Department of Physics and Astronomy
Heidelberg University

Bachelor Thesis in Physics
submitted by

Simon Groß-Bölting

born in Heidelberg (Germany)

November 2021

Measurement of the Λ_c^+ production in proton-proton collisions for $\Lambda_c^+ \to pK_s^0$ at $\sqrt{s} = 5.02\,\text{TeV}$ with the ALICE detector

This Bachelor Thesis has been carried out by Simon Groß-Bölting at the
Physikalisches Institut of the University of Heidelberg
under the supervision of
Prof. Dr. Silvia Masciocchi

Abstract

In this thesis the Λ_c^+ baryon production cross section at midrapidity in proton–proton (pp) collisions at $\sqrt{s} = 5.02\,\text{TeV}$ is measured with the ALICE detector using the decay channel $\Lambda_c^+ \rightarrow \text{pK}_s^0$. The Λ_c^+ candidates will be reconstructed in the transverse momentum range $0 < p_\text{T} < 24\,\text{GeV/c}$, extending the p_T range down to $p_\text{T} = 0\,\text{GeV/c}$ which has never been done before for pp collisions at $\sqrt{s} = 5.02\,\text{TeV}$ at the LHC. With the Λ_c^+ baryon production cross section and earlier measurements of the D^0 meson production cross section the baryon-to-meson ratio as the Λ_c^+/D^0 ratio is calculated. The resulting Λ_c^+/D^0 ratio is in agreement with previous measurements in pp collisions. The integrated Λ_c^+/D^0 ratio in pp collisions is $\left(\Lambda_c^+/\text{D}^0\right)_{\text{pp}} = 0.47 \pm 0.03(\text{stat.}) \pm 0.04(\text{syst.})$, which is inconsistent with the measurements from e^-e^+ collisions, challenging the assumed universality of charm fragmentation.
The results are compared to those of earlier publications and will serve as a reference for the Λ_c^+ production in nucleus-nucleus collisions in which the Quark-Gluon Plasma is investigated.

Zusammenfassung

In dieser Arbeit wird der Λ_c^+ Produktions Wirkungsquerschnitt in Proton-Proton Kollisionen bei einer Schwerpunktsenergie von $\sqrt{s} = 5.02\,\text{TeV}$ im ALICE Detektor gemessen. Hierbei wird der Zerfallskanal $\Lambda_c^+ \to \text{pK}_s^0$ betrachtet und die Rekonstruktion für Transversalimpulse im Bereich zwischen $0 < p_\text{T} < 24\,\text{GeV/c}$ durchgeführt. Die Erweiterung der Rekonstruktion in solch niedrigen Impulsbereichen $p_\text{T} = 0\,\text{GeV/c}$ wurde bei Proton-Proton Kollisionen bei $\sqrt{s} = 5.02\,\text{TeV}$ am LHC noch nie durchgeführt. Mit dem Wirkungsquerschnitt der Λ_c^+ Baryonen und früheren Messungen des Wirkungsquerschnittes der D^0 Mesonen wird dann das Baryon-zu-Meson Verhältnis als Λ_c^+/D^0 berechnet, um auch diese Messung zu erweitern. Das integrierte Λ_c^+/D^0 Verhältnis ist $\left(\Lambda_c^+/\text{D}^0\right)_{\text{pp}} = 0.47 \pm 0.03(\text{stat.}) \pm 0.04(\text{syst.})$ und ist somit inkonsistent mit den Messergebnissen in e^-e^+ Kollisionen. Diese Inkonsistenz fechtet die Universalität der Charm Fragmentierung an.

Die Ergebnisse dieser Arbeit werden mit denen von früheren Publikationen verglichen. Außerdem dienen die Ergebnisse der Λ_c^+ Baryon Produktion als Referenz, um in Nukleus-Nukleus Kollisionen das dort entstehende Quark-Gluon Plasma zu untersuchen.

Table of Contents

1 Introduction

The purpose of high-energy collisions at the Large Hadron Collider (LHC) is to produce and investigate particles under extreme conditions, like high energy densities and temperatures, and explore the strong interaction, the weak interaction and the electromagnetic interaction. With the gravitational interaction they make up the four fundamental forces which are described by the *Standard Model of particle physics*. The particles investigated in the experiments at the LHC are hadronised quarks, leptons and bosons.

One specific goal is to study the physics of strongly interacting matter, described by *Quantum Chromodynamics (QCD)*, at the highest energy densities reached so far in the laboratory. In such conditions, an extreme state of matter, called *Quark–Gluon Plasma (QGP)*, is formed. Our universe is thought to have been in such a primordial state for the first few millionths of a second after the Big Bang, before quarks and gluons were bound together to form protons and neutrons [1]. Recreating this primordial state of matter in the laboratory and understanding how it evolves will allow us to shed light on questions about how matter is organised and the mechanisms that confine quarks and gluons.

At the LHC, nucleus–nucleus collisions, like lead–lead (Pb–Pb) collisions, are primarily used to study the formation of a QGP. Smaller collision systems, namely proton–proton (pp) and proton–nucleus collisions, can be used as a reference to compare against, as they have lower energy density and are not expected to form a QGP.

In this thesis the production of Λ_c^+ in proton–proton (pp) collisions at $\sqrt{s} = 5.02\,\text{TeV}$ is investigated. The decay channel of interest is the $\Lambda_c^+ \rightarrow \text{pK}_s^0$ and subsequent $\text{K}_s^0 \rightarrow \pi^+\pi^-$ decay. The transverse momentum (p_T)-differential cross section and the baryon-to-meson ratio are calculated. The transverse momentum p_T is the component of the momentum that the Λ_c^+ particle has in the plane transverse to the direction of the colliding beams. Until now the production of the Λ_c^+ in pp collisions could only be investigated in the p_T region ranging from 1–12 GeV/c as shown in Figure 1 [2]. Therefore calculations of the p_T-integrated cross section relied on models to investigate the Λ_c^+ production down to $p_\text{T} = 0\,\text{GeV/c}$. These models are affected by very large uncertainties which means that losing the model dependency would bring a big improvement on the measurement. It was since shown in the analysis of proton–lead (p–Pb) collisions [3] that the usage of a more sophisticated vertexing technique with the KFParticle package combined with the usage of machine learning (ML) improved upon the previous published results, allowing Λ_c^+ baryons to be measured down to $p_\text{T} = 0\,\text{GeV/c}$. Hence the same analysis strategy is adopted here in order to also measure the Λ_c^+ down to $p_\text{T} = 0\,\text{GeV/c}$ in pp collisions, thus removing the dependence on the models.

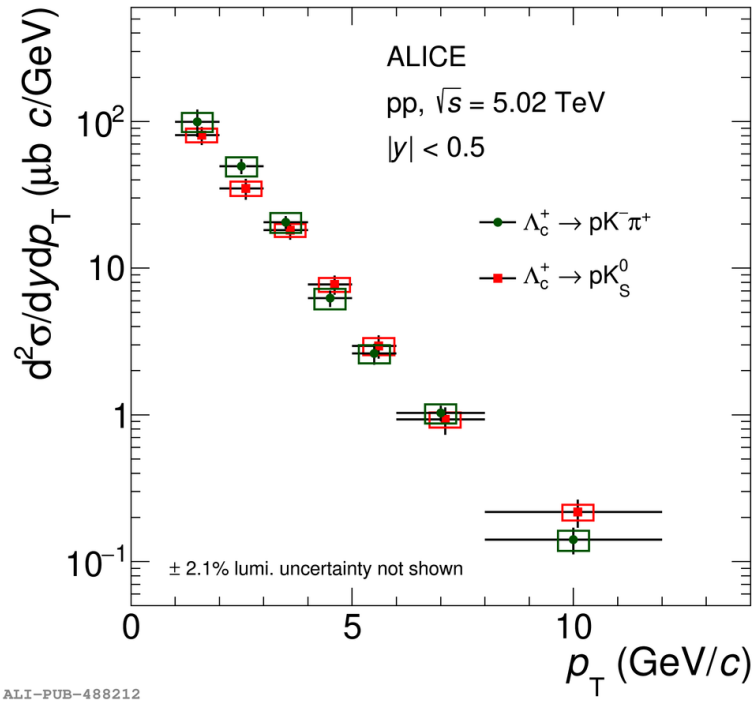


Figure 1: Comparison of the p_T-differential cross section for different Λ_c^+ decay channels [2]

1.1 Charm Production

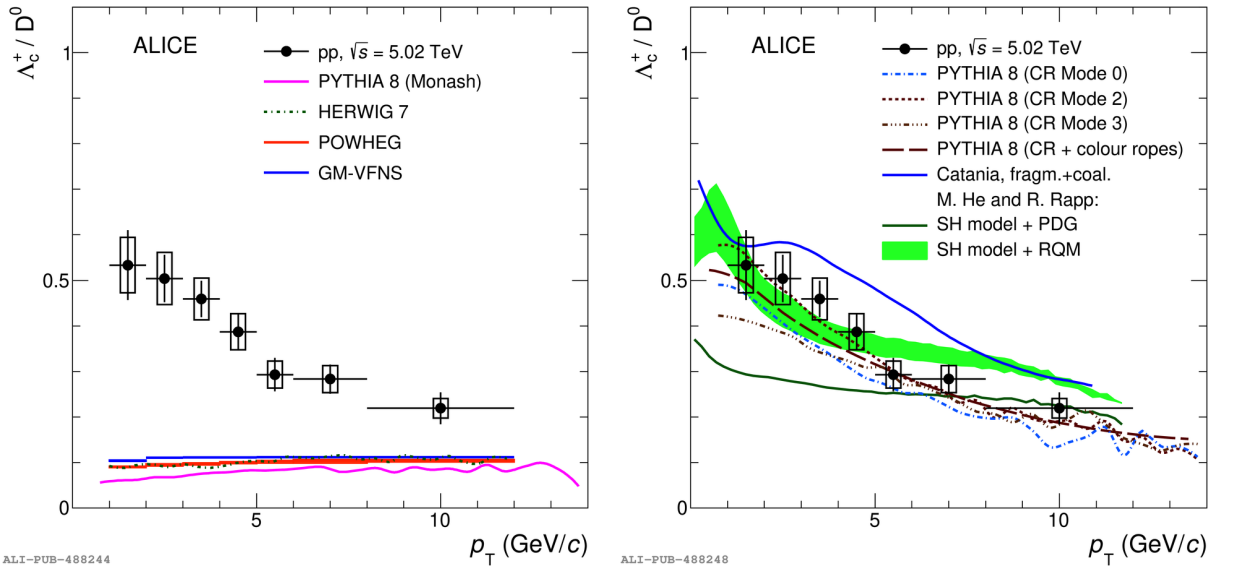
Heavy Flavour (HF) physics is dedicated to particles that consist of at least one heavy quark (a charm (c) or beauty (b)) and other lighter quarks (such as up (u), down (d) and strange (s)). The top quark is not considered as its lifetime is too short to bind into a hadronic state. Because of their large masses ($m_c \approx 1.27\,\mathrm{GeV/c^2}$, $m_b \approx 4.18\,\mathrm{GeV/c^2}$ [4]) heavy quarks are mainly created in hard scattering processes with large momentum transfer $Q^2 > 4m_{b,c}^2$. They are also produced shortly after the collision ($\tau_{b,c} \approx 0.01$–$0.1\,\mathrm{fm/c}$) before the QGP can form ($\tau_{QGP} \approx 0.1$–$1\,\mathrm{fm/c}$). Therefore they can experience the full evolution of the system and interact with the QGP, making heavy quarks the perfect tool to probe and measure the properties of the QGP. In the pp collisions investigated in this thesis no QGP is expected to form. But the measurements in pp collisions will serve as a reference to collision systems with heavy ions.

The probability of a collision producing a specific HF hadron can be calculated with perturbative Quantum Chromodynamics (pQCD). Calculations based on the QCD factorisation approach describe the HF hadron production cross section as a convolution of the parton distribution functions of the colliding hadrons, the parton hard-scattering cross sections and

fragmentation functions [2; 5]:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}p_{\mathrm{T}}}^{pp\to H_q} = \underbrace{f_i(x_1, Q^2) f_j(x_2, Q^2)}_{\text{Parton Distribution Function}} \cdot \overbrace{\frac{\mathrm{d}\sigma^q}{\mathrm{d}p_{\mathrm{T}}}}^{\text{Partonic Cross Section}} \cdot \underbrace{D_{q\to H_q}}_{\text{Fragmentation Function}} \tag{1}$$

The Parton Distribution Function (PDF) describes the probability that a parton (quark or gluon) with a specific fraction x_1 or x_2 of the total momentum (or energy) Q^2 is found in the momentum region of a colour-neutral hadron [6]. The partonic cross section describes the production probability of the parton q itself and the Fragmentation Function (FF) describes the probability that a colour-carrying quark q hadronises into a colour-neutral hadron H_q [6]. The FF is usually thought to be universal among different collision systems.



(a) Models with fragmentation processes tuned on $\mathrm{e^-p}$ collisions

(b) Improved models with enhanced baryon formation

Figure 2: $\Lambda_{\mathrm{c}}^+/\mathrm{D}^0$ for different models and measurements in pp collisions at $\sqrt{s} = 5.02\,\mathrm{TeV}$ [2]

Under the factorisation approach, the $\Lambda_{\mathrm{c}}^+/\mathrm{D}^0$ ratio is sensitive to the FF, as the charm production and PDF terms cancel out. This means that it is particularly useful for studying hadronisation processes. Previous analyses by the ALICE Collaboration, shown in Figure 2, found a substantially higher $\Lambda_{\mathrm{c}}^+/\mathrm{D}^0$ ratio in pp collision than measured at low energies in electron-positron ($\mathrm{e^-e^+}$) and electron-proton ($\mathrm{e^-p}$) collisions [2]. This challenged the assumption that the FFs are universal among different collision systems.

Additionally several Monte Carlo (MC) generators in which the fragmentation process is tuned on $\mathrm{e^-e^+}$ and $\mathrm{e^-p}$ data, significantly underestimate the ratio by a factor of about

5. This underestimation can be seen in Figure 2a). All models (PYTHIA 8 (Monash)[7], HERWIG [8], POWHEG [9] and GM-VFNS [10]) predict a value of the Λ_c^+/D^0 of around 0.1, with a mild p_T dependence. The right panel, Figure 2b), shows a comparison to the models which include processes that enhance baryon production. This modification causes a significant enhancement of the Λ_c^+/D^0 ratio. The models compared there are from the MC generator PYTHIA 8 with tunes that implement Colour Reconnection (CR) beyond the leading-colour approximation, which implement different mechanisms of distribution of the colors of the quarks prior to the hadronisation [11], a phenomenological approach such as the Statistical Hadronisation (SH) model [12] as well as the "Catania" model which implements hadronisation via coalescence and fragmentation [13].
To get a clearer view on how the Λ_c^+/D^0 spectrum evolves the Λ_c^+/D^0 needs to be measured over a wider p_T range. Therefore to complete the results of the previous analyses, the Λ_c^+/D^0 will be calculated down to $p_T = 0\,\mathrm{GeV/c}$, which will be done by coupling the vertexing KFParticle package with a ML algorithm.

1.2 Λ_c^+ Decay

The Λ_c^+ particle contains one up quark, one down quark and a charm quark. It has a relatively short lifetime ($\tau_{\Lambda_c^+} \approx (202.4 \pm 3.1) \cdot 10^{-15}\,\mathrm{s}$ [4]) which corresponds to a proper decay length of $c\tau_{\Lambda_c^+} \approx 60\,\mu\mathrm{m}$, after which it will decay. For this thesis the decay channel $\Lambda_c^+ \to pK_s^0$ and $K_s^0 \to \pi^+\pi^-$ is investigated.

