(\pi^+, \pi^-) - M(K^0_S)$	< 0.03 GeV/c^2
$M(\pi^+, \pi^-) - M(\Lambda^0)$	> 0.05 GeV/c^2

Table 4.7: Prefilter cuts applied.

4.4 Machine learning algorithm

To perform the measurement with the highest possible precision, i.e. maximizing the signal-to-background (S/B) ratio and the statistical significance, the cuts were optimized with a machine learning (ML) technique based on Boosted Decision Tree (BDT) classification algorithm from the analysis package XGBoost (Extreme Gradient

Boosting) [71]. BDT is a supervised machine learning algorithm that takes in input a labelled dataset (training data sample) and builds a model able to classify the data in signal and background candidates. The "Decision Tree" is built relying on a set of input variables (training variables) that exploit physics knowledge of the signal and background. At each step the labelled sample is splitted in "branches" requiring that a condition on one of the training variables is fulfilled. Candidates passing the cut condition go one way and those failing go the other. A probability to be a signal candidate, p_i, is associated to the candidates in each branch by computing the branch purity (S/(S+B)). The points in which the tree is splitted are called "nodes". Each split at a node is chosen to maximize the separation between signal and background candidates, i.e. the branch purity. This is done defining the so called "loss function" that has to be minimized after each split. In this analysis, a logistic-loss functions was utilized to train the model, according to the following definition.

$$L = \sum_i \left[y_i \ln(p_i) \; + \; (1 - y_i) \ln(1 - p_i) \right] \tag{4.8}$$

where:

y_i = label of the candidate: in our binary classification, it can take the values 0 and 1, respectively, for the background and the signal candidates;

p_i = probability that the candidate i-th is a signal one.

When the observation belongs to class 1 the first part of the formula becomes active and the second part vanishes and vice versa in the case in which the actual class of the observation is 0. So, the loss function assumes the following shape:

$$L = \begin{cases} \sum_i \ln(p_i) & \text{if } y_i = 1 \\ \sum_i \ln(1 - p_i) & \text{if } y_i = 0 \end{cases} \tag{4.9}$$

After each split, the loss function, L, is evaluated for the produced left and the right branches. The gain is computed as the difference between the value that the loss function assumes before fulfilling the condition on the training variables, L_{tot}, and

the loss functions obtained after the split.

$$\text{Gain} = L_{\text{tot}} - (L_{\text{left}} + L_{\text{right}}) \tag{4.10}$$

The algorithm iterates over all features and values per each feature and evaluates each possible split loss reduction. The split with the highest gain will be selected. A decision tree example is shown in Fig.4.10. The growth of the tree is regulated by the XGBoost hyperparameters. This parameters help to control the complexity of the model, avoiding an overfitting[4] of the training data sample. Precisely for this reason, the boosting method was employed: many weak learners (trees) were trained and then combined into a strong classifier, adding up the predictions of multiple trees together. In this way, the stability of decision tree was enhanced improving also the performance of the method but without the risk of overtraining.

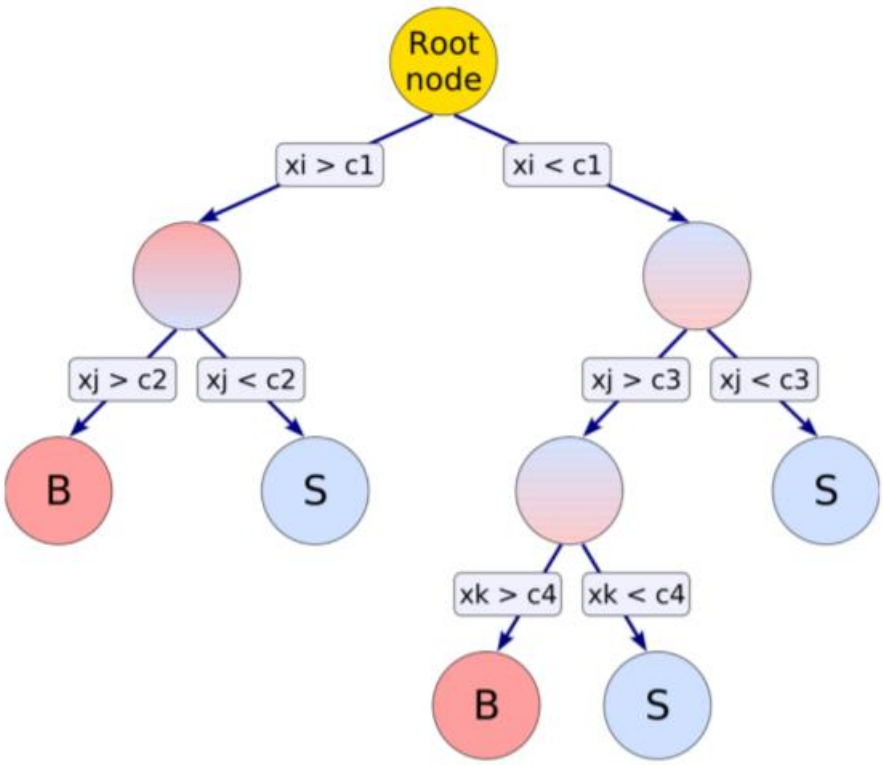

Figure 4.10: Schematic of a decision tree. The tree is built splitting the sample depending on the values assumed by the three training variables X = [x_i, x_j, x_k].

[4]The overtraining (or overfitting) is the production of a too complex model that classifies very well the training set of data but cannot generalize to new data, leading to poor performance.

The hyperparameters used to control the growing of tree and the boosting procedure are:

- min-child-weight: minimum sum of the factor $p_i(1 - p_i)$ needed in a child. If the tree partition step results in a leaf node with this sum less than min-child-weight, the building process will give up further partitioning. This parameter imposes a stop to the splitting once a certain degree of purity in a node is reached.
- max depth: maximum number of splits (nodes) of a tree. Increasing this value will make the model more complex and more likely to overfit.
- subsample: subsample fraction of the training dataset used to build the decision tree. It helps to prevent overfitting. Subsampling will occur once in every boosting iteration.
- colsample-bytree: fraction of features (randomly selected) that will be used to train each tree.
- learning-rate: weighting factor applied to the new trees when added to the model to slow down the learning in the gradient boosting.
- n-estimators: number of weak learners combined.

The BDT's hyperparameters were optimised with a grid search scan: different hyperparameter combinations were tested and the Area-Under-Curve (AUC) was used to judge on the performance of the different configurations. The scan was performed for each Λ_c^+ p_T interval; the optimal configuration was found to be the same for all Λ_c^+ p_T bins. The chosen set of XGBoost hyperparameters is reported in Table 4.8.

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

Table 4.8: The set of XGBoost BDT hyperparameters used to train the model. The same parameters were used for all Λ_c p_T intervals.

At the end of the process the variable phase space is thus split into many regions, called leaves. The BDT returns the probability value that a candidate in a given leaf is a signal one. A candidate is selected based on whether its assigned probability exceeds a chosen working point (WP) and flagged as background if the assigned value is below that.

4.4.1 BDT training

The training sample was assembled considering the background from the sidebands of the candidate invariant-mass distribution in data ($M_{\rm inv}(\Lambda_c) < 2.24$ GeV/c^2 or $M_{\rm inv}(\Lambda_c) > 2.33$ GeV/c^2), and the prompt[5] $\Lambda_c^+ \to {\rm pK}^0_{\rm s} \to {\rm p}\pi^+\pi^-$ signal candidates from MC simulations based on the `PYTHIA` Monash event generator [35]. A dedicated heavy-flavour enriched MC production, in which at least one Λ_c decaying in $p{\rm K}^0_{\rm S}$ within $|y_{\Lambda_c}| < 1.0$ is produced, was utilized to increase the data sample size available for the training procedure. A maximum of 500k candidates of each source was considered. Indeed, if the data sample size is sufficient, the model converges and adding more input data does not lead to further improvements but only to an increase of the required training time. At higher p_T, where the sample size for the signal is less, the number of selected background candidate was lowered to match the

[5]The prompt $\Lambda_c^\pm$ are the $\Lambda_c^\pm$ baryons that are produced from a c quark hadronisation, while the feed-down $\Lambda_c^\pm$ are the ones that come from the decay of a B meson.

available signal. Two independent sets were prepared, namely a training and a test sample, containing 80% and 20% of available samples, respectively. The training sample is used to train the model. The performance of the trained model were then studied on the test sample. Similar performance are expected in the two cases if the model is properly trained. Indeed, our goal is to create a model that generalizes well to new data avoiding overfitting of the training dataset. A dedicated model is constructed for each analysed p_{T} bin using the multiplicity integrated MB MC and data samples. The training data sample size was found to be insufficient when splitting it in multiplicity bins.

Topological variables as well as the PID information of each candidate daughter were exploited in the model training. All used features are listed in the following:

- $m_{inv}(\mathbf{V^0})$: invariant mass of the V^0 candidate;
- $\mathbf{DCA}(\pi^+\pi^-)$: distance of closest approach between the two prongs of the V^0 decay;
- $d_{len}(\mathbf{V^0})$: V^0 primary-to-secondary vertex distance;
- $c\tau(\mathbf{V^0})$: $d_{len}(\mathrm{V}^0)$ * $\mathrm{K}^0_{\mathrm{S}}$ mass (0.497 GeV/c) over V^0 momentum;
- $\mathbf{cos}\ \theta_{point}(\mathbf{V^0})$: cosine of the pointing angle of the V^0 candidate;
- $\mathbf{arm}(\mathbf{V^0})$: transverse momentum of the positive daughter particle with respect to the reconstructed mother particle's (V^0 candidate) momentum divided by the V^0 longitudinal momentum asymmetry α[6];
- b_{V^0}: V^0 candidate impact parameter, i.e. distance of closest approach of the V^0 candidate to the primary vertex;

[6]The V^0 longitudinal momentum asymmetry α is defined as:

$$\alpha = \frac{p_L^+ - p_L^-}{p_L^+ + p_L^-} \tag{4.11}$$

where p_L^+ and p_L^- are the longitudinal momentum of the positive and negative daughter particles, respectively. The $\mathrm{K}^0_{\mathrm{S}}$ decays into two particles with the same mass so their momenta are distributed symmetrically. This is not true for the V^0 from Λ^0 ($\overline{\Lambda^0}$) decays in $p\pi^+$ ($\overline{p}\pi^-$) in which the proton takes a larger part of the momentum. Therefore, the Armenteros variable allows discriminating the V^0 from $\mathrm{K}^0_{\mathrm{S}}$ from those that come from Λ^0 and $\overline{\Lambda^0}$ decays.

- b_p: prong0 impact parameter, i.e. distance of closest approach of the bachelor candidate to the primary vertex;
- b_{π^+}: prong1 impact parameter, i.e. distance of closest approach of the positive V^0 prong to the secondary vertex;
- b_{π^-}: prong2 impact parameter, i.e. distance of closest approach of the negative V^0 prong to the secondary vertex.
- **cos** θ^*: cosine of the proton emission angle in the rest frame of the mother particle;
- $\mathbf{n\sigma_{TOF}(p)}$: number of sigma for proton hypothesis in the TOF;
- $\mathbf{n\sigma_{TPC}(p)}$: number of sigma for proton hypothesis in the TPC.

Since some of the low p_T tracks cannot reach the TOF, the $n\sigma_{TOF}(p)$ is not available for all the candidates. XGBoost is able to manage independently missing features values, computing both the assignments and choosing the one that minimizes the loss. The distributions of the training variables for signal and background candidates are shown in Fig. 4.11. Also the separation power p_{sep} is reported for each variable. It is defined as:

$$p_{\text{sep}} = \frac{1}{2} \sum_{i}^{N_{bins}} \frac{(s_i - b_i)^2}{s_i + b_i} \tag{4.12}$$

where:

N_{bins} = number of bins of the histogram;

s_i = signal entries in the bin i;

b_i = background entries in the bin i.

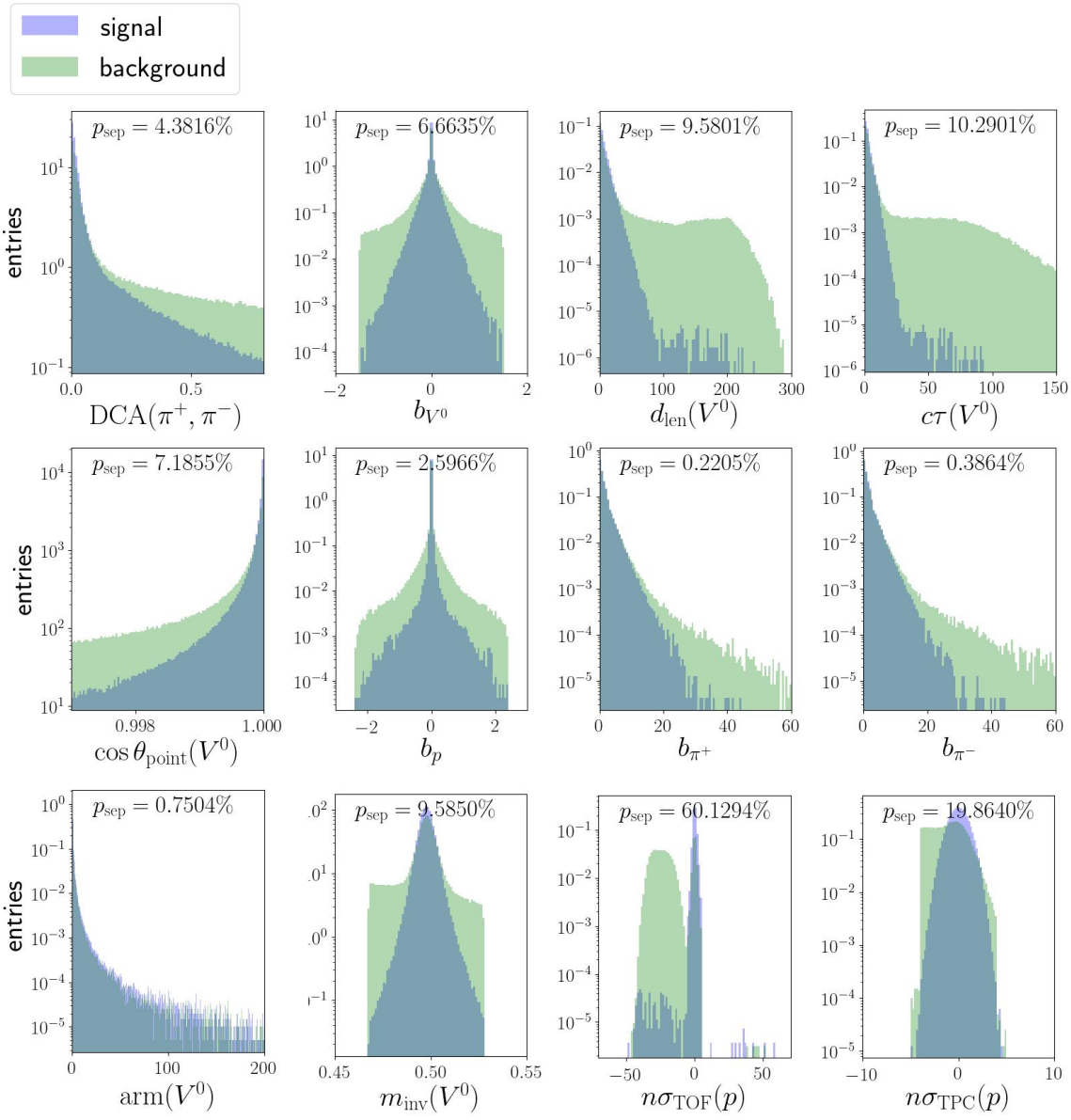

Figure 4.11: Distributions of training variables for prompt Λ_c baryons, and combinatorial background candidate in the $1 \leq p_T < 2$ GeV/c interval.

The separation power provides an indication of the discriminatory power of the training variables. A more precise way of establishing the weight of the training variables on the final model are the SHAP (SHapley Additive exPlanations) plots reported in Figs 4.12, 4.13 and 4.14, respectively for Λ_c p_T bins [1,2), [2,4), [4,6), [6,8), [8,12) and [12,24) GeV/c. Each dot in the plot corresponds to one candidate and for each candidate is estimated a SHAP value that gives information on how much a feature drives the prediction in the positive or negative direction. The SHAP

value for a training variable A is defined as:

$$SHAP(A) = \sum_{S \subseteq N \setminus A} \frac{|S|!\ (n - |S| - 1)!}{n!}\ (p(N) - p(S)) \tag{4.13}$$

where:

N = feature set;

n = number of features. $n!$ represent the total number of features combination;

$S \subseteq N \setminus A$ = the sum is extended on all the feature subsets S that can be built excluding the features A;

$|S|$ = number of features included in the subset S;

p(N) = probability associated to the candidate when all the feature are considered in the training;

p(S) = probability associated to the candidate when only the features in the subset S are utilized for the training.

The SHAP estimator quantifies the discrimination power of a specific feature A by evaluating the prediction accuracy of the model when such feature is removed. A high (low) shap value means that a feature value tends to produce a more signal-like (background-like) prediction. For the points with a shap value around 0, the particular feature do not really have an impact on whether it is a signal or background decision. The colour indicates the magnitude of a feature for a given candidate. The red colour points to "high value" while the blue one to "low value" of the feature. The final importance for a given feature is then obtained adding up the squared sum of the single SHAP values for all the candidate related to the feature and it determines how likely a given feature impacts the prediction overall. The features in the plot are sorted by decreasing importance, from top to bottom. As expected, the most important variable in the training is the nσ_{TOF}(p). Low nσ_{TOF}(p) values tend to impact the prediction towards background. The candidates with a low nσ_{TOF}(p) are indeed constituted by pions contamination that will be rejected by the BDT cut on this variable.

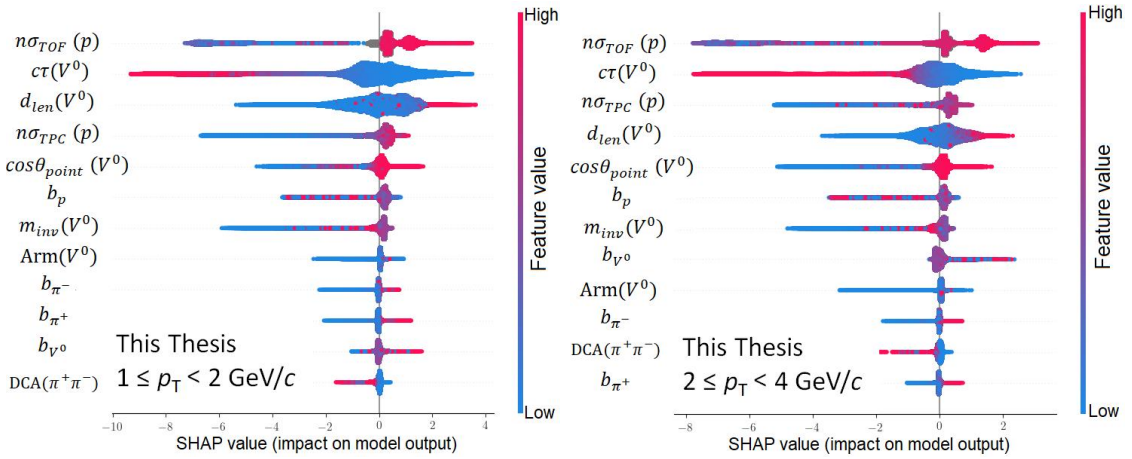

Figure 4.12: SHAP feaures importance in the $1 \leq p_{\mathrm{T}} < 2$ GeV/c (left) and $2 \leq p_{\mathrm{T}} < 4$ GeV/c (right) intervals.

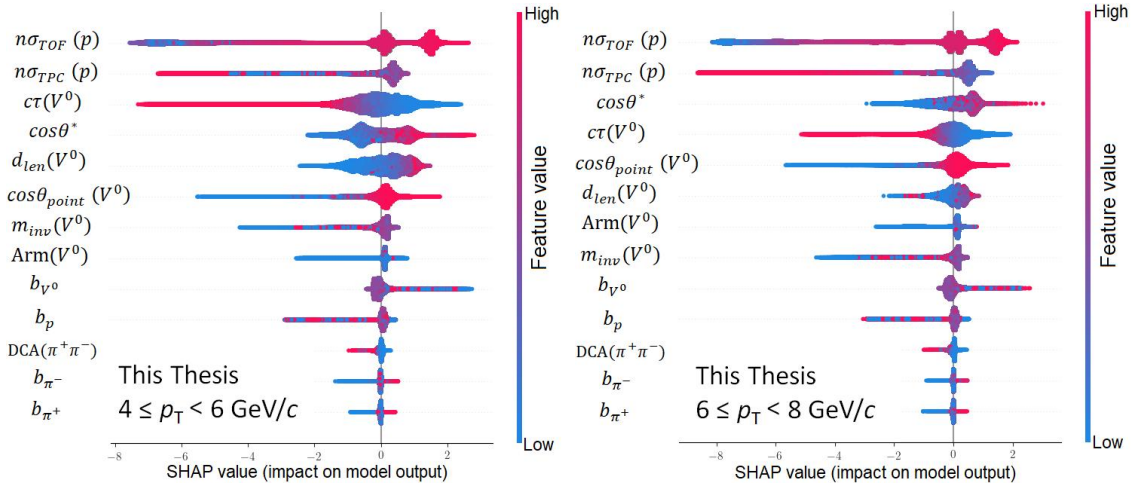

Figure 4.13: SHAP feaures importance in the $4 \leq p_{\mathrm{T}} < 6$ GeV/c (left) and $6 \leq p_{\mathrm{T}} < 8$ GeV/c (right) intervals.

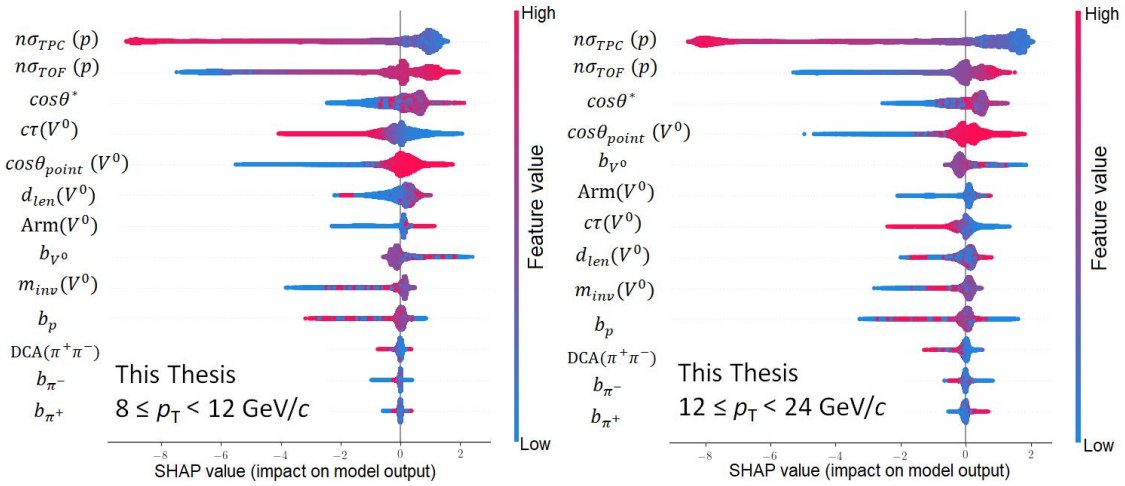

Figure 4.14: SHAP feaures importance in the $8 \leq p_{\mathrm{T}} < 12$ GeV/c (left) and $12 \leq p_{\mathrm{T}} < 24$ GeV/c (right) intervals.

The linear correlations between the variables used in the training were studied in detail and are reported for the two classes (signal and background) of candidates in Fig. 4.15-4.20. The correlation coefficient measures the strength and direction of a linear relationship. In particular, the value 1 indicates a perfect positive correlation, -1 a perfect negative correlation (anti-correlation) and 0 completely mutual independence, i.e. that there is no relationship between the different variables.

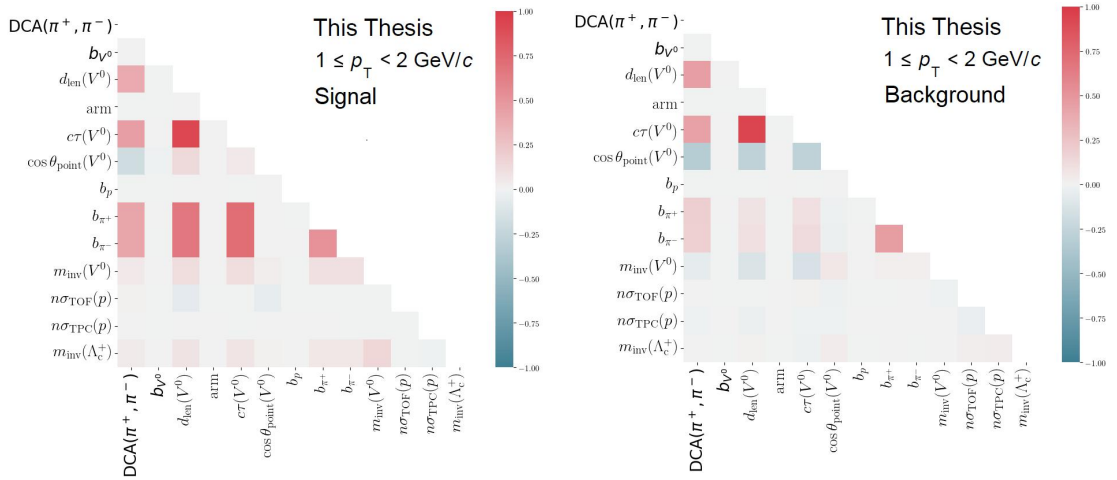

Figure 4.15: Features correlations for signal (left) and background (right) candidates in the $1 \leq p_{\mathrm{T}} < 2$ GeV/c interval.

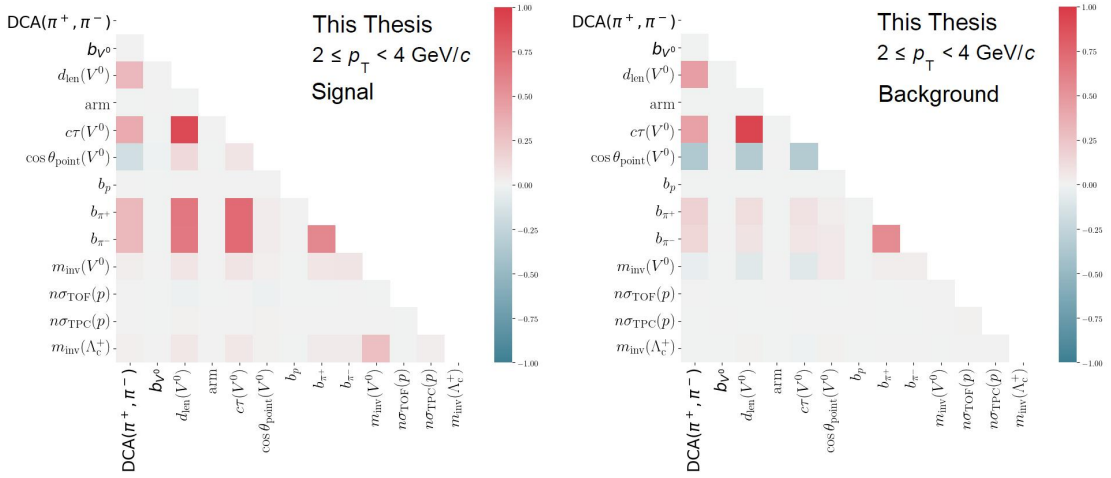

Figure 4.16: Features correlations for signal (left) and background (right) candidates in the $2 \leq p_{\mathrm{T}} < 4$ GeV/c interval.

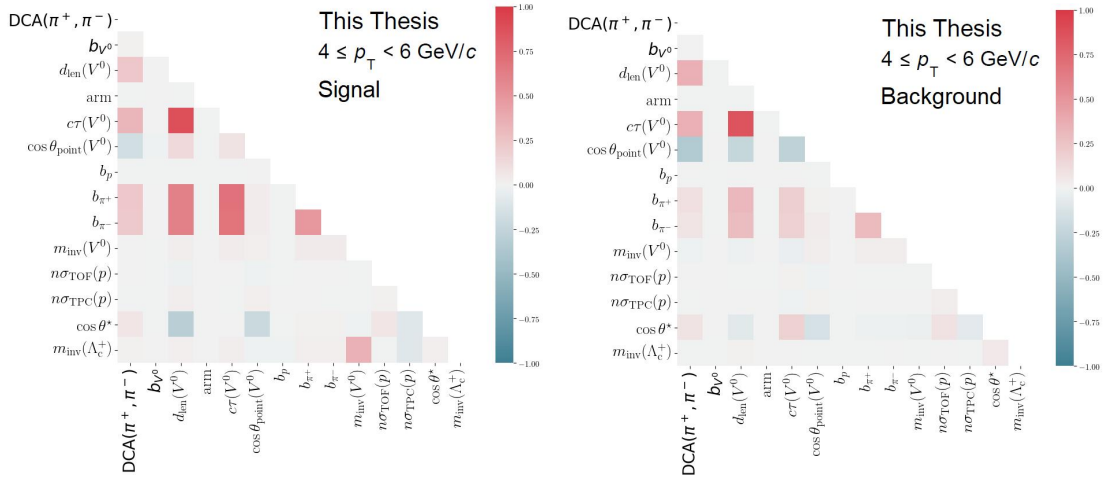

Figure 4.17: Features correlations for signal (left) and background (right) candidates in the $4 \leq p_{\mathrm{T}} < 6$ GeV/c interval.

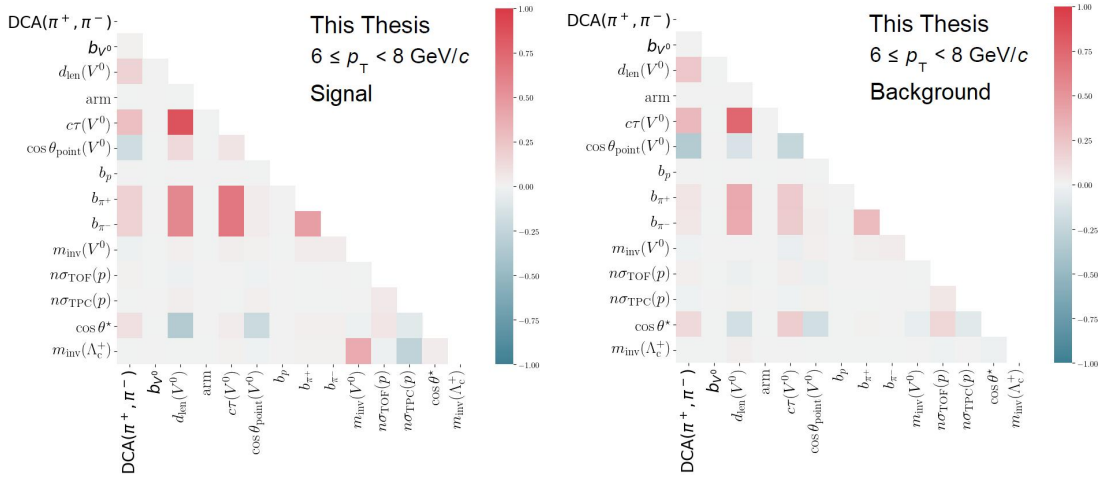

Figure 4.18: Features correlations for signal (left) and background (right) candidates in the $6 \leq p_{\mathrm{T}} < 8$ GeV/c interval.

