Charmed-strange meson production in proton-proton collisions with the ALICE experiment at the LHC

Arthur Gal

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DE PHYSIQUE ET CHIMIE-PHYSIQUE

Institut Pluridisciplinaire Hubert Curien (IPHC), UMR 7178

THÈSE

présentée par :

Arthur Gal

soutenue le : **3 décembre 2021**

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Discipline/ Spécialité : Physique

Charmed-strange meson production in proton-proton collisions with the ALICE experiment at the LHC

THÈSE dirigée par :

M. BELIKOV Iouri — Directeur de recherche, IPHC, Strasbourg

M. RAMI Fouad — Chargé de recherche, IPHC, Strasbourg

RAPPORTEURS :

M. GRANIER DE CASSAGNAC Raphaël — Directeur de recherche, Laboratoire Leprince-Ringuet, Palaiseau

M. STASZEL Pawel — Professeur, Jagiellonian University, Institute of Physics, Cracovie

---

AUTRES MEMBRES DU JURY :

Mme LE BIHAN Anne-Catherine — Chargée de recherche, IPHC, Strasbourg

M. GOSSIAUX Pol-Bernard — Professeur, Laboratoire de physique subatomique et des technologies associées, Nantes

Table of contents

Résumé étendu

Introduction

La chromodynamique quantique, théorie décrivant l'interaction forte entre les partons (quarks et gluons), prédit une transition de phase de la matière hadronique où les partons initialement confinés au sein d'hadrons, comme le proton ou le neutron, se déconfinent perdant ainsi leur identité d'origine et formant un plasma de quarks et de gluons (PQG). Cet état de la matière, qui aurait existé quelques micro-secondes après le Big Bang, est étudié expérimentalement dans les collisions ultra-relativistes de noyaux atomiques lourds. Les collisions ultra-relativistes engendrent des diffusions à grands transferts d'impulsion permettant ainsi la production de quarks lourds (quarks charmés et beaux). Dans le plasma de quarks et de gluons, les quarks lourds peuvent interagir avec ses constituants par processus collisionnels et radiatifs conduisant à des pertes d'énergies. La modélisation de ces pertes d'énergie permet de caractériser certaines propriétés du PQG, comme les coefficients de transport. Dans ces conditions, les quarks lourds représentent donc des sondes privilégiées pour étudier les propriétés du milieu. La durée de vie du PQG étant très courte, de l'ordre de 10^{-23} s, l'étude expérimentale de ses propriétés s'effectue de manière indirecte via la détection des hadrons produits lors de la collision. Parmi les hadrons charmés, les mésons D représentent près de 90% de la population de quarks charmés produite dans les collisions hadroniques.

La caractérisation du plasma de quarks et de gluons, observé dans les collisions de noyaux lourds, nécessite l'étude rigoureuse des collisions, proton-proton et proton-noyaux, afin de découpler les contributions PQG et non-PQG aux mécanismes de production des hadrons. Dans les collisions proton-proton, la production des hadrons charmés permet l'étude du contenu partonique des protons, de la production des quarks lourds et de leur hadronisation qui sont des processus se réalisant à différentes échelles d'énergie. L'étude des collisions proton-proton, au-delà de leur fonction de système de référence, est motivée par l'observation récente d'effets collectifs de type PQG dans les petits systèmes à haute multiplicité de particules produites, telle que l'augmentation de la production de l'étrangeté [1]. L'étude de la production des mésons charmés-étranges D_s^+, méson composé d'un quark charmé et d'un quark étrange, permet de contribuer à une meilleure compréhension de l'origine des effets collectifs observés dans les petits systèmes, via le prisme de l'augmentation de l'étrangeté.

Cadre et objectifs de la thèse

La thèse a été effectuée dans le cadre de l'expérience ALICE (A Large Ion Collider Experiment) dédiée à l'étude de la chromodynamique quantique et particulièrement du plasma de quarks et de gluons auprès du collisioneur LHC (Large Hadron Collider) du CERN. Elle porte sur l'étude de la production du méson charmé-étrange D_s^+ dans les collisions proton-proton à une énergie dans le centre de masse $\sqrt{s} = 13$ TeV en analysant les données collectées par le détecteur ALICE pendant le run II du LHC (mars 2015 - décembre 2018). Le run II du LHC a permis d'enregistrer environ 2 milliards de collisions à cette énergie. La disponibilité de ces données de haute statistique permet de réaliser, pour la première fois, des mesures détaillées de la production des mésons D_s^+ en fonction de la multiplicité en particules chargées de la collision.
L'objet de la thèse est de fournir des mesures de la production des mésons D_s^+.

- une mesure de la section efficace différentielle de production, $\mathrm{d}^2\sigma/(\mathrm{d}p_\mathrm{T}\,\mathrm{d}y)$, des mésons $\mathrm{D_s^+}$ à rapidité centrale ($|y| < 0.5$).
- une mesure des taux de production des mésons $\mathrm{D_s^+}$ en fonction de la multiplicité en particules chargées de la collision.

Travaux réalisés

La faible distance de désintégration du méson $\mathrm{D_s^+}$, $c\tau \simeq 151.2$ μm oblige d'effectuer sa mesure de manière indirecte à partir de ses produits de désintégration. Dans le canal étudié, $\mathrm{D_s^+} \to \phi(1020)(\to \mathrm{K^+K^-})\pi^+$, le détecteur ALICE mesure trois particules filles, un pion ($\pi^\pm$) et une paire de kaons ($\mathrm{K^\pm}$). L'une des parties critiques de l'analyse des données consiste en la sélection du signal parmi l'importante quantité de bruit de fond combinatoire. Cette sélection s'appuie sur la cinématique, la topologie de désintégration du méson $\mathrm{D_s^+}$ et l'identification des particules filles.

L'approche standard de sélection du signal consiste en l'application successive de conditions de sélection uniques déterminées à partir d'une étude basée sur des simulations Monte Carlo. La limite principale de cette approche vient du fait qu'elle tire peu parti des potentielles corrélations entre les différentes variables de sélection. Une méthode plus avancée, employant des arbres de décision boostés (BDT, un algorithme de machine learning) a été développée afin d'optimiser les performances de sélection du signal et de rejet du bruit de fond combinatoire. Cette méthode consiste à utiliser un algorithme identifiant les règles de décision, prenant la forme d'opérations conditionnelles "if-then-else", à partir des caractéristiques des données afin de prédire de manière probabiliste la nature des mésons $\mathrm{D_s^+}$ (signal ou bruit de fond). La représentation graphique des ces règles de décision prend la forme d'une structure en forme d'arbre. Le modèle BDT, un ensemble d'arbres de décision, ainsi construit de manière supervisée peut effectuer une tâche de classification binaire sur les candidats mésons $\mathrm{D_s^+}$. La prédiction du modèle pour un candidat est un nombre flottant compris entre 0 et 1. Les valeurs proches de 1 (0) sont associées à une probabilité haute d'être un candidat signal (bruit de fond). La distribution des prédictions du modèle BDT utilisé pour les mésons $\mathrm{D_s^+}$ dans l'intervalle en impulsion transverse $2 < p_\mathrm{T} < 4$ GeV/c est présentée sur la figure 1. Au total, six modèles BDT sont entraînés et utilisés couvrant l'intervalle $1 < p_\mathrm{T} < 24$ GeV/c.

L'objectif de la phase d'entraînement des modèles est d'extraire les caractéristiques générales et corrélations sous-jacentes des données d'entraînement afin de les réutiliser sur de nouvelles données du même type. Cette capacité est appelée "généralisation". Un soin particulier a donc été apporté à la construction des échantillons de données d'entraînement afin qu'ils soient représentatifs de l'ensemble des données traitées. La phase d'entraînement des modèles est exposée au compromis dit de "biais-variance". Le biais correspond à la distance entre la prédiction du modèle et la "vraie" valeur (0 ou 1). Un modèle hautement biaisé est un signe de sous-apprentissage, les caractéristiques pertinentes n'ont pas été saisies due à une description trop simpliste des données. La variance caractérise la sensibilité du modèle aux fluctuations présentes dans les données. Un modèle avec une haute variance est susceptible d'être sur-entraîné, il décrit ainsi les fluctuations aléatoires et non les caractéristiques générales des données due à une description trop complexe. La capacité de généralisation des modèles entraînés à été contrôlée en mesurant les performances de sélection des mésons $\mathrm{D_s^+}$ pour différents échantillons de données. La méthode BDT a été validée par comparaison des résultats à différentes étapes de l'analyse avec ceux obtenus avec la méthode standard considérée comme référence.

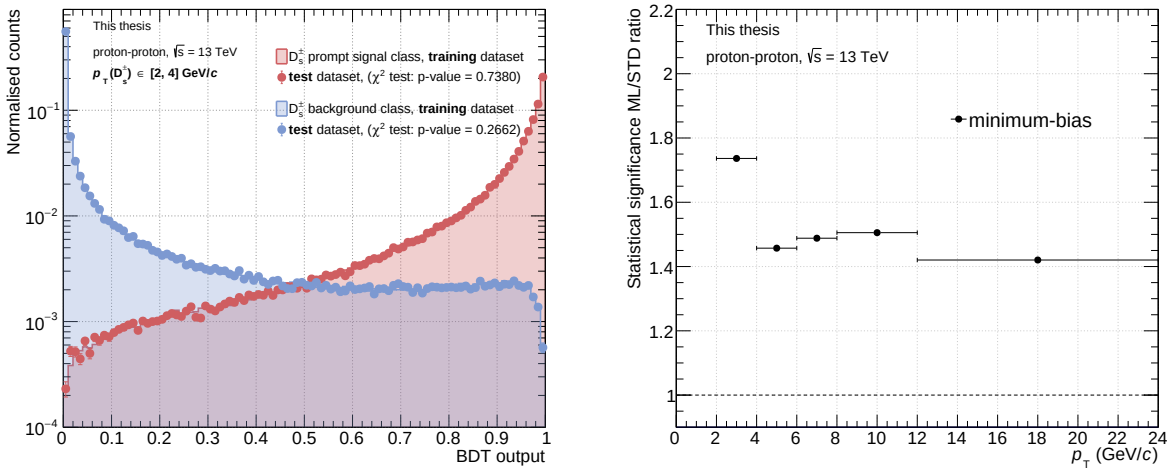


FIGURE 1 – Panneau gauche : distribution des prédictions du modèle BDT pour les candidats ayant une impulsion transverse entre $2 < p_{\mathrm{T}} < 4$ GeV/c issus de l'échantillon de données d'entraînement et de validation du modèle. Panneau droit : comparaison de la significance statistique obtenue avec les candidats sélectionnés par le modèle BDT et avec la méthode standard de sélection pour différents intervalles en p_{T}.

Une part importante du travail présenté a consisté à developper l'environnement d'analyse de données dans lequel s'insère cet algorithme et d'optimiser ses paramètres. L'utilisation des modèles BDT a permis une amélioration significative de l'efficacité de sélection des mésons $\mathrm{D_s^+}$, notamment aux basses impulsions transverses, p_{T}, où le rapport signal sur bruit est le plus faible. La significance statistique du signal ainsi obtenue a été augmentée de façon significative (entre 40% - 70% selon l'intervalle en p_{T}) par rapport à la méthode standard. Le rapport des significances obtenues avec les deux méthodes en fonction de l'impulsion transverse des mésons $\mathrm{D_s^+}$ est montré sur le panneau droit de la figure 1.

La sélection des candidats permet de reconstruire leur distribution en masse invariante comme présenté sur le panneau gauche de la figure 2. La distribution est ajustée par la somme de deux fonctions gaussiennes décrivant les pics des mésons $\mathrm{D_s^+}$ et des mésons $\mathrm{D^+}$ se désintégrant dans le même canal ainsi qu'une fonction exponentielle décroissante décrivant le bruit de fond. Le nombre brut de particules signal mesurées est extrait en intégrant la fonction d'ajustement sur une plage de $\pm 3\ \sigma$ autour de la masse du méson $\mathrm{D_s^+}$.

Ce nombre est ensuite corrigé pour tenir compte de l'acceptance géométrique du détecteur et de l'efficacité de reconstruction des mésons $\mathrm{D_s^+}$. Le facteur d'acceptance dépend uniquement de la cinématique de désintégration des mésons $\mathrm{D_s^+}$. Il est évalué à partir de simulations Monte Carlo où la rapidité et l'impulsion transverse des mésons $\mathrm{D_s^+}$ générés sont tirées aléatoirement suivant les prédictions du modèle FONLL [2,3]. Le calcul de l'efficacité de reconstruction nécessite des simulations plus approfondies dans lesquelles les particules sont propagées dans le détecteur ALICE dont la géométrie et la réponse sont prises en compte. L'évolution de l'efficacité en fonction de l'impulsion transverse des mésons $\mathrm{D_s^+}$ dans les collisions proton-proton à $\sqrt{s} = 13$ TeV est montrée sur le panneau droit de la figure 2. L'efficacité de reconstruction des mésons $\mathrm{D_s^+}$ augmente avec leur impulsion transverse, variant entre ∼1% et ∼50%. En effet, la distance de vol des mésons $\mathrm{D_s^+}$ est plus importante à haute impulsion, le vertex de désintégration s'éloigne du point de collision et sa position devient mieux contrainte.

Le nombre brut de mésons $\mathrm{D_s^+}$ reconstruits est composé de mésons produits au point

de collision ainsi que ceux issus de la désintégration de particules plus lourdes (hadrons beaux). La fraction de ces derniers a été estimée par une méthode basée sur les prédictions FONLL [2, 3] de la section efficace de production des hadrons beaux. La cinématique de désintégration des hadrons beaux en mésons D_s^+ a été effectuée avec PYTHIA 8 [4].

Le nombre corrigé de mésons D_s^+ ainsi obtenu a permis, après normalisation, d'extraire une mesure de la section efficace différentielle de production ainsi que la mesure des taux de production pour différents intervalles en multiplicité de particules chargées produites lors de la collision.

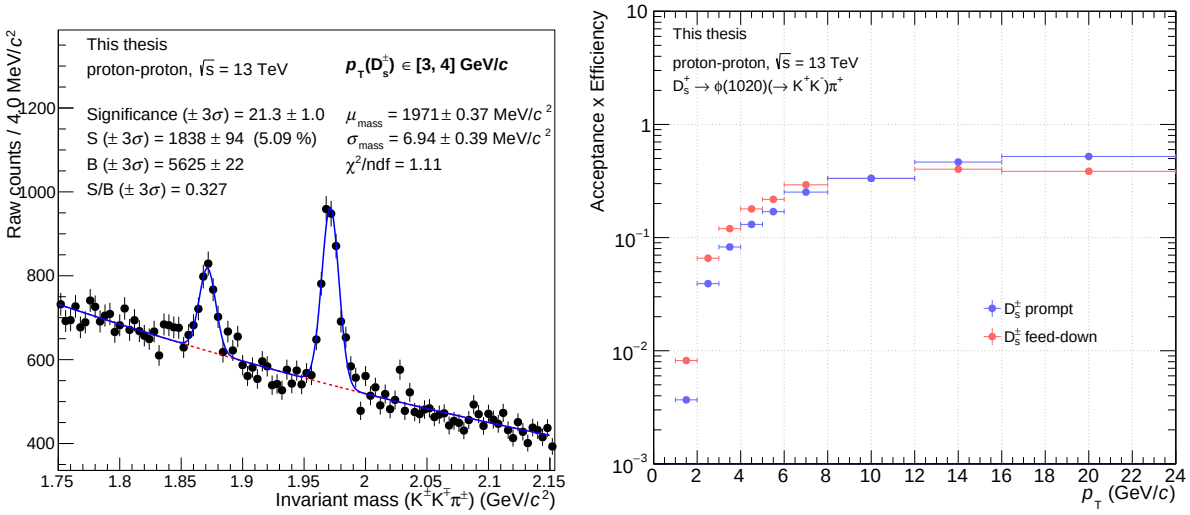


FIGURE 2 – Panneau gauche : histogramme de masse invariante des mésons D_s^+ pour l'intervalle $3 < p_T < 4$ GeV/c. Panneau droit : Produit du facteur d'acceptance et du facteur l'efficacité de reconstruction des mésons D_s^+ en fonction de leur impulsion transverse p_T.

Une étude détaillée a été menée pour évaluer les incertitudes systématiques associées aux imperfections du détecteur et aux méthodes de reconstruction et d'analyse. Les principales sources d'incertitudes sont liées aux méthodes de sélection des candidats D_s^+, d'extraction du signal brut, de reconstruction et sélection des trajectoires des particules dans le détecteur et la soustraction de la contribution des mésons provenant de désintégrations des hadrons beaux. L'exploration des différentes sources d'incertitudes et de leurs potentielles corrélations a représenté une partie importante du travail d'analyse.

Principaux résultats

Section efficace de production des mésons D_s^+

La mesure de la section efficace différentielle de production des mésons D_s^+ en fonction de leur impulsion transverse dans les collisions proton-proton à $\sqrt{s} = 13$ TeV est présentée sur la figure 3. La comparaison avec des mesures effectuées précédemment à $\sqrt{s} = 5.02$ TeV [5] et $\sqrt{s} = 7$ TeV [6] montre une augmentation de la production des mésons D_s^+ avec l'énergie de collision. La luminosité intégrée enregistrée à $\sqrt{s} = 13$ TeV, $\mathcal{L}_{\rm int} = 31.7 \pm 0.51$ nb^{-1}, supérieure comparée aux autres énergies de collision ($\mathcal{L}_{\rm int}(7\text{ TeV}) = 6 \pm 0.21$ nb^{-1}, $\mathcal{L}_{\rm int}(5.02\text{ TeV}) = 19.3 \pm 0.40$ nb^{-1}) permet d'étendre la mesure aux faibles impulsions transverses (0 - 2 GeV/c)

avec des incertitudes statistiques réduites comme montré sur le panneau inférieur gauche de la figure 3.

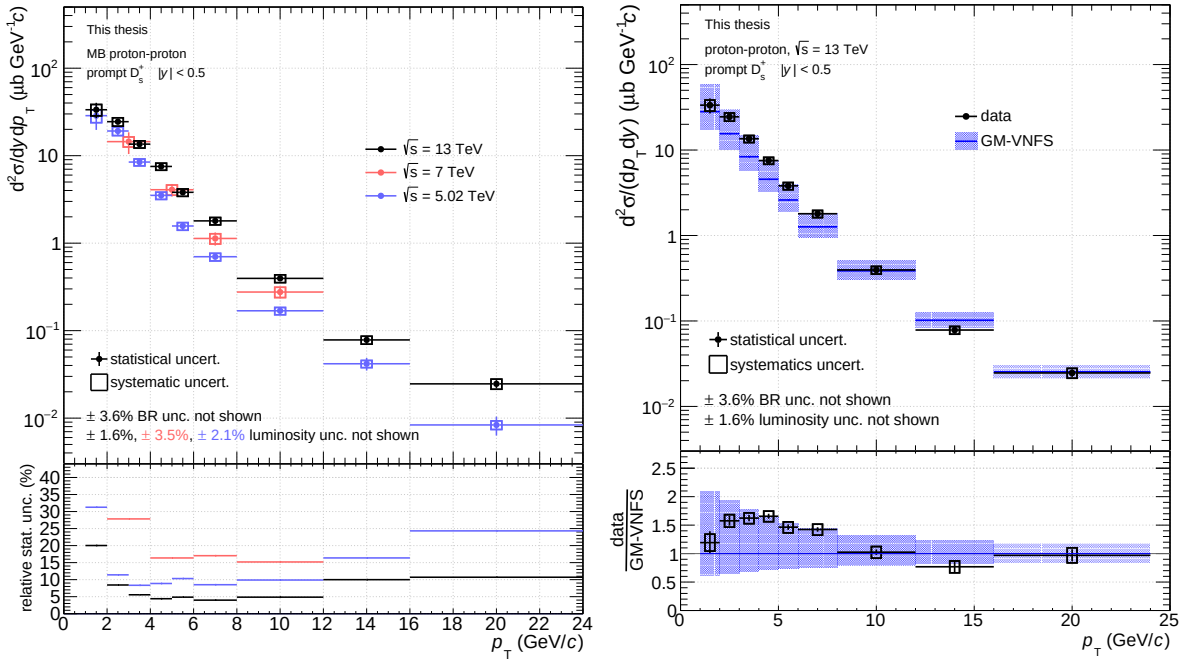


FIGURE 3 – Panneau gauche : section efficace différentielle de production des mésons D_s^+ dans les collisions proton-proton à $\sqrt{s} = 13$ TeV, 7 TeV [6] et 5.02 TeV [5]. Panneau droit : comparaison des résultats aux prédictions des calculs GM-VNFS [7–9].

Sur le panneau droit de la figure 3, la section efficace obtenue dans ce travail est comparée avec les prédictions GM-VNFS basées sur des calculs de QCD perturbative [7–9]. Dans ces calculs, l'expression générale de la section efficace différentielle de production des mésons D_s^+ est donnée par le théorème de factorisation [10] permettant de réunir la description du contenu partonique des protons incidents (fonctions de distribution des partons), la production des quarks charmés ainsi que son hadronisation (fonction de fragmentation). La production des quarks charmés dans les processus de diffusions à grands transferts d'impulsion est décrite par la QCD perturbative [11]. Les descriptions du contenu partonique initial et de l'hadronisation des quarks charmés sont extraites à partir d'analyses de données expérimentales [12–15].
Dans la limite des incertitudes, les prédictions GM-VNFS sont compatibles avec les résultats expérimentaux. Néanmoins, aux impulsions transverses intermédiaires, la mesure se situe systématiquement sur la partie supérieure de l'incertitude des prédictions. Par rapport à la précision de la mesure expérimentale, les incertitudes systématiques des prédictions sont larges, en particulier à bas p_T. Le développement perturbatif de la section efficace de production du quark charmé, restreint aux premiers ordres (n<3), limite la précision des prédictions. Les incertitudes systématiques associées sont réduites en construisant les rapports de la section efficace des mésons D_s^+ sur celle des mésons D ainsi que les rapports de la section efficace des mésons D_s^+ à différentes énergies de collision. En effet, les corrélations entre les différentes sources d'incertitudes systématiques pour ces rapports entraînent une réduction de l'incertitude totale permettant une meilleure contrainte des prédictions par la mesure expérimentale.

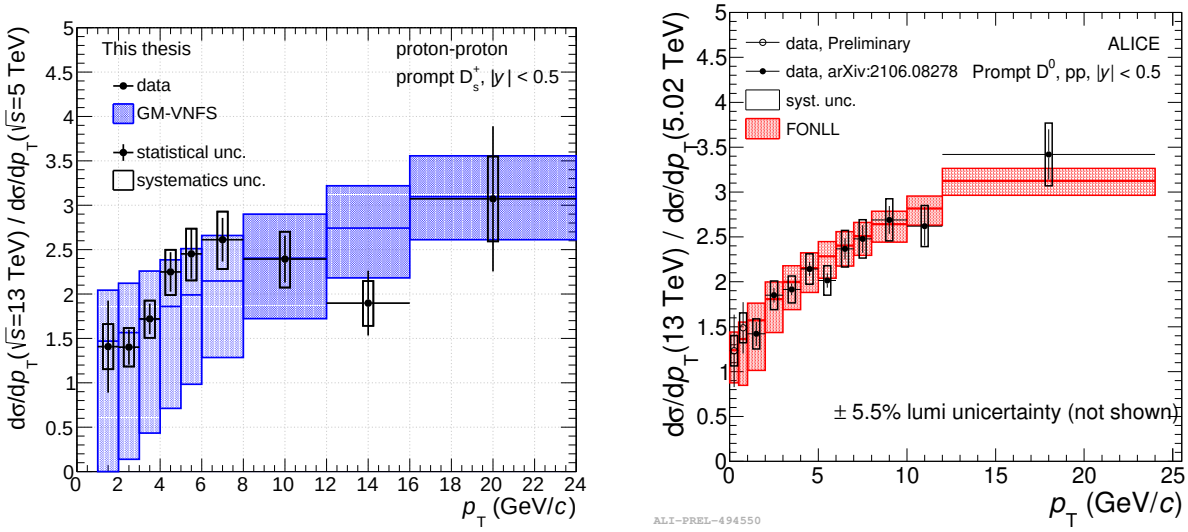


FIGURE 4 – Rapport de la section efficace de production des mésons D_s^+, mesurée dans ce travail à $\sqrt{s} = 13$ TeV, sur celle obtenue précédemment à une énergie $\sqrt{s} = 5.02$ TeV (panneau gauche). Le panneau droit montre le même rapport pour pour des mésons D^0. Les mesures sont comparées aux prédictions des calculs GM-VNFS [7–9] et FONLL [2,3,16].

Le rapport de la section efficace de production des mésons D_s^+ à $\sqrt{s} = 13$ TeV sur celle obtenue précédemment à $\sqrt{s} = 5.02$ TeV est présenté sur le panneau gauche de la figure 4 et comparé avec les prédictions des calculs GM-VFNS [7–9]. Ce même rapport pour les mésons D^0, comparé aux prédictions des calculs FONLL [2,3,16], est présenté sur le panneau droit. Les deux rapports, supérieurs à l'unité, augmentent avec l'impulsion transverse de la particule indiquant une production plus importante de mésons D_s^+ et D^0 de haute impulsion transverse à $\sqrt{s} = 13$ TeV par rapport à $\sqrt{s} = 5.02$ TeV. Dans la limite des incertitudes, les deux mesures sont compatibles avec les calculs de QCD pertubative.

Les contributions d'ordres supérieurs à la section efficace de production des quarks charmés (n>3) ne dépendent pas fortement de l'énergie de la collision $\sqrt{s}$ [16]. Les calculs perturbatifs à deux énergies différentes sont donc fortement corrélés. Pour les prédictions GM-VNFS le degré de correlation n'est pas connu, il a donc été choisi d'assigner l'incertitude relative minimum entre les prédictions à $\sqrt{s} = 13$ TeV et $\sqrt{s} = 5.02$ TeV. En revanche, pour les prédictions FONLL, dont les calculs sont détaillés dans [16], cette corrélation est considérée dans la propagation des incertitudes du rapport. Notamment, les incertitudes liées aux contributions d'ordres supérieurs s'annulent et l'incertitude totale est désormais dominée par les incertitudes liées aux fonctions de distribution de partons des protons. Les rapports de la section efficace des mésons D_s^+ à différentes énergies de collision permettent donc de contraindre la description de l'état initial de la collision pour ce type de calculs.

La mesure des rapports de la section efficace des mésons D_s^+ sur celle des mésons D non-étranges permet d'étudier la description du mécanisme de fragmentation du quark charmé. La fraction de fragmentation du quark charmé en mésons D_s^+, f_s, décrit la probabilité qu'un quark charmé fragmente en un méson D_s^+. Le rapport des fractions de fragmentation $f_s/(f_u + f_d)$ estimé à partir du rapport des sections efficaces de production $D_s^+/(D^0 + D^+)$ mesuré à $\sqrt{s} = 13$ TeV est présenté sur le panneau gauche de la figure 5. La valeur du rapport à $\sqrt{s} = 13$ TeV est extraite en ajustant une fonction constante sur la distribution (figure 5) sous l'hypothèse que le rapport est constant sur l'intervalle en impulsion transverse mesuré. Le

résultat de l'ajustement donne,

$$\frac{f_s}{f_u + f_d} = 0.140 \pm 0.004(\text{stat})$$

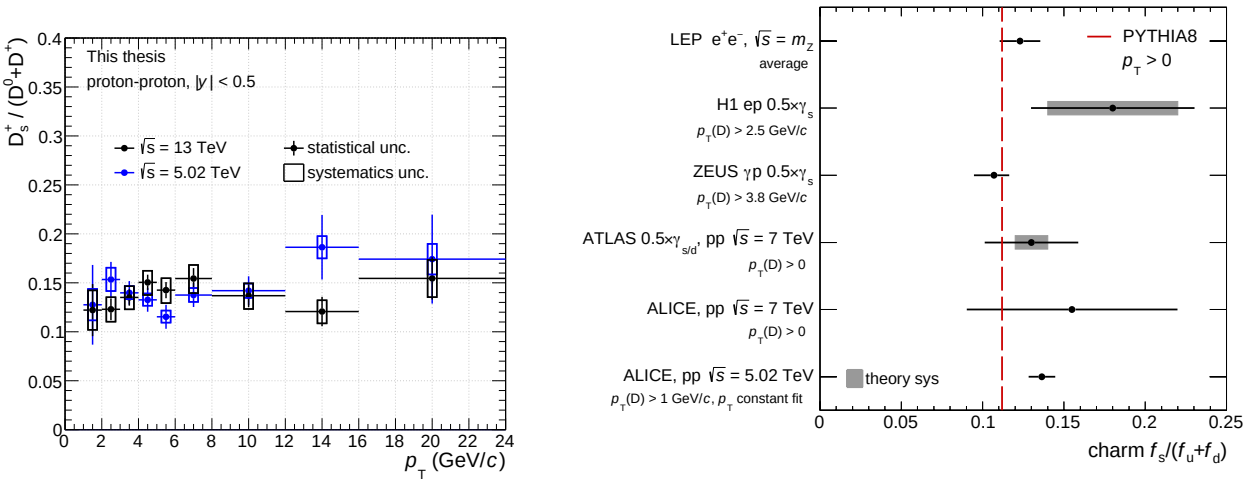


FIGURE 5 – Panneau gauche : Rapport de la section efficace de production des mésons D_s^+ sur la somme des sections efficaces des mésons D^0 et D^+ dans les collisions proton-proton à $\sqrt{s} = 13$ TeV et $\sqrt{s} = 5.02$ TeV [5]. Panneau droit : rapport des fractions de fragmentation du quark charmé $f_s/(f_u+f_d)$ mesurées par les expériences ALICE [5], [17], ATLAS [18], H1 [19] et ZEUS [20] ainsi que la moyenne des mesures du LEP [21]. La figure est extraite de [5].

Différentes mesures du facteur de suppression de l'étrangeté, $\gamma_s = 2\ f_s/(f_u + f_d)$, réalisées pour les collisions proton-proton au LHC (ALICE [5], [17] et ATLAS [18]), pour les collisions proton-électron (H1 [19] et ZEUS [20]) ainsi que pour les collisions électron-positron (LEP [21]) sont résumées sur le panneau droit de la figure 5. La comparaison avec la valeur extraite dans le présent travail pour les collisions proton-proton à $\sqrt{s} = 13$ TeV montre que les mesures sont compatibles pour différents systèmes de collision et à différentes énergies suggérant l'universalité de la fragmentation du quark charmé en mésons D.

Production des mésons D_s^+ en fonction de la multiplicité de la collision

La figure 6 montre le taux de production des mésons D_s^+ en fonction de l'impulsion transverse pour différents intervalles en multiplicité de particules chargées de la collision. La multiplicité de la collision exprimée par la densité moyenne de particules chargées produite à rapidité centrale $\langle \mathrm{d}N_{\mathrm{ch}}/\mathrm{d}\eta \rangle_{|\eta|<0.5}$, est déterminée en utilisant deux estimateurs expérimentaux différents. Sur le panneau gauche de la figure, la multiplicité ($N_{tracklets}$) est estimée à rapidité centrale ($|\eta| < 1$) alors que sur le panneau droit, la multiplicité (V0M) est estimée dans les regions dites "à l'avant" (-3.7 $< \eta <$ -1.7 et 2.8 $< \eta <$ 5.1). La mesure couvre de 0.5 à 5.5 fois la valeur de la multiplicité moyenne ($\langle \mathrm{d}N_{\mathrm{ch}}/\mathrm{d}\eta \rangle_{|\eta|<0.5} = 6.9$). Les résultats montrent une augmentation de la production des mésons D_s^+ avec la multiplicité de la collision. Le taux de production est deux ordres de grandeur plus élevé à haute multiplicité qu'à basse multiplicité. Cette observation est en accord avec les mesures effectuées pour les autres mésons D à $\sqrt{s} = 7$ TeV [22] montrant la corrélation entre la production des mésons D_s^+ et la multiplicité de la collision. De plus, une augmentation d'environ 32% de l'impulsion transverse moyenne

$\langle p_{\mathrm{T}} \rangle$ des mésons $\mathrm{D}_{\mathrm{s}}^{+}$ est observée entre l'intervalle le plus bas et le plus haut en multiplicité. Cette augmentation est également observé dans le secteur des hadrons beaux [23] et des hadrons légers [24].

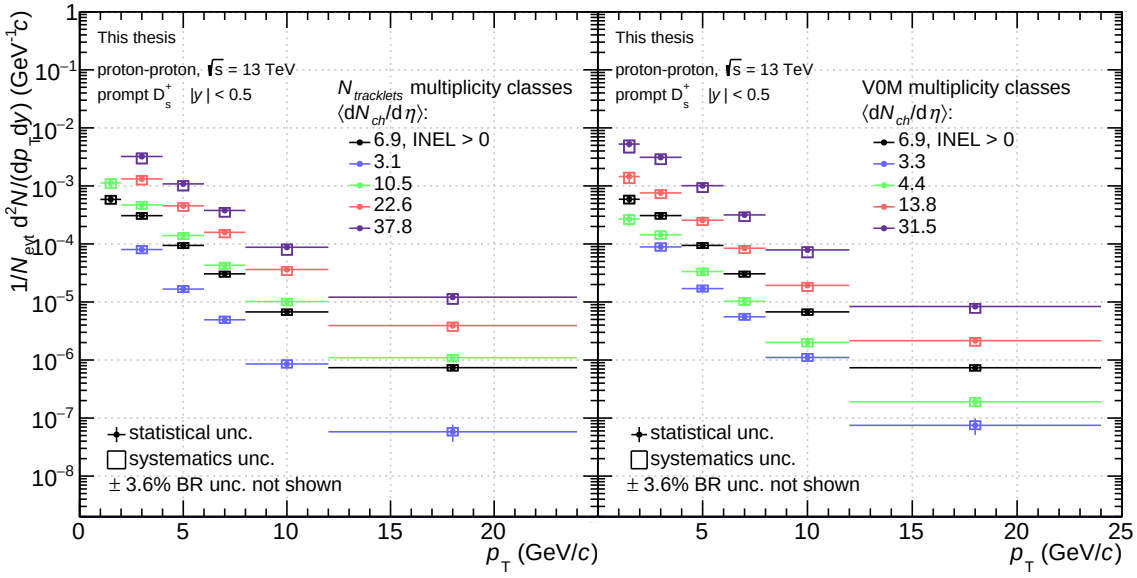


FIGURE 6 – Taux de production des mésons $\mathrm{D}_{\mathrm{s}}^{+}$ en fonction de l'impulsion transverse pour différents intervalles en multiplicité de particules chargées dans les collisions proton-proton à $\sqrt{s} = 13$ TeV. Les résultats sont montrés pour deux estimateurs de la multiplicité (voir texte pour plus de détails).

Le taux de production normalisés des mésons $\mathrm{D}_{\mathrm{s}}^{+}$ en fonction de la multiplicité relative de particules chargées est présenté sur la figure 7 pour différents intervalles en impulsion transverse. Le taux de production, $\mathrm{d}^2N/\mathrm{d}p_{\mathrm{T}}\mathrm{d}y$, est normalisé par la densité de particule moyenne $\langle \mathrm{d}N_{\mathrm{ch}}/\mathrm{d}\eta \rangle_{|\eta|<0.5} = 6.9$. La multiplicité relative est définie par $\mathrm{d}N_{ch}/\mathrm{d}\eta / \langle \mathrm{d}N_{\mathrm{ch}}/\mathrm{d}\eta \rangle_{|\eta|<0.5}$ avec $\mathrm{d}N_{ch}/\mathrm{d}\eta$ la densité moyenne de particules chargées dans un intervalle en multiplicité. La mesure montre une augmentation de la production des mésons $\mathrm{D}_{\mathrm{s}}^{+}$ plus importante par rapport à une augmentation linéaire montrée par la droite en pointillé sur la figure 7. Le quark charmé est produit dans des processus à grands transferts d'impulsion engendrant de fortes émissions de gluons, ce qui se traduit par une production élevée de particules chargées. La production des mésons $\mathrm{D}_{\mathrm{s}}^{+}$, et plus généralement des hadrons lourds, est donc corrélée à la multiplicité de la collision. Si l'estimation de la multiplicité est effectuée dans le même intervalle que celui de la reconstruction des mésons $\mathrm{D}_{\mathrm{s}}^{+}$, dont la production est issue de la fragmentation du quark charmé, la mesure est exposée à un biais dit "d'auto-corrélations". En effet, un biais peut intervenir lors de la classification des collisions puisqu'un certain nombre de particules chargées sont susceptibles d'être produites dans la direction de propagation du méson $\mathrm{D}_{\mathrm{s}}^{+}$. La reconstruction des mésons $\mathrm{D}_{\mathrm{s}}^{+}$ étant faite dans la région centrale $|\eta| < 0.5$, l'utilisation de deux estimateurs de la multiplicité dans deux régions de rapidité différentes (centrale et "à l'avant") permet d'étudier ces potentiels biais. La figure 7 montre la même tendance dans l'évolution des taux de production en fonction de la multiplicité pour les deux estimateurs. La même tendance est observée pour les mésons D non-étranges et le méson J/Ψ [22].

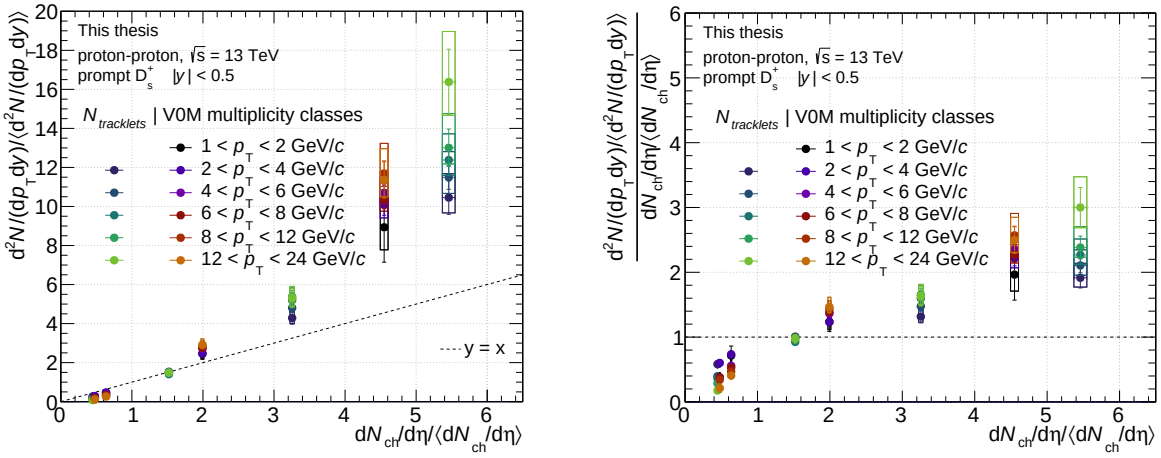


FIGURE 7 – Taux de production normalisé des mésons D_s^+ en fonction de la multiplicité de particules chargées dans les collisions proton-proton à $\sqrt{s} = 13$ TeV. Les résultats sont montrés pour différents intervalles en impulsion transverse. Sur le panneau droit, le taux normalisé est divisé par la densité relative de particules chargées.

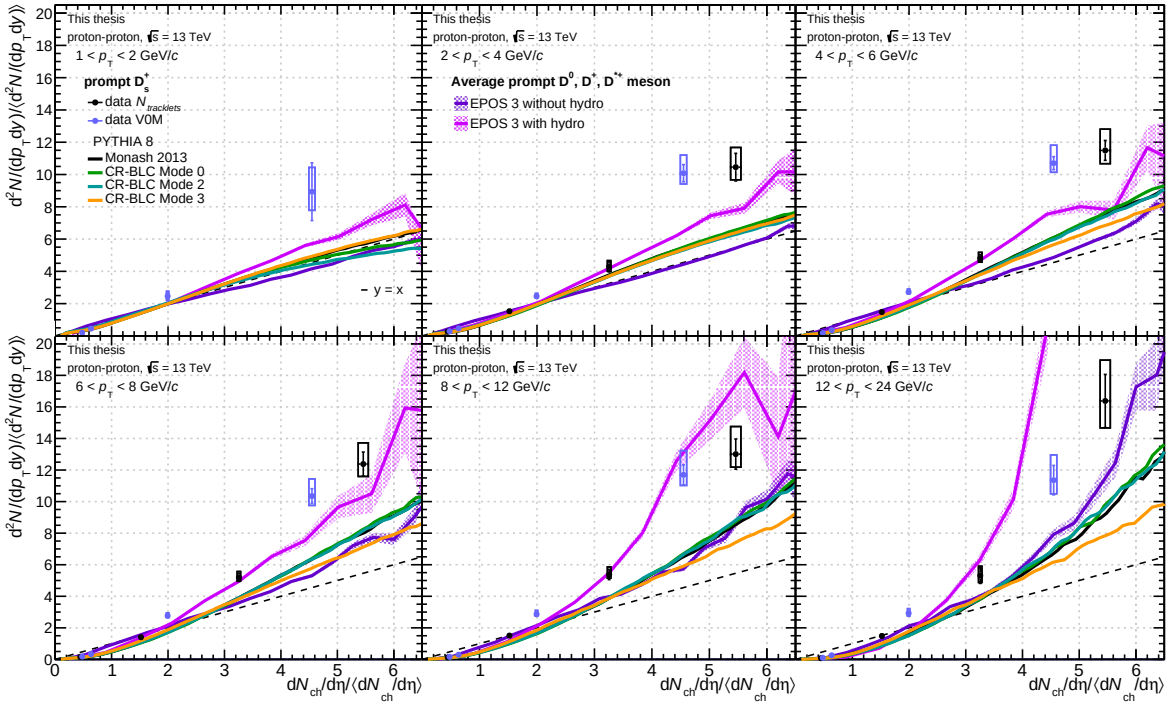


FIGURE 8 – Taux de production normalisé des mésons D_s^+ en fonction de la multiplicité de particules chargées pour différents intervalles d'impulsion transverse dans les collisions proton-proton à $\sqrt{s} = 13$ TeV. La mesure est comparée aux prédictions des modèles PYTHIA 8 [4, 25] et EPOS 3 [26, 27].

La description de la production des mésons D en fonction de la multiplicité est examinée dans le cadre des modèles PYTHIA 8 [4, 25] et EPOS 3 [26, 27]. Les prédictions issues

de PYTHIA 8 pour les paramétrisations Monash 13 [4] et CR-BLC [25] sont présentées sur la figure 8 et comparées aux résultats expérimentaux. Les deux modèles décrivent les interactions partoniques multiples se produisant dans les collisions proton-proton. Le nombre d'interactions partoniques est proportionnel au nombre de particules chargées produites à rapidité centrale dans EPOS 3 [28] ainsi que dans PYTHIA 8 si le mécanisme de reconnection de couleur, responsable de l'interaction entre partons venant d'interactions partoniques multiples, est désactivé. Les différentes implémentations de PYTHIA 8 et EPOS 3 donnent des prédictions similaires. Pour des impulsions transverses $p_{\mathrm{T}} < 4$ GeV/c, le taux de production des mésons D croît linéairement en fonction de la multiplicité. Au delà, la production s'éloigne de cette tendance. Dans l'ensemble, les prédictions sous-estiment la mesure pour des multiplicités relatives supérieures à ~ 2. En ajoutant une description hydrodynamique du milieu partonique dans EPOS 3, les prédictions décrivent mieux la mesure. La description hydrodynamique n'affecte pas directement la production des mésons D mais réduit le nombre de particules chargées produites à haute multiplicité en raison de la redistribution de l'énergie disponible contribuant à l'écoulement collectif. Dans ce contexte, l'évolution de la production des mésons D avec la multiplicité est interprétée comme un signe de collectivité dans les collisions à hautes multiplicités. Expérimentalement, la mesure effectuée avec l'estimateur de la multiplicité à l'avant peut être encore sensible aux biais d'auto-corrélations comme exploré dans [29].

Rapport $\sigma(\mathrm{D_s^+})/\sigma(\mathrm{D^0})$

La comparaison de la production des mésons charmés-étranges par rapport aux mésons non-étranges mesurés dans les petits systèmes en fonction de la multiplicité permet d'étudier l'effet de l'augmentation de l'étrangeté et du mécanisme de coalescence. La figure 9 montre le rapport du taux de production des mésons $\mathrm{D_s^+}$ sur celui des mésons $\mathrm{D^0}$ (constitué d'un quark charmé et d'un quark up) pour deux intervalles en multiplicité : à basse, $\sim 0.5 \cdot <\mathrm{d}N_{ch}/\mathrm{d}\eta>$, (panneau gauche) et haute, $\sim 5.5 \cdot <\mathrm{d}N_{ch}/\mathrm{d}\eta>$, (panneau droit) multiplicité. À l'inverse des mesures du rapport du taux de production des hadrons légers (multi)-étranges sur celui des mésons $\pi^{\pm}$ [1, 30], les résultats ne montrent pas une dépendance du rapport $\sigma(\mathrm{D_s^+})/\sigma(\mathrm{D^0})$ avec la multiplicité de la collision dans la limite des incertitudes expérimentales.

Les résultats sont comparés avec des prédictions issues des modèles PYTHIA 8 et CE-SH. Les prédictions issues de PYTHIA 8 sont en accord avec la mesure expérimentale suggérant qu'aucun mécanisme supplémentaire de production du quark étrange n'est nécéssaire pour décrire qualitativement le rapport $\sigma(\mathrm{D_s^+})/\sigma(\mathrm{D^0})$. En revanche, la description de la production des hadrons légers (multi)-étranges nécessite l'introduction de développements théoriques au-delà du mécanisme de fragmentation de cordes comme effectué pour le générateur Monte Carlo DIPSY [31]. Le modèle CE-SH [32] est un modèle thermique statistique effectif décrivant les différentes abondances de hadrons à partir d'un système interprété comme un gaz de hadrons et de résonances à (ou proche de) l'équilibre thermique. Alors qu'à basse multiplicité ce modèle est compatible avec la mesure, le rapport $\sigma(\mathrm{D_s^+})/\sigma(\mathrm{D^0})$ à haute multiplicité est surestimé. Dans ce modèle, la réduction de la production des mésons $\mathrm{D_s^+}$ à basse multiplicité (*i.e.* volume du gaz réduit) est la conséquence de la stricte conservation de l'étrangeté.

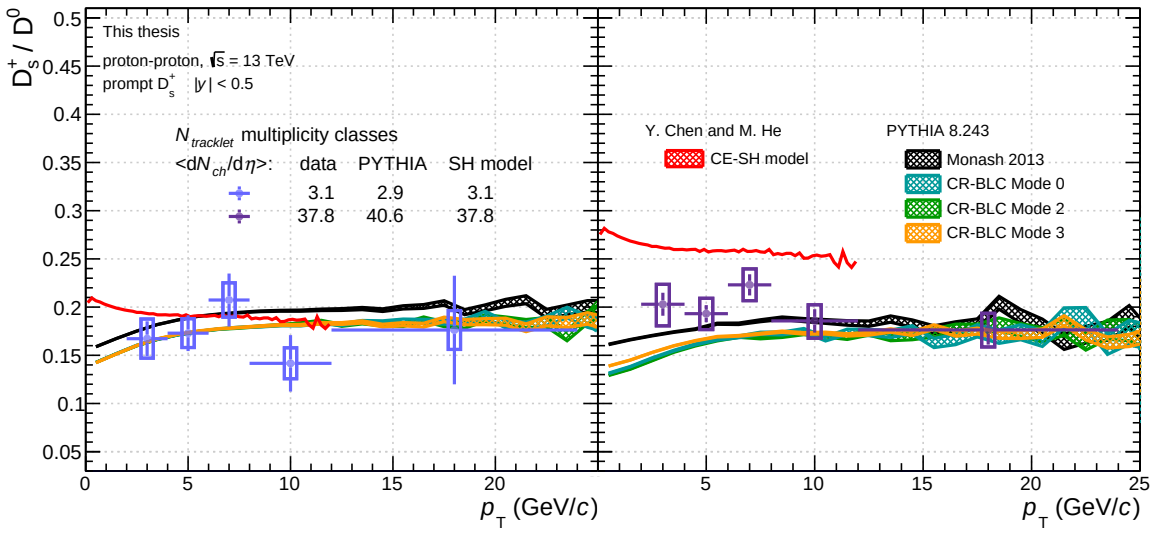


FIGURE 9 – Rapport du taux de production des mésons $\mathrm{D_s^+}$ sur celui des mésons $\mathrm{D^0}$ en fonction de l'impulsion transverse p_{T}. Les résultats sont présentés pour des basses (panneau gauche) et des hautes (panneau droit) multiplicités en particules chargées. Les mesures sont comparées aux prédictions du générateur Monte Carlo PYTHIA [4, 25] ainsi qu'aux prédictions du modèle CE-SH [32].

Conclusion

La production des mésons charmés-étranges $\mathrm{D_s^+}$ à rapidité centrale ($|y| < 0.5$) dans les collisions proton-proton à $\sqrt{s} = 13$ TeV a été mesurée, pour la première fois en fonction de la densité de particules chargées de la collision, dans l'expérience ALICE au CERN.

La reconstruction des mésons $\mathrm{D_s^+}$ a été effectuée dans le canal $\mathrm{D_s^+} \to \phi(1020)(\to \mathrm{K^+K^-})\pi^+$. La sélection du signal a été réalisée en utilisant des arbres de décision boostés, un algorithme de machine learning, exploitant la topologie de désintégration des mésons $\mathrm{D_s^+}$ et l'identification des particules filles. Cette méthode, bien plus optimisée par rapport à la méthode dite standard qui consiste en l'application successives de coupures de sélection uniques, permet de prendre en compte de manière plus approfondie les corrélations entre les variables de sélection. En conséquence, la précision statistique des mesures est plus fine et la couverture en impulsion transverse plus étendue.

Le présent travail de thèse a permis d'obtenir une mesure de la section efficace différentielle de production des mésons $\mathrm{D_s^+}$, à rapidité centrale ($|y| < 0.5$), dans les collisions proton-proton à $\sqrt{s} = 13$ TeV, énergie la plus élevée jamais atteinte par un accélérateur. C'est la mesure la plus précise à ce jour, comparée à celles réalisées précédemment à des énergies plus basses [5, 6]. Les résultats sont globalement bien décrits par les prédictions des modèles de QCD perturbative qui postulent l'universalité de la fragmentation des quarks charmés. La comparaison avec les résultats obtenus pour les mésons non-étranges montre que la production des mésons charmés étranges est réduite d'un facteur ∼3.6. Ce facteur est en accord avec les observations faites dans le cas de systèmes de collisions plus légers tels que électron-positron et électron-proton à différentes énergies de collision confortant l'hypothèse d'une fragmentation universelle des quarks charmés en mésons D. En revanche, cette hypothèse est remise en cause par les mesures des baryons charmés ouverts tels que le Λ_{c}^+, $\Xi_{\mathrm{c}}^{0,+}$ et $\Sigma_{\mathrm{c}}^{0,+,++}$ [33, 34].

La mesure obtenue dans ce travail pour la production des mésons D_s^+ en fonction de la multiplicité en particules chargées s'étend sur un intervalle couvrant de 0.5 à 5.5 fois la multiplicité moyenne observées dans les collisions. Elle contribue à l'effort de caractérisation des effets de type PQG dans les petits systèmes via le prisme de l'augmentation de l'étrangeté. La mesure du rapport des taux de production $\sigma(D_s^+)/\sigma(D^0)$ a été discutée à la lumière des récentes observations de l'augmentation de la production des hadrons (multi)-étranges à haute multiplicité dans les petits systèmes [1]. Dans la limite des incertitudes expérimentales, la mesure ne montre pas une dépendance de la production de l'étrangeté avec la multiplicité de la collision dans le secteur du charme. Le mécanisme de reconnection de couleur tel qu'implémenté dans PYTHIA 8 [25] décrit simultanément la dépendance en multiplicité des rapports $\sigma(D_s^+)/\sigma(D^0)$ et $\sigma(\Lambda_c^+)/\sigma(D^0)$. Pour ce dernier, une production accrue du baryon Λ_c^+ par rapport au D^0 est observée à haute multiplicité.

La mesure actuelle à haute multiplicité ne permet pas d'atteindre des multiplicités suffisamment élevées pour pouvoir effectuer des comparaisons avec les résultats obtenus pour les collisions Pb-Pb semi-centrales (30%-50%) [35] dans lesquelles une augmentation de la production des mésons D_s^+ est observées. Des mesures étendues sont requises pour mettre en lumière une éventuelle continuité du rapport $\sigma(D_s^+)/\sigma(D^0)$ en fonction de la multiplicité pour différents systèmes de collisions comme observée pour les hadrons légers (multi)-étranges. Ces mesures seront rendues possibles grâce aux futures améliorations de l'expérience ALICE notamment de son trajectographe interne.

Bibliographie

[1] ALICE Collaboration, "Enhanced production of multi-strange hadrons in high-multiplicity proton–proton collisions," *Nature Physics*, vol. 13, no. 6, p. 535–539, Apr 2017, DOI link, arXiv:1606.07424

[2] M. Cacciari, M. Greco, and P. Nason, "The p_{T} spectrum in heavy-flavour hadroproduction," *Journal of High Energy Physics*, vol. 1998, no. 05, p. 007–007, May 1998, DOI link, arXiv:hep-ph/9803400

[3] M. Cacciari, S. Frixione, and P. Nason, "The p_{T} spectrum in heavy-flavour photoproduction," *Journal of High Energy Physics*, vol. 2001, no. 03, p. 006–006, Mar 2001, DOI link, arXiv:hep-ph/0102134

[4] P. Skands, S. Carrazza, and J. Rojo, "Tuning pythia 8.1 : the monash 2013 tune," *The European Physical Journal C*, vol. 74, no. 8, Aug 2014, DOI link, arXiv:1404.5630

[5] ALICE Collaboration, "Measurement of beauty and charm production in pp collisions at $\sqrt{s} = 5.02$ TeV via non-prompt and prompt D mesons," 2021, arXiv:2102.13601

[6] ALICE Collaboration, "Measurement of D-meson production at mid-rapidity in pp collisions at $\sqrt{s} = 7$ TeV," *The European Physical Journal C*, vol. 77, no. 8, Aug 2017, DOI link, arXiv:1702.00766

[7] M. Benzke, M. V. Garzelli, B. A. Kniehl, G. Kramer, S. Moch, and G. Sigl, "Prompt neutrinos from atmospheric charm in the general-mass variable-flavor-number scheme," *Journal of High Energy Physics*, vol. 2017, no. 12, Dec 2017, DOI link, arXiv:1705.10386

[8] G. Kramer and H. Spiesberger, "Study of heavy meson production in p-Pb collisions at $\sqrt{s}$= 5.02 TeV in the general-mass variable-flavour-number scheme," *Nucl. Phys. B*, vol. 925, pp. 415–430, 2017, DOI link, arXiv:1703.04754

[9] P. Bolzoni and G. Kramer, "Inclusive charmed-meson production from bottom hadron decays at the lhc," *Journal of Physics G : Nuclear and Particle Physics*, vol. 41, no. 7, p. 075006, May 2014, DOI link, arXiv:1310.2924

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

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

[12] T. J. Hou, J. Gao, J. Hobbs, K. Xie, S. Dulat, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, and et al., "New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC," *Physical Review D*, vol. 103, no. 1, Jan 2021, DOI link, arXiv:1912.10053

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

[14] B. Mele and P. Nason, "The Fragmentation function for heavy quarks in QCD," *Nucl. Phys. B*, vol. 361, pp. 626–644, 1991, DOI link, [Erratum : Nucl.Phys.B 921, 841–842 (2017)].

[15] B. A. Kniehl and G. Kramer, "Charmed-hadron fragmentation functions from CERN LEP1 revisited," *Physical Review D*, vol. 74, no. 3, Aug 2006, DOI link, arXiv:hep-ph/0607306

[16] M. Cacciari, M. L. Mangano, and P. Nason, "Gluon PDF constraints from the ratio of forward heavy quark production at the LHC at $\sqrt{s} = 7$ and 13 TeV," *Eur. Phys. J. C*, vol. 75, p. 610. 21 p, Jul 2015, DOI link, arXiv:1507.06197

[17] ALICE Collaboration, "D_s^+ meson production at central rapidity in proton-proton collisions at $\sqrt{s} = 7$ TeV," *Physics Letters B*, vol. 718, no. 2, p. 279–294, Dec 2012, DOI link, arXiv:1208.1948

[18] A. Collaboration, "Measurement of $D^{*\pm}$, $D^{*\pm}$ and $D_s^{\pm}$ meson production cross sections in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector," *Nuclear Physics B*, vol. 907, p. 717–763, Jun 2016, DOI link, arXiv:1512.02913

[19] H. Collaboration, "Inclusive production of D^+, D^0, D_s^+ and D^{*+} mesons in deep inelastic scattering at HERA," *The European Physical Journal C*, vol. 38, no. 4, p. 447–459, Jan 2005, DOI link, arXiv:hep-ex/0408149

[20] ZEUS Collaboration, "Measurement of charm fragmentation fractions in photoproduction at HERA," 2013, arXiv:1306.4862

[21] L. Gladilin, "Fragmentation fractions of c and b quarks into charmed hadrons at LEP," *The European Physical Journal C*, vol. 75, no. 1, Jan 2015, DOI link, arXiv:1404.3888

[22] A. Collaboration, "Measurement of charm and beauty production at central rapidity versus charged-particle multiplicity in proton-proton collisions at $\sqrt{s} = 7$ TeV," *Journal of High Energy Physics*, no. 9, 2015, DOI link, arXiv:1505.00664

[23] CMS Collaboration, "Investigation into the event-activity dependence of Υ(nS) relative production in proton-proton collisions at $\sqrt{s} = 7$ TeV," *Journal of High Energy Physics*, vol. 2020, no. 11, Nov 2020, DOI link, arXiv:2007.04277

[24] ALICE Collaboration, "Multiplicity dependence of π, k, and p production in pp collisions at $sqrts = 13$ tev," *The European Physical Journal C*, vol. 80, no. 8, Aug 2020, DOI link, arXiv:2003.02394v2

[25] J. R. Christiansen and P. Z. Skands, "String formation beyond leading colour," *Journal of High Energy Physics*, vol. 08, p. 003, 2015, DOI link, arXiv:1505.01681

[26] H. Drescher, M. Hladik, S. Ostapchenko, T. Pierog, and K. Werner, "Parton-based Gribov–Regge theory," *Physics Reports*, vol. 350, p. 93–289, 2001, DOI link, arXiv:0007198

[27] K. Werner, B. Guiot, I. Karpenko, and T. Pierog, "Analyzing radial flow features in p-Pb and p-p collisions at several TeV by studying identified-particle production with the event generator EPOS3," *Physical Review C*, vol. 89, no. 6, 2014, DOI link, arXiv:1312.1233

[28] K. Werner, B. Guiot, I. Karpenko, T. Pierog, and G. Sophys, "Charm production in high multiplicity pp events," *Journal of Physics : Conference Series*, vol. 736, p. 012009, Aug 2016, DOI link, arXiv:1602.03414v1

[29] S. G. Weber, A. Dubla, A. Andronic, and A. Morsch, "Elucidating the multiplicity dependence of J/ψ production in proton–proton collisions with PYTHIA8," *The European Physical Journal C*, vol. 79, no. 1, Jan 2019, DOI link, arXiv:1811.07744

[30] ALICE Collaboration, "Multiplicity dependence of (multi-)strange hadron production in proton-proton collisions at $\sqrt{s} = 13$ TeV," *The European Physical Journal C*, vol. 80, no. 2, Feb 2020, DOI link, arXiv:1908.01861

[31] C. Bierlich, G. Gustafson, L. Lönnblad, and A. Tarasov, "Effects of Overlapping Strings in pp Collisions," *JHEP*, vol. 03, p. 148, 2015, DOI link, arXiv:1412.6259

[32] Y. Chen and M. He, "Charged-particle multiplicity dependence of charm-baryon-to-meson ratio in high-energy proton-proton collisions," *Physics Letters B*, vol. 815, p. 136144, Apr 2021, DOI link, arXiv:2011.14328

[33] ALICE Collaboration, "Λ_c^+ production in pp and in p-pb collisions at $\sqrt{s_{NN}} = 5.02$ tev," 2021, arXiv:2011.06079

[34] ALICE Collaboration, "Measurement of the cross sections of Ξ_c^0 and Ξ_c^+ baryons and branching-fraction ratio $\mathrm{BR}(\Xi_c^0 \to \Xi^- e^+ \nu_e)/\mathrm{BR}(\Xi_c^0 \to \Xi^- \pi^+)$ in pp collisions at 13 TeV," 2021, arXiv:2105.05187

[35] ALICE Collaboration, "Measurement of prompt D_s^+-meson production and azimuthal anisotropy in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV," 2021, arXiv:2110.10006

Introduction

The quantum chromodynamics theory (QCD) describing the strong interaction between quarks and gluons predicts a phase transition from hadronic matter to a colour-deconfined medium called the quark-gluon plasma (QGP) [1–4]. This state of nuclear matter, presumably the state of the early universe up to a few micro-seconds after the Big-Bang, is created and studied in ultra-relativistic heavy-ion experiments. Experimental signatures of the QGP were first detected at the Super Proton Synchrotron (SPS), where experiments provided evidence for a new state of matter [5]. At the Relativistic Heavy Ion Collider (RHIC), the quark-gluon plasma was characterised as a highly opaque medium behaving like a nearly perfect fluid [6–9]. Extended characterisations of the QGP thermodynamic properties in ultra-relativistic collisions are performed at RHIC and since 2009 also at the Large Hadron Collider (LHC).

A Large Ion Collider Experiment (ALICE) is a dedicated QGP experiment at the LHC at CERN. The main physics goal of the ALICE collaboration is to study the properties of the QGP in ultra-relativistic heavy-ion collisions. In the context of these studies, small systems (proton-proton and proton-A) are taken as "non-QGP" reference systems used mainly for the purposes of comparison with heavier colliding systems. However, recent LHC measurements performed in small systems at high collision charged-particle density exhibit some effects similar to those observed in heavy-ion collisions, such as the presence of positive elliptic flow of unidentified charged particles [10], and the enhanced production of multi-strange hadrons [11].

The work presented in this thesis contributes to the experimental effort aimed at the investigation of QGP-like effects in small systems in the open heavy-flavour sector. Open heavy-flavour hadrons are composed of at least one heavy quark (charm or bottom) and light quark(s). At LHC energies, heavy quarks are produced in hard scattering processes at the early stage of collision (before the QGP formation). They are therefore sensitive to the full evolution of the partonic medium. Their interactions with the QGP, such as energy losses, are described in theoretical models. Thus, the characterisation of heavy-flavour hadron production allows for an indirect study of the QGP properties. About ninety percent of the charm quark population produced in collisions hadronise into open-charm hadrons [12]. Among them, the $\mathrm{D_s^+}(c\bar{s})$ meson, is composed by a charm and a strange quark. Strangeness enhancement is one of the experimental signatures of the QGP [13,14]. In central Pb-Pb collisions, a hint of an enhanced strange charmed meson production has been observed [15]. Due to its strangeness content, the $\mathrm{D_s^+}$ meson is expected to be sensitive to the in-medium hadronisation mechanisms. In small systems, the observation of an enhanced production of multi-strange hadrons [11] in the light flavour sector motivates performing similar studies in the heavy-flavour sector.

In this thesis, the production of $\mathrm{D_s^+}$ mesons in proton-proton collisions was studied at the centre-of-mass energy of $\sqrt{s} = 13$ TeV, using the data collected by the ALICE experiment during the LHC run II. The $\mathrm{D_s^+}$ mesons were reconstructed in the $\mathrm{D_s^+} \rightarrow \phi(1020)(\rightarrow \mathrm{K^+K^-})\pi^+$ decay channel, exploiting their displaced vertex topology. The high integrated luminosity recorded during the LHC run II allowed for detailed measurements of $\mathrm{D_s^+}$ p_T-differential cross section, as a function

of the collision charged particle density.

The manuscript consists of six chapters. In the first part of Chapter I, a brief review of the ideas from theoretical physics and experimental results contributing to the development of the quantum chromodynamics theory is given. The most fundamental properties of QCD and the QCD phase diagram are described. In the second part, a focus is made on the experimental study of the QGP in ultra-relativistic heavy-ion collisions. A few experimental QGP signatures are presented. Chapter II is dedicated to open heavy-flavour physics. The theoretical models describing the production of open heavy-flavour hadrons are presented in the first part of the chapter. Measurements of the heavy-flavour hadron production in proton-proton, proton-Pb and Pb-Pb are then presented and discussed in the second part. The motivations for the study of the production of strange charmed mesons as a function of the charged particle density are introduced in the last part of the chapter. In Chapter III, an overview of the ALICE detectors with a focus on the sub-systems of interest for the present thesis work, is given. The detectors performance and the main principles of the ALICE event reconstruction are presented as well.

The last three chapters (IV, V and VI) are dedicated to the presentation of the analysis work and the results obtained in this thesis. The identification of D_s^+ mesons is performed by the topological reconstruction of the $D_s^+ \to \phi(1020)(\to K^+K^-)\pi^+$ decay channel. One of the most critical parts of the analysis is the extraction of the D_s^+ signal from the large combinatorial background originating from the reconstruction procedure. In Chapter IV, two methods for suppressing the background are considered. The first one applies independent selection cuts, whereas the second is based on a machine learning technique. The application of this second method allows for a significant improvement of the reconstruction performances (signal significance), which is important for the measurements at low and high collision charged particle density. After correction of the extracted raw yield, the production cross section and the production yields for different multiplicity intervals are obtained. A dedicated study of the systematic uncertainties is performed in Chapter V allowing for a discussion of subtle aspects of the data analysis and the evaluation of its limits. The results are presented and discussed in Chapter VI. A comparison to the non-strange D mesons production is shown. The dependance of the D_s^+ production cross section on the collision energy is studied as well. The transverse-momentum integrated cross section is also evaluated. The measurement of the production of D_s^+ mesons as a function of the charged particle density is presented and discussed in comparison to the non-strange D mesons production and with theoretical models. Finally, the measurement of the ratio of production yields of D_s^+ over the D^0 is discussed in the context of the recent observations related to the presence of QGP-like effects in small systems. The discussion is then extended to baryon over meson ratio both in the light- and heavy-flavour sectors.

A general conclusion ends this manuscript, giving a summary of the main results and the perspectives to this thesis work.

Chapter I

Heavy-ion physics

In this chapter, a brief review is given on the ideas from theoretical physics and experimental results contributing to the development of the quantum chromodynamics theory (QCD), the current admitted quantum field theory describing the strong interactions of quarks and gluons. The fundamental properties of the QCD will be introduced as the colour confinement and asymptotic freedom. The QCD predicts a phase transition from hadronic matter to a colour-deconfined medium called the quark-gluon plasma (QGP) [1–4]. The QGP is created and studied experimentally in ultra-relativistic heavy ion collisions. In the last part of the chapter, an overview of selected experimental results from heavy-ion physics will be presented and discussed.

I.1 Strong interaction

I.1.1 Quark model

The term "strong" interaction inherits from nuclear physics as the strong nuclear force binding nucleons in a nuclei. Protons and neutrons were the first discovered strongly interacting particles, called hadrons. With the variety of hadrons measured in cosmic rays and particle accelerator experiments as the pions, kaons, hyperons (Λ, Ξ, Σ, Ω) and resonance particles, the idea came that hadrons are not fundamental particles.

In an attempt to organise them, Y. Ne'eman [16] and M. Gell-Man [17, 18] used group theory and developed a hadron classification scheme based on the SU(3) symmetry. In this model known as the "eightfold way", hadrons are grouped in multiplets (octets and decuplets) sharing the same spin and parity. The spin 1/2 hadron octet, including the proton and the neutron, is represented on the middle panel of figure I.1.1. One of the most successful result of the eightfold way is the prediction of the Ω^- baryon, experimentally discovered three years later (see right panel of figure I.1.1). Despite the success of the model, no fundamental physical explanation of the origin of SU(3) symmetry was given.

In 1964, M. Gell-Man [19] and G. Zweig [20] independently developed the quark model, offering a fundamental description of hadrons constructed from three fundamental particles called quarks. The quarks have fractional electrical charges, $+\frac{2}{3}e$ or $-\frac{1}{3}e$, and a spin 1/2; they form a fundamental representation of the global symmetry $SU(3)_{flavour}$ which is broken since quarks don't have the same masses. Hadrons are classified into two families: mesons, as quark-antiquark ($q\bar{q}$) bound states and baryons, as qqq (or $\bar{q}\bar{q}\bar{q}$, anti-baryon) bound states. At that time only three quarks (up, down and strange) were needed to explain the hadron quantum numbers (baryon number, charge, spin...) of the measured hadron mass spectrum.

Quarks are also included in field theory toy models where their properties are studied through specific properties of their weak and electromagnetic interactions. From such models, M. Gell-Mann derived algebraic symmetry relations between measurable quantities leading to the development of the "current algebra" [18, 21].

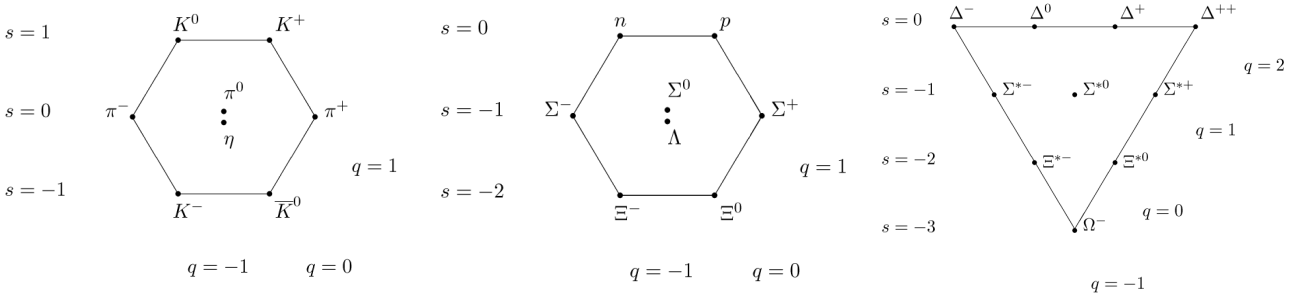


Figure I.1.1: Mesons octet (left), spin 1/2 hadron octet (middle) and spin 3/2 hadron decuplet (right). Hadrons on the same horizontal lines share the same strangeness quantum number s while hadrons on the same diagonal share the same electric charge q.

I.1.2 Parton model

Further insights on the hadron structure came with deep inelastic lepton-hadron scattering experiments. The inclusive cross section expression of a deep electron (e^-) - hadron (H) scattering process, $e^- + H \rightarrow e' + X$, (final state particles X not measured), can be parametrised in term of structure functions which are experimental observables. Such functions are derived from the theoretical description of the electroweak interactions of the lepton with the target hadron, typically a proton or a neutron. They depends on the energy transferred to the hadron by the scattering electron in its rest frame (ν) and the 4-momentum transfer squared (Q^2) of the process.

Pursuing his previous work [22], J. D. Bjorken [23] proposed that in the infinite momentum transfer limit, the structure functions would not depend on ν and Q^2 independently but on the Bjorken scaling variable $x = \frac{Q^2}{2M\nu}$ with M the mass of the nucleon. Bjorken scaling as the Callan-Gross relation between structure functions [24] were confirmed experimentally by SLAC-MIT deep inelastic scattering experiments [25, 26].

The parton model, introduced in 1969 by R. Feynman [27] and by J. D. Bjorken and E. A. Paschos [28], explained the experimental results by describing a hadron at high energy as a collection of non-interacting point-like particles called partons. In the model, the deep inelastic scattering inclusive cross section of a hadron with a particle is described as an incoherent sum of the elastic scattering cross sections of the quasi-free partons with the particle. Each partons carries a fraction x of the hadron longitudinal momentum which corresponds to the Bjorken scaling variable x, the probability density function for finding a parton with a certain fraction of the proton momentum is called parton distribution function (PDF). Structure functions are expressed as sums of the parton PDFs.

Deep inelastic lepton-hadron scattering experiments showed that only half of the measured proton momentum is recovered if the proton is assumed to be only constituted by three valence quarks. At that time, the missing fraction was attributed to hypothetical electrically neutral constituents, the "quark glue", the gluons.

The Bjorken scaling behaviour and results from the partons model were reproduced by the extension of Gell Mann's current algebra, the light cone current algebra [29], abstracted from a free quark triplet theory and from a interacting single gluon field theory.

I.1.3 Colour charge

There were many signs showing the limits of the quark model. An example is the Δ^{++} a spin 3/2 baryon composed of three up quarks, for which the Fermi statistics is violated since its total wave function is expected to be antisymmetric. To reconcile with the Pauli exclusion principle, O. W. Greenberg [30], introduced an additional internal quantum number to quarks using a Fermi-Dirac "para-statistics" model. This quark quantum number, latter called "colour charge" comes with three distinct indices: "red" ($\bar{r}$), "green" ($\bar{g}$) and "blue" ($\bar{b}$). Antiquarks carry an anti-colour index.

In 1965, M. Y. Han and Y. Nambu proposed a model [31] with three triplets of integer electric charged quarks based on a double SU(3) symmetry. The model tried to overcome some of the limits of the quark model: why only a fraction of the hadrons predicted were observed and the lack of a quark confinement dynamical mechanism. They postulated a strong force coupled to colour, mediated by eight gauge vector fields, responsible for a mass splitting between colour-neural hadrons and coloured hadrons. Colour states being more massive and unstable than colour-neutral states, they would be difficult to observe in experiment.

In order to form a colour-neutral hadrons, colour indices of the constituent quarks must cancel out. Colour-neutral mesons are made of a valence quark-antiquark pair, the antiquark anti-colour matching the quark colour. Colour-neutral baryon are made of three valence quarks, each of them carrying a different colour index: rgb edifices (rgb = white i.e "neutral colour"). For anti-baryon it must be $\bar{r}\bar{g}\bar{b}$ edifices.

Including colour into their field theory toy model, H. Fritzsch, W. A. Bardeen and M. Gell-Mann developed a quark-gluon model [29] containing three triplets of spin 1/2 quarks, i.e. three flavour quarks coming in three colour indices. The $\mathrm{SU}(3)_{colour}$ symmetry is responsible for colour indices transformations, physical states are required to be colour-neutral (also called colour singlet). Using the derived light cone algebra, they successfully predicted the $\pi^0 \to 2\gamma$ decay rate amplitude and the ratio of cross-sections of electron-positron annihilation into hadrons and muons.

I.1.4 Quantum chromodynamics

One of the main ideas leading to a complete description of the dynamics of quark and gluon strong interactions was to extend the global $\mathrm{SU}(3)_{colour}$ symmetry of the quark-gluon field theory Lagrangian to a local symmetry. In order to achieve the local invariance, *i.e.* the physics remains unchanged in different space-time points, one can introduce gauge fields into the theory making it a gauge theory. Fundamental interactions can be described in gauge theories as it was already done at that time for the electromagnetism theory where the introduced gauge field was identified as the photon responsible of the electromagnetic interaction between charged particles.

As gluons were suggested to be the mediators of the strong interaction between quarks, H. Fritzsch and M. Gell-Mann in [32, 33], interpreted them as gauge fields of a $\mathrm{SU}(3)_{colour}$ local gauge symmetry. The corresponding non-abelian gauge theory, a Yang-Mills theory [34], implies that gauge fields self-interact. The strong interaction "paradox", quarks and gluons being strongly confined into hadrons at low energy (quark model picture) and behaving like quasi-free particles at high energy (parton model picture) was explained by the asymptotic freedom property of non-abelian gauge theories. The property was discovered by D. J. Gross, F. Wilczek [35] and H. D. Politzer [36], they showed that self interactions of the gauge fields of such theories approaches a free-field theory in the high energy limit.

Quantum chromodynamics (QCD) is the current admitted quantum field theory describing the strong interactions of quarks and gluons. Quarks are massive, spin 1/2, fermions with fractional electric charges, they comes in six flavours: up, down, strange, charm, bottom and top, each flavour having three colour charge indices. Quarks properties are summarised in Table I.1.1. The three colour indices form a fundamental representation of $SU(3)_{colour}$ gauge symmetry which is an exact symmetry. Eight gluons, massless spin 1 bosons, couples to the colour charge. Gluons being colour charge carriers, they can self-interact. The fundamental coupling "constant" (*i.e.* the strength) of the interaction $\alpha_s(Q^2)$ depends on the squared momentum transfer (Q^2), this property generates the full underlying dynamics of hadron physics from colour confinement to asymptotic freedom. Quark masses and the coupling constant are the fundamental parameters of QCD, therefore they cannot be predicted by the theory. They are extracted from comparisons of theoretical predictions with experimentally measurements of various physical observables.

	up	**charm**	**top**
mass	$m_u = 2.16^{+0.49}_{-0.26}$ MeV/c^2	$m_c = 1.27^{+0.02}_{-0.02}$ GeV/c^2	$m_t = 172.9^{+0.4}_{-0.4}$ GeV/c^2
charge	2/3 e	2/3 e	2/3 e
spin	1/2	1/2	1/2
	down	**strange**	**bottom**
	$m_d = 4.67^{+0.48}_{-0.17}$ MeV/c^2	$m_s = 93^{+11}_{-5}$ MeV/c^2	$m_b = 4.18^{+0.03}_{-0.02}$ GeV/c^2
	-1/3 e	-1/3 e	-1/3 e
	1/2	1/2	1/2

Table I.1.1: Quark flavours table

$\alpha_s(Q^2)$ running and asymptotic freedom

The QCD vacuum is made of fluctuating virtual quark-antiquark pairs and gluons, making it behave like a polarised medium. Considering a colour charge in the vacuum, virtual $q\bar{q}$ contributes to the colour screening effect, the effective colour charge decreases with increasing distance. Virtual gluon clouds contributes to the effective colour charge with an anti-screening effect which is the dominant contribution. Screening and anti-screening effects leads to the change of the interaction coupling strength as a function of distance (or equivalently of momentum). The evolution of the coupling α_s as a function of the momentum scale Q is governed by the renormalisation group β-function of QCD.

$$Q^2 \frac{\partial \alpha_s}{\partial Q^2} = \beta(\alpha_s) = -\frac{1}{2\pi}(11 - \frac{2}{3}n_f)\alpha_s^2 + ...$$

with n_f the number of active flavours at the scale Q.
In $(11 - \frac{2}{3}n_f)$, the first term comes from gluon anti-screening while the factor $\frac{2}{3}$ counts for $q\bar{q}$ screening, for $n_f \leqslant 16$, the anti-screening contribution is dominant. Therefore, the interaction strength decreases with increasing momentum transfer.
In the one-loop approximation, the solution gives the relation of the coupling between two different scales:

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu_0^2)}{1 + \alpha_s(\mu^2)\beta_0 \ln \frac{Q^2}{\mu_0^2}}$$

One can substitute the arbitrary starting scale μ_0 by introducing Λ_{QCD}.

$$\alpha_s(Q^2) = \frac{1}{\beta_0 \ln \frac{Q^2}{\Lambda_{QCD}}}$$

For Q $\gg \Lambda_{QCD}$, $\alpha_s(Q) \to 0$, the asymptotic freedom regime is reached. For Q below the Λ_{QCD} value, $\alpha_s(Q \to \Lambda_{QCD})$ diverge, the strength of the strong interaction becomes infinite. Λ_{QCD} acts as the limit between the perturbative and non-perturbative regimes. The Λ_{QCD} value depends on the renormalisation scheme and number of active quark flavours considered, the "modified minimal subtraction scheme" ($\overline{MS}$) is a common choice [37], for $n_f = 3$, $\Lambda_{QCD} \simeq 250$ MeV.

In the asymptotic freedom regime, perturbative theory can be applied to QCD where observables are expressed as a perturbative series of α_s. Quantitative predictions for high momentum transfer physical processes (hard processes) can be derived. The coupling constant evolution as a function of the energy scale is not measured directly but is extracted from hard scattering process measurements. Heavy quarkonia ($c\bar{c}$ and $b\bar{b}$ mesons) decay width, deep inelastic scattering processes, jet production rates are examples of observables for which perturbative QCD provide predictions. Figure I.1.2 show the evolution of α_s as a function of the energy scale Q, it goes from ~0.3 at the GeV scale to ~0.12 at the Z boson mass scale. By comparison, the fundamental QED coupling constant evolve from ~1/137 to ~1/127 over the same range. Such amplitude of the coupling constant variation is a unique feature of the QCD.

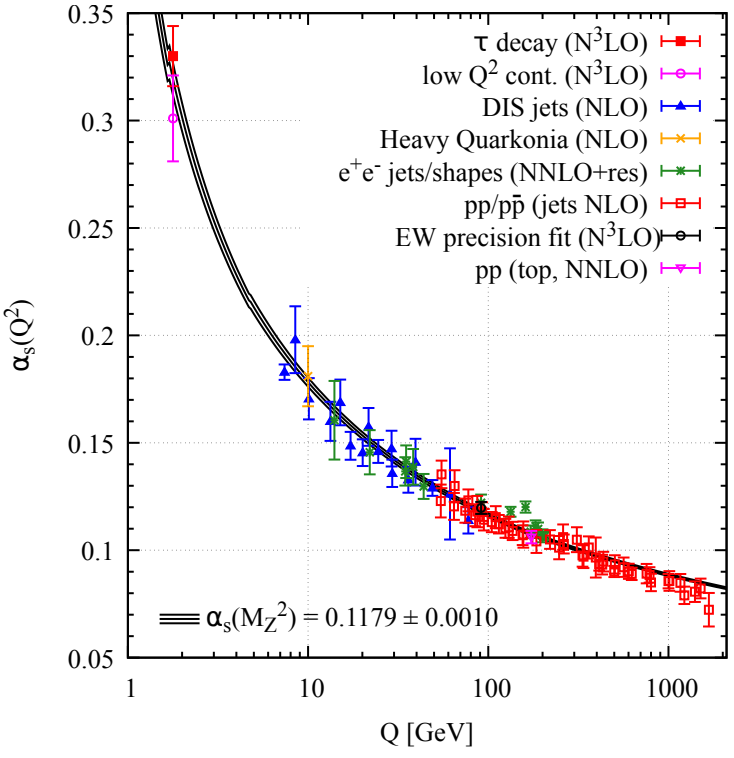


Figure I.1.2: Measurements of α_s a function of the energy scale Q. The respective degree of QCD perturbation theory used in the extraction of α_s are indicated in brackets [37]. The value of α_s at an energy scale equal to the Z boson mass is shown.

One of the important predictions of perturbative QCD is the violation of Bjorken scaling for the nucleon structure functions at low x ($<10^{-2}$), experimentally confirmed at HERA electron-proton collider [38]. The original parton model gives a static view of the nucleon structure at high

energy as the nucleon is made of quasi-free particles. The spatial resolution (Q^2) dependence of the nucleon structure was brought by QCD introducing valence quark interactions. Bjorken scaling violation of the nucleon structure functions is explained by the gluon emission of valence quarks and by the sea of virtual $q\bar{q}$ pairs coming from those gluons. The nucleon parton distribution functions evolution as a function of Q^2 are governed by the Dokshitzer, Gribov, Lipatov, Altarelli and Parisi (DGLAP) renormalisation equations [39–42].

Colour confinement

The statement that isolated free coloured particles does not exist in nature at energy scales below Λ_{QCD} is supported by experiment since no evidence of individual free parton as been made. The QCD SU(3) symmetry impose confined quarks and gluons into colour-neutral hadrons. While perturbative QCD established the validity of QCD at high energy, the theory is not applicable in the regime of strong coupling and in consequence cannot explain the colour confinement mechanism.

A new formulation of QCD called Lattice QCD was proposed by K. G. Wilson [43] in which the continuum gauge theory is formulated on a discrete lattice in Euclidian space-time with preservation of the local gauge invariance. Quarks are placed on lattice site while gluons represents the link in between. Considering a static $q\bar{q}$ pair and assuming a sufficiently large value of the coupling, K. G. Wilson showed that the probability of finding the pair in a separated configuration decreases with increasing distance. The quark and the antiquark are bound by a potential linearly increasing as a function of the separation distance, an infinite distance separation would require an infinite amount of energy.

The static quark potential can be phenomenologically described as a sum of a Coulomb potential and a linear potential [44]:

$$V_{QCD}(r) \propto -\frac{4}{3}\frac{\alpha_s(r)}{r} + \sigma.r \qquad \text{with } \sigma \text{ the string tension}$$

The Coulomb term, dominant at small r, comes from the fractional electric charge of the quarks while the linear term, dominant a large r, arises from gluon interactions between the quarks. The strong coupling, i.e the gluon field between the quarks, can be interpreted as a string. The string picture comes from the colour flux squeezed into a thin one-dimensional flux tube (a string) leading to a constant energy density per unit of length and a linearly increasing potential as a function of the distance. Beyond a string length of 1 fm, string breaking is provoked by colour charge screening effect from a $q\bar{q}$ pair creation.

Chiral symmetry

The Lagrangian of QCD is invariant under the chiral group transformations, $SU(N_f)_R$ x $SU(N_f)_L$ (N_f, number of quark flavours) group, if quarks are massless. Considering only the light quarks, u, d and s, having small masses, the chiral symmetry is only approximate. In nature, this symmetry would result in the existence of parity doublets of hadron states with close mass, same spin but opposite parity. Since parity doublets are not experimentally observed, an additional mechanism for the symmetry breaking is considered, the spontaneous symmetry breaking.

Lagrangian invariance under a symmetry group of transformations does not imply that the ground state, the vacuum state, fulfil this invariance. In that case, the symmetry is said spontaneously broken and its manifestation is realised in the existence of spin 0 particles of zero mass as postulated by the Goldstone theorem [45, 46]. The eight predicted Nambu-Goldstone bosons acquire mass because of the additional explicit symmetry breaking induced by u, d and s quark

masses. They are identified with the $\pi^{+,-,0}$, $K^{+,-}$, K^0 , $\bar{K}^0$ and η mesons. The QCD vacuum is characterised by the chiral quark condensate, a Bose-Einstein condensate of quark-antiquark pairs, which is an order parameter of the chiral symmetry. A non vanishing value of the chiral quark condensate is an indication of the chiral symmetry breaking.

Quarks can be interpreted in two different ways, as current quarks (bare quarks) or as constituent quarks (valence quarks). Current quarks are the quarks of the QCD Lagrangian (see Table I.1.1). Constituent quarks are the quark "dressed" by their interactions with the QCD vacuum ($m_u \sim 300$ MeV). Taking the proton as example, one can easily notice that its mass ($\sim$ 938 MeV) does not correspond to the sum of its bare quark masses ($\sim$10 MeV). Most of the proton mass is generated dynamically by QCD. Mass of light and charmed baryons obtained form different lattice QCD calculations compared with experimental data are presented on Figure I.1.3.