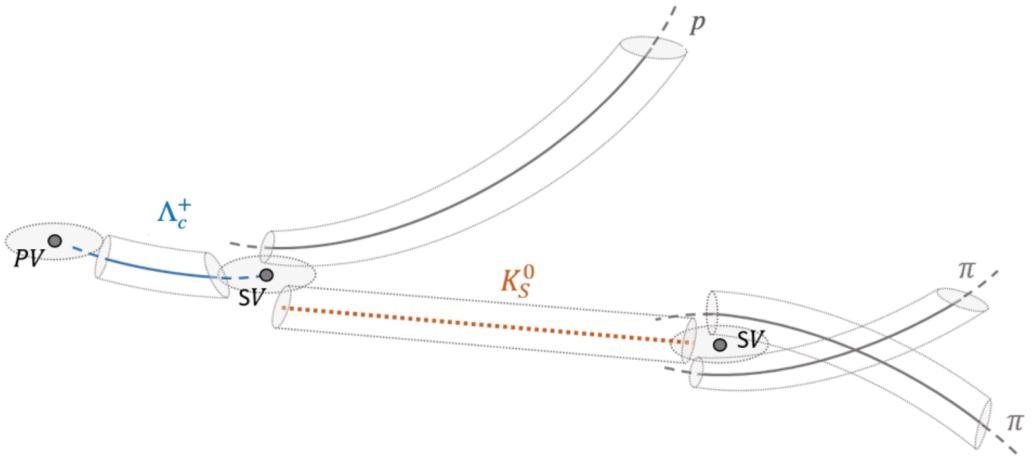


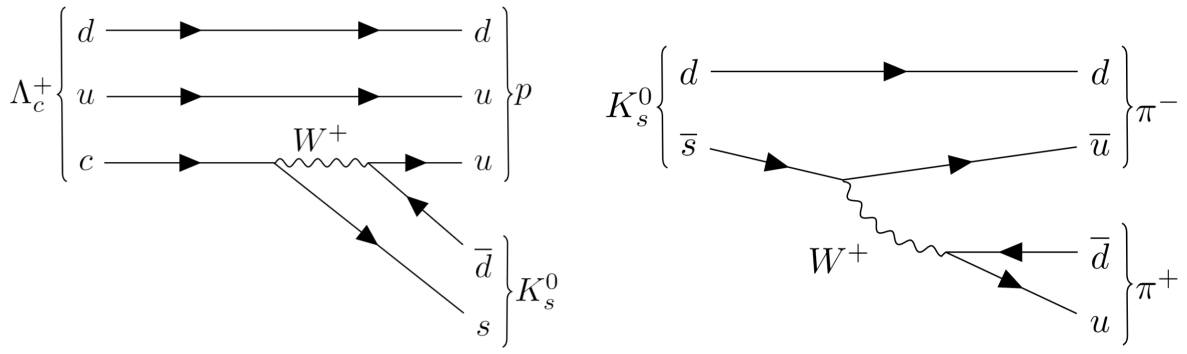
Figure 3: Sketch of the Λ_c^+ decay investigated, including the uncertainties of all particle tracks and on the primary vertex (PV) and secondary vertices (SV) [3]

The decay schematics of the $\Lambda_c^+ \to pK_s^0$ decay is shown in Figure 3. The aim is on Λ_c^+ that

originate from the Primary Vertex (PV), the point where the pp collision occurred, and then the Λ_c^+ decays into a proton (p) and a kaon (K_s^0) at the Secondary Vertex (SV). Another decay vertex is the point where the K_s^0 decays into two pions (π^+ and π^-). The secondary vertices are the decay vertices of short-lived particles and production points of their decay products.

The decay of the Λ_c^+ happens through the weak interaction by emitting a W^+ boson, which will cause the charm quark to transmute into a strange quark, and the W^+ boson will then subsequently decay into a up (u) and a anti-down ($\bar{d}$) quark. These will hadronise with the other leftover quarks into a proton (uud) and a K^0 which consists of a superposition: $\frac{s\bar{d}+d\bar{s}}{\sqrt{2}}$. This decay can be seen in Figure 4a).

In the other diagram (Figure 4b)) the decay of the K_s^0 can be seen. The K_s^0 will also decay via the weak interaction by emitting a W^+ boson, which will cause the anti-strange quark ($\bar{s}$) to transmute into a anti-up quark ($\bar{u}$) which combines with the remaining down quark into a pion (π^-). The W^+ boson will then decay into a anti-down ($\bar{d}$) and up (u) quark which will combine into a pion (π^+). These decays are similar for the respective charge conjugates.



(a) $\Lambda_c^+ \to pK_s^0$ **(b)** $K_s^0 \to \pi^+\pi^-$

Figure 4: Feynman diagrams of the two sequential decays

The branching ratio, which is a measurement of how likely it is that a particle will decay in a specific decay channel, is for the $\Lambda_c^+ \to pK_s^0$ decay BR = $(1.59 \pm 0.08)\%$ and the branching ratio for the $K_s^0 \to \pi^+\pi^-$ decay is BR = $(69.2 \pm 0.05)\%$ [4].

The aim of this analysis is to investigate only Λ_c^+ candidates that are all directly produced in the hard scattering process at the PV. Such candidates are call *prompt* candidates. In contrast to that a charm quark can also come from the decay of a beauty quark and therefore some Λ_c^+ candidates will be not directly produced in the PV. Those candidates are called *feed-down* candidates. Because the number of feed-down candidates is not known the calculations

will be done with both types of candidates and later the estimated number of feed-down candidates will be accounted for. This is done by estimating the fraction of prompt and feed-down candidates through MC simulations of pp collisions.

2 The ALICE Experiment

A Large Ion Collider Experiment (ALICE) is dedicated to study the collisions of nuclei at ultra-relativistic energies provided by the LHC. It is one of the four major experiments at the LHC. ALICE is designed to study strongly interacting matter at the highest energy densities reached so far in the laboratory. The ALICE detector itself is 26 meters long, has a diameter of 16 meters and weighs roughly 10.000 tonnes. The detector sits in a vast cavern 56 meters underground near the town of St. Genis-Pouilly [14].

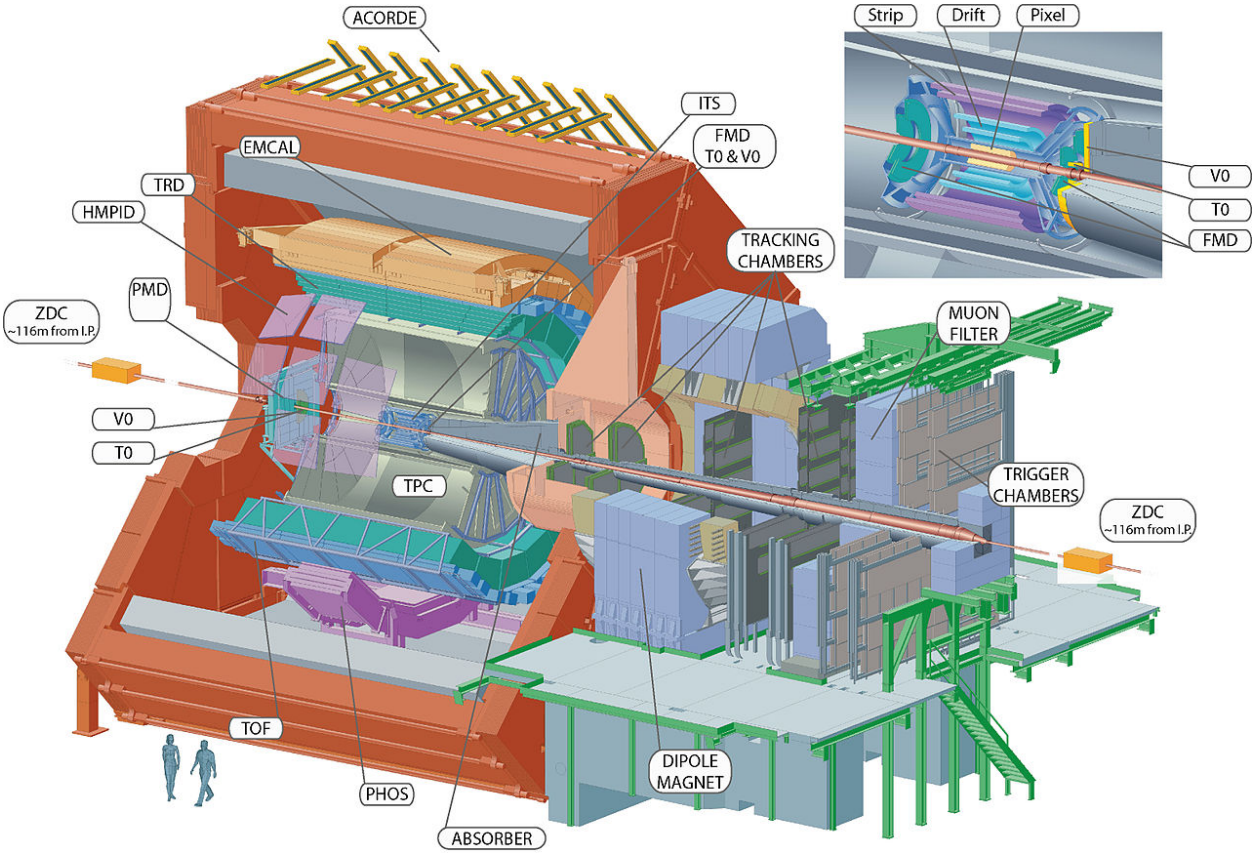


Figure 5: Schematic view of ALICE showing the sub-detector systems of the experiment [15]

The collisions that are studied in the experiments are created with bunches of particles, like lead ions and protons, which are accelerated, travel through the LHC with almost the speed of light and collide at the centre of the detector. The particles created in a collision and their decay products will travel through the detector and interacting with its material they will create a measurable signal.

ALICE uses an operational magnetic field aligned with the beam axis of 0.5 T. Charged particles will bend when experiencing a magnetic field. The bent trajectory is used to

calculate the transverse momentum of the particle itself. It is also used to differentiate the charges of the particles. Opposite charged particles and antiparticles will curve in opposite directions. Compared to other LHC experiments the magnetic field is very low. A low magnetic field is chosen in order to measure particles with very low transverse momentum. If the magnetic field was higher a charged particle with a low p_{T} might bend too much and consequently it might not reach any of the detector layers.

ALICE uses a right-handed orthogonal coordinate system with the z-axis being parallel to the beam pipe. The x and y axis are perpendicular to the beam pipe with the x-axis being aligned with the local horizon pointing to the centre of the accelerator and the y-axis perpendicular to the x-axis pointing upward. The xy-plane is therefore the plane perpendicular to the beam pipe [16].

There are many detectors used in the ALICE experiment (see Figure 5) with each one having the purpose to detect different aspects of the collision making them all complementary. The three main sub-detector systems used in this thesis are the Inner Tracking System (ITS), which is the innermost detector closest to the beam pipe, the Time Projection Chamber (TPC), which is situated around the ITS and the Time-of-Flight detector (TOF) which is located further from the centre.

2.1 Inner Tracking System (ITS)

The Inner Tracking System (ITS) is the innermost detector in ALICE and it is located directly around the beam pipe. Its main purpose is to measure the Primary Vertex (PV) with a resolution better than 100 µm and to reconstruct Secondary Vertices (SV) from particle decays. It is also used for high resolution tracking and momentum measurements. The ITS consists of six cylindrical layers of silicon detectors which have radii ranging from 4 cm to 43 cm.

The two innermost detector layers are Silicon Pixel Detectors (SPD) and are used to provide a first estimate of the PV. This is done by reconstructing the tracks and vertices of the particles using clusters from both detector layers. Sometimes there may occur more than one collision when two particle bunches cross and collide and therefore the PV is defined as the one where most tracks originate from.

The third and fourth layers are Silicon Drift Detectors (SDD) and the last two are double-sided Silicon micro-Strip Detectors (SSD). Those four outer layers are relevant because they will provide extra tracking information and can be used for particle identification (see Section 2.4) via the specific energy loss $\mathrm{d}E/\mathrm{d}x$ measurements [14].

2.2 Time Projection Chamber (TPC)

The Time Projection Chamber (TPC) is the main tracking detector in ALICE and is used to provide precise charged-particle momentum measurements with good two-track separation, vertex determination and Particle Identification (PID) (see Section 2.4). The pseudo-rapidity covered by the TPC is $|\eta| < 0.9$ and it can cover a p_{T} range for PID ranging from about 0.1 GeV/c2 up to 100 GeV/c [17].

The detector contains of a gas mixture situated in a uniform electromagnetic field. Charged particles will ionise the gas in the chamber as they pass through, and the electric field will drift the stripped electrons to the detection site which consists of Multi-Wire Proportional Chambers (MWPC) and the ions to the central electrodes.

The PID information provided by the TPC is based on the specific energy loss ($\mathrm{d}E/\mathrm{d}x$), the charge of a particle and its momentum p as will be further discussed in Section 2.4.

2.3 Time-of-Flight Detector (TOF)

The Time-of-Flight detector (TOF) is made up of Multi-gap Resistive-Plate Chambers (MRPC) covering radii ranging from 370 cm to 399 cm.

In the TOF traversing particles are identified by measuring their particles flight times t over the length of the track's trajectory L. Together with the momentum p of the particle the particle's velocity can be determined by [18]

$$\beta = \frac{v}{c} = \frac{L}{tc} = \frac{1}{\sqrt{\left(\frac{mc}{p}\right)^2 + 1}}. \tag{2}$$

Thus the particle's mass can be calculated from measuring L, t and p:

$$m = \frac{p}{c}\sqrt{\frac{c^2t^2}{L^2} - 1}. \tag{3}$$

The TOF covers a pseudo-rapidity range of $|\eta| < 0.9$ and for Particle Identification it has a time resolution down to 60 ps [19; 20].

2.4 Particle Identification (PID)

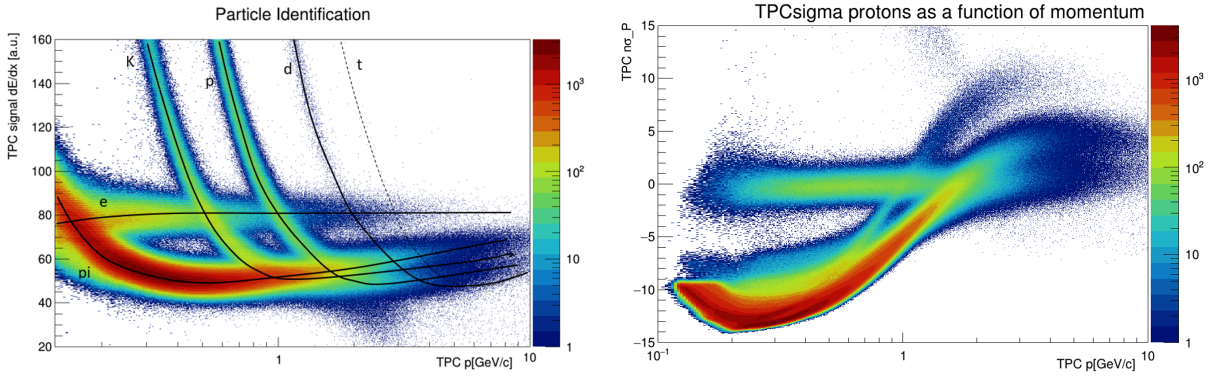
In this analysis three different Particle Identification (PID) variables were used to identify the particles and a systematic study was performed to investigate which of these is the best suited to the $\Lambda_{\mathrm{c}}^{+} \rightarrow \mathrm{pK}_{\mathrm{s}}^{0}$ analysis. The three different PID variables are the n_σ TPC, the

n_σ Combined and the Bayesian PID. These three variables are all referred on the proton, because it is most important to reconstruct a Λ_c^+ and pions are the most abundant and therefore an explicit identification does not bring much better discrimination. They all are also based on the same discriminating variable, the $n_{\sigma_\alpha^i}$.

It is a measurement of the deviation of the measured signal in a given detector α from the expected value for the particle species i. The deviation is measured in factors (n) of σ_α^i, the width of the Gaussian that describes the expected signal distribution in the detector based on the resolution. Therefore with a given signal S_α^i, the reference values $\langle S_\alpha^i \rangle$ and resolution σ_α^i, the deviation can be calculated with:

$$n_{\sigma_\alpha^i} = \frac{S_\alpha^i - \langle S_\alpha^i \rangle}{\sigma_\alpha^i} \tag{4}$$

An example of this PID process with the n_σ is given in the next subsection (n_σ TPC).
The n_σ method is not limited to the TPC but can also be done by other detectors by using their different PID information.