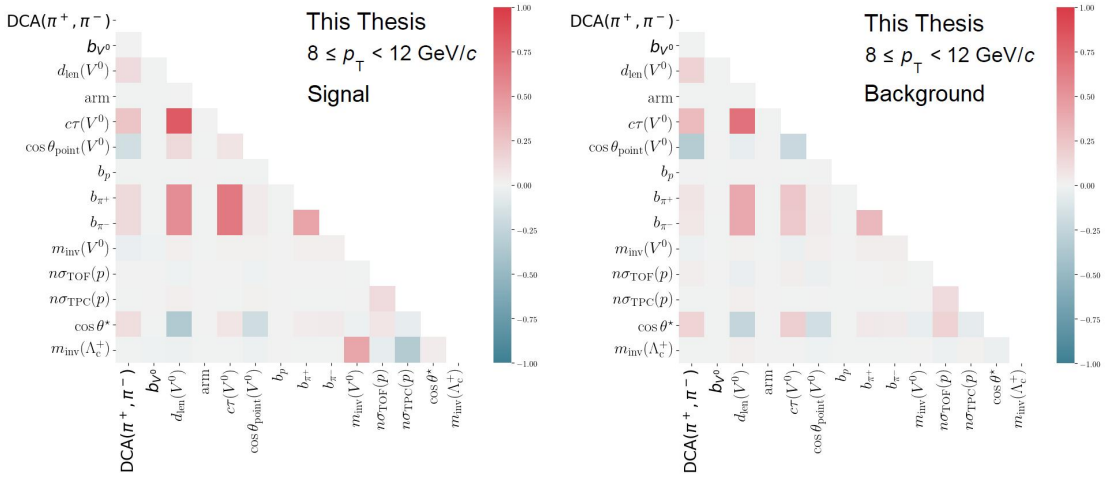

Figure 4.19: Features correlations for signal (left) and background (right) candidates in the $8 \leq p_{\mathrm{T}} < 12$ GeV/c interval.

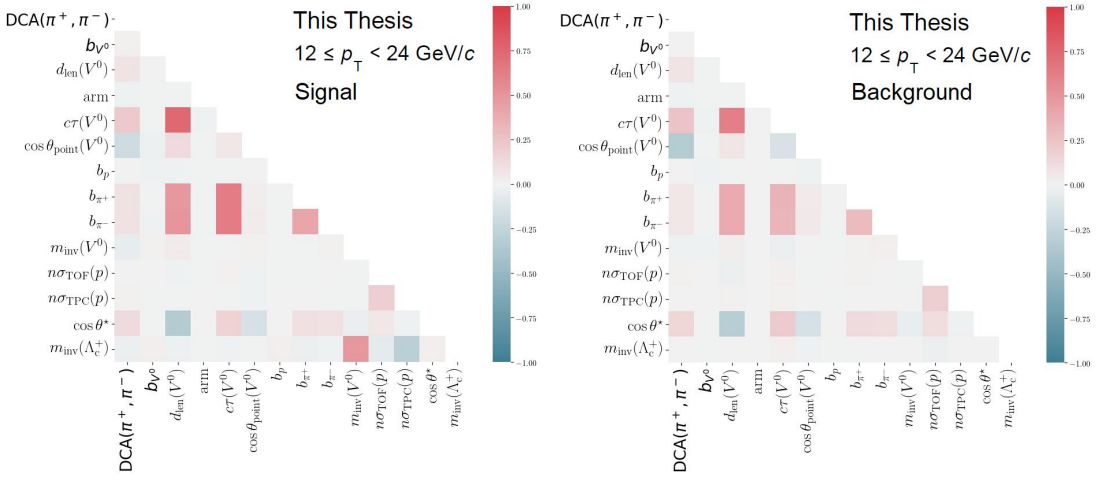

Figure 4.20: Features correlations for signal (left) and background (right) candidates in the $12 \leq p_{\mathrm{T}} < 24$ GeV/c interval.

Variables that carry the same physical information, such as those related to the V^0 decay topology ($b_{\pi^{\pm}}$, $c\tau(\mathrm{V}^0)$, $d_{len}(\mathrm{V}^0)$), are strongly correlated as expected. Moreover, there are some differences in the variable correlations between signal and background candidates, which could be exploited by the model to discriminate signal from background. It is also useful to control the presence of correlations between the training variables and the invariant mass of the candidates. In particular, it is preferable that the selection based on the model predictions does not significantly modify the invariant mass distributions. In fact, a complex shape of the background invariant mass distribution could require a function with more free parameters to be described, therefore the yield extraction could be more affected by eventual background fluctuation. This is the reason why the cos θ^* was removed from the training procedure in the two lowest p_{T} bins. No further correlations were observed between the training variables and the invariant mass, thus no deformations of the background invariant mass distribution are expected.

4.4.2 BDT performance

After the training, the model was applied to both the training and test sets to check the performance and the agreement between the two sets. The BDT model assigns

to each candidate a BDT score that represents the probability of a candidate to be a prompt Λ_c signal. The BDTs scores for the training and test sample are reported in Figs. 4.21, 4.22, 4.23. The quality of the training is confirmed by the very similar scores obtained on the two sets.

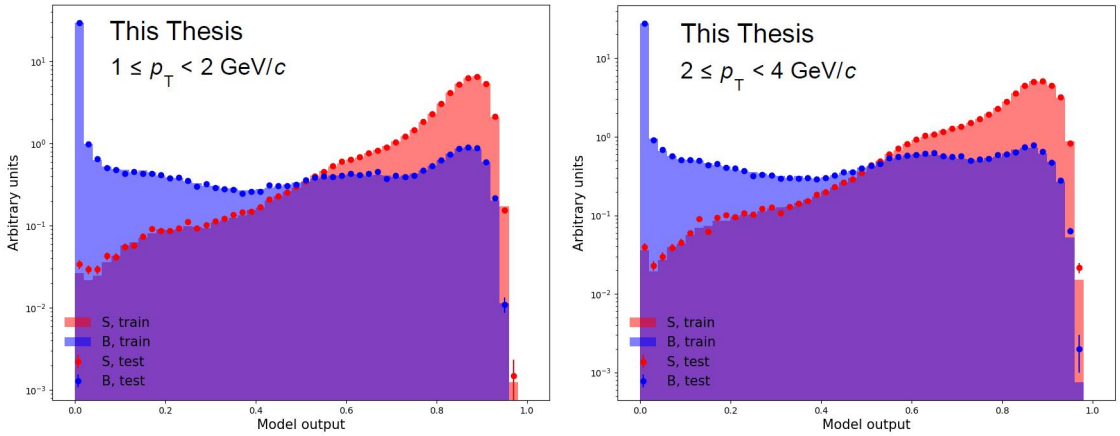

Figure 4.21: Distributions of ML output scores for signal (red) and combinatorial background (blue) candidates, for the training and the test sets in the $1 \leq p_{\rm T} < 2$ GeV/c (left) and $2 \leq p_{\rm T} < 4$ GeV/c (right) intervals.

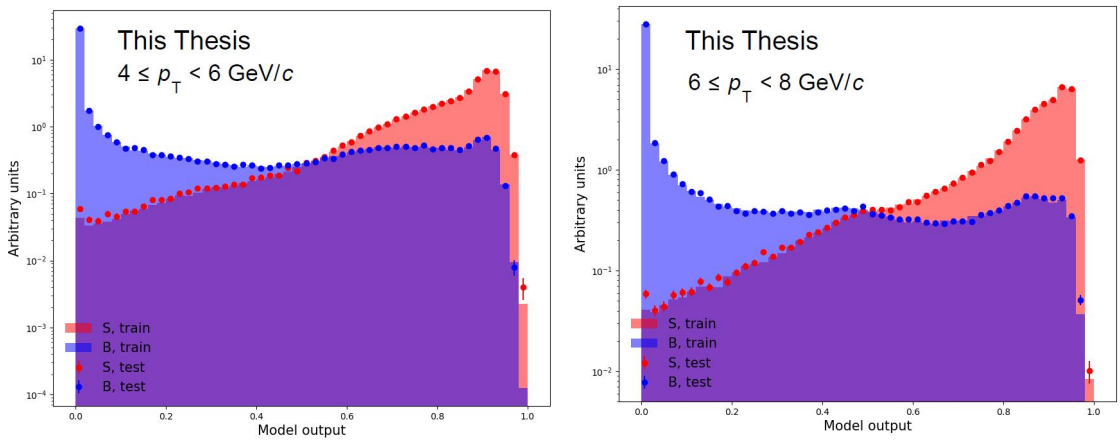

Figure 4.22: Distributions of ML output scores for signal (red) and combinatorial background (blue) candidates, for the training and the test sets in the $4 \leq p_{\rm T} < 6$ GeV/c (left) and $6 \leq p_{\rm T} < 8$ GeV/c (right) intervals.

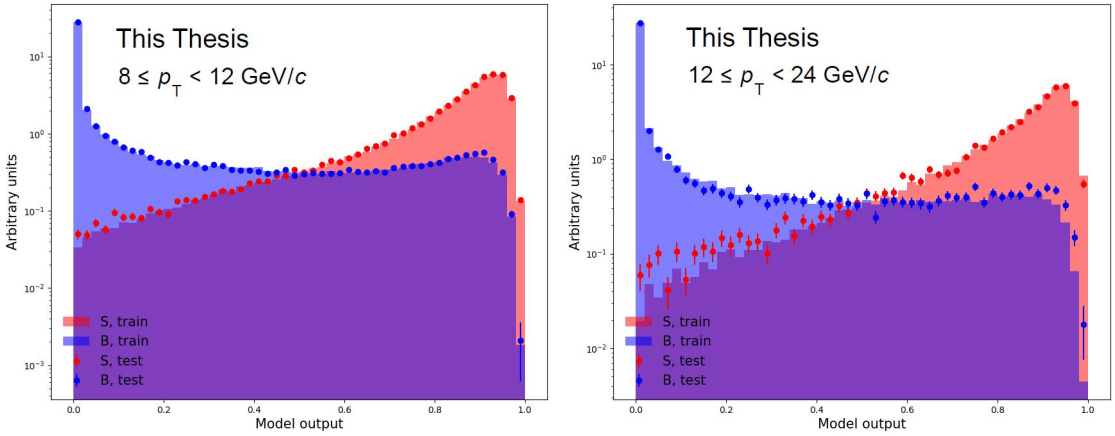

Figure 4.23: Distributions of ML output scores for signal (red) and combinatorial background (blue) candidates, for the training and the test sets in the $8 \leq p_{\mathrm{T}} < 12$ GeV/c (left) and $12 \leq p_{\mathrm{T}} < 24$ GeV/c (right) intervals.

For a more quantitative estimation of model performance, the area under the Receiver Operating Characteristic curve (ROC AUC) is considered. The ROC curve is obtained plotting the signal selection efficiency (the True Positive Rate (TPR) in more general terms) as function of the background selection efficiency (False Positive Rate (FPR)), for various threshold settings on the model output. The TPR and the FPR are defined as in the following:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad , \quad \mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{4.14}$$

where:

TP (True Positive) = true signal candidate that has been selected;

FN (False Negative) = true signal candidate that has been rejected;

FP (False Positive) = background candidate that has been selected;

TN (True Negative) = background candidate that has been rejected.

The possible values of the ROC AUC range in [0.5, 1] where 0.5 corresponds to a random classification and 1 to a perfect discrimination between the two hypothesis. The ROC AUC gives a global estimation of the model performance, i.e. not related

to the threshold value that will be chosen. The ROC curves were estimated for both train and test samples and are reported in Figs. 4.24, 4.25, 4.26. The ROC AUC in the two cases are very similar: the model is not overfitting the training sample and has a good generalization capability. In addition, the ROC AUC values are close to 1 in all the p_{T} intervals, pointing to a proper identification of the candidates.

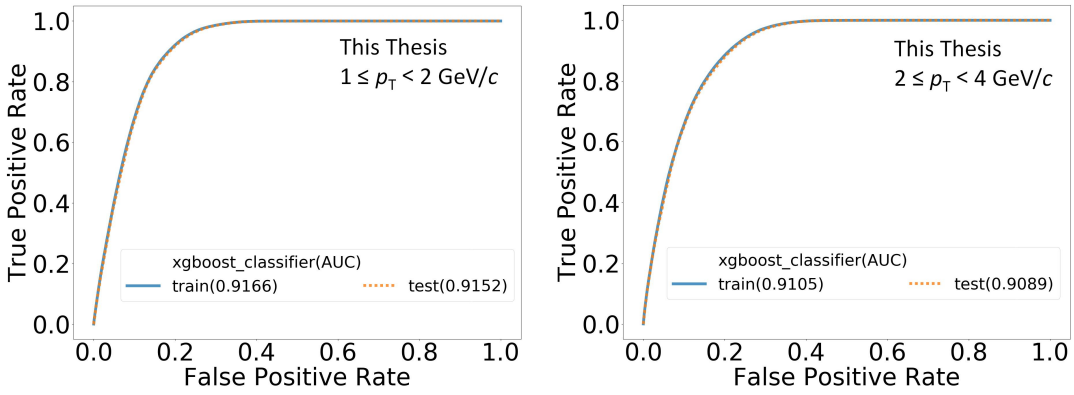

Figure 4.24: ROC curves for the training and the test sets in the $1 \leq p_{\mathrm{T}} < 2$ GeV/c (left) and $2 \leq p_{\mathrm{T}} < 4$ GeV/c (right) intervals.

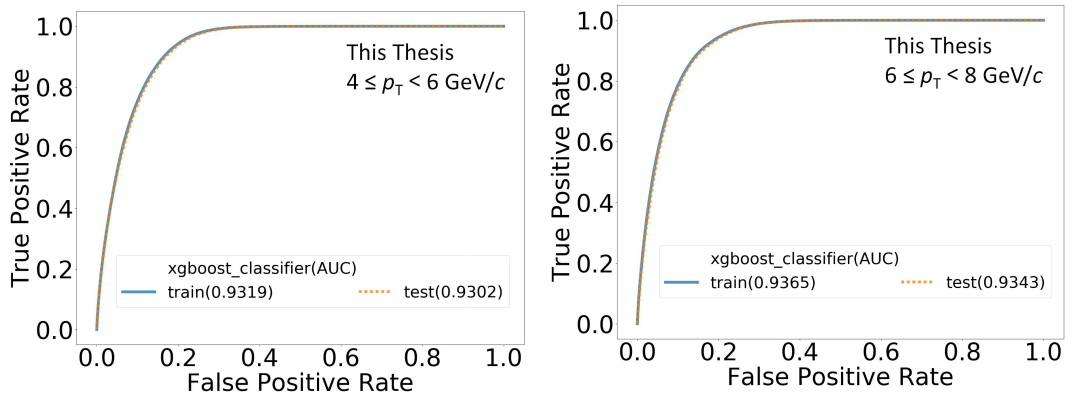

Figure 4.25: ROC curves for the training and the test sets in the $4 \leq p_{\mathrm{T}} < 6$ GeV/c (left) and $6 \leq p_{\mathrm{T}} < 8$ GeV/c (right) intervals.

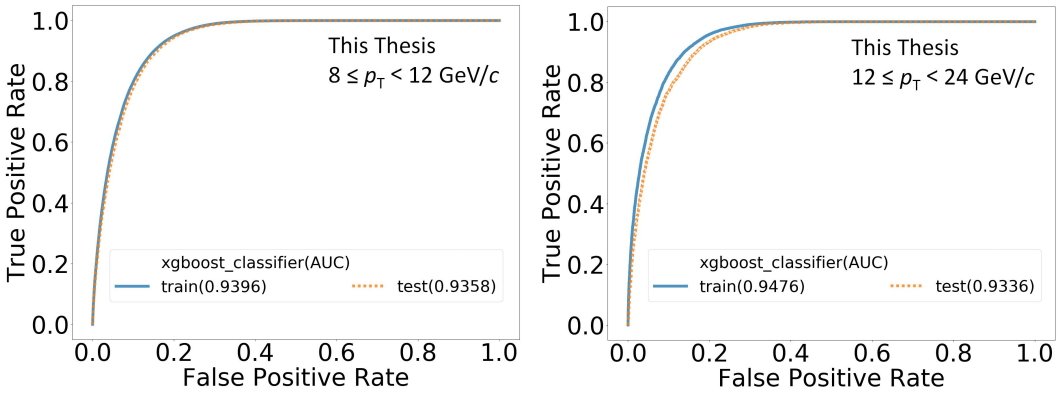

Figure 4.26: ROC curves for the training and the test sets in the $8 \leq p_{\mathrm{T}} < 12$ GeV/c (left) and $12 \leq p_{\mathrm{T}} < 24$ GeV/c (right) intervals.

4.4.3 Working Point determination

To improve the purity of our sample, a cut was applied on the BDT output. The criterion employed to fix our threshold value (or working point WP) was to select the BDT cut that maximises the significance. To achieve our goal, the expected significance, defined in eq. (4.15), was computed versus the BDT threshold:

$$\text{Significance} = \frac{S}{\sqrt{S+B}} \tag{4.15}$$

where S and B are, respectively, the expected signal yield and background one in the signal invariant mass region.

The signal term was retrieved according to the following formula:

$$S_{\mathrm{exp}} = 2 \left(\frac{\mathrm{d}\sigma}{\mathrm{d}p_{\mathrm{T}}} \right)^{\mathrm{FONLL}}_{\mathrm{prompt}} \cdot (\mathrm{Acc} \times \mathrm{Eff})_{\mathrm{prompt}} \cdot c_{\Delta y} \cdot \Delta p_{\mathrm{T}} \cdot \mathrm{BR} \cdot \mathcal{L}_{\mathrm{int}} \cdot \frac{1}{f_{\mathrm{prompt}}} \tag{4.16}$$

where:

- the factor 2 is needed to take into account particle and antiparticles;
- $\left(\frac{\mathrm{d}\sigma}{\mathrm{d}p_{\mathrm{T}}}\right)^{\mathrm{FONLL}}_{\mathrm{prompt}}$ is the `FONLL` calculation for the $\Lambda_{\mathrm{c}}^{+} \rightarrow \mathrm{pK}_{\mathrm{s}}^{0} \rightarrow \mathrm{p}\pi^{+}\pi^{-}$ p_{T}-differential cross section;
- Δp_{T} is the width of the p_{T} interval;

- $c_{\Delta y}$ is the correction factor for the rapidity coverage (needed because of the fiducial-acceptance selection);
- BR is the branching ratio of the Λ_c in the hadronic channel of interest for the analysis;
- $f_{\rm prompt}$ is the $\Lambda_c^+ \to {\rm pK_s^0} \to {\rm p}\pi^+\pi^-$ prompt fraction;
- $\mathcal{L}_{\rm int}$ is the integrated luminosity;
- $({\rm Acc} \times {\rm Eff})_{\rm prompt}$ is the acceptance times efficiency correction (described in detail in sec. 4.6.1) computed on the training data sample as the ratio between the signal candidates that survive to the applied BDT cut and the total number of signal candidate. The behaviour of the $({\rm Acc} \times {\rm Eff})_{\rm prompt}$ in the analysed $p_{\rm T}$ intervals can be found in Figs. 4.27, 4.28 and 4.29.

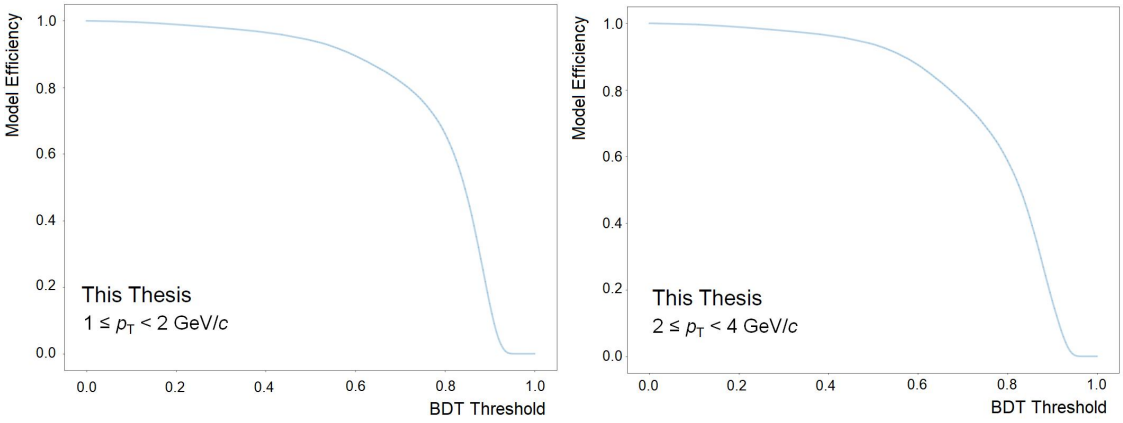

Figure 4.27: Efficiency trend versus the BDT threshold in the $1 \leq p_{\rm T} < 2$ GeV/c (left) and $2 \leq p_{\rm T} < 4$ GeV/c (right) intervals.

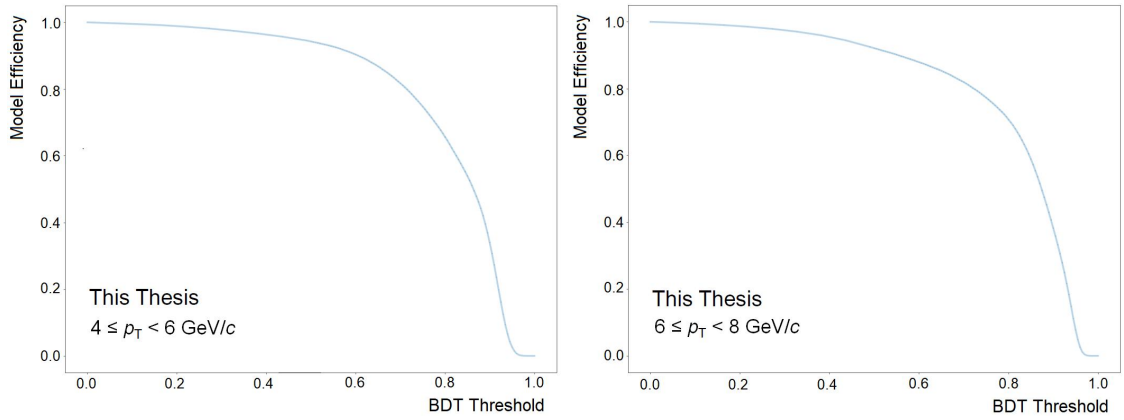

Figure 4.28: Efficiency trend versus the BDT threshold in the $4 \leq p_{\mathrm{T}} < 6$ GeV/c (left) and $6 \leq p_{\mathrm{T}} < 8$ GeV/c (right) intervals.

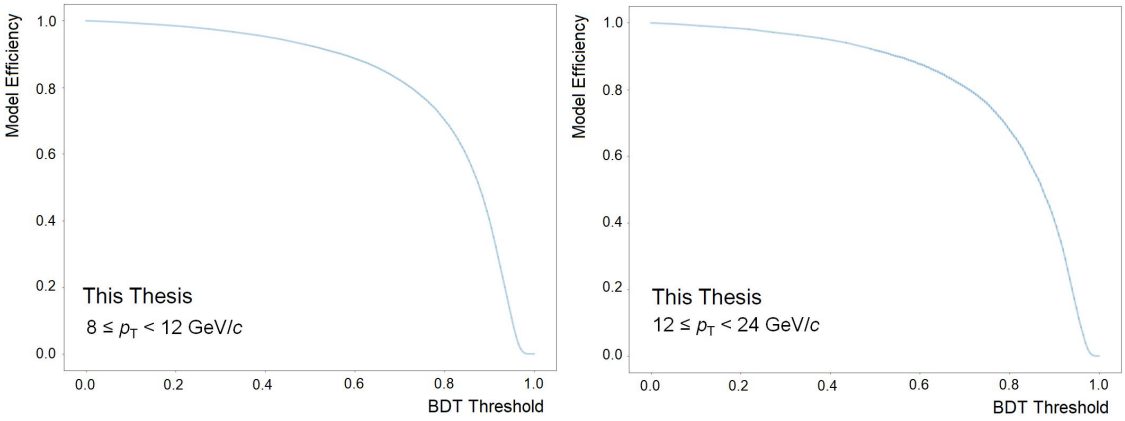

Figure 4.29: Efficiency trend versus the BDT threshold in the $8 \leq p_{\mathrm{T}} < 12$ GeV/c (left) and $12 \leq p_{\mathrm{T}} < 24$ GeV/c (right) intervals.

The expected background under the signal peak was, instead, evaluated on a fraction of the data by fitting the sidebands of the invariant-mass distribution and, subsequently, scaling it to match the one in the full data sample.

The optimal WP was selected as the value of probability that guarantees a high significance and a sufficiently high selection efficiency. Very low values of efficiencies are not ideal since they usually imply large values of systematics coming from cut variation (see section 4.8.2), due to the limited residual statistics after the BDT selection. In Figs. 4.30, 4.31 and 4.32 the results of these scans are reported for all the p_{T} intervals. The optimal working point was chosen to be 0.7 across all p_{T} bins.

BDT Cut					
$1 \leq p_T < 2$	$2 \leq p_T < 4$	$4 \leq p_T < 6$	$6 \leq p_T < 8$	$8 \leq p_T < 12$	$12 \leq p_T < 24$
0.7	0.7	0.7	0.7	0.7	0.7

Table 4.9: Optimal cut on the BDT probability applied in each p_T interval for the $\Lambda_c^+ \to pK_s^0 \to p\pi^+\pi^-$ analysis.

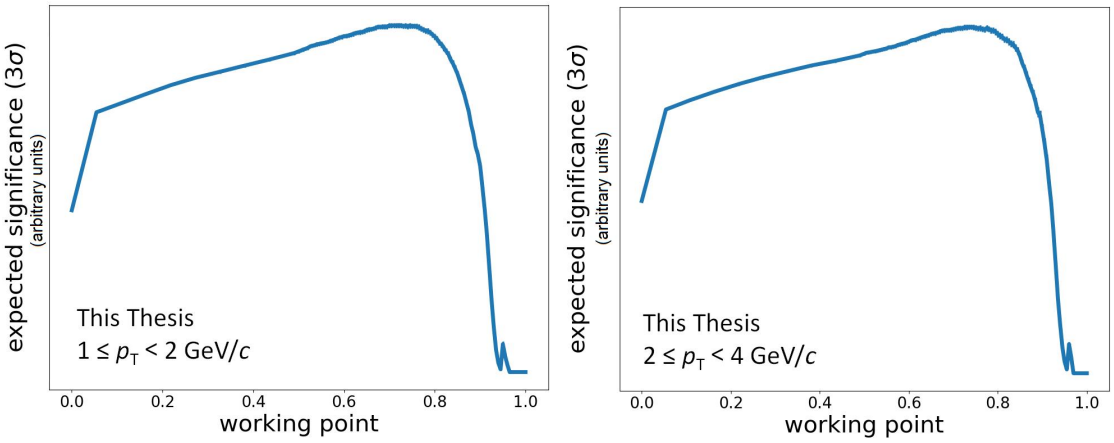

Figure 4.30: Significance as a function of the probability cut for Λ_c candidates in the $1 \leq p_T < 2$ GeV/c (left) and $2 \leq p_T < 4$ GeV/c (right) intervals.

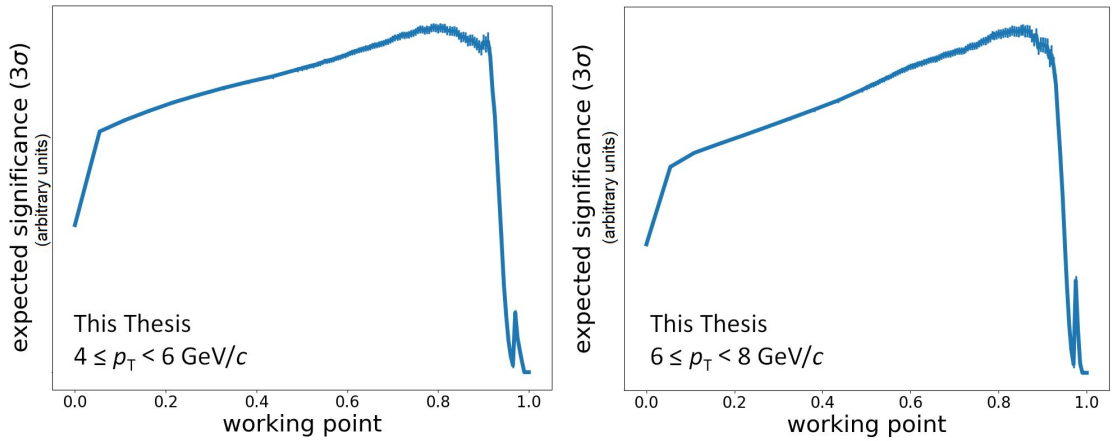

Figure 4.31: Significance as a function of the probability cut for Λ_c candidates in the $4 \leq p_T < 6$ GeV/c (left) and $6 \leq p_T < 8$ GeV/c (right) intervals.

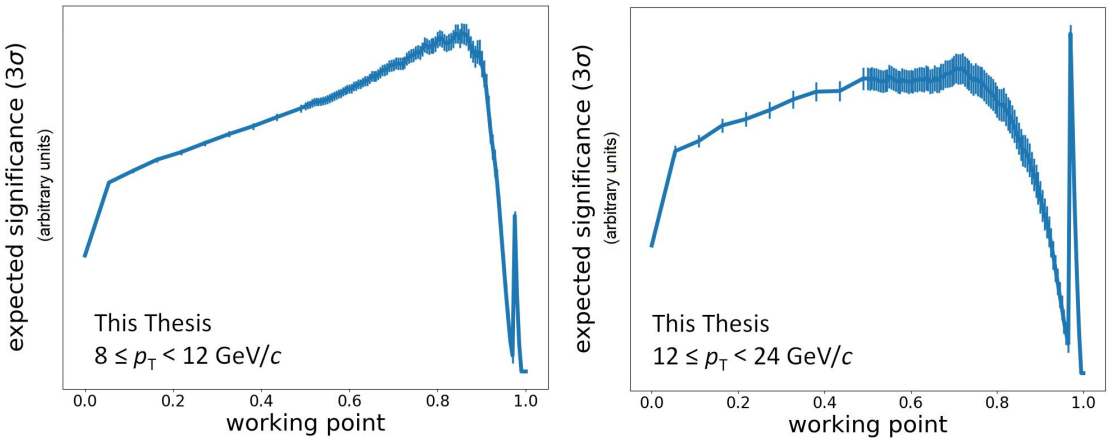

Figure 4.32: Significance as a function of the probability cut for Λ_c candidates in the $8 \leq p_T < 12$ GeV/c (left) and $12 \leq p_T < 24$ GeV/c (right) intervals.

4.5 Raw yield extraction

The trained model was applied to the data sample to extract the Λ_c raw yields. Λ_c candidates that pass the selection cuts were used to fill invariant mass distributions for each analysed p_T and multiplicity interval. The signal extraction was performed via binned maximum-likelihood fits to the invariant mass distributions. A gaussian function was used to describe the signal peak while a 2^{nd} order polynomial function was utilized to model the background. In order to improve the fit stability, the standard deviation of the gaussian signal function was fixed to the value obtained from simulations in the integrated multiplicity case, reported in Fig 4.33. The systematic effect of this choice is accounted for in the estimation of the systematic uncertainty of the raw yield extraction (section 4.8.1).

