Dissertation

submitted to the

Combined Faculty of Natural Sciences and Mathematics

for the degree of

Doctor of Natural Sciences

Put forward by

Benedikt Völkel

born in Hildesheim, Germany

Oral examination on June 30th, 2021

Λ_c^+ production in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ALICE experiment and Virtual Monte Carlo developments for LHC Run 3 and 4

Referees: Prof. Dr. Johanna Stachel
Prof. Dr. Ulrich Uwer

Abstract

The production of prompt Λ_c^+ hadrons at midrapidity $|y| < 0.5$ in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ as a function of charged-particle multiplicity is presented. Also analogous measurements of D^0 and D_s^+ production are currently being carried out and the D_s^+/D^0 as well as the Λ_c^+/D^0 production ratios are discussed. In contrast to the D_s^+/D^0 production ratio, the Λ_c^+/D^0 production ratio shows an increase with increasing multiplicity. In addition, its ratio is significantly underestimated by predictions where fragmentation is tuned to e^+e^- and e^-p measurements. On the other hand, comparisons with a canonical approach of a statistical hadronisation model with augmented baryon production as well as with a fragmentation model implementing colour reconnection beyond leading colour approximation are shown to qualitatively describe the shape and multiplicity dependence. The Λ_c^+/D^0 ratio is similar in shape and magnitude compared to the light-flavour ratio Λ/K_S^0 measured at comparable multiplicities. Together, these results indicate (i) hadronisation mechanisms beyond pure in-vacuum fragmentation in pp collisions as well as (ii) potential common mechanisms of baryon formation in the light-flavour and heavy-flavour sector.

-

The Virtual Monte Carlo (VMC) detector simulation framework which is used by the ALICE and FAIR collaboration has been extended to support track partitioning among multiple different simulation engines. Before, it was only possible to use one chosen engine for the entire event simulation. Especially in view of the coming LHC Run 3 and 4, faster simulations are crucial to cope with the expected increase of experimental data. The enhanced VMC framework is now capable of running full-simulation together with fast-simulations. Based on specified user conditions, the simulation of single tracks can be transferred from full-simulations to be handled by fast-simulations when feasible to speed-up the detector simulation.

Zusammenfassung

Es wird die Λ_c^+ Produktion bei zentraler Rapidität $|y| < 0.5$ in Proton–Proton (pp) Kollisionen bei $\sqrt{s} = 13\,\text{TeV}$ als Funktion der Multiplizität geladener Teilchen mit dem ALICE Detektor vorgestellt. Analog werden derzeit ebenfalls die Messungen der D^0 and D_s^+ Produktionsraten durchgeführt und die Produktionsverhältnisse D_s^+/D^0 und Λ_c^+/D^0 werden diskutiert. Im Gegensatz zu dem Produktionsverhältnis D_s^+/D^0 zeigt das Verhältnis Λ_c^+/D^0 einen Anstieg mit wachsender Multiplizität. Modellvorhersagen, deren Fragmentationsprozess an Messungen in e^+e^- und e^-p Kollisionen angpasst ist, unterschätzen das Verhältnis deutlich. Vorhersagen eines Ansatzes basierend auf der kanonischen Behandlung durch ein statistisches Hadronisierungsmodell, unter Hinzunahme weiterer Baryonzustände, als auch eines Fragmentationsmodells, das Farbladungsverbindungen in höherer Ordnung implementiert, können den Anstieg und die Abhängigkeit mit der Multiplizität hingegen qualitativ beschreiben. Darüberhinaus zeigt das Produktionsverhältnis Λ_c^+/D^0 einen ähnlichen Verlauf und eine ähnliche Größenordnung im Vergleich mit dem light-flavour Λ/K_S^0 Produktionsverhältnis. Zusammen deuten diese Ergebnisse auf (i) Hadronisierungsmechanismen hin, die über die reine Fragmentation in pp Kollisionen hinausgehen und (ii) legen möglicherweise gemeinsame Mechanismen im heavy-flavour und light-flavour Sektor nahe.

-

Das Virtual Monte Carlo (VMC) Detektorsimulations-Framework, das sowohl von ALICE als auch von der FAIR Kollaboration genutzt wird, wurde erweitert, um die Partitionierung einzelner Tracks zwischen unterschiedlichen Simulationengines zu ermöglichen. Schnellere Detektorsimlationen sind vor allem im Hinblick auf LHC Run 3 und 4 ausschlaggebend, um den erwarteten signifikanten Anstieg experimenteller Daten zu bewältigen. Mit dem weiterentwickelten VMC-Framework ist es nun möglich, Full-Simulationengines und Fast-Simulationengines zu kombinieren. Die Simulation einzelner Tracks kann basierend auf individuellen Bedingungen von Full- an Fast-Simulationengine abgegeben werden um die Detektorsimulation zu beschleunigen.

Contents

1 Introduction

1.1 Λ_c^+ production in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$

The Standard Model (SM) [1–8] of particle physics is the theoretical foundation of high-energy particle and heavy-ion physics. It encodes the properties and interactions of elementary particles and to an impressive degree, the SM successfully describes a wide range of phenomena. Among others, it covers the theory of electrodynamics, builds the basis for nuclear physics, it explains how fundamental particles acquire their bare masses and it breaks the vast number of observed particles and interactions down to a few fundamental principles.

The journey of the SM began in 1897 with the first discovery of an elementary particle, namely the electron [9]. Over more than 100 years, other elementary particles were found and the theoretical description was more and more refined. In 2012, the Higgs boson as theorised by the SM [5–7] was found [10, 11] completing the theory. Several fundamental properties of this new particle are still to be explored such as the predicted triple and quartic self-coupling [5].

Until today, measuring the parameters and implications of the SM to higher and higher precision is still an ongoing effort.

On the other hand, the SM lacks the description of well-known observations. For instance, neutrinos are considered to be massless. The observations of neutrino flavour oscillations imply, however, that they are indeed massive and different mass-generation mechanisms are under consideration. The SM also does not incorporate a description of what is called dark matter or dark energy. But these observations imply that certain theoretical descriptions are missing rather than stating that the SM as an effective theory is wrong. A more severe issue arises from the fact that the mathematically solidly based and well-tested theory of General Relativity [12] can, as of today, not be unified with the SM. This indeed shows that there are inconsistencies of two fundamental theories.

But despite some shortcomings, (i) the SM impressively comprises the theoretical description of more than 100 years of research, (ii) it is one of the most precisely tested theories in physics, (iii) it is still not explored to its entirety and (iv) it drives new ideas concerning beyond-SM physics (see also [13] for a comprehensive review of the SM).

As one particular cornerstone, the theory of Quantum Chromodynamics (QCD) describes the strong interaction of colour-charged partons such as quarks and gluons.

At ordinary energy scales, partons are bound into hadrons, carrying a net-colour charge of zero, and there is no direct observation of free colour charges as of today. This property of QCD is called colour confinement.

Since QCD is a non-Abelian gauge theory [4], its coupling parameter α_S depends on the energy scale it is evaluated at. It decreases with increasing energy and therefore, despite the colour-confinement property, perturbative calculations become feasible at high energy scales [14] which is referred to as asymptotic freedom.

In heavy-ion collisions, the created energy densities can be large enough that partons overcome the force that binds them together to evolve to a thermalised system called Quark-Gluon-Plasma (QGP). Initially, it was thought to be a system of weakly interacting colour charges. Contrary to that expectation, already early measurements at the Relativistic Heavy Ion Collider (RHIC) have established more complex properties such as jet quenching [15], long-range particle correlations [16] and collective flow [17] all of which imply a strongly rather than a weakly coupled phase of the evolving medium. These observation are supported by following measurements conducted at RHIC as well as at the Large Hadron Collider (LHC).

Although these were assumed to be unique features of heavy-ion collisions, recent measurements provide evidence for collectivity even in the comparably small systems such as in proton–proton (pp) collisions (see [18, 19] for a comprehensive overview). Different from heavy-ion collisions, there is no expectation of a strongly coupled medium.

Another recent observation is a larger enhancement of baryon production compared to meson production which also increases with the event multiplicity. Measurements in the light-flavour (LF) sector conducted by ALICE have revealed this by studying a wide range of multiplicities in lead–lead (Pb–Pb) [20, 21], p–Pb [22–24] and pp [25, 26] collision systems at various centre-of-mass energies. This enhancement is hence seen even in pp collisions which is an indication for altered hadronisation mechanisms of baryons beyond pure in-vacuum fragmentation of partons. Predictions based on baryon fragmentation tuned to measurements in e^+e^- and e^-p are also seen to clearly underestimate the ratio and do not reproduce the observed shape [27, 28]. In heavy-ion collisions on the other hand, additional descriptions of hadronisation are based on the development of a thermalised system and comprise for instance partonic coalescence [29] or an effective approach based on statistical hadronisation models (SHMs) [30].

Together with the evidence of collectivity in small collision systems, these results lead to the questions of (i) whether there are additional common mechanisms in small and large collision systems, (ii) whether different mechanisms result in the same or at least similar observations, or (iii) whether there is indeed a medium formed also in small collision systems.

First goal of this thesis

The first goal of this work is the measurement presented in Ch. 5 to shed additional light on previous observations of an enhanced baryon-over-meson production. This is done in the heavy-flavour (HF) sector by measuring the prompt Λ_c^+ baryon production at midrapidity $|y| < 0.5$ in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ as a function of the average primary charged-particle density. The production ratio Λ_c^+/D^0 is discussed and also compared to the D_s^+/D^0 ratio. In addition, theoretical predictions of a canonical approach of a statistical hadronisation model with augmented baryon production as well as a fragmentation model implementing colour reconnection beyond leading colour approximation are presented and compared to the measurements. A comparison to the ratio Λ/K_S^0 as obtained in the LF sector is provided as well.

1.2 Virtual Monte Carlo developments for LHC Run 3 and 4

Detector simulation is a crucial tool for collider experiments such as ALICE and it provides, among others, the necessary basis to

- estimate event selection or particle selection efficiencies,
- conduct detector calibrations,
- estimate systematic uncertainties,
- search for new physics,
- develop new detectors by for instance estimate expected performances,

and is hence at the heart of detector developments and physics analyses.

The main task of detector simulation engines is the transport of particles through a given detector geometry consisting of various sub-detectors of different shapes and materials. These engines implement physics models in order to describe the particles' interactions with the material and their trajectories in external magnetic or electric fields. The set of interactions usually comprises electromagnetic and nuclear interactions including particle decays where secondary particles are produced.

Full event simulation is a highly resource demanding process. As the available experimental data increases, MC samples are eventually required to increase in statistics as well, in particular in order to avoid that the precision of analyses depend significantly on the otherwise comparably low MC statistics. Already during Run 2, simulation processes required more than 60 % of the ALICE computing resources. Thus, it is not possible to scale up the production by a significant factor while relying on the Run 2 hardware or software infrastructure. For ALICE it is expected to collect up to 100 times more data in Run 3 and 4 compared to Run 2 which makes hardware

and software upgrades inevitable in order to reliably provide sufficient MC samples. The simulation requirements are expected to increase by a factor 20 [31].

So far, commonly used transport codes are for instance provided by GEANT and FLUKA [32, 33]. The Fortran-based GEANT3 simulation toolkit [34] was used by ALICE during Run 2 and it was also employed at previous collider experiments at CERN. Nowadays, most experiments rely on its successor GEANT4 [35–37] which is entirely written in C++.

To be able to interface with different detector simulation engines, ALICE exploits the Virtual Monte Carlo (VMC) package [38]. Previously, however, it was only capable of interfacing to one engine at a time for the event simulation. Moreover, to reach a faster simulation on the level of particle transport, it is desirable to interface a full simulation with what is generically referred to as fast-simulation.

Second goal of this thesis

As its second goal, this thesis presents the extensions of the VMC framework in order to be capable of *partitioning* the event simulation among multiple different simulation engines. In that way, it will become possible to run a full simulation where required but to dispatch particles to fast-simulation engines where feasible.

1.3 Thesis outline

The following Ch. 2 introduces the basic theoretical framework focusing on key features of the SM in Sec. 2.1. In particular, central properties and implications of QCD are outlined in Sec. 2.2 covering the QCD phase diagram and the evolution of the QGP. Sec. 2.3 provides a selection of observables usually characterising heavy-ion collisions and they are compared to recent results in smaller collision systems such as pp and p–Pb. Ch. 3 delivers the motivation for the specific interest of conducting the analysis in the HF sector. The theoretical basis for HF quark production in pp collisions is given in Sec. 3.1 and various hadronisation models are introduced in Sec. 3.2. After that, the LHC and the ALICE detector are presented in Ch. 4, focusing on the sub-detectors and aspects in Sec. 4.2 which are crucial for the presented analysis in Ch. 5.

In Ch. 6 the Virtual Monte Carlo developments are presented. The basic building blocks of detector simulation engines and an overview of fast-simulations used at different LHC experiments are introduced in Sec. 6.2 and Sec. 6.3. Afterwards, Sec. 6.4 contains details of how detector simulation is embedded in the ALICE software stack. The VMC framework is introduced in more detail as it devises the common interface layer to transport codes. Finally, the work of how the VMC framework has been extended to provide the event-partitioning among different engines is presented and discussed in Sec. 6.5.

This thesis is concluded with final remarks and perspectives in Ch. 7.

2 Theory and physics overview

2.1 Brief introduction to the Standard Model

The Standard Model (SM) of elementary particle physics is a Yang-Mills, Lorentz-invariant and fully renormalisable Quantum Field Theory (QFT) [1–8] based on the gauge group

$$\mathrm{U}(1) \times \mathrm{SU}(2) \times \mathrm{SU}(3). \tag{2.1}$$

Symmetry transformations are performed by the generators t^a which satisfy the commutation relations

$$[t^a, t^b] = i f^{abc} t^c \tag{2.2}$$

where the fully antisymmetric f^{abc} are the structure constants of the group. A theory with non-commuting generators is called *non-Abelian* and this is the case for the SM given the group structure in Eq. (2.1). The symmetry requirement implies the existence of corresponding spin-1 *gauge* fields A^a_μ where μ denotes the Lorentzian space-time index and a refers to a certain generator.

Dirac-*spinors* Ψ represent the SM matter fields of spin-$\frac{1}{2}$, also called *fermions*. While the number and properties of the gauge fields are implied by Eq. (2.1), the matter fields are not and can only be established by experimental observations.

The SM is formulated in terms of a Lagrangian density $\mathcal{L}$ which is invariant under any of the required symmetry transformations and incorporates all elementary particles and the interactions between matter and gauge fields are encoded in the SM Lagrangian density which in its generic form reads*

$$\mathcal{L} = -\frac{1}{4} \left(F^a_{\mu\nu}\right)^2 + \bar{\Psi} \left(i \not{D} - m\right) \Psi \tag{2.3}$$

with the *field strength tensor* $F^a_{\mu\nu}$. Gauge symmetry is ensured by the *covariant derivative* given as

$$D_\mu = \partial_\mu - i g A^a_\mu t^a$$

*Note that explicit mass terms are not compatible with the SU(2) symmetry. This is resolved by the mechanism of spontaneous symmetry breaking induced by the Higgs boson. This not only generates the mass terms but also entangles the U(1) × SU(2) symmetry such that the observable gauge fields in nature are superpositions of the original fields. The terms involving the Higgs are omitted since they are not important for the further arguments made. It is referred to [1–3] detailing the calculations and implications.

and the field strength tensor reads

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c - \nu \qquad \text{where} \qquad \not{D} \qquad = \gamma_\mu D^\mu.$$

μ and ν again denote the space-time indices, a, b, c refer to the respective generator and g is a dimensionless coupling parameter. $\not{D}$ is the contraction of the covariant derivative with the corresponding Dirac matrix γ_μ. Writing out the interaction terms involving g results in

$$\begin{aligned}
\mathcal{L} =& \mathcal{L}_{\text{kin}} && (2.4)\\
&+ g A^a_\mu \bar{\Psi} \gamma^\mu t^a \Psi && (2.5)\\
&- g f^{abc} \left(\partial_\mu A^a_\nu\right) A^{b,\mu} A^{c,\nu} && (2.6)\\
&- \frac{1}{4} g^2 \left(f^{abc} A^b_\mu A^c_\nu\right) \left(f^{ade} A^{d,\mu} A^{e,\nu}\right). && (2.7)
\end{aligned}$$

This form reveals the building blocks of possible interactions in the SM as shown in the Feynman-diagram notation in Fig. 2.1 which are couplings between

- two matter and one gauge field,
- three gauge fields,
- four gauge fields.

The gauge fields couple to the charge carried by the specific fields which can be electric, weak or colour charge.

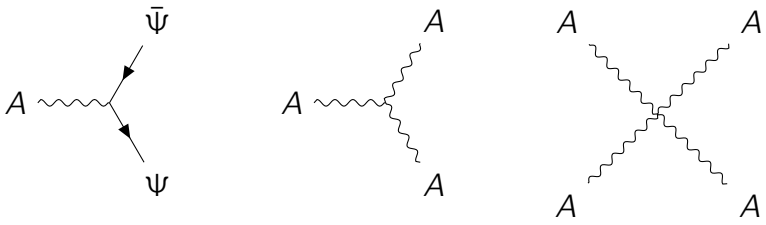


Figure 2.1: Generic Feynman diagrams sketching the possible couplings between gauge and matter fields in the SM.

The SM incorporates 12 gauge fields which are the massless photon (γ), the massive $W^\pm$ and Z bosons and eight coloured-charged massless gluons. Furthermore, there are 12 *flavours* of fermions, namely six quarks and six leptons. Fermions can be characterised by the third component of their weak isospin I_3 which corresponds to the charge connected to the SU(2) symmetry. It is $\pm\frac{1}{2}$ for left-chiral and 0 for right-chiral fermions. The interaction with $W^\pm$ bosons is fully parity-violating as they only couple to fermion dublets of left-handed chirality. Tab. 2.1 summarises the properties of gauge and matter fields. Only left-chiral fermions are included

field		mass $[\mathrm{MeV}/c^2]^*$	electric charge $[e]$	colour charge	weak isospin	gauge field coupling
gauge bosons	gluons	0	0	yes	0	gluons
	photon γ			no		$W^\pm$
	Z	91.1876				
	$W^\pm$	80.379	± 1		± 1	γ, $W^\pm$, Z
quarks	u_L	2.16	$+\frac{2}{3}$	yes	$+\frac{1}{2}$	gluons γ, $W^\pm$, Z
	d_L	4.67	$-\frac{1}{3}$		$-\frac{1}{2}$	
	c_L	1270	$+\frac{2}{3}$		$+\frac{1}{2}$	
	s_L	93	$-\frac{1}{3}$		$-\frac{1}{2}$	
	t_L	$172.76 \cdot 10^3$	$+\frac{2}{3}$		$+\frac{1}{2}$	
	b_L	4180	$-\frac{1}{3}$		$-\frac{1}{2}$	
leptons	$\nu_{e,\mathrm{L}}$	$\gtrsim 0$	0	no	$+\frac{1}{2}$	$W^\pm$, Z
	$\nu_{\mu,\mathrm{L}}$	$\gtrsim 0$	0			
	$\nu_{\tau,\mathrm{L}}$	$\gtrsim 0$	0			
	e_L	0.51	-1		$-\frac{1}{2}$	γ, $W^\pm$, Z
	μ_L	105.66	-1			
	τ_L	1768.86	-1			

Table 2.1: Summary of gauge field and left-chiral quarks properties and the coupling to gauge fields. Leptons and right-chiral quarks are omitted. Properties are extracted from [13]. Additional information is provided in the text.

(indicated by sub-script L). Note, that in the case of neutrinos ν, no right-handed component has been observed yet. In addition, each matter field has its counterpart carrying opposite charges and also those are not explicitly listed.

Feynman diagrams describing a SM process consist of multiple blocks as shown in Fig. 2.1 attached to one another. They then represent the probability a specific process to occur in nature and there is a well-defined procedure to translate those diagrams into a corresponding calculation. Every possible process has an infinite number of contributing Feynman diagrams as indicated by Fig. 2.2 showing the fusion of two gluons and the subsequent production of a quark-antiquark pair. In the top-left, the *leading order* (LO) contribution is shown. A LO contribution is in general given by the diagram with the least number of internal nodes. Each node results in an additional factor of the coupling and *next-to-leading* order (NLO) contributions are characterised in powers of g. Due to the infinite number of contributions an exact calculation is not feasible. For sufficiently low values of the coupling (well below 1), an approximate result can be obtained *perturbatively* up to a fixed order in g taking only the corresponding terms into account (see [14, 39, 40] for details). For larger coupling values, on the other hand, non-perturbative approaches are required.

Internal loops result in another subtlety, namely it needs to be accounted for any possible momentum flow within these loops, indicated as k and p in the top-right

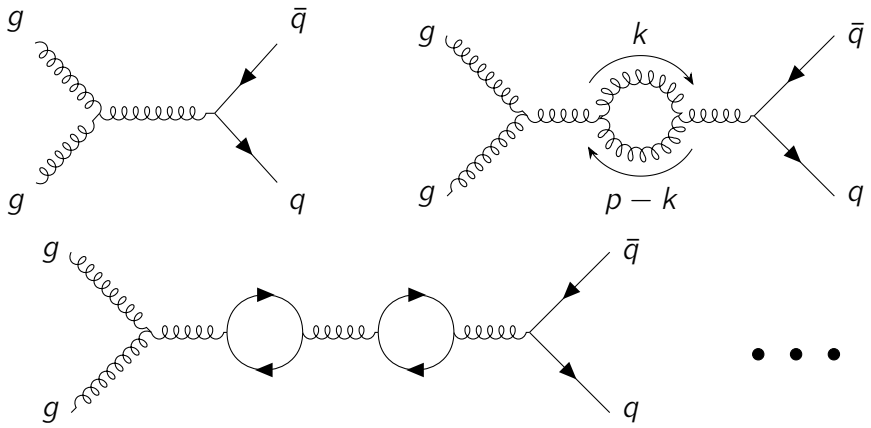


Figure 2.2: Examples of Feynman diagrams contributing to quark-antiquark pair production via gluon fusion showing the leading-order (top left) and two possible higher-order contributions. In the *top-left* diagram the internal momentum flow is indicated by p and k.

diagram in Fig. 2.2. That gives rise to divergencies which are an artefact of the fixed order calculations. It can be shown, however, that this feature can be handled by a well-defined *regularisation* and *renormalisation* procedure in which these divergencies can be absorbed and subtracted to yield finite values. The SM has been shown to be fully renormalisable [8].

Higher-order contributions of gauge fields coupling to one another cause the coupling to depend on the energy scale Q it is evaluated at. This property will be discussed in the next section in the context of QCD since only due to this property perturbative calculations become applicable in that sector.

2.2 A brief introduction to Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the sector of the SM describing the strong interaction of colour-charged quarks and gluons. Before it was established, the large set of different hadron states observed already in the early 1960s could be grouped into an $SU(3)_{\text{flavour}}$ symmetry [41–43]. Gell-Mann introduced the term "quarks" [41] for the assigned flavours of the hadrons and the quarks have not then been recognised as actual elementary particles. However, a problem for this model arose with the discovery of the Δ^{++} with the—at that time—assigned quantum numbers uuu. Without an additional quantum number, this would violate Pauli's exclusion principle based on the Fermi-Dirac statistics for spin-$\frac{1}{2}$ particles. In order to resolve this and to end up with an anti-symmetric wave function for the whole system, Gell-Mann introduced a new quantum number to be carried by either of the quarks [4]. Due

to the fact that all hadrons are singlets w.r.t. this quantum number and there are baryons with three quarks, it was called *colour*, so a blue, green and red state would result in *zero* colour. Mesons on the other hand carry a colour and the corresponding anticolour. The $\mathrm{SU}(3)_{\mathrm{colour}}$ colour symmetry at the same time implies *gluons* as the mediators which couple to colour charge and carry colour charge themselves. After *deep-inelastic scattering* (DIS) experiments in electron–proton ($e^- p$) collisions have revealed the substructure of hadrons, quarks were established as elementary particles. The representation of colour-triplets was introduced and the colour symmetry group $\mathrm{SU}(3)_{\mathrm{colour}}$ was implemented as the gauge group of QCD.

At the time it was introduced, the assignment of 3 different colours was an ad-hoc assumption which is nowadays known to be carried by partons. Although partons cannot be accessed directly, the number of colours can be inferred from the measurement of the centre-of-mass energy-dependent R-value in e^+e^- collisions which at LO of the strong coupling α_{S} is given as

$$R\left(\sqrt{\mathrm{s}}\right)_{\mathrm{LO}} = \frac{\sigma_{e^+e^-\to f\bar{f}}\left(\sqrt{\mathrm{s}}\right)_{\mathrm{LO}}}{\sigma_{e^+e^-\to\mu^+\mu^-}\left(\sqrt{\mathrm{s}}\right)_{\mathrm{LO}}} = N_{\mathrm{C}} \sum_{f\in\mathrm{flavours}} Q_f^2. \tag{2.8}$$

The numerator is the production cross section of a quark-antiquark pair of flavour f divided by the production cross section of a muon-antimuon pair. If the centre-of-mass energy is well above the production threshold for certain quark-antiquark pairs of flavour f and also if muons can be considered as massless, the LO computation yields the sum over the squared electric charges of the accounted quark flavours Q_f^2 and the factor N_{C} which denotes the number of QCD colours. Fig. 2.3 shows the ratio extracted from various measurements at different centre-of-mass energies and N_{C} is found to be 3 which is in agreement with the SU(3) prediction. The ratio increases whenever the centre-of-mass energy reaches the threshold of twice the value of the next-heavier quark mass while resonant bound states and excited states manifest themselves as excesses in the measurements.

Although partons are confined in hadronic states, their interactions can be described perturbatively at large enough energy scales due to the property of asymptotic freedom. Defining $\alpha_{\mathrm{S}} = g_{\mathrm{S}}^2/4\pi$ based on the QCD gauge coupling parameter g_{S}, the energy dependence is described by the renormalisation group equation according to

$$\mu^2\frac{\partial\alpha_{\mathrm{S}}\left(\mu^2\right)}{\partial\mu^2} = \beta\left(\alpha_{\mathrm{S}}\right), \text{ where } \beta\left(\alpha_{\mathrm{S}}\right) = -b\alpha_{\mathrm{S}}^2\left[\sum_{i=0}^{N}\alpha_{\mathrm{S}}^i b_i + \mathcal{O}\left(\alpha_{\mathrm{S}}^{N+1}\right)\right], \tag{2.9}$$

such that the value of α_{S} at scale μ can be calculated from a known value at another scale. To the first order, the energy scale dependence is

$$\alpha_{\mathrm{S}}(\mu^2) = \frac{\alpha_{\mathrm{S}}(Q^2)}{1-\alpha_{\mathrm{S}}(Q^2)\frac{11n_{\mathrm{C}}-2n_f}{12\pi}\ln\left(\frac{\mu^2}{Q^2}\right)} \tag{2.10}$$

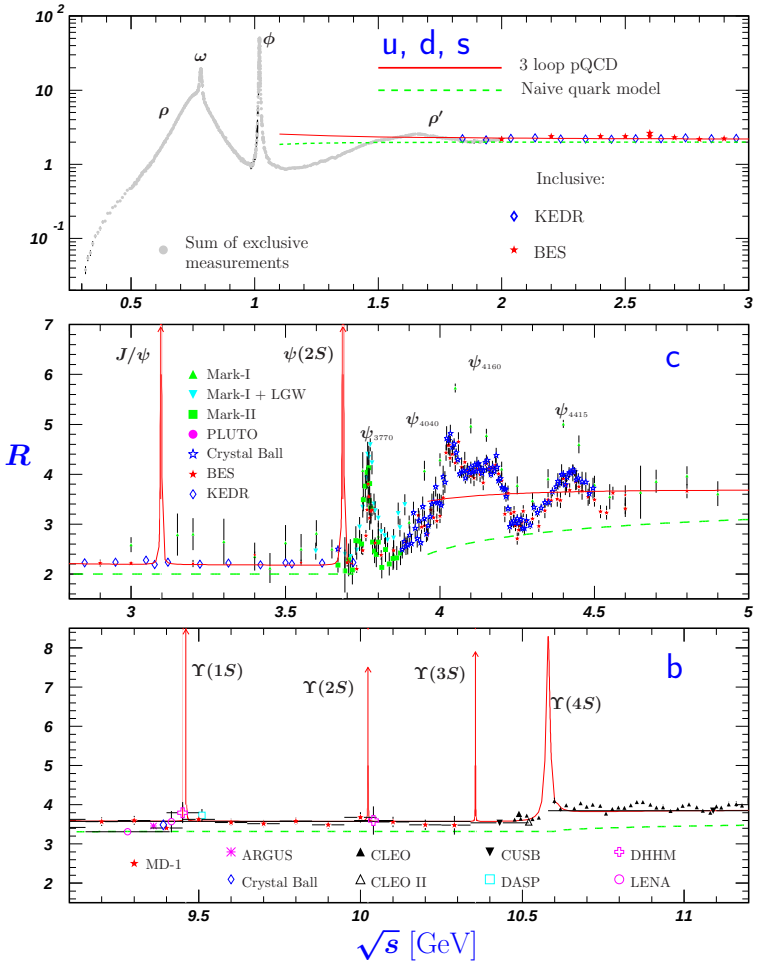


Figure 2.3: Cross section ratio of quark-antiquark pair production w.r.t. $\mu^+\mu^-$ production in e^+e^- collisions as a function of the centre-of-mass energy from various measurements. The top row contains measurements where only the lightest quark flavours u, d and s were accessible while with increasing centre-of-mass energy the c (middle row) and b quark (bottom row) are produced as well. The prediction shown in red is derived from perturbative QCD calculations containing up to three loops [13].

extrapolating to the value at scale μ from a known value at Q. n_C and n_f are the number of colours and quark flavours with $m_f^2 \ll \mu^2$. It is evident that the QCD coupling decreases with increasing energy hence allowing for perturbative calculations at large enough scales. Fig. 2.4 shows extracted values based on different experimental approaches and the perturbative order of the involved calculations is given in the brackets. The world average of the strong coupling at the Z-boson mass m_Z was computed in [13] to be $\alpha_S(m_Z) = 0.1179 \pm 0.0010$.

The theoretical prediction extrapolated from α_S at the Z mass pole is in good agreement with the data. It can be concluded that perturbative calculations start to

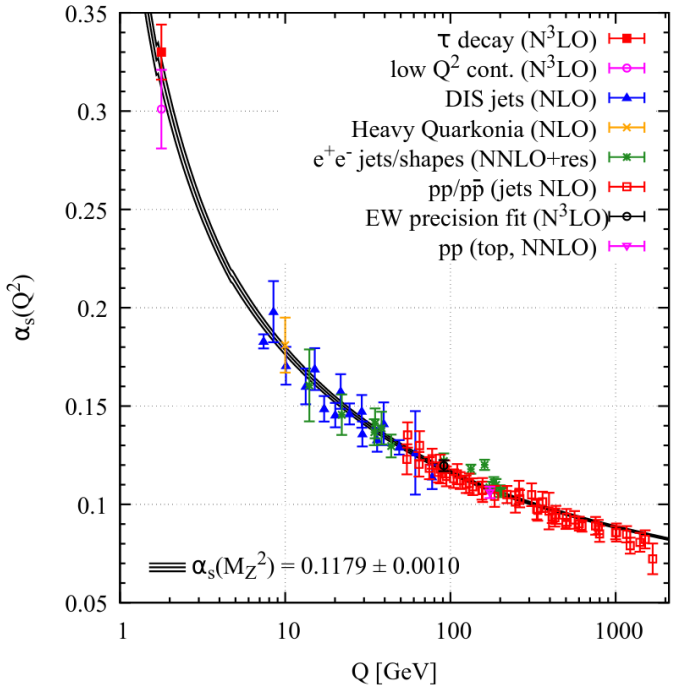


Figure 2.4: Determination of the strong coupling α_S at different energy scales from various experimental input [13]. In brackets the perturbation order in α_S is indicated up to which the calculations have been carried out to extract it.

become inapplicable at energy scales below 1 GeV but are valid for values well above which is the case in the high-relativistic particle collisions provided by the LHC.

The low energy regime of QCD is inaccessible through perturbative calculations but *lattice* QCD (lQCD) approaches have been proven feasible (see reviews in [13, 44]). In lQCD the theory is discretised onto a hypercubic lattice with spacing d. The lattice spacing induces the energy scale $\propto d^{-1}$ which thus implies the bare value of the coupling. This enables for numerical calculations starting from first principles of the theory and lQCD calculations and the parametrisation of the QCD potential is found according to

$$V_{\mathrm{QCD}}(r) \propto \sigma r - \frac{\alpha_S}{r} \tag{2.11}$$

for large distances where σ is the so-called string tension between quark-antiquark pairs [45, 46]. The linear increase with distance r can be understood as a developing colour-tube of constant energy density. Hence, the energy increases with larger distances leading to an increased attraction between quarks as they move apart.

lQCD can therefore be applied to derive the equation of state for a phase of unconfined quarks and gluons to investigate its properties and also the transition between a confined and unconfined state can be studied. Fig. 2.5 shows lattice

predictions for the normalised pressure, energy and entropy density of the system as a function of temperature including 2 light quarks and the strange quark [47]. The critical temperature of the phase transition is estimated to be $T_c = 154(9)\,\mathrm{MeV}$ as indicated in the plot.

In addition, several phenomenological models have been developed over the course of the years to describe confinement and the phase transition. Among them a very intuitive one is the *MIT bag model* [48, 49]. Quarks are thought of being placed inside a spherical bag of radius R surrounded by a QCD vacuum filled with gluons which apply a bag pressure B onto the bag working against the kinetic pressure p. The total energy in such a system is

$$E = E_{\mathrm{bag}} + E_{\mathrm{kin}} = \frac{4\pi}{3} B R^3 + \frac{2.04 n_f}{R} \tag{2.12}$$

and the equilibrium where the bag has a constant radius is found by $\frac{\partial E}{\partial R} = 0$. By approximating the bag radius by an experimentally determined size of a typical hadron, for instance the proton, the corresponding bag pressure p can be derived. At the same time, the assumption of a relativistic and massless quark gas yields the equation of state

$$p = \left[g_{\mathrm{g}} + \frac{7}{8} \left(g_q + g_{\bar{q}} \right) \right] \frac{\pi T^4}{90}, \tag{2.13}$$

where g_{g}, $g_q(g_{\bar{q}})$ are the degrees of freedom for gluons and (anti)quarks, respectively. The critical temperature at which hadrons dissolve can be estimated to be $T_{\mathrm{crit}} \approx 144\,\mathrm{MeV}$. This is indeed compatible with the value $T_c = 154(9)\,\mathrm{MeV}$ derived from lQCD as seen above.

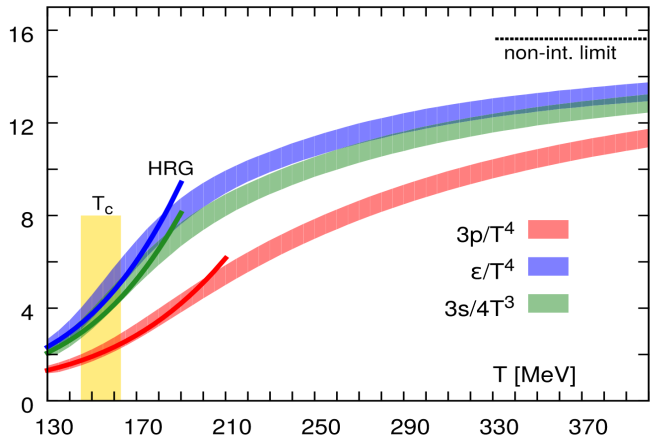


Figure 2.5: Pressure, energy and entropy density as a function of temperature derived from $(2+1)$ flavour lQCD along with predictions from the Hadron-Resonance-Gas model shown as solid lines [47].

2.2.1 The QCD phase diagram and the evolution of the Quark-Gluon-Plasma

Studying the phases and transitions of QCD matter provides further crucial understanding of the theory. Fig. 2.6 shows the QCD phase diagram with the temperature T on the vertical and the baryon chemical potential μ_B on the horizontal axis. At low temperature and baryon potential, partons are found to be confined in hadrons as it is the case at ordinary temperatures in nature. An increasing net-baryon density away from the vacuum at low temperature decreases the spatial distance between partons which eventually results in a transition to unconfined quarks and gluons strongly interacting with one another. Starting again at the vacuum at low μ_B, higher temperatures are expected to let the partons overcome the confinement as well, according to the aforementioned models, resulting in the Quark-Gluon-Plasma (QGP) phase but now with a lower interaction probability among partons. Such a state can be reproduced in heavy-ion collisions, for instance by colliding lead nuclei at the LHC which allows measurements in the region of $\mu_B \approx 0$. The fraction of highest net-baryon density resulting from such a collision, however, moves in the very forward region along the initial direction of the colliding nuclei and cannot yet be studied in current midrapidity experiments such as those placed at the LHC. This makes systems with a large μ_B inaccessible at the LHC. On the other hand, a decrease in energy leads to a larger contribution of the baryon number of the incident particles in the medium which effectively increases μ_B. Hence, systems with a larger baryon chemical potential are produced for instance at the NA61/SHINE experiment [50] at CERN. In particular, the Compressed Baryonic Matter (CBM) collaboration [51] at the FAIR facility in Darmstadt, Germany, explicitly declares the exploration of the QCD phase diagram at high μ_B as one of their main goals. It is planned to explore nucleus–nucleus collisions with a per-nucleon energy range of 2 GeV to 45 GeV. In particular and as indicated in Fig. 2.6, the phase transition to the QGP is foreseen to happen on the other side of a potential critical point, if it exists. By doing so, this might help to address the question of whether the phase transition is of first or second order, depending on whether a critical point can be determined

In the initial state of a heavy-ion collision, the nuclei are highly boosted and can be described as disc-like as a consequence of the Lorentz contraction which overlap during the collision as sketched in the left panel of Fig. 2.7. Since they are moving with almost the speed of light, the incoming beams are on the light-like diagonals in the space-time diagram in the right panel. When the two nuclei collide, the interacting partons of the participating nucleons constitute a system which expands rapidly with almost the speed of light filling the entire light-cone. The system reaches an equilibrium stage after less than 1 fm/c, forming the QGP which can be treated as a relativistic hydrodynamic state. As the systems expands and cools down, hadrons are formed and during further cooling, hadrons still interact elastically as well as inelastically until reaching the chemical freeze-out temperature

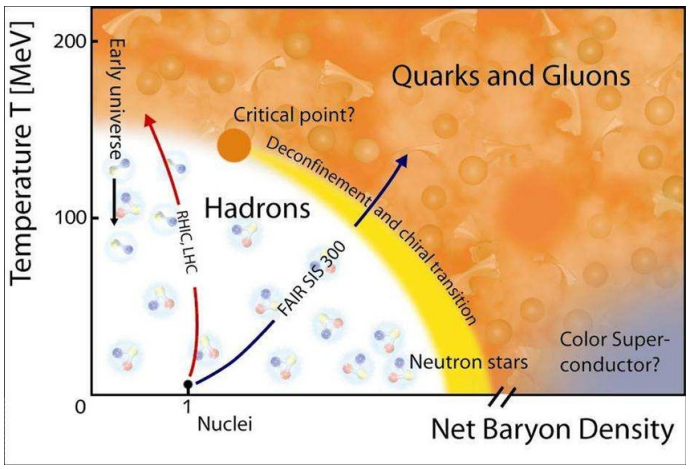


Figure 2.6: The phase diagram of QCD. Phases are characterised by given temperature and baryon chemical potential. Changing one of them eventually results in a phase transition.

at $T_{\text{fo}} = 156.5(15)\,\text{MeV}$ [52]. At this point, hadron abundances are fixed. During the further expansion, hadrons still interact elastically until the kinetic freeze-out is reached. Of course, some particles might then decay to lighter ones. Created particles can then be measured in detector experiments and although this entire evolution cannot be observed directly, it leaves its fingerprint in the kinematics and abundances of detectable particles.

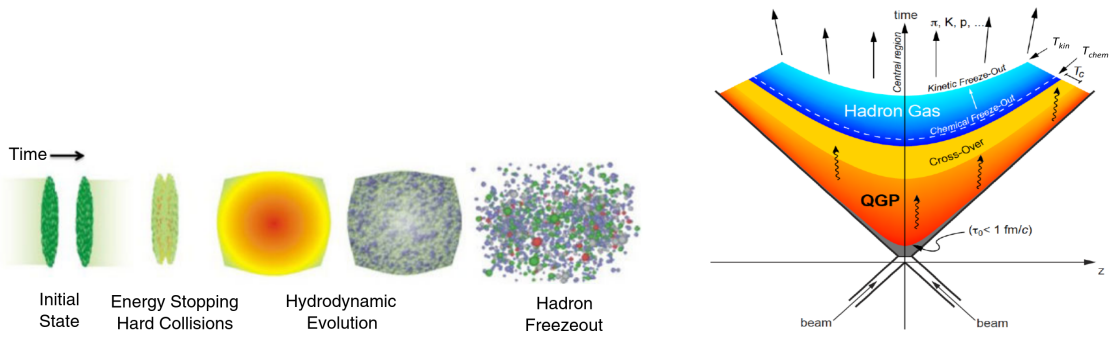


Figure 2.7: Evolution of a heavy-ion collision. *Left:* Depiction of the collision where highly Lorentz-contracted nuclei collide and initiate the creation and evolution of a QGP [53]. *Right:* Corresponding space-time diagram indicating of a heavy-ion collision with incoming beams indicated in the light-like diagonals and showing the system evolution in the beam direction [54].

2.3 Selected observables in large and small collision systems

A selection of observations in heavy-ion collisions is presented in the following to highlight the implications of a QCD medium. In particular, also results obtained in pp collisions are shown to emphasize that even those small systems show characteristics which were previously only expected to be present in heavy-ion collisions where a QCD medium evolves. A comprehensive discussion can be found in [18, 19].

2.3.1 Flow

In heavy-ion collisions the nuclei collide with a different overlap which varies on an event-by-event basis. A collision is called *peripheral* if the nuclei only overlap at their boundaries and a depiction of such a collision is given in Fig. 2.8 showing the almond-shaped overlap region of the nucleons. As the overlap region grows, it is called *semi-peripheral* and eventually *central*. The overlap is given in terms of *centrality* and ranges from 0 % (full overlap) to 100 % (no overlap).

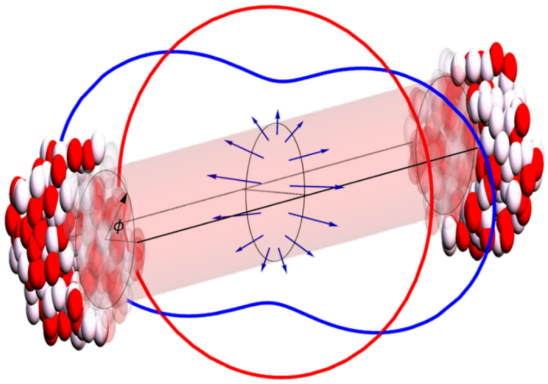


Figure 2.8: Sketch of a peripheral heavy-ion collision where the incoming nuclei have an almost elliptical overlap region as indicated. Taken from [55].

The initial anisotropic distribution of nucleons in the overlap region and the pressure gradients caused by the interacting particles right after the collision lead to an anisotropic momentum distribution of final state particles. This is accessible and characterised by the so-called *flow coefficients* v_n which are connected to the particle multiplicity's azimuthal distribution $\frac{\mathrm{d}N}{\mathrm{d}\varphi}$ by

$$\frac{\mathrm{d}N}{\mathrm{d}\varphi} \propto 1 + 2\sum_{n=1}^{\infty} v_n \cos\left(n(\varphi - \Psi_n)\right). \tag{2.14}$$

φ is the azimuthal angle in the laboratory frame while Ψ_n denotes the common reaction plane of the n'th component [56, 57]. In particular, v_2, v_3 and v_4 are called

elliptic, triangular and quadrangular flow. v_2 is the dominant coefficient in semi-central collisions reflecting the almond-shaped like overlapping region of colliding nuclei. Finite higher order coefficients reflect event-by-event fluctuations of the spatial anisotropy.

Non-vanishing flow coefficients are an evidence for *collective* effects in the sense that the partons produced in the collision are significantly interacting along the system development. The evolution of that medium as a collectively developing system is then imprinted on the final state configuration in which case a heavy-ion collision cannot be described as the independent development of binary nucleon-nucleon interactions.

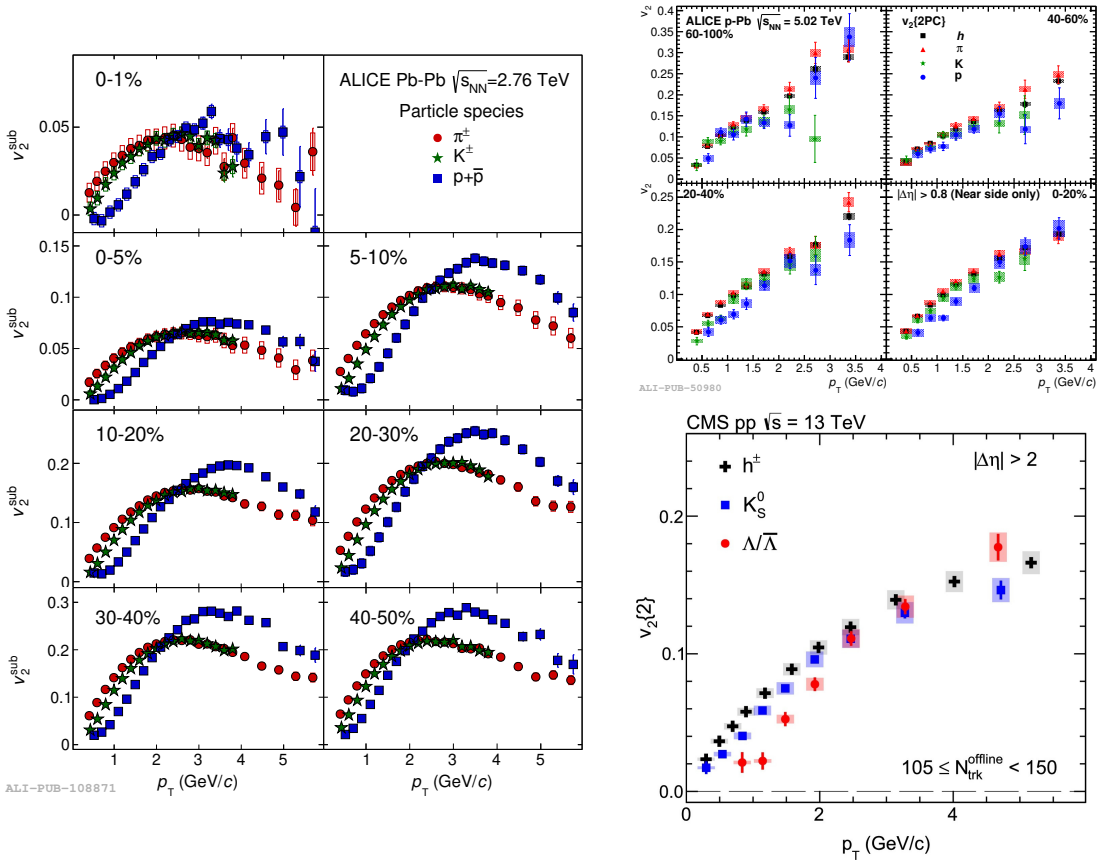


Figure 2.9: Elliptic flow as a function of particle p_T in different collision systems, measured in Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76\,\text{TeV}$ [58] *(left)* as well as in p–Pb collisions at $\sqrt{s_{NN}} = 5.02\,\text{TeV}$ [59] *(top right)* and pp collisions at $\sqrt{s} = 13\,\text{TeV}$ [60] *(bottom right)*. Finite elliptical flow and a shift to higher p_T for heavier particles is observed in all systems.

Finite elliptic flow of charged particles has been measured already in the year 2001 in gold–gold (Au–Au) collisions at $\sqrt{s_{NN}} = 130\,\text{GeV}$ at RHIC [17]. The first measurement of v_3, v_4 and v_5 of charged particles in Pb–Pb collisions at $\sqrt{s_{NN}} =$

2.76 TeV was presented by ALICE in 2011 [61] and also v_6 has been measured to be finite [62].

Results from measurements of elliptic flow at $\sqrt{s_{NN}} = 2.76\,\text{TeV}$ conducted by ALICE [58] in central and semi-central collisions are shown in the left panel of Fig. 2.9. The overall decrease of the elliptic flow with increasing centrality reflects the fact of the overlap region becoming more and more symmetric approaching a circle. The distribution of protons (blue) appears shifted ($\lesssim 3\,\text{GeV}$) w.r.t. the lighter pions and kaons shown in red and green, respectively. This is expected since higher masses imply higher momenta given a common velocity field in a thermalised system. Finite elliptic flow as well as the shift of more massive particles is also observed in proton–lead (p–Pb) collisions at $\sqrt{s_{NN}} = 5.02\,\text{TeV}$ [59], shown in the top-right panel. Although initially only expected in heavy-ion collisions, there is evidence of finite elliptic flow in pp collisions as measured by CMS at $\sqrt{s} = 13\,\text{TeV}$ [60] in events with a large final state particle multiplicity as shown in the bottom-right plot of Fig. 2.9. As one can see, there is the indication that also the mass ordering is preserved as the red points of Λ appear shifted to higher momenta compared to the inclusive hadron (black) and K^0_S (blue) data points. This might indicate an early anisotropy and a collective evolution of even small collision systems. However, despite the evidence of collectivity, there is no direct evidence yet that a medium emerges in pp collisions or and it is not clear whether the measured flow could be caused by another mechanism.

2.3.2 Jet quenching

A high energetic parton produced in a high-energetic collision with large momentum transfer between the colliding particles radiates further partons collinearly initiating a so-called parton shower (see also Sec. 3.3 for a parton shower description). In heavy-ion collisions forming a QGP, these partons traverse the medium. By doing so, they lose energy by elastic and inelastic interactions with the thermalised partons in the medium. At the same time, the transverse momentum of interacting partons would be expected to be modified by the interactions which would lead to a broadening of observed jets. These mechanisms are summarised under the term of *jet quenching* (see [63] for a review). The radial size of jets is given by

$$R = \sqrt{(\Delta\varphi)^2 + (\Delta\eta)^2}, \tag{2.15}$$

where $\Delta\varphi$ and $\Delta\eta$ are azimuthal and pseudorapidity distances. A jet is then defined as the collection of hadrons inside a chosen R using appropriate jet-finding algorithms [64].

To investigate potential medium effects, jet production is studied Pb–Pb collisions in in pp systems as a reference and the nuclear modification factor is defined as

$$R_{AA} = \frac{d^2 N^{AA}_{jets}/dp_T d\eta}{\langle T_{AA}\rangle d^2\sigma^{pp}_{jets}/dp_T d\eta}, \tag{2.16}$$

where $\mathrm{d}N^{\mathrm{AA}}_{\mathrm{jets}}/\mathrm{d}p_{\mathrm{T}}\mathrm{d}\eta$ is the measured jet spectrum in Pb–Pb collisions and $\mathrm{d}^2\sigma^{\mathrm{pp}}_{\mathrm{jets}}/\mathrm{d}p_{\mathrm{T}}\mathrm{d}\eta$ is the jet cross section in the reference pp system. $\langle T_{\mathrm{AA}}\rangle = \langle N_{\mathrm{coll}}\rangle/\sigma^{\mathrm{NN}}$ scales the pp cross section by the expected average number of binary nucleon-nucleon interactions $\langle N_{\mathrm{coll}}\rangle$ which is calculated by a Glauber model [65] estimating the probability of the spatial configuration of participating nucleons within the nuclei A.

Fig. 2.10 contains the jet production $\mathrm{R_{AA}}$ for six intervals of increasing centrality from top-left to bottom-right as measured by CMS in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76\,\mathrm{TeV}$ for three different jet radii [66]. As one can see, the suppression of jet production in Pb–Pb increases with centrality. This is hence a clear indication for a strongly interacting medium in which the traversing partons lose their energy and thus, jet production is suppressed. On the other hand, the measurements are compatible for all jet radii within each centrality interval and it cannot be concluded on a potential p_{T}-broadening given the current uncertainties.

Different from observations in flow analyses, no indication for jet quenching in pp collisions has been observed yet.

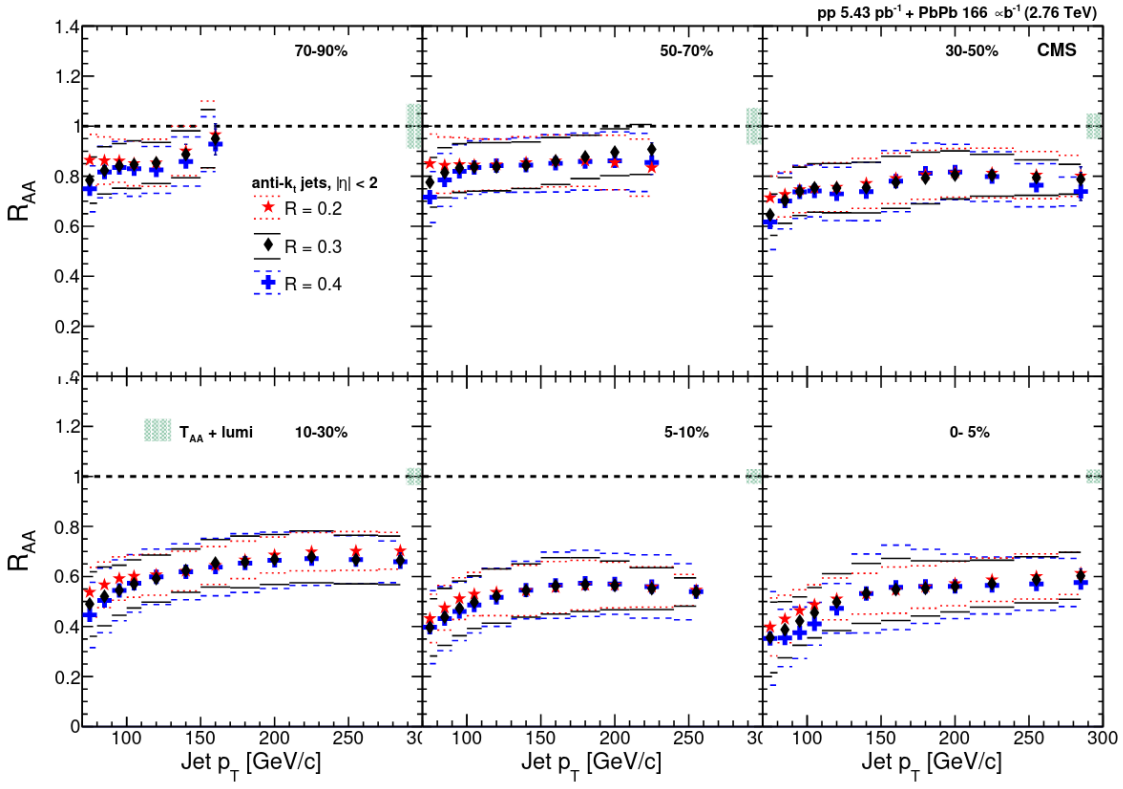


Figure 2.10: Jet production $\mathrm{R_{AA}}$ measured in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76\,\mathrm{TeV}$ [66] in intervals of increasing centrality from *top-left* to *bottom right*. The measurement was conducted for three different jet radii (red, black, blue).

3 Heavy-flavour physics in proton–proton collisions

Heavy-flavour (HF) physics covers the study of properties of hadrons containing charm or beauty quarks* as well as properties of collision systems and their development using those hadrons as probes. In heavy-ion collisions at the LHC, the initial temperatures reach values of several $\sim 100\,\mathrm{MeV}$ which is still significantly below the charm- and beautyquark bare masses of $m_\mathrm{c} = 1.3\,\mathrm{GeV}$ and $m_\mathrm{b} = 4.2\,\mathrm{GeV}$ [13], respectively. In pp or heavy-ion collisions, HF quarks are hence predominantly produced in the hard scattering within about $0.02\,\mathrm{fm}/c$ and $0.1\,\mathrm{fm}/c$ [67], respectively, when large momenta are exchanged between the colliding protons. This happens well before the formation of a medium (within $0.3\,\mathrm{fm}/c$ to $1.5\,\mathrm{fm}/c$ [68]) in heavy-ion collisions and their thermal production can be neglected [69]. In addition, a flavour-changing decay can only occur through a weak decay radiating off a $\mathrm{W}^\pm$ on a time scale which is 15 orders of magnitude above the one of the strong interaction. For these reasons, the abundance of HF quarks can be assumed to be constant in time starting from the collision up to their hadronisation. On the other hand, this is not the case for LF quarks as the above argument is particularly based on the quark bare masses. Consequently, HF hadrons can be used as a proxy to measure these abundances. HF hadrons therefore devise unique probes, which in particular can help to test hadronisation processes and perturbative QCD calculations.

According to the factorisation theorem [70], the probability of the inclusive production of a certain hadron h in the collisions of two protons, $p + p \rightarrow h + Y$ with Y denoting the additional production of hadrons and leptons, is given by the corresponding *cross section* which can be written as

$$\sigma_{p+p\rightarrow h+Y}(\sqrt{\mathrm{s}}) = \sum_{a,b} \int_0^1 \mathrm{d}x_1 f_a(x_1, \mu_\mathrm{f}) \int_0^1 \mathrm{d}x_2 f_b(x_2, \mu_\mathrm{f}) \\ \hat{\sigma}_{a+b\rightarrow c+\hat{Y}}(x_1, x_2, \sqrt{\hat{\mathrm{s}}}, \mu_\mathrm{f}, \mu_\mathrm{r}) D_{c\rightarrow h}(x_c, \mu_\mathrm{f}). \tag{3.1}$$

a and b indicate the partons originating from either of the protons with energy fractions x_1 and x_2 w.r.t. the entire energy carried by the corresponding proton. The partonic cross section $\hat{\sigma}$ is assumed to be calculable perturbatively at the renormalisation scale μ_r and devises the probability of the production of parton c accompanied by the potential additional production of other partons and leptons denoted by $\hat{Y}$.