(a) Specific energy loss

(b) n_σ proton

Figure 6: TPC signal and $n_{\sigma_{\mathrm{TPC}}^{\mathrm{p}}}$ as a function of the momentum p

n_σ TPC

The TPC is used for the PID via the particle's specific energy loss dE/dx as a function of momentum p. Each particle species has a unique energy loss as a function of the momentum shown in Figure 6a) as is theoretically described by the Bethe-Bloch-Formula [21]:

$$-\frac{dE}{dx} = \frac{4\pi n z^2}{m_e c^2 \beta^2} \cdot \left(\frac{e^2}{4\pi\epsilon_0}\right)^2 \cdot \left[\ln\left(\frac{2m_e c^2 \beta^2}{I \cdot (1-\beta^2)}\right) - \beta^2\right] \tag{5}$$

Here z is the charge of the particle, I is the average ionisation potential of the target, ϵ_0 the vacuum permittivity, with e and m_e being the electric charge and rest mass of an electron. $\beta = v/c$ with v being the velocity of the particle and c the speed of light and n describes the electron density of the detector material and can be further calculated with $n = \frac{N_A \cdot Z \cdot \rho}{A \cdot M_u}$, here ρ is the density of the detectors material, Z is the atomic number, A the relative atomic mass, N_A the Avogadro number and M_u is the molar mass constant.
For lower momentum the different particle species can be differentiated well but in the higher momentum regions the relativistic rise in the energy loss causes the distributions to all lie in the same region, making the PID more difficult.
The separation of the particle species is done with the $n_{\sigma^{\mathrm{i}}_{\mathrm{TPC}}}$ distribution and an example of the identification of protons with $n_{\sigma^{\mathrm{p}}_{\mathrm{TPC}}}$ can be seen in figure 6b). It shows the deviation of the measured signal from the calibrated TPC response for protons divided by the expected resolution, as a function of momentum. A small $n_{\sigma^{\mathrm{p}}_{\mathrm{TPC}}}$ ($|n| \lessapprox 5$) means that the particle's energy loss was close to the expectation for a proton and a large $n_{\sigma^{\mathrm{p}}_{\mathrm{TPC}}}$ ($|n| \gtrapprox 5$) means that it is most likely not a proton.

n_σ Combined

The n_σ Combined is a combination of the n_σ of protons from the TPC and the TOF. The n_σ of the TOF is based on the time and velocity of the particles:

$$n\sigma^p_{\mathrm{Comb}} = \begin{cases} |n\sigma^p_{\mathrm{TPC}}| & \text{proton tracks only in TPC;} \\ |n\sigma^p_{\mathrm{TOF}}| & \text{proton tracks only in TOF;} \\ \frac{1}{\sqrt{2}}\sqrt{(n\sigma^p_{\mathrm{TPC}})^2 + (n\sigma^p_{\mathrm{TOF}})^2} & \end{cases}$$

When there is only a signal in the TPC then only the n_σ in the TPC will be analysed and when only a signal in the TOF is detected only the TOF will be used. The latter does almost never happen as a proton must pass through the TPC to reach the TOF. In case both detectors have a signal a combination of both detectors will be used.

Bayesian PID

The Bayesian PID uses a different approach to assign a particle species to a detected particle. For a given detector (α) with a Gaussian uncertainty, the conditional probability that a

particle of species H_i will produce a signal S is given by:

$$P(S|H_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}n_\sigma^2} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(S_\alpha^i - \langle S_\alpha^i \rangle)^2}{2{\sigma_\alpha^i}^2}} \tag{6}$$

The probabilities can then be combined considering all detectors where the particle was measured by a product to:

$$P(\vec{S}|H_i) = \prod_{\alpha=\text{ITS,TPC,...}} P_\alpha(S_\alpha|H_i), \tag{7}$$

Now the variable of interest $P(H_i|\vec{S})$, the conditional probability that the particle is of the species H_i given a measured detector signal, can be expressed using Bayes' theorem:

$$P(H_i|\vec{S}) = \frac{P(\vec{S}|H_i)C(H_i)}{\sum_{e,\mu,\pi,...} P(\vec{S}|H_k)C(H_k)} \tag{8}$$

With $C(H_i)$ being the *a priori* probability of measuring the particle species H_i. The priors are defined by the relative abundances of the particle species in data. More details on the Bayesian PID technique are reported in [22].

3 Methods

3.1 Λ_c^+ Candidate Reconstruction and Preselection

A charged particle will leave a signal in the layers of the detectors as it passes through. Combining these signals makes it possible to reconstruct the trajectory (*track*) that the particle followed. When combining the information of different detectors each track can be identified and characterised by the particle's mass M, its momentum p, its energy E and its charge q.

The Λ_c^+ has a very short proper decay length ($c\tau_{\Lambda_c^+} \approx 60\,\mu\text{m}$) causing it to decay before leaving the beam pipe ($r \approx 3\,\text{cm}$). Therefore its decay products are needed to reconstruct the Λ_c^+ baryon. Because the K_s^0 is a neutral particle it can not be detected by the detectors and therefore also its decay products are needed to reconstruct it. The reconstruction is done by finding two pions that originate from the same SV which would be the decay vertex of the kaon. Then the information of the pions is combined to reconstruct the kaon's track. Together with a proton another SV needs to be found where the kaon and proton originated from. This would serve as the decay vertex of the Λ_c^+ and with the information of the proton and kaon also the Λ_c^+ can be reconstructed.

In this analysis the **KFParticle** package [23] is used to reconstruct the full decay chain of the $\Lambda_c^+ \rightarrow pK_s^0$ and the $K_s^0 \rightarrow \pi^+\pi^-$. It uses a Kalman filter approach to improve the reconstruction of the decay chains. This method has been shown to significantly improve the reconstruction of the Λ_c^+ in the analysis of p–Pb collisions [3] and the Ξ_c^0 reconstruction in pp collisions [24].

The Λ_c^+ reconstruction with the KFParticle package was done in advance by others and was saved in a *ROOT* tree which is then used for this analysis. The reconstructed data are candidates that might be of real Λ_c^+ origin (*signal*) or are from combinatorial *background*. To reduce the contamination in the data set preselection criteria are used to remove a large part of candidates with a high probability of being background. Because this will also affect signal candidates the fraction of falsely removed signal candidates must be accounted for with the preselection efficiency (Section 3.1).

The first selection criteria is the rapidity range. Only candidates with a rapidity smaller than $|y| < 0.8$ are accepted. This selection is applied to reduce possible edge effects of the detector, which has an acceptance of $|\eta| < 0.9$. Another preselection criteria is that $0 < \chi^2_{\text{topo}} < 50$. This variable is computed by the KFParticle package in the process of the reconstruction. It is a measurement of the fitting quality of how well the Λ_c^+ candidate is

compatible with the hypothesis of belonging or emerging from the PV.
The other preselection criteria all use the Particle Identification (PID) on the protons and pions. Here the TPC and the TOF are used. In the TPC a tight selection of $3\sigma^i_{\mathrm{TPC}}$ is used, and if the particle raches the TOF and has a TOF signal assigned it has to lay within $5\sigma^i_{\mathrm{TOF}}$. This selection on the deviation to the expected proton and pion values is done to filter out the combinatorial background introduced by wrong article track combination. The preselection combines then to be (with $i \in [p, \pi^+, \pi^-]$):

$$|y| < 0.8\,,\ 0 < \chi^2_{\mathrm{topo}} < 50\,,\ |n^i_{\sigma,\mathrm{TPC}}| < 3 \text{ and } \left(|n^i_{\sigma,\mathrm{TOF}}| < 5 \text{ or no signal in TOF}\right) \quad (9)$$

3.2 Model Training

To separate the Λ_c^+ signal candidates from the combinatorial background a machine learning (ML) approach is used. The model that is used is a Boosted Decision Tree (BDT) trained by the **XGBoost** classifier [25]. It must be noted that the training of the classifiers is done for each p_{T} interval individually.
A ML algorithm learns to differentiate between signal and background candidates by training a network of nodes, with each node having a different purpose on differentiating the input variables into either signal or background. A decision tree has a tree-like node structure, which is optimised to predict the value of a target variable by learning decision rules based on the data features. The boosting referred to in BDT means combining the learning algorithm of multiple decision trees sequentially to create a improved model
The inputs to the model are signal and background samples. The signal is taken from a Monte Carlo (MC) simulation of proton–proton (pp) collisions. The conditions of the ALICE detectors in terms of active channels, gain, noise level and alignment and their evolution with time during the data taking were taken into account in the simulation. The data is generated such that there is at least one Λ_c^+ per event. Therefore the simulated data can be used as a proxy of the real signal.
The background candidates come from the sidebands of the invariant mass distribution of the real data as can be seen in Figure 7. The sidebands have been chosen to be at least 5σ away from the MC invariant mass peak. To get the width of the invariant mass distribution the Λ_c^+ peak in MC was fitted with a Gaussian and the mean and width extracted. Choosing 5σ can be justified by the characteristics of a Gaussian distribution that more than 99.999% of points that follow a normal distribution lie inside the 5σ range. Therefore almost no candidates are expected to lie outside that region, as can be seen in Figure 7.

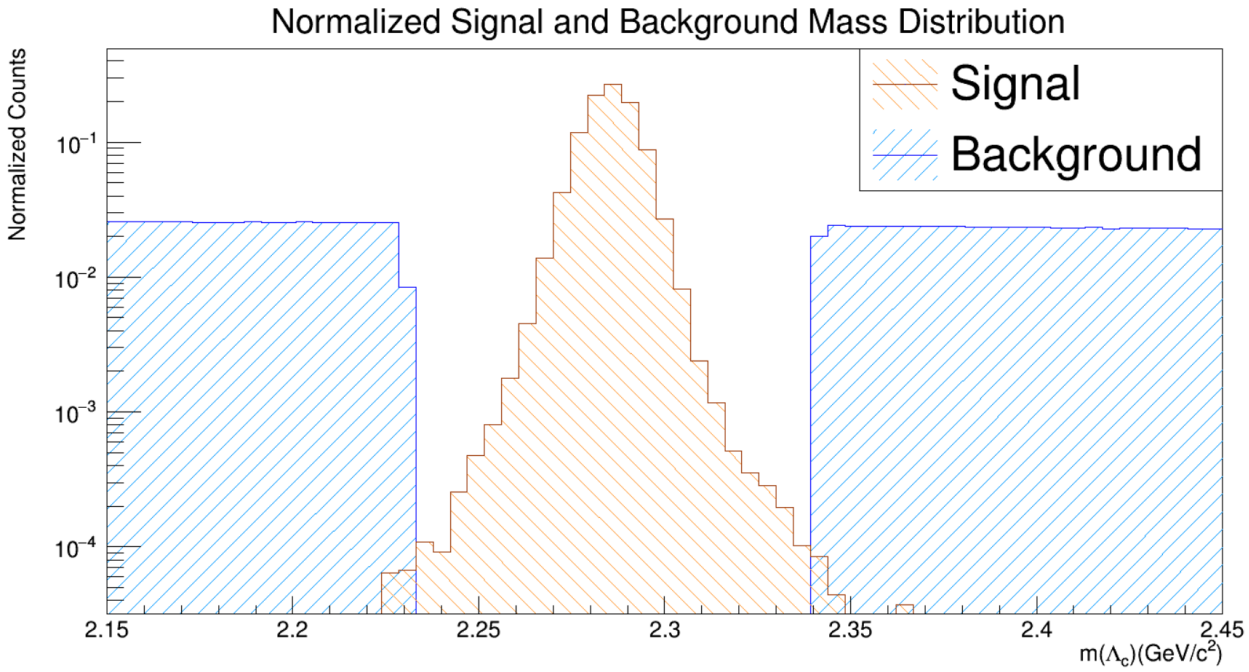


Figure 7: Mass distributions of signal (orange) and background (blue) used in the model training in the p_{T} region 0–1 GeV/c

The data samples containing signal and background candidates are divided into equal sized training and test sets. The model is trained by iterating with different and uncorrelated samples [26] of the training set. After every trial the model will change the selection behavior of each node to increase the quality of the predictions.
The other half, the test set, is used to test the stability of the model and test if any overfitting or underfitting has occurred. Overfitting occurs when the model is trained too closely to the training set and therefore gives many wrong predictions when used on different test sets. Underfitting occurs when the model could not find any useful information in the variables and therefore the predictions made by the model are not as good as they could be.

p_{T}[GeV/c]	0–1	1–2	2–3	3–4	4–5	5–6	6–8	8–12	12–24
Signal	177796	297242	350521	328100	235342	150962	152834	86116	(23124)
Background	355592	594484	701042	328100	146238	50713	84108	22349	(2994)

Table 1: Number of signal and background candidates used for training and testing

For the model training not all background candidates were used. This is done to avoid autocorrelation when the model is applied to the real data sample. Autocorrelation occurs when the model correlates variables with themselves and not with other variables, leading

to wrong predictions. For the signal candidates 60% of the MC data after the preselection was used and for the background candidates only 10% (for $p_T < 6\,\mathrm{GeV/c}$) and 30% (for $p_T > 6\,\mathrm{GeV/c}$) of the data after the preselection were used. The 40% of the MC data that was not used will be used to determine the preselection efficiency as further explained in Section 3.1. In Table 1 the number of signal and background candidates used in the testing and training of the ML model are shown.

3.2.1 Training Variables

The training variables were selected by comparing their distributions between signal and background, as shown in Figure 8. It shows the distribution of values of the different vari-

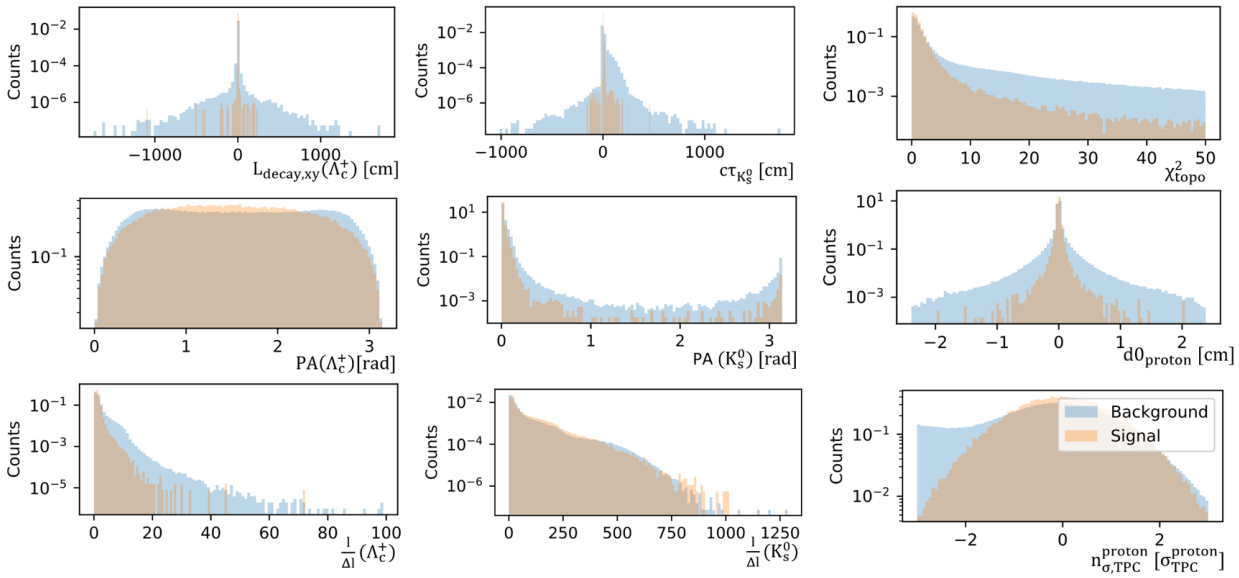
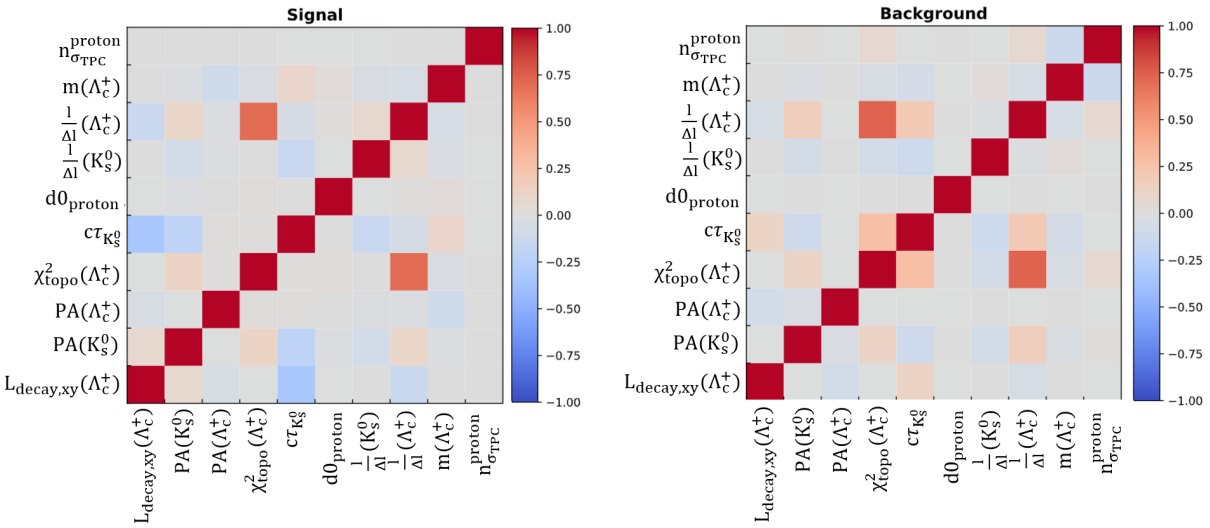


Figure 8: Normalized variable distributions in signal (orange) and background (blue) for the variables used in the model training (from p_T interval 0–1 GeV/c)

ables in signal and background normalized. Variables which show a larger difference between signal and background are more likely to lead to stronger background discrimination in the trained model. The variables selected for the training are all shown here. Other variables that have less difference in both have already been sorted out. Therefore those variables in Figure 8 are the ones with the highest separation. A good separation is visible when there is little overlap between signal and background and a difference is visible in the distributions. This can then be used by the algorithm to find a difference between a real candidate and combinatorial background.