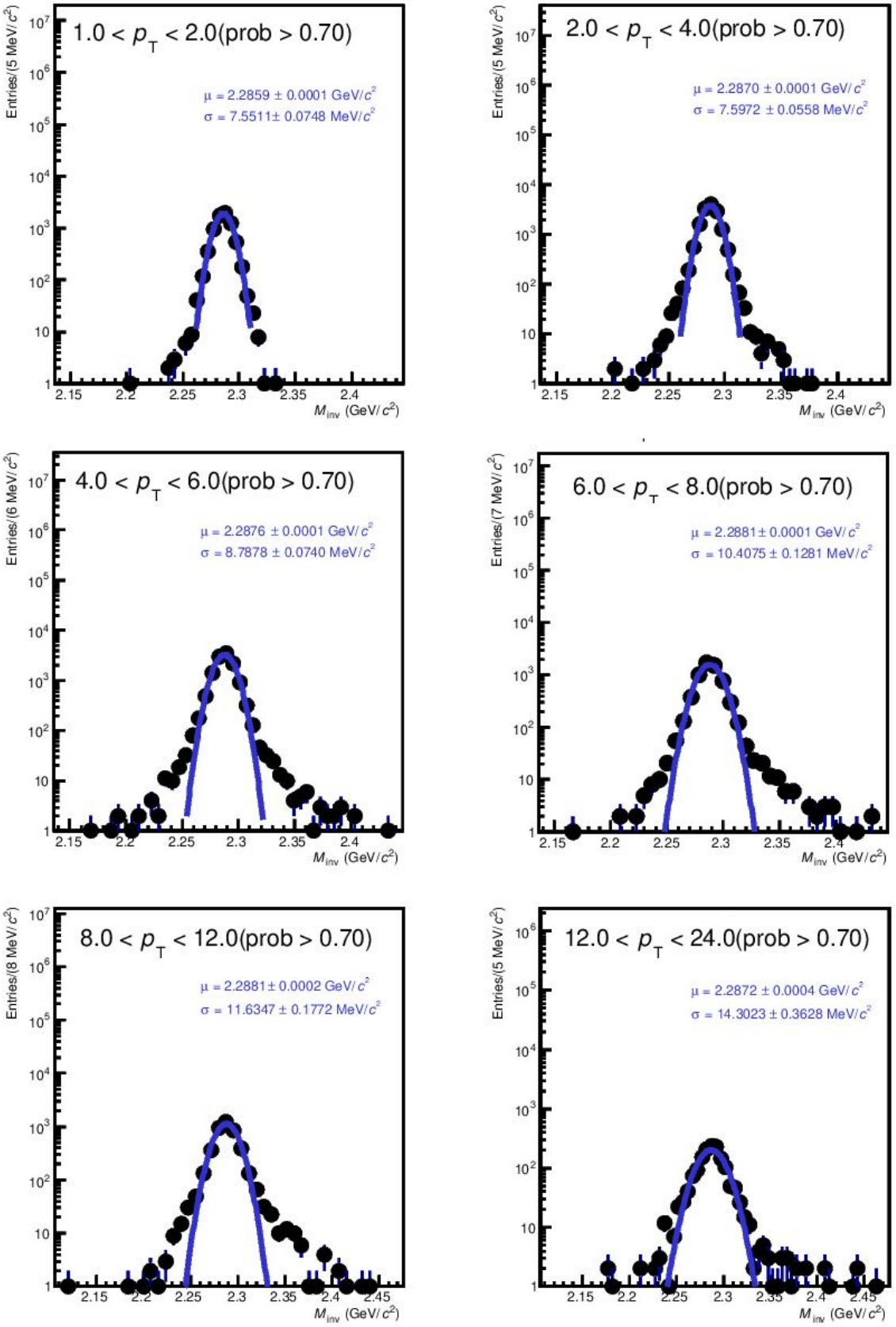

Figure 4.33: Width of the signal gaussian peak obtained from MC simulations in the integrated multiplicity case.

The widths obtained in the integrated multiplicity bin for data and simulation were compared to check the agreement between the two; it was observed that the sigma values in data are typically compatible with the simulation within 2 standard deviations. The peak width increases with increasing p_{T}, as expected from the transverse-momentum resolution of the decay tracks. The comparison is represented in Fig 4.34.

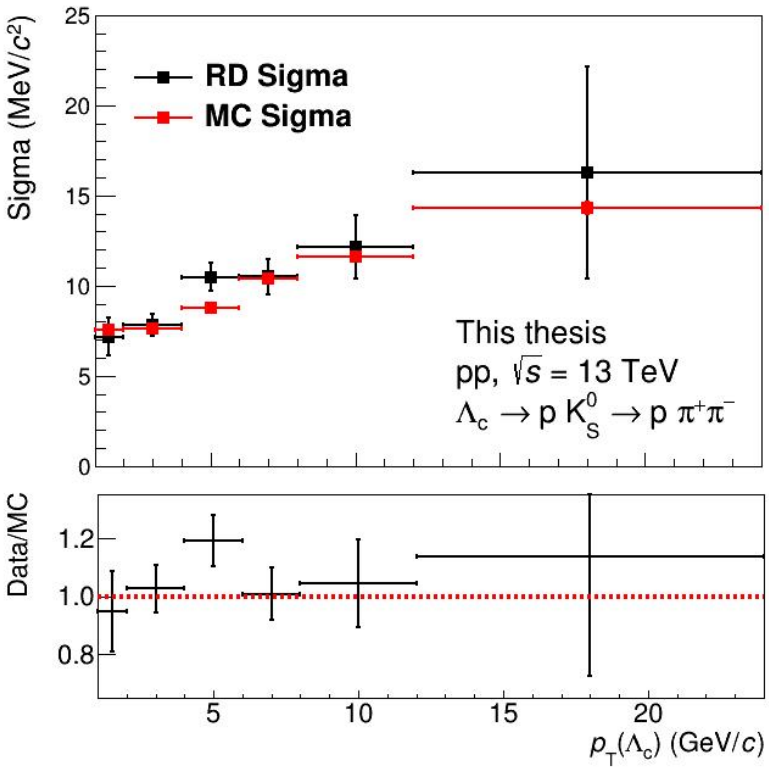

Figure 4.34: Comparison of the extracted width of the signal gaussian peak in data and in MC simulations in the multiplicity integrated case.

Due to the limited number of candidates in some multiplicity classes and the large combinatorial background, it was not possible to extract the raw yield in the full p_{T} range for all the multiplicity intervals: the range $12 < p_{\mathrm{T}} < 24$ GeV/c in the low multiplicity and high multiplicity classes is indeed not doable. The signal was determined as the integral of the gaussian function between $\pm 3\sigma$ and the background is determined as the integral of the background function in the same mass range. The Λ_{c} invariant mass plots for each p_{T} bin and multiplicity class are shown in Figs. 4.35- 4.39.

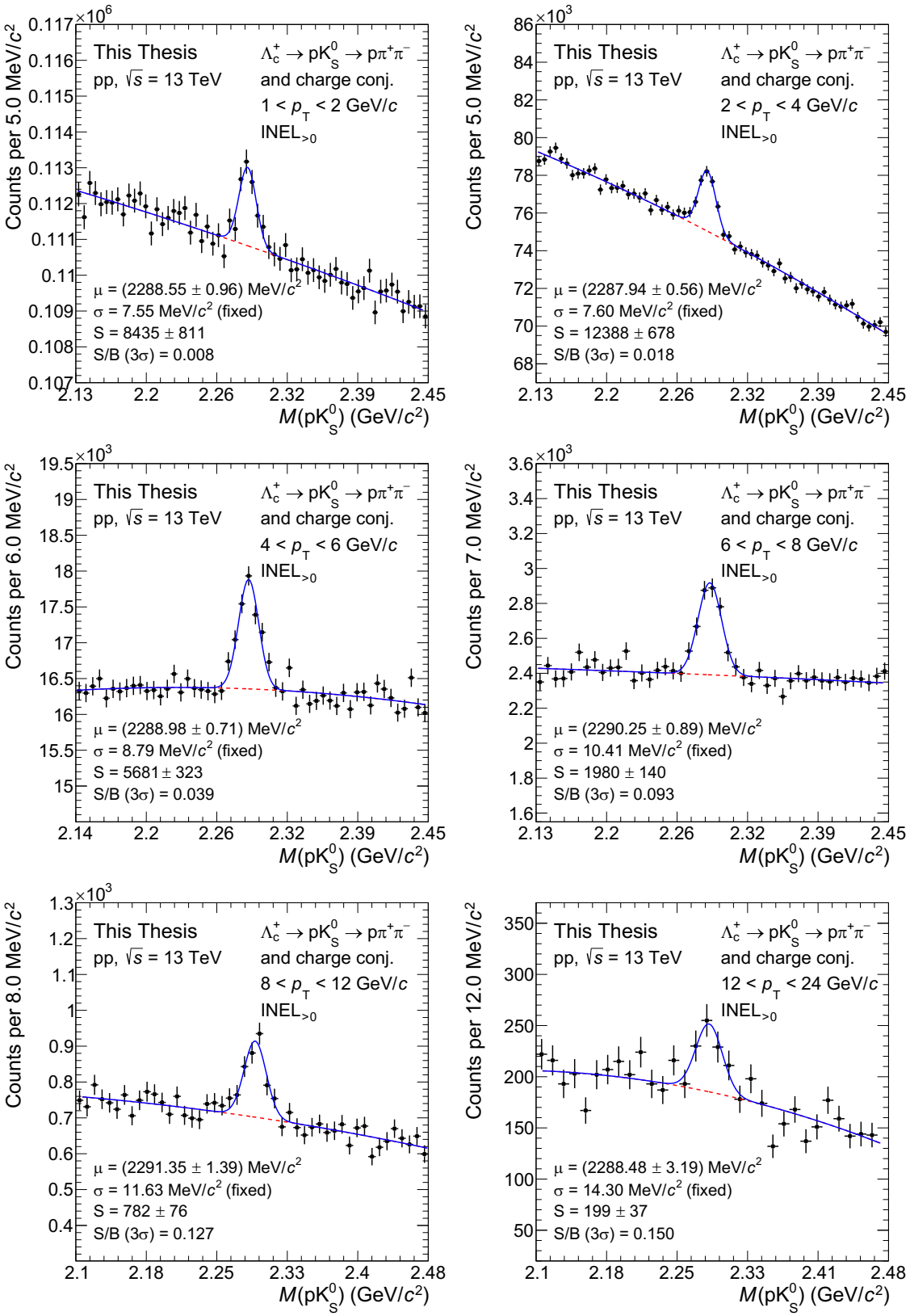

Figure 4.35: Invariant mass spectra for the Λ_c baryon selected in the decay chain $\Lambda_c^+ \to pK_s^0 \to p\pi^+\pi^-$ for MB triggered data collected with the MB trigger.

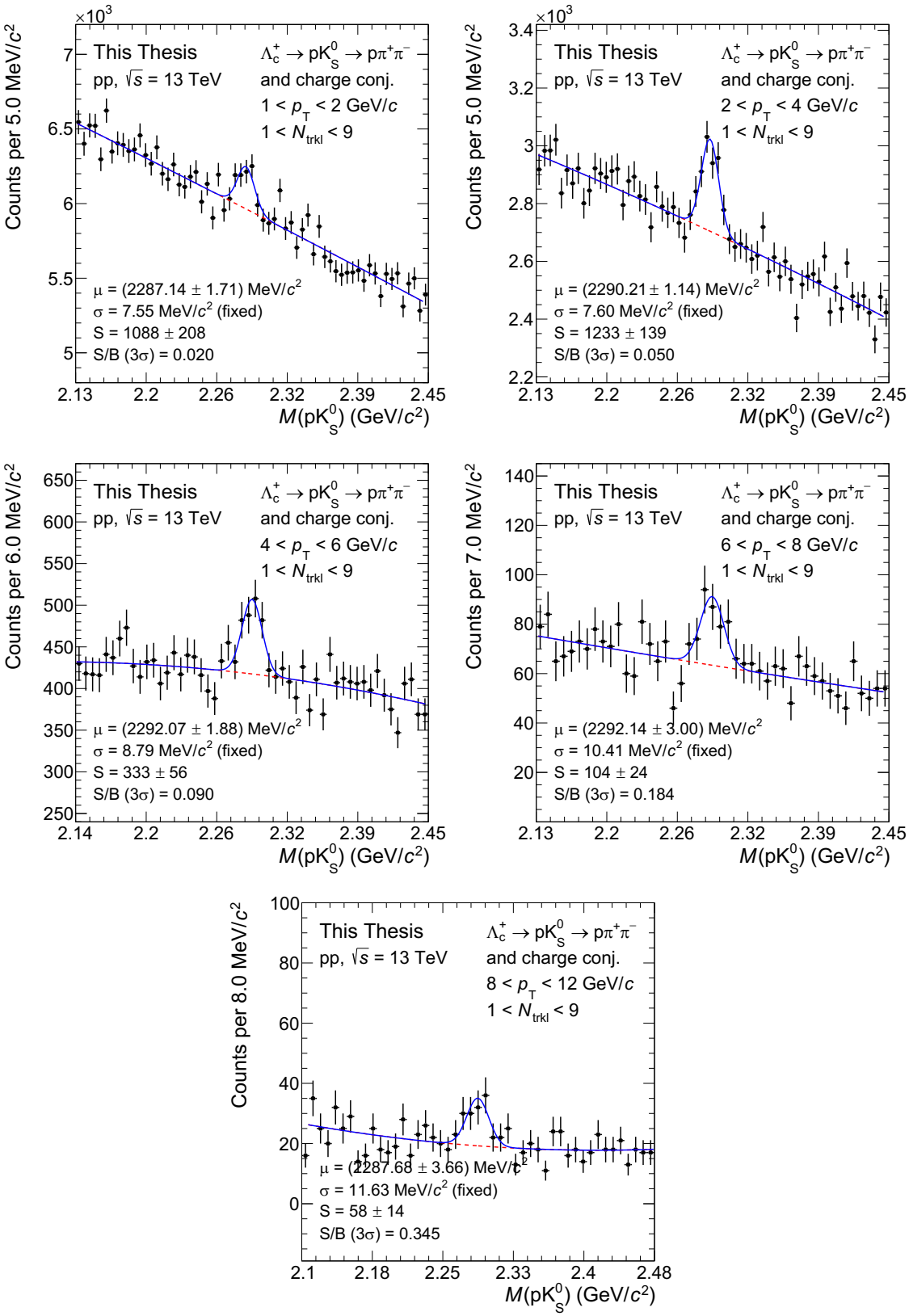

Figure 4.36: Invariant mass spectra for the Λ_c baryon selected in the decay chain $\Lambda_c^+ \to pK_s^0 \to p\pi^+\pi^-$ in the multiplicity class [1-9] collected with the MB trigger.

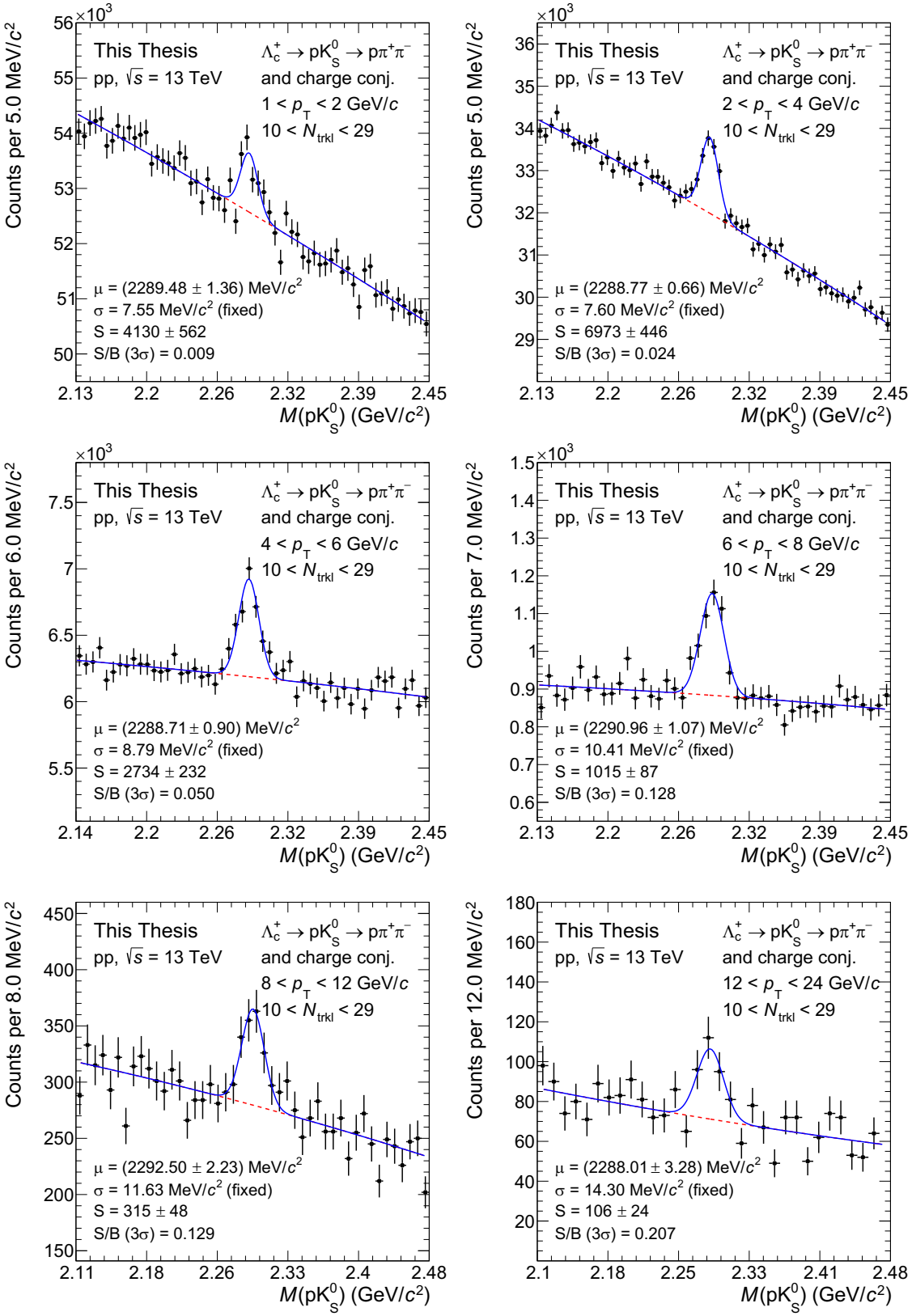

Figure 4.37: Invariant mass spectra for the Λ_c baryon selected in the decay chain $\Lambda_c^+ \to pK_s^0 \to p\pi^+\pi^-$ in the multiplicity class [10-29] collected with the MB trigger.

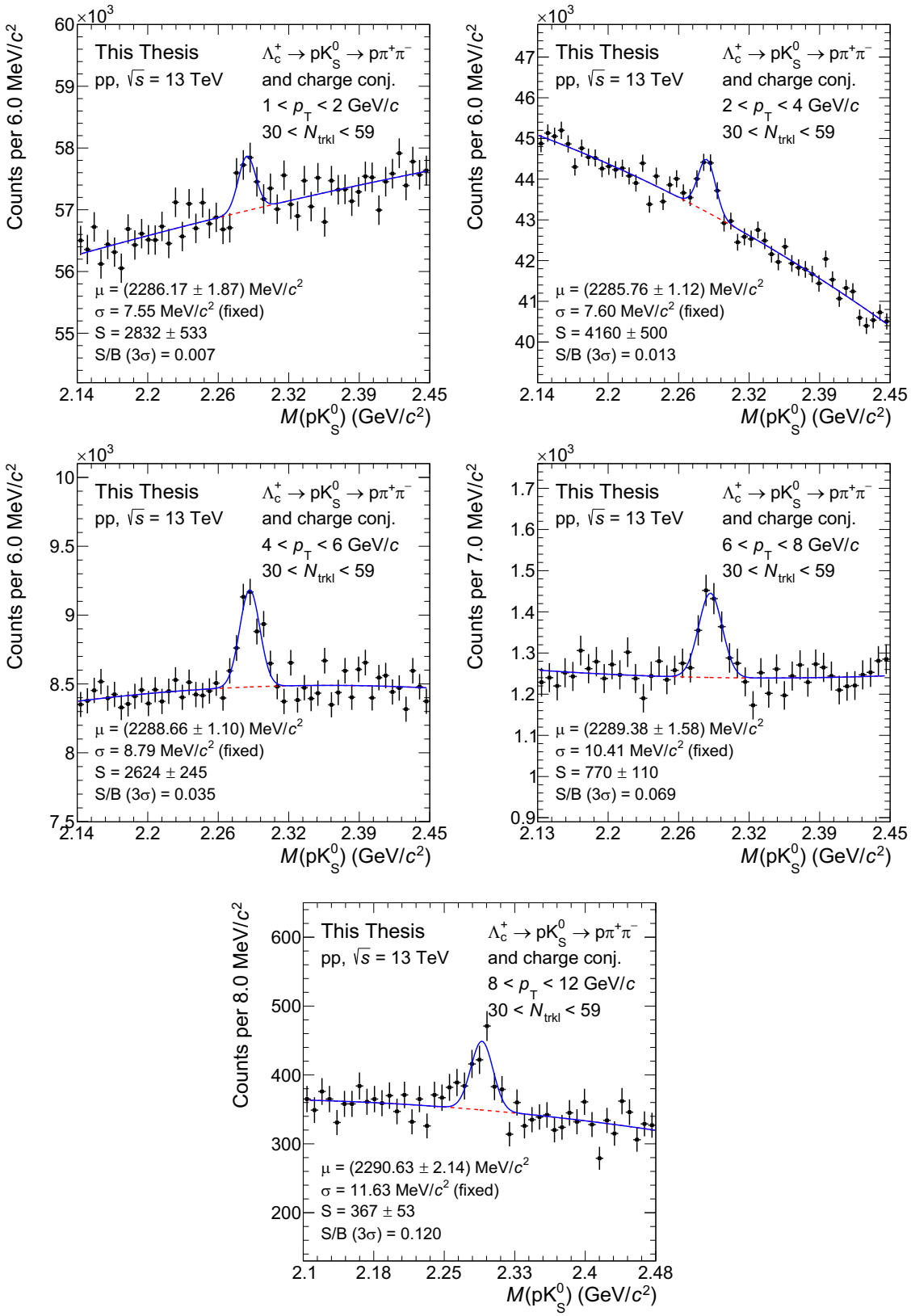

Figure 4.38: Invariant mass spectra for the Λ_c baryon selected in the decay chain $\Lambda_c^+ \rightarrow pK_s^0 \rightarrow p\pi^+\pi^-$ in the multiplicity class [30-59] collected with the MB trigger.

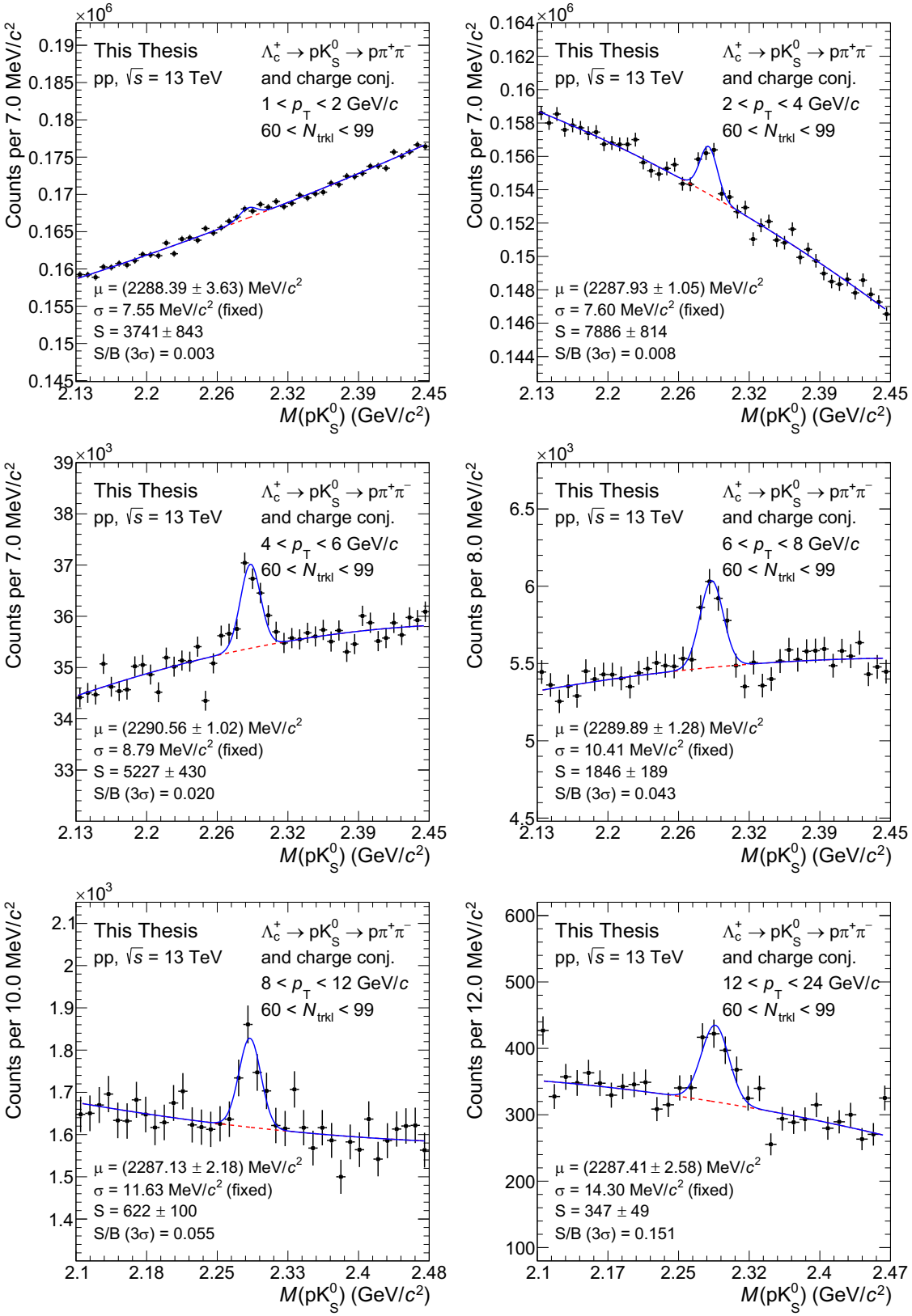

Figure 4.39: Invariant mass spectra for the Λ_c baryon selected in the decay chain $\Lambda_c^+ \to pK_s^0 \to p\pi^+\pi^-$ in the multiplicity class [60-99] collected with the HMSPD trigger.

The extracted raw yields are summarized in Tab. 4.10.

	p_{T} (Λ_{c}) (GeV/c)					
Mult.	1-2	2-4	4-6	6-8	8-12	12-24
MB	8435 ± 811	12388 ± 678	5681 ± 323	1980 ± 140	782 ± 76	199 ± 37
1-9	1088 ± 208	1233 ± 139	333 ± 56	104 ± 24	58 ± 14	/
10-29	4130 ± 562	6973 ± 446	2734 ± 232	1015 ± 87	315 ± 48	106 ± 24
30-59	2832 ± 533	4160 ± 500	2464 ± 245	770 ± 110	367 ± 53	/
60-99	3741 ± 843	7886 ± 814	5227 ± 430	1846 ± 189	622 ± 100	347 ± 49

Table 4.10: Summary of the extracted raw yield in each p_{T} and multiplicity class.

4.6 Correction

To retrieve the number of produced prompt Λ_{c} at midrapidity ($|y| < 0.5$), the measured Λ_{c} raw yield needs to be corrected for two different factors:

- the Acceptance $\times$ Efficiency;
- the feed-down subtraction.

The two correction factors are described in detail in the following subsections.

4.6.1 Acceptance $\times$ Efficiency correction

The raw counts of Λ_{c} baryons, extracted in each p_{T} and multiplicity interval after all the selections applied, were corrected for the detector acceptance and the reconstruction and selection efficiency. The acceptance factor allows one to correct for the Λ_{c} generated outside the geometrical acceptance of the detector and it is defined as the ratio between the Λ_{c} produced in the fiducial acceptance, i.e. $|y| < y_{\mathrm{fid}}(p_{\mathrm{T}})$, and the Λ_{c} generated in $|y| < 0.5$. Only events with the z coordinate of the generated primary-vertex position within $\pm$ 10 cm from the nominal IP ($|z_{\mathrm{vtx}}| < 10$ cm) were considered in the computation. The efficiency, ϵ, corrects the reconstructed yield

for the Λ_c fraction that is rejected because of the selection done (single track cuts, prefiltering selections on kinematical, topological and PID variables, BDT cut). It is defined as the ratio between the Λ_c surviving the cut selections and the Λ_c generated in the fiducial acceptance. The total correction factor, (Acc $\times$ ϵ), is thus defined as:

$$(\mathrm{Acc} \times \epsilon)_{\mathrm{mult}}(p_{\mathrm{T}}) = \frac{\Lambda_{\mathrm{creco,mult}}(p_{\mathrm{T}})}{\Lambda_{\mathrm{cgen,mult}}(p_{\mathrm{T}})\ |_{|y|<0.5}} \tag{4.17}$$

The (Acc $\times$ ϵ) was determined using Monte Carlo simulations of pp collisions generated with the `PYTHIA` 8 with Colour Reconnection Beyond Leading Colour (CR-BLC) Mode 2 tune [42]. The simulations were configured with a detailed description of the ALICE apparatus geometry and detector response. They were tuned to reproduce the position and width of the interaction vertex distribution, the number of active electronic channels, noise level and the accuracy of the detector calibration, as well as their time evolution within the pp data taking periods considered for the analysis. In order to improve the precision of the computation, each simulated event is required to contain either a $c\bar{c}$ or $b\bar{b}$ pair and the Λ_c baryons were forced to decay in the hadronic channels of interest for the analysis ($\Lambda_c^+ \rightarrow \mathrm{pK}_s^0 \rightarrow \mathrm{p}\pi^+\pi^-$). The particles were propagated through the apparatus using the `GEANT3` transport code [63].

The reconstruction and selection efficiency might depend on the multiplicity of the charged particles produced in the collision. Indeed, the primary vertex resolution improves at high multiplicity; as a consequence, the resolution of the selection variables that make use of the primary vertex position are expected to improve with increasing multiplicity, introducing a multiplicity dependence in the selection efficiency. However, as explained in sec. 4.3, in this analysis only one training variable that depends on the primary vertex position has been used, i.e. the impact parameter of the bachelor candidate, b_p. This could reflect in a weak dependence of the reconstruction efficiency versus multiplicity. An in-depth study of the multiplicity dependence was performed plotting the efficiency versus the number of tracklets. The results are reported for both prompt and feed-down efficiency, respectively in

Figs. 4.40, 4.41, 4.42 and Figs. 4.43, 4.44, 4.45.

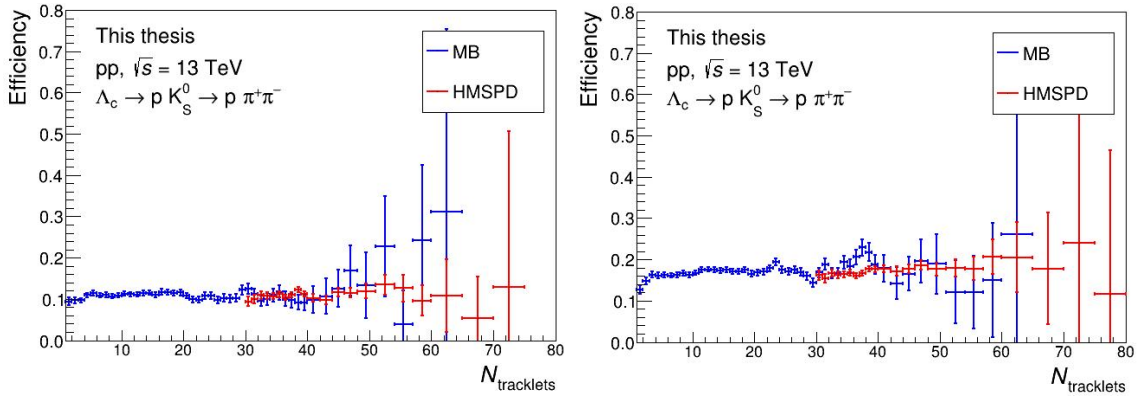

Figure 4.40: Prompt Λ_c efficiency vs the $N_{\text{tracklets}}$ in the $1 \leq p_T < 2$ GeV/c (left) and $2 \leq p_T < 4$ GeV/c (right) intervals for the MB (blue) and HM (red) MC productions.

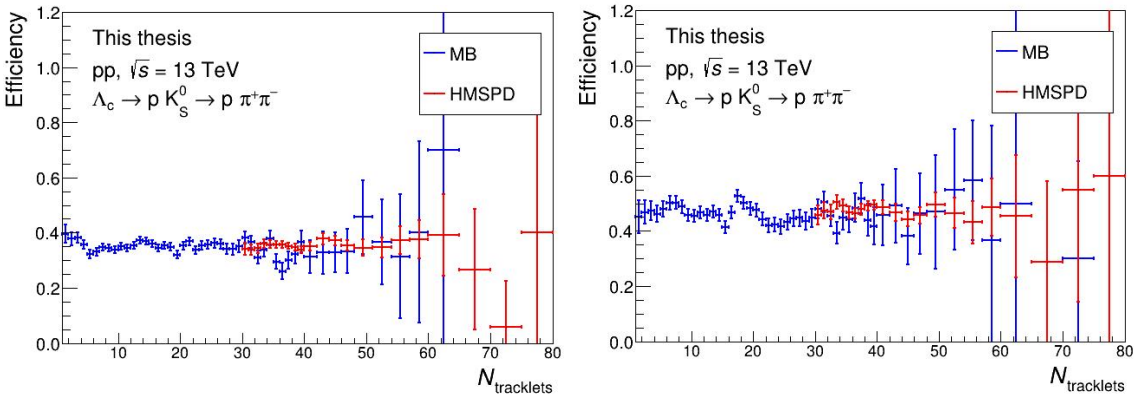

Figure 4.41: Prompt Λ_c efficiency vs the $N_{\text{tracklets}}$ in the $4 \leq p_T < 6$ GeV/c (left) and $6 \leq p_T < 8$ GeV/c (right) intervals for the MB (blue) and HM (red) MC productions.