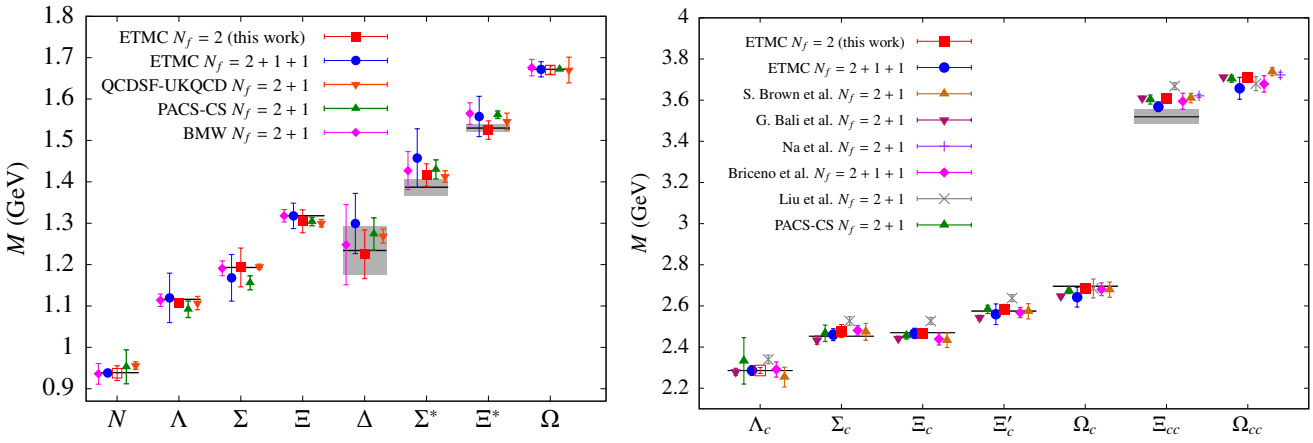


Figure I.1.3: Baryon masses obtained at physical point or extrapolated to the physical point of different lattice calculations [47]. The results are compared with experimental masses shown by the horizontal bands [48].

I.1.5 Hadronic matter under extreme conditions

Hadronic matter under extreme conditions is of particular interest since its macroscopic and thermodynamics properties derives directly from the underlying fundamental strong interactions.

The heating of hadron gas can be studied within the Statistical Bootstrap Model (SBM) [49] to study statistical thermodynamics of strong interactions at high energies. In the SBM, resonance and hadron states are identified as clusters. Clusters are composed of lighter clusters and can be potential constituents of heavier clusters. The strong interaction is generated due to the presence of clusters, as a consequence, a volume of strongly interacting hadrons is described as a mixture of ideal cluster gases. Considering an increasing temperature of the gas system, heavier hadronic states are being produced. R. Hagedorn showed an exponential rise of the mass spectrum of the hadronic states. While massive hadron resonance production keeps going on, at a certain point, the temperature of the system reaches a limit, the Hagedorn temperature T_H $\sim$170 MeV, representing the temperature limit of hadronic matter. The Hagedorn temperature was re-interpreted by N. Cabibbo and G. Parisi [50] as an indication of a phase transition between an hadronic gas and a deconfined quark-gluon phase.

Pursuing previous works on hyperon stars, N. Itoh studied the hydrodynamical equilibrium of super-dense stars made of a quark matter [51]. In the context of neutron stars, J. C. Collins and M. J. Perry [52], postulated that matter at densities higher than the nuclear ($n_0 \simeq 0.17$

nucleon.fm^{-3}) consists of a "quark soup". As matter is compressed, hadrons get closer, up to a point where they overlap. Quarks from a given hadron get screened by other quarks. As consequence, quark interactions weaken and "loose" their original hadron identity leading to their de-confinement.

In the QCD framework, asymptotic freedom suggests that at high temperatures and/or densities the relevant degree of freedom of hadronic matter are the deconfined quarks and gluons.

I.1.6 Quantum chromodynamics phase diagram

The QCD phase diagram represent the hadronic matter states in different temperature and baryon chemical potential (μ_B) regions. The baryon chemical potential is a measure of the energy needed to increase by one unit the baryonic number of the system, it relates to the net baryon density of the system. The schematic phase diagram presented on figure I.1.4 highlights the three main domains of hadronic matter separated by critical lines: the hadron gas, the QGP and the colour super-conducting quark matter. The quark-gluon plasma phase occupying the high temperature regime is a strong candidate for the state of the early universe a few micro-seconds after the Big-Bang. At low temperature and high density, super-conducting quark matter and derived states are hypothetical states of dense astrophysical objects as neutron stars [53–55].

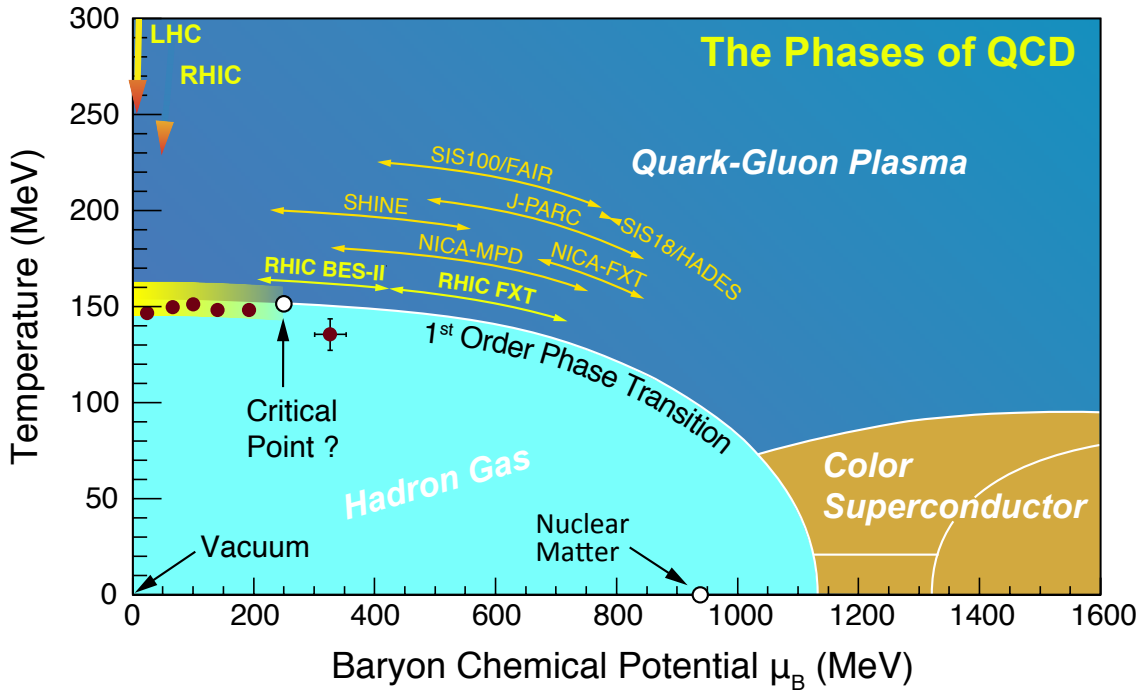


Figure I.1.4: Quantum chromodynamics phase diagram.

Lattice QCD has been extensively used to study the QCD phase transition at finite temperature and vanishing baryon chemical potential. For region of the phase diagram with $\mu_B > 0$, where lattice QCD is not applicable, extrapolations from lattice are made, other models and perturbative QCD calculations are used. The predictions showed that the general shape of the diagram and the order of the phase transition is dictated by the global symmetries of the QCD Lagrangian, the number of quark flavour considered and the quark masses as summarised below.

- **massless quarks QCD**: the chiral symmetry being spontaneously broken in the hadron gas, it is restored in the QGP phase. The transition phase can be of first or second order.
- **2 massless quarks + 1 massive strange quark QCD**: two transition phase domains appears. The second order and first order domains are separated by a critical point which position depends on the strange quark mass and μ_B.
- **3 massive quarks QCD**: a first order transitions are identified in the vanishing and infinite quark mass limits. At intermediate quark masses, a crossover transition takes place.

For three quarks (up, down and strange) with physical masses, a crossover phase transition occurring at $T_c \sim 155$ MeV and $\mu_B = 0$ MeV is predicted by lattice QCD calculations [1–4]. During the crossover transition the hadronic gas and the QGP phases coexist in an indistinguishable manner. The critical end point position, delimiting the crossover and the first order phase transition, is still under investigation.

In order to describe the crossover phase transition, one has to determine the equation of state of hadronic matter. The lattice QCD calculations, with three quarks with physical masses, are presented on figure I.1.5. On the left panel of figure I.1.5, the temperature evolution of the trace anomaly (also called interaction measure) is shown. It is defined as $(\epsilon - 3p)/T^4$ with ϵ the energy density, p the pressure and T the temperature of the system. The trace anomaly inflexion point highlights the temperature of the crossover transition between a hadron gas and the QGP.

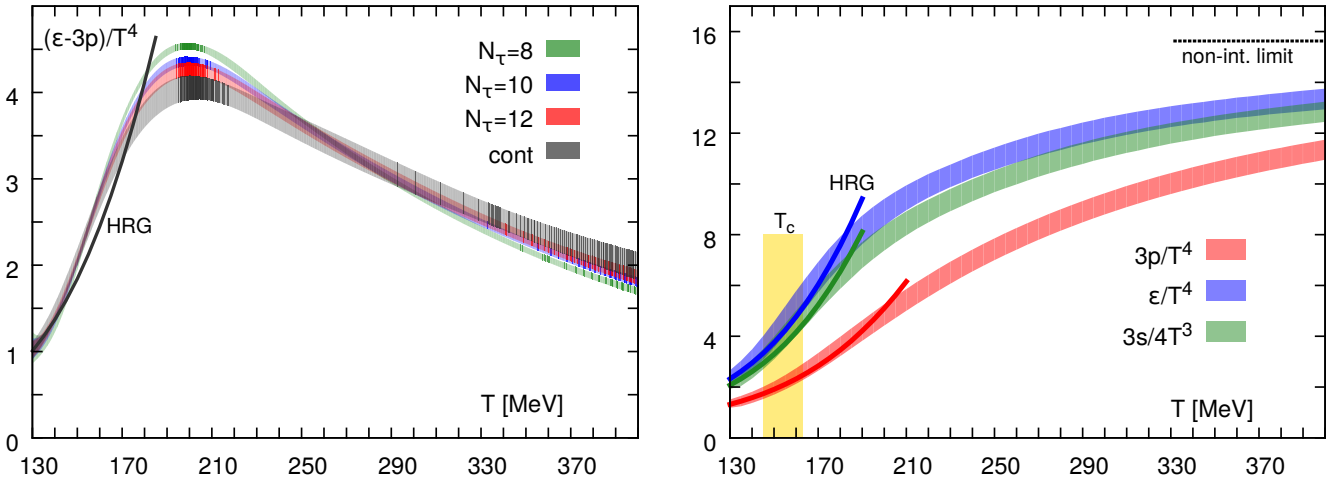


Figure I.1.5: Left panel: Spline fits to the trace anomaly for several values of the lattice spacing ($aT = 1/N_\tau$) and the result of the continuum extrapolation in grey [3].
Right panel: the normalised pressure ($3p/T^4$), energy density (ϵ/T^4), and entropy density ($3s/T^4$) as a function of the temperature. The dark lines show the prediction of the HRG model. The horizontal dashed line corresponds to the Stefan-Boltzmann ideal gas limit of massless quarks and gluons and the yellow vertical band marks the crossover region [3].

The normalised pressure, energy density, and entropy density as a function of temperature are shown on the right panel of figure I.1.5. The steep rise of the thermodynamical quantities in the critical temperature region reflects the increase of the number of degrees of freedom. The quarks and gluons being confined in the hadron gas below T_c, do not contribute to the bulk thermodynamic of the system. As the temperature rise above T_c, quarks and gluons de-confinement leads to their

participation and they become the relevant degree of freedom. At low temperature, the lattice QCD results are compared to the hadron resonance gas (HRG) model, a thermal statistical model of non-interacting hadrons and resonances ([56, 57]).

On the right panel of figure I.1.5, the energy density of a non-interacting relativistic massless gas composed of three quark flavours and gluons is represented by a dashed line. The fact that even at $T \gg T_c$, the QGP thermodynamic quantities are far below this non-interacting parton gas limit shows that the QGP is still a strongly interacting medium. One of the main properties of the QGP formation is the rise of the colour screening with temperature leading to quarks and gluons de-confinement.

I.2 Ultra-relativistic heavy-ion collisions

The first experimental heavy-ion programs started in the mid 1980s at the Brookhaven National Laboratory (BNL, New York, USA) and at CERN (Geneva, Switzerland). Heavy-ion collisions enter in the ultra-relativistic regime for collisions of centre-of-mass energy per nucleon pair $\sqrt{s_{\rm NN}}$ above 10 GeV where the critical energy density $\epsilon_c \sim 1\text{GeV/fm}^3$ is reached allowing for the formation of a quark-gluon plasma [7].

The experimental signatures of the quark-gluon plasma were discovered and extensively studied in various heavy-ion colliding systems at the Super Proton Synchrotron (SPS, $\sqrt{s_{\rm NN}}$ = 6.2 - 17.3 GeV) [5] and at the Relativistic Heavy Ion Collider (RHIC, $\sqrt{s_{\rm NN}}$ = 7.0 - 200 GeV) [6–9]. The RHIC experiments characterised the quark-gluon plasma as a highly opaque medium behaving like a nearly perfect fluid. Studies of the QGP thermodynamic properties are still performed at RHIC and since 2009 also at the Large Hadron Collider (LHC) where lead-lead collisions up to $\sqrt{s_{\rm NN}}$ = 5.02 TeV are delivered. Beam energy scans from $\sqrt{s_{NN}}$ = 200 GeV down to 7.7 GeV performed at RHIC allowed to explore the QCD phase diagram from μ_B = 20 MeV to 400 MeV while searching for the critical end point. At LHC energies and top RHIC energies, the vanishing baryon chemical potential region is explored.

I.2.1 Heavy-ion collision space-time evolution

Relativistic fluid dynamics has been successfully used as an effective model to describe the space-time evolution of relativistic heavy-ion collisions. In the high energy collision regime, J. D. Bjorken proposed a space-time evolution scenario of the hadronic matter at central rapidity [58]. This evolution is schematised on the figure I.2.1, where the horizontal axis represents the longitudinal spatial direction of the beams (z) and the vertical axis the time direction. The time unit, the fm/c, correspond to roughly 10^{-23} second. The time before the collision is represented in the lower part of the figure. At τ = 0 fm/c the two heavy nuclei collide, the system created will go through different phases as described further below.

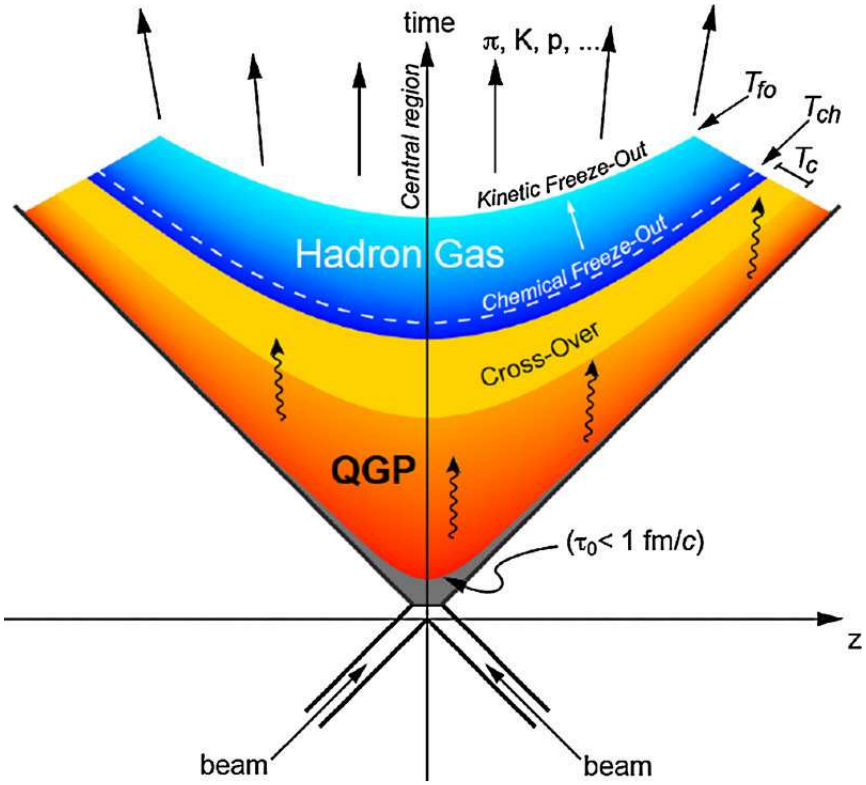


Figure I.2.1: Space-time evolution of a relativistic heavy-ion collision of two nuclei colliding at time t=0 and longitudinal position z=0 [59]. See text for a full description of the space-time evolution.

At centre of mass energy $\sqrt{s_{NN}}$ above 100 GeV, in the Bjorken regime, the particle production as a function of the rapidity y is assumed to be flat around $y = 0$. In the laboratory frame, the two highly Lorentz-contracted incoming nuclei are seen as flat disks. At LHC, where the lead beams are accelerated up to an energy of 2.51 TeV, the Lorentz γ factor is around 3000.

At the collision time τ_{coll}, the two nuclei overlap, multiple parton-parton interactions take place leading to particle production. The initial energy density reached at mid-rapidity is expressed as:

$$\epsilon_{BJ} = \frac{1}{A.\tau_0}\frac{dE_{\mathrm{T}}}{dy}\bigg|_{y=0}$$

with A the transverse overlap surface, τ_0 the formation time and $dE_{\mathrm{T}}/dy|_{y=0}$ the transverse energy rapidity density at mid-rapidity. With $\tau_0 \sim 1$ fm/c and using experimental measurement of the transverse energy rapidity density in the final state, at LHC, ϵ_{BJ} ~12-14 GeV/fm^3.

The energy density exceeding the critical value for colour de-confinement ($\sim$ 1 GeV/c), a partonic system is created, reaching a local thermal equilibrium at τ_0 ~1 fm/c. The thermalised hot partonic medium, the QGP, assimilated as a perfect fluid will undergo a longitudinal and radial hydrodynamic expansion of its original volume.

As the QGP expands it cools down. When the temperature of the medium drops below the critical temperature region, parton hadronisation starts and a hadron gas is formed. At chemical freeze-out, τ ~10 fm/c (T_{ch} on figure I.2.1), inelastic interactions between hadrons stops, fixing their relative abundances. The gas expansion continues, hadrons interacts elastically until kinematic freeze-out (T_{fo} on figure I.2.1) defining the final hadron momentum spectra. The hadrons

will stream freely reaching the first detectors. The time scale of the QGP formation and evolution is far below the capabilities of the detector technologies, therefore, the experimental QGP studies trough the produced particle out of the collisions are indirect.

Another exotic state of the hadronic mater studied in ultra-relativistic heavy-ion collisions is the Colour-Glass Condensate (CGC) [60]. The CGC describes the heavy-ion at high energy as a low-momentum gluon saturated system were gluon splitting and gluon recombination processes are balanced. The ultra-relativistic collision of two colour-glass sheets, as seen in the laboratory frame due to Lorentz contraction, leads to their melting into quarks and gluons and to the formation of the pre-equilibrated partonic phase preceding the QGP phase.

I.2.2 Heavy-ion collision geometry

The partonic medium de-confinement and thermalisation are driven by the energy density reached at collision time but one needs to find a way to study it experimentally. The charged particle production is a candidate to measure the event activity. Due to the finite size and the composite structure of the heavy-ion nuclei, it is natural to consider that the particle production depends on the geometry of the collision. The geometry of a collision is generally characterised by the impact parameter, the number of participating nucleons (N_{part}) and the number of inelastic binary nucleon-nucleon interactions (N_{coll}).

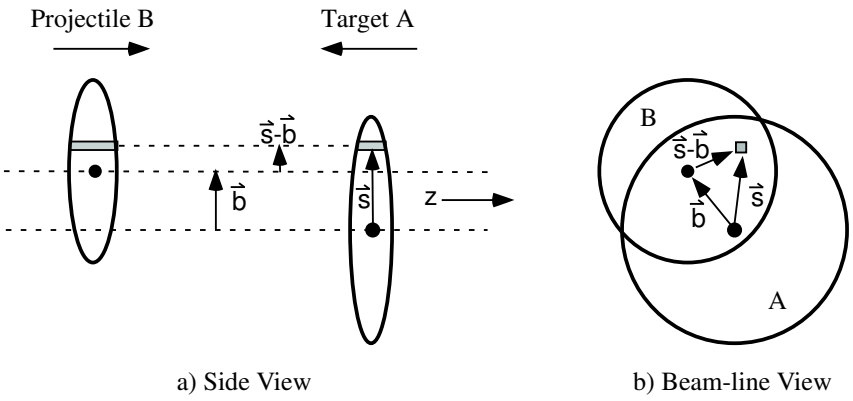


Figure I.2.2: Schematic representation of the Optical Glauber Model geometry [61]. Transverse view (a) and longitudinal view (b).

The Glauber models [61] describe the nucleus-nucleus interaction using considerations on multiple high energy scattering of their nucleons. They have been developed to extract geometry quantities from experimental data. The two main experimental inputs are the nuclear charge densities parametrised with a modified Woods-Saxon function and the inelastic nucleon-nucleon cross section.

In the optical limit approximation, a collision of two nuclei is described as a superposition of independent inelastic binary nucleon-nucleon interactions and the nucleon trajectories are not affected by the interactions. Additionally, as a nucleon might participate to several binary interactions, the cross-section does not depend on its previous number of interactions. On figure I.2.2, picturing a representation of the Optical Glauber Model geometry, the impact parameter is indicated by the vector $\vec{b}$ corresponding to the distance between the two centre of the two colliding nuclei. The position of two overlapping "flux tubes" (represented in grey) is indicated by the vector

$\vec{s}$. The thickness function $\widehat{T}_{AB}(\vec{b})$ corresponds to the effective overlap area in which nucleons from the target and the projectile can interacts:

$$\widehat{T}_{AB}(\vec{b}) = \int \widehat{T}_A(\vec{s})\,\widehat{T}_B(\vec{s}-\vec{b})\,d^2(s) \quad \text{with} \quad \widehat{T}_i(\vec{s}) = \int \widehat{\rho}_i(\vec{s},z_i)\,dz_i \qquad i = A \text{ or } B \text{ with } \vec{s} \leftarrow (\vec{s}-\vec{b})$$

$\widehat{T}_i(\vec{s})$ is the probability to find a nucleon from the nuclei i at a given position in the flux tube and $\widehat{\rho}_i(\vec{s},z_i)$ the corresponding nuclear charge density. In the optical limit, N_{part} and N_{coll} can be derived analytically ([61]):

$$N_{part}(\vec{b}) = A\int \widehat{T}_A(\vec{s})\left\{1-\left[1-\widehat{T}_B(\vec{s}-\vec{b})\,\sigma_{inel}^{NN}\right]^B\right\} d^2s + B\int \widehat{T}_B(\vec{s}-\vec{b})\left\{1-\left[1-\widehat{T}_A(\vec{s})\,\sigma_{inel}^{NN}\right]^A\right\} d^2s$$

$$N_{coll}(\vec{b}) = A\,B\,\widehat{T}_{AB}(\vec{b})\,\sigma_{inel}^{NN}$$

with A and B the number of nucleons in their corresponding nucleus.
The optical approach suffers from the treatment of continuous density distributions which are discrete in nature as the nucleon density distributions. As consequence, a description of the local density fluctuations is missing. This is tackled in the Glauber Monte Carlo approach, where nucleon positions are randomly sampled from the nucleon density distributions for each collisions. The impact parameter is also randomly sampled. By simulating an important number of collisions, one can determine the average number of participating nucleons, the average number of inelastic binary nucleon-nucleon interactions but also the charged particle multiplicity distribution.

Heavy-ion observables are often measured in different collision centrality intervals or classes. The centrality of the collision is characterised by the impact parameter, a 0% collision centrality corresponds to the full overlap of the two nuclei, collision centralities close to 100% correspond to peripheral collisions. Studies of the centrality in different heavy-ion collision systems are done by the ALICE Collaboration as presented in [62].

I.3 Experimental study of the quark-gluon plasma

In this section, a few experimental key signatures of the formation of the quark-gluon plasma in heavy-ion collisions are highlighted. This is not meant to be an exhaustive list of all the signatures, a focus is given on the experimental characterisation of the QGP bulk properties as in the particle production and strange particle production enhancement measurements and the hydrodynamics description of the hot medium. The study of the QGP microscopic properties through high-momentum parton energy losses is presented.

I.3.1 Particle production

The particle production measured in heavy-ion collisions are remarkably well described by statistical hadronisation models (SHM) over a wide range of collision systems and collision energies ($\sqrt{s_{\rm NN}} \simeq$ 2 GeV - 5.02 TeV). In such models, particle production is realised from the decays of fireball(s) at statistical equilibrium produced in heavy-ion collisions. The particle abundances being fixed at chemical freeze-out, the thermodynamic properties of the system at this stage are encoded in the particle yields.

The partition function of thermodynamic QCD is approximated in the framework of an ideal hadron-resonance gas treated in the grand canonical ensemble to account for conservation laws as the baryon number, the strangeness and the electric charge. This description is compatible with the equation of state of hadronic matter below the critical temperature derived from lattice QCD. The density of a given hadron specie is given as,

$$n_i(T,\vec{\mu}) = \frac{\langle N_i \rangle}{V} = \frac{T g_i}{2\pi^2} \sum_{k=1}^{\infty} \frac{(\pm 1)^{k+1}}{k} \lambda_i^k m_i^2 K_2(\frac{k m_i}{T})$$

- $\vec{\mu} = (\mu_B, \mu_S, \mu_Q)$ chemical potentials related to the baryonic number B, strangeness S and electric charge Q
- $\lambda_i(T,\vec{\mu}) = \exp \frac{B_i\mu_B + S_i\mu_S + Q_i\mu_Q}{T}$
- g_i the spin degeneracy factor of the hadron specie i
- m_i mass of the hadron specie i
- (+) for bosons, (-) for fermions
- $K_2(x)$ the modified Bessel function

In their most simple formulation, SHM models describe the hadron abundances with three free parameters: the temperature and volume of the system at chemical freeze-out and the baryon chemical potential μ_B. Additional free parameters, called γ saturation parameters, are added to the model to describe the deviation from chemical non-equilibrium of quark flavours as for strangeness γ_s and charm γ_c.

Particle production experimental data are fitted in order to estimate the temperature and the volume of the system at chemical freeze-out. The figure I.3.1 present thermal fits of identified particle yield measured in 0-10% central Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV performed with THERMUS [63], GSI-Heidelberg [64] and SHARE [65]. The chemical freeze-out temperature extracted, T $\sim$ 153 MeV, is consistent among the different implementations of the SHM model and close to the phase transition critical temperature $T_c \sim 155$ MeV from lattice QCD. For the presented thermal fits with THERMUS and GSI-Heidelberg models, μ_B is set to 0, the γ_s parameter is set to unity and γ_c is set to 20 to consider the charm yield abundance above hadron gas chemical-equilibrium. In the SHARE model, a light quarks (u, d) chemical non-equilibrium can be introduced by the $\gamma_q \neq 1$ parameter, however, in order to allow a fair comparison between the models, here it is fixed to 1 as for γ_s. Additionally to non-equilibrium thermodynamics, deviation from SHM results can be investigated in terms of inelastic hadron interactions leading to a sequential freeze-out of hadrons.

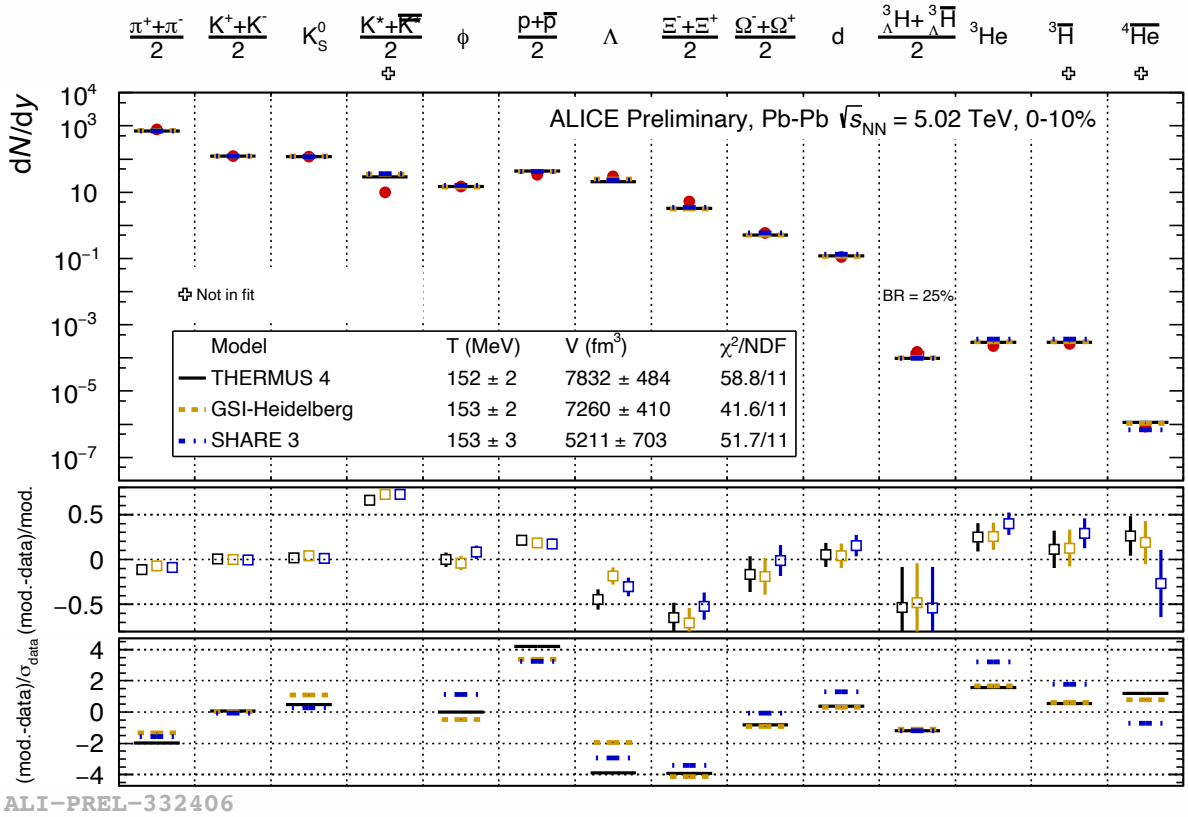


Figure I.3.1: Grand canonical thermal fit of identified particle yields measured in 0-10% central Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV with THERMUS [63], GSI-Heidelberg [64] and SHARE [65].

I.3.2 Strangeness enhancement

Strangeness enhancement was one of the first signature proposed for the formation of a QGP in heavy-ion collisions. In 1982, J. Rafelski and B. Müller [13] suggested that the QGP formation greatly affects the strange hadron production, particularly, multi-strange baryons. In heavy-ion collisions, as strange valence quarks are not present in the initial state (the two colliding nuclei), they are produced by the initial partonic interactions and during the latter QGP phase. At the lowest QCD order, strange quark pairs are produced in $q\bar{q} \rightarrow s\bar{s}$ and $g\bar{g} \rightarrow s\bar{s}$ processes. In a hot deconfined medium, the gluon fusion is the dominant process due to its higher production rate compared to quark annihilation. The relaxation time toward chemical equilibrium of $g\bar{g} \rightarrow s\bar{s}$ is comparable with the lifetime of the QGP (a few fm/c), for an energy density above $\sim$ 1 GeV/c, strange quarks reach chemical equilibrium. As a consequence, and with respect to hadron gas model calculations, strange baryon production is expected to be enhanced due to $s\bar{s}$ saturation in the QGP produced by deconfined thermal gluon fusion.

The first observation of multi-strange (anti-)baryons enhancement in heavy-ion (AA) collisions with respect to proton-A collisions were performed at SPS by the WA85 and WA94 Collaborations [14, 66, 67]. More recently, the ALICE Collaboration measured for the fist time an enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions [11]. The Figure I.3.2 presents ratios of p_T-integrated yields of strange and multi-strange hadrons over the charged pions as a function of the charged-particle density at mid-rapidity for different collision systems.

Globally all the hadron ratios increases continuously from low multiplicity density proton-proton collisions to lead-lead collisions following a universal trend. The amplitude of the enhancement grows with the strange quark content of the hadrons, this is seen for example by comparing the $\Lambda(uds)$ versus $\Omega(sss)$ ratios. Focusing on the proton-proton collision system, the yield ratios increases with the multiplicity density and are compatible with proton-lead results.

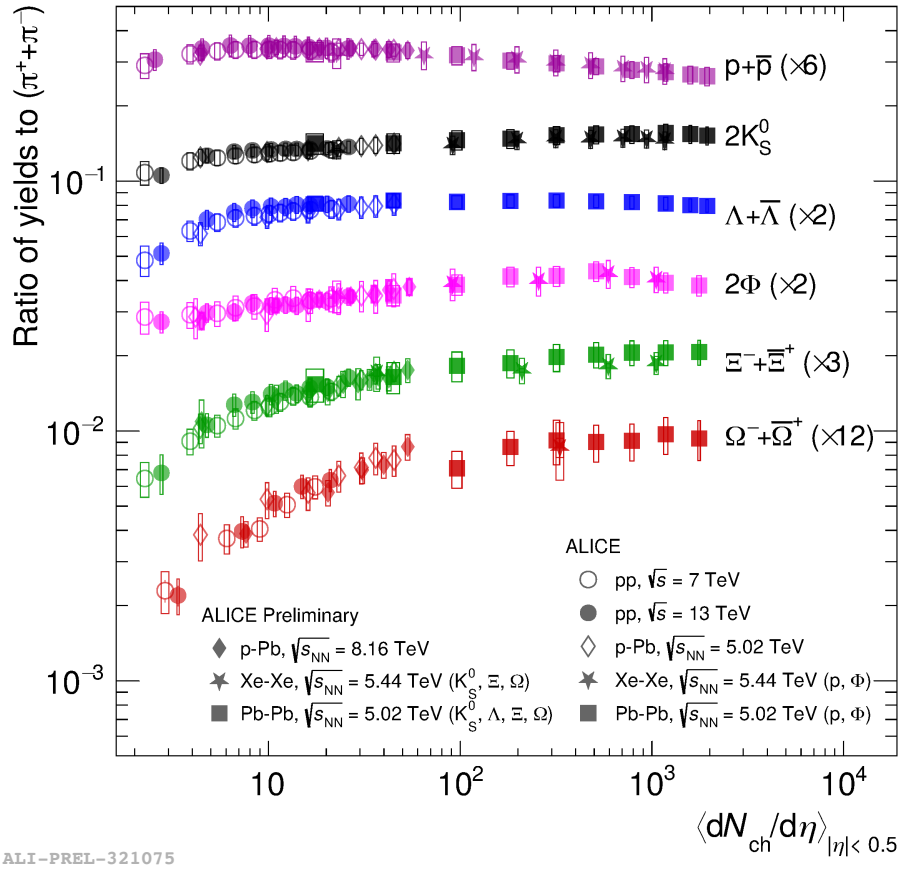


Figure I.3.2: Ratios of p_T-integrated yields of the proton, K^0_S, Λ, Ξ, Ω and ϕ over the charged pions as a function of the charged-particle density at mid-rapidity measured by the ALICE Collaboration for different collision systems and energies.

Strangeness production in proton-proton collisions is studied in Monte Carlo and statistical hadronisation models. In the microscopic Monte Carlo models as PYTHIA 8 [68], the hadronisation is performed via the string fragmentation model. The model describe the fragmentation of colour flux tubes arising from $q\bar{q}$ pair constituents going in opposite directions. In high multiplicity density collisions, strings can overlap to form colour ropes leading to an increase of the strangeness production with respect to low multiplicity collisions [69].

In the picture of the statistical hadronisation, strangeness is suppressed at low multiplicity. As described in section I.3.1, particle production in heavy-ion collisions is treated in the grand canonical statistical ensemble. In peripheral heavy-ion collisions or in small collision systems (proton-proton and proton-A), particle production is treated in the canonical ensemble to account for the low number of particle produced. In the canonical ensemble strangeness conservation must be exact, while in the grand canonical ensemble it is on average. Such strict restriction reduces the phase space for open strangeness production called the canonical suppression.

I.3.3 Particle flow

The quark-gluon plasma space-time evolution can be described by relativistic viscous hydrodynamics driven by the equation of state and transport coefficients of the system. The shear viscosity η characterising the resistance of the fluid to deformation and bulk viscosity ξ the resistance to the change of volume are two transport coefficients studied when describing the QGP as quasi-ideal fluid.

As discussed in section I.2.2, the geometry of an heavy-ion collision depends on the impact parameter. In non-central AA collisions, the initial spatial anisotropy and the initial fluctuations of the energy density are translated to an anisotropic pressure gradient due to the collective behaviour of the thermalised constituent of the QGP. The pressure gradient, mainly directed along the impact parameter axis, produce an anisotropic flow reflected later in the momentum azimuthal distribution of the produced hadrons. By performing a Fourier decomposition of the momentum dependent azimuthal distribution of final state particles one can extract the harmonic coefficients ν_{n} quantifying the magnitude of anisotropic flow effects.

$$E\frac{\mathrm{d}^3N}{\mathrm{d}p^3} = \frac{1}{2\pi}\frac{1}{p_{\mathrm{T}}}\frac{\mathrm{d}^2N}{\mathrm{d}p_{\mathrm{T}}\mathrm{d}y}\{1 + 2\sum_{n=1}^{\infty}\nu_{\mathrm{n}}\cos(\mathrm{n}(\phi - \Psi_{\mathrm{n}}))\} \qquad \nu_{\mathrm{n}} = \langle\cos(\mathrm{n}(\phi - \Psi_{\mathrm{n}}))\rangle$$

with E the energy, p the momentum, p_T the transverse momentum , ϕ the azimuthal angle, y the rapidity of the particle. If there are no fluctuations of the initial geometry, i.e. different initial state shapes for the same impact parameter, the n-th harmonic collision symmetry plane angle Ψ_{n} corresponds to the reaction plane angle, defined by the impact parameter and the beam direction vectors. While Ψ_{n} is measured event per event, the harmonic coefficients ν_{n} are averaged over the number of particle in a given rapidity and p_{T} interval and over the events of a given centrality class. The harmonic coefficients ν_{n} and the collision symmetry planes are determined experimentally using different methods as reviewed in [70].

The asymmetries of non-central AA collisions contributes mostly to the second order coefficient ν_2, called the elliptic flow, due to the elliptic shape of the initial overlapping region. Higher order coefficients corresponds to higher-order deformations. The figure I.3.3 present an elliptic flow measurement as a function of p_{T} for different identified hadrons performed in different centrality classes in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. From central (0-1%) to semi-central (30-40%) Pb-Pb collisions, the initial spatial anisotropy get more pronounced resulting in an overall increasing $\nu_2(p_{\mathrm{T}})$. As the collisions get more peripheral, the density of the system decreases, the pressure gradient are weaker resulting in a smaller $\nu_2(p_{\mathrm{T}})$. Across centrality classes, three p_{T} regions of interest highlight different physical mechanisms in play. The radial expansion of the QGP phase cause an isotropic boost of the velocity of all resulting hadrons, the resulting radial flow explain the mass-ordered $\nu_2(p_{\mathrm{T}})$ mass ordering in the $p_{\mathrm{T}} <$ 2-3 GeV/c region. Heavier hadrons are further pushed to high p_{T} compared to lighter hadrons. For 3 $< p_{\mathrm{T}} <$ 8-10 GeV/c, the $\nu_2(p_{\mathrm{T}})$ comes in two groups depending on their constituent quark number, the baryon/meson grouping. This is explained by the quark coalescence hadronisation mechanism in which quarks close in space and momenta form hadrons. Quark coalescence is an indication of collective behaviour. At higher p_{T} ($>$ 10 GeV/c), the p_{T} dependence is weaker, all particle $\nu_2(p_{\mathrm{T}})$ seems to be compatible within uncertainties.

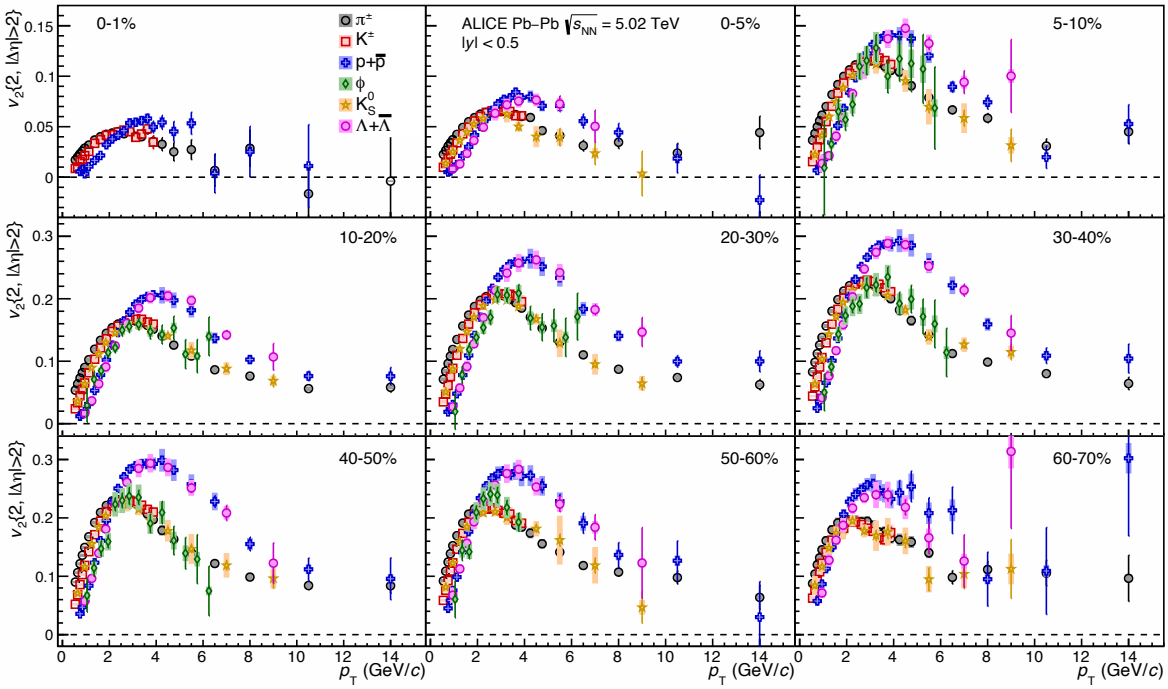


Figure I.3.3: Elliptic flow coefficient as a function of p_{T} for pion, kaons, protons, ϕ, K^0_s and Λ in different centrality classes in Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV, measured by the ALICE Collaboration [71].

Measurements of the harmonic coefficients are compared to relativistic viscous hydrodynamics model predictions from which properties of the medium can be extracted as the equation of states and transport coefficients. A shear viscosity over entropy density ratio η/s of about 0.12 is used to reproduce Au-Au collisions at RHIC energies results and $\eta/s = 0.2$ for Pb-Pb collisions at LHC energies results [72].

Recent measurements has been made in proton-proton and proton-A collisions allowing the study of collective effects in small systems where the formation of a QGP is not expected. In high multiplicity proton-proton collisions, a ridge-like structure emerge in long range azimuthal two-particle correlations [73] similar as the one observed in AA collisions. Non-zero values of anisotropic flow coefficients are measured in small systems as a function of multiplicity [10]. The elliptic flow $\nu_2(p_{\mathrm{T}})$ measured as a function of p_{T} in p-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV [74] present the mass ordering behaviour at low p_{T} followed by the baryon/meson grouping at intermediate p_{T} both indicating collective effects. Flow measurement are also performed at RHIC for various proton-A collision systems at $\sqrt{s_{\mathrm{NN}}} = 200$ GeV [75] where a dependence on the system size is highlighted. The measured flow coefficients are compatible with predictions from relativistic viscous hydrodynamic evolution of the hot medium with initial state fluctuation [76] and with initial state colour correlations computed in the Colour-Glass Condensate framework (CGC) [77].

I.3.4 High p_{T} parton in-medium energy losses

Large transverse momentum quarks and gluons are produced in hard-scattering processes taking place at the early stages of the collision before the formation of the QGP. These off mass-shell high-p_{T} partons radiate gluons as their virtuality decreases developing a parton shower.

When non-perturbative scale is reached (of the order of Λ_{QCD}), hadronisation take place turning the parton shower into a collimated jet of hadrons.

Parton shower constituents loose energy as they interact with the hot partonic medium via multiple scattering and medium-induced gluon radiation processes. The hard parton momentum broadening (scattering) and energy loss (radiative emission) are governed by the transport coefficient $\hat{q}$ also called the jet quenching parameter, defined as the transverse momentum squared per unit length exchanged between the medium and the traversing hard parton. The resulting modification of the energy distribution among the jet partons by the hot and dense medium cause a jet quenching effect compared to the corresponding unmodified jet propagation in the vacuum. Jet quenching is studied experimentally with the jet nuclear modification factor defined as,

$$R_{\rm AA} = \frac{\frac{1}{N_{event}} \frac{{\rm d}^2N}{dp_{\rm T,jet}{\rm d}\eta_{\rm jet}}\Big|_{\rm AA}}{\langle T_{\rm AA}\rangle \frac{{\rm d}^2\sigma}{dp_{\rm T,jet}{\rm d}\eta_{\rm jet}}\Big|_{\rm pp}} \quad , \text{with } \langle T_{\rm AA}\rangle = \frac{\langle N_{\rm coll}\rangle}{\sigma^{NN}_{inel}} \tag{I.1}$$

with $\frac{{\rm d}^2\sigma}{dp_{\rm T,jet}{\rm d}\eta_{\rm jet}}\Big|_{\rm pp}$ the $p_{\rm T,jet}$ and $\eta_{\rm jet}$ differential jet cross-section in proton-proton collisions, $\langle T_{\rm AA}\rangle$ the nuclear overlap density defined as the ratio of the average number of binary nucleon-nucleon collisions over the inelastic nucleon-nucleon cross-section and $\frac{{\rm d}^2N}{dp_{\rm T,jet}{\rm d}\eta_{\rm jet}}\Big|_{\rm AA}$ the jet corrected spectra.

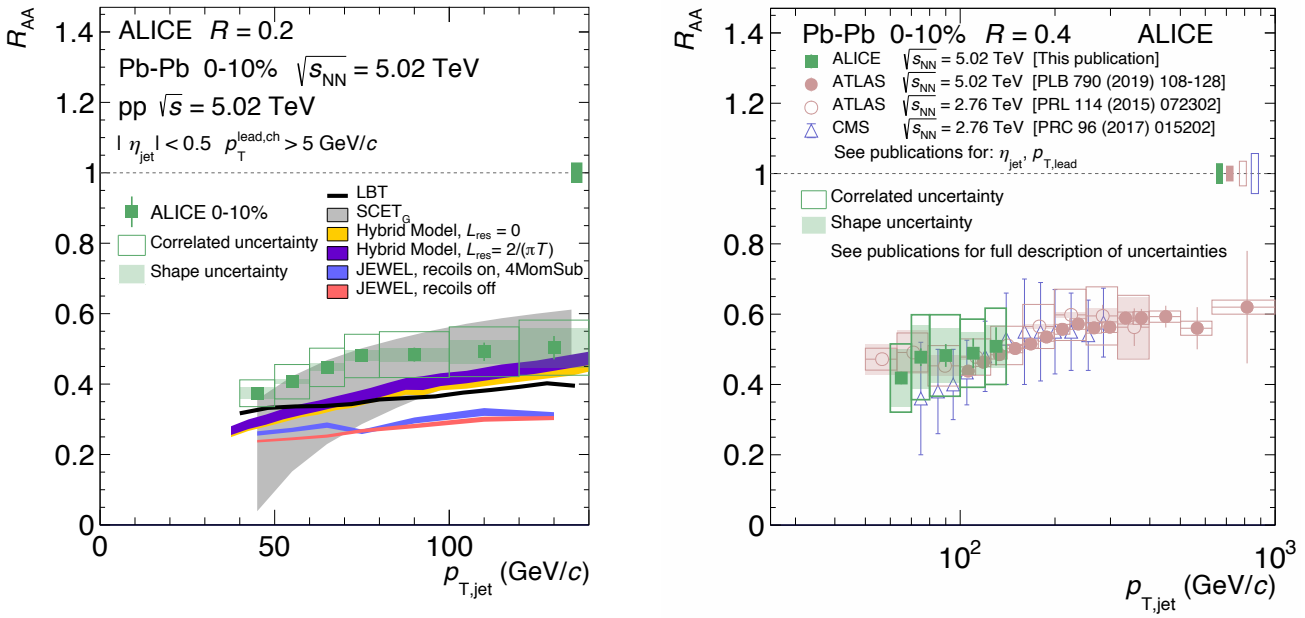


Figure I.3.4: Left panel: Jet $R_{\rm AA}$ for R = 0.2 in 0-10% central Pb-Pb collisions at $\sqrt{s_{\rm NN}}$ = 5.02 TeV measured by the ALICE Collaboration [78] and compared to LBT, SCETG, Hybrid model and JEWEL predictions. Right panel: Jet $R_{\rm AA}$ in 0-10% central Pb-Pb collisions at $\sqrt{s_{\rm NN}}$ = 5.02 TeV for R = 0.4 measured by the ALICE, CMS and ATLAS collaborations [78].

The figure I.3.4 present the jet $R_{\rm AA}$ measurement performed by the ALICE collaboration in 0-10% central Pb-Pb collisions at $\sqrt{s_{\rm NN}}$ = 5.02 TeV [78]. On the left (right) panel, the hard jet in Pb-Pb collisions are identified by requiring at least one charged track with $p_{\rm T} > 5$ (7) GeV/c for a jet resolution parameter R = 0.2 (0.4). The results are compared with different models: the

Linear Boltzmann Transport (LBT) model [79, 80], Soft Collinear Effective Theory with Glauber gluons (SCET_G) [81, 82], the Hybrid model [83–86] and JEWEL [87, 88].

For the hole p_{T} range covered, $\sim$ 30 to 900 GeV/c, the jet nuclear modification factor is below unity indicating the jet yield suppression in Pb-Pb with respect to the proton-proton measurement. The R_{AA} evolution as a function of the jet p_{T} shows that high p_{T} jet are less sensible to in-medium energy losses compared to low p_{T} jet. The models calculations predict also the jet suppression and produce the same jet p_{T} evolution of the R_{AA}.

Chapter II

Open heavy-flavour physics overview

In this chapter, the motivation to study the open heavy-flavour hadrons, *i.e.* particles made of at least a heavy quark and lighter quark(s), in hadronic collisions are discussed. The heavy quarks are produced in hard scattering processes occurring in the early stage of the collision making them sensitive to the full evolution of the system created by the collision. The description of the mechanisms of heavy quark production and hadronisation within the perturbative QCD framework and in Monte Carlo generators are presented. Measurements of open heavy-flavour hadrons production in the three main collisions systems delivered by the LHC, *i.e.* proton-proton, proton-Pb and Pb-Pb, and their comparison to models are discussed.

II.1 From heavy quarks to heavy-flavour hadrons

The bare masses of the charm, bottom and top quarks ($m_{\rm Q}$, Q = c, b, t) are large with respect to the Λ_{QCD} (∼200 MeV) scale. As the coupling constant $\alpha_s(m_{\rm Q}) \propto \ln^{-1}(m_{\rm Q}^2/\Lambda_{\rm QCD}^2)$ at the heavy quark mass energy scale is well below unity (asymptotic freedom regime), the perturbative QCD theory can be used to make quantitative predictions on their production [89]. Due to their large masses, heavy quark production cross section calculations are performed down to low heavy quark transverse momenta $p_{\rm T,Q}$ providing total and $p_{\rm T,Q}$ differential cross sections. The access to theoretical predictions from perturbative QCD allow us to verify its validity with experimental measurements.

In the heavy-ion collisions context, heavy quark production occurs at time scales t $\sim 1/2m_{c,b,t}$ (0.07 fm/c for the charm and 0.02 fm/c for bottom) shorter than the quark-gluon plasma formation time (∼ 0.1-1 fm/c). At LHC, in central Pb-Pb collisions, the temperature of the QGP created is about 300 MeV which is below the energy needed to produce heavy quark pairs from thermal partons [90]. Since the total heavy quark production receive negligible contribution from thermal production in the QGP and initial hard scattering are not affected by the QGP formation, heavy quark production is independent from the QGP phase. As briefly introduced in section I.3.4, partons produced in hard scattering processes are expected to loose energy by interacting with the QGP, making the heavy quark an external probe of the hot deconfined medium.

Heavy quark hadronise into heavy-flavour hadrons. Open heavy-flavour hadrons are composed by charm or bottom quarks and lighter quarks (u, d, s), as for example, the D mesons: $\rm D^0(\bar{u}c)$, $\rm D^+(\bar{d}c)$, $\rm D_s^+(\bar{s}c)\ldots$, the B mesons: $\rm B^0(d\bar{b})$, $\rm B^+(u\bar{b})$, $\rm B_s^0(s\bar{b})\ldots$, and baryons: $\Lambda_c^+(udc)$, $\Lambda_b^0(udb)$, $\Sigma_c^0(ddc)$, $\Sigma_b^+(uub)$, $\Xi_c^+(usc)\ldots$ Hidden heavy-flavour hadrons, also called quarkonia, are bound states of one $c\bar{c}$ pair (J/Ψ, $\Psi(2S)\ldots$) or one $b\bar{b}$ pair ($\Upsilon(1S)$, $\Upsilon(2S)\ldots$).

The different heavy quark energy loss and hadronisation mechanisms at interplay, greatly

depend on the surrounding partonic and hadronic environment produced by the collision system. Such mechanisms are characterised by studying inclusive and differential heavy-flavour hadron production in different collision systems, from proton-proton to heavy-ion collisions, and as a function of the event activity, leading to the verification of the QCD theory and the characterisation of the quark-gluon plasma properties.

II.2 Open heavy-flavour hadroproduction

The theoretical calculation of heavy-flavour hadroproduction cross section is a multi-energy scale problem requiring the description of the initial state, *i.e.* the partonic content of two colliding hadrons, the heavy quark production and its further hadronisation. The perturbative QCD theory is employed to describe the heavy quark production in hard parton scattering processes. However, the initial state and the heavy quark hadronisation processes are characterised by a lower-energy scale, making the comparison of theoretical calculations with cross section measurements not straight-forward. The connection between perturbative QCD and experiment is made either by constructing infrared safe observables, *i.e.* independent from the soft physics, or relying on factorisation theorems, separating the long (non perturbative) and short (perturbative) distance physics [91]. The collinear factorisation theorem gives the general expression of the differential cross section of the inclusive open heavy-flavour H hadroproduction as the following convolution [92]:

$$\begin{aligned} \mathrm{d}\sigma^{\mathrm{A+B\to H+X}} \simeq \sum_{i,j,k} & f_i^{\mathrm{A}}(x_1, \mu_{fact}^2) \otimes f_j^{\mathrm{B}}(x_2, \mu_{fact}^2) \\ & \otimes \mathrm{d}\hat{\sigma}_{ij\to k\mathrm{X}}(\tau_1, \tau_2, \alpha_s(\mu_R), \mu_{fact}^2, \mu_{frag}^2) \\ & \otimes D_k^{\mathrm{H}}(z, \mu_{frag}^2) \end{aligned} \tag{II.1}$$

The sum is performed over all $i + j \to \mathrm{k} + \mathrm{X}$ possible partonic processes, where i, j and k can be gluons, light quarks and heavy quarks.
• $f_i^{\mathrm{A}}(x_1, \mu_{fact}^2)$ the parton distribution function at the factorisation scale μ_{fact}, describing the probability to find a parton inside the hadron A, carrying a fraction x_1 of the initial hadron momentum.
• $\mathrm{d}\hat{\sigma}_{ij\to k\mathrm{X}}(\tau_1, \tau_2, \alpha_s(\mu_R), \mu_{fact}^2, \mu_{frag}^2)$ the heavy quark partonic cross section at a given fixed order, $\tau_1 = \frac{p_{\mathrm{T,k}} e^{-y}}{x_2\sqrt{s}}$, $\tau_2 = \frac{p_{\mathrm{T,k}} e^{y}}{x_1\sqrt{s}}$ with p_{T}, y the transverse momentum and rapidity of the parton k and $\sqrt{s}$ the centre of mass energy of the collision. The term also depends on the factorisation scale μ_{fact}, the fragmentation scale μ_{frag} and the strong coupling constant α_s evaluated at the renormalisation scale μ_R. In the partonic kinematics, the masses of the partons are neglected, however, when a heavy quark pair is produced from light partons, the heavy quark mass m_{Q} is accounted by replacing $p_{\mathrm{T,k}}$ by $m_{\mathrm{T}} = \sqrt{p_{\mathrm{T,Q}}^2 + m_{\mathrm{Q}}^2}$.
• $D_k^{\mathrm{H}}(z, \mu_{frag}^2)$ the heavy quark fragmentation function describe the probability of the parton k fragmenting into the heavy-flavour hadron H at the fragmentation scale μ_{frag} with $z = E_{\mathrm{H}}/E_k$, the fraction of partonic energy carried by the outgoing hadron.

The parton distribution functions and the fragmentation function are considered to be universal, *i.e.* they do not depend on the partonic process, here the hard scattering leading to heavy quarks production. As they are non-perturbative terms, both functions receive input from experiments. The factorisation μ_{fact} and fragmentation μ_{frag} scales defines the boundary between the short and long distance physics. The ultraviolet divergences coming from heavy flavour quark production loop diagrams, integrating over infinite momentum space, are removed by applying a

renormalisation procedure to the theory at the renormalisation scale μ_R. The $\mu_{R,fact,frag}$ scale values are set around the relevant energy scale of the hard process, $\mu_R = \mu_{fact} = \mu_{frag} = \sqrt{p_{\mathrm{T,Q}}^2 + m_{\mathrm{Q}}^2}$ is a common default choice. As they scales are not part of the QCD theory but are artefacts of the renormalisation and factorisation procedures, the calculated heavy-flavour production cross section must be scale independent. This is the case if all orders of the perturbative expansion are included in the calculation. In fixed order perturbative QCD, the perturbative series is truncated at a given fixed order resulting in a residual scale dependence of the numerical result. The scale dependence of the calculated cross section is controlled by varying the different scales independently around the default value measuring the stability of the result.

II.2.1 Parton distribution functions

In the parton model, the parton distribution functions (PDF) are interpreted as probability densities for finding a parton with a certain fraction of the proton momentum (see section I.1.2). When considering perturbative QCD corrections beyond leading order coming from initial state collinear parton emission, PDF get redefined and acquire an energy scale dependence. In the collinear QCD factorisation framework, PDF represent the internal structure of hadron, describing the momentum distributions of the quarks and gluons hadron constituents. They depends on two variables: the momentum fraction x and the momentum transfer energy scale $\mathcal{Q}^2$. In perturbative QCD, their scale evolution is governed by the Dokshitzer, Gribov, Lipatov, Altarelli and Parisi (DGLAP) renormalisation equations, a set of $2n_f + 1$ equations of the form [39–42]:

$$\mathcal{Q}^2 \frac{\partial f_i(x_i, \mathcal{Q}^2)}{\partial \mathcal{Q}^2} = \sum_{f}^{2n_f+1} \int_x^1 \frac{dy}{y} P_{f'f}(y, \mathcal{Q}^2)\, f_j(\frac{x}{y}, \mathcal{Q}^2) \tag{II.2}$$

with n_f the number of quark flavour considered, $P_{f'f}(y, \mathcal{Q}^2)$ the splitting function interpreted as the probability that a parton j radiate a collinear parton i carrying a fraction y of the parton j original momentum.

The parton distribution function x-dependence can be determined by performing a global fit analysis on measurements using QCD factorisation theorems or obtained from lattice QCD calculations based on QCD first principles [93]. Currently, the global fit analysis method gives much more precise results and it is widely used. As heavy-flavour production predictions discussed later in this chapter are derived using PDF from a global fit analysis, the approach is briefly explained.

In global fit analysis, the PDFs are first parametrised, trying to describe their momentum fraction x-dependence at a given initial scale $\mathcal{Q}_0$ ($\sim$ 1-2 GeV). Two popular sets of PDFs are determined by the CTEQ global fit analysis [94] using flexible functional forms and the NNPDF Collaboration [95] which use artificial neural networks for the PDF parameterisations. Each quark flavour and gluon PDFs are then evolved up to an energy scale of interest using the DGLAP equations. The evolved PDF together with prediction calculations from perturbative QCD are convoluted into observables (using factorisation) to be compared to measurements. The predictions are then fitted to data, the optimal PDF parameters are determined via a χ^2 minimisation procedure. Different measurement datasets are used to constrain all PDF flavours in the different x kinematic ranges. The main ones are structure functions from deep inelastic scattering, neutral-current ($e^{\pm}p \to e^{\pm}X$) and charged-current ($e^{\pm}p \to \bar{\nu_e}(\nu_e)X$) cross sections, Drell-Yan production ($q\bar{q} \to l^+l^-$), weak boson production and inclusive jet productions measured in fixed-target, HERA I+II, Tevatron and LHC experiments.

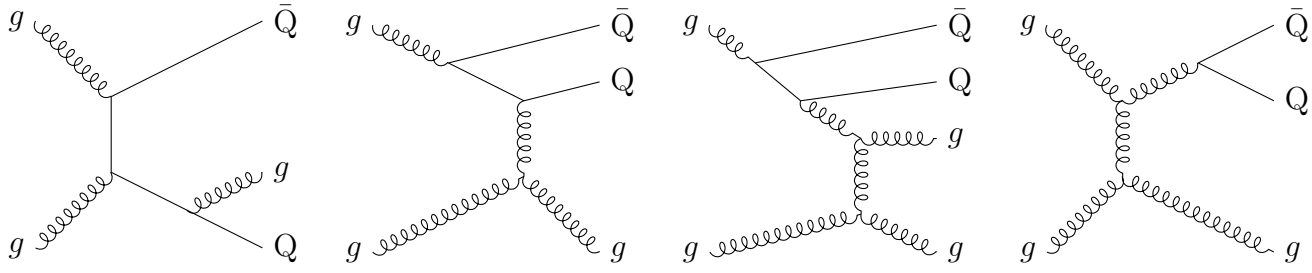
The treatment of heavy quarks in the initial state, *i.e.* heavy flavour PDF, is necessary for the complete description of their overall contribution to a given observable. Since these consideration depend on the energy scale of the process, to describe them over the full kinematic range from $\mathcal{Q}^2 < m_{\mathrm{Q}}^2$ to $\mathcal{Q}^2 \gg m_{\mathrm{Q}}^2$, different formalisms are employed where heavy quarks can be considered as massive or massless particles. For processes occurring below and near the heavy quark mass threshold, $\mathcal{Q}^2 \lesssim m_{\mathrm{Q}}^2$, the heavy quarks are considered as massive final state particles and not as initial hadron constituents. In the $\mathcal{Q}^2 \gg m_{\mathrm{Q}}^2$ regime, the heavy quarks are treated as massless particles as the light flavour quarks and are considered as initial partons. Their distributions are not extracted from data but are generated via radiated gluon splitting according to the DGLAP evolution of light flavour quark and gluon distributions. This is a purely extrinsic perturbative contribution. The charm and bottom non-perturbative intrinsic component of the hadron are also studied [96].

The parton distribution function uncertainties originates from the statistical and systematic uncertainties of experimental data, the global fit analysis methodology and from theoretical systematic uncertainties. The uncertainties from the PDF parameterisation and the fitting procedure are kept under control by performing closure tests [97]. The choice of the renormalisation scheme and the perturbative order in the perturbative calculation make the largest contributions to theoretical uncertainties [98].

II.2.2 Heavy quark production

Any physical quantity calculated in perturbative QCD is expressed as a perturbative series of $\alpha_s(\mu_R)$ expanded to all orders. The heavy quark partonic cross section is characterised by the heavy quark mass m_{Q} and the hard scattering energy scale identified with the transverse momentum of the heavy quark $p_{\mathrm{T,Q}}$.

At the leading order (LO) $\mathcal{O}(\alpha_s^2)$, heavy quarks are produced from gluon fusion, $gg \to \mathrm{Q\bar{Q}}$ and quark-antiquark $q\bar{q} \to \mathrm{Q\bar{Q}}$ annihilation processes. At LHC energies, the inclusive heavy quark production cross section is dominated by the gluon fusion channel (∼85%) [99]. Next to leading order (NLO) contributions $\mathcal{O}(\alpha_s^3)$ are also included in the calculation as the heavy quark pair creation with an emitted gluon in the final state (figure II.2.1a); flavour excitation, *i.e.* the scattering of a heavy flavour quark coming from the parton distribution of one hadron with a parton from the other hadron (figure II.2.1b); gluon splitting in the initial and final states, represented on figures II.2.1c and II.2.1d respectively; $q\bar{q} \to \mathrm{Q\bar{Q}}g$, $gq \to \mathrm{Q\bar{Q}}g$ and virtual loop corrections to the LO processes. The LO and NLO contributions constitute the foundation of all the current heavy quark hadroproduction predictions from semi-analytical approaches and Monte Carlo generators. More recently, next-to next-to leading order (NNLO) contribution calculations have been performed for the bottom [100] and top [101] quarks.



(a) pair creation with gluon emission (b) flavour excitation (c) gluon splitting of flavour excitation character (d) gluon splitting

Figure II.2.1: Example of next to leading order heavy flavour quark production diagrams.

The NLO contributions to the cross section, in form of leading logarithm (LL) terms $\alpha_s^n \ln^n(p_{\mathrm{T,Q}}^2/m_{\mathrm{Q}}^2))$ and next to leading logarithm (NLL) term $\alpha_s^n \ln^{n-1}(p_{\mathrm{T,Q}}^2/m_{\mathrm{Q}}^2))$ appear at all orders of the calculation. As discussed in the section II.2.1, the treatment of heavy quarks in factorisation theorems is energy-scale dependent, with the three main kinematic regimes of heavy quark production defined as, $p_{\mathrm{T,Q}} < m_{\mathrm{Q}}$, $p_{\mathrm{T,Q}} \sim m_{\mathrm{Q}}$ and $p_{\mathrm{T,Q}} \gg m_{\mathrm{Q}}$.

Near the quark mass threshold, $0 \leq p_{\mathrm{T,Q}} \lesssim m_{\mathrm{Q}}$, the heavy quark production is described in the n_f-Fixed Flavour Number (n_f-FFN) schemes, for $\mathrm{Q} = \{c, b, t\}$, $n_f = \{3, 4, 5\}$ respectively. A heavy quark is considered as a massive final state particle in the perturbative calculation while the light flavour quarks are treated as massless and are considered as active partons in the initial state. In these schemes, flavour excitation and initial state gluon splitting processes are suppressed. Because the LL and NLL terms are large, the mass dependence of the calculation becomes divergent as $p_{\mathrm{T,Q}}$ increases above m_{Q}^2. At $p_{\mathrm{T,Q}} \gg m_{\mathrm{Q}}$, the NLO gluon fusion processes engender different sources of collinear divergences: the collinear gluon emission from a final state heavy quark and gluon splitting into a collinear heavy quark pair both in the initial and final states. The corresponding LL and NLL terms become quite significant in this regime, limiting the precision on the n_f-FFN calculations. The convergence of the calculation is regained by resumming the logarithm series to all orders using DGLAP equations and absorbed in universal terms. Heavy-quark parton distribution functions are therefore introduced resumming initial state collinear divergences while final state divergences are absorbed in the fragmentation functions.

The full description of heavy quark hadroproduction from low to high momenta including heavy quark mass(es) and the resummation of divergent terms are performed in Variable Flavour Number (VFN) schemes. The VFN schemes are composed by a sequence of n_f-FFN schemes with different number of active quark flavours in distinctive energy scale regions. Each scale region is matched to the next one at its boundary, defined by the matching scale μ_{n_f} between the n_f and n_f+1 schemes. At the charm quark transition point from $n_f = 3$ to $n_f = 3$+1, the matching scale $\mu_{n_f=3}$ is set around the charm quark mass; at the bottom quark transition point, $n_f = 4$ to $n_f = 4$+1, $\mu_{n_f=5} \sim m_b$. The heavy quark mass effects are therefore introduced at each transition point in these perturbative calculations. The simplest realisation of the VFN schemes is done in the approximation were all quarks are considered massless, the Zero Mass Variable Flavour Number (ZM-VFNS) scheme covers the $p_{\mathrm{T,Q}} \gg m_{\mathrm{Q}}$ kinematic region.

II.2.3 Heavy quark fragmentation

The fragmentation is a particular mechanism of the hadronisation of a coloured parton into a colourless hadron. This mechanism is described by the non-perturbative heavy quark fragmenta-

tion function, $D_{\mathrm{Q}}^{\mathrm{H}}(z)$. Examples of commonly accepted parameterisations of this function can be found in [102–107], the function parameters are extracted by performing fits to experimental data on electron-positron annihilation, semi-inclusive deep inelastic scattering and hadron production measurements in high-energy hadronic collisions. A significant part of the charm and bottom production data is provided by the OPAL [108], ALEPH [109], DELPHI [110], BABAR [111] and BELLE [112] Collaborations.

As discussed in section II.2.2, final collinear divergences rising in the hard scattering cross section at $\mathcal{Q}^2 \gg m_{\mathrm{Q}}^2$ must be dealt with. The heavy quark fragmentation can be described in a formalism where the fragmentation is a two-steps process [113, 114]. First the perturbative off-shell parton fragmentation into a heavy quark and then the non-perturbative heavy quark fragmentation into a hadron. In this formalism, all quark flavours are considered as massless in the scattering process. As their virtuality decreases by gluon emissions, they fragment into massive heavy quarks. This transition is described by the perturbative fragmentation functions (PFF) $D_k^{\mathrm{Q}}(z, \mu_{frag})$, derived from perturbative QCD. The parton k can be a gluon, a light- or a heavy- flavour quark. Their evolution driven by DGLAP equations resums to all order the final state collinear gluon emission logarithm terms. The final FF are taken as the factorisation of the perturbative and non-perturbative contributions, $D_k^{\mathrm{H}}(z, \mu_{frag}) = D_k^{\mathrm{Q}}(z, \mu_{frag}) \otimes D_{\mathrm{Q}}^{\mathrm{H}}(z)$.

In the Binnewies, Kniehl and Kramer (BKK) approach [115], the non-perturbative parton FF $D_k^{\mathrm{H}}(z, \mu_{frag})$, is parametrised at an initial scale $\mu_0 = m_c$ for gluon, light flavour and charm quark and to $\mu_0 = m_b$ for the bottom quark. Then it is extrapolated to higher energy scales using DGLAP equations and fitted to experimental data applying a procedure similar to what is done for parton distribution functions.

Beside semi-analytical calculations, the perturbative contribution to the parton fragmentation are also modelled in Monte Carlo generators. The parton shower describes the successive parton splitting $q \to qg$, $g \to gg$ and $g \to q\bar{q}$ in a probabilistic approach. The splitting functions describing the processes are obtained as solutions of the DGLAP equations. Splitting probabilities are computed by summing the splitting functions for all possible final states over a given $\mathcal{Q}^2$ range. This iterative approach is applied for complex partonic system calculations such as hadron-hadron collisions, especially for the description of secondary partons which are relatively soft and nearly collinear to the primary partons. As the parton shower branches, the parton virtuality decreases down to the virtuality cutoff $\mathcal{Q}_0 \sim 1$ GeV where the parton shower stops and the non-perturbative hadronisation of the partons takes over.

In proton-proton collisions, fragmentation is presumably the dominating mechanism of hadronisation. Two popular approaches are commonly used in Monte Carlo generators. The string fragmentation hadronisation model [68, 116, 117] describe the dynamics of a string representing the colour flux between a quark-antiquark pair. Considering a colour connected quark-antiquark (colour dipole) and supposing a linear confinement, at large distance r, the inter-quark potential is written as V(r) = Kr, representing a string with tension K $\sim$ 1GeV/c. As the quarks are moving away from each other, the string tension grows up to a point where the string breaks by producing quark-antiquark pairs via tunnelling. Strings close in space can interact and form ropes. In the case of multiple hard scattering, partons in the final state coming from the different processes can form colour connections during hadronisation [118]. The cluster model [119–121] considers isotropic soft gluons splitting into quark-antiquark pairs at the end of the parton shower. The pairs form colour singlet clusters whose mass distribution is independent from the hard scattering process. Quark-antiquark clusters which can be seen as hadron resonance, decays into final hadrons or splits into lighter clusters before decaying.

Contrary to semi-analytical approaches, Monte Carlo generators aim to describe the full partonic and hadronic evolution of a collision and, in the case of heavy-flavour hadroproduction, provide tools for studying the heavy quark hadronisation.

II.2.4 Measurements of heavy-flavour production cross sections

On figure II.2.2 the transverse momentum (p_{T}) differential production cross sections for the prompt and non-prompt D^0, D_s^+ mesons in proton-proton collisions at $\sqrt{s} = 5.02$ TeV measured by the ALICE Collaboration [122] are presented and compared to predictions from semi-analytical calculations.

Both the FONLL [123–125] and the GM-VFNS [126–128] formalisms perform a match of the massive FFN scheme and the massless ZM-VFN scheme for a description of heavy quark production over the whole $\mathcal{Q}^2 \lesssim m_{\mathrm{Q}}^2$ to $\mathcal{Q}^2 \gg m_{\mathrm{Q}}^2$ range. The main differences between the two approaches are found in their matching procedure between the relatively low and high energy scale regimes of heavy quark hadroproduction and the heavy quark fragmentation where FONLL uses the "two-steps" fragmentation approach while GM-VNFS implement the extended BKK approach (see section II.2.3).

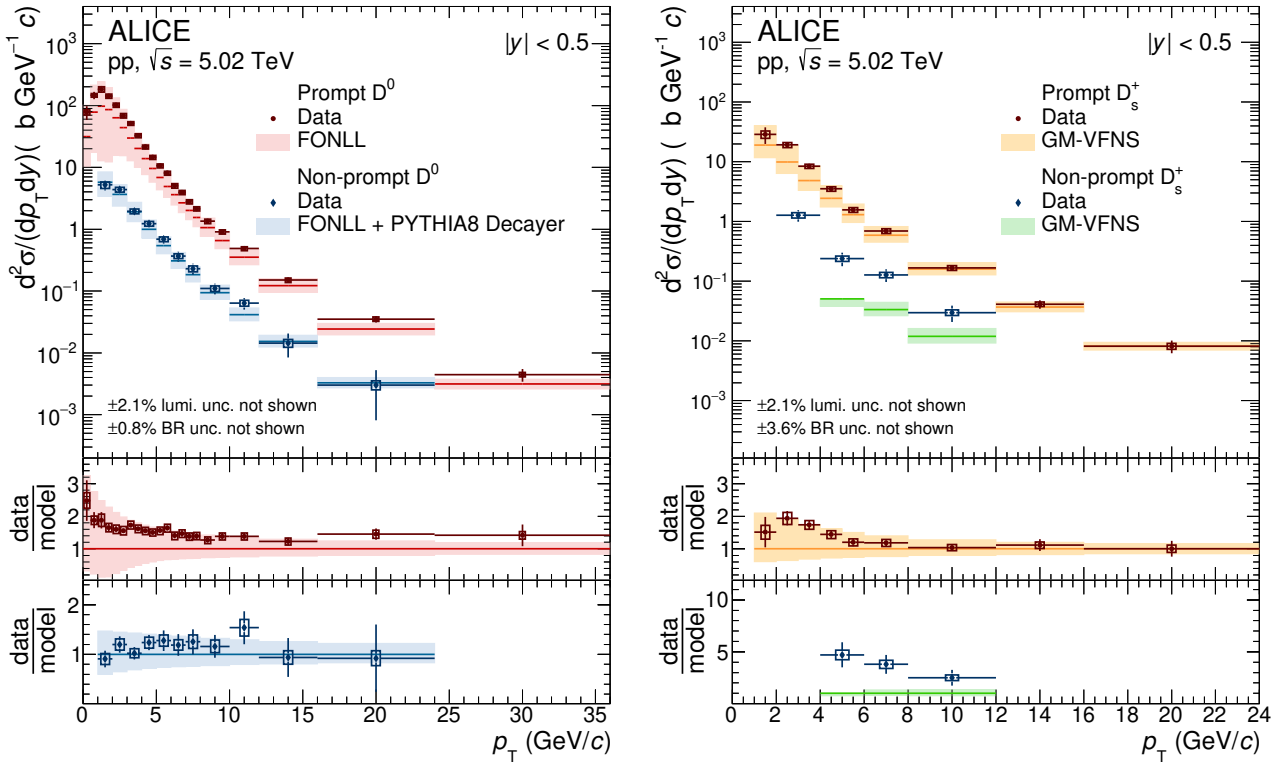


Figure II.2.2: The p_{T}-differential production cross sections of prompt and non-prompt D^0 (left panel), D_s^+ (right panel) mesons in proton-proton collisions at $\sqrt{s} = 5.02$ TeV measured by the ALICE Collaboration. The measurements are compared to FONLL calculations combined with PYTHIA 8 for the $\mathrm{H}_b \to \mathrm{D+X}$ decay kinematics and predictions obtained using GM-VFNS calculations.

The FONLL and GM-VNFS predictions are compatible within uncertainties with the prompt D mesons cross section measurements. The central prediction points are systematically below the

experimental data in the low and intermediate p_{T} regions. This could be explained by the missed contributions to the heavy quark production cross section from higher order diagrams (NNLO), not currently available for the charm quark. The non-prompt D mesons cross sections measurements are well reproduced by the FONLL + PYTHIA 8 H_b decayer. The GM-VNFS predictions greatly underestimate the non-prompt production, for the D_s^+, from a factor 5 to 2 as a function of p_{T}. The effect of the fragmentation approach used in the cross section calculation is explored in [128] where the "two-steps" and the extended BKK approaches are studied with the GM-VNF scheme. In the reference [122] extended D mesons comparisons with the two formalisms are presented.

The measurements of the $\mathrm{B}^{\pm}$ and B_s^0 mesons p_{T}-differential cross sections in proton-proton collisions at $\sqrt{s} = 5.02$ TeV are shown on figure II.2.3 and compared to FONLL calculations [129, 130]. The FONLL predictions are obtained by scaling the total b-quark production [123–125] by the world-average production fractions of $\mathrm{B}^{\pm}$ and B_s^0 [131]. Within uncertainties, the predictions are consistent with the measurements.

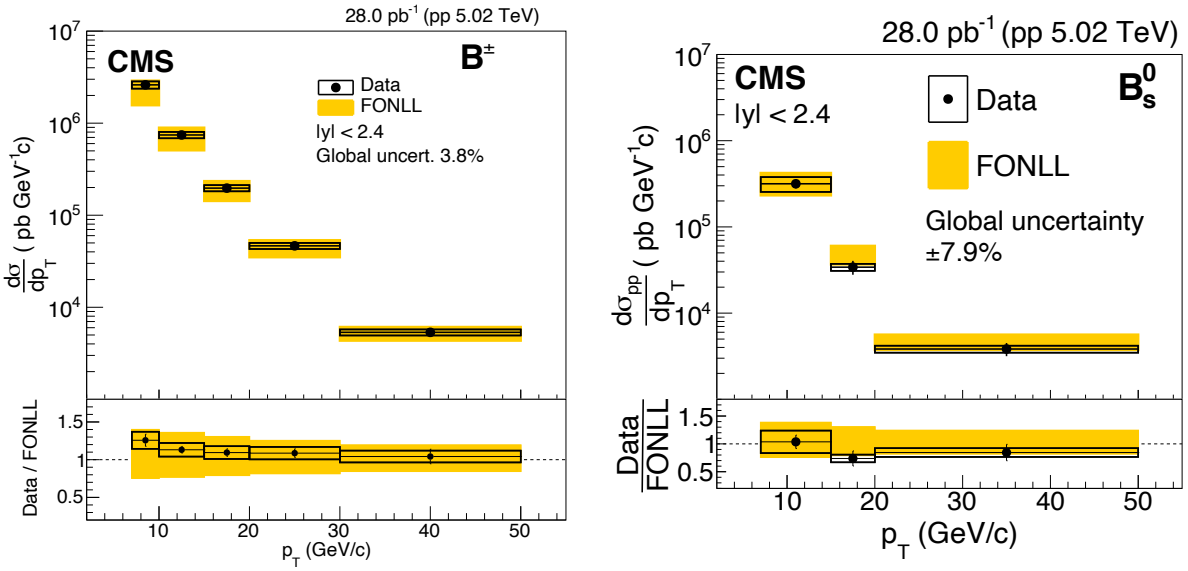


Figure II.2.3: The p_{T}-differential production cross sections of $\mathrm{B}^{\pm}$ (left panel) and B_s^0 (right panel) mesons in proton-proton collisions at $\sqrt{s} = 5.02$ TeV measured by the CMS Collaboration [129, 130]. The measurements are compared to predictions from FONLL [123–125].

The measurement of the Λ_c^+ p_{T}-differential cross section in proton-proton collisions at $\sqrt{s} = 5.02$ TeV is shown on figure II.2.4 and compared to perturbative QCD and Monte Carlo predictions [132]. The POWHEG framework [133] allows for interfacing perturbative QCD NLO heavy quark calculation with parton showers Monte Carlo generators. Here, the PYTHIA 6 Monte Carlo generator [134] is used to generate the parton shower and the parton hadronisation, the parton distribution functions are taken from [135]. Different tunes of the PYTHIA 8 Monte Carlo generator are used for comparison, a popular LHC specific tune, the Monash 2013 tune [136], and tunes implementing the colour reconnection modes [118]. Colour reconnection models try to address the question of colour interactions between partons coming from multiple parton interactions in hadronic collisions.

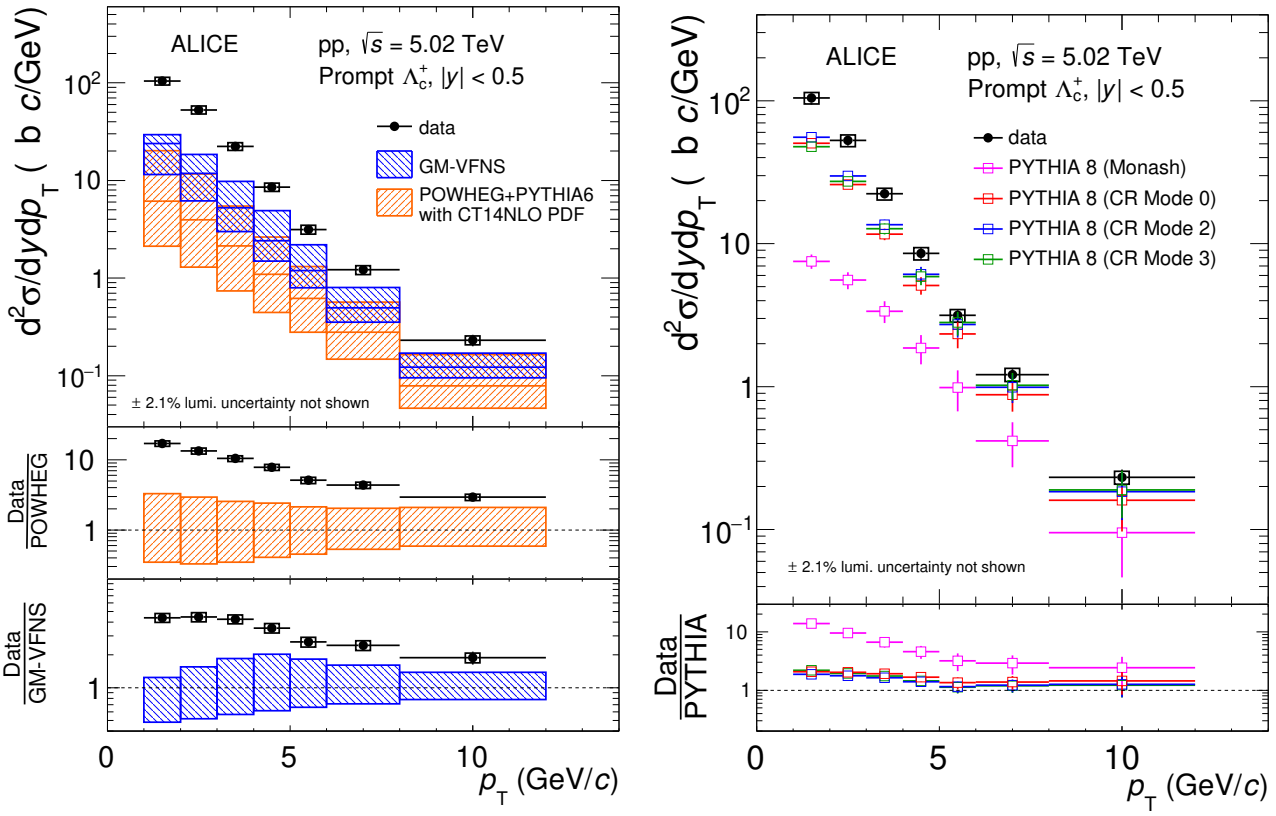


Figure II.2.4: The p_{T}-differential production cross sections of prompt Λ_c^+ baryon in proton-proton collisions at $\sqrt{s}$ = 5.02 TeV measured by the ALICE Collaboration. The measurements are compared to predictions obtained using GM-VFNS calculations, POWHEG+PYTHIA 6 with CT14NLO parton distribution functions and PYTHIA 8 predictions.

While the GM-VNFS prediction describes reasonably well the prompt D meson production cross section, it is not the case for the prompt Λ_c^+ (right panel of figure II.2.4). This suggests that charm baryons hadronise differently with respect to charm mesons in the hadronic environment, breaking the universal fragmentation picture from the factorisation theorems. This idea is implemented in more advanced colour reconnection models describing the complex and colour rich environment surrounding the charm quark which tend to be more in agreement with data compared to the PYTHIA 8 Monash 2013 tune or the POWHEG + PYTHIA 6 predictions.

This is further supported by the $\Xi_c^{0,+}$ and $\Sigma_c^{0,+,++}$ charm baryon production measurements [137, 138], not described by the perturbative QCD calculations and models based on a fragmentation hadronisation mechanism. The charm baryon to meson ratios measured in proton-proton collision are significantly higher than the value found in e^+e^- and e^+p collision systems. These results contribute to the understanding of charm baryon production via the colour reconnection mechanism, the coalescence mechanism and feed-down contribution from unobserved higher-mass charm baryon states.

II.3 Influence of the cold nuclear environment

In the presence of a nucleus in the collision system, additional QCD dynamics effects arise with respect to hadron-hadron collisions, so called cold nuclear matter (CNM) effects. They need to be disentangled from the hot medium effect present in heavy-ion collisions. Different theoretical approaches are employed to describe such effects: the modification of the parton distribution functions in the nuclei, the colour glass condensate (CGC) and the multiple scattering frameworks.

The nuclear parton distribution functions (nPDFs) are introduced as a modification of the free nucleon PDFs by the nuclear modification factor $R_i^A(x, Q^2)$ with x being the parton momentum fraction and Q^2 the momentum transfer energy scale. At the initial parametrisation scale Q_0^2, the parton momentum fraction x dependence of the the nuclear modification factor is parametrised considering different modifications regions as shown on figure II.3.1.