*Even though the top quark is even more massive, it does not fall into this category since it decays before it forms hadrons [13].

The long distance/low energy features are *factorised* by the factorisation scale μ_f*. At long distances, the *parton distribution functions* (PDFs) encapsulate the probability of finding partons a and b within either of the protons carrying the energy fractions x_1 and x_2. PDFs are determined experimentally at specific energy scales because they cannot be derived from first principles due to the fact that perturbative calculations are not applicable to describe partons in the confined state of the proton. To do so, results from *deep inelastic scattering* (DIS) experiments can be used where the sub-structure of protons can be probed by the interaction with high energetic electrons [71]. Fig. 3.1 shows an example of PDFs for a proton probed at a scale of $Q = 10\,\text{GeV}$. Once determined, a PDF given at one scale can be developed to another scale by means of the DGLAP equations [72–74].

The final ingredient to be accounted for on the long distance scale is the fragmentation function (FF) $D_{c\to h}(x_c, \mu_f)$ which yields the probability that parton c fragments into hadron h of interest carrying the energy fraction z of the initially produced parton. In the following, HF quark production and different hadronisation mechanisms are introduced, the latter also covering FFs.

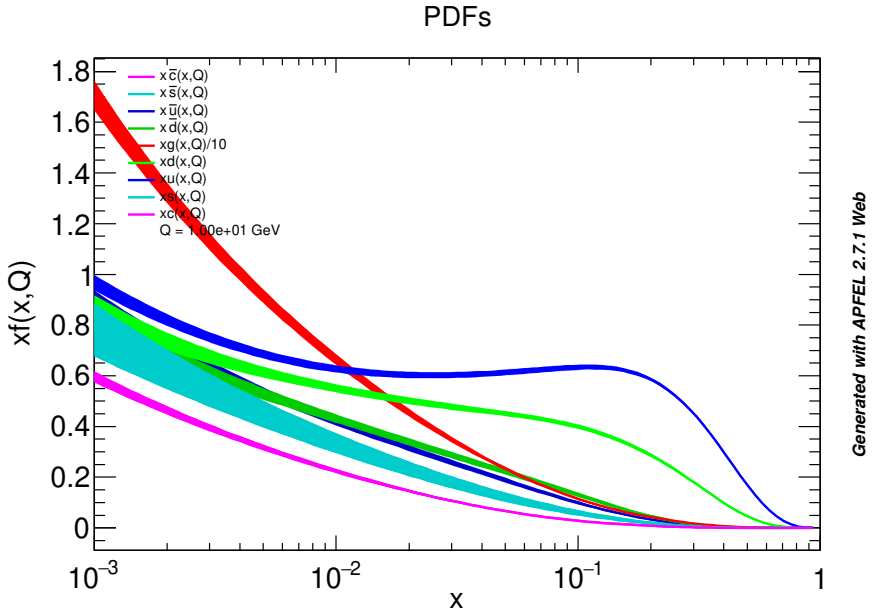


Figure 3.1: Different parton distributions within a proton probed at the energy scale of 10 GeV compiled using APPLE [75, 76].

*The hadronic cross section $\sigma_{p+p\to h+Y}(\sqrt{s})$ cannot depend on any of the introduced scales and they are, in principle, arbitrary as they are only introduced for computational reasons in order to obtain a finite result. They are commonly chosen to be of the order of the hard process $\hat{\sigma}$ under study.

3.1 Heavy-flavour quark production

The following overview is based on the discussion in [27] and it is referred to it (and references therein) for a more in-depth discussion.

Fig. 3.2 sketches examples of Feynman diagrams for LO and a possible NLO contribution of a HF quark pair ($Q\bar{Q}$) in pp collisions. The NLO contribution contains an additional gluon radiated off either in the initial or final state of the process. The perturbative calculations diverge for branching angles $\to 0$ between the quark and the radiated gluon. Therefore, theoretical calculation have to account for these *collinear singularities* which correspond to additional real emissions of further gluons and quarks, in case radiated gluons split into quark-antiquark pairs. If all parton masses involved can be neglected compared to the energy scale of the process, these singularities can be fully factorised into the hadron PDFs and the FFs and do not enter the perturbative calculations. This approach is commonly called *zero-mass variable-flavour-number scheme* (ZM-VFNS). As the factorisation scale μ_f (see Eq. (3.1)) crosses a certain value such that the next-heavier quark flavour can be assumed massless, also that is treated in the perturbative calculations, hence motivating the term *variable-flavour*. This is also the case when calculating quark production at high momenta.

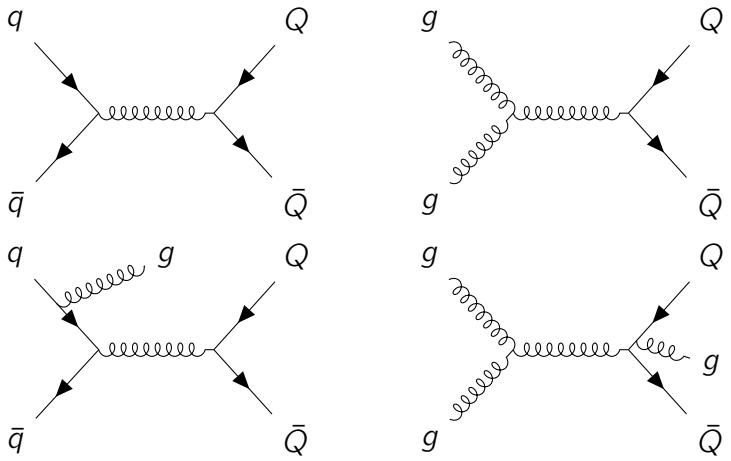


Figure 3.2: Examples of Feynman diagrams contributing to HF quark pair production in pp collisions. The top row shows LO contributions, namely quark annihilation and gluon fusion, whereas the bottom row shows possible NLO contributions with additional gluon radiation in the initial and final state, respectively.

On the other hand, in the *fixed-flavour-number scheme* (FFNS) the HF quark mass m_HF introduces a finite large scale throughout the calculation and acts as a cut-off shielding the collinear divergence.

Both approaches complement each other and can be used to calculate the cross

sections for inclusive HF production: the ZM-VFNS covering the high-p_{T} region while FFNS-based calculations are applied at low p_{T}. Both approaches can be combined in the *general-mass variable-flavour-number scheme* (GM-VFNS). In that case, the FFNS calculations are conducted in the limit $m_{\mathrm{HF}} \to 0$ to derive appropriate subtraction terms by comparing it to the ZM-VFNS calculations. By doing so, it can be fully accounted for the masses and the subtraction terms ensure the correct behaviour when the mass becomes negligible.

Fixed-Order plus Next-to-Leading-Logarithms (FONLL) [77, 78]* calculations take care of this by an exact calculation at fixed-order (FO) NLO, fully incorporating finite mass effects of the heavy quark, as well as including the resummed (RS) next-to-leading logarithm (NLL) contribution to all orders in which the quark is treated as a light degree-of-freedom. Since the terms at a given FO are calculated *exactly*, the corresponding zero-mass limit FOM0 has to be subtracted from the RS terms in order to avoid adding this contribution twice. In a simplified notation adapted from [77] the calculation of the partonic cross section can be expressed as

$$\hat{\sigma}_{\mathrm{FONLL}} = \hat{\sigma}_{\mathrm{FO}} + (\hat{\sigma}_{\mathrm{RS}} - \hat{\sigma}_{\mathrm{FOM0}}) \cdot G(m, p_{\mathrm{T}}), \tag{3.2}$$

where the function $G(m, p_{\mathrm{T}})$ is in principle arbitrary but constrained to $\lim_{\frac{m}{p_{\mathrm{T}}} \to 0} = 1$.

3.2 Hadronisation models

Hadronisation is the process in which partons that carry free colour charge are bound into colour-neutral hadrons. When considering each parton to hadronise independently, this can be factorised from the perturbative QCD calculations as described at the beginning of this chapter. In the following paragraph, different theoretical approaches currently used to describe the mechanisms of hadronisation in vacuum and those applied in the presence of a QCD medium are presented. For the latter, a microscopic approach providing a description on partonic level, namely the *partonic coalescence model*, and the *Statistical Hadronisation Model* effectively describing hadronisation by a macroscopic approach, will be outlined.

3.2.1 Fragmentation functions to describe in-vacuum fragmentation

Fragmentation functions (FFs) are typically indicated as $D_{c \to h}(x_c, \mu_{\mathrm{f}})$ (see also Eq. (3.1)) to describe the probability that a certain parton c fragments into a hadron h that carries a fraction x_c of the initial parton momentum. There are various proposed parametrisations describing their functional form such as the Peterson [82] or Bowler [83] parametrisations.

*See [79–81] for FONLL calculations being compared to different experimental measurements.

The approach of FFs in general does not take any interaction of partons into account but each one of those is treated to fragment independently. FFs are derived from measurements in small collision systems such as e^+e^- or pp and it is generally assumed that they are independent of the collision system. A summarising measurement of D-meson and Λ_c^+ FFs is presented in [84] where they are calculated from a combined analysis of different collision systems, namely e^+e^-, e^-p and pp, as the ratio

$$FF(c \to h) = \frac{\sigma(h)}{\sigma(c)}. \tag{3.3}$$

Here, $\sigma(h)$ is the production cross section of hadron h and $\sigma(c)$ the inclusive charm production cross section. The central result is shown in Fig. 3.3 where, for instance, $D_{c\to\Lambda_c^+} \lesssim 0.1$.

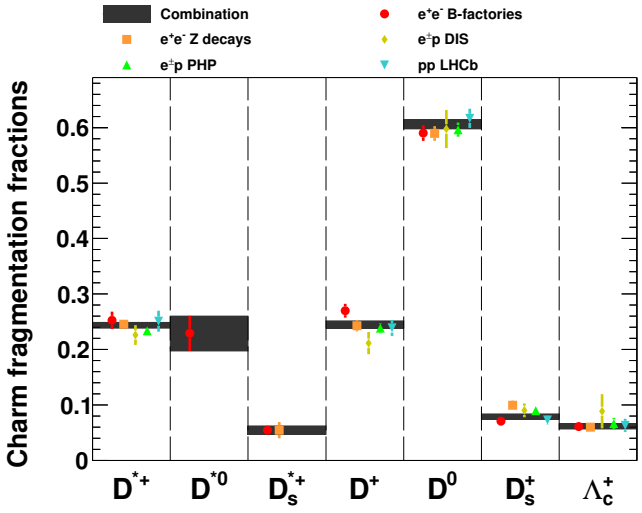


Figure 3.3: Inclusive charm fragmentation functions for D-mesons and the Λ_c^+ baryon obtained from a combined analysis of various experimental results [84].

3.2.2 Quark coalescence

Another model describing hadronisation on parton level is given by partonic coalescence models and different from the aforementioned in-vacuum fragmentation, the presence of a medium is assumed.

Originally, the coalescence approach was put forward to describe the formation of light nuclei from nucleons which are close in phase space during the kinetic freeze out [85]. Later, however, it was also started to be used to describe hadron formation in heavy-ion collisions via coalescence in a thermalised partonic medium (see for instance [86]). Following the discussion in [29], the number of light-flavour mesons

hadronising via coalescence can be written as

$$N_{\mathrm{M}} = g_{\mathrm{M}} \int \prod_{i=1}^{2} \cdot \frac{p_i \cdot \mathrm{d}\sigma_i \mathrm{d}^3\boldsymbol{p}_i}{(2\pi)^3 E_i} f_q(x_1, p_1) f_{\bar{q}}(x_2, p_2) f_{\mathrm{M}}(x_1, p_1; x_2, p_2). \tag{3.4}$$

$f_{q,\bar{q}}(x, p)$ are the invariant phase space distribution functions of quarks/antiquarks and f_{M} denotes the probability of forming meson M given their respective positions $x_{1/2}$ and momenta $p_{1/2}$ in phase space. g_{M} is the corresponding probabilistic factor accounting for the spin and colour quantum numbers of the quark-antiquark pair forming the colourless meson M which for the π-meson is $g_\pi = 1/36$ [29]. $\mathrm{d}\sigma_{1/2}$ is an infinitesimal space-like hypersurface element. This approach can be generalised to treat the production of baryons where three quarks instead of a quark-antiquark pair are involved as described in the previously given reference.

Evidently and analogous to the original coalescence of light nuclei, this model follows an approach where the probability of a quark-antiquark pair forming a meson is proportional to the overlap $f_q(x_1, p_1) f_{\bar{q}}(x_2, p_2) f_{\mathrm{M}}(x_1, p_1; x_2, p_2)$ in phase space.

3.2.3 Statistical hadronsisation model

In contrast to the aforementioned macroscopic models acting on parton level, a fully thermalised system can be attempted to be described statistically on the basis of its macroscopic properties. This approach is chosen by so-called statistical hadronisation models (SHMs). Hadron abundances are predicted from the volume V, the temperature T and the chemical potential μ of the thermalised system at the chemical freeze-out of the system development in heavy-ion collisions.

Hence, assuming an equilibrated system, its macroscopic properties can be derived from the grand canonical ensemble with the partition function

$$Z(T, V, \mu) = \mathrm{Tr}\left[\exp\left(-\beta\left(H - \sum_i \mu_{Q_i} Q_i\right)\right)\right]. \tag{3.5}$$

Here, $\beta = 1/T$, μ_{Q_i} are the chemical potentials to conserve the charge Q_i on average in the system and H is the Hamiltonian to describe the hadron resonance gas (see for instance [30] for further details). Conserved quantities in such a system are the electrical charge, strangeness and the baryon number with their chemical potentials denoted as μ_{Q}, μ_{S} and μ_{B}, respectively. As argued in [87] (see also references therein), the formation of the single hadrons is independent and the overall partition function can be approximated as the product of the partition functions Z_h of each hadron species h according to the product

$$Z \approx \prod_h Z_h. \tag{3.6}$$

The fraction of each species per volume V can then be written as

$$n_h = \frac{N_h}{V} = \frac{T}{V}\left(\frac{\partial \ln Z_h}{\partial \mu_h}\right)_{V,T}, \tag{3.7}$$

where μ_h is the sum over the chemical potentials applicable to h. Strangeness is conserved to be 0 in the overall system and the electrical charge is conserved and given from the studied collision system in a specific rapidity region [88]. In a modelling approach, T, μ and V are free parameters which can be obtained from a fit to measured data. Once determined, SHMs can be used to predict the particle abundances at a given chemical freeze-out temperature T_{cfo}. Note that, in order to obtain the correct production ratio of each hadron species h, contributions from higher mass resonant states decaying to h have to be taken into account as well.

The red data points in Fig. 3.4 indicate hadron abundances obtained from measurements in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76\,\mathrm{TeV}$ [89–92]. A fit yields $T_{\mathrm{cfo}} = 156.5(15)\,\mathrm{MeV}$, $\mu_{\mathrm{B,cfo}} = 0.7(38)$ (compatible with 0 as expected for a system evolving at midrapidity) and $V = 5280(410)\,\mathrm{fm}^3$ and it shows a good agreement compared to the measurements and covers the hadrons abundances across nine orders of magnitude and within a range of 11 different particles and antiparticles such as light and strange hadrons as well as light nuclei [87].

Being hence well applicable to heavy-ion collision systems, SHM models were also found to be capable of describing particle abundances in small collision systems such as e^+e^- and pp as discussed in [93] (and references therein). However, as of now there is no common ground reached yet for how to interpret these observations in terms of a thermalised system in such cases.

3.3 Cross section calculations with Monte Carlo event generators

Calculations of cross sections can be computed by means of Monte Carlo event generators which are programs carrying out the calculations necessary to cover the entire computation given in Eq. (3.1) numerically by means of the Monte Carlo approach.

Different solutions are available, either designed as *multi-purpose* event generators or as those made for more specific use cases. The first group comprises among others PYTHIA 8 [94], Herwig7 [95, 96] or Sherpa [97, 98]. On the other hand, for instance the event generator JEWEL [99] is specialised to treat the jet evolution in a QCD medium.

The following discussion in particular applies to PYTHIA 8 as its predictions are also used as comparisons in Ch. 5.

PYTHIA 8 usually calculates cross sections at LO in α_{S}. Beyond that, it can be interfaced with *Matrix Element* (ME) calculation tools such as MadGraph [100] or

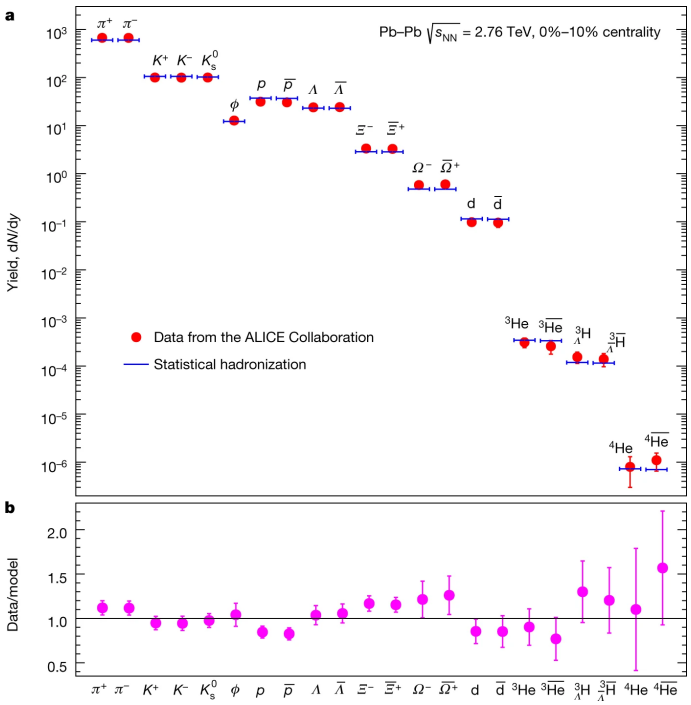


Figure 3.4: Particle abundances measured in Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76\,\text{TeV}$ compared to predictions from a statistical hadronisation model with the parameters chemical freeze-out temperature T_{cfo}, volume V_{cfo} and baryonic chemical potential $\mu_{B,cfo}$ obtained from a fit to data [87].

POWHEG BOX [101] to compute cross sections at NLO in QCD and some even to NLO in QED.

After the particles have been produced from the ME calculation, partons are developed through parton shower (PS) algorithms [102] devising the resummation of higher-order contributions to the hard scattering process. This results in multiple branchings in the collinear limit of the outgoing partons which reflect the additional higher-order real emissions of the underlying process. The calculations of the single branchings in the collinear regime are done using the splitting kernels according to the DGLAP equations [72–74]. Since PSs are based on pQCD calculations, a cut-off value is chosen where the development stops.

At this point and due to the non-perturbative nature at these energy scales, phenomenological models are employed to described the hadronisation process. By default, PYTHIA 8 uses the Lund string fragmentation model, which is introduced in the following section, to compute the hadronic final state of the collision.

3.3.1 The Lund string model

The Lund string fragmentation model [103, 104] acts on the microscopic parton level and is invoked after the final state PS. By default, the model is tuned to FFs (see

also Sec. 3.2.1).

The approach is based on the development of a colour field of constant energy density emerging between quark-antiquark pairs as discussed in Sec. 2.2 which are called *colour strings*. While moving apart, eventually another quark-antiquark pair is created from the energy stored in the string when its energetically favoured. Between each parton of the newly created pair and the initial quarks new strings are created and the initial string is broken up into two new ones. Gluons carry a colour and an anticolour and are connected to a quark-antiquark pair. A simplified picture of colour strings between quark-antiquark pairs is shown in the left panel of Fig. 3.5* with the time dimension on the vertical and the spatial dimension on the horizontal axis, respectively. The oscillating structure arises since the quarks are attracted, always being pulled closer to each other until they are eventually bound together.

Also a *junction* can be formed with three colour strings each with one quark on one end and the other ends are attached to each other. This can also be interpreted as two colours being connected to a third. Between each quark and the junction, new $q\bar{q}$ pairs can be created. The three quarks left closest to the junction will finally be combined into a baryon. In the same way in which two colours are connected to the third in a junction, a di-partonic system with two different colours can lead to a colour string of the third colour as sketched in the right panel of Fig. 3.5 which eventually also allows for baryon formation.

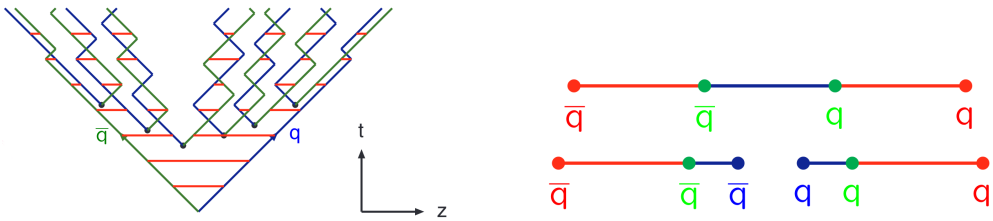


Figure 3.5: *Left:* Development of colour strings in space (horizontal axis) and time (vertical axis) showing multiple break-ups of colour strings. Eventually, the partons are attracted to each other, causing an oscillation-pattern. *Right:* Colour string formation between a di-partonic system. Both sketches taken from [105].

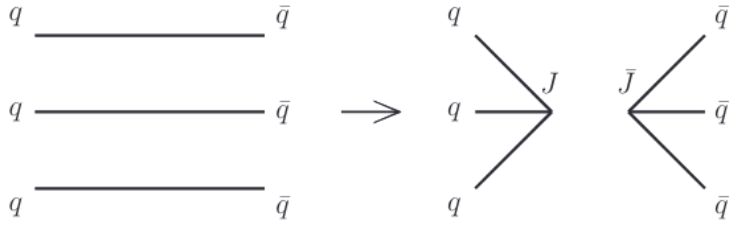
This baseline approach assumes that colour strings develop independently of each other and there are no overlaps at similar space-time points. This is called *leading-colour* (LC) approximation. In high-energetic pp interactions, one might expect an overlap of strings. For instance, not only one but multiple partons of the colliding protons can interact leading to what is called *multi-parton interaction* (MPI). All of the developing systems thereafter are close in space-time and colour strings could be considered between partons originating from different primary interactions. Such an

*The colours are introduced to distinguish the quark from the antiquark as they move along. They should not be taken as the colour of the quarks.

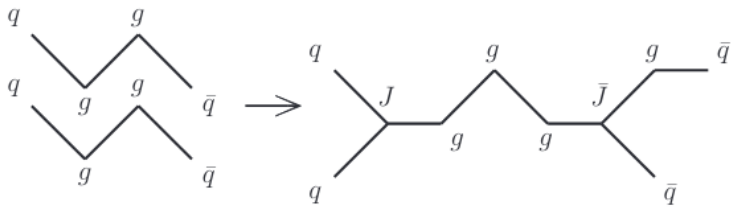
approach is called *colour-reconnection* (CR). The implementation of *colour reconnection beyond leading colour* (CR-BLC) in PYTHIA 8 is discussed in detail in [106] and briefly outined in the following. After first building up the LC strings following the final state PS, it is searched for allowed *reconnections* and possible configurations are sketched in Fig. 3.6. Via the type II and III reconnection, this enhances baryon production stronger than meson production since colour strings that would lead to meson formation are re-configured to those containing junctions which give rise to baryon formation.

(a) Type I: ordinary dipole-style reconnection

(b) Type II: junction-style reconnection



(c) Type III: baryon-style junction reconnection



(d) Type IV: zipper-style junction reconnection

Figure 3.6: Different colour reconnections allowed in the model developed in [106]. Reconnections of type II and III eventually cause additional baryon formation. Taken from the aforementioned reference.

3.3.2 Fragmentation plus coalescence

The fragmentation process can be combined with the partonic coalescence approach and used in MC event generators to model the hadronisation. In the following, a study is presented where PYTHIA 8 is interfaced with a coalescence hadronisation model applied in the HF sector of heavy-ion collisions.

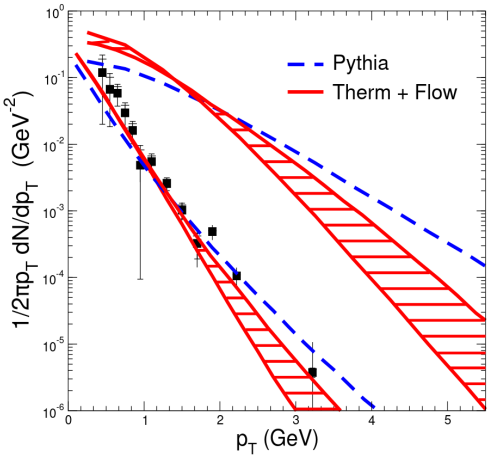


Figure 3.7: p_{T} spectrum of D-mesons [107]. Blue curves show the prediction neglecting the re-interaction of charm quarks in the medium while the red curves assume fully thermalised charm quarks. For details see text.

In addition to the assumption of a thermalised system of LF quarks, the partonic coalescence model has been further developed for charmed mesons [107] assuming however, that the abundance of charm quarks is determined from their production during the hard interaction. Two approaches have been investigated, namely no additional interactions of the produced charm quarks inside the medium and the scenario of fully thermalised charm quarks, which yield significant differences in the p_{T} spectrum of D-mesons as shown in Fig. 3.7. The blue and red curves are the predictions without interactions and fully thermalised charm quarks, respectively. The upper two curves correspond to the D-meson yield and the lower two curves are obtained after their semileptonic decay to $e^{\pm}$. The latter is compared to experimental results obtained at PHENIX studying gold–gold collisions at $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$ [108]. Given the precision of the measurements it is not possible to differentiate whether the data favours one prediction or the other. Overall, however, it can be seen that the model without re-interaction of charm quarks develops a harder spectrum.

3.4 Meson and baryon production in the heavy- and light-flavour sector

In general and as discussed before, hadronisation in pp collisions is described based on in-vacuum fragmentation being independent of the centre-of-mass energies. Furthermore, fragmentation is modelled independently of the collision system. In this section, experimental results are discussed that actually indicate modifications of this universal hadronisation mechanism in pp collisions, affecting mesons and baryons *differently*.

Two observables are considered in the following. One is the nuclear modification factor previously defined for di-jet production in Eq. (2.16). In this case it is defined as

$$\mathrm{R_{AA}} = \frac{\mathrm{d}^2 N_{\mathrm{h}}^{\mathrm{AA}}/\mathrm{d}p_{\mathrm{T}}\mathrm{d}\eta}{\langle T_{\mathrm{AA}}\rangle \mathrm{d}^2\sigma_{\mathrm{h}}^{\mathrm{pp}}/\mathrm{d}p_{\mathrm{T}}\mathrm{d}\eta}, \tag{3.8}$$

where $N_{\mathrm{h}}^{\mathrm{AA}}$ is the hadron multiplicity of species h in heavy-ion collision with nuclei of type A and $\sigma_{\mathrm{h}}^{\mathrm{pp}}$ is the production cross section in the pp reference collision system.

The second observable considered is the production of a given hadron h as a function of p_{T}. It is defined as

$$n^h = \frac{1}{N^{\mathrm{ev}}}\frac{\mathrm{d}N^{\mathrm{h}}}{\mathrm{d}p_{\mathrm{T}}}, \tag{3.9}$$

where N^{ev} is the number of events and $\mathrm{d}N^{\mathrm{h}}/\mathrm{d}p_{\mathrm{T}}$ is the p_{T}-differential hadron abundance. To infer on potential different hadron production mechanisms, the ratios based on the previous definition are defined as

$$\mathcal{N}^{\mathrm{h_1,h_2}} = n^{\mathrm{h_1}}/n^{\mathrm{h_2}} \tag{3.10}$$

with hadron species $\mathrm{h_1}$ and $\mathrm{h_2}$.

3.4.1 Meson ratios in the heavy-flavour sector

The partonic coalescence model has the ability to enhance strange hadron production in heavy-ion collisions relative to non-strange hadrons. On the other hand, no such enhancement is expected in pp collisions. This was studied by ALICE comparing the $\mathrm{R_{AA}}$ and production of $\mathrm{D_s^+}$ and non-strange D-mesons in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ [109] and discussed in the following.

The left panel of Fig. 3.8 shows the comparison of the $\mathrm{D_s^+}$ and non-strange D-meson $\mathrm{R_{AA}}$ as a function of the meson p_{T} for most central Pb–Pb collisions. The central points of the $\mathrm{D_s^+}$ $\mathrm{R_{AA}}$ (orange) are slightly higher, however, compatible with the $\mathrm{R_{AA}}$ of the non-strange D-mesons (black). Thus, within uncertainties, the production of both the $\mathrm{D_s^+}$ and non-strange D-mesons affected in the same way in presence of a medium.

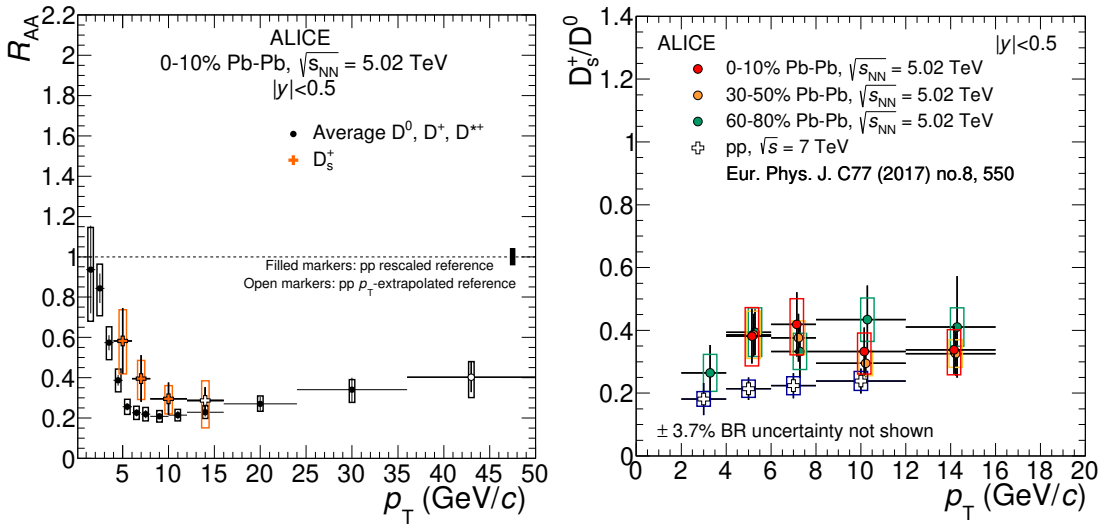


Figure 3.8: *Left:* Comparison of the D_s^+ R_{AA} to the non-strange D-meson R_{AA} as a function of hadron p_T measured in Pb–Pb collisions in the 0 % to 10 % centrality class. *Right:* Comparison of D_s^+/D^0 production ratio shown for three centrality classes compared to the value obtained in pp collisions. In both cases the measurements in different classes are compatible within uncertainties, respectively. Both plots taken from [109].

In the right panel, $\mathcal{N}^{D_s^+,D^0}$ is shown as measured in three centrality classes of Pb–Pb collisions and compared to pp results. All ratios are flat within uncertainties and the Pb–Pb ratios are compatible with one another. At the same time, the central points might indicate otherwise but within uncertainties the Pb–Pb ratios are compatible with the pp ratios. From these result, no conclusion can be drawn on a difference of the hadronisation mechanisms for strange and non-strange D-mesons.

To test the dependence of hadronisation on the centre-of-mass energy, various different D-meson ratios such as D^+/D^0, D^{*+}/D^0, D_s^+/D^0, and D_s^+/D^+ were measured by ALICE in pp collisions at $\sqrt{s}$ = 7 TeV and $\sqrt{s}$ = 5.02 TeV [110, 111]. The results are shown in Fig. 3.9. The ratios are seen to be compatible at different centre-of-mass energies and no dependence for different HF mesons is observed.

So far, no modifications of the hadronisation process, which would have a different impact on different HF mesons, have been observed, neither in Pb–Pb nor in pp collisions. The aforementioned observations will thus be complemented by measurements of HF baryon production and its ratio to mesons.

3.4.2 Baryon-to-meson ratios in the heavy-flavour sector

Predictions for the Λ_c^+/D^0 production ratio in heavy-ion collisions have been made available where a different combinations of fragmentation and partonic coalescence [112] were studied. Fig. 3.10 shows a prediction for three scenarios (i) using fragmentation

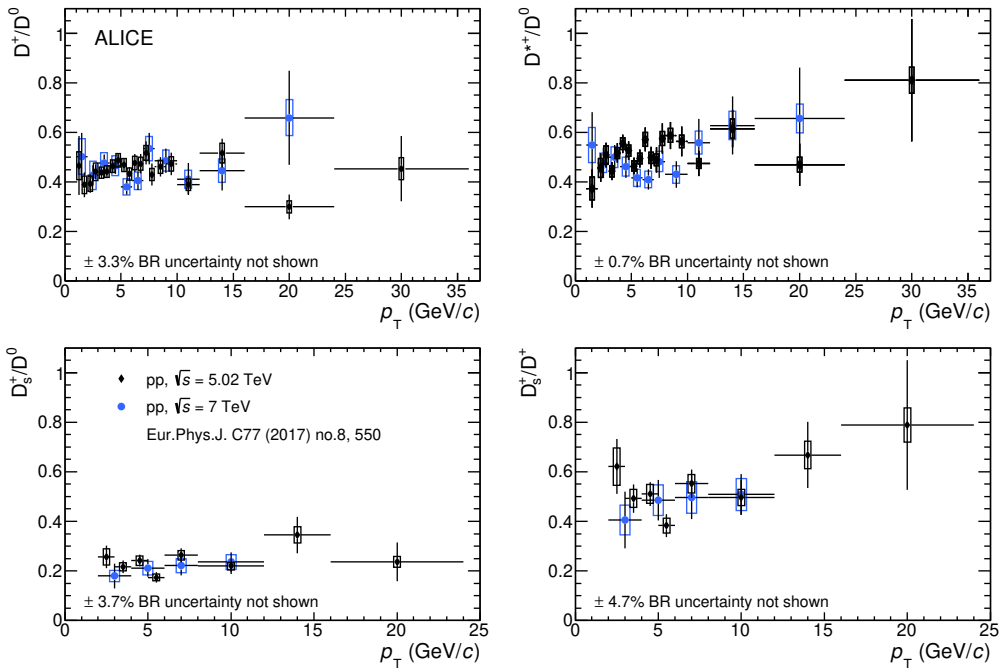


Figure 3.9: D^+/D^0, D^{*+}/D^0, D_s^+/D^0 and D_s^+/D^+ production ratios measured in pp collisions at $\sqrt{s} = 5.02\,\text{TeV}$ [111] compared to results obtained at $\sqrt{s} = 7\,\text{TeV}$ [110] from *top left* to *bottom right*. The comparison agrees within uncertainties and no p_T dependence is observed.

and coalescence (green), (ii) only coalescence (black) and (iii) like (i) but normalising the fragmentation to the fraction obtained by PYTHIA 8 using the Angantyr model for heavy-ion collisions [113]. The two plots show the predictions for Au–Au collisions at RHIC ($\sqrt{s_{NN}} = 200\,\text{GeV}$) and Pb–Pb collisions at LHC ($\sqrt{s_{NN}} = 2.76\,\text{TeV}$) in the left and right panel, respectively. The model using only coalescence is comparable in magnitude and shape in both cases. On the other hand, the coalescence plus fragmentation scenario predicts a lower ratio for the LHC-scenario. However, these models indeed predict an enhanced Λ_c^+/D^0 ratio compared to models using only fragmentation and are tuned to fragmentation functions as obtained in e^+e^- and e^-p systems. In that case, the ratio is predicted to be flat and $\simeq 0.1$. On top of that, the predicted ratio when incorporating partonic coalescence shows a shape increasing from low to intermediate p_T and then decreasing for higher p_T.

Although the first predictions of the Λ_c^+/D^0 production ratio were made for nucleon–nucleon collisions, it was a striking result to see an enhancement in pp collisions at $\sqrt{s} = 5.02\,\text{TeV}$ [115]. The central result plot is shown in Fig. 3.11 where the experimental results are compared to different model predictions. It can clearly be seen that the ratio exceeds 0.1 and also has a shape as a function of p_T.

In contrast to the discussed results of HF mesons in the previous section, this result is an indication for hadronisation beyond pure in-vacuum fragmentation. The

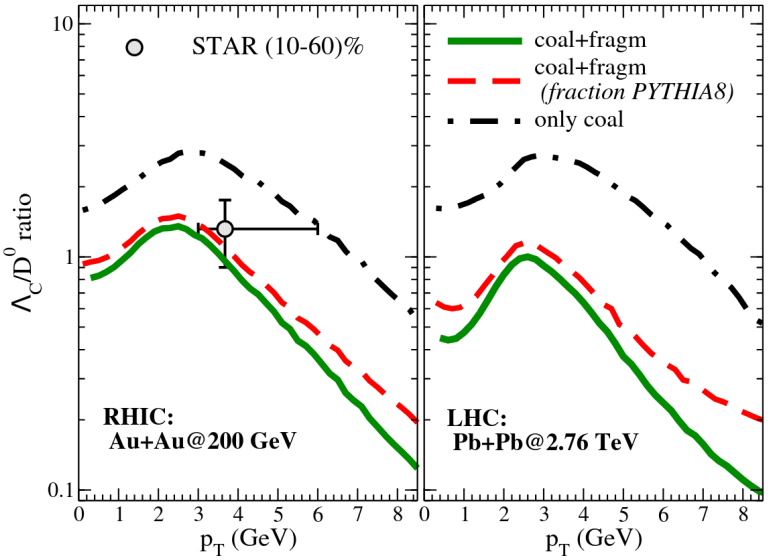


Figure 3.10: Prediction of the Λ_c^+/D^0 production ratio in Au–Au collisions at RHIC energies of $\sqrt{s_{NN}} = 200\,\text{GeV}$ and in Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76\,\text{TeV}$ as probed at the LHC [112]. Experimental data obtained by STAR is taken from [114].

predictions from PYTHIA 8 (Monash tune [116]) and HERWIG7 [95] with charm fragmentation tuned to measurements in e^+e^- and e^-p collisions [27, 28] therefore clearly underestimate the data (dashed green and purple lines) predicting an overall flat ratio.

Two additional approaches are based on statistical hadronisation [117]. The two shown predictions are based either on a charmed baryon spectrum taken from PDG [13] (SH model + PDG, dark green) or were augmented to include additional excited baryon states predicted by the Relativistic Quark Model (RQM) [118] (SH model + RQM, light green). The further enhancement of the latter is due to the additional baryon states since those decaying to Λ_c^+ baryons further enhance their abundance. At the same time it predicts a decrease of the ratio with increasing p_T which is also seen in data.

The PYTHIA 8 tune implementing colour reconnection beyond leading colour approximation [106] (as introduced also in Sec. 3.3) is also able to reproduce the shape qualitatively and predicts an enhanced ratio (dashed red line).

Another model which predicts an enhancement and the falling trend with increasing p_T is based on the assumption of additional hadronisation via coalescence also in pp collisions [112] (blue line). The predictions are based on the model as introduced in the beginning of this section.

Given the current precision of the measurements and multiple predictions based on different model assumptions that qualitatively described the data to a similar degree, it is yet unclear what the actual underlying hadronisation processes are that

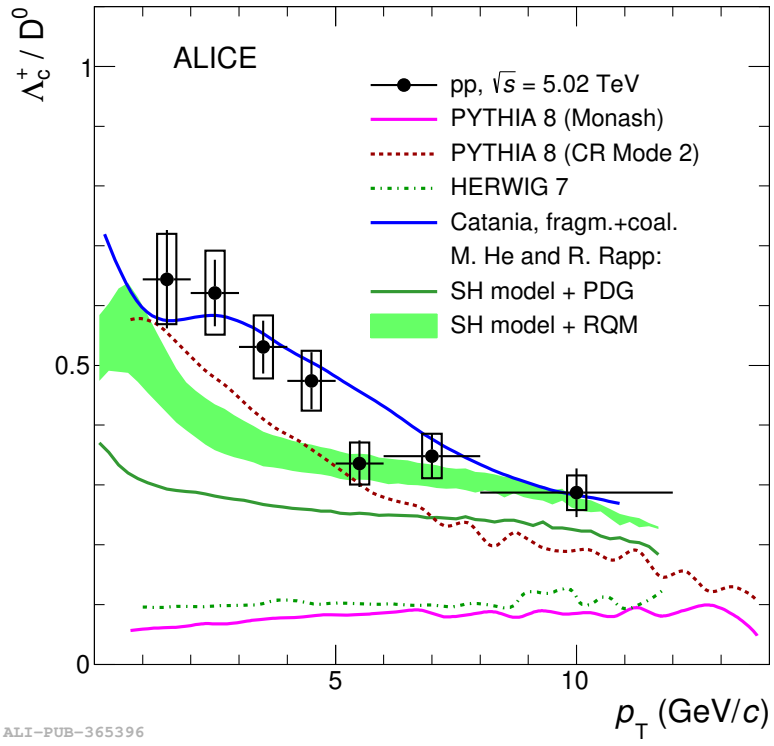


Figure 3.11: Λ_c^+/D^0 production ratio extracted in pp collisions at $\sqrt{s}$ = 5.02 TeV [115] compared to different model predictions (see text for further explanations).

is implemented by nature and cause the difference in charmed baryon and meson production. To draw further conclusions, additional measurements are necessary, especially to enhance the precision. The current experimental picture of production ratios in the HF sector is complemented by spanning the discussion also to the light-flavour sector in order to broaden the picture.

Extensive studies of baryon and meson production as well as their corresponding ratios have been conducted in the light-flavour sector by ALICE in Pb–Pb [20, 21], p–Pb [22–24] and pp [25] collision systems in intervals of charged particle multiplicity. A central result is shown in Fig. 3.12 where each row contains another particle ratio as a function of p_T in different collision systems per column (pp, p–Pb and Pb–Pb from left to right). The studied ratios are $(p+\bar{p})/\phi$, $(K^+ + K^-)/(\pi^+ + \pi^-)$, $(p+\bar{p})/(\pi^+ + \pi^-)$, and Λ/K^0_S. Each panel includes the ratios obtained in the lowest and highest multiplicity class in the studied collision system*.

Each ratio shows qualitatively the same trend across different collision system. In addition, the baryon-to-meson ratios, $(p+\bar{p})/(\pi^+ + \pi^-)$ and Λ/K^0_S show a significant enhancement in the high multiplicity interval across all studied collision systems.

*An extensive discussion of a potential medium evolution and its impact in pp collisions as well as comparisons to larger collision systems can be found in [25].

The ratio at low p_{T} rises until reaching a maximum around $p_{\mathrm{T}} \approx 2 - 3\,\mathrm{GeV}/c$ and then falling off towards higher momenta. This trend is visible in both multiplicity classes but more pronounced in the highest. This enhancement can be seen as an indication for a hadronisation mechanism modifying the pure vacuum fragmentation affecting baryons and mesons differently. While this can be understood in terms of a thermalised system created in heavy-ion collision systems, it is more puzzling how this comes about in smaller systems and even in pp collisions.

The $(p + \bar{p})/\phi$ ratio falls with p_{T} in pp collisions but becomes more and more flat starting from low p_{T} in the highest multiplicity class in p–Pb. This trend can be seen even in the lowest multiplicity class in Pb–Pb and even stronger in the highest. The authors of [25] interpret this as a possible signature for radial flow in large collision systems.

The ratios were also analysed as a function of multiplicity as shown in Fig. 3.13, with increasing intervals of p_{T} from left to right and different particle ratios per row. Overall, an approximately smooth transition from small to large collision systems is observed. In particular, the baryon-to-meson ratios, $(p + \bar{p})/(\pi^+ + \pi^-)$ and $\Lambda/\mathrm{K}^0_{\mathrm{S}}$, in the two bottom rows show a strong decreasing (increasing) trend with multiplicity in the first (second) p_{T} interval while this is much less pronounced in the other two ratios in the top rows. At high p_{T} the ratio tends to become stable. The baryon-to-meson ratio which shows a dependence on multiplicity could finally indicate that there is no qualitative differences with respect to underlying hadronisation mechanisms between small and large hadronic collision systems. It could be indeed their peculiarity which varies with multiplicity or rather with the complexity of the system whose development *results* in the overall charged particle multiplicity.

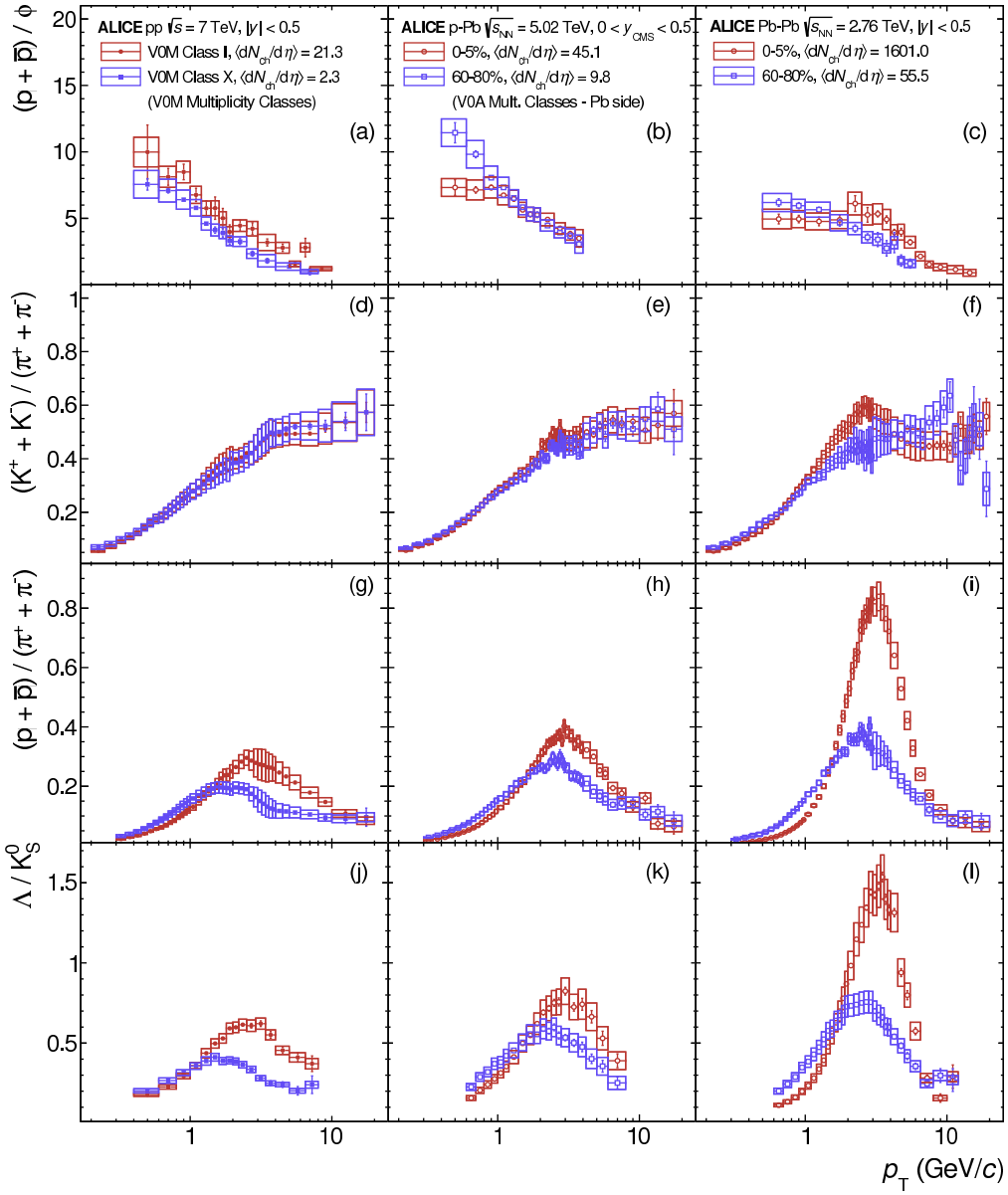


Figure 3.12: Different particle ratios $\mathcal{N}^{h_1,h_2}$ measured in the LF sector per row and in different collision systems pp, p–Pb and Pb–Pb from left to right. Each panel shows the lowest and highest multiplicity interval considered in the corresponding collision system [25].

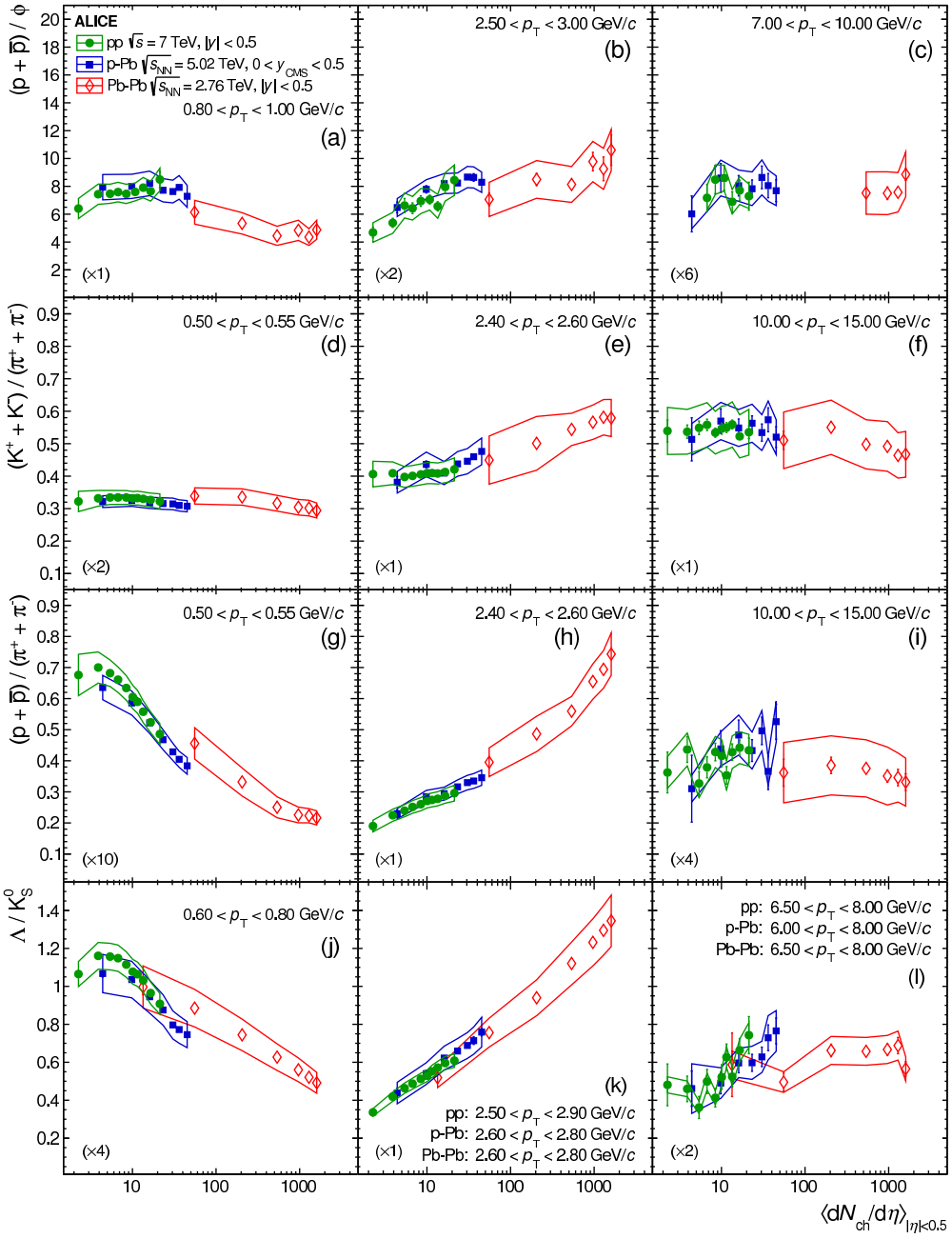


Figure 3.13: Different particle ratios $\mathcal{N}^{h_1,h_2}$ as a function of multiplicity in the LF sector per row and in increasing p_T intervals from left to right. Each panel shows the ratio obtained in different collision systems pp, p–Pb and Pb–Pb shown in green, bue and red, respectively [25].

4 The ALICE experiment at the LHC

In the following, the experimental setup of the Large Hadron Collider (LHC), which delivers the particle collisions, and the ALICE detector will be explained in Sec. 4.1 and Sec. 4.2. The focus lies in particular on the features and functionalities employed in the data analysis presented in Ch. 5. For further reading and detailed information it is referred to the provided references.

4.1 The LHC

The Large Hadron Collider [119] at CERN is a circular particle collider and was built as the successor of the Large Electron-Positron Collider (LEP) and uses the same 26.7 km long tunnel system for the beam line. It is situated around 100 m underground below the boarder region of France and Switzerland near Geneva.

At a hadron collider, much higher collision energies are accessible compared to lepton colliders which, in the case of a e^+e^- collider, are mostly limited by the energy loss due to synchrotron radiation which scales with m^{-4} of the accelerated particle mass. The highest centre-of-mass energy reached at LEP was 209 GeV in the year 2000. The LHC was hence built to yield the highest particle collision energies ever reached in an experimental environment to study potential new physics, in particular to be able to find or exclude the predicted SM Higgs boson. It also provides energy densities as they have existed in the early universe by not only colliding protons but also heavy ions such as lead during a yearly Pb–Pb run. In addition, there was a one-day run of Xenon–Xenon collisions on the 12 October 2017 with a per-nucleon centre-of-mass energy of 5.44 TeV.

The beams are guided in two separate beam pipes which are crossed at four points where the beams can be brought to collisions. At these points the four main experiments, namely ALICE [120], ATLAS [121], CMS [122] and LHCb [123] are installed, each following a particular physics programme. The ALICE experiment is dedicated to physics concerning the QCD sector by investigating pp and heavy-ion collisions. It was designed in particular to study the dynamics in large collision systems such as the properties of the Quark-Gluon-Plasma and to explore the QCD phase diagram. The detector will be introduced in more detail in the next section together with a brief overview of the upgrades implemented for the upcoming Run 3 and their physics motivations. LHCb focuses on the study of *beauty*-physics, for instance in order to gain important understandings of the CP-violating properties of and beyond the SM. The ATLAS and CMS experiments were built as multi-purpose

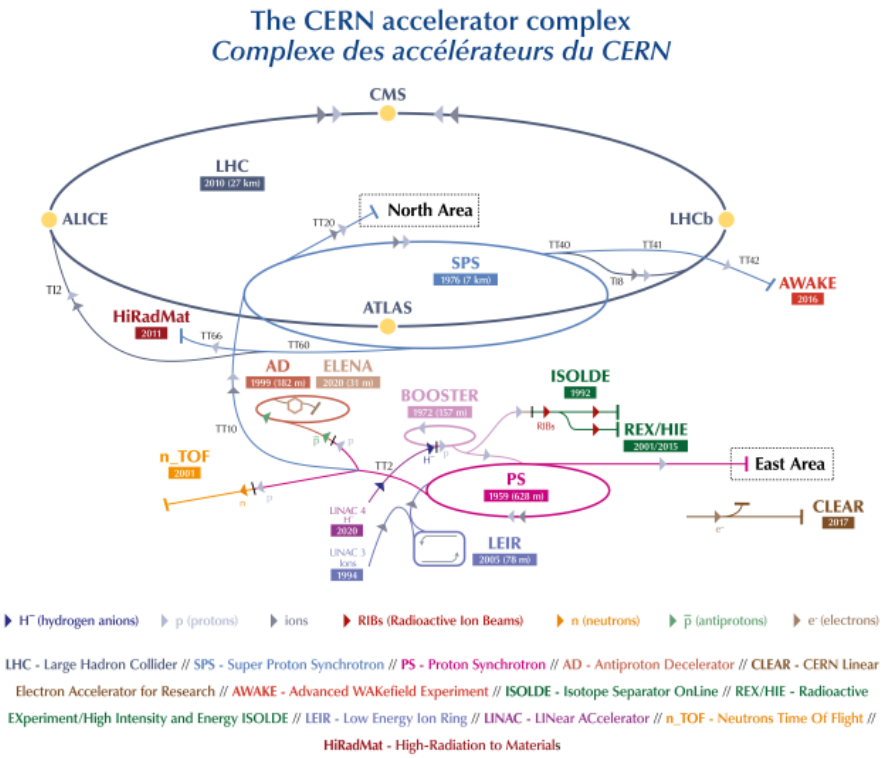


Figure 4.1: Accelerator chain at CERN.

experiments and were in particular optimised to search for the SM Higgs boson.

The LHC ring is the final stage in an acceleration chain which makes use of previously installed accelerators as shown in Fig. 4.1. Lead ions are obtained from heating a ^{208}Pb sample and ionising the produced gas. During the acceleration through LINAC3 the electrons are stripped off entirely and the lead nuclei are accelerated to a per-nucleon energy of 4.5 MeV, then forwarded to the Low Energy Ion Ring (LEIR) which they leave as four bunches with a per-nucleon energy of 72 MeV. Via the Proton Synchrotron (PS) followed by Super Proton Synchrotron (SPS) they reach the LHC and are injected with a per-nucleon energy of 177.4 GeV in opposite directions.

Proton beams are initiated at different linear accelerators. Until the end of Run 2, the LINAC2 facility was used. Protons were extracted from ionised hydrogen gas and accelerated to 50 MeV to be injected into the PS where the particles are organised into bunches. With an energy of 26 GeV these were forwarded to the SPS to be accelerated up to 450 GeV before they were injected into the LHC. As of Run 3, the new LINAC4 facility will be used.