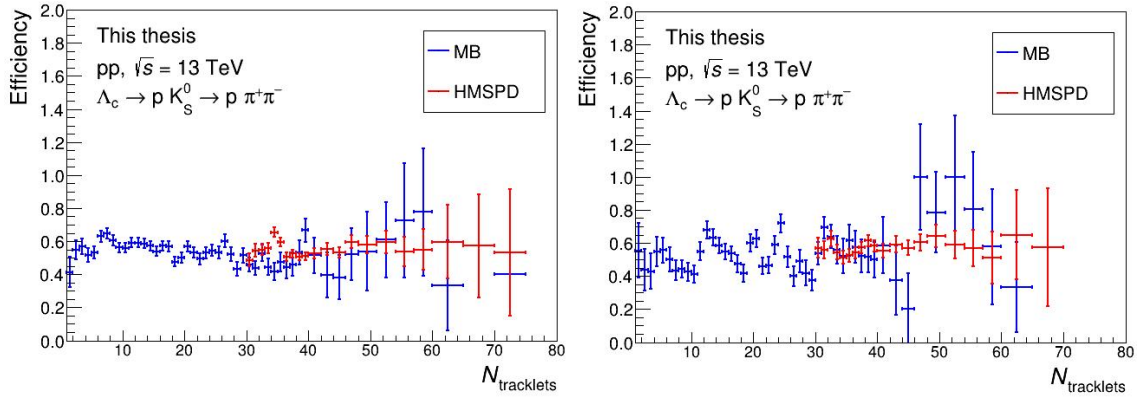

Figure 4.42: Prompt Λ_c efficiency vs the $N_{\text{tracklets}}$ in the $8 \leq p_T < 12$ GeV/c (left) and $12 \leq p_T < 24$ GeV/c (right) intervals for the MB (blue) and HM (red) MC productions.

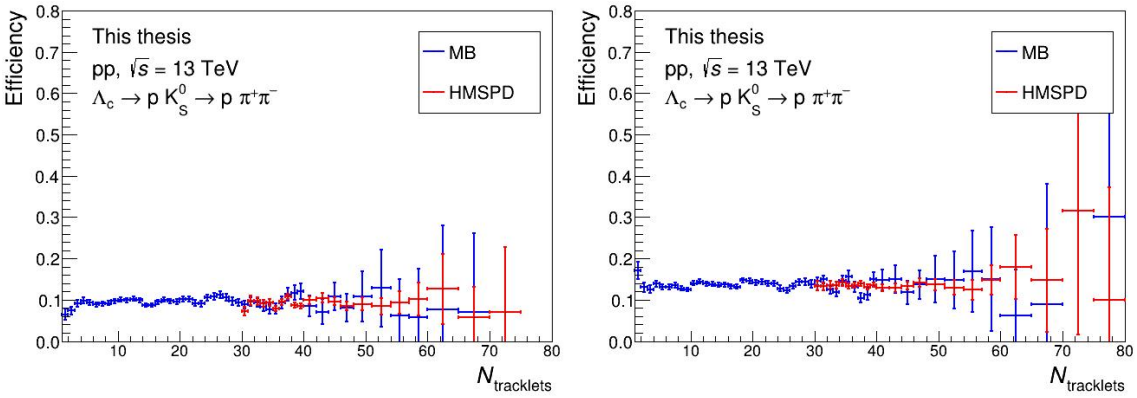

Figure 4.43: Feed-down Λ_c efficiency vs the $N_{\text{tracklets}}$ in the $1 \leq p_T < 2$ GeV/c (left) and $2 \leq p_T < 4$ GeV/c (right) intervals for the MB (blue) and HM (red) MC productions.

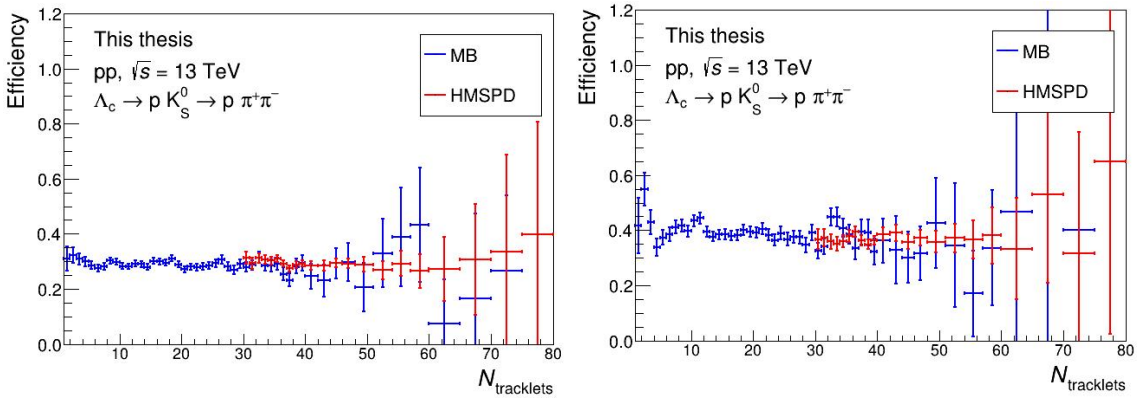

Figure 4.44: Feed-down Λ_c efficiency vs the $N_{\text{tracklets}}$ in the $4 \leq p_T < 6$ GeV/c (left) and $6 \leq p_T < 8$ GeV/c (right) intervals for the MB (blue) and HM (red) MC productions.

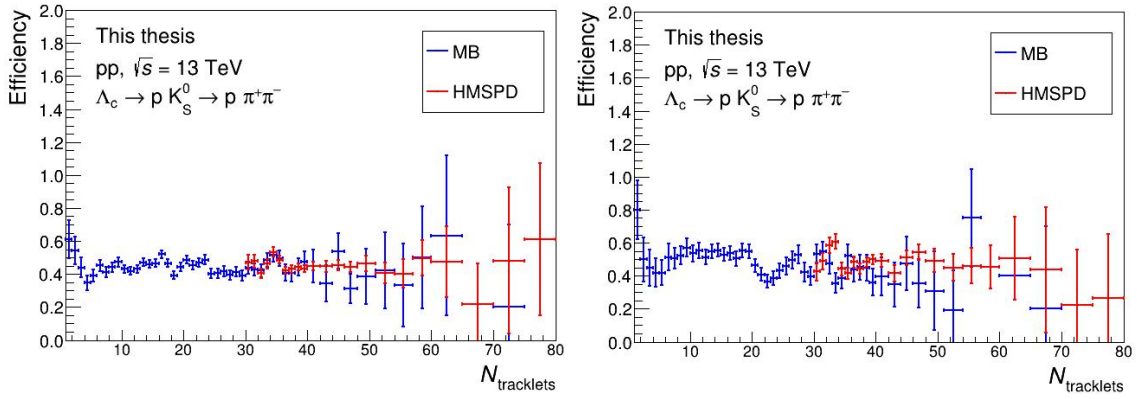

Figure 4.45: Feed-down Λ_c efficiency vs the $N_{\text{tracklets}}$ in the $8 \leq p_T < 12$ GeV/c (left) and $12 \leq p_T < 24$ GeV/c (right) intervals for the MB (blue) and HM (red) MC productions.

The efficiency trend is almost flat in all Λ_c p_T bins for both MB and HM MC productions. A very mild dependence is observed at the very low multiplicity, for $N_{\text{tracklets}}$ <4. So, results confirm the expectation for the higher multiplicity bins, i.e. [10,29], [30,59] and [60,99]. Concerning the first multiplicity class, [1-9], in which a multiplicity dependence of the efficiency is observed, it is important, for a proper efficiency estimation, to verify whether the multiplicity distributions in the Monte Carlo samples reproduce the ones in data. In the left panel of Fig 4.46 the comparison between the two samples is shown separately for each year; it can be observed that the barrel multiplicity is not sufficiently well reproduced in the MC simulations. The $N_{\text{tracklets}}$ distributions in each MC sample are then re-weighted to the distribution from data. The MC weights are generally obtained per year according to the following steps:

- the full event selection was applied on both MC and data;
- only the events in MC and data with at least one candidate were considered. All the candidate selection criteria were applied and the candidate invariant mass was required to be at most ± 20 MeV/c^2 off the PDG value;
- the normalised $N_{\text{tracklets}}$ distribution was extracted for MC and data;
- the obtained distributions from data were divided by the MC ones to get the event weights, ω_i, in each barrel multiplicity bin, i.

The computed weights ω_i for the three MB and the HM productions are represented in the right panel of Fig 4.46. The re-weighted efficiencies for the lower multiplicity class were estimated as:

$$(\text{Acc} \times \epsilon)(p_T)_{corr} = \frac{\sum_i \Lambda_{\text{creco}}\, \omega_i}{\sum_i \Lambda_{\text{cgen}}\, \omega_i} \tag{4.18}$$

where i represents the i-th multiplicity bin.

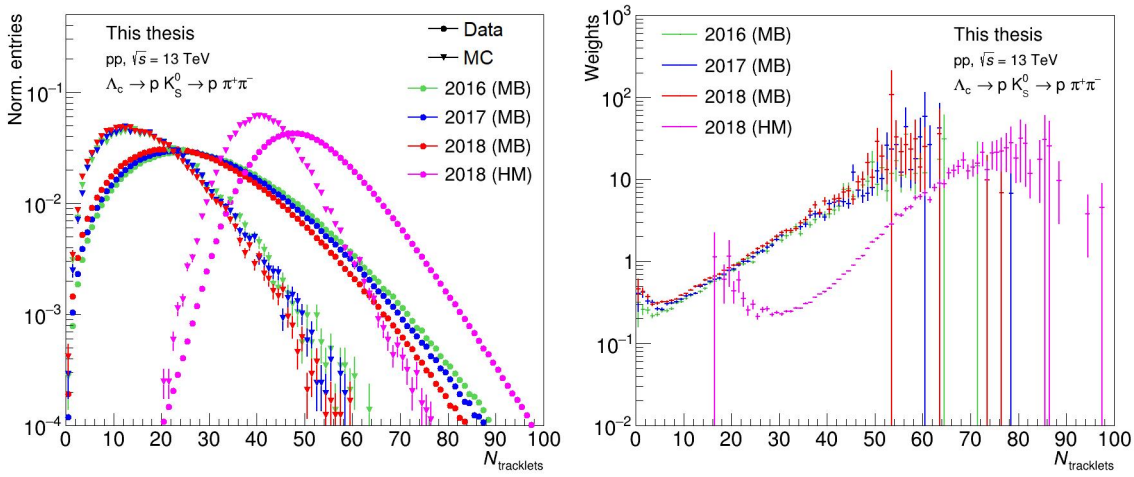

Figure 4.46: Left: Data and MC distributions for the MB and HM samples. Right: Multiplicity weights computes as the ratio of $N_{\text{tracklets}}$ distributions in data and MC for all three years.

The final efficiencies, computed following the procedure described above, are reported in Fig 4.47 for Λ_c prompt (left) and feed-down (right). The relative statistical uncertainty of the efficiency is assigned as systematic on the final corrected yield measurement.

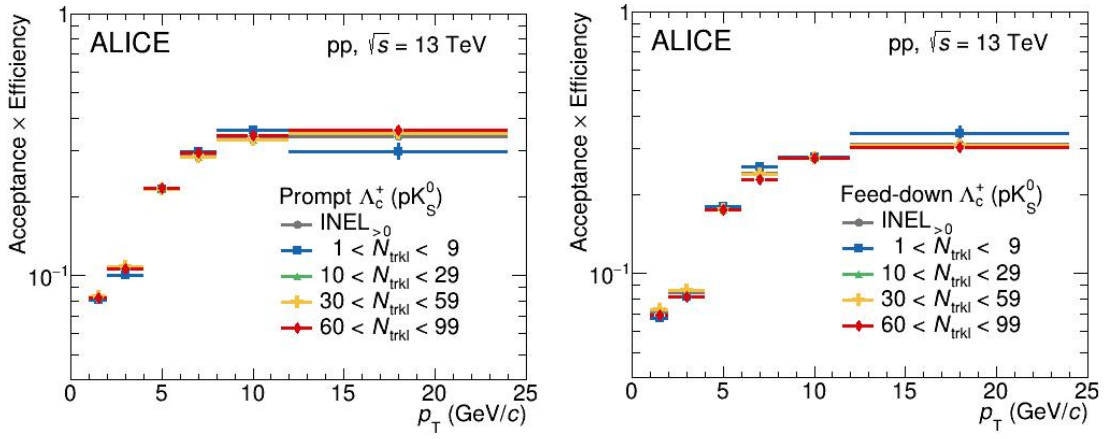

Figure 4.47: Acceptance × efficiency of prompt (left) and feed-down (right) Λ_c baryons in the analysed multiplicity classes.

The (Acc × ϵ) varies as a function of p_{T} because of the track-reconstruction efficiency and the topological selections applied. In addition, it is in general different for prompt and feed-down Λ_c baryons: in particular, the feed-down Λ_c are more

displaced w.r.t. the primary vertex and so the impact parameter of the bachelor candidate b_p assumes higher values for Λ_c from beauty-hadron decays than for Λ_c from charm hadronisation. The BDT cut can thus reject part of the feed-down yield with a selection on this variable, lowering more Λ_c feed-down efficiency than the prompt one.

4.6.2 Feed-down subtraction

The prompt Λ_c production yield in pp collisions were obtained by subtracting the contribution of Λ_c from beauty-hadron decays from the inclusive raw yield. In the absence of a measured beauty-hadron production cross section at central rapidity, we use perturbative QCD calculations for the production of beauty hadron and detailed Monte Carlo simulations to estimate the fraction of Λ_c stemming from the decay of beauty hadrons in our raw yields. In detail, the Λ_c-baryon feed-down contribution was estimated with the so-called N_b method:

$$\begin{aligned} f_{\text{prompt}} &= 1 - \frac{N^{\Lambda_c \text{ feed-down, raw}}}{N^{\Lambda_c \text{ raw}}} = \\ &= 1 - \left(\frac{\mathrm{d}^2\sigma}{\mathrm{d}y\,\mathrm{d}p_{\mathrm{T}}}\right)^{\texttt{FONLL+PYTHIA8}}_{\text{feed-down}} \cdot \frac{(\text{Acc} \times \epsilon)_{\text{feed-down}} \cdot \Delta y\, \Delta p_{\mathrm{T}} \cdot \mathrm{BR_b} \cdot \mathcal{L}_{\text{int}}}{N^{\Lambda_c \text{ raw}}/2} \end{aligned} \tag{4.19}$$

where :

- $N^{\Lambda_c \text{ raw}}/2$ is the raw yield divided by a factor of two to account for particles and antiparticles;
- $\left(\frac{\mathrm{d}^2\sigma}{\mathrm{d}y\,\mathrm{d}p_{\mathrm{T}}}\right)^{\texttt{FONLL+PYTHIA8}}_{\text{feed-down}}$ is the production cross section of Λ_c from Λ_b-baryon decays, calculated using the beauty production cross section from the `FONLL` calculation [34], the fraction of beauty quarks that fragment into Λ_b estimated from LHCb measurements [72] and the $\mathrm{H}_b \to \Lambda_c + \mathrm{X}$ decay kinematics from `PYTHIA` 8 simulations [73];
- $(\text{Acc} \times \epsilon)_{\text{feed-down}}$ is the acceptance times efficiency correction for feed-down

Λ_c, computed with Monte Carlo simulation;

- BR_b is the branching ratio for a Λ_b decay in Λ_c + X.

The FONLL calculations are multiplicity independent and so the N_b method is in principle not very suitable for a multiplicity analysis. However, considering that the feed-down fraction is not expected to be (heavily) dependent on the multiplicity and the feed-down efficiencies do not vary significantly between the different multiplicity bins, the $f_{\rm prompt}$ fraction for the multiplicity classes is taken from the one extracted in the multiplicity integrated analysis. A systematic source has been added to take into account eventual dependency by the multiplicity of the $f_{\rm prompt}$ and it is discussed in section 4.8.3. The Λ_c $f_{\rm prompt}$ obtained with the N_b method is shown in Fig 4.48.

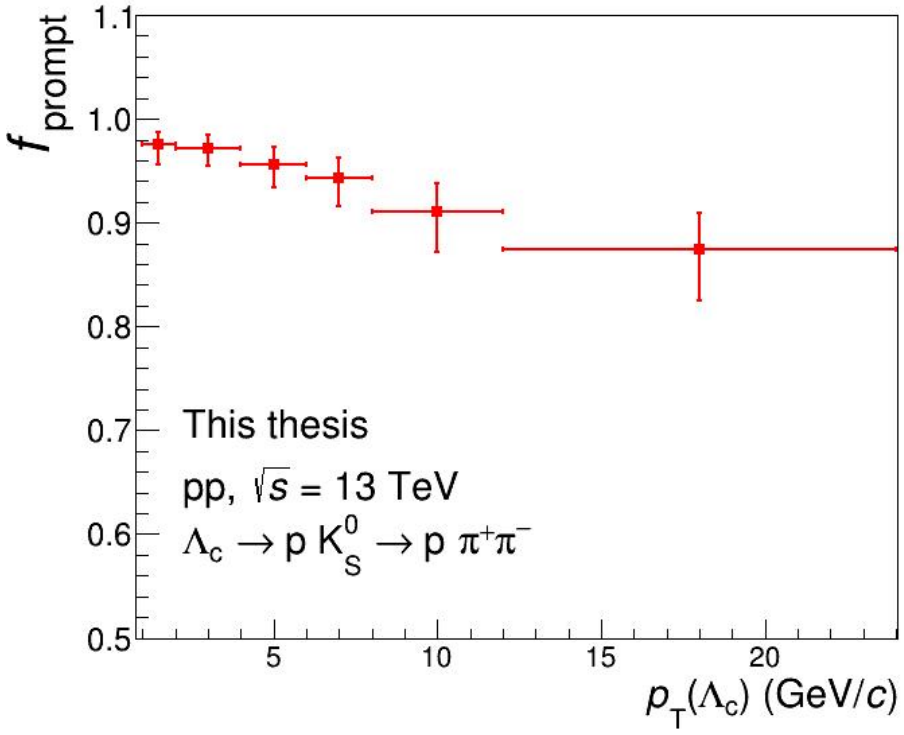

Figure 4.48: Λ_c prompt fraction computed with the N_b method.

The Λ_c $f_{\rm prompt}$ calculated with this method is around 0.98 in the lowest $p_{\rm T}$ bin; it decreases as a function of $p_{\rm T}$ till 0.88 in the highest $p_{\rm T}$ interval, corresponding to a non-prompt fraction of the raw yield ranging in [2 - 12]%.

4.7 Corrected Yields

The corrected per-event yields were computed for each p_{T} and multiplicity interval as:

$$\frac{1}{N_{\mathrm{mult}}^{\mathrm{ev}}}\frac{\mathrm{d}N_{\mathrm{mult}}^{\Lambda_{\mathrm{c}}^{+}}}{\mathrm{d}p_{\mathrm{T}}} = \frac{\epsilon_{\mathrm{mult}}^{\mathrm{INEL}}}{N_{\mathrm{mult}}^{\mathrm{ev}}}\frac{1}{c_{\Delta y}(p_{\mathrm{T}})\cdot\Delta p_{\mathrm{T}}}\frac{1}{\mathrm{BR}}\frac{f_{\mathrm{prompt}}(p_{\mathrm{T}})\cdot\frac{1}{2}\cdot N_{\mathrm{mult}}^{\Lambda_{\mathrm{c}},\mathrm{raw}}(p_{\mathrm{T}})\Big|_{|y|<y_{\mathrm{fid}}(p_{\mathrm{T}})}}{(\mathrm{Acc}\times\epsilon)_{\mathrm{prompt,mult}}(p_{\mathrm{T}})} \tag{4.20}$$

where :

- $N_{\mathrm{mult}}^{\Lambda_{\mathrm{c}},\mathrm{raw}}$ is the raw yield (sum of particles and antiparticles) extracted in a given p_{T} and multiplicity interval;
- f_{prompt} is the Λ_{c} prompt fraction that allows one to correct for the corresponding beauty-hadron decay contribution;
- $(\mathrm{Acc}\times\epsilon)_{\mathrm{prompt,mult}}$ is the multiplicity-dependent prompt acceptance-times-efficiency correction;
- the factor 1/2 is needed to obtain the charge-averaged yield;
- BR is the branching ratio of the decay channel under study;
- Δp_{T} is the p_{T} interval width;
- $c_{\Delta y}$ is the correction for the rapidity coverage, computed as the ratio between the generated hadron yield in $\Delta y = 2y_{\mathrm{fid}}$ and that in $|y| < 0.5$;
- $N_{\mathrm{mult}}^{\mathrm{ev}}$ denotes the number of recorded events in each multiplicity class;
- the number of recorded events needs to be corrected for the trigger efficiency, $\epsilon_{\mathrm{mult}}^{\mathrm{INEL}}$, defined as the fraction of INEL > 0 events that were not selected by the trigger.

4.8 Systematic sources estimation

The evaluation of the systematic uncertainties on the corrected yield measurements, that arise from uncertainties on the different ingredients in (4.20), are described in this section. The following systematic uncertainties have been considered in the analysis:

- **raw yield extraction**;
- **cut variation**;
- **tracking efficiency**;
- **feed-down subtraction**;
- **trigger correction**;
- **MC $p_{\rm T}$ shape**;
- **$z_{\rm vtx}$ position**;
- **branching ratio**.

Each uncertainty source is assumed to be uncorrelated with all other sources so the final uncertainty is determined summing in quadrature all systematic uncertainties. The relative systematic uncertainties for each source are summarised at the end of this section.

4.8.1 Raw yield extraction

The raw yield extraction systematic uncertainty is related to the fit procedure that could depend on the chosen values for the fit parameters. The yield extraction systematic was then evaluated with a multi-trial approach, i.e. the fit to each invariant-mass distribution is repeated many times varying: the invariant mass fit range (five different lower and five different upper limits), the functional form of the combinatorial background (linear, second and third order polynomials and

exponential), the function configuration for the signal (leaving the peak position and the peak width as a free parameter, fixing it/them to the values extracted from the MC simulation). All the possible combinations of the aforementioned fit configurations were considered. The performed trials are summarized in Tab. 4.11.

Sigma, Mean configs.	fixed sigma & free mean fixed sigma & fixed mean sigma +15% & free mean sigma −15% & free mean
Lower bound fit range	$\{2.14, 2.13, 2.12, 2.15, 2.16\}$
Upper bound fit range	$\{2.436, 2.426, 2.416, 2.446, 2.456\}$
Background functions	{ Exponential and first, second, third order polynomial }

Table 4.11: Fit parameter variations employed to compute the $\Lambda_c^+ \to pK_s^0 \to p\pi^+\pi^-$ raw-yield extraction systematic uncertainty.

As additional stability check, the results were compared to the ones obtained with bin-counting methods. This last approach consists in integrating the invariant-mass distribution within $\pm 3\sigma$ and $\pm 5\sigma$, after subtracting the combinatorial background estimated, respectively, from a fit to the side-band distributions (green trials in the bottom panels of Fig 4.49) and from the total fit function (brown trials in the bottom panels of Fig 4.49). This test allows having an estimation of the raw yield independent of the signal shape chosen for the fit. The systematic uncertainty was then evaluated by considering the RMS of the trials plus the shift with respect to the raw-yield obtained with the nominal fit parameters configuration. Only the fits having $\chi^2/\mathrm{ndf} < 2$ and a significance > 3 have been considered in the computation. In Fig. 4.49 an example of raw-yield distribution for the interval $1 \leq p_{\mathrm{T}} < 2$ GeV/c in the multiplicity integrated class, evaluated with the multi-trial study from the fit and the bin-counting method, is shown. The same procedure has been used for each multiplicity and p_{T} interval and the estimated raw-yield extraction systematic uncertainties are reported in Tab. 4.12.

Systematic on raw yield extraction [%]						
Mult. \ p_T	1 - 2	2 - 4	4 - 6	6 - 8	8 - 12	12 - 24
MB	3	4	4	4	4	8
1-9	4	5	5	6	5	-
10-29	3	4	4	4	4	8
30-59	5	6	6	6	6	-
60-99	7	6	5	5	5	8

Table 4.12: Systematic uncertainties associated to raw yield extraction for $\Lambda_c^+ \to pK_s^0 \to p\pi^+\pi^-$ estimated for each p_T and multiplicity interval.

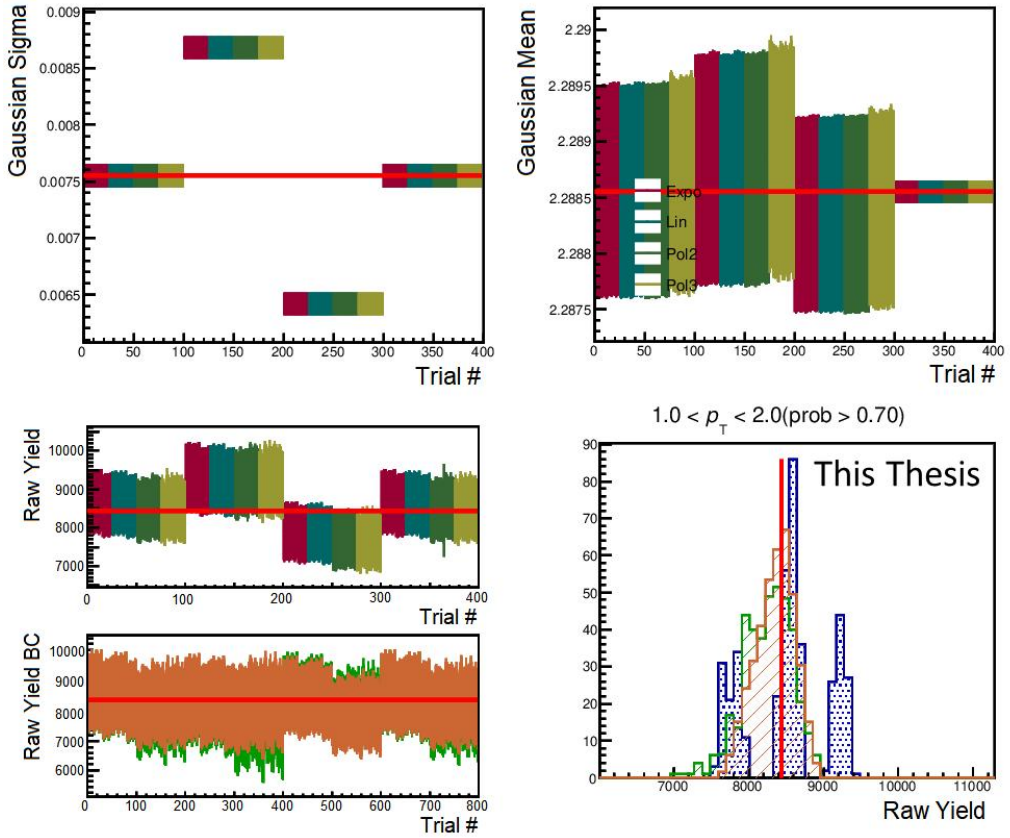

Figure 4.49: Output of the MultiTrial fitter for the $\Lambda_c^+ \to pK_s^0 \to p\pi^+\pi^-$ decay for the bin $1 \leq p_T < 2$ GeV/c in the multiplicity integrated case. It shows the gaussian mean values (top right), gaussian sigma values (top left), the extracted raw yield from fit and bin-counting (bottom left) and the histogram of the raw yield values (bottom right).

4.8.2 Cut Variation

The data has been corrected for the detector acceptances and efficiencies, which are determined from Monte Carlo simulations. Differences between data and MC in the distribution of one or more of the variables used in the candidates selection could introduce biases in the extraction of the corrected yields. To quantify the effect and to estimate a corresponding systematic uncertainty, the corrected yield were obtained using different BDT working points. Twenty looser and twenty tighter equidistantly BDT cut were scanned, varying the efficiency till ±25% around the nominal one. In order to reduce as much as possible the contribution from statistical fluctuations and to disentangle this systematic source to the raw yield extraction one, the gaussian mean and width were fixed for all the trial to the value extracted in the central fit. The full analysis was repeated for each of the chosen working points. The trials that do not provide a good Λ_c signal extraction were rejected requiring a significance $>$ 3. The ratios with respect to the nominal values were computed and the assigned systematic is based on the RMS, computed w.r.t. unity, of the variations. The Λ_c raw yield and efficiency ratios are reported, respectively, on the left and right panel of Fig. 4.50 while the variation on the Λ_c corrected yield is shown in Fig. 4.51 for the multiplicity integrated class.

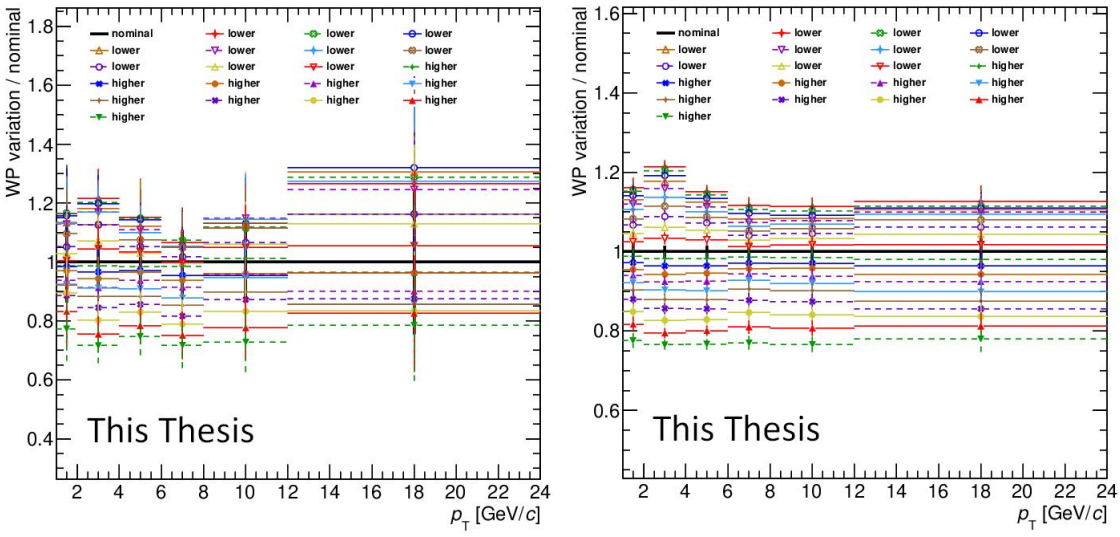

Figure 4.50: Ratio of Λ_c raw yield (left) and prompt efficiency (right) extracted with different BDT cuts w.r.t. the central value for the multiplicity integrated class.

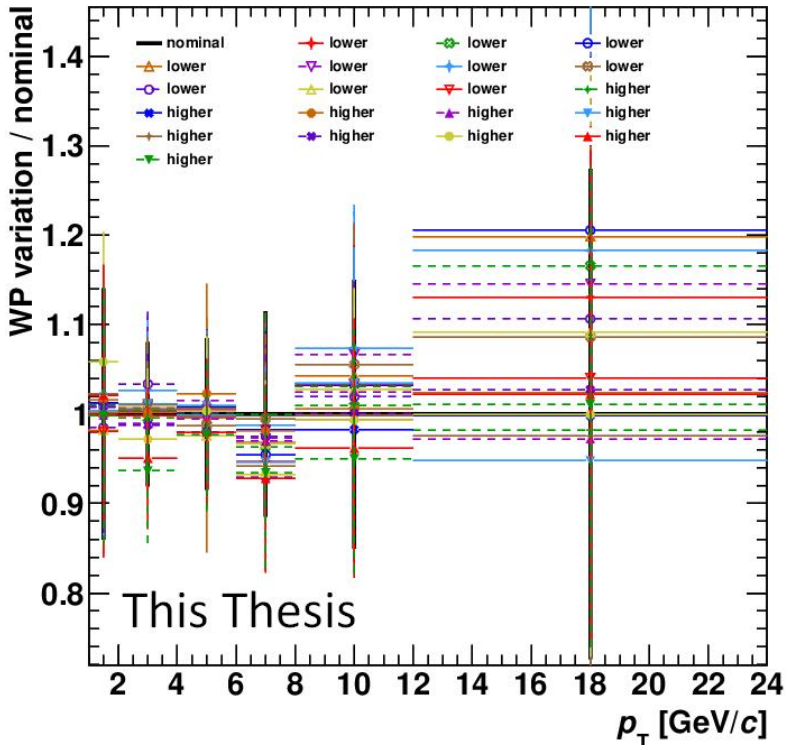

Figure 4.51: Ratio of corrected yields extracted with different BDT cuts w.r.t. the central value for the multiplicity integrated class [1-9999] $N_{\text{tracklets}}$.