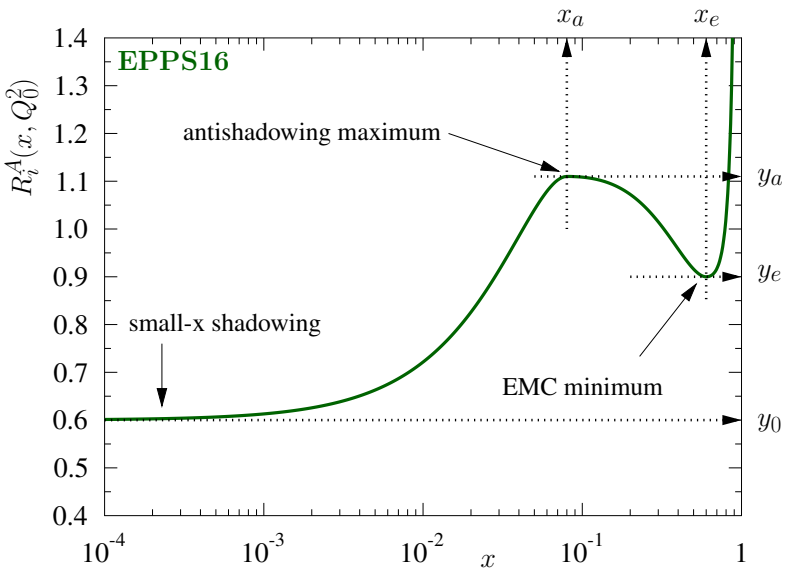


Figure II.3.1: The EPPS16 fit function of the nuclear modification factor $R_i^A(x, Q_0^2)$ at the initial parametrisation scale Q_0^2 [139].

In the shadowing region, x (<0.1) and $R_i^A(x, Q_0^2) < 1$, the gluon PDF is the dominant component of the nucleons. Gluon interactions and especially the gluon fusion process are expected to play an important role. The gluon fusion engender a depopulation of the low x region in favour of the anti-shadowing region, at $0.1 \leq x \leq 0.25$ where $R_i^A(x, Q_0^2)$ is above unity.

The $0.25 \leq x \leq 0.8$ region correspond to the European Muon Collaboration (EMC) region where the x dependence of the iron over deuterium nucleon structure functions F_2^N ratio was measured [140], showing a smaller average momentum of a quark in a bound nucleon compared to a free one. The EMC effect was first explained by the larger volume available to the quarks in a bound nucleon causing a decrease of their average momentum due to the uncertainty principle. The average nuclear density dependence of the quark PDFs is studied with quark-cluster models, through the Q^2 dependence of the confinement scale and the influence of the nuclear binding [141]. Local density modification of the nuclear matter as the nucleon fluctuations into short range correlated pairs is also suggested to be possibly correlated to the EMC effect [142].

At $x > 0.8$, in the Fermi motion region, the $R_i^A(x, Q_0^2)$ goes again above unity. This regime is

due to thermal motion of nucleons in the nuclei leading to an increased transverse momentum of the partons with respect to the free nucleon case. The nuclear PDFs are parametrised and fitted to various hard processes experimental by performing global analysis [143, 144]. Deep inelastic scattering, Drell-Yan production and heavy-flavours production in proton-nucleus are the main experimental data inputs.

At low x, the rise of the gluon distribution in a hadron is driven by the Balitsky, Fadin, Kuraev and Lipatov (BFKL) perturbative QCD equations [145–147]. This low x evolution at fixed Q^2 reach a saturation scale Q_{sat} at which hadrons can be identified as dense systems of gluons. The colour glass condensate (CGC) is an effective theory describing the properties of the saturated gluons in the Reggea-Gribov limit (fixed Q^2, $x \to 0$, centre of mass energy squared $s \to \infty$), aiming to provide a unified description of high energy hadrons and nuclei. The framework is based on the reformulation of the BFKL equations required for the treatment of large gluon density appearing at high energy [60]. The term "glass" is employed because of the large gluon evolution time scale compared to the natural time scale $t \sim Q_{sat}$. The CGC framework can be applied to describe the initial state of hadron-hadron, hadron-nucleus and nucleus-nucleus ultra-relativistic collisions.

Another common approach to treat CNM effects is to study the hard parton energy loss mechanisms in the nucleus. In both the initial and final states, *i.e.* before and after the heavy quark pair creation, parton energy loss can occur via a collisional process, the parton scatter on partons of the nuclear medium [148, 149], and radiative process, the parton goes under multiple soft scattering inducing gluon emission [150, 151]. The Cronin effect [152] was attributed to the enhanced production of hadrons at intermediate p_{T} (2-6 GeV/c) in proton-nucleus collisions with respect to proton-proton collisions. As the parton scatters in the initial or final states, it gains in transverse momentum leading to a transverse momentum broadening.

Experimentally, the CNM effects are quantified using the nuclear modification factor, with a similar definition as equation I.1 for nucleus-nucleus collisions but with the appropriate normalisation to binary collisions,

$$R_{\mathrm{pA}}(p_T, y) = \frac{1}{A} \frac{d^2\sigma_{pA}/dp_T dy}{d^2\sigma_{pp}/dp_T dy} \tag{II.3}$$

with A the mass number of the nucleus.

The figure II.3.2 presents the prompt D mesons nuclear modification factor measurement in p-Pb collisions measured by the ALICE Collaboration [153]. The R_{pPb} is compatible with unity within a 2σ uncertainty over the whole transverse momentum interval measured, showing no clear modification of the D meson production in p-Pb collisions with respect to the production in proton-proton collisions. The measurement is compared to predictions from the CGC framework [154], FONLL [125] with the EPPS16 nuclear PDF [139], the Vitev et al. leading order perturbative QCD calculation including many-body QCD scattering effects as the Cronin effect, CNM energy loss and dynamical shadowing [155], and the Kang et al. higher-twist calculation allowing the study of incoherent multiple scatterings in the initial and final states [156]. All the theoretical approaches predict significant CNM effects in the low p_{T} region, typically below 5 GeV/c. With the current precision of the measurement and the model predictions, no firm conclusion can be made on the specific origin and amplitude of the CNM effects. However the data do not seem to be in agreement with the positive contributions from incoherent double scatterings to the heavy-flavour meson cross section for $p_{\mathrm{T}} < 3$ GeV/c obtained with the Kang et al. calculation.

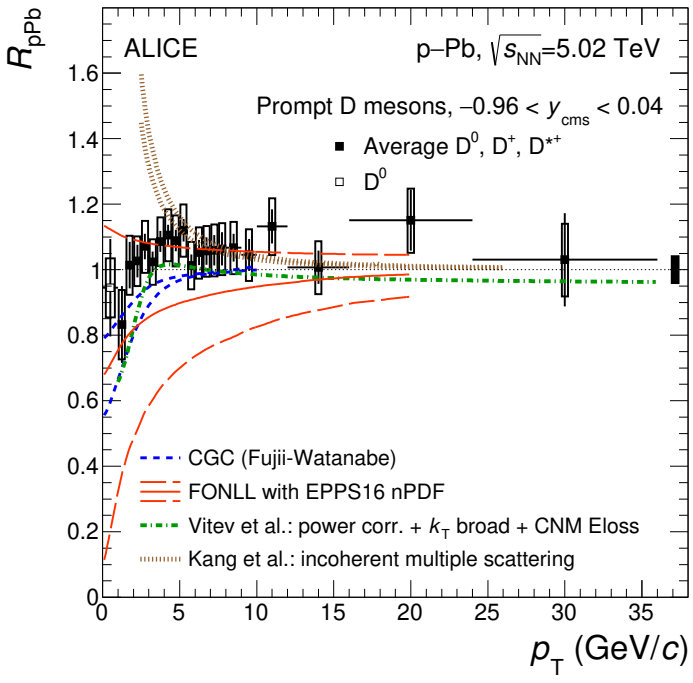


Figure II.3.2: The nuclear modification factor $R_{\rm pPb}$ of prompt $\rm D^0$ meson and averaged $R_{\rm pPb}$ over $\rm D^0$, $\rm D^+$ and $\rm D^{*+}$ mesons measured in p-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV by the ALICE Collaboration [153]. The measurements are compared to theoretical predictions including CNM effects.

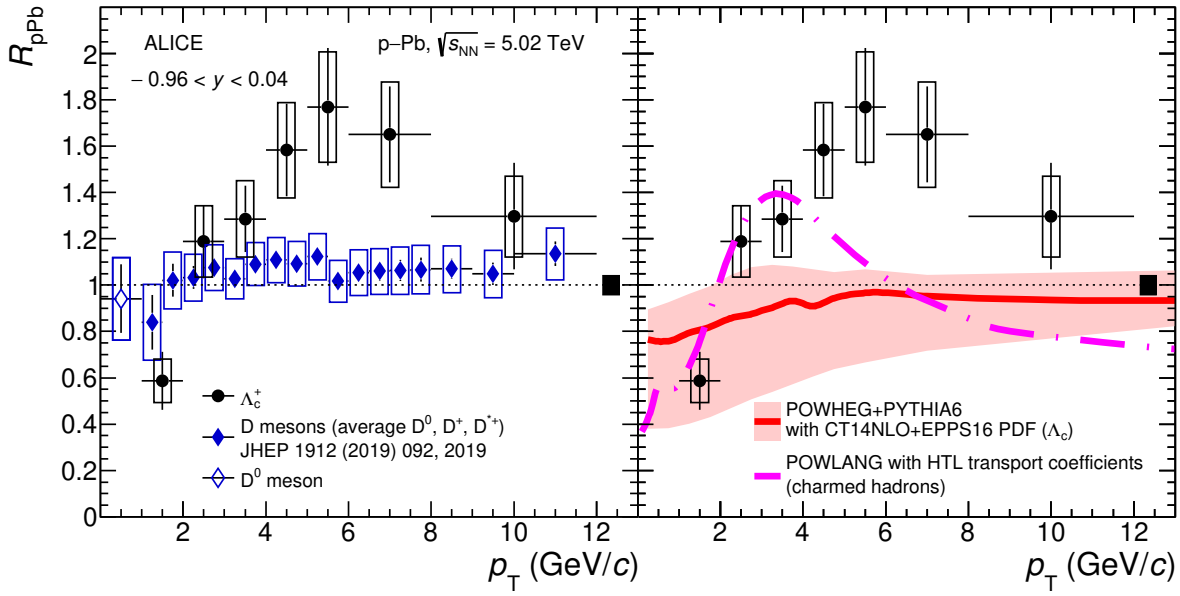


Figure II.3.3: The nuclear modification factor $R_{\rm pPb}$ of prompt Λ_c^+ in p-Pb collisions at $\sqrt{s_{\rm NN}}$ = 5.02 TeV measured by the ALICE Collaboration [132]. A comparison to the POWEGH+PYTHIA 6 with EPPS16 nuclear PDFs and POWLANG model prediction is shown.

.

The nuclear modification factor of the Λ_c^+ is compared to the corresponding D mesons measurement on the left panel of figure II.3.3. The trend observed at intermediate p_{T} with R_{pPb} values systematically above unity and higher than the D meson results indicates an increased production of the open-charm baryon in p-Pb collisions with respect to proton-proton collisions and also differences in the production of open-charm baryon and meson in this p_{T} region. The POWEGH + PYTHIA 6 [133] with the EPPS16 nuclear PDF [139] predictions show a similar trend as the one for the D mesons R_{pPb} (see figure II.3.2). While the POWLANG model [157], that assumes the formation of hot deconfined medium, predicts an enhancement of the R_{pPb} in the low-intermediate p_{T} region but does not fully describe the data.

II.4 Hot deconfined medium effects

As heavy quarks are produced in hard-scattering processes happening at time scales t$\sim 1/2m_{c,b}$ (0.07 fm/c for the charm quark and 0.02 fm/c for the bottom quark), shorter than the QGP formation time ($\sim$0.1-1 fm/c), they are potential probe of the whole space-time evolution of the medium. Since their mass is well above the temperature of the medium, their production is restricted to the initial hard scattering processes. Having a relaxation time comparable or larger than the lifetime of the QGP [158, 159], heavy quark equilibration with the surrounding medium is not guaranteed. As they propagate through the QGP, heavy quarks will interact with the medium constituents via two main mechanisms: elastic collisional $q(g)Q \rightarrow q(g)Q$ and radiative $q(g)Q \rightarrow q(g)Qg$ processes modifying their original momentum distribution. For heavy quark momentum $p_Q \ll m_Q$, collisional energy loss due to elastic scattering of the heavy quark with the medium constituents is expected to be non-negligible and is related to the mean free path, the Debye mass (the inverse of the screening length of the colour electric fields in the medium) and the transport coefficients of the medium. For heavy quark momentum $p_Q \gg m_Q$, gluon radiation process is the dominant source of heavy quark energy loss. After hadronisation, these modifications are transferred to the final hadron momentum spectra allowing for studying the parton-medium interactions.

II.4.1 Heavy quark interactions

The heavy quark diffusion in the medium can be described the Boltzmann transport equation governing the space-time evolution of the heavy quark phase space distribution function f_Q [160].

$$\left[\frac{\partial}{\partial t} + \frac{\vec{p}}{E_p}\frac{\partial}{\partial \vec{x}} + \vec{F}\frac{\partial}{\partial \vec{p}}\right] f_Q(t, \vec{x}, \vec{p}) = C[f_Q] \tag{II.4}$$

with $E_p = \sqrt{m_Q^2 + \vec{p}^2}$ the heavy quark energy, $\vec{F}$ a force induced by an external field (e.g. electromagnetic or colour field) and $C[f_Q]$ the collision integral containing the heavy quark interaction mechanisms.

Since the thermal momentum exchange $q_0 \sim T$ from the medium is small compared to the characteristic thermal momentum of the heavy quark $p_{th} \simeq \sqrt{3m_{\mathrm{Q}}T}$, with m_{Q} the mass of the heavy quark and T the temperature of the medium, the heavy quark propagation in the QGP can be assimilated to the diffusion of a heavy particle plunged in a system of lighter particle fluid therefore treated as a Brownian motion. In this regime, the Fokker-Planck equation is derived from the Boltzmann equation, where the in-medium interactions are encoded in transport coefficients.

Its expression in the non-relativistic regime is as follows [161],

$$\frac{\partial}{\partial t} f_{\mathrm{Q}}(t,p) = \gamma \frac{\partial}{\partial p_i}[p_i f_{\mathrm{Q}}(t,p)] + D_p \Delta_{\vec{p}} f_{\mathrm{Q}}(t,p) \tag{II.5}$$

with γ being the drag coefficient and D_p the momentum diffusion coefficient. The spatial diffusion coefficient, D_s, describing the evolution in time of the spatial distribution is related to the drag and momentum diffusion coefficient through:

$$D_s = \frac{T}{m_{\mathrm{Q}}\gamma} = \frac{T^2}{D_p} \tag{II.6}$$

The equation can be implemented via a Langevin process where the heavy quark momentum evolve following the Langevin equation [158],

$$\frac{d\vec{p}}{\mathrm{d}t} = -\eta_D(p)\vec{p} + \vec{\xi}(t) \tag{II.7}$$

with $\eta_D(p)$ the drag coefficient representing the fractional momentum loss per unit of time. Additional random kicks uncorrelated in time are added trough the stochastic term, $\vec{\xi}(t)$, containing the momentum diffusion coefficients. The Boltzmann approach describe a medium composed of quasi-particle and allows for a description of off-equilibrium effects in the medium [160]. While the Langevin approach does not require the existence of quasi-particles and is particularly adapted for the heavy quark propagation in a strongly coupled medium [160].

The first evaluations of the heavy quark transport coefficients from elastic interactions were obtained from perturbative QCD at leading order calculations [162–164]. The main contributions to the interactions comes from the gluon t-channel exchange. A review of heavy quark elastic interactions in QCD matter is given in [161].

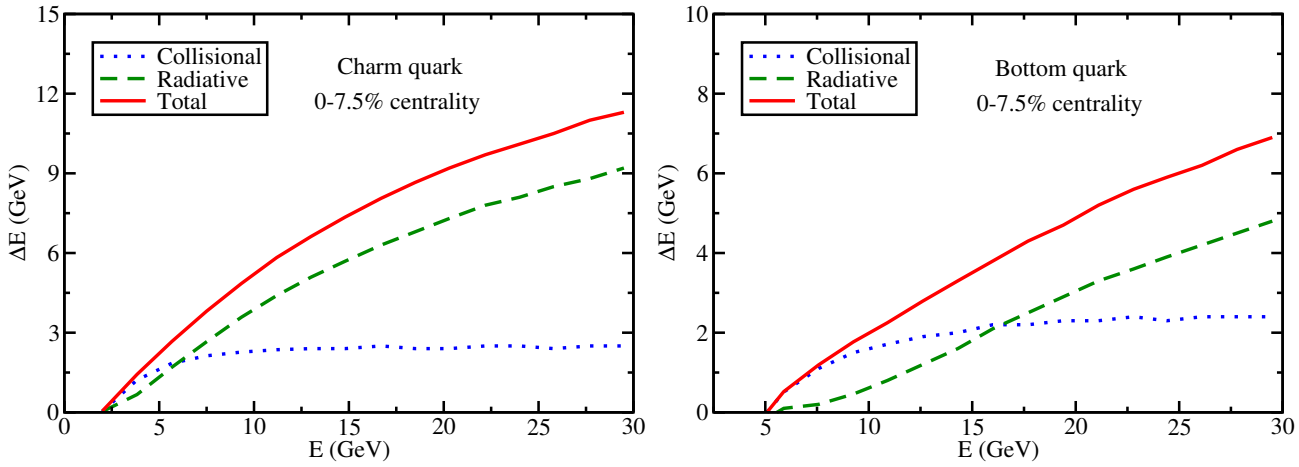


Figure II.4.1: Radiative and collisional energy losses for charm (left panel) and for bottom (right panel) quarks as a function of the initial quark energy traversing a QGP medium created in Pb-Pb ultra-relativistic collisions at LHC energies [165].

At high heavy quark momentum, $p_Q \gg m_Q$, radiative energy loss come into play and is expected to be the dominant mechanism. The formalisms to describe energy loss via medium-induced gluon emission are generally based on perturbative QCD and are divided in four main

classes: the path-integral formalism, the opacity expansion approach, the higher-twist approach, and the finite temperature field theory approach. In the following, only a few aspects of the different formalisms and their differences are highlighted, a complete review is given in [166]. Considering a single gluon emission, the energy loss is expressed as:

$$\Delta E = \int d\omega \omega \frac{\mathrm{d}N}{\mathrm{d}\omega}$$

with ω the energy of the emitted gluon and $\frac{\mathrm{d}N}{\mathrm{d}\omega}$ the gluon emission spectrum per energy unit.

In the Baier-Dokshitzer-Mueller-Peigne-Schiff and Zakharov (BDMPS-Z) formalism [167–169], based on the path-integral formalism, the radiative energy loss is calculated considering multiple soft scattering of the heavy quark of a given energy E on static centres provoking a single soft gluon emission with an energy $\omega \ll E$. The total radiative energy loss of a high momentum parton traversing a medium of path length L is proportional to

$$\Delta E \propto \alpha_s C_R \hat{q} L^2$$

with C_R the colour Casimir constant, $\hat{q}$ the diffusion coefficient related to the transverse momentum broadening. The opacity expansion take a similar approach where the gluon emission spectrum is usually obtained assuming a single hard scattering. While the multiple soft scattering formalism is suited for the energy loss description in a thick medium, the opacity expansion formalism is more adapted for a thin medium. In the higher-twist approach, the heavy quark fragmentation in the vacuum receives an additional contribution encoding the medium effects on the parton shower. The average energy loss is extracted by fitting experimental data. Unlike the BDMPS-Z approach, the finite temperature field theory approach describes the medium as a collection of moving thermal partons instead of static scattering centres.

The gluon radiation from a heavy quark is suppressed at angle $\theta < \theta_0 = m_Q/E$ with m_Q being the mass and E the energy of the heavy quark, known as the dead cone effect [170]. Heavy quarks are then expected to loose less energy than lighter quarks and gluons. The figure II.4.1 presents the radiative and collisional energy losses for charm (left panel) and for bottom (right panel) quarks traversing a QGP medium as a function of their initial energy. The heavy quark propagation is described by the Langevin equation including both quasi-elastic and radiative energy losses [165]. For both heavy quarks, the quasi-elastic scattering energy loss is dominant at low energy while above 6 GeV for the charm and 16 GeV for the bottom quark, the radiative energy loss becomes dominant. Compared to the charm quark, the bottom quark loose less energy as expected from the dead cone effect.

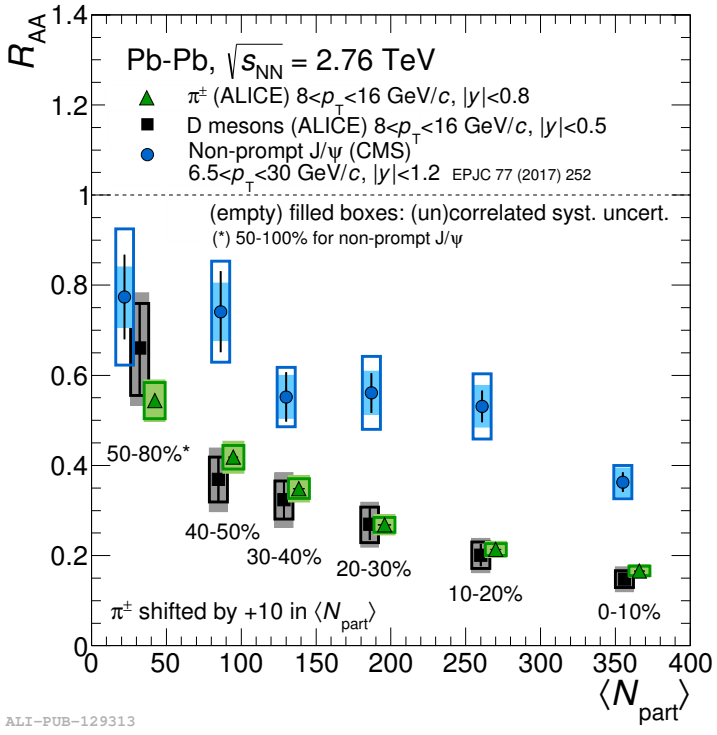


Figure II.4.2: Average $R_{\mathrm{AA}}(p_{\mathrm{T}})$ of the prompt D^0, D^+ and D^{*+} mesons and the R_{AA} of charged particle in Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV measured by the ALICE Collaboration [171, 172] compared with the R_{AA} of non-prompt J/Ψ mesons measured by the CMS Collaboration [173].

Experimentally, heavy quark energy loss is probed by measuring the nuclear modification factor R_{AA} in nucleus-nucleus collisions (eq. I.1). If the R_{AA} is compatible with unity, *i.e.* the heavy quark yield scale with the nuclear overlap density, it would indicate that the produced medium is transparent to heavy quarks. The figure II.4.2 show a comparison of the nuclear modification factors of prompt D mesons, charged particle and non-prompt J/Ψ as a function of the centrality expressed as the average number of nucleons participating in the Pb-Pb collision ($\langle N_{\mathrm{part}} \rangle$). The production yield of the final-sate particles get further suppressed as the centrality of the collision increases, reaching a suppression factor of ∼2.5 for the non prompt J/Ψ and ∼5 for the prompt D mesons and charged particle in the most central collisions. This strong suppression demonstrate the effects of the heavy quark interactions with the surrounding medium constituents on the production yields. The R_{AA} of the prompt D mesons and charged particles (dominated by pions) are compatible within uncertainties in all centrality classes. The non-prompt J/Ψ R_{AA} is much higher with respect to the prompt D mesons (3.4σ effect in the 0-10% and 10-20% centrality classes) suggesting a different suppression of the charm and bottom hadron productions. The comparison with models performed in [172] supports this observation and leads to the interpretation that the difference is manly due to the decrease of in-medium heavy quark radiative energy loss with increasing quark mass.

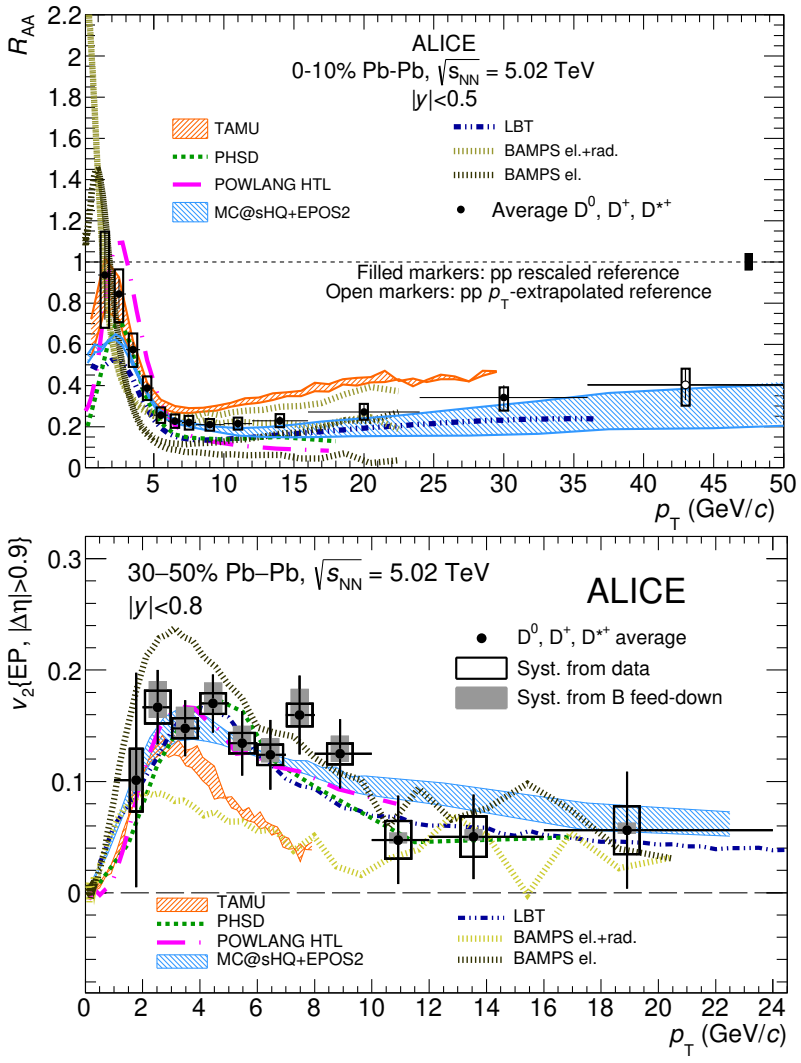


Figure II.4.3: Average $R_{\rm AA}(p_{\rm T})$ of the prompt $\rm D^0$, $\rm D^+$ and $\rm D^{*+}$ mesons in the 0-10% centrality class (top) and the average elliptic flow ν_2 in the 30-50% centrality class (bottom) in Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV measured by the ALICE Collaboration [15, 174]. Both measurement are compared to the models of table II.4.1.

As discussed in the section I.3.3, the initial spatial anisotropy of a nucleus-nucleus collision can be translated into final state particles momentum anisotropy due to the collective dynamic of the expanding medium. This anisotropy is experimentally measured through the Fourier coefficient ν_n of the momentum dependent azimuthal distribution of final state particles. In the low $p_{\rm T}$ region, heavy-flavour hadron elliptic flow ν_2 measurements allow for studying the flow transferred from the medium to the heavy quark which is an indication on the degree of heavy quark thermalisation [175]. The dependence of the heavy quark energy loss on the path length is reflected in the elliptic flow at high $p_{\rm T}$ [176, 177]. A non-vanishing elliptic flow of the prompt D mesons is observed suggesting that the heavy quark is affected by the medium via multiple interactions with the medium constituents. The coalescence production mechanism of heavy-flavour hadrons and its impact on the $R_{\rm AA}$ and ν_2 are discussed in the next section.

The figure II.4.3 present the $R_{\rm AA}$ of prompt D mesons in the 0-10% central collisions and the elliptic flow ν_2 in the 30-50% centrality class as a function of $p_{\rm T}$ in Pb-Pb collisions [15, 174]. The

$R_{\rm AA}$ is compatible with unity at low $p_{\rm T}$ but decrease rapidly to reach the maximum suppression factor in the $6 < p_{\rm T} < 10$ GeV/c region where the radiative energy loss mechanism is dominant. The measurements are compared to different models describing initial heavy quark production with possible CNM effects, the space time evolution of the QGP up to the freeze-out stage, the charm quark energy losses, the hadronisation mechanism(s) and the possible D mesons final state interactions in the hadronic phase. The access to precise measurements of the heavy-flavour $R_{\rm AA}$ and ν_2 help to constrain the model parameters such as the heavy quark spatial diffusion coefficient D_s. A brief overview of the different energy loss mechanisms discussed above and implemented in different models is presented in table II.4.1 along with their respective D_s values.

The TAMU model overestimate the nuclear modification factor for $p_{\rm T} > 8$ GeV/c and underestimate the elliptic flow for $p_{\rm T} > 3$ GeV/c. By comparison with the other models, the discrepancies could be explained by the missing radiative energy loss mechanism. The BAMPS elastic model underestimate the $R_{\rm AA}$ for $p_{\rm T} > 3$ GeV/c and overshoot the maximum flow. The $R_{\rm AA}$ is better described when adding the radiative energy loss mechanism causing a drastic decrease of the ν_2 predictions. Over the whole $p_{\rm T}$ range, the PHSD, POWLANG, MC@sHQ+EPOS2, and LBT models describe reasonably well the ν_2 measurement. The model comparison performed in [174] led to the estimation of the charm quark thermalisation time, $\tau_{\rm charm} = \frac{m_{\rm charm}}{T} D_s(T) \simeq 3 - 14$ fm/c, with $T = T_c \simeq 155$ MeV and $m_{\rm charm} = 1.5$ GeV/c^2, which is similar to the decoupling time of the medium estimated in [178].

models	**heavy quark interaction**	**heavy quark hadronisation**	$\boldsymbol{2\pi T D_s(T)}$ $[T_c$ - $2T_c]$
TAMU [179]	transport with Langevin eq. collisional energy loss diffusion in hadronic phase	fragmentation coalescence	4 - 10 [180]
PHSD [181]	Parton-Hadron-String Dynamics collisional + radiative energy loss	fragmentation coalescence	4 - 9 [181]
POWLANG [182]	transport with Langevin eq. collisional energy loss	fragmentation coalescence	7 - 18 [160]
MC@sHQ+EPOS2 [183]	transport with Boltzmann eq. collisional + radiative energy loss	fragmentation coalescence	1.5 - 4.5 [180]
LBT [184]	transport with Boltzmann eq. collisional + radiative energy loss	fragmentation coalescence	2 - 6 [185]
BAMPS [186] BAMPS el. BAMPS el.+rad.	transport with Boltzmann eq. collisional energy loss collisional + radiative energy loss	fragmentation	 1 - 2 6 - 10

Table II.4.1: Overview of the heavy quark in-medium interactions and heavy quark hadronisation mechanisms in different models. The dimensionless quantity $2\pi T D_s(T)$ is extracted for the interval $T \in [T_c, 2T_c]$ with $T_c \simeq 155$ MeV [187] being the QGP formation critical temperature.

In the open beauty sector, the CMS Collaboration measured the nuclear modification factor of the $\rm B^+$, $\rm B_s^0$ and $\rm B_c^+$ mesons as shown on the figure II.4.4 [129, 130, 188]. The $R_{\rm AA}(p_{\rm T})$ of the $\rm B^+$ mesons shows a similar production suppression as for the $\rm D^0$ in Pb-Pb collisions. Compared to the $\rm B^+$, both the $\rm B_s^0$ and $\rm B_c^+$ measurements provide a hint of a possible production enhancement. An enhanced production could originate from a contribution of recombination of beauty

with strange and charm quarks processes competing with suppression mechanisms in heavy-ion collisions. However, the uncertainties are quite large and the results are still compatible with a suppression scenario. More precise measurements of beauty hadron $R_{\mathrm{AA}}(p_{\mathrm{T}})$ could help to disentangle the different mechanisms at play during the heavy-quark propagation in the QGP.

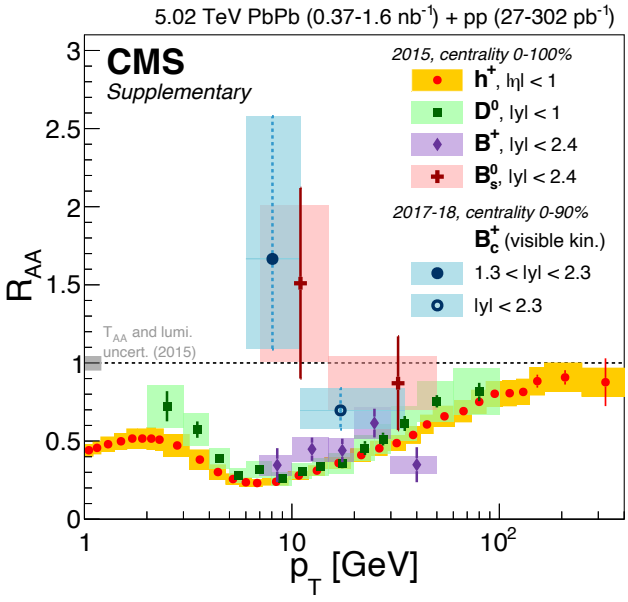


Figure II.4.4: $R_{\mathrm{AA}}(p_{\mathrm{T}})$ of the D^0, B^+, B^0_s and B^+_c mesons compared to that of light charged hadrons as a function of the measured transverse momentum, measured by the CMS Collaboration [129, 130, 188].

II.4.2 Hadronisation mechanisms

The presence of the hot medium can manifest itself by changes in the particle production mechanisms. In addition to the fragmentation, open heavy-flavour hadrons can be produced by the recombination or coalescence of a heavy quark with quarks from the surrounding deconfined medium. This in-medium hadronisation mechanism require hadrons to be formed from quarks that are close in phase space (position and momentum). In the instantaneous coalescence model [189], the initial phase space filled with partons undergoes a very brief freeze-out phase, projecting the parton states into hadron states in a probabilistic approach. The resonance recombination model [190], describes the coalescence hadronisation through the formation of in-medium resonant states.

The heavy quark cross section decreases with increasing transverse momentum $p_{\mathrm{T,\,Q}}$, therefore the baryon production in the fragmentation scenario is penalised with respect to meson production with the same momentum $p_{\mathrm{T,\,H}}$ since more quark must be produced out of the vacuum ($p_{\mathrm{T,\,H}} < p_{\mathrm{T,\,Q}}$). In the coalescence mechanism is it the opposite, baryons are composed of quarks with momenta of $\sim 30\%$ of $p_{\mathrm{T,\,H}}$ and mesons of quarks with momenta of $\sim 50\%$ of $p_{\mathrm{T,\,H}}$ ($p_{\mathrm{T,\,H}} > p_{\mathrm{T,\,Q}}$). At low momentum, the hadron production is expected to receive a dominant contribution from the coalescence mechanism while at high momentum it is the fragmentation mechanism [191]. With respect to a pure hadronisation via quark fragmentation, the addition of the coalescence mechanism is likely to result in an increase of the baryon over meson production ratio at intermediate p_{T}. As the recombination of the charm quark with flowing light-flavour quarks is expected to increase heavy-

flavour elliptic flow [192], the manifestation of the coalescence mechanism can also be reflected in the D meson elliptic flow measurements.

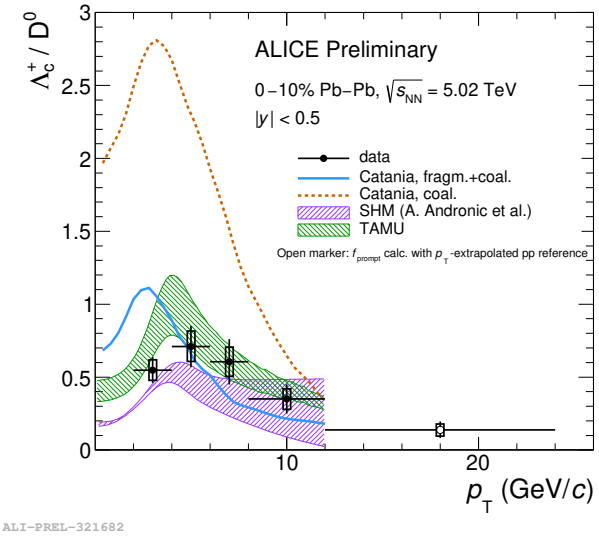


Figure II.4.5: Ratio of the p_{T}-differential production yield of the Λ_c^+ over the D^0 in central (0-10%) Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV measured by the ALICE Collaboration. The result is compared to theoretical predictions from the Catania group [191], the GSI/Hd+BW model [193], and the TAMU model [194].

The figure II.4.5 show the Λ_c^+ over D^0 production yield ratio in central Pb-Pb collisions measured by the ALICE Collaboration. The Catania group predictions [191] including only the coalescence hadronisation mechanism overshoot drastically the measurement due to the overestimated Λ_c^+ production. With the inclusion of the fragmentation mechanism, the charm hadronisation into the D^0 meson gets more balanced with respect to the Λ_c^+, leading to a better description of the data. However the maximum of the predicted ratio spectrum at intermediate p_{T}, highlighting the coalescence region, seems to be more shifted toward lower p_{T}. Unlike the Catania model, the TAMU model [194] use the resonance approach to coalescence and include charm baryon excited states predicted by the Relativistic Quark Model [195] and lattice QCD [196].

Hadronisation mechanisms of charm quarks are also studied with the measurement of the D_s^+ over D^0 production yield ratio as a function of p_{T} [197] in proton-proton collision and in different centrality classes of Pb-Pb collisions as shown on figure II.4.6. In particular, an enhancement of the D_s^+ production in nucleus-nucleus collisions at low and intermediate momenta, with respect to proton-proton, is expected if the main contribution to the D meson production is the recombination mechanism, due to $s\bar{s}$ saturation in the QGP (see section I.3.2). In the $2 < p_{\mathrm{T}} < 8$ GeV/c region, the $\mathrm{D}_s^+/\mathrm{D}^0$ ratio values are higher in Pb-Pb collisions compared to proton-proton, the effect is about 2.3σ (0-10% class) and 2.4σ (30-50% class) of the total uncertainty.

The results in proton-proton and Pb-Pb collisions are described by the Catania [191] an PHSD [181, 198] models. Both model implement the fragmentation and the coalescence hadronisation mechanisms in Pb-Pb collisions. The TAMU [194] and LGR [199] predicts a peak in the $3 < p_{\mathrm{T}} < 4$ GeV/c region of the double ratio $(\mathrm{D}_s^+/\mathrm{D}^0)_{\text{Pb-Pb}}/(\mathrm{D}_s^+/\mathrm{D}^0)_{\text{pp}}$ (see bottom panels of figure II.4.6) while the Catania and PHSD predictions are flat for $p_{\mathrm{T}} < 3$ GeV/c and then decrease to reach unity. An interplay between a collective radial expansion of the system and the coalescence

mechanism can explain the modification of the p_{T} shape of the ratio due to the different masses of the up and strange quarks.

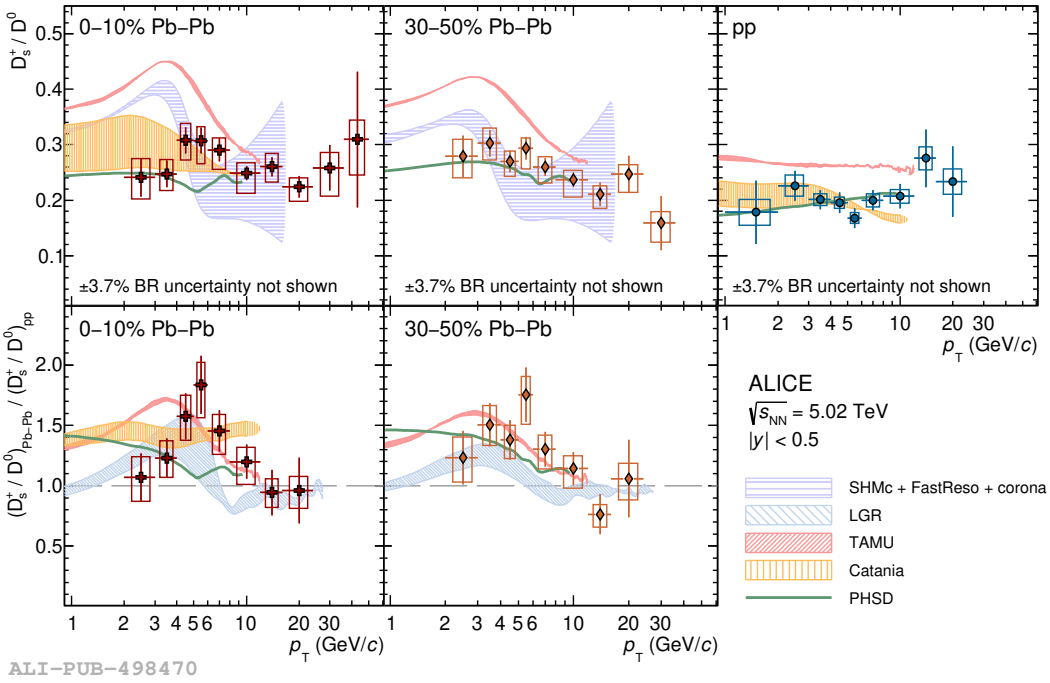


Figure II.4.6: Ratio of the p_{T}-differential production yield of the D_s^+ over the D^0 in the 0-10% and 30-50% centrality classes in Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV and in proton-proton collisions at the same centre-of-mass energy, measured by the ALICE Collaboration [197]. The results are compared to predictions from the SHM [193], LGR [199], TAMU [194], Catania [191] and PHSD [181, 198] models.

The measurement of the B_s^0 over B^+ production yield ratio as a function of p_{T} in Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV compared to the measurement in proton-proton at $\sqrt{s} = 7$ TeV [200] is shown on figure II.4.7. The result in Pb-Pb collisions is compatible within uncertainties with the proton-proton measurement, however, the points lays systematically above the proton-proton data. More precise measurements are needed to better characterise the beauty hadron production and the p_{T}-dependence of the B_s^0 over B^+ ratio in Pb-Pb collisions.

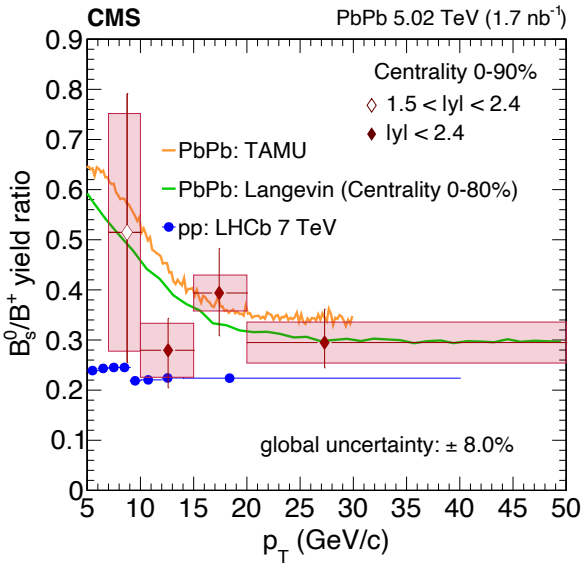


Figure II.4.7: Ratio of the p_{T}-differential production yield of the B_s^0 over the B^+ in Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV measured by the CMS Collaboration [201] compared to the measurement in proton-proton collisions at $\sqrt{s} = 7$ TeV performed by the LHCb Collaboration [200]. The predictions from the TAMU model [179] and a Langevin hydrodynamics model [202].

II.5 Studies as a function of the partonic density in small collision systems

The charged-particle multiplicity produced in a collision is correlated to the energy of the initial parton interactions and the particle production mechanisms. The study of the hard and soft QCD process contributions to the production of particles in hadron-hadron collisions is important in order to better characterise the related mechanisms. The transverse partonic structure of hadrons allows for studying of the interplay of QCD processes in the impact parameter representation of hadron-hadron collisions. The centrality of the collision can be geometrically characterised by the impact parameter and the transverse spatial distribution of partons as shown on figure II.5.1. Due to the Gribov diffusion in the partonic wave function in the transverse plane [203], the transverse spatial distribution of partons increases at high energies, however, the effect is suppressed for hard partons. Two classes of collisions can be distinguished. Small impact parameter collisions where hard parton distribution are likely to overlap are therefore associated to hard parton scattering processes. In peripheral collisions, due to the restricted spatial extension of hard parton distribution with respect to the soft parton distribution, the probability of hard scattering processes becomes very low. Soft scattering processes represents the dominant contribution to the total inelastic cross section [204–206].

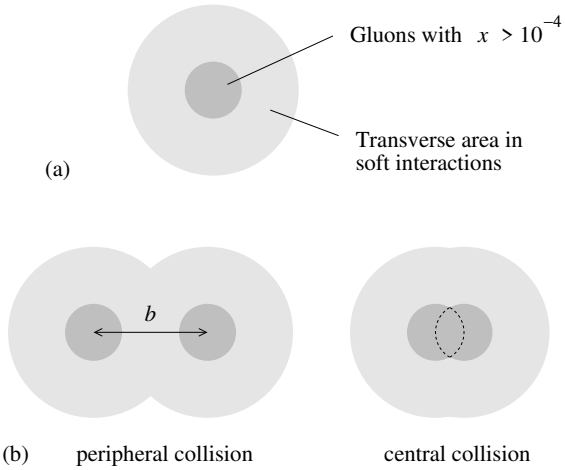


Figure II.5.1: Geometrical shape of the transverse hadron structure at high energy (a) and the two derived collision classes (b) [205].

II.5.1 Heavy-flavour hadron production in small systems

In hadron-hadron collisions, the underlying event (UE) is defined as the additional activity coexisting with any hard scattering. The underlying event receives contributions from spectator partons, *i.e.* the beam remnants, subsequent parton interactions called multiple parton interactions (MPI) and the associated initial/final state gluon radiation. At LHC energies, the underlying event is the dominant hadronic activity in hadron-hadron collisions. Particularly, the multiple particle interactions are not restricted to the underlying events, they can happen at semi-hard and hard energy scale. The multiple hard parton interactions and the large amount of gluon radiation associated with hard scattering processes induce a correlation between heavy-flavour production and the event multiplicity [207]. The study of heavy-flavour production as a function of the event multiplicity provide therefore insights on hard and soft contributions interplay at both partonic and hadronisation levels.

The figure II.5.2 show the prompt D meson relative yield, defined as the ratio of the production yield in a given multiplicity interval over the production yield in the integrated multiplicity interval, as a function of the relative charged-particle multiplicity of the collisions. The relative yields are measured in inelastic proton-proton collisions at $\sqrt{s} = 7$ TeV with at least one charged particle in $|\eta| < 1.0$, reaching a multiplicity up to 6 times the average charged particle, $\langle \mathrm{d}N_{ch}/\eta \rangle = 6.01 \pm 0.01\ (\mathrm{stat.})^{+0.20}_{-0.12}\ (\mathrm{syst.})$ [208]. The results show the augmentation of the D meson production with multiplicity which is steeper than a linear increase (highlighted by the dotted line on the figure).

The D meson production as a function of multiplicity is studied in different models. The percolation model [209, 210] characterises hadron-hadron collisions by the colour source exchanges between the colliding projectiles. The transverse spatial extension of the sources is inversely proportional to their transverse mass defined by their quark content and transverse momentum. The soft sources have therefore larger transverse size compared to hard ones making them much more sensible to source interactions. Due to coherence effects, the source interactions lead to a reduction of their effective number. The collision multiplicity, being mostly driven by the soft QCD interactions, will then heavily depend on the number of produced soft sources. The coherence effects in a dense medium will impact much more the soft source effective number rather than the hard sources. Therefore, the linear increase of the D meson production at low multiplicity becomes

more steep at high multiplicity due to the charged particle multiplicity reduction. The formation of a dense medium in proton-proton collisions at high multiplicity could explain the observed faster-than-linear increasing production of heavy-flavour as a function of multiplicity.

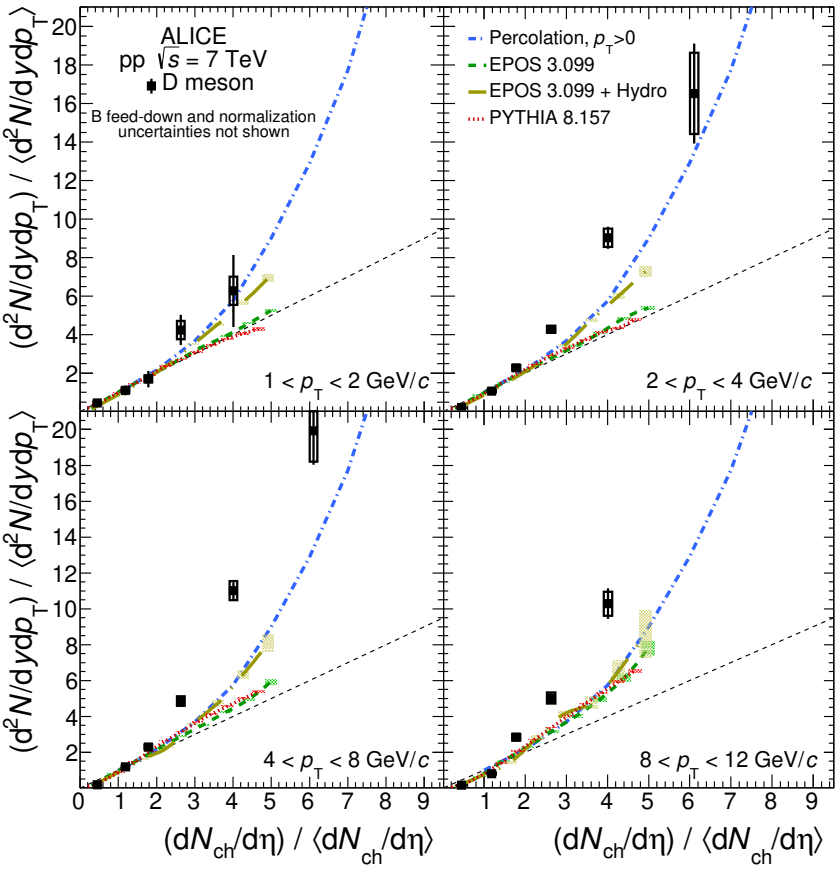


Figure II.5.2: Average prompt D meson (D^0, D^+, D^{*+}) relative yield as function of the relative charged-particle multiplicity in proton-proton collisions at $\sqrt{s} = 7$ TeV in different p_T intervals, measured by the ALICE Collaboration [208]. The measurement is compared to predictions from the percolation model, EPOS 3 and PYTHIA 8.157 models.

The EPOS 3 [211] model proposes a unified approach across collision systems from hadron-hadron to nucleus-nucleus collisions. A modified Gribov-Regge theory is used to describe the initial conditions of the collision. The parton-based Gribov-Regge theory [212] combines the multiple scattering theory and perturbative QCD where elementary hadron-hadron interactions, single scattering, are called pomerons. In EPOS3, a pomeron takes the form of a parton ladder including initial and final state radiations representing a quasi-longitudinal colour electric field treated as a relativistic string. At high multiplicity, a high string density is reached, each string is classified as being part of a low density region, the corona, or a high density region, the core. The corona hadronisation is performed via string fragmentation. The core is assumed to be thermalised and follow a hydrodynamical expansion. The hadronisation is then realised statistically.

In the Monte Carlo generator PYTHIA 8 [136], hard MPI are subsequent to the first hardest scattering. Heavy quarks production follows three main processes: 2→2 hard partonic processes, gluon splitting from hard processes and gluon splitting from initial/final state radiation. The colour

reconnection model tries to address the question of colour interaction (string) between partons coming from MPI. Partons coming from lower p_{T} interactions can be reconnected to harder ones if the reconnection probability exceed a given threshold. The heavy quark hadronisation is performed via the string fragmentation model (see section II.2.3). The contributions of the different charm quark production mechanisms to the total D meson production is studied in [208]. Initial and final state radiations are found to be the main contributors to the total D meson production (∼62%), MPI comes in second (∼21%) and then the first hardest scattering contribution (∼11%). The D meson production from the MPI and the gluon splitting processes increase with multiplicity. However for the D meson production from the hardest scattering production process, the increase reach a saturation at high multiplicity (see figure 9 in [208]).

II.5.2 Heavy-ion phenomenology effects

Recently, a few effects initially observed in heavy-ion collisions were also detected in small collision systems at high multiplicity, such as the positive elliptic flow of unidentified charged particles in proton-proton [10] and the enhanced production of multi-strange hadrons in proton-proton and proton-Pb [11].

The hint of more complex partonic interactions in dense QCD medium created at high multiplicity and their effects on particle production are studied in models, like for example with the colour reconnection mechanisms in general purpose Monte Carlo generators. The partonic density reached in small systems is such that hydrodynamic description of the medium [211] are proposed, allowing for the study of collectivity-like effects. More generally and in an attempt to provide an unified picture across collisions systems, the measurements performed in small systems at high multiplicity are bridging the gap between the studies of the hard and soft QCD production mechanisms of particles in hadron-hadron collisions and the studies of the quark-gluon plasma phenomenology (hydrodynamics, statistical physics principles, energy loss...) in heavy-ion collisions.

A potential presence of heavy-ion phenomenology effects in the heavy-flavour sector in small systems is investigated with measurements of charm and beauty hadron long-range correlations in small systems. The figure II.5.3 presents the elliptic flow $\nu_2^{\mathrm{sub}}(p_{\mathrm{T}})$ of D^0 mesons, J/Ψ, and light flavours for comparison purpose, in proton-proton and proton-Pb collisions at high multiplicity of charged particles [213]. In proton-proton collisions, a positive ν_2^{sub} is observed in the $2 < p_{\mathrm{T}} < 6$ GeV/c for prompt D^0, providing indications on the charm quark participation to possible collective effects. The magnitude of the ν_2^{sub} is found to be below the one of the K^0_s and Λ. A similar observation is made in proton-Pb collisions at high multiplicity where the prompt D^0 meson results are compatible with those of the prompt J/Ψ within uncertainties.

The Colour Glass Condensate (CGC) framework predicts a flavour hierarchy between $\nu_2^{\mathrm{sub}}(p_{\mathrm{T}})$ of prompt and non-prompt hadrons. The collectivity effects arise from the correlations between partons in the initial stage of the collision described as a gluon saturated system. The predictions for the prompt D^0 and J/Ψ describe the measurement, suggesting that collectivity might be generated by initial-state effects.

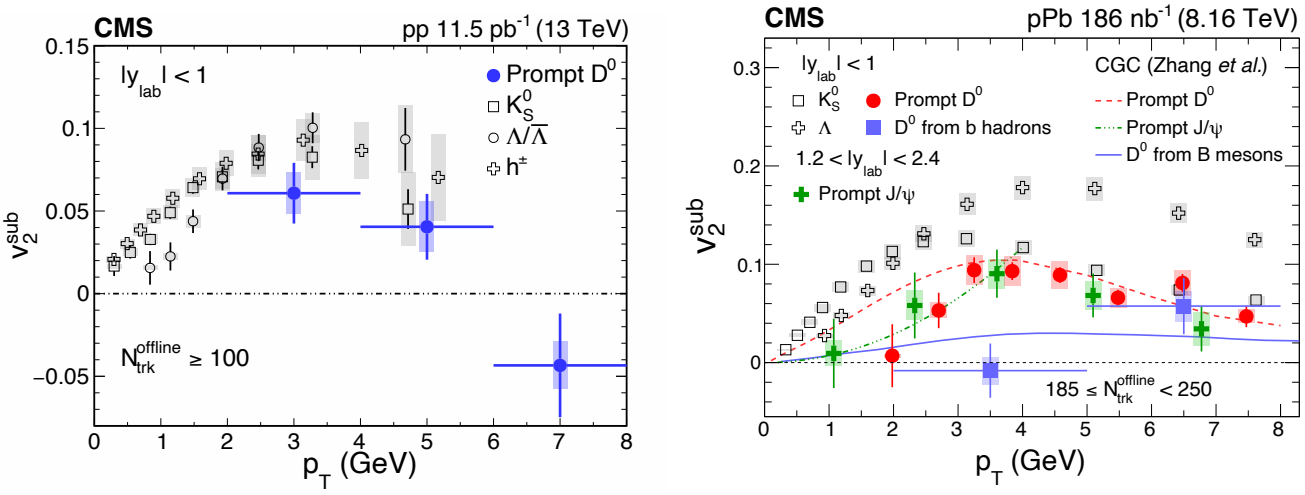


Figure II.5.3: Left panel: Elliptic flow $\nu_2^{\rm sub}$ of prompt $\rm D^0$ mesons, $\rm K^0_s$, Λ, and charged hadrons as a function of the particle transverse momentum in proton-proton collisions at $\sqrt{s}$ = 13 TeV for a multiplicity $N_{\rm trk^{offline}} \geq 100$ [213]. Right panel: Elliptic flow $\nu_2^{\rm sub}$ of prompt and non-promt $\rm D^0$ mesons, $\rm K^0_s$, Λ, and prompt $\rm J/\Psi$ as a function of the particle transverse momentum in proton-Pb collisions at $\sqrt{s}$ = 8.16 TeV for a multiplicity $185 \leq N_{\rm trk^{offline}} \leq 250$ [213].

The $\rm D_s^+$ meson

An increase of strange and multi-strange particle production is expected in the presence of a QGP statured in strange quark pairs (see section I.3.2). The ALICE Collaboration measured the relative strange particle production in the light-flavour sector in small systems where high multiplicity results reach the relative production in Pb-Pb collisions [11]. As discussed in section I.3.2, an enhanced strangeness production in small systems requires additional QCD interactions at high multiplicity such as colour ropes [69], hadronisation and collectivity assumptions similar to those made in heavy-ion collisions [211] or the use of statistical hadronisation models. The measurement of the charm-strange $\rm D_s^+(c\bar{s})$ meson production with respect to non-strange charm meson allows to probe strangeness enhancement in the heavy-flavour sector. By comparison with non-strange D meson results and model predictions, the measurement of the $\rm D_s^+$ over $\rm D^0$ $p_{\rm T}$-dependent production yield ratio as a function of multiplicity in small collision systems could provide additional insights on the charm hadronisation with strange quark in low- and high- density partonic systems.

Chapter III

The ALICE experiment at LHC

III.1 The Large Hadron Collider

The Large Hadron Collider [214] (LHC) is a hadron accelerator and collider installed at the European Organisation for Nuclear Research (CERN). It is composed of two rings (with counter-rotating beams) of 26.7 km of circumference designed to accelerate and deliver proton-proton collisions at a maximum centre-of-mass energy of $\sqrt{s} = 13$ TeV at a top instantaneous luminosity of $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1} and Pb-Pb collisions at a maximum centre-of-mass energy per nucleon of $\sqrt{s_{\rm NN}}$ = 5.5 TeV at a top instantaneous luminosity of $\mathcal{L} = 10^{27}$ cm^{-2}s^{-1}. The run I physics program of the LHC started from November 2009 to February 2013. During this period, the LHC delivered proton-proton collisions at $\sqrt{s} = 0.9$, 2.76, 5.02, 7 and 8 TeV, proton-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV and Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 2.76$ TeV. During the run II physics program (March 2015 - December 2018), the LHC delivered proton-proton collisions at $\sqrt{s} = 5.02$, 13 TeV, proton-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ and 8.16 TeV, Xe-Xe collisions at $\sqrt{s_{\rm NN}} = 5.44$ TeV and Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV.

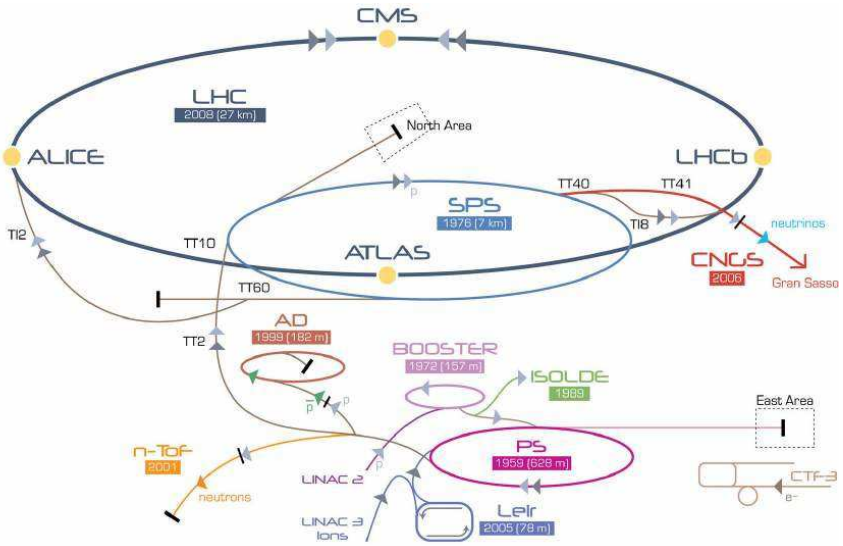


Figure III.1.1: Acceleration complex at CERN [215].

Four major experiments are located at different beam intersection points around the LHC as represented by the full yellow circles on figure III.1.1. The ATLAS (A Toroidal LHC Apparatus)

and CMS (Compact Muon Solenoid) experiments were designed to verify the predictions of the Standard Model of particle physics, particularly the characterisation of the Higgs boson properties discovered in 2012 [216, 217], and the search of physics beyond the Standard Model. The LHCb (Large Hadron Collider beauty) experiment is specialised in different aspects of beauty quark physics such as measurements of the CP violation in the B-meson sector. Additionally to their main physics program, the three experiments have also a heavy-ion physics program. ALICE (A Large Ion Collider Experiment) is the dedicated experiment to heavy-ion physics at LHC.

III.2 The ALICE experiment

The ALICE detector [218] is a general-purpose detector aiming to study the physics of strongly interacting matter and the QGP in ultra-relativistic heavy-ion collisions. The specificities of its design are,

- a charged particle tracking down to $p_{\mathrm{T}} \sim 0.1$ GeV/c allowed by the high granularity of the detectors and the low material budget which is about 13% of X_0, the radiation length, at the outer radius of the Time Projection Chamber sub-detector (see below). By comparison, the tracking detectors at central rapidity of CMS and ATLAS have respectively material budgets of 36% and 47% of X_0.
- good particle identification capabilities with multiple identification methods (energy losses, time-of-flight, transition radiation, ring-imaging Cherenkov, electromagnetic calorimeter).
- the ability to deal with the high charged particle multiplicity at mid-rapidity reached in central heavy-ion collisions, up to $\mathrm{d}N/\mathrm{d}\eta = 4000$ [218].

The overall dimensions of the ALICE apparatus are of 16x16x26 m^3 for a total weight of roughly 10 000 t. It is composed of 19 detector systems as shown on figure III.2.1. The ALICE coordinate system is a right-handed Cartesian system with the z-axis parallel to the beam direction pointing in the opposite direction to the muon spectrometer and the y-axis is oriented upward. The azimuthal angle ϕ is defined in the transverse plane (x, y), the pseudo-rapidity η = - ln(tan(θ/2)) is defined with respect to the z-axis direction.

The central barrel detectors are located in the pseudo-rapidity region $|\eta| < 0.9$ inside the solenoid magnet (L3 magnet on the figure) generating a maximum magnetic field of $B = 0.5$ T parallel to the beam direction. Starting from the Inner Tracking System (ITS), closest to the interaction point, the successive detectors are: the Time Projection Chamber (TPC) detector, the Transition Radiation Detector (TRD), the Time Of Flight (TOF) detector, the ring-imaging Cherenkov (HMPID) detector, and two electromagnetic calorimeters (EMCal and PHOS). To the exception of the EMCal and HMPID, all central barrel detectors covers the full azimuth. At forward and backward rapidities, the Photon Multiplicity Detector (PMD), the Forward Multiplicity Detector (FMD), the quartz Cherenkov detector T0, the plastic scintillator detector V0 and the Zero Degree Calorimeter (ZDC) are installed. The muon spectrometer is composed by the succession of the muon tracker, muon wall and muon trigger. The exhaustive list of the ALICE detectors with their respective acceptance, position, technology and main purposes is given in table 1 of [219].

In the following sections an emphasis is made on the detectors used to perform the data analysis presented in chapter IV.

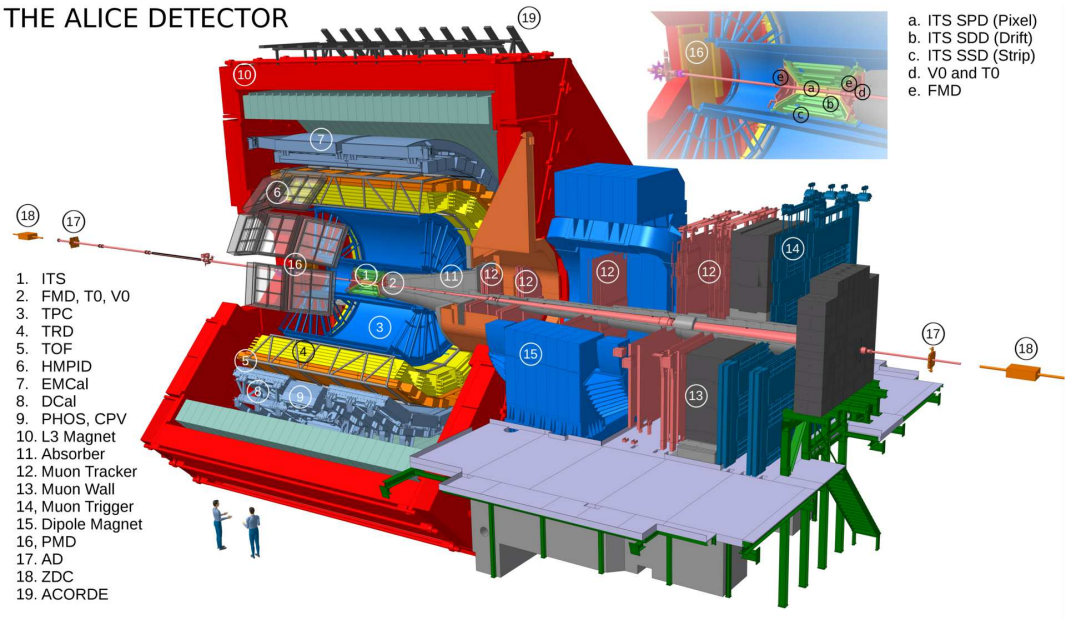


Figure III.2.1: The ALICE detector. The top right panel is a zoom on the sub-detectors at the collision point. Figure from ALICE figure repository ©.

III.2.1 Inner Tracking System

The Inner Tracking System is composed of 3 x 2 cylindrical layers of silicon detectors of different technologies. It is located at a radial position between 3.9 and 43 cm covering the pseudo-rapidity range $|\eta| < 0.9$ and full azimuth for vertices within $\pm$ 5.3 cm along the beam direction around the detector centre. The low material budget of the ITS (7.7% of X_0 for $|\eta| = 0$ particles) limits the multiple scattering of low momentum particles needed for efficient track finding and high impact-parameter resolution of secondary tracks.

The Silicon Pixel Detector (SPD) consists of the two first sensitive layers of the ITS, situated at a the transverse positions $r = 3.9$ and 7.6 cm away from the beam line where the charged particle density is expected to reach 50 tracks/cm^2. It is composed of 9.8 x 10^6 pixel cells of size 50 μm ($r\phi$) x 425 μm (z) and a thickness of 200 μm. The intrinsic spatial resolution is of 12 μm ($r\phi$) x 100 μm (z) and a two-track separation resolution of 100 μm ($r\phi$) x 850 μm (z). When a pixel cell (or a group of them) detects a particle signal above threshold, a Fast-OR digital pulse is generated. Due to its fast response time, about 100 ns, the SPD can be used as an input for the trigger system. The SPD is fundamental for the primary vertex reconstruction and the determination of the impact parameter of secondary tracks originating from the weak decays of strange, charm, and beauty hadrons.

Next to the (SPD) comes the Silicon Drift Detector (SDD), forming the two intermediate layers of the ITS at $r = 15$ and 23.9 cm where the charged particle density is expected to reach 7 tracks/cm^2. It is made of sensors having a thickness of 300 μm and a sensitive are of $\sim$ 70 mm ($r\phi$) x 75 mm (z) split in two drift regions. The position of the crossing particle is calculated from the drift time of the electrons originating from the interaction of the particle with the detector, with respect to the trigger time. The spatial resolution of the detector is of 35 μm ($r\phi$) x 25 μm (z) and a two-track resolution of 200 μm ($r\phi$) x 600 μm (z). In addition to the tracking, the SDD provides 2 of the 4 energy loss samples needed for the particle identification (PID) with the ITS.

Finally the Silicon Strip Detectors (SSD) complete the tracking and PID of the ITS. The two

outer layers, essential for the matching of tracks from the TPC to the ITS, are located at r = 38 and 43 cm. The sensors are of thickness 300 μm and have a sensitive are of 73 mm ($r\phi$) x 40 mm (z) composed of 768 strips on each side with a pitch of 95 μm. The spatial resolution of the detector is of 20 μm ($r\phi$) x 830 μm (z) and a two-track resolution of 300 μm ($r\phi$) x 2400 μm (z). A stereo angle of 35 mrad is given between the strips of the two sides, this choice is a compromise between the resolution on the $r\phi$ direction and two-track separation at high particle densities. Overall, the ITS is designed to provide :

- primary vertices reconstruction with a spatial resolution below 100μm.
- reconstruction of secondary vertices coming from the weak decay of strange, charmed and bottom hadrons.
- track reconstruction efficiency above 90%.
- identification of low momentum particles $p_t < 200$ MeV/c.

III.2.2 Time Projection Chamber

The Time Projection Chamber is the main tracking device of the central barrel providing also energy loss information for particle identification. It covers the pseudo-rapidity range $|\eta| < 0.9$ for a full radial track length with matching in the ITS, TRD and TOF detectors and full azimuth to the exception of dead zones. It is a 500 cm long cylindric detector of inner radius 85 cm and outer radius 250 cm with a sensitive volume of a 90 m^3 filled with a gas mixture, a 100 kV central electrode, and two end plates of 18 trapezoidal readout sectors each where multi-wire proportional chambers are mounted. The trapezoidal sectors, each of them covering a 20° angle, are composed of 159 pad rows, a pad corresponding to a detection cell. The uniform electrostatic field generates a 400 V.cm^{-1} drift field with a maximum drift time of 92 μs. The gas mixture composition evolved with time, at the start of the run II a mixture of $Ar/CO_2/N_2$ (90/10/5) was used and in 2017 the argon was replaced back by neon, the original composition of the LHC run I, due to larger space-charge distortions observed. From the inner to outer radii, the spatial resolution of the detector is, $\sim$ 1100 μm ($r\phi$) x 1250 μm (z) to $\sim$ 800 μm ($r\phi$) x 1100 μm (z). The momentum resolution is about 2% for charged particles with $p_{\mathrm{T}} \sim 10$ GeV/c. The resolution on the linear energy loss on tracks reconstructed with 159 clusters is of 5% for isolated tracks (proton-proton collision conditions) and of 7% in central Pb-Pb collisions.

III.2.3 Time Of Flight detector

The Time-Of-Flight detector is a cylindrical detector covering the pseudo-rapidity range $|\eta| < 0.9$, with an inner radius of 370 cm and an outer radius of 400 cm. The detector is based on the Multi-gap Resistive-Plate Chamber technology. This detector is used for particle identification at intermediate momentum, up to 2.5 GeV/c for pions and kaons and 4 GeV/c for protons with a pion-kaon and proton-kaon separation resolution more than 3 times the time-of-flight resolution. The time of flight of a particle is calculated from the difference between the hit time measured by the detector and the start time of the event determined by the T0 detector or the particle arrival times at the TOF [220]. The overall time resolution of the detector is about 100 ps in proton-proton collisions and of 60-80 ps in Pb-Pb collisions. The lower resolution in proton-proton collisions is explained by the larger uncertainty on the event start time.

III.2.4 V0 detector

The V0 detector is composed of two arrays of scintillators (V0A and V0C) located on either side of the interaction point. The V0A is on the opposite side of the muon spectrometer, at 340 cm from the central barrel geometric centre, covering the pseudo-rapidity region $2.8 < \eta < 5.1$. The V0C is in front of the hadronic absorber, at 90 cm from the central barrel geometric centre, covering the pseudo-rapidity region $3.7 < \eta < 1.7$. Each V0 is made of 32 counters arranged in 4 rings and 8 sectors of 45% covering the full azimuth. The V0 detector is used as an input for the trigger system, together with the SPD detector it defines the minimum-bias trigger (see sections IV.1.1 and IV.1.2). With a time resolution of the order of a nano-second, beam-residual gas background interactions event are discarded. It is used to estimate the multiplicity of events using the correlations between the amplitude of the V0 signals and the number of charged particles produced in a collision.

III.3 Track and vertex reconstruction

The track and vertex reconstruction procedures process the data recorded by the detectors with the aim to reconstruct the collision interaction point and the kinematics of particles in the collision. In ALICE, for the LHC run I and II, this reconstruction was performed offline.

III.3.1 Primary vertex reconstruction

The starting point of the primary vertex reconstruction is the conversion from hits in the ITS detector to clusters with an associated position (gravity centre of the hit cloud), signal amplitude and timing information and their related uncertainties. Clusters in the two layers of the SPD within an azimuthal window are associated by pairs to form track segments, called *tracklets*. The first evaluation of the primary vertex is done by finding the space point to which a maximum number of tracklets converge. The spatial coordinates of the collision vertex are determined by minimisation of the distance between the group of N associated tracklets and the vertex. The procedure is performed iteratively, the clusters of the previously reconstructed vertex being discarded for the iteration. The vertex attached to the highest number of tracklets is defined as the primary vertex. Alternatively, if no convergence point is found, as it is possible in low multiplicity events, the collision vertex z position is obtained from the distribution of the points of closest approach of the tracklets to the average position of the beam in the transverse plane. The preliminary primary vertex position is later refined by using the global tracks reconstructed with both the ITS and the TPC detectors.

On the left panel of figure III.3.1, the evolution of the transverse resolution of the primary vertex position distribution with the charged particle multiplicity of proton-proton collisions is shown. The resolution is a combination of the resolution of the luminous region σ_{D} and the vertex resolution $\alpha/\sqrt{\mathrm{d}N/\mathrm{d}\eta}$. It is about 200 μm at low multiplicity, $\mathrm{d}N/\mathrm{d}\eta \sim 2$, and about 50 μm at high multiplicity, $\mathrm{d}N/\mathrm{d}\eta \sim$35-40 for primary vertices reconstructed with global tracks. As the multiplicity of the collision increases, the primary vertex position becomes more constrained as more tracks contribute to the minimisation procedure. A better spatial resolution is obtained by using global tracks due to taking into account the bending of the particle trajectories in the magnetic field.

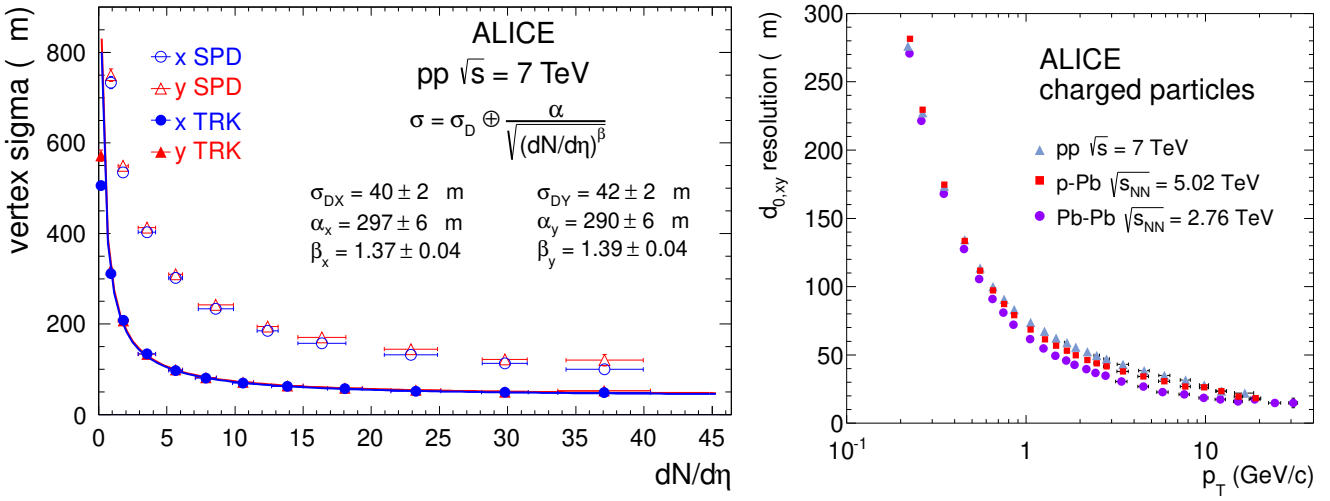


Figure III.3.1: Left panel: Transverse resolution of the primary vertex position distribution as a function of the charged particle multiplicity in proton-proton collisions at $\sqrt{s}$ = 7 TeV. The vertices are reconstructed using the SPD tracklets (open points) and the global track (full points) algorithms. Right panel: Resolution of the transverse impact parameter of the global ITS-TPC tracks as a function of the track transverse momentum p_{T} in proton-proton, proton-Pb and Pb-Pb collisions. Figures taken from [221].

III.3.2 Track reconstruction

The tracks finding and reconstruction is performed using the Kalman filter algorithm [222], an iterative two-phase procedure. During the prediction phase, the track parameters estimated from a given cluster are extrapolated to the next one. The extrapolation includes the multiple scattering and energy loss effects from the particle-matter interactions. In the filtering phase, based on proximity criteria, the cluster closest to the extrapolation point is attached to the track. The track parameters are then recalculated including this new cluster.

The first stage of the track reconstruction is performed in the "inward" direction from the outer wall of the TPC to the estimated primary vertex [223]. The starting point of the track search is done at large TPC radii where the track density is relatively low. The TPC trapezoidal sectors being made of 159 pad rows (see section III.2.2), in the ideal case, a track crossing the TPC volume produce 159 clusters. Combining two TPC clusters with the primary vertex, track seeds are build. Following the Kalman filter procedure, the seeds are propagated inward, updating iteratively the track parameters as clusters are added to the "growing" track. Additional quality criteria are applied to reject tracks sharing a large fraction of its cluster with another track(s) (between 25% and 50% [219]). Tracks having at least 20 clusters attached and at least 50% of the clusters expected for their given track positions are kept. The TPC tracks are then matched to hits in the outer layer of the SSD and used as seeds for their propagation in the ITS. Detection inefficiency due to ITS dead zones are considered in the layer-by-layer propagation by penalising tracks with missing clusters. At the end of the track reconstruction in the ITS, the tree of possible track prolongations associated to each TPC track is sorted by their reduced χ^2. Only the branch with the best χ^2 is retained.

During the second tracking stage, the track are propagated in the outward direction, from their point of closest approach to the preliminary primary vertex towards the clusters reconstructed in

in the TRD and TOF detectors. Inside the ITS and TPC, the Kalman filter procedure re-uses the already associated clusters. Tracks reaching the TRD are updated using the information form the TRD clusters and then are matched with the clusters in the TOF detector.

A final inward track propagation from the outer radius of the TPC is done to refine the track parameters estimation as the final track position, direction, inverse curvature and the associate covariance matrix are determined.

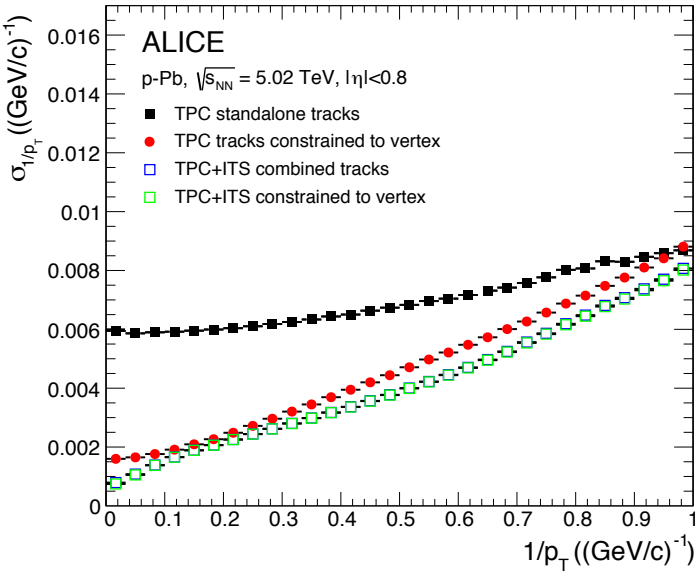


Figure III.3.2: The resolution on the track transverse momentum p_{T} for standalone TPC and global ITS-TPC tracks reconstructed with and without the constraint to the vertex as a function of the track inverse momentum in p-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV [221].

The right panel of figure III.3.1 shows the evolution of the resolution on track impact parameter in the transverse plane as a function of the track transverse momentum in proton-proton, proton-Pb and Pb-Pb collisions. As the p_{T} of the tracks increases, the resolution on the impact parameter improves from 75 μm for $p_{\mathrm{T}} = 1$ GeV/c to 20 μm at $p_{\mathrm{T}} = 20$ GeV/c in proton-proton collisions. The improved resolution in proton-Pb and Pb-Pb collision systems is due to the more precise determination of the primary vertex position at higher multiplicities.

The figure III.3.2 show the resolution on the inverse p_{T} of the reconstructed TPC standalone tracks and global ITS-TPC tracks in proton-Pb collisions. The quantity is directly extracted from the track parameters covariance matrix and is related to the relative momentum resolution via,

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} = \frac{\sigma_{1/p_{\mathrm{T}}}}{1/p_{\mathrm{T}}}$$

Overall the p_{T} resolution degrades with increasing p_{T} as the track curvature reduces reaching the asymptotic straight line. As expected, the resolution on the global tracks is better with respect to the TPC standalone tracks. However, by constraining the track parameters with the primary vertex position during the Kalman filter procedure, the resolution of the TPC standalone track greatly improves reaching the global track resolution for $p_{\mathrm{T}} < 5$ GeV/c. For global tracks, the vertex constraint effect is negligible, the relative p_{T} resolution evolve from 1 to 10% between 1 and 50 GeV/c. Below 1 GeV/c the relative resolution degrades, increasing up to 5%, due to the multiple scattering effect becoming important.

III.4 Particle identification

Several ALICE systems are used to perform particle identification (PID) on charged hadrons and leptons. Absorbers are installed in front of the muon spectrometer to stop all particles except muons. Electrons are identified with the electromagnetic calorimeter EMCal and in the TRD via transition radiation measurements. The charged particle PID with the ITS is performed through the particle charge deposit in the SDD and SSD layers providing energy loss $\mathrm{d}E/\mathrm{d}x$ measurements mainly relevant for particle with $p_{\mathrm{T}} \leq 0.7$ GeV/c. With the High-Momentum Particle Identification Detector, the PID is performed via the Cherenkov angle measurement. The TPC and the TOF detectors are the key detectors of the hadron identification in the central barrel region. As they were used in the analysis presented in this thesis (chapter IV), their PID strategies are described in the following sections.

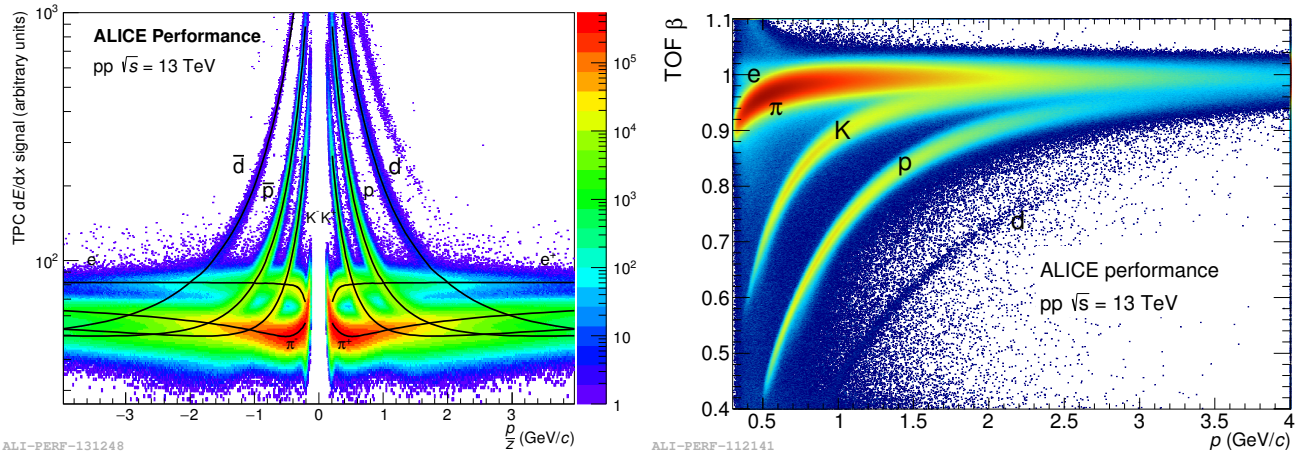


Figure III.4.1: Left panel: particle energy losses in the TPC as a function of the rigidity in proton-proton collisions at $\sqrt{s} = 13$ TeV. The curves show the Bethe-Bloch parametrisation of the expected mean energy loss for electron (e), pion (π), kaon (K), proton (p), and deuteron (d). Right panel: velocity β measured by the TOF detector as a function of the particle momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV. Figure from ALICE figure repository © [224].

III.4.1 TPC particle identification

In the TPC, the particle identification information is provided by the simultaneous measurements of the charge deposits along the trajectory and the momentum of the particle crossing the gas volume. The PID is based on the measured energy loss of a particle at a given momentum compared to the expected value for a given particle hypothesis. The expected value is derived from a parametrisation of the mean energy loss per distance of unit, described by the Bethe-Bloch formula, proposed by the ALEPH collaboration [225]:

$$f(\beta\gamma) = \frac{P_1}{\beta^{P_4}}\left(P_2 - \beta^{P_4} - \ln\left[P_3 + \frac{1}{(\beta\gamma)^{P_5}}\right]\right) \quad \text{with} \quad \beta\gamma = \frac{p}{M \cdot c} \tag{III.1}$$

with p the momentum, M the mass, and $\beta = v/c$ the relative velocity of the particle, γ the Lorentz factor, and P_{1-5} free parameters determined from fit.

The particle energy loss in the TPC as a function of the particle rigidity (momentum over charge ratio) measured in proton-proton collision at $\sqrt{s} = 13$ TeV is shown on figure III.4.1 together with the expected mean energy loss curves for electrons, pions, kaons, protons and deuterons. In the $p/z \leq 1$ GeV/c region, the clear separation between the different particle species allows for track-by-track particle identification basis. The separation between the measured and the expected energy loss is expressed in number of σ, the expected resolution for a given particle specie at a given momentum. At higher momenta, in the relativistic rise region ($p/z \geq 2$ GeV/c), particle identification is performed with a statistical approach via multi-Gaussian fits of the energy loss distributions in momentum intervals. The resolution on the energy loss is about 5% in proton-proton collision for isolated tracks reconstructed with 159 clusters [219].