The LHC is designed to house up to 2808 proton and 592 lead bunches per beam which can be accelerated up to 7 TeV or 2.759 TeV per nucleus, respectively, by a series of radio frequency cavities. A magnetic dipole field of $|B| = 8.33\,\mathrm{T}$ forces the beams onto the desired nominal circular orbit. Quadrupole magnets control

the transverse dimensions of the beam and higher order multipole magnets are used to apply finer corrections to the beam orbit. If necessary, the beams can safely be dumped either for safety reasons or when a re-fill is required due to degrading luminosity which is mainly caused by its decay due to collisions.

The *instantaneous luminosity* is defined as

$$\mathcal{L} = \frac{N_1 N_2 n_\mathrm{b}}{4\pi\sigma_x\sigma_y} f, \tag{4.1}$$

where N_i is the number of particles per bunch in beam i, n_b is the number of bunches in each beam, $\sigma_{x/y}$ specify the transverse spread of the beams and f is the evolution frequency of the bunches. Since the number of occurrences of a process of interest N_proc with cross section σ_proc is directly proportional to the integrated luminosity, $N_\mathrm{proc} = \sigma_\mathrm{proc} \int \mathcal{L} \mathrm{d}t$, high luminosities increase the probability of collecting events containing rare processes. The design peak luminosity of the LHC in pp collisions was $\mathcal{L}_{pp} = 1 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ [119] but more than twice this value was reached in 2018 [124]. The highest possible values are reached at the collision points of the ATLAS and CMS experiments. The instantaneous luminosity is instead levelled [125] for ALICE and LHCb by introducing an offset between the incoming beams, effectively decreasing the instantaneous luminosity. For instance in the case of ALICE, the drift time in the Time Projection Chamber sets the limit on the number recordable collisions per time. With time, the beam offset can be decreased resulting in a constant luminosity. The peak luminosity for Pb–Pb collisions is $\mathcal{L}_{PbPb} = 10 \times 10^{27}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

In the left panel, Fig. 4.2 shows the integrated luminosity $L = \int \mathcal{L} \mathrm{d}t$ over time acquired by the ALICE experiment in pp collisions during Run 2 which spanned from 2016 to 2018. A constant increase employing different triggers was achieved over all data-taking periods resulting in, for instance, around 3 billion recorded minimum-bias events and 1.7 billion high-multiplicity events collected with the V0 trigger. The right panel contains the corresponding performance achieved during Pb–Pb data-taking runs in Run 2 where overall more then 600 million events were collected.

After Run 2 the LHC as well as the experiments went into the period of long-shutdown (LS) 2 which is used for maintenance work and upgrades. Run 3 is expected to start in April 2022 and shall last 3 years as indicated by the schedule shown in Fig. 4.3. The delivered integrated luminosity is foreseen to be doubled aiming for $350\,\mathrm{fb}^{-1}$ in pp collisions. During the final LS3, the LHC shall be upgraded to become the High-Luminosity LHC to deliver up to 7.5 times the Run 3 luminosity and is expected to operate until 2040.

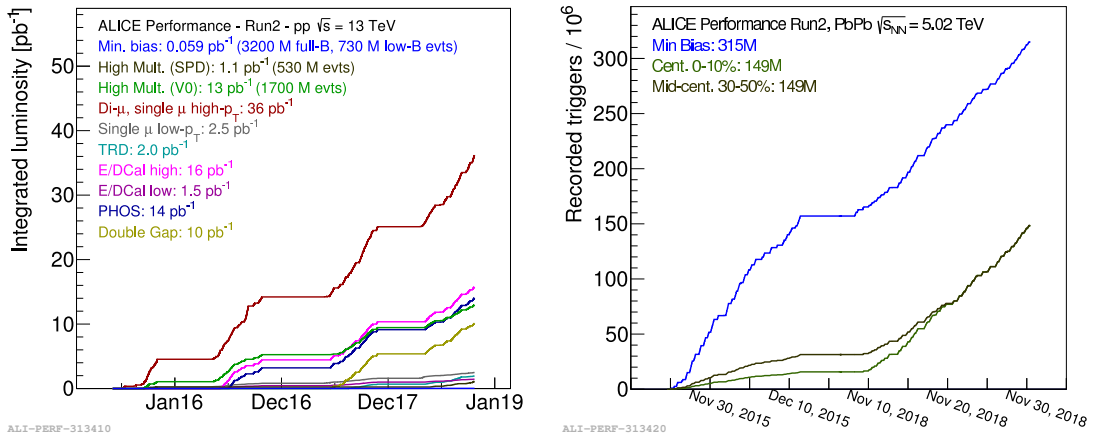


Figure 4.2: pp *(left)* and Pb–Pb *(right)* integrated luminosities per trigger recorded by ALICE during Run 2.

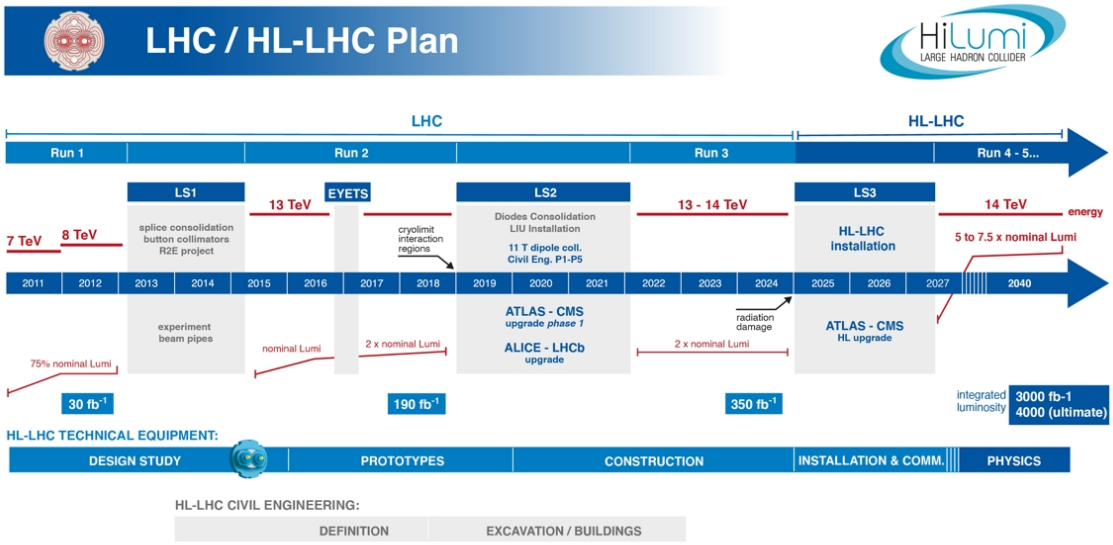


Figure 4.3: LHC schedule and planned upgrades towards High-Luminosity LHC.

4.2 The ALICE experiment

The ALICE detector [120] has been built as a multi-purpose detector with the particular focus on the physics of QCD sector of the SM. Therefore, it is optimised for highly dense systems as produced in heavy-ion collisions. A value of 8000 particles per rapidity unit at midrapidity was first estimated from data taken at the SPS [126] but was then known to be around values of $2000 - 3200$ after the first data was published from RHIC [127]. Indeed values up to 1943 ± 54 have been measured in the ALICE experiment [128]. To be able to resolve all tracks in such a dense environment, the choice was made to employ the conservative design of a time projection chamber as the main tracking device. This implied the trade-off of a limited readout rate of up to 1 kHz during Run 1 and 2.

The ALICE apparatus is capable of tracking charged particles over a wide momentum range starting from few tens of MeV/c up to hundreds of GeV/c. At the same time, reliable particle identification (PID) can be provided for all tracked particles using the information of different subdetectors. These excellent tracking and PID capabilities are required to explore the QCD phase diagram, for instance by means of particle production and correlations measurements.

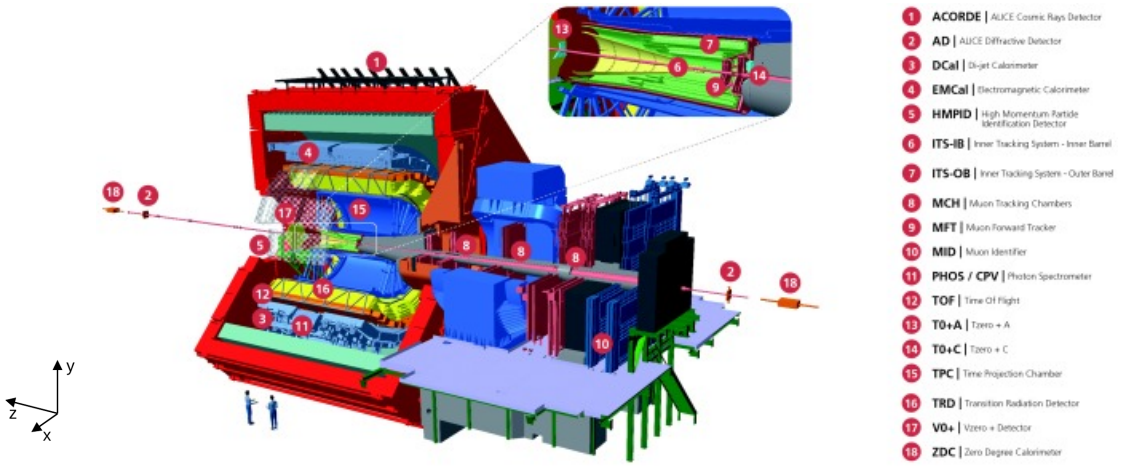


Figure 4.4: The ALICE detector as configured during Run 3. Its global structure is given by the barrel and the forward muon arm. Components such as the ITS, the TPC and the TOF detectors cover the full azimuth and are embedded in a solenoid magnet. Forward detectors such as the V0, T0 and ZDC detectors are placed close to the beam line.

These requirements have been met by the detector design as sketched in Fig. 4.4 showing the configuration for Run 3. However, the overall design and placement of detector components covers to a large extend the previous configuration consisting of the central barrel and the dedicated forward muon arm. A summary of the dimensions, rapidity coverage and number of readout channels as of Run 2 of the detectors

detector	acceptance η	distance [m]	dimension [m^2]	channels
ITS layer 1, 2 (SPD)	± 2, ± 1.4	0.039, 0.076	0.21	9.8 M
ITS layer 3, 4 (SDD)	± 0.9, ± 0.9	0.150, 0.239	1.31	133000
ITS layer 5, 6 (SSD)	± 0.97, ± 0.97	0.380, 0.430	5.0	2.6 M
TPC	± 0.9 at $r = 2.8$ m ± 1.5 at $r = 1.4$ m	0.848, 2.466	readout 32.5 volume 90 m	557568
TOF	± 0.9	3.78	144	157248
V0A	$2.8 < \eta < 5.1$	3.4	0.548	32
V0C	$-3.7 < \eta < 1.7$	-0.897	0.315	32

Table 4.1: Solid angle coverage, spatial dimensions and number of readout channels of detectors relevant for the analysis discussed in Ch. 5 adapted from [120]. All listed detectors cover the full azimuth. The distance is given as the radial distance to the beam pipe except for the V0 detectors where the longitudinal distance from the nominal interaction point is given. For the TPC the volume is specified in addition to the readout area.

crucial for the analysis in Ch. 5 can be found in Tab. 4.1 which has been adapted from [120].

The largest volume of the barrel consists of low material-budget detectors for tracking and PID purposes, embedded in a moderate solenoid magnetic field of 0.5 T pointing along the beam direction to support good momentum resolution of low-p_{T} tracks of better than 10 % up to around 100 GeV/c (see also Sec. 4.2.5). The main tracking and PID detectors such as the Inner Tracking System (ITS), the Time Projection Chamber (TPC), the Transition Radiation Detector (TRD) and Time Of Flight (TOF) have a cylindrical shape covering the full azimuth and in their entirety the rapidity region of approximately $|\eta| < 0.9$. The HMPID systems adds additional hadron PID capabilities for high-momentum hadrons in a limited solid angle covering around 10 % of the barrel acceptance. Further PID capabilities and jet energy measurements are devised by the electromagnetic and di-jet calorimeter, also with a limited azimuthal and rapidity coverage. The forward muon arm is devoted to the reconstruction and identification of muons, covering $-4.0 < \eta < -2.5$. In order to track and identify muons, a system of hadron absorbers culminating in a iron wall is used. Tracking detectors based on Resistive Plate Chambers (RPCs) (Muon Chambers, or MCH) are situated after the front absorber while the Muon Trigger system (MTR) is situated after the last iron wall. Note that what is called MID in the sketch corresponds to the MTR in Run 2.

Additional forward detector systems have been installed for trigger purposes and the determination of event characteristics. The two Cherenkov counters of the T0 detector are used for event time measurements and are installed on either side of the interaction point (IP) at a distance of 3.75 m and -0.727 m, respectively.

Two scintillator arrays of the V0 detector devise triggering capabilities as well as event multiplicity estimation and are installed at longitudinal distances of 3.4 m and −0.897 m on each side of the IP. The ZDC system detects collision remnants at very forward rapidity and provides trigger information through two pairs of compact calorimeter systems, namely the electromagnetic (ZDC:ZEM) and hadronic calorimetry (ZDC:ZN and ZDC:ZP) at distances of ±7 m and ±113 m, respectively.

In the following, the components crucial for this thesis are presented in further detail as installed during Run 2 focusing on their performance in pp collision. This is followed by the introduction of the tracking and PID procedures in Sec. 4.2.5 and Sec. 4.2.6, respectively. Sec. 4.2.7 gives a brief overview of the trigger system and Sec. 4.2.8 introduces the multiplicity estimation in pp collisions. Future physics perspectives provided by the upgraded ALICE detector are outlined in Sec. 4.2.9.

4.2.1 Inner Tracking System

The ITS as sketched in Fig. 4.5 is the first detector symmetrically surrounding the beam pipe at the IP. It consists of three pairs of silicon pixel detectors (SPD), silicon drift detectors (SDD) and silicon strip detectors (SSD), respectively, from inside-out. The first SPD layer has a radius of 3.9 cm and the outermost SSD layer is 43 cm away from the beam pipe. The ITS covers $|\eta| < 0.9$ of pseudorapidity with an extended coverage of the innermost layer of $|\eta| < 2$. Accounting for the thermal shield, support structures as well as the sensitive detector and readout material, it has an overall low effective thickness of 7.18 % of a radiation length, 7.26 % X_0 including air, hence minimising its effect on traversing particles.

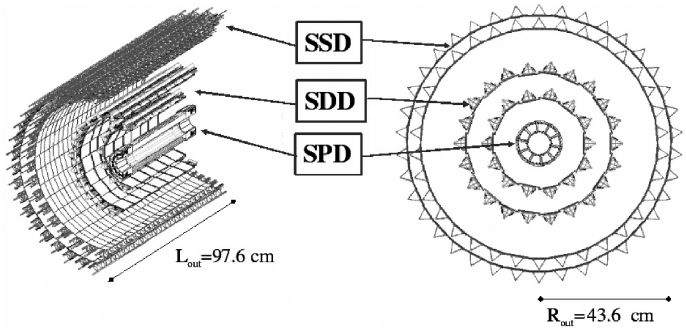


Figure 4.5: Technical drawing of the ITS showing the three pairs of SPD, SSD and SDD layers covering the full azimuth around the beam line.

The segmentation of the layers allows for the simultaneous finding of 15000 tracks which corresponds to the estimate of 8000 particles per unit rapidity at midrapidity. Overall, two tracks can be resolved at distances lower than 100 µm in the transverse plane and at the order of 100 µm in the longitudinal direction. The aforementioned specifications are based on its two main purposes to determine the tracks' impact

parameter as well as to find primary and secondary vertices, both crucial for heavy-flavour particle reconstruction. For tracks with $p_T > 1\,\mathrm{GeV}/c$ the impact parameter resolution is $\sim 10\,\mu\mathrm{m}$. It serves as a tracking detector, either stand-alone and specifically for low-momentum tracks or together with the TPC detector. Additional stand-alone PID capabilities for low-p_T particles below $200\,\mathrm{MeV}/c$ are provided by the analogue readout of the SDD and SSD layers.

4.2.2 Time Projection Chamber

The TPC detector as sketched in Fig. 4.6 surrounds the ITS with an inner radius of 0.85 m, an outer radius of 2.50 m and a length of 5 m along the beam direction and devises the principal tracking apparatus. While covering the full azimuth, it has a forward coverage of $|\eta| < 0.9$ containing the full radial track length and extends to $|\eta| < 1.5$ for 1/3 of the radial track length. The central high-voltage electrode has a thickness of $\approx 22\,\mu\mathrm{m}$ to maintain a minimal material budget at $\eta = 0$ and establishes a uniform electrostatic field between either of the end plates. The drift volume is filled with a mixture of Ne, CO_2 and N_2 with ratios 16 : 2 : 1 to maximise the drift speed while keeping the diffusion and radiation length small, the latter being approximately 3.5 % X_0 at $\eta = 0$. The readout of drift electrons is devised by multi-wire proportional chambers mounted on each end plate. An azimuthal segmentation into 18 trapezoidal sectors was chosen per end plate, each radially divided into 2 to deal with higher track densities towards the centre. The sectors are divided into 159 pad rows hence each track traversing the TPC volume can potentially induce a signal in each of the rows. In addition to tracking it is also a main PID device given its excellent energy loss $\mathrm{d}E/\mathrm{d}x$ resolution of 5 % which is further outlined in Sec. 4.2.6.

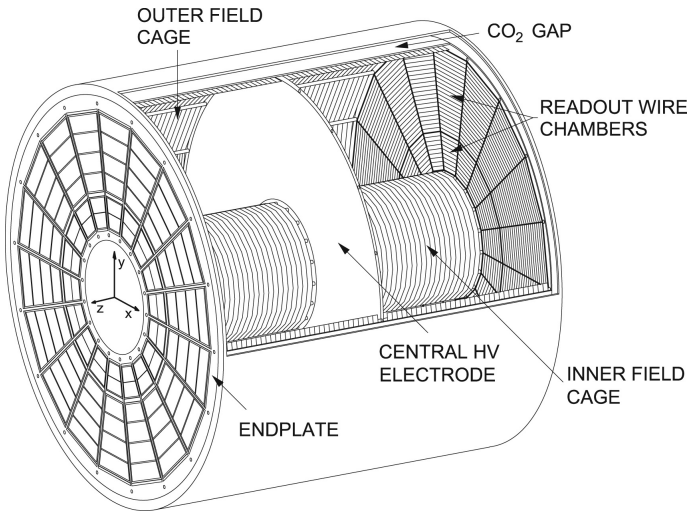


Figure 4.6: Sketch of the TPC volume with the high-voltage electrode in the centre plane and the multi-wire proportional chamber mounted on 18 segments per end plate.

4.2.3 Time Of Flight

The TOF detector is built from 18×5 segments in $\phi \times z$, each of which is made of 10-gap double stack Multi-gap Resistive Plate-Chamber (MRPC) strips and inscribed into a cylindrical volume covering the full azimuth and the pseudorapidity interval of $|\eta| < 0.9$. It has an inner and outer radius of 37.0 m and 39.9 m, respectively. As the outermost barrel PID detector a higher effective thickness of approximately 30 % X_0 is acceptable. The occupancy at highest estimated multiplicities is calculated to be below 15 % and 10^5 readout channels were estimated to be sufficient. For tracks up to $\approx 1\,\mathrm{GeV}/c$ the TOF detector together with the ITS and TPC can be included in the tracking and vertex finding as well as in the $\mathrm{d}E/\mathrm{d}x$ determination. As one of the principal PID detectors it can identify pions and kaons with momenta $\lesssim 2.5\,\mathrm{GeV}/c$ and protons with momenta $\lesssim 4\,\mathrm{GeV}/c$.

4.2.4 V0

The V0 detector consists of two scintillator arrays denoted as V0A and V0C, displaced by 3.4 m and -0.897 m from the IP along the beam line. They cover pseudorapidity regions of $2.8 < \eta < 5.1$ and $-3.7 < \eta < -1.7$, respectively. The V0 detector devises event multiplicity estimation in the forward region by measuring the deposited charge induced by particles coming from the interaction and travelling along the beam line. At the same time, it provides triggering capabilities. Both the triggering and multiplicity estimation are further described in Sec. 4.2.7 and Sec. 4.2.8.

4.2.5 Vertex and track reconstruction

Vertex and track reconstruction is done in three iterations as described in [129, 130] and the following outlines the procedure in particular as employed in pp collisions. In a first step, vertices are estimated from hits in the SPD layers. Two hits in the first and second SPD layer are combined to form tracklets and vertices are estimated from their minimal spatial distance. The vertex with the most tracklets connected to it is assumed to be the primary vertex, other potential vertices are flagged as pile-up vertices.

During the first iteration, tracks are reconstructed from outside-in, starting from seeds at the outer pad-rows of the TPC. Two clusters in combination with the vertex position are used as a first seed, later followed by a seed constructed from three clusters without the vertex position. An inward propagation finds the next cluster compatible with the previous one. A Kalman filter is used to update the track parameters. If clusters are shared among multiple track candidates, those tracks with worse reconstruction quality are rejected by defining a maximum fraction of allowed shared clusters. Each track is then propagated further to the innermost pad row and finally accepted when at least 20 clusters could be assigned to it and the ratio of the number of clusters over crossed rows is larger than 0.5. At this point, found

tracks are stored as TPC-only tracks. The left panel in Fig. 4.7 shows the simulated TPC track finding efficiency as a function of track p_{T} and is found to be $\gtrsim 80\,\%$ for track momenta of $1\,\mathrm{GeV}/c$ to $10\,\mathrm{GeV}/c$ which is compatible across different collision topologies of central and peripheral Pb–Pb as well as pp collisions. It drops significantly below $p_{\mathrm{T}} \lesssim 500\,\mathrm{MeV}/c$ due to energy loss caused by the interaction with the detector material. After a maximum at low p_{T} the efficiency drops slightly caused by dead zones in the TPC.

The inward position is used as the initial seed for the propagation into the ITS down to the vertex which is as well based on a Kalman filter approach to find the next clusters compatible with the previous one. In the end, the combination of clusters inside the ITS with the best quality parameters is chosen to build the final track candidate.

Due to a lower performance for particles of $p_{\mathrm{T}} < 200\,\mathrm{MeV}/c$ caused by multiple scattering, an additional ITS-only tracking is performed using two clusters of the three innermost ITS layers and the primary vertex position. The used algorithm is analogous to the one employed during the TPC track reconstruction mentioned before but now starting from the vertex towards the outer layers of the ITS.

The second iteration starts from the ITS vertex and tracks are refitted outwards again using a Kalman filter based algorithm up to the outer radius of the TPC. From there, the tracks are further extrapolated to the TRD and trying to be matched with TRD tracklets. If the TRD information is not available, the algorithm attempts to match the tracks with clusters found in the TOF detector. During this iteration the integrated track length as well as timing information are computed and the tracks are extrapolated to the following detectors. As the final third step the tracks are re-fitted again from outside-in using all available cluster information obtained from the previous passes. The vertices are fitted once more reducing the impact parameter w.r.t. each reconstructed track according to a χ^2 fit as described in detail in [130].

This procedure results in a set of SPD vertices, a set of re-fitted vertices built from tracks, both including potential pile-up vertices, a set of TPC-only tracks (and potentially ITS-only tracks for low-p_{T} particles) and a set of tracks reconstructed using all of the previously mentioned detectors.

The right panel of Fig. 4.7 shows the momentum resolution of reconstructed tracks in p–Pb collisions $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ as a function of $1/p_{\mathrm{T}}$, either using the TPC stand-alone or in combination with the ITS. Using

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} = p_{\mathrm{T}}\sigma_{1/p_{\mathrm{T}}} \tag{4.2}$$

it is seen that the resolution reaches values of about 10 % for momenta of the order $100\,\mathrm{GeV}/c$ and enhances to $\approx 1\,\%$ for momenta of $\approx 1\,\mathrm{GeV}/c$.

The decays of neutral long-lived particles such as the $\mathrm{K}^0_{\mathrm{S}}$ or Λ give rise to secondary decay vertices. These are reconstructed by iterating over all pairs of oppositely-charged tracks which have an impact parameter of $> 0.5\,\mathrm{mm}$ w.r.t. the primary vertex. These are combined into V^0 candidates in case the point of closest approach

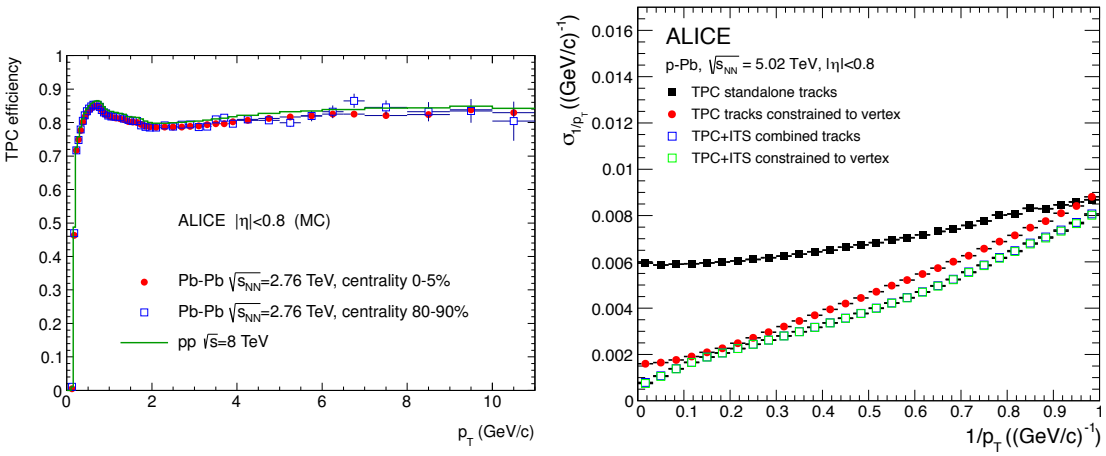


Figure 4.7: *Left:* Simulated TPC track finding efficiency in central (red) and peripheral (blue) Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76\,\text{TeV}$ as well as in pp collisions (green) at $\sqrt{s} = 7\,\text{TeV}$. *Right:* Momentum resolution of tracks using the TPC stand-alone (black and red) as well as together with the ITS (blue and gree) measured in p–Pb collisions at $\sqrt{s_{NN}} = 5.02\,\text{TeV}$. Both plots from [129].

(PCA) is below 1.5 cm and lies between the primary vertex and innermost hit of either of the potential V^0 tracks. Since they are expected to originate from the primary vertex or from an immediate decay of a particle produced in the collision, the cosine of the pointing angle θ_p is required to be > 0.9 for candidate momenta of $> 1.5\,\text{GeV}/c$. Below those values the criterion is relaxed due to possible multiple scattering.

4.2.6 Particle identification

Based on its particular observables, each detector α with PID capabilities assigns a signal S^α to each seen track. A signal S^α follows the conditional probability densities $P(S^\alpha|H_i)$ for a given particle hypothesis H_i with standard deviation σ_i^α. The weighted difference of the signal w.r.t. the expectation $\hat{S}_i^\alpha$ is given as

$$n_{\sigma_i^\alpha} = \frac{S^\alpha - \hat{S}_i^\alpha}{\sigma_i^\alpha}, \tag{4.3}$$

which can be used to cut on different PID hypotheses where a tighter cut corresponds to a stricter requirement on a certain hypothesis H_i. A combination of different detectors can be obtained by building a likelihood from the various probability densities and computing the posterior probability $P(S^\alpha|H_i)$ using Bayes' theorem according to

$$P(S^\alpha|H_i) = \frac{L(H_i|S^\alpha)\cdot\pi(H_i)}{\sum\limits_{j\in\{\text{particles}\}} L(H_j|S^\alpha)\cdot\pi(H_j)}, \tag{4.4}$$

where $\pi(H_i)$ is the prior for hypothesis H_i. Detailed information On the choice of priors can be found in [131].

The employed PID in the later presented analysis is based on the TPC and TOF detectors which are mainly used to tag protons. The TPC PID estimation is based on simultaneous momentum, charge and the energy loss determination $\mathrm{d}E/\mathrm{d}x$ which can be described by the Bethe-Bloch formula and according to the parametrisation introduced by the ALEPH collaboration [132] is given as

$$f(\beta\gamma) = \frac{P_1}{\beta^2 P_4}\left[P_2 - \beta^{P_4} - \ln\left(P_3 + \frac{1}{(\beta\gamma)^{P_5}}\right)\right], \tag{4.5}$$

where β and γ are the relativistic velocity and Lorentz factor, respectively, and P_i are free fit parameters. The measured $\mathrm{d}E/\mathrm{d}x$ as a function of the rigidity in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ is given in the left panel of Fig. 4.8 along with the fitted curves for electrons, pions, kaons, protons and deuterons which can be clearly separated over a wide range of rigidity, with a $\mathrm{d}E/\mathrm{d}x$ resolution of 5.2 % up to momenta of $20\,\mathrm{GeV}/c$ (see also [129]).

The TOF detector provides additional PID information based on the time-of-flight t_{TOF} and the travelled distance L as determined during the track reconstruction according to

$$\beta c = \frac{t_{\mathrm{TOF}}}{L}. \tag{4.6}$$

The relativistic velocity β measured by the TOF detector is shown in the right panel of Fig. 4.8 as a function of the momentum measured in pp collisions at $\sqrt{s} = 5.02\,\mathrm{TeV}$. The start time of the event comes from the T0 detector as described in [133] but can also be estimated from a combinatorial algorithm taking all mass hypothesis into account and testing them against tracks reaching the TOF detector. With 30 tracks it reaches a resolution of 30 ps [129].

4.2.7 Trigger system

The event recoding is based on a trigger chain which consists of three hardware-based stages, the Level0 (L0), Level1 (L1) and Level2 (L2) triggers and decisions are made by the Central Trigger Processor (CTP) [129] every 25 ns. Among other detectors, inputs from the V0 and T0 are considered by the L0 trigger to provide decisions around 0.9 µs after a collision. Accepted events are forwarded to L1, making a decision approximately 6.5 µs after L0 accounting for the latency of the TRD and EMCal required for computation as well as for the signal propagation from the ZDC detectors which are located 113 m away from the IP. Positive decisions by L0 and L1 cause the data buffering and following a positive signal from L2 after 100 µs (to account for the TPC drift time) the data is forwarded in parallel to the High Level Trigger (HLT) and Data Acquisition (DAQ) system.

Physics analyses can hence be based on data according to different conditions reflected by specific trigger inputs and those employed in the presented physics analysis

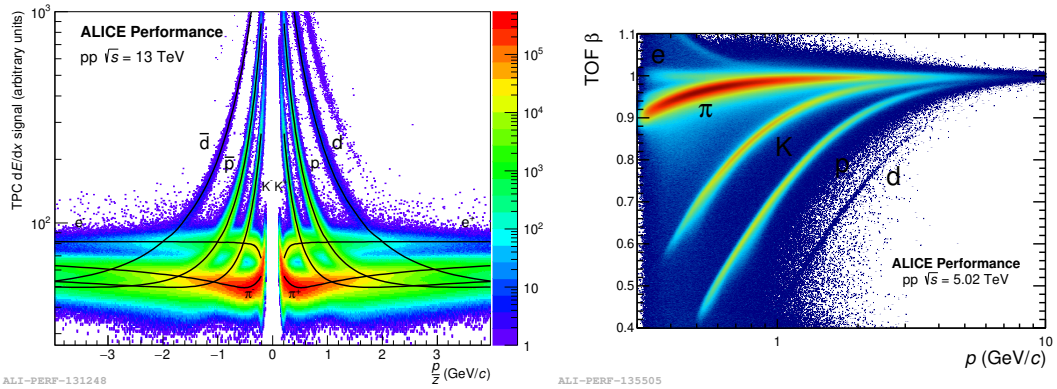


Figure 4.8: *Left:* The TPC dE/dx as a function of the rigidity in pp collisions at $\sqrt{s} = 13\,\text{TeV}$, clearly separating the curves for low- and high-p_T electrons, pions, kaons, protons and deuterons allowing for PID. *Right:* Velocity β measured in the TOF detector as a function of track momentum in pp collisions at $\sqrt{s} = 5.02\,\text{TeV}$ adding complementary PID capabilities in the low and intermediate track momentum region.

are outlined in more detail. The minimum-bias (MB) trigger requires coincident signals in the V0A and V0C scintillators compatible with the beam arrival time. These few conditions reduce potential biases introduced by any additional trigger input. On the other hand, the High-Multiplicity V0 trigger (HMVZERO) requires an increased charge deposit in the V0 detectors. In order to do so, the summed charge deposit V0M from both scintillator arrays is calculated and the threshold corresponds to $\text{V0M}/\langle\text{V0M}\rangle \approx 5$, where $\langle\text{V0M}\rangle$ is the mean charge deposit per run depending on the particular run conditions.

4.2.8 Multiplicity estimation in pp collisions

In order to conduct an analysis as a function of the event multiplicity, suitable multiplicity estimators have to be defined. There are different approaches based on information from different detectors and two of them are briefly outlined in view of the later presented analysis.

In the central barrel, it can be estimated by simply counting the number of SPD tracklets within the acceptance region. Events can then be classified based on the number of tracklets.

In the forward region the total charge deposit V0M can be employed. However, due to aging of the electronics over time, the charge amplitudes decrease and the estimated event multiplicities extracted at different data taking runs are not comparable. This is accounted for by building *percentile* values on a run-by-run basis. The V0M values are first normalised to the mean amplitude of the respective run to account for the aging. This is followed by slicing the resulting distribution into

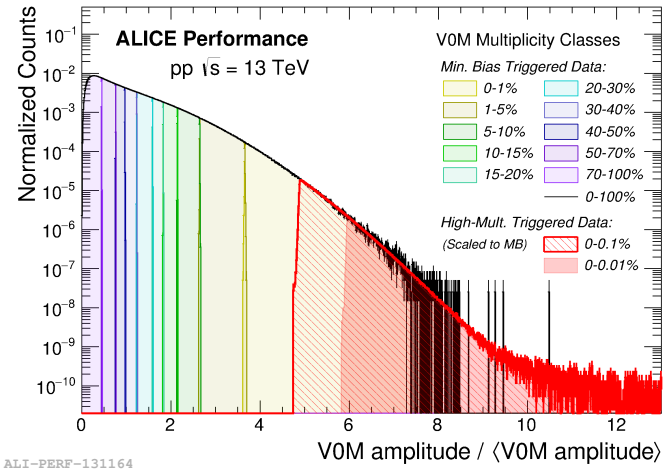


Figure 4.9: V0M amplitude normalised per run according to the corresponding average. The distribution is sliced into intervals containing a certain fraction of events used to assign a percentile value $p_{\mathrm{V0M}} \in [0\,\%, 100\,\%]$ to each event.

intervals that contain a certain fraction of events as shown in Fig. 4.9. Based on that slicing, each event is assigned a percentile value $p_{\mathrm{V0M}} \in [0\,\%, 100\,\%]$. The multiplicity is therefore highest at 0 % and decreases towards 100 %.

4.2.9 Future perspectives after the LS2 upgrades

During LS2, major detector and software upgrades have been implemented by the collaboration with a comprehensive physics programme in mind initially proposed in the Letter Of Intent (LoI) [134]. A new beam pipe with a smaller diameter has been installed [135], the readout electronics of the TRD, TOF and the muon spectrometer have been replaced and the forward trigger detectors such as the V0 and ZDC have been upgraded.

The major upgrades comprise a new ITS detector [135] with reduced material budget and significantly higher resolution, the replacement of the TPC MWPCs with Gas Electron Multiplier (GEM) detectors as well as the implementation of new readout electronics [136, 137], the installation of the new Muon Forward Tracker (MFT) detector [138], the upgrade of the readout and trigger system [139], and the implementation of the entirely new online-offline (O^2) computing system [31].

The upgrades are driven by the future physics programme outlined in the aforementioned LoI [134] and referenced technical design reports. The goal is to cover a wide range of different physics topics and to increase the precision of current measurements significantly. Among those, two major aspects of HF physics shall be targeted. The first one concerns the investigation of thermalisation of HF quarks, possible thermal production of charm quarks and the studies of the elliptic flow of

charm and beauty. This is also the focus of the following introduction, in particular emphasising new perspectives of D-meson and Λ_c^+ reconstruction and their production ratios. Closely related to the first point, the second area covers the in-medium energy loss of HF quarks accessible via the nuclear modification factor. For further information it is referred to the aforementioned LoI and technical design reports.

The next section will briefly outline the major upgrades concerning the ITS and TPC. This is followed by a selection of physics perspectives concerning the HF programme.

Key aspects of the ITS and TPC upgrade

The upgrade of the ITS is a cornerstone, in particular to achieve the goals of the future HF physics programme. To improve the impact parameter resolution, the upgraded ITS has its first layer closer to the beam pipe, namely 22 mm instead of the previous 39 mm. It now consists of seven pixel layers covering the full azimuth whose spatial parameters are summarised in Tab. 4.2. The outermost layer covers the rapidity region $|\eta| < 1.3$ which increases towards the beampipe and is $|\eta| < 2.5$ for the innermost layer. By using Monolithic Active Pixel Sensors (MAPS), the material budget is lowered overall by almost 70 % compared to the previous ITS configuration to around 2.1 % of radiation length (≈ 0.3 % per layer).

detector	acceptance η	distance (min,max) [mm]	dimension [cm^2]
layer 0	$\pm$2.5	22.4, 26.7	421
layer 1	$\pm$2.3	30.1, 34.6	562
layer 2	$\pm$2.0	37.8, 42.1	702
layer 3	$\pm$1.5	194.4, 197.7	10483
layer 4	$\pm$1.4	243.9, 247.0	13104
layer 5	$\pm$1.4	342.3, 345.4	32105
layer 6	$\pm$1.3	391.8, 394.9	36691

Table 4.2: Spatial parameters of the ITS layers as installed after LS2. Each layer covers the full azimuth in addition to the mentioned pseudorapidity coverage.

The predicted improvements of the impact parameter in the longitudinal direction and transverse plane are shown in Fig. 4.10 and are compared to the previous resolution. The upgraded ITS is expected to enhance the longitudinal resolution by a significant amount of almost one order of magnitude over the entire momentum range and even by a larger amount at highest p_T values. Furthermore, the longitudinal resolution shall become compatible with the resolution in the transverse plane which is predicted to improve significantly as well.

In addition to the significantly enhanced resolution of the ITS detector, the data yield to be collected by ALICE shall also be increased. This will be achieved first by

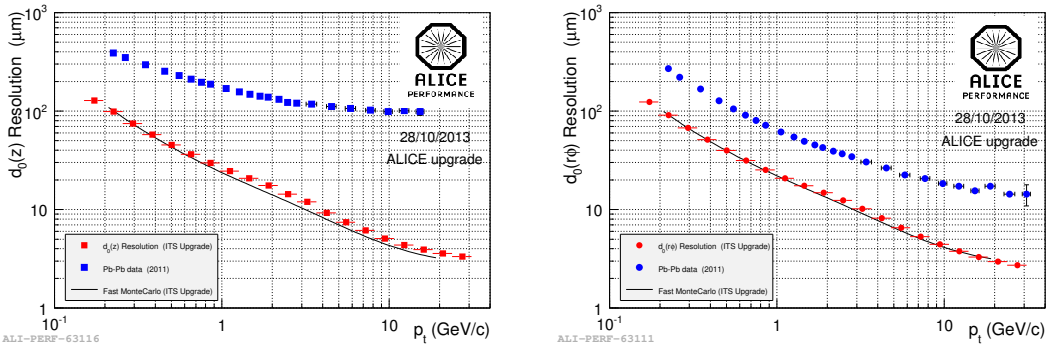


Figure 4.10: Longitudinal (left) and transverse (right) primary impact parameter resolution of the ITS before Run 3 shown in blue and its expectation after the upgrade shown in red for primary charged pions in Pb–Pb collisions as presented in [135].

the improved capabilities of the new ITS which will allow for data taking rates of up to 100 kHz in Pb–Pb and even 400 kHz in pp collisions. A higher data taking rate is also allowed by the aforementioned upgrade of the readout electronics of the TRD and TOF detector and the muon spectrometer and they will all be capable of dealing with an event rate of 50 kHz expected in future Pb–Pb collisions.

The major challenge to achieve such high data rates is raised by the TPC detector whose upgrade strategy is presented in detail in [136, 137]. The upgrade features the replacement of the MWPC with GEM detectors which suppress the accumulation of large charges in the drift volume by blocking back-drifting ions from the chambers. This is of extreme importance to ensure a stable electric field in the drift volume which is challenged with a high occupancy caused by the high interaction rate. The installation of GEMs implies the need of new readout electronics which has been installed as well. Since the ion back-flow cannot be prevented via a gating grid as previously used in the MWPCs, the readout of the GEMs also provides the necessary capability of continuous and untriggered readout where acquired data is forwarded in parallel while new signals are readout. At 50 kHz interaction rate and a maximum electron drift time of 100 μs approximately 5 events are expected to be superimposed in the drift volume at all times on average.

The tracking performance in terms of the momentum resolution is slightly lower relative to the one obtained by the previous TPC as seen by comparing the TPC-only tracking performance in the left and right panel of Fig. 4.11 shown in red. However, combined with the ITS the track momentum resolution is expected to be compatible with the previous performance. Furthermore, it was shown that also the tracking efficiency is compatible with the previous TPC. At the same time, the dE/dx resolution is expected to degrade slightly due to the high occupancy and compared to the previous TPC to 5.5 % in single pp collisions up to 7.5 % in central Pb–Pb collisions.

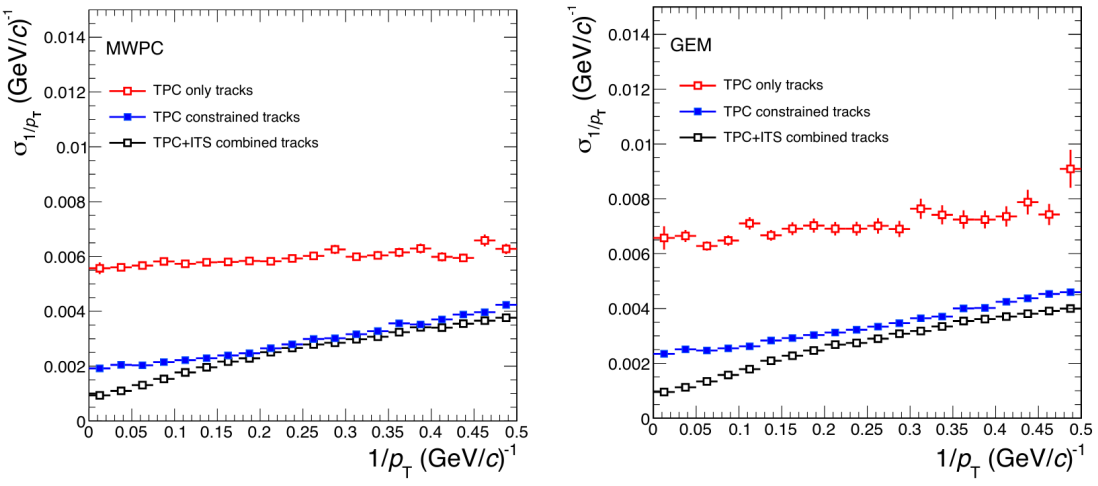


Figure 4.11: p_T resolution for track reconstruction showing the TPC stand-alone performance in red. The performance for fully contained tracks is shown in blue and the combined performance of the ITS +TPC is shown in black. The performance of the previous TPC as installed during Run 2 is shown on the *left* while the expected performance of the newly installed TPC is given on the *right*.

Selected topics of the HF physics programme perspectives

Enhanced impact parameter and vertex reconstruction capabilities will allow for the reconstruction of D-mesons to very low p_T, the D^0 even towards zero p_T in Pb–Pb collisions. This will allow to add low-p_T data points to the previous measurement of the D-meson elliptic flow obtained by ALICE in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02\,\text{GeV}$ [140]. The current results averaged over different D-mesons are shown in Fig. 4.12 in the two centrality classes 10 % to 30 % and 30 % to 50 %. On the other hand, the B-meson v_2 is as of now not accessible due to low statistics but shall become accessible in Run 3 and Run 4. It is expected to be significantly smaller than the D-meson v_2 and in order to achieve a significant resolution and access the momentum range below $2\,\text{GeV}/c$, a maximum absolute uncertainty of 0.02 and 0.05 on D- and B-meson measurements is required. Such a resolution is achieved per meson in the case of D^0 and D_s^+ as shown in the left panel of Fig. 4.13.

At the same time, access to the low-p_T regime is necessary to study the potential thermal production of $c\bar{c}$ in Pb–Pb collisions. Based on next-to-leading order QCD calculations [143] it is predicted that this mechanism increases the total charm yield by 30 % to 80 % depending on the initial medium temperature in the range of 700 MeV to 750 MeV and assuming a charm quark mass of 1.3 GeV. The right panel of Fig. 4.13 shows the $c\bar{c}$ production in Pb–Pb collisions at LHC energies as a function of the proper time of the system evolution for 3 different temperatures. For lower temperatures of $\lesssim 630\,\text{MeV}$ or larger charm-quark masses of $\gtrsim 1.5\,\text{GeV}$

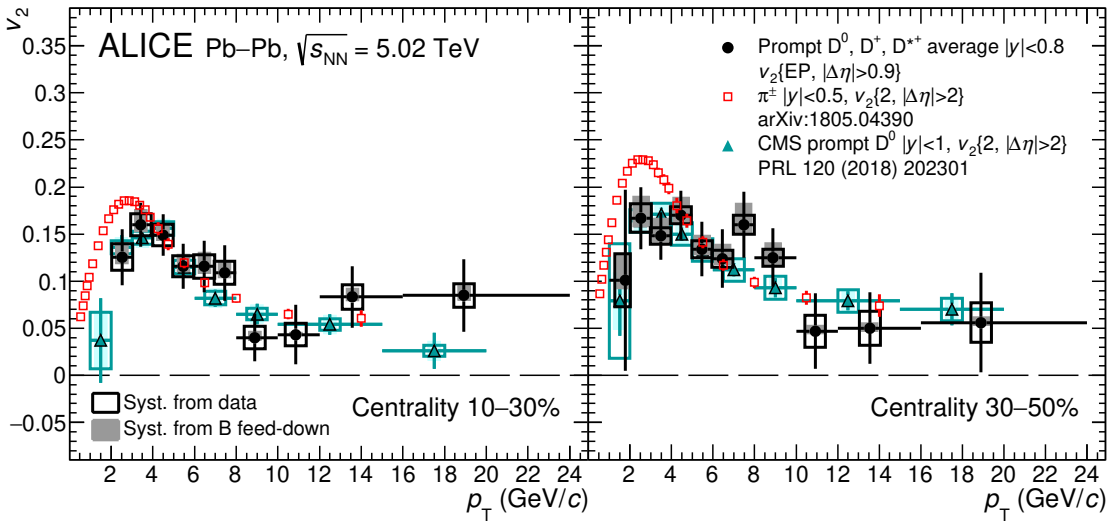


Figure 4.12: Elliptic flow averaged over non-strange D-mesons measured in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02\,\text{GeV}$ by ALICE [140]. Also shown are the corresponding results from CMS [141] and the v_2 of $\pi^{\pm}$ [142].

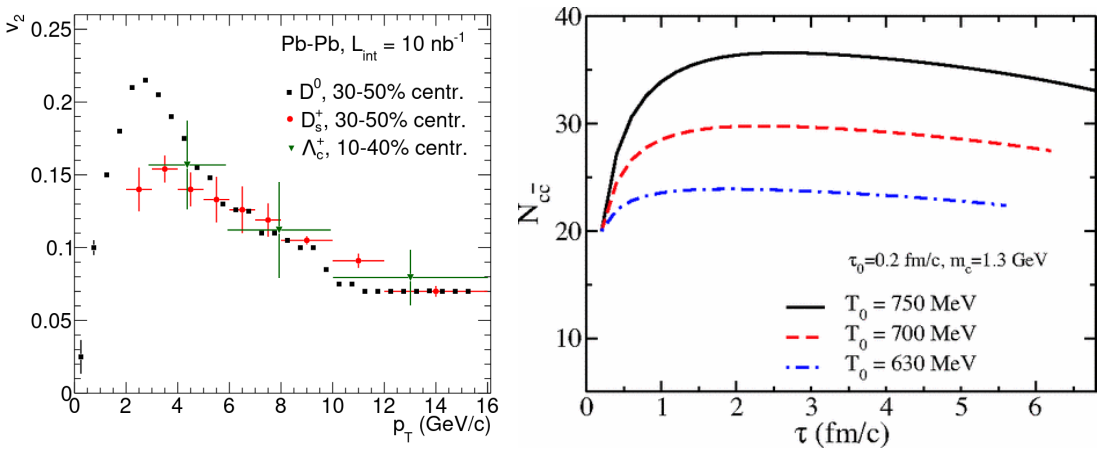


Figure 4.13: *Left:* Expected measurements of elliptic flow with estimated statistical uncertainties of D^0, D_s^+ in the 30 % to 50 % and Λ_c^+ in the 10 % to 40 % centrality region in Pb–Pb collisions. *Right:* Predicted $c\bar{c}$ production in Pb–Pb collisions at LHC energies as a function of proper time for three different initial medium temperatures and a charm quark mass of 1.3 GeV [143].

the predicted enhancement vanishes. Also in this case the momentum resolution down to very low p_T and enhanced statistics are necessary to enable for quantitative measurements.

Detector upgrades will also have an impact on the sensitivity to test hadronisation mechanisms of the D_s^+ meson. As already discussed in Sec. 3.4 the measurement

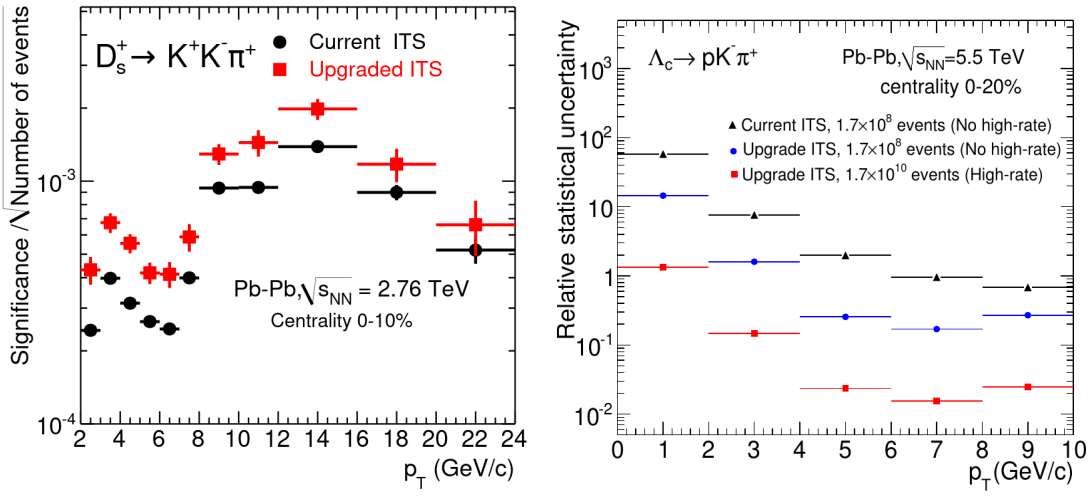


Figure 4.14: *Left:* Expected significance of the D_s^+ reconstruction comparing the performance with the previous and new ITS detector. *Right:* Statistical precision of the Λ_c^+ as a function of p_T of the previous setup of the ITS (black) compared to new ITS in two run scenarios running at no high-rate (blue) and at high rate (red). The predictions are made for central Pb–Pb at $\sqrt{s_{NN}} = 5.5$ TeV [134].

of its nuclear modification factor compared to non-strange D-mesons indicates an enhancement which could be due to enhanced strangeness production in the medium as well as enhanced in-medium hadronisation of the charm quark. In addition, there is the potentially connected indication of an increasing D_s^+/D^0 ratio in Pb–Pb collisions compared to pp. The left panel of Fig. 4.14 shows an expected enhancement on the significance of a factor of ≈ 2 in the D_s^+ reconstruction. A higher sensitivity will allow firmer conclusions on whether or not D_s^+ might additionally be produced inside the medium, for instance by hadronisation via coalescence.

At the same time, the detector upgrades will further enhance the sensitivity on the Λ_c^+ baryon in Pb–Pb collisions, in particular towards low p_T. In the right panel of Fig. 4.14 the expected statistical precision of the Λ_c^+ baryon in the decay channel $\Lambda_c^+ \to pK\pi$ in central Pb–Pb collisions is shown for the previous ITS setup in black and for two scenarios involving the new ITS. In case of running at expected high rates of 50 kHz the precision is expected to enhance by more than one order of magnitude.

Higher precision Λ_c^+/D^0 ratio measurements as a function of p_T will complement the measurement presented in [144] where as of now there is only one data point available in the p_T interval 6 GeV/c to 12 GeV/c in the large centrality interval 0 % to 80 %. A more differential measurement in terms of momentum and centrality will be of great interest to further test the hadronisation mechanisms in large collision systems but then also in comparison with small collision systems such as pp. The additional decay channel $\Lambda_c^+ \to pK_S^0$ will add to the precision.

5 Λ_c^+ production in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$

The presented physics analysis investigates the prompt Λ_c^+ production at midrapidity $|y| < 0.5$ as a function of the candidate p_T and in intervals of forward multiplicity in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ exploiting the decay chain $\Lambda_c^+ \to pK_S^0 \to p\pi^+\pi^-$. The final results will be expressed as a function of the average primary charged-particle density $\langle dN_{ch}/d\eta \rangle$ in order to provide a suitable observable for comparisons to other experiments and theory predictions.

This work is part of an analysis paper currently being in preparation which in addition comprises the analogous measurements of D^0 and D_s^+ production as well as the Λ_c^+ production in the decay channel $\Lambda_c^+ \to pK^-\pi^+$. All measurements are in addition conducted using the number of SPD tracklets as the multiplicity estimator. For both multiplicity estimators, the production ratios Λ_c^+/D^0 and D_s^+/D^0 are computed. The results of this work are therefore presented and discussed covering all different measurements and computed production ratios.

The ITS and TPC detectors were used for vertex reconstruction and track reconstruction while PID information were obtained from the TPC and TOF detectors. Multiplicities were estimated employing the V0 detector and deriving the percentile multiplicity p_{V0M}.

In the following Sec. 5.1 the properties of the Λ_c^+ baryon are briefly reviewed. The data and simulation samples used for the presented analysis are summarised and explained in Sec. 5.2 followed by the event selection in Sec. 5.2.2. The Λ_c^+ reconstruction and candidate selection is presented in Secs. 5.2.3 and 5.2.4. Sec. 5.3 contains the detailed outline of the analysis method and the computation of the raw Λ_c^+ candidate yield as well as required corrections necessary for the yield-per-event calculation. Systematic uncertainties on the central extracted values are discussed in Sec. 5.4. In Sec. 5.5 the final results are presented and discussed. The results are examined to highlight potential deviations from universal hadronisation mechanisms of charmed *baryons* and *meson* production via their ratios. A comparison to the production ratio of Λ/K_S^0 measured in the LF sector [145] is provided as well. In addition, the comparison of the ratios with a canonical approach of a statistical hadronisation model with augmented baryon production [118, 146] as well as with a fragmentation model implementing colour reconnection beyond leading colour approximation [106] are discussed.

5.1 The Λ_c^+ baryon

The Λ_c^+ is the lightest open-charm baryon with a mass of $2.286\,46(14)\,\text{GeV}/c^2$ as measured by the BABAR experiment [147], has a mean lifetime of $\tau = 202.4(31) \times 10^{-15}\,\text{s}$ [13], which, in the laboratory frame is given as $\gamma_{\text{rel}}\tau$ where γ_{rel} is the relativistic Lorentz factor. The proper decay length is given as $c\tau$ which for the Λ_c^+ baryon results in a value of $\approx 60\,\mu\text{m}$. For the current datasets this is below the resolution which was provided by the detector. It has various weak decay modes with branching ratios (BRs) of the order of 1 %, one of which is $\Lambda_c^+ \to \text{p}K_S^0$ channel with BR = 1.59(8) % [13]. Its decay topology is sketched in Fig. 5.1. The proton is stable while the K_S^0 in turn has a mean lifetime of $89.54(400) \times 10^{-12}\,\text{s}$ and predominantly decays into a $\pi^+\pi^-$ pair with $\text{BR}(K_S^0 \to \pi^+\pi^-) = 69.20(5)\,\%$ [13]. This V^0 decay causes a displaced vertex w.r.t. the primary vertex. Tracks are reconstructed in the ITS and TPC hence those originating from the decay of interest must be compatible with the proton hypothesis and two additional oppositely charged particles.

Topological, kinematic and PID properties are used for theΛ_c^+ reconstruction and selection. Note also that in the following, the term *secondary vertex* is not used to refer to the Λ_c^+ decay vertex (which was not reconstructed due to the detector resolution) but to the reconstructed K_S^0 decay vertex and it is used interchangeably.

