



**UNIVERSITÀ  
DI TORINO**

**Università degli Studi di Torino**  
Dipartimento di Fisica

Corso di Laurea Magistrale in Fisica Nucleare e Subnucleare

**Ratio of strange to non-strange D-meson  
production in proton-proton collisions at  
 $\sqrt{s} = 13.6 \text{ TeV}$**

**Supervisor:**

Francesco Prino

**Candidate:**

Fabrizio Chinu

**Co-supervisors:**

Fabrizio Grosa

Stefano Politanò

**Referee:**

Roberto Covarelli

Anno Accademico 2023/2024



## Abstract

Quantum Chromo-Dynamics (QCD) is the theory describing the strong interactions between quarks and gluons, the fundamental constituents of hadrons. Lattice QCD calculations predict a transition from colour-neutral hadrons to a colour-deconfined state called Quark-Gluon Plasma (QGP) under extreme temperature and energy density conditions, which can be reached in the laboratory by colliding high-energy heavy ions.

Heavy-flavour quarks (charm and beauty) are produced in the earliest stages of the collision and interact with the formed medium, losing energy through interactions with its constituents, making heavy-flavour hadrons excellent probes of the properties of the QGP. The hadronisation process is the transition from colour-charged partons (quarks and gluons) produced in a collision into colour-neutral hadrons. It is typically parametrised using fragmentation functions, assuming the universality of the process (i.e., independence from the collision system and energy) and the independence of the quark hadronisation via fragmentation. This approach fails in describing the measured charm-baryon production measured in proton-proton (pp) and p–Pb collisions at the Large Hadron Collider (LHC), suggesting that the hadronisation process is not universal. The hadronisation mechanism is expected to be modified by the presence of the QGP, as a novel process, called recombination, is expected to occur. In this process, the produced heavy-flavour quarks combine with other quarks from the medium to form hadrons. The observed baryon enhancement in pp collisions is qualitatively described by models that include QGP droplets formation, as well as by event generators such as PYTHIA 8 in which colour reconnections are considered in modelling the parton shower and the fragmentation in the parton-rich environment created in pp collisions at the LHC.

In the presence of QGP, the production of strange quarks is expected to be enhanced due to the increase in their thermal production owing to the high temperatures reached in the medium. An increased production of strange hadrons relative to pions is observed in Pb–Pb collisions with respect to pp collisions, where the production of QGP is not expected. However, a smooth increase of strange-hadron relative abundances with the number of charged-particles produced in the collision has been observed in pp collisions, raising the question of whether small droplets of QGP could also be produced in high-multiplicity pp collisions.

Therefore, because of the recombination mechanism and the enhanced production of strange quarks, an increase in the strange over non-strange  $D_s^+/D^+$  production yield ratio is expected in the presence of QGP.

The ALICE detector installed at the LHC is designed to address the physics of strongly-interacting matter and QGP produced in ultra-relativistic heavy-ion collisions. During the LHC Long Shutdown 2 (2019–2021) the detector was upgraded by enhancing the tracking performance and increasing the readout rate to collect larger data samples, improving its capabilities to probe the QGP with heavy-flavours.

This Thesis is devoted to the precise measurement of the transverse-momentum- ( $p_T$ -)differential  $D_s^+/D^+$  production-yield ratio at midrapidity ( $|y| < 0.5$ ) in pp collisions at  $\sqrt{s} = 13.6$  TeV with the data collected by the ALICE experiment during

the ongoing LHC Run 3 data-taking period. This ratio of strange over non-strange charm meson yields allows for a direct access to information on charm-quark hadronisation mechanisms. Due to their small lifetime ( $\tau \sim 100 - 300 \mu\text{m}/c$ ),  $D_s^+$  and  $D^+$  mesons cannot be directly detected and are reconstructed through their decay products. They are reconstructed through the same hadronic decay channel

$$D_s^+, D^+ \rightarrow \phi\pi^+ \rightarrow K^+K^-\pi^+ ,$$

allowing for the cancellation of some of the systematic uncertainties related to the measurement. Multiclass Machine Learning (ML) algorithms have been employed to suppress the large combinatorial background arising from the combination of three independent tracks produced in the pp collision and increase the statistical significance of the measurement. Additionally, the ML-based selections were used to increase the relative contribution of prompt  $D_s^+$  and  $D^+$  mesons (i.e., those directly produced in the hadronisation of a charm quark or through the strong decay of a directly produced excited charm hadron or charmonium state) in the selected sample. The signal is extracted in 14  $p_T$  intervals within the  $0.5 < p_T < 24 \text{ GeV}/c$  range by fitting the invariant mass distribution of candidates passing the ML selections.

The extracted signal is then corrected for the geometrical acceptance of the ALICE detector, the selection efficiency, and the residual non-prompt contamination arising from  $D$  mesons produced in the decay of a beauty hadron and surviving the ML selections.

The measured  $D_s^+/D^+$  production-yield ratio is compared to results obtained by the ALICE Collaboration with the data collected during the LHC Run 2 data-taking period at the different centre-of-mass energies of  $\sqrt{s} = 5.02, 7$ , and  $13 \text{ TeV}$  and with measurements performed by the LHCb Collaboration at the LHC in the forward-rapidity range  $2.0 < y < 4.5$  in pp collisions at  $\sqrt{s} = 13 \text{ TeV}$ . The results are compatible with those obtained in Run 2 by both ALICE and LHCb Collaborations, indicating no significant dependence of the  $D_s^+/D^+$  ratio on the centre-of-mass energy and rapidity. Thanks to the larger data samples collected during the LHC Run 3 data-taking period, a more precise and granular measurement of the  $D_s^+/D^+$  production-yield ratio than that measured in Run 2 is achieved. Furthermore, the  $p_T$  reach of the measurement has been extended to lower values, reaching as low as  $0.5 \text{ GeV}/c$ . These measurements provide state-of-the-art results on the production of charm-strange mesons in pp collisions.

Lastly, to study the performance of the ALICE experiment in heavy-ion collisions, the reconstruction of  $D_s^+$  and  $D^+$  mesons was performed in Pb–Pb collisions at  $\sqrt{s_{\text{NN}}} = 5.36 \text{ TeV}$  for different centrality intervals, which represent the degree of overlap of the two colliding nuclei, using rectangular selection criteria. The results in the 10–30% centrality class have been used as benchmarks of the performance of the upgraded ALICE experiment in heavy-ion collisions. These results will provide a solid baseline for the study of the  $D_s^+/D^+$  production-yield ratio in Pb–Pb collisions, to be performed in the future. By comparing the results obtained in Pb–Pb collisions and the measurements in pp collisions described in this Thesis, insights into the hadronisation mechanisms of charm quarks in the presence of the QGP, where strange quarks are more abundant, will be obtained.

# Contents

<b>1 Quarks, gluons and the Quark-Gluon Plasma</b>	<b>1</b>
1.1 Quantum Chromodynamics . . . . .	1
1.1.1 Running coupling constant . . . . .	4
1.2 Confinement . . . . .	6
1.2.1 MIT bag model . . . . .	7
1.2.2 Lattice QCD . . . . .	8
1.3 Quark–Gluon Plasma . . . . .	10
1.3.1 High-energy heavy-ion collisions . . . . .	11
1.3.2 Hadron species abundances . . . . .	15
1.3.3 Radial flow . . . . .	17
1.3.4 High- $p_T$ hadrons and jet quenching . . . . .	20
1.3.5 Strangeness enhancement . . . . .	22
<b>2 Open heavy-flavour production in proton-proton collisions</b>	<b>25</b>
2.1 Factorisation theorem . . . . .	25
2.1.1 Parton Distribution Functions . . . . .	26
2.1.2 Partonic cross-section . . . . .	29
2.1.3 Fragmentation Functions . . . . .	30
2.2 Hadronisation: microscopic and macroscopic descriptions . . . . .	31
2.2.1 Independent fragmentation . . . . .	31
2.2.2 String model . . . . .	32
2.2.3 Cluster model . . . . .	35
2.2.4 Coalescence model . . . . .	36
2.2.5 Core-corona model . . . . .	38
2.2.6 Statistical hadronisation model . . . . .	38
2.2.7 Baryon enhancement: $\Lambda_c^+/\bar{D}^0$ ratio . . . . .	39
<b>3 A Large Ion Collider Experiment</b>	<b>41</b>
3.1 The Large Hadron Collider . . . . .	41
3.2 The ALICE experiment . . . . .	43
3.2.1 ALICE upgrades overview . . . . .	44
3.2.2 Inner Tracking System . . . . .	46
3.2.3 Time Projection Chamber . . . . .	47
3.2.4 Time-of-Flight . . . . .	49
3.2.5 Fast Interaction Trigger . . . . .	50
<b>4 <math>D_s^+</math> and <math>D^+</math> reconstruction strategy in proton-proton collisions</b>	<b>52</b>

4.1	Data sample and event selection . . . . .	54
4.2	$D_s^+$ and $D^+$ reconstruction workflow . . . . .	54
4.3	Vertex reconstruction and D-meson selection . . . . .	56
4.3.1	Track-level selections . . . . .	57
4.3.2	Topological selections . . . . .	58
4.3.3	Kinematic selections . . . . .	62
4.3.4	Particle identification selections . . . . .	63
<b>5</b>	<b><math>D_s^+</math> and <math>D^+</math> signal extraction</b>	<b>66</b>
5.1	Machine Learning . . . . .	66
5.1.1	Supervised learning . . . . .	66
5.2	$D_s^+$ and $D^+$ selection using Machine Learning . . . . .	68
5.2.1	Data preparation . . . . .	69
5.2.2	Boosted Decision Trees . . . . .	74
5.2.3	Tuning the model's hyperparameters . . . . .	76
5.2.4	Evaluation of the model's performance . . . . .	77
5.2.5	Interpretation of the model's output: Feature importance . . . . .	80
5.2.6	Optimisation of the selection thresholds on the BDT scores . . . . .	83
5.3	$D_s^+$ and $D^+$ raw-yield extraction . . . . .	85
<b>6</b>	<b>Evaluation of the corrected <math>D_s^+/D^+</math> yield ratio</b>	<b>97</b>
6.1	Acceptance and efficiency correction . . . . .	97
6.2	Prompt fraction correction . . . . .	100
6.3	Branching ratio correction . . . . .	104
6.4	Systematic uncertainties . . . . .	105
6.4.1	BDT selection efficiency . . . . .	105
6.4.2	Imperfections in the description of the tracking performance . . . . .	107
6.4.3	Prompt fraction . . . . .	108
6.4.4	Generated Monte Carlo $p_T$ shape . . . . .	109
6.4.5	Summary of systematic uncertainties . . . . .	111
<b>7</b>	<b><math>D_s^+/D^+</math> production-yield ratio</b>	<b>112</b>
7.1	Dependence on centre-of-mass energy . . . . .	113
7.2	Dependence on rapidity . . . . .	116
7.3	Comparison to models . . . . .	119
<b>8</b>	<b><math>D_s^+</math> and <math>D^+</math> mesons reconstruction in Pb–Pb collisions</b>	<b>121</b>
8.1	Event selection and data sample . . . . .	122
8.2	$D_s^+$ and $D^+$ mesons reconstruction . . . . .	123
8.3	Signal extraction . . . . .	125
<b>9</b>	<b>Summary, conclusions, and perspectives</b>	<b>130</b>
<b>A</b>	<b>Figures for the systematic uncertainty on the raw yield extraction</b>	<b>135</b>
<b>B</b>	<b>Figures for the data-driven <math>D_s^+</math> and <math>D^+</math> prompt fraction evaluation</b>	<b>140</b>

---

List of Contents

C Figures for the systematic uncertainty on the BDT selection efficiency	149
Bibliography	154
Ringraziamenti	181

# Chapter 1

## Quarks, gluons and the Quark-Gluon Plasma

### 1.1 Quantum Chromodynamics

In the mid-20<sup>th</sup> century, the realm of particle physics underwent a transformative phase, marked by the discovery of a seemingly endless variety of subatomic particles. This era witnessed the unveiling of numerous hadron species, which left physicists with the necessity of developing a framework that could describe the behaviour of these particles and their interactions. This led to the development of the static quark model, which emerged in the 1960s as a groundbreaking conceptual framework to categorise the various observed particles. Developed independently by Murray Gell-Mann [1] and George Zweig [2, 3], this model postulated the existence of fundamental constituents of hadrons called quarks, which, in order to reflect the experimental findings, had to be fermions (to describe baryons with spin 1/2 and 3/2) with fractional electric charge. The quark model beautifully explained the organisation of the known hadrons in terms of three quarks ( $u$ ,  $d$ , and  $s$ ), leading to the development of a more structured and coherent classification of particles.

Despite the phenomenological success of the static quark model, it had two problems: it introduced particles with fractional charge, which had never been observed before, and, most importantly, it gave rise to a violation of the Fermi-Dirac statistics. The  $\Delta^{++}$ ,  $\Delta^-$ , and  $\Omega^-$  baryons, in fact, have symmetric orbital, spin, and flavour wavefunctions, which defied the Pauli exclusion principle that should have implied antisymmetric wavefunctions for these particles. To resolve these inconsistencies, a new degree of freedom, the *colour*, was introduced. Hadron wavefunctions were assumed to be totally antisymmetric in colour quantum numbers, effectively implementing the Pauli exclusion principle.

The simplest model of colour would be to assign quarks to the fundamental representation of a global  $SU(3)$  symmetry. Each quark carries a colour index:  $q_i$ , where  $i = 1, 2, 3$ , and transforms under the fundamental (3) representation of  $SU(3)$ , while antiquarks,  $\bar{q}_i$ , transform in the  $\bar{3}$  representation. Introducing the totally antisymmetric tensor  $\varepsilon^{ijk}$ , possible compositions of quarks that give rise to colour singlets are

$$\bar{q}^i q_i, \quad \varepsilon^{ijk} q_i q_j q_k, \quad \varepsilon^{ijk} \bar{q}_i \bar{q}_j \bar{q}_k,$$

which are the quark compositions of mesons, baryons, and antibaryons, respectively.

One of the tests supporting the existence of colour and fractional electric charge came in the form of the ratio  $R$ , of the total cross-section of hadron production in  $e^+e^-$  annihilations to the cross-section of a pair of muons produced from the same annihilation process. The virtual photon emitted in the annihilation can produce all electrically charged pairs of particles and antiparticles, as shown in Fig. 1.1.

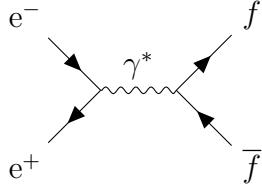


Figure 1.1:  $e^+e^-$  annihilation to a fermion-antifermion pair.

The ratio  $R$  is given by

$$R = \frac{\sigma(e^+e^- \rightarrow \text{hadrons})}{\sigma(e^+e^- \rightarrow \mu^+\mu^-)} = N_c \sum_f Q_f^2 ,$$

where  $N_c$  represents the number of existing colours and  $Q_f$  is the electric charge of the quark flavour  $f$ . Notably, this ratio depends on centre-of-mass energy ( $\sqrt{s}$ ) and encompasses all possible quark flavours that can be produced by the virtual photon at that specific energy level. While at low energies (few GeV) only light quarks ( $u, d, s$ ) are produced, at higher energies, heavier quarks like charm ( $c$ ,  $M_c = 1.2730 \pm 0.0046$  GeV/c $^2$  [4]) and beauty ( $b$ ,  $M_b = 4.183 \pm 0.007$  GeV/c $^2$  [4]) can be produced as well. The experimental measurements of  $R$ , shown in Fig. 1.2, exhibit a remarkable agreement with the predictions of the three-colour model, thereby providing compelling evidence for the existence of colour and fractional electric charge of quarks.

The final step that propelled the development of Quantum Chromodynamics (QCD) as a comprehensive theory of the strong force was the insight into the mechanism that ensured all hadron wavefunctions to be colour singlets. This emerged from the discovery of asymptotic freedom, a phenomenon observed in deep-inelastic scattering experiments. Non-Abelian gauge theories, often referred to as Yang-Mills theories, were identified as having this unique characteristic. This realisation led to the formulation of QCD by elevating the global colour  $SU(3)$  symmetry to a local one, allowing the 8 quanta of the  $SU(3)$  gauge field, called *gluons*, to mediate the strong force, successfully describing the confinement and behavior of quarks and gluons within hadrons.

The QCD Lagrangian density can be written as

$$\mathcal{L}_{QCD} = -\frac{1}{4} F_{\mu\nu}^a F_a^{\mu\nu} + \sum_f \bar{q}_f^i (i\gamma^\mu (\mathcal{D}_\mu)_{ij} - m_f \delta_{ij}) q_f^j , \quad (1.1)$$

where  $F_{\mu\nu}^a$  is the field strength tensor defined in terms of the gluon field  $A_\mu^a$  and the  $SU(3)$  structure constant  $f^{abc}$  as

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g_s f^{abc} A_\mu^b A_\nu^c \quad (1.2)$$

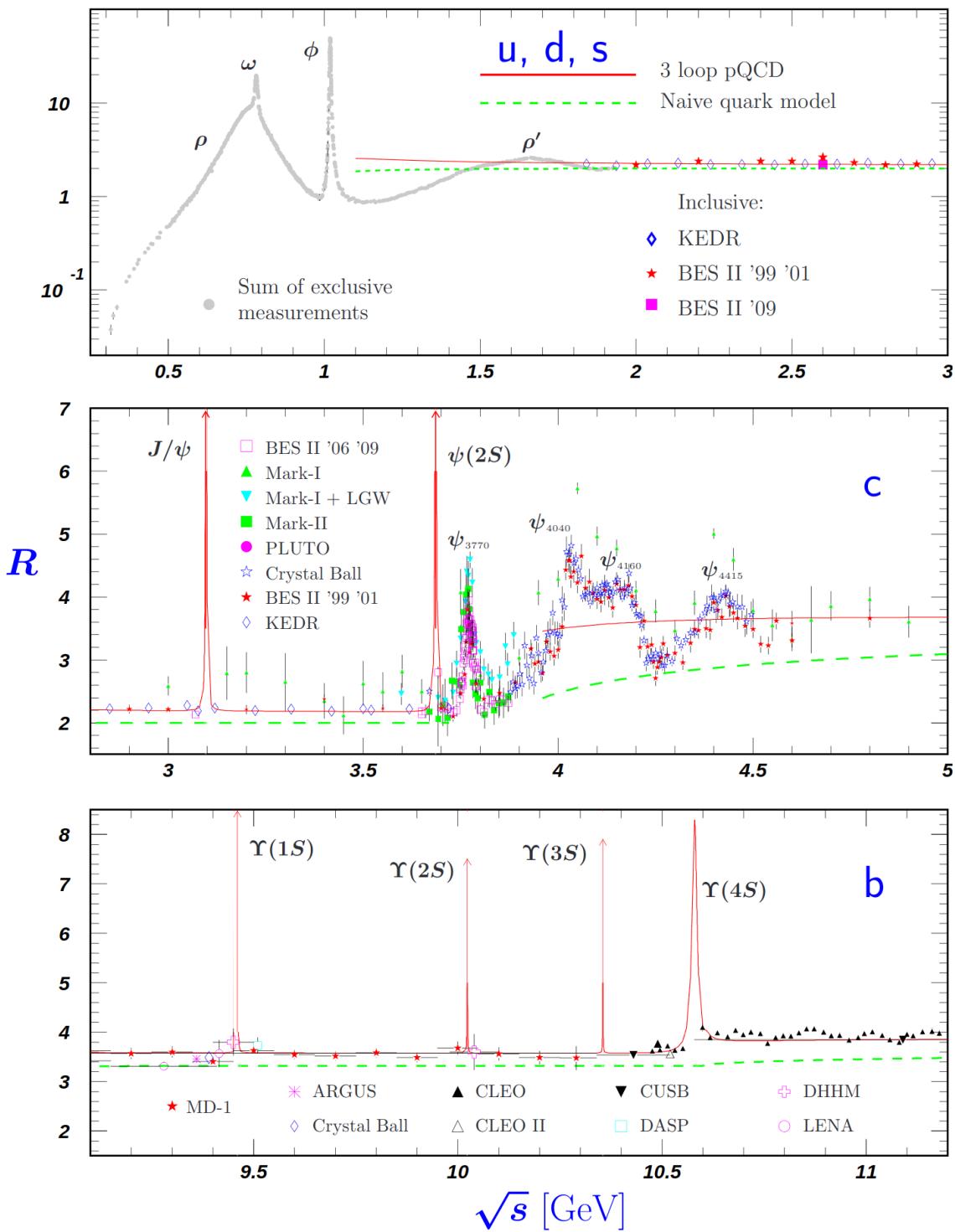


Figure 1.2:  $R$  as a function of  $\sqrt{s}$  in the light-flavour, charm, and beauty threshold regions taken from Ref. [4]. The green curve is a naive quark-parton model prediction, while the red one is a 3-loops perturbative QCD prediction. Breit-Wigner parameterisations of  $J/\psi$ ,  $\psi(2S)$ , and  $\Upsilon(nS)$ ,  $n = 1, 2, 3, 4$  are also shown.

and  $(\mathcal{D}_\mu)_{ij}$  is the covariant derivative:

$$(\mathcal{D}_\mu)_{ij} = \partial_\mu \delta_{ij} - ig_s(t^a)_{ij} A_\mu^a ,$$

with  $t^a$  being one of the generators of the  $SU(3)$  representation.

The last term in Eq. 1.2 is peculiar to non-Abelian theories, and gives rise to triplet and quartic gluon self-interactions illustrated in Fig. 1.3.  $g_s$  is a coupling parameter related to the strong coupling constant  $\alpha_s$ , which determines the strength of the interaction between the coloured particles.

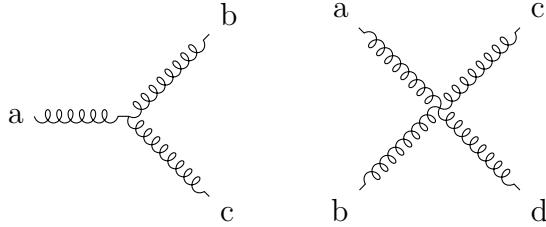


Figure 1.3: Feynman diagrams for gluon self-interactions.

The second term of Eq. 1.1 describes the interactions between quarks and gluons, sketched in Fig. 1.4, and contains the mass term for the fermions. It is noteworthy to observe that the interaction between quarks and gluons is diagonal in flavour, meaning that the strong interaction conserves the flavour of quarks. In contrast, colour mixing is allowed within the framework of QCD.

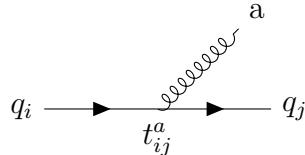


Figure 1.4: Feynman diagram for quark-gluon interaction.

### 1.1.1 Running coupling constant

If one considers a dimensionless physical observable, denoted in the following as  $R$ , which solely depends on a single energy scale,  $Q$ , one might naturally expect  $R$  to maintain a constant value, independent of the specific energy scale chosen. However, this does not hold true when loop diagrams are studied: the necessity of renormalisation introduces a new energy scale denoted as  $\mu$ . This scale, known as the renormalisation scale, is the point at which the subtraction of the ultraviolet divergences is carried out. Critically,  $\mu$  is an arbitrary parameter and, as such, is non-physical. Consequently,  $R$  becomes dependent on the ratio  $Q^2/\mu^2$  and the renormalised coupling  $\alpha_s = g_s^2/4\pi$ :  $R \equiv R\left(\frac{Q^2}{\mu^2}, \alpha_s\right)$ . The  $\mu$  independence of  $R$  (which is an essential requirement given  $\mu$ 's arbitrariness) can be expressed as

$$\mu^2 \frac{dR\left(\frac{Q^2}{\mu^2}, \alpha_s\right)}{d\mu^2} = \mu^2 \left[ \frac{\partial}{\partial \mu^2} + \frac{\partial \alpha_s}{\partial \mu^2} \frac{\partial}{\partial \alpha_s} \right] R\left(\frac{Q^2}{\mu^2}, \alpha_s\right) = 0 , \quad (1.3)$$

a fundamental equation known as the renormalisation group equation. This equation is exactly true in the case of a prediction that considers all perturbative orders. If one limits the expansion at a fixed order  $\alpha_s^N$ , then a dependence of  $R$  from  $\mu$  is observed at the  $\alpha_s^{N+1}$  order. By introducing

$$t \equiv \log(Q^2/\mu^2), \quad \beta(\alpha_s) \equiv \mu^2 \frac{d\alpha_s}{d\mu^2} ,$$

Eq. 1.3 can be written as

$$\left( -\frac{\partial}{\partial t} + \beta(\alpha_s) \frac{\partial}{\partial \alpha_s} \right) R(e^t, \alpha_s) = 0 .$$

This first-order partial differential equation can be solved by defining a new function: the running coupling  $\alpha_s(Q^2)$ , which evolves as a function of  $Q$ .

$$t = \log(Q^2/\mu^2) \equiv \int_{\alpha_s}^{\alpha_s(Q^2)} \frac{dx}{\beta(x)}, \quad \text{with } \alpha_s = \alpha_s(\mu^2) . \quad (1.4)$$

By differentiating Eq. 1.4 with respect to  $t$  and  $\alpha_s$ , one gets:

$$\beta(\alpha_s(Q^2)) = \frac{\partial \alpha_s(Q^2)}{\partial t}, \quad \frac{d\alpha_s(Q^2)}{d\alpha_s} = \frac{\beta(\alpha_s(Q^2))}{\beta(\alpha_s)} . \quad (1.5)$$

It results from this last set of equations that  $R(1, \alpha_s(Q^2))$  satisfies Eq. 1.3; hence, the running coupling constant has absorbed the  $\mu$  scale dependence of  $R$ . As a consequence, the knowledge of  $R(1, \alpha_s)$ , which can be evaluated in fixed-order perturbation theory, allows knowing the dependence of  $R$  from  $Q^2$ , the physical scale at which the coupling is gauged, by simply substituting  $\alpha_s \rightarrow \alpha_s(Q^2)$ .

### The $\beta$ function

The running of the coupling constant is determined by the  $\beta(\alpha_s)$  function, which is evaluated from loop corrections to the bare vertices of QCD. As of the time of the writing of this Thesis, the  $\beta$  function has been evaluated up to 5 loops [5]. In Fig. 1.5, the 1-loop Feynman diagrams contributing to the  $\beta$  function evaluation are reported.

By limiting the calculations at the first order in the perturbative expansion, one gets:

$$\beta(\alpha_s) = -\alpha_s^2 \frac{11N_c - 2N_f}{12\pi} + \mathcal{O}(\alpha_s^3) \equiv -\alpha_s^2 \beta_0 + \mathcal{O}(\alpha_s^3) , \quad (1.6)$$

where  $N_c$  is the number of colours (3), while  $N_f$  is the number of quark flavours that can be considered massless at the physical scale  $Q^2$  at which the coupling is being measured. From Eqs. 1.5 and 1.6, one can extract the  $Q^2$  dependency of the running coupling constant:

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu^2)}{1 + \alpha_s(\mu^2) \beta_0 \log(Q^2/\mu^2)} , \quad (1.7)$$

Notably, since  $\beta_0$  is positive in a six-quark-flavour framework, the strong coupling constant exhibits a monotonic decreasing trend as a function of  $Q^2$ . This behaviour

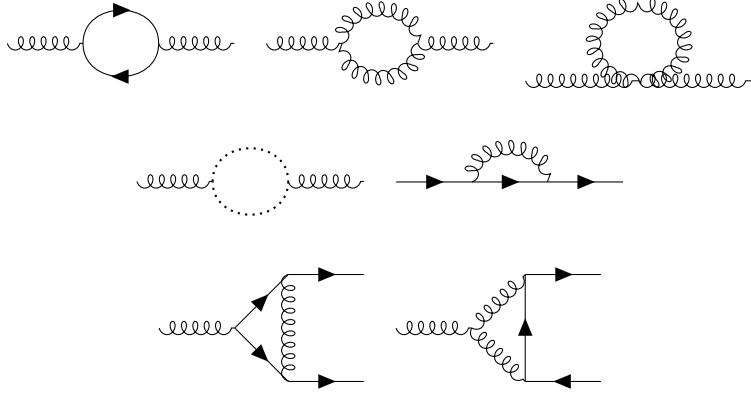


Figure 1.5: 1-loop Feynman diagrams contributing to the  $\beta$  function evaluation.

differs from the one of the electromagnetic coupling constant, which increases with the energy scale due to the screening effect of vacuum polarisation. For QCD, the running of the coupling constant is a direct consequence of the non-Abelian nature of the theory, allowing for gluon self-interactions, which give rise to an anti-screening effect. The idea is that the emission of virtual gluons by static colour sources causes their colour charges to 'leak out' into the surrounding vacuum. Since the interaction between distributions of charges is weaker than the one between point-like charges when the distributions overlap, the effective coupling constant decreases at short distances. This behaviour is known as asymptotic freedom, a key feature of QCD that allows for the perturbative expansion of the theory at high energy scales, where the strong coupling constant is small. At the same time, the running of the coupling constant implies that the theory is non-perturbative at low energy scales, and phenomenological models are required to describe the strong interaction in this regime.

Instead of using the renormalisation scale  $\mu$  as a free parameter, one can use the running coupling constant to define a physical scale,  $\Lambda_{\text{QCD}}$ , which is the energy scale at which the coupling constant would diverge, if extrapolated outside the perturbative regime. Using Eq. 1.7, one can write

$$\alpha_s(\Lambda_{\text{QCD}}) = \frac{1}{\beta_0 \log(Q^2/\Lambda_{\text{QCD}}^2)} .$$

The value of  $\Lambda_{\text{QCD}}$  is determined by the specific definition being used. However, to obtain the value of the coupling constant measured at  $Q^2 = M_Z^2$ , an approximate value of  $\Lambda_{\text{QCD}}$  of around 200 MeV can be used.

Measurements of the running of the coupling constant at different values of  $Q$  are illustrated in Fig. 1.6 and compared to the theoretical prediction at 5 loops. The agreement between the experimental data and the theoretical prediction is remarkable, confirming the validity of the QCD framework at high energy scales.

## 1.2 Confinement

The concept of confinement is one of the most intriguing aspects of QCD. It is the phenomenon by which quarks and gluons are never observed as free particles,

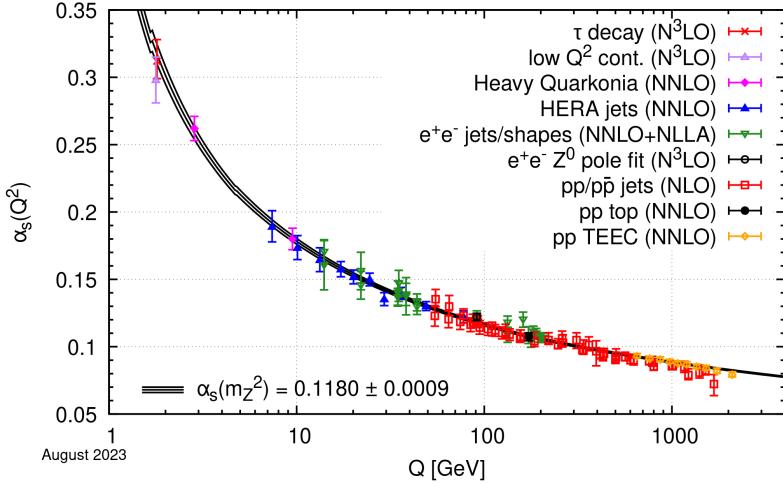


Figure 1.6: Summary of measurements of  $\alpha_s$  as a function of the energy scale  $Q$ , compared to the running of the coupling computed at five loops, taking as an input the current PDG average,  $\alpha_s(M_Z^2) = 0.1180 \pm 0.0009$   $\text{GeV}/c^2$ . Figure taken from Ref. [4].

but are always confined within colour-neutral hadrons. The confinement of quarks and gluons is a direct consequence of the non-Abelian nature of the theory, which, as described in the previous Section, is characterised by an increase in the strong coupling constant at low energy scales. The confinement of quarks and gluons is a non-perturbative effect, and despite extensive research, a comprehensive theoretical description of confinement in QCD remains elusive. Phenomenological models like the MIT bag model have been proposed, but a full comprehension of confinement is still lacking. Lattice QCD simulations are the most successful approach to study the non-perturbative regime of the theory, and they have provided a wealth of information on the properties of hadrons and the strong interaction at low energy scales.

### 1.2.1 MIT bag model

The MIT bag model [6] is a phenomenological model of confinement, which describes hadrons as bound states of quarks and gluons confined within a finite volume, called *bag*. The model was developed in the 1970s by A. Chodos, R. L. Jaffe, K. Johnson, C. B. Thorn, and V. F. Weisskopf, and it has been widely used to study the properties of hadrons and strong interactions. In the MIT bag model,  $N$  non-interacting massless fermions are confined within a spherical cavity of radius  $R$ , which is the bag radius. The confinement arises from a balance between the pressure due to the kinetic energy of the fermions inside the bag and an ad hoc external pressure, which is introduced to confine the fermions within the bag. The fermions are described by the Dirac equation for massless fermions:

$$i\gamma^\mu \partial_\mu \psi = 0 \quad ,$$

where  $\psi$  is the fermion field, and  $\gamma^\mu$  are the Dirac matrices. The solution to the Dirac equation is given in terms of the spherical Bessel functions of zeroth and first order,  $j_0(p_0 r)$  and  $j_1(p_0 r)$ :

$$\psi = \mathcal{N} e^{-ip_0 t} \begin{pmatrix} j_0(p_0 r) \chi^+ \\ \vec{\sigma} \cdot \hat{r} j_1(p_0 r) \chi^- \end{pmatrix} ,$$

where  $p_0$  is the energy of the fermion,  $\chi^+$  and  $\chi^-$  are the two components of the fermion four-dimensional spinor  $\psi$ , and  $\vec{\sigma}$  are the Pauli matrices. The colour flux at a point  $r$  inside the bag is given by

$$j_{ab}^\mu(r) = \bar{\psi}_a(r) \gamma^\mu \psi_b(r) ,$$

where  $a$  and  $b$  are the colour indices of the fermions. If the quantum numbers are not to be lost through the surface of the bag, which is the definition of confinement, then, on the surface:

$$n_\mu j_{ab}^\mu(r) = \bar{\psi}_a(r) \gamma \cdot n \psi_b(r) = 0 ,$$

where  $n_\mu$  is a unit space-like vector perpendicular to the surface. Using the gamma properties,  $(i\gamma \cdot n)^2 = 1$ , so that by assuming that  $i\gamma \cdot n = +1$ , the boundary condition on the surface of the bag is given by

$$\bar{\psi}(R)\psi(R) = 0 ,$$

leading to the solution of the Dirac equation in the bag:

$$[j_0(p_0 R)]^2 - [j_1(p_0 R)]^2 = 0 ,$$

with solution  $p_0 R = 2.04$ . The total energy inside the bag is given by

$$E = \frac{2.04 N}{R} (\hbar c) + \frac{4\pi}{3} R^3 B ,$$

where the first term is the kinetic energy of the fermions, and the second term is the energy due to the presence of an external pressure  $B$  which keeps the fermions confined in the bag. The bag pressure is a phenomenological parameter of the model, and it is introduced to confine the fermions within the bag. It can be extracted by minimising the energy of the system with respect to the bag radius  $R$ , which yields  $B = 234$  MeV/fm<sup>3</sup> for a baryon with  $R = 0.8$  fm.

### 1.2.2 Lattice QCD

Lattice QCD is a numerical technique used to study the non-perturbative regime of QCD. The method is based on the discretisation of space-time on a four-dimensional lattice, and the evaluation of the path integral of the theory using Monte Carlo methods, i.e. by sampling possible configurations of the quark and gluon fields according to the probability distribution given by the QCD Lagrangian. The lattice spacing is a parameter of the method, and allows one to avoid the ultraviolet divergences of the theory, which are typical in perturbative QCD, by introducing a cutoff on the momenta of the quark and gluon fields.

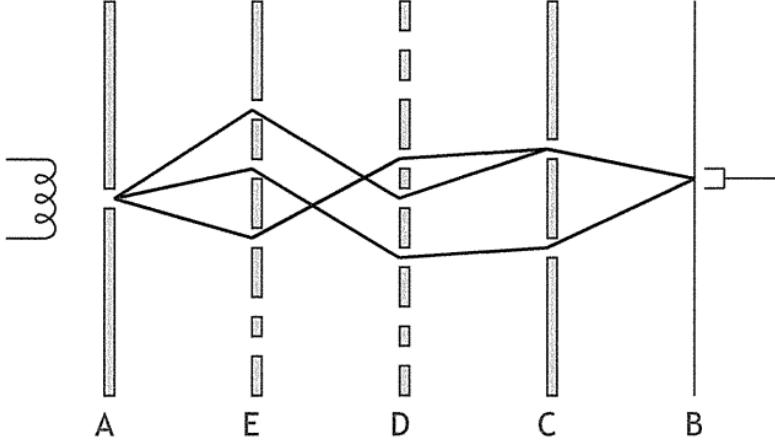


Figure 1.7: Feynman introduction to path integrals. Here, a particle emitted from a source at  $x_a$  is detected at  $x_b$ . A finite number of screens, each with a finite number of holes, is placed between the source and the detector. The probability amplitude for the particle to hit the detector is given by the sum of the probabilities of moving from the source to the detector through all possible paths. By adding an infinite amount of screens with an infinite number of holes, and by also considering the time at which the particle passes through each screen, the sum becomes an integral over all possible paths, called a *path integral*.

The Lattice QCD simulations are based on the path integral formalism of quantum field theory [7], developed by R. Feynman in the 1940s. The path integral provides a natural extension to quantum mechanics of the least action principle of classical mechanics, and it allows one to calculate the probability amplitude of a particle to move from one point to another in space-time, considering the evolution of the system over all possible paths. The transition amplitude from the state  $(x_a, t_a)$  to the state  $(x_b, t_b)$  is given by

$$A[(x_a, t_a) \rightarrow (x_b, t_b)] = \langle x_b, t_b | e^{-iH(t_b-t_a)} | x_a, t_a \rangle = \sum_{\text{paths}} e^{iS[x(t)]} , \quad (1.8)$$

where  $H$  is the Hamiltonian of the system,  $S[x(t)]$  is the action of the system for a given path  $x(t)$ , and the sum runs over all possible paths from  $(x_a, t_a)$  to  $(x_b, t_b)$ . By taking the continuum limit on space-time, one obtains an integration over all the possible space-time paths of the system:

$$\sum_{\text{paths}} e^{iS[x(t)]} \rightarrow \int_{x_a}^{x_b} [\mathcal{D}x(t)] e^{iS[x(t)]} , \quad (1.9)$$

where the right-hand side term is a functional integral over all possible paths. It is interesting to note that by combining Eqs. 1.8 and 1.9, a quantity resembling the partition function of a statistical system is obtained:

$$\mathcal{Z} = \sum_{x_a} \langle x_a, t_a | e^{\beta H} | x_a, t_a \rangle .$$

It is possible to express the partition function in terms of a path integral by applying a Wick rotation to the time variable,  $t \rightarrow -i\tau$ , with  $\tau_a = 0 \leq \tau \leq \tau_b = \beta$  and considering the Euclidean action in place of the Minkowskian one,  $S_E = iS$ . Furthermore, since the state at  $\tau_a$  is the same as the one at  $\tau_b$  in the partition function definition, a periodic boundary condition is imposed:  $x(\tau_a) = x(\tau_b)$ . With these considerations, the partition function can be expressed as

$$\mathcal{Z} = \int [\mathcal{D}x(\tau)] e^{-S_E[x(\tau)]} .$$

This formalism, which was here developed for a single particle, can be extended to a quantum field theory, and in particular to QCD.

Lattice QCD simulations are computationally intensive, and they require large supercomputers to perform the calculations. To limit the computational costs, calculations are often performed at larger up and down quark masses than in nature, drastically reducing the number of virtual quark-antiquark loops that have to be taken into account. Because of the employed Monte Carlo approach, only a finite number of configurations can be considered, leading to statistical uncertainties in the lattice QCD results. In order to obtain physical results, several limits have to be taken: i) the continuum limit, i.e. the extrapolation of the lattice spacing to zero, ii) the infinite-volume limit, i.e. the extrapolation of the lattice size to infinity, and iii) the physical quark-mass limit, i.e. the extrapolation to physical quark masses. Many present-day lattice calculations are already performed directly at, or very close to, the physical values of the quark masses [8, 9, 10], so that the latter extrapolation becomes less of an issue.

The results of the lattice QCD simulations are in good agreement with the experimental data as shown in Fig. 1.8 for the spectrum of hadron masses obtained from lattice QCD simulations, taken from Ref. [8].

### 1.3 Quark–Gluon Plasma

The concept of deconfinement refers to the transition from a confined state to a state where quarks and gluons are no longer confined within colour-neutral hadrons, but can move in a larger volume. As modelled by the MIT bag model, non-perturbative QCD effects can be described in terms of an external pressure, which confines quarks and gluons within a finite volume. If the external pressure is overcome by the pressure due to the kinetic energy of the quarks and gluons, then the hadrons constituents are no longer confined, and a transition to a state called Quark-Gluon Plasma (QGP) occurs. Lattice QCD calculations are used to understand the properties of the QGP, and they predict that a strongly interacting system with zero net baryon density evolves smoothly from a confined (hadronic) towards a deconfined (partonic) state when its temperature is increased up to  $\sim 155$  MeV [11, 12] ( $1.8 \times 10^{12}$  K, around  $10^6$  times the temperature of the core of the Sun), reaching energy densities of  $\sim 0.5$  GeV/fm<sup>3</sup> [10].

It is believed that the Universe underwent a phase of deconfinement in the early stages of its evolution, few microseconds after the Big Bang [14]. Direct observation of the primordial QGP (i.e. that created just after the Big Bang) would provide

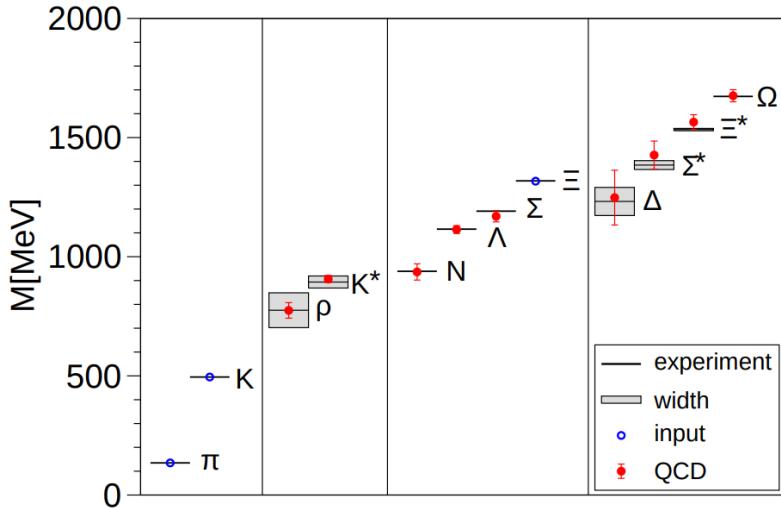


Figure 1.8: The light hadron mass spectrum of QCD. Horizontal lines and bands are the experimental values with their decay widths. Lattice QCD results [8] are shown by solid circles. Vertical error bars represent the combined statistical and systematic error estimates.  $\pi$ ,  $K$ , and  $\Xi$  have no error bars, because they are used to set the light quark mass, the strange quark mass, and the overall scale, respectively.

a wealth of information on the early Universe; however, the Universe experienced a phase in which electrons were not bound into atoms (electromagnetic plasma), making it opaque to electromagnetic radiation, and denying us the possibility of directly observing the primordial QGP. Once the Universe cooled enough (3000 K) to allow electrons to bind to nuclei, the electromagnetic radiation decoupled with a black body spectrum of around 3000 K. Since then, as the Universe expanded, this electromagnetic radiation has redshifted to a temperature of around 2.7 K, and is denoted as the Cosmic Microwave Background (CMB). The CMB is the oldest light in the Universe and provides a snapshot of the Universe when it was 300'000 years old, long after the QGP phase. Although recent studies evidenced the presence of QGP in massive neutron star cores [15], a more controlled environment for studying the QGP is achieved by recreating it in laboratories, through the collision of heavy ions at ultra-relativistic energies. In the past decades, several experiments [16, 17, 18, 19] have been carried out to study the properties of this state of matter, and the results have provided valuable insights into the properties of the strongly-interacting matter.

### 1.3.1 High-energy heavy-ion collisions

Heavy-ion collisions are the most suitable environment to study the properties of the QGP in the laboratory. In these collisions, two heavy ions, such as lead or gold nuclei, are accelerated to ultra-relativistic energies and made to collide. Given the large amount of energy deposited in the collision, the system reaches high energy densities (of around  $16 \text{ GeV}/\text{fm}^3$  after  $1 \text{ fm}/c$  for central collisions at the LHC [21]), which allow for the production of the QGP for temperatures  $T > T_{pc}$ , where  $T_{pc}$  is the pseudo-critical temperature for the transition to a deconfined state, as shown in

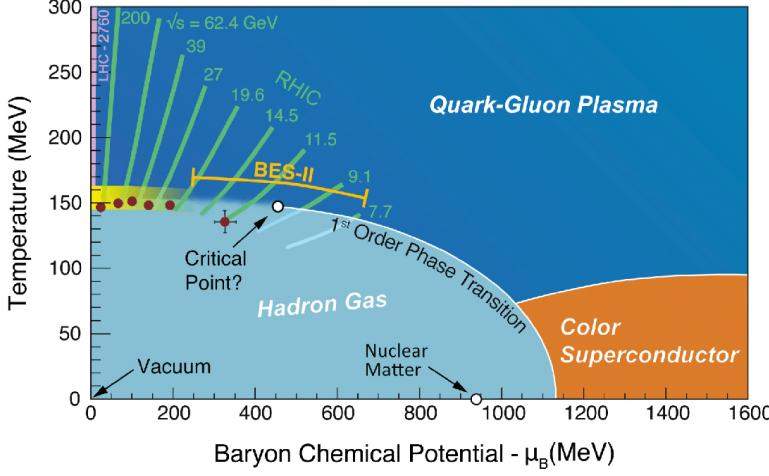


Figure 1.9: QCD phase diagram in the temperature-baryon chemical potential plane. Figure taken from Ref. [13].

Fig. 1.9. As nuclei are objects made of many nucleons (i.e. protons and neutrons), the interaction volume (called *fireball*) is larger and longer-lived than in proton-proton collisions, allowing the usage of thermodynamics and hydrodynamics to describe the system created in the collision.

It is possible to distinguish between two different collision regimes, which probe different phase spaces in the QCD diagram reported in Fig. 1.9. The ordinary nuclear matter is located at approximately  $T \sim 0$  and  $\mu_B = 1$  GeV, where  $\mu_B$  is the baryo-chemical potential, expressing the energy needed to increase the baryon number of the system by one unit, as described in Sec. 1.3.2. When the centre-of-mass energy per nucleon pair ( $\sqrt{s_{NN}}$ ) is below a few GeV, the nucleons of the colliding nuclei are stopped in the collision as they lose energy and momentum. Due to conserved currents, the quantum numbers of the initial state are preserved, so that, for example, the net baryon production (i.e. the difference between the number of produced baryons and antibaryons) is positive. This regime is called *stopping regime*. In this regime, the baryo-chemical potential is increased by about 5–10 times with respect to that of nuclear matter. When the  $\sqrt{s_{NN}}$  increases, the initial state baryon number is carried away by the receding nucleons, and the net baryon number in the fireball at midrapidity ( $y \sim 0$ ) is zero. The rapidity is defined as

$$y = \frac{1}{2} \log \left( \frac{E + p_z}{E - p_z} \right) ,$$

where  $E$  is the energy and  $p_z$  is the longitudinal momentum of the particle. In this regime, denoted as *Bjorken regime*, or *transparency regime*, high temperatures of a few hundred MeV are reached, and  $\mu_B$  is close to 0.

The collision of two nuclei is a rather complex process, with a space-time evolution that can be divided into several stages, as depicted in Fig. 1.10. They can be studied by measuring different final-state observables that are sensitive to such stages. One of the first descriptions of the fireball evolution was given by J. Bjorken [22], who proposed a simple model to describe the expansion of the sys-

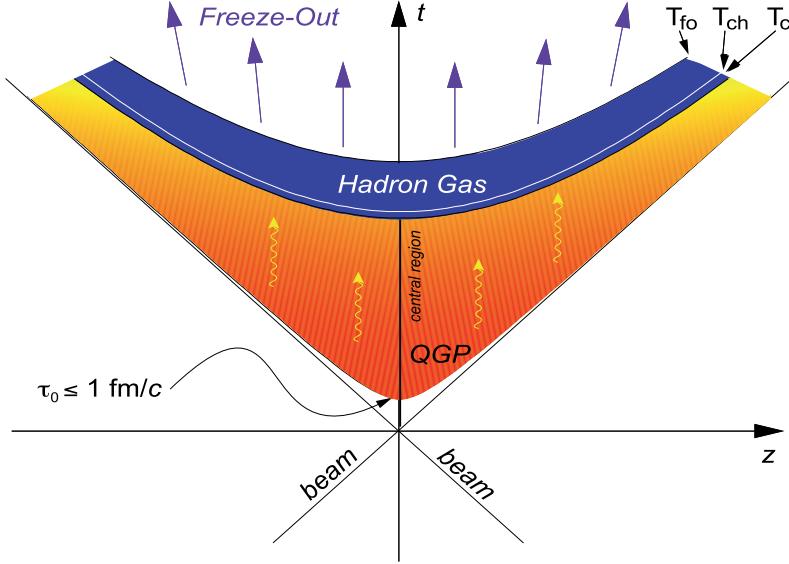


Figure 1.10: Space-time evolution of a heavy-ion collision. Figure taken from Ref. [20].

tem in the longitudinal direction. In this model, it is assumed that there exists a central-plateau structure in the inclusive particle productions as a function of rapidity. In other terms, the Bjorken model assumes that the system is boost-invariant, i.e. independent of the longitudinal velocity. According to this model, the evolution of the system is described by the following stages: i) collision of the two nuclei, ii) pre-equilibrium, iii) QGP formation and expansion, iv) hadronisation and hadron gas expansion, and v) freeze-out.

## Collision

The two nuclei are accelerated to ultra-relativistic energies and are thus Lorentz-contracted in the direction of motion. The collision takes place in a very short time  $\tau_{\text{coll}} = 2R/\gamma$ , where  $R$  is the nucleus radius and  $\gamma$  the Lorentz factor. At LHC energies, the crossing time of the nuclei is of  $\sim 0.005 \text{ fm}/c$ , which is much smaller than the time scale of the strong interaction  $\tau_{\text{strong}} \sim 1/\Lambda_{\text{QCD}} \sim 1 \text{ fm}/c$ . Particles produced in the collision through parton interactions mediated by the strong force are thus created once the colliding nuclei have already passed through each other and moved away from the interaction region. During the collision, a large amount of energy is deposited in the interaction region, leading to the formation of a fireball with high energy densities.

## Pre-equilibrium

After  $\tau_{\text{coll}}$ , secondary particles are produced from the energy deposited in the collision. In the Bjorken regime, it is possible to evaluate the energy density of the system as a function of time by measuring the transverse energy of the particles produced at midrapidity, where the net baryon density is zero. The energy density

is given by

$$\varepsilon_{\text{Bjorken}} = \frac{\langle m_T \rangle}{\tau A} \left. \frac{dN}{dy} \right|_{y=0} = \frac{1}{\tau A} \left. \left\langle \frac{dE_T}{dy} \right\rangle \right|_{y=0},$$

where  $A$  is the transverse area collision region,  $\langle m_T \rangle$  is the mean transverse mass, defined as  $m_T = \sqrt{m^2 + p_T^2}$ ,  $N$  is the number of secondary particles, and  $\langle dE_T/dy \rangle$  is the mean transverse energy density. It is interesting to evaluate the energy densities reached at the time of particle formation, which can be estimated using Heisenberg's uncertainty principle:  $\tau = \hbar/m_T$ . The mean transverse mass of the particles produced in the collision at the proper time of formation  $\tau_f$  can be evaluated using the approximate formula:

$$\langle m_T \rangle = \frac{\frac{dE_T(\tau_f)}{dy}}{\frac{dN(\tau_f)}{dy}} \sim \left. \frac{\frac{dE_T}{d\eta}}{\frac{dN}{d\eta}} \right|_{\text{final state}},$$

leading to a formation time of  $\sim 0.35 \text{ fm}/c$  in Au–Au collisions at  $\sqrt{s_{\text{NN}}} = 200 \text{ GeV}/c$  at RHIC [23], to which corresponds an energy density of  $\sim 15 \text{ GeV}/\text{fm}^3$ . Measurements at LHC energies yield a higher energy density than RHIC by about a factor of 2.3 [24].

## Quark-gluon plasma formation

The system of produced particles reaches thermal equilibrium through multiple scatterings at a proper time  $\tau_{\text{eq}}$ , and the QGP is formed. Its evolution can be described using relativistic hydrodynamics. Studies at RHIC allow to constrain  $\tau_{\text{eq}}$  in the range of  $0.6 < \tau_{\text{eq}} < 1 \text{ fm}/c$ , corresponding to energy densities of  $5.4 < \varepsilon_{\text{Bjorken}}(\tau_{\text{eq}}) < 9 \text{ GeV}/\text{fm}^3$ , well above the threshold for QGP formation.

## Hadronisation

The deconfined system expands and cools down, until the temperature decreases below the pseudo-critical value for the transition crossover ( $T_{\text{pc}} \sim 155 - 165 \text{ MeV}$ ). The QGP undergoes a transition and hadronises, producing an expanding gas of colour-neutral particles. At LHC energies, this is expected to happen  $\sim 10 \text{ fm}/c$  [25] after the QGP formation. This transition is associated with a sharp decrease in the entropy density of the system.

## Freeze-out

During the hadron gas expansion, the particle density decreases to a point where the inelastic interactions cease. The moment in which this happens is called *chemical freeze-out*, and takes place at around  $T_{\text{chem.}} \sim 160 \text{ MeV}$  [26]. This temperature value is very close to  $T_{\text{pc}}$ , suggesting that the chemical freeze-out takes place very close to, or simultaneously with, the hadronisation. The chemical composition of the hadron gas cannot vary anymore, although elastic interactions, which change the momentum spectrum of the produced particles still occur. Once the temperature decreases below  $T_{\text{therm.}} \sim 130 \text{ MeV}$ , the elastic interactions cease as well, and this is referred to as *thermal freeze-out*. The particles then keep expanding without interactions (*free-streaming*), and are detected by the experimental apparatus.

### 1.3.2 Hadron species abundances

Measurements of the multiplicity of identified hadron species, i.e., the chemical composition of the system, allow for the investigation of various properties of the system at the time of chemical freeze-out. The relative abundances of different hadron species produced in hadronic collisions are found to be well described by the *Statistical Hadronisation Model* (SHM) [27]. The SHM is based on the assumption that the fireball created in an ultra-relativistic heavy-ion collision is in thermal and chemical equilibrium at the time of chemical freeze-out, which is assumed to be characterised by the same temperature for all the hadronic species.

A system is in thermal equilibrium when the distribution of particle velocities follows a Maxwell-Boltzmann distribution, which depends solely on the temperature of the system. Chemical equilibrium is achieved when the relative multiplicities of the different hadron species are described by their mass, spin degeneracy, and the temperature of the system. The hypothesis of thermal and chemical equilibrium is postulated, and allows the use of statistical mechanics to describe the system by defining a partition function that characterises the statistical properties of a system in equilibrium. The validity of this assumption is then verified a posteriori by comparing the model predictions with the experimental data. The equilibrium at the chemical freeze-out is purely statistical, i.e. the particle abundances are determined through the principle of maximum entropy, which suggests that hadronisation fills the phase space in the most probable configuration. Furthermore, the SHM does not take into account the dynamics of the system, but only its state at the time of the chemical freeze-out, without any assumption on the presence of a partonic phase.

The system is described in the SHM picture in terms of a non-interacting gas of hadrons and resonances, which corresponds, in a Yukawa approach, to a hadron gas interacting through the exchange of resonances. This framework was found to be consistent with the equation of state of strongly-interacting matter from lattice QCD for temperatures below the pseudo-critical temperature of the QCD phase transition [11].

The ensemble used to describe the system is the grand canonical ensemble, which describes a system exchanging energy and particles with the environment. In a grand canonical ensemble, the energy and the quantum numbers are conserved on average, but not exactly and locally as in a micro canonical ensemble. The fireball produced in an ultra-relativistic heavy-ion collision can be described in terms of small clusters that exchange energy and particles with the rest of the system. Smaller collision systems such as pp and  $e^+e^-$  collisions, where fewer particles are produced, are described by the canonical ensemble, where the quantum numbers are conserved exactly and locally, while the energy is conserved on average.

The density of the hadron species  $i$  in the system can be calculated from the partition function  $Z_i^{GC}$  and is given by

$$n_i = \frac{1}{V} \frac{\partial(TZ_i^{GC})}{\partial\mu_i} = \frac{g_i T^2}{2\pi^2} \sum_{k=1}^{\infty} \frac{(\pm 1)^{k+1}}{k} \lambda_i^k m_i^2 K_2 \left( \frac{km_i}{T} \right) ,$$

where  $V$  is the volume of the system,  $g_i$ ,  $\lambda_i = e^{\mu_i/T}$  and  $m_i$  are the degeneracy factor,

the fugacity, and the mass of the hadron species i, respectively. The + sign holds for bosons, while the – sign is for fermions.  $K_2$  is the second-order modified Bessel function of the second kind. The chemical potential  $\mu_i$  of the hadron species i is related to the conservation (on average, on a large volume) of the quantum numbers of the system. It is given by  $\mu = \sum_j \mu_{Qj} Q_j$ , where  $Q_j$  are the conserved quantum numbers (third component of isospin, strangeness, and baryon number), and  $\mu_{Qj}$  are the chemical potentials, which guarantee the conservation of the conserved charges.  $\mu_i$  is thus the energy needed to add a particle with quantum numbers  $Q_j$  to the system. The obtained multiplicities of the hadron species are then corrected with a feed-down contribution from decays of short-lived particles, which cannot be detected directly.

Of the five free parameters of the SHM ( $T$ ,  $\mu_B$ ,  $\mu_s$ ,  $\mu_{I_3}$ , and  $V$ ), two can be fixed through conservation of the initial state quantum numbers. Particle abundances are typically fitted in terms of the temperature  $T$ , the baryon chemical potential  $\mu_B$  and the volume of the system  $V$ .

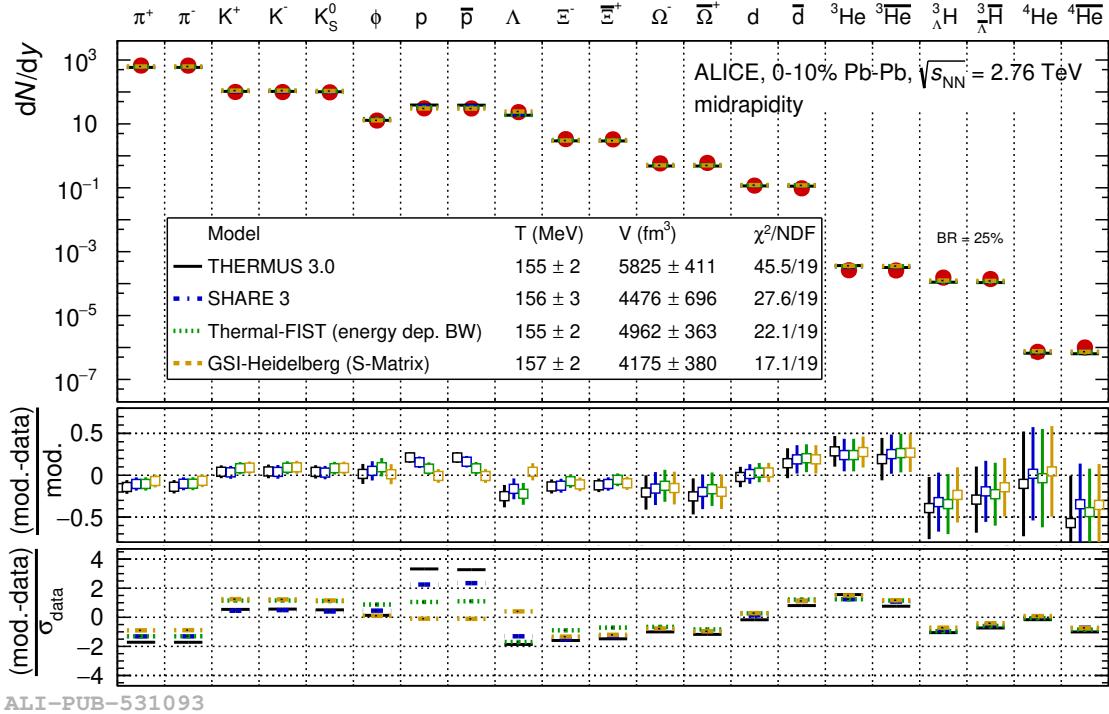


Figure 1.11: Multiplicity per unit of rapidity of different hadron species and light nuclei measured by ALICE compared to SHM fits from THERMUS [28], SHARE [29], Thermal-FIST [30], and GSI-Heidelberg [31]. Differences between the model calculations and the measured yields are shown in the bottom panels. Figure taken from Ref. [16].

Figure 1.11 illustrates the yields of different identified hadron species measured by the ALICE Collaboration in the 10% most central Pb–Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV. They are compared to predictions from various implementations of the SHM, including THERMUS [28], GSI-Heidelberg [31], SHARE [29], and

Thermal-FIST [30]. These models make different assumptions regarding the equilibrium condition and/or conservation laws at the chemical freeze-out stage. The SHM is capable of describing the different particle abundances, which span a large range of 9 orders of magnitude. The yields of exotic hadron species, such as hyperons and light nuclei are precisely described too. This confirms the assumption that the system is in thermal and chemical equilibrium at the time of the chemical freeze-out, and that the outcome of the hadronisation process is well described by a statistical approach.

The chemical freeze-out temperature is found to be of  $\sim 155$  MeV, very close to that predicted by lattice QCD calculations for QCD phase transition. Most recent results from the ALICE Collaboration [32] show that the measured electric charge and baryon chemical potentials at midrapidity at the LHC energies are compatible with 0, indicating that the system created in Pb–Pb collisions at the LHC is on average baryon free and electrically neutral.

### 1.3.3 Radial flow

After the chemical freeze-out, inelastic interactions between hadrons do not take place anymore. However, elastic interactions can still occur in the hadron gas, leading to a modification of the momentum distribution of the particles until the system reaches the thermal freeze-out. By studying the transverse momentum spectra of the particles produced in the collision, it is possible to extract information on the properties of the produced particles at the thermal freeze-out, and on the collective expansion (*flow*) of the system.

For a stationary thermal source with a temperature  $T$ , the Lorentz-invariant momentum distribution of the particles is given by

$$E \frac{d^3N}{dp^3} = \frac{d^3N}{p_T dp_T d\varphi dy} = \frac{g_i V}{(2\pi)^3} E \frac{1}{e^{(E-\mu_i)/T} \pm 1} ,$$

where  $E$  is the energy of the particle,  $g_i$  and  $\mu_i$  are the degeneracy factor and chemical potential of the particles of species  $i$ , respectively,  $V$  is the volume of the source, and the  $+1$  term is for fermions while the  $-1$  is for bosons. By using the relation

$$\frac{dN}{p_T dp_T} = \frac{dN}{m_T dm_T} ,$$

and by assuming that  $e^{(E-\mu_i)/T} \gg 1$ , the transverse momentum spectrum of particles can be obtained by integrating over the azimuthal angle  $\varphi$  and the rapidity  $y$ , leading to

$$\frac{dN}{m_T dm_T} = \frac{g_i V}{(2\pi)^2} m_T e^{\mu_i/T} K_1 \left( \frac{m_T}{T} \right) \xrightarrow{m_T \gg T} V' \sqrt{m_T} e^{-m_T/T} . \quad (1.10)$$

From Eq. 1.10 emerges that for a stationary thermal particle source with a temperature  $T$ , the transverse mass spectrum of the particles follows an exponential distribution with a slope parameter  $T$ , and is independent of the particle species. This is known as the  $m_T$ -scaling of the transverse mass spectra, and is observed in pp and small-system collisions at low  $\sqrt{s}$ , with a slope parameter of  $T \sim 167$  MeV [33]. However, in heavy-ion collisions, this description is not valid, as the QGP, which is

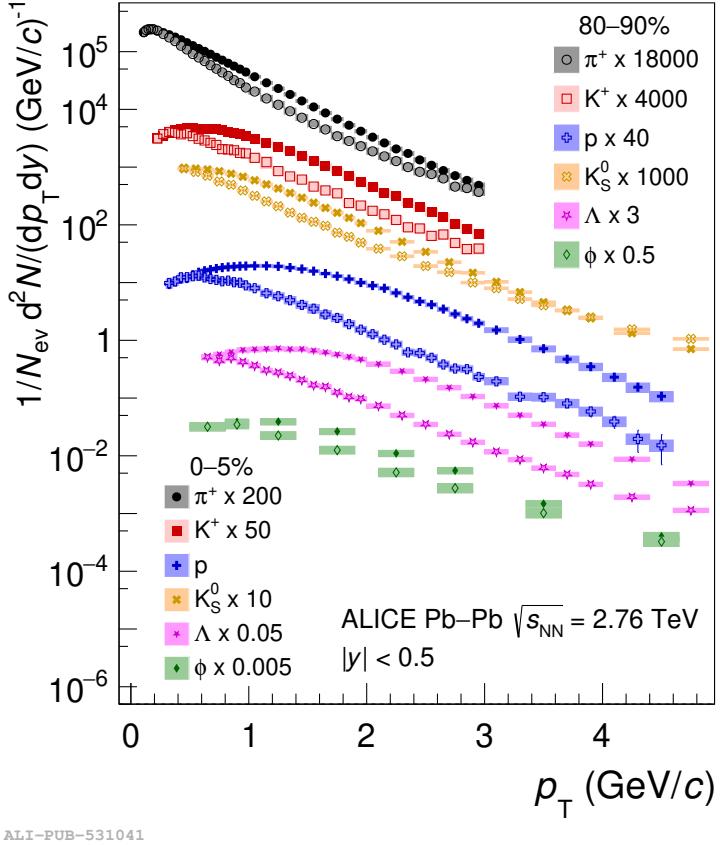


Figure 1.12: Transverse momentum distributions of  $\pi^+$ ,  $K^+$ ,  $p$ ,  $K_S^0$ ,  $\Lambda$ , and  $\phi$  mesons for the 0–5% and 80–90% centrality intervals in Pb–Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV measured by the ALICE Collaboration. The data points are scaled by various factors for better visibility. Figure taken from Ref. [16].

a thermalised system of deconfined quarks and gluons, has a thermal pressure. The pressure difference between the QGP and the surrounding vacuum leads to a collective expansion of the fireball, which is called *radial flow*. This causes the particles to be pushed in the transverse direction, causing a modification of the transverse momentum spectra of the particles, which is characterised by a shift of the spectra towards higher  $p_T$  values, with a more pronounced effect for heavier particles, as shown in Fig. 1.12, where the  $p_T$  distributions measured by the ALICE Collaboration [16] of different hadron species are shown for central (0–5%) and peripheral (80–90%) Pb–Pb collisions. The radial flow, as expected, is more pronounced in central collisions (due to the larger pressure gradients) which is manifest in the harder  $p_T$  spectra for central collisions as compared to peripheral ones. The radial flow is determined by a correlation between the velocity of an element of the system and its space-time position, which is superimposed on the random thermal motion of the particles.

The radial flow can be described by using phenomenological models, such as the blast-wave model [35], which assumes that the particles are emitted from a thermal

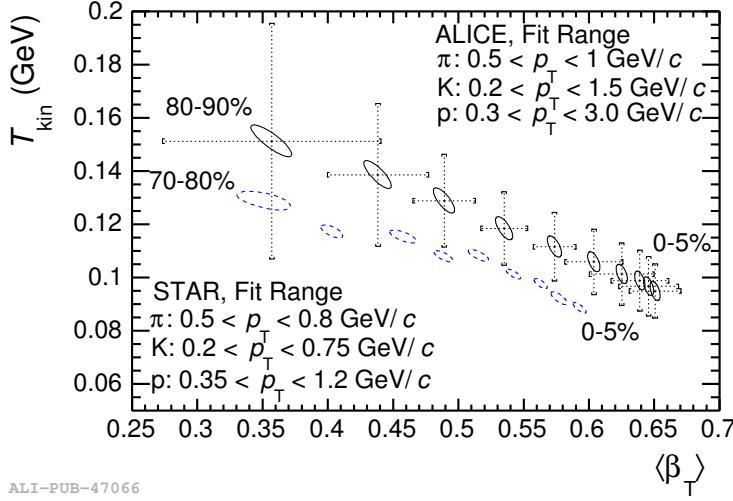


Figure 1.13: Results of blast-wave fits obtained by the ALICE Collaboration at the LHC, compared to similar fits at RHIC energies from the STAR Collaboration, for different centrality intervals. Figure taken from Ref. [34]

source with a collective velocity field. It relies on the Cooper-Frye prescription [36], which assumes an instantaneous freeze-out of the particles in the radial direction, to calculate the transverse momentum spectra of the particles in terms of the temperature of the source, the collective velocity field, and the transverse velocity profile. Results from SPS, RHIC, and LHC experiments show that the transverse momentum spectra of the particles produced in heavy-ion collisions can be well described by the blast-wave model, with a temperature of the source (which corresponds to the thermal freeze-out temperature  $T_{\text{therm.}}$ ) of around 110–120 MeV, almost independent of  $\sqrt{s_{\text{NN}}}$ , and a collective velocity field of around 0.5–0.6  $c$  in the most central collisions, with a growing trend with  $\sqrt{s_{\text{NN}}}$ , as shown in Fig. 1.13.

Hydrodynamics-based models provided a more robust description of the system’s expansion as compared to blast-wave fits. They rely on energy and momentum conservation laws, which need to be expressed in a relativistic form due to the relativistic nature of the system:

$$\partial_\mu T^{\mu\nu} = 0 \quad , \text{ with } T^{\mu\nu} = (\varepsilon + p)u^\mu u^\nu - pg^{\mu\nu} \quad ,$$

where  $T^{\mu\nu}$  is the energy-momentum tensor,  $\varepsilon$  is the energy density,  $p$  is the pressure,  $u^\mu$  is the four-velocity of the fluid element, and  $g^{\mu\nu}$  is the metric tensor. An additional condition comes from the conservation of the baryon number, which is expressed as

$$\partial_\mu(n_B u^\mu) = 0 \quad ,$$

where  $n_B$  is the baryon density. A further equation is needed to close the system of equations, and is provided by the equation of state of the QGP, which relates the pressure to the energy density and the baryon density. As the fireball is in a non-equilibrium state in the early stages of the collision, the hydrodynamics equations need an initial condition to describe the system at the time  $\tau_{\text{eq}}$  in which it reaches

the equilibrium and hydrodynamics can be applied. This is usually obtained via the Glauber [37] or T<sub>R</sub>ENTo [38] models, which describe the collision of the two nuclei as a superposition of nucleon-nucleon collisions, and provide the initial energy density profile of the system. Models based on the Colour Glass Condensate framework such as IP-Glasma [39], or Monte Carlo simulations based on the description of partonic showers such as AMPT [40] can also be used to provide the initial conditions for the hydrodynamics calculations.

The hydrodynamics calculations provide a good description of the transverse momentum spectra of the particles produced in heavy-ion collisions, and the emerging parameters (freeze-out temperature, flow velocity) are similar to those obtained with the blast-wave model. In addition, hydrodynamics calculations and their extensions, such as viscous hydrodynamics, can be used to study other observables related to the collective expansion of the system, such as the elliptic flow, which is a measure of the anisotropy of the particle emission in the transverse plane.

### 1.3.4 High- $p_T$ hadrons and jet quenching

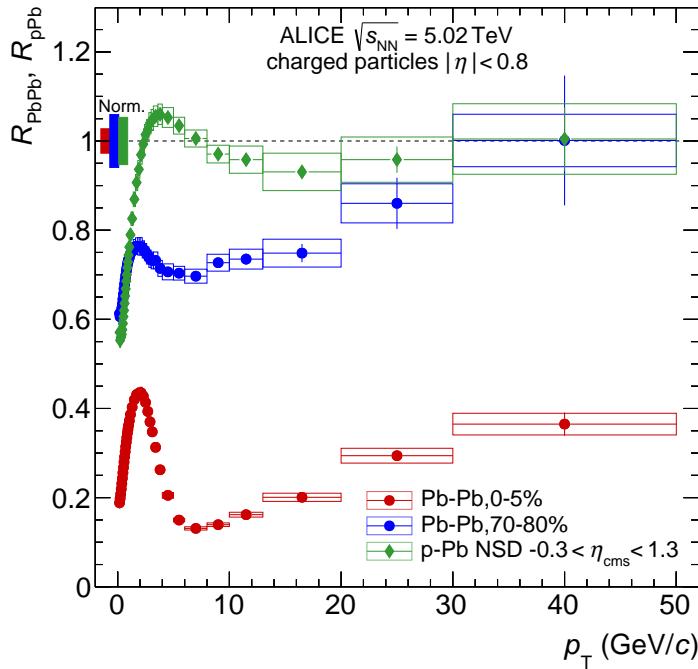


Figure 1.14: Nuclear modification factors measured by ALICE in central (0 – 5%) and peripheral (70 – 80%) Pb–Pb collisions at  $\sqrt{s_{\text{NN}}} = 5.02$  TeV and in p–Pb collisions at  $\sqrt{s_{\text{NN}}} = 5.02$  TeV. Figure taken from Ref. [41].

