Department of Physics and Astronomy
University of Heidelberg

Bachelor Thesis in Physics
submitted by

Christoph Otte

born in Osnabrück (Germany)

2017

Identifying charmed hadrons from b hadron decays at LHCb

This Bachelor Thesis has been carried out by Christoph Otte at
Physikalisches Institut in Heidelberg
under the supervision of
Prof. Dr. Stephanie Hansmann-Menzemer

Abstract

This thesis documents the development of a classification algorithm for the identification of charmed hadrons coming from b decays. For this purpose, several boosted decision trees were trained on 2012 LHCb data from the channels $\Lambda_b^0 \to \Lambda_c^+\pi^-$, $\overline{B}_s^0 \to D_s^+\pi^-$ and $B^- \to D^0\pi^-$ respectively. In this multivariate method, all information of the charm hadron and its daughters is used in order to take any possible correlation into account automatically. The BDTs reduce the selection of charm hadron to a single variable and effectively provide charm particle ID.

Zusammenfassung

Diese Arbeit dokumentiert die Entwicklung eines Klassifikationsalgorithmus zur Identifikation von charm Hadronen, die von B Zerfällen stammen. Zu diesen Zweck wurden mehrere Boosted Decision Trees mit LHCb Daten von 2012 der Kanäle $\Lambda_b^0 \to \Lambda_c^+\pi^-$, $\overline{B}_s^0 \to D_s^+\pi^-$ und $B^- \to D^0\pi^-$ trainiert. In dieser multivariaten Methode wurde die vollständige Information des charm Hadrons und seiner Tochterteilchen verwendet um jede mögliche Korrelation automatisch zu berücksichten. Die BDTs reduzieren die Selektion eines charm Hadrons zu einer einzigen Variable und stellen effektiv eine Teilchen ID zur Verfügung.

Contents

1 Introduction

In mid July, 2015, the LHCb collaboration published a paper, which documents a Dalitz analysis of the decay $\Lambda_b^0 \to J/\psi K^- p$ [1]. Two intermediate resonances are observed in the $J/\psi p$ system. They are interpreted as particles composed of five quarks. Both of these so-called *pentaquarks* are measured with a significance of more than nine standard deviations (for comparison, a discovery is typically claimed with a significance of five standard deviations). The pentaquarks therefore seem to be a well-established experimental fact. In order to learn more about the nature of these exotic objects, various channels are investigated [2].

The lowest-order Feynman diagram representing the discovery channel is sketched in Figure 1.1. In addition, another channel $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ is shown possessing the same quarks but forming different final state hadrons. In particular, the $c\bar{c}$ pair forming the J/ψ in the discovery channel hadronises in different final state particles in $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$. The Λ_c^+ ($\overline{D}^0$) contains the c ($\bar{c}$) in the other channel. This type of decay with non-zero charm-ness is referred to as *open charm*. This difference makes $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ interesting for a search for pentaquarks.

Studying the decay $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ requires a selection algorithm which on one hand provides a large and clean signal yield, on the other hand a well-understood efficiency e. g. in view of a Branching Ratio measurement or a Dalitz plot analysis. The particle

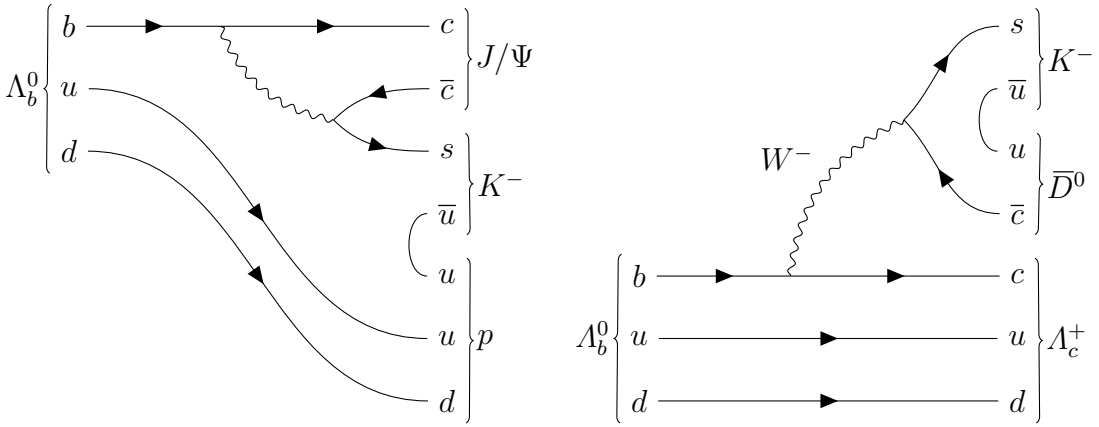

Figure 1.1: *Lowest-order Feynman diagrams representing the discovery channel $\Lambda_b^0 \to J/\psi K^- p$ (left) and analysis channel $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ (right) under current investigation.*

identification system at LHCb provides a powerful tool to identify . However, applying PID selection criteria on multiple daughter particles lead to correlations which are not well understood.