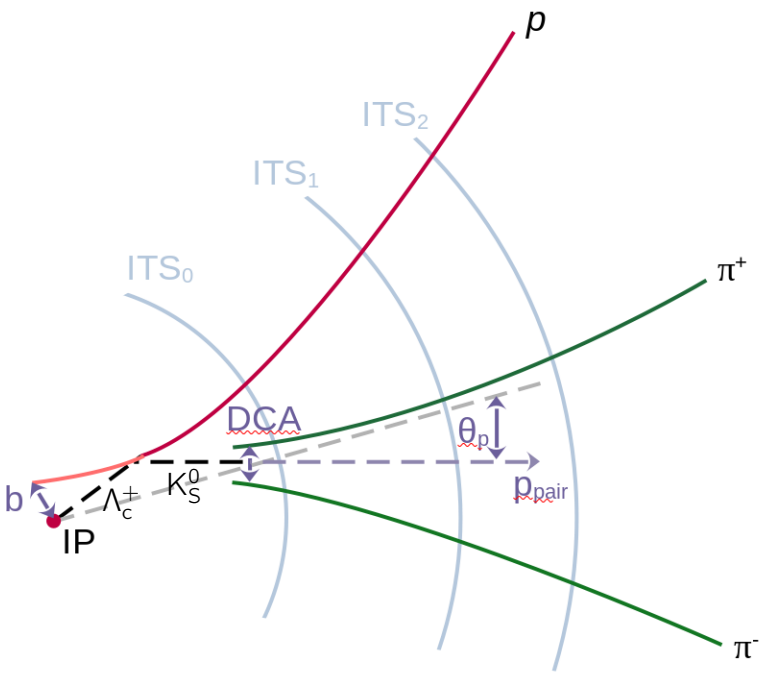


Figure 5.1: Sketch of the $\Lambda_c^+ \to \text{p}K_S^0 \to \text{p}\pi^+\pi^-$ decay topology inscribed into the ITS detector. The spatial proportions have been adjusted for better visibility and do not correspond to the real scales. Adapted from [129].

The impact parameter b is defined as the minimum distance of a reconstructed track to the primary vertex, while the distance of closest approach (DCA) denotes the closest approach of two tracks. If not stated otherwise, it refers to the DCA of the reconstructed $\pi^\pm$ tracks. The pointing angle θ_p is the angle between the p_T-direction of the reconstructed V^0 candidate and the direction of the connection between its decay and primary vertex in the transverse plane as indicated in the sketch. Both the proton and the K_S^0 masses are of the same order of magnitude and they are both expected to be emitted close to the direction of the parent Λ_c^+. θ_p can hence be expected to be small for correctly reconstructed and selected $\pi^\pm$ decay tracks.

The invariant mass of the V^0 candidate m_{V^0} can be reconstructed from the p_T and pseudorapidity which can then also be combined with the proton to yield the invariant mass of the Λ_c^+ candidate $m_{\Lambda_c^+}$. d_{V^0} denotes the decay distance between the primary and secondary vertices and $c\tau_{V^0}$ is the proper lifetime of the V^0 candidate before its decay.

The Armenteros-Podolanski value v_{ap} shall be defined as

$$v_{ap} = \frac{q_T}{|\alpha_{asym}|} \quad \text{with} \quad \alpha_{asym} = \frac{p_L^+ - p_L^-}{p_L^+ + p_L^-} \quad \text{and} \quad q_T = \frac{|\vec{p}^+ \times \vec{p}_{V^0}|}{|\vec{p}_{V^0}|}. \tag{5.1}$$

q_T is hence the transverse momentum of the positively charged V^0 decay track w.r.t. the momentum axis of its parent particle. α_{asym} on the other hand yields the longitudinal momentum asymmetry of the decay tracks. In case these originate from particles of the same mass, this observable is expected to be distributed symmetrically around $\alpha_{asym} = 0$ as it can be seen for the K_S^0 in Fig. 5.2. On the other hand, the proton from a $\Lambda \to \pi^- p$ is expected to acquire on average the higher momentum which shifts α_{asym} away from $\alpha_{asym} = 0$. Photon conversion* yields a symmetric distribution around $\alpha_{asym} = 0$ as well making it indistinguishable from $K_S^0 \to p\pi^+\pi^-$ considering only α_{asym}. The ratio as computed by v_{ap}, however, tends to high values for K_S^0 and by choosing an appropriate lower bound of the ratio around ≈ 1.8, the desired signal can be well discriminated against background.

*Photon conversion is not a V^0 decay in the strict sense as those are defined as weakly decaying electrically neutral mesons. However, the qualitative signature of a positively and negatively decay track is the same.

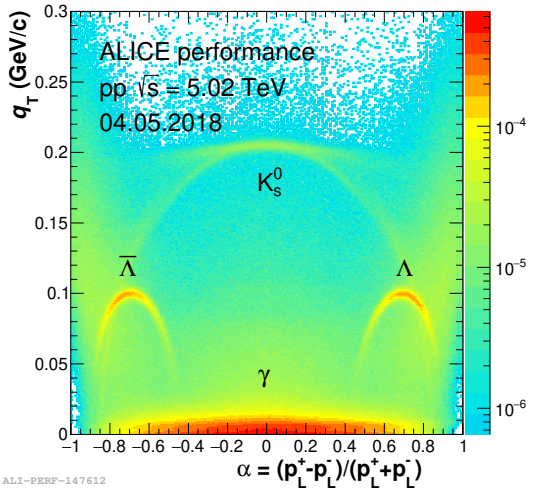


Figure 5.2: Armenteros-Podolanski plot showing the 2-dimensional distribution of q_T against α_{asym}.

5.2 Data samples, event and candidate selection

5.2.1 Samples

Data samples

The analysis is based on data obtained in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ recorded during LHC Run 2 in the years 2016, 2017 and 2018 selected with the MB or HMVZERO triggers as introduced in Sec. 4.2.7. The latter enhances the statistics of events with highest multiplicities.

Monte Carlo samples

Monte Carlo (MC) events were generated using PYTHIA 6 [148] with the Perugia-2011 tune [149] and the production of a $c\bar{c}$ or $b\bar{b}$ quark pair at midrapidity $|y_{HF}| < 1.0$ was requested. Primary particles were propagated through the simulated detector geometry with GEANT3 [34]. These samples were used to derive the acceptance and efficiency corrections of prompt and feed-down Λ_c^+ signal candidates as explained in Sec. 5.3.2.

MC production were employed each anchored to a different year of data taking. Furthermore, there is one MC sample per trigger used in data. For the HMVZERO trigger, the simulated average multiplicity is enhanced by a factor of ~ 4.

Boosted Decision Trees are employed for the candidate selection in this analysis. To fit them reliably, a larger statistics of $\Lambda_c^+ \to pK_S^0$ is required which is provided by three additional MC samples, one for each year of data taking. Instead of HF quark pairs, the production of Λ_c^+ baryons with the decay channel $\Lambda_c^+ \to pK_S^0$ was enforced.

Hence, there are in total nine different MC sample, six of which are used for the efficiency calculation (one per trigger and data-taking year) and three for the fitting of the Boosted Decision Trees.

5.2.2 Offline event selection

Events were selected to satisfy the MB or HMVZERO trigger conditions. At least one SPD tracklet was required within $|\eta| < 1.0$ to reduce the contamination from diffractive events. It will be referred to this selection criterion as INEL > 0 and it corresponds to $\sim 75\,\%$ of the inelastic cross section at $\sqrt{s} = 13\,\text{TeV}$ [150].

Events were rejected in which no primary vertex could be reconstructed. The contribution from pile-up was reduced by requiring the longitudinal distance between the primary vertex reconstructed from ITS-TPC matched tracks and the ITS-only reconstructed primary vertex to be less than 0.2 cm. It was further suppressed by rejecting events where another vertex could be reconstructed with at least 5 tracks or SPD tracklets associated to it or if the χ^2 of a set of tracks fitted to another potential vertex is less than 5. Also, in case the weighted difference

$$d_{i,\text{pu}} = (\vec{v}_i - \vec{v}_{\text{prim}})^{\mathsf{T}} \boldsymbol{C}_{\text{prim}}^{-1} (\vec{v}_i - \vec{v}_{\text{prim}}) \tag{5.2}$$

between the primary vertex position $\vec{v}_{\text{prim}}$ with covariance matrix $\boldsymbol{C}_{\text{prim}}$ and any potential pile-up vertex i with position $\vec{v}_i$ exceeds 15, this event was excluded. Finally, the longitudinal primary vertex position was required to lie within $|z_{\text{vtx}}| < 10\,\text{cm}$ around the nominal collision point to ensure a compatible acceptance among events.

The number of selected events N_{sel} was corrected for those where no vertex could be reconstructed, N_{noVtx}, or where the vertex was outside of the longitudinal acceptance region of $|z_{\text{vtx}}| < 10\,\text{cm}$, N_{vtxOut}. The corrected number of events N_{ev} is then given as

$$N_{\text{ev}} = (N_{\text{sel}} + N_{\text{noVtx}}) - \frac{N_{\text{noVtx}} N_{\text{vtxOut}}}{N_{\text{sel}} + N_{\text{vtxOut}}}. \tag{5.3}$$

Those events were assigned into 4 intervals of V0M percentile p_{V0M} as summarised in Tab. 5.1. The trigger used for a specific interval is indicated as well. In the following, the motivation for choosing the percentile based on the raw V0M amplitude rather than the amplitude itself is given.

p_{V0M} interval [%]	selected events [10^9]	trigger
$[0, 100]$	1.690	MB
$[30, 100]$	1.189	MB
$[0.1, 30]$	0.499	MB
$[0, 0.1]$	0.471	HMVZERO

Table 5.1: Selected number of events per multiplicity interval.

Due to the high radiation dose, the scintillators of the V0 detectors degraded with time. To study this and at the same time investigate the impact of the longitudinal vertex position z_{vtx} on the measured amplitude, the average amplitude $\langle \text{V0M} \rangle$ is shown in Fig. 5.3 per period for the data-taking years 2016 to 2018 from top-left to bottom-right. Indeed, $\langle \text{V0M} \rangle$ decreased from period to period due to the degrading performance of the scintillator arrays. The dependency with the vertex position however is mild and of $\approx 1\,\%$.

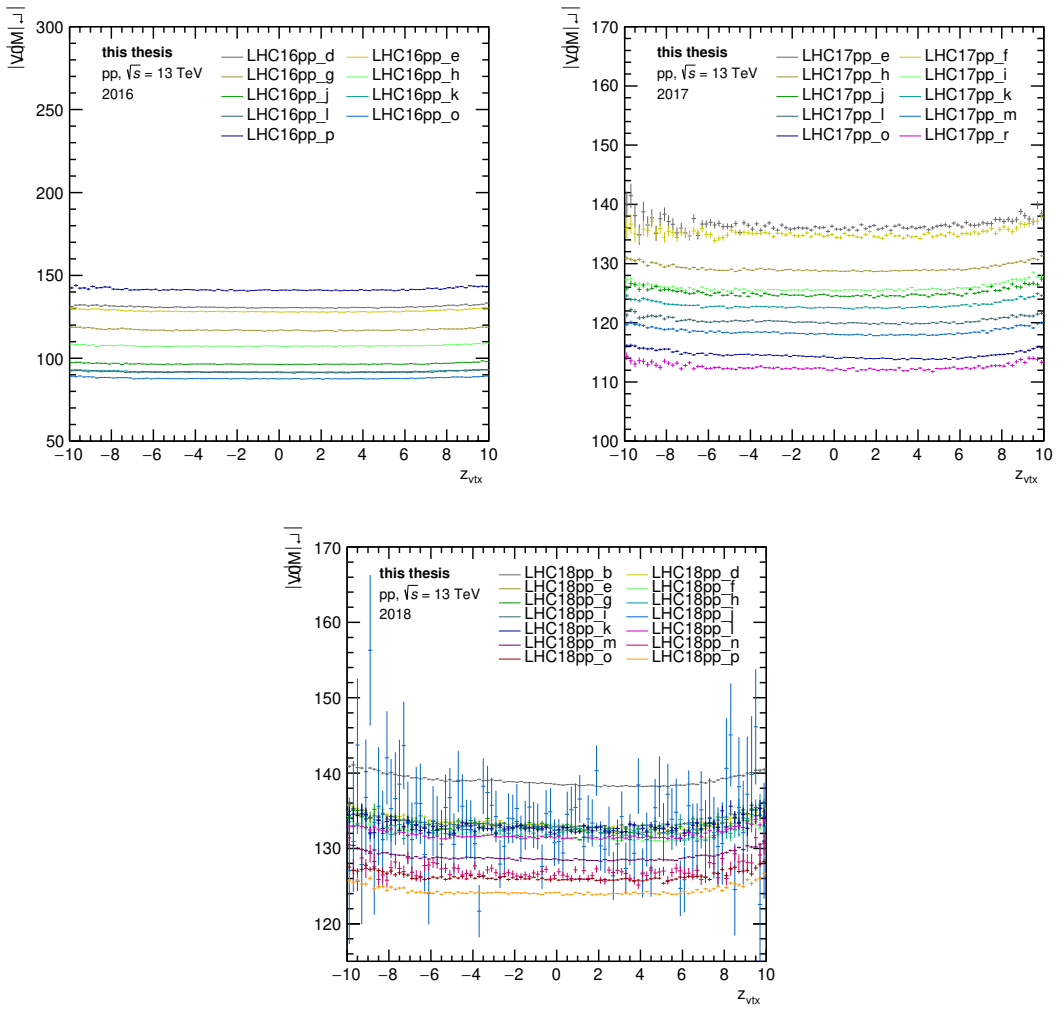


Figure 5.3: $\langle \text{V0M} \rangle$ distributions as a function of the longitudinal vertex position z_{vtx} per period and data-taking year 2016 to 2018 *(from top-left to bottom)*.

Using the definition of the percentile p_{V0M} as introduced in Sec. 4.2.8, first each amplitude obtained in an event was normalised to the average of the corresponding period. The percentile values were then computed per period to obtain a suitable multiplicity observable across periods.

First, it was investigated whether the percentile distributions are flat according

to the expectation from its definition. The distributions for MB triggered events are shown in Fig. 5.4 per period in each data-taking year and the respective plots for HMVZERO triggered events are given in Fig. 5.5. For MB triggered events the distributions are flat within 5 %. In the case of HMVZERO triggered data the distributions reach a plateau for $p_{\mathrm{V0M}} < 0.1\,\%$ and hence the HMVZERO trigger was assumed to be fully efficient in that percentile regime. This observation is indeed the motivation to choose the highest multiplicity interval as specified above in Tab. 5.1. There were periods where the HMVZERO trigger input was not available, hence those periods were not taken into account and are not shown in the figure.

Certain periods such as period *j* of the year 2018 show larger statistical errors which arise due to a significantly lower number of events in that period compared to the other periods. However, they are still seen to reach the plateau for HMVZERO triggered events and therefore, also this period was included in the analysis.

To investigate whether the observed degeneration of the V0 scintillator arrays was compensated by building the percentile, Fig. 5.6 shows the average p_{V0M} as a function of the longitudinal vertex position per period of the data-taking years 2016 to 2018 from top-left to bottom-right. As seen from the plots, it has been compensated and hence, the percentile was considered to be a suitable parameter to estimate the forward multiplicity given a compatible parameter across data-taking periods. A mild dependence on the vertex position of $\lesssim 1\,\%$ can be seen caused by the different distances and acceptances of the V0A and V0C scintillator arrays w.r.t. the interaction point. This very mild dependency was assumed to be negligible in the further analysis. Furthermore, the usage of the V0M percentile disentangles the multiplicity estimation done in the forward region from the candidate reconstruction and selection in the barrel region of the detector.

5.2.3 Λ_{c}^{+} candidate reconstruction and pre-selection

Note first, that for the candidate reconstruction and selection it is not differentiated between the Λ_{c}^{+} and its antiparticle Λ_{c}^{-}. Both are implied in the following when mentioning the Λ_{c}^{+}.

Tracks were reconstructed in the pseudorapidity region $|\eta| < 0.8$. Note that the following discussion implies always the reconstruction and selection of the charged-conjugate $\Lambda_{\mathrm{c}}^{-} \rightarrow \bar{\mathrm{p}}\mathrm{K}_{\mathrm{S}}^{0} \rightarrow \bar{\mathrm{p}}\pi^{+}\pi^{-}$. In that case an antiproton instead of a proton results from the Λ_{c}^{-} decay. It will be accounted for that in the yield-per-event calculation as discussed in Sec. 5.3.

Primary tracks and V^{0} decay tracks were required to have minimum transverse momenta of $0.3\,\mathrm{GeV}/c$ and $0.1\,\mathrm{GeV}/c$, respectively. V^{0} decay tracks did not have to pass the combined ITS and TPC refit but only had to pass the latter due to its secondary vertex decay topology which may lie outside of the SPD layers. At least 70 crossed rows in the TPC (see Sec. 4.2.2) had to be associated to each track and the ratio of findable TPC clusters over crossed rows was required to be at least 0.8.

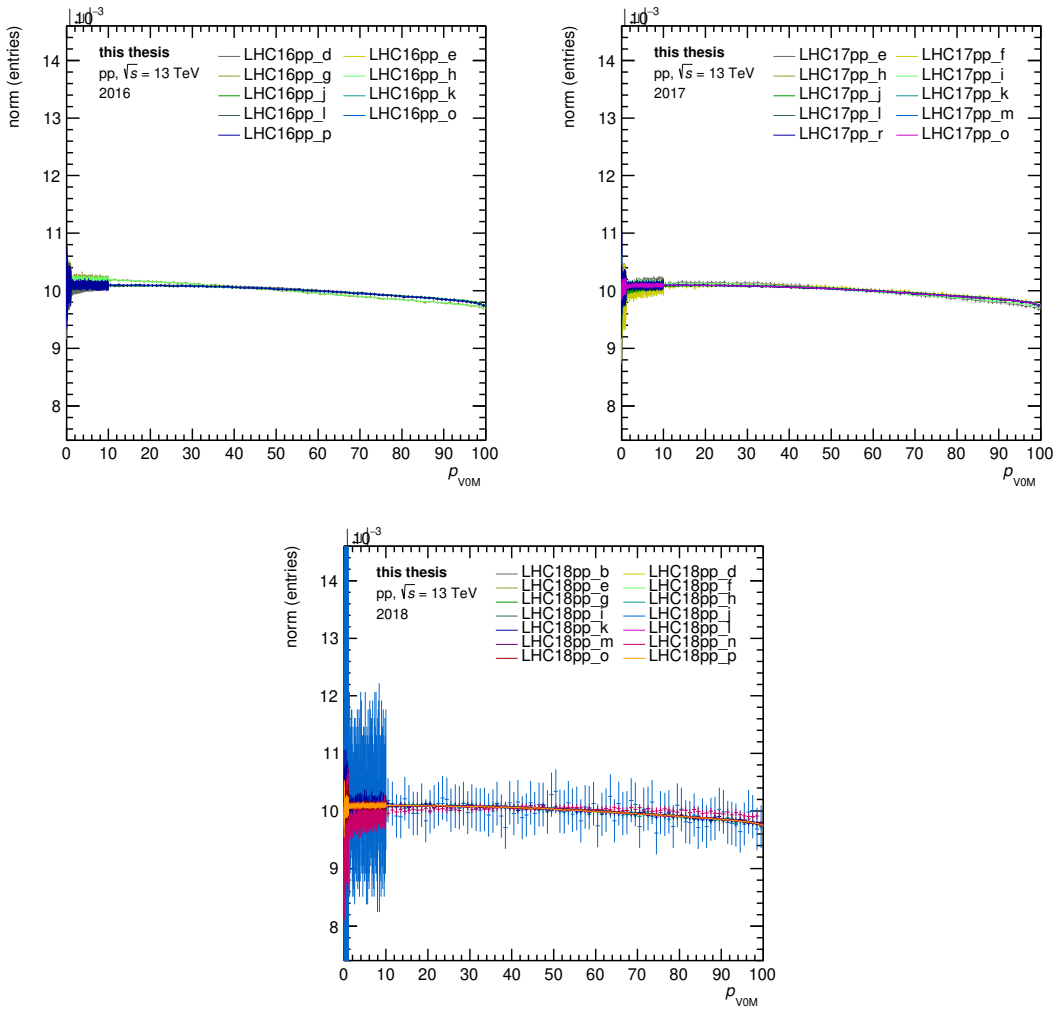


Figure 5.4: p_{V0M} distributions per data-taking year and period *(from top-left to bottom)*. Distributions are shown for MB-triggered events. They are flat over the entire range of p_{V0M} within $< 5\,\%$ and hence the trigger is assumed to be fully efficient over the entire range

A candidate proton track had to leave one hit in either of the SPD layers as it is expected from the early Λ_c^+ decay.

Additional loose selection cuts were applied to the reconstructed V^0 and proton candidates which are summarised in Tab. 5.2 along with their explanation. The selection cuts were motivated based on the Λ_c^+ decay topology as discussed in Sec. 5.1.

After the reconstruction and pre-selection candidates were grouped into 6 intervals depending on their transverse momentum as summarised in Tab. 5.3.

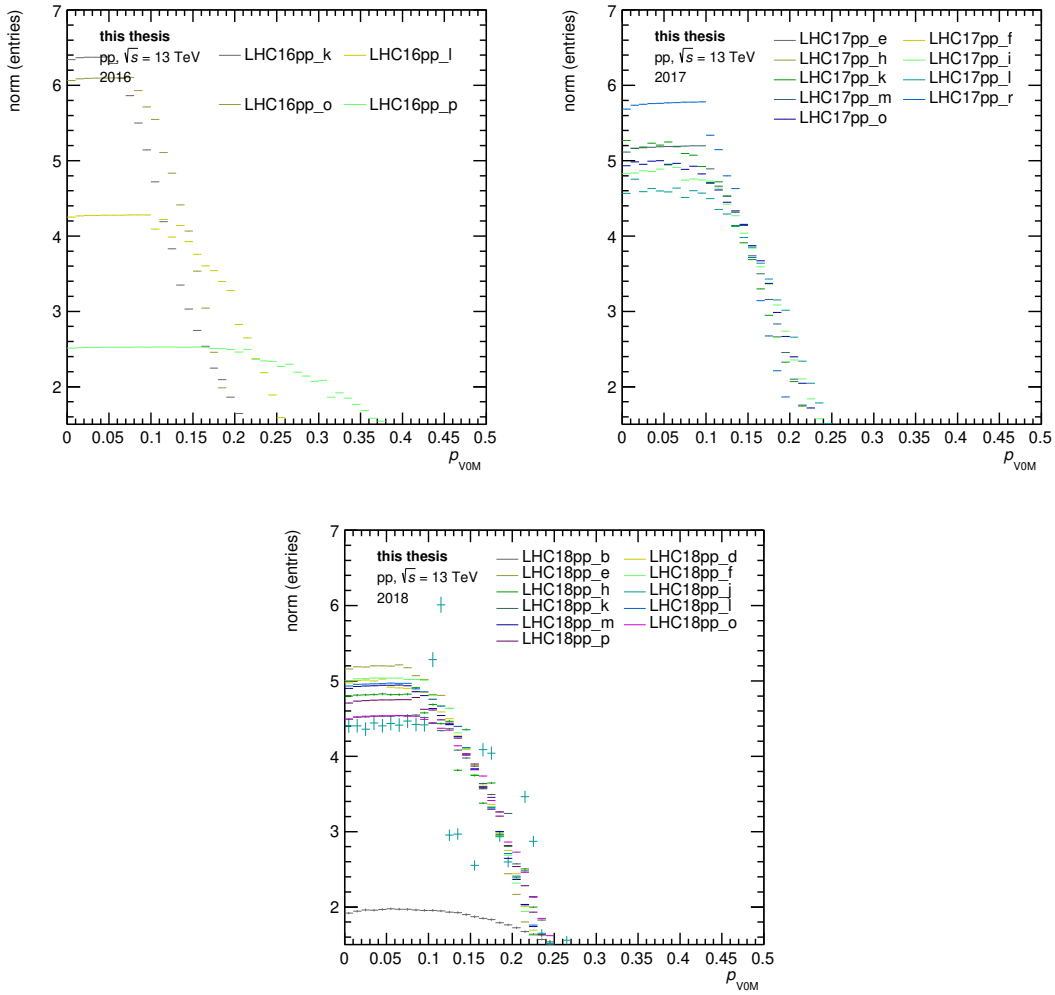


Figure 5.5: p_{V0M} distributions per data-taking year and period *(from top-left to bottom)*. Distributions are shown for HMVZERO-triggered events. The HMVZERO trigger is seen to be fully efficient for $p_{\mathrm{V0M}} < 0.1\,\%$. Periods with inactive HMVZERO trigger and those which do not reach a plateau for $p_{\mathrm{V0M}} < 0.1\,\%$ are not shown and were excluded in the analysis of the highest multiplicity interval.

5.2.4 Λ_{c}^{+} candidate selection using Boosted Decision Trees

The selection criteria applied so far provide a sample which is still dominated by background. In those cases, three charged tracks were associated which did not originate from a Λ_{c}^{+} decay. For instance, the set of candidates includes cases where three light hadron tracks from the primary interaction were associated or some tracks from a Λ_{c}^{+} decay were mixed with those not originating from that decay.

To further reduce the background contamination, the final Λ_{c}^{+} candidate selection

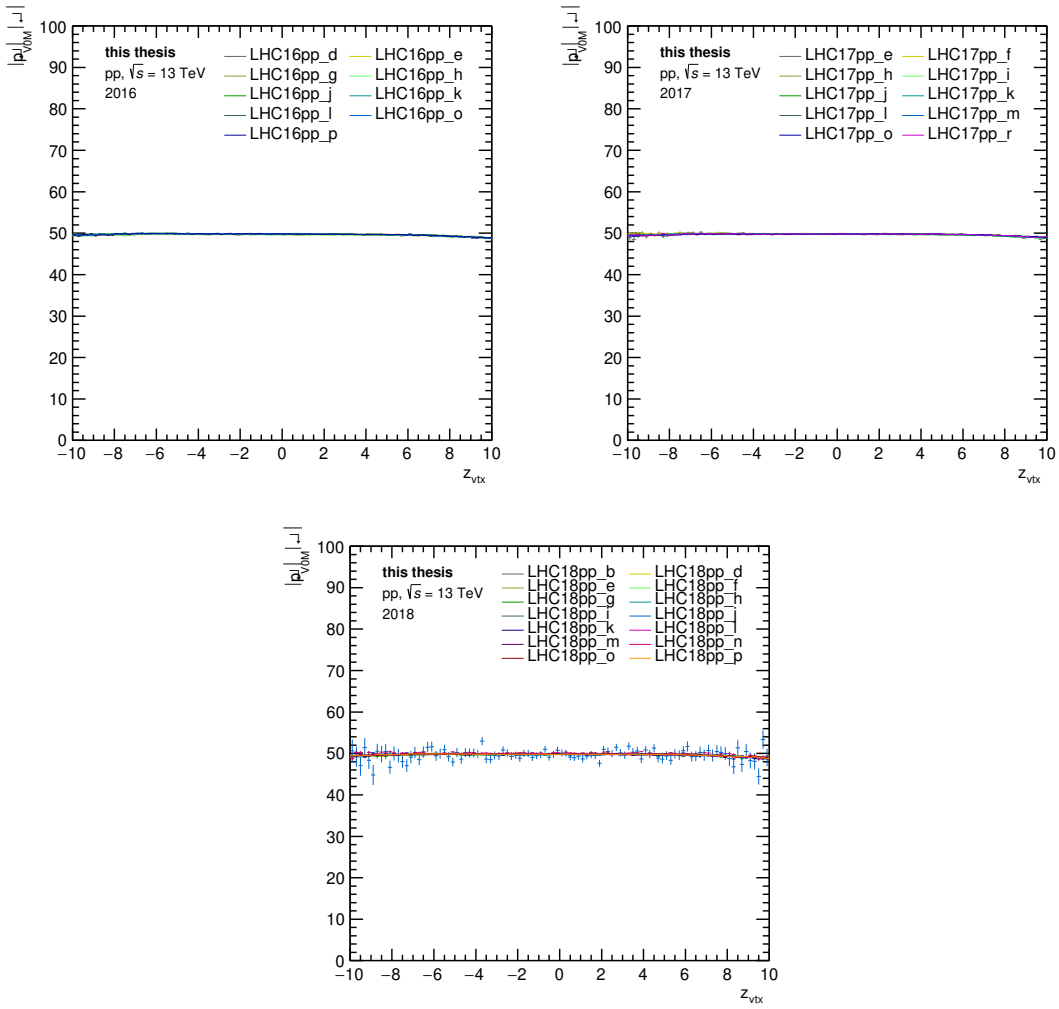


Figure 5.6: Average p_{V0M} as a function of the longitudinal vertex position z_{vtx} per data-taking year and period *(from top-left to bottom)* for MB-triggered events. The degeneracy of the V0 detector with increasing period has been compensated.

was based on a Boosted Decision Tree (BDT) algorithms using the XGBoost package [151] where one model was fitted per Λ_c^+ candidate p_T interval. Passing a set of topological, kinematical and PID properties as *features* per pre-selected candidate i, those are combined to assign a signal probability $p_{sig,i} \in [0, 1]$ reflecting how likely this candidate is considered to originate from the $\Lambda_c^+ \to pK_S^0 \to p\pi^+\pi^-$ decay chain. A working point p_{WP} was chosen such that each candidate i with $p_{sig,i} > p_{WP}$ was treated as signal and background otherwise by which the procedure devises a binary decision. For each of the specified Λ_c^+ p_T interval one BDT was constructed.

parameter	comment	requirement
$\Delta m_{\Lambda_c^+}$	difference of reconstructed Λ_c^+ candidate mass to PDG value	$< 0.2\,\mathrm{GeV}/c^2$
$\Delta m_{K_S^0}$	difference of reconstructed K_S^0 candidate mass to PDG value	$< 0.03\,\mathrm{GeV}/c^2$
DCA(V^0)	DCA of V^0 decay tracks	$< 10\,\mathrm{m}$
n_σ(DCA(V^0))	number of standard deviations of reconstructed positions at the DCA(V^0) between V^0 decay tracks	< 0.8
b_p	impact parameter of proton track	$< 3\,\mathrm{cm}$
b_{V^0}	impact parameter of reconstructed V^0	$< 1.5\,\mathrm{cm}$
$\cos\theta_p$	cosine of V^0 pointing angle	> 0.997
v_{ap}	Armenteros-Podolanski value as defined in (5.1)	$> 1.8\,\mathrm{GeV}/c$
$\lvert n_{\sigma_{TPC}}(p)\rvert$ $\vee$ $\lvert n_{\sigma_{TOF}}(p)\rvert$	minimum absolute value of deviation from the proton hypothesis when applied to the proton candidate track candidate track	< 4

Table 5.2: Pre-selection criteria for the Λ_c^+ candidates in the $\Lambda_c^+ \to pK_S^0 \to p\pi^+\pi^-$ decay channel.

p_T intervals [GeV/c]
[1, 2)
[2, 4)
[4, 6)
[6, 8)
[8, 12)
[12, 24)

Table 5.3: p_T intervals the pre-selected Λ_c^+ candidates were assigned to.

Binary classification with Boosted Decision Trees

A binary decision for sample i with a BDT is obtained by applying it to the sample's feature vector $\boldsymbol{x}_i$ to yield the signal probability $p_{\mathrm{sig},i}$. Before doing so, the BDTs had to be fitted to a training set of samples with known target values $y_i \in \{0, 1\}$ where 0 denote and 1 denote background and signal Λ_c^+ candidates. During the fitting procedure, the BDT parameters are fitted to accurately predict the targets of the training set. This is done by choosing a *loss function*

$$L(\{\hat{y}\}, \{y\}) = L(\{\hat{y}\}, \{y\}). \tag{5.4}$$

where

$$\hat{y}_i = f(\boldsymbol{x}_i; \boldsymbol{\omega}), \tag{5.5}$$

which can be seen as the function f associating a value $\hat{y}_i$ to each feature vector $\boldsymbol{x}_i$ with function parameters $\boldsymbol{\omega}$ to be fitted. A BDT is the sum of single regression trees t

$$f(\boldsymbol{x}) = \sum_t f_t(\boldsymbol{x}; \boldsymbol{\omega}_t) \tag{5.6}$$

Fig. 5.7 shows an example sketch of such a single tree. At each node a new branching is introduced by cutting on a given feature such that each sample finally ends up in one of the final leaves. The score assigned to a sample is the weight w_k of that leaf k.

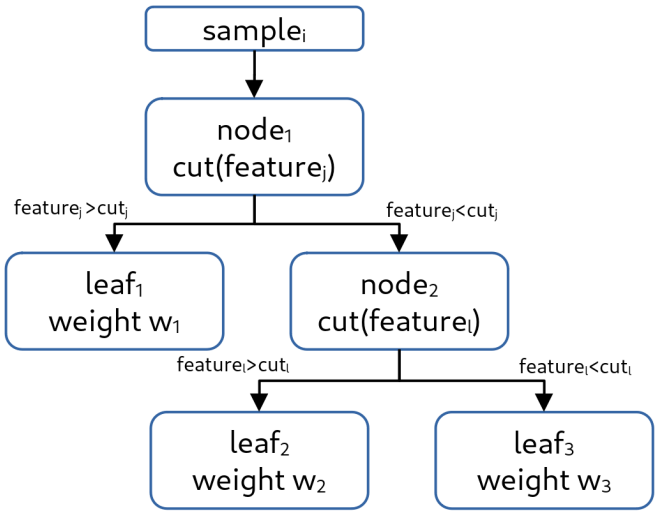


Figure 5.7: Example sketch of a possible structure of a single tree in the BDT ensemble.

The optimal parameters of the t'th tree is built by minimising the loss function

$$L_t(\{\boldsymbol{y}\}, \{\hat{\boldsymbol{y}}\}, f_t) = \sum_i \ell(\boldsymbol{y}_i, \hat{\boldsymbol{y}}_i) + \Omega(f_t), \tag{5.7}$$

where $\ell(\boldsymbol{y}_i, \hat{\boldsymbol{y}}_i)$ computes the partial loss per sample i and

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_k \omega_k^2 \tag{5.8}$$

is a regularisation term that increases with the number of leaves T and the squared sum of the leaf weights, which are controlled by the parameters γ and λ, respectively. Hence, it penalises more complex trees as well as those with large weights. By doing so, the single trees tend to have smaller weights of similar absolute magnitude and similar number of leaves. In order to improve the performance with each newly constructed tree, the new optimal tree structure has to be determined. This is done

based on an iterative procedure and yields the actual *gradient boosting* strategy. Writing the loss function of the t'th tree as

$$L_t(s) = -\frac{1}{2}\sum_{k=1}^{T}\frac{\left(\sum_{i\in I_k} g_i\right)^2}{\sum_{i\in I_k} h_i + \lambda} + \gamma T \tag{5.9}$$

with

$$g_i = \partial_{\hat{\boldsymbol{y}}_i}\ell\left(\boldsymbol{y}_i, \hat{\boldsymbol{y}}_i\right)\big|_{\hat{\boldsymbol{y}}_i^{(t-1)}} \tag{5.10}$$

and the *Hessian*

$$h_i = \partial^2_{\hat{\boldsymbol{y}}_i}\ell\left(\boldsymbol{y}_i, \hat{\boldsymbol{y}}_i\right)\big|_{\hat{\boldsymbol{y}}_i^{(t-1)}}\,, \tag{5.11}$$

the latter two are both evaluated at the predictions of the previous tree $t-1$. I_k contains the indices of samples which would end up on leaf k. In this way, the loss function can be minimised with respect to the tree structure and the optimal tree can be found by setting the decision parameters at each node such that the samples end up in the optimal leaf based on the performance of the previous tree.

Since not all possible configurations can be tested due to resource limitations, the tree is actually built by an iterative *branching* procedure where the corresponding loss function is optimised to minimise the loss w.r.t. the next splitting at the current node. As a final step the leaf weights are scaled by the *learning rate* $\eta \in (0, 1]$. This is done to lower and control the impact of each new tree on the BDT ensemble hence approaching better and better tree performance in more fine-grained steps of the loss minimisation. It therefore decreases potential fluctuations of the performance of single or subsequent trees.

For further reading and detailed explanation of the algorithm it is referred to [151].

Sample preparation

Signal samples were taken from simulation by using the MC truth information to ensure only true prompt Λ_c^+ baryons were injected. Background samples were taken from the sidebands of the invariant mass distribution in data of the respective pre-selected candidate p_T. This was done to ensure a negligible signal contamination in the background samples.

Each sample was labelled with its target value, 1 for signal and 0 in case of background. Due to the availability of the target values, a supervised-learning approach was chosen and one BDT was constructed per Λ_c^+ p_T interval. All samples were shuffled and then split into a *training* set containing 80 % and the *test* set with 20 % of all samples. The number of signal and background samples contained in the training and test sets are summarised in Tab. 5.4 per p_T interval. As one can see, the statistics of the samples decreases with increasing p_T which is due to the fact of a lower available number of signal samples in MC at high p_T. Fitting the models for the higher p_T intervals with significantly less samples was still considered sufficient due to the better vertex and track resolution (see also Sec. 4.2.5). This will also be discussed later when the performance of the BDTs will be discussed.

p_T interval [GeV/c]	training		testing	
	signal	background	signal	background
[1, 2]	399800	400200	100200	99800
[2, 4]	399800	400200	100200	99800
[4, 6]	399800	400200	100200	99800
[6, 8]	269870	292705	67163	73481
[8, 12]	173546	186564	43254	46774
[12, 24]	63577	33282	15936	8279

Table 5.4: Sizes of signal and background samples for training and testing of the BDT models in each p_T interval.

Feature selection

Each observable taken into account as a feature was chosen to have a different distribution in signal and background, hence already providing some discriminating power on its own. The correlation with the reconstructed invariant mass is required to be negligible. Otherwise, a BDT might be biased to weight such a feature highly which effectively could result in an implicit selection on the invariant mass in order to select the signal.

Tab. 5.5 summarises the chosen features.

feature	description
b_p	proton candidate impact parameter w.r.t. the primary vertex
$b_{\pi^\pm}$	positive/negative decay track impact parameter w.r.t. the secondary vertex
b_{V^0}	V^0 impact parameter w.r.t. the primary vertex
DCA(V^0)	closest approach between the two V^0 decay tracks
d_{V^0}	V^0 decay length before secondary vertex
$c\tau_{V^0}$	V^0 proper lifetime
$\cos\theta^\star$	cosine of the proton emission angle in the parent particle's rest frame (not used in lowest two p_T intervals)
$\cos\theta_p$	cosine of V^0 pointing angle
v_{ap}	Armenteros-Podolanski of V^0 tracks as defined in Sec. 5.1
m_{V^0}	invariant mass of the V^0 decay tracks
$n_{\sigma_{TPC/TOF}}(p)$	number of sigmas of the proton candidate track fulfilling the proton hypothesis in the TPC or TOF

Table 5.5: Chosen features for the BDT classification.

Fig. 5.8 shows the normalised feature distributions in the lowest p_T interval, each of them already providing some discriminating power on its own. In addition, the invariant mass distribution of the reconstructed Λ_c^+ candidates is shown on the bottom-

right, indicating the background is taken from the sideband region. Sharp edges in the plots correspond to pre-selection cuts.

The correlation among features in the signal set are shown in Fig. 5.9. The invariant mass of the reconstructed Λ_c^+ candidates, which was not used as a feature as explained, was added to the plots to show its negligible correlation with the used features. Hence, all shown observables are considered to be valid features. The full compilation of feature correlation plots also including correlations of the background samples is provided in App. B.2. In all cases, the correlation with the invariant mass of the reconstructed Λ_c^+ candidates is considered to be negligible.

The proper lifetime and decay length of the V^0 candidate show a high correlation as it can be expected. Also the impact parameters of the $\pi^\pm$ candidates show a higher correlation, especially among the signal samples. A correlation with the DCA($\pi^+\pi^-$) can be expected as well since a larger impact parameter w.r.t. the secondary vertex is likely to enlarge the closest approach of the V^0 decay tracks. The Armenteros-Podolanski value shows negligible correlations with all other features and also the TPC and TOF PID of the proton candidate track are not correlated with any other feature.

Performance and hyperparameter optimisation

A BDT algorithm is fitted to yield the optimal tree structure to ensure high classification performance on given training set. To evaluate its generalisation capabilities, the classification performance is evaluated on the test set which has not yet been presented to the BDT. The performance measure used in the following is given by the so-called *receiver operating characteristics* (ROC) which yields the background rejection depending on the signal efficiency. The *area under curve* AUC is computed as the integral of the ROC; a value of 0.5 reflects a performance compatible to randomly draw values uniformly from $[0, 1]$ while a maximum AUC of 1 would reflect an algorithm that perfectly separates all signal from background samples. The generalisation capabilities can hence be estimated from the compatibility of the ROC and AUC of the training and test samples.

As seen in the previous section, the loss function has free parameters such as γ scaling the regularisation in the loss function or the learning rate η to scale the leaf weights. Furthermore, the complexity of a BDT increases with each branching and is $\mathcal{O}(2^d)$ for a tree of depth d. Packages such as XGBoost provide hence the possibility to choose these *hyperparameters*. The hyperparameters taken into account for the optimisation are summarised and described in Tab. 5.6.

The optimisation of each BDT model w.r.t. the hyperparameters is conducted to find the optimal set within a defined search space. The latter is summarised for each model in the first column of Tab. 5.7. An approach based on a grid search was employed which builds a BDT for each combination of hyperparameters, fits it and evaluates the ROC and AUC on the training and test set. The model configuration

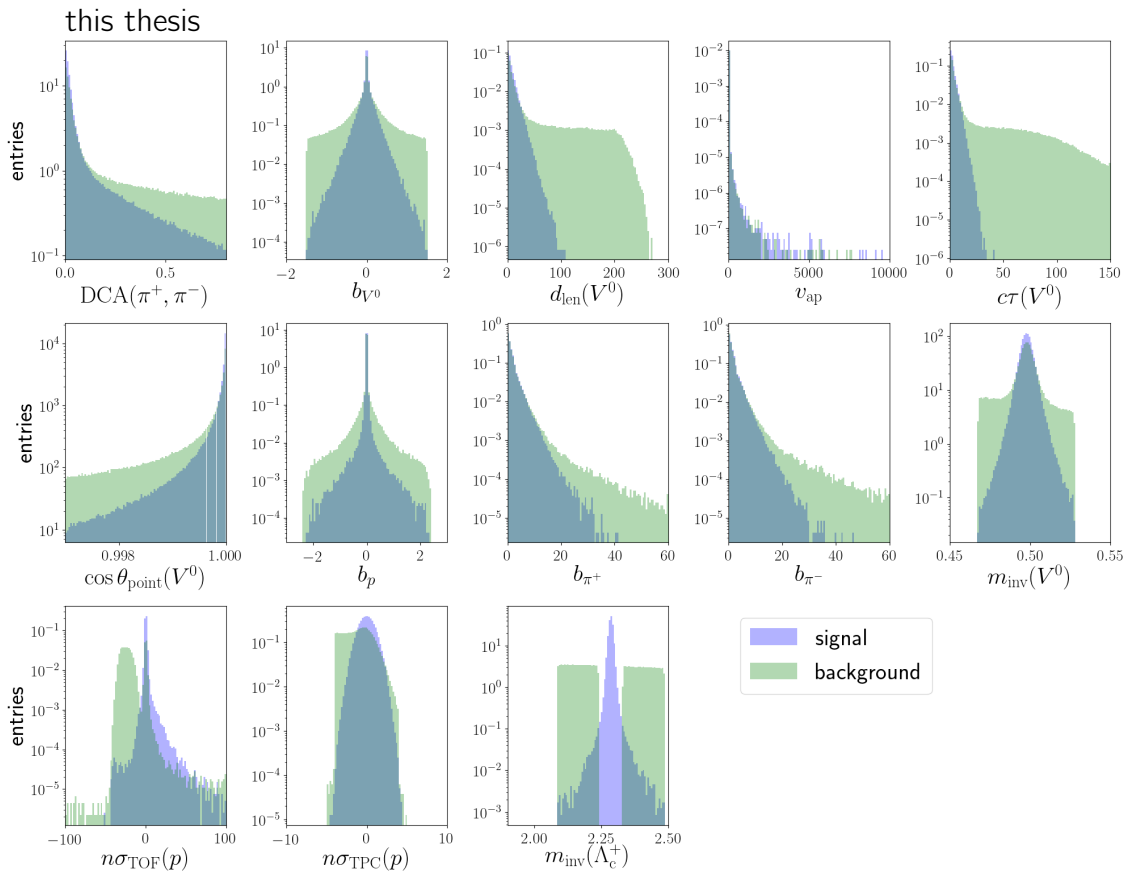


Figure 5.8: Feature distributions of signal (blue) and background (green) samples in the 1 GeV/c to 2 GeV/c momentum interval. Note, that the single histograms are not stacked but that the colours overlap in certain regions to identify the different shapes of the signal and background histograms. In addition to the feature distributions the invariant mass distribution of the Λ_c^+ candidates is shown as well.

was chosen which showed the highest test AUC for all p_T models. As a boundary condition, the relative deviation between the training and test AUC was required to be below 2 % at the most, hence $|1 - \text{AUC}_{\text{train}}/\text{AUC}_{\text{test}}| < 2\,\%$. The obtained configuration is given in Tab. 5.7 as well as the respective performance in each p_T interval.

The corresponding ROC curves of the optimised BDT structures are shown in Fig. 5.10. For all p_T intervals the performance of the model is compatible for the training set (dotted blue) and test set (dashed orange) and the AUC is of similar magnitude $91.01\,\% < \text{AUC}_{\text{test}} < 93.80\,\%$. Only in the highest p_T interval the corresponding model starts to show a slight deviation between the train and test performance, however, the relative deviation of the AUC for training and test is $\approx 1.7\,\%$ which was considered small. It can therefore also be concluded that the lower number of available samples for the higher p_T models did not have a significant

this thesis

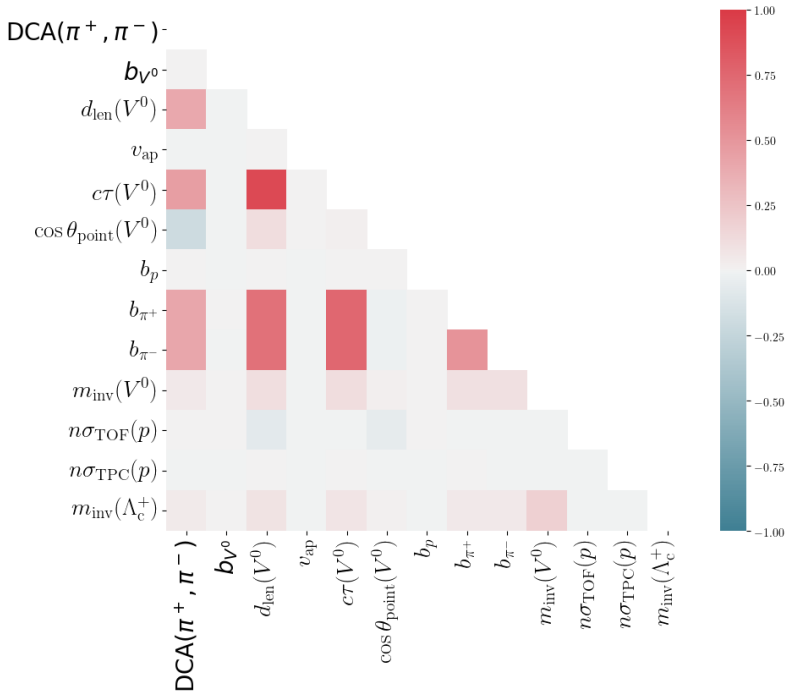


Figure 5.9: Correlations among the chosen features and the Λ_c^+ reconstructed invariant mass in the 1 GeV/c to 2 GeV/c momentum interval for signal samples. The colour code goes from red (fully positively correlated) to blue (fully negatively correlated) as indicated by the colour bar on the right.

parameter	description
h_{min}	minimum sum of Hessian instance weights (see Eq. (5.11)) required in a node to continue further branching
d_{max}	maximum number of consecutive splits
γ	minimum loss reduction required to split again at a given node
f_{samples}	ratio of samples to be used for fitting which is sampled once per boosting step
$f_{\text{col,tree}}$	ratio of features used for fitting which is sampled once per boosting step
η	learning rate
$n_{\text{estimators}}$	number of single fitted trees per BDT

Table 5.6: Hyperparameters to be optimised for the BDT configuration.

impact on the BDT performances (see also Tab. 5.4).

For the optimised hyperparameters, the relative importance of each feature used for the BDTs' decision making was evaluated again. For all models, the relative feature importance is given in Fig. 5.11. PID information obtained from the TPC

parameter	search space	optimal value					
		1	2	3	4	5	6
h_{min}	$\{1, 3\}$	3					
d_{max}	$\{2, 3, 6\}$	3					
γ	$\{0, 0.2\}$	0					
f_{samples}	$\{0.6, 0.8, 0.9\}$	0.8					
$f_{\text{col,tree}}$	$\{0.6, 0.8, 0.9\}$	0.8					
η	$\{0.05, 0.1, 0.5\}$	0.1					
$n_{\text{estimators}}$	$\{500, 850, 1000\}$	850					
AUC_{test}	/	0.9209	0.9103	0.9308	0.9367	0.9380	0.9323
$\text{AUC}_{\text{train}}$		0.9210	0.9124	0.9326	0.9376	0.9399	0.9480
$\Delta\text{AUC}/\text{AUC}_{\text{test}}[\%]$		0.01	0.2	0.2	0.1	0.2	1.7

Table 5.7: Optimal set of hyperparameters obtained via a grid search which evaluates the performances for each hyperparameter combination.

and TOF detectors were in all cases of highest importance. In the two lowest p_{T} intervals this is followed by topological features such as the proper decay length $c\tau$ and the pointing angle θ_{p} of the V^0 candidate as well as the proton impact parameter b_p. Hence, in this case the overall 3-prong topology was exploited by the BDTs. The Armenteros-Podolanski value v_{ap} gains importance for the two intermediate p_{T} intervals which reflects the decision making based on disentangling K_{S}^0 from background contributions introduced by Λ. For even higher p_{T}, the emission angle of the proton in its parent's rest frame gains importance. If correctly reconstructed and associated to a Λ_c^+, this emission angle is expected to be isotropic. With increasing p_{T} and an incorrectly reconstructed Λ_c^+, the proton candidate is boosted w.r.t. the parent system.

Choosing the BDT working points

The signal probability value assigned to each sample by the BDT will be used as a cut selection parameter. To do so, a *working point* p_{WP} was chosen and candidates assigned a probability below that were treated as background and above as signal candidates. The procedure of determining the working point is outlined in the following.

The final observable as introduced in the next section is proportional to the extracted Λ_c^+ signal yield S. The contribution of the relative statistical error from the extracted signal scales with $\sqrt{S}/S$. The BDT working points were hence chosen to reach a high expected significance $S_{\text{exp}}/\sqrt{S_{\text{exp}} + B_{\text{exp}}}$ for the selection.

The expected signal S_{exp} was estimated from FONLL predictions where the p_{T} shape of prompt Λ_c^+ baryons has been computed (see App. B.1 for further explana-

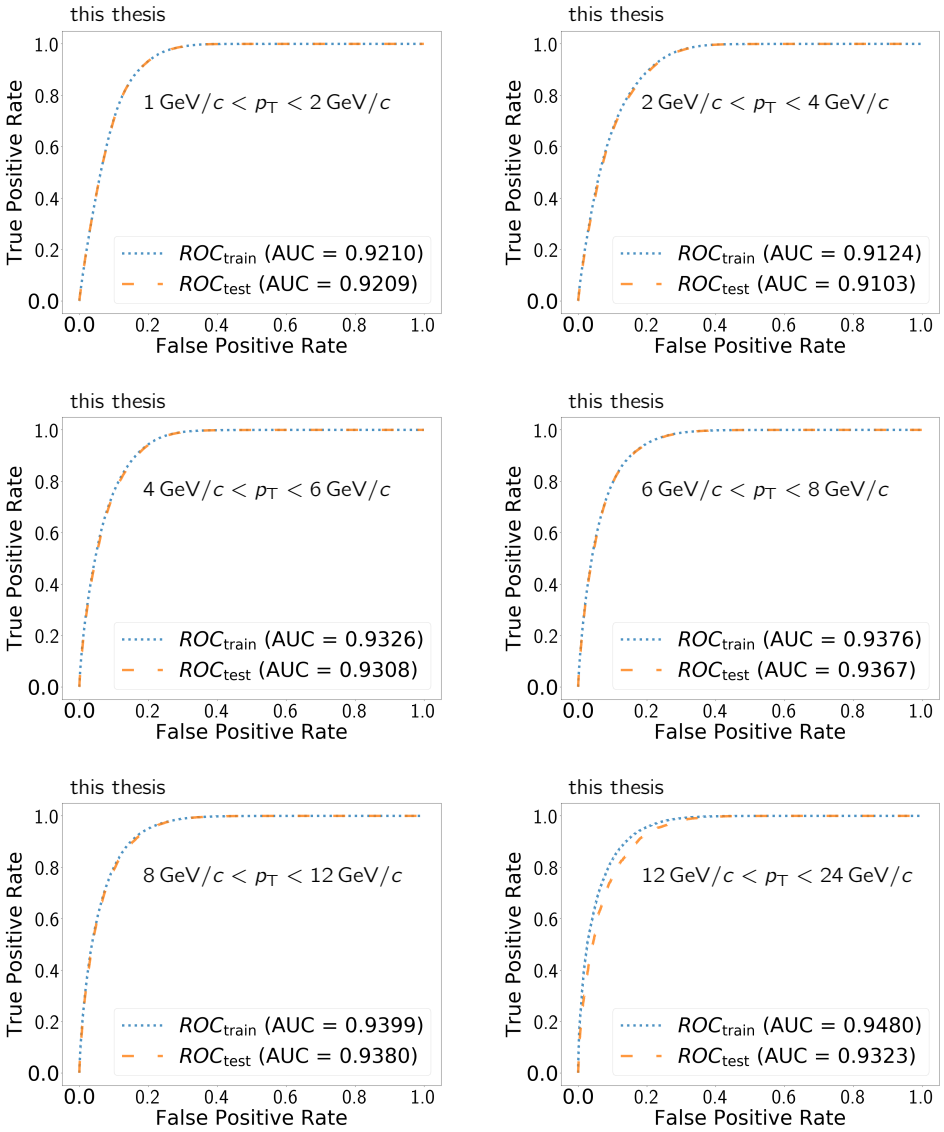


Figure 5.10: ROC curves for all models with optimised hyperparamaters from lowest to highest p_T interval from *top-left* to *bottom-right*.

tion) according to

$$S_{exp} = 2 \left(\frac{d\sigma}{dp_T} \right)^{FONLL}_{prompt} \cdot (Acc \times \varepsilon)_{prompt} c_{\Delta y} \Delta p_T \cdot BR \cdot \mathcal{L} / \tilde{f}_{prompt}. \quad (5.12)$$

Since the FONLL results are given as the cross section of prompt Λ_c^+ production, it has to be scaled to what is expected from the candidate selection obtained by using the BDT. Hence, the above equation is scaled back to the decay channel $\Lambda_c^+ \rightarrow pK_S^0 \rightarrow p\pi^+\pi^-$ and divided by the f_{prompt} fraction to re-introduce potential feed-down contributions from $\Lambda_b^0 \rightarrow \Lambda_c^+ + X$ which would also be selected by the BDT. Furthermore, it is multiplied by the expected Λ_c^+ selection efficiency (Acc $\times$

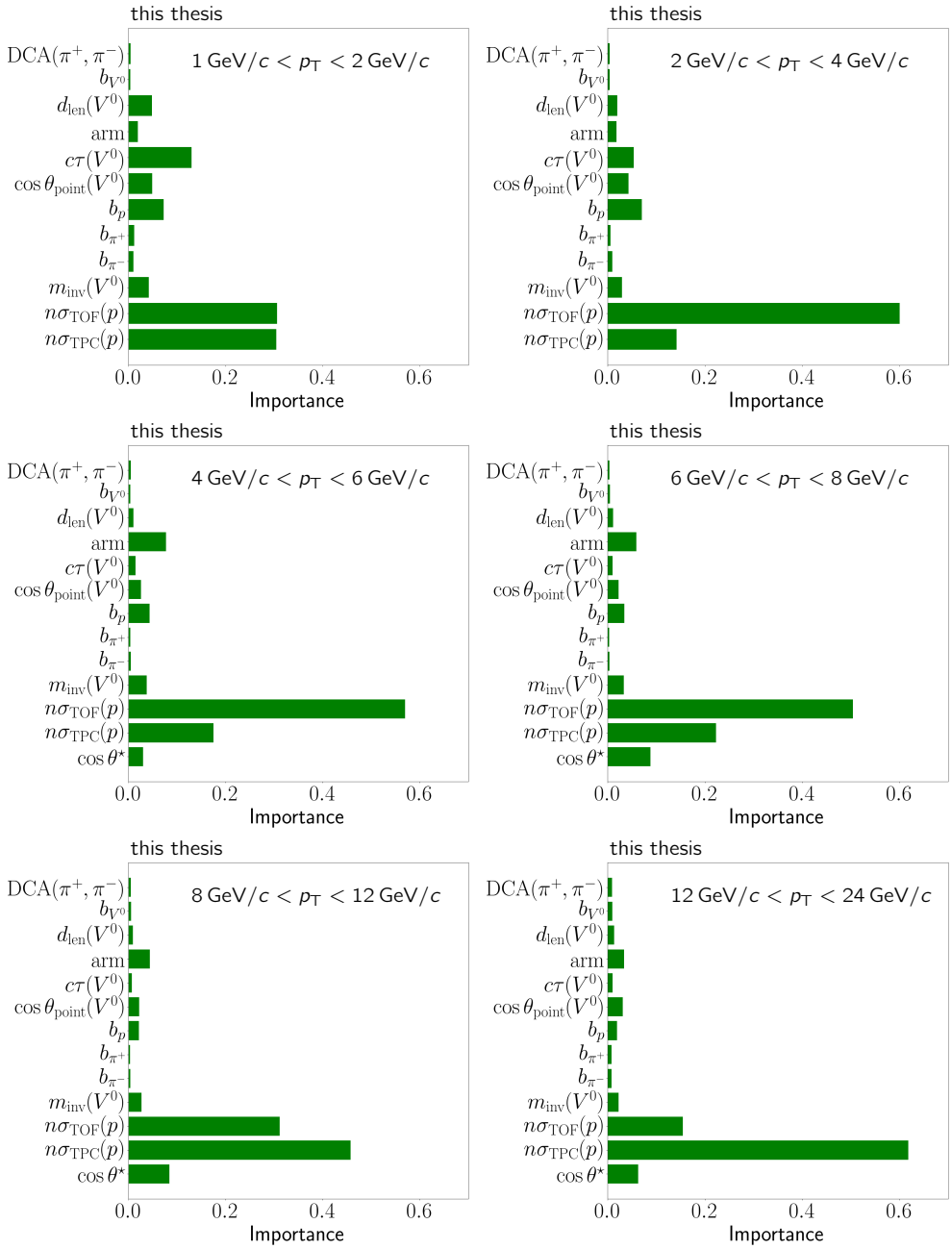


Figure 5.11: Feature importance for all models with optimised hyperparamaters from lowest to highest p_T interval from *top-left* to *bottom-right*.

