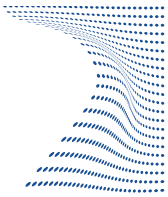

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR



Doctoral Dissertation
Doctoral Program in Physics (34th cycle)

# Strange D-meson measurements in pp and Pb–Pb collisions exploiting machine learning with ALICE at the LHC

**Fabio Catalano**

* * * * * *

**Supervisors**
Prof. Stefania Bufalino, Supervisor
Dr. Francesco Prino, Co-supervisor

**Doctoral Examination Committee:**

Prof. Donatella Lucchesi, Referee, Università degli Studi di Padova

Dr. Patrick Robbe, Referee, Laboratoire de Physique des 2 Infinis Irène Joliot-Curie

Prof. Andrea Lavagno, Politecnico di Torino

Prof. Massimo Masera, Università degli Studi di Torino

Prof. Ernesto Migliore, Università degli Studi di Torino

Politecnico di Torino
May 19, 2022

I hereby declare that, the contents and organisation of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

. . . . . .

Fabio Catalano
Turin, May 19, 2022

**Abstract**

The ALICE experiment at the CERN Large Hadron Collider is focused on the study of the quark-gluon plasma (QGP), a deconfined state of the strongly-interacting matter which is formed in ultrarelativistic heavy-ion collisions. Heavy flavours, i.e., charm and beauty quarks, are produced in hard-scattering processes on a shorter timescale than the one of the QGP formation. They experience the full evolution of the system, propagating through the deconfined medium and losing energy via elastic and inelastic scatterings with the medium constituents. Therefore, heavy flavours are ideal probes of the QGP, and the measurements of open heavy-flavour hadrons provide relevant information on the QGP properties. Moreover, the comparison of the relative abundances of heavy-flavour hadron species with and without strange-quark content is a tool to investigate the role of the heavy-quark hadronisation via recombination with light quarks from the medium.

In proton–proton (pp) collisions, measurements of the production of hadrons containing charm or beauty quarks are an important benchmark of perturbative quantum chromodynamics (QCD) calculations. They test the validity of the factorisation theorem, where the heavy-flavour hadron production is described as a convolution of perturbative and non-perturbative QCD processes. Finally, these measurements provide insights on the heavy-flavour hadronisation via fragmentation in the vacuum.

This Thesis is devoted to the measurement of $\mathrm{D_s^+}$ mesons, containing a charm and an antistrange quark, in pp and Pb–Pb collisions. They are reconstructed in the fully-hadronic decay channel $\mathrm{D_s^+ \to \phi\pi^+ \to K^-K^+\pi^+}$. Supervised machine-learning techniques are adopted to improve the measurement precision and transverse-momentum coverage. In addition, these techniques are exploited to separate prompt $\mathrm{D_s^+}$ mesons, produced in the charm-quark hadronisation or in the decays of excited charm-hadron states, and non-prompt ones, produced in beauty-hadron decays.

The measurements of prompt and non-prompt $\mathrm{D_s^+}$-meson production in pp collisions at a centre-of-mass energy of $\sqrt{s} = 5.02$ TeV are described by perturbative QCD calculations. The fragmentation fraction of heavy quarks into strange mesons divided by the one into non-strange mesons, $f_\mathrm{s}/(f_\mathrm{u} + f_\mathrm{d})$, is found to be compatible for charm and beauty quarks and with previous measurements at different centre-of-mass energies and collision systems, supporting a universal fragmentation of heavy-flavour quarks into mesons. The estimated $\mathrm{c\overline{c}}$ and $\mathrm{b\overline{b}}$ production cross sections at midrapidity are in agreement with theoretical calculations.

The production of prompt $\mathrm{D_s^+}$ mesons is measured in central and semicentral Pb–Pb collisions at a centre-of-mass energy per nucleon pair of $\sqrt{s_\mathrm{NN}} = 5.02$ TeV, and the $p_\mathrm{T}$-integrated yield is compatible with the prediction of the statistical hadronisation model for charm hadrons. A strong suppression of the $\mathrm{D_s^+}$-meson

nuclear modification factor $R_{\rm AA}$ at intermediate and high $p_{\rm T}$, resulting from the charm-quark energy loss in the QGP, is observed. The $\rm D_s^+/D^0$ production-yield ratio measured in Pb–Pb collisions is on average higher than that in pp collisions in the interval $2 < p_{\rm T} < 8$ GeV/$c$. This enhancement is expected if charm quarks hadronise also via recombination in the deconfined medium, where the production of strange quarks is enhanced. The azimuthal anisotropy of prompt $\rm D_s^+$ mesons was studied in semicentral Pb–Pb collisions, and the measured $\rm D_s^+$-meson elliptic flow $v_2$ is found to be compatible with that of non-strange D mesons. The main features of the measured $R_{\rm AA}$, $\rm D_s^+/D^0$ ratio, and $v_2$ as a function of $p_{\rm T}$ are described by theoretical calculations of charm-quark transport in a hydrodynamically expanding QGP including the hadronisation via the recombination mechanism.

The results presented in this Thesis were approved by the ALICE Collaboration and were presented in various international conferences. The measurements of the production of prompt and non-prompt $\rm D_s^+$ mesons in pp collisions have been published, as well as the measurements of the prompt $\rm D_s^+$-meson $R_{\rm AA}$ and $v_2$ in Pb–Pb collisions.

Keywords: heavy-flavour, charm, beauty, strange D mesons, nuclear modification factor, elliptic flow, hadronisation, strangeness enhancement, quark-gluon plasma, ALICE experiment, machine learning

# List of Publications

1. **Measurement of beauty and charm production in pp collisions at $\sqrt{s} = 5.02$ TeV via non-prompt and prompt D mesons**
   ALICE Collaboration
   Paper committee members: Mengke Cai, Fabio Catalano (chair), Fabrizio Grosa, Xinye Peng, Francesco Prino, Andrea Rossi
   DOI: 10.1007/JHEP05(2021)220
   arXiv: 2102.13601

2. **Measurement of prompt $\mathrm{D_s^+}$-meson production and azimuthal anisotropy in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV**
   ALICE Collaboration
   Paper committee members: Fabio Catalano (chair), Fabrizio Grosa, Francesco Prino
   DOI: 10.1016/j.physletb.2022.136986
   arXiv: 2110.10006

3. **Measurement of D-meson production and flow in Pb–Pb collisions with ALICE at the LHC**
   Fabio Catalano (on behalf of the ALICE Collaboration)
   DOI: 10.1142/S0218301320400042

4. **Open heavy-flavour production from small to large collision systems with ALICE at the LHC**
   Fabio Catalano (on behalf of the ALICE Collaboration)
   DOI: 10.22323/1.387.0055
   arXiv: 2007.13419

# Acknowledgements

I would like to acknowledge all the people who played a role in my professional and personal development during these Ph.D. years. First, I would like to thank my supervisors Francesco Prino and Stefania Bufalino, for all the support they provided to my activities and the huge amount of things they taught me about Physics, analysis techniques, and research. A special thank goes to Michelangelo Agnello, who involved me in his teaching activities and encouraged me to attend conferences. Then, I would like to acknowledge Fabrizio Grosa and Stefano Politanò who shared with me the endeavour of exploring machine learning techniques applied to heavy-flavour measurements. Thanks also to all my colleagues of the ALICE Turin group and D2H for the fruitful collaborations and the nice time spent together. Last but not least, I would like to thank my family and Chiara, without which I would not have reached this achievement.

# Contents

# Chapter 1

# Quark-gluon plasma and open heavy flavours

In the Standard Model, the strong interaction among quarks and gluons, which are the elementary constituents of hadronic matter, is described by the Quantum Chromodynamics (QCD) field theory. At extremely high temperatures and energy densities, lattice QCD (lQCD) calculations and effective phenomenological models based on QCD predict a transition of the hadronic matter to a colour-deconfined medium called quark-gluon plasma (QGP). This exotic state of hadronic matter is created and studied in the laboratory by colliding heavy ions at ultrarelativistic energies [1]. In this Chapter, an introduction to the topic of high-energy nuclear physics is presented, together with a selection of the principal experimental results providing signatures of the QGP formation. Subsequently, the study of open-heavy flavour hadrons is introduced. Open-heavy flavour hadrons contain at least a charm or a beauty quark, and are powerful probes of the QGP [2].

## 1.1 Quantum chromodynamics

Quantum Chromodynamics is a non-Abelian quantum gauge field theory, where the Lagrangian is invariant under local $SU(3)$-group transformations [3, 4]. This symmetry of the Lagrangian leads to the introduction of a conserved charge denominated *colour* which can assume three different values, i.e., *red*, *blue*, and *green*. In the theory, quarks are characterised by their colour charge, and the interactions are mediated by massless vector gauge bosons called gluons, which carry a colour and an anticolour charge. The QCD Lagrangian density can be written as

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

where the sum runs over the six quark flavours (up, down, strange, charm, beauty, and top), and $q_f^l$ is a quark-field with flavour $f$, colour charge $l$, and mass $m_f$. The

covariant derivative $D^{\mu}_{kl}$ introduces the interaction between quarks and gluons, and it is expressed as

$$D^{\mu}_{kl} = \partial^{\mu}\delta_{kl} - ig_s\frac{\lambda^a_{kl}}{2}A^{\mu}_a, \tag{1.2}$$

where $A^{\mu}_a$ are the gluon fields, corresponding to the eight possible colour configurations, and $\lambda^a_{kl}$ are the Gell-Mann matrices. The constant $g_s$ quantifies the intensity of the interaction, and it is related to the *strong coupling constant* $\alpha_s$ by $g_s = \sqrt{4\pi\alpha_s}$. The second term of Eq. (1.1) describes the kinematics and dynamics of the gluons, and $G^a_{\mu\nu}$ can be expanded as

$$G^a_{\mu\nu} = \partial_{\mu}A^a_{\nu} - \partial_{\nu}A^a_{\mu} + g_s f_{abc}A^b_{\mu}A^c_{\nu}, \tag{1.3}$$

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

The peculiar characteristics of the strong interaction can be understood by studying the dependence of the strong coupling constant $\alpha_s$ on the energy scale of the interaction. The evolution of the $\alpha_s$ value as a function of the squared transferred momentum $Q^2$ is expressed by

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

where $\mu$ is the mass scale of the renormalisation, and $n_f$ is the number of quark flavours with mass smaller than $\mu$. This $Q^2$ dependence of $\alpha_s$ is related to the renormalisation process used to remove divergences in the perturbative approach to QCD (pQCD), and it is due to the inclusion of quantum corrections, such as vertex corrections, self-energy, and vacuum polarisation, in the coupling-constant value.

The $\alpha_s$ evolution described by Eq. (1.4) is confirmed by experimental measurements as shown in Fig. 1.1. The value of $\alpha_s$ decreases with increasing transferred momentum, and two different regimes can be distinguished: *asymptotic freedom* and *confinement*. At high $|Q|$ values, $\alpha_s$ is small and asymptotic freedom is reached, i.e., quarks and gluons constituting hadrons can be considered free and weakly interacting. In this condition, pQCD calculations of scattering amplitudes based on a perturbative expansion in $\alpha_s$ are possible. For $|Q| < 1$ GeV/$c$, the strong coupling constant diverges, and quarks and gluons are strongly bound in hadrons leading to the confinement regime. The large $\alpha_s$ value prevents the application of the perturbative technique in this regime, and QCD calculations can be performed on a discretised space-time grid with the lattice QCD approach [5].

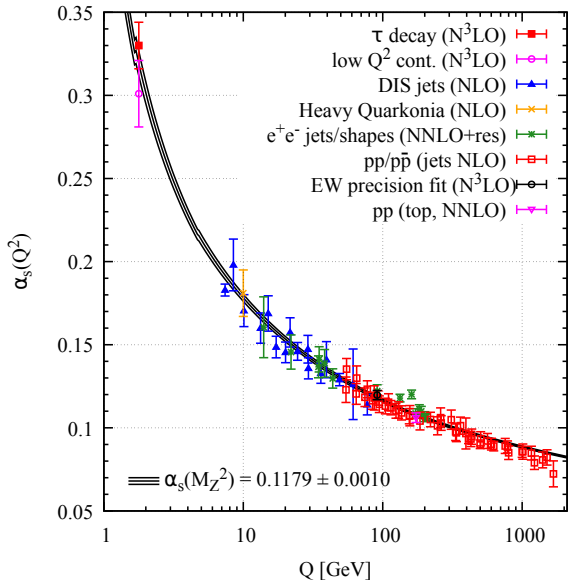



Figure 1.1: Summary of $\alpha_s$ measurements as a function of the transferred momentum $Q$. The respective degree of QCD perturbation theory used in the extraction of $\alpha_s$ is indicated in parentheses. Figure from Ref. [6].

## 1.2 Deconfinement of hadronic matter

The strong dependence of the $\alpha_s$ value on $Q^2$ points to the existence of distinct phases of the strongly-interacting matter, depending on the average momentum exchanged in the interactions between its constituents. The properties of strongly-interacting matter can be studied through lQCD calculations as a function of the system temperature $T$ and baryon-chemical potential $\mu_B$ [7, 8], which is defined as the energy needed to increase the total baryon number by one unit and is proportional to the net baryon density.

Figure 1.2 displays the phase diagram of QCD matter as a function of the $T$ and $\mu_B$ values. The ordinary nuclear matter resides in the region with approximatively $T \approx 0$ and $\mu_B \approx 1$ GeV. At higher values of the baryon-chemical potential, the nuclear matter can be described as a degenerate gas of neutrons with a structure similar to the atomic nucleus one. This matter is expected to compose the core of neutron stars. Further increasing the $\mu_B$ value, a low-temperature gas in which quarks are no more confined inside hadrons is predicted, and it is referred to as a colour-superconductive state [10]. The nuclear matter converts to a hadron gas at high temperatures, and above $T \approx 100$–$150$ MeV, it is expected to undergo a transition to the quark-gluon plasma. The QGP is a state of strongly-interacting matter in which the colour degrees of freedom are not confined in colour-singlet hadrons, and quarks and gluons are free to move over a larger volume of space. The nature of the transition from hadronic matter to QGP can be investigated

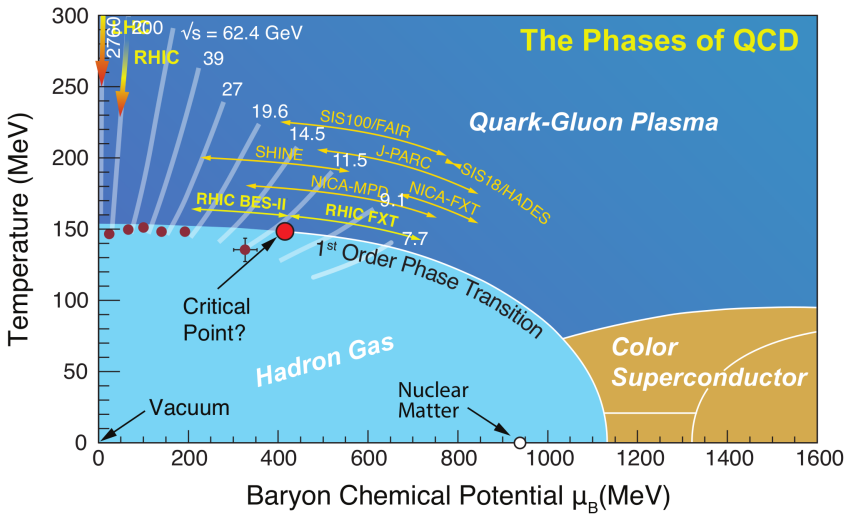



Figure 1.2: The QCD phase diagram as a function of the temperature and the baryon-chemical potential. Figure from Ref. [9].

using lattice QCD calculations. In the low-$\mu_B$ limit, a crossover between the two phases is expected as a consequence of the chiral-symmetry restoration at high energies [11]. At higher $\mu_B$, a phase transition of the first order is predicted to occur, implicating the presence of a critical point due to the coexistence with the crossover.

For vanishing baryon-chemical potentials, $\mu_B \approx 0$, the QGP is predicted to be formed at the pseudo-critical temperature $T_{pc} \approx 154$–$158$ MeV, which is equivalent to about $1.8 \times 10^{12}$ K and corresponds to an energy density $\epsilon_{pc}$ of about 0.42 GeV/fm$^3$ [11, 12]. These conditions are attained in the initial stages of ultrarelativistic heavy-ion collisions, and existing measurements indicate the QGP as a strongly-coupled liquid-like system characterised by a remarkably low shear viscosity to entropy density ratio $\eta/s \approx 1/4\pi$ [1].

## 1.3 Heavy-ion collisions

As mentioned in the previous Section, the conditions required for the QGP formation can be achieved experimentally via heavy-ion collisions at ultrarelativistic energies. The main experiments where high-energy heavy-ion collisions are currently studied are located at the Relativistic Heavy Ion Collider (RHIC) of the Brookhaven National Laboratory (BNL), and at the Large Hadron Collider (LHC) of the European Organisation for Nuclear Research (CERN). The maximum values of the centre-of-mass energy per nucleon pair reached are $\sqrt{s_{\mathrm{NN}}} = 200$ GeV in Au–Au collisions at the RHIC, and $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV in Pb–Pb collisions at the LHC. For such high energies, the colliding nuclei are not stopped at the collision point,

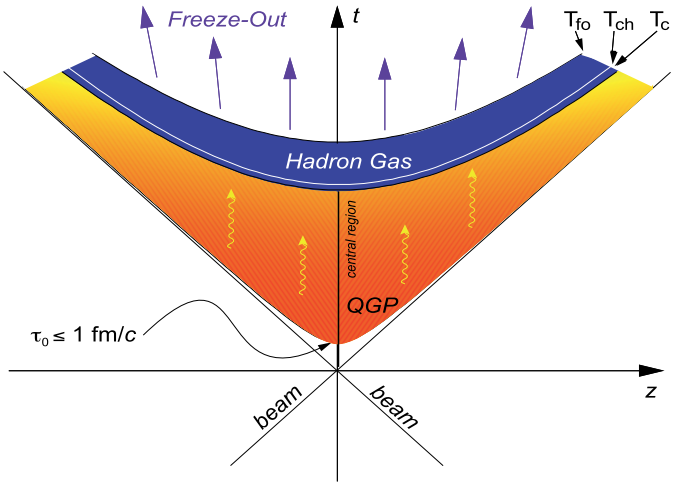



Figure 1.3: Space-time diagram representing the evolution of the system formed in ultrarelativistic heavy-ion collisions. Figure from Ref. [17]

and the colliding nucleons lose a fraction of their kinetic energy corresponding on average to about two units of rapidity[1] [13]. At LHC energies, the beam rapidity is $y \simeq 8$, giving rise to a net baryon-density distribution characterised by a double bump structure with bumps of hot baryonic matter at $y \approx \pm 6$ units, and a region around midrapidity[2] where the net baryon density is very small. Consequently, the QCD phase-diagram region with $\mu_B$ close to zero is explored. This condition is called *Bjorken* or *transparency regime* [14]. At lower centre-of-mass energies, the *stopping regime* is realised, in which the colliding ions do not move away from the collision region and the net baryon density is high. The low collision-energy regime was investigated at the Super Proton Synchrotron (SPS) and at the Alternating Gradient Synchrotron (AGS) in the second half of the 1980s, and subsequently at many other facilities [15, 16]. In these facilities various fixed-target and collider measurements were performed spanning centre-of-mass energies from 7.7 to 17.3 GeV, and exploring the strongly-interacting matter for baryon-chemical potentials of about 200–500 MeV.

### 1.3.1 Space-time evolution of the collision

The system produced in ultrarelativistic heavy-ion collisions is characterised by a complex space-time evolution, which is shown in Fig. 1.3. If the Bjorken condition is satisfied, the system evolution can be divided in the following stages [8, 14]:

[1]The rapidity of a particle with energy $E$ and momentum component along the beam direction $p_z$ is defined as $y = 1/2 \times \ln[(E + p_z)/(E - p_z)]$.

[2]The term midrapidity indicates the region around $y = 0$.

- The collision, in which the two nuclei cross each other. It lasts for a crossing time $\tau_{cross} = 2R/\gamma$, being $R$ the nucleus radius and $\gamma$ the Lorentz factor. In Pb–Pb collisions at $\sqrt{s} = 5.02$ TeV at the LHC, $\tau_{cross}$ is about $6 \times 10^{-3}$ fm/$c$, which is much smaller than the strong-interaction timescale $\tau_{QCD} \simeq 1$ fm/$c$.

- The pre-equilibrium stage, where particles are produced from the energy released in the interaction region by the colliding nuclei, after a formation time $\tau_0 > \tau_{cross}$. In the midrapidity region, the net baryon density is close to zero and the system energy density can be estimated as

$$\epsilon = \frac{\langle E_{\mathrm{T}} \rangle}{\pi R^2 \tau_0} \left. \frac{\mathrm{d}N}{\mathrm{d}y} \right|_{y=0}, \tag{1.5}$$

  where $\pi R^2$ gives the collision-region transverse area, $\langle E_{\mathrm{T}} \rangle$ is the average transverse energy[3] of the measured particles, and $\mathrm{d}N/\mathrm{d}y$ is the particle multiplicity per rapidity unit [14, 17]. Considering the particle-multiplicity values measured at the LHC and assuming $\tau_0 \sim 1$ fm/$c$, the energy densities reached at this collider are larger than 12 GeV/fm$^3$ [18], well above the pseudo-critical value for the QGP formation mentioned in Section 1.2.

- The QGP formation and evolution. After the pre-equilibrium stage, the QGP forms, and it evolves according to relativistic viscous hydrodynamics laws [19]. The QGP droplet expands driven by large pressure gradients and consequently cools down. The QGP lifetime for the energy densities reached at the LHC is of the order of 10 fm/$c$ [20].

- The QGP hadronisation, which occurs when the system temperature decreases below the pseudo-critical value for the crossover to hadronic matter. In this condition, the QGP constituents hadronise and produce a gas of interacting hadrons which continues to expand.

- The system freeze-out. As the hadron gas expands, its temperature reduces. When inelastic scatterings are no longer possible, the abundances of the hadron species are fixed, and the *chemical freeze-out* is reached. At the LHC, the chemical freeze-out occurs at $T_{ch} \simeq 153$ MeV, as shown in Section 1.4.2, shortly after the hadronisation. Finally, when also elastic scatterings cease, the system undergoes the *thermal freeze-out*, and the particle momentum spectra are defined.

[3] The transverse energy $E_{\mathrm{T}}$ of a particle is defined as $E_{\mathrm{T}} = \sqrt{m^2 + p_{\mathrm{T}}^2}$, where $m$ is the particle mass and $p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}$ is the particle momentum in the plane transverse to the direction of the beam.

### 1.3.2 The Glauber model

Due to the extended size of the nuclei involved in heavy-ion collisions, a modelling of the collision geometry is required to interpret the experimental measurements. Such a description can be obtained from the Glauber model [21]. This model is able to compute the nucleus–nucleus interaction cross section, together with average estimates of the number of *binary nucleon–nucleon interactions* $N_{\text{coll}}$ and *participant nucleons* $N_{\text{part}}$ involved in the nucleus–nucleus collision. The Glauber model is based on a semiclassical description of the nucleus–nucleus collision as an incoherent superposition of constituent-nucleon interactions. The model works in the so-called *optical limit*, and the following assumptions are adopted: nucleons are point-like and move independently inside the nucleus, the momentum of the nucleons is large and thus their trajectories are not deflected in the collision, protons and neutrons are indistinguishable, and the nucleon–nucleon interaction cross section does not vary during the collision. Under these assumptions, the collision between two nuclei can be described analytically starting from their nuclear density functions and the inelastic nucleon–nucleon cross section $\sigma_{\text{NN}}^{\text{inel}}$.

Figure 1.4 reports a sketch of the collision between a nucleus A, containing $A$ nucleons, and a nucleus B, formed by $B$ nucleons. The coordinate system origins in the centre of A, and the relative position of the two nuclei in the plane transverse to the direction of motion is defined by the *impact parameter* $\boldsymbol{b}$, which is the vector connecting their centres. The position in the transverse plane of a nucleon inside A is expressed by the vector $\boldsymbol{s}$, the same position with respect to the centre of B is given by $\boldsymbol{s} - \boldsymbol{b}$. The probability per unit area of finding a nucleon with coordinate $\boldsymbol{s}$ in the nucleus A is quantified by the *nuclear thickness function* $T_A(\boldsymbol{s})$, which is

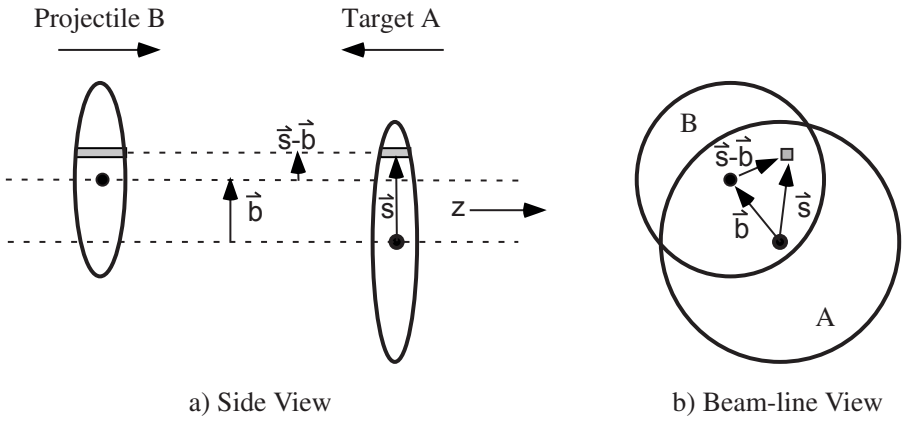



Figure 1.4: Schematic representation of a nucleus–nucleus collision in the Glauber-model approach. The transverse (a) and longitudinal (b) views are shown. Figure from Ref. [21]

defined as

$$T_A(\boldsymbol{s}) = \int \rho_A(\boldsymbol{s}, z_A)\mathrm{d}z_A, \tag{1.6}$$

where $\rho_A(\boldsymbol{s}, z_A)$ is the probability per unit volume of having the nucleon at the position $(\boldsymbol{s}, z_A)$, obtained dividing the nuclear density function by the number of nucleons $A$. A similar consideration holds for the nucleus B, and the probability per unit area of having two nucleons overlapping, one from each nucleus, is given by the *nuclear overlap function*

$$T_{AB}(\boldsymbol{b}) = \int T_A(\boldsymbol{s})T_B(\boldsymbol{s}-\boldsymbol{b})\mathrm{d}^2s. \tag{1.7}$$

Therefore, the probability of a nucleon–nucleon interaction to occur is $p(\boldsymbol{b}) = \sigma_{\mathrm{NN}}^{\mathrm{inel}} T_{AB}(\boldsymbol{b})$, and the probability of having $n$ nucleon–nucleon interactions in the nucleus–nucleus collision is obtained from a binomial distribution

$$P_{n,AB}(\boldsymbol{b}) = \binom{AB}{n} \left[\sigma_{\mathrm{NN}}^{\mathrm{inel}} T_{AB}(\boldsymbol{b})\right]^n \left[1 - \sigma_{\mathrm{NN}}^{\mathrm{inel}} T_{AB}(\boldsymbol{b})\right]^{AB-n}. \tag{1.8}$$

At this point, useful quantities about the collision can be computed. The total probability for a nucleus–nucleus collision $p_{AB}^{\mathrm{inel}}(\boldsymbol{b})$ can be calculated by requiring at least one nucleon–nucleon interaction as

$$\frac{\mathrm{d}^2\sigma_{AB}^{\mathrm{inel}}}{\mathrm{d}b^2} \equiv p_{AB}^{\mathrm{inel}}(\boldsymbol{b}) = 1 - P_{0,AB}(\boldsymbol{b}) = 1 - \left[1 - \sigma_{\mathrm{NN}}^{\mathrm{inel}} T_{AB}(\boldsymbol{b})\right]^{AB}. \tag{1.9}$$

It is possible then to obtain the total nucleus–nucleus inelastic cross section by integrating the above equation

$$\sigma_{AB}^{\mathrm{inel}} = \int_0^{+\infty} 2\pi b \left\{1 - \left[1 - \sigma_{\mathrm{NN}}^{\mathrm{inel}} T_{AB}(\boldsymbol{b})\right]^{AB}\right\} \mathrm{d}b. \tag{1.10}$$

The average number of binary collisions is provided by the mean value of the binomial probability distribution of Eq. (1.8) as

$$\langle N_{\mathrm{coll}} \rangle (\boldsymbol{b}) = AB \times \sigma_{\mathrm{NN}}^{\mathrm{inel}} T_{AB}(\boldsymbol{b}), \tag{1.11}$$

and the average number of participant nucleons can be computed, with a more complex derivation, as

$$\begin{aligned} \langle N_{\mathrm{part}} \rangle (\boldsymbol{b}) = A \times \int T_A(\boldsymbol{s}) \left\{1 - \left[1 - \sigma_{\mathrm{NN}}^{\mathrm{inel}} T_B(\boldsymbol{s}-\boldsymbol{b})\right]^B\right\} \mathrm{d}^2s \\ + B \times \int T_B(\boldsymbol{s}-\boldsymbol{b}) \left\{1 - \left[1 - \sigma_{\mathrm{NN}}^{\mathrm{inel}} T_A(\boldsymbol{s})\right]^A\right\} \mathrm{d}^2s. \end{aligned} \tag{1.12}$$

If hard-scattering processes between nucleons are considered, for which the production cross section is small, the Eq. (1.9) can be rewritten using a Taylor expansion at the first order as

$$\frac{\mathrm{d}^2\sigma_{AB}^{\mathrm{hard}}}{\mathrm{d}b^2} \simeq AB \times \sigma_{\mathrm{NN}}^{\mathrm{hard}} T_{AB}(\boldsymbol{b}) \propto \sigma_{\mathrm{NN}}^{\mathrm{hard}} \langle N_{\mathrm{coll}} \rangle (\boldsymbol{b}), \tag{1.13}$$

where $\sigma_{\mathrm{NN}}^{\mathrm{inel}}$ is replaced by the hard-scattering cross section $\sigma_{\mathrm{NN}}^{\mathrm{hard}}$, and the Eq. (1.12) is used in the last passage. Consequently, the Glauber model predicts that the yield of particles produced in nucleus–nucleus collisions through hard-scattering processes is proportional to the number of binary interactions $\langle N_{\mathrm{coll}} \rangle$. This property is referred to as *binary scaling*, and it is typical also of open heavy-flavour hadron production which will be discussed in Section 1.5.

The Glauber model successfully computes many quantities of interest related to the heavy-ion collision geometry. However, since it is an analytical model based on continuos nuclear density functions, it is only able to provide average quantities and does not account for event-by-event fluctuations, which arise from the discrete distribution of nucleons inside the nucleus. More realistic results are obtained from Glauber Monte Carlo simulations. In these simulations, the nucleon positions are randomly sampled from the nuclear density function, and a random impact parameter is generated according to the distribution $\mathrm{d}\sigma/\mathrm{d}b \propto 2\pi b$. The nucleus–nucleus collision is then simulated as an ensemble of independent nucleon–nucleon interactions, which occur if the following condition is satisfied

$$d \leq \sqrt{\frac{\sigma_{\mathrm{NN}}^{\mathrm{inel}}}{\pi}}, \tag{1.14}$$

where $d$ is the distance between the considered nucleon pair in the transverse plane. In this approach, the $N_{\mathrm{part}}$ and $N_{\mathrm{coll}}$ values are obtained for each event, and average quantities are computed simulating many nucleus–nucleus collisions. As will be discussed in Section 2.4, Glauber Monte Carlo simulations are adopted to determine the heavy-ion collision centrality starting from experimental measurements, and for central (0–10%) Pb–Pb collision at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV at the LHC, they estimate $\langle N_{\mathrm{part}} \rangle \simeq 360$ and $\langle N_{\mathrm{coll}} \rangle \simeq 1600$ [22].

## 1.4 Characterisation of the QGP

This Section presents a selection of the main observables and experimental results that lead to the current understanding of the properties of the short-lived QGP medium produced in ultrarelativistic heavy-ion collisions.

### 1.4.1 Charged-particle multiplicity

The number of charged particles, also called *multiplicity*, generated in heavy-ion collisions provides relevant information about the collision geometry and the properties of the created medium. In particular, the measurement of the particle multiplicity at midrapidity can be used to estimate the energy density characterising the pre-equilibrium phase, as described in Section 1.3.1.

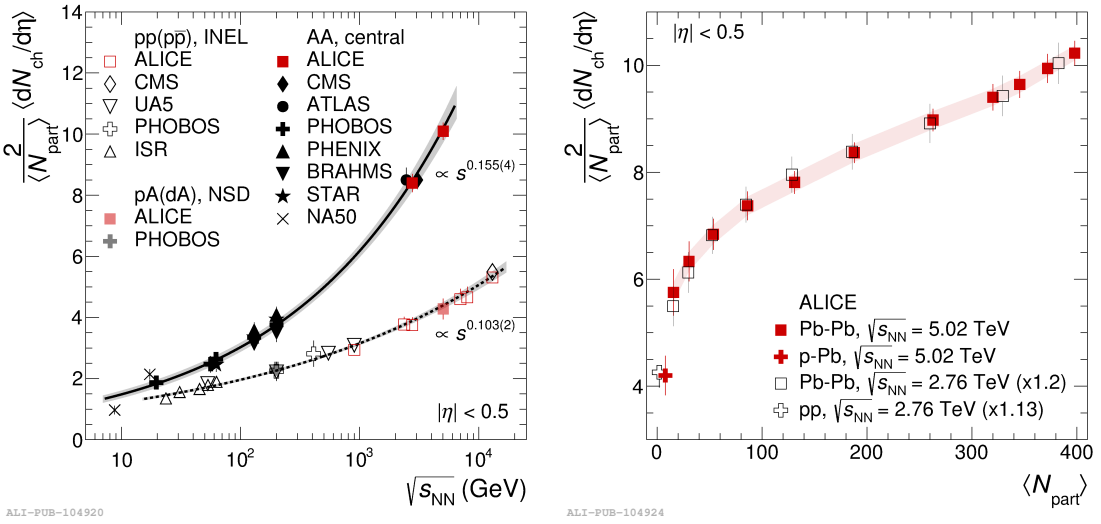



Figure 1.5: Charged-particle pseudorapidity density per participant-nucleon pair $\frac{2}{\langle N_{\rm part}\rangle}\langle {\rm d}N_{\rm ch}/{\rm d}\eta\rangle$ for pp and central A–A collisions as a function of $\sqrt{s_{\rm NN}}$ (left panel) and $\langle N_{\rm part}\rangle$ at LHC energies (right panel). The values measured at midrapidity in different collision systems and at different energies are reported. Figures from Ref. [23].

The left panel of Fig. 1.5 reports the measured average charged-particle pseudorapidity[4] density $\langle {\rm d}N_{\rm ch}/{\rm d}\eta\rangle$ at midrapidity divided by the number of participant-nucleon pairs $\langle N_{\rm part}\rangle/2$, as a function of the centre-of-mass energy per nucleon pair $\sqrt{s_{\rm NN}}$ [23]. The total number of particles generated in a collision is dominated by low-$p_{\rm T}$ light-flavour hadrons, i.e., pions, kaons, and protons, whose production is expected to scale with $N_{\rm part}$. Therefore, the $\langle N_{\rm part}\rangle/2$ factor is introduced to compare measurements in different colliding systems and at different energies, and it is computed from Glauber Monte Carlo simulations. The increase of the charged-particle density per participant-nucleon pair $\frac{2}{\langle N_{\rm part}\rangle}\langle {\rm d}N_{\rm ch}/{\rm d}\eta\rangle$ with the energy is steeper in central nucleus–nucleus (A–A) compared to proton–proton (pp) collisions. The measurements in p–A and d–A collisions lie on the same curve of the pp results, suggesting that the behaviour observed in A–A collisions is not only due to multiple interactions between the participant nucleons, which are also present in p–A and d–A collisions.

In the right panel of Fig. 1.5, the $\frac{2}{\langle N_{\rm part}\rangle}\langle {\rm d}N_{\rm ch}/{\rm d}\eta\rangle$ values measured by the ALICE Collaboration in pp, p–Pb, and Pb–Pb collisions are shown as a function

[4]The particle pseudorapidity is defined as $\eta = -\ln[\tan(\theta/2)]$, where $\theta$ is the particle emission angle relative to the beam direction, i.e., the polar angle. The pseudorapidity is used when the measured-particle mass is unknown, and it is equivalent to the rapidity $y$ in the limit $p \gg m$.

of the average number of participant nucleons $\langle N_{\rm part} \rangle$ [23]. The charged-particle density per participant-nucleon pair increases steeply from pp and p–Pb collisions to peripheral Pb–Pb collisions. The slope is reduced in Pb–Pb collisions, where the measured quantity increases by a factor of about 1.8 from the most peripheral, small $\langle N_{\rm part} \rangle$, to the most central collisions, large $\langle N_{\rm part} \rangle$. The observed trend does not depend on the collision energy and is described by models including initial-stage gluon saturation that limits the particle production in small systems [23].

The energy density of the system created in a heavy-ion collision can be estimated from the measured charged-particle multiplicity by using Eq. (1.5). The energy density obtained in the most central Pb–Pb collisions at the LHC is larger than 12 GeV/fm$^3$ [18], well above the pseudo-critical value for the QGP formation.

### 1.4.2 Identified-hadron abundances

By measuring the production yields of different hadron species, it is possible to estimate the temperature $T_{ch}$ and the volume $V$ of the system at the moment of the chemical freeze-out. These quantities can be obtained by comparing the measured yields of different hadron species with the predictions of the statistical hadronisation model (SHM) [24]. The SHM model assumes that the system composed of strongly-interacting hadrons is in thermal and chemical equilibrium at the chemical freeze-out, forming an ensemble which for A–A collisions at collider energies can be treated using the gran-canonical ensemble formalism. In these conditions, the hadron production is determined by the system thermodynamical properties. In particular, the abundance $N_i$ of a particle species $i$ is given by

$$N_i(T_{ch}, V, \mu_i) = \frac{g_i V}{2\pi^2} \int_0^{+\infty} \frac{p^2 \mathrm{d}p}{\exp\left[(E - \mu_i)/T_{ch}\right] \pm 1}, \tag{1.15}$$

where $V$ is the system volume, $p$ and $E$ are respectively the particle momentum and energy, $g_i$ quantifies the particle spin degeneracy, and $\mu_i$ is the chemical potential for the considered species which guarantees the conservation on average of the pertinent quantum numbers. The difference in the description of bosons and fermions is accounted for by the $\pm$ sign. An additional parameter, $\gamma_{\rm s}$, is in some cases introduced to account for the non-perfect chemical equilibrium of strange quarks. This quantity is used to scale the thermal predictions and is smaller than unity in case of non-equilibrium, as in peripheral A–A collisions or in the smaller p–A and pp collision systems. A further parameter, $\gamma_{\rm q}$, is introduced in the SHARE model, where a non-equilibrium QGP undergoing a sudden hadronisation without additional hadronic re-scatterings is considered [25].

In Fig. 1.6, the total yields per rapidity unit of different particle species measured by the ALICE Collaboration in the 10% most central Pb–Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV are reported. They are compared to the values obtained by the THERMUS 4 [26], GSI-Heidelberg [27], and SHARE 3 [25] models, which perform

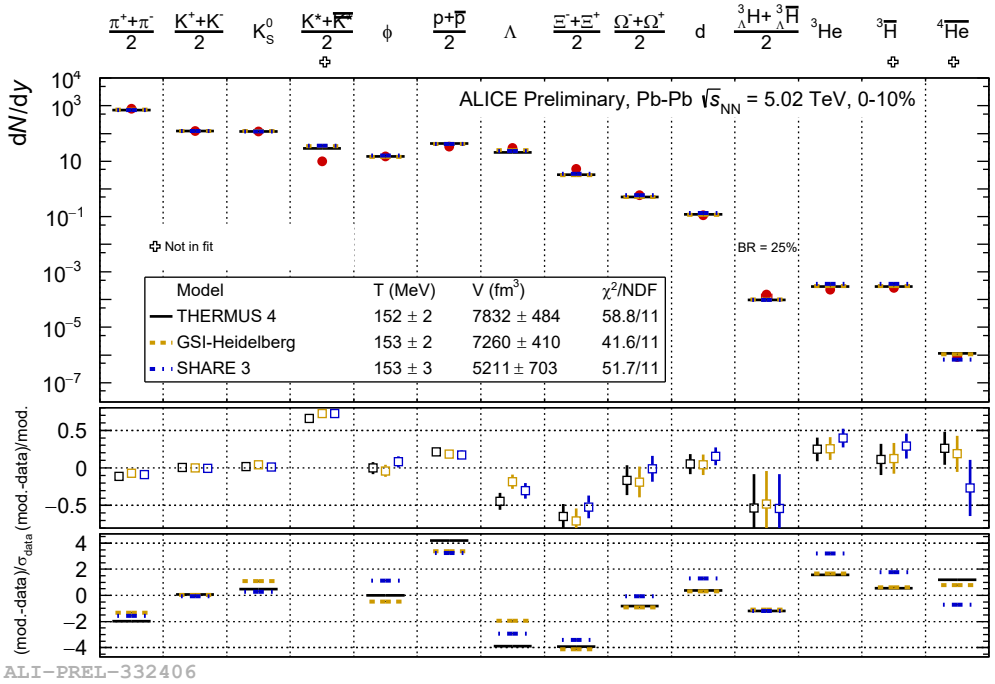



Figure 1.6: Simultaneous thermal fit to the identified-particle yields measured by the ALICE Collaboration in the 10% most central Pb–Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV. The results obtained with the THERMUS 4 [26], GSI-Heidelberg [27], and SHARE 3 [25] models are reported. Figure from the ALICE figure repository ©.

a simultaneous fit to the particle yields to determine the $T_{ch}$ and $V$ parameters. The models are based on different assumptions for the system-equilibrium condition and the conservation laws at the chemical freeze-out. They describe the observed particle abundances over nine orders of magnitude with a $\chi^2/ndf$ varying between about 3.8 and 5.3. A significant tension with the measurement is present for the proton yield, while excited kaons are excluded from the fit since they are strongly affected by rescatterings and regeneration in the hadronic phase. The estimated chemical freeze-out temperature is about 153 MeV, and it is compatible among the different models. This value is close to the expected pseudo-critical temperature for the crossover transition from the hadronic matter to the QGP.

### 1.4.3 Strangeness enhancement

A larger production of hadrons with strange-quark content in nucleus–nucleus collisions with respect to pp collisions is predicted to be a signature of the hadronic-matter deconfinement, and it is denominated *strangeness enhancement* [28]. In fact, the absence of valence strange quarks in the colliding nuclei implies that they must be produced in the collision or during the QGP phase. Below the critical

temperature for the phase transition, the production of $s\bar{s}$ pairs is suppressed, since the strange-quark effective mass implies a production threshold of about 1 GeV. In the QGP, the effective mass of strange quarks reduces by a factor up to 10 due to the chiral-symmetry restoration [29]. Therefore, $s\bar{s}$ pairs are expected to be abundantly produced via gluon-fusion processes in the QGP, where the gluon density is large, resulting in an increase of the strangeness production in heavy-ion collisions.

In an alternative approach based on the statistical hadronisation model, the strangeness enhancement can be explained as a suppression of the strangeness production in small collision systems, referred to as *canonical suppression* [30]. The gran-canonical description, where quantum numbers are conserved on average over relatively large volumes, cannot be applied in pp and p–A collisions due to the small number of produced particles. Instead, the canonical formulation must be adopted, which requires an exact quantum-number conservation. Therefore, the phase space available for particle production is reduced, and the strange-hadron production is suppressed. In this context, the strangeness enhancement in heavy-ion collisions is interpreted as the lifting of the canonical suppression typical of small systems.

An enhanced production of strange hadrons was observed in heavy-ion collisions at SPS [31], RHIC [32], and LHC [33] energies. Recently, it was observed for the first time also in high-multiplicity pp collisions by the ALICE Collaboration [34]. In Fig. 1.7, the $p_{\mathrm{T}}$-integrated yields of strange and multi-strange hadrons divided by that of charged pions as a function of the charged-particle multiplicity at midrapidity are reported for different collision systems and centre-of-mass energies per nucleon pair. The measured strange-hadron production is larger in Pb–Pb collisions with respect to low-multiplicity pp collisions, and it increases smoothly with increasing charged-particle multiplicity. The observed enhancement is more pronounced for hadrons with a higher strange-quark content. The production yields are compatible in collisions characterised by similar final state multiplicities, showing no significant dependence on the centre-of-mass energy or the collision system. This points to an origin of the strangeness production in hadronic collisions driven by the properties of the final state.

### 1.4.4 Radial flow

The properties of the system formed in heavy-ion collisions at the moment of the kinetic freeze-out can be studied by measuring the momentum spectra of produced hadrons. Assuming the system in thermal equilibrium, the particle transverse-momentum spectrum at low $p_{\mathrm{T}}$ can be described at first order by a Boltzmann-Gibbs law as

$$\frac{1}{p_{\mathrm{T}}}\frac{\mathrm{d}^2N}{\mathrm{d}p_{\mathrm{T}}\mathrm{d}y} \propto e^{-m_{\mathrm{T}}/T_{kin}}, \tag{1.16}$$

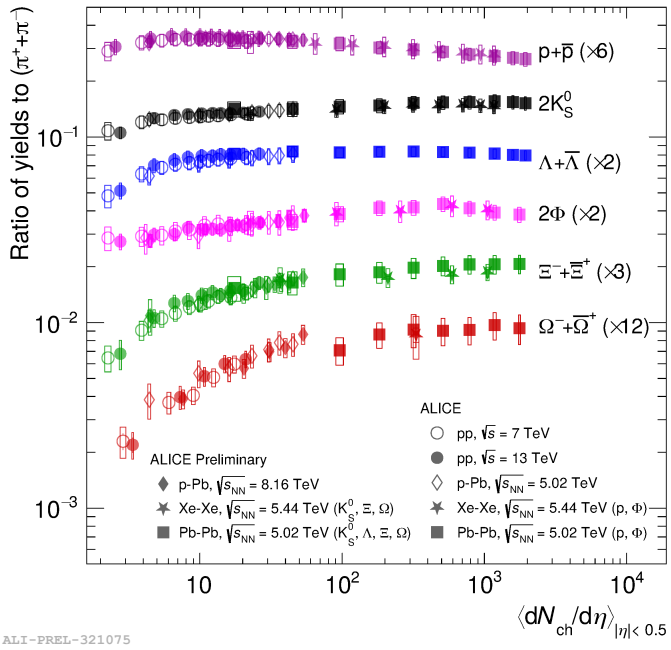



Figure 1.7: Ratios of the $p_{\mathrm{T}}$-integrated yield of hadrons with strange-quark content and charged pions as a function of the charged-particle multiplicity at midrapidity, measured by the ALICE Collaboration in various collision systems and at different energies. Figure from the ALICE figure repository ©.

where $m_{\mathrm{T}}$ is the particle transverse mass[5], and $T_{kin}$ is the temperature at the kinetic freeze-out. According to the above equation, the spectra of all produced hadron species follow the same exponential trend, where the slope depends on the species mass. This behaviour is usually denominated $m_{\mathrm{T}}$-*scaling* [35].

Deviations from the $m_{\mathrm{T}}$-scaling are observed for the hadron $p_{\mathrm{T}}$ spectra measured in Pb–Pb collisions at SPS, RHIC, and LHC energies [36, 37]. They are caused by the QGP internal pressure gradients, which generate a collective expansion of the system in the transverse plane. This collective motion is called *radial flow* and superimposes to the particle random thermal motion. The radial flow is considered in the Boltzmann-Gibbs Blast-Wave model [38], where the spectral shapes of the different hadron species are obtained under the assumption of particle emission from a thermal source at rest in the fluid reference frame, which is boosted by the collective fluid-expansion velocity $\beta_{\mathrm{T}}$. The velocity $\beta_{\mathrm{T}}$ is not constant along the radial direction due to the system expansion, and it is parametrised as a function

[5]The particle transverse mass is defined as $m_{\mathrm{T}} = \sqrt{m^2 + p_{\mathrm{T}}^2}$, where $m$ and $p_{\mathrm{T}}$ are the particle mass and transverse momentum, respectively. It is equivalent to the transverse energy $E_{\mathrm{T}}$.

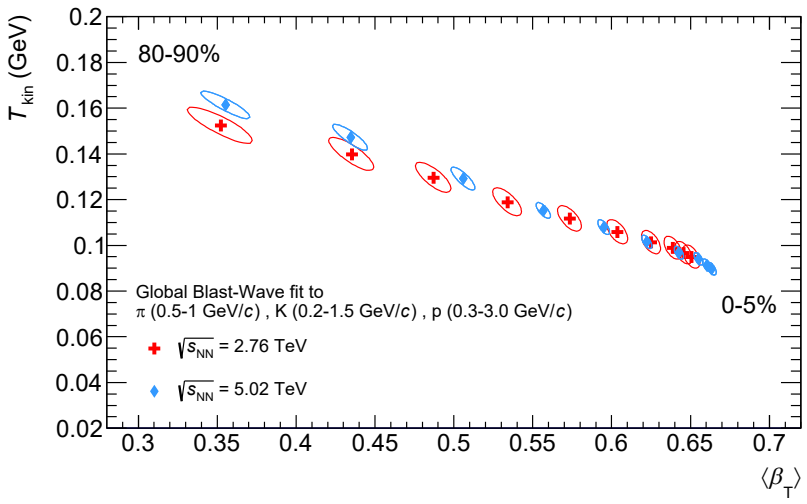



Figure 1.8: Progression of the average radial expansion velocity $\langle\beta_{\mathrm{T}}\rangle$ and the kinetic freeze-out temperature $T_{kin}$ resulting from the simultaneous Boltzmann-Gibbs Blast-Wave fit to $\pi^{\pm}$, $\mathrm{K}^{\pm}$, $\mathrm{p}(\overline{\mathrm{p}})$ spectra measured in Pb–Pb collisions. The ellipses report the $1\sigma$ correlated uncertainties from the fit. Figure from Ref. [37].

of the radial distance $r$ from the system centre as

$$\beta_{\mathrm{T}}(r) = \beta_s \left(\frac{r}{R}\right)^n, \tag{1.17}$$

being $\beta_s$ the velocity of the expanding-system surface, $R$ the system maximum radius, and $n$ a dimensionless parameter. In the Blast-Wave model, the produced particle spectrum results from the superposition of infinitesimal sources in thermal equilibrium and with different radial expansion velocities. This model provides an estimate of the kinetic freeze-out temperature and the average radial expansion velocity through fits to the $p_{\mathrm{T}}$-differential distributions of different hadron species, where $T_{kin}$, $\langle\beta_{\mathrm{T}}\rangle$, and $n$ are free parameters.

The correlation between the $T_{kin}$ and $\langle\beta_{\mathrm{T}}\rangle$ parameters measured by the ALICE Collaboration in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV [36] and 5.02 TeV [37] is displayed in Fig. 1.8. The parameter values are obtained from a simultaneous Blast-Wave fit to the $p_{\mathrm{T}}$ spectra of pions, kaons, and protons at low transverse momentum, and are measured for different collision centralities. No significant discrepancies are observed between the two different collision energies. The average radial expansion velocity $\langle\beta_{\mathrm{T}}\rangle$ increases from more peripheral to more central collisions, and reaches a maximum of about 0.66 in the 5% most central Pb–Pb collisions. This indicates that the system expansion becomes more rapid with increasing collision centrality. The kinetic freeze-out temperature $T_{kin}$ instead decreases from about 161 MeV in the most peripheral collisions to about 90 MeV in the most central ones, possibly pointing to a delayed kinetic decoupling of the system in central collisions. It is

worth mentioning that the $T_{kin}$ and $\langle\beta_{\mathrm{T}}\rangle$ values obtained from the simultaneous fit gain a physical meaning only if they are reproduced by theoretical models describing the dynamical expansion of the system. These models are based on the relativistic hydrodynamics formalism, which connects the collective expansion of the medium with its equation of state.

### 1.4.5 Anisotropic flow

The geometry of a nucleus–nucleus collision is characterised by an anisotropy of the overlap region in the transverse plane, which is caused by the finite impact parameter and increases with increasing impact-parameter values, as shown in Fig. 1.4. The spatial anisotropy of the initial state produces anisotropic pressure gradients in the QGP, which affect the collective flow of the medium. This ultimately results in an azimuthal anisotropy in the momentum distribution of the produced hadrons. The anisotropy can be studied by measuring the particle production as a function of the azimuthal angle $\varphi$, which can be decomposed in a Fourier expansion [39] as

$$E\frac{\mathrm{d}^3N}{\mathrm{d}p^3} = \frac{1}{2\pi}\frac{\mathrm{d}^2N}{p_{\mathrm{T}}\mathrm{d}p_{\mathrm{T}}\mathrm{d}y}\left\{1 + 2\sum_{n=1}^{\infty} v_n(p_{\mathrm{T}})\cos[n(\varphi - \Psi_n)]\right\}. \tag{1.18}$$

In the above equation, $n$ is the order of the harmonic, $v_n(p_{\mathrm{T}})$ is the $p_{\mathrm{T}}$-dependent Fourier coefficient, also referred to as *flow coefficient*, and $\Psi_n$ is the symmetry-plane angle. The coefficient $v_n(p_{\mathrm{T}})$ is defined as

$$v_n(p_{\mathrm{T}}) = \langle\cos[n(\varphi - \Psi_n)]\rangle\,, \tag{1.19}$$

where the brackets indicate the average over all particles in a given $p_{\mathrm{T}}$ interval, and over all events within a given centrality interval. The symmetry-plane angle $\Psi_n$ is defined by the spatial distribution of the nucleons participating in the collision. In the case of $n = 2$ and neglecting fluctuations of the participant distribution, it identifies the reaction plane defined by the impact-parameter vector and the direction of the colliding nuclei.

The second-harmonic coefficient $v_2$, called *elliptic flow*, is the dominant term of the Fourier expansion in non-central collisions, due to the almond-like shape of the interaction region. The measurement of a large elliptic flow is a direct signature of a collective motion. Higher order coefficients, $v_3$ and $v_4$, are mainly generated by fluctuations in the initial distribution of participants in the overlap region, and by the non-linear hydrodynamic response of the medium [40, 41]. The comparison of the measured $v_n$ coefficients with viscous hydrodynamic models of the QGP provides information about the system equation of state and transport coefficients, such as the shear and bulk viscosity to entropy density ratios [42, 43].

Figure 1.9 reports the $p_{\mathrm{T}}$-differential elliptic-flow coefficient $v_2$ of charged pions, charged and neutral kaons, $\phi$ mesons, protons, and $\Lambda$ hyperons measured by the

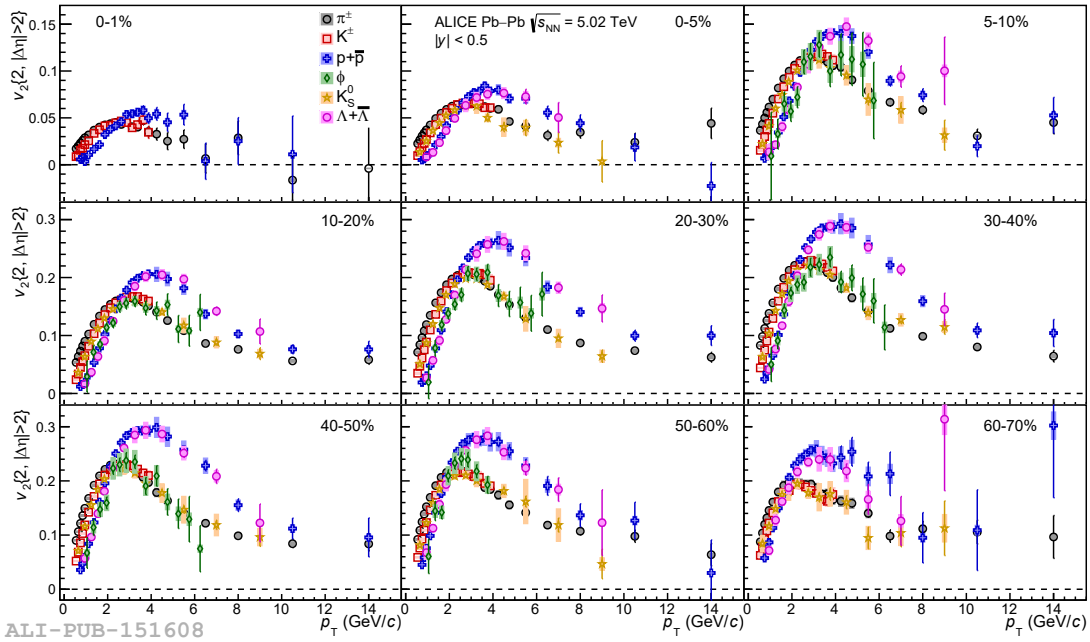



Figure 1.9: $p_{\mathrm{T}}$-differential elliptic-flow coefficient $v_2$ of charged pions, charged and neutral kaons, $\phi$ mesons, protons, and $\Lambda$ hyperons for several centrality intervals in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. Figure from Ref. [44].

ALICE Collaboration in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV in several centrality intervals [44]. The elliptic-flow coefficient increases from the most central (0–1%) to semicentral (30–40%) collisions, due to the increasing spatial anisotropy of the overlapping region between the nuclei. In more peripheral collisions, it decreases as a consequence of the smaller pressure gradients originating from the smaller density of the produced medium. The $v_2$ coefficient shows a mass ordering in the low-$p_{\mathrm{T}}$ region, $p_{\mathrm{T}} < 2$–$3$ GeV/$c$, where lighter particles have a larger $v_2$ than the heavier ones. This is understood as an effect of the radial flow, which imposes an equal velocity boost to all particles, thus pushing more massive hadrons to higher $p_{\mathrm{T}}$. In the transverse-momentum region $3 < p_{\mathrm{T}} < 8$ GeV/$c$, the $v_2$ coefficients show a grouping by baryons and mesons, with the values of baryons being larger than those of mesons. Notably, the $\phi$ meson, whose mass is similar to the proton one, has a $v_2$ value compatible with that of the other mesons in this $p_{\mathrm{T}}$ interval. This implies a dependence on the number of constituent quarks rather than the hadron mass. This behaviour is in agreement with the hypothesis of quark hadronisation via coalescence [45], where the elliptic-flow of each constituent quark contributes to the $v_2$ of the final-state hadron.

### 1.4.6 Jet quenching

Partons with high $p_{\mathrm{T}}$ are produced in hard-scattering processes, i.e., scatterings in which the momentum exchanged between the interacting quarks and gluons is large, occurring in the early stages of the heavy-ion collision. They subsequently propagate in the QGP medium and interact with the medium constituents via elastic scatterings and gluon radiation. As a consequence, the parton shower originating from the hard-scattered parton and the particles produced in its hadronisation will be modified by the interactions with the medium. This phenomenon is called *jet quenching*. One of the manifestations of jet quenching is the parton energy loss in the QGP, which results in a softening of the $p_{\mathrm{T}}$ distributions of the final-state hadrons.

According to the Glauber model introduced in Section 1.3.2, the yield of particles originating from hard processes in nucleus–nucleus collisions is expected to be proportional to the yield observed in pp collisions by the average number of binary nucleon–nucleon interactions $\langle N_{\mathrm{coll}} \rangle$. The validity of this binary scaling can be investigated by measuring the *nuclear modification factor* $R_{\mathrm{AA}}$, which is defined as

$$R_{\mathrm{AA}}(p_{\mathrm{T}}, y) = \frac{1}{\langle N_{\mathrm{coll}} \rangle} \times \frac{\mathrm{d}^2 N_{\mathrm{AA}}/\mathrm{d}p_{\mathrm{T}}\mathrm{d}y}{\mathrm{d}^2 N_{\mathrm{pp}}/\mathrm{d}p_{\mathrm{T}}\mathrm{d}y}, \tag{1.20}$$

where $\mathrm{d}^2 N_{\mathrm{AA(pp)}}/\mathrm{d}p_{\mathrm{T}}\mathrm{d}y$ is the $p_{\mathrm{T}}$- and rapidity-differential yield of hadrons observed in A–A (pp) collisions. The $R_{\mathrm{AA}}$ is expected to be equal to unity if the binary-scaling assumption holds. Instead, a deviation from unity arises from a modification of the $p_{\mathrm{T}}$ distribution of produced hadrons, implying the possible presence of effects induced by the QGP medium. However, *cold nuclear matter* (CNM) effects, which are caused by having nuclei as colliding objects and are not related to the QGP formation, can also result in a $R_{\mathrm{AA}}$ different from unity. These effects are typically assessed by measuring the nuclear modification factor in p–A collisions, and will be discussed in Section 1.5.2. For light-flavour hadrons at low $p_{\mathrm{T}}$, whose production is dominated by soft processes and scales with the average number of participant nucleons $\langle N_{\mathrm{part}} \rangle$, a $R_{\mathrm{AA}}$ below unity is expected even in absence of medium effects.

In the left panel of Fig. 1.10, the $R_{\mathrm{AA}}$ of charged particles as a function of $p_{\mathrm{T}}$ measured by the ALICE [46] and CMS [48] Collaborations in p–Pb and in central (0–5%) Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV is reported. At intermediate $p_{\mathrm{T}}$, $3 < p_{\mathrm{T}} < 10$ GeV/$c$, the nuclear modification factor is near unity in p–Pb collisions while it is significantly suppressed in central Pb–Pb collisions, in agreement with the picture of partons losing energy in the deconfined medium. In addition, the $R_{\mathrm{AA}}$ of photons, W and Z bosons, which do not carry colour charge, is compatible with unity in the same $p_{\mathrm{T}}$ region in Pb–Pb collisions [49–51]. This indicates that the energy loss in the QGP originates from strong interactions of high-energy quarks and gluons with the medium constituents, not affecting electroweak probes. The charged-particle $R_{\mathrm{AA}}$ shows an increasing trend for $p_{\mathrm{T}} \gtrsim 10$ GeV/$c$. This behaviour

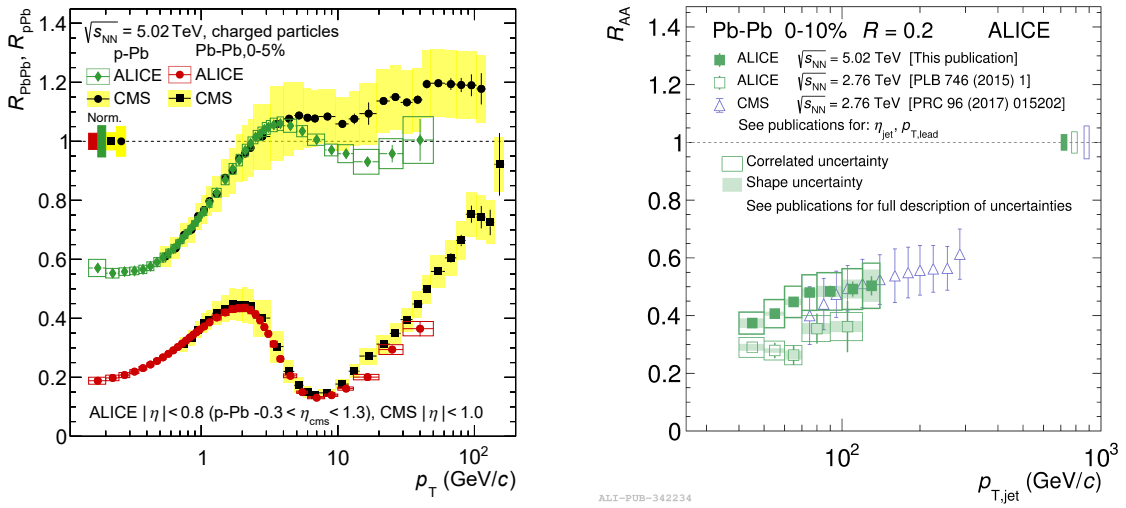



Figure 1.10: Left panel: charged-particle nuclear modification factor in p–Pb and in central (0–5%) Pb–Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV measured by the ALICE and CMS Collaborations. Figure from Ref. [46]. Right panel: jet $R_{\rm AA}$ in central (0–10%) Pb–Pb collisions at $\sqrt{s_{\rm NN}} = 2.76$ and 5.02 TeV. Figure from Ref. [47].

is less pronounced studying the production of jets, which are collimated sprays of hadrons produced by the fragmentation of high-energy partons. The jet $R_{\rm AA}$ is approximatively flat in the common $p_{\rm T}$ region and has a plateau at a value of about 0.5 for $p_{\rm T}$ above 100 GeV/$c$, as displayed in the right panel of Fig. 1.10. The figure reports the nuclear modification factor of jets in central (0–10%) Pb–Pb collisions measured at $\sqrt{s_{\rm NN}} = 2.76$ by the ALICE [52] and CMS [53] Collaborations, and at $\sqrt{s_{\rm NN}} = 5.02$ TeV by ALICE [47]. The jet $R_{\rm AA}$ lower than unity indicates that in-medium interactions transport a portion of the parton energy at large angles from the hard-scattered parton direction, i.e., out of the jet cone, where it does not contribute to the jet momentum. A comparison of the jet and hadron $R_{\rm AA}$ is not straightforward, since hadrons and jets at any given $p_{\rm T}$ originate from partons with different energies, and the selection of high-$p_{\rm T}$ hadrons imposes a bias on the jet fragmentation.

### 1.4.7 Quarkonium suppression

The study of $\rm c\bar{c}$ and $\rm b\bar{b}$ bound states, called *quarkonium* states, is an excellent tool to investigate the QGP properties. In particular, the suppression of quarkonium production in ultrarelativistic heavy-ion collisions is a probe of the strongly-interacting matter deconfinement, and it is sensitive to the QGP temperature [54, 55]. The binding energy of a heavy quark-antiquark pair in the vacuum

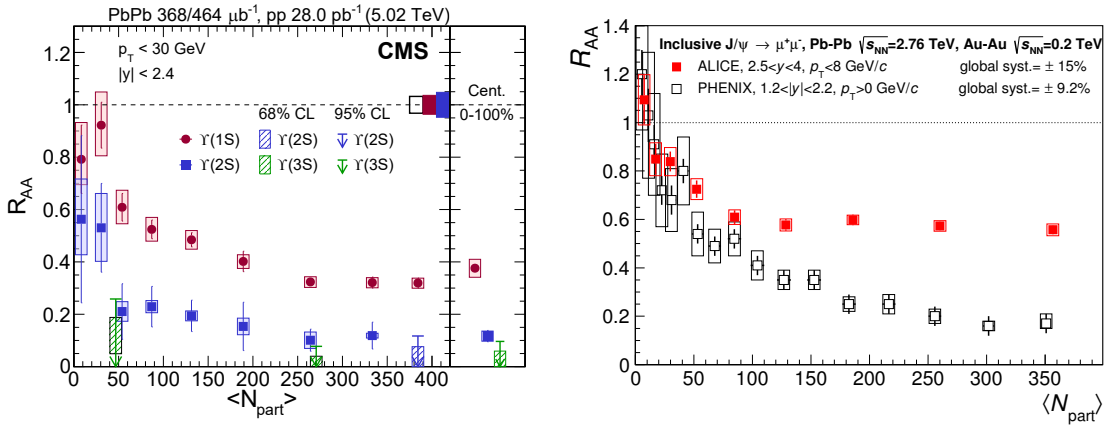



Figure 1.11: Left panel: $R_{\mathrm{AA}}$ of $\Upsilon(1\mathrm{S})$, $\Upsilon(2\mathrm{S})$, and $\Upsilon(3\mathrm{S})$ mesons as a function of the average number of participant nucleons $\langle N_{\mathrm{part}} \rangle$ in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. Figure from Ref. [56]. Right panel: J/ψ-meson $R_{\mathrm{AA}}$ as a function of $\langle N_{\mathrm{part}} \rangle$ in Au–Au collisions at $\sqrt{s_{\mathrm{NN}}} = 200$ GeV and in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV. Figure from Ref. [57].

can be derived from the Cornell potential

$$V(r) = kr - \frac{\alpha}{r}, \tag{1.21}$$

where $r$ is the distance between the two quarks, $k$ represents the string tension, and $\alpha$ is a Coulomb-like constant. The Coulomb-like term dominates at low $r$ values, while the linear term is relevant at high distances and is responsible for the quark-pair confinement. In presence of the deconfined medium, the linear term vanishes since colour charges are free in the QGP, and the potential becomes

$$V^{QGP}(r) = -\frac{\alpha}{r} e^{-r/\lambda_D}. \tag{1.22}$$

The exponential term in the above equation introduces a screening effect similar to the one experienced by electric charges inside a plasma of electrons and ions, which depends on the $\lambda_D$ parameter, i.e., the *Debye length*. This parameter quantifies the typical extension of the colour-screening effect in the deconfined medium, determining the maximum distance between two quarks for the bound-state existence. The $\lambda_D$ value decreases as the medium temperature increases, thus leading to a suppression of quarkonium states with characteristic radius larger than it.

The existence of many quarkonium states with different binding energies provides the possibility to study the phenomenon of the sequential suppression. This phenomenon consists in the progressive depletion of the different quarkonium resonances as the temperature of the system increases, resulting in a decrease of the measured production yields. The right panel of Fig. 1.11 reports the nuclear modification factor $R_{\mathrm{AA}}$ of different bottomonium ($\mathrm{b\bar{b}}$) states measured as a function

of the average number of participant nucleons $\langle N_{\mathrm{part}} \rangle$ by the CMS Collaboration in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV [56]. The observed $R_{\mathrm{AA}}$ decreases as the centrality of the collision increases, which corresponds to a hotter QGP medium. Moreover, the $R_{\mathrm{AA}}$ of the bottomonium ground state, $\Upsilon$(1S), is less suppressed than the one of the excited states, $\Upsilon$(2S), and $\Upsilon$(3S), which have smaller binding energies. These observations support the assumption of a medium-induced colour screening and the consequent sequential melting of quarkonium states.

The J/$\psi$-meson $R_{\mathrm{AA}}$ measured as a function of the collision centrality by the ALICE Collaboration in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV [57] and by the PHENIX Collaboration in Au–Au collisions at $\sqrt{s_{\mathrm{NN}}} = 200$ GeV [58] is shown in the right panel of Fig. 1.11. The nuclear modification factor of J/$\psi$ mesons measured at $\sqrt{s_{\mathrm{NN}}} = 200$ GeV is smaller than the one observed at a centre-of-mass energy per nucleon pair higher by more than one order of magnitude, contradicting the expectation of a larger charmonium ($\mathrm{c\overline{c}}$) suppression caused by the higher QGP temperature reached. This observation can be explained by the mechanism of charmonium regeneration, in which the statistical recombination of charm and anticharm quarks originating from uncorrelated hard scatterings partially compensates the suppression due to the QGP presence [59]. The regeneration is larger at the LHC since the number of $\mathrm{c\overline{c}}$ pairs per event produced is higher than the one at RHIC of about a factor ten, thus explaining the larger $R_{\mathrm{AA}}$ observed. This regeneration is in principle possible also for bottomonium states. However, at LHC energies, the small number of $\mathrm{b\overline{b}}$ pairs per event produced prevents the observation of this effect.

## 1.5 Open heavy-flavour hadrons

Charm and beauty quarks, also called *heavy flavours*, are produced in hard-scattering processes between the partons of the interacting hadrons. In nucleus–nucleus collisions, they are generated on a shorter timescale than the QGP formation time, and they experience all the phases of the system evolution interacting with the medium constituents via inelastic and elastic scatterings. Therefore, open heavy-flavour hadrons, such as D and B mesons, are powerful tools to investigate the properties of the QGP created in heavy-ion collisions [2]. In addition, measurements of the production of open heavy-flavour hadrons in pp collisions are useful benchmarks of perturbative-QCD calculations [60] and provide the reference for the studies in nucleus–nucleus collisions. In this Section, the production, interaction with the medium, and hadronisation of heavy flavours are discussed.

### 1.5.1 Production in hadronic collisions

Due to their large mass, $m_{\mathrm{c}} \simeq 1.3$ GeV/$c^2$ and $m_{\mathrm{b}} \simeq 4.2$ GeV/$c^2$ [6], charm and beauty quarks are always produced in processes characterised by large transferred momentum, i.e., $Q^2 \gtrsim 4m_{\mathrm{c,b}}^2$. As shown in Section 1.1, the strong coupling constant $\alpha_s$ is smaller than unity in this regime, and pQCD calculations can describe the heavy-flavour production down to $p_{\mathrm{T}} = 0$. The partonic cross section $\sigma_{ij\to\mathrm{Q\overline{Q}}}$ for the production of a heavy quark-antiquark pair is calculated with a perturbative expansion in $\alpha_s$. The main contributing processes at leading order, $\mathcal{O}(\alpha_s^2)$, are the quark-antiquark annihilation $\mathrm{q\overline{q}} \to \mathrm{Q\overline{Q}}$ and the gluon fusion $\mathrm{gg} \to \mathrm{Q\overline{Q}}$ [61]. Higher-order corrections introduce a dependence of the cross section on the energy scale chosen for the renormalisation of ultraviolet divergences, usually referred to as the *renormalisation scale* $\mu_{\mathrm{R}}$.

According to the factorisation theorem [62], the production cross section of open heavy-flavour hadrons in pp collisions can be computed starting from the partonic cross section as

$$\sigma_{\mathrm{pp}\to\mathrm{H_Q+X}} = \sum_{i,j=\mathrm{q},\overline{\mathrm{q}},\mathrm{g}} f_i(x_i, Q^2) f_j(x_j, Q^2) \otimes \sigma_{ij\to\mathrm{Q\overline{Q}}} \otimes D_{\mathrm{Q}\to\mathrm{H_Q}}(z, Q^2). \tag{1.23}$$

In the above equation, $f_i(x_i, Q^2)$ is the *parton distribution function* (PDF), which provides the probability of finding a parton $i$ in the interacting proton with a fraction $x_i$ of the proton momentum. The term $D_{\mathrm{Q}\to\mathrm{H_Q}}(z, Q^2)$ is the *fragmentation function* (FF) describing the probability of a heavy-quark Q to hadronise into a particular hadron species $\mathrm{H_Q}$, where $z$ is the quark-momentum fraction inherited by the hadron. Both the PDF and FF describe non-perturbative processes and thus cannot be treated in the pQCD framework. The PDFs are typically extracted from deep-inelastic scattering measurements [63–65], from which different parametrisations are derived depending on the parton nature. The FFs are obtained from measurements in $\mathrm{e^+e^-}$ collisions [66, 67], and are assumed to be independent of the collision system. The integrated values of the FFs are the *fragmentation fractions* $f(\mathrm{Q} \to \mathrm{H_Q})$, which give the total probability of heavy quarks to produce specific hadrons and have been measured both in $\mathrm{e^+e^-}$ [68] and hadronic [6, 69] collisions. The parton distribution and the fragmentation functions depend on the energy scale of the considered process, related to the *factorisation scale* $\mu_{\mathrm{F}}$. Therefore, they are determined at a given energy scale $Q_0^2$, and are then evolved to the $Q^2$ value of interest employing the DGLAP equations [70]. The renormalisation and factorisation scales are normally set to the same order of the hard-process momentum transfer, i.e, $\mu_{\mathrm{R}} \approx \mu_{\mathrm{F}} \approx \sqrt{m_{\mathrm{Q}}^2 + p_{\mathrm{T,Q}}^2}$.

State-of-the-art predictions of the production cross section of open heavy-flavour hadrons are provided by the FONLL [60, 72, 73] and GM-VFNS calculations [74–76]. Both schemes are based on the collinear factorisation approach, and perform next-to-leading-order (NLO) pQCD calculations with resummation of next-to-leading logarithms. However, they adopt different choices for the matching of

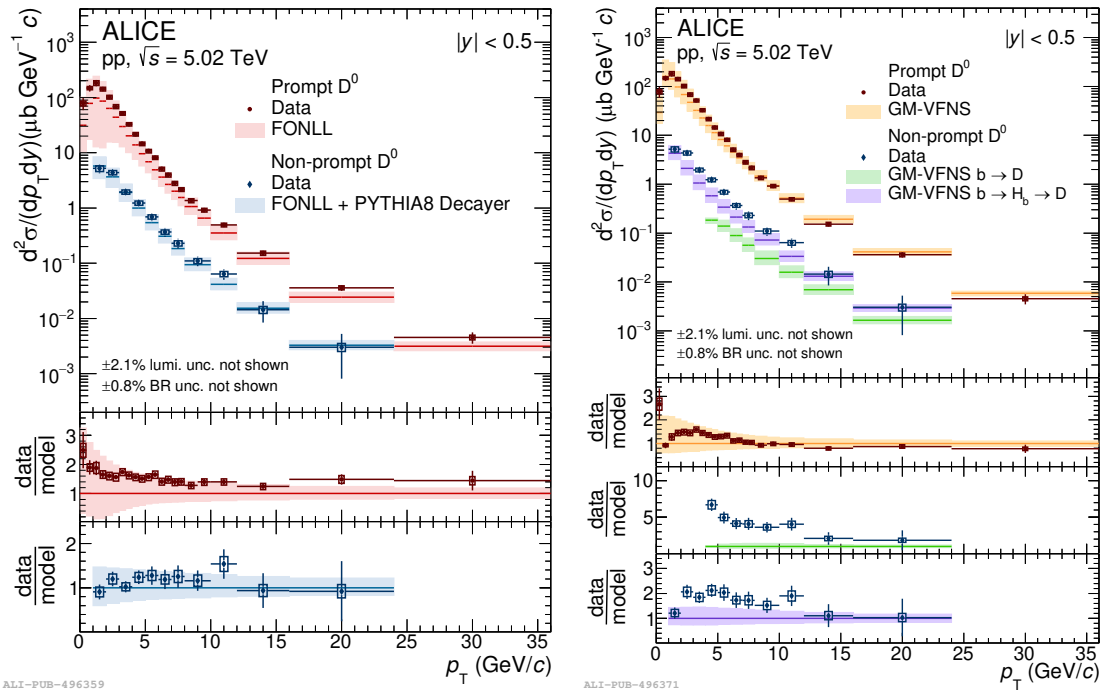



Figure 1.12: Comparison of the FONLL (left panel) and GM-VFNS (right panel) calculations with the measured $p_{\mathrm{T}}$-differential production cross section of prompt and non-prompt $\mathrm{D}^0$ mesons in pp collisions at $\sqrt{s} = 5.02$ TeV. Figure from Ref. [71].

the NLO and resummation calculations, the PDF and FF sets used, and the values of the charm- and beauty-quark masses. In Fig. 1.12, the FONLL and GM-VFNS calculations are compared to the $p_{\mathrm{T}}$-differential production cross section of prompt $\mathrm{D}^0$ mesons, i.e., coming directly from the charm-quark hadronisation or from decays of excited charm-hadron states, in pp collisions at $\sqrt{s} = 5.02$ TeV [71]. The comparison with non-prompt $\mathrm{D}^0$ mesons, which originate from beauty-hadron decays, is also reported. The pQCD calculations are in agreement with the prompt $\mathrm{D}^0$-meson measurement within uncertainties, indicating that the factorisation approach coupled with the fragmentation functions extracted from $\mathrm{e}^+\mathrm{e}^-$ collisions is successful in the meson sector. This is not the case in the baryon sector as will be discussed in Section 1.5.4. The non-prompt $\mathrm{D}^0$ production is well described by FONLL, while GM-VFNS calculations underestimate the measured values. These discrepancies are driven by the description of the beauty-quark fragmentation into a beauty-hadron and its subsequent decay to a $\mathrm{D}^0$ meson. In particular, a better agreement is observed if the two processes are described separately ($\mathrm{b} \rightarrow \mathrm{H_b} \rightarrow \mathrm{D}$), as done to obtain the FONLL predictions, with respect to considering them in a single step ($\mathrm{b} \rightarrow \mathrm{D}$) [77].

Calculations of the production of open heavy-flavour hadrons are also provided by the $k_{\mathrm{T}}$-factorisation model [78, 79], where an implementation of the factorisation

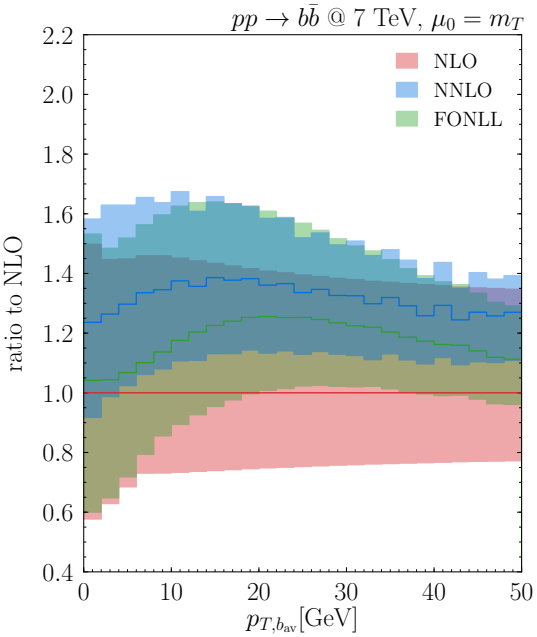



Figure 1.13: Comparison among the NLO, FONLL and NNLO predictions for the $p_{\mathrm{T}}$-differential beauty-quark production cross section in pp collisions at $\sqrt{s} = 7$ TeV. Figure from Ref. [80].

alternative to the collinear approach is considered. These calculations are performed at the leading order and include higher-order corrections. The $k_{\mathrm{T}}$-factorisation predictions describe the measurements at low and intermediate $p_{\mathrm{T}}$, but overshoot the data for transverse momenta above about 10 GeV/$c$.