Therefore, this thesis presents an identification method based on a multivariate algorithm that is able to provide an optimal and well calibrated selection for any charm hadron and is thus as well useful for the decay $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ for the investigation of pentaquarks.

2 Theoretical fundamentals

2.1 The strong interaction

The particles in the channels mentioned in the introduction are called *hadrons*. All hadrons can be formed from only a few different elementary and electrically charged particles, called *quarks*. They come across in three *generations*, each contains two types, called *flavours*. The quarks are listed in Table 2.1. The force that is responsible for binding quarks into hadrons is the *strong interaction*, which is one of four fundamental forces in physics. The strong interaction is described by a quantum field theory named *Quantum Chromodynamics*. In quantum field theory, the interaction between particles is explained by their exchange of other particles, so-called *gauge bosons*. Each of the four fundamental forces is attributed such a characteristic "force-carrying" particle. The strong interaction is mediated by the massless *gluon*. To be more precisely, the strong interaction between two particles will only take place if both carry *colour charge*. Besides other charges, the quarks especially carry the colour charge, which either can be red, green or blue (anti-quarks carry the respective anti-colour charge, namely anti-red, anti-green or anti-blue).

Table 2.1: *The six quark flavours. For each quark there is a corresponding anti-quark. The electric charge Q is given in units of the elementary charge $e = 1.6 \times 10^{-19}$ C [3].*

Generation	Particle		Q	mass [GeV]
First	up	(u)	+2/3	0.005
	down	(d)	-1/3	0.003
Second	charm	(c)	+2/3	1.3
	strange	(s)	-1/3	0.1
Third	top	(t)	+2/3	174
	beauty	(b)	-1/3	4.5

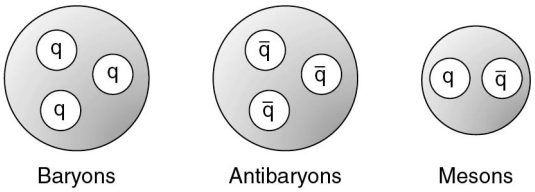

Figure 2.1: *The three conventional types of hadrons* [3].

2.2 Colour confinement

Up to this point, the quarks were considered as the "building blocks" of other particles. However, free quarks that are not bound in any hadrons are not observed so far. This non-observation of free quarks is explained in Quantum Chromodynamics by introducing the concept of *colour confinement*. According to this, every particle carrying the colour charge needs to be bound together with other colour-charged particles into a colour-neutral state. Thus, to observe a particle the colour charges of its constituents need to compensate themselves. Although there is no analytical proof of this concept, a possible assumption of its origin is the strong interaction of gluons with other gluons, which is allowed since they also carry the colour charge [3].

The three "conventional" types of hadrons are shown in Figure 2.1. A *baryon* consists of three quarks (q), an antibaryon of the respective antiquarks ($\overline{q}$), and a meson is formed from a quark-antiquark ($q\overline{q}$) pair. All of these hadrons provide colour-neutral states. Various combinations of different quark flavours have been observed so far, and each of them were categorized as either baryon or meson. The quark model allows for the possibility of colour-neutral combinations provided by four quarks and one antiquark ($qqqq\overline{q}$) [4]. However, these *pentaquarks* were not observed in nature since introducing the quark model in the 1960s till their discovery in 2015 by the LHCb experiment.

3 The LHCb experiment

The LHCb experiment is one of four large experiments at the Large Hadron Collider (LHC). It is designed to study heavy flavour physics, in particular b hadron decays. The search for indirect evidence of new physics in CP violation and rare decays of beauty and charm hadrons are the main goals of LHCb [5]. Deep inelastic proton-proton (pp) collisions at high energies are investigated providing a large $b\bar{b}$ production cross section. The main process for $b\bar{b}$ production is gluon-gluon fusion. Not the quarks but the gluons within the protons interact and create a $b\bar{b}$ pair.