(a) Signal **(b)** Background

Figure 9: Example of the Correlation Matrices from p_T interval 0–1 GeV/c

The different variables describe different features of the reconstructed Λ_c^+ decay (Figure 10). The $l/\Delta l$ describes the distance from the PV to the decay vertex of a particle, normalised by its uncertainty of the distance and this is done for the Λ_c^+ and K_s^0. They can be seen in the bottom left and middle.

Another variable used for the training is the Pointing Angle (PA) of the Λ_c^+ and the K_s^0. The variable distributions can be seen in the middle and middle left of Figure 8. They describe the angle between the momentum vector of the particle and the line that connects the production and decay vertices, as shown in the sketch in Figure 10. Both range from 0 to π and the distributions show only a small difference in signal and background.

The impact parameter of the proton track (d_0) in the transverse plane with respect to the PV is also used and shown in the middle right plane. Because the Λ_c^+ is a short-lived particle the displacement between the PV and its decay vertex is only $c\tau_{\Lambda_c^+} \approx 60\,\mu$m on average and therefore the impact parameter of the proton track to the PV is expected to be small. The distribution is also expected to be narrower in higher p_T regions due to the higher velocity of the protons and their less bent tracks in those region.

The decay lengths of the Λ_c^+ on the xy-plane in the laboratory system in the top left and the proper decay length of the K_s^0 in the middle at the top are used. Both variables seem to have a relatively good separation as the signal distributions are narrow and the background distributions are widespread. This can be explained by the fact that the decay lengths of both particles are calculated wrongly when it comes to the combinatorial background as par-

ticles are combined to other particles which they did not originate from, leading to wrongly calculated energies and decay lengths.

In the top right corner the χ^2_{topo} is shown, which is a product of the reconstruction with the KFParticle package and a measurement of the compatibility of the Λ_c^+ track hypothesis belonging to the PV. When watching closely it shows a slightly higher distribution for signal candidates at low values. This can be explained by the fact that candidates of real Λ_c^+ origin are reconstructed correctly and therefore also really come from the PV whereas background candidates are combined wrongly leading to a false reconstruction of the Λ_c^+ track.

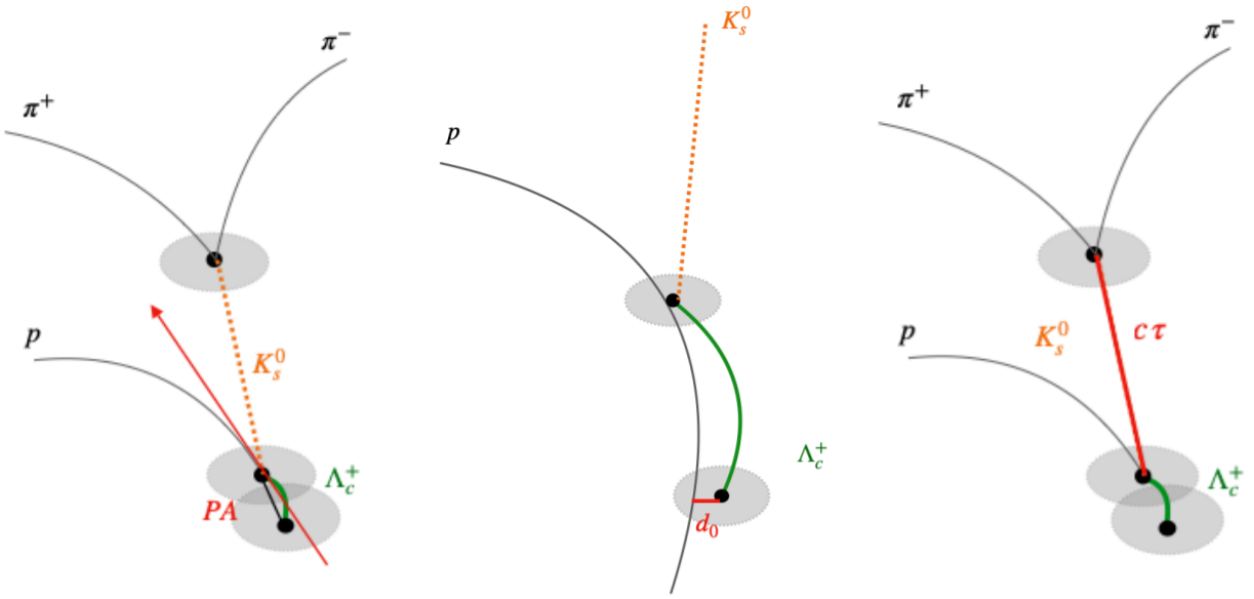


Figure 10: Sketch of the different variables used in the training [3]

In addition to the eight variables a Particle Identification (PID) variable is also used in the training process. This PID variable is the proton n_σ of the TPC, called $n^{proton}_{\sigma,TPC}$, which is shown in the bottom right corner and will be further discussed in Section (3.3). In that section a comparison with two other PID methods is reported.

In order to better differentiate between signal and background, it is important to pay attention to the correlations between the variables. The correlation matrices of the variables for signal and background are shown in Figure 9. A difference in the correlation and anti-correlation between pairs of variables for signal and background can generally be exploited by the model to better differentiate between the two classes. In particular, the selected variables should not have a strong correlation or anti-correlation with the mass of the reconstructed Λ_c^+. If they are correlated then the natural shape of the invariant mass spectrum could be modified, which would lead to an artificial enhancement or reduction of the extracted signal.

3.2.2 Hyperparameter Optimisation

p_{T}[GeV/c]	0–1	1–2	2–3	3–4	4–5	5–6	6–8	8–12	12–24
Learning rate	0.06	0.05	0.04	0.04	0.08	0.06	0.09	0.05	–
Depth	3	3	3	3	3	3	2	2	–
Number of estimators	761	763	759	759	760	761	757	763	–

Table 2: Optimised Hyperparameters

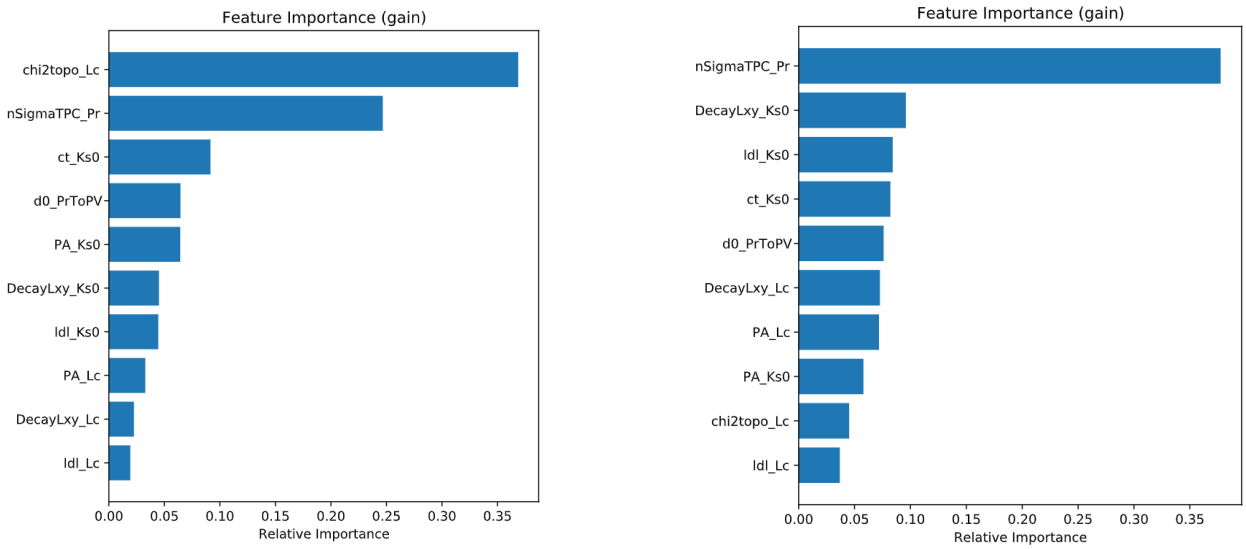
The hyperparameters which define the structure of the model were optimised in every p_{T} interval. This was done to improve the quality of the model predictions. The hyperparameters are the *number of estimators*, the *depth of the model* and the *learning rate*. The optimisation of the hyperparameters happens with a Bayesian iteration procedure by training the model and then checking for the prediction quality. Then small changes in the hyperparameters are made several times, trying to improve the prediction quality with every iteration. The *number of estimators* represents the number of nodes used in the decision tree. The more nodes there are, the more complex the tree becomes. When there are too many estimators the model will overfit, while a too small number of estimators will tend to underfit. The same is true for the *depth of the model*. This parameter describes how many layers the tree will have, with larger numbers of layers leading to a greater complexity in each tree. A more complex tree can be helpful for a training with more complex relationships between variables. For this analysis a maximum depth of three was chosen, as overfitting was observed in deeper trees. For higher p_{T} regions the maximum depth for the optimisation was set even lower because of the reduced background sample available which caused the model to overfit. For the depth and the number of estimators, if the number was set too low the model could not train enough. The range used for optimisation was set to 100–1000 for the number of estimators. For the tree depth, the range was set to 1–3 for low p_{T} regions ($p_{\mathrm{T}} < 6\,\mathrm{GeV/c}$) and 1–2 for higher p_{T}. The third and last optimised hyperparameter is the *learning rate*. It describes how fast the model will change the nodes based on prediction errors. If the learning rate was set too low a local maximum of the prediction quality could be reached causing the model not to reach its full potential. A too-high learning rate would cause the model to overshoot the target. The learning rate was optimised in the range 0.01 to 0.1. The optimised hyperparameters are listed in Table 2. All parameters seem to be consistent for every model and also when comparing them to the hyperparameters used in the earlier p–Pb analysis [3] they are in the same range advocating the models stability.
The model for the p_{T} interval 12–24 GeV/c could not be trained because there were too few

background candidates (see Table 1). Therefore the model from 8–12 GeV/c was used for that p_{T} interval.

3.2.3 Model Prediction

After the model is trained it is then applied to the full data set to differentiate between signal and background candidates and it will predict and return a probability for each candidate to be a signal candidate.

In Figure 11 two examples of the feature importance are shown. These diagrams represent the ranked importance of the features in the trained model when it comes to predicting the probability of a candidate being of real Λ_{c}^{+} origin. It is clear that the Particle Identification variable is important and is having a large impact on the prediction quality, as it is one of the highest ranked features. It was observed that in all p_{T} intervals the PID variable was one of the highest ranked features. Especially in high p_{T} regions the PID is ranked very important to the model in contrast to the other variables as shown in Figure 11b). In contrast, the χ^2_{topo} was seen to have high importance in low p_{T} regions, while having less impact at high p_{T}. The high ranking at low p_{T} demonstrates the gain in performance from using the KFParticle package in this kinematic region.



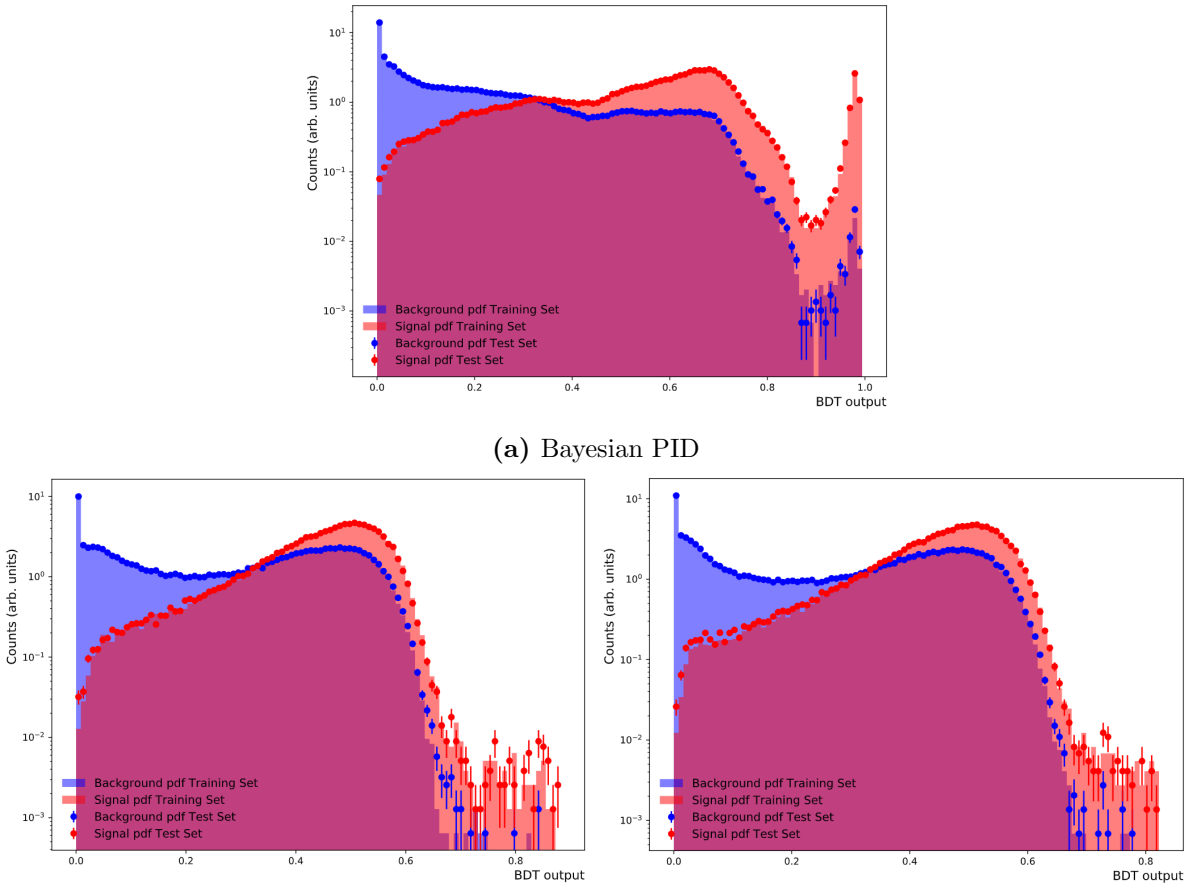
(a) p_{T} interval 0–1 GeV/c

(b) p_{T} interval 8–12 GeV/c

Figure 11: Feature Importance plots from the n_σ TPC only model

The probability that the model will assign each candidate is called the BDT-output probability. A low BDT-output value corresponds to a low likelihood for a candidate to be a real

Λ_c^+ baryon. A high BDT-output value corresponds with a high chance that a reconstructed candidate is a real Λ_c^+. The BDT-output probability from the model are shown in Figure 12, which shows the normalised distribution of BDT-output probabilities that the model assigned for signal in red and background candidates in blue. It shows the training set, as the shaded area, and the test set, as the dots and there is little deviation between those two, indicates a well trained model. The variation for low and high BDT-output values is caused by small differences in the test and training set.



(a) Bayesian PID

(b) n_σ Combined

(c) n_σ TPC

Figure 12: BDT-output plots for different models (p_T interval 0–1 GeV/c)

If the model would be ideal and the features could be separated ideally the background candidates would all have a value of zero and all signal candidates a value of one. But because there is overlap between the features in signal and background the model is not be able differentiate between those two perfectly and therefore the model will overlap their BDT-output probabilities. The model of interest is the one in Figure 12c) as this model will be the one used for further analysis. The rise of the signal and fall of the background

distribution toward higher BDT-output values is a sign that the model has properly learned to differentiate between signal and background candidates. The overall fall for higher BDT-output values is expected because in the case of this analysis the model can not be fully sure that a candidate is a real Λ_c^+.