Recently, predictions for the $\mathrm{c\overline{c}}$ and $\mathrm{b\overline{b}}$ production cross section employing next-to-next-to-leading-order (NNLO) pQCD calculations became available [80, 81], providing improved results. Figure 1.13 reports the $p_{\mathrm{T}}$-differential beauty-quark production cross sections in pp collisions at $\sqrt{s} = 7$ TeV from NLO, FONLL and NNLO calculations divided by the one resulting from computations at the NLO [80]. The central value and uncertainty band obtained with FONLL calculations match the NLO ones at low transverse momentum. This is expected, since the contribution of the logarithm resummation is relevant at high $p_{\mathrm{T}}$, where it leads to an increase of the FONLL central prediction and a reduction of the uncertainties with respect to NLO calculations. The cross section resulting from NNLO calculations is systematically higher than the FONLL one over all the considered $p_{\mathrm{T}}$ range and is contained in the FONLL uncertainty band. It is also affected by a smaller uncertainty, particularly in the low- and intermediate-$p_{\mathrm{T}}$ regions, which highlight the improved precision of NNLO calculations.

An alternative approach for the description of the heavy-flavour hadron production is provided by general purpose Monte Carlo generators, such as PYTHIA [82]

and HERWIG [83]. These generators are characterised by a more complete representation of the final state than pQCD calculations, since they include also soft and multi-parton interactions, initial- and final-state parton showers, and a description of the hadron decays. Although, only leading-order processes are implemented exactly, and higher-order effects are considered with coarse-grained approaches. Monte Carlo event generators with next-to-leading order accuracy, such as MC@NLO [84] and POWHEG [85], are also available. They provide predictions for differential observables and jets, and can be consistently matched to the parton showers generated by PYTHIA and HERWIG.

### 1.5.2 Cold nuclear matter effects

As mentioned in Section 1.4.6, cold nuclear matter effects influence the heavy-flavour hadron production in nucleus–nucleus collisions, and they must be considered to properly interpret the measurements. These effects are typically investigated in p–A and d–A collisions, and by studying the production of electroweak bosons in A–A collisions, where phenomena due to the QGP formation can be neglected.

A relevant CNM effect at LHC energies is the modification of the PDFs of the nucleons bound in the colliding nuclei with respect to free nucleons. Even considering the nucleus as a simple collection of nearly-free nucleons, a variation of the PDFs is expected due the quantum Fermi motion of the nucleons inside the nucleus. In fact, the distributions of the nucleon-momentum fraction, called *Bjorken* $x$, carried by the partons of nucleons bound in a nucleus are the convolution of the distributions for a free nucleon at rest and the nucleon Fermi motion. In addition, deep-inelastic scattering experiments observed various other PDF modification effects, which depend on the interval of $x$ considered. The PDF of a nucleon bound in a nucleus can be expressed as [86]

$$f_i^{\mathrm{A}}(x, Q^2) = R_i^{\mathrm{A}}(x, Q^2) f_i(x, Q^2), \tag{1.24}$$

where $R_i^{\mathrm{A}}(x, Q^2)$ quantifies the nuclear modification of the free-nucleon parton distribution function $f_i(x, Q^2)$, and $i$ indicates the parton species.

The trend of $R_i^{\mathrm{A}}$ as a function of $x$ is exemplified in Fig. 1.14. Deviations of this quantity from unity are present in all the $x$ range, while a value equal to unity is expected in absence of nuclear effects. For $x \lesssim 0.05$, a suppression of the PDFs is observed due to the *shadowing* effect [88]. This behaviour can be explained by the spatial overlap of the partons from different nucleons giving rise to a saturation of the density of gluons, which are the main nucleon constituents at low $x$ values. An *anti-shadowing* region can be identified in the $0.05 \lesssim x \lesssim 0.3$ interval, where the enhancement of the $R_i^{\mathrm{A}}$ above unity is due to momentum conservation, and is a consequence of the shadowing effect depleting the low Bjorken-$x$ region. Another suppression region is present for $0.3 \lesssim x \lesssim 0.8$, and is usually denoted as the *EMC*

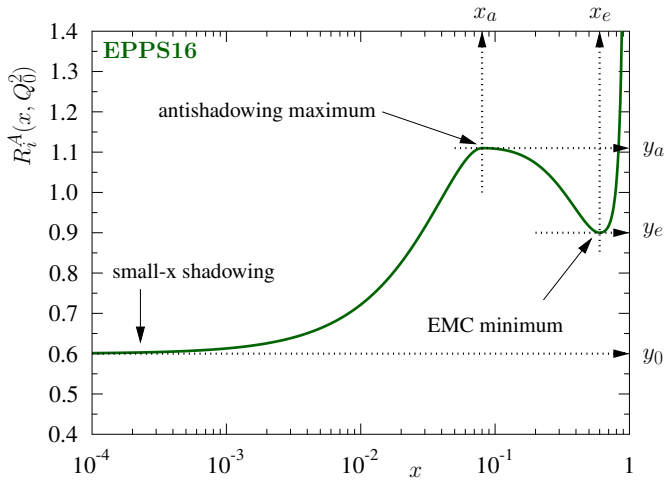



Figure 1.14: Representation of the nuclear PDF modification as a function of $x$ observed in deep-inelastic scattering measurements. Figure from Ref. [87].

*effect* [89]. It results from the modification of the quark and gluon distributions in bound nucleons by the nuclear environment. At higher $x$ values, the $R_i^{\mathrm{A}}$ increases due to the aforementioned Fermi motion. The heavy-flavour production in heavy-ion collisions at LHC energies is principally influenced by the shadowing effect. In fact, $\mathrm{c\overline{c}}$ and $\mathrm{b\overline{b}}$ pairs at midrapidity originate in hard scatterings between partons with Bjorken-$x$ values around $10^{-4}$ for charm quarks, and $10^{-3}$ for beauty quarks, which increase with increasing $p_{\mathrm{T}}$ [90]. Therefore, their yield normalised to the number of nucleon–nucleon collisions, which is dominated by the low-$p_{\mathrm{T}}$ region, is suppressed as compared to pp interactions. Similarly to the free-proton PDFs used for theoretical calculations in pp collisions, the nuclear PDFs are parametrised by using global fits to several experimental measurements of deep-inelastic scatterings and Drell-Yan processes [87, 91].

Another CNM effect is the Cronin enhancement [92]. This effect results in nuclear modification factors slightly larger than unity in p–A collisions at intermediate $p_{\mathrm{T}}$, due to the shift of the hadron $p_{\mathrm{T}}$ distributions towards higher values as compared to pp collisions. The Cronin enhancement can be explained by assuming that the partons of the projectile nucleon perform multiple elastic scatterings with the constituents of the target nucleus, before the hard-scattering process ($k_{\mathrm{T}}$-*broadening*). The multiple interactions transfer an initial transverse momentum to the partons, thus increasing the $p_{\mathrm{T}}$ of the produced particles. As discussed in Sec. 1.4.4, in presence of the QGP, a similar shift of the light-flavour hadron $p_{\mathrm{T}}$ spectra towards higher values is caused by the radial flow, which could also influence the $p_{\mathrm{T}}$ distributions of heavy-flavour hadrons. Since collective-like behaviours were observed in p–A collisions, hypotheses about the existence of hot-medium effects in

small systems have been advanced and are currently debated [93–95]. The presence of the radial flow in p–A collisions could constitute an alternative or complementary explanation of the Cronin enhancement with respect to the $k_{\mathrm{T}}$-broadening.

### 1.5.3 Interactions with the QGP

Heavy flavours propagating through the QGP interact with the deconfined medium constituents performing multiple elastic and inelastic scattering, which are usually referred to as *collisional* and *radiative* processes, respectively. The typical momentum exchanged in the scatterings is small compared to the charm- and beauty-quark mass. Therefore, the heavy-flavour propagation in the medium can be described as a Brownian motion characterised by several small momentum kicks [2]. The interactions with the QGP constituents bring low- and intermediate-$p_{\mathrm{T}}$ heavy quarks towards the thermal equilibrium with the medium and transfer to them the system collective flow. The heavy-flavour thermalisation time in the medium is comparable to the QGP lifetime or larger [96], thus the momentum spectra of heavy-flavour hadrons retain memory of the interaction history, providing information on the coupling strength to the medium.

The interactions with the constituents of the medium result in an energy loss for the partons traversing it. Fig. 1.15 displays the collisional and radiative energy loss of charm and beauty quarks in the QGP as a function of their initial energy predicted by the model of Ref. [97] for central Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV. According to these theoretical calculations, the collisional energy loss is the dominant mechanism for low-$p_{\mathrm{T}}$ heavy quarks. At high transverse momentum instead, the largest fraction of energy is lost via radiative processes, in which gluon emission induced by the medium presence occurs.

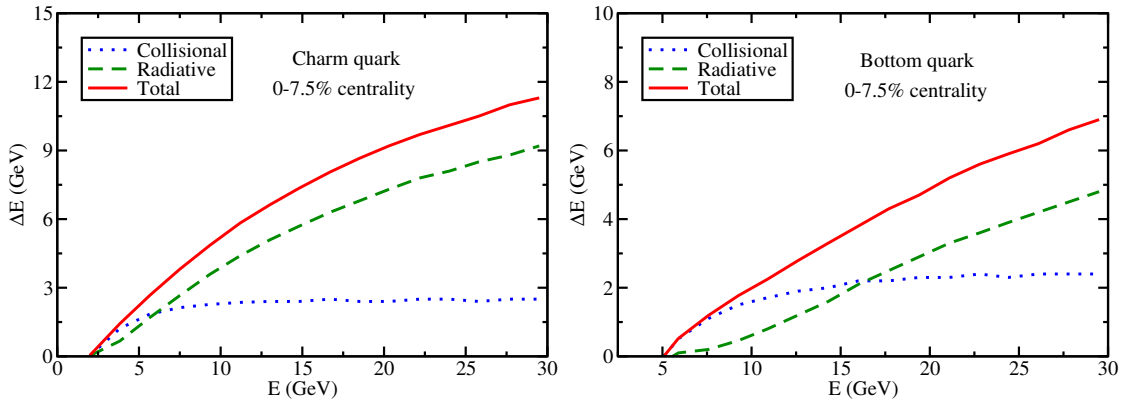



Figure 1.15: Comparison of radiative and collisional energy loss for charm quarks (left panel) and for bottom quarks (right panel) in the QGP formed in central Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV. Figure from Ref. [97].

### Collisional processes

The multiple elastic scatterings of heavy quarks with the medium constituents result in a collisional energy loss $\langle \Delta E_{coll} \rangle$ dependent on the quark mass [98]. In the limit of high quark energy, $E_{\mathrm{Q}} \gg m_{\mathrm{Q}}^2 T$, this energy loss is predicted to increase logarithmically as a function of the initial parton energy and to linearly depend on the medium thickness [99]. For low-energy quarks, the energy-momentum exchanges with the medium result in the thermalisation of heavy quarks in the QGP. Another consequence of the multiple soft interactions is the transfer to charm and beauty quarks of the radial and anisotropic flow of the system.

The dynamic and space-time evolution of heavy flavours undergoing multiple elastic scatterings in the QGP can be described in the transport formalism by the Boltzmann equation [2]

$$\left[\frac{\partial}{\partial t} + \frac{\boldsymbol{p}}{E_p}\frac{\partial}{\partial \boldsymbol{x}} + \boldsymbol{F}\frac{\partial}{\partial \boldsymbol{p}}\right] f_{\mathrm{Q}}(t, \boldsymbol{x}, \boldsymbol{p}) = C\left[f_{\mathrm{Q}}\right], \tag{1.25}$$

where $E_p$ is the quark energy, and $f_{\mathrm{Q}}(t, \boldsymbol{x}, \boldsymbol{p})$ indicates the heavy-quark phase-space distribution function. The interactions with the medium are encoded in the force $\boldsymbol{F}$ induced by the mean colour and electromagnetic fields, and in the collisional integral $C\left[f_{\mathrm{Q}}\right]$ containing the parton scattering amplitudes. In the Brownian-motion limit, $m_{\mathrm{Q}} \gg T$, which is valid for charm and beauty quarks as discussed above, the Boltzmann equation can be approximated by the Fokker-Planck equation

$$\frac{\partial}{\partial t} f_{\mathrm{Q}}(t, \boldsymbol{p}) = \frac{\partial}{\partial p_i}\left\{A_i(\boldsymbol{p}) f_{\mathrm{Q}}(t, \boldsymbol{p}) + \frac{\partial}{\partial p_j}\left[B_{ij}(\boldsymbol{p}) f_{\mathrm{Q}}(t, \boldsymbol{p})\right]\right\}, \tag{1.26}$$

which is a differential equation of the phase-space distribution function and depends only on the transport parameters $A_i(\boldsymbol{p})$ and $B_{ij}(\boldsymbol{p})$. The $A_i(\boldsymbol{p})$ coefficient quantifies the average heavy-flavour momentum variation per unit time, thus encoding the medium-induced friction, while the $B_{ij}(\boldsymbol{p})$ coefficients represent the average momentum broadening per unit time, i.e., the momentum diffusion.

In the non-relativistic limit, the transport coefficients can be assumed to be momentum independent, and Eq. (1.26) is further simplified in

$$\frac{\partial}{\partial t} f_{\mathrm{Q}}(t, p) = \gamma \frac{\partial}{\partial p_i}\left[p_i f_{\mathrm{Q}}(t, p)\right] + D_p \nabla_{\boldsymbol{p}} f_{\mathrm{Q}}(t, p), \tag{1.27}$$

which is the Langevin equation. In the equation, the $\gamma \equiv A_i(\boldsymbol{p})$ and $D_p \equiv B_{ij}(\boldsymbol{p})$ constants are the drag and momentum-diffusion coefficients, respectively. The drag coefficient $\gamma$ is the inverse of the heavy-flavour thermalisation (relaxation) time $\tau_{\mathrm{Q}}^{eq}$. It is also linked to the spatial-diffusion coefficient $D_s$, which quantifies the broadening of the spatial distribution with time and is defined as

$$D_s = \frac{T}{m_{\mathrm{Q}}\gamma} = \frac{T}{m_{\mathrm{Q}}}\tau_{\mathrm{Q}}^{eq}. \tag{1.28}$$

Common values of $\tau_{\mathrm{Q}}^{eq}$ estimated for charm quarks from data-to-model comparisons are of the order of 5 fm/$c$, while they are of the order of 15 fm/$c$ for beauty quarks due to their larger mass [2]. Therefore, the heavy-flavour thermalisation time is comparable to the lifetime of the QGP formed in heavy-ion collisions at LHC energies.

### Radiative processes

Partons propagating through the QGP medium characterised by free colour charges lose energy radiating gluons. Various calculations of the average radiative energy loss $\langle \Delta E_{rad} \rangle$, which adopt different assumptions, are available. In the BDMPS model [100], the medium is considered as composed of static scattering centres with which the parton performs multiple interactions acquiring a transverse momentum $k_{\mathrm{T}}$. An on-shell gluon is then emitted when the momentum acquired by the parton is large enough to decohere the gluon from the parton wave function. According to this picture, the model predicts the following dependence of the mean radiative energy loss

$$\langle \Delta E_{rad} \rangle \propto \alpha_s C_R \hat{q} L^2. \tag{1.29}$$

In the above equation, $C_R$ is the Casimir factor which is equal to 4/3 for the quark-gluon coupling and to 3 for the gluon-gluon one, $L$ is the distance traversed by the parton in the medium, and $\hat{q} = \langle k_{\mathrm{T}}^2 \rangle / L$ is the transport coefficient quantifying the average squared transverse momentum transferred to the parton per path-length unit. In this model, the average energy loss does not depend on the parton energy and is larger by a factor 9/4 for gluons than for quarks. Therefore, the effect of the energy loss is expected to be larger for light-flavour hadrons, which mainly originate from the hadronisation of gluons, than for heavy-flavour hadrons. The values of the transport coefficient normalised to the medium temperature $\hat{q}/T^3$ resulting from data-to-model studies are comprehended between about 1.5 and 4 [101].

The radiative energy loss is also affected by the so-called *dead cone effect* [102, 103], where the emission of gluons from a quark is forbidden at angles around its direction of motion smaller than a critical value $\Theta_c = m_{\mathrm{Q}}/E_{\mathrm{Q}}$. Hence, the energy loss is expected to decrease with increasing quark mass $m_{\mathrm{Q}}$, and to increase as the quark energy $E_{\mathrm{Q}}$ increases. The dead-cone effect and the dependence described in Eq. 1.29 imply that the radiative energy loss of partons in the QGP is expected to follow the hierarchy $\Delta E(\mathrm{b}) < \Delta E(\mathrm{c}) < \Delta E(\mathrm{u,d,s}) < \Delta E(\mathrm{g})$.

As visible in Fig. 1.15, the radiative energy loss is expected to be dominant with respect to the collisional one at high $p_{\mathrm{T}}$. Therefore, the radiative energy loss is the main effect leading to the suppression of the hadron yield at high $p_{\mathrm{T}}$, and comparing the $R_{\mathrm{AA}}$ of heavy- and light-flavour hadrons it is possible to probe the colour-charge and quark-mass dependence of the energy-loss mechanism. In Fig. 1.16, the average $R_{\mathrm{AA}}$ of prompt $\mathrm{D}^0$, $\mathrm{D}^+$, and $\mathrm{D}^{*+}$ mesons [104] is reported together with the one of charged pions [37], charged particles [46], and inclusive

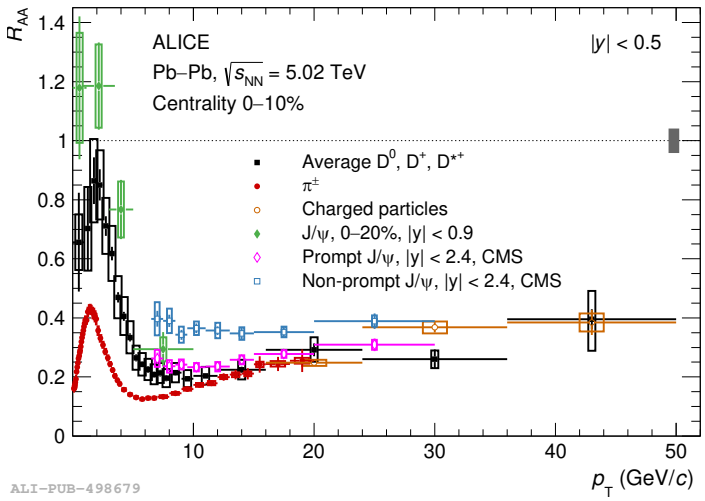



Figure 1.16: Measurements of the $R_{\mathrm{AA}}$ of prompt D mesons, charged pions, charged particles, and J/ψ mesons in central Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. Figure from Ref. [104].

J/ψ mesons [105] measured by the ALICE Collaboration in central Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. The $R_{\mathrm{AA}}$ of prompt and non-prompt J/ψ mesons measured by the CMS Collaboration [106] is also shown. The $R_{\mathrm{AA}}$ of D mesons and charged pions are similar for $p_{\mathrm{T}} \gtrsim 8$ GeV/$c$. This is due to the interplay of the larger energy loss expected for gluons and light-flavour quarks than for charm quarks, and the harder charm-quark $p_{\mathrm{T}}$ distribution and fragmentation function. Moreover, the nuclear modification factor of non-prompt J/ψ mesons from beauty-hadron decays is higher than the prompt D-meson one, as predicted by the quark-mass dependence of the radiative energy loss.

The dependence of the radiative energy loss on the distance travelled by the parton in the medium is the origin of the positive elliptic flow observed at high $p_{\mathrm{T}}$ in non-central collisions. In this regime, the elliptic flow results from the path-length asymmetries arising from the spatial anisotropy of the fireball in the first instants of the collision, which lead to different energy losses depending on the heavy-flavour azimuthal direction [107, 108].

### Theoretical description of heavy-flavour measurements

The prediction of experimental observables by theoretical models describing the heavy-flavour transport in the hydrodynamically expanding QGP is a challenging task. This endeavour requires the modelling of many complex ingredients, such as: the initial conditions produced in the heavy-ion collisions, the QGP evolution and collective behaviours, the initial momentum and position distribution of the charm and beauty quarks, the interactions of heavy flavours in the medium and

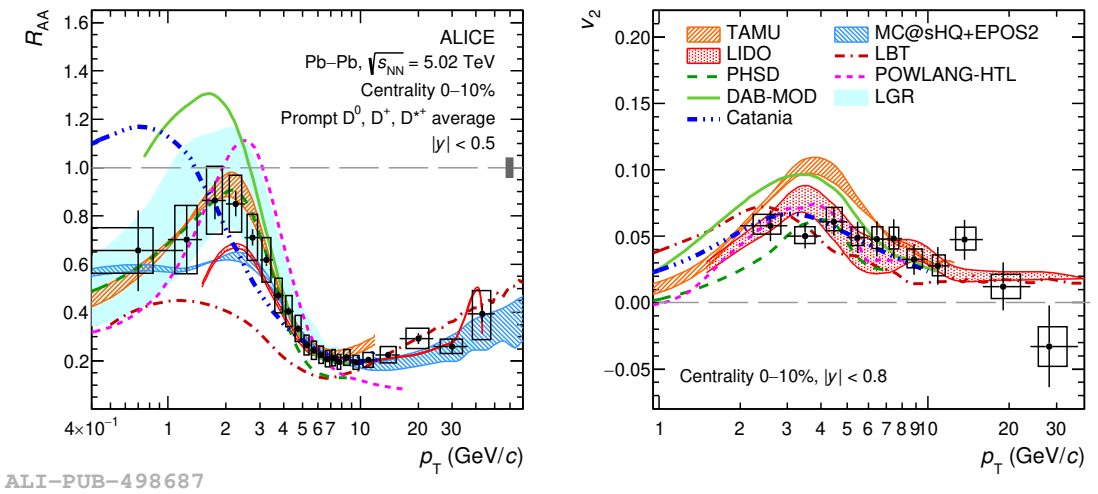



Figure 1.17: Average $R_{AA}$ (left panel) and $v_2$ (right panel) of prompt $D^0$, $D^+$, and $D^{*+}$ mesons in central Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, compared with predictions of models implementing the charm-quark transport in a hydrodynamically expanding medium. Figure from Ref. [104].

their hadronisation, and possibly the interactions of heavy-flavour hadrons in the hadronic phase. In particular, the simultaneous description of the nuclear modification factor $R_{AA}$ and the elliptic flow $v_2$ of heavy-flavour hadrons is a stringent test of model predictions and provides the possibility to constrain the model ingredients and parameters.

In Fig. 1.17, the average $R_{AA}$ [104] and $v_2$ [109] of prompt $D^0$, $D^+$, and $D^{*+}$ mesons as a function of $p_T$ in central Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV measured by the ALICE Collaboration are compared to transport-model predictions. The TAMU [110], POWLANG-HTL [111, 112], DAB-MOD [113], and LGR [114] models describe the heavy-flavour transport in the QGP by using the Langevin equation. The Catania [115, 116], MC@sHQ+EPOS2 [117], LBT [118, 119], and LIDO [120] models are instead based on the Boltzmann equation. The parton transport in PHSD [121] is described with the dynamical quasi-particle model, which reproduces lattice QCD results for the QGP in thermodynamic equilibrium by using effective propagators for quarks and gluons. The heavy-quark interactions with the medium are treated by considering only collisional processes in the case of TAMU, POWLANG-HTL, PHSD, and Catania, while in the other theoretical calculations radiative processes are also included. The mentioned models, except for DAB-MOD, adopt nuclear PDFs to describe initial-state effects affecting the heavy-quark initial $p_T$ distribution. Finally, all the models include the possibility for charm and beauty quarks to hadronise via the recombination mechanism, which will be discussed in the next Section, in addition to the hadronisation by

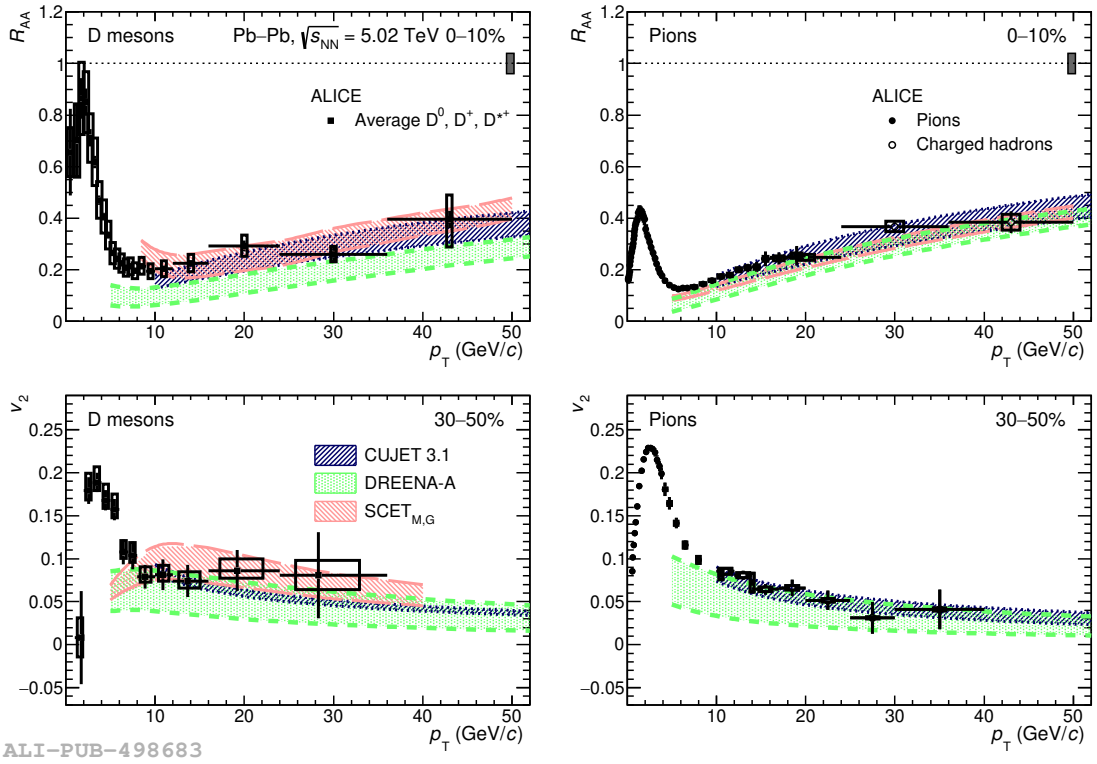



Figure 1.18: $R_{\rm AA}$ (top panels) and $v_2$ (bottom panels) of prompt D mesons, charged pions, and charged hadrons in Pb–Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV, compared with predictions from models based on pQCD calculations. Figure from Ref. [104].

fragmentation.

The agreement between the measurements and the theoretical predictions is evaluated by computing the $\chi^2/ndf$. The models describing fairly well both the measured $R_{\rm AA}$ and $v_2$ are TAMU, MC@sHQ+EPOS2, LIDO, LGR, and Catania, where a $\chi^2/ndf$ below 5 and 2 is imposed as a threshold in the $R_{\rm AA}$ and $v_2$ comparison, respectively. Considering these models, it is possible to constrain the value of the heavy-flavour spatial diffusion coefficient $D_s$ to the range $1.5 < 2\pi D_s < 4.5$ at the pseudo-critical temperature $T_{pc} = 155$ MeV, which is in agreement with lattice QCD calculations [104].

Figure 1.18 shows the average nuclear modification factor $R_{\rm AA}$ and elliptic flow $v_2$ of prompt D$^0$, D$^+$, and D$^{*+}$ mesons [104, 109], together with those of charged pions [37, 44] and charged hadrons [46], measured by the ALICE Collaboration in Pb–Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV. The measurements are compared with the predictions of the CUJET3.1 [122–124], DREENA-A [125–127], and SCET$_{\rm M,G}$ [128] models, which implement pQCD calculations of the energy loss of high-$p_{\rm T}$ partons. The comparison of these models with the measurements provides a quantitative

understanding of the parton energy loss in the medium. In the CUJET3.1 framework, the jet-medium interactions are described with the DGLV opacity expansion model [129]. The DREENA-A model is based on a dynamical energy-loss formalism [130], where jet-medium interactions are modelled in a QCD medium of finite size and temperature composed of dynamical scattering centres. These frameworks include both the collisional and radiative energy loss mechanisms. The $\mathrm{SCET_{M,G}}$ model is instead based on a soft-collinear effective theory [131] and implements the medium-induced gluon radiation via modified splitting functions. All models are in agreement with both the $R_{\mathrm{AA}}$ and $v_2$ of D mesons and charged pions for $p_{\mathrm{T}} > 10$ GeV/$c$, where the radiative energy loss is expected to be the dominant interaction mechanism. This indicates that the colour-charge and quark-mass dependence of the radiative energy loss in the medium, as well as its path-length dependence, are reasonably well described in the model calculations.

### 1.5.4 Hadronisation mechanisms

The charm and beauty quarks produced in the initial hard partonic scatterings undergo a process to generate colour-singlet hadrons in the final state, which is referred to as the *hadronisation*. In pp collisions, the heavy quarks hadronise in the vacuum via the *fragmentation* mechanism producing a spray of lower-momentum hadrons. The heavy-quark fragmentation into mesons is found to be the same in $\mathrm{e^+e^-}$, $\mathrm{e^{\pm}p}$, and pp collisions, pointing to a universal mechanism independent of the collision system [132]. A common approach to describe the fragmentation mechanism is the Lund string model [133, 134], which is adopted by the default tune of the PYTHIA event generator [135]. In this model, the hadronisation occurs via the breaking of colour strings with the production of $\mathrm{q\overline{q}}$ pairs from the vacuum. However, this picture fails to describe the $\Lambda_{\mathrm{c}}^{+}$-baryon production, which is significantly larger in pp collisions than what is expected by adopting fragmentation functions constrained by $\mathrm{e^+e^-}$ and $\mathrm{e^{\pm}p}$ data. Alternative theoretical approaches to describe this observation are currently under discussion [69, 136].

In heavy-ion collisions, the hadronisation process is predicted to be modified by the presence of the deconfined QGP medium. When the system reaches the pseudo-critical temperature for the transition to a hadron gas, heavy quarks can hadronise by combining with light-quarks of the medium close in phase space [137, 138]. This mechanism is called *recombination* or *coalescence*, and produces hadrons with higher momentum than the one of the hadronising charm and beauty quarks. The recombination mechanism is in competition with the fragmentation and its contribution is relevant for the production of low- and intermediate-$p_{\mathrm{T}}$ hadrons, while the fragmentation mechanism is dominant at high $p_{\mathrm{T}}$.

The inclusion of the hadronisation via recombination in the theoretical models is crucial for the description of open heavy-flavour hadron observables. In Fig. 1.19, the predictions of the PHSD [121], POWLANG-HTL [111, 112], and

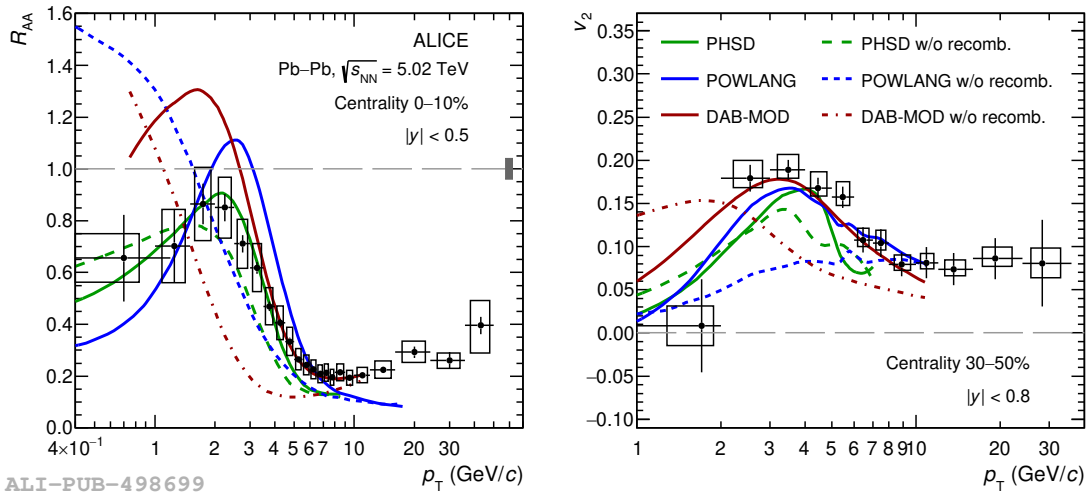



Figure 1.19: Prompt D-meson $R_{\mathrm{AA}}$ in the 0–10% centrality class (left panel) and $v_2$ in the 30–50% centrality class (right panel) in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV, compared with model predictions including (solid lines) and not including (dashed lines) the hadronisation via recombination. Figure from Ref. [104].

DAB-MOD [113] models, both including and not including the hadronisation via recombination, are compared to the prompt D-meson $R_{\mathrm{AA}}$ [104] and $v_2$ [109] measured by the ALICE Collaboration in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. The calculations considering only the fragmentation process underestimate both the observed D-meson $R_{\mathrm{AA}}$ and $v_2$. Hence, the recombination mechanism is needed to push the predictions closer to the experimental data, indicating that the charm-quark coalescence with light-flavour quarks of the medium plays a relevant role. This is also consistent with J/ψ-meson measurements, where the recombination of charm and anticharm quarks is fundamental to describe the data, as discussed in Section 1.4.7. Finally, when including the hadronisation via recombination, the $R_{\mathrm{AA}}$ shows a bump at intermediate $p_{\mathrm{T}}$ originating from the radial flow of the medium, and the $v_2$ increases since the elliptic flow of light-flavour quarks is added to the charm-quark one.

Different implementations of the quark-recombination mechanism are adopted in the various theoretical frameworks. The majority of the models mentioned in the previous Section uses an instantaneous coalescence at the phase boundary, which is implemented through the Wigner-function formalism [139]. In the PHSD model instead, the hadronisation is described via a Monte Carlo simulation of the recombination process in competition to the fragmentation. The POWLANG-HTL model implements the recombination through an in-medium string formation approach. In the TAMU model, the hadronisation via coalescence occurs via the formation of resonant states when approaching the pseudo-critical temperature within the

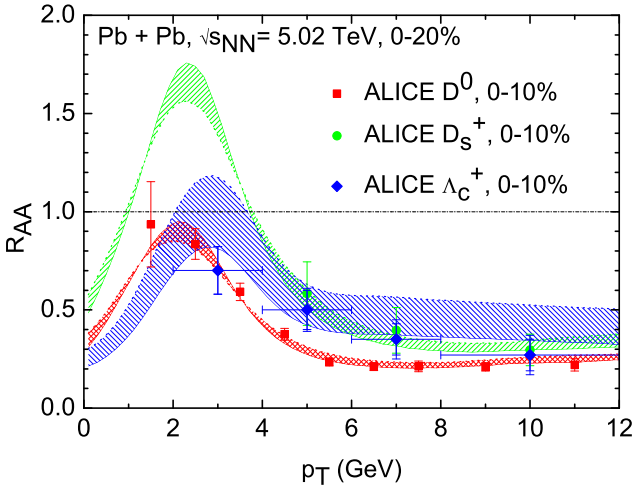



Figure 1.20: Predictions of the TAMU model for the $R_{\rm AA}$ of prompt $\rm D^0$, $\rm D_s^+$ and $\Lambda_c^+$ hadrons in central Pb–Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV compared to the measurements performed by the ALICE Collaboration. Figure from Ref. [110].

resonance-recombination model formalism [140].

As a further consequence of the charm-quark hadronisation via recombination, the $\rm D_s^+$-meson production is expected to be enhanced with respect to the non-strange D-meson one in nucleus–nucleus collisions at low and intermediate $p_{\rm T}$ in comparison with pp interactions, due to the large abundance of strange quarks in the QGP medium [141, 142]. This results in a smaller suppression of the $\rm D_s^+$-meson $R_{\rm AA}$ compared to D mesons without strange-quark content in the $p_{\rm T}$ region where the coalescence mechanism is relevant. Moreover, the hadronisation via recombination gives rise to an enhancement also of the baryon-over-meson yield ratios at intermediate $p_{\rm T}$. Figure 1.20 reports the TAMU model predictions for the nuclear modification factor of prompt $\rm D^0$, $\rm D_s^+$ and $\Lambda_c^+$ hadrons in central Pb–Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV compared to the measurements performed by the ALICE Collaboration[6] [110]. The model predictions show a hierarchy in the $R_{\rm AA}$ of the three charm-hadron species at low transverse momentum, while the calculations for the different hadrons converge to similar values at high $p_{\rm T}$, where the fragmentation is the dominant hadronisation mechanism. The larger nuclear modification factor predicted by the model for the $\rm D_s^+$ mesons is motivated by the interplay of the charm-quark coalescence in the medium and the strangeness enhancement.

[6]The measurements used in the paper are not the latest available from the ALICE Collaboration.

### 1.5.5 SHM for charm hadrons

As mentioned in Section 1.4.2, the statistical hadronisation model (SHM) successfully describe the production of light-flavour hadrons in ultrarelativistic heavy-ion collisions. The SHM model is extended to describe the charm-hadron production by the GSI-Heidelberg SHMc model [143], where charm quarks are treated as "impurities" of the QGP medium characterised by thermal distributions. The charm-quark thermal production in the QGP is strongly suppressed at LHC energies, since the medium temperature reached is substantially lower than the $\mathrm{c\bar{c}}$-pair mass. Hence, the charm-quark content of the deconfined medium is determined by the initial production in hard-scattering processes, and it is an input parameter for the model. The SHMc model assumes that charm quarks thermalise in the QGP due to their interactions with the medium constituents. Therefore, they can be distributed into hadrons with thermal weights at the moment of the chemical freeze-out, and the abundances of the different charm-hadron species can be computed starting from the thermal parameters, i.e., $T_{ch}$, V, and $\mu_B$, extracted from the fit to the light-flavour hadron yields.

In addition to the total yields of the charm-hadron species, the SHMc model provides predictions also for their $p_{\mathrm{T}}$ spectra, which are modelled with a core-corona approach [144]. In this approach, the low-$p_{\mathrm{T}}$ region is dominated by the core contribution, where charm quarks are considered in local thermal equilibrium with the QGP formed in the collision, and their $p_{\mathrm{T}}$ distribution is described by a Blast-Wave function. The $p_{\mathrm{T}}$-spectra modification due to resonance decays is also accounted for by using the FastReso package [145]. In the corona, which is the volume corresponding to the outer region of the collision zone, the energy density is much smaller than the maximum value achieved in the collision and the assumption of no QGP formation is made. Therefore, the corona contribution is parametrised from measurements in pp collisions scaled by the number of nucleon-nucleon interactions occurring in the corona. The final charm-hadron $p_{\mathrm{T}}$ spectra are obtained by summing the thermal component originating from the core to the one from the corona. This latter contribution, being harder, is relevant at high $p_{\mathrm{T}}$.

Figure 1.21 shows the SHMc predictions for the $p_{\mathrm{T}}$-differential production yield of prompt $\mathrm{D^0}$ mesons and $\Lambda_{\mathrm{c}}^{+}$ baryons in central Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. The $\mathrm{D^0}$-meson prediction is compared with the measurement performed by the ALICE Collaboration in the same collision system and at the same centre-of-mass energy [146]. The model describes the data within uncertainties at low transverse momentum, while it undershoots the $\mathrm{D^0}$ yield for $p_{\mathrm{T}} \gtrsim 4$ GeV/$c$. Even if the SHMc model does not catch the measurement at intermediate and high $p_{\mathrm{T}}$, it provides relevant predictions for the D-meson total production yields, which are dominated by the low-$p_{\mathrm{T}}$ contribution, and for the $p_{\mathrm{T}}$-differential yield ratios between the different D-meson species.

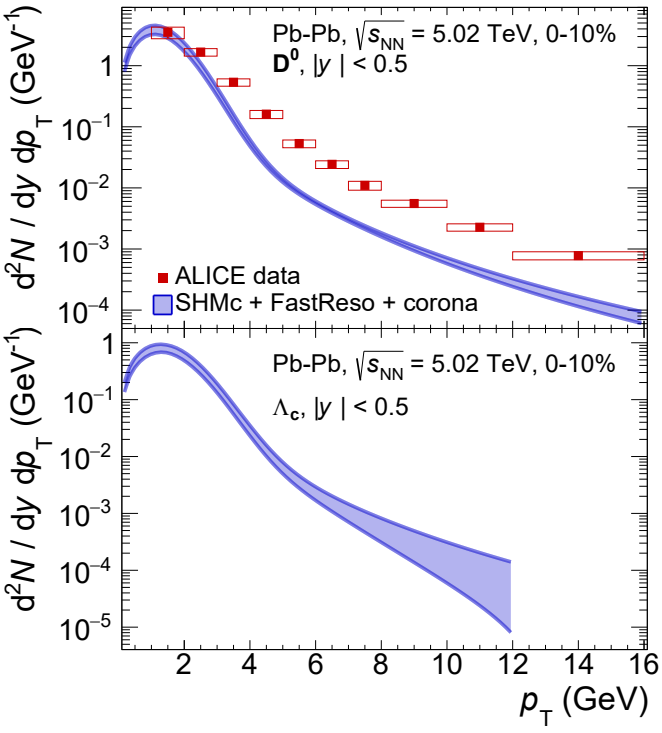



Figure 1.21: SHMc predictions for the $p_{\mathrm{T}}$-differential production yield of prompt $\mathrm{D}^0$ mesons (top panel) and $\Lambda_{\mathrm{c}}^+$ baryons (bottom panel) in central Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. The $\mathrm{D}^0$-meson prediction is compared with the measurement performed by the ALICE Collaboration. Figure from Ref. [143].

### 1.5.6 Motivation for $\mathrm{D_s^+}$-meson measurements

The measurement of the $\mathrm{D_s^+}$-meson production in pp collisions complements those of non-strange D mesons, i.e., $\mathrm{D}^0$ and $\mathrm{D}^+$, and contributes to provide information on the charm- and beauty-quark production cross section, via the measurement of prompt and non-prompt $\mathrm{D_s^+}$ mesons, respectively. In addition, from the measurement of the $\mathrm{D_s^+}/(\mathrm{D}^0 + \mathrm{D}^+)$ yield ratios is possible to derive the fragmentation fraction of charm and beauty quarks into strange mesons divided by the one into non-strange mesons $f_{\mathrm{s}}/(f_{\mathrm{u}} + f_{\mathrm{d}})$. Therefore, the heavy-quark hadronisation via fragmentation in the vacuum can be probed.

As mentioned in Section 1.5.4, the measurement of the $\mathrm{D_s^+}$-meson production in Pb–Pb collisions and its comparison with the non-strange D-meson one is a powerful tool to investigate the role of the charm-quark hadronisation via recombination in the deconfined medium, where the strangeness production is enhanced. In particular, precise measurements are needed to confirm the enhancement of the charm-strange meson production in heavy-ion collision at low and intermediate $p_{\mathrm{T}}$, and to constrain the predictions of models describing the charm-quark transport

and hadronisation in the QGP, as well as those of the SHMc. Moreover, measurements of the $\mathrm{D_s^+}$-meson elliptic flow in non-central heavy-ion collisions have not yet reached the precision required to claim that a positive $v_2$ is observed, as for the non-strange D mesons. A difference between the elliptic flow of $\mathrm{D_s^+}$ and non-strange D mesons is also predicted due to the different D-meson masses, the charm-quark hadronisation via coalescence with strange quarks of the medium instead of light quarks, and different interactions in the hadronic phase [142, 147]. Hints of this possible difference could be obtained from measurements with improved precision.

# Chapter 2

# A Large Ion Collider Experiment — ALICE

The ALICE (A Large Ion Collider Experiment) [148] experiment is one of the four main experiments installed at the CERN Large Hadron Collider (LHC) [149] near Geneva. ALICE design was optimised to study heavy-ion collisions at ultrarelativistic centre-of-mass energies. Its main focus is the characterisation of the quark-gluon plasma formed at extreme values of temperature and energy density in Pb–Pb collisions. The physics program of the experiment also includes the study of lighter-ions, p–Pb, and pp collisions to address various QCD topics. In this Chapter, the detectors composing the ALICE experiment are introduced, together with some details about their performance. The computing and software framework of the ALICE Collaboration is then outlined. Finally, the data acquisition and reconstruction strategies are briefly discussed.

## 2.1 The Large Hadron Collider

The LHC [149] is a hadron accelerator and collider installed in the laboratory of the European Organisation for Nuclear Research (CERN). It is situated underground, at a depth varying between 45 m and 170 m, in the tunnel where the Large Electron Positron (LEP) collider [150] was located. The LHC is currently the largest and most powerful particle accelerator in the world, thanks to a circumference of about 26.7 km and the adoption of superconducting magnets. The LHC is designed to accelerate protons, and provides pp collisions with a centre-of-mass energy up to $\sqrt{s} = 14$ TeV and with a peak instantaneous luminosity of $\mathcal{L} = 10^{34}$ s$^{-1}$cm$^{-2}$. It can also operate with lead ions. In this case, Pb-Pb collisions with a centre-of-mass energy per nucleon pair up to $\sqrt{s_{\mathrm{NN}}} = 5.5$ TeV and a luminosity of about $\mathcal{L} = 10^{27}$ s$^{-1}$cm$^{-2}$ can be achieved.

Before entering the LHC ring, beams of protons and lead ions are prepared

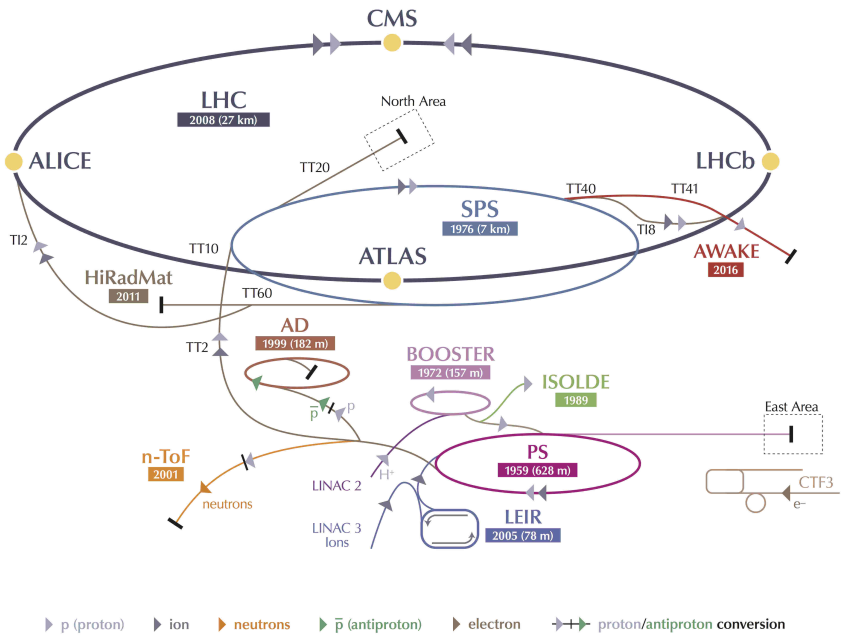



Figure 2.1: Schematic of the LHC accelerator and its injection system. The four main LHC experiments are also shown. Figure adapted from the original [153].

and accelerated by a complex injection system [151, 152]. This system is composed of reconverted accelerators and dedicated facilities, and it is schematised in Fig. 2.1. Protons are obtained by ionising hydrogen atoms, and are brought to an energy of about 50 MeV by the LINear ACcelerator 2 (LINAC2). They are subsequently accelerated by the Proton Synchrotron Booster (PSB) to about 1.4 GeV, and then are injected in the Proton Synchrotron (PS) where they reach an energy of about 25 GeV. Finally, the Super Proton Synchrotron accelerates the protons to 450 GeV before the injection in the LHC ring. Lead ions are produced from the evaporation of metallic lead which is then ionised. They initially pass through the LINear ACcelerator 3 (LINAC3), after which they reach an energy of about 4.2 MeV/nucleon. Then, they are accelerated by the Low Energy Ion Ring (LEIR) to about 72 MeV/nucleon. At this point, the lead ions are injected in the PS and follow the same path of the protons before being delivered to the LHC.

The LHC provides the beams for four major experiments, whose positions are schematically shown in Fig. 2.1. The ATLAS (A Toroidal LHC ApparatuS) [154] and CMS (Compact Muon Solenoid) [155] experiments are designed to exploit the high luminosity provided by the LHC in pp collisions. They discovered the Higgs boson in 2012 [156, 157] and have a wide physics program comprehending various topics, such as the top-quark and Higgs-coupling measurements, and the search for physics beyond the Standard Model. The LHCb (Large Hadron

Collider beauty) [158] experiment is devoted to heavy-flavour physics, studying charge-parity violation and rare decays of hadrons containing charm and beauty quarks. The ALICE experiment is mainly dedicated to the physics of high-energy heavy-ion collisions, and will be described in detail in the following Section. It is worth mentioning that all the four main experiments have a dedicated heavy-ion program.

## 2.2 The ALICE experiment

ALICE is a general-purpose detector designed to cope with the large multiplicity of charged particles produced at midrapidity in Pb–Pb collisions at the LHC [148, 159]. The apparatus was optimised considering a value of charged-particle multiplicity per pseudorapidity unit, $\mathrm{d}N/\mathrm{d}\eta$, of about 4000, and the experiment detectors are characterised by a high granularity. To reconstruct particles down to very low transverse momentum, about 80 MeV/$c$ for pions, a magnetic field of 0.5 T parallel to the beam direction is used. Moreover, the tracking detectors were designed minimising the material budget to guarantee a good $p_{\mathrm{T}}$ resolution. At high $p_{\mathrm{T}}$, a satisfactory resolution is provided by the large tracking lever arm of about 3.5 m. The physics studied by the ALICE experiment requires particle identification (PID) in a wide momentum range, i.e., from low $p_{\mathrm{T}}$ to about 20 GeV/$c$. This is obtained thanks to detectors with complementary capabilities which exploit different PID techniques.

The ALICE experiment is 26 m long, 16 m high, 16 m wide and weights approximatively 10000 t. The layout of the ALICE apparatus is displayed in Fig. 2.2, where the different components are highlighted. The apparatus is formed by two main sections, i.e., a central barrel and a forward muon spectrometer. The central barrel covers the pseudorapidity interval $-0.9 < \eta < 0.9$, and it is contained in the large solenoid magnet previously employed in the L3 experiment at LEP. Moving from the point where the beams collide outward, the barrel includes: the Inner Tracking System (ITS) which is made of silicon detectors of different technologies; a Time Projection Chamber (TPC); the Transition Radiation (TRD), the Time-of-Flight (TOF), and the Ring Imaging Cherenkov (HMPID) detectors dedicated to particle identification; and three electromagnetic calorimeters (PHOS, EMCal, and DCal). These cylindrical detectors are specialised in the measurement of hadrons, electrons, and photons. The ITS, TPC, TOF and TRD detectors cover the full azimuth. The forward muon spectrometer is dedicated to the triggering and tracking of muons, and it is composed of various absorbers, a dipole magnet, ten planes of cathode pad chambers, and four planes of resistive plate chambers. The muon spectrometer covers the pseudorapidity range $-4 < \eta < -2.5$ and the full azimuth. In addition, several detectors of small dimensions (PMD, FMD, T0, V0, ZDC) are located at forward and backward rapidity. They are mainly used for global event

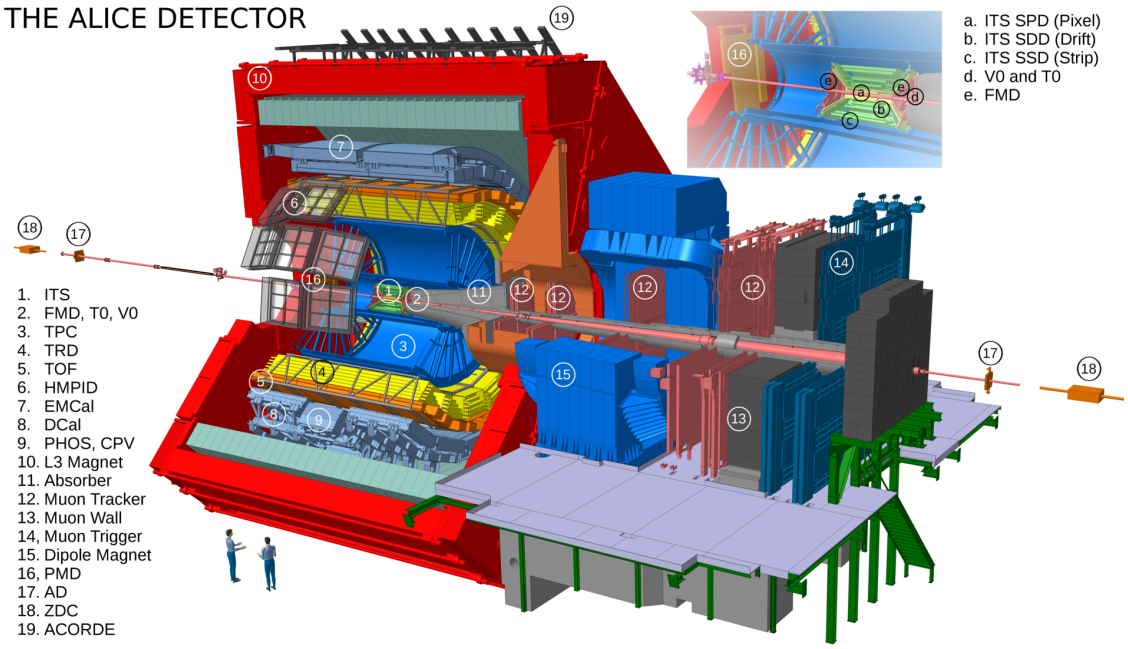



Figure 2.2: Detector layout of the ALICE experiment. Figure from the ALICE figure repository ©.

characterisation and triggering. An array of scintillators (ACORDE) is located on top of the L3 magnet, and it is employed to trigger on muons originating from cosmic rays, which can be used for alignment purposes and cosmic-ray measurements. In the following Sections, the ALICE detectors used in the analyses presented in this Thesis will be described more thoroughly.

The ALICE experiment adopts a right-handed orthogonal Cartesian coordinate system which originates in the LHC Interaction Point 2 (IP2), where the experiment is located. The $\hat{z}$ direction is parallel to the direction of the beams and points towards the ATLAS experiment. The $\hat{x}$ direction is horizontal and points towards the centre of the LHC ring, while the $\hat{y}$ direction is vertical pointing upwards. Moreover, ALICE uses also a spherical coordinate system, where the azimuthal angle $\varphi$ is defined in the $xy$ plane and increases counter-clockwise starting from the $\hat{x}$ direction. The polar angle $\theta$ is instead defined in the $yz$ plane.

### 2.2.1 Inner Tracking System

The Inner Tracking System is the ALICE detector closest to the interaction point. The ITS, which is schematised in Fig. 2.3, is composed of six cylindrical layers of silicon detectors placed coaxially around the beam vacuum tube. The layers are instrumented with sensors based on different technologies. The two innermost layers are silicon pixel detectors (SPD), the third and fourth ones are silicon drift

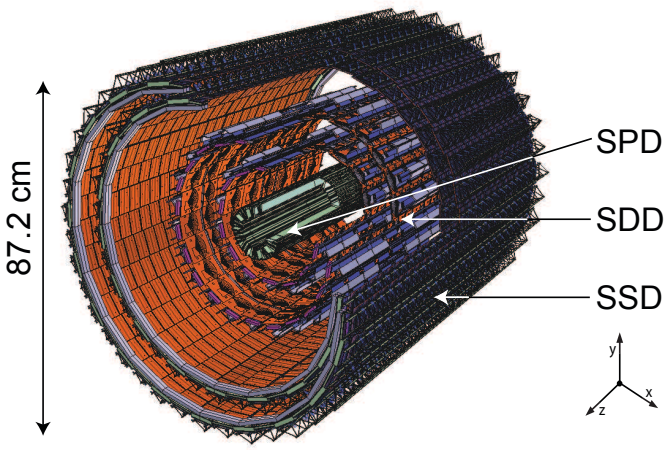



Figure 2.3: Structure of the Inner Tracking System with its six silicon layers. Figure from Ref. [160].

detectors (SDD), and the last two layers are double-sided silicon strip detectors (SSD). The details about the position, spatial extension, material budget, and spatial resolution of each layer are reported in Tab. 2.1. The ITS is characterised by a low total material budget, which reduces the particle multiple scattering with the detector structures. A particle crossing the ITS at $\eta = 0$ traverses an amount of material corresponding to about 7.2% of a radiation length $X_0$, including the support structures and the thermal shields.

The SPD detector is highly segmented to sustain the large charged-particle multiplicity of heavy-ion collisions. Each sensor matrix composing the detector contains $256 \times 160$ cells measuring 50 µm ($r\varphi$) $\times$ 425 µm ($z$), for a total sensor

Table 2.1: Properties of the six layers of the ITS detector [148].

| Layer | Detector | $r$ (cm) | $\pm z$ (cm) | Material budget ($\%X_0$) | Spatial resolution $r\varphi \times z$ (µm$^2$) |
|---|---|---|---|---|---|
| 1 | SPD | 3.9 | 14.1 | 1.14 | $12 \times 100$ |
| 2 | SPD | 7.9 | 14.1 | 1.14 | $12 \times 100$ |
| 3 | SDD | 15.0 | 22.2 | 1.13 | $35 \times 25$ |
| 4 | SDD | 23.9 | 29.7 | 1.26 | $35 \times 25$ |
| 5 | SSD | 38.0 | 43.1 | 0.83 | $20 \times 830$ |
| 6 | SSD | 43.0 | 48.9 | 0.86 | $20 \times 830$ |

active area of 12.8 mm ($r\varphi$) × 70.7 mm ($z$). The thickness of the sensor is 200 µm. The SPD has a binary readout and its fast response provides the opportunity to use its signal in a trigger. The SDD detector is formed by 300 µm thick sensors with a sensitive area of 70.17 mm ($r\varphi$) × 75.26 mm ($z$), which is split into two separated drift regions by a central cathode strip. Each drift region has 256 anodes for charge collection with a pitch of 294 µm. The SDD has an excellent multitrack capability, and thanks to an analogue readout delivers two of the four measurements of ionisation energy loss per unit length $\mathrm{d}E/\mathrm{d}x$ employed for particle identification by the ITS. The sensors of the SSD detector have 768 strips on each side with 95 µm pitch, an active area of 75 mm ($r\varphi$) × 42 mm ($z$), and are 300 µm thick. The SSD is essential for the prolongation of tracks from the TPC to the ITS, which will be discussed in Section 2.5, and provides the other two $\mathrm{d}E/\mathrm{d}x$ samples needed for particle identification.

The ITS performs the following main tasks:

- improvement of the spatial, angular, and momentum resolution for tracks reconstructed by the TPC;
- localisation of the collision position, i.e., the primary vertex, with a resolution better than 100 µm in pp collisions and 3–4 µm in central Pb–Pb collisions;
- reconstruction of the secondary vertices which originate from the decays of short-lived particles as heavy-flavour hadrons ($c\tau(\mathrm{D_s^+}) \simeq 151$ µm, $c\tau(\mathrm{B^0}) \simeq 455$ µm [6]), resolving them from the primary one;
- tracking and identification of particles with momentum below 200 MeV/$c$, which do not reach the TPC.

### 2.2.2 Time Projection Chamber

The Time Projection Chamber is the primary detector for track reconstruction in the central barrel. The TPC is a chamber filled with high-purity gas designed to transport ionisation electrons, which are produced by charged particles traversing the gas, over long distances and to collect them at the readout end plates. Figure 2.4 shows the layout of the detector, which is composed of a cylindrical field cage with an overall length of about 5 m along the beam direction, and an inner and outer radius of about 0.85 m and 2.5 m, respectively. The total active volume of the detector is about 90 $\mathrm{m}^3$, being the largest TPC mounted on an experimental apparatus to date.

The field cage of the TPC is formed by a central high-voltage electrode, which separates the gas volume into two specular drift regions, and by two opposite axial resistive potential dividers, which generate a highly uniform electrostatic field of about 400 V/cm. The maximum drift time in the chamber is 90 µs. During the

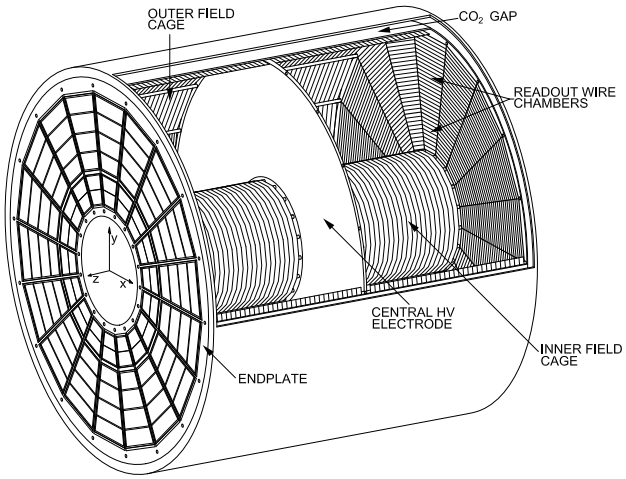



Figure 2.4: Layout of the Time Projection Chamber. Figure from Ref. [161].

Run 2 data-taking period (2015–2018), the gas contained in the cage was composed of a mixture of Ar, $CO_2$, and $N_2$ (90/10/5), except in 2017 when the Ar component was replaced with Ne. The TPC end plates are segmented into 18 trapezoidal sectors in the azimuthal direction. Each sector is divided in an inner and an outer region, which are instrumented with multi-wire proportional chambers (MWPC). The two MWPC together contains 159 readout pad rows corresponding to the cathodes collecting the avalanche created by the drifting electrons. The spatial resolution of the TPC detector is 1.1 mm ($r\varphi$) $\times$ 0.8 mm ($z$) (1.25 mm ($r\varphi$) $\times$ 1.1 mm ($z$)) for the inner (outer) readout chambers.

In addition to the tracking, the TPC is optimised to identify charged particles via their $\mathrm{d}E/\mathrm{d}x$, and to determine the primary-vertex position together with the ITS. The TPC provides relevant PID information at low $p_{\mathrm{T}}$, i.e., up to about 0.8 GeV/$c$ for pions and kaons, and below 1.6 GeV/$c$ for protons. Further information on the PID capabilities of the TPC detector is reported in Section 2.6.

### 2.2.3 Time-of-Flight detector

The Time of Flight detector is positioned outside the TRD and has a large active area of about 141 m$^2$. The detector is contained in a cylindrical shell with a length of 7.4 m, an internal radius of 3.7 m and an external one of 4 m. It is divided in 18 sectors in the azimuthal direction and 5 segments along the $z$ axis. The TOF is instrumented with 1593 strips of multi-gap resistive plate chambers (MRPC) in double-stack configuration, each one with an active area of 120 $\times$ 7.4 cm$^2$. The MRPC are filled with a gas mixture of 90% $C_2H_2F_4$, 5% i-$C_4H_{10}$ and 5% $SF_6$. A high and uniform electric field is present in their volume, thus the ionisation produced by a particle crossing the device starts an avalanche which generates a

detectable signal. The MRPC are characterised by an intrinsic time resolution of about 40 ps. The TOF complements the particle-identification capabilities of the TPC in the intermediate momentum range, i.e., below 3 GeV/$c$ for pions and kaons, and up to 5 GeV/$c$ for protons [162]. More details are provided in Section 2.6.

### 2.2.4 V0 detector

The V0 detector is composed of two arrays of scintillator counters denominated V0A and V0C, which are segmented into 4 rings in the radial direction and 8 sectors in the azimuthal one. The V0A is positioned at 340 cm distance from the nominal interaction point along the beam axis, on the opposite side to the muon spectrometer, whereas the V0C is located in front of the hadronic absorber at 90 cm from the interaction point. They cover the full azimuth and the pseudorapidity ranges $2.8 < \eta < 5.1$ (V0A) and $-3.7 < \eta < -1.7$ (V0C). The V0 provides a Minimum Bias trigger (MB) via a logic AND between the V0A and V0C signals. Its timing information is used to reject background events arising from beam-gas interactions and the pileup in the SPD readout time. Furthermore, the V0 signals are employed to estimate the collision centrality and the reaction plane in Pb–Pb collisions.

### 2.2.5 T0 detector

The T0 detector consists of two modules, T0A and T0C, instrumented with Cherenkov counters. They are placed on opposite sides of the interaction point, and cover the pseudorapidity intervals $4.61 < \eta < 4.92$ (T0A) and $-3.28 < \eta < -2.97$ (T0C). The T0 main purpose is the determination of a start time for the TOF detector, which complements the one computed by using the TOF itself, as will be discussed in Section 2.6.2. It has a time resolution, estimated by the time difference measured in the T0A and T0C, of about 40 ps in pp collisions and 25 ps in Pb–Pb collisions. The T0 is the principal detector for the online luminosity monitoring. It also provides redundancy to the V0 counters with the possibility of generating MB and multiplicity triggers.

### 2.2.6 Zero Degree Calorimeter

The Zero Degree Calorimeter (ZDC) detector consists of six quartz-fibre sampling calorimeters: two hadronic calorimeters for the detection of protons (ZP), two for the detection of neutrons (ZN), and two electromagnetic calorimeters (ZEM). The ZP and ZN calorimeters are located on the opposite sides of the interaction point at a distance of about 112.5 m. The ZN calorimeters are placed between the two beam vacuum tubes, while the ZP ones are positioned externally to the vacuum tubes on the side where positive particles are deflected by the LHC magnets. The

electromagnetic calorimeters are located on either side of the vacuum tube at about 7 m distance from the interaction point, in the opposite direction with respect to the muon arm. They cover the pseudorapidity range $4.8 < \eta < 5.7$.

The ZDC is employed to determine the event centrality in p–Pb and Pb–Pb collisions by measuring the energy deposited by spectator nucleons, which do not interact in the collision and emerge at zero degrees relative to the beam direction. The ZEMs measure the energy deposited by photons and $\pi^0$ decays at forward rapidity, to solve the ambiguities in the centrality determination due to the nuclear fragments which remain in the beam vacuum tubes and are not collected by the hadronic calorimeters [163]. Time measurements with the ZDC allow the rejection of background events produced by collisions of main bunches with satellite bunches during the Pb–Pb data tacking. The ZDC can also provide an estimate of the reaction plane in nuclear collisions.

## 2.3 The ALICE offline framework

The ALICE offline framework composed of AliRoot [164] and AliPhysics [165] is based on ROOT [166, 167], a scientific software toolkit for data analysis developed at CERN. ROOT is an object-oriented framework written in C++ designed for big data processing, which provides I/O handling functionalities, visualisation and statistical analysis tools. It is also integrated with other languages such as Python and R. AliRoot provides a set of software classes and macros for the detector alignment and calibration, and for the reconstruction and visualisation of the data collected by the experiment, as well as to steer Monte Carlo (MC) simulations. To perform full simulations, AliRoot is interfaced with event-generator codes, such as PYTHIA 6 [82] and PYTHIA 8 [168] for pp collisions, and HIJING [169] for heavy-ion collisions. The particle interactions with the detectors and support structures can be simulated with different transport codes, namely GEANT3 [170], GEANT4 [171], and FLUKA [172]. AliRoot includes a detailed description of the detector geometry, material budget, and response, which is implemented in independent modules dedicated to the different detectors. The code devoted to the physics analyses of reconstructed data from the experiment or from simulations is contained in the AliPhysics repository. This code consists of multiple tasks with a predefined structure built on top of ROOT and AliRoot, and it is progressively updated by the data-analysis groups.

The size of raw data collected by the ALICE experiment in the last years sums up to more than 160 PB. Therefore, its processing and analysis necessitate of a large amount of computing and storage resources. They are provided by the Worldwide LHC Computing Grid (WLCG) [173, 174], which will also be referred to as the *Grid* in the following. The WLCG is a global infrastructure coordinated by CERN, and it is the largest computing grid in the world with its 170 computing centres

spread in 42 different countries. The Grid is hierarchically divided in three *Tiers* granting different levels of data processing and storage. The Tier-0 centres keep one replica and perform the first reconstruction of the raw data, they are the CERN computing centre and the Wigner Research Centre for Physics in Budapest. The thirteen Tier-1 centres store a second replica of the raw and reconstructed data, and are involved in the data reconstruction and reprocessing. Smaller Tier-2 centres run the MC simulations and the user data analyses. The access to the collected and simulated data from anywhere in the Grid is ensured by the AliEn (ALICE Environment) [175, 176] middleware software. It provides a dedicated interface for data access, and for the execution of reconstruction, simulation, and analysis tasks on the Grid, together with the monitoring of their status.

The physical information for the reconstruction of the collisions from real data and MC simulations is stored in the Event Summary Data (ESD) files. They are mainly used for detector calibration, alignment, and performance studies. The information relevant for physics analyses is stored in a more compact data format, i.e., the Analysis Object Data (AOD), which is the one typically processed by the analysis tasks. This latter format has a size smaller by a factor of about six than the ESD one. The average AOD size is about 400 kB/event and 2.2 MB/event for pp and Pb–Pb collisions, respectively.

## 2.4 Centrality determination in Pb–Pb collisions

Since heavy ions are extended objects, their collisions are usually classified according to the collision centrality, which is proportional to the volume of the overlapping region and is directly related to the collision impact parameter $b$ introduced in Section 1.3.2. The centrality is normally expressed as a percentage $c$ of the total hadronic interaction cross section $\sigma_{\mathrm{AA}}$,

$$c(b) = \frac{\int_0^b \frac{\mathrm{d}\sigma}{\mathrm{d}b'}\mathrm{d}b'}{\int_0^{+\infty} \frac{\mathrm{d}\sigma}{\mathrm{d}b'}\mathrm{d}b'} = \frac{1}{\sigma_{\mathrm{AA}}} \int_0^b \frac{\mathrm{d}\sigma}{\mathrm{d}b'}\mathrm{d}b'. \tag{2.1}$$

The impact parameter of a collision cannot be accessed experimentally, thus the collision centrality is obtained by exploiting the relation between the centrality, the number of particles produced in the collision, and the number of spectator nucleons. The ALICE experiment measures two main observables related to the collision geometry, i.e., the energy deposited by spectator nucleons in the ZDC detector, and the amplitude of the signals produced by charged particles in the V0 detector, which is proportional to the particle multiplicity $N_{\mathrm{ch}}$ [22, 163]. The latter approach is the one employed in the analyses presented in this Thesis and will be discussed in the following.

Assuming that the particle multiplicity decreases monotonically with increasing impact parameter, the centrality can be experimentally computed as

$$c \simeq \frac{1}{\sigma_{\mathrm{AA}}} \int_{N_{\mathrm{ch}}}^{+\infty} \frac{\mathrm{d}\sigma}{\mathrm{d}N'_{\mathrm{ch}}} \mathrm{d}N'_{\mathrm{ch}} = \frac{1}{n_{\mathrm{ev}}^{\mathrm{tot}}} \int_{N_{\mathrm{ch}}}^{+\infty} \frac{\mathrm{d}n_{\mathrm{ev}}}{\mathrm{d}N'_{\mathrm{ch}}} \mathrm{d}N'_{\mathrm{ch}}. \tag{2.2}$$

In the above equation, the differential cross section $\mathrm{d}\sigma$ and the total hadronic cross section $\sigma_{\mathrm{AA}}$ can be replaced, respectively, by the number of observed events $\mathrm{d}n_{\mathrm{ev}}$ within a given multiplicity interval $\mathrm{d}N'_{\mathrm{ch}}$, and by the total number of events $n_{\mathrm{ev}}^{\mathrm{tot}}$ corrected for the trigger efficiency. The percentile of the hadronic cross section is determined for a given value of the V0 amplitude, which is proportional to the particle multiplicity, by integrating the distribution of the measured V0 amplitudes normalised at an *anchor point*. The anchor point is defined as the V0 amplitude corresponding to 90% of the hadronic cross section, and it determines the absolute scale of the centrality. Events with a measured multiplicity lower than the anchor point are strongly contaminated by background electromagnetic collisions and are not considered in the physics analyses.

The anchor point is determined by fitting the distribution of V0 amplitudes measured in Pb–Pb collisions with a parametrisation based on a Glauber Monte Carlo simulation. For each simulated event, the number of participant nucleons $N_{\mathrm{part}}$ and the number of binary collisions among nucleons $N_{\mathrm{coll}}$ is obtained from Glauber-model calculations [21]. The multiplicity produced in a single nucleon–nucleon interaction is generated by a negative binomial distribution (NBD)

$$P_{\mu,k}(n) = \frac{\Gamma(n+k)}{\Gamma(n+1)\Gamma(k)} \times \frac{(\mu/k)^n}{(\mu/k+1)^{n+k}}, \tag{2.3}$$

which provides the probability of measuring $n$ particles for a single emitting source, called *ancestor*. In Eq. (2.3), $\mu$ is the mean multiplicity per ancestor and $k$ is a parameter controlling the width of the distribution. The number of ancestors is parametrised as $fN_{\mathrm{part}} + (1-f)N_{\mathrm{coll}}$. This linear combination is inspired by two-component models [177, 178] which decompose the interactions occurring in a nucleus–nucleus collision into soft processes, producing a particle multiplicity proportional to $N_{\mathrm{part}}$, and hard processes which are characterised by a probability proportional to $N_{\mathrm{coll}}$. The average simulated V0 amplitude for the event, which is related to the number of particles hitting the detector, is computed by sampling the NBD a number of times equal to the number of ancestors. The distribution of the V0 amplitudes is simulated with an ensemble of events, and the $\mu$, $k$, and $f$ parameters are obtained from a minimisation procedure.

In Fig 2.5, the distribution of the V0 amplitudes measured in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV is shown, together with the fitted NBD-Glauber parametrisation. The bands report the limits of the centrality classes which are defined in terms of percentiles of the integral of the distribution up to the anchor point.

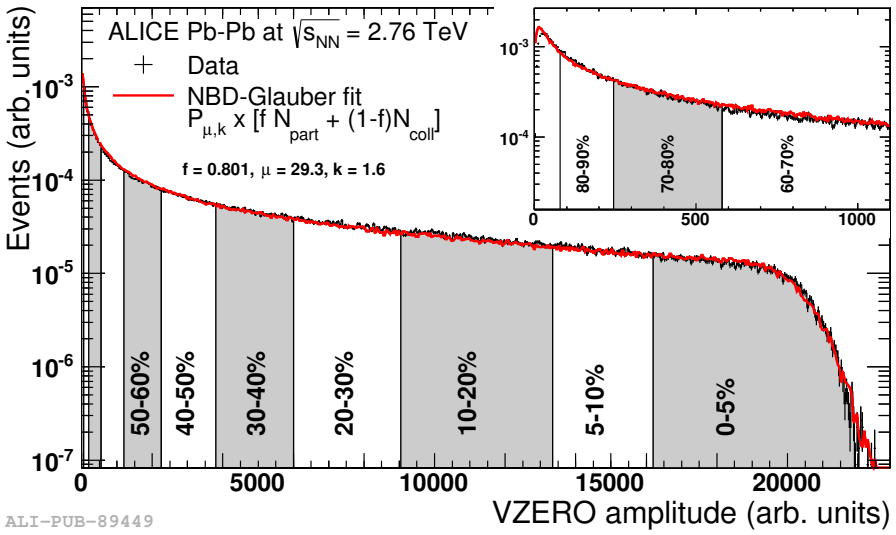



Figure 2.5: Distribution of the V0 amplitudes measured in Pb–Pb collisions at $\sqrt{s_{\rm NN}} = 2.76$ TeV. The red line reports the NBD-Glauber fit, while the bands show the different centrality classes. Figure from Ref. [163].

## 2.5 Track and vertex reconstruction

The track reconstruction and vertex finding are performed offline using the information acquired by the central barrel detectors [159]. First, for each detector, the collected raw data are converted into *clusters*, which correspond to groups of hits produced by a single particle interacting with a detector element. The clusters are characterised by various quantities such as their positions and signal amplitudes, and by the related uncertainties. Subsequently, a first estimation of the interaction-vertex position is obtained by using the clusters in the two innermost layers of the ITS detector, i.e., the SPD ones. Track segments, called *tracklets*, are constructed by associating pairs of clusters, one for each SPD layer, falling in a predefined azimuthal acceptance window. The primary vertex is obtained as the point in space where the largest number of tracklets converge. If a point of convergence is not found, condition occurring more frequently in low-multiplicity pp collisions, the interaction-vertex position is determined as the maximum along the $z$ axis of the distribution of the points of closest approach of the tracklets to the measured beam-axis position. This latter method is also adopted in Pb–Pb collisions for computational time reasons.

The finding and fitting of the tracks is carried out by using the Kalman filter algorithm [179] and a three-step procedure based on an inward-outward-inward propagation scheme. The first step starts with the track finding in the TPC. Track seeds are built from the clusters of the two outermost TPC pad rows and the primary-vertex position, and then using three clusters without any constraint to

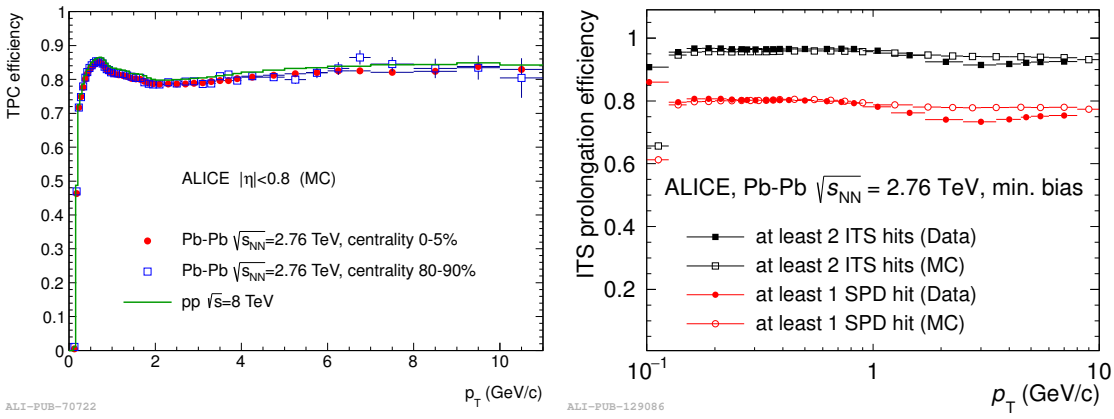



Figure 2.6: Left panel: TPC track-reconstruction efficiency from MC simulations in pp collisions at $\sqrt{s} = 8$ TeV and in Pb–Pb collisions at $\sqrt{s_{\rm NN}} = 2.76$ TeV. Right panel: TPC-to-ITS prolongation efficiency as a function of $p_{\rm T}$ for data (filled markers) and MC simulations (open markers) in Pb–Pb collisions at $\sqrt{s_{\rm NN}} = 2.76$ TeV. Figures from Ref. [159].

the vertex. The track seeds are then propagated inward by attaching clusters along the trajectory satisfying specific proximity criteria and by updating the track. The track propagation continues until the TPC inner radius is reached, and only tracks with at least 20 out of a maximum of 159 possible clusters are kept. A dedicated algorithm is also employed to reject multiple reconstructions of the same physical track partially sharing the same TPC clusters. In the left panel of Fig. 2.6, the track-reconstruction efficiency in the TPC as a function of $p_{\rm T}$ is reported in pp collisions at $\sqrt{s} = 8$ TeV and in Pb–Pb collisions at $\sqrt{s_{\rm NN}} = 2.76$ TeV for the 0–5% and 80–90% centrality classes, showing almost no dependence on the detector occupancy. The efficiency drops steeply to zero for $p_{\rm T} < 0.5$ GeV/$c$ due to the track deviation from the helicoidal path of a charged particle in a magnetic field, which is induced by multiple scattering and energy loss in the detector material. The efficiency shape observed at higher $p_{\rm T}$ is caused by the loss of clusters in the dead zones between the TPC readout sectors.

The tracks reconstructed in the TPC are propagated to the outermost layer of the ITS detector, and are used as seeds for the ITS track finding, which follows a procedure similar to the one described above. Tracks are extended toward the interaction point by attaching at each ITS layer all the clusters within a proximity cut, until the innermost layer is reached. This procedure is repeated using or not the primary-vertex position. Multiple track hypotheses in the ITS are then obtained for each TPC track. Finally, only the highest-quality track candidate is kept according to its $\chi^2$ value and cluster-sharing considerations. The prolongation efficiency of tracks from the TPC to the ITS is reported in the right panel of Fig. 2.6 for Pb–Pb collisions at $\sqrt{s_{\rm NN}} = 2.76$ TeV compared to MC simulations.

In the Figure, the results obtained with different requirements on the ITS hits are displayed. The prolongation efficiency shows a mild dependence from the track $p_{\mathrm{T}}$, and the values resulting from data and MC simulations are in good agreement. The smaller efficiency observed when requiring two hits in the ITS detector of which at least one in the SPD layers is due to the presence of inactive modules in the tracking layers. The clusters not associated to successfully-prolonged TPC–ITS tracks are employed to perform a standalone reconstruction with the ITS. In this way, particles decaying before reaching the TPC or traversing the TPC dead zones can be reconstructed. The standalone tracking enables the reconstruction of particles down to a $p_{\mathrm{T}}$ of about 80 MeV/$c$ for pions, thus extending the low transverse-momentum reach provided by the TPC.

Before the start of the tracking-procedure second step, tracks are extrapolated to their point of closest approach to the primary vertex. Then, they are refitted in the outward direction using the ITS and TPC clusters previously found. When the tracks reach the radii of the TRD and TOF detectors, an attempt is pursued to match them to the respective detector clusters. A further matching with the signals of the HMPID and calorimeter detectors is also performed. The propagation to the detectors placed at a radius larger than the TPC one is not used to update the measured track parameters, instead it is needed for particle-identification purposes. In the third and final step, all tracks are propagated from the TPC outer radius toward the interaction point, a last refit is performed, and the final track parameters and covariance matrix are computed.

If two or more tracks are reconstructed with the procedure detailed above, the position of the primary vertex is recomputed using the TPC–ITS tracks, which are also referred to as *global tracks*. Similarly to the method adopted using the SPD tracklets, global tracks are extrapolated to the point of closest approach to the measured beam-line position, and then a vertex finding and fitting is performed. Tracks are then propagated to the reconstructed vertex position. In Fig. 2.7, the transverse width of the distribution of primary-vertex positions reconstructed from global tracks is compared to the one from SPD tracklets as a function of the charged-particle multiplicity per unit of pseudorapidity in pp collisions at $\sqrt{s} = 7$ TeV. Both quantities scale with the inverse of the square root of the particle multiplicity, and an improvement of more than a factor two is observed for the global-track method. In particular, the value achieved by the more precise approach varies between about 500 and 50 μm in pp collisions, improving with increasing multiplicity, and it is better than 5 μm in central Pb–Pb collisions.