In these events, a large number of hadrons is produced. This so-called hadronisation process is a consequence of colour confinement, which implies that quarks do not propagate freely but are always observed as colour-neutral objects. Thus, if a beauty quark is produced, there will be one hadron, which contains the beauty quark and is therefore called b *hadron*. The b hadrons decay due to the weak interaction and therefore possess a long lifetime, which allows them to travel a measurable lifetime from the production vertex. Although being heavier than the c quark and thus providing a larger phase space volume for the decay, b hadrons live longer since their CKM favoured transition to a t quark is forbidden by energy conservation. The long lifetime in combination with the boost of quarks allows that b hadrons can travel several millimetres before it decays. In summary, from an experimental point of view, b hadrons are very suitable to study because of the large $b\bar{b}$ production cross section and their relatively distinct experimental signature provided that both vertices can be resolved.

3.1 The LHCb detector

In contrast to other experiments at LHC where the detectors are designed to surround the entire collision point, the LHCb detector covers a pseudorapidity range of $2 < \eta < 5$ with $\eta = -\ln\left(\tan\frac{\theta}{2}\right)$ and θ denotes the opening angle with respect to the beam axis. This design is justified by the fact that at high energies the b hadrons are predominantly produced with a forward boost along the direction of the incoming protons. In addition, the production of b and $\bar{b}$ hadrons at high energies is highly correlated, which implies the same amount of b hadrons in the forward and backward direction [5]. The layout of the LHCb detector is shown in Figure 3.1. The following sections will provide a brief introduction to the systems that are needed for the analysis presented in this thesis.

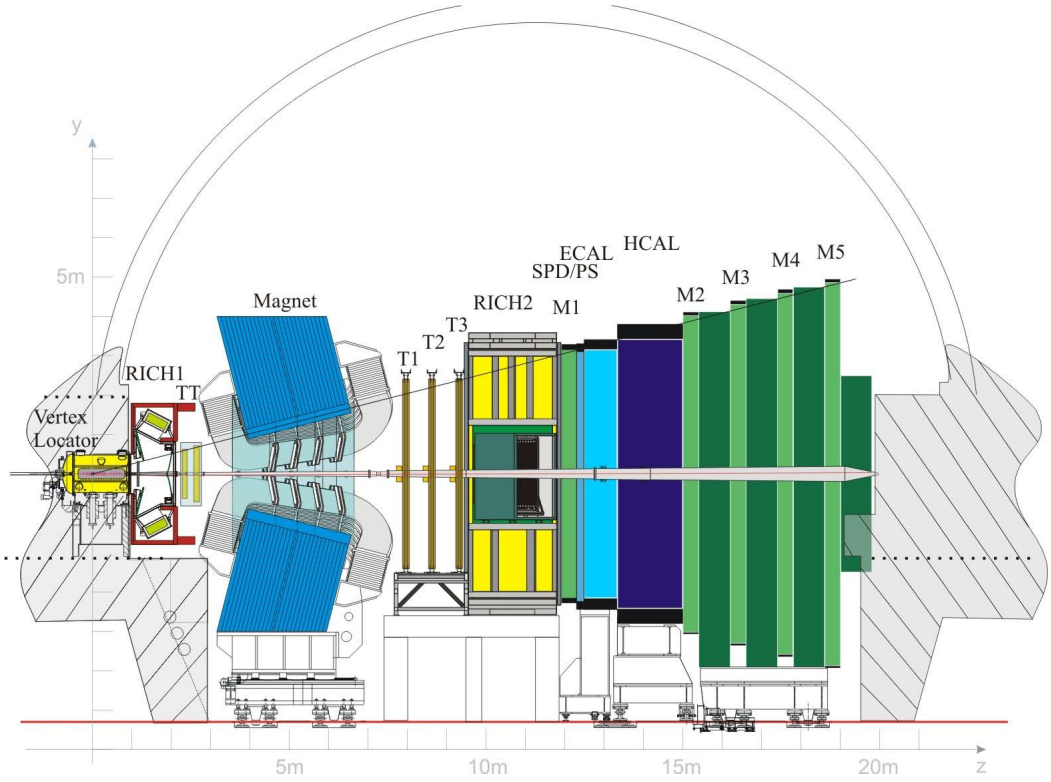

Figure 3.1: *LHCb detector [6].*

3.1.1 Tracking system

Tracking systems are able to record the trajectories of charged particles [7]. The tracking combines the information from several subdetectors.