High- $p_T$  partons are produced in hard-scattering processes, i.e., those with a high momentum transfer. As such, they are produced in the early stages of the collision and experience the whole fireball evolution. They can therefore be used to probe the properties of the earliest stages of the QGP. The Glauber model predicts the

production cross-section of hard-scattering processes to scale with the number of binary nucleon-nucleon collisions. To test this prediction, the nuclear modification factor  $R_{\text{AA}}$  is defined as the ratio of the  $p_{\text{T}}$ -differential hadron yield in heavy-ion collisions to the one in proton-proton collisions, scaled by the mean number of binary nucleon-nucleon collisions  $\langle N_{\text{coll}} \rangle$ :

$$R_{\text{AA}} = \frac{dN_{\text{AA}}/dp_{\text{T}}}{\langle N_{\text{coll}} \rangle dN_{\text{pp}}/dp_{\text{T}}} .$$

The nuclear modification factor of particles produced in hard-scattering processes (jets, high  $p_{\text{T}}$  hadrons, heavy flavours) is expected to be equal to unity, namely *binary scaling*, in the absence of nuclear effects. However, experimental results [41] show that the  $R_{\text{AA}}$  of charged particles measured in the most central heavy-ion collisions exhibits a suppression at low  $p_{\text{T}}$ , a peak at intermediate  $p_{\text{T}}$  ( $\sim 2 \text{ GeV}/c$ ), and a significant suppression by a factor of  $\sim 5$  at high  $p_{\text{T}}$ , as shown in Fig. 1.14. The suppression at low  $p_{\text{T}}$  and the peak at intermediate  $p_{\text{T}}$  are mostly attributed to soft particle production, which dominates at low  $p_{\text{T}}$  and does not scale with  $N_{\text{coll}}$ . In addition, other initial-state effects, such as the modification of parton distribution functions in the nucleus [42], and the Cronin effect [43] modify the spectra in Pb–Pb (and p–Pb) collisions as compared to pp. In addition, the radial flow causes a modification of the spectral shapes at low  $p_{\text{T}}$  and affects the  $R_{\text{AA}}$ . The initial-state effects can be studied in small collision systems [44], where the production of a QGP is not expected, or through the measurement of the  $R_{\text{AA}}$  for colour-neutral particles, such as photons [45], which are not affected by the presence of a deconfined medium. The high- $p_{\text{T}}$  suppression of  $R_{\text{AA}}$  increases with collision centrality and is related to the formation of the QGP. Partons traversing the medium lose energy through elastic scatterings with the QGP constituents and via gluon radiation, which is the dominant process for high-energy partons. This energy loss can be described in the BDMPS formalism [46], which assumes that the radiated gluon becomes de-coherent from the emitting parton through multiple soft scatterings with the medium constituents. The energy loss of the partons is quantified as

$$\Delta E = \frac{1}{4} \alpha_s C_R \hat{q} L^2 ,$$

Where  $\alpha_s$  is the strong coupling constant,  $C_R$  is the Casimir factor, which is 3 for gluon-gluon couplings and  $4/3$  for quark-gluon interactions,  $\hat{q}$  is the transport coefficient, and  $L$  is the path length of the parton in the medium. The transport coefficient is related to the energy density of the medium, as  $\hat{q} \propto \varepsilon^{3/4}$  [47], so that energy loss measurements can be used to infer the properties of the QGP. Typical energy losses of high- $p_{\text{T}}$  partons traversing the whole QGP are of the order of 40 GeV, which is a significant fraction of the parton energy. The energy loss of the partons leads to a suppression of high- $p_{\text{T}}$  hadron yield, which consequently leads to a suppression of the nuclear modification factor at high  $p_{\text{T}}$ .

At leading order, partons are produced in back-to-back pairs, forming what are known as di-jets. In the presence of the QGP, the jets are expected to be modified due to the interactions with the medium, an effect known as *jet quenching*. For example, the opening angle of the jets is observed to be broadened [49], and the  $p_{\text{T}}$

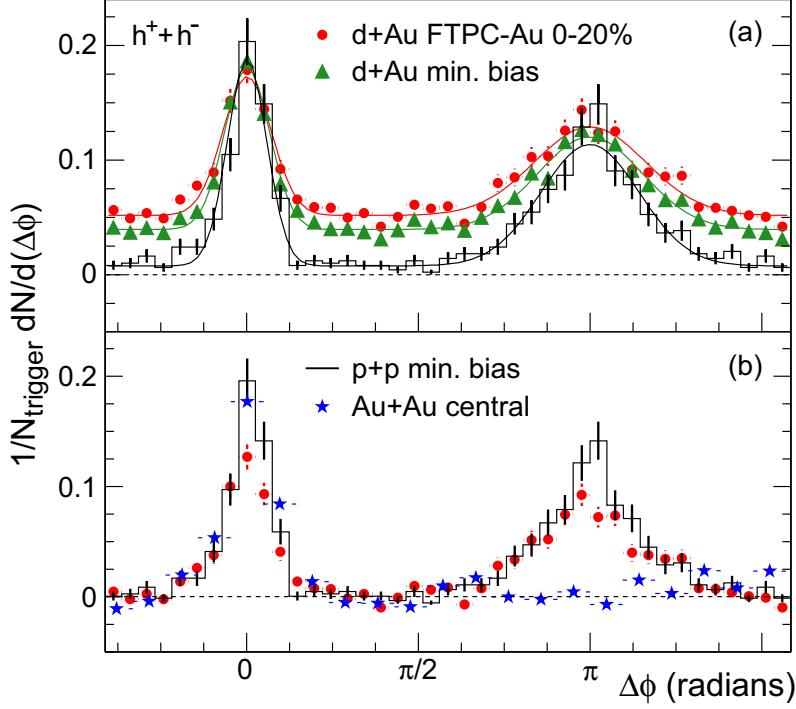


Figure 1.15: Top: Two-particle azimuthal correlations for minimum bias and central d–Au collisions, and for proton-proton collisions at  $\sqrt{s_{\text{NN}}} = 200$  GeV measured by the STAR Collaboration. Bottom: Comparison of pedestal-subtracted two-particle azimuthal correlations for central Au–Au collisions to those seen in proton-proton collisions. Figure taken from Ref. [48].

distribution of the leading hadron (i.e., the one with the highest  $p_T$ ) is expected to be softer than that measured in pp collisions [50], therefore yielding an  $R_{\text{AA}}$  smaller than unity at high  $p_T$ .

Further insights into the jet quenching mechanism can be obtained by measuring two-particle azimuthal correlations. For each event, the particle with the highest  $p_T$  above a certain threshold is selected as the trigger particle, and the azimuthal angle of other high- $p_T$  particles is measured with respect to the first one. For proton-proton and proton-nucleus collisions, the azimuthal distribution presents two peaks: the *near-side peak*, which is found at  $\varphi = 0$  and is due to the hadrons produced in the same jet as the trigger particle, and the *away-side peak*, which is found at  $\varphi = \pi$  and is due to the hadrons produced in the opposite jet. In heavy-ion collisions, the away-side peak is suppressed due to the energy loss in the QGP, as shown in Fig. 1.15. In fact, for a di-jet surface emission, the parton shower generated by the jet pointing in the direction of the medium is expected to lose more energy than the one pointing in the opposite direction, yielding the suppression of the away-side peak.

### 1.3.5 Strangeness enhancement

One of the first proposed signatures of the QGP formation is the strangeness enhancement [51], which refers to the observation of an increased production of strange

hadrons in heavy-ion collisions compared to proton-proton collisions. At a microscopic level, the enhancement of strange hadrons is a direct consequence of the high temperatures reached in the QGP, which allow for the thermal production of strange quarks and antiquarks. Strange quarks are not present as valence quarks in the initial state, and can be produced in hard  $2 \rightarrow 2$  scatterings ( $gg \rightarrow s\bar{s}$ ,  $q\bar{q} \rightarrow s\bar{s}$ ), or via gluon splitting ( $g \rightarrow s\bar{s}$ ) during the evolution of the system. While these processes are dominant for the production of strange hadrons with high  $p_T$ , at low  $p_T$  the production of strangeness is dominated by non-perturbative processes. At a macroscopic level, the enhancement of the production of strange hadrons in the presence of QGP can be explained using the statistical approach of the SHM. The description of hadron production in smaller colliding systems, such as  $e^+e^-$  and pp collisions, requires the use of a canonical ensemble, which is used to describe a system that exchanges only energy with a reservoir, and requires the exact (local) conservation of quantum numbers. This reduces the phase space available for strange particle production [52], leading to a suppression of strange quark production as compared to the grand canonical ensemble, called *canonical suppression*.

The strangeness enhancement in heavy-ion collisions has been observed by several Collaborations, such as the WA97 [53] and NA57 [54] at the SPS, STAR at RHIC [55], and ALICE at the LHC [56]. In addition, an enhancement of strange hadrons has been observed in small systems, such as pp [57] and p-Pb [58, 59] collisions, in high-multiplicity events, and follows a hierarchy determined by the hadron strangeness. This observation is intriguing, as QGP production is not expected in such systems. Furthermore, several effects such as azimuthal correlations and mass-dependent hardening of  $p_T$  distributions, which are typically associated with QGP formation in nuclear collisions, have been observed in these systems. Could small droplets of QGP be formed in these collisions? Could other mechanisms explain the observed effects? These are still open questions in the field of high-energy nuclear physics, and they are the subjects of ongoing research.

Figure 1.16 shows the  $p_T$ -integrated yield ratios of strange ( $K_s^0$ ,  $\Lambda$ ) and multi-strange ( $\Xi^\pm$ ,  $\Omega^\pm$ ) hadrons to pions ( $\pi^+ + \pi^-$ ) as a function of the charged particle multiplicity density,  $\langle dN_{ch}/d\eta \rangle$ , measured in  $|y| < 0.5$  in different collision systems (pp, p-Pb and Pb-Pb) by the ALICE Collaboration at the LHC [57]. The data show a clear enhancement of strange to non-strange hadron production with a smooth trend with increasing multiplicity. At the highest multiplicities of Pb-Pb collisions, the yield ratios saturate at values that are described by SHM calculations with a grand canonical ensemble. This can be interpreted as a consequence of the formation of a QGP in Pb-Pb collisions, where the  $s\bar{s}$  is expected to rapidly reach its equilibrium value.

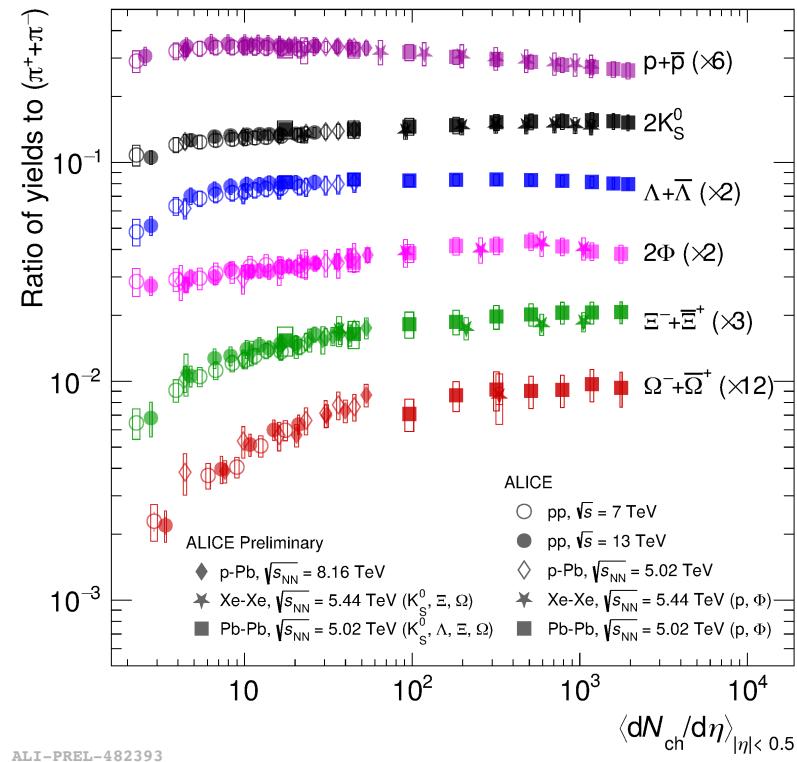


Figure 1.16:  $p_T$ -integrated yield ratios of strange ( $K_s^0$ ,  $\Lambda$ ) and multi-strange ( $\Xi^\pm$ ,  $\Omega^\pm$ ) hadrons to pions ( $\pi^+ + \pi^-$ ) as a function of  $\langle dN_{ch}/d\eta \rangle$  measured in pp, p-Pb and Pb-Pb collisions at midrapidity ( $|y| < 0.5$ ). Figure taken from Ref. [60].

# Chapter 2

## Open heavy-flavour production in proton-proton collisions

Open heavy-flavour hadrons, composed of a heavy quark (charm or beauty) along with lighter quarks, are exclusively formed in high-momentum transfer processes due to the large masses of approximately  $1.3 \text{ GeV}/c^2$  and  $4.2 \text{ GeV}/c^2$  [4] for charm and beauty quarks, respectively. As a result, heavy quarks are created in the early stages of the collision, and their production cross-section in the partonic interaction can be evaluated perturbatively using QCD. Studying the production of open heavy-flavour hadrons in proton-proton (pp) collisions not only provides a crucial test of the perturbative QCD framework, but also allows to set constraints on heavy-quark hadronisation models. Furthermore, measurements in proton-proton collisions, where the production of an extended deconfined medium is not expected, are a necessary reference for the study of heavy-ion collisions, where the properties of the QGP can be investigated.

### 2.1 Factorisation theorem

The production of open heavy-flavour hadrons in proton-proton collisions can be described using the factorisation theorem [61], which allows for the separation of short-distance, perturbative behaviour from long-distance, non-perturbative phenomena. The total production cross-section can be expressed as

$$\sigma_{\text{pp}}^{\text{H}} = \sum_{a,b=g,q,\bar{q}} \int dx_1 dx_2 f_{a/A}(x_1, \mu_F^2) f_{b/B}(x_2, \mu_F^2) \hat{\sigma}_{ab \rightarrow Q\bar{Q}}(x_1, x_2, \mu_F^2, \mu_R^2) D_{Q \rightarrow H_Q}(z, \mu_F^2) , \quad (2.1)$$

i.e., the convolution of: i) the Parton Distribution Functions (PDFs)  $f_{a/A}(x_1, \mu_F^2)$  and  $f_{b/B}(x_2, \mu_F^2)$ , describing the initial-state probability of finding a parton  $a$  in the proton  $A$  carrying a fraction  $x_1$  of the proton's momentum, and a parton  $b$  in the proton  $B$  carrying a fraction  $x_2$  of its momentum, respectively; ii) the hard partonic scattering cross-section  $\hat{\sigma}_{ab \rightarrow Q\bar{Q}}(x_1, x_2, \mu_F^2, \mu_R^2)$ , defining the probability of producing the  $Q\bar{Q}$  final state from the collision of partons  $a$  and  $b$ , where  $Q$  is either a charm or beauty quark; and iii) the Fragmentation Functions (FFs)  $D_{Q \rightarrow H_Q}(z, \mu_F^2)$ , which describe the probability of the heavy quark  $Q$  to fragment into a heavy-flavour

hadron  $H_Q$  carrying a fraction of the parent parton momentum  $z$ .  $\mu_F$  and  $\mu_R$  are the factorisation and renormalisation scales, respectively. The former is the scale at which the non-perturbative processes are separated from the perturbative ones, while the latter is the scale at which the ultraviolet divergences, emerging from the calculations of scattering amplitudes involving loop diagrams, are subtracted. While the PDFs and FFs are non-perturbative quantities, parametrised from experimental data and regarded as universal across different collision systems, the hard partonic scattering cross-section can be perturbatively calculated using QCD, but needs specific evaluations for each process.

The factorisation theorem has been widely employed to describe the production of open heavy-flavour hadrons in proton-proton collisions, and has proven to be successful in modeling experimental data. Figure 2.1 shows the production cross-section of prompt and non-prompt  $D^0$ -mesons in proton-proton collisions at  $\sqrt{s} = 5.02$  TeV measured at midrapidity ( $|y| < 0.5$ ) as a function of the transverse momentum ( $p_T$ ) by the ALICE Collaboration [62], compared to Fixed-Order plus Next-to-Leading Logarithms (FONLL) perturbative QCD predictions [63] combined with the PYTHIA 8 [64] event generator for the  $H_b \rightarrow D^0 + X$  decay kinematics. The calculation is performed at next-to-leading order, including all the  $\alpha_s^2$  and  $\alpha_s^3$  terms, with the resummation of the next-to-leading logarithmic terms. These include terms of the form  $\alpha_s^2 \alpha_s^k \log^k(p_T/m)$  and  $\alpha_s^3 \alpha_s^k \log^k(p_T/m)$ , where  $m$  is the heavy-quark mass. The term *prompt* refers to charm-hadrons directly produced in the hadronisation of a charm quark or through the strong decay of a directly produced excited charm-hadron or charmonium state, while *non-prompt* charm hadrons are produced in the decay of a hadron containing a beauty quark. The FONLL predictions are in good agreement with the non-prompt  $D^0$ -meson production cross-section, whereas the prompt contribution lies at the upper edge of the theoretical uncertainty band, albeit being described within the uncertainties. Similar trends are observed in the production of other open heavy-flavour hadrons across different experimental facilities, such as the Tevatron [65], RHIC [66], and LHC [62].

### 2.1.1 Parton Distribution Functions

#### Deep inelastic scattering

The PDFs are non-perturbative quantities describing the probability of finding a parton carrying a fraction  $x$  of the proton's momentum in the initial state of a process. The first experimental evidence revealing the partonic structure of the proton emerged from deep inelastic scattering experiments carried out at the Stanford Linear Accelerator Center (SLAC) in the 1960s [67], where an electron was scattered off a proton with momentum  $P$ , and the transferred momentum  $q$  was measured. The cross-section for deep inelastic scattering can be defined in terms of the Lorentz invariant variables  $Q^2 = -q^2$  and  $x = \frac{Q^2}{2P \cdot q}$ , yielding

$$\frac{d^2\sigma}{dx dQ^2} = \frac{4\pi\alpha^2}{xQ^4} [(1-y) F_2(x, Q^2) - xy^2 F_1(x, Q^2)] \quad ,$$

where  $y = Q^2/(sx)$ ,  $s = (P + p_e)^2$  denotes the centre-of-mass energy of the electron-proton system, and the structure functions  $F_1(x, Q^2)$  and  $F_2(x, Q^2)$  represent an

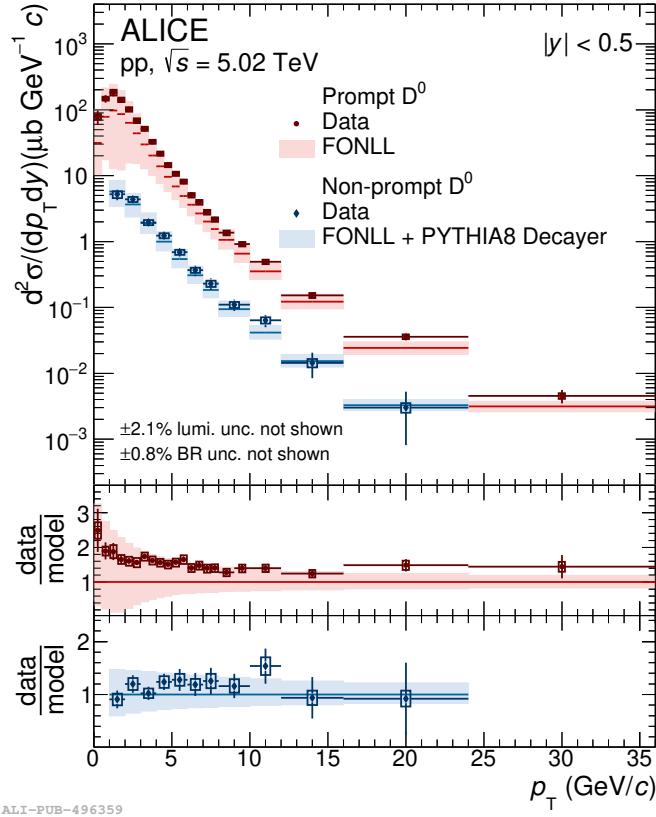


Figure 2.1:  $p_T$ -differential production cross-section of prompt and non-prompt  $D^0$ -mesons measured at midrapidity ( $|y| < 0.5$ ) in pp collisions by the ALICE Collaboration at  $\sqrt{s} = 5.02$  TeV [62] compared to predictions obtained with FONLL calculations [63] combined with PYTHIA 8 [64] for the  $H_b \rightarrow D^0 + X$  decay kinematics. Figure taken from Ref. [62]

extension of the form factors for elastic scattering. The first measurements of high-energy inclusive inelastic scattering experiments were performed using a 20 GeV linear accelerator at SLAC, and showed that the structure functions  $F_1(x, Q^2)$  and  $F_2(x, Q^2)$  were independent of  $Q^2$  at fixed  $x$  within the studied  $1 < Q^2 < 10 \text{ GeV}/c^2$  range. This was in contrast with the behavior observed for the proton elastic form factors, where a decrease of two orders of magnitude was observed within the same  $Q^2$  interval. This independence of the structure functions from  $Q^2$  in deep inelastic scatterings was predicted by Bjorken in 1968 for  $Q^2 \rightarrow \infty$  [68], and is known as *Bjorken scaling*. A physical interpretation of this phenomenon arrived just one year later, in 1969, with Feynman's parton model [69], which described the interaction in terms of elastic scattering of the probe off a point-like constituent (parton) within the proton. This model explains the scale-invariance property of the proton structure functions, as the scattering centres are assumed to be structure-less. In this picture, the Bjorken variable  $x$  acquires a new interpretation as the fraction of the proton momentum carried by the struck parton. The parton model also offers a straightforward definition of the structure functions in terms of the parton

distribution functions  $f_a(x)$ :

$$F_2(x, Q^2) = \sum_a e_a^2 x f_a(x) \quad ,$$

where the sum is over partons with electric charge  $e_a$ , and  $f_a$  are unknown, but universal functions for a given hadron, describing the probability of finding a parton of type a with a fraction  $x$  of the proton's momentum.

To explore the spin properties of the partons, the structure functions  $F_1$  and  $F_2$  were studied at different centre-of-mass energies. By investigating the relationship between the two structure functions, it was established that the partons have spin 1/2, as the Callan-Gross relation [70], which holds true for point-like Dirac particles, was found to be satisfied:

$$F_2(x, Q^2) = 2x F_1(x, Q^2) \quad .$$

In the next years, it became clear that additional constituents within the proton carry momentum but lack electric or weak charge, as the so-called momentum sum rule was not saturated by the measured PDFs in electron and neutrino scatterings. This missing momentum was attributed to gluons, which were discovered in the 1970s and are the field quanta of the strong force.

### Bjorken scaling violation

By the late 1970s, measurements of the structure functions at larger  $Q^2$  values taken at CERN and DESY revealed that Bjorken scaling was violated, i.e., the structure functions were not  $Q^2$  independent. Figure 2.2 shows measurements of the proton structure functions  $F_2(x, Q^2)$  as a function of  $Q^2$  for various values of  $x$  taken from different experiments [4]. It is clear from the plot that structure functions present an increasing trend as a function of  $Q^2$  at low  $x$ , and a decreasing trend as a function of  $Q^2$  at high  $x$ .

The parton model fails to explain this behaviour, as it relies on the assumption that the transferred energy is sufficiently large to neglect the proton and its constituents' masses, as well as the interactions among partons. In particular, the partons' transverse momentum with respect to the proton momentum is neglected. The key to understanding Bjorken scaling violation comes from QCD and the realisation that the parton's transverse momentum is not necessarily restricted to be small. A quark, for instance, can emit a gluon and acquire large transverse momentum  $k_T$  with a probability proportional to  $\alpha_s dk_T/k_T^2$  at large  $k_T$ . The integral extends up to the kinematic limit  $k_T \sim Q^2$ , giving rise to contributions proportional to  $\alpha_s \log Q^2$ , which break the scaling. The evolution of PDFs with  $Q^2$  from a parametrisation at a given  $Q_0^2$  can be perturbatively described using the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations [71, 72, 73], which require the introduction of a new arbitrary scale, at which the factorisation of the non-perturbative processes happens: the factorisation scale  $\mu_F$ . There exists a wide range of PDF parametrisations, such as the NNPDF [74], CTEQ [75], and MMHT [76], which are determined from global fits to a wide range of experimental data, including deep inelastic scattering, Drell-Yan, and jet production.

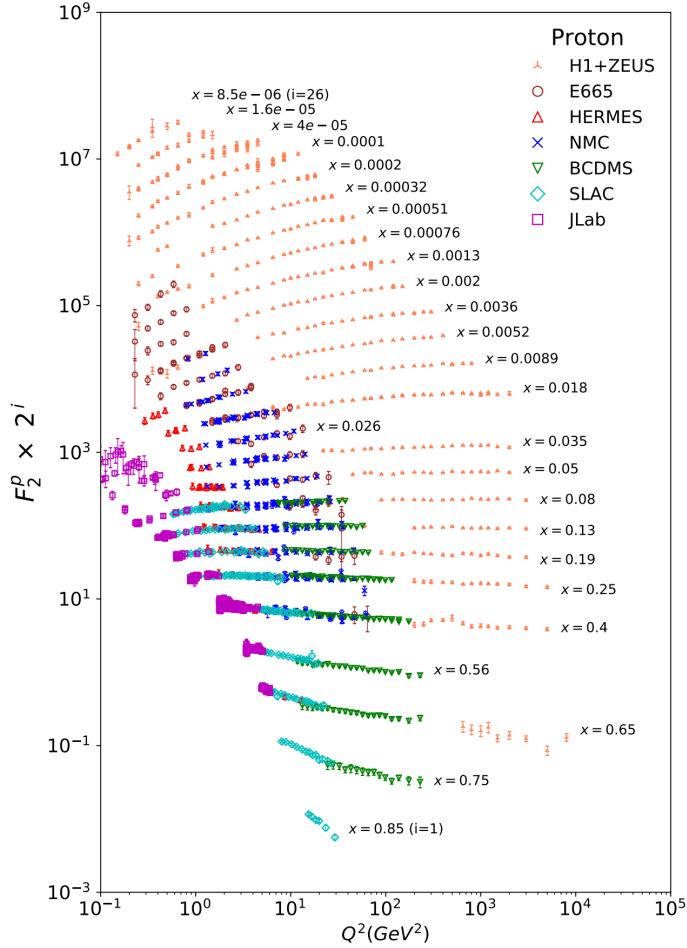


Figure 2.2: The proton structure function  $F_2^p$  measured in electromagnetic scattering of electrons and positrons on protons, and for electrons/positrons and muons on a fixed target [4].

### 2.1.2 Partonic cross-section

Because of their large masses, heavy quarks can only be produced through hard-scattering processes, characterised by momentum transfers of the order of  $Q^2 \geq 4m_{b,c}^2$ . In this regime, the strong coupling constant is significantly smaller than unity, allowing for the perturbative calculation of the heavy quark production cross-section from partonic scattering using QCD. While predictions at next-to-next-to-next-to-leading order ( $N^3\text{LO}$ ) are available for certain processes, such as Higgs production [77, 78], the current state-of-the-art calculations for charm-quark production are at next-to-leading order (NLO) with all-order resummation to next-to-leading logarithmic (NLL) accuracy in the limit where the  $p_T$  of a heavy quark is much larger than its mass [63]. The contributions arising at the NLO include 1-loop virtual corrections to the Born process and real emission of a gluon or a quark-antiquark pair, and are depicted in Fig 2.3. Next-to-next-to-leading order (NNLO) QCD radiative corrections to the production of bottom-quark pairs have also been computed [79].

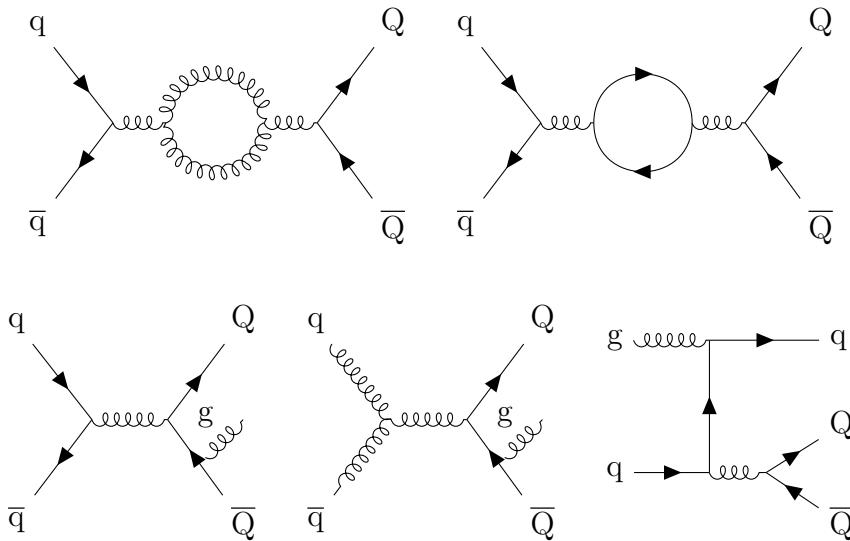


Figure 2.3: Feynman diagrams contributing to the first order corrections of the heavy-flavour production cross-section calculations.

### 2.1.3 Fragmentation Functions

Quarks and gluons produced in hard-scattering processes ultimately give rise to colourless observable hadrons. The associated process, known as hadronisation, is non-perturbative since the partonic processes involved are characterised by relatively small energy scales ( $\sim \Lambda_{\text{QCD}}$ ), where the strong coupling constant is large. The hadronisation process is described by the Fragmentation Functions (FFs)  $D_{Q \rightarrow H_Q}(z, \mu_F^2)$ , which parametrise the probability of the heavy quark  $Q$  to fragment into a heavy-flavour hadron  $H_Q$  with a fraction of the parent parton momentum  $z$ . They provide a simple “effective” description of the parton-to-hadron transition by mapping the quark-momentum-differential cross-section into the hadron-momentum differential one. FFs are typically determined from experimental data, usually by analyzing the final-state hadrons produced in electron-positron collisions where the initial momenta are well-known. These FFs are then applied in the evaluation of cross-sections in other colliding systems, assuming that the relevant hadronisation processes are “universal”, i.e., independent of the collision energy and system. This assumption implies the hypothesis of *independent fragmentation*, which states that the hadronisation of the parton produced in the hard-scattering process is independent of the production mechanism and the partonic environment which surrounds it. Differently from the fragmentation of light quarks and gluons, kinematic considerations on heavy-quark fragmentation [80, 81] suggest that heavy hadrons tend to carry a larger fraction of the parent parton momentum. Fragmentation functions for heavy quarks are therefore shifted towards higher values of  $z$  compared to those of light quarks and gluons [82]. Many FFs have been determined from global fits to data, such as the NNFF1.1h [83], DSS [84], and KKP [85] parametrisations for light flavours, and typically differ in the data sets used for the fit, the treatment of the data, and the functional form of the FFs. Parametrisations for heavy quark fragmentation include that of Peterson et al. [82], Kartvelishvili et al. [86], and

Breetan et al. [87]. Similarly to PDFs, FFs also evolve with the energy scale of the interaction, and this evolution is described perturbatively by the DGLAP equations.

## 2.2 Hadronisation: microscopic and macroscopic descriptions

Despite being a very powerful effective tool for describing heavy-flavour hadron production, the parametrisation of fragmentation functions used in a factorisation-theorem approach is not the only method used to describe the hadronisation of energetic quarks in high-energy collisions. Over the years, other models have been developed to describe hadron production starting from a partonic stage (e.g., quarks and gluons from the initial hard-scattering process, and those being radiated from the initial partons) and implementing a colour neutralisation procedure that groups quarks and gluons into hadrons.

The standard approach for describing the complex event topologies in these models, which are typically implemented in Monte Carlo event generators, begins with a matrix-element calculation for the production of a few well-separated partons, followed by the application of a parton shower. The *parton shower* provides an approximate perturbative treatment of QCD dynamics, which is then combined with a non-perturbative model for the hadronisation process at a certain infrared cut-off scale, typically taken to be of the order of 1 GeV. The basic idea of the parton shower relies on the Sudakov form factor [88], which expresses the probability of a parton not radiating another parton in a given phase space region. If a parton does radiate, the newly-produced parton becomes the source of a new cascade, continuing until the parton shower terminates at the hadronisation scale. At this point, partons are allowed to fragment into hadrons through hadronisation models.

Although several hadronisation models have been developed, each implementing a different approach to the description of this non-perturbative process, a common feature is the hypothesis of local parton-hadron duality [89], which states that the flow of momentum and quantum numbers at the hadron level tends to follow the flow established at the parton level.

### 2.2.1 Independent fragmentation

The description of the hadronisation process using FFs relies on the assumption of *independent fragmentation*, i.e., the probability of a parton fragmenting into a hadron is considered independent of the other partons produced in the same collision, effectively considering the hadronisation process as universal, i.e., independent of the collision energy and system. In the original scheme proposed by R. Field and R. Feynman [90], the fragmenting quark combines with an antiquark from a  $q\bar{q}$  pair produced from the vacuum to create a meson with energy fraction  $z$ . The remaining quark, with energy fraction  $(1 - z)$ , fragments in the same way, continuing until an energy cut-off is reached. The probability distribution of  $z$  is the fragmentation function. The assumption of independent fragmentation is valid in the case of low-multiplicity  $e^+e^-$  collision events, where the number of produced partons is small

and no hadronic remnants are present.

However, recent results on the production of strange and heavy-flavour baryons in proton-proton and proton-lead collisions at the LHC [91, 92] show that a description of the hadronisation process based on independent fragmentation is not sufficient to describe the data, as it significantly underestimates the baryon production. Figure 2.4 shows the  $\Lambda_c^+ / D^0$  production-yield ratio measured at midrapidity ( $|y| < 0.5$ ) in pp collisions at  $\sqrt{s} = 5.02$  TeV by the ALICE Collaboration [91] as a function of  $p_T$ , compared to theoretical predictions. The predictions are obtained from: i) PYTHIA 8 with Monash tune [93]; ii) HERWIG 7 [94]; iii) POWHEG [95] NLO pQCD calculations, matched with PYTHIA 6 to generate the parton shower; iv) General-Mass Variable-Flavour-Number Scheme [96] (GM-VFNS) NLO pQCD calculation with next-to-leading-log resummation. All these models, whose details are described in the following sections, implement fragmentation processes tuned on results of charm production measurements in  $e^+e^-$  collisions, and predict an almost  $p_T$ -independent  $\Lambda_c^+ / D^0$  ratio of around 0.1, significantly underestimating the values measured in pp collisions by a factor of about 5 at low  $p_T$ .

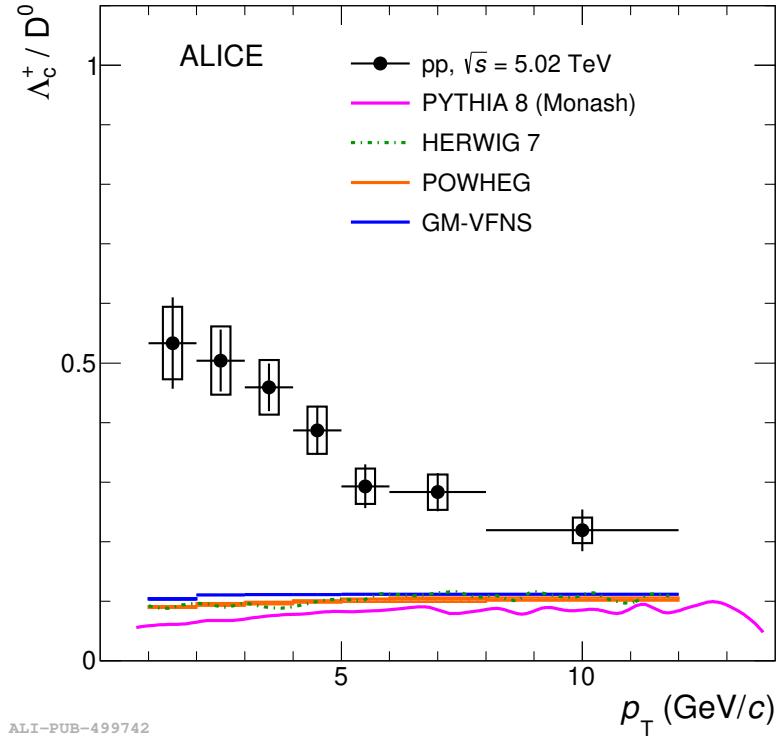


Figure 2.4:  $\Lambda_c^+ / D^0$  production-yield ratio measured at  $\sqrt{s} = 5.02$  TeV by the ALICE Collaboration as a function of  $p_T$ , compared to theoretical predictions. Figure taken from Ref. [91].

### 2.2.2 String model

The most widely used model for the description of the hadronisation process in Monte Carlo event generators is the Lund string model [97], employed in the PYTHIA

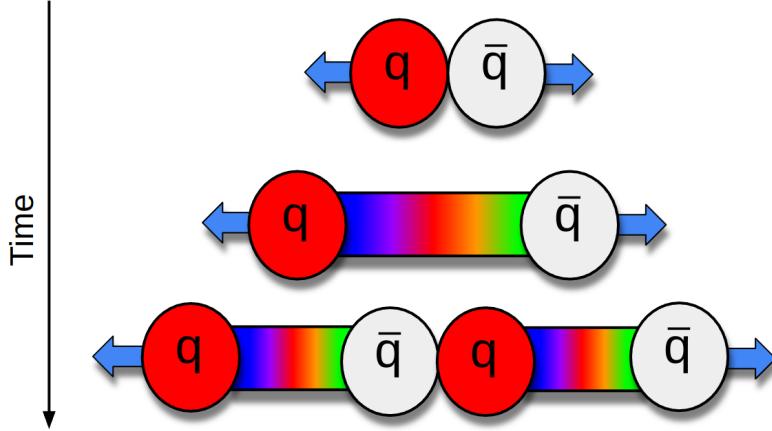


Figure 2.5: Schematic representation of the Lund string model hadronisation process.

event generator [98]. In this model, the strong force between quarks and gluons is modelled in terms of a colour string with energy given by the Cornell potential [99],

$$V(r) = -\frac{A(r)}{r} + \kappa r \quad .$$

Since the linear term is dominant, the Cornell potential is typically approximated as  $V(r) = \kappa r$ , with  $\kappa \sim 1$  GeV/fm. This implies that a constant force is exerted between the quarks, leading to a linear increase in the potential energy with increasing distance between the quarks. When back-to-back quarks are produced in a hard-scattering process, they move apart, causing the string to stretch and accumulate energy. When the energy stored in the string becomes large enough, it becomes energetically convenient to materialise a quark-antiquark pair with mass  $m_q$  from the colour flux tube, making the string break. The probability of string breaking is given by [100]

$$P \propto \exp\left(-\frac{\pi m_{T,q}^2}{\kappa}\right) = \exp\left(-\frac{\pi m_q^2}{\kappa}\right) \exp\left(-\frac{\pi p_{T,q}^2}{\kappa}\right) \quad . \quad (2.2)$$

This string-breaking process continues until the energy of the string is below the threshold for producing new quark-antiquark pairs, effectively modelling the parton fragmentation mechanism and applying the infrared cut-off introduced in Sec. 2.2.

The mass dependence of Eq. 2.2 leads to a Gaussian suppression factor that limits the probability for the colour field to produce quark-antiquark pairs for heavier flavours. The  $s\bar{s}$  and  $c\bar{c}$  pair production probability can be estimated with respect to that for  $u\bar{u}$  and  $d\bar{d}$  pairs [101], yielding

$$u : d : s : c \sim 1 : 1 : 1/3 : 10^{-11} \quad .$$

The production of  $c\bar{c}$  pairs in string-breaking (non-perturbative) processes results significantly suppressed, by a factor of  $10^{-11}$ , compared to  $u\bar{u}$  pairs. Such quarks are therefore almost exclusively produced in hard-scattering processes, quantitatively confirming the qualitative discussion in Sec. 2.1.2 on the feasibility of using perturbative QCD to describe the production of heavy-flavour quarks.

The fragmentation model described above shares many similarities with the independent fragmentation model discussed in Sec. 2.2.1, with the main difference being the pictorial description of interactions with a colour string connecting the quarks. However, the gluon radiation has been neglected until now. A significant difference emerges once the gluon bremsstrahlung is considered [102].

In this framework, gluons are represented as carrying both a colour and an anticolour charge. Therefore, in a three-parton system ( $qg\bar{q}$ ), the quark is colour-connected to the anticolour index of the gluon, and the colour index of the gluon is connected to the antiquark. The interaction between the quark-antiquark pair is suppressed by a factor  $1/N_c$ , where  $N_c$  is the number of colours, and can be neglected in a leading-colour approximation ( $N_c \rightarrow \infty$ ). The presence of the gluon produces a corner, or “kink”, on the string. The Lorentz boost of a string causes the hadrons formed through its breaking to go preferentially in the direction of its motion. Most hadrons that the  $q$ - $g$  string segment produces will go between the quark and the gluon, while hadrons from the  $\bar{q}$ - $g$  string will go between the gluon and the antiquark. Only very few hadrons will go between the quark and the antiquark. This behaviour leads to an angular distribution of hadrons in  $e^+e^-$  three-jet final states which differs from that predicted by independent fragmentation and is found to be in better agreement with experimental results.

### Non-independent fragmentation and multiple parton interactions

To fully describe hadron production in high-energy collisions, the description of the fragmentation process through colour-string breaking in a parton shower, coupled with a hadronisation step, is not sufficient. A definition of the ways in which the partons produced in the hard-scattering process can be connected to hadronise is also needed. At leading-colour approximation, in the partonic final states, each quark is colour-connected to a single other parton in the event. Since gluons carry both a colour and an anticolour charge, they are connected to two other partons. With this approximation, the Lund model provides a good description of measurements in  $e^+e^-$  collisions, where a parton-rich environment is not produced. However, recent results on the production of strange and heavy-flavour baryons in proton-proton and proton-lead collisions at the LHC [103, 92] show that a description of the hadronisation process based on independent fragmentation is not sufficient to describe the data, as it significantly underestimates the charm-baryon production.

In hadronic collisions, one must consider that to achieve a comprehensive description of a given process, a description of coloured initial-state partons and the associated beam remnants should be taken into account, as they hadronise and may potentially interact with each other. Furthermore, new insights on the underlying event and soft-physics processes occurring in a hadronic collision suggest that such events are dominated by Multiple Parton Interactions (MPI), where two or more distinct hard parton interactions occur in a single hadron-hadron collision. MPI produce densely-populated partonic environments in which colour string can be formed between partons from different hard-scattering processes. Phase-space overlaps among the partons produced in the MPIs become more likely as the number of MPI increases, and are expected to play a more important role in pp collisions with higher particle multiplicities. Their non-independent hadronisation should be

considered for a more complete description of the process.

New models based on colour-reconnection beyond the leading-colour approximation [104] have been developed, where the leading-colour connections produced in the partonic showers are rearranged to form new subleading topologies that would have been present in a full-colour treatment. New string configurations emerge in the beyond-leading-colour picture, as junction topologies and gluon loops are formed in addition to the standard  $q\bar{q}$  and  $q$ - $g$  strings. In particular, junctions allow for an enhanced production of baryons, to account for the observed enhancement in the production of baryons in proton-proton and proton-lead collisions. Different configurations for the colour reconnections (*CR modes*) are implemented in the PYTHIA event generator, applying different constraints on the possible colour reconstructions.

### 2.2.3 Cluster model

A different approach to the description of the hadronisation process is the cluster model [105], which is implemented in the HERWIG event generator [106]. It is based on the *preconfinement* of colour [107], which arises from the observation that by following the colour structure of the parton shower and studying the colour-singlet pairs of colour-connected quark-antiquark states, one finds that they tend to end up close in phase space. This suggests that quarks and gluons produced in this evolution become organised in clusters of colour singlets with finite masses, i.e., the mass distribution of these clusters is independent of the hard scattering process and its centre-of-mass energy. The confinement can then convert these singlets of small mass into hadrons.

The first step of the cluster hadronisation model is to non-perturbatively split the gluons left at the end of the parton shower into quark-antiquark pairs. Each gluon is allowed to decay into any of the accessible quark flavours with a probability given by the available phase space for the decay.

Then, the hadronisation process takes place. The key idea is based on the ansatz that the cluster mass spectrum is both universal and steeply falling at high masses. Therefore, the clusters can be regarded as highly excited hadron resonances and decayed, according to phase space, into the observed hadrons. Before the actual cluster decays, a few heavier clusters are split into lighter clusters (*cluster fission*), for a more reasonable agreement with experimental results. A cluster is split into two clusters if its mass,  $M$ , is such that

$$M^{\text{Cl}_{\text{pow}}} \geq \text{Cl}_{\text{max}}^{\text{Cl}_{\text{pow}}} + (m_1 + m_2)^{\text{Cl}_{\text{pow}}} \quad ,$$

where  $\text{Cl}_{\text{max}}$  and  $\text{Cl}_{\text{pow}}$  are parameters of the model, and  $m_{1,2}$  are the masses of the constituent partons of the cluster. For clusters that need to be split, a  $q\bar{q}$  pair is produced from the vacuum. Only up, down, and strange quarks are chosen with probabilities given by other model parameters. Once a  $q\bar{q}$  pair is produced, the cluster is split into two new clusters with one of the original partons in each cluster.

Finally, the cluster is decayed into a pair of hadrons. For a cluster of a given flavour  $(q_1, \bar{q}_2)$ , a quark-antiquark or diquark-antidiquark pair is extracted from the vacuum and a pair of hadrons with flavours  $q_1$  and  $q_2$  is formed. The hadrons are selected from all the possible hadrons with the appropriate flavour based on

the available phase space, spin and flavour. As a consequence, heavier hadrons are suppressed, leading to a natural description of the baryon and strangeness suppression. The cluster model was found to describe data reasonably well, with far fewer parameters than the string model [108].

### 2.2.4 Coalescence model

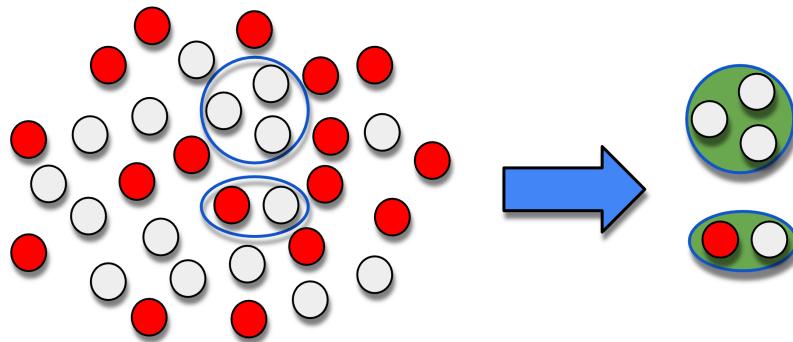


Figure 2.6: Representation of hadron production via recombination.

As outlined in Sec. 1.3.4, the hadronisation process can be modified in the presence of a strongly-interacting deconfined medium. A novel mechanism for the production of hadrons, the recombination (also called coalescence), can take place in a thermalised system such as the QGP. When recombination occurs, two or three quarks close in the space-velocity phase space can bind together to form a hadron, which will have a higher (transverse-) momentum than the quarks that combined to form the hadron. This is different from the case of fragmentation, in which an energetic parton produces hadrons with lower momentum than the parent quark. As a consequence, the  $p_T$  distribution of hadrons formed by coalescence differs from the one expected from fragmentation. In particular, the hadron  $p_T$  is shifted to larger values as compared to fragmentation. Due to the steeply-falling quark  $p_T$  spectrum, this leads to a depletion of the low- $p_T$  hadron yield, and an enhancement of the intermediate- $p_T$  hadron yield. In addition, due to the relatively high abundance of low- $p_T$  quarks and the sparsity of high- $p_T$  partons, an enhancement of baryon-to-meson production-yield ratios at intermediate  $p_T$  is expected in proton-proton collisions when compared to  $e^+e^-$  collisions as a natural consequence of the recombination mechanism.

Models implementing the coalescence mechanism can effectively describe the production of hadrons in heavy-ion collisions. Intriguingly, models implementing recombination that successfully describe the production and azimuthal anisotropy of charm hadrons in these collisions, fail in doing so when this process is de-activated, as shown in Fig. 2.7 for the case of prompt D mesons in the heavy-flavour sector. The D-meson  $R_{AA}$  measured in the 0–10% centrality interval and the second coefficient  $v_2$  of the Fourier transform of the azimuthal distribution of D mesons in the 30–50% centrality range measured by the ALICE Collaboration at midrapidity ( $|y| < 0.5$  for the  $R_{AA}$  measurement and  $|y| < 0.8$  for that of the  $v_2$  coefficients) at  $\sqrt{s_{NN}} = 5.02$  TeV [109] are illustrated in the left and right panels, respectively. The

measurements are compared to the Parton-Hadron-String Dynamics [110] (PHSD), POWLANG [111], and DAB-MOD [112] predictions, which are provided with (solid line) and without (dashed line) the implementation of the recombination process in the hadronisation mechanism. The calculations performed including the recombination of charm quarks with light quarks provide a good description of the experimental results, while they fail to reproduce the data when recombination is turned off. This suggests that the recombination mechanism plays a relevant role in the description of heavy-flavour hadron production in heavy-ion collisions.

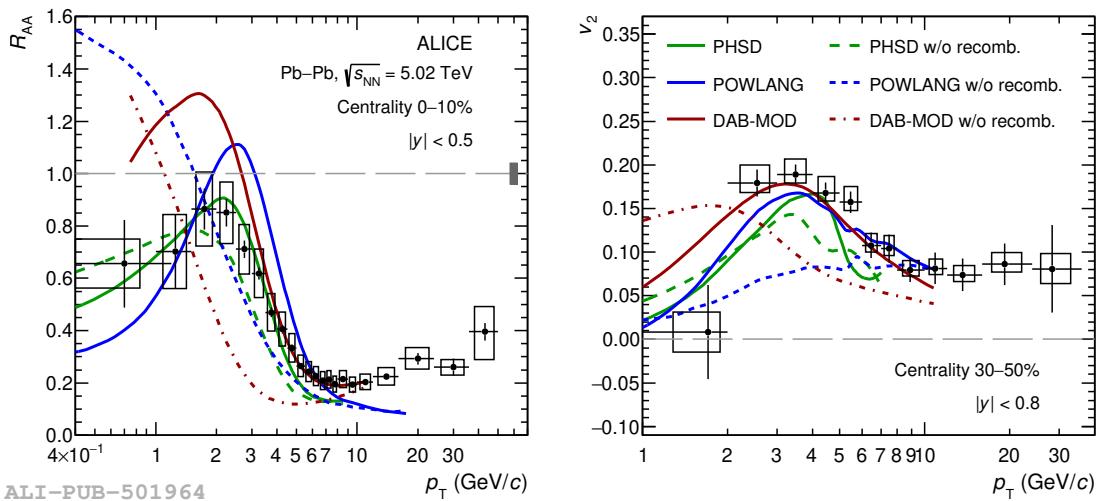


Figure 2.7: Prompt D-meson  $R_{AA}$  in the 0–10% centrality class (left panel) and  $v_2$  in the 30–50% centrality class (right panel) compared with predictions obtained with and without including hadronisation via recombination. Figure taken from Ref. [109].

New experimental measurements in pp and p–Pb collisions at the LHC [92, 57, 91, 113, 114, 115] provided a wealth of results suggesting that certain phenomena that are observed in nuclear collisions, such as strangeness enhancement and flow, may also occur in smaller collision systems. These effects are typically related to the presence of an expanding medium, thus raising the question of whether the coalescence hadronisation mechanism could also play a role in small collision systems. A hint that this could be the case comes from the observation that the production of charm baryons in pp and p–Pb collisions [103, 92] is significantly larger than expected from models based on a fragmentation tuned on  $e^+e^-$  collisions.

Various models have been developed in order to describe the observed charm-baryon enhancement in proton-proton and proton-lead collisions with the assumption of a recombination mechanism. These include the Quark (re-)Combination Mechanism [116] (QCM) model, where coalescence between a charm quark and equal-velocity light quarks takes place at all momenta, and thermal weights are applied to account for relative production of scalar and vector mesons. In this model, the quark spectra are parametrised based on measurements of light-flavour hadrons and D-mesons. Other models adopt a description of the expanding system using a

hydrodynamic approach and implementing hadronisation of the medium via a combination of coalescence and independent fragmentation. The Catania coalescence model [117] describes a thermalised system of u, d, s quarks, and gluons, where the charm quark can hadronise via either fragmentation or coalescence with light quarks from the bulk. The POWLANG model [118] predicts the formation of a small, deconfined, and expanding fireball in proton-proton collisions, where charm quarks are subject to rescattering and hadronisation, and can recombine with light quarks as in heavy-ion collisions.

Each of these models provides a different hadronisation mechanism in proton-proton collisions compared to  $e^+e^-$  ones, and independent hadronisation is no longer assumed. They improve the description of the data, achieving a better agreement with experimental results such as the  $p_T$  dependence of  $\Lambda_c^+/D^0$  and  $\Xi_c^0/D^0$  baryon-to-meson production-yield ratios.

### 2.2.5 Core-corona model

Core-corona models [119] offer an intermediate approach between the two presented above. In these models, the fireball is divided into a high-density core part, treated with hydrodynamics and thermal modelling of particle production, similarly to that implemented in the SHM introduced in Sec. 1.3.2, and a corona part, described using the Lund string model. This framework is based on the assumption that small QGP droplets may only be produced once the density is high enough. The EPOS Monte Carlo event generator [120] implements such a picture, and provides a very competitive description of data.

### 2.2.6 Statistical hadronisation model

The last model presented in this section exploits the idea of a thermalised system, which is by definition in thermal equilibrium. A macroscopic approach to hadronisation, based on a statistical model, can be used to describe the charm-hadron production in hadronic collisions. The principles of thermodynamics can be applied to the hadronisation process, where the hadron yields are determined by the hadron’s mass and the temperature and baryon chemical potential of the system. As described in Sec. 1.3.2, this approach has been successfully adopted to describe the light and strange hadron production in heavy-ion collisions and smaller systems.

Different approaches are used to predict the production of light- and heavy-flavour hadrons with the SHM. Due to the large masses of charm and beauty quarks, which are significantly larger than the temperature of the QGP, their thermal production in the medium is strongly suppressed. Therefore, the production of charm and beauty quarks is determined by their initial production in the hard-scattering processes. Their distribution into hadrons is then predicted with thermal weights. A charm balance equation is used to ensure the conservation of the initially produced charm quarks into the final-state hadrons via a fugacity factor  $g_c$ , which accounts for the absence of chemical equilibrium for heavy quarks. In addition, the description of heavy-flavour hadron production and the baryon enhancement in this framework requires the inclusion of a strong feed-down from an augmented set of

excited charm- or beauty-baryon states, not included in the PDG [4]. Recent model developments [121, 122] have extended the Statistical Hadronisation Model (SHM) to include a set of 18  $\Lambda_c$ 's, 42  $\Sigma_c$ 's, 62  $\Xi_c$ 's, and 34  $\Omega_c$ 's excited states up to a mass of 3.5 GeV, predicted by the Relativistic Quark Model [123]. Such models provide a good description of the measured  $\Lambda_c^+/\bar{D}^0$  production yield ratio and the  $p_T$  spectra of D mesons and  $\Lambda_c^+$  at midrapidity.

### 2.2.7 Baryon enhancement: $\Lambda_c^+/\bar{D}^0$ ratio

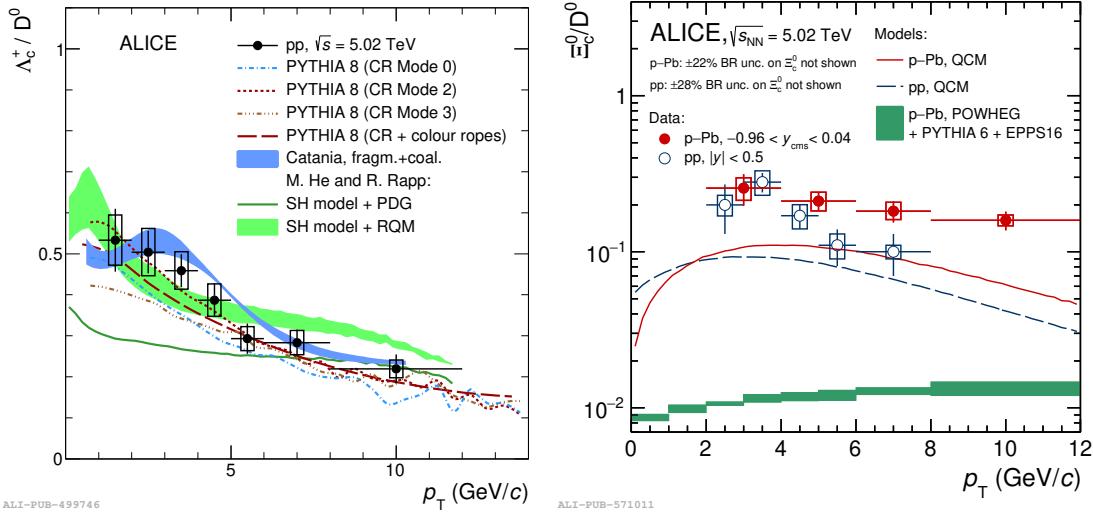


Figure 2.8: Left panel:  $\Lambda_c^+/\bar{D}^0$  production-yield ratio measured at midrapidity ( $|y| < 0.5$ ) in pp collisions at  $\sqrt{s} = 5.02$  TeV by the ALICE Collaboration [91] as a function of  $p_T$ , compared to theoretical predictions. Right panel:  $\Xi_c^0/\bar{D}^0$  production-yield ratio measured at midrapidity ( $-0.96 < y < 0.05$ ) in p–Pb collisions at  $\sqrt{s_{\text{NN}}} = 5.02$  TeV by the ALICE Collaboration [91] as a function of  $p_T$ , compared to theoretical predictions. Figures taken from Refs. [91] and [92].

The hadronisation process can be experimentally studied through the measurement of production-yield ratios of hadrons. Since the initial state of the collision and the charm production cross-section is the same for all charm hadrons, the measurement of the production-yield ratio of different charm hadrons can be expressed, using the factorisation theorem, as the ratio of the fragmentation functions of the hadrons, which describe the hadronisation process.

The left panel of Fig. 2.8 shows the  $\Lambda_c^+/\bar{D}^0$  baryon-to-meson production-yield ratio measured at midrapidity ( $|y| < 0.5$ ) in pp collisions at  $\sqrt{s} = 5.02$  TeV by the ALICE Collaboration as a function of  $p_T$ , compared to models that include mechanisms to enhance baryon production compared to those in which the hadronisation is implemented through independent fragmentation and parametrised on results from  $e^+e^-$  collisions, already presented in Fig. 2.4. They include i) PYTHIA 8 simulations with colour-reconnection beyond the leading-colour approximation [104] (three colour reconnection modes (0,2,3), which apply different constraints on the allowed reconnection are considered); ii) PYTHIA 8 with colour-reconnection plus

rope hadronisation [124] where colour charges can act coherently to form a rope, increasing the effective string tension; iii) Catania coalescence model [117], where a QGP is formed in pp collisions and hadronisation occurs through both recombination and fragmentation; iv) SHM [27] where the underlying charm baryon spectrum is either taken from the PDG [4], or augmented to include additional excited baryon states, which have not yet been observed but are predicted by the RQM [123]. These models are capable of describing both the magnitude and the  $p_T$  dependence of the  $\Lambda_c^+/\bar{D}^0$  ratio, suggesting that the hadronisation process in proton-proton collisions might differ from that in  $e^+e^-$  collisions. This conclusion is also confirmed by the measurement of the  $\Xi_c^0/\bar{D}^0$  baryon-to-meson production-yield ratio at midrapidity ( $-0.96 < y < 0.04$ ) in p–Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV performed by the ALICE Collaboration [92], shown in the right panel of Fig. 2.8. The measured ratio is compared to POWHEG [95] NLO pQCD calculations, matched with PYTHIA 6 [125] to generate the parton shower and fragmentation, which is tuned on  $e^+e^-$  collisions, and to the EPPS16 parametrisation of the modification of the PDFs in the nuclear environment of the lead ion [126]. It predicts a slightly increasing trend with  $p_T$ , and undershoots the measured  $\Xi_c^0/\bar{D}^0$  by a large factor of around 20. Predictions from the QCM model [116] implementing a coalescence hadronisation mechanism provide a better description of both the  $p_T$  dependence and the magnitude of the data.

The influence of the surrounding environment in the hadronisation of the charm quark could also potentially affect the relative abundances of strange and non-strange charm hadrons with increasing multiplicity from  $e^+e^-$  collisions to high multiplicity proton-proton and proton-nucleus collisions. Enhanced production of charm-strange hadrons may result from the coexistence of numerous strange quarks (which could be thermally produced in an environment with  $T \gg m_s$ , as explained in Sec. 1.3.5) within the same region as the charm quark, thereby increasing the probability of recombination between them. This phenomenon can be studied through the measurement of the production-yield ratio of charm-strange hadrons to non-strange charm hadrons, such as the  $D_s^+$ -meson to  $D^+$ -meson ratio, which constitutes the primary focus of this Thesis.

# Chapter 3

## A Large Ion Collider Experiment

ALICE (A Large Ion Collider Experiment) is a general-purpose experiment at the CERN LHC [127]. It was conceived and built to focus on the study of QCD and heavy-ion collisions. It is designed to address the physics of strongly interacting matter and the quark-gluon plasma at extreme values of energy density and temperature in nucleus-nucleus collisions. Nonetheless, its physics programme also includes proton-proton and proton-nucleus collisions, which provide reference data for the heavy-ion programme and address a number of specific strong-interaction topics for which ALICE is complementary to the other LHC experiment. During the LHC Long Shutdown 2, between 2019 and 2021, ALICE underwent a major upgrade to improve its capabilities to probe the QGP with heavy-flavour quarks, and to enable completely new measurements of the thermal emission of dielectron pairs [128].

### 3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a two-ring superconducting hadron accelerator and collider based at CERN [129], near Geneva, on the border between Switzerland and France. The LHC is the world’s largest circular particle accelerator with a 26.7-kilometre circumference housed in a 3.8-metre-wide tunnel buried 50 to 175 metres underground, capable of accelerating protons up to energies of  $\sqrt{s} = 14$  TeV and heavy ions (e.g.,  $^{208}\text{Pb}^{82+}$ ) up to a centre-of-mass energy per nucleon pair of  $\sqrt{s_{\text{NN}}} = 5.5$  TeV.

The LHC, which is located in the same tunnel previously used to host the Large Electron-Positron Collider [130] (LEP), is the final stage of a chain of accelerators that provide protons and heavy ions with sufficient energy to be injected into the LHC, illustrated in Fig. 3.1. Protons are firstly collected by stripping electrons from a gaseous hydrogen source with an electric field and then grouped into bunches. The Linear accelerator 4 (Linac 4), the first linear accelerator in the LHC injection chain, accelerates negative hydrogen ions to 160 MeV. The ions are then stripped of their two electrons, and protons are subsequently injected into the Proton Synchrotron Booster (PSB), which accelerates them to 1.4 GeV, and then into the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS), reaching 25 and 450 GeV, respectively. The protons are finally injected in the LHC, where they are accelerated

up to a maximum energy of 7 TeV. In the LHC, the beams are forced to follow the circular path of the ring by strong magnetic fields ( $\sim 8$  T) produced by superconducting dipole magnets. Other magnets are used to steer and focus the beams in the eight different interaction points.

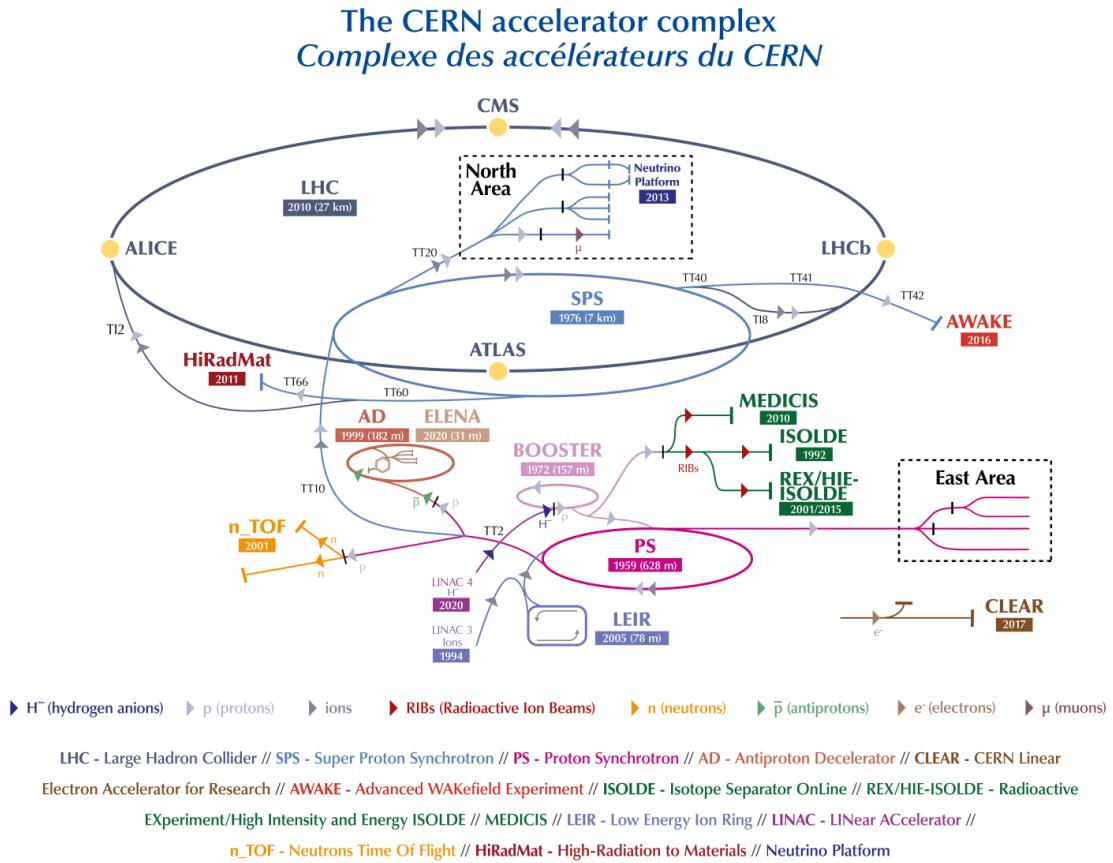


Figure 3.1: The Large Hadron Collider (LHC) accelerator complex at CERN. Figure taken from Ref. [131].

Heavy-ion acceleration follows a similar path. Pb ions ( $\text{Pb}^{25+}$  –  $\text{Pb}^{28+}$ ) are extracted in the Electron-Cyclotron Resonance (ECR), and are focused and accelerated in a Radio-Frequency Quadrupole (RFQ) using only electric fields.  $\text{Pb}^{28+}$  ions with momentum of  $2.5 \text{ keV}/c$  are selected with a spectrometer, and are accelerated to  $4.2 \text{ MeV}/c$  in the Linear accelerator 3 (Linac 3). A first  $0.5 \mu\text{m}$  copper stripping foil is then used to increase the charge state of the ions to  $\text{Pb}^{53+}$ . The Low Energy Ion Ring (LEIR) groups the ions into bunches and accelerates them to 72 MeV. The PSB and PS then accelerate the ions to 95.4 MeV and 5.09 GeV. After the PS, a copper stripper ( $\sim 1 \text{ mm}$ ) is used to further ionise the Pb ions to  $\text{Pb}^{82+}$ , which are then accelerated to 158 GeV in the SPS. The ions are finally injected into the LHC, where they are accelerated up to 2.76 TeV per nucleon.

The LHC can collide protons with protons, protons with heavy ions, and heavy ions with heavy ions. The LHC has four main interaction points, where the four main experiments are located: A Large Ion Collider Experiment (ALICE),

A Toroidal LHC ApparatuS (ATLAS), Compact Muon Solenoid (CMS), and LHC-beauty (LHCb). The four experiments were conceived and built to address different physics topics: ATLAS and CMS are general-purpose detectors designed to study the Higgs boson, which they discovered in 2012 [132, 133] and to search for physics beyond the Standard Model; LHCb is dedicated to the measurement of beauty quarks as proxy for CP violation studies, and to the study of matter-antimatter asymmetry. ALICE is the only detector at the LHC that is dedicated to the study of QCD rather than the electroweak sector of the Standard Model. A detailed description of the ALICE apparatus is given in the following sections.

## 3.2 The ALICE experiment

The physics programme of ALICE was designed to study the properties of strongly interacting matter at the extreme values of energy density and temperature reached in ultra-relativistic heavy-ion collisions. This unique environment poses several challenges, the most stringent one being the need to carry out measurements in a very high multiplicity environment. Originally, estimates for the charged particle multiplicity density at mid-rapidity in central Pb–Pb collisions ranged from  $dN/d\eta = 2000$  up to almost  $dN/d\eta = 8000$ . Detectors with high granularity, fast readout capabilities, and high radiation hardness were therefore needed. The ALICE detector was designed to meet these requirements, and to provide a comprehensive set of measurements to study the properties of the QGP. The ALICE apparatus is shown in Fig. 3.2. It is based on a central barrel, located inside a magnet which provides a magnetic field of  $\sim 0.5$  T, covering full azimuth ( $0 < \varphi < 2\pi$ ) and the pseudorapidity region  $|\eta| < 0.9$ , and a forward muon system with a dipole magnet providing a total bending power of  $B\rho = 3$  Tm, covering full azimuth and pseudorapidity interval  $-4.0 < \eta < -2.5$ . The pseudorapidity  $\eta$  is defined from the polar angle  $\theta$  as  $\eta \equiv -\log(\tan(\theta/2))$ .

The ALICE coordinate system is a right-handed system with the  $z$ -axis pointing along the beam direction, in the direction away from the muon arm, the  $y$ -axis pointing vertically up, and the  $x$ -axis pointing horizontally towards the center of the LHC. The nominal interaction point is the origin of the coordinate system. The two sides of the apparatus along the beam axis are referred to as the C-side, where the muon arm is positioned, and the A-side, where the FV0 is positioned. The polar angle  $\theta$  is defined with respect to the  $z$ -direction, while the azimuthal angle  $\varphi$  increases counter-clockwise starting from the  $x$ -axis towards the CMS side.

The central barrel is enclosed in the L3 solenoid, which has an internal length of 12.1 m and a radius of 5.75 m, providing a magnetic field of 0.5 T. The L3 experiment was a detector at the Large Electron-Positron Collider (LEP) at CERN, which was built in the cavern where ALICE is now located, operating from 1989 to 2000. The central barrel detector system is designed for efficient tracking in the high track-density environment of heavy-ion collisions, covering transverse momenta from  $\sim 100$  MeV/ $c$  to  $\sim 100$  GeV/ $c$  with excellent hadron and electron identification capabilities. It also reconstructs primary and secondary vertices. This allows for precise measurements of ground-state heavy-flavour hadrons, which typically decay at a distance of  $\sim 100$   $\mu$ m from the primary vertex. The L3 magnet hosts the Inner Tracking

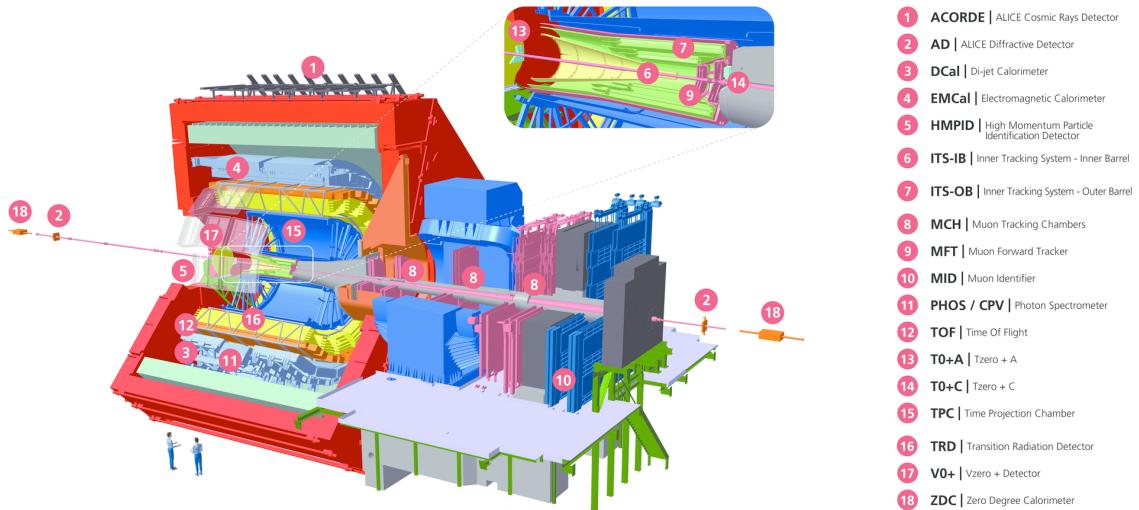


Figure 3.2: The ALICE experimental apparatus. The top right panel shows a zoom of the ITS, FT0-A, FT0-C, and MFT detectors. Figure taken from the ALICE figure repository [60].