The procedure was performed for each multiplicity class and the final cut variation systematic uncertainties, obtained after a smoothing, are summarised in Tab. 4.13.

Systematic on cut variation [%]						
p_T / Mult.	1 - 2	2 - 4	4 - 6	6 - 8	8 - 12	12 - 24
MB	6	6	6	6	6	6
1-9	6	6	6	6	6	-
10-29	6	6	6	6	6	6
30-59	6	6	6	6	6	-
60-99	6	6	6	6	6	6

Table 4.13: Systematic uncertainties associated to the BDT cut variation for the $\Lambda_c^+ \to pK_s^0 \to p\pi^+\pi^-$ in the analysed multiplicity ranges.

4.8.3 Feed-down subtraction

The Λ_c prompt fraction was computed with the N_b method, based on FONLL predictions. In addition, the assumption that the relative contribution of beauty hadron decays to the Λ_c yield is multiplicity independent and it is equal to that in the multiplicity integrated sample was made. For this reason, two different feed-down subtraction systematics were quoted: the first one is related to the FONLL theoretical uncertainties while the second one takes care of the possible multiplicity dependence of the Λ_c prompt fraction. The systematic uncertainty on the estimation of feed-down correction from beauty-hadrons was performed by varying the parameters used for the FONLL B predictions. In particular, the b quark mass and the factorisation and renormalisation scales in the FONLL calculations were varied. The envelope of these variations, corresponding to the asymmetric error bars in Fig. 4.48, was taken and assigned directly as systematic. The values are reported in Tab. 4.14.

Systematic on feed-down from FONLL [%]						
p_T interval	1 - 2	2 - 4	4 - 6	6 - 8	8 - 12	12 - 24
MB	+1 -2	+1 -2	+2 -2	+2 -3	+3 -4	+3 -4

Table 4.14: Systematic uncertainty associated to the feed-down from B subtraction, estimated from FONLL predictions.

To estimate the uncertainty related to the assumption of the f_{prompt} independence on multiplicity, PYTHIA simulations were employed. Different PYTHIA 8 tunes, i.e. Monash [35] and CR-BLC [42], were investigated. The Λ_c prompt fractions with the different tunes are reported in the left panel of Fig. 4.52. The Λ_c $f_{\text{feed-down}}$ was then estimated as 1-f_{prompt}. The ratios between the Λ_c $f_{\text{feed-down}}$ obtained with the different PYTHIA tunes and the one estimated with the N_b method on the multiplicity integrated sample $f_{\text{feed-down}}$(MB), shown in the right panel of Fig. 4.52, were used to define a set of upper and lower variations, dependent on the event multiplicity, for the $f_{\text{feed-down}}$(MB).

They are reported in Tab. 4.15 and shown in Fig. 4.53.

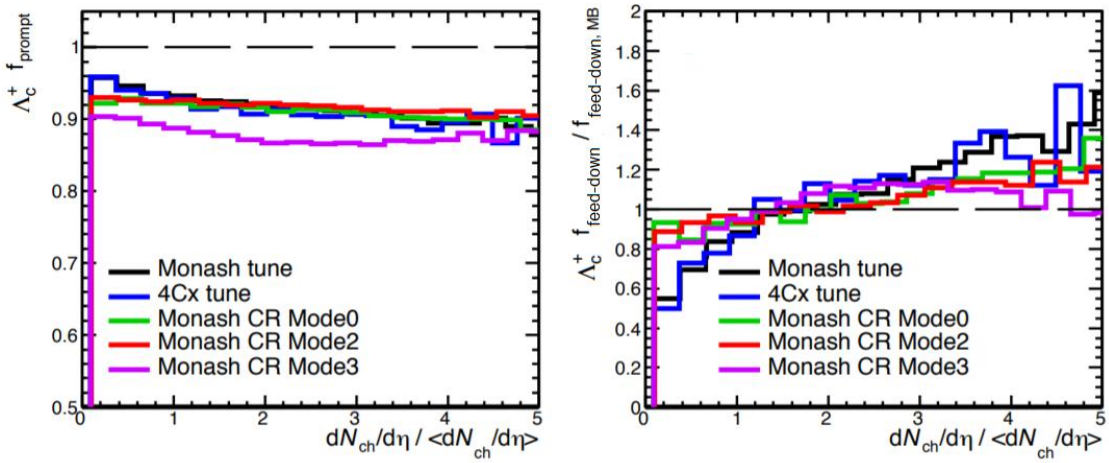

Figure 4.52: Left: Λ_c prompt fractions obtained with different PYTHIA tunes. Right: Ratios between the Λ_c $f_{\text{feed-down}}$ obtained with the different PYTHIA tunes and the central one estimated with the N_b method.

Multiplicity interval	$f_{\text{feed-down}}$ variation
1-9	[0.7 - 1.0]
10-29	[0.9 - 1.2]
30-59	[1.0 - 1.4]
60-99	[1.0 - 1.6]

Table 4.15: Variations on $f_{\text{feed-down}}$ applied to the central value to estimate the multiplicity dependent feed-down subtraction systematic uncertainty.

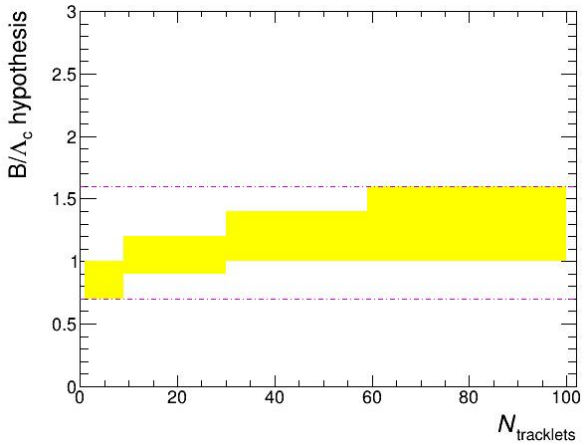

Figure 4.53: B/Λ_c variations applied to the central value to estimate the multiplicity dependent feed-down subtraction systematic uncertainty.

An additional feed-down systematic source was associated to each measured multiplicity class and the evaluation procedure consists in the following steps:

- The $f_{\mathrm{feed-down}}$ was estimated from the upper and lower values of the uncertainties on $f_{\mathrm{prompt}}(\mathrm{MB})$ as:

$$f_{\mathrm{feed-down}} = 1 - f_{\mathrm{prompt}}(\mathrm{MB})$$

- The central $f_{\mathrm{feed-down}}$ is varied in each multiplicity bin by the factors shown in Tab. 4.15 to get the minimum ($f_{\mathrm{feed-down,min}}$) and maximum ($f_{\mathrm{feed-down,max}}$) feed-down fraction expected.

$$f_{\mathrm{feed-down,min}} = f_{\mathrm{feed-down}} \cdot \mathrm{var}^{\mathrm{low}}$$
$$f_{\mathrm{feed-down,max}} = f_{\mathrm{feed-down}} \cdot \mathrm{var}^{\mathrm{up}}$$

- The feed-down fraction range is converted in the minimum and maximum f_{prompt} fraction:

$$f_{\mathrm{prompt,min}} = 1 - f_{\mathrm{feed-down,max}}$$
$$f_{\mathrm{prompt,max}} = 1 - f_{\mathrm{feed-down,min}}$$

- The f_{prompt} range is converted into a relative asymmetric uncertainty by computing the ratios $f_{\mathrm{prompt,min}}/f_{\mathrm{prompt}}(\mathrm{MB})$ and $f_{\mathrm{prompt,max}}/f_{\mathrm{prompt}}(\mathrm{MB})$.

The computed ratios are reported in Fig. 4.53 per each Λ_{c} p_{T} bin and the estimated multiplicity dependent feed-down systematic uncertainties are summarised in Tab. 4.16.

The procedure was repeated to estimate the multiplicity dependent uncertainty on the baryon-over-meson ratio. In this case the `FONLL` uncertainty part cancels out in the ratio and so, in the first step, only the central $f_{\mathrm{prompt}}(\mathrm{MB})$ was taken into account. The final feed-down subtraction systematic values for the baryon-over-meson ratio

are reported in Tab. 4.17.

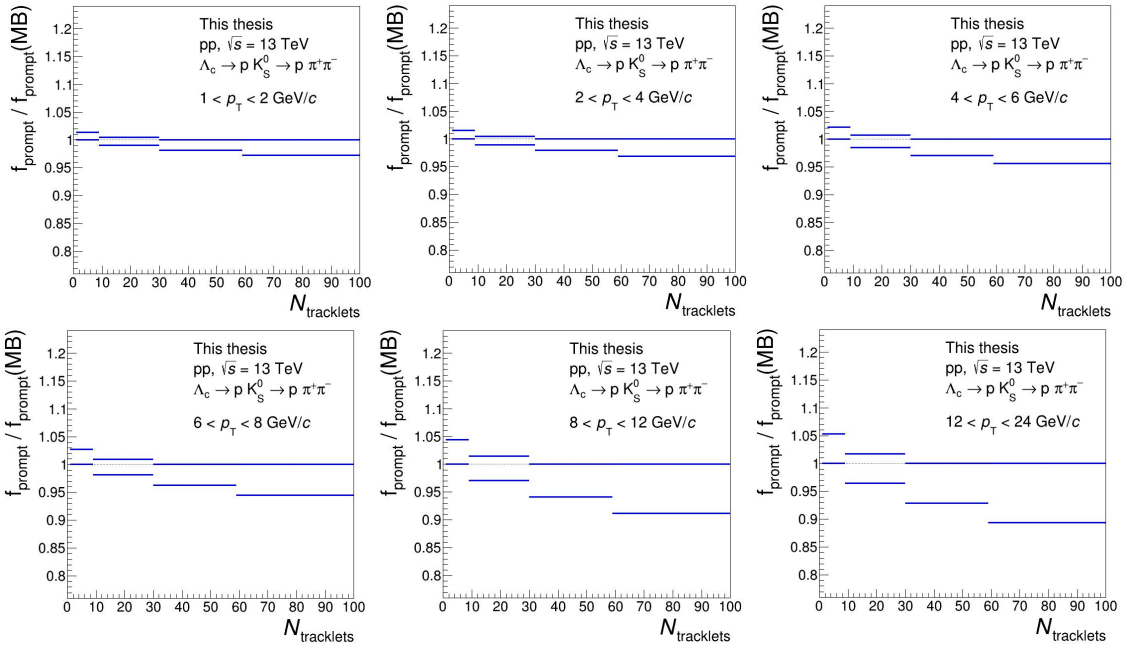

Figure 4.54: Λ_c $f_{\mathrm{prompt,min}}/f_{\mathrm{prompt}}(\mathrm{MB})$ and $f_{\mathrm{prompt,max}}/f_{\mathrm{prompt}}(\mathrm{MB})$ ratios.

Systematic on feed-down subtraction - Λ_c spectra [%]						
Mult. \ p_T	1 - 2	2 - 4	4 - 6	6 - 8	8 - 12	12 - 24
1-9	+1	+2	+2	+3	+5	-
	-0	-0	-0	-0	-0	-
10-29	+1	+1	+1	+1	+2	+2
	-1	-1	-2	-2	-3	-4
30-59	+0	+0	+0	+0	+0	-
	-2	-2	-3	-4	-6	-
60-99	+0	+0	+0	+0	+0	+0
	-3	-3	-4	-6	-9	-11

Table 4.16: Multiplicity dependent feed-down subtraction systematic uncertainties in each p_T and multiplicity bin associated to the Λ_c spectra.

Systematic on feed-down subtraction - Λ_c/D^0 ratio [%]						
Mult. \ p_T	1 - 2	2 - 4	4 - 6	6 - 8	8 - 12	12 - 24
1-9	+1 -0	+1 -0	+2 -0	+2 -0	+3 -0	- -
10-29	+1 -1	+1 -1	+1 -1	+1 -1	+1 -2	+1 -3
30-59	+0 -1	+0 -1	+0 -2	+0 -3	+0 -4	- -
60-99	+0 -2	+0 -2	+0 -3	+0 -4	+0 -6	+0 -7

Table 4.17: Multiplicity dependent feed-down subtraction systematic uncertainties in each p_T and multiplicity bin associated to the Λ_c^+/D^0 ratio.

4.8.4 Tracking efficiency

The tracking efficiency systematic uncertainty includes the effects arising from the track propagation from the TPC to the ITS (ITS-TPC matching efficiency) and from the applied track quality selection. It is assumed to be independent of the multiplicity intervals and is therefore estimated for the MB case only.

ITS-TPC matching efficiency

The ITS-TPC matching efficiency was computed as the ratio between the number of tracks successfully reconstructed with the Kalman filter [67] in the TPC and ITS with at least one hit in the SPD layers and the number of reconstructed tracks in the TPC. In the $\Lambda_c^+ \rightarrow pK_s^0 \rightarrow p\pi^+\pi^-$ analysis the systematic on ITS-TPC matching efficiency has to be taken into account only for the bachelor track since it is the only primary decay particle involved in the Λ_c decay (the pions decay out of the SPD most of the times). The systematic uncertainty on its determination arises from

discrepancies in the tracking performance between data and the MC simulation. The ITS-TPC matching efficiency is expected to be higher for primary particles than for secondary particles. More precisely, primary particles are defined as particles produced in the collision, including decay products except those from weak decays of strange particles, while the secondary particles are those produced in the interactions with the material or in decays of strange hadrons and thus with secondary vertices likely out of SPD. For this reason, the ITS-TPC matching efficiency is calculated separately for primary and secondary tracks in MC. The `PYTHIA` event generator and the `GEANT3` transport package do not perfectly reproduce the relative abundance of primary and secondary particles. Therefore, to account for this data-MC matching efficiency discrepancy, the real fraction of track types in data were evaluated for both primary and secondary particles (f_{primary} and $f_{\text{secondary}}$, respectively) and used to re-weight the respective MC efficiencies to obtain a corrected inclusive MC efficiency, $\epsilon^{\text{MC}}_{\text{inclusive}}$, which is computed as:

$$\epsilon^{\text{MC}}_{\text{inclusive}} = f_{\text{primary}} \times \epsilon^{\text{MC}}_{\text{primaries}} + (1 - f_{\text{primary}}) \times \epsilon^{\text{MC}}_{\text{secondary}} \tag{4.21}$$

The f_{primary} was evaluated with a data-driven technique based on a fit to the distribution of the measured track impact parameter on the xy plane, d_0^{xy}. A selection on the tracks requiring at least one hit in either of the two SPD layers was used to assure good enough resolution to separate primary and secondary d_0^{xy} distributions. The associated ITS-TPC matching efficiency systematic uncertainty is then evaluated as:

$$\frac{\epsilon^{\text{Data}}_{\text{inclusive}} - \epsilon^{\text{MC}}_{\text{inclusive}}}{\epsilon^{\text{Data}}_{\text{inclusive}}} \tag{4.22}$$

The Monte Carlo productions used for the study were the general purpose minimum bias MCs anchored to each period. The systematic uncertainties for the different periods are then averaged into the final values reported in Tab. 4.18.

Systematic on ITS-TPC matching efficiency [%]						
p_{T} interval	1 - 2	2 - 4	4 - 6	6 - 8	8 - 12	12 - 24
MB	1.8	2.7	2.3	2.4	3	2.9

Table 4.18: ITS-TPC matching efficiency systematic uncertainty associated to the bachelor track.

Track quality selection

A discrepancy between the efficiency of the track-quality selections in data and in the MC simulations need to be considered. This uncertainty was estimated by varying the track selection criteria for both the bachelor and V^0 daughters tracks. The following variations were tested:

- Central cut: at least 70 associated cross rows in the TPC and ratio of crossed rows over findable clusters in the TPC > 0.8.
- At least 70 associated cross rows in the TPC and ratio of crossed rows over findable clusters in the TPC > 0.9.
- Additional cut on the number of TPC crossed rows $> 120 - (5/p_{\mathrm{T}})$.
- Number of TPC clusters $> 0.65 \times$ number of TPC crossed rows.

The systematic uncertainties were estimated as the variation of the corrected yields with the different trials with respect to the central set of cuts and reported in Tabs. 4.19 and 4.20.

Track quality selection - Protons [%]							
p_{T} interval	0 - 1	1 - 2	2 - 4	4 - 6	6 - 8	8 - 12	12 - 24
MB	1	1	1	1	2	2	2

Table 4.19: Systematic uncertainties associated to the track quality selections for the bachelor tracks.

Track quality selection - Pions [%]							
p_T interval	0 - 1	1 - 2	2 - 4	4 - 6	6 - 8	8 - 12	12 - 24
MB	1	1	1	1	1	1.5	1.5

Table 4.20: Systematic uncertainties associated to the track quality selections for the V^0 daughters tracks.

Propagation to candidate level

Both the ITS-TPC matching efficiency and track quality selection were estimated at the track level as a function of the track p_T. A MC simulation was used to propagate the uncertainty at the track level to the Λ_c baryon level, accounting for the daughter's kinematic in the hadron p_T range of our analysis. The final assigned uncertainties are reported in Tab. 4.21.

Systematic on tracking efficiency [%]						
p_T interval	1 - 2	2 - 4	4 - 6	6 - 8	8 - 12	12 - 24
MB	4.5	5.5	6	7	7	7

Table 4.21: Systematic uncertainties associated to the Λ_c candidates tracking efficiency.

4.8.5 Multiplicity interval

Systematic effects due to the dependence of the efficiency on the $N_{\text{tracklets}}$ interval limits were studied for the multiplicity class [1-9] $N_{\text{tracklets}}$. The higher multiplicity classes were not considered in this computation because, as shown in 4.6.1, a multiplicity dependence of the efficiency is not observed for $N_{\text{tracklets}}$>4. These effects were a consequence of removing the reconstructed candidate's decay tracks from the multiplicity in data, described in section 4.2.2, but not in MC, where the efficiency weakly depends on multiplicity. The systematic uncertainty was evaluated by comparing the efficiency computed in a $N_{\text{tracklets}}$ interval shifted by one unit, [2-10] $N_{\text{tracklets}}$, with the one in the default intervals. The ratio is shown in Fig. 4.55 and

the estimated multiplicity interval systematic uncertainty is reported in Tab. 4.22.

Systematic on multiplicity interval [%]					
p_{T} interval	1 - 2	2 - 4	4 - 6	6 - 8	8 - 12
[1-9]	2	2	2	2	2

Table 4.22: Systematic uncertainties associated to the multiplicity interval [1-9] $N_{\mathrm{tracklets}}$.

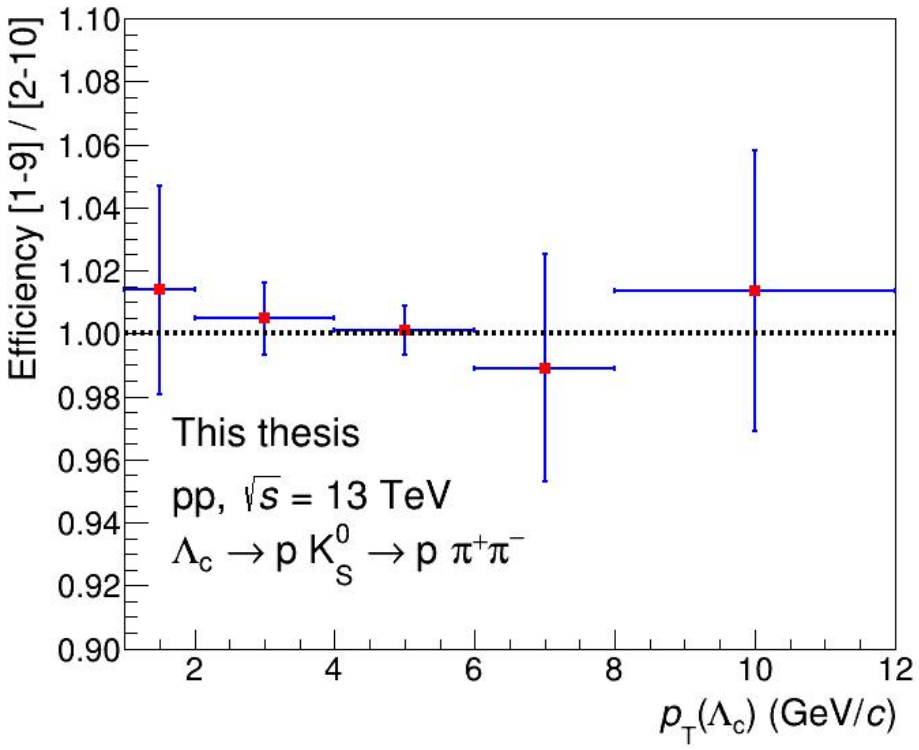

Figure 4.55: Ratio between the efficiency computed in the multiplicity range [1-9] $N_{\mathrm{tracklets}}$ and the one in [2-10] $N_{\mathrm{tracklets}}$.

4.8.6 HMSPD trigger correction

As described in section 4.2.3, a data-driven event reweighting procedure was applied for the HMSPD triggered data sample to account for the trigger inefficiency. The trigger turn-on-curve was normalized to 1 dividing the HMSPD/MB $N_{\mathrm{tracklets}}$ distribution by the observed ratio at $N_{\mathrm{tracklets}} = 70$. A systematic uncertainty was assigned to the reweighing procedure normalizing the ratio with different strategies. In particular, the following ways were adopted:

1. fit to a constant (pol0) in the range $65 < N_{\text{tracklets}} < 80$.

2. normalizing at $N_{\text{tracklets}} = 65$ and reweighing at 1 for $N_{\text{tracklets}} > 65$.

3. using the value at $N_{\text{tracklets}} = 70$ but selecting only the event with a Λ_c.

The corrected yield were estimated with the different normalization strategies listed above and the systematic was assumed to be the ratio between the nominal yield and the ones estimated with the alternative normalizations. The ratios are shown in Fig. 4.56 and the systematic was assumed to be 1% for all the p_T bins.

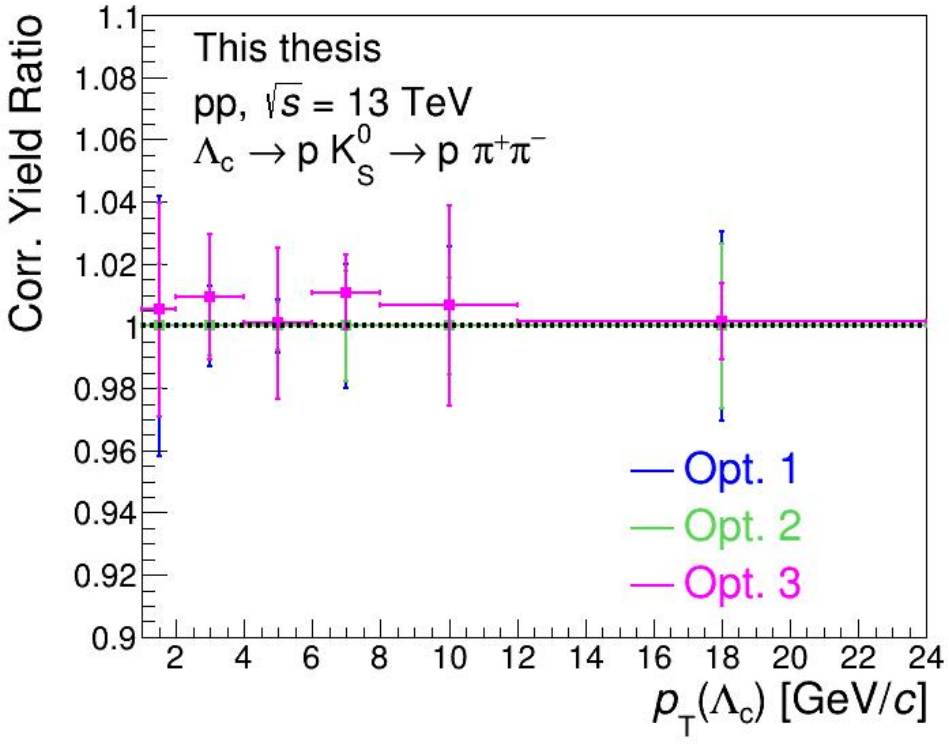

Figure 4.56: Ratios between the corrected yields in the HM class [60-99] measured with the nominal turn-on-curve normalization and those obtained with the different normalization strategies listed above.

4.8.7 MC p_T shape

Λ_c baryons are generated in the MC simulation with a p_T shape given by `PYTHIA` with the CR Mode 2. A discrepancy between the generated shape and the true Λ_c p_T shape leads to a systematic shift of the calculated efficiencies. To estimate the effect

arising from a different Λ_c p_T shape, the efficiency was calculated reweighting the Λ_c p_T distribution in the MC in order to reproduce the p_T shape predicted by FONLL calculations. The systematic effect was then estimated looking at the ratio between the efficiencies obtained using the different p_T shapes tested. The final computed systematic uncertainty is reported in Tab. 4.23.

Systematic on Λ_c generated p_T-shape [%]						
p_T interval	1 - 2	2 - 4	4 - 6	6 - 8	8 - 12	12 - 24
MB	1	0.5	0.5	0.5	0.5	1

Table 4.23: Systematic uncertainties associated to the Λ_c generated p_T shape.

4.8.8 z_{vtx} distribution

Possible differences between the primary-vertex position distributions along the beam axis, z_{vtx}, in simulations and in data were investigated, since a slight dependence of the efficiencies with z_{vtx} is observed. Hence, a further data-driven reweighting procedure was performed, taking this effect into account. The final computed systematic uncertainty was found to be 0.5% in the first three p_T bins and negligible elsewhere. The values are summarised in Tab. 4.24.

Systematic on z_{vtx} distribution [%]						
p_T interval	1 - 2	2 - 4	4 - 6	6 - 8	8 - 12	12 - 24
MB	0.5	0.5	0.5	0	0	0

Table 4.24: Systematic uncertainties associated to the z_{vtx} distribution.

4.8.9 Branching ratio

The results have a global systematic uncertainty due to the branching ratio of the Λ_c decay channel taken into account in this analysis. Its value is not dependent on the p_T bin and multiplicity class and it has been quoted as 5.5% according to the more recent PDG value [4].

4.9 Results

Fig. 4.57 shows the p_T-differential corrected yields per event at mid-rapidity ($|y| < 0.5$) for the Λ_c^+ baryons in pp collisions at $\sqrt{s} = 13$ TeV for the inelastic events INEL > 0 and in interval of event multiplicity, computed with (4.20). In the bottom panel, the ratios to the INEL > 0 class is presented. The vertical bars represent the statistical uncertainty, while the boxes the total systematic uncertainty. In the ratio, the efficiency and the multiplicity-dependent systematic sources were considered as uncorrelated among different multiplicity classes while the contributions of the tracking, the MC p_T shape, the z_{vtx} distribution, the beauty feed-down from `FONLL` and the branching ratio as fully correlated. The values utilized for the multiplicity dependent feed-down systematic uncertainty are reported in Tab. 4.17. The systematic uncertainties related to the BDT selection and to the raw-yield extraction were considered partially correlated with respect to the measurement performed in the INEL > 0 class. The differences between the multiplicity bins and the multiplicity integrated case are likely to be of statistical origin. So, for these two systematic sources the largest of the two was assigned to the ratio. Concerning the HM trigger efficiency systematic, it was assigned directly to the HM/INEL > 0 corrected yield ratio. The statistical uncertainty is also partially correlated because part of the data sample is in common. The assigned statistical uncertainty on the ratio is then estimated from a Barlow test [74].

A clear trend is visible with multiplicity: the measured p_T-differential yields increase from the lowest to the highest multiplicity class. Their ratios to INEL > 0 increase (decrease) with increasing p_T for the highest (lowest) multiplicity class, suggesting a plateau towards p_T>10 GeV/c, as recently observed also for the light-flavour hadrons [75, 76, 77]. The observed trend of the corrected yields can be explained by a higher charm quark production in high multiplicity events and to a hardening of the measured p_T spectra (shift of the spectra toward higher p_T) with increasing $\langle dN_{ch}/d\eta \rangle$, which contributes to the increase of the average transverse momentum with multiplicity.

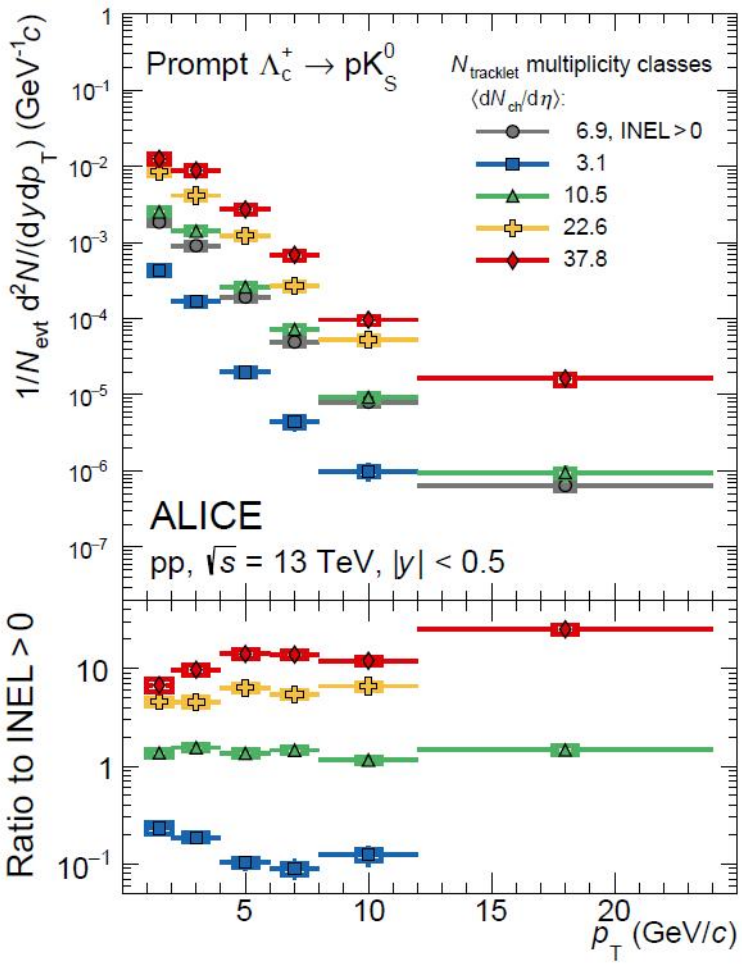

Figure 4.57: Transverse-momentum spectra of Λ_c^+ hadrons in the decay channel $\Lambda_c^+ \rightarrow pK_s^0 \rightarrow p\pi^+\pi^-$ measured in pp collisions at $\sqrt{s} = 13$ TeV for different SPD multiplicity classes selected at midrapidity. The corresponding ratios to INEL > 0 are shown in the bottom panel.

4.9.1 Combination of $\Lambda_c^+ \rightarrow pK_s^0 \rightarrow p\pi^+\pi^-$ and $\Lambda_c^+ \rightarrow pK^-\pi^+$ measurements

The ALICE collaboration has also performed the Λ_c^+ corrected yields measurement via the reconstruction of its three prongs decay channel, $\Lambda_c^+ \rightarrow pK^-\pi^+$, with a BR $= (6.28 \pm 0.32)$ %, on the same data. The $\Lambda_c^+ \rightarrow pK^-\pi^+$ has a higher branching ratio that guarantees a higher production yield. Besides, its decay in three charged prongs: (i) allows one to reconstruct the secondary vertex and to exploit the Λ_c^+ topology for the selections (Λ_c decay length, cosine of the pointing angle, etc.); (ii) makes it more

sensitive to the tracking efficiency than the $\Lambda_c^+ \to pK_s^0 \to p\pi^+\pi^-$. The comparison between the two analysed Λ_c^+ decay channels is reported in Fig. 4.58.