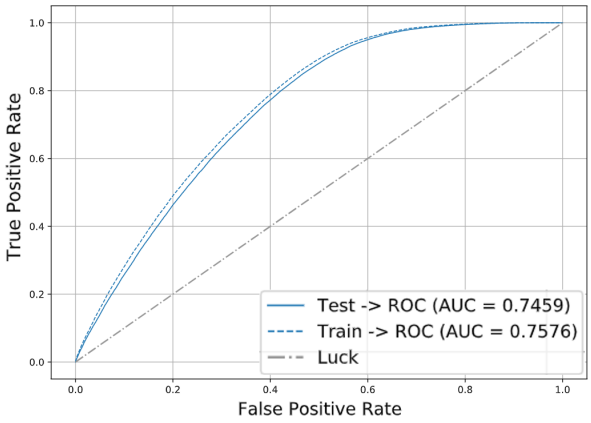
3.3 Systematic Studies of Particle Identification

Since it has been observed that the PID is a highly ranked feature, it was decided to perform a systematic study about the three PID techniques previously explained, to check which one is most suited for this analysis. In particular, the Bayesian PID method had been shown to introduce a bias into the analysis of p–Pb collisions [3] and so its performance was checked against the other PID methods. The training and study of PID was done for every p_T interval individually.

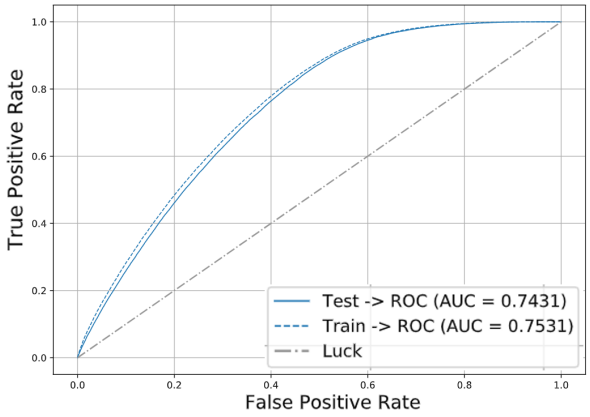
The bias in the Bayesian PID is evident in the behavior shown in Figure 12a), where the BDT output shows a peak for signal at intermediate values before falling and then rising again at higher values. The bias was observed by others in p–Pb as a difference in the efficiency of the data and the MC. This leads to an inconsistency in the description of the selection efficiency in MC, leading to a bias in the final corrected cross section. This behavior was observed in every p_T interval for every model when using the Bayesian PID. The observed behaviour would bias the analysis and therefore the Bayesian PID was excluded from this analysis.

The next step is the comparison of the n_σ Combined and the n_σ TPC. In both cases almost the same BDT-output probability distribution was observed in every p_T interval as can be seen in Figure 12b) and c). Therefore the *Receiver Operating Characteristic curve (ROC-curve)* was used as another way to compare the performance between the two models. The ROC-curve is created by plotting the *true positive rate* against the *false positive rate*. The *AUC-score* is the area underneath the curve and the closer to unity the better the classification. Separate curves are plotted for the test and training sets in order to check for overfitting or underfitting. A large deviation between the two curves would imply issues in the training of the model. All training ROC-curves in Figure 13 and also in other p_T intervals do not deviate much from their test curves. This implies a well trained model for both PID methods in every p_T interval.

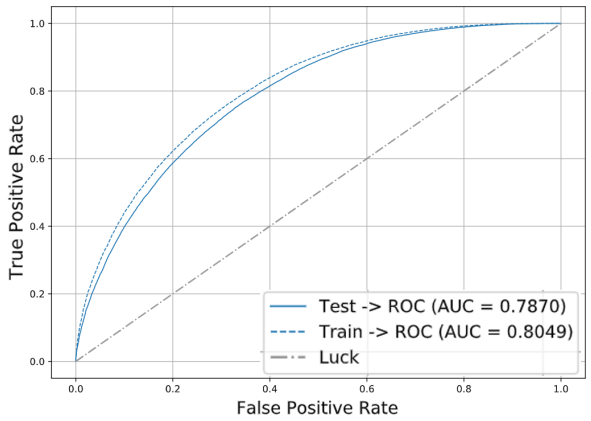
But there was also hardly any deviation in the ROC-curves and AUC-values between the n_σ TPC model and the n_σ Combined model. This implies that the classification via the n_σ Combined method is dominated by the n_σ TPC, with little contribution from the TOF.



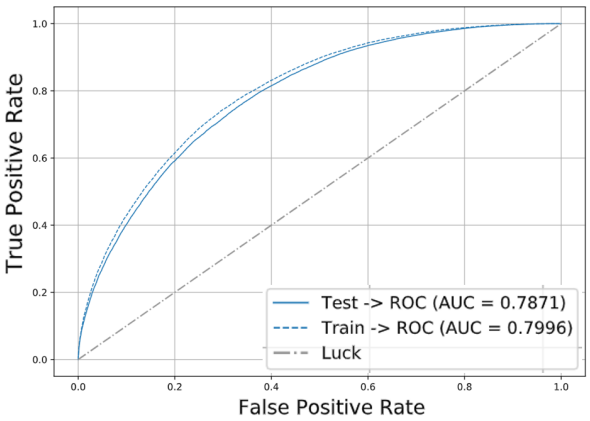
(a) n_σ TPC (p_T interval 0–1 GeV/c)



(b) n_σ Combined (p_T interval 0–1 GeV/c)



(c) n_σ TPC (p_T interval 6–8 GeV/c)



(d) n_σ Combined (p_T interval 6–8 GeV/c)

Figure 13: ROC-curves

This means that the TOF PID information seems not to be helpful for the classification. This consistent behaviour between the n_σ Combined and the n_σ TPC is observed in every p_T interval.

Therefore, in order to simplify the efficiency description in the model and to reduce sources of systematic uncertainties, the n_σ TPC signal alone was used in the model training for this analysis.

4 Analysis

4.1 Model Performance

Considering all AUC values for every p_T interval in Table 3 shows that there is no overfitting or underfitting occurring in the training of the model. A small deviation between the train and test set is expected because the model will not be able to separate between signal and background perfectly. Also it can be observed that for higher p_T intervals the deviation

p_T[GeV/c]	0–1	1–2	2–3	3–4	4–5	5–6	6–8	8–12	12–24
Train-set	0.76	0.77	0.76	0.75	0.76	0.79	0.80	0.86	–
Test-set	0.75	0.76	0.75	0.74	0.75	0.77	0.79	0.82	–

Table 3: AUC-values of the ROC-curves

between the train and test AUC-values rises. This is caused by the limited number of background candidates available in the training. Therefore the model is not able to differentiate between signal and background candidates with the same precision as in lower p_T intervals. Another factor is also the relativistic rise of the TPC signal distribution of the different particle species which is shown in Figure 6. This causes the different particle species to overlap making it harder to differentiate between them. And because the n_σ of the TPC is used as the main feature, as shown in Figure 11b), the differentiation between signal and background is getting more difficult in the higher p_T intervals. However the AUC values also seem to rise for higher p_T which indicates a good separation, introduced by other variables with a relatively low feature importance like $l/\Delta l$ and the decay length of the reconstructed kaon. Therefore the models are trained well and can be used for further analysis.

4.2 Working Point Determination

After the model is trained it is applied to the full data sample for each p_T interval individually. A selection is made on the BDT output score in order to classify signal and background. The BDT selection value is optimised to give the most precise signal extraction possible. The precision is described by the statistical significance:

$$S = \frac{s}{\sqrt{s+b}}, \tag{10}$$

where s is the signal and b the background. To avoid a bias in choosing the selection with the highest significance by accidentally catching a statistical fluctuation the significance of the

invariant mass distribution of the real Λ_c^+ can not be used. Instead the pseudo-significance defined by an approximation of the number of signal and background candidates is used.
To approximate the number of background candidates, an extrapolation of the side bands for the expected invariant Λ_c^+ mass region is used, as shown as an example for the p_T region 0–1 GeV/c in Figure 14a). The side band region contains only candidates that are at least 5σ away from the expected Λ_c^+ mass peak and the extrapolation is done by fitting with a second degree polynomial. The number of background candidates is then given by the integral underneath the extrapolation in the 3σ region of the Λ_c^+ mass peak. This is done for every BDT selection value. In Figure 14a) the background fit may not be perfect but it is not far off the real distribution and is also only used as an approximation of the background.
The estimated number of signal candidates can be approximated using the previously measured p_T differential cross section values [2]. The expected signal for this analysis can be approximated by rearranging the formula of the differential cross section Formula (15):

$$N_{\text{raw}}^{\Lambda_c^+}(p_T) = \frac{2 \cdot \Delta y \cdot \Delta p_T \cdot \text{BR} \cdot L_{\text{int}}}{f_{\text{prompt}}(p_T)} \cdot \frac{\mathrm{d}^2\sigma}{\mathrm{d}p_T \mathrm{d}y} \cdot (Acc \times \epsilon)_{\text{prompt}}(p_T) \tag{11}$$

Here the differential cross section values that are used to approximate the signal in the p_T regions 1–12 GeV/c are measurements from an earlier publication [2]. In the p_T regions 0–1 GeV/c and 12–24 GeV/c no measurements of the differential cross section exist until now, therefore extrapolated values of the differential cross section based on models are used in those regions.
The rapidity range is given by $\Delta y = 1.6$. This is because the preselection filtered out the candidates with a rapidity range of $|y| > 0.8$ caused by the measurement boundaries due to the detector acceptance. The p_T range Δp_T is the width of each p_T interval and the BR is the probability of the subsequent decays of Λ_c^+ and K_s^0 happening ($\text{BR} = (1.59\% \times 69.2\%)$). Another factor is the integrated Luminosity L_{int} which is the luminosity of a pp collision integrated over all analysed events ($N_{\text{events}} \approx 900\,\text{M}$). The integrated luminosity is defined in terms of N_{events} and the inelastic scattering cross section of a pp collision $\sigma_{pp,\text{inel}} = 50.9\,\text{b}$:

$$L_{\text{int}} = L_{pp} \cdot N_{\text{events}} = \frac{N_{\text{events}}}{\sigma_{pp,\text{inel}}} = 17.68\,\text{nb}^{-1} \tag{12}$$

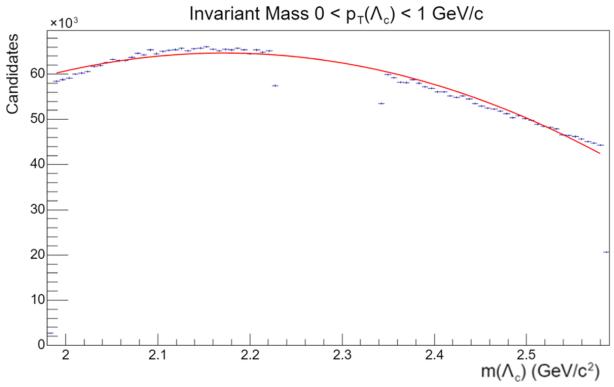
The $f_{\text{prompt}}(p_T)$ is the prompt fraction which is a crucial part because the p_T differential cross section for Λ_c^+ will only be calculated for the prompt candidates. It is the fraction of prompt Λ_c^+ candidates of all Λ_c^+ candidates. A prompt candidate is a Λ_c^+ particle which directly originated from a charm quark produced in the initial collision. By contrast, feed-down candidates are defined as Λ_c^+ that originate from the decay of a beauty hadron. Therefore

the fraction of charm quarks which developed from a beauty quark need to be accounted for and will be get rid of. With the number of feed-down cadidates taken from the MC simulation, the feed-down fraction is calculated as in [27]:

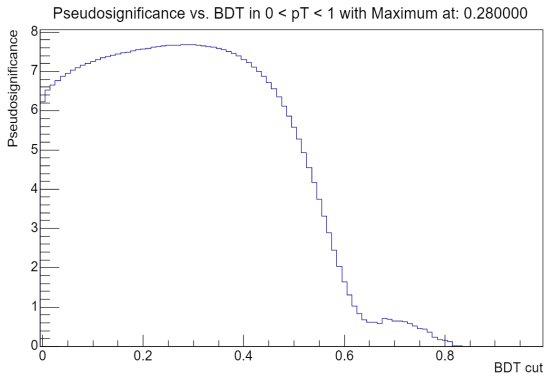
$$f_{\text{prompt}} = 1 - \frac{N^{\Lambda_c,\text{feed-down}}}{N^{\Lambda_c}} \tag{13}$$

$(Acc \times \epsilon)_{\text{prompt}}(p_T)$ is the acceptance times the total efficiency for prompt candidates. The total efficiency contains the preselection efficiency and the BDT efficiency. The acceptance and preselection efficiency is calculated by taking the ratio of the generated and selected MC candidates and the number of all candidates after the preselection. The BDT efficiency is calculated by taking the ratio of the candidates after each BDT selection to the number of candidates after the preselection. With that, $(Acc \times \epsilon)_{\text{prompt}}(p_T)$ is calculated for each BDT selection value. This will also be described in more detail in Section 4.4.

From these ingredients the pseudo-significance depending on the BDT selection value is calculated and an example for the p_T region 0–1 GeV/c is shown in Figure 14b). The specific shape of the pseudo-significance comes from the BDT selection. At first the selection is rejecting mostly background candidates. Meanwhile, the rejection of signal happens at a slower rate, causing the pseudo-significance to rise. The pseudo-significance then reaches a maximum, after which the rate of signal rejection becomes faster than that of the remaining background. This causes the pseudo-significance to fall and eventually reach zero when all candidates are rejected due to a too high BDT selection. The maximum in the pseudo-significance distribution is chosen as the working point. This procedure is repeated for each p_T interval. Table 4 summarises the chosen BDT-output probability selections for all p_T intervals. It shows that the selected values rise with increasing p_T values.



(a) Invariant mass side-bands with second degree polynomial fit



(b) Pseudo-significance as a function of the BDT selection value

Figure 14: Distributions considered for the Working Point (WP) determination

p_{T}[GeV/c]	0–1	1–2	2–3	3–4	4–5	5–6	6–8	8–12	12–24
BDT-output	0.28	0.34	0.31	0.45	0.59	0.74	0.69	0.84	0.67

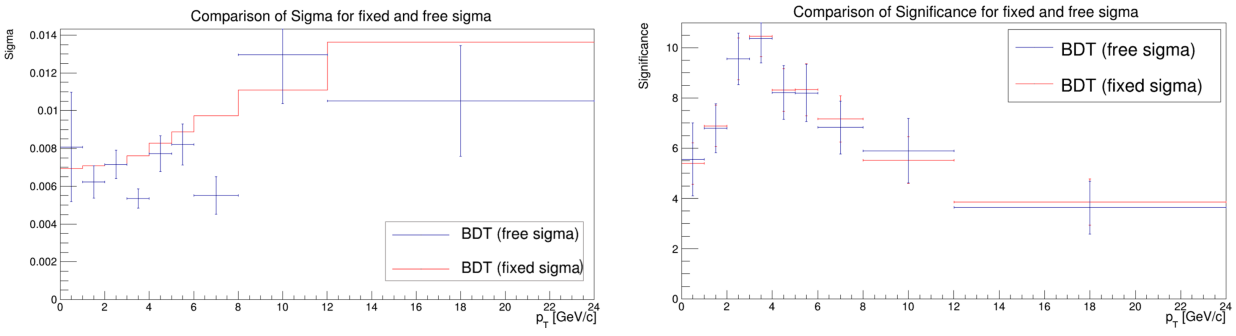
Table 4: BDT-output probability selection

4.3 Signal Extraction

After extracting and applying the best BDT selection values for each p_{T} region the signal extraction for the cross-section calculation of the Λ_{c}^{+} can be performed. To fit the signal a Gaussian function is used and the background is fitted with a second-degree polynomial. Before fitting the signal in data with a Gaussian function, the invariant mass distribution of signal in MC is fitted with a separate Gaussian function and its mean and standard deviation (sigma) are then used for the invariant mass fit of the real data. This is done to get a first estimate on where the peak is expected and to verify whether the Gaussian shape is the correct one to be used. For the fitting procedure the **AliRoot** fitter "*AliHFInvMassFitter*" was used.

As a first estimate of the background the sidebands of the invariant mass distribution are fitted with a second degree polynomial as shown as the gray curve in the invariant mass distributions in Figure 16. Then the estimated background function and the sigma from the MC fit are used for a Gaussian function merged with a second degree polynomial to fit the whole invariant mass distributions. Here the parameters of the background function and the mean values are free parameters. After that the signal and background are extracted by taking the integral in the three-sigma region of the peak. The background is given as the area underneath the polynomial (red curve) and the signal is given by the area enclosed by the Gaussian plus polynomial fit (blue curve) and the background polynomial alone.