III.4.2 TOF particle identification

The TOF detector performs particle identification by measuring the time intervals between moments a collision happens and a particles arrives to this detector. The collision time can be estimated online with the T0 detector having a time resolution of ∼40 ps in proton-proton collisions and 20-25 ps in PbPb collisions. When no timing information from the T0 is available, the start time of the event is determined offline averaging the calculated arrival times for at least three particles. With 30 particle the event start time resolution is about 30 ps [220]. Knowing the starting time and arrival time of a particle, its velocity $\beta = v/c$ is determined with the travelled distance obtained from its reconstructed trajectory. The TOF detector resolution includes the uncertainty on the starting time, the tracking and momentum resolutions.

The performances of the detector are shown on the right panel of figure III.4.1 with the measured velocity β distribution as a function of the particle momentum. In the momentum $0.2 < p < 4$ GeV/c interval, a good separation of the particle species is obtained. Especially, the proton-koan separation is at more than 3σ level in this interval.

III.5 The ALICE offline framework

The ALICE collaboration's offline software framework, called AliRoot [226], is based on the ROOT framework dedicated to large scale data analysis [227]. AliRoot is used for the simulation, alignment, calibration, reconstruction, and analysis of the experimental data collected by the experiment during the LHC run I and II. For the LHC run III an IV, the new ALICE Online-Offline (O^2) framework will be used [228].

The AliRoot framework provides,

- an interface to external Monte Carlo generators used to generates collision events. The data obtained contains the identification, kinematics and the decay chain of each generated particle. The PYTHIA [134,136] and HERWIG [121] Monte Carlo generators are commonly used for proton-proton collisions.
- the complete description of the ALICE detector geometry and response. The code for simulation and reconstruction for each sub-detector are done separately. The transport Monte Carlo framework GEANT 3 [229], GEANT 4 [230] and FLUKA [231] are example of frameworks used to perform the particle transport through the ALICE detector.
- frameworks for detector alignment and calibration.

- an event reconstruction framework.
- a data analysis framework with code dedicated to the different physics studies performed within the collaboration.
- data formats: the RAW data format as coming out from the readout electronics of the detectors, the Event Summary Data (ESD) containing the results of event reconstruction and the Analysis Object Data (AOD) containing reconstructed data filtered for specific physics analyses.

The processing of the recorded data is performed on the Worldwide LHC Computing Grid (WLCG), a hierarchical computing infrastructure shared by the LHC experiments [232]. The Tier-0 computing centre is a reunion of the CERN computing centre and the computing centre at the Wigner Research Centre for Physics in Budapest. For the ALICE collaboration, the Tier-0 centre copy, store and perform the first reconstruction of the RAW data. The Tier-1 centre regroup different large regional computing centres scattered across the world where a copy of the RAW data is stored and additional reconstruction processing are performed. Smaller centres forming the Tier-2 dedicated to the production of Monte Carlo simulations and the physics analysis of the data.

Chapter IV

Data analysis

In this chapter, the measurements of $\mathrm{D_s^+}$ meson production at central rapidity ($|y| < 0.5$) in proton-proton collisions at the energy of $\sqrt{s} = 13$ TeV are presented. The transverse-momentum differential production cross section for $\mathrm{D_s^+}$ meson is expressed as,

$$\frac{\mathrm{d}^2\sigma^{\mathrm{D_s^+prompt}}}{\mathrm{d}p_\mathrm{T}\mathrm{d}y} = \frac{1}{\mathcal{L}_\mathrm{int}\cdot \mathrm{B.R.}}\frac{1}{\Delta p_\mathrm{T}}\frac{f_\mathrm{prompt}(p_\mathrm{T})\cdot\frac{1}{2}\cdot N_\mathrm{raw}^{\mathrm{D_s^\pm}}(p_\mathrm{T})\Big|_{|y|<y_{fid}(p_\mathrm{T})}}{(Acc\cdot\epsilon)_\mathrm{prompt}(p_\mathrm{T})\cdot c_{\Delta y}(p_\mathrm{T})} \tag{IV.1}$$

The transverse-momentum differential production yield per event for $\mathrm{D_s^+}$ meson in a given charged particle multiplicity of the collision interval, is expressed as,

$$\frac{1}{N_\mathrm{mult}^\mathrm{evt}}\frac{\mathrm{d}^2N_\mathrm{mult}^{\mathrm{D_s^+prompt}}}{\mathrm{d}p_\mathrm{T}\mathrm{d}y} = \frac{\epsilon_\mathrm{mult}^\mathrm{INEL}}{N_\mathrm{mult}^\mathrm{evt}}\frac{1}{\mathrm{B.R.}}\frac{1}{\Delta p_\mathrm{T}}\frac{f_\mathrm{prompt}(p_\mathrm{T})\cdot\frac{1}{2}\cdot N_\mathrm{raw,mult}^{\mathrm{D_s^\pm}}(p_\mathrm{T})\Big|_{|y|<y_{fid}(p_\mathrm{T})}}{(Acc\cdot\epsilon)_\mathrm{prompt,mult}(p_\mathrm{T})\cdot c_{\Delta y}(p_\mathrm{T})} \tag{IV.2}$$

Starting from the extracted raw signal $N_\mathrm{raw}^{\mathrm{D_s^\pm}}$ in a p_T interval of width Δp_T (section IV.7), the quantity is corrected by the $\mathrm{D_s^+}$ meson reconstruction and selection efficiencies $(Acc\cdot\epsilon)_\mathrm{prompt}$ and by the rapidity coverage correction factor $c_{\Delta y}$ (section IV.8). The prompt fraction f_prompt calculation is detailed in section IV.9. The corrected yields are divided by the branching ratio, B.R. = 2.24 ± 0.08 %, of the studied $\mathrm{D_s^+} \to \phi(1020)(\to \mathrm{K^+K^-})\pi^+$ decay channel [37]. The cross section and the production yield formula differs in their normalisation. The cross section, eq. IV.1, is normalised by the integrated luminosity $\mathcal{L}_\mathrm{int} = 31.7 \pm 0.51\ \mathrm{nb}^{-1}$ of the data sample. In each charged particle multiplicity interval, the production yields, eq. IV.2, are obtained normalising by the trigger efficiency $\epsilon_\mathrm{mult}^\mathrm{INEL}$ and the number of collisions $N_\mathrm{mult}^\mathrm{evt}$ (section IV.10).

IV.1 Data samples and event selection

IV.1.1 Data samples and trigger strategies

The presented analyses are performed on the data samples coming from proton-proton collisions at mass centre energy of $\sqrt{s} = 13$ TeV recorded during the LHC run II over a period of three years from spring 2016 to autumn 2018.

The total proton-proton inelastic collision cross-section includes the single diffractive, double diffractive and non-diffractive interactions. Diffractive interactions generally involve a relatively small momentum transfer between the two colliding protons, after interaction, the proton(s)

break(s) up into multi-particle final states with the same quantum numbers as the initial protons. Non-diffractive interactions are dominant in the high momentum transfer regime.

The integrated luminosity collected in proton-proton collisions at $\sqrt{s} = 13$ TeV for different ALICE triggers over the whole LHC run II period is shown on figure IV.1.1. The plateaus on the curves correspond to the LHC winter shutdown periods. The triggers of interest for this analysis are the minimum-bias (blue curve), the high multiplicity SPD (dark-green curve) and the high multiplicity V0 (light-green curve) triggers [221, 233]. The corresponding number of collisions recorded, *i.e.* events, are mentioned on the figure.

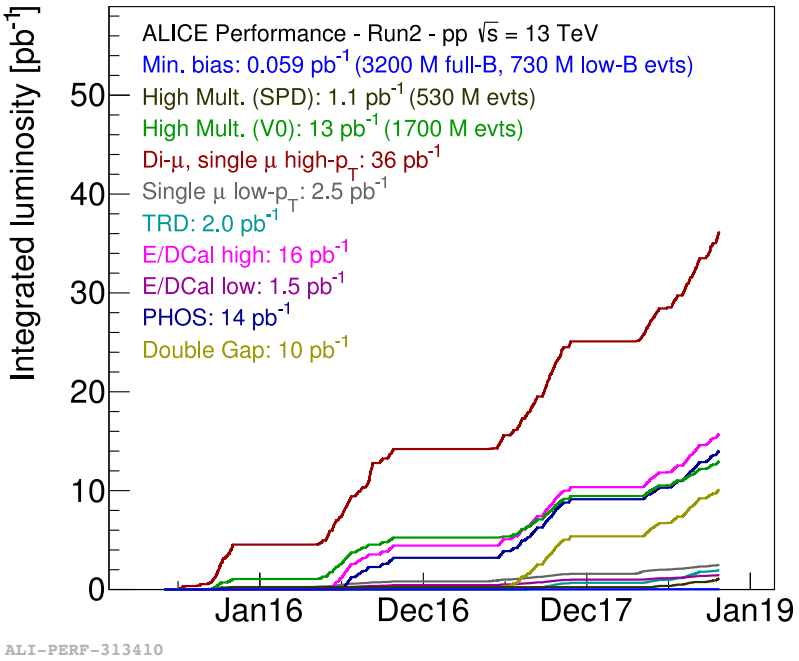


Figure IV.1.1: Integrated luminosity for various triggers in proton-proton collisions at $\sqrt{s} = 13$ TeV during LHC run II (2015-2018).

The following trigger acronyms will be employed in the text: MB for the minimum bias, HMSPD for the high multiplicity SPD and HMV0M for the high multiplicity V0 trigger. The table IV.1.1 summarises the different data samples used to perform the analysis. Each year of data taking is divided in periods themselves divided in runs. The interaction rate, *i.e.* the frequency of inelastic proton-proton collisions, varies frequently from run to run depending on the filling scheme and other parameters of the LHC. For the selected periods, it goes from 1.5 kHz to 240 kHz in 2016, from 1 kHz to 200 kHz in 2017, and from 2 kHz to 250 kHz in 2018. The minimum-bias and the HMV0M triggers were available during the full data taking periods while the HMSPD was only introduced in 2018. About 1.71 billion of minimum-bias events were selected for the total minimum-bias data sample. Periods with non-optimal trigger calibration are removed from the data samples leading to fewer number of periods kept for the high multiplicity triggered data samples. The number of events selected for the high multiplicity data samples will be discussed after introducing the multiplicity of event definitions and corrections discussed in the next sections.

The minimum bias trigger is used to perform an event selection avoiding any bias regarding the original physical process. Most of the beam induced background interactions are rejected while

performing the offline event selection (section IV.1.2), by requiring coincident signals in the V0A and V0C detectors, *i.e.* happening within a 8 ns window around the beam collision time [234]. Beam induced background interactions originates from inelastic interactions of the beam with collimators close to the ALICE detector (protecting the sensible sub-systems) and from interactions with the beam-residual gas in the vacuum pipe.

The dedicated high multiplicity triggers [221, 233] were used to enrich the statistics of high multiplicity events. They both require a certain threshold indicating a (high) number of charged particles in the event. The energy deposit in the two V0 scintillators is proportional to the number of primary charged particles produced in this detector's pseudo-rapidity range ($3.7 < \eta < 1.7$ and $2.8 < \eta < 5.1$). The high multiplicity V0 triggers are issued when the V0 amplitude reaches a certain threshold value around ∼2700 (arbitrary unit). The high multiplicity SPD trigger is based on the number of fired FAST-OR ships of the Silicon Pixel Detector outer layer (see section III.2.1). A multiplicity threshold is set around 80-100 hits correspond roughly to 60-80 tracklets. The threshold value chosen is a compromise between the reduction of low-multiplicity event rate and minimum contamination from pile-up events.

year	**trigger**	**periods**	**number of events**
2016	MB	d, e, g, h, j, k, l, o, p	$424 \cdot 10^6$
	HMV0	k, l, o, p	
2017	MB	e, f, h, i, j, k, l, m, o, r	$580 \cdot 10^6$
	HMV0	e, f, h, i, k, l, m, o, r	
2018	MB	b, d, e, f, g, h, i, j, k, l, m, n, o, p	$706 \cdot 10^6$
	HMV0	b, d, e, f, h, j, k, l, m, o, p	
	HMSPD	f, h, j, k, l, m, o, p	

Table IV.1.1: Summary of the different periods included in the data samples used for the data analysis.

IV.1.2 Offline event selection

An additional event selection is performed offline in order to refine the rejection of beam-gas interaction background events using the differences of time of flight of particles in the two V0 scintillators [234].

Different selection procedures are also applied to discard multiple interaction vertices in the same bunch crossing (in-bunch pileup events) affecting every detectors in the same way and across bunch crossing (out-of-bunch pileup events) depending on different integration time of the detectors. The pile-up effects can bias the estimation of the charge particle multiplicity produced in the event which are critical for data analyses. The "multi-vertexer" algorithm with global tracks (reconstructed in both the ITS and the TPC detectors) is used to tag in-bunch and out-of-bunch pileup by finding multiple primary vertices. Once the vertex with the highest number of attached tracks is identified (the "main" vertex), additional pile-up vertices are searched using the remaining tracks not pointing to the main vertex. When available, the time of flight information on the track is used to check if it can contribute to the vertex based on the bunch crossing time. A given event can be tagged as pileup event by applying selection criteria on the additional vertices found. The tagging efficiency, *i.e.* the proportion of events successfully tagged as pileup, and the false positive rate, *i.e.* the proportion of event wrongly tagged as pileup, depend both on the selection criteria

and the event multiplicity.
For the analysis presented in this thesis, an event is kept if the following criteria are met. They are obtained from dedicated studies performed within the ALICE Collaboration:

- a minimum number of 5 tracks associated with pile-up vertex candidates.
- a $\chi^2/\text{ndf} < 5$ per tracks (obtained from the tracking procedure) in order to avoid combinatorial background and potential additional vertices contributions.
- the minimum weighted longitudinal distance between the main and secondary vertex candidates, $(z_{\text{main}} - z_{\text{pile-up}})/\sqrt{\sigma^2_{\text{main}} + \sigma^2_{\text{pile-up}}} > 15$.

Out-of-bunch pileup occurs when collisions from different bunches happened within the integration time of the detectors. Integration times is quite different across ALICE detectors, for example, the SPD readout time is ~ 300 ns and for the TPC it is ~ 90 μs. This is to compare with a typical LHC proton bunch spacing time of 25 ns. A past-future protection is implemented to remove out-of-bunch pileup by applying selection cuts based on the correlations between detectors with different timing windows and on a direct detection of pile-up with fast detectors (TOF, V0, T0 detectors, see section III.2). An event is rejected if there is any additional V0 activity during the SPD integration time within $\pm$ 10 bunch crossings around the triggered event. Residual out-of-bunch pileup events are further discarded with offline correlation studies between the SPD and V0 detectors. The impact of potentially remaining pile-up events is on the percent level and does not influence the final results.

The selected events are required to have their collision vertex reconstructed with global tracks in the longitudinal window $|z_{vertex}| < 10$ cm around the nominal position in order to assure a pseudo-rapidity coverage of $|\eta| < 0.9$ for tracks in the central barrel detectors and avoid acceptance border effects.

IV.2 Multiplicity definition and corrections

Two different multiplicity estimators are used: the number of SPD tracklets $N_{tracklets}$ within the pseudo-rapidity interval $|\eta| < 1$ and the V0M signal amplitude from the V0A and V0C detectors located in the two pseudo-rapidity regions: $3.7 < \eta < 1.7$ and $2.8 < \eta < 5.1$, respectively. Since the D_s^+ mesons are reconstructed and selected at mid-rapidity, potential auto-correlation effects can arise with the estimation of the event multiplicity in the same rapidity window (here with the $N_{tracklets}$ estimator). By using multiplicity estimators defined in common and different rapidity regions with respect to the particle-of-interest measurement, one can study such effects.

IV.2.1 Multiplicity estimation with the SPD detector

Correction to the number of tracklets

The measured distributions of the raw number of SPD tracklets in minimum-bias collisions in the three different years of data recording (2016, 2017 and 2018) are shown on the right panel of figure IV.2.1. The differences between the multiplicity distribution in the 2016 data sample with respect to the 2017 and 2018 data samples are shown in the distribution ratios on the right panel of figure IV.2.1. The decreasing trends of the multiplicity distribution ratios is a reflection of the decreasing acceptance of the SPD detector over time (some of the pixel chips becoming dead). The

effect increases as a function of multiplicity, reaching a value of ∼20% and ∼50% at $N_{tracklets}$ = 80 for the 2017 and 2018 distributions respectively.

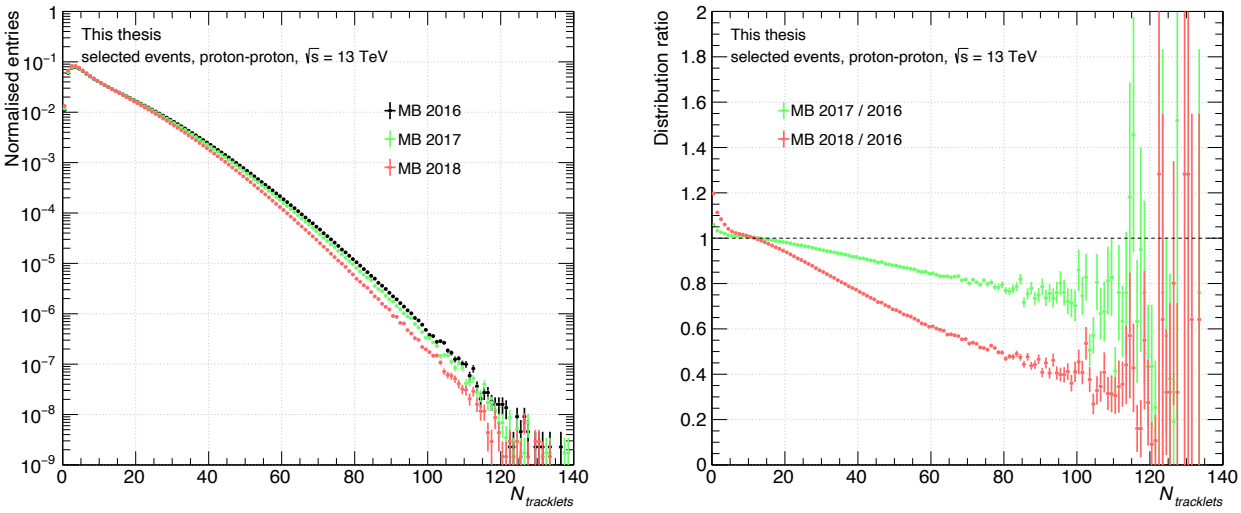


Figure IV.2.1: The measured raw number of SPD tracklets distribution in minimum-bias proton-proton collisions at $\sqrt{s}$ = 13TeV in the three different years of data taking. On the right panel, the ratio of the distribution in the 2017 (green) and 2018 (red) data samples to the one obtained in 2016.

One can also observe the evolution of the measured $N_{tracklets}$ distribution as a function of the longitudinal collision vertex position, z_{vertex} in the different years of data taking. The $N_{tracklets}$ versus z_{vertex} 2D distributions are presented the figure IV.2.2 for the 2016, 2017, 2018 minimum-bias collisions data samples with their corresponding tracklet profiles, i.e. the mean number of tracklets as a function of the vertex position, plotted in red. The profiles show an overall decrease of the $N_{tracklets}$ going from positive to negative z_{vertex} values with an acceptance falling more steeply close to the longitudinal window limits $|z_{vertex}| \gtrsim 5.5$ cm. The trend is due to the inhomogeneous distribution and dead areas of the SPD layers and the overall evolution of the number of active modules during each data taking period.

The uncorrected tracklet profiles per periods per years are shown on the top panels of the figure IV.2.3. The per-period tracklets profiles differ within a same year. The spread of profiles seems negligible in 2016, while it becomes more important in 2017 and 2018. In order to correct the distributions for both the SPD ageing over time and the z_{vertex} dependence of the acceptance, a data driven equalisation procedure is applied event-by-event and for different periods.
The correction is applied as follow:

$$N_{corr} = N_{raw} \pm Poisson\left[N_{raw} \cdot \left(\frac{\langle N_{ref}\rangle}{\langle N_{period}(z)\rangle} - 1\right)\right] \tag{IV.3}$$

with $\langle N_{ref}\rangle$ the reference number of tracklet defined as the maximum measured over all the periods considered. For the full minimum-bias data sample, $\langle N_{ref}\rangle = 12.25$ at $z_{vertex} = 5.14$ cm, the value is found in the 2016 h period. $\langle N_{period}(z)\rangle$ correspond to the mean number of tracklets at a given z_{vertex} for the current period considered.

The corrective term is randomly computed following the Poisson statistics. The corrected tracklet profiles are on the bottom panels of the figure IV.2.3 show the effect of the correction procedure. All periods present an equalised and flat z_{vertex} distribution of the $N_{tracklets}$ with a mean value sitting around $N_{tracklets} = 12.75$ and compatible within 2% along z_{vertex}.

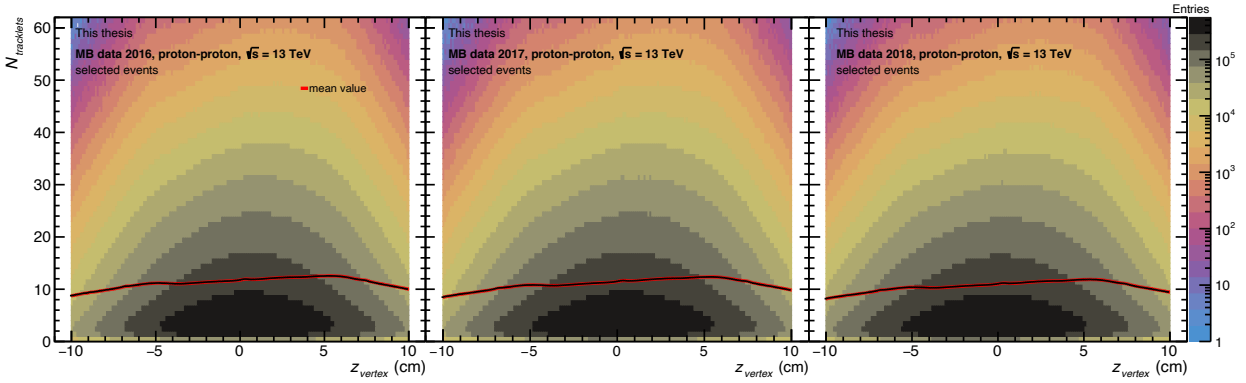


Figure IV.2.2: Distributions of the raw number of tracklet as a function of z_{vertex} for the 2016 (left panel), 2017 (middle panel), and 2018 (right panel) minimum-bias collisions data samples. The tracklet profiles are draw in red.

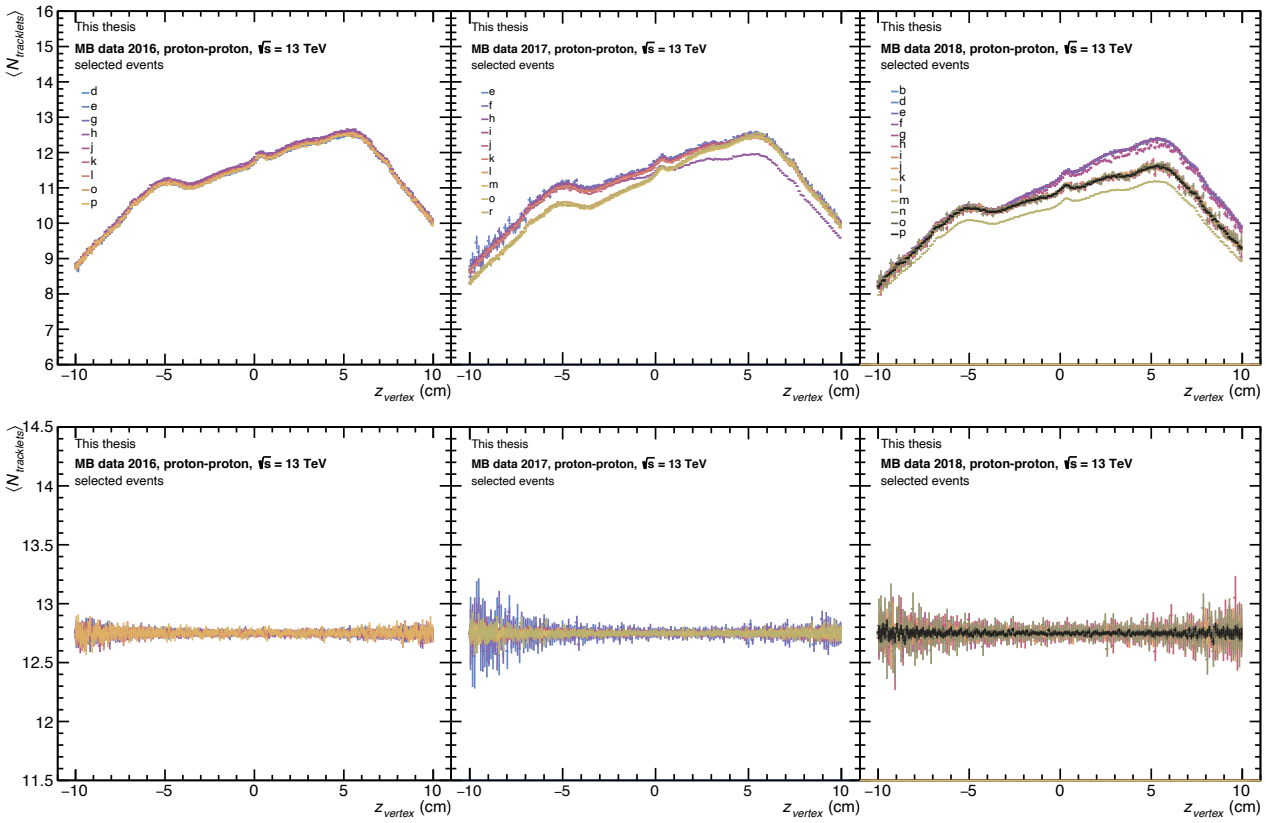


Figure IV.2.3: Raw tracklet profiles (top panels) and their corresponding corrected distributions (bottom panels) as a function of z_{vertex} for the different period of the 2016, 2017, and 2018 minimum-bias collisions data samples.

The effect of the correction procedure on the global $N_{tracklets}$ distributions per year can be seen on the figure IV.2.4. In this analysis, the total minimum-bias collisions data sample is exploited in the $N_{tracklets} \in [1, 59]$ interval. The zoomed plot of the left panel on figure IV.2.4 in the region of interest is presented on the right panel, showing an overall change of $\pm 1\%$ in the 2017 data sample and of $\pm$ 3% for the 2018 data sample.

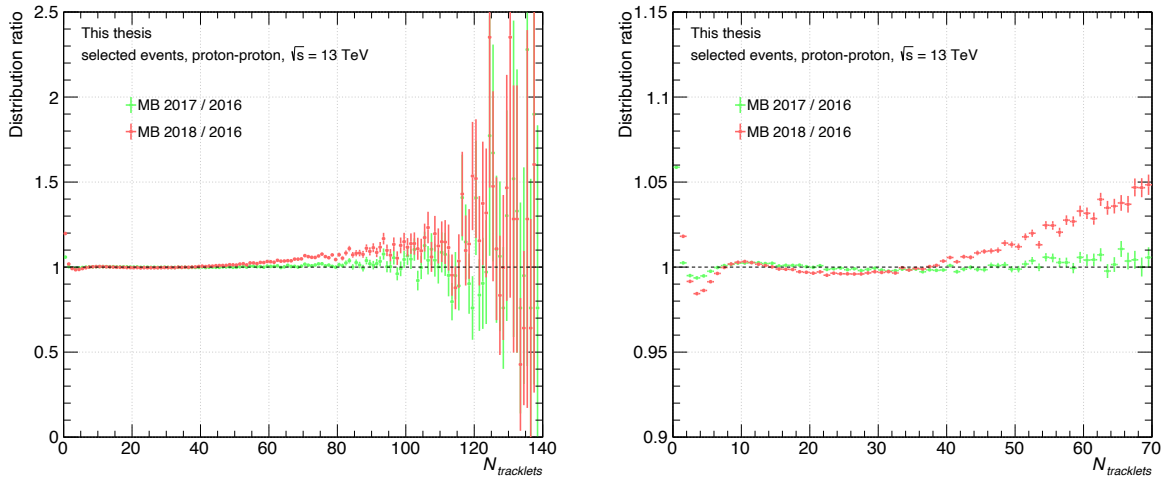


Figure IV.2.4: The ratio of the corrected $N_{tracklets}$ distribution in the 2017 (green) and 2018 (red) data samples to the one obtained in 2016. On the left panel, a zoomed plot in the 0 to 70 $N_{tracklets}$ interval.

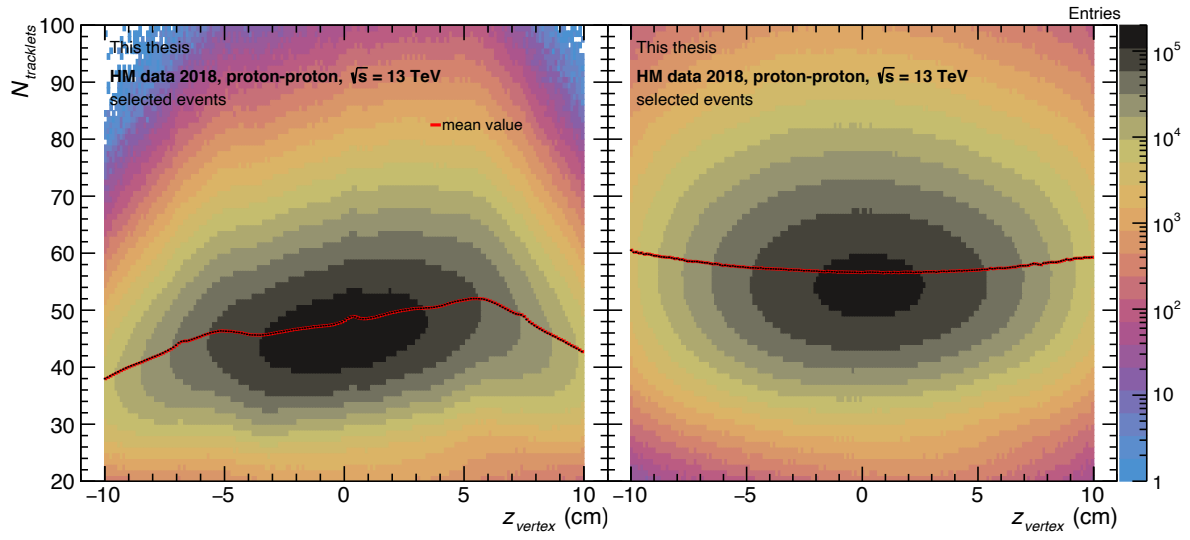


Figure IV.2.5: Distributions of the raw (left panel) and corrected (right panel) number of tracklet as a function of z_{vertex} for the high multiplicity collisions data sample. The tracklet profiles are draw in red.

The uncorrected tracklet distribution of the high multiplicity collision data sample, shown on the left panel of figure IV.2.5, exhibit a very similar trend as seen in the minimum-bias data

sample. Therefore, the $N_{tracklets}$ distributions are corrected using the profiles and the reference of the minimum-bias data sample. On the right panel of figure IV.2.5, one can see the effect on the corrected $N_{tracklets}$ distribution. The $N_{tracklets}$ profile, plotted in red, is not perfectly equalised as a function of z_{vertex} particularly in the $|z_{vertex}| > 5$ cm region where the profile slowly rises. However most of the statistics lay in the $|z_{vertex}| < 5$ cm region where the profile is reasonably flat. This correction choice, preferred over a dedicated high multiplicity equalisation, is further justified by the dedicated high multiplicity trigger efficiency correction discussed in section IV.6.

Estimation of the charged particle density

The multiplicity estimated with $N_{tracklets}$ in $|\eta| < 1$ is then converted to the average charge particle density $\langle \mathrm{d}N_{ch}/\mathrm{d}\eta \rangle$ to allow for the comparison with model predictions independent from the ALICE experiment. The conversion is based on the correlation between the number of tracklets and the number of primary charged particle, N_{ch}, in the same pseudo-rapidity region. A primary charged particle is defined by the ALICE Collaboration as a charged particle with a mean proper lifetime $\tau > 1\mathrm{cm}/c$ which is either produced directly in the collision interaction, or from the decays of particles with $\tau < 1\mathrm{cm}/c$, excluding particles produced in interactions with material [235]. The correlation from $N_{tracklets}$ to N_{ch} is obtained with Monte Carlo simulations. The $N_{tracklets}$–N_{ch} 2D correlation plot is shown on the figure IV.2.6 for both the uncorrected $N_{tracklets}$ (left panel) and the corrected $N_{tracklets}$ (right panel). The equalisation of $N_{tracklets}$ versus z_{vertex} is consistently performed as in the data, using the same $N_{tracklets}$ profiles and reference value. The 2D distribution with equalised $N_{tracklets}$ is more symmetric with respect to the linear fit performed on the data and show a clear linear correlation between the two quantities.

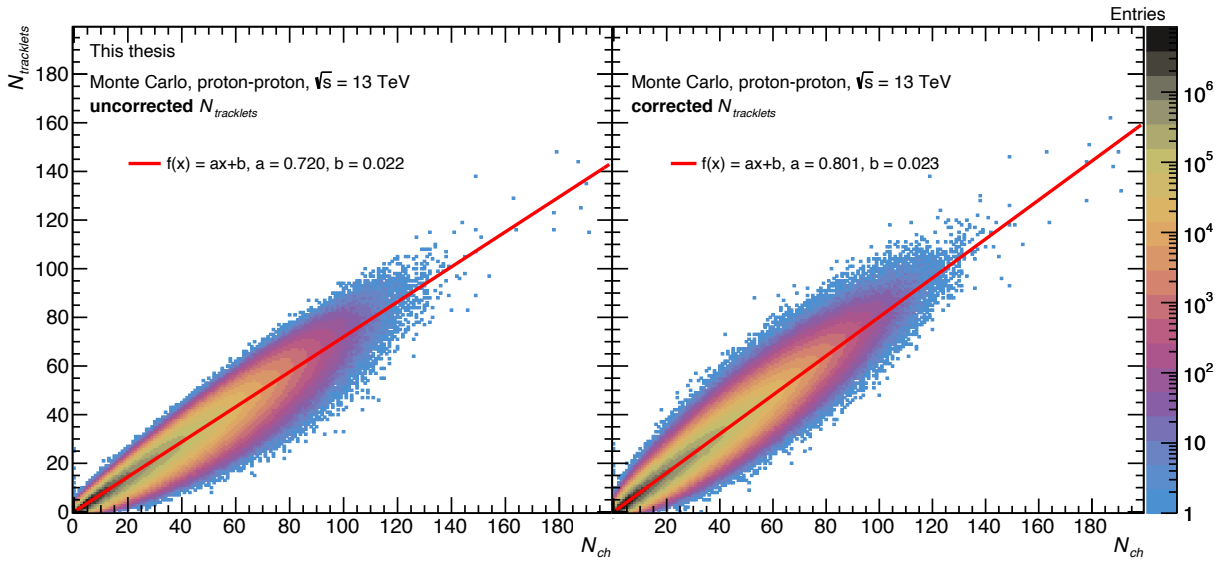


Figure IV.2.6: Correlation distribution between the uncorrected $N_{tracklets}$ (left panel) and corrected $N_{tracklets}$ (right panel) versus the number of primary charged particle N_{ch} within $|\eta| < 1$. A linear fit is performed and plotted in red on the two histograms.

The mean and the root mean square (RMS) values of the number of primary charged particle in each multiplicity interval are obtained by projecting the corrected $N_{tracklets}$ correlation distribution. These projections in the different multiplicity intervals are shown on the left panel of figure IV.2.7

for the whole analysed data sample. The uncertainties associated to the N_{ch} mean and RMS value of the global projection distributions are estimated by applying the same procedure to each of the data-taking periods. The obtained values are summarised in the table IV.2.1.

The profile distribution of the corrected $N_{tracklets}$–N_{ch} correlation distribution shown on the right panel of figure IV.2.7 is used to extract the N_{ch} minimum and maximum values for each corrected $N_{tracklets}$ interval considered. The central points and the uncertainties of the points correspond to the $N_{tracklets}$ mean and standard deviation obtained in the different N_{ch} bins. Two different fitting strategies are considered to convert the corrected $N_{tracklets}$ interval limits, (a) a global linear fit with the function $f(x) = ax + b$ over the whole N_{ch} range and (b) linear fits with the function $f(x) = ax$ in each corrected $N_{tracklets}$ interval. The fits are shown on the right panel of figure IV.2.7. The minimum and maximum values extracted with the two strategies are summarised in the table IV.2.1.

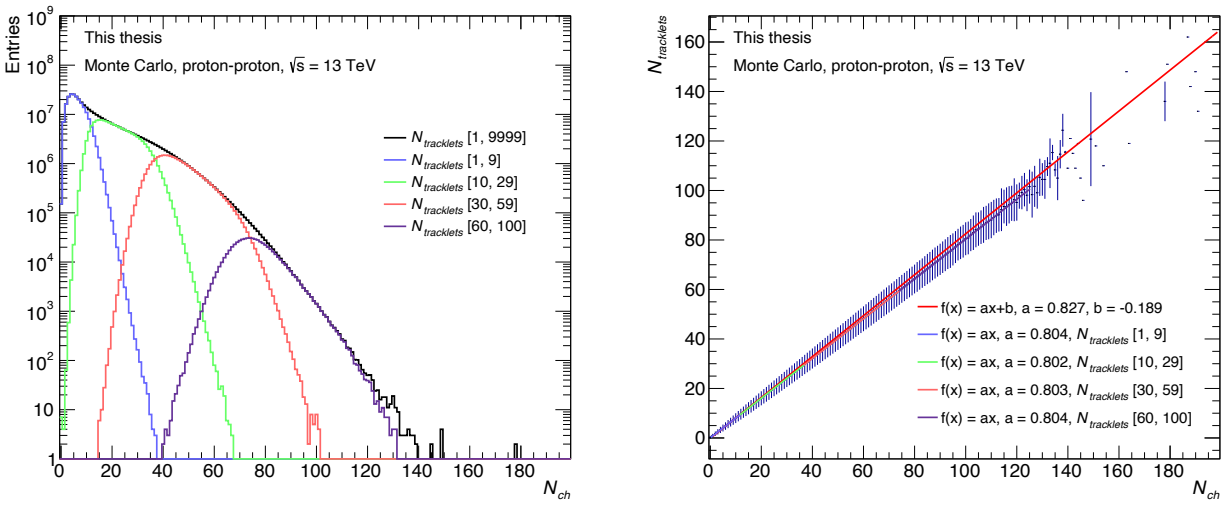


Figure IV.2.7: Left panel: Distributions of the primary N_{ch} for different corrected $N_{tracklets}$ intervals. Right panel: The mean $N_{tracklets}$ distribution as a function of N_{ch}, the uncertainties on the mean values correspond to the $N_{tracklets}$ standard deviation in each of the N_{ch} bins. Linear fit are performed on the distribution in different multiplicity ranges (see text for details).

corrected $N_{tracklets}$	$\mathrm{d}N_{ch}/\mathrm{d}\eta$			
interval	**mean**	**RMS**	**[min, max] (a)**	**[min, max] (b)**
[1, 999]	7.48 ± 0.02	6.49 ± 0.00	/	/
[1, 9]	3.10 ± 0.02	1.64 ± 0.02	[0.72, 5.56]	[0.62, 5.59]
[10, 29]	10.54 ± 0.09	3.77 ± 0.04	[6.16, 17.65]	[6.23, 18.08]
[30, 59]	22.56 ± 0.21	4.51 ± 0.06	[18.25, 35.79]	[18.69, 36.75]
[60, 100]	37.83 ± 0.62	4.52 ± 0.12	[36.40, 60.58]	[37.33, 62.21]

Table IV.2.1: Summary of the conversion from the corrected $N_{tracklets}$ intervals to the density of charged particle in $|\eta| < 1$.

IV.2.2 Multiplicity estimation with the V0A and V0C detectors

The multiplicity estimator at forward rapidity is based on the sum of the energy deposits in the V0A and V0C detectors. The distribution of the V0M amplitude normalised by its average value $\langle \text{V0M} \rangle$ is shown on figure IV.2.8 for minimum-bias and high multiplicity triggered proton-proton collisions at $\sqrt{s} = 13$ TeV. The V0M multiplicity classes are defined by a percentile interval which can interpreted as the fraction of the proton-proton inelastic collision cross section triggered by the V0A and V0C detectors (*i.e.* the visible cross section). High charged-particle multiplicity collisions, will fall in the low percentile value classes as shown on the figure IV.2.8 with the high multiplicity triggered data. Relatively low charged-particle multiplicity collisions, which constitute the majority of the V0M minimum-bias visible cross section, will get assigned to larger percentile value classes.

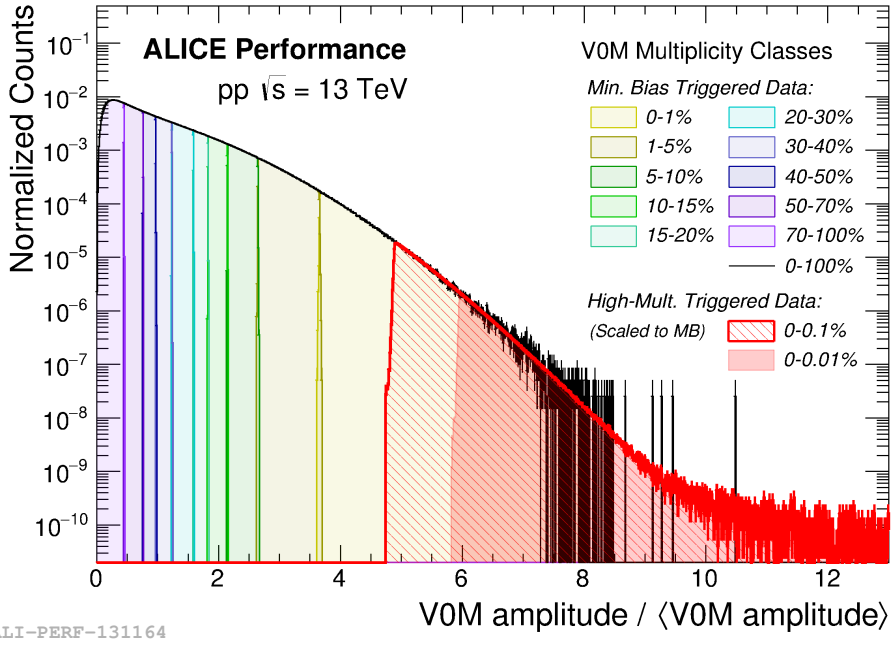


Figure IV.2.8: V0M amplitude distribution normalised by the average value $\langle \text{V0M} \rangle$ in minimum-bias and high multiplicity triggered proton-proton collisions at $\sqrt{s} = 13$ TeV.

The figure IV.2.9 shows the average $\langle \text{V0M} \rangle$ amplitude distribution as a function of the longitudinal collision vertex position in the different periods of the minimum-bias collision data-sample. No significant z_{vertex} dependence is observed, therefore no correction is applied when using this multiplicity estimator. Still, one can see the spread of the $\langle \text{V0M} \rangle$ amplitude measured across the periods of one single year of data taking, reflecting the ageing of the detector. This is why the percentile of the V0M amplitude is preferred as multiplicity estimator over the single $\langle \text{V0M} \rangle$ amplitude, since it is a relative quantity computed run per run, they are self corrected from the degrading performance of the detectors over time.

Since the measured V0M percentiles are independent from their data taking period, the normalised V0M percentile distributions in the different periods of the collision data samples are expected to be flat. Their distribution is shown on the top panels, for the minimum-bias collision data sample, and on the bottom panels, for the high-multiplicity collision data sample, of figure IV.2.10. A slight decrease of the V0M percentile minimum-bias distributions is observed with

increasing V0M percentile, however they can be considered flat within 5% band over the whole range. The large uncertainties of the period 2018 is due to the low statistics collected. For the V0M percentile high multiplicity distributions, going to the low percentile values, the distribution rise and becomes flat in the 0-0.1% region, indicating the full efficiency of the high multiplicity trigger. Because of a lack of proper calibration, several period were rejected form the high multiplicity data sample with respect to the minimum-bias data sample.

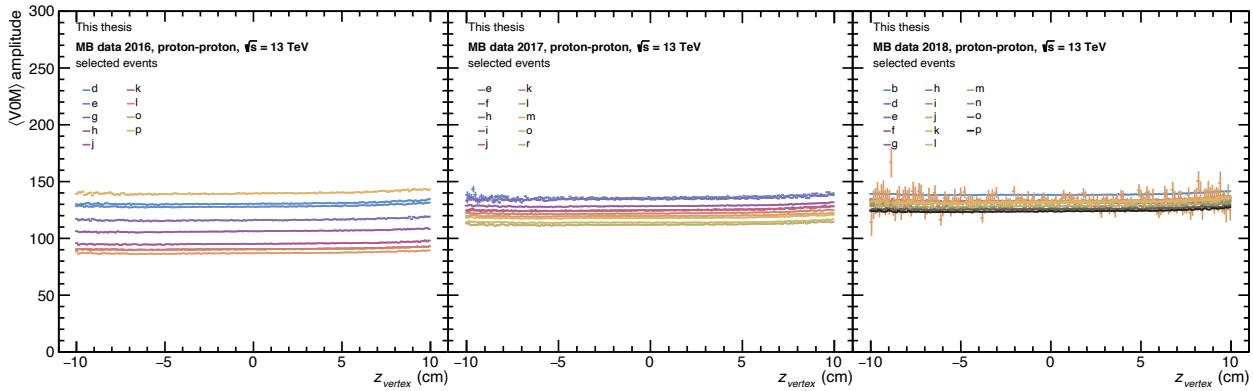


Figure IV.2.9: Distribution of the average ⟨V0M⟩ amplitude as a function of z_{vertex} in the different periods on the 2016 (left), 2017 (center) and 2018 (right) minimum-bias collision data samples.

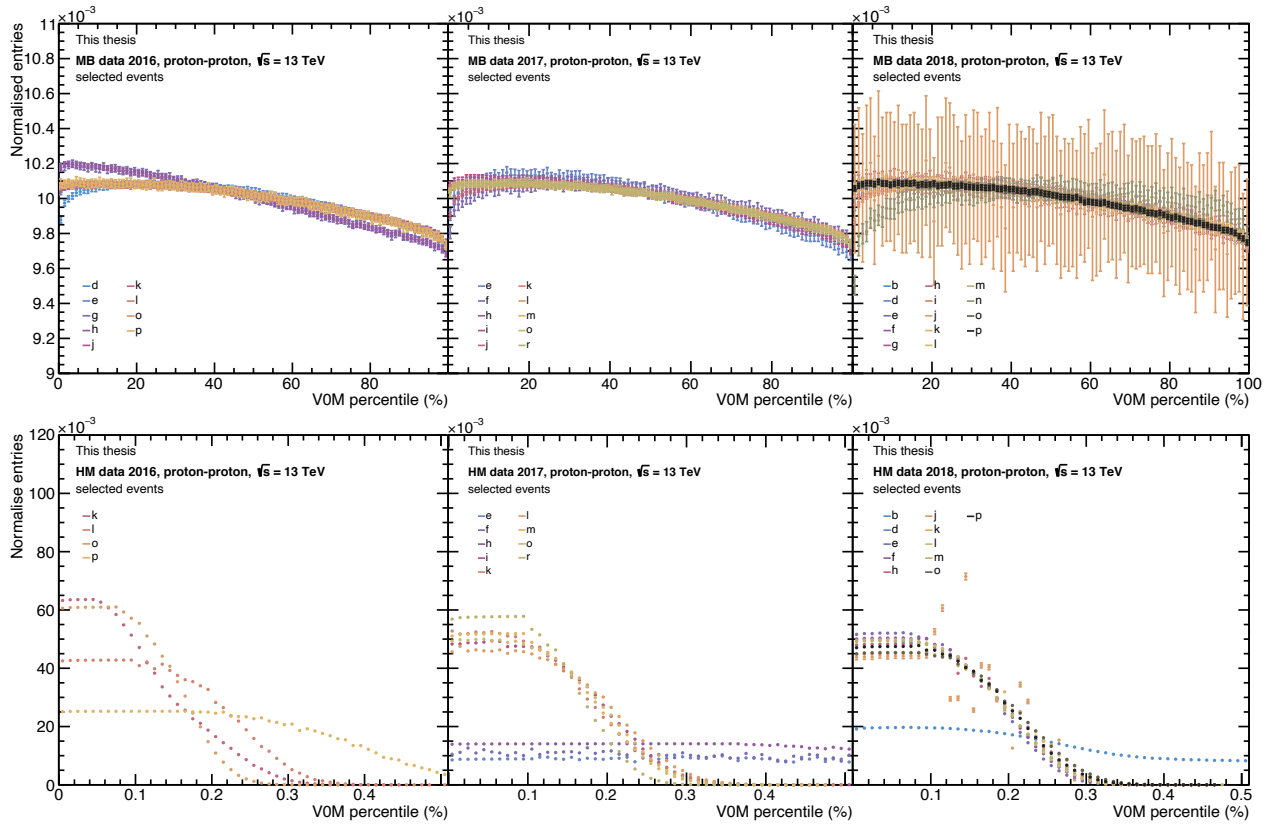


Figure IV.2.10: Normalised V0M percentile distributions in the different periods of the minimum-bias (top panels) and high-multiplicity (bottom panels) proton-proton collision data samples.

IV.2.3 Multiplicity intervals

The different multiplicity intervals with the two estimators are summarised in table IV.2.2. The INEL>0 multiplicity class correspond to inelastic collision having at least one charged particle in the pseudo-rapidity interval $|\eta| < 1$. This class of events correspond to about 75% of the total inelastic proton-proton cross section [236]. Experimental measurements done with the INEL>0 event selection facilitate the comparison with theoretical models.

In the data analysis, the triggered and selected events are required to have at least one SPD tracklet in $|\eta| < 1$. While it is done per definition in the different multiplicity intervals of the $N_{tracklets}$ analysis, in the V0M percentiles analysis the additional requirement is made and indicated with the INEL>0 subscript. To correct for the fraction of event fulfilling the INEL>0 condition but not triggered, the trigger efficiency factor, $\epsilon_{\text{mult}}^{\text{INEL}}$, is applied to the number of event $N_{\text{mult}}^{\text{evt}}$ used for normalisation. They are computed in dedicated analyses performed within the collaboration [236]. In the high multiplicity $N_{tracklets}$ [60, 100] interval, the high multiplicity SPD trigger is not fully efficient over the whole interval, the trigger efficiency correction is done following a data driven procedure detailed in section IV.6. In the V0M percentile [0, 0.1] interval the high multiplicity V0M trigger is fully efficient as seen on figure IV.2.10 so no trigger correction is applied.

The multiplicity intervals are converted to the corresponding mean charged particle density in $|\eta| < 0.5$. The mean charged particle densities, $\langle \mathrm{d}N_{ch}/\mathrm{d}\eta \rangle$, corresponding to the V0M percentile intervals with the additional INEL>0 selection were computed by performing a weighted average of the values obtained in finer intervals from the dedicated study [236]. For the $N_{tracklets}$ analysis, the conversion is done following the procedure described in section IV.2.1. In the $N_{tracklets}$ [1, 9999] interval, the procedure gave $\langle \mathrm{d}N_{ch}/\mathrm{d}\eta \rangle = 7.48 \pm 0.02$ which is different form the INEL>0 event class value 6.93 ± 0.09 since INEL>0 events with $N_{tracklets} = 0$ were not taken into account. The later value is retained since the measurement in the multiplicity intervals are corrected for INEL>0 events. The two multiplicity intervals: $N_{tracklets}$ [1, 9999] and V0M percentile $[0, 100]_{\text{INEL>0}}$, are equivalent as they both correspond to a multiplicity integrated interval with the INEL > 0 requirement.

As it will be discussed in the section IV.4, the reconstructed tracks corresponding to the $\mathrm{D_s^+}$ meson decay daughter charged particles are required to have a hit in the SPD detector. Therefore, the $N_{tracklets} > 0$ selection is equivalent to the INEL>0 when classifying the $\mathrm{D_s^+}$ mesons in the different multiplicity intervals. Moreover, the signal detection efficiency correction is negligible since no signal loss is expected due to both requirements on the selected events and the reconstructed $\mathrm{D_s^+}$ mesons.

The number of events used for the normalisation of the corrected yield is computed as follows:

$$\begin{aligned} N_{\text{mult}}^{\text{evt}} &= N_{\text{mult}}^{\text{reco vtx}}(|z_{vertex}| < 10 \text{ cm}) + N_{\text{mult}}^{\text{no vtx}}(|z_{vertex}| < 10 \text{ cm}) \\ &= N_{\text{mult}}^{\text{reco vtx}}(|z_{vertex}| < 10 \text{ cm}) + N_{\text{mult}}^{\text{no vtx}} - N_{\text{mult}}^{\text{no vtx}}(|z_{vertex}| > 10 \text{ cm}) \\ &= N_{\text{mult}}^{\text{reco vtx}}(|z_{vertex}| < 10 \text{ cm}) + N_{\text{mult}}^{\text{no vtx}} - N_{\text{mult}}^{\text{no vtx}} \cdot \frac{N_{\text{mult}}^{\text{reco vtx}}(|z_{vertex}| > 10 \text{ cm})}{N_{\text{mult}}^{\text{reco vtx}}} \end{aligned} \tag{IV.4}$$

with $N_{\text{mult}}^{\text{reco vtx}}$ and $N_{\text{mult}}^{\text{no vtx}}$ being the number of events with and without a primary vertex reconstructed respectively. $N_{\text{mult}}^{\text{reco vtx}}(|z_{vertex}| < 10 \text{ cm})$ correspond to the number of events satisfying the criteria described in section IV.1.2. The ratio of the $N_{\text{mult}}^{\text{evt}}$ normalisation factor over the number of selected event is shown on figure IV.2.11 for the minimum-bias event interval (i.e. event selected without any multiplicity or $N_{tracklets}$ requirement) and for the $N_{tracklets}$ and V0M percentile intervals. The event correction factor to $N_{\text{mult}}^{\text{reco vtx}}(|z_{vertex}| < 10 \text{ cm})$ is of the order of 4% at low multiplicity and becomes negligible in the highest multiplicity intervals.

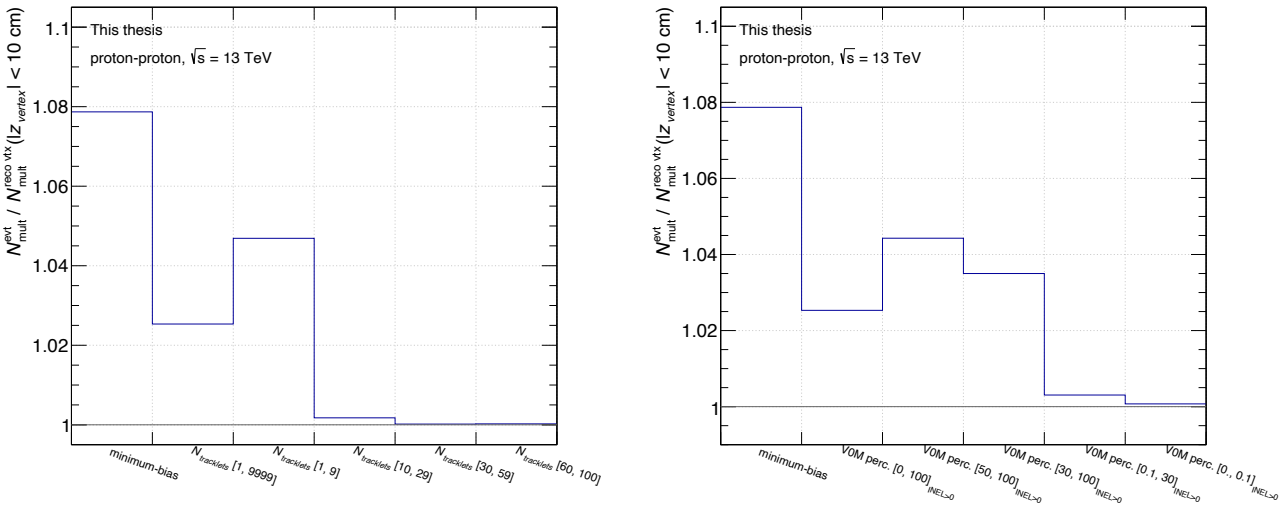


Figure IV.2.11: Ratio of the number of event $N_{\mathrm{mult}}^{\mathrm{evt}}$ (see text for the definition) over the number of selected events $N_{\mathrm{mult}}^{\mathrm{reco\ vtx}}(|z_{vertex}| < 10$ cm) in the different $N_{tracklets}$ (left panel) and V0M percentile intervals (right panel). The ratio in the minimum-bias analysis is shown in the first bin of the two histograms.

$N_{tracklets}$	$\langle \mathrm{d}N_{ch}/\mathrm{d}\eta \rangle$	trigger	trigger efficiency ($\epsilon_{\mathrm{mult}}^{\mathrm{INEL}}$)	$N_{\mathrm{mult}}^{\mathrm{evt}}$ (x 10^9)
[1, 9999]	6.93 ± 0.09	MB	0.92 ± 0.003	1.724
[1, 9]	3.10 ± 0.02	MB	0.862 ± 0.015	0.919
[10, 29]	10.54 ± 0.09	MB	0.997 ± 0.002	0.667
[30, 59]	22.56 ± 0.21	MB	1.0 ± 0.0	0.135
[60, 100]	37.83 ± 0.62	HMSPD	1.0 ± 0.0	0.119
V0M percentile (%)				
$[0, 100]_{\mathrm{INEL>0}}$	6.93 ± 0.09	MB	0.92 ± 0.003	1.724
$[50, 100]_{\mathrm{INEL>0}}$	3.32 ± 0.04	MB	0.862 ± 0.016	0.871
$[30, 100]_{\mathrm{INEL>0}}$	4.41 ± 0.05	MB	0.897 ± 0.013	1.214
$[0.1, 30]_{\mathrm{INEL>0}}$	13.81 ± 0.14	MB	0.997 ± 0.001	0.509
$[0, 0.1]_{\mathrm{INEL>0}}$	31.53 ± 0.38	HMV0M	1.0 ± 0.0	0.483

Table IV.2.2: Summary table of the $N_{tracklets}$ and V0M percentile multiplicity intervals with the corresponding mean charged particle densities in $|\eta| < 0.5$, trigger efficiencies and the number of events for normalisation.

IV.3 Monte Carlo simulations

Data samples of simulated proton-proton collisions are used at different steps of the real data analysis. The Monte Carlo simulations are done proportionally to the amount of recorded data, reproducing the position and width of the interaction vertex distribution and the status (active channels, gain, noise level, and alignment) of all detectors in different runs of data taking within a

given period. The proton-proton collisions at $\sqrt{s} = 13$ TeV are generated with PYTHIA 8.2 with the Monash 13 tune [136] requiring at least a $c\bar{c}$ or a $b\bar{b}$ pair produced at mid-rapidity ($|y| < 1.5$) in the event. The produced D_s^+ meson are forced to decay into the channel of interest in order to increase the statistics of generated events. The propagation of the generated particles through the ALICE detector is performed using the GEANT3 transport package [229].

For detector acceptance and efficiency correction, around 480 millions of generated minimum-bias proton-proton collisions at $\sqrt{s} = 13$ TeV are associated to the minimum-bias data sample. As the high multiplicity triggers are not implemented in simulations, a multiplicity threshold was set requiring at least 65 charged particles produced in $|\eta| <1$ in the events for the 150 millions and 170 millions associated to the HMSPD and HMV0M triggered data respectively. Additionally, Monte Carlo samples are produced for machine learning studies of the D_s^+ meson selection.

The quality assurance analysis on the simulations are done by comparing global event observable distributions (such as the track multiplicity of the event, the position of the primary vertex...) with their counter parts measured in data.

IV.4 D_s^+ meson reconstruction and selection

The D_s^+ meson and its charge conjugate are studied in the $D_s^+ \to \phi(1020)(\to K^+K^-)\pi^+$ decay channel. Its main properties, together with those of most abundant beauty hadrons decaying into the D_s^+ meson, are summarised in table IV.4.1. The proper lifetime of the D_s^+ and beauty hadrons are quite low with respect to the distance of the first sensitive layers of the ALICE detector from the collision vertex position (primary vertex). In the D_s^+ transverse momentum range presented in this analysis, the probability of survival of the D_s^+ mesons beyond the first sensitive layers is negligible. Therefore, the D_s^+ mesons are reconstructed from their decay products.

hadrons	**mass** (MeV/c^2)	$c\tau$ (μm)	**decay channel**	**branching ratio** (%)
D_s^+	1968.34 ± 0.07	151.2	$\phi(1020)(\to K^+K^-)\pi^+$	2.27 ± 0.08
B^+	5279.32 ± 0.14	491.1	D_s^+ X	$7.9^{+1.4}_{-1.3}$
			D_s^- X	$1.10^{+0.40}_{-0.32}$
B^0	5279.63 ± 0.15	455.7	D_s^+ X	$10.3^{+2.1}_{-1.8}$
			D_s^- X	< 2.6
B_s^0	5366.89 ± 0.19	452.4	D_s^-	93 ± 25
Λ_b^0	5619.60 ± 0.17	440.7	Λ_c^+ D_s^-	1.10 ± 0.10

Table IV.4.1: The mass and proper lifetime of the D_s^+ and main beauty hadrons decaying into the D_s^+ meson with their respective branching ratios [237].

A D_s^+ candidate is defined as a triplet of charged particles with the charge-sign configuration (+, −, +) and (−, +, −) for the D_s^-, coming from a common decay vertex. The reconstructed

invariant mass formula of the candidate is computed as follows:

$$\begin{aligned} M_{\text{inv}}^2 =& \Big(\sum_{i=0}^{2} E_i \Big)^2 - \Big\| \sum_{i=0}^{2} \vec{p}_i \Big\|^2 \\ =& \Big(\sqrt{m_0^2 + p_0^2} + \sqrt{m_1^2 + p_1^2} + \sqrt{m_2^2 + p_2^2} \Big)^2 - \Big\| \vec{p}_0 + \vec{p}_1 + \vec{p}_2 \Big\|^2 \end{aligned} \tag{IV.5}$$

with E_i the energy, m_i the mass, and p_i the momentum of the daughter particle i.
In the decay channel studied, the daughter track with opposite charge with respect to the D_s^+ candidate charge must be a kaon. When computing the invariant mass of a $D_s^\pm$ candidate, the kaon mass is assigned to the opposite charge daughter track and the remaining two mass hypotheses are considered one after other, e.g., for a D_s^+: $i = \{\pi^\pm, K^\mp, K^\pm\}$ and $i = \{K^\pm, K^\mp, \pi^\pm\}$. A significant fraction of the candidates reconstructed are pure combinatorial background. Several selection criteria are applied to the daughter tracks, the candidate topology and to the daughter particle identification in order to enhance the signal to background ratio which is critical for the D_s^+ meson raw yield extraction.

IV.4.1 Decay track selection

The daughter particles coming from the D_s^+ meson decay are reconstructed in the ALICE detector as charged-particle tracks. For the analysis, global tracks (reconstructed in both TPC and ITS) are used in order to get the optimal track parameters maximising the precision on the decay vertex position. It is ensured by performing a final fit of the track with the Kalman filter based tracking algorithm using the full particle identification information and optimal track parameters determined at the collision vertex, requiring a reduced chi-square χ^2/ndf below 2 of the momentum fit in the TPC (with ndf, the number of degree of freedom of the fitting procedure). Additionally, the tracks must have at least one cluster in any of the two SPD layers of the ITS, ensuring a reasonable resolution on the track impact parameter and the position of the collision and decay vertices while keeping a high reconstruction efficiency. On the TPC side, the track must be associated with at least 70 crossed pad rows out of the maximal number of 159 and having a ratio of the number of crossed pad rows over the number of findable clusters associated given the track geometry larger than 0.8. Global tracks with $p_{\text{T}} < 1\text{GeV}/c$, satisfying the quality criteria, have a transverse momentum resolution below 1% and increase to reach 2% for $p_{\text{T}} \sim 10\text{GeV}/c$ (see figure III.3.2), and global tracks with $p_{\text{T}} > 0.3$ GeV/c have an impact parameter resolution in the transverse plane below 130 μm (see figure III.3.1).

The D_s^+ meson candidates are built with tracks selected within the kinematic range $|\eta| < 0.8$ and $p_{\text{T}} > 0.3$ GeV/c. The rapidity acceptance of D_s^+ meson candidates with $p_{\text{T}} < 2$ GeV/c start to decrease as shown on figure IV.4.1. To compensate the drop of acceptance and potential border effects, a fiducial acceptance y_{fid} is introduced, defined as a 2nd order polynomial function:

$$y_{fid}(p_{\text{T}}) = 0.5 + \frac{1.9}{15}\, p_{\text{T}} - \frac{0.2}{15}\, p_{\text{T}}^2 \tag{IV.6}$$

$y_{fid}(p_{\text{T}})$ is drawn in red on the figure. The D_s^+ meson candidates outside the fiducial region are rejected.

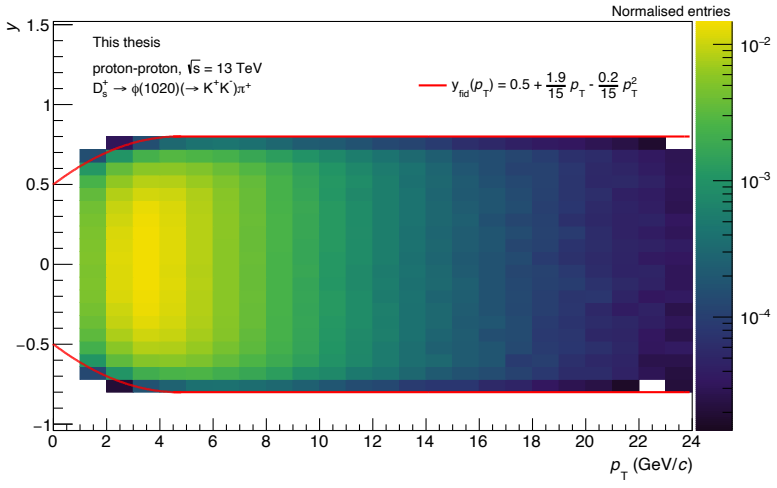


Figure IV.4.1: Reconstructed rapidity distribution of D_s^+ in proton-proton collision at $\sqrt{s} = 13$ TeV. The fiducial acceptance region is defined as $|y| < y_{\rm fid}$ delimited in red.

The contribution of D_s^+ daughter tracks to the $N_{tracklets}$ multiplicity estimation

The inclusion of the tracks from the decay daughter of the D_s^+ meson candidates affect the multiplicity estimation of the event and the resolution on the collision vertex reconstruction. The resolution of the collision vertex further affect the resolution on the D_s^+ meson decay topological variables, especially at low charged-particle multiplicity. For this analysis, the decay daughter tracks are removed before determining the collision vertex position in the data. The D_s^+ meson daughter tracks having a hit in each of the two layers of the SPD, *i.e.* attached to a tracklet, are subtracted from the corrected $N_{tracklets}$ of the event.

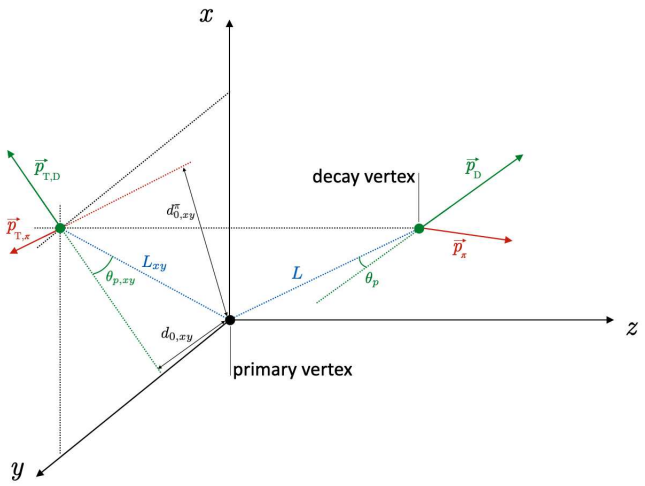


Figure IV.4.2: Schematic representation of the D_s^+ meson decay with the main topological variables used to perform the candidate selection.

IV.4.2 Topological and kinematical selections

The main topological variables related to the $D_s^+ \to \phi(1020)(\to K^+K^-)\pi^+$ decay channel are presented on figure IV.4.2. The centre of the coordinates system is located at the collision position, the primary vertex. The longitudinal z axis is parallel to the proton beams. The x and y axes, orthogonal to the z axis, define the transverse plane (x, y). The topological variables described below can be defined in both the 3D space and in the transverse plane. Their distributions are studied, using Monte Carlo simulations, for both true D_s^+ meson decays and background D_s^+ candidates. In the following parts of the analysis, reconstructed D_s^+ mesons produced at the collision vertex are denoted as prompt D_s^+ candidates and reconstructed D_s^+ mesons produced by the decay of a B hadron (see table IV.4.1) are called feed-down candidates.

Decay length

The decay length L is defined as the distance between the reconstructed decay vertex and the primary vertex. The distributions for the D_s^+ prompt, feed-down and background candidates having a p_T (D_s^+) in the 4 to 6 GeV/c interval are shown on the left panel of Figure IV.4.3. The D_s^+ prompt decay vertices are expected to be displaced by a few hundred μm from the primary vertex, they have larger values of decay length compared to background candidates which for the most part are expected to be built from primary tracks. The background candidate distribution is not centred to zero due to the limited resolution on the impact parameter. The D_s^+ feed-down candidates decay vertices are further displaced due to the additional B hadron proper lifetime. The distribution in two different p_T intervals are shown in the middle panel of the figure. Due to the Lorentz boost, the high p_T distribution is shifted towards higher decay length making the topological variable more efficient. The right panel of Figure IV.4.3 shows the decay length in the transverse plane L_{xy} normalised by its uncertainty $\sigma(L_{xy})$ for D_s^+ candidates in the 4 to 6 GeV/c p_T interval. Since the detector resolution is better in the transverse plane than in the 3D space, L_{xy} is used to perform a refined background candidate rejection.

Impact parameter

The impact parameter d_0 of a D_s^+ meson candidate is defined as the distance of closest approach between the reconstructed flight line of the candidate and the primary vertex. For prompt D_s^+ meson candidates, the impact parameter is expected to be null. But due to the limited resolution on the primary vertex position and the reconstructed tracks, the distribution is spread around zero as shown on the left panel of figure IV.4.4 for the impact parameter in the transverse plane $d_{0,xy}$. The finer resolution obtained at higher candidate momentum is reflected in a narrower distribution as seen on the middle panel of figure IV.4.4. The expected impact parameter of a daughter track i can be estimated from L_{xy} and the angle θ^i_{xy} between the D_s^+ candidate momentum and the one of the daughter track in the transverse plane.

$$d^i_{0,xy}(\text{exp.}) = L_{xy} \cdot \sin\theta^i_{xy} \tag{IV.7}$$

The maximum normalised difference between the measured and the expected daughter track impact parameters in the transverse plane, $\Delta d_{0,xy}$, is defined as follow,

$$\Delta d_{0,xy} = \max_{i=0,1,2} \left\{ \frac{d^i_{0,xy}(\text{meas.}) - d^i_{0,xy}(\text{exp.})}{\sqrt{\sigma^2_{d^i_{0,xy}}(\text{meas.}) + \sigma^2_{L_{xy}} \cdot \sin^2\theta^i_{xy}}} \right\} \tag{IV.8}$$

with $d^i_{0,xy}$ the impact parameter of the track i in the transverse plane with its associated uncertainty $\sigma_{d^i_{0,xy}}$. The corresponding distribution is shown on the right panel of the figure IV.4.4.

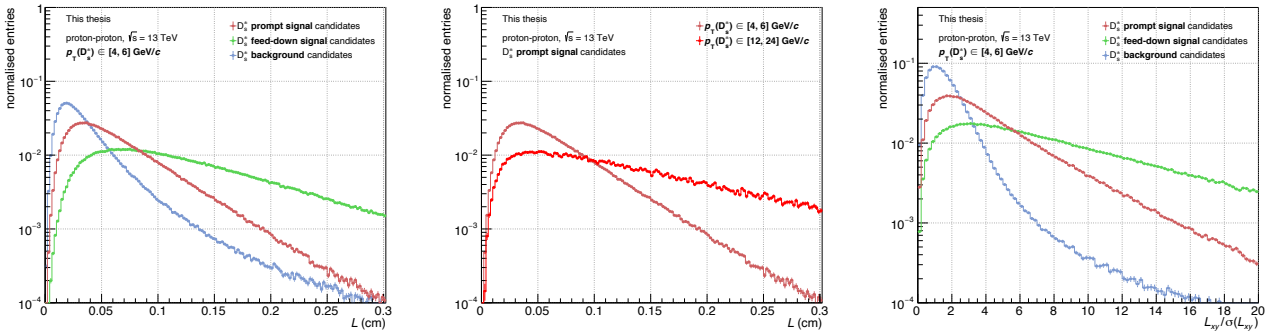


Figure IV.4.3: Left panel: distribution of the decay length for D_s^+ prompt, feed-down and background candidates in the p_T (D_s^+) $\in$ [4, 6] GeV/c interval. Middle panel: distribution of the decay length for D_s^+ prompt candidates in the p_T (D_s^+) $\in$ [4, 6] and [12, 24] GeV/c intervals. Right panel: distribution of the normalised decay length in the transverse plane for D_s^+ prompt, feed-down and background candidates in the p_T (D_s^+) $\in$ [4, 6] GeV/c interval.

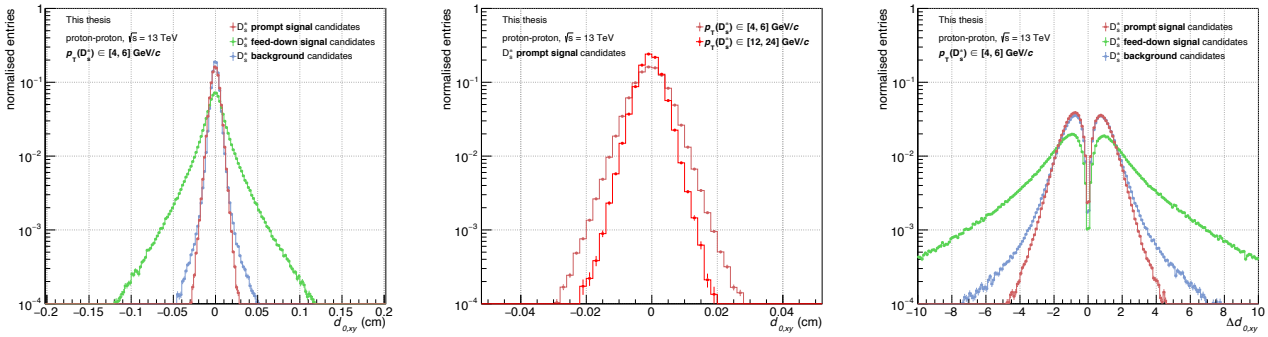


Figure IV.4.4: Left panel: distribution of the impact parameter in the transverse plane for prompt, feed-down and background D_s^+ meson candidates in the p_T (D_s^+) $\in$ [4, 6] GeV/c interval. Middle panel: distribution of the impact parameter in the transverse plane for prompt D_s^+ meson candidates in the p_T (D_s^+) $\in$ [4, 6] and [12, 24] GeV/c intervals. Right panel: distribution of the $\Delta d_{0,xy}$ variable (see text for details) for prompt, feed-down and background D_s^+ meson candidates in the p_T (D_s^+) $\in$ [4, 6] GeV/c interval.

Daughter track dispersion around the decay vertex

The dispersion of the daughter tracks around the decay vertex is quantified by σ_{vtx},

$$\sigma_{\text{vtx}} = \sqrt{\sum_{i=0}^{2} d_i^2}$$

with d_i the distance of closest approach between the daughter track i and the decay vertex. Non-zero σ_{vtx} values are expected for the true D_s^+ meson candidates due to the limited tracking and vertexing resolutions. As shown on the left panel of figure IV.4.6, the true D_s^+ distribution maxima are closer to zero and decrease faster compared to the background candidate distribution.

Pointing angle

The pointing angle θ_{p} is defined as the angle between the direction of the reconstructed D_s^+ meson momentum and the line connecting the primary and decay vertices. It is expected to be close to zero for prompt candidates due to the finite resolution on the momentum and primary vertex position. The distribution of the cosine of the variable in the transverse plan $\theta_{\text{p, xy}}$ is shown on the left panel of figure IV.4.5 and is peaked at unity for all the D_s^+ candidates. However, the prompt D_s^+ candidate distribution decrease faster with the cosine of the pointing angle.

Kaon emission angle

The angle $\theta_{\text{K},\phi}$ is defined as the angle between the propagation direction of the $\phi(1020)$ meson, estimated with the one of the pion, and the propagation direction of the kaon in the ϕmeson decay frame. The emission of the kaon is isotropic in the $\phi(1020)$ meson rest-frame, therefore the distribution of $|\cos^3\theta_{\text{K},\phi}|$ is expected to be flat for true D_s^+ candidates as shown on the right panel of figure IV.4.5.

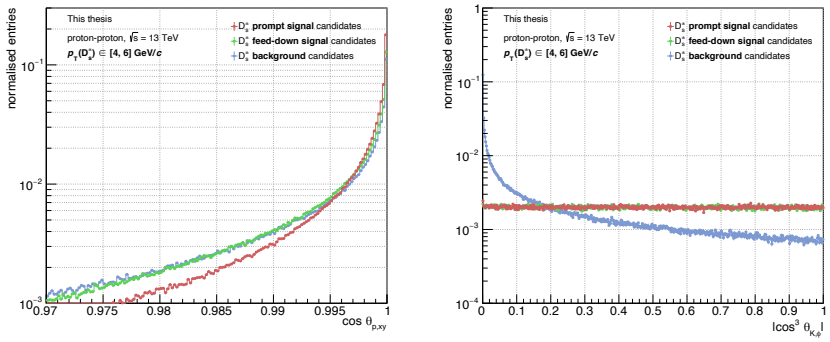


Figure IV.4.5: Left panel: distribution of the cosine of the pointing angle in the transverse plane for prompt, feed-down and background D_s^+ meson candidates in the p_{T} $(D_s^+) \in [4, 6]$ GeV/c interval. Right panel: distribution of $|\cos^3\phi_{\pi,K}|$ for prompt, feed-down and background D_s^+ meson candidates in the p_{T} $(D_s^+) \in [4, 6]$ GeV/c interval.

Kaon pair invariant mass

In the studied decay channel, the kaon pair must originate from a decay of a $\phi(1020)$ meson. The absolute difference between the reconstructed invariant mass of the kaon pair and the $\phi(1020)$ meson mass is therefore exploited to reject background candidates. Since the reconstructed D_s^+ meson candidates are kept in their two mass hypotheses configuration, the candidates with a kaon pair associated to the wrong mass hypothesis will be discarded. The distribution of the mass difference is shown on the right panel of figure IV.4.6 where the true D_s^+ candidate distributions are peaked around zero while the background candidate distribution is more uniform.

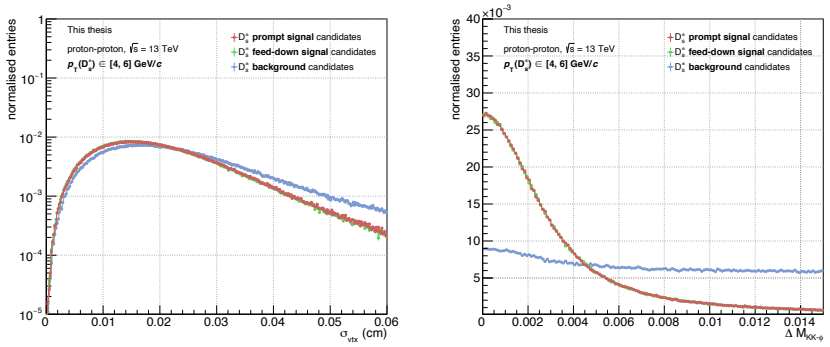


Figure IV.4.6: Left panel: distribution of the decay track dispersion around the secondary vertex for prompt, feed-down and background D_s^+ meson candidates in the p_T $(D_s^+) \in [4, 6]$ GeV/c interval. Right panel: distribution of the difference between the reconstructed invariant mass of the kaon pair and the $\phi(1020)$ meson mass for prompt, feed-down and background D_s^+ meson candidates in the p_T $(D_s^+) \in [4, 6]$ GeV/c interval.

IV.4.3 Particle identification

The particle identification (PID) information on the D_s^+ candidate decay tracks are provided by the measurement of energy losses in the TPC detector and the time of flight measured with the TOF detector. The selection variable, applied per a decay track, is based on the difference between the measured S_{meas} and expected $S_{\text{exp}}^{\pi,\text{K,p}}$ signal for a given particle specie hypothesis. The n_σ variable is defined as,

$$\text{n}\sigma^{\pi,\text{K,p}} = \left| \frac{S_{\text{meas}} - S_{\text{exp}}^{\pi,\text{K,p}}}{\sigma_{\pi,\text{K,p}}} \right| \tag{IV.9}$$

with $\sigma_{\pi,\text{K,p}}$ the detector resolution for a particle specie. For the decay channel studied, the pion and kaon hypothesis are tested and the n_σ variables from the TOF and TPC are combined to account for tracks that do not have hits in the TOF detector. Depending on the availability of the PID information for a given track in the two detectors,

$$\text{n}\sigma_{\text{comb}}^{\pi,\text{K}} = \begin{cases} |\text{n}\sigma_{\text{TPC}}^{\pi,\text{K}}| & \text{TPC information only} \\ |\text{n}\sigma_{\text{TOF}}^{\pi,\text{K}}| & \text{TOF information only} \\ \frac{1}{\sqrt{2}}\sqrt{(n\sigma_{\text{TPC}}^{\pi,\text{K}})^2 + (n\sigma_{\text{TOF}}^{\pi,\text{K}})^2} & \text{TPC and TOF informations} \end{cases} \tag{IV.10}$$

A total of 3 (daughter tracks) $\cdot$ 2 (particle specie hypothesis) = 6 particle identification related variables are used to discriminate signal D_s^+ candidate from the combinatorial background. The $\text{n}\sigma_{\text{text}}$ distribution for the pion and the kaon hypotheses of the opposite charge sign daughter track (with respect to the candidate charge) are shown on figure IV.4.7. In the D_s^+ decay channel studied, the opposite charge sign track of the candidate triplet must be a kaon. In the candidate reconstruction procedure, the opposite charge sign track happened to be stored in second position in the triplet. Therefore the $\text{n}\sigma_{\text{comb}}^{\pi}$ distribution of the second track of true D_s^+ candidates is shifted toward higher value, *i.e.* less probable to be a pion, with respect to the background D_s^+ candidate distributions while for the $\text{n}\sigma_{\text{comb}}^{\text{K}}$ distribution it is the opposite.

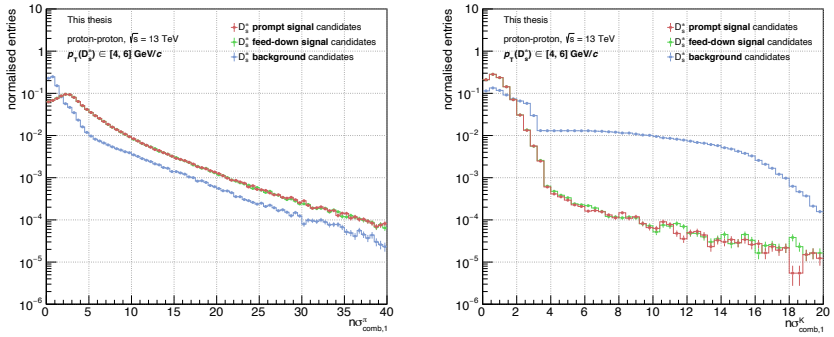


Figure IV.4.7: Distribution of the $n\sigma^{\pi}_{comb}$ (left panel) and $n\sigma^{K}_{comb}$ (right panel) of the second daughter track for prompt, feed-down and background D_s^+ meson candidates in the p_T (D_s^+) $\in$ [4, 6] GeV/c interval.

IV.4.4 Linear correlations among selection variables

The Pearson correlation coefficient is used to compute the linear correlation between the different topological, kinematical and particle identification D_s^+ candidate selection variables. The linear correlation coefficient measures the strength of the relationship between two variables. The absolute value of the coefficient increases as the two-dimensional distribution, *i.e.* the correlation distribution, of two variables approaches to a linear function. The Pearson or product-moment correlation coefficient between the X and X variable distributions is defined as,

$$\rho = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum_i^n (x_i - \langle X \rangle)(y_i - \langle Y \rangle)}{\sqrt{\sum_i^n (x_i - \langle X \rangle)^2 \sum_i^n (y_i - \langle Y \rangle)^2}} \tag{IV.11}$$

with cov(X, Y) the covariance, σ_X, σ_Y and $\langle X \rangle$, $\langle Y \rangle$ the standard deviations and the means of the X and Y distributions respectively. The Pearson coefficient is a real number between $-1 \leqslant \rho \leqslant 1$. Positive ρ values indicate a linear correlation while negative value a linear anti-correlation. If $\rho = 0$ there is no linear correlation between the two variables. The results obtained for the D_s^+ candidate selection variables are presented in the form of correlation matrices on the figures IV.4.8 and IV.4.9 for prompt and background D_s^+ candidates in two different p_T intervals. As expected, the same topological variables defined in the transverse plan and in 3D, such as the (normalised) decay length L and the cosine of the pointing angle θ_p, are strongly to fully correlated since they carry the same physical information. The same observation is true for the combined PID informations in the kaon and pion hypotheses for a given daughter track.

In addition to taking into account the differences in the distributions of selection variables for prompt D_s^+ and background candidates, the candidate selection model can exploit the differences between the correlations. By comparing the panels of the figure IV.4.8, stronger linear correlations are present between the σ_{vtx}, the decay length and the pointing angle for the background candidates with respect to the signal candidates. Stronger linear correlation between the decay length and the pointing angle are present for the prompt D_s^+ signal candidates. A candidate selection model based on machine learning algorithm (see section IV.5) allows us to exploit correlations of more complex than linear shapes.