The left panel of Fig. 2.8 reports the resolution on the track impact parameter in the transverse plane, defined as the distance of closest approach to the primary vertex in the plane orthogonal to the beam direction, for charged particles reconstructed with global tracks in pp, p–Pb, and Pb–Pb collisions. The contribution due to the primary-vertex resolution is included. The impact-parameter resolution for tracks with $p_{\mathrm{T}} = 1$ GeV/$c$ varies between about 60 and 75 μm improving

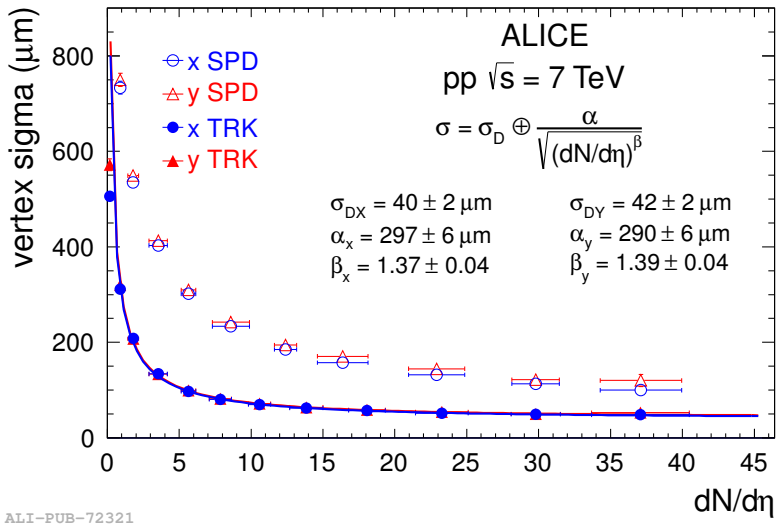



Figure 2.7: Transverse width of the primary-vertex position distribution as a function of the charged-particle multiplicity per unit of pseudorapidity in pp collisions at $\sqrt{s} = 7$ TeV. The values obtained using SPD and global tracks (TRK) for the primary-vertex determination are shown. Figure from Ref. [159].

for heavier collision systems. The resolution improvement is due to the more precise determination of the primary-vertex position induced by the larger particle multiplicities produced in larger collisions systems. The resolution on $1/p_{\mathrm{T}}$[1] for TPC-standalone and global tracks in p–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV is shown in the right panel of Fig 2.8. The effect due to the constraint of the tracks to the primary vertex is also displayed. The vertex constraint does not provide a relevant improvement when global tracks are considered. Moreover, the ITS hits used in the global tracks improve the $1/p_{\mathrm{T}}$ resolution similarly as the vertex constraint does for the TPC-standalone tracks. The resolution on $1/p_{\mathrm{T}}$ is obtained directly from the covariance matrix resulting from the Kalman filter fit, and it is related to the resolution on the transverse momentum $\sigma_{p_{\mathrm{T}}}$ as $\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} = p_{\mathrm{T}} \times \sigma_{1/p_{\mathrm{T}}}$. The $p_{\mathrm{T}}$ resolution for global tracks is better than 1% at $p_{\mathrm{T}} = 1$ GeV/$c$ and degrades to about 2% at $p_{\mathrm{T}} = 10$ GeV/$c$.

### 2.5.1 Secondary-vertex reconstruction

The reconstruction of secondary vertices originating from heavy-flavour hadron decays is performed at the analysis stage. Tracks are approximated as straight lines

[1]The transverse momentum $p_{\mathrm{T}}$ of a track is determined from the measurement of the track curvature in the magnetic field present in the central barrel.

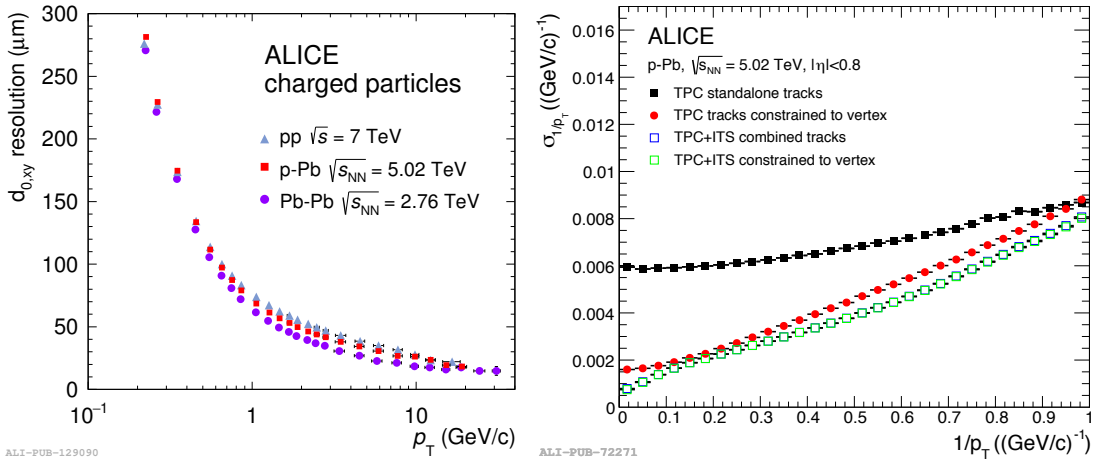



Figure 2.8: Left panel: resolution on the impact parameter in the transverse plane for global tracks corresponding to charged particles in pp, p–Pb, and Pb–Pb collisions. The contribution from the vertex resolution is included. Right panel: resolution on $1/p_{\mathrm{T}}$ as a function of $1/p_{\mathrm{T}}$ for TPC-only and global tracks with and without vertex constraint in p–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. Figures from Ref. [159].

in the proximity of the primary vertex, and multiplets of tracks are explored. The secondary-vertex position is defined as the space point minimising the distance $D$ among the considered tracks, which is calculated as

$$D = \sqrt{\sum_{i=1}^{N} \left[ \left( \frac{x_i - x_0}{\sigma_{x_i}} \right)^2 + \left( \frac{y_i - y_0}{\sigma_{y_i}} \right)^2 + \left( \frac{z_i - z_0}{\sigma_{z_i}} \right) \right]^2}, \tag{2.4}$$

where $N$ is the number of tracks, $(x_i, y_i, z_i)$ are the $i^{\mathrm{th}}$-track coordinates at the point of the closest approach to the vertex position $(x_0, y_0, z_0)$, and $(\sigma_{x_i}, \sigma_{y_i}, \sigma_{z_i})$ are the uncertainties from the covariance matrix of the $i^{\mathrm{th}}$ track. The momentum associated to the secondary vertex is calculated as the sum of track momenta evaluated at the point along each track closest to the secondary-vertex position. The invariant mass related to the secondary vertex is computed using the momentum and mass hypothesis of the decay tracks. The resolution on the position of secondary vertices originating from $\mathrm{D_s^+}$-meson decays is of about 100 μm with a slight dependence on the transverse momentum.

## 2.6 Particle identification

The ALICE experiment is characterised by excellent particle-identification capabilities in a wide momentum range, which result from the different and complementary techniques used by the experiment detectors. Charged hadrons are

identified by exploiting measurements of their specific energy loss in the ITS and TPC detectors, via particle time-of-flight measurements with the TOF detector, and through angle measurements of the Cherenkov light produced traversing the HMPID detector material. Electrons and positrons are identified via transition-radiation measurements with the TRD detector, and by comparing the energy deposited in the PHOS and EMCal calorimeters with their momentum. These latter two detectors can also reconstruct photons. The analyses presented in this Thesis employ the PID information provided by the TPC and TOF detectors, whose particle-identification techniques and performances are described in more details in the following Sections.

### 2.6.1 TPC particle identification

The particle identification with the TPC is based on the measurement of the specific energy loss $\mathrm{d}E/\mathrm{d}x$, the charge, and the momentum of the particles traversing the detector gas. The $\mathrm{d}E/\mathrm{d}x$ of a track is calculated as a truncated mean of the 65% lowest-amplitude measurements of the charge deposited on up to 159 pad rows. This choice rejects secondary-ionisation contributions. Then, the PID is derived from the comparison of the measured $\mathrm{d}E/\mathrm{d}x$ and the expected one for a defined particle species with a certain momentum. The expected $\mathrm{d}E/\mathrm{d}x$ is obtained from a data-driven parametrisation based on splines or on the function originally adopted by the ALEPH collaboration [180] deriving from the Bethe-Bloch formula

$$f(\beta\gamma) = \frac{P_1}{\beta^{P_4}} \left[ P_2 - \beta^{P_4} - \ln(P_3 + (\beta\gamma)^{-P_5}) \right], \tag{2.5}$$

where $\beta$ is the particle velocity, $\gamma$ is the Lorenz-boost factor, and $P_i$ are the fit parameters.

The left panel of Fig. 2.9 shows the specific energy loss of various particle species measured in the TPC as a function of their momentum in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. The expected mean energy losses of electrons, pions, kaons, protons, deuterons, tritons, and $^3$He nuclei are superimposed to the data. The relative $\mathrm{d}E/\mathrm{d}x$ resolution of the TPC is of about 5% in pp collisions and degrades to 6% in high-multiplicity heavy-ion collisions [181]. The TPC provides good pion-kaon and kaon-proton separations for transverse momenta below about 0.8 GeV/$c$ and 1.6 GeV/$c$, respectively. Nevertheless, the relative abundances of particle species can still be extracted in the relativistic-rise region, $2 < p_{\mathrm{T}} < 20$ GeV/$c$, utilizing a statistical approach.

### 2.6.2 TOF particle identification

The TOF identifies the particle species measuring their time of flight from the interaction point to the detector chambers. The start time for the TOF measurement is the time of the primary-collision event, which is estimated by using the

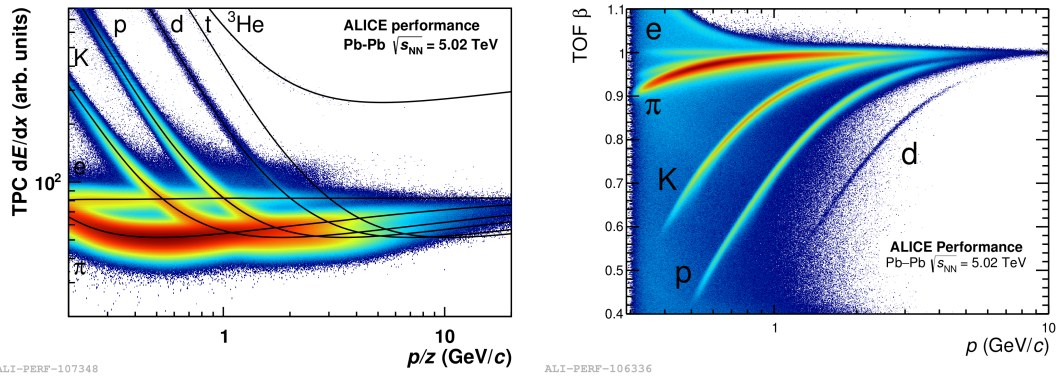



Figure 2.9: Left panel: specific energy loss $\mathrm{d}E/\mathrm{d}x$ of different particle species in the TPC as a function of their momentum in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. The parametrisation of the expected mean energy loss of the different species is shown by the continuous lines. Right panel: particle velocity $\beta$ measured by the TOF detector as a function of momentum in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. Figures from the ALICE figure repository ©.

signal from the T0 detector or the particle arrival times at the TOF detector. In the latter case, a combinatorial minimisation algorithm accounting for all the possible mass hypotheses is used. If both methods are available, the start time is the average of the T0 and TOF information weighted by their respective resolution.

The distribution of particle velocities $\beta$ measured by the TOF as a function of momentum, which is provided by the TPC, is displayed in the right panel of Fig. 2.9 for Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. The background is caused by tracks which are incorrectly matched to TOF hits. The electron, pion, kaon, proton, and deuteron bands are clearly separated. The overall TOF resolution is of about 56 ps for pions with a momentum of 1.5 GeV/$c$ in Pb–Pb collisions [162]. This value includes the intrinsic detector resolution, the contribution from electronics and calibration, the uncertainty on the event start time, and the tracking and momentum resolution. The TOF detector provides a good pion-kaon separation for $p_{\mathrm{T}} < 3$ GeV/$c$ and a satisfactory kaon-proton discrimination up to $p_{\mathrm{T}} \simeq 5$ GeV/$c$. At $p_{\mathrm{T}} < 1$ GeV/$c$, the fraction of tracks matched to the TOF detector is small. Therefore, in the analyses, low-$p_{\mathrm{T}}$ tracks without the TOF particle-identification information are usually identified using only the TPC.

# Chapter 3

# Machine learning — introduction and tools

The term *machine learning* (ML) comprehends a multitude of algorithms that are able to perform a specific task without being explicitly programmed to do so. These algorithms have a broad range of applications such as e.g., natural language processing or the generation of synthetic images. In recent years, the subject has become very popular due to the diffusion of the so-called *deep learning* [182, 183]. This field is based on algorithms which weakly resemble the functioning of the human brain, i.e., the neural networks, and are able to provide superhuman performances in various domains [184, 185]. The deep-learning diffusion has been possible thanks to the large computational power provided by modern Graphic Processing Units [186, 187], which permit to perform the expensive calculations required by deep-learning algorithms in a reasonable amount of time.

Over the past twenty years, machine-learning techniques have been increasingly adopted in the high-energy physics community, mainly to solve classification and regression problems, but also for event reconstruction and detector simulations [188–190]. Notably, they were also used in the discovery of the Higgs boson [157]. These methods show significant performance improvements compared to traditionally-used approaches in various applications of experimental interest [188, 191, 192].

At the moment of writing this Thesis, only a handful of published ALICE measurements employed machine learning [71, 136, 193–203], and the usage is limited to classification tasks for signal selection. Nevertheless, such techniques are progressively becoming more common in the ALICE Collaboration. The application of machine-learning tools to the quality control of the data collected by the experiment and to the calibration of the TPC space-charge distortions [204] is being investigated for the next data taking periods.

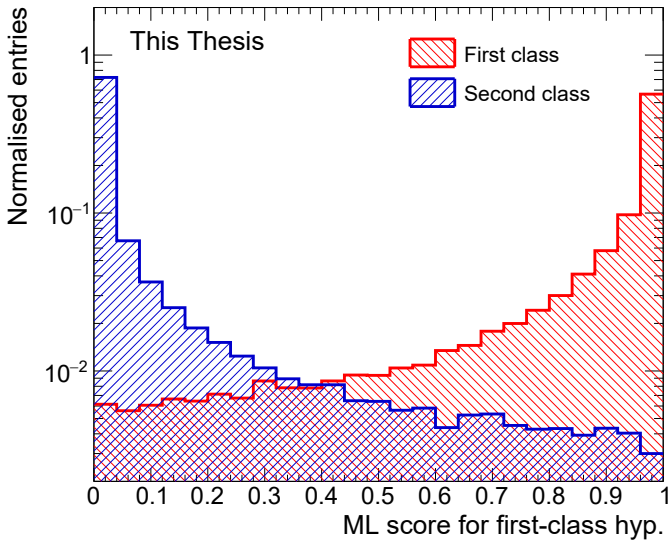



Figure 3.1: Example of the output of a supervised ML model for a binary classification problem. The score related to the hypothesis of belonging to the first class is reported for instances of the two classes, in blue and red respectively.

## 3.1 Supervised learning

*Supervised learning* is one of the main branches of machine-learning problems, together with unsupervised and reinforcement learning [182]. In supervised learning the desired output of the task to be accomplished is known and it is used to guide the ML algorithm "learning" process. Such problems are very common in particle physics, where e.g., interesting phenomena, the signal, must be separated from unimportant ones, the background. This is also the case for the work presented in this Thesis, where supervised machine-learning models are used to classify particle candidates into background and different kind of signals.

In the context of a classification problem, supervised algorithms learn how to label the data starting from a set of instances, where the belonging class is known. This is referred to as the *training set*. In a procedure called *training*, which is different for each algorithm, the internal parameters of the model are tuned using the training set. The objective of this operation is to obtain the parameter configuration that leads to the best discrimination of the training data. Ideally, the model learns general patterns related to the different class typologies and does not memorise the noisy peculiarities of each single instance of the training set. Therefore, after the training, the ML model can be employed to label unknown samples.

The output of a supervised ML model is a score evaluated from the instance properties, normally called *features*. This score is a numerical value related to the instance probability of belonging to the different classes. In Fig. 3.1, an example

of the output of a supervised model for a binary classification problem is shown. To define which instances are members of a class and which of the other, it is necessary to select a threshold value. Instances with a score below this threshold will be classified as belonging to the blue class and those above will be labelled as red class members, considering the output of Fig. 3.1. From the particle physics point of view, this delimiting value could be defined in relation to the analysis characteristics and requirements. Such as, e.g., the background rejection and the signal selection efficiency needed.

Before the application of the trained model on unknown data, it is important to assess its performance. This cannot be done using the training set, since the algorithm is more likely to classify correctly data points that has already "seen" during the training process. Therefore, the performance evaluation would be biased. A separate dataset, the *test set*, is needed to guarantee an unbiased performance estimation. As for the training set, also for the test set the correct label of the instances must be known.

Finally, machine-learning algorithms have also external parameters in addition to the internal ones. These parameters are not set during the training and must be chosen by the user according to the specific problem considered, and especially to the size of the available training set. They are called *hyperparameters* and are optimised repeating the model training with different combinations of their values and selecting the one that provides the best performance. The hyperparameter optimisation is discussed in more detail in Section 3.1.3. This procedure is not performed evaluating the model on the test set, because the model would adapt to give better results on it and the final performance estimation would be biased. Hence, a portion of the available data is used to build a *validation set*, which is then employed for the hyperparameter optimisation with the aim of maintaining unbiased the performance evaluation on the test set. An alternative approach is the so-called *cross-validation* method, which is described in Section 3.1.2.

### 3.1.1 Receiver Operating Characteristic curve

To compare the performances of different algorithms or to assess which set of hyperparameters gives the best result, it is necessary to define a numerical quantity related to the predictive capability of the machine-learning model. Many evaluation metrics are available to this end, each one with its pros and cons. The *area under the Receiver Operating Characteristic curve* (ROC AUC) is a metric well suited for binary classification problems [205] and can be extended to multi-class classification [206]. Moreover, the ROC AUC is independent of the relative abundances of class instances in the dataset used for its evaluation. This is an important feature when facing classification problems in particle physics, where usually the interesting signal in a real data sample is much less abundant than the uninteresting instances. In these cases, for the training and testing of the supervised model, it is handy and

can lead to better results to work with datasets enriched of true signals, which do not reproduce the real class unbalance.

For the aforementioned reasons, the ROC AUC is used as evaluation metric in the studies presented in this Thesis.

### Binary classification case

In the case of the discrimination between a positive class $P$ and a negative one $N$, it is possible to study the model performance on a dataset looking how instances of the two classes are classified. In the high-energy physics terminology the $P$ class would be the signal and the $N$ class would correspond to the background.

Considering a set of instances of the $P$ class, those labelled by the model as member of $P$ are defined true positive $TP$, while those incorrectly labelled as members of $N$ are called false negative $FN$. Whereas, samples of the $N$ class can be classified as true negative $TN$ or false positive $FP$. The discrimination capability of a machine-learning model can be summed up in a confusion matrix, which is shown in Tab. 3.1. From the number of samples falling in each element of the aforementioned matrix, the true positive rate $TPR$ and the false positive rate $FPR$ are computed. They correspond respectively to the signal and background selection efficiency in the particle-physics analogy and are defined as

$$TPR = \frac{TP}{TP + FN}, \qquad FPR = \frac{FP}{FP + TN}. \tag{3.1}$$

The *Receiver Operating Characteristic* (ROC) curve, shown in the left panel of Fig. 3.2, is obtained plotting the $TPR$ as a function of the $FPR$ for various threshold values on the model output. Then, the area under the ROC curve can be computed and it is used as a metric of the supervised model performance. The possible values of the ROC AUC range between 0.5 and 1, where the former value corresponds to a random classification and the latter to a perfect discrimination of the two class samples. The ROC AUC provides a global estimation of the model performance, i.e., not related to the final threshold value that will be chosen. Finally, it is possible to notice from Eq. (3.1) that the ROC AUC is independent of

Table 3.1: Confusion matrix

| | | **Real Class** | |
|---|---|---|---|
| | | ***P*** | ***N*** |
| **Predicted Class** | ***P*** | $TP$ | $FP$ |
| | ***N*** | $FN$ | $TN$ |

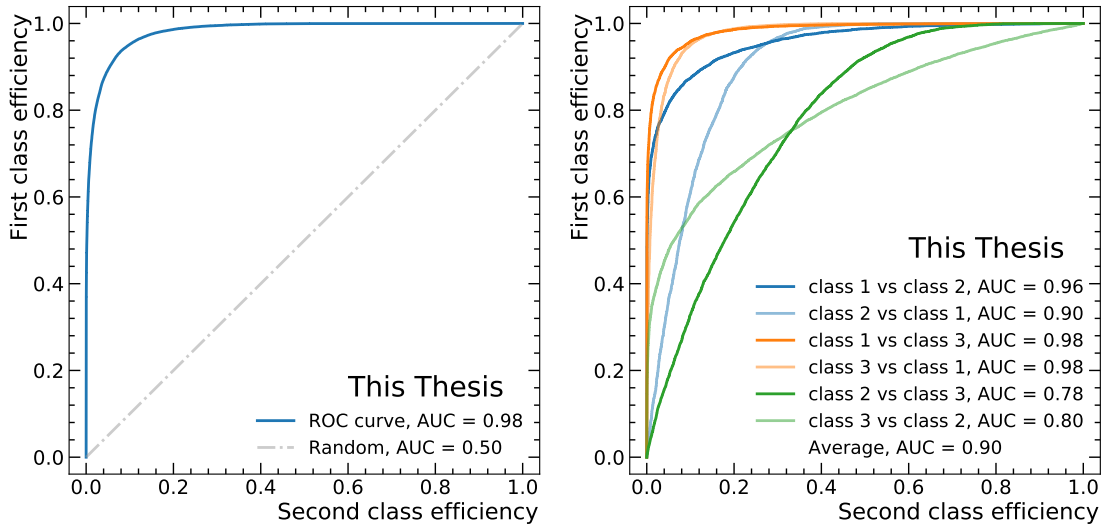



Figure 3.2: Left panel: example of ROC curve for a binary classification problem. The curve obtained with a trained model is reported in blue. The result expected from a random classification is represented by the dashed grey line. Right panel: example of ROC curves for a three-class classification problem using the one-vs-one approach. Curves of different nuances of the same colour are obtained swapping the same two classes.

the relative number of positive and negative instances present in the sample used for the performance evaluation.

### Extension to multi-class classification

The area under the ROC curve cannot be natively defined for a multi-class classification problem, i.e., in the case in which it is needed to perform a discrimination between more than two classes. However, an extension to an arbitrary number of classes has been proposed in Ref. [206] and has been implemented in the Python library for machine learning `scikit-learn` [207, 208].

In a classification task with $N$ classes, the outputs of a supervised model are $N$ scores related to the instance probability to belong to each one of the different classes. A simple way to estimate the model discrimination power using the ROC-curve method is the *one-vs-one* approach. In this approach, all the possible $N(N-1)$ pairwise ROC AUCs are computed. Considering two classes $i$ and $j$, the area under the ROC curve $AUC(i,j)$ is calculated adopting $i$ as the signal class and $j$ as the background class and using the score related to the class $i$. In general $AUC(i,j) \neq AUC(j,i)$, i.e., swapping $i$ and $j$ produces different results, since in the latter case the score related to the class $j$ is used. This can be observed in the example reported in the right panel of Fig. 3.2.

A single value that represents the supervised-model discrimination power is obtained averaging the pairwise ROC AUCs as

$$\text{Average ROC AUC} = \frac{1}{N(N-1)} \sum_{i \neq j} AUC(i,j) \tag{3.2}$$

where the sum runs over all the possible combinations. The one-vs-one method preserves the independence of the average ROC AUC from the relative abundances of the class instances in the sample used for the estimation. Another possible way to tackle the same problem is to compute the ROC curve considering the $i$ class as signal and the remaining $N-1$ classes as background. This approach is called *one-vs-rest* and only $N$ ROC AUCs are needed. However, this method is not independent of the relative abundances of the classes. For this reason, the one-vs-one strategy is adopted for the model performance evaluation in Chapter 5.

### 3.1.2 Cross validation

As already introduced in Section 3.1, to optimise the hyperparameters and evaluate the performance of a supervised ML model, the available data needs to be split in a training, a test, and a validation set. Therefore, a considerable portion of the data sample is not used in the algorithm learning process. This could be a problem especially when the size of the available data sample is small. Since the performance of a supervised ML model is related to the number of instances used for its training and, in general, the larger is their number the better. To address this issue a procedure called cross validation [205] is employed. Thanks to this procedure it is possible to perform the hyperparameter optimisation of the model without using a separate validation set. Hence, a larger fraction of the dataset can be allocated for the training.

In the cross-validation technique, the training set is divided in $k$ parts referred to as *folds*, for this reason often the term *k-fold cross validation* is used. For a fixed combination of hyperparameters, the model training is performed on $k-1$ folds and the validation of the model performance is done on the remaining training-set portion. This operation is repeated $k$ times permuting the folds at every step, in order to evaluate the model predictions on each fold only one time. In this way $k$ values of the metric, which quantify the performance of the model, are obtained, and their mean value can be used as the final estimator. The dispersion of the $k$ values around the mean gives an estimation of the performance-evaluation error. An exemplification of this approach is shown in Fig. 3.3 in the case of $k = 5$.

The aforementioned procedure is repeated for all the hyperparameter configurations that need to be explored. When the configuration that yields the best result is found, the training is performed again utilising all the folds to obtain the final model. Another possibility is to build the final model as an ensemble of the $k$ models trained during the cross-validation procedure, i.e., combining their predictions

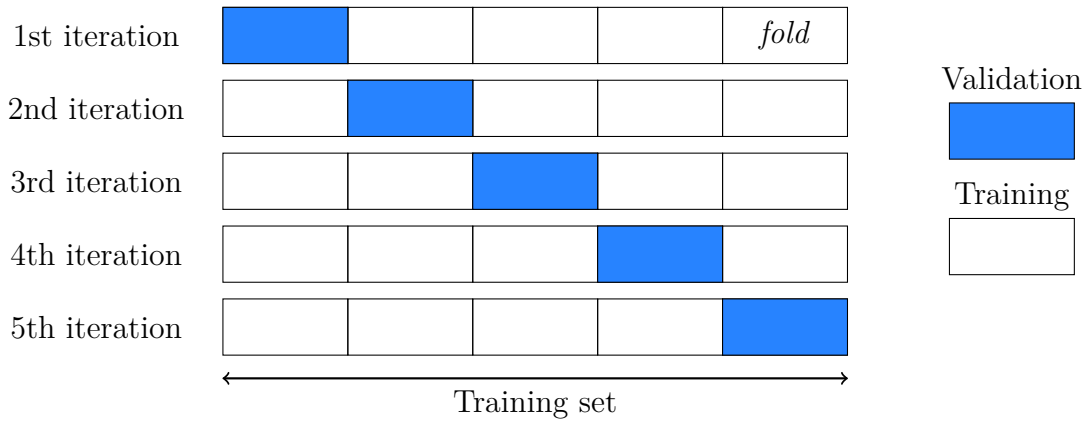



Figure 3.3: Schematisation of the 5-fold cross-validation procedure.

in order to obtain the final one.

Concerning how many folds should be employed, a number of folds equal to the number of instances in the training set would be the best choice of $k$. In this way, the algorithm trained in each step would be as close as possible to the final one. Thus, the estimation of the final model performance would be almost unbiased. However, a limit on the value of $k$ is imposed by the available computational power. The number of total trainings that need to be performed is given by the hyperparameter combinations tested times the number of folds and can easily become huge. Therefore, a trade-off between the accuracy of the cross-validation procedure and the time required for the computation must be accepted. Typical values of $k$ are five or ten [205].

### 3.1.3 Hyperparameter optimisation

A common way to perform the hyperparameter optimisation is the *grid search*, which is a simple brute-force search over the hyperparameter space. In this approach, a grid of points in the $n$-dimensional hyperparameter space is manually defined. The grid-search method performs poorly when the dimensionality of the space to be explored is high, i.e., when the number of hyperparameters to be tuned is large. Another simple solution for the hyperparameter optimisation is the *random search* [209], where the points to be probed are selected randomly in the hyperparameter space. It has been demonstrated that the random-search method can outperform the grid search when few of the hyperparameters considered are important for the machine-learning model performance. Both the aforementioned approaches are embarrassingly parallel and easily scale on multicore CPU.

A more refined method to explore the space of hyperparameters, which is employed in this Thesis, is the *Bayesian optimisation* [210, 211]. The Bayesian optimisation is a procedure to find the minimum of an unknown function $f(\boldsymbol{x})$ on a

bounded set $\mathcal{X}$. To this end a posterior distribution on functions is built to describe $f(\boldsymbol{x})$ starting from multivariate Gaussian distributions. This posterior distribution is used to determine the next point on $\mathcal{X}$ where to evaluate $f(\boldsymbol{x})$ and the result is used to update the posterior. As the number of observations grows, the posterior distribution improves and the optimisation algorithm gets better in exploring the more interesting regions of $\mathcal{X}$. The procedure is designed to minimise the number of steps required to find a point that is close to the minimum.

In the case of interest, the Bayesian optimisation is used to maximise the unknown function that describes the performance of the machine-learning model as a function of the hyperparameters. The evaluation of this unknown function is very expensive since it requires to run the full training of the machine-learning algorithm. This especially applies when using the cross-validation procedure introduced in the previous Section. Therefore, it is more efficient to spend computational time making better guesses about where to look for the best parameter combinations. Finally, Bayesian optimisation has been shown to obtain better results with fewer evaluations compared to grid search and random search [210], thanks to the possibility to account for the quality of the trials before they are performed.

## 3.2 Boosted Decision Trees

A family of popular machine-learning algorithms for classification and regression problems are the *Boosted Decision Trees* (BDT) [212, 213]. BDTs have been extensively used to analyse data in high-energy-physics experiments [188, 189], and have been adopted to perform the measurements presented in this Thesis. They are well suited when the inputs of the problem of interest are composed of high-level features, which results from a manual engineering of the most-useful quantities for the problem resolution based on field-specific knowledge. This is often the case in particle-physics studies, where the employed features are obtained from complex elaborations of the raw data collected by the experiment and usually have a physical meaning.

The building block of a BDT is the *Decision Tree* (DT), which is an algorithm that combines a sequence of simple tests to classify an instance depending on its features. Each test confronts a numeric attribute versus a threshold value or a categorical attribute against a set of possible values. An example of a simple DT is reported in Fig. 3.4 for a classification problem between two classes, $A$ and $B$, using numerical features. The features are labelled $f_i$ while the threshold values for each *node* of the tree are reported as $\mathrm{t}_\alpha$. The logical rules followed by this kind of algorithms are relatively easy to interpret. Thus, the decision tree has an advantage in terms of comprehensibility with respect to other machine-learning algorithms, such as the neural networks which are more similar to a "black-box".

The training of a DT consists in the automatic procedure by which the tree

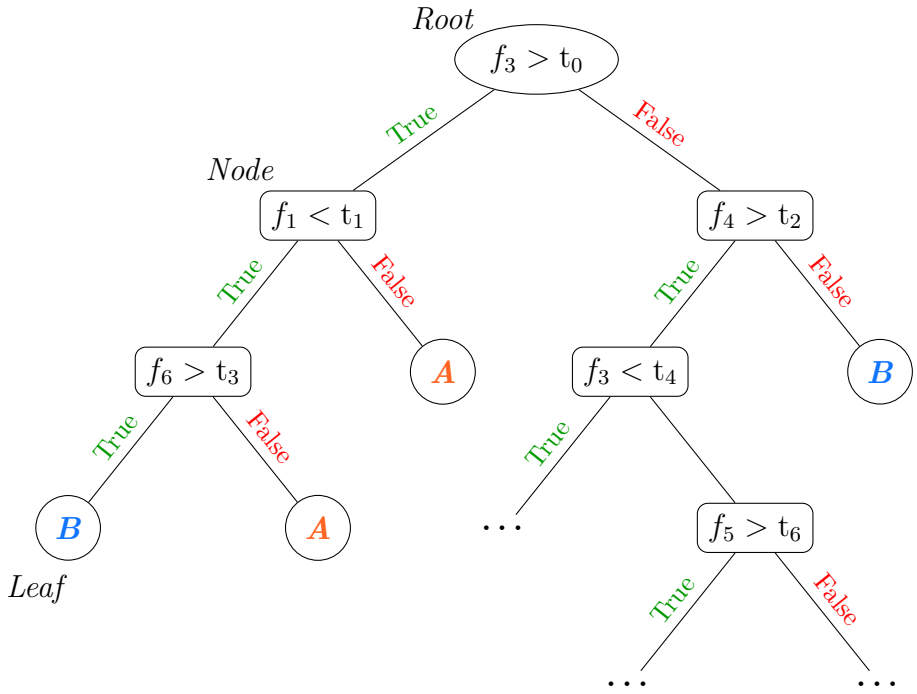



Figure 3.4: Structure of a simple Decision Tree for a binary classification problem with numerical features.

structure is created. The tree is built recursively utilising the training set and attempting to find patterns in the data. At each tree node, the feature and its value maximising the separation between the classes is selected, and the training data is partitioned accordingly. Various measures of the partition goodness can be used, such as, e.g., the Gini index or the cross entropy [205]. The process is repeated over the partitioned data until all instances in a node are of the same class or a stopping condition, imposed by the user, is reached. In this case the node is called a *leaf* of the tree.

It is possible to build large decision trees with many nodes that can classify well the training set. However, they generalize poorly, namely, their performance on unknown data is poor. For this reason small trees containing few nodes are used. They are characterized by a low discrimination power but their performance on the training and test set is similar. Thus, small trees are able to learn some general pattern from the data and not the data itself as the larger ones. With the purpose of having a model with a good classification power, numerous decision trees are combined, usually in the order of hundreds. This is done with a procedure called *boosting*, where the trees are constructed sequentially taking care of compensating the shortcomings of a tree with the others. A simple way to perform the boosting is the one implemented in the *AdaBoost* algorithm [214]. In this approach, during the

training of a DT, higher importance is given to the correct classification of samples that were misclassified by the preceding tree. When all the decision trees have been built, a prediction can be obtained combining their outputs. The ensemble of these decision trees forms the BDT model. The AdaBoost algorithm is a special case of the *gradient boosting* technique [215], an implementation of which will be presented in the next Section.

### 3.2.1 XGBoost

An implementation of the BDT algorithm is available in the TMVA package [216] contained in the ROOT framework [166, 167]. Even if TMVA would provide a seamless integration of the machine-learning tools with the ALICE experiment analysis software, which is also based on ROOT, a different approach using the Python ecosystem has been adopted in this Thesis. The reason for this choice is that Python is one of the main programming languages employed in the machine-learning community. Therefore, the amount of documentation and advanced techniques available is much larger than those provided by TMVA. A more in-depth discussion of the software packages used in this Thesis is presented in Section 3.3.

The BDT implementation provided by the XGBoost [217] library is employed in the analyses presented in this Thesis. This library is an open-source package and its algorithm achieved state-of-the-art results in many machine-learning competitions. Moreover, XGBoost is characterized by the possibility to parallelise the execution on CPUs and GPUs [218], which lead to a reduction of the training times.

The term XGBoost stands for *Extreme Gradient Boosting* and refers to the fact that this BDT is based on the gradient boosting method. The idea behind this method is to define an objective function and to run the machine-learning algorithm performing a gradient descent of this function, to create the model that minimises it. The objective function is composed of a training loss, which measures the difference between the predictions and the target values on the training set, and by a regularization term used to control the BDT complexity. The optimisation of the objective function is performed iteratively and at each iteration the tree that minimises the function for the current model configuration is built. In the building of the tree, at each node, the feature and its value that maximise the *gain*, i.e., the reduction of the objective function, is used to partition the data.

In the case of XGBoost, each tree assigns to an instance a score related to its input features. The output of the BDT is given by the sum of all the tree scores, as exemplified in Fig. 3.5 for a BDT composed of two trees.

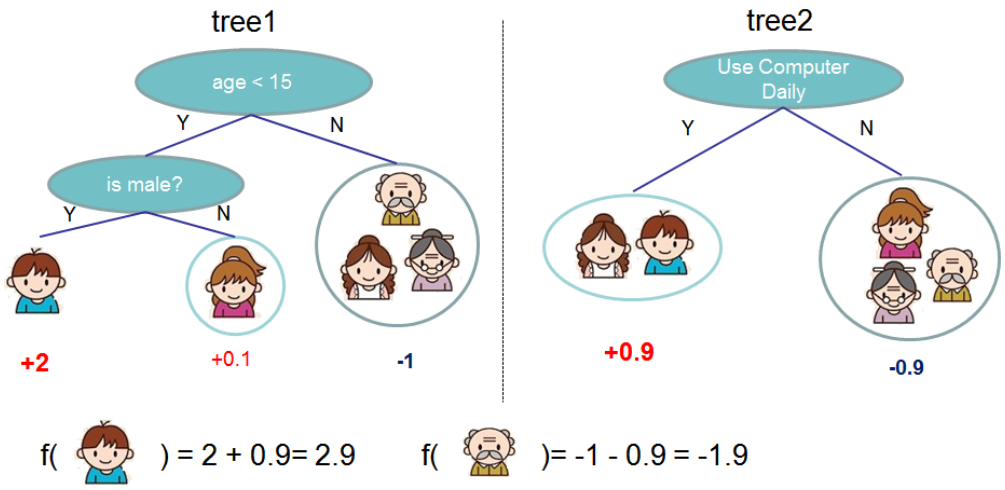



Figure 3.5: Evaluation of unknown samples for a simple XGBoost model composed of two trees. Figure from Ref. [217].

### XGBoost hyperparameters

The algorithm provided by the XGBoost library has many hyperparameters that can be tuned to control the BDT model complexity and improve the performance for the problem of interest [219]. They are also used to prevent the *overfitting* [205], a condition in which the supervised-model performance on the test set is much lower than the one on the training set. This condition occurs when the model "memorises" the training data and does not learn general patterns. The full list of available hyperparameters, together with their default values and a description of their meaning can be found in the XGBoost documentation [219].

The hyperparameters that have been considered during the BDT optimisation in this Thesis work are introduced in the following.

- **`n_estimators`**, the number of trees that compose the BDT.
- **`learning_rate`**, a weighting factor applied to the outputs given by new trees added to the model. Generally low learning-rate values lead to better model performance. However, the training procedure converges more slowly and more trees must be added to the ensemble, thus increasing the computation cost.
- **`max_depth`**, the maximum depth of the trees, i.e., the maximum number of edges from an extremal node, *leaf*, to the root node of the tree. Larger maximum-depth values lead to bigger and more complex models.
- **`min_child_weight`**, related to the minimum number of instances needed in a node in order to be split. This parameter controls the stopping condition

during the tree building.

- **subsample**, the fraction of training instances used to grow each tree. For the construction of each tree a different subsampling of the training set will be performed.
- **colsample_bytree**, the fraction of randomly-chosen instance features used in the training of each tree.

### 3.2.2 Feature importance

When developing a machine-learning model to perform a given task, it is useful to investigate how much each of the instance features influences the model predictions [220]. The objectives of this investigation, also referred to as measuring the *feature importance*, are three-fold: to improve the predictions selecting the variables more relevant to accomplish the task, to reduce the number of features used obtaining faster and simpler models, and to understand better the decision process of the model. The latter objective is of particular interest in particle-physics analyses, where usually the features are well understood quantities and have a clear physical meaning. Thus, the feature importance can be used to check that what is relevant for the model predictions matches the user expectations and insights on the problem.

The relatively simple structure of the BDTs gives the possibility to investigate their decision process straightforwardly. For example, the feature importance can be estimated by counting the number of times a particular variable is used to split a node, for all trees in the ensemble, thus obtaining the frequency with which each variable is employed [217]. The idea is that features utilised more frequently should be more useful in the decision process. It is also possible to consider the average of the gain values, which have been introduced in Section 3.2.1, of each feature. This information is more accurate than the frequency in estimating the importance of a variable because it also accounts for the contribution to the objective-function minimisation. Finally, another commonly used approach is to randomly permute the values of a feature in the test set and measure the worsening of the model performance [212]. If a feature is important, this will cause a significant decrease of the model prediction power. In fact, the permutation destroys the information associated with that feature.

An alternative way to study the predictions of a machine-learning model is provided by SHAP (SHapley Additive exPlanations) [221, 222]. In the SHAP approach, a prediction is explained assuming that each feature value of an instance is a player in a game where the model output is the payout. Then, Shapley values, a method from cooperative game theory, are used to fairly distribute the payout among the features based on their contribution in the game. It has been demonstrated that SHAP is more reliable than the aforementioned techniques. In addition to compute

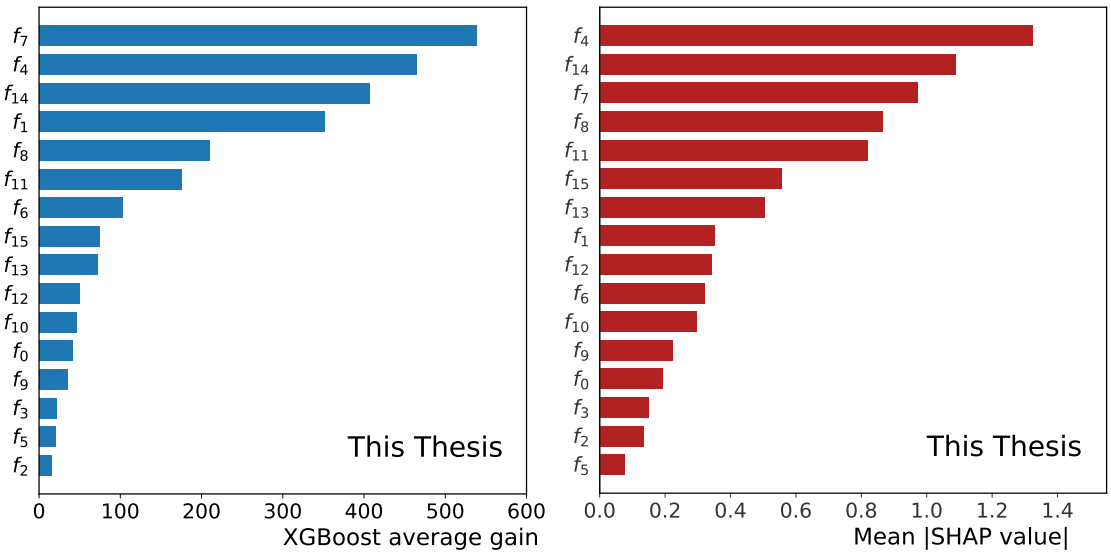



Figure 3.6: Examples of feature importance estimation for the same XGBoost trained model and set of features. Left panel: average XGBoost gain values of the features. Right panel: mean absolute SHAP values.

the global importance value of a feature, it offers more advanced tools to investigate the interactions between different features and to assign feature-importance values for each single prediction. In Fig. 3.6 the feature-importance measurements performed with the gain method and SHAP are reported for the same trained BDT model and the same set of features $\{f_i\}$. Although the results of the two estimation methods are different, the groups of relevant and unimportant features are similar. In particular, the top three $\{f_4, f_7, f_{14}\}$ and bottom five $\{f_0, f_2, f_3, f_5, f_9\}$ classified features are the same for both methods.

## 3.3 Software tools

As mentioned in Section 3.2.1, the training and performance assessment of the BDT models used in this Thesis are performed in Python. The complete and easy-to-use available software packages simplify the task of developing a good machine-learning model. However, there is the need to bridge the part of the analysis written in Python with the one executed with the standard ALICE tools introduced in Section 2.3, which are based on the C++ language and the ROOT framework. To this end, the ROOT bindings for Python [167] are employed, which permit using ROOT classes and functions in a Python script. The other important piece in the workflow integration is the `uproot` [223] package, which reads and converts the ROOT data to a format friendly to Python libraries. This package is used to stream data extracted in ROOT format from the ALICE AODs into the machine-learning

algorithms, which require a tabular data format.

The `hipe4ml` (Heavy-Ion Physics Environment for Machine Learning) [224] package has been developed to simplify the application of machine-learning techniques for analyses similar to those performed in this Thesis. The library is built on top of commonly used Python packages. It automatises the typical steps required by a supervised classification task applied to particle physics, such as the data preparation, the model training, and the hyperparameter optimisation.

The code used in the studies presented in this Thesis is available in the software repository of the ALICE Collaboration `AliPhysics`[1] and in the public repository `DmesonAnalysis`[2] [225]. Hereafter, the list of Python packages employed in this Thesis is reported.

- `bayesian-optimization` [211]
- `hipe4ml` [224]
- `matplotlib` [226]
- `numpy` [227]
- `pandas` [228, 229]
- `scikit-learn` [207]
- `scipy` [230]
- `shap` [221]
- `uproot` [223]
- `xgboost` [217]

### 3.3.1 Interface with the ALICE Grid

An important technical point for the analyses presented in this Thesis is how to perform the *inference* step after having obtained the trained models, i.e., how to apply them to the data that need to be classified.

For the $\mathrm{D_s^+}$-meson measurement in pp collisions presented in Chapter 5, the full data sample collected by the ALICE experiment has a size of few GBs, after a suitable selection and data reduction. Therefore, the inference step can be performed in a short time, generally few minutes, on a personal computer or a small workstation. When data from Pb–Pb collisions are studied instead, as for the $\mathrm{D_s^+}$-meson $R_{\mathrm{AA}}$ measurement reported in Chapter 6, the size of the data sample to analyse is of the order of few TBs. Thus, managing and applying the machine-learning models on the full data sample require a server-grade machine. Moreover, for more complex analyses requiring also other information about the collision properties, as the one carried out in Chapter 7 to measure the $\mathrm{D_s^+}$-meson elliptic flow, it is cumbersome to store and manage all the data locally.

For the aforementioned reasons, it is useful to be able to profit from the distributed computing resources provided by the Grid. To this end, the `treelite` [231] library is employed to perform the inference of a BDT, trained with Python software, in the ALICE framework based on C++. This library converts a BDT model

[1] https://github.com/alisw/AliPhysics/tree/master/PWGHF/vertexingHF

[2] https://github.com/DmesonAnalysers/DmesonAnalysis

stored in the XGBoost internal data format in a C++ function, which is then compiled at runtime and used to provide the BDT output given the input features. The `treelite` package has been integrated in the ALICE framework and a simple interface has been developed[3]. This approach permits to easily integrate the BDT inference in the analysis code of any ALICE measurement and to fully exploit the distributed computing power of the Grid.

To conclude, it is worth mentioning that the ROOT project is working to provide the possibility to perform the inference of XGBoost models in its framework [232]. However, this feature is still in an experimental state.

[3] https://github.com/alisw/AliPhysics/tree/master/ML

# Chapter 4

# $D_s^+$-meson reconstruction in pp and Pb–Pb collisions

In this Chapter, the data samples employed in this Thesis are described and the strategy for the reconstruction of $D_s^+$ mesons and their antiparticles at midrapidity ($|y| < 0.5$) is detailed. The $D_s^+$ mesons, composed of a charm quark and a strange antiquark, cannot be directly observed since their short mean proper decay length, $c\tau(D_s^+) = (151 \pm 1)$ µm [6], does not allow them to reach the detectors. Therefore, the $D_s^+$ mesons are measured by fully reconstructing their decay products, and the hadronic decay channel $D_s^+ \to \phi\pi^+ \to K^-K^+\pi^+$ with branching ratio BR = $(2.24 \pm 0.08)\%$ [6] is the one considered in this Thesis. The analysis is based on the reconstruction and selection of secondary decay vertices displaced from the primary interaction vertex, exploiting the good spatial resolution of the ALICE detectors. The particle-identification capabilities of the TPC and TOF detectors are also used. Then, supervised machine-learning models perform the $D_s^+$-meson signal selection and the background rejection. Finally, an invariant-mass analysis of candidates passing the selection is utilised to extract the raw-signal yield, which is then corrected for the reconstruction and selection efficiencies from simulations.

## 4.1 Data samples and event selection

The results reported in this Thesis were obtained analysing the data sample of pp collisions at $\sqrt{s} = 5.02$ TeV collected with the ALICE experiment in 2017 and that of Pb–Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV collected in 2018. Both data samples were recorded with a minimum bias (MB) trigger which required coincident signals in the V0A and V0C detectors. In Pb–Pb collisions, a further online event selection based on the signal amplitudes of the V0 detectors was exploited to enrich the sample of central (0–10%) and semicentral (30–50%) collisions. The strategy for the determination of the centrality in Pb–Pb collisions is detailed in Section 2.4.

Various offline event selections were applied. The background events due to the interaction between the beams and the residual gas in the beam vacuum tubes were rejected using the V0-detector timing information [159]. In pp collisions, the correlation between the number of clusters and tracklets in the SPD layers of the ITS detector was additionally used for the same scope. The background induced by collisions of main bunches with satellite bunches is relevant in the Pb–Pb data taking, and it was suppressed thanks to the time measurements of the ZDC detector.

Only events with a primary vertex reconstructed using global tracks, out of the two possible strategies introduced in Section 2.5, were considered. This choice provides the best resolution on the primary vertex position, which is crucial for the D-meson measurements based on the displaced decay-vertex topology, at the cost of a lower vertex-reconstruction efficiency for low-multiplicity pp events. However, low-multiplicity events without a reconstructed vertex with global tracks are very unlikely to contain D mesons produced at midrapidity and decaying in the hadronic channels of interest.

Events with multiple primary vertices due to pileup of several collisions are common in pp and Pb–Pb interactions. The large pileup occurring in the TPC readout time of about 90 µs is rejected at the analysis level. In pp collisions, the pileup contained in the readout time of the SPD detector, about 300 ns, is discarded by a vertexing algorithm which exploits the SPD information. The vertex-finding algorithm is run on the SPD tracklets not associated to the main vertex, which is the one with the largest multiplicity. Then, an event is considered pileup and removed if another vertex with at least five contributors is found in the region outside ±0.8 cm from the main vertex. The pileup probability in the SPD readout time is negligible in Pb–Pb collisions given the low interaction rate. To guarantee a uniform acceptance in the pseudorapidity region $|\eta| < 0.8$, the selected events were required to have a reconstructed collision vertex within ±10 cm from the detector centre along the beam-line direction. This requirement further rejects main-satellite collisions in Pb–Pb interactions.

In Tab. 4.1, the number of events $N_{\text{events}}$ and the corresponding integrated luminosity $\mathcal{L}_{\text{int}}$, after the selections described above, are reported for each data sample and centrality class. The visible cross sections for the MB trigger were measured with van der Meer scans and correspond to $\sigma_{\text{vis}}^{\text{pp}} = (50.9 \pm 0.9)$ mb [233] and $\sigma_{\text{vis}}^{\text{Pb–Pb}} = (7.67 \pm 0.24)$ b [234] for pp and Pb–Pb collisions, respectively.

## 4.2 $D_s^+$ decay-vertex reconstruction

The $D_s^+$ mesons and their charge conjugates are reconstructed through the hadronic decay channel $D_s^+ \to \phi\pi^+ \to K^-K^+\pi^+$ with BR = (2.24 ± 0.08)% [6]. The analysis strategy is based on the selection of fully reconstructed decay-vertex topologies displaced from the primary vertex. For prompt $D_s^+$ mesons, which are

Table 4.1: Data samples used for the measurements reported in this Thesis. The centre-of-mass energy, the trigger approach, the number of events after the selections, and the integrated luminosity are detailed.

| System | $\sqrt{s_{\rm NN}}$ (TeV) | Trigger | $N_{\rm events}$ | $\mathcal{L}_{\rm int}$ ($\mu b^{-1}$) | Year |
|---|---|---|---|---|---|
| pp | 5.02 | MB | $990 \times 10^6$ | $19500 \pm 400$ | 2017 |
| Pb–Pb | 5.02 | Central | $100 \times 10^6$ | $130 \pm 4$ | 2018 |
| | 5.02 | Semicentral | $85 \times 10^6$ | $55 \pm 2$ | 2018 |

produced in the charm-quark hadronisation or from decays of excited open and hidden charm-hadron states, weak decays induce a displacement of a few hundreds of μm, $c\tau(D_s^+) = (151 \pm 1)$ μm [6]. This average displacement from the primary vertex enlarges with increasing $D_s^+$-meson momentum due to the Lorentz boost. Non-prompt $D_s^+$ mesons, produced in beauty-hadron decays, on average are more displaced from the primary vertex than prompt $D_s^+$ mesons due to the larger proper decay length of beauty hadrons, $c\tau(H_b) \simeq 500$ μm [6]. The main characteristics of $D_s^+$ mesons and the beauty hadrons of interest are reported in Tab. 4.2. Considering calculations based on the branching ratios reported in the Table and on the beauty-quark fragmentation fractions from $Z \to b\bar{b}$ decays [6], around half of non-prompt $D_s^+$ mesons originate from $B^0$- and $B^+$-meson decays while the other half is produced from $B_s^0$ mesons. The contribution due to $\Lambda_b^0$-baryon decays is negligible.

The $D_s^+$-meson candidates are built by combining triplets of tracks with the correct charge-sign arrangement, i.e., $(+,-,+)$ for $D_s^+$ mesons and $(-,+,-)$ for their charge conjugates. The invariant mass and momentum of the $D_s^+$-meson candidate are computed from the masses and momenta of the measured decay products, where each track momentum is evaluated at the point along the track closest to the decay vertex. The kaon invariant mass is assigned unambiguously to the track with the opposite charge sign with respect to the $D_s^\pm$-meson candidate. The two possible assignments of the kaon and pion masses are considered for the like-sign tracks. As an example, for a triplet with a global positive charge sign, two candidate configurations are constructed with mass assignments ($K^+$, $K^-$, $\pi^+$) and ($\pi^+$, $K^-$, $K^+$).

Numerous candidates are created given the large amount of track combinations, the vast majority of them being combinatorial background. The average number of reconstructed $D_s^+$-meson candidates per event corresponds to about 3 in pp collisions, and $10^4$ ($10^3$) in central (semicentral) Pb-Pb collisions, after the application of loose filtering selections. Stricter selections are needed to increase the signal-to-background ratio and the statistical significance of the measurement.

Table 4.2: Principal characteristics of the $D_s^+$ mesons and their decay relevant for this Thesis studies, together with main beauty hadrons that decay into $D_s^+$ mesons [6].

| Particle | Mass (MeV/$c$) | $c\tau$ (μm) | Decay channel | BR (%) |
|---|---|---|---|---|
| $D_s^+$ ($c\bar{s}$) | $1968.34 \pm 0.07$ | $151 \pm 1$ | $\phi(1020)[\rightarrow K^+K^-]\pi^+$ | $2.24 \pm 0.08$ |
| $B^+$ ($u\bar{b}$) | $5279.34 \pm 0.12$ | $491 \pm 1$ | $D_s^+ + X$ | $7.9^{+1.4}_{-1.3}$ |
| | | | $D_s^- + X$ | $1.10^{+0.40}_{-0.32}$ |
| $B^0$ ($d\bar{b}$) | $5279.65 \pm 0.12$ | $455 \pm 1$ | $D_s^+ + X$ | $10.3^{+2.1}_{-1.8}$ |
| | | | $D_s^- + X$ | $< 2.6$ |
| $B_s^0$ ($s\bar{b}$) | $5366.88 \pm 0.14$ | $458 \pm 3$ | $D_s^- + X$ | $93 \pm 25$ |
| $\Lambda_b^0$ (udb) | $5619.60 \pm 0.17$ | $441 \pm 3$ | $\Lambda_c^+ D_s^-$ | $1.10 \pm 0.10$ |

These selections can be divided in: single-track selections, particle identification of the decay products, and selections based on the $D_s^+$-meson decay kinematics and topology.

### 4.2.1 Single-track selections

The tracks used to construct the $D_s^+$-meson candidates were required to fulfil several quality and kinematic criteria. These requirements were:

- successful fit in the last step of the tracking procedure based on the Kalman-filter algorithm and described in Section 2.5;
- pseudorapidity interval $|\eta| < 0.8$;
- $p_T > 0.3$ GeV/$c$ in pp collisions and $p_T > 0.6$ (0.4) GeV/$c$ in central (semi-central) Pb–Pb collisions;
- $\chi^2/ndf < 2$ (1.25) for the TPC track-momentum fit in pp (Pb–Pb) collisions[1], where $ndf$ is the number of degrees of freedom entering the track-fit procedure and corresponding to two times the number of TPC clusters;

[1]In the reconstruction of the data sample of Pb–Pb collisions collected in 2018, a new parametrisation of the uncertainties on the TPC clusters was adopted. This resulted in a smaller width of the $\chi^2/ndf$ distribution compared to the 2017 data sample of pp collisions. The two different threshold values reported were chosen to provide the same track rejection.

- at least 70, out of a maximum of 159, crossed rows in the TPC;
- ratio of crossed rows over findable clusters in the TPC, i.e., the number of hit TPC pad rows divided by the pad rows which clusters can be found based on the track geometry, larger than 0.8;
- at least two hits in the ITS, out of a maximum of six, of which at least one in either of the two SPD layers.

This latter condition rejects the tracks originating from pileup collisions in the TPC readout time. In addition, it is particularly relevant for measurements of heavy-flavour hadrons, since a detected point close to the primary and decay vertices provides a good track impact-parameter and decay-vertex position resolution. Tracks that satisfy these selections have a transverse-momentum resolution better than 1% for $p_{\mathrm{T}} = 1$ GeV/$c$ and of about 2% for $p_{\mathrm{T}} = 10$ GeV/$c$, and the resolution on the track impact-parameter in the transverse plane is better than 75 µm for $p_{\mathrm{T}} > 1$ GeV/$c$, as shown in Fig. 2.8.

In Pb–Pb collisions, at least 50 clusters available for particle identification in the TPC were additionally required. Moreover, a selection on the impact parameter $|d_0| < 60$ (25) µm of tracks with $p_{\mathrm{T}} < 2$ GeV/$c$ was applied in central (semicentral) Pb–Pb collisions to suppress the large combinatorial background selecting tracks displaced from the primary vertex.

The aforementioned track-selection criteria reduce the rapidity region in which $\mathrm{D_s^+}$ mesons can be studied. As shown in Fig. 4.1, the acceptance in rapidity drops steeply to zero for $|y| > 0.5$ at low $p_{\mathrm{T}}$ and for $|y| > 0.8$ at $p_{\mathrm{T}} > 5$ GeV/$c$. Therefore, a $p_{\mathrm{T}}$-dependent fiducial acceptance cut $|y| < y_{\mathrm{fid}}(p_{\mathrm{T}})$ was imposed on the $\mathrm{D_s^+}$-meson rapidity, with $y_{\mathrm{fid}}(p_{\mathrm{T}})$ defined as

$$y_{\mathrm{fid}}(p_{\mathrm{T}}) = \begin{cases} 0.5 + \frac{1.9}{15} p_{\mathrm{T}} - \frac{0.2}{15} {p_{\mathrm{T}}}^2 & 0 < p_{\mathrm{T}} < 5 \text{ GeV}/c \\ 0.8 & p_{\mathrm{T}} > 5 \text{ GeV}/c. \end{cases} \tag{4.1}$$

### 4.2.2 Particle-identification selections

To reduce the combinatorial background contribution, especially in the low $p_{\mathrm{T}}$ region, particle identification requirements were applied to the decay tracks. The PID selections are based on the measurement of the specific energy loss $\mathrm{d}E/\mathrm{d}x$ in the TPC detector and of the time of flight from the interaction vertex to the TOF detector, as introduced in Section 2.6.

To assess if a track is compatible with a specific particle species, a pion or a kaon in the case of $\mathrm{D_s^+}$-meson decays, the difference between the measured signal

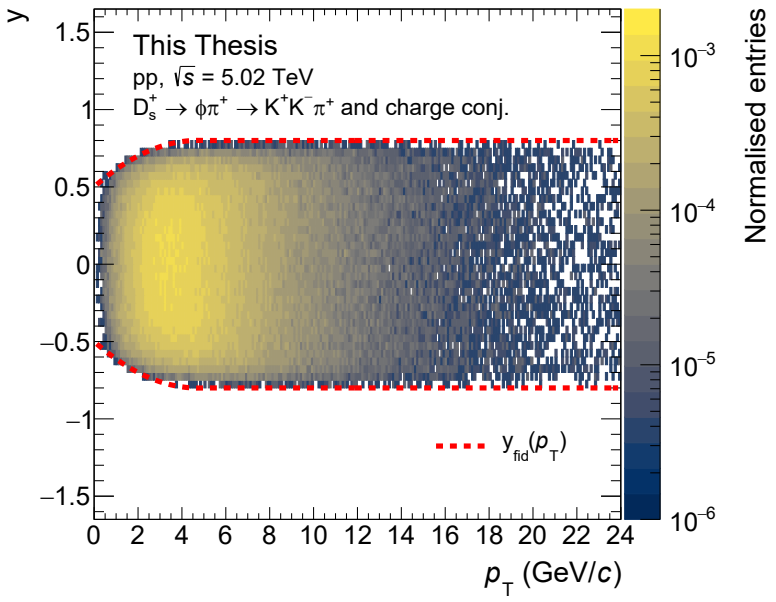



Figure 4.1: Rapidity distribution versus transverse momentum for $D_s^+$ mesons reconstructed in pp collisions at $\sqrt{s} = 5.02$ TeV. The fiducial acceptance region is defined by $|y| < y_{\rm fid}(p_{\rm T})$.

$S_{meas}$ and the expected one $S_{exp}^{\pi,\rm K}$ is computed for a given mass hypothesis. Then it is expressed in units of detector resolution

$$n\sigma^{\pi,\rm K} = \frac{|S_{meas} - S_{exp}^{\pi,\rm K}|}{\sigma^{\pi,\rm K}}, \tag{4.2}$$

where $\sigma^{\pi,K}$ is the resolution on the energy-loss or the time-of-flight signals for each particle species.

The strategy employed to select the tracks starting from the TPC and TOF information is schematised in Fig. 4.2. For both the TPC and TOF detectors, a response value of $-1$, 0, or $+1$ is associated to each mass hypothesis. If the detector signal is not available, a response value of 0 is given. This choice preserves the large fraction of low-$p_{\rm T}$ tracks that do not reach the TOF detector and thus have no TOF signal. A response value of $-1$ is assigned for both detectors if $n\sigma > 3$. For the TPC detector, an additional comparison is performed between the measured $n\sigma_{\rm TPC}$ and a threshold value $n\sigma_{\rm TPC}^{max}$, which depends on the track transverse momentum. The chosen values of $n\sigma_{\rm TPC}^{max}$ are reported in Tab. 4.3 and are the same for the two PID strategies adopted, denominated *conservative PID* and *strong PID*. In the case of the conservative PID strategy, a response value of 0 for the TOF detector is assigned to tracks with $n\sigma_{\rm TOF} < 3$ and $p_{\rm T} > 1.5$ GeV/$c$. The aim is to reject tracks not fulfilling the TPC PID requirements in a $p_{\rm T}$ region where the TOF discrimination power between the different mass hypotheses is low. A track is

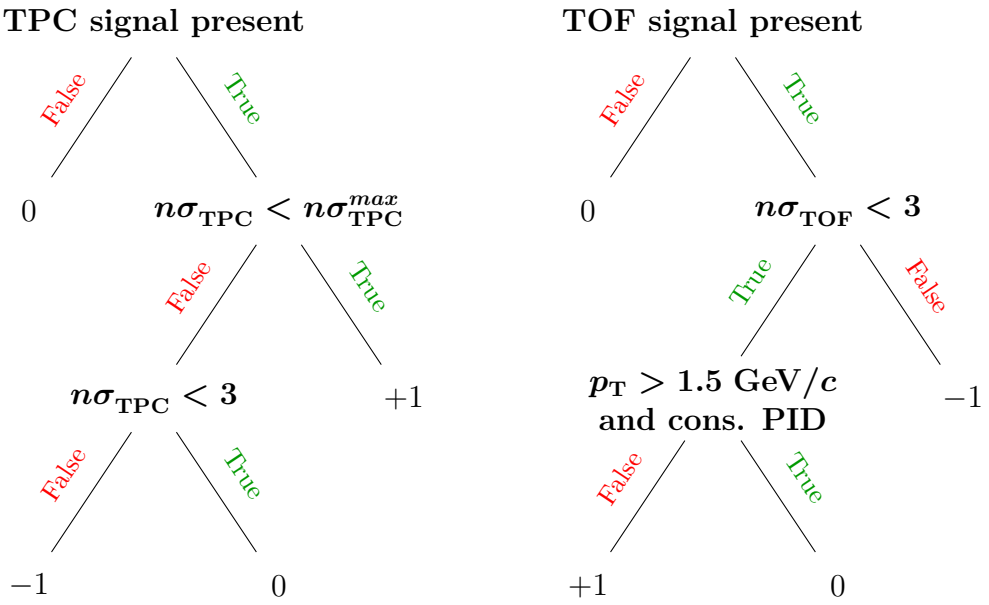



Figure 4.2: Schematisation of the PID selection criteria for a given mass hypothesis. More information is reported in the text.

considered compatible with a given mass hypothesis if the sum of the TPC and TOF response values is greater than or equal to zero (one) for the conservative (strong) PID strategy.

A $D_s^+$-meson candidate is rejected if at least one of the decay tracks is not compatible with either of the pion or kaon hypotheses. In addition, the track with the opposite charge sign with respect to the $D_s^+$-meson candidate must satisfy the kaon hypothesis. For what concerns the like-sign tracks, one of them must be compatible with a kaon and the other one with a pion. These PID selections preserve about 99% (85%) of the signal for the conservative (strong) PID strategy, and reject up to 90% of the background depending on the transverse momentum. The PID requirements are also useful to reduce the contribution of particle candidates with

Table 4.3: Values of $n\sigma_{\mathrm{TPC}}^{max}$ for the conservative and strong PID strategies as a function of the decay-track transverse momentum.

| Track $p_{\mathrm{T}}$ (GeV/$c$) | $n\sigma_{\mathrm{TPC}}^{max}$ |
|---|---|
| $p_{\mathrm{T}} < 0.6$ | 2 |
| $0.6 < p_{\mathrm{T}} < 0.8$ | 1 |
| $p_{\mathrm{T}} > 0.8$ | 2 |

the wrong mass assigned to the like-sign tracks, mentioned in Section 4.2. The conservative PID selection together with a selection on the kaon-pair invariant mass, which will be introduced in Section 4.2.3, discards about 98% of the candidates with the wrong mass hypothesis.

In the measurements presented in the next Chapters, the particle-identification information is also provided to the machine-learning models used to perform the candidate selection. For each one of the three $D_s^+$-meson decay tracks, its $n\sigma$ difference with respect to the kaon and pion hypotheses, independently for the TPC and TOF detectors, is available. Therefore, a total of twelve variables related to the candidate PID can be employed. To keep low the number of features used in the model training and to account for tracks that do not have the TOF signal, the $n\sigma^{\pi,K}$ obtained from the TPC and TOF measurements are combined as follows

$$n\sigma^{\pi,\mathrm{K}}_{\mathrm{comb}} = \begin{cases} \frac{1}{\sqrt{2}}\sqrt{\left(n\sigma^{\pi,\mathrm{K}}_{\mathrm{TPC}}\right)^2 + \left(n\sigma^{\pi,\mathrm{K}}_{\mathrm{TOF}}\right)^2} & \\ n\sigma^{\pi,\mathrm{K}}_{\mathrm{TPC}} & \text{only TPC PID information} \\ n\sigma^{\pi,\mathrm{K}}_{\mathrm{TOF}} & \text{only TOF PID information.} \end{cases} \tag{4.3}$$

This approach leads to six $n\sigma^{\pi,\mathrm{K}}_{\mathrm{comb}}$ features, two for each track, that encapsulate the PID information of the $D_s^+$-meson candidate.

### 4.2.3 Topological and kinematic variables

As mentioned at the beginning of this Chapter, a supervised machine-learning technique is used to perform the $D_s^+$-meson candidate selection. The variables provided as input to the machine-learning algorithm are mainly based on the displaced decay-vertex topology of $D_s^+$ mesons and the kinematic properties of the decay chain, which contain a $\phi$ meson as intermediate state. In fact, combinatorial background candidates are principally built from tracks produced at the primary vertex and have an impact parameter different from zero only due to the detector resolution. Moreover, the large mean proper decay length of beauty hadrons, compared to the $D_s^+$-meson one, and their more complex decay chain induce different characteristics between prompt and non-prompt $D_s^+$ mesons.

In Fig. 4.3, the variables employed in the analyses presented in this Thesis are sketched for the case of prompt $D_s^+ \to \phi\pi^+ \to K^-K^+\pi^+$ decays. They are described in more detail below. The distributions of prompt and non-prompt $D_s^+$ mesons, shown in Figs. 4.4, 4.5, 4.6, and 4.7, are obtained from Monte Carlo simulations, while those of the combinatorial background from data excluding the invariant-mass region of the signal.

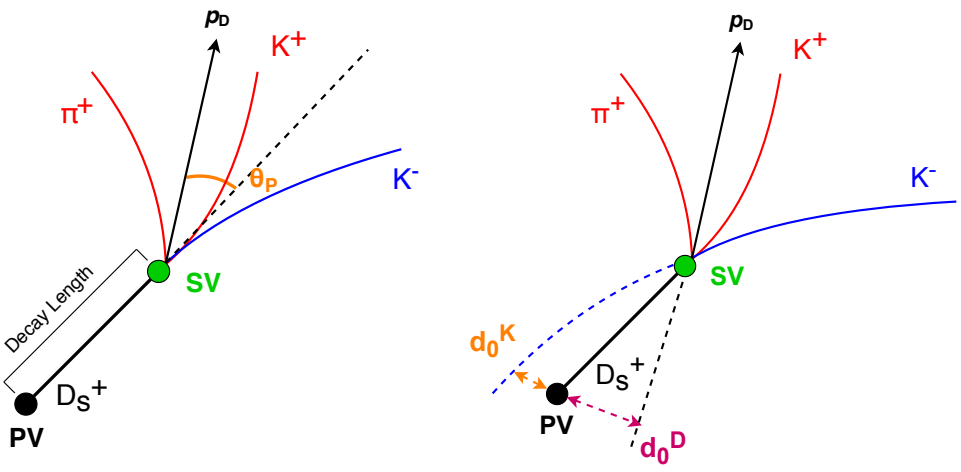



Figure 4.3: Sketch of the $D_s^+ \to \phi\pi^+ \to K^-K^+\pi^+$ decay for promptly produced $D_s^+$ mesons with some topological quantities used in the selections.

### Decay length ($DL$, $DL^{xy}$)

The decay length is measured as the distance between the primary and secondary vertex. It is an approximation of the real decay length due to the curvature of the $D_s^+$-meson trajectory in the magnetic field. However, the effect is negligible since the mean proper decay length of $D_s^+$ mesons is small. The decay-length distributions of prompt and non-prompt $D_s^+$ mesons, and combinatorial background are reported in the top left panel of Fig. 4.4 for pp collisions at $\sqrt{s} = 5.02$ TeV in the $3 < p_T < 4$ GeV/$c$ interval. The decay length of non-prompt $D_s^+$ mesons is on average larger than the one of prompt $D_s^+$ mesons since it includes the distance covered by the beauty hadron before decaying. Moreover, the mean decay length of signal candidates increases with the transverse momentum due to the Lorentz boost, as shown for prompt $D_s^+$ mesons in the left panel of Fig. 4.5.

The decay length projected in the plane normal to the beam direction is also considered in the candidate selection. This projection exploits the better detector resolution along the x and y coordinates than in the z coordinate.

### Normalised decay length in the transverse plane ($DL^{xy}/\sigma_{DL^{xy}}$)

The $DL^{xy}/\sigma_{DL^{xy}}$ variable is obtained dividing the decay length measured in the transverse plane by its uncertainty. In the top right panel of Fig. 4.4, the $DL^{xy}/\sigma_{DL^{xy}}$ distributions of prompt $D_s^+$ mesons, non-prompt $D_s^+$ mesons and combinatorial background are shown in the $3 < p_T < 4$ GeV/$c$ interval. Similarly to what observed for the $DL$ distributions, the $DL^{xy}/\sigma_{DL^{xy}}$ is on average larger for signal $D_s^+$ mesons compared to the combinatorial background, and the non-prompt $D_s^+$-meson distribution is broader than the prompt one.

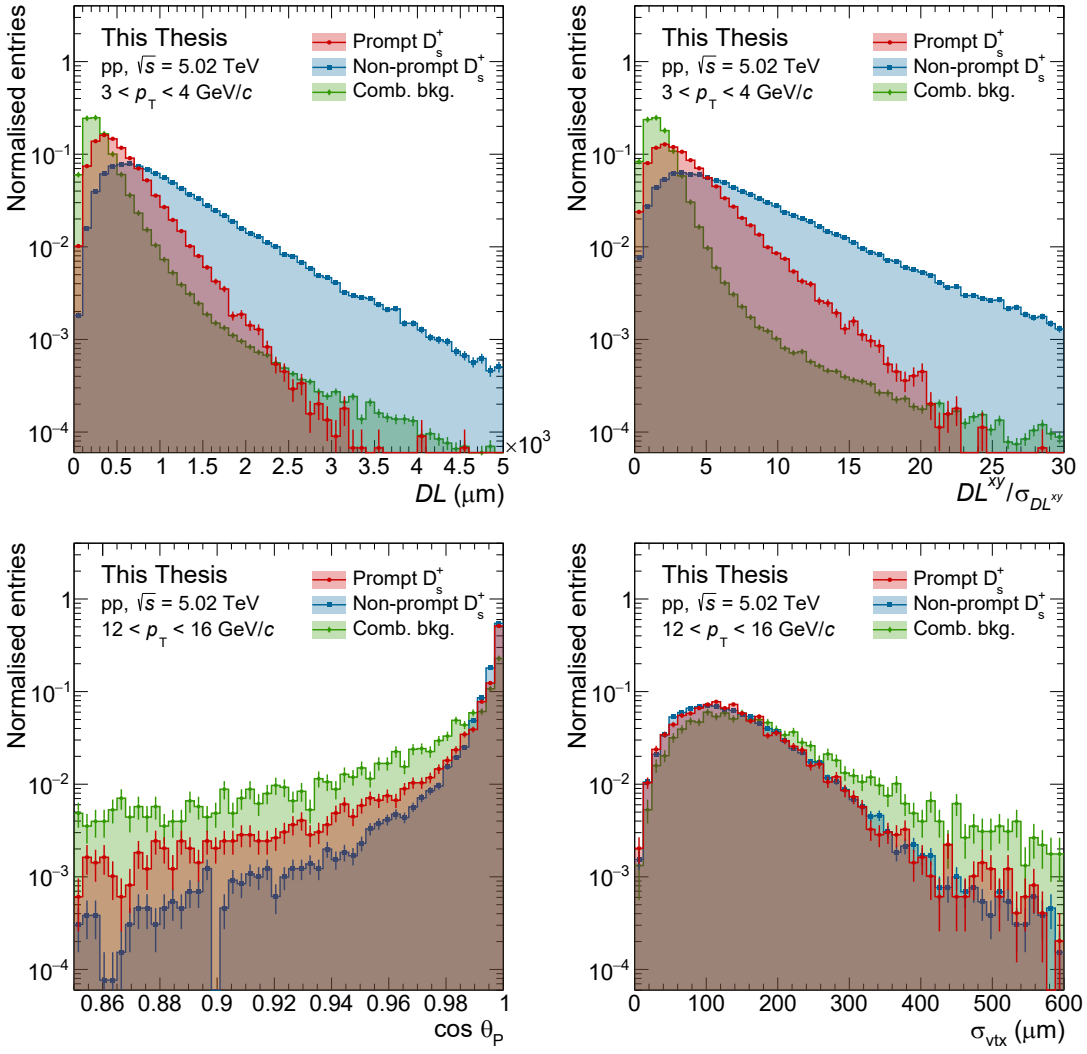



Figure 4.4: Distributions of $DL$, $DL^{xy}/\sigma_{DL^{xy}}$, $\cos\theta_{\mathrm{P}}$, and $\sigma_{\mathrm{vtx}}$ for prompt (red) and non-prompt (blue) $\mathrm{D_s^+}$ mesons, and combinatorial background (green) in pp collisions at $\sqrt{s} = 5.02$ TeV in the $3 < p_{\mathrm{T}} < 4$ GeV/$c$ and $12 < p_{\mathrm{T}} < 16$ GeV/$c$ transverse-momentum intervals.

### Cosine of pointing angle ($\cos\theta_{\mathrm{P}}$, $\cos\theta_{\mathrm{P}}^{xy}$)

The pointing angle $\theta_{\mathrm{P}}$ is the angle between the momentum of the reconstructed $\mathrm{D_s^+}$ meson and the line connecting the primary and secondary vertices. Neglecting the experimental resolution on the vertex positions and the momentum, the pointing angle is expected to be exactly zero for prompt $\mathrm{D_s^+}$ mesons and close to zero for non-prompt $\mathrm{D_s^+}$ mesons. Since in the latter case, the $\mathrm{D_s^+}$ meson does not originate directly from the primary vertex but from an intermediate decay of a beauty

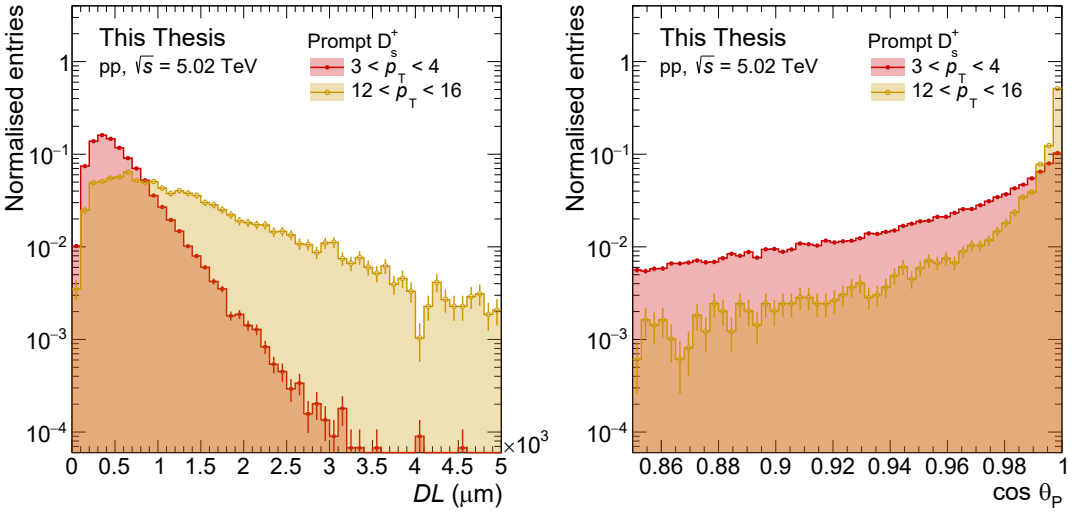



Figure 4.5: Comparison of $DL$ and $\sigma_{\rm vtx}$ distributions between the $3 < p_{\rm T} < 4$ GeV/$c$ (red) and $12 < p_{\rm T} < 16$ GeV/$c$ (yellow) intervals for prompt $D_s^+$ mesons in pp collisions at $\sqrt{s} = 5.02$ TeV.

hadron. The distribution of $\cos\theta_{\rm P}$ is peaked at one for the signal candidates while it is broader for the background, as shown in the bottom left panel of Fig. 4.4 in the $12 < p_{\rm T} < 16$ GeV/$c$ interval. The distribution of non-prompt $D_s^+$ mesons is narrower than the prompt one due to their better pointing-angle resolution, which is caused by the on average larger secondary-vertex displacement.

The right panel of Fig. 4.5 shows the evolution of the $\cos\theta_{\rm P}$ distribution of prompt $D_s^+$ mesons with $p_{\rm T}$, induced by the variation of the decay length, and the vertexing and momentum resolutions. As for the decay length, the $\theta_{\rm P}$ angle can be measured also in the transverse plane to exploit the better resolution that can be achieved.

### Daughter-track dispersion around the decay vertex ($\sigma_{\rm vtx}$)

The $\sigma_{\rm vtx}$ quantity is computed as

$$\sigma_{\rm vtx} = \sqrt{\sum_{i=1}^{3} d_i^2} \ , \tag{4.4}$$

where $d_i$ is the distance of minimal approach between the decay track $i$ and the secondary vertex. For a real $D_s^+$ meson, all the tracks originate from the decay vertex and $\sigma_{\rm vtx}$ is zero, neglecting the vertexing and tracking resolution. The $\sigma_{\rm vtx}$ distribution is broader for the combinatorial background, as shown in the bottom right panel of Fig. 4.4.

### Impact parameter in the transverse plane ($d_{0,\mathrm{D}}^{xy}$)

The $d_{0,\mathrm{D}}^{xy}$ variable is the $\mathrm{D_s^+}$-meson candidate impact parameter with respect to the primary vertex projected in the transverse plane. Values of $d_{0,\mathrm{D}}^{xy}$ closely spread around zero characterise prompt $\mathrm{D_s^+}$ mesons which really originate from the primary vertex. In addition to reject the combinatorial background, the $d_{0,\mathrm{D}}^{xy}$ is also useful to discriminate between prompt and non-prompt D mesons. In fact, the distribution of this variable is much larger for non-prompt $\mathrm{D_s^+}$ mesons than prompt ones due to the beauty-hadron decay length, as displayed in the left panel of Fig. 4.6, where the absolute value of $d_{0,\mathrm{D}}^{xy}$ is shown for the $3 < p_{\mathrm{T}} < 4$ GeV/$c$ transverse-momentum interval.