System (ITS), the Time Projection Chamber (TPC), the Transition Radiation Detector (TRD), the Time-Of-Flight (TOF) detector, the High-Momentum Particle Identification Detector (HMPID), the ElectroMagnetic Calorimeter (EMCal), and the PHOton Spectrometer (PHOS). The central barrel also includes forward detectors such as the Muon Forward Tracker (MFT) and the Fast Interaction Trigger (FIT). The muon arm covers the forward pseudorapidity range  $-4.0 < \eta < -2.5$  and consists of absorbers, a large dipole magnet, planes of triggering Resistive Plate Chambers (RPC) and tracking MultiWire Proportional Chambers (MWPC).

In the following sections, the main components of the ALICE experiment used for the measurements presented in this Thesis are described in detail, ordered from the innermost layer to the outermost detector, following the path of a particle produced in a hadronic collision at mid-rapidity. The upgrades of the ALICE detectors during the LHC Long Shutdown 2 will be highlighted.

### 3.2.1 ALICE upgrades overview

The ALICE detector underwent a major upgrade during the LHC Long Shutdown 2, between 2019 and 2021, which renewed the experiment, now referred to as ALICE 2. The upgrade aimed at significantly improving the capabilities of ALICE to probe the QGP with heavy-flavour quarks, and enabling new measurements of the thermal emission of dielectron pairs. To achieve these goals, the ALICE detector was improved by enhancing the tracking capabilities of the central barrel detectors and increasing the readout rate of the detectors to collect larger data samples.

The Inner Tracking System (ITS) was replaced with a new detector, the ITS 2, a thinner and lighter detector with the first layer closer to the interaction point, which improves the pointing resolution up to a factor of 3 in the transverse direction and a factor 5 in the longitudinal direction as compared to the old ITS, which operated from 2009 to 2018, as shown in Fig. 3.3.

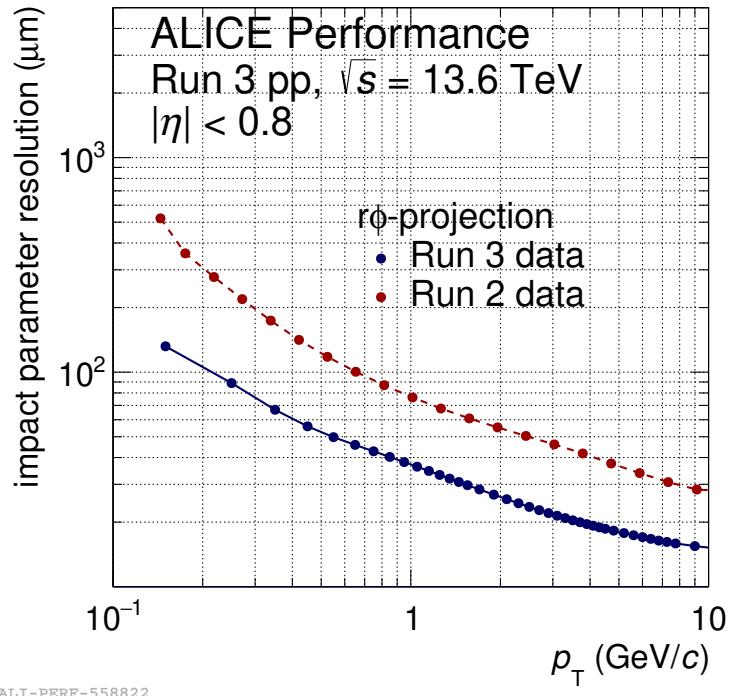


Figure 3.3: Impact parameter resolution in  $r\phi$  as a function of  $p_T$  in pp collisions at  $\sqrt{s} = 13.6 \text{ TeV}$  from Run 3 data compared with the same quantity measured in collisions at  $\sqrt{s} = 13 \text{ TeV}$  from Run 2 data. Figure taken from ALICE figure repository [60].

The upgrade strategy for the enhanced readout rate is based on the LHC plans to increase the luminosity of Pb–Pb collisions progressively after the LHC Long Shutdown 2, eventually reaching an interaction rate of about 50 kHz (from less than 1 kHz during LHC Run 1 and Run 2 data-taking periods), i.e. instantaneous luminosity of  $\mathcal{L} = 6 \times 10^{27} \text{ cm}^{-2} \text{ s}^{-1}$ . At these high interaction rates, each TPC drift time period of  $\sim 100 \mu\text{s}$  will contain on average 5 Pb–Pb events. It was therefore decided to use a continuous, untriggered readout strategy. With the MWPC used in ALICE 1 for the TPC readout, the ion backflow into the drift region was suppressed by active gating, limiting the readout rate to about 700 Hz for Pb–Pb collisions. This is overcome in ALICE 2 by using a readout based on Gas Electron Multiplier [134] (GEM) foils, which reduces the ion backflow and resulting space charge in the TPC to a level that can be corrected for while operating the detector with Pb–Pb interaction rates up to 50 kHz.

To synchronise the continuous data stream across all readout and processing branches, the data stream is divided into time frames (TF) of nominal length of 128 LHC orbits ( $\sim 11 \text{ ms}$ ). Each TF is subdivided into heartbeat frames (HBF) with a length corresponding to an orbit of  $\sim 89.4 \mu\text{s}$ . The detector data are time stamped with a precision of an LHC bunch crossing of 25 ns.

A new software framework has been developed to allow for distributed and efficient processing of this unprecedented amount of data. Because of the continuous

readout in Run 3, the vertex-to-track association is no longer unambiguous. Therefore, in place of the analysis data model used in Runs 1 and 2, based on the hierarchical structure of the "event content", the analysis data model for Run 3 is based on a columnar data format, where collisions and tracks are represented as separate tables, connected by an index. This maps naturally on a "flattened" structure-of-arrays data format. Hence, the columnar data format provided by Apache Arrow [135] was chosen. The new Online-Offline ( $\mathcal{O}^2$ ) framework ensures a unified and coherent computing environment from data taking up to analysis. Throughout all stages of the data processing, the timeframe represents the minimal processing unit.

The result of asynchronous reconstruction is the Analysis Object Data (AOD) format, with the best calibration available. The content of an AOD timeframe is kept contiguously in shared memory allowing efficient application of parallel execution and pipelining. AODs are stored as ROOT [136] TTrees, exploiting their columnar storage format, and are then used to perform physics analyses.

### 3.2.2 Inner Tracking System

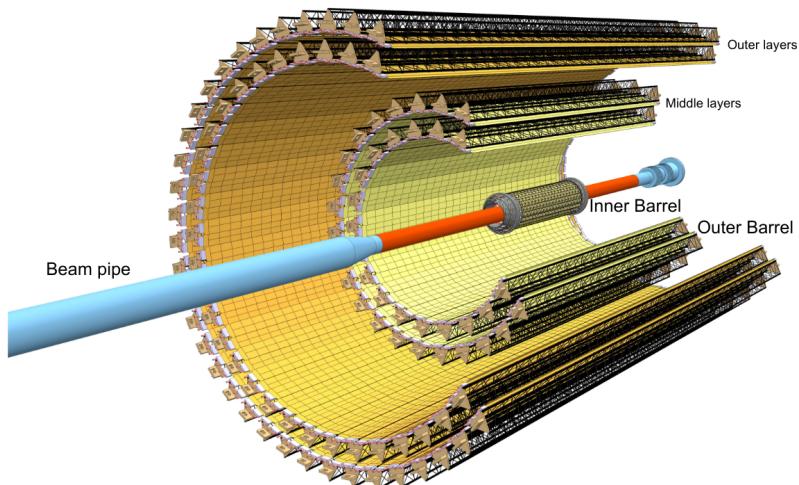


Figure 3.4: Schematic view of the ALICE ITS 2. Figure taken from Ref. [128].

The Inner Tracking System (ITS) is the innermost detector of the ALICE apparatus, and is designed to provide precise tracking and vertexing capabilities in the high-multiplicity environment of heavy-ion collisions. The ITS upgrade is one of the key improvements of the ALICE 2 detector. The ITS 1 was a six-layer silicon detector, with two layers of Silicon Pixel Detectors (SPD), two layers of Silicon Drift Detectors (SDD), and two layers of Silicon Strip Detectors (SSD). The ITS 1 has been replaced by the ITS 2, which is fully constructed with ALPIDE [137] chips, i.e., Monolithic Active Pixel Sensors [138] (MAPS) implemented in a 180 nm CMOS technology for imaging sensors provided by TowerJazz [139].

MAPS are an evolution of the hybrid pixel sensors which have been widely used in the past by several experiments [127, 140, 141, 142]. Hybrid pixel sensors are made of an active layer, typically made of silicon, and a readout layer with the front-end electronics, connected to the active layer with bump-bonds. On top of being

expensive and complex, the bump-bonding operation introduces additional material in the detector, therefore increasing the material budget of the detector. MAPS, on the other hand, are monolithic sensors, where the active layer and the front-end electronics are integrated into the same silicon wafer. This significantly reduces the material budget of the detector. The ALPIDE chip is a  $15\text{ mm} \times 30\text{ mm}$  chip, with a pixel size of  $29.24\text{ }\mu\text{m} \times 26.88\text{ }\mu\text{m}$ , and a total of  $512 \times 1024$  pixels. Each pixel cell contains a sensing diode, a front-end amplifier, a shaping stage, a discriminator, and a digital section. The digital section includes a multi-event buffer with three hit storage registers and a pixel mask register. At the time of writing this Thesis, the ITS 2 is the largest-scale application of MAPS in a high-energy physics experiment.

The ITS 2 brought a plethora of improvements: i) a new beryllium beam pipe with an outer radius reduced from 28 mm to 18 mm allows for an innermost detector layer closer to the interaction point, from 39 mm to 22.4 mm; ii) an increased granularity for all layers, which are now silicon pixel detectors with a cell size of  $29.24\text{ }\mu\text{m} \times 26.88\text{ }\mu\text{m}$ ; iii) the number of layers for the inner barrel was increased from two to three, raising the total number of layers from six to seven; iv) the material budget was reduced to  $0.36\% X_0$  ( $1.10\% X_0$ ) per layer for the innermost (outer) layers. A schematic view of the ITS 2 is presented in Fig. 3.4, while the main layout parameters are summarised in Table 3.1.

Table 3.1: Main layout parameters of the new ITS2. Table taken from Ref. [128].

Layer no.	Average radius (mm)	Stave length (mm)	No. of staves	No. of HICs/ stave	Total no. of chips
0	23	271	12	1	108
1	31	271	16	1	144
2	39	271	20	1	180
3	196	844	24	8	2688
4	245	844	30	8	3360
5	344	1478	42	14	8232
6	393	1478	48	14	9408

### 3.2.3 Time Projection Chamber

The Time Projection Chamber (TPC) is the main tracking detector of the ALICE experiment, and is designed to provide particle identification via  $dE/dx$  of charged particles on a track-by-track basis up to  $p_T \sim 1\text{ GeV}/c$  and precise momentum measurements. The TPC is a large cylindrical detector with a length and outer diameter of about 5 m, resulting in a volume of  $88\text{ m}^3$  filled with a gas mixture of Ne-CO<sub>2</sub>-N<sub>2</sub> (90-10-5). The choice of a neon-based gas mixture in place of an argon-based one, used during the LHC Run 2 data-taking period to guarantee the stability of the readout chambers [143], was made to reduce the space-charge distortions in the drift volume. The higher ion mobility of the Ne-based mixture compared to similar Ar-based mixtures directly reduces the magnitude of the space-charge distortions by

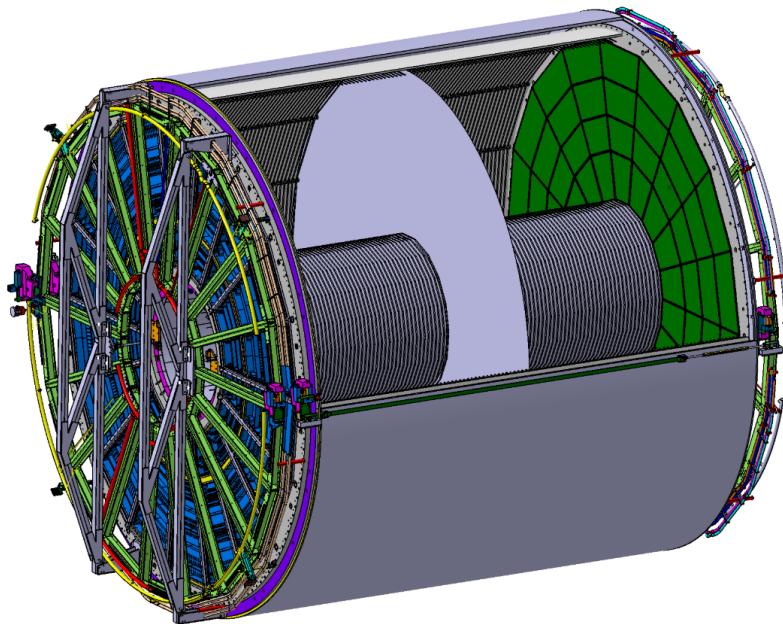


Figure 3.5: Schematic view of the ALICE TPC. Figure taken from Ref. [128].

nearly a factor of two [144]. The TPC covers a symmetric pseudorapidity interval around midrapidity ( $|\eta| < 0.9$ ) at full azimuth. The field cage has a high-voltage (100 kV) electrode in its center, which divides the active volume into two halves. The inner diameter of the central field cage drum is 114 cm, providing the necessary space for the installation of the ITS. Each of the two endplates houses 18 inner and outer readout chambers (IROCs and OROCs), which are arranged into pairs to form 18 equal azimuthal sectors.

During Runs 1 and 2, the readout chambers were based on MWPCs, which have to be operated with an active ion gating grid in order to prevent the ions produced in the amplification region from reaching the active drift volume, which could lead to space-charge distortions. The ambitious physics program of the ALICE experiment for Run 3 requires the elimination of the intrinsic trigger rate limitation of about 3 kHz of the original MWPC-based TPC, imposed by the operation of the active ion gating grid.

Operating the TPC at a collision rate of 50 kHz implies that on average five collision events pile up within the TPC readout time window of about  $\sim 100 \mu\text{s}$ . This excludes triggered operation and defines the need for continuous readout, requiring the exploitation of gas amplification techniques capable of providing sufficient ion blocking without an active gate. At the same time, the readout system must ensure that the  $dE/dx$  resolution of the TPC is preserved.

GEMs represent a valid solution to overcome the limitations of MWPCs. Sufficient ion blocking can be achieved by stacking four GEM foils using standard (S, 140  $\mu\text{m}$ ) and large (LP, 250  $\mu\text{m}$ ) hole pitch in an S-LP-LP-S configuration and by adjusting the gain share among the four layers. The latter optimisation allows for efficiently blocking the ions, most of which are produced in the last amplification step, i.e., in the layer closest to the readout pad. A careful choice of hole patterns is made to avoid the accidental alignment of holes in subsequent layers.

Thanks to this design, the detector is capable of satisfying the requirements for Run 3 operation, with an ion backflow below 2% and a local energy resolution at the  $^{55}\text{Fe}$ -peak below 14% [145].

### 3.2.4 Time-of-Flight

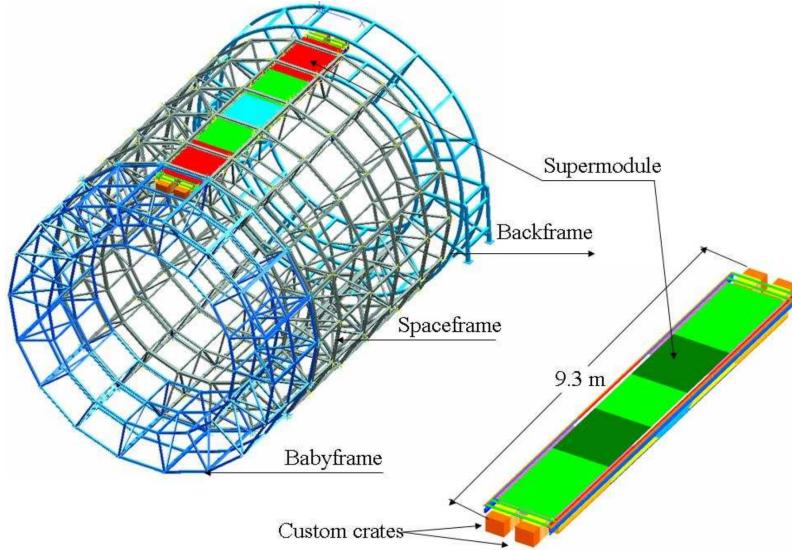


Figure 3.6: Schematic view of the ALICE TOF. Figure taken from Ref. [146].

The Time-of-Flight (TOF) detector is a large area detector covering the central pseudo-rapidity region ( $|\eta| < 0.9$ ), extending from an inner radius of 370 cm to an external one of 399 cm. Its main purpose is the particle identification in the intermediate momentum range, achieving a  $\pi/\text{K}$  and  $\text{K}/\text{p}$  separation better than 3 times the time-of-flight resolution up to a momentum of about  $2.5 \text{ GeV}/c$  for pions, and up to  $4 \text{ GeV}/c$  for protons. The TOF detector is built with a modular structure corresponding to 18 sectors in  $\varphi$  and five modules along the beam direction. It is based on Multi-gap Resistive Plate Chambers (MRPCs) [147], which are capable of providing a time resolution of about 40 ps [127]. The gas mixture used in the MRPCs is  $\text{C}_2\text{H}_2\text{F}_4(90\%)$ ,  $\text{i-C}_4\text{H}_{10}(10\%)$ ,  $\text{SF}_6(5\%)$ , which proved to be a good solution as no significant ageing effects were observed after 3.5 times the dose foreseen in the first 10 years of operation.

Together with the ITS and TPC, the TOF is the detector in charge of providing event-by-event identification of large samples of pions, kaons, and protons. The mass of a given particle is estimated from the measurement of its time-of-flight:

$$m = p \cdot \sqrt{\left(\frac{t_{\text{flight}}}{L}\right)^2 - 1} \quad , \quad (3.1)$$

where  $p$  is the particle's momentum, measured from the curvature of the track trajectory,  $t_{\text{flight}}$  is the time-of-flight, and  $L$  is the track length. The time-of-flight is evaluated as the difference between the time of arrival of the particle at the TOF

( $t_{\text{hit}}$ ) and the time of the collision ( $t_0$ ), estimated using a timing signal from the FT0 detector:  $t_{\text{flight}} = t_{\text{hit}} - t_0$ .

### 3.2.5 Fast Interaction Trigger

The Fast Interaction Trigger (FIT) serves as the main forward trigger, luminometer, and interaction-time detector. It also provides an initial indication of the vertex position and is used to determine the multiplicity, centrality, and reaction plane of heavy-ion collisions. The FIT consists of five distinct detector stations, positioned at different locations along the beam line: the FT0 (-A and -C), the FV0, and the FDD (-A and -C). An illustration of the FIT is shown in Fig. 3.7.

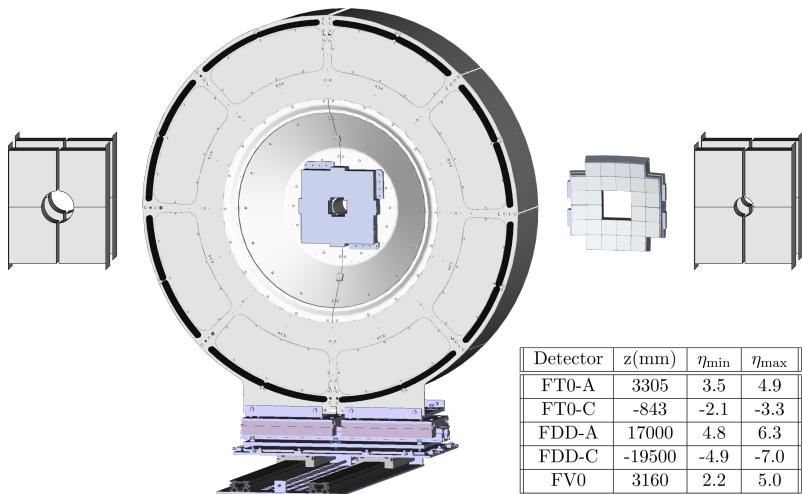


Figure 3.7: View of the FIT detectors illustrating the relative sizes of each component. From left to right, the FDD-A, FT0-A, FV0, FT0-C, and FDD-C are shown. The FT0-A and FV0 systems have a common mechanical support: the former is the small quadrangular structure in the centre of the large, circular FV0 support. The inset table lists the distance from the interaction point and the pseudorapidity coverage for each component. Figure taken from Ref. [128].

## FT0

The FT0 is made of two arrays of quartz Cherenkov radiators, FT0-A and FT0-C, coupled to MicroChannel Plate-based photomultipliers (MCP). Its main task is to determine the vertex position with an accuracy of a few centimetres, and to provide the interaction time with the precision needed by the TOF system for event-by-event particle identification. An excellent time resolution (< 50 ps) is therefore required. The FT0-A is located at 3.3 m from the nominal interaction point, while the FT0-C is only 84 cm from the interaction point. To provide a reliable indication of the time of a collision, the FT0-C has a convex shape, to ensure that each of the 112 quartz radiators is positioned at a distance of 84 cm from the nominal interaction point. On the contrary, the FT0-A is characterised by a planar geometry, made of 96 quartz radiators. To achieve the best possible timing resolution, the signal

path from each MCP anode to the front-end electronics has the same length. The intrinsic time resolution of each quadrant is  $\sigma_t \sim 13$  ps

### FV0

The FV0 is a large, segmented scintillator disk. It achieves a single MIP time resolution of about 200 ps with a very uniform response across the entire detection surface thanks to its light collection scheme, shown in Fig. 3.8. The active element of FV0 is a 4 cm-thick plastic scintillator divided into five concentric rings of equal pseudorapidity coverage. The outer diameter of the largest ring is 144 cm and the inner diameter of the smallest is 8 cm. The four inner rings are subdivided into eight sectors of 45 degrees each, while the outermost ring, due to its large area, has 16 sectors. A grid of equal-length fibers is connected to the back side of the scintillator, where they are attached to the photomultiplier tubes (PMTs). Each sector is read out by an independent PMT to provide a minimum bias and multiplicity trigger, when coupled to the FT0 information.



Figure 3.8: Photograph of one half of the FV0 detector. Figure taken from Ref. [128].

### FDD

The FDD is made of two similar arrays, FDD-A and FDD-C, surrounding the beam pipe on opposite sides of the nominal interaction point, and consisting of eight rectangular scintillator pads each, arranged into two overlapping layers of four sectors. A quadrant was removed from the innermost corner of each scintillator plate to allow for the passage of the beam pipe. Since the FDD covers a large pseudorapidity interval, and is sensitive to the presence of even a single MIP, it is an ideal system to tag interactions characterised by large rapidity gaps as those from photon-induced ultra-peripheral collisions or diffractive processes.

# Chapter 4

## D<sub>s</sub><sup>+</sup> and D<sup>+</sup> reconstruction strategy in proton-proton collisions

Due to their mean proper decay lengths ( $c\tau$ ) of 151.2 μm and 309.8 μm respectively [4], D<sub>s</sub><sup>+</sup> (cs) and D<sup>+</sup> (cd) mesons and their charge conjugates cannot be directly detected, as they typically decay before reaching the innermost layer of the ALICE detector. Consequently, their production is measured through an invariant mass analysis based on the full reconstruction of the decay topology. This analysis exploits their hadronic decays into D<sub>s</sub><sup>+</sup>, D<sup>+</sup> →  $\phi\pi^+ \rightarrow K^+K^-\pi^+$  and their charge conjugates, with branching ratios of  $(2.21 \pm 0.06) \times 10^{-2}$  and  $(2.69^{+0.07}_{-0.08}) \times 10^{-3}$  [4], respectively. An additional hadronic decay channel of the D<sup>+</sup> meson with a larger BR of  $(9.38 \pm 0.16) \times 10^{-2}$  [4] is usually exploited for the analyses of the production of D<sup>+</sup> mesons [148]. However, the reconstruction of the two D-mesons species in the same decay channel allows for the cancellation of many of the systematic uncertainties affecting the measurement, potentially leading to a more precise D<sub>s</sub><sup>+</sup>/D<sup>+</sup> production yield ratio measurement. The choice of reconstructing D mesons through their decays into a hadronic final state allows for the full reconstruction of the decay topology, therefore providing access to the D-meson kinematics, and the reconstruction of the secondary vertex, which is used in the selection of the signal against the large combinatorial background. Semileptonic decay measurements are in fact affected by larger uncertainties due to the presence of neutrinos in the final state, which are not detected. In addition, the decay into the resonant  $\phi(1020)$  state opens up the possibility of exploiting the invariant mass of the  $\phi$  meson for the selection of the D-meson candidates. A sketch of the decay topology of D<sub>s</sub><sup>+</sup> and D<sup>+</sup> mesons into  $\phi\pi^+ \rightarrow K^+K^-\pi^+$  is shown in Fig. 4.1.

D<sub>s</sub><sup>+</sup> and D<sup>+</sup> mesons (and their charge conjugates) are reconstructed in three different steps: i) firstly, charged tracks are reconstructed at midrapidity ( $|\eta| < 0.8$ ) exploiting the information of the ITS and TPC detectors; ii) D<sub>s</sub><sup>+</sup> and D<sup>+</sup> candidates are constructed by combining triplets of tracks with the appropriate charge signs, i.e., (+, -, +) for D<sub>s</sub><sup>+</sup> and D<sup>+</sup> mesons, and (-, +, -) for their antiparticles; iii) lastly, the D<sub>s</sub><sup>+</sup> and D<sup>+</sup> candidates are selected by applying a set of topological, kinematical, and Particle-IDentification (PID) selections employing Machine Learning algorithms. Given the large number of tracks produced in a pp collision and the comparatively small production cross-section of charmed hadrons as well as the

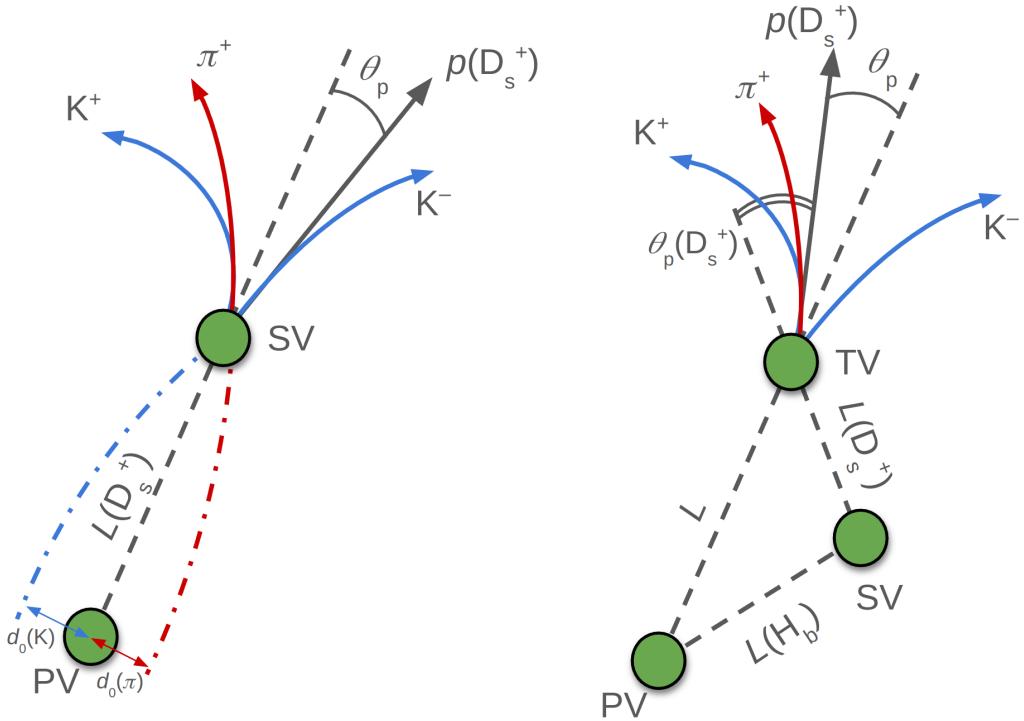


Figure 4.1: Sketch of the decay topology of  $D_s^+$  and  $D^+$  mesons into  $\phi\pi^+ \rightarrow K^+K^-\pi^+$  for prompt (left) and non-prompt (right)  $D_s^+$  meson.

small BRs of the considered decay modes, the vast majority of the constructed  $D_s^+$  and  $D^+$  candidates are obtained from the combination of uncorrelated tracks, which do not originate from the decay of a  $D_s^+$  or  $D^+$  meson. This results in a large *combinatorial background*, which has to be suppressed in order to extract the signal of the D mesons.

The spatial resolution capabilities of the ALICE detector described in Chapter 3 enable the separation of the secondary decay vertices of D mesons from the primary interaction vertex, which consents the development of an analysis based on the reconstruction and selection of secondary-vertex topologies characterised by relatively large separations from the primary interaction vertex. Furthermore, the PID of the decay products can be further exploited to improve the selection of D mesons.

Two distinct categories of D mesons emerge based on their production mechanism: *prompt*, directly produced in the hadronisation of a charm quark or through the strong decay of a directly produced excited charm-hadron or charmonium state, and *non-prompt* (also referred to as *feed-down*), produced in the decay of a hadron containing a beauty quark. Decay vertices of feed-down D mesons are on average more displaced from the primary interaction point with respect to promptly-produced ones, due to the larger mean proper decay length of beauty hadrons ( $c\tau \sim 500 \mu\text{m}$  [4]) as compared to charm hadrons. Therefore, by exploiting selection criteria based on displaced decay-vertex topologies, it is possible not only to separate D-meson signals from the combinatorial background, but also to discriminate between feed-down and prompt D mesons.

## 4.1 Data sample and event selection

The analysis reported in this Thesis is performed on a dataset of pp collisions at a centre-of-mass energy of  $\sqrt{s} = 13.6$  TeV, collected by the ALICE detector during the 2022 data-taking period. The data sample is collected using a Minimum-Bias trigger (the **Se18** trigger), which selects events satisfying the requirement of having a signal coincidence in the FT0-A and FT0-C detectors. Furthermore, since it is observed that the readout of the TPC (which is performed at the end of each time frame (TF)) causes a drop in the track reconstruction efficiency, it was decided to exclude the collisions read out during this time interval (*TF border selection*). An additional requirement is imposed on the primary vertex position along the beam axis,  $|z_{\text{vtx}}| < 10$  cm, to ensure that the point where protons have collided is located within the region of the detector where the tracking efficiency is optimal. Of the total 59.2 billion collected events, only 54.7 billion have been analysed, corresponding to an integrated luminosity of  $\mathcal{L}_{\text{int}} \sim 1 \text{ pb}^{-1}$ . Around 3% of the events are rejected due to the minimum bias trigger, about 1% due to the TF border selection, and 4% of the events are rejected due to the requirement on the  $z_{\text{vtx}}$  position. A summary of the event selection is shown in Fig. 4.2.

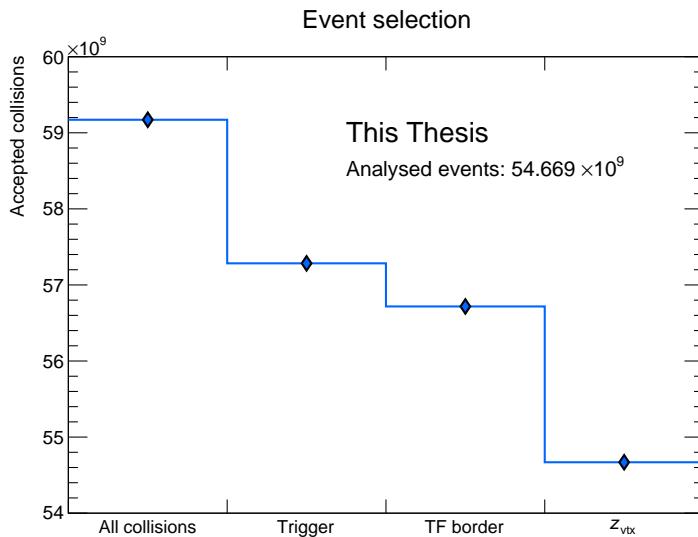


Figure 4.2: Summary of the event selection criteria applied to the data sample.

## 4.2 $D_s^+$ and $D^+$ reconstruction workflow

The reconstruction of  $D_s^+$  and  $D^+$  mesons is a complex process performed in many subsequent steps. As stated in Chapter 3, the reconstructed data are stored in AOD format. This format contains essential information about the reconstructed tracks, which are parameterised at the innermost update point, i.e., the closest point to the primary vertex with a detected track signal. Since the track parametrisation is not the same for each track, as missing hits in the ITS can lead to different radii

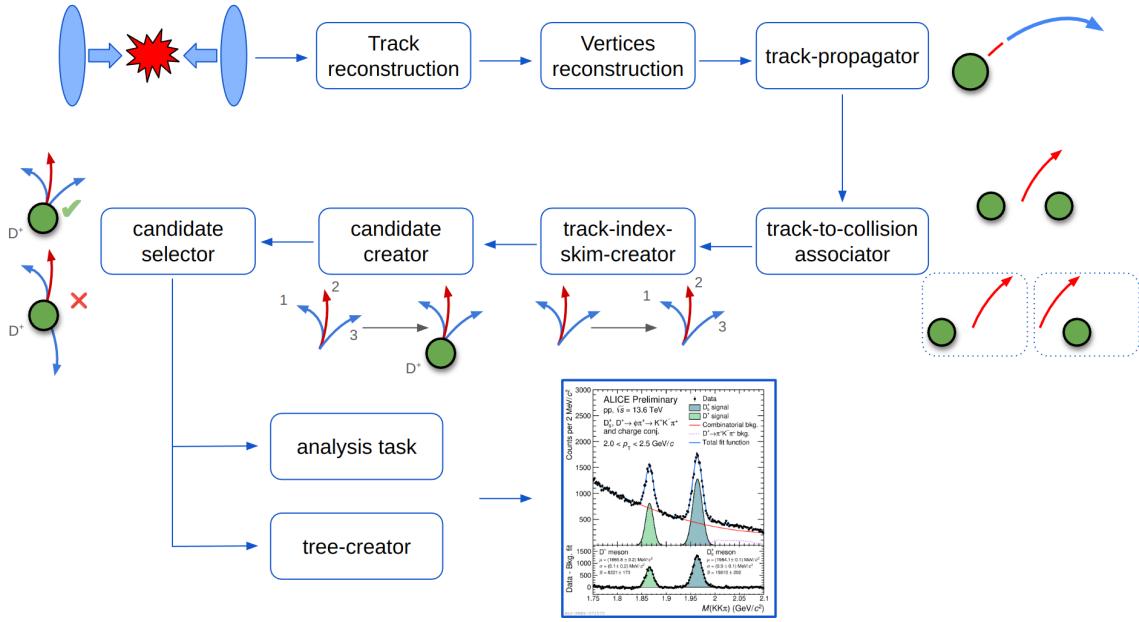


Figure 4.3: Schematic representation of the reconstruction workflow of  $D_s^+$  and  $D^+$  mesons.

of the innermost update points, the tracks are propagated to the point of closest approach to the primary vertex, and their parameters are updated accordingly. This procedure is performed in the **track-propagator** workflow, and only tracks passing a set of predefined quality requirements are selected.

Because of the continuous readout, there exists the possibility that a single track is found to be compatible with multiple collisions (these tracks are named *ambiguous tracks*). In the AOD files, the ambiguous tracks are only associated with the first space-time-compatible collision. In order to increase the reconstruction efficiency, the **track-to-collision associator** workflow associates each track with all compatible collisions. This process significantly increases the reconstruction efficiency of decaying particles.

For each considered collision, combinations of two or three tracks are built in the **track-index-skim-creator** workflow, producing candidates of charmed hadrons, in particular for the hadronic decays of  $D^0 \rightarrow K^-\pi^+$ ,  $D^{*+} \rightarrow D^0\pi^+$ ,  $D^+ \rightarrow \pi^+K^-\pi^+$ ,  $D_s^+(D^+) \rightarrow K^+K^-\pi^+$ ,  $\Lambda_c^+ \rightarrow pK^-\pi^+$ ,  $\Xi_c^+ \rightarrow pK^-\pi^+$  and their charge conjugates. The workflow produces a table of track indices for the combinations passing a set of loose selection criteria, which depends on the considered hadron species. Candidates for each particle species and their decay vertices are created in the **candidate-creator** workflow starting from the track indices of combinations saved by the **track-index-skim-creator**, and those satisfying the more stringent analysis selection criteria are flagged in the **candidate-selector** workflow. The **tree-creator** (**analysis-task**) workflow produces the final output of the analysis, which consists of a ROOT [136] TTree (several ROOT histograms) containing the properties of selected candidates.

## 4.3 Vertex reconstruction and D-meson selection

The track and primary vertex reconstruction is performed both in the “synchronous” reconstruction, performed quasi-online for quality-control and detector-calibration purposes, and in the “asynchronous” reconstruction performed on the stored raw data a few days after the data taking and after the detector calibrations have been performed offline, using information from the central barrel detectors. Firstly, the hits recorded in the ITS and TPC detectors are converted into *clusters*, corresponding to groups of hits produced by a single particle interacting with a detector element. They contain information about position, signal, timing, and the corresponding uncertainties. Then, a first estimation of the interaction-vertex position is obtained by using clusters in the ITS inner barrel. Track segments, called tracklets, are constructed by associating pairs of clusters from different layers within a predefined azimuthal acceptance window. The primary vertex is obtained as the point in space where the largest number of tracklets converge.

Tracks are constructed using ITS [149] and TPC [150] standalone reconstructions using a Kalman filter-based algorithm [151]. Track information in the two detectors is then matched within compatible time windows (ITS-TPC matching) to build the *global* tracks. These are subsequently refitted in the outward direction and matched to outer detectors, e.g., TRD, TOF, and HMPID [152]. The propagation to detectors located beyond the TPC radius does not update the kinematic track parameters, but it is used to attach the information needed for particle identification to the track. Tracks are then re-fitted towards the interaction point using the previously found clusters, and the final track parameters and covariance matrix are computed and stored at the innermost hit. The primary vertex is then determined using the tracks reconstructed in the previous steps.

The decay vertex of the candidate is reconstructed through the minimisation of a  $\chi^2$ -like quantity, denoted as  $D$ :

$$D = \sqrt{\sum_{i=1}^3 \left[ \left( \frac{x_i - x_0}{\sigma_{x_i}} \right)^2 + \left( \frac{y_i - y_0}{\sigma_{y_i}} \right)^2 + \left( \frac{z_i - z_0}{\sigma_{z_i}} \right)^2 \right]} ,$$

where  $(x_i, y_i, z_i)$  and  $(\sigma_{x_i}, \sigma_{y_i}, \sigma_{z_i})$  represent the position and the uncertainty of the  $i$ -th track at the point of closest approach, respectively, while  $(x_0, y_0, z_0)$  denotes the position of the reconstructed vertex. The invariant mass and momentum of the  $D_s^+$  and  $D^+$  candidates are computed from the energy and momentum of the measured tracks evaluated at the point of closest approach to the decay vertex. The momentum of the candidate is defined as the sum of the momenta of the three tracks. For the invariant mass computation, the kaon mass is always assigned to the track with opposite charge sign with respect to the D-meson candidate (*opposite-sign track*). For the two *like-sign tracks*, the two pion-kaon mass hypothesis combinations (i.e.  $(K^+ K^- \pi^+)$  and  $(\pi^+ K^- K^+)$  for positively charged candidates) are considered.

The vast majority of the created  $D_s^+$ - and  $D^+$ -meson candidates belong to the large combinatorial background. To increase the signal-over-background ratio and the statistical significance of the measurement, tight selections are required to suppress background candidates. The analyses presented in this Thesis exploit several

selection criteria, which can be divided into:

- i Track-level selections
- ii Selections based on the decay topology and kinematics
- iii Particle identification of the decay products

In the following, the variables used to select D mesons and some of the applied selections are described in more detail.

### 4.3.1 Track-level selections

Only tracks that successfully pass strict quality and kinematical requirements are considered eligible for inclusion in the construction of  $D_s^+$ - and  $D^+$ -meson candidates. In particular, only ITS-TPC tracks with at least 70 (out of a maximum of 159) TPC crossed rows (i.e., the total number of hit TPC pad rows) are selected. To improve the vertex reconstruction procedure, at least one hit in the 3 innermost layers of the ITS, which compose the ITS inner barrel, was required. Track-quality requirements of  $\chi^2$  per TPC cluster smaller than 4 and  $\chi^2$  per ITS cluster smaller than 36 have been applied.

Secondary vertices of  $D_s^+$ - and  $D^+$ -meson candidates are constructed using tracks having  $|\eta^{\text{dau}}| < 0.8$  and  $p_T^{\text{dau}} > 0.3 \text{ GeV}/c$ . These kinematical selections limit the rapidity acceptance of D mesons, which steeply decreases for  $|y| > 0.5$  at low  $p_T$  and for  $|y| > 0.8$  for  $p_T \gtrsim 5 \text{ GeV}/c$ , as shown in Fig. 4.4. The applied track-quality selection criteria are summarised in Table 4.1.

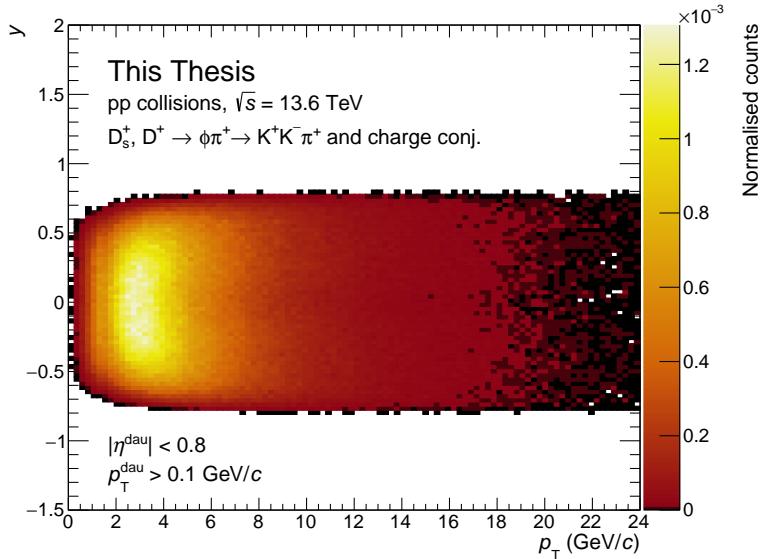


Figure 4.4: Rapidity and transverse momentum distribution of reconstructed  $D_s^+$  and  $D^+$  mesons from MC simulations of pp collisions at  $\sqrt{s} = 13.6 \text{ TeV}$ . The kinematical selections applied on the daughter tracks, requiring  $|\eta^{\text{dau}}| < 0.8$  and  $p_T^{\text{dau}} > 0.3 \text{ GeV}/c$  cause the steep decrease of the rapidity acceptance for the D-meson transverse momentum range  $p_T \lesssim 2 \text{ GeV}$ .

Single-track selection	Value
Number of TPC crossed-rows >	70
$ \eta  <$	0.8
$p_T >$	0.1 $\text{GeV}/c$
$\chi_{\text{TPC}}^2/\text{TPC clusters} <$	4
$\chi_{\text{ITS}}^2/\text{ITS clusters} <$	36
ITS matching	At least 1 cluster in L0, L1, L2

Table 4.1: Applied single-track selection criteria.

### 4.3.2 Topological selections

$D_s^+$  and  $D^+$  mesons exhibit a displaced decay vertex topology, which can be used to separate the signal from the uninteresting combinatorial background. Moreover, promptly produced D mesons exhibit different topological features compared to feed-down D mesons, enabling further discrimination between the two production mechanisms. This differentiation potentially offers insights into beauty-quark production through the measurement of non-prompt open-charm states.

A cut on the daughter-track transverse impact parameter projection in the transverse plane  $d_{0,\text{dau}}^{xy}$  was applied, requiring  $d_{0,\text{dau}}^{xy} > 25 \mu\text{m}$  for tracks with  $p_T < 2 \text{ GeV}/c$ . This selection allows for reducing the combinatorial background by exploiting the displacement of the tracks arising from the decay of a D meson. Most of the tracks making up the combinatorial background are produced directly in the primary vertex, and therefore have a small impact parameter. Since this selection is performed directly on the tracks, it can be applied at a very early stage of the reconstruction process, allowing for a reduction of the number of candidates to be processed in the subsequent steps and significantly reducing the computational time.

The topological selections on the D-meson candidates are tuned as a function of  $p_T$  in order to increase the signal-over-background and the statistical significance of the measurement. The different variables used in the selections are presented herein. The corresponding distributions are shown for the signal, which is divided into prompt and feed-down contributions and obtained from Monte Carlo simulations, as well as for the combinatorial background, which is obtained from real data in an invariant-mass region away from the signal region of the  $D_s^+$  and  $D^+$  mesons, denoted as *sidebands* and chosen as  $1.7 < M < 1.75 \text{ GeV}/c^2$  or  $2.1 < M < 2.15 \text{ GeV}/c^2$ . The invariant mass of the D-meson candidate is defined as the mass of the track triplet used to build the candidate. For each track triplet, both the mass hypotheses on the like-sign tracks (i.e.,  $K^+K^-\pi^+$  and  $\pi^+K^-K^+$ ) are considered, and eventually rejected if any of the selection criteria applied to select signal candidates is not satisfied. With this selection of the invariant-mass region employed for the definition of the sidebands, the candidates in the combinatorial background are well separated from the signal.  $D_s^+$  and  $D^+$  mesons are characterised by masses of  $(1968.35 \pm 0.07) \text{ MeV}/c^2$  and  $(1869.66 \pm 0.05) \text{ MeV}/c^2$  [4], respectively, and the width of their reconstructed invariant-mass distributions  $\sigma$  ranges from 7 to 30  $\text{MeV}/c^2$ , depending on  $p_T$ , as reported in Chapter 5. Therefore, a separation of more than  $4.5\sigma$  from the

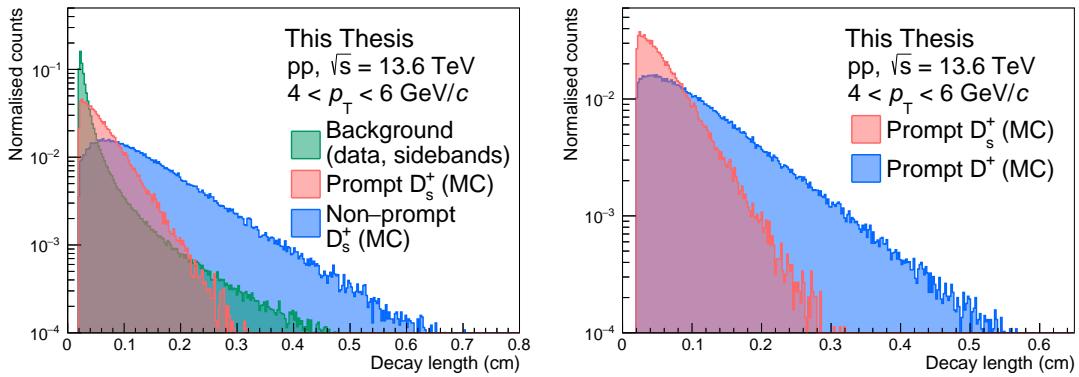


Figure 4.5: Distributions of decay length for background (green), prompt (red), and non-prompt (blue)  $D_s^+$ -meson signals (left panel) and for prompt  $D_s^+$ - (red) and  $D^+$ - (blue) meson signals (right panel) in pp collisions at  $\sqrt{s} = 13.6 \text{ TeV}$  in the  $4 < p_T < 6 \text{ GeV}/c$  D-meson transverse momentum interval. For prompt and non-prompt  $D_s^+$  and  $D^+$  mesons, the distributions are taken from Monte Carlo simulations, whereas for combinatorial background they are taken from the data sidebands.

$D_s^+$  and  $D^+$  signal peaks is achieved across the whole analysed  $p_T$  range, ensuring negligible signal contamination in the sidebands.

## Decay Length

The decay length  $L$  is defined as the spatial distance between the primary and secondary vertices. It provides an approximation of the actual decay length of  $D_s^+$  and  $D^+$  mesons, as the particle's curvature resulting from the motion in the presence of a magnetic field is not considered. However, given the small mean proper decay length of  $\sim 150$  ( $310$ )  $\mu\text{m}$  of  $D_s^+$  ( $D^+$ ) mesons, this effect can be neglected.

This is one of the most important variables used to distinguish between signal and combinatorial background, since the displaced topology of the signal shifts the decay length distribution towards larger values. Furthermore, since beauty-hadron decay vertices are not reconstructed, the measured decay length of non-prompt D-mesons also accounts for the decay length of their parent hadrons. As a consequence, the resulting decay length distribution for feed-down D mesons is particularly shifted towards larger values also compared to prompt D mesons, allowing for an easier separation of this class of signal from the background. The distribution of the decay length also depends on the D-meson  $p_T$ , because of the Lorentz boost which contributes to an increase in the distance flown by the hadron measured in the laboratory reference frame. Moreover, since the mean proper decay length of  $D^+$  meson is about twice as large as that of the  $D_s^+$  meson, the decay length distribution of the former is shifted towards larger values. Therefore, the same decay length-based selection criterion will yield different selection efficiencies for the two D-meson species.

The distributions of the decay length for combinatorial background, prompt  $D_s^+$  and  $D^+$  mesons, and for non-prompt  $D_s^+$  mesons are shown in Fig. 4.5 for candidates reconstructed in the  $4 < p_T < 6 \text{ GeV}/c$  D-meson transverse momentum interval.

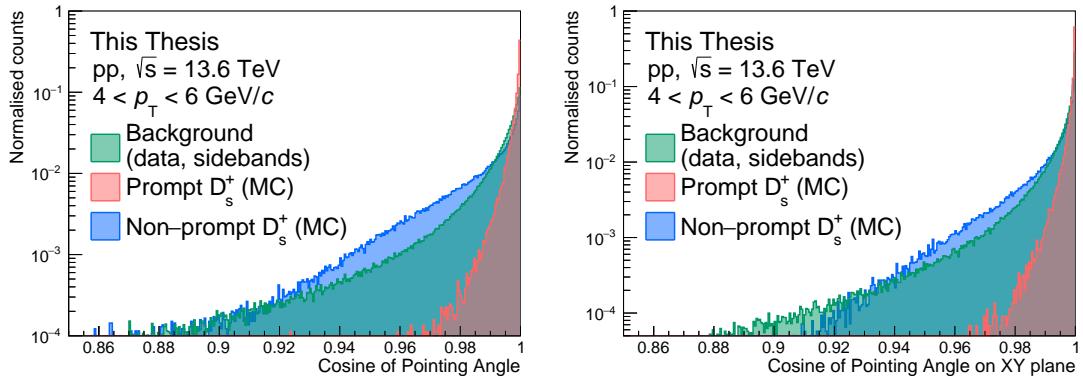


Figure 4.6: Distributions of the cosine of the pointing angle (left panel) and its projection on the transverse plane (right panel) for  $D_s^+$  mesons in pp collisions at  $\sqrt{s} = 13.6 \text{ TeV}$  in the  $4 < p_T < 6 \text{ GeV}/c$  D-meson transverse momentum interval. The distributions are shown for prompt  $D_s^+$  mesons (red), non-prompt  $D_s^+$  mesons (blue), and combinatorial background (green). For prompt and non-prompt  $D_s^+$  mesons, the distributions are taken from Monte Carlo simulations, whereas for combinatorial background they are taken from the data sidebands.

In addition to the decay length, its projection on the transverse plane  $L^{xy}$  can also be used to select  $D_s^+$  and  $D^+$  signals. Previous measurements of D mesons also leveraged the normalised decay length, defined as the ratio between the decay length and its uncertainty, to further improve the signal extraction. Although this variable demonstrated a good separation power, it was not used in this analysis, due to possible biases introduced by the uncertainty description in the Monte Carlo simulations.

### Cosine of pointing angle

The pointing angle  $\theta_p$  is defined as the angle between the flight direction of the D meson, obtained from the line connecting the primary and secondary vertices, and the direction of the reconstructed D-meson momentum. This variable can also be defined using only the transverse components of these quantities ( $\theta_p^{xy}$ ), as done for the decay length. In an ideal scenario where particles' momentum and the primary and secondary vertices are perfectly reconstructed, the cosine of the pointing angle for promptly-produced D mesons would be equal to 1, while that of feed-down D mesons would be distributed around 1, but with a tail towards lower values as the flight direction of the D meson is not equal to that of the parent beauty hadron. On the contrary, the pointing angle for combinatorial background could assume any value with the same probability. The pointing angle is therefore particularly useful to separate the signal from the combinatorial background. The distributions of the cosine of the pointing angle and its projection on the transverse plane are shown in Fig. 4.6 for signal and combinatorial background candidates reconstructed in the  $4 < p_T < 6 \text{ GeV}/c$  D-meson transverse momentum interval. Due to the finite resolution of the tracking detectors, the pointing angle is not perfectly reconstructed, leading to a distribution of the cosine of the pointing angle for prompt  $D_s^+$  that is

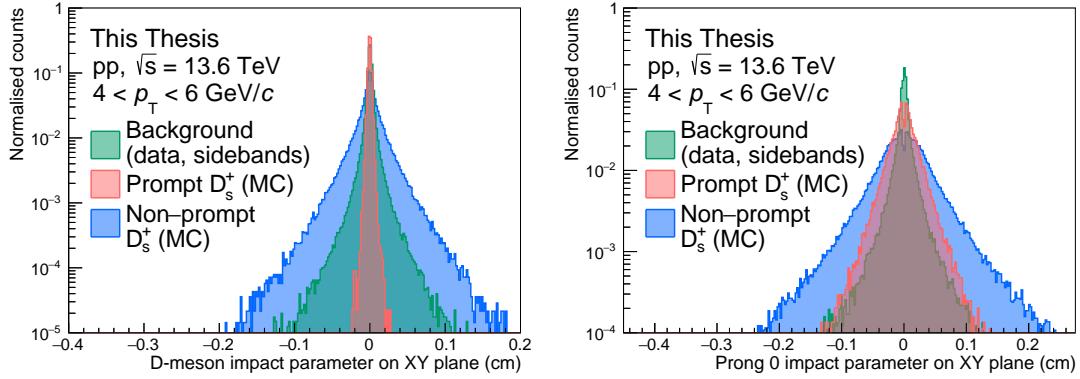


Figure 4.7: Distributions of the projection of the impact parameter on the transverse plane for  $D_s^+$  mesons (left panel) and for one of the like-sign daughter tracks (right panel) in pp collisions at  $\sqrt{s} = 13.6$  TeV in the  $4 < p_T < 6$  GeV/ $c$  D-meson transverse momentum interval. The distributions are shown for prompt  $D_s^+$  mesons (red), non-prompt  $D_s^+$  mesons (blue), and combinatorial background (green). For prompt and non-prompt  $D_s^+$  mesons, the distributions are taken from Monte Carlo simulations, whereas for combinatorial background they are taken from the data sidebands.

peaked at 1, but with a tail towards lower values. The background distribution peaks at 1 as well. However, this effect is due to other applied selection criteria which select only candidates with similar features as those of the signal.

Similarly to the decay length, also  $\theta_p$  evolves with the D-meson transverse momentum. Because of the Lorentz boost, the direction of the daughter particles of the D meson is more collimated with the D-meson flight direction at higher  $p_T$ . This results in a distribution of the cosine of the pointing angle that is more peaked at 1 for higher  $p_T$  values.

### Impact parameter in the transverse plane

The projection of the D-meson impact parameter on the transverse plane  $d_0^{xy}$  is defined as the distance of closest approach between the reconstructed flight line of the D-meson and the primary vertex, projected in the  $xy$  plane. It is expected to be very close to zero for promptly-produced D mesons, with any deviation due to the detector resolution. Thanks to the better momentum and vertexing resolution, its distribution becomes narrower at high  $p_T$ . It is a powerful variable not only to separate the signal from the combinatorial background, but also to discriminate between prompt and feed-down D mesons, as the latter have a much broader impact parameter distribution, as shown in the left panel of Fig 4.7. In addition, also the projection of the daughter tracks' impact parameter to the primary vertex projected on the transverse plane is used to select the signal. Since combinatorial-background candidates are mostly composed of tracks originating from the primary vertex, a narrow distribution around 0 is expected for the prongs' impact parameters of such candidates. On the contrary, the displaced decay vertex of the signal candidates leads to a broader distribution of the impact parameter for prompt D-meson candidates.

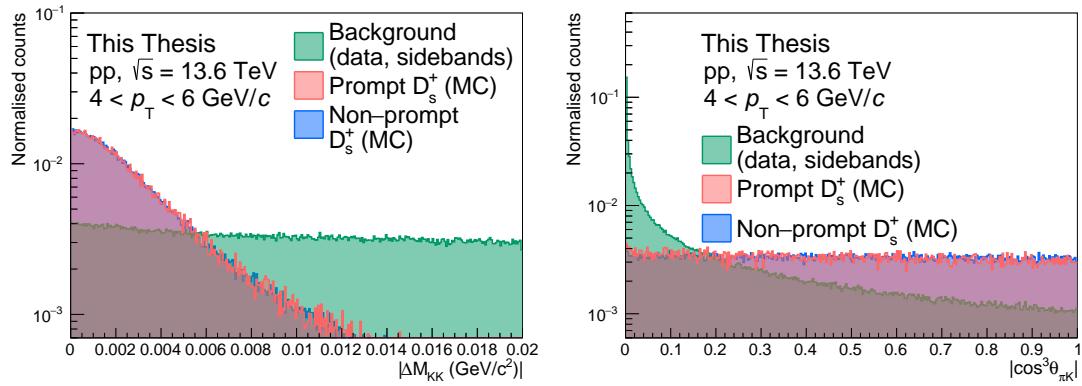


Figure 4.8: Distributions of the difference between the reconstructed  $K^+K^-$  invariant mass and the PDG mass of the  $\phi$  meson (left panel) and the cosine cubed of the  $K-\pi$  angle in the KK rest frame (right panel) for  $D_s^+$  mesons in pp collisions at  $\sqrt{s} = 13.6 \text{ TeV}$  in the  $4 < p_T < 6 \text{ GeV}/c$  D-meson transverse momentum interval. The distributions are shown for prompt  $D_s^+$  mesons (red), non-prompt  $D_s^+$  mesons (blue), and combinatorial background (green). For prompt and non-prompt  $D_s^+$  mesons, the distributions are taken from Monte Carlo simulations, whereas for combinatorial background they are taken from the data sidebands.

dates, and to an even broader one for the non-prompt contribution to the signal, as shown in the right panel of Fig. 4.7.

### 4.3.3 Kinematic selections

In addition to the selections based on the decay topology, kinematic selections are also applied to increase the signal-over-background ratio and the statistical significance of the measurement.

#### Difference between reconstructed and PDG mass of the $\phi$ meson

This selection exploits the production of an intermediate resonant  $\phi(1020)$ -meson state in the considered decay channel of the D mesons, which is characterised by an invariant mass of  $M_\phi = (1019.461 \pm 0.016) \text{ MeV}/c^2$  and a narrow resonance peak of  $\Gamma_\phi = 4.249 \pm 0.013 \text{ MeV}/c^2$  [4]. Therefore, the difference between the reconstructed invariant mass of the  $K^+K^-$  pair and the PDG mass of the  $\phi$  meson  $|\Delta M(\text{KK})|$  is expected to be close to zero for signal candidates, while the distribution for combinatorial background is expected to be uniformly distributed.

Moreover, two possible  $K^+K^-$  pairs can be built for each triplet of tracks, depending on the mass hypothesis assigned to the like-sign tracks (e.g., for a  $D_s^+$  decay, both the  $K^+K^-\pi^+$  and  $\pi^+K^-\bar{K}^+$  mass hypotheses can be considered). Both hypotheses are considered in the reconstruction, and the selection on  $|\Delta M(\text{KK})|$  results extremely helpful not only in rejecting the combinatorial background, but also the reflections, i.e., D mesons with wrongly-assigned mass hypothesis to the decay tracks.

The distribution of this variable is shown in the left panel of Fig. 4.8 for signal

and combinatorial background candidates reconstructed in the  $4 < p_T < 6 \text{ GeV}/c$  D-meson transverse momentum interval.

### Cosine cubed of the K- $\pi$ angle in the KK rest frame

The decay of a (pseudoscalar) D-meson to a (vector)  $\phi$ -meson and a (pseudoscalar)  $\pi^+$  final state results in an alignment of the spin of the  $\phi$  meson with respect to the direction of motion of the  $\phi$  relative to the D-meson [153]. As a consequence, the distribution of  $\cos(\theta'(\text{K}))$ , where  $\theta'(\text{K})$  is the angle between one of the kaons and the pion in the  $K^+K^-$  rest frame, follows a  $\cos^2(\theta'(\text{K}))$  shape, which in turn implies a flat distribution for the  $\cos^3(\theta'(\text{K}))$  variable, in case of signal. In contrast, the combinatorial background has a flat distribution for  $\cos(\theta'(\text{K}))$ , and its  $\cos^3(\theta'(\text{K}))$  distribution peaks at zero. This variable is particularly useful to separate the signal from the combinatorial background, as shown in the right panel of Fig. 4.8.

#### 4.3.4 Particle identification selections

One of the distinctive features of the ALICE detector is the excellent PID capabilities in a wide range of momentum, achieved by combining the information from different detectors. The TPC is the main PID detector in the ALICE experiment, as it provides a simultaneous measure of the specific energy loss via the charge produced in the drift volume and deposited in its 159 pad rows, the charge, and the momentum of each particle traversing the detector gas. A truncated mean of the  $dE/dx$  samples is calculated discarding the 40% highest-charge clusters, which ensures the removal of the Landau tail of the  $dE/dx$  distribution caused by  $\delta$ -rays, which are produced from the ionisation of the gas by the electrons produced in the primary ionisation. The particle identification is based on the comparison of the measured  $dE/dx$  with the expectation for a specific particle species with a certain momentum  $p$ . The expected  $dE/dx$  is described by the Bethe-Bloch formula, and is typically parametrised with a function originally proposed by the ALEPH collaboration [154]:

$$f(\beta\gamma) = \frac{P_1}{\beta^{P_4}} \left[ P_2 - \beta^{P_4} - \log \left( P_3 + \frac{1}{(\beta\gamma)^{P_5}} \right) \right] .$$

In this parametrisation,  $\beta$  and  $\gamma$  are the particle velocity and Lorentz factor, respectively, and  $P_{1-5}$  are the parameters of the fit function. The  $dE/dx$  distribution in the TPC is shown in the right panel of Fig. 4.9 for pp collisions at  $\sqrt{s} = 13.6 \text{ TeV}$ . The curves obtained with a fit using the ALEPH parametrisation of the Bethe-Bloch formula for electrons, muons, pions, kaons, protons, deuterium, and tritium are also shown. While in the non-relativistic low momentum region ( $p \leq 1 \text{ GeV}/c$ )  $p$ ,  $K$ , and  $\pi$  can be separated on a track-by-track basis, at higher momenta, in the relativistic rise region, particles can still be identified on a statistical basis via multi-Gaussian fits [146]. This is possible because the truncated-mean method produces a  $dE/dx$  peak with a Gaussian shape down to three orders of magnitude for long tracks (at least 130 clusters) [146].

The TOF detector is dedicated to PID in the intermediate momentum range, up to  $2.5 \text{ GeV}/c$  for pions and kaons, and up to  $4 \text{ GeV}/c$  for protons. In this higher- $p_T$

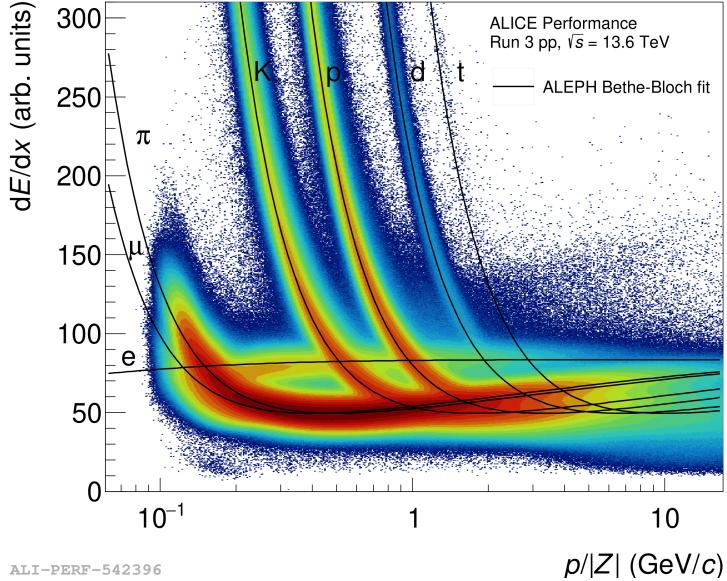


Figure 4.9: Specific energy loss distribution  $dE/dx$  in the TPC for pp collisions at  $\sqrt{s} = 13.6 \text{ TeV}$ . The curves obtained with a fit using the ALEPH parametrisation of the Bethe-Bloch formula for electrons, muons, pions, kaons, protons, deuterium, and tritium are also shown. Figure taken from the ALICE figure repository [60].

region, the TPC is not able to provide a sufficient separation between the different particle species. The TOF detector is based on the measurement of the time of flight of charged particles between the interaction point and the detector. The start time is provided by the FT0 forward-rapidity detectors, and a total time resolution of about 80 ps is achieved, allowing for a separation of the different particle species based on the time of flight. The TOF PID performance is shown in Fig. 4.10 for pp collisions at  $\sqrt{s} = 13.6 \text{ TeV}$ . The large background is due to tracks that are incorrectly matched to TOF hits in high-multiplicity pp collisions.

The PID information on the decay tracks can be exploited in the D-meson selection to improve the background rejection and thereby the signal extraction. A track is considered compatible with a mass hypothesis (i.e. a hadron species) depending on the difference between the measured signal  $S_{\text{meas}}^{\text{TPC,TOF}}$  and the expected signal  $S_{\text{exp}(\pi,K,p)}^{\text{TPC,TOF}}$  for the given hypothesis. The measured signal is the truncated mean of the  $dE/dx$  distribution for the TPC detector, and the time of flight for the TOF detector. The expected signal is extracted from the ALEPH parametrisation of the Bethe-Bloch formula described above for the TPC, and from the expected value of time of flight for a given particle species based on the measured track length and track momentum, from a parametrisation of the distribution of  $\beta$  as a function of  $p$  for the TOF. The difference between the measured and expected signals is then normalised by the uncertainty on the measured signal  $\sigma_{\text{TPC,TOF}}$ :

$$n\sigma_{\pi,K,p}^{\text{TPC,TOF}} = \frac{|S_{\text{meas}}^{\text{TPC,TOF}} - S_{\text{exp}(\pi,K,p)}^{\text{TPC,TOF}}|}{\sigma_{\text{TPC,TOF}}} .$$

Tracks characterised by a poor determination of  $dE/dx$  in the TPC and not matched to a hit in the TOF are not considered in this analysis. A loose preselection is applied

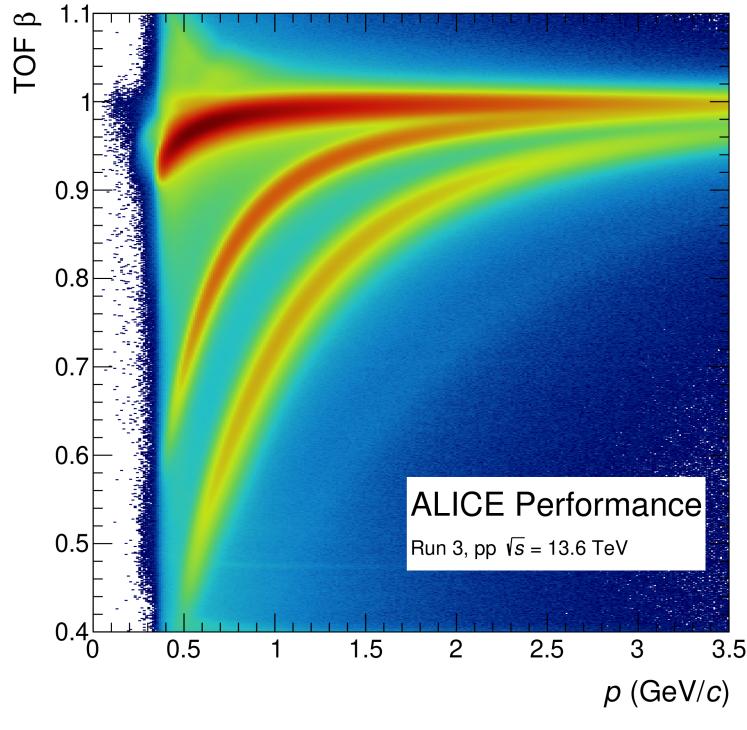


Figure 4.10:  $\beta$  parameter measured with the TOF detector as a function of the momentum  $p$  measured by the TPC detector in pp collisions at  $\sqrt{s} = 13.6$  TeV. Figure taken from the ALICE figure repository [60].

to the candidates based on PID information, requiring the three tracks to have a  $n\sigma_{\text{TPC}} < 5$ , or  $n\sigma_{\text{TOF}} < 5$  for the considered (pion or kaon) mass hypothesis. This selection allows rejecting a large number of background candidates and, moreover, it effectively reduces the number of reflections in the data sample, as the wrongly-assigned mass hypothesis is often rejected.

Due to a TPC–TOF matching efficiency which is significantly lower than unity, especially at low  $p_T$ , the PID information from the two detectors is typically combined to obtain a single PID variable  $n\sigma_{\pi,\text{K},p}^{\text{comb}}$ , which is used in the final selection. The PID information is combined as

$$n\sigma_{\pi,\text{K},p}^{\text{comb}} = \begin{cases} n\sigma_{\pi,\text{K},p}^{\text{TOF}} & \text{only TOF information is available} \\ n\sigma_{\pi,\text{K},p}^{\text{TPC}} & \text{only TPC information is available} \\ \frac{1}{\sqrt{2}} \sqrt{(n\sigma_{\pi,\text{K},p}^{\text{TPC}})^2 + (n\sigma_{\pi,\text{K},p}^{\text{TOF}})^2} & \text{otherwise} \end{cases}$$

This reduces the number of features used for the Machine Learning analysis described in the following Chapters, and ensures a proper handling of sparse data.

# Chapter 5

## $D_s^+$ and $D^+$ signal extraction

The main ingredient for the evaluation of the  $D_s^+/D^+$  production-yield ratio are the  $D_s^+$ - and  $D^+$ -meson raw yields, i.e., the number of reconstructed and selected  $D_s^+$  and  $D^+$  mesons. Due to the vast amount of combinatorial background, the extraction of the raw yields through a simple candidate counting method is not feasible. Instead, the raw yield is obtained on a statistical basis by fitting the invariant-mass distribution of the  $D_s^+$ - and  $D^+$ -meson candidates passing tight selection criteria. To reduce the combinatorial background and preserve the highest possible efficiency of D-meson selection, Machine Learning algorithms have been employed. The following sections describe the procedure for the extraction of the raw yield of  $D_s^+$  and  $D^+$  mesons.

### 5.1 Machine Learning

The term *Machine Learning* (ML) is a broad and versatile concept, encompassing a wide range of algorithms that grant computers the capacity to learn and adapt without being explicitly programmed to do so [155]. A more comprehensive definition characterises ML as the study of algorithms that enhance their performance at a specific task through the accumulation of experience [156]. In recent years, ML techniques have witnessed widespread adoption across diverse fields, with significant impacts realised especially with the emergence of generative models such as GPT [157]. ML algorithms have found extensive applications in the high-energy physics field, primarily for the task of distinguishing interesting signals from the vast background present in particle-collision data. Furthermore, these algorithms have been employed as triggers, aiding in the rapid identification of events of interest, and have also been instrumental in event reconstruction. Notably, ML algorithms were used in the discovery of the Higgs boson [133], one of the most significant achievements in the field of particle physics of the last decades.

#### 5.1.1 Supervised learning

Supervised learning is one of the main branches of machine learning, along with unsupervised and reinforcement learning. Machine learning tasks are usually described

in terms of how the machine learning system should process an example, which is a collection of features  $\mathbf{x}$  that have been quantitatively measured from some object or event that one wants the machine learning system to process. In the case of supervised learning, each example is coupled with a corresponding label or target,  $\mathbf{y}$ . The objective of supervised learning is to learn to predict or infer  $\mathbf{y}$  based on the associated features,  $\mathbf{x}$ , assuming that there exists a functional relationship  $\mathbf{y} = f(\mathbf{x})$  between the two. The goal of the ML system is to produce an approximation  $\hat{f}(\mathbf{x})$  of the true function  $f(\mathbf{x})$  by minimising a given loss function, which quantifies the discrepancy between the predicted and true labels. Supervised learning problems are further segmented into two distinct sub-categories: classification and regression. In the former, the label  $\mathbf{y}$  assumes values from a finite and discrete set of categories, often representing distinct classes or groups. In the latter, the label  $\mathbf{y}$  takes the form of one or more continuous variables, necessitating the learning system to deduce a continuous function or mapping between  $\mathbf{x}$  and  $\mathbf{y}$ .

The usage of ML algorithms in classification problems, such as the one presented in this Thesis, allows for the definition of multi-dimensional non-linear decision boundaries, which are not available with traditional selection methods based on linear selections on some cut variables. This is particularly important as it provides more efficient selections and a larger purity of the selected data sample.

The application of a supervised learning algorithm to a dataset involves the following steps: i) the model's hyperparameters are optimised through reduced training phases; ii) the optimised model is trained on a set of labelled data, i.e., the value of  $\mathbf{y}$  is known for each example in the training set; iii) the model is tested on a separate set of labelled data, known as the test set, to evaluate its performance; iv) the model is then used to make predictions on new, unseen data.