$\varepsilon)_{\text{prompt}}$*. The factor 2 accounts for the selection of charged-conjugates and finally, the cross section is scaled by the collected integrated luminosity $\mathcal{L} = \sigma_{\text{MB}} N_{\text{ev}}$ in the $[0\,\%, 100\,\%]$ multiplicity interval computed from the total minimum-bias cross section σ_{MB} and the number of selected events N_{ev}.

The number of expected background candidates B_{exp} is estimated from a second-

*Both the calculation of the f_{prompt} fraction and $(\text{Acc} \times \varepsilon)_{\text{prompt}}$ are outlined in detail in Sec. 5.3.2.

order polynomial fit in the invariant-mass sideband region in a subset of the events which is then scaled to the total size of the selected data. Both B_{exp} as well as $(\text{Acc} \times \varepsilon)_{\text{prompt}}$ were computed for each scanned working point and it was done so for each BDT model.

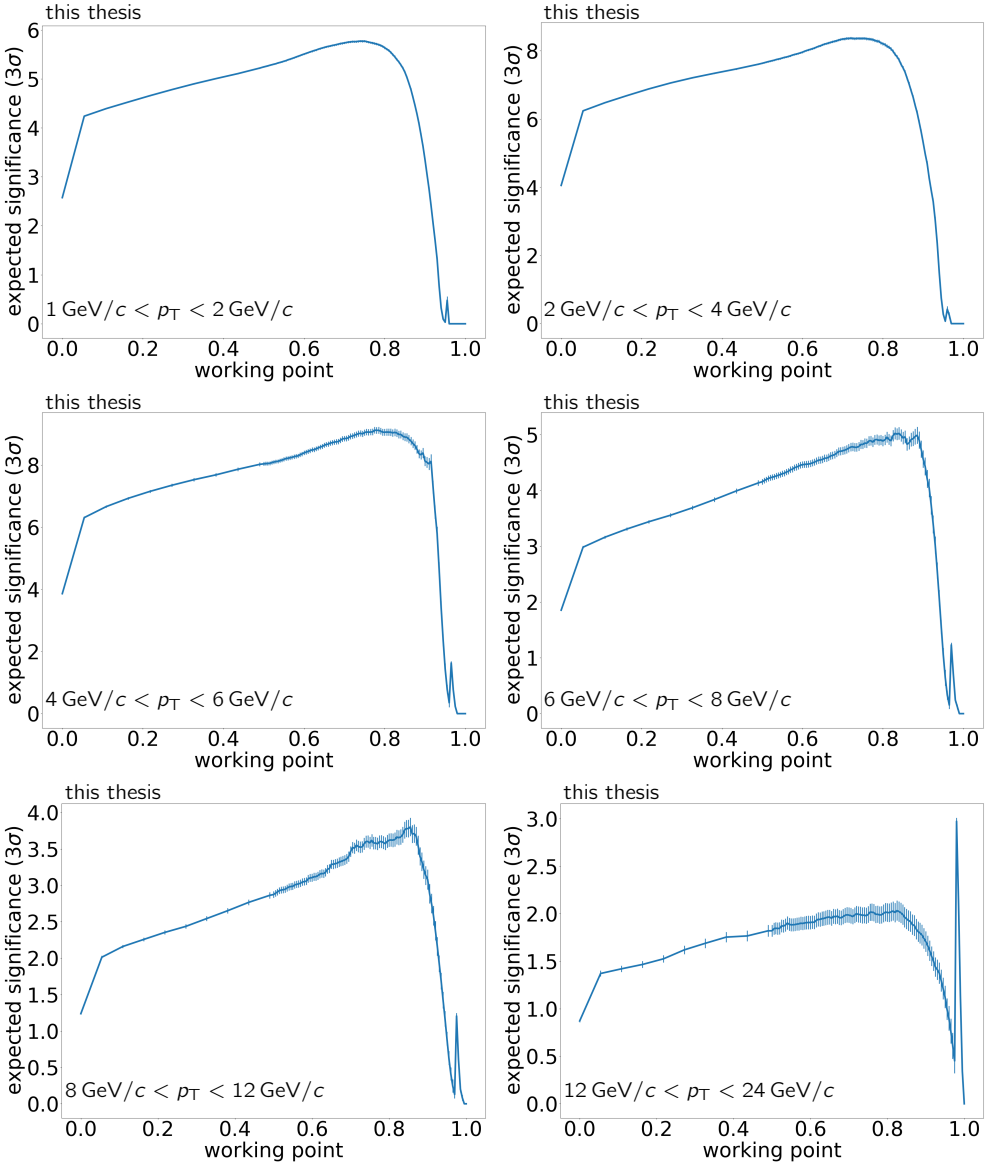


Figure 5.12: Expected statistical significance of the BDT selection as a function of the chosen working point for lowest to highest p_{T} interval model from *top-left* to *bottom-right*.

The expected significance was then estimated as a function of the BDT working point for each p_{T} model. Fig. 5.12 shows the corresponding curves from the lowest to the highest momentum intervals from top-left to bottom-right. In both cases the expected significance constantly rises with an increasing working point until it

reaches a maximum at ≈ 0.8. This is followed by a sharp drop of the working point when almost all signal gets lost. For all p_T intervals, the working point was chosen to be 0.7 for three reasons: (i) to guarantee a stable performance around the working point and to avoid entering the region of dropping significance, (ii) to maintain a high signal yield, (iii) to not cut too tight in the case of MC which would introduce a larger impact of statistical uncertainties for the efficiency estimation.

5.3 Analysis strategy

The final observable is the p_T spectrum of the prompt Λ_c^+ yield per event in intervals of forward multiplicity p_{V0M} given as

$$\frac{1}{N_{\text{mult}}^{\text{ev}}} \frac{\mathrm{d}N_{\text{mult}}^{\Lambda_c^+}}{\mathrm{d}p_T} = \frac{\varepsilon_{\text{mult}}^{\text{INEL}}}{N_{\text{mult}}^{\text{ev,sel}}} \frac{1}{c_{\Delta y}(p_T) \cdot \Delta p_T} \frac{1}{\text{BR}} \cdot \frac{f_{\text{prompt}}(\Delta p_T) \cdot \frac{1}{2} \cdot N_{\text{mult}}^{\Lambda_c^+,\text{raw}}(\Delta p_T)\Big|_{|y|<y_{\text{fid}}(p_T)}}{(\text{Acc} \times \varepsilon)_{\text{prompt,mult}}(\Delta p_T)}.$$

It is calculated from the number of selected events per multiplicity interval $N_{\text{mult}}^{\text{ev,sel}}$ and the extracted raw number of Λ_c^+ baryons $N_{\text{mult}}^{\Lambda_c^+,\text{raw}}$ per multiplicity interval.

The number of selected events is corrected by an efficiency factor $\varepsilon_{\text{INEL}}$ to account for the INEL > 0 criterion and hence for those events fulfilling the INEL > 0 requirement but were not selected by the trigger and event selection criteria. The raw number of selected Λ_c^+ candidates is divided by a factor of 2 to obtain the average particle-antiparticle yield. Dividing by BR accounts for the fact that the selection was done in the specific $\Lambda_c^+ \to \text{p}K_S^0 \to \text{p}\pi^+\pi^-$ decay channel while the overall Λ_c^+ production is of interest. To account for feed-down contribution of non-prompt Λ_c^+ baryons originating from beauty-hadron decays, the raw yield is scaled by the prompt fraction f_{prompt} which was obtained from FONLL calculations (see App. B.1 for further explanations). The raw yield is further corrected for the acceptance and efficiency $(\text{Acc} \times \varepsilon)_{\text{prompt,mult}}$ of the Λ_c^+ yield which is extracted from MC simulations. The fiducial rapidity acceptance region for the Λ_c^+ reconstruction in the ALICE detector depends on the transverse momentum. It increases smoothly from $|y| < 0.5$ to $|y| < 0.8$ for $0 < p_T < 5\,\text{GeV}/c$ and is constant at $|y| < 0.8$ for larger transverse momenta. A correction factor $c_{\Delta y}(p_T)$ was applied to account for this selection (see also discussion in [110]).

The here presented analysis is at the same time also analogously being prepared using the number of SPD tracklets n_{trkl} within $|\eta| < 1$ for the event multiplicity estimation. The multiplicity intervals of the here presented analysis as well as, for comparison and since they will be used later, the intervals of the barrel multiplicity analysis are summarised in Tab. 5.8 together with the estimated values of the average

Mult. estimator	Mult. interval	$\langle dN_{ch}/d\eta \rangle$
p_{V0M} [%]	[30, 100]	$4.51^{+0.07}_{-0.05}$
	[0.1, 30]	$14.04^{+0.18}_{-0.15}$
	[0, 0.1]	$31.87^{+0.60}_{-0.54}$
n_{trkl}	[1, 9]	3.10 ± 0.02
	[10, 29]	10.54 ± 0.01
	[30, 59]	22.56 ± 0.07
	[60, 99]	37.83 ± 0.06
INEL > 0		$7.06^{+0.10}_{-0.08}$

Table 5.8: Summary of the multiplicity event classes at midrapidity (n_{trkl}) and forward rapidity (p_{V0M} [%]). The average charged-particle densities $\langle dN_{ch}/d\eta \rangle$ are shown.

primary charged-particle density $\langle dN_{ch}/d\eta \rangle$ per interval. The last row shows the values for the entire INEL > 0 set obtained with the MB trigger.

The $\langle dN_{ch}/d\eta \rangle$ values corresponding to the studied percentile multiplicity intervals in this work were taken from the comprehensive analysis studying primary charged-particle distributions as a function of mid- and forward-rapidity estimators conducted by ALICE [150]. The primary charged-particle density per event was obtained by the analysers from the SPD tracklet density $dn_{trkl}/d\eta$ following the procedure in [152] according to

$$dN_{ch}/d\eta = \alpha(1-\beta)dn_{trkl}/d\eta. \tag{5.13}$$

The correction factor α denotes the acceptance and efficiency for a primary charged particle to produce a tracklet and β corresponds to the number of SPD hits combined to tracklets that were not produced by the same primary particle. The corrections were obtained from MC simulation as a function of the longitudinal primary vertex position as well as the pseudorapidity of the tracklet.

In the next section, the raw yield extraction is presented, which is followed by the detailed extraction and discussion of the corrections mentioned above.

5.3.1 Raw yield extraction

The raw yield was extracted per p_T and multiplicity interval by means of a binned likelihood fit to the invariant mass distribution of the selected Λ_c^+ candidates. A Gaussian function was chosen to describe the signal and for the background a second order polynomial function was used. To ensure a stable fit procedure, the Gaussian widths per p_T interval were fixed parameters. They are obtained per p_T interval from a Gaussian fit to the MC signal distribution in a region of 3 RMS around the maximum of the corresponding distribution to avoid picking up the non-Gaussian tails of the distributions. These *pre*-fits are shown in Fig. 5.13 for increasing p_T interval

from the top-left to the bottom-right. The widths obtained from an unconstrained fit to the invariant mass distribution in data from MB-triggered events in the 0 % to 100 % multiplicity region were compared to the ones obtained from MC in Fig. 5.14. It can be seen how the fits' widths directly obtained from data fluctuate around the MC values. Hence, it was decided to fix the Gaussian widths for the final data fits to that obtained from MC.

The Gaussian mean and width were initialised with the parameters from the MC fit. While the width was then fixed, all other parameters including the Gaussian mean and the background parameters were left unconstrained. The signal S and background contamination B were extracted in the region of 3 Gaussian sigmas around the maximum. The fits of the 0 % to 100 % multiplicity region are shown in Fig. 5.15, the other ones can be found in the App. B.3. In blue the full fit function is shown, red indicates the background fit over the whole region and grey corresponds to an initial fit of the background component in the sideband region defined as ± 4 Gaussian sigmas away from the initial mean parameter. For each fit the significance defined as

$$p_{\text{signif}} = \frac{S}{\sqrt{S+B}}, \tag{5.14}$$

was computed. A fit was defined to have failed and was rejected in a given p_{T} and multiplicity interval if $p_{\text{signif}} < 3$. In that case, the corresponding interval was not used in subsequent steps of the analysis. Fig. 5.16 summarises the raw signal yields and the significances in the left and right panel, respectively. The significance in the highest p_{T} interval of the lowest multiplicity region is below 3 and was hence discarded from the further analysis. The raw yield increases with the event multiplicity according to the expectation that an increasing number of Λ_c^+ baryons is produced. Note, that a higher number of candidates is extracted in the 0 % to 0.1 % multiplicity region compared to the full region. This is due to the fact that these are two samples selected with different triggers so the first is not fully contained in the latter.

Tab. 5.9 summarises the raw yield per p_{T} and multiplicity interval.

mult. interval [%]	p_{T} region [GeV/c]					
	[1, 2]	[2, 4]	[4, 6]	[6, 8]	[8, 12]	[12, 24]
[0, 100]	8034	12109	5463	1724	720	224
[30, 100]	2184	3328	1284	394	150	-
[0.1, 30]	5888	8653	4196	1322	564	214
[0, 0.1]	14617	30063	16640	5007	2499	758

Table 5.9: Extracted raw Λ_c^+ yield per p_{T} and multiplicity interval.

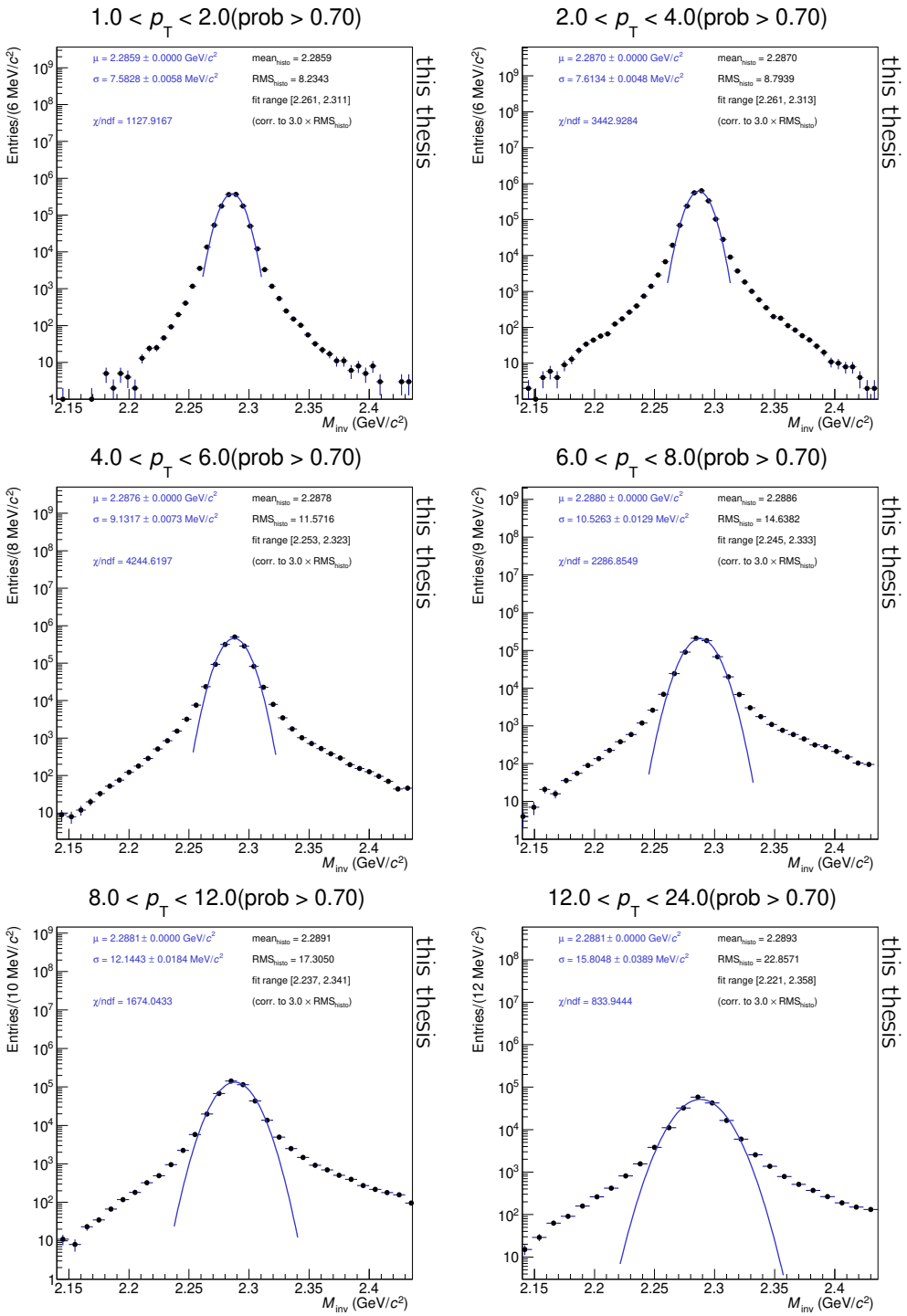


Figure 5.13: Gaussian fit to MC invariant mass distribution in the 3 RMS region around the maximum.

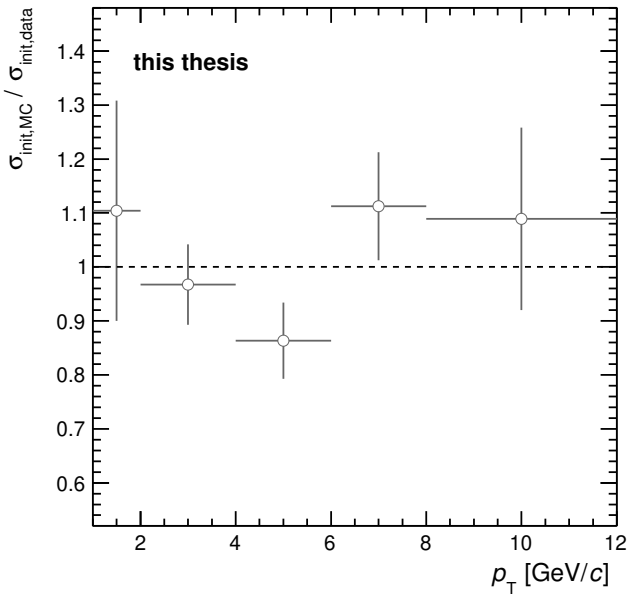


Figure 5.14: Ratio of widths derived from an unconstrained fit to data in the MB-triggered sample in the 0 % to 100 % multiplicity region and those obtained from fits in the signal region taken from simulation.

5.3.2 Corrections

In the following, the various required corrections according to Eq. (5.13) will be discussed with a main focus on the estimation of the acceptance and efficiency as well as on the f_{prompt} fraction per p_T region.

Acceptance and efficiency

The acceptance and efficiency of the Λ_c^+ selection was estimated using MC simulations. To do so, generated events were transported through the ALICE detector geometry by means of GEANT3 [34]. Except for the selection on p_{V0M} due to the missing calibration, the same selection criteria as applied in data were used to select candidates in simulation. To account for the multiplicity selection in simulation, a correction was applied as explained in the following.

For each forward-multiplicity interval, the *barrel* multiplicity distribution in data was determined as the number of SPD tracklets. That can be seen in Fig. 5.17 for the years 2016 to 2018 separately. The top row shows the normalised tracklet distributions per p_{V0M} interval in data obtained using the MB trigger along with the anchored MC distribution and the second row shows the corresponding HMVZERO selected data and anchored MC. Since the MC samples are obtained by requesting either a $c\bar{c}$ or $b\bar{b}$ quark pair, the barrel-multiplicity distributions in data are extracted

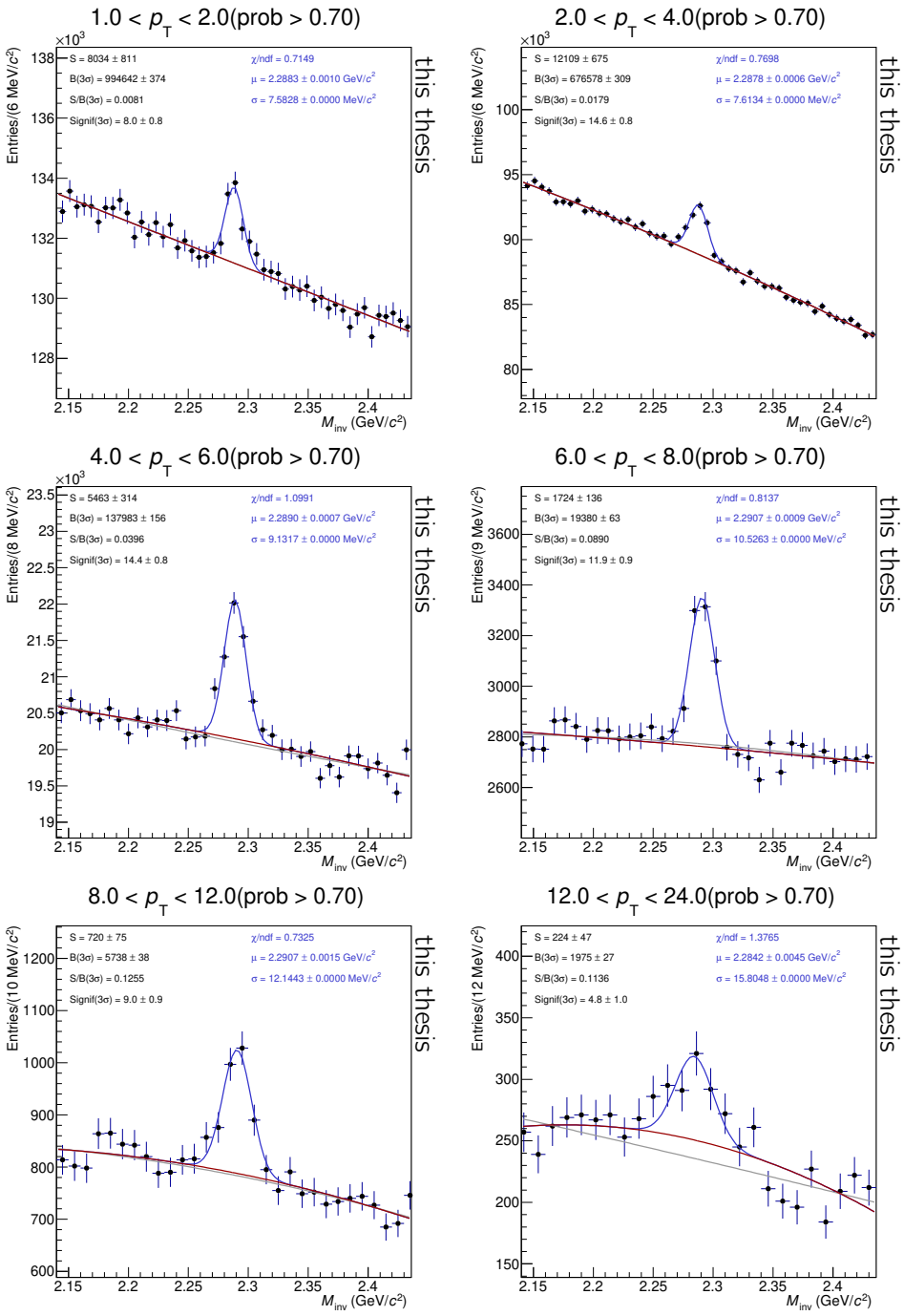


Figure 5.15: Binned maximum-likelihood fit to the invariant mass distribution of selected Λ_c^+ candidates per p_T interval *(from top-left to bottom-right)* in the $[0\,\%, 100\,\%]$ multiplicity interval. The initial sideband fit is shown in grey while the full fit is shown in blue. In red the sideband part of the full fit is shown.

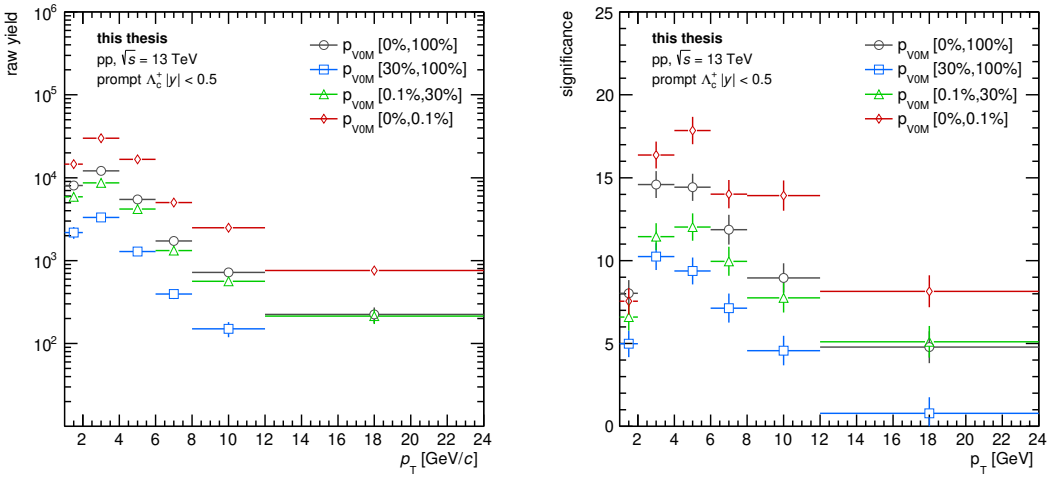


Figure 5.16: Significances of fits for the yield extraction *(left)* and extracted yields *(right)* as a function of the Λ_c^+ candidate p_T per multiplicity interval. Since the significance of the highest p_T interval in the [30 %, 100 %] has a significance below 3, it is not shown.

by requesting at least one Λ_c^+ candidate selected according to the candidate selection criteria explained in Secs. 5.2.3 and 5.2.4. In addition, the candidate was requested to have an invariant mass which differs at most by $20\,\text{MeV}/c^2$ from the value extracted by the Particle Data Group [13]. Clearly, the simulated barrel multiplicities were not sufficiently modelled compared to data. To account for this, event weights were built by dividing the data distributions by the MC distributions, separately for the MB and HMVZERO trigger scenario. The obtained weights are shown in the bottom row of Fig. 5.17. The efficiency-times-acceptance for the prompt Λ_c^+ selection is hence given as

$$(\text{Acc} \times \varepsilon)(\Delta p_T)^{\text{prompt,mult}} = \frac{\sum_i N_i^{\text{sel}}(\Delta p_T)\omega_i^{\text{mult}}}{\sum_i N_i^{\text{gen}}(\Delta p_T)\omega_i^{\text{mult}}}, \tag{5.15}$$

where i specifies the barrel multiplicity bin, ω_i is the weight of that bin as derived above, and N_i^{sel} and N_i^{gen} are the selected and generated numbers of prompt Λ_c^+ baryons in simulation, respectively. The superscript mult refers to the forward-multiplicity interval in terms of p_{V0M}. Analogously, the acceptance-times-efficiency of feed-down (FD) contributions from beauty hadrons decaying to Λ_c^+ is determined from MC simulations. Fig. 5.18 summarises the prompt and FD acceptance-times-efficiency per multiplicity interval in the left and right panel, respectively. Note, that without these corrections, the efficiency would have been the same for all multiplicity intervals.

The efficiencies increase with increasing p_T caused by a better separation of candidate tracks and an increasing vertex resolution since the impact parameter resolution increases significantly with p_T (see also Fig. 4.10). The efficiencies for the highest

multiplicity interval show a larger fluctuation w.r.t. the [0 %, 100 %] interval in particular due to the available statistics in the HMVZERO-anchored MC samples. The trends of the differential multiplicity intervals around the [0 %, 100 %] interval are comparable per p_{T} interval for the prompt and feed-down acceptance and efficiency.

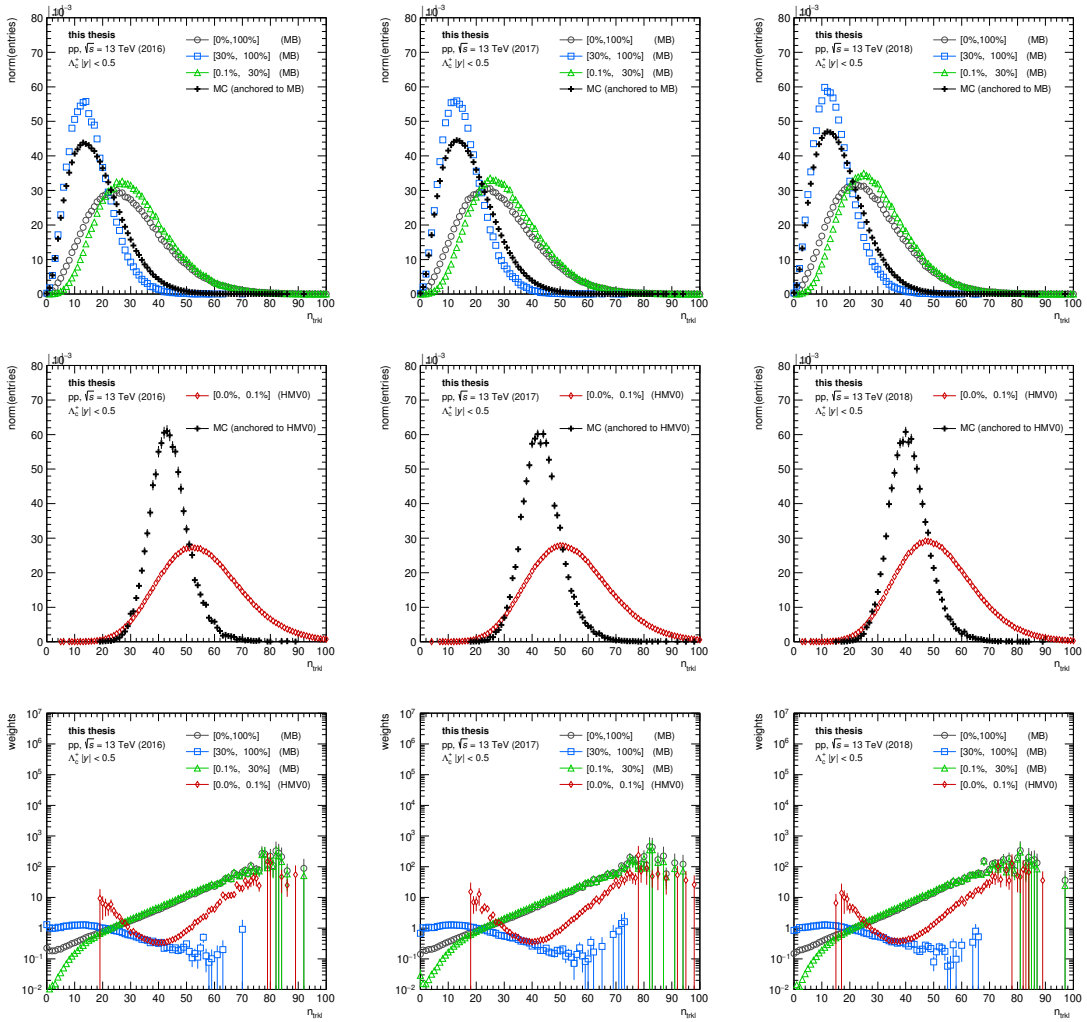


Figure 5.17: Distributions for different data-taking years from left to right. *Top row:* Normalised SPD tracklet distribution in slices of p_{V0M} corresponding to the chosen analysis intervals selected with the MB trigger along with the distribution obtained in the anchored simulation. *Middle row:* Normalised SPD tracklet distribution in the highest p_{V0M} interval selected with the HMVZERO trigger along with the distribution obtained in the anchored simulation. *Bottom row:* MC event weights derived from the ratio of distributions obtained in data and MC for each forward-multiplicity interval separately.

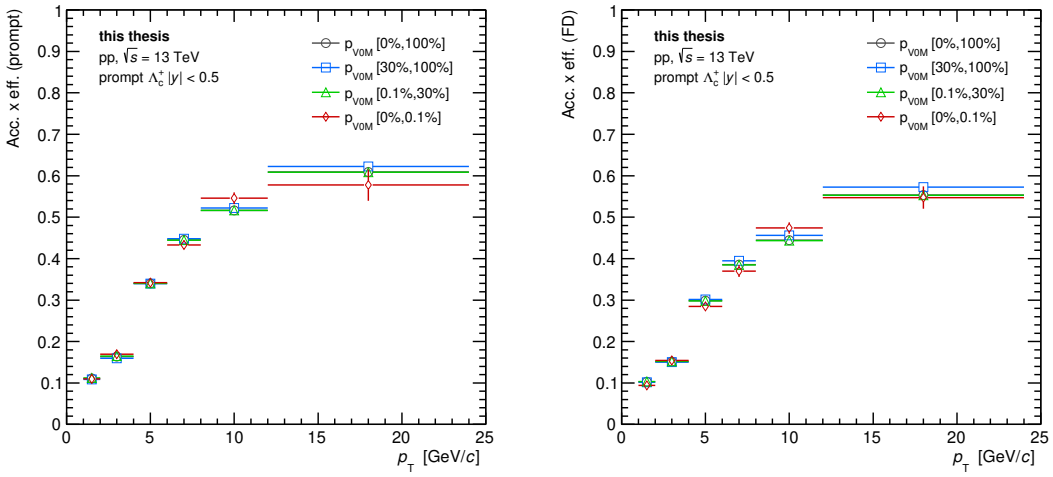


Figure 5.18: Acceptance-times-efficiency as a function of the candidate p_T for prompt Λ_c^+ *(left)* and candidates from $\Lambda_b^0 \to \Lambda_c^+ + X$ feed-down *(right)* per multiplicity interval.

Λ_c^+ prompt fraction from beauty feed-down

The raw Λ_c^+ yield was corrected for FD contribution originating from beauty hadrons decaying to Λ_c^+ by computing the f_{prompt} fraction. Generically, it is computed by subtracting the FD fraction according to

$$f_{\text{prompt}} = 1 - \left(\frac{N^{\text{FD,raw}}_{[0\,\%,100\,\%]}}{N^{\Lambda_c^+,\text{raw}}_{[0\,\%,100\,\%]}} \right). \tag{5.16}$$

An approach based on FONLL predictions was chosen in this analysis since a data-driven approach is not feasible due to low statistics*. Using the FONLL approach, the FD fraction in the above equation is given as

$$\left(\frac{N^{\text{FD,raw}}_{[0\,\%,100\,\%]}}{N^{\Lambda_c^+,\text{raw}}_{[0\,\%,100\,\%]}} \right) = \left(\frac{\mathrm{d}^2\sigma}{\mathrm{d}y\mathrm{d}p_T} \right)^{\text{FONLL+Pythia8}}_{\text{FD}} \cdot \frac{(\text{Acc} \times \varepsilon)_{\text{FD}} \Delta y \Delta p_T \cdot \text{BR} \cdot N^{\text{ev}}_{[0\,\%,100\,\%]}}{N^{\Lambda_c^+,\text{raw}}_{[0\,\%,100\,\%]}/2}. \tag{5.17}$$

The differential cross section given as the first term on the right-hand side of the above equation is estimated from FONLL predictions of beauty-quark production.

*An analysis of D-meson production in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ and p–Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02\,\text{TeV}$ conducted by ALICE [153] tested a data-driven approach by making use of the different shapes of of the impact parameter distribution of prompt and feed-down D-mesons in the transverse plane. This approach has shown compatible uncertainties with the FONLL-based method. In the lowest p_T interval of $1\,\text{GeV}/c$ to $2\,\text{GeV}/c$ the data driven-approach was not applicable due to low precision and in the highest p_T interval of $16\,\text{GeV}/c$ to $24\,\text{GeV}/c$ it could only be extracted for the D^{*+} which provided sufficient precision.

Measurements of fragmentation fraction to Λ_b^0 as obtained by LHCb in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ in the pseudorapidity region $2 < |\eta| < 5$ [154] where used. While showing a p_T dependence, no rapidity dependence was observed and therefore the results could be used for this analysis where the selection was done at central rapidity. The decay kinematics of $\Lambda_b^0 \to \Lambda_c^+ + X$ were handled by PYTHIA 8 [94].

The FONLL predictions were scaled by the branching ratio BR($\Lambda_b^0 \to \Lambda_c^+ + X$) as well as by the raw charge-averaged Λ_c^+ yield per event in the INEL > 0 data set. The f_{prompt} fraction was therefore assumed to be independent of the multiplicity interval but as explained in Sec. 5.4, multiplicity dependent systematic uncertainties were derived. The derived f_{prompt} fraction per p_T interval is shown in Fig. 5.19. The theoretical uncertainties arise from variations of the beauty-quark mass as well as of the renormalisation and factorisation scale in the FONLL calculations.

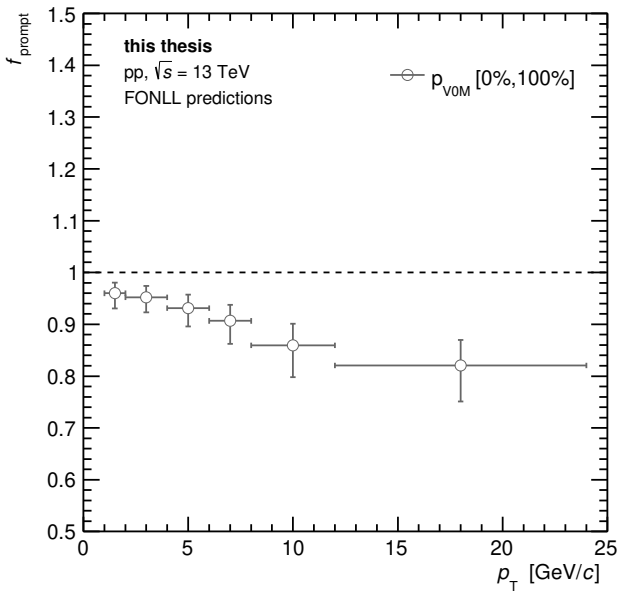


Figure 5.19: Estimated prompt Λ_c^+ fraction per p_T interval accounting for feed-down from $\Lambda_b^0 \to \Lambda_c^+ + X$ according to Eq. (5.16).

Branching ratio

As mentioned already in Sec. 5.1, the branching ratio of the full decay channel of $\Lambda_c^+ \to pK_S^0 \to p\pi^+\pi^-$ is BR = 1.59(8) % [13]. The nominal value is used here and the uncertainty is treated as a systematic uncertainty which will further be discussed in Sec. 5.4.

Trigger and event selection correction

The efficiency correction factor $\varepsilon_{\text{INEL}}$ as specified in the table has been extracted centrally by the ALICE collaboration [150] to correct for those events fulfilling the INEL > 0 and event selection requirements but were not selected by the triggers. The values for each multiplicity interval are summarised in Tab. 5.10. The uncertainties will be discussed in Sec. 5.4.8.

mult. interval [%]	$\varepsilon_{\text{INEL}}$
[0, 100]	0.920 ± 0.003
[30, 100]	0.897 ± 0.031
[0.1, 30]	0.997 ± 0.001
[0, 0.1]	1.000 ± 0.000

Table 5.10: Trigger efficiency correction per multiplicity interval to correct for events not selected by the trigger but fulfilling the INEL > 0 criterion.

5.4 Systematic uncertainties

Different systematic uncertainties are considered in this analysis. Two of those are assumed to be independent of the multiplicity interval, namely the MC p_{T} shape uncertainty and the uncertainty on the TPC track efficiency. The latter is known to depend on multiplicity in Pb–Pb collisions but this dependence becomes negligible at the significantly lower multiplicities produced in pp collisions. After discussing these multiplicity-independent sources of systematic uncertainties, those depending on the multiplicity will be discussed. In the following section each uncertainty will be explained and at the end all uncertainties will be summarised together. Some of the following systematic uncertainties have been adopted from the analysis of $\Lambda_c^+ \to \text{pK}^-\pi^+$, namely those for the MC p_{T} shape, the track reconstruction efficiency and uncertainty on the z_{vtx} distribution.

5.4.1 MC p_{T} shape

The MC p_{T} shape uncertainty accounts for a potential bias in the p_{T} distribution as modelled in simulation which might be different to the one in data. This uncertainty was estimated using a standard procedure adopted in the ALICE collaboration.

The p_{T} distribution as modelled by PYTHIA 6 and the one from FONLL predictions were each normalised and then divided in order to calculate p_{T} weights. Those were applied when calculating the prompt Λ_c^+ acceptance and efficiency (the weights are applied analogously to the multiplicity weights according to Eq. (5.15)). The normalised distributions are shown in the left panel of Fig. 5.20 followed by the

obtained weights shown on the right and the bottom row contains the efficiency ratio. The biggest impact is seen in the lowest p_{T} interval where the relative uncertainty is around 3 % where the simulated shape is steeper compared to the shape obtained from FONLL. It then decreases with increasing p_{T} and is assumed to be negligible in the 4 highest intervals.

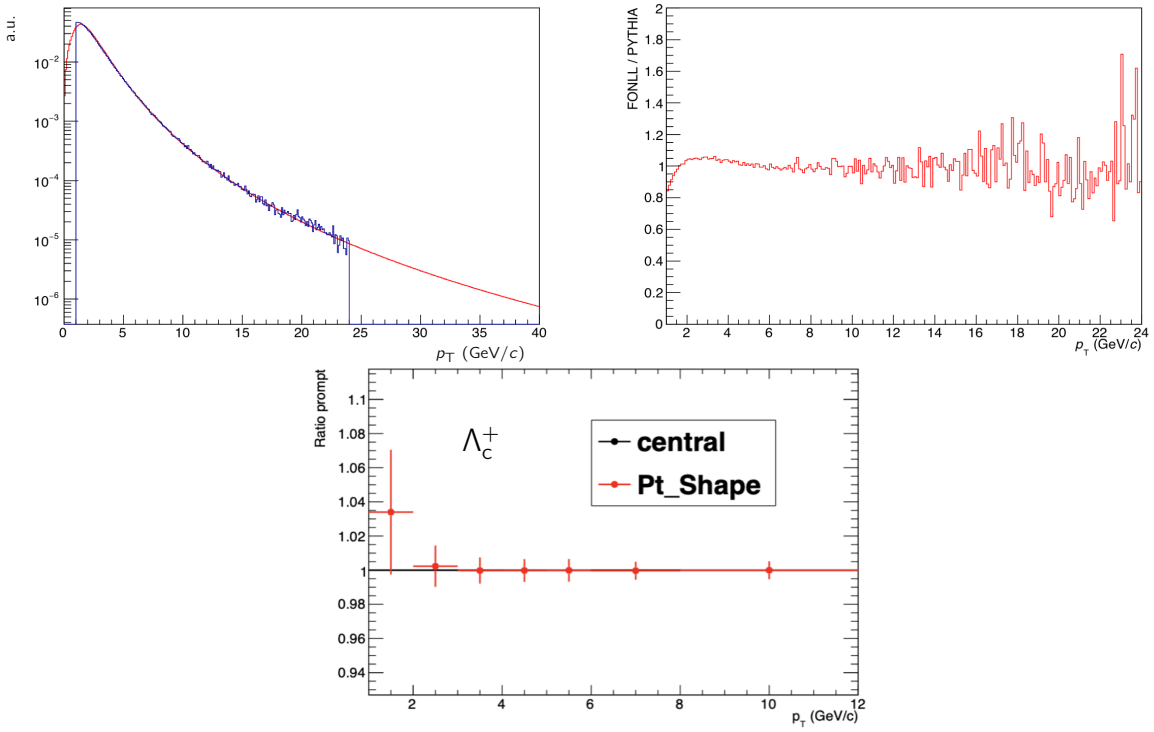


Figure 5.20: *Top-left:* Normalised p_{T} distributions obtained from PYTHIA 6 simulation (blue) and FONLL predictions (red). *Top-right:* Weights to be applied in the calculation of the prompt Λ_{c}^{+} acceptance and efficiency. *Bottom:* Relative systematic MC p_{T} shape uncertainty.

5.4.2 Track reconstruction

Different track reconstruction parameters might have a different impact on data and MC which could lead to a bias introduced by the acceptance and efficiency. Furthermore, the ITS–TPC matching efficiency could differ in data and MC. Therefore, these are sources of a potential systematic uncertainty.

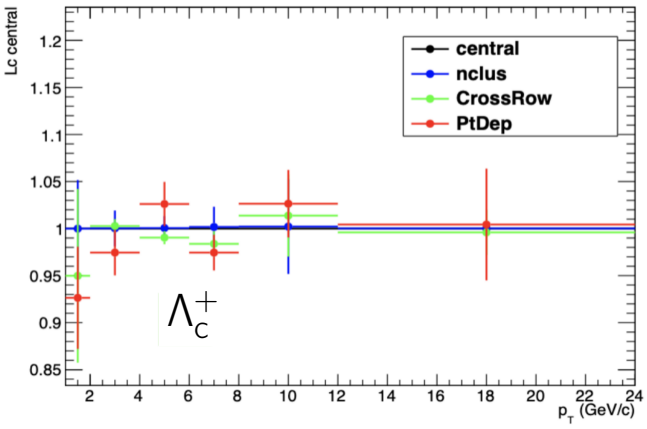


Figure 5.21: Relative deviation of the corrected yield per event from the nominal value for three different scenarios of varied TPC track selection criteria.

TPC track reconstruction

To estimate the uncertainty on the TPC track reconstruction, three additional scenarios of different track selection cuts were employed:

1. number of crossed rows $n_{\text{xrows}}^{\text{TPC}}$ has to be > 120,
2. $n_{\text{xrows}}^{\text{TPC}} > 120 - \frac{5\,\text{GeV}/c}{p_{\text{T}}}$,
3. number of TPC clusters to be at least $0.65 \cdot n_{\text{xrows}}^{\text{TPC}}$.

For all variations the ratio of $n_{\text{xrows}}^{\text{TPC}}$ over findable clusters has to be > 0.9. Due to the relatively low occupancy in pp collisions compared to Pb–Pb collisions this uncertainty could be evaluated independent of the studied multiplicity interval. Fig. 5.21 shows the ratio of the corrected yield per event for each scenario to the nominal result to obtain the relative uncertainties. They are largest for the lowest p_{T} interval and everywhere of the order of 1 %.

ITS–TPC matching efficiency

The uncertainty in the ITS–TPC track matching efficiency $\varepsilon_{\text{match}}$ is estimated centrally by the ALICE Data Preparation Group and the method is briefly summarised here.

It is defined as a function of p_{T} intervals of tracks as

$$\varepsilon_{\text{match}}(\Delta p_{\text{T}}) = \frac{N_{\text{tracks}}^{\text{ITS,TPC}}(\Delta p_{\text{T}})}{N_{\text{tracks}}^{\text{TPC}}(\Delta p_{\text{T}})}, \tag{5.18}$$

where $N_{\text{tracks}}^{\text{dets}}$ is the number of tracks with clusters in detectors dets. They are selected requiring impact parameters in the transverse and longitudinal direction of

$|b_{xy}| < 2.4\,\text{cm}$ $|b_z| < 3.2\,\text{cm}$, respectively, within $|\eta| < 0.8$. In addition, for the numerator, at least one cluster in either of the SPD layers is required. An uncertainty arises from potential different efficiencies in data and simulation. Hence, efficiencies are first estimated separately for those and the final relative uncertainty is estimated as

$$\sigma^{\text{rel}}_{\text{match}} = \frac{\varepsilon^{\text{MC}}_{\text{match}} - \varepsilon^{\text{data}}_{\text{match}}}{\varepsilon^{\text{data}}_{\text{match}}}. \tag{5.19}$$

The efficiencies in simulation are estimated separately for primary and secondary tracks in MC since a lower efficiency is expected for secondary tracks which lie either outside of the SPD layers or arise from material interaction. To avoid a bias due to potential different primary and secondary track fractions in data and simulation, the primary track fraction f_{prim} is first estimated from data. Templates of the impact parameter distributions are derived from MC for primary and secondary tracks and then fit to the obtained distributions in data. f_{prim} is then rescaled to the total number of tracks in the TPC since the primary template was derived for tracks with at least one cluster in one of the SPD layers to obtain f'_{prim}. Finally, the MC efficiency is derived according to

$$\varepsilon^{\text{MC}}_{\text{match}} = f'_{\text{prim}} \cdot \varepsilon^{\text{MC}}_{\text{match,prim}} + (1 - f'_{\text{prim}}) \cdot \varepsilon^{\text{MC}}_{\text{match,sec}}. \tag{5.20}$$

The entire procedure has been conducted for each data taking period and year separately. The uncertainties for this analysis were obtained by averaging the single values from the different periods and are of the order of 1 %.

5.4.3 z_{vtx} distribution

The acceptance and efficiency shows a slight dependency on the longitudinal position of the primary vertex z_{vtx}. Analogously to the procedure to account for the discrepancy of the multiplicity distribution in simulation (see Eq. (5.15) in Sec. 5.3.2), weights were derived from the ratios of the z_{vtx} distribution in data and simulation. Those were applied to MC events and the efficiencies were derived for all multiplicity intervals. The relative systematic uncertainty was then taken as the ratio w.r.t. the central result of the efficiencies and values of $\approx 1\,\%$ were estimated.

5.4.4 Raw yield extraction

The raw yield as extracted in Sec. 5.3.1 is potentially sensitive to the chosen fit parameters and hence a particular choice can introduce a bias in the raw yield extraction. To estimated the impact of the fit parameters on the extracted raw yield, a set of variations was defined as summarised in Tab. 5.11. For each variation, the raw yield extraction from the fit in the invariant mass distribution of selected Λ_c^+ candidates was repeated and the root-mean-square (RMS) deviation w.r.t. the mean value was extracted.

The up and down variation of the Gaussian width by 15 % is motivated from the fluctuation of the widths obtained from the unconstrained data and MC fits in the [0 %, 100 %] multiplicity interval (see also Sec. 5.3.1). The fit range was enlarged and shrunk by $\sim 1\,\%$. In order to not overestimate this uncertainty, fits from variations were rejected if the reduced χ^2 was above 2 or if the significance was below 3.

parameter	variations
Gaussian parameters	fixed mean and fixed width
	free mean and fixed sigma
	0.85·nominal width and free mean
	1.15·nominal width and free mean
lower bound fit range [GeV/c]	{2.12, 2.13, 2.14, 2.15, 2.16}
upper bound fit range [GeV/c]	{2.416, 2.426, 2.436, 2.446, 2.456}
background model functions	{exponential and first, second, third order polynomial}

Table 5.11: Variations used for the fits to the invariant mass distribution of selected Λ_c^+ candidates to obtain the systematic uncertainty on the raw yield extraction.

Fig. 5.22 shows the ratio of the mean value obtained from the variations w.r.t. the nominal value for all multiplicity intervals from top-left to bottom-right and it is seen that in all cases the deviation is well below 10 %. To estimate the variation of the yield with the variation of the fit parameters, Fig. 5.23 in addition shows the RMS of the yield variations over the mean value for all intervals. This was finally taken as the relative systematic uncertainty. For all intervals it ranges between 5 % to 8 % except for the highest p_T interval where the uncertainty reaches 16 % in the integrated multiplicity interval.

5.4.5 MC $(\text{Acc} \times \varepsilon)_{\text{prompt}}$

The estimation of the efficiency re-weighting as described in Sec. 5.3.2 via Eq. (5.15) introduces a source for a systematic uncertainty caused by the candidate selection in data. This impact is estimated by loosening the selection criteria to build the weights. In this case, the selection on the Λ_c^+ mass window in data was not applied, otherwise the derivation of the weights followed the exact same procedure as outlined before. As seen in Fig. 5.24 the deviations in all p_T and multiplicity intervals are well below 1 % and were therefore considered negligible.

5.4.6 Selection and PID efficiency

All BDT models contain the $n\sigma$ of the TPC and TOF detectors for the proton candidate track and they are entangled in the candidate selection with the BDT.

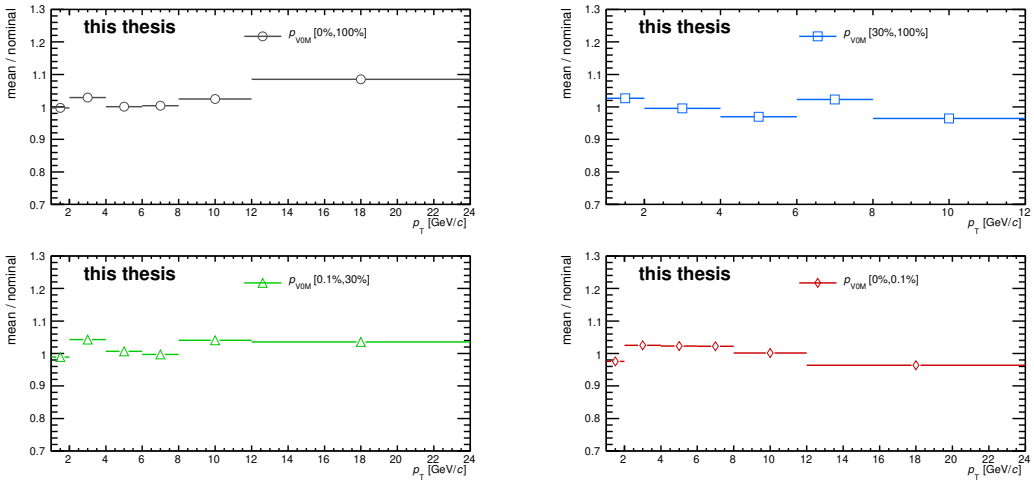


Figure 5.22: Ratio of mean value of extracted raw yield for each variation w.r.t. the corresponding nominal value for all multiplicity and p_{T} intervals.

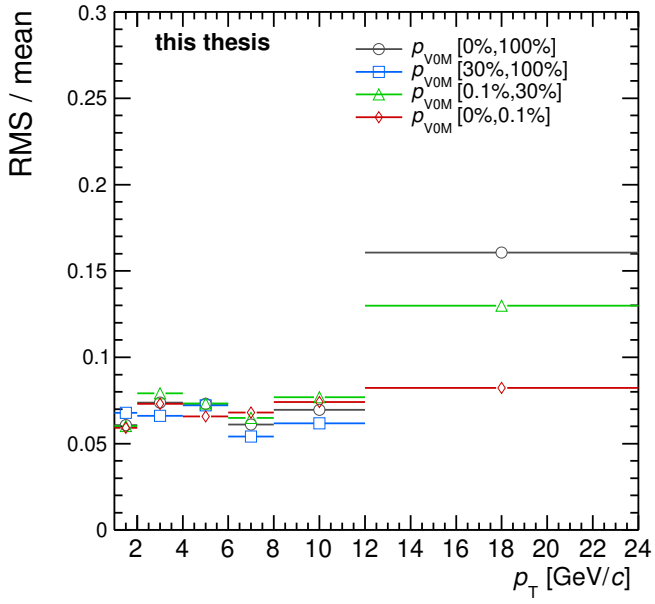


Figure 5.23: RMS over mean of extracted yield variations for all multiplicity and p_{T} intervals.

The PID efficiency uncertainty is therefore contained in the BDT selection efficiency. To estimate it, the BDT working points were varied in 20 equidistant steps within the range of ± 0.2 while a maximum relative deviation of $\pm 25\,\%$ from the nominal signal efficiency was required to avoid large fluctuations which could lead to an overestimation of the uncertainty. The signal efficiency was estimated using MC samples. For each variation the corrected yield was computed and the ratio w.r.t. the

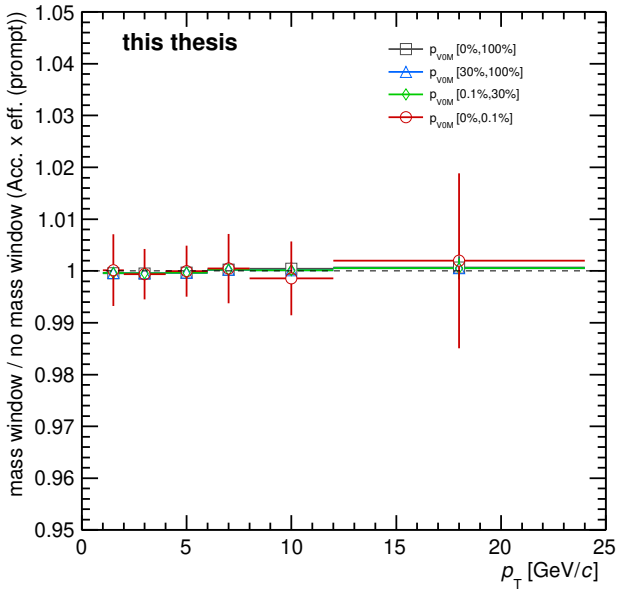


Figure 5.24: Relative deviation of the prompt Λ_c^+ acceptance and efficiency from the nominal value applying varied multiplicity weights.

nominal value in each p_T and multiplicity region was calculated. No significant trend of the ratio as a function of the working point was seen. From that the root-mean-square of the ratios around 1 was determined which is shown for all multiplicity and p_T intervals in Fig. 5.25. In particular in the highest p_T and for the lowest multiplicity interval the deviations are large compared to the other intervals. In order to account for fluctuations caused by statistics, it was decided to apply a conservative smoothing to the deviations per multiplicity interval over the entire p_T range as indicated by the dotted lines in the figure and a conservative estimation was considered. These smoothened values are then taken as the final relative systematic uncertainty.