### Maximum normalised residual between the reconstructed and expected daughter-track impact parameters in the transverse plane ($D_0^{xy}$)

The $D_0^{xy}$ variable is computed as

$$D_0^{xy} = \max_{i=1}^{3} \left\{ |\Delta d_{0,i}^{xy}| \right\} \times sign\left(\Delta d_{0,max}^{xy}\right), \tag{4.5}$$

where

$$\Delta d_{0,i}^{xy} = \frac{d_{0,i}^{xy}(\mathrm{reco}) - d_{0,i}^{xy}(\mathrm{exp})}{\sqrt{\sigma^2_{d_{0,i}^{xy}(\mathrm{reco})} + \sigma^2_{d_{0,i}^{xy}(\mathrm{exp})}}} \tag{4.6}$$

is the difference between the reconstructed impact-parameter in the transverse plane $d_{0,i}^{xy}(\mathrm{reco})$ and the expected value $d_{0,i}^{xy}(\mathrm{exp})$, for the decay track $i$, divided by the total uncertainty. The expected impact parameter $d_{0,i}^{xy}(\mathrm{exp})$ is computed from the $\mathrm{D_s^+}$-meson $DL^{xy}$ and the angle $\theta_i^{xy}$ between the decay-track and the mother-particle momentum in the transverse plane as $d_{0,i}^{xy}(\mathrm{exp}) \simeq DL^{xy} \sin\theta_i^{xy}$, assuming linear trajectories and neglecting the presence of the magnetic field. The uncertainty on $d_{0,i}^{xy}(\mathrm{exp})$ is estimated as $\sigma_{d_{0,i}^{xy}(\mathrm{exp})} \simeq \sigma_{DL^{xy}} \sin\theta_i^{xy}$, not considering the $\theta_i^{xy}$ uncertainty.

This variable is correlated with the $\mathrm{D_s^+}$-meson $d_{0,\mathrm{D}}^{xy}$, and it has different distributions between combinatorial background, prompt and non-prompt $\mathrm{D_s^+}$ mesons, as illustrated in the right panel of Fig. 4.6. The Figure shows the absolute value of $D_0^{xy}$ exploiting the symmetry of the variable distribution with respect to zero.

### $\cos^3\theta'(\mathrm{K})$

The $\theta'(\mathrm{K})$ angle is the angle between one of the kaons and the pion in the $\mathrm{K^+K^-}$ rest frame. In the $\mathrm{D_s^+} \to \phi\pi^+ \to \mathrm{K^-K^+}\pi^+$ decay, the spin of the $\phi$ vector meson is aligned to its direction of motion with respect to the $\mathrm{D_s^+}$-meson flight line. Therefore, it can be demonstrated that the $\cos^3\theta'(\mathrm{K})$ distribution is flat for real $\mathrm{D_s^+}$-meson candidates, and it is peaked at zero for the combinatorial background [235], as shown in the left panel of Fig. 4.7.

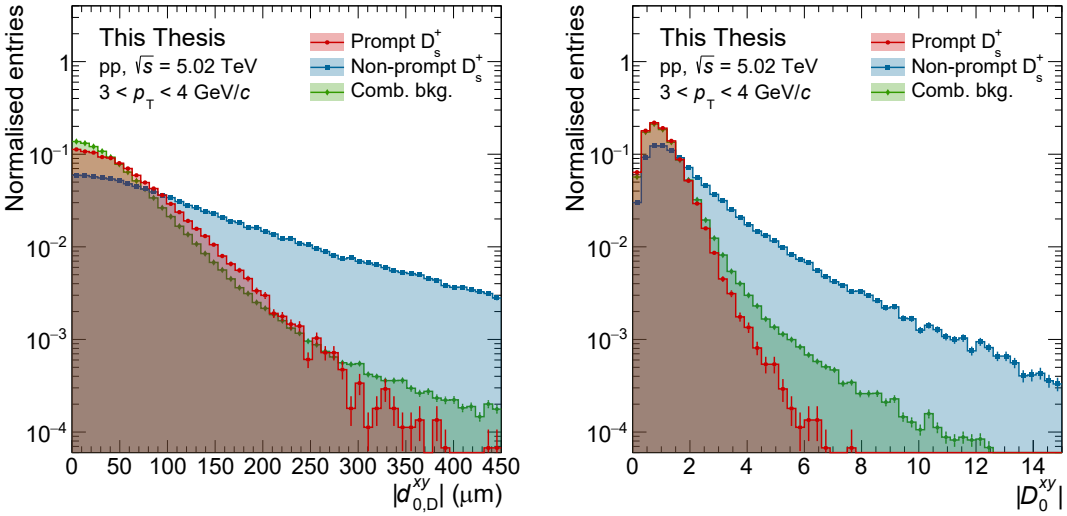



Figure 4.6: Distributions of $d_{0,\mathrm{D}}^{xy}$ and $D_0^{xy}$ absolute values for prompt $\mathrm{D_s^+}$ mesons (red), non-prompt $\mathrm{D_s^+}$ mesons (blue) and combinatorial background (green) in the $3 < p_\mathrm{T} < 4$ GeV/$c$ interval in pp collisions at $\sqrt{s} = 5.02$ TeV.

### Absolute difference between the kaon-pair invariant mass and the $\phi$-meson mass ($\Delta M_{\mathrm{KK}}$)

The choice of reconstructing the $\mathrm{D_s^+}$ mesons in the $\mathrm{D_s^+} \to \phi\pi^+ \to \mathrm{K^-K^+\pi^+}$ decay channel enables also the possibility to exploit the narrow width of the $\phi$-meson resonance. To this end, the invariant mass of the reconstructed $\mathrm{K^+K^-}$ pair is required to be compatible with the $\phi$-meson mass, $M(\phi) = (1019.461 \pm 0.016)$ MeV/$c^2$ [6], and the variable

$$\Delta M_{\mathrm{KK}} = |M_{\mathrm{inv}}(\mathrm{KK}) - M(\phi)| \tag{4.7}$$

is computed. $\mathrm{D_s^+}$-meson signal candidates usually have $\Delta M_{\mathrm{KK}}$ values smaller than a few tens of MeV/$c^2$, due to the $\phi$-meson width convoluted with the detector resolution. The combinatorial background is instead equally spread on a wide range of values, as displayed in the right panel of Fig. 4.7 for the $3 < p_\mathrm{T} < 4$ GeV/$c$ interval. Furthermore, a selection on $\Delta M_{\mathrm{KK}}$ contributes to reject, together with the PID requirements, the $\mathrm{D_s^+}$-meson candidates with the wrong mass assignment as discussed in Section 4.2.2. Finally, $\mathrm{D_s^+}$ mesons decaying in the $\mathrm{K^-K^+\pi^+}$ final state through the $\mathrm{D_s^+} \to \overline{\mathrm{K}}^{*0}\mathrm{K^+} \to \mathrm{K^-K^+\pi^+}$ channel with BR $= (2.58 \pm 0.08)\%$ [6] are rejected by the aforementioned selections, and their residual contamination is smaller than 1%.

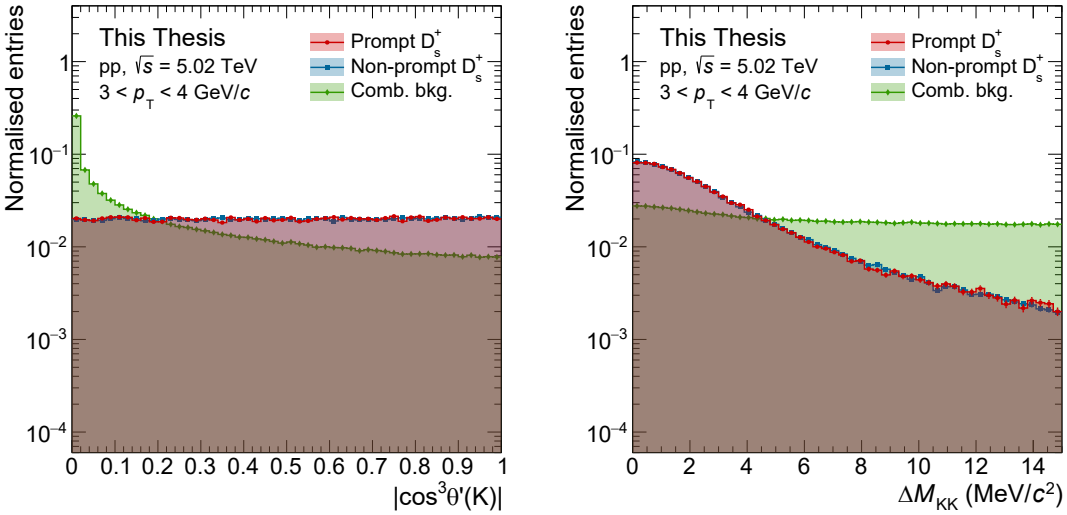



Figure 4.7: Distributions of the $\cos^3\theta'(\mathrm{K})$ absolute value and $\Delta M_{\mathrm{KK}}$ for prompt $\mathrm{D_s^+}$ mesons (red), non-prompt $\mathrm{D_s^+}$ mesons (blue) and combinatorial background (green) in the $3 < p_{\mathrm{T}} < 4$ GeV/$c$ interval in pp collisions at $\sqrt{s} = 5.02$ TeV.

### 4.2.4 Variable distributions in pp and Pb–Pb collisions

The distributions of the topological and kinematic variables introduced in the previous Section may differ between pp and Pb–Pb collisions. These differences are driven by the improved resolution on the primary-vertex position available in Pb–Pb interactions and caused by the larger multiplicity of produced tracks with respect to pp collisions. Moreover, stricter selection criteria are applied in Pb–Pb collisions than in pp ones to reject the huge combinatorial background and to keep the data size manageable for the analysis.

In Fig. 4.8, the distributions of the $DL^{xy}/\sigma_{DL^{xy}}$ and the absolute value of the $d^{xy}_{0,\mathrm{D}}$ for prompt $\mathrm{D_s^+}$ mesons in pp collisions at $\sqrt{s} = 5.02$ TeV, and in central and semicentral Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV in the $3 < p_{\mathrm{T}} < 4$ GeV/$c$ interval are shown. Samples of signal prompt $\mathrm{D_s^+}$ mesons for the different collision systems are obtained from Monte Carlo simulations, while the selections employed will be detailed in Sections 5.1.1 and 6.1.1 for pp and Pb–Pb collisions, respectively. The $DL^{xy}/\sigma_{DL^{xy}}$ distributions in central and semicentral Pb–Pb collisions start from higher values than in pp collisions showing the effect of the applied selections. In addition, the reduced uncertainty on the primary-vertex position leads to narrower distributions of $d^{xy}_{0,\mathrm{D}}$ in Pb–Pb collisions than in pp. For the same reason, the $DL^{xy}/\sigma_{DL^{xy}}$ distributions are slightly larger in central and semi-central Pb–Pb interactions than in pp ones.

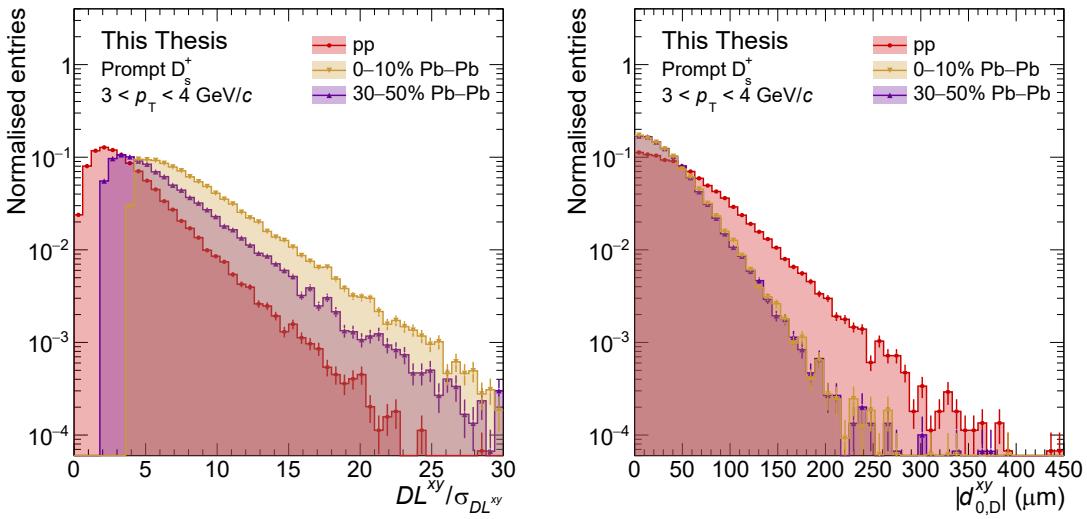



Figure 4.8: Distributions of the $DL^{xy}/\sigma_{DL^{xy}}$ and the $d^{xy}_{0,\mathrm{D}}$ absolute value for prompt $\mathrm{D_s^+}$ mesons in pp collisions at $\sqrt{s} = 5.02$ TeV (red), and in central (yellow) and semicentral (purple) Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV in the $3 < p_{\mathrm{T}} < 4$ GeV/$c$ interval.

## 4.3 Monte Carlo simulations

Precise Monte Carlo (MC) simulations of the kinematics and topology of the $\mathrm{D_s^+} \to \phi\pi^+ \to \mathrm{K^-K^+}\pi^+$ decays are fundamental for the measurement of the $\mathrm{D_s^+}$-meson production, in particular for the training of the machine-learning models, and the estimation of the acceptance and efficiency corrections. To this end, MC simulations must also reproduce the conditions, tracking resolution, and PID performance of the experiment detectors.

The PYTHIA 8.243 event generator [82, 168] with the Monash-13 tune [135] was used to simulate proton–proton collisions at $\sqrt{s} = 5.02$ TeV. A $\mathrm{c\bar{c}}$- or $\mathrm{b\bar{b}}$-quark pair was required for each simulated PYTHIA pp event and the $\mathrm{D_s^+}$ mesons were forced to decay in the hadronic channel of interest for the analysis. In the case of Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV, the underlying event was simulated with the HIJING 1.36 event generator [169]. In each HIJING simulated event, $\mathrm{D_s^+}$-meson signals were injected through multiple $\mathrm{c\bar{c}}$- and $\mathrm{b\bar{b}}$-quark pairs generated with PYTHIA 8, and the same requirements as for pp collisions were applied. For both pp and Pb–Pb collisions, the generated particles were propagated through the ALICE detectors using the GEANT3 transport package [170].

The MC simulations employed in this Thesis included a detailed description of the ALICE experiment geometry and detector response. They were configured to reproduce the position and width of the interaction-vertex distribution, and the status of all ALICE detectors for what concerns the number of active channels,

gain, noise level and the precision of the calibration. The evolution with time of the aforementioned parameters during the pp and Pb–Pb data taking periods was also taken into account. The simulation and reconstruction of MC data were performed with the AliRoot framework, which interfaces with the event generators and the GEANT3 package. Moreover, the reconstruction of simulated and real data was executed using the same code. To guarantee a satisfactory agreement between real data and MC simulations of the topological and kinematic quantities of interest for this Thesis studies, a dedicated procedure was employed to tune the single-track parameters and uncertainties for MC-simulated particles. A detailed description of this procedure is reported in Ref. [236]. Similarly, in MC simulations, the $n\sigma$ PID variables were sampled to match the distributions observed in data.

### 4.3.1 Acceptance and efficiency correction

The efficiency and acceptance correction factors were calculated from the MC simulations detailed above. These corrections are necessary to compute the production yield of $\mathrm{D_s^+}$ mesons at midrapidity, $|y| < 0.5$, starting from the measured raw yield obtained after the candidate reconstruction and the application of topological, kinematic, and PID selections. The efficiency was estimated as a function of $p_{\mathrm{T}}$ with the following formula

$$\epsilon\ (p_{\mathrm{T}}) = \frac{N_{\mathrm{reco}}(p_{\mathrm{T}})\big|_{|y|<y_{\mathrm{fid}}(p_{\mathrm{T}})}}{N^{\mathrm{acc}}_{\mathrm{gen}}(p_{\mathrm{T}})\Big|_{|y|<y_{\mathrm{fid}}(p_{\mathrm{T}})}}, \tag{4.8}$$

where the numerator is the number of reconstructed $\mathrm{D_s^+}$ mesons after having applied the selections. The equation denominator is the number of generated $\mathrm{D_s^+}$ mesons in the fiducial rapidity region $|y| < y_{\mathrm{fid}}$, decaying in the decay channel of interest for the analysis, and with all the three daughters fulfilling the requirements of $p_{\mathrm{T}} > 0.1$ GeV/$c$ and $|\eta| < 0.8$. Only events with a generated primary-vertex position within $\pm 10$ cm from the detector centre in the $z$ direction were considered. The efficiency varies as a function of $p_{\mathrm{T}}$ due to the track-reconstruction efficiency and the selection employed to reject the combinatorial background. The acceptance correction factor was computed as

$$\mathrm{Acc}\ (p_{\mathrm{T}}) = \frac{N^{\mathrm{acc}}_{\mathrm{gen}}(p_{\mathrm{T}})\Big|_{|y|<y_{\mathrm{fid}}(p_{\mathrm{T}})}}{N_{\mathrm{gen}}(p_{\mathrm{T}})\big|_{|y|<y_{\mathrm{fid}}(p_{\mathrm{T}})}}, \tag{4.9}$$

where $N_{\mathrm{gen}}(p_{\mathrm{T}})|_{|y|<y_{\mathrm{fid}}(p_{\mathrm{T}})}$ is the number of generated $\mathrm{D_s^+}$ mesons with $|y| < y_{\mathrm{fid}}$, decaying in the decay channel of interest, and produced in events with a generated primary vertex that satisfies the requirement on its position mentioned above.

An additional correction factor $c_{\Delta y}$ was considered to account for the differences in the rapidity coverage between the fiducial region defined by Eq. (4.1) and the

$|y| < 0.5$ interval. It was defined as the ratio of the generated D-meson yield in $\Delta y = 2y_{\rm fid}$ and that in $|y| < 0.5$. The $c_{\Delta y}$ factor was estimated from FONLL pQCD calculations [72, 73], which provide a good description of the rapidity dependence of the D-meson cross section in pp collisions [237, 238]. As a consistency check, $c_{\Delta y}$ was also computed using simulations based on the PYTHIA 8 event generator, and the obtained result was in agreement with the FONLL-based calculation within 1%. Finally, to account for possible differences of the D-meson production in Pb–Pb collisions and as an extreme variation, a flat rapidity distribution was considered. The discrepancies with respect to the FONLL-based calculation were negligible in comparison to the other sources of systematic uncertainty in this collision system, which will be described in Section 6.6.

# Chapter 5

# Non-prompt and prompt $\mathrm{D_s^+}$-meson production in pp collisions at $\sqrt{s} = 5.02$ TeV

In this Chapter, the measurements of the $p_{\mathrm{T}}$-differential production cross section of prompt and non-prompt $\mathrm{D_s^+}$ mesons at midrapidity ($|y| < 0.5$) in pp collisions at $\sqrt{s} = 5.02$ TeV are presented. The prompt and non-prompt $\mathrm{D_s^+}$-meson cross sections in pp collisions are a sensitive test for pQCD calculations [60], and provide insights into the hadronisation mechanism of charm and beauty quarks. They are also a reference for the measurements of the nuclear modification factor in Pb–Pb collisions. The measurements reported in this Chapter were published in Ref. [71].

The $p_{\mathrm{T}}$-differential cross section of prompt (non-prompt) $\mathrm{D_s^+}$ mesons was calculated for each $p_{\mathrm{T}}$ interval as

$$\left.\frac{\mathrm{d}^2\sigma}{\mathrm{d}p_{\mathrm{T}}\mathrm{d}y}\right|_{|y|<0.5} = \frac{\frac{1}{2}\, f_{\text{(non-)prompt}}(p_{\mathrm{T}}) N^{\mathrm{D_s^\pm,raw}}(p_{\mathrm{T}})\Big|_{|y|<y_{\mathrm{fid}}(p_{\mathrm{T}})}}{(\mathrm{Acc}\times\epsilon)^{\text{(non-)prompt}}(p_{\mathrm{T}}) c_{\Delta y}(p_{\mathrm{T}}) \Delta p_{\mathrm{T}}} \times \frac{1}{\mathrm{BR}\mathcal{L}_{\mathrm{int}}}. \tag{5.1}$$

The measured raw yield $N^{\mathrm{D_s^\pm,raw}}$, containing both $\mathrm{D_s^+}$ and $\mathrm{D_s^-}$ mesons, was divided by a factor two and multiplied by the prompt (non-prompt) fraction $f_{\text{(non-)prompt}}$ to obtain the charge-averaged yield of prompt (non-prompt) $\mathrm{D_s^+}$ mesons. It was then divided by the acceptance times efficiency of prompt (non-prompt) $\mathrm{D_s^+}$ mesons $(\mathrm{Acc}\times\epsilon)^{\text{(non-)prompt}}$, the rapidity-coverage correction factor $c_{\Delta y}$ introduced in Section 4.3.1, the $p_{\mathrm{T}}$-interval width $\Delta p_{\mathrm{T}}$, the decay-channel branching ratio BR reported in Tab. 4.2, and the integrated luminosity $\mathcal{L}_{\mathrm{int}}$. The integrated luminosity was computed as $\mathcal{L}_{\mathrm{int}} = N_{\mathrm{evt}}/\sigma_{\mathrm{vis}}^{\mathrm{pp}}$, where $N_{\mathrm{evt}}$ is the number of analysed events which is about 990 millions, and $\sigma_{\mathrm{vis}}^{\mathrm{pp}} = (50.9 \pm 0.9)$ mb is the cross section for the MB trigger condition [233].

In the following Sections, the steps needed to obtain the $\mathrm{D_s^+}$-meson BDT classification and the resulting selections, the raw-yield extraction approach, and the

corrections necessary to compute the $p_{\mathrm{T}}$-differential cross section are discussed. Finally, the study of the systematic-uncertainty contributions and the results are presented.

## 5.1 Machine-learning selections

As mentioned in Section 3.2, the aim of this Thesis is to explore the application of machine-learning techniques to particle-candidate selection. Therefore, BDTs were employed in the analysis process, replacing the linear selections traditionally used in previous ALICE D-meson measurements [146, 238–241].

The analysis strategy is based on the selection of decay vertices displaced from the interaction point. In addition to the rejection of the combinatorial background, it is also possible to discriminate between prompt and non-prompt $\mathrm{D_s^+}$ mesons thanks to their different characteristics (see Section 4.2.3). To this end, the same multi-class classification BDTs were employed to measure both prompt and non-prompt $\mathrm{D_s^+}$ mesons. Selections on the BDT outputs were applied to reduce the combinatorial background and to enhance the fraction of prompt or non-prompt $\mathrm{D_s^+}$ mesons, depending on the measurement to be performed.

### 5.1.1 Data preparation

The data samples used for the BDT training and performance assessment were composed of prompt and non-prompt $\mathrm{D_s^+}$-meson signal candidates and by the combinatorial background. The signal candidates were obtained from the MC simulations with the PYTHIA 8 event generator introduced in Section 4.3. Combinatorial background candidates were extracted from the sidebands of the $\mathrm{D_s^+}$-meson invariant-mass distribution in the data collected by the experiment. The idea behind this decision was to use real data, where feasible, to avoid losses in the BDT predictive power caused by possible shortcomings of MC simulations. In fact, the quality of the combinatorial-background description is deeply influenced by the modelling of the pp collision by the MC event generator in terms of: particle multiplicity, kinematic distributions of the particle produced, and relative abundances of the different particle species.

In addition to the single-track requirements described in Section 4.2.1, the $\mathrm{D_s^+}$-meson candidates used for the training were required to satisfy the loose kinematic and topological selections reported in Tab. 5.1. They were also required to fulfil the conservative PID strategy. The resulting reconstruction efficiency for prompt and non-prompt $\mathrm{D_s^+}$ mesons as a function of $p_{\mathrm{T}}$ is reported in Fig. 5.1. The efficiencies rise with increasing $p_{\mathrm{T}}$ due to the topological selections applied. The non-prompt $\mathrm{D_s^+}$-meson efficiency is larger than the prompt one due to the narrower $\cos\theta_{\mathrm{P}}$ and $\cos\theta_{\mathrm{P}}^{xy}$ distributions, which are driven by the larger displacement of non-prompt

Table 5.1: Topological and kinematic selections applied on $D_s^+$-meson candidates before the machine-learning step.

| $p_T$ (GeV/$c$) | 1–24 |
|---|---|
| $\sigma_{vtx}$ (μm) < | 600 |
| $\cos\theta_P$ > | 0.85 |
| $\cos\theta_P^{xy}$ > | 0.85 |
| $\Delta M_{KK}$ (MeV/$c^2$) < | 15 |

$D_s^+$ mesons and by resolution effects, as explained in Section 4.2.3.

The BDT training was performed in larger transverse-momentum intervals than the analysis ones to guarantee large enough training samples, especially in the high-$p_T$ region. All the prompt and non-prompt $D_s^+$-meson candidates available in a dedicated MC production, where events containing $c\bar{c}$- or $b\bar{b}$-quark pairs were generated in equal number, were used. The number of combinatorial-background candidates was chosen to be roughly the same as the signal, at low $p_T$. While, at high $p_T$, fewer background candidates were employed to keep small the fraction of

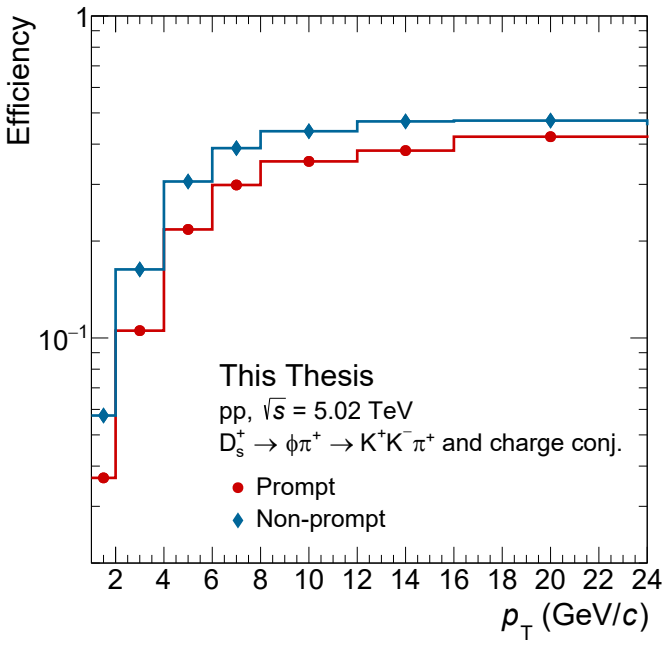



Figure 5.1: Efficiency for the reconstruction of prompt (red) and non-prompt (blue) $D_s^+$ mesons as a function of $p_T$ after the selections used to prepare the data sample for the BDT training and test.

Table 5.2: Number of prompt $\mathrm{D_s^+}$, non-prompt $\mathrm{D_s^+}$ and combinatorial-background candidates used to build the training and test sets in the $p_\mathrm{T}$ intervals in which the training was performed.

| $p_\mathrm{T}$ (GeV/$c$) | 1–2 | 2–4 | 4–6 | 6–12 | 12–24 |
|---|---|---|---|---|---|
| prompt $\mathrm{D_s^+}$ ($\times 10^3$, approx.) | 11 | 52 | 43 | 33 | 5 |
| non-prompt $\mathrm{D_s^+}$ ($\times 10^3$, approx.) | 19 | 110 | 95 | 77 | 11 |
| background ($\times 10^3$, approx.) | 60 | 162 | 41 | 11 | 1 |

real data used in the training. The aim was to avoid possible biases since these candidates were also used for the physics measurement. The approximate number of prompt and non-prompt $\mathrm{D_s^+}$-meson, and combinatorial-background candidates, composing the training and test sets, is reported in Tab. 5.2 for the $p_\mathrm{T}$ intervals in which the training was performed. The training sets were formed by 80% of the available data samples, while the remaining 20% constituted the test sets.

### 5.1.2 BDT training and performance evaluation

The input features provided to the BDTs were the $\mathrm{D_s^+}$-meson candidate topological, kinematic, and PID variables introduced in Sections 4.2.2 and 4.2.3. They are listed hereafter, where the index of the $n\sigma_\mathrm{comb}^{\pi,\mathrm{K}}$ variables indicates the corresponding $\mathrm{D_s^+}$-meson decay track.

- $DL$
- $DL^{xy}$
- $DL^{xy}/\sigma_{DL^{xy}}$
- $\cos\theta_\mathrm{P}$
- $\cos\theta_\mathrm{P}^{xy}$
- $\sigma_\mathrm{vtx}$
- $d_{0,\mathrm{D}}^{xy}$
- $D_0^{xy}$
- $\cos^3\theta'(\mathrm{K})$
- $\Delta M_\mathrm{KK}$
- $n\sigma_\mathrm{comb}^{\pi,0}$
- $n\sigma_\mathrm{comb}^{\pi,1}$
- $n\sigma_\mathrm{comb}^{\pi,2}$
- $n\sigma_\mathrm{comb}^{\mathrm{K},0}$
- $n\sigma_\mathrm{comb}^{\mathrm{K},1}$
- $n\sigma_\mathrm{comb}^{\mathrm{K},2}$

The correlations between the training variables were studied by computing the Pearson correlation coefficient $\rho$ [242] for each variable pair. The linear-correlation coefficients among the training variables, and the $\mathrm{D_s^+}$-meson invariant mass $M$ and transverse momentum $p_\mathrm{T}$, which were not employed in the training, are presented in Fig. 5.2 for the prompt and non-prompt $\mathrm{D_s^+}$ mesons, and the combinatorial background in the $2 < p_\mathrm{T} < 4$ GeV/$c$ interval. Variables that carry the same physical information, such as those related to the candidate decay length, pointing angle, and impact parameter, are strongly correlated among each other, as expected. Signal candidates, both prompt and non-prompt, show a complex correlation pattern between the variables containing the PID information, which is different and less pronounced for background candidates. Small discrepancies are present also

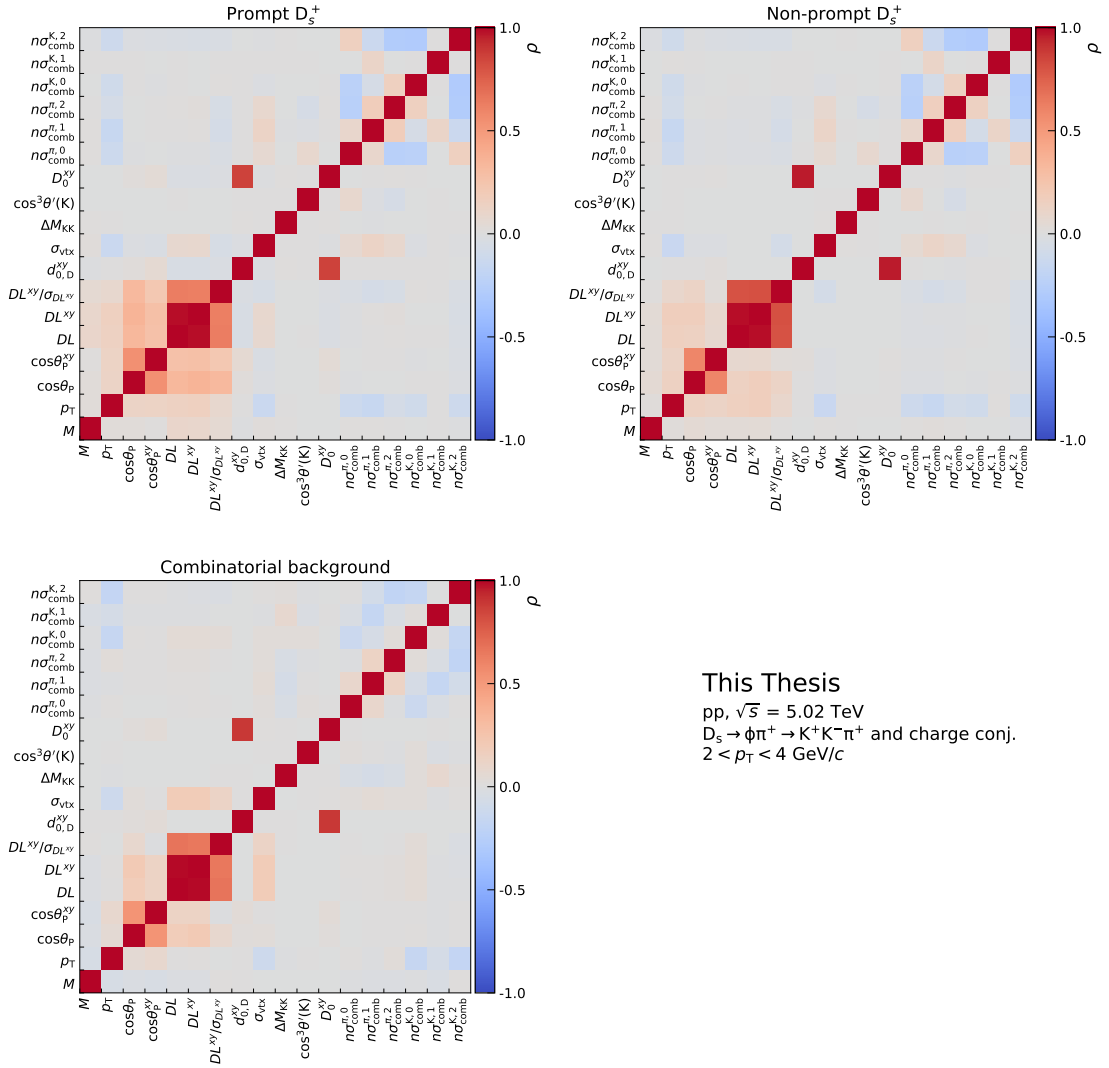



Figure 5.2: Linear-correlation coefficients $\rho$ between the training variables, and the candidate invariant mass $M$ and $p_{\mathrm{T}}$ for prompt (top left panel) and non-prompt (top right panel) $\mathrm{D_s^+}$ mesons, and the combinatorial background (bottom left panel) in the $2 < p_{\mathrm{T}} < 4$ GeV/$c$ interval. The colours from red (fully correlated) to blue (fully anti-correlated) indicate the grade of correlation.

between signal prompt and non-prompt candidates, such as for the correlation between the variables related to the decay length and the cosine of pointing angle. These differences can be exploited by the BDT to discriminate between prompt and non-prompt $\mathrm{D_s^+}$ mesons, and the combinatorial background. No relevant correlations are observed between the training variables and the $\mathrm{D_s^+}$-meson invariant mass or $p_{\mathrm{T}}$. This indicates that the BDT selection is unlikely to modify significantly

Table 5.3: Hyperparameter configurations for the final BDTs in the different $p_{\mathrm{T}}$ intervals used in the training procedure.

| $p_{\mathrm{T}}$ (GeV/$c$) | 1–2 | 2–4 | 4–6 | 6–12 | 12–24 |
|---|---|---|---|---|---|
| `n_estimators` | 665 | 674 | 674 | 672 | 352 |
| `learning_rate` | 0.029 | 0.092 | 0.092 | 0.070 | 0.086 |
| `max_depth` | 4 | 4 | 4 | 4 | 3 |
| `min_child_weight` | 2.7 | 7.0 | 7.0 | 8.7 | 8.6 |
| `subsample` | 0.90 | 0.90 | 0.90 | 0.87 | 0.96 |
| `colsample_bytree` | 0.97 | 0.81 | 0.81 | 0.82 | 0.95 |

the shapes of the invariant-mass and transverse-momentum distributions of $\mathrm{D_s^+}$-meson candidates, and to lead to the formation of structures in the invariant-mass distribution.

The BDT hyperparameters that were tuned for this analysis are those reported in Section 3.2.1, while the other parameters were kept at their default values. The Bayesian optimisation approach was adopted, using a 5-fold cross validation to evaluate the BDT performance for a given set of hyperparameters. The average one-vs-one ROC AUC was employed as the performance metric. The hyperparameter combinations providing the highest ROC AUC values were chosen and are detailed in Tab. 5.3 for the different $p_{\mathrm{T}}$ intervals considered in the training. The final BDTs, used to perform the $\mathrm{D_s^+}$-meson selection in the analyses, were trained with the reported hyperparameter configurations.

The BDT outputs are three scores related to the candidate probability to be a prompt $\mathrm{D_s^+}$ meson, a non-prompt $\mathrm{D_s^+}$ meson, and combinatorial background. The sum of these three scores for a particle candidate is unity by construction. In Fig. 5.3, the BDT-output distributions for prompt and non-prompt $\mathrm{D_s^+}$ mesons, and the combinatorial background, obtained for the test-set data in the $2 < p_{\mathrm{T}} < 4$ GeV/$c$ transverse-momentum region, are reported. As expected, real non-prompt $\mathrm{D_s^+}$ mesons have a high score related to the hypothesis of being non-prompt, and the same is true for background candidates looking at the corresponding score. Interestingly, the distribution of the prompt $\mathrm{D_s^+}$-meson score is not peaked at unity for real prompt $\mathrm{D_s^+}$ mesons. This is due to the implementation details of the multi-class classification in the `xgboost` library [217], in which a number of BDTs equal to the number of classes are trained, each one performing a one-vs-rest classification task, and the final outputs are the scores predicted by the different models appropriately normalised [243]. The score related to the prompt $\mathrm{D_s^+}$-meson hypothesis is obtained from a classification of prompt $\mathrm{D_s^+}$ (one) versus non-prompt $\mathrm{D_s^+}$ and combinatorial

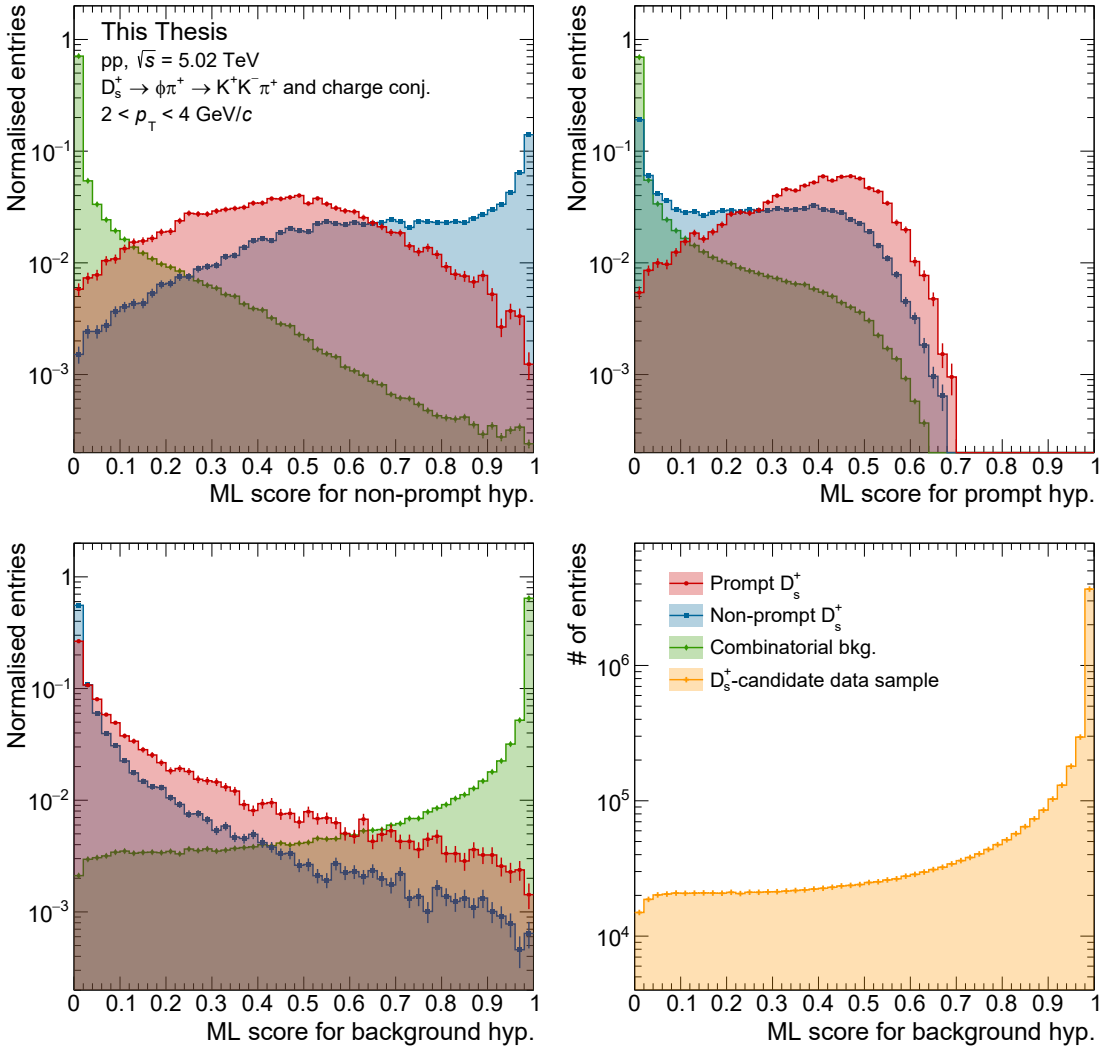



Figure 5.3: Top and bottom left panels: distributions of the three BDT-model outputs for prompt (red) and non-prompt (blue) $D_s^+$ mesons, and the combinatorial background (green) in the $2 < p_T < 4$ GeV/$c$ interval. Bottom right panel: distribution of the BDT output corresponding to the background hypothesis for the sample of $D_s^+$-meson candidates collected by the experiment with $2 < p_T < 4$ GeV/$c$ and invariant mass $1.72 < M(KK\pi) < 2.22$ GeV/$c^2$.

background (rest). This mix of background and non-prompt $D_s^+$-meson candidates, which have different characteristics, is more difficult to discriminate with respect to the other combinations. In any case, real prompt $D_s^+$ mesons have on average higher values of the corresponding score than non-prompt $D_s^+$ and combinatorial background, as expected.

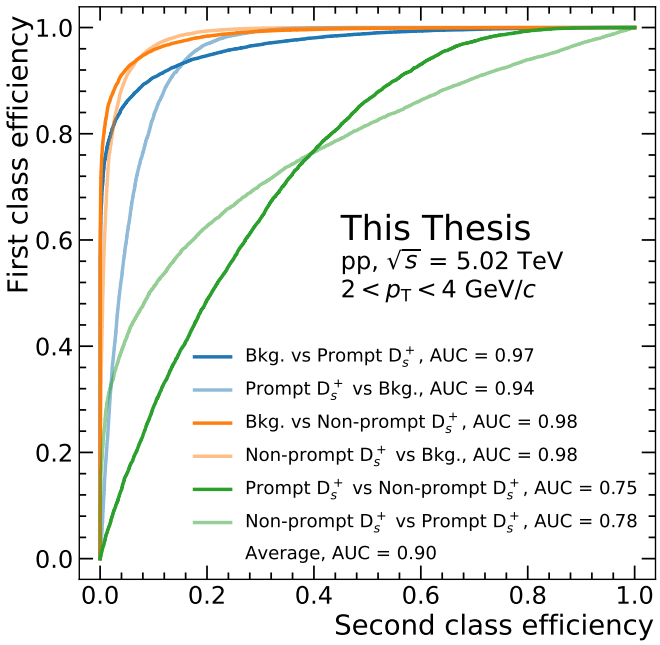



Figure 5.4: One-vs-one ROC curves and the average ROC AUC values, estimated using the test set, for the BDT model trained in the $2 < p_T < 4$ GeV/$c$ interval. Curves of the same colour are obtained swapping the same two classes.

In the bottom right panel of Fig. 5.3, the BDT-output distribution corresponding to the background hypothesis for the sample of $D_s^+$-meson candidates collected by the experiment in the $2 < p_T < 4$ GeV/$c$ interval and in the $1.72 < M(KK\pi) < 2.22$ GeV/$c^2$ invariant-mass region is displayed. The shape of the distribution matches almost exactly the one of the combinatorial background shown in the bottom left panel, due to the very small fraction of signal $D_s^+$-meson candidates in the collected data sample.

After the training, the BDT performance was evaluated estimating the average ROC AUC value on the test set. The one-vs-one ROC curves, their area-under-the-curve values, and the average ROC AUC value are shown in Fig. 5.4 for the BDT model trained in the $2 < p_T < 4$ GeV/$c$ interval. High ROC AUC values are obtained for the discrimination between signal $D_s^+$ mesons, prompt or non-prompt, and the combinatorial background. Lower values of the metric are achieved for the discrimination between prompt and non-prompt $D_s^+$ mesons. This is expected, since the two signal classes are more similar between each other than signal and background candidates. The ROC curves obtained swapping the same two classes are not equivalent between each other, since different BDT outputs were used to compute them, as explained in Section 3.1.1. The estimated average ROC AUC values for the trained BDTs range between about 0.87 and 0.90 depending on the $p_T$ interval, and are reported in Tab. 5.4. These values were compared to those

Table 5.4: Average ROC AUC values for the final BDT models in the different $p_{\mathrm{T}}$ intervals in which the training was performed.

| $p_{\mathrm{T}}$ (GeV/$c$) | 1–2 | 2–4 | 4–6 | 6–12 | 12–24 |
|---|---|---|---|---|---|
| Average ROC AUC | 0.874 | 0.900 | 0.901 | 0.899 | 0.875 |

resulting from the same calculation performed by using the training-set data, to check for a possible overfitting (see Section 3.2.1). The average ROC AUC values obtained from the training set are slightly larger than the test-set ones of about 2-7% depending on $p_{\mathrm{T}}$. This behaviour is foreseen, and the small discrepancies indicate a good generalisation power of the trained BDT models.

The BDT-model predictions were studied computing the SHAP values using the test-set data sample. The mean absolute SHAP values of the input features for the three classes are reported in the left panel of Fig. 5.5, for the BDT trained in the $2 < p_{\mathrm{T}} < 4$ GeV/$c$ interval. They measure the global importance of each variable for the predictions. The most important features are those with the higher mean absolute SHAP values, and correspond to: the variables related to the $\mathrm{D_s^+}$-meson decay length and PID information, $\Delta M_{\mathrm{KK}}$, and $\cos^3\theta'(\mathrm{K})$. This matches the physical

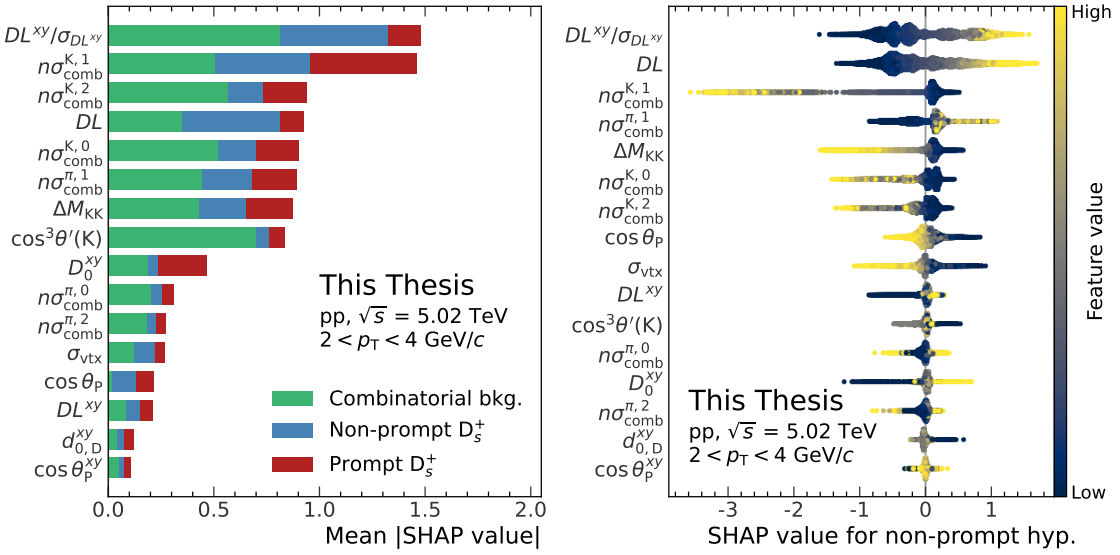



Figure 5.5: Feature-importance plots for the BDT model trained in the $2 < p_{\mathrm{T}} < 4$ GeV/$c$ interval. Left panel: mean absolute SHAP values of features for the prompt $\mathrm{D_s^+}$ (red), non-prompt $\mathrm{D_s^+}$ (blue), and combinatorial background (green) classes. Right panel: SHAP values of features for each point in the test set for the non-prompt $\mathrm{D_s^+}$-meson hypothesis (more details in the text).

knowledge of the problem and the differences in the variable distributions observed between prompt and non-prompt $\mathrm{D_s^+}$ mesons, and the combinatorial background, as shown in Figs. 4.4, 4.6, and 4.7. The most important variable among the PID ones is $n\sigma_{\mathrm{comb}}^{\mathrm{K},1}$. This is expected since the decay-track with index one is always a kaon for signal candidates by construction, because it is the track with opposite charge sign with respect to the candidate, while the other two tracks can be either a kaon or a pion. Finally, $\Delta M_{\mathrm{KK}}$ and $\cos^3\theta'(\mathrm{K})$ are more important for the background-class predictions than for the signal-class ones due to their almost identical distributions between prompt and non-prompt $\mathrm{D_s^+}$ mesons.

Another possibility offered by the `shap` library [221], to investigate the BDT predictions, is to look at the contribution of each variable of a particle candidate to the score assigned to it. In the right panel of Fig. 5.5, the SHAP-value distributions of the input features for the non-prompt $\mathrm{D_s^+}$-meson hypothesis are shown. The input-feature values are described by a heatmap ranging from blue (low values) to yellow (high values). Large SHAP values, reported on the x-axis of the panel, relate to a large BDT-output score and vice versa. For instance, particle candidates with high $DL^{xy}/\sigma_{DL^{xy}}$ values or low $n\sigma_{\mathrm{comb}}^{\mathrm{K},1}$ values are more probably predicted to be non-prompt $\mathrm{D_s^+}$ mesons, as expected.

### 5.1.3 Working-point optimisation

Threshold values on the BDT-output scores need to be chosen to define the $\mathrm{D_s^+}$-meson candidate selection. To this end, scans were performed varying the threshold values on the BDT outputs simultaneously. For each combination, the expected statistical significance $S/\sqrt{(S+B)}$, the signal-to-background ratio $S/B$, the acceptance-times-efficiency (Acc $\times$ $\epsilon$) factors for prompt and non-prompt $\mathrm{D_s^+}$ mesons, and the fraction of prompt $\mathrm{D_s^+}$ mesons were estimated. The aforementioned $S$ and $B$ quantities are the expected number of signal and background candidates, respectively, which are defined in the following.

The (Acc $\times$ $\epsilon$) factors were computed from samples of MC simulations not used in the BDT training. The expected fraction of prompt $\mathrm{D_s^+}$ mesons was estimated as

$$f_{\mathrm{prompt}} = \left[1 + \frac{(\mathrm{Acc}\times\epsilon)^{\mathrm{non\text{-}prompt}}\ \frac{\mathrm{d}^2\sigma}{\mathrm{d}p_{\mathrm{T}}\mathrm{d}y}\Big|_{\mathrm{non\text{-}prompt}}^{\mathrm{FONLL+PYTHIA\ 8}}}{(\mathrm{Acc}\times\epsilon)^{\mathrm{prompt}}\ \frac{\mathrm{d}^2\sigma}{\mathrm{d}p_{\mathrm{T}}\mathrm{d}y}\Big|_{\mathrm{prompt}}^{\mathrm{FONLL}}}\right]^{-1}, \tag{5.2}$$

where $\mathrm{d}^2\sigma/\mathrm{d}y\mathrm{d}p_{\mathrm{T}}$ is the $p_{\mathrm{T}}$-differential production cross section in the $|y| < 0.5$ rapidity interval obtained from FONLL calculations [72, 73]. In the case of prompt $\mathrm{D_s^+}$ mesons, since FONLL calculations are not directly available, the cross-section value was computed as done in Ref. [238] using the FONLL prediction for the charm-quark $p_{\mathrm{T}}$-differential cross section, the charm fragmentation functions from Ref. [244], and the fragmentation fractions $f(\mathrm{c} \to \mathrm{D_s^+})$ and $f(\mathrm{c} \to \mathrm{D_s^{*+}})$ from

Table 5.5: Fragmentation fractions of beauty quarks into beauty-hadron species in $Z \to b\bar{b}$ decays, and in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV [6].

| b-hadron | Fraction at Z (%) | Fraction at $p\bar{p}$ (%) |
|---|---|---|
| $B^0$, $B^+$ | $40.8 \pm 0.7$ | $34.4 \pm 2.1$ |
| $B_s^0$ | $10.0 \pm 0.8$ | $11.5 \pm 1.3$ |
| $\Lambda_b^0$ | $8.4 \pm 1.1$ | $19.8 \pm 4.6$ |

ALEPH measurements [245]. The non-prompt $D_s^+$-meson cross section was obtained from the $p_T$-differential beauty-hadron cross sections predicted by FONLL, while the description of the $H_b \to D_s^+ + X$ decay kinematics and branching ratios was performed with PYTHIA 8 [82, 168]. The fragmentation fractions of beauty quarks into beauty hadrons $f(b \to H_b)$ measured in $Z \to b\bar{b}$ decays [6], which are reported in Tab. 5.5, were used to weight the contributions from the different beauty-hadron species. These fragmentation fractions provide a satisfying normalisation for the B-meson measurements conducted by the ATLAS, CMS, and LHCb Collaborations [246–248]. The fraction of non-prompt $D_s^+$ mesons was estimated as $f_{\text{non-prompt}} = 1 - f_{\text{prompt}}$.

The expected signal $S$, containing both prompt and non-prompt candidates, was computed starting from theoretical predictions as

$$S = \frac{\mathrm{d}^2\sigma^{D_s^+}}{\mathrm{d}p_T \mathrm{d}y}\bigg|^{\text{FONLL, (non-)prompt}}_{|y|<0.5} \times \frac{2(\text{Acc} \times \epsilon)^{\text{(non-)prompt}} c_{\Delta y} \Delta p_T \text{BR} \mathcal{L}_{\text{int}}}{f_{\text{(non-)prompt}}}, \tag{5.3}$$

where the prompt (non-prompt) $D_s^+$-meson $(\text{Acc} \times \epsilon)$, fraction, and predicted cross section were employed in the case of the prompt (non-prompt) measurement. Real data was not utilised to avoid possible biases due to statistical fluctuations of the background under the $D_s^+$-meson peak.

The background $B$ was estimated by fitting the invariant-mass distribution of $D_s^+$-meson candidates obtained from data. An exponential function was used and only the region away from the peaks of the $D_s^+$ and $D^+$ mesons was considered. Then, the background in the signal region was computed integrating the fit function in the invariant-mass interval within $\pm 3\sigma$ from the $D_s^+$-meson world-average mass [6]. The width $\sigma$ of the $D_s^+$-meson signal peak was obtained from MC simulations, and it depends on the transverse momentum as will be shown in Fig. 5.9. To avoid correlations between statistical fluctuations in the estimated significance and the measurement, only about 10% of the available data was used in the computation. The obtained background value was thus scaled to match the full data sample collected by the experiment.

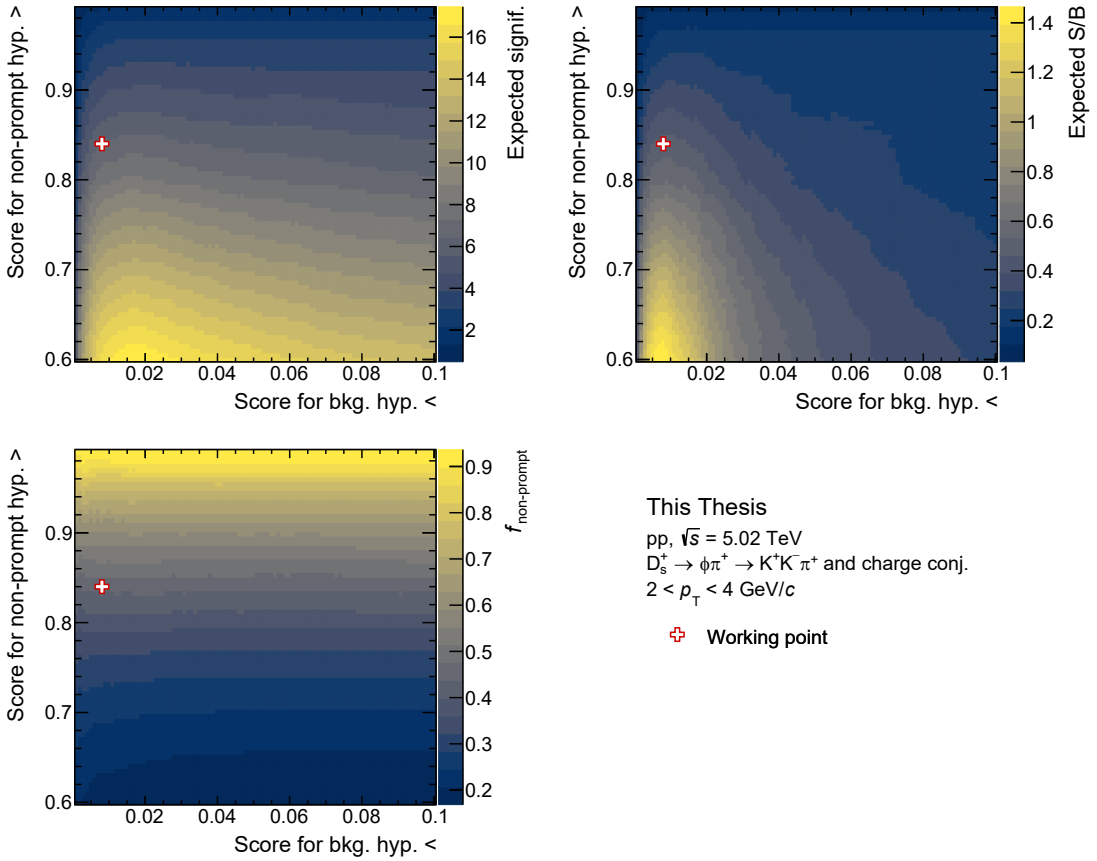



Figure 5.6: Expected significance (top left panel), S/B (top right panel) and $f_{\text{non-prompt}}$ fraction (bottom left panel) as a function of the thresholds on the BDT outputs related to the non-prompt and background hypotheses. The selected working point for the measurement of non-prompt $D_s^+$ mesons is shown by the red cross.

Since the three BDT scores sum to unity for a particle candidate, as mentioned in the previous Section, there are two degrees of freedom and the scan can be performed considering only two scores. In the case of the non-prompt $D_s^+$-meson measurement, which will be detailed in Section 5.2, the selection was performed using the scores related to the probability to be combinatorial background and non-prompt $D_s^+$ signal. Figure 5.6 reports an example of the scan results for the $2 < p_T < 4$ GeV/$c$ transverse-momentum interval. The expected significance, $S/B$, and $f_{\text{non-prompt}}$ values are shown in colour scale as a function of the thresholds on the background and non-prompt scores. The thresholds on the BDT outputs were tuned to obtain the highest possible fraction of non-prompt $D_s^+$ mesons in the extracted raw yields, while maintaining a good statistical significance of the signal. The chosen working point for the $2 < p_T < 4$ GeV/$c$ interval is shown in

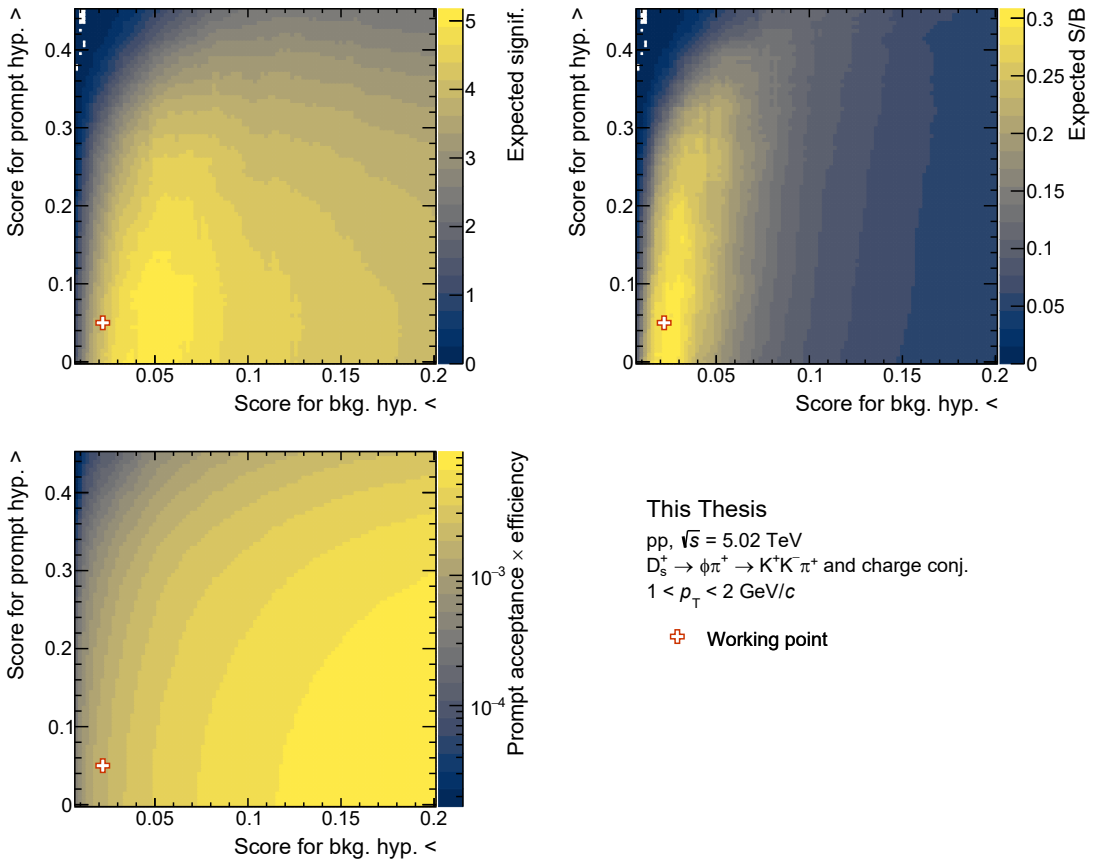



Figure 5.7: Expected significance (top left panel), S/B (top right panel) and prompt (Acc $\times$ $\epsilon$) (bottom left panel) as a function of the thresholds on the BDT outputs related to the prompt and background hypotheses. The selected working point for the measurement of prompt $\mathrm{D_s^+}$ mesons is shown by the orange cross.

the Figure by a red cross. The selections applied required a BDT score for the non-prompt $\mathrm{D_s^+}$-meson hypothesis above 0.70–0.84, and a score for the combinatorial background hypothesis below 0.008–0.040 depending on $p_{\mathrm{T}}$.

For the measurement of prompt $\mathrm{D_s^+}$ mesons, further discussed in Section 5.3, the BDT outputs corresponding to the prompt $\mathrm{D_s^+}$-meson and background hypotheses were considered. The threshold values were chosen to maximise the expected statistical significance of the extracted signal, while keeping a high selection efficiency and signal-to-background ratio to reduce the systematic uncertainties as much as possible. This resulted in requiring BDT scores for the prompt $\mathrm{D_s^+}$-meson hypothesis above 0.05–0.10, and scores for the combinatorial background hypothesis below 0.018–0.050 depending on $p_{\mathrm{T}}$. The loose selection on the prompt $\mathrm{D_s^+}$-meson score fulfils the purpose of reducing the contamination of non-prompt candidates in the measured raw yields. As an example, the results of the threshold scan performed

considering the scores linked to the probability to be combinatorial background and prompt $\mathrm{D_s^+}$ signal for the $1 < p_{\mathrm{T}} < 2$ GeV/$c$ interval are displayed in Fig. 5.7. The expected significance, $S/B$, and prompt (Acc$\times\epsilon$) values are reported in colour scale as a function of the threshold values, while the chosen working point is represented by the orange cross.

## 5.2 Non-prompt $\mathrm{D_s^+}$-meson analysis

### 5.2.1 Raw-yield extraction

A sample enriched in non-prompt $\mathrm{D_s^+}$-meson candidates, resulting from the selection introduced in the previous Section, was used to extract the raw yields for the measurement of the non-prompt $\mathrm{D_s^+}$-meson cross section. The raw yields were obtained from binned maximum-likelihood fits to the candidate invariant-mass $M(\mathrm{KK\pi})$ distributions, including both particles and antiparticles. Four different transverse-momentum intervals between 2 and 12 GeV/$c$ were considered. The fit function was constituted by a Gaussian to describe the signal peak and by an exponential term to model the background. An additional Gaussian was used to describe the peak caused by the decay $\mathrm{D^+ \to \phi\pi^+ \to K^-K^+\pi^+}$, which has a branching ratio of $(2.69^{+0.07}_{-0.08}) \times 10^{-3}$, present at lower invariant-mass values than the $\mathrm{D_s^+}$-meson peak, $M(\mathrm{D^+}) = (1869.65 \pm 0.05)$ MeV/$c^2$ [6]. To improve the fit stability, the widths of the $\mathrm{D_s^+}$-meson signal peak were fixed to the values resulting from a sample enriched in prompt candidates, which was obtained applying the prompt-measurement selection, given the naturally larger abundance of prompt $\mathrm{D_s^+}$ mesons with respect to non-prompt ones. The widths of the $\mathrm{D^+}$-meson peak $\sigma(\mathrm{D^+})$ were fixed to the $\mathrm{D_s^+}$-meson ones $\sigma(\mathrm{D_s^+})$ multiplied by their ratio $\sigma(\mathrm{D^+})^{\mathrm{MC}}/\sigma(\mathrm{D_s^+})^{\mathrm{MC}}$ computed from MC simulations. This assumed that the relative variation with $p_{\mathrm{T}}$ of the mass resolution of the $\mathrm{D_s^+}$- and $\mathrm{D^+}$-meson peaks is described in the simulations. The value of the $\sigma(\mathrm{D^+})^{\mathrm{MC}}/\sigma(\mathrm{D_s^+})^{\mathrm{MC}}$ ratio was found to be about 0.9 independently of the transverse-momentum interval considered.

The invariant-mass spectra are shown in Fig. 5.8, together with the fit functions and the resulting values for the mean, width, raw yield, and significance of the signal peak. A good statistical significance of the observed signal was obtained, it varies between about 7 and 9 depending on the $p_{\mathrm{T}}$ interval.

In the left panel of Fig. 5.9, the observed mean of the Gaussian function fitting the $\mathrm{D_s^+}$-meson peak is compared to the values resulting from fits in which the width was left free to vary, from MC simulations, and from the PDG. The values of the $\mathrm{D_s^+}$-meson peak position extracted from the fixed- and free-width fits are in good agreement. Moreover, they are compatible with the peak positions obtained from MC simulations within two standard deviations. The shift of the mass values observed in data from the MC-simulation and PDG ones is common to all the

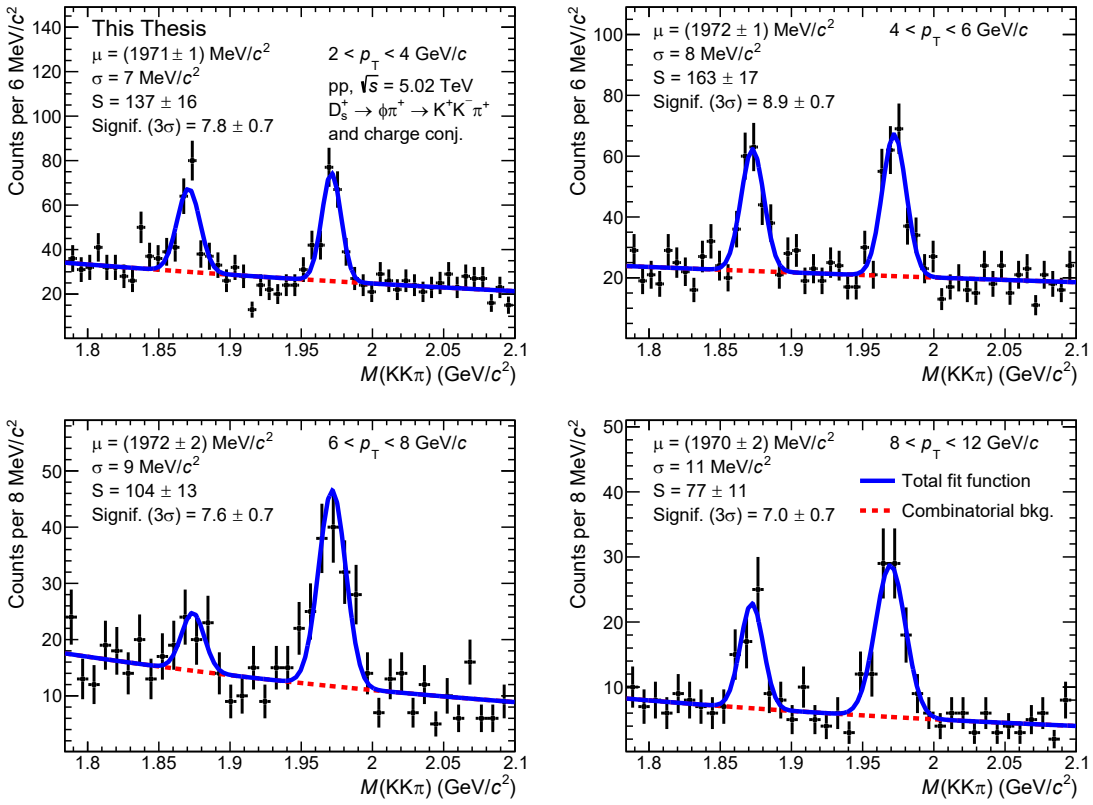



Figure 5.8: Invariant-mass distributions of $D_s^+$-meson candidates and their charge conjugates in the $p_T$ intervals of the non-prompt analysis. The blue line shows the total fit function and the red dashed line is the combinatorial-background component. The values of the mean ($\mu$), width ($\sigma$), raw yield (S), and significance (Signif.) of the signal peak are reported. The reported uncertainties are only the statistical uncertainties from the fit.

ALICE D-meson measurements in Run 2. The Gaussian width extracted fitting the prompt-enriched sample is confronted to those obtained from the free-width fit to the non-prompt-enriched sample, and from MC simulations in the right panel of Fig. 5.9. The $D_s^+$-meson width values are compatible between the three considered cases within uncertainties. Except for the $6 < p_T < 8$ GeV/$c$ interval, where the width resulting from the fit to the non-prompt-enriched sample deviates from the prompt-enriched and MC-simulation values due to statistical fluctuations. The $D_s^+$-meson peak width shows an increasing trend with the transverse momentum due to the $p_T$ resolution of the decay tracks.

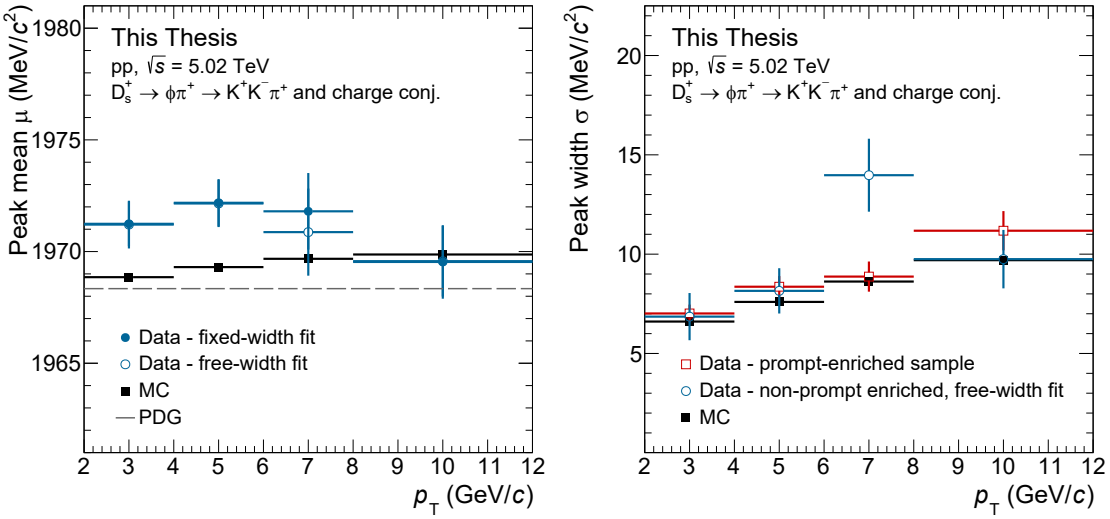



Figure 5.9: Left panel: $D_s^+$-meson signal peak position obtained from the fixed-width fit (blue filled circles) compared to the free-width fit (blue open circles), MC simulation (black squares), and the PDG [6] values. Right panel: $D_s^+$-meson signal peak width from the fit to the prompt-enriched sample (red open squares), the free-width fit to the non-prompt-enriched sample (blue open circles), and MC simulations (black squares).

### 5.2.2 Efficiency and acceptance corrections

The acceptance-times-efficiency (Acc$\times\epsilon$) correction factors were computed from MC simulations as described in Section 4.3.1. Different samples than those employed in the BDT training were used to avoid possible biases caused by the ML-model memorisation of the training-data peculiarities. The (Acc $\times$ $\epsilon$) correction factors for prompt and non-prompt $D_s^+$ mesons as a function of $p_T$ are reported in Fig. 5.10, together with the ratios of the non-prompt over prompt correction factors. The BDT selection applied to obtain the non-prompt-enriched sample strongly suppresses the prompt $D_s^+$-meson selection efficiency, while the geometrical acceptance is the same between prompt and non-prompt $D_s^+$ mesons. The non-prompt $D_s^+$-meson acceptance times efficiency is larger than the one of prompt $D_s^+$ mesons by a factor ranging from about 5 to 17, depending on transverse momentum. The ratio $(\text{Acc}\times\epsilon)^{\text{non-prompt}}/(\text{Acc}\times\epsilon)^{\text{prompt}}$ is larger at low $p_T$ and decreases with increasing transverse momentum, since at high $p_T$ looser selections were applied due to the smaller $D_s^+$-meson yield to guarantee a good signal extraction. This leads to a smaller difference between the prompt and non-prompt (Acc$\times\epsilon$) correction factors with increasing $p_T$.

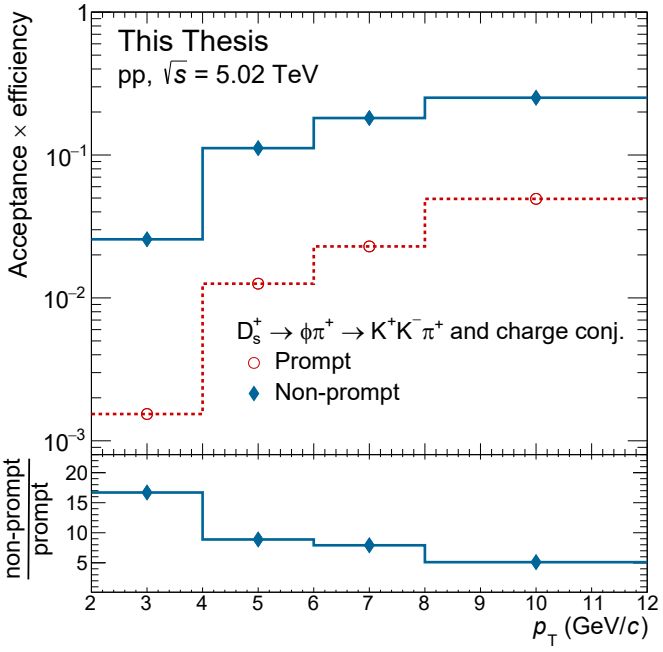



Figure 5.10: Acceptance-times-efficiency correction factors for non-prompt (blue) and prompt (red) $D_s^+$ mesons as a function of $p_T$. The bottom panel shows the non-prompt over prompt (Acc × $\epsilon$) ratio.

### 5.2.3 Fraction of non-prompt $D_s^+$ mesons

The measured raw yields, although dominated by non-prompt $D_s^+$ mesons, still contain a residual contribution of prompt candidates which satisfy the BDT-based selection. For each $p_T$ interval of the analysis, the fraction of non-prompt $D_s^+$ mesons in the raw yield, $f_{\text{non-prompt}}$, was estimated with a data-driven approach, which is described in the following.

A set of different selections $\{i\}$ was defined, each one leading to a raw yield with different contributions of prompt and non-prompt $D_s^+$ mesons. These selections were obtained varying the threshold on the BDT output related to the candidate probability of being a non-prompt $D_s^+$ meson. Instead, the threshold on the BDT output for the combinatorial-background hypothesis was fixed to the value used in the central selection detailed in Section 5.1.3. The resulting raw yields $Y_i$ can be connected to the *true* yields of prompt $N_{\text{prompt}}$ and non-prompt $N_{\text{non-prompt}}$ $D_s^+$ mesons in the data sample as follows

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

where $(\text{Acc} \times \epsilon)_i^{\text{prompt}}$ and $(\text{Acc} \times \epsilon)_i^{\text{non-prompt}}$ are the acceptance-times-efficiency factors computed for prompt and non-prompt $D_s^+$ mesons, respectively. The definition

of a set of $n$ selections leads to the system of equations

$$\begin{cases} (\mathrm{Acc} \times \epsilon)_1^{\text{prompt}} N_{\text{prompt}} + (\mathrm{Acc} \times \epsilon)_1^{\text{non-prompt}} N_{\text{non-prompt}} = Y_1 \\ \dots \\ (\mathrm{Acc} \times \epsilon)_n^{\text{prompt}} N_{\text{prompt}} + (\mathrm{Acc} \times \epsilon)_n^{\text{non-prompt}} N_{\text{non-prompt}} = Y_n \end{cases}, \tag{5.5}$$

which can be exactly solved in the case of $n = 2$ to obtain the prompt and non-prompt true yields. For $n > 2$, the system is likely to be overdetermined and the $N_{\text{prompt}}$ and $N_{\text{non-prompt}}$ parameters are obtained from a minimisation procedure. To this end, Eq. (5.5) can be rewritten as

$$\begin{pmatrix} (\mathrm{Acc} \times \epsilon)_1^{\text{prompt}} & (\mathrm{Acc} \times \epsilon)_1^{\text{non-prompt}} \\ \vdots & \vdots \\ (\mathrm{Acc} \times \epsilon)_n^{\text{prompt}} & (\mathrm{Acc} \times \epsilon)_n^{\text{non-prompt}} \end{pmatrix} \times \begin{pmatrix} N_{\text{prompt}} \\ N_{\text{non-prompt}} \end{pmatrix} - \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_n \end{pmatrix}, \tag{5.6}$$

where $\delta_i$ is a residuum that accounts for the equations not holding exactly due to the fluctuations of $Y_i$, $(\mathrm{Acc} \times \epsilon)_i^{\text{prompt}}$, and $(\mathrm{Acc} \times \epsilon)_i^{\text{non-prompt}}$.

The prompt and non-prompt corrected yields were estimated minimising the $\chi^2$, which is defined as

$$\chi^2 = \boldsymbol{\delta}^T \boldsymbol{C}^{-1} \boldsymbol{\delta}. \tag{5.7}$$

In the above formula, $\boldsymbol{\delta}$ is the vector of residuals and $\boldsymbol{C}$ is the covariance matrix which contains the uncertainties inherent to each equation. The variances $\sigma_i^2$ were computed from the raw-yield and efficiency statistical uncertainties as

$$\sigma_i^2 = \sigma_{Y_i}^2 + N_{\text{prompt}} \sigma^2_{(\mathrm{Acc}\times\epsilon)_i^{\text{prompt}}} + N_{\text{non-prompt}} \sigma^2_{(\mathrm{Acc}\times\epsilon)_i^{\text{non-prompt}}}. \tag{5.8}$$

Since the true yields are unknown variables, the variances were evaluated with an iterative procedure. In the first iteration only the uncertainty on the raw yields was considered, while from the second iteration onward the values of $N_{\text{prompt}}$ and $N_{\text{non-prompt}}$ resulting from the previous step were used to compute the uncertainty contribution deriving from the efficiency-times-acceptance factors. Only a few iterations are necessary to obtain a stable result. The covariance terms were computed as $\sigma_{i,j} = \sigma_i \sigma_j \rho_{i,j}$ and the correlation coefficient $\rho_{i,j}$ was assumed to be

$$\rho_{i,j} = \frac{\sigma_i}{\sigma_j} \quad \text{with } i \subset j. \tag{5.9}$$

This assumption is motivated by the fact that the threshold on the BDT output was varied monotonically, making the selection more stringent with increasing $i$. Therefore, the $i^{\text{th}}$ selected sample is completely included in the $(i-1)^{\text{th}}$ one by construction. Finally, the prompt and non-prompt $\mathrm{D_s^+}$-meson corrected yields were calculated as

$$\boldsymbol{N} = \boldsymbol{Cov}(\boldsymbol{N}) \boldsymbol{\epsilon}^T \boldsymbol{C}^{-1} \boldsymbol{Y}. \tag{5.10}$$

In the equation, $\boldsymbol{N}$ and $\boldsymbol{Y}$ are the vectors of corrected and raw yields respectively, $\boldsymbol{\epsilon}$ is the matrix of efficiency-times-acceptance factors, and $\boldsymbol{Cov(N)} = (\boldsymbol{\epsilon}^T \boldsymbol{C}^{-1} \boldsymbol{\epsilon})^{-1}$ is the covariance matrix containing the statistical uncertainties on the corrected yields and their correlation.