## Hyperparameters tuning

Before the training phase, hyperparameter tuning is performed to optimise the model's performance. *Hyperparameters* are parameters that are not learned during the training process, but rather define the model's architecture and the training process itself. Hyperparameter tuning is usually performed through a grid search, random search, or with a more efficient bayesian optimisation [158, 159, 160]. Several combinations of hyperparameters are tested on a dedicated labelled dataset: the validation set. Models with different hyperparameter sets are trained with a reduced training phase, and the one yielding the best performance is then selected for a full training phase.

## Training

During the training process, the model learns (i.e., adjusts its internal parameters) to map the input features  $\mathbf{x}$  to the corresponding labels  $\mathbf{y}$  by minimizing a given loss function. Typically used loss functions include the Mean Squared Error (MSE) for regression tasks and the Cross-Entropy loss [161] for classification tasks. The loss function is minimised through an optimisation algorithm, usually stochastic gradient descent [162], which iteratively updates the model parameters to reduce the loss. Since an over-optimisation of the model on the training data can lead to poor

generalisation on unseen data (the model is said to be *overfitting*), a regularisation term is often added to the loss function to penalise overly complex models. The training process continues until the model reaches a satisfactory level of performance on the training data, or until its performance does not improve further.

### Cross-validation

With the strategy defined above to optimise the hyperparameters and train the model, the training dataset is divided into two subsets: the validation set and the training set, which are used for hyperparameter tuning and training, respectively. When small datasets are involved, this division can lead to a suboptimal model, as the model's performance can be highly dependent on the specific examples in the training and validation sets. Furthermore, this approach limits the amount of data available for training the model, which can lead to poor generalisation. To mitigate this issue, cross-validation [163] is often employed. This term refers to a set of techniques that allow for a more robust estimate of the model's generalisation performance by using the entire dataset for training and validation. The most common cross-validation technique is the  $k$ -fold cross-validation. It consists in dividing the training sample into  $k$  subsets of equal size, called *folds*. Then, the ML algorithm is trained  $k$  times, each time using  $k - 1$  folds as training set, while the remaining fold is used as validation set. The model's performance is then averaged over the  $k$  folds to obtain a more robust estimate of this quantity. This operation is repeated for each hyperparameter configuration to be considered. The hyperparameter configuration minimising the loss function or other metrics used to evaluate the model's performance is then chosen as the optimal configuration.

### Testing

After the model has been trained, its performance is evaluated on a dataset that was not used during the training process, known as the test set. Like the training and validation sets, also the test set contains labelled examples, but the labels are not provided to the model. While the model is optimised to minimise the loss function during the training phase, the test set is used to estimate the model's generalisation error, i.e., how well the model performs on unseen data. The model's performance is evaluated using metrics that are specific to the task at hand, such as accuracy for classification tasks, or Mean Squared Error (MSE) for regression tasks. Once the model achieves satisfactory performance on the test set, it is ready to be used for making predictions on unlabelled data.

## 5.2 $D_s^+$ and $D^+$ selection using Machine Learning

The task of extracting  $D_s^+$ - and  $D^+$ -meson signals from the vast combinatorial background is a challenging one, due to the large amount of background compared to signal. It is however an excellent example of classification problem, and ML algorithms can therefore be exploited to enhance the efficiency of the selection.

### 5.2.1 Data preparation

In order to train a ML model, a labelled dataset with a well-defined set of features is required. The dataset used for training the ML algorithms employed in this Thesis is composed of a number of signal and background examples. To obtain a pure sample of signal candidates, Monte Carlo (MC) simulations are used to generate  $D_s^+$  and  $D^+$  mesons and transport their decay products through a simulation of the ALICE apparatus. Proton-proton collisions are simulated using the PYTHIA 8 event generator [98] with colour-reconnection Mode 2 [104], and the generated particles are propagated through the ALICE detector using the GEANT4 transport simulation toolkit [164]. A dataset enriched of heavy-flavour hadrons is obtained by only selecting (*triggering*) events where at least a  $c\bar{c}$  or  $b\bar{b}$  pair is produced. The produced  $D_s^+$  and  $D^+$  mesons are then forced to decay into the studied decay channels, i.e.:  $D_s^+ \rightarrow \phi\pi \rightarrow K^+K^-\pi^+$ ,  $D_s^+ \rightarrow K^+K^{*0} \rightarrow K^+K^-\pi^+$ , characterised by the same  $K^+K^-\pi^+$  final state,  $D^+ \rightarrow \pi^+K^-\pi^+$ , and  $D^+ \rightarrow \phi\pi \rightarrow K^+K^-\pi^+$ .

Only prompt and non-prompt  $D_s^+ \rightarrow \phi\pi \rightarrow K^+K^-\pi^+$  decays are used to train the model, as  $D^+$  mesons decay into the same final state as  $D_s^+$  mesons, and selections optimised to reconstruct  $D_s^+$  mesons are also effective for  $D^+$  mesons.

Background candidates are obtained from real data, as MC simulations may not be able to reproduce the complexity of the jet fragmentation and soft processes occurring in the underlying event, therefore limiting the dependence on MC simulations. Background examples are obtained by selecting candidates from a subsample of the full data sample (corresponding to its 3%) in an invariant-mass region away from both the  $D_s^+$ - and  $D^+$ -meson mass peaks, where  $1.7 < M < 1.75 \text{ GeV}/c^2$  or  $2.1 < M < 2.15 \text{ GeV}/c^2$ , as shown in Fig. 5.1.

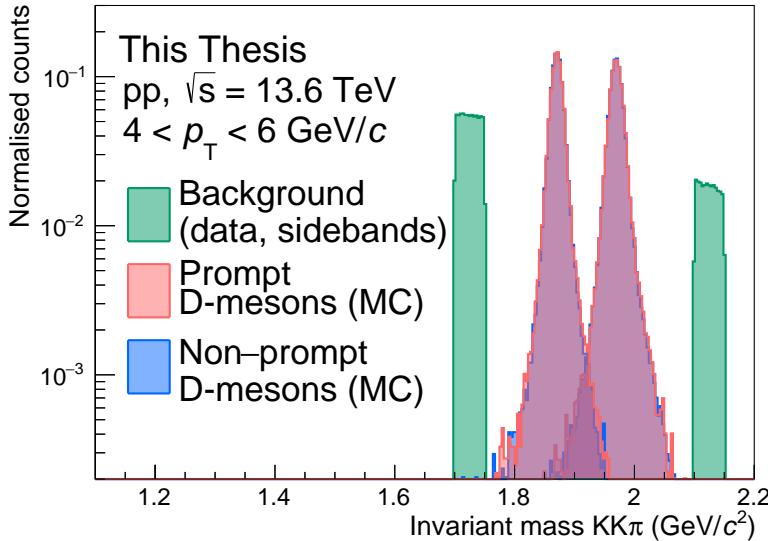


Figure 5.1: Invariant mass distribution of prompt and non-prompt D mesons (red and blue, respectively), taken from MC simulations, and of the background candidates taken from real data used to train the ML model (green) in the  $4 < p_T < 6 \text{ GeV}/c$  interval. Background candidates are selected in the  $1.7 < M < 1.75 \text{ GeV}/c^2$  or  $2.1 < M < 2.15 \text{ GeV}/c^2$  invariant-mass interval.

To reduce the amount of combinatorial background in the data samples prior to the application of ML-based selections, some loose selections were applied. These pre-selections are very effective at rejecting the combinatorial background, while preserving the selection efficiency for  $D_s^+$  and  $D^+$  mesons, and are reported in Table 5.1. These criteria include selections on the invariant mass of the  $\text{KK}\pi$  triplet to reduce the number of candidates to those necessary for the invariant mass analysis. To separate signal and background, topological variables such as the decay length  $L$  and the cosine of the pointing angle  $\cos \theta_p$ , introduced in Sec. 4.3.2, and PID variables, introduced in Sec. 4.3.4 were applied. The selection criterion on the  $n\sigma_{\text{TOF}}$  variable is applied only when a TOF cluster is matched to the track, i.e., tracks with no TOF signal are not rejected. In addition, a logical OR is applied to the conditions on the PID variables, meaning that a candidate is accepted if, for all daughter tracks, at least one of the conditions on the  $n\sigma_{\text{TOF}}$  and  $n\sigma_{\text{TPC}}$  is satisfied. A selection is also applied to the  $\chi^2_{\text{PCA}}$  variable, quantifying the dispersion of the decay tracks around the secondary vertex, normalised to their uncertainty.

Table 5.1: Topological and kinematical pre-selections applied to the  $D_s^+$ - and  $D^+$ -mesons candidates. The selection criterion on the  $n\sigma_{\text{TOF}}$  variable is applied only when a TOF cluster is matched to the track, i.e., tracks with no TOF signal are not rejected. In addition, a logical OR is applied to the conditions on the  $n\sigma_{\text{TOF}}$  and  $n\sigma_{\text{TPC}}$  PID variables, meaning that a candidate is accepted if, for all daughter tracks, at least one of the conditions is satisfied.

Variable	$p_T$ interval ( $\text{GeV}/c$ )		
	0.0–1.5	1.5–12	12–24
$ M - M_{D_s^+}^{\text{PDG}}  (\text{MeV}/c^2) <$	400	400	400
$p_T(\pi, K) (\text{GeV}/c) >$	0.3	0.4	0.4
Decay length (cm) >	0.02	0.02	0.03
Normalized decay length XY >	2	2	2
$\cos \theta_p >$	0.85	0.85	0.85
$\cos \theta_p^{xy} >$	0.85	0.85	0.85
$M_{\text{inv}}^\phi - M_{\text{inv}}^{\phi \text{ PDG}} (\text{MeV}/c^2) <$	20	20	20
$\chi^2_{\text{PCA}} <$	10	10	10
$ n\sigma_{\text{TPC}}  <$	5	5	5
$ n\sigma_{\text{TOF}}  <$	5	5	5

Labels are assigned as a numerical value to each candidate, with 0 indicating a background candidate, 1 a prompt  $D_s^+$  meson, and 2 a non-prompt  $D_s^+$  meson.

The dataset is then divided into two different subsamples. The first comprehends 80% of the data, and is used to train the model, while the remaining 20% is used to test its performance. In addition, since the D-meson decay topology can significantly differ depending on the  $p_T$  of the meson due to different Lorentz boosts, the dataset is divided into several  $p_T$  intervals, and the model is trained and tested separately for each of them. To achieve a better performance of the ML models, they are

Table 5.2: Number of candidates within the  $p_T$  intervals used to train and test the model.

$p_T$ (GeV/ $c$ )	Candidates
0–1.5	$\sim 4.6 \times 10^3$
1.5–2	$\sim 6.1 \times 10^3$
2–3	$\sim 26 \times 10^3$
3–4	$\sim 34 \times 10^3$
4–5	$\sim 31 \times 10^3$
5–6	$\sim 24 \times 10^3$
6–8	$\sim 22 \times 10^3$
8–12	$\sim 10 \times 10^3$
12–24	$\sim 2.6 \times 10^3$

trained in broader  $p_T$  intervals than those used for the analysis, to ensure that enough data is available to train a well-performing model. The total number of candidates available for training and testing the model is reported in Table 5.2 for the considered  $p_T$  intervals.

To produce a balanced dataset, the number of candidates in each class is equalised to the number of examples in the minority class. This is achieved by randomly selecting a subset of the majority classes. The balanced dataset is then used to train the model.

The choice of features used to separate signal from background is crucial, as they must be able to discriminate between signal and background candidates, and must be chosen in such a way that no bias is introduced in the final result. The variables used to train the model were introduced in Chapter 4, and are a mixture of topological, kinematical, and PID variables. The key idea is to exploit the displaced topology of the D-meson decay, which is a distinctive feature of the signal candidates, the kinematic properties of the D-meson decay, and the PID information of the daughter tracks to discriminate between signal and background candidates. The features used to train the model are reported in Table 5.3. The number in parenthesis after  $n\sigma$  indicates the prong identification number. The tracks are ordered based on their charge sign, so that the first and third prongs (prongs 0 and 2) are the like-sign tracks (i.e., their charge sign is the same as that of the D meson), while the second prong (prong 1) is the opposite-sign track.

The invariant mass of the candidate and its  $p_T$  are excluded from the model training to prevent any bias in the final results. Including these variables would cause the model to preferentially select candidates within a specific invariant mass range (corresponding to that of  $D_s^+$  and  $D^+$  mesons) or  $p_T$ , thereby skewing both candidate selection and the  $p_T$  distribution of the final sample, resulting in a biased  $p_T$ -differential yield. While it is possible to correct for a  $p_T$ -dependent selection using MC simulations, this approach is not taken here to avoid biases from any inaccuracies in the  $p_T$  distribution of the D mesons in the simulations.

Some of the variables used to train the model may be correlated with the invariant

Table 5.3: Candidate features used to train the ML model.

Variable
Cosine of pointing angle ( $\cos \theta_p$ )
Cosine of pointing angle XY ( $\cos \theta_p^{xy}$ )
Decay length ( $L$ )
Decay length XY ( $L^{xy}$ )
Candidate impact parameter XY ( $d^{xy}$ )
$ \cos^3 \theta'(K) $
Prong 0 impact parameter XY ( $d_0^{xy}$ )
Prong 1 impact parameter XY ( $d_1^{xy}$ )
Prong 2 impact parameter XY ( $d_2^{xy}$ )
$n\sigma_{\text{comb}}^\pi(0)$
$n\sigma_{\text{comb}}^\pi(1)$
$n\sigma_{\text{comb}}^\pi(2)$
$n\sigma_{\text{comb}}^K(0)$
$n\sigma_{\text{comb}}^K(1)$
$n\sigma_{\text{comb}}^K(2)$

mass of the candidate, and the ML may learn to discriminate the signal from the background by exploiting this correlation with the  $D_s^+$  meson mass and transverse momentum, rather than the physical properties of the signal and background. To exclude this possibility, the correlation between the features used to train the model is studied. To quantitatively describe the correlation between the variables, the Pearson correlation coefficient  $\rho$  is evaluated for each pair of variables. It is defined as the ratio between the covariance of two variables  $x$  and  $y$  and the product of their standard deviations,  $\rho(x, y) = \text{cov}(x, y) / (\sigma_x \sigma_y)$ . It expresses the strength and direction of a linear correlation between two variables, ranging from  $\rho = 1$  (perfect positive linear correlation) to  $\rho = -1$  (perfect negative linear relationship).  $\rho = 0$  indicates the absence of linear correlation.

The correlation matrix of the features used to train the model is shown in Fig. 5.2 for the prompt  $D_s^+$ , non-prompt  $D_s^+$  and background classes, in the  $2 < p_T < 3$  GeV/ $c$  interval of D-meson transverse momentum. The correlation with the invariant mass and the transverse momentum is also reported. The Pearson coefficient is encoded in the colour of the cell, with red indicating a positive correlation, blue a negative correlation, and grey no linear correlation. The correlation matrix shows that the variables used to train the model are not linearly correlated with the invariant mass of the candidate, suggesting that a ML model should not modify the invariant-mass distribution of the selected candidates.

Variables carrying similar physical information, such as those related to the candidate decay length, pointing angle, and impact parameter, are strongly correlated among each other, as expected. Different degrees of correlation between the same variable pairs are observed for the different classes. The ML model can exploit these differences to discriminate between the three classes of candidates.

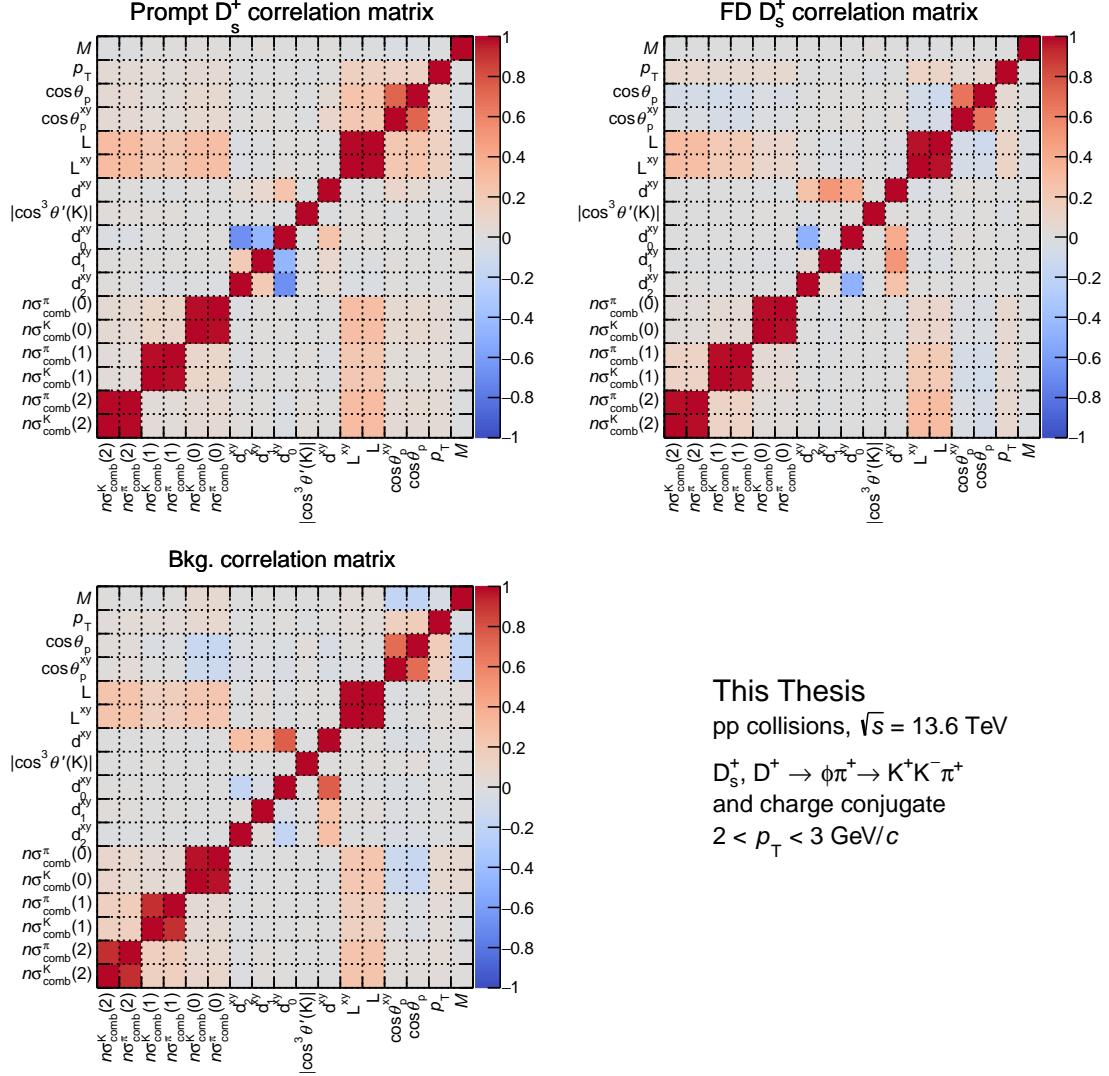


Figure 5.2: Correlation matrices of the features used to train the ML model for prompt  $D_s^+$  (top-left), non-prompt  $D_s^+$  (top-right), and background (bottom-left) candidates in the  $2 < p_T < 3 \text{ GeV}/c$  interval. The correlation with the invariant mass and the transverse momentum is also reported. The Pearson coefficient is encoded in the colour of the cell, with red indicating a positive correlation, blue a negative correlation, and grey no linear correlation.

### 5.2.2 Boosted Decision Trees

Once the training dataset has been composed and the features have been selected, the ML architecture has to be chosen. Several algorithms are available, each with its own strengths and weaknesses. The choice of the algorithm depends on the specific problem to solve, the size of the dataset, and the computational resources available.

Boosted decision trees [165, 166] (BDTs) are a family of machine learning algorithms employed in different fields, including high-energy physics. Their building blocks are decision trees, which are a versatile type of supervised learning algorithm that can be used for both classification and regression tasks. A decision tree is made of many *nodes*, each containing conditions that split the data into two [167] or more [168] children nodes. The first node of the tree, which receives all the data, is called *root* node, while nodes that do not further split the data are called *leaves*, and contain the output of the tree. The model is trained by considering the Gini index, which measures the impurity of the node:

$$G = 1 - \sum_{i=1}^n p_i^2 \quad ,$$

where  $p_i$  is the fraction of samples in the node that belong to the class  $i$ . The Gini index therefore provides an indication of the quality of the split, with  $G = 0$  indicating a perfect split. A commonly used algorithm to build *binary* decision trees (i.e., each node contains binary-output conditions, and is split into two children nodes) is the *Classification And Regression Tree* (CART) algorithm [167], which recursively splits the dataset into subsets based on a single feature  $k$  and a threshold  $t_k$  that minimises the impurity of the subsets (weighted by their size). The cost function that the algorithm tries to minimise is given by

$$J(k, t_k) = \frac{m_{\text{left}}}{m} G_{\text{left}} + \frac{m_{\text{right}}}{m} G_{\text{right}} \quad ,$$

where  $m_{\text{left}}$  and  $m_{\text{right}}$  are the number of samples in the left and right nodes, respectively, summing up to the total number of samples  $m$ , and  $G_{\text{left}}$  and  $G_{\text{right}}$  are the Gini indices of the left and right nodes. The tree is grown until a stopping criterion is met, such as a maximum depth, a minimum number of samples in a node, or a minimum impurity decrease. These are all hyperparameters that can be tuned to optimise the model's performance.

Given their simplicity, decision trees are fairly easy to interpret, and are often called *white-box* models (in contrast to BDTs and neural networks, where the decision-making process is less transparent, therefore called *black-box* models). An additional strength of decision trees is that they require very little data preparation, e.g., they do not require feature scaling or centering, making them a very powerful yet simple tool for data analysis. However, they are prone to overfitting, as they can grow to a large depth, capturing the noise in the training data. To mitigate this issue, their depth can be constrained, but this may lead to a model with limited discrimination power. To build a robust model with a good discrimination power, ensemble methods may be used. Several decision trees can be trained, and the final prediction is obtained by combining the outcome of all the trees.

## XGBoost

In this work, the Extreme Gradient Boosting [169] (XGBoost) Boosted Decision Trees (BDT) algorithm is used. It has achieved state-of-the-art results in a number of machine learning and data mining challenges (for example in Ref. [170]). In addition, this algorithm, which is available as an open-source package, can be easily parallelised on CPUs and GPUs [171], thereby reducing the training and application time.

The term *boosting* refers to any ensemble method combining several weak learners into a strong learner. The general idea of most boosting methods is to train many predictors sequentially, each trying to correct its predecessor [172]. The function estimate  $\hat{f}(x)$  is parametrised with an additive functional form:

$$\hat{f}(x) = \sum_{k=0}^M \hat{f}_k(x) \quad ,$$

where  $M$  is the number of iterations,  $\hat{f}_0(x)$  is the initial prediction, and  $\hat{f}_i(x)$  is the function increment at the  $i$ -th iteration, also called *boost*. To reduce the loss function, a new weak learner, whose functional form is parametrised as  $h(x, \theta)$ , can be added to the ensemble:

$$\hat{f}_t(x) \leftarrow \hat{f}_{t-1}(x) + \rho_t h(x, \theta_t) \quad .$$

$\rho_t$  is the step size, which is optimised for each iteration  $t$ , together with the parameters  $\theta_t$  of the weak learner:

$$(\rho_t, \theta_t) = \arg \min_{\rho, \theta} \sum_{i=1}^N L(y_i, \hat{f}_{t-1}(x_i) + \rho h(x_i, \theta)) \quad ,$$

where  $L$  is the loss function, and  $y_i$  is the true label of the  $i$ -th example. Despite having a well-defined set of equations for minimising the loss function, the optimisation of the parameters is not trivial, as the loss function is non-convex and the search space is high-dimensional. Therefore, the optimisation is usually performed using a gradient-based algorithm [165, 173], where  $h(x, \theta_t)$  is chosen as the most parallel function to the negative gradient of the loss function with respect to the previous prediction  $g_t(x)$ :

$$g_t(x) = E_y \left[ \frac{\partial L(y, \hat{f}_{t-1}(x))}{\partial \hat{f}_{t-1}(x)} \middle| x \right] \quad ,$$

where  $E_y$  is the expectation over the true labels. The parameters are then optimised by minimising the difference between the negative gradient and the weak learner prediction:

$$(\rho_t, \theta_t) = \arg \min_{\rho, \theta} \sum_{i=1}^N [-g_t - \rho h(x_i, \theta)]^2 \quad .$$

Through the iterative addition of weak learners, the loss function is reduced and the model learns the complex patterns of data. The final prediction is obtained

by summing the predictions of all the weak learners. In the XGBoost algorithm, the weak learners are decision trees. The output consists of a numerical score for each class, ranging from 0 to 1 and summing up to unity. Each score represents the confidence of the model in the prediction, which can be interpreted as the probability of the example belonging to that class.

### 5.2.3 Tuning the model’s hyperparameters

The XGBoost algorithm has several hyperparameters [174] that can be tuned to optimise the model’s performance. The most important hyperparameters are:

- `eta` or `learning_rate`, which is the step size shrinkage of the gradient descent algorithm. To reduce the risk of overfitting, this factor multiplies the weak-learner prediction ( $\rho_t h(x_i, \theta) \rightarrow \text{eta} \cdot \rho_t h(x_i, \theta)$ ), and is usually set to a small value, such as 0.1;
- `max_depth`, which is the maximum depth of a single decision tree. A large depth can lead to overfitting, while a small depth can lead to a model with limited discrimination power. Usually, this parameter is set to around 5;
- `n_estimators`, which defines the number of trees to train. A large number of weak learners can lead to overfitting, while a small number can lead to a model with limited discrimination power. Usually, the number of weak learners is set to around 1000;
- `subsample`, which is the fraction of the training data to be used to train each tree at each iteration;
- `min_child_weight`, which is the minimum sum of instance weight needed in a child. It is related to the purity in a node, and it is used to stop the tree growth;
- `colsample_bytree`, which is the fraction of features to be used to train each tree at each iteration;
- `tree_method`, which defines the algorithm used to build the trees. The `hist` option uses an optimised histogram-based algorithm and is usually the fastest.

The hyperparameters are optimised using the Optuna framework [175], which proved to be a powerful tool thanks to its state-of-the-art algorithms for sampling the hyperparameter space and for efficiently pruning unpromising trials. The Tree-Structured Parzen Estimator [176], is used in this Thesis. It is a Bayesian optimisation algorithm able to explore the hyperparameter space efficiently. The aim of a Bayesian optimisation is to maximise (or minimise, depending on the task) an objective function  $f(\mathbf{x})$  by iteratively sampling a bounded hyperparameter space,  $\chi$ . The algorithm builds a probabilistic model of the objective function, and uses it to decide which hyperparameters to sample next. The model is updated at each iteration, and the hyperparameters that are most likely to improve the model’s performance are sampled. The Optuna algorithm is also able to prune unpromising

Table 5.4: Optimised hyperparameter configuration for some representative  $p_T$  intervals considered in the model training.

Hyper-parameter	$p_T$ interval ( $\text{GeV}/c$ )					
	0–1.5	1.5–2	2–3	4–5	6–8	12–24
<code>max_depth</code>	3	3	3	3	3	3
<code>learning_rate</code>	0.04	0.068	0.065	0.091	0.070	0.030
<code>n_estimators</code>	473	339	1352	1256	1142	1188
<code>min_child_weight</code>	1	3	10	10	3	5
<code>subsample</code>	0.87	0.95	0.84	0.95	0.81	0.88
<code>colsample_bytree</code>	0.91	0.98	0.90	0.96	0.88	0.89
<code>lambda</code>	$8.0 \times 10^{-4}$	$4.8 \times 10^{-4}$	$9.1 \times 10^{-4}$	$3.0 \times 10^{-4}$	$1.9 \times 10^{-4}$	$6.7 \times 10^{-4}$
<code>tree_method</code>	hist	hist	hist	hist	hist	hist

trials, reducing the computational cost of the optimisation. The optimisation is performed using a 5-fold cross-validation, and the hyperparameters that maximise the macro-averaged one-vs-one Receiver Operating Characteristic Area Under the Curve (ROC AUC) metric (described in detail in Sec. 5.2.4) are chosen as the optimal configuration. The hyperparameters optimised for the XGBoost model are reported in Table 5.4 for some representative  $p_T$  intervals considered in the model training. An additional hyperparameter, `lambda`, which is the  $L_2$  regularisation term, is also optimised. It helps to prevent overfitting by penalising overly complex models. The optimal hyperparameters are then used to train the model on the full training dataset.

#### 5.2.4 Evaluation of the model’s performance

After training the model, its performance is evaluated on the test dataset. The model’s performance can be assessed using a *confusion matrix*, which summarises the number of examples for a given class (the true label) that are classified by the model as belonging to any of the available classes (the predicted label). A good model should provide a high number of correctly-classified examples (reported on the diagonal of the confusion matrix), and a low number of misclassified examples (off-diagonal elements of the confusion matrix). The confusion matrix also allows an understanding of which classes are more difficult to classify, and which classes are more likely to be confused with each other. An example of a confusion matrix is shown in Fig. 5.3 for the XGBoost model trained on the  $2 < p_T < 3 \text{ GeV}/c$  interval.

Despite providing a lot of information on the model’s performance, more concise metrics of the model’s performance are usually used, for a more direct comparison between different models. In addition, the confusion matrix provides a threshold-dependent measure of the model’s performance, as the classification threshold (i.e., the threshold on the model output that defines the separation between the different classes) can be varied to increase the number of correctly classified signal candidates at the expense of the number of correctly classified background candidates, and vice versa.

In binary classification tasks, where only two classes are available (a positive and

		True labels		
		Bkg.	Prompt $D_s^+$	FD $D_s^+$
Predicted labels	Bkg.	3901	273	584
	Prompt $D_s^+$	200	3772	982
	FD $D_s^+$	187	383	2795

Figure 5.3: Confusion matrix for the BDT model trained in the  $2 < p_T < 3$  GeV/ $c$  interval. Candidates are classified as the class with the highest score.

a negative class), several metrics can be defined from the elements of the confusion matrix. The  $2 \times 2$  confusion matrix contains four entries: the true positives (TP), which are the number of correctly classified positive candidates, the false positives (FP), which are the number of negative candidates being mistakenly classified as positives, and the analogously defined true negatives (TN) and false negatives (FN). One of the most used tools for binary classifiers is the *Receiver Operating Characteristic* (ROC) curve, which represents the true positive rate (TPR) against the false positive rate (FPR) for different threshold values. The TPR is the fraction of correctly classified positive candidates ( $\text{TPR} = \text{TP}/(\text{TP} + \text{FN})$ ), while the FPR is the fraction of incorrectly classified negative candidates ( $\text{FPR} = \text{FP}/(\text{FP} + \text{TN})$ ).

The output of the ML model is a single value ranging from 0 to 1 for a binary classification task, and can be interpreted as the probability of the candidate belonging to the positive class. If positive candidates are selected as those with a score greater than a certain threshold  $t$ , when  $t = 0$  all candidates are classified as positive, and both the TPR and FPR will be equal to 1. On the other hand, if  $t = 1$ , no candidate is classified as positive, and the TPR and FPR will both equal 0. Therefore, the different values of  $t$  between 0 and 1, which yield different TPR and FPR values, will trace the ROC curve, going from the point (1,1) to (0,0) as  $t$  increases. The ROC *Area Under the Curve* (AUC), is used to measure the model's ability to discriminate between positive and negative candidates, for any given threshold. The ROC AUC ranges from 0 to 1. A random classifier has a ROC AUC of 0.5, while a perfect classifier has a ROC AUC of 1. The ROC AUC is a threshold-independent measure of the model's performance, and is often used to compare different models.

In a multiclass classification task, where more than two classes are available, a generalisation of the ROC curve and the ROC AUC metric is required. In this case, the *One-vs-One* ROC curve [177] can be defined as a plot of the TPR against the FPR for a given pair of classes. The One-vs-One ROC AUC can be averaged to

the *macro-averaged* One-vs-One ROC AUC, which is the average of the ROC AUC for each pair of classes and can provide a measurement of the model’s ability to discriminate between all the classes. For a classification problem with  $N$  classes, there are  $N(N - 1)$  possible pairs of ordered classes, and therefore of One-vs-One ROC curves.

The One-vs-One ROC curves for the model trained on the  $2 < p_T < 3 \text{ GeV}/c$  interval are shown in Fig. 5.4. The ROC AUC is calculated for each class pair, and is reported in the legend. Given that the classification task is a three-class problem, the One-vs-One ROC AUC is calculated for the three possible pairs of classes. Furthermore, despite showing an approximate symmetry around the line  $\text{TPR} = 1 - \text{FPR}$ , the ROC curve for the A-vs-B pair is not exactly the symmetrical curve of the ROC curve of the B-vs-A pair when more than two classes are considered, due to the different probability of misclassification with the remaining classes. Therefore, two curves are shown for each pair of classes, and present different ROC AUC values. The metric is evaluated on both the training and test sets to test the model’s generalisation power. The model’s performance is excellent, with a macro-averaged One-vs-One ROC AUC value of 0.965, very close to that of an ideal classifier of 1. In addition, little overfitting is observed, as the ROC AUC values for the training and test sets are similar. The macro-averaged One-vs-One ROC AUC presents similar values in the other  $p_T$  intervals, indicating that the model’s performance is stable across the studied  $p_T$  range. A slight increase in the ROC AUC of a few percent is observed for increasing  $p_T$ , due to the more displaced decay topology, which facilitates the separation of the three classes. In the highest  $p_T$  intervals the ROC AUC decreases due to the limited number of candidates in the training sample, which can lead to a less performant model. The model is then used to select  $D_s^+$ - and  $D^+$ -meson candidates from the full dataset.

In addition to the ROC AUC, the model’s performance can be evaluated by studying the distribution of the probability of belonging to a given class assigned to labelled candidates. The score distributions for the model trained on the  $2 < p_T < 3 \text{ GeV}/c$  interval are shown in the three panels of Fig. 5.5 for the background, prompt  $D_s^+$ -meson, and non-prompt  $D_s^+$ -meson classes. For each class, the score distribution is shown for candidates belonging to the three different classes and to both the training and test sets. The distribution of the true background score provides interesting information on the model’s performance. The score distribution for the background candidates peaks at high values, while the score distribution for the signal candidates (both prompt and non-prompt  $D_s^+$  meson) peaks at low values. This highlights that the model has effectively learned to discriminate between signal and background candidates, with good separation power. Furthermore, the score distributions for the training and test sets are fairly similar, indicating that the model generalises well to unseen data. Since non-prompt  $D_s^+$  mesons present a more displaced topology than promptly produced ones, due to the large lifetime of beauty-hadrons, the separation between non-prompt  $D_s^+$  mesons and background is noticeable in the non-prompt  $D_s^+$ -meson score distribution, where the score of true background candidates peaks at zero and the one of non-prompt  $D_s^+$ -meson candidates peaks at one. The distribution of prompt  $D_s^+$  mesons, which have a smaller displacement as compared to non-prompt ones falls in between those of background

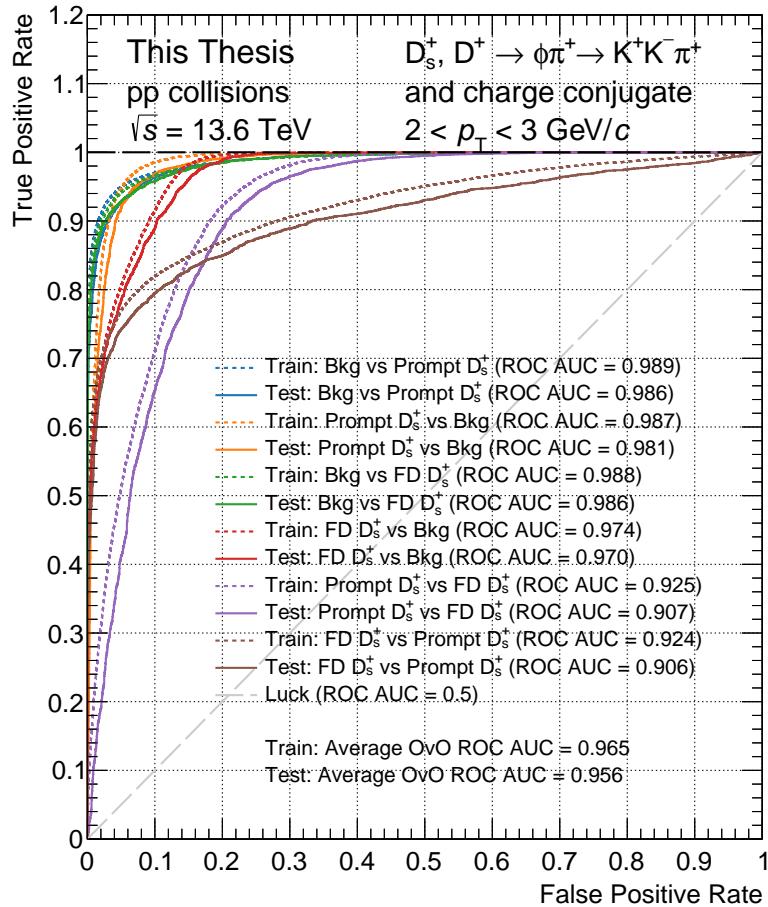
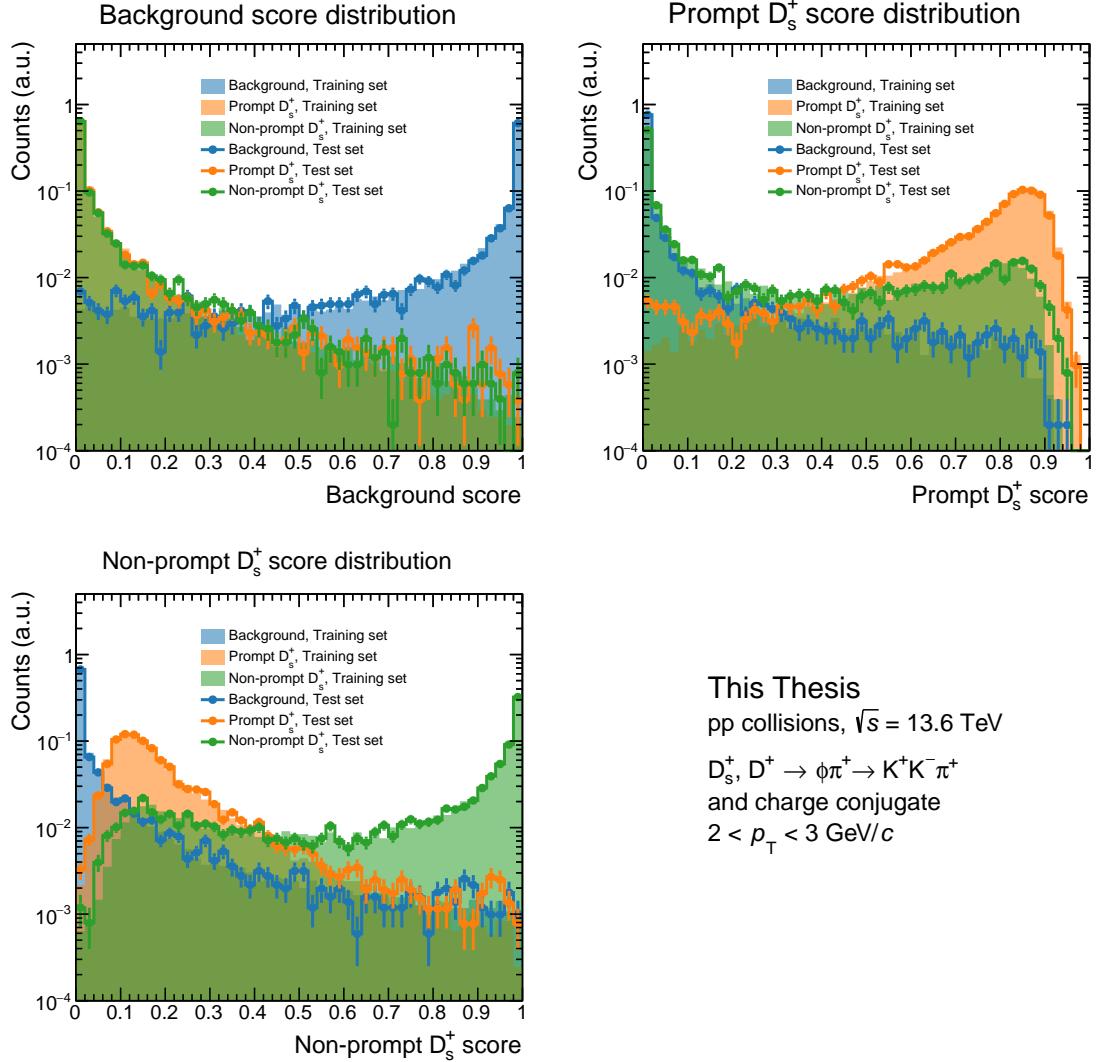


Figure 5.4: ROC curves for the model trained on the  $2 < p_T < 3 \text{ GeV}/c$  interval. The One-vs-One ROC AUC metric is calculated for each class pair and reported in the legend.

and non-prompt  $D_s^+$  mesons. Lastly, the separation between the three classes is less pronounced in the prompt  $D_s^+$ -meson score distribution, where the prompt  $D_s^+$ -meson distribution peaks at values significantly lower than one. Similar trends are observed in the other studied  $p_T$  intervals.

### 5.2.5 Interpretation of the model’s output: Feature importance

The usage of ML algorithms usually provides a better performance in terms of signal-to-background separation as compared to approaches based on “rectangular” cuts, but it also introduces a level of complexity in the selection process. One of the most difficult aspects of using ML models is the interpretation of their output. To understand how the model makes its decisions, the feature importance can be studied. This allows the understanding of which features are more important for the model’s decision-making process, and the optimisation of the feature selection. In addition, the feature importance can be used to check whether the model is learning



This Thesis  
 pp collisions,  $\sqrt{s} = 13.6 \text{ TeV}$   
 $D_s^+, D^+ \rightarrow \phi\pi^+ \rightarrow K^+K^-\pi^+$   
 and charge conjugate  
 $2 < p_T < 3 \text{ GeV}/c$

Figure 5.5: Score distributions for the model trained on the  $2 < p_T < 3 \text{ GeV}/c$  interval. The score distributions related to the probability of belonging to the background (top-left panel), prompt  $D_s^+$ -meson (top-right panel), and non-prompt  $D_s^+$ -meson (bottom-left panel) classes are shown. For each class, the score distribution is shown for candidates belonging to the different classes and to both the training (filled area) and test sets (markers).

on the correct features in terms of the physics of the problem.

One of the most used algorithms for feature importance studies is the SHapley Additive exPlanations [178] (SHAP) algorithm. SHAP is a game-theoretic approach to explain the output of any machine learning model. It is based on the Shapley value [179] from cooperative game theory, which requires retraining the model on all feature subsets  $\mathcal{S} \subseteq \mathcal{F}$ , where  $\mathcal{F}$  is the set of all features. An importance value is assigned to each feature  $i$ , representing the effect on the model prediction of including that feature. To compute this effect, a model  $\hat{f}_{\mathcal{S} \cup \{i\}}$  is trained with that feature present, and another model  $\hat{f}_{\mathcal{S}}$  is trained with the feature withheld. Then, predictions from the two models are compared on the current input  $\hat{f}_{\mathcal{S} \cup \{i\}}(x) - \hat{f}_{\mathcal{S}}(x)$ . The Shapley values are then computed as the weighted average of all possible differences:

$$\phi_i = \sum_{\mathcal{S} \subseteq \mathcal{F} \setminus \{i\}} \frac{|\mathcal{S}|!(|\mathcal{F}| - |\mathcal{S}| - 1)!}{|\mathcal{F}|!} \left[ \hat{f}_{\mathcal{S} \cup \{i\}}(x) - \hat{f}_{\mathcal{S}}(x) \right] .$$

Since most models cannot handle arbitrary patterns of missing input values,  $\hat{f}(z_S)$  is approximated with  $E[\hat{f}(z)|z_S]$ , where  $z_S$  is the input missing the features in  $\mathcal{S}$ . SHAP values therefore explain how to get from the base value  $E[\hat{f}(z)]$ , which would be predicted if no features were known, to the output  $\hat{f}(x)$ .

A beeswarm-style SHAP feature importance plot for the prompt  $D_s^+$ -meson probability predicted by the model trained in the  $2 < p_T < 3$  GeV/ $c$  interval is shown in Fig. 5.6. In this plot, each instance is represented by a single dot on each feature row. The position of the dot along the horizontal axis is determined by the SHAP value of that feature, and the dots “pile up” along each feature row, to provide information on the distribution of the SHAP values. The colour of the dot represents the value of the feature, with blue indicating a low value and red a high value. Positive values indicate that the feature is pushing the model’s prediction towards the prompt  $D_s^+$ -meson class, while negative values indicate that the model is less likely to classify the candidate as a prompt  $D_s^+$  meson. Feature rows are ordered from top to bottom based on the mean absolute value of the SHAP values for that feature, with the most important features, i.e., those with the highest impact on the model’s decision, at the top. The most important features are the cosine of pointing angle  $\cos \theta_p$ , the decay length  $L$ , the decay length in the XY plane  $L^{xy}$ , the absolute value of the cosine cubed of the K- $\pi$  angle in the KK rest frame  $|\cos^3 \theta'(K)|$ , and the PID information on the prong 1. As discussed in Chapter 4, the first three features are related to the displaced decay topology of  $D_s^+$  mesons, and are therefore expected to be the most important variables in the model decisions. It is also expected that the prong 1 PID information resulted as the most important PID variable, as this is the opposite sign track, which is always a kaon in the considered decay channel. On the contrary, prongs 0 and 2 could be either kaons or pions, resulting in a lower importance of the PID information for these prongs. This check provides a good indication that the model is making decisions based on features that are expected to be relevant for the analysed physics process.

For each  $p_T$  interval, the SHAP values for the different features employed in the model can be evaluated for the three classes. The overall feature impact on the model’s decision can be evaluated by computing the average absolute SHAP value

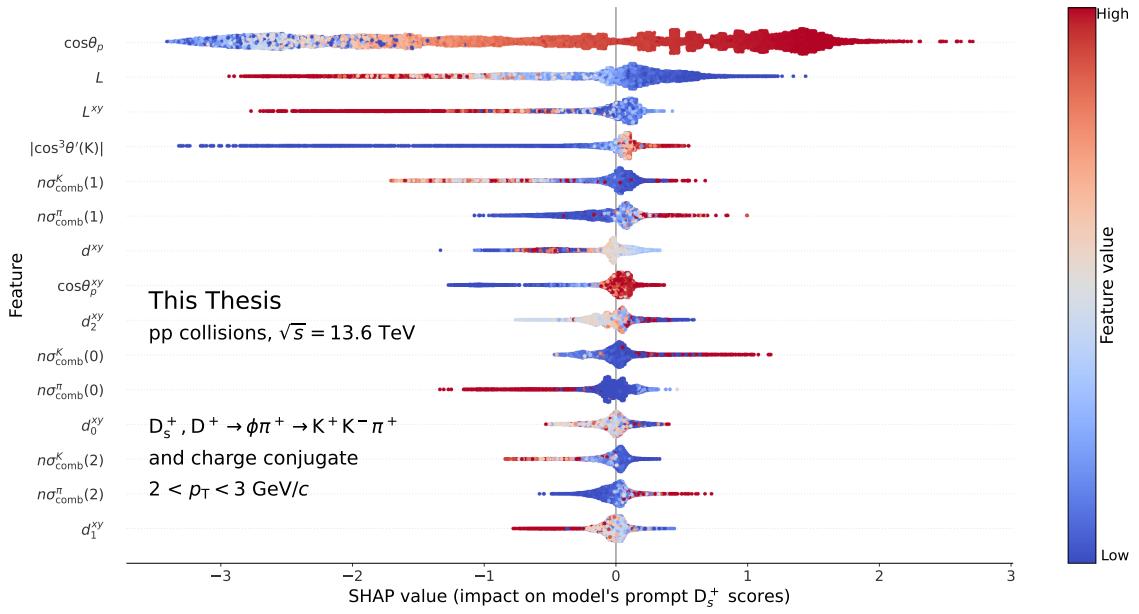


Figure 5.6: Beeswarm-style SHAP feature importance plot for the prompt  $D_s^+$ -meson score of the XGBoost model trained on the  $2 < p_T < 3$  GeV/ $c$  interval.

for each feature. The average  $|SHAP\ value|$  feature importance for the BDT models trained on the  $2 < p_T < 3$  GeV/ $c$  and  $8 < p_T < 12$  GeV/ $c$  intervals are shown in Fig. 5.7, providing insights on the evolution of the feature importance across the different  $p_T$  intervals. The features are ordered from top to bottom based on their importance, evaluated as the sum of the mean absolute SHAP values for the three classes. Consistently with Fig. 5.6, in the  $2 < p_T < 3$  GeV/ $c$  interval the overall most impactful features are the cosine of pointing angle  $\cos \theta_p$ , the decay length  $L$  and the PID information on the prong 1. Differences in the feature importance order are due to the different importance for the classes other than the prompt  $D_s^+$  meson reported in Fig. 5.6. In the  $8 < p_T < 12$  GeV/ $c$  interval, the decay length  $L$  and its projection on the transverse plane acquire a significantly higher importance. This is due to the fact that at higher  $p_T$ , the decay length of the  $D_s^+$  mesons is larger owing to the higher Lorentz boost, and therefore the selection of both prompt and non-prompt based on their decay length becomes more effective.

### 5.2.6 Optimisation of the selection thresholds on the BDT scores

Once the model performance has been validated, a set of selection criteria on the BDT output scores has to be chosen to select the candidates. This is a crucial step of the analysis, as it will define the signal selection efficiency and the background contamination. Since the model's output consists of a score related to the probability of belonging to each class, and the three probabilities sum up to unity, the selection criteria have a total of two degrees of freedom. A first selection is applied on the maximum probability to be a background candidate, and rejects most of the contamination from the combinatorial background. The second one is applied on

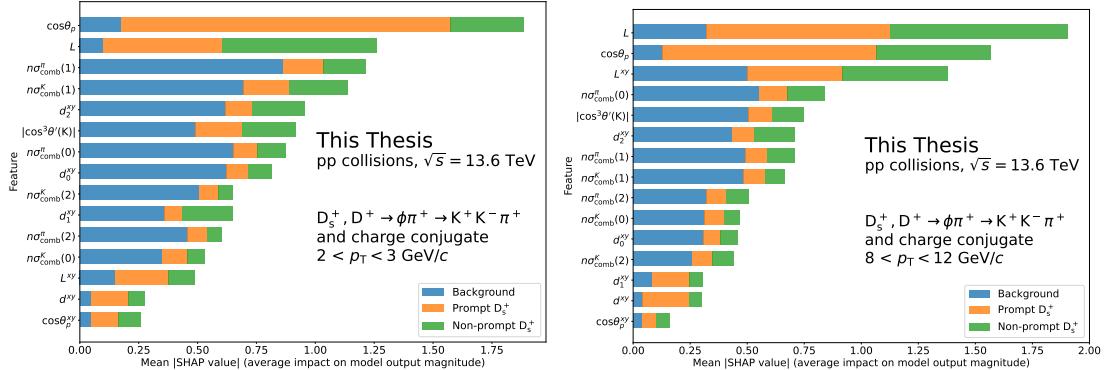


Figure 5.7: Average  $|\text{SHAP value}|$  feature importance for the XGBoost model trained on the  $2 < p_T < 3 \text{ GeV}/c$  (left panel) and  $8 < p_T < 12 \text{ GeV}/c$  (right panel) intervals. The feature importance for the background, prompt  $D_s^+$ -meson, and non-prompt  $D_s^+$ -meson classes are shown with blue, orange, and green bars, respectively.

the minimum probability of being a prompt  $D_s^+$ -meson candidate, and suppresses the signal contribution arising from non-prompt  $D_s^+$ -meson candidates.

A first indication of the optimal selection criteria can be obtained by studying the model's output distributions, e.g., those reported in Fig. 5.5. A good working point can be chosen as the point where a good separation between the three classes is achieved. For this analysis, however, the statistical significance of the extracted inclusive signal (i.e., of both prompt and non-prompt  $D$  mesons) is used to define the optimal selection criteria. For each  $p_T$  interval of the analysis, the signal  $S$  and the background  $B$  are evaluated by fitting the invariant mass distribution of candidates from the data passing the different ML selections considered in the optimisation process. Only a subset of the full dataset, corresponding to approximately 3% of the available data sample is used in this process. The working point is chosen using different considerations. Firstly, the selection criteria are chosen to maximise the statistical significance of the signal, defined as  $S/\sqrt{S+B}$ . This definition of statistical significance is equivalent to the inverse of the relative statistical uncertainty on the signal, for a Poisson distribution of candidate counts. Thus, maximising the statistical significance corresponds to minimising the relative statistical uncertainty on the signal. In order to avoid an over-optimisation of the model to enhance the statistical significance (which could bias the results if statistical fluctuations are not properly taken into account), the considered threshold values on the BDT output scores are multiples of 0.05. Additionally, the efficiency of the selection is checked to ensure that it is kept at sufficiently high levels, to reduce possible biases in the final results due to possible imperfections in the MC description of the data. A smooth increasing trend with  $p_T$  of the efficiency is also ensured to provide consistency in the tightness of the selection criteria across the different  $p_T$  intervals.

Usually, the maximisation of the statistical significance is avoided in the choice of the selection criteria, as it can lead to a bias in the final results. However, in this case, the optimisation is performed on a very small fraction of the data thanks to the large dataset available, and the bias is expected to be negligible. A different approach to select an optimal working point consists in the optimisation of

a pseudo-significance, defined as the ratio of the expected signal and the square root of the sum of the expected signal and the expected background. For each considered selection criterion, the selection efficiency is evaluated on MC simulations, and the expected signal is estimated from the D-meson production cross section provided by FONLL [63] theoretical predictions. The expected background is estimated from a fit of the invariant mass distribution of the candidates in the sidebands of the signal region. This approach avoids the introduction of a bias in the final results, as it is blind to the candidates in the signal region. However, it relies on FONLL predictions and MC simulations. To avoid the introduction of biases in the optimisation due to shortcomings of FONLL predictions or MC simulations, the optimisation of the pseudo-significance is not performed in this analysis.

A result of the working point optimisation is presented in Fig. 5.8 for the  $2.0 < p_T < 2.5 \text{ GeV}/c$  interval. The statistical significance of the signal is shown as a function of the BDT output score threshold for the probability of being a prompt  $D_s^+$  meson and a background candidate. The chosen set of selection criteria is shown with a green cross.

The optimal selection criteria for each  $p_T$  interval considered in the analysis are reported in Table 5.5. Because of the large combinatorial background and the small Lorentz boost of the  $D_s^+$  mesons, which results in a small decay length and a less effective selection based on the displaced topology, the selection criteria in the lowest  $p_T$  intervals are chosen to be stringent to extract the signal with a large enough signal-to-background ratio. At higher  $p_T$ , the selection criteria are loosened, as the larger decay length of the  $D_s^+$  mesons results in a more effective separation between signal and background candidates. In addition, at higher  $p_T$  the combinatorial background is reduced, and the selection criteria can be loosened to increase the signal efficiency. The threshold on the prompt  $D_s^+$ -meson probability is kept at around 0.2 for all  $p_T$  intervals. This selection allows a significant reduction of the contamination from non-prompt D mesons in the selected sample, while it does not considerably influence the statistical significance of the signal. In the  $5.5 < p_T < 6.0 \text{ GeV}/c$  interval, the threshold is increased to 0.25 to guarantee an increasing trend of the efficiency with  $p_T$ .

### 5.3 $D_s^+$ and $D^+$ raw-yield extraction

After the working point for the BDT algorithm has been defined, the raw yields of  $D_s^+$  and  $D^+$  mesons are extracted in each  $p_T$  interval. They are defined as the sum of particles and antiparticles and are measured in 14  $p_T$  intervals in the range  $0.5 < p_T < 24 \text{ GeV}/c$ . The raw yield is extracted by fitting the invariant mass distribution of the selected candidates.

In several analyses performed by the ALICE Collaboration during the Run 2 data-taking period [62, 180, 181], the raw yield of  $D_s^+$  mesons was extracted by fitting the invariant mass distribution of selected candidates with a probability density function constructed as the sum of a function describing the shape of the combinatorial background (usually an exponential function or a low-order ( $< 3$ ) polynomial) and of two Gaussian distributions to model the  $D_s^+$ - and  $D^+$ -meson peaks. The raw yields for the two D-meson species are then obtained by integrating the signal

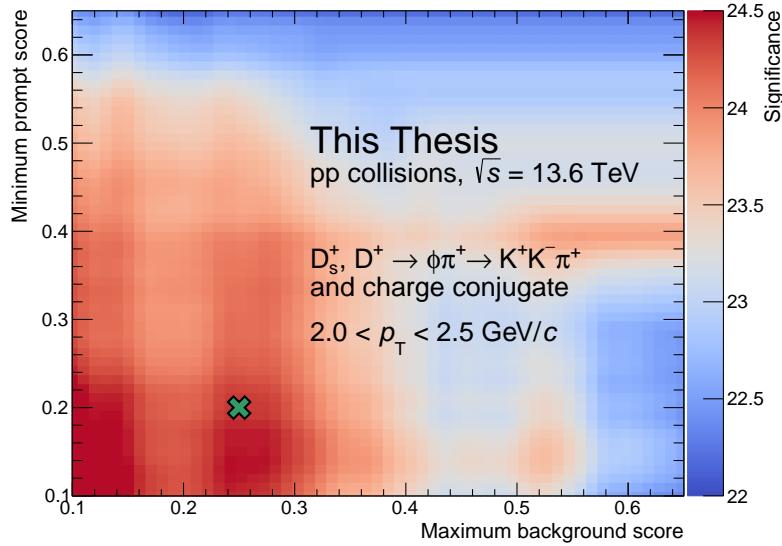


Figure 5.8: Statistical significance of the signal as a function of the selection criteria applied to the model output for the  $2.0 < p_T < 2.5 \text{ GeV}/c$  interval. The chosen set of selection criteria is shown with a green cross.

function.

Figure 5.9 shows the fit to the invariant mass distribution of the selected candidates in the  $1.5 < p_T < 2.0 \text{ GeV}/c$  interval using the approach described above. Due to the concavity-changing shape of the background, the function chosen to describe the background is a third-order polynomial. This change in the concavity of the background invariant-mass distribution was not observed in previous analyses performed on pp collisions data collected during the LHC Run 2 data-taking period. With the upgrade of the ALICE experimental apparatus, the amount of data

Table 5.5: Selection criteria applied to enhance the significance of the  $D_s^+$  meson signal in the  $p_T$  intervals considered for the analysis.

$p_T$ interval ( $\text{GeV}/c$ )	Probability to be background <	Probability to be prompt $D_s^+$ meson >
0.5–1.0	0.01	0.20
1.0–1.5	0.05	0.20
1.5–2.0	0.15	0.20
2.0–2.5	0.25	0.20
2.5–3.0	0.30	0.20
3.0–3.5	0.20	0.20
3.5–4.0	0.20	0.20
4.0–4.5	0.20	0.20
4.5–5.0	0.20	0.20
5.0–5.5	0.30	0.20
5.5–6.0	0.30	0.25
6–8	0.45	0.20
8–12	0.50	0.20
12–24	0.55	0.20

collected during the ongoing LHC Run 3 data-taking period is significantly larger than that collected during Run 2, as described in Chapter 3. With the increased number of candidates available, the statistical precision of the data sample is higher, and the concavity-changing shape of the background becomes more evident. The small number of candidates available in the Run 2 data sample did not allow for the observation of this feature, as statistical uncertainties covered the shape of the background. On top of the previously-unobserved peculiar shape of the background, the fitting function is not able to describe the data accurately between the two peaks, overestimating the data in this invariant mass region.

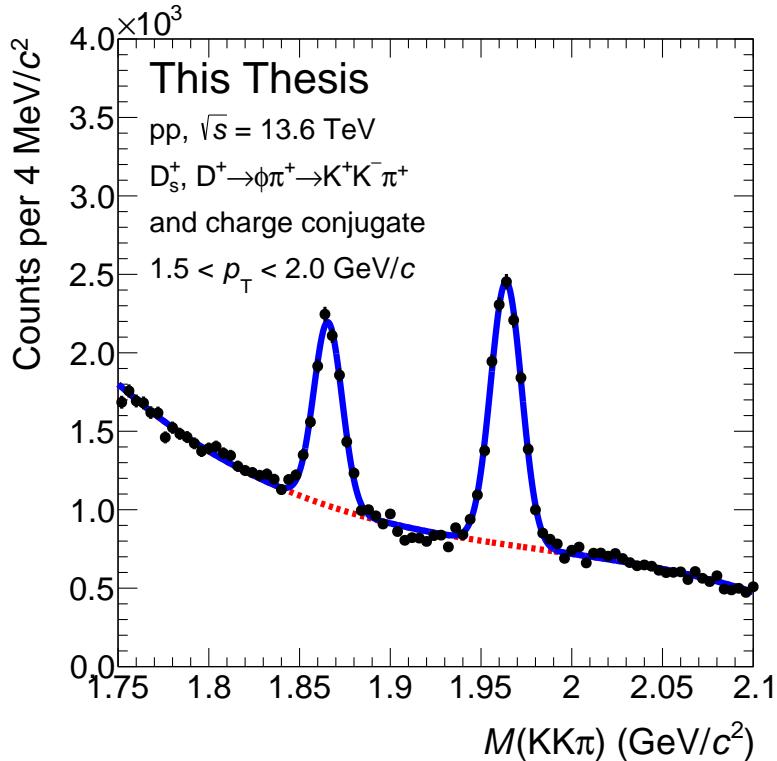


Figure 5.9: Fit to the invariant mass distribution of selected candidates in the  $1.5 < p_T < 2.0$  GeV/c interval. The fit function is shown as a solid line, while the signal and background components are shown as dashed lines. The background is modelled with a third-order polynomial function.

These two features can be understood as due to the fact that the background does not solely arise from the combination of independent tracks, i.e., the combinatorial background. Other physics processes can contribute to the contamination of the data sample, giving rise to a *correlated background*. One such process is the decay of  $D^+$  mesons into the  $D^+ \rightarrow \pi^+K^-\pi^+$  decay channel, where one of the pions is misidentified as a kaon. Despite being suppressed by the applied PID and ML selections, this contribution can give rise to a noticeable contribution due to the large BR of 9.38% [4].

To validate this hypothesis, a simulation of  $D^+$ -meson decays into  $D^+ \rightarrow \pi^+K^-\pi^+$  was run using PYTHIA 8 [98]. Ten billion  $D^+$  mesons were produced with a uniformly distributed  $p_T$  spectrum in the  $0 < p_T < 24$  GeV/c interval.

The  $D^+$  mesons were then forced to decay into the  $D^+ \rightarrow \pi^+ K^- \pi^+$  decay channel. For each decay, two invariant masses were evaluated, by assigning the kaon mass to one of the two pions to simulate the misidentification of one pion, while the correct masses were assigned to the other two prongs. The invariant mass distributions obtained from the simulation are shown in Fig. 5.10 for different  $p_T$  intervals. The distribution is characterised by a peak at  $\sim 2$  GeV/ $c^2$ , well inside the fit range used to extract the raw yield of  $D_s^+$  and  $D^+$  mesons in Fig 5.9. In addition, the invariant mass distribution evolves with the  $p_T$  of the  $D^+$  meson, with a tail towards higher invariant masses for higher  $p_T$  values. This follows naturally from the kinematic properties of the decay. Other sources of correlated background can be present (e.g., from  $\Lambda_c^+$  decays), but their contribution was found to be negligible.

To account for this contribution, a template contribution is included in the fit function, in addition to the combinatorial background and the two signal peaks. This method involves fitting the invariant mass distribution using a predefined template to model the misidentified  $D^+ \rightarrow \pi^+ K^- \pi^+$  decay, whose shape is fixed, and the only adjustable parameter is the normalisation. The shape of the template is taken from the same simulation used to train the BDT model, as described in Sec. 5.2.1, where  $D^+ \rightarrow \pi^+ K^- \pi^+$  mesons reconstructed as  $D^+ \rightarrow K^+ K^- \pi^+$ , passing PID selections as  $K^+ K^- \pi^+$ , are selected. The distribution is fixed to that obtained before applying any ML model, as it was studied that the applied ML selections do not affect the shape of the correlated background. This allows for reducing statistical fluctuations in the template shape. The final fit function is then constructed as the sum of a parabolic function to describe the combinatorial background, the template function described above for the correlated  $D^+ \rightarrow \pi^+ K^- \pi^+$  background source, and two Gaussian functions to describe the  $D_s^+$ - and  $D^+$ -meson peaks. The signal parameters (mean, width, normalisation), as well as the parameters of the combinatorial background and the normalisation of the template fit are left free in the fit in the  $p_T$  intervals below 8 GeV/ $c$ . Because of the observed momentum resolution worsening at higher  $p_T$ , and because of the smaller number of candidates available in the high- $p_T$  intervals, for the  $p_T > 8$  GeV/ $c$  range the widths of the  $D^+$ -meson signal peaks are fixed to those of the  $D_s^+$  meson divided by a factor 1.2, which is the observed ratio of the peak widths at low  $p_T$ , with a flat trend with  $p_T$ , as illustrated in Fig. 5.11. In the left panel, the evolution of the peak widths of the  $D_s^+$  and  $D^+$  mesons as a function of  $p_T$  is shown for both data and MC simulations. A clear increasing trend of the peak widths with  $p_T$  is observed because of the degrading momentum resolution due to the decreasing track curvature. MC simulations underestimate the peak widths extracted from the data for both  $D_s^+$  and  $D^+$  mesons, across the studied  $p_T$  range. In the right panel of Fig. 5.11, the ratio between the peak widths of the  $D_s^+$  and  $D^+$  mesons is shown as a function of  $p_T$ , for both data and MC simulations. The ratio is observed to be almost constant across the analysed  $p_T$  range, with a value of  $\sim 1.2$ . The ratio is underestimated by the MC simulations, which predict a value of  $\sim 1.1$  across the studied  $p_T$  range.

In the  $p_T$  interval above 8 GeV/ $c$ , where the peak width of the  $D^+$  meson is fixed to that of the  $D_s^+$  meson divided by a factor of 1.2, a first fit is performed by keeping both peak widths as free parameters. Then, the  $D^+$ -meson peak width is fixed to that of the  $D_s^+$  divided by 1.2, and the fit is repeated,

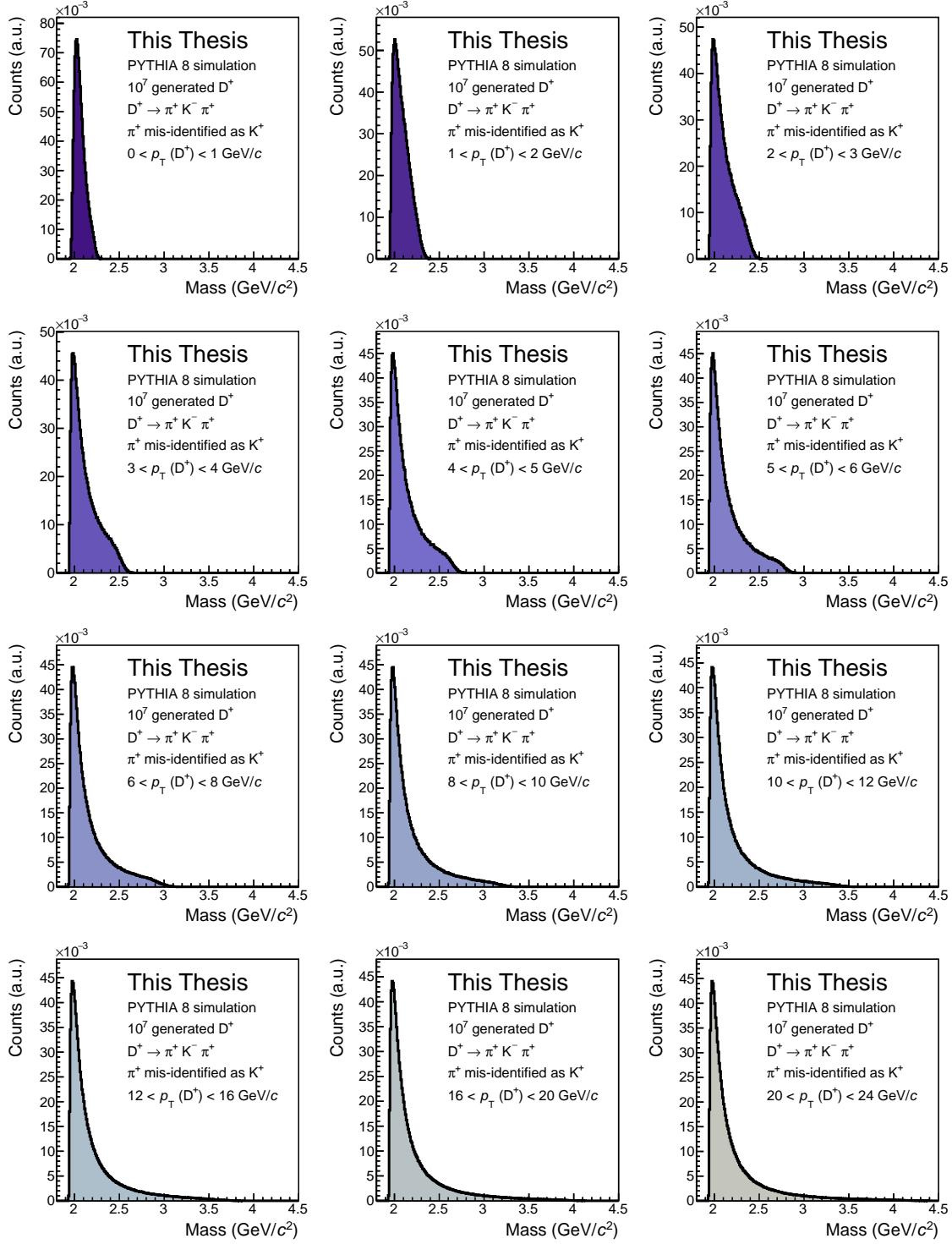


Figure 5.10: Invariant-mass distribution of simulated decays of  $D^+$  mesons into  $D^+ \rightarrow \pi^+ K^- \pi^+$ , where one of the pions produced in the decay is misidentified as a kaon. Distributions for different  $p_T$  intervals are shown.

keeping the remaining parameters free. Additionally, in order to correctly describe the background shape, the fit is performed in the invariant mass range  $1.73 < M(KK\pi) < 2.15 \text{ GeV}/c^2$  for  $p_T > 5 \text{ GeV}/c$ , while a narrower invariant mass window,  $1.75 < M(KK\pi) < 2.1 \text{ GeV}/c^2$ , is used at lower  $p_T$ . The invariant-mass bin width has been fixed to  $2 \text{ MeV}/c^2$ . The fits to the invariant mass distributions of selected D-meson candidates are performed using the `flarefly` package [182], which provides a flexible Python interface for performing fits.

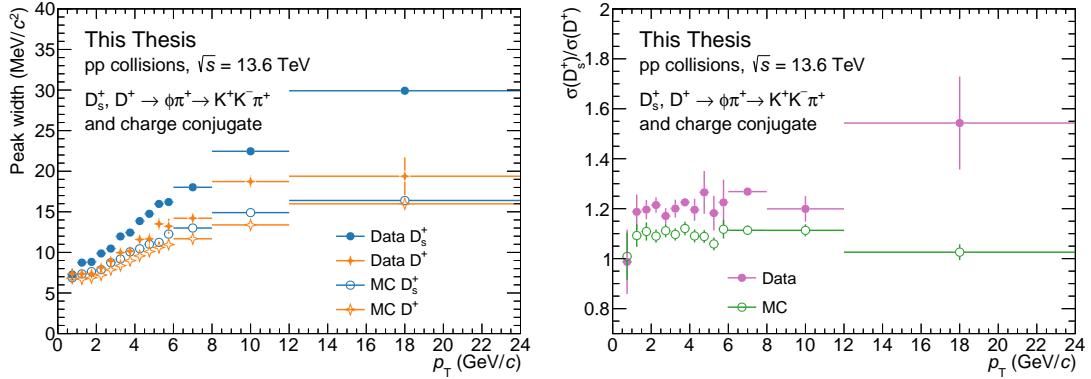


Figure 5.11: Left panel: peak width of the  $D_s^+$  and  $D^+$  mesons as a function of  $p_T$ , before fixing the width of the  $D^+$  meson to that of the  $D_s^+$  meson divided by a factor of 1.2 for  $p_T > 8 \text{ GeV}/c$ . Both the peak widths from fits to data and MC simulations are reported. Right panel: ratio of the peak widths of the  $D_s^+$  and  $D^+$  mesons as a function of  $p_T$ , for both data and MC simulations.

The fit to the invariant mass distribution of candidates passing the ML selections is shown in Fig. 5.12 for one representative  $p_T$  interval of  $2.0 < p_T < 2.5 \text{ GeV}/c$ . The total fit function is shown as a solid blue line, the signal contributions are shown as filled green and azure areas for  $D^+$  and  $D_s^+$  mesons, respectively, the combinatorial background is represented with a solid red line, while the correlated background is shown as a dashed violet line. The fit is able to describe the data accurately, as can be deduced from the distribution of the difference between the data and the background fit function, shown in the bottom panels for the two  $p_T$  bins.