5.4.7 Feed-down subtraction

Theoretical uncertainties on the f_{prompt} fraction arise from variations of the beauty-quark mass as well as of the renormalisation and factorisation scale in the FONLL calculations. For the nominal results the f_{prompt} fraction was assumed to be the same in all multiplicity intervals. However, to account for a potential difference between B-hadron and prompt Λ_c^+ production depending on the multiplicity, variations of the FD fraction $f_B = 1 - f_{\text{prompt}}$ were considered.

PYTHIA 8 simulations with five different tunes, namely with Monash and the 4Cx tune [116] as well as with three colour reconnection tunes [106] have been investigated. For each tune the ratio f_B^{mult}/f_B in bins of $\mathrm{d}N_{\text{ch}}/\mathrm{d}\eta/\langle \mathrm{d}N_{\text{ch}}/\mathrm{d}\eta \rangle_{\text{PYTHIA8 tune}}$ has been computed as shown in Fig. 5.26. For each single multiplicity interval analysed

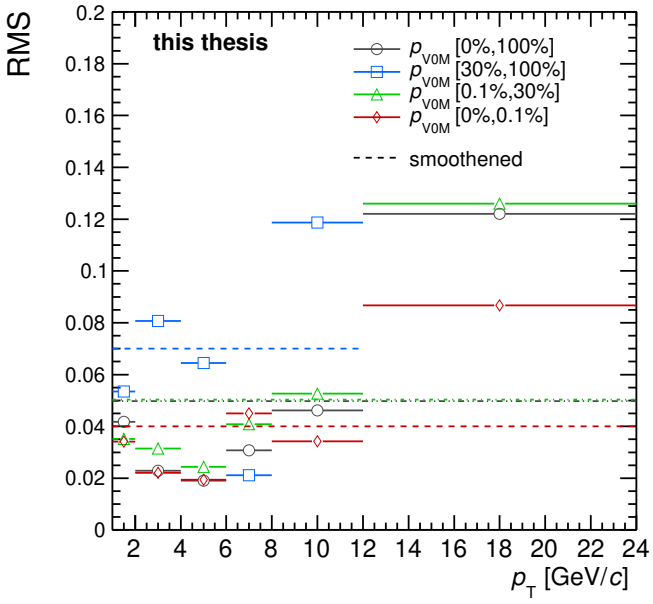


Figure 5.25: *Left:* Signal efficiency of the BDT selection obtained for each tested working point. A maximum variation of 25 % from the nominal efficiency was chosen to vary the BDT working point around its nominal value. *Right:* Ratio of the yield per event for each tested working point relative to the nominal result.

in this work, the ratio of the corresponding mean charged-particle density w.r.t. the overall value of 7.06 in the INEL $>$ 0 is taken. The matching bin in the shown histogram was chosen and the envelope of the tunes around 1 was taken to be the up and down variation of the $f_{\mathrm{B}}^{\mathrm{mult}}/f_{\mathrm{B}}$ fraction. These were used as scale factors $s_{\pm}^{\mathrm{mult}}$ for f_{B} which was first calculated from the f_{prompt} derived in Sec. 5.3.2 as $f_{\mathrm{B}} = 1 - f_{\mathrm{prompt}}$. The up and down scale factors $s_{\pm}^{\mathrm{mult}}$ per multiplicity interval are summarised in Tab. 5.12. The prompt fraction in a given multiplicity interval was then derived as $f_{\mathrm{prompt}}^{\mathrm{mult}} = 1 - s_{\pm}^{\mathrm{mult}} \cdot f_{\mathrm{B}}$ and the relative systematic uncertainty in each interval was computed as the ratio $f_{\mathrm{prompt}}^{\mathrm{mult}}/f_{\mathrm{prompt}}$.

Fig. 5.27 summarises the resulting relative lower and upper variations of the prompt fraction as shown for the lowest to the highest p_{T} interval from the top-left to bottom-right. The shown error bars correspond to the relative systematic uncertainties which are asymmetric due to the procedure of taking the envelope described by the different tunes. The magnitude of the relative uncertainties is of the order of 1 % except for the highest two p_{T} intervals in the highest multiplicity region where it reaches values of -10 % and -13 %.

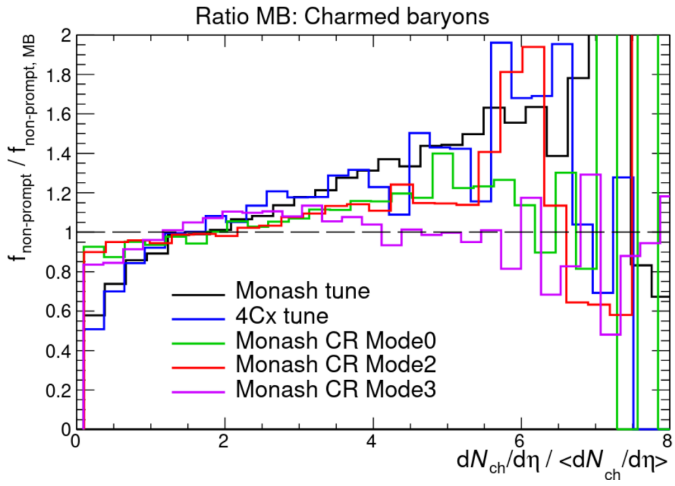


Figure 5.26: Ratio of f_B^{mult}/f_B as a function of the $\mathrm{d}N_{\text{ch}}/\mathrm{d}\eta/\langle \mathrm{d}N_{\text{ch}}/\mathrm{d}\eta \rangle_{\text{PYTHIA8 tune}}$ for different PYTHIA 8 tunes.

p_{V0M} interval	s_-^{mult}	s_+^{mult}
[30 %, 100 %]	0.7	1.0
[0.1 %, 30 %]	0.9	1.2
[0 %, 0.1 %]	1.0	1.4

Table 5.12: Up and down scale factors of the FD fraction f_B for analysed multiplicity intervals.

5.4.8 Trigger efficiency

The uncertainties as estimated in Tab. 5.10 are quoted as systematic uncertainties.

5.4.9 Summary of uncertainties

The relative systematic uncertainties are summarised in Tab. 5.13. The uncertainty on the MC p_T shape is ±2 % in the lowest p_T interval and negligible in all others. The tracking uncertainty has a sizeable effect of 4.5 % in the lowest up to ±7 % in the highest p_T interval.

Across all multiplicity intervals, the systematic uncertainties of the ML working point and the yield extraction are of similar size and vary overall between ±5 % and ±8 %. Only in the highest p_T interval of the [0 %, 100 %] and [0.1 %, 30 %] multiplicity intervals the yield uncertainty even reaches values of ±16 % and ±13 %, respectively.

The following leading contributions arise from the uncertainties on the feed-down with a maximum upper value of +8 % in the highest accessible p_T interval of the [30 %, 100 %] multiplicity region and a minimum lower value of −13 % in the highest p_T interval of the [0 %, 0.1 %] multiplicity region.

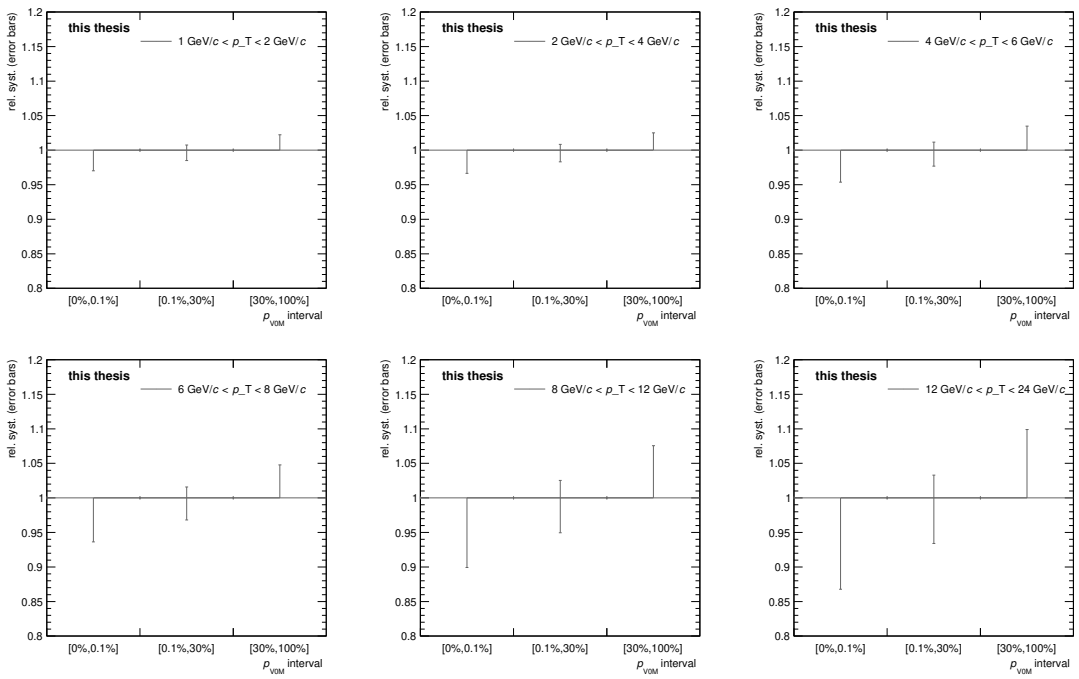


Figure 5.27: Relative systematic uncertainties in differential multiplicity regions for all p_{T} intervals from top-left to bottom-right.

Systematic uncertainties on the trigger and event selection reach a maximum value of $\pm 1.4\,\%$ in the $[30\,\%, 100\,\%]$ multiplicity region and shrink to value of $\pm 0.3\,\%$ and $\pm 0.1\,\%$ in the $[0.1\,\%, 30\,\%]$ and $[0\,\%, 100\,\%]$ multiplicity interval, respectively.

The uncertainty of the efficiency weight was found to be negligible across all p_{T} and multiplicity intervals.

relative systematic uncertainties [%]							
p_{V0M} interval [%]	source	p_{T} interval [GeV/c]					
		[1, 2)	[2, 4)	[4, 6)	[6, 8)	[8, 12)	[12, 24)
[0, 100]	tracking	4.5	5.5	6	7	7	7
	MC p_{T} shape	2	0	0	0	0	0
	trigger	0.3	0.3	0.3	0.3	0.3	0.3
	yield	6	7	7	6	7	16
	efficiency weights	0	0	0	0	0	0
	z_{vtx} position	1.0	0.5	0.5	0.5	0.5	0.5
	working point	5	5	5	5	5	5
	FD	-	-	-	-	-	-
[30, 100]	tracking	4.5	5.5	6	7	7	
	MC p_{T} shape	2	0	0	0	0	
	trigger	1.4	1.4	1.4	1.4	1.4	
	yield	7	7	7	5	6	/
	efficiency weights	0	0	0	0	0	
	z_{vtx} position	1.0	0.5	0.5	0.5	0.5	
	working point	7	7	7	7	7	
	FD	$^{+2}_{-0}$	$^{+2}_{-0}$	$^{+3}_{-0}$	$^{+5}_{-0}$	$^{+8}_{-0}$	
[0.1, 30]	tracking	4.5	5.5	6	7	7	7
	MC p_{T} shape	2	0	0	0	0	0
	trigger	0.1	0.1	0.1	0.1	0.1	0.1
	yield	6	8	7	7	8	13
	efficiency weights	0	0	0	0	0	0
	z_{vtx} position	1.0	0.5	0.5	0.5	0.5	0.5
	working point	5	5	5	5	5	5
	FD	$^{+1}_{-1}$	$^{+2}_{-1}$	$^{+2}_{-1}$	$^{+3}_{-2}$	$^{+5}_{-3}$	$^{+6}_{-3}$
[0, 0.1]	tracking	4.5	5.5	6	7	7	7
	MC p_{T} shape	2	0	0	0	0	0
	trigger	0	0	0	0	0	0
	yield	6	7	7	7	7	8
	efficiency weights	0	0	0	0	0	0
	z_{vtx} position	1.5	1.0	1.0	0.5	0.5	0.5
	working point	4	4	4	4	4	4
	FD	$^{+0}_{-3}$	$^{+0}_{-3}$	$^{+0}_{-4}$	$^{+0}_{-6}$	$^{+0}_{-10}$	$^{+0}_{-13}$

Table 5.13: Summary of relative systematic uncertainties. If only one value is quoted it means a symmetric up and down variation.

5.5 Results

The p_T spectrum of the prompt Λ_c^+ yield per event in $|y| < 0.5$ has been extracted in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ in intervals of the event multiplicity estimated via the V0M percentile p_{V0M}. Statistical as well as systematic uncertainties as estimated in the previous Sec. 5.4 are taken into account and are shown as vertical bars and boxes in the following plots, respectively. The results are shown in Fig. 5.28.

The yield per event is seen to increase with multiplicity, the data of the low multiplicity of [30 %, 100 %] lying below while the measured values of the [0.1 %, 30 %] and [0 %, 0.1 %] intervals lying above the integrated multiplicity interval. This agrees with the expectation that in general a higher (lower) number per particle species is produced with increasing (decreasing) multiplicity since it is expected that the abundance of all particle species scales with multiplicity by the same amount in pp collisions.

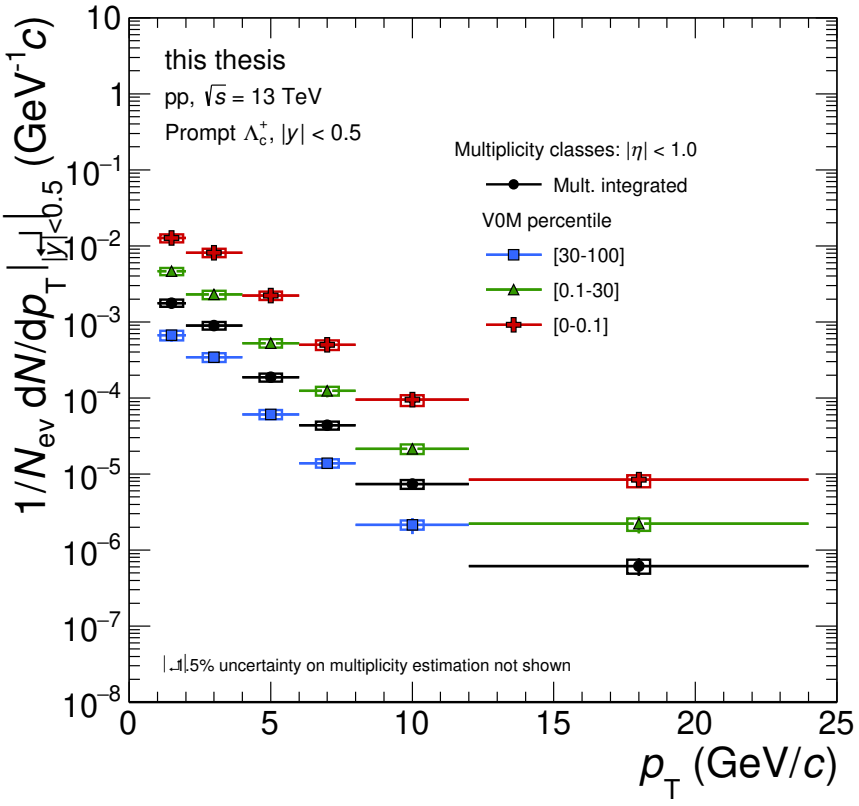


Figure 5.28: Prompt Λ_c^+ yield per event as a function of candidate p_T and in intervals of V0M percentile p_{V0M}. Statistical and systematic uncertainties are indicated by vertical bars and boxes, respectively.

Fig. 5.29 shows the analogous results for the charmed mesons D^0 and D_s^+ as

well as the result obtained from averaging the measurements obtained in this thesis with a measurement exploiting the decay channel $\Lambda_c^+ \to pK^-\pi^+$. The shown ratios below are obtained by dividing the spectra of the multiplicity classes to the result obtained in the integrated sample. The V0M percentile multiplicities were converted to the average charged-particle density $\langle dN_{ch}/d\eta \rangle$ as described in Sec. 5.3. In this ratio, systematic uncertainties related to the multiplicity binning were assumed to be uncorrelated among different multiplicity intervals. The tracking and MC p_T shape uncertainty were assumed to be fully correlated and cancel and for the same reason, the uncertainty on the branching ratio was assumed to cancel. This is due to the fact that those were estimated in the integrated multiplicity interval assumed to be shared among multiplicity intervals as argued in Sec. 5.4. Uncertainties related to the raw yield extraction as well as that of the selection and PID efficiency were assumed to be partially correlated.

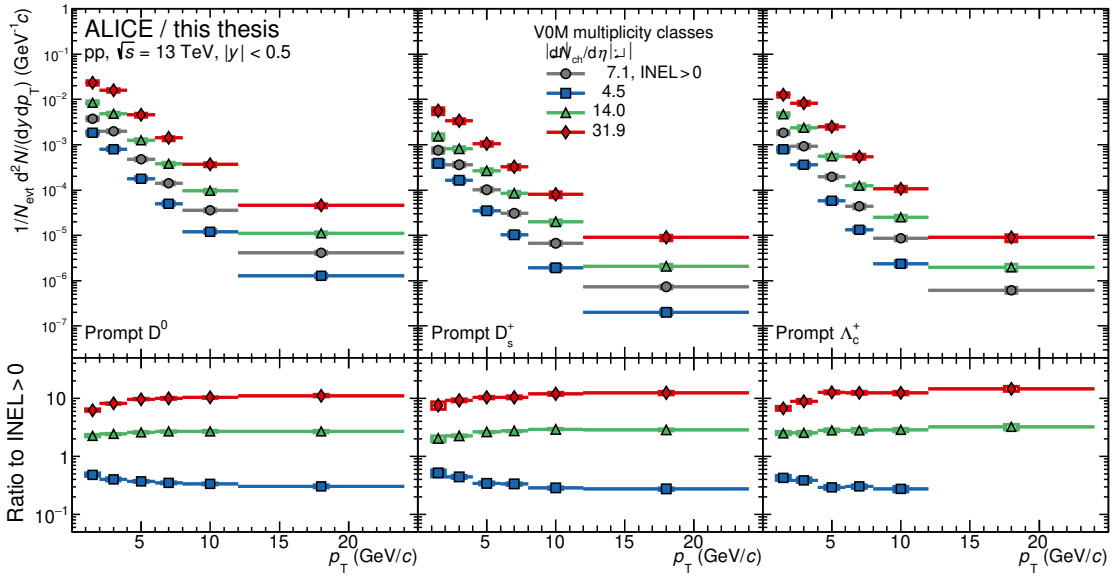


Figure 5.29: Prompt D^0, D_s^+ and Λ_c^+ yield per event as a function of candidate p_T and in intervals of $\langle dN_{ch}/d\eta \rangle$. The Λ_c^+ results were obtained by averaging the results in the channels $\Lambda_c^+ \to pK_S^0 \to p\pi^+\pi^-$ and $\Lambda_c^+ \to pK^-\pi^+$ [155].

The increase of the relative yield with multiplicity is observed for the charmed mesons D^0 and D_s^+ as well. Furthermore, the ratio w.r.t. the integrated interval increases (decreases) for higher (lower) multiplicities. It cannot be concluded from the shown measurements, however, if there might be an indication for the ratios to reach a plateau for $p_T \gtrsim 10\,\text{GeV}/c$ as it was observed for strange and non-strange LF hadrons in [26, 156].

A comparison of the relative yields among hadrons can reveal differences in the production and hence in the underlying hadronisation mechanisms as discussed earlier. Therefore, the ratios D_s^+/D^0 and Λ_c^+/D^0 were derived as shown in Fig. 5.30 for both

multiplicity estimators, using the number of SPD tracklets in the left column and the forward multiplicity estimated by the V0M percentile in the right column.

Within uncertainties, the D_s^+/D^0 ratio is independent of the momentum as well as of the multiplicity for both estimators.

This is different for the ratio Λ_c^+/D^0 which shows a significant p_T dependence as it decreases with increasing momentum. Furthermore, an increasing ratio can be seen with increasing multiplicity. In particular, this observation strongly suggests that the hadronisation into charmed baryons is enhanced compared to charmed mesons which becomes more pronounced at higher particle densities. Therefore, this observation is an indication for mechanisms that increase the baryon production stronger than the meson production. This was already seen in the LF sector as discussed in Sec. 3.4, thus this observation indeed hints at different baryon and meson hadronisation processes in general.

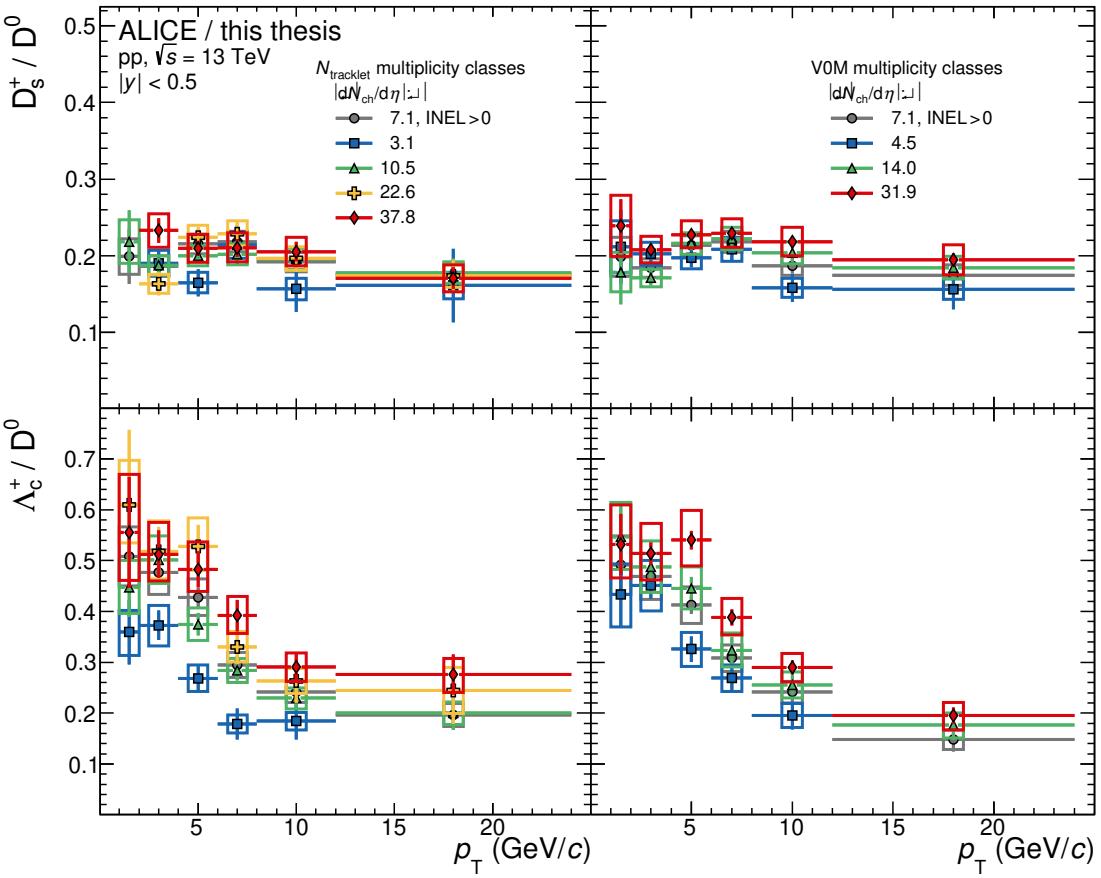


Figure 5.30: Production ratio of prompt D_s^+/D^0 *(top row)* and Λ_c^+/D^0 *(bottom row)* for two different multiplicity estimations, one exploiting the number of SPD tracklets *(left column)* and the other using the V0M percentile *(right column)* [155].

A comparison with the published results of the $(\Lambda + \bar{\Lambda})/(2K_S^0)$ ratio in the LF

sector as measured in [26] is shown in Fig. 5.31, again for both multiplicity estimators. Although the particle densities within each of the multiplicity estimations are not quite the same (for instance looking at the V0M estimation where one sees a comparison between 4.5 and 4.6 as well as 31.9 and 25.5) they are assumed to be comparable. The ratios obtained in the LF and HF sector are compatible within the low and high multiplicity regions for both estimators. However, looking at the V0M estimation in the right panel, the trend with momentum in the LF sector indicates a stronger decrease with p_T.

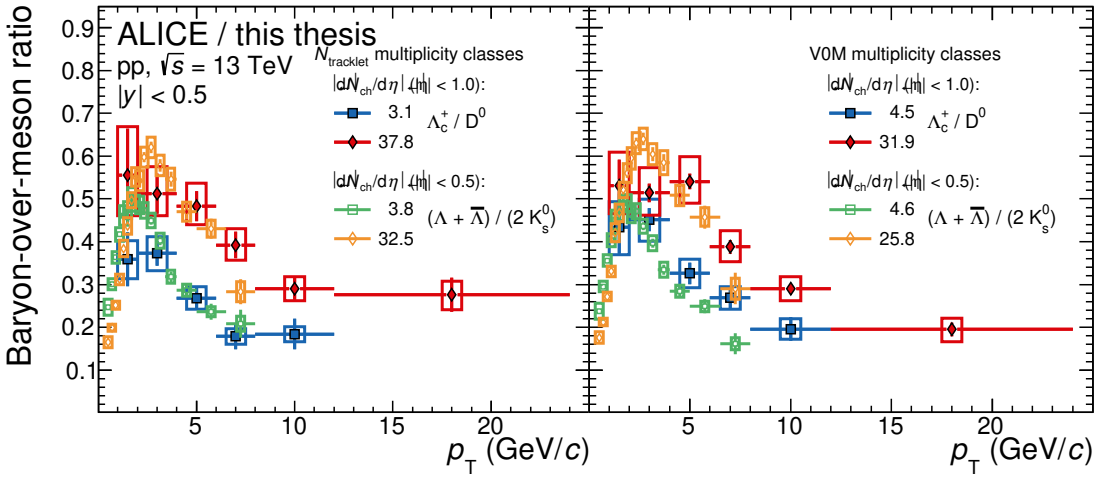


Figure 5.31: Comparison of the HF production ratio Λ_c^+/D^0 and the LF production ratio $(\Lambda + \bar{\Lambda})/(2K_S^0)$ as derived from the measurements in [26]. The respective highest and lowest multiplicity intervals are shown using the number of SPD tracklets *(left)* and the V0M percentile *(right)* for the multiplicity estimation. Statistical and systematic uncertainties are shown as vertical bars and boxes, respectively [155].

Two model predictions are shown in Fig. 5.32 in comparison with data obtained using the SPD tracklet estimator. The highest and lowest multiplicity regions are shown in the left and right column, while the top row shows the data and model predictions for the D_s^+/D^0 ratio and the bottom row contains the Λ_c^+/D^0 comparisons.

Firstly, a statistical hadronisation model [146] (CE-SHM) was used to compare the data to. In the case of the Λ_c^+/D^0 ratio, the underlying charmed-baryon spectrum has been augmented to include additional excited baryon states predicted by the Relativistic Quark Model (RQM) [118], while in the case of the D_s^+/D^0 ratio the spectrum was based on measurements collected by the Particle Data Group [13].

The comparison of the D_s^+/D^0 ratio is in good agreement in the low multiplicity region but tends to overshoot the data in the high multiplicity region. With an increasing resolution, future measurements could help to either support or rule out this model to be able to describe the D_s^+/D^0 ratio.

The comparison to the Λ_c^+/D^0 data shows good agreement in shape and magnitude

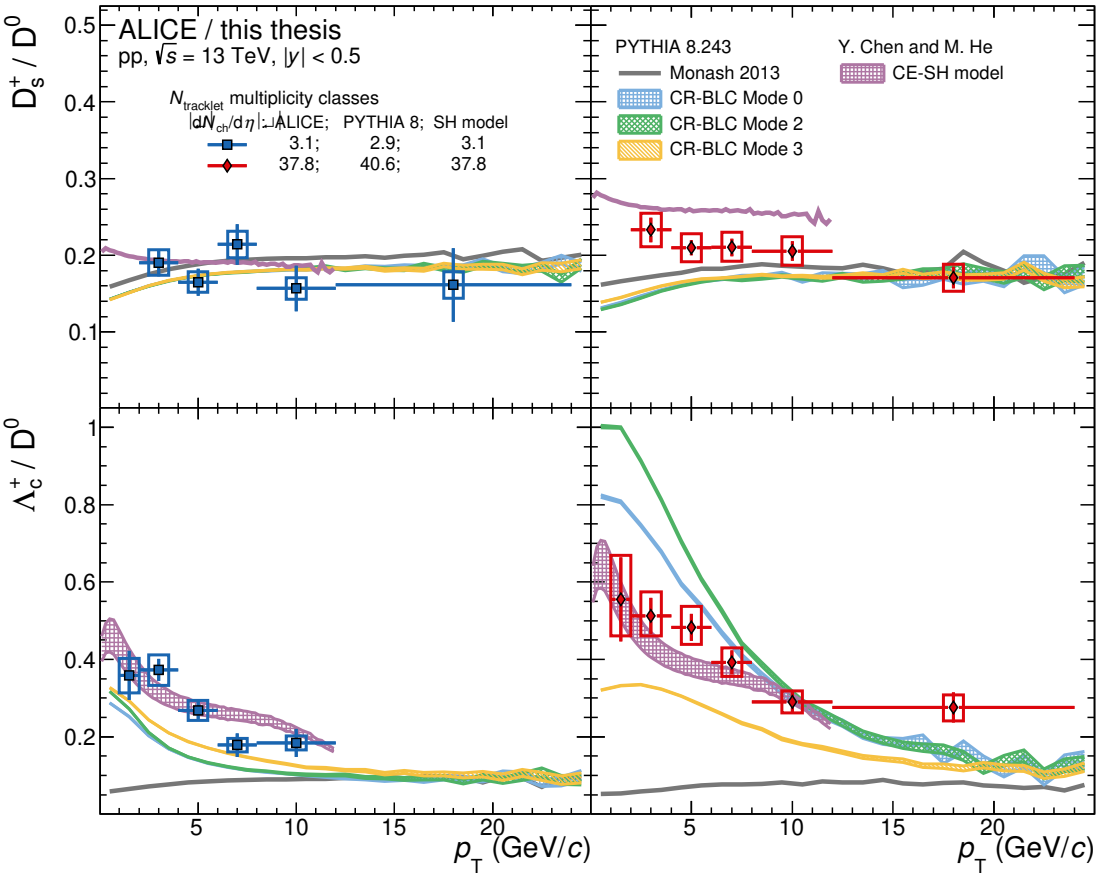


Figure 5.32: Theory predictions for the production ratios of D_s^+/D^0 *(top)* and Λ_c^+/D^0 *(bottom)* in the lowest *(left)* and highest *(right)* multiplicity interval [155]. For further information it is referred to the text.

in both multiplicity intervals. In this case, additional baryonic states drive the Λ_c^+ yield up with multiplicity. But also here, the decrease of uncertainties in the future will help to draw more firm conclusions concerning the compatibility of the model with the data.

Both ratios were also compared to different PYTHIA 8 tunes, namely Monash [116] and three different configurations of the CR-BLC tune [106] implementing colour reconnection mechanisms as introduced in Sec. 3.3.1. The D_s^+/D^0 ratio is described by all tunes equally well.

In the case of the Λ_c^+/D^0 ratio, the prediction of the Monash tune is overall flat as it is tuned to charm fragmentation derived from e^+e^- collisions and can therefore not describe a p_T or multiplicity dependence. The CR-BLC tunes on the other hand develop a similar shape as the data as a function of p_T but are not able to predict the right order of magnitude.

From the discussed observation above it cannot be firmly concluded what the underlying mechanism is that drives the increase of charmed-baryon production. How-

ever, model predictions were presented that are able to qualitatively describe the enhancement with multiplicity seen in Λ_c^+/D^0. On the other hand, the CE-SHM is capable of describing the p_T and multiplicity dependence but fails to get the magnitude of the D_s^+/D^0 ratio at high multiplicity right.

Nonetheless and apart from the theory comparisons, it is a remarkable observation that baryon production is seen to be enhanced compared to meson production in various measurements. It has been observed in the LF sector [25, 26], but also in Λ_c^+ production measurements conducted in p–Pb and Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02\,\text{TeV}$ [115, 157, 158]. Taken these and the measurements discussed in Sec. 3.4 together, there are strong indications to assume hadronisation mechanisms beyond vacuum fragmentation not only in heavy-ion collision systems but also in smaller systems such as pp.

5.6 Updates for the BDT optimisation strategy

Updated MC productions from PYTHIA 8 [94] with Monash tune [116] are available and furthermore, new FONLL predictions based on the latest parameters as extracted by the Particle Data Group [13] have been computed.

When the datasets will become fully available, an updated BDT fitting and optimisation procedure can be investigated in view of the expected paper publication. This procedure is described and briefly discussed w.r.t. the previous approach used in the current analysis as presented in Sec. 5.2.4.

The updated procedure presented in the following is not expected to improve the precision of the results significantly, but it should be tested and can be implemented to ensure more rigorous and reproducible BDT fitting and optimisation procedures.

5.6.1 Feature selection

The feature selection can be revised by exploiting a metric to estimate the separation potential of each feature before fitting a BDT. This can be done by defining a metric according to

$$p_{\text{sep}} = \frac{1}{2} \sum_{i}^{n_{\text{bins}}} \frac{(s_i - b_i)^2}{s_i + b_i}, \tag{5.21}$$

where s_i and b_i are the bin contents of signal and background, respectively, in bin i. For a given feature, $p_{\text{sep}} \in [0, 1]$ tends to larger values the more the shapes of signal and background candidates differ. Note, however, that this procedure only provides an initial estimate since p_{sep} depends on the choice of the binning. In the extreme case of only one bin, the separation potential would be computed to be 0 whereas for the number of bins approaching infinity the value would tend towards 1.

Fig. 5.33 shows the same features as used before in Sec. 5.2.4 for the lowest p_T interval but now the value of the separation metric is indicated as well. For instance,

the impact parameters of the V^0 decay tracks as well as the Armenteros-Podolanski value in this case show a comparably low separation potential w.r.t. other features. This observation is also in agreement with the feature importance obtained after the fitting (see Fig. 5.11) where these features were rated comparably low w.r.t. others.

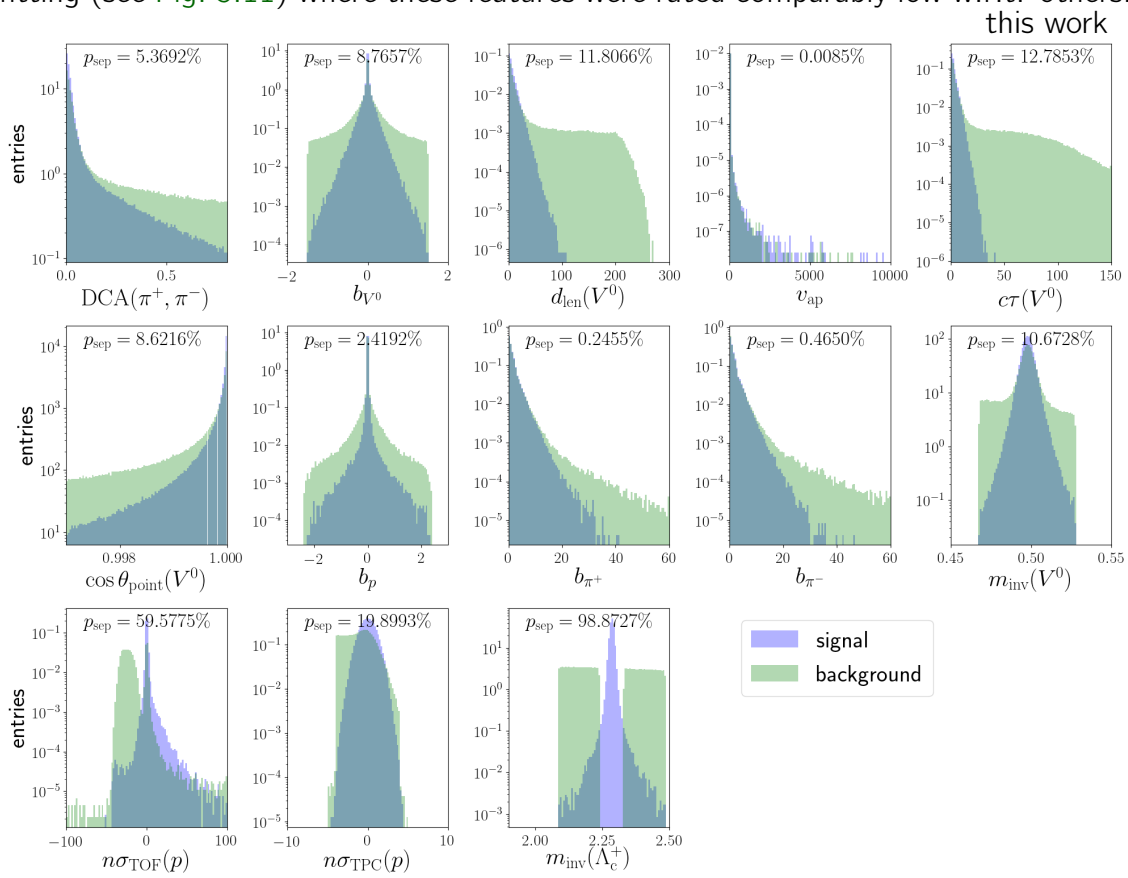


Figure 5.33: Feature distributions of signal (blue) and background (green) samples in the 1 GeV/c to 2 GeV/c momentum interval. Note, that the single histograms are not stacked but that the colours overlap in certain regions to identify the different shapes of the signal and background histograms. In addition to the feature distributions the invariant mass distribution of the Λ_c^+ candidates is shown as well.

Hence, an estimate on the separation potential supports the finding of features from the very beginning.

5.6.2 Hyperparameter optimisation

In contrast to the grid search conducted to find suitable hyperparameters, the hyperparameters optimisation could profit from a Bayesian optimisation approach for instance exploiting packages such as [159]. Rather than scanning a fixed set of hyperparameters independent of the performance of each configuration, a Bayesian

approach draws a new set of hyperparameters for the next evaluation step based on the performance of the previous iteration. This also allows for a significantly larger space of hyperparameter values to be scanned.

The optimisation procedure could follow what is sketched in Fig. 5.34. The training and test sets are combined, shuffled and subsequently, at the beginning of each iteration, split into k disjoint sets. $k - n$ sets are combined to form the new training set and the left n sets build the test set. For each possible combination of forming the training and test sets the performance is evaluated and after the last iteration the average score is built. This cross validation (CV) approach is an additional measure to ensure the generalisation capabilities since for all splits the performance is expected to be compatible for a suitably fitted BDT. In case large fluctuations of the performance are observed for different splits it has to be investigated if the model configuration was to complex or whether the training and test set contain artefacts in which case the choice of features or the sample preparation might need to be revised.

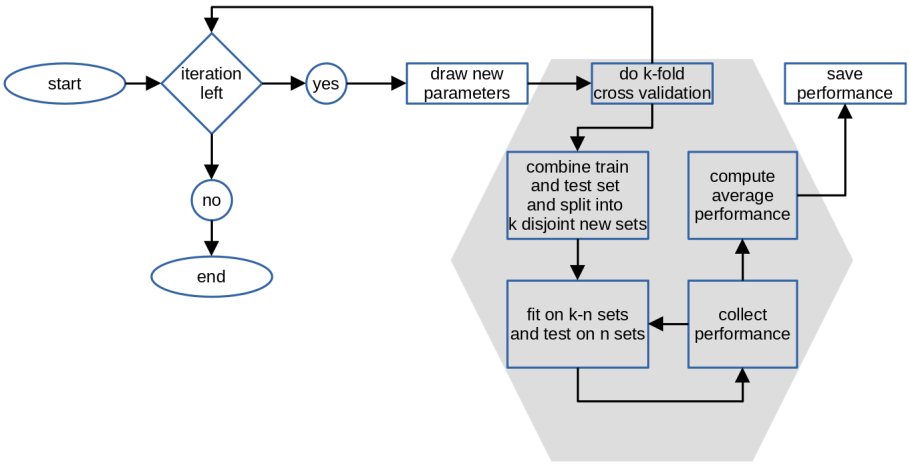


Figure 5.34: Workflow of Bayesian optimisation combined with a cross validation approach.

Summarised in Tab. 5.14 is an enlarged set of possible hyperparameters to be investigated in the case of XGBoost [151]. In addition to the previously conducted grid search it now also contains regularisation parameters such as γ, λ and α. Furthermore, all available $f_{\text{col,i}}$ parameters can be taken into account whereas before only $f_{\text{col,tree}}$ was used (see also the beginning of Sec. 5.2.4 for the principal of BDT fitting). These parameters work cumulatively, for instance if each one is 1/2, there is the fraction of 1/8 of all features to choose from on a given node. In a first attempt, the AUC can be chosen as the performance metric and in addition, a maximum relative deviation between the AUC values evaluated on the training and test set can be requested to once more ensure a good generalisation capability of each model.

parameter	description
h_{min}	minimum sum of instance weights required in each leaf to continue further branching
d_{max}	maximum number of consecutive splits
γ	minimum loss reduction required to split again at a given node
f_{samples}	ratio of samples to be used for fitting which is sampled once per boosting step
$f_{\text{col,tree}}$	ratio of features used for fitting which is sampled once per boosting step
$f_{\text{col,level}}$	ratio of features to choose from at each new level of depth
$f_{\text{col,node}}$	ratio of features to choose from which is sampled for each new node
η	learning rate
$n_{\text{estimators}}$	number of single fitted trees per BDT
λ	regularisation term as given in Eq. (5.9)
α	regularisation term analogue to λ, however, penalising with $\alpha \sum_i \lvert\omega_i\rvert$

Table 5.14: Hyperparameters to be optimised for the BDT configuration.

5.6.3 Feature importance evaluation

Different from the overall feature importance as it was evaluated for the models described in Sec. 5.2.4, the SHAP package [160] can be employed. In Fig. 5.35 the importance of each feature for the decision making of the BDT is shown for the lowest p_{T} interval. Features are on the vertical axis and from top to bottom the feature importance decreases. Each training sample is evaluated separately and the SHAP value is given on the horizontal axis: Higher values indicate that a feature of a given sample had a positive impact on the signal-like classification while lower values correspond to a higher impact on the background classification. The size of feature values is indicated by a colour scale from red to blue, red indicating high values and blue low values. For instance, large cosine values of the pointing angle $\cos\theta_{\text{p}}$ enhance the signal-like classification.

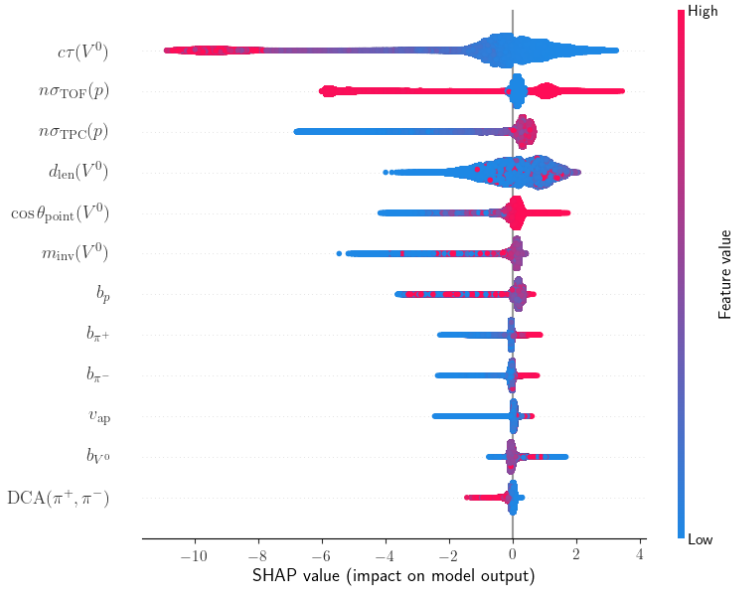


Figure 5.35: SHAP importance of features and their impact on the signal- and background-like classification of the 1 GeV/c to 2 GeV/c BDT model.

6 Simulation developments for LHC Run 3 and 4

As motivated in the beginning of this work, this chapter presents the extensions of the Virtual Monte Carlo (VMC) framework in order to be capable of *partitioning* the event simulation among multiple different simulation engines. In that way, it will become possible to run a full simulation where required but to dispatch particles to a fast-simulation engine where feasible.

The further structure of this chapter is organised as follows: In the next section the working principle of full-simulation MC engines is briefly presented. The basic building blocks of the implementation of such engines are introduced in Secs. 6.2 and 6.3. Afterwards, Sec. 6.4 contains details of how the detector simulation is embedded in the ALICE software stack. In particular, the VMC framework will be introduced as it devises the common interface layer to transport codes in ALICE. Finally, the work of how the VMC framework has been extended to provide the event-partitioning among multiple engines is presented in detail in Sec. 6.5 as the topic of this thesis.

6.1 The foundations of Monte Carlo detector simulation

The smallest granularity in MC detector simulations is a *step* and the basic procedure is described in the following. For a given particle transported in the material i, first the mean-free path λ_j for a given process j is computed according to

$$\lambda_j = \left(\sum_i n_i \cdot \sigma_j \left(Z_i, E \right) \right), \tag{6.1}$$

where n_i denotes the material density and σ_j is the process' cross section given the atomic number Z_i of material i and the particle energy E. The mean-free path is used to compute the no-interaction probability over a path length Δx for process j as

$$P_j(\Delta x) = \exp\left(- \int_{x_0}^{x_1} \frac{\mathrm{d}x}{\lambda_j(x)} \right). \tag{6.2}$$

This is then used to sample the number of mean-free paths n_{λ_j} of the particle for process j according to

$$-\ln P(\Delta x) = \int_{x_0}^{x_1} \frac{\mathrm{d}x}{\lambda_j(x)} = n_{\lambda_j}. \tag{6.3}$$

These calculations are conducted for all possible processes j and for each one of those the step length $s(x)$ is derived as

$$s_j(x) = n_{\lambda_j} \cdot \lambda_j(x). \tag{6.4}$$

The process with the shortest step length $s_j(x)$ is invoked and the actual computation of the resulting kinematics, energy loss, potential radiation or decay is conducted.

All primary particles as well as secondary particles produced during the transport are processed until they reach the boundary of the overall simulated volume, the so-called *world volume*. The collection of steps of the same particle will be referred to as a *track* in the following. Note therefore, that this is not the same as a reconstructed track on the level of a detector experiment. In the case of a simulated pp or Pb–Pb interaction this requires to transport 1000 up to 10000 primary tracks per event, respectively, as well as all produced secondary tracks.

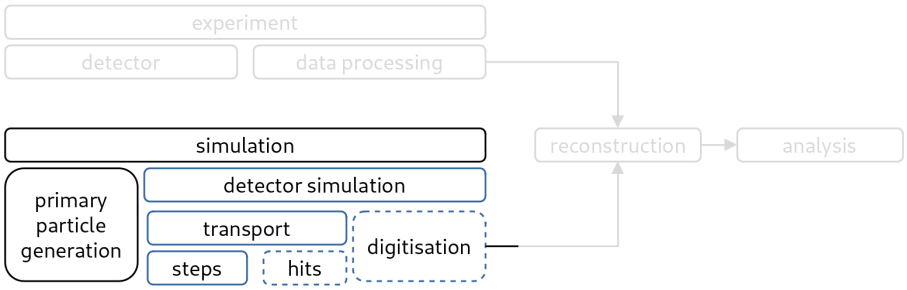


Figure 6.1: Sketch of stages during the simulation chain shown on the bottom. The experimental data taking and processing stages at the top are not shown in further detail. After digitisation both simulated and experimental data is treated the same in view of reconstruction and analysis.

In addition to the transport, there is a chain of further stages involved as depicted in Fig. 6.1. What is shown in black boxes is not discussed in further detail while the blue boxes show individual stages of the detector simulation. Note therefore, that the focus lies on the simulation, the experimental data taking and processing stages have been simplified to a minimum. Stages in the solid blue boxes are taken care of by the detector simulation engine and the dashed ones require additional user implementations.

Each step in a *sensitive* volume might potentially leave a *hit*. Sensitive volumes are sensor parts of a detector where a readout signal maybe induced as a consequence

to the interaction with traversing particles. A hit is produced by the interaction with the sensitive material leading to a certain amount of deposited energy induced by the interaction. Hits are afterwards run through a *digitisation* procedure which models the electronic amplification and readout process. Digitised signals are used to build up the raw event data which, at this point, has the same structure as events taken with the physical detector experiment. The following stages such as reconstruction and potentially the treatment as part of a physics analysis are common to both the simulated and experimental data.

6.2 Central building blocks of a detector simulation

In the GEANT4 documentation [161] two principal building blocks are identified, namely (i) the part of defining and modelling the detector geometry and its response as well as (ii) the tracking and physics simulation. In the following, these conventions from GEANT4 will be used to briefly explain the basic building blocks of detector simulation transport codes. Their principles apply equally well to GEANT3 [34] or FLUKA [32, 33].

6.2.1 Detector definition and response

First of all, the detector definition and response comprises the description of the detector geometry by providing the implementations for *physical* detector volumes. Those are the objects corresponding to the different parts of the real detector. As shown in Fig. 6.2, a *logical* volume is associated to each physical one. A logical volume is an aggregation of a specific *shape* and a *material* defining the physics properties. Shapes and materials can be shared among multiple logical volumes and in turn, the same logical volumes can underlie different physical volumes. In addition, a *sensitive detector* can be associated to logical volumes to model the hit creation and therefore the detector response. The same sensitive detector can be assigned to multiple logical volumes.

The geometry is defined inside a so-called *world volume* which defines the overall volume and coordinate system, hence when tracks reach its boundaries their transport is considered to be complete. Physical volumes are placed at specific coordinates relative to the world volume coordinates and might be rotated accordingly. In the same way physical volumes are placed inside the world volume, other physical volumes can be placed *inside* another physical volume which results in a complex geometry hierarchy in the case of the simulation of large detector experiments such as those at the LHC.

A *navigator* implements the functionality to determine physical volumes corresponding to a particle's global coordinates. Usually, a navigator finds the deepest volume in a hierarchy. Once found, the navigator also provides information about distances to volume boundaries as well as neighbouring boundaries for both those

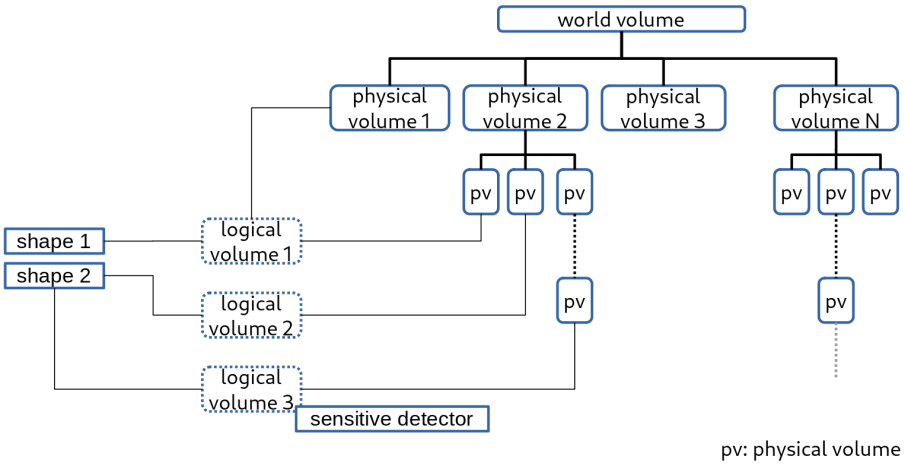


Figure 6.2: Sketch of a geometry tree. Physical volumes build a hierarchy with the world volume as the physical volume containing all other physical volumes. Each physical volume aggregates a logical volume which is defined by a shape and material. A sensitive detector can be associated with a logical volume to register hits.

spatially next to the current volume and those next to it in the volume hierarchy. As such, navigation is a core part in view of the particle transport itself which will be discussed in the next section.

6.2.2 Tracking and physics

The physics simulation is done within an event loop for a certain number of events as sketched in Fig. 6.3. For each event, generated primaries are transported through the detector geometry. Primaries can be provided by different sources. For instance, a user can explicitly define specific particles in order to conduct detector studies. On the other hand, to study the physics at collider experiments, primary particles are usually provided by MC event generators such as PYTHIA 8 [94], HERWIG7 [95, 96] or SHERPA [97, 98] which devise the quantum-field-theoretical calculations during the collisions to compute the produced hadron and lepton kinematics to be fed to the detector simulation.

Primary and secondary particles are stored on a *particle stack* which is used by the engine to push to and pop from.

The physics modelling can be based on different descriptions. For instance, GEANT4 uses so-called *physics lists* providing electromagnetic and hadronic cross sections for various processes such as bremsstrahlung, photon conversion, particle decays, nuclear interactions and so forth. The modelling of certain processes can differ among those physics lists, for instance w.r.t. accuracy, depending on the specific physics a user is interested in. Therefore, before the transport is initiated, the

physics to be modelled has to be specified.

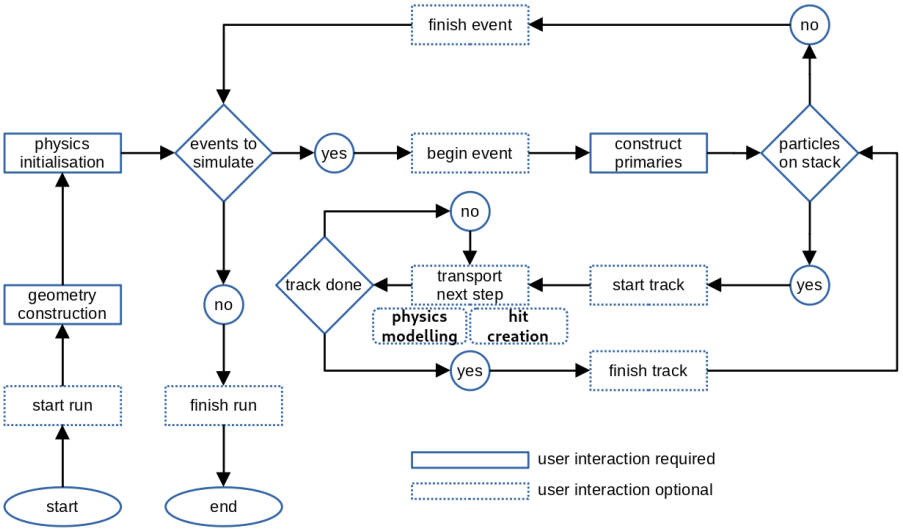


Figure 6.3: Sketch of the general workflow of a detector simulation run.

The boxes in Fig. 6.3 indicate additional implemented user routines. Apart from the hit creation, they can be injected before and after tracks are transported, during the interaction of the engines with the particle stack or before and after each event. As such, these user routines allow for the interaction with the software framework the simulation is embedded in. Some of the routines are required, for instance in order to define the geometry or provide primary particle at the beginning of each event.

In addition to the hits, usually the MC *truth* information is stored such that a reconstructed track can be associated with the actual simulated particle. Such information can be used to compute acceptance and efficiencies of detector experiments usually required for physics analyses (see for instance Sec. 5.3.2).

Simulations which follow the above principle of a step-by-step transport will also be referred to as *full-simulations* hereafter. Due to the increasing demand on physics simulation during the upcoming LHC runs, so-called *fast-simulation* approaches are investigated by the LHC experiments. The next section thus provides a brief overview of fast-simulation solutions.

6.3 Brief introduction of detector fast-simulation approaches used at the LHC

The term *fast-simulation* in the following is meant to refer to the principle of using techniques that significantly speed-up the detector simulation. Usually, these performance gains come at the cost of accuracy due to less detailed modelling of the

underlying physics or using a simplified detector geometry. Typically, this may be achieved through parametrisations or approaches based on selecting outcomes from previously generated and finite lookup-tables. Another possible way is provided by the application of specific machine-learning algorithms to fast produce corresponding energy deposits, hits or digits. A general integration of fast-simulation engines might follow a flow as depicted in Fig. 6.4 where the overall simulation is a mixture of full- and fast-simulation and it is dispatched to the latter under certain conditions. Those conditions could depend on the geometry, particle type or phase space and the user decides in which case speed can or should be favoured over accuracy.

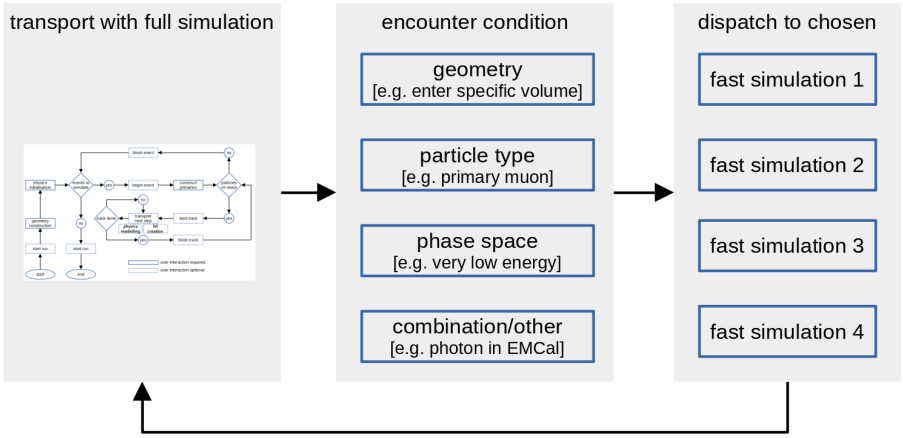


Figure 6.4: Mixing full- and fast-simulation. Dispatch to fast-simulation if certain user conditions are met.