For an arbitrary selection $k$, the fraction of non-prompt $\mathrm{D_s^+}$ mesons in the raw yield can be computed from the corrected yields resulting from the $\chi^2$ minimisation as

$$f^k_{\text{non-prompt}} = \frac{(\text{Acc} \times \epsilon)^{\text{non-prompt}}_k N_{\text{non-prompt}}}{(\text{Acc} \times \epsilon)^{\text{non-prompt}}_k N_{\text{non-prompt}} + (\text{Acc} \times \epsilon)^{\text{prompt}}_k N_{\text{prompt}}}. \tag{5.11}$$

The final values of the non-prompt $\mathrm{D_s^+}$-meson cross section were not determined from the $N_{\text{non-prompt}}$ parameter resulting from the minimisation of the $\chi^2$ in Eq. (5.7). Instead, they were obtained by choosing a selection providing a large non-prompt component and a good signal extraction, as described in Section 5.1.3, and by calculating the corresponding $f_{\text{non-prompt}}$ fraction according to Eq. (5.11). This approach facilitates the determination of the systematic uncertainty. Moreover, it is not necessary that the considered central selection is one of those used in the data-driven procedure.

Figure 5.11 shows an example of the minimisation-procedure ingredients and results for the $2 < p_{\mathrm{T}} < 4$ transverse-momentum region. In the top left panel of the Figure, the acceptance-times-efficiency factors for prompt and non-prompt $\mathrm{D_s^+}$ mesons as a function of the BDT-based selection are reported. The leftmost data point corresponds to the loosest selection on the BDT output related to the candidate probability of being a non-prompt $\mathrm{D_s^+}$ meson. The rightmost one corresponds to the strictest selection, which is expected to preferentially select non-prompt $\mathrm{D_s^+}$ mesons. In the top right panel, the correlation coefficients $\rho$, defined in Eq. (5.9), between the samples obtained with the different selections are displayed. Their values range between 0.5 and 1. The raw-yield distribution as a function of the BDT-based selection is shown in the bottom left panel of Fig. 5.11. It is compared to the sum of the prompt and non-prompt components obtained from the minimisation procedure as $(\text{Acc} \times \epsilon)^{\text{prompt}}_i N_{\text{prompt}}$ and $(\text{Acc} \times \epsilon)^{\text{non-prompt}}_i N_{\text{non-prompt}}$, which are represented by the stacked red and blue histograms, respectively. Finally, the estimated fraction of prompt and non-prompt $\mathrm{D_s^+}$ mesons for the different selections is reported in the bottom right panel. As expected, stricter selections result in higher fractions of non-prompt candidates in the extracted raw yields. For completeness, the raw-yield distributions as a function of the BDT-based selection for the other transverse-momentum intervals of the analysis are reported in Fig. 5.12. In all the cases, they are compatible within uncertainties with the sum of the estimated prompt and non-prompt components, signalling that the minimisation procedure provided a satisfactory result.

The $\mathrm{D_s^+}$-meson $f_{\text{non-prompt}}$ fraction, corresponding to the data-sample enriched in non-prompt candidates used for the final measurement, is shown as a function

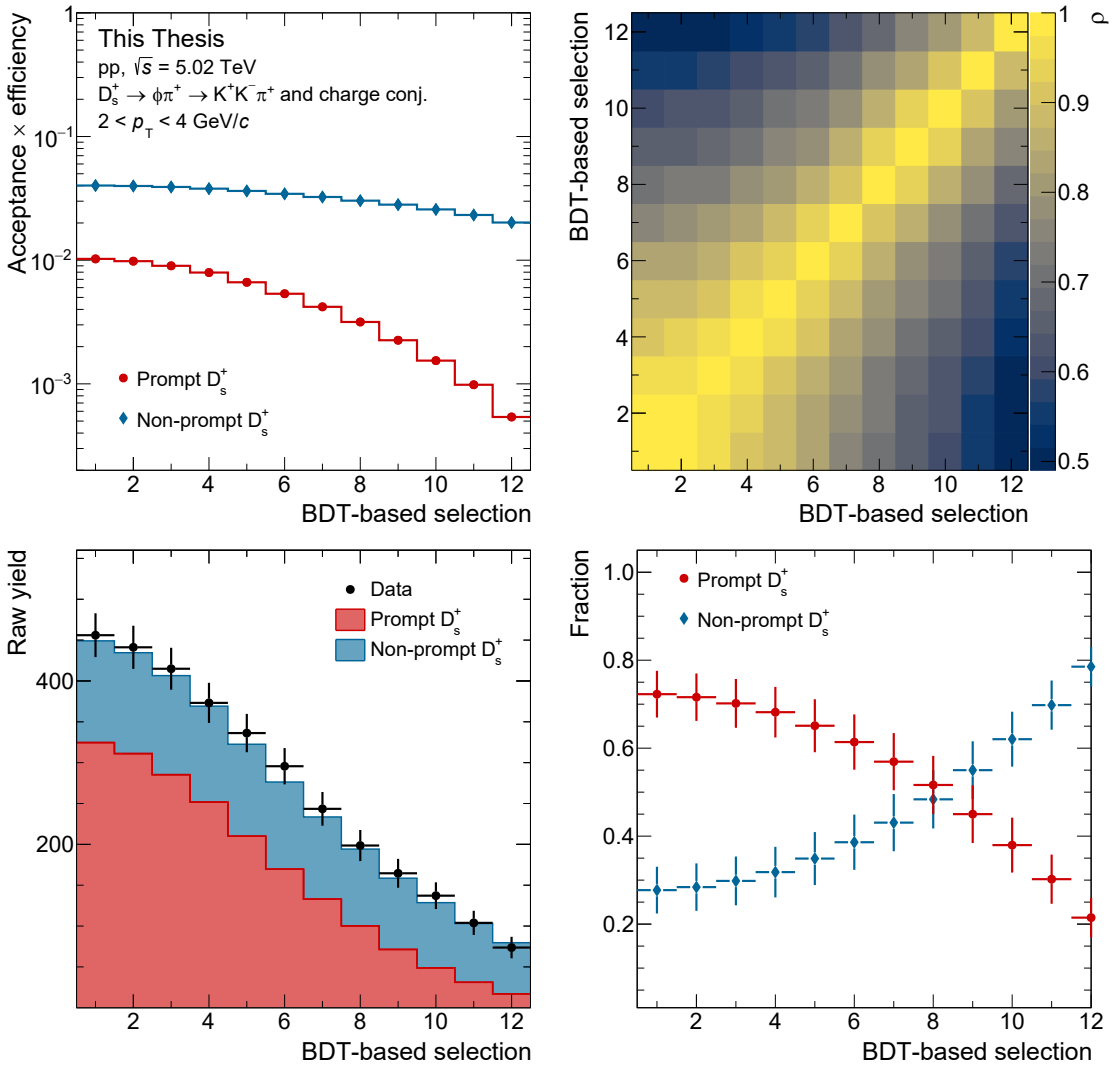



Figure 5.11: Top left panel: (Acc × $\epsilon$) factors for prompt (red) and non-prompt (blue) $D_s^+$ mesons for the BDT-based selections used in the data-driven $f_{\text{non-prompt}}$ fraction estimation in the $2 < p_T < 4$ GeV/$c$ interval. Top right panel: correlation coefficients between the samples obtained with the different selections. Bottom left panel: raw-yield distribution as a function of the selection. Bottom right panel: estimated fraction of prompt and non-prompt $D_s^+$ mesons as a function of the selection.

of $p_T$ in the bottom right panel of Fig. 5.12. The vertical bars report the statistical uncertainties computed propagating the uncertainties on the corrected yields resulting from the $\chi^2$-minimisation procedure. The correlation between $N_{\text{prompt}}$ and $N_{\text{non-prompt}}$ is accounted for in the uncertainty propagation. The systematic uncertainties, which are described hereafter, are represented by the open boxes. The

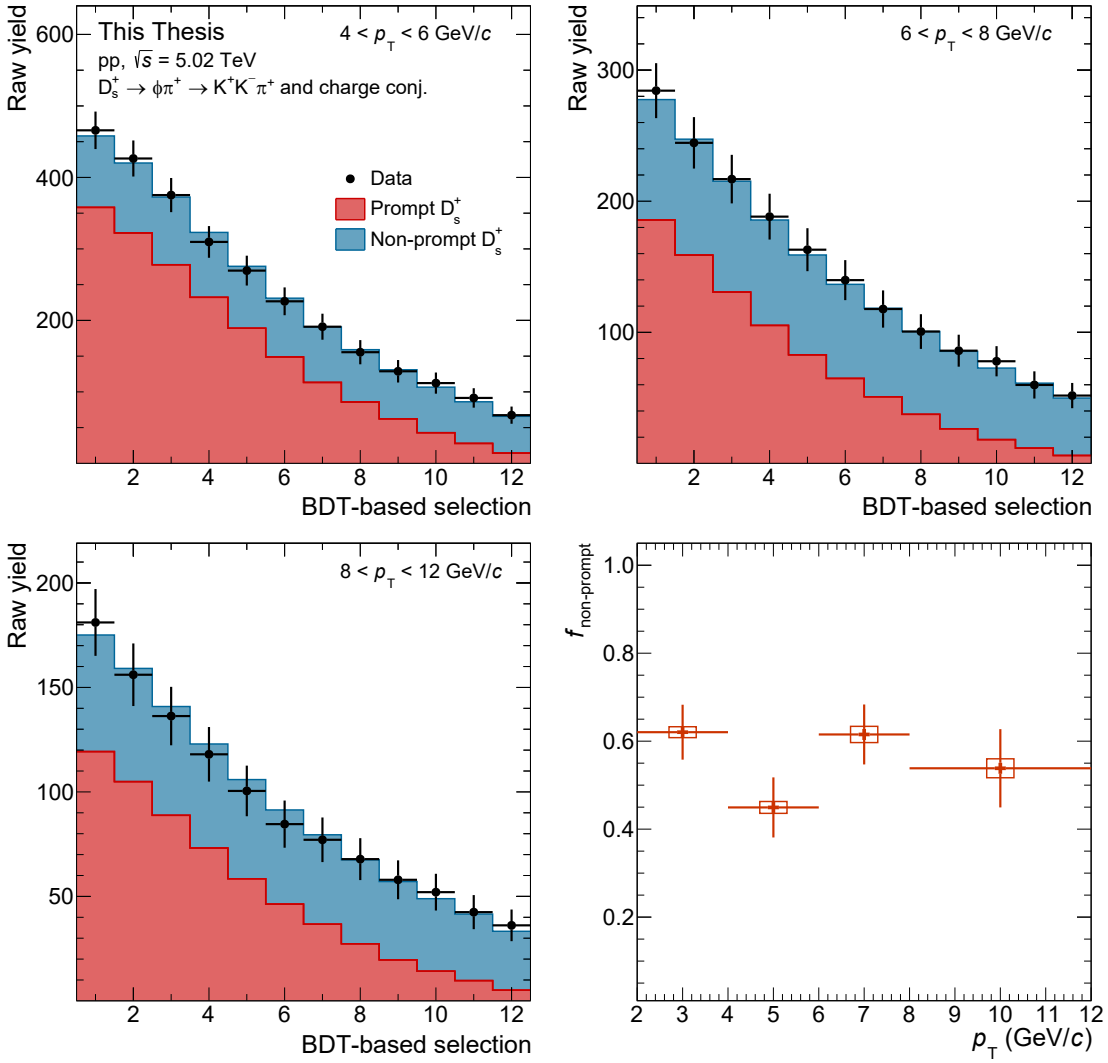



Figure 5.12: Top and bottom left panels: raw-yield distribution as a function of the BDT-based selection compared to the sum of the estimated prompt (red) and non-prompt (blue) components for the $p_T$ intervals in the $4 < p_T < 12$ GeV/$c$ region. Bottom right panel: fraction of non-prompt $D_s^+$ mesons in the extracted raw yields as a function of $p_T$ in pp collisions at $\sqrt{s} = 5.02$ TeV. The vertical bars (open boxes) report the statistical (systematic) uncertainties.

$f_{\text{non-prompt}}$ fraction spans the interval 0.45–0.65, and does not show a trend with the transverse-momentum.

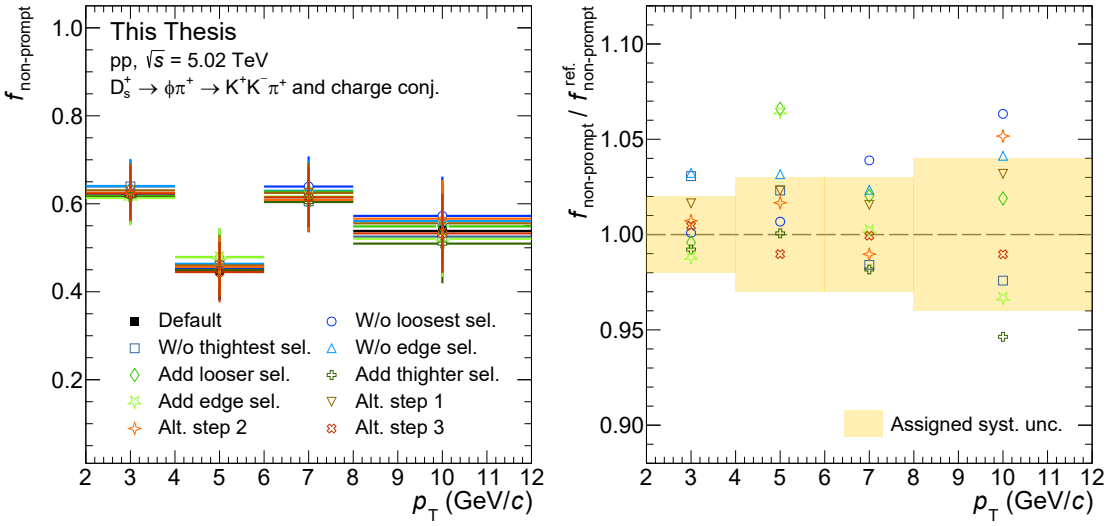



Figure 5.13: Left panel: $f_{\text{non-prompt}}$ fractions as a function of $p_{\text{T}}$ resulting from the data-driven estimation procedure with different set of selections. Right panel: ratio of $f_{\text{non-prompt}}$ values as a function of $p_{\text{T}}$ obtained from the various procedure configurations with respect to the reference value. The orange band reports the assigned systematic uncertainty.

### Systematic uncertainty

The systematic uncertainty on the measured non-prompt $D_s^+$-meson fraction was estimated by repeating the data-driven procedure described above with various set of selections. These selection sets were obtained removing the loosest (tightest) selection from the set used in the reference-value computation, adding a looser (tighter) selection, and changing the step with which the candidate probability of being a non-prompt $D_s^+$-meson is sampled. In the left panel of Fig. 5.13, the values of $f_{\text{non-prompt}}$ fraction resulting from the different selection sets are reported for each $p_{\text{T}}$ interval and compared with those deriving from the default configuration. In the right panel, the $f_{\text{non-prompt}}$ ratios with respect to the reference value for the different configurations are displayed as a function of $p_{\text{T}}$. The systematic uncertainty was assigned considering the RMS and the shift of the $f_{\text{non-prompt}}$-ratio distribution. The values assigned range from 2% to 4% depending on the transverse-momentum interval, and are shown in the plot by the orange band.

In the data-driven approach, the raw yields and efficiencies are extracted from the same data and MC-simulation samples used to obtain the final cross-section measurement, and employing the same procedures. Therefore, the systematic uncertainty on the $f_{\text{non-prompt}}$ fraction could be correlated to the systematic uncertainties due to the raw-yield extraction and selection-efficiency determination. This

possibility was investigated in detail, and more information is reported in Sections 5.4.1 and 5.4.2.

## 5.3 Prompt $D_s^+$-meson analysis

### 5.3.1 Raw-yield extraction

The prompt $D_s^+$-meson raw yields were measured in nine different $p_T$ intervals spanning the $1 < p_T < 24$ GeV/$c$ transverse-momentum region. This extended the measurement to lower $p_T$ with respect to the previously published result [238], which was based on linear selections of the $D_s^+$-meson candidates. A data sample enriched in prompt $D_s^+$ mesons, which was obtained applying the selection introduced in Section 5.1.3, was employed to extract the raw yields. The fit strategy adopted is similar to the one used for the non-prompt $D_s^+$-meson measurement detailed in Section 5.2.1. The only difference is that the width of the $D_s^+$-meson signal peak was left free in the fits.

Figure 5.14 reports the invariant-mass distributions for the various $p_T$ intervals, together with the fit functions and the resulting values for the mean, width, raw yield, and significance of the signal peak. The statistical significance of the observed signals varies from about 4 to 14 depending on $p_T$, while the $S/B$ values are comprehended between 0.31 and 3.1. The adopted BDT-based selections show a performance improvement with respect to the previous study [238]. A detailed comparison will be presented in Section 5.3.4.

### 5.3.2 Efficiency and acceptance corrections

Similarly to the non-prompt $D_s^+$-meson measurement, the acceptance-times-efficiency factors were estimated from MC simulations with the PYTHIA 8 event generator [82, 168], employing different samples than those used in the BDT training. The acceptance-times-efficiency factors for prompt and non-prompt $D_s^+$ mesons as a function of transverse momentum are shown in the left panel of Fig. 5.15. The prompt $D_s^+$-meson efficiency is higher than the non-prompt one for $p_T > 8$ GeV/$c$ due to the BDT-based selection applied, which preferentially selects prompt $D_s^+$ mesons while rejecting the more displaced non-prompt $D_s^+$ mesons.

The (Acc $\times$ $\epsilon$) factors obtained with the PYTHIA 8 event generator are compared to those resulting from MC simulations employing the PYTHIA 6.425 event generator [82] with the Perugia-11 tune [249]. In both cases, the same BDT models and thresholds on the BDT outputs were used. The PYTHIA 8 over PYTHIA 6 acceptance-times-efficiency ratios for prompt and non-prompt $D_s^+$ mesons are reported in the right panel of Fig. 5.15. For prompt $D_s^+$ mesons, the efficiency estimated employing PYTHIA 8 is systematically higher than the PYTHIA 6 one of about 3%. For non-prompt $D_s^+$ mesons, the ratio is significantly larger than unity

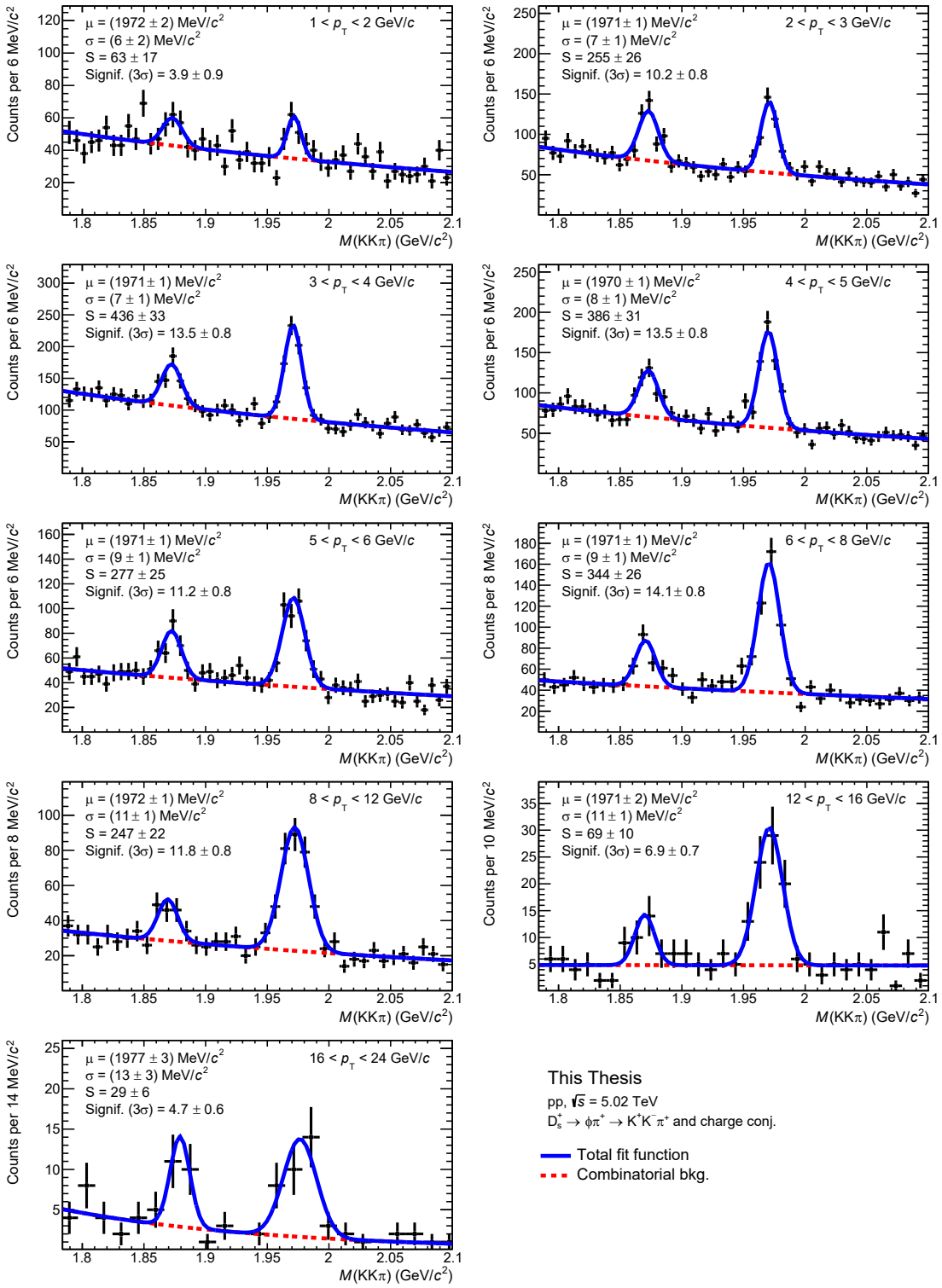



Figure 5.14: Invariant-mass distributions of $D_s^+$-meson candidates and charge conjugates in the $p_T$ intervals of the prompt analysis, together with the fit results.

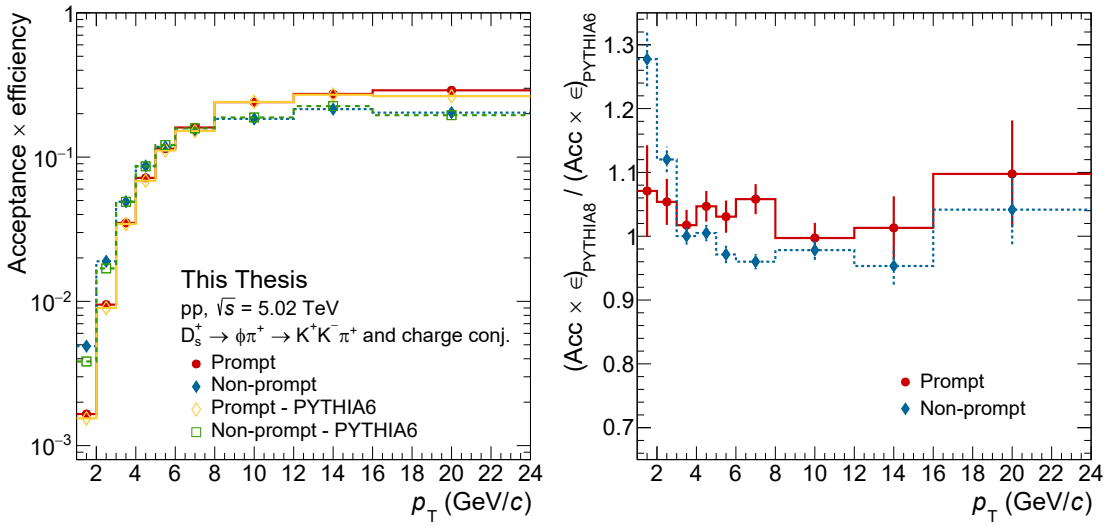



Figure 5.15: Left panel: acceptance-times-efficiency factors of prompt and non-prompt $D_s^+$ mesons as a function of $p_T$ obtained from MC simulations with the PYTHIA 6 (open markers) and PYTHIA 8 (filled markers) event generators. Right panel: acceptance-times-efficiency factor ratios between the two PYTHIA versions for prompt (red) and non-prompt (blue) $D_s^+$ mesons.

for $p_T < 3$ GeV/$c$ and slightly below unity above. These discrepancies can be explained by the different values of the mean proper decay length $c\tau$ of heavy-flavour hadrons implemented in the two versions of the PYTHIA event generator.

In Tab. 5.6, the $c\tau$ of $D^0$, $D^+$, and $D_s^+$ mesons, and beauty hadrons used by the PYTHIA 6 and PYTHIA 8 event generators are compared to the PDG values [6]. The PYTHIA 8 values are more similar to those reported by the PDG. In the PYTHIA 8 generator, the $D_s^+$-meson mean proper decay length is larger by about 7% than the one in PYTHIA 6, explaining the observed difference in the prompt acceptance times efficiency. The discrepancies in the non-prompt (Acc × $\epsilon$) arise from the combination of the larger proper lifetimes of $D_s^+$ and $B^+$ mesons, and of the smaller mean proper decay lengths of $B_s^0$ and $B^0$ mesons in PYTHIA 8 compared to PYTHIA 6. Moreover, the beauty-hadron decay kinematics, which depends on the decay tables implemented in the two versions of the event generator, also plays a role.

### 5.3.3 Fraction of prompt $D_s^+$ mesons

The fraction of prompt $D_s^+$ mesons, $f_{\text{prompt}}$, present in the measured raw yields was not estimated with the data-driven approach introduced in Section 5.2.3. This method requires the reliable extraction of raw yields with different contributions of prompt and non-prompt candidates, which was not possible in all the $p_T$ intervals

Table 5.6: Comparison of the PYTHIA 6.425[1], PYTHIA 8.243[2], and PDG [6] values for the mean proper decay length of $\mathrm{D^0}$, $\mathrm{D^+}$, and $\mathrm{D_s^+}$ mesons, and beauty hadrons. (*) In the ALICE MC productions using PYTHIA 8, the $\Lambda_\mathrm{b}^0$ mean proper decay length was set to the PDG value.

| Particle | $c\tau$ (μm) | | |
|---|---|---|---|
| | PYTHIA 6.425 | PYTHIA 8.243 | PDG 2020 |
| $\mathrm{D^0}$ | 124.4 | 122.9 | $122.9 \pm 0.4$ |
| $\mathrm{D^+}$ | 317 | 312 | $312 \pm 2$ |
| $\mathrm{D_s^+}$ | 140 | 150 | $151 \pm 1$ |
| $\mathrm{B^+}$ | 462 | 491 | $491 \pm 1$ |
| $\mathrm{B^0}$ | 468 | 459 | $455 \pm 1$ |
| $\mathrm{B_s^0}$ | 483 | 439 | $458 \pm 3$ |
| $\Lambda_\mathrm{b}^0$ | 342 | 441* | $441 \pm 3$ |

of the prompt $\mathrm{D_s^+}$-meson measurement due to the small size of the analysed data sample. Instead, the $f_\mathrm{prompt}$ fraction was computed using an approach based on FONLL calculations, similarly to what was done in Ref. [238], as

$$f_\mathrm{prompt} = 1 - \frac{N^{\mathrm{D_s^\pm,raw}}_\mathrm{non\text{-}prompt}}{N^{\mathrm{D_s^\pm,raw}}} = 1 - \left.\frac{\mathrm{d}^2\sigma}{\mathrm{d}p_\mathrm{T}\mathrm{d}y}\right|^{\mathrm{FONLL+PYTHIA\ 8}}_\mathrm{non\text{-}prompt} \times \times \frac{2(\mathrm{Acc}\times\epsilon)^\mathrm{non\text{-}prompt} c_{\Delta y} \Delta p_\mathrm{T} \mathrm{BR} \mathcal{L}_\mathrm{int}}{N^{\mathrm{D_s^\pm,raw}}}, \tag{5.12}$$

where $N^{\mathrm{D_s^\pm,raw}}$ is the measured raw yield and $N^{\mathrm{D_s^\pm,raw}}_\mathrm{non\text{-}prompt}$ is the expected raw yield of non-prompt $\mathrm{D_s^+}$ mesons. This last ingredient was calculated from the acceptance-time-efficiency factor of non-prompt $\mathrm{D_s^+}$ mesons extracted from MC simulations, and the theoretical prediction for the non-prompt $\mathrm{D_s^+}$-meson production cross section. The non-prompt $\mathrm{D_s^+}$-meson production was estimated using the beauty-hadron production cross sections from FONLL calculations, and the $\mathrm{H_b} \to \mathrm{D_s^+} + \mathrm{X}$ decay kinematics and branching ratios from the PYTHIA 8 decayer, as explained in Section 5.1.3.

[1] https://github.com/alisw/AliRoot/blob/f4cbc687b98df260c0dd252f31769071a4dd7244/PYTHIA6/AliPythia6/AliDecayerPythia.cxx#L1138

[2] https://github.com/alisw/AliRoot/blob/master/PYTHIA8/pythia8243/share/Pythia8/htmldoc/ParticleData.html

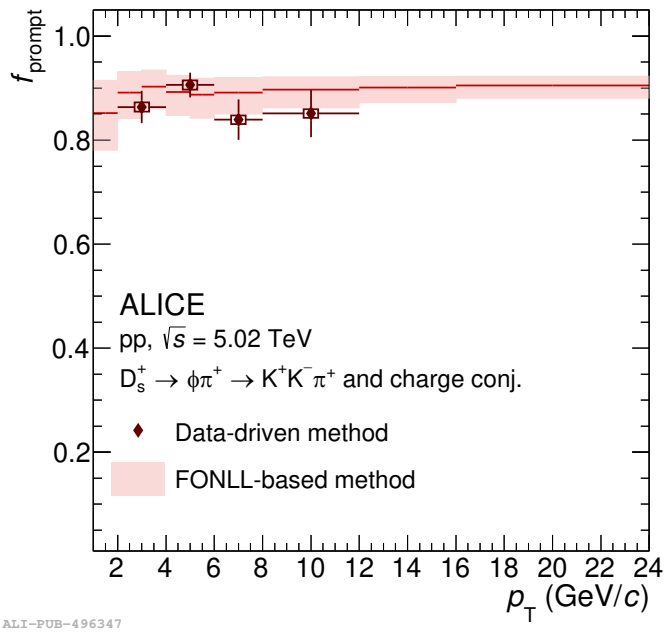



Figure 5.16: $D_s^+$-meson $f_{prompt}$ obtained with the FONLL-based approach compared with the data-driven one as a function of $p_T$. The results from the data-driven method are shown as diamond markers with the error bars (boxes) representing the statistical (systematic) uncertainty. For the FONLL-based approach, the continuous lines show the $f_{prompt}$ central values and the uncertainty is represented by the shaded boxes. Figure from Ref. [71].

The $D_s^+$-meson $f_{prompt}$ values resulting from the FONLL-based method range between 0.86 and 0.91 depending on the $p_T$ interval. They are reported in Fig. 5.16, and are compared with the $f_{prompt}$ fractions obtained with the data-driven method in the $p_T$ range of the non-prompt $D_s^+$-meson measurement, where a good reliability of the method can be guaranteed. The fractions of prompt $D_s^+$-meson yield estimated with the two different strategies are in good agreement within the statistical and systematic uncertainties in the common $p_T$ intervals. The evaluation of the $f_{prompt}$-fraction systematic uncertainty for the FONLL-based method is discussed hereafter.

### Systematic uncertainty

The systematic uncertainty on the $f_{prompt}$ fraction was estimated by varying the FONLL parameters as suggested in Ref. [60]. These parameters are the heavy-quark masses, and the factorisation $\mu_F$ and renormalisation $\mu_R$ scales. The value of the beauty-quark mass was varied between 4.5 and 5 GeV/$c^2$. The factorisation and renormalisation scales were varied in the range $0.5 < \mu_{F,R}/\mu_0 < 2$, where $\mu_0 =$

$\sqrt{p_\mathrm{T}^2 + m_\mathrm{b}^2}$ is the default value of the scales used in the FONLL central predictions. The constraint of a $\mu_\mathrm{F}/\mu_\mathrm{R}$ ratio comprehended between 0.5 and 2 was additionally imposed. The uncertainty on the parton-distribution functions employed in the FONLL calculations, which are the CTEQ6.6 PDFs [63], was also considered.

The relative values of the $f_\mathrm{prompt}$-fraction systematic uncertainty range between $^{+7.4}_{-8.4}\%$ and $^{+2.0}_{-2.7}\%$, decreasing in magnitude with increasing transverse momentum.

### 5.3.4 Comparison with linear-selection approach

The performance of the BDT-based selection was compared to the one obtained using the traditional linear selections. To this end, the prompt $\mathrm{D_s^+}$-meson measurement published in Ref. [238], which employed the same data sample of pp collisions at $\sqrt{s} = 5.02$ TeV, was considered. In the left panel of Fig. 5.17, the acceptance-times-efficiency factors computed from MC simulations with the PYTHIA 6 event generator for the BDT-based and linear selection strategies are reported, together with their ratios in the common $p_\mathrm{T}$ intervals of the two measurements. The (Acc $\times$ $\epsilon$) factor for prompt (non-prompt) $\mathrm{D_s^+}$ mesons obtained with the BDT-based selection, developed in this Thesis, is 1.5–2.2 (0.8–1.9) times larger than the one deriving from the previous-study linear selections, depending on $p_\mathrm{T}$. This larger prompt efficiency, which derives from the better separation of signal from background provided by the BDT selection, implies smaller correction factors

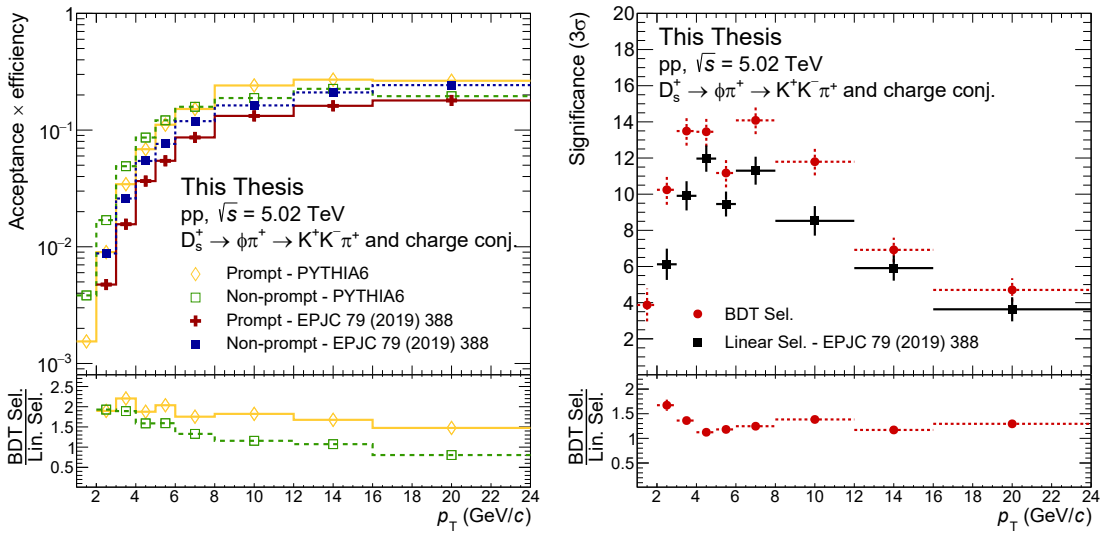



Figure 5.17: Left panel: acceptance-time-efficiency factors for prompt and non-prompt $\mathrm{D_s^+}$ mesons obtained with the BDT selection (open markers) using MC simulations with PYTHIA 6 compared with those obtained using the linear selections [238] (filled markers). Right panel: comparison of the significance of the extracted signal between the BDT (red) and linear (black) selection approaches.

and thus a potential reduction of the associated systematic uncertainties with respect to the linear-selection approach. The ratios of the (Acc $\times$ $\epsilon$) factor obtained with the BDT selection over the one resulting from the linear selections decrease with increasing transverse momentum for both non-prompt and prompt mesons. For $p_{\mathrm{T}} \gtrsim 6$ GeV/$c$, the efficiency ratio for non-prompt $\mathrm{D_s^+}$ mesons is smaller than the prompt one due to the better non-prompt contribution rejection of the BDT selection. The significances of the extracted $\mathrm{D_s^+}$-meson signal as a function of $p_{\mathrm{T}}$ resulting from the two different candidate-selection approaches are shown in the right panel of Fig. 5.17. With the BDT-based selection, an improvement of the statistical significance by a factor 1.2–1.7 depending on $p_{\mathrm{T}}$ is observed, implying a reduction of the statistical uncertainties by the same factor. Moreover, the adoption of machine-learning selections extended the transverse-momentum range of the measurement to lower $p_{\mathrm{T}}$. The two cross-section measurements are in agreement within uncertainties, and their comparison will be discussed in Section 5.5.1.

## 5.4 Systematic uncertainties

In this Section, the procedures used to estimate the systematic uncertainties on the measurements of the $p_{\mathrm{T}}$-differential prompt and non-prompt $\mathrm{D_s^+}$-meson cross sections are discussed. The sources of systematic uncertainty are the following: (i) the raw-yield extraction from the invariant-mass distributions; (ii) the BDT-selection and PID efficiencies; (iii) the generated $\mathrm{D_s^+}$-meson $p_{\mathrm{T}}$-shape in the MC simulations; (iv) the track-reconstruction efficiency; and (v) the prompt and non-prompt fraction estimation. This last contribution has been already discussed in Sections 5.2.3 and 5.3.3 for non-prompt and prompt $\mathrm{D_s^+}$ mesons, respectively.

### 5.4.1 Raw-yield extraction

The systematic uncertainty due to the raw-yield extraction procedure was estimated for each $p_{\mathrm{T}}$ interval varying the fit configuration and observing the consequent variation of the raw yield. In this multi-trial approach, all the possible combinations resulting from five different upper and lower values of the fit range, three combinatorial-background functional forms (linear, exponential, and parabolic), and seven different bin widths of the invariant-mass distribution were considered. In the case of the prompt $\mathrm{D_s^+}$-meson measurement, the width of the Gaussian function used to model the signal peak was a free parameter of the fit. For the non-prompt measurement, the width was varied within the uncertainty of the value resulting from the fit to the prompt-enriched sample. As an additional check, to test the sensitivity to the line-shape of the signal, the raw yields resulting from the fits were compared to those obtained from bin-counting estimates. In the bin-counting approach, the signal yield is obtained by integrating the invariant-mass

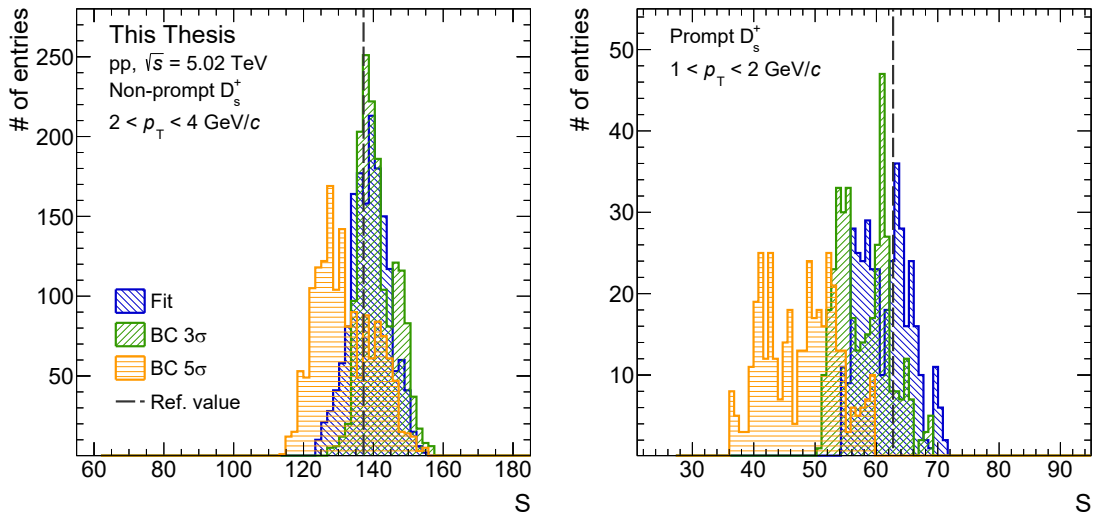



Figure 5.18: Raw-yield distributions from the multi-trial study for the fit and the bin-counting (BC) approaches in the $2 < p_{\mathrm{T}} < 4$ GeV/$c$ interval of the non-prompt $\mathrm{D_s^+}$-meson measurement (left panel) and in the $1 < p_{\mathrm{T}} < 2$ GeV/$c$ interval of the prompt $\mathrm{D_s^+}$-meson measurement (right panel).

distribution after subtracting the background estimated from a side-band fit. Two different integration regions were considered spanning the $\pm 3\sigma$ and $\pm 5\sigma$ intervals around the $\mathrm{D_s^+}$-meson mass, respectively. The multi-trial configurations providing a poor fit quality, i.e., with $\chi^2/ndf > 2$, were not considered in the estimation of the systematic uncertainty.

As an example, the raw-yield distributions resulting from the multi-trial procedure for the fit and bin-counting methods are reported in Fig. 5.18 for the $2 < p_{\mathrm{T}} < 4$ GeV/$c$ interval of the non-prompt $\mathrm{D_s^+}$-meson measurement and the $1 < p_{\mathrm{T}} < 2$ GeV/$c$ interval of the prompt $\mathrm{D_s^+}$-meson measurement. In this latter case, a relevant discrepancy was observed between the raw-yield distribution obtained with the $5\sigma$ bin-counting method and those deriving from the other two approaches. This is due to fluctuations in the invariant-mass bins near the $\mathrm{D_s^+}$-meson peak, as can be seen in the top left panel of Fig. 5.14, which are picked up when the bin counting is performed in a larger invariant-mass region.

The systematic uncertainty was assigned considering the RMS of the raw-yield distribution, resulting from the fit method, and the shift of its mean from the reference value when relevant. The distributions resulting from the bin-counting approach were used as a cross-check and did not enter directly in the systematic-uncertainty estimation. In Fig. 5.19, the sum in quadrature of the RMS and mean shift of the yield distribution relative to the reference value, obtained with the default fit configuration, is reported as a function of $p_{\mathrm{T}}$ for prompt and non-prompt

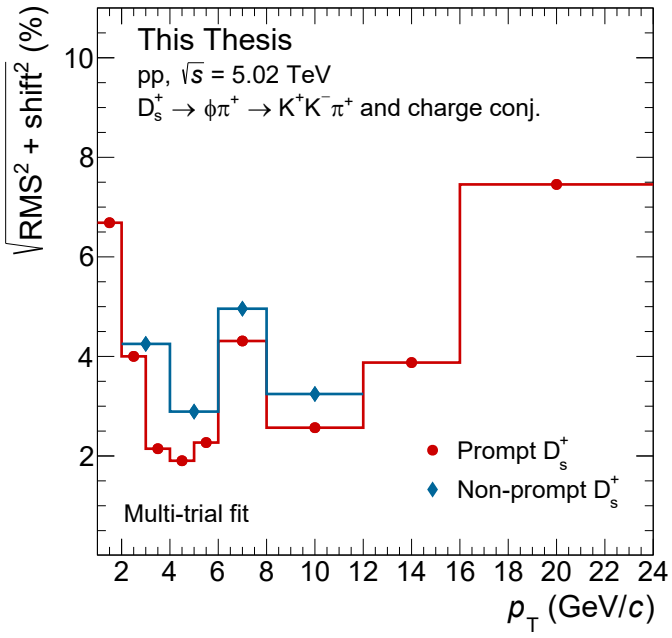



Figure 5.19: Sum in quadrature of the RMS and shift of the mean with respect to the reference value of the raw-yield distributions resulting from the fit approach, for prompt (red) and non-prompt (blue) $D_s^+$ mesons.

$D_s^+$ mesons. The values of the uncertainty are higher for the non-prompt $D_s^+$-meson analysis than for the prompt one in the comparable transverse-momentum intervals due to the more challenging signal extraction of the former measurement. The assigned relative uncertainty values range between 2% (3%) and 7% (4%) depending on $p_T$ for the prompt (non-prompt) $D_s^+$ mesons measurement.

For the non-prompt $D_s^+$-meson measurement, the presence of a correlation between the systematic uncertainty on the $f_{\text{non-prompt}}$ fraction and that due to the raw-yield extraction was explored, as mentioned in Section 5.2.3. To this end, the data-driven computation of the $f_{\text{non-prompt}}$ fraction was repeated varying the fit configurations used for the raw-yield extraction as described above. Only a representative subset of the possible combinations was used for simplicity. The linear correlation coefficient $\rho$ was computed from the relative variations of the $f_{\text{non-prompt}}$ fraction and the raw yield, resulting from the different fit configurations, with respect to the reference values. All the $p_T$ intervals of the analysis were considered together in the computation. Points of the distribution identified as outliers by the unsupervised clustering algorithm DBSCAN [250] were removed from the correlation-coefficient estimation. A positive correlation was observed with a $\rho$ value of about 0.69. This was taken into account in the propagation of the estimated systematic uncertainties to the measured non-prompt $D_s^+$-meson cross section.

### 5.4.2 BDT-selection and PID efficiency

Discrepancies between MC simulations and real data in the description of the variables provided as input to the BDT models can lead to biases in the measured cross sections. These discrepancies arise from imperfections in the modelling of the detector resolutions and alignments in the simulations. The associated systematic uncertainty was estimated by repeating the analysis varying the BDT-selection criteria, and comparing the cross-section values resulting from the different sets of selections. The thresholds on the BDT outputs related to the candidate probability of being combinatorial background and prompt (non-prompt) $D_s^+$ mesons, for the prompt (non-prompt) measurement, were varied. Both the variation of a single threshold keeping the other one fixed to the reference value and the simultaneous variation of both thresholds were considered. The ranges spanned by the thresholds were chosen to provide a maximum variation of the selection efficiency of about $\pm 40\%$. The systematic uncertainty on the BDT-selection efficiency was assigned considering the RMS and shift from unity of the distribution of cross-section ratios between the values obtained from the tested selection criteria and the reference one.

Figure 5.20 reports the results of the BDT selection variation for the non-prompt $D_s^+$-meson measurement in the $2 < p_T < 4$ interval. The evolution as a function of the selection criteria of the raw yield (top left panel), the statistical significance (top middle panel), the acceptance-times-efficiency (top right panel), and the fraction of prompt and non-prompt $D_s^+$ mesons (bottom left panel) is displayed. The selection criteria with labels from 1 to 20 were obtained by enlarging monotonically the threshold on the BDT-output related to the background hypothesis, thus loosening the selection. Those labelled with numbers from 21 to 40 resulted from different thresholds on the BDT-output related to the non-prompt hypothesis, and provide a larger prompt $D_s^+$-meson rejection with increasing label number. Finally, the selection criteria with labels from 41 onward were obtained varying simultaneously the two thresholds.

The cross-section values obtained with the different BDT selections are shown in the bottom middle panel of the Fig. 5.20, and they are compared with the reference value and the assigned systematic uncertainty. The largest discrepancies of the cross section with respect to the reference value are observed for the selection criteria which reject the largest fraction of prompt $D_s^+$ mesons, and correspond to low raw-yield and statistical-significance values, thus indicating that these discrepancies are driven by a poor signal extraction. In the bottom right panel of the Figure, the distribution of the ratios of the cross sections resulting from the tested selection criteria and the reference value is displayed. The sum in quadrature of the distribution RMS and shift from unity is indicated by the orange band, while the blue band reports the systematic uncertainty assigned after smoothing out the obtained value looking at the neighbour $p_T$ intervals.

Figure 5.21 displays the sum in quadrature of the RMS and shift from unity of

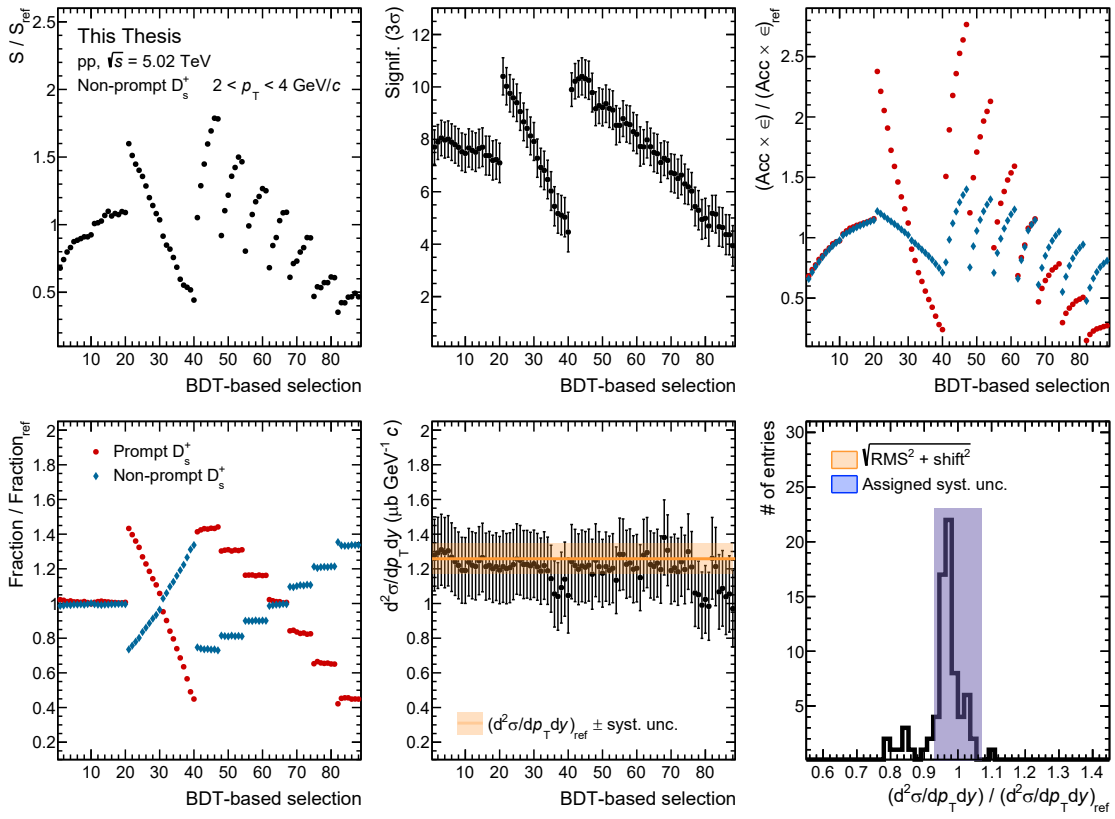



Figure 5.20: Results of the selection-criterion variation for the non-prompt $D_s^+$-meson measurement in the $2 < p_T < 4$ interval. The variation of the raw yield (top left panel), significance (top middle), (Acc $\times \epsilon$) (top right), fraction of prompt and non-prompt $D_s^+$ mesons (bottom left), and cross section (bottom middle) as a function of the BDT-based selection is reported. The bottom right panel shows the distribution of the cross-section ratio with respect to the reference value. The orange and blue bands display the sum in quadrature of the distribution RMS and shift from unity and the assigned systematic uncertainty, respectively.

the distribution of cross-section ratios as a function $p_T$ for the prompt and non-prompt $D_s^+$-meson measurements. The values are slightly larger for the non-prompt measurement than the prompt one in the common $p_T$ region. The uncertainty estimation at high $p_T$ is affected by fluctuations in the raw-yield extraction, thus it was not considered in the systematic-uncertainty assignment. In fact, a larger systematic uncertainty is not expected at high transverse momenta where the selection criteria are less restrictive. The assigned relative systematic uncertainty vary between 3% and 8% decreasing with increasing $p_T$ and depending on the kind of $D_s^+$ meson considered.

Another systematic uncertainty can originate from the PID selections used to

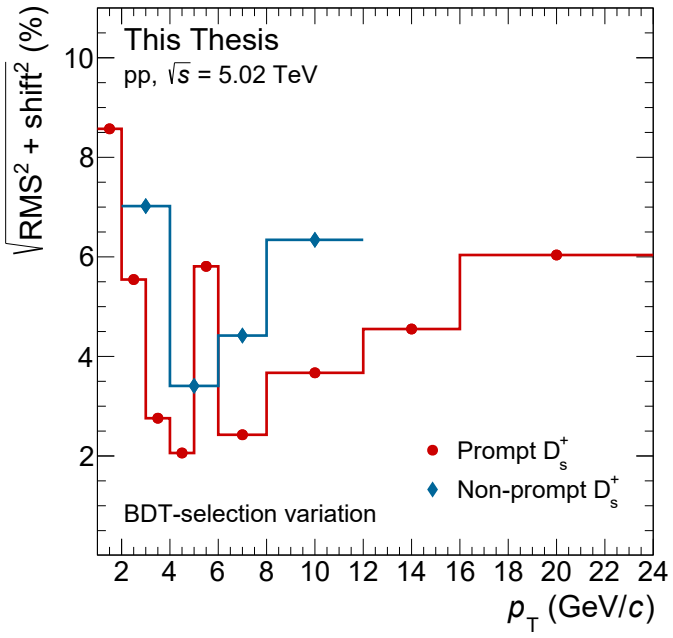



Figure 5.21: Sum in quadrature of the RMS and shift from unity of the distribution of cross-section ratios with respect to the reference value as a function $p_{\mathrm{T}}$ for the prompt (red) and non-prompt (blue) $\mathrm{D_s^+}$ mesons.

prepare the data sample for the machine-learning model training and application mentioned in Sec. 5.1.1. This source of systematic uncertainty was estimated similarly to previous $\mathrm{D_s^+}$-meson studies [238, 251]. The PID-selection efficiencies for pions and kaons were compared in data and in MC simulations. To this end, pure samples of pions were obtained in data from $\Lambda$ and $\mathrm{K_s^0}$ decays, while samples of kaons in the TPC (TOF) were selected applying strict PID requirements using the TOF (TPC) information. The observed per-track differences as a function of $p_{\mathrm{T}}$ were then combined using the $\mathrm{D_s^+}$-meson decay kinematics. No significant discrepancies were observed for the conservative PID strategy, which is described in Sec. 4.2.2. Therefore, no systematic uncertainty was assigned. Possible additional systematic effects due to the employment of the PID information as input to the BDT models are already accounted for in the variation of the BDT-selection criteria described above.

In the case of the non-prompt $\mathrm{D_s^+}$-meson measurement, the possible correlation between the systematic uncertainty on the $f_{\text{non-prompt}}$ fraction and the one on the selection efficiency, which was introduced in Section 5.2.3, was studied. To emulate the effect induced by discrepancies between simulations and data, the $f_{\text{non-prompt}}$ computation was repeated modifying the prompt and non-prompt acceptance times efficiencies employed in the data-driven procedure (see top panel of Fig. 5.11). Shifts of the efficiencies to upper and lower values were considered, together with

modifications of the slope of the efficiency as a function of the BDT selection. These variations were performed both independently and simultaneously for prompt and non-prompt $\mathrm{D_s^+}$ mesons. No clear correlation or anti-correlation between the variations and the resulting $f_{\text{non-prompt}}$ fractions was observed. Therefore, the two sources of systematic uncertainty were considered as uncorrelated.

### Comparison of $\mathrm{D_s^+}$-meson variables between data and MC

To further investigate the systematic uncertainty arising from the BDT-based selection, the distributions of the $\mathrm{D_s^+}$-meson variables provided as input to the BDT models were compared between MC simulations and real data. The variable distributions for signal $\mathrm{D_s^+}$ mesons were obtained from the data sample collected by the experiment as follows. The invariant-mass distribution of $\mathrm{D_s^+}$-meson candidates was fitted, according to the procedure described in Section 5.2.1, to extract the mean, width, and signal-over-background ratio of the signal peak. A *signal-region sample*, containing both $\mathrm{D_s^+}$-meson and combinatorial-background candidates, was obtained from the invariant-mass region spanning $\pm 3\sigma$ around the signal peak position, while a *background-only sample* was obtained from the sidebands of the invariant-mass distribution. The variable distributions for signal $\mathrm{D_s^+}$ mesons were computed by subtracting the distributions obtained from the background-only sample from those resulting from the signal-region sample. In the computation, the distributions were weighted to account for the different sizes of the two samples and the observed signal-over-background ratio of the signal peak. The variable distributions from MC simulations were calculated by summing those of prompt and non-prompt $\mathrm{D_s^+}$ mesons, which were weighted to account for their relative abundance observed in data. The same loose selections reported in Sec. 5.1.1 were employed to prepare the data and MC samples. Moreover, a value of $n\sigma_{\text{comb}}^{\mathrm{K},1}$ below 5 was also required. This additional selection criterium was imposed to suppress the residual background contamination present in the real-data distributions after the background-subtraction procedure, and it is expected to not affect the signal $\mathrm{D_s^+}$-meson candidates for which the $n\sigma_{\text{comb}}^{\mathrm{K},1}$ value is in general well below the required threshold.

In Fig. 5.22, the distributions of $DL^{xy}/\sigma_{DL^{xy}}$, $DL$, $n\sigma_{\text{comb}}^{\mathrm{K},1}$, and the absolute value of $\cos^3\theta'(\mathrm{K})$ for $\mathrm{D_s^+}$ mesons are compared between MC simulations and real data in the $6 < p_{\mathrm{T}} < 8$ GeV/$c$ interval. The variables reported in the Figure are those with the highest mean absolute SHAP values, thus those affecting more the BDT-based selection, for the model used to perform the analysis-level selection in the considered $p_{\mathrm{T}}$ interval. For all the considered variables, the distribution resulting from MC simulations is in good agreement with the real data one. However, large uncertainties affect the latter due to the relative small data sample available and the challenges of extracting a pure sample of $\mathrm{D_s^+}$ mesons from data. Therefore, the precision currently achieved prevents from drawing conclusions about the systematic uncertainty related to the BDT-selection, since the assigned uncertainty

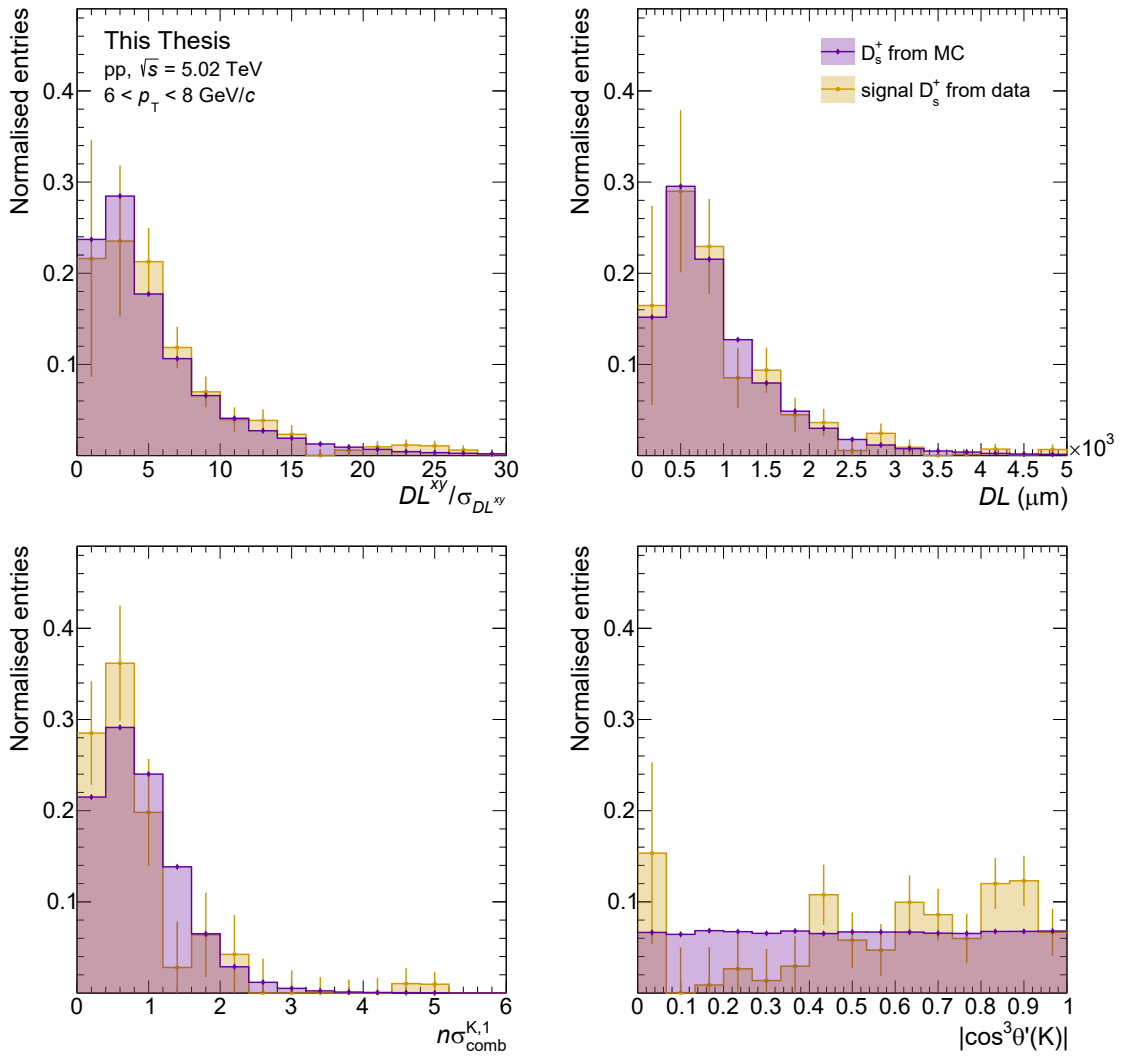



Figure 5.22: Distributions of $DL^{xy}/\sigma_{DL^{xy}}$, $DL$, $n\sigma^{\mathrm{K,1}}_{\mathrm{comb}}$, and the $\cos^3\theta'(\mathrm{K})$ absolute value for signal $\mathrm{D_s^+}$ mesons from MC simulations (purple) and real data (yellow) in pp collisions at $\sqrt{s} = 5.02$ TeV in the $6 < p_{\mathrm{T}} < 8$ GeV/$c$ interval.

values are of the order of a few percent. Finally, it was not possible to perform this exercise for transverse momenta below 6 GeV/$c$, where the analysis-level selections are stricter and thus more sensible to possible discrepancies between data and simulations.

### 5.4.3 Generated Monte Carlo $p_{\mathrm{T}}$ shape

A systematic effect on the efficiency can originate from differences between the real and simulated $\mathrm{D_s^+}$-meson transverse-momentum distributions, due to the

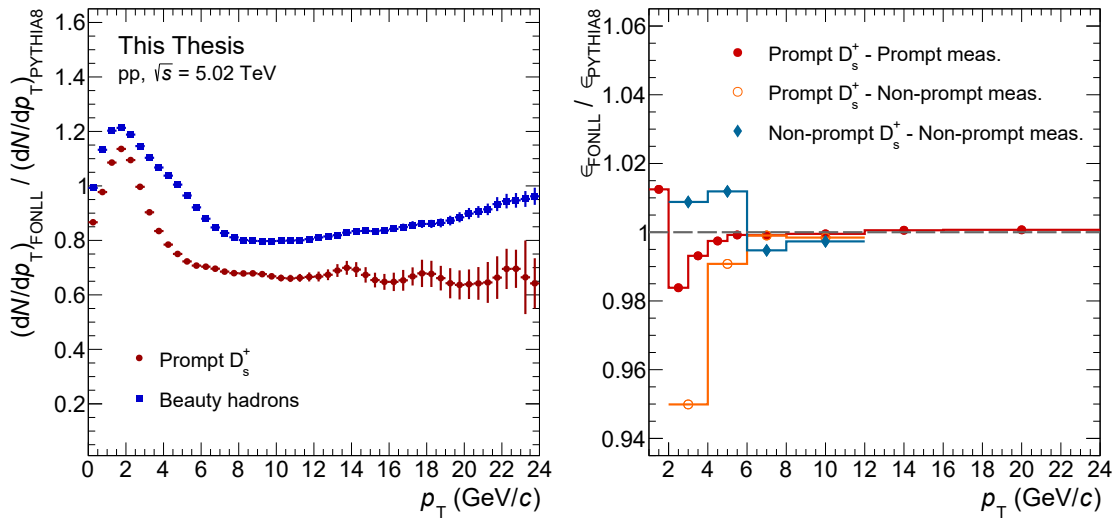



Figure 5.23: Left panel: ratio of normalised $p_{\mathrm{T}}$-differential yields from FONLL calculations and PYTHIA 8 simulations for prompt $\mathrm{D_s^+}$ mesons (red) and beauty hadrons (blue). Right panel: ratio of the efficiencies obtained using the FONLL and PYTHIA 8 $p_{\mathrm{T}}$ shapes for prompt and non-prompt (blue) $\mathrm{D_s^+}$ mesons. The prompt-efficiency ratio obtained for the BDT-selection used in the prompt (non-prompt) analysis is reported in red (orange).

finite width of the analysis transverse-momentum intervals and the variation of the efficiency within the $p_{\mathrm{T}}$ interval. This uncertainty was estimated by computing the efficiency after weighting the $p_{\mathrm{T}}$ shape from the PYTHIA 8 generator to match the one from FONLL calculations. In the left panel of Fig. 5.23, the ratio between the normalised $p_{\mathrm{T}}$-differential yield obtained from FONLL calculations and the one from PYTHIA 8 simulations is reported for prompt $\mathrm{D_s^+}$ mesons and beauty hadrons. The FONLL predictions for the prompt $\mathrm{D_s^+}$-meson production were obtained using the recipe described in Section 5.1.3.

The weights were applied to the $p_{\mathrm{T}}$ distributions of prompt $\mathrm{D_s^+}$ mesons and to the parent beauty hadrons in case of non-prompt $\mathrm{D_s^+}$ mesons. The ratio of the efficiencies obtained using the FONLL and PYTHIA 8 $p_{\mathrm{T}}$ shapes is shown for prompt and non-prompt $\mathrm{D_s^+}$ mesons in the right panel of Fig. 5.23. In the case of the prompt $\mathrm{D_s^+}$-meson measurement, a small variation of the prompt efficiency is observed and a 1% relative uncertainty was assigned below $p_{\mathrm{T}} = 3$ GeV/$c$, while the effect was considered negligible for $p_{\mathrm{T}} > 3$ GeV/$c$. For the non-prompt measurement, the variation of the non-prompt $\mathrm{D_s^+}$-meson efficiency is small, i.e., about 1% for $p_{\mathrm{T}} < 6$ GeV/$c$. However, a larger variation is present for the prompt efficiency at low $p_{\mathrm{T}}$. This is due to the small efficiencies resulting from the selections applied to reject the prompt $\mathrm{D_s^+}$ mesons, which are more sensible to the generated $p_{\mathrm{T}}$ shape. The prompt efficiency does not enter directly in the cross-section computation, but

it is used to estimate the $f_{\text{non-prompt}}$ fraction. Therefore, the $f_{\text{non-prompt}}$ fraction was calculated using the efficiency obtained from the FONLL $p_{\mathrm{T}}$ shape and compared to the reference value. The resulting discrepancy varied between zero and 1% depending on $p_{\mathrm{T}}$ and was summed linearly to the one observed for the efficiency. A relative systematic uncertainty of 1% was assigned in all the non-prompt $\mathrm{D_s^+}$-meson measurement $p_{\mathrm{T}}$ range.

### 5.4.4 Track-reconstruction efficiency

The systematic uncertainty on the track-reconstruction efficiency includes effects due to the track-quality selections, and the track finding and prolongation from the TPC to the ITS detector.

The contribution due to the applied track-quality requirements, which are reported in Section 4.2.1, was inherited from previous D-meson measurements performed on the same data sample [238]. It was evaluated measuring the prompt D-meson cross section varying three different track-quality selection criteria, which include: (i) a number of crossed rows in the TPC above $120-5/(p_{\mathrm{T}}\ [\mathrm{GeV}/c])$ instead of 70 which is the default value, (ii) a number of TPC clusters larger than 0.65 times the TPC crossed-row number, and (iii) a ratio of crossed rows over findable clusters in the TPC above 0.9, while the default values is 0.8. The cross-section variation was found to be of about 1.5% for D mesons decaying into three charged particles independently of $p_{\mathrm{T}}$, and this value was assigned as systematic uncertainty.

The TPC-to-ITS prolongation efficiency is computed as the ratio of the number of tracks prolonged from the TPC to the ITS detector with at least one hit in the SPD layers, and the number of tracks reconstructed in the TPC. Discrepancies in the tracking performance between data and MC simulations induce a systematic uncertainty. This uncertainty was estimated as the deviation from unity of the prolongation efficiency obtained from MC simulations over the one observed in data, and is reported as a function of the track $p_{\mathrm{T}}$ in the left panel of Fig. 5.24. The prolongation efficiency for primary particles, which are produced in the collision or in strong and weak decays of heavy-flavour hadrons, is higher than the one for secondary particles, originating from material interactions and strange-hadron decays [252]. In addition, the relative abundances of primary and secondary particles are not correctly reproduced in the MC simulations based on the PYTHIA 8 event generator and the GEANT 3 transport package. Therefore, the prolongation efficiency from simulations was weighted to account for the abundances observed in data, which were determined via fits to the inclusive $d_0^{xy}$ distributions of the tracks [236].

The per-track systematic uncertainty on the TPC-to-ITS prolongation efficiency was propagated to the $\mathrm{D_s^+}$ mesons through the kinematics of the decay tracks, i.e., taking into account the correlation between the $\mathrm{D_s^+}$-meson transverse momentum

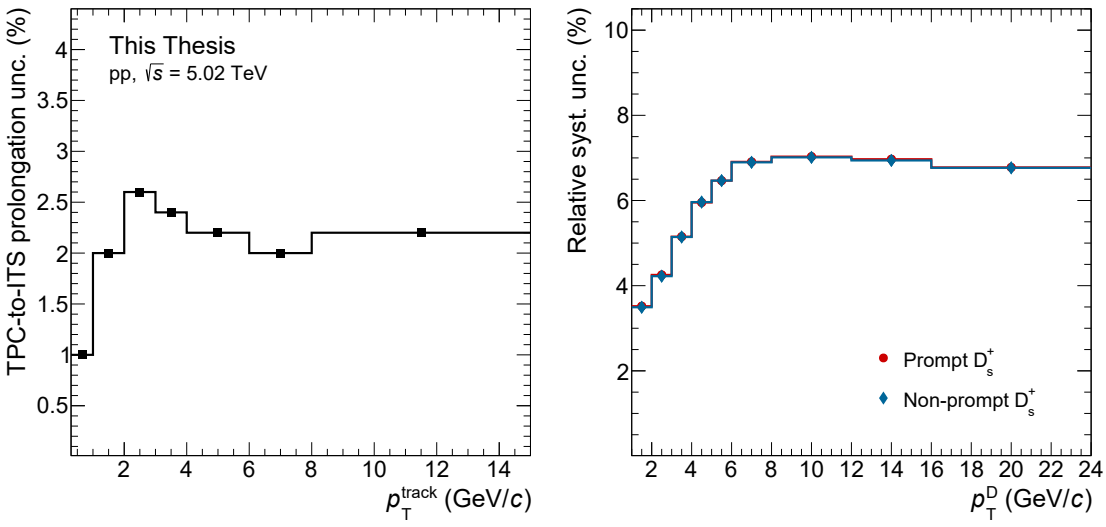



Figure 5.24: Left panel: relative systematic uncertainty on the TPC-to-ITS prolongation efficiency for single tracks as a function of $p_{\mathrm{T}}$. Right panel: relative systematic uncertainty due to tracking for prompt (red) and non-prompt (blue) $\mathrm{D_s^+}$ mesons, including the contributions of track-quality selections and prolongation efficiency.

and the $p_{\mathrm{T}}$ of the decay particles. In this calculation, the track-prolongation uncertainty was considered as fully correlated among the three particles that constitute the candidate. Therefore, the uncertainties on the single tracks were summed linearly. Finally, this contribution was summed in quadrature to the uncertainty due to the track-quality selection variation. The relative systematic uncertainty on the track-reconstruction efficiency is shown as a function of $p_{\mathrm{T}}$ for prompt and non-prompt $\mathrm{D_s^+}$ mesons in the right panel of Fig. 5.24. The uncertainty values are the same for prompt and non-prompt $\mathrm{D_s^+}$ mesons.

### 5.4.5 Systematic uncertainty summary

The systematic uncertainties on the measured cross sections are summarised in Tabs. 5.7 and 5.8 for non-prompt and prompt $\mathrm{D_s^+}$ mesons, respectively. In addition to the sources mentioned in the previous Sections, a global normalisation uncertainty was considered. It is due to the branching-ratio uncertainty for the $\mathrm{D_s^+ \to \phi\pi^+ \to K^-K^+\pi^+}$ decay [6] and the systematic uncertainty on the integrated luminosity [233]. The total systematic uncertainties were obtained summing in quadrature the different sources reported in the Tables. For non-prompt $\mathrm{D_s^+}$ mesons, the systematic uncertainties on the non-prompt fraction estimation and the raw-yield extraction were propagated as partially correlated following what discussed in Section 5.4.1.

The systematic uncertainties on the measurement were considered to be fully correlated between the various $p_{\mathrm{T}}$ intervals, except for the raw-yield extraction uncertainty. This contribution was considered as uncorrelated due to the substantial variations of the combinatorial-background shape and the $S/B$ ratio as a function of $p_{\mathrm{T}}$. Considering the prompt and non-prompt $\mathrm{D_s^+}$-meson cross sections, the tracking-efficiency and global-normalisation systematic uncertainties were considered as fully correlated between the two measurements, while all the other sources of uncertainty were treated as uncorrelated.

Table 5.7: Summary of the relative systematic uncertainties on the non-prompt $\mathrm{D_s^+}$-meson $p_{\mathrm{T}}$-differential cross section in pp collisions at $\sqrt{s} = 5.02$ TeV.

| $p_{\mathrm{T}}$ (GeV/$c$) | 2–4 | 4–6 | 6–8 | 8–12 |
|---|---|---|---|---|
| Yield extraction (%) | 4.0 | 3.0 | 3.0 | 3.0 |
| Selection efficiency (%) | 7.0 | 5.0 | 5.0 | 5.0 |
| Non-prompt fraction (%) | 2.0 | 3.0 | 3.0 | 4.0 |
| MC $p_{\mathrm{T}}$ shape (%) | 1.0 | 1.0 | 1.0 | 1.0 |
| Tracking efficiency (%) | 5.5 | 6.5 | 7.0 | 7.0 |
| Branching ratio (%) | 3.6 | | | |
| Luminosity (%) | 2.1 | | | |
| Total (%) | 11.3 | 10.8 | 11.1 | 11.6 |

## 5.5 Results

### 5.5.1 Production cross sections

In Fig. 5.25, the prompt $\mathrm{D_s^+}$-meson cross section measured in this Thesis is compared with the previous result from Ref. [238], which was obtained using the same data sample of pp collisions at $\sqrt{s} = 5.02$ TeV and utilising linear selections in the analysis. The $p_{\mathrm{T}}$ reach is extended to $p_{\mathrm{T}} = 1$ GeV/$c$, while the previous measurement was performed in the $2 < p_{\mathrm{T}} < 24$ GeV/$c$ interval, and the total uncertainty is smaller by a factor varying from 1.07 to 1.40 depending on $p_{\mathrm{T}}$. These improvements are due to the updated analysis technique based on machine-learning selections.

The ratio of the two cross-section measurements is reported in the bottom panel

Table 5.8: Summary of the relative systematic uncertainties on the prompt $\mathrm{D_s^+}$-meson $p_\mathrm{T}$-differential cross section in pp collisions at $\sqrt{s} = 5.02$ TeV.

| $p_\mathrm{T}$ (GeV/$c$) | 1–2 | 2–3 | 3–4 | 4–5 | 5–6 | 6–8 | 8–12 | 12–16 | 16–24 |
|---|---|---|---|---|---|---|---|---|---|
| Yield extraction (%) | 7.0 | 4.0 | 2.0 | 2.0 | 3.0 | 3.0 | 3.0 | 3.0 | 7.0 |
| Selection efficiency (%) | 8.0 | 5.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 |
| Prompt fraction (%) | $^{+7.4}_{-8.4}$ | $^{+4.6}_{-5.6}$ | $^{+3.5}_{-4.8}$ | $^{+3.6}_{-5.1}$ | $^{+3.5}_{-5.1}$ | $^{+3.2}_{-4.6}$ | $^{+2.7}_{-3.9}$ | $^{+2.3}_{-3.2}$ | $^{+2.0}_{-2.7}$ |
| MC $p_\mathrm{T}$ shape (%) | 1.0 | 1.0 | 1.0 | negl. | negl. | negl. | negl. | negl. | negl. |
| Tracking efficiency (%) | 3.5 | 4.5 | 5.5 | 6.0 | 6.5 | 7.0 | 7.0 | 7.0 | 7.0 |
| Branching ratio (%) | | | | | 3.6 | | | | |
| Luminosity (%) | | | | | 2.1 | | | | |
| Total (%) | $^{+14.1}_{-14.6}$ | $^{+10.0}_{-10.5}$ | $^{+8.6}_{-9.2}$ | $^{+8.9}_{-9.6}$ | $^{+9.5}_{-10.2}$ | $^{+9.7}_{-10.3}$ | $^{+9.6}_{-10.0}$ | $^{+9.5}_{-9.7}$ | $^{+11.3}_{-11.5}$ |

of the Figure. In the ratio, the statistical uncertainties were propagated as partially correlated considering the sample of $\mathrm{D_s^+}$ mesons obtained with the linear selections fully contained in the one resulting from the BDT selection. This choice is motivated by the 1.5–2 times larger raw yields extracted in the analysis based on machine learning. The systematic uncertainties due to the prompt-fraction estimation, the tracking efficiency, and the normalisation were propagated as fully correlated, while the remaining sources were considered as uncorrelated in the ratio. The two measurements are compatible within the combined statistical and systematic uncertainties. However, their ratio is systematically below unity of about 10%. This discrepancy is understood considering that the prompt $\mathrm{D_s^+}$-meson efficiency obtained from MC simulations based on the PYTHIA 8 generator, employed in the analysis presented in this Thesis, is larger of a few percent than the one resulting from simulations with PYTHIA 6, used in the previous result, as shown in Section 5.3.2. Another systematic shift towards lower cross-section values is caused by the updated branching ratios of the beauty-hadron decays $\mathrm{H_b} \to \mathrm{D_s^+} + \mathrm{X}$ employed to compute the non-prompt $\mathrm{D_s^+}$-meson contribution, which lead to a larger correction and thus to a smaller $f_\mathrm{prompt}$ fraction. Finally, the discrepancies in the $6 < p_\mathrm{T} < 8$ GeV/$c$ and $12 < p_\mathrm{T} < 16$ GeV/$c$ intervals are understood as fluctuations in the raw-yield extraction of the previous measurement. They were caused by the width of the $\mathrm{D_s^+}$-meson peak resulting from the fit, which was larger than the one obtained from MC simulations [251].

The $p_\mathrm{T}$-differential production cross sections of prompt and non-prompt $\mathrm{D^0}$, $\mathrm{D^+}$, and $\mathrm{D_s^+}$ mesons measured in the rapidity interval $|y| < 0.5$ are reported in the left panel of Fig. 5.26. In the right panel of the Figure, the ratios of the $p_\mathrm{T}$-differential cross sections of non-prompt and prompt D mesons is displayed [71].

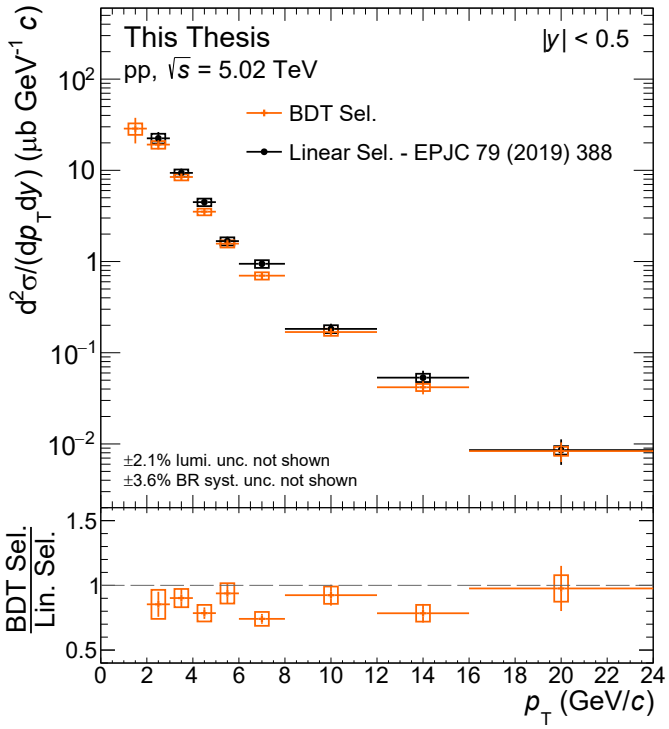



Figure 5.25: $p_{\mathrm{T}}$-differential cross section of prompt $\mathrm{D_s^+}$ mesons in pp collisions at $\sqrt{s} = 5.02$ TeV obtained in this Thesis (orange) compared with the previous result on the same data sample obtained using linear selections [238] (black). The ratio of the two measurements is reported in the bottom panel. Statistical (vertical bars) and systematic uncertainties (boxes) are shown.

The statistical uncertainties on the prompt and non-prompt measurements were propagated as uncorrelated in the ratios, since the fraction of D-meson candidates in common between the two samples is small. For what concerns the systematic uncertainty, the contributions due to the tracking efficiency, the luminosity, and the branching ratio were considered as fully correlated in the propagation, while all the other sources were considered as uncorrelated. The ratio shows an increasing trend with $p_{\mathrm{T}}$ for the three D-meson species up to $p_{\mathrm{T}} = 12$ GeV/$c$, as expected due to the harder $p_{\mathrm{T}}$ distribution of beauty hadrons with respect to D mesons. The $\mathrm{D_s^+}$-meson ratio is systematically higher than those observed for $\mathrm{D^0}$ and $\mathrm{D^+}$ mesons, suggesting a larger contribution of beauty-hadron decays to $\mathrm{D_s^+}$ than to $\mathrm{D^0}$ and $\mathrm{D^+}$ mesons. Nevertheless, the large uncertainties prevent from drawing strong conclusions.