The complete set of fits to the invariant mass distributions in the 14 studied  $p_T$  intervals is shown in Fig. 5.13. All the fits present a good level of agreement between the data and the fit function. Figure 5.13 also shows that the contribution of the correlated  $D^+ \rightarrow \pi^+ K^- \pi^+$  background evolves with  $p_T$ , with a larger contribution at lower  $p_T$  values. The raw yield of  $D_s^+$  and  $D^+$  mesons is then extracted by integrating the signal function.

The evolution of the mean and width of the  $D_s^+$  and  $D^+$  meson peaks as a function of  $p_T$  is shown in Fig. 5.14. In the left panel, the means of the Gaussian functions used to describe the  $D_s^+$ - and  $D^+$ -meson peaks are shown as a function of  $p_T$ . They present an increasing trend with  $p_T$  lying below the values reported in the PDG [4], represented with dashed lines. The shift of the mass values observed in data from those reported in the PDG is common to all analyses of D mesons due to an imperfect description of the detector misalignments in MC simulations. In the right panel, the peak widths of the  $D_s^+$  and  $D^+$  mesons, defined as the standard deviation of the

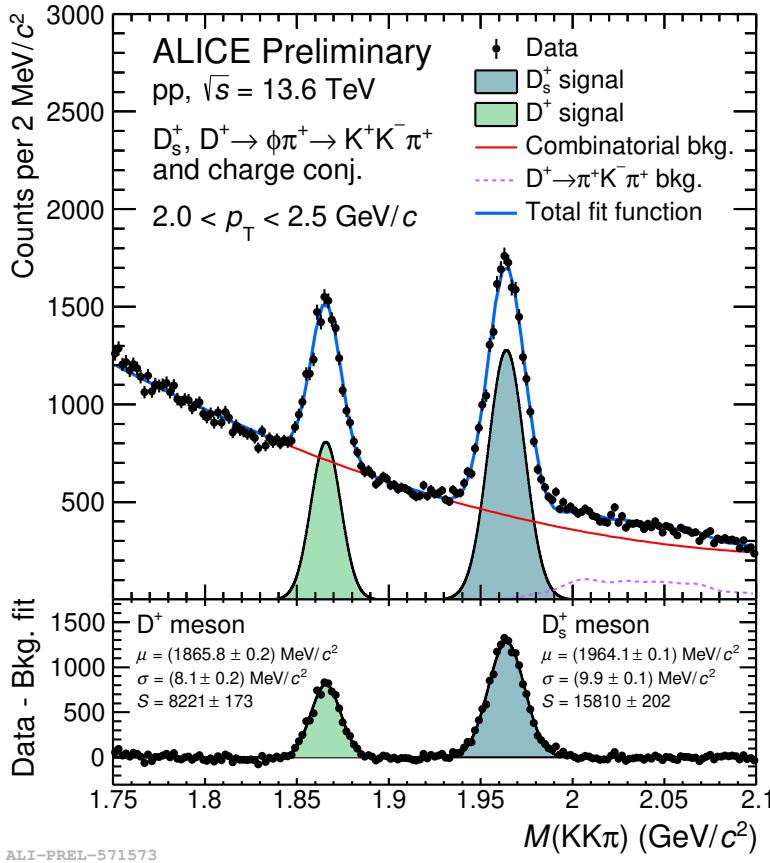


Figure 5.12: Fit to the invariant mass distribution of selected candidate in the  $2.0 < p_T < 2.5$  GeV/c interval. The total fit function is shown as a solid blue line, while background components are shown as solid red lines (correlated background) and dashed violet lines (dashed violet lines). The signal contributions are shown as filled green and azure areas for  $D^+$  and  $D_s^+$  mesons, respectively. The bottom panels show the distribution of the difference between the data and the background fit function.

Gaussian functions used to describe their invariant mass distributions, are shown as a function of  $p_T$ . An increasing trend with  $p_T$  is observed for both mesons, as detailed above for Fig. 5.11, with the peak width of the  $D_s^+$  meson being larger than that of the  $D^+$  meson across the studied  $p_T$  range.

The relative contribution of the  $D^+ \rightarrow \pi^+K^-\pi^+$  correlated background to the total fitting function is shown in the left panel of Fig. 5.15 as a function of  $p_T$ . At low  $p_T$ , where the PID is more effective, the contribution of the correlated background is observed to be smaller. It then increases up to 2 GeV/c and then decreases due to the kinematic properties of the decay as expected from Fig. 5.10. Some residual contribution is observed at higher  $p_T$ , related to fluctuations in the fitting procedure. In the right panel of Fig. 5.15, the statistical significance of the extracted  $D_s^+$ - and  $D^+$ -meson signals is shown as a function of  $p_T$ . Due to the large combinatorial background at low  $p_T$ , the statistical significance presents a decreasing trend with decreasing  $p_T$  for  $p_T \lesssim 2$  GeV/c, for both D mesons. The maximum statistical

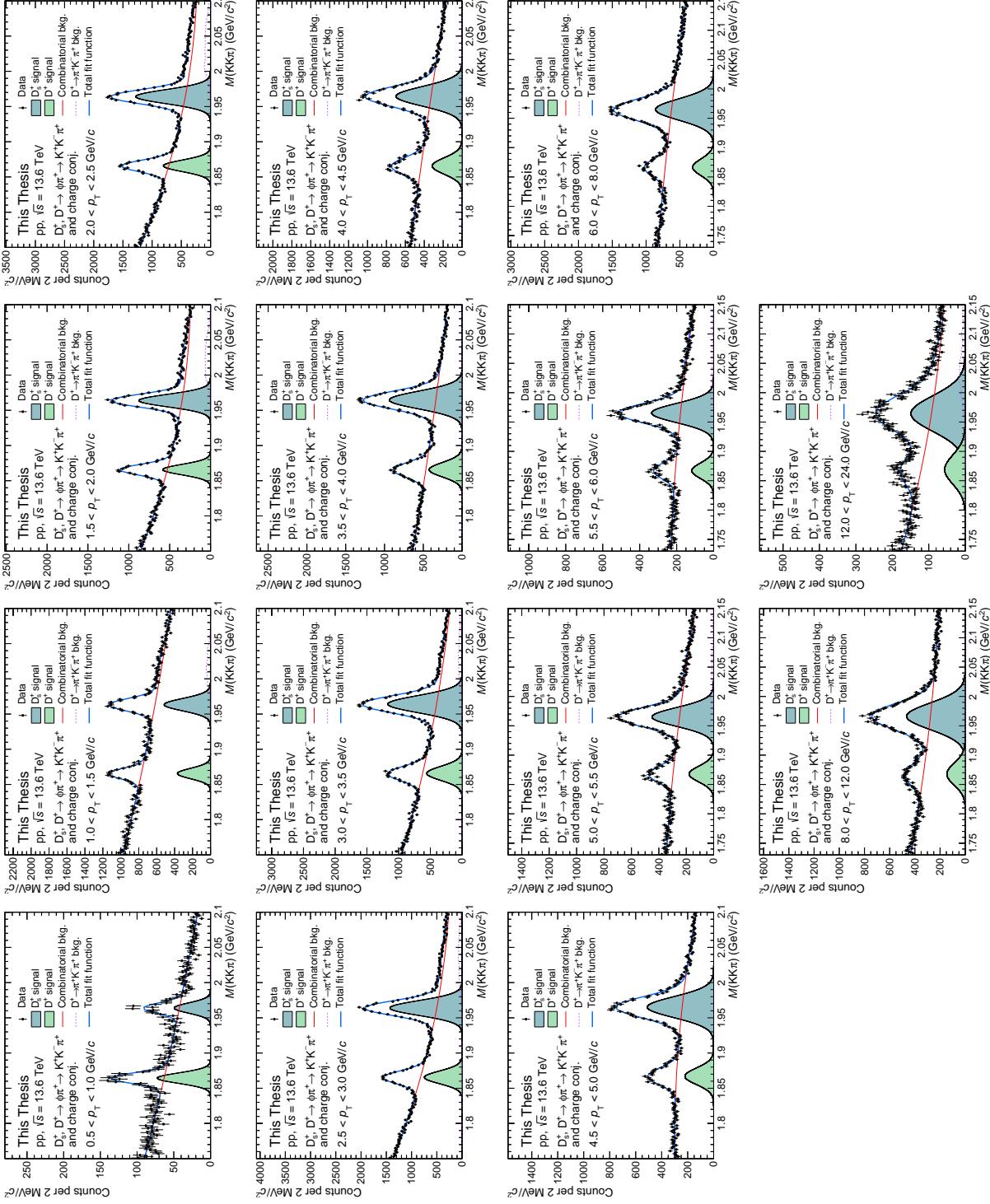


Figure 5.13: Fit to the invariant mass distribution of selected candidate in the 14 studied  $p_T$  intervals in the  $0.5 < p_T < 24$  GeV/ $c$  range. The total fit function is shown as a solid blue line, while background components are shown as solid red lines (correlated background) and dashed violet lines (dashed background). The signal contributions are shown as filled green and azure areas (dashed green and dashed purple lines) for  $D^+$  and  $D_s^+$  mesons, respectively.

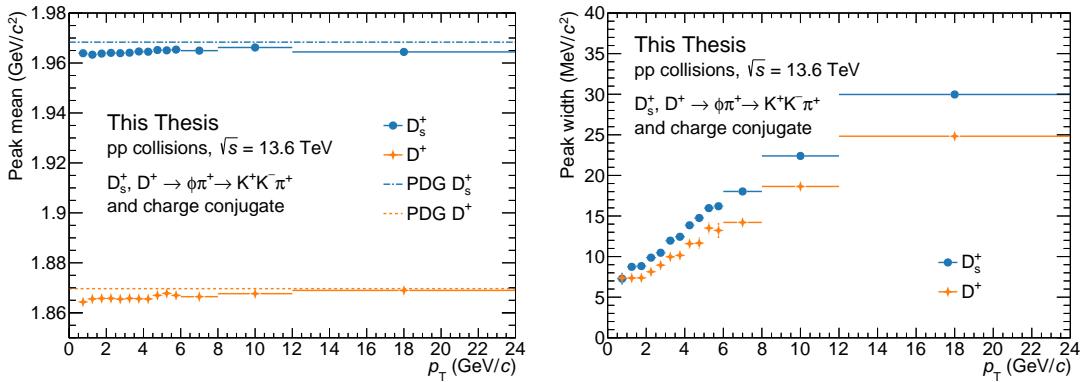


Figure 5.14: Evolution of the mean (left panel) and width (right panel) of the Gaussian functions used to describe the  $D_s^+$  and  $D^+$  meson peaks as a function of  $p_T$ . The dashed lines in the left panel represent the mass values reported in the PDG [4] for the two D-meson species.

significance is observed in the  $2.5 < p_T < 3.0$  GeV/ $c$  interval for the  $D_s^+$  meson and in the  $2.0 < p_T < 2.5$  GeV/ $c$  range for the  $D^+$  meson, where significances of 99 and 50 are reached for the two meson species, respectively. At higher  $p_T$ , the statistical significance decreases due to the smaller number of produced D mesons, and the larger peak widths of the  $D_s^+$  and  $D^+$  mesons, which result in a larger background contribution in the signal region. A sudden increase in the statistical significance is observed in the  $6 < p_T < 8$  GeV/ $c$  interval, and is due to an increase of the width of the  $p_T$  interval for the signal extraction, from 0.5 GeV/ $c$  to 2 GeV/ $c$ , which results in a larger number of candidates available in the signal region.

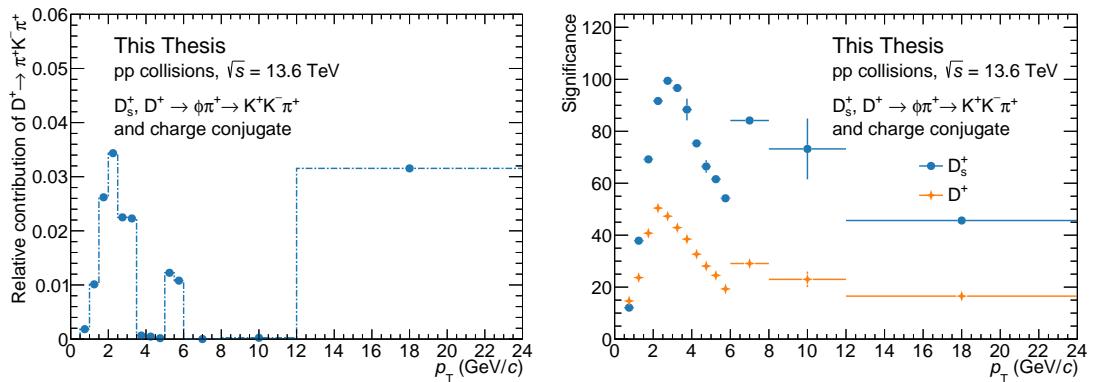


Figure 5.15: Evolution of the  $D^+ \rightarrow \pi^+K^-\pi^+$  relative contribution to the fitting function (left panel) and statistical significance of the extracted  $D_s^+$ - and  $D^+$ -meson signal (right panel) as a function of  $p_T$ .

The extraction of the raw yields is affected by several arbitrary choices, for example, the functional description of the background, the choice of the fit range, the choice of fixing the  $D^+$  width at high  $p_T$ , and the invariant-mass bin width. Changes in these choices can lead to variations in the extracted raw yields. To estimate the effect of such arbitrary choices in the final observable (the  $D_s^+/D^+$

production-yield ratio), and estimate a systematic uncertainty associated with the raw yield extraction procedure, the fit is repeated several times by varying the fit range, the bin width, and the functional form of the background. In the low  $p_T$  region ( $p_T < 4 \text{ GeV}/c$ ), the lower fit limit is varied between 1.71 and 1.77  $\text{GeV}/c^2$ , while the upper limit is varied between 2.08 and 2.14  $\text{GeV}/c^2$ . The functions considered to describe the background are a second-order polynomial and an exponential function. At higher  $p_T$  (up to 8  $\text{GeV}/c$ ), where the signal peaks become broader, the lower limit is varied between 1.71 and 1.75  $\text{GeV}/c^2$ , while the maximum mass is varied between 2.13 and 2.19  $\text{GeV}/c^2$ . The functions considered to describe the background are a second- and a third-order polynomial, as the exponential function was observed not to reproduce the background distribution. The bin width is varied between 1 and 4  $\text{MeV}/c^2$  across the studied  $p_T$  interval. At  $p_T > 8 \text{ GeV}/c^2$ , where the  $D^+$ -meson peak width is fixed to that of the  $D_s^+$ -meson divided by 1.2, the peak width of the  $D^+$  meson is changed by varying the dividing factor by  $\pm 10\%$ . For each possible combination of these variations, the signal is extracted for the two D-meson species, and the ratio between them is calculated. This results in 128, 96, and 288 trials for the  $0.5 < p_T < 4 \text{ GeV}/c$ ,  $4 < p_T < 8 \text{ GeV}/c$ , and  $8 < p_T < 24 \text{ GeV}/c$  intervals, respectively. As a quality check, trials with a  $\chi^2/\text{ndf}$  greater than 2 are discarded. The distribution of the  $D_s^+/D^+$  raw-yield ratio is then obtained for the different trials, and the mean and standard deviation are calculated. The difference  $\Delta$  between the mean of this distribution and the ratio extracted with the default configuration is calculated. The systematic uncertainty is then evaluated as the sum in quadrature of the standard deviation and  $\Delta$ :  $\sqrt{\text{RMS}^2 + \Delta^2}$ .

The result of this multi-trial approach for the evaluation of the systematic uncertainty on the raw-yields extraction is shown in Fig. 5.16 for the  $1.0 < p_T < 1.5 \text{ GeV}/c$  interval. The complete set of figures on the results of the evaluation of the systematic uncertainty on the raw-yields extraction is provided in Appendix A. In the top-left panel, the raw yields extracted from the fit to the invariant mass distribution are reported for the  $D_s^+$  and  $D^+$  mesons for the different trials. As a cross-check, the raw yields are also extracted by summing the counts of a distribution obtained by subtracting the background fit function (composed of the function considered in the trial for the combinatorial background and of a template function for the  $D^+ \rightarrow \pi^+ K^- \pi^+$  correlated background) from the invariant mass distribution of the candidates passing the ML selections. The bin contents are summed within 3 standard deviations from the peak mean. The extracted raw yields are stable within uncertainty across all the trials for the  $D^+$  meson. For the  $D_s^+$  meson, the  $3\sigma$  bin counting method presents higher raw yields when compared to the ones obtained by integrating the signal function. However, this is considered as related to fluctuations in the invariant mass distribution rather than a systematic shift of the raw yields due to a different definition of the observable. The raw yields extracted with the default configuration described above are shown as dashed lines. In the top-right panel, the  $\chi^2/\text{ndf}$  of the fit to the invariant mass distribution is shown for the different trials. Only trials where the signal is extracted by integration of the fitted signal function are shown in this panel. In this  $p_T$  interval, all the different trials present a  $\chi^2/\text{ndf} < 2$ , and no trial is rejected because of poor fit quality. However, at higher  $p_T$  a few trials do not survive the quality check on the  $\chi^2/\text{ndf}$ , and are

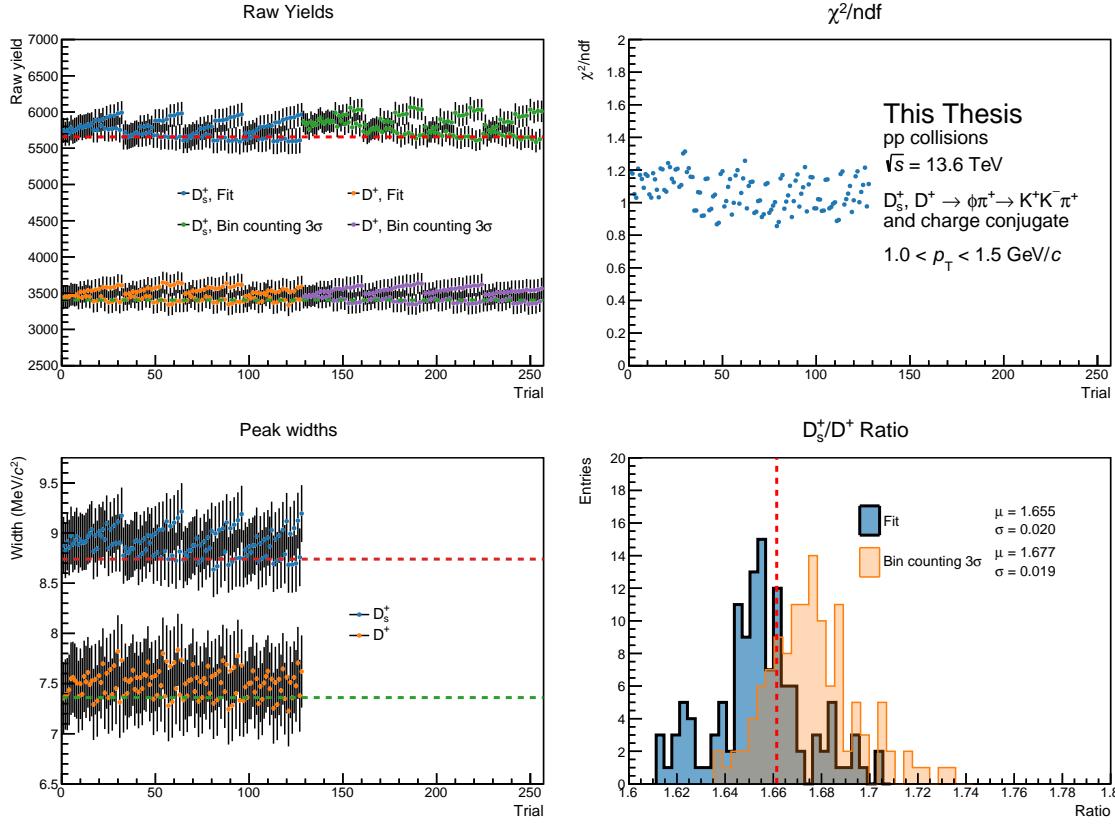


Figure 5.16: Results from the multi-trial approach employed for estimating the systematic uncertainty related to the raw yield extraction in the  $1.0 < p_T < 1.5 \text{ GeV}/c$  interval.

discarded. These results illustrate that the fit function is able to describe the data accurately, independently of the configuration of the fit parameters. In the bottom-left panel, the peak widths for  $D_s^+$  and  $D^+$  mesons are reported, and present a stable behaviour across the different trials. The peak widths for the default configuration are also reported as dashed lines. The bottom-right panel shows the distribution of the  $D_s^+/D^+$  raw-yield ratio for the different trials, obtained using the two methods previously introduced (namely, integration of the signal fit function and bin counting). The  $3\sigma$  bin counting method presents a higher raw-yield ratio in this  $p_T$  interval when compared to the other method, because of the higher  $D_s^+$ -meson yields. The systematic uncertainty is defined as the sum in quadrature of the standard deviation of the distribution of  $D_s^+/D^+$  raw-yield ratio obtained by integrating the signal function and the difference  $\Delta$  between the default value reported with a dashed red line and the mean of this distribution. This quantity ranges from 1% to 10% of the central  $D_s^+/D^+$  raw-yield ratio, depending on the considered  $p_T$  interval. The systematic uncertainty is then assigned after smoothing the  $p_T$  dependence of the obtained values. The assigned systematic uncertainty is reported in Table 5.6.

Table 5.6: Systematic uncertainty on the raw-yield extraction for the ratio of the  $D_s^+$  and  $D^+$  mesons yields.

$p_T$ (GeV/ $c$ )	$\sqrt{\text{RMS}^2 + \Delta^2}/(\text{central } D_s^+/\text{D}^+) (\%)$	Assigned systematic uncertainty (%)
0.5–1	4	3
1–1.5	1	3
1.5–2	2	3
2–2.5	3	3
2.5–3	3	3
3–3.5	3	3
3.5–4	3	3
4–4.5	6	5
4.5–5	7	5
5–5.5	3	5
5.5–6	5	5
6–8	8	8
8–12	9	9
12–24	10	10

# Chapter 6

## Evaluation of the corrected $D_s^+ / D^+$ yield ratio

To evaluate the ratio between the production yield ratio of the two D-meson species, a number of corrections must be applied to the extracted raw yields. These corrections are necessary to account for the acceptance of the detector, the selection efficiency of  $D_s^+$  and  $D^+$  mesons, the branching ratio of the decay channels, and the feed-down from beauty-hadron decays. The prompt  $D_s^+ / D^+$  production yield ratio is then obtained by dividing the corrected yield of  $D_s^+$  mesons by the corrected yield of  $D^+$  mesons:

$$D_s^+ / D^+ = \frac{N_{\text{raw}}^{D_s^+} \cdot f_{\text{prompt}}^{D_s^+}}{(\text{Acc} \times \varepsilon)_{\text{prompt}}^{D_s^+} \cdot \text{BR}^{D_s^+}} \cdot \left( \frac{N_{\text{raw}}^{D^+} \cdot f_{\text{prompt}}^{D^+}}{(\text{Acc} \times \varepsilon)_{\text{prompt}}^{D^+} \cdot \text{BR}^{D^+}} \right)^{-1}, \quad (6.1)$$

where  $(\text{Acc} \times \varepsilon)_{\text{prompt}}$  is the product of the detector acceptance and the efficiency of the selections for promptly-produced D mesons,  $f_{\text{prompt}}$  is the fraction of prompt D mesons in the extracted raw yield, and BR is the branching ratio of the considered decay channel.

In the following sections, a detailed description of the corrections applied to the raw yields is given.

### 6.1 Acceptance and efficiency correction

The first correction takes into account that only a fraction Acc (the detector's *acceptance*) of the produced D mesons can be detected by the ALICE apparatus, due to the geometry of the detector. As the tracks are reconstructed in the pseudorapidity region  $|\eta| < 0.8$ , the acceptance is limited to  $|y| < 0.8$  at high  $p_T$  ( $\gtrsim 5 \text{ GeV}/c$ ) and to  $|y| < 0.5$  for  $p_T \sim 0$ , as shown in Fig. 4.4. Due to the very similar decay topology and kinematics of the  $D_s^+$  and  $D^+$  mesons in the considered decay channel, this term is expected to cancel out in the ratio of the two species' production yields. Additionally, only a fraction  $\varepsilon$  (the reconstruction and selection *efficiency*) of the detectable D mesons passes the selection criteria described in Chapters 4 and 5. The efficiency term  $\varepsilon$  also includes the reconstruction efficiency of the D meson daughter particles.

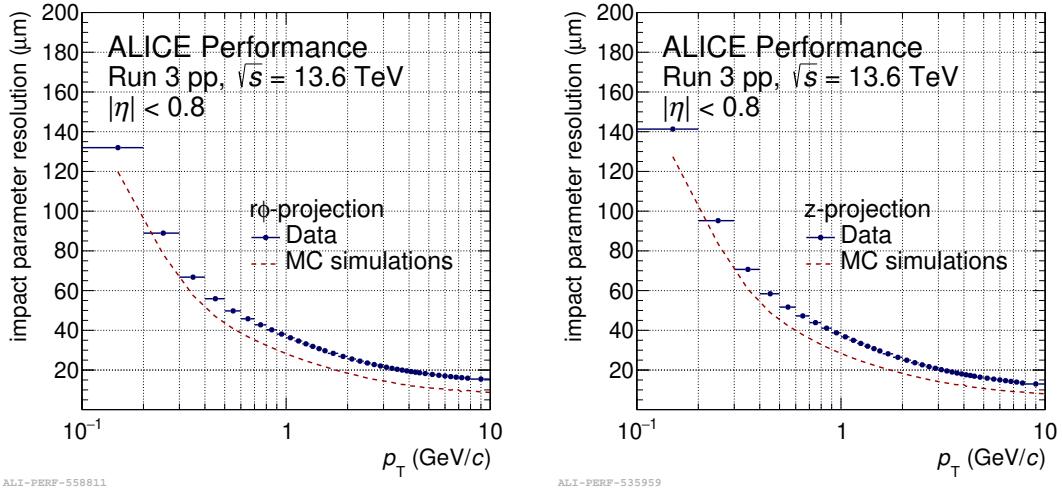


Figure 6.1: Impact-parameter resolution in the transverse plane (left panel) and beam direction (right panel) of primary particles in data and MC. Figure taken from the ALICE figure repository [60].

These corrections are evaluated using a sample of  $D_s^+$  and  $D^+$  mesons from Monte Carlo (MC) simulations. To avoid the introduction of biases, a different data sample is used for the evaluation of the acceptance and efficiency corrections compared to the one used for the ML model training and performance evaluation. Proton-proton collisions are simulated using the PYTHIA 8 event generator [98] with colour-reconnection Mode 2 [104]. As for the MC simulations employed for the ML training, the MC dataset is enriched of heavy-flavour hadrons by only selecting events where at least a  $c\bar{c}$  or  $b\bar{b}$  pair is produced. The generated particles are propagated through the ALICE experimental apparatus using the GEANT4 transport simulation toolkit [164]. Due to the displaced topology of heavy-flavour decays and the continuous readout employed by the ALICE detector, events with charm or beauty hadrons are frequently affected by “fake” primary vertices arising from the association of displaced decay tracks, thereby degrading the reconstruction of heavy-flavour hadrons. The inefficiency in the reconstruction due to the presence of many heavy-flavour hadrons in nearby events is worse in MC simulations than in data, as heavy-flavour hadrons are produced more frequently than in data and the time resolution on the tracks does not always allow for their non-ambiguous association to a single collision. To overcome this problem, minimum bias events are generated between charm- or beauty-triggered ones (*gap-triggered* approach). Studies performed using different gap sizes have shown that a gap of 4 minimum bias events between heavy-flavour-triggered events reduces the number of fake vertices to an acceptable level, while keeping the simulation time reasonable.

Special care is taken to ensure that the MC simulation reproduces the experimental conditions and the reconstruction configuration used for data. A slightly worse impact-parameter resolution is obtained in data, probably due to a residual misalignment of the ITS, as shown in Fig. 6.1, where the impact parameter resolutions in the

transverse plane and along the beam direction are shown for both data and MC simulations. To improve the MC description of the spatial resolution measured in data, a smearing of the track impact parameters (i.e., the distance of closest approach to the primary vertex) is applied via a dedicated workflow (**track-tuner**). The workflow has been tuned by comparing the impact-parameter resolution in data and MC obtained by fitting their distributions for primary particles. The **track-tuner** workflow allows the MC simulations to reproduce the impact parameter resolution observed in data.

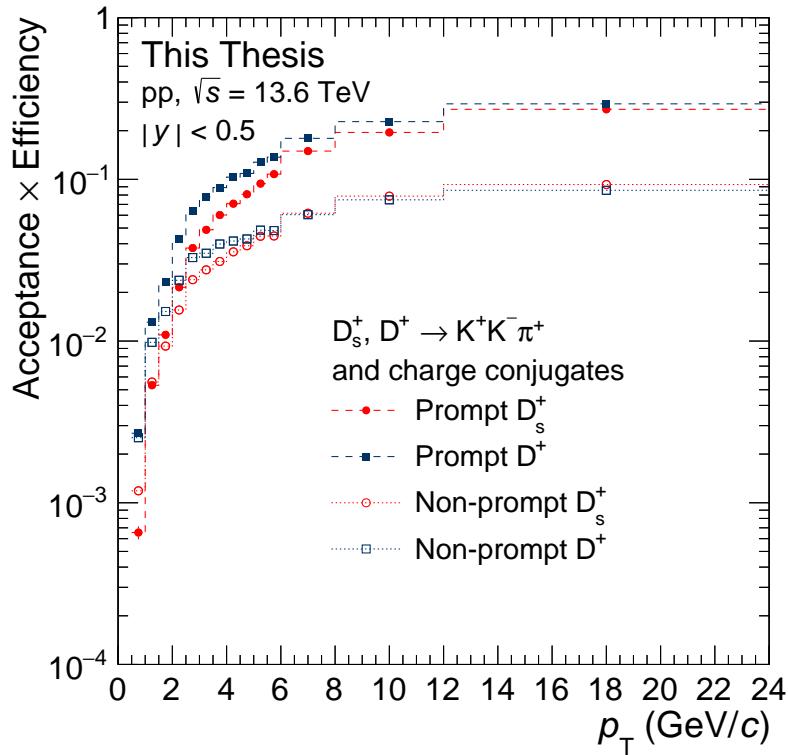


Figure 6.2: Acceptance-times-efficiency correction factor as a function of  $p_T$  for prompt and non-prompt  $D_s^+$  and  $D^+$  mesons.

The  $(\text{Acc} \times \varepsilon)$  factor is defined as

$$(\text{Acc} \times \varepsilon) = \frac{N_{\text{sel.}}^{|y|<0.8}}{N_{\text{gen.}}^{|y|<0.5}} ,$$

where  $N_{\text{gen.}}^{|y|<0.5}$  is the number of generated D mesons with  $|y| < 0.5$ , and  $N_{\text{sel.}}^{|y|<0.8}$  is the number of D mesons with  $|y| < 0.8$  that pass the applied selection criteria. The  $(\text{Acc} \times \varepsilon)$  factor is illustrated in Fig. 6.2 for both  $D_s^+$  and  $D^+$  mesons as a function of  $p_T$ . Due to the different decay topologies of  $D_s^+$  and  $D^+$  mesons, with the second decaying on average at a larger distance from the primary vertex, the efficiency is lower for  $D_s^+$  mesons than for  $D^+$  mesons, while the acceptance for the two meson species is the same. For the same reason, the  $(\text{Acc} \times \varepsilon)$  factor is larger for promptly-produced D mesons than for non-prompt ones. The correction presents a

strong dependence on  $p_T$ . The larger Lorentz boost of D mesons at high  $p_T$  results in a more displaced decay vertex, facilitating the selection of  $D_s^+$  and  $D^+$  mesons. To extract a statistically significant signal at low  $p_T$ , where a larger number of background tracks (and therefore combinatorial background candidates) is present, the selection criteria are tightened, reducing the efficiency of the selection. The  $(\text{Acc} \times \varepsilon)$  correction ranges from about 0.3 in the highest  $p_T$  interval down to about  $7 \times 10^{-4}$  in the lowest considered  $p_T$  interval.

## 6.2 Prompt fraction correction

The fraction of prompt  $D_s^+$  and  $D^+$  mesons in the measured raw yield ( $f_{\text{prompt}}^{D_s^+}$  and  $f_{\text{prompt}}^{D^+}$ ) is estimated using a data-driven method that involves varying the BDT probability threshold for candidate selection. This variation changes the balance between prompt and non-prompt contributions to the signal, exploiting their different trends with the BDT output score to disentangle their relative abundances in the extracted raw yield. Introduced in Ref. [62], this approach leverages the relationship between the raw yield and the number of produced prompt ( $N_{\text{prompt}}$ ) and non-prompt ( $N_{\text{non-prompt}}$ ) D mesons decaying into the considered decay channel:

$$N_{\text{raw}} = N_{\text{prompt}} \times (\text{Acc} \times \varepsilon)_{\text{prompt}} + N_{\text{non-prompt}} \times (\text{Acc} \times \varepsilon)_{\text{non-prompt}} . \quad (6.2)$$

Eq. 6.2 holds for every chosen BDT selection criterion, defining the  $(\text{Acc} \times \varepsilon)_{\text{prompt}}$  and  $(\text{Acc} \times \varepsilon)_{\text{non-prompt}}$  correction factors and the raw yield  $N_{\text{raw}}$ . Since the  $(\text{Acc} \times \varepsilon)$  factors can be determined from MC simulations (as described in the previous section), and  $N_{\text{raw}}$  is extracted through a fit to the invariant mass distribution of selected candidates from data (as detailed in Chapter 5), Eq. 6.2 contains only two unknown variables,  $N_{\text{prompt}}$  and  $N_{\text{non-prompt}}$ . They can be evaluated by varying the BDT working point, extracting the corresponding raw yield, and estimating the  $(\text{Acc} \times \varepsilon)_{\text{prompt}}$  and  $(\text{Acc} \times \varepsilon)_{\text{non-prompt}}$  factors. However, a single variation of the selection criteria may lead to large uncertainties, as a small change in the  $(\text{Acc} \times \varepsilon)$  correction adds limited information on  $N_{\text{prompt}}$  and  $N_{\text{non-prompt}}$ . To overcome this issue, the BDT threshold values are varied several times to span a wide range of  $(\text{Acc} \times \varepsilon)_{\text{prompt}}$  (and  $(\text{Acc} \times \varepsilon)_{\text{non-prompt}}$ ) and significantly change the  $f_{\text{prompt}}$  (and  $f_{\text{non-prompt}}$ ) in the extracted raw yield. For each selection criterion  $i$ , the acceptance-times-efficiency factor is calculated for prompt  $(\text{Acc} \times \varepsilon)_{\text{prompt}}^i$  and non-prompt  $(\text{Acc} \times \varepsilon)_{\text{non-prompt}}^i$  D mesons, and the raw yields  $N_{\text{raw}}^i$  are extracted by fitting the invariant mass distribution of candidates passing the selections. By including the entire set of selection criteria, a system of equations can be defined:

$$\begin{cases} N_{\text{raw}}^1 = N_{\text{prompt}} \times (\text{Acc} \times \varepsilon)_{\text{prompt}}^1 + N_{\text{non-prompt}} \times (\text{Acc} \times \varepsilon)_{\text{non-prompt}}^1 \\ N_{\text{raw}}^2 = N_{\text{prompt}} \times (\text{Acc} \times \varepsilon)_{\text{prompt}}^2 + N_{\text{non-prompt}} \times (\text{Acc} \times \varepsilon)_{\text{non-prompt}}^2 \\ \vdots \\ N_{\text{raw}}^n = N_{\text{prompt}} \times (\text{Acc} \times \varepsilon)_{\text{prompt}}^n + N_{\text{non-prompt}} \times (\text{Acc} \times \varepsilon)_{\text{non-prompt}}^n \end{cases} .$$

For  $n > 2$ , the system of equations is overconstrained, and the solution can be found through a minimisation procedure. The system of equations can be written in matrix

notation as

$$\begin{pmatrix} (\text{Acc} \times \varepsilon)_{\text{prompt}}^1 & (\text{Acc} \times \varepsilon)_{\text{non-prompt}}^1 \\ (\text{Acc} \times \varepsilon)_{\text{prompt}}^2 & (\text{Acc} \times \varepsilon)_{\text{non-prompt}}^2 \\ \vdots & \vdots \\ (\text{Acc} \times \varepsilon)_{\text{prompt}}^n & (\text{Acc} \times \varepsilon)_{\text{non-prompt}}^n \end{pmatrix} \begin{pmatrix} N_{\text{prompt}} \\ N_{\text{non-prompt}} \end{pmatrix} - \begin{pmatrix} N_{\text{raw}}^1 \\ N_{\text{raw}}^2 \\ \vdots \\ N_{\text{raw}}^n \end{pmatrix} = \begin{pmatrix} \delta^1 \\ \delta^2 \\ \vdots \\ \delta^n \end{pmatrix},$$

where  $\delta^i$  are the residuals due to the uncertainty in the raw yield and  $(\text{Acc} \times \varepsilon)$  correction factors for the  $i$ -th selection. For each selection criterion, the uncertainty on  $\delta^i$  is estimated by propagating the uncertainties on the raw yield and  $(\text{Acc} \times \varepsilon)$  factors as

$$\sigma_i^2 = \sigma_{N_{\text{raw}}^i}^2 + N_{\text{prompt}} \times \sigma_{(\text{Acc} \times \varepsilon)_{\text{prompt}}^i}^2 + N_{\text{non-prompt}} \times \sigma_{(\text{Acc} \times \varepsilon)_{\text{non-prompt}}^i}^2,$$

where  $\sigma_{N_{\text{raw}}^i}$  is the uncertainty on the extracted raw yield, and  $\sigma_{(\text{Acc} \times \varepsilon)_{\text{prompt}}^i}$  and  $\sigma_{(\text{Acc} \times \varepsilon)_{\text{non-prompt}}^i}$  are the uncertainties on the prompt and non-prompt  $(\text{Acc} \times \varepsilon)$  factors, respectively. Given that the corrected prompt and non-prompt yields are unknown variables, an iterative procedure is used to define the total uncertainty: initially,  $N_{\text{prompt}}$  and  $N_{\text{non-prompt}}$  are set to zero and only the uncertainty on the raw yields is taken into account (as  $\sigma_{(\text{Acc} \times \varepsilon)_{\text{prompt}}^i}$  and  $\sigma_{(\text{Acc} \times \varepsilon)_{\text{non-prompt}}^i}$  are not relevant in the evaluation of  $\sigma_i$  if  $N_{\text{prompt}} = N_{\text{non-prompt}} = 0$ ). From the second iteration, the corrected yields  $N_{\text{prompt}}$  and  $N_{\text{non-prompt}}$  obtained in the previous step are also used, along with the uncertainties on the  $(\text{Acc} \times \varepsilon)$  factors. This iteration is repeated until the difference between the corrected yields in successive iterations falls below a predefined threshold.

The solution of the system of equations is found by minimising the residuals with a least-squares method, with a  $\chi^2$  defined as

$$\chi^2 = \boldsymbol{\delta}^T \mathbf{C}^{-1} \boldsymbol{\delta},$$

where  $\boldsymbol{\delta}$  is the vector of residuals and  $\mathbf{C}$  is the covariance matrix of the residuals.

In this analysis, the BDT threshold is fixed for the background score, at a tighter value than the one used to extract the central raw yields to ensure the convergence of the invariant mass fits. The minimum probability for being a non-prompt D meson is varied in different ranges for the two D meson species, to guarantee that a large enough variation in the  $f_{\text{non-prompt}}$  is achieved. Indeed, due to the larger proper mean decay length of  $D^+$  mesons, the usage of the selection criteria employed for the  $D_s^+$  meson would lead to a smaller variation in the  $f_{\text{non-prompt}}$  factor for the  $D^+$  meson. Therefore, the BDT threshold for the non-prompt score is varied in a wider range for  $D^+$  mesons than for  $D_s^+$  mesons, reaching larger threshold values on the minimum non-prompt BDT score for  $D^+$  mesons.

Because the selection criterion is increasingly tightened, the  $i+1$ -th selected sample will be entirely contained in the  $i$ -th one. The residuals  $\delta^i$  will therefore exhibit a degree of correlation. The off-diagonal elements  $\sigma_{i,j}$  of the covariance matrix  $\mathbf{C}$ , which are the covariance between the residuals of the  $i$ -th and  $j$ -th selection criteria, can be estimated as

$$\sigma_{i,j} = \rho_{i,j} \sigma_i \sigma_j,$$

where it can be demonstrated [183] that the correlation coefficient  $\rho_{i,j}$  is given by

$$\rho_{i,j} = \frac{\sigma_i}{\sigma_j}$$

if the measurement  $i$  is made on a dataset that is fully included in the one used for the measurement  $j$ .

The minimisation procedure described above leads to the determination of the true number of prompt and non-prompt D mesons decaying into the considered decay channel ( $N_{\text{prompt}}$  and  $N_{\text{non-prompt}}$ , respectively), which are independent of the applied selection criteria. The results for the evaluation of the  $D_s^+ f_{\text{prompt}}$  correction factor in the  $1.5 < p_T < 2.0 \text{ GeV}/c$  interval are shown in Fig. 6.3. The complete set of figures on the results of the evaluation of the  $f_{\text{prompt}}^{D_s^+}$  and  $f_{\text{prompt}}^{D^+}$  is provided in Appendices B and C, respectively. In the top-left panel, the correlation factor  $\rho$  between the residuals  $\delta^i$  is shown for the different selection criteria; in the top-right panel, the  $(\text{Acc} \times \varepsilon)$  factors are shown for both prompt and non-prompt  $D_s^+$  mesons as a function of the index used to refer to the BDT threshold value applied to the non-prompt BDT output score. The leftmost data point of each distribution corresponds to the loosest selection on the BDT non-prompt score, while the rightmost one corresponds to the strictest selection, which is expected to preferentially select non-prompt D mesons. In the bottom-left panel, the relative contribution of the prompt and non-prompt components to the extracted  $D_s^+$  raw yields is shown as a function of the ML selections. Lastly, in the bottom-right panel, the extracted yields are fitted using a template fit, utilising the  $(\text{Acc} \times \varepsilon)_{\text{prompt}}^{D_s^+}$  and  $(\text{Acc} \times \varepsilon)_{\text{non-prompt}}^{D_s^+}$  evolution as a function of the BDT selection criterion. The minimisation procedure described above can in fact be interpreted as a template fit to the extracted raw yields, where the templates are the  $(\text{Acc} \times \varepsilon)$  factors as a function of the BDT selection. The prompt and non-prompt components of the raw yield, obtained for each BDT-based selection from the minimisation procedure as  $(\text{Acc} \times \varepsilon)_{\text{prompt}} \times N_{\text{prompt}}$  and  $(\text{Acc} \times \varepsilon)_{\text{non-prompt}} \times N_{\text{non-prompt}}$  respectively, are represented with red and blue filled histograms. Their sum is reported by the green histogram. Given the small  $\chi^2/\text{ndf}$  value obtained from the fit, the  $f_{\text{prompt}}$  factor is considered to be well determined.

The  $f_{\text{prompt}}$  factor is then calculated for each D-meson species and for any given selection criterion with efficiencies  $(\text{Acc} \times \varepsilon)_{\text{prompt}}$  and  $(\text{Acc} \times \varepsilon)_{\text{non-prompt}}$  for prompt and non-prompt D mesons respectively, as

$$f_{\text{prompt}} = \frac{N_{\text{prompt}} \times (\text{Acc} \times \varepsilon)_{\text{prompt}}}{N_{\text{prompt}} \times (\text{Acc} \times \varepsilon)_{\text{prompt}} + N_{\text{non-prompt}} \times (\text{Acc} \times \varepsilon)_{\text{non-prompt}}} .$$

As a cross-check, the  $f_{\text{prompt}}$  factor is also estimated using a theory-driven approach [148], which relies on FONLL [63] predictions for beauty-hadron production and the decay kinematic description of PYTHIA 8 [98] for estimating the  $p_T$  differential cross section of non-prompt D mesons. The  $f_{\text{prompt}}$  factor is then calculated for each D-meson species as

$$f_{\text{prompt}} = 1 - \frac{d^2\sigma}{dp_T dy} \Big|_{\text{FONLL+PYTHIA 8}}^{\text{non-prompt}} \times \frac{(\text{Acc} \times \varepsilon)_{\text{non-prompt}} \Delta y \Delta p_T \text{BR} \mathcal{L}_{\text{int}}}{\frac{1}{2} \times N_{\text{raw}}} ,$$

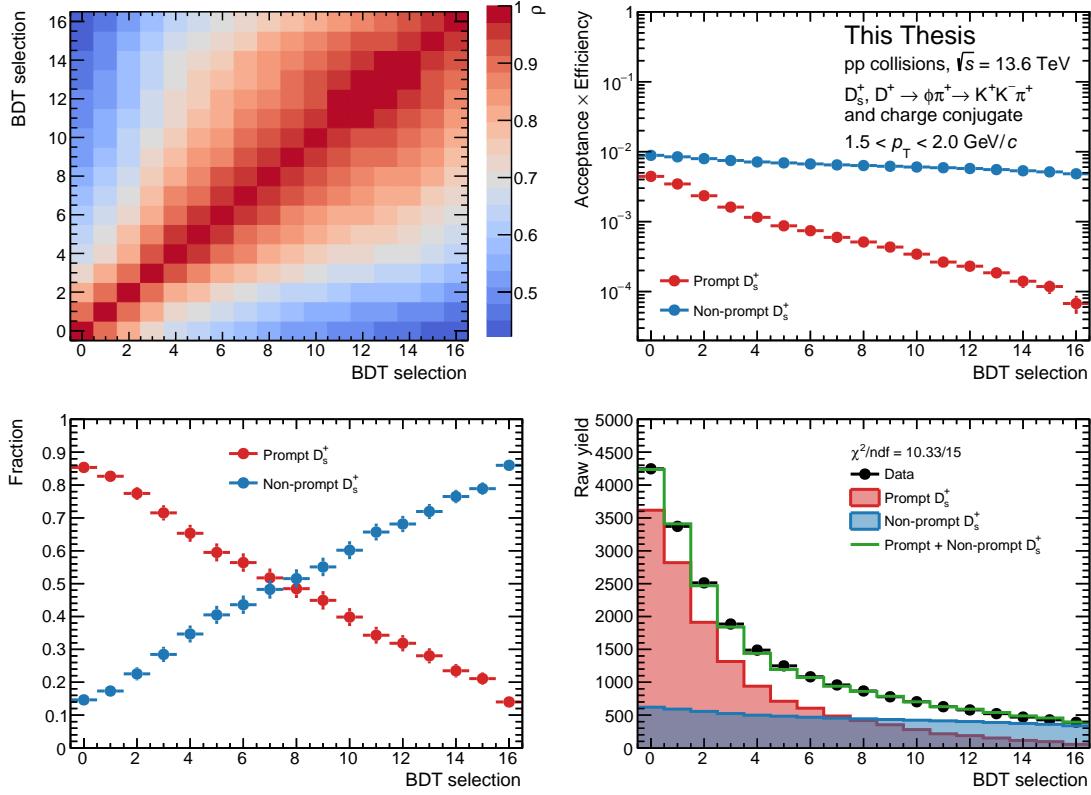


Figure 6.3: Results for the evaluation of the  $D_s^+ f_{\text{prompt}}$  correction factor in the  $1.5 < p_T < 2.0$  GeV/c interval. The correlation factor  $\rho$  (top-left panel), ( $\text{Acc} \times \varepsilon$ ) factors for both prompt and non-prompt  $D_s^+$  mesons (top-right panel), prompt and non-prompt fraction in the extracted  $D_s^+$  raw yields (bottom-left panel), and contribution of prompt and non-prompt  $D_s^+$  mesons to the extracted raw yields (bottom-right panel) are shown for the different BDT selection criteria applied on the non-prompt probability.

where  $\Delta y$  and  $\Delta p_T$  are the rapidity and  $p_T$  intervals, respectively,  $\mathcal{L}_{\text{int}}$  is the integrated luminosity of the data sample, which is  $1 \text{ pb}^{-1}$ ,  $N_{\text{raw}}$  is the extracted raw yield of the considered D meson species, and the factor  $1/2$  takes into account that both particle and antiparticles are selected, while FONLL calculations include only particles. The uncertainty on the  $f_{\text{prompt}}$  factor is propagated through a Gaussian approach, taking into account the uncertainties on the FONLL predictions, raw yield, and  $(\text{Acc} \times \varepsilon)_{\text{non-prompt}}$  factor.

A comparison between the  $f_{\text{prompt}}$  correction factors for  $D_s^+$  and  $D^+$  mesons in the different  $p_T$  intervals of the analysis obtained with the two methods is illustrated in the left panel of Fig. 6.4. A good level of agreement is observed between the two results, which are compatible within their uncertainties across the studied  $p_T$  range, with the exception of the highest  $p_T$  interval, where fluctuations in the yield extraction with the data-driven method lead to a small discrepancy between the two methods. For the evaluation of the central  $f_{\text{prompt}}$  values, the data-driven method is used, as it is not sensitive to possible shortcomings in the theoretical description

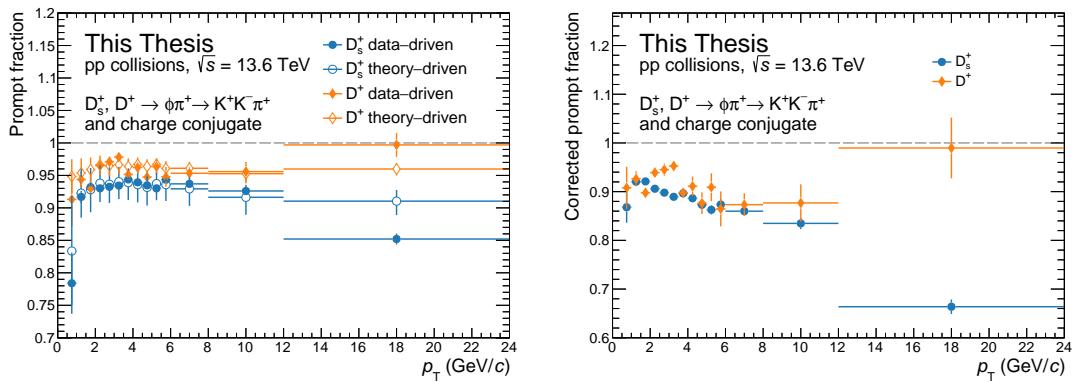


Figure 6.4: Left panel:  $f_{\text{prompt}}$  correction factor for  $D_s^+$  (blue) and  $D^+$  (orange) mesons as a function of  $p_T$ . The results obtained with the data-driven (filled markers) method are compared with those exploiting a theory-driven approach (void markers). Right panel: corrected  $f_{\text{prompt}}$  factors for  $D_s^+$  (blue) and  $D^+$  (orange) mesons as a function of  $p_T$ , obtained with the data-driven method.

of  $b\bar{b}$  production from FONLL which may affect the theory-driven approach. The  $f_{\text{prompt}}$  factors result to be larger for  $D^+$  mesons than for  $D_s^+$  mesons, although being compatible within uncertainties. This is due to the larger mean proper decay length of the former, which results in a larger fraction of non-prompt D mesons in the extracted raw yield, as the selection of signal candidates is based on the displaced decay topology. Additionally, the  $f_{\text{prompt}}$  correction factor is found to be almost constant as a function of  $p_T$ . This differs from the “natural” (corrected) prompt fraction of the two D-meson species, which can be estimated as  $N_{\text{prompt}}/(N_{\text{prompt}} + N_{\text{non-prompt}})$  using the data-driven approach described above, reported in the right panel of Fig. 6.4. A decreasing trend with  $p_T$  is present in the corrected prompt fraction for both  $D_s^+$  and  $D^+$  mesons. The different trend with  $p_T$  of the  $f_{\text{prompt}}$  correction factor as compared to the corrected  $f_{\text{prompt}}$  factor is due to the BDT selection criteria, which are loosened at high  $p_T$  and more promptly produced D mesons are selected.

### 6.3 Branching ratio correction

The branching ratio correction is applied to account for the fact that of the produced  $D_s^+$  and  $D^+$  mesons, only those decaying into a  $\phi\pi^+ \rightarrow K^+K^-\pi^+$  final state are reconstructed. The values of the branching ratios for the considered decays are taken from Ref. [4], and correspond to  $\text{BR}(D_s^+) = (2.21 \pm 0.06) \times 10^{-2}$  and  $\text{BR}(D^+) = ((2.69^{+0.07}_{-0.08}) \times 10^{-3})$ . The uncertainty, taken from the PDG, is propagated to the  $D_s^+/D^+$  production yield ratio with a Gaussian approach, yielding an asymmetric  $^{+3.7\%}_{-4.0\%}$  uncertainty on the ratio.

## 6.4 Systematic uncertainties

Measurements of ratios between different particles’ production yields allow for the cancellation of some of the systematic uncertainties which have to be taken into account for the measurement of production cross sections, such as that related to luminosity. In addition, the reconstruction of  $D_s^+$  and  $D^+$  mesons through the same decay channel allows for the cancellation of supplementary sources of systematic uncertainties, such as those related to tracking and PID efficiency, thereby enhancing the precision of the results. Nonetheless, the measurement is still affected by several sources of systematic uncertainties, due to arbitrary choices made in the analysis, or the need to rely on MC simulations, which could not perfectly reproduce the data.

The first systematic uncertainty, related to the extraction of the  $D_s^+$ - and  $D^+$ -meson raw yields, has already been discussed in Chapter 5. In the next sections, other sources of systematic uncertainty, related to the correction factors presented above, are discussed.

### 6.4.1 BDT selection efficiency

The choice of the set of BDT threshold values used to extract the raw yields is arbitrary, albeit driven by a defined criterion through the optimisation of the significance on a subsample of the analysed dataset. Variations in the ML selection criteria may lead to differences in the extracted raw yields, and therefore in the  $D_s^+/D^+$  production yield ratio, due to possible imperfections in the MC description of the topological, kinematic, and PID variables used in the training of the BDT model. The systematic uncertainty on the BDT selection efficiency is assessed by repeating the analysis varying the BDT threshold values on prompt and background probabilities, independently and then simultaneously, for a total of 48 variations. Three tighter and three looser BDT thresholds on both the background and prompt D meson scores are used on top of those employed for the central case. To avoid extreme variations in the selection criteria, the  $(\text{Acc} \times \varepsilon)_{\text{prompt}}$  for both D-meson species is required to be within 30% of that obtained for the central selections. In addition, to ensure a good fit quality, the  $\chi^2/\text{ndf}$  of the invariant mass fits is required to be below 2, and the statistical significance of the extracted signal is required to be larger than 3 and than half of that obtained with the central selections, for both  $D_s^+$  and  $D^+$  mesons.

The results for the estimation of the systematic uncertainty on the BDT selection efficiency are shown in Fig. 6.5 for the  $4.5 < p_T < 5.0 \text{ GeV}/c$   $p_T$  interval. In the top row, the raw yields of  $D_s^+$  and  $D^+$  mesons, the prompt and non-prompt  $D_s^+$  selection efficiency, the prompt and non-prompt  $D^+$  selection efficiency, and the prompt fraction of  $D_s^+$  and  $D^+$  mesons are shown for the different BDT selections used to estimate the systematic uncertainty, normalised to their respective central values. In the bottom row, the statistical significance of the  $D_s^+$  and  $D^+$  extracted signals, the signal-over-background ratio for the two meson species, and the  $D_s^+/D^+$  production yield ratio are reported for the considered BDT selections. The different results are ordered such that the central value is the leftmost one (BDT selection 0). In the following six variations, the background score threshold is fixed to the

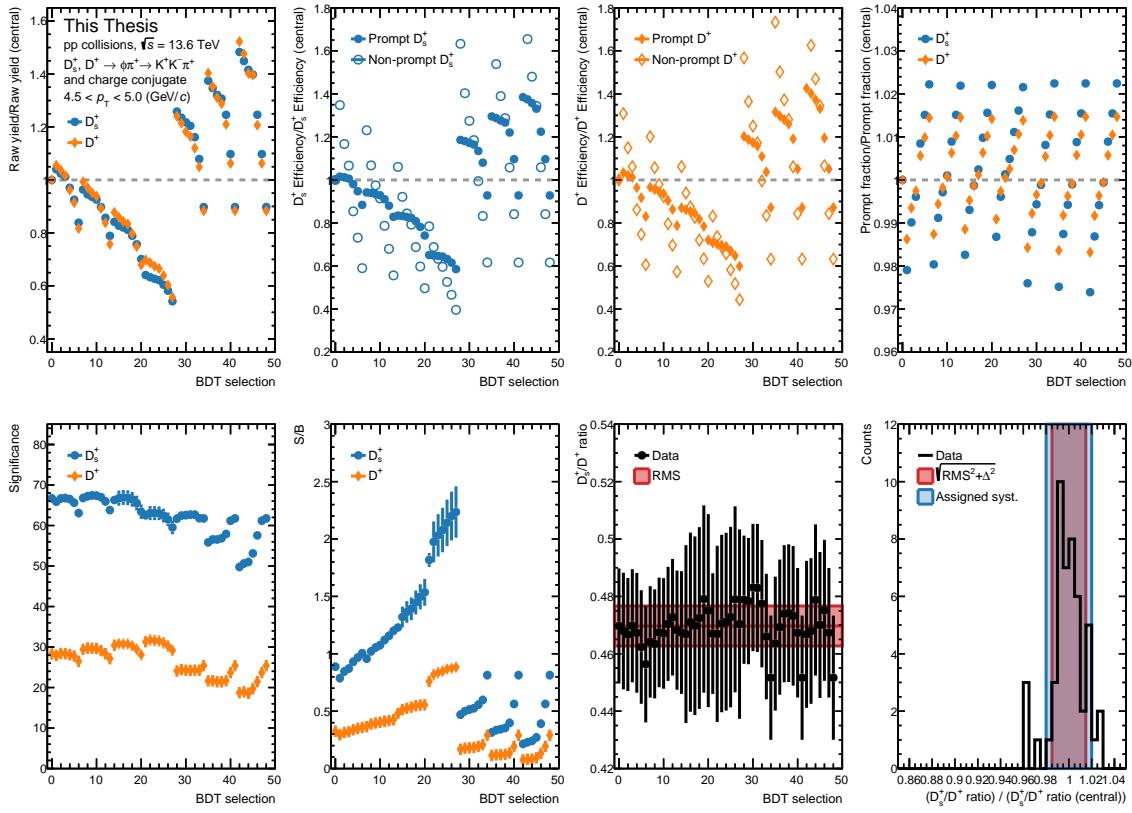


Figure 6.5: Results for the evaluation of the systematic uncertainty on the BDT selection efficiency for the  $4.5 < p_T < 5.0$  GeV/c  $p_T$  interval. In the top row, the raw yields of  $D_s^+$  (blue) and  $D^+$  (orange) mesons, the prompt (filled markers) and non-prompt (void markers)  $D_s^+$  selection efficiency, the prompt and non-prompt  $D^+$  selection efficiency, and the prompt fraction of  $D_s^+$  and  $D^+$  mesons are shown for the different BDT selections used to estimate the systematic uncertainty, normalised to their respective central values. In the bottom row, the statistical significance of the  $D_s^+$  and  $D^+$  extracted signals, the signal-over-background ratio for the two meson species, and the  $D_s^+/D^+$  production yield ratio are reported for the different BDT selections. A red band is shown in this last panel around the central value, whose width represents the RMS of the measurements. The rightmost bottom panel shows the distribution of the  $D_s^+/D^+$  production yield ratio for the different BDT selections divided by the central value. A red band is shown around the central value, representing the sum in quadrature of the standard deviation of the  $D_s^+/D^+$  distribution and of the difference between its mean and the default value. The blue band displays the assigned systematic uncertainty.

central case, while the prompt score threshold is varied through the whole set of considered values. Then, in the next 21 variations, three different tighter thresholds are applied to the background score, while the prompt score is varied. The last 21 variations are obtained by applying three looser thresholds to the background score and varying the prompt score. In the rightmost bottom panel, the distribution of

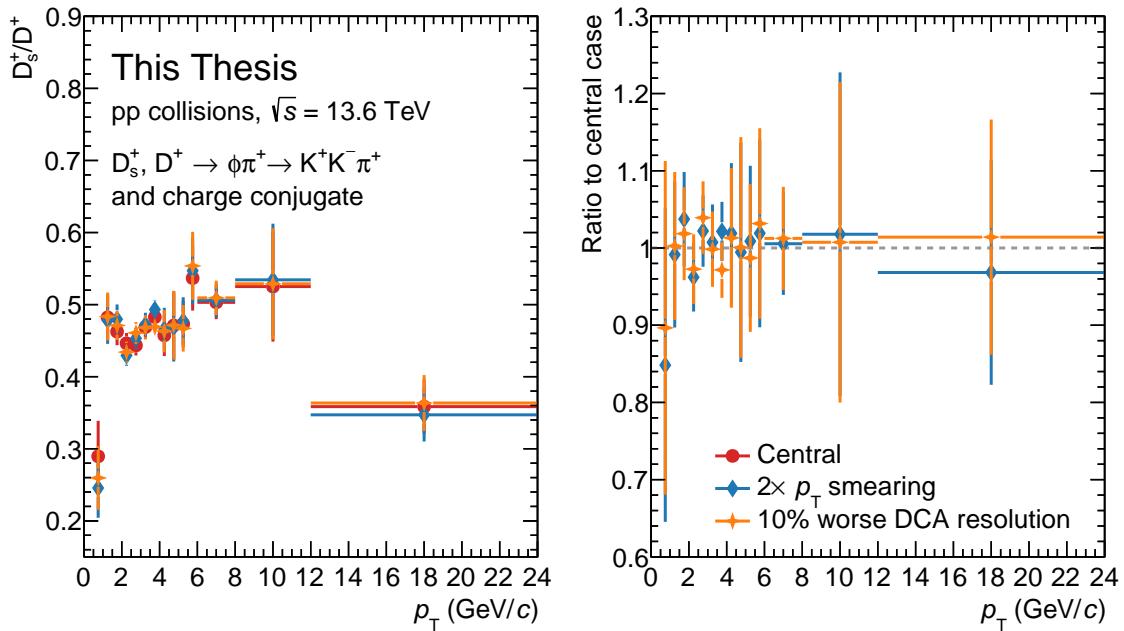


Figure 6.6: Results for the evaluation of the systematic uncertainty on a possible imperfect description of the tracking performance in MC simulations. The  $D_s^+/D^+$  production yield ratio is shown in the left panel for the central result (red) and for the two different configurations of the `track-tuner`: with a factor 2 worse  $p_T$  resolution (blue) and a 10% worse resolution on the track impact parameter (distance of closest approach (DCA), orange). In the right panel, the ratio of the different  $D_s^+/D^+$  factors to that obtained in the central case is reported.

the  $D_s^+/D^+$  production yield ratio for the different BDT selections divided by the central value is shown. The systematic uncertainty on the BDT selection efficiency is estimated as the sum in quadrature of the standard deviation of the  $D_s^+/D^+$  distribution and of the difference ( $\Delta$ ) between its mean and the value of the ratio obtained with the default selections. The relative uncertainty is then rounded to the nearest 1% value, and to avoid the inclusion of statistical fluctuations in the assigned systematic uncertainty, the  $p_T$  dependence of the systematic uncertainty is smoothed. The assigned systematic uncertainty is reported in Table 6.1 and ranges from 1% to 10% depending on  $p_T$ . At low  $p_T$ , where the efficiency is lower and the sensitivity to possible shortcomings in the MC description of data increases, the uncertainty is larger, reaching 10% in the lowest  $p_T$  interval. The uncertainty then follows a decreasing trend with  $p_T$ , reaching 2% in the highest  $p_T$  intervals.

#### 6.4.2 Imperfections in the description of the tracking performance

As discussed above, an imperfect description of the topological variables in the MC simulations may imply a systematic uncertainty in the determination of the selection efficiencies of  $D_s^+$  and  $D^+$  mesons. A contribution to this uncertainty was

tested by varying the topological selections applied and evaluating the stability of the result against the variations, as discussed in Sec. 6.4.1. In addition to this, the sensitivity of the measurement to possible discrepancies in the impact parameter resolution between data and MC was evaluated by using different configurations of the **track-tuner** workflow. In particular, a configuration with a 10% worse resolution than the default one was considered. Moreover, a version of the **track-tuner** implementing a smearing of the track  $p_T$  at the innermost update point was also used. In this case, a factor 2 worse  $p_T$  resolution was used to account for the discrepancy between the widths of the  $D_s^+$ - and  $D^+$ -meson mass peaks in data and MC. The analysis was repeated using the two different configurations of the **track-tuner** for estimating the correction factors, and the resulting  $D_s^+/D^+$  production yield ratios are shown in Fig. 6.6, together with the ratio to the central result. The systematic uncertainty is estimated for each  $p_T$  interval as the semi-dispersion of the three results. As for the other sources of systematic uncertainty, the  $p_T$  dependence of the systematic uncertainty is smoothed. The assigned systematic uncertainty is reported in Table 6.1, and ranges from 1% to 8% depending on  $p_T$ . It is maximum at low  $p_T$ , where possible imperfections in the MC simulations have a larger impact on the measurement, and decreases with increasing  $p_T$ .

### 6.4.3 Prompt fraction

The choice of the set of selection criteria employed for evaluating the  $f_{\text{prompt}}$  correction via the data-driven procedure described in Sec. 6.2 may introduce a systematic uncertainty on the  $D_s^+/D^+$  production yield ratio, due to possible imperfections in the description of the efficiency evolution as a function of the applied selections. To estimate the magnitude of this uncertainty, the  $f_{\text{prompt}}$  correction is evaluated using different sets of BDT threshold values. Of the 17 variations on the non-prompt score threshold used to estimate the prompt fraction correction as described in Sec. 6.2, a subsample of 13 is used to estimate the systematic uncertainty. The  $f_{\text{prompt}}$  estimation is repeated using only the first, last, and central selections, corresponding to the loosest, tightest, and intermediate selection criteria. For each set of BDT threshold values, the  $f_{\text{prompt}}^{D_s^+}$  and  $f_{\text{prompt}}^{D^+}$  are extracted using the same procedure as in Sec. 6.2. The correction factor applied to Eq. 6.1 is then calculated for each  $p_T$  interval as the ratio between the  $f_{\text{prompt}}^{D_s^+}$  and  $f_{\text{prompt}}^{D^+}$  correction factors, considering all the 16 possible combinations of results from the four different sets of BDT selections for  $D_s^+$  and  $D^+$  mesons (also the central values are considered). A total of 16 values of the  $f_{\text{prompt}}^{D_s^+}/f_{\text{prompt}}^{D^+}$  ratio are obtained for each  $p_T$  interval, as shown in the left panel of Fig. 6.7. On the right panel, the ratio of the different  $f_{\text{prompt}}^{D_s^+}/f_{\text{prompt}}^{D^+}$  factors to that obtained in the central case is reported. The systematic uncertainty on the  $f_{\text{prompt}}$  correction is then estimated as the standard deviation of the  $f_{\text{prompt}}^{D_s^+}/f_{\text{prompt}}^{D^+}$  distribution. To avoid the inclusion of statistical fluctuations in the systematic uncertainty, the  $p_T$  dependence of the systematic uncertainty is smoothed. The assigned systematic uncertainty is reported in Table 6.1, and ranges from 1% to 4% depending on  $p_T$ . It is minimum at intermediate  $p_T$ , and increases both at low  $p_T$ , where the efficiency is lower leading to a larger sensitivity to the imperfections in the MC simulations, and at high  $p_T$ , where it is more difficult to vary by a large

amount the prompt fraction in the selected data sample, and therefore the  $N_{\text{prompt}}$  and  $N_{\text{non-prompt}}$  result determined with less precision.

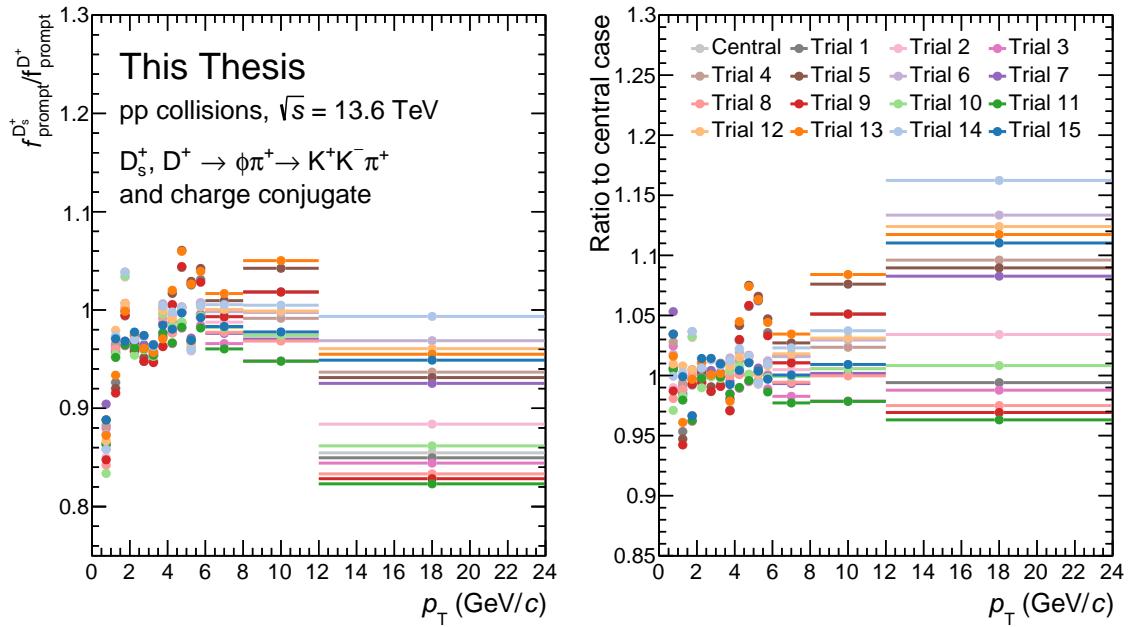


Figure 6.7: Results for the evaluation of the systematic uncertainty on the  $f_{\text{prompt}}$  correction. In the left panel, the different  $f_{\text{prompt}}^{D_s^+}/f_{\text{prompt}}^{D^+}$  corrections obtained with different sets of BDT selections are shown. In the right panel, their ratio to that obtained in the central case is reported.

#### 6.4.4 Generated Monte Carlo $p_T$ shape

A non-realistic  $p_T$  shape of the generated D mesons in the MC simulations could lead to a bias in the determination of the efficiency because of the  $p_T$  dependence of the  $(\text{Acc} \times \varepsilon)$  corrections within the  $p_T$  intervals considered in the analysis. This source of systematic uncertainty is expected to be negligible, as: i) it was observed to have almost no impact for D mesons in previous analyses (e.g., in Refs. [62] and [180]), ii) because possible biases in the  $p_T$  shape of the generated D mesons would affect both  $D_s^+$  and  $D^+$  mesons in a similar way, canceling out in the ratio, and iii) because very narrow  $p_T$  intervals are considered in the analysis. Nonetheless, the efficiency steeply increases with increasing  $p_T$  at low  $p_T$ , as shown in Fig. 6.2. Therefore, it was checked that changes in the  $p_T$  distribution of generated D mesons in the MC simulations do not affect the results. The systematic uncertainty is estimated by changing the  $p_T$  distribution of the generated D mesons in the MC simulations used for the evaluation of the  $(\text{Acc} \times \varepsilon)$  corrections, which are based on the PYTHIA 8 [98] with colour-reconnection Mode 2 [104] event generator, as described in Sec. 6.1. The  $p_T$  distribution of the generated D mesons is modified to that predicted by FONLL [63]. Both PYTHIA 8 with colour-reconnection Mode 2 and FONLL predictions describe the  $p_T$ -differential cross-section of D mesons measured by the ALICE Collaboration

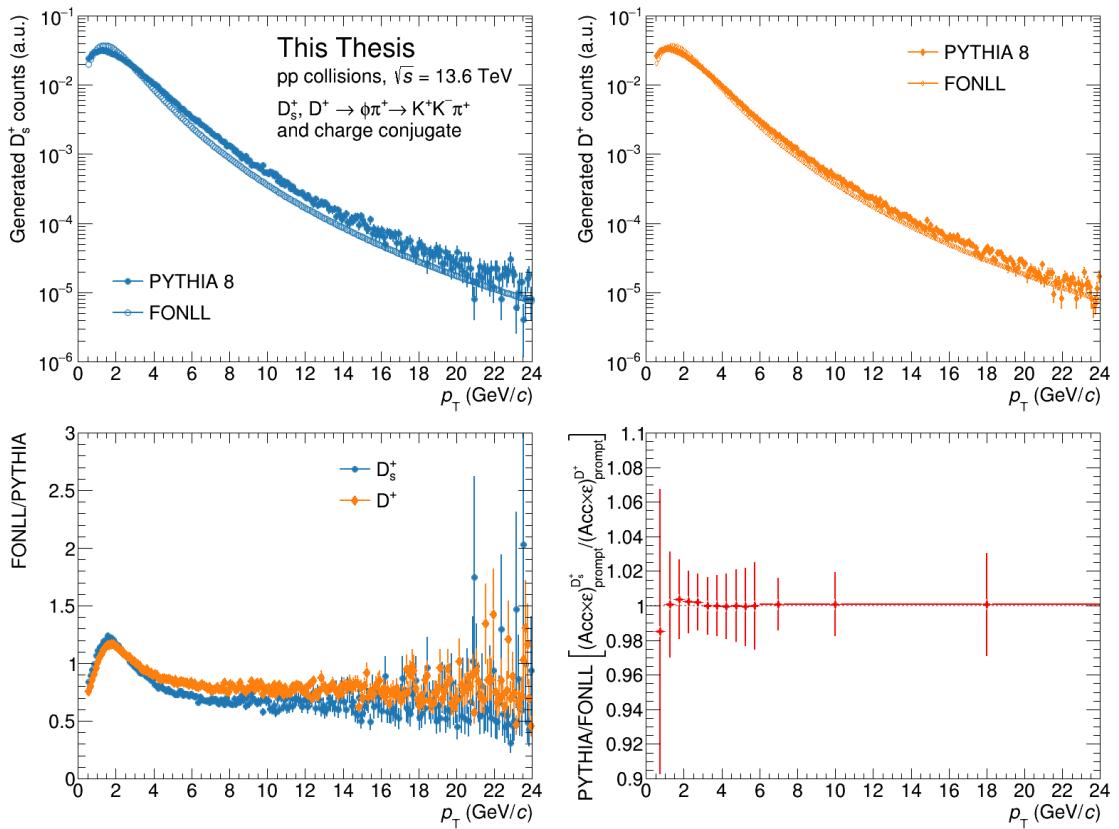


Figure 6.8: Comparison of the  $p_T$  distributions of generated  $D_s^+$  (left panel) and  $D^+$  (right panel) mesons in the MC simulations used for evaluating the  $(\text{Acc} \times \varepsilon)$  corrections. The distributions obtained with the PYTHIA 8 event generator are shown with filled markers, while those obtained with FONLL are shown with void markers.

in pp collisions at  $\sqrt{s} = 13$  TeV [62], and are therefore suitable models to describe the  $p_T$  shape of the D mesons.