In the following, different concrete approaches are presented. The first one is more general and based on simplified geometry and parametrisation approaches and can be used for different collider experiments. The others have been specifically developed dedicated to certain collider experiments.

A generic parametrisation approach is chosen by the DELPHES framework [162]. It assumes a detector configuration of tracking volumes surrounding the beam pipe followed by a calorimeter system. Long-lived particles are directly given as tracks traversing the tracking volumes and a user-defined momentum smearing is applied to the track as well as a reconstruction probability. Hence, no dedicated step-by-step simulation is done at this point. Neutral particles such as photons are hence modelled as straight paths reaching the calorimetry. Usually it is made of an electromagnetic (EMCal) followed by a hadronic (HCal) calorimeter segmented in φ and η. The cells of the EMCal and HCal are perfectly aligned in the radial direction. Photons and electrons are considered to deposit all their energy in the EMCal. Long-lived hadrons on the other hand deposit their energy in the HCal (kaons and Λ particles are assumed to have a share between the EMCal and the HCal). Muons and neutrinos are immediately propagated without leaving any energy deposit. The sum of deposited energy from all particles reaching a calorimeter cell is finally summed up. DELPHES

also implements a reconstruction procedure and hence skips the detailed hit creation and digitisation.

As another example, a dedicated fast-simulation development used in the CMS experiment combines a simplified geometry of the tracking detector together with simplified treatment of the energy deposited in the calorimetry [163, 164]. They report an acceleration of the event simulation by a factor up to 100 compared to a GEANT4-based full simulation. Charged-particle tracks leave a "simulated hit" in each traversed tracking layer which is computed to be a "reconstructed hit" with a certain efficiency. The energy deposit of particles in the calorimeters is then parametrised.

In the case of ATLAS, the calorimeter simulation took around 75 % of the full simulation time in previous software while the simulation itself took around 34 % of the overall CPU time [165] needed by the simulation. Since the projected resource demands for simulation are significant, ATLAS aims to replace the calorimeter shower simulation with a parametrised fast-simulation approach which shall also involve Machine-Learning [166]. According to the ATLAS collaboration it might be possible to even implement a fast tracking procedure to reach an even higher increase in speed by using a simplified geometry as well as parametrisation [167]. At the same time, ATLAS has developed a fast-simulation chain [168] in order to modularise the single steps of transport, digitisation and reconstruction such that at each individual stage a fast or full simulation approach can be chosen as indicated in Fig. 6.5. The green module depicts the Integrated Simulation Facility (ISF) [169] which, based on specific conditions, can combine different full and fast simulation approaches chosen for the particle transport.

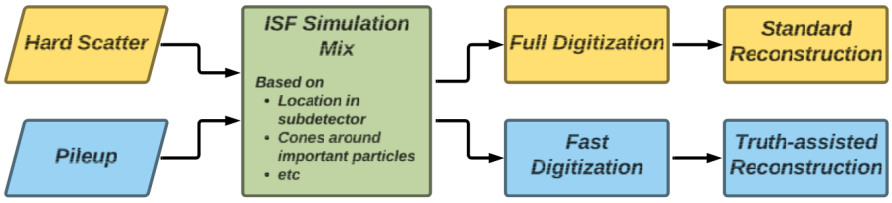


Figure 6.5: Fast simulation chain developed by ATLAS, adapted from [168].

A faster simulation procedure can also be achieved by other means, for instance by simplifying the detector geometry or by so-called embedding techniques. The latter can be approached by

1. re-use simulated background and inject specific signal simulation, either during primary event generation or by merging digitised data,

2. inject signal simulation into an event taken from experimental data.

A specific implementation called *ReDecay* has been developed by the LHCb collaboration [170] and an increase in speed of one order of magnitude, while conserving

the accuracy of full simulation, is reported. This approach is based on separating the particles participating in the signal of interest from all other particles. As an example, candidate signal processes as studied in the above mentioned reference are $D^0 \to K^-\pi^+$ or $B^0 \to D^{\star-}\tau^+\nu_\tau$. In such a case, the full event is simulated but the signal particles are removed and only the rest of the event is transported. This is followed by only simulating and transporting the signal N times taking the original vertex into account which yields the "re-decay" of the signal particle of interest and is embedded into the previously simulated background.

Fast-simulation approaches have not yet been used by ALICE in large productions, however, different ones are being tested. At the same time, it will be discussed, how fast-simulations can be interfaced with full-simulations via the extension of the VMC [38] package which is used by ALICE to interface to transport codes. These developments, which follow an analogous approach as the ATLAS ISF framework [169] does, are the topic of this work and will be presented in Sec. 6.5.

6.4 The Virtual Monte Carlo framework and detector simulation in ALICE

6.4.1 Brief introduction to the ALICE Online–Offline framework

For the upcoming LHC runs, the ALICE detector underwent an extensive upgrade programme as briefly outlined in Sec. 4.2.9. At the same time, an entirely new computing model (see [31] for the corresponding technical design report) is being deployed which exploits the WLCG computing GRID as well as a newly installed Online–Offline (O^2) computing facility which is located at LHC Point 2. The latter will be in charge of data compression and the online reconstruction and calibration. The facility will have to cope with up to $3.5\,\mathrm{TB\,s^{-1}}$ with which the data is shipped from the detector to the First Level Processors (FLPs) in Computing Room 1 (CR1). After collecting the data from all detectors and after a first reduction, it is further transferred to the Event Processing Nodes (EPNs) located at CR0 on the surface of Point 2 with $625\,\mathrm{GB\,s^{-1}}$. In addition to the real-time processing of the data during the synchronous (online) stage, the EPNs run in the asynchronous stage (offline), for instance when there is no beam, to perform additional and more complex processing such as the full event reconstruction. The data is finally organised into Analysis Data Objects (AODs) and transferred to the Analysis Facilities (AFs) where they are ready to be used for physics analyses.

In the same way the online and offline data processing is bound tightly together, the newly developed O^2 software framework will be put in place combining all necessary components to handle all tasks within one common framework. The basic building block is made of general libraries such as ROOT [171], the VMC framework [38], GEANT4 [35–37] or build tools such as CMake [172]. In addition, O^2

relies on ALFA [173] and FairRoot [174]. ALFA is a joint effort of the ALICE and FAIR collaborations which in particular devises the common implementation for the multi-processing approach aimed for by ALICE. This approach has been favoured by ALICE to meet the high-throughput parallelisation and offers high flexibility for heterogeneous computing. At the same time the system is fully capable of handling multi-threaded execution per process if required. The communication among processes is realised based on FairMQ [175] which is also a part of ALFA. It provides asynchronous, message-based communication between processes and is designed for large-scale systems but also as lightweight as to support small infrastructures such as personal computers and laptops.

The utilisation of the VMC detector simulation framework is provided via FairRoot. In addition, it provides the base implementation of the simulated detector geometry defining additional properties of volumes or sensitive detectors but also implements the interface class `FairMCApplication` the O^2 simulation builds upon which devises the steering and handling of the detector simulation. The embedding of the VMC framework into O^2 is further described in the next section.

6.4.2 Introduction to the Virtual Monte Carlo framework

The VMC [38] package is designed as a C++ interface library separating the concrete implementation of detector simulation transport code from the user code. In the following, the VMC package and its workflow is discussed assuming only one engine to be used. The extensions introduced in Sec. 6.5 are shown to conserve this logic but extend it to the possibility of having multiple engines running at the same time.

At its core the VMC package consists of three purely virtual classes, namely `TVirtualMCApplication`, `TVirtualMCStack` and `TVirtualMC`. The class dependency is shown in Fig. 6.6. Realisations of `TVirtualMC` devise the interface to concrete simulation engines and in that way, changes in the engine code or its native API are entirely hidden from the user. It aggregates an instance of `TVirtualMCStack` to abstract the internal stacking procedure of the engine code. An implementation of `TVirtualMCApplication` in turn aggregates `TVirtualMC` as well as `TVirtualMCStack` and builds the principal interface with the user code.

Realisations of the engine interfaces `TEngine` are maintained by the developers. For instance, the GEANT interfaces and the code of the abstract VMC package can be found at [176] while the documentation is available at [177]. On the other hand, the implementations of `TVirtualMCApplication` and `TVirtualMCStack` have to be provided by the user to be tailored to specific needs. In particular, the realisations require the implementation of various pure virtual methods.

The workflow of the VMC package follows that introduced in Sec. 6.2.2 (see also Fig. 6.3). All user routines are implemented in the realisation of `TVirtualMCApplication` such that they can be called by `TVirtualMC`. For instance, the implementation of `TVirtualMCApplication::Stepping()` would allow for the hit construction based on sin-

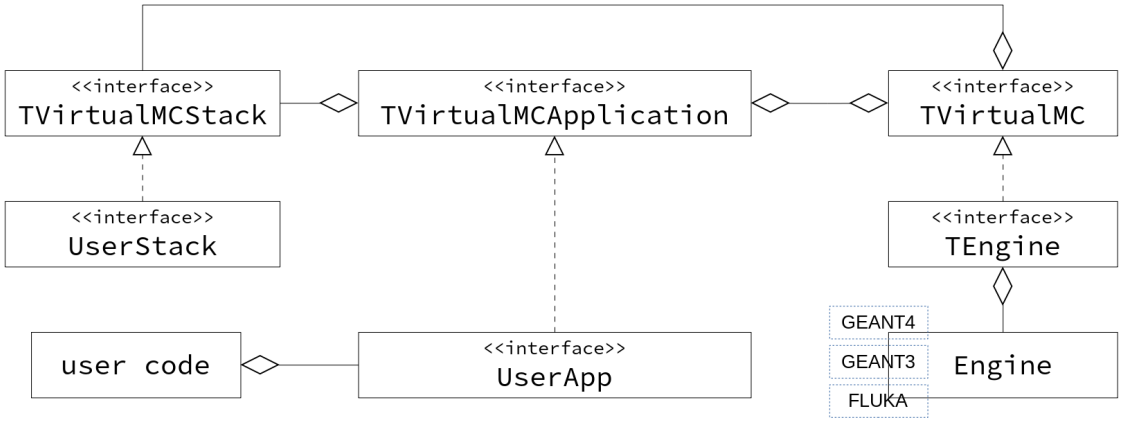


Figure 6.6: Class diagram of the core VMC classes and their realisations.

gle steps.

The ALICE O^2 software framework utilises the VMC package via the FairRoot software as mentioned in the previous section and sketched in the diagram in Fig. 6.7. It does so by further specialising the interface class `FairMCApplication` and by implementing its own stack interface based on `FairGenericStack`.

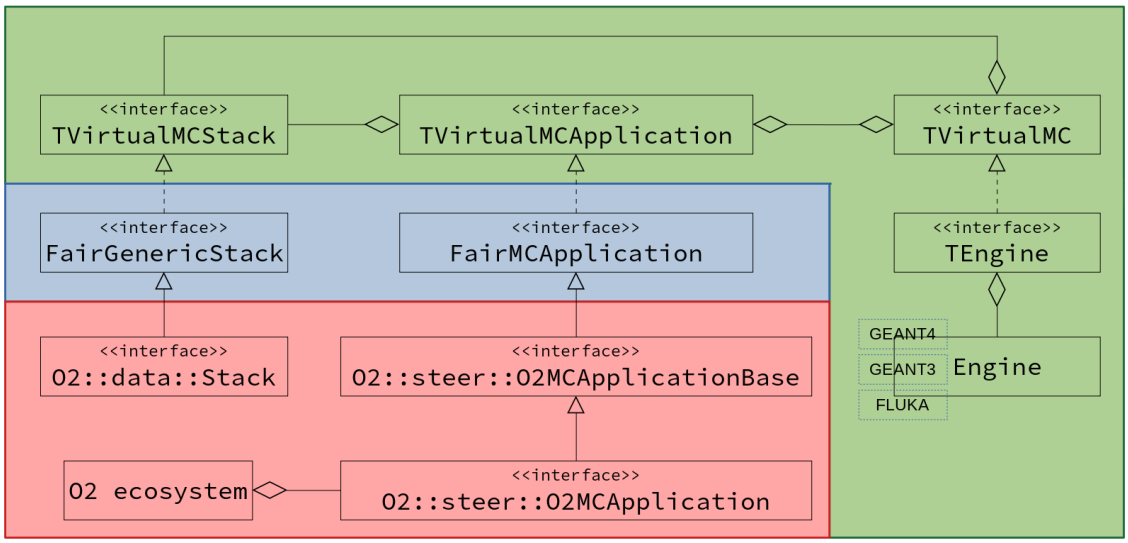


Figure 6.7: Class diagram of the VMC implementation in the O^2 software framework.

As mentioned in the previous section, the O^2 software framework is designed to support heavy multi-processing. Thus, the `O2MCApplication` exploits the features of FairMQ in order to run simulation processes concurrently and asynchronously within multiple processes. The resulting output from all processes is then merged afterwards.

The modular and abstract design of the VMC package hence allows for either

choosing a different engine for different simulation runs or to use the same engine for different implementations of `TVirtualMCApplication` and `TVirtualMCStack` in a transparent manner. In addition, new developed engines can be used without changes of the user defined classes or of the user code itself which offers large flexibility. All engine interfaces support the geometry modelling provided by ROOT which adds another level of abstraction so that only one geometry definition is required. Therefore, it can also be made use of ROOT's integrated geometry navigation. Specifically the aforementioned properties were exploited in the developments presented in this thesis.

6.5 Virtual Monte Carlo developments

In the following, the extensions and modifications of the VMC framework will be presented to enable the possibility of partitioning the event simulation among multiple different engines. The new central class `TMCManager` has been introduced providing the communication and management between the engines and offering the necessary user interfaces to steer the run.

The structure of the implementations will be sketched including short listings which provide insight into critical parts of the implementation. It should give a clear idea of how the integration into the user code can be achieved. For detailed explanations, the reader is referred to the code repositories and documentation at [176, 177].

The VMC developments where accompanied by the requirement to ensure *backward-compatibility*. This was done to avoid breaking with already existing realisations of `TVirtualMCApplication` and `TVirtualMCStack`. Therefore, a seamless transition was guaranteed in particular for FAIR and ALICE software.

Furthermore, the developments had a computing model in mind in which the focus lies on a multi-processing approach shared by FAIR and ALICE via the usage of the FairMQ framework.

The core developments are summarised in the following section.

6.5.1 Core developments

The developments described in the following present a way of how to partition the simulation per event among multiple different engines. The user specifies the conditions of how tracks are distributed between them. Possible conditions can be based on the detector geometry, particle species or phase space as it was already motivated in Sec. 6.3 (see also Fig. 6.4 therein).

The core developments comprise seamless extensions of the `TVirtualMC` and the `TVirtualMCApplication` classes. Furthermore, the new central class `TMCManager` was introduced along with the `TMCManagerStack`. The class dependencies are sketched in Fig. 6.8. In the case where only one engine is chosen, this dependency graph collapses

to the one presented earlier in Sec. 6.4.2. Therefore, users do not need to change their implementations in case the new event partitioning feature is not desired.

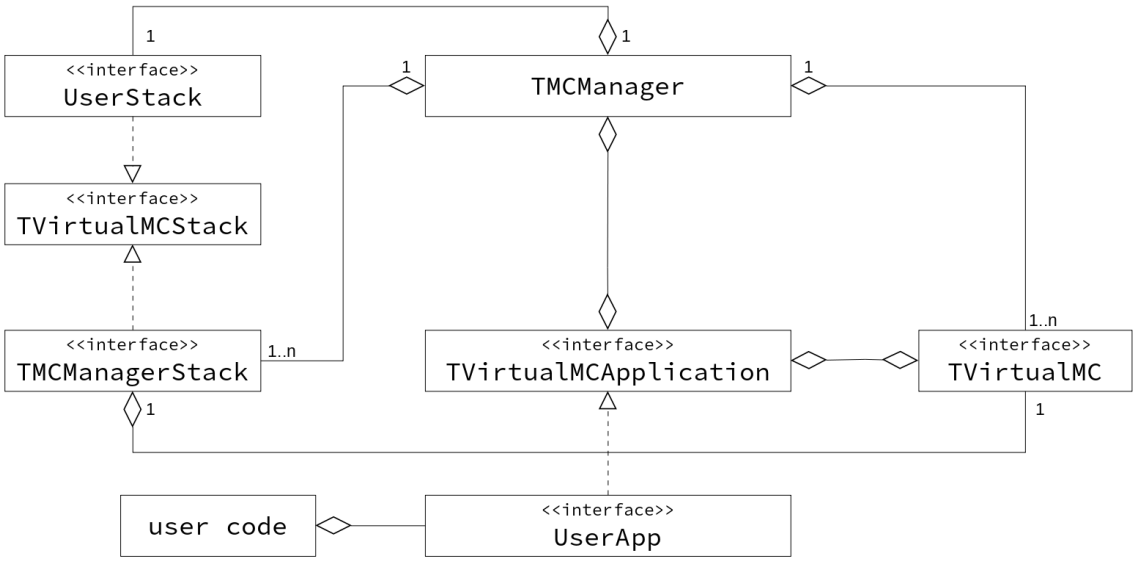


Figure 6.8: Organisation of VMC core classes after the extensions by the `TMCManager` and `TMCManagerStack`.

The `TMCManager` aggregates all instantiated objects of `TVirtualMC` automatically. When the run is initialised, it creates the same number of `TMCManagerStack` each associated to one engine interface. These stack classes are also realisations of `TVirtualMCStack` with specifically implemented behaviour of the interfaces. A user stack has to be registered with the `TMCManager` as well before the run is started such that it can synchronise the overall history on that stack with the partitioned run to provide a coherent history to the user. In that way, it looks to the user as if one single engine was running.

A more detailed description of the implementation and their working principle is given in the next two sections. First, the general interplay with the `TMCManager` is described. This is followed by the explanations of how the event or track partitioning is achieved.

Interaction between the `TMCManager`, the `TVirtualMCApplication` and the `TVirtualMC`

A run with multiple engines has to be requested explicitly by the user since otherwise no manager object is constructed and a single engine run is assumed. Any instantiation of further engines will lead to a fatal error. This ensures backward-compatibility.

The creation of the manager object can be requested in the implementation of `TVirtualMCApplication`. If done so, a *singleton* object of type `TMCManager` is constructed which is accessible from the member `TVirtualMCApplication::fMCManager`. At

the same time, each object of TVirtualMC registers itself automatically to the manager during construction if the manager is present. The additional aggregation of the TVirtualMCApplication allows the manager to invoke user routines when required such as the geometry initialisation before a run or the creation of primary particles for each new event.

It is ensured that

1. an instance of TVirtualMCApplication is present before any TVirtualMC can be instantiated (which was the default behaviour of the framework before),
2. in case of a run with multiple engines, the TMCManager object has been created,
3. only one engine can be constructed if no TMCManager has been created and therefore, the framework falls back to the exact workflow as presented in Sec. 6.4.2.

After all engines have been instantiated, the TMCManager object has to be initialised. This first invokes the construction of the geometry which has to be defined via ROOT's TGeoManager. Further, the initialisation of each registered engine is invoked including the physics initialisation. In addition, each engine refers to the same geometry tree and navigator entity. The used interface is TMCManager::Init() which is overloaded in addition by a templated version which accepts an object that in turn can be applied to each single engine. This is shown as an example in List. 6.1. The initialisation procedure also creates one object of TMCManagerStack per engine.

```
template <typename F>
void Init(F initFunction)
{
  //...
  Init();
  for (auto &mc : fEngines) {
     // Set to current engine and call user init procedure
     UpdateEnginePointers(mc);
     initFunction(mc);
  }
  //...
}
```

Listing 6.1: Example of how custom initialisation steps can be applied to engines during TMCManager initialisation.

The method TMCManager::Run(Int_t nEvents) launches the event loop for a specified number of events. This is fully controlled by the manager object and sketched in Fig. 6.9. For each event the manager centrally requests the primary generation. The first engine with particles to transport is steered and runs the event loop. As soon as all particles have been transported, the next engine is steered. This is done until all particles have been transported. During the transport, tracks might be transferred to another engine, for instance in case a volume has been reached for which a certain fast-simulation engine should take over the simulation. This track transfer is one of

the most important ingredient implemented during this work and described in the next section.

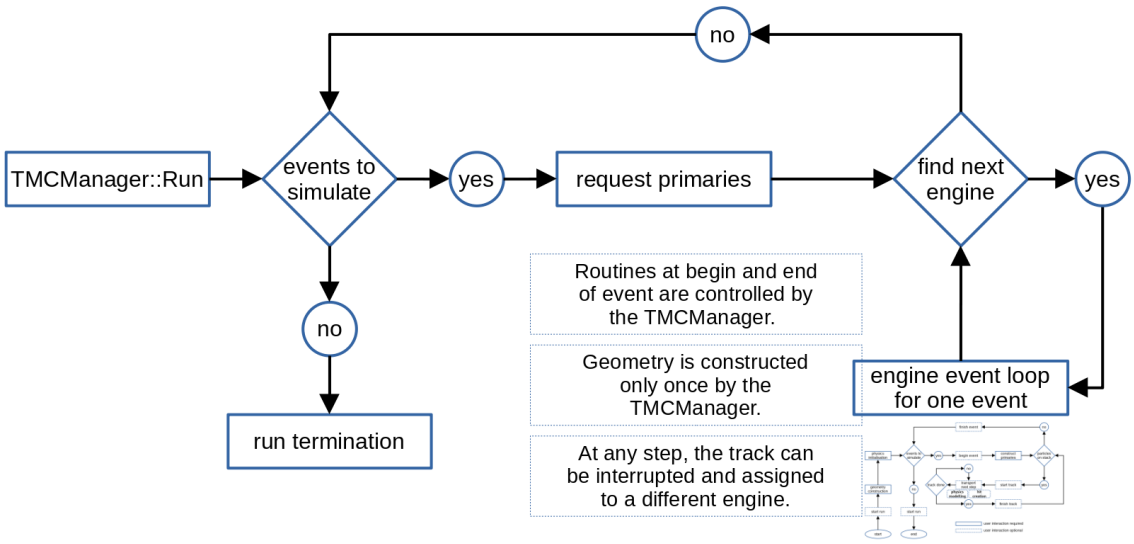


Figure 6.9: Run cycle managed by the `TMCManager`.

Transferring tracks between engines

In order to handle tracks of one event transported by different engines, a reliable stacking logic had to be implemented. The following considerations drove the development:

1. Copying of particle objects is avoided.
2. The user is the owner of constructed particle objects.
3. Cache geometry states for tracks transferred between engines to avoid the need fo the navigator to search the geometry when a track is resumed.

The first criterion is met by collecting pointers to the particle objects centrally in the `TMCManager` object. All `TMCManagerStack` objects hold a pointer to this collection and in addition they are only informed about the indices in this collection referring to the particles to be transported. In addition, a lightweight object storing intermediate kinematic and geometry information is stored for each track that has been transferred to another engine is provided. This accounts for the third consideration mentioned above since this object is used to re-initialise the navigator when a particle transport is resumed to avoid searching a potentially complex geometry tree.

The second criterion is realised by forwarding each request for the construction of a new particle to the user stack. Once constructed, the manager is informed

about it by using the interface `TMCManager::ForwardTrack`. The workflow is shown in Fig. 6.10. The target engine can be specified, otherwise the currently running engine that pushed the particle is assumed to be responsible for the transport.

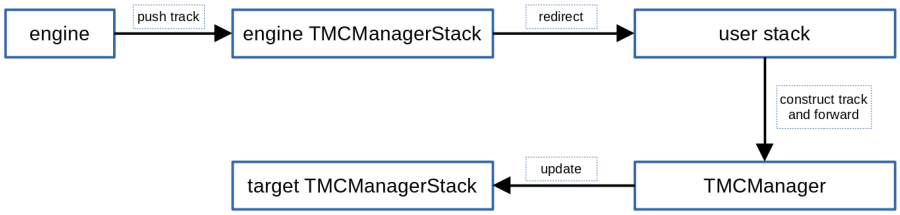


Figure 6.10: Stacking logic to ensure the user is the owner of constructed track objects.

At this point, the last core functionality is invoked by the interface `TMCManager::TransferTrack` as indicated in Fig. 6.11. Most likely, this is invoked in the user routine called during a MC step is done and the target engine is specified. If the track meets the criteria to be moved to another engine, this interface interrupts the simulation in the current engine and the engines' stacks are updated accordingly. The target engine acquires a new particle to be transported in addition with the stored geometry information. According to the event loop shown above, this particle will be transported as soon as the target engine's transport is invoked again.

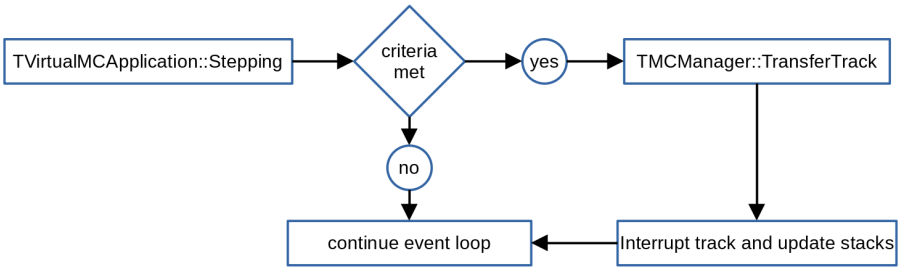


Figure 6.11: Logic of how tracks can be requested to be transferred from one engine to another.

Core principles of the implementation to handle multiple different engines have been presented. For further details of the `TMCManager` interfaces and their usage the reader is referred to [178], to the documentation [177] and the code repositories [176]. In order to use these new features, the VMC interface implementations had to be adapted accordingly which is briefly described in the following section.

Adapting the GEANT3_VMC and GEANT4_VMC interfaces

As discussed previously, GEANT4 is foreseen to be used as the default transport engine by ALICE, however, in order to ensure a fully backward-compatible development,

the GEANT3 interface was adapted as well. The main reason for the adaptation was the direct communication with the `TVirtualMCApplication` which was called by the interface implementation directly to invoke a potential user routine as explained in Sec. 6.4. Hence, in the case of multiple engines, that could either circumvent the `TMCManager` or could potentially lead to multiple redundant calls of the same user routine and would therefore corrupt the event loop.

The modifications ensure that an engine is prevented from, for instance, calling routines at the beginning of a track in case that track has been already transported partially by a previous engine. Similarly, end-of-track routines are avoided in case it was only interrupted to be transferred to another engine. This applies analogously to routines invoked for primary tracks or at the beginning and end of each event.

6.5.2 Performance

This section comprises performance evaluations and proof-of-principle examples to show that the new software functionality is ready to be used reliably.

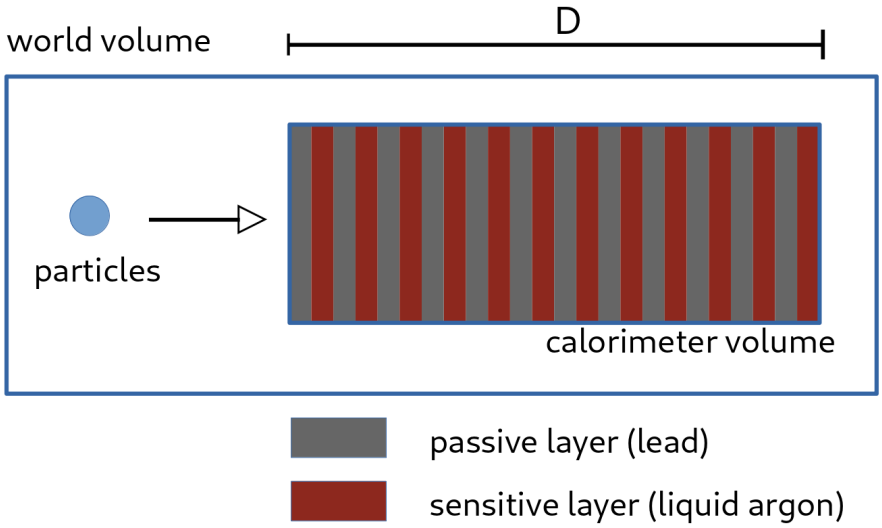


Figure 6.12: Geometry setup of a sampling calorimeter used for performance benchmarks.

A baseline scenario of a sampling calorimeter of depth D with N layers as sketched in Fig. 6.12* is chosen to show that

1. the partitioning according to different conditions works,
2. the track transfer does not introduce overhead w.r.t. the speed of the simulation,

*This has been adopted from the available example `E03c` at https://github.com/vmc-project/geant4_vmc/tree/master/examples/E03/E03c.

3. additional storage of track information (in particular to re-initialise the geometry states) does not lead to significant overhead w.r.t. timing and memory consumption.

All of these criteria have to be met in order to guarantee the reliable performance of each single engine and to ensure predictable speed and accuracy. This is particularly important when fast-simulations would be injected since its increase in speed must not depend in the workflow it is employed in. To demonstrate the correct working of the extensions it was decided to use the well established VMC engines interfacing to GEANT3 and GEANT4.

For the evaluation of the above criteria, GEANT3 and GEANT4 are used as they provide two fully functioning simulation engines. Note again that this section is not meant to introduce fast-simulations but focuses on the reliability of the VMC framework extensions.

Partitioning

To demonstrate that the partitioning works as expected, 6 partitioning conditions were chosen as summarised in Tab. 6.1. The scenarios are sorted by decreasing precedence from top to bottom. The simulated geometry is the one introduced above. 100 events each containing 100 primary protons of an energy of 20 GeV were simulated and the protons were shot head-on onto the geometry as indicated in the above figure.

scenario	condition	use engine
1	photon	GEANT3
2	$p < 50\,\mathrm{MeV}/c$	GEANT4
3	sensitive layer	GEANT4
4	sensitive layer and proton	GEANT3
5	passive layer	GEANT4
6	anything else	GEANT3

Table 6.1: Partitioning conditions to demonstrate the event partitioning and track transfer between engines. The conditions are sorted by precedence from top to bottom.

The relative number of steps simulated per engine and scenario are summarised in Tab. 6.2. As one can see, the steps were made exactly according to the conditions. In this case, most steps were simulated in GEANT3 for photons (scenario 1) which are produced in electromagnetic interactions of the proton with the material (for instance Bremsstrahlung). Since many secondary particles are produced at low energies, the second most steps were made for particles with momenta below $50\,\mathrm{MeV}/c$ with GEANT4 (scenario 2). Furthermore, GEANT4 transported all particles inside the

sensitive material (scenario 3), however, protons are transported by GEANT3 in that material (scenario 4). The transport in the passive material is overall handled by GEANT4 (scenario 5) and if none of the previous conditions applied, the transport is taken over by GEANT3 (scenario 6).

scenario	relative number of steps [%]	
	GEANT3	GEANT4
1	$\lesssim 0.1$	0
2	0	5.8
3	1.7	0
4	0	0.7
5	35.7	0
6	55.8	0

Table 6.2: Relative number of steps made per engine and scenario.

Performance of track transfer

In order to show that no overhead is introduced by the track transfer, the geometry shown above is used. Here, for simplicity, the partitioning is solely based on geometry: GEANT3 simulates the passive material while GEANT4 is used for the sensitive one. While keeping the depth D of the calorimeter constant, the number of layers was increased, starting off with 10 up to 1000 layers. 4 different particle species were shot onto the calorimeter, namely electron neutrinos, electrons, protons and neutrons. This was done to cover different physics processes. Neutrinos are expected to traverse the geometry without any interaction whereas electrons interact electromagnetically mostly initiating electromagnetic showers. For neutrons, nuclear interactions play a crucial role while protons participate in both electromagnetic as well as nuclear interactions. Per species, 100 events with 10 primary particles each were simulated. The setup is summarised in Tab. 6.3.

parameter	values
number of layers	$\{10, 20, 50, 100, 250, 500, 1000\}$
particle species	electron neutrino, electron ($2\,\mathrm{GeV}/c^2$), neutron ($20\,\mathrm{GeV}/c^2$), proton ($20\,\mathrm{GeV}/c^2$)
number of events	100
particles per event	10
partition condition	passive material: GEANT3, sensitive material GEANT4

Table 6.3: Simulation parameters to study the timing performance of event partitioning.

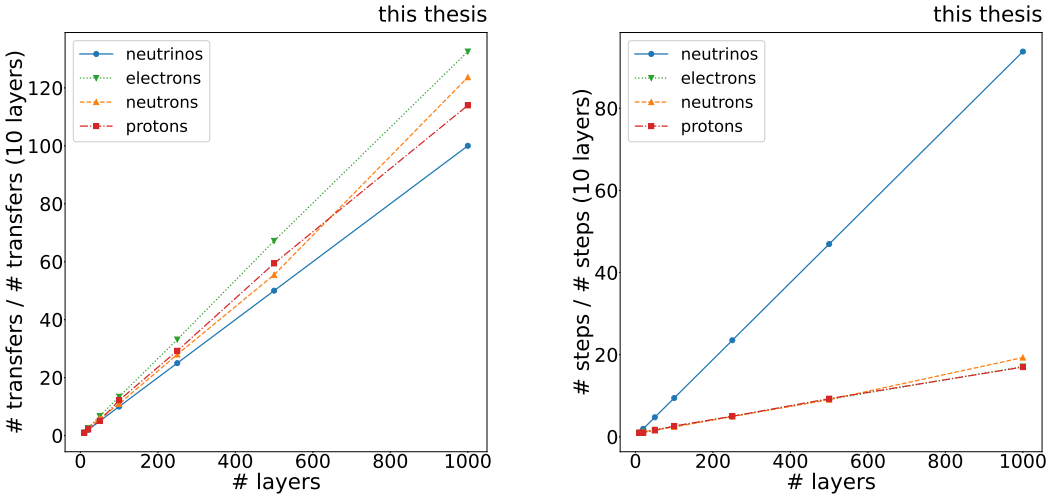


Figure 6.13: *Left:* Number of track transfers as a function of the number of layers in the test scenarios for different primary particles. The number of transfers is normalised to the number of transfers in the case of 10 layers. *Right:* Analogous plot for the number of steps.

The left panel of Fig. 6.13 shows the number of track transfers between engines as a function of the number of layers where the number of transfers was scaled to the number obtained for 10 layers. Neutrinos did not undergo any further interaction and in particular do not produce any secondary particles, thus, compared to the setting with 10 layers, the number of track transfers increases by the factor of the layer increase.

For all other particle species, physics processes are computed when they traverse the materials which also gives rise to secondary particles. This causes a stronger linear increase compared to the neutrino case since also some secondary particles traverse multiple layers. Overall, the number of transferred tracks scales with an increasing number of layers as expected.

The right panel in Fig. 6.13 shows the steps as a function of the number of layers and counted steps are scaled to the number of steps obtained for 10 layers. Again for the neutrino case, the number of steps scales directly with the number of layers. Since there are no physics interactions for the neutrino, it travels from boundary to boundary.

Fig. 6.14 shows the ratio of the number of steps in the partitioned scenario w.r.t. the simulation only conducted by GEANT4. All layers are inscribed into the overall calorimeter volume as indicated above in Fig. 6.12. This introduces an offset of steps due to additional boundary steps of particles entering or leaving the calorimeter. With the increasing number of layers this offset becomes negligible. Due to yet additional steps at the boundaries in the partitioned scenario, the impact of this offset is already

lowered for a smaller number of layers which is the reason for the mild turn-on visible in the ratios.

The observation that the ratio is lower for electrons, neutron and protons comes from the fact of a more coarse-grained simulation of GEANT3 in the passive layers where the GEANT4-only simulation is more detailed. Hence, the number of steps due to physics processes causes the overall number of steps to be lower in the partitioned scenario. For the neutrino on the other hand, no physics processes are invoked and hence neglecting the offset discussed above, the number of steps in the partitioned scenario is 1.5 times as large as in the GEANT4-only scenario: GEANT4 makes one step at each boundary and one step to transport the neutrino through the volume resulting in 2 steps per layer. Due to one additional step at each volume boundary during the track transfer, this number increases to 3 steps per layer.

Since in all particle cases the ratio reaches a plateau, the scaling of the number of steps with the number of layers in the partitioned scenario is negligible.

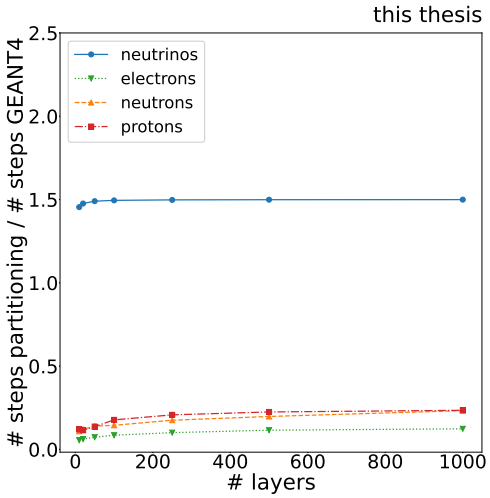


Figure 6.14: Ratio of number of steps of the partitioned scenario w.r.t. GEANT4 handling the entire simulation.

In a next step, the time the simulation takes is evaluated. Only the time spent in the event loop is considered hence excluding for instance the geometry construction or physics initialisation. An increase of the simulation time with an increasing number of layers is shown in the left panel of Fig. 6.15. For neutrinos, the transport depends only on volume boundaries and the increase is seen to be strongest. This is in line with the previous discussion. When physics processes are computed in addition, the increase of the simulation time is still linear in the number of layers, however, significantly milder.

In order to judge whether this increase is due to the transfer of tracks between engines, the calorimeter was also simulated using GEANT4 for the entire simulation.

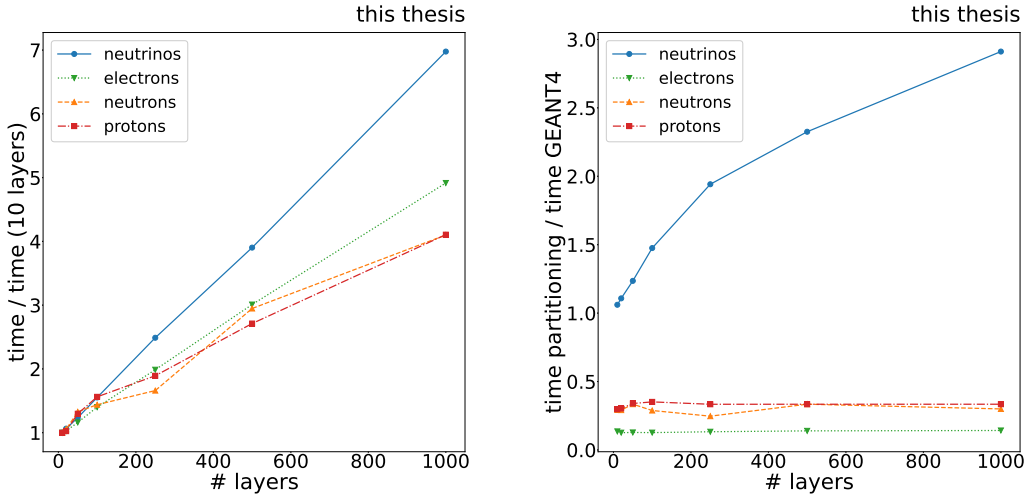


Figure 6.15: *Left:* Simulation time as a function of the number of layers in the test scenarios for different primary particles. The time is normalised to the time in the case of 10 layers. *Right:* Ratio of simulation time of partitioned scenario w.r.t. GEANT4 doing the entire simulation.

The ratio of the simulation time of the partitioned scenario to the GEANT4-only scenario is shown in the right panel of Fig. 6.15. For neutrinos it can indeed be seen that, although in both scenarios it is transported between volume boundaries, the ratio increases. In that case, the track transfer introduces computational overhead as it can be expected because it is the only additional computational procedure added in the partitioned case. On the other hand, the overhead from the transfer decreases with an increasing number of layers and therefore the number of transfers.

When compared to the transport of electrons, neutrons and protons, the ratio is flat and does not depend on the number of changes between engines. Therefore, it can be stated that the transfer of tracks between engines can be neglected compared to the physics simulation of full-simulation engines such as GEANT4.

As it can be seen as well from the right panel of Fig. 6.15 the ratio is not only flat but the simulation time was reduced by a factor of ≈ 3 (proton and neutron case) and ≈ 9 (electron case) compared to the GEANT4-only scenario. This is due to the more accurate simulation of the passive material when simulated also with GEANT4 which in the partitioned scenario was simulated with GEANT3. The partitioned scenario can therefore be understood as treating GEANT3 as a quasi-fast-simulation used together with GEANT4. A more accurate simulation would be used for the sensitive material where hits would be processed while the passive material would be simulated faster with GEANT3.

To illustrate a scenario where indeed yet another simulation engine can take over the transport from a full-simulation, the next section shows an example of partitioning

the event simulation by including a pseudo-fast-simulation engine.

A proof-of-principal scenario

The purpose of this section is to show that other *custom* realisations of `TVirtualMC` can be run together with a full-simulation. To do so, a VMC interface was written to mimic a fast-simulation engine to be applied to the calorimeter test case scenario as introduced above. The purpose of this engine is to compute the total energy deposit in the calorimeter cells as soon as the particle enters the first layer of the calorimeter.

Note, that this is a simplified scenario and not meant to introduce a fast-simulation ready to be used for actual physics simulation. Instead it is meant to demonstrate the partitioning with a custom VMC implementation.

GEANT4 is mixed with this engine as follows: First, GEANT4 is in charge of the transport of electrons within the world volume up to the calorimeter. As soon as that is reached, the track is transferred to the pseudo-fast-simulation engine which computes the deposited energy. To do so, it drew a value from a previously defined distribution which was obtained from a Gaussian fit to the actual energy distribution determined from a full GEANT4 simulation. No secondary particles are produced by the pseudo-fast-simulation.

Fig. 6.16 shows the energy distribution obtained by GEANT4 in red and the Gaussian fit is shown in grey. The blue points then show the distribution of the energy deposit when using a fast-simulation engine that draws values according to the fitted Gaussian.

By this scenario it has been shown that other custom realisations of `TVirtualMC` can be used in a partitioned simulation scenario*. In addition, in the previous section it was demonstrated that the event partitioning between different simulation engines is fully functional and no runtime overhead is introduced by transferring tracks between engines. Hence, the performance increase when partitioning to a custom fast-simulation solution will mainly depend on its implementation and stand-alone performance. A performance penalty from the interplay with other engines should be negligible as discussed.

It can be concluded that the enhanced VMC framework is ready to be used in order to inject fast-simulations into full-simulations or even to run different full-simulation engines or a combination of multiple full- and fast-simulations.

The code used to produce the performance plots is based on the example `E03c` which can be found in the GEANT4_VMC repository [181].

*The code of this scenario is based on [179, 180].

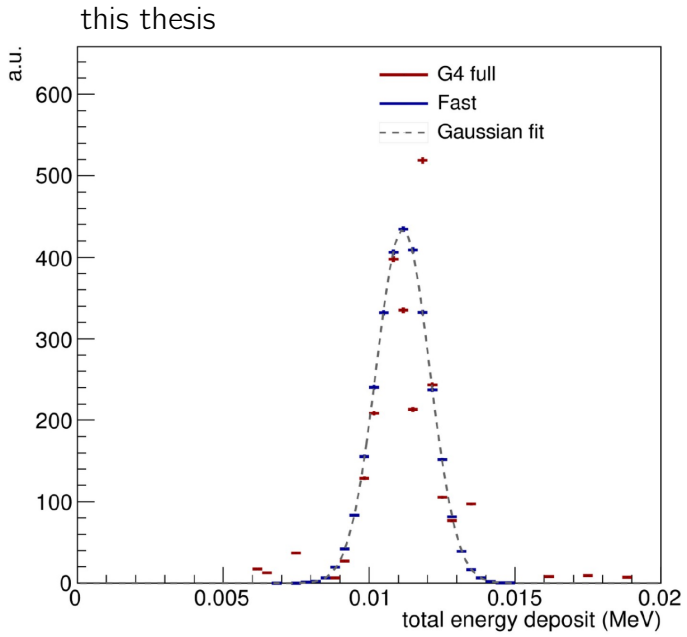


Figure 6.16: Energy distribution (dashed grey) fitted to the energy distribution obtained in a GEANT4 full-simulation scenario (red). A pseudo fast-simulation engine draws values from the fitted distribution to model the total energy deposit in the calorimeter (blue).

7 Conclusion

Two major results were obtained and presented in this work, namely

1. the extraction of prompt Λ_c^+ production at midrapidity $|y| < 0.5$ in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ as a function of charged-particle density (see Ch. 5),
2. a major extension of the VMC framework to be capable of partitioning tracks in an event during the simulation among multiple different engines (see Ch. 6).

7.1 Λ_c^+ production in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$

The result is used for the computation of the Λ_c^+/D^0 production ratio to further investigate an enhanced baryon-over-meson production in pp collisions in the HF sector. The presented results support what has previously been measured in pp collisions at $\sqrt{s} = 5.02\,\text{TeV}$ [115]: Model predictions only based on in-vacuum fragmentation (i) underestimate the measured ratio and (ii) do not reproduce the shape as a function of p_T. Even more striking, the current results obtained at $\sqrt{s} = 13\,\text{TeV}$ reveal that the enhancement increases with the charged-particle density in the events. In comparison with the discussed meson ratio D_s^+/D^0, which is seen to be flat in p_T and does not show a dependence on multiplicity, but also in comparison with baryon-to-meson ratios obtained in the LF sector [25, 26], the presented results give a strong indication for hadronisation mechanisms beyond pure in-vacuum fragmentation in the case of baryon production in pp collisions. The increase of the baryon-to-meson ratios with multiplicity are furthermore an indication for hadronisation mechanisms that might be driven by the particle density produced in pp collisions.

Model predictions based on the Lund string model implementing colour reconnection and a prediction based on a statistical hadronisation approach are qualitatively able to describe the shape and the increase of the Λ_c^+/D^0 production ratio with increasing multiplicity. A similar quality of compatibility is seen by comparing the measured ratio to a model that includes partonic coalescence in addition to pure fragmentation in [115]. Therefore, despite the fact that the pure fragmentation is altered, the current precision of the results does not allow to pinpoint a model prediction.

The precision of the ratios is limited by a similar amount by systematic and statistical uncertainties. Recently, new simulation samples based on PYTHIA 8 predictions became available and since systematic uncertainties such as the tracking efficiency,

the MC p_{T} shape, the efficiency weights and the BDT working point selection arise from the modelling of MC predictions (see Sec. 5.4), these uncertainties shall be re-evaluated. In addition, the fitting procedure of the involved Boosted Decision Trees can potentially be refined as outlined in Sec. 5.6. Although it is not expected to increase the precision significantly, especially the refined BDT optimisation strategy provides a clearly defined procedure. An estimation of the potential separation refines the feature selection further. On top of that, the evaluation of the feature importance based on the SHAP values [160] provides additional information compared to the simple importance evaluation. In particular, the hyperparameter optimisation based on a Bayesian approach [159] allows to scan a significantly larger space of parameters compared to a grid-search approach. Such an optimisation procedure should therefore be favoured above a grid search also in other future analyses if possible.

A significant improvement of the measurements can be expected when the data of Run 3 and 4 will become available in the future. For instance, the statistical precision on the raw yield extraction can be expected to improve significantly and a combined analysis of data from both runs might yield an improvement by a factor of $\lesssim 10$. If also the systematic uncertainties can be refined and improved, future measurements could provide further concrete and valuable insight into the hadronisation processes at work in pp collisions, in particular w.r.t. the comparison with model predictions.

Furthermore, an additional measurement of Λ_{c}^{+} production in Pb–Pb collisions over the similar p_{T} range would provide an important comparison to the measurements presented in this work. Comparisons might reveal or rule out common properties in heavy-ion and pp collisions.

As a final remark, it will be attempted to extrapolate the production analysis of prompt Λ_{c}^{+} down to 0 momentum in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$. If that turns out to be feasible with sufficient precision, the extraction of integrated ratios over p_{T} could help to answer the question of whether there is a shift of the ratio in p_{T} with increasing multiplicity or whether there is indeed an overall enhancement of the ratio over the entire p_{T} range.

7.2 Virtual Monte Carlo developments for LHC Run 3 and 4

Thanks to the extensions, the VMC framework is now ready to support multiple collaborating engines during transport simulation. This feature enables the combination of fast-simulation engines with full-simulation and tracks can be transferred among the engines to transport particles by full-simulations where necessary and by fast-simulations where feasible. It has been shown that no overhead is introduced by the transfer between engines and in addition, the developments fully conserve compatibility with previously implemented user classes.

As of Run 3 and 4, the ALICE collaboration will move to GEANT4 as the default full-simulation engine. At the same time, full-simulation is foreseen to be the default approach for detector simulations. Due to the requirements, fast-simulation approaches such as embedding are currently being investigated. If then also the need arises for developing and injecting fast-simulations into the transport or hit computation, the presented enhancement of the VMC package devises the necessary interfaces and functionality to do so.

Acknowledgements

The last three years have been quite an exciting journey and I am very thankful to everyone who accompanied and guided me during this time.

I therefore thank Prof. Dr. Johanna Stachel who made this journey possible in the first place. Although Geneva and Heidelberg are not exactly close, I received help, valuable feedback and support wherever possible or necessary.

For his guidance, help, encouragement and just the fact that I could find my place in the ALICE group at CERN, I am very thankful to Dr. Jan Fiete Grosse-Oetringhaus.

I would not have this understanding of physics I have today without Dr. Gian Michele Innocenti and I am grateful for his patience and all the discussions.

The same is true for the programming side. I am in particular thankful to Dr. Sandro Wenzel who encouraged me to give a talk at the CHEP conference in 2019.

And of course it was a pleasure to share an office with Dr. Maximiliano Puccio.

I also thank the entire CERN ALICE group. I really felt being part of productive and creative team.

I thank Prof. Dr. Ulrich Uwer to agree and find the time to be a referee for this thesis.

I owe great thanks to everyone who lived with me through this strange last year and made my life easier. Without the support of good friends that would have been quite a difficult time.

And finally I thank Imke for always listening, supporting me but also pushing me when it was necessary. I believe that especially thanks to her, I achieved things that I would not have otherwise.

A Lists

A.1 List of Figures

A.2 List of Tables

B Analysis

B.1 FONLL

For the used FONLL prediction in this work the production of $c\bar{c}$ and $b\bar{b}$ quark pairs is modelled by means of FONLL approach as described in Sec. 3.1 (see also [77, 182]).

The FONLL predictions to retrieve the expected significance of the Boosted Decision Trees (see Sec. 5.2.4) provide the predicted p_{T} shape of prompt Λ_{c}^{+} baryons. This was also used to estimate the modelling uncertainty of the simulated Λ_{c}^{+} p_{T} distribution for the estimation of the corresponding systematic uncertainty of the MC p_{T} shape (see Sec. 5.4).

Predictions for the p_{T} distribution of Λ_{c}^{+} baryons arising from $\Lambda_{\mathrm{b}}^{0} \to \Lambda_{\mathrm{c}}^{+} + \mathrm{X}$ feed-down where obtained by applying the p_{T} dependent Λ_{b}^{0} fraction as measured by the LHCb collaboration in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ in the pseudorapidity region $2 < |\eta| < 5$ [154]. As discussed in the reference, no rapidity dependence was observed and therefore it was assumed to be valid to be used for this analysis where the selection was done at central rapidity. The decay kinematics of the Λ_{c}^{+} was then handled by PYTHIA 8 [94]. These feed-down predictions were used to obtain the prompt fraction f_{prompt} of Λ_{c}^{+} candidates to correct the extracted raw yield as described in Sec. 5.3.

B.2 BDT feature correlations

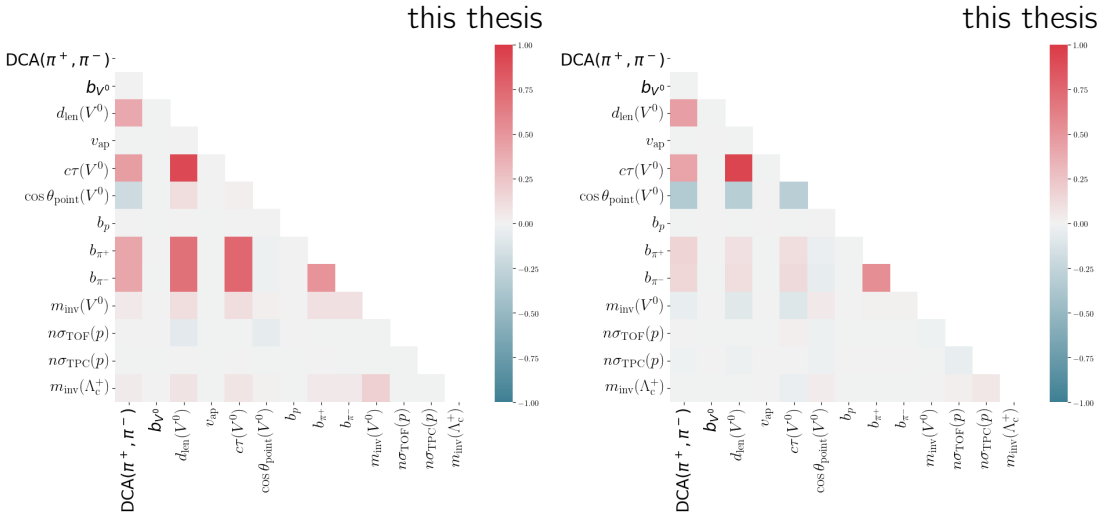


Figure B.1: Correlations among the chosen features and the Λ_c^+ reconstructed invariant mass in the 1 GeV/c to 2 GeV/c momentum interval for signal *(left)* and background *(right)*. The colour code goes from red (fully positively correlated) to blue (fully negatively correlated) as indicated by the colour bar on the right side of each plot.

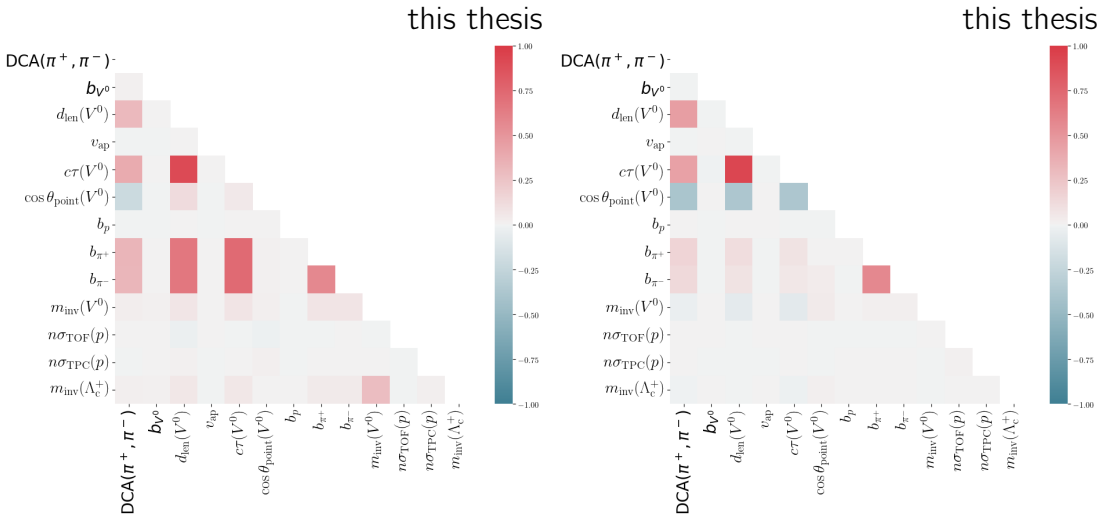


Figure B.2: Correlations among the chosen features and the Λ_c^+ reconstructed invariant mass in the 2 GeV/c to 4 GeV/c momentum interval for signal *(left)* and background *(right)*. The colour code goes from red (fully positively correlated) to blue (fully negatively correlated) as indicated by the colour bar on the right side of each plot.

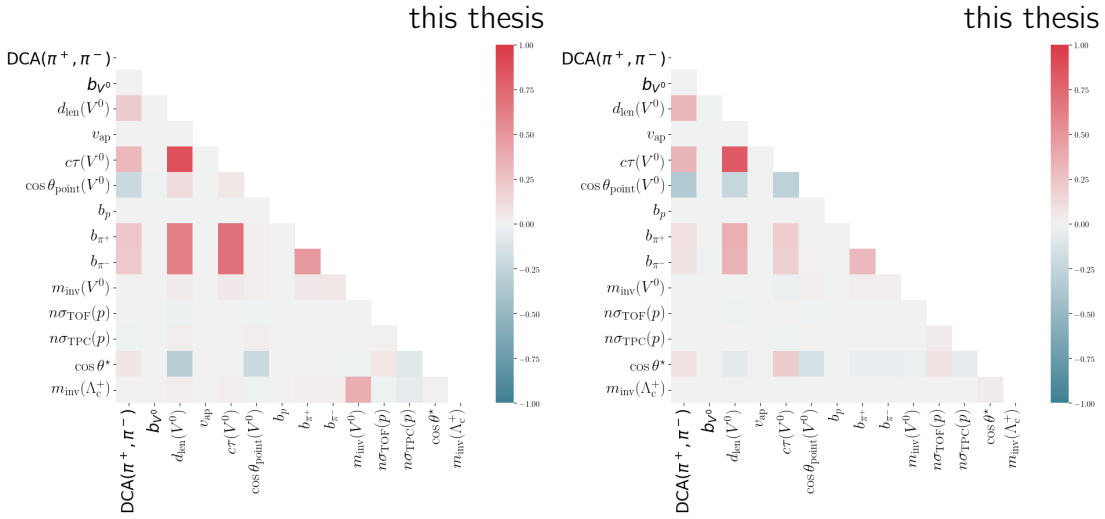


Figure B.3: Correlations among the chosen features and the Λ_c^+ reconstructed invariant mass in the 4 GeV/c to 6 GeV/c momentum interval for signal *(left)* and background *(right)*. The colour code goes from red (fully positively correlated) to blue (fully negatively correlated) as indicated by the colour bar on the right side of each plot.