In Fig. 5.27, the prompt and non-prompt $\mathrm{D_s^+}$-meson cross sections are compared to predictions resulting from FONLL [60, 72, 73] and GM-VFNS [75, 76] pQCD calculations. FONLL predictions for the production cross section of prompt $\mathrm{D_s^+}$

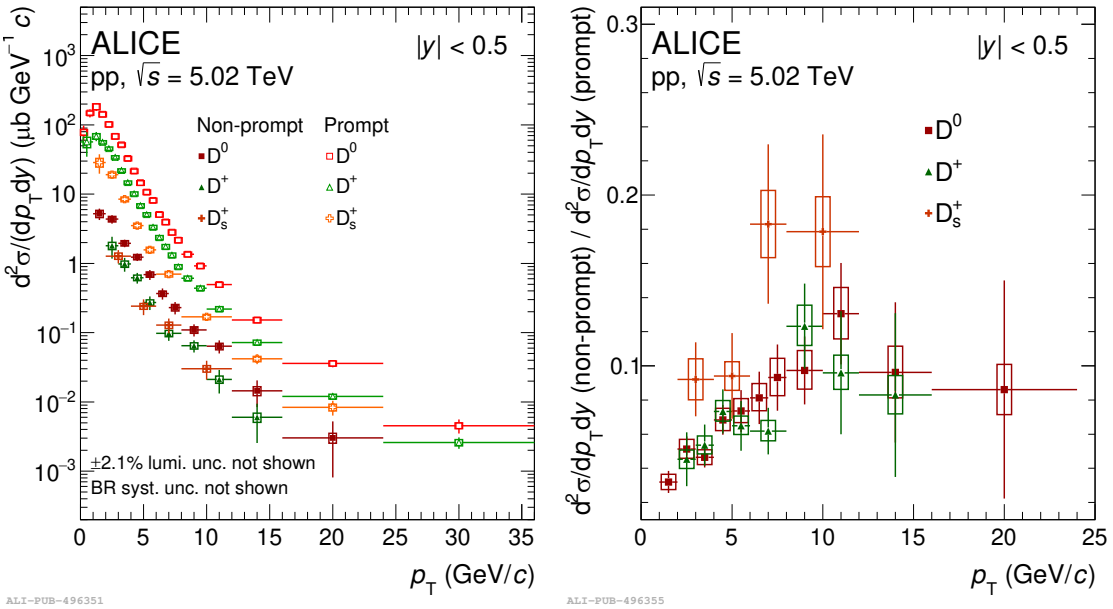



Figure 5.26: Left panel: $p_{\mathrm{T}}$-differential production cross sections of prompt and non-prompt $\mathrm{D}^0$, $\mathrm{D}^+$, and $\mathrm{D}_\mathrm{s}^+$ mesons in pp collisions at $\sqrt{s} = 5.02$ TeV. Right panel: cross-section ratios of non-prompt and prompt D mesons. Statistical (vertical bars) and systematic uncertainties (boxes) are shown. Figures from Ref. [71].

mesons are not available. For non-prompt $\mathrm{D}_\mathrm{s}^+$ mesons, predictions were obtained from FONLL calculations of the beauty-hadron production cross section, and the description of the $\mathrm{H_b} \to \mathrm{D}_\mathrm{s}^+ + \mathrm{X}$ decay kinematics and branching ratios was performed with PYTHIA 8 [82, 168], as explained in Section 5.1.3. In the GM-VFNS framework, the non-prompt $\mathrm{D}_\mathrm{s}^+$-meson cross section results from a single-step description of the transition from beauty quarks to charm hadrons. A set of fragmentation functions for $\mathrm{b} \to \mathrm{D}_\mathrm{s}^+ + \mathrm{X}$ obtained from measurements in $\mathrm{e^+e^-}$ collisions was adopted [253]. The FONLL uncertainty band includes the uncertainties due to the choice of renormalisation and factorisation scales, and heavy-quark masses, and the uncertainties on the PDFs. More details on the uncertainty calculation are provided in Section 5.3.3. In the GM-VFNS uncertainties, only the renormalisation scale is varied and the CTEQ14 PDFs [64] uncertainties are considered.

The $p_{\mathrm{T}}$-differential prompt $\mathrm{D}_\mathrm{s}^+$-meson cross section is described within uncertainties by the GM-VFNS predictions. The measured values lie on the upper edge of the theory uncertainty band at low $p_{\mathrm{T}}$, and are in agreement with the central values of GM-VFNS predictions for $p_{\mathrm{T}} \gtrsim 5$ GeV/$c$. The measured cross section for non-prompt $\mathrm{D}_\mathrm{s}^+$ mesons is well compatible with the FONLL+PYTHIA 8 calculations, while it is underestimated by the GM-VFNS predictions by a factor ranging from about 5 to 2 depending on $p_{\mathrm{T}}$. GM-VFNS predictions better describing the non-prompt measurements are available for the $\mathrm{D}^0$ and $\mathrm{D}^+$ mesons, and are based

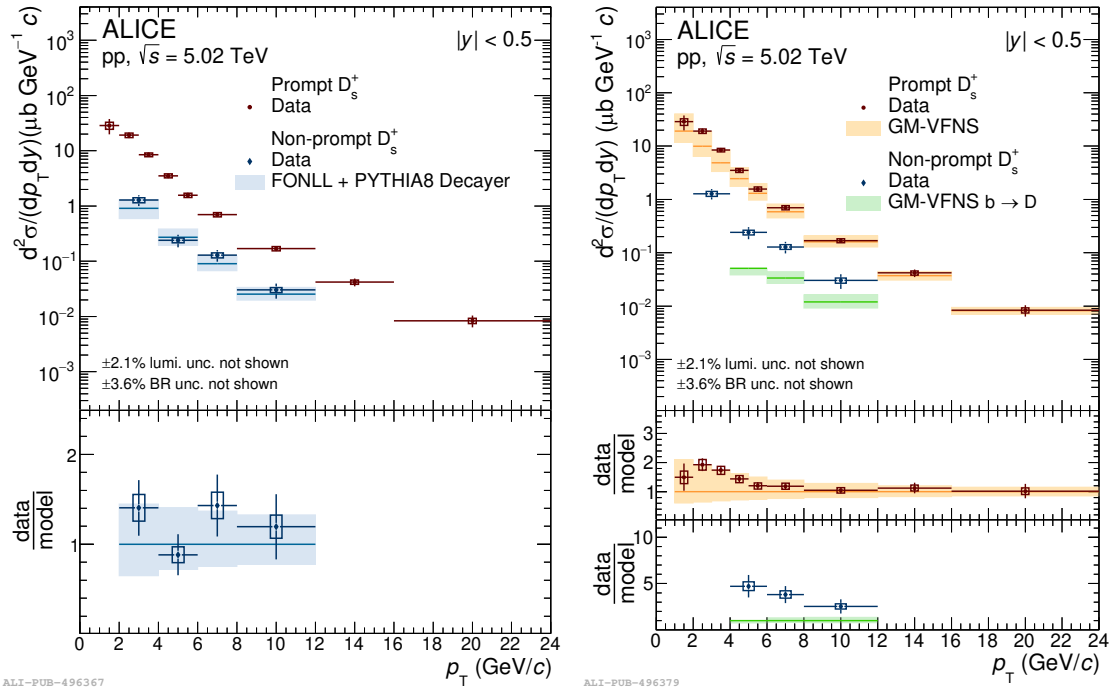



Figure 5.27: $p_{\mathrm{T}}$-differential cross sections of prompt and non-prompt $D_s^+$ mesons compared with predictions obtained with FONLL [60, 72, 73] and PYTHIA 8 [82, 168] calculations (left panel) and to GM-VFNS [75, 76] predictions (right panel). Figures from Ref. [71].

Table 5.9: $p_{\mathrm{T}}$-integrated visible cross sections of prompt and non-prompt $D_s^+$ mesons in pp collisions at $\sqrt{s} = 5.02$ TeV, in the measured $p_{\mathrm{T}}$ range and in $|y| < 0.5$.

| $D_s^+$ mesons | $p_{\mathrm{T}}$ range (GeV/$c$) | Visible cross section (μb) |
|---|---|---|
| Prompt | 1–24 | $64 \pm 9(\mathrm{stat})^{+6}_{-7}(\mathrm{syst}) \pm 1(\mathrm{lumi}) \pm 2(\mathrm{BR})$ |
| Non-prompt | 2–12 | $3.4 \pm 0.6(\mathrm{stat}) \pm 0.3(\mathrm{syst}) \pm 0.1(\mathrm{lumi}) \pm 0.1(\mathrm{BR})$ |

on a double-step approach similar to the one adopted in the FONLL+PYTHIA 8 calculations [71].

The visible cross sections of prompt and non-prompt $D_s^+$ mesons are reported in Tab. 5.9. They were calculated by integrating the $p_{\mathrm{T}}$-differential cross sections in the measured transverse-momentum ranges. The systematic uncertainties on the $p_{\mathrm{T}}$-differential cross sections were propagated as fully correlated among the measured $p_{\mathrm{T}}$ intervals, except for the raw-yield extraction uncertainty as mentioned in Section 5.4.5.

The total production cross sections of prompt and non-prompt $D_s^+$ mesons in

Table 5.10: Total production cross sections of prompt and non-prompt $\mathrm{D_s^+}$ mesons in the $|y| < 0.5$ range in pp collisions at $\sqrt{s} = 5.02$ TeV, together with the computed extrapolation factor to $p_{\mathrm{T}} > 0$.

| $\mathrm{D_s^+}$ mesons | Extr. factor | $\mathrm{d}\sigma/\mathrm{d}y\|_{\|y\|<0.5}$ (μb) |
|---|---|---|
| Prompt | $1.28^{+0.35}_{-0.12}$ | $82 \pm 12(\mathrm{stat}) \pm 8(\mathrm{syst}) \pm 2(\mathrm{lumi}) \pm 3(\mathrm{BR})^{+23}_{-8}(\mathrm{extr})$ |
| Non-prompt | $2.03^{+0.04}_{-0.15}$ | $6.9 \pm 1.2(\mathrm{stat}) \pm 0.7(\mathrm{syst}) \pm 0.1(\mathrm{lumi}) \pm 0.2(\mathrm{BR})^{+0.1}_{-0.5}(\mathrm{extr})$ |

the $|y| < 0.5$ rapidity interval were obtained by multiplying the visible cross sections by an extrapolation factor, to account for the yield in the unmeasured $p_{\mathrm{T}}$ region. They are reported in Tab. 5.10. The extrapolation factor was calculated as the ratio between the total production cross section in $|y| < 0.5$ and that in the measured $p_{\mathrm{T}}$ range using the FONLL central predictions. For prompt $\mathrm{D_s^+}$ mesons, FONLL calculations are not available, and the predictions were computed from the FONLL charm-quark $p_{\mathrm{T}}$-differential cross section as detailed in Section 5.1.3. The systematic uncertainty on the extrapolation factor was estimated varying the FONLL parameters as discussed in Section 5.3.3. In the case of non-prompt $\mathrm{D_s^+}$ mesons, the extrapolation factor was computed using the FONLL predictions for the beauty-hadron production and the PYTHIA 8 description of the $\mathrm{H_b} \to \mathrm{D_s^+} + \mathrm{X}$ decay kinematics. In addition to the FONLL uncertainties, two other systematic-uncertainty contributions to the non-prompt extrapolation factor were accounted for. They were the uncertainties on the beauty-quark fragmentation fractions $f(\mathrm{b} \to \mathrm{H_b})$, and on the branching ratios of the $\mathrm{H_b} \to \mathrm{D_s^+} + \mathrm{X}$ decays. The former contribution was estimated considering the alternative set of fragmentation fractions measured in $\mathrm{p\overline{p}}$ collisions [6] and reported in Tab. 5.5. For the latter contribution, the branching ratios implemented in PYTHIA 8 were weighted to match the measured PDG values [6] shown in Tab. 4.2. The prompt $\mathrm{D_s^+}$-meson cross section is compatible with that obtained in previous studies [238]. However, the total uncertainty is reduced due to the improved precision of the $p_{\mathrm{T}}$-differential measurement and the extended $p_{\mathrm{T}}$ range, which implies a smaller fraction of extrapolated cross section.

### 5.5.2 Prompt and non-prompt $\mathrm{D_s^+}/(\mathrm{D^0} + \mathrm{D^+})$ ratios

The ratios of the production cross section of different D-meson species constrain theoretical calculations, where these ratios are sensitive mainly to the fragmentation functions and the implemented hadronisation model. Moreover, the comparison of the production of strange and non-strange mesons allows the determination of the ratio $f_{\mathrm{s}}/(f_{\mathrm{u}} + f_{\mathrm{d}})$, i.e., the fragmentation fraction of charm and beauty quarks to strange mesons divided by the one to non-strange mesons.

The ratios between the $p_{\mathrm{T}}$-differential cross sections of $\mathrm{D_s^+}$ mesons and the

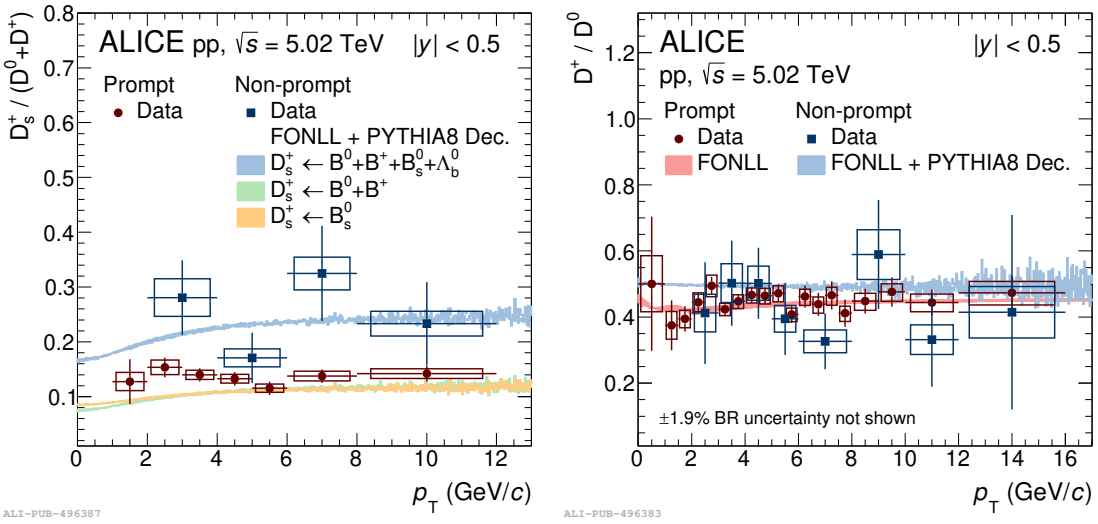



Figure 5.28: Ratios of the $p_{\mathrm{T}}$-differential cross sections of $\mathrm{D_s^+}$ and the sum of $\mathrm{D^0}$ and $\mathrm{D^+}$ mesons (left panel), and of $\mathrm{D^+}$ and $\mathrm{D^0}$ mesons (right panel) compared with predictions obtained from FONLL [60, 72, 73] and PYTHIA 8 [82, 168] calculations. For the non-prompt $\mathrm{D_s^+}/(\mathrm{D^0}+\mathrm{D^+})$ ratio, the predictions for the $\mathrm{D_s^+}$ from $\mathrm{B_s^0}$-meson and from non-strange B-meson decays are also displayed separately. Figures from Ref. [71].

sum of $\mathrm{D^0}$- and $\mathrm{D^+}$- meson ones are shown in the left panel of Fig. 5.28. In the right panel, the $\mathrm{D^+}$-over-$\mathrm{D^0}$-meson cross section ratios for prompt and non-prompt mesons are shown [71]. In the ratio calculation, the systematic uncertainties due to the tracking-efficiency, luminosity, and $f_{\mathrm{prompt}}$-fraction estimation, in the case of the prompt D-meson measurements, were propagated as fully correlated between the D-meson species. The other sources of systematic uncertainty were treated as uncorrelated. The observed D-meson ratios do not depend on $p_{\mathrm{T}}$ in the measured transverse-momentum range, considering the current experimental uncertainties. The $\mathrm{D^+}/\mathrm{D^0}$ ratio is compatible between prompt and non-prompt D mesons, while a higher $\mathrm{D_s^+}/(\mathrm{D^0}+\mathrm{D^+})$ ratio is observed for non-prompt mesons than for prompt ones. This enhancement of the final states with a $\mathrm{D_s^+}$ meson is qualitatively expected from $\mathrm{b} \to \mathrm{c\bar{c}s}$ and $\mathrm{\bar{b}} \to \mathrm{c\bar{c}s}$ weak decays, and it is consistent with LEP measurements [68]. The prompt $\mathrm{D^+}/\mathrm{D^0}$ ratio is in good agreement with FONLL calculations, while the non-prompt D-meson ratios are compatible with the FONLL+PYTHIA 8 predictions. The contributions of $\mathrm{B_s^0}$ and non-strange B mesons decaying into $\mathrm{D_s^+}$ mesons are displayed separately in the left panel of Fig. 5.28, to show the relevant fraction of non-prompt $\mathrm{D_s^+}$ mesons produced in non-strange B-meson decays.

From the prompt $\mathrm{D_s^+}/(\mathrm{D^0}+\mathrm{D^+})$ ratio it is possible to compute the charm-quark fragmentation fraction to charm-strange mesons $f_{\mathrm{s}}$ divided by the one to

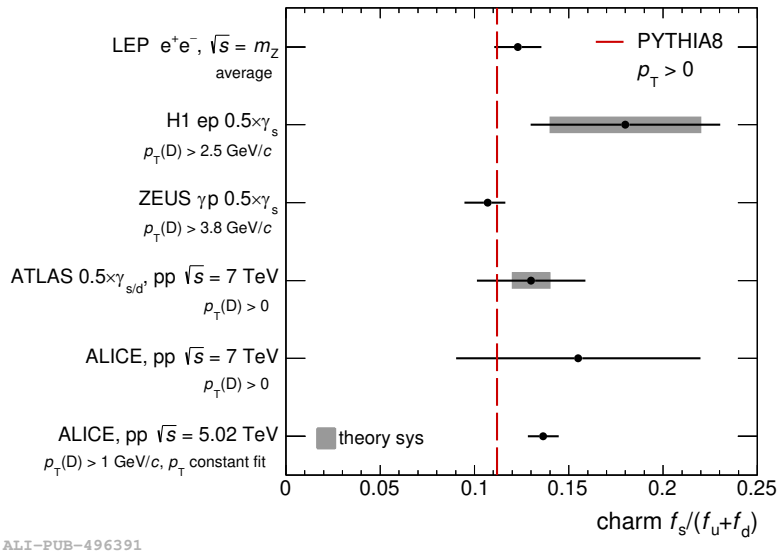



Figure 5.29: Charm-quark $f_{\rm s}/(f_{\rm u}+f_{\rm d})$ ratio compared with previous measurements performed by the ALICE [254], H1 [255], ZEUS [256], and ATLAS [257] Collaborations and to the average of LEP measurements [68]. The experimental (bars) and theoretical (shaded boxes) uncertainties are shown. The red line reports the value obtained from PYTHIA 8 simulations with Monash-13 tune [135]. Figure from Ref. [71].

non-strange charm mesons $f_{\rm u}+f_{\rm d}$, since all the relevant D-meson excited states decay to $\rm D^0$, $\rm D^+$, or $\rm D_s^+$ mesons. The charm-quark $f_{\rm s}/(f_{\rm u}+f_{\rm d})$ ratio was obtained by fitting the data with a constant function, given that no $p_{\rm T}$ dependence was observed for the D-meson ratios . The resulting value is

$$\left(\frac{f_{\rm s}}{f_{\rm u}+f_{\rm d}}\right)_{\rm charm} = 0.136 \pm 0.005({\rm stat}) \pm 0.006({\rm syst}) \pm 0.005({\rm BR}). \qquad (5.13)$$

All the sources of systematic uncertainty were treated as fully correlated between the different $p_{\rm T}$ intervals, except for the one related to the raw-yield extraction. The same quantity can be estimated starting from the total production cross-sections of $\rm D^0$, $\rm D^+$, and $\rm D_s^+$ mesons. In this case the computed $f_{\rm s}/(f_{\rm u}+f_{\rm d})$ ratio is $0.128 \pm 0.020({\rm stat}) \pm 0.010({\rm syst}) \pm 0.005({\rm BR})^{+0.035}_{-0.012}({\rm extr})$ [71], which is compatible with the value obtained by fitting the prompt $\rm D_s^+/(D^0+D^+)$ ratio. However, it is affected by larger uncertainties since the estimation is dominated by the low $p_{\rm T}$ region, where the measured cross sections are less precise and extrapolations based on theoretical calculations are needed.

Figure 5.29 reports the comparison between the charm-quark fragmentation-fraction ratio $f_{\rm s}/(f_{\rm u}+f_{\rm d})$ obtained and previous measurements of the strangeness suppression factor $\gamma_{\rm s} = 2{\rm D_s^+}/({\rm D^0}+{\rm D^+})$ from the ALICE [254], H1 [255], ZEUS [256],

and ATLAS [257] Collaborations. These values were divided by a factor two accounting for the difference between $\gamma_\mathrm{s}$ and $f_\mathrm{s}/(f_\mathrm{u} + f_\mathrm{d})$. The measurements are compatible within uncertainties, and are in agreement with the average of the LEP measurements [68]. They are also compatible with the $f_\mathrm{s}/(f_\mathrm{u} + f_\mathrm{d})$ ratio computed using PYTHIA 8 simulations with the Monash-13 tune [135]. Although, considering the result presented in this Thesis, a tension of about $2.7\sigma$ is observed between the PYTHIA 8 result and the measurement including both statistical and systematic uncertainties. Overall, the charm-quark fragmentation fractions into D mesons, and in particular the $f_\mathrm{s}/(f_\mathrm{u} + f_\mathrm{d})$ ratio, are compatible between different collision systems and centre-of-mass energies, supporting the observation that the fragmentation of charm quarks into mesons is "universal".

The beauty-quark fragmentation fraction to beauty-strange mesons divided by the one to non-strange beauty mesons was calculated from the non-prompt $\mathrm{D_s^+}/(\mathrm{D^0} + \mathrm{D^+})$ ratio following a procedure similar to that detailed above. An additional correction factor was applied considering the fraction of non-prompt $\mathrm{D^0}$ and $\mathrm{D^+}$ mesons not produced in non-strange B-meson decays and that of non-prompt $\mathrm{D_s^+}$ mesons not originating from $\mathrm{B_s^0}$-meson decays. The correction factor was obtained from FONLL+PYTHIA 8 calculations, and the systematic uncertainty was estimated by varying the fragmentation fractions and the beauty-hadron branching ratios as discussed in Sec. 5.5.1. The contribution of $\mathrm{B_s^0}$ and non-strange B mesons to the non-prompt $\mathrm{D_s^+}$-meson cross section is similar, as shown in the left panel of Fig. 5.28. Instead, the non-prompt $\mathrm{D^0}$- and $\mathrm{D^+}$-meson production is almost exclusively due to decays of non-strange B mesons. The $p_\mathrm{T}$-differential beauty-quark fragmentation-fraction ratio was computed as

$$\left.\frac{f_\mathrm{s}}{f_\mathrm{u} + f_\mathrm{d}}\right|_\mathrm{beauty} = \left[\frac{N(\mathrm{D_s^+} \leftarrow \mathrm{B_s^0})}{N(\mathrm{D_s^+} \leftarrow \mathrm{H_b})} \times \frac{N(\mathrm{D^0}, \mathrm{D^+} \leftarrow \mathrm{H_b})}{N(\mathrm{D^0}, \mathrm{D^+} \leftarrow \mathrm{B^{0,+}})}\right]^{\text{FONLL+PYTHIA 8}} \times \times \left.\frac{\mathrm{D_s^+}}{\mathrm{D^0} + \mathrm{D^+}}\right|_\mathrm{non-prompt}, \tag{5.14}$$

where the $p_\mathrm{T}$-dependence of the different terms is omitted. The correction factor determined from FONLL+PYTHIA 8 calculations varies between about 0.49 and 0.50 and decreases with increasing transverse momentum. The values resulting from Eq. (5.14) were then fitted with a constant function to obtain the final result, which is

$$\left(\frac{f_\mathrm{s}}{f_\mathrm{u} + f_\mathrm{d}}\right)_\mathrm{beauty} = 0.127 \pm 0.036(\mathrm{stat}) \pm 0.012(\mathrm{syst}) \pm 0.005(\mathrm{BR}) \pm 0.005(\mathrm{th}). \tag{5.15}$$

The beauty-quark $f_\mathrm{s}/(f_\mathrm{u} + f_\mathrm{d})$ ratio has an additional theoretical-uncertainty contribution compared to the charm-quark one, which is due to the correction factor.

The beauty-quark fragmentation-fraction ratio is compared with measurements in pp and $\mathrm{p\overline{p}}$ collisions from the CDF [258], LHCb [259, 260], and ATLAS [261]

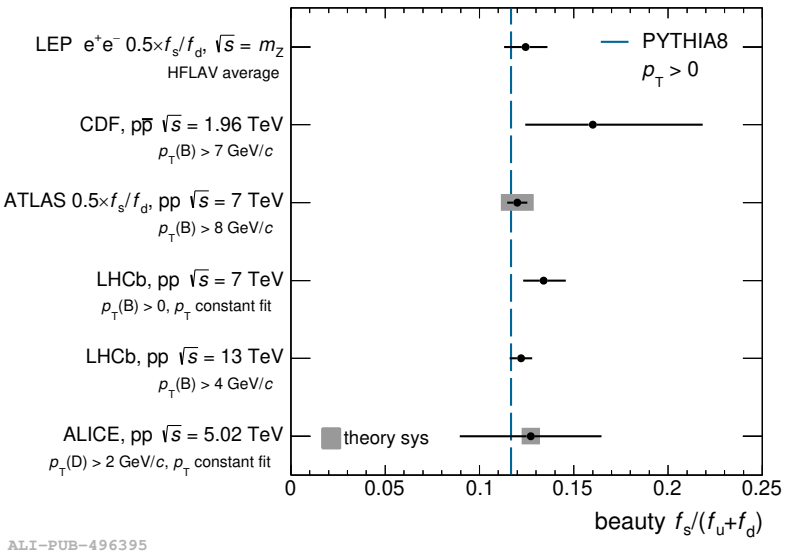



Figure 5.30: Beauty-quark $f_{\rm s}/(f_{\rm u}+f_{\rm d})$ ratio compared with previous measurements performed by the CDF [258], LHCb [259, 260], and ATLAS [261] Collaborations and to the average of LEP measurements [262]. The experimental (bars) and theoretical (shaded boxes) uncertainties are shown. The blue line reports the value obtained from PYTHIA 8 simulations with Monash-13 tune [135]. Figure from Ref. [71].

Collaborations in Fig. 5.30. The ATLAS measurement of $f_{\rm s}/f_{\rm u}$ was divided by a factor two assuming isospin symmetry for the u and d quarks, which implies $f_{\rm u} = f_{\rm d}$. The considered measurements of the $f_{\rm s}/(f_{\rm u} + f_{\rm d})$ ratio are compatible with the LEP average provided by the HFLAV Collaboration [262], and with the value resulting from PYTHIA 8 simulations with Monash-13 tune. Finally, it is possible to observe that the fragmentation-fraction ratios $f_{\rm s}/(f_{\rm u}+f_{\rm d})$ of charm and beauty quarks are similar between each other.

### 5.5.3 $\mathrm{b\overline{b}}$ production cross section

The $\mathrm{b\overline{b}}$ production cross section per unit of rapidity was estimated from the measured visible non-prompt $\mathrm{D_s^+}$-meson cross section at midrapidity. The approach utilised is similar to that employed to compute the $p_{\rm T}$-integrated total production cross section of $\mathrm{D_s^+}$ mesons, which is reported in Section 5.5.1. An extrapolation factor $\alpha_{\rm extr}^{\rm b\overline{b}}$ was calculated as

$$\alpha_{\rm extr}^{\rm b\overline{b}} = \frac{{\rm d}\sigma_{\rm b\overline{b}}/{\rm d}y|_{|y|<0.5}^{\rm FONLL}}{\sigma_{\rm b\to D_s^+}^{\rm FONLL+PYTHIA\ 8}(p_{\rm T}^{\rm min} < p_{\rm T} < p_{\rm T}^{\rm max}, |y| < 0.5)}. \tag{5.16}$$

In the above equation, ${\rm d}\sigma_{\rm b\overline{b}}/{\rm d}y|_{|y|<0.5}^{\rm FONLL}$ is the FONLL prediction for the $\mathrm{b\overline{b}}$ production cross section corrected for the different rapidity-distribution shapes between

beauty hadrons and $\mathrm{b\overline{b}}$ pairs. The denominator is the non-prompt $\mathrm{D_s^+}$-meson cross section from FONLL+PYTHIA 8 calculations in the $p_{\mathrm{T}}$ and rapidity ranges of the measurement. The correction factor for the $\mathrm{b\overline{b}}$ rapidity distribution comprehends two terms. The first one considers the different beauty-hadron and single beauty-quark rapidity distributions, and it was estimated to be equal to unity in the rapidity range of interest using FONLL calculations. A 1% systematic uncertainty on this term was assigned from the difference between FONLL and PYTHIA 8 estimations. The second term of the correction factor is the $(\mathrm{d}\sigma_{\mathrm{b\overline{b}}}/\mathrm{d}y)/(\mathrm{d}\sigma_{\mathrm{b}}/\mathrm{d}y)$ ratio, and it was computed from the NLO pQCD calculations of POWHEG [85] as $\mathrm{d}\sigma_{\mathrm{b\overline{b}}}^{|y|<0.5}/\mathrm{d}\sigma_{\mathrm{b}}^{|y|<0.5} = 1.06$ [3]. A 1% uncertainty on this term was assigned considering the variation of the factorisation and renormalisation scales in the POWHEG calculation and using different sets of PDFs, namely CT10NLO [263] and CT14NLO [64]. The value of the extrapolation factor obtained from Eq. (5.16) is $\alpha_{\mathrm{extr}}^{\mathrm{b\overline{b}}} = 11.9^{+0.7}_{-1.0}$. The $\mathrm{b\overline{b}}$ production cross section resulting from the measurement of non-prompt $\mathrm{D_s^+}$ mesons presented in this Thesis is

$$\left.\frac{\mathrm{d}\sigma_{\mathrm{b\overline{b}}}}{\mathrm{d}y}\right|_{|y|<0.5}^{\mathrm{b}\to\mathrm{D_s^+}} = 40.5 \pm 6.9(\mathrm{stat}) \pm 3.8(\mathrm{syst}) \pm 0.8(\mathrm{lumi}) \pm 1.4(\mathrm{BR})^{+2.7}_{-3.3}(\mathrm{extr}) \pm 0.6(\mathrm{rap.\ shape})\ \mu\mathrm{b}. \tag{5.17}$$

Figure 5.31 reports the $\mathrm{b\overline{b}}$ cross section extrapolated from the non-prompt $\mathrm{D_s^+}$-meson measurement compared to those obtained from non-prompt $\mathrm{D^0}$ and $\mathrm{D^+}$ mesons [71], and dielectrons [264]. The cross sections resulting from the three D-meson species are compatible within uncertainties between each other and with those provided by the other two ALICE measurements. The average of the non-prompt D-meson measurements $\mathrm{d}\sigma_{\mathrm{b\overline{b}}}/\mathrm{d}y|_{|y|<0.5} = 34.5 \pm 2.4(\mathrm{stat})^{+4.7}_{-2.9}(\mathrm{tot.\ syst})\ \mu\mathrm{b}$ is also reported in the Figure. All the ALICE measurements are in agreement with FONLL [60, 72, 73] and NNLO [80] calculations. The inclusion of next-to-next-to-leading-order corrections provides smaller theoretical uncertainties and a larger central value, which is in better agreement with the result obtained from the non-prompt D-meson average, compared to the FONLL prediction.

The $\mathrm{b\overline{b}}$ cross section per rapidity unit at midrapidity resulting from the non-prompt D-meson average is also shown in Fig. 5.32, together with the other existing measurements in pp collisions by the ALICE [265–268] and PHENIX [269] Collaborations at different centre-of-mass energies, and in $\mathrm{p\overline{p}}$ collisions by the CDF [270] and UA1 [271] Collaborations. The experimental results are compatible with FONLL and NNLO calculations over all the explored interval of centre-of-mass energies.

[3] In the calculations, the rapidity of the $\mathrm{b\overline{b}}$ pair was defined as $y_{\mathrm{b\overline{b}}} = 1/2 \times \ln[(E_1 + E_2 + p_{z,1} + p_{z,2})/(E_1 + E_2 - p_{z,1} - p_{z,2})]$, where $E_i$ and $p_{z,i}$ are the energy and the momentum component along the beam direction, respectively, of the beauty and antibeauty quarks.

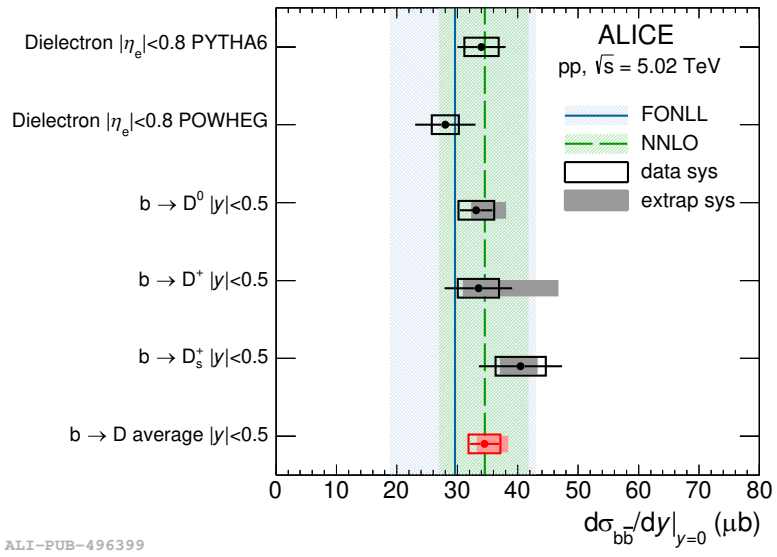



Figure 5.31: Estimates of $\mathrm{d}\sigma_{\mathrm{b\bar{b}}}/\mathrm{d}y$ at midrapidity from dielectrons [264] and non-prompt $\mathrm{D}^0$, $\mathrm{D}^+$, and $\mathrm{D}_\mathrm{s}^+$ mesons measured in pp collisions at $\sqrt{s} = 5.02$ TeV compared to FONLL [60, 72, 73] (blue) and NNLO [80] (green) predictions. The average $\mathrm{d}\sigma_{\mathrm{b\bar{b}}}/\mathrm{d}y$ of the estimates from the single D-meson species is reported in red. Figure from Ref. [71].

Similarly to what described above, the $\mathrm{c\bar{c}}$ cross section per unit of rapidity in pp collisions at $\sqrt{s} = 5.02$ TeV can be computed from production measurements of the prompt ground-state charm hadrons. i.e., $\mathrm{D}^0$, $\mathrm{D}^+$, and $\mathrm{D}_\mathrm{s}^+$ mesons, and $\Lambda_\mathrm{c}^+$, and $\Xi_\mathrm{c}^0$ baryons. The $\mathrm{c\bar{c}}$ cross-section value obtained from the measurements of the ALICE Collaboration at midrapidity, including the $\mathrm{D}_\mathrm{s}^+$-meson one reported in this Thesis, is $\mathrm{d}\sigma_{\mathrm{c\bar{c}}}/\mathrm{d}y|_{|y|<0.5} = 1165 \pm 44(\mathrm{stat})^{+134}_{-101}(\mathrm{tot.\ syst})$ μb [69], and it is in agreement within uncertainties with FONLL and NNLO [81] calculations.

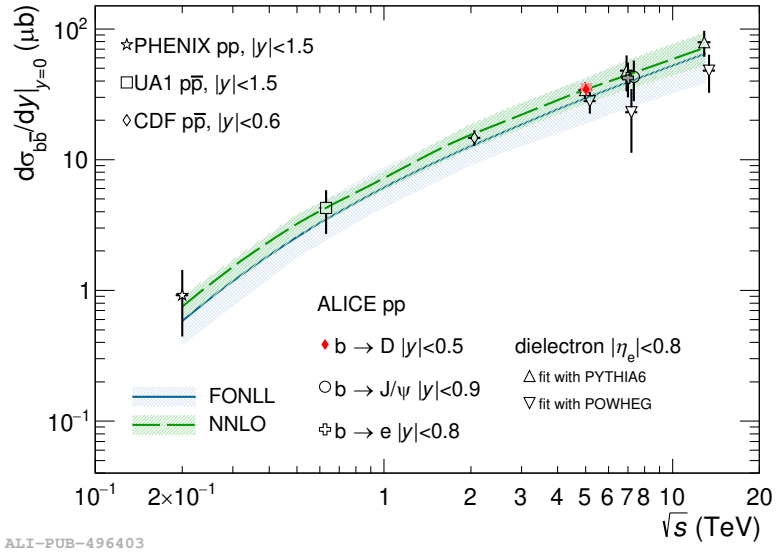



Figure 5.32: Measurements of the $b\bar{b}$ production cross section as a function of the centre-of-mass energy in pp collisions by the ALICE [265–268] and PHENIX [269] Collaborations and in $p\bar{p}$ collisions by the CDF [270] and UA1 [271] Collaborations. The ALICE data points are shifted in the $\sqrt{s}$-axis for better visibility. The FONLL [60, 72, 73] (blue) and NNLO [80] (green) calculations with their uncertainties are reported.

# Chapter 6

# Prompt $\mathrm{D_s^+}$-meson nuclear modification factor in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV

The measurements of the prompt $\mathrm{D_s^+}$-meson $p_{\mathrm{T}}$-differential corrected yield and nuclear modification factor $R_{\mathrm{AA}}$ in central and semicentral Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV in the $|y| < 0.5$ rapidity interval are discussed in this Chapter. The study of the prompt $\mathrm{D_s^+}$-meson production in Pb–Pb collisions is a powerful tool to probe the charm-quark interaction with the expanding dense and hot QGP medium [272]. Moreover, comparing the production of $\mathrm{D_s^+}$ and non-strange D mesons in heavy-ion collisions, it is possible to investigate the interplay of the charm-quark hadronisation via recombination with quarks from the medium and the hadronisation due to the fragmentation mechanism [137]. Finally, possible differences between non-strange D and $\mathrm{D_s^+}$ mesons can arise from their different coupling to the hadronic phase [142]. The measurements shown in this Chapter were published in Ref. [193].

The $p_{\mathrm{T}}$-differential corrected yield of prompt $\mathrm{D_s^+}$ mesons was computed in each $p_{\mathrm{T}}$ interval as

$$\left.\frac{\mathrm{d}N_{\mathrm{AA}}}{\mathrm{d}p_{\mathrm{T}}}\right|_{|y|<0.5} = \frac{\frac{1}{2}\, f_{\mathrm{prompt}}(p_{\mathrm{T}}) N^{\mathrm{D_s^\pm},raw}(p_{\mathrm{T}})\Big|_{|y|<y_{\mathrm{fid}}(p_{\mathrm{T}})}}{(\mathrm{Acc}\times\epsilon)^{\mathrm{prompt}}(p_{\mathrm{T}}) c_{\Delta y}(p_{\mathrm{T}}) \Delta p_{\mathrm{T}}} \times \frac{1}{\mathrm{BR} N_{\mathrm{evt}}}. \tag{6.1}$$

The equation ingredients are the same as those introduced for the cross-section formula of Eq. (5.1). The only exception is the normalisation term $N_{\mathrm{evt}}$ which is the number of analysed events, corresponding to about 100 and 85 million events for the 0–10% and 30–50% centrality classes, respectively. The centrality-class definition is discussed in Section 2.4.

From the $p_{\mathrm{T}}$-differential corrected yield measured in Pb–Pb collisions, the nuclear modification factor was calculated as

$$R_{\mathrm{AA}}(p_{\mathrm{T}}) = \frac{1}{\langle T_{\mathrm{AA}} \rangle} \times \frac{\mathrm{d}N_{\mathrm{AA}}/\mathrm{d}p_{\mathrm{T}}}{\mathrm{d}\sigma_{\mathrm{pp}}/\mathrm{d}p_{\mathrm{T}}}, \tag{6.2}$$

where $\mathrm{d}\sigma_{\mathrm{pp}}/\mathrm{d}p_{\mathrm{T}}$ is the prompt $\mathrm{D_s^+}$-meson cross section measured in pp collisions, and $\langle T_{\mathrm{AA}} \rangle$ is the average value of the nuclear overlap function[1] for the centrality class considered. The $\langle T_{\mathrm{AA}} \rangle$ values are $(23.26 \pm 0.17)$ mb$^{-1}$ and $(3.92 \pm 0.06)$ mb$^{-1}$ for the central and semicentral event intervals, respectively [22].

## 6.1 Machine-learning selections

In the measurements presented in this Chapter, machine-learning techniques were adopted to perform the selection of prompt $\mathrm{D_s^+}$-meson candidates. This choice was motivated by the performance improvement shown by BDT classifiers with respect to linear selections for the prompt $\mathrm{D_s^+}$-meson measurement in pp collisions, which is quantified in Section 5.3.4.

Since the data sample of $\mathrm{D_s^+}$-meson candidates to analyse is much larger in Pb–Pb than in pp collisions, the machine-learning inference step was executed on the ALICE-experiment Grid using the tools introduced in Section 3.3.1. Binary classification BDTs were employed, and the two classes were constituted by prompt $\mathrm{D_s^+}$ mesons and combinatorial background, respectively. In fact, differently from the studies performed in pp collisions, only prompt $\mathrm{D_s^+}$ mesons were measured. These machine-learning models are simpler and smaller than multi-class classification BDTs with the same hyperparameter configuration [243]. Therefore, their inference step requires less computing resources, which are a concern when large data samples are processed on the distributed-computing infrastructure of the ALICE experiment. It is also worth noticing that the trade-off for the better performance of the machine-learning selections is a larger footprint on the execution time than the traditional linear selections.

### 6.1.1 Data preparation

The BDT training and performance assessment were performed exploiting data samples constituted by prompt $\mathrm{D_s^+}$-meson signal candidates and by combinatorial background. The signal candidates were obtained from MC simulations based on the PYTHIA 8 [82, 168] and HIJING [169] event generators, which are detailed

[1] In this context, the nuclear overlap function $T_{AB}$ is not normalised to obtain $\int T_{AB}(\boldsymbol{b})\mathrm{d}b^2 = 1$ as done in Section 1.3.2. Its integrated value is instead equal to $AB$, where $A$ and $B$ are the number of nucleons composing the two colliding nuclei, respectively.

Table 6.1: Topological and kinematic selections, as a function of $p_{\mathrm{T}}$, applied on $\mathrm{D_s^+}$-meson candidates before the machine-learning step for the 0–10% and 30–50% centrality classes.

| Centrality | 0–10% | | | 30–50% | |
|---|---|---|---|---|---|
| $p_{\mathrm{T}}$ (GeV/$c$) | 2–4 | 4–6 | 6–50 | 2–6 | 6–36 |
| $\sigma_{\mathrm{vtx}}$ (µm) < | 400 | 400 | 600 | 450 | 600 |
| $DL$ (µm) > | 500 | 300 | 200 | 200 | 200 |
| $DL^{xy}/\sigma_{DL^{xy}}$ (µm) > | 5 | 4 | 2 | 2 | 2 |
| $\cos\theta_{\mathrm{P}}$ > | 0.97 | 0.97 | 0.92 | 0.96 | 0.92 |
| $\cos\theta_{\mathrm{P}}^{xy}$ > | 0.96 | 0.96 | 0.90 | 0.96 | 0.90 |
| $\Delta M_{\mathrm{KK}}$ (MeV/$c^2$) < | 10 | 10 | 15 | 15 | 15 |

in Section 4.3. The combinatorial-background candidates were extracted from the sidebands of the $\mathrm{D_s^+}$-meson invariant-mass distribution in the data collected by the experiment. Differently from the studies conducted in pp collisions, non-prompt $\mathrm{D_s^+}$ mesons were not included in the training sample. The mix of prompt and non-prompt candidates in the signal class would bias the selection towards secondary vertices with larger displacements, while disfavouring the selection of prompt $\mathrm{D_s^+}$ mesons. This would lead to a large contribution of non-prompt $\mathrm{D_s^+}$-meson candidates in the measured raw yields.

The daughter tracks of the $\mathrm{D_s^+}$-meson candidates fulfilled the single-track requirements reported in Section 4.2.1 and the conservative PID strategy. Moreover, different kinematic and topological selections were applied depending on the centrality class and the transverse momentum. They are reported in Tab. 6.1. These selections are less stringent than those used to extract the $\mathrm{D_s^+}$-meson signal in previous studies based on linear selections [251, 273]. They were used to partly reduce the large combinatorial background typical of Pb–Pb collisions, while keeping the candidates that have a higher probability of being signal for the machine-learning application.

The BDT training was performed separately for the 0–10% and the 30–50% centrality classes, and in larger $p_{\mathrm{T}}$ intervals than the analysis ones. The aim was to keep small the number of BDTs used in the inference step on the Grid. All the prompt $\mathrm{D_s^+}$-meson candidates available in dedicated MC productions were used. At low $p_{\mathrm{T}}$, the number of combinatorial-background candidates was chosen to be the double of the signal ones. While, at high $p_{\mathrm{T}}$, fewer background candidates were employed to keep small the fraction of real data used in the training, since these

Table 6.2: Number of prompt $D_s^+$-meson and combinatorial-background candidates used to build the training and test set in the $p_T$ intervals in which the training was performed, for the 0–10% and 30–50% centrality classes.

| Centrality | 0–10% | | | | 30–50% | | | |
|---|---|---|---|---|---|---|---|---|
| $p_T$ (GeV/$c$) | 2–4 | 4–6 | 6–12 | 12-50 | 2–4 | 4–6 | 6–12 | 12-36 |
| prompt $D_s^+$ ($\times 10^3$, approx.) | 94 | 141 | 381 | 501 | 336 | 265 | 365 | 374 |
| background ($\times 10^3$, approx.) | 188 | 282 | 381 | 125 | 672 | 530 | 365 | 22 |

candidates were also used for the physics measurement. The approximate number of prompt $D_s^+$-meson and combinatorial-background candidates, composing the training and test sets, is reported in Tab. 6.2 for the 0–10% and 30–50% centrality classes. The training sets were formed by 80% of the available data samples, while the remaining 20% composed the test sets.

### 6.1.2 BDT training and performance evaluation

The $D_s^+$-meson topological, kinematic, and PID features provided as input to the BDTs were the same as those employed for the studies in pp collisions, which are reported in Section 5.1.2. Moreover, the same procedure based on the Bayesian optimisation and a 5-fold cross validation was employed to tune the model hyperparameters. In this case, the performance metric chosen was the ROC AUC since the classification problem is binary. Particular attention was dedicated to the definition of the hyperparameter intervals explored by the optimisation algorithm, with the aim of keeping small the size of the final models in view of the inference step on the Grid. The hyperparameter configurations adopted in the training of the final BDTs, which were then used to perform the $D_s^+$-meson selection in the analysis, are reported in Tab. 6.3 as a function of $p_T$ and centrality class. These configurations provided the highest ROC AUC values given the imposed constrains.

The output of the BDT is a score related to the candidate probability of being a real $D_s^+$ meson, and the score values range between zero and unity. An example of the BDT-score distributions for prompt and non-prompt $D_s^+$ mesons, and the combinatorial background is shown in Fig. 6.1. They were obtained from the test-set data sample in the $2 < p_T < 4$ GeV/$c$ interval for the 0–10% centrality class. Combinatorial-background candidates have a low score related to the candidate probability of being signal, and their distribution is peaked at zero. Instead, the distribution of real prompt $D_s^+$ mesons shows a peak at unity as expected. Non-prompt $D_s^+$-meson candidates, which are not included in the training, have a wider distribution than prompt $D_s^+$ mesons with a small secondary peak for BDT outputs

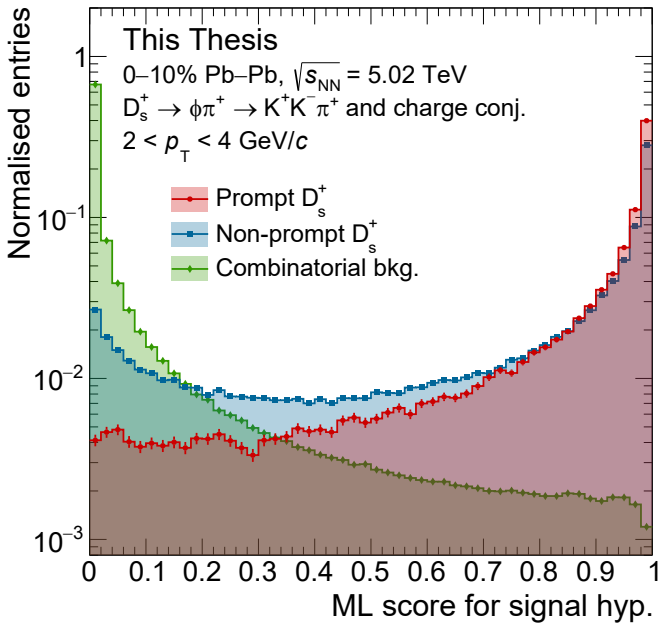



Figure 6.1: Distributions of the binary-BDT output score for prompt (red) and non-prompt (blue) $D_s^+$ mesons, and the combinatorial background (green) in the $2 < p_T < 4$ GeV/$c$ interval for the 0–10% centrality class.

close to zero. This indicates that the model is partially sensible to the differences between prompt and non-prompt $D_s^+$ mesons, and on average assigns to the latter smaller scores.

Table 6.3: Hyperparameter configurations for the final BDTs in the different $p_T$ intervals used in the training procedure, for the 0–10% and 30–50% centrality classes.

| Centrality | 0–10% | | | | 30–50% | | | |
|---|---|---|---|---|---|---|---|---|
| $p_T$ (GeV/$c$) | 2–4 | 4–6 | 6–12 | 12-50 | 2–4 | 4–6 | 6–12 | 12-36 |
| `n_estimators` | 883 | 1043 | 994 | 815 | 833 | 1099 | 1098 | 833 |
| `learning_rate` | 0.084 | 0.066 | 0.075 | 0.081 | 0.081 | 0.046 | 0.060 | 0.081 |
| `max_depth` | 4 | 5 | 6 | 5 | 4 | 5 | 5 | 4 |
| `min_child_weight` | 5.7 | 1.1 | 6.1 | 4.5 | 5.6 | 5.2 | 7.3 | 5.6 |
| `subsample` | 0.90 | 0.85 | 0.90 | 0.82 | 0.81 | 0.89 | 0.92 | 0.81 |
| `colsample_bytree` | 0.90 | 0.84 | 0.82 | 0.95 | 0.89 | 0.90 | 0.89 | 0.89 |

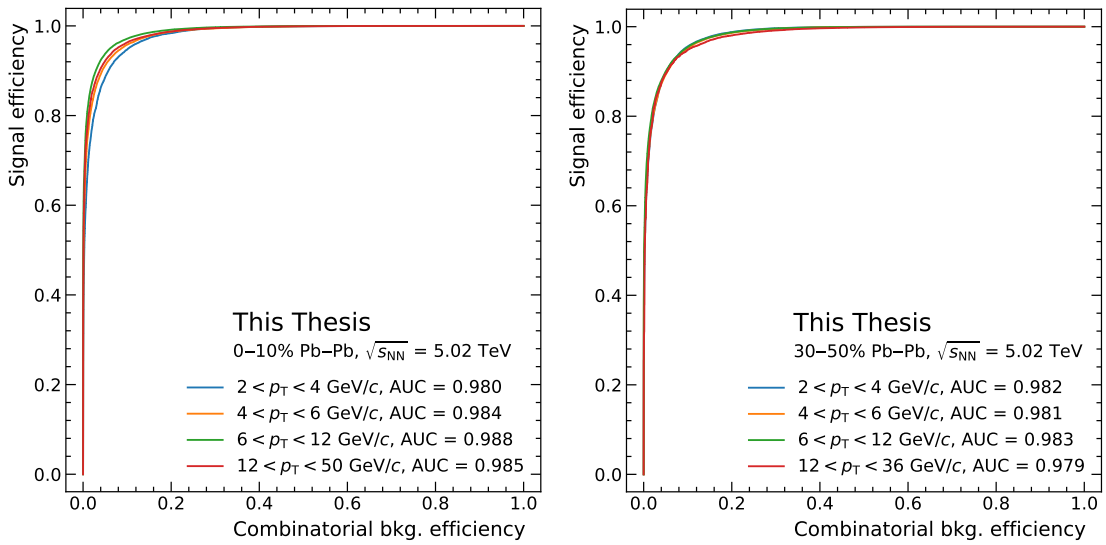



Figure 6.2: ROC curves and the corresponding ROC AUC values, estimated using the test set, for the trained BDT models in the different $p_{\mathrm{T}}$ intervals for the 0–10% (left panel) and 30–50% (right panel) centrality classes.

The performance of the trained BDT models was estimated computing their ROC AUC values on the test set. In Fig. 6.2, the ROC curves and the corresponding ROC AUC values are displayed for the 0–10% and 30–50% centrality classes, and for the different transverse-momentum intervals in which the training was performed. The ROC AUC values range between about 0.98 and 0.99, and are very similar between the two centrality classes considered. Moreover, no significant dependence on the $p_{\mathrm{T}}$ interval of the training is observed.

### 6.1.3 Working-point optimisation

The procedure used to define the prompt $\mathrm{D_s^+}$-meson candidate selection followed closely what was performed for the measurements in pp collisions, which is detailed in Section 5.1.3. It was executed independently for the 0–10% and 30–50% centrality classes. The only differences with respect to the studies carried-on in pp collisions were the necessity to account for the nuclear effects typical of Pb–Pb collisions, and the fact that only one threshold value was needed in each $p_{\mathrm{T}}$ interval, since binary BDT models were employed in this case.

The expected signal $S$ was calculated, considering the contribution of both prompt and non-prompt $\mathrm{D_s^+}$ mesons, as

$$S = R_{\mathrm{AA}} \langle T_{\mathrm{AA}} \rangle \left. \frac{\mathrm{d}^2\sigma^{\mathrm{D_s^+}}}{\mathrm{d}p_{\mathrm{T}}\mathrm{d}y} \right|^{\text{FONLL, prompt}}_{|y|<0.5} \times \frac{2(\mathrm{Acc} \times \epsilon)^{\text{prompt}} c_{\Delta y} \Delta p_{\mathrm{T}} \mathrm{BR} N_{\mathrm{evt}}}{f_{\text{prompt}}}. \quad (6.3)$$

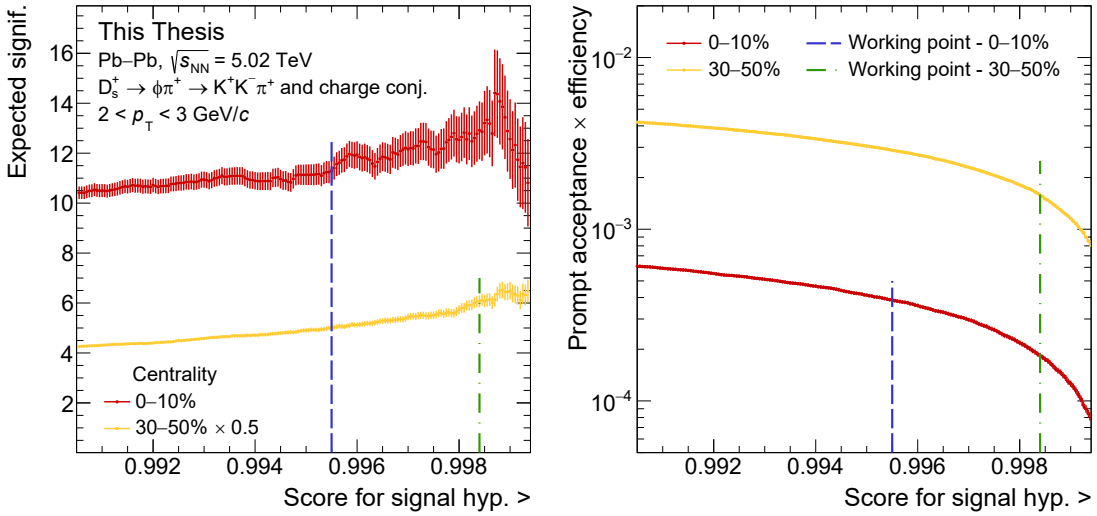



Figure 6.3: Expected significance (left panel) and prompt $D_s^+$-meson (Acc$\times\epsilon$) factor (right panel) as a function of the threshold on the BDT output in the $2 < p_T < 3$ GeV/$c$ interval for the 0–10% (red) and 30–50% (orange) centrality classes. The selected working points are shown by the dashed lines.

The procedure is analogous to the one adopted in pp collisions. Compared to Eq. (5.3), the integrated luminosity is replaced by the number of analysed events, and the FONLL calculations [72, 73] are multiplied by the average nuclear overlap function and by the $R_{AA}$ from TAMU model predictions [110]. The expected fraction of prompt $D_s^+$ mesons was computed from FONLL calculations according to Eq. (5.2), where the $p_T$-differential production cross sections of prompt and non-prompt $D_s^+$ mesons in pp collisions were multiplied by the corresponding nuclear modification factors predicted by the TAMU model. The (Acc $\times$ $\epsilon$) factors for prompt and non-prompt $D_s^+$ mesons were estimated from samples of MC simulations not used in the BDT training. The background under the signal peak $B$ was estimated, as described in Section 5.1.3, by fitting the $D_s^+$-meson invariant-mass distribution for a fraction of the collected data, and then by scaling the result to obtain the expectation value for the full data sample.

In Fig. 6.3, the expected significance $S/\sqrt{(S+B)}$ of the extracted signal, which contains prompt and non-prompt $D_s^+$ mesons, and the acceptance-time-efficiency factor for prompt $D_s^+$ mesons are reported as a function of the threshold value on the BDT output for the 0–10% and 30–50% centrality classes in the $2 < p_T < 3$ GeV/$c$ interval. The reported uncertainty on the expected significance is the statistical uncertainty due to the estimated combinatorial-background component. The final threshold values were chosen to provide a good statistical significance of the extracted signal, and a high selection efficiency and signal-to-background ratio. The

applied selections required a BDT score for the signal hypothesis above 0.9200-0.9985 and 0.9600-0.9984 for the 0–10% and 30–50% centrality classes, respectively, depending on $p_{\mathrm{T}}$. Threshold values very close to unity were necessary at low $p_{\mathrm{T}}$, since the large charged-particle multiplicity of Pb–Pb collisions leads to a huge combinatorial background, especially in central collisions. From Fig. 6.3, it is possible to notice that the chosen threshold value in the $2 < p_{\mathrm{T}} < 3$ GeV/$c$ interval is lower for central collisions than for semicentral ones, even if a larger combinatorial background is expected in the first case. This is due to the tighter topological and kinematic selections that were applied in central Pb–Pb collisions to prepare the data sample for the machine-learning step, which are reported in Tab. 6.1.

## 6.2 Raw-yield extraction

The $\mathrm{D_s^+}$-meson raw yields were extracted by fitting the $M(\mathrm{KK}\pi)$ invariant-mass distributions, which include both particles and antiparticles, resulting from the selections described in the previous Section. The raw yields were measured in 10 (9) transverse-momentum intervals spanning the $2 < p_{\mathrm{T}} < 50$ GeV/$c$ ($2 < p_{\mathrm{T}} < 36$ GeV/$c$) region for central (semicentral) Pb–Pb collisions. The same fit strategy as for the measurements performed in pp collisions was employed, which is detailed in Sections 5.2.1 and 5.3.1. The signal peak was described by a Gaussian function, while the background was modelled by an exponential term. An additional Gaussian function was introduced to account for the peak arising from the $\mathrm{D^+} \to \phi\pi^+ \to \mathrm{K^-K^+}\pi^+$ decay. The parameters of the $\mathrm{D_s^+}$-meson signal peak were left free in the fits. The only exception was the $2 < p_{\mathrm{T}} < 3$ GeV/$c$ interval for the 0–10% centrality class, where the peak width was fixed to the value obtained from MC simulations, to improve the signal extraction in the most challenging region for the measurement. In the other $p_{\mathrm{T}}$ intervals of the $\mathrm{D_s^+}$-meson measurement and for $\mathrm{D^0}$ and $\mathrm{D^+}$ mesons, it was verified that MC simulations describe the peak width observed in data within about ±10%.

Figure 6.4 shows the invariant-mass distributions for some representative $p_{\mathrm{T}}$ intervals of the 0–10% and 30–50% centrality classes. The fit functions and the resulting values for the mean, width, raw yield, and significance of the signal peak are also reported. The observed statistical significance in the $2 < p_{\mathrm{T}} < 3$ GeV/$c$ interval is lower than the expected one, reported in the left panel of Fig. 6.3, by a factor around 1.7 (1.4) for central (semicentral) collisions. This is caused by the large $\mathrm{D_s^+}$-meson $R_{\mathrm{AA}}$ predicted by the TAMU model, which overestimates the measurement in the low-$p_{\mathrm{T}}$ region, as will be shown in Fig. 6.17, and thus the expected significance.

In Fig. 6.5, the raw yield per event, the signal-to-background ratio, and the statistical significance of the extracted signal are displayed as a function of $p_{\mathrm{T}}$ for

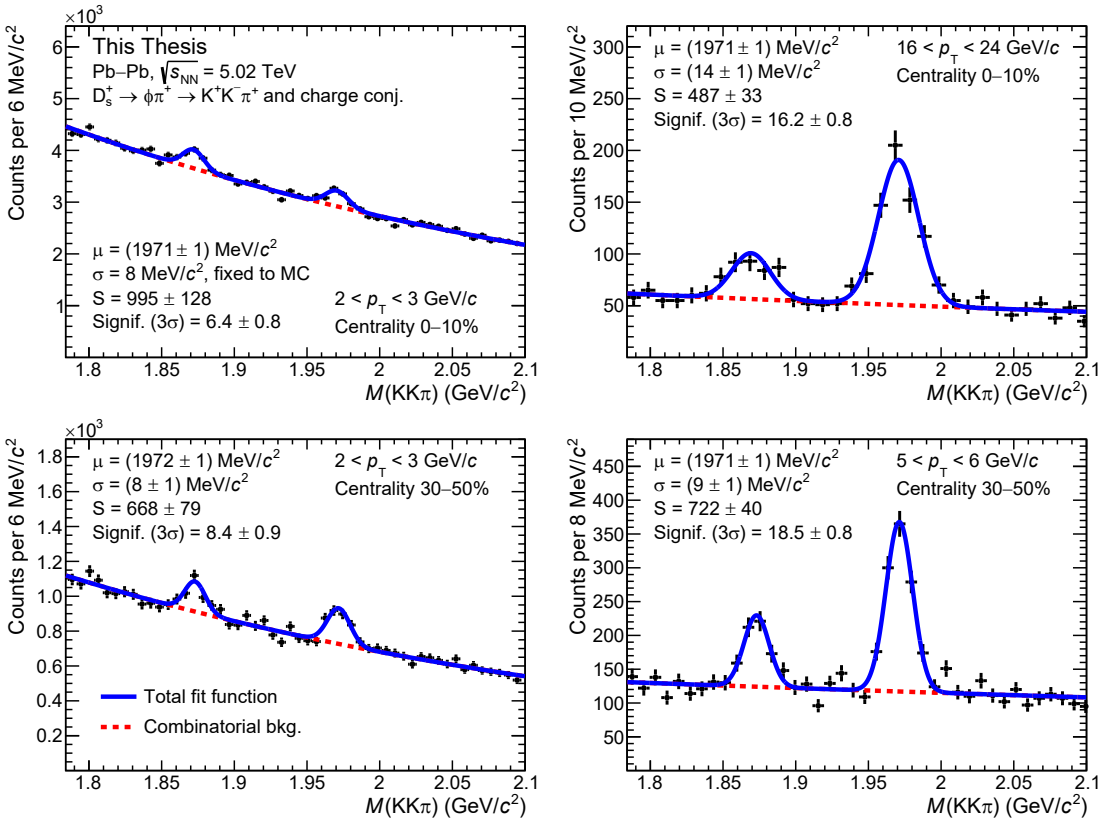



Figure 6.4: Invariant-mass distributions of $D_s^+$-meson candidates and their charge conjugates in representative $p_T$ intervals for the 0–10% (top panels) and 30–50% (bottom panels) centrality classes. The blue line shows the total fit function and the red dashed line is the combinatorial-background component. The values of the mean (μ), width ($\sigma$), raw yield (S), and significance (Signif.) of the signal peak are reported. The displayed uncertainties are only the fit statistical uncertainties.

central and semicentral Pb–Pb collisions. The raw yield per event is larger in the 0–10% centrality class than in the 30–50% one, despite the tighter selections applied, due to the larger number of nucleon–nucleon collisions occurring in the Pb–Pb interaction. The signal-to-background ratio is larger for semicentral collisions than for central ones for $p_T < 12$ GeV/$c$, due to the smaller combinatorial background present in more peripheral Pb–Pb collisions. An analogous trend is observed for the statistical significance at low transverse momentum, while at high $p_T$ the larger $D_s^+$-meson yield gives rise to larger significance values for the 0–10% centrality class.

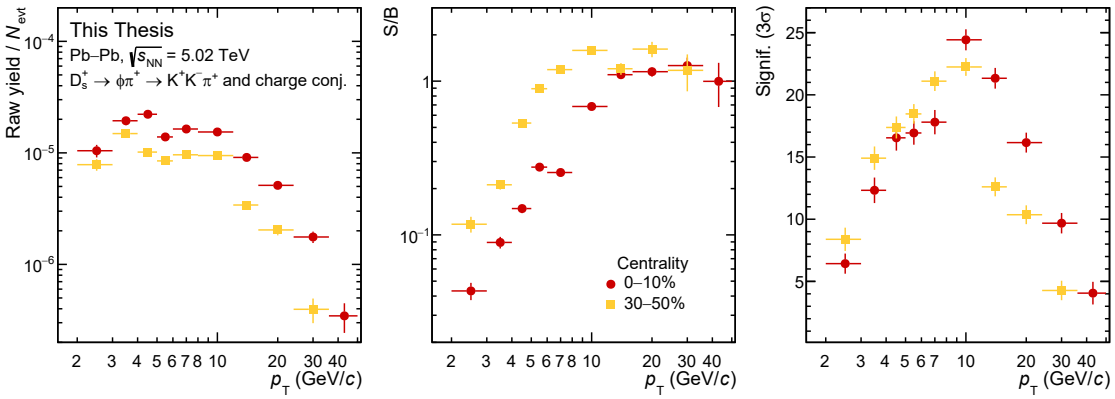



Figure 6.5: Raw yield per event (left panel), signal-to-background ratio (middle panel), and statistical significance (right panel) as a function of $p_{\mathrm{T}}$ for the 0–10% (red) and 30–50% (orange) centrality classes.

## 6.3 Efficiency and acceptance corrections

The acceptance-times-efficiency correction factors were computed from the MC simulations based on the PYTHIA 8 and HIJING event generators introduced in Section 4.3. To avoid possible biases, MC samples different from those employed in the BDT training were used. The $\mathrm{D_s^+}$-meson $p_{\mathrm{T}}$ distribution from simulations was weighted to match the shape given by FONLL calculations multiplied by the $\mathrm{D_s^+}$-meson $R_{\mathrm{AA}}$ predicted by the TAMU model [110]. This procedure accounted for hot-medium effects, as the charm-quark energy loss in the QGP and the charm-quark hadronisation via recombination, and produced more realistic transverse-momentum distributions for the determination of the (Acc × $\epsilon$) factors, which depend on $p_{\mathrm{T}}$.

In Fig. 6.6, the (Acc × $\epsilon$) correction factors as a function of $p_{\mathrm{T}}$ for prompt and non-prompt $\mathrm{D_s^+}$ mesons in the 0–10% and 30–50% centrality classes are reported. The $\mathrm{D_s^+}$-meson selection efficiency is higher for semicentral collisions than for central ones, by up to a factor two at low $p_{\mathrm{T}}$, since less stringent selections can be adopted thanks to the smaller combinatorial background. The non-prompt $\mathrm{D_s^+}$ mesons are more efficiently selected than prompt ones for $p_{\mathrm{T}} < 8$ GeV/$c$, given their on average larger displacement from the primary vertex. For $p_{\mathrm{T}} > 12$ GeV/$c$, where the candidate decay length is less relevant in discriminating signal from background, the BDT selections are able to suppress the non-prompt efficiency with respect to the prompt one.

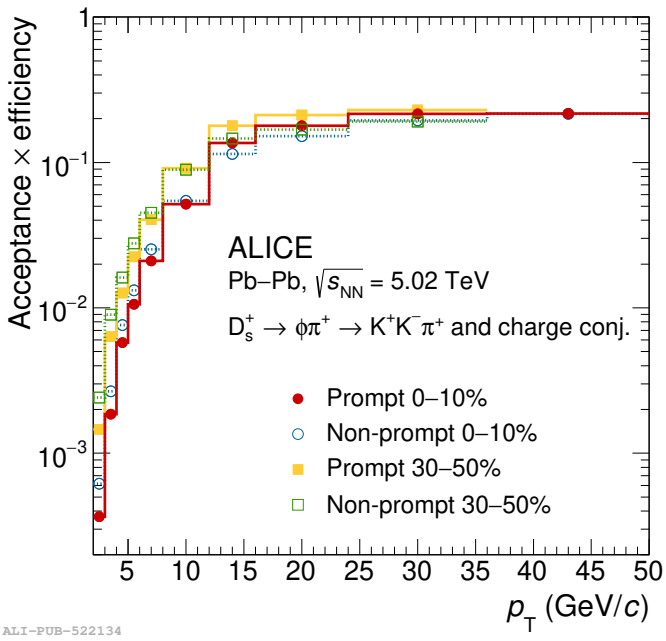



Figure 6.6: Acceptance-times-efficiency factors for $\mathrm{D_s^+}$ mesons as a function of $p_\mathrm{T}$. The (Acc $\times\,\epsilon$) factors for prompt (red) and non-prompt (blue) $\mathrm{D_s^+}$ mesons in Pb–Pb collisions for the 0–10% centrality class are shown, together with those for prompt (orange) and non-prompt (green) $\mathrm{D_s^+}$ mesons for the 30–50% centrality class. Figure from Ref. [193].

## 6.4 Fraction of prompt $\mathrm{D_s^+}$ mesons

The fraction of prompt $\mathrm{D_s^+}$ mesons, $f_\mathrm{prompt}$, in the measured raw yields was estimated using the procedure developed in previous studies of D-meson production in Pb–Pb collisions [104, 146, 274]. Similarly to what was performed for the prompt $\mathrm{D_s^+}$-meson measurement in pp collisions described in Section 5.3.3, the prompt fraction was computed as

$$\begin{aligned} f_\mathrm{prompt} &= 1 - \frac{N^{\mathrm{D_s^\pm,raw}}_\mathrm{non\text{-}prompt}}{N^{\mathrm{D_s^\pm,raw}}} = \\ &= 1 - \langle T_\mathrm{AA}\rangle\, R_\mathrm{AA}^\mathrm{non\text{-}prompt} \left.\frac{\mathrm{d}^2\sigma}{\mathrm{d}p_\mathrm{T}\mathrm{d}y}\right|^\mathrm{FONLL+PYTHIA\ 8}_\mathrm{non\text{-}prompt} \times \\ &\times \frac{2(\mathrm{Acc}\times\epsilon)^\mathrm{non\text{-}prompt} c_{\Delta y} \Delta p_\mathrm{T} \mathrm{BR} N^\mathrm{norm}_\mathrm{evt}}{N^{\mathrm{D_s^\pm,raw}}}. \end{aligned} \tag{6.4}$$

With respect to Eq. (5.12), the non-prompt $\mathrm{D_s^+}$-meson $p_\mathrm{T}$-differential cross section from FONLL and PYTHIA 8 calculations was multiplied by the average nuclear overlap function $\langle T_\mathrm{AA}\rangle$, and by the nuclear modification factor of non-prompt $\mathrm{D_s^+}$

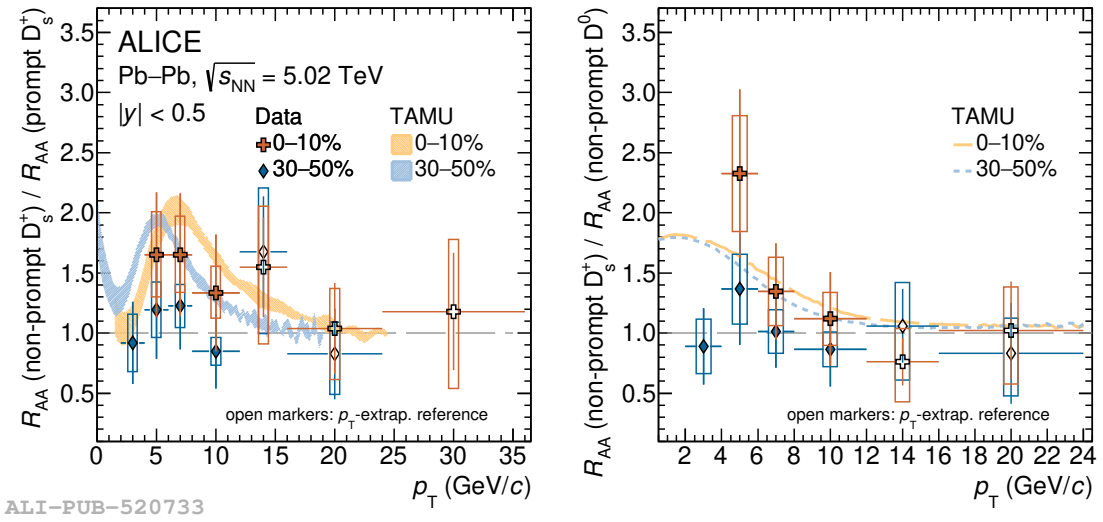



Figure 6.7: Non-prompt $\mathrm{D_s^+}$-meson $R_{\mathrm{AA}}$ divided by the one of prompt $\mathrm{D_s^+}$ mesons (left panel) and non-prompt $\mathrm{D^0}$ mesons [202] (right panel) in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV in the 0–10% and 30–50% centrality classes. The measurements are compared to TAMU model predictions [279]. Figure from Ref. [203].

mesons $R_{\mathrm{AA}}^{\text{non-prompt}}$. Moreover, the integrated luminosity $\mathcal{L}_{\mathrm{int}}$ was replaced by the number of analysed events $N_{\mathrm{evt}}$.

In the computation above, the nuclear modification factor of non-prompt $\mathrm{D_s^+}$ mesons was assumed to be equal to the one of prompt $\mathrm{D_s^+}$ mesons, $R_{\mathrm{AA}}^{\text{non-prompt}} = R_{\mathrm{AA}}^{\text{prompt}}$. Measurements of prompt and non-prompt $\mathrm{J/\psi}$ mesons in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV by the ATLAS and CMS Collaborations [106, 275], and the comparison of prompt non-strange D mesons [104] with non-prompt $\mathrm{D^0}$ mesons [202, 276] indicate a smaller suppression of non-prompt charmed hadrons than prompt ones. However, possible modifications of the heavy-flavour hadronisation mechanism in the QGP medium, together with the enhancement of the strange-quark production, are predicted to induce a higher $\mathrm{D_s^+}$-meson $R_{\mathrm{AA}}$ compared to D mesons without strange-quark content, and can affect differently charm- and beauty-strange hadrons [110, 114, 121, 141, 277, 278]. Finally, a recent measurement of the non-prompt $\mathrm{D_s^+}$-meson $R_{\mathrm{AA}}$ in central and semicentral Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV by the ALICE Collaboration [203] supports the adopted hypothesis. The $R_{\mathrm{AA}}$ ratios between non-prompt and prompt $\mathrm{D_s^+}$ mesons are displayed in the left panel of Fig. 6.7, where the prompt measurements are the ones obtained in this Thesis and shown in Fig. 6.16. The ratios are compatible with unity in all the measurement $p_{\mathrm{T}}$ range given the current statistical and systematic uncertainties, even if a hint of a larger $R_{\mathrm{AA}}$ for non-prompt $\mathrm{D_s^+}$ mesons than prompt ones is present in the 0–10% centrality class for $p_{\mathrm{T}} < 8$ GeV/$c$.

To estimate the systematic uncertainty of $f_{\mathrm{prompt}}$, which will be discussed in

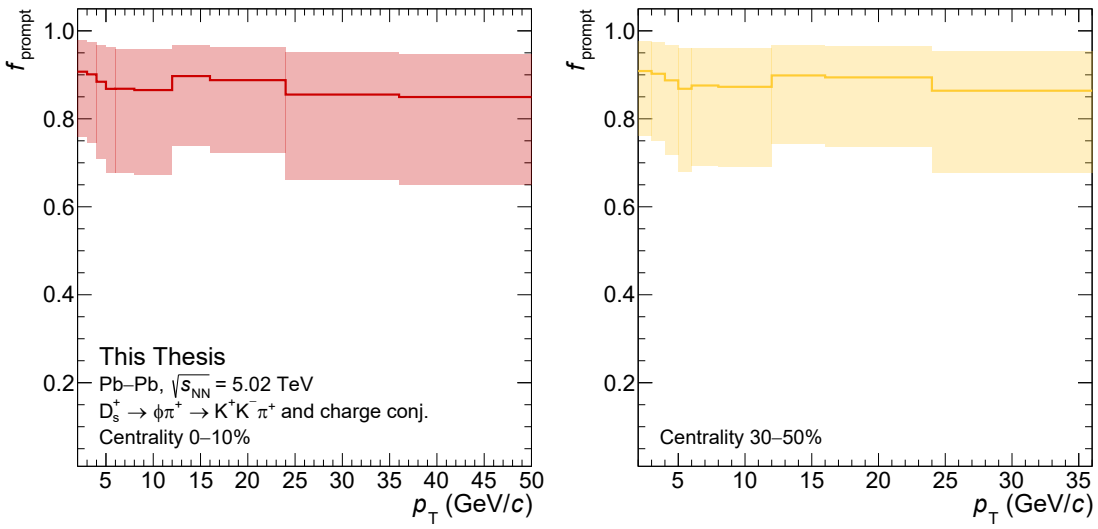



Figure 6.8: Fraction of prompt $\mathrm{D_s^+}$ mesons in the measured raw yields as a function of $p_{\mathrm{T}}$ for central (left panel) and semicentral (right panel) Pb–Pb collisions. The continuous lines show the $f_{\mathrm{prompt}}$ central values and the uncertainty is represented by the shaded boxes.

Sec. 6.6.3, the hypothesis for the $R_{\mathrm{AA}}$ of non-prompt $\mathrm{D_s^+}$ mesons was varied within 1/3 and 3 times the prompt $\mathrm{D_s^+}$-meson $R_{\mathrm{AA}}$, according to what performed in previous $\mathrm{D_s^+}$-meson measurements [146]. This variation range accounts for the model and the available data uncertainties, and well comprehends the results and the TAMU model predictions reported in Fig. 6.7. The estimated values of $f_{\mathrm{prompt}}$ are shown as a function of $p_{\mathrm{T}}$ in Fig. 6.8 together with their systematic uncertainty. They range between 0.86 and 0.91 depending on the transverse-momentum interval and the centrality class.

## 6.5 Proton-proton reference

The measurement of the $p_{\mathrm{T}}$-differential production cross section of prompt $\mathrm{D_s^+}$ mesons in pp collisions at $\sqrt{s} = 5.02$ TeV, which is presented in this Thesis in Chapter 5, was used as a reference for the $R_{\mathrm{AA}}$ calculation. The prompt $\mathrm{D_s^+}$-meson measurement in pp collisions reaches up to $p_{\mathrm{T}} = 24$ GeV/$c$. Therefore, it was extrapolated to the $24 < p_{\mathrm{T}} < 50$ GeV/$c$ interval using FONLL calculations. According to the procedure employed in Ref. [280], the extrapolation was performed scaling the FONLL predictions to match the measured values at lower transverse momenta. The scaling factor is computed by fitting the ratio between the measured $p_{\mathrm{T}}$-differential cross section and the predicted one with a constant function. The prompt $\mathrm{D_s^+}$-meson pp reference with the extrapolated points for the $24 < p_{\mathrm{T}} <$

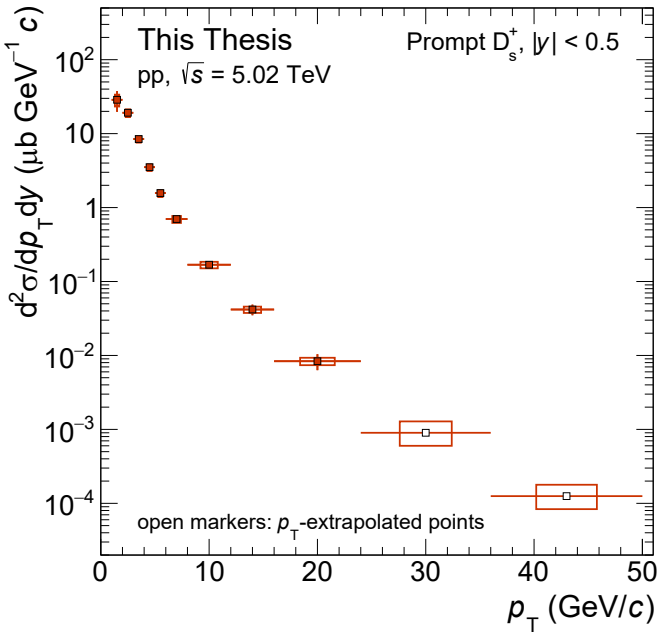



Figure 6.9: Prompt $D_s^+$-meson pp reference at $\sqrt{s} = 5.02$ TeV in the $|y| < 0.5$ interval, the open markers indicate the $p_T$-extrapolated points.