The *vertex locator* (VELO) is the innermost part of the LHCb detector [8]. Only inside the VELO the pp collision will take place and $b\bar{b}$ pairs are produced. It is designed to distinguish B hadrons from daughter particles, which are produced in the collision. It has a precise spatial resolution and is thus able to distinguish between charged particles originating from the pp collision and those, which came from displaced vertices such as the B decay vertices.

T stations with twelve layers are placed behind the magnet. A fine-grained silicon tracker is placed in the interior area whilst in the exterior area there is a straw-tube detector. Measurements before and behind the the magnet allow for a determination of the three-momentum of the track.

3.1.2 Particle identification

The position and three-momentum of stable[1] charged particles (p, π, K, e, μ) is obtained from the tracking system. In order to reconstruct a four-vector, which provide the full information of the particle, the mass is needed and therefore its identity. Several sub-

[1] The lifetimes are large enough that the particles can be detected at LHCb.

detectors are responsible to provide the particle identification (PID) information, namely the calorimeters, the muon station and both RICH detectors.

The identification of charged hadrons is provided by the *Ring Imaging Cherenkov* (RICH) detectors [9]. As the name implies, the underlying physical phenomenon is the Cherenkov effect which is exploited for this purpose. A charged particle traversing a medium of refractive index n will produce photons, if its velocity is greater than the speed of light in that medium. This Cherenkov radiation is emitted at an angle θ relative to the particle's trajectory which produces a ring-like structure in the detector. The angle depends on the particle's velocity

$$\cos\theta = \frac{1}{n\beta}. \tag{3.1}$$

One can show that only particles of momentum p with mass

$$m < p\sqrt{n^2 - 1} \tag{3.2}$$

will produce Cherenkov radiation.

Energy measurement of photons, electrons and hadrons is achieved by *calorimeters* [10]. A high-energy electron radiates bremsstrahlung when interacting electromagnetically with a nuclues of the traversed detector material. The bremsstrahlung photon will create an e^+e^- pair where for example the resulting e^- again can produce bremsstrahlung. The process continues until the average energy of all created particles falls below a critical energy. The energy deposit of the particle due to this *electromagnetic shower* can be investigated by the *electromagnetic calorimeter* (ECAL). A similar process is possible for charged hadrons. Furthermore, both charged and neutral hadrons can interact with the nucleus via the strong interaction. The energy deposit of the particles produced in these *hadronic shower* is analysed by the *hadronic calorimeter* (HCAL).

The muon stations are responsible for the detection of muons [11, 12]. Ionisation is the dominant energy-loss process for muons with energies below 100 GeV. Muons can travel large distances compared to other particles. Therefore, the muon stations are located at the very end of the LHCb detector.

The information from the various PID systems is combined in a single PID variable. The two standard types of PID variables are DLL and Prob_{NN}.

For defining a DLL (delta-log-likelihood) variable a likelihood function $\mathcal{L}$ is used. In order to determine the particle species for each track for example in the RICH detector, the Cherenkov angles for each track are simultaneously reconstructed using the three-momentum obtained in the tracking system. Firstly, an overall event-likelihood $\mathcal{L}(\pi)$ is calculated assuming that all tracks in the RICH detector represent pions since they are the most abundant particles in the detector. Afterwards, the mass hypothesis is changed to e, μ, K and p for which the likelihood is recomputed. Comparing all tracks, the change in mass hypothesis, which causes the largest increase of event likelihood can

be identified. The mass hypothesis of that particle is then its most-likely hypothesis. This procedure continues until every track is set to its optimal mass hypothesis. The differences $\Delta \log \mathcal{L}(X - \pi) \equiv \log \mathcal{L}(X)/\log \mathcal{L}(\pi)$ in the log-likelihood values, where X denotes another mass hypothesis (e, μ, K or p), are used for particle identification of a certain track assuming that all other tracks have been set to their optimal mass hypothesis. Similar likelihood informations provided by other detectors can then added together to form so-called *combined* DLLs. An even better performance of PID is provided by the ProbNN variables, which are based on multivariate algorithms, namely neural networks. They use about 30 variables providing a most-sophisticated combination of these variables. A similar ansatz is presented in this thesis, as described later. However, this variable is based on a per charm hadron and not a per-track basis.