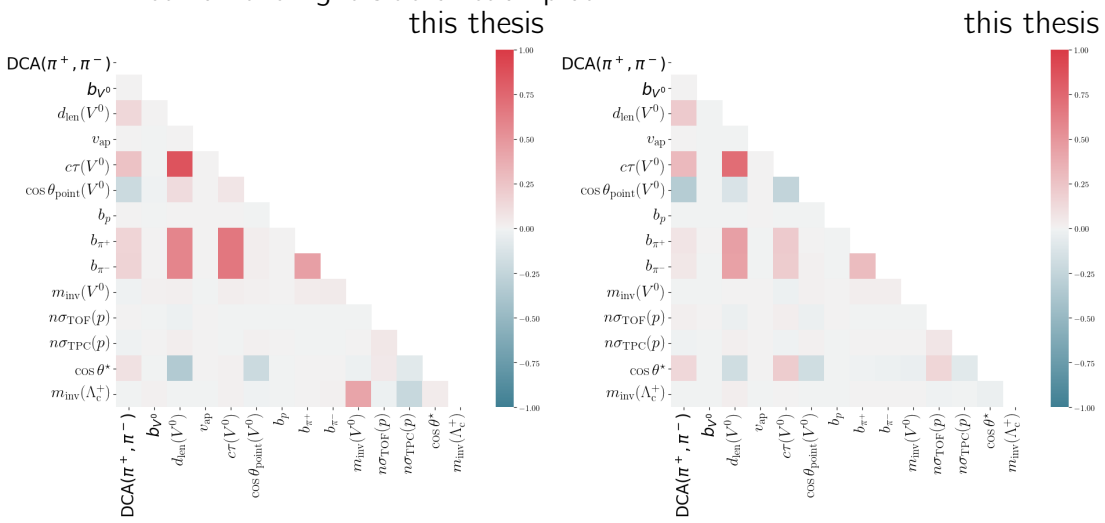


Figure B.4: Correlations among the chosen features and the Λ_c^+ reconstructed invariant mass in the 6 GeV/c to 8 GeV/c momentum interval for signal *(left)* and background *(right)*. The colour code goes from red (fully positively correlated) to blue (fully negatively correlated) as indicated by the colour bar on the right side of each plot.

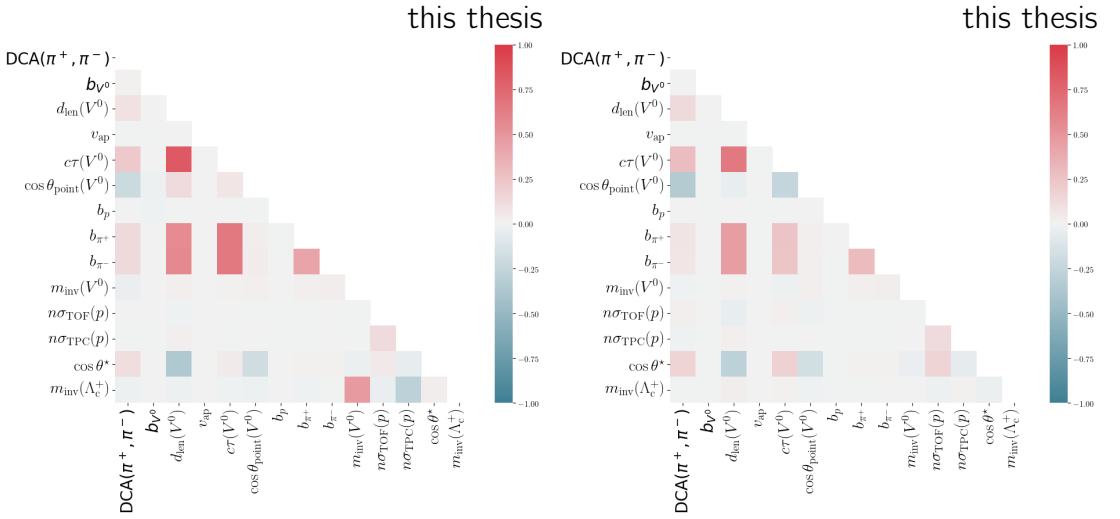


Figure B.5: Correlations among the chosen features and the Λ_c^+ reconstructed invariant mass in the 8 GeV/c to 12 GeV/c momentum interval for signal *(left)* and background *(right)*. The colour code goes from red (fully positively correlated) to blue (fully negatively correlated) as indicated by the colour bar on the right side of each plot.

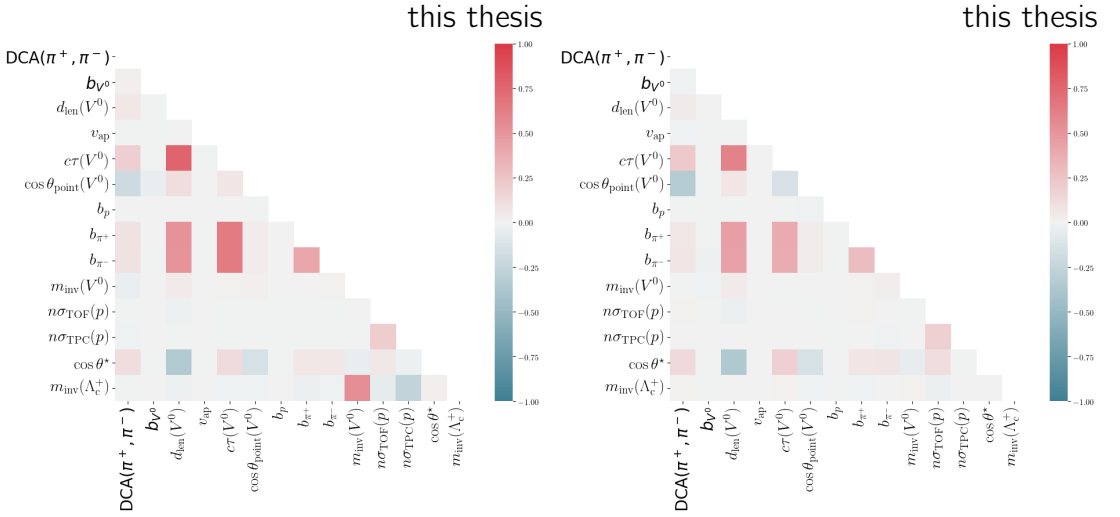


Figure B.6: Correlations among the chosen features and the Λ_c^+ reconstructed invariant mass in the 12 GeV/c to 24 GeV/c momentum interval for signal *(left)* and background *(right)*. The colour code goes from red (fully positively correlated) to blue (fully negatively correlated) as indicated by the colour bar on the right side of each plot.

B.3 Λ_c^+ invariant mass fits

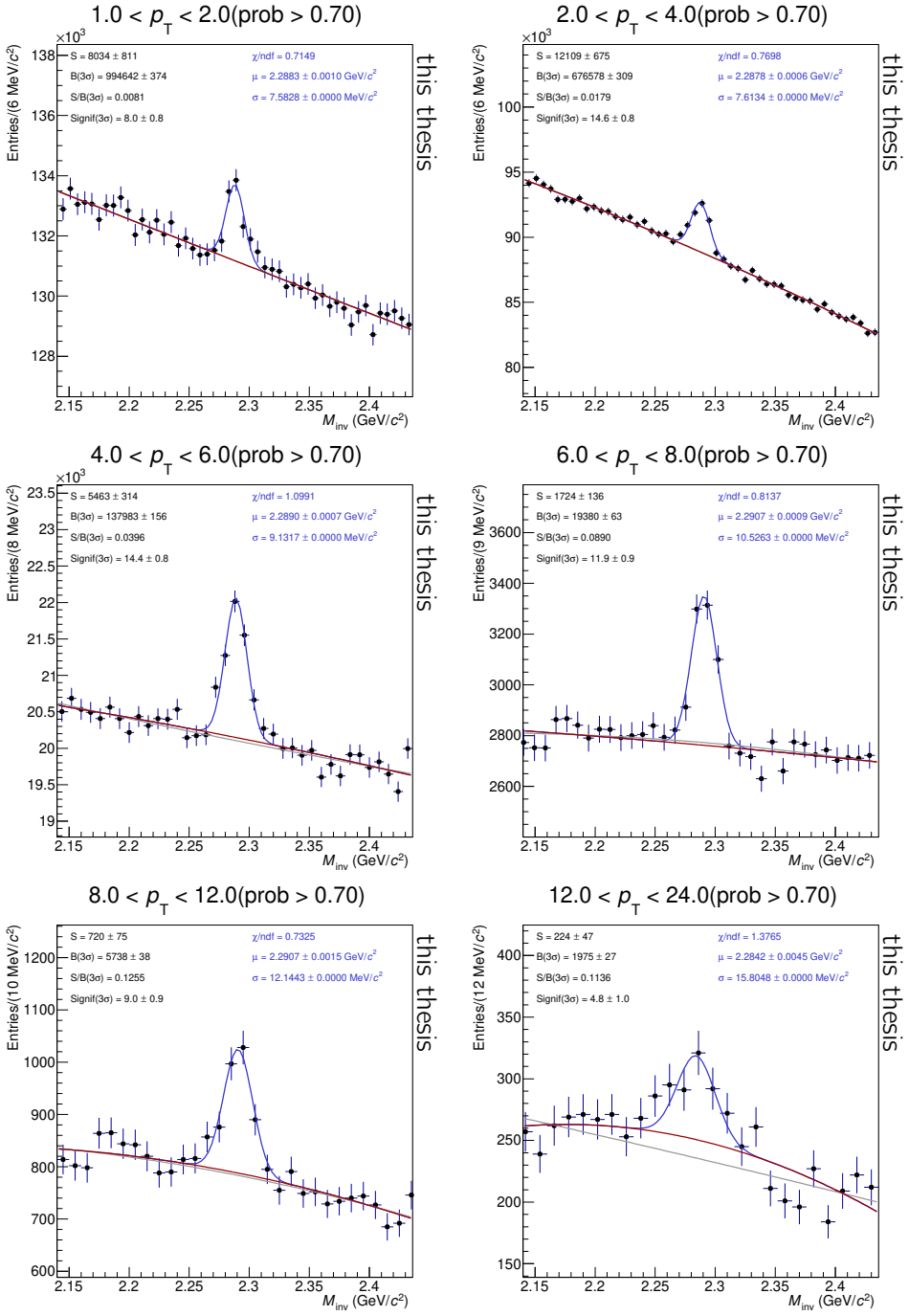


Figure B.7: Binned maximum-likelihood fit to the invariant mass distribution of selected Λ_c^+ candidates per p_T interval *(from top-left to bottom-right)* in the [0 %, 100 %] multiplicity interval. The initial sideband fit is shown in grey while the full fit is shown in blue. In red the sideband part of the full fit is shown.

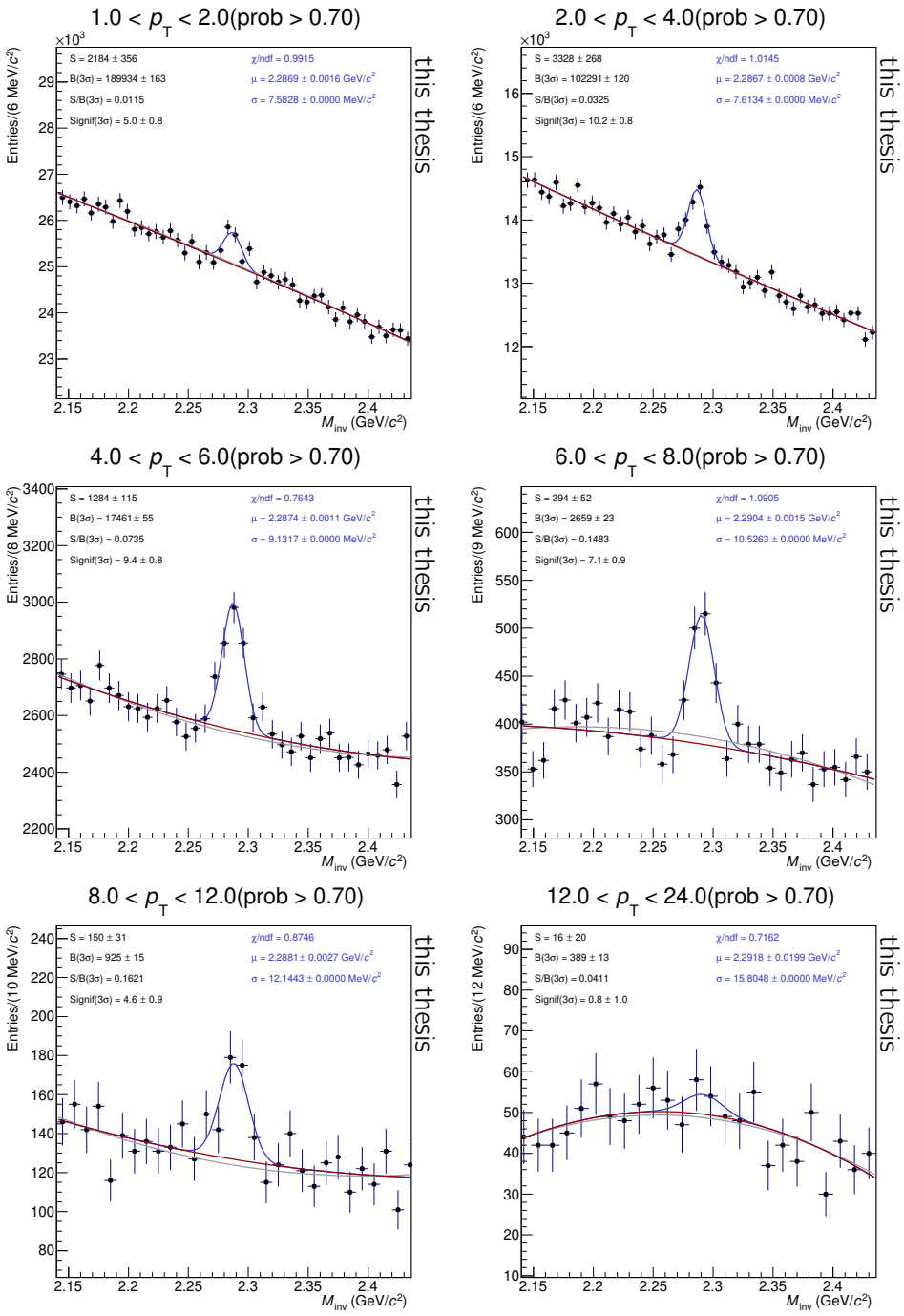


Figure B.8: Binned maximum-likelihood fit to the invariant mass distribution of selected Λ_c^+ candidates per p_T interval *(from top-left to bottom-right)* in the [30 %, 100 %] multiplicity interval. The initial sideband fit is shown in grey while the full fit is shown in blue. In red the sideband part of the full fit is shown.

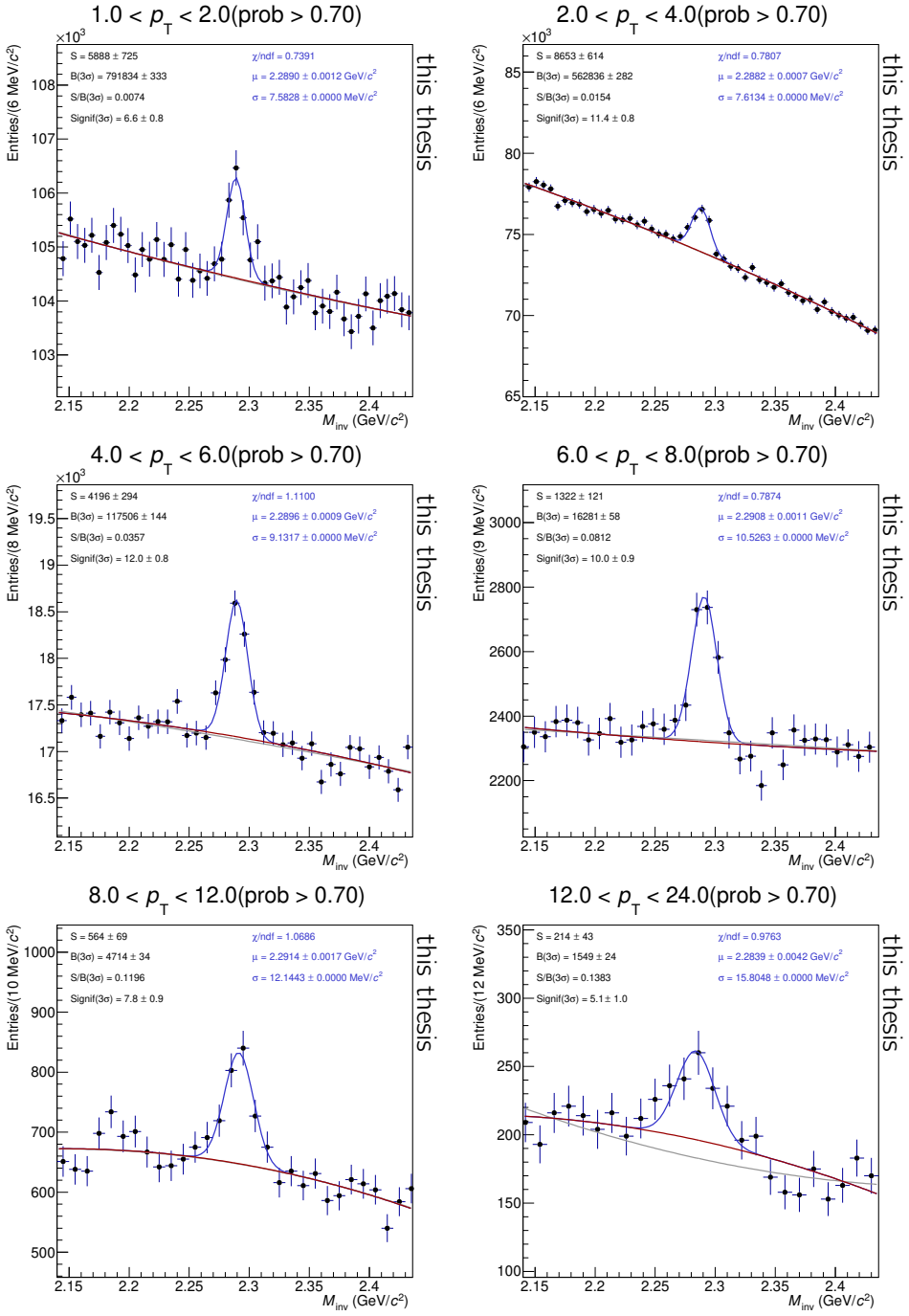


Figure B.9: Binned maximum-likelihood fit to the invariant mass distribution of selected Λ_c^+ candidates per p_T interval *(from top-left to bottom-right)* in the [0.1 %, 30 %] multiplicity interval. The initial sideband fit is shown in grey while the full fit is shown in blue. In red the sideband part of the full fit is shown.

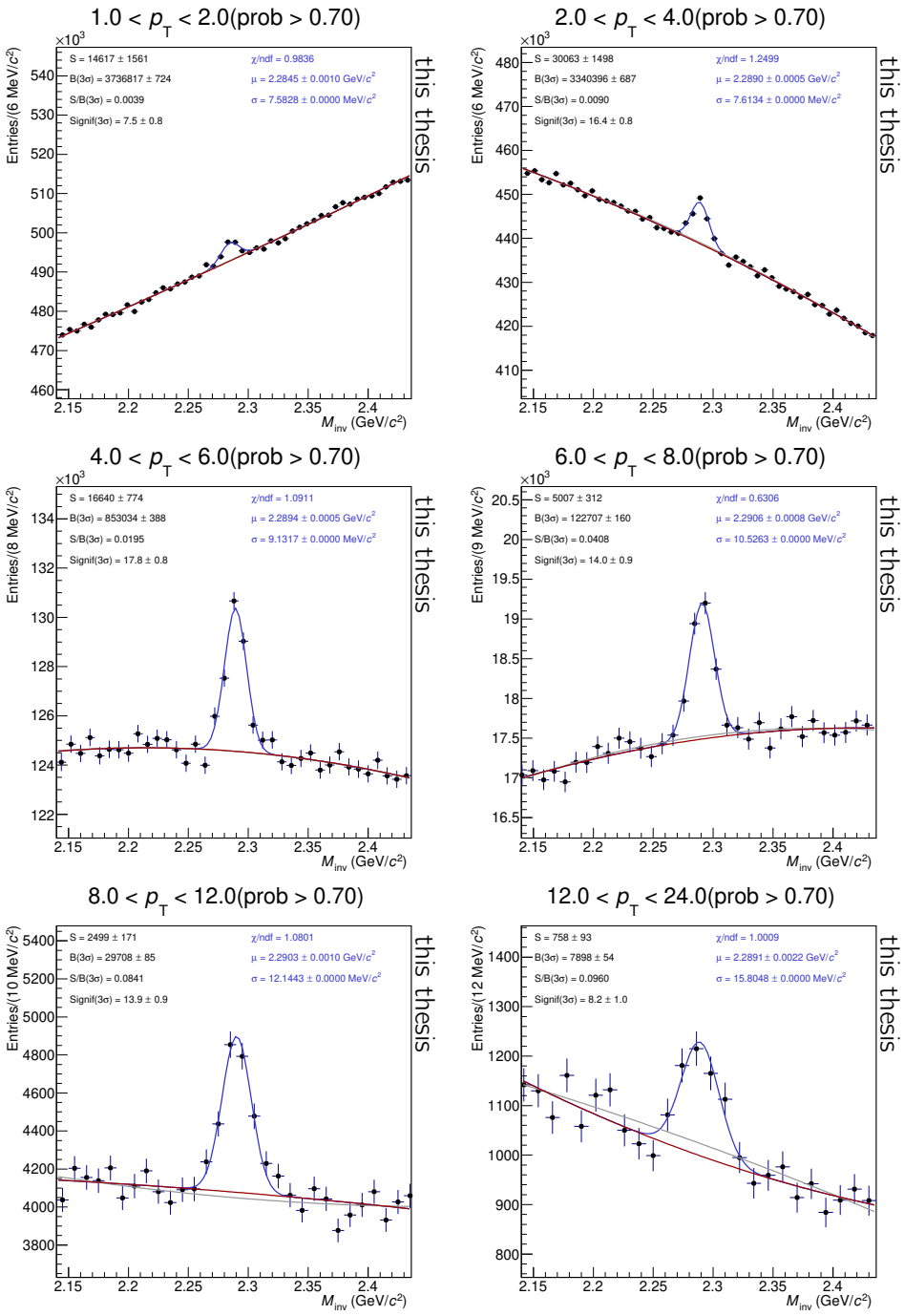


Figure B.10: Binned maximum-likelihood fit to the invariant mass distribution of selected Λ_c^+ candidates per p_T interval *(from top-left to bottom-right)* in the $[0\,\%, 0.1\,\%]$ multiplicity interval. The initial sideband fit is shown in grey while the full fit is shown in blue. In red the sideband part of the full fit is shown.

C Bibliography

[1] S. L. Glashow, "Partial-symmetries of weak interactions", *Nuclear Physics* **22** no. 4, (1961) 579 – 588.

[2] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266.

[3] A. Salam, "Weak and Electromagnetic Interactions", *Conf. Proc. C* **680519** (1968) 367–377.

[4] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, "Advantages of the color octet gluon picture", *Physics Letters B* **47** no. 4, (1973) 365 – 368.

[5] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509.

[6] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323.

[7] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587.

[8] G. 't Hooft and M. Veltman, "Regularization and renormalization of gauge fields", *Nuclear Physics B* **44** no. 1, (1972) 189 – 213.

[9] J. J. T. M. F.R.S., "XL. Cathode Rays", *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **44** no. 269, (1897) 293–316, https://doi.org/10.1080/14786449708621070.

[10] G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Physics Letters B* **716** no. 1, (2012) 1 – 29.

[11] S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Physics Letters B* **716** no. 1, (2012) 30 – 61.

[12] A. Einstein, "The Foundation of the General Theory of Relativity", *Annalen Phys.* **49** no. 7, (1916) 769–822.

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

[14] J. C. Collins and D. E. Soper, "The Theorems of Perturbative QCD", *Annual Review of Nuclear and Particle Science* **37** no. 1, (1987) 383–409, https://doi.org/10.1146/annurev.ns.37.120187.002123.

[15] **STAR** Collaboration, C. Adler *et al.*, "Disappearance of back-to-back high p_T hadron correlations in central Au+Au collisions at $\sqrt{s_{NN}}$ = 200-GeV", *Phys. Rev. Lett.* **90** (2003) 082302, arXiv:nucl-ex/0210033.

[16] **STAR** Collaboration, B. I. Abelev *et al.*, "Long range rapidity correlations and jet production in high energy nuclear collisions", *Phys. Rev. C* **80** (2009) 064912, arXiv:0909.0191 [nucl-ex].

[17] **STAR** Collaboration, K. H. Ackermann *et al.*, "Elliptic flow in Au + Au collisions at (S(NN))**(1/2) = 130 GeV", *Phys. Rev. Lett.* **86** (2001) 402–407, arXiv:nucl-ex/0009011.

[18] K. Dusling, W. Li, and B. Schenke, "Novel collective phenomena in high-energy protonproton and protonnucleus collisions", *International Journal of Modern Physics E* **25** no. 01, (Jan, 2016) 1630002.

[19] J. L. Nagle and W. A. Zajc, "Small System Collectivity in Relativistic Hadronic and Nuclear Collisions", *Annual Review of Nuclear and Particle Science* **68** no. 1, (Oct, 2018) 211235.

[20] B. Abelev, J. Adam, D. Adamová, A. M. Adare, M. M. Aggarwal, G. Aglieri Rinella, M. Agnello, A. G. Agocs, A. Agostinelli, Z. Ahammed, and et al., "K^0_S and Λ Production in Pb-Pb Collisions at $\sqrt{s_{NN}} = 2.76$ TeV", *Physical Review Letters* **111** no. 22, (Nov, 2013) .

[21] B. Abelev, J. Adam, D. Adamová, M. Aggarwal, G. Aglieri Rinella, M. Agnello, A. Agostinelli, N. Agrawal, Z. Ahammed, N. Ahmad, and et al., "Production of charged pions, kaons and protons at large transverse momenta in pp and PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV", *Physics Letters B* **736** (Sep, 2014) 196207.

[22] B. Abelev, J. Adam, D. Adamová, A. Adare, M. Aggarwal, G. Aglieri Rinella, M. Agnello, A. Agocs, A. Agostinelli, Z. Ahammed, and et al., "Multiplicity dependence of pion, kaon, proton and lambda production in pPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV", *Physics Letters B* **728** (Jan, 2014) 2538.

[23] A. Collaboration, "Multi-strange baryon production in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV", 2016.

[24] J. Adam, D. Adamová, M. Aggarwal, G. Aglieri Rinella, M. Agnello, N. Agrawal, Z. Ahammed, S. Ahmad, S. Ahn, S. Aiola, and et al., "Multiplicity dependence of charged pion, kaon, and (anti)proton production

at large transverse momentum in pPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV", *Physics Letters B* **760** (Sep, 2016) 720735.

[25] S. Acharya, F. T.-. Acosta, D. Adamová, A. Adler, J. Adolfsson, M. M. Aggarwal, G. Aglieri Rinella, M. Agnello, N. Agrawal, Z. Ahammed, and et al., "Multiplicity dependence of light-flavor hadron production in pp collisions at $\sqrt{s} = 7$ TeV", *Physical Review C* **99** no. 2, (Feb, 2019) .

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

[27] B. A. Kniehl, G. Kramer, I. Schienbein, and H. Spiesberger, "Collinear subtractions in hadroproduction of heavy quarks", *Eur. Phys. J. C* **41** (2005) 199–212, arXiv:hep-ph/0502194.

[28] B. A. Kniehl, G. Kramer, I. Schienbein, and H. Spiesberger, "Inclusive charmed-meson production at the CERN LHC", *The European Physical Journal C* **72** no. 7, (Jul, 2012) .

[29] V. Greco, C. M. Ko, and P. Levai, "Parton coalescence at RHIC", *Phys. Rev. C* **68** (2003) 034904, arXiv:nucl-th/0305024.

[30] P. Braun-Munzinger, K. Redlich, and J. Stachel, "Particle production in heavy ion collisions", arXiv:nucl-th/0304013.

[31] P. Buncic, M. Krzewicki, and P. Vande Vyvre, "Technical Design Report for the Upgrade of the Online-Offline Computing System",.

[32] A. Ferrari, P. R. Sala, A. Fasso, and J. Ranft, "FLUKA: A multi-particle transport code (Program version 2005)",.

[33] T. T. Böhlen, F. Cerutti, M. P. W. Chin, A. Fassò, A. Ferrari, P. G. Ortega, A. Mairani, P. R. Sala, G. Smirnov, and V. Vlachoudis, "The FLUKA Code: Developments and Challenges for High Energy and Medical Applications", *Nucl. Data Sheets* **120** (2014) 211–214.

[34] R. Brun, F. Bruyant, M. Maire, A. C. McPherson, and P. Zanarini, *GEANT 3: user's guide Geant 3.10, Geant 3.11; rev. version*. CERN, Geneva, 1987. https://cds.cern.ch/record/1119728.

[35] **GEANT4** Collaboration, S. Agostinelli *et al.*, "GEANT4–a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250–303.

[36] J. Allison *et al.*, "Geant4 developments and applications", *IEEE Trans. Nucl. Sci.* **53** (2006) 270.

[37] J. Allison *et al.*, "Recent developments in Geant4", *Nucl. Instrum. Meth. A* **835** (2016) 186–225.

[38] I. Hrivnacova, D. Adamova, V. Berejnoi, R. Brun, F. Carminati, A. Fasso, E. Futo, A. Gheata, I. Gonzalez Caballero, and A. Morsch, "The Virtual Monte Carlo",. https://cds.cern.ch/record/619573. Talk from the 2003 Computing in High Energy and Nuclear Physics (CHEP03), La Jolla, Ca, USA, March 2003, 8 pages, LaTeX, 6 eps figures. PSN THJT006. See http://root.cern.ch/root/vmc/VirtualMC.html.

[39] D. J. Gross and F. Wilczek, "Ultraviolet Behavior of Non-Abelian Gauge Theories", *Phys. Rev. Lett.* **30** (Jun, 1973) 1343–1346.

[40] H. D. Politzer, "Reliable Perturbative Results for Strong Interactions?", *Phys. Rev. Lett.* **30** (Jun, 1973) 1346–1349.

[41] M. Gell-Mann, "A Schematic Model of Baryons and Mesons", *Phys. Lett.* **8** (1964) 214–215.

[42] G. Zweig, "An SU_3 model for strong interaction symmetry and its breaking; Version 1", Tech. Rep. CERN-TH-401, CERN, Geneva, Jan, 1964. https://cds.cern.ch/record/352337.

[43] G. Zweig, "An SU_3 model for strong interaction symmetry and its breaking; Version 2",. https://cds.cern.ch/record/570209.

[44] R. Soltz, C. DeTar, F. Karsch, S. Mukherjee, and P. Vranas, "Lattice QCD Thermodynamics with Physical Quark Masses", *Annual Review of Nuclear and Particle Science* **65** no. 1, (2015) 379–402, https://doi.org/10.1146/annurev-nucl-102014-022157.

[45] M. Lüscher, K. Symanzik, and P. Weisz, "Anomalies of the free loop wave equation in the WKB approximation", *Nuclear Physics B* **173** no. 3, (1980) 365 – 396.

[46] M. Lüscher, "Symmetry-breaking aspects of the roughening transition in gauge theories", *Nuclear Physics B* **180** no. 2, (1981) 317 – 329.

[47] **HotQCD Collaboration** Collaboration, A. Bazavov, T. Bhattacharya, C. DeTar, H.-T. Ding, S. Gottlieb, R. Gupta, P. Hegde, U. M. Heller, F. Karsch, E. Laermann, L. Levkova, S. Mukherjee, P. Petreczky, C. Schmidt, C. Schroeder, R. A. Soltz, W. Soeldner, R. Sugar, M. Wagner, and P. Vranas, "Equation of state in $(2+1)$-flavor QCD", *Phys. Rev. D* **90** (Nov, 2014) 094503.

[48] A. Chodos, R. L. Jaffe, K. Johnson, and C. B. Thorn, "Baryon structure in the bag theory", *Phys. Rev. D* **10** (Oct, 1974) 2599–2604.

[49] T. DeGrand, R. L. Jaffe, K. Johnson, and J. Kiskis, "Masses and other parameters of the light hadrons", *Phys. Rev. D* **12** (Oct, 1975) 2060–2076.

[50] **NA61** Collaboration, N. Abgrall *et al.*, "NA61/SHINE facility at the CERN SPS: beams and detector system", *JINST* **9** (2014) P06005, arXiv:1401.4699 [physics.ins-det].

[51] **CBM** Collaboration, J. M. Heuser, "The Compressed Baryonic Matter Experiment at FAIR", *EPJ Web Conf.* **13** (2011) 03001.

[52] B. V. Jacak and B. Müller, "The Exploration of Hot Nuclear Matter", *Science* **337** no. 6092, (2012) 310–314, https://science.sciencemag.org/content/337/6092/310.full.pdf.

[53] S. S. Bass, April, 2021.

[54] B. Hippolyte, April, 2021.

[55] W. Busza, K. Rajagopal, and W. van der Schee, "Heavy Ion Collisions: The Big Picture, and the Big Questions", *Ann. Rev. Nucl. Part. Sci.* **68** (2018) 339–376, arXiv:1802.04801 [hep-ph].

[56] S. Voloshin and Y. Zhang, "Flow study in relativistic nuclear collisions by Fourier expansion of azimuthal particle distributions", *Zeitschrift für Physik C Particles and Fields* **70** no. 4, (May, 1996) 665671.

[57] A. M. Poskanzer and S. A. Voloshin, "Methods for analyzing anisotropic flow in relativistic nuclear collisions", *Physical Review C* **58** no. 3, (Sep, 1998) 16711678.

[58] J. Adam, D. Adamová, M. M. Aggarwal, G. Aglieri Rinella, M. Agnello, N. Agrawal, Z. Ahammed, S. Ahmad, S. U. Ahn, and et al., "Higher harmonic flow coefficients of identified hadrons in Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV", *Journal of High Energy Physics* **2016** no. 9, (Sep, 2016) .

[59] B. Abelev, J. Adam, D. Adamová, A. Adare, M. Aggarwal, G. Aglieri Rinella, M. Agnello, A. Agocs, A. Agostinelli, Z. Ahammed, and et al., "Long-range angular correlations of , K and p in pPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV", *Physics Letters B* **726** no. 1-3, (Oct, 2013) 164177.

[60] V. Khachatryan, A. Sirunyan, A. Tumasyan, W. Adam, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, J. Erö, and et al., "Evidence for collectivity in pp collisions at the LHC", *Physics Letters B* **765** (Feb, 2017) 193220.

[61] **ALICE** Collaboration, K. Aamodt *et al.*, "Higher harmonic anisotropic flow measurements of charged particles in Pb-Pb collisions at $\sqrt{s_{NN}}$=2.76 TeV", *Phys. Rev. Lett.* **107** (2011) 032301, arXiv:1105.3865 [nucl-ex].

[62] **ATLAS** Collaboration, G. Aad *et al.*, "Measurement of the azimuthal anisotropy for charged particle production in $\sqrt{s_{NN}} = 2.76$ TeV lead-lead collisions with the ATLAS detector", *Phys. Rev. C* **86** (2012) 014907, arXiv:1203.3087 [hep-ex].

[63] S. Cao and X.-N. Wang, "Jet quenching and medium response in high-energy heavy-ion collisions: a review", *Rept. Prog. Phys.* **84** no. 2, (2021) 024301, arXiv:2002.04028 [hep-ph].

[64] R. Atkin, "Review of jet reconstruction algorithms", *J. Phys. Conf. Ser.* **645** no. 1, (2015) 012008.

[65] M. L. Miller, K. Reygers, S. J. Sanders, and P. Steinberg, "Glauber modeling in high energy nuclear collisions", *Ann. Rev. Nucl. Part. Sci.* **57** (2007) 205–243, arXiv:nucl-ex/0701025.

[66] **CMS** Collaboration, V. Khachatryan *et al.*, "Measurement of inclusive jet cross sections in *pp* and PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV", *Phys. Rev. C* **96** no. 1, (2017) 015202, arXiv:1609.05383 [nucl-ex].

[67] 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* **76** no. 3, (Feb, 2016) .

[68] F.-M. Liu and S.-X. Liu, "Quark-gluon plasma formation time and direct photons from heavy ion collisions", *Physical Review C* **89** no. 3, (Mar, 2014) .

[69] P. Braun-Munzinger, "Quarkonium production in ultra-relativistic nuclear collisions: suppression versus enhancement", *Journal of Physics G: Nuclear and Particle Physics* **34** no. 8, (Jul, 2007) S471–S477.

[70] J. C. Collins, D. E. Soper, and G. F. Sterman, "Factorization of Hard Processes in QCD", *Adv. Ser. Direct. High Energy Phys.* **5** (1989) 1–91, arXiv:hep-ph/0409313.

[71] M. Breidenbach, J. I. Friedman, H. W. Kendall, E. D. Bloom, D. H. Coward, H. DeStaebler, J. Drees, L. W. Mo, and R. E. Taylor, "Observed Behavior of Highly Inelastic Electron-Proton Scattering", *Phys. Rev. Lett.* **23** (Oct, 1969) 935–939.

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

[73] V. N. Gribov and L. N. Lipatov, "Deep inelastic e p scattering in perturbation theory", *Sov. J. Nucl. Phys.* **15** (1972) 438–450.

[74] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.", *Sov. Phys. JETP* **46** (1977) 641–653.

[75] V. Bertone, S. Carrazza, and J. Rojo, "APFEL: A PDF Evolution Library with QED corrections", *Comput. Phys. Commun.* **185** (2014) 1647–1668, arXiv:1310.1394 [hep-ph].

[76] S. Carrazza, A. Ferrara, D. Palazzo, and J. Rojo, "APFEL Web: a web-based application for the graphical visualization of parton distribution functions", *J. Phys. G* **42** no. 5, (2015) 057001, arXiv:1410.5456 [hep-ph].

[77] M. Cacciari, M. Greco, and P. Nason, "The P(T) spectrum in heavy flavor hadroproduction", *JHEP* **05** (1998) 007, arXiv:hep-ph/9803400.

[78] M. Cacciari, S. Frixione, and P. Nason, "The p(T) spectrum in heavy flavor photoproduction", *JHEP* **03** (2001) 006, arXiv:hep-ph/0102134.

[79] M. Cacciari and P. Nason, "Is there a significant excess in bottom hadroproduction at the Tevatron?", *Phys. Rev. Lett.* **89** (2002) 122003, arXiv:hep-ph/0204025.

[80] S. Frixione and P. Nason, "Phenomenological study of charm photoproduction at HERA", *JHEP* **03** (2002) 053, arXiv:hep-ph/0201281.

[81] M. Cacciari and P. Nason, "Charm cross-sections for the Tevatron Run II", *JHEP* **09** (2003) 006, arXiv:hep-ph/0306212.

[82] C. Peterson, D. Schlatter, I. Schmitt, and P. M. Zerwas, "Scaling Violations in Inclusive e+ e- Annihilation Spectra", *Phys. Rev. D* **27** (1983) 105.

[83] M. G. Bowler, "e+ e- Production of Heavy Quarks in the String Model", *Z. Phys. C* **11** (1981) 169.

[84] M. Lisovyi, A. Verbytskyi, and O. Zenaiev, "Combined analysis of charm-quark fragmentation-fraction measurements", *EPJ Web Conf.* **120** (2016) 03002.

[85] R. Scheibl and U. W. Heinz, "Coalescence and flow in ultrarelativistic heavy ion collisions", *Phys. Rev. C* **59** (1999) 1585–1602, arXiv:nucl-th/9809092.

[86] V. Greco, C. M. Ko, and P. Lévai, "Partonic coalescence in relativistic heavy ion collisions", *Phys. Rev. C* **68** (Sep, 2003) 034904.

[87] A. Andronic, P. Braun-Munzinger, K. Redlich, and J. Stachel, "Decoding the phase structure of QCD via particle production at high energy", *Nature* **561** no. 7723, (2018) 321–330, arXiv:1710.09425 [nucl-th].

[88] A. Andronic, P. Braun-Munzinger, J. Stachel, and M. Winn, "Interacting hadron resonance gas meets lattice QCD", *Phys. Lett. B* **718** (2012) 80–85, arXiv:1201.0693 [nucl-th].

[89] **ALICE** Collaboration, B. Abelev *et al.*, "Centrality dependence of π, K, p production in Pb-Pb collisions at $\sqrt{s_{NN}}$ = 2.76 TeV", *Phys. Rev. C* **88** (2013) 044910, arXiv:1303.0737 [hep-ex].

[90] **ALICE** Collaboration, B. B. Abelev *et al.*, "K^0_S and Λ production in Pb-Pb collisions at $\sqrt{s_{NN}}$ = 2.76 TeV", *Phys. Rev. Lett.* **111** (2013) 222301, arXiv:1307.5530 [nucl-ex].

[91] "Corrigendum to Multi-strange baryon production at mid-rapidity in PbPb collisions at $\sqrt{s_{NN}}$=2.76 TeV [Phys. Lett. B 728 (2014) 216-227]", *Physics Letters B* **734** (2014) 409–410.

[92] **ALICE** Collaboration, B. B. Abelev *et al.*, "$K^*(892)^0$ and (1020) production in Pb-Pb collisions at $\sqrt{sNN}$ = 2.76 TeV", *Phys. Rev. C* **91** (2015) 024609, arXiv:1404.0495 [nucl-ex].

[93] F. Becattini, P. Castorina, J. Manninen, and H. Satz, "The Thermal Production of Strange and Non-Strange Hadrons in e+ e- Collisions", *Eur. Phys. J. C* **56** (2008) 493–510, arXiv:0805.0964 [hep-ph].

[94] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852–867, arXiv:0710.3820 [hep-ph].

[95] M. Bähr, S. Gieseke, M. A. Gigg, D. Grellscheid, K. Hamilton, O. Latunde-Dada, S. Plätzer, P. Richardson, M. H. Seymour, A. Sherstnev, and et al., "Herwig++ physics and manual", *The European Physical Journal C* **58** no. 4, (Nov, 2008) 639707.

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

[97] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, "Event generation with SHERPA 1.1", *JHEP* **02** (2009) 007, arXiv:0811.4622 [hep-ph].

[98] **Sherpa** Collaboration, E. Bothmann *et al.*, "Event Generation with Sherpa 2.2", *SciPost Phys.* **7** no. 3, (2019) 034, arXiv:1905.09127 [hep-ph].

[99] K. C. Zapp, "JEWEL 2.0.0: directions for use", *Eur. Phys. J. C* **74** no. 2, (2014) 2762, arXiv:1311.0048 [hep-ph].

[100] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, arXiv:1405.0301 [hep-ph].

[101] C. Oleari, "The POWHEG-BOX", *Nucl. Phys. B Proc. Suppl.* **205-206** (2010) 36–41, arXiv:1007.3893 [hep-ph].

[102] S. Höche, "Introduction to parton-shower event generators", in *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*, pp. 235–295. 2015. arXiv:1411.4085 [hep-ph].

[103] B. Andersson, *The Lund model*, vol. 7. Cambridge University Press, 7, 2005.

[104] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, "Parton Fragmentation and String Dynamics", *Phys. Rept.* **97** (1983) 31–145.

[105] A. Buckley *et al.*, "General-purpose event generators for LHC physics", *Phys. Rept.* **504** (2011) 145–233, arXiv:1101.2599 [hep-ph].

[106] J. R. Christiansen and P. Z. Skands, "String formation beyond leading colour", *Journal of High Energy Physics* **2015** no. 8, (Aug, 2015) .

[107] V. Greco, C. M. Ko, and R. Rapp, "Quark coalescence for charmed mesons in ultrarelativistic heavy ion collisions", *Phys. Lett. B* **595** (2004) 202–208, arXiv:nucl-th/0312100.

[108] **PHENIX** Collaboration, R. Averbeck, "Single leptons from heavy flavor decays at RHIC", *Nucl. Phys. A* **715** (2003) 695–698, arXiv:nucl-ex/0209016.

[109] S. Acharya, F. T.-. Acosta, D. Adamová, J. Adolfsson, M. M. Aggarwal, G. Aglieri Rinella, M. Agnello, N. Agrawal, Z. Ahammed, and et al., "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 Physics* **2018** no. 10, (Oct, 2018) .

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

[111] S. Acharya, D. Adamová, S. P. Adhya, A. Adler, J. Adolfsson, M. M. Aggarwal, G. Aglieri Rinella, M. Agnello, Z. Ahammed, and et al., "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* **79** no. 5, (May, 2019) .

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

[113] C. Bierlich, G. Gustafson, L. Lönnblad, and H. Shah, "The Angantyr model for Heavy-Ion Collisions in PYTHIA8", *JHEP* **10** (2018) 134, arXiv:1806.10820 [hep-ph].

[114] **STAR** Collaboration, L. Zhou, "Measurements of Λ_c^+ and D_s^+ productions in Au+Au collisions at $\sqrt{s_{NN}}$ = 200 GeV from STAR", *Nucl. Phys. A* **967** (2017) 620–623, arXiv:1704.04364 [nucl-ex].

[115] A. Collaboration, "Λ_c^+ production and baryon-to-meson ratios in pp and p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV at the LHC", 2020.

[116] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 tune", *The European Physical Journal C* **74** no. 8, (Aug, 2014) .

[117] M. He and R. Rapp, "Charm-baryon production in proton-proton collisions", *Physics Letters B* **795** (Aug, 2019) 117121.

[118] D. Ebert, R. N. Faustov, and V. O. Galkin, "Spectroscopy and Regge trajectories of heavy baryons in the relativistic quark-diquark picture", *Phys. Rev. D* **84** (2011) 014025, arXiv:1105.0583 [hep-ph].

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

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

[121] A. Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** no. 08, (Aug, 2008) S08003–S08003.

[122] C. Collaboration, "The CMS experiment at the CERN LHC", *Journal of Instrumentation* **3** no. 08, (Aug, 2008) S08004–S08004.

[123] L. Collaboration, "The LHCb Detector at the LHC", *Journal of Instrumentation* **3** no. 08, (Aug, 2008) S08005–S08005.

[124] R. Steerenberg, M. Albert, R. Alemany-Fernández, T. Argyropoulos, E. Bravin, G. Crockford, J.-C. Dumont, K. Fuchsberger, R. Giachino, M. Giovannozzi, G.-H. Hemelsoet, W. Höfle, D. Jacquet, M. Lamont, E. Métral, D. Nisbet, G. Papotti, M. Pojer, L. Ponce, S. Redaelli, B. Salvachua, M. Schaumann, M. Solfaroli Camillocci, R. Suykerbuyk, G. Trad, J. Uythoven, S. Uznanski, D. Walsh, J. Wenninger, and M. Zerlauth, "Operation and performance of the CERN Large Hadron Collider during proton Run 2",. http://cds.cern.ch/record/2696126.

[125] B. Muratori and T. Pieloni, "Luminosity levelling techniques for the LHC", in *ICFA Mini-Workshop on Beam-Beam Effects in Hadron Colliders*, pp. 177–181. 2014. arXiv:1410.5646 [physics.acc-ph].

[126] K. Eskola, "On predictions of the first results from RHIC", *Nuclear Physics A* **698** no. 1, (2002) 78–87. 15th Int. Conf. on Ultra-Relativistic Nucleus-Nucleus Collisions (Quark Matter 2001).

[127] M. Harrison, T. Ludlam, and S. Ozaki, "RHIC project overview", *Nucl. Instrum. Meth. A* **499** (2003) 235–244.

[128] **ALICE** Collaboration, J. Adam *et al.*, "Centrality dependence of the charged-particle multiplicity density at midrapidity in Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}}$ = 5.02 TeV", *Phys. Rev. Lett.* **116** no. 22, (2016) 222302, arXiv:1512.06104 [nucl-ex].

[129] **ALICE** Collaboration, B. B. Abelev *et al.*, "Performance of the ALICE Experiment at the CERN LHC", *Int. J. Mod. Phys. A* **29** (2014) 1430044, arXiv:1402.4476 [nucl-ex].

[130] E. Bruna, A. Dainese, M. Masera, and F. Prino, "Vertex reconstruction for proton-proton collisions in ALICE",. https://cds.cern.ch/record/1225497.

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

[132] W. Blum, L. Rolandi, and W. Riegler, *Particle detection with drift chambers*. Particle Acceleration and Detection. 2008.

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

[134] **ALICE** Collaboration, B. Abelev *et al.*, "Upgrade of the ALICE Experiment: Letter Of Intent", *J. Phys. G* **41** (2014) 087001.

[135] **ALICE** Collaboration, B. Abelev *et al.*, "Technical Design Report for the Upgrade of the ALICE Inner Tracking System", *J. Phys. G* **41** (2014) 087002.

[136] **ALICE** Collaboration, C. Lippmann, "Upgrade of the ALICE Time Projection Chamber",.

[137] "Addendum to the Technical Design Report for the Upgrade of the ALICE Time Projection Chamber", Tech. Rep. CERN-LHCC-2015-002. ALICE-TDR-016-ADD-1, Feb, 2015. https://cds.cern.ch/record/1984329.

[138] "Technical Design Report for the Muon Forward Tracker", Tech. Rep. CERN-LHCC-2015-001. ALICE-TDR-018, Jan, 2015. https://cds.cern.ch/record/1981898.

[139] **ALICE** Collaboration, "Upgrade of the ALICE Readout & Trigger System",.

[140] **ALICE** Collaboration, S. Acharya *et al.*, "Event-shape engineering for the D-meson elliptic flow in mid-central Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV", *JHEP* **02** (2019) 150, arXiv:1809.09371 [nucl-ex].

[141] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Measurement of prompt D^0 meson azimuthal anisotropy in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV", *Phys. Rev. Lett.* **120** no. 20, (2018) 202301, arXiv:1708.03497 [nucl-ex].

[142] **ALICE** Collaboration, S. Acharya *et al.*, "Anisotropic flow of identified particles in Pb-Pb collisions at $\sqrt{s}_{\rm NN} = 5.02$ TeV", *JHEP* **09** (2018) 006, arXiv:1805.04390 [nucl-ex].

[143] B.-W. Zhang, C.-M. Ko, and W. Liu, "Thermal charm production in a quark-gluon plasma in Pb-Pb collisions at s**(1/2)(NN) = 5.5-TeV", *Phys. Rev. C* **77** (2008) 024901, arXiv:0709.1684 [nucl-th].

[144] **ALICE** Collaboration, S. Acharya *et al.*, "$\Lambda_{\rm c}^{+}$ production in Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV", *Phys. Lett. B* **793** (2019) 212–223, arXiv:1809.10922 [nucl-ex].

[145] S. Acharya, F. T.-. Acosta, D. Adamová, A. Adler, J. Adolfsson, M. M. Aggarwal, G. Aglieri Rinella, M. Agnello, N. Agrawal, Z. Ahammed, and et al., "Multiplicity dependence of light-flavor hadron production in pp collisions at s=7TeV", *Physical Review C* **99** no. 2, (Feb, 2019) .

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

[147] **BaBar** Collaboration, B. Aubert *et al.*, "A precision measurement of the Λ_c^+ baryon mass", *Phys. Rev. D* **72** (2005) 052006, arXiv:hep-ex/0507009.

[148] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175.

[149] P. Z. Skands, "Tuning Monte Carlo Generators: The Perugia Tunes", *Phys. Rev. D* **82** (2010) 074018, arXiv:1005.3457 [hep-ph].

[150] **ALICE** Collaboration, S. Acharya *et al.*, "Pseudorapidity distributions of charged particles as a function of mid and forward rapidity mutiplicities in pp collisions at $\sqrt{s}$ = 5.02, 7 and 13 TeV", arXiv:2009.09434 [nucl-ex].

[151] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System", in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pp. 785–794. ACM, New York, NY, USA, 2016.

[152] **ALICE** Collaboration, J. Adam *et al.*, "Pseudorapidity and transverse-momentum distributions of charged particles in proton–proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B* **753** (2016) 319–329, arXiv:1509.08734 [nucl-ex].

[153] **ALICE** Collaboration, J. Adam *et al.*, "*D*-meson production in *p*-Pb collisions at $\sqrt{s_{\rm NN}}$ =5.02 TeV and in pp collisions at $\sqrt{s}$ =7 TeV", *Phys. Rev. C* **94** no. 5, (2016) 054908, arXiv:1605.07569 [nucl-ex].

[154] **LHCb** Collaboration, R. Aaij *et al.*, "Measurement of *b* hadron fractions in 13 TeV *pp* collisions", *Phys. Rev. D* **100** no. 3, (2019) 031102, arXiv:1902.06794 [hep-ex].

[155] **ALICE** Collaboration, "Observation of multiplicity dependence of charmed baryon-to-meson ratios in proton–proton collisions at $\sqrt{s} = 13$ TeV",. (to be published).

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

[157] **ALICE** Collaboration, S. Acharya *et al.*, "Λ_c^+ production in pp and in p-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV", arXiv:2011.06079 [nucl-ex].

[158] S. Acharya, F. Acosta, D. Adamová, S. Adhya, A. Adler, J. Adolfsson, M. Aggarwal, G. Aglieri Rinella, M. Agnello, N. Agrawal, and et al., "Λ_c^+ production in Pb–Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV", *Physics Letters B* **793** (Jun, 2019) 212223.

[159] J. Bergstra, D. Yamins, and D. Cox, "Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures", in *Proceedings of the 30th International Conference on Machine Learning*, S. Dasgupta and D. McAllester, eds., vol. 28 of *Proceedings of Machine Learning Research*, pp. 115–123. PMLR, Atlanta, Georgia, USA, 17–19 jun, 2013.

[160] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions", in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds., pp. 4765–4774. Curran Associates, Inc., 2017. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

[161] G. Collaboration, "GEANT4 Book For Application Developers, Release 10.7." https://geant4-userdoc.web.cern.ch/UsersGuides/ForApplicationDeveloper/fo/BookForApplicationDevelopers.pdf, accessed 2021-04-18.

[162] **DELPHES 3** Collaboration, J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, "DELPHES 3, A modular framework for fast simulation of a generic collider experiment", *JHEP* **02** (2014) 057, arXiv:1307.6346 [hep-ex].

[163] R. Rahmat, R. Kroeger, and A. Giammanco, "The fast simulation of the CMS experiment", *J. Phys. Conf. Ser.* **396** (2012) 062016.

[164] A. Giammanco, "The Fast Simulation of the CMS Experiment", *J. Phys. Conf. Ser.* **513** (2014) 022012.

[165] Chapman, John, Cranmer, Kyle, Gadatsch, Stefan, Golling, Tobias, Ghosh, Aishik, Gray, Heather M., Lari, Tommaso, Pascuzzi, Vincent R., Raine, John A., Rousseau, David, Salamani, Dalila, and Schaarschmidt, Jana, "Fast simulation methods in ATLAS: from classical to generative models", *EPJ Web Conf.* **245** (2020) 02035.

[166] **ATLAS Collaboration** Collaboration, "Deep generative models for fast shower simulation in ATLAS", Tech. Rep. ATL-SOFT-PUB-2018-001, CERN, Geneva, Jul, 2018. https://cds.cern.ch/record/2630433.

[167]

[168] **ATLAS** Collaboration, A. Basalaev and Z. Marshall, "The Fast Simulation Chain for ATLAS", *J. Phys. Conf. Ser.* **898** no. 4, (2017) 042016.

[169] E. Ritsch, "The ATLAS Integrated Simulation Framework",. https://cds.cern.ch/record/1532476.

[170] D. Müller, M. Clemencic, G. Corti, and M. Gersabeck, "ReDecay: A novel approach to speed up the simulation at LHCb", *Eur. Phys. J. C* **78** no. 12, (2018) 1009, arXiv:1810.10362 [hep-ex].

[171] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework", *Nucl. Instrum. Meth. A* **389** (1997) 81–86.

[172] "CMake Software build system." https://cmake.org/, accessed 2021-04-18.

[173] M. Al-Turany *et al.*, "ALFA: The new ALICE-FAIR software framework", *J. Phys. Conf. Ser.* **664** no. 7, (2015) 072001.

[174] F. Group, "FairRoot." https://github.com/FairRootGroup/FairRoot, accessed 2021-04-18.

[175] F. Group, "FairMQ." https://github.com/FairRootGroup/FairMQ, accessed 2021-04-18.

[176] R. Brun, F. Carminati, I. Hrivnacova, A. Morsch, A. Gheata, and B. Volkel, "VMC." https://github.com/vmc-project, accessed 2021-04-18.

[177] R. Brun, F. Carminati, I. Hrivnacova, A. Morsch, A. Gheata, and B. Volkel, "VMC Project." https://vmc-project.github.io/, accessed 2021-04-18.

[178] Volkel, Benedikt, Morsch, Andreas, Hrivnácová, Ivana, Grosse-Oetringhaus, Jan Fiete, and Wenzel, Sandro, "Using multiple engines in the Virtual Monte Carlo package", *EPJ Web Conf.* **245** (2020) 02008.

[179] B. Volkel, "VMCFastSim." https://github.com/benedikt-voelkel/VMCFastSim, accessed 2021-04-18.

[180] B. Volkel, "FastShower." https://github.com/benedikt-voelkel/FastShower, accessed 2021-04-18.

[181] B. Volkel, "GEANT4_VMC multiple engines example." https://github.com/vmc-project/geant4_vmc/tree/master/examples/E03/E03c, accessed 2021-04-18.

[182] M. Cacciari, S. Frixione, N. Houdeau, M. L. Mangano, P. Nason, and G. Ridolfi, "Theoretical predictions for charm and bottom production at the LHC", *JHEP* **10** (2012) 137, arXiv:1205.6344 [hep-ph].