The  $p_T$  distributions of  $D_s^+$  ( $D^+$ ) mesons are shown in the top-left (top-right) panel of Fig. 6.8 for PYTHIA 8 (filled markers) and FONLL (void markers) predictions. In the bottom-left panel, the ratio between the FONLL and PYTHIA  $p_T$  distributions is shown for  $D_s^+$  and  $D^+$  mesons. Lastly, the bottom-right panel shows the ratio between the FONLL- and PYTHIA 8-based  $(\text{Acc} \times \varepsilon)_{\text{prompt}}^{D_s^+} / (\text{Acc} \times \varepsilon)_{\text{prompt}}^{D^+}$  factors, which is the quantity that is used to correct the  $D_s^+ / D^+$  production yield ratio in Eq. 6.1. The systematic uncertainty is estimated as the difference between the central result (obtained with PYTHIA 8 simulations) and the one obtained with the FONLL  $p_T$  shape. It resulted to be  $< 1\%$  for the considered  $p_T$  intervals, with the exception of the lowest interval, where a 1% effect is present. However, given that this uncertainty is much smaller than the other sources of systematic uncertainty, it is not considered in the final result.

### 6.4.5 Summary of systematic uncertainties

The total systematic uncertainty assigned to the different  $p_T$  intervals considered in the analysis is evaluated by summing in quadrature the contributions from the different sources of systematic uncertainty, as the sources are considered to be uncorrelated. The assigned systematic uncertainty ranges from 4% to 13% depending on  $p_T$ . A summary of the considered systematic uncertainties, together with the total systematic uncertainty on the  $D_s^+/D^+$  production yield ratio, is reported in Table 6.1.

Table 6.1: Summary of the assigned systematic uncertainties on the  $D_s^+/D^+$  production yield ratio.

$p_T$ (GeV/ $c$ )	Raw yield extraction (%)	BDT selection efficiency (%)	Tracking performance (%)	Prompt fraction (%)	Tot. syst. unc. (%)	BR (%)
0.5–1	3	10	8	2	13	
1–1.5	3	5	2	2	7	
1.5–2	3	3	2	2	5	
2–2.5	3	3	2	1	5	
2.5–3	3	2	2	1	4	
3–3.5	3	2	1	1	4	
3.5–4	3	1	1	1	4	
4–4.5	5	1	1	2	6	$+3.7$
4.5–5	5	1	1	2	6	$-4.0$
5–5.5	5	2	1	2	6	
5.5–6	5	2	1	2	6	
6–8	8	2	1	2	8	
8–12	9	2	1	3	10	
12–24	10	2	2	4	13	

# Chapter 7

## $D_s^+/D^+$ production-yield ratio

The  $p_T$ -differential  $D_s^+/D^+$  production-yield ratio can be evaluated using the raw yields and the correction factors obtained in Chapters 5 and 6. The  $D_s^+/D^+$  production-yield ratio is defined using Eq. 6.1, here reported for convenience:

$$D_s^+/D^+ = \frac{N_{\text{raw}}^{D_s^+} \cdot f_{\text{prompt}}^{D_s^+}}{(Acc \times \varepsilon)_{\text{prompt}}^{D_s^+} \cdot BR^{D_s^+}} \cdot \left( \frac{N_{\text{raw}}^{D^+} \cdot f_{\text{prompt}}^{D^+}}{(Acc \times \varepsilon)_{\text{prompt}}^{D^+} \cdot BR^{D^+}} \right)^{-1}.$$

Previous measurements of this ratio have been performed in pp collisions at different centre-of-mass energies and rapidity windows, and it is therefore interesting to investigate whether the  $D_s^+/D^+$  production-yield ratio depends on either of these variables. In addition, several models have been developed to describe the production of the two D-meson species in pp collisions. They are all based on perturbative QCD calculations for the description of the charm-quark production in the hard-scattering processes (although through different approaches), but differ in the description of the parton shower and hadronisation process, as discussed in Chapter 2. The prompt  $D_s^+/D^+$  production-yield ratio is a sensitive probe of the hadronisation mechanism of charm quarks, and the comparison between the measured values and the predictions from these models can be used to test their validity. Furthermore, these results provide a baseline for analogous measurements in Pb–Pb collisions.

Using the factorisation theorem [61], from Eq. 2.1, the  $D_s^+/D^+$  production-yield ratio is sensitive to the fragmentation functions of the charm quark into the two D-meson species, as the description of the parton distribution functions of the colliding protons and the charm production in the partonic scattering is the same for the two D-meson species. The ratio of the  $p_T$ -differential cross-sections of the two D-meson species can be expressed as:

$$\frac{\frac{d^2\sigma}{dp_T dy} \Big|_{\text{prompt}}^{D_s^+}}{\frac{d^2\sigma}{dp_T dy} \Big|_{\text{prompt}}^{D^+}} = \frac{\int dz D_{c \rightarrow D_s^+}(z, \mu_F^2)}{\int dz D_{c \rightarrow D^+}(z, \mu_F^2)} = \frac{f_{c \rightarrow D_s^+}}{f_{c \rightarrow D^+}},$$

where  $D_{c \rightarrow D_s^+(D^+)}(z, \mu_F^2)$  are the fragmentation functions of the charm quark for the hadronisation into a  $D_s^+$  ( $D^+$ ) meson and their integrals with respect to  $z$ ,  $f_{c \rightarrow D_s^+(D^+)}$ , are called *fragmentation fractions*. These studies therefore provide information on

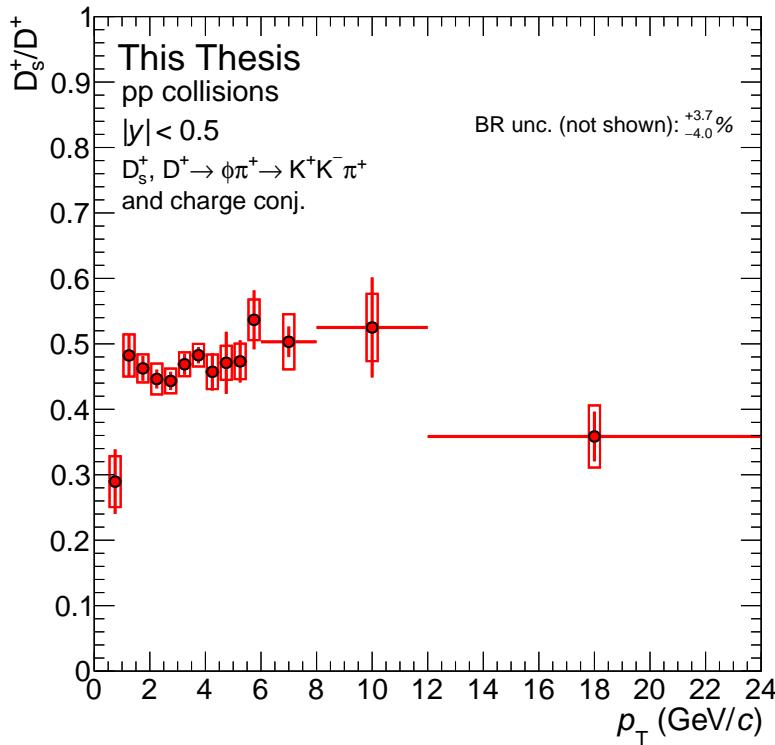


Figure 7.1:  $p_T$ -differential  $D_s^+/D^+$  production-yield ratio measured in this Thesis at midrapidity ( $|y| < 0.5$ ) in pp collisions at  $\sqrt{s} = 13.6$  TeV with the ALICE experimental apparatus.

the energy and rapidity dependence of the fragmentation functions, and, potentially, constraints for these quantities, which are crucial ingredients for the perturbative calculation of the open-charm hadron production cross-section in pp collisions in the framework of the factorisation theorem.

The  $p_T$ -differential  $D_s^+/D^+$  production-yield ratio measured in this Thesis at midrapidity ( $|y| < 0.5$ ) in pp collisions at  $\sqrt{s} = 13.6$  TeV with the ALICE detector is shown in Fig. 7.1. The statistical uncertainties are reported using vertical bars, while the systematic uncertainties are shown as boxes. The results present a steeply-increasing  $D_s^+/D^+$  production-yield ratio between the first two  $p_T$  intervals, followed by a flat trend with  $p_T$  for  $p_T > 1$  GeV/c.

## 7.1 Dependence on centre-of-mass energy

Measurements performed in heavy-ion collisions at the SPS observed an enhancement in the production of strange hadrons with respect to non-strange hadrons [53, 54], which was attributed to the modification of the hadronisation process in the QGP and the augmented thermal production of strange quarks as compared to pp collisions owing to the high temperatures reached in the medium. At the LHC, measurements performed in pp, p–Pb, and Pb–Pb collisions by the ALICE Collaboration [57, 56, 59] showed that the phenomenon of strangeness enhancement

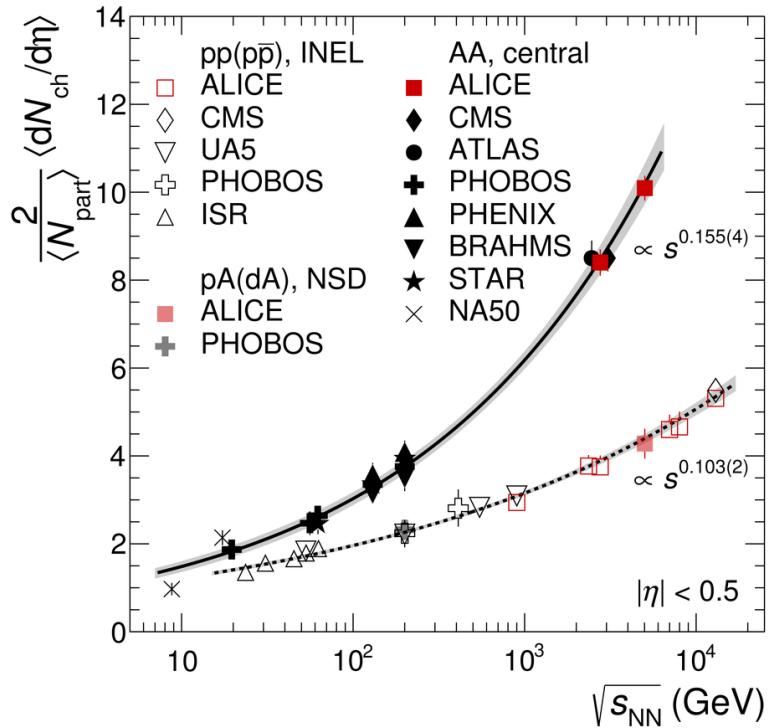


Figure 7.2: Charged-particle multiplicity density, normalised to the number of nucleons participating in the collision divided by a factor 2 ( $\langle N_{\text{part}} \rangle / 2$ ) as a function of the centre-of-mass energy per nucleon pair of the collision for central nucleus-nucleus, proton-nucleus, and pp collisions. The  $s_{\text{NN}}$ -dependence of the charged-particle multiplicity density is parametrised using a power-law function, and is shown with solid and dashed lines for central nucleus-nucleus collisions and smaller collision systems, respectively. Figure taken from Ref. [184]

is also present in small collision systems, in high charged-particle multiplicity collisions. The  $K_s^0/\pi$ ,  $\Lambda/\pi$ ,  $\Xi/\pi$ , and  $\Omega/\pi$  production-yield ratios were observed to follow a smoothly increasing trend with multiplicity from pp to Pb–Pb collisions. This behaviour is described by different models (such as PYTHIA with colour-reconnection plus rope hadronisation [124], Statistical Hadronisation Model [26], and the Catainia coalescence model [117]), which implement different approaches to describe the hadronisation process, but all assume the establishment of a parton-rich environment in the collision.

The charged-particle multiplicity of a collision is observed to increase with the centre-of-mass energy of the event [184, 185] because the larger energy available can be used to produce more particles. This is shown in Fig. 7.2, where the charged-particle multiplicity density, normalised to the number of nucleons participating in the collision divided by a factor 2 ( $\langle N_{\text{part}} \rangle / 2$ ), is shown as a function of the centre-of-mass energy per nucleon pair of the collision ( $\sqrt{s_{\text{NN}}}$ ) for central nucleus-nucleus, proton-nucleus, and pp collisions. The  $s_{\text{NN}}$ -dependence of the charged-particle multiplicity density is parametrised using a power-law function, and is shown with solid and dashed lines for central nucleus-nucleus collisions and smaller collision

systems, respectively.

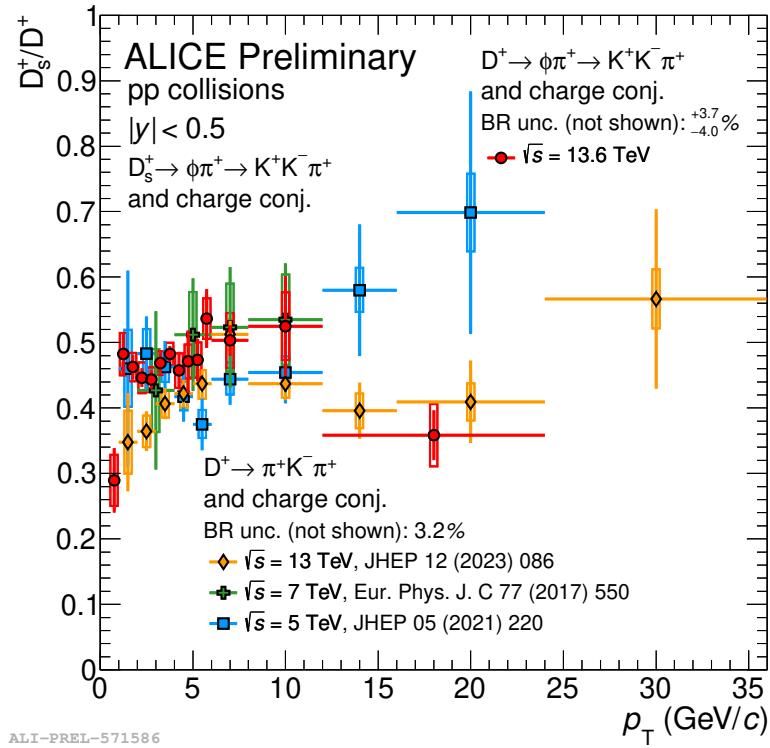


Figure 7.3:  $p_T$ -differential  $D_s^+/D^+$  production-yield ratio measured at midrapidity ( $|y| < 0.5$ ) in pp collisions by the ALICE Collaboration. The results obtained in this Thesis at  $\sqrt{s} = 13.6$  TeV (red) are compared with previous measurements in pp collisions at  $\sqrt{s} = 5$ , 7, and 13 TeV (blue, green, and orange, respectively).

Therefore, an enhancement in the production of strange hadrons might be expected with the increase in the energy available in the collision. The dependence on the centre-of-mass energy of the  $D_s^+/D^+$  production-yield ratio can be studied by comparing the results obtained in this Thesis in pp collisions at  $\sqrt{s} = 13.6$  TeV with those obtained, in the same rapidity window of  $|y| < 0.5$ , by the ALICE Collaboration in pp collisions at the centre-of-mass energies of  $\sqrt{s} = 5$ , 7, and 13 TeV (taken from Refs. [62, 148, 180], respectively). Given the small difference of a few TeV between the considered measurements, a very small effect on the  $D_s^+/D^+$  production-yield ratio is expected. Using the parametrisations of the charged-particle multiplicity density as a function of the centre-of-mass energy per nucleon pair of the collision reported in Ref [184] and shown in Fig. 7.2, a factor of about  $(13.6/5.02)^{0.206} \sim 1.2$  is expected between the multiplicities measured at  $\sqrt{s} = 13.6$  TeV and  $\sqrt{s} = 5.02$  TeV (since the number of participating nucleons  $N_{\text{part}}$  in a pp collision is always 2: the two protons).

The comparison is shown in Fig. 7.3, with red markers being used for the measurement performed in this Thesis. Thanks to the upgrade to the ALICE experimental apparatus [128], a very large number of minimum-bias events was recorded. In one year of Run 3 data-taking, an integrated luminosity larger than that of the entire Run 2 was collected. In addition, previous measurements of the  $D_s^+/D^+$  production-

yield ratio reconstructed the  $D^+$  meson through the  $D^+ \rightarrow \pi^+ K^- \pi^+$  decay channel, which differs from that exploited in this analysis. Despite being characterised by a larger BR of  $(9.38 \pm 0.16) \times 10^{-2}$  [4], which leads to a larger amount of reconstructed  $D^+$  mesons, the reconstruction through the same decay channel used for the  $D_s^+$  meson,  $D^+ \rightarrow \phi \pi^+ \rightarrow K^+ K^- \pi^+$ , allows for the cancellation of some of the systematic uncertainties in the  $D_s^+/D^+$  ratio. Thereby, a high-precision measurement of the  $D_s^+/D^+$  production-yield ratio was achieved in this Thesis. Both statistical and systematic uncertainties have been significantly reduced with respect to previous results. The  $p_T$  reach of the measurement has been extended to lower values, down to  $p_T = 0.5 \text{ GeV}/c$ , and narrower  $p_T$  intervals have been analysed. At higher  $p_T$ , the worse  $p_T$  resolution on the tracks due to the increase of the electric field distortions in the TPC is more impactful, and the uncertainties increase. For the same reason, the  $p_T$  reach of the measurement is limited to  $p_T < 24 \text{ GeV}/c$ , with no improvements in the granularity of the measurement at high  $p_T$ . The comparison with previous measurements at midrapidity performed by the ALICE Collaboration at the centre-of-mass energies of  $\sqrt{s} = 5, 7$ , and  $13 \text{ TeV}$  shows no significant dependence on the centre-of-mass energy, with results being compatible within uncertainties across the whole analysed  $p_T$  and  $\sqrt{s}$  ranges. The comparison with the results obtained at  $\sqrt{s} = 13 \text{ TeV}$  shows that a slightly larger  $D_s^+/D^+$  production-yield ratio is measured at low  $p_T$  at  $\sqrt{s} = 13.6 \text{ TeV}$ . However, the large uncertainties in the previous measurement prevent any firm conclusion on any energy dependence of the observable. In addition, large fluctuations may be present in the  $D_s^+$  and  $D^+$  measurements at  $\sqrt{s} = 13 \text{ TeV}$ , as can be deduced from the different trends of their production-yield ratio to the  $D^0$  meson compared to that measured at  $\sqrt{s} = 5.02 \text{ TeV}$ .

## 7.2 Dependence on rapidity

Measurements of the  $D_s^+/D^+$  production-yield ratio in different rapidity intervals can provide further insights into the hadronisation mechanism. At forward rapidities, the production of particles closer to the beam remnants is probed. In this phase space region, interactions may occur between the fragments of the colliding hadrons and the partons produced in the hard-scattering processes. Therefore, the parton shower and the hadronisation might differ from those at midrapidity. Measurements performed at forward rapidity by colliding a beam of  $340 \text{ GeV}/c \pi^-$  on a nuclear target at the SPS by the WA82 Collaboration [186] and by colliding a beam of  $\pi^\pm, K^\pm$ , or protons on a nuclear target at the low energy of  $250 \text{ GeV}$  by the E769 Collaboration at Fermilab [187] indicated the presence of an asymmetric production of leading and subleading charm particles. The term *leading* particle refers in this context to a charm particle that shares at least one light quark or antiquark flavour with the beam particle. On the contrary, recent measurements performed by the ALICE Collaboration demonstrated a balance in the production of particles and antiparticles at midrapidity [32]. This suggests that the hadronisation may undergo a change in the vicinity of the beam remnants. These asymmetries at rapidities close to the beam rapidity can be interpreted in the framework of hadronisation via coalescence, by assuming that the charm quark hadronises via recombination with some other quarks belonging to the beam remnants [188, 189]. At LHC energies,

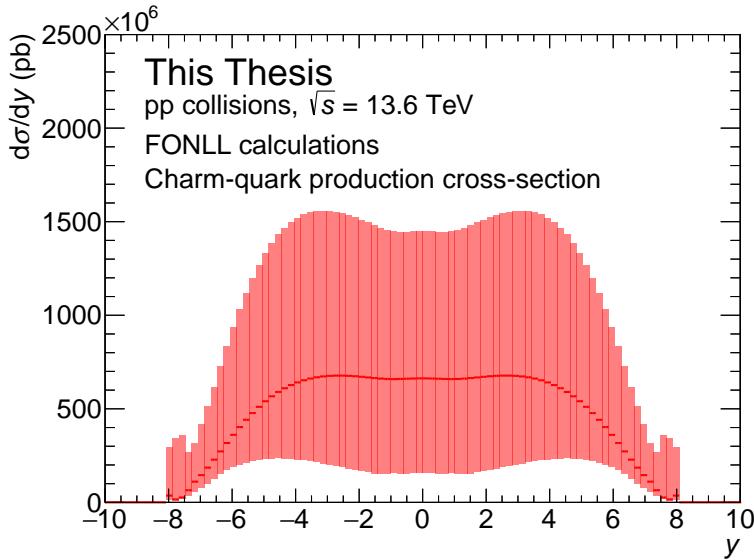


Figure 7.4:  $y$ -differential charm-quark production cross-section in pp collisions at  $\sqrt{s} = 13.6$  TeV predicted by FONLL [63] calculations.

rapidities of around 10 in the laboratory frame are reached by both colliding protons. FONLL calculations [63] for the  $y$ -differential charm-quark production cross-section in pp collisions at  $\sqrt{s} = 13.6$  TeV predict an almost rapidity-independent cross-section for charm hadron production in  $|y| < 4$  and a steeply-falling trend with rapidity for  $|y| > 4$ , as shown in Fig. 7.4. The LHCb experimental apparatus is able to perform measurements at forward rapidity in  $2.0 < y < 4.5$ , but does not reach the large  $y$  values of the beam remnants. Therefore, it is not expected to observe significant differences between measurements performed in the ALICE and LHCb rapidity ranges.

The rapidity dependence of the  $D_s^+/D^+$  production-yield ratio can be investigated by comparing the measurement at midrapidity ( $|y| < 0.5$ ) performed in this Thesis in pp collisions at  $\sqrt{s} = 13.6$  TeV with those performed at forward rapidities by the LHCb Collaboration [190] at a similar centre-of-mass energy of  $\sqrt{s} = 13$  TeV. Given the energy-independence of the observable which has been discussed above, it is possible to directly compare the two measurements. Figure 7.5 shows the comparison between the  $D_s^+/D^+$  production-yield ratio measured in this work compared to that from the LHCb Collaboration measured in pp collisions in the  $2.0 < y < 2.5$  and  $2.5 < y < 3.0$  rapidity intervals (left panel), and in the  $3.0 < y < 3.5$ ,  $3.5 < y < 4.0$ , and  $4.0 < y < 4.5$  rapidity intervals (right panel). Thanks to the large dataset available and the reconstruction of the two D-meson species through the same decay channel, the uncertainties of the measurement are significantly smaller than those of the results obtained by the LHCb Collaboration across a wide  $p_T$  range. In addition, the  $p_T$  reach of the measurement is extended to lower values. The results obtained at midrapidity and forward rapidity show a similar trend, with the  $D_s^+/D^+$  production-yield ratios being compatible within uncertainties across the whole studied  $p_T$  and  $y$  range. The comparison between the results obtained at midrapidity and forward rapidity shows no significant dependence of the  $D_s^+/D^+$  production-yield ratio on

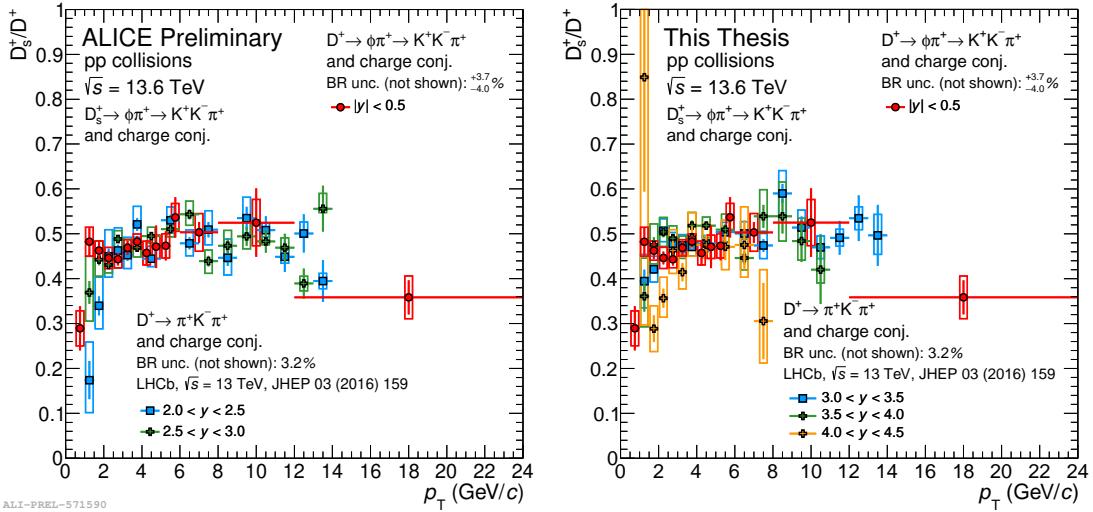


Figure 7.5:  $p_T$ -differential  $D_s^+/D^+$  production-yield ratio measured at midrapidity ( $|y| < 0.5$ ) in pp collisions at  $\sqrt{s} = 13.6$  TeV by the ALICE Collaboration (red), obtained in this Thesis, compared with previous measurement in pp collisions at  $\sqrt{s} = 13$  TeV performed by the LHCb Collaboration in the forward-rapidity intervals of  $2.0 < y < 2.5$  (blue) and  $2.5 < y < 3.0$  (green) in the left panel and the  $3.0 < y < 3.5$  (blue),  $3.5 < y < 4.0$  (green), and  $4.0 < y < 4.5$  (orange) rapidity intervals in the right panel.

the rapidity of the measurement.

Other measurements performed at forward rapidity by the LHCb Collaboration in p–Pb collisions at  $\sqrt{s_{NN}} = 8.16$  TeV [191] present an enhancement of the  $D_s^+/D^+$  production-yield ratio as a function of the charged-particle multiplicity across the studied  $2 < p_T < 12$  GeV/ $c$  interval. These results were obtained by determining the charged-particle multiplicity of the collision using a detector capable of measuring the number of produced particles in the same rapidity interval in which the D-meson production is measured. This may lead to possible biases in the measurement, as heavy-flavour hadrons are typically produced in jets, and auto-correlations between the charged-particle multiplicity and the D-meson production may be present. This could also cause the different slope of the measured  $D_s^+/D^+$  production between forward and backward rapidities. The ALICE apparatus overcomes these limitations, as the charged-particle multiplicity of events containing D mesons in the central barrel can be measured with forward-rapidity detectors, such as the FV0 and the two arrays of the FT0 detector. Therefore, future measurements of the multiplicity dependence of the  $D_s^+/D^+$  production-yield ratio in pp collisions performed by the ALICE Collaboration are expected to provide a more accurate determination of the observable, and to be less affected by biases due to the presence of auto-correlations between the D-meson production and the charged-particle multiplicity of the collision.

### 7.3 Comparison to models

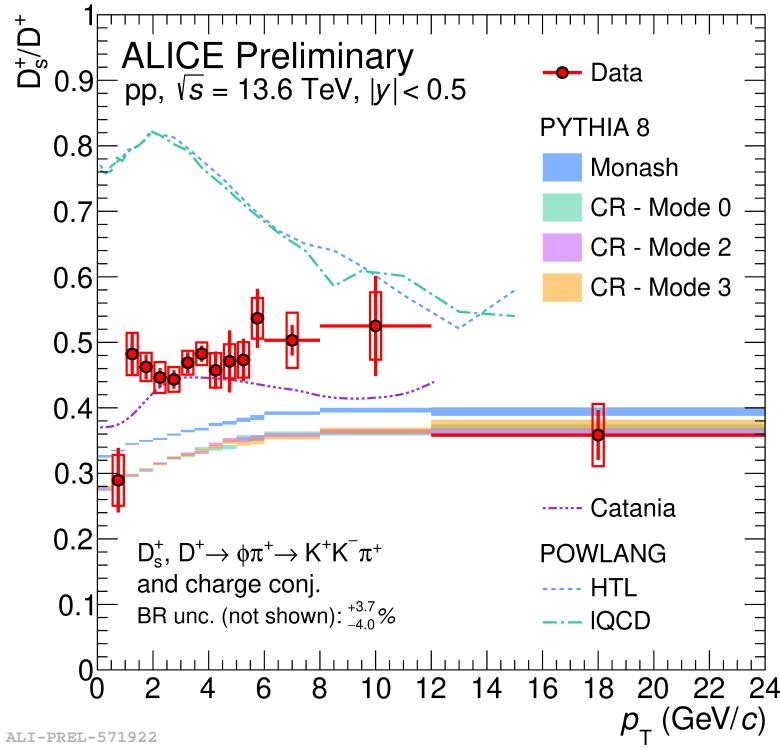


Figure 7.6:  $p_T$ -differential  $D_s^+/D^+$  production-yield ratio measured at midrapidity ( $|y| < 0.5$ ) in pp collisions at  $\sqrt{s} = 13.6$  TeV by the ALICE Collaboration. The results obtained in this Thesis (red) are compared with different theoretical predictions.

The  $D_s^+/D^+$  production-yield ratio can be used to test the validity of models describing the production of the two D-meson species, and set constraints on their description of the hadronisation process. Figure 7.6 presents a comparison between the  $D_s^+/D^+$  production-yield ratio measured in this Thesis at midrapidity ( $|y| < 0.5$ ) in pp collisions at  $\sqrt{s} = 13.6$  TeV with the ALICE experiment (red markers) and theoretical predictions.

Predictions from PYTHIA 8 [98] are reported using the Monash 2013 tune [93] (blue filled boxes), which tunes its hadronisation description on results from  $e^+e^-$  collisions, and CR tunes implementing colour-reconnections beyond the leading-colour approximation [104]. The width of the PYTHIA 8 bands in Fig. 7.6 represents the statistical uncertainty due to the limited amount of simulated pp collision events. Three colour reconnection modes (Mode 0,2,3) are reported in green, violet, and orange, and present similar trends with  $p_T$ , with values compatible within uncertainties. PYTHIA 8 predictions with Monash 2013 tune present a similar dependence on  $p_T$ , with a predicted  $D_s^+/D^+$  production-yield ratio that is larger than that with colour-reconnections beyond the leading-colour approximation. Nonetheless, the four predictions underestimate the measured  $D_s^+/D^+$  ratio, by a factor of about 1.5 at intermediate  $p_T$ . This discrepancy is also observed in the measurement

of the  $p_T$  integrated  $D_s^+/D^0$  and  $D^+/D^0$  production-yield ratio in pp collisions at 5.02 TeV performed by the ALICE Collaboration [192], where the same PYTHIA 8 predictions were found to slightly underestimate the  $D_s^+$ -meson production and overestimate that of  $D^+$  mesons.

Predictions from the Catania model [117], which assumes the production of small droplets of QGP in pp collisions, implementing both the coalescence and fragmentation mechanisms for charm quark hadronisation, are reported with a violet dot-dashed line. The model provides a good description of the  $D_s^+/D^+$  production-yield ratio, managing to reproduce both the magnitude and the general trend with transverse momentum in the range  $2 < p_T < 6 \text{ GeV}/c$ , showing some tension with the data at low and high  $p_T$ .

Lastly, predictions from POWLANG [118], are also reported. Similarly to the Catania model, POWLANG assumes the formation of a small deconfined system in pp collisions, and the same in-medium hadronisation mechanism developed for heavy-ion collisions, based on local colour neutralisation via the formation of clusters of quarks and diquarks, is employed. Two sets of predictions are reported, employing transport coefficients calculated with weak-coupling calculations [193] (Hard-Thermal-Loop, HTL) and lattice-QCD simulations [194], using blue and green dashed lines, respectively. Both predictions overestimate the measured  $D_s^+/D^+$  production-yield ratio, by a factor of about 1.5 at intermediate  $p_T$ . In addition, a strong decreasing trend with  $p_T$  is predicted in the  $2 < p_T < 12 \text{ GeV}/c$ , but not observed in the data. These discrepancies are also observed in the measurement of the  $p_T$ -differential  $D_s^+/D^0$  production-yield ratio in pp collisions at 5.02 TeV performed by the ALICE Collaboration [118], where the POWLANG predictions overestimate the  $D_s^+$ -meson production, with similar differences in the description of the  $p_T$ -dependence, while an accurate description of the  $D^+/D^0$  production-yield ratio is achieved.

# Chapter 8

## D<sub>s</sub><sup>+</sup> and D<sup>+</sup> mesons reconstruction in Pb–Pb collisions

The work presented in this Thesis so far has focused on the study of the D<sub>s</sub><sup>+</sup>/D<sup>+</sup> production-yield ratio in pp collisions at  $\sqrt{s} = 13.6$  TeV, which can provide information on the hadronisation of charm quarks. This serves also as a baseline for studies in pp collisions as a function of charged-particle multiplicity, to search for the effects of a possible production of small QGP droplets in small collision systems. Moreover, it also provides a precise reference for measurements in heavy-ion collisions, where an extended deconfined phase is expected to be formed. Measurements of D<sub>s</sub><sup>+</sup>-meson production in Pb–Pb collisions have been performed by the ALICE Collaboration at the centre-of-mass energies per nucleon-nucleon collision of  $\sqrt{s_{\text{NN}}} = 5.02$  TeV [181, 195] and  $\sqrt{s_{\text{NN}}} = 2.76$  TeV [196]. The measured  $R_{\text{AA}}$  of prompt D<sub>s</sub><sup>+</sup> mesons in central and semicentral Pb–Pb collisions is compatible within uncertainties with that of other non-strange D mesons (D<sup>0</sup>, D<sup>+</sup>, and D<sup>\*+</sup>) for  $p_{\text{T}} \gtrsim 10$  GeV/ $c$ , where the hadronisation process is expected to occur mainly via fragmentation. At lower  $p_{\text{T}}$ , where the coalescence mechanism is expected to play a more relevant role, the measured  $R_{\text{AA}}$  of prompt D<sub>s</sub><sup>+</sup> mesons is systematically higher than that of non-strange D mesons, although the results are compatible within one standard deviation. In addition, while the measured production-yield ratio of non-strange over non-strange D mesons (D<sup>+</sup>/D<sup>0</sup> and D<sup>\*+</sup>/D<sup>0</sup>) is found to be compatible in Pb–Pb and pp collisions, indicating no significant modification of their relative abundances as a function of  $p_{\text{T}}$  and centrality, the production-yield ratio of strange over non-strange D mesons (D<sub>s</sub><sup>+</sup>/D<sup>+</sup> and D<sub>s</sub><sup>+</sup>/D<sup>0</sup>) is found to be larger in central (semicentral) Pb–Pb collisions than in pp collisions, but the measurements in the two systems are compatible within about 2.3 (2.4) standard deviations. The large uncertainties of the measurements in Pb–Pb collisions available so far did not allow for drawing firm conclusions on the production mechanism of D<sub>s</sub><sup>+</sup> mesons in heavy-ion collisions. Additionally, the measurement is performed for  $p_{\text{T}} > 2$  GeV/ $c$ . About 70% of the produced D mesons have  $p_{\text{T}}$  below this threshold [109], and therefore the results cannot disentangle a possible modification of the  $p_{\text{T}}$  spectrum from an enhancement of the D<sub>s</sub><sup>+</sup> meson production in Pb–Pb collisions with respect to pp collisions.

Thanks to the upgrade of the ALICE detector during the LHC Long Shutdown 2,

the experiment significantly enhanced its readout rate in Pb–Pb collisions to more than 50 times that achieved during the LHC Run 2 data-taking period, and improved its spatial resolution by a factor of about 2, allowing for a better separation of displaced decay vertices [128]. These upgrades will allow for enhancing the precision of the measurements of rare probes, such as heavy-flavour hadrons, and for completely new insights into the properties of the QGP, such as the exclusive reconstruction of beauty-hadron decays. With the improvements in the detector capabilities, the strange over non-strange D-meson production-yield ratios can be measured with unprecedented precision in Pb–Pb collisions, providing more significant evidence, or the lack thereof, of the enhancement of  $D_s^+$ -meson production in the hadronisation of the QGP, due to the enhancement of strange-antistrange quark pairs and the hadronisation of charm quarks via recombination.

At the time of starting the work presented in this Thesis, the ALICE Collaboration was performing the first collisions of ultra-relativistic lead ions with the upgraded apparatus. Therefore, rapid feedback on the quality of the data was required. Since the upgrade of the ALICE detector mainly focused on the improvement of its capabilities for heavy-flavour measurements, particular attention was given to the performance of the reconstruction of D mesons. Hence, the first studies carried out for this Thesis were devoted to the reconstruction of D mesons in Pb–Pb collisions. In this Chapter, the analysis strategy and the results of the reconstruction of  $D_s^+$  and  $D^+$  mesons in Pb–Pb collisions at  $\sqrt{s_{\text{NN}}} = 5.36$  TeV are presented.

## 8.1 Event selection and data sample

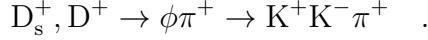
The analysis reported in this Chapter is performed on a dataset of Pb–Pb collisions at a centre-of-mass energy per nucleon pair of  $\sqrt{s_{\text{NN}}} = 5.36$  TeV, collected by the ALICE Collaboration during the 2023 data-taking period. Since the calibration of the tracking detectors was still ongoing at the time of the analysis, and because the reconstruction of Pb–Pb events is very resource-intensive, only a fraction of 20% of the total data sample collected during the 2023 data-taking period was made available and therefore analysed. Only once the quality of the reconstruction of Pb–Pb events was validated, the full dataset was processed.

The data sample is collected using a Minimum-Bias trigger (the **Se18** trigger), which selects events that satisfy the requirement of having a signal coincidence in the FT0-A and FT0-C detectors. The centrality of the collision is determined using the signal amplitude of the forward-rapidity FT0-C detector, which is proportional to the number of particles produced in the pseudorapidity interval  $-3.3 < \eta < -2.1$ . The centrality classes are defined as percentiles of the inelastic cross-section and are determined starting from the charged-particle multiplicity distribution obtained with the FT0-C detector (which is calibrated for each data-taking period). Collisions in the 0–10% centrality class are the most central, with the largest number of produced charged particles.

The analysed centrality classes are the 10–30%, 30–50%, and 60–90%, but no signal was visible in the last centrality class on the small data sample used for these checks.

## 8.2 $D_s^+$ and $D^+$ mesons reconstruction

$D_s^+$  and  $D^+$  mesons are reconstructed in the same decay channel exploited for the measurement of the  $D_s^+/D^+$  production-yield ratio in pp collisions presented in the previous Chapters:



In the most central Pb–Pb collisions, around 21500 charged particles are produced on average [197]. Therefore, the reconstruction of D mesons in ultra-relativistic heavy-ion collisions is more challenging than in pp collisions since a much larger combinatorial background from the combination of three uncorrelated tracks is produced. On the other hand, the higher number of tracks in Pb–Pb collisions allows for a more precise reconstruction of the primary vertex position. Thereby, a slight improvement in the separation of the signal from the background, by exploiting the displaced decay topology of the D mesons, is achieved as compared to pp collisions.

Furthermore, with the replacement of the TPC MultiWire Proportional Chambers readout chambers with Gas Electron Multipliers performed during the LHC Long Shutdown 2 upgrade, the amount of backflow of ions in the TPC has significantly increased as compared to the LHC Run 2 data-taking period. The residual ions in the TPC gas volume can distort the electric field, leading to a distortion of the drift paths of the electrons produced by the ionising particles that are intended to be detected. This effect is particularly relevant in the track reconstruction, as it worsens the spatial and momentum resolution and the track reconstruction efficiency. Despite a great effort in the calibration of the TPC, the backflow of ions is still not fully corrected for, and affects the track reconstruction performance. This leads to a worse quality in the reconstruction of  $D_s^+$  and  $D^+$  mesons, as the broadening of their signal peaks, due to the non-optimal momentum resolution causes a worsening of the signal-to-background ratio and may lead to an overlap of their invariant mass distributions given the small mass difference of about  $100 \text{ MeV}/c^2$ .

Lastly, at the time this work was carried out, Monte Carlo simulations of Pb–Pb collisions enriched with heavy-flavour hadrons were not available. Therefore, there was no possibility of employing Machine Learning (ML) techniques to select the D-meson candidates, as it is done for the analysis of the  $D_s^+/D^+$  production yield ratio in pp collisions described in the previous Chapters, since a pure data sample of signal candidates needed to train the ML model was not available.

Therefore, the selection of D-meson candidates is performed using rectangular selection criteria exploiting the displaced decay topology of heavy-flavour hadrons and their decay kinematics. Additionally, particle identification techniques of the decay products are employed to reduce the combinatorial background and select  $D_s^+$  and  $D^+$  candidates. Selections based on the same variables used in the pp analysis, such as the decay length, pointing angle, and kinematical variables based on the production of a  $\phi$  meson in the considered decay channel of the  $D_s^+$  and  $D^+$  mesons ( $|\Delta M(KK)|$ ,  $\cos(\theta'(K))$ ) described in Chapter 4 are applied. Since the number of particles produced in a heavy-ion collision varies drastically with the centrality of the collision, the selection criteria employed for the selection of D-meson signals were optimised independently for each centrality class.

For each centrality interval, the selection criteria are optimised by testing different sets of cut values and maximising the statistical significance of the extracted  $D_s^+$ -meson signal, for each considered  $p_T$  interval. This strategy is used as no physics observables are meant to be extracted from the analysis, and the main goal is to validate the reconstruction framework and provide feedback on its performance. In fact, the maximisation of the statistical significance of the signal on the same sample used for the yield extraction should not be pursued as a strategy for the selection of D-meson candidates in physics analyses, as statistical fluctuations can lead to a bias in the extracted yields.

Table 8.1: Variation range and number of steps for the variables used in the optimisation of the selection criteria for the selection of  $D_s^+$ - and  $D^+$ -meson candidates in Pb–Pb collisions at  $\sqrt{s_{\text{NN}}} = 5.36$  TeV.

Variable	N. steps	$p_T$ (GeV/ $c$ )				
		2–4	4–6	6–8	8–12	12–24
$L(\mu\text{m}) >$	3	min	200	200	200	200
		max	600	800	800	800
$\cos \theta_p >$	4	min	0.9900	0.9980	0.9980	0.9800
		max	0.9990	0.9995	0.9995	0.9950
$\cos \theta_p^{xy} >$	4	min	0.9900	0.9970	0.9980	0.9800
		max	0.9990	0.9985	0.9995	0.9950
$ \Delta M(\text{KK}) (\text{MeV}/c^2) <$	3	min	5	5	5	5
		max	9	9	9	9
$ \cos^3(\theta'(\text{K}))  >$	3	min	0.1	0.1	0.1	0.1
		max	0.3	0.3	0.3	0.3
$\chi^2_{\text{PCA}} <$	4	min	3	3	3	3
		max	12	12	12	12

The configuration used for the optimisation of the set of selection criteria employed for the selection of  $D_s^+$  and  $D^+$  meson candidates is reported in Table 8.1, where the variables used, the considered variation range, and the number of intermediate steps are reported. A total of 1728 selection criteria were considered for each  $p_T$  interval of the different centrality classes. In addition to those reported in Table 8.1, a few additional variables are used to further improve the selection of  $D_s^+$  and  $D^+$  mesons, but their thresholds are not optimised. The normalised decay length in the transverse plane is required to be larger than 2, and the PID information on the three daughter tracks from both the TPC and TOF detectors is exploited by requiring that the  $|n\sigma|$  with respect to the pion (kaon) expected signal is at most equal to 4. As for the analysis performed in pp collisions, the information from the TOF detector was used when available, but not required.

Figure 8.1 illustrates the results for the optimisation of the  $D_s^+$  meson statistical significance in the 10–30% centrality class for the 2–4 GeV/ $c$   $p_T$  interval, by reporting the statistical significance as a function of the set of applied selections identified by an index. The large steps in the statistical significance dividing the graph into three parts are due to the different thresholds applied to the candidate decay length.

Smaller periodic modulations of the  $D_s^+$  statistical significance are also observed and are due to the different thresholds applied to the other considered variables.

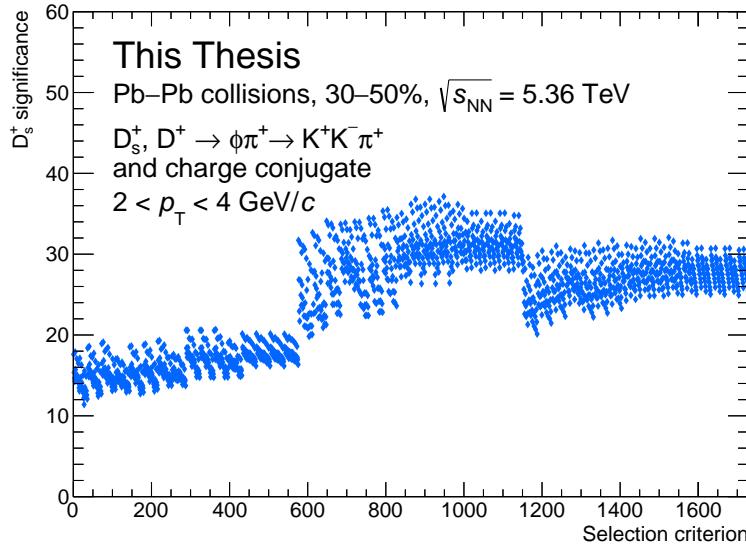


Figure 8.1:  $D_s^+$ -meson signal statistical significance as a function of the selection criteria for the  $2 < p_T < 4$  GeV/ $c$  interval in the 30–50% centrality class.

The optimal selection criteria for the selection of  $D_s^+$  meson candidates in the 10–30% and 30–50% centrality classes are reported in the Tables 8.2 and 8.3, respectively. The optimal selection criteria for the 60–90% centrality class are not reported, as no signal was extracted in this centrality class.

Table 8.2: Optimal selection criteria for the selection of  $D_s^+$  meson candidates in the 10–30% centrality class.

Variable	$p_T$ interval (GeV/ $c$ )				
	2–4	4–6	6–8	8–12	12–24
$L(\mu\text{m}) >$	600	200	500	200	200
$\cos \theta_p >$	0.9900	0.9980	0.9985	0.9800	0.9800
$\cos \theta_p^{xy} >$	0.9990	0.9980	0.9980	0.9900	0.9950
$ \Delta M(\text{KK}) (\text{MeV}/c^2) <$	7	9	5	7	7
$ \cos^3(\theta'(\text{K}))  >$	0.3	0.2	0.2	0.2	0.3
$\chi^2_{\text{PCA}} <$	3.0	6.0	3.0	3.0	9.0

### 8.3 Signal extraction

Once the selection criteria are optimised for each  $p_T$  interval and centrality class, the signal is extracted by fitting the invariant mass distribution of the selected candidates. The fit function is composed of an exponential function for describing the

Table 8.3: Optimal selection criteria for the selection of  $D_s^+$  meson candidates in the 30–50% centrality class.

Variable	$p_T$ interval ( $\text{GeV}/c$ )				
	2–4	4–6	6–8	8–12	12–24
$L(\mu\text{m}) >$	400	200	200	200	200
$\cos \theta_p >$	0.996	0.998	0.998	0.980	0.995
$\cos \theta_p^{xy} >$	0.996	0.997	0.998	0.985	0.980
$ \Delta M(\text{KK}) (\text{MeV}/c^2) <$	7	9	9	5	5
$ \cos^3(\theta'(\text{K}))  >$	0.1	0.1	0.1	0.1	0.2
$\chi^2_{\text{PCA}} <$	3.0	3.0	12.0	12.0	9.0

background shape and two Gaussian functions for the  $D_s^+$  and  $D^+$  signal peaks. The correlated background from  $D^+ \rightarrow \pi^+ K^- \pi^+$  introduced in Chapter 5 is not included, as the absence of MC simulations precludes the estimation of its shape. The fit of the invariant mass distributions of the candidates passing the selection criteria described above is performed in the invariant mass range of  $1.7 < M < 2.1 \text{ GeV}/c^2$ . A clear signal peak is observed in both the 10–30% and 30–50% centrality classes, with statistical significances ranging from 14 to 35 for  $D_s^+$  mesons and from 4 to 14 for  $D^+$  mesons depending on  $p_T$  in the 10–30% centrality interval, and from 9 to 37 for  $D_s^+$  mesons and from 4 to 21 for  $D^+$  mesons in the 30–50% centrality class. Figure 8.2 shows the normalised invariant mass distribution of D-meson candidates passing the applied selections in the  $2 < p_T < 4 \text{ GeV}/c$  (top-left panel) and  $6 < p_T < 8 \text{ GeV}/c$  (bottom-left panel) intervals of the 10–30% centrality class and in the  $4 < p_T < 6 \text{ GeV}/c$  (top-right panel) and  $8 < p_T < 12 \text{ GeV}/c$  (bottom-right panel) intervals of the 30–50% centrality class.

As mentioned above, one of the consequences of the backflow of ions in the TPC is the distortion of the drift paths of the ionised electrons, which leads to a worsening of the track momentum resolution. An increase of the peak widths with increasing  $p_T$  is already visible in Fig. 8.2: while the  $D_s^+$  and  $D^+$  peaks are well separated in both the  $2 < p_T < 4 \text{ GeV}/c$  and  $4 < p_T < 6 \text{ GeV}/c$  intervals, at the higher  $p_T$  intervals of  $6 < p_T < 8 \text{ GeV}/c$  and  $8 < p_T < 12 \text{ GeV}/c$  the peak widths are larger, and the two peaks are partially overlapping. Nevertheless, a worsening in the resolution with increasing  $p_T$  is expected due to the reduced track curvature in the presence of the magnetic field. Therefore, to test the validity of the distortion corrections in the reconstruction framework it is useful to study the resolution of the  $D_s^+$  and  $D^+$  mesons peaks, and compare it with that obtained during the LHC Run 2 data-taking period, where fewer distortions were present thanks to the reduced interaction rate and the active ion gates. Similar studies are being performed for the  $K_s^0 \rightarrow \pi^+ \pi^-$  peaks. However, the measurement of D meson widths allows the test of a different  $p_T$  range of the tracks and the effects on the charged kaon reconstruction, which is not possible with the  $K_s^0$  mesons.

The resolution of the  $D_s^+$  and  $D^+$  mesons peaks is extracted from the fits of the invariant mass distribution of the selected candidates as the standard deviation of the Gaussian function used to describe the D-meson signals. In Figure 8.3, the

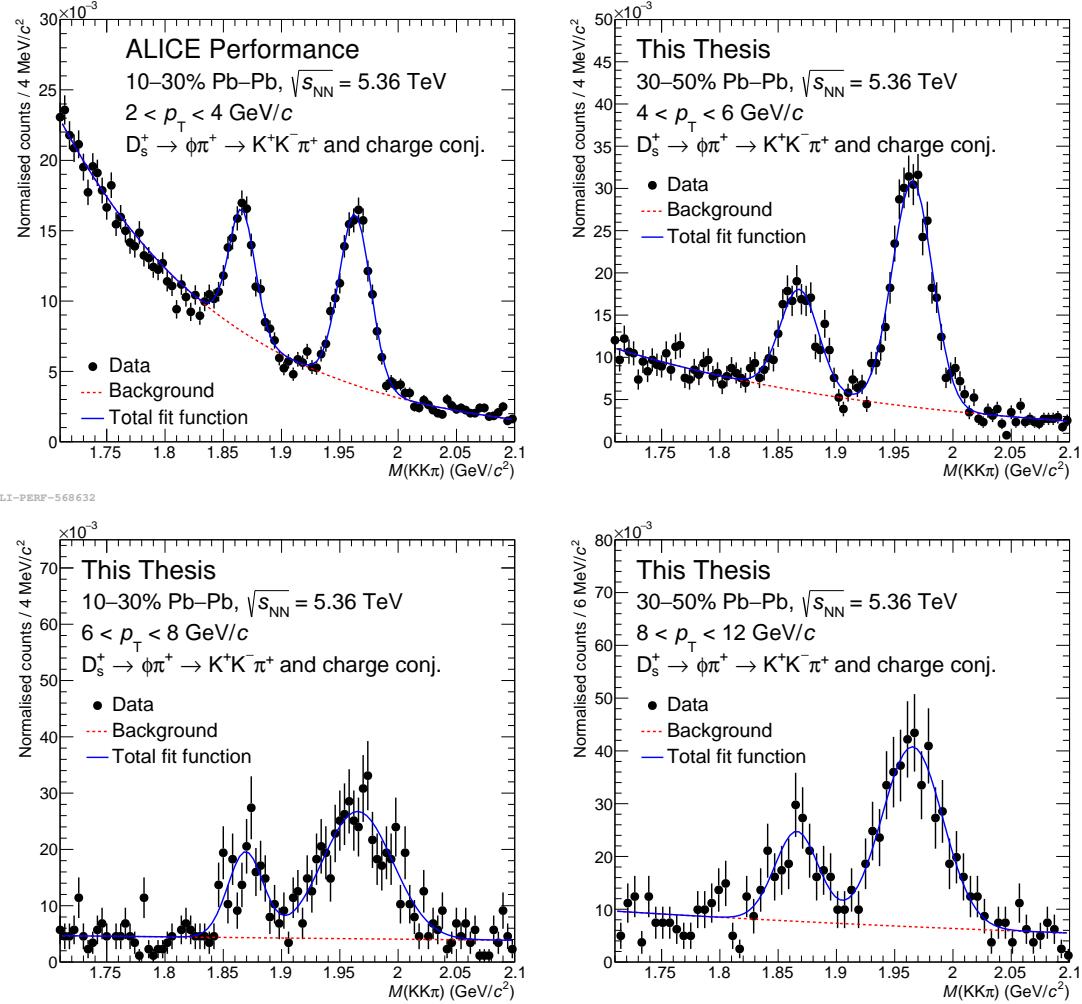


Figure 8.2: Normalised invariant mass distribution of the selected D-meson candidates in the  $2 < p_{\text{T}} < 4 \text{ GeV}/c$  (top-left panel) and  $6 < p_{\text{T}} < 8 \text{ GeV}/c$  (bottom-left panel) intervals of the 10–30% centrality class and in the  $4 < p_{\text{T}} < 6 \text{ GeV}/c$  (top-right panel) and  $8 < p_{\text{T}} < 12 \text{ GeV}/c$  (bottom-right panel) intervals of the 30–50% centrality class.

peak width of the  $D_s^+$  meson signal is shown for the analysis performed in this Thesis using the 30–50% centrality class of Pb–Pb collisions recorded in the LHC Run 3 data-taking period (blue markers) as a function of  $p_T$ . A clear increasing trend of the peak width with  $p_T$  is observed, and is attributed to the worsening of the momentum resolution with increasing  $p_T$  due to the smaller curvature of high-momentum tracks. The peak widths of the  $D_s^+$ -meson signals extracted from the Pb–Pb collisions data collected during the LHC Run 2 data-taking period at  $\sqrt{s_{\text{NN}}} = 5.02$  TeV in the 30–50% centrality class [181] (shown as green markers in Fig. 8.3) are smaller than those found for Run 3 samples. This is due to the distortions of the electric field in the TPC, which were smaller in Run 2 because the TPC ion backflow was significantly suppressed using active gating grids. Lastly, the peak widths of the  $D_s^+$ -meson signals extracted from pp collisions data recorded during the LHC Run 3 data-taking period at  $\sqrt{s} = 13.6$  TeV in this Thesis (as described in Chapter 5) are reported as a function of  $p_T$  with red markers. These results show a magnitude of the width and a trend with  $p_T$  which are consistent with what is observed in Pb–Pb collisions at  $\sqrt{s_{\text{NN}}} = 5.36$  TeV, confirming that the observed worsening of the momentum resolution is due to the backflow of ions in the TPC, which affect similarly the different samples of pp and Pb–Pb collisions collected during the LHC Run 3 data-taking period. The widths extracted in pp collisions are systematically smaller than those from Pb–Pb collisions, albeit being compatible within uncertainties. This effect may be due to a newer reconstruction performed in the pp collisions data, which may have improved the momentum resolution. Studies are currently ongoing to further improve the calibration of the reconstruction and improve the momentum resolution in both pp and Pb–Pb collisions, allowing for more precise measurements of the production of D mesons.

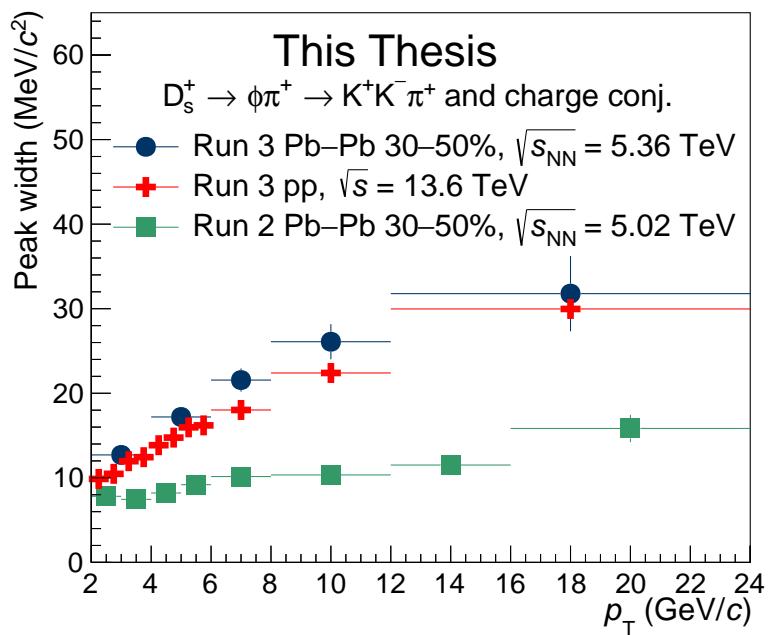


Figure 8.3: Width of the  $D_s^+$ -meson signal peak as a function of  $p_T$  measured in Pb–Pb collisions at  $\sqrt{s_{NN}} = 5.36$  TeV in the 30–50% centrality class and pp collisions at  $\sqrt{s} = 13.6$  TeV recorded during the LHC Run 3 data-taking period (blue and red markers, respectively) and in the 30–50% centrality class of Pb–Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV recorded during the LHC Run 2 data-taking period (green markers).

# Chapter 9

## Summary, conclusions, and perspectives

This Thesis presented the measurement of the  $p_T$ -differential  $D_s^+/D^+$  production-yield ratio in proton-proton collisions at  $\sqrt{s} = 13.6$  TeV, performed with the upgraded ALICE detector and the data collected during the LHC Run 3 data-taking period. This allowed the study of the charm-quark hadronisation (i.e., the transition from a colour-charged quark into colour neutral hadrons) via strange over non-strange charm meson production-yield ratios.

The analysis was performed via the full reconstruction of displaced decay topologies through the same hadronic decay channel  $D_s^+, D^+ \rightarrow \phi\pi^+ \rightarrow K^+K^-\pi^+$ . Multi-class Machine Learning models using the Boosted Decision Tree algorithm provided by the XGBoost library were employed to enhance the selection efficiency of the signal candidates and to reduce the combinatorial background. Additionally, they were used to increase the relative contribution of prompt  $D_s^+$  and  $D^+$  mesons (i.e., those directly produced in the hadronisation of a charm quark or through the strong decay of a directly produced excited charm-hadron or charmonium state) in the selected sample. The measurement was performed in 14 transverse-momentum intervals in the range  $0.5 < p_T < 24$  GeV/ $c$ , and extended the  $p_T$  coverage at low  $p_T$  with respect to previous measurements performed by the ALICE Collaboration at  $\sqrt{s} = 5.02, 7,$  and  $13$  TeV, reported in Refs. [62, 148, 180], respectively. Additionally, thanks to the larger data sample collected during the LHC Run 3 data-taking period, and the reconstruction of both D-meson species in the same decay channel, the results presented in this Thesis significantly reduced both the statistical and systematic uncertainties of the measurement, and improved its granularity with respect to the measurements performed by both the ALICE and LHCb [190] Collaborations at mid and forward rapidities, respectively.

The measurements performed at midrapidity ( $|y| < 0.5$ ) at the centre-of-mass energies of  $\sqrt{s} = 5.02, 7, 13$  and  $13.6$  TeV by the ALICE Collaboration indicate no significant energy-dependence of the  $D_s^+/D^+$  production-yield ratio. Furthermore, the comparison of the results presented in this Thesis, performed at midrapidity ( $|y| < 0.5$ ) at  $\sqrt{s} = 13.6$  TeV with the ALICE experimental apparatus, with those obtained by the LHCb Collaboration in the  $2.0 < y < 4.5$  interval at forward rapidity at  $\sqrt{s} = 13$  TeV, shows a good agreement within the uncertainties. This

indicates no significant dependence of the  $D_s^+/D^+$  ratio on the rapidity within the range covered by the ALICE and LHCb measurements.

This measurement provides state-of-the-art results for the understanding of the hadronisation mechanisms of charm quarks in high-energy hadronic collisions. As described in Chapter 2, the hadronisation mechanism is expected to be modified in the presence of a deconfined medium, the Quark-Gluon Plasma (QGP), which is formed in high-energy nuclear collisions. In a thermalised deconfined medium, charmed hadrons can be produced through coalescence of charm quarks, which are produced before the QGP is formed, with light quarks from the medium. Additionally, in the presence of a QGP, the production of strange quarks is expected to be enhanced, as the high temperatures reached in the medium allow for the thermal production of strange-antistrange quark pairs. The measurement of the  $D_s^+/D^+$  production-yield ratio is a powerful tool to investigate the hadronisation mechanisms of charm quarks, and is a sensitive probe to the phenomenon of strangeness enhancement.

Several measurements confirmed the formation of the QGP in high-energy nuclear collisions, which are summarised in Refs. [16, 198, 23]. The production of an extended QGP phase is not expected in proton-proton collisions. However, recent measurements performed at the LHC [199, 200, 201] provide evidence for the presence of some effects, such as collective behaviours and strangeness enhancement, in small collision systems, as proton-proton and proton-lead collisions, which are usually associated with the formation of a QGP. The ALICE Collaboration reported the observation of a smooth increase in the strange- and multi-strange-hadron production with the charged-particle multiplicity in proton-proton collisions at  $\sqrt{s} = 7$  TeV [57], reaching, for the highest multiplicity classes, values compatible with those measured in the most central lead-lead collisions.

Previous measurements of the multiplicity-dependence of the strange over non-strange  $D_s^+/D^0$  production-yield ratio in different multiplicity intervals performed at midrapidity ( $|y| < 0.5$ ) by the ALICE Collaboration [202] did not provide a clear indication of the strangeness enhancement in proton-proton collisions at  $\sqrt{s} = 13$  TeV. The results are illustrated in Fig. 9.1, where the measured  $D_s^+/D^0$  production-yield ratio is shown as a function of  $p_T$  for two different multiplicity classes. The results show a slight increase of the  $D_s^+/D^0$  production-yield ratio with the charged-particle multiplicity, although the two  $p_T$ -differential measurements are compatible within their uncertainties. The methodologies and results presented in this Thesis provide a solid foundation for future studies of the strangeness enhancement in the heavy-flavour sector. The strategy of using the  $D^+$  meson as a reference for the non-strange production, as well as the reconstruction of both  $D_s^+$  and  $D^+$  mesons in the same hadronic decay channel allows for a significant reduction of the systematic uncertainties of the measurement. The measurement of the multiplicity dependence of the  $D_s^+/D^+$  production-yield ratio along with the larger data samples that will be available from the LHC Run 3 data-taking period may provide more precise results than those achieved through the usage of the  $D^0$  meson as a reference for non-strange D meson production, and, thereby, provide a more sensitive probe to the phenomenon of strangeness enhancement in proton-proton collisions. Additionally, the much larger data sample collected during the LHC Run 3 data-taking period

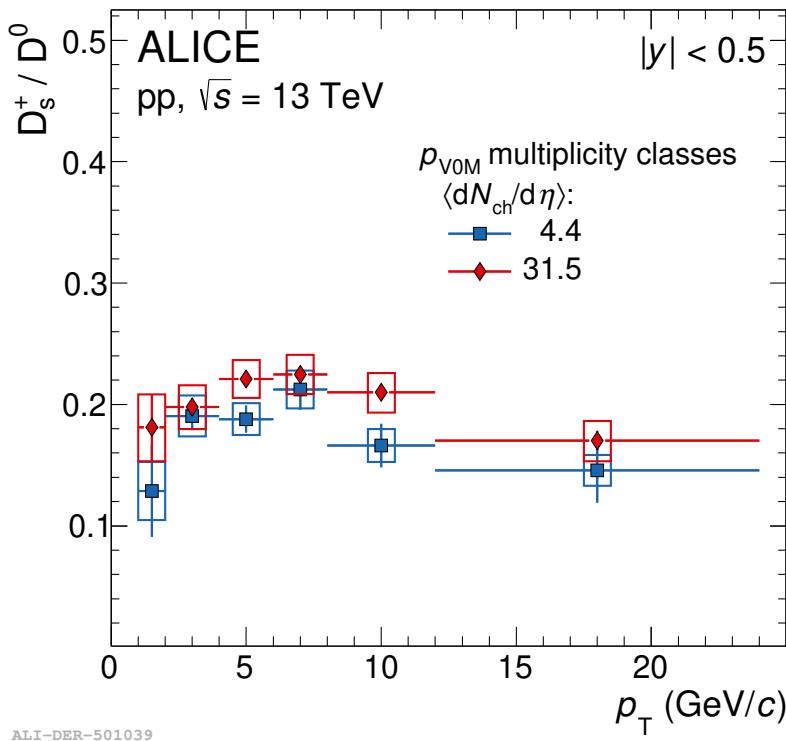


Figure 9.1: Strange over non-strange  $D_s^+ / D^0$  production-yield ratio as a function of  $p_T$  for two different multiplicity classes measured at midrapidity ( $|y| < 0.5$ ) in proton-proton collisions at  $\sqrt{s} = 13$  TeV by the ALICE Collaboration [202]. Figure taken from the ALICE figure repository [60].

will allow for the extension of the measurement to lower  $p_T$  values and perform the measurement in finer intervals of  $p_T$  and multiplicity.

The multiplicity phase space can be further explored by measuring the strange over non-strange D meson production-yield ratio in Pb–Pb collisions, where much higher charged-particle multiplicities are reached. Previous measurements performed by the ALICE experiment at a centre-of-mass energy per nucleon pair of  $\sqrt{s_{NN}} = 5.02$  TeV [181] did not allow to draw firm conclusions on the possible enhancement of strange-hadron production in the charm sector.

The results from the measurement of the  $D_s^+ / D^0$  production-yield ratios in central (0–10%) Pb–Pb collisions compared to pp collisions are shown in Fig. 9.2. While the measured ratio is consistent within the uncertainties between the two collision systems for  $p_T < 4$  GeV/c and  $p_T > 8$  GeV/c, the average values of the measured  $D_s^+ / D^0$  production-yield ratio in the  $2 < p_T < 8$  GeV/c interval in Pb–Pb collisions are larger than those in pp collisions by about  $2.3\sigma$  of the combined statistical and systematic uncertainties.

The results are also compared to different theoretical predictions. The Catania model [203, 204], already introduced in Chapter 2 for pp collisions, assumes that a colour-deconfined state of matter is formed in both pp and Pb–Pb collisions and implements the heavy-quark transport via the Boltzmann equation. The hadronisation can occur via instantaneous coalescence, in addition to the fragmentation. In the

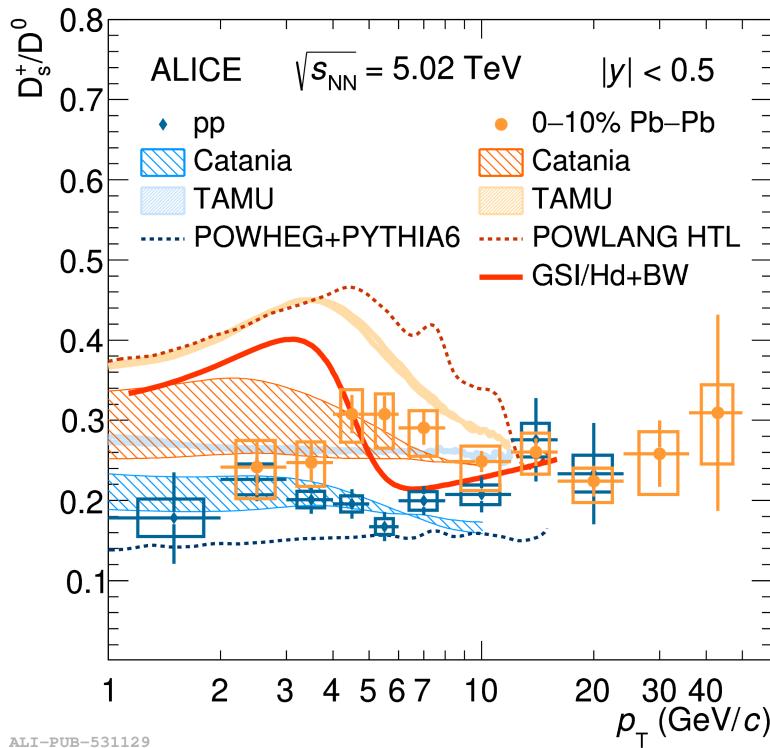


Figure 9.2:  $D_s^+/D_0^0$  production-yield ratio as a function of  $p_T$  in pp collisions and in the 10% most central Pb–Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV compared to different model calculations. Figure taken from Ref. [16].

TAMU [205] model, a combined recombination and fragmentation approach is implemented. The former is realized in Pb–Pb via the Resonance Recombination Model (RRM) [206] where the recombination probability for the two-body case is controlled by resonance amplitudes and is expressed as a relativistic Breit-Wigner cross-section. In pp collisions, the abundances of the different charm-hadron species are instead determined with a statistical hadronisation approach [121]. The charm-quark transport in a hydrodynamically expanding medium is described by the Langevin equation. POWHEG [95] NLO pQCD calculations for the charm-quark production, matched with PYTHIA 6 to generate the parton shower are reported for pp collisions. Predictions from the GSI-Heidelberg Statistical Hadronisation Model [207] and POWLANG [111] with transport coefficients calculated with weak-coupling calculations [193] (Hard-Thermal-Loop, HTL) are shown for Pb–Pb collisions. In the former, the  $p_T$  spectra of charm hadrons are modelled with a core-corona approach. In the low- $p_T$  region, the charm production is dominated by the core contribution, described with a Blast Wave function, which is used to describe the velocity profile of the collectively expanding system, as introduced in Chapter 1. The corona contribution is parametrised from measurements in pp collisions and is relevant at high  $p_T$ . The  $p_T$ -spectra modification due to resonance decays is computed using the FastReso package [208]. In the latter, a Langevin-based transport of heavy quarks in the QGP is followed by in-medium hadronisation. At the hadronisation stage charm quarks are recombined with light thermal quark or di-quark states from the

medium into colour-singlet clusters.

The Catania model reproduces within the uncertainties the measured  $D_s^+/D^0$  production-yield ratios both in pp and Pb–Pb collisions. The TAMU model overestimates the measured  $D_s^+/D^0$  ratio by a similar amount in the two colliding systems. A similar magnitude and  $p_T$  shape is predicted by the POWLANG model, while POWHEG calculations with PYTHIA 6 slightly underestimate the  $D_s^+/D^0$  production-yield ratio in pp collisions. The GSI-Heidelberg SHMc model provides a similar  $p_T$  shape for the  $D_s^+/D^0$  production-yield ratio as that provided by the TAMU and POWLANG models. The results do not allow for drawing firm conclusions on the phenomenon of strangeness enhancement in Pb–Pb collisions given the large uncertainties. However, they provide indications about the role of the charm-quark hadronisation via coalescence in the QGP.