The PID performance is evaluated on control samples. These so-called PID *calibration samples* are found without using PID information on data. There are several decay channels, which are used for this calibration, such as $K^0_S \to \pi^+\pi^-$, $\Lambda \to p\pi^-$ and $D^{*+} \to D^0 (K^-\pi^+)\pi^-$, already showing a high purity by only exploiting kinematic properties. For a general introduction to the DLL variables and the calibration, see the PIDCalib package [9, 13].

3.2 Data flow

In a typical collider experiment, the accelerated protons are grouped into bunches and brought into collision. At LHCb, there are 40 million collisions of bunches per second where in average 1.4 pp interactions take place per bunch crossing [14]. Due to the detector geoemtry the LHCb detector registers about 10 million pp collisions per second [15]. This large amount of data must be filtered since it is impossible neither to store, nor to handle it due to limited storage space and computer performance. Only a small subset of the recorded data is selected, while the main part is rejected.

A trigger system is used to select that subset that contains the most valuable data for the LHCb experiment, namely beauty and charm decays. The *Level-0* (L0) trigger implemented in hardware selects 1 million events recorded in real-time by the VELO, the calorimeters and the muon stations. Up to this point the dataset corresponds to a storage size of 35 GB per second. The *High-Level-Trigger* (HLT) implemented in software further reduces dataset to a more manageable size of 2000 events per second.

During the reconstruction the detector signal is transformed into physical objects such as tracks, calorimeter clusters or PID variables and stored into an output file. In principle, the output files can be used for analysis but they are still to large for individual people to handle. Further filtering called *stripping* are performed in order to reduce the data. In addition, similar events are grouped together (so-called *stripping lines*) and stored into the same output file. This allows the analysts to efficiently access the subset of data they

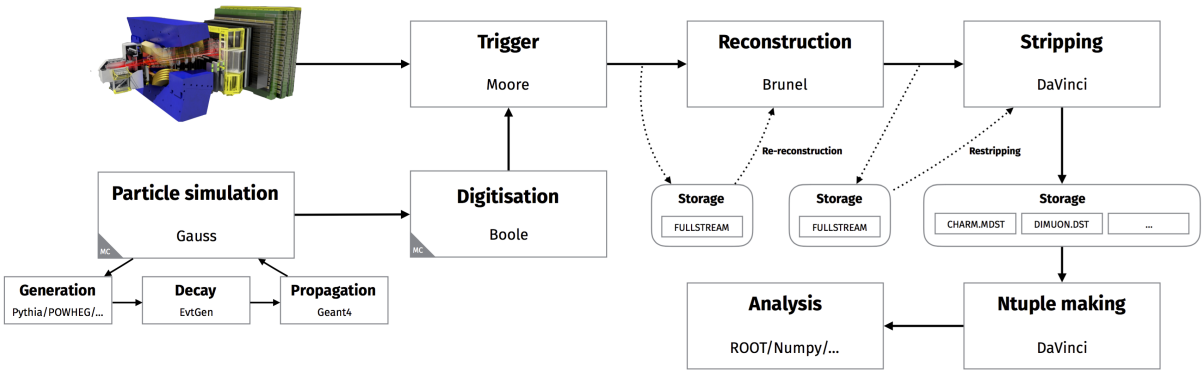

Figure 3.2: *LHCb data flow [16].*

are interested in.

Since the reconstruction is computationally very intensive, it is only done once directly after data taking or when major improvements due to a new alignment or calibration are expected. The stripping is carry out more often.

4 Charm hadron selection for the decay $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$

4.1 Correlation of efficiencies

Studying a certain decay (regardless of the type of analysis), a selection of events related to this decay is required. In the case at hand, a classification algorithm needs to be developed having the ability to identify charmed hadrons from b decays. A classification that is only based on kinematic properties would not lead to an optimal selection. The *efficiency* for selecting signal events at a given background rejection level can be further increased by using PID variables (see 3.1.2). Since one PID variable is designed to identify a single stable particle, cuts on PID variables of six final state hadrons would be necessary. The problem is to calculate the resulting efficiency of a selection that is based on cuts on six variables. This is closely related to the question how to obtain the correlation of the efficiencies of each variable. This problem becomes difficult for non-factorising efficiencies where the total efficiency ε_{tot} cannot be simply written as the product of the single efficiencies ε_k