As a cross-check of the fit, a second case is considered where also the sigma of the Gaussian fit is left as a free parameter. It was observed that in all cases the "free sigma" adjusted by the fitting function would return a value that was within the fitting uncertainties in agreement with the one given by the MC simulation (Figure 15a)). Even bigger deviations as in the p_{T} regions 2–3 GeV/c and 6–8 GeV/c are within the uncertainties in agreement. This means that on the one hand the fitting procedure is stable and on the other hand that the MC is describing the data well. It can also be seen in Figure 15a) that the "free sigma" fluctuates around the "fixed sigma" values. Therefore in this analysis the "fixed sigma" has been chosen because it better stabilizes the signal extraction. Figure 15b) shows that the



(a) Sigma value per p_{T} interval

(b) Significance per p_{T} interval

Figure 15: Distributions for the comparison of "fixed" and "free sigma" fitting procedures

significance of the two different methods is consistent. All p_{T} intervals have a significance higher than three, this indicates a statistically precise result.
The invariant mass fits are displayed in Figure 16. For every plot the signal *"S"*, the background in the 3σ range of the mean of the invariant mass distribution *"B(3σ)"*, the signal to background ratio *"S/B(3σ)"*and the significance *"Signif(3σ)"* are listed. The uncertainties on the signal and background extraction come from the uncertainties of the fitting procedure. The final polynomial function describing the background (red curve) has no big deviation to the first estimate (grey curve) in almost every p_{T} region. This can be seen as a sign that the fitting procedure is stable.

p_{T} interval 0–1 GeV/c

This is the fist time that it was possible to extract a Λ_{c}^{+} signal down to $p_{\mathrm{T}} = 0\,\mathrm{GeV/c}$ in pp collisions at $\sqrt{s} = 5.02\,\mathrm{TeV}$ at the LHC. It was made possible by the combination of a ML approach together with the KFParticle package.
The signal fit is shown in Figure 16 in the top left corner. The peak of the invariant mass distribution is clearly visible at the expected mass of the Λ_{c}^{+} and the Gaussian function fits the peak well. It has also a significance larger than five.

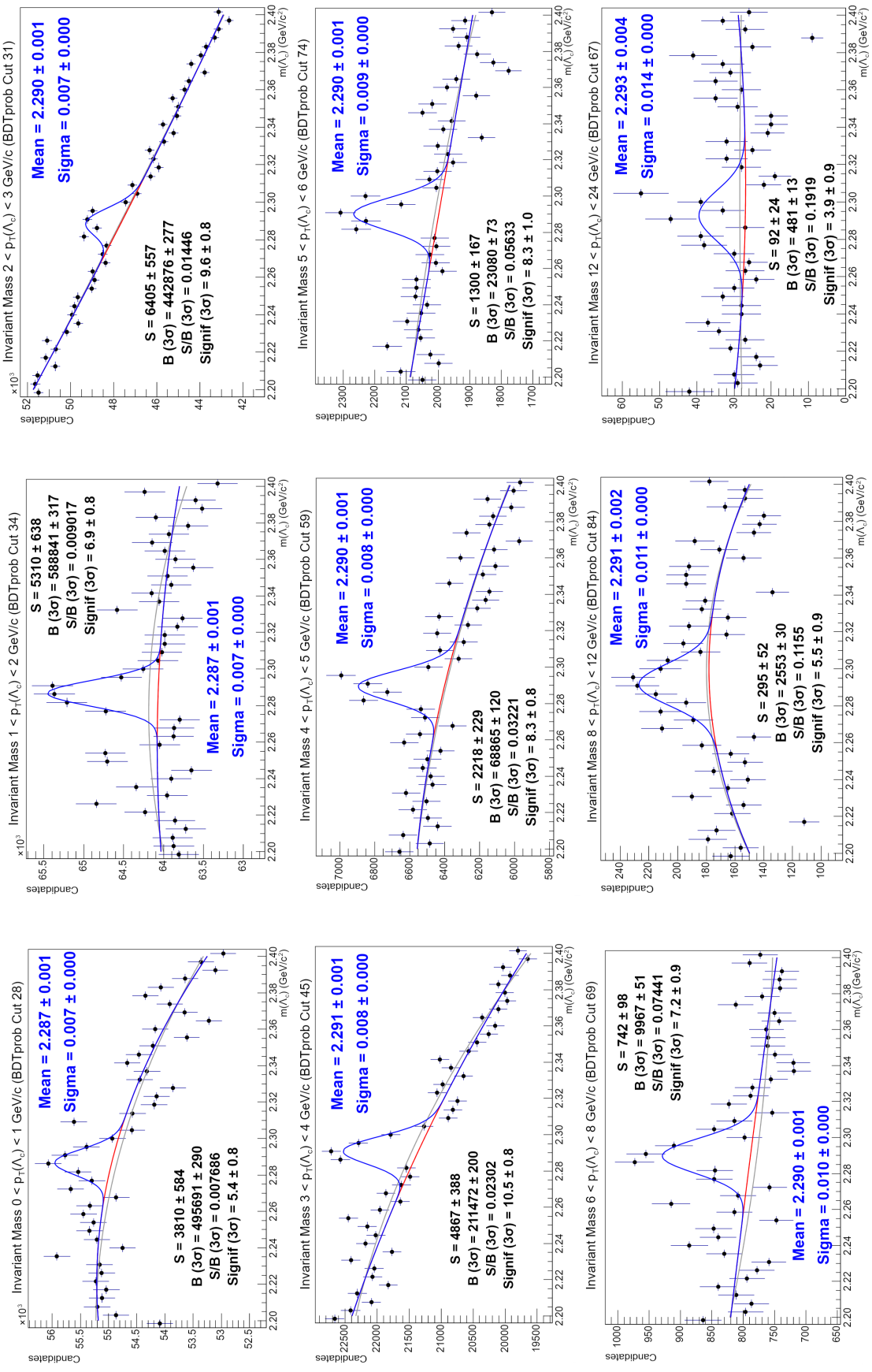


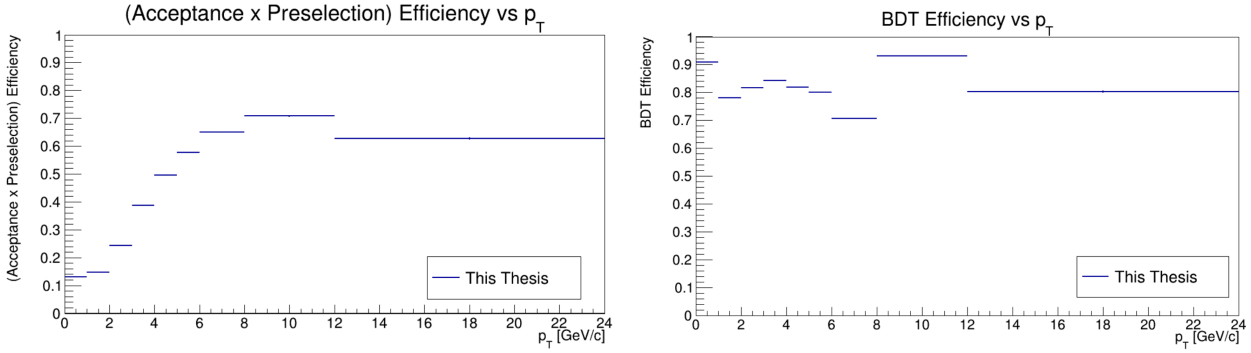
Figure 16: Invariant mass distributions for each p_{T} region after BDT cut and with Gaussian Λ_{c}^{+} mass and background fit

4.4 Efficiency Correction

The efficiency correction accounts for the fraction of candidates that have been incorrectly filtered out in the process of the preselection and for the candidates that were rejected due to the BDT probability selection applied in the analysis. The efficiencies are calculated using Λ_c^+ candidates from MC. The 40 % of the MC that had not been previously used for the training and testing are chosen for this, in order to avoid a bias from reusing the same MC candidates.

Preselection Efficiency

The preselection efficiency is given by the fraction of the generated candidates that remain after the preselection. Therefore the preselection efficiency is influenced by the acceptance and the detector efficiency. The preselection efficiencies for every p_T region are shown in Figure 17a). It can be seen that for low p_T the preselection efficiency is rather low. This is caused by the acceptance because a low p_T particle will bend and can go out of the acceptance more easily than a high p_T particle. Hence a low acceptance reduces the efficiency at low p_T. The efficiency rises towards higher p_T. This behavior of a low efficiency for low p_T and then the rise for higher p_T was expected as it was also observed in [2]. The one aspect that was not expected is the small drop in the p_T region 12–24 GeV/c. This small but noticeable effect could be due to fluctuations because of the small MC sample in this p_T region.



(a) Acceptance × Preselection Efficiency

(b) BDT Efficiency

Figure 17: Efficiency vs. p_T plots

BDT Efficiency

The BDT efficiency corrects for the loss of signal candidates that have been rejected and declared as background by the applied BDT selection criteria. To compute the BDT efficiency the number of MC candidates after the BDT selection used as the WP is divided by the number of signal candidates after the preselection. The BDT efficiency distribution as a function of p_{T} is shown in Figure 17b). The BDT efficiencies are around $\epsilon_{\mathrm{BDT}} \approx 0.8$.

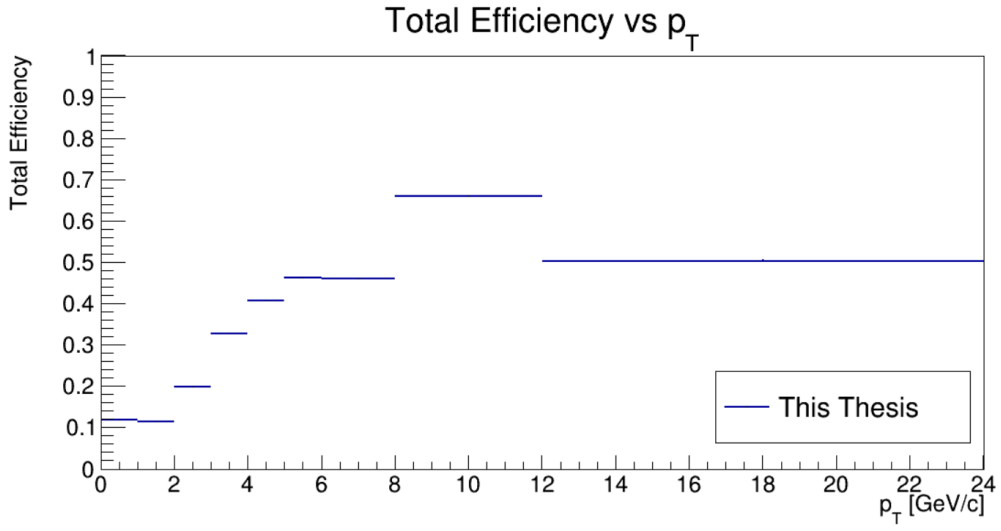


Figure 18: Total acceptance and efficiency for the prompt Λ_{c}^{+} candidates for every p_{T} region

Total Efficiency

The overall acceptance and efficiency $(Acc \times \epsilon)_{\mathrm{prompt}}$ can be calculated by multiplying the acceptance and preselection efficiency by the BDT efficiency. Therefore the total efficiency will be:

$$(Acc \times \epsilon_{\mathrm{tot}})_{\mathrm{prompt}} = (Acc \times \epsilon)_{\mathrm{presel}} \cdot \epsilon_{\mathrm{BDT}} \tag{14}$$

The total efficiency is shown in Figure 18. The efficiency drops for low p_{T} because of the acceptance efficiency. For higher p_{T} the efficiency increases. The high value in the p_{T} region 8–12 GeV/c seems to come from a coincidentally high BDT efficiency and a high preselection efficiency and acceptance. The low efficiency value in the p_{T} region 12–24 GeV/c is coming from both efficiencies being rather low. The low BDT efficiency may be explained by the model, because the chosen model was not trained on data in that region but from the p_{T} region below. Therefore the model predictions may not be as good as they could be if it was trained with data coming from the same p_{T} region.

5 Results and Discussion

5.1 p_{T}-Differential Cross Section

The measurement of the p_{T}-differential Λ_{c}^{+} cross section is important because it represents the likelihood of Λ_{c}^{+} particles being produced in a pp collision in a specific range. In this analysis the production cross section of the prompt Λ_{c}^{+} baryons and their charge conjugate were measured. The decay channel $\Lambda_{\mathrm{c}}^{+} \rightarrow \mathrm{pK}_{\mathrm{s}}^{0}$ with the subsequent decay $\mathrm{K}_{\mathrm{s}}^{0} \rightarrow \pi^{+}\pi^{-}$ with an overall branching ratio of $BR = (1.10\% \pm 0.06\%)$ [4] were analysed at midrapidity in pp collisions at $\sqrt{s} = 5.02\,\mathrm{TeV}$ in the range $0 < p_{\mathrm{T}} < 24\,\mathrm{GeV/c}$. The cross section is calculated via [2]:

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}p_{\mathrm{T}}\mathrm{d}y} = \frac{1}{2} \cdot \frac{f_{\mathrm{prompt}}(p_{\mathrm{T}}) \cdot N_{\mathrm{raw}}^{\Lambda_{\mathrm{c}}^{+}}(p_{\mathrm{T}})}{\Delta y \cdot \Delta p_{\mathrm{T}} \cdot (Acc \times \epsilon)_{\mathrm{prompt}}(p_{\mathrm{T}}) \cdot \mathrm{BR} \cdot L_{\mathrm{int}}} \tag{15}$$

Almost all terms of the equation are the same as described in Section 4.2. The f_{prompt} is the prompt fraction, $N_{\mathrm{raw}}^{\Lambda_{\mathrm{c}}^{+}}$ is the raw yield as extracted in Section 4.3, the Δp_{T} is the width of the analysed p_{T} interval. The Δy is the rapidity range, $(Acc \times \epsilon)_{\mathrm{prompt}}$ is the prompt efficiency correction factor, BR is the branching ratio and L_{int} is the integrated luminosity. The factor $\frac{1}{2}$ accounts for the particles and antiparticles.

p_{T}[GeV/c]	0–1	1–2	2–3	3–4	4–5	5–6	6–8	8–12	12–24
This Analysis [µb c/GeV]	47 ± 7	68 ± 8	46.7 ± 4.2	21.5 ± 1.8	7.7 ± 0.8	4.0 ± 0.5	1.11 ± 0.16	0.20 ± 0.04	0.021 ± 0.006
Reference values [µb c/GeV]	69 ± 14	87 ± 10	43.0 ± 4.1	19.4 ± 1.6	7.0 ± 0.7	2.77 ± 0.24	1.0 ± 0.11	0.179 ± 0.028	0.014 ± 0.006
Deviation	1.8σ	3.1σ	4.1σ	2.5σ	1.7σ	2.8σ	0.9σ	0.7σ	2.6σ

Table 5: p_{T}-differential cross section value comparison of this analysis and the published and extrapolated values (light gray) as reference

The measured p_{T} -differential cross section is shown in Figure 19 and in Table 5. Thanks to the KFParticle package and the ML approach with XGBoost it was possible to extract a signal down to $p_{\mathrm{T}} = 0\,\mathrm{GeV/c}$ and calculate the p_{T} -differential cross section in that region. This was never done before in the evaluation of Λ_{c}^{+} in pp collisions at $\sqrt{s} = 5.02\,\mathrm{TeV}$ at ALICE, because the signal was dominated by combinatorial background at low transverse momenta when using standard selections.