The results discussed above do not allow for drawing firm conclusions on the phenomenon of strangeness enhancement in Pb–Pb collisions given the large uncertainties. The increase in the  $D_s^+/D^0$  production-yield ratio in the  $2 < p_T < 8 \text{ GeV}/c$  interval is compatible with the expectations from the modification of the charm-quark hadronisation in the presence of a deconfined medium and the phenomenon of strangeness enhancement. However, as also discussed in Chapter 8, a measurement performed in the  $p_T > 2 \text{ GeV}/c$  does not allow the disentanglement of an enhancement in the  $D_s^+$  meson production from a difference in the momentum spectra in the two collision systems. Furthermore, the collective radial expansion of the medium may play a role in the increase of the  $D_s^+/D^0$  production-yield ratio in Pb–Pb collisions because of the different masses of the  $D_s^+$  and  $D^0$  mesons. Additionally, the studied centrality classes do not provide a complete picture of a possible trend in the production of strange hadrons as a function of the centrality of the collision (which, in turn, provides information on the partonic densities reached in the formed medium). A more comprehensive study would require to also explore both the intermediate centrality (10–30%) and the most peripheral collisions (50–100%), which could not be studied with the LHC Run 2 samples because of the limited number of recorded events.

The measurement of the double ratio of  $D_s^+/D^+$  production-yield ratios in Pb–Pb and pp collisions will provide a clearer insight into the charm-quark hadronisation in a strangeness rich medium. This measurement would doubly benefit from the reduction of the systematic uncertainties from the reconstruction of both D meson species in the same decay channel, as the improvement would affect both the numerator (results in Pb–Pb collisions) and the denominator (results in pp collisions) of the double ratio. Additionally, the statistical uncertainties of the measurement will be significantly reduced thanks to the improved spatial resolution of the upgraded Inner Tracking System (ITS 2) and the larger data samples collected during the LHC Run 3. This will allow the extension of the measurement to lower  $p_T$  values, where the effects of the hadronisation via recombination are expected to be more pronounced, and to explore a wide multiplicity interval from pp collisions up to the most central Pb–Pb collisions. With these perspectives, the ALICE Collaboration will provide a state-of-the-art measurement of the strangeness enhancement in the heavy-flavour sector, with a comprehensive study of the hadronisation mechanisms of charm quarks in high-energy nuclear collisions.

# **Appendix A**

## **Figures for the systematic uncertainty on the raw-yield extraction**

In this Appendix, the results of the multi-trial approach employed for estimating the systematic uncertainties related to the raw-yield extraction are reported. The description of the used methodologies and the content of the Figures can be found in Chapter 5.

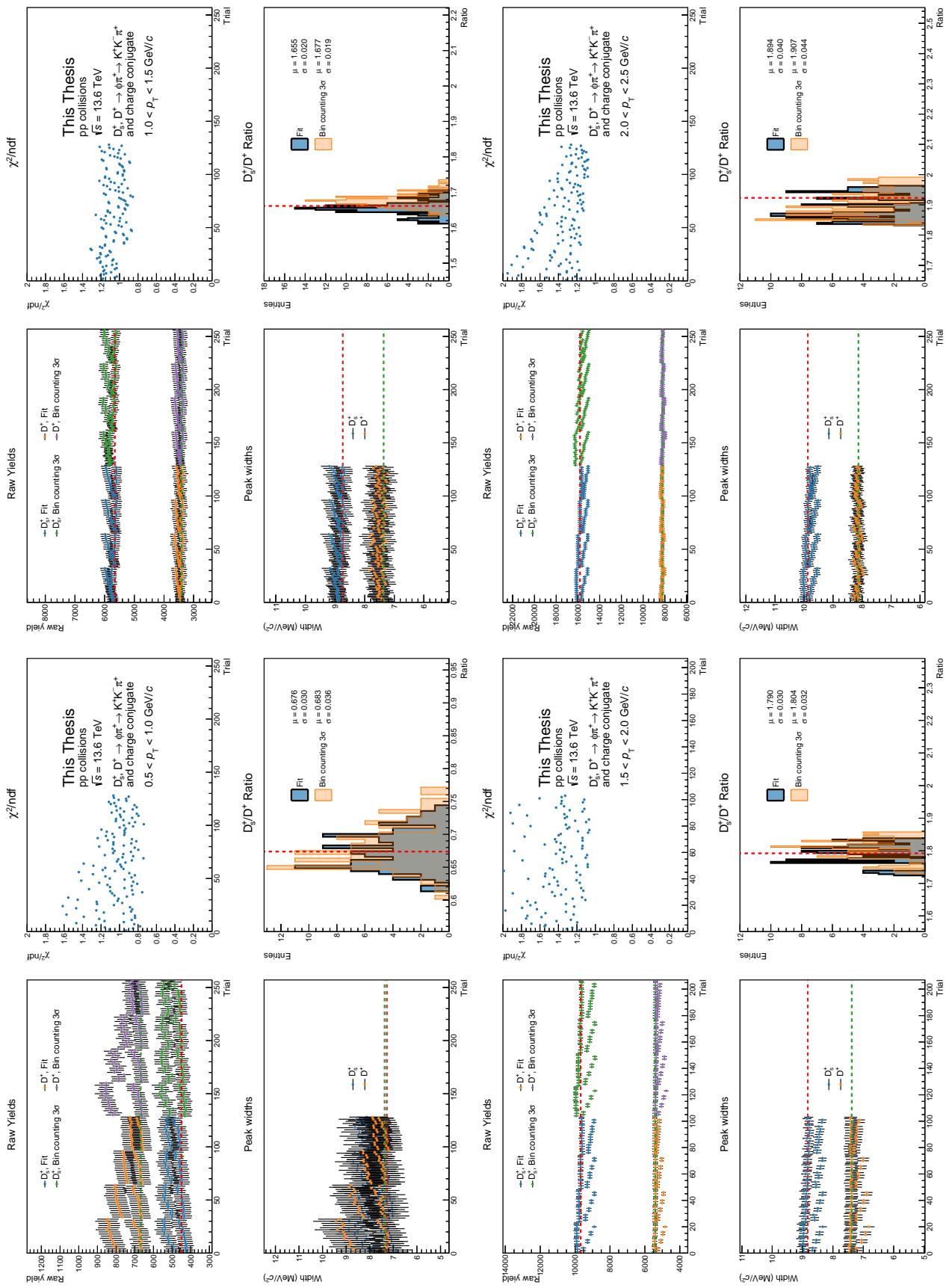


Figure A.1: Systematic uncertainties on the raw-yield extraction for the  $0.5 < p_T < 1$  GeV/c (top left),  $1 < p_T < 1.5$  GeV/c (top right),  $1.5 < p_T < 2$  GeV/c (bottom left) and  $2 < p_T < 2.5$  GeV/c (bottom right) intervals.

Appendix A – Figures for the systematic uncertainty on the raw yield extraction

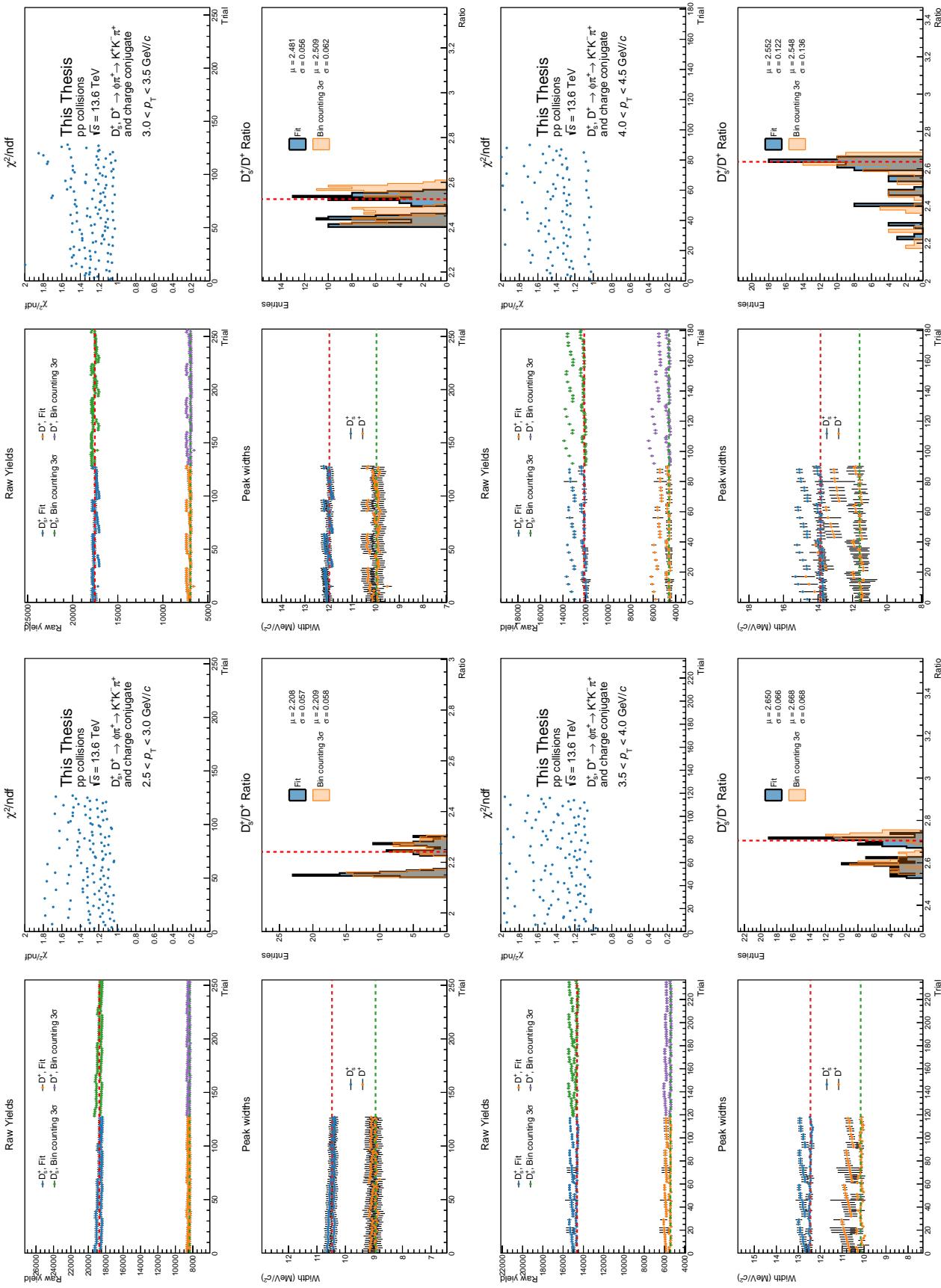


Figure A.2: Systematic uncertainties on the raw yield extraction for the  $2.5 < p_T < 3 \text{ GeV}/c$  (top left),  $3 < p_T < 3.5 \text{ GeV}/c$  (top right),  $3.5 < p_T < 4 \text{ GeV}/c$  (bottom left) and  $4 < p_T < 4.5 \text{ GeV}/c$  (bottom right) intervals.

Appendix A – Figures for the systematic uncertainty on the raw yield extraction

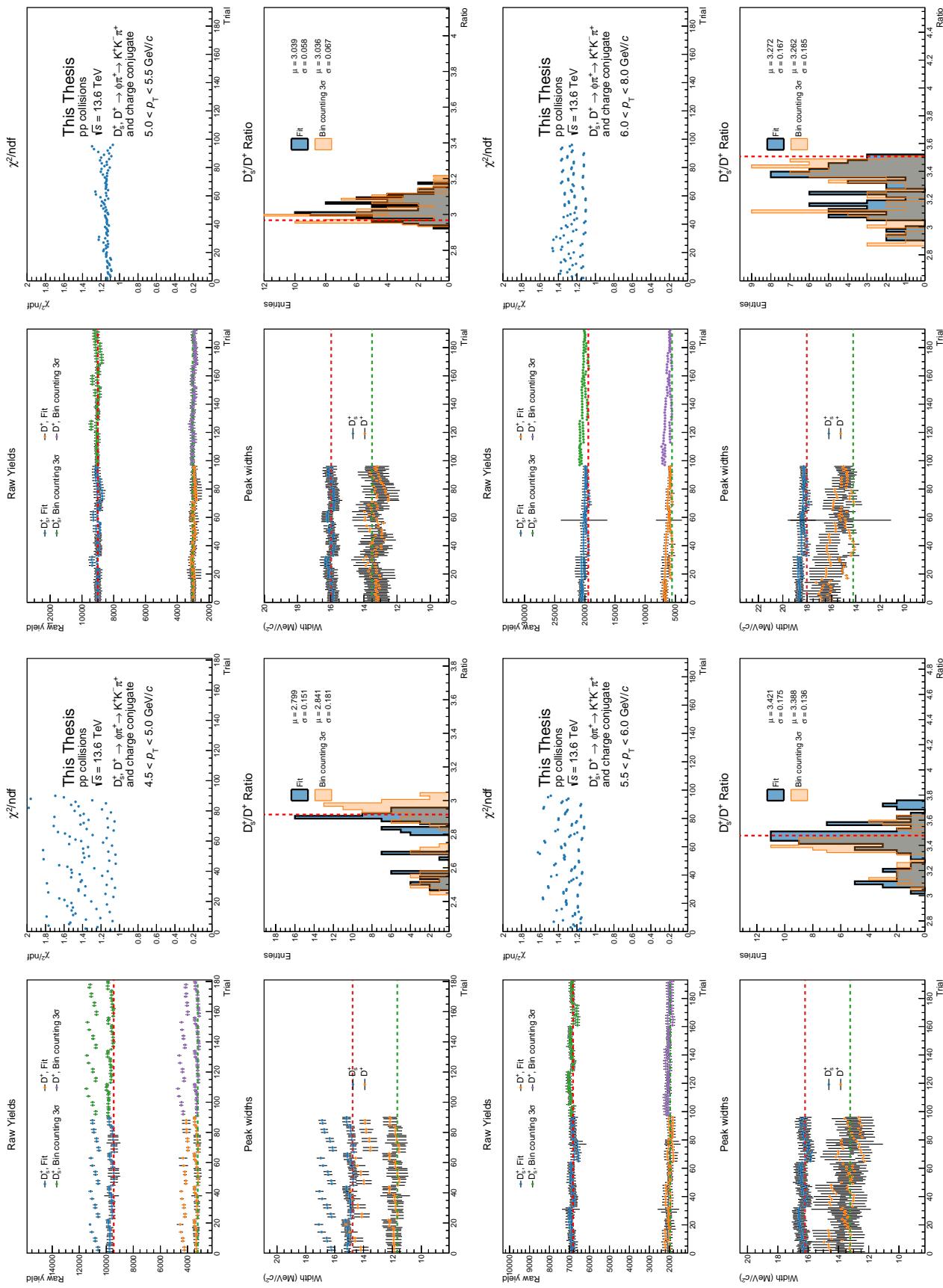


Figure A.3: Systematic uncertainties on the raw-yield extraction for the  $4.5 < p_T < 5 \text{ GeV}/c$  (top left),  $5 < p_T < 5.5 \text{ GeV}/c$  (top right),  $5.5 < p_T < 6 \text{ GeV}/c$  (bottom left) and  $6 < p_T < 8 \text{ GeV}/c$  (bottom right) intervals.

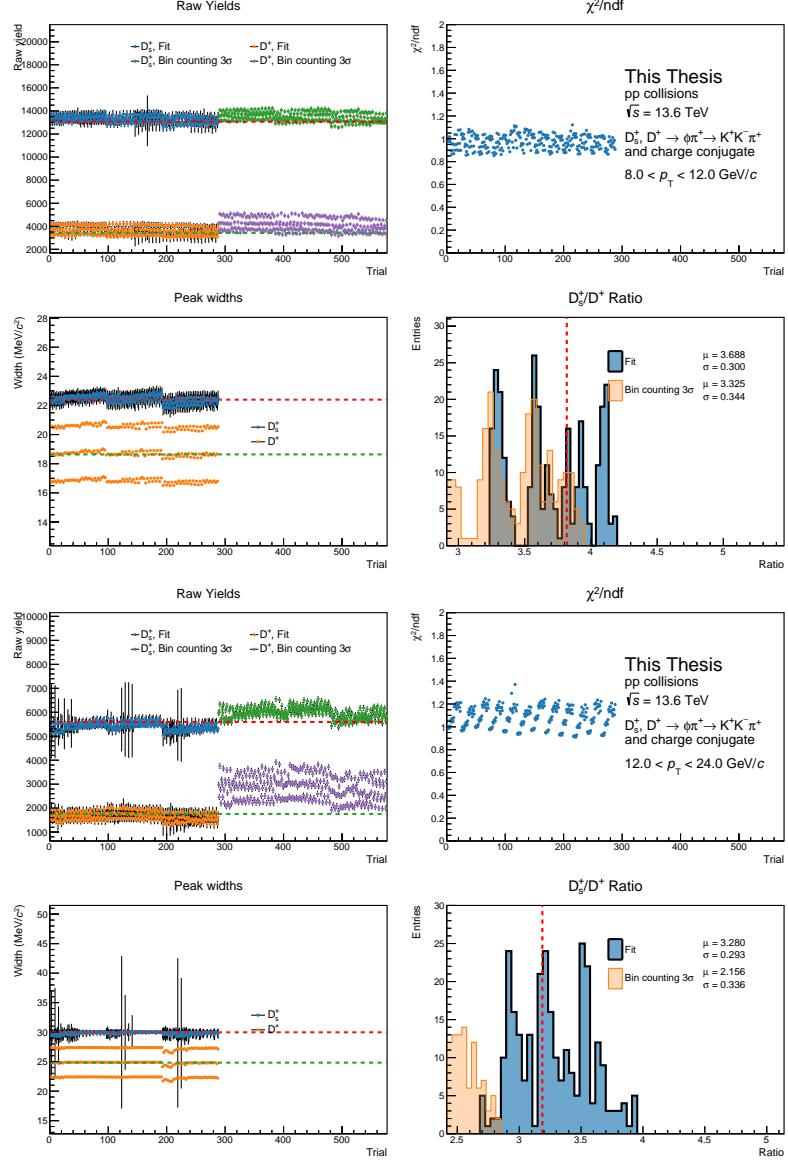


Figure A.4: Systematic uncertainties on the raw-yield extraction for the  $8 < p_T < 12 \text{ GeV}/c$  (left) and  $12 < p_T < 24 \text{ GeV}/c$  (right) intervals.

## Appendix B

### Figures for the data-driven $D_s^+$ and $D^+$ prompt fraction evaluation

In this Appendix, the results of the data-driven approach employed for estimating the  $f_{\text{prompt}}^{D_s^+}$  and  $f_{\text{prompt}}^{D^+}$  correction factors are reported. The description of the used methodologies and the content of the Figures can be found in Chapter 6.

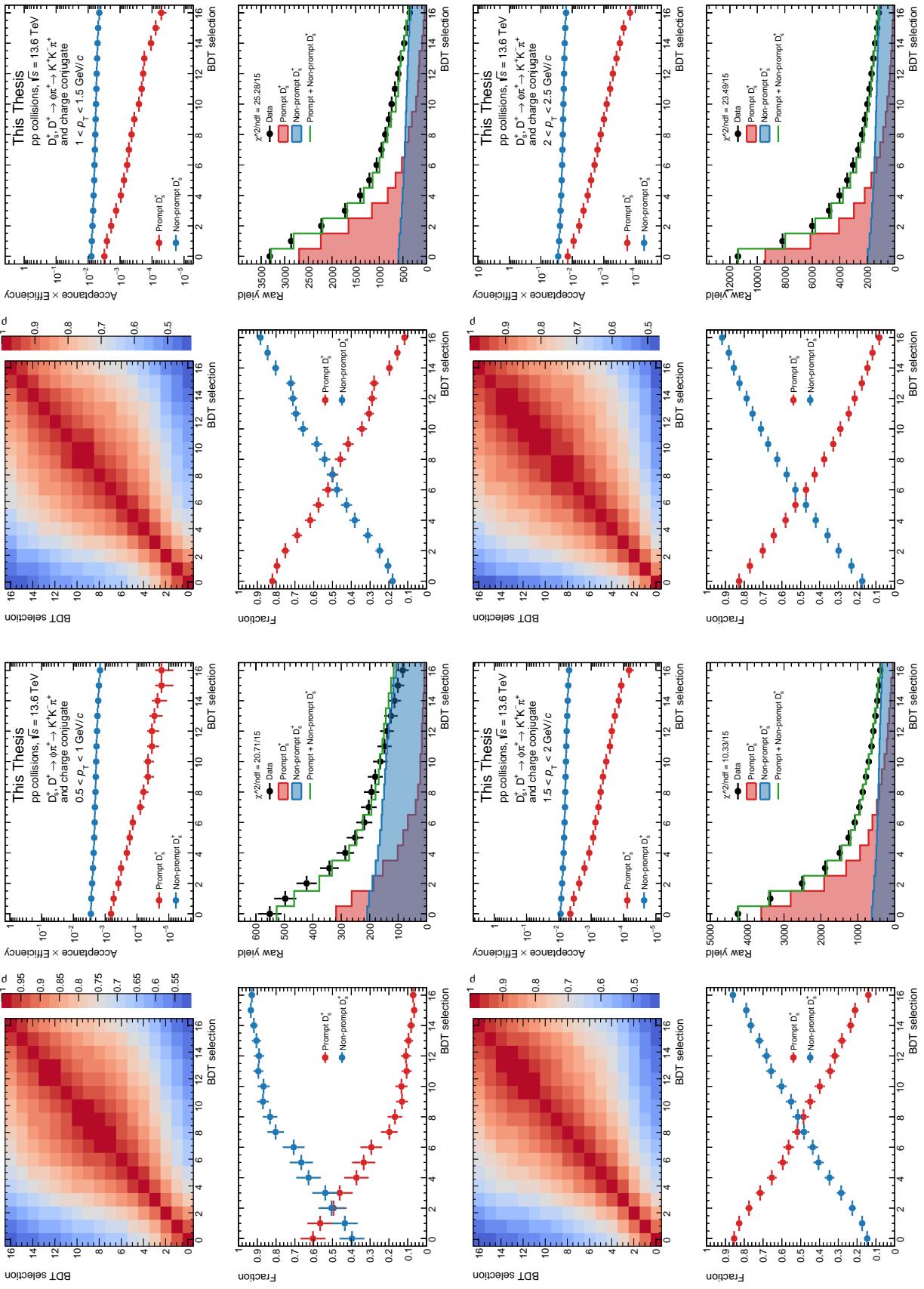


Figure B.1: Results for the evaluation of the  $f^{D_s^+}$  correction factor in the  $0.5 < p_T < 1.0 \text{ GeV}/c$  (top left),  $1.0 < p_T < 1.5 \text{ GeV}/c$  (top right),  $1.5 < p_T < 2.0 \text{ GeV}/c$  (bottom left) and  $2.0 < p_T < 2.5 \text{ GeV}/c$  (bottom right) intervals.

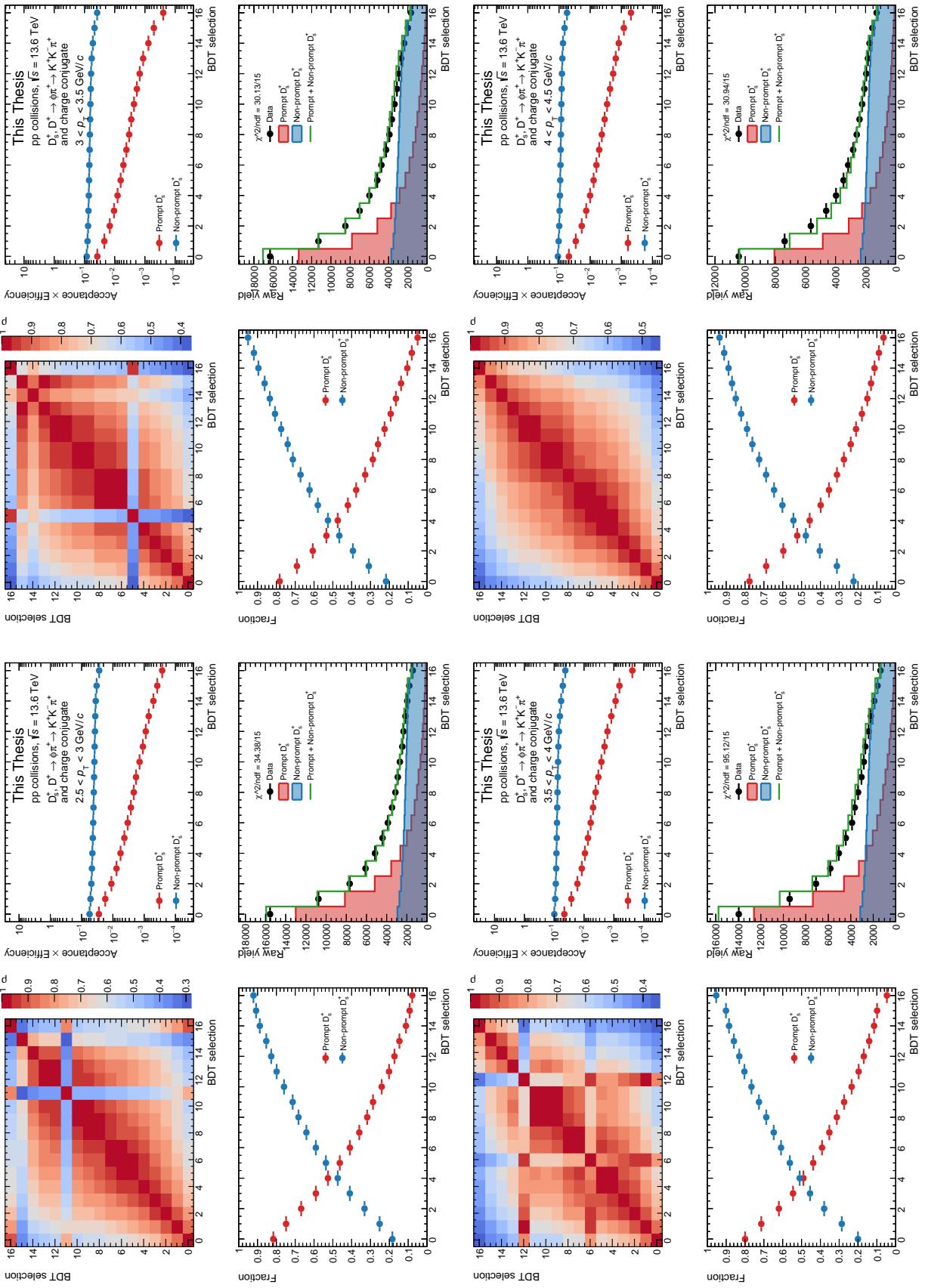


Figure B.2: Results for the evaluation of the  $f^{D_s^+}$  prompt correction factor in the  $2.5 < p_T < 3.0 \text{ GeV}/c$  (top left),  $3.0 < p_T < 3.5 \text{ GeV}/c$  (top right),  $3.5 < p_T < 4.0 \text{ GeV}/c$  (bottom left) and  $4.0 < p_T < 4.5 \text{ GeV}/c$  (bottom right) intervals.

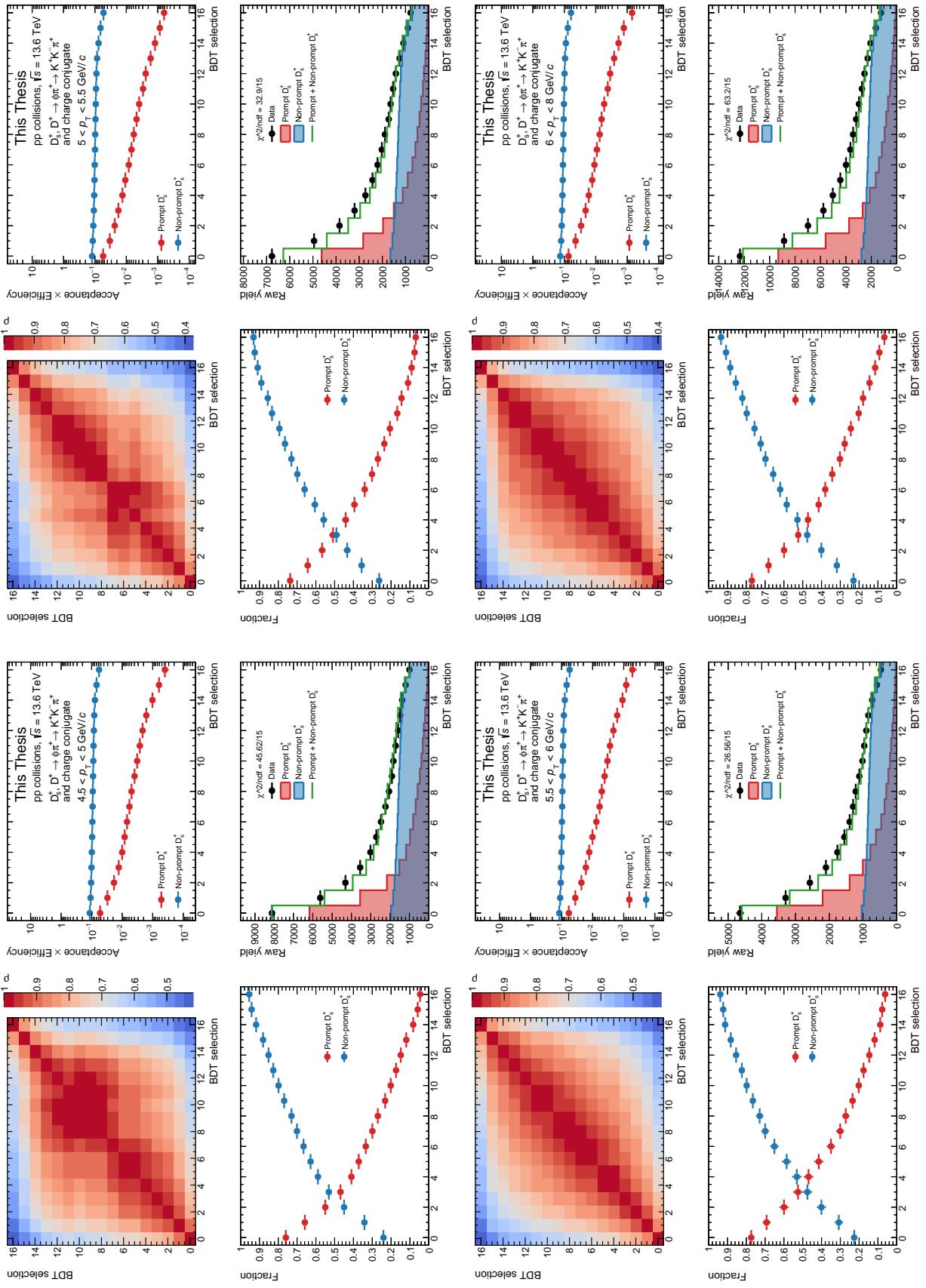


Figure B.3: Results for the evaluation of the  $f^{D_s^+}$  correction factor in the  $4.5 < p_T < 5.0 \text{ GeV}/c$  (top left),  $5.0 < p_T < 5.5 \text{ GeV}/c$  (top right),  $5.5 < p_T < 6.0 \text{ GeV}/c$  (bottom left) and  $6.0 < p_T < 8.0 \text{ GeV}/c$  (bottom right) intervals.

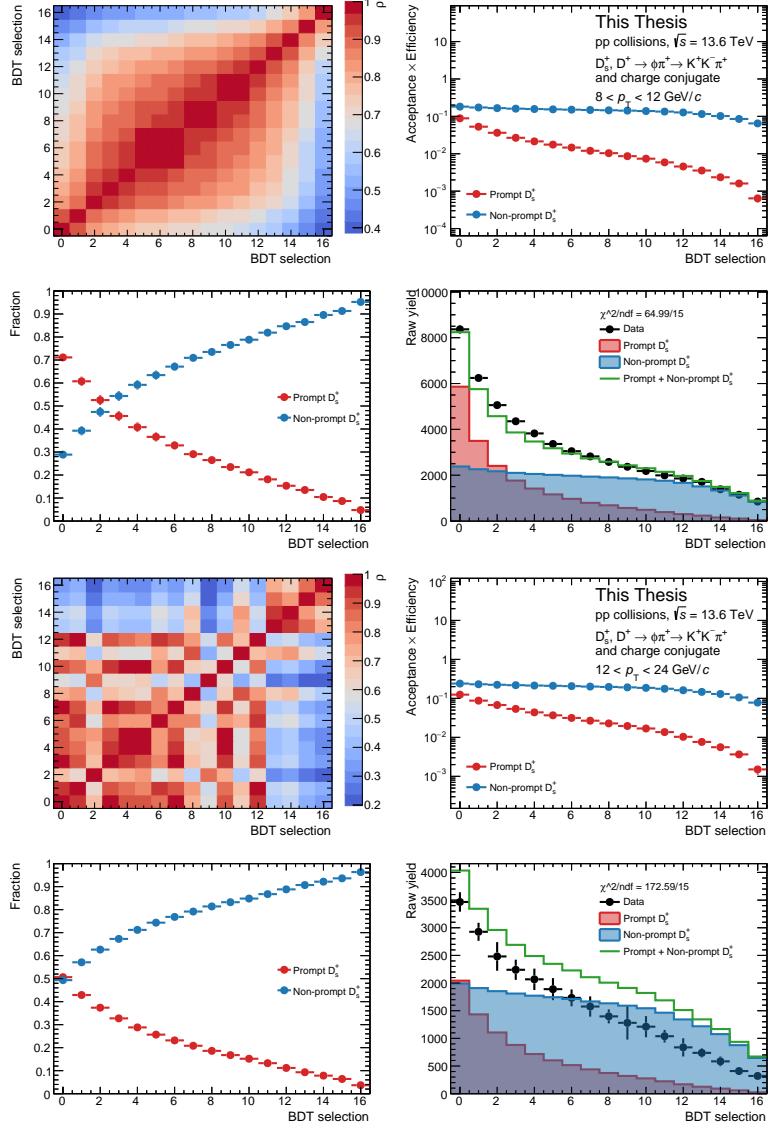


Figure B.4: Results for the evaluation of the  $f_{\text{prompt}}^{D_s^+}$  correction factor in the  $8.0 < p_T < 12.0 \text{ GeV}/c$  (top) and  $12.0 < p_T < 24.0 \text{ GeV}/c$  (bottom) intervals.

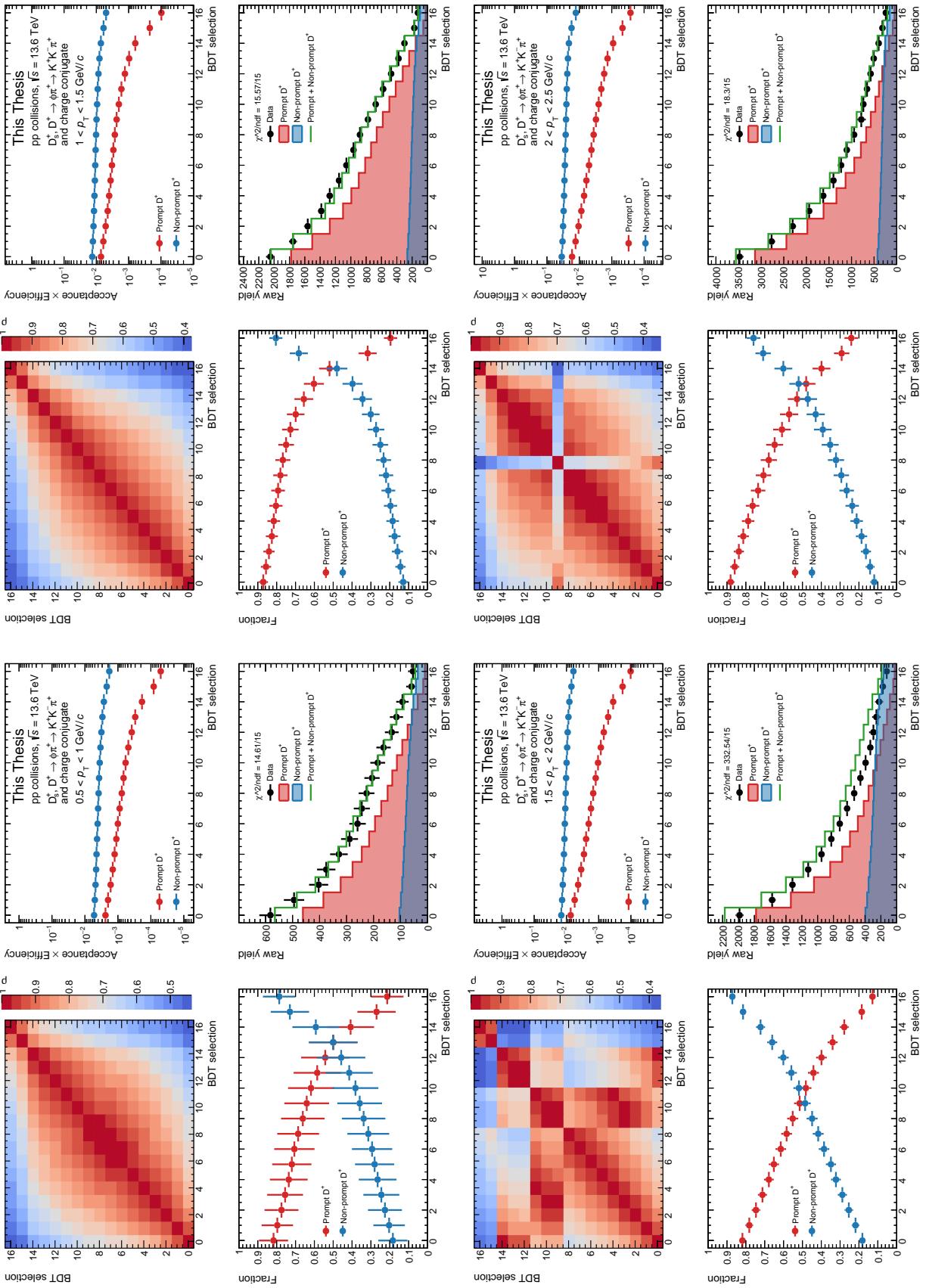


Figure B.5: Results for the evaluation of the  $f^{D^+}$  correction factor in the  $0.5 < p_T < 1.0 \text{ GeV}/c$  (top left),  $1.0 < p_T < 1.5 \text{ GeV}/c$  (top right),  $1.5 < p_T < 2.0 \text{ GeV}/c$  (bottom left) and  $2.0 < p_T < 2.5 \text{ GeV}/c$  (bottom right) intervals.

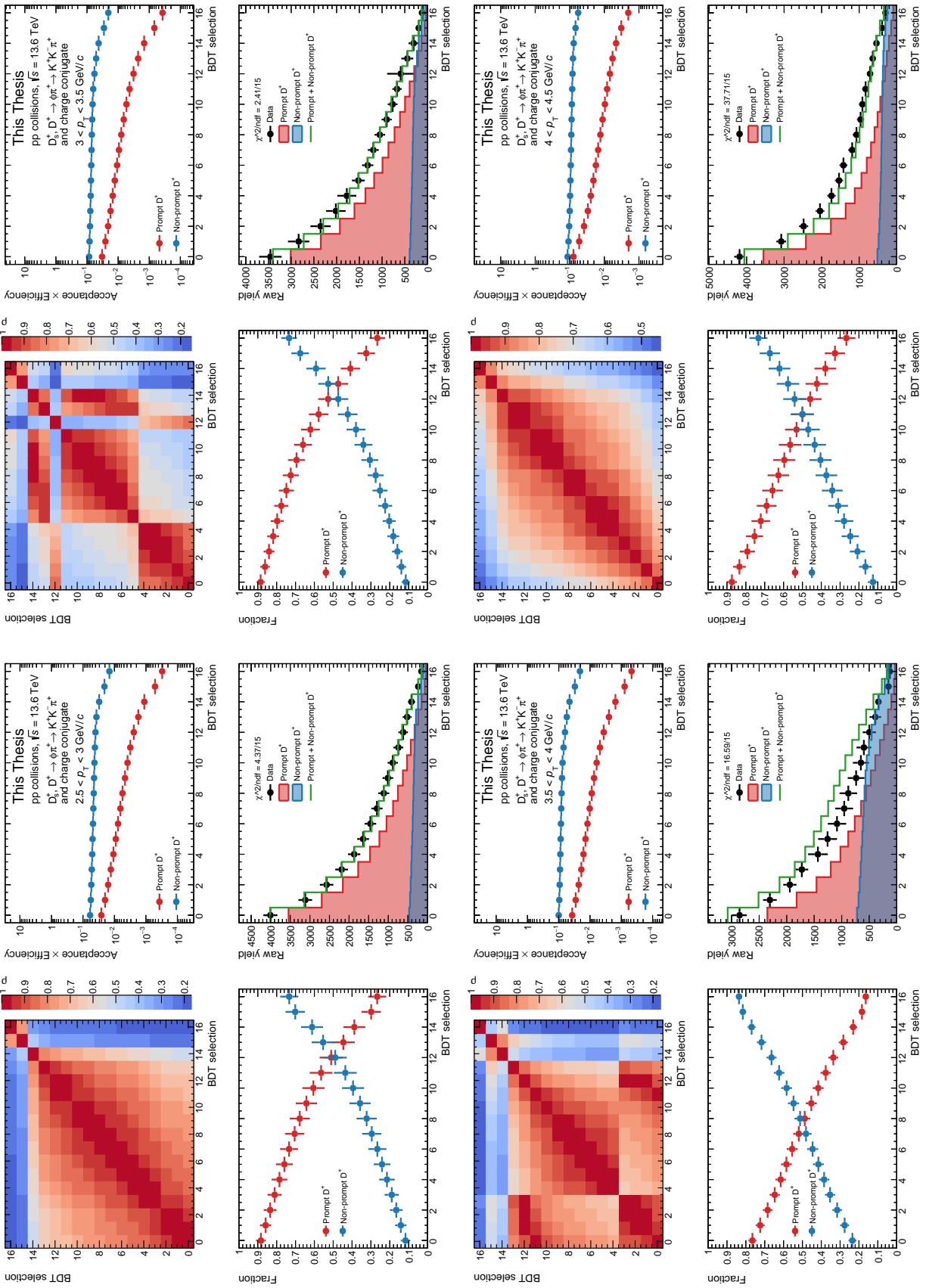


Figure B.6: Results for the evaluation of the  $f^{D^+}$  correction factor in the  $2.5 < p_T < 3.0 \text{ GeV}/c$  (top left),  $3.0 < p_T < 3.5 \text{ GeV}/c$  (top right),  $3.5 < p_T < 4.0 \text{ GeV}/c$  (bottom left) and  $4.0 < p_T < 4.5 \text{ GeV}/c$  (bottom right) intervals.

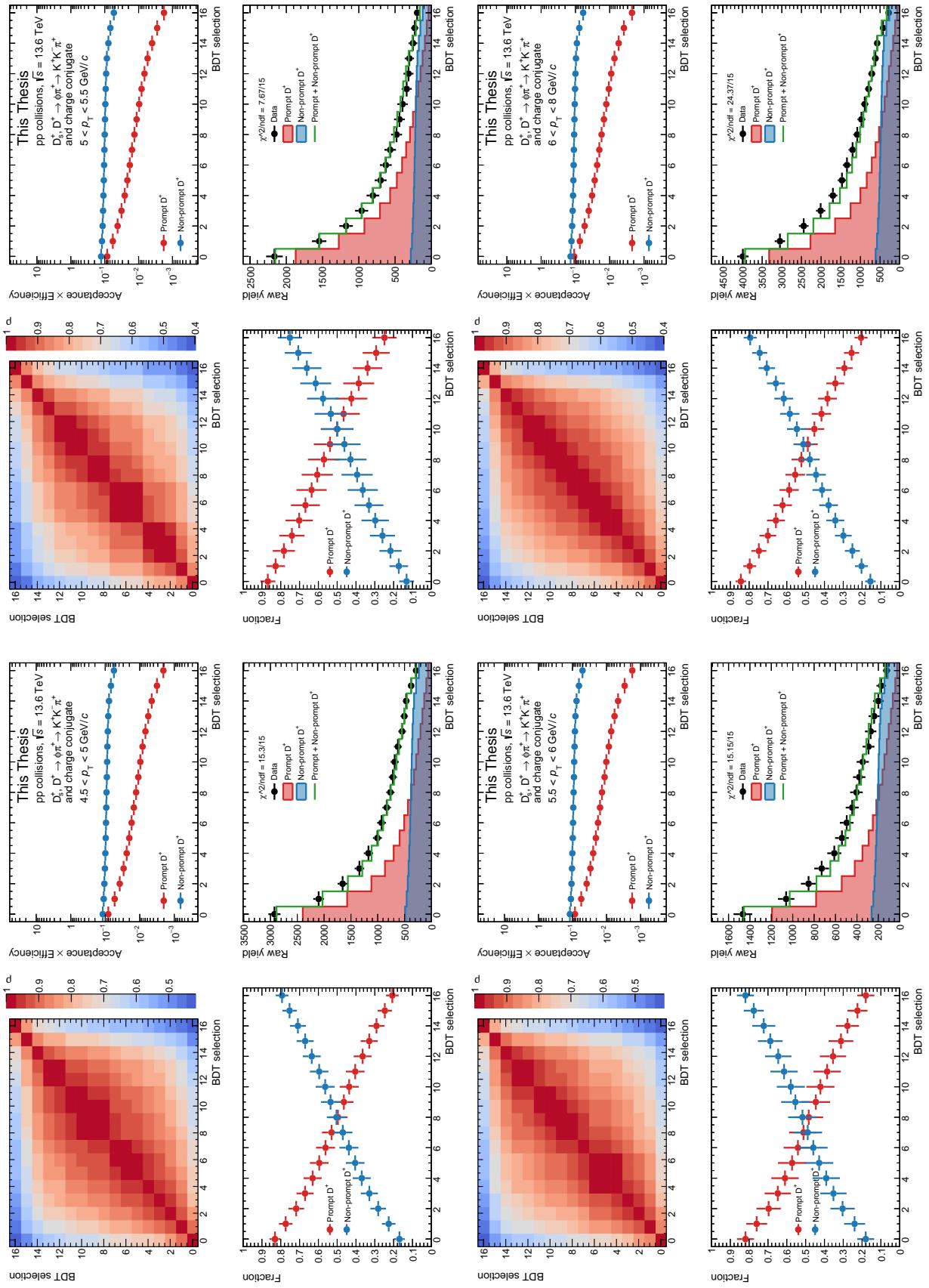


Figure B.7: Results for the evaluation of the  $f^{D^+}$  correction factor in the  $4.5 < p_T < 5.0 \text{ GeV}/c$  (top left),  $5.0 < p_T < 5.5 \text{ GeV}/c$  (top right),  $5.5 < p_T < 6.0 \text{ GeV}/c$  (bottom left) and  $6.0 < p_T < 8.0 \text{ GeV}/c$  (bottom right) intervals.

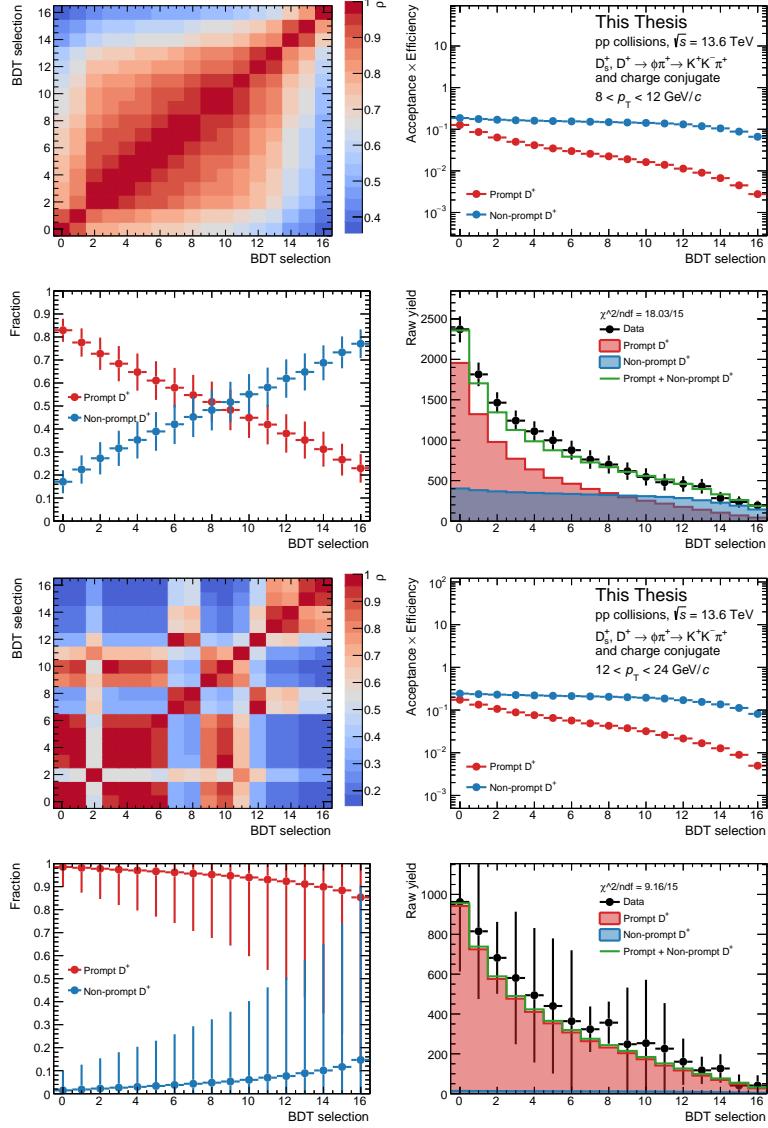


Figure B.8: Results for the evaluation of the  $f_{\text{prompt}}^{D^+}$  correction factor in the  $8.0 < p_T < 12.0 \text{ GeV}/c$  (top) and  $12.0 < p_T < 24.0 \text{ GeV}/c$  (bottom) intervals.

## **Appendix C**

### **Figures for the systematic uncertainty on the BDT selection efficiency**

In this Appendix, the results for the evaluation of the systematic uncertainty on the BDT selection efficiency are reported. The description of the used methodologies and the content of the figures can be found in Chapter 6.

## Appendix C – Figures for the systematic uncertainty on the BDT selection efficiency

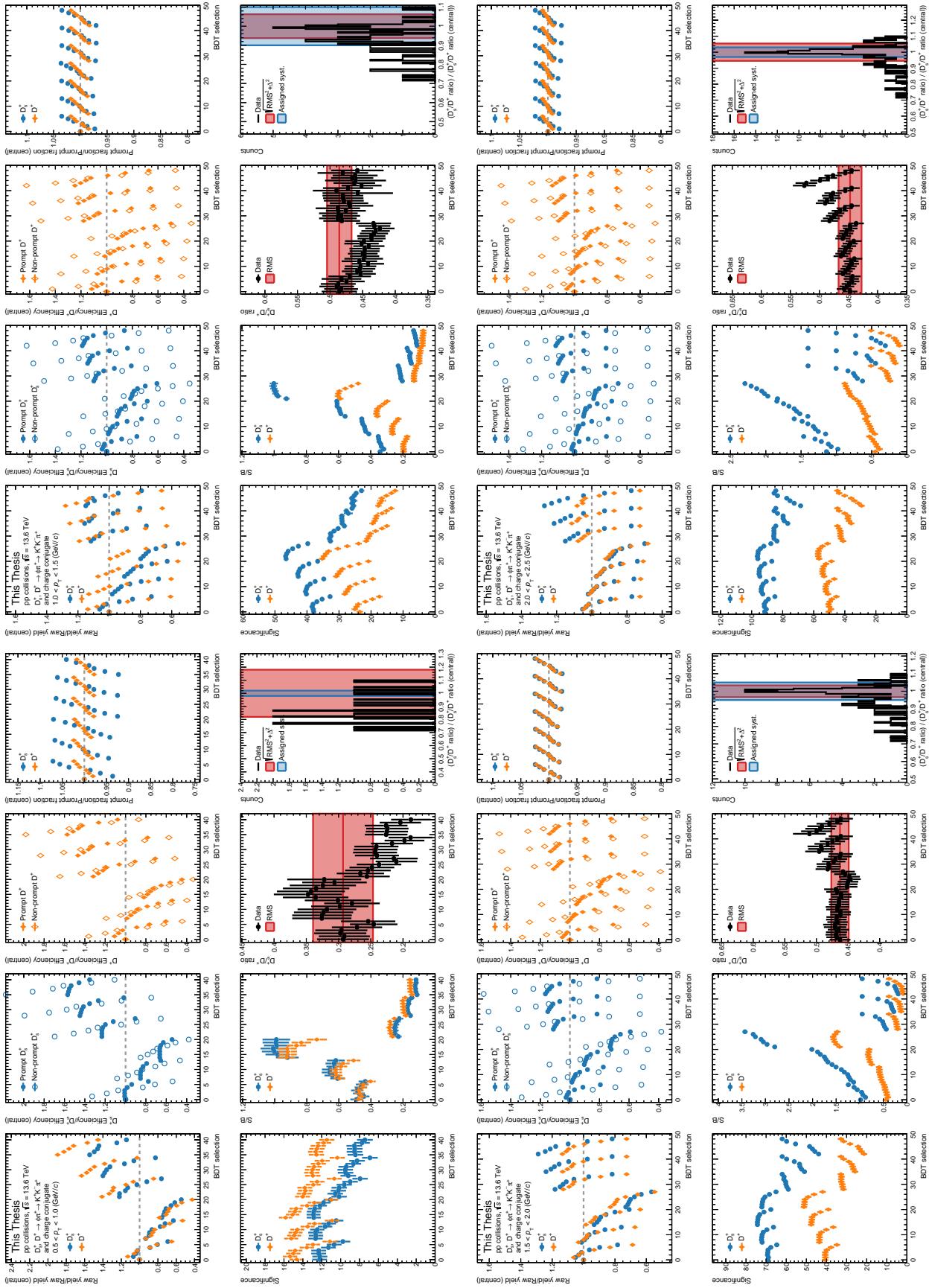


Figure C.1: Results for the evaluation of the systematic uncertainty on the BDT selection efficiency in the  $0.5 < p_T < 1.0 \text{ GeV}/c$  (top left),  $1.0 < p_T < 1.5 \text{ GeV}/c$  (top right),  $1.5 < p_T < 2.0 \text{ GeV}/c$  (bottom left) and  $2.0 < p_T < 2.5 \text{ GeV}/c$  (bottom right) intervals.

## Appendix C – Figures for the systematic uncertainty on the BDT selection efficiency

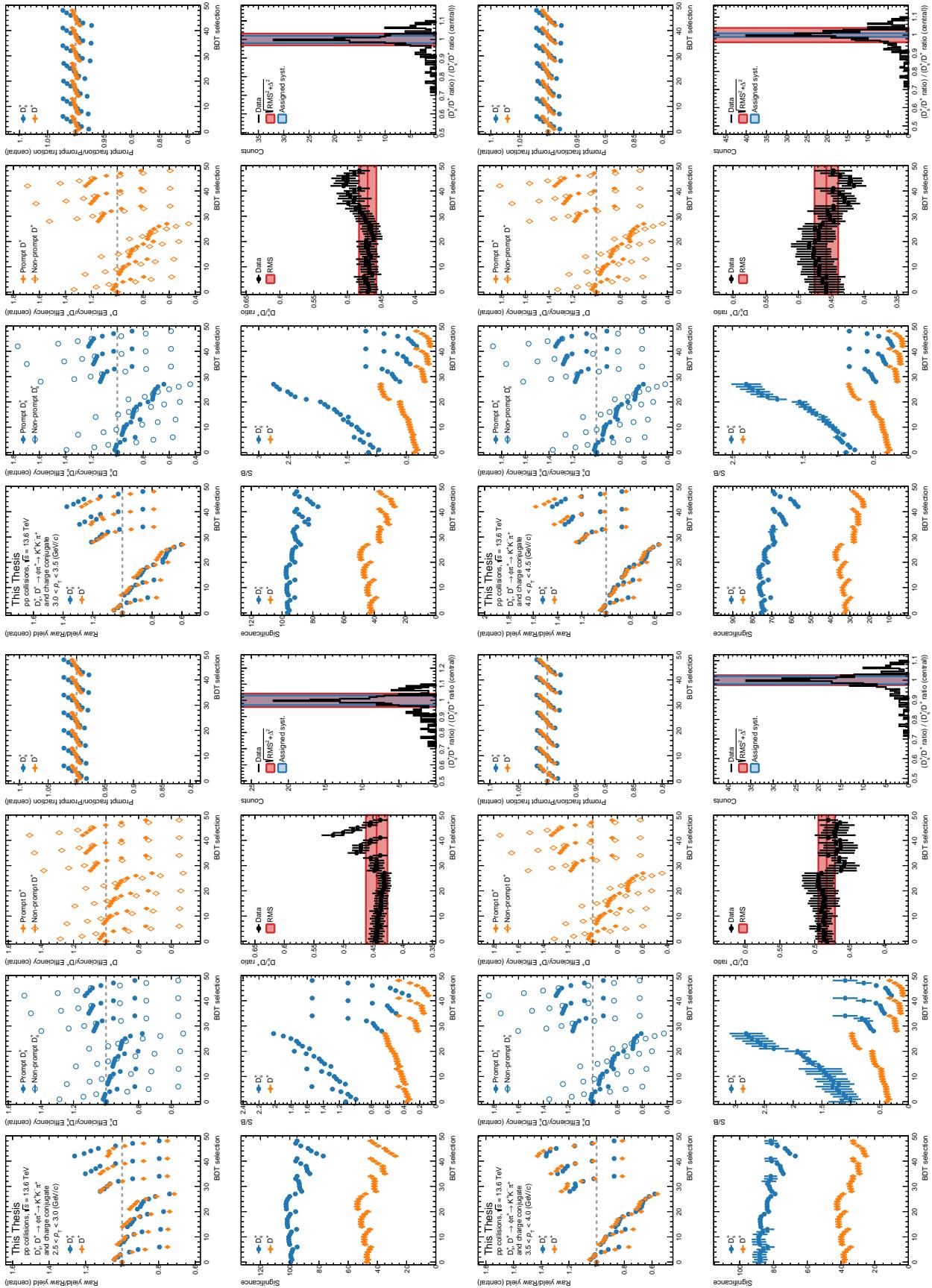


Figure C.2: Results for the evaluation of the systematic uncertainty on the BDT selection efficiency in the  $2.5 < p_T < 3.0 \text{ GeV}/c$  (top left),  $3.0 < p_T < 3.5 \text{ GeV}/c$  (top right),  $3.5 < p_T < 4.0 \text{ GeV}/c$  (bottom left) and  $4.0 < p_T < 4.5 \text{ GeV}/c$  (bottom right) intervals.

## Appendix C – Figures for the systematic uncertainty on the BDT selection efficiency

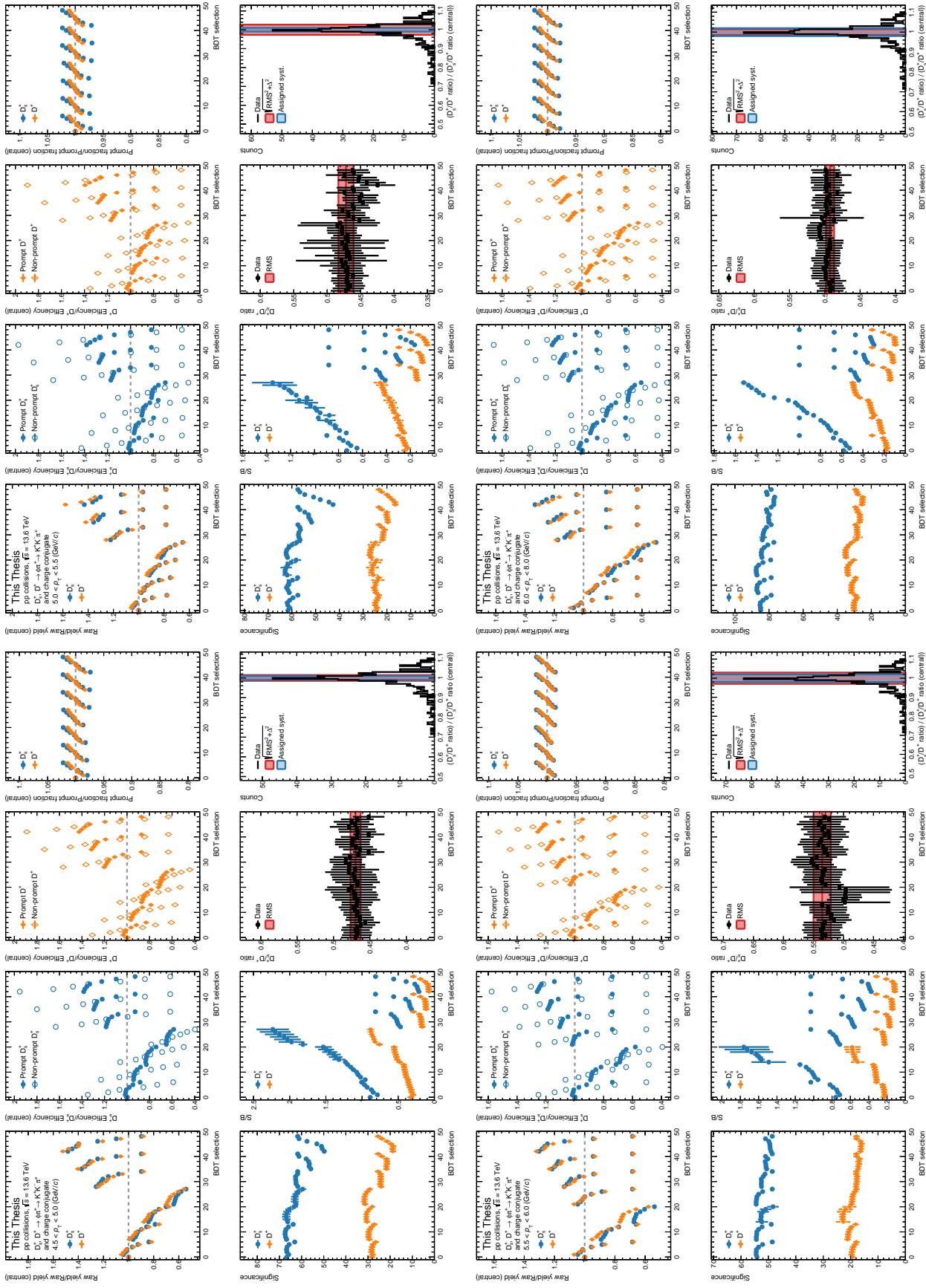


Figure C.3: Results for the evaluation of the systematic uncertainty on the BDT selection efficiency in the  $4.5 < p_T < 5.0 \text{ GeV}/c$  (top left),  $5.0 < p_T < 5.5 \text{ GeV}/c$  (top right),  $5.5 < p_T < 6.0 \text{ GeV}/c$  (bottom left) and  $6.0 < p_T < 8.0 \text{ GeV}/c$  (bottom right) intervals.

Appendix C – Figures for the systematic uncertainty on the BDT selection efficiency

---

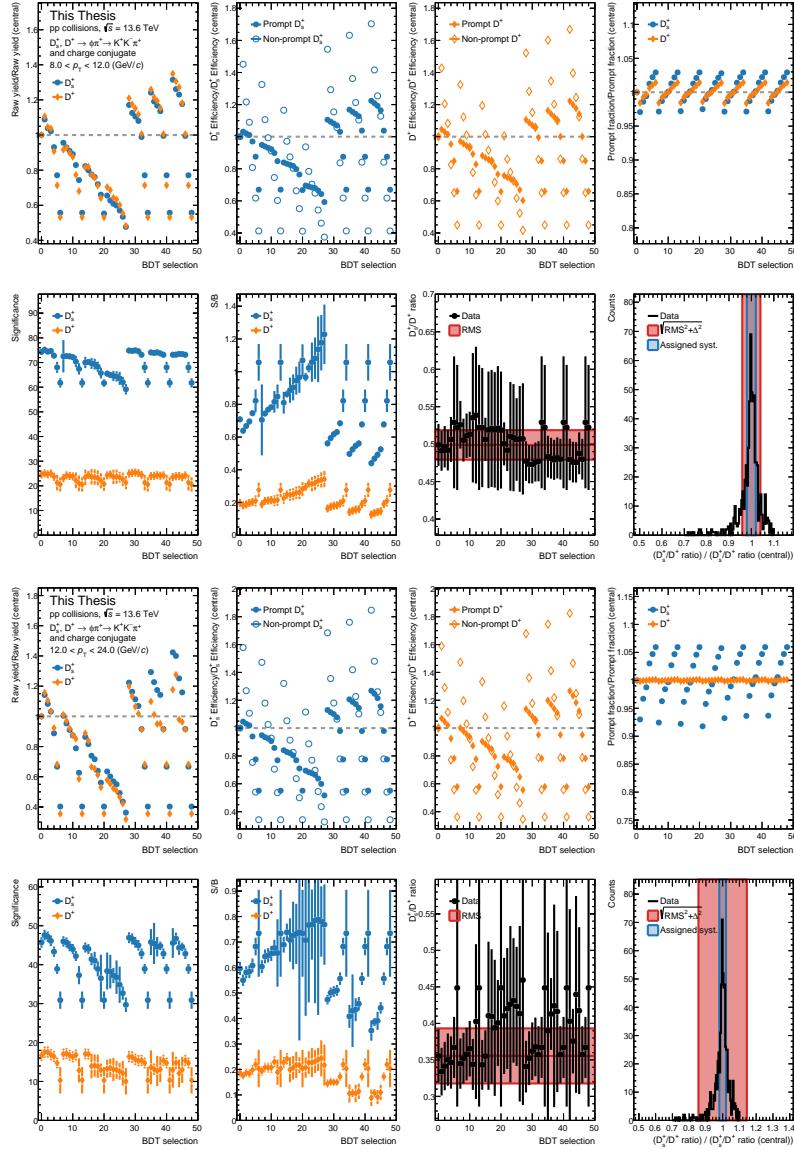


Figure C.4: Results for the evaluation of the systematic uncertainty on the BDT selection efficiency in the  $8.0 < p_T < 12.0$  GeV/ $c$  (top) and  $12.0 < p_T < 24.0$  GeV/ $c$  (bottom) intervals.

# Bibliography

- [1] M. Gell-Mann, “A Schematic Model of Baryons and Mesons”, *Phys. Lett.* **8** (1964) 214–215.
- [2] G. Zweig, *An SU(3) model for strong interaction symmetry and its breaking. Version 2*, pp. 22–101. 2, 1964.
- [3] H. Fritzsch and M. Gell-Mann, “Current algebra: Quarks and what else?”, *eConf C720906V2* (1972) 135–165, [arXiv:hep-ph/0208010](https://arxiv.org/abs/hep-ph/0208010).
- [4] **Particle Data Group** Collaboration, R. L. Workman and Others, “Review of Particle Physics”, *PTEP* **2022** (2022) 083C01.
- [5] F. Herzog, B. Ruijl, T. Ueda, J. A. M. Vermaseren, and A. Vogt, “The five-loop beta function of Yang-Mills theory with fermions”, *JHEP* **02** (2017) 090, [arXiv:1701.01404 \[hep-ph\]](https://arxiv.org/abs/1701.01404).
- [6] K. Johnson, “The M.I.T. Bag Model”, *Acta Phys. Polon. B* **6** (1975) 865.
- [7] R. P. Feynman, “Space-time approach to non-relativistic quantum mechanics”, *Rev. Mod. Phys.* **20** (Apr, 1948) 367–387.  
<https://link.aps.org/doi/10.1103/RevModPhys.20.367>.
- [8] **BMW** Collaboration, S. Durr *et al.*, “Ab-Initio Determination of Light Hadron Masses”, *Science* **322** (2008) 1224–1227, [arXiv:0906.3599 \[hep-lat\]](https://arxiv.org/abs/0906.3599).
- [9] **BMW** Collaboration, S. Borsanyi *et al.*, “Ab initio calculation of the neutron-proton mass difference”, *Science* **347** (2015) 1452–1455, [arXiv:1406.4088 \[hep-lat\]](https://arxiv.org/abs/1406.4088).
- [10] A. Bazavov *et al.*, “The QCD Equation of State to  $\mathcal{O}(\mu_B^6)$  from Lattice QCD”, *Phys. Rev. D* **95** (2017) 054504, [arXiv:1701.04325 \[hep-lat\]](https://arxiv.org/abs/1701.04325).
- [11] **HotQCD** Collaboration, A. Bazavov *et al.*, “Equation of state in ( 2+1 )-flavor QCD”, *Phys. Rev. D* **90** (2014) 094503, [arXiv:1407.6387 \[hep-lat\]](https://arxiv.org/abs/1407.6387).
- [12] S. Borsanyi, Z. Fodor, C. Hoelbling, S. D. Katz, S. Krieg, and K. K. Szabo, “Full result for the QCD equation of state with 2+1 flavors”, *Phys. Lett. B* **730** (2014) 99–104, [arXiv:1309.5258 \[hep-lat\]](https://arxiv.org/abs/1309.5258).

- [13] CERN Courier, “Quark-matter fireballs hashed out in Protvino”, 2021.
- [14] J. Rafelski and J. Birrell, “Traveling through the universe: Back in time to the quark-gluon plasma era”, in *Journal of Physics: Conference Series*, vol. 509, p. 012014, IOP Publishing. 2014.
- [15] E. Annala, T. Gorda, A. Kurkela, J. Näättilä, and A. Vuorinen, “Evidence for quark-matter cores in massive neutron stars”, *Nature Phys.* **16** (2020) 907–910, [arXiv:1903.09121 \[astro-ph.HE\]](https://arxiv.org/abs/1903.09121).
- [16] **ALICE** Collaboration, “The ALICE experiment – A journey through QCD”, [arXiv:2211.04384 \[nucl-ex\]](https://arxiv.org/abs/2211.04384).
- [17] **NA38, NA50** Collaboration, M. C. Abreu *et al.*, “Dimuon and charm production in nucleus-nucleus collisions at the CERN SPS”, *Eur. Phys. J. C* **14** (2000) 443–455.
- [18] **NA50** Collaboration, M. C. Abreu *et al.*, “Anomalous J / psi suppression in Pb - Pb interactions at 158 GeV/c per nucleon”, *Phys. Lett. B* **410** (1997) 337–343.
- [19] R. Nouicer, “Formation of Dense Partonic Matter in High Energy Heavy-Ion Collisions: Highlights of RHIC Results”, in *Advanced Studies Institute on Symmetries and Spin (SPIN-Praha-2008)*. 1, 2009. [arXiv:0901.0910 \[nucl-ex\]](https://arxiv.org/abs/0901.0910).
- [20] R. Stock, “Relativistic Nucleus-Nucleus Collisions and the QCD Matter Phase Diagram”, [arXiv:0807.1610 \[nucl-ex\]](https://arxiv.org/abs/0807.1610).
- [21] **ALICE** Collaboration, C. Loizides, “Charged-particle multiplicity and transverse energy in Pb-Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV with ALICE”, *J. Phys. G* **38** (2011) 124040, [arXiv:1106.6324 \[nucl-ex\]](https://arxiv.org/abs/1106.6324).
- [22] J. D. Bjorken, “Highly Relativistic Nucleus-Nucleus Collisions: The Central Rapidity Region”, *Phys. Rev. D* **27** (1983) 140–151.
- [23] **PHENIX** Collaboration, K. Adcox *et al.*, “Formation of dense partonic matter in relativistic nucleus-nucleus collisions at RHIC: Experimental evaluation by the PHENIX collaboration”, *Nucl. Phys. A* **757** (2005) 184–283, [arXiv:nucl-ex/0410003](https://arxiv.org/abs/nucl-ex/0410003).
- [24] **ALICE** Collaboration, J. Adam *et al.*, “Measurement of transverse energy at midrapidity in Pb-Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV”, *Phys. Rev. C* **94** (2016) 034903, [arXiv:1603.04775 \[nucl-ex\]](https://arxiv.org/abs/1603.04775).
- [25] **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 \[nucl-ex\]](https://arxiv.org/abs/1012.4035).

- [26] A. Andronic, P. Braun-Munzinger, K. Redlich, and J. Stachel, “Decoding the phase structure of QCD via particle production at high energy”, *Nature* **561** (2018) 321–330, [arXiv:1710.09425 \[nucl-th\]](#).
- [27] P. Braun-Munzinger, K. Redlich, and J. Stachel, “Particle production in heavy ion collisions”, [arXiv:nucl-th/0304013](#).
- [28] S. Wheaton and J. Cleymans, “THERMUS: A Thermal model package for ROOT”, *Comput. Phys. Commun.* **180** (2009) 84–106, [arXiv:hep-ph/0407174](#).
- [29] G. Torrieri, S. Steinke, W. Broniowski, W. Florkowski, J. Letessier, and J. Rafelski, “SHARE: Statistical hadronization with resonances”, *Comput. Phys. Commun.* **167** (2005) 229–251, [arXiv:nucl-th/0404083](#).
- [30] V. Vovchenko and H. Stoecker, “Thermal-FIST: A package for heavy-ion collisions and hadronic equation of state”, *Comput. Phys. Commun.* **244** (2019) 295–310, [arXiv:1901.05249 \[nucl-th\]](#).
- [31] 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](#).
- [32] **ALICE** Collaboration, S. Acharya *et al.*, “Measurements of chemical potentials in Pb-Pb collisions at  $\sqrt{s_{\text{NN}}} = 5.02 \text{ TeV}$ ”, [arXiv:2311.13332 \[nucl-ex\]](#).
- [33] 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](#).
- [34] **ALICE** Collaboration, B. Abelev *et al.*, “Centrality dependence of  $\pi$ , K, p production in Pb-Pb collisions at  $\sqrt{s_{\text{NN}}} = 2.76 \text{ TeV}$ ”, *Phys. Rev. C* **88** (2013) 044910, [arXiv:1303.0737 \[hep-ex\]](#).
- [35] 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](#).
- [36] F. Cooper and G. Frye, “Comment on the Single Particle Distribution in the Hydrodynamic and Statistical Thermodynamic Models of Multiparticle Production”, *Phys. Rev. D* **10** (1974) 186.
- [37] 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](#).
- [38] J. S. Moreland, J. E. Bernhard, and S. A. Bass, “Alternative ansatz to wounded nucleon and binary collision scaling in high-energy nuclear collisions”, *Phys. Rev. C* **92** (2015) 011901, [arXiv:1412.4708 \[nucl-th\]](#).