$$\varepsilon_{\text{tot}} = \prod_{k=1}^{n} \varepsilon_k. \tag{4.1}$$

The decay $\Lambda_b^0 \to \Lambda_c^+ \, (pK^-\pi^+) D_s^- \, (K^+K^-\pi^-)$ has the same final state particles as $\Lambda_b^0 \to \Lambda_c^+ \, (pK^-\pi^+)\overline{D}^0 \, (K^+\pi^-)K^-$, and the correlation of the PID efficiencies has been investigated on this channel. By applying a few cuts on kinematic variables, this channel already provides a high purity. This allows for a signal definition without using any PID information, which is used to obtain the signal yield for the calculation of the efficiency. Afterwards, the efficiency of every single cut was calculated. After this, the efficiency of all cuts, ε_{tot}, assuming no correlation as given by eq. 4.1 is computed. The signal yields are extracted with all cuts applied and then corrected for ε_{tot}. Ideally, the efficiency-corrected yields are identically to the yield without any PID cut. Figure 4.1 shows the comparison between the efficiency-corrected yields and the yields without PID cuts clearly demonstrating that the PID efficiencies of this (complex) decay with six final state hadrons do not simply factorise; the correlation seems to be significant.

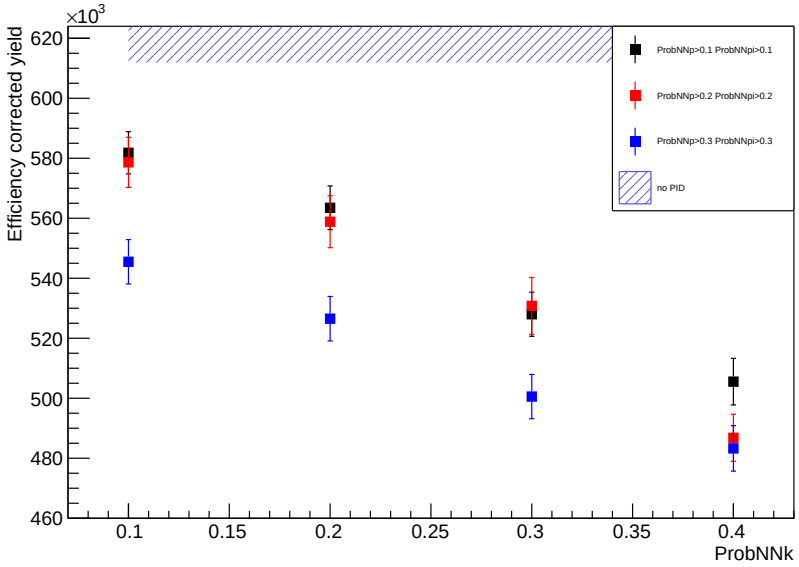

Figure 4.1: *Efficiency-corrected yields for the decay $\Lambda_b^0 \to \Lambda_c^+ D_s^+$ in comparison to the yields without PID cuts. In case of uncorrelated PID efficiencies the corrected yields are expected to lie in the blue band.*

4.2 Analysis strategy

Since the correlation between six PID efficiencies is not well-understood, the usage of PID variables in such a complex case is avoided. Therefore, alternative PID-like variables are going to be defined and calibrated identifying charm hadrons coming from beauty hadrons. In contrast to the standard PID variables, which are defined and calibrated *per track*, these new PID-like variables are supposed to be defined and calibrated *per charm hadron*. The identification of a c hadron is achieved by combining the information from several tracks related to its daughter particles simultaneously. Exploiting the large statistic modes $\Lambda_b^0 \to \Lambda_c^+ \pi^-$, $\overline{B}_s^0 \to D_s^+ \pi^-$ and $B^- \to D^0 \pi^-$, the efficiencies are then calibrated with $\Lambda_b^0 \to \Lambda_c^+ D_s^-$.

The PID-like variables can be defined by developing a classifier using multivariate algorithms. Several methods from the *Toolkit for Multivariate Analysis* (TMVA) [17] can be employed for this problem. A well-tested and highly performed classification is provided by *Boosted Decision Trees* (BDTs). More information explaining the idea and functionality of a BDT is given in 5.3.1.