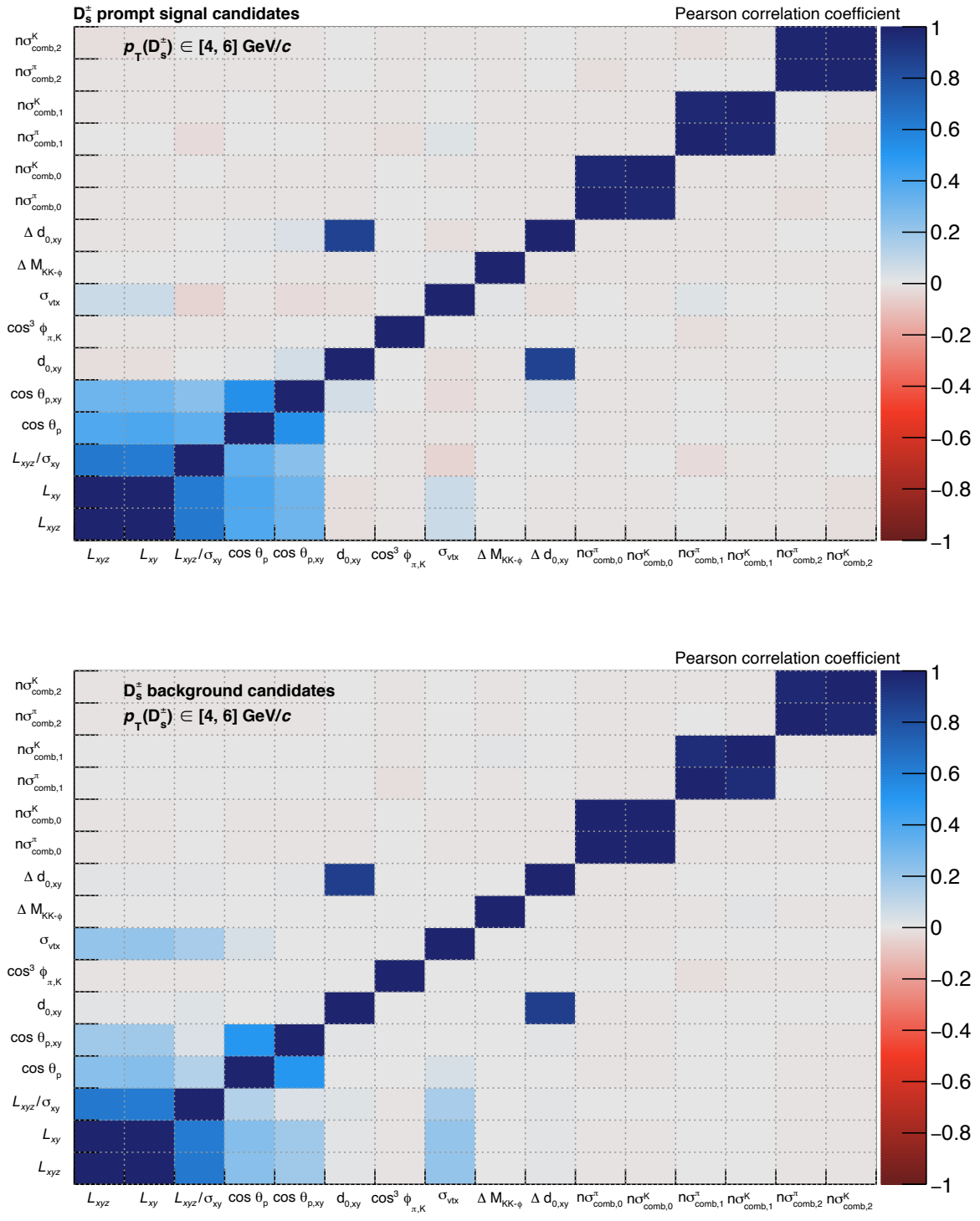


Figure IV.4.8: Linear correlation matrices for prompt (top) and background (bottom) D_s^+ candidates in the 4 to 6 GeV/c p_T interval.

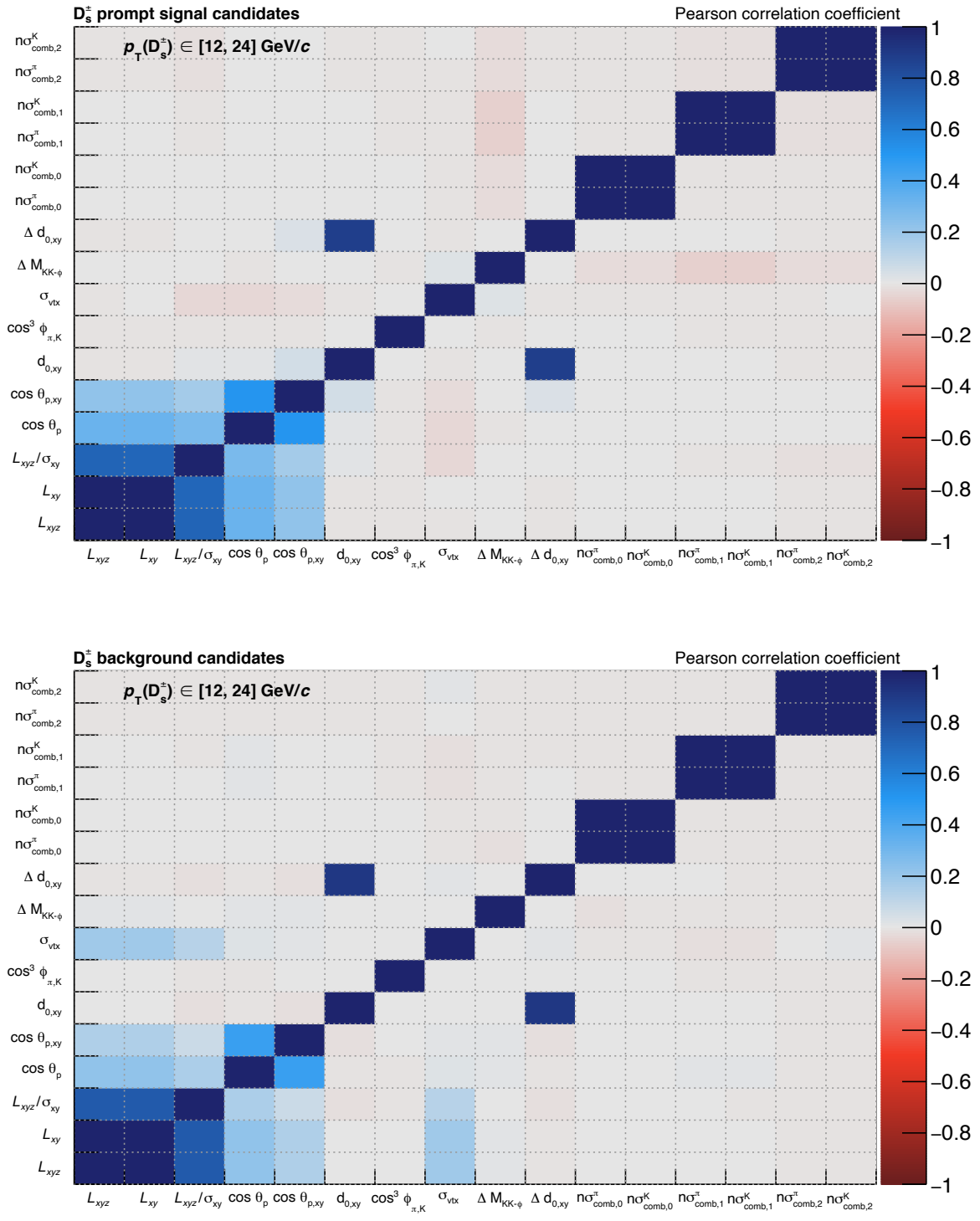


Figure IV.4.9: Linear correlation matrices for prompt (top) and background (bottom) D_s^+ candidates in the 12 to 24 GeV/c p_T interval.

IV.5 Candidate selection using boosted decision trees

The goal of the D_s^+ candidate selection is to perform a binary classification of the candidates reconstructed in the data sample, selecting prompt D_s^+ candidate and rejecting background D_s^+ candidates by exploiting the D_s^+ meson decay topology and the particle identification variables. In the "standard" approach, a single decision value with an associated comparison operator is determined for each topological and kinematical variables from their distributions in Monte Carlo simulations. For example, if we want to select D_s^+ candidates with $\cos\theta_p > 0.9$, ">" is the comparison operator and 0.9 is the decision value. For the particle identification, the selection strategy is slightly more refined and depends on the availability of PID information in the TPC and the TOF detector. As the distribution of the variables evolve with the transverse momentum of the D_s^+ meson, different decision values are attributed in different p_T intervals. By applying sequentially the predefined decision values to the candidate data sample, the related selection variable phase space is sliced to select regions with high signal-to-background ratio. A retained D_s^+ candidate must satisfy every selection criteria. This approach, based essentially on a single-variable decisions does not take into account the correlations among selection variables.

The candidate selection based on a machine learning approach aims to perform more complex and refined selection decisions by training a model to learn a mapping function from the set of selection variables to a probabilistic prediction on the nature of the candidate (signal or background). By refining the selection of signal candidates in the selection variable phase space, a better signal selection efficiency is expected allowing for performing the data analysis in extended kinematic regions, especially at low D_s^+ transverse momentum, where the "standard" selection procedure reaches its limits. The gain in selection performance obtained with the machine learning approach will be discussed in the next paragraphs.

A definition of a machine learning algorithm is proposed by Tom Mitchell [238]: *"A computer program is said to learn from experience E with respect to some class of tasks in T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E".*

IV.5.1 Decision tree and the gradient boosting method

The machine learning algorithm chosen to perform the D_s^+ candidate selection is the boosted decision tree. The decision tree is a supervised learning method to perform regression and classification tasks. The model is constructed by an algorithm identifying the decision rules, taking the form of if-then-else operations, from the features of data to predict the value of a target variable. As its name suggests, the graphical representation of the model is a tree graph structure where nodes represent the decision making rules. The edges of the node represent the node decision outcomes which splits the data recursively in subsets as data is propagated from the root to the final nodes. The final nodes, called leaves, represent the target variable output. Unlike more complex machine learning algorithms such as neural networks, the decision trees allows for a much easier interpretation, because they can be visualised.

The gradient boosting method consists in combining several trees ("weak learner") built sequentially to constitute a tree ensemble model that becomes a "strong learner". For a given model composed of K trees, the predicted output for a given observation is expressed as,

$$\hat{y}_i = \sum_{k=1}^{K} f_k(\mathbf{x_i}) \tag{IV.12}$$

with f_k an independent tree and $\mathbf{x_i}$ the feature vector of the i-th observation. In our case, the feature vector is composed by the D_s^+ candidate selection variables and an observation is a D_s^+ candidate. Each tree f_k has its own structure and a leaf weight vector ω containing the leaf scores. The prediction of a given tree on a given observation, $f_k(\mathbf{x_i})$, is computed by summing the corresponding leaf scores [239]. The final model prediction on a D_s^+ candidate is a floating number between 0 and 1, values close to 1 (0) are associated to a high probability to be a signal (background) candidate.
The boosting procedure is performed iteratively by minimisation of an objective function. The objective function is the sum of the loss function measuring the distance between the truth and the model prediction for each observation of the data sample. The regularisation function is introduced to penalise the structure complexity of the model. Since we are performing a binary classification task, D_s^+ signal versus background classification, the loss function used is the the binary logistic loss,

$$L(y_i, \hat{y}_i) = -\frac{1}{n}\sum_{i}^{n}[y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \text{ with } p_i = \frac{1}{1 + e^{-\hat{y}_i}} \tag{IV.13}$$

with y_i the true label taking the value 1 for prompt D_s^+ signal candidates and 0 for background and p_i the sigmoid of the model prediction for the i-th observation in the data sample of size n. The general expression of the regularisation function is,

$$\Omega(F) = \sum_{k=1}^{K}\left[\gamma T_k + \frac{1}{2}\lambda||\omega_k||^2\right] \tag{IV.14}$$

with F the model, T_k the number of leaves and ω_k the leaf weight vector of the k-th tree. For the model configuration chosen for this analysis, the gain threshold, γ, is set to zero and the regularisation coefficient λ is set to unity.

Starting from an initial prediction, the following steps are repeated until the maximum number of trees in the model is reached, which is an hyper-parameter of the model. Hyper-parameters values are specified by the user as opposed to the "learnable parameters", for which the values are determined from the minimisation of the objective function.

- step 1: Compute the residuals of the model for each observation in the data sample. The residual of a model for a given observation is the difference between the model prediction and the true label value.
- step 2: A decision tree is build from the residuals data sample by constructing the decision rules maximising the information gain at each node split until a stop condition is met. The information gain is based on the decrease of the entropy after the data sample is split with respect to a given variable. The maximum depth of the tree limiting its growth is an hyper-parameter of the model and it is usually used as a stop condition.
- step 3: Update the model by adding the built decision tree to the tree ensemble.

By doing so, the decision tree built on the next iteration will try to correct the prediction mistakes from the current iteration decision tree. For the D_s^+ candidate selection task, the boosted decision tree (BDT) model of the XGBoost library [239] is used.

IV.5.2 Model training and validation

The BDT is a supervised learning method so the true labels of the D_s^+ candidates are provided for the model training phase. The prompt D_s^+ signal candidates are taken from dedicated Monte Carlo simulation data sample and background candidates from the sidebands of the D_s^+ candidate invariant-mass distribution in real data (see section IV.7). The simulated and real data correspond to minimum-bias proton-proton collisions at $\sqrt{s} = 13$ TeV, no selection on the event charged particle multiplicity is performed. The model training is performed in six D_s^+ meson p_T intervals, [1, 2], [2, 4], [4, 6], [6, 8], [8, 12], and [12, 24] GeV/c.

variables	selection criteria
Δ M$_{inv}$(D_s^+)	< 0.3 GeV/c^2
daughter track p_T	> 0.3 GeV/c
σ_{vtx}	< 0.06 cm
$\cos\theta_p$	> 0.85
$\cos\theta_{p,xy}$	> 0.85
Δ M$_{KK\text{-}\phi}$	< 15 MeV/c^2

Table IV.5.1: Pre-selection decision values for the topological and kinematical D_s^+ candidates selection variables. The same criteria are applied in all p_T intervals considered.

The candidates are filtered with the daughter track selection criteria described in section IV.4.1 and loose pre-selection cuts on the topological and kinematical D_s^+ candidates selection variables summarised in table IV.5.1. For the particle identification selection, a D_s^+ candidate is selected if all its daughter tracks have a 3σ compatibility between the measured signal in the TPC or in the TOF detectors and the desired mass hypothesis (n$\sigma_{TPC}^{\pi,K}$, n$\sigma_{TPC}^{\pi,K}$). The opposite charge sign track with respect to the candidate charge has to be compatible with the kaon hypothesis and the like sign daughter tracks with a pion and a kaon or vice-versa. This selection on the track particle identification is called "conservative PID strategy".

The D_s^+ candidate data samples are balanced, composed by an equal fraction of signal and background candidates in each p_T (D_s^+) intervals considered. The fraction of background candidate used for the BDT model training with respect to the total amount of real data analysed is well below 1‰ in low p_T intervals and around 1% in the highest p_T interval. These fraction are negligible which is important to avoid any bias when performing the D_s^+ candidate selection on the whole real data sample. The data samples are split in training data samples (80%) and a test data sample (20%) in order to cross-check the D_s^+ selection performances reached during the training and with an unseen test data sample. The number of candidates per data sample used to train and test the different models as the BDT hyper-parameters are summarised in table IV.5.2. Only a fraction of the available BDT hyper-parameters are discussed, their names are kept as the ones of the XGBoost library documentation,

- **min_child_weight** ($[0, \infty[$): the minimum sum of observation weights contained in a node in order to be further split. In the binary logistic loss, the sum of the observation weights is expressed as $\sum_i p_i(1-p_i)$ (see equation IV.13), the sum is null if all the D_s^+ candidates in the node are predicted as signal (p_i close to 1) or background (p_i close to 0). Indeed no further split is needed if the candidate population of the node is pure. If the sum of weights in the node is below the threshold, the node becomes a leaf. By choosing high threshold values, the tree becomes more conservative, helping to reduce the model complexity.

- **max_depth** ([0, ∞[): the maximum depth of a tree, limiting is growth.
- **subsample** ([0, 1]): the fraction of observations randomly selected out of the training sample and used to train each tree. The subsampling is done at each boosting iteration.
- **colsample_bytree** ([0, 1]): fraction of features (the D_s^+ selection variables) randomly selected and used to build each tree. It is is done at each boosting iteration.
- **learning_rate** ([0, 1]): this factor is introduced to weight the corrections of the new tree when it is added to the model to prevent overfitting.
- **lambda** (0 or 1): the regularisation term, see equation IV.14.
- **gamma** ([0, ∞[): the gain threshold specifies the minimum reduction of the loss function required to make a split.
- **n_estimators** ([0, ∞[): the number of trees in the model.

$p_{\mathbf{T}}$ **intervals** (GeV/c)	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
# candidates (x 10^3)	300	800	800	500	400	200
hyper-parameters						
min_child_weight	7.17	5.14	7.3	6.4	7.13	6.42
max_depth	4	4	4	4	4	4
subsample	1	1	1	1	1	1
colsample_bytree	0.51	0.74	0.64	0.43	0.51	0.47
learning_rate	0.085	0.089	0.093	0.95	0.085	0.1
n_estimators	400	400	400	400	400	400

Table IV.5.2: Summary table of the number of candidate used to train the models. The model hyper-parameters are also reported.

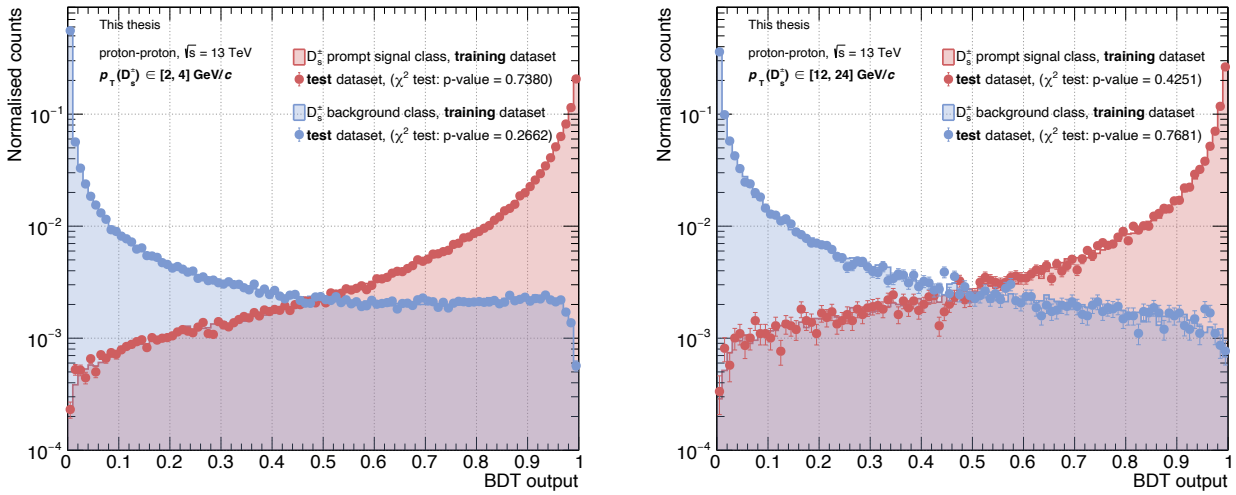


Figure IV.5.1: BDT model output distributions for the signal and background candidates in the training (shaded area) and test (points) data samples in the 2 to 4 (left)and 12 to 24 (right) GeV/c p_{T} intervals.

Since the signal candidate were labelled with the target value 1 and background candidates with 0, the model is expected to provide predictions close to unity for signal-like candidates and predictions close to zero for background-like candidates. The BDT output distributions in two different p_{T} intervals is shown on figure IV.5.1, representing the model predictions for the candidates in the training and test data samples. As expected, we obtain two peaked distributions around their respective label value. In real data, where the two distributions cannot be distinguished, the user define a decision value in the [0, 1] range delimiting the signal-like region from the background-like region. One could choose the decision value corresponding to the intersection of the two distributions, around 0.5, however, due to the very low signal to background ratio in real data, the decision value is set closer to unity to increase the background rejection.

The goal of the training phase is to make the model learn the general features and underlying patterns of the training data and reuse them on unseen data. This capability is called "generalisation". Related to the generalisation, the model training phase is exposed to the variance-bias tradeoff. The bias corresponds to the distance between the model prediction and the true value. A highly biased model is a sign of under-fitting, relevant data features have been missed due to an unreasonably simplistic description of the data. The variance characterise the sensitivity of the model to small fluctuations in the data. A high variance model is likely to overfit, potentially describing random fluctuations and not general features due to an overly complex description of the data. The model generalisation capability can be checked by measuring and comparing its performances in the training and the test data samples.

One way to measure the model global performances is to compute the receiver operating characteristic (ROC) curve. The ROC curve is built with the confusion matrix, for the $\mathrm{D_s^+}$ classification task it is defined in table IV.5.3.

model prediction \ **true label**	signal	background
signal	true positive	false positive
background	false negative	true negative

Table IV.5.3: Confusion matrix of the D_s^+ candidate binary classification task.

- True positive (TP) candidates are candidates predicted and labelled as signal.
- False positive (FP) candidates are candidates predicted as signal but labelled as background.
- True negative (TN) candidates are candidates predicted and labelled as background.
- False negative (FN) candidates are candidates predicted as background but labelled as signal.

The true positive rate, *i.e.* the model efficiency, is defined as the ratio of the number of successfully predicted signal candidate over all candidates labelled as signal. The false positive rate, *i.e.* the background contamination, is defined as the ratio of the number of candidate wrongly predicted as signal over all candidates labelled as background.

$$\text{TP rate} = \frac{\sum\limits_{i} \text{TP}_i}{\sum\limits_{i} \text{TP}_i + \sum\limits_{j} \text{FN}_j} \qquad \text{FP rate} = \frac{\sum\limits_{k} \text{FP}_k}{\sum\limits_{k} \text{FP}_k + \sum\limits_{l} \text{TN}_l} \tag{IV.15}$$

The ROC curve is defined by the set of TP and FP rate points computed for different decision values. For the models trained in the 2 to 4 and 12 to 24 GeV/c p_{T} intervals, they are presented on the figure IV.5.2. The bottom left part of the curve is composed by points obtained with decisions value close to unity where the fraction of true signal among the selected candidates is high but the signal selection efficiency is low, as the decision value decreases, both the background contamination and the model efficiency increases. The ideal classification point is found at the (0, 1) coordinate, as the ROC curve moves away from the linear line y = x, associated to a random classification performances, and get closer to the ideal point, the global performances of the model improves. This is quantified by the area under the curve, noted AUC, reported on the legend of the figure IV.5.2.

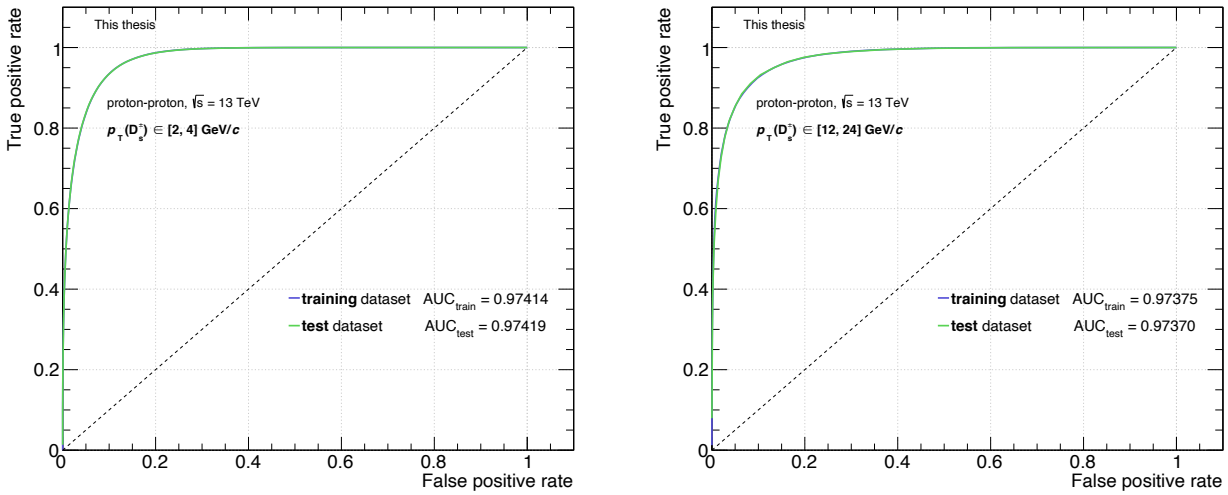


Figure IV.5.2: ROC curves in the training (blue) and test (green) data samples in the 2 to 4 (left) and 12 to 24 (right) GeV/c p_{T} intervals. The area under the curve (AUC) are reported.

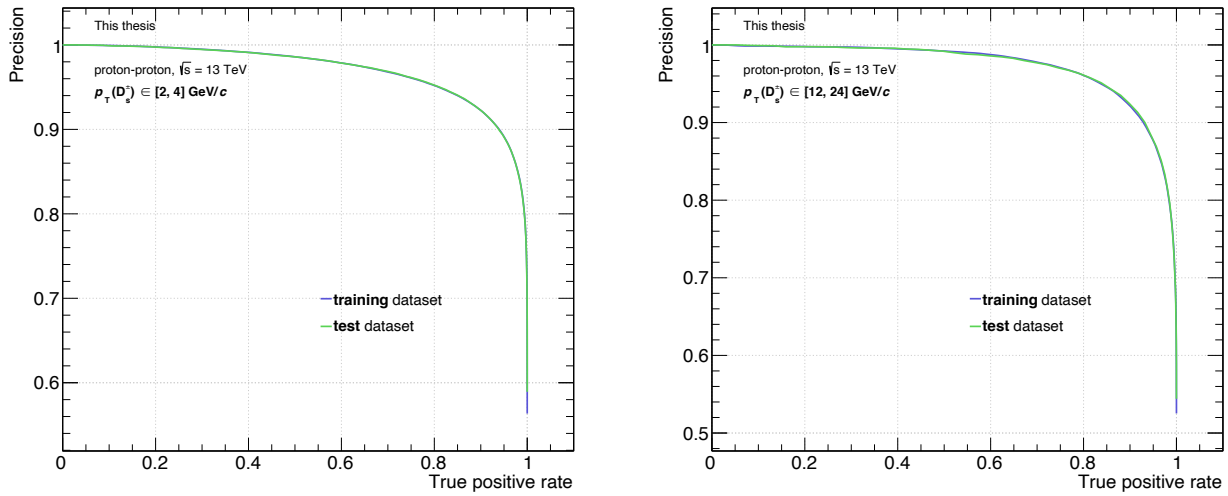


Figure IV.5.3: Precision curves in the training (blue) and test (green) data samples in the 2 to 4 (left) and 12 to 24 (right) GeV/c p_{T} intervals.

The area under the curve value of the ROC curve in the training and test data samples are summarised in table IV.5.4. For different considered models, the ROC AUC is around 0.97, in the first and the last p_{T} interval, where the training data sample statistics is lower. The AUC score in the test data sample is below the one obtained in the training data sample. For the other p_{T} intervals, it is the opposite. However the differences are below 1‰, which is an indication of similar performances obtained in the two data samples.

The performance can also be quantified by comparing the precision curves as on figure IV.5.3. In this case, the curve of the training and test data sample overlap. Here, the precision is defined

as the ratio of the number of successfully predicted signal candidate over all predicted signal candidates.

$$\text{Precision} = \frac{\sum_i \text{TP}_i}{\sum_i \text{TP}_i + \sum_k \text{FP}_k} \tag{IV.16}$$

The higher precision is, the less the predicted signal candidate sample is contaminated by wrongly classified background candidates. The ideal classification point is at (1, 1). As for the ROC curves, the precision curves show similar global performances both being much superior to random classification.

p_{T} **intervals** (GeV/c)	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
ROC AUC$_{\mathbf{train}}$	0.97782	0.97414	0.96963	0.97037	0.97402	0.97375
ROC AUC$_{\mathbf{test}}$	0.97764	0.97419	0.96986	0.97096	0.97429	0.97370
$\frac{\lvert\text{AUC}_{\text{train}}-\text{AUC}_{\text{test}}\rvert}{\text{AUC}_{\text{train}}}$ (%)	0.018	0.005	0.024	0.06	0.028	0.005

Table IV.5.4: Summary of the area under the ROC curves obtained with the different models.

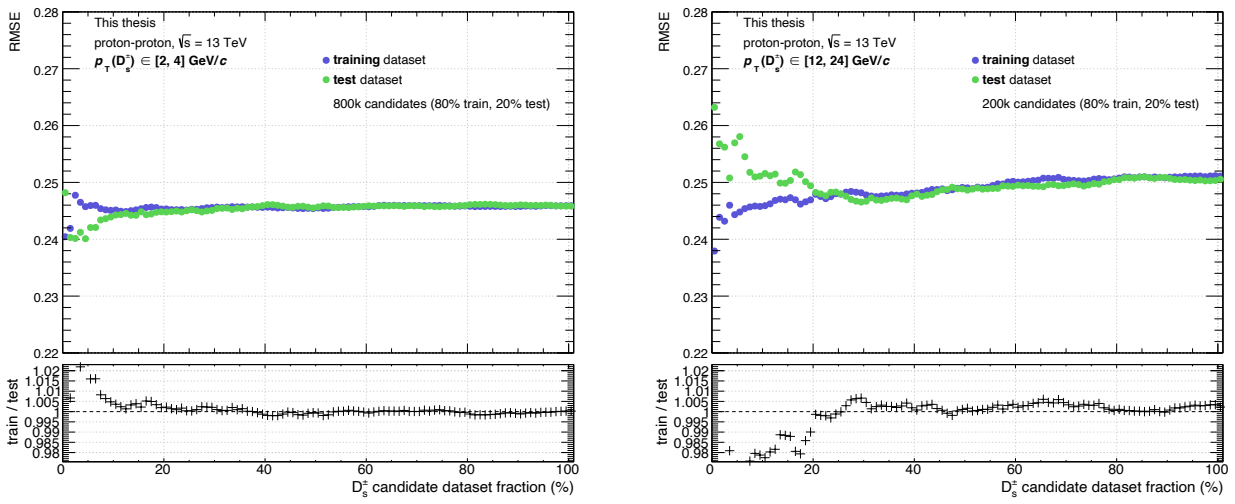


Figure IV.5.4: Evolution of the root mean square error (RMSE) as a function of the fraction of predictions in the training and test data samples. On the bottom panels of the figure, the ratio between the two curves is shown.

On figure IV.5.4 the evolution of the root mean square error (RMSE), *i.e.* the standard deviation of the prediction errors, as a function of the fraction of predictions in the training and test data samples.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2} \tag{IV.17}$$

with y_i the true label and $\hat{y}_i$ the model prediction of the i-th candidate. As the number of predictions increases, the RMSE converge. In the low p_{T} interval, where 800k candidates were used, the ratio of the curve obtained in the training and test data samples, shown in the bottom panel,

the two RMSE curves converge when more than 40% of both data samples is used, reaching a RMSE = 0.246. For the high p_{T} interval, a reduced number of candidates were used (200k) due to the constraint on the fraction of background candidate to be taken from real data. The RMSE of the two curves reach a plateau for a candidate fraction of candidates above 80%. The number of candidate used to train the model is sufficient to obtain stabilised performances.

Even if the corresponding risk is reduced, having similar model performances in the training and test data samples does not guarantee that no bias is introduced during the training phase. As an additional cross-check, a comparison of the production yield results obtained with the standard $\mathrm{D_s^+}$ candidate selection method and the machine learning method is done and discussed in section IV.10.

Model hyper-parameters tuning

A tuning of the model hyper-parameters was performed using a Bayesian optimisation approach. The Bayesian optimisation procedure is used to maximise the AUC value over the hyper-parameters phase space. Starting from an initial point, a probabilistic mapping function is built and exploited to make a decision about where in the hyper-parameter space the next evaluation of the function should be done. Iteratively, the phase space is explored taking into account the previous observations. An overview of the Bayesian optimisation approach is given in [240]. The range of hyper-parameters explored are presented in table IV.5.5.

hyper-parameters	ranges
min_child_weight	[1, 10]
max_depth	[2, 5]
colsample_bytree	[0.2, 0.8]
learning_rate	[0.001, 0.1]
n_estimators	[200, 1500]

Table IV.5.5: Ranges of the model hyper-parameters explored with the Bayesian optimisation procedure.

The ROC AUC score of the model is evaluated on a validation data sample following a 5-fold cross-validation procedure. Tuning the hyper-parameters of a model, cannot be done on the test data sample as it is supposed to be used for the evaluation of the final performances. The risk is to bias the estimation of the generalisation capability of the model. The so called k-fold cross-validation procedure is introduced to address the problem. For example for k=5, the training dataset is randomly shuffled and split in five sub-samples. One sub-sample is taken as validation data sample while the four others are used for the training, the procedure is repeated successively until all sub-samples have been used as validation data sample. The average performance of the five model trained is used to compute the AUC score for a given hyper-parameter configuration. The Bayesian optimisation of the model hyper-parameters is preferred of the "brute force" grid search approach that is computationally expensive. The hyper-parameter sets of the final models in the different p_{T} intervals do not vary much (see table IV.5.2), since the same classification task is carried out using the same set of selection variables.

Selection variable impact on the model decision

To get insights on the model performance for the D_s^+ candidate selection, the selection variable relative importances are estimated using the so called SHAP value. The SHapley Additive exPlanations [241] assign to each feature an importance value for a particular prediction. Considering that all selection variables of a D_s^+ candidate are contributors to the final prediction, we want to know how much each individual variable contributed to it. An importance value is assigned to a feature by measuring the effect of including the specific feature on the model predictions. The average marginal contribution of the feature is computed from the difference between model predictions obtained with and without including the feature over all possible feature sets. In practice, a Monte Carlo sampling technique [242] is used to approximate the feature SHAP values, which minimises the required computing power.

The resulting feature importance plots of the D_s^+ candidate selection variables for the models trained in the 2 to 4 GeV/c and 12 to 24 GeV/c intervals are shown on figure IV.5.5. The five most important selection variables are the same in every p_T intervals, however their relative importance evolve. In low p_T intervals, the particle identification variable related to the opposite charged sign daughter track with respect to the D_s^+ candidate charge, $n\sigma_{comb,\,1}$, are the most important selection variables. The PID performances decreases for higher p_T tracks and as prompt D_s^+ mesons gets more displaced from the primary vertex at higher p_T, the normalised decay length becomes the most important variable in the high p_T intervals. The distributions of the kaon emission angle related variable $|\cos^3\theta_{K,\phi}|$, figure IV.4.5, and the mass difference between the reconstructed kaon pair and the $\phi(1020)$ phi meson $\Delta M_{KK-\phi}$, figure IV.4.6, are very different for prompt D_s^+ signal and background candidates. That is why they end up to be highly discriminative variables in the model.

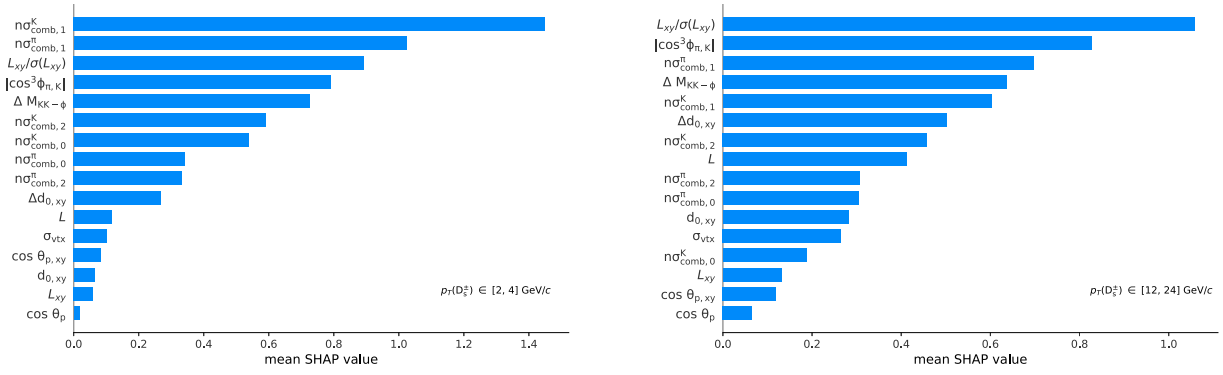


Figure IV.5.5: Feature importance plot for the D_s^+ candidate selection variables using SHAP values of the model trained in the 2 to 4 GeV/c and 12 to 24 GeV/c intervals.

IV.5.3 Optimal model decision value

The choice of the model's optimal decision value is made based on the expected significance of the obtained D_s^+ signal. The expected significance are derived using the FONLL predictions [123–125] avoiding any bias in the extracted raw yield due to a "fine tuning" of the decision value when extracting the signal from real data. The expected quantity of prompt D_s^+ meson signal in each p_T intervals is obtained by isolating the raw signal quantity of the D_s^+ production cross

section formula, eq. IV.1,

$$S_{\text{exp}} = 2 \cdot \frac{\mathrm{d}^2\sigma}{\mathrm{d}y\mathrm{d}p_{\text{T}}}\Big|_{\text{prompt}}^{\text{FONLL}} \cdot \frac{1}{f_{\text{prompt}}} \cdot (Acc \cdot \epsilon)_{\text{prompt}} \cdot c_{\Delta y} \cdot \Delta p_{\text{T}} \cdot \text{B.R.} \cdot \mathcal{L}_{\text{int}} \tag{IV.18}$$

with $\mathrm{d}\sigma/\mathrm{d}p_{\text{T}}\,|_{\text{prompt}}^{\text{FONLL}}$ the p_{T}-differential production cross-section of $\mathrm{D_s^+}$ mesons, $f_{\text{prompt}} = 0.9$ the prompt fraction assumed to be constant across p_{T} (see section IV.9), $(Acc \cdot \epsilon)_{\text{prompt}} \cdot c_{\Delta y}$ the prompt $\mathrm{D_s^+}$ meson efficiency, Δp_{T} the width of the p_{T} interval, B.R. the branching ratio of the studied decay channel and $\mathcal{L}_{\text{int}}$ the integrated luminosity. The background is extracted by fitting the sidebands of the invariant-mass distribution. The expected significance, $S_{\text{exp}}/\sqrt{S_{\text{exp}} + B}$ is shown of the figure IV.5.6 as function of the BDT model output in the [0.8, 1] region. For the four first p_{T} intervals the expected significance sightly increase and reach a maximum value around 0.985 for [1, 2] and [2, 4] GeV/c p_{T} intervals, 0.95 for [4, 6] GeV/c, and 0.92 for [6, 8] GeV/c before decreasing for higher threshold values. For the two latest p_{T} intervals, the expected significance smoothly decreases with increasing threshold values.

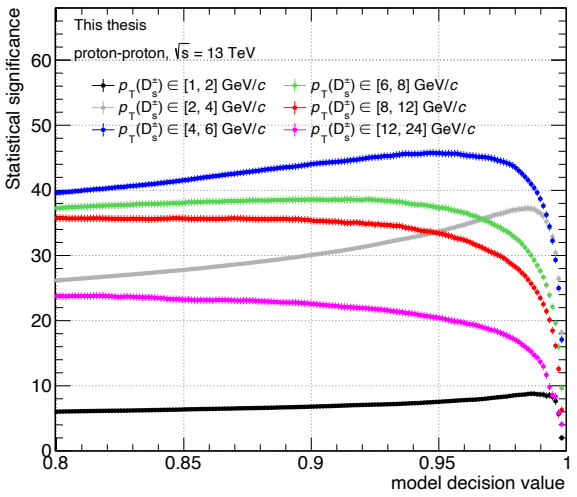


Figure IV.5.6: The expected significance as a function of the BDT model output in the different p_{T} intervals.

p_{T} **intervals** (GeV/c)	**Decision value**
[1, 2]	0.99
[2, 4]	0.96
[4, 6]	0.95
[6, 8]	0.92
[8, 12]	0.92
[12, 24]	0.82

Table IV.5.6: BDT model optimal decision value for the different p_{T} intervals.

The optimal decision values stability is checked by analysing the deviations of the corrected yield variations obtained with variations of the decision value in a $\mathrm{D_s^+}$ selection efficiency interval

of ± 50% around the optimal central value. The procedure is discussed in detail in the dedicated systematic study section. In this way, we obtain decision values guarantying an optimal statistical significance while keeping a reasonably high D_s^+ selection efficiency and avoiding unstable regions of low efficiencies. The table IV.5.6 report the optimal decision values used in the analysis.

IV.6 High multiplicity SPD trigger correction

The high multiplicity SPD triggered collision data sample corresponds to the $N_{tracklets} \in [60, 100]$ interval. Since a full trigger efficiency is not reached in this region, the turn-on curve is used to correct the number of D_s^+ meson signal candidates extracted per event in the data sample.

In the 2018 [f, h, j, k, l, m, o, p] set of periods, both the minimum-bias and the high multiplicity SPD trigger were used during the data taking. The corrected $N_{tracklets}$ distributions for the collisions events having a selected D_s^+ meson candidate, *i.e.* passing the signal selection cut, for both triggers in the same data sample, are shown on the left panel of figure IV.6.1. By requiring a selected D_s^+ meson in the event, possible biases in the trigger response due to the presence of a reconstructed heavy-flavour hadron are avoided. The high multiplicity tracklet profiles are corrected for their z_{vertex} dependence (section IV.2.1). Two options were investigated, a correction using the $N_{tracklets}$ profiles and reference value from the minimum-bias data sample, in blue on the figure, and a correction with the high multiplicity profile, in red on the figure. On the right panel of figure IV.6.1, the ratio of the HMSPD $N_{tracklets}$ distributions to the $N_{tracklets}$ minimum-bias distribution shows the "turn-on curve" of the HMSPD. The colour code employed are consistent between the two panels. As more and more charged particles are produced in the event, the HMSPD trigger efficiency smoothly increases up to the full efficiency above a certain $N_{tracklets}$ threshold. The smooth increase of the efficiency is due to the smearing of the $N_{tracklets}$ with respect to the number of SPD FAST-OR ships fired.

The statistics enhancement obtained at high multiplicity, of a factor $\simeq 120$, is due to the larger live-time of the HMSPD trigger and/or different trigger pre-scaling factors between MB and HMSPD triggers. For the $\text{HMSPD}_{\text{MB corr}}$ turn-on curve, the full efficiency is reached for $N_{tracklets} = 80$ and form a plateau at high multiplicity. For the $\text{HMSPD}_{\text{HM corr}}$ curve, the full efficiency is reached around $N_{tracklets} = 65$ and the curve start to decrease at higher multiplicity. The different values of $N_{tracklets}$ reached and the curve trends are direct consequences of the two correction strategies. A decreasing trend of the turn-on curve at high multiplicity is not expected, it was therefore decided to use the minimum-bias tracklet profiles to correct the $N_{tracklets}$ in the HMSPD data sample.

The weights are obtained by renormalising to unity the plateau region of the HMSPD turn-on curve. In the default case, the ratio value found in the plateau region of the fitted curve is used for the renormalisation but alternative approaches were explored and are discussed in a dedicated section of the systematic uncertainties study chapter V.5. Each D_s^+ meson candidate selected is re-weighted by the inverse of the normalised turn-on curve value at the $N_{tracklets}$ corresponding to the associated event. The same correction is performed when counting the number of events used to compute the event normalisation factor (section IV.2.3). The derived weights are shown on figure IV.6.2. The number of selected event in $N_{tracklets}$ [60, 100] increase from 108 millions to 119 millions after this correction.

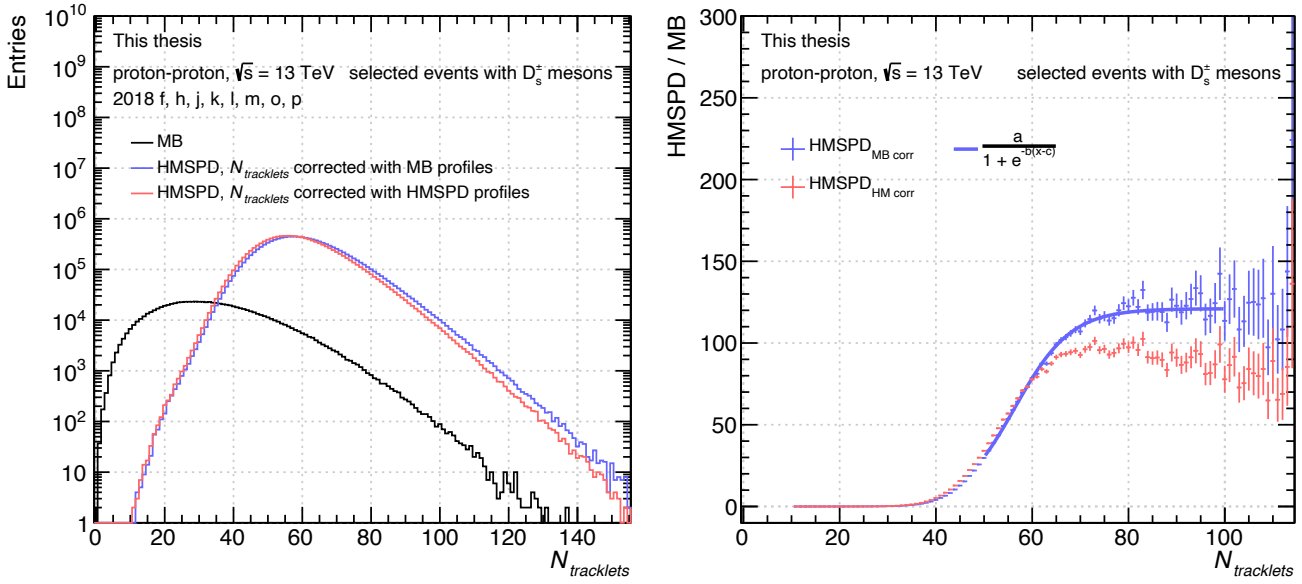


Figure IV.6.1: On the left panel, the $N_{tracklets}$ distribution in the 2018 [f, h, j, k, l, m, o, p] data sample for HMSPD and MB collisions events containing a selected D_s^+ meson candidate. The blue curve corresponds to the $N_{tracklets}$ distribution corrected with the minimum-bias tracklets profile, while for the red curve the distribution is corrected with the HMSPD profiles. On the right panel, the corresponding $N_{tracklets}$ distribution ratio HMSPD in the two configuration over MB.

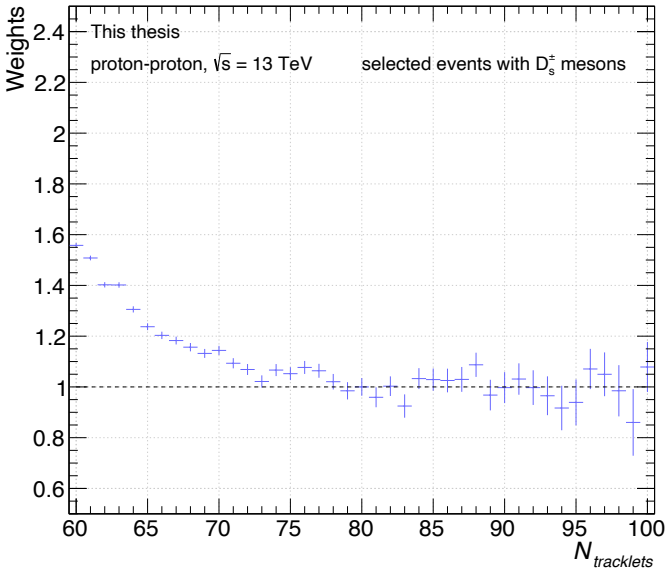


Figure IV.6.2: High-multiplicity SPD trigger correction weights.

IV.7 Raw yield extraction

The raw yield extraction procedure is performed by fitting the invariant-mass distribution of the selected $\mathrm{D_s^+}$ meson candidates as showed on the figures IV.7.2, IV.7.6, IV.7.7, IV.7.8 and IV.7.9. As described in the $\mathrm{D_s^+}$ meson reconstruction section IV.4, the candidate are built out of triplets of candidates and they are kept for two mass hypotheses: $i = \{\pi^\pm, \mathrm{K}^\mp, \mathrm{K}^\pm\}$ and $i = \{\mathrm{K}^\pm, \mathrm{K}^\mp, \pi^\pm\}$. This may leads to a modification of the invariant-mass distribution and to a possible double counting of the $\mathrm{D_s^+}$ candidates in the raw yield extraction procedure. However, a study in Monte Carlo simulations shows that after applying the decay track criteria and performed the candidate selection with the BDT model, the contribution of the $\mathrm{D_s^+}$ candidates with the wrong mass hypothesis assigned to the pion and kaon daughters is strongly suppressed. Representing around 1‰ of the raw yield extracted value, the contribution is considered negligible. The suppression of these so called reflected candidates, is particularly due to the use of the particle identification informations on the daughter tracks and the absolute difference between the reconstructed invariant mass of the kaon pair and the $\phi(1020)$ meson mass selection criteria.

In the following paragraphs and sections:

- The minimum-bias analysis refers to the analysis performed on the full minimum-bias triggered data sample without any requirement on the multiplicity of the collision. The analysis leads to the $\mathrm{D_s^+}$ production p_T-differential cross section (eq. IV.1) measurement.
- The multiplicity analyses are done in the collision multiplicity intervals defined in section IV.2.3 for both the minimum-bias and high multiplicity triggered data samples. The $\mathrm{D_s^+}$ corrected production yield (eq. IV.2) will be built out of this analyses.

The analyses covers the 1 to 24 GeV/c $\mathrm{D_s^+}$ p_T interval, however, due to the limited statistics in the low and high multiplicity intervals the raw yield extraction is not feasible in the 1 to 2 GeV/c p_T interval. The p_T binning between the minimum-bias and multiplicity analyses are also different for the same reason.

The fitting function used for representation of the signal consists of a sum of two Gaussians describing the $\mathrm{D_s^+}$ meson peak and the peak of the $\mathrm{D^+}$ meson decaying in the same channel ($\mathrm{m_{PDG}(D^+)} = 1869.65 \pm 0.05$ MeV [37]). The background $\mathrm{D_s^+}$ candidate distribution is described by a decaying exponential function for all the p_T and multiplicity intervals considered. In the minimum-bias analysis, the function parameters are left free during the fit, most importantly, the Gaussian mean and width of the $\mathrm{D_s^+}$ and $\mathrm{D^+}$ peaks. The raw yield quantity is obtained by integration of the $\mathrm{D_s^+}$ meson Gaussian function in the invariant-mass range $\pm 3\sigma_{\mathrm{D_s^+}}$ around the mean value $\mu_{\mathrm{D_s^+}}$. The quantity of background is obtained by integration of the related decaying exponential function in the same invariant mass range.

For the minimum-bias analysis, the total fit functions in blue are shown on the invariant mass distributions of the figure IV.7.2. The fitted background function correspond to the red dotted curve. A summary of the fitting procedure is presented in table IV.7.1 where the raw yield, the signal to background ratio, the statistical significance of the signal and the reduced χ^2 are reported for the nine p_T intervals. The low p_T region is the most affected by the large fraction of combinatorial background as shown by the low S/B ratios, the statistical significance goes from 6 to 14 in the two first p_T intervals. At intermediate p_T the statistical significance is between 25 and 30. At high p_T, the $\mathrm{D_s^+}$ production yield is low, causing a decrease of the statistical significance.

Despite the different performances obtained across the p_{T} intervals, overall, the reduced χ^2 of the fits are around unity, reflecting the reasonable description of the invariant mass distributions and the statistical significance is above 3 which is the usual chosen limit of the feasibility of the raw yield extraction for this type of analysis.

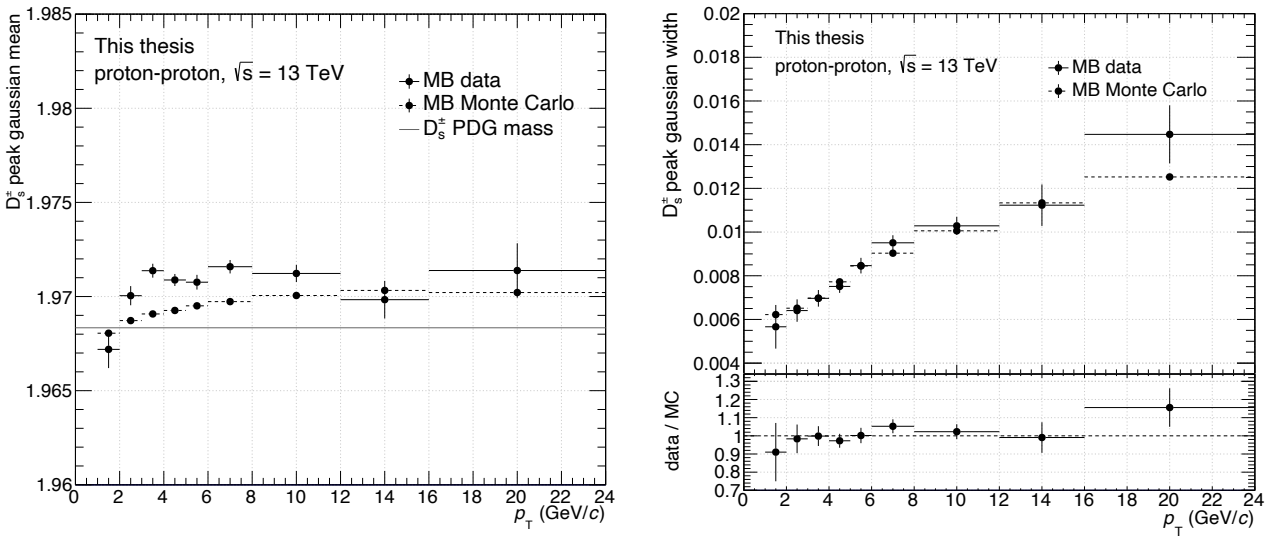


Figure IV.7.1: Gaussian mean value (left) and width (right) of the fitted $\mathrm{D_s^+}$ meson peak for the p_{T} intervals in minimum-bias proton-proton collisions at $\sqrt{s} = 13$ TeV. The $\mathrm{D_s^+}$ meson PDG mass is reported [37] and draw in grey.

p_{T} **intervals** (GeV/c)	**Raw yield**	**S/B** (3σ)	**Significance** (3σ)	χ^2**/ndf**
[1, 2]	215 ± 35	0.23	6.3 ± 0.9	1.15
[2, 3]	1598 ± 115	0.14	13.9 ± 0.9	0.77
[3, 4]	1847 ± 88	0.32	21.4 ± 0.9	1.10
[4, 5]	1621 ± 62	0.75	26.4 ± 0.8	0.85
[5, 6]	1075 ± 45	1.24	24.4 ± 0.7	0.79
[6, 8]	1497 ± 51	1.51	30.1 ± 0.7	0.95
[8, 12]	899 ± 37	2.37	25.2 ± 0.7	0.97
[12, 16]	254 ± 21	1.76	12.7 ± 0.7	0.84
[16, 24]	168 ± 16	2.75	11.1 ± 0.7	0.96

Table IV.7.1: The raw yield extracted, signal to background ratio, statistical significance and reduced χ^2 of the $\mathrm{D_s^+}$ invariant-mass fit for the p_{T} intervals in minimum-bias proton-proton collisions at $\sqrt{s} = 13$ TeV.

The values of the Gaussian mean (left panel) and the width (right panel) are shown on figure IV.7.1 from the fit are compared to the values obtained in Monte Carlo simulations. In Monte Carlo simulations, the invariant mass distribution is build with only the selected $\mathrm{D_s^+}$ candidate reconstructed from true prompt $\mathrm{D_s^+}$ mesons. The distributions are fitted with a single Gaussian function. In the first low p_{T} interval and at high p_{T}, the Gaussian mean are compatible within

1σ of uncertainty, however in the intermediate p_{T} region, the gaussian mean values in data are systematically above the one obtained in Monte Carlo simulations. The effect of this discrepancy on the extracted raw yield will be discussed when evaluating the related systematic uncertainties. The Gaussian width values in data and Monte Carlo simulations are in agreement within uncertainties as shown of the ratio of the right panel of figure IV.7.1. The increase of the peak width as a function of p_{T} is caused by the deterioration of the p_{T}-resolution on the daughter track momenta.

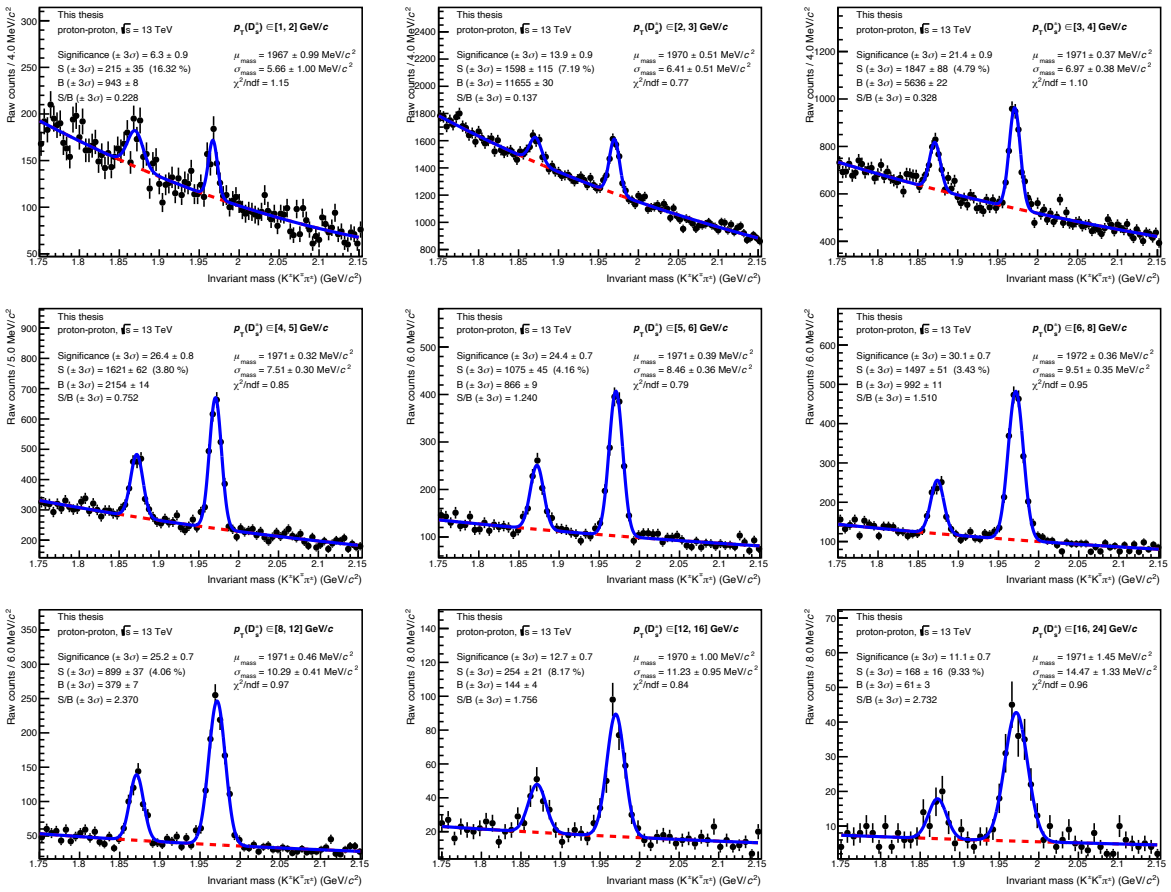


Figure IV.7.2: Fitted invariant-mass distribution of $\mathrm{D_s^+}$ meson candidates for p_{T} intervals in minimum-bias proton-proton collisions at $\sqrt{s} = 13$ TeV. The right peak corresponds to the $\mathrm{D_s^+}$ and the left peak to the $\mathrm{D^\pm}$ in the same decay channel.

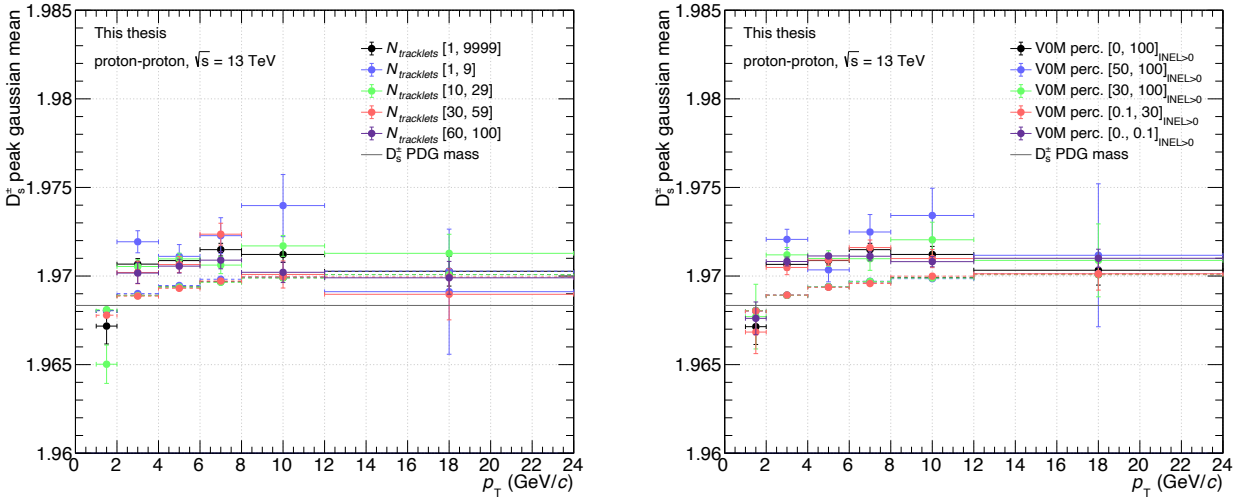


Figure IV.7.3: Gaussian mean value of the fitted D_s^+ meson peak for the p_T intervals in proton-proton collisions at $\sqrt{s} = 13$ TeV for the $N_{tracklets}$ (left) and V0M percentiles (right) intervals. The D_s^+ meson PDG mass [37] is shown with the grey horizontal line.

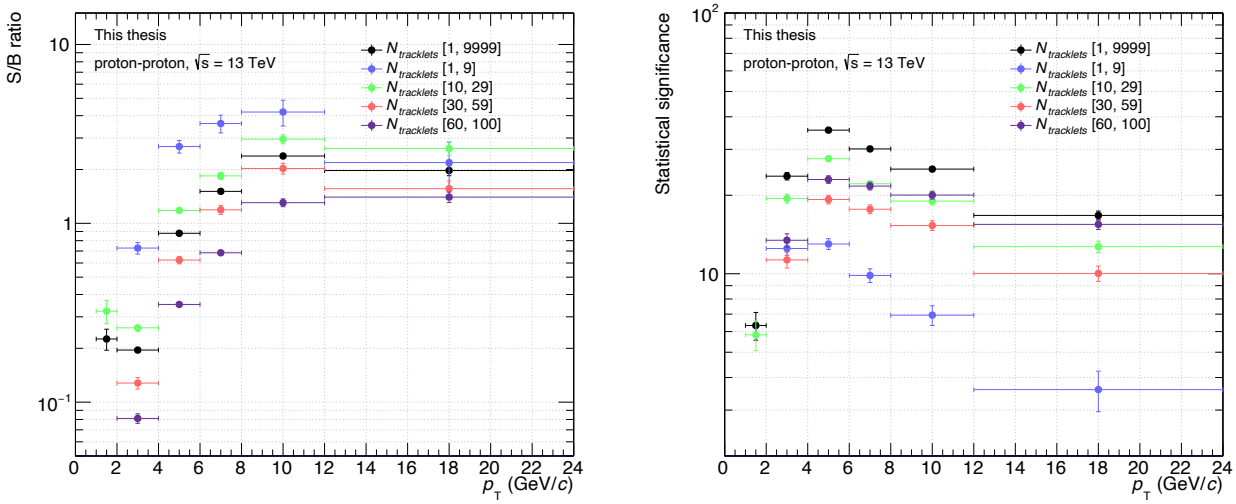


Figure IV.7.4: The signal to background ratio (left) and the statistical significance (right) for the p_T and $N_{tracklets}$ intervals in proton-proton collisions at $\sqrt{s} = 13$ TeV.

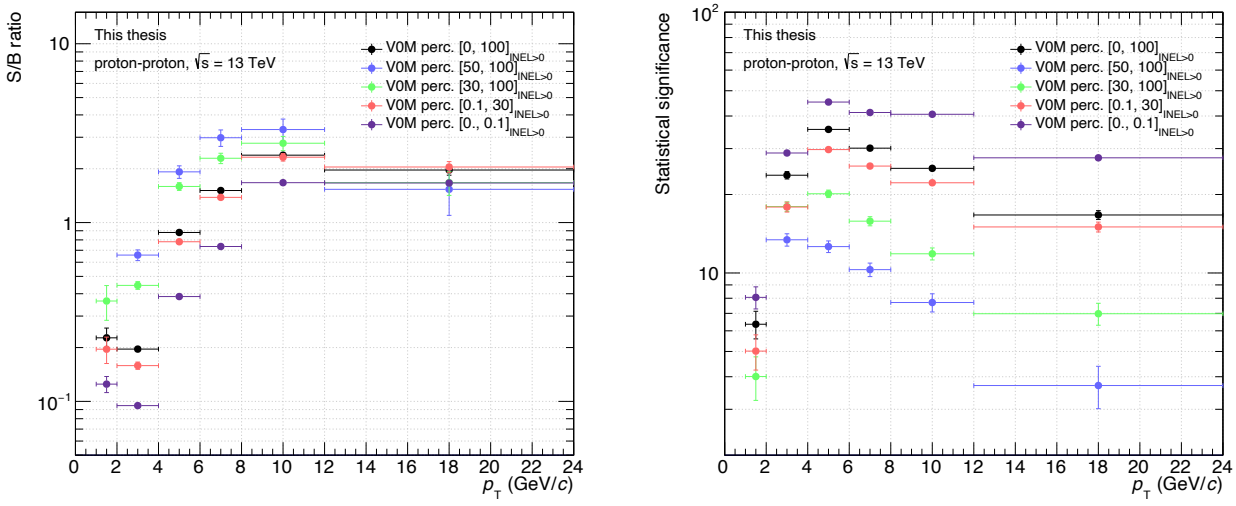


Figure IV.7.5: The signal to background ratio (left) and the statistical significance (right) for the p_{T} and V0M percentiles intervals in proton-proton collisions at $\sqrt{s} = 13$ TeV.

The same fit procedure is applied in the multiplicity analyses, with the exception of the $\mathrm{D_s^+}$ gaussian width parameter which is fixed to the values found in the minimum-bias analysis, since we do not expect a variation of the width as a function of multiplicity. This choice will be rediscussed when estimating the related systematic uncertainties. The fitted invariant-mass distributions in the lowest and the highest multiplicity intervals are shown on the figures IV.7.6, IV.7.7 for the $N_{tracklets}$ analysis and on figures IV.7.8 and IV.7.9 for the V0M percentile analysis. The gaussian mean value obtained are reported on figure IV.7.3, on the left panel for the $N_{tracklets}$ analysis and on the right panel for the V0M percentile analysis. Compared to the minimum-bias analysis, similar Gaussian mean values are observed.

The signal to background ratio and the statistical significance extracted from the fits in the different $N_{tracklets}$ intervals are shown on figure IV.7.4 and in the different V0M percentiles intervals on figure IV.7.5. A clear hierarchy of the S/B ratio is observed as a function of multiplicity in all p_{T} intervals, low multiplicity intervals have a higher S/B than the high multiplicity one. As more tracks are present in high multiplicity events, more background $\mathrm{D_s^+}$ candidates are built from random combinations of tracks still passing the BDT model selection. Overall, the significance is above three for all p_{T} and multiplicity intervals. It increases as a function of multiplicity and reaches maximum values at intermediate p_{T}. Because of the lack of statistics, the analysis was not performed in the 1 to 2 GeV/c p_{T} interval for the [1, 9], [30, 59] and [60, 100] $N_{tracklets}$ intervals and the $[50, 100]_{\mathrm{INEL>0}}$ V0M percentiles interval.

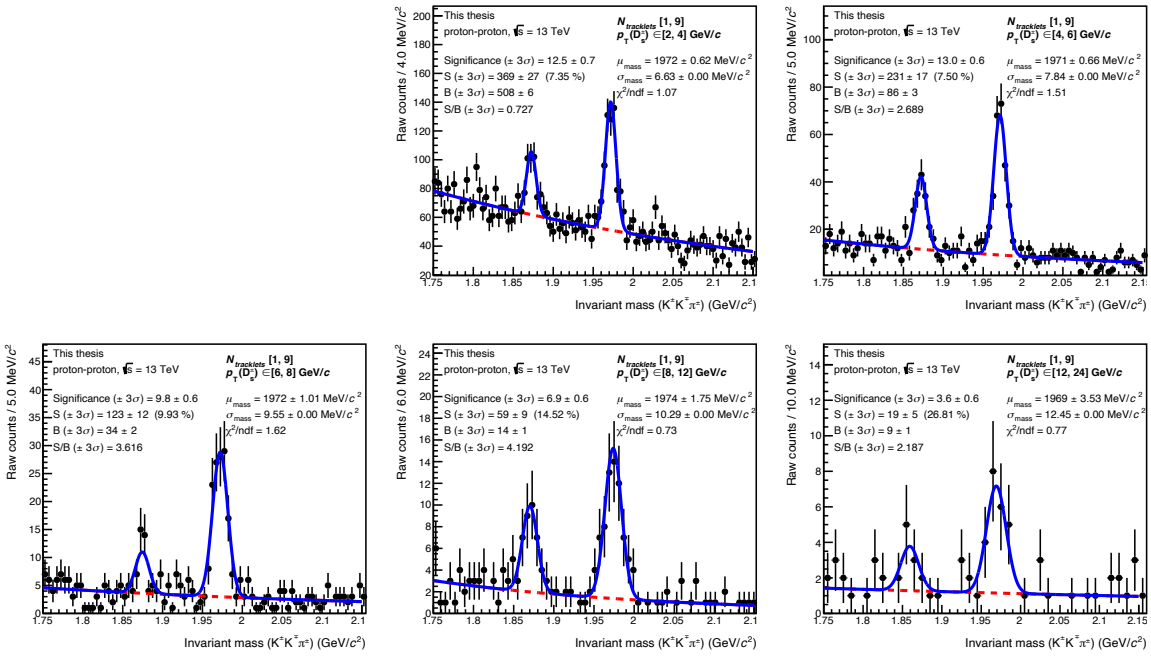


Figure IV.7.6: Fitted invariant-mass distribution of D_s^+ meson candidates for the p_T intervals in minimum-bias proton-proton collisions at $\sqrt{s} = 13$ TeV in the $N_{tracklets}$ [1, 9] interval. The right peak corresponds to the D_s^+ and the left peak to the $D^\pm$ in the same decay channel.

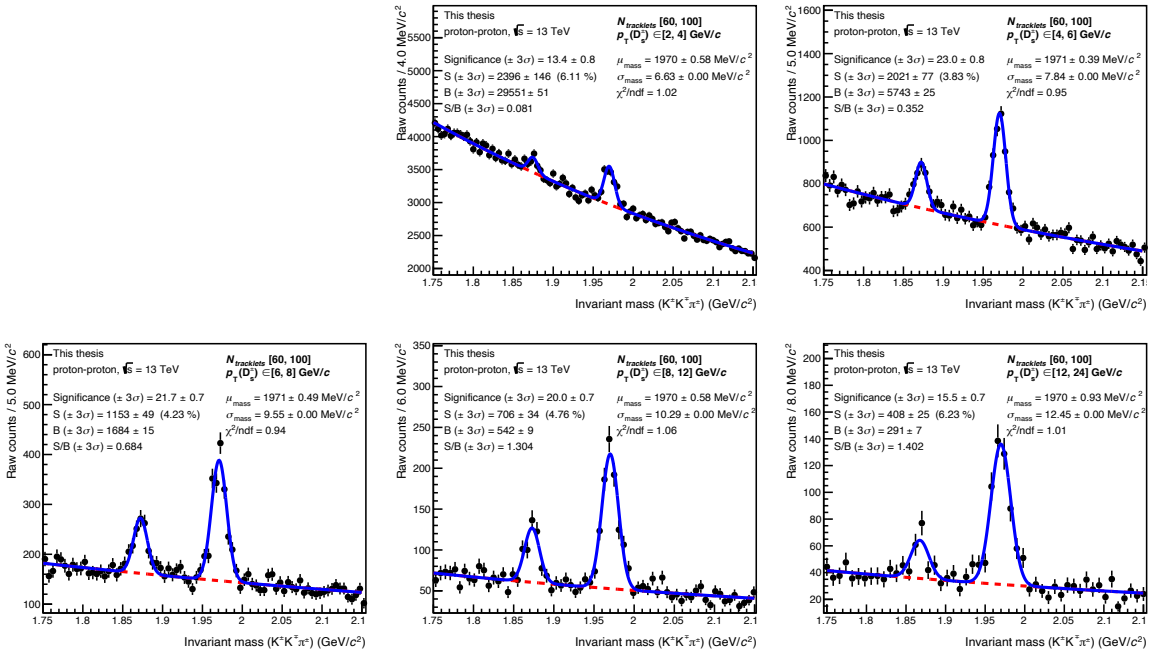


Figure IV.7.7: Fitted invariant-mass distribution of D_s^+ meson candidates for the p_T intervals in high multiplicity proton-proton collisions at $\sqrt{s} = 13$ TeV in the $N_{tracklets}$ [60, 100] interval. The right peak corresponds to the D_s^+ and the left peak to the $D^\pm$ in the same decay channel.

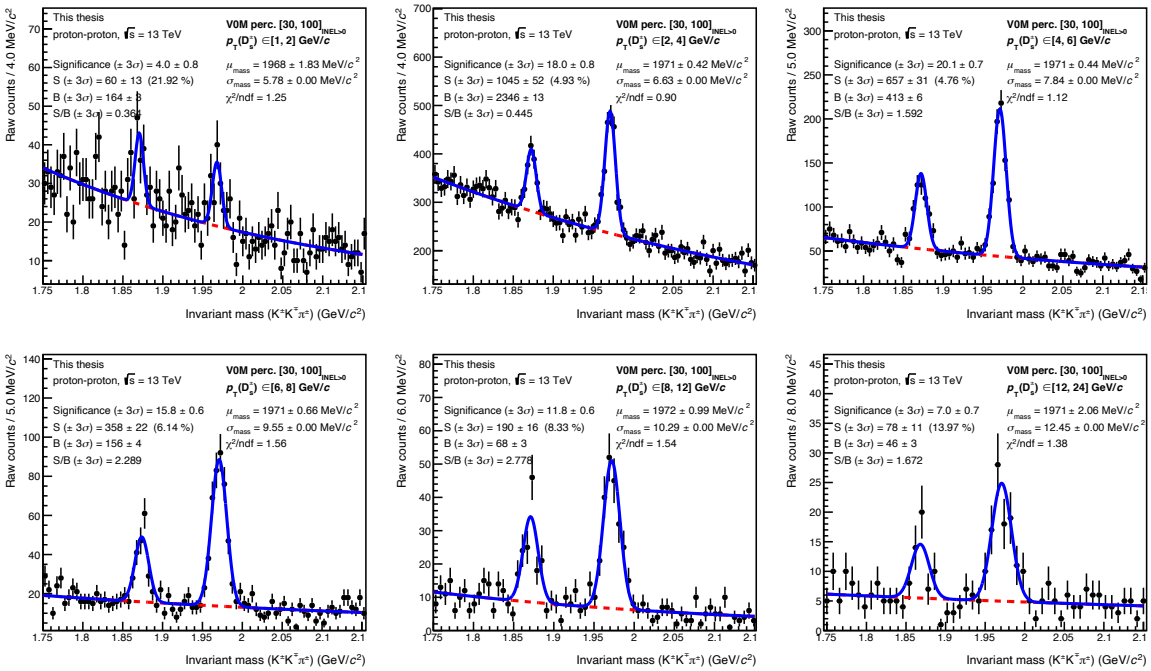


Figure IV.7.8: Fitted invariant-mass distribution of $\mathrm{D_s^+}$ meson candidates for the p_{T} intervals in minimum-bias proton-proton collisions at $\sqrt{s} = 13$ TeV in the V0M percentile $[30, 100]_{\mathrm{INEL>0}}$ interval. The right peak corresponds to the $\mathrm{D_s^+}$ and the left peak to the $\mathrm{D^\pm}$ in the same decay channel.

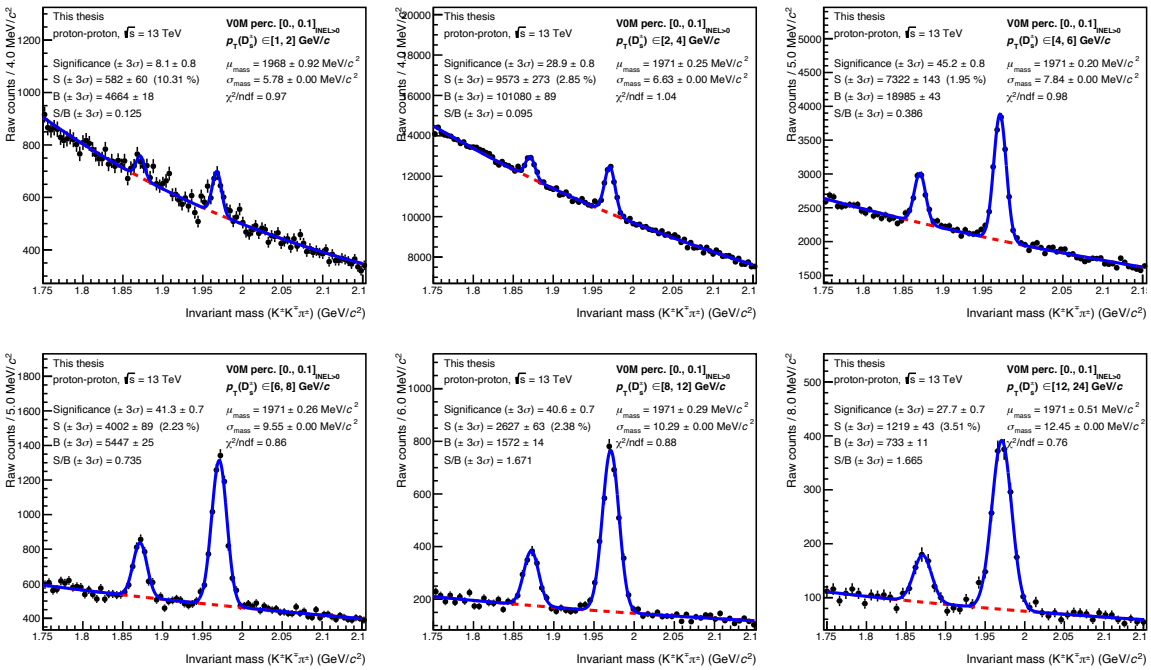


Figure IV.7.9: Fitted invariant-mass distribution of $\mathrm{D_s^+}$ meson candidates for the p_{T} intervals in high multiplicity proton-proton collisions at $\sqrt{s} = 13$ TeV in the V0M percentile $[0., 0.1]_{\mathrm{INEL>0}}$ interval. The right peak corresponds to the $\mathrm{D_s^+}$ and the left peak to the $\mathrm{D^\pm}$ in the same decay channel.

Comparison of the D_s^+ meson candidate selection approaches

The raw yield extraction performances obtained with the machine learning based D_s^+ candidate selection approach are compared to those obtained from the standard sequential selection approach. Since the $N_{tracklets}$ analysis was performed first, the comparison is made in this analysis. The selection criteria of the standard approach are summarised in table IV.7.2 and the conservative PID strategy is performed for the particle identification selection as described in section IV.5.2.

p_T (GeV/c)	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
L_{xy} (cm)	> 0.02	> 0.03	> 0.03	> 0.04	> 0.04	> 0.04
$L_{xy}/\sigma(L_{xy})$	> 3	> 3	> 4	> 4	> 4	> 4
σ_{vtx} (cm)	< 0.03	< 0.03	< 0.04	< 0.04	< 0.04	< 0.04
$\cos\theta_{\mathrm{p}}$	> 0.9	> 0.92	> 0.92	> 0.92	> 0.92	> 0.92
$\cos\theta_{\mathrm{p,xy}}$	> 0.9	> 0.92	> 0.92	> 0.92	> 0.92	> 0.92
$\lvert\cos^3\phi_{\pi,K}\rvert$	> 0.1	> 0.1	> 0.1	> 0.1	> 0.1	> 0.1
$\Delta d_{0,xy}$	$< 3.$	$< 3.$	$< 3.$	$< 3.$	$< 3.$	$< 3.$
Δ $M_{KK\text{-}\phi}$ (MeV)	< 0.007	< 0.007	< 0.007	< 0.007	< 0.007	< 0.007

Table IV.7.2: Selection criteria for the topological and kinematical D_s^+ candidates selection variables.

The statistical significance ratio between the two procedures is shown on figure IV.7.10 as a function of p_T in the different $N_{tracklets}$ intervals. Globally, the gain in statistical significance is above 20% when using the machine learning approach. In the 2 to 4 GeV/c p_T interval, the gain reaches at least 60%, showing that the BDT model seems really to make the difference in regions where the combinatorial background is particularly important. The standard approach performance can be however improved by being more selective on the PID variable at low p_T. In the 1 to 2 GeV/c p_T interval, the standard selection approach, even if possible, performs poorly and is less stable than the BDT selection.

One concern when using the machine learning method is to unintentionally exploit hidden correlations between the selection variables and the invariant mass of the D_s^+ candidates. This may lead to a more complex shape of the invariant-mass distribution for the background candidates. However, with both methods applied here, the resulting background candidate distribution is well described by the same fitting function in every p_T and multiplicity intervals. The final validation of the machine learning approach will be discussed in the result section IV.10 where the production yields will be compared.

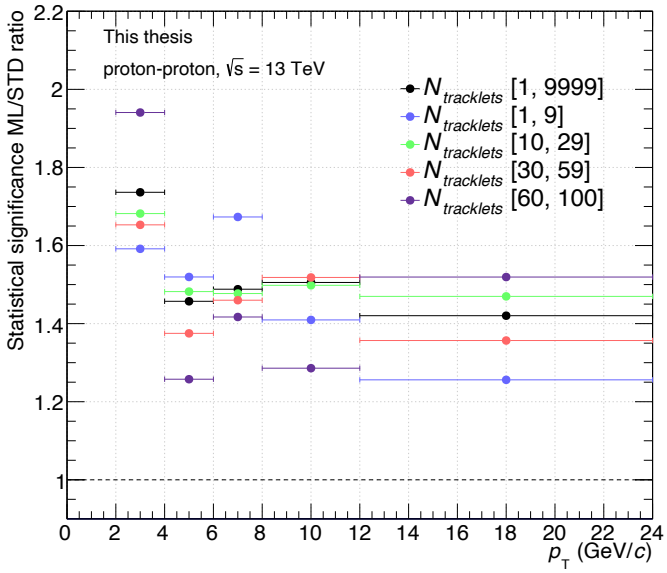


Figure IV.7.10: Comparison of the statistical significance obtained from the D_s^+ meson candidates selection with the BDT models and with the standard sequential selection approach for the p_T and $N_{tracklets}$ intervals in minimum-bias proton-proton collisions at $\sqrt{s}$ = 13 TeV.

IV.8 Acceptance and efficiency corrections

The raw yield is corrected by the detector acceptance determined by the D_s^+ meson decay kinematics in the rapidity region of interest and by the detection efficiency, extracted from dedicated Monte Carlo simulations. The acceptance-times-efficiency correction factor is defined as,

$$Acc^* \cdot \epsilon = \frac{N_{gen}^{daughters\,in\,acc}|_{|y|<y_{fid}}}{N_{gen}|_{|y|<0.5}} \cdot \frac{N_{reco,sel}|_{|y|<y_{fid}}}{N_{gen}^{daughters\,in\,acc}|_{|y|<y_{fid}}} \tag{IV.19}$$

with $N_{gen}^{daughters\,in\,acc}|_{|y|<y_{fid}}$ the number of generated D_s^+ mesons in the fiducial acceptance with daughter tracks in $|\eta| < 0.9$ and $p_T > 0.1$ GeV/c, $N_{reco,sel}|_{|y|<y_{fid}}$ the number of reconstructed D_s^+ mesons in the fiducial acceptance passing the candidate selection, and $N_{gen}|_{|y|<0.5}$ the number of generated D_s^+ mesons in $|y| < 0.5$ decaying in the decay channel of interest. The superscript on the acceptance term is there to notify the additional correction factor for the rapidity coverage, $Acc^* = Acc \cdot c_{\Delta y}$. Only events generated with a longitudinal position of the primary vertex within a $|z_{vertex}| < 10$ cm region around the nominal collision point are considered. The efficiency is transverse-momentum dependent due to the different track-reconstruction efficiencies and applied D_s^+ candidate selections.

The reconstruction efficiency and the resolution on the position of the collision vertex (see figure III.3.1) depend on the number of tracks used to perform the vertex reconstruction. Since the selection of the D_s^+ candidates is heavily based on the displaced decay-vertex topology, the reconstruction and the selection efficiencies of the D_s^+ mesons are expected to depend on the multiplicity of charged particles produced in the collision. As the primary vertex position is better constrained at high multiplicity, an increase of the selection efficiency is expected as the multiplicity

goes up. In order to compensate the discrepancies between the measured multiplicity distribution in data and in Monte Carlo simulations, the D_s^+ meson efficiencies are re-weighted. The left panel of figure IV.8.1 shows the $N_{tracklets}$ distributions of events having a selected D_s^+ candidate within a mass window ± 20 MeV in data and Monte Carlo simulations in different proton-proton collision data samples. The requirement on the D_s^+ candidate is needed for a fair comparison of the distributions, the generated Monte Carlo events are enriched with D_s^+ mesons. Also, the multiplicity distributions in Monte Carlo simulations are underestimating the multiplicity observed in data.

The multiplicity weights shown on figure IV.8.1 are defined as the ratio of the $N_{tracklets}$ distribution in the data over the one in simulations. They are used to re-weight the efficiencies in the minimum-bias and $N_{tracklets}$ multiplicity analyses. As expected, the weight distributions associated to the three minimum-bias data samples are very similar.

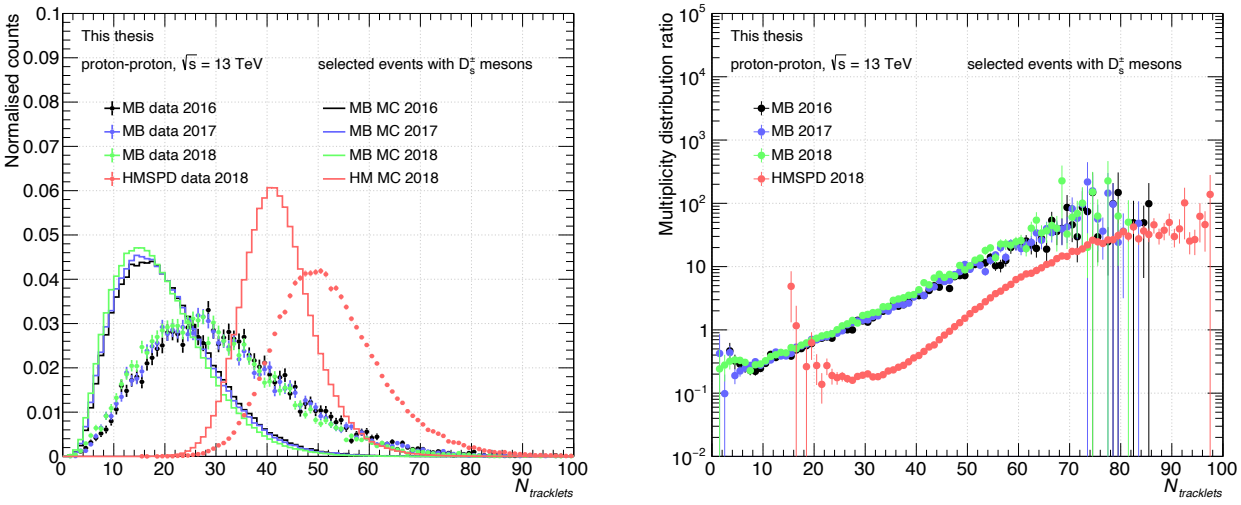


Figure IV.8.1: Left panel: $N_{tracklets}$ distributions of events having a selected D_s^+ candidate within a mass window ± 20 MeV in data and Monte Carlo simulations in proton-proton collisions at $\sqrt{s} = 13$ TeV. Right panel: The corresponding data/MC distribution ratios defining the multiplicity weights.