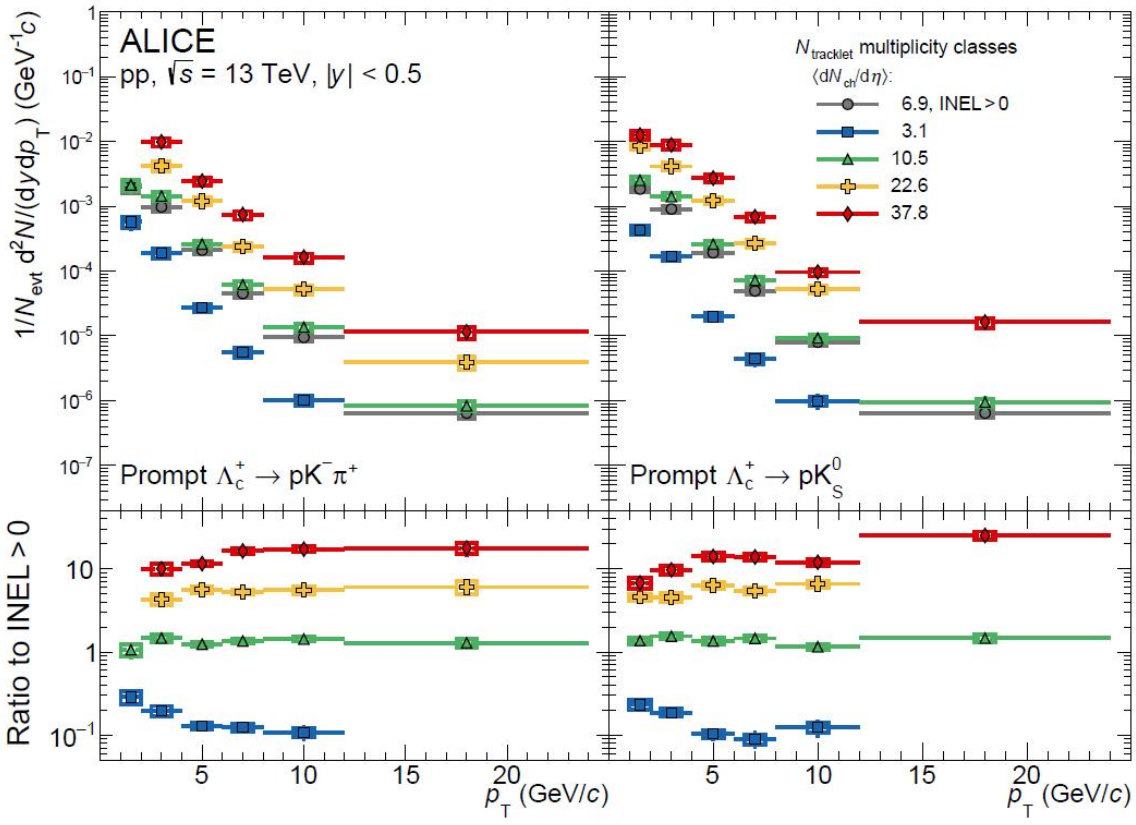

Figure 4.58: Transverse-momentum spectra of Λ_c^+ hadrons in the decay channels $\Lambda_c^+ \to pK^-\pi^+$ (left) and $\Lambda_c^+ \to pK_s^0 \to p\pi^+\pi^-$ (right) measured in pp collisions at $\sqrt{s} = 13$ TeV for different SPD multiplicity classes selected at midrapidity. The corresponding ratios to INEL > 0 are shown in the bottom panels.

To obtain a more precise measurement of the p_T-differential Λ_c^+-baryon corrected yields, the results from the two decay channels, which are statistically uncorrelated, were combined. A weighted average was computed using the inverse of the quadratic sum of the relative statistical and uncorrelated systematic uncertainties, $1/(w_i^{uncorr})^2$, as weights.

$$w_i = \frac{1}{(w_i^{uncorr})^2} \tag{4.23}$$

$$\text{Merged CorrYield} = \frac{\sum_i (\text{CorrYield}_i \cdot w_i)}{\sum_i w_i} \tag{4.24}$$

The statistically uncorrelated uncertainties were added quadratically while the correlated ones linearly, as:

$$\Delta\sigma^{\text{uncorr}} = \frac{\sqrt{\sum_{\text{i}}(\Delta\sigma_{\text{i}}^{\text{uncorr}} \cdot \text{w}_{\text{i}})^2}}{\sum_i \text{w}_{\text{i}}} \tag{4.25}$$

$$\Delta\sigma^{\text{corr}} = \frac{\sum_i(\Delta\sigma_{\text{i}}^{\text{corr}} \cdot \text{w}_{\text{i}})}{\sum_i \text{w}_{\text{i}}} \tag{4.26}$$

The merged Λ_{c}^{+} corrected yields are reported in Fig. 4.59. All the results that will be shown from now on are obtained merging the two Λ_{c}^{+} decay channels.

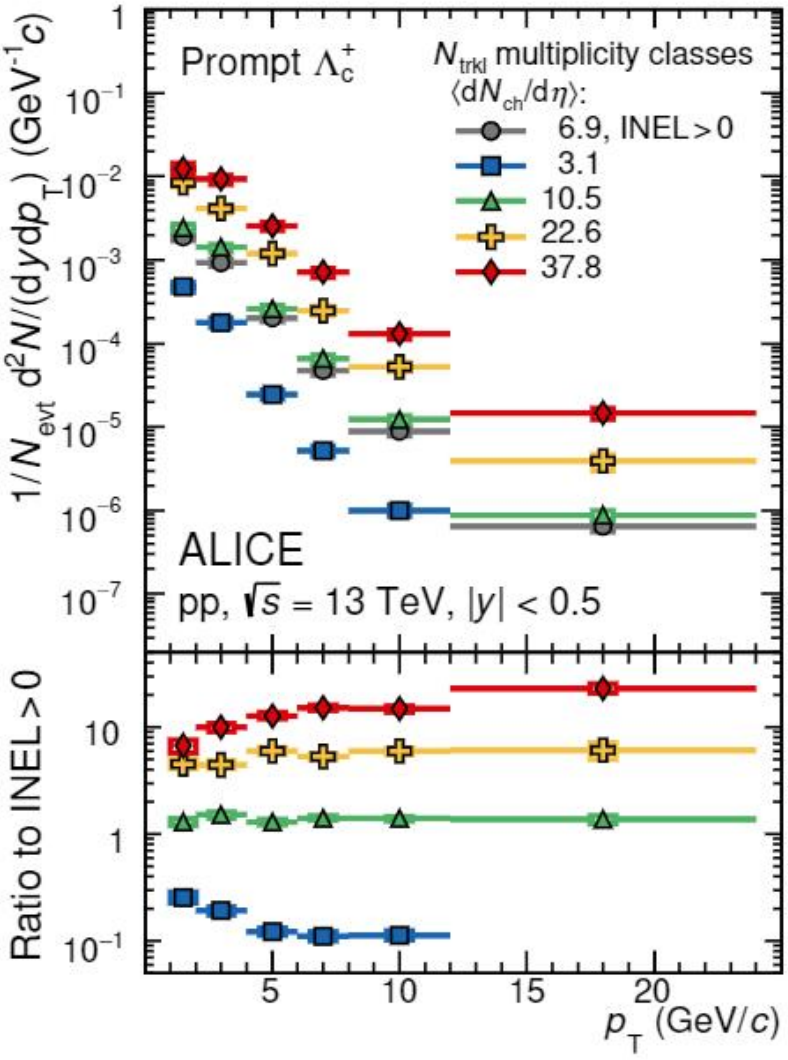

Figure 4.59: Merged $\Lambda_{\text{c}}^{+} \to \text{pK}^{-}\pi^{+}$ and $\Lambda_{\text{c}}^{+} \to \text{pK}_{\text{s}}^{0} \to \text{p}\pi^{+}\pi^{-}$ transverse-momentum spectra measured in pp collisions at $\sqrt{s} = 13$ TeV for different SPD multiplicity classes selected at midrapidity. The corresponding ratios to INEL > 0 are shown in the bottom panel.

The systematic uncertainties treated as uncorrelated between the different decay channels include those due to the raw-yield extraction, the cut and PID variations and the (Acc $\times$ ϵ) statistical uncertainties. The systematic uncertainties due to the tracking efficiency, the generated Λ_c^+ p_T spectrum, the trigger and both the beauty feed-down uncertainties were treated as correlated between the two decay channels. The branching fraction is partially correlated and the correlation coefficient $\langle \delta x_i \delta x_j \rangle = 50\%$ is taken from the PDG [4]. In addition, the multiplicity-dependent systematic sources were considered as correlated between the two decay channels.

4.9.2 Baryon-to-meson ratio

The baryon-to-meson ratio, Λ_c^+/D^0, was measured in order to study the multiplicity dependence of the D^0 and Λ_c^+ productions. Furthermore, it allows one to further investigate the already observed enhancement in pp collisions at $\sqrt{s} = 5.02$ and $\sqrt{s} = 7$ TeV [40, 78] with the respect to e^+e^- collisions. The p_T-differential Λ_c^+/D^0 ratio in pp collisions at $\sqrt{s} = 13$ TeV in different event multiplicity classes is shown in Fig. 4.60.

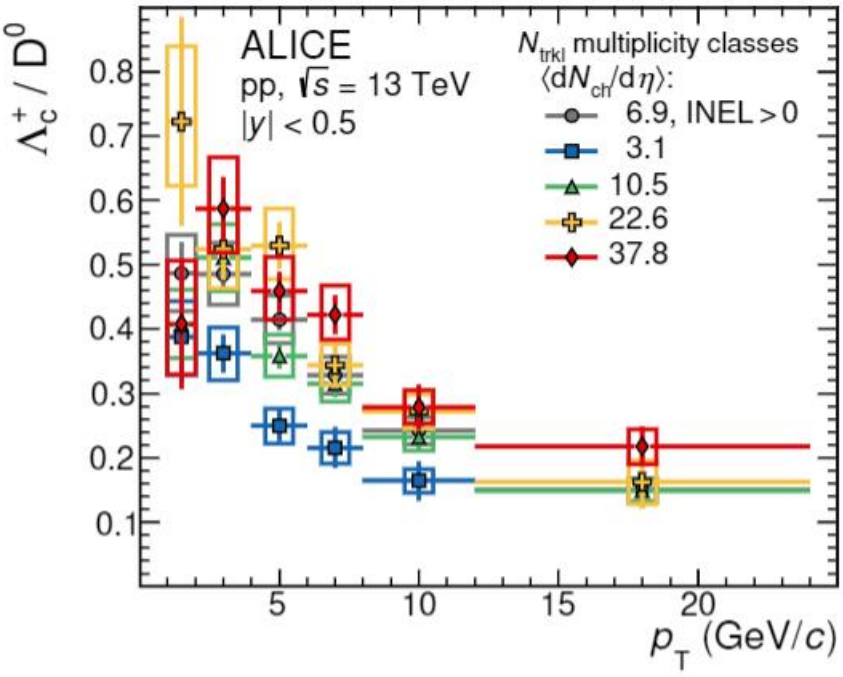

Figure 4.60: Λ_c^+/D^0 ratios measured in pp collisions at $\sqrt{s} = 13$ TeV for different SPD multiplicity classes.

The sources of uncertainty assumed to be uncorrelated between different charm-hadron species include the raw-yield extraction, the cut variation, the shape of the MC p_{T} spectra, the z_{vtx} distribution and the branching ratio. The systematic uncertainty deriving from the variation of the multiplicity-interval limits was propagated as partially correlated[7], while the other systematic uncertainties were assumed to be fully correlated.

The p_{T}-differential $\Lambda_{\mathrm{c}}^{+}/\mathrm{D}^{0}$ ratio shows an evident dependence on multiplicity, and a hierarchy is observed going from the lowest to the highest multiplicity interval. The measured $\Lambda_{\mathrm{c}}^{+}/\mathrm{D}^{0}$ ratio in the lowest multiplicity class is still higher, in the measured p_{T} range, than the average of corresponding ratios measured in $\mathrm{e}^{+}\mathrm{e}^{-}$ collisions at LEP [79]. Moreover, the increase at intermediate p_{T} has a similar behaviour to those observed in p-Pb and Pb-Pb collisions shown in Fig.2.8. It suggests that modifications of the hadronisation mechanisms seem to be a common feature of all hadronic collisions, and further theoretical and experimental investigations are needed to shed the light on this topic.

In order to estimate a significance level for the difference observed in the two extreme multiplicity classes at midrapidity, the two highest multiplicity (HM) over the lowest multiplicity (LM) $\Lambda_{\mathrm{c}}^{+}/\mathrm{D}^{0}$ ratios were computed and shown in Fig. 4.61. This estimate was performed taking into account statistical and systematic uncertainties, for which the raw-yield extraction, the cut variation, the shape of the MC p_{T} spectra, and the z_{vtx} distribution sources were considered as uncorrelated, the systematic uncertainty deriving from the variation of the multiplicity-interval limits as partially correlated, while the other sources cancelled out in the double ratio.

[7] The Λ_{c}^{+} decays in 3 prongs while the D^{0} in 2 prongs. It was assumed that the 2 prongs part cancel out in the ratio and 1/3 of the numerator systematic uncertainty was applied to the ratio.

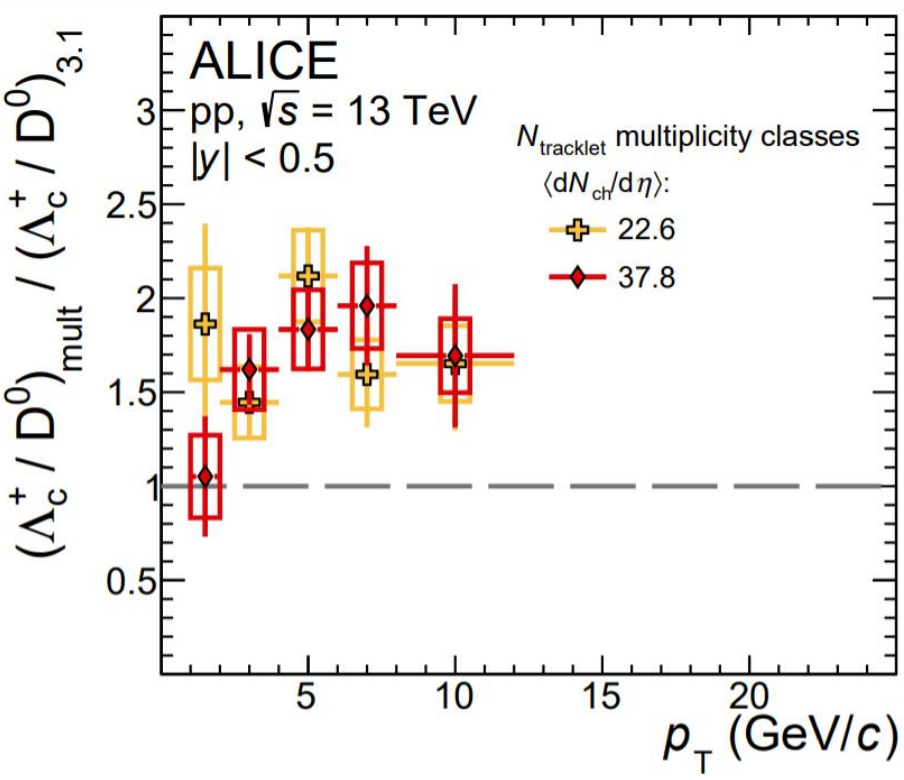

Figure 4.61: $(\Lambda_c^+/D^0)_{HM}/(\Lambda_c^+/D^0)_{LM}$ double-ratios measured in pp collisions at $\sqrt{s} = 13$ TeV for different SPD multiplicity classes.

The probability of the measured double-ratio DR $= (\Lambda_c^+/D^0)_{HM}/(\Lambda_c^+/D^0)_{LM} > 1$ corresponds to a significance of 5.3σ[8] in the $1 \leq p_T < 12$ GeV/c interval for the highest multiplicity class, considering as null hypothesis DR $= 1$. With the aim of investigating the least favourable case, in the significance computation, the measured

[8]The probability that the measured double-ratio DR $= (\Lambda_c^+/D^0)_{HM}/(\Lambda_c^+/D^0)_{LM} > 1$ was estimated via the cumulative distribution function (CDF) of the of the gaussian distribution:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}}\, du \tag{4.27}$$

where, in our case, x is defined as:

$$x = \frac{\mathrm{DR} - 1}{\Delta(\mathrm{DR})} \tag{4.28}$$

and Δ(DR) is the statistical uncertainty associated to DR. The CDF, $\Phi(x)$, of a random variable Z normally distributed with mean 0, represents the probability that Z will take a value less than or equal to x. So, $p_i = (1 - \Phi(x))$ is the probability that the measurement is $> x$. The probability p_i was calculated for each measured p_T bin. The total probability, p, was then estimated multiplying all of them and expressed in term of statistical significance via the probit function, i.e. the inverse of the gaussian cumulative distribution function:

$$\Phi^{-1}(p) = \sqrt{2}\mathrm{erf}^{-1}(2p - 1) \tag{4.29}$$

values in all p_{T} intervals were shifted down by one standard deviation, by considering the sources of systematic uncertainties correlated with p_{T} that do not cancel out in the double ratio, i.e. those arising from the cut variation and the generated p_{T} spectra.

The measured charm-hadron ratios for the lowest and highest multiplicity classes were compared to model predictions from PYTHIA MC generators and a statistical hadronisation model in Fig. 4.62, in the left and right panels, respectively. The simulations with the PYTHIA event generator were performed with the Monash [35] and the Colour-Reconnection Beyond the Leading Colour approximation (CR-BLC) tunes [42]. In the Monash tune the fragmentation functions are tuned on $\mathrm{e^+e^-}$ data and it is meant at describing the in-vacuum hadronisation. Three different configurations (or modes) for the CR simulations were considered. For each mode, different constraints on the allowed reconnection among colour sources are imposed, in particular concerning the causality connection among strings involved in a reconnection and time dilation caused by relative boosts of the strings [42]. The Monash tune, however, does not reproduce the $\Lambda_{\mathrm{c}}^{+}/\mathrm{D}^0$ ratio and, furthermore, it does not show a multiplicity dependence. By contrast, the CR-BLC tunes describe the $\Lambda_{\mathrm{c}}^{+}/\mathrm{D}^0$ decreasing trend versus p_{T} and are closer to the overall magnitude, as already observed in minimum-bias pp collisions at $\sqrt{s} = 5.02$ and $\sqrt{s} = 13$ TeV [41, 80]. The CR-BLC tunes show a clear dependence with multiplicity, qualitatively reproducing the trend observed in data. The measurements in Fig. 4.62 are also compared with the predictions of a canonical-ensemble statistical hadronisation (CE-SH) model [81], where the grand-canonical statistical hadronisation model (SHM) [24] of charm-hadron production was generalized to the case of canonical SHM, with a fixed number of particles. The CE-SH model explains the multiplicity dependence as deriving from the reduced volume size of the formalism towards smaller multiplicity, where a decrease of the $\Lambda_{\mathrm{c}}^{+}/\mathrm{D}^0$ ratio is a consequence of the strict baryon-number conservation. The version of the SHM model based on the measured charm-baryon spectrum reported in the PDG [4] was observed to strongly underestimate the $\Lambda_{\mathrm{c}}^{+}/\mathrm{D}^0$ measurements in

minimum-bias pp collisions [41]. For this reason, for the Λ_c^+/D^0 case, the underlying charm-baryon spectrum in the calculations is augmented to include additional excited baryon states predicted by the Relativistic Quark Model (RQM), Tab. 2-6 in [55]. The model calculations describe the measured Λ_c^+/D^0 ratios and reproduce the multiplicity dependence. This measurement provides new experimental constrain to those yet-undiscovered charmed baryon states. The Catania model [56], in which the hadronisation is implemented via both coalescence and fragmentation, provides a good description of the Λ_c^+/D^0 ratio measured in pp collisions at $\sqrt{s} = 5.02$ [40] and $\sqrt{s} = 13$ TeV [80] and it may indicate that charm hadronisation also in pp collisions involves coalescence of charm and light quarks. Since the Catania model is not able to provide predictions as a function of the event charged-particle multiplicity, the comparison with the multiplicity-dependent results reported in this thesis is not available.

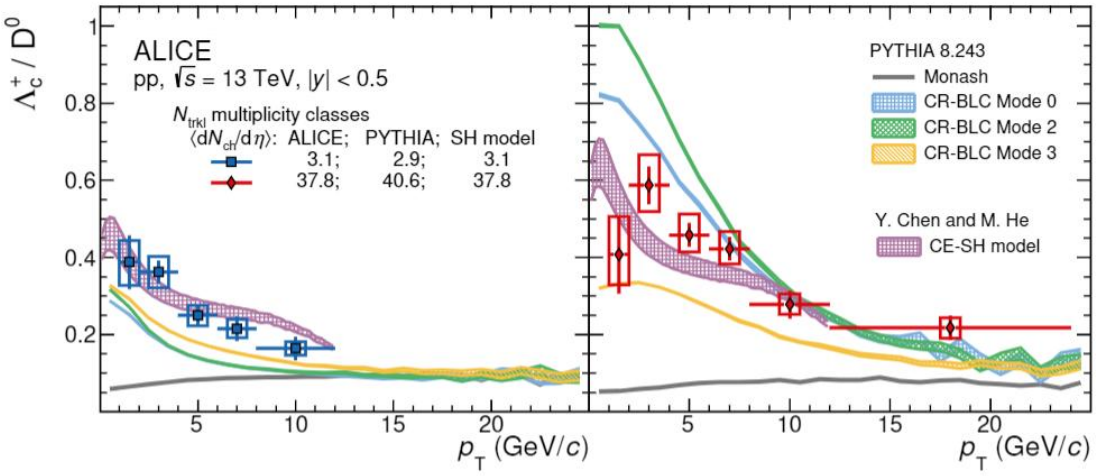

Figure 4.62: The Λ_c^+/D^0 ratios measured at midrapidity in pp collisions at $\sqrt{s} =$ 13 TeV for the lowest (left) and highest (right) multiplicity classes. The measurements are compared to `PYTHIA` predictions with the Monash [35] and CR-BLC tunes [42], and with the CE-SH model [81]. For the `PYTHIA` predictions the bands are the statistical uncertainties on the simulations, while for the CE-SH model they refer to the variation of the branching ratios of the additional charm-baryon states from RQM [55].

Similarities in the baryon-to-meson ratio between the heavy- and light-flavour sectors were observed in pp and p–Pb collisions at $\sqrt{s} = 5.02$ TeV in the multiplicity integrated case [40]. In particular, the Λ_c^+/D^0 and the Λ/K_S^0 ratios were observed to be consistent, in terms of both shape and magnitude, within uncertainties. These similarities between heavy-flavour and light-flavour measurements hint a potential common mechanism for light- and charm-baryon formation in hadronic collisions at LHC energies. To investigate if this common behaviour extends in different multiplicity regions, the comparison of the Λ_c^+/D^0 and the Λ/K_S^0 [75] baryon-to-meson ratios as a function of p_T in pp collisions at $\sqrt{s} = 13$ TeV, in similar low and high multiplicity classes, reported in Fig 4.63, were studied.

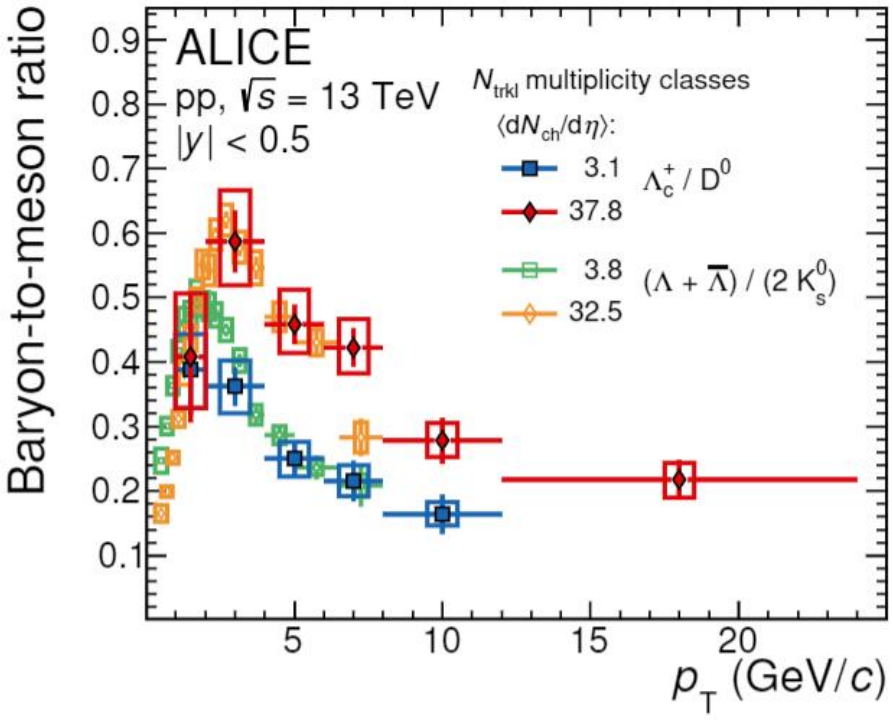

Figure 4.63: The Λ/K_S^0 baryon-to-meson ratios [75] compared to the Λ_c^+/D^0 ratios measured in pp collisions at $\sqrt{s} = 13$ TeV for similar low- and high-multiplicity classes.

In the vacuum-fragmentation scenario, the light-flavour hadron production has a significant contribution from gluon fragmentation, whereas heavy-flavour hadrons are primarily produced through the fragmentation of a charm quark, which is in turn

produced in the initial hard scattering. In addition, at low p_T, light-flavour hadrons originate mainly from small-momentum soft scattering processes. Despite these differences, the light- and heavy-flavour baryon-to-meson ratios, Λ_c^+/D^0 and Λ/K^0_S, show a remarkably similar trend as a function of $\langle dN_{ch}/d\eta \rangle$. The measurements also suggest a similar shift of the baryon-to-meson ratio peaks towards higher momenta, with increasing multiplicity.

4.9.3 p_T integrated results

To access the total hadron yield, the p_T-integrated yields of Λ_c^+ and D^0 were computed by integrating the p_T-differential spectra in their measured range and extrapolating them down to $p_T = 0$ in each multiplicity interval. In the integration, the systematic uncertainties were propagated considering the uncertainties due to the raw-yield extraction, the statistical uncertainty on the efficiency as fully uncorrelated and all the other sources as fully correlated among p_T intervals. The `PYTHIA` predictions with CR-BLC Mode 2 were used for the extrapolation in each multiplicity interval, for both Λ_c^+ and D^0. The procedure is reported in the following:

- Compute the integrated corrected yield for Λ_c^+ and D^0 in the measured (visible) range, i.e. $1 < p_T < 24$ GeV/c ($1 < p_T < 12$ GeV/c for the lower multiplicity class).
- Extrapolate the Λ_c^+ and D^0 spectra down to $p_T = 0$ using `PYTHIA` Mode 2 as central method.
- The extrapolation factor was computed as the ratio of the `PYTHIA` spectrum integrated in the full p_T range, $0 < p_T < 24$ GeV/c, to the integral in the visible p_T range, $1 < p_T < 24$ GeV/c.

$$\text{Extrap.Factor}\,\Lambda_c^+ = \frac{\Lambda_c^{+\,\text{PYTHIA}}_{(0-24)}}{\Lambda_c^{+\,\text{PYTHIA}}_{(1-24)}} \tag{4.30}$$

$$\text{Extrap.Factor}\,D^0 = \frac{{D^0}^{\text{PYTHIA}}_{(0-24)}}{{D^0}^{\text{PYTHIA}}_{(1-24)}} \tag{4.31}$$

- The Λ_c^+ and D^0 yields in the full p_T range were obtained by integrating the yield in the visible p_T interval and scaling by the extrapolation factor.

$$\Lambda_c^{+\,\text{Integrated}}_{(0-24)} = \Lambda_c^{+\,\text{Visible}}_{(1-24)} * \text{Extrap.Factor}\,\Lambda_c^+ \tag{4.32}$$

$$D^{0\,\text{Integrated}}_{(0-24)} = D^{0\,\text{Visible}}_{(1-24)} * \text{Extrap.Factor}\,D^0 \tag{4.33}$$

- Compute the systematics for the extrapolation considering the difference among Λ_c^+ and D^0 with `PYTHIA` Mode2 and `PYTHIA` Mode 0 and with the extrapolation via corrected yield fit with a Tsallis-Lévy and a power law functions. The fits were performed considering the statistical and p_T-uncorrelated sources of systematic uncertainties and also shifting up and down the data by one sigma of the p_T-correlated systematic uncertainties. The envelope of the extrapolation factors obtained with all the trials was assigned as the extrapolation uncertainty on Λ_c^+ and D^0 and it was propagated to the Λ_c^+/D^0 ratio.

The p_T-integrated Λ_c^+/D^0 yield ratio as a function of $\langle dN_{ch}/d\eta \rangle$ is shown in Fig. 4.64, where the systematic uncertainties from the extrapolation (shaded boxes, assumed to be uncorrelated among multiplicity intervals) are drawn separately from the other sources of systematic uncertainties (empty boxes). The sources related to the raw-yield extraction, the multiplicity-interval limits, the high-multiplicity triggers, the multiplicity-independent prompt fraction assumption and the statistical uncertainties on the efficiencies are also considered uncorrelated with multiplicity. The other systematic uncertainties are assumed to be correlated. The measurements performed in pp and p-Pb collisions at $\sqrt{s} = 5.02$ Tev [41] are also shown in Fig 4.64.

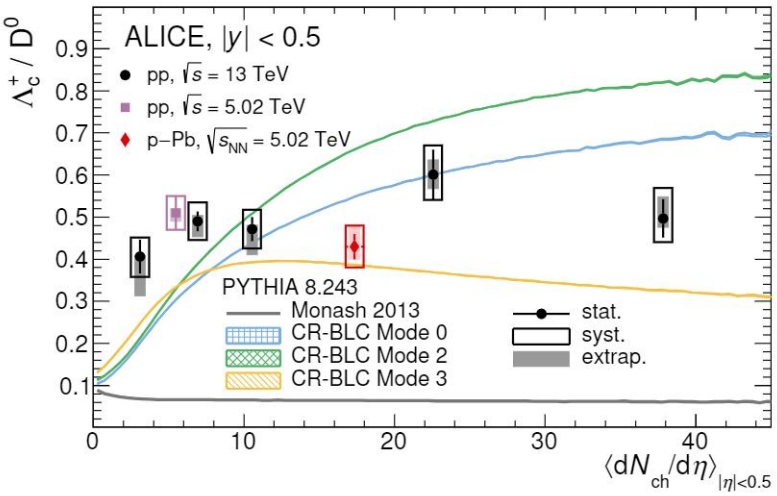

Figure 4.64: p_{T}-integrated $\Lambda_{\mathrm{c}}^{+}/\mathrm{D}^{0}$ ratios as a function of $\langle \mathrm{d}N_{\mathrm{ch}}/\mathrm{d}\eta \rangle$ in pp collisions at $\sqrt{s} = 13$ TeV. Measurements performed in pp and p–Pb collisions at $\sqrt{s} = 5.02$ TeV [41] and the PYTHIA predictions are also shown. Statistical and systematic uncertainties are shown by error bars and empty boxes, respectively. Shaded boxes represent the extrapolation uncertainties.

The result does not favour an increase of the yield ratios with multiplicity, as also observed for the $\Lambda/\mathrm{K}^{0}_{\mathrm{S}}$ ratio in Ref. [75], and the trend is compatible with a constant function. This suggests that the observed trend for the $\Lambda_{\mathrm{c}}^{+}/\mathrm{D}^{0}$ ratio in the $1 < p_{\mathrm{T}} < 24$ GeV/c range could come from a re-distribution of p_{T} that acts differently for baryons and mesons. The results are compared to the p_{T}-integrated PYTHIA predictions in Fig. 4.64. The measurements again exclude the Monash prediction in the whole multiplicity range and tend to be significantly below the CR-BLC Mode 2 for the three highest multiplicity intervals. The results support the hypothesis that flow-like effects or modified hadronisation mechanisms as coalescence could contribute already in pp collisions to enhance the mean baryon p_{T}, although more accurate measurements are needed to establish their role in pp collisions.