- [39] T. Lappi and L. McLerran, “Some features of the plasma”, *Nucl. Phys. A* **772** (2006) 200–212, [arXiv:hep-ph/0602189](#).
- [40] Z.-W. Lin, C. M. Ko, B.-A. Li, B. Zhang, and S. Pal, “A Multi-phase transport model for relativistic heavy ion collisions”, *Phys. Rev. C* **72** (2005) 064901, [arXiv:nucl-th/0411110](#).
- [41] **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 \[nucl-ex\]](#).
- [42] M. Arneodo, “Nuclear effects in structure functions”, *Phys. Rept.* **240** (1994) 301–393.
- [43] B. Z. Kopeliovich, J. Nemchik, A. Schafer, and A. V. Tarasov, “Cronin effect in hadron production off nuclei”, *Phys. Rev. Lett.* **88** (2002) 232303, [arXiv:hep-ph/0201010](#).
- [44] **PHENIX** Collaboration, U. A. Acharya *et al.*, “Systematic study of nuclear effects in  $p + \text{Al}$ ,  $p + \text{Au}$ ,  $d + \text{Au}$ , and  ${}^3\text{He} + \text{Au}$  collisions at  $\sqrt{s_{NN}} = 200$  GeV using  $\pi^0$  production”, *Phys. Rev. C* **105** (2022) 064902, [arXiv:2111.05756 \[nucl-ex\]](#).
- [45] **PHENIX** Collaboration, S. Afanasiev *et al.*, “Measurement of Direct Photons in Au+Au Collisions at  $\sqrt{s_{NN}} = 200$  GeV”, *Phys. Rev. Lett.* **109** (2012) 152302, [arXiv:1205.5759 \[nucl-ex\]](#).
- [46] R. Baier, Y. L. Dokshitzer, A. H. Mueller, S. Peigne, and D. Schiff, “Radiative energy loss of high-energy quarks and gluons in a finite volume quark - gluon plasma”, *Nucl. Phys. B* **483** (1997) 291–320, [arXiv:hep-ph/9607355](#).
- [47] R. Baier, “Jet quenching”, *Nucl. Phys. A* **715** (2003) 209–218, [arXiv:hep-ph/0209038](#).
- [48] **STAR** Collaboration, J. Adams *et al.*, “Evidence from d + Au measurements for final state suppression of high  $p_T$  hadrons in Au+Au collisions at RHIC”, *Phys. Rev. Lett.* **91** (2003) 072304, [arXiv:nucl-ex/0306024](#).
- [49] **ALICE** Collaboration, J. Adam *et al.*, “Centrality dependence of pion freeze-out radii in Pb-Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV”, *Phys. Rev. C* **93** (2016) 024905, [arXiv:1507.06842 \[nucl-ex\]](#).
- [50] **ATLAS** Collaboration, M. Aaboud *et al.*, “Measurement of jet  $p_T$  correlations in Pb+Pb and  $pp$  collisions at  $\sqrt{s_{NN}} = 2.76$  TeV with the ATLAS detector”, *Phys. Lett. B* **774** (2017) 379–402, [arXiv:1706.09363 \[hep-ex\]](#).

- [51] 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)].
- [52] K. Redlich and A. Tounsi, “Strangeness enhancement and energy dependence in heavy ion collisions”, *Eur. Phys. J. C* **24** (2002) 589–594, [arXiv:hep-ph/0111261](https://arxiv.org/abs/hep-ph/0111261).
- [53] **WA97** Collaboration, E. Andersen *et al.*, “Strangeness enhancement at mid-rapidity in Pb Pb collisions at 158-A-GeV/c”, *Phys. Lett. B* **449** (1999) 401–406.
- [54] **NA57** Collaboration, F. Antinori *et al.*, “Strangeness enhancements at central rapidity in 40 A GeV/c Pb-Pb collisions”, *J. Phys. G* **37** (2010) 045105, [arXiv:1001.1884](https://arxiv.org/abs/1001.1884) [nucl-ex].
- [55] **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](https://arxiv.org/abs/0705.2511) [nucl-ex].
- [56] **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](https://arxiv.org/abs/1307.5543) [nucl-ex]. [Erratum: Phys.Lett.B 734, 409–410 (2014)].
- [57] **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](https://arxiv.org/abs/1606.07424) [nucl-ex].
- [58] **ALICE** Collaboration, B. B. Abelev *et al.*, “Multiplicity Dependence of Pion, Kaon, Proton and Lambda Production in p-Pb Collisions at  $\sqrt{s_{NN}} = 5.02$  TeV”, *Phys. Lett. B* **728** (2014) 25–38, [arXiv:1307.6796](https://arxiv.org/abs/1307.6796) [nucl-ex].
- [59] **ALICE** Collaboration, J. Adam *et al.*, “Multi-strange baryon production in p-Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV”, *Phys. Lett. B* **758** (2016) 389–401, [arXiv:1512.07227](https://arxiv.org/abs/1512.07227) [nucl-ex].
- [60] **ALICE** Collaboration, “ALICE Figure.” <https://alice-figure.web.cern.ch/>.
- [61] 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](https://arxiv.org/abs/hep-ph/0409313).
- [62] **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](https://arxiv.org/abs/2102.13601) [nucl-ex].
- [63] M. Cacciari, M. Greco, and P. Nason, “The  $p_T$  spectrum in heavy-flavour hadroproduction.”, *JHEP* **05** (1998) 007, [arXiv:hep-ph/9803400](https://arxiv.org/abs/hep-ph/9803400).

- [64] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, “An introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* **191** (2015) 159–177, [arXiv:1410.3012 \[hep-ph\]](https://arxiv.org/abs/1410.3012).
- [65] **CDF** Collaboration, D. Acosta *et al.*, “Measurement of prompt charm meson production cross sections in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96$  TeV”, *Phys. Rev. Lett.* **91** (2003) 241804, [arXiv:hep-ex/0307080](https://arxiv.org/abs/hep-ex/0307080).
- [66] **STAR** Collaboration, L. Adamczyk *et al.*, “Measurements of  $D^0$  and  $D^*$  Production in  $p + p$  Collisions at  $\sqrt{s} = 200$  GeV”, *Phys. Rev. D* **86** (2012) 072013, [arXiv:1204.4244 \[nucl-ex\]](https://arxiv.org/abs/1204.4244).
- [67] J. I. Friedman and H. W. Kendall, “Deep inelastic electron scattering”, *Ann. Rev. Nucl. Part. Sci.* **22** (1972) 203–254.
- [68] J. D. Bjorken, “Asymptotic Sum Rules at Infinite Momentum”, *Phys. Rev.* **179** (1969) 1547–1553.
- [69] R. P. Feynman, “Very high-energy collisions of hadrons”, *Phys. Rev. Lett.* **23** (1969) 1415–1417.
- [70] C. G. Callan, Jr. and D. J. Gross, “High-energy electroproduction and the constitution of the electric current”, *Phys. Rev. Lett.* **22** (1969) 156–159.
- [71] V. N. Gribov and L. N. Lipatov, “Deep inelastic e p scattering in perturbation theory”, *Sov. J. Nucl. Phys.* **15** (1972) 438–450.
- [72] Y. L. Dokshitzer, “Calculation of the Structure Functions for Deep Inelastic Scattering and  $e^+e^-$  Annihilation by Perturbation Theory in Quantum Chromodynamics.”, *Sov. Phys. JETP* **46** (1977) 641–653.
- [73] G. Altarelli and G. Parisi, “Asymptotic Freedom in Parton Language”, *Nucl. Phys. B* **126** (1977) 298–318.
- [74] **NNPDF** Collaboration, R. D. Ball *et al.*, “The path to proton structure at 1% accuracy”, *Eur. Phys. J. C* **82** (2022) 428, [arXiv:2109.02653 \[hep-ph\]](https://arxiv.org/abs/2109.02653).
- [75] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan, “New parton distribution functions from a global analysis of quantum chromodynamics”, *Phys. Rev. D* **93** (2016) 033006, [arXiv:1506.07443 \[hep-ph\]](https://arxiv.org/abs/1506.07443).
- [76] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, “Parton distributions in the LHC era: MMHT 2014 PDFs”, *Eur. Phys. J. C* **75** (2015) 204, [arXiv:1412.3989 \[hep-ph\]](https://arxiv.org/abs/1412.3989).
- [77] C. Anastasiou, C. Duhr, F. Dulat, F. Herzog, and B. Mistlberger, “Higgs Boson Gluon-Fusion Production in QCD at Three Loops”, *Phys. Rev. Lett.* **114** (2015) 212001, [arXiv:1503.06056 \[hep-ph\]](https://arxiv.org/abs/1503.06056).

- [78] C. Anastasiou, C. Duhr, F. Dulat, E. Furlan, T. Gehrmann, F. Herzog, A. Lazopoulos, and B. Mistlberger, “High precision determination of the gluon fusion Higgs boson cross-section at the LHC”, *JHEP* **05** (2016) 058, [arXiv:1602.00695 \[hep-ph\]](https://arxiv.org/abs/1602.00695).
- [79] S. Catani, S. Devoto, M. Grazzini, S. Kallweit, and J. Mazzitelli, “Bottom-quark production at hadron colliders: fully differential predictions in NNLO QCD”, *JHEP* **03** (2021) 029, [arXiv:2010.11906 \[hep-ph\]](https://arxiv.org/abs/2010.11906).
- [80] J. D. Bjorken, “Properties of Hadron Distributions in Reactions Containing Very Heavy Quarks”, *Phys. Rev. D* **17** (1978) 171–173.
- [81] M. Suzuki, “Fragmentation of Hadrons from Heavy Quark Partons”, *Phys. Lett. B* **71** (1977) 139–141.
- [82] C. Peterson, D. Schlatter, I. Schmitt, and P. M. Zerwas, “Scaling Violations in Inclusive e+ e- Annihilation Spectra”, *Phys. Rev. D* **27** (1983) 105.
- [83] **NNPDF** Collaboration, V. Bertone, N. P. Hartland, E. R. Nocera, J. Rojo, and L. Rottoli, “Charged hadron fragmentation functions from collider data”, *Eur. Phys. J. C* **78** (2018) 651, [arXiv:1807.03310 \[hep-ph\]](https://arxiv.org/abs/1807.03310). [Erratum: Eur.Phys.J.C 84, 155 (2024)].
- [84] D. de Florian, R. Sassot, and M. Stratmann, “Global analysis of fragmentation functions for pions and kaons and their uncertainties”, *Phys. Rev. D* **75** (2007) 114010, [arXiv:hep-ph/0703242](https://arxiv.org/abs/hep-ph/0703242).
- [85] B. A. Kniehl, G. Kramer, and B. Potter, “Fragmentation functions for pions, kaons, and protons at next-to-leading order”, *Nucl. Phys. B* **582** (2000) 514–536, [arXiv:hep-ph/0010289](https://arxiv.org/abs/hep-ph/0010289).
- [86] 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.
- [87] 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](https://arxiv.org/abs/hep-ph/9409316).
- [88] V. V. Sudakov, “Vertex parts at very high-energies in quantum electrodynamics”, *Sov. Phys. JETP* **3** (1956) 65–71.
- [89] Y. I. Azimov, Y. L. Dokshitzer, V. A. Khoze, and S. I. Troyan, “Similarity of Parton and Hadron Spectra in QCD Jets”, *Z. Phys. C* **27** (1985) 65–72.
- [90] R. D. Field and R. P. Feynman, “Quark Elastic Scattering as a Source of High Transverse Momentum Mesons”, *Phys. Rev. D* **15** (1977) 2590–2616.
- [91] **ALICE** Collaboration, S. Acharya *et al.*, “ $\Lambda_c^+$  production in  $pp$  and in  $p\text{-Pb}$  collisions at  $\sqrt{s_{NN}}=5.02$  TeV”, *Phys. Rev. C* **104** (2021) 054905, [arXiv:2011.06079 \[nucl-ex\]](https://arxiv.org/abs/2011.06079).

- [92] **ALICE** Collaboration, S. Acharya *et al.*, “Measurement of the production cross section of prompt  $\Xi_c^0$  baryons in p–Pb collisions at  $\sqrt{s_{\text{NN}}} = 5.02$  TeV”, [arXiv:2405.14538 \[nucl-ex\]](#).
- [93] P. Skands, S. Carrazza, and J. Rojo, “Tuning PYTHIA 8.1: the Monash 2013 Tune”, *Eur. Phys. J. C* **74** (2014) 3024, [arXiv:1404.5630 \[hep-ph\]](#).
- [94] J. Bellm *et al.*, “Herwig 7.0/Herwig++ 3.0 release note”, *Eur. Phys. J. C* **76** (2016) 196, [arXiv:1512.01178 \[hep-ph\]](#).
- [95] 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 \[hep-ph\]](#).
- [96] B. A. Kniehl, G. Kramer, I. Schienbein, and H. Spiesberger, “Collinear subtractions in hadroproduction of heavy quarks”, *Eur. Phys. J. C* **41** (2005) 199–212, [arXiv:hep-ph/0502194](#).
- [97] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, “Parton Fragmentation and String Dynamics”, *Phys. Rept.* **97** (1983) 31–145.
- [98] C. Bierlich *et al.*, “A comprehensive guide to the physics and usage of PYTHIA 8.3”, *SciPost Phys. Codeb.* **2022** (2022) 8, [arXiv:2203.11601 \[hep-ph\]](#).
- [99] E. Eichten, K. Gottfried, T. Kinoshita, J. B. Kogut, K. D. Lane, and T.-M. Yan, “The Spectrum of Charmonium”, *Phys. Rev. Lett.* **34** (1975) 369–372. [Erratum: *Phys. Rev. Lett.* **36**, 1276 (1976)].
- [100] B. Andersson, G. Gustafson, and B. Soderberg, “A General Model for Jet Fragmentation”, *Z. Phys. C* **20** (1983) 317.
- [101] S. Ferreres-Solé and T. Sjöstrand, “The space–time structure of hadronization in the Lund model”, *Eur. Phys. J. C* **78** (2018) 983, [arXiv:1808.04619 \[hep-ph\]](#).
- [102] T. Sjöstrand, “Jet Fragmentation of Nearby Partons”, *Nucl. Phys. B* **248** (1984) 469–502.
- [103] **ALICE** Collaboration, S. Acharya *et al.*, “First measurement of  $\Lambda_c^+$  production down to  $p_T = 0$  in pp and p–Pb collisions at  $\sqrt{s_{\text{NN}}} = 5.02$  TeV”, *Phys. Rev. C* **107** (2023) 064901, [arXiv:2211.14032 \[nucl-ex\]](#).
- [104] J. R. Christiansen and P. Z. Skands, “String Formation Beyond Leading Colour”, *JHEP* **08** (2015) 003, [arXiv:1505.01681 \[hep-ph\]](#).
- [105] B. R. Webber, “A QCD Model for Jet Fragmentation Including Soft Gluon Interference”, *Nucl. Phys. B* **238** (1984) 492–528.
- [106] M. Bahr *et al.*, “Herwig++ Physics and Manual”, *Eur. Phys. J. C* **58** (2008) 639–707, [arXiv:0803.0883 \[hep-ph\]](#).

- [107] D. Amati and G. Veneziano, “Preconfinement as a Property of Perturbative QCD”, *Phys. Lett. B* **83** (1979) 87–92.
- [108] M. H. Seymour and M. Marx, “Monte Carlo Event Generators”, in *69th Scottish Universities Summer School in Physics: LHC Physics*, pp. 287–319. 4, 2013. [arXiv:1304.6677 \[hep-ph\]](#).
- [109] **ALICE** Collaboration, S. Acharya *et al.*, “Prompt  $D^0$ ,  $D^+$ , and  $D^{*+}$  production in Pb–Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV”, *JHEP* **01** (2022) 174, [arXiv:2110.09420 \[nucl-ex\]](#).
- [110] W. Cassing and E. L. Bratkovskaya, “Parton transport and hadronization from the dynamical quasiparticle point of view”, *Phys. Rev. C* **78** (2008) 034919, [arXiv:0808.0022 \[hep-ph\]](#).
- [111] A. Beraudo, A. De Pace, M. Monteno, M. Nardi, and F. Prino, “Heavy flavors in heavy-ion collisions: quenching, flow and correlations”, *Eur. Phys. J. C* **75** (2015) 121, [arXiv:1410.6082 \[hep-ph\]](#).
- [112] R. Katz, C. A. G. Prado, J. Noronha-Hostler, J. Noronha, and A. A. P. Suaide, “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** (2020) 024906, [arXiv:1906.10768 \[nucl-th\]](#).
- [113] **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** (2021) 202301, [arXiv:2011.06078 \[nucl-ex\]](#).
- [114] **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  $\text{BR}(\Xi_c^0 \rightarrow \Xi^- e^+ \nu_e)/\text{BR}(\Xi_c^0 \rightarrow \Xi^- \pi^+)$  in pp collisions at 13 TeV”, *Phys. Rev. Lett.* **127** (2021) 272001, [arXiv:2105.05187 \[nucl-ex\]](#).
- [115] **CMS** Collaboration, V. Khachatryan *et al.*, “Measurement of long-range near-side two-particle angular correlations in pp collisions at  $\sqrt{s}=13$  TeV”, *Phys. Rev. Lett.* **116** (2016) 172302, [arXiv:1510.03068 \[nucl-ex\]](#).
- [116] J. Song, H.-h. Li, and F.-l. Shao, “New feature of low  $p_T$  charm quark hadronization in pp collisions at  $\sqrt{s} = 7$  TeV”, *Eur. Phys. J. C* **78** (2018) 344, [arXiv:1801.09402 \[hep-ph\]](#).
- [117] V. Minissale, S. Plumari, and V. Greco, “Charm hadrons in pp collisions at LHC energy within a coalescence plus fragmentation approach”, *Phys. Lett. B* **821** (2021) 136622, [arXiv:2012.12001 \[hep-ph\]](#).
- [118] A. Beraudo, A. De Pace, D. Pablos, F. Prino, M. Monteno, and M. Nardi, “Heavy-flavor transport and hadronization in pp collisions”, *Phys. Rev. D* **109** (2024) L011501, [arXiv:2306.02152 \[hep-ph\]](#).

- [119] K. Werner, “Core-corona separation in ultra-relativistic heavy ion collisions”, *Phys. Rev. Lett.* **98** (2007) 152301, [arXiv:0704.1270 \[nucl-th\]](#).
- [120] S. Porteboeuf and K. Werner, “Generation of complete events containing very high-p(T) jets”, *Eur. Phys. J. C* **62** (2009) 145–150.
- [121] M. He and R. Rapp, “Charm-Baryon Production in Proton-Proton Collisions”, *Phys. Lett. B* **795** (2019) 117–121, [arXiv:1902.08889 \[nucl-th\]](#).
- [122] M. He and R. Rapp, “Bottom Hadrochemistry in High-Energy Hadronic Collisions”, *Phys. Rev. Lett.* **131** (2023) 012301, [arXiv:2209.13419 \[hep-ph\]](#).
- [123] D. Ebert, R. N. Faustov, and V. O. Galkin, “Spectroscopy and Regge trajectories of heavy baryons in the relativistic quark-diquark picture”, *Phys. Rev. D* **84** (2011) 014025, [arXiv:1105.0583 \[hep-ph\]](#).
- [124] C. Bierlich, G. Gustafson, L. Lönnblad, and A. Tarasov, “Effects of Overlapping Strings in pp Collisions”, *JHEP* **03** (2015) 148, [arXiv:1412.6259 \[hep-ph\]](#).
- [125] T. Sjostrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual”, *JHEP* **05** (2006) 026, [arXiv:hep-ph/0603175](#).
- [126] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, “EPPS16: Nuclear parton distributions with LHC data”, *Eur. Phys. J. C* **77** (2017) 163, [arXiv:1612.05741 \[hep-ph\]](#).
- [127] **ALICE** Collaboration, K. Aamodt *et al.*, “The ALICE experiment at the CERN LHC”, *JINST* **3** (2008) S08002.
- [128] **ALICE** Collaboration, “ALICE upgrades during the LHC Long Shutdown 2”, [arXiv:2302.01238 \[physics.ins-det\]](#).
- [129] “LHC Machine”, *JINST* **3** (2008) S08001.
- [130] S. Myers, “The LEP collider, from design to approval and commissioning”,.
- [131] E. Lopienska, “The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022”,.  
<https://cds.cern.ch/record/2800984>. General Photo.
- [132] **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 \[hep-ex\]](#).
- [133] **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 \[hep-ex\]](#).

- [134] F. Sauli, “GEM: A new concept for electron amplification in gas detectors”, *Nucl. Instrum. Meth. A* **386** (1997) 531–534.
- [135] Apache Arrow, “Apache Arrow, a cross-language development platform for in-memory analytics.” <https://arrow.apache.org/>.
- [136] R. Brun and F. Rademakers, “ROOT: An object oriented data analysis framework”, *Nucl. Instrum. Meth. A* **389** (1997) 81–86.
- [137] **ALICE** Collaboration, G. Aglieri Rinella, “The ALPIDE pixel sensor chip for the upgrade of the ALICE Inner Tracking System”, *Nucl. Instrum. Meth. A* **845** (2017) 583–587.
- [138] W. Snoeys, “CMOS monolithic active pixel sensors for high energy physics”, *Nucl. Instrum. Meth. A* **765** (2014) 167–171.
- [139] S. Senyukov, J. Baudot, A. Besson, G. Claus, L. Cousin, A. Dorokhov, W. Dulinski, M. Goffe, C. Hu-Guo, and M. Winter, “Charged particle detection performances of CMOS pixel sensors produced in a  $0.18\mu\text{m}$  process with a high resistivity epitaxial layer”, *Nucl. Instrum. Meth. A* **730** (2013) 115–118, [arXiv:1301.0515 \[physics.ins-det\]](https://arxiv.org/abs/1301.0515).
- [140] **CMS** Collaboration, V. Karimäki, “The CMS tracker system project: Technical Design Report”,.
- [141] G. Aad *et al.*, “ATLAS pixel detector electronics and sensors”, *JINST* **3** (2008) P07007.
- [142] **LHCb** Collaboration, I. Bediaga, “LHCb VELO Upgrade Technical Design Report”,.
- [143] **ALICE** Collaboration, C. Lippmann, “Upgrade of the ALICE Time Projection Chamber”,.
- [144] A. Deisting, C. Garabatos, A. Szabo, and D. Vranic, “Measurements of ion mobility in argon and neon based gas mixtures”, *Nucl. Instrum. Meth. A* **845** (2017) 215–217, [arXiv:1603.07638 \[physics.ins-det\]](https://arxiv.org/abs/1603.07638).
- [145] **ALICE TPC** Collaboration, J. Adolfsson *et al.*, “The upgrade of the ALICE TPC with GEMs and continuous readout”, *JINST* **16** (2021) P03022, [arXiv:2012.09518 \[physics.ins-det\]](https://arxiv.org/abs/2012.09518).
- [146] **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 \[nucl-ex\]](https://arxiv.org/abs/1402.4476).
- [147] Y. Wang and Y. Yu, “Multigap Resistive Plate Chambers for Time of Flight Applications”, *Appl. Sciences* **11** (2020) 111.
- [148] **ALICE** Collaboration, S. Acharya *et al.*, “Measurement of D-meson production at mid-rapidity in pp collisions at  $\sqrt{s} = 7 \text{ TeV}$ ”, *Eur. Phys. J. C* **77** (2017) 550, [arXiv:1702.00766 \[hep-ex\]](https://arxiv.org/abs/1702.00766).

- [149] M. Concas, “Primary vertex reconstruction using GPUs for the upgrade of the Inner Tracking System of the ALICE experiment at LHC”, 2020.  
<https://cds.cern.ch/record/2878385>. Presented 11 Sep 2020.
- [150] **ALICE** Collaboration, D. Rohr, S. Gorbunov, M. O. Schmidt, and R. Shahoyan, “Track Reconstruction in the ALICE TPC using GPUs for LHC Run 3”, in *4th International Workshop Connecting The Dots 2018*. 11, 2018. [arXiv:1811.11481](https://arxiv.org/abs/1811.11481) [physics.ins-det].
- [151] R. Fruhwirth, “Application of Kalman filtering to track and vertex fitting”, *Nucl. Instrum. Meth. A* **262** (1987) 444–450.
- [152] D. Rohr, S. Gorbunov, M. O. Schmidt, and R. Shahoyan, “GPU-based Online Track Reconstruction for the ALICE TPC in Run 3 with Continuous Read-Out”, *EPJ Web Conf.* **214** (2019) 01050, [arXiv:1905.05515](https://arxiv.org/abs/1905.05515) [physics.ins-det].
- [153] **ZEUS** Collaboration, J. Breitweg *et al.*, “Measurement of inclusive D<sup>+</sup>-(s) photoproduction at HERA”, *Phys. Lett. B* **481** (2000) 213–227, [arXiv:hep-ex/0003018](https://arxiv.org/abs/hep-ex/0003018).
- [154] W. Blum, L. Rolandi, and W. Riegler, *Particle detection with drift chambers. Particle Acceleration and Detection*. 2008.
- [155] A. L. Samuel, “Some studies in machine learning using the game of checkers”, *IBM Journal of Research and Development* **3** (1959) 210–229.
- [156] T. M. Mitchell, “Machine learning”, 1997.
- [157] OpenAI, “GPT-4 Technical Report”, *arXiv e-prints* (Mar., 2023) [arXiv:2303.08774](https://arxiv.org/abs/2303.08774) [cs.CL].
- [158] P. I. Frazier, “A tutorial on bayesian optimization”, *arXiv preprint arXiv:1807.02811* (2018) .
- [159] J. Snoek, H. Larochelle, and R. P. Adams, “Practical bayesian optimization of machine learning algorithms”, *Advances in neural information processing systems* **25** (2012) .
- [160] J. Mockus, “The bayesian approach to global optimization”, in *System Modeling and Optimization: Proceedings of the 10th IFIP Conference New York City, USA, August 31–September 4, 1981*, pp. 473–481, Springer. 2005.
- [161] A. Mao, M. Mohri, and Y. Zhong, “Cross-entropy loss functions: Theoretical analysis and applications”, in *International Conference on Machine Learning*, pp. 23803–23828, PMLR. 2023.
- [162] J. Kiefer and J. Wolfowitz, “Stochastic Estimation of the Maximum of a Regression Function”, *The Annals of Mathematical Statistics* **23** (1952) 462 – 466. <https://doi.org/10.1214/aoms/1177729392>.

- [163] M. Stone, “Cross-validatory choice and assessment of statistical predictions”, *Journal of the royal statistical society: Series B (Methodological)* **36** (1974) 111–133.
- [164] GEANT4 Collaboration, S. Agostinelli *et al.*, “GEANT4—a simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250–303.
- [165] J. H. Friedman, “Greedy function approximation: a gradient boosting machine”, *Annals of statistics* (2001) 1189–1232.
- [166] 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** (1997) 119–139.
- [167] L. Breiman, *Classification and regression trees*. Routledge, 2017.
- [168] J. R. Quinlan, “Induction of decision trees”, *Machine learning* **1** (1986) 81–106.
- [169] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system”, *CoRR* **abs/1603.02754** (2016) , 1603.02754.  
<http://arxiv.org/abs/1603.02754>.
- [170] B. Kegl, CecileGermain, ChallengeAdmin, ClaireAdam, D. Rousseau, Djabbz, fradav, G. Cowan, Isabelle, and joycenv, “Higgs boson machine learning challenge”, 2014.  
<https://kaggle.com/competitions/higgs-boson>.
- [171] R. Mitchell and E. Frank, “Accelerating the xgboost algorithm using gpu computing”, *PeerJ Computer Science* **3** (2017) e127.
- [172] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow.* ” O'Reilly Media, Inc.”, 2022.
- [173] A. Natekin and A. Knoll, “Gradient boosting machines, a tutorial”, *Frontiers in neurorobotics* **7** (2013) 21.
- [174] XGBoost Documentation, “Xgboost parameters.”  
<https://xgboost.readthedocs.io/en/stable/parameter.html>.
- [175] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A next-generation hyperparameter optimization framework”, in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2623–2631. 2019.
- [176] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, “Algorithms for hyper-parameter optimization”, *Advances in neural information processing systems* **24** (2011) .
- [177] 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** (2001) 171–186.

- [178] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions”, *Advances in neural information processing systems* **30** (2017) .
- [179] S. Lipovetsky and M. Conklin, “Analysis of regression in game theory approach”, *Applied stochastic models in business and industry* **17** (2001) 319–330.
- [180] **ALICE** Collaboration, S. Acharya *et al.*, “Charm production and fragmentation fractions at midrapidity in pp collisions at  $\sqrt{s} = 13$  TeV”, *JHEP* **12** (2023) 086, [arXiv:2308.04877 \[hep-ex\]](https://arxiv.org/abs/2308.04877).
- [181] **ALICE** Collaboration, S. Acharya *et al.*, “Measurement of prompt  $D_s^+$ -meson production and azimuthal anisotropy in Pb–Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV”, *Phys. Lett. B* **827** (2022) 136986, [arXiv:2110.10006 \[nucl-ex\]](https://arxiv.org/abs/2110.10006).
- [182] F. Grossa, S. Politanò, and A. Bigot, “flarefly”, Jan., 2023. <https://doi.org/10.5281/zenodo.7579657>.
- [183] G. Cowan, *Statistical data analysis*. Oxford university press, 1998.
- [184] **ALICE** Collaboration, J. Adam *et al.*, “Centrality Dependence of the Charged-Particle Multiplicity Density at Midrapidity in Pb-Pb Collisions at  $\sqrt{s_{NN}} = 5.02$  TeV”, *Phys. Rev. Lett.* **116** (2016) 222302, [arXiv:1512.06104 \[nucl-ex\]](https://arxiv.org/abs/1512.06104).
- [185] **ALICE** Collaboration, S. Acharya *et al.*, “Pseudorapidity distributions of charged particles as a function of mid- and forward rapidity multiplicities in pp collisions at  $\sqrt{s} = 5.02, 7$  and 13 TeV”, *Eur. Phys. J. C* **81** (2021) 630, [arXiv:2009.09434 \[nucl-ex\]](https://arxiv.org/abs/2009.09434).
- [186] **WA82** Collaboration, M. Adamovich *et al.*, “Study of D+ and D- Feynman’s x distributions in pi- nucleus interactions at the SPS”, *Phys. Lett. B* **305** (1993) 402–406.
- [187] **E769** Collaboration, G. A. Alves *et al.*, “Forward cross-sections for production of D+, D0, D(s), D\*+ and Lambda(c) in 250-GeV pi+-, K+-, and p - nucleon interactions”, *Phys. Rev. Lett.* **77** (1996) 2388–2391. [Erratum: *Phys.Rev.Lett.* 81, 1537 (1998)].
- [188] E. Norrbin and T. Sjostrand, “Production mechanisms of charm hadrons in the string model”, *Phys. Lett. B* **442** (1998) 407–416, [arXiv:hep-ph/9809266](https://arxiv.org/abs/hep-ph/9809266).
- [189] E. Norrbin and T. Sjostrand, “Production and hadronization of heavy quarks”, *Eur. Phys. J. C* **17** (2000) 137–161, [arXiv:hep-ph/0005110](https://arxiv.org/abs/hep-ph/0005110).
- [190] **LHCb** Collaboration, R. Aaij *et al.*, “Measurements of prompt charm production cross-sections in pp collisions at  $\sqrt{s} = 13$  TeV”, *JHEP* **03** (2016) 159, [arXiv:1510.01707 \[hep-ex\]](https://arxiv.org/abs/1510.01707). [Erratum: *JHEP* 09, 013 (2016), Erratum: *JHEP* 05, 074 (2017)].

- [191] **LHCb** Collaboration, R. Aaij *et al.*, “Observation of strangeness enhancement with charmed mesons in high-multiplicity  $p\text{Pb}$  collisions at  $\sqrt{s_{\text{NN}}} = 8.16 \text{ TeV}$ ”, [arXiv:2311.08490 \[hep-ex\]](#).
- [192] **ALICE** Collaboration, S. Acharya *et al.*, “Charm-quark fragmentation fractions and production cross section at midrapidity in  $\text{pp}$  collisions at the LHC”, *Phys. Rev. D* **105** (2022) L011103, [arXiv:2105.06335 \[nucl-ex\]](#).
- [193] E. Braaten and R. D. Pisarski, “Soft Amplitudes in Hot Gauge Theories: A General Analysis”, *Nucl. Phys. B* **337** (1990) 569–634.
- [194] **HotQCD** Collaboration, L. Altenkort, O. Kaczmarek, R. Larsen, S. Mukherjee, P. Petreczky, H.-T. Shu, and S. Stendebach, “Heavy Quark Diffusion from 2+1 Flavor Lattice QCD with 320 MeV Pion Mass”, *Phys. Rev. Lett.* **130** (2023) 231902, [arXiv:2302.08501 \[hep-lat\]](#).
- [195] **ALICE** Collaboration, S. Acharya *et al.*, “Measurement of  $D^0$ ,  $D^+$ ,  $D^{*+}$  and  $D_s^+$  production in  $\text{Pb-Pb}$  collisions at  $\sqrt{s_{\text{NN}}} = 5.02 \text{ TeV}$ ”, *JHEP* **10** (2018) 174, [arXiv:1804.09083 \[nucl-ex\]](#).
- [196] **ALICE** Collaboration, J. Adam *et al.*, “Measurement of  $D_s^+$  production and nuclear modification factor in  $\text{Pb-Pb}$  collisions at  $\sqrt{s_{\text{NN}}} = 2.76 \text{ TeV}$ ”, *JHEP* **03** (2016) 082, [arXiv:1509.07287 \[nucl-ex\]](#).
- [197] **ALICE** Collaboration, J. Adam *et al.*, “Centrality dependence of the pseudorapidity density distribution for charged particles in  $\text{Pb-Pb}$  collisions at  $\sqrt{s_{\text{NN}}} = 5.02 \text{ TeV}$ ”, *Phys. Lett. B* **772** (2017) 567–577, [arXiv:1612.08966 \[nucl-ex\]](#).
- [198] **STAR** Collaboration, J. Adams *et al.*, “Experimental and theoretical challenges in the search for the quark gluon plasma: The STAR Collaboration’s critical assessment of the evidence from RHIC collisions”, *Nucl. Phys. A* **757** (2005) 102–183, [arXiv:nucl-ex/0501009](#).
- [199] **CMS** Collaboration, V. Khachatryan *et al.*, “Evidence for collectivity in  $\text{pp}$  collisions at the LHC”, *Phys. Lett. B* **765** (2017) 193–220, [arXiv:1606.06198 \[nucl-ex\]](#).
- [200] **CMS** Collaboration, V. Khachatryan *et al.*, “Observation of Long-Range Near-Side Angular Correlations in Proton-Proton Collisions at the LHC”, *JHEP* **09** (2010) 091, [arXiv:1009.4122 \[hep-ex\]](#).
- [201] **ALICE** Collaboration, S. Acharya *et al.*, “Investigations of Anisotropic Flow Using Multiparticle Azimuthal Correlations in  $\text{pp}$ ,  $\text{p-Pb}$ ,  $\text{Xe-Xe}$ , and  $\text{Pb-Pb}$  Collisions at the LHC”, *Phys. Rev. Lett.* **123** (2019) 142301, [arXiv:1903.01790 \[nucl-ex\]](#).
- [202] **ALICE** Collaboration, S. Acharya *et al.*, “Observation of a multiplicity dependence in the  $\text{pT}$ -differential charm baryon-to-meson ratios in proton–proton collisions at  $s=13 \text{ TeV}$ ”, *Phys. Lett. B* **829** (2022) 137065, [arXiv:2111.11948 \[nucl-ex\]](#).

- [203] S. Plumari, V. Minissale, S. K. Das, G. Coci, and V. Greco, “Charmed Hadrons from Coalescence plus Fragmentation in relativistic nucleus-nucleus collisions at RHIC and LHC”, *Eur. Phys. J. C* **78** (2018) 348, [arXiv:1712.00730 \[hep-ph\]](https://arxiv.org/abs/1712.00730).
- [204] F. Scardina, S. K. Das, V. Minissale, S. Plumari, and V. Greco, “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** (2017) 044905, [arXiv:1707.05452 \[nucl-th\]](https://arxiv.org/abs/1707.05452).
- [205] M. He and R. Rapp, “Hadronization and Charm-Hadron Ratios in Heavy-Ion Collisions”, *Phys. Rev. Lett.* **124** (2020) 042301, [arXiv:1905.09216 \[nucl-th\]](https://arxiv.org/abs/1905.09216).
- [206] L. Ravagli and R. Rapp, “Quark Coalescence based on a Transport Equation”, *Phys. Lett. B* **655** (2007) 126–131, [arXiv:0705.0021 \[hep-ph\]](https://arxiv.org/abs/0705.0021).
- [207] A. Andronic, P. Braun-Munzinger, M. K. Köhler, A. Mazeliauskas, K. Redlich, J. Stachel, and V. Vislavicius, “The multiple-charm hierarchy in the statistical hadronization model”, *JHEP* **07** (2021) 035, [arXiv:2104.12754 \[hep-ph\]](https://arxiv.org/abs/2104.12754).
- [208] A. Mazeliauskas, S. Floerchinger, E. Grossi, and D. Teaney, “Fast resonance decays in nuclear collisions”, *Eur. Phys. J. C* **79** (2019) 284, [arXiv:1809.11049 \[nucl-th\]](https://arxiv.org/abs/1809.11049).

# List of Figures

1.1	$e^+e^-$ annihilation to a fermion-antifermion pair. . . . .	2
1.2	$R$ as a function of $\sqrt{s}$ in the light-flavour, charm, and beauty threshold regions taken from Ref. [4]. The green curve is a naive quark-parton model prediction, while the red one is a 3-loops perturbative QCD prediction. Breit-Wigner parameterisations of $J/\psi$ , $\psi(2S)$ , and $\Upsilon(nS)$ , $n = 1, 2, 3, 4$ are also shown. . . . .	3
1.3	Feynman diagrams for gluon self-interactions. . . . .	4
1.4	Feynman diagram for quark-gluon interaction. . . . .	4
1.5	1-loop Feynman diagrams contributing to the $\beta$ function evaluation. . . . .	6
1.6	Summary of measurements of $\alpha_s$ as a function of the energy scale $Q$ , compared to the running of the coupling computed at five loops, taking as an input the current PDG average, $\alpha_s(M_Z^2) = 0.1180 \pm 0.0009 \text{ GeV}/c^2$ . Figure taken from Ref. [4]. . . . .	7
1.7	Feynman introduction to path integrals. Here, a particle emitted from a source at $x_a$ is detected at $x_b$ . A finite number of screens, each with a finite number of holes, is placed between the source and the detector. The probability amplitude for the particle to hit the detector is given by the sum of the probabilities of moving from the source to the detector through all possible paths. By adding an infinite amount of screens with an infinite number of holes, and by also considering the time at which the particle passes through each screen, the sum becomes an integral over all possible paths, called a <i>path integral</i> . . . . .	9
1.8	The light hadron mass spectrum of QCD. Horizontal lines and bands are the experimental values with their decay widths. Lattice QCD results [8] are shown by solid circles. Vertical error bars represent the combined statistical and systematic error estimates. $\pi$ , $K$ , and $\Xi$ have no error bars, because they are used to set the light quark mass, the strange quark mass, and the overall scale, respectively. . . . .	11
1.9	QCD phase diagram in the temperature-baryon chemical potential plane. Figure taken from Ref. [13]. . . . .	12
1.10	Space-time evolution of a heavy-ion collision. Figure taken from Ref. [20]. . . . .	13
1.11	Multiplicity per unit of rapidity of different hadron species and light nuclei measured by ALICE compared to SHM fits from THERMUS [28], SHARE [29], Thermal-FIST [30], and GSI-Heidelberg [31]. Differences between the model calculations and the measured yields are shown in the bottom panels. Figure taken from Ref. [16]. . . . .	16

1.12 Transverse momentum distributions of $\pi^+$ , $K^+$ , $p$ , $K_s^0$ , $\Lambda$ , and $\phi$ mesons for the 0–5% and 80–90% centrality intervals in Pb–Pb collisions at $\sqrt{s_{\text{NN}}} = 2.76$ TeV measured by the ALICE Collaboration. The data points are scaled by various factors for better visibility. Figure taken from Ref. [16]. . . . .	18
1.13 Results of blast-wave fits obtained by the ALICE Collaboration at the LHC, compared to similar fits at RHIC energies from the STAR Collaboration, for different centrality intervals. Figure taken from Ref. [34] . . . . .	19
1.14 Nuclear modification factors measured by ALICE in central (0 – 5%) and peripheral (70 – 80%) Pb–Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV and in p–Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV. Figure taken from Ref. [41]. . . . .	20
1.15 Top: Two-particle azimuthal correlations for minimum bias and central d–Au collisions, and for proton-proton collisions at $\sqrt{s_{\text{NN}}} = 200$ GeV measured by the STAR Collaboration. Bottom: Comparison of pedestal-subtracted two-particle azimuthal correlations for central Au–Au collisions to those seen in proton-proton collisions. Figure taken from Ref. [48]. . . . .	22
1.16 $p_{\text{T}}$ -integrated yield ratios of strange ( $K_s^0$ , $\Lambda$ ) and multi-strange ( $\Xi^\pm$ , $\Omega^\pm$ ) hadrons to pions ( $\pi^+ + \pi^-$ ) as a function of $\langle dN_{\text{ch}}/d\eta \rangle$ measured in pp, p–Pb and Pb–Pb collisions at midrapidity ( $ y  < 0.5$ ). Figure taken from Ref. [60]. . . . .	24
2.1 $p_{\text{T}}$ -differential production cross-section of prompt and non-prompt $D^0$ -mesons measured at midrapidity ( $ y  < 0.5$ ) in pp collisions by the ALICE Collaboration at $\sqrt{s} = 5.02$ TeV [62] compared to predictions obtained with FONLL calculations [63] combined with PYTHIA 8 [64] for the $H_b \rightarrow D^0 + X$ decay kinematics. Figure taken from Ref. [62]. . . . .	27
2.2 The proton structure function $F_2^p$ measured in electromagnetic scattering of electrons and positrons on protons, and for electrons/positrons and muons on a fixed target [4]. . . . .	29
2.3 Feynman diagrams contributing to the first order corrections of the heavy-flavour production cross-section calculations. . . . .	30
2.4 $\Lambda_c^+/D^0$ production-yield ratio measured at $\sqrt{s} = 5.02$ TeV by the ALICE Collaboration as a function of $p_{\text{T}}$ , compared to theoretical predictions. Figure taken from Ref. [91]. . . . .	32
2.5 Schematic representation of the Lund string model hadronisation process. . . . .	33
2.6 Representation of hadron production via recombination. . . . .	36
2.7 Prompt D-meson $R_{\text{AA}}$ in the 0–10% centrality class (left panel) and $v_2$ in the 30–50% centrality class (right panel) compared with predictions obtained with and without including hadronisation via recombination. Figure taken from Ref. [109]. . . . .	37

2.8	Left panel: $\Lambda_c^+/\text{D}^0$ production-yield ratio measured at midrapidity ( $ y  < 0.5$ ) in pp collisions at $\sqrt{s} = 5.02$ TeV by the ALICE Collaboration [91] as a function of $p_T$ , compared to theoretical predictions. Right panel: $\Xi_c^0/\text{D}^0$ production-yield ratio measured at midrapidity ( $-0.96 < y < 0.05$ ) in p–Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV by the ALICE Collaboration [91] as a function of $p_T$ , compared to theoretical predictions. Figures taken from Refs. [91] and [92]. . . . .	39
3.1	The Large Hadron Collider (LHC) accelerator complex at CERN. Figure taken from Ref. [131]. . . . .	42
3.2	The ALICE experimental apparatus. The top right panel shows a zoom of the ITS, FT0-A, FT0-C, and MFT detectors. Figure taken from the ALICE figure repository [60]. . . . .	44
3.3	Impact parameter resolution in $r\phi$ as a function of $p_T$ in pp collisions at $\sqrt{s} = 13.6$ TeV from Run 3 data compared with the same quantity measured in collisions at $\sqrt{s} = 13$ TeV from Run 2 data. Figure taken from ALICE figure repository [60]. . . . .	45
3.4	Schematic view of the ALICE ITS 2. Figure taken from Ref. [128]. . .	46
3.5	Schematic view of the ALICE TPC. Figure taken from Ref. [128]. . .	48
3.6	Schematic view of the ALICE TOF. Figure taken from Ref. [146]. . .	49
3.7	View of the FIT detectors illustrating the relative sizes of each component. From left to right, the FDD-A, FT0-A, FV0, FT0-C, and FDD-C are shown. The FT0-A and FV0 systems have a common mechanical support: the former is the small quadrangular structure in the centre of the large, circular FV0 support. The inset table lists the distance from the interaction point and the pseudorapidity coverage for each component. Figure taken from Ref. [128]. . . . .	50
3.8	Photograph of one half of the FV0 detector. Figure taken from Ref. [128]. . . . .	51
4.1	Sketch of the decay topology of $D_s^+$ and $D^+$ mesons into $\phi\pi^+ \rightarrow K^+K^-\pi^+$ for prompt (left) and non-prompt (right) $D_s^+$ meson. . . . .	53
4.2	Summary of the event selection criteria applied to the data sample. .	54
4.3	Schematic representation of the reconstruction workflow of $D_s^+$ and $D^+$ mesons. . . . .	55
4.4	Rapidity and transverse momentum distribution of reconstructed $D_s^+$ and $D^+$ mesons from MC simulations of pp collisions at $\sqrt{s} = 13.6$ TeV. The kinematical selections applied on the daughter tracks, requiring $ \eta^{\text{dau}}  < 0.8$ and $p_T^{\text{dau}} > 0.3$ GeV/c cause the steep decrease of the rapidity acceptance for the D-meson transverse momentum range $p_T \lesssim 2$ GeV. . . . .	57

---

4.5	Distributions of decay length for background (green), prompt (red), and non-prompt (blue) $D_s^+$ -meson signals (left panel) and for prompt $D_s^+$ - (red) and $D^+$ - (blue) meson signals (right panel) in pp collisions at $\sqrt{s} = 13.6$ TeV in the $4 < p_T < 6$ GeV/ $c$ D-meson transverse momentum interval. For prompt and non-prompt $D_s^+$ and $D^+$ mesons, the distributions are taken from Monte Carlo simulations, whereas for combinatorial background they are taken from the data sidebands.	59
4.6	Distributions of the cosine of the pointing angle (left panel) and its projection on the transverse plane (right panel) for $D_s^+$ mesons in pp collisions at $\sqrt{s} = 13.6$ TeV in the $4 < p_T < 6$ GeV/ $c$ D-meson transverse momentum interval. The distributions are shown for prompt $D_s^+$ mesons (red), non-prompt $D_s^+$ mesons (blue), and combinatorial background (green). For prompt and non-prompt $D_s^+$ mesons, the distributions are taken from Monte Carlo simulations, whereas for combinatorial background they are taken from the data sidebands.	60
4.7	Distributions of the projection of the impact parameter on the transverse plane for $D_s^+$ mesons (left panel) and for one of the like-sign daughter tracks (right panel) in pp collisions at $\sqrt{s} = 13.6$ TeV in the $4 < p_T < 6$ GeV/ $c$ D-meson transverse momentum interval. The distributions are shown for prompt $D_s^+$ mesons (red), non-prompt $D_s^+$ mesons (blue), and combinatorial background (green). For prompt and non-prompt $D_s^+$ mesons, the distributions are taken from Monte Carlo simulations, whereas for combinatorial background they are taken from the data sidebands.	61
4.8	Distributions of the difference between the reconstructed $K^+K^-$ invariant mass and the PDG mass of the $\phi$ meson (left panel) and the cosine cubed of the $K-\pi$ angle in the KK rest frame (right panel) for $D_s^+$ mesons in pp collisions at $\sqrt{s} = 13.6$ TeV in the $4 < p_T < 6$ GeV/ $c$ D-meson transverse momentum interval. The distributions are shown for prompt $D_s^+$ mesons (red), non-prompt $D_s^+$ mesons (blue), and combinatorial background (green). For prompt and non-prompt $D_s^+$ mesons, the distributions are taken from Monte Carlo simulations, whereas for combinatorial background they are taken from the data sidebands.	62
4.9	Specific energy loss distribution $dE/dx$ in the TPC for pp collisions at $\sqrt{s} = 13.6$ TeV. The curves obtained with a fit using the ALEPH parametrisation of the Bethe-Bloch formula for electrons, muons, pions, kaons, protons, deuterium, and tritium are also shown. Figure taken from the ALICE figure repository [60].	64
4.10	$\beta$ parameter measured with the TOF detector as a function of the momentum $p$ measured by the TPC detector in pp collisions at $\sqrt{s} = 13.6$ TeV. Figure taken from the ALICE figure repository [60].	65

---

5.1	Invariant mass distribution of prompt and non-prompt D mesons (red and blue, respectively), taken from MC simulations, and of the background candidates taken from real data used to train the ML model (green) in the $4 < p_T < 6 \text{ GeV}/c$ interval. Background candidates are selected in the $1.7 < M < 1.75 \text{ GeV}/c^2$ or $2.1 < M < 2.15 \text{ GeV}/c^2$ invariant-mass interval. . . . .	69
5.2	Correlation matrices of the features used to train the ML model for prompt $D_s^+$ (top-left), non-prompt $D_s^+$ (top-right), and background (bottom-left) candidates in the $2 < p_T < 3 \text{ GeV}/c$ interval. The correlation with the invariant mass and the transverse momentum is also reported. The Pearson coefficient is encoded in the colour of the cell, with red indicating a positive correlation, blue a negative correlation, and grey no linear correlation. . . . .	73
5.3	Confusion matrix for the BDT model trained in the $2 < p_T < 3 \text{ GeV}/c$ interval. Candidates are classified as the class with the highest score. . . . .	78
5.4	ROC curves for the model trained on the $2 < p_T < 3 \text{ GeV}/c$ interval. The One-vs-One ROC AUC metric is calculated for each class pair and reported in the legend. . . . .	80
5.5	Score distributions for the model trained on the $2 < p_T < 3 \text{ GeV}/c$ interval. The score distributions related to the probability of belonging to the background (top-left panel), prompt $D_s^+$ -meson (top-right panel), and non-prompt $D_s^+$ -meson (bottom-left panel) classes are shown. For each class, the score distribution is shown for candidates belonging to the different classes and to both the training (filled area) and test sets (markers). . . . .	81
5.6	Beeswarm-style SHAP feature importance plot for the prompt $D_s^+$ -meson score of the XGBoost model trained on the $2 < p_T < 3 \text{ GeV}/c$ interval. . . . .	83
5.7	Average $ \text{SHAP value} $ feature importance for the XGBoost model trained on the $2 < p_T < 3 \text{ GeV}/c$ (left panel) and $8 < p_T < 12 \text{ GeV}/c$ (right panel) intervals. The feature importance for the background, prompt $D_s^+$ -meson, and non-prompt $D_s^+$ -meson classes are shown with blue, orange, and green bars, respectively. . . . .	84
5.8	Statistical significance of the signal as a function of the selection criteria applied to the model output for the $2.0 < p_T < 2.5 \text{ GeV}/c$ interval. The chosen set of selection criteria is shown with a green cross. . . . .	86
5.9	Fit to the invariant mass distribution of selected candidates in the $1.5 < p_T < 2.0 \text{ GeV}/c$ interval. The fit function is shown as a solid line, while the signal and background components are shown as dashed lines. The background is modelled with a third-order polynomial function. . . . .	87
5.10	Invariant-mass distribution of simulated decays of $D^+$ mesons into $D^+ \rightarrow \pi^+ K^- \pi^+$ , where one of the pions produced in the decay is misidentified as a kaon. Distributions for different $p_T$ intervals are shown. . . . .	89

5.11 Left panel: peak width of the $D_s^+$ and $D^+$ mesons as a function of $p_T$ , before fixing the width of the $D^+$ meson to that of the $D_s^+$ meson divided by a factor of 1.2 for $p_T > 8 \text{ GeV}/c$ . Both the peak widths from fits to data and MC simulations are reported. Right panel: ratio of the peak widths of the $D_s^+$ and $D^+$ mesons as a function of $p_T$ , for both data and MC simulations. . . . .	90
5.12 Fit to the invariant mass distribution of selected candidate in the $2.0 < p_T < 2.5 \text{ GeV}/c$ interval. The total fit function is shown as a solid blue line, while background components are shown as solid red lines (correlated background) and dashed violet lines (dashed violet lines). The signal contributions are shown as filled green and azure areas for $D^+$ and $D_s^+$ mesons, respectively. The bottom panels show the distribution of the difference between the data and the background fit function. . . . .	91
5.13 Fit to the invariant mass distribution of selected candidate in the 14 studied $p_T$ intervals in the $0.5 < p_T < 24 \text{ GeV}/c$ range. The total fit function is shown as a solid blue line, while background components are shown as solid red lines (correlated background) and dashed violet lines (dashed violet lines). The signal contributions are shown as filled green and azure areas for $D^+$ and $D_s^+$ mesons, respectively. . . . .	92
5.14 Evolution of the mean (left panel) and width (right panel) of the Gaussian functions used to describe the $D_s^+$ and $D^+$ meson peaks as a function of $p_T$ . The dashed lines in the left panel represent the mass values reported in the PDG [4] for the two D-meson species. . . . .	93
5.15 Evolution of the $D^+ \rightarrow \pi^+ K^- \pi^+$ relative contribution to the fitting function (left panel) and statistical significance of the extracted $D_s^+$ - and $D^+$ -meson signal (right panel) as a function of $p_T$ . . . . .	93
5.16 Results from the multi-trial approach employed for estimating the systematic uncertainty related to the raw yield extraction in the $1.0 < p_T < 1.5 \text{ GeV}/c$ interval. . . . .	95
6.1 Impact-parameter resolution in the transverse plane (left panel) and beam direction (right panel) of primary particles in data and MC. Figure taken from the ALICE figure repository [60]. . . . .	98
6.2 Acceptance-times-efficiency correction factor as a function of $p_T$ for prompt and non-prompt $D_s^+$ and $D^+$ mesons. . . . .	99
6.3 Results for the evaluation of the $D_s^+$ $f_{\text{prompt}}$ correction factor in the $1.5 < p_T < 2.0 \text{ GeV}/c$ interval. The correlation factor $\rho$ (top-left panel), $(\text{Acc} \times \varepsilon)$ factors for both prompt and non-prompt $D_s^+$ mesons (top-right panel), prompt and non-prompt fraction in the extracted $D_s^+$ raw yields (bottom-left panel), and contribution of prompt and non-prompt $D_s^+$ mesons to the extracted raw yields (bottom-right panel) are shown for the different BDT selection criteria applied on the non-prompt probability. . . . .	103

6.4	Left panel: $f_{\text{prompt}}$ correction factor for $D_s^+$ (blue) and $D^+$ (orange) mesons as a function of $p_T$ . The results obtained with the data-driven (filled markers) method are compared with those exploiting a theory-driven approach (void markers). Right panel: corrected $f_{\text{prompt}}$ factors for $D_s^+$ (blue) and $D^+$ (orange) mesons as a function of $p_T$ , obtained with the data-driven method. . . . .	104
6.5	Results for the evaluation of the systematic uncertainty on the BDT selection efficiency for the $4.5 < p_T < 5.0 \text{ GeV}/c$ $p_T$ interval. In the top row, the raw yields of $D_s^+$ (blue) and $D^+$ (orange) mesons, the prompt (filled markers) and non-prompt (void markers) $D_s^+$ selection efficiency, the prompt and non-prompt $D^+$ selection efficiency, and the prompt fraction of $D_s^+$ and $D^+$ mesons are shown for the different BDT selections used to estimate the systematic uncertainty, normalised to their respective central values. In the bottom row, the statistical significance of the $D_s^+$ and $D^+$ extracted signals, the signal-over-background ratio for the two meson species, and the $D_s^+/D^+$ production yield ratio are reported for the different BDT selections. A red band is shown in this last panel around the central value, whose width represents the RMS of the measurements. The rightmost bottom panel shows the distribution of the $D_s^+/D^+$ production yield ratio for the different BDT selections divided by the central value. A red band is shown around the central value, representing the sum in quadrature of the standard deviation of the $D_s^+/D^+$ distribution and of the difference between its mean and the default value. The blue band displays the assigned systematic uncertainty. . . . .	106
6.6	Results for the evaluation of the systematic uncertainty on a possible imperfect description of the tracking performance in MC simulations. The $D_s^+/D^+$ production yield ratio is shown in the left panel for the central result (red) and for the two different configurations of the <b>track-tuner</b> : with a factor 2 worse $p_T$ resolution (blue) and a 10% worse resolution on the track impact parameter (distance of closest approach (DCA), orange). In the right panel, the ratio of the different $D_s^+/D^+$ factors to that obtained in the central case is reported. . . . .	107
6.7	Results for the evaluation of the systematic uncertainty on the $f_{\text{prompt}}$ correction. In the left panel, the different $f_{\text{prompt}}^{D_s^+}/f_{\text{prompt}}^{D^+}$ corrections obtained with different sets of BDT selections are shown. In the right panel, their ratio to that obtained in the central case is reported. . . . .	109
6.8	Comparison of the $p_T$ distributions of generated $D_s^+$ (left panel) and $D^+$ (right panel) mesons in the MC simulations used for evaluating the $(\text{Acc} \times \varepsilon)$ corrections. The distributions obtained with the PYTHIA 8 event generator are shown with filled markers, while those obtained with FONLL are shown with void markers. . . . .	110
7.1	$p_T$ -differential $D_s^+/D^+$ production-yield ratio measured in this Thesis at midrapidity ( $ y  < 0.5$ ) in pp collisions at $\sqrt{s} = 13.6 \text{ TeV}$ with the ALICE experimental apparatus. . . . .	113

7.2	Charged-particle multiplicity density, normalised to the number of nucleons participating in the collision divided by a factor 2 ( $\langle N_{\text{part}} \rangle / 2$ ) as a function of the centre-of-mass energy per nucleon pair of the collision for central nucleus-nucleus, proton-nucleus, and pp collisions. The $s_{\text{NN}}$ -dependence of the charged-particle multiplicity density is parametrised using a power-law function, and is shown with solid and dashed lines for central nucleus-nucleus collisions and smaller collision systems, respectively. Figure taken from Ref. [184] . . . . .	114
7.3	$p_{\text{T}}$ -differential $D_s^+ / D^+$ production-yield ratio measured at midrapidity ( $ y  < 0.5$ ) in pp collisions by the ALICE Collaboration. The results obtained in this Thesis at $\sqrt{s} = 13.6$ TeV (red) are compared with previous measurements in pp collisions at $\sqrt{s} = 5, 7$ , and $13$ TeV (blue, green, and orange, respectively). . . . .	115
7.4	$y$ -differential charm-quark production cross-section in pp collisions at $\sqrt{s} = 13.6$ TeV predicted by FONLL [63] calculations. . . . .	117
7.5	$p_{\text{T}}$ -differential $D_s^+ / D^+$ production-yield ratio measured at midrapidity ( $ y  < 0.5$ ) in pp collisions at $\sqrt{s} = 13.6$ TeV by the ALICE Collaboration (red), obtained in this Thesis, compared with previous measurement in pp collisions at $\sqrt{s} = 13$ TeV performed by the LHCb Collaboration in the forward-rapidity intervals of $2.0 < y < 2.5$ (blue) and $2.5 < y < 3.0$ (green) in the left panel and the $3.0 < y < 3.5$ (blue), $3.5 < y < 4.0$ (green), and $4.0 < y < 4.5$ (orange) rapidity intervals in the right panel. . . . .	118
7.6	$p_{\text{T}}$ -differential $D_s^+ / D^+$ production-yield ratio measured at midrapidity ( $ y  < 0.5$ ) in pp collisions at $\sqrt{s} = 13.6$ TeV by the ALICE Collaboration. The results obtained in this Thesis (red) are compared with different theoretical predictions. . . . .	119
8.1	$D_s^+$ -meson signal statistical significance as a function of the selection criteria for the $2 < p_{\text{T}} < 4$ GeV/ $c$ interval in the 30–50% centrality class. . . . .	125
8.2	Normalised invariant mass distribution of the selected D-meson candidates in the $2 < p_{\text{T}} < 4$ GeV/ $c$ (top-left panel) and $6 < p_{\text{T}} < 8$ GeV/ $c$ (bottom-left panel) intervals of the 10–30% centrality class and in the $4 < p_{\text{T}} < 6$ GeV/ $c$ (top-right panel) and $8 < p_{\text{T}} < 12$ GeV/ $c$ (bottom-right panel) intervals of the 30–50% centrality class. . . . .	127
8.3	Width of the $D_s^+$ -meson signal peak as a function of $p_{\text{T}}$ measured in Pb–Pb collisions at $\sqrt{s_{\text{NN}}} = 5.36$ TeV in the 30–50% centrality class and pp collisions at $\sqrt{s} = 13.6$ TeV recorded during the LHC Run 3 data-taking period (blue and red markers, respectively) and in the 30–50% centrality class of Pb–Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV recorded during the LHC Run 2 data-taking period (green markers). . . . .	129
9.1	Strange over non-strange $D_s^+ / D^0$ production-yield ratio as a function of $p_{\text{T}}$ for two different multiplicity classes measured at midrapidity ( $ y  < 0.5$ ) in proton-proton collisions at $\sqrt{s} = 13$ TeV by the ALICE Collaboration [202]. Figure taken from the ALICE figure repository [60].	132

9.2	$D_s^+ / D^0$ production-yield ratio as a function of $p_T$ in pp collisions and in the 10% most central Pb–Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV compared to different model calculations. Figure taken from Ref. [16]. . .	133
A.1	Systematic uncertainties on the raw-yield extraction for the $0.5 < p_T < 1 \text{ GeV}/c$ (top left), $1 < p_T < 1.5 \text{ GeV}/c$ (top right), $1.5 < p_T < 2 \text{ GeV}/c$ (bottom left) and $2 < p_T < 2.5 \text{ GeV}/c$ (bottom right) intervals. . . . .	136
A.2	Systematic uncertainties on the raw-yield extraction for the $2.5 < p_T < 3 \text{ GeV}/c$ (top left), $3 < p_T < 3.5 \text{ GeV}/c$ (top right), $3.5 < p_T < 4 \text{ GeV}/c$ (bottom left) and $4 < p_T < 4.5 \text{ GeV}/c$ (bottom right) intervals. . . . .	137
A.3	Systematic uncertainties on the raw-yield extraction for the $4.5 < p_T < 5 \text{ GeV}/c$ (top left), $5 < p_T < 5.5 \text{ GeV}/c$ (top right), $5.5 < p_T < 6 \text{ GeV}/c$ (bottom left) and $6 < p_T < 8 \text{ GeV}/c$ (bottom right) intervals. . . . .	138
A.4	Systematic uncertainties on the raw-yield extraction for the $8 < p_T < 12 \text{ GeV}/c$ (left) and $12 < p_T < 24 \text{ GeV}/c$ (right) intervals. . . . .	139
B.1	Results for the evaluation of the $f_{\text{prompt}}^{D_s^+}$ correction factor in the $0.5 < p_T < 1.0 \text{ GeV}/c$ (top left), $1.0 < p_T < 1.5 \text{ GeV}/c$ (top right), $1.5 < p_T < 2.0 \text{ GeV}/c$ (bottom left) and $2.0 < p_T < 2.5 \text{ GeV}/c$ (bottom right) intervals. . . . .	141
B.2	Results for the evaluation of the $f_{\text{prompt}}^{D_s^+}$ correction factor in the $2.5 < p_T < 3.0 \text{ GeV}/c$ (top left), $3.0 < p_T < 3.5 \text{ GeV}/c$ (top right), $3.5 < p_T < 4.0 \text{ GeV}/c$ (bottom left) and $4.0 < p_T < 4.5 \text{ GeV}/c$ (bottom right) intervals. . . . .	142
B.3	Results for the evaluation of the $f_{\text{prompt}}^{D_s^+}$ correction factor in the $4.5 < p_T < 5.0 \text{ GeV}/c$ (top left), $5.0 < p_T < 5.5 \text{ GeV}/c$ (top right), $5.5 < p_T < 6.0 \text{ GeV}/c$ (bottom left) and $6.0 < p_T < 8.0 \text{ GeV}/c$ (bottom right) intervals. . . . .	143
B.4	Results for the evaluation of the $f_{\text{prompt}}^{D_s^+}$ correction factor in the $8.0 < p_T < 12.0 \text{ GeV}/c$ (top) and $12.0 < p_T < 24.0 \text{ GeV}/c$ (bottom) intervals. . . . .	144
B.5	Results for the evaluation of the $f_{\text{prompt}}^{D_s^+}$ correction factor in the $0.5 < p_T < 1.0 \text{ GeV}/c$ (top left), $1.0 < p_T < 1.5 \text{ GeV}/c$ (top right), $1.5 < p_T < 2.0 \text{ GeV}/c$ (bottom left) and $2.0 < p_T < 2.5 \text{ GeV}/c$ (bottom right) intervals. . . . .	145
B.6	Results for the evaluation of the $f_{\text{prompt}}^{D_s^+}$ correction factor in the $2.5 < p_T < 3.0 \text{ GeV}/c$ (top left), $3.0 < p_T < 3.5 \text{ GeV}/c$ (top right), $3.5 < p_T < 4.0 \text{ GeV}/c$ (bottom left) and $4.0 < p_T < 4.5 \text{ GeV}/c$ (bottom right) intervals. . . . .	146
B.7	Results for the evaluation of the $f_{\text{prompt}}^{D_s^+}$ correction factor in the $4.5 < p_T < 5.0 \text{ GeV}/c$ (top left), $5.0 < p_T < 5.5 \text{ GeV}/c$ (top right), $5.5 < p_T < 6.0 \text{ GeV}/c$ (bottom left) and $6.0 < p_T < 8.0 \text{ GeV}/c$ (bottom right) intervals. . . . .	147
B.8	Results for the evaluation of the $f_{\text{prompt}}^{D_s^+}$ correction factor in the $8.0 < p_T < 12.0 \text{ GeV}/c$ (top) and $12.0 < p_T < 24.0 \text{ GeV}/c$ (bottom) intervals. . . . .	148

C.1	Results for the evaluation of the systematic uncertainty on the BDT selection efficiency in the $0.5 < p_T < 1.0 \text{ GeV}/c$ (top left), $1.0 < p_T < 1.5 \text{ GeV}/c$ (top right), $1.5 < p_T < 2.0 \text{ GeV}/c$ (bottom left) and $2.0 < p_T < 2.5 \text{ GeV}/c$ (bottom right) intervals. . . . .	150
C.2	Results for the evaluation of the systematic uncertainty on the BDT selection efficiency in the $2.5 < p_T < 3.0 \text{ GeV}/c$ (top left), $3.0 < p_T < 3.5 \text{ GeV}/c$ (top right), $3.5 < p_T < 4.0 \text{ GeV}/c$ (bottom left) and $4.0 < p_T < 4.5 \text{ GeV}/c$ (bottom right) intervals. . . . .	151
C.3	Results for the evaluation of the systematic uncertainty on the BDT selection efficiency in the $4.5 < p_T < 5.0 \text{ GeV}/c$ (top left), $5.0 < p_T < 5.5 \text{ GeV}/c$ (top right), $5.5 < p_T < 6.0 \text{ GeV}/c$ (bottom left) and $6.0 < p_T < 8.0 \text{ GeV}/c$ (bottom right) intervals. . . . .	152
C.4	Results for the evaluation of the systematic uncertainty on the BDT selection efficiency in the $8.0 < p_T < 12.0 \text{ GeV}/c$ (top) and $12.0 < p_T < 24.0 \text{ GeV}/c$ (bottom) intervals. . . . .	153

# List of Tables

3.1	Main layout parameters of the new ITS2. Table taken from Ref. [128].	47
4.1	Applied single-track selection criteria.	58
5.1	Topological and kinematical pre-selections applied to the $D_s^+$ - and $D^+$ -mesons candidates. The selection criterion on the $n\sigma_{TOF}$ variable is applied only when a TOF cluster is matched to the track, i.e., tracks with no TOF signal are not rejected. In addition, a logical OR is applied to the conditions on the $n\sigma_{TOF}$ and $n\sigma_{TPC}$ PID variables, meaning that a candidate is accepted if, for all daughter tracks, at least one of the conditions is satisfied.	70
5.2	Number of candidates within the $p_T$ intervals used to train and test the model.	71
5.3	Candidate features used to train the ML model.	72
5.4	Optimised hyperparameter configuration for some representative $p_T$ intervals considered in the model training.	77
5.5	Selection criteria applied to enhance the significance of the $D_s^+$ meson signal in the $p_T$ intervals considered for the analysis.	86
5.6	Systematic uncertainty on the raw-yield extraction for the ratio of the $D_s^+$ and $D^+$ mesons yields.	96
6.1	Summary of the assigned systematic uncertainties on the $D_s^+/D^+$ production yield ratio.	111
8.1	Variation range and number of steps for the variables used in the optimisation of the selection criteria for the selection of $D_s^+$ - and $D^+$ -meson candidates in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.36$ TeV.	124
8.2	Optimal selection criteria for the selection of $D_s^+$ meson candidates in the 10–30% centrality class.	125
8.3	Optimal selection criteria for the selection of $D_s^+$ meson candidates in the 30–50% centrality class.	126

# Ringraziamenti

Con questa Tesi si conclude un capitolo importante del mio percorso accademico. Un percorso lungo, che ha richiesto impegno, dedizione e sacrifici, ma che ha anche portato molte soddisfazioni. Nel raggiungere questo traguardo, ho avuto la fortuna di essere supportato da tante persone che mi hanno sostenuto, e a cui vorrei dedicare queste pagine di ringraziamenti.

Vorrei innanzitutto ringraziare il mio relatore, Francesco Prino, per avermi guidato in questo percorso di ricerca, fornendomi sempre preziosi consigli e supportando la mia crescita come ricercatore. Soprattutto, un grazie per la pazienza e la disponibilità nel rispondere in modo sempre puntuale a tutte le mie domande, e per avermi trasmesso la passione per la fisica nucleare delle alte energie. Ringrazio anche i miei correlatori, Fabrizio e Stefano, per tutti i consigli e l'aiuto fornito, per il supporto che mi hanno dato alla stesura di questa Tesi, ma anche per tutti i "caffettoni", le partite a padel, e i momenti di svago dentro e fuori l'ufficio.

Grazie anche a tutti i colleghi del gruppo ALICE Torino, e in particolare dello "stanzone": i tre Luca, Giacomo, Giulia e Mario, con cui ho condiviso le mie giornate lavorative, ma anche tante risate. Un ringraziamento va anche al gruppo ALICE CERN per i due mesi passati insieme: il mio compagno d'ufficio Mario, Fabio, Mattia, Isabella, Ivan, Matteo, Max, Francesco, Chiara, e Andrea. Grazie per aver reso il mio soggiorno piacevole e per l'ambiente di lavoro stimolante che avete creato.

Un pensiero va anche a tutti coloro che hanno sostenuto la mia crescita professionale, trasmettendomi la passione per la scienza e per la ricerca: tutti i professori, ricercatori di ALICE e non, e il gruppo di lavoro D2H, grazie per tutti i suggerimenti che mi avete fornito e per le conoscenze trasmesse.

Un grazie speciale va ai miei compagni di avventura, che da ormai cinque anni mi accompagnano in questo percorso: Vittorio, Lorenzo, Arianna, Luca, Sara, e i miei compagni di laboratorio Marcello, Roberta, e Giorgio. Mi avete supportato nei momenti difficili e durante miei studi, ma mi avete sempre garantito anche tanti momenti di svago e di spensieratezza.

Ai miei cari amici Elena, Matteo, Elisa e Claudio, mille grazie per tutte le uscite fatte insieme, per avermi permesso di staccare la spina e di avermi permesso di confidarmi con voi. Spero in tanti altri momenti da passare insieme.

Un ringraziamento speciale va alla mia famiglia, per avermi sempre supportato e sostenuto in ogni mia scelta, e per avermi trasmesso i valori che mi hanno permesso di crescere e di diventare la persona che sono oggi. Uno speciale pensiero va a mia cugina, o meglio, alla mia "sorellina" Michela, che mi ha sempre (e che ho sempre) sopportato, e che voglio congratulare per il traguardo da poco raggiunto.

Grazie a mia madre e a mio padre, che mi hanno lasciato libero di studiare ciò

che più mi piace, che mi hanno sempre spinto a dare il massimo, a cui devo tutto quello che sono.

Infine, l'ultima dedica va ad Alessia, la mia stella polare da ormai quattro anni. Grazie per essere stata il mio rifugio nei momenti difficili e scintilla di gioia in quelli felici. Grazie per avermi supportato e sostenuto in ogni mia scelta, per avermi sempre spronato a dare il meglio di me stesso, e per avermi sempre fatto sentire speciale. Spero di essere riuscito a fare altrettanto per te.

*Grazie!*