The BDTs in this analysis are trained using a large number of variables related to the daughter particles of the charmed hadrons. *Training* refers to the phase where the BDT learns how to discriminate between signal and background. By doing this, the separation power of single variables, but also any possible correlation can be exploited for discriminating events, hence, for identifying Λ_c^+, D_s^+ and $\overline{D}^0$ respectively. The success of using so many variables for this case is described in [18]. The BDTs effectively provide the desired

Table 4.1: *Channels used for the training of the BDTs.*

Λ_b^0	$\rightarrow$	$\Lambda_c^+\pi^-$	with	Λ_c^+	$\rightarrow$	$pK^-\pi^+$
$\overline{B}_s^0$	$\rightarrow$	$D_s^+\pi^-$	with	D_s^+	$\rightarrow$	$K^+K^-\pi^+$
B^-	$\rightarrow$	$D^0\pi^-$	with	D^0	$\rightarrow$	$K^-\pi^+$

PID-like variable. Since one single BDT is supposed to identify candidates related to one particular decay channel, three BDTs are thus required to take all the charmed particles, namely Λ_c^+, $\overline{D}^0$ and D_s^+, from both the signal and reference channel into account. PID variables will be used for the identification of K^- in the decay $\Lambda_b^0 \rightarrow \Lambda_c^+\overline{D}^0K^-$. Table 4.1 shows the training channels. When the training is finished, the resulting BDT cut can be applied to a dataset with unknown sample composition. For instance, for the classification of the reference channel $\Lambda_b^0 \rightarrow \Lambda_c^+D_s^-$ both the BDT cut for Λ_c^+ and D_s^- needs to be applied to the dataset containing events related to the reference channel. The advantage of using BDT cuts is to reduce the number of possibly correlated efficiencies to a small set of well-understood efficiencies. In addition, the BDTs provide a near optimal selection by combining information of many variables in a clever way.

For each channel a trained BDT already exists which is presended in [19]. However, these BDTs were trained using data from a previous stripping version. The task of this thesis is to retrain and calibrate the PID-like variable for the most recent stripping version.

5 Retraining of Boosted Decision Trees

This chapter is dedicated to the retraining phase of the BDTs, which is the main part of this thesis. The description of the pre-selection is followed up by the extraction of the signal yields using the ${}_s\mathcal{P}lot$ technique two-dimensional fits to the b and c hadron mass distributions. Once the pure signal and background sample are obtained, the BDT training is documented including the training configuration and the used variables.

5.1 Pre-selection

The BDTs are supposed to provide an open charm hadron selection from beauty decays. Therefore, $X_b \rightarrow X_c\pi$ decays possessing large statistics are used to define and calibrate PID-like variables. A clear X_b signal in the $X_c\pi$ mass spectrum is needed. Thus, a pre-selection on top of the stripping selection, which enables a clear labelling of signal and background but also retains a large sample size, has to be defined. The used stripping lines do not include any PID cuts. The channels already show a very high purity by further applying cuts on kinematic variables (described below). No PID information is necessary to define a signal sample. The pre-selection cuts on the training channels are necessary to provide a higher training performance of the BDTs. In order to provide a *generalised* PID-like variabe that is able to identify any charmed hadrons in b decays, the pre-selection must not bias the charm mass distribution by altering the background shape. Therefore, no cuts on variables related to charmed hadrons are applied. Analyses concerned with the investigation of other b decays including charmed hadrons are expected to apply the same pre-selection cuts. Therefore, only two cuts were applied in the pre-selection of the decays $X_b \rightarrow X_c\pi$, namely on DIRA and χ^2_{IP} of the b hadron and on one ProbNN variable of the bachelor pion (the pion in the $X_b \rightarrow X_c\pi$ decays). The cuts are listed in Table 5.1 and explained later.

Table 5.1: *Pre-selection cuts for $\Lambda_b^0 \rightarrow \Lambda_c^+\pi^-$ and $\overline{B}{}_s^0 \rightarrow D_s^+\pi^-$ candidates.*

Particle	Cut
Λ_b^0, B_s^0	DIRA > 0.999
	$\chi^2_{\text{IP}} < 10$
π^-	$\text{Prob}_{\text{NN}}(\pi) > 0.1$

5.1.1 Selection variables

The following section gives a brief introduction to the selection variables used in this thesis.

- The *transverse momentum* (p_T) denotes the component of the momentum transversal to the beam axis and can be calculated by