Chapter 5

Prospects: the ALICE 3 experiment

Building on the very successful Run 1 and 2 campaigns, the LHC experiments are gearing up in preparation for the upcoming Runs 3 and 4, in which significant further progress is expected. The ALICE collaboration has just completed, in view of the LHC Run 3, a major upgrade programme specifically targeted at the physics of ultra-relativistic nuclear collisions, focussing on the improvement of its event-rate capabilities and on a significant enhancement of the tracking accuracy at low momentum, achieved with the installation of the new Inner Tracking System (ITS 2) [82].

In spite of such an ambitious scientific programme for the present decade, crucial questions will still remain unanswered with the present detector concepts, and will require a novel experimental approach. Addressing what remains inaccessible in Run 3 and 4 requires qualitative steps in luminosity and detector performance. For this reason, the design of a new detector has been proposed. Great efforts for the preparation of the Letter Of Intent of this new experiment are underway. The ALICE 3 detector will start his operation during the Run 5 of LHC, foreseen for 2035-2036. Some of the main topic that will be investigated by the ALICE 3 experiment are:

- **Multi-charm heavy-flavour hadrons**: provide a unique system for testing

models of quantum chromodynamics. Their in-vacuum production from the same hard scattering is highly suppressed but a large increase is expected via coalescence with charm from different hard scattering. They provide a very strong probe to detect the QGP formation and to study the hadronisation mechanism in presence of a QGP medium.

- **Dileptons**: they are produced at all stages of the collision with negligible final-state interactions, so they carry information from the whole space-time evolution of the system, allowing precision measurement of the plasma temperature and of the chiral symmetry restoration.

- **In-medium interaction**: understanding of the rich phenomenology of QCD matter, connecting parton energy loss, collective flow, hadronisation and electromagnetic radiation in a unified description.

5.1 The ALICE 3 detector

The access to the physics observables discussed above determine the requirements on the detector design:

- Tracker with excellent vertexing, able to push the impact parameter resolution down to very low transverse momentum.

- Extensive particle identification over a very wide transverse momentum range.

- Large rapidity acceptance.

- Kinematic range down to very low p_{T}.

- Operate at rates that significantly exceed the capabilities of the present ALICE detector.

The final detector design is still under discussion. The first proposed layout, fulfilling the requirements listed above, is reported in Fig. 5.1.

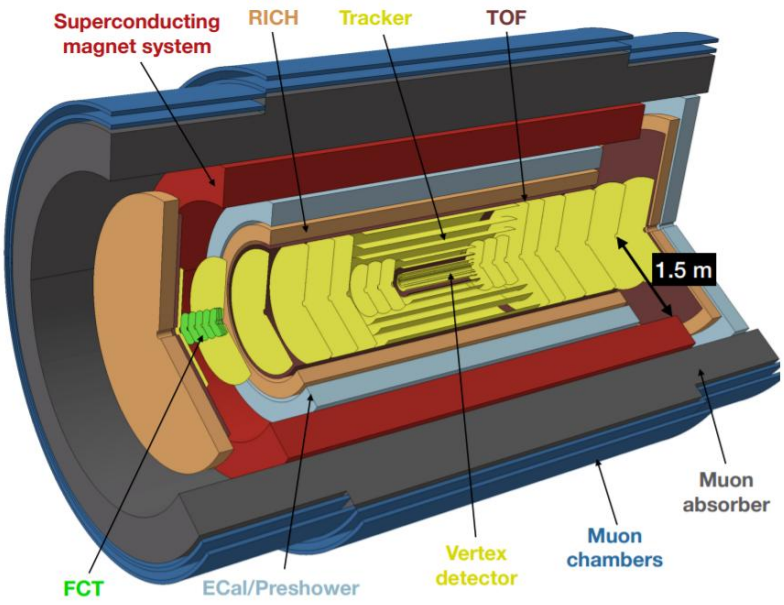

Figure 5.1: ALICE 3 detector layout: a silicon tracker composed of cylinders and disks serves for track reconstruction in the magnetic field provided by a super-conducting magnet system. The vertex tracker is contained within the beam pipe. For particle identification a time-of-flight detector, RICH detector, photon detector and a muon system are employed.

The charged particle reconstruction is based on a silicon pixel tracker, with sensors arranged in barrel layers and forward disks. The momentum is reconstructed from the curvature in the magnetic field provided by a superconducting magnet system, for which different magnetic field configurations are under study. The tracker consists of 11 cylindrical layers and 12 forward discs on either side of the IP. In order to achieve the required pointing resolution, the first hit must be measured as close as possible to the IP and with as little material as possible in front of the first layer to reduce the impact of multiple scattering. The minimal radial distance from the IP is determined by the aperture required for the LHC beam. While this amounts to $\sim$5 mm at top energy, $\sim$15 mm are necessary at injection energy. Therefore, having the first detection layer at a radial distance of $\sim$5 mm for data taking is possible only with a detector that can be retracted for injection and approached for data taking at collision energy (IRIS tracker). This implies that the IRIS vertex detector, made by

3 layers, must be installed within the beam pipe. A schematic representation of the IRIS tracker is shown in Fig. 5.2.

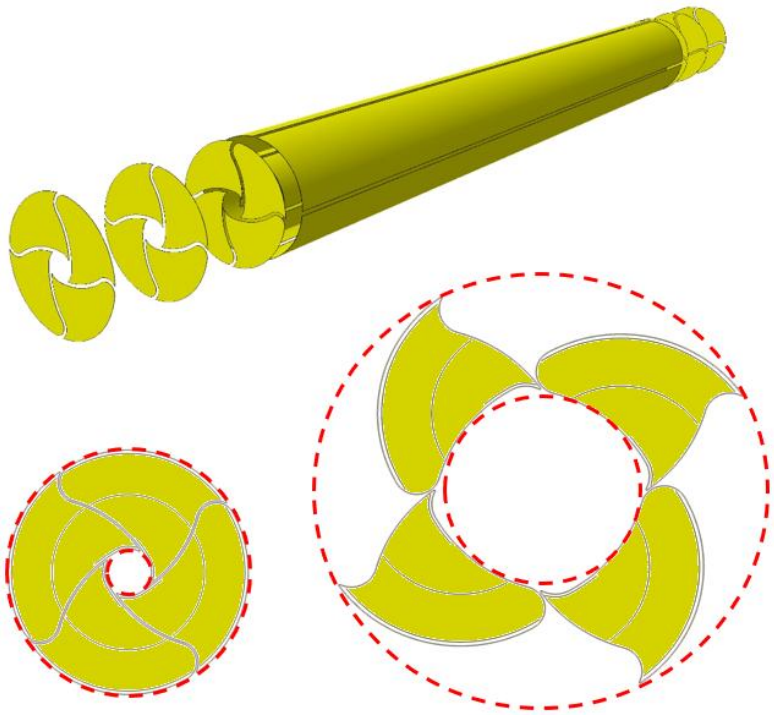

Figure 5.2: IRIS tracker schematic representation. On the bottom of the figure the IRIS detector is reported closed (left) and retracted (right).

The tracker is complemented by systems for particle identification. A time-of-flight layer outside of the tracker at a radius of 85 cm allows the identification of electrons and hadrons up to transverse momenta of 500 MeV/c and 2 GeV/c (π/K separation), respectively. For particles below 300 MeV/c, which do not reach this TOF layer, an inner TOF layer is foreseen at a radius of 20 cm. In the forward direction, TOF disks are installed following the last tracking disk.

To further extend the PID capabilities beyond the momentum reach of the TOF, a Cherenkov detector is considered behind it. In the central barrel, a Ring Imaging Cherenkov (RICH) detector enables the separation of electrons and pions up to 2 GeV/c and of protons from e, π, K up to 4 GeV/c.

For the identification of muons, a steel-absorber of about 70 cm thickness would be installed outside of the magnet. Two layers of muon detectors are used to detect and

match the muon tracklets to tracks in the silicon pixel tracker, which will provide the information on the transverse momentum.

An electromagnetic calorimeter (ECal) is foreseen for the reconstruction of photons with very low transverse momenta, down to 1 MeV/c.

The tracking of $e^{\pm}$ pairs from photon conversion could be achieved with a Forward Conversion Tracker (FCT), i.e. an array of silicon pixel disks installed in the forward direction ($3 < \eta < 5$). In this way, both the photon direction and energy can be reconstructed precisely.

5.2 Detector performance

The Λ_c baryons in pp collisions were used as a baseline for the performance studies of this new detector. In particular, the expected significance of the Λ_c in the decay channel $\Lambda_c^+ \rightarrow pK^-\pi^+$ was estimated to benchmark the detector configuration. The $\Lambda_c^+ \rightarrow pK^-\pi^+$ is indeed very sensitive to the impact parameter resolution and to the PID and can provide a starting point for the multi-charm analyses. Moreover, this channel has been chosen over the $\Lambda_c^+ \rightarrow pK_s^0 \rightarrow p\pi^+\pi^-$ because it is more sensitive to the tracking resolution of the detector. The studies presented were performed by means of fast Monte Carlo simulations based on the `DELPHES` software package [83]. The aim of the analysis is to provide a significance estimation for the $\Lambda_c^+ \rightarrow pK^-\pi^+$ decay channel with different configurations of the detector and including or not the PID in the computation. As a first test, the tracking performance of the ALICE 3 barrel were studied, computing the Λ_c expected significance with and without the retractable IRIS detector. The two layouts considered are:

- Layout v1, with the IRIS detector and the tracker innermost layer at R = 5 mm.
- Without the IRIS detector and the tracker innermost layer at R = 15 mm.

In the computation, the following MC simulations were employed:

- a Minimum Bias MC produced with PYTHIA 8.2 [84] with the Monash tune [35], from which the Λ_c coming from the combinatorial background were taken.
- a $\Lambda_c^+ \rightarrow pK^-\pi^+$ enriched MC, made with PYTHIA 8.2 with the Monash tune and in which at least one $c\bar{c}$ pair is required and the Λ_c are forced to decay in the channel under study, utilized to increase the Λ_c signal candidates needed for the computation.

Only the candidates in the rapidity interval $|y| < 1.44$ were considered. The same magnetic field already employed in the Run 2 set-up (B = 0.5 T) was used for this test as well. Detailed studies of the signal and background distributions were performed to choose a set of prefiltering cuts with a high background rejection and the variables that can be useful to discriminate signal from background. Examples of the variable distributions, i.e. the cosine of the Λ_c pointing angle, the Λ_c decay length and the impact parameted for the Λ_c proton track (prong0), for low ($<$ 2 GeV/c) and high ($>$ 4 GeV/c) Λ_c transverse momentum bins are shown in Figs. 5.3-5.8.

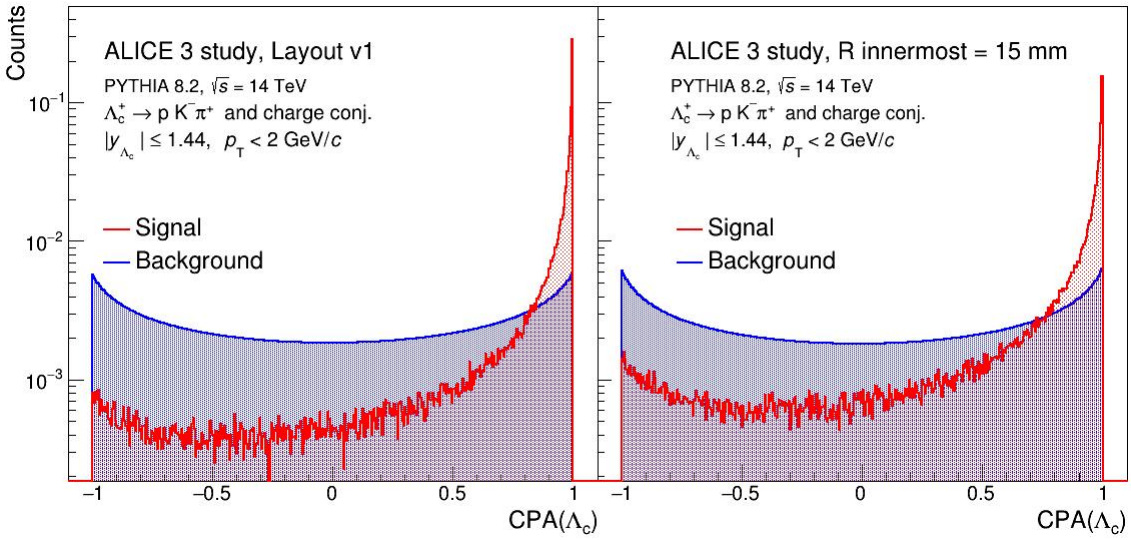

Figure 5.3: Signal and background distributions of cosine of the pointing angle of $\Lambda_c^+ \rightarrow pK^-\pi^+$ candidates with (left) and without (right) the IRIS detector at low p_T ($<$ 2 GeV/c). The distributions are normalised by their integral.

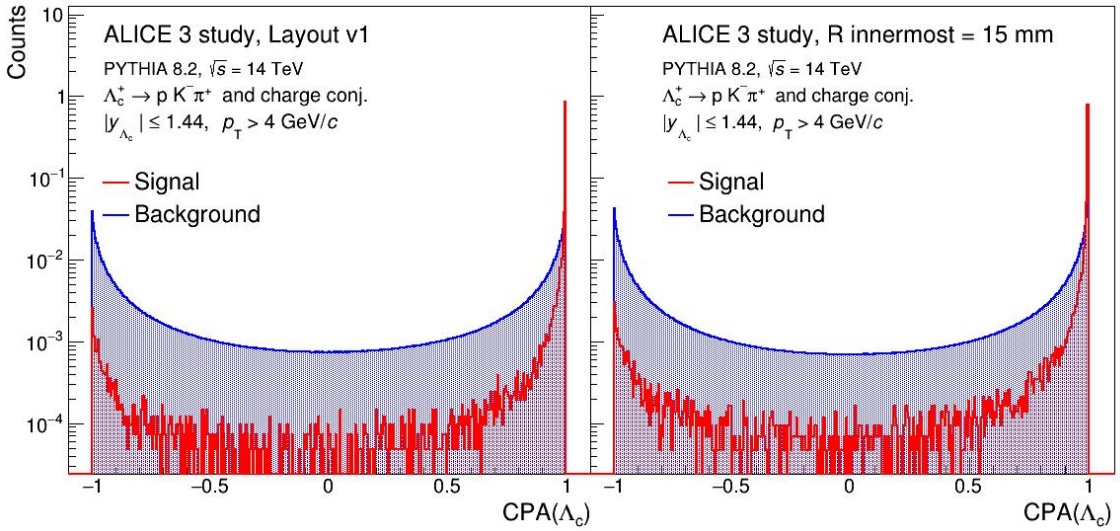

Figure 5.4: Signal and background distributions of cosine of the pointing angle of $\Lambda_c^+ \rightarrow pK^-\pi^+$ candidates with (left) and without (right) the IRIS detector at high p_T ($>$ 4 GeV/c). The distributions are normalised by their integral.

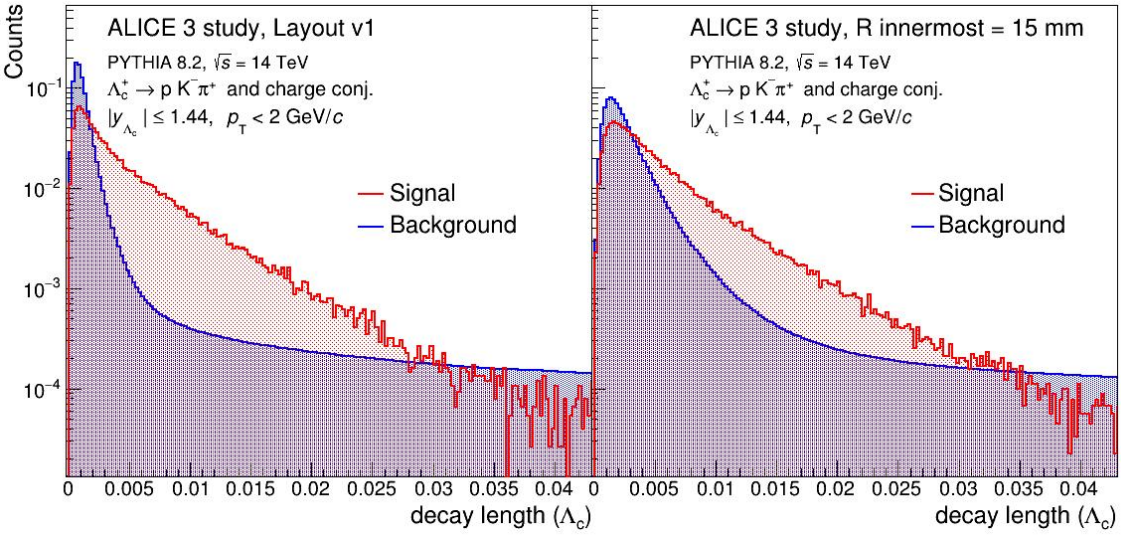

Figure 5.5: Signal and background distributions of decay length of $\Lambda_c^+ \rightarrow pK^-\pi^+$ candidates with (left) and without (right) the IRIS detector at low p_T ($<$ 2 GeV/c). The distributions are normalised by their integral.

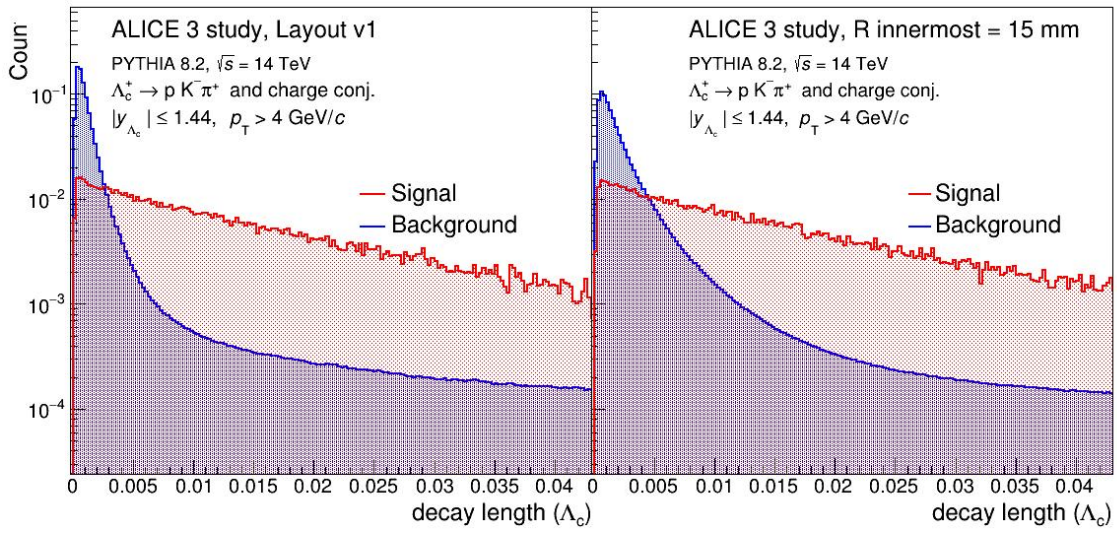

Figure 5.6: Signal and background distributions of decay length of $\Lambda_c^+ \rightarrow pK^-\pi^+$ candidates with (left) and without (right) the IRIS detector at high p_T ($>$ 4 GeV/c). The distributions are normalised by their integral.

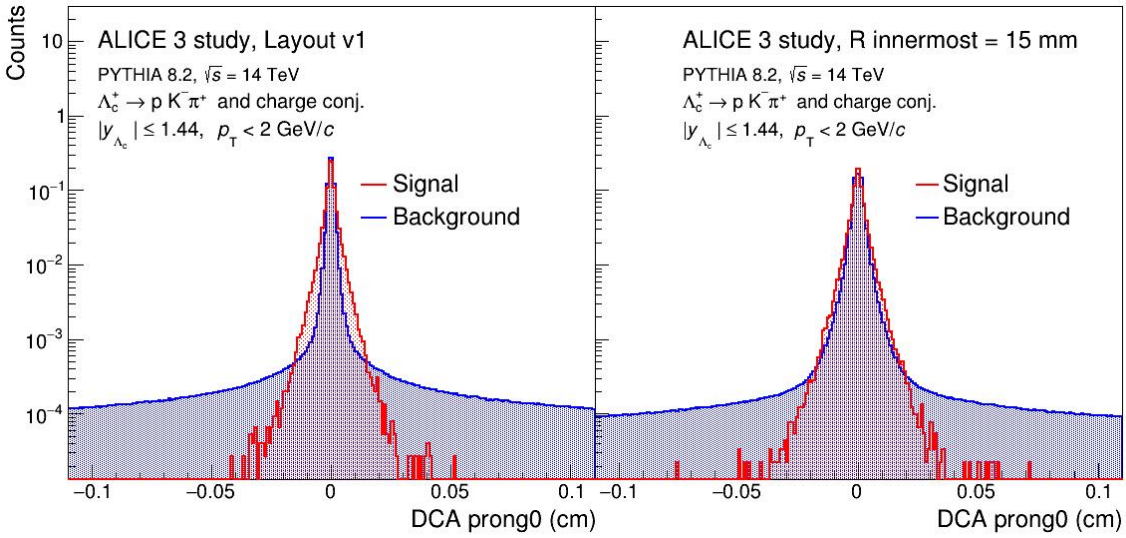

Figure 5.7: Signal and background distributions of the first prong impact parameter of $\Lambda_c^+ \rightarrow pK^-\pi^+$ candidates with (left) and without (right) the IRIS detector at low p_T ($<$ 2 GeV/c). The distributions are normalised by their integral.

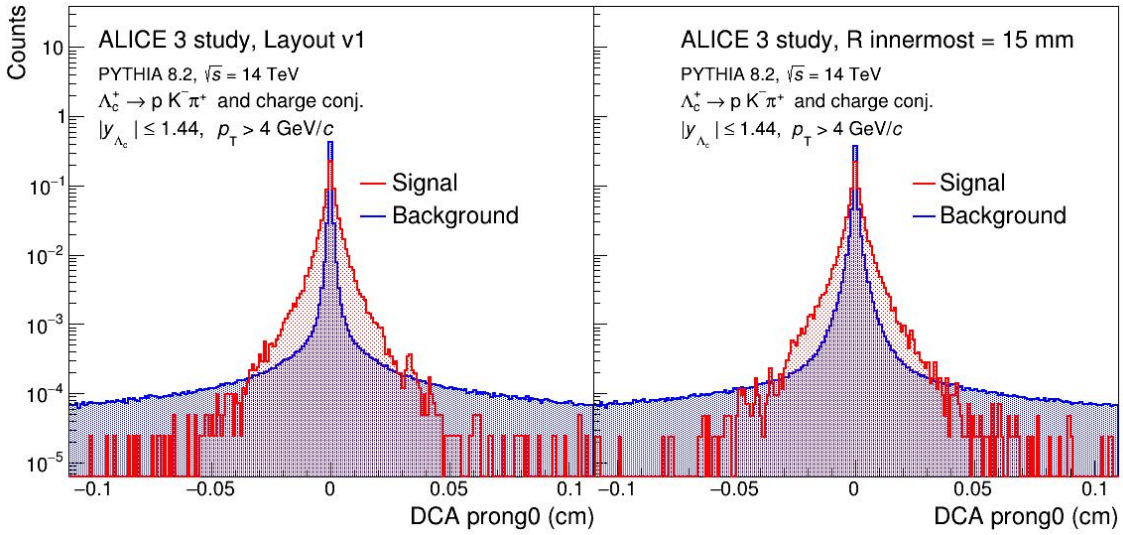

Figure 5.8: Signal and background distributions of the first prong impact parameter of $\Lambda_c^+ \rightarrow pK^-\pi^+$ candidates with (left) and without (right) the IRIS detector at high p_T ($>$ 4 GeV/c). The distributions are normalised by their integral.

As expected, the separation between the signal and background in the topological variables is more pronounced in the layout v1, in which the resolution on the impact parameter is better because the first layer of the tracker is very close to the IP. Going towards higher p_T, the distributions in the two layouts get similar. The reason is that the impact parameter resolution improves with the p_T and saturates, giving the same results in both the scenarios.

The same machine-learning approach with Boosted Decision Trees, described in section 4.4, was employed to perform the analysis. The training sample was assembled considering the background from the sidebands of the candidate invariant-mass distribution in MB MC simulation and the $\Lambda_c^+ \rightarrow pK^-\pi^+$ signal candidates from enriched MC simulations based on the `PYTHIA` 8 event generator. Independent BDTs were trained for each p_T interval. Since the first goal is to test the tracking performance of the detector, only topological variables were considered in the training. The features used are listed in the following:

- Cosine of the Λ_c pointing angle.
- Cosine of the Λ_c pointing angle in the xy plane.

- Λ_c decay length.
- Impact parameters of the Λ_c prongs (proton, kaon and pion).

The SHAP feature importance in the p_T interval $1 \leq p_T < 2$ GeV/c without and with the IRIS tracker are reported in Figs. 5.9 and 5.10, respectively.

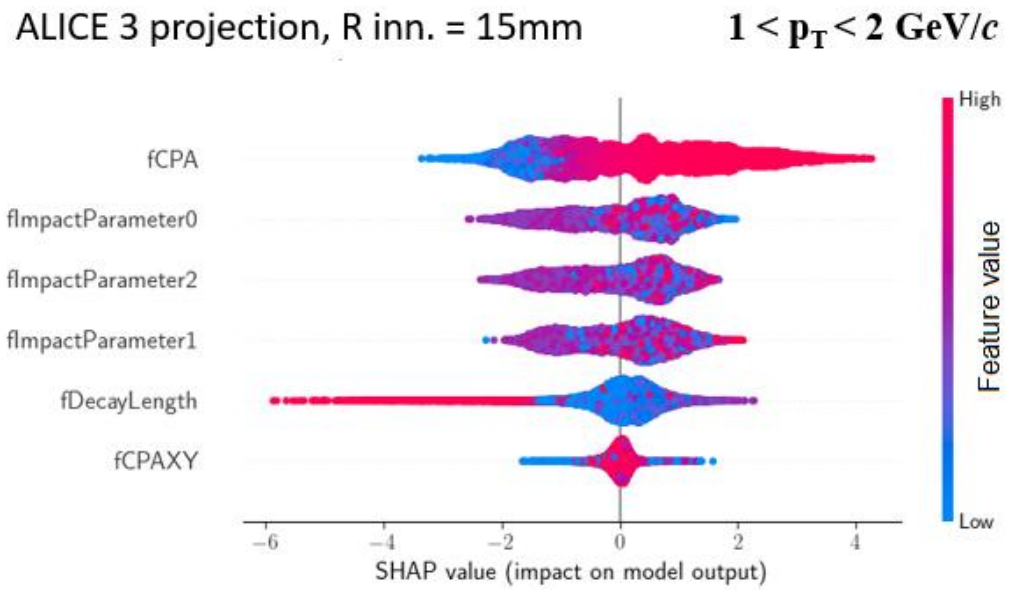

Figure 5.9: $\Lambda_c^+ \to pK^-\pi^+$ training variables discrimination power without the IRIS tracker.

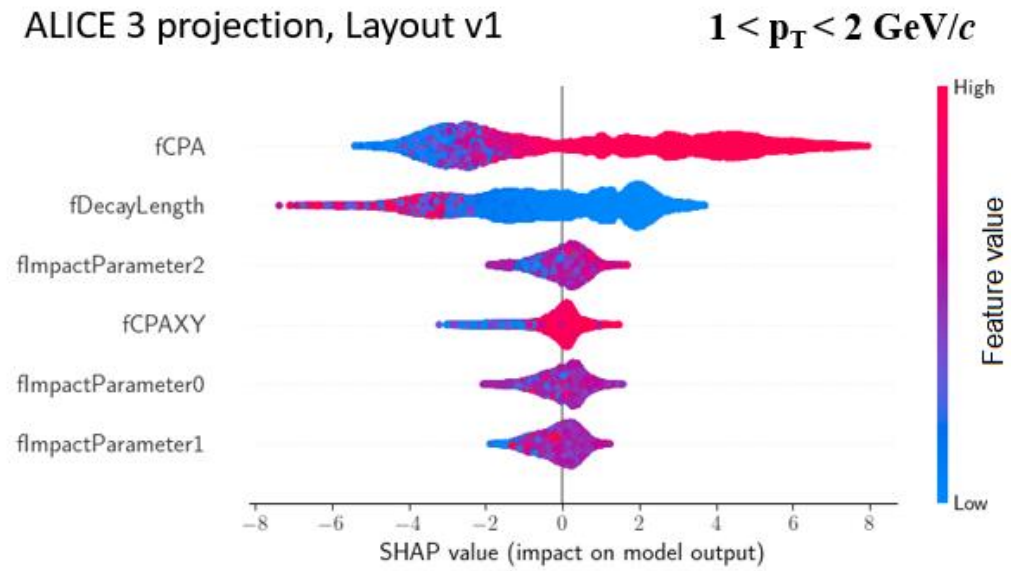

Figure 5.10: $\Lambda_c^+ \to pK^-\pi^+$ training variables discrimination power with the IRIS tracker.

It is worth noting how the variable importance plots change between the two cases. For the layout v1, the number of points around 0 for the Λ_c CPA and decay length are much less than the scenario with the innermost radius at 15 mm. Thanks to the better impact parameter resolution gained with the IRIS tracker in layout v1, the CPA and the decay length are useful variables to discriminate signal from background for most of the candidates, while their discriminating power is much lower in the other case and a lot of candidates are distributed around 0.

The trained model is then applied to the MC samples. To increase the purity of the sample the candidate with a BDT probability less than 50% were rejected.
The expected background under the signal peak was estimated from the MB MC production. The sidebands of the invariant mass distribution were fitted excluding the signal region $[M(\Lambda_c) - 3\sigma,\ M(\Lambda_c) + 3\sigma]$, where σ is the gaussian width of the $\Lambda_c^+ \to pK^-\pi^+$ signal peak, obtained fitting the invariant mass distribution in the $\Lambda_c^+ \to pK^-\pi^+$ enriched MC. The fit function is then extended to the whole range and integrated in the signal range $[M(\Lambda_c) - 3\sigma,\ M(\Lambda_c) + 3\sigma]$. An example of the invariant mass sidebands fit and the background per event under the signal region in the two considered layouts are reported in Figs. 5.11 and 5.12, respectively.

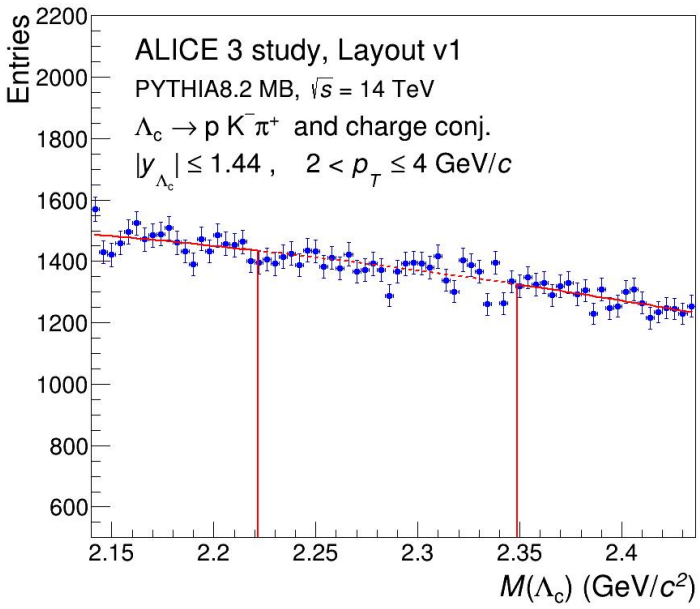

Figure 5.11: $\Lambda_c^+ \to pK^-\pi^+$ fit of the side-bands of the invariant mass distribution in the p_T range $2 \leq p_T < 4$ GeV/c .