In the minimum-bias analysis, the acceptance is computed with a toy Monte Carlo and the efficiency with the full simulations, while the full simulations are used to compute both correction in the case of the multiplicity-dependent analyses.

IV.8.1 Minimum-bias analysis

In order to reduce the statistical uncertainty, the acceptance factor is computed with a toy Monte Carlo. Since it is a pure geometrical term, the CPU time needed to compute the factor is negligible compared to the full simulations. The rapidity and transverse momentum of the D_s^+ mesons are sampled from the prompt D^0 meson and feed-down D mesons rapidity and p_T distributions predicted by FONLL [123–125]. The use of FONLL predictions is motivated by their good description of D mesons production in proton-proton collisions at LHC energies [122, 243]. The D_s^+ meson predictions are not available in FONLL, however, as it will be discussed in the

dedicated systematic uncertainty study section, the acceptance correction factor computed using the rapidity and p_{T} distributions of FONLL and using the full simulation with PYTHIA 8 are compatible within a percent. The acceptance correction factor for the prompt and feed-down $\mathrm{D_s^+}$ mesons, decaying in the $\mathrm{D_s^+} \to \phi(1020)(\to \mathrm{K^+K^-})\pi^+$ decay channel, as a function of their p_{T} are shown on the left panel of figure IV.8.2. For both type of $\mathrm{D_s^+}$ mesons, the distributions are similar and increases with p_{T}, starting from ∼0.6 to reach ∼1.55 at high p_{T} due to the Lorentz boost.

On the right panel of figure IV.8.2, the full acceptance-times-efficiency corrective term is shown for the prompt and feed-down $\mathrm{D_s^+}$ mesons. The proper decay length of the $\mathrm{D_s^+}$ meson is increasing with its p_{T}, therefore, the decay vertex becomes further displaced from the primary vertex with a better resolution on its measured position. The $\mathrm{D_s^+}$ candidate topological variables becomes more selective resulting in the increase of the selection efficiency starting below the percent in the first p_{T} interval up to 50% at high p_{T}. As feed-down $\mathrm{D_s^+}$ mesons are more displaced due to the B meson life-time, their selection efficiency is higher with respect to the prompt $\mathrm{D_s^+}$ mesons. The drop of the feed-down $\mathrm{D_s^+}$ meson efficiency in the last two p_{T} bins could be caused by the rejection of very displaced $\mathrm{D_s^+}$ mesons by the BDT since the feed-down $\mathrm{D_s^+}$ candidates were not considered in the model training as signal candidates to avoid a low fraction of prompt $\mathrm{D_s^+}$ meson in the extracted raw yield.

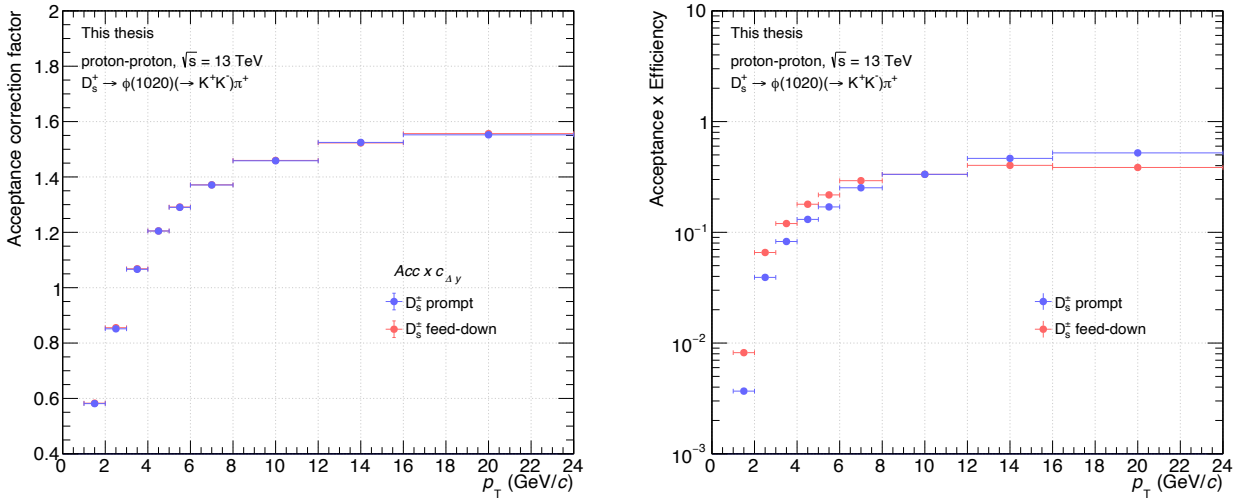


Figure IV.8.2: Acceptance factor (left panel) and acceptance-times-efficiency correction (right panel) of prompt and feed-down $\mathrm{D_s^+}$ mesons in minimum-bias proton-proton collisions at $\sqrt{s} = 13$ TeV. See text for definitions.

IV.8.2 Multiplicity analyses

When using the $N_{tracklets}$ multiplicity estimator, we observed a multiplicity dependence of the rapidity distribution of the D_s^+ mesons in proton-proton collisions generated in the Monte Carlo simulations as shown on the left panel of figure IV.8.3. This evolution of the shape of the distribution cannot be reproduced by the toy Monte Carlo simulations. Therefore, the full acceptance time efficiency term is computed using the full simulations. No multiplicity dependence on the rapidity distribution of the D_s^+ mesons is observed when selecting multiplicity intervals with V0M percentile estimator.

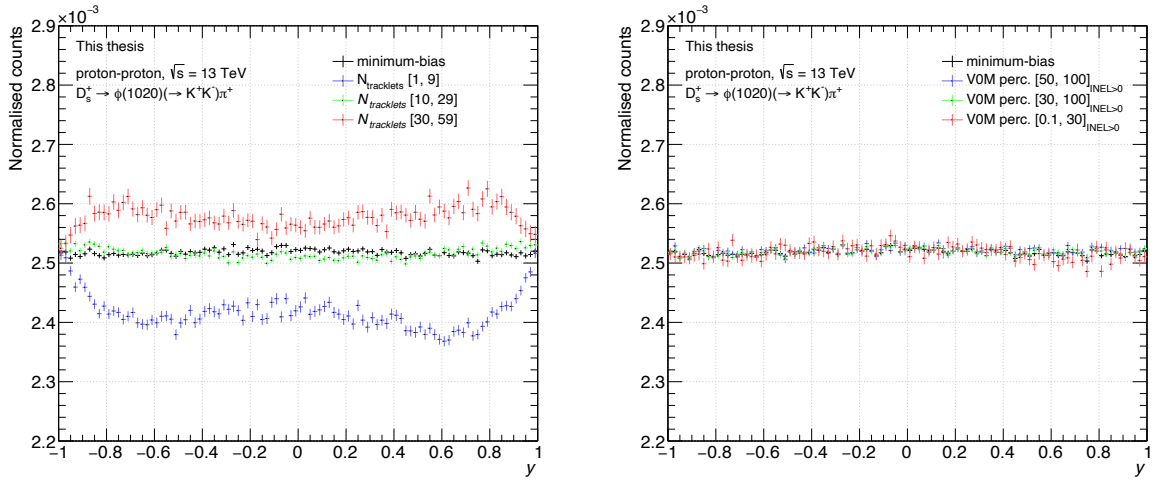


Figure IV.8.3: Rapidity distribution of D_s^+ mesons in $N_{tracklets}$ (left panel) and V0M percentiles (right panel) intervals in proton-proton collisions at $\sqrt{s} = 13$ TeV generated with PYTHIA 8.

Multiplicity weights with the V0M percentile estimator

The V0M percentile calibration procedure performed in data is not reproduced in Monte Carlo simulation, the selection in multiplicity interval is done by re-weighting the efficiencies in a similar manner as done for the minimum-bias and $N_{tracklets}$ analyses. In data, the $N_{tracklets}$ distributions of events having a selected D_s^+ candidate within a mass window ± 20 MeV are sliced in the different V0M percentiles intervals as shown on the top panels of figure IV.8.4. The Monte Carlo multiplicity distributions with the same requirement on the events are integrated in multiplicity. The V0M percentile [0, 0.1] distribution corresponding to HMV0M triggered events is not comparable to the HM MC distribution since the HMV0 trigger is not simulated and the high multiplicity selection is done at mid rapidity with $N_{tracklets}$ (see section IV.3).

The multiplicity weights in each multiplicity interval is defined as the ratio of the sliced multiplicity distribution in data over the integrated distribution in simulations. The weights are shown on the bottom panels of the figure IV.8.4, across the different years of data taking, they are quite similar.

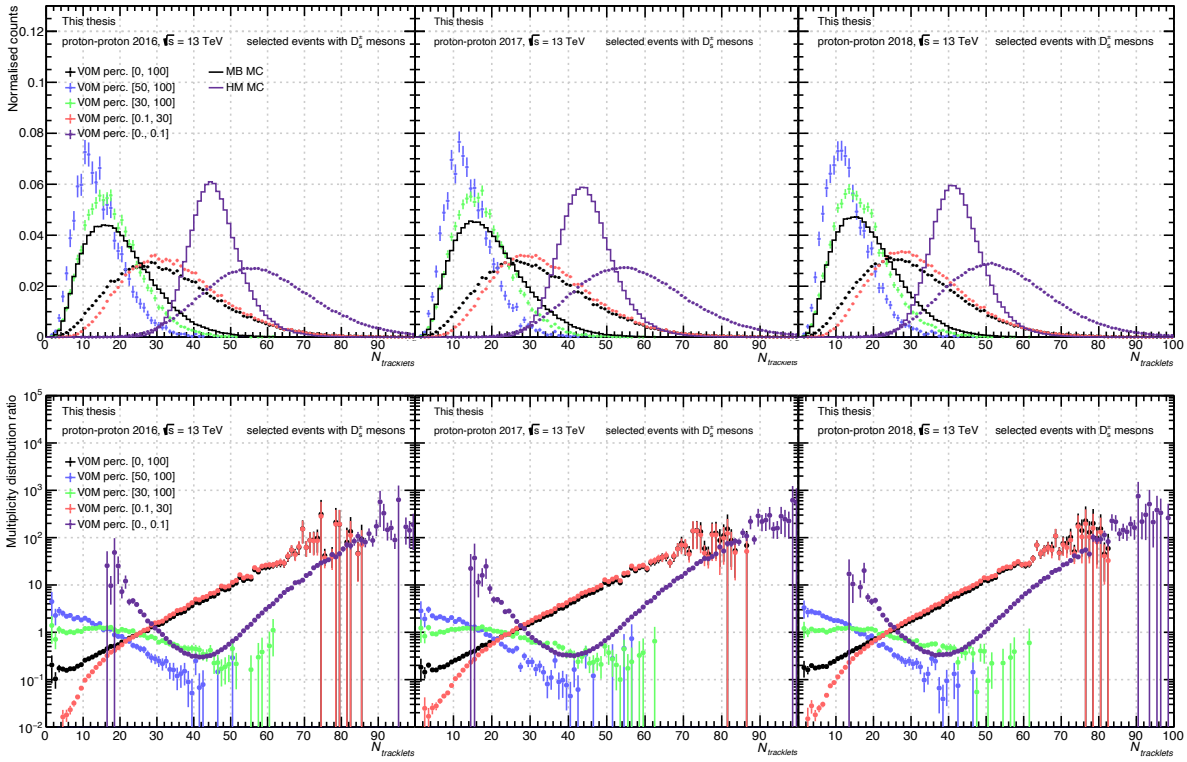


Figure IV.8.4: Top panels: $N_{tracklets}$ distributions of events having a selected $\mathrm{D_s^+}$ candidate within a mass window ± 20 MeV, for different V0M percentile intervals, in proton-proton collisions at $\sqrt{s}$ = 13 TeV. In the the Monte Carlo simulations, the distributions, MB and HMV0M MC are integrated in multiplicity. Each column corresponds to one year of data taking. Bottom panels: The corresponding data/MC distribution ratios defining the multiplicity weights.

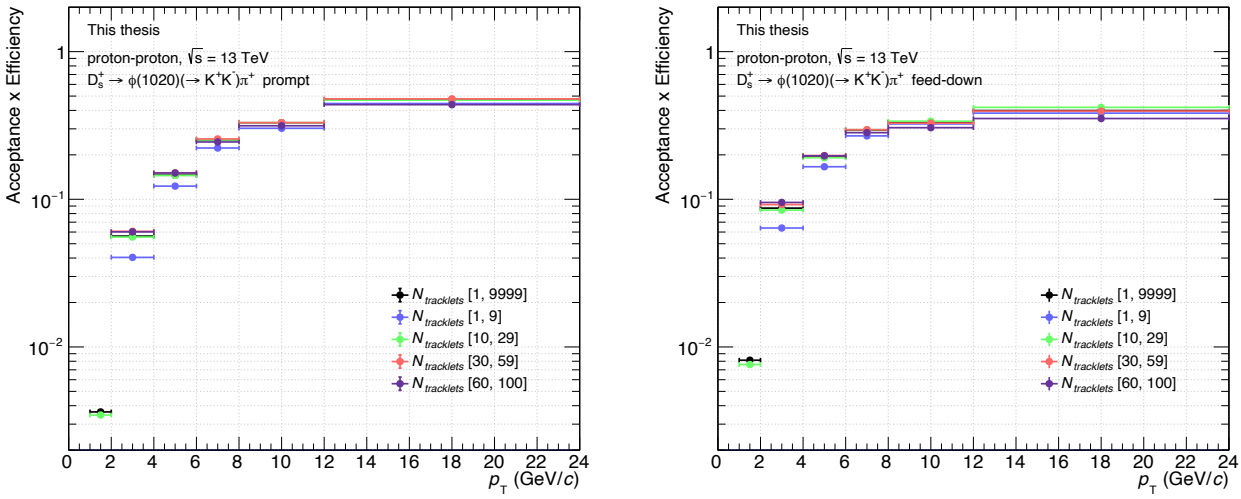


Figure IV.8.5: Acceptance-times-efficiency correction as a function of p_{T}, in different $N_{tracklets}$ multiplicity intervals, of prompt (left panel) and feed-down (right panel) $\mathrm{D_s^+}$ mesons in proton-proton collisions at $\sqrt{s}$ = 13 TeV.

The resulting acceptance-times-efficiency corrections

The acceptance-times-efficiency term as a function of p_{T}, versus multiplicity, for prompt and feed-down $\mathrm{D_s^+}$ mesons are shown in $N_{tracklets}$ intervals on figure IV.8.5 and in V0M percentile intervals on figure IV.8.6. For $p_{\mathrm{T}} < 6$ GeV/c intervals, the efficiencies for the two analyses increases as a function of multiplicity for both prompt and feed-down $\mathrm{D_s^+}$ mesons. At high p_{T}, the efficiencies converge to the exception of the high multiplicity intervals, *i.e.* $N_{tracklets}$ [60, 100] and V0M perc, [0, 0.1] intervals, for which the values are systematically below the lower multiplicity intervals. This is due to the fact that the high multiplicity triggers were available only for a subset of the minimum-bias triggered periods in which the reconstruction efficiency is slightly lower at high p_{T}. The ratio of the prompt $\mathrm{D_s^+}$ mesons efficiency in a given multiplicity interval divided by the minimum-bias interval both obtained in a common set of periods (corresponding to the high multiplicity ones) are shown on the left panel of figure IV.8.7 for the $N_{tracklets}$ analysis and on the right panel of the same figure for the V0M percentiles analysis. In order to reduce the statistical fluctuations due to the limited statistics of low and intermediate multiplicity events in the subset of common periods considered, the model optimal decision value has been lowered. The ratios in the two panels shows an increase of the efficiency as a function of multiplicity with some statistical fluctuations in the $N_{tracklets}$ intervals. The effect seems to saturate at high multiplicity as the resolution on the reconstructed primary vertex, see figure IV.8.7.

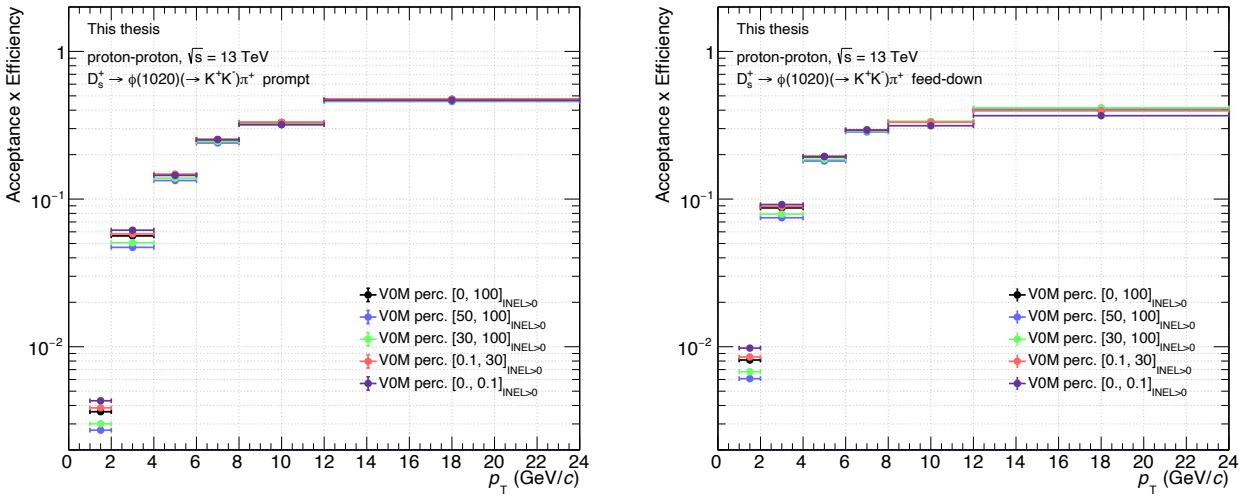


Figure IV.8.6: Acceptance-times-efficiency correction as a function of p_{T}, in different V0M percentile intervals, of prompt (left panel) and feed-down (right panel) $\mathrm{D_s^+}$ mesons in proton-proton collisions at $\sqrt{s} = 13$ TeV

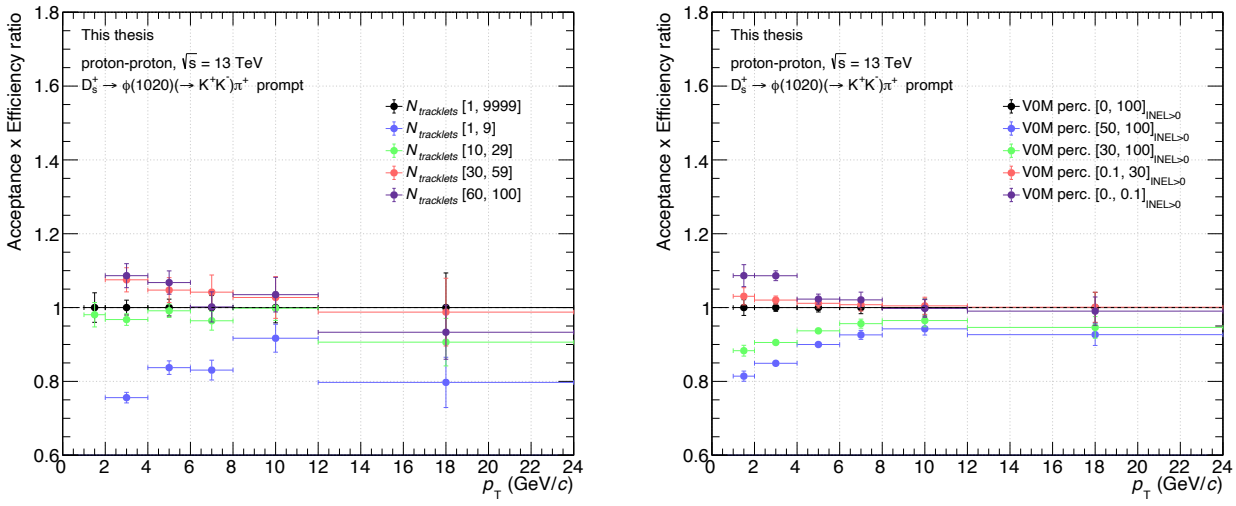


Figure IV.8.7: Acceptance-times-efficiency correction ratio to the minimum-bias efficiency as a function of p_{T}, in different $N_{tracklets}$ (left panel) and V0M percentile (right panel) intervals, of prompt $\mathrm{D_s^+}$ mesons in proton-proton collisions at $\sqrt{s} = 13$ TeV. Common periods of minimum-bias and hight multiplicity triggered event have been used.

IV.9 Fraction of prompt $\mathrm{D_s^+}$ meson

The fraction of prompt $\mathrm{D_s^+}$ mesons in the measured raw yield is estimated with a theory-driven method. The $\mathrm{D_s^+}$ prompt fraction is defined as,

$$\begin{aligned} f_{\mathrm{prompt}} &= 1 - f_{\text{feed-down}} = 1 - \frac{N_{\mathrm{raw}}^{\text{feed-down } \mathrm{D_s^\pm}}}{N_{\mathrm{raw}}^{\mathrm{D_s^\pm}}} \\ &= 1 - \frac{1}{2 \cdot N_{\mathrm{raw}}^{\mathrm{D_s^\pm}}} \cdot \frac{\mathrm{d}^2\sigma}{\mathrm{d}y\mathrm{d}p_{\mathrm{T}}}\Big|_{\text{feed-down}}^{\text{FONLL+PYTHIA8}} \cdot (Acc \cdot \epsilon)_{\text{feed-down}} \cdot c_{\Delta y} \cdot \Delta p_{\mathrm{T}} \cdot \mathrm{B.R.} \cdot \mathcal{L}_{\mathrm{int}} \end{aligned} \tag{IV.20}$$

The $\mathrm{D_s^+}$ feed-down fraction is estimated using FONLL calculations [125] for the beauty hadrons production cross section. The beauty hadron decay kinematics, $\mathrm{H}_b \rightarrow \mathrm{D} + \mathrm{X}$, is done via the PYTHIA8 decayer. The $(Acc \cdot \epsilon)_{\text{feed-down}} \cdot c_{\Delta y}$ is the feed-down $\mathrm{D_s^+}$ efficiency obtained from simulations presented in section IV.8. The main contribution to the total systematic uncertainty on f_{prompt} (error bars on figure IV.9.1) comes from parameter variations in the FONLL calculations [125].

The prompt $\mathrm{D_s^+}$ fraction obtained for the minimum-bias analysis is shown on figure IV.9.1 as a function of p_{T}. Within uncertainties, the fraction is about 85% over the whole p_{T} range. The FONLL predictions are restricted to minimum-bias collisions and are not available as a function of multiplicity. For the computation of the central values of the production corrected yield, the relative contributions from beauty hadron to D mesons are assumed to be constant across multiplicity bins. The f_{prompt} values are then taken from the minimum-bias analysis. The limits of this assumption are discussed by investigating the feed-down fraction evolution as a function of multiplicity with PYTHIA 8 in the dedicated systematic uncertainty study section V.6.

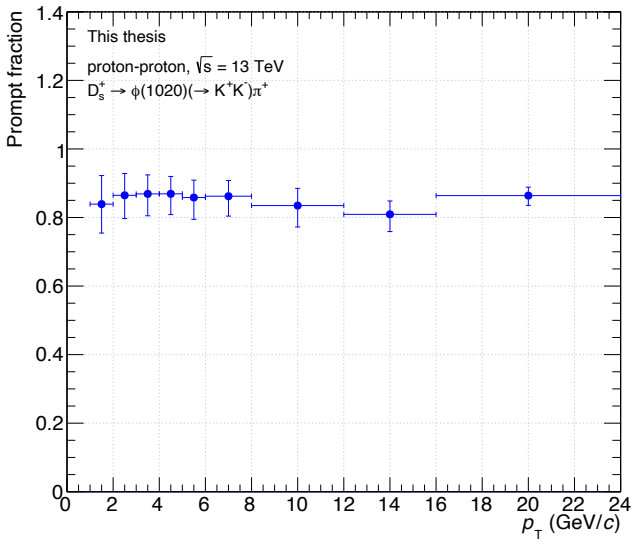


Figure IV.9.1: Prompt D_s^+ fraction as a function of p_T.

IV.10 Cross section and corrected yields

In the cross section, eq. IV.1, and corrected yield formula, eq. IV.2, the factor 1/2 in front of the raw yield quantity is applied to take into account the anti-particle contribution. The D_s^+-D_s^- asymmetry measured by the LHCb experiment in proton-proton collisions is about 0.5% [244], the effect is well below the statistical precision of the measurements presented. The branching ratio for the $D_s^+ \to \phi(1020)(\to K^+K^-)\pi^+$ decay channel is B.R. = 2.24 ± 0.08 % [37].

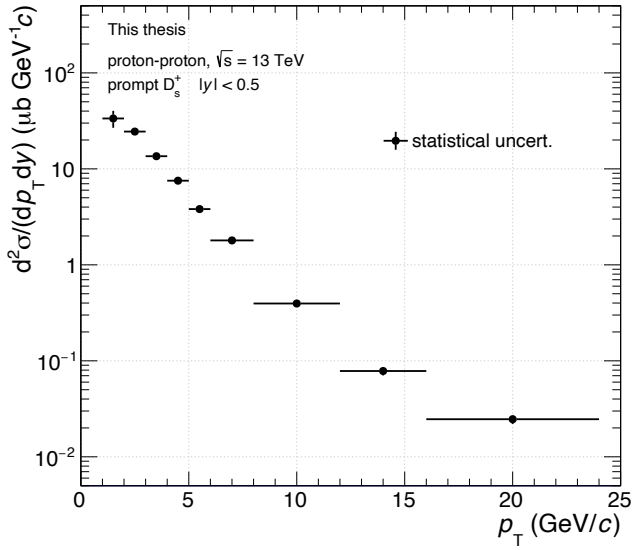


Figure IV.10.1: The transverse momentum-differential production cross section of D_s^+ meson at central rapidity ($|y| < 0.5$) in proton-proton collisions at $\sqrt{s} = 13$ TeV.

The final step to the p_{T}-differential production cross section for $\mathrm{D_s^+}$ meson is the normalisation by the integrated luminosity $\mathcal{L}_{\mathrm{int}} = N_{\mathrm{evt}}/\sigma_{\mathrm{pp}}^{\mathrm{V0}}$ with N_{evt} the number of events triggered by the minimum-bias trigger (section IV.1.1) and $\sigma_{\mathrm{pp}}^{\mathrm{V0}}$ the proton-proton inelastic interaction cross section seen by the V0A and V0C detectors [221]. The minimum-bias trigger is fully efficient for $\mathrm{D_s^+}$ meson in the p_{T} range covered at mid-rapidity $|y| < 0.5$. For the full LHC run II proton-proton collisions at $\sqrt{s} = 13$ TeV data sample, $N_{\mathrm{evt}} = 1.84\ 10^9$, $\sigma_{\mathrm{pp}}^{\mathrm{V0}} = 57.95 \pm 0.93$ mb, $\mathcal{L}_{\mathrm{int}} = 31.7 \pm 0.51$ nb^{-1}. The p_{T}-differential production cross section is shown on figure IV.10.1. The p_{T}-differential production yield per event for $\mathrm{D_s^+}$ meson in different multiplicity intervals are shown on figure IV.10.2. An increase of the $\mathrm{D_s^+}$ production with multiplicity is observed.

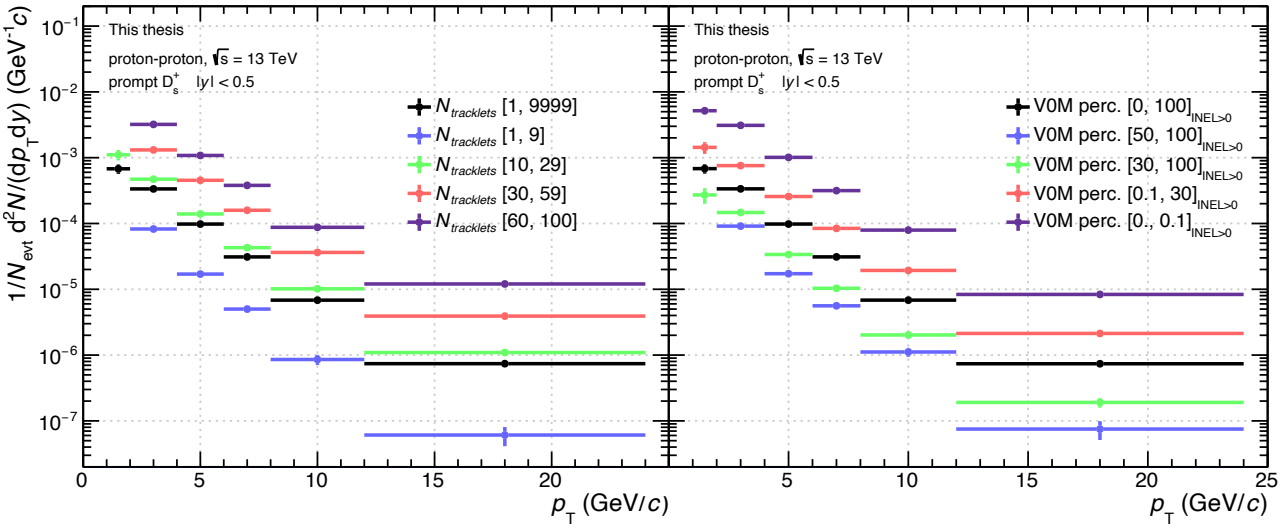


Figure IV.10.2: The transverse momentum-differential production yield per event of $\mathrm{D_s^+}$ meson at central rapidity ($|y| < 0.5$) in $N_{tracklets}$ (left panel) and V0M percentiles (right panel) multiplicity intervals, in proton-proton collisions at $\sqrt{s} = 13$ TeV.

The final comparison of the machine learning (ML) and the standard approach (STD) to the $\mathrm{D_s^+}$ candidate selection is shown on figure IV.10.3 in the form of the corrected yield ratio ML / STD. Each sub-plot of the figure correspond to one $N_{tracklets}$ interval. As mentioned earlier, the ML approach was validated with the $N_{tracklets}$ analysis since this analysis was performed first. The uncertainties associated to the ratio values are the relative statistical uncertainties obtained with the ML approach while on the ratio = 1 line, the error band correspond to the relative statistical uncertainties obtained with the STD approach. The results are compatible within ± 10%, no trend is observed in the ratios as a function of p_{T} and across multiplicity intervals. At low p_{T} and depending on the multiplicity interval, the statistical precision on the measurements is increased (up to a factor 2) by using the ML approach. With this, the method is considered valid as it does not introduced any specific bias with respect to the standard approach.

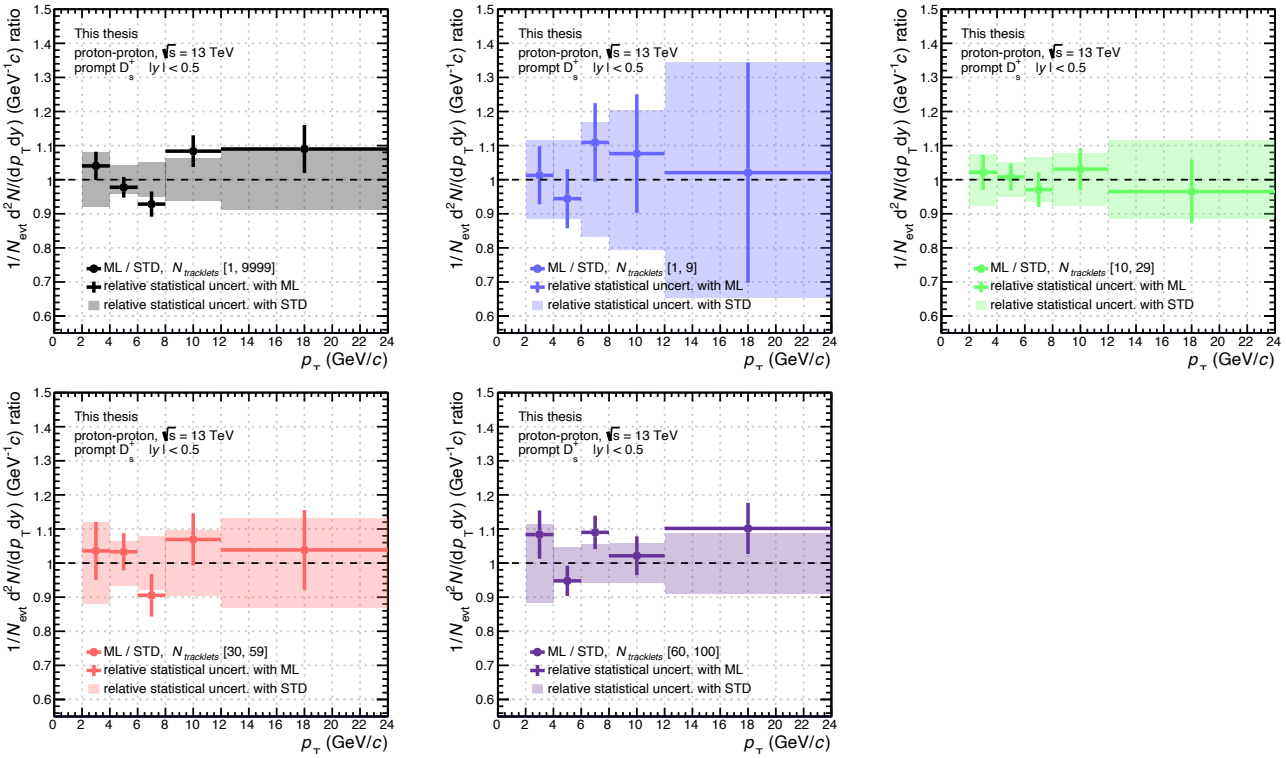


Figure IV.10.3: Ratio of the transverse momentum-differential production yield per event of $\mathrm{D_s^+}$ meson at central rapidity ($|y| < 0.5$) in $N_{tracklets}$ multiplicity intervals, in proton-proton collisions at $\sqrt{s} = 13$ TeV obtained with the standard and machine learning analyses.

Chapter V

Studies of systematic uncertainties

The sources of systematic effects considered for the data analysis are investigated in this chapter. They arise from the strategies adopted at the different steps of the data analysis and are quantified by their related systematic uncertainties. Together with the statistical uncertainties, they reflect the precision of measurements. The following main sources are considered:

- candidate selection efficiency, section V.1
- raw yield extraction, section V.2
- track reconstruction efficiency, section V.3
- Monte Carlo simulation corrections, section V.4
- high multiplicity SPD trigger correction, section V.5
- estimation of the prompt D_s^+ meson fraction, section V.6
- branching ratio and normalisation, section V.7

V.1 Candidate selection efficiency

The measurements are corrected for D_s^+ reconstruction and selection efficiencies, therefore, the precise reproduction of the D_s^+ decay vertex geometry, topological and particle identification variables in Monte Carlo simulations is critical. To quantify the possible impact of discrepancies between the data and Monte Carlo simulations on the final results, the corrected yield obtained with different D_s^+ candidate selection criteria, *i.e.* BDT model decision values, are compared to the central values obtained with nominal parameter values.

The D_s^+ candidate selection is a two step procedure, the procedure of selection-efficiency scan described below is related to the D_s^+ candidate selection performed with the Boosted Decision Trees. First, a pre-filtering procedure is performed using selection criteria on the D_s^+ kinematics, topological and particle identification variables as described in section IV.5.2. No systematic uncertainty is associated to this pre-filtering step as the selection criteria employed are very loose and do not affect the global D_s^+ candidate selection procedure.

A scan of model decision value is performed within a $\pm$ 50% D_s^+ selection efficiency window around the central value defined by the efficiency obtained with the optimal decision value. Let's suppose we want to perform a scan in the decision value range $[a, b]$ around the optimal decision

value $c = a+0.5(b-a)$. With a (b) the lower (higher) decision value limit corresponding to the +50% (-50%) selection efficiency variation (as the decision value is decreased (increased), the efficiency increase (decrease)). By defining the N decision-value variations set by starting from the optimal point and decreasing or increasing in equal steps, i-th step $= c+i\cdot\frac{b-c}{N}$, there is no guaranty that the corresponding D_s^+ efficiency variations will also be uniformly distributed. In order to avoid clusters of low efficiency variations and asymmetric variations in the ± 50% efficiency range explored, the choice of decision values variations is done based on uniform efficiency variations. By doing so, the chance to probe statistical fluctuations between several consecutive decision values biasing the systematic uncertainty estimation is then minimised.

The results of the scan in the minimum-bias data sample for the p_T intervals matching the BDT model training p_T intervals are shown on figure V.1.1 illustrating the variations on the prompt and feed-down D_s^+ efficiencies and the raw yield extracted. On the left panel of the figure, the prompt D_s^+ selection efficiencies variations are presented in the form of ratios to the central efficiency value for the 20 tighter + 20 looser decision values tested. In the last p_T interval, the [0.40, 0.95] decision value range is explored (the optimal point is 0.82) and the resulting efficiency variations stay within ±25%. Given the amplitude of the decision value explored in this interval, the efficiency variations are considered sufficient. The corresponding feed-down D_s^+ selection efficiency variation amplitudes, shown on the middle panel of the figure, are about ±40% at low p_T and slightly increase to reach ±50% at intermediate p_T. For each variation, the invariant mass distributions were fitted requiring a reduced $\chi^2 < 2$ and a statistical significance above 3 as quality criteria. The resulting extracted raw yield ratios on the right panel, present a ±50% variation amplitude at intermediate p_T and decrease to ±40% at low p_T. As mentioned in the data analysis chapter, the BDT decision value for a given p_T interval is kept across multiplicity intervals. By using the same decision value variations from the selection efficiency scan performed in the minimum-bias data sample, no systematic effects are found as a function of multiplicity, the prompt and feed-down efficiencies as the extracted raw yield variations are quite uniformly distributed and the amplitude explored stay within ±10% around the 50% efficiency variation.

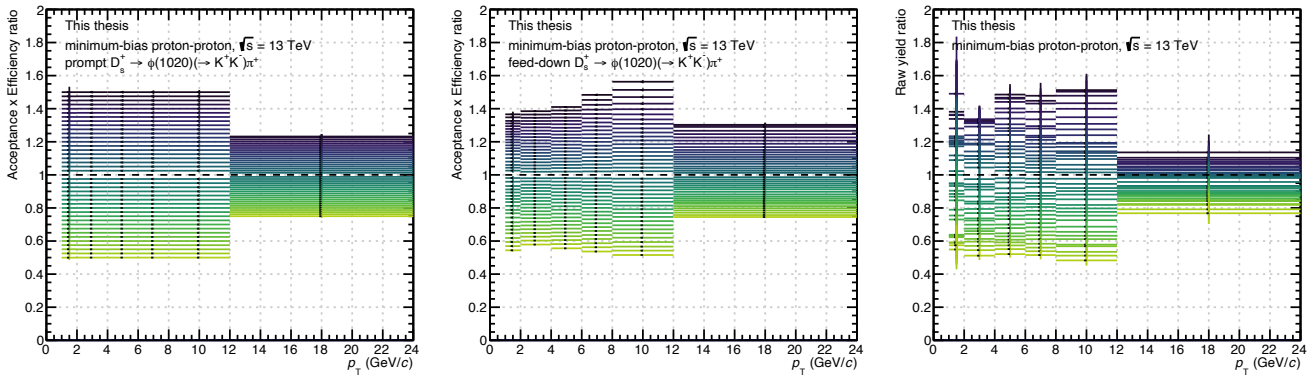


Figure V.1.1: Results of the decision-value scan performed on the prompt (left panel) and feed-down (middle panel) D_s^+ selection efficiencies and the extracted raw yield (right panel) in the different p_T intervals.

The systematic uncertainty related to the candidate selection efficiency is determined from the the cross section variations in the minimum-bias analysis and from the production yield variations for the versus multiplicity analyses. For illustration, the results obtained in the different p_T interval in the minimum-bias analysis are presented on the figure V.1.2 and on figure V.1.3 and V.1.4 for

the $N_{tracklets}$ [10, 29] and V0M percentile [0.1, 30] interval respectively. From low to high decision value number (cut #) the decision value is increasing. In each p_{T} interval, the plotted red line correspond to the central result obtained with the optimal decision value. By projection on the y axis, the distribution of the results (cross section or yield) is obtained from which the mean and root mean square (RMS) values are extracted. The red band associated to the central value is the quadratic sum of the RMS and the shift to the central value. The shift is taken as the difference between the mean value of distribution and the central value. The quadratic sum is a first evaluation of the systematic uncertainty related to the $\mathrm{D_s^+}$ candidate selection. The final systematic uncertainty is refined by applying a smoothing between adjacent p_{T} intervals removing statistical fluctuation contributions, especially at low and high p_{T}. For the multiplicity analyses, the computed systematic uncertainty in low statistics multiplicity intervals are cross checked with high statistics intervals to perform a smoothing across multiplicity.

Each tighter decision cut variation is correlated with the previous ones as they share a common fraction of selected $\mathrm{D_s^+}$ candidates. However, this fraction decreases as the "distance" between two variations increases. No specific treatment of these correlations is done, which is one point to be improved on the computation of the systematic uncertainty.

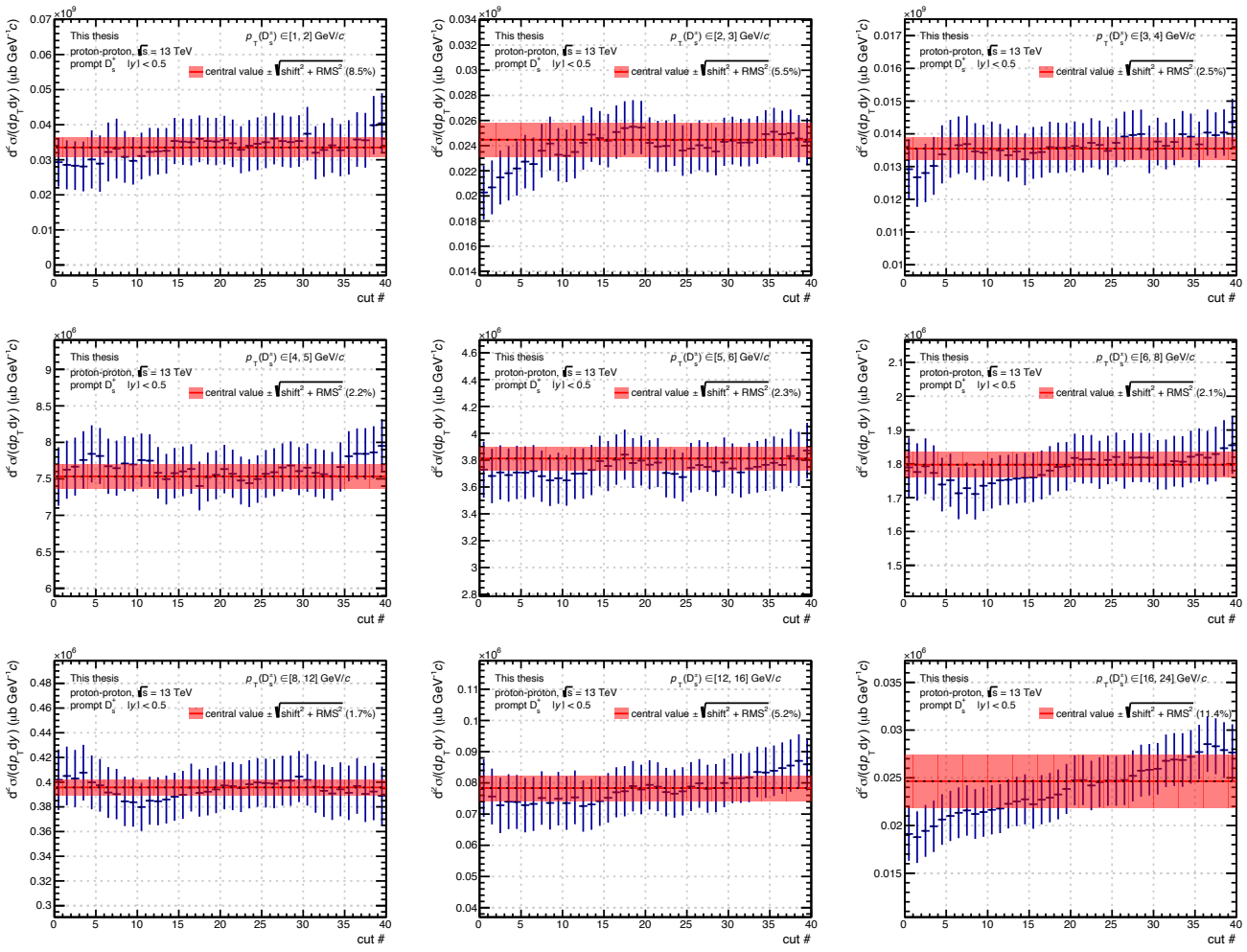


Figure V.1.2: $\mathrm{D_s^+}$ meson production cross section ratio for each decision value variation over the central value in each p_{T} interval considered in the minimum-bias analysis.

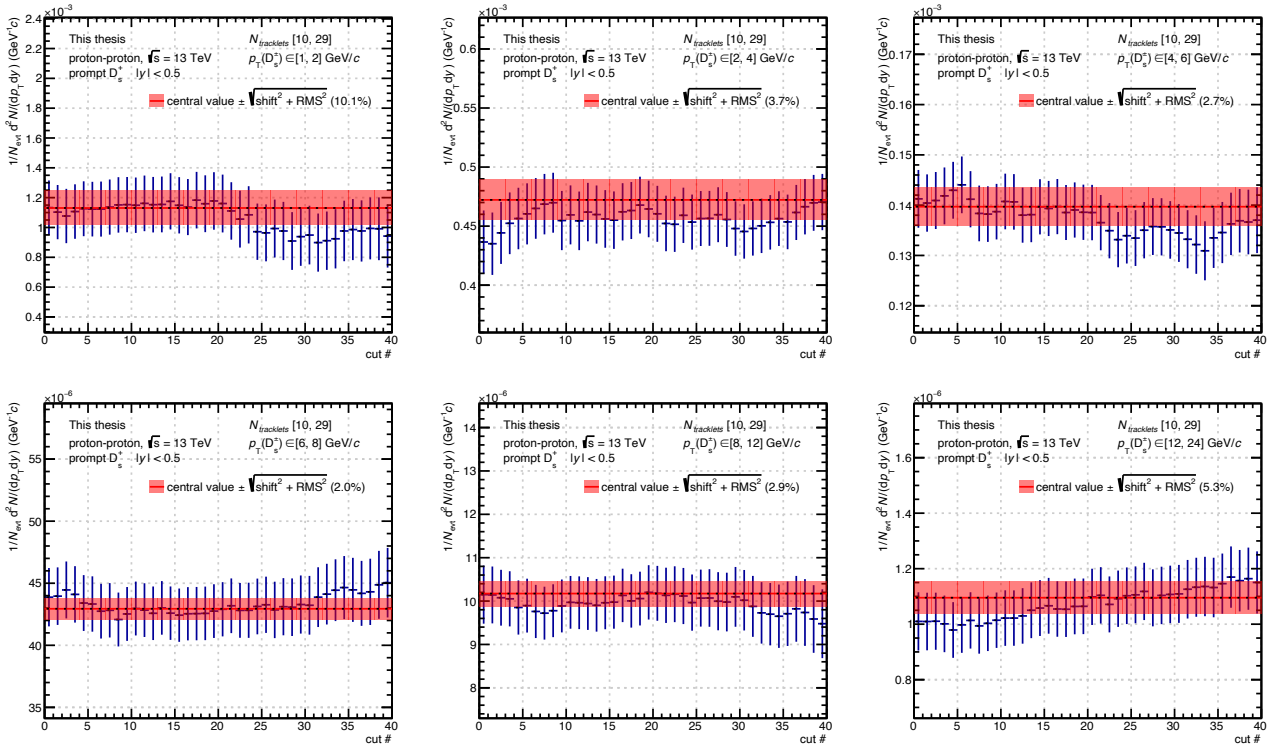


Figure V.1.3: D_s^+ meson production yield ratio for each decision value variation over the central value in each p_T interval considered in the $N_{tracklets}$ multiplicity analysis.

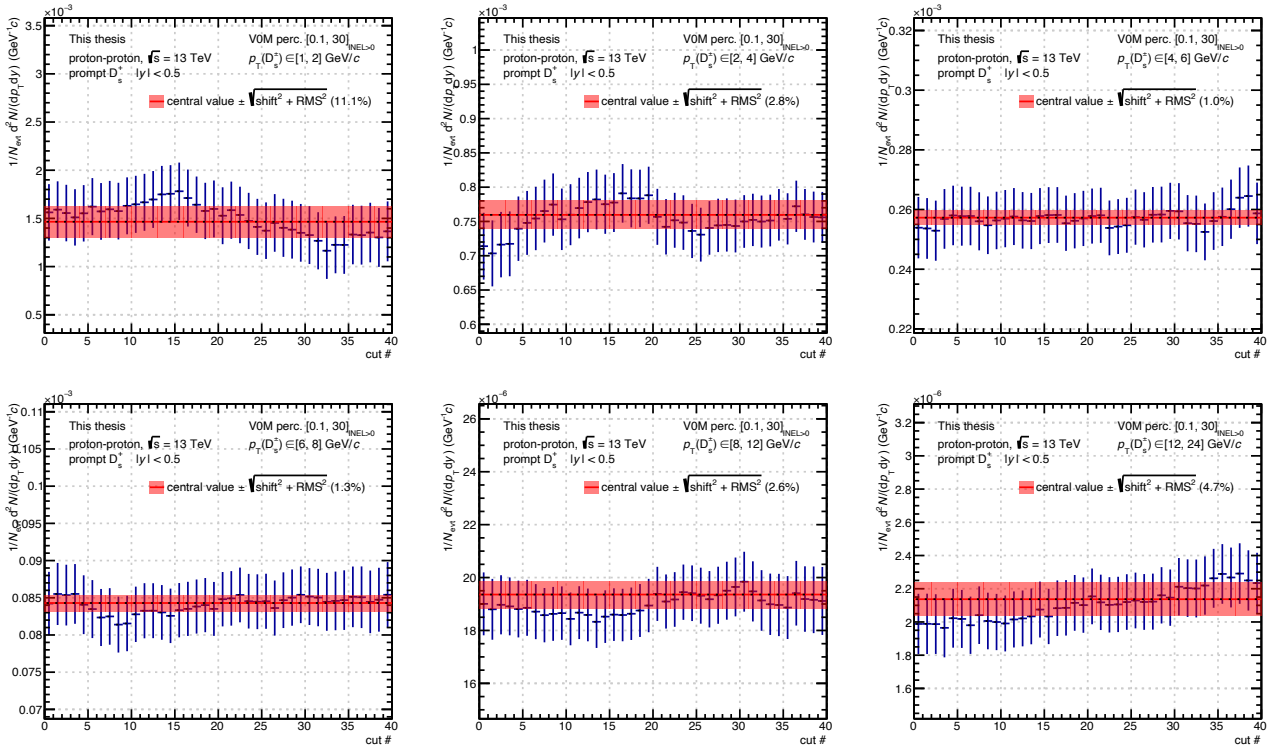


Figure V.1.4: D_s^+ meson production yield ratio for each decision value variation over the central value in each p_T interval considered in the V0M percentile multiplicity analysis.

The systematic uncertainty ranges from 10% to 3% depending on the p_{T} and the multiplicity intervals, they are summarised in the tables V.7.1, V.7.2, and V.7.3 for the different analyses. Highest uncertainty values are found in low and high p_{T} intervals, at intermediate p_{T} the computed cross sections and yield seem to be more stable against decision value variations.

V.2 Raw yield extraction

Sources of systematic effects in the raw yield extraction procedure are explored by performing multiple fits to the invariant mass distributions varying the fit conditions. The following variations, or trials, on the function fitting the background distribution, the fit range and the histogram binning are considered:

- 2^{nd} and 3^{rd} order polynomials and the exponential functions
- 25 variations of the fit range coming from 5 different lower fit limits (from 1.69 to 1.81 GeV/c) and 5 different upper fit limits (from 2.09 to 2.21 GeV/c).
- ± 1 MeV/c^2 bin width variations around the default value.

resulting in 3 x 25 x 3 = 225 fit variations.

The raw yield obtained by integration of the Gaussian part of the total function are cross checked by a bin-counting method. The bin-counting method is used to have yet another estimation of the raw yield independent from the fit of the $\mathrm{D_s^+}$ signal part. It consists in integrating the signal by counting the signal in a $\pm 3\sigma$ and $\pm 5\sigma$ regions around the $\mathrm{D_s^+}$ mass after subtracting the background estimated from a sideband fit of the invariant mass distribution. The integration in $\pm 5\sigma$ is expected to be less sensitive to statistical fluctuations with respect to the integration in $\pm 3\sigma$. For each fit a reduced $\chi^2 < 2.5$ is required as quality criteria.

In the minimum-bias analysis, due to the known discrepancy between the Gaussian mean value of the $\mathrm{D_s^+}$ peak extracted in data and in Monte Carlo simulations (see section IV.7), an additional variation is explored by fixing the Gaussian function mean parameter to the value found in Monte Carlo. A summary of the multi-trial fit procedure is shown on the top panel of figure V.2.1 for the p_{T} [5, 6] GeV/c interval presenting from the left to the right and from top to bottom: the fitted or fixed Gaussian mean, the Gaussian width, the reduced χ^2 of the fit and the extracted raw yield for each of the 2 x 225 fit combinations. The blue, red and green data points correspond to fits performed with the exponential, 2^{nd} and 3^{rd} order polynomials functions respectively. The central values of each quantity are indicated by a magenta line. The raw yield distribution, on the bottom middle plot, is the projection on the y axis of the extracted raw yield per trial plot (bottom left plot). The distribution in black corresponds to the raw yield distribution extracted from the Gaussian function integration, in brown and orange to the 3σ and 5σ bin counting method. The shift of the mean to the central raw yield value and the RMS of the different raw yield distributions are summarised in the bottom right part of the panel.

The same procedure is performed in the multiplicity-dependent analyses. Since in the default fit configuration, the Gaussian width is fixed to the value found in the minimum-bias analysis (see section IV.7), variations where it is left free are envisaged. In total, 225 x $\{(\mu_{\mathrm{Gaus}}$ free, σ_{Gaus} free), (μ_{Gaus} fixed, σ_{Gaus} free), (μ_{Gaus} free, σ_{Gaus} fixed), (μ_{Gaus} fixed, σ_{Gaus} fixed)$\}$ = 900 fit variations are tried. For illustration, the results for the p_{T} [1, 2] GeV/c and p_{T} [12, 24] GeV/c intervals are shown in the middle and bottom panel in the V0M percentile [0.1, 30] and $N_{tracklets}$ [10, 29] multiplicity intervals respectively.

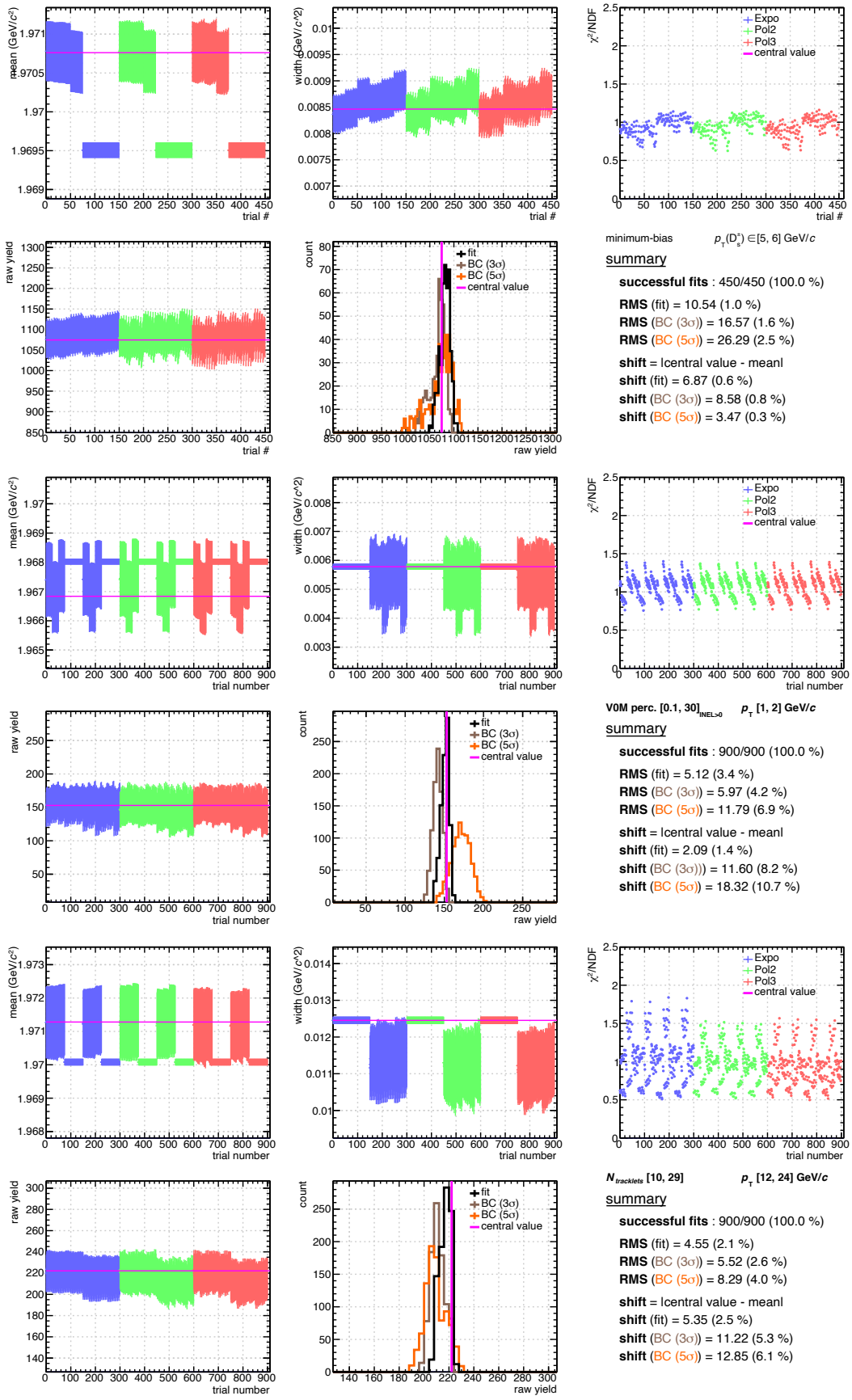


Figure V.2.1: Results of the multi-trial approach in the p_{T} [5, 6] GeV/c in the minimum-bias analysis (top panel), in the p_{T} [1, 2] GeV/c and V0M percentile [0.1, 30] interval (middle panel) and in the p_{T} [12, 24] GeV/c and $N_{tracklets}$ [10, 29] interval (bottom panel).

The systematic uncertainty on the raw yield extraction is based on the quadratic sum of the shift to the central value and the RMS of the raw yield distribution extracted from the total function (background + signal) fit. Depending on the shift of the bin counting distributions with respect to the central value, the uncertainty is enlarged. The systematic uncertainty ranges from 1% to 5% depending on the p_{T} and the multiplicity intervals, lowest values are found at intermediate p_{T} where the raw yield extraction is the most stable. The values are summarised in tables V.7.1, V.7.2, and V.7.3 for the different analyses.

V.3 Track reconstruction efficiency

The track reconstruction efficiency is sensitive to systematic effects related to the track quality selections in the TPC detector and the track propagation from the ITS to the TPC. The systematic uncertainties are estimated in the minimum-bias analysis as a function of p_{T} and considered independent of the multiplicity of the collision. Therefore, the same computed values are assigned in the multiplicity-dependent analyses.

The variations (with respect to the default criteria described in section IV.4.1) of track selections in the TPC are listed below:

- a ratio of the number of crossed pad rows over the number of findable clusters associated given the track geometry > 0.9 (default value 0.8).
- additional selection: number of TPC clusters $> 0.65 \cdot$ number of TPC crossed rows.
- additional selection: number of TPC crossed rows $>$ 120 - 5(GeV/c)/p_{T}.

The production cross section is re-evaluated from the three different selection criteria and the corresponding variation is found to be 3%. For the case of the decay $\mathrm{D}^0 \rightarrow \mathrm{K}^-\pi^+$, where the statistics is higher, the same procedure lead to a 2% variation of the cross section. Since the variation between a two-body and three-body D-meson decay differs by one percent, we attribute a 1% systematic uncertainty related to the quality selections in the TPC detector per $\mathrm{D}_{\mathrm{s}}^+$ decay track independently of the p_{T} interval considered.

Potential systematic effects in the track propagation from the ITS to the TPC detector are evaluated from the comparison of the track matching efficiency in data and general purpose Monte Carlo simulations (*i.e.* not enriched with heavy quark pairs). The track matching efficiency corresponds to the fraction of reconstructed tracks in both the ITC and the TPC with at least a cluster in one of the two SPD layers divided by the total number of reconstructed tracks with clusters in the TPC. The matching efficiencies are estimated for primary tracks, *i.e.* charged particles with a mean proper lifetime $\tau > 1$ cm/c that are produced in beam-beam interactions or come from the decays of particles with $\tau < 1$ cm/c [235], and secondary tracks as track coming from strange hadron decay or produced in interactions with detector materials. As the matching efficiencies for primary and secondary tracks are expected to be different, the proportion of primary tracks in Monte Carlo simulations is reweighed to match the one found in the data in order to avoid any bias in the estimation. The matching efficiency in Monte Carlo simulations is expressed as,

$$\epsilon^{\mathrm{MC}}_{\mathrm{inclusive}} = f^{\mathrm{data}}_{\mathrm{primary}} \cdot \epsilon^{\mathrm{MC}}_{\mathrm{primary}} + (1 - f^{\mathrm{data}}_{\mathrm{primary}}) \cdot \epsilon^{\mathrm{MC}}_{\mathrm{secondary}} \tag{V.1}$$

with $f^{\mathrm{data}}_{\mathrm{primary}}$ the fraction of primary track estimated by fitting the impact parameter distribution of tracks in data using template fits for the primary and secondary track contributions determined

in Monte Carlo simulations. More information about the procedure can be found in section 4.8.4 of [245]. The systematic uncertainties on the track-matching efficiency is taken as the relative difference between the values found in data and Monte Carlo simulations, $(\epsilon^{\text{data}}_{\text{inclusive}} - \epsilon^{\text{MC}}_{\text{inclusive}})/\epsilon^{\text{data}}_{\text{inclusive}}$. They are determined as a function of the track p_{T} in dedicated analyses performed inside the ALICE collaboration. In the analysed periods of the proton-proton collisions at $\sqrt{s} = 13$ TeV data sample, the averaged systematic uncertainty ranges from 2% for $p_{\text{T}} \sim$ 1-2GeV/c to 3% for $p_{\text{T}} > 8$ GeV/c.

The uncertainties on the decay tracks are propagated to the D_s^+ meson level using Monte Carlo simulations reproducing the D_s^+ meson decay kinematics. The 2D correlation distribution of the p_{T} of the daughter tracks with p_{T} (D_s^+) is shown on the left panel of figure V.3.1. The resulting systematic uncertainty on the matching efficiency is shown on the right panel of figure V.3.1 ranging from 5.25% at low p_{T} (D_s^+) to 7.8% at high p_{T} (D_s^+).

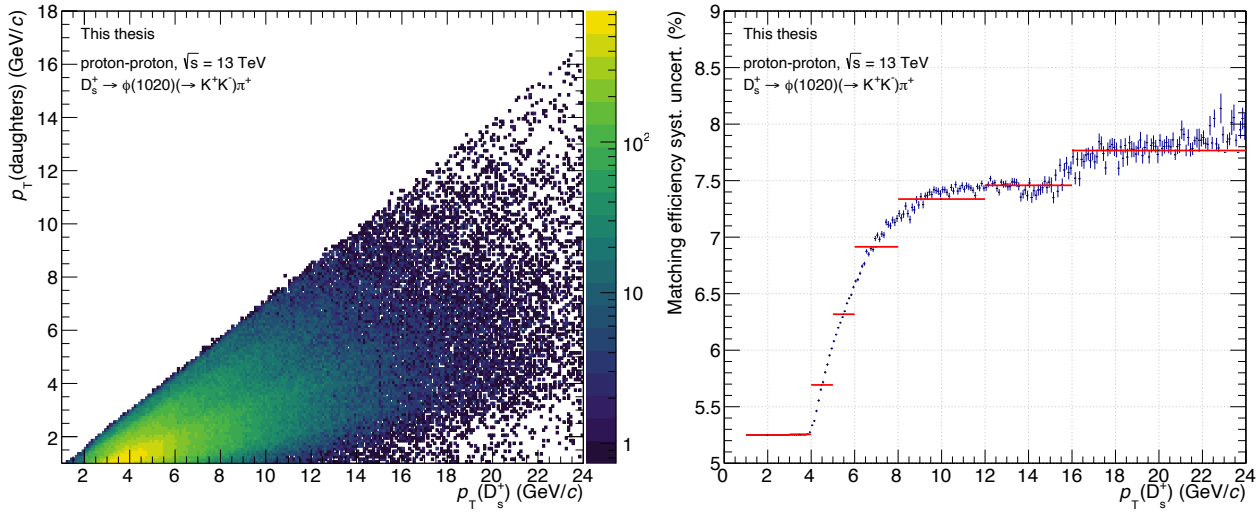


Figure V.3.1: Left panel: Distribution of the p_{T} of the daughter tracks as a function of the D_s^+ meson p_{T}. Right panel: Systematic uncertainty (%) related to the matching efficiency as a function of the D_s^+ meson p_{T}.

The total systematic uncertainties on the track reconstruction efficiency, going from 6% to 8% with increasing p_{T}, are obtained by summing in quadrature the quality selection criteria and the matching efficiency systematic uncertainties. The values are summarised in tables V.7.1, V.7.2, and V.7.3 for the different analyses.

V.4 Monte Carlo simulation corrections

In this section, additional sources of systematic effects related to the acceptance and efficiency corrections are explored. They complete the D_s^+ candidate selection efficiency systematic study done in section V.1. The p_{T} shape of the D_s^+ spectra, the multiplicity re-weighting procedure, the z_{vertex} distributions and considerations on the statistical precision of the efficiency are discussed in the following paragraphs.

V.4.1 Generated p_{T} shape

The strong p_{T} dependence of the $\mathrm{D_s^+}$ candidate reconstruction and selection efficiency (see section IV.8) makes the measurement sensitive to the $\mathrm{D_s^+}$ meson p_{T} spectrum generated in Monte Carlo simulations. The simulations used to compute the efficiency correction are performed with PYTHIA 8.2 as described in section IV.3. Due to the finite width of the p_{T} intervals in the analysis, a non-realistic p_{T} shape of the generated $\mathrm{D_s^+}$ meson could introduce a bias in the variations of the efficiency within a given p_{T} interval.

Another hypothesis on the shape of the $\mathrm{D^0}$ meson p_{T}-spectrum is obtained using FONLL calculations. Predictions of the $\mathrm{D_s^+}$ production cross section are not available in FONLL. The prompt $\mathrm{D_s^+}$ spectrum p_{T} shape in PYTHIA is therefore compared to the p_{T} shape of the prompt $\mathrm{D^0}$ meson from FONLL predictions which describe well the p_{T} shapes of the measured D-meson cross sections. For the feed-down $\mathrm{D_s^+}$ mesons, the p_{T} shape is obtained from predictions computed with the same procedure as described in section IV.9 using the FONLL calculations for the beauty hadrons production cross section. The p_{T} spectra are shown on the left panel of figure V.4.1. The ratios of the FONLL distributions to the PYTHIA 8 distributions for prompt and feed-down $\mathrm{D_s^+}$ mesons are shown on the right panel.

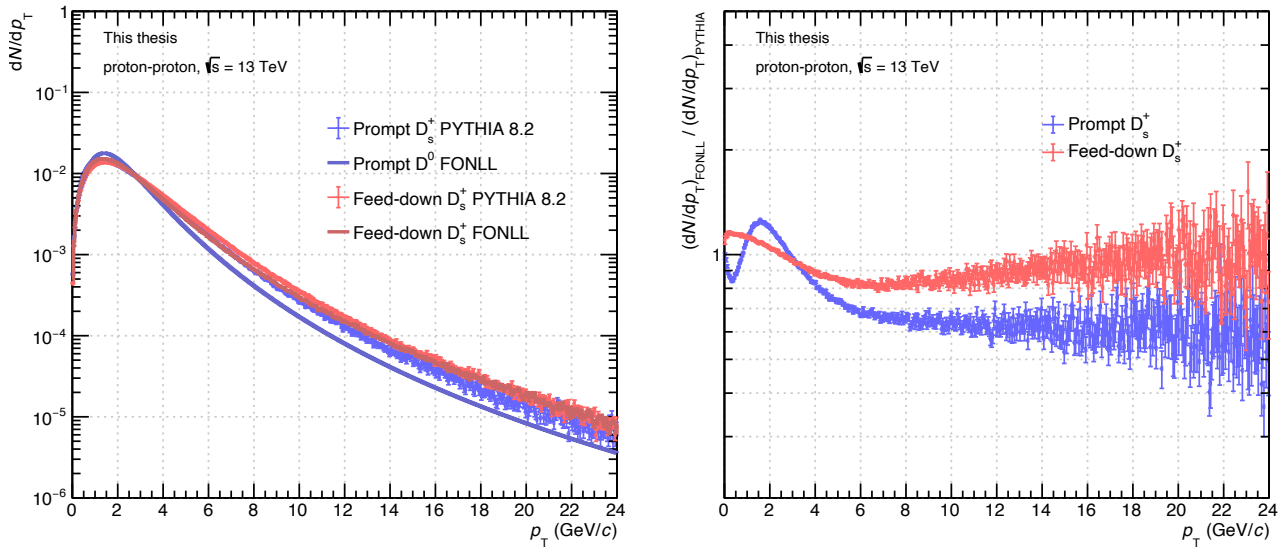


Figure V.4.1: Left panel: Normalised p_{T} spectra of the prompt and feed-down $\mathrm{D_s^+}$ mesons in PYTHIA 8 simulations and from FONLL calculations. Right panel: Ratio between the FONLL spectra over the PYTHIA 8 spectra.

The systematic uncertainty is derived by re-weighting the p_{T} spectra in the Monte Carlo simulations to match the one from the FONLL predictions and comparing the result obtained with the one using the PYTHIA 8.2 p_{T} shape. The ratio of the production cross section with the two different p_{T} shape hypotheses is shown on the left panel of figure V.4.2. The same check is performed as a function of multiplicity for which some of the p_{T} intervals are broader, the production yield ratios are shown on the middle panel of figure V.4.2 for the $N_{tracklets}$ analysis and on the right panel for the V0M percentile analysis. Only the central value of the ratios are considered for the estimation of the systematic uncertainty. The systematic effects related to the change of generated $\mathrm{D_s^+}$ mesons p_{T} shape are mostly seen in the p_{T} 2 - 6 GeV/c region where the p_{T} weights vary steeply. In the multiplicity analyses, no strong multiplicity-dependent effect is

observed. By comparing the results obtained in the minimum-bias and multiplicity analyses, the $N_{tracklets}$ >0 multiplicity selection seems to cause an increase of the discrepancy in the p_{T} [2, 4] GeV/c interval.

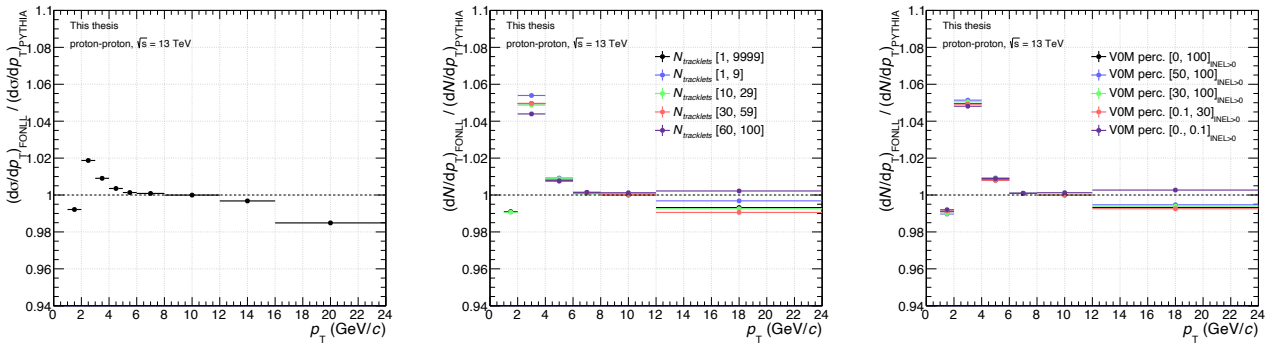


Figure V.4.2: Ratio of the production cross section (left panel) and production yield (middle and right panel) considering the two $\mathrm{D_s^+}$ p_{T}-spectra hypothesis in PYTHIA simulations and from FONLL calculations.

No systematic uncertainties is assigned in the intermediate p_{T} region, it reaches a maximum of 2% in the p_{T} [2, 3] GeV/c interval in the minimum-bias analysis and a maximum of 5% in the p_{T} [2, 4] GeV/c interval of the multiplicity analyses. The values are summarised in tables V.7.1, V.7.2, and V.7.3 for the different analyses.

V.4.2 Multiplicity weights

The default strategy for the multiplicity re-weighting of the efficiencies in the different analyses is to compute the weights from the ratio of the multiplicity distribution in data and Monte Carlo simulations requiring a selected $\mathrm{D_s^+}$ candidate in the event having a reconstructed invariant mass within a ± 20 MeV window around the $\mathrm{D_s^+}$ meson mass [37]. This requirement is applied to allow for a fair comparison of the distributions since the simulated events are enriched in $c\bar{c}$ and $b\bar{b}$ pairs (see section IV.3).

To study the potential systematic effects related to this choice, the multiplicity weights were recomputed removing the requirement on the invariant mass window on the selected $\mathrm{D_s^+}$ candidates in the events. The systematic uncertainty is determined by comparison of the results obtained between the two re-weighting strategies, taking the form of production cross section and production yield ratios. The results are shown on the left panel of figure V.4.3 for the minimum-bias analysis. The removal of the invariant mass requirement provoke a decrease of the efficiency directly translated to the cross section since the other terms of the cross section equation are equivalent between the two re-weighting strategies. The effects is of the order of 2% at low p_{T} and decreases as the p_{T} increases. Due to the small efficiencies at low p_{T} it is expected that the re-weighting procedure affects more this region.

A multiplicity dependence of the effect is shown on the middle and right panels of figure V.4.3 where the same comparative ratios were computed in the different multiplicity intervals. In both multiplicity analyses, the change of re-weighting strategy affects more the low multiplicity intervals. In the $N_{tracklets}$ analysis, where higher mean value of the charged particle multiplicity is reached with respect to the V0M percentile analysis, negligible deviations from unity of the yield ratios are found in the the two highest multiplicity intervals. By comparison with the minimum-bias analysis, the lowest multiplicity intervals are more affected, resulting in a deviation from 2% to

3% at low p_{T}. The deviation of the ratios from unity is assigned as the corresponding systematic uncertainty, with the values summarised in tables V.7.1, V.7.2, and V.7.3 for the different analyses.

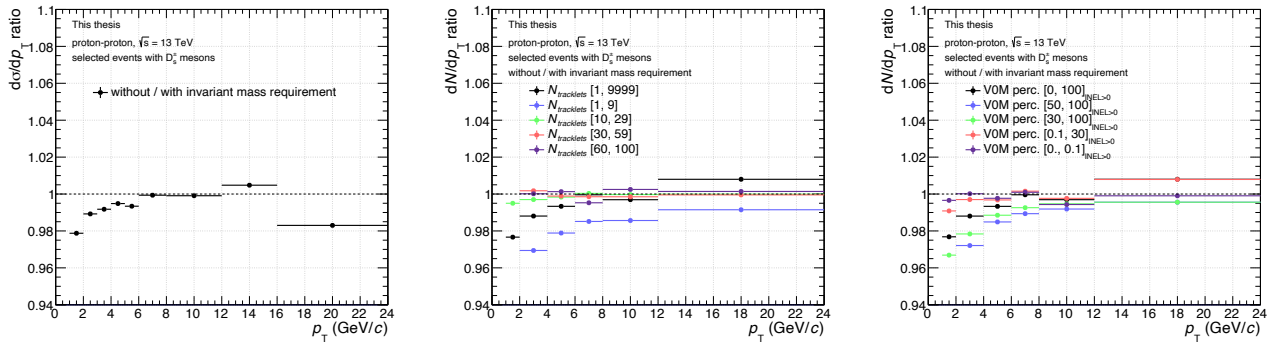


Figure V.4.3: Ratio of the production cross section (left panel) and the production yield (middle and right panels) with the efficiency correction re-weighted without and with the invariant mass window requirement.

V.4.3 Distribution of the primary vertex position

The multiplicity weights computed to correct the efficiencies are based on the uncorrected $N_{tracklets}$ distributions in data and Monte Carlo simulations. This choice is motivated by the possible correlation of the $\mathrm{D_s^+}$ meson reconstruction and selection efficiency between the longitudinal positions of the primary vertex z_{vertex} position and the multiplicity of the event. The z_{vertex} equalisation procedure corrects the multiplicity of the event depending on its z_{vertex} position and the correction is more important for event with $z_{vertex} < 0$ values (see section IV.2.1). By using multiplicity weights based on equalised $N_{tracklets}$ distributions, the z_{vertex} dependence of the efficiency is compensated. A bias could be introduced in the efficiency estimation by systematically re-weighting events with a given $\mathrm{D_s^+}$ meson selection efficiency in low or high multiplicity regions.

The discrepancies between the z_{vertex} distributions in data and Monte Carlo for events with a selected $\mathrm{D_s^+}$ candidate are studied with the ratio of the z_{vertex} distributions. The results obtained in the $N_{tracklets}$ analysis are shown on the top panels of figure V.4.4 for the different $N_{tracklets}$ intervals. The three panels of the figures correspond to the 2016, 2017 and 2018 data samples. No significant deviation from unity is present in the distribution ratio of the $N_{tracklets} > 0$ interval, however for narrower event multiplicity selections, a trend in the ratios is observed. The data distributions are overestimated in the $z_{vertex} < 2$ cm region and underestimated the $0 < z_{vertex} < 2$ cm region in Monte Carlo simulations. The results for V0M percentile analysis are shown on the bottom panels of figure V.4.4 where the selection in multiplicity intervals in only done in data since the V0M percentiles are not defined in the Monte Carlo simulations. As mentioned in section IV.3, in the simulated Monte Carlo sample associated to the high multiplicity V0 triggered data sample, a multiplicity threshold was on the number of charged particle produced in the central region ($|\eta| < 1$). The effect of two different selection strategies employed is directly reflected on the z_{vertex} distribution ratio of the high multiplicity interval (V0M percentile $[0, 0.1]_{\mathrm{INEL>0}}$) deviating from the rather uniform distribution ratios in the lower multiplicity intervals.

The z_{vertex} distribution ratios are used to correct the discrepancies between the Monte Carlo simulated and real data. The systematic uncertainties are determined from the ratio of the production yields with and without the z_{vertex} re-weighted efficiency. The results are shown on the left

panel of figure V.4.5 for the $N_{tracklets}$ analysis and on the right panel for the V0M percentile analysis. In the $N_{tracklets}$ analysis, to the exception of the $N_{tracklets} > 0$ interval, a global 1% systematic uncertainty is assigned in each p_{T} and multiplicity interval. In the V0M percentile analysis, no effect is observed due to the z_{vertex} re-weighting procedure in the minimum-bias data sample. A 1% systematic uncertainty is assigned in the high multiplicity interval due to the clear deviation from the central values.

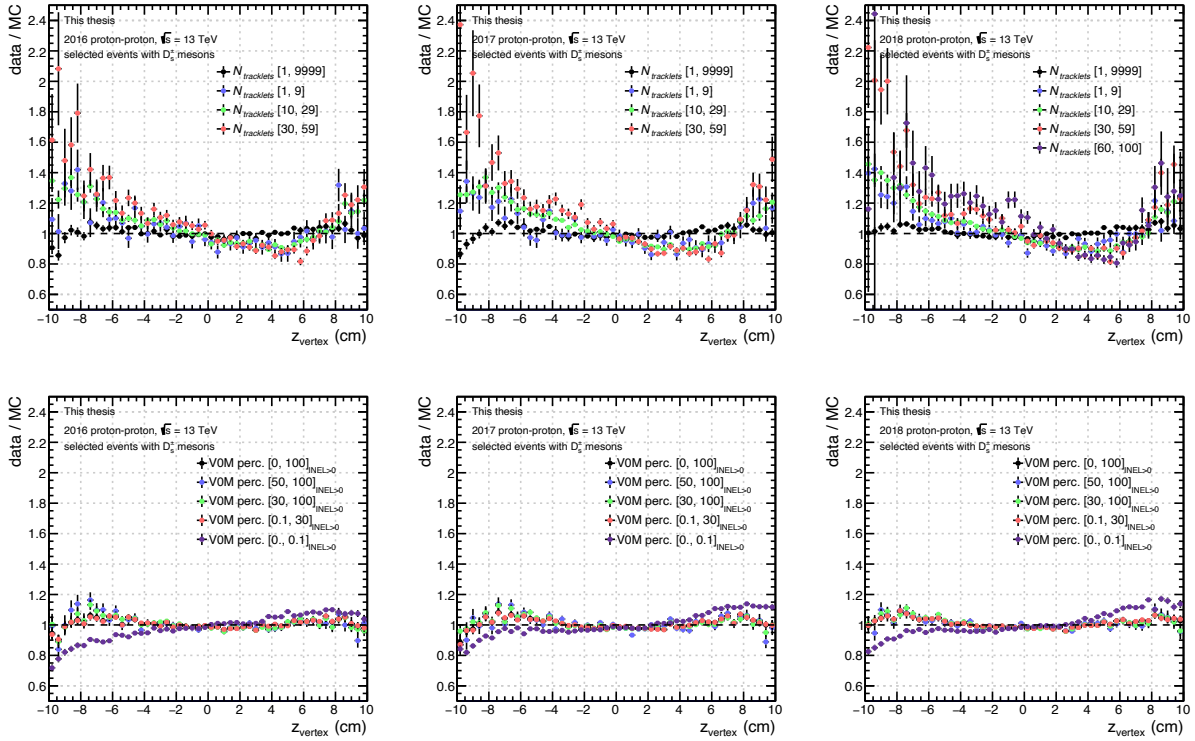


Figure V.4.4: Ratio of the z_{vertex} distribution of events with a selected $\mathrm{D_s^+}$ meson in data over the distribution in Monte Carlo simulations in the 2016, 2017 and 2018 data samples for the $N_{tracklets}$ analysis (top panels) and V0M percentile analysis (bottom panels).

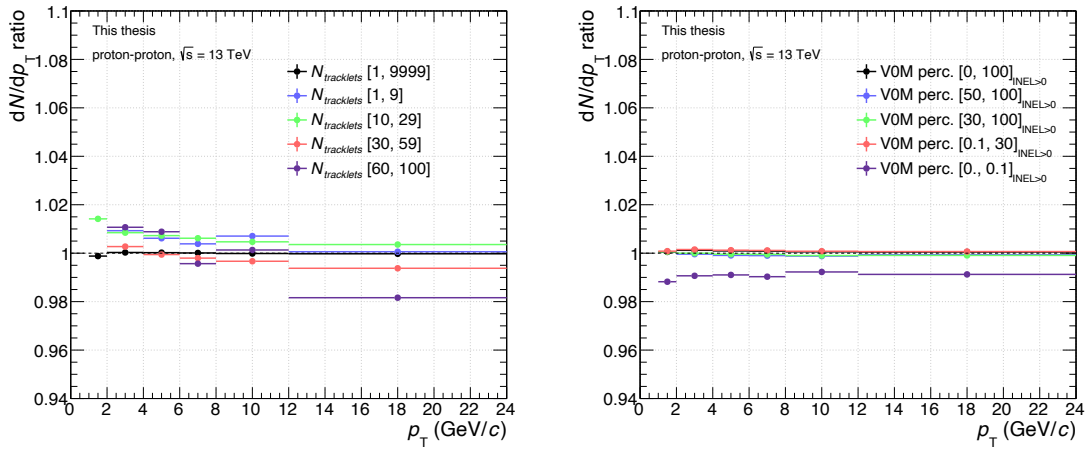


Figure V.4.5: Ratio of the production yields with a z_{vertex} re-weighted efficiency over the non re-weighed production yield in the different $N_{tracklets}$ (left panel) and V0M percentiles (right panel) multiplicity intervals.

V.4.4 The size of simulated data sample

An additional systematic uncertainty was assigned due to the limited statistics of the Monte Carlo simulated samples. On the left panel of figure V.4.6 the relative uncertainty on the prompt D_s^+ efficiencies are shown for the two high multiplicity intervals, $N_{tracklets}$ [60, 100] and V0M percentile $[0, 0.1]_{\mathrm{INEL>0}}$, for which dedicated Monte Carlo samples with limited statistics were produced. For comparison, the relative uncertainties are also shown in other $N_{tracklets}$ intervals. In the $N_{tracklets}$ analysis, unlike the V0M percentile analysis, a selection in multiplicity intervals is performed in Monte Carlo simulations. As a result, the statistical uncertainty on the efficiencies in low $N_{tracklets}$ intervals is higher with respect to low multiplicity V0M percentile intervals.

The statistical precision of the production yields are computed considering only the statistical uncertainties of the extracted raw yields in data. They are shown on the right panel of figure V.4.6. The statistical uncertainty of the prompt D_s^+ efficiencies, is quite sizeable, especially at low and high multiplicity and in the p_T [1, 2] GeV/c interval. Therefore, the relative statistical uncertainty on the prompt D_s^+ efficiencies are directly assigned as systematic uncertainty on the simulation sample sizes.