The published and extrapolated values are used as a reference to verify the results of this analysis. The values come from the recent published results by ALICE in pp collisions at $\sqrt{s} = 5.02\,\mathrm{TeV}$ [2]. The uncertainties of the published and extrapolated values are of statistical and systematic origin. The uncertainties of this analysis are only of statistical nature, because no systematic uncertainty analysis has been made. A detailed uncertainty

analysis would therefore be a major point of interest for future analysis. Because the data set analysed in this thesis is the same as was analysed in the reference publication [2], the uncertainties of both analysis are fully correlated. Hence the deviation of the differential cross section values from this analysis (A) to the published and extrapolated values (P) are calculated according to [28] with:

$$n_\sigma = \frac{E_A - E_P}{\sqrt{|\sigma_A^2 - \sigma_P^2|}} \tag{16}$$

With σ_A and σ_P being the uncertainties. The deviations in each p_T region are also shown

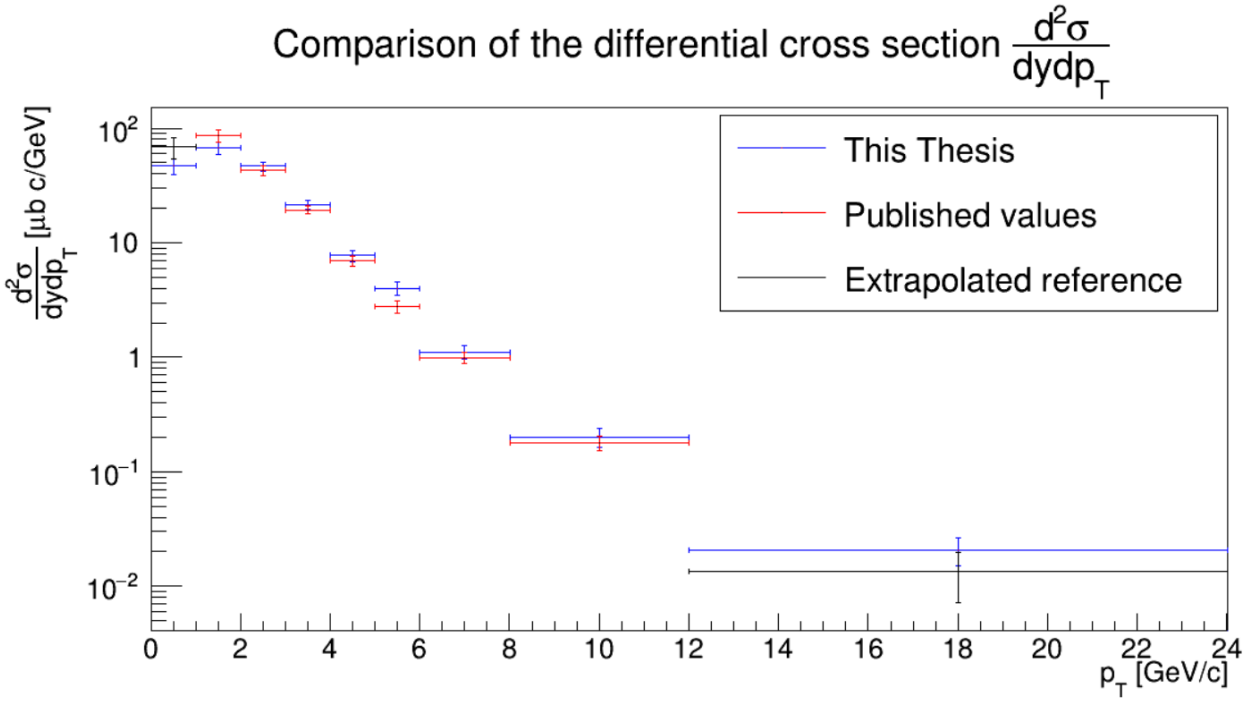


Figure 19: Measured prompt $\Lambda_\mathrm{c}^+ \rightarrow \mathrm{pK}_\mathrm{s}^0$ p_T differential cross section in pp collisions at $\sqrt{s} = 5.02\,\mathrm{TeV}$ for the transverse momentum range $0 < p_\mathrm{T} < 24\,\mathrm{GeV/c}$, compared to previous analysis and extrapolation [2]

in Table 5. The uncertainties of this analysis are underestimated because they are only of statistical nature. Therefore the deviations shown in Table 5 are overestimated.

The results of the differential cross section are within the 3σ of the published values in most p_T intervals. This validates the use of the XGBoost and KFParticle package when analysing outside of the previously visible region. In addition, the new measured values appear to be consistent with the extrapolated cross section values in the p_T region 0–1 GeV/c and 12–24 GeV/c.

Figure 20 shows the ratios of the p_T-differential cross section values of this analysis to the

values of the publication [2]. The ratios for 0–1 GeV/c and 12–24 GeV/c are taken with the extrapolated values and are therefore not as expressive as the ratios to the published values and are used as validity checks. For $p_{\mathrm{T}} < 2\,\mathrm{GeV/c}$ the values of this analysis are underestimated by about 20 %. For p_{T} values greater than $p_{\mathrm{T}} = 2\,\mathrm{GeV/c}$ the values of this analysis are all a factor of about 10 % greater than the published values. Nevertheless the high ratios will need further investigation.

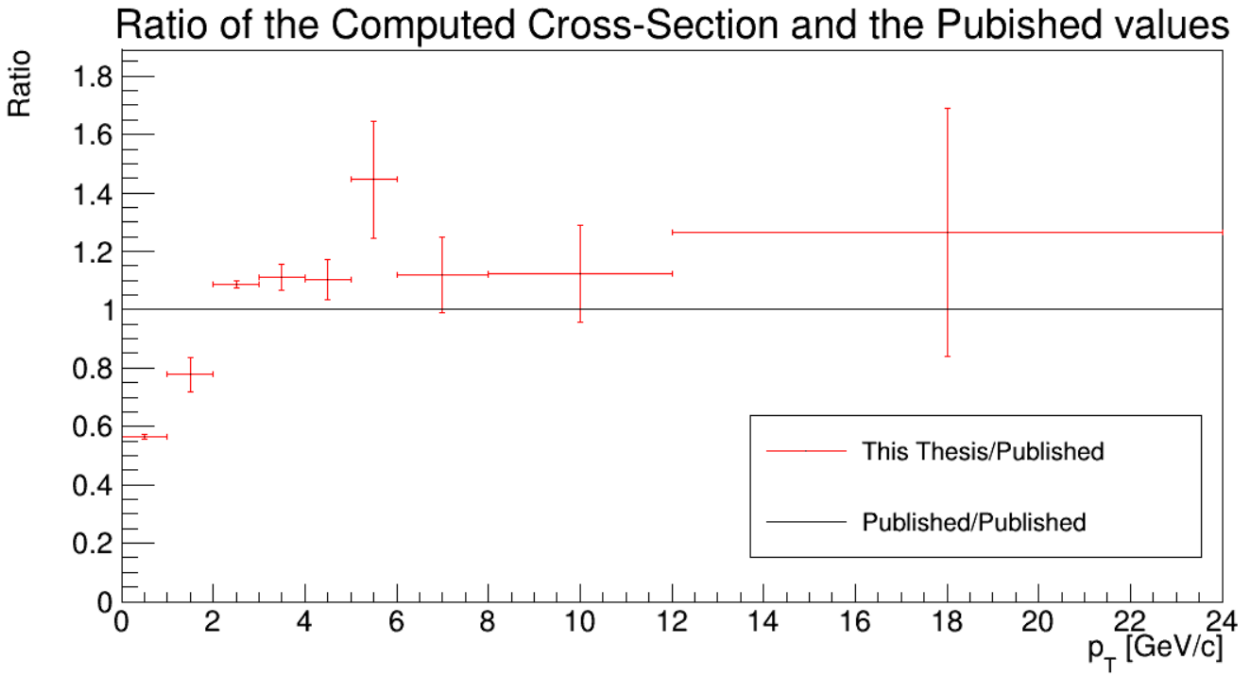


Figure 20: Ratios of the Prompt Λ_{c}^{+} cross section in pp collisions at $\sqrt{s} = 5.02\,\mathrm{TeV}$ for the transverse momentum range $0 < p_{\mathrm{T}} < 24\,\mathrm{GeV/c}$ [2]

It should also be considered that the published and extrapolated values are averages of a differential cross section measurement containing two decay channels. The decay channels are the $\Lambda_{\mathrm{c}}^{+} \rightarrow \mathrm{pK}^{-}\pi^{+}$ and the channel evaluated in this analysis $\Lambda_{\mathrm{c}}^{+} \rightarrow \mathrm{pK}_{\mathrm{s}}^{0}$, as shown in Figure 1. Therefore the comparison of the analysis done in this thesis and the published values in Figure 19 and Table 5 should be handled with care. Figure 1 also shows that the differential cross section values of the decay channel $\Lambda_{\mathrm{c}}^{+} \rightarrow \mathrm{pK}_{\mathrm{s}}^{0}$ would mostly come closer to the values of this analysis reducing the deviation between them. It is crucial that the deviation for the measurements are consistent with the already published values to presume stability of the used methods.

In the p_{T} region 2–3 GeV/c the deviation is about 4σ which is a rather big deviation. It is caused by the properties of the deviation measurement of variables which are fully correlated,

which causes the combined uncertainty to be very small when the individual uncertainties lie close to each other (as described in Formula (16)). And because the uncertainties of this analysis are underestimated the deviation in terms of number of sigma to the published values would shrink. But still a deeper analysis of those p_{T} regions would bring more insight on the correctness of the measured values.

p_{T}-integrated cross section

To evaluate the p_{T}-integrated Λ_{c}^{+} cross section at $\sqrt{s} = 5.02\,\mathrm{TeV}$ the p_{T}-differential cross section needs to be integrated over the transverse momentum. To get the total Λ_{c}^{+} cross section earlier publications had to rely on models to extrapolate the cross section for $p_{\mathrm{T}} < 1\,\mathrm{GeV/c}$ and $p_{\mathrm{T}} > 12\,\mathrm{GeV/c}$. This left the total Λ_{c}^{+} cross section to be [2]:

$$\mathrm{d}\sigma^{\Lambda_{\mathrm{c}}^{+}}_{\mathrm{pp,5.02\,TeV}}/\mathrm{d}y_{|y|<0.5} = 230 \pm 16(\mathrm{stat.}) \pm 20(\mathrm{syst.}) \pm 5(\mathrm{lumi.})^{+5}_{-10}(\mathrm{extrap.})\mu\mathrm{b} \tag{17}$$

With the extension of the p_{T}-differential cross section in the scope of this thesis the p_{T}-integrated cross section for Λ_{c}^{+} in pp collisions can also be measured down to $p_{\mathrm{T}} = 0\,\mathrm{GeV/c}$. This is important because about 20% of the p_{T}-integrated cross section is contained in the p_{T} range 0–1 GeV/c. The p_{T}-integrated Λ_{c}^{+} cross section in the transverse momentum range $0 < p_{\mathrm{T}} < 24\,\mathrm{GeV/c}$ for pp collisions at $\sqrt{s} = 5.02\,\mathrm{TeV}$ in ALICE at midrapidity is:

$$\mathrm{d}\sigma^{\Lambda_{\mathrm{c}}^{+}}_{\mathrm{pp,5.02\,TeV}}/\mathrm{d}y^{0<p_{\mathrm{T}}<24}_{|y|<0.5} = 209 \pm 11(\mathrm{stat.}) \pm 18(\mathrm{syst.})\mu\mathrm{b} \tag{18}$$

The systematic uncertainties for the p_{T} regions 0–1 GeV/c and 12–24 GeV/c are not yet evaluated and included. But that will be done in the future for the finalisation of such important measurements. The Λ_{c}^{+} cross section values of this analysis (18) are within the uncertainties in good agreement with the total p_{T}-integrated cross section (17). The value of the total p_{T}-integrated cross section is larger than the Λ_{c}^{+} cross section evaluated in this analysis. This is due to the relatively big deviation of the differential cross section measurement to the extrapolation in the p_{T} region 0–1 GeV/c.

5.2 Baryon-to-Meson Ratio

The baryon-to-meson ratio is sensitive to the hadronisation process and thus to the charm FF. The FF is thought to be universal among collision systems and is typically tuned on data from $\mathrm{e^{-}e^{+}}$ and $\mathrm{e^{-}p}$ collisions. Therefore measuring and comparing the $\Lambda_{\mathrm{c}}^{+}/\mathrm{D}^{0}$ to models is a direct test of the universality of the FF. The $\Lambda_{\mathrm{c}}^{+}/\mathrm{D}^{0}$ is calculated by taking the ratio

of the Λ_c^+ to the D^0 p_T-differential cross section. The D^0 cross section is taken from [29]. The uncertainties of the cross section measurement of the Λ_c^+ and D^0 are assumed to be uncorrelated. The ratios are compared to already published and extrapolated values from the publication [2]. The Λ_c^+/D^0 ratios are shown in Figure 21 and in Table 6. The uncertainties of the Λ_c^+/D^0 ratios from this thesis are only of statistical nature whereas the uncertainties of the published values are the square root of the quadratic sum of the statistic and systematic uncertainties. All Λ_c^+/D^0 are within the uncertainties in good agreement with the previously published values. The decreasing trend for higher p_T is clearly visible. The extrapolation gives a Λ_c^+/D^0 ratio that is increasing when moving from high to low p_T. In contrast to that the values measured in this analysis are decreasing for $p_T < 1\,\mathrm{GeV/c}$. A drop in the baryon-

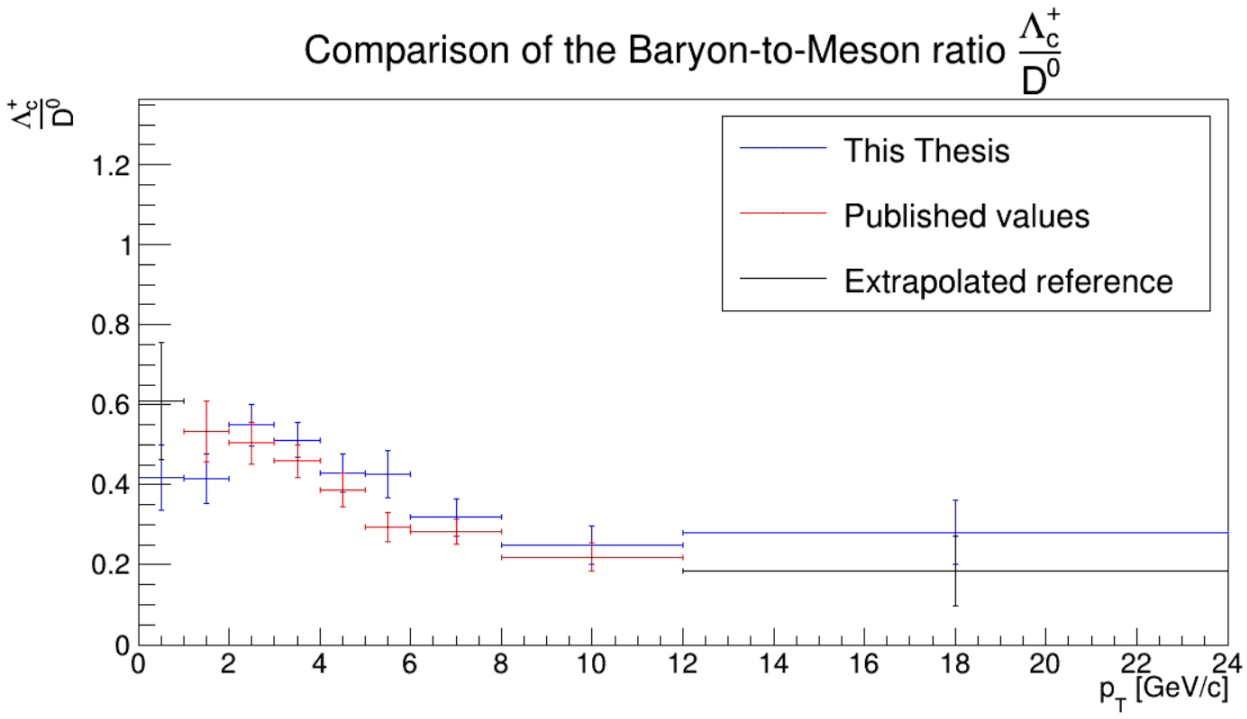


Figure 21: Comparison of the Baryon-to-Meson ratio for this analysis with the previous analysis [2; 3] for pp collisions at $\sqrt{s} = 5.02\,\mathrm{TeV}$ in the transverse momentum range $0 < p_T < 24\,\mathrm{GeV/c}$

to-meson ratio when moving from high to low p_T was also observed in the $(p+\overline{p})/(\pi^+ + \pi^-)$ ratio [30], in the Λ/K_s^0 ratio in pp collisions as well as in p–Pb collision [31; 32] and in the Λ_c^+/D^0 ratio in p–Pb collisions at $\sqrt{s} = 5.02\,\mathrm{TeV}$ [2; 3], which is also shown in Figure 22. It also shows that the Λ_c^+/D^0 ratio for the p–Pb collisions has its maximum in the p_T range 4–6 GeV/c, whereas in pp collisions the maximum is at about 1–2 GeV/c, therefore the Λ_c^+/D^0 distribution is shifted. Figure 22 shows only the p_T region 0–1 GeV/c from this

analysis and uses in the other p_{T} region the values from earlier publications. The decreasing trend for p_{T} approaching $p_{\mathrm{T}} = 0\,\mathrm{GeV/c}$ therefore needs to be handled with care. A high precision measurement in future analysis especially in the low p_{T} region could confirm the trend.