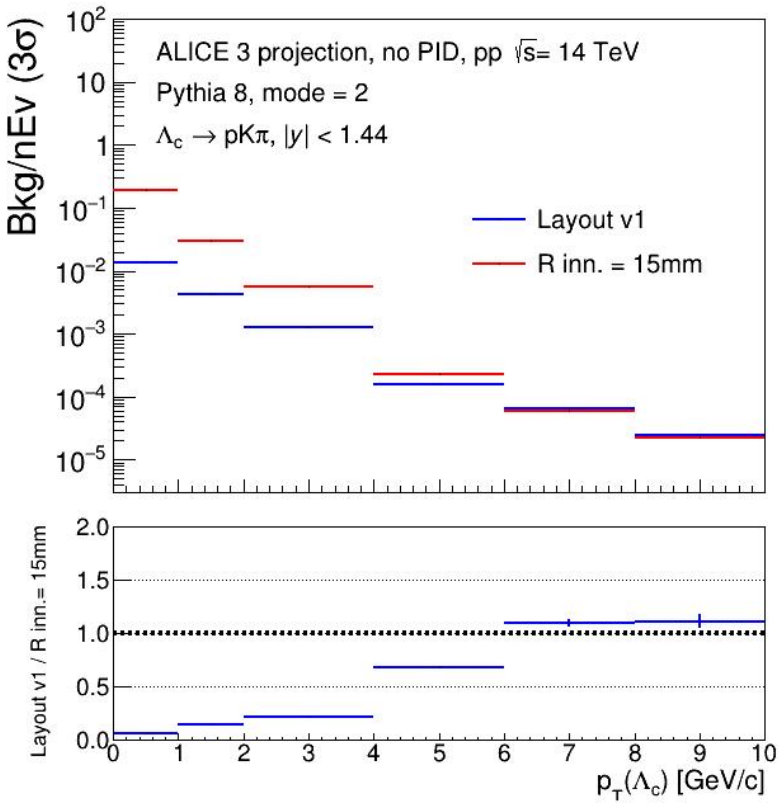

Figure 5.12: $\Lambda_c^+ \rightarrow pK^-\pi^+$ expected background normalized by the number of events with (blue) and without (red) the IRIS detector. In the bottom panel, the ratio between the two scenarios is reported.

The signal efficiency was estimated from the $\Lambda_c^+ \rightarrow pK^-\pi^+$ dedicated MC productions and the comparison between the different layouts are shown in Fig. 5.13. Looking at Figs. 5.12 and 5.13 can be concluded that the IRIS tracker provides a very strong background suppression, of about one order of magnitude at low p_T. On the other hand, the price in term of efficiency is quite limited: less than 20% for $p_T < 1$ GeV/c and less than 5% for $1 \leq p_T < 4$ GeV/c. As expected, the discrepancies between the two scenarios are negligible at high p_T.

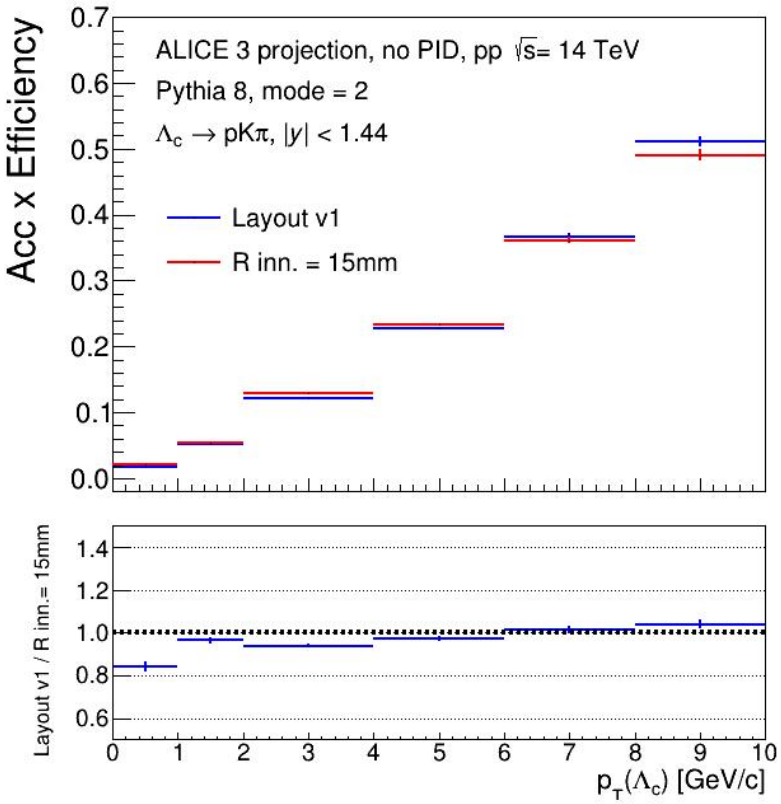

Figure 5.13: $\Lambda_c^+ \rightarrow pK^-\pi^+$ expected efficiency with (blue) and without (red) the IRIS detector. In the bottom panel, the ratio between the two scenarios is reported.

PYTHIA 8 simulations were used to estimate the $\Lambda_c^+ \rightarrow pK^-\pi^+$ cross section, assuming the integrated luminosity of 3 fb^{-1} per operational year, that is the expected one in pp collisions. Higher pp luminosities would rule out some detector and technology choices, which are crucial to realise the best possible detector for the heavy-ion programme. The expected $\Lambda_c^+ \rightarrow pK^-\pi^+$ signal can be retrieved correcting the simulated cross section by the computed expected efficiency. The signal over background per number of events and the significance computed in the two scenarios, obtained with the procedure described, are shown in Fig. 5.14, respectively on the left and right panels.

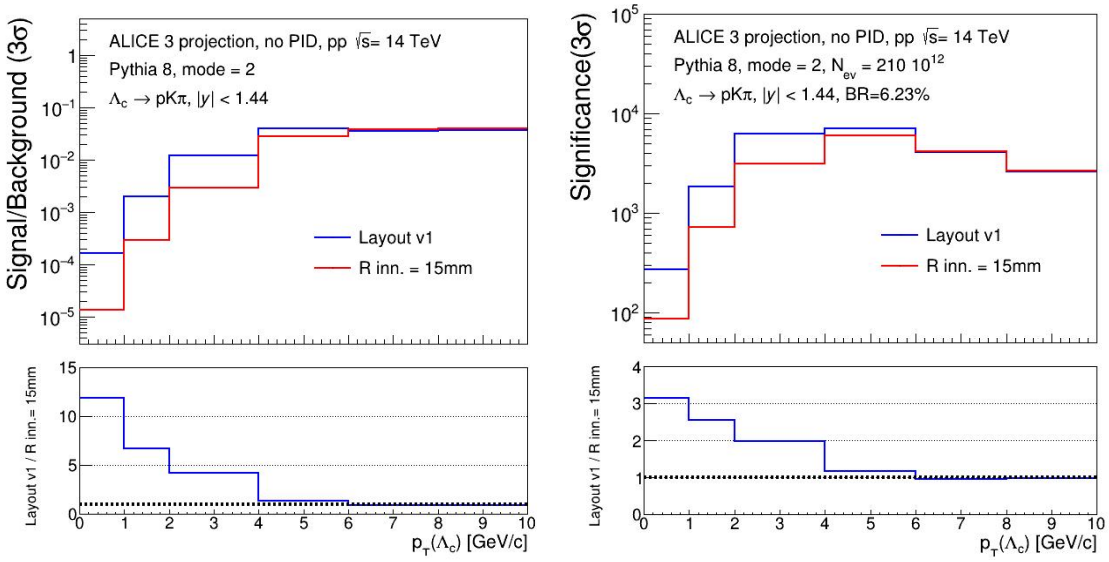

Figure 5.14: $\Lambda_c^+ \rightarrow pK^-\pi^+$ expected signal over background ratio (left) and significance (right) with (blue) and without (red) the IRIS detector. In the bottom panel, the ratio between the two scenarios is reported.

The improvement achieved thanks to the IRIS tracker, especially at low p_T, is really huge, with an increase of the S/B and of the significance by a factor 12 and 3, respectively, and similar performance at high p_T. This improvement should justify the challenging and expensive building of these additional retractable layers very close to the IP.

The computation is then repeated including the TOF to test the PID capabilities of the detector. Only the outer TOF is considered in this analysis. A flat 3σ cut on the TOF nσ distributions for the proton, the kaon and the pion hypothesis was applied on top of the BDT selection when the TOF information is available for the considered prong. The performance obtained with the layout v1 were compared with those achieved with the additional PID selections, in Figs. 5.15 and 5.16. The background per event is stongly reduced, in particular at low p_T, while the efficiency loss is less than 10%. The S/B per event and the significance are increased by a factor ∼14 and ∼3.5, respectively, at low p_T, showing the impressive PID background rejection obtained including only the outer TOF in the computation. At very high p_T the PID cannot separate anymore the different particles species and the performance

are compatible in the two cases.

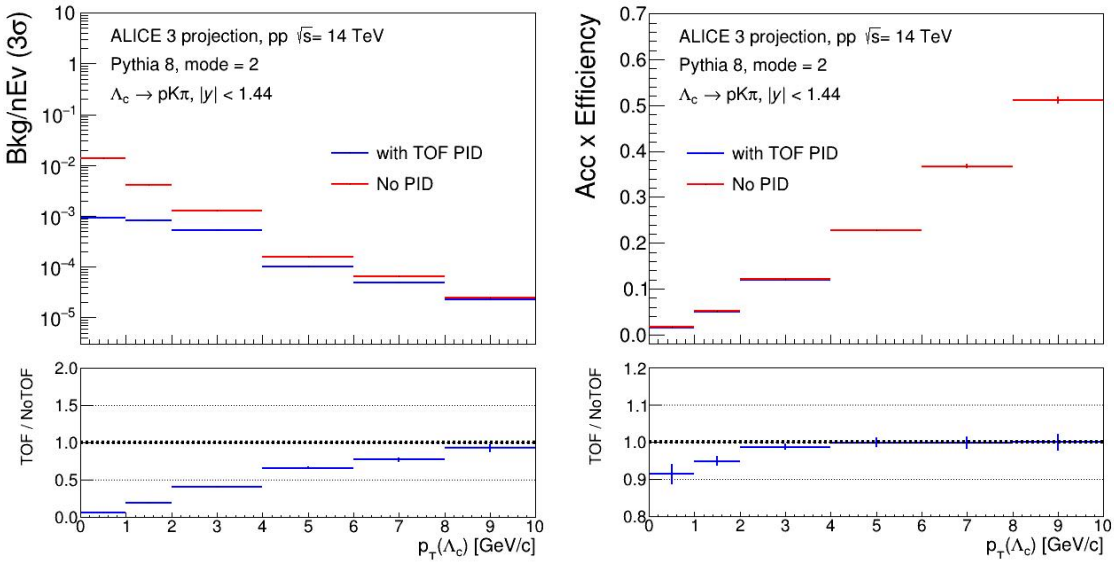

Figure 5.15: Left: $\Lambda_c^+ \to pK^-\pi^+$ expected background normalized by the number of events with (blue) and without (red) the TOF PID. Right: $\Lambda_c^+ \to pK^-\pi^+$ expected efficiency with (blue) and without (red) the TOF PID. In the bottom panels, the ratio between the two scenarios is reported.

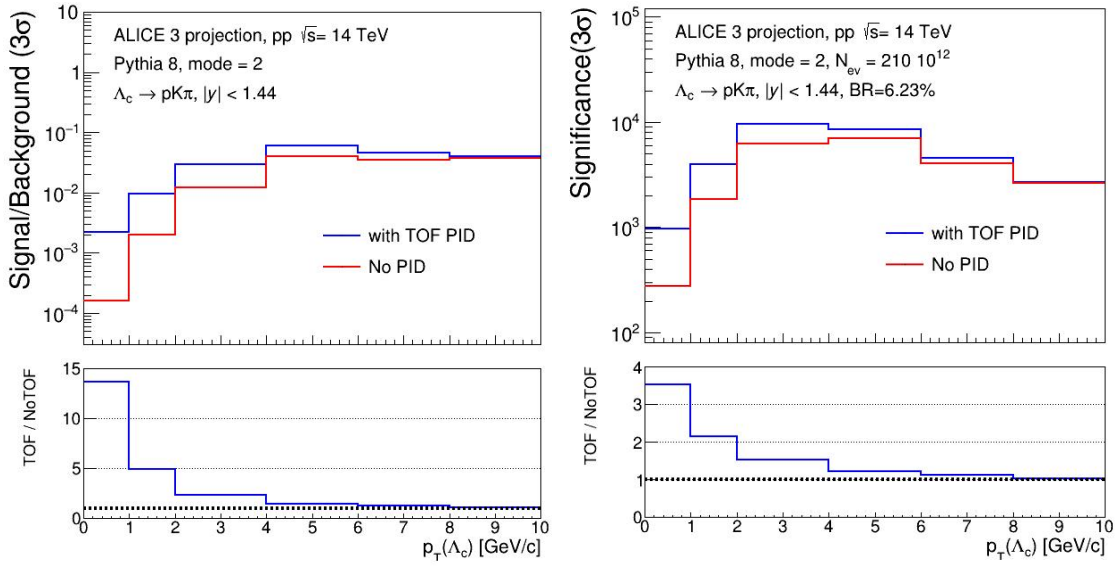

Figure 5.16: $\Lambda_c^+ \to pK^-\pi^+$ signal over background ratio (left) and expected signficance (right) with (blue) and without (red) the TOF PID. In the bottom panels, the ratio between the two scenarios is reported.

The studies carried out have shown how far the Λ_c^+ measurements can be pushed with the extraordinary vertexing and tracking resolution achieved with the IRIS tracker and the PID capabilities of the ALICE 3 detector. More differential studies will be feasible, as, for instance, the measurement versus the Λ_c^+ rapidity. The detector will be able to perform multi-charm analyses that could finally shed lights on the charm hadronisation mechanisms in small and large hadronic collisions systems.

Conclusions

In this thesis, the analysis of the Λ_c^+ production in pp collisions at $\sqrt{s} = 13$ TeV has been performed through the reconstruction of the decay channel $\Lambda_c^+ \rightarrow pK_s^0 \rightarrow p\pi^+\pi^-$. This work provided qualitatively new constraints into the mechanisms of hadronisation of heavy quarks in small systems. In particular, for the first time, the Λ_c^+/D^0 ratio has been measured as a function of the charged-particle pseudorapidity density. The p_T-differential Λ_c^+/D^0 ratios show an evident dependence on the event multiplicity, with the ratio in the low multiplicity class that is still higher than that measured in e^+e^- collisions. The ratio is strongly underestimated and the multiplicity dependence is not reproduced by the models tuned on e^+e^- measurements, pointing to additional or different hadronisation mechanisms that are not present in e^+e^- collisions. The results are qualitatively reproduced by models in which additional colour reconnection topologies in string fragmentation are allowed and by the statistical hadronisation model when the underlying charm-baryon spectrum in the calculations is augmented to include additional excited baryon states predicted by the Relativistic Quark Model. The p_T-integrated Λ_c^+/D^0 ratio has been studies as well. In this case a clear enhancement with the charged-particle density is not observed, suggesting that the increase in the p_T-differential ratio could originate from a modification of the transverse momentum distribution that acts differently for baryons and mesons.

An innovative machine learning technique based on a BDT algorithm, which combines particle identification with TOF and TPC and topological selection, has been developed and used to enhance the purity of the reconstructed spectra and to reduce the statistical uncertainty of the measurement. Such a technique was key in

accessing the very low p_{T} region, and allowed a more accurate extrapolation down to $p_{\mathrm{T}} = 0$. The results presented in this thesis have been approved by the ALICE Collaboration and published in Physics Letters B [85].

The current precision of the measurements does not allow drawing final conclusions. Several theoretical models are under investigation but more differential and precise measurements are needed for a clear discrimination. The required sensitivity could be reached during the imminent LHC Run 3 (2026-2029) and Run 4 (2031-2035) data taking periods, owing to a substantial upgrade of the ALICE detector. In particular, during the just concluded Long Shut-down 2 (2020-2021), the Inner-Tracking-System (ITS) has been replaced with the ITS 2, that will improve the resolution on the primary and secondary vertices determination and, consequently, the accuracy of the reconstruction of charmed-baryons. In general, more exotic charm heavy-flavour particles ($\Xi_c^{0,+}$, $\Sigma_c^{0,+,++}$, Ω_c^0) could be measured and studied with higher precision. They could provide further constraints to the charm hadronisation mechanisms. In this context, studies that are critical for the design of a new experiment, ALICE 3, have been carried out. One of the main physics goals of the ALICE 3 experiment will be the analysis of multi-charmed baryons. The studies that I performed using the Λ_{c}^+ reconstruction via its decay channel $\Lambda_{\mathrm{c}}^+ \rightarrow \mathrm{pK}^-\pi^+$ were critical for the design and the optimization of ALICE 3 detector setup and demonstrated its unique capabilities in the study of heavy-flavour baryons.

Bibliography

[1] W. Cottingham and D. Greenwood, *An introduction to the standard model of particle physics*. Cambridge University Press, United Kingdom, 1998.

[2] **Belle** Collaboration, S.-K. Choi *et al.*, "Observation of a Narrow Charmoniumlike State in Exclusive $B^{\pm} \to K^{\pm}\pi^{+}\pi^{-}J/\psi$ Decays", *Phys. Rev. Lett.* **91** (Dec, 2003) 262001.

[3] **LHCb** Collaboration, R. Aaij *et al.*, "Observation of $J/\Psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\Psi K^{-} p$ Decays", *Phys. Rev. Lett.* **115** (Aug, 2015) 072001.

[4] **Particle Data Group** Collaboration, P. Zyla *et al.*, "Review of Particle Physics", *PTEP* **2020** no. 8, (2020) 083C01.

[5] R. Gupta, "Introduction to lattice QCD: Course", in *Les Houches Summer School in Theoretical Physics, Session 68: Probing the Standard Model of Particle Interactions*, pp. 83–219. 7, 1997. `arXiv:hep-lat/9807028`.

[6] G. Martinez, "Advances in quark gluon plasma", 2013.

[7] O. Kaczmarek and F. Zantow, "Static quark-antiquark interactions in zero and finite temperature QCD: I. Heavy quark free energies, running coupling, and quarkonium binding", *Phys. Rev. D* **71** (Jun, 2005) 114510.

[8] L. Husdal, "On effective degrees of freedom in the early universe", *Galaxies* **4** no. 4, (2016) .

[9] R. D. Pisarski and F. Wilczek, "Remarks on the Chiral Phase Transition in Chromodynamics", *Phys. Rev. D* **29** (1984) 338–341.

[10] F. Karsch, "Lattice QCD at finite temperature and density", *Nucl. Phys. B Proc. Suppl.* **83** (2000) 14–23, arXiv:hep-lat/9909006.

[11] J. Adams *et al.*, "Experimental and theoretical challenges in the search for the quark–gluon plasma: The STAR Collaboration's critical assessment of the evidence from RHIC collisions", *Nuclear Physics A* **757** no. 1, (2005) 102–183.

[12] K. Adcox *et al.*, "Formation of dense partonic matter in relativistic nucleus–nucleus collisions at RHIC: Experimental evaluation by the PHENIX Collaboration", *Nuclear Physics A* **757** no. 1, (2005) 184–283.

[13] R. Nouicer, "New State of Nuclear Matter: Nearly Perfect Fluid of Quarks and Gluons in Heavy Ion Collisions at RHIC Energies", *Eur. Phys. J. Plus* **131** no. 3, (2016) 70, arXiv:1512.08993 [nucl-ex].

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

[15] M. L. Miller *et al.*, "Glauber Modeling in High-Energy Nuclear Collisions", *Annual Review of Nuclear and Particle Science* **57** no. 1, (2007) 205–243.

[16] **ALICE** Collaboration, S. Acharya *et al.*, "Transverse momentum spectra and nuclear modification factors of charged particles in pp, p-Pb and Pb-Pb collisions at the LHC", *JHEP* **11** (2018) 013, arXiv:1802.09145 [nucl-ex].

[17] **ALICE** Collaboration, B. Abelev *et al.*, "$K^*(892)^0$ and $\phi(1020)$ production in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV", *Phys. Rev. C* **91** (Feb, 2015) 024609.

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

[19] **NA50** Collaboration, M. C. Abreu *et al.*, "Evidence for deconfinement of quarks and gluons from the J/Ψ suppression pattern measured in Pb-Pb collisions at the CERN SPS", *Phys. Lett. B* **477** (2000) 28–36.

[20] B. Abelev *et al.*, "Centrality, rapidity and transverse momentum dependence of J/Psi suppression in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV",.

[21] **PHENIX** Collaboration, A. Adare *et al.*, "J/ψ suppression at forward rapidity in Au + Au collisions at $\sqrt{s_{NN}} = 200$ GeV", *Phys. Rev. C* **84** (Nov, 2011) 054912.

[22] S. Sarkar *et al.*, *The Physics of the Quark-Gluon Plasma*. Springer, Berlin, Heidelberg, 2010.

[23] **STAR** Collaboration, J. Adams *et al.*, "Evidence from d + Au measurements for final state suppression of high p_{T} hadrons in Au+Au collisions at RHIC", *Phys. Rev. Lett.* **91** (2003) 072304, arXiv:nucl-ex/0306024.

[24] M. He and R. Rapp, "Charm-baryon production in proton–proton collisions", *Phys. Lett. B* **795** (2019) 117–121, arXiv:1902.08889 [nucl-th].

[25] A. Andronic, P. Braun-Munzinger, K. Redlich, and J. Stachel, "Hadron yields, the chemical freeze-out and the QCD phase diagram", *Journal of Physics: Conference Series* **779** (Jan, 2017) 012012.

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

[27] J. Pumplin *et al.*, "New Generation of Parton Distributions with Uncertainties from Global QCD Analysis", *Journal of High Energy Physics* **2002** no. 07, (Jul, 2002) 012–012. https://doi.org/10.1088/1126-6708/2002/07/012.

[28] **NNPDF** Collaboration, R. Ball *et al.*, "Parton distributions from high-precision collider data", *Eur. Phys. J. C* **77** (Jun, 2017) 663. 95 p, arXiv:1706.00428.

[29] X. Ji, "Parton Physics from Large-Momentum Effective Field Theory", *Sci. China Phys. Mech. Astron.* **57** (2014) 1407–1412, arXiv:1404.6680 [hep-ph].

[30] V. Kartvelishvili, A. Likhoded, and V. Petrov, "On the fragmentation functions of heavy quarks into hadrons", *Physics Letters B* **78** no. 5, (1978) 615–617.

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

[32] K. Eskola, P. Paakkinen, H. Paukkunen, and C. Salgado, "EPPS16: nuclear parton distributions with LHC data", *The European Physical Journal C* **77** (12, 2016) .

[33] B. Kopeliovich *et al.*, "Cronin effect in hadron production off nuclei", *Physical review letters* **88** (07, 2002) 232303.

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

[35] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 Tune", *Eur. Phys. J.* **C74** no. 8, (2014) 3024, arXiv:1404.5630 [hep-ph].

[36] G. Marchesini *et al.*, "HERWIG 5.1 - a Monte Carlo event generator for simulating hadron emission reactions with interfering gluons", *Computer Physics Communications* **67** no. 3, (1992) 465–508.

[37] S. Frixione, G. Ridolfi, and P. Nason, "A positive-weight next-to-leading-order monte carlo for heavy flavour hadroproduction", *Journal of High Energy Physics* **2007** no. 09, (Sep, 2007) 126–126.

[38] X.-N. Wang and M. Gyulassy, "Hijing: A monte carlo model for multiple jet production in pp, pA, and AA collisions", *Phys. Rev. D* **44** (Dec, 1991) 3501–3516.

[39] S. Ferreres-Solé and T. Sjöstrand, "The space–time structure of hadronization in the Lund model", *The European Physical Journal C* **78** (11, 2018) .

[40] **ALICE** Collaboration, S. Acharya *et al.*, "Λ_c^+ production and Baryon-to-Meson Ratios in pp and p–Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV at the LHC", *Phys. Rev. Lett.* **127** (2021) 202301, arXiv:2011.06078 [nucl-ex].

[41] **ALICE** Collaboration, S. Acharya *et al.*, "Λ_c^+ production in pp and in p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV", *Phys. Rev. C* **104** (2021) 054905, arXiv:2011.06079 [nucl-ex].

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

[43] S. Plumari, V. Minissale, S. K. Das, G. Coci, and V. Greco, "Charmed Hadrons from Coalescence plus Fragmentation in relativistic nucleus-nucleus collisions at RHIC and LHC", *Eur. Phys. J. C* **78** no. 4, (2018) 348, arXiv:1712.00730 [hep-ph].

[44] V. Minissale, F. Scardina, and V. Greco, "Hadrons from coalescence plus fragmentation in AA collisions at energies available at the BNL Relativistic Heavy Ion Collider to the CERN Large Hadron Collider", *Phys. Rev. C* **92** (Nov, 2015) 054904.

[45] V. Minissale, F. Scardina, and V. Greco, "Hadrons from coalescence plus fragmentation in A A collisions at energies available at the BNL Relativistic Heavy Ion Collider to the CERN Large Hadron Collider", *Physical Review C* **92** (11, 2015) .

[46] **ALICE** Collaboration, S. Acharya *et al.*, "First measurement of Ξ_c^0 production in pp collisions at $\sqrt{s} = 7$ TeV", *Physics Letters B* **781** (2018) 8–19.

[47] **ATLAS** Collaboration, Y. Chen, "Open heavy flavour production via semi-leptonic decay muons in lead lead collisions at $\sqrt{s_{NN}} = 2.76$ TeV with the ATLAS detector at the LHC", *Nucl. Phys. A* **910-911** (2013) 351–354, arXiv:1209.6282 [nucl-ex].

[48] **ALICE** Collaboration, S. Acharya *et al.*, "Measurement of the production of charm jets tagged with D^0 mesons in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **08** (2019) 133, arXiv:1905.02510 [nucl-ex].

[49] **ALICE** Collaboration, F. Catalano, "Open heavy-flavour production from small to large collision systems with ALICE at the LHC", *PoS* **HardProbes2020** (Jul, 2020) 055. 4 p, arXiv:2007.13419.

[50] **ALICE** Collaboration, B. Abelev *et al.*, "Measurement of electrons from semileptonic heavy-flavor hadron decays in *pp* collisions at $\sqrt{s} = 2.76$ TeV", *Phys. Rev. D* **91** (Jan, 2015) 012001.

[51] B. Kniehl, G. Kramer, I. Schienbein, and H. Spiesberger, "Collinear subtractions in hadroproduction of heavy quarks", *European Physical Journal C - EUR PHYS J C* **41** (02, 2005) .

[52] R. Maciuła and A. Szczurek, "Open charm production at the LHC: k_t-factorization approach", *Phys. Rev. D* **87** (May, 2013) 094022.

[53] Lisovyi, Mykhailo, Verbytskyi, Andrii, and Zenaiev, Oleksandr, "Combined analysis of charm-quark fragmentation-fraction measurements", *EPJ Web of Conferences* **120** (2016) 03002. https://doi.org/10.1051/epjconf/201612003002.

[54] **ALICE** Collaboration, S. Acharya *et al.*, "Constraining hadronization mechanisms with Λ_c^+/D^0 production ratios in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV", arXiv:2112.08156 [nucl-ex].

[55] D. Ebert, R. N. Faustov, and V. O. Galkin, "Spectroscopy and regge trajectories of heavy baryons in the relativistic quark-diquark picture", *Phys. Rev. D* **84** (Jul, 2011) 014025.

[56] V. Minissale, S. Plumari, and V. Greco, "Charm hadrons in pp collisions at LHC energy within a coalescence plus fragmentation approach", *Physics Letters B* **821** (2021) 136622.

[57] **ALICE** Collaboration, "The ALICE experiment at the CERN LHC", *Journal of Instrumentation* **3** no. 08, (Aug, 2008) S08002–S08002.

[58] L. Evans and P. Bryant, "LHC machine", *Journal of Instrumentation* **3** no. 08, (Aug, 2008) S08001–S08001.

[59] **ALICE** Collaboration, J. Adam *et al.*, "Determination of the event collision time with the ALICE detector at the LHC", *The European Physical Journal Plus* **132** (02, 2017) .

[60] **ALICE** Collaboration, B. Abelev *et al.*, "Centrality determination of Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV with ALICE", *Phys. Rev. C* **88** (Oct, 2013) 044909.

[61] M. L. Miller, K. Reygers, S. J. Sanders, and P. Steinberg, "Glauber modeling in high-energy nuclear collisions", *Annual Review of Nuclear and Particle Science* **57** no. 1, (2007) 205–243.

[62] R. Brun and F. Rademakers, "Root — an object oriented data analysis framework", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **389** no. 1, (1997) 81–86. New Computing Techniques in Physics Research V.

[63] R. Brun *et al.*, *GEANT: Detector Description and Simulation Tool*. CERN Program Library. CERN, Geneva, 1993. `https://cds.cern.ch/record/1082634`. Long Writeup W5013.

[64] S. Agostinelli *et al.*, "Geant4—a simulation toolkit", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** no. 3, (2003) 250–303.

[65] A. Ferrari *et al.*, "FLUKA: A multi-particle transport code (Program version 2005)",.

[66] P. Saiz *et al.*, "Alien—alice environment on the grid", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **502** no. 2, (2003) 437–440. Proceedings of the VIII International Workshop on Advanced Computing and Analysis Techniques in Physics Research.

[67] R. Frühwirth, "Application of kalman filtering to track and vertex fitting", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **262** no. 2, (1987) 444–450.

[68] **ALICE** Collaboration, S. Acharya *et al.*, "Pseudorapidity distributions of charged particles as a function of mid- and forward rapidity multiplicities in pp collisions at $\sqrt{s} = 5.02$, 7 and 13 TeV", *Eur. Phys. J. C* **81** no. 7, (2021) 630, arXiv:2009.09434 [nucl-ex].

[69] **ALICE** Collaboration, J. Adam *et al.*, "Charged-particle multiplicities in proton–proton collisions at $\sqrt{s} = 0.9$ to 8 TeV", *Eur. Phys. J. C* **77** no. 1, (2017) 33, arXiv:1509.07541 [nucl-ex].

[70] **ALICE** Collaboration, "The ALICE definition of primary particles",. https://cds.cern.ch/record/2270008.

[71] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System", KDD '16, p. 785–794. Association for Computing Machinery, New York, NY, USA, 2016. https://doi.org/10.1145/2939672.2939785.

[72] **LHCb** Collaboration, R. Aaij *et al.*, "Measurement of b hadron fractions in 13 TeV pp collisions", *Phys. Rev. D* **100** (Aug, 2019) 031102.

[73] T. Sjöstrand, S. Mrenna, and P. Skands, "A brief introduction to pythia 8.1", *Computer Physics Communications* **178** no. 11, (2008) 852–867.

[74] O. Behnke, K. Kröninger, T. Schörner-Sadenius, and G. Schott, eds., *Data analysis in high energy physics*. Wiley-VCH, Weinheim, Germany, 2013. http://www.wiley-vch.de/publish/dt/books/ISBN3-527-41058-9.

[75] **ALICE** Collaboration, S. Acharya *et al.*, "Multiplicity dependence of (multi-)strange hadron production in proton–proton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **80** no. 2, (2020) 167, arXiv:1908.01861 [nucl-ex].

[76] **ALICE** Collaboration, S. Acharya *et al.*, "Multiplicity dependence of π, K, and p production in pp collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **80** no. 8, (2020) 693, arXiv:2003.02394 [nucl-ex].

[77] **ALICE** Collaboration, S. Acharya *et al.*, "Multiplicity dependence of light-flavor hadron production in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. C* **99** no. 2, (2019) 024906, arXiv:1807.11321 [nucl-ex].

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

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

[80] **ALICE** Collaboration, S. Acharya *et al.*, "Measurement of prompt D^0, Λ_c^+, and $\Sigma_c^{0,++}(2455)$ production in pp collisions at $\sqrt{s} = 13$ TeV", arXiv:2106.08278 [hep-ex].

[81] Y. Chen and M. He, "Charged-particle multiplicity dependence of charm-baryon-to-meson ratio in high-energy proton–proton collisions", *Phys. Lett. B* **815** (2021) 136144, arXiv:2011.14328 [hep-ph].

[82] B. Abelev *et al.*, "Technical design report for the upgrade of the ALICE inner tracking system", *Journal of Physics G: Nuclear and Particle Physics* **41** no. 8, (Jul, 2014) 087002.

[83] M. Selvaggi *et al.*, "Delphes 3: A modular framework for fast-simulation of generic collider experiments", *Journal of Physics: Conference Series* **523** (06, 2014) 012033.

[84] "An introduction to pythia 8.2", *Computer Physics Communications* **191** (2015) 159–177.

[85] S. Acharya *et al.*, "Observation of a multiplicity dependence in the p_{T}-differential charm baryon-to-meson ratios in proton–proton collisions at $\sqrt{s} = 13$ TeV", *Physics Letters B* **829** (2022) 137065.