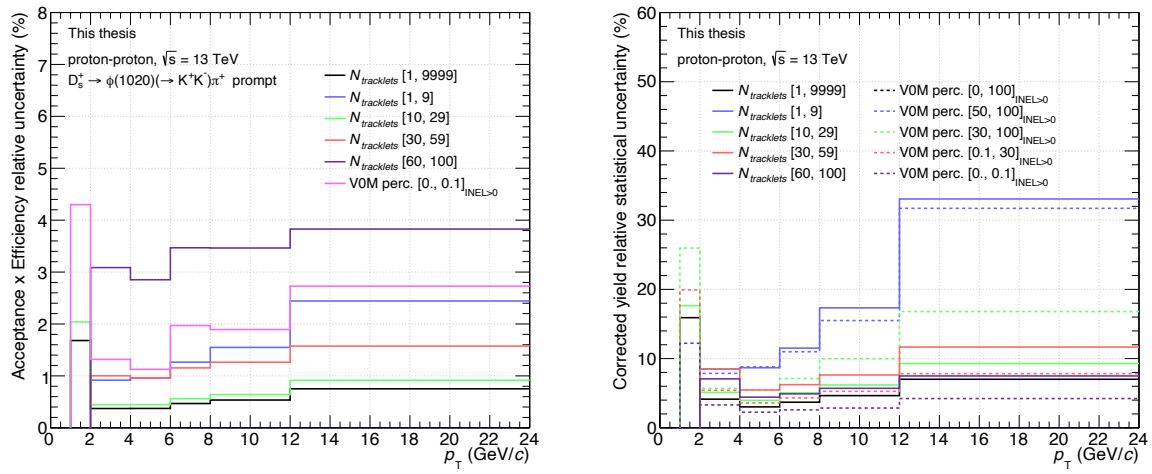


Figure V.4.6: Left panel: Relative statistical uncertainty of the prompt D_s^+ acceptance x efficiency in the different $N_{tracklets}$ intervals and in the high multiplicity V0M percentiles interval. Right panel: Relative statistical uncertainty of the production yield in the different multiplicity intervals.

V.5 High multiplicity SPD trigger correction

The strategy for the high multiplicity SPD trigger correction is to re-weight events based on a data driven estimated turn-on curve. In the default case, the HMSPD trigger turn-on curve is defined by the ratio of the $N_{tracklets}$ distribution of events having a selected D_s^+ candidate triggered by the HMSPD trigger over the corresponding minimum-bias triggered distribution. The turn-on curve is then normalised to saturate at unity at high multiplicity ($N_{tracklets} > 80$).

Sources of biases in the HMSPD trigger correction can originate from the selection criteria of events used to compute the turn-on curve and from the normalisation procedure. Variations from the central selection procedure are done by requiring that the selected D_s^+ has its invariant

mass within a $\pm$ 20 MeV/c^2 invariant mass window around the PDG value [37] for different BDT decision values: 0.4 and 0.7 (constant across p_T intervals) and the central decision values reported table IV.5.6. New weights are computed for each selection variations.

By re-weighting events, the trigger correction modifies the number of selected events used to normalise the production yields and the number of selected D_s^+ candidates extracted from invariant mass histogram fits. Variations up to 10-11% on the number of selected events and extracted raw yields are found by comparing the quantities corrected with weights computed from different selection strategies. Different strategies for the turn on curve normalisation were explored: normalisation to the $N_{tracklets} = 80$ value, to the value found with a the constant fit of the high multiplicity part of the turn on curve for different ranges, $N_{tracklets}$ [65, 100] and [80 ,100]. However the effects on the production yields are well below the percent level.

The variations obtained on the production yields due to the different trigger corrections applied are constant across p_T intervals. A maximum deviation of 3% is observed in the production yield ratio to the default case, and so this value is assigned as systematic uncertainty on the high multiplicity trigger correction procedure.

V.6 Fraction of prompt D_s^+ mesons

As briefly described in section IV.9, the main contribution to the systematic uncertainty on the prompt D_s^+ meson fraction in the minimum-bias analysis comes from variations in the parameters of FONLL calculations for the production of beauty hadrons. In the multiplicity-dependent analyses, a constant prompt D_s^+ mesons fraction as a function of multiplicity is assumed. In the two next sections, more details are provided regarding the systematic uncertainty coming from FONLL calculations and from the assumption of a constant fraction of prompt D_s^+ meson.

Systematic uncertainties from FONLL calculations

The factorisation μ_F and renormalisation μ_R scales are essential parameters of perturbative QCD calculations of open heavy flavour hadron production (see section II.2). In FONLL [125], the central predictions are computed with $\mu_R = \mu_F = \mu_0 = \sqrt{p_{T,Q}^2 + m_Q^2}$ with $p_{T,Q}$ the transverse momentum and m_Q the mass of the heavy quark considered, being here the beauty quark. Independent variations of the two scales are performed in the $0.5 \leq \mu_{R,F}/\mu_0 \leq 2$ interval and the envelope of the variations is taken as systematic uncertainty. The second source of systematic effect is the value of bottom quark mass which is a free parameter of the calculation. Three bottom quark mass values are considered, m = 4.75, m = 4.5 and 5 GeV/c^2. The envelope of the computed prediction variations keeping $\mu_{R,F}/\mu_0 = 1$ is used as systematic uncertainty. The final contribution is the uncertainties coming from the CTEQ6.6 [246] parton density functions (see section II.2.1). The total systematic uncertainty on the predictions of the beauty hadrons production cross section is taken as the quadrature sum of the three mentioned sources [125]. Combined with other systematic uncertainties from the terms in the expression for $f_{\rm prompt}$ IV.20, the total uncertainties on $f_{\rm prompt}(p_T)$ ranges from 11% to 3% depending on the transverse momentum interval. The values for the minimum-bias analysis are summarised in table V.7.1.

Fraction of prompt D_s^+ as a function of multiplicity

The multiplicity dependence of the prompt D_s^+ fraction is studied via the feed-down D_s^+ fraction evolution as a function of the density of charged particle in collisions simulated with

the PYTHIA 8.2 Monte Carlo generator. The Monash 13 tune [136] and the colour reconnection beyond leading colour (CR-BLC) tunes [118] are considered for this study. The results are presented on figure V.6.1 with the evolution of the feed-down fraction of charmed mesons (left panel) and D_s^+ meson (right panel) divided to the INEL>0 value as a function of the normalised charged particle density in proton-proton collisions at $\sqrt{s} = 13$ TeV. The differences between the Monash 13 and the CR-BLC tunes lies in the different consideration and parameterisations of a string fragmentation model describing the interactions between colour sources in the final state, the colour reconnection, in dense partonic environment as high multiplicity collisions. The Monash 13 and CR-BLC Mode 0, both predictions show an increasing feed-down fraction to the total charm production with increasing multiplicity. For multiplicities above 2 times the mean charged particle density $\langle \mathrm{d}N_{\mathrm{ch}}/\mathrm{d}\eta \rangle$, the CR-BLC Mode 3 predict a decreasing feed-down fraction. The different modes of the CR-BLC tune provide different restrictions on the formation of the colour reconnections [118]. The predictions are similar between the charmed meson and D_s^+ meson feed-down fractions.

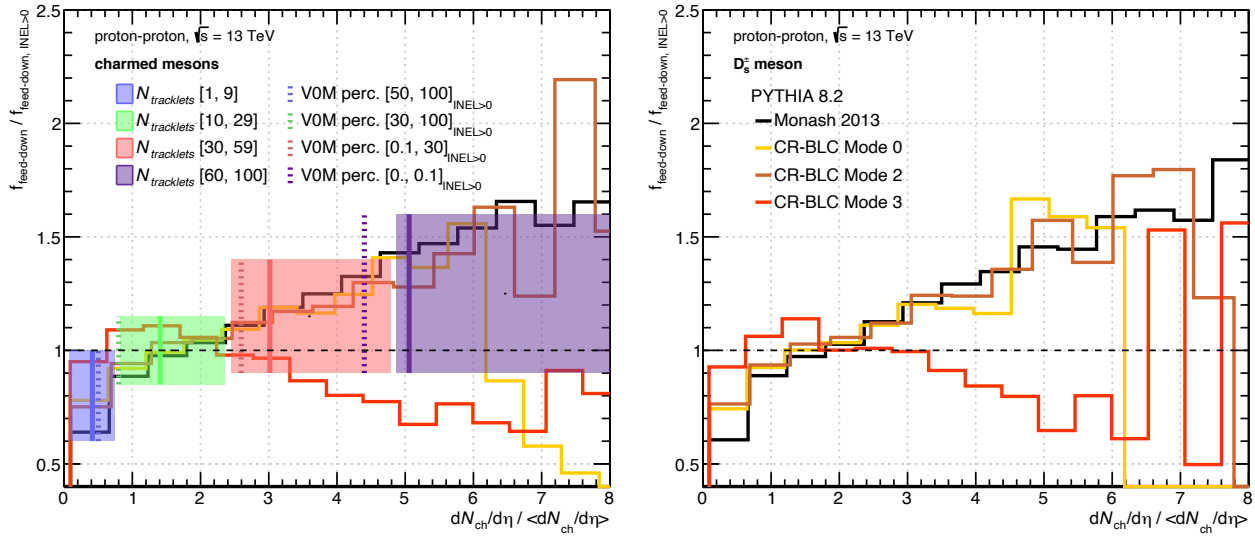


Figure V.6.1: Distribution of the ratio of the feed-down fraction to the INEL>0 value as a function of normalised charged particle density for charmed mesons (left panel) and D_s^+ mesons (right panel). The distributions are simulated with the Monte Carlo generator PYTHIA 8.2 [68] using the Monash 13 [136] and the string formation beyond leading colour tunes [118].

Based on the PYTHIA 8.2 results, amplitude variations of the feed-down D_s^+ fraction, $f_{\text{feed-down}}$, are defined in each multiplicity interval. The $N_{tracklets}$ intervals converted to their corresponding normalised charged particle density intervals (see section IV.2.1) the width of the coloured bands on the figure V.6.1. The height of the bands delimits the amplitude variations based on the prediction deviations from unity, more weight is given to the Monash 13 and Mode 0, 2 trends with respect to the Mode 3. The mean charged particle density values from table IV.2.2 are reported and shown as vertical plain lines for the $N_{tracklets}$ interval and dotted lines for the V0M percentile intervals. The variations explored are reported on table V.6.1.

Multiplicity interval	$f_{\text{feed-down}}$ variation
[1, 9]	[0.6, 1]
[10, 29]	[0.85, 1.15]
[30, 59]	[0.9, 1.4]
[60, 100]	[0.9, 1.6]
$[50, 100]_{\text{INEL>0}}$	[0.6, 1]
$[30, 100]_{\text{INEL>0}}$	[0.6, 1.15]
$[0.1, 30]_{\text{INEL>0}}$	[0.9, 1.4]
$[0, 0.1]_{\text{INEL>0}}$	[0.9, 1.6]

Table V.6.1: Summary table of the feed-down D_s^+ meson fraction variations performed in the different multiplicity bins.

The systematic uncertainty purely related to the multiplicity dependence of the prompt D_s^+ meson fraction is derived from the variation of the central value of f_{prompt}^{MB} obtained from the minimum-bias analysis. The procedure is as follows,

- from the central value of $f_{\text{prompt}}^{\text{MB}}$, the corresponding $f_{\text{feed-down}}^{\text{MB}} = 1 - f_{\text{prompt}}^{\text{MB}}$ is obtained
- the variation factors in table V.6.1 are then applied to the $f_{\text{feed-down}}^{\text{MB}}$ leading to a variation range $[f_{\text{feed-down}}^{\text{min}}, f_{\text{feed-down}}^{\text{max}}]$.
- the range is converted back to $f_{\text{prompt}}^{\text{min, max}}$ values with $f_{\text{prompt}} = 1 - f_{\text{feed-down}}$

The relative systematic uncertainty related to the variations are computed by making the ratios $f_{\text{prompt}}^{\text{min, max}}/f_{\text{prompt}}^{\text{MB}}$ in each multiplicity intervals. The values are summarised in the tables V.7.2 and V.7.3. For the production yield measurements, the variations are performed on the upper and lower f_{prompt} values in order to take into account the FONLL related systematic uncertainties. The envelope of the variations is taken as total systematic uncertainty on the prompt D_s^+ meson fraction. For the self-normalised yield measurements, *i.e.* production yield normalised to the INEL>0 value, only the multiplicity related systematic uncertainty as described above is considered.

V.7 Other sources of systematics and summary

In addition to the systematic uncertainties described in the previous sections, systematic uncertainties related to the trigger efficiency, the normalisation and the branching ratio are considered as well. In the minimum-bias analysis, the systematic uncertainty on the integrated luminosity of 1.6% is estimated from Van der Meer scan dedicated studies [247]. In the multiplicity analyses, the systematic uncertainty on the trigger efficiency in the different multiplicity intervals are summarised on table IV.2.2. The overall systematic uncertainty on the $\mathrm{D_s^+} \to \phi(1020)(\to \mathrm{K^+K^-})\pi^+$ branching ratio of 3.6% is taken from [37]. The systematic uncertainties associated to the different sources discussed in the previous sections are summarised on tables V.7.1, V.7.2, and V.7.3.

For the production cross section and yields measurements, the different sources are considered independent and therefore summed in quadrature. The yield measurements are used to produce other quantities such as cross section ratios and self-normalised yields. The uncertainty propagation strategy, that depends on the correlation between systematic sources, will be discussed in the sections of the chapter VI, presenting the results.

p_{T} **intervals** (GeV/c)	[1, 2]	[2, 3]	[3, 4]	[4, 5]	[5, 6]	[6, 8]	[8, 12]	[12, 16]	[16, 24]
Selection efficiency (%)	8	5	3	3	3	3	3	5	9
Raw yield extraction (%)	5	2	2	1	1	1	1	1	3
Tracking efficiency (%)	6	6	6	6	7	7	8	8	8
Generated p_{T} shape (%)	1	2	1	0	0	0	0	0	1
Multiplicity weights (%)	2	1	1	1	1	0	0	0	2
Prompt fraction (%)	$^{+11}_{-11}$	$^{+8}_{-9}$	$^{+7}_{-8}$	$^{+6}_{-7}$	$^{+6}_{-7}$	$^{+5}_{-7}$	$^{+6}_{-7}$	$^{+4}_{-6}$	$^{+3}_{-3}$

Table V.7.1: Summary table of the systematic uncertainties considered in the minimum-bias analysis.

$N_{tracklets}$ **interval**	p_{T} **intervals** (GeV/c)					
[1, 9999]	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
Selection efficiency (%)	8	3	3	3	3	5
Raw yield extraction (%)	4	1	1	1	1	2
Tracking efficiency (%)	6	6	7	7	8	8
Generated p_{T} shape (%)	1	5	1	0	0	1
Multiplicity weights (%)	2	1	1	0	0	1
z_{vertex} distribution (%)	0	0	0	0	0	1
Prompt fraction (%)	$^{+11}_{-11}$	$^{+7}_{-8}$	$^{+6}_{-7}$	$^{+5}_{-7}$	$^{+6}_{-7}$	$^{+4}_{-5}$
[1, 9]	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
Selection efficiency (%)		6	5	5	7	8
Raw yield extraction (%)		5	3	3	4	5
Tracking efficiency (%)		6	7	7	8	8
Generated p_{T} shape (%)		5	1	0	0	1
Multiplicity weights (%)		3	2	2	1	1
z_{vertex} distribution (%)		1	1	1	1	1
Prompt fraction (mult.) (%)		$^{+6}_{-0}$	$^{+6}_{-0}$	$^{+6}_{-0}$	$^{+8}_{-0}$	$^{+8}_{-0}$
[10, 29]	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
Selection efficiency (%)	10	4	3	3	3	5
Raw yield extraction (%)	4	1	1	1	1	3
Tracking efficiency (%)	6	6	7	7	8	8
Generated p_{T} shape (%)	1	5	1	0	0	1
Multiplicity weights (%)	0.5	0	0	0	0	0
z_{vertex} distribution (%)	1	1	1	1	1	1
Prompt fraction (mult.) (%)	$^{+3}_{-3}$	$^{+2}_{-2}$	$^{+2}_{-2}$	$^{+2}_{-2}$	$^{+3}_{-3}$	$^{+3}_{-3}$
[30, 59]	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
Selection efficiency (%)		6	4	3	3	7
Raw yield extraction (%)		2	2	2	2	3
Tracking efficiency (%)		6	7	7	8	8
Generated p_{T} shape (%)		5	1	0	0	1
Multiplicity weights (%)		0	0	0	0	0
z_{vertex} distribution (%)		1	1	1	1	1
Prompt fraction (mult.) (%)		$^{+1}_{-6}$	$^{+1}_{-6}$	$^{+2}_{-6}$	$^{+2}_{-8}$	$^{+2}_{-8}$
[60, 100]	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
Selection efficiency (%)		5	5	3	3	7
Raw yield extraction (%)		3	2	2	2	3
Tracking efficiency (%)		6	7	7	8	8
Generated p_{T} shape (%)		5	1	0	0	1
Multiplicity weights (%)		0	0	0	0	0
z_{vertex} distribution (%)		1	1	1	1	1
HMSPD trigger (%)		3	3	3	3	3
Prompt fraction (mult.) (%)		$^{+1}_{-9}$	$^{+1}_{-9}$	$^{+2}_{-9}$	$^{+2}_{-12}$	$^{+2}_{-12}$

Table V.7.2: Summary table of the systematic uncertainties considered in the $N_{tracklets}$ multiplicity analysis.

V0M percentile interval	p_{T} **intervals** (GeV/c)					
$[50, 100]_{\mathrm{INEL>0}}$	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
Selection efficiency (%)		9	4	4	7	10
Raw yield extraction (%)		4	2	2	3	5
Tracking efficiency (%)		6	7	7	8	8
Generated p_{T} shape (%)		5	1	0	0	1
Multiplicity weights (%)		3	2	1	1	0
z_{vertex} distribution (%)		0	0	0	0	0
Prompt fraction (mult.) (%)		$^{+6}_{-0}$	$^{+6}_{-0}$	$^{+6}_{-0}$	$^{+8}_{-0}$	$^{+8}_{-0}$
$[30, 100]_{\mathrm{INEL>0}}$	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
Selection efficiency (%)	7	5	3	4	5	10
Raw yield extraction (%)	5	2	2	2	3	4
Tracking efficiency (%)	6	6	7	7	8	8
Generated p_{T} shape (%)	1	5	1	0	0	1
Multiplicity weights (%)	3	2	1	1	0	0
z_{vertex} distribution (%)	0	0	0	0	0	0
Prompt fraction (mult.) (%)	$^{+7}_{-3}$	$^{+6}_{-2}$	$^{+6}_{-2}$	$^{+6}_{-2}$	$^{+8}_{-3}$	$^{+8}_{-3}$
$[0.1, 30]_{\mathrm{INEL>0}}$	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
Selection efficiency (%)	7	3	3	3	3	5
Raw yield extraction (%)	5	2	1	1	1	3
Tracking efficiency (%)	6	6	7	7	8	8
Generated p_{T} shape (%)	1	5	1	0	0	1
Multiplicity weights (%)	1	0	0	0	0	1
z_{vertex} distribution (%)	0	0	0	0	0	0
Prompt fraction (mult.) (%)	$^{+2}_{-7}$	$^{+1}_{-6}$	$^{+1}_{-6}$	$^{+2}_{-6}$	$^{+2}_{-8}$	$^{+2}_{-8}$
$[0, 0.1]_{\mathrm{INEL>0}}$	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
Selection efficiency (%)	6	5	3	3	3	3
Raw yield extraction (%)	5	2	2	2	2	2
Tracking efficiency (%)	6	6	7	7	8	8
Generated p_{T} shape (%)	1	5	1	0	0	1
Multiplicity weights (%)	1	1	1	1	1	1
z_{vertex} distribution (%)	1	1	1	1	1	1
Prompt fraction (mult.) (%)	$^{+2}_{-11}$	$^{+1}_{-9}$	$^{+1}_{-9}$	$^{+2}_{-9}$	$^{+2}_{-12}$	$^{+2}_{-12}$

Table V.7.3: Summary table of the systematic uncertainties considered in the V0M percentile multiplicity analysis.

Chapter VI

Results and discussion

VI.1 D_s^+ meson production in minimum-bias proton-proton collisions

VI.1.1 Production cross-section

The p_T-differential production cross section of D_s^+ meson in proton-proton collisions at $\sqrt{s}$ = 13 TeV is shown on figure VI.1.1 together with the same measurements performed at $\sqrt{s}$ = 5.02 TeV [122] and $\sqrt{s}$ = 7 TeV [243], showing an increasing production of the D_s^+ meson with the collision energy. The total relative systematic uncertainties computed for the $\sqrt{s}$ = 13 TeV analysis presented in this thesis are summarised in table VI.1.1 where each source of systematic is considered to be independent and summed in quadrature (see previous chapter V) in the final result. With the increased collision energy and integrated luminosity of the different data samples analysed:

$$\mathcal{L}_{\rm int}(7\ {\rm TeV}) = 6 \pm 0.21\ {\rm nb}^{-1} \quad \mathcal{L}_{\rm int}(5.02\ {\rm TeV}) = 19.3 \pm 0.40\ {\rm nb}^{-1} \quad \mathcal{L}_{\rm int}(13\ {\rm TeV}) = 31.7 \pm 0.51\ {\rm nb}^{-1}$$

the p_T coverage of the measurements can be extended towards low p_T (1 - 2 GeV/c) representing a large fraction of the total D_s^+ meson production cross section. Consequently, the statistical precision of the measurements is also increased as shown on the lower panel of figure VI.1.1 where the relative statistical uncertainties are shown. For the 2 to 12 GeV/c p_T interval, common across the measurements, the statistical uncertainties are reduced by a factor three for the measurement made at $\sqrt{s}$ = 13 TeV compared to the measurement performed at $\sqrt{s}$ = 7 TeV.

$p_{\rm T}$ **intervals** (GeV/c)	[1, 2]	[2, 3]	[3, 4]	[4, 5]	[5, 6]	[6, 8]	[8, 12]	[12, 16]	[16, 24]
Statistical unc. (%)	±20	±9	±6	±5	±5	±4	±5	±10	±11
Systematic unc. (%)	±15	±11	$^{+9}_{-10}$	$^{+9}_{-10}$	$^{+9}_{-10}$	$^{+9}_{-10}$	$^{+9}_{-10}$	±10	±13

Table VI.1.1: Summary table of the relative statistical and systematic uncertainties of the p_T-differential cross section of D_s^+ meson in proton-proton collisions at $\sqrt{s}$ = 13 TeV.

With the exception of the first and last p_T intervals, the total systematic uncertainties are about 10% over the whole p_T range and are similar to those of the $\sqrt{s}$ = 5.02 TeV measurement.

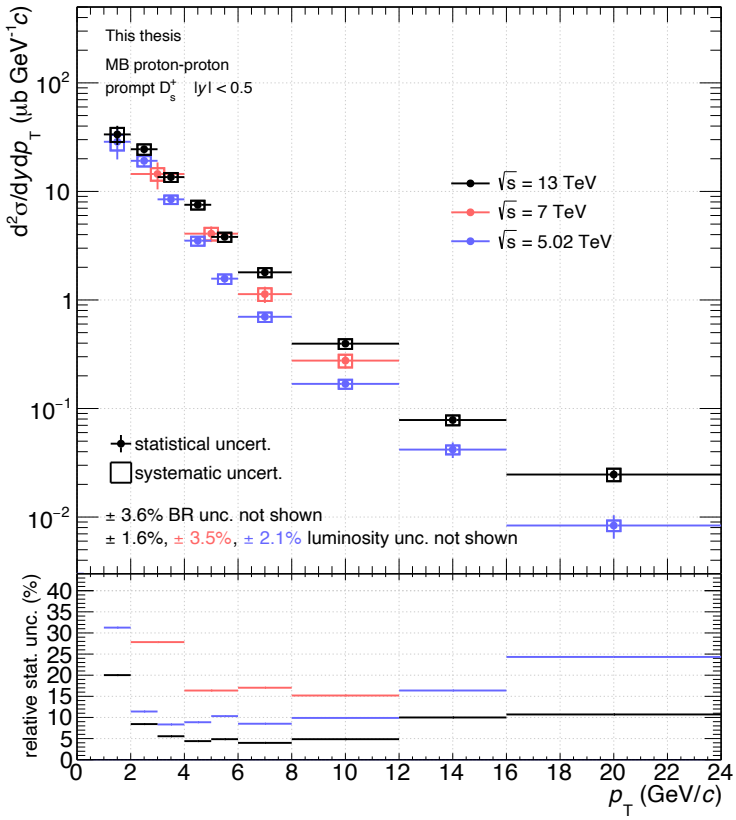


Figure VI.1.1: p_T-differential cross section of D_s^+ meson in proton-proton collisions at $\sqrt{s}$ = 13 TeV, 7 TeV [243] and 5 TeV [122]. On the lower panel, the relative statistical uncertainties of the measurements are compared.

In the following sections, this measurement is compared to two different approaches to the description of the D-mesons production in proton-proton collisions: with perturbative QCD calculations (pQCD) here performed with the GM-VNFS framework and with the PYTHIA 8 Monte Carlo generator.

Comparison to perturbative QCD calculations

A brief introduction to open heavy-flavour hadron prediction computations with perturbative QCD based frameworks is given in section II.2. In the GM-VNFS framework [126–128], the treatment of the heavy quark hadroproduction is performed within the Variable Flavour Number (VFN) scheme including heavy quark mass effects in the resummation procedure of divergent terms emerging from the perturbative QCD calculation (see section II.2.2). The CT14 NLO [135] parton density function set is used in the framework. Parton density functions are determined from a global fit analysis (see section II.2.1) on experimental data from lepton deep-inelastic scattering experiments, the full list of experimental inputs is given on table 1 of [135]. Heavy quark production processes are included up to the next to leading order (NLO) in the perturbative QCD calculation of the charm quark production cross section. The D_s^+ fragmentation function is taken from [115] and determined from fits of e^+e^- measurements performed by the OPAL Collaboration.

On the left panel of figure VI.1.2, the measurement of the p_T-differential cross section is compared to predictions from the GM-VNFS framework. The systematic uncertainties of the predictions are computed by varying the renormalisation scale μ_R fixed at the initial value μ_0 =

$\sqrt{p_{T,c}^2 + 4m_c^2}$ in the $0.5 \leq \mu_R/\mu_0 \leq 2$ interval. The ratio of the measurement over the prediction is shown on the lower part of the panel. The statistical and systematic uncertainties assigned to the ratio correspond to the relative uncertainties of the measurement. The central values from GM-VNFS are set to unity and the blue uncertainty band correspond to the relative systematic uncertainties.

The measurement and the predictions are compatible within uncertainties, however, at $p_T < 8$ GeV/c, the data points appear systematically in the upper part of the predictions. With respect to the precision of the measurement, the amplitude of the pQCD calculation systematic uncertainties are quite large at low p_T. This is due to the renormalisation scale dependence of the calculation rising from the missing higher order contributions to the charm quark production cross section. Currently, the leading order (LO) and NLO calculations represent the state of art of the heavy-flavour hadroproduction description. Recently, differential next-to-next-to leading order (NNLO) calculations have been performed for the bottom quark production [100] but no NNLO predictions are yet available for the charm quark. By including higher order contributions, the calculation becomes less dependent on the choice of the renormalisation and factorisation scales. In order to reduce the current associated systematic uncertainties, cross section ratios between different D meson species and at different collisions energies are performed to better constrains the calculations, exploiting the advantage of cancelling correlated sources of uncertainties. Additionally to the renormalisation scale dependence, the systematic uncertainties from the parton density function, the choice of charm quark mass value and from the D_s^+ fragmentation function are expected to be included.

Comparison to PYTHIA 8

Unlike the perturbative QCD frameworks, Monte Carlo generators aim to generate collision events and simulate all associated physics processes from the initial scattering interactions to the multi-particle final state. In the PYTHIA 8 Monte Carlo generator [68, 136], heavy quark production receives contributions from the initial hardest partonic scattering, possible subsequent hard multi-parton interactions (MPI), gluon splitting from hard scattering and gluon splitting from initial or final state radiations. The heavy quark production processes implemented are mainly leading order (LO) processes. The colour interactions between partons produced in MPI are governed by the colour reconnection model and the heavy quark hadronisation is performed via the string fragmentation model (see section II.2.3).

On the right panel of figure VI.1.2, the p_T-differential cross section is compared to predictions from the PYTHIA 8 with the Monash 13 tune [136], a set of parameters tuned to describe minimum-bias proton-proton collision LHC data. Additionally to the default configuration, the production cross section was computed by switching off the initial state radiation (ISR), the final state radiation (FSR), the multi-parton interactions and the colour reconnection (CR). For each prediction, a sample of 200 millions proton-proton events was simulated. The statistical precision of the predictions is considered to be negligible compared to the precision of the measurement and no systematic uncertainties are available, therefore, only the central points are compared to the measurement.

The Monash 13 tune describes reasonably well the data, the central points being in agreement within the experimental uncertainties over the full measured p_T range. The absence of the colour reconnection mechanism translates in a shift towards the low p_T (< 1 GeV/c) of the D_s^+ production cross section. Indeed, in the 2 to 8 GeV/c p_T range, a reduction of 30% of the p_T-differential cross section is observed but the p_T-integrated cross section values obtained from Monash 13 with and without including the mechanism are compatible ($d\sigma/dy \sim 95$ μb). As expected, switching off

initial and final state radiations or multi-parton interactions greatly affects the D_s^+ production causing a drop up to 60% of the p_T-integrated cross section.

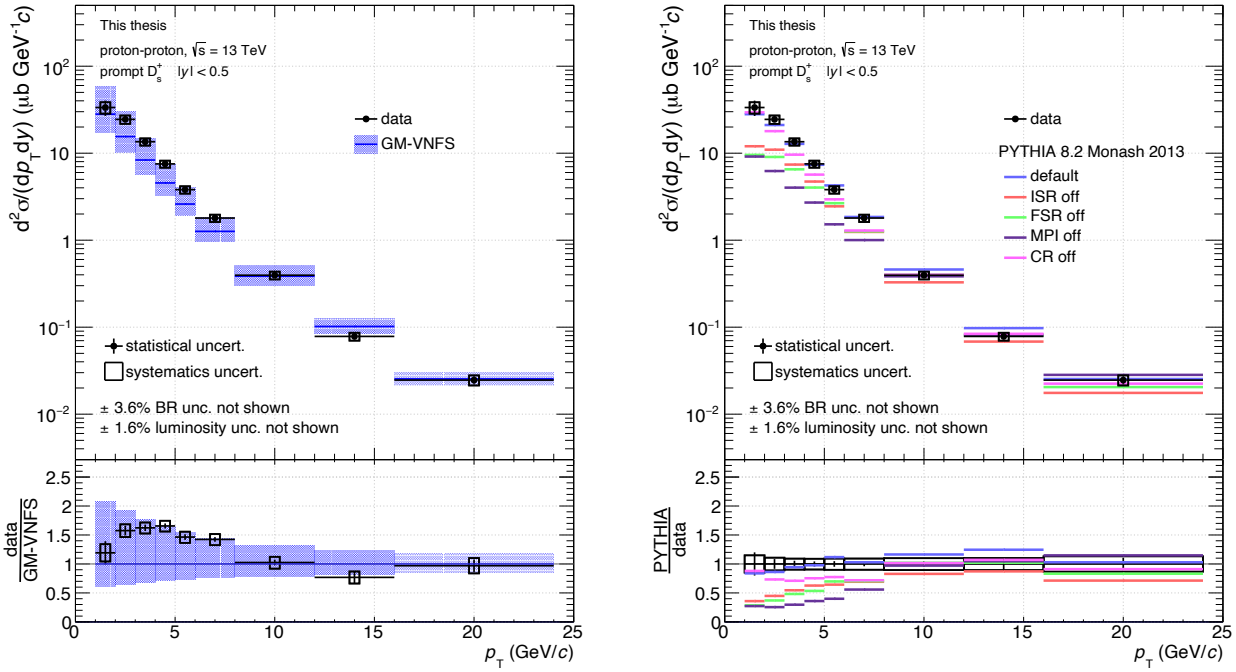


Figure VI.1.2: p_T-differential cross section of D_s^+ meson in proton-proton collisions at $\sqrt{s} = 13$ TeV compared to predictions from the GM-VNFS framework [126–128] on the left panel and from PYTHIA 8 with the Monash 13 tune [136] in different configurations on the right panel (see text for details). On the lower panels, the ratio between the measurement and the different predictions are shown.

VI.1.2 D_s^+ over D mesons cross section ratio

The D meson fragmentation process can be studied by measuring the proportions of the up, down and strange quarks produced from the charm quark fragmentation. The fragmentation fraction of charm quarks to charm-strange mesons f_s relatively to the one to non-strange charm mesons $f_u + f_d$ is estimated from the $D_s^+/(D^0+D^+)$ p_T-differential cross section ratio shown on figure VI.1.3 together with the measurements performed at $\sqrt{s} = 5.02$ TeV [122]. The D^0 and D^+ mesons measurements are taken from data analyses performed on the 2016 proton-proton collisions at $\sqrt{s} = 13$ TeV data sample ($\mathcal{L}_{\rm int} = 8.6 \pm 0.4$ nb^{-1}) by other members of the ALICE Collaboration.

The propagation of the systematic uncertainties is done depending on the degree of correlation between sources as summarised in table VI.1.2. The tracking efficiency and the integrated luminosity systematic uncertainties are considered as correlated source of systematic uncertainties and therefore cancel out in the ratio. For the $\sqrt{s} = 5.02$ TeV measurement, the systematic uncertainty related to the prompt fraction estimation is also considered as fully correlated between D meson species. A more conservative choice is adopted in this analysis where the largest relative systematic uncertainty on the prompt fraction between the measurements is considered in the propagation. The other sources of systematic uncertainties considered as uncorrelated were

summed in quadrature.

	correlated	partially correlated	uncorrelated
Selection efficiency			✓
Raw yield extraction			✓
Tracking efficiency	✓		
Generated p_T shape			✓
Prompt fraction		✓	
Branching ratio			✓
Integrated luminosity	✓		

Table VI.1.2: Degree of correlation between sources of systematic uncertainty between different D meson species measurements.

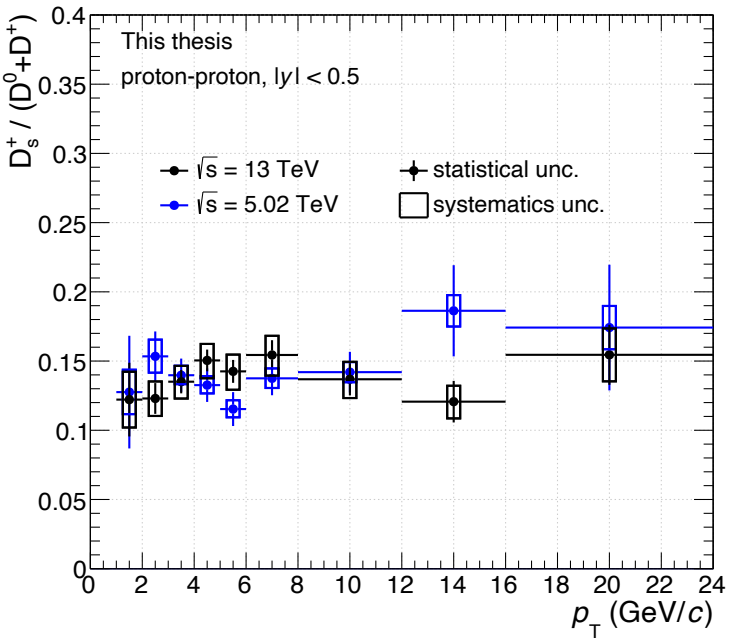


Figure VI.1.3: Ratio of the p_T-differential cross section of D_s^+ meson over the sum of the D^0 and D^+ cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV and 5.02 TeV [122].

The D_s^+, D^0 and D^+ mesons measurements include the contributions from the excited state D^{*0}, D^{*+} and D_s^{*+} mesons since the first two species fully decay into a D^0 or a D^+ and the latter into D_s^+ mesons [37]. The charm quark fragmentation fraction ratio is extracted by performing a constant fit of the cross section ratio shown on figure VI.1.3 under the hypothesis that the ratio is constant over the measured p_T range. The fit result gives,

$$\frac{f_s}{f_u + f_d} = 0.140 \pm 0.004(\text{stat})$$

Measurements of the strangeness suppression factor γ_s were performed in proton-proton collisions: ALICE [122, 248], ATLAS [249], e-proton scattering: H1 [250], ZEUS [251] and e^+e^-

annihilation: LEP [12] experiments. Presented on figure VI.1.4, the value was divided by two to match the definition of the charm quark fragmentation fraction ratio. The measurement of the ALICE Collaboration in proton-proton collision at $\sqrt{s} = 5.02$ TeV [122] is obtained by a constant fit of the cross section ratio (see figure VI.1.3) as performed in this analysis at $\sqrt{s} = 13$ TeV. Overall, the measurements performed in different collision systems and at different energies are compatible within uncertainties, indicating the universality of the charm fragmentation.

By neglecting the contribution of the decay of excited charm-strange mesons heavier than the D_s^{*+} to the D^0 and D^+ mesons yields, the charm-strange meson production is suppressed by a factor ~ 3.6 ($1/(2\gamma_s) = 3.57$) in the fragmentation of charm quarks.

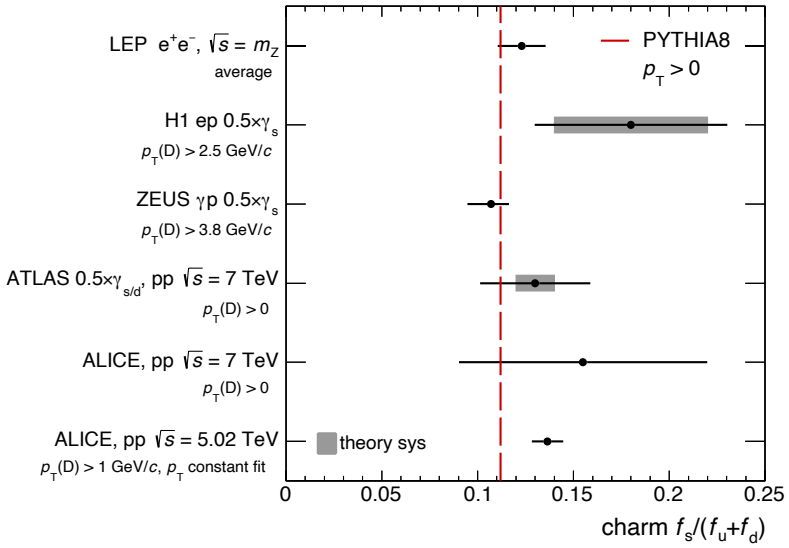


Figure VI.1.4: Charm quark fragmentation fraction ratio $f_s/(f_u+f_d)$ compared to measurements performed by the ALICE [122], [248], ATLAS [249], H1 [250], and ZEUS [251] Collaborations and to the averaged LEP [12] measurements. The total experimental uncertainties (bars) and the theoretical uncertainties (shaded boxes) are shown. The value obtained with PYTHIA 8 with the Monash 2013 tune is shown by the red dotted line [136]. The figure is taken from [122].

	correlated	partially correlated	uncorrelated
Selection efficiency			✓
Raw yield extraction			✓
Tracking efficiency			✓
Generated p_T shape			✓
Prompt fraction		✓	
Branching ratio	✓		
Integrated luminosity			✓

Table VI.1.3: Degree of correlation between sources of systematic uncertainty between measurement performed at different collision energies.

VI.1.3 Collision energy dependence

The production cross section measured at $\sqrt{s} = 13$ TeV is compared to the same measurement performed at $\sqrt{s} = 5.02$ TeV in the form of a cross section ratio. All the sources of systematic uncertainties are considered as uncorrelated between the two measurements except for the branching ratio systematic uncertainty considered fully correlated and the prompt-fraction systematic uncertainty considered partially correlated (see table VI.1.3).

The cross section ratio is compared to the GM-VNFS predictions [126–128] on the left panel of figure VI.1.5. The ratio of the D^0 meson at $\sqrt{s} = 13$ TeV and 5.02 TeV compared to FONLL predictions [123, 124, 252] is shown on the right panel. The D_s^+ and D^0 cross section ratios are increasing as a function of the particle transverse momentum in a similar trend. At low p_T the D_s^+ and D^0 productions are increased by a factor 1.5 at $\sqrt{s} = 13$ TeV with respect to the production at 5.02 TeV and reach up to a factor 3 at high p_T. Within uncertainties, the measurements are consistently described by the predictions from perturbative QCD.

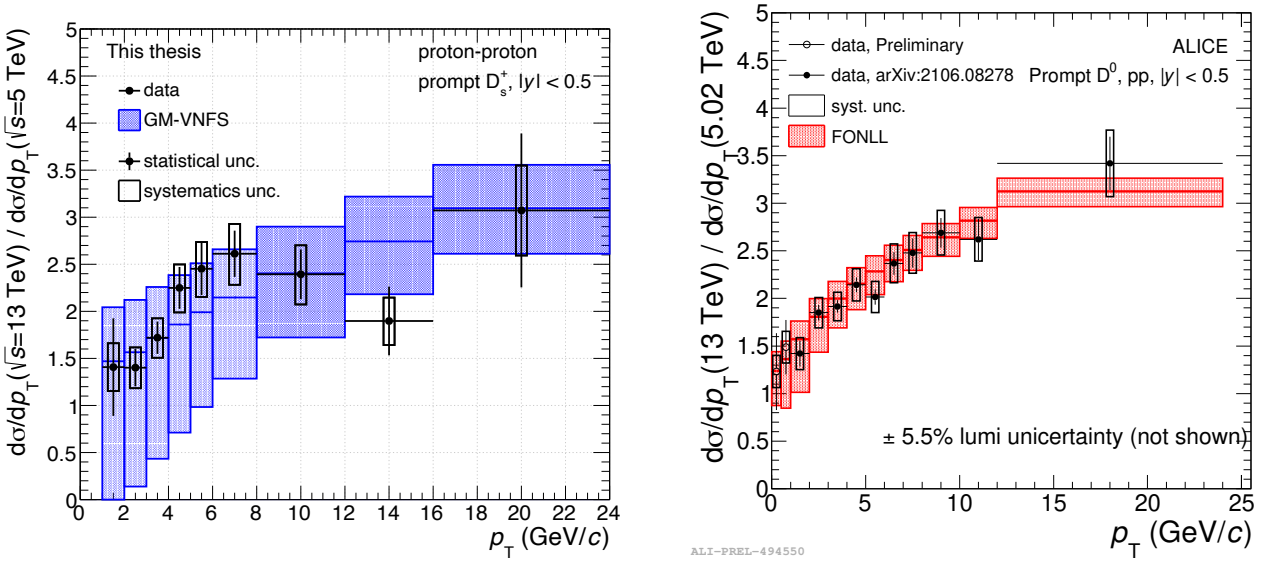


Figure VI.1.5: Ratio of the p_T-differential cross section of D_s^+ meson (left panel) and prompt D^0 meson (right panel) in proton-proton collisions at $\sqrt{s} = 13$ TeV and 5.02 TeV. The measurements are compared to the GM-VNFS [126–128] and FONLL perturbative QCD calculations [123, 124, 252].

The main contributions to the systematic uncertainty of perturbative QCD calculations originate from the dependence of the predictions on the renormalisation and factorisation scales and the choice of the heavy quark mass value. The missing higher order contributions to the charm quark production cross section are not expected to be strongly dependent on the collision energy $\sqrt{s}$. The choice of the scales at two different energies are therefore correlated leading to a reduction of the associated systematic uncertainty in the cross section ratio [252]. The detailed degree of correlation between the systematic uncertainties of the GM-VNFS predicted cross sections were not available. As a conservative choice, the minimum relative uncertainty between the measurements at $\sqrt{s} = 13$ TeV and 5.02 TeV is assigned. For the FONLL predictions, the correlation of the scale choices were taken into account as detailed in [252] leading to a great reduction of the systematic uncertainty.

As shown in sections II.2.4 and VI.1.1, the D meson p_T-differential production cross section measurements are more precise than their corresponding perturbative QCD predictions especially at low p_T. Additionally to the correlation between the uncertainty related to renormalisation and factorisation scale dependence, the parton density functions, the heavy quark mass and the D meson fragmentation function systematic uncertainties are also expected to be highly correlated. The resulting reduction of the total systematic uncertainty makes the perturbative QCD predictions more accurate.

VI.1.4 Transverse momentum integrated cross section

The measured visible cross sections of D_s^+ mesons are obtained by integration of the measurement in $p_T \in [1, 24]$ kinematic range. All the sources of systematic uncertainties are considered p_T-correlated to the exception of the uncertainty related to the raw yield extraction source. The p_T-correlated systematic uncertainties are summed linearly across p_T intervals while the p_T-uncorrelated uncertainty is summed in quadrature.

$$\sigma_{\text{vis}}(1 < p_T < 24\text{GeV}/c, |y| < 0.5) = 89 \pm 7(\text{stat}) \pm 8(\text{syst})\,\mu\text{b}$$

The visible p_T-differential cross section is extrapolated down to $p_T = 0$ using the *adhoc* Lévy-Tsallis and power law functions fitted on the spectra. The choice is motivated by their success in the description of light flavour hadron production cross sections [253]. Their definitions are as follow,

Lévy-Tsallis,

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}p_T\mathrm{d}y} = \frac{(n-1)(n-2)}{nC[nC + m(n-2)]} \cdot \frac{\mathrm{d}\sigma}{\mathrm{d}p_T} \cdot p_T \cdot \left(1 + \frac{m_T - m}{nc}\right)^{-n} \tag{VI.1}$$

with m the D_s^+ mass fixed to the PDG mass [37] and $m_T = \sqrt{m^2 + p_T^2}$ the D_s^+ transverse mass.

Power law,

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}p_T\mathrm{d}y} = \frac{\mathrm{d}\sigma}{\mathrm{d}p_T} \cdot p_T \cdot \left[1 + \left(\frac{p_T}{B}\right)^D\right]^{-C} \tag{VI.2}$$

The central fit is performed on the cross section with p_T-uncorrelated uncertainties. The p_T-uncorrelated uncertainty is taken as the quadrature sum of the statistical uncertainty and the systematic uncertainty related to the raw yield extraction. All other sources of systematic uncertainties are considered as p_T-correlated. As variations to the central fit, two additional fits are performed on the shifted data points by $\pm\ \sigma_{p_T\ \text{corr.}}$, the total p_T-correlated uncertainty. The envelope of the fits are shown on the figure VI.1.6, for the Lévy-Tsallis function on the left panel and for the power law function on the right panel. For $p_T < 2$ GeV/c, the functions shows different trends. While the Lévy-Tsallis function lays on the upper part of the data, the power law function goes via the lower part. Since most of the D_s^+ production happens at low p_T (50% of the D_s^+ at $p_T < 2$ GeV/c, 2% at $p_T > 8$ GeV/c) the p_T-integrated cross section absolute value is expected to be mostly influenced by this kinematic region.

Fit function	extrapolation factor to $p_T > 0$	$d\sigma/dy\ |_{|y|<0.5}$ (μb)
Lévy-Tsallis	$1.24^{+0.037}_{-0.046}$	110 $\pm$7(stat) $\pm$8(syst) $^{+3}_{-4}$(extrap)
Power law	$1.17^{+0.043}_{-0.111}$	104 $\pm$7(stat) $\pm$8(syst) $^{+4}_{-10}$(extrap)

Table VI.1.4: Extrapolation factors extracted from the fit and production cross sections of D_s^+ mesons.

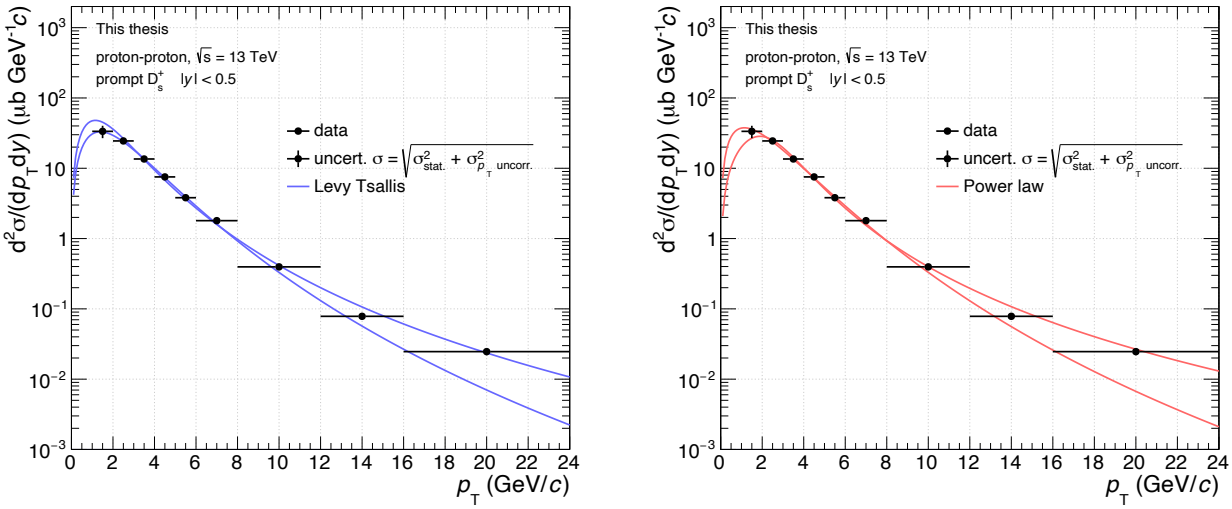


Figure VI.1.6: Fit of the p_T-differential cross section of D_s^+ meson with the Lévy-Tsallis (left panel) and the power law (right panel) functions.

The extrapolation factor is defined as the ratio of the integral over the full p_T range over the interval in the visible range. The associated uncertainty is computed considering the variations from the central fit and the uncertainties on the integral obtained from the fit procedure. The p_T-integrated cross sections were evaluated by multiplying the visible cross section by the extrapolation factor. The missing fraction of the visible cross section to reach the p_T-integrated cross section is about 20% which motivate the extension of the measurement toward $p_T = 0$, the fraction of the D_s^+ production at $p_T > 24$ GeV/c being below 1‰. The results obtained with the two fit functions and presented in table VI.1.4. The following p_T-integrated cross section at $\sqrt{s} = 13$ TeV is obtained by taking the weighted average of the two values, the measurement at $\sqrt{s} = 5.02$ TeV is taken from [122] and at $\sqrt{s} = 7$ TeV from [243],

$$\left.\frac{d\sigma}{dy}\right|_{|y|<0.5}^{\mathbf{13\ TeV}} = 108 \pm 7(\text{stat}) \pm 8(\text{syst})\, ^{+4}_{-10}(\text{extrap})\,\mu\text{b} \quad \text{(VI.3)}$$

$$\left.\frac{d\sigma}{dy}\right|_{|y|<0.5}^{\mathbf{7\ TeV}} = 89 \pm 18(\text{stat}) \pm 11(\text{syst})\, ^{+28}_{-26}(\text{extrap})\,\mu\text{b} \quad \text{(VI.4)}$$

$$\left.\frac{d\sigma}{dy}\right|_{|y|<0.5}^{\mathbf{5.02\ TeV}} = 82 \pm 12(\text{stat}) \pm 8(\text{syst})\, ^{+23}_{-8}(\text{extrap})\,\mu\text{b} \quad \text{(VI.5)}$$

The p_T-integrated cross section at $\sqrt{s} = 13$ TeV, is about 18% higher than the measurement at $\sqrt{s} = 7$ TeV [243] extrapolated from the visible cross section measured in the $2 < p_T < 12$ GeV/c

interval and about 24% higher than the measurement at $\sqrt{s} = 5.02$ TeV [243] extrapolated from the visible cross section measured in the $1 < p_T < 24$ GeV/c interval. However, the extrapolation uncertainties are still quite large, further motivating the extension of the measurement toward $p_T = 0$.

The extrapolation procedure employed in this analysis does not follow the procedure employed in the ALICE papers [122, 243] where perturbative QCD calculations are used to compute the extrapolation factor. The systematic uncertainties from pQCD are then propagated to the extrapolation systematic uncertainties.

VI.2 D_s^+ meson production as a function of the proton-proton collision multiplicity

VI.2.1 Production yields

The p_T-differential production yields of D_s^+ meson are shown on figure VI.2.1 for different $N_{tracklets}$ and V0M percentiles multiplicity event classes. The mean charged particle densities $\langle dN_{ch}/d\eta \rangle$ in $|\eta| < 0.5$ are quoted for each multiplicity intervals (see table IV.2.2). The lowest and highest charged particle densities are reached with the SPD tracklet multiplicity estimator with respectively $\langle dN_{ch}/d\eta \rangle_{|\eta|<0.5} = 3.1$ and $\langle dN_{ch}/d\eta \rangle_{|\eta|<0.5} = 37.8$ corresponding to ∼0.5 to ∼5.5 times the average multiplicity of the multiplicity integrated (INEL>0) interval.

As for the production cross section measurement, the sources of systematic uncertainties described in chapter V on the production yields are considered to be independent and summed in quadrature. They are summarised in tables VI.2.1 and VI.2.2. The total systematic uncertainties are dominated by the multiplicity dependence of the uncertainty on the estimation of the prompt D_s^+ fraction (see section V.6).

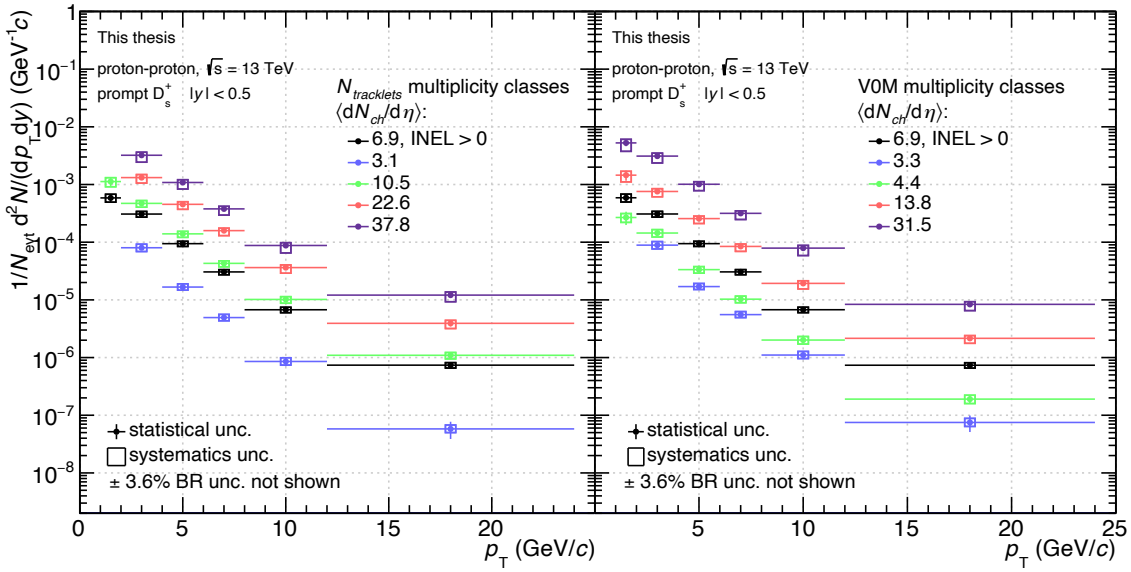


Figure VI.2.1: p_T-differential production yield of D_s^+ meson in proton-proton collisions at $\sqrt{s} = 13$ TeV in different multiplicity event classes. $N_{tracklets}$ classes (left), V0M classes (right).

The measurement shows an increase of the production of D_s^+ mesons with charged particle multiplicity at mid-rapidity. The Lévy-Tsallis fits of the spectra obtained for the different V0M multiplicity classes, similarly to the procedure described in section VI.1.4, present an increase of 32% of the average transverse momentum $\langle p_T \rangle$ value between the lowest and the highest multiplicity class. The hardening of the p_T spectra with increasing multiplicity is also observed for bottomia [254] and light-flavour [255] measurements with a mass ordering effect. The transverse momentum increases more rapidly with the multiplicity as the mass of the particle increases.

$\langle dN_{ch}/d\eta \rangle$	p_T intervals (GeV/c)					
3.1	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
Statistical unc. (%)		± 8	± 9	± 12	± 17	± 33
Systematic unc. (%)		$^{+15}_{-14}$	$^{+14}_{-12}$	$^{+14}_{-12}$	$^{+16}_{-13}$	$^{+17}_{-14}$
10.5	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
Statistical unc. (%)	± 20	± 5	± 4	± 5	± 6	± 9
Systematic unc. (%)	$^{+17}_{-19}$	$^{+12}_{-14}$	$^{+10}_{-13}$	$^{+11}_{-13}$	$^{+12}_{-14}$	$^{+12}_{-13}$
22.6	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
Statistical unc. (%)		± 9	± 5	± 6	± 8	± 12
Systematic unc. (%)		$^{+13}_{-20}$	$^{+11}_{-18}$	$^{+10}_{-18}$	$^{+11}_{-20}$	$^{+13}_{-19}$
37.8	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
Statistical unc. (%)		± 7	± 4	± 5	± 6	± 7
Systematic unc. (%)		$^{+13}_{-24}$	$^{+12}_{-23}$	$^{+11}_{-22}$	$^{+12}_{-26}$	$^{+13}_{-23}$

Table VI.2.1: Summary table of the relative statistical and systematic uncertainties of the p_T-differential production yield of D_s^+ meson in proton-proton collisions at $\sqrt{s} = 13$ TeV for the $N_{tracklets}$ multiplicity analysis.

$\langle dN_{ch}/d\eta \rangle$	p_T intervals (GeV/c)					
3.3	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
Statistical unc. (%)		± 8	± 9	± 11	± 15	± 32
Systematic unc. (%)		$^{+15}_{-14}$	$^{+13}_{-11}$	$^{+13}_{-11}$	$^{+16}_{-13}$	$^{+18}_{-15}$
4.4	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
Statistical unc. (%)	± 26	± 6	± 6	± 7	± 10	± 17
Systematic unc. (%)	$^{+17}_{-18}$	$^{+15}_{-15}$	$^{+13}_{-13}$	$^{+13}_{-13}$	$^{+15}_{-15}$	$^{+17}_{-16}$
13.8	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
Statistical unc. (%)	± 20	± 5	± 4	± 4	± 5	± 8
Systematic unc. (%)	$^{+17}_{-24}$	$^{+12}_{-19}$	$^{+10}_{-18}$	$^{+10}_{-18}$	$^{+11}_{-20}$	$^{+12}_{-18}$
31.5	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
Statistical unc. (%)	± 12	± 3	± 2	± 3	± 3	± 4
Systematic unc. (%)	$^{+15}_{-29}$	$^{+13}_{-24}$	$^{+11}_{-23}$	$^{+11}_{-22}$	$^{+11}_{-26}$	$^{+11}_{-22}$

Table VI.2.2: Summary table of the relative statistical and systematic uncertainties of the p_T-differential production yield of D_s^+ meson in proton-proton collisions at $\sqrt{s} = 13$ TeV for the V0M percentiles multiplicity analysis.

VI.2.2 Self-normalised yields

The self-normalised production yields of D_s^+ meson as a function of relative charged-particle multiplicity at central rapidity ($|\eta|<0.5$) are shown on figure VI.2.2 for both the $N_{tracklets}$ and V0M multiplicity classes. They are defined as the ratios of the production yields, $d^2N/dp_T dy$, normalised to the multiplicity integrated yield noted $\langle d^2N/dp_T dy\rangle$. The relative charged-particle density is defined as $dN_{ch}/d\eta/\langle dN_{ch}/d\eta\rangle$ with $dN_{ch}/d\eta$ the mean charged particle density in a given multiplicity interval and $\langle dN_{ch}/d\eta\rangle = 6.93$ the INEL>0 value (see table IV.2.2).

Since a fraction of the statistics is shared between the multiplicity intervals and the multiplicity integrated interval, the statistical uncertainties, as the systematic uncertainty related to the selection efficiency and to the raw yield extraction, are partially correlated in the minimum-bias triggered data sample. The statistical uncertainty on the self-normalised yields are taken as the difference between the statistical uncertainty on the measurement in a given multiplicity interval and in INEL>0, following the recommendations given on the page 280 of the *Data analysis in high energy physics* book [256]. For the highest multiplicity interval, where a different trigger strategy was used (see section IV.1.1), these uncertainties are considered as uncorrelated and therefore summed in quadrature. The multiplicity dependent part of the systematic uncertainty on the estimation of the prompt D_s^+ fraction is considered as uncorrelated. The systematic uncertainties on the tracking efficiency, the generated p_T shape in Monte Carlo simulation and the branching ratio are considered as fully correlated. The degree of correlation of the sources are summarised in table VI.2.3. The total systematic uncertainties are represented by boxes on the following figures.

	correlated	partially correlated	uncorrelated
Selection efficiency		✓ (MB trigger)	✓ (HM triggers)
Raw yield extraction		✓ (MB trigger)	✓ (HM triggers)
Tracking efficiency	✓		
Generated p_T shape	✓		
Multiplicity weights			✓
z_{vertex} distribution	✓ (MB trigger)		✓ (HM triggers)
simulation stat. unc.			✓
Prompt fraction		✓	
Branching ratio	✓		
Trigger efficiency			✓

Table VI.2.3: Degree of correlation between sources of systematic uncertainty for the self-normalised production yield measurements.

The measurements show a stronger than linear increase of the production of the D_s^+ as a function of charged particle multiplicity of the collision as indicated by the slope of the dotted straight line drawn on the figure. The increasing trend seems to be more pronounced for high p_T intervals compared with low p_T ones. The results for the two multiplicity estimators are merged on the left panel of figure VI.2.3. In order to better appreciate the trend of the data points for the low multiplicity part of the spectra, the self-normalised yields are divided by their corresponding relative charged particle density on the right panel of the figure.

For the two first multiplicity intervals where the average charged particle density is half the INEL>0 value, *i.e.* $dN_{ch}/d\eta/\langle dN_{ch}/d\eta\rangle \sim 0.5$, the production of D_s^+ is suppressed by a factor ∼3.7 for $2 < p_T(D_s^+) < 4$ GeV/c and by a factor ∼10 for $12 < p_T(D_s^+) < 24$ GeV/c compared to the

production in INEL>0. As the multiplicity of the events increases and goes above the INEL>0 charged particle density, the low precision of the measurements does not allow for drawing a firm conclusion on the p_T ordering as a function of the multiplicity as the results are compatible within uncertainties. For the highest multiplicity interval, $dN_{ch}/d\eta/\langle dN_{ch}/d\eta\rangle = 5.46$, a clear separation of the low and high p_T intervals is visible suggesting the hardening of the D_s^+ yield as a function of multiplicity contributing to the increase of the average transverse momentum $\langle p_T\rangle$ with multiplicity. Studies performed with PYTHIA 8 in [257] show that the number of high-p_T jets increases with the event multiplicity in the context of multi-partonic interactions as the probability of finding a hard parton is expected to be higher at high multiplicity.

As the charm quark is produced in hard parton interactions associated to large amount of gluon radiation leading to a large production of charged particle, heavy-flavour hadrons production is correlated to the multiplicity of the event (see section II.5). By measuring the charged particle multiplicity in the same pseudo-rapidity window as the D_s^+ meson, originating from the fragmentation of charm quarks, one might introduce a bias in the event multiplicity classification as more charged particles are expected to be produced in the vicinity of the direction of the propagation of the charm quark. This effect is known as "auto-correlation" effect. It is supported by measurements of the azimuthal correlations of D mesons with charged particles in proton-proton collisions [258] showing a "near-side" peak $(\Delta\phi, \Delta\eta) = (0, 0)$ and an "away-side" peak $\Delta\phi = \pi$ structure lying on top of the underlying event baseline. A similar increasing trend is observed in the D_s^+ yields as a function of the multiplicity estimated in the same mid-rapidity region ($|\eta|<1$, $N_{tracklets}$ estimator) and at forward rapidity ($3.7 < \eta < 1.7$ and $2.8 < \eta < 5.1$, V0M estimator) where a η gap is introduced between the D_s^+ and the multiplicity measurement regions.

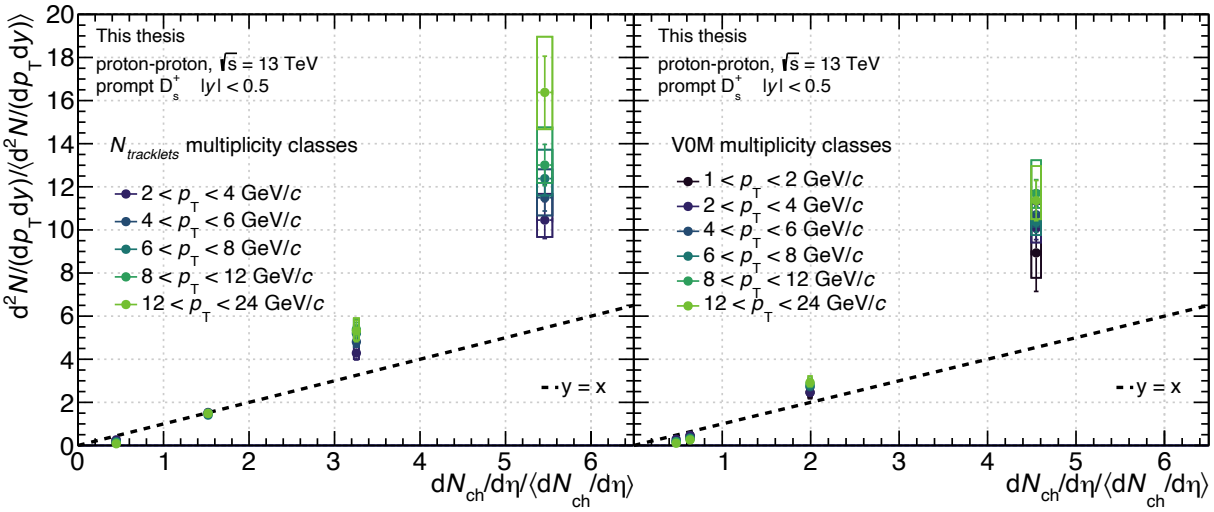


Figure VI.2.2: Self-normalised production yield of D_s^+ meson as a function of the relative charged-particle density at central rapidity in proton-proton collisions at $\sqrt{s} = 13$ TeV in different p_T intervals.

The ALICE preliminary results of the self-normalised yield of averaged D^0, D^+, D^{*+} mesons in proton-proton collisions at $\sqrt{s} = 13$ TeV is shown on figure VI.2.4. The analysis is done in different multiplicity intervals using the SPD tracklets as estimator, the intervals are summarised in table VI.2.4.

On the figure VI.2.5, the self-normalised yields of the D_s^+ meson and averaged D^0, D^+, D^{*+} mesons measurements are shown together for each p_T interval in separated panels. The ratios to the relative charged particle density are presented on the figure VI.2.6. Despite the slightly different first low and two highest multiplicity intervals, the corresponding charged particle density is quite similar (table VI.2.4). In the D_s^+ meson analysis, the $\langle dN_{ch}/d\eta \rangle = 6.93$ value is used in the computation of the relative charges particle densities while in the averaged D meson analysis the value $\langle dN_{ch}/d\eta \rangle = 7.42$ is used. The difference between the two values comes from the method employed for the conversion from the $N_{tracklets}$ interval to the charged particle density (see sections IV.2.1 and IV.2.3). As a result, the relative charged particle density are shifted between the two analyses. The increasing trend of the self-normalised yields as a function of multiplicity of the charm-strange D_s^+ meson is similar to the averaged non-strange D mesons. For the comparable charged particle density intervals, the results are compatible within uncertainties. The same measurements performed for the J/ψ also exhibit a similar trend [208].

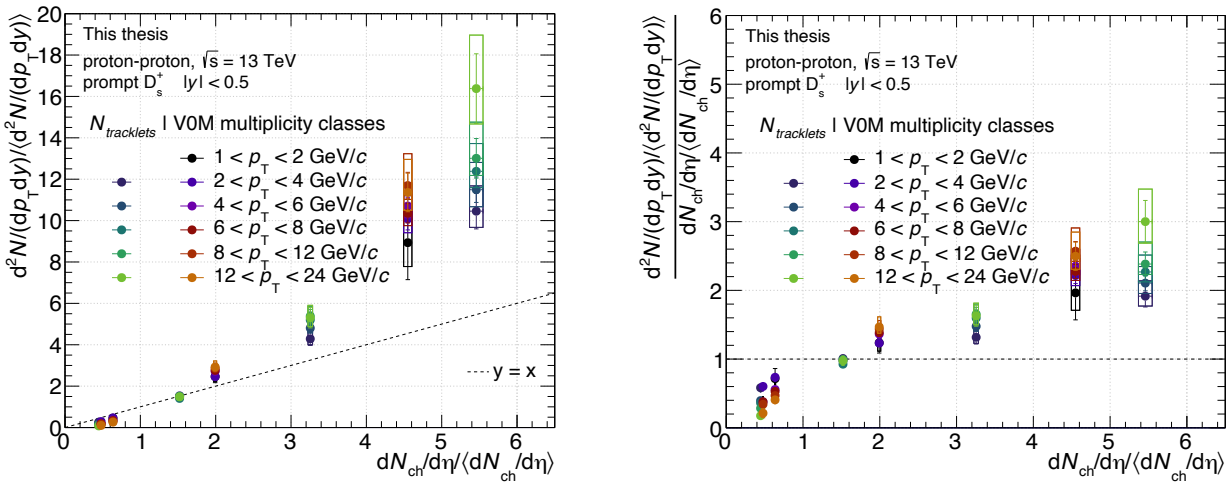


Figure VI.2.3: Self-normalised production yield of D_s^+ meson as a function of the relative charged-particle multiplicity at central rapidity in proton-proton collisions at $\sqrt{s} = 13$ TeV in different p_T intervals. On the right panel the results are divided by the relative charged-particle density.

The measurement at $\sqrt{s} = 13$ TeV extends the results obtained at $\sqrt{s} = 7$ TeV [208] where the study of the D_s^+ meson as a function of multiplicity was not possible due to the lack of statistics. The observation of an enhanced production of multi-strange light hadrons in high multiplicity proton-proton collisions [11, 259] with respect to pions motivated the measurement of the production of the D_s^+ meson. By comparison to the production of non-strange D mesons, different results could have indicated a change in the production of the D_s^+ related to the dense partonic system created at high multiplicity. The similarity of the results obtained with open and hidden charm hadrons suggests that this stronger-than-linear increase of the charm hadron production originates from the charm quark production rather than from charm quark hadronisation mechanism. Because the D-mesons production in proton-proton collisions is described by the fragmentation of the charm quark which is a quite different mechanism, compared with the description of the evolution of the $c\bar{c}$ pair into physical quarkonium states (see [180] for a review).

The description of D meson production as a function of multiplicity is explored with PYTHIA 8 [118, 136] and EPOS 3 [211, 260]. On the figures VI.2.7 and VI.2.8, the predictions from the

PYTHIA 8 Monte Carlo generator parametrised with different tunes [118] are shown together with the measurements. In PYTHIA 8, heavy quarks are produced in gluon fusion, quark-antiquark annihilation, flavour excitation and gluon splitting mechanisms (see section II.5.1 for the different proportions to the total D meson production). The tunes differ in the implementation of the colour reconnection mechanism responsible for the colour interaction between partons coming from multiple-partonic interactions (MPI). In the standard colour reconnection mechanism implemented in Monash 2013 [136], partons produced in MPIs can interact (they "reconnect") with partons produced from harder interactions. Partons, mostly gluons, are added to the colour string scheme representing the interaction, the string configuration minimising the total string length is retained. The effect of the colour reconnection model on the correlation between the number of MPI and the charged particle multiplicity at central rapidity is shown on figure 2 of [261]. Without the colour reconnection mechanism, the MPI are independent from each other and lead to a roughly linear increase of the charged particle multiplicity with the number of MPIs per collision.

A more advanced colour reconnection model [118] was developed to describe particle spectra and ratios in proton-proton collisions. In this model, colour string reconnection is based on three main considerations. The first consideration is the application of SU(3) colour rules for colour string compatibility. The second is a space-time check if two strings are connected by causality. The last consideration gives a preference to string topologies minimising the total string length. In addition to simple quark-antiquark colour connection, other reconnections topologies creating multi-quark junction structures are considered.

Three different parametrisations (mode 0, 2 and 3) of the beyond leading colour reconnection model (CR-BLC) are shown on the figures VI.2.7 and VI.2.8. The different implementations of the colour reconnection models in PYTHIA 8 give similar predictions on the production of D_s^+ mesons. For $p_T < 4$ GeV/c, the predicted D_s^+ production increase linearly as a function of multiplicity, as the p_T of particle further increases, a stronger than linear trend arises. Overall, PYTHIA 8 predictions describe qualitatively the data up to a relative charged particle density $\sim$3. At higher multiplicity the D_s^+ production is systematically underestimated.

In EPOS 3 [211, 260], the initial stage of collisions is described by a multiple scattering approach, the parton-based Gribov-Regge theory [212]. The produced parton matter follows a collective evolution driven by string momentum and local string density, the core-corona approach. Strings in low density regions (corona) escape from the core and hadronise through fragmentation while the high density region (core) undergo a hydrodynamic expansion and hadronise statistically. The EPOS 3 predictions for the average D meson production presented on figures VI.2.7 and VI.2.8 were computed with and without the hydrodynamic evolution of the core. The multiplicity trend of the averaged D meson predictions without using the hydrodynamic expansion are similar to the PYTHIA 8 D_s^+ predictions. By "turning on" the hydrodynamic contribution, the predictions move further away from the linear production to better describe the measured stronger than linear increase of the D meson production as a function of multiplicity.

Both models provide a description of multiple scattering interactions in proton-proton collisions either with the MPI implementation in PYTHIA 8 or with pomerons in the parton-based Gribov-Regge theory in EPOS 3. The number of parton interactions is roughly proportional to the number of charged particle produced at mid-rapidity in EPOS 3 [260] and in PYTHIA 8 if no colour reconnection mechanism is present [261]. The comparison with the measurements show that it is also roughly proportional to the number of produced D mesons even if at $p_T >$ 4-6 GeV/c the increase is slightly more than linear. The hydrodynamic description in EPOS 3 does not affect the D meson production however it greatly reduces the number of produced charged particles at high multiplicities. The reduction is explained as the re-distribution of the available

energy going into flow rather than particle production [260]. The stronger than linear increase of the D meson production is then interpreted as a sign of collective effects at high multiplicity. Experimentally, the rapidity gap introduced between the particles used for multiplicity estimation and the measured D mesons might be not sufficient to remove the auto-correlation effects of heavy flavour production with charged particle multiplicity, they are further explored in [261].

Averaged prompt D^0, D^+, D^{*+} mesons		**prompt D_s^+ meson**	
$N_{tracklets}$	$\langle dN_{ch}/d\eta \rangle$	$N_{tracklets}$	$\langle dN_{ch}/d\eta \rangle$
[1, 200]	7.42 ± 0.02	INEL>0	6.93 ± 0.09
[1, 8]	2.87 ± 0.02	[1, 9]	3.10 ± 0.02
[9, 13]	6.59 ± 0.04	[10, 29]	10.54 ± 0.09
[14, 19]	9.88 ± 0.06		
[20, 30]	14.59 ± 0.1		
[31, 59]	22.64 ± 0.17	[30, 59]	22.56 ± 0.21
[60, 99]	37.77 ± 0.47	[60, 100]	37.83 ± 0.62

Table VI.2.4: Summary table of the $N_{tracklets}$ intervals and the corresponding mean charge particle densities in $|\eta| < 0.5$ for the averaged prompt D^0, D^+, D^{*+} mesons and prompt D_s^+ meson analyses.

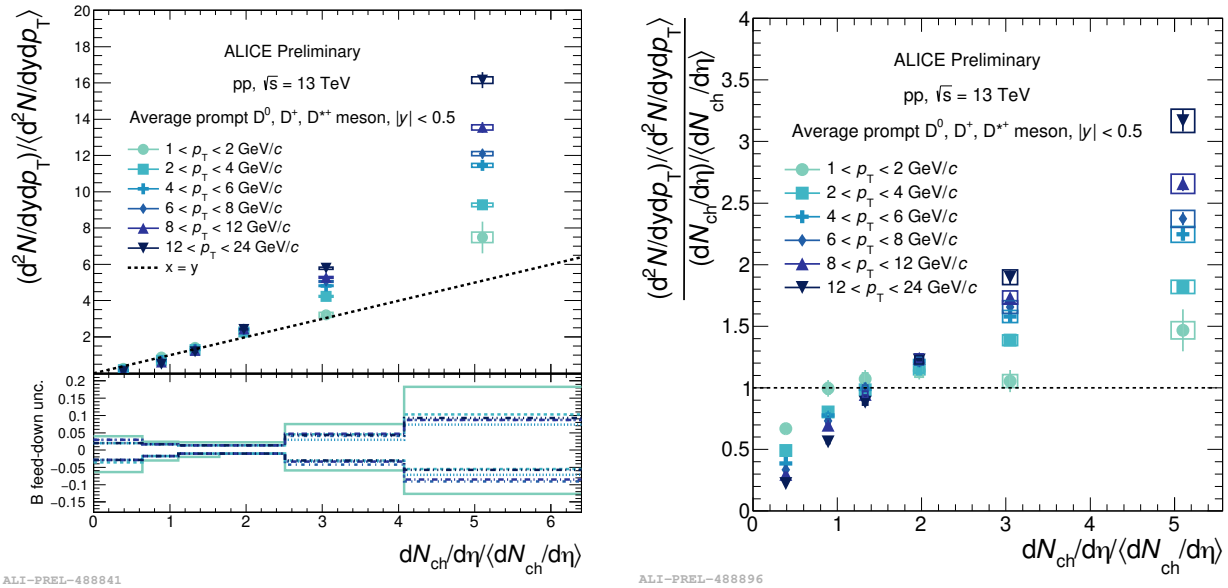


Figure VI.2.4: The average prompt D^0, D^+, D^{*+} self-normalised yields over relative charged-particle multiplicity as a function of the relative charged-particle multiplicity at central rapidity in proton-proton collisions at $\sqrt{s} = 13$ TeV in different p_T intervals. The results are presented in the top panel with statistical (vertical bars) and systematic (boxes) uncertainties. The feed-down fraction uncertainty is drawn in the bottom panel. On the right panel the results are divided by the relative charged-particle density.

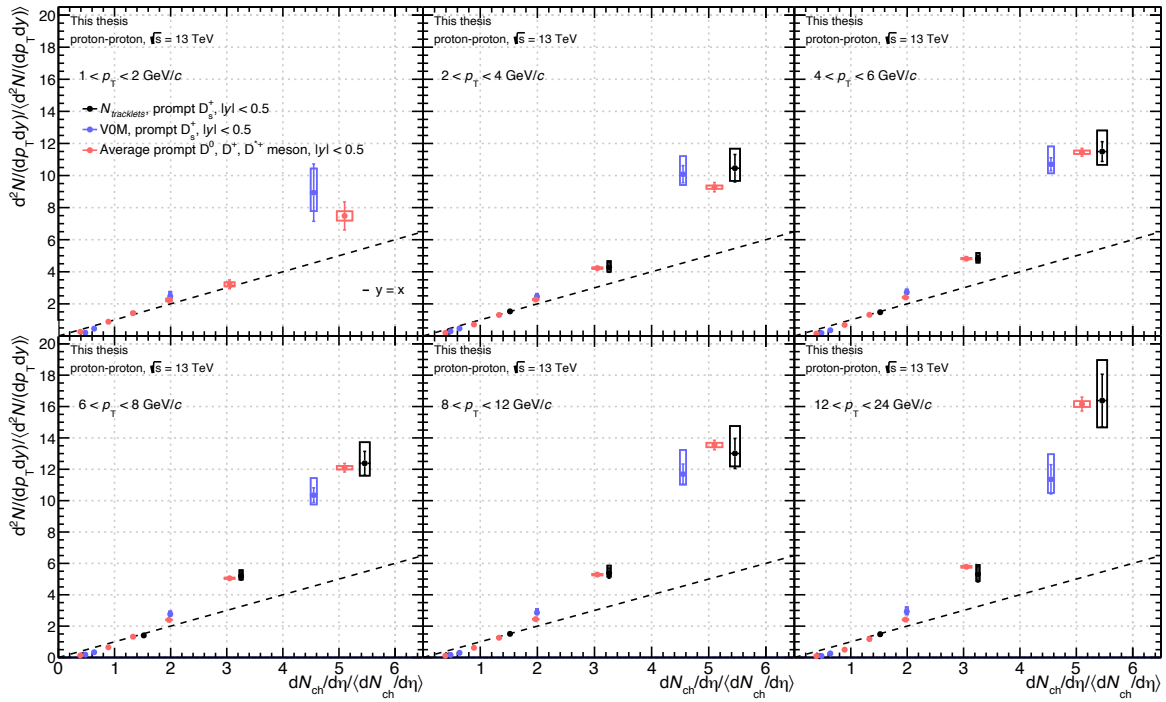


Figure VI.2.5: Self-normalised production yield of D_s^+ and averaged D^0, D^+, D^{*+} mesons as a function of the relative charged-particle multiplicity at central rapidity in proton-proton collisions at $\sqrt{s} = 13$ TeV in different p_T intervals.

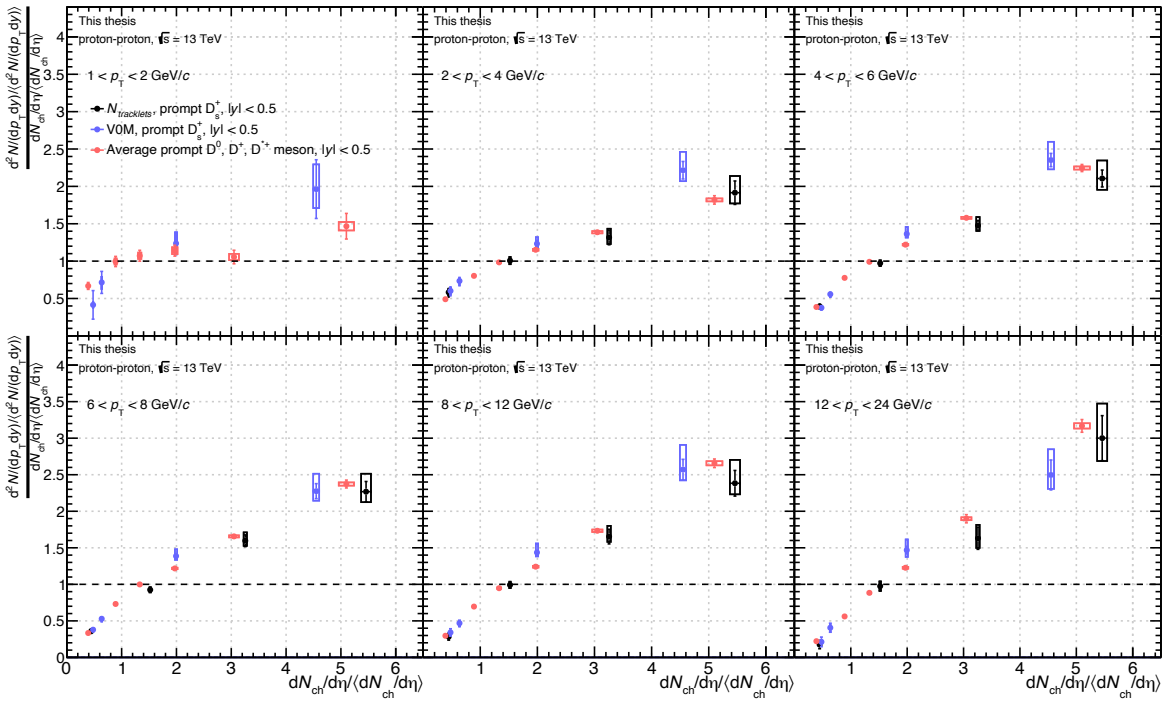


Figure VI.2.6: Self-normalised production yield of D_s^+ and averaged D^0, D^+, D^{*+} mesons divided by the relative charged particle density as a function of the relative charged-particle multiplicity at central rapidity in proton-proton collisions at $\sqrt{s} = 13$ TeV in different p_T intervals.

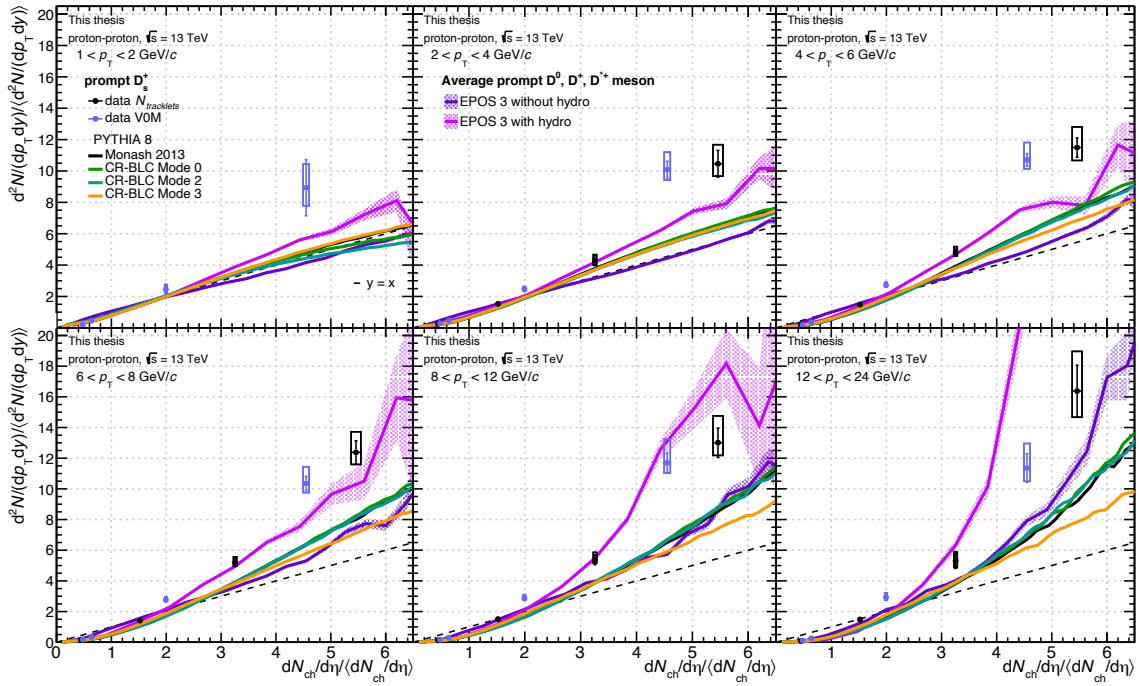


Figure VI.2.7: Self-normalised production yield of D_s^+ meson as a function of the relative charged-particle multiplicity at central rapidity in proton-proton collisions at $\sqrt{s} = 13$ TeV in different p_T intervals. The measurement is compared to predictions from PYTHIA 8 [118, 136] and EPOS 3 [211, 212].