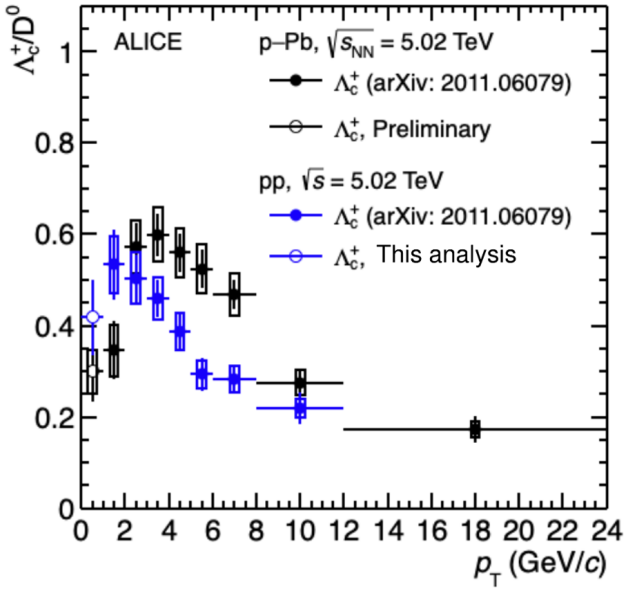


Figure 22: Comparison of the Baryon-to-Meson ratios in pp and p–Pb collisions [2; 3]

p_{T}[GeV/c]	0–1	1–2	2–3	3–4	4–5	5–6	6–8	8–12	12–24
This Analysis	0.42 ± 0.08	0.42 ± 0.06	0.55 ± 0.05	0.51 ± 0.04	0.43 ± 0.05	0.43 ± 0.06	0.32 ± 0.05	0.25 ± 0.05	0.28 ± 0.08
Reference values	0.61 ± 0.14	0.53 ± 0.08	0.5 ± 0.05	0.46 ± 0.04	0.39 ± 0.04	0.29 ± 0.04	0.28 ± 0.03	0.22 ± 0.04	0.18 ± 0.09
Deviation	1.6σ	2.5σ	3.8σ	2.7σ	1.7σ	2.9σ	1.0σ	0.9σ	2.9σ

Table 6: Baryon-to-Meson ratio comparison of this analysis and the published and extrapolated values (light gray) as reference [2]

p_{T}-integrated Baryon-to-Meson ratio

The p_{T}-integrated baryon-to-meson ratio makes it possible to compare the baryon-to-meson production in different collision systems. The p_{T}-integrated baryon-to-meson ratio measured in $\mathrm{e^-e^+}$ collisions at $\sqrt{s} = 10.55\,\mathrm{GeV}$ by CLEO is [33]:

$$\left(\frac{\Lambda_c^+}{D^0}\right)_{\mathrm{e^-e^+}} = 0.119 \pm 0.021(\mathrm{stat.}) \pm 0.019(\mathrm{syst.}) \tag{19}$$

And the p_{T}-integrated baryon-to-meson ratio evaluated in earlier pp collision measurements at $\sqrt{s} = 5.02\,\mathrm{TeV}$ with extrapolations based on models for $p_{\mathrm{T}} < 1\,\mathrm{GeV/c}$ and $p_{\mathrm{T}} > 12\,\mathrm{GeV/c}$

was evaluated to be [2]:

$$\left(\frac{\Lambda_c^+}{D^0}\right)_{\text{pp,extrapolated}} = 0.51 \pm 0.04(\text{stat.}) \pm 0.04(\text{syst.})^{+0.01}_{-0.02} \tag{20}$$

Using the measurements of this thesis in the p_{T} region $p_{\mathrm{T}} < 1\,\mathrm{GeV/c}$ and $p_{\mathrm{T}} > 12\,\mathrm{GeV/c}$ together with the published results between p_{T} 1–12 GeV/c the p_{T}-integrate $\Lambda_{\mathrm{c}}^+/\mathrm{D}^0$ ratio measured in pp collisions at $\sqrt{s} = 5.02\,\mathrm{TeV}$ in the transverse momentum range $0 < p_{\mathrm{T}} < 24\,\mathrm{GeV/c}$ is:

$$\left(\frac{\Lambda_c^+}{D^0}\right)_{\text{pp}} = 0.47 \pm 0.03(\text{stat.}) \pm 0.04(\text{syst.}) \tag{21}$$

The systematic uncertainties come from earlier measurements [2] because no systematic uncertainty analysis has been done in the p_{T} regions $p_{\mathrm{T}} < 1\,\mathrm{GeV/c}$ and $p_{\mathrm{T}} > 12\,\mathrm{GeV/c}$. This will be done in future analysis.

The p_{T}-integrated $\Lambda_{\mathrm{c}}^+/\mathrm{D}^0$ ratio of this thesis (21) is within the uncertainties in good agreement with the published value (20) which depended on models. The p_{T}-integrated $\Lambda_{\mathrm{c}}^+/\mathrm{D}^0$ ratio in pp collisions is regarding the uncertainties more than 5σ away from the ratio measured in $\mathrm{e^-e^+}$ collisions. This implies an enhancement by a factor of about 4 in pp collisions compared to $\mathrm{e^-e^+}$ collisions, indicating that the fragmentation of the charm quark into hardons is dependent on the collision system and is no a universal process. The uncertainties of the p_{T}-integrated $\Lambda_{\mathrm{c}}^+/\mathrm{D}^0$ ratio from this thesis decrease with respect to earlier publications because the uncertainties are inverse proportional to the square root of the number of measurements. Therefore increasing the number of p_{T} regions while decreasing their uncertainties will reduce the overall uncertainty. The new p_{T}-integrated $\Lambda_{\mathrm{c}}^+/\mathrm{D}^0$ ratio in pp collisions allows it to compare it to the $\Lambda_{\mathrm{c}}^+/\mathrm{D}^0$ in p–Pb collisions [3]:

$$\left(\frac{\Lambda_c^+}{D^0}\right)_{\text{pPb}} = 0.43 \pm 0.03(\text{stat.}) \pm 0.05(\text{syst.}) \tag{22}$$

Both $\Lambda_{\mathrm{c}}^+/\mathrm{D}^0$ ratios (21) and (22) are within the uncertainties consistent with each other. Both $\Lambda_{\mathrm{c}}^+/\mathrm{D}^0$ ratios are significantly greater than the $\Lambda_{\mathrm{c}}^+/\mathrm{D}^0$ ratio measured in $\mathrm{e^-e^+}$ collisions. This indicates that the fragmentation fraction of charm quark into baryons is different in pp and p–Pb collisions with respect to $\mathrm{e^-e^+}$ collisions but similar in terms of the total production between the two hadronic collision systems.

6 Conclusion and Outlook

The goal of this thesis was to extend the measurement of the Λ_c^+ production in pp collisions at $\sqrt{s} = 5.02\,\text{TeV}$ with ALICE. This was done by implementing new methods like the KFParticle package and a machine learning approach via XGBoost. The production cross section of the prompt charmed baryon $\Lambda_c^+ \rightarrow \text{pK}_s^0$ and its charge conjugate was measured at midrapidity in the transverse momentum range $0 < p_\text{T} < 24\,\text{GeV/c}$. The measurements of this thesis were compared to previously measured averages in $\Lambda_c^+ \rightarrow \text{pK}_s^0$ and $\Lambda_c^+ \rightarrow \text{pK}^-\pi^+$ [2] showing a good agreement in the full p_T spectrum. The reconstruction via the KFParticle package made it possible to perform the measurement down to $p_\text{T} = 0\,\text{GeV/c}$, which was never done before in pp collisions at $\sqrt{s} = 5.02\,\text{TeV}$ in ALICE before and made it possible to no longer be dependent on model calculations for $p_\text{T} < 1\,\text{GeV/c}$.

To investigate the universality of the FF the Λ_c^+/D^0 baryon-to-meson ratio was calculated with the p_T-differential Λ_c^+ and D^0 cross section in the transverse momentum range $0 < p_\text{T} < 24\,\text{GeV/c}$. It was observed that the measured Λ_c^+/D^0 ratio is in good agreement with the results from earlier publications and that there is large deviation to model predictions made by models tuned on e^-e^+ collisions, underlining the challenge of the assumption that the FF is universal. The p_T-integrated Λ_c^+/D^0 ratio is in good agreement with earlier publications of the Λ_c^+/D^0 ratio in pp collisions that relied on models to extrapolate the Λ_c^+ cross section and with Λ_c^+/D^0 ratios from p–Pb collisions that measured the Λ_c^+ cross section in the same p_T range as in this thesis. This indicated the difference of the FF in e^-e^+ collisions compared to pp and p–Pb collisions.

Because no systematic uncertainty analysis was done in the scope of thesis the uncertainties of all measurements were underestimated. Therefore a wider uncertainty analysis would be a major point of interest for future analysis. Also a comparison to only the decay channel investigated in this thesis would bring more insight on the deviation of the measurements in this thesis and the already published values.

List of Acronyms

ALICE	A Large Ion Collider Experiment
AUC	Area under the curve
BDT	Boosted Decision Tree
BR	Branching Ratio
CR	Colour Reconnection
e^-e^+	electron-positron
e^-p	electron-proton
FF	Fragmentation Function
HF	Heavy Flavour
ITS	Inner Tracking System
LHC	Large Hadron Collider
MC	Monte Carlo
ML	machine learning
MRPC	Multi-gap Resistive-Plate Chambers
MWPC	Multi-Wire Proportional Chambers
p–Pb	proton–lead
PA	Pointing Angle
Pb–Pb	lead–lead
PDF	Parton Distribution Function
PID	Particle Identification
pp	proton–proton
pQCD	perturbative Quantum Chromodynamics
PV	Primary Vertex
QCD	Quantum Chromodynamics
QGP	Quark–Gluon Plasma
ROC	Receiver Operating Characteristic curve
SDD	Silicon Drift Detectors
SH	Statistical Hadronisation
SPD	Silicon Pixel Detector
SSD	double-sided Silicon micro-Strip Detectors
SV	Secondary Vertex
TOF	Time-of-Flight detector
TPC	Time Projection Chamber

WP Working Point

References

[1] John Ellis. *From Little Bangs to the Big Bang*, volume 50. IOP Publishing, Nov 2006. doi: 10.1088/1742-6596/50/1/002. URL http://dx.doi.org/10.1088/1742-6596/50/1/002.

[2] ALICE Collaboration. Λ_c^+ *production in pp and in p-Pb collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV*. 2021. URL https://arxiv.org/abs/2011.06079.

[3] Annalena Kalteyer. *Reconstruction of* Λ_c^+ *in p-Pb collisions at* $\sqrt{s_{NN}} = 5.02\,\mathrm{eV}$ *with the ALICE detector*. May 2021. URL https://www.physi.uni-heidelberg.de/Publications/MasterThesis_AnnalenaKalteyer.pdf.

[4] Particle Data Group. *Review of Particle Physics*. August 2020. URL https://doi.org/10.1093/ptep/ptaa104.

[5] John C. Collins, Davison E. Soper, and George Sterman. *Factorization of Hard Processes in QCD*. September 2004.

[6] Andreas Metz and Anselm Vossen. *Parton Fragmentation Functions*. July 2016. doi: 10.1016/j.ppnp.2016.08.003.

[7] Peter Skands, Stefano Carrazza, and Juan Rojo. *Tuning PYTHIA 8.1: the Monash 2013 Tune*. April 2014. doi: 10.1140/epjc/s10052-014-3024-y.

[8] Johannes Bellm et al. *Herwig 7.0 / Herwig++ 3.0 Release Note*. December 2015. doi: 10.1140/epjc/s10052-016-4018-8.

[9] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. *A Positive-Weight Next-to-Leading-Order Monte Carlo for Heavy Flavour Hadroproduction*. July 2007. doi: 10.1088/1126-6708/2007/09/126.

[10] B. A. Kniehl, G. Kramer, I. Schienbein, and H. Spiesberger. $\Lambda_c^\pm$ *production in pp collisions with a new fragmentation function*. April 2020. doi: 10.1103/PhysRevD.101.114021.

[11] Jesper R. Christiansen and Peter Z. Skands. *String Formation Beyond Leading Colour*. May 2015. doi: 10.1007/JHEP08(2015)003.

[12] Min He and Ralf Rapp. *Charm-Baryon Production in Proton-Proton Collisions*. February 2019. doi: 10.1016/j.physletb.2019.06.004.

[13] Salvatore Plumari, Vincenzo Minissale, Santosh K. Das, Gabriele Coci, and Vincenzo Greco. *Charmed Hadrons from Coalescence plus Fragmentation in relativistic nucleus-nucleus collisions at RHIC and LHC*. December 2017. doi: 10.1140/epjc/s10052-018-5828-7.

[14] The ALICE Collaboration et al. *The ALICE experiment at the CERN LHC*, volume 3. IOP Publishing, aug 2008. doi: 10.1088/1748-0221/3/08/s08002. URL https://doi.org/10.1088/1748-0221/3/08/s08002.

[15] Pcharito. *Schematic cut-away of ALICE*. Wikipedia, CC BY-SA 3.0. URL https://commons.wikimedia.org/w/index.php?curid=31365856.

[16] L. Betev et al. *Definition of the ALICE coordinate system and basic rules for subdetector components numbering*. 2003. URL https://edms.cern.ch/document/406391/2.

[17] G. Dellacasa et al. *ALICE time projection chamber: Technical Design Report*. Technical design report. ALICE. CERN, Geneva, 2000. URL http://cds.cern.ch/record/451098.

[18] Christian Lippmann. *Particle identification*, volume 666. Elsevier BV, feb 2012. doi: 10.1016/j.nima.2011.03.009.

[19] LHC Experiments Committee. *ALICE Time-Of-Flight system (TOF): Technical Design Report*. Technical design report. ALICE. CERN, Geneva, 2000. URL http://cds.cern.ch/record/430132.

[20] Francesca Carnesecchi. *Performance of the ALICE Time-Of-Flight detector at the LHC*. June 2018. doi: 10.1088/1748-0221/14/06/C06023.

[21] Peter Sigmund. *Particle penetration and radiation effects : general aspects and stopping of swift point charges*. Springer, Berlin New York, 2008. ISBN 3540317139.

[22] ALICE Collaboration. *Particle identification in ALICE: a Bayesian approach*. February 2016. doi: 10.1140/epjp/i2016-16168-5.

[23] Ivan Kisel, Igor Kulakov, and Maksym Zyzak. *Standalone First Level Event Selection Package for the CBM Experiment*, volume 60. 10 2013. doi: 10.1109/TNS.2013.2265276.

[24] ALICE Collaboration. *Measurement of the cross sections of Ξ_c^0 and Ξ_c^+ baryons and branching-fraction ratio BR($\Xi_c^0 \to \Xi^- e_e^+$)/BR($\Xi_c^0 \to \Xi^{-+}$) in pp collisions at 13 TeV*. May 2021.

[25] XGBoost developers. *XGBoost Documentation*. 2021.

[26] F. Pedregosa et al. *Scikit-learn: Machine Learning in Python*, volume 12. 2011.

[27] ALICE Collaboration. Λ_c^+ *production in pp collisions at* $\sqrt{s} = 7$ *TeV and in p-Pb collisions at* $\sqrt{s_{\mathrm{NN}}} = 5.02$ *TeV*. December 2017. doi: 10.1007/JHEP04(2018)108.

[28] Lara De Nardo. *On the Propagation of Statistical Errors*. PHENIX, March 2002. URL https://www.phenix.bnl.gov/WWW/publish/elke/EIC/Files-for-Wiki/lara.02-008.errors.pdf.

[29] Alice Collaboration. *Measurement of* D^0, D^+, D^{*+} *and* D^+_s *production in pp collisions at* $\sqrt{s} = 5.02$ *TeV with ALICE*. January 2019. doi: 10.1140/epjc/s10052-019-6873-6.

[30] ALICE Collaboration. *Multiplicity dependence of charged pion, kaon, and (anti)proton production at large transverse momentum in p-Pb collisions at* $\sqrt{s_{\mathrm{NN}}} = 5.02$ *TeV*. January 2016. doi: 10.1016/j.physletb.2016.07.050.

[31] ALICE Collaboration. K^0_{S} *and* Λ *production in Pb-Pb collisions at* $\sqrt{s_{\mathrm{NN}}} = 2.76$ *TeV*. July 2013. doi: 10.1103/PhysRevLett.111.222301.

[32] ALICE Collaboration. Multiplicity dependence of pion, kaon, proton and lambda production in p-pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ tev. *Phys. Lett. B 728 (2014) 25-38*, July 2013. doi: 10.1016/j.physletb.2013.11.020.

[33] P. Avery et al. *Inclusive production of the charmed baryon* ${\Lambda_c}^+$ *from* e^+e^- *annihilations at* $\sqrt{s} = 10.55$ *GeV*, volume 43. American Physical Society, Jun 1991. doi: 10.1103/PhysRevD.43.3599. URL https://link.aps.org/doi/10.1103/PhysRevD.43.3599.

Declaration of Authorship

Ich versichere, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den 11.11.2021