36 GeV/$c$ and $36 < p_T < 50$ GeV/$c$ intervals is reported in Fig. 6.9.

The systematic uncertainty on the extrapolated cross section was estimated by propagating the statistical and systematic uncertainties of the measurement to the scaling factor. In this calculation, the systematic uncertainties on the measurement were considered as fully correlated between the different $p_T$ intervals. The systematic uncertainty contribution due to the FONLL predictions was evaluated repeating the computation of the scaling factor varying the factorisation and renormalisation scales, the charm-quark mass and considering the PDF uncertainties. The resulting total systematic uncertainty on the $p_T$-extrapolated cross section is $^{+42}_{-33}$%, and it is the same in the two transverse-momentum intervals where the extrapolation was performed.

## 6.6 Systematic uncertainties

The measurements of the prompt $D_s^+$-meson corrected yields and $R_{AA}$ are affected by the following sources of systematic uncertainties: (i) the raw-yield extraction, (ii) the BDT-selection and PID efficiencies, (iii) the prompt fraction estimation, (iv) the generated $D_s^+$-meson $p_T$ shape in the MC simulation, (v) the track-reconstruction efficiency, and (vi) the normalisation. The majority of these systematic-uncertainty sources are the same as those affecting the measurement of

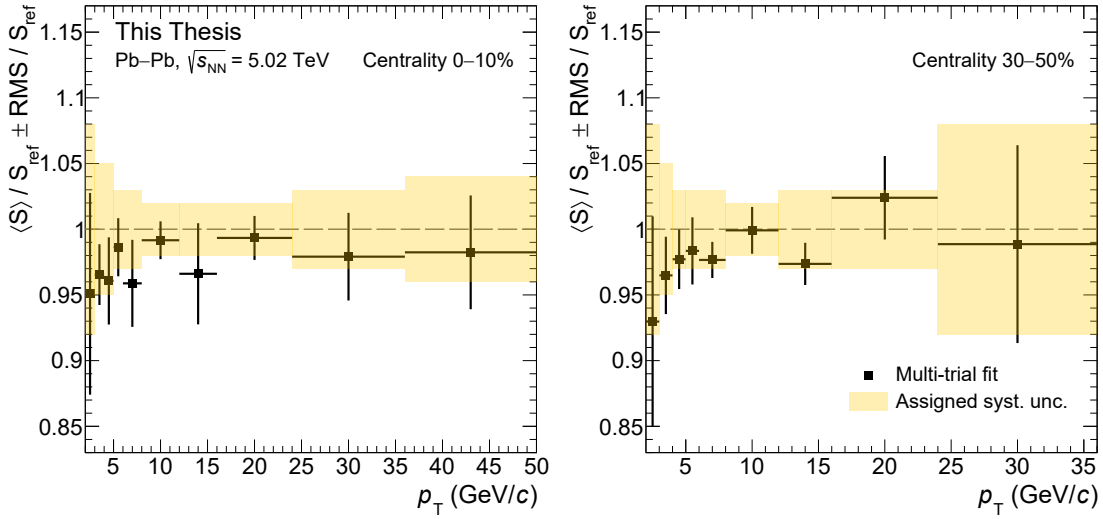



Figure 6.10: Mean and RMS of the multi-trial raw-yield distribution divided by the reference value as a function of $p_T$ for the 0–10% (left panel) and 30–50% (right panel) centrality classes. The yellow band indicates the assigned systematic uncertainty.

the $D_s^+$-meson cross section in pp collisions, which are detailed in Section 5.4. Thus, this Section will be focused on the differences with respect to the analysis in pp collisions, and on the dependence of the systematic uncertainties on the centrality class.

### 6.6.1 Raw-yield extraction

The systematic uncertainty affecting the raw-yield extraction was studied with a multi-trial approach, i.e., by repeating the fit of the invariant-mass distribution varying the functional form of the background fit function (linear, exponential, and parabolic), the lower and upper limits of the fit range, and the bin width. The width of the Gaussian function used to describe the signal peak was a free parameter of the fit, unless for the $2 < p_T < 3$ GeV/$c$ interval in the 0–10% centrality class, where the Gaussian width was fixed to the one obtained from MC simulations, as mentioned in Section 6.2. In this latter case the multi-trial procedure was also repeated by fixing the width to the value resulting from simulations ±10%, which approximatively corresponds to the maximum difference between MC simulations and data observed in the other $p_T$ intervals.

The systematic uncertainty was assigned considering the RMS of the raw-yield distribution, obtained from the different fit configurations, and the shift of its mean from the reference value when relevant. In Fig. 6.10, the mean of the raw-yield distribution divided by the raw-yield extracted using the default fit configuration is

reported as a function of $p_{\mathrm{T}}$ for the 0–10% and 30–50% centrality classes. The error bars show the RMS of the raw-yield distributions while the yellow band indicates the assigned systematic uncertainty. The relative systematic uncertainty ranges from 2% to 8% depending on $p_{\mathrm{T}}$. Large uncertainties characterise the lowest and highest $p_{\mathrm{T}}$ intervals where the raw-yield extraction is more challenging, due to the large combinatorial background in the first case and to the small $\mathrm{D_s^+}$-meson signal in the latter one. The assigned systematic-uncertainty values are similar in the two centrality classes considered in the analysis.

### 6.6.2 BDT-selection and PID efficiency

The imperfections in the description of the detector resolutions and alignments in the MC simulations can result in a systematic effect on the selection efficiency. The associated systematic uncertainty was estimated with a procedure analogous to the one described in Section 5.4.2, i.e., by performing the analysis with different BDT-selection criteria and by comparing the resulting corrected yields. However, in this case, the various BDT-selection criteria were obtained varying only a single threshold, given that a binary classification was performed. Tighter and looser thresholds on the BDT output were considered, which resulted in up to 50% higher and lower efficiencies with respect to the central values.

The systematic uncertainty was assigned considering the sum in quadrature of the RMS and shift from unity of the distribution of corrected-yield ratios, which were computed dividing the values resulting from the tested selection criteria by the reference one. Figure 6.11 reports the sum in quadrature of the RMS and shift from unity of the resulting corrected-yield distributions as a function $p_{\mathrm{T}}$ for the 0–10% and 30–50% centrality classes. The systematic uncertainty values were assigned based on the RMS and shift after smoothing out the values looking at the neighbour $p_{\mathrm{T}}$ intervals. For both centralities, the uncertainty estimation in the high-$p_{\mathrm{T}}$ region is affected by fluctuations in the raw-yield extraction caused by the smaller $\mathrm{D_s^+}$-meson yield. Therefore, it was not considered in the systematic-uncertainty assignment, in fact a larger systematic uncertainty is not expected at high transverse momenta where the selection criteria are less stringent. The assigned systematic uncertainty ranges from 3% to 9% decreasing with increasing $p_{\mathrm{T}}$, and it is smaller in the 30–50% centrality class than in the 0–10% one due to the looser selections applied.

Possible systematic effects arising from the conservative PID selection, applied to prepare the data sample for the machine-learning model training and application, were investigated comparing the PID selection efficiencies for pions and kaons in data and in simulations. For this study, a pure sample of pions was selected from $\Lambda$ and $\mathrm{K_s^0}$ decays, while samples of kaons in the TPC (TOF) were obtained applying a strict PID selection using the TOF (TPC) information. No significant discrepancies were noticed, in agreement to what observed for the measurement of

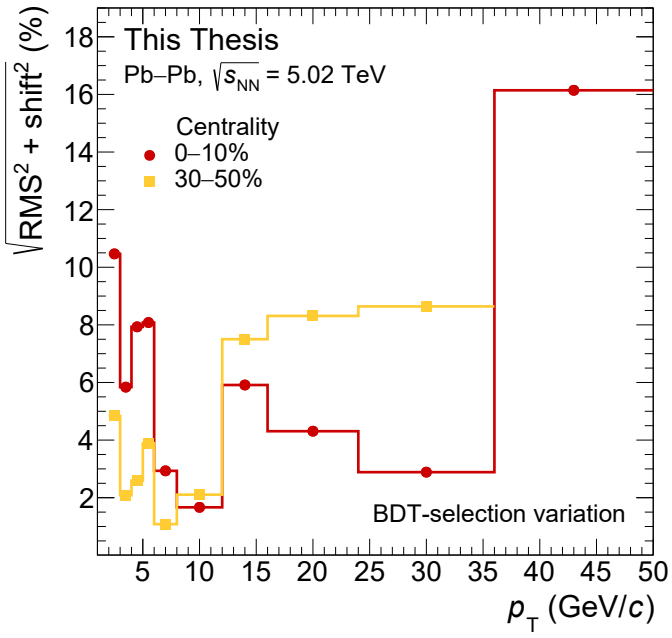



Figure 6.11: Sum in quadrature of the RMS and shift from unity of the distribution of corrected-yield ratios with respect to the reference value as a function $p_{\mathrm{T}}$ for the 0–10% (red) and 30–50% (orange) centrality classes.

non-strange D mesons in the same data sample [104]. In addition, possible systematic effects arising from the use of the PID information as input to the BDT models are already accounted for by the variation of the BDT-selection criteria. Therefore, no systematic uncertainty was assigned.

### 6.6.3 Prompt fraction estimation

The procedure used to estimate the $f_{\mathrm{prompt}}$ fraction described in Section 6.4 is affected by two sources of systematic uncertainty. The first one is the uncertainty due to the FONLL theoretical calculations, and it is common to the measurement of the prompt $\mathrm{D_s^+}$-meson cross section in pp collisions. This contribution was evaluated by computing the $f_{\mathrm{prompt}}$ fraction varying the FONLL parameters for what concerns the beauty-quark mass, the factorisation and renormalisation scales, and considering the PDF uncertainties. More details are reported in Section 5.3.3.

The second source of systematic uncertainty arises from the assumption on the non-prompt $\mathrm{D_s^+}$-meson nuclear modification factor. This contribution was estimated by varying the $R_{\mathrm{AA}}^{\text{non-prompt}}/R_{\mathrm{AA}}^{\text{prompt}}$ ratio within the uncertainties of the measurements and model predictions considered in the choice of the central value, i.e., $R_{\mathrm{AA}}^{\text{non-prompt}}/R_{\mathrm{AA}}^{\text{prompt}} = 1$. As mentioned in Section 6.4, the hypothesis on the $R_{\mathrm{AA}}$ ratio was varied between 1/3 and 3 according to what performed in previous

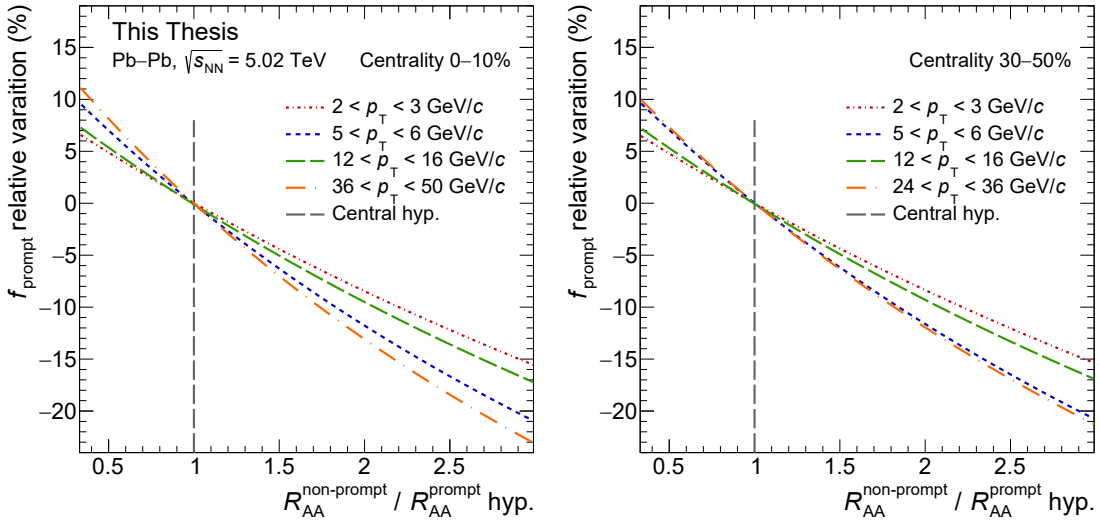



Figure 6.12: Relative variation of the $f_{\mathrm{prompt}}$ fraction as a function of the hypothesis on the $R_{\mathrm{AA}}^{\mathrm{non\text{-}prompt}}/R_{\mathrm{AA}}^{\mathrm{prompt}}$ ratio in representative $p_{\mathrm{T}}$ intervals in the 0–10% (left panel) and 30–50% (right panel) centrality classes.

$\mathrm{D_s^+}$-meson measurements [146]. Figure 6.12 displays the relative variation of the prompt $\mathrm{D_s^+}$-meson fraction as a function of the $R_{\mathrm{AA}}^{\mathrm{non\text{-}prompt}}/R_{\mathrm{AA}}^{\mathrm{prompt}}$ hypothesis for some representative $p_{\mathrm{T}}$ intervals in the 0–10% and 30–50% centrality classes. The magnitude of the $f_{\mathrm{prompt}}$ variation does not show a monotonous $p_{\mathrm{T}}$ dependence, and it is similar between the two analysed centrality classes.

The assigned systematic uncertainty was obtained by summing in quadrature the uncertainty values resulting from the two sources described above, and it ranges between $^{+8}_{-16}$% and $^{+12}_{-23}$% depending on the $p_{\mathrm{T}}$ interval and the centrality class. When computing the $R_{\mathrm{AA}}$, the source of prompt-fraction uncertainty due to the variation of the FONLL parameters was considered to be fully correlated between the measurements in pp and Pb–Pb collisions. Therefore, the two contributions partially cancel out in the ratio of Eq. (6.2).

### 6.6.4 Generated Monte Carlo $p_{\mathrm{T}}$ shape

Differences between the $\mathrm{D_s^+}$-meson $p_{\mathrm{T}}$ distribution assumed in Section 6.3 and the real one can lead to systematic effects on the selection-efficiency correction, since the selection efficiency varies with transverse momentum and the analysis is performed in intervals with a finite $p_{\mathrm{T}}$ width. This systematic-uncertainty contribution was estimated by calculating the selection efficiency using alternative $\mathrm{D_s^+}$-meson $p_{\mathrm{T}}$ shapes than the one adopted for the final result, which was obtained multiplying the shape given by FONLL calculations by the prompt $\mathrm{D_s^+}$-meson $R_{\mathrm{AA}}$ predicted by

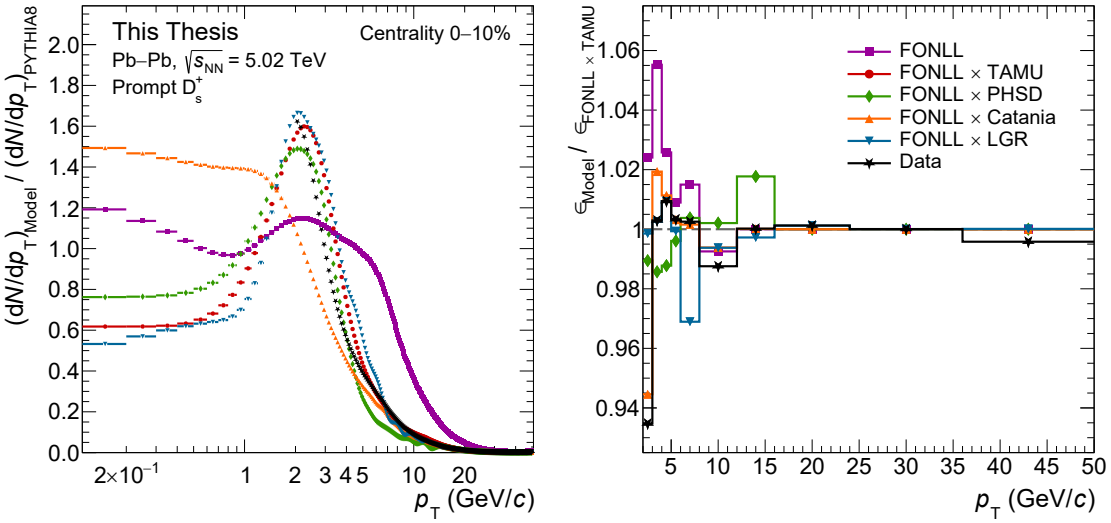



Figure 6.13: Left panel: normalised $p_T$-differential yields from FONLL calculations and $R_{AA}$ predictions of different models divided by the one from PYTHIA 8 simulations for the 0–10% centrality class. Right panel: ratios between the efficiency computed using the $p_T$ shapes provided by the alternative models and the one from FONLL calculations multiplied by the $R_{AA}$ predicted by the TAMU model. The values resulting from the fit to the measured yield are shown by the black stars.

the TAMU model [110]. The FONLL predictions for the prompt $D_s^+$-meson production were obtained using the recipe described in Section 5.1.3. The alternative $p_T$ shapes were computed from the FONLL $p_T$ distribution and the $R_{AA}$ predicted by theoretical models with different approaches to the description of the charm-quark propagation and hadronisation in the QGP medium. The models considered were LGR [114], PHSD [277], and Catania [115]. The $R_{AA}$ predictions were extrapolated to transverse momenta above the model reach when necessary, using the last available point. As an extreme variation, also the case of no medium effects was taken into account by adopting the FONLL transverse-momentum shape without further parametrisations.

The $p_T$ shapes obtained using FONLL calculations and different model predictions for the $R_{AA}$ divided by the one from PYTHIA 8 simulations are shown in the left panel of Fig. 6.13, for the 0–10% centrality class. In the right panel of Fig. 6.13, the ratios of the prompt $D_s^+$-meson efficiencies obtained using the alternative $p_T$ shapes and the one employed for the final result are displayed for the 0–10% centrality class. The resulting uncertainty was estimated considering the discrepancy of the efficiency ratios in each $p_T$ interval. Higher uncertainty values were observed in central collisions than in semicentral ones due to the tighter selections applied. The assigned systematic uncertainty is 5% (3%) for the 0–10% (30–50%) centrality class in the lowest $p_T$ interval, where the efficiency varies steeply with transverse

momentum, and decreases to zero above 12 (8) GeV/$c$.

As a cross-check, the measured $p_{\mathrm{T}}$ shape was extracted by fitting the prompt $\mathrm{D_s^+}$-meson $p_{\mathrm{T}}$-differential yield, which is shown in Fig. 6.15, with a power-law function. It was then used to weight the prompt $\mathrm{D_s^+}$-meson efficiency resulting from simulations. In the left panel of Fig. 6.13, the $p_{\mathrm{T}}$ shape obtained for the 0–10% centrality class divided by the one from PYTHIA 8 simulations is shown. The ratio between the corresponding efficiency and the one used for the final result is reported in the right panel of the Figure. The observed discrepancies of the efficiency ratios are similar to those obtained employing the different model predictions, and are within the assigned systematic-uncertainty values for both central and semicentral Pb–Pb collisions. The only exception is the $2 < p_{\mathrm{T}} < 3$ GeV/$c$ interval in the 0–10% centrality class, where the discrepancy is slightly larger than the assigned uncertainty value. However, this $p_{\mathrm{T}}$ interval is at the edge of the measured transverse-momentum range, thus the $p_{\mathrm{T}}$-shape extracted from the fit is not completely trustable. For this reason, no further uncertainty was assigned.

### 6.6.5 Track-reconstruction efficiency

The systematic uncertainty on the tracking efficiency was computed following the same procedure adopted for the measurements in pp collisions and detailed in Section 5.4.4. Furthermore, it was estimated independently for the two centrality classes considered in this analysis. Two different sources contribute to the systematic uncertainty, i.e., the track-quality criteria applied and the prolongation of the TPC tracks to the ITS.

To estimate the contribution due to the first source, the track-quality selections listed in Section 4.2.1 were varied, and for each configuration the prompt $\mathrm{D_s^+}$-meson corrected yield was computed. The alternative track-quality selection criteria considered were: (i) a number of crossed rows in the TPC larger than $120 - 5/(p_{\mathrm{T}}\ [\mathrm{GeV}/c])$ instead of 70 which is the default value, (ii) a ratio of crossed rows over findable clusters in the TPC larger than 0.9, while the default value is 0.8, (iii) a number of TPC clusters larger than 0.65 times the TPC crossed-row number, and (iv) a number of clusters available for the particle identification in the TPC larger than 40 or 60, respectively, instead of 50 which is the default value. A variation of the corrected yield of about 3.5% and 2% was observed for the 0–10% and 30–50% centrality classes, respectively, independently of $p_{\mathrm{T}}$. The observed values were slightly smaller than those obtained for the measurement of non-strange D mesons in the same data sample [104], where an uncertainty value of 4.5% (3%) in central (semicentral) events was adopted for D mesons decaying into three charged particles. These latter values were assigned as systematic uncertainty due to the higher statistical precision of the non-strange D-meson measurement with respect to the $\mathrm{D_s^+}$-meson one.

The second source of systematic uncertainty was studied by comparing the

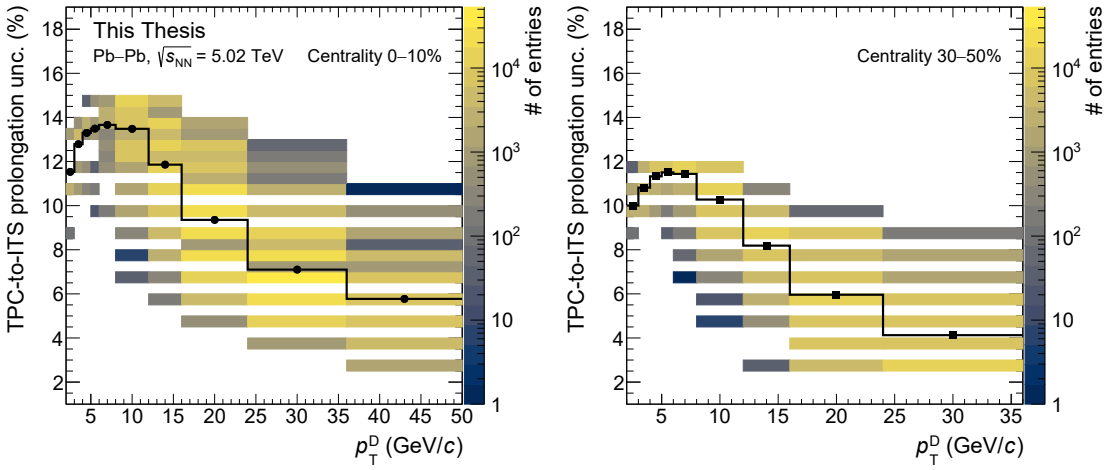



Figure 6.14: Relative systematic uncertainty on the TPC-to-ITS prolongation probability as a function of the $D_s^+$-meson $p_T$ for the 0–10% (left panel) and 30–50% (right panel) centrality classes. The average value in each transverse-momentum interval is shown by the filled markers.

TPC-to-ITS prolongation probability of tracks in data and simulation. As discussed in Section 5.4.4, the comparison was performed after weighting the relative abundances of primary and secondary particles in the simulation to match those observed in data. The resulting per-track systematic uncertainty was propagated to the $D_s^+$ mesons via their decay kinematics, summing linearly the contribution of each track. In Fig. 6.14, the distributions of the relative uncertainty values for the $D_s^+$-meson candidates are reported as a function of $p_T$ for the 0–10% and 30–50% centrality classes, together with their average in each $p_T$ interval. The average relative systematic uncertainty on the TPC-to-ITS prolongation probability was summed in quadrature to the track-quality contribution to obtain the final values. The relative systematic uncertainty due to the tracking ranges between 5% and 14% depending on the $p_T$ interval and the centrality class. Larger uncertainty values were assigned for central Pb-Pb collisions than for semicentral ones.

### 6.6.6 Normalisation

The systematic uncertainty on the normalisation of the $D_s^+$-meson corrected yields contains a contribution of 3.6% due to the uncertainty on the BR of the $D_s^+ \to \phi\pi^+ \to K^-K^+\pi^+$ decay channel used to measure the $D_s^+$ mesons [6]. An additional contribution is due to the uncertainty on the hadronic cross-section fraction used in the Glauber fit to define the centrality intervals, and it was estimated to be $< 0.1\%$ and 2% for the 0–10% and 30–50% centrality classes, respectively [280].

In the case of the nuclear-modification-factor measurement, the normalisation

uncertainty is the quadratic sum of: the uncertainty due to the centrality-interval definition mentioned above, the integrated-luminosity uncertainty on the pp reference amounting to 2.1% [233], and the $\langle T_{\mathrm{AA}} \rangle$ uncertainty corresponding to 0.7% (1.5%) for the 0–10% (30–50%) centrality class [22].

### 6.6.7 Systematic uncertainty summary

The summary of the systematic uncertainties on the measured corrected yields are reported in Tab. 6.4 for the 0–10% and 30–50% centrality classes. The total systematic uncertainty resulting from the sum in quadrature of the various contributions is also shown. The systematic uncertainties of the measurement were considered to be fully correlated between the different $p_{\mathrm{T}}$ intervals, except for the raw-yield extraction uncertainty. This last contribution was considered as uncorrelated due to the relevant variations of the signal-to-background ratio and the shape of the combinatorial-background distribution as a function of $p_{\mathrm{T}}$.

Table 6.5 reports the total systematic uncertainty on the measurement of the nuclear modification factor for the 0–10% and 30–50% centrality classes, together with the contribution due to the normalisation. The total systematic-uncertainty values were computed propagating the uncertainties on the prompt $\mathrm{D_s^+}$-meson corrected yields and on the pp reference cross section according to Eq. 6.2. The BR uncertainty cancels out in the ratio, thus was not considered in the calculation. Moreover, as discussed in Section 6.6.3, the contribution to the prompt-fraction uncertainty due to the variation of the FONLL parameters was propagated as fully correlated. The other systematic uncertainties were treated as uncorrelated between the Pb–Pb corrected yields and the pp cross section.

## 6.7 Results

In Fig. 6.15, the $p_{\mathrm{T}}$-differential corrected yields of prompt $\mathrm{D_s^+}$ mesons measured in central and semicentral Pb–Pb collisions are reported. The results are compared with the $\mathrm{D_s^+}$-meson pp reference cross section, introduced in Section 6.5, scaled by the average nuclear overlap function $\langle T_{\mathrm{AA}} \rangle$ for the corresponding centrality class. The $\mathrm{D_s^+}$-meson yields are strongly suppressed with respect to the binary-scaled pp reference for $p_{\mathrm{T}} > 3$–4 GeV/$c$, due to the modification of the charm-quark $p_{\mathrm{T}}$ spectra induced by the interactions with the QGP medium. A similar suppression is observed also for non-strange D mesons [104]. Compared to the previous measurement by the ALICE Collaboration in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV [146], the production yield of $\mathrm{D_s^+}$ mesons was measured down to lower transverse momentum and in finer $p_{\mathrm{T}}$ intervals. Moreover, the statistical uncertainties are smaller by a factor ranging from about 2.2 to 4.5 depending on the $p_{\mathrm{T}}$ interval and the centrality class, while the systematic uncertainties are similar between

Table 6.4: Summary of the relative systematic uncertainties of the prompt $\mathrm{D_s^+}$-meson $p_\mathrm{T}$-differential corrected yield in Pb–Pb collisions at $\sqrt{s_\mathrm{NN}} = 5.02$ TeV in the 0–10% and 30–50% centrality classes.

| $p_\mathrm{T}$ (GeV/$c$) | 2–3 | 3–4 | 4–5 | 5–6 | 6–8 | 8–12 | 12–16 | 16–24 | 24–36 | 36–50 |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0–10% centrality class | | | | | | | | | |
| Yield extraction (%) | 8.0 | 5.0 | 5.0 | 3.0 | 3.0 | 2.0 | 2.0 | 2.0 | 3.0 | 4.0 |
| Selection efficiency (%) | 9.0 | 6.0 | 6.0 | 6.0 | 4.0 | 4.0 | 4.0 | 4.0 | 4.0 | 4.0 |
| Prompt fraction (%) | $^{+7.8}_{-16.4}$ | $^{+8.0}_{-17.3}$ | $^{+9.4}_{-19.6}$ | $^{+10.7}_{-22.0}$ | $^{+10.4}_{-21.9}$ | $^{+10.8}_{-22.2}$ | $^{+7.8}_{-17.6}$ | $^{+8.4}_{-18.6}$ | $^{+11.2}_{-22.8}$ | $^{+11.5}_{-23.4}$ |
| MC $p_\mathrm{T}$ shape (%) | 5.0 | 3.0 | 3.0 | 1.0 | 1.0 | 1.0 | negl. | negl. | negl. | negl. |
| Tracking efficiency (%) | 12.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 13.0 | 10.0 | 9.0 | 7.0 |
| Branching ratio (%) | | | | | | 3.6 | | | | |
| Centrality limits (%) | | | | | | $< 0.1$ | | | | |
| Total (%) | $^{+19.7}_{-24.4}$ | $^{+18.5}_{-24.0}$ | $^{+19.2}_{-25.8}$ | $^{+19.2}_{-27.2}$ | $^{+18.5}_{-26.8}$ | $^{+18.6}_{-26.9}$ | $^{+16.2}_{-22.6}$ | $^{+14.3}_{-21.9}$ | $^{+15.6}_{-25.2}$ | $^{+15.0}_{-25.4}$ |
| | 30–50% centrality class | | | | | | | | | |
| Yield extraction (%) | 8.0 | 5.0 | 3.0 | 3.0 | 3.0 | 2.0 | 3.0 | 3.0 | 8.0 | — |
| Selection efficiency (%) | 5.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | — |
| Prompt fraction (%) | $^{+7.5}_{-16.1}$ | $^{+7.9}_{-17.0}$ | $^{+9.0}_{-19.2}$ | $^{+10.6}_{-21.8}$ | $^{+9.8}_{-20.8}$ | $^{+10.0}_{-20.9}$ | $^{+7.6}_{-17.3}$ | $^{+7.9}_{-17.7}$ | $^{+10.3}_{-21.6}$ | — |
| MC $p_\mathrm{T}$ shape (%) | 3.0 | 3.0 | 2.0 | 1.0 | 1.0 | negl. | negl. | negl. | negl. | — |
| Tracking efficiency (%) | 10.0 | 11.0 | 12.0 | 12.0 | 12.0 | 11.0 | 9.0 | 7.0 | 5.0 | — |
| Branching ratio (%) | | | | | | 3.6 | | | | |
| Centrality limits (%) | | | | | | 2.0 | | | | |
| Total (%) | $^{+16.5}_{-21.8}$ | $^{+15.6}_{-21.7}$ | $^{+16.2}_{-23.4}$ | $^{+17.1}_{-25.6}$ | $^{+16.6}_{-24.7}$ | $^{+15.8}_{-24.3}$ | $^{+13.1}_{-20.4}$ | $^{+12.1}_{-19.9}$ | $^{+14.8}_{-24.1}$ | — |

the two measurements. These improvements were possible thanks to the adoption of machine-learning selections and to the larger data sample collected by the experiment.

In Fig. 6.16, the nuclear modification factor of prompt $\mathrm{D_s^+}$ mesons is compared to the average $R_\mathrm{AA}$ of prompt $\mathrm{D^0}$, $\mathrm{D^+}$, and $\mathrm{D^{*+}}$ mesons in the 0–10% and 30–50% centrality classes from Ref. [104]. The systematic uncertainties related to the tracking efficiency and the prompt-fraction estimation are considered as fully correlated between the different D-meson species, and are reported separately from the other sources of systematic uncertainty which are uncorrelated. Both nuclear modification factors show a minimum value of about 0.2 (0.4) for $p_\mathrm{T} \approx 10$ GeV/$c$ in central (semicentral) Pb–Pb collisions. For $p_\mathrm{T} < 10$ GeV/$c$, the $R_\mathrm{AA}$ increases with decreasing $p_\mathrm{T}$ and reaches about unity for $p_\mathrm{T} \approx 2$–3 GeV/$c$. At $p_\mathrm{T} \gtrsim 10$ GeV/$c$, the nuclear

Table 6.5: Total relative systematic uncertainty of the prompt $\mathrm{D_s^+}$-meson $p_{\mathrm{T}}$-differential $R_{\mathrm{AA}}$ in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV in the 0–10% and 30–50% centrality classes. The normalisation uncertainty is also reported.

| $p_{\mathrm{T}}$ (GeV/$c$) | 2–3 | 3–4 | 4–5 | 5–6 | 6–8 | 8–12 | 12–16 | 16–24 | 24–36 | 36–50 |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0–10% centrality class | | | | | | | | | |
| Normalisation (%) | 2.2 | | | | | | | | | |
| Total (%) | $^{+20.9}_{-25.2}$ | $^{+19.2}_{-24.3}$ | $^{+19.8}_{-25.9}$ | $^{+20.0}_{-27.4}$ | $^{+19.7}_{-27.1}$ | $^{+19.6}_{-27.3}$ | $^{+17.9}_{-23.8}$ | $^{+21.0}_{-26.6}$ | $^{+36.7}_{-49.3}$ | $^{+36.8}_{-49.6}$ |
| | 30–50% centrality class | | | | | | | | | |
| Normalisation (%) | 3.3 | | | | | | | | | |
| Total (%) | $^{+18.0}_{-22.7}$ | $^{+16.5}_{-22.1}$ | $^{+17.2}_{-23.8}$ | $^{+18.2}_{-25.9}$ | $^{+18.0}_{-25.3}$ | $^{+17.3}_{-24.9}$ | $^{+15.3}_{-21.7}$ | $^{+19.6}_{-25.0}$ | $^{+36.5}_{-48.8}$ | — |

modification factors of prompt $\mathrm{D_s^+}$ and non-strange D mesons are compatible within uncertainties in both centrality classes. In this $p_{\mathrm{T}}$ region, the hadronisation is expected to be dominated by the in-vacuum fragmentation mechanism, and the most relevant effect leading to the observed suppression is the charm-quark energy loss in the medium. Therefore, the $R_{\mathrm{AA}}$ of heavy- and light-flavour hadrons are similar, and the relative abundances of the different hadronic species are not modified compared to pp collisions. At lower $p_{\mathrm{T}}$, the prompt $\mathrm{D_s^+}$-meson $R_{\mathrm{AA}}$ is systematically higher than the non-strange D-meson one. However, they are compatible within about one standard deviation considering the combination of statistical and systematic uncertainties.

The $R_{\mathrm{AA}}$ of prompt $\mathrm{D_s^+}$ and non-strange D mesons in the 0–10% and 30–50% centrality classes are compared with theoretical calculations [110, 114, 115, 121, 277] in Fig. 6.17. All the models considered implement the charm-quark transport in the QGP and include an enhancement of the production of $\mathrm{s\bar{s}}$ pairs in the deconfined medium. The charm-quark hadronisation is implemented both via recombination with light quarks of the QGP medium, dominant at low $p_{\mathrm{T}}$, and via fragmentation, which is dominant at high transverse momentum. The description of the D-meson $p_{\mathrm{T}}$ spectra in pp collisions, which is necessary for the $R_{\mathrm{AA}}$ determination, is based on FONLL calculations of the charm-quark $p_{\mathrm{T}}$ spectra [60, 72, 73] for all the models. The PHSD and LGR models adopt only the charm-quark hadronisation via fragmentation in pp collisions, while the Catania model includes also the mechanism of the hadronisation via coalescence [281]. Instead, the TAMU model describes the hadronisation in pp collisions with a statistical approach, in which the strangeness production is suppressed with respect to heavy-ion collisions introducing a suppression factor for strange particles $\gamma_{\mathrm{s}} = 0.6$ [282].

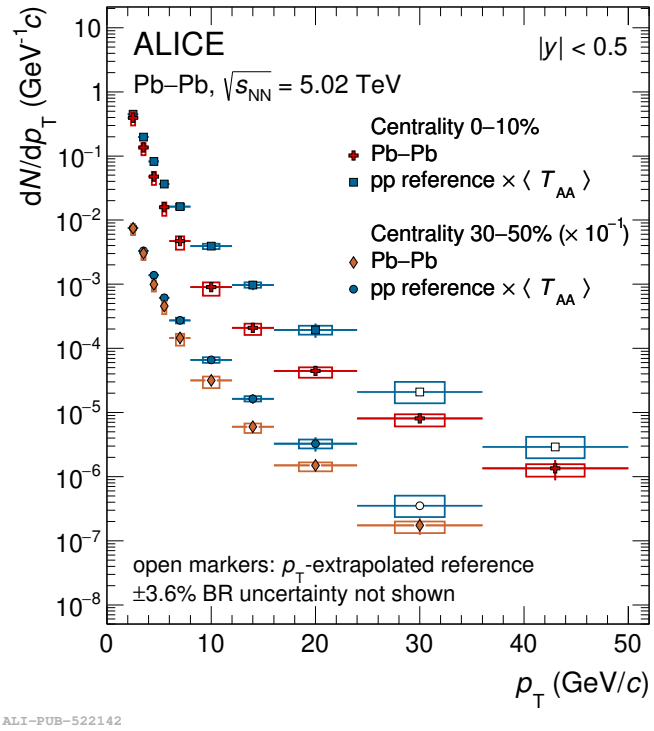



Figure 6.15: $p_{\mathrm{T}}$-differential production yields of prompt $\mathrm{D_s^+}$ mesons in the 0–10% and 30–50% centrality classes in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV compared to the pp reference scaled by the corresponding centrality-class $\langle T_{\mathrm{AA}} \rangle$. The open markers indicate where the pp reference is extrapolated. The measurements in the 30–50% centrality class are divided by a factor of ten for better visibility. Statistical (bars) and systematic (boxes) uncertainties are shown. Figure from Ref. [193].

The considered models predict a smaller suppression of the nuclear modification factor of strange D mesons compared to non-strange D mesons. Moreover, they reproduce qualitatively the $p_{\mathrm{T}}$ trend of the measured $R_{\mathrm{AA}}$ of prompt $\mathrm{D_s^+}$ and non-strange D mesons. The TAMU model describes the measurements within uncertainties, with a tension for the prompt $\mathrm{D_s^+}$-meson measurement in the $2 < p_{\mathrm{T}} < 3$ GeV/$c$ interval of about $1.5\sigma$ of the combined statistical and systematic uncertainties. The PHSD model describes well the measured non-strange D-meson $R_{\mathrm{AA}}$, especially in the 0–10% centrality class, while it overestimates the $\mathrm{D_s^+}$-meson measurements in the $2 < p_{\mathrm{T}} < 3$ GeV/$c$ interval and underestimates them for $p_{\mathrm{T}} > 5$ GeV/$c$. Instead, the Catania model underestimates the measurements in the $2 < p_{\mathrm{T}} < 5$ GeV/$c$ interval by about $2\sigma$ of the combined statistical and systematic uncertainties, while it overestimates the non-strange D-meson $R_{\mathrm{AA}}$ in the 0–10% centrality class for $p_{\mathrm{T}} < 1.5$ GeV/$c$, where no measurement is available for the $\mathrm{D_s^+}$ mesons. The aforementioned models are not expected to describe the D-meson nuclear modification

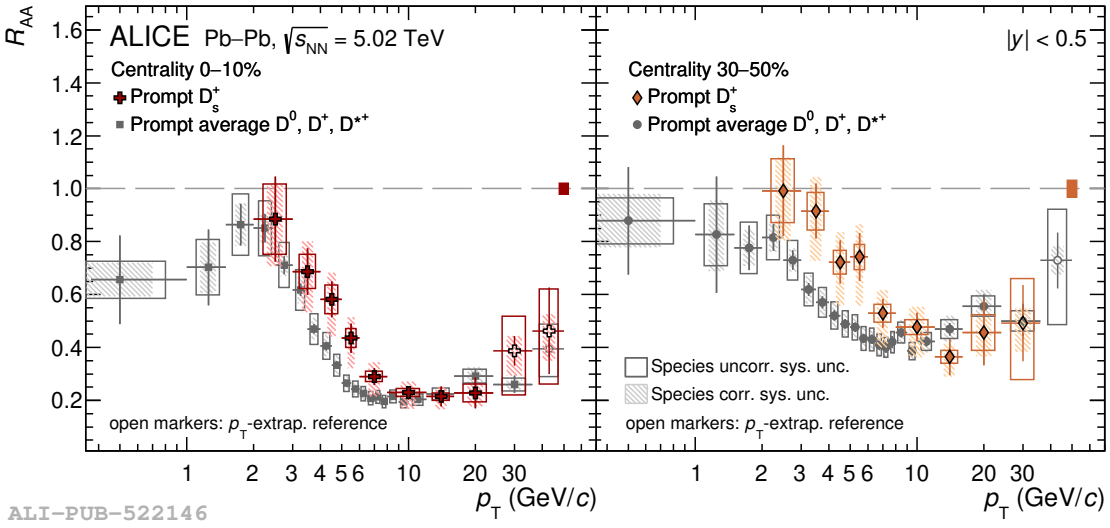



Figure 6.16: $R_{AA}$ of prompt $D_s^+$ mesons in the 0–10% (left panel) and 30–50% (right panel) centrality classes in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV compared with the one of prompt non-strange D mesons [104]. The empty (filled) boxes represent the species uncorrelated (correlated) systematic uncertainties. The normalisation uncertainty is represented by a filled box at $R_{AA} = 1$. Figure from Ref. [193].

factor for $p_T > 6$–8 GeV/$c$, since they do not include the charm-quark interactions with the medium constituents via radiative processes. The LGR model, which in contrast includes gluon-radiation processes, provides a good description of the $R_{AA}$ up to high transverse momenta.

To investigate the possible enhancement of the $D_s^+$-meson production with respect to D mesons without strange-quark content, the ratio between the $p_T$-differential yields of prompt $D_s^+$ mesons and those of prompt $D^0$ mesons [104] was computed. The systematic uncertainties due to the tracking-efficiency determination and those related to the estimation of the $f_{prompt}$ fraction were considered as fully correlated between the measurements of $D_s^+$ and $D^0$ mesons, while the other sources of systematic uncertainty were propagated as uncorrelated in the ratios. The $D_s^+/D^0$ yield ratios in Pb–Pb collisions in the 0–10% and 30–50% centrality classes compared to the same quantity measured in pp collisions [71] and to the predictions of theoretical models are shown in the top row of Fig. 6.18. In the bottom row of the Figure, the $D_s^+/D^0$ ratios in Pb–Pb collisions divided by the ones observed in pp collisions are reported. The values of the $D_s^+/D^0$ ratios in the $2 < p_T < 8$ GeV/$c$ region are higher in Pb–Pb collisions than in pp collisions by about $2.3\sigma$ and $2.4\sigma$ for central and semicentral events, respectively, considering the combined statistical and systematic uncertainties.

The Catania model [278, 281] describes within uncertainties the $D_s^+/D^0$ ratios in central Pb–Pb and in pp collisions, while the PHSD model [121, 277] provides

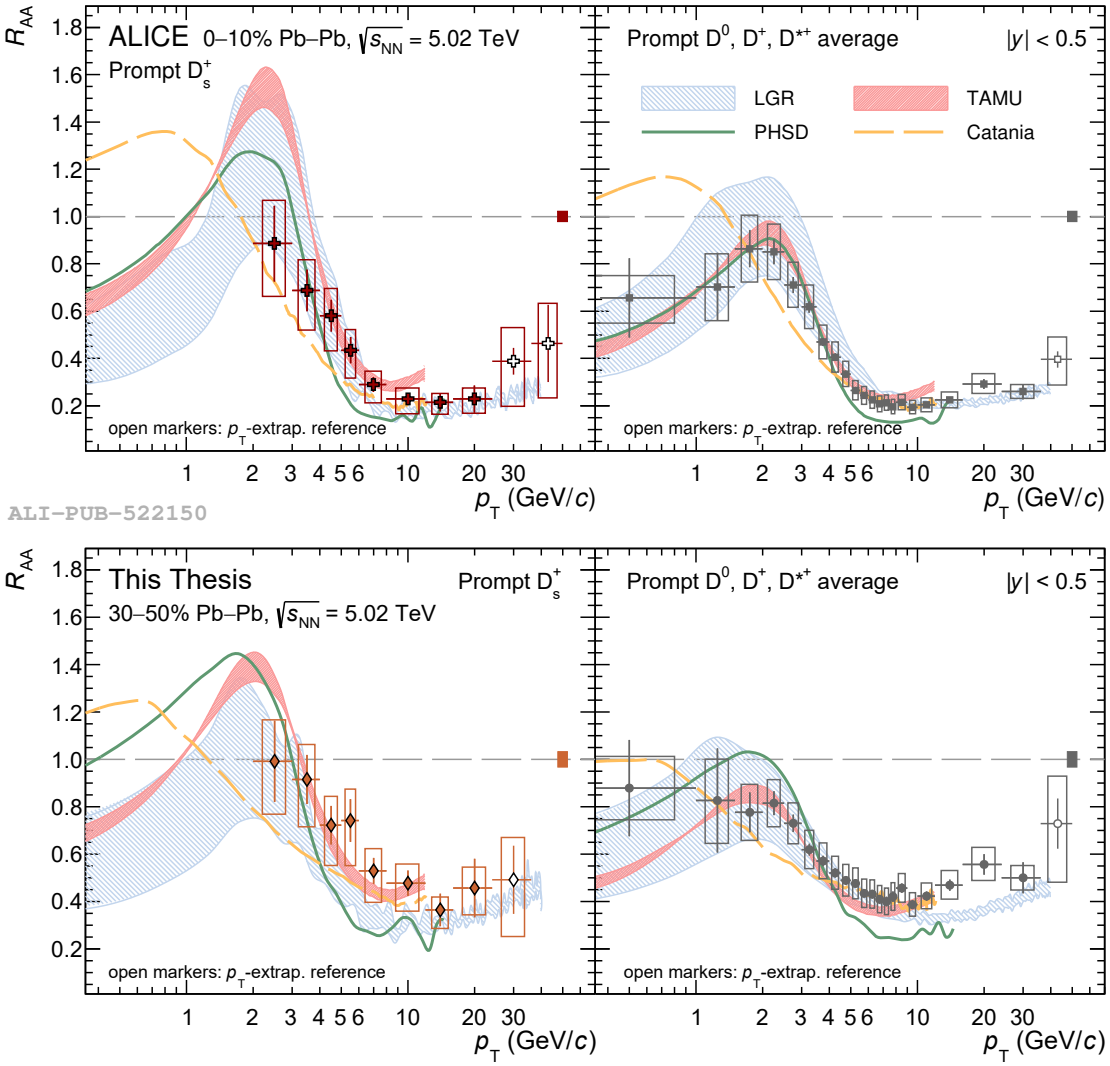



Figure 6.17: $R_{AA}$ of prompt $D_s^+$ mesons (left panels) and non-strange D mesons [104] (right panels) in the 0–10% (top row) and 30–50% (bottom row) centrality classes in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV compared with theoretical calculations [110, 114, 115, 121, 277]. The boxes represent the total systematic uncertainties. The coloured bands report the theoretical uncertainties when available. Top figure from Ref. [193].

a good description of the measurements in both centrality classes and collision systems. Instead, the measurements are significantly overestimated by the TAMU model [110]. The size of this overestimation is similar between the two colliding systems, leading to a good description of the $D_s^+/D^0$ double ratios between Pb–Pb and pp collisions, as displayed in the bottom panels of Fig. 6.18. This is consistent with the good description by the TAMU model of the D-meson $R_{AA}$, which is a

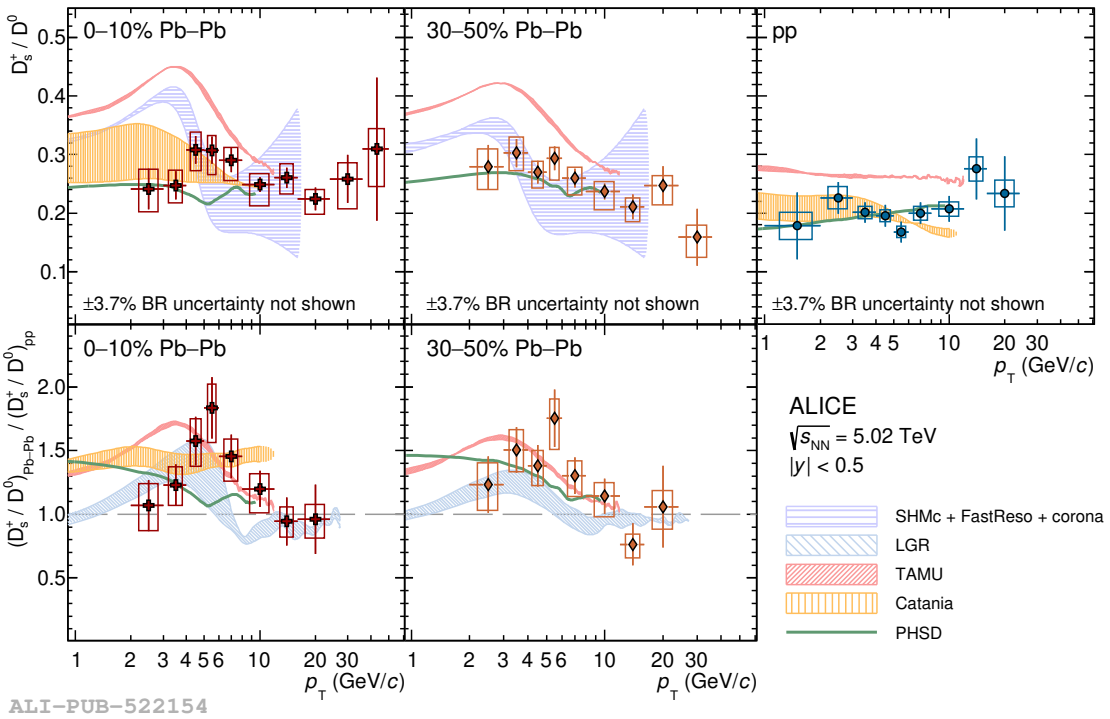



Figure 6.18: Top panels: $D_s^+/D^0$ $p_T$-differential production ratios in the 0–10% (left panel) and 30–50% (middle panel) centrality classes in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV and in pp collisions (right panel) at the same centre-of-mass energy compared with theoretical calculations [110, 114, 121, 143, 277, 278, 281, 282]. Bottom panels: $D_s^+/D^0$ $p_T$-differential ratios in Pb–Pb collisions divided by those in pp collisions compared with theoretical calculations. Figure from Ref. [193].

ratio of the measurements in Pb–Pb and pp collisions. The PHSD and Catania models predict a $D_s^+/D^0$ ratio in Pb–Pb collisions almost independent of $p_T$ for $p_T < 3$ GeV/$c$, which decreases mildly towards the pp values at higher transverse momenta. A peak of the $D_s^+/D^0$ ratio around $p_T \approx 3$–5 GeV/$c$ is instead predicted by the TAMU and LGR [114] models, for this latter model only the double-ratio predictions are available. Such a peak can be motivated by the different $D_s^+$- and $D^0$-meson masses and by the collective radial expansion of the system with a common flow-velocity profile, which imposes an equal velocity boost to all particles in case of complete thermalisation. Moreover, in the high-$p_T$ region, the hadronisation via fragmentation becomes dominant, and consequently the values of the $D_s^+/D^0$ ratio are similar in Pb-Pb and pp collisions. An analogous $p_T$ shape is predicted by the GSI-Heidelberg statistical hadronisation model (SHMc) [143] shown in the top panels of Fig. 6.18. In the SHMc model, the $p_T$ spectra of charm hadrons are modelled with a core-corona approach. The low-$p_T$ region is dominated by the core contribution described with a Blast-Wave function, while the corona contribution

Table 6.6: $p_{\mathrm{T}}$-integrated visible production yields of prompt $\mathrm{D_s^+}$ mesons in the $|y| < 0.5$ interval in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV.

| Centrality | $p_{\mathrm{T}}$ range (GeV/$c$) | $\mathrm{d}N/\mathrm{d}y\|_{\|y\|<0.5}$ |
|---|---|---|
| 0–10% | 2–50 | $0.61 \pm 0.07(\mathrm{stat})^{+0.13}_{-0.16}(\mathrm{syst}) \pm 0.02(\mathrm{BR})$ |
| 30–50% | 2–36 | $0.120 \pm 0.010(\mathrm{stat})^{+0.021}_{-0.028}(\mathrm{syst}) \pm 0.004(\mathrm{BR})$ |

Table 6.7: Total production yields of prompt $\mathrm{D_s^+}$ mesons in the $|y| < 0.5$ interval in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV compared to the predictions of the SHMc model [143].

| Centrality | $\mathrm{d}N/\mathrm{d}y\|_{\|y\|<0.5}$ | SHMc |
|---|---|---|
| 0–10% | $1.89 \pm 0.07(\mathrm{stat})^{+0.13}_{-0.16}(\mathrm{syst})^{+0.36}_{-0.55}(\mathrm{extr}) \pm 0.07(\mathrm{BR})$ | $2.22 \pm 0.38$ |
| 30–50% | $0.34 \pm 0.01(\mathrm{stat})^{+0.02}_{-0.03}(\mathrm{syst})^{+0.11}_{-0.09}(\mathrm{extr}) \pm 0.01(\mathrm{BR})$ | $0.344 \pm 0.056$ |

is parametrised from measurements in pp collisions and is dominant at high $p_{\mathrm{T}}$. The modification of the $p_{\mathrm{T}}$ spectra due to resonance decays is described using the FastReso package [145].

Due to the current uncertainties of the measurement in Pb–Pb collisions, no strong conclusions can be made on the $p_{\mathrm{T}}$ shape of the $\mathrm{D_s^+/D^0}$ ratio at low and intermediate $p_{\mathrm{T}}$. However, these results provide important information on the role of the charm-quark hadronisation via recombination in the QGP medium, and are complementary to those deriving from the simultaneous comparison of the D-meson $R_{\mathrm{AA}}$ and $v_n$ coefficients [104, 109].

The visible production yield of prompt $\mathrm{D_s^+}$ mesons was calculated by integrating the $p_{\mathrm{T}}$-differential corrected yield over the transverse-momentum interval of the measurement. The systematic uncertainties were propagated as fully correlated between the different $p_{\mathrm{T}}$ intervals, except for the uncertainty on the raw-yield extraction, as detailed in Section 6.6.7. The resulting visible production yields are reported in Tab. 6.6 for the 0–10% and 30–50% centrality classes. In addition, the corrected yield was extrapolated in the $0 < p_{\mathrm{T}} < 2$ GeV/$c$ interval to obtain the total $p_{\mathrm{T}}$-integrated production yield. To this end, the measured $p_{\mathrm{T}}$-differential $\mathrm{D_s^+/D^0}$ ratio was fitted using the shape predicted by the PHSD model and keeping the normalisation free to vary. The resulting $\mathrm{D_s^+/D^0}$ ratio extrapolated to $p_{\mathrm{T}} < 2$ GeV/$c$ is about 0.29 (0.27) in central (semicentral) Pb–Pb collisions. It was then multiplied by the $\mathrm{D^0}$-meson yield measured in the same $p_{\mathrm{T}}$ region [104], which is about 4.38 (0.84), to compute the extrapolated $\mathrm{D_s^+}$-meson yield, which

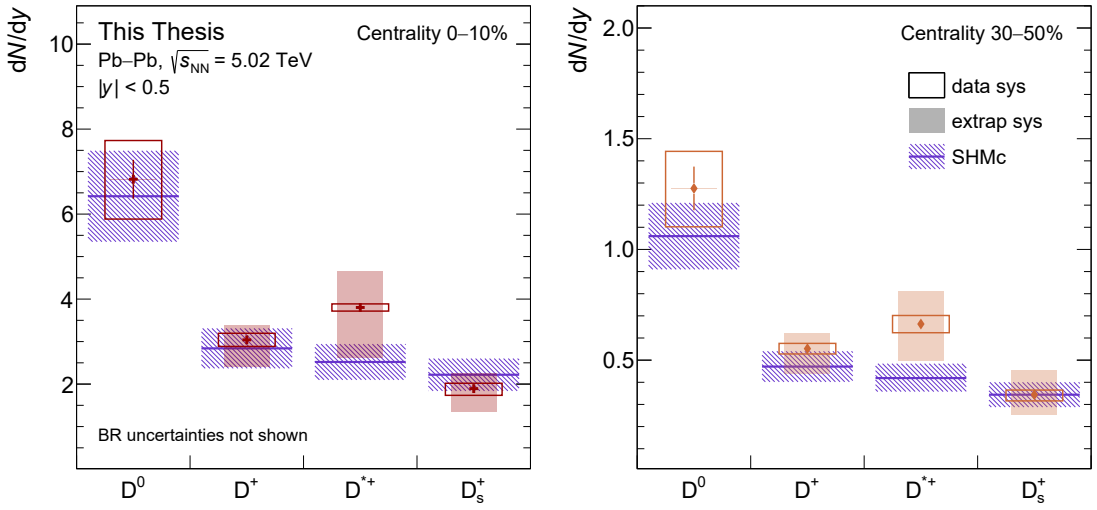



Figure 6.19: Total $p_T$-integrated production yields of prompt $D^0$, $D^+$, $D^{*+}$ [104], and $D_s^+$ mesons in central (left panel) and semicentral (right panel) Pb–Pb collisions compared to SHMc model predictions [143]. Statistical (bars), systematic (empty boxes) and extrapolation (shaded boxes) uncertainties are shown for the data. The continuous lines report the central values of the SHMc predictions, while the uncertainties are represented by the filled boxes.

corresponds to 1.28 (0.22) for the 0–10% (30–50%) centrality class. The extrapolated $D_s^+$-meson yield in the $0 < p_T < 2$ GeV/$c$ interval accounts for about 70% of the total production yield. An uncertainty was assigned to the extrapolation procedure. It was estimated by repeating the extrapolation of the $D_s^+/D^0$ ratio using the $p_T$ shapes provided by the TAMU, Catania and SHMc models. The total $p_T$-integrated production yields of prompt $D_s^+$ mesons at midrapidity for the 0–10% and 30–50% centrality classes are reported in Tab. 6.7. They are compatible within uncertainties with the predictions of the SHMc model, similarly to what observed for the $D^0$, $D^+$, and $D^{*+}$ mesons [104]. Figure 6.19 reports the total production yields of prompt $D^0$, $D^+$, $D^{*+}$, and $D_s^+$ mesons in central and semicentral Pb–Pb collisions compared with the respective SHMc model calculations. The agreement between the measurements and the SHMc predictions suggests that low-$p_T$ charm quarks, which determine the total yield, could be thermalised in the QGP.

# Chapter 7

# Prompt $\mathrm{D_s^+}$-meson elliptic flow in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV

In this Chapter, the measurement of the prompt $\mathrm{D_s^+}$-meson elliptic flow, $v_2$, in semicentral Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV is reported. The measurement of the azimuthal anisotropy of prompt D mesons provides insights into the degree of thermalisation of charm quarks in the QGP medium and their participation in the fireball collective motion. Moreover, the comparison of the elliptic flow of $\mathrm{D_s^+}$ and non-strange D mesons could provide sensitivity to the transport properties of the hadronic-phase, since $\mathrm{D_s^+}$ mesons are expected to decouple early from the hadron gas and thus to not acquire additional $v_2$ in the late hadronic stages [142]. The measurement shown in this Chapter was published in Ref. [193]. The analysis procedure for the $v_2$ determination, which is described in the following, is the same as the one adopted for the measurement of the non-strange D-meson elliptic flow reported in Ref. [109].

## 7.1 Scalar-product method for the elliptic-flow measurement

The azimuthal anisotropy of the particle-yield distribution is usually studied through a Fourier decomposition in the particle azimuthal angle $\varphi$ with respect to the initial-state symmetry-plane angle. The elliptic flow $v_2$ is the second coefficient of this decomposition, and it is defined as

$$v_2 = \langle \cos[2(\varphi - \Psi_2)] \rangle, \tag{7.1}$$

where $\Psi_2$ is the $2^{\mathrm{nd}}$-harmonic symmetry-plane angle, and $\langle\rangle$ signals the average over all particles and events.

The initial-state symmetry-plane angle $\Psi_2$ is not experimentally accessible. It can be estimated by the determination of the event-plane angle for the 2$^{\mathrm{nd}}$ harmonic $\psi_2$, which is calculated starting from the so-called flow vector $\boldsymbol{Q}_2$. The $\boldsymbol{Q}_2$ vector is calculated from the measured azimuthal angles of particles produced in the event as

$$\boldsymbol{Q}_2 = \sum_{j=1}^{M} w_j e^{i2\varphi_j}. \tag{7.2}$$

When the $\boldsymbol{Q}_2$ vector is obtained using a tracking detector, as the TPC, the sum runs over the reconstructed tracks, $\varphi_j$ is the azimuthal emission angle of the $j^{\mathrm{th}}$ track, and $w_j$ is the weight assigned to the track. In the case of segmented detectors, as the V0A and V0C, the sum runs over the various segments, $\varphi_j$ is the azimuthal angle of the $j^{\mathrm{th}}$ detector segment, and $w_j$ is the measured signal amplitude, which is proportional to the number of charged particles collected in that segment. The event-plane angle $\psi_2$ is computed from the $\boldsymbol{Q}_2$-vector components as $\psi_2 = \tan^{-1}(Q_{2,y}/Q_{2,x})/2$.

In the analysis presented in this Thesis, the elliptic flow was estimated using the scalar-product (SP) method, which does not rely on the explicit calculation of the event-plane angle. Instead, this method provides the $v_2$ coefficient directly from the correlation between the $\boldsymbol{Q}_2$ vector and the direction of the particle of interest [283, 284]. The three-subevent formula was used to estimate the resolution on the $\psi_2$ or equivalently on the $\boldsymbol{Q}_2$. The $\mathrm{D_s^+}$-meson elliptic flow was computed as

$$v_2\{\mathrm{SP}\} = \frac{\left\langle\left\langle \boldsymbol{u}_2 \cdot \dfrac{\boldsymbol{Q}_2^{\mathrm{A}*}}{M^{\mathrm{A}}} \right\rangle\right\rangle}{\sqrt{\dfrac{\left\langle \dfrac{\boldsymbol{Q}_2^{\mathrm{A}}}{M^{\mathrm{A}}} \cdot \dfrac{\boldsymbol{Q}_2^{\mathrm{B}*}}{M^{\mathrm{B}}} \right\rangle \left\langle \dfrac{\boldsymbol{Q}_2^{\mathrm{A}}}{M^{\mathrm{A}}} \cdot \dfrac{\boldsymbol{Q}_2^{\mathrm{C}*}}{M^{\mathrm{C}}} \right\rangle}{\left\langle \dfrac{\boldsymbol{Q}_2^{\mathrm{B}}}{M^{\mathrm{B}}} \cdot \dfrac{\boldsymbol{Q}_2^{\mathrm{C}*}}{M^{\mathrm{C}}} \right\rangle}}}, \tag{7.3}$$

where $\boldsymbol{u}_2 = e^{i2\varphi_{\mathrm{D}}}$ is the unit flow vector of the $\mathrm{D_s^+}$-meson candidate with azimuthal angle $\varphi_{\mathrm{D}}$, $\boldsymbol{Q}_2^k$ is the 2$^{\mathrm{nd}}$-harmonic flow vector for the subevent $k$, and $M^k$ represents the subevent multiplicity. For tracking detectors, $M^k$ corresponds to the particle multiplicity in the considered subevent, while for segmented detectors it is the sum of the amplitudes measured in the channels that compose the subevent $k$. The single bracket $\langle\rangle$ in the Equation denotes the average over all events, wile the double brackets $\langle\langle\rangle\rangle$ indicate the average over all particles in the $p_{\mathrm{T}}$ interval of interest and all events.

## 7.2 Subevent configuration and resolution

The detectors used to estimate the $\boldsymbol{Q}_2$ vectors for the prompt $\mathrm{D_s^+}$-meson $v_2$ measurement were the TPC and V0 detectors. The three independent subevents required for the computation of Eq. (7.3) were defined from:

- the charged-particle tracks reconstructed in the TPC, which has a pseudorapidity coverage of $|\eta| < 0.8$;
- the amplitudes in the V0A detector, which covers the $2.8 < \eta < 5.1$ interval;
- the amplitudes in the V0C detector, which covers the $-3.7 < \eta < -1.7$ interval.

The V0C was chosen to define the main subevent, indicated as A in Eq. (7.3), while the TPC and V0A detectors were used to define the other two subevents, B and C, which enter only in the equation denominator. The adoption of the V0C detector guarantees a pseudorapidity gap larger than 0.9 units, $|\Delta\eta| > 0.9$, between the pseudorapidity interval of the particles entering the definition of the main $\boldsymbol{Q}_2$ vector and that of the $\mathrm{D_s^+}$ mesons, $|\eta| < 0.8$, which were measured at midrapidity. This pseudorapidity gap assures a substantial reduction of non-flow contributions which could bias the elliptic-flow measurement. Non-flow contributions are caused by correlations between the emission azimuthal angles of particles induced by jet production and particle decays, thus not deriving from the collective expansion of the system which is the phenomenon of interest. The V0A detector would provide a larger pseudorapidity gap, $|\Delta\eta| > 2$, than the V0C and in principle a possible further suppression of non-flow contributions. However, its adoption was observed to lead to larger statistical uncertainties, as will be discussed in Section 7.5.2. Finally, the subevent obtained from the TPC detector covers the same pseudorapidity interval as the measured $\mathrm{D_s^+}$ mesons, and cannot be employed as main subevent.

The $\boldsymbol{Q}_2$ vectors were corrected to account for the non-uniform acceptance of the detectors in the azimuthal direction, which could introduce artificial correlations and bias the measurement [285]. For the subevent based on the TPC detector, a uniform track distribution as a function of $\varphi$ was obtained by assigning per-track weights, which were estimated independently for each data-taking run to consider the variation of the detector conditions. In the case of the V0A and V0C detectors, the raw signal amplitudes of each channel were standardised through a gain equalisation. Moreover, the $\boldsymbol{Q}_2$ vectors were shifted to centre at zero the distributions of the $Q_{2,x}$ and $Q_{2,y}$ components. These corrections were applied independently for each data-taking run and as a function of the primary-vertex position in the $z$ direction. More details on the $\boldsymbol{Q}_2$-vector calibration can be found in Ref. [251].

The denominator of Eq. (7.3), which will be referred to as the *scalar-product resolution* $R_2$ in the following, was computed using the subevent configuration

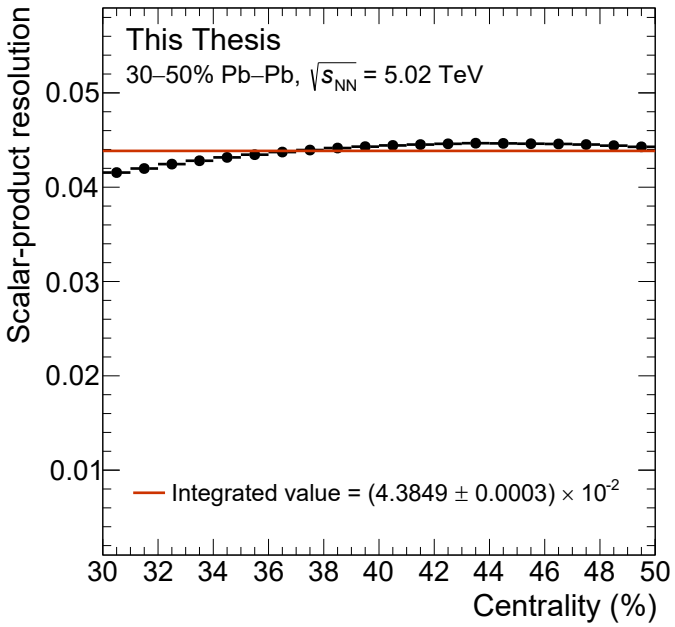



Figure 7.1: Scalar-product resolution of the V0C detector as a function of the centrality sub-interval for Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV in the 30–50% centrality class. The red line displays the centrality-integrated SP resolution with its statistical uncertainty.

described above and the sample of semicentral Pb–Pb collisions collected by the experiment. The SP-resolution values estimated in 1%-wide centrality intervals within the 30–50% centrality class are reported in Fig. 7.1, together with the resolution obtained considering the full sample of semicentral events, whose value amounts to $(4.3849 \pm 0.0003) \times 10^{-2}$. The SP resolution of the V0C detector shows a slight centrality dependence within the 30–50% centrality class, and increases from more central to more peripheral collisions. This trend is determined by the interplay between the event multiplicity and the event eccentricity. The first one decreases from more central to more peripheral collisions inducing a reduction of the SP resolution. The second one increases instead, leading to an opposite effect.

The possible systematic uncertainty due to the centrality dependence of the SP resolution was estimated by calculating the resolution as a weighted average of the values obtained in the 1% centrality intervals. The weights were defined by the D-meson raw yields measured in the narrow centrality intervals. A relative systematic uncertainty of 0.5% was assigned by considering the difference between the SP resolution resulting from the weighted average and the one calculated using all the events in the 30–50% centrality class [109, 251].

## 7.3 Extraction of the $\mathrm{D_s^+}$-meson $v_2$

The elliptic flow of $\mathrm{D_s^+}$ mesons in the 30–50% centrality class was measured for six transverse-momentum intervals in the range $2 < p_{\mathrm{T}} < 24$ GeV/$c$. The elliptic flow was extracted from a data sample of $\mathrm{D_s^+}$-meson candidates that satisfy the single-track selection criteria listed in Section 4.2.1, and the conservative PID strategy, which is detailed in Section 4.2.2. The large combinatorial-background contribution was suppressed by employing the same BDTs trained for the $R_{\mathrm{AA}}$ measurement in the 30–50% centrality class, and by adopting the same threshold values on the BDT output. More details are reported in Section 6.1. As for the measurement of the nuclear modification factor, the application of the BDT models on the full data sample of Pb–Pb collisions collected by the ALICE experiment was performed on the Grid.

The sample of selected $\mathrm{D_s^+}$-meson candidates was composed of both signal and background, thus it was not possible to obtain the $\mathrm{D_s^+}$-meson elliptic flow directly with the scalar-product method. Therefore, the anisotropic flow coefficient $v_2^{\mathrm{tot}}$, relative to both signal and background candidates, was measured using Eq. (7.3) in narrow invariant-mass intervals. It was then expressed as a weighted sum of the elliptic flow of real $\mathrm{D_s^+}$-meson decays, $v_2^{\mathrm{sig}}$, and that of the combinatorial background, $v_2^{\mathrm{bkg}}$ [286], according to the following equation

$$v_2^{\mathrm{tot}}(M) = \frac{N^{\mathrm{sig}}(M)v_2^{\mathrm{sig}} + N^{\mathrm{bkg}}(M)v_2^{\mathrm{bkg}}(M) + N^{\mathrm{D^+}}(M)v_2^{\mathrm{D^+}}}{N^{\mathrm{sig}}(M) + N^{\mathrm{bkg}}(M) + N^{\mathrm{D^+}}(M)}, \tag{7.4}$$

where $N^{\mathrm{sig}}$ and $N^{\mathrm{bkg}}$ are the signal and background raw yields, respectively. To account for the contribution of the $\mathrm{D^+} \to \phi\pi^+ \to \mathrm{K^-K^+}\pi^+$ decays to the measured $v_2^{\mathrm{tot}}$ distribution, another free parameter $v_2^{\mathrm{D^+}}$ with the corresponding raw yield $N^{\mathrm{D^+}}$ was included. The background $v_2$ is defined as a function of $M$, since it usually depends on the invariant mass.

The elliptic flow coefficients were extracted for each $p_{\mathrm{T}}$ interval from a simultaneous fit to the invariant-mass distribution of $\mathrm{D_s^+}$-meson candidates and the $v_2^{\mathrm{tot}}$ values as a function of the invariant mass. The fit of the invariant-mass distribution was performed with two Gaussian functions to describe the $\mathrm{D_s^+}$-meson signal peak and the peak deriving from the $\mathrm{D^+} \to \phi\pi^+ \to \mathrm{K^-K^+}\pi^+$ decays, while an exponential term was used to describe the background. The $v_2^{\mathrm{sig}}$ was obtained from the fit to the $v_2^{\mathrm{tot}}(M)$ distribution by using the function defined in Eq. (7.4), in which $N^{\mathrm{sig}}(M)$, $N^{\mathrm{bkg}}(M)$, and $N^{\mathrm{D^+}}(M)$ are the parameters in common with the invariant-mass distribution in the simultaneous fit. In the fit, a linear function was employed to describe $v_2^{\mathrm{bkg}}(M)$. Figure 7.2 reports the simultaneous fits to the invariant-mass and $v_2^{\mathrm{tot}}(M)$ distributions of $\mathrm{D_s^+}$ mesons in the $p_{\mathrm{T}}$ intervals in which the analysis was performed. A wider binning was adopted for $v_2^{\mathrm{tot}}(M)$ in the sidebands, to reduce the statistical fluctuations in a region where the elliptic flow is dominated by the background component.

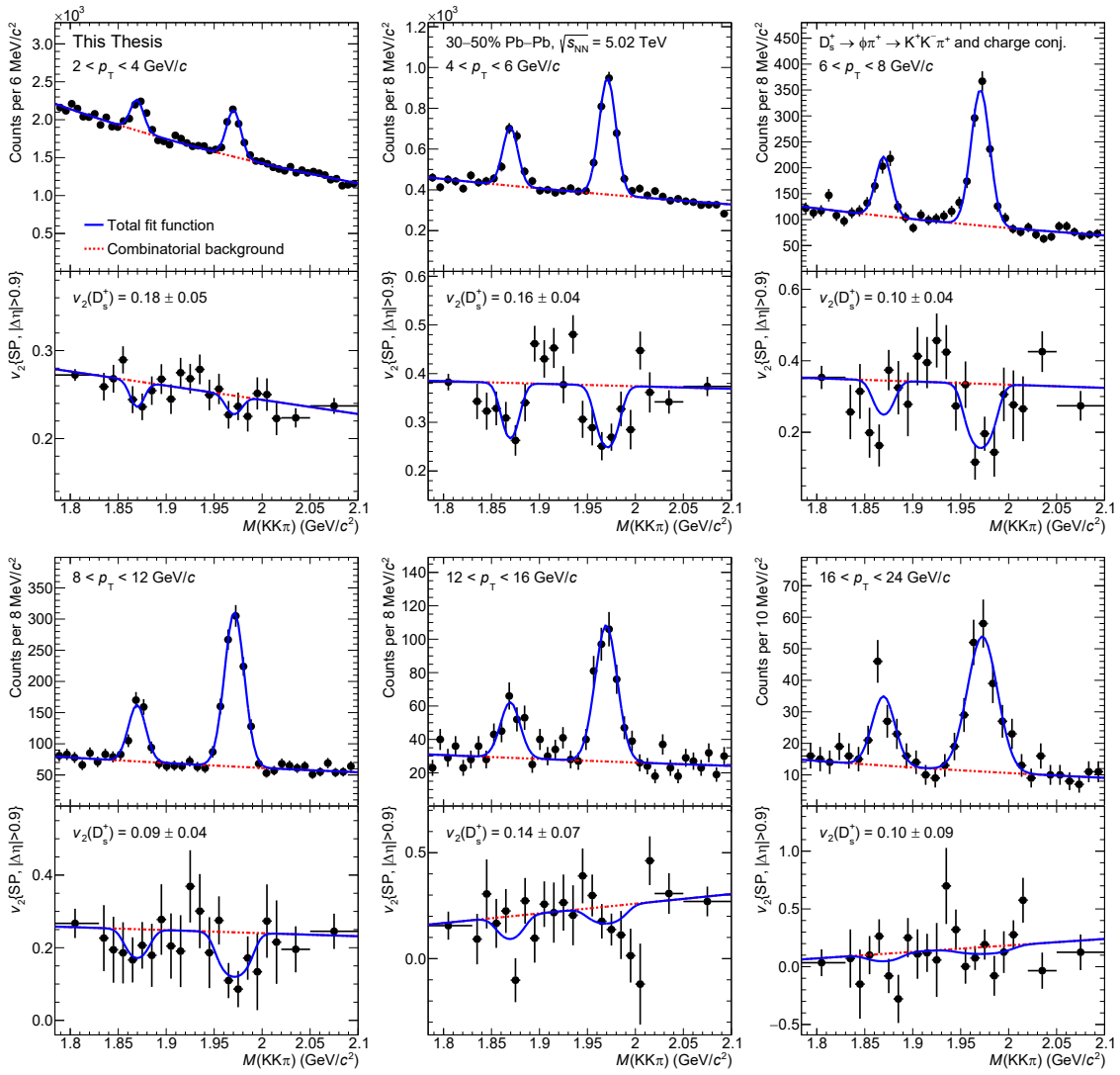



Figure 7.2: Invariant-mass and $v_2^{\rm tot}(M)$ distributions of $D_s^+$-meson candidates in the 30–50% centrality class for the $p_T$ intervals in which the analysis was performed. The solid blue and dotted red curves show the total and combinatorial-background fit functions, respectively. The extracted values of $v_2^{\rm sig}$ are also reported together with their statistical uncertainties.

## 7.4 Correction for non-prompt $D_s^+$-meson contribution

The extracted $D_s^+$-meson elliptic flow is a mixture of the contribution of prompt mesons and the one of non-prompt mesons from beauty-hadron decays. Hence, the measured $v_2^{\rm sig}$ can be expressed as a linear combination of the two contributions as

follows

$$v_2^{\mathrm{sig}} = f_{\mathrm{prompt}} v_2^{\mathrm{prompt}} + (1 - f_{\mathrm{prompt}}) v_2^{\mathrm{non\text{-}prompt}}, \tag{7.5}$$

where $v_2^{\mathrm{prompt}}$ and $v_2^{\mathrm{non\text{-}prompt}}$ indicate the prompt and non-prompt $\mathrm{D_s^+}$-meson elliptic flow, respectively, and $f_{\mathrm{prompt}}$ is the fraction of prompt $\mathrm{D_s^+}$ mesons in the extracted signal. To solve the equation, an assumption is necessary for the $v_2$ of non-prompt $\mathrm{D_s^+}$-mesons, which was not measured so far. For the calculation of the central value of $v_2^{\mathrm{prompt}}$, the assumption $v_2^{\mathrm{non\text{-}prompt}} = v_2^{\mathrm{prompt}}/2$ was considered. This hypothesis was based on the elliptic-flow measurements of prompt and non-prompt $\mathrm{J/\psi}$ mesons by the ATLAS and CMS Collaborations [287, 288], and on the predictions of theoretical models [289–291], which indicate a $v_2^{\mathrm{non\text{-}prompt}}$ comprehended between zero and $v_2^{\mathrm{prompt}}$. Using the aforementioned assumption, the $v_2$ of prompt $\mathrm{D_s^+}$ mesons was obtained from the measured $v_2^{\mathrm{sig}}$ rewriting Eq. (7.5) as

$$v_2^{\mathrm{prompt}} = \frac{2 v_2^{\mathrm{sig}}}{1 + f_{\mathrm{prompt}}}. \tag{7.6}$$

The systematic uncertainty related to the choice of the $v_2^{\mathrm{non\text{-}prompt}}$ value was estimated by considering a flat distribution of the non-prompt elliptic flow in the $0 < v_2^{\mathrm{non\text{-}prompt}} < v_2^{\mathrm{prompt}}$ interval. Therefore, the central value of $v_2^{\mathrm{non\text{-}prompt}}$ was varied by $\pm v_2^{\mathrm{prompt}}/\sqrt{12}$, which corresponds to one standard deviation.

The procedure adopted to estimate the prompt fraction was the same as that used for the measurement of the prompt $\mathrm{D_s^+}$-meson nuclear modification factor in Pb–Pb collisions, which is described in Section 6.4. The $f_{\mathrm{prompt}}$ fraction was computed for the different $p_{\mathrm{T}}$ intervals of the analysis employing the following ingredients: the beauty-hadron production cross section from FONLL calculations [72, 73], the $f(\mathrm{b} \rightarrow \mathrm{H_b})$ fragmentation fractions measured in $\mathrm{Z} \rightarrow \mathrm{b\overline{b}}$ decays [6], the description of the $\mathrm{H_b} \rightarrow \mathrm{D_s^+} + \mathrm{X}$ decays from the PYTHIA 8 package [82, 168], the acceptance-times-efficiency factor for non-prompt $\mathrm{D_s^+}$ mesons estimated from the MC simulations introduced in Section 4.3, and the assumption on the nuclear modification factor $R_{\mathrm{AA}}^{\mathrm{non\text{-}prompt}} = R_{\mathrm{AA}}^{\mathrm{prompt}}$. The systematic uncertainty on the $f_{\mathrm{prompt}}$ fraction was computed varying the parameters of the FONLL calculations, and the assumed non-prompt $\mathrm{D_s^+}$-meson $R_{\mathrm{AA}}$ between 1/3 and 3 times the prompt $\mathrm{D_s^+}$-meson one. More details are reported in Section 6.6.3.

In the left panel of Fig. 7.3, the $f_{\mathrm{prompt}}$ fraction is shown as a function of $p_{\mathrm{T}}$, together with its systematic uncertainty. The $v_2^{\mathrm{sig}}$ and $v_2^{\mathrm{prompt}}$ values are reported as a function of $p_{\mathrm{T}}$ in the right panel of the Figure. The systematic uncertainties deriving from the correction for the non-prompt $\mathrm{D_s^+}$-meson contribution are also shown. This correction leads to a larger prompt $v_2$ than the extracted one by about 5–7%, while the absolute values of the systematic uncertainty range between $^{+0.020}_{-0.005}$ and $^{+0.039}_{-0.009}$ depending on $p_{\mathrm{T}}$.