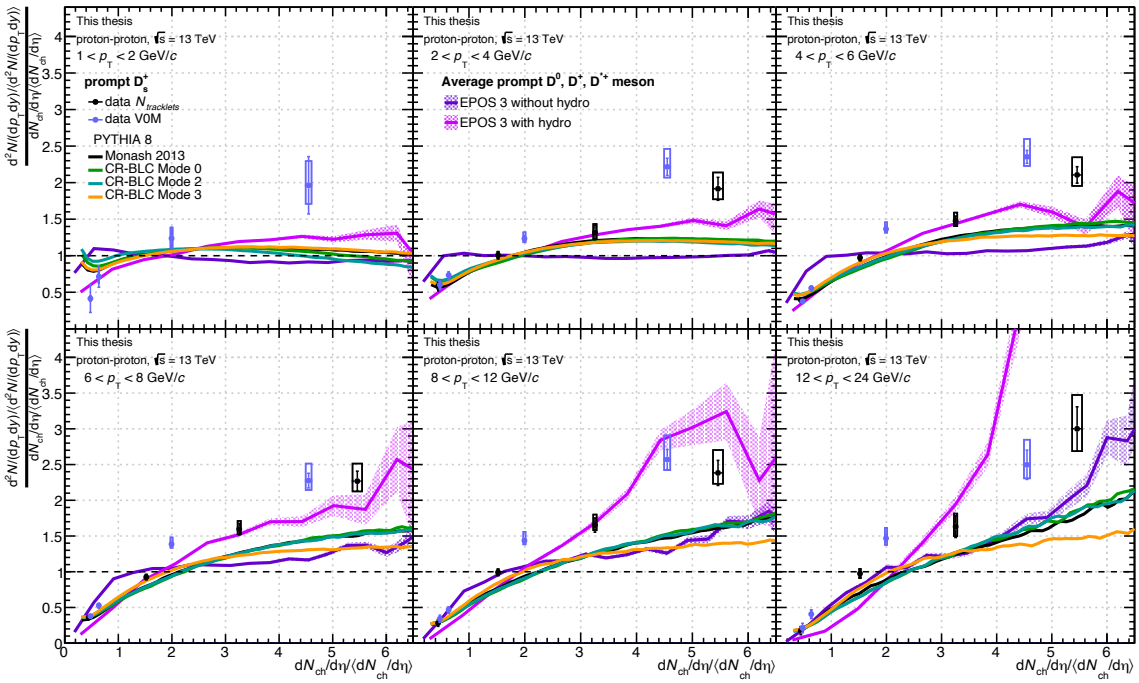


Figure VI.2.8: Self-normalised production yield of D_s^+ meson divided by the relative charged particle density as a function of the relative charged-particle multiplicity at central rapidity in proton-proton collisions at $\sqrt{s} = 13$ TeV in different p_T intervals. The measurement is compared to predictions from PYTHIA 8 [118, 136] and EPOS 3 [211, 212].

VI.2.3 D_s^+ over D^0 ratio

The measured ratio of production yields of D_s^+ over the D^0 as a function of p_T for the different $N_{tracklets}$ and V0M charged particle multiplicity classes are shown on figure VI.2.9. The source of systematics related to the selection efficiency, raw yield extraction, generated p_T shape, the z_{vertex} distribution and the branching ratio are considered to be uncorrelated and therefore added in quadrature. The other sources are considered to be correlated.

The measurements performed in this thesis do not show much dependence of the D_s^+/D^0 ratio as a function of the charged particle multiplicity. This results is also compatible with the averaged p_T-integrated measurements performed in e^+e^- collisions at LEP, 0.17 ± 0.03, [12]. On figure VI.2.10, the results for the lowest and highest $N_{tracklets}$ multiplicity classes are compared to predictions from the Monash 2013 tune [136] and the beyond leading colour reconnection tune [118] of PYTHIA 8. As discussed in the previous section VI.2.2 a similar increasing production of D_s^+ meson and non-strange D mesons is observed as a function of multiplicity. Despite the underestimation of the D meson production at high multiplicity, the colour reconnection mechanism implemented in PYTHIA 8 describe the D_s^+/D^0 ratio at both low and high charged particle multiplicities.

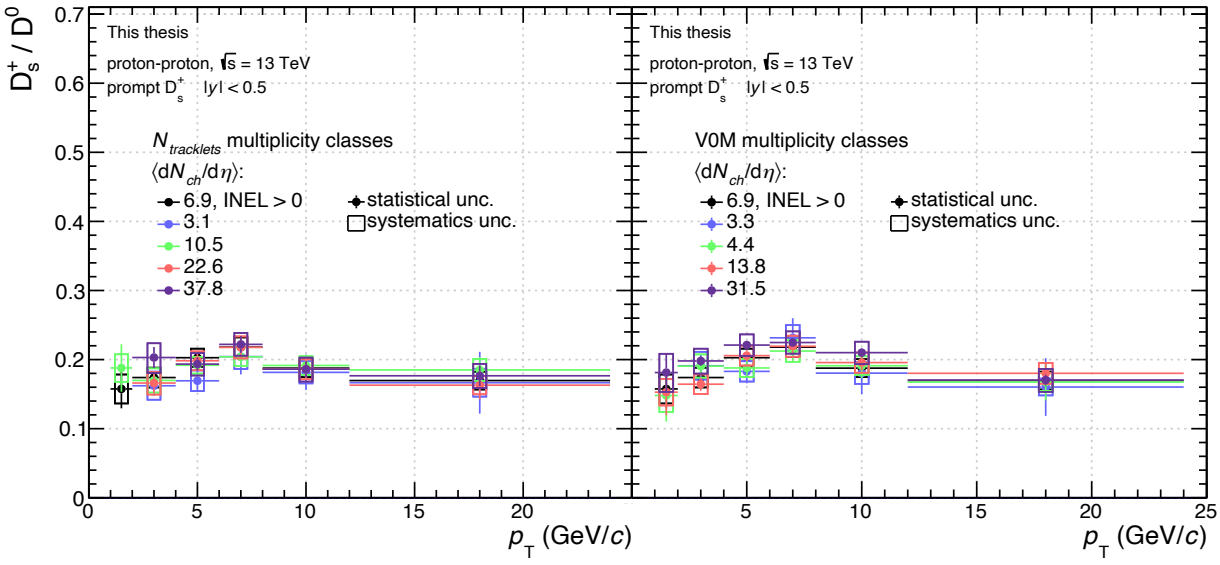


Figure VI.2.9: Production yield ratio of D_s^+ over the prompt D^0 as a function of p_T for different $N_{tracklets}$ and V0M charged particle multiplicity classes.

As was discussed in section I.3.1, particle production in heavy-ion collisions is successfully described by statistical hadronisation models (SHM). The canonical ensemble SHM (CE-SHM) [262] proposed as a generalisation of the grand canonical SHM allows for dealing with small reaction volumes such as found in proton-proton collisions. In such models, the hadron abundances are not obtained through the hadronisation of quarks but follow a probability law of an ideal hadron-resonance gas. The description of such a thermodynamical system depends on the provided hadron mass spectrum, *i.e.* the list of available hadrons, and the decay properties of resonance states. For the prediction obtained from [262] and compared to the measurement on figure VI.2.10, the charm hadron list from the PDG [237] is used as input. Statistical hadronisation models provide predictions for total, *i.e.* p_T-integrated, hadron production. The charm hadron p_T-differential spectra are obtained with the help of perturbative QCD calculations. The D^0 and D_s^+ p_T-spectra are derived from the charm quark p_T-spectrum and the charm quark fragmentation functions from

the perturbative QCD framework FONLL [123,124]. The fragmentation function for a given charm hadron H_c is normalised by a coefficient so that the p_T-integrated yield of H_c matches its relative density to the total density of all charm hadrons for a given volume in the CE-SHM [262]. In order to obtain results at low and high charged particle multiplicity, a correlation is made between the volume of the system in CE-SHM and the particle density per unit of rapidity. The description of the system in the canonical ensemble require the exact conservation of the quantum numbers: Q the electric charge, B the baryonic number, S the strangeness and C the charm quantum numbers. The system was chosen to be completely neutral in the CE-SHM calculation, $(Q, B, S, C) = (0, 0, 0, 0)$. From large to small system volume, the thermal density of the D_s^+ and the D^0 decreases due to the charm conservation reducing the phase space of open charm hadron production (see table 3 of [262]). The decrease of the thermal density ratio D_s^+/D^0 is mainly driven by the additional strangeness conservation restriction on the D_s^+ meson with respect to the non-strange D^0, this effect is known as the canonical strangeness suppression. At low multiplicity, the prediction from the CE-SHM is compatible with the measurement and the PYTHIA 8 predictions, while at high multiplicity, it tend to overshoot the measurement.

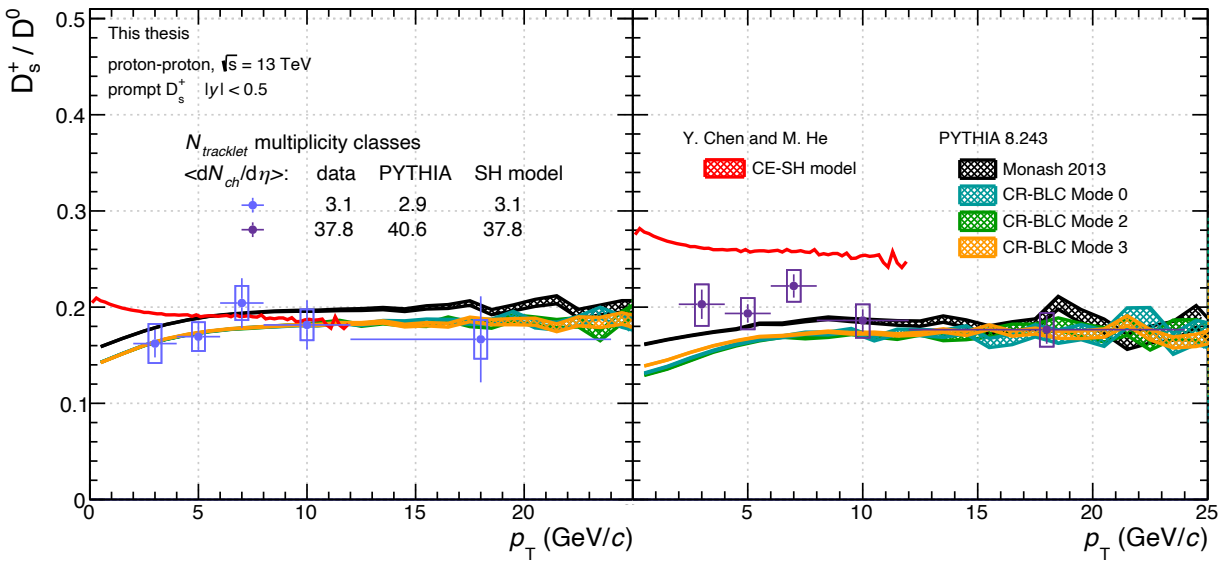


Figure VI.2.10: Production yield ratio of D_s^+ over the D^0 as a function of p_T for the lowest (left panel) and highest (right panel) $N_{tracklets}$ multiplicity classes. The ratios are compared to predictions from PYTHIA 8 [118,136] and a canonical ensemble statistical hadronisation (CE-SH) model [262].

VI.2.4 Discussion

The results presented in this thesis and charm baryon over meson ratio measured in the ALICE Collaboration are discussed in the context of the observation of collectivity-like effects in small systems at high multiplicity. As introduced in section II.5.2, recent measurements [10], [11] performed in small collision systems at high charged particle multiplicity, provide indications of collectivity-like effects originally associated with the phenomenology of heavy-ion collisions. The strangeness enhancement, measured in central heavy-ion collisions, is one of the signature of the production of a quark-gluon plasma (see section I.3.2). The measured enhanced production of

light strange hadrons in high multiplicity proton-proton collisions shows a steeper increase of the p_T-integrated production as a function of multiplicity with the strangeness content of the hadron and follow a universal trend across collision systems from low multiplicity proton-proton collisions to central Pb-Pb collisions (figure 4 of [11]). The particle production from multi-partonic interactions in PYTHIA 8 [136] predicts a flat multiplicity evolution of the strange over non-strange hadron in proton-proton collisions and therefore does not reproduce the measurement. The qualitative description of the results requires further development of the string fragmentation model of PYTHIA 8 as done in the DIPSY Monte Carlo generator [69] where in high density proton-proton collisions, strings overlap and form ropes increasing the tension between colour source leading to an increased production of strange quark (figure 2 of [11]). The canonical strangeness suppression at low multiplicity from the statistical hadronisation models (figure 15 of [263]) and the hybrid "core-corona" approach to hadronisation in EPOS [211] describe also qualitatively the data (figure 2 of [11]). The measurement of the $K/\pi = 2K_s^0/(\pi^+ + \pi^-)$ p_T-integrated production yield ratio reveals the increasing kaon production (strangeness quantum number $S = 1$) with respect to pions ($S = 0$) as a function of multiplicity as shown on figure 2 of [11].

The comparison of the charm-strange D_s^+ meson production with respect to non-strange charm meson allow to probe strangeness production in the heavy-flavour sector. The measurement of the D_s^+/D^0 p_T-differential production yield ratio do not show a multiplicity dependence within uncertainties. The reached precision of the D_s^+/D^0 p_T-differential production yield ratio as a function of multiplicity does not allow for drawing firm conclusions about a possible increase of the charm-strange meson ($S = 1$) production with respect to non strange D mesons ($S = 0$) as the results are compatible within uncertainties. The PYTHIA 8 predictions suggest that no additional mechanism of strange quark production is needed to describe qualitatively the strange-charm meson production. The current result at $\langle dN_{ch}/d\eta \rangle_{|\eta|<0.5} \sim 40$ does not reach yet the multiplicity of the measurement at low centrality in Pb-Pb collisions [15]. As the measurement in Pb-Pb collisions indicates an enhancement of the D_s^+/D^0 ratio (see figure 4 of [15]), further precise measurements are needed to get insight on a the possible continuous evolution of the D_s^+/D^0 ratio with multiplicity across collision systems as observed in the light flavour sector.

It is interesting to note here, that several measurement of baryon over meson ratios show a different trend as a function multiplicity. The measurement of the p_T-differential production yield ratio of Λ_c^+/D^0 in proton-proton collisions at $\sqrt{s} = 13$ TeV at low $\langle dN_{ch}/d\eta \rangle_{|\eta|<0.5} \sim 4$ and high $\langle dN_{ch}/d\eta \rangle_{|\eta|<0.5} \sim 45$ multiplicity is shown on the left panel of figure VI.2.11 together with PYTHIA 8 predictions. The Λ_c^+/D^0 ratio shows a dependence with the collision multiplicity, the results at high multiplicity being significantly larger than at low multiplicity (5.3 σ effect). The predictions from the Monash 2013 tune [136] do not describe the p_T trend of the ratio nor the multiplicity dependence while the CR-BLC Mode 2 tune [118] reproduces qualitatively both. As discussed in section VI.2.2, the CR-BLC model further develops the colour reconnection mechanism, particularly, the multi-quark junction reconnections topologies enhancing baryon production. The improvement of the description of the Λ/K_s^0 ratio are shown on figure 20(b) of [118] where the CR-BLC model called "new CR model" on the figure better describe the measurement. Similarities between light- and heavy- flavour baryon over meson ratios are observed as shown on the right panel of the figure VI.2.11 and both are qualitatively described by colour reconnection which is a collective phenomenon.

The figure 7 of [263] shows that the baryon to meson ratios p/π and Λ/K_s^0 at high multiplicity present a similar depletion at $p_T \lesssim 1.5$ GeV/c and an enhancement at intermediate p_T ($\sim$3 GeV/c) with respect to the low multiplicity measurements and the trend is also observed in p-Pb and Pb-Pb

collisions. In Pb-Pb collisions, the behaviour is interpreted as radial flow and parton coalescence (see section I.3.3) where baryons are further pushed to higher momenta due to their larger mass with respect to mesons. In Ref. [264], an enhanced Λ/K^0_s ratio is measured in small systems in the underlying region, *i.e.* not associated to the hard scattering process, with respect to the in-jet region. The results indicate that the baryon over meson ratio enhancement can be linked to collective behaviour phenomena at high multiplicity in small systems. However the link with heavy flavour baryon to meson ratio is not straightforward as the production mechanism of light and heavy hadrons are much different. Charm quark production is always produced in initial hard scattering while light flavour quark are mainly produced in gluon string breaking or soft scattering processes.

The results obtained in this thesis also indicate that features appearing in the baryon/meson ratios should be attributed to specific mechanism of baryon production, rather than to straightforward coalescence effects common for production of both baryons and mesons.

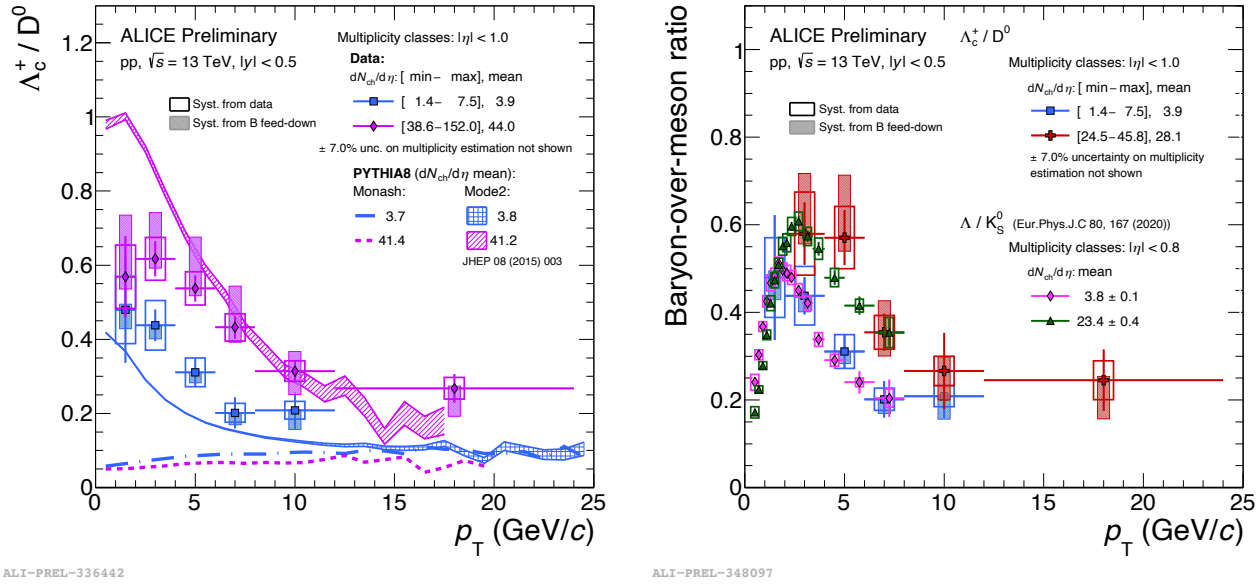


Figure VI.2.11: Production yield ratio of Λ_c^+ over the D^0 as a function of p_T in proton-proton collisions at $\sqrt{s} = 13$ TeV at low and high charged particle multiplicity. The ratios are compared to predictions from PYTHIA 8 [118, 136] on the left panel and to the measurement of the Λ to K^0_s ratio [259] on the right panel.

Conclusion and perspectives

The production of the strange charmed mesons $\mathrm{D_s^+}$, in proton-proton collisions has been studied at the centre-of-mass energy of $\sqrt{s} = 13$ TeV, using the ALICE detector at the CERN LHC. The $\mathrm{D_s^+}$ meson p_T-differential production yield was measured, for the first time, as a function of the collision charged particle density, thanks to the high statistics data collected during the LHC run II.

The $\mathrm{D_s^+}$ meson reconstruction was performed in the $\mathrm{D_s^+} \to \phi(1020)(\to \mathrm{K^+K^-})\pi^+$ decay channel exploiting the displaced decay-vertex topology. A machine learning algorithm, the Boosted Decision Tree algorithm, was employed to improve the $\mathrm{D_s^+}$ meson signal extraction from the combinatorial background with respect to the standard approach that relies on applying independent selection cuts. The use of a machine learning algorithm allows for a gain in statistical precision over an extended transverse-momentum range.

$\mathrm{D_s^+}$ meson production in minimum-bias proton-proton collisions

The measurement of the p_T-differential production cross section of $\mathrm{D_s^+}$ mesons, at mid-rapidity ($|y| < 0.5$), was performed over a broad transverse momentum range going from $p_\mathrm{T} = 1$ GeV/c to $p_\mathrm{T} = 24$ GeV/c, extending the previous measurements at $\sqrt{s} = 7$ TeV [243] and at $\sqrt{s} = 5.02$ TeV [265]. The data at $\sqrt{s} = 5$ TeV were recently re-analysed [122], using a machine learning technique, leading to a significant gain in the precision of the measurement.

With respect to the recent measurement at $\sqrt{s} = 5.02$ TeV [122], the statistical precision of the measurement has been improved by a factor 1.2 to 2 depending on the p_T interval. The results are described by perturbative QCD calculations [127,128] based on the factorisation theorem assuming the universality of the charm quark fragmentation. By comparison to the production of non-strange D mesons, the charm-strange meson production is suppressed by a factor ~ 3.6, this result is compatible with the measurements made in $\mathrm{e^+e^-}$ collisions [12] and proton-proton collisions [122,248,249] at different collision energies indicating the universality of the charm fragmentation into charm-strange mesons.

The dependence of the $\mathrm{D_s^+}$ meson production cross section on the collision energy was studied by making the cross section ratio $\mathrm{d}\sigma/\mathrm{d}p_\mathrm{T}$ (13 TeV)/$\mathrm{d}\sigma/\mathrm{d}p_\mathrm{T}$ (5.02 TeV). The result shows an increasing trend as a function of p_T, similar to that observed for the non-strange $\mathrm{D^0}$ meson. The ratio is described within uncertainties by perturbative QCD calculations [123,124,127,128] where the theoretical systematic uncertainties are greatly reduced as shown in [252].

The p_T-differential cross section was extrapolated from the measured p_T region $1 < p_\mathrm{T} < 24$ GeV/c to the full kinematic range. From the extrapolation procedure, a 20% missing fraction of the $\mathrm{D_s^+}$ production from the $p_\mathrm{T} < 1$ GeV/c region to the total production cross section was estimated. The p_T-integrated cross section, $\mathrm{d}\sigma/\mathrm{d}y|_{|y|<0.5} = 108 \pm 7(\mathrm{stat}) \pm 8(\mathrm{syst}) \,^{+4}_{-10}(\mathrm{extrap})$ μb, is found to be about 24% (18%) higher than the measurement at $\sqrt{s} = 5.02$ TeV (7 TeV) [122,243].

D_s^+ meson production as a function of multiplicity

The high statistics collected during the LHC run II allows for the first time for the study of the D_s^+ production as a function of the collision charged-particle multiplicity. The measurement of the production yields of D_s^+ meson at mid-rapidity ($|y| < 0.5$), was performed in different multiplicity intervals from 0.5 to 5.5 times the minimum-bias mean charged particle density at mid-rapidity.

The measurement shows a stronger than linear increase of the D_s^+ production with the charged particle density of the collisions. The trend is observed for two different multiplicity estimators used to evaluate potential auto-correlations effects between the multiplicity estimation and the D_s^+ meson reconstruction. A similar stronger than linear production is also observed for non-strange D mesons and the J/ψ [208] indicating that this trend originates from common charm quark production rather than from specific hadronisation mechanism effects.

The predictions from the PYTHIA 8 Monte Carlo generator fail to describe the observed heavy flavour hadron production at high multiplicity. Collectivity effects affecting heavy flavour hadron production, such as an hydrodynamical evolution of the dense part of the partonic medium, are suggested by the EPOS model [211], in order to better describe the data.

The measurement of the p_{T}-differential D_s^+/D^0 ratio is discussed in the light of the measurement of the enhanced production of light strange hadrons in small systems at high multiplicity [11]. The current results, probing the strangeness production in the heavy-flavour sector, do not show a multiplicity dependence within experimental uncertainties. Moreover, the results are compatible with the measurement in e^+e^- collisions [12]. A similar conclusion was made from the D_s^+/D^+ measurement in proton-Pb collisions at $\sqrt{s} = 5.02$ TeV [153]. The predictions from PYTHIA 8 describe the results in a scenario of a D meson production via fragmentation. The colour reconnection beyond leading colour mechanism [118] describe simultaneously the p_{T}-differential D_s^+/D^0 and Λ_c^+/D^0 ratios, where for the latter, an enhanced production of charm baryon with respect to charm meson is observed at high multiplicity.

Due to its strangeness content, the D_s^+ meson is sensitive to the mechanisms of hadronisation in the quark-gluon plasma such as the quark coalescence mechanism. The results presented in this thesis contribute to these studies in small systems. In the heavy-flavour sector, the measurement of the D_s^+/D^0 ratio presented in this thesis does not allow for a firm conclusion on the strangeness enhancement as a function of multiplicity. However, similarities are observed between light- and heavy- flavour baryon over meson ratios.

The present results call for further improvements in the precision of the measurements and the extension of the explored phase space, which will be made possible thanks to the future upgrades of the ALICE experiment. The development of more performant analysis methods could also contribute to improving the accuracy of the measurement.

Perspectives

The use of a machine learning algorithm for the signal from background discrimination leads to improvements of the signal extraction which results in a better statistical precision of the measured cross section. At the same time, this approach offers the possibility to discriminate prompt from feed-down D mesons. In the ALICE Collaboration, this improved analysis technique allows for measuring the non-prompt D meson production in proton-proton collisions at $\sqrt{s}$= 5.02 TeV as was recently reported in [122]. This approach could be used for improving the evaluation of the relative contribution of non-prompt D_s^+ mesons, which is one of the dominant sources of systematic uncertainty in the results presented in this thesis.

During the second long shutdown of the LHC (2019 - 2022), different systems of the ALICE detector are upgraded, particularly, the Inner Tracking System. The six silicon layers of the Inner Tracking System (ITS) are replaced by the new 7 layers of silicon pixels (12.6 x 10^9 pixels) with the first layer closer to the beam pipe (2.34 cm versus 3.9 cm). Compared to the material budget of 1.14% X_0 per layer of the SPD detector, the material budget of the two innermost layers of the new ITS is reduced down to ∼0.3% X_0 and ∼0.9% X_0 in the middle and outer layers. This upgrade allows for much faster readout (up to a few 100 kHz in pp), an improved reconstruction of the collision vertex and heavy flavour decay vertices, as well as a better tracking of low-momentum particles [266]. This opens new possibilities for heavy-flavour hadron production measurements in proton-proton collisions.

Precise measurements of charm and beauty hadron production require an extended measured p_T range, down to $p_T = 0$.

Charm baryon production measurements in proton-proton collisions at the LHC challenge the universal fragmentation picture, commonly accepted as a dominating quark hadronisation mechanism. The measured Λ_c^+, $\Xi_c^{0,+}$ and $\Sigma_c^{0,+,++}$ production at mid-rapidity in proton-proton collisions [137, 138] are not described by the perturbative QCD calculations using fragmentation function tuned on e^+e^- data and models based on a fragmentation hadronisation mechanisms. The observed increase of charm baryon production in proton-proton collisions with respect to e^+e^- collisions, contributes to the understanding of its production via the colour reconnection mechanism, the coalescence mechanism and feed-down contribution from unobserved higher-mass charm baryon states.

Concerning the production of strange charmed mesons, precise measurements at higher multiplicities in small systems and at lower centrality in Pb-Pb collisions are needed to further investigate the observed hint of an enhanced strange charmed meson production in central Pb-Pb collisions [15]. A study of the D_s^+ meson production in jets, associated to hard scattering and fragmentation, and outside the jet cone, in the soft underlying event region, could provide additional insights on the dynamics of strange charmed meson production.

The results expected from the analysis of the LHC run III data will provide constraints for theoretical calculations in the charm sector and will allow for studying the beauty hadron production with a better than ever precision.

Bibliography

[1] L. Susskind, "Lattice models of quark confinement at high temperature," *Physical Review D*, vol. 20, pp. 2610–2618, 1979, DOI link

[2] Y. Aoki, G. Endrdi, Z. Fodor, S. D. Katz, and K. K. Szabó, "The order of the quantum chromodynamics transition predicted by the standard model of particle physics," *Nature*, vol. 443, p. 675678, 2006, DOI link, arXiv:hep-lat/0611014

[3] A. Bazavov *et al.*, "Equation of state in (2+1)-flavor QCD," *Phys. Rev. D*, vol. 90, p. 094503, Nov 2014, DOI link, arXiv:1407.6387

[4] T. Bhattacharya, "The QCD Equation of State," 2015, arXiv:1501.07652

[5] U. W. Heinz and M. Jacob, "Evidence for a new state of matter: An Assessment of the results from the CERN lead beam program," *CERN Announcement*, 1 2000, arXiv:nucl-th/0002042

[6] STAR Collaboration, "Experimental and theoretical challenges in the search for the quarkgluon plasma: The STAR Collaboration's critical assessment of the evidence from RHIC collisions," *Nuclear Physics A*, vol. 757, no. 1, pp. 102–183, 2005, DOI link, arXiv:nucl-ex/0501009

[7] PHENIX Collaboration, "Formation of dense partonic matter in relativistic nucleus-nucleus collisions at RHIC: Experimental evaluation by the PHENIX Collaboration," *Nuclear Physics A*, vol. 757, no. 1, pp. 184–283, 2005, DOI link, arXiv:nucl-ex/0410003

[8] PHOBOS Collaboration, "The PHOBOS perspective on discoveries at RHIC," *Nuclear Physics A*, vol. 757, no. 1, pp. 28–101, 2005, DOI link, arXiv:nucl-ex/0410022

[9] BRAHMS Collaboration, "Quark-gluon plasma and color glass condensate at RHIC? The perspective from the BRAHMS experiment," *Nuclear Physics A*, vol. 757, no. 1, pp. 1–27, 2005, DOI link, arXiv:nucl-ex/0410020

[10] CMS Collaboration, "Evidence for collectivity in pp collisions at the LHC," *Physics Letters B*, vol. 765, p. 193220, Feb 2017, DOI link, arXiv:1606.06198

[11] ALICE Collaboration, "Enhanced production of multi-strange hadrons in high-multiplicity protonproton collisions," *Nature Physics*, vol. 13, no. 6, p. 535539, Apr 2017, DOI link, arXiv:1606.07424

[12] L. Gladilin, "Fragmentation fractions of c and b quarks into charmed hadrons at LEP," *The European Physical Journal C*, vol. 75, no. 1, Jan 2015, DOI link, arXiv:1404.3888

[13] J. Rafelski and B. Muller, "Strangeness production in the Quark-Gluon Plasma," *Physical Review Letters*, vol. 48, p. 1066, 1982, DOI link, [Erratum: Phys.Rev.Lett. 56, 2334 (1986)].

[14] WA94 Collaboration, "Hyperon production in proton-sulphur collisions at 200 GeV/c," *Physics Letters B*, vol. 400, no. 1, pp. 239–244, 1997, DOI link

[15] ALICE Collaboration, "Measurement of D^0, D^+, D^{*+} and D_s^+ production in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV," *Journal of High Energy Pysics*, vol. 2018, no. 10, 2018, DOI link, arXiv:1804.09083

[16] Y. Ne'eman, "Derivation of strong interactions from a gauge invariance," *Nuclear Physics*, vol. 26, pp. 222–229, aug 1961, DOI link

[17] M. Gell-Mann, "The Eightfold Way: A Theory of strong interaction symmetry," 1961, DOI link

[18] G.-M. Murray, "Symmetries of Baryons and Mesons," *Physical Review*, vol. 125, pp. 1067–1084, feb 1962, DOI link

[19] M. Gell-Mann, "A schematic model of baryons and mesons," *Physics Letters*, vol. 8, pp. 214–215, 1964, DOI link

[20] G. Zweig, "An SU(3) model for strong interaction symmetry and its breaking. Version 2," in *Developments in the Quark Theory of Hadrons'. Vol. 1. 1964-1978*, D. B. Lichtenberg and S. P. Rosen, Eds. Nonantum, Usa: Hadronic Press, 1964.

[21] M. Gell-Mann, "The symmetry group of vector and axial vector currents," *Physics Physique Fizika*, vol. 1, pp. 63–75, 1964, DOI link

[22] J. D. Bjorken, "Inequality for Electron and Muon Scattering from Nucleons," *Physical Review Letters*, vol. 16, p. 408, 1966, DOI link

[23] J. D. Bjorken, "Asymptotic Sum Rules at Infinite Momentum," *Physical Review*, vol. 179, pp. 1547–1553, 1969, DOI link

[24] C. G. Callan, Jr. and D. J. Gross, "High-Energy Electroproduction and the Constitution of the Electric Current," *Physical Review Letters*, vol. 22, pp. 156–159, jan 1969, DOI link

[25] J. Prentki and J. Steinberger, Eds., *Proceedings, 14th International Conference on High-Energy Physics (ICHEP 68): Vienna, Austria, 28 Aug-5 Sep, 1968*. Geneva: CERN, 1968.

[26] E. D. Bloom *et al.*, "High-Energy Inelastic e-p Scattering at 6° and 10°," *Physical Review Letters*, vol. 23, pp. 930–934, 1969, DOI link

[27] R. Feynman, "The behavior of hadron collisions at extreme energies," *Conf. Proc. C*, vol. 690905, pp. 237–258, 1969.

[28] J. D. Bjorken and E. A. Paschos, "Inelastic Electron-Proton and γ-Proton Scattering and the Structure of the Nucleon," *Physical Review*, vol. 185, pp. 1975–1982, 1969, DOI link

[29] W. A. Bardeen, H. Fritzsch, and M. Gell-Mann, "Light cone current algebra, π^0 decay, and e^+e^- annihilation," in *Topical Meeting on the Outlook for Broken Conformal Symmetry in Elementary Particle Physics*, 1972, arXiv:hep-ph/0211388

[30] O. W. Greenberg, "Spin and Unitary Spin Independence in a Paraquark Model of Baryons and Mesons," *Physical Review Letters*, vol. 13, pp. 598–602, 1964, DOI link

[31] M. Y. Han and Y. Nambu, "Three-Triplet Model with Double SU(3) Symmetry," *Physical Review*, vol. 139, pp. B1006–B1010, 1965, DOI link

[32] H. Fritzsch and M. Gell-Mann, "Current algebra: Quarks and what else ?" *eConf*, vol. C720906V2, pp. 135–165, 1972, arXiv:hep-ph/0208010

[33] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, "Advantages of the Color Octet Gluon Picture," *Phys. Lett. B*, vol. 47, pp. 365–368, 1973, DOI link

[34] C. N. Yang and R. L. Mills, "Conservation of Isotopic Spin and Isotopic Gauge Invariance," *Physical Review*, vol. 96, pp. 191–195, Oct 1954, DOI link

[35] D. J. Gross and F. Wilczek, "Ultraviolet Behavior of Non-Abelian Gauge Theories," *Physical Review Letters*, vol. 30, pp. 1343–1346, 1973, DOI link

[36] H. Politzer, "Reliable Perturbative Results for Strong Interactions?" *Physical Review Letters*, vol. 30, pp. 1346–1349, 1973, DOI link

[37] P. D. Group, "Review of Particle Physics," *PTEP*, vol. 2020, no. 8, p. 083C01, 2020, DOI link

[38] H1 and ZEUS Collaborations, "Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data," *Eur. Phys. J. C*, vol. 75, no. 12, p. 580, 2015, DOI link, arXiv:1506.06042

[39] V. N. Gribov and L. N. Lipatov, "Deep inelastic e p scattering in perturbation theory," *Sov. J. Nucl. Phys.*, vol. 15, pp. 438–450, 1972.

[40] L. N. Lipatov, "The parton model and perturbation theory," *Sov. J. Nucl. Phys.*, vol. 20, pp. 94–102, 1975.

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

[42] 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*, vol. 46, pp. 641–653, 1977.

[43] K. G. Wilson, "Confinement of quarks," *Phys. Rev. D*, vol. 10, pp. 2445–2459, Oct 1974, DOI link

[44] D. G. Wolfgang Lucha, Franz F. Schöberl, "Bound states of quarks," *Physics Reports*, vol. 200, pp. 127–240, 1991, DOI link

[45] J. Goldstone, "Field theories with "Superconductor" solutions," *Il Nuovo Cimento (1955-1965)*, vol. 19, pp. 154–164, 1961, DOI link

[46] J. Goldstone, A. Salam, and S. Weinberg, "Broken Symmetries," *Phys. Rev.*, vol. 127, pp. 965–970, Aug 1962, DOI link

[47] C. Alexandrou and C. Kallidonis, "Low-lying baryon masses using $N_f = 2$ twisted mass clover-improved fermions directly at the physical pion mass," *Phys. Rev. D*, vol. 96, p. 034511, Aug 2017, DOI link, arXiv:1704.02647

[48] P. D. Group, "Review of Particle Physics," *Chin. Phys. C*, vol. 38, p. 090001, 2014, DOI link

[49] R. Hagedorn, "Statistical thermodynamics of strong interactions at high energies," *Nuovo Cimento, Suppl.*, vol. 3, no. CERN-TH-520, pp. 147–186, 1965.

[50] N. Cabibbo and G. Parisi, "Exponential hadronic spectrum and quark liberation," *Physics Letters B*, vol. 59, pp. 67–69, 1975, DOI link

[51] N. Itoh, "Hydrostatic Equilibrium of Hypothetical Quark Stars," *Progress of Theoretical Physics*, vol. 44, no. 1, pp. 291–292, 07 1970, DOI link

[52] J. C. Collins and M. J. Perry, "Superdense Matter: Neutrons Or Asymptotically Free Quarks," *Phys. Rev. Lett.*, vol. 34, p. 1353, 1975, DOI link

[53] B. C. Barrois, "Superconducting Quark Matter," *Nucl. Phys. B*, vol. 129, pp. 390–396, 1977, DOI link

[54] D. Bailin and A. Love, "Superfluidity and superconductivity in relativistic fermion systems," *Physics Reports*, vol. 107, no. 6, pp. 325–385, 1984, DOI link

[55] M. Alford, K. Rajagopal, and F. Wilczek, "Color-flavor locking and chiral symmetry breaking in high density QCD," *Nuclear Physics B*, vol. 537, no. 1-3, p. 443458, Jan 1999, DOI link, arXiv:hep-ph/9804403

[56] P. Braun-Munzinger, K. Redlich, and J. Stachel, "Particle production in heavy ion collisions," *QuarkGluon Plasma 3*, p. 491599, Jan 2004, DOI link, arXiv:nucl-th/0304013

[57] P. Braun-Munzinger, V. Koch, T. Schäfer, and J. Stachel, "Properties of hot and dense matter from relativistic heavy ion collisions," *Physics Reports*, vol. 621, pp. 76–126, 2016, DOI link, arXiv:1510.00442

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

[59] P. Braun-Munzinger and B. Dönigus, "Loosely-bound objects produced in nuclear collisions at the LHC," *Nuclear Physics A*, vol. 987, 04 2019, DOI link, arXiv:1809.04681

[60] F. Gelis, E. Iancu, J. Jalilian-Marian, and R. Venugopalan, "The Color Glass Condensate," *Ann. Rev. Nucl. Part. Sci.*, vol. 60, pp. 463–489, 2010, DOI link, arXiv:1002.0333

[61] M. L. Miller, K. Reygers, S. J. Sanders, and P. Steinberg, "Glauber Modeling in High-Energy Nuclear Collisions," *Annual Review of Nuclear and Particle Science*, vol. 57, no. 1, pp. 205–243, 2007, DOI link, arXiv:nucl-ex/0701025

[62] ALICE Collaboration, "Centrality determination in heavy ion collisions," 2018.

[63] S. Wheaton and J. Cleymans, "Statistical-thermal model calculations using THERMUS," *Journal of Physics G: Nuclear and Particle Physics*, vol. 31, no. 6, p. S1069S1073, May 2005, DOI link, arXiv:hep-ph/0412031

[64] A. Andronic, P. Braun-Munzinger, K. Redlich, and J. Stachel, "Decoding the phase structure of QCD via particle production at high energy," *Nature*, vol. 561, no. 7723, p. 321330, Sep 2018, DOI link, arXiv:1710.09425

[65] M. Petran, J. Letessier, J. Rafelski, and G. Torrieri, "SHARE with CHARM," *Computer Physics Communications*, vol. 185, no. 7, p. 20562079, Jul 2014, DOI link, arXiv:1310.5108

[66] WA85 Collaboration, "Enhancement of strange and multi-strange baryons and anti-baryons in SW interactions at 200 GeV/c," *Physics Letters B*, vol. 447, no. 1, pp. 178–182, 1999, DOI link

[67] WA94 Collaboration, "A study of cascade and strange baryon production in sulphur-sulphur interactions at 200 GeV/c per nucleon," *Physics Letters B*, vol. 354, no. 1, pp. 178–182, 1995, DOI link

[68] 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," *Computer Physics Communications*, vol. 191, p. 159177, 2015, DOI link, arXiv:1410.3012

[69] C. Bierlich, G. Gustafson, L. Lönnblad, and A. Tarasov, "Effects of Overlapping Strings in pp Collisions," *JHEP*, vol. 03, p. 148, 2015, DOI link, arXiv:1412.6259

[70] S. A. Voloshin, A. M. Poskanzer, and R. Snellings, "Collective Phenomena in Non-Central Nuclear Collisions," *Landolt-Börnstein - Group I Elementary Particles, Nuclei and Atoms*, vol. 23, pp. 293–333, 2010, copyright 2010 Springer-Verlag Berlin Heidelberg, DOI link, arXiv:0809.2949

[71] ALICE Collaboration, "Anisotropic flow of identified particles in Pb-Pb collisions at $\sqrt{s}_{\mathrm{NN}} =$ 5.02 TeV," *Journal of High Energy Physics*, vol. 2018, no. 9, DOI link, arXiv:1805.04390

[72] C. Gale, S. Jeon, B. Schenke, P. Tribedy, and R. Venugopalan, "Event-by-Event Anisotropic Flow in Heavy-ion Collisions from Combined Yang-Mills and Viscous Fluid Dynamics," *Physical Review Letters*, vol. 110, no. 1, Jan 2013, DOI link, arXiv:1209.6330

[73] CMS Collaboration, "Observation of long-range, near-side angular correlations in proton-proton collisions at the LHC," *Journal of High Energy Physics*, vol. 2010, no. 9, Sep 2010, DOI link, arXiv:1009.4122

[74] V. Pacík, "Elliptic flow of identified hadrons in small collisional systems measured with ALICE," *Nuclear Physics A*, vol. 982, p. 451454, Feb 2019, DOI link, arXiv:1807.04538

[75] PHENIX Collaboration, "Creation of quarkgluon plasma droplets with three distinct geometries," *Nature Physics*, vol. 15, no. 3, p. 214220, Dec 2018, DOI link, arXiv:1805.02973

[76] G.-Y. Qin and B. Müller, "Elliptic and triangular flow anisotropy in deuteron-gold collisions at $\sqrt{s_{\mathrm{NN}}} = 200$ GeV at RHIC and in proton-lead collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV at the LHC," *Physical Review C*, vol. 89, no. 4, Apr 2014, DOI link, arXiv:1306.3439

[77] M. Mace, V. V. Skokov, P. Tribedy, and R. Venugopalan, "Hierarchy of Azimuthal Anisotropy Harmonics in Collisions of Small Systems from the Color Glass Condensate," *Physical Review Letters*, vol. 121, no. 5, Jul 2018, DOI link, arXiv:1805.09342

[78] ALICE Collaboration, "Measurements of inclusive jet spectra in pp and central Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}}$=5.02 TeV," *Physical Review C*, vol. 101, no. 3, Mar 2020, DOI link, arXiv:1909.09718

[79] Y. He, T. Luo, X.-N. Wang, and Y. Zhu, "Linear Boltzmann transport for jet propagation in the quark-gluon plasma: Elastic processes and medium recoil," *Phys. Rev. C*, vol. 91, p. 054908, May 2015, DOI link, arXiv:1503.03313

[80] Y. He, S. Cao, W. Chen, T. Luo, L.-G. Pang, and X.-N. Wang, "Interplaying mechanisms behind single inclusive jet suppression in heavy-ion collisions," *Phys. Rev. C*, vol. 99, p. 054911, May 2019, DOI link, arXiv:1809.02525

[81] A. Idilbi and A. Majumder, "Extending soft collinear effective theory to describe hard jets in dense QCD media," *Phys. Rev. D*, vol. 80, p. 054022, Sep 2009, DOI link, arXiv:0808.1087

[82] G. Ovanesyan and I. Vitev, "An effective theory for jet propagation in dense QCD matter: jet broadening and medium-induced bremsstrahlung," *Journal of High Energy Physics*, vol. 80, 2011, DOI link, arXiv:1103.1074

[83] J. Casalderrey-Solana, D. C. Gulhan, J. G. Milhano, D. Pablos, and K. Rajagopal, "A hybrid strong/weak coupling approach to jet quenching," *Journal of High Energy Physics*, vol. 2014, no. 10, Oct 2014, DOI link, arXiv:1405.3864

[84] J. Casalderrey-Solana, D. C. Gulhan, J. G. Milhano, D. Pablos, and K. Rajagopal, "Predictions for boson-jet observables and fragmentation function ratios from a hybrid strong/weak coupling model for jet quenching," *Journal of High Energy Physics*, vol. 2016, no. 3, Mar 2016, DOI link, arXiv:1508.00815

[85] J. Casalderrey-Solana, D. C. Gulhan, J. G. Milhano, D. Pablos, and K. Rajagopal, "Angular structure of jet quenching within a hybrid strong/weak coupling model," *Journal of High Energy Physics*, vol. 2017, no. 3, Mar 2017, DOI link, arXiv:1609.05842

[86] Z. Hulcher, D. Pablos, and K. Rajagopal, "Resolution effects in the hybrid strong/weak coupling model," *Journal of High Energy Physics*, vol. 2018, no. 3, Mar 2018, DOI link, arXiv:1707.05245

[87] K. C. Zapp, F. Krauss, and U. A. Wiedemann, "A perturbative framework for jet quenching," *Journal of High Energy Physics*, vol. 2013, no. 3, Mar 2013, DOI link, arXiv:1212.1599

[88] R. K. Elayavalli and K. C. Zapp, "Simulating V+jet processes in heavy ion collisions with JEWEL," *The European Physical Journal C*, vol. 76, no. 12, Dec 2016, DOI link, arXiv:1608.03099

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

[90] B.-W. Zhang, C. M. Ko, and W. Liu, "Thermal charm production in a quark-gluon plasma in Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.5$TeV," *Physical Review C*, vol. 77, no. 2, Feb 2008, DOI link

[91] S. G. Collins J. C., Soper D.E., "Factorization of Hard Processes in QCD," *Advanced Series on Directions in High Energy Physics*, vol. 5, pp. 1–91, 1989, DOI link, arXiv:hep-ph/0409313

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

[93] H.-W. Lin, E. R. Nocera, F. Olness, K. Orginos, J. Rojo, A. Accardi, C. Alexandrou, A. Bacchetta, G. Bozzi, J.-W. Chen, and et al., "Parton distributions and lattice QCD calculations: A community white paper," *Progress in Particle and Nuclear Physics*, vol. 100, p. 107160, May 2018, DOI link, arXiv:1711.07916

[94] T. J. Hou, J. Gao, J. Hobbs, K. Xie, S. Dulat, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, and et al., "New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC," *Physical Review D*, vol. 103, no. 1, Jan 2021, DOI link, arXiv:1912.10053

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

[96] S. J. Brodsky, A. Kusina, F. Lyonnet, I. Schienbein, H. Spiesberger, and R. Vogt, "A Review of the Intrinsic Heavy Quark Content of the Nucleon," *Advances in High Energy Physics*, vol. 2015, p. 112, 2015, DOI link, arXiv:1504.06287

[97] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, and et al., "Parton distributions for the LHC run II," *Journal of High Energy Physics*, vol. 2015, no. 4, Apr 2015, DOI link, arXiv:1410.8849

[98] The NNPDF Collaboration and Rabah Abdul Khalek and Richard D. Ball and Stefano Carrazza and Stefano Forte and Tommaso Giani and Zahari Kassabov and Rosalyn L. Pearson and Emanuele R. Nocera and Juan Rojo and Luca Rottoli and Maria Ubiali and Cameron Voisey and Michael Wilson, "Parton Distributions with Theory Uncertainties: General Formalism and First Phenomenological Studies," 2019, arXiv:1906.10698

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

[100] S. Catani, S. Devoto, M. Grazzini, S. Kallweit, and J. Mazzitelli, "Bottom-quark production at hadron colliders: fully differential predictions in NNLO QCD," *Journal of High Energy Physics*, vol. 2021, no. 3, Mar 2021, DOI link, arXiv:2010.11906

[101] S. Catani, S. Devoto, M. Grazzini, S. Kallweit, and J. Mazzitelli, "Top-quark pair hadroproduction at NNLO: differential predictions with the $\overline{\text{MS}}$ mass," *Journal of High Energy Physics*, vol. 2020, no. 8, Aug 2020, DOI link

[102] C. Peterson, D. Schlatter, I. Schmitt, and P. M. Zerwas, "Scaling violations in inclusive e^+e^- annihilation spectra," *Phys. Rev. D*, vol. 27, pp. 105–111, Jan 1983, DOI link

[103] V. G. Kartvelishvili, A. K. Likhoded, and V. A. Petrov, "On the fragmentation functions of heavy quarks into hadrons," *Physics Letters B*, vol. 78, no. 5, pp. 615–617, 1978, DOI link

[104] P. D. B. Collins and T. P. Spiller, "The Fragmentation of Heavy Quarks," *J. Phys. G*, vol. 11, p. 1289, 1985, DOI link

[105] P. Nason and B. R. Webber, "Non-perturbative corrections to heavy quark fragmentation in e^+e^- annihilation," *Physics Letters B*, vol. 395, no. 3, pp. 355–363, 1997, DOI link, arXiv:hep-ph/9612353

[106] M. G. Bowler, "e^+e Production of heavy quarks in the string model," *Zeitschrift für Physik C Particles and Fields*, vol. 11, pp. 169–174, 1981, DOI link

[107] E. Braaten, K. Cheung, S. Fleming, and T. C. Yuan, "Perturbative QCD fragmentation functions as a model for heavy-quark fragmentation," *Physical Review D*, vol. 51, no. 9, p. 48194829, May 1995, DOI link, arXiv:9409316

[108] OPAL Collaboration, "A study of charmhadron production in $Z^0 \to c\bar{c}$ and $Z^0 \to b\bar{b}$ decays at LEP," *Z. Phys. C*, vol. 72, 1996, DOI link

[109] ALEPH Collaboration, "Study of charm production in Z decays," *Eur. Phys. J. C*, vol. 16, pp. 597–611, 2000, DOI link, arXiv:hep-ex/9909032

[110] DELPHI Collaboration, "Measurements of the Z partial decaywidth into $c\bar{c}$ and multiplicity of charm quarks per b decay," *Eur. Phys. J. C*, vol. 12, pp. 225–241, 2000, DOI link

[111] BaBar Collaboration, "Inclusive Λ_c^+ production in e^+e^- annihilations at $\sqrt{s} = 10.54$ GeV and in $\Upsilon(4S)$ decays," *Physical Review D*, vol. 75, p. 012003, 2007, DOI link, arXiv:hep-ex/0609004

[112] Belle Collaboration, "Charm hadrons from fragmentation and B decays in e^+e^- annihilation at $\sqrt{s} = 10.6$ GeV," *Physical Review D*, vol. 73, p. 032002, Feb 2006, DOI link

[113] B. Mele and P. Nason, "The Fragmentation function for heavy quarks in QCD," *Nucl. Phys. B*, vol. 361, pp. 626–644, 1991, DOI link, [Erratum: Nucl.Phys.B 921, 841–842 (2017)].

[114] J. Ma, "Perturbative prediction for parton fragmentation into heavy hadron," *Nuclear Physics B*, vol. 506, no. 1-2, p. 329347, Nov 1997, DOI link, arXiv:hep-ph/9705446

[115] B. A. Kniehl and G. Kramer, "Charmed-hadron fragmentation functions from CERN LEP1 revisited," *Physical Review D*, vol. 74, no. 3, Aug 2006, DOI link, arXiv:hep-ph/0607306

[116] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, "Parton fragmentation and string dynamics," *Physics Reports*, vol. 97, pp. 31–145, 1983, DOI link

[117] S. Ferreres-Solé and T. Sjöstrand, "The spacetime structure of hadronization in the Lund model," *The European Physical Journal C*, vol. 78, 2018, DOI link, arXiv:1808.04619

[118] J. R. Christiansen and P. Z. Skands, "String formation beyond leading colour," *Journal of High Energy Physics*, vol. 08, p. 003, 2015, DOI link, arXiv:1505.01681

[119] B. R. Webber, "A QCD model for jet fragmentation including soft gluon interference," *Nuclear Physics B*, vol. 238, pp. 492–528, 1984, DOI link

[120] G. Marchesini and B. R. Webber, "Monte Carlo simulation of general hard processes with coherent QCD radiation," *Nuclear Physics B*, vol. 310, pp. 461–526, 1988, DOI link

[121] J. Bellm, G. Bewick, S. Ferrario Ravasio, S. Gieseke, D. Grellscheid, P. Kirchgaeßer, F. Loshaj, M. R. Masouminia, G. Nail, A. Papaefstathiou, and et al., "Herwig 7.2 release note," *The European Physical Journal C*, vol. 80, 2020, DOI link

[122] ALICE Collaboration, "Measurement of beauty and charm production in pp collisions at $\sqrt{s} = 5.02$ TeV via non-prompt and prompt D mesons," 2021, arXiv:2102.13601

[123] M. Cacciari, M. Greco, and P. Nason, "The p_{T} spectrum in heavy-flavour hadroproduction," *Journal of High Energy Physics*, vol. 1998, no. 05, p. 007007, May 1998, DOI link, arXiv:hep-ph/9803400

[124] M. Cacciari, S. Frixione, and P. Nason, "The p_{T} spectrum in heavy-flavour photoproduction," *Journal of High Energy Physics*, vol. 2001, no. 03, p. 006006, Mar 2001, DOI link, arXiv:hep-ph/0102134

[125] M. Cacciari, S. Frixione, N. Houdeau, M. L. Mangano, P. Nason, and G. Ridolfi, "Theoretical predictions for charm and bottom production at the LHC," *Journal of High Energy Physics*, vol. 2012, no. 10, Oct 2012, DOI link, arXiv:1205.6344

[126] M. Benzke, M. V. Garzelli, B. A. Kniehl, G. Kramer, S. Moch, and G. Sigl, "Prompt neutrinos from atmospheric charm in the general-mass variable-flavor-number scheme," *Journal of High Energy Physics*, vol. 2017, no. 12, Dec 2017, DOI link, arXiv:1705.10386

[127] G. Kramer and H. Spiesberger, "Study of heavy meson production in p-Pb collisions at $\sqrt{s}=$ 5.02 TeV in the general-mass variable-flavour-number scheme," *Nucl. Phys. B*, vol. 925, pp. 415–430, 2017, DOI link, arXiv:1703.04754

[128] P. Bolzoni and G. Kramer, "Inclusive charmed-meson production from bottom hadron decays at the lhc," *Journal of Physics G: Nuclear and Particle Physics*, vol. 41, no. 7, p. 075006, May 2014, DOI link, arXiv:1310.2924

[129] CMS Collaboration, "Measurement of the $\mathrm{B}^{\pm}$ meson nuclear modification factor in Pb-Pb Collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV," *Physical Review Letters*, vol. 119, no. 15, Oct 2017, DOI link, arXiv:1705.04727

[130] CMS Collaboration, "Measurement of B_s^0 meson production in pp and PbPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV," *Physics Letters B*, vol. 796, p. 168190, Sep 2019, DOI link, arXiv:1810.03022

[131] P. D. Group, "Review of particle physics," *Chinese Physics C*, vol. 40, no. 10, p. 100001, Oct 2016, DOI link

[132] ALICE Collaboration, "Λ_{c}^+ production in pp and in p-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV," 2020, arXiv:2011.06079

[133] S. Frixione, G. Ridolfi, and P. Nason, "A positive-weight next-to-leading-order monte carlo for heavy flavour hadroproduction," *Journal of High Energy Physics*, vol. 2007, no. 09, p. 126126, Sep 2007, DOI link, arXiv:0707.3088

[134] T. Sjöstrand, S. Mrenna, and P. Skands, "Pythia 6.4 physics and manual," *Journal of High Energy Physics*, vol. 2006, no. 05, p. 026026, May 2006, DOI link, arXiv:hep-ph/0603175

[135] 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," *Physical Review D*, vol. 93, no. 3, Feb 2016, DOI link, arXiv:1506.07443

[136] P. Skands, S. Carrazza, and J. Rojo, "Tuning pythia 8.1: the monash 2013 tune," *The European Physical Journal C*, vol. 74, no. 8, Aug 2014, DOI link, arXiv:1404.5630

[137] ALICE Collaboration, "Measurement of Prompt D^0, Λ_c^+, and $\Sigma_c^{0,++}(2455)$ production in proton-proton collisions at $\sqrt{s} = 13$ TeV," *Physical Review Letters*, vol. 128, no. 1, Jan 2022, DOI link, arXiv:2106.08278

[138] ALICE Collaboration, "Measurement of the cross sections of Ξ_c^0 and Ξ_c^+ baryons and branching-fraction ratio $BR(\Xi_c^0 \to \Xi^- e^+ \nu_e)/BR(\Xi_c^0 \to \Xi^- \pi^+)$ in pp collisions at 13 TeV," 2021, arXiv:2105.05187

[139] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, "EPPS16: Nuclear parton distributions with LHC data," *Eur. Phys. J. C*, vol. 77, no. 3, p. 163, 2017, DOI link, arXiv:1612.05741

[140] J. J. Aubert *et al.*, "The ratio of the nucleon structure functions F_2^N for iron and deuterium," *Phys. Lett. B*, vol. 123, pp. 275–278, 1983, DOI link

[141] D. F. Geesaman, K. Saito, and A. W. Thomas, "The nuclear emc effect," *Annual Review of Nuclear and Particle Science*, vol. 45, no. 1, pp. 337–390, 1995, DOI link

[142] O. Hen, G. A. Miller, E. Piasetzky, and L. B. Weinstein, "Nucleon-nucleon correlations, short-lived excitations, and the quarks within," *Reviews of Modern Physics*, vol. 89, no. 4, Nov 2017, DOI link, arXiv:1611.09748

[143] P. Paakkinen, "Nuclear parton distribution functions," 2018, arXiv:1802.05927

[144] D. de Florian, R. Sassot, P. Zurita, and M. Stratmann, "Global analysis of nuclear parton distributions," *Physical Review D*, vol. 85, 2012, DOI link, arXiv:1112.6324

[145] L. N. Lipatov, "Reggeization of the Vector Meson and the Vacuum Singularity in Nonabelian Gauge Theories," *Sov. J. Nucl. Phys.*, vol. 23, pp. 338–345, 1976.

[146] E. A. Kuraev, L. N. Lipatov, and V. S. Fadin, "The Pomeranchuk Singularity in Nonabelian Gauge Theories," *Sov. Phys. JETP*, vol. 45, pp. 199–204, 1977.

[147] I. I. Balitsky and L. N. Lipatov, "The Pomeranchuk Singularity in Quantum Chromodynamics," *Sov. J. Nucl. Phys.*, vol. 28, pp. 822–829, 1978.

[148] B. Z. Kopeliovich, J. Nemchik, A. Schäfer, and A. V. Tarasov, "Cronin effect in hadron production off nuclei," *Physical Review Letters*, vol. 88, no. 23, May 2002, DOI link, arXiv:hep-ph/0201010

[149] M. Lev and B. Petersson, "Nuclear effects at large transverse momentum in a qcd parton model," *Z. Phys. C*, vol. 21, p. 155, 1983, DOI link

[150] I. Vitev, T. Goldman, M. B. Johnson, and J. W. Qiu, "Open charm tomography of cold nuclear matter," *Physical Review D*, vol. 74, 2006, DOI link

[151] I. Vitev, "Non-Abelian energy loss in cold nuclear matter," *Physical Review C*, vol. 75, 2007, DOI link, arXiv:0703002

[152] A. Accardi, "Cronin effect in proton-nucleus collisions: a survey of theoretical models," 2002, arXiv:hep-ph/0212148

[153] ALICE Collaboration, "Measurement of prompt D^0, D^+, D^{*+}, and D_s^+ production in pPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV at the LHC," *Journal of High Energy Physics*, vol. 2019, no. 12, Dec 2019, DOI link, arXiv:1906.03425

[154] H. Fujii and K. Watanabe, "Nuclear modification of forward D production in pPb collisions at the LHC," 2018, arXiv:1706.06728

[155] R. Sharma, I. Vitev, and B.-W. Zhang, "Light-cone wave function approach to open heavy flavor dynamics in QCD matter," *Physical Review C*, vol. 80, no. 5, Nov 2009, DOI link, arXiv:0904.0032

[156] Z.-B. Kang, I. Vitev, E. Wang, H. Xing, and C. Zhang, "Multiple scattering effects on heavy meson production in p+A collisions at backward rapidity," *Physics Letters B*, vol. 740, p. 2329, Jan 2015, DOI link, arXiv:1409.2494

[157] A. Beraudo, A. De Pace, M. Monteno, M. Nardi, and F. Prino, "Heavy-flavour production in high-energy d-Au and p-Pb collisions," *Journal of High Energy Physics*, vol. 2016, no. 3, Mar 2016, DOI link, arXiv:1512.05186

[158] G. D. Moore and D. Teaney, "How much do heavy quarks thermalize in a heavy ion collision?" *Physical Review C*, vol. 71, no. 6, Jun 2005, DOI link, arXiv:0412346

[159] S. Cao and S. A. Bass, "Thermalization of charm quarks in infinite and finite quark-gluon plasma matter," *Phys. Rev. C*, vol. 84, p. 064902, Dec 2011, DOI link

[160] F. Prino and R. Rapp, "Open heavy flavor in QCD matter and in nuclear collisions," vol. 43, no. 9, p. 093002, aug 2016, DOI link

[161] R. Rapp and H. V. Hees, "Heavy quarks in the quark-gluon palsma," *Quark-Gluon Plasma 4*, p. 111206, Feb 2010, DOI link, arXiv:0903.1096

[162] B. L. Combridge, "Associated Production of Heavy Flavor States in p p and anti-p p Interactions: Some QCD Estimates," *Nucl. Phys. B*, vol. 151, pp. 429–456, 1979, DOI link

[163] B. Svetitsky, "Diffusion of charmed quarks in the quark-gluon plasma," *Phys. Rev. D*, vol. 37, pp. 2484–2491, May 1988, DOI link

[164] M. G. Mustafa, D. Pal, and D. K. Srivastava, "Propagation of charm quarks in equilibrating quark-gluon plasma," *Phys. Rev. C*, vol. 57, pp. 889–898, Feb 1998, DOI link

[165] S. Cao, G.-Y. Qin, and S. A. Bass, "Heavy-quark dynamics and hadronization in ultrarelativistic heavy-ion collisions: Collisional versus radiative energy loss," *Physical Review C*, vol. 88, no. 4, Oct 2013, DOI link, arXiv:1308.0617

[166] N. Armesto, B. Cole, C. Gale, W. A. Horowitz, P. Jacobs, S. Jeon, M. van Leeuwen, A. Majumder, B. Müller, G.-Y. Qin, and et al., "Comparison of jet quenching formalisms for a quark-gluon plasma brick," *Physical Review C*, vol. 86, no. 6, Dec 2012, DOI link, arXiv:1106.1106

[167] R. Baier, Y. Dokshitzer, S. Peigné, and D. Schiff, "Induced gluon radiation in a qcd medium," *Physics Letters B*, vol. 345, no. 3, p. 277286, Feb 1995, DOI link, arXiv:hep-ph/9411409

[168] R. Baier, Y. Dokshitzer, A. Mueller, S. Peigné, and D. Schiff, "Radiative energy loss of high energy quarks and gluons in a finite-volume quark-gluon plasma," *Nuclear Physics B*, vol. 483, no. 1-2, p. 291320, Jan 1997, DOI link, arXiv:hep-ph/9607355

[169] R. Baier, Y. Dokshitzer, A. Mueller, S. Peigné, and D. Schiff, "Radiative energy loss and p-broadening of high energy partons in nuclei," *Nuclear Physics B*, vol. 484, no. 1-2, p. 265282, Jan 1997, DOI link, arXiv:hep-ph/9608322

[170] Y. Dokshitzer and D. Kharzeev, "Heavy-quark colorimetry of qcd matter," *Physics Letters B*, vol. 519, no. 3-4, p. 199206, Nov 2001, DOI link, arXiv:hep-ph/0106202

[171] ALICE Collaboration, "Centrality dependence of high-p_T D meson suppression in Pb-Pb collisions at $\sqrt{s_{\rm NN}}$ = 2.76 TeV," *Journal of High Energy Physics*, vol. 2015, no. 11, Nov 2015, DOI link, arXiv:1506.06604

[172] "Addendum to JHEP 11 (2015) 205: Centrality dependence of high-$p_{\rm T}$ D-meson suppression in Pb-Pb collisions at $\sqrt{s_{\rm NN}}$ = 2.76TeV," May 2017. [Online]. Available: https://cds.cern.ch/record/2265113

[173] CMS Collaboration, "Suppression and azimuthal anisotropy of prompt and nonprompt J/Ψ production in PbPb collisions at $\sqrt{s_{\rm NN}}$ = 2.76TeV," *The European Physical Journal C*, vol. 77, no. 4, Apr 2017, DOI link, arXiv:1610.00613

[174] ALICE Collaboration, "D-meson Azimuthal Anisotropy in Midcentral Pb-Pb collisions at $\sqrt{s_{\rm NN}}$ = 5.02 TeV," *Physical Review Letters*, vol. 120, no. 10, p. 102301, 2018, DOI link, arXiv:1707.01005

[175] S. Batsouli, S. Kelly, M. Gyulassy, and J. Nagle, "Does the charm flow at rhic?" *Physics Letters B*, vol. 557, no. 1-2, p. 2632, Mar 2003, DOI link, arXiv:nucl-th/0212068

[176] M. Gyulassy, I. Vitev, and X.-N. Wang, "Highptazimuthal asymmetry in noncentrala+aat rhic," *Physical Review Letters*, vol. 86, no. 12, p. 25372540, Mar 2001, DOI link, arXiv:nucl-th/0012092

[177] E. V. Shuryak, "Azimuthal asymmetry at largeptseem to be too large for a pure jet quenching," *Physical Review C*, vol. 66, no. 2, Aug 2002, DOI link, arXiv:nucl-th/0112042

[178] ALICE Collaboration, "Two-pion boseeinstein correlations in central pb-pb collisions at $\sqrt{s_{\rm NN}}$ = 2.76tev," *Physics Letters B*, vol. 696, no. 4, p. 328337, Feb 2011, DOI link, arXiv:1012.4035

[179] M. He, R. J. Fries, and R. Rapp, "Heavy flavor at the large hadron collider in a strong coupling approach," *Physics Letters B*, vol. 735, p. 445450, Jul 2014, DOI link, arXiv:1401.3817

[180] A. Andronic, F. Arleo, R. Arnaldi, A. Beraudo, E. Bruna, D. Caffarri, Z. C. del Valle, J. G. Contreras, T. Dahms, A. Dainese, and et al., "Heavy-flavour and quarkonium production in the lhc era: from protonproton to heavy-ion collisions," *The European Physical Journal C*, vol. 76, no. 3, Feb 2016, DOI link, arXiv:1506.03981

[181] T. Song, H. Berrehrah, D. Cabrera, W. Cassing, and E. Bratkovskaya, "Charm production in Pb + Pb collisions at energies available at the CERN Large Hadron Collider," *Physical Review C*, vol. 93, no. 3, Mar 2016, DOI link, arXiv:1512.00891

[182] A. Beraudo, A. De Pace, M. Monteno, M. Nardi, and F. Prino, "Heavy flavors in heavy-ion collisions: quenching, flow and correlations," *The European Physical Journal C*, vol. 75, no. 3, Mar 2015, DOI link, arXiv:1410.6082

[183] M. Nahrgang, J. Aichelin, P. B. Gossiaux, and K. Werner, "Influence of hadronic bound states above T_c on heavy-quark observables in Pb+Pb collisions at the CERN Large Hadron Collider," *Physical Review C*, vol. 89, no. 1, Jan 2014, DOI link, arXiv:1305.6544

[184] S. Cao, T. Luo, G.-Y. Qin, and X.-N. Wang, "Heavy and light flavor jet quenching at RHIC and LHC energies," *Physics Letters B*, vol. 777, p. 255259, Feb 2018, DOI link, arXiv:1703.00822

[185] Y. Xu, M. Nahrgang, J. E. Bernhard, S. Cao, and S. A. Bass, "A data-driven analysis of the heavy quark transport coefficient," *Nuclear Physics A*, vol. 967, p. 668671, Nov 2017, DOI link, arXiv:1704.07800

[186] J. Uphoff, O. Fochler, Z. Xu, and C. Greiner, "Elastic and radiative heavy quark interactions in ultra-relativistic heavy-ion collisions," *Journal of Physics G: Nuclear and Particle Physics*, vol. 42, no. 11, p. 115106, Oct 2015, DOI link, arXiv:1408.2964

[187] S. Borsányi, Z. Fodor, C. Hoelbling, S. D. Katz, S. Krieg, C. Ratti, and K. K. Szabó, "Is there still any T c mystery in lattice QCD? Results with physical masses in the continuum limit III," *Journal of High Energy Physics*, vol. 2010, no. 9, Sep 2010, DOI link, arXiv:1005.3508

[188] CMS Collaboration, "Observation of the B_c^+ meson in PbPb and pp collisions at $\sqrt{s_{NN}} = 5.02$ TeV," 2022, arXiv:2201.02659

[189] R. J. Fries, B. Müller, C. Nonaka, and S. A. Bass, "Hadron production in heavy ion collisions: Fragmentation and recombination from a dense parton phase," *Physical Review C*, vol. 68, no. 4, Oct 2003, DOI link, arXiv:nucl-th/0306027

[190] L. Ravagli and R. Rapp, "Quark coalescence based on a transport equation," *Physics Letters B*, vol. 655, no. 3-4, p. 126131, Nov 2007, DOI link, arXiv:0705.0021

[191] 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," *The European Physical Journal C*, vol. 78, no. 4, Apr 2018, DOI link, arXiv:1712.00730

[192] R. Rapp, P. Gossiaux, A. Andronic, R. Averbeck, S. Masciocchi, A. Beraudo, E. Bratkovskaya, P. Braun-Munzinger, S. Cao, A. Dainese, and et al., "Extraction of heavy-flavor transport coefficients in qcd matter," *Nuclear Physics A*, vol. 979, p. 2186, Nov 2018, DOI link, arXiv:1803.03824

[193] 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," 2021, arXiv:2104.12754

[194] M. He and R. Rapp, "Hadronization and charm-hadron ratios in heavy-ion collisions," *Physical Review Letters*, vol. 124, no. 4, Jan 2020, DOI link, arXiv:1905.09216

[195] M. He and R. Rapp, "Charm-baryon production in proton-proton collisions," *Physics Letters B*, vol. 795, p. 117121, Aug 2019, DOI link, arXiv:1902.08889

[196] D. Ebert, R. N. Faustov, and V. O. Galkin, "Spectroscopy and regge trajectories of heavy baryons in the relativistic quark-diquark picture," *Physical Review D*, vol. 84, no. 1, Jul 2011, DOI link, arXiv:1105.0583

[197] ALICE Collaboration, "Measurement of prompt D_s^+-meson production and azimuthal anisotropy in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV," 2021, arXiv:2110.10006

[198] T. Song, H. Berrehrah, D. Cabrera, J. M. Torres-Rincon, L. Tolos, W. Cassing, and E. Bratkovskaya, "Tomography of the quark-gluon plasma by charm quarks," *Physical Review C*, vol. 92, no. 1, Jul 2015, DOI link, arXiv:1503.03039

[199] S. Li and J. Liao, "Data-driven extraction of heavy quark diffusion in quark-gluon plasma," *The European Physical Journal C*, vol. 80, no. 7, Jul 2020, DOI link, arXiv:1912.08965

[200] R. Aaij, C. Abellán Beteta, B. Adeva, M. Adinolfi, C. Aidala, Z. Ajaltouni, S. Akar, P. Albicocco, J. Albrecht, F. Alessio, M. Alexander, A. Alfonso Albero, G. Alkhazov, and et al., "Measurement of f_s/f_u variation with proton-proton collision energy and B-meson kinematics," *Physical Review Letters*, vol. 124, no. 12, Mar 2020, DOI link, arXiv:1910.09934

[201] C. Collaboration, "Observation of b_s^0 mesons and measurement of the b_s^0/b^+ yield ratio in pbpb collisions at $\sqrt{s_{NN}} = 5.02$ tev," 2021, arXiv:2109.01908

[202] S. Cao, K.-J. Sun, S.-Q. Li, S. Y. Liu, W.-J. Xing, G.-Y. Qin, and C. M. Ko, "Charmed hadron chemistry in relativistic heavy-ion collisions," *Physics Letters B*, vol. 807, p. 135561, Aug 2020, DOI link, arXiv:1911.00456

[203] V. N. Gribov, "Asymptotic behaviour of the scattering amplitude at high energies," *Nucl. Phys.*, vol. 22, pp. 249–261, 1961, DOI link

[204] L. Frankfurt, M. Strikman, and C. Weiss, "Dijet production as a centrality trigger forppcollisions at CERN LHC," *Physical Review D*, vol. 69, no. 11, Jun 2004, DOI link, arXiv:hep-ph/0311231

[205] L. Frankfurt, M. Strikman, and C. Weiss, "Transverse nucleon structure and diagnostics of hard parton-parton processes at LHC," *Physical Review D*, vol. 83, no. 5, Mar 2011, DOI link, arXiv:1009.2559 [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.83.054012

[206] M. Azarkin, I. Dremin, and M. Strikman, "Jets in multiparticle production in and beyond geometry of protonproton collisions at the LHC," *Physics Letters B*, vol. 735, pp. 244–249, 2014, DOI link

[207] M. Aguilar-Benitez *et al.*, "Comparative Properties of 400-GeV/c Proton - Proton Interactions With and Without Charm Production," *Z. Phys. C*, vol. 41, p. 191, 1988, DOI link

[208] A. Collaboration, "Measurement of charm and beauty production at central rapidity versus charged-particle multiplicity in proton-proton collisions at $\sqrt{s} = 7$ TeV," *Journal of High Energy Physics*, no. 9, 2015, DOI link, arXiv:1505.00664

[209] E. G. Ferreiro and C. Pajares, "High multiplicity pp events and J/ψ production at energies available at the CERN Large Hadron Collider," *Physical Review C*, vol. 86, p. 034903, 2012, DOI link, arXiv:1203.5936

[210] E. G. Ferreiro and C. Pajares, "Open charm production in high multiplicity proton-proton events at the LHC," 2015, arXiv:1501.03381

[211] K. Werner, B. Guiot, I. Karpenko, and T. Pierog, "Analyzing radial flow features in p-Pb and p-p collisions at several TeV by studying identified-particle production with the event generator EPOS3," *Physical Review C*, vol. 89, no. 6, 2014, DOI link, arXiv:1312.1233

[212] H. Drescher, M. Hladik, S. Ostapchenko, T. Pierog, and K. Werner, "Parton-based GribovRegge theory," *Physics Reports*, vol. 350, p. 93289, 2001, DOI link, arXiv:0007198

[213] CMS Collaboration, "Studies of charm and beauty hadron long-range correlations in pp and pPb collisions at LHC energies," *Physics Letters B*, vol. 813, p. 136036, 2021, DOI link, arXiv:2009.07065

[214] L. Evans and P. Bryant, "LHC machine," *Journal of Instrumentation*, vol. 3, no. 08, pp. S08 001–S08 001, aug 2008, DOI link

[215] "LHC Guide," Mar 2017. [Online]. Available: http://cds.cern.ch/record/2255762

[216] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Physics Letters B*, vol. 716, no. 1, p. 129, Sep 2012, DOI link, arXiv:1610.03055v2

[217] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Physics Letters B*, vol. 716, no. 1, p. 3061, Sep 2012, DOI link

[218] ALICE Collaboration, "The ALICE experiment at the CERN LHC," *Journal of Instrumentation*, vol. 3, no. 08, pp. S08 002–S08 002, aug 2008, DOI link

[219] ALICE Collaboration, "Performance of the ALICE experiment at the CERN LHC," *International Journal of Modern Physics A*, vol. 29, no. 24, p. 1430044, Sep 2014, DOI link, arXiv:1402.4476v4

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

[221] ALICE Collaboration, "Performance of the ALICE experiment at the CERN LHC," *International Journal of Modern Physics A*, vol. 29, no. 24, p. 1430044, Sep 2014, DOI link, arXiv:1402.4476

[222] R. Frühwirth, "Application of Kalman filtering to track and vertex fitting," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 262, no. 2, pp. 444–450, 1987, DOI link

[223] "Track reconstruction in high density environment," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 566, no. 1, pp. 70–74, 2006, tIME 2005.

[224] ALICE Collaboration. ALICE Figure repository. [Online]. Available: https://alice-figure.web.cern.ch/

[225] W. Blum and W. R. L. Rolandi, *Particle Detection with Drift Chambers*. Berlin, Heidelberg: Springer, 2008, DOI link

[226] ALICE Collaboration. AliRoot. [Online]. Available: https://alice-offline.web.cern.ch/

[227] R. Brun and F. Rademakers, "ROOT An object oriented data analysis framework," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 389, no. 1, pp. 81–86, 1997, new Computing Techniques in Physics Research V.

[228] ALICE Collaboration. AliRoot. [Online]. Available: https://alice-o2-project.web.cern.ch/

[229] R. Brun, F. Bruyant, F. Carminati, S. Giani, M. Maire, A. McPherson, G. Patrick, and L. Urban, *GEANT: Detector Description and Simulation Tool; Oct 1994*, ser. CERN Program Library. Geneva: CERN, 1993, long Writeup W5013. [Online]. Available: https://cds.cern.ch/record/1082634

[230] GEANT4 Collaboration, "GEANT4. A Simulation toolkit," *Nucl. Instrum. Methods Phys. Res., A*, vol. 506, pp. 250–303. 54 p, Jul 2002, DOI link

[231] A. Ferrari, P.R. Sala, A. Fasso, J. Ranft. FLUKA: A multi-particle transport code. [Online]. Available: https://inspirehep.net/files/e8cb3847561a920327bd8b03bc518617

[232] Worldwide LHC Computing Grid. [Online]. Available: https://wlcg.web.cern.ch/

[233] P. Antonioli, A. Kluge, and W. Riegler, "Upgrade of the ALICE Readout & Trigger System," Tech. Rep., Sep 2013. [Online]. Available: https://cds.cern.ch/record/1603472

[234] ALICE Collaboration, "Performance of the ALICE VZERO system," *Journal of Instrumentation*, vol. 8, no. 10, pp. P10 016–P10 016, oct 2013.

[235] "The ALICE definition of primary particles," Jun 2017. [Online]. Available: https://cds.cern.ch/record/2270008

[236] ALICE Collaboration, "Pseudorapidity distributions of charged particles as a function of mid- and forward rapidity multiplicities in pp collisions at $\sqrt{s} = 5.02$, 7 and 13 TeV," *The European Physical Journal C*, vol. 81, no. 7, Jul 2021, DOI link, arXiv:2009.09434

[237] P. D. Group, "Review of Particle Physics," *Phys. Rev. D*, vol. 98, p. 030001, Aug 2018, DOI link

[238] T. Mitchell, *Machine Learning*, ser. McGraw-Hill International Editions. McGraw-Hill, 1997.

[239] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug 2016, DOI link, arXiv:1603.02754

[240] E. Brochu, V. M. Cora, and N. de Freitas, "A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," 2010, arXiv:1012.2599

[241] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," 2017, arXiv:1705.07874

[242] E. Strumbelj and I. Kononenko, "Explaining prediction models and individual predictions with feature contributions," *Knowledge and Information Systems*, vol. 41, pp. 647–665, 2013, DOI link

[243] ALICE Collaboration, "Measurement of D-meson production at mid-rapidity in pp collisions at $\sqrt{s} = 7$ TeV," *The European Physical Journal C*, vol. 77, no. 8, Aug 2017, DOI link, arXiv:1702.00766

[244] LHCb Collaboration, "Measurement of $D_s^{\pm}$ production asymmetry in pp collisions at $\sqrt{s}$ = 7 and 8 TeV," *Journal of High Energy Physics*, vol. 2018, no. 8, Aug 2018, DOI link, arXiv:1805.09869

[245] A. M. Barbano, "PhD thesis: Prompt D_s^+ meson production in pp, p-Pb and Pb-Pb collisions at LHC with ALICE," Dec 2018. [Online]. Available: http://cds.cern.ch/record/2309930

[246] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, D. Stump, W.-K. Tung, and C.-P. Yuan, "Implications of cteq global analysis for collider observables," *Physical Review D*, vol. 78, no. 1, Jul 2008, DOI link, arXiv:0802.0007

[247] "ALICE 2016-2017-2018 luminosity determination for pp collisions at $\sqrt{s} = 13$ TeV," Jul 2021. [Online]. Available: https://cds.cern.ch/record/2776672

[248] ALICE Collaboration, "D_s^+ meson production at central rapidity in proton-proton collisions at $\sqrt{s} = 7$ TeV," *Physics Letters B*, vol. 718, no. 2, p. 279294, Dec 2012, DOI link, arXiv:1208.1948

[249] A. Collaboration, "Measurement of $D^{*\pm}$, $D^{*\pm}$ and $D_s^{\pm}$ meson production cross sections in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector," *Nuclear Physics B*, vol. 907, p. 717763, Jun 2016, DOI link, arXiv:1512.02913

[250] H. Collaboration, "Inclusive production of D^+, D^0, D_s^+ and D^{*+} mesons in deep inelastic scattering at HERA," *The European Physical Journal C*, vol. 38, no. 4, p. 447459, Jan 2005, DOI link, arXiv:hep-ex/0408149

[251] ZEUS Collaboration, "Measurement of charm fragmentation fractions in photoproduction at HERA," 2013, arXiv:1306.4862

[252] M. Cacciari, M. L. Mangano, and P. Nason, "Gluon PDF constraints from the ratio of forward heavy quark production at the LHC at $\sqrt{s} = 7$ and 13 TeV," *Eur. Phys. J. C*, vol. 75, p. 610. 21 p, Jul 2015, DOI link, arXiv:1507.06197

[253] ALICE Collaboration, "Measurement of pion, kaon and proton production in protonproton collisions at $\sqrt{s} = 7$ TeV," *The European Physical Journal C*, vol. 75, no. 5, May 2015, DOI link, arXiv:1504.00024

[254] CMS Collaboration, "Investigation into the event-activity dependence of Υ(nS) relative production in proton-proton collisions at $\sqrt{s} = 7$ TeV," *Journal of High Energy Physics*, vol. 2020, no. 11, Nov 2020, DOI link, arXiv:2007.04277

[255] ALICE Collaboration, "Multiplicity dependence of π, k, and p production in pp collisions at *sqrts* = 13 tev," *The European Physical Journal C*, vol. 80, no. 8, Aug 2020, DOI link, arXiv:2003.02394v2

[256] O. Behnke, K. Kröninger, T. Schörner-Sadenius, and G. Schott, Eds., *Data analysis in high energy physics: A practical guide to statistical methods*. Weinheim, Germany: Wiley-VCH, 2013.

[257] A. Ortiz, G. Bencédi, and H. Bello, "Revealing the source of the radial flow patterns in protonproton collisions using hard probes," *Journal of Physics G: Nuclear and Particle Physics*, vol. 44, no. 6, p. 065001, Apr 2017, DOI link, arXiv:1608.04784v2

[258] ALICE Collaboration, "Azimuthal correlations of prompt D mesons with charged particles in pp and p-Pb collisions at $\sqrt{s} = 5.02$ TeV," 2019, arXiv:1910.14403

[259] ALICE Collaboration, "Multiplicity dependence of (multi-)strange hadron production in proton-proton collisions at $\sqrt{s} = 13$ TeV," *The European Physical Journal C*, vol. 80, no. 2, Feb 2020, DOI link, arXiv:1908.01861

[260] K. Werner, B. Guiot, I. Karpenko, T. Pierog, and G. Sophys, "Charm production in high multiplicity pp events," *Journal of Physics: Conference Series*, vol. 736, p. 012009, Aug 2016, DOI link, arXiv:1602.03414v1

[261] S. G. Weber, A. Dubla, A. Andronic, and A. Morsch, "Elucidating the multiplicity dependence of J/ψ production in protonproton collisions with PYTHIA8," *The European Physical Journal C*, vol. 79, no. 1, Jan 2019, DOI link, arXiv:1811.07744

[262] Y. Chen and M. He, "Charged-particle multiplicity dependence of charm-baryon-to-meson ratio in high-energy proton-proton collisions," *Physics Letters B*, vol. 815, p. 136144, Apr 2021, DOI link, arXiv:2011.14328

[263] ALICE Collaboration, "Multiplicity dependence of light-flavor hadron production in pp collisions at $\sqrt{s} = 7$ TeV," *Physical Review C*, vol. 99, no. 2, Feb 2019, DOI link, arXiv:1807.11321v2

[264] ALICE Collaboration, "Production of Λ and K^0_S in jets in p-Pb collisions at $\sqrt{s_{NN}} = 5$ TeV and pp collisions at $\sqrt{s} = 7$ TeV," 2021, arXiv:2105.04890

[265] ALICE Collaboration, "Measurement of D^0, D^+, D^{*+} and D_s^+ production in pp collisions at $\sqrt{s} = 5.02$ TeV with ALICE," *The European Physical Journal C*, vol. 79, no. 5, May 2019, DOI link, arXiv:1901.07979

[266] ALICE Collaboration, "Technical Design Report for the Upgrade of the ALICE Inner Tracking System," *J. Phys. G*, vol. 41, p. 087002, 2014, DOI link

Arthur Gal

Charmed-strange meson production in proton-proton collisions with the ALICE experiment at the LHC

Abstract

The thesis was performed within the ALICE experiment, which is dedicated to the study of strongly interacting matter physics and the quark-gluon plasma (QGP) in ultra-relativistic heavy-ion collisions at the LHC at CERN. The thesis aims at the study of the production of charmed-strange D_S^+ meson, composed of a charm and a strange quark, in proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV. The measurement of the p_T-differential production cross section of D_S^+ mesons at mid-rapidity ($|y| < 0.5$) is presented and discussed in comparison with non-strange D meson measurements and with predictions from theory.
A study of the D_S^+ meson production as a function of the charged-particle density is presented. The result of this study contribute to the characterisation of QGP-like effects in proton-proton collisions by probing strangeness enhancement with the production of heavy-flavour hadrons.

Keywords : quark-gluon plasma, strong interaction, heavy flavour, D mesons, ALICE, LHC, proton–proton collisions, production cross-section, production yield as a function of multiplicity

Résumé

La thèse a été effectuée dans le cadre de l'expérience ALICE dédiée à l'étude de la chromodynamique quantique et particulièrement du plasma de quarks et de gluons (PQG) auprès du collisioneur LHC du CERN. Elle porte sur l'étude de la production du méson charmé-étrange D_S^+, composé d'un quark charmé et d'un quark étrange, dans les collisions proton-proton à une énergie dans le centre de masse $\sqrt{s} = 13$ TeV. La thèse présente la mesure de la section efficace différentielle de production des mésons D_S^+ à rapidité centrale ($|y| < 0.5$) en fonction de l'impulsion transverse. Cette mesure est discutée et comparée aux mesures des mésons D non-étranges et aux prédictions de modèles théoriques. Une étude détaillée de la production des mésons D_S^+ en fonction de la multiplicité en particules chargées de la collision est présentée. Cette mesure contribue à la caractérisation des effets collectifs, de type PQG, dans les collisions proton-proton via le prisme de l'augmentation de l'étrangeté.

Mots clés : plasma de quarks et de gluons, interaction forte, saveurs lourdes, mésons D, ALICE, LHC, collisions proton–proton, section efficace de production, taux de production en fonction de la multiplicité