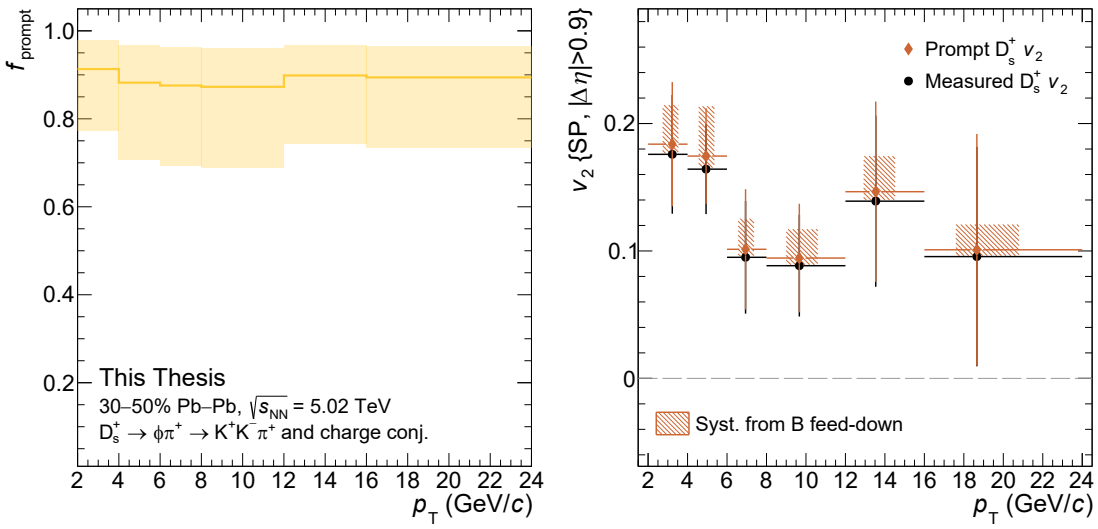



Figure 7.3: Left panel: $f_{\mathrm{prompt}}$ fraction used to correct the measured elliptic flow as a function of $p_{\mathrm{T}}$. The orange band indicates the systematic uncertainties. Right panel: comparison of the measured $v_2$, containing both prompt and non-prompt, and the prompt $\mathrm{D_s^+}$-meson $v_2$ as a function of $p_{\mathrm{T}}$. The systematic uncertainties related to the correction for the non-prompt contribution are reported by the boxes.

## 7.5 Systematic uncertainties

The sources of systematic uncertainty affecting the measurement of the prompt $\mathrm{D_s^+}$-meson $v_2$ coefficients are the following: (i) the elliptic-flow extraction from the fits to the invariant-mass and $v_2^{\mathrm{tot}}(M)$ distributions, (ii) the correction for the presence of non-prompt $\mathrm{D_s^+}$ mesons, and (iii) the centrality dependence of the scalar-product resolution. The estimation procedures and the assigned systematic-uncertainty values for the latter two contributions were presented in Sections 7.4 and 7.2, respectively. Moreover, no dependence of the selection efficiency from the $\mathrm{D_s^+}$-meson azimuthal direction was observed, thus no contribution was added to the systematic uncertainty. A $\varphi$-dependent selection efficiency would introduce an artificial modulation of the measured particle-yield distribution in the azimuthal direction, leading to a biased result. Finally, possible systematic effects due to non-flow contributions and the residual miscalibration of the $\boldsymbol{Q}_2$ vectors were investigated, and they were found to be negligible as will be discussed in the following.

### 7.5.1 Fitting procedure

The systematic uncertainty due to the elliptic-flow extraction was estimated with a multi-trial approach, i.e., by repeating the simultaneous fit to the invariant-mass and $v_2^{\mathrm{tot}}(M)$ distributions using various fit configurations. The alternative fit

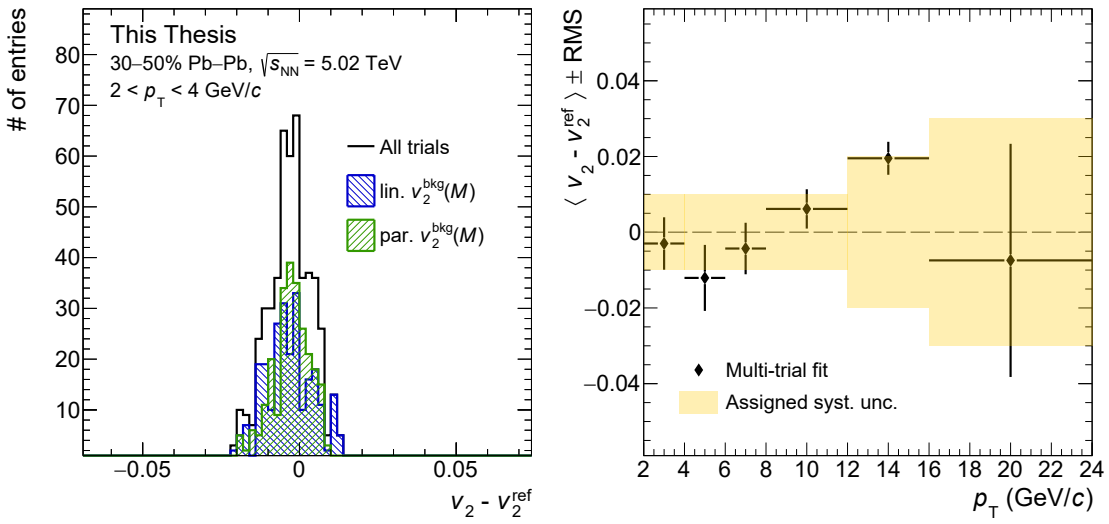



Figure 7.4: Left panel: distribution of the $v_2$ residuals, $v_2 - v_2^{\mathrm{ref}}$, obtained from the different trials in the $2 < p_{\mathrm{T}} < 4$ GeV/$c$ interval for semicentral Pb–Pb collisions. The distributions obtained using a linear (blue) or parabolic (green) function to describe the background elliptic-flow $v_2^{\mathrm{bkg}}(M)$ are reported separately. Right panel: mean and RMS of the multi-trial residual distribution as a function of $p_{\mathrm{T}}$. The yellow band indicates the assigned systematic-uncertainty values.

configurations were obtained considering: five different upper and lower limits of the invariant-mass range, five different bin widths of the invariant-mass and $v_2^{\mathrm{tot}}(M)$ distributions, three combinatorial-background functional forms (linear, exponential, and parabolic) for the invariant-mass fit, and linear and parabolic functions to describe the $v_2(M)$ of the combinatorial background.

The $\mathrm{D_s^+}$-meson elliptic flow was extracted for each alternative fit configuration and the residual with respect to the reference value, $v_2 - v_2^{\mathrm{ref}}$, was computed. As an example, the residual distribution obtained from the different trials in the $2 < p_{\mathrm{T}} < 4$ GeV/$c$ interval is shown in the left panel of Fig. 7.4. The trials which presented a poor fit quality, i.e., $\chi^2/ndf > 2$, were not included in the distribution and were not considered for the systematic-uncertainty estimation. In the Figure, the distributions of the trials with a linear and parabolic description of the background $v_2$ are reported separately, and are similar between each other. The systematic uncertainty was assigned taking into account the RMS and shift from zero, when it was relevant, of the residual distribution in each transverse-momentum interval. In the right panel of Fig. 7.4, the mean of the residual distribution is displayed as a function of $p_{\mathrm{T}}$, while the distribution RMS is represented by the vertical bars. The absolute systematic uncertainties assigned are reported as a yellow band, and range between 0.01 and 0.03 depending on $p_{\mathrm{T}}$.

### 7.5.2 Residual miscalibration and non-flow effects

A systematic bias of the elliptic-flow measurement due to spurious correlations can be induced by residual miscalibrations in the corrections for the detector non-uniform acceptance, which have been presented in Section 7.2. The effect of the residual miscalibrations was studied by calculating the product of the orthogonal components between the unitary flow vector $\boldsymbol{u}_2$ of $\mathrm{D_s^+}$ mesons and the main $\boldsymbol{Q}_2$ vector, which was obtained using the V0C information. The computed quantities were $u_{2,x} \times Q_{2,y}^{\mathrm{V0C}}$ and $u_{2,y} \times Q_{2,x}^{\mathrm{V0C}}$, which are expected to be equal to zero for symmetry reasons if the $\boldsymbol{Q}_2$ vectors are perfectly calibrated. The average values of the aforementioned products were obtained through simultaneous fits to the invariant-mass distribution of $\mathrm{D_s^+}$-meson candidates and their values in invariant-mass intervals, with a procedure similar to the one used to extract the $\mathrm{D_s^+}$-meson elliptic flow. In the left panel of Fig. 7.5, the sum of the average values of the two products divided by the SP resolution $R_2$ is shown as a function of $p_{\mathrm{T}}$. The measured values are compatible with zero considering their uncertainties. Larger discrepancies are observed at high transverse momenta where statistical fluctuations are more relevant. The same study was performed for the measurement of the elliptic flow of $\mathrm{D^0}$, $\mathrm{D^+}$, and $\mathrm{D^{*+}}$ mesons, which are more abundantly produced and thus are characterised by smaller statistical uncertainties, showing no systematic effects [109, 251]. This confirms what is observed for the $\mathrm{D_s^+}$ mesons. Therefore, the bias due to residual miscalibrations was considered to be negligible, and no systematic uncertainty was assigned.

Another possible source of systematic bias on the prompt $\mathrm{D_s^+}$-meson $v_2$ is the non-flow contribution. This possible bias was investigated by repeating the measurement of the elliptic flow defining the main subevent using the V0A detector instead of the V0C detector. The V0A provides a pseudorapidity gap of $|\Delta\eta| > 2$ with respect to the $\mathrm{D_s^+}$ mesons reconstructed at midrapidity, which is larger than the one obtained with the V0C, as discussed in Section 7.2. Thus, its employment should further reduce the non-flow effects if present. The prompt $\mathrm{D_s^+}$-meson elliptic flow measured with the two configurations are compared in the right panel of Fig. 7.5. The two measurements are compatible within uncertainties. This observation is in agreement with measurements of the elliptic flow of light-flavour hadrons and non-strange D mesons, which have a much higher statistical precision, where no differences between the 0.9 and 2 pseudorapidity gap are noticed [44, 109]. For the aforementioned reasons, no systematic uncertainty was assigned. The $v_2$ obtained with the V0A detector has larger statistical uncertainties of about 10–20% than the V0C one depending on $p_{\mathrm{T}}$. This is due to the worse SP resolution obtained using the V0A detector with respect to the V0C one. The V0A detector cover a more forward pseudorapidity region where the particle multiplicity is smaller. Therefore, the number of particles within its acceptance and available for computing the $\boldsymbol{Q}_2$ vector is lower than for the V0C detector.

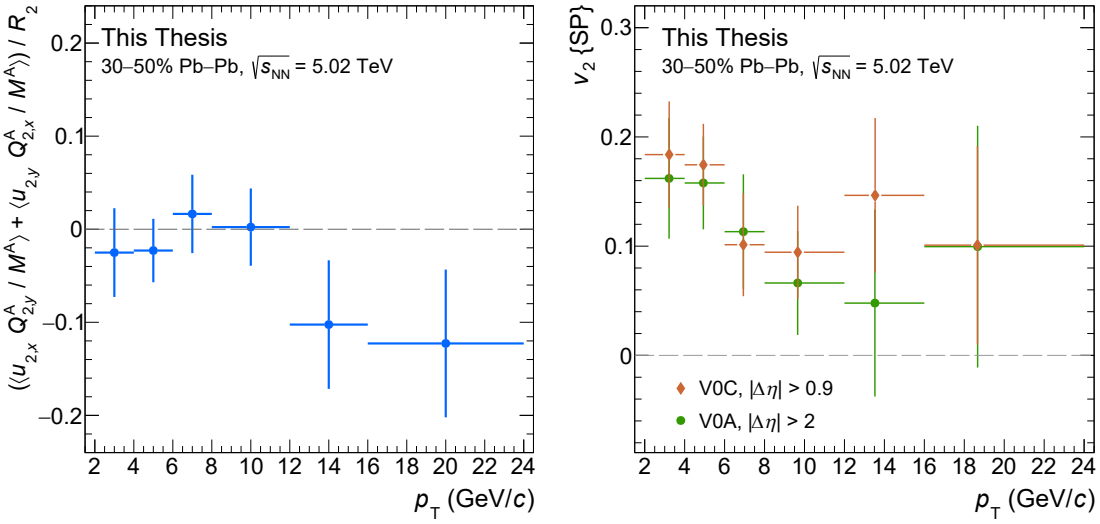



Figure 7.5: Left panel: estimated values of the bias due to residual miscalibrations of the $\boldsymbol{Q}_2$ vectors as a function of $p_{\mathrm{T}}$. Right panel: comparison of the prompt $\mathrm{D_s^+}$-meson elliptic flow measured using the V0C (orange) and V0A (green) detectors to build the main subevent in semicentral Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV.

### 7.5.3 Systematic uncertainty summary

The systematic uncertainties on the measured prompt $\mathrm{D_s^+}$-meson elliptic flow in semicentral Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV are summarised in Tab. 7.1. The absolute values of the total systematic uncertainty are also reported. They were obtained by summing in quadrature the different contributions.

Table 7.1: Summary of the systematic uncertainties of the prompt $\mathrm{D_s^+}$-meson $v_2$ in semicentral Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. The total absolute systematic uncertainty is also reported.

| $p_{\mathrm{T}}$ (GeV/$c$) | 2–4 | 4–6 | 6–8 | 8–12 | 12–16 | 16–24 |
|---|---|---|---|---|---|---|
| $M$ and $v_2$ fits | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 | 0.03 |
| Non-prompt contribution | $^{+0.031}_{-0.007}$ | $^{+0.039}_{-0.009}$ | $^{+0.024}_{-0.006}$ | $^{+0.023}_{-0.005}$ | $^{+0.028}_{-0.006}$ | $^{+0.020}_{-0.005}$ |
| SP resolution | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| Total | $^{+0.032}_{-0.013}$ | $^{+0.041}_{-0.014}$ | $^{+0.026}_{-0.012}$ | $^{+0.025}_{-0.012}$ | $^{+0.035}_{-0.022}$ | $^{+0.037}_{-0.032}$ |

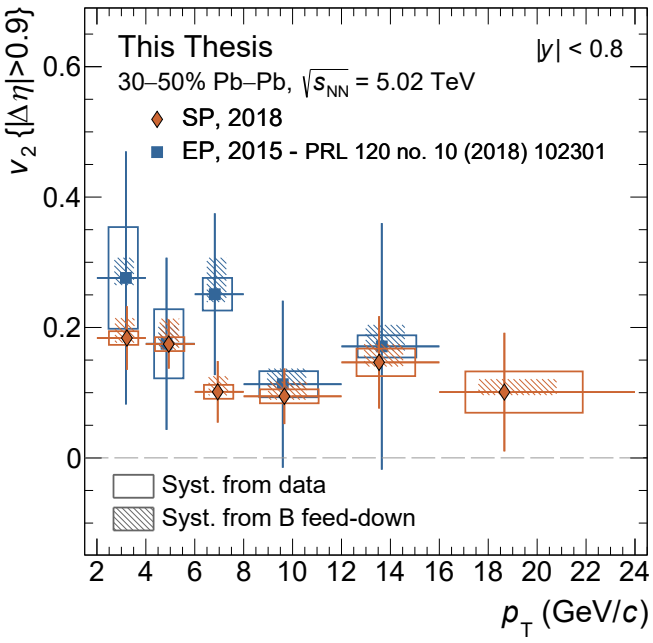



Figure 7.6: Prompt $D_s^+$-meson elliptic flow measured in semicentral Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV obtained in this Thesis (orange diamonds) compared with the previous measurement of the ALICE Collaboration (blue squares) in the same collision system and centrality class [240].

## 7.6 Results

The measured $v_2$ of prompt $D_s^+$ mesons in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV in the 30–50% centrality class is shown in Fig. 7.6. It is compared to the previous measurement of the ALICE Collaboration [240] in the same collision system and centrality class, which was obtained analysing the data sample of Pb–Pb collisions collected in 2015. The two results are in agreement within the statistical and systematic uncertainties. Compared with the previous measurement, the $D_s^+$-meson elliptic flow was measured in a larger transverse-momentum region, which extends up to 24 GeV/$c$. Furthermore, the statistical uncertainties are greatly reduced by up to a factor four thanks to two different factors. The first one is the larger data sample of semicentral Pb–Pb collisions collected in 2018 of about 85 million events, while the 2015 data sample was composed of about 21 million events only. The second factor is the more advanced analysis technique based on machine learning employed for the $D_s^+$-meson candidate selection. The systematic uncertainties are also significantly smaller at low $p_T$ due to the more stable signal extraction.

A positive prompt $D_s^+$-meson elliptic flow is observed in the $2 < p_T < 8$ GeV/$c$ interval with a significance of $6.4\sigma$, considering as null hypothesis $v_2 = 0$. This significance value was obtained by shifting down the measured $v_2$ by the systematic

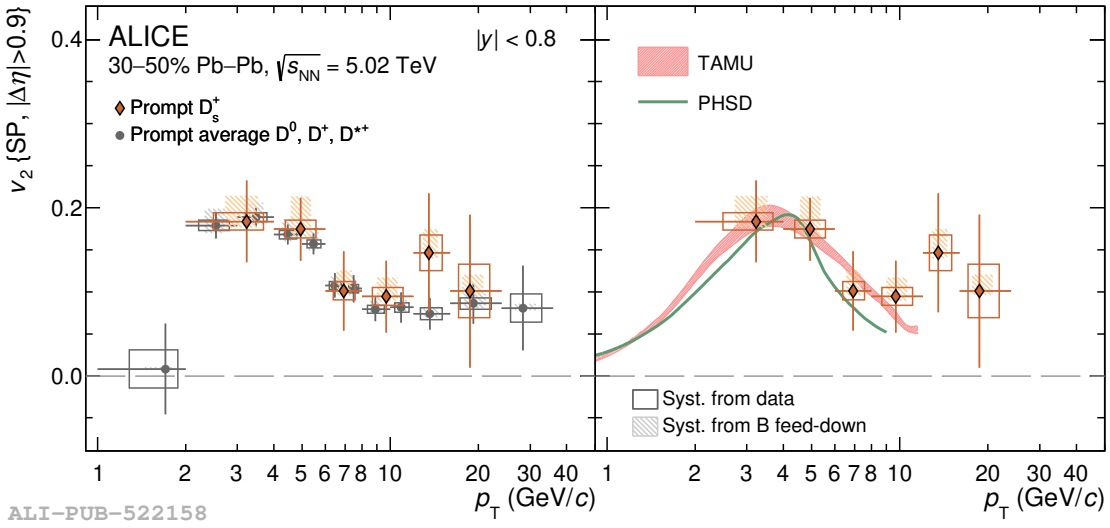



Figure 7.7: Left panel: elliptic flow of prompt $D_s^+$ mesons (orange diamonds) in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV in the 30–50% centrality class compared with that of non-strange D mesons [109] (grey circles). Right panel: prompt $D_s^+$-meson $v_2$ compared with theoretical calculations of the TAMU [110] (red) and PHSD [121] (green) models. Figure from Ref. [193].

uncertainties due to the non-prompt $D_s^+$-meson contribution and the scalar-product resolution, which were considered to be fully correlated among the different $p_T$ intervals. Then, the probability to observe the measured $v_2$ given the null hypothesis was computed summing in quadrature the statistical uncertainty and the systematic one due to the fitting procedure, which was considered as uncorrelated with $p_T$. Finally, the probabilities in each $p_T$ interval were combined to obtain the final result.

Figure 7.7 displays the prompt $D_s^+$-meson elliptic flow compared with the one of prompt non-strange D mesons [109], average of $D^0$, $D^+$, and $D^{*+}$, and with the predictions of the TAMU [110] and PHSD [121] models. The observed positive $v_2$ of promptly produced strange and non-strange D mesons confirms that charm quarks participate in the collective expansion of the system, and that they are at least partially thermalised in the QGP medium. The prompt $D_s^+$-meson elliptic flow is in agreement with the one observed for the non-strange D mesons, and within the current uncertainties no conclusion can be drawn on the possible difference between the two. This difference would be motivated by the different D-meson masses, the charm-quark hadronisation via coalescence with strange quarks of the medium instead of light quarks [147], and different interactions in the hadronic phase [142]. The theoretical model of Ref. [142] predicts a difference between the strange and non-strange D-meson $v_2$ of up to 30%, which is smaller than the current measurement uncertainties. Therefore, to obtain evidence on the possible effect of

the hadronic interactions, the larger data samples of Pb–Pb collisions that will be collected by the ALICE experiment in Run 3 and 4 are needed.

The measured $\mathrm{D_s^+}$-meson elliptic flow is in agreement within uncertainties with the predictions of the TAMU and PHSD models which implement the charm-quark transport in a hydrodynamically-expanding medium, and that include the charm-quark recombination with flowing strange quarks of the medium. These models would significantly underestimate the measured $v_2$ if the hadronisation via recombination would not be considered, and only the charm-quark hadronisation via fragmentation was adopted. Hence, the measurements of the $\mathrm{D_s^+}$-meson production and elliptic flow provide a consistent indication of the role of the hadronisation via recombination at low and intermediate $p_{\mathrm{T}}$.

# Chapter 8

# Conclusions

In this Thesis, the measurements of the production cross section of prompt and non-prompt (i.e., from beauty hadron decays) $\mathrm{D_s^+}$ mesons in pp collisions at a centre-of-mass energy of $\sqrt{s} = 5.02$ TeV were presented, together with the measurements of the prompt $\mathrm{D_s^+}$-meson production yield and azimuthal anisotropy in Pb–Pb collisions at centre-of-mass energy per nucleon pair of $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. The $\mathrm{D_s^+}$ mesons were reconstructed at midrapidity in the $\mathrm{D_s^+} \to \phi\pi^+ \to \mathrm{K^-K^+}\pi^+$ decay channel, and the analyses leveraged on the displacement of the decay vertices from the interaction point. Supervised machine-learning algorithms, namely Boosted Decision Trees, were adopted to perform the particle-candidate selection by rejecting the combinatorial background and, in the case of the analysis in pp collisions, by separating prompt from non-prompt $\mathrm{D_s^+}$ mesons. These machine-learning tools proved crucial to fully exploit the potential of the data samples collected by the ALICE experiment, and they enabled performing the non-prompt $\mathrm{D_s^+}$-meson measurement.

The measurement of the prompt $\mathrm{D_s^+}$-meson $p_{\mathrm{T}}$-differential cross section in pp collisions has significantly smaller uncertainties and is extended to lower $p_{\mathrm{T}}$ with respect to the previous result obtained from the same data sample, which was based on linear selections. The prompt $\mathrm{D_s^+}$-meson cross section is described by pQCD calculations. The non-prompt one is well in agreement with FONLL+PYTHIA 8 calculations, while it is underestimated by the GM-VFNS predictions due to the description of the beauty-hadron decays to $\mathrm{D_s^+}$ mesons. The charm- and beauty-quark fragmentation fraction ratios, $f_{\mathrm{s}}/(f_{\mathrm{u}}+f_{\mathrm{d}})$, are compatible between each other and with previous measurements at different centre-of-mass energies and in different collision systems. This supports the picture of a universal fragmentation of heavy-flavour quarks into mesons, which instead does not hold in the case of fragmentation into baryons. The measured $\mathrm{b\overline{b}}$ production cross section at midrapidity is in agreement with FONLL and NNLO calculations, similarly to what observed for the $\mathrm{c\overline{c}}$ production cross section.

In Pb–Pb collisions, a strong suppression of the $\mathrm{D_s^+}$-meson nuclear modification

factor at intermediate and high $p_{\rm T}$ is observed due to the charm-quark energy loss in the QGP. The suppression increases with increasing collision centrality. The $\rm D_s^+/D^0$ production-yield ratio is higher than that measured in pp collisions in the $2 < p_{\rm T} < 8$ GeV/$c$ interval with a significance slightly larger than $2\sigma$ in both centrality classes. This hints to an enhancement of the $\rm D_s^+$-meson production in heavy-ion collisions due to charm-quark hadronisation via recombination with light-flavour quarks in the strangeness-rich medium. The measurements of the $\rm D_s^+$-meson $R_{\rm AA}$ and the $\rm D_s^+/D^0$ ratio are described by charm-quark transport models that include the hadronisation via recombination, while the $p_{\rm T}$-integrated yield of prompt $\rm D_s^+$ mesons is compatible with the prediction of the statistical hadronisation model for charm quarks (SHMc). In semicentral Pb–Pb collisions, a positive prompt $\rm D_s^+$-mesons elliptic flow is observed with a significance of $6.4\sigma$ in the $2 < p_{\rm T} < 8$ GeV/$c$ interval, thanks to the improved measurement precision compared to the previous result. This observation points to a participation of charm quarks in the QGP collective motions. The prompt $\rm D_s^+$-meson $v_2$ is also compatible with the non-strange D-meson one. Moreover, it is described by TAMU and PHSD model calculations, which include the charm-quark transport in the QGP and the hadronisation via recombination.

The results reported in this Thesis contribute to give insights into the production of charm and beauty mesons in pp collisions. Moreover, they represent the most precise measurement of prompt $\rm D_s^+$-meson production in heavy-ion collisions at LHC energies to date, providing valuable information on the modification of the hadronisation mechanism in the deconfined medium. However, the current precision of the measurement does not permit final conclusions on the enhancement of the charm-strange meson production in heavy-ion collisions. The larger data samples that will be collected by the upgraded ALICE detector in Run 3, accompanied by machine-learning techniques similar to those adopted in this Thesis, will have the potential to further improve and extend to lower $p_{\rm T}$ the measurements of prompt and non-prompt $\rm D_s^+$ mesons in all collision systems. In addition, direct studies of B-meson production in pp and Pb–Pb collisions will be feasible.

# Bibliography


[1] W. Busza, K. Rajagopal, and W. van der Schee, "Heavy Ion Collisions: The Big Picture, and the Big Questions," *Ann. Rev. Nucl. Part. Sci.* **68** (2018) 339–376, arXiv:1802.04801 [hep-ph].

[2] F. Prino and R. Rapp, "Open Heavy Flavor in QCD Matter and in Nuclear Collisions," *J. Phys. G* **43** no. 9, (2016) 093002, arXiv:1603.00529 [nucl-ex].

[3] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.

[4] K. Melnikov, "Lectures on QCD for hadron colliders," *CERN Yellow Rep. School Proc.* **3** (2018) 37–82.

[5] K. G. Wilson, "Confinement of Quarks," *Phys. Rev. D* **10** (1974) 2445–2459.

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

[7] H.-T. Ding, F. Karsch, and S. Mukherjee, "Thermodynamics of strong-interaction matter from Lattice QCD," *Int. J. Mod. Phys. E* **24** no. 10, (2015) 1530007, arXiv:1504.05274 [hep-lat].

[8] G. Martinez, "Advances in Quark Gluon Plasma," arXiv:1304.1452 [nucl-ex].

[9] H. Caines, "The Search for Critical Behavior and Other Features of the QCD Phase Diagram – Current Status and Future Prospects," *Nucl. Phys. A* **967** (2017) 121–128.

[10] M. G. Alford, A. Schmitt, K. Rajagopal, and T. Schäfer, "Color superconductivity in dense quark matter," *Rev. Mod. Phys.* **80** (2008) 1455–1515, arXiv:0709.4635 [hep-ph].

[11] **HotQCD** Collaboration, A. Bazavov et al., "Chiral crossover in QCD at zero and non-zero chemical potentials," *Phys. Lett. B* **795** (2019) 15–21, arXiv:1812.08235 [hep-lat].

[12] S. Borsanyi et al., "QCD Crossover at Finite Chemical Potential from Lattice Simulations," *Phys. Rev. Lett.* **125** no. 5, (2020) 052001, arXiv:2002.02821 [hep-lat].

[13] **BRAHMS** Collaboration, I. C. Arsene et al., "Nuclear stopping and rapidity loss in Au+Au collisions at $\sqrt{s_{\rm NN}} = 62.4$ GeV," *Phys. Lett. B* **677** (2009) 267–271, arXiv:0901.0872 [nucl-ex].

[14] J. D. Bjorken, "Highly Relativistic Nucleus-Nucleus Collisions: The Central Rapidity Region," *Phys. Rev. D* **27** (1983) 140–151.

[15] G. Odyniec, "Beam Energy Scan Program at RHIC (BES I and BES II) — Probing QCD Phase Diagram with Heavy-Ion Collisions," *PoS* **CORFU2018** (2019) 151.

[16] **NICA** Collaboration, A. N. Sissakian, V. D. Kekelidze, and A. S. Sorin, "The nuclotron-based ion collider fAcility at the Joint Institute for Nuclear Research," *Nucl. Phys. A* **827** (2009) 630C–637C.

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

[18] **ALICE** Collaboration, J. Adam et al., "Measurement of transverse energy at midrapidity in Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 2.76$ TeV," *Phys. Rev. C* **94** no. 3, (2016) 034903, arXiv:1603.04775 [nucl-ex].

[19] P. Romatschke and U. Romatschke, *Relativistic Fluid Dynamics In and Out of Equilibrium*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 5, 2019. arXiv:1712.05815 [nucl-th].

[20] **ALICE** Collaboration, K. Aamodt et al., "Two-pion Bose-Einstein correlations in central Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV," *Phys. Lett. B* **696** (2011) 328–337, arXiv:1012.4035 [nucl-ex].

[21] M. L. Miller, K. Reygers, S. J. Sanders, and P. Steinberg, "Glauber modeling in high energy nuclear collisions," *Ann. Rev. Nucl. Part. Sci.* **57** (2007) 205–243, arXiv:nucl-ex/0701025.

[22] **ALICE** Collaboration, "Centrality determination in heavy ion collisions,". https://cds.cern.ch/record/2636623.

[23] **ALICE** Collaboration, J. Adam et al., "Centrality dependence of the charged-particle multiplicity density at midrapidity in Pb-Pb collisions at $\sqrt{s_{\rm NN}} =$ 5.02 TeV," *Phys. Rev. Lett.* **116** no. 22, (2016) 222302, arXiv:1512.06104 [nucl-ex].

[24] U. W. Heinz, "Concepts of heavy ion physics," in *2nd CERN-CLAF School of High Energy Physics*, pp. 165–238. 7, 2004. arXiv:hep-ph/0407360.

[25] G. Torrieri et al., "SHARE: Statistical hadronization with resonances," *Comput. Phys. Commun.* **167** (2005) 229–251, arXiv:nucl-th/0404083.

[26] S. Wheaton and J. Cleymans, "THERMUS: A Thermal model package for ROOT," *Comput. Phys. Commun.* **180** (2009) 84–106, arXiv:hep-ph/0407174.

[27] A. Andronic, P. Braun-Munzinger, and J. Stachel, "Hadron production in central nucleus-nucleus collisions at chemical freeze-out," *Nucl. Phys. A* **772** (2006) 167–199, arXiv:nucl-th/0511071.

[28] J. Rafelski and B. Muller, "Strangeness Production in the Quark-Gluon Plasma," *Phys. Rev. Lett.* **48** (1982) 1066. [Erratum: Phys.Rev.Lett. 56, 2334 (1986)].

[29] O. Kaczmarek and F. Zantow, "Static quark anti-quark interactions in zero and finite temperature QCD. I. Heavy quark free energies, running coupling and quarkonium binding," *Phys. Rev. D* **71** (2005) 114510, arXiv:hep-lat/0503017.

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

[31] **NA57** Collaboration, F. Antinori et al., "Enhancement of hyperon production at central rapidity in 158-A-GeV/c Pb-Pb collisions," *J. Phys. G* **32** (2006) 427–442, arXiv:nucl-ex/0601021.

[32] **STAR** Collaboration, B. I. Abelev et al., "Enhanced strange baryon production in Au + Au collisions compared to p + p at $\sqrt{s_{NN}} = 200$ GeV," *Phys. Rev. C* **77** (2008) 044908, arXiv:0705.2511 [nucl-ex].

[33] **ALICE** Collaboration, B. B. Abelev et al., "Multi-strange baryon production at mid-rapidity in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV," *Phys. Lett. B* **728** (2014) 216–227, arXiv:1307.5543 [nucl-ex]. [Erratum: Phys.Lett.B 734, 409–410 (2014)].

[34] **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].

[35] G. Gatoff and C. Y. Wong, "Origin of the soft p(T) spectra," *Phys. Rev. D* **46** (1992) 997–1006.

[36] **ALICE** Collaboration, B. Abelev et al., "Pion, Kaon, and Proton Production in Central Pb–Pb Collisions at $\sqrt{s_{NN}} = 2.76$ TeV," *Phys. Rev. Lett.* **109** (2012) 252301, arXiv:1208.1974 [hep-ex].

[37] **ALICE** Collaboration, S. Acharya et al., "Production of charged pions, kaons, and (anti-)protons in Pb-Pb and inelastic pp collisions at $\sqrt{s_{NN}} = 5.02$ TeV," *Phys. Rev. C* **101** no. 4, (2020) 044907, arXiv:1910.07678 [nucl-ex].

[38] E. Schnedermann, J. Sollfrank, and U. W. Heinz, "Thermal phenomenology of hadrons from 200-A/GeV S+S collisions," *Phys. Rev. C* **48** (1993) 2462–2475, arXiv:nucl-th/9307020.

[39] S. Voloshin and Y. Zhang, "Flow study in relativistic nuclear collisions by Fourier expansion of Azimuthal particle distributions," *Z. Phys. C* **70** (1996) 665–672, arXiv:hep-ph/9407282.

[40] B. H. Alver, C. Gombeaud, M. Luzum, and J.-Y. Ollitrault, "Triangular flow in hydrodynamics and transport theory," *Phys. Rev. C* **82** (2010) 034913, arXiv:1007.5469 [nucl-th].

[41] D. Teaney and L. Yan, "Non linearities in the harmonic spectrum of heavy ion collisions with ideal and viscous hydrodynamics," *Phys. Rev. C* **86** (2012) 044908, arXiv:1206.1905 [nucl-th].

[42] S. McDonald et al., "Hydrodynamic predictions for Pb+Pb collisions at 5.02 TeV," *Phys. Rev. C* **95** no. 6, (2017) 064913, arXiv:1609.02958 [hep-ph].

[43] W. Zhao, H.-j. Xu, and H. Song, "Collective flow in 2.76 A TeV and 5.02 A TeV Pb+Pb collisions," *Eur. Phys. J. C* **77** no. 9, (2017) 645, arXiv:1703.10792 [nucl-th].

[44] **ALICE** Collaboration, S. Acharya et al., "Anisotropic flow of identified particles in Pb-Pb collisions at $\sqrt{s}_{\mathrm{NN}} = 5.02$ TeV," *JHEP* **09** (2018) 006, arXiv:1805.04390 [nucl-ex].

[45] D. Molnar and S. A. Voloshin, "Elliptic flow at large transverse momenta from quark coalescence," *Phys. Rev. Lett.* **91** (2003) 092301, arXiv:nucl-th/0302014.

[46] **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].

[47] **ALICE** Collaboration, S. Acharya et al., "Measurements of inclusive jet spectra in pp and central Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV," *Phys. Rev. C* **101** no. 3, (2020) 034911, arXiv:1909.09718 [nucl-ex].

[48] **CMS** Collaboration, V. Khachatryan et al., "Charged-particle nuclear modification factors in PbPb and pPb collisions at $\sqrt{s_{\mathrm{N\,N}}} = 5.02$ TeV," *JHEP* **04** (2017) 039, arXiv:1611.01664 [nucl-ex].

[49] **ALICE** Collaboration, J. Adam et al., "Direct photon production in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV," *Phys. Lett. B* **754** (2016) 235–248, arXiv:1509.07324 [nucl-ex].

[50] **ATLAS** Collaboration, G. Aad et al., "Measurement of $Z$ boson Production in Pb+Pb Collisions at $\sqrt{s_{NN}} = 2.76$ TeV with the ATLAS Detector," *Phys. Rev. Lett.* **110** no. 2, (2013) 022301, arXiv:1210.6486 [hep-ex].

[51] **CMS** Collaboration, S. Chatrchyan et al., "Study of $W$ boson production in PbPb and $pp$ collisions at $\sqrt{s_{NN}} = 2.76$ TeV," *Phys. Lett. B* **715** (2012) 66–87, arXiv:1205.6334 [nucl-ex].

[52] **ALICE** Collaboration, J. Adam et al., "Measurement of jet suppression in central Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 2.76$ TeV," *Phys. Lett. B* **746** (2015) 1–14, arXiv:1502.01689 [nucl-ex].

[53] **CMS** Collaboration, V. Khachatryan et al., "Measurement of inclusive jet cross sections in $pp$ and PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV," *Phys. Rev. C* **96** no. 1, (2017) 015202, arXiv:1609.05383 [nucl-ex].

[54] T. Matsui and H. Satz, "$J/\psi$ Suppression by Quark-Gluon Plasma Formation," *Phys. Lett. B* **178** (1986) 416–422.

[55] A. Mocsy and P. Petreczky, "Can quarkonia survive deconfinement?," *Phys. Rev. D* **77** (2008) 014501, arXiv:0705.2559 [hep-ph].

[56] **CMS** Collaboration, A. M. Sirunyan et al., "Measurement of nuclear modification factors of $\Upsilon$(1S), $\Upsilon$(2S), and $\Upsilon$(3S) mesons in PbPb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV," *Phys. Lett. B* **790** (2019) 270–293, arXiv:1805.09215 [hep-ex].

[57] **ALICE** Collaboration, J. Adam et al., "Differential studies of inclusive J/$\psi$ and $\psi$(2S) production at forward rapidity in Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 2.76$ TeV," *JHEP* **05** (2016) 179, arXiv:1506.08804 [nucl-ex].

[58] **PHENIX** Collaboration, A. Adare et al., "$J/\psi$ suppression at forward rapidity in Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV," *Phys. Rev. C* **84** (2011) 054912, arXiv:1103.6269 [nucl-ex].

[59] X. Du and R. Rapp, "Sequential Regeneration of Charmonia in Heavy-Ion Collisions," *Nucl. Phys. A* **943** (2015) 147–158, arXiv:1504.00670 [hep-ph].

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

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

[62] J. C. Collins, D. E. Soper, and G. F. Sterman, "Factorization of Hard Processes in QCD," *Adv. Ser. Direct. High Energy Phys.* **5** (1989) 1–91, arXiv:hep-ph/0409313.

[63] J. Pumplin et al., "New generation of parton distributions with uncertainties from global QCD analysis," *JHEP* **07** (2002) 012, arXiv:hep-ph/0201195.

[64] S. Dulat et al., "New parton distribution functions from a global analysis of quantum chromodynamics," *Phys. Rev. D* **93** no. 3, (2016) 033006, arXiv:1506.07443 [hep-ph].

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

[66] V. G. Kartvelishvili, A. K. Likhoded, and V. A. Petrov, "On the Fragmentation Functions of Heavy Quarks Into Hadrons," *Phys. Lett. B* **78** (1978) 615–617.

[67] M. Cacciari, P. Nason, and C. Oleari, "A Study of heavy flavored meson fragmentation functions in e+ e- annihilation," *JHEP* **04** (2006) 006, arXiv:hep-ph/0510032.

[68] 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].

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

[70] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language," *Nucl. Phys. B* **126** (1977) 298–318.

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

[72] M. Cacciari, M. Greco, and P. Nason, "The $p_{\mathrm{T}}$ spectrum in heavy-flavor hadroproduction," *JHEP* **05** (1998) 007, arXiv:hep-ph/9803400 [hep-ph].

[73] M. Cacciari, S. Frixione, and P. Nason, "The $p_{\mathrm{T}}$ spectrum in heavy-flavor photoproduction," *JHEP* **03** (2001) 006, arXiv:hep-ph/0102134 [hep-ph].

[74] B. A. Kniehl, G. Kramer, I. Schienbein, and H. Spiesberger, "Inclusive Charmed-Meson Production at the CERN LHC," *Eur. Phys. J. C* **72** (2012) 2082, arXiv:1202.0439 [hep-ph].

[75] G. Kramer and H. Spiesberger, "Study of heavy meson production in p–Pb collisions at $\sqrt{S}$= 5.02 TeV in the general-mass variable-flavour-number scheme," *Nucl. Phys. B* **925** (2017) 415–430, arXiv:1703.04754 [hep-ph].

[76] M. Benzke et al., "Prompt neutrinos from atmospheric charm in the general-mass variable-flavor-number scheme," *JHEP* **12** (2017) 021, arXiv:1705.10386 [hep-ph].

[77] P. Bolzoni and G. Kramer, "Inclusive charmed-meson production from bottom hadron decays at the LHC," *J. Phys. G* **41** (2014) 075006, arXiv:1310.2924 [hep-ph].

[78] R. Maciuła and A. Szczurek, "Production of $\Lambda_c$ baryons at the LHC within the $k_T$-factorization approach and independent parton fragmentation picture," *Phys. Rev. D* **98** no. 1, (2018) 014016, arXiv:1803.05807 [hep-ph].

[79] R. Maciula and A. Szczurek, "Consistent treatment of charm production in higher-orders at tree-level within $k_T$-factorization approach," *Phys. Rev. D* **100** no. 5, (2019) 054001, arXiv:1905.06697 [hep-ph].

[80] S. Catani et al., "Bottom-quark production at hadron colliders: fully differential predictions in NNLO QCD," *JHEP* **03** (2021) 029, arXiv:2010.11906 [hep-ph].

[81] D. d'Enterria and A. M. Snigirev, "Triple parton scatterings in high-energy proton-proton collisions," *Phys. Rev. Lett.* **118** no. 12, (2017) 122001, arXiv:1612.05582 [hep-ph].

[82] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175.

[83] G. Marchesini et al., "HERWIG: A Monte Carlo event generator for simulating hadron emission reactions with interfering gluons. Version 5.1 – April 1991," *Comput. Phys. Commun.* **67** (1992) 465–508.

[84] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations," *JHEP* **06** (2002) 029, arXiv:hep-ph/0204244.

[85] S. Frixione, P. Nason, and G. Ridolfi, "A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction," *JHEP* **09** (2007) 126, arXiv:0707.3088 [hep-ph].

[86] H. Paukkunen, "Nuclear PDFs in the beginning of the LHC era," *Nucl. Phys. A* **926** (2014) 24–33, arXiv:1401.2345 [hep-ph].

[87] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, "EPPS16: Nuclear parton distributions with LHC data," *Eur. Phys. J. C* **77** no. 3, (2017) 163, arXiv:1612.05741 [hep-ph].

[88] N. Armesto, "Nuclear shadowing," *J. Phys. G* **32** (2006) R367–R394, arXiv:hep-ph/0604108.

[89] K. Rith, "Present Status of the EMC effect," *Subnucl. Ser.* **51** (2015) 431–449, arXiv:1402.5000 [hep-ex].

[90] N. Carrer and A. Dainese, "Charm and beauty production at LHC," arXiv:hep-ph/0311225. ALICE-INT-2003-019.

[91] K. Kovarik et al., "nCTEQ15 — Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework," *Phys. Rev. D* **93** no. 8, (2016) 085037, arXiv:1509.00792 [hep-ph].

[92] J. W. Cronin et al., "Production of hadrons with large transverse momentum at 200, 300, and 400 GeV," *Phys. Rev. D* **11** (1975) 3105–3123.

[93] A. M. Sickles, "Possible Evidence for Radial Flow of Heavy Mesons in d+Au Collisions," *Phys. Lett. B* **731** (2014) 51–56, arXiv:1309.6924 [nucl-th].

[94] A. Beraudo et al., "Heavy-flavour production in high-energy d-Au and p-Pb collisions," *JHEP* **03** (2016) 123, arXiv:1512.05186 [hep-ph].

[95] Y. Xu et al., "Heavy-flavor dynamics in relativistic p-Pb collisions at $\sqrt{S_{NN}} = 5.02$ TeV," *Nucl. Part. Phys. Proc.* **276-278** (2016) 225–228, arXiv:1510.07520 [nucl-th].

[96] G. D. Moore and D. Teaney, "How much do heavy quarks thermalize in a heavy ion collision?," *Phys. Rev. C* **71** (2005) 064904, arXiv:hep-ph/0412346.

[97] S. Cao, G.-Y. Qin, and S. A. Bass, "Heavy-quark dynamics and hadronization in ultrarelativistic heavy-ion collisions: Collisional versus radiative energy loss," *Phys. Rev. C* **88** (2013) 044907, arXiv:1308.0617 [nucl-th].

[98] H. van Hees, V. Greco, and R. Rapp, "Heavy-quark probes of the quark-gluon plasma at RHIC," *Phys. Rev. C* **73** (2006) 034913, arXiv:nucl-th/0508055.

[99] S. Peigne and A. Peshier, "Collisional energy loss of a fast heavy quark in a quark-gluon plasma," *Phys. Rev. D* **77** (2008) 114017, arXiv:0802.4364 [hep-ph].

[100] R. Baier et al., "Radiative energy loss and p(T) broadening of high-energy partons in nuclei," *Nucl. Phys. B* **484** (1997) 265–282, arXiv:hep-ph/9608322.

[101] **JETSCAPE** Collaboration, S. Cao et al., "Determining the jet transport coefficient $\hat{q}$ from inclusive hadron suppression measurements using Bayesian parameter estimation," *Phys. Rev. C* **104** no. 2, (2021) 024905, arXiv:2102.11337 [nucl-th].

[102] Y. L. Dokshitzer and D. E. Kharzeev, "Heavy quark colorimetry of QCD matter," *Phys. Lett. B* **519** (2001) 199–206, arXiv:hep-ph/0106202.

[103] **ALICE** Collaboration, S. Acharya et al., "Direct observation of the dead-cone effect in QCD," arXiv:2106.05713 [nucl-ex].

[104] **ALICE** Collaboration, S. Acharya et al., "Prompt $D^0$, $D^+$, and $D^{*+}$ production in Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV," arXiv:2110.09420 [nucl-ex].

[105] **ALICE** Collaboration, S. Acharya et al., "Centrality and transverse momentum dependence of inclusive J/$\psi$ production at midrapidity in Pb–Pb collisions at sNN=5.02 TeV," *Phys. Lett. B* **805** (2020) 135434, arXiv:1910.14404 [nucl-ex].

[106] **CMS** Collaboration, A. M. Sirunyan et al., "Measurement of prompt and nonprompt charmonium suppression in PbPb collisions at 5.02 TeV," *Eur. Phys. J. C* **78** no. 6, (2018) 509, arXiv:1712.08959 [nucl-ex].

[107] M. Gyulassy, I. Vitev, and X. N. Wang, "High p(T) azimuthal asymmetry in noncentral A+A at RHIC," *Phys. Rev. Lett.* **86** (2001) 2537–2540, arXiv:nucl-th/0012092.

[108] E. V. Shuryak, "The Azimuthal asymmetry at large p(t) seem to be too large for a 'jet quenching'," *Phys. Rev. C* **66** (2002) 027902, arXiv:nucl-th/0112042.

[109] **ALICE** Collaboration, S. Acharya et al., "Transverse-momentum and event-shape dependence of D-meson flow harmonics in Pb–Pb collisions at $\sqrt{s_{NN}}$ = 5.02 TeV," *Phys. Lett. B* **813** (2021) 136054, arXiv:2005.11131 [nucl-ex].

[110] M. He and R. Rapp, "Hadronization and Charm-Hadron Ratios in Heavy-Ion Collisions," *Phys. Rev. Lett.* **124** no. 4, (2020) 042301, arXiv:1905.09216 [nucl-th].

[111] A. Beraudo et al., "Heavy flavors in heavy-ion collisions: quenching, flow and correlations," *Eur. Phys. J. C* **75** no. 3, (2015) 121, arXiv:1410.6082 [hep-ph].

[112] A. Beraudo et al., "Development of heavy-flavour flow-harmonics in high-energy nuclear collisions," *JHEP* **02** (2018) 043, arXiv:1712.00588 [hep-ph].

[113] R. Katz et al., "Sensitivity study with a D and B mesons modular simulation code of heavy flavor RAA and azimuthal anisotropies based on beam energy, initial conditions, hadronization, and suppression mechanisms," *Phys. Rev. C* **102** no. 2, (2020) 024906, arXiv:1906.10768 [nucl-th].

[114] S. Li and J. Liao, "Data-driven extraction of heavy quark diffusion in quark-gluon plasma," *Eur. Phys. J. C* **80** no. 7, (2020) 671, arXiv:1912.08965 [hep-ph].

[115] F. Scardina et al., "Estimating the charm quark diffusion coefficient and thermalization time from D meson spectra at energies available at the BNL Relativistic Heavy Ion Collider and the CERN Large Hadron Collider," *Phys. Rev. C* **96** no. 4, (2017) 044905, arXiv:1707.05452 [nucl-th].

[116] S. Plumari et al., "Heavy — light flavor correlations of anisotropic flows at LHC energies within event-by-event transport approach," *Phys. Lett. B* **805** (2020) 135460, arXiv:1912.09350 [hep-ph].

[117] M. Nahrgang, J. Aichelin, P. B. Gossiaux, and K. Werner, "Influence of hadronic bound states above $T_c$ on heavy-quark observables in Pb + Pb collisions at the CERN Large Hadron Collider," *Phys. Rev. C* **89** no. 1, (2014) 014905, arXiv:1305.6544 [hep-ph].

[118] S. Cao, T. Luo, G.-Y. Qin, and X.-N. Wang, "Linearized Boltzmann transport model for jet propagation in the quark-gluon plasma: Heavy quark evolution," *Phys. Rev. C* **94** no. 1, (2016) 014909, arXiv:1605.06447 [nucl-th].

[119] S. Cao, T. Luo, G.-Y. Qin, and X.-N. Wang, "Heavy and light flavor jet quenching at RHIC and LHC energies," *Phys. Lett. B* **777** (2018) 255–259, arXiv:1703.00822 [nucl-th].

[120] W. Ke, Y. Xu, and S. A. Bass, "Modified Boltzmann approach for modelling the splitting vertices induced by the hot QCD medium in the deep Landau-Pomeranchuk-Migdal region," *Phys. Rev. C* **100** no. 6, (2019) 064911, arXiv:1810.08177 [nucl-th].

[121] T. Song et al., "Tomography of the Quark-Gluon-Plasma by Charm Quarks," *Phys. Rev. C* **92** no. 1, (2015) 014910, arXiv:1503.03039 [nucl-th].

[122] J. Xu, J. Liao, and M. Gyulassy, "Consistency of Perfect Fluidity and Jet Quenching in semi-Quark-Gluon Monopole Plasmas," *Chin. Phys. Lett.* **32** no. 9, (2015) 092501, arXiv:1411.3673 [hep-ph].

[123] J. Xu, J. Liao, and M. Gyulassy, "Bridging Soft-Hard Transport Properties of Quark-Gluon Plasmas with CUJET3.0" *JHEP* **02** (2016) 169, arXiv:1508.00552 [hep-ph].

[124] S. Shi, J. Liao, and M. Gyulassy, "Global constraints from RHIC and LHC on transport properties of QCD fluids in CUJET/CIBJET framework," *Chin. Phys. C* **43** no. 4, (2019) 044101, arXiv:1808.05461 [hep-ph].

[125] S. Stojku, B. Ilic, M. Djordjevic, and M. Djordjevic, "Extracting the temperature dependence in high-$p_\perp$ particle energy loss," *Phys. Rev. C* **103** no. 2, (2021) 024908, arXiv:2007.07851 [nucl-th].

[126] D. Zigic et al., "DREENA-B framework: first predictions of $R_{AA}$ and $v_2$ within dynamical energy loss formalism in evolving QCD medium," *Phys. Lett. B* **791** (2019) 236–241, arXiv:1805.04786 [nucl-th].

[127] D. Zigic et al., "DREENA-C framework: joint $R_{AA}$ and $v_2$ predictions and implications to QGP tomography," *J. Phys. G* **46** no. 8, (2019) 085101, arXiv:1805.03494 [nucl-th].

[128] Z.-B. Kang, F. Ringer, and I. Vitev, "Effective field theory approach to open heavy flavor production in heavy-ion collisions," *JHEP* **03** (2017) 146, arXiv:1610.02043 [hep-ph].

[129] M. Gyulassy, P. Levai, and I. Vitev, "Reaction operator approach to non-Abelian energy loss," *Nucl. Phys. B* **594** (2001) 371–419, arXiv:nucl-th/0006010.

[130] M. Djordjevic, "Collisional energy loss in a finite size QCD matter," *Phys. Rev. C* **74** (2006) 064907, arXiv:nucl-th/0603066.

[131] C. W. Bauer, S. Fleming, D. Pirjol, and I. W. Stewart, "An Effective field theory for collinear and soft gluons: Heavy to light decays," *Phys. Rev. D* **63** (2001) 114020, arXiv:hep-ph/0011336.

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

[133] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, "Parton Fragmentation and String Dynamics," *Phys. Rept.* **97** (1983) 31–145.

[134] S. Ferreres-Solé and T. Sjöstrand, "The space–time structure of hadronization in the Lund model," *Eur. Phys. J. C* **78** no. 11, (2018) 983, arXiv:1808.04619 [hep-ph].

[135] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 Tune," *Eur. Phys. J. C* **74** no. 8, (2014) 3024, arXiv:1404.5630 [hep-ph].

[136] **ALICE** Collaboration, S. Acharya et al., "$\Lambda_c^+$ Production and Baryon-to-Meson Ratios in pp and p-Pb Collisions at $\sqrt{s_{NN}}$=5.02 TeV at the LHC," *Phys. Rev. Lett.* **127** no. 20, (2021) 202301, arXiv:2011.06078 [nucl-ex].

[137] R. J. Fries, B. Muller, C. Nonaka, and S. A. Bass, "Hadronization in heavy ion collisions: Recombination and fragmentation of partons," *Phys. Rev. Lett.* **90** (2003) 202303, arXiv:nucl-th/0301087.

[138] V. Greco, C. Ko, and P. Levai, "Parton coalescence and anti-proton / pion anomaly at RHIC," *Phys. Rev. Lett.* **90** (2003) 202302, arXiv:nucl-th/0301093.

[139] C. B. Dover, U. W. Heinz, E. Schnedermann, and J. Zimanyi, "Relativistic coalescence model for high-energy nuclear collisions," *Phys. Rev. C* **44** (1991) 1636–1654.

[140] L. Ravagli and R. Rapp, "Quark Coalescence based on a Transport Equation," *Phys. Lett. B* **655** (2007) 126–131, arXiv:0705.0021 [hep-ph].

[141] I. Kuznetsova and J. Rafelski, "Heavy flavor hadrons in statistical hadronization of strangeness-rich QGP," *Eur. Phys. J. C* **51** (2007) 113–133, arXiv:hep-ph/0607203.

[142] M. He, R. J. Fries, and R. Rapp, "$\mathbf{D_s}$-Meson as Quantitative Probe of Diffusion and Hadronization in Nuclear Collisions," *Phys. Rev. Lett.* **110** no. 11, (2013) 112301, arXiv:1204.4442 [nucl-th].

[143] A. Andronic et al., "The multiple-charm hierarchy in the statistical hadronization model," *JHEP* **07** (2021) 035, arXiv:2104.12754 [hep-ph].

[144] A. Andronic et al., "Transverse momentum distributions of charmonium states with the statistical hadronization model," *Phys. Lett. B* **797** (2019) 134836, arXiv:1901.09200 [nucl-th].

[145] A. Mazeliauskas, S. Floerchinger, E. Grossi, and D. Teaney, "Fast resonance decays in nuclear collisions," *Eur. Phys. J. C* **79** no. 3, (2019) 284, arXiv:1809.11049 [nucl-th].

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

[147] D. Molnar, "Charm elliptic flow from quark coalescence dynamics," *J. Phys. G* **31** (2005) S421–S428, arXiv:nucl-th/0410041.

[148] **ALICE** Collaboration, K. Aamodt et al., "The ALICE experiment at the CERN LHC," *JINST* **3** (2008) S08002.

[149] L. Evans and P. Bryant, "LHC Machine," *JINST* **3** (2008) S08001.

[150] CERN, "LEP design report,". http://cds.cern.ch/record/102083.

[151] K. Schindl, "The injector chain for the LHC," in *9th LEP Performance Workshop*, pp. 47–52. 3, 1999.

[152] A. Beuret et al., "The LHC Lead Ion Injector Chain," in *9th European Particle Accelerator Conference (EPAC 2004)*. 7, 2004.

[153] J. Haffner, "The cern accelerator complex," 2013. https://cds.cern.ch/record/1621894/.

[154] **ATLAS** Collaboration, G. Aad et al., "The ATLAS experiment at the CERN large hadron collider," *JINST* **3** (2008) S08003.

[155] **CMS** Collaboration, S. Chatrchyan et al., "The CMS experiment at the CERN LHC," *JINST* **3** (2008) S08004.

[156] **ATLAS** Collaboration, G. Aad et al., "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys. Lett. B* **716** (2012) 1–29, arXiv:1207.7214 [hep-ex].

[157] **CMS** Collaboration, S. Chatrchyan et al., "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC," *Phys. Lett. B* **716** (2012) 30–61, arXiv:1207.7235 [hep-ex].

[158] **LHCb** Collaboration, A. A. Alves et al., "The LHCb detector at the LHC," *JINST* **3** (2008) S08005.

[159] **ALICE** Collaboration, B. B. Abelev et al., "Performance of the ALICE Experiment at the CERN LHC," *Int. J. Mod. Phys. A* **29** (2014) 1430044, arXiv:1402.4476 [nucl-ex].

[160] **ALICE** Collaboration, K. Aamodt et al., "Alignment of the ALICE Inner Tracking System with cosmic-ray tracks," *JINST* **5** (2010) P03003, arXiv:1001.0502.

[161] J. Alme et al., "The ALICE TPC, a large 3-dimensional tracking device with fast readout for ultra-high multiplicity events," *Nucl. Instrum. Meth. A* **622** (2010) 316–367, arXiv:1001.1950.

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

[163] **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** no. 4, (2013) 044909, arXiv:1301.4361 [nucl-ex].

[164] **ALICE** Collaboration, Vv. Aa., "AliRoot," 1998–. https://alice-offline.web.cern.ch/.

[165] **ALICE** Collaboration, Vv. Aa., "AliPhysics." https://github.com/alisw/AliPhysics.

[166] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework," in *New computing techniques in physics research V. Proceedings, 5th International Workshop, AIHENP '96, Lausanne, Switzerland, September 2-6, 1996*, vol. A389, pp. 81–86. 1997.

[167] F. Rademakers et al., "root-project/root: v6.20/04," Apr., 2020. https://doi.org/10.5281/zenodo.3895855.

[168] T. Sjöstrand et al., "An introduction to PYTHIA 8.2" *Comput. Phys. Commun.* **191** (2015) 159–177, arXiv:1410.3012 [hep-ph].

[169] 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** (1991) 3501–3516.

[170] R. Brun et al., *GEANT: Detector Description and Simulation Tool; Oct 1994*. CERN Program Library. CERN, Geneva, 1993. http://cds.cern.ch/record/1082634. Long Writeup W5013.

[171] **GEANT4** Collaboration, S. Agostinelli et al., "GEANT4: A Simulation toolkit," *Nucl. Instrum. Meth.* **A506** (2003) 250–303.

[172] A. Ferrari, P. R. Sala, A. Fassò, and J. Ranft, *FLUKA: A multi-particle transport code (program version 2005)*. CERN Yellow Reports: Monographs. CERN, Geneva, 2005. https://cds.cern.ch/record/898301.

[173] K. Bos et al., *LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)*. Technical design report. LCG. CERN, Geneva, 2005. https://cds.cern.ch/record/840543.

[174] Vv. Aa., "Worldwide lhc computing grid." https://wlcg-public.web.cern.ch/.

[175] **ALICE** Collaboration, P. Saiz et al., "AliEn - ALICE environment on the GRID," *Nucl. Instrum. Meth. A* **502** (2003) 437–440.

[176] **ALICE** Collaboration, Vv. Aa., "Alien environment." https://alien.web.cern.ch/.

[177] D. Kharzeev, E. Levin, and M. Nardi, "Color glass condensate at the LHC: Hadron multiplicities in pp, pA and AA collisions," *Nucl. Phys. A* **747** (2005) 609–629, arXiv:hep-ph/0408050.

[178] W.-T. Deng, X.-N. Wang, and R. Xu, "Hadron production in p+p, p+Pb, and Pb+Pb collisions with the HIJING 2.0 model at energies available at the CERN Large Hadron Collider," *Phys. Rev. C* **83** (2011) 014915, arXiv:1008.1841 [hep-ph].

[179] R. Fruhwirth, "Application of Kalman filtering to track and vertex fitting," *Nucl. Instrum. Meth. A* **262** (1987) 444–450.

[180] W. Blum, W. Riegler, and L. Rolandi, *Particle detection with drift chambers; 2nd ed.* Springer, Berlin, 2008. https://cds.cern.ch/record/1105920.

[181] C. Lippmann, "Performance of the ALICE Time Projection Chamber," *Phys. Procedia* **37** (2012) 434–441.

[182] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[183] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM* **60** no. 6, (May, 2017) 84–90.

[184] O. Russakovsky et al., "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vis.* **115** no. 3, (2015) 211–252.

[185] D. Silver et al., "Mastering the game of Go with deep neural networks and tree search," *Nature* **529** no. 7587, (Jan., 2016) 484–489.

[186] Y. E. Wang, G.-Y. Wei, and D. Brooks, "Benchmarking TPU, GPU, and CPU Platforms for Deep Learning," arXiv:1907.10701 [cs.LG].

[187] M. Svedin et al., "Benchmarking the Nvidia GPU Lineage: From Early K80 to Modern A100 with Asynchronous Memory Transfers," arXiv:2106.04979 [cs.DC].

[188] A. Radovic et al., "Machine learning at the energy and intensity frontiers of particle physics," *Nature* **560** no. 7716, (2018) 41–48.

[189] K. Albertsson et al., "Machine Learning in High Energy Physics Community White Paper," *J. Phys. Conf. Ser.* **1085** no. 2, (2018) 022008, arXiv:1807.02876 [physics.comp-ph].

[190] D. Guest, K. Cranmer, and D. Whiteson, "Deep Learning and its Application to LHC Physics," *Ann. Rev. Nucl. Part. Sci.* **68** (2018) 161–181, arXiv:1806.11484 [hep-ex].

[191] J. Lin, M. Freytsis, I. Moult, and B. Nachman, "Boosting $H \to b\bar{b}$ with Machine Learning," *JHEP* **10** (2018) 101, arXiv:1807.10768 [hep-ph].

[192] R. Di Sipio, M. Faucci Giannelli, S. Ketabchi Haghighat, and S. Palazzo, "DijetGAN: A Generative-Adversarial Network Approach for the Simulation of QCD Dijet Events at the LHC," *JHEP* **08** (2019) 110, arXiv:1903.02433 [hep-ex].

[193] **ALICE** Collaboration, S. Acharya et al., "Measurement of prompt $\mathrm{D_s^+}$-meson production and azimuthal anisotropy in Pb–Pb collisions at $\sqrt{s_{NN}}$=5.02TeV," *Phys. Lett. B* **827** (2022) 136986, arXiv:2110.10006 [nucl-ex].

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

[195] **ALICE** Collaboration, S. Acharya et al., "$\Lambda_c^+$ production in pp and in p-Pb collisions at $\sqrt{s_{NN}}$=5.02 TeV," *Phys. Rev. C* **104** no. 5, (2021) 054905, arXiv:2011.06079 [nucl-ex].

[196] **ALICE** Collaboration, S. Acharya et al., "Measurement of the Cross Sections of $\Xi_c^0$ and $\Xi_c^+$ Baryons and of the Branching-Fraction Ratio $\mathrm{BR}(\Xi_c^0 \to \Xi^- e^+ \nu_e)/\mathrm{BR}(\Xi_c^0 \to \Xi^- \pi^+)$ in pp collisions at 13 TeV," *Phys. Rev. Lett.* **127** no. 27, (2021) 272001, arXiv:2105.05187 [nucl-ex].

[197] **ALICE** Collaboration, S. Acharya et al., "Measurement of Prompt $\mathrm{D^0}$, $\Lambda_c^+$, and $\Sigma_c^{0,++}(2455)$ Production in Proton–Proton Collisions at $\sqrt{s} = 13$ TeV," *Phys. Rev. Lett.* **128** no. 1, (2022) 012001, arXiv:2106.08278 [hep-ex].

[198] **ALICE** Collaboration, S. Acharya et al., "Hypertriton production in p-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV," arXiv:2107.10627 [nucl-ex].

[199] **ALICE** Collaboration, S. Acharya et al., "Observation of a multiplicity dependence in the $p_{\mathrm{T}}$-differential charm baryon-to-meson ratios in proton–proton collisions at s=13 TeV," *Phys. Lett. B* **829** (2022) 137065, arXiv:2111.11948 [nucl-ex].

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

[201] **ALICE** Collaboration, S. Acharya et al., "First study of the two-body scattering involving charm hadrons," arXiv:2201.05352 [nucl-ex].

[202] **ALICE** Collaboration, S. Acharya et al., "Measurement of beauty production via non-prompt $D^0$ mesons in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV," arXiv:2202.00815 [nucl-ex].

[203] **ALICE** Collaboration, "Measurement of beauty-strange meson production in Pb−Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV via non-prompt $D_s^+$ mesons," arXiv:2204.10386 [nucl-ex].

[204] S. Gorbunov et al., "Deep neural network techniques in the calibration of space-charge distortion fluctuations for the ALICE TPC," *EPJ Web Conf.* **251** (2021) 03020.

[205] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics. Springer, 2009.

[206] D. J. Hand and R. J. Till, "A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems," *Machine Learning* **45** no. 2, (Nov., 2001) 171–186.

[207] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research* **12** no. 85, (2011) 2825–2830. https://www.jmlr.org/papers/v12/pedregosa11a.html.

[208] Vv. Aa., "Scikit-learn documentation," 2021. http://scikit-learn.org/stable/user_guide.html.

[209] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research* **13** no. 10, (2012) 281–305. http://jmlr.org/papers/v13/bergstra12a.html.

[210] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in Neural Information Processing Systems*, vol. 25. 2012. https://proceedings.neurips.cc/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf.

[211] F. Nogueira, "Bayesian Optimization: Open source constrained global optimization tool for Python," 2014–. https://github.com/fmfn/BayesianOptimization.

[212] L. Breiman, "Random Forests," *Machine Learning* **45** no. 1, (Oct., 2001) 5–32.

[213] S. B. Kotsiantis, "Decision trees: a recent overview," *Artificial Intelligence Review* **39** no. 4, (Apr., 2013) 261–283.

[214] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences* **55** no. 1, (1997) 119–139.

[215] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," *The Annals of Statistics* **28** no. 2, (2000) 337 – 407.

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

[217] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pp. 785–794. New York, NY, USA, 2016.

[218] R. Mitchell and E. Frank, "Accelerating the xgboost algorithm using gpu computing," *PeerJ Computer Science* **3** (2017) e127.

[219] Vv. Aa., "Xgboost documentation," 2021. https://xgboost.readthedocs.io/en/latest/.

[220] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.* **3** no. null, (Mar., 2003) 1157–1182. https://www.jmlr.org/papers/v3/guyon03a.html.

[221] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30*, pp. 4765–4774. 2017. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

[222] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," *Nature Machine Intelligence* **2** no. 1, (Jan., 2020) 56–67.

[223] J. Pivarski et al., "scikit-hep/uproot4: 4.0.11," June, 2021. https://doi.org/10.5281/zenodo.5047600.

[224] L. Barioglio et al., "hipe4ml/hipe4ml," July, 2021. https://doi.org/10.5281/zenodo.5070132.

[225] F. Catalano et al., "DmesonAnalysis," 2019–. https://github.com/DmesonAnalysers/DmesonAnalysis.

[226] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering* **9** no. 3, (2007) 90–95.

[227] C. R. Harris et al., "Array programming with NumPy," *Nature* **585** no. 7825, (Sept., 2020) 357–362.

[228] Wes McKinney, "Data Structures for Statistical Computing in Python," in *Proceedings of the 9th Python in Science Conference*, pp. 56 – 61. 2010.

[229] J. Reback et al., "pandas-dev/pandas: Pandas 1.1.5," Dec., 2020. https://doi.org/10.5281/zenodo.4309786.

[230] P. Virtanen et al., "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods* **17** (2020) 261–272.

[231] Vv. Aa., "treelite," 2017–. https://github.com/dmlc/treelite.

[232] K. Albertsson et al., "Fast Inference for Machine Learning in ROOT/TMVA," in *European Physical Journal Web of Conferences*, vol. 245, p. 06008. Nov., 2020.

[233] **ALICE** Collaboration, "ALICE 2017 luminosity determination for pp collisions at $\sqrt{s} = 5.02$ TeV,". http://cds.cern.ch/record/2648933.

[234] **ALICE** Collaboration, "ALICE luminosity determination for Pb−Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV" arXiv:2204.10148 [nucl-ex].

[235] **ZEUS** Collaboration, S. Chekanov et al., "Measurement of charm fragmentation ratios and fractions in photoproduction at HERA," *Eur. Phys. J. C* **44** (2005) 351–366, arXiv:hep-ex/0508019.

[236] M. Faggin, *Measurement of heavy-flavour decay electrons and heavy-flavour baryon production with ALICE experiment at LHC*. PhD thesis, Università degli Studi di Padova, 2021. https://cds.cern.ch/record/2798337.

[237] **LHCb** Collaboration, R. Aaij et al., "Measurements of prompt charm production cross-sections in pp collisions at $\sqrt{s} = 5$ TeV," *JHEP* **06** (2017) 147, arXiv:1610.02230 [hep-ex].

[238] **ALICE** Collaboration, S. Acharya et al., "Measurement of $\rm D^0$ , $\rm D^+$ , $\rm D^{*+}$ and $\rm D_s^+$ production in pp collisions at $\sqrt{s} = 5.02$ TeV with ALICE," *Eur. Phys. J. C* **79** no. 5, (2019) 388, arXiv:1901.07979 [nucl-ex].

[239] **ALICE** Collaboration, J. Adam et al., "Measurement of $\rm D_s^+$ production and nuclear modification factor in Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 2.76$ TeV," *JHEP* **03** (2016) 082, arXiv:1509.07287 [nucl-ex].

[240] **ALICE** Collaboration, S. Acharya et al., "$D$-meson azimuthal anisotropy in midcentral Pb-Pb collisions at $\sqrt{\mathbf{s_{\rm NN}}} = \mathbf{5.02}$ TeV," *Phys. Rev. Lett.* **120** no. 10, (2018) 102301, arXiv:1707.01005 [nucl-ex].

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

[242] D. Freedman, R. Pisani, and R. Purves, *Statistics (international student edition)*. W. W. Norton & Company, 2007.

[243] Vv. Aa., "Xgboost forum: Probabilities returned by multi:softprob," 2020. https://discuss.xgboost.ai/t/probabilities-returned-by-multi-softprob/1745.

[244] E. Braaten, K.-m. Cheung, S. Fleming, and T. C. Yuan, "Perturbative QCD fragmentation functions as a model for heavy quark fragmentation," *Phys. Rev. D* **51** (1995) 4819–4829, arXiv:hep-ph/9409316.

[245] **ALEPH** Collaboration, R. Barate et al., "Study of charm production in Z decays," *Eur. Phys. J. C* **16** (2000) 597–611, arXiv:hep-ex/9909032.

[246] **ATLAS** Collaboration, G. Aad et al., "Measurement of the differential cross-section of $B^+$ meson production in pp collisions at $\sqrt{s} = 7$ TeV at ATLAS," *JHEP* **10** (2013) 042, arXiv:1307.0126 [hep-ex].

[247] **CMS** Collaboration, A. M. Sirunyan et al., "Measurement of the $B^{\pm}$ Meson Nuclear Modification Factor in Pb-Pb Collisions at $\sqrt{s_{NN}} = 5.02$ TeV," *Phys. Rev. Lett.* **119** no. 15, (2017) 152301, arXiv:1705.04727 [hep-ex].

[248] **LHCb** Collaboration, R. Aaij et al., "Measurement of the $B^{\pm}$ production cross-section in $pp$ collisions at $\sqrt{s} = 7$ TeV," *JHEP* **04** (2012) 093, arXiv:1202.4812 [hep-ex].

[249] P. Z. Skands, "Tuning Monte Carlo Generators: The Perugia Tunes," *Phys. Rev. D* **82** (2010) 074018, arXiv:1005.3457 [hep-ph].

[250] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, pp. 226–231. AAAI Press, 1996.

[251] F. Grosa, *Strange and non-strange D-meson production in pp, p-Pb, and Pb-Pb collisions with ALICE at the LHC*. PhD thesis, Politecnico di Torino, 2020. http://cds.cern.ch/record/2713513.

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

[253] B. A. Kniehl and G. Kramer, "Charmed-hadron fragmentation functions from CERN LEP1 revisited," *Phys. Rev. D* **74** (2006) 037502, arXiv:hep-ph/0607306.

[254] **ALICE** Collaboration, B. Abelev et al., "$D_s^+$ meson production at central rapidity in proton–proton collisions at $\sqrt{s} = 7$ TeV," *Phys. Lett. B* **718** (2012) 279–294, arXiv:1208.1948 [hep-ex].

[255] **H1** Collaboration, A. Aktas et al., "Inclusive production of D+, D0, D+(s) and D*+ mesons in deep inelastic scattering at HERA," *Eur. Phys. J. C* **38** (2005) 447–459, arXiv:hep-ex/0408149.

[256] **ZEUS** Collaboration, H. Abramowicz et al., "Measurement of charm fragmentation fractions in photoproduction at HERA," *JHEP* **09** (2013) 058, arXiv:1306.4862 [hep-ex].

[257] **ATLAS** Collaboration, G. Aad et al., "Measurement of $D^{*\pm}$, $D^{\pm}$ and $D_s^{\pm}$ meson production cross sections in $pp$ collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector," *Nucl. Phys. B* **907** (2016) 717–763, arXiv:1512.02913 [hep-ex].

[258] **CDF** Collaboration, T. Aaltonen et al., "Measurement of Ratios of Fragmentation Fractions for Bottom Hadrons in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV," *Phys. Rev. D* **77** (2008) 072003, arXiv:0801.4375 [hep-ex].

[259] **LHCb** Collaboration, R. Aaij et al., "Measurement of $b$ hadron production fractions in 7 TeV pp collisions," *Phys. Rev. D* **85** (2012) 032008, arXiv:1111.2357 [hep-ex].

[260] **LHCb** Collaboration, R. Aaij et al., "Measurement of $b$ hadron fractions in 13 TeV $pp$ collisions," *Phys. Rev. D* **100** no. 3, (2019) 031102, arXiv:1902.06794 [hep-ex].

[261] **ATLAS** Collaboration, G. Aad et al., "Determination of the ratio of $b$-quark fragmentation fractions $f_s/f_d$ in $pp$ collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector," *Phys. Rev. Lett.* **115** no. 26, (2015) 262001, arXiv:1507.08925 [hep-ex].

[262] **HFLAV** Collaboration, Y. S. Amhis et al., "Averages of $b$-hadron, $c$-hadron, and $\tau$-lepton properties as of 2018," arXiv:1909.12524 [hep-ex].

[263] H.-L. Lai et al., "New parton distributions for collider physics," *Phys. Rev. D* **82** (2010) 074024, arXiv:1007.2241 [hep-ph].

[264] **ALICE** Collaboration, S. Acharya et al., "Dielectron production in proton-proton and proton-lead collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV," *Phys. Rev. C* **102** no. 5, (2020) 055204, arXiv:2005.11995 [nucl-ex].

[265] **ALICE** Collaboration, B. Abelev et al., "Measurement of prompt $J/\psi$ and beauty hadron production cross sections at mid-rapidity in $pp$ collisions at $\sqrt{s} = 7$ TeV," *JHEP* **11** (2012) 065, arXiv:1205.5880 [hep-ex].

[266] **ALICE** Collaboration, B. Abelev et al., "Measurement of electrons from beauty hadron decays in $pp$ collisions at $\sqrt{s} = 7$ TeV," *Phys. Lett. B* **721** (2013) 13–23, arXiv:1208.1902 [hep-ex]. [Erratum: Phys.Lett.B 763, 507–509 (2016)].

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

[268] **ALICE** Collaboration, S. Acharya et al., "Dielectron and heavy-quark production in inelastic and high-multiplicity proton–proton collisions at $\sqrt{s_{NN}} = 13$TeV," *Phys. Lett. B* **788** (2019) 505–518, arXiv:1805.04407 [hep-ex].

[269] **PHENIX** Collaboration, A. Adare et al., "Measurement of Bottom Versus Charm as a Function of Transverse Momentum with Electron-Hadron Correlations in $p+p$ Collisions at $\sqrt{s} = 200$ GeV," *Phys. Rev. Lett.* **103** (2009) 082002, arXiv:0903.4851 [hep-ex].

[270] **CDF** Collaboration, D. Acosta et al., "Measurement of the $J/\psi$ meson and $b-$hadron production cross sections in $p\bar{p}$ collisions at $\sqrt{s} = 1960$ GeV," *Phys. Rev. D* **71** (2005) 032001, arXiv:hep-ex/0412071.

[271] **UA1** Collaboration, C. Albajar et al., "Beauty production at the CERN $p\bar{p}$ collider," *Phys. Lett. B* **256** (1991) 121–128. [Erratum: Phys.Lett.B 262, 497 (1991)].

[272] R. Rapp and H. van Hees, "Heavy Quark Diffusion as a Probe of the Quark-Gluon Plasma," arXiv:0803.0901 [hep-ph].

[273] A. M. Barbano, *Prompt $D_s^+$ meson production in pp, p-Pb and Pb-Pb collisions at LHC with ALICE*. PhD thesis, Università degli Studi di Torino, 2018. https://cds.cern.ch/record/2309930.

[274] **ALICE** Collaboration, B. Abelev et al., "Suppression of high transverse momentum D mesons in central Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV," *JHEP* **09** (2012) 112, arXiv:1203.2160 [nucl-ex].

[275] **ATLAS** Collaboration, M. Aaboud et al., "Prompt and non-prompt $J/\psi$ and $\psi$(2S) suppression at high transverse momentum in 5.02 TeV Pb+Pb collisions with the ATLAS experiment," *Eur. Phys. J. C* **78** no. 9, (2018) 762, arXiv:1805.04077 [nucl-ex].

[276] **CMS** Collaboration, A. M. Sirunyan et al., "Studies of Beauty Suppression via Nonprompt $D^0$ Mesons in Pb-Pb Collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV," *Phys. Rev. Lett.* **123** no. 2, (2019) 022001, arXiv:1810.11102 [hep-ex].

[277] T. Song et al., "Charm production in Pb + Pb collisions at energies available at the CERN Large Hadron Collider," *Phys. Rev. C* **93** no. 3, (2016) 034906, arXiv:1512.00891 [nucl-th].

[278] S. Plumari et al., "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].

[279] M. He, R. J. Fries, and R. Rapp, "Heavy Flavor at the Large Hadron Collider in a Strong Coupling Approach," *Phys. Lett. B* **735** (2014) 445–450, arXiv:1401.3817 [nucl-th].

[280] **ALICE** Collaboration, J. Adam et al., "Transverse momentum dependence of D-meson production in Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 2.76$ TeV," *JHEP* **03** (2016) 081, arXiv:1509.06888 [nucl-ex].

[281] V. Minissale, S. Plumari, and V. Greco, "Charm Hadrons in pp collisions at LHC energy within a Coalescence plus Fragmentation approach," arXiv:2012.12001 [hep-ph].

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

[283] S. A. Voloshin, A. M. Poskanzer, and R. Snellings, "Collective phenomena in non-central nuclear collisions," *Landolt-Bornstein* **23** (2010) 293–333, arXiv:0809.2949 [nucl-ex].

[284] M. Luzum and J.-Y. Ollitrault, "Eliminating experimental bias in anisotropic-flow measurements of high-energy nuclear collisions," *Phys. Rev. C* **87** no. 4, (2013) 044907, arXiv:1209.2323 [nucl-ex].

[285] I. Selyuzhenkov and S. Voloshin, "Effects of non-uniform acceptance in anisotropic flow measurement," *Phys. Rev. C* **77** (2008) 034904, arXiv:0707.4672 [nucl-th].

[286] N. Borghini and J. Y. Ollitrault, "Azimuthally sensitive correlations in nucleus-nucleus collisions," *Phys. Rev. C* **70** (2004) 064905, arXiv:nucl-th/0407041.

[287] **CMS** Collaboration, V. Khachatryan et al., "Suppression and azimuthal anisotropy of prompt and nonprompt J/$\psi$ production in PbPb collisions at $\sqrt{s_{\rm NN}} = 2.76$ TeV," *Eur. Phys. J. C* **77** no. 4, (2017) 252, arXiv:1610.00613 [nucl-ex].

[288] **ATLAS** Collaboration, M. Aaboud et al., "Prompt and non-prompt $J/\psi$ elliptic flow in Pb+Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ Tev with the ATLAS detector," *Eur. Phys. J. C* **78** no. 9, (2018) 784, arXiv:1807.05198 [nucl-ex].

[289] J. Uphoff, O. Fochler, Z. Xu, and C. Greiner, "Open Heavy Flavor in Pb+Pb Collisions at $\sqrt{s} = 2.76$ TeV within a Transport Model," *Phys. Lett. B* **717** (2012) 430–435, arXiv:1205.4945 [hep-ph].

[290] J. Aichelin, P. B. Gossiaux, and T. Gousset, "Radiative and Collisional Energy Loss of Heavy Quarks in Deconfined Matter," *Acta Phys. Polon. B* **43** (2012) 655–662, arXiv:1201.4192 [nucl-th].

[291] V. Greco, H. van Hees, and R. Rapp, "Heavy-quark kinetics at RHIC and LHC," in *23rd International Nuclear Physics Conference (INPC 2007)*. 9, 2007. arXiv:0709.4452 [hep-ph].

This Ph.D. thesis has been typeset by means of the TEX-system facilities. The typesetting engine was pdfLATEX. The document class was `toptesi`, by Claudio Beccari, with option `tipotesi=scudo`. This class is available in every up-to-date and complete TEX-system installation.