$$p_T = \sqrt{p_x^2 + p_y^2}. \tag{5.1}$$

- The *direction angle* (DIRA) is defined as the cosine of the angle between the momentum of the particle and the direction vector from the production vertex to the decay vertex (end-vertex, ED) of the particle. It is supposed to be close to one if the vertices and all momenta of the considered particle and its daughter particles are well reconstructed, and no particle is missing in the decay chain.
- The *impact parameter* (IP) of the b hadron is the smallest distance from a given reference vertex to the reconstructed particle trajectory. The Vertex Fitter provides a χ^2 value, which is defined as $\chi^2_{\text{IP}} = \text{IP}/\Delta\text{IP}$. Analogous χ^2 values for other selections variables like flight distance, end vertex (EV) etc. are possible as well.
- The *flight distance* (FD) is the distance from a given reference vertex to another decay vertex. $N_{\text{shared}}(\mu)$ denotes the number of measurements in the muon system, which are shared with other *proto-particles*. The proto-particles are reconstructed tracks without any assigned mass hypothesis.
- $\log \mathcal{L}_{\text{Muon}}(\mu)$ denotes a DLL variable provided by the muon stations for the identification of muons. $\log \mathcal{L}_{\text{Muon}}(\text{bkg})$ denotes the DLL variable from the muon stations to be background.
- $\log \mathcal{L}_{\text{RICH}}(p)$ denotes a DLL variable provided by the RICH detector for the identification of protons. $\log \mathcal{L}_{\text{RICH}}(\text{bt})$ denotes a DLL variable from the RICH detector for the below threshold hypothesis.
- $\text{Prob}_{\text{NN}}(\pi)$ denotes a ProbNN variable for the identification of pions.
- $b_{\text{thresh}}(K)$ marks a boolean variable, which is true if the kaon is above threshold to radiate Cherenkov photons. The treshold is given by eq. 3.2. $b_{\text{RICH1}}(\pi)$ marks a boolean variable, which is true if the π causes a measurement in the first RICH detector.
- *Ghosts* are reconstructed tracks that do not correspond to a real particle, which passed through the detector. $\text{Prob}_{\text{NN}}(\text{Ghost})$ denotes the probability that the considered track is actually a ghost.
- Q_{Velo} denotes the charge that is deposited in the VELO system.
- $b_{\text{RICH,1}}$ denotes a boolean, which is true if there is measurement in RICH1. $b_{\text{RICH,aerogel}}$ denotes a boolean, which is true if there is a measurement in a certain area of RICH1, which is made of aerogel.
- $b_{\text{PID}}(\mu)$ denotes a boolean, which is true if a muon was measured in the muon system.

It is a PID variable, which comes from the muon PID system.

- χ^2_{Match}, $\chi^2_{\text{T,NDOF}}$ and $\chi^2_{\text{Velo,NDOF}}$ denote the track quality computed from the track fit.

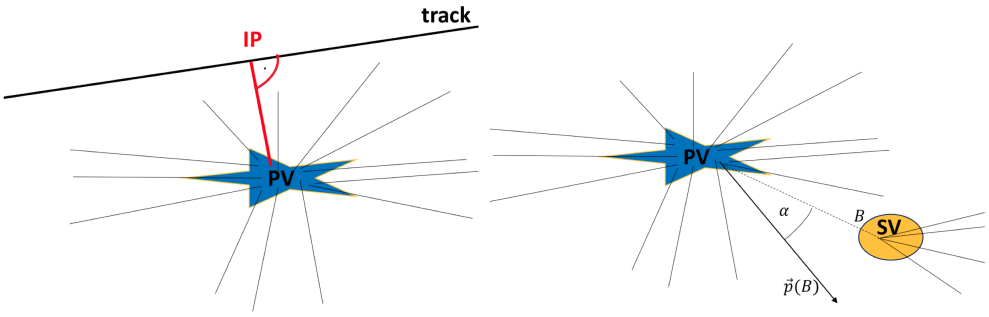

Figure 5.1: *Schematic view of the impact parameter (left) and the direction angle (right)* [20].

5.1.2 Vetos

Since no PID information for separation has been used yet, the usage of *vetos* due to mis-assignment of the particle's identity becomes important. For instance, it is possible that the proton coming from the Λ_c^+ is actually a positively charged kaon. Hence, instead of $\Lambda_c^+ \to pK^-\pi^+$, there are actually D_s^+ candidates decaying via $D_s^+ \to K^+K^-\pi^+$. In order to be able to veto these decays, the kaon mass is assigned to the proton candidate. Those events related to the D_s^+ candidates can then be rejected by applying a simple cut on a properly chosen region around the D_s^+ mass in the invariant mass representation with swapped particle hypothesis. Hence, the background, which is caused by this additional channel, is eliminated by applying veto-cuts. In the following, several examples are given demonstrating the effect of vetos.

5.1.3 Results of the pre-selection

Figure 5.2 shows the invariant mass distributions of $\Lambda_b^0 \to \Lambda_c^+\pi^-$ and $\Lambda_c^+ \to pK^-\pi^+$ candidates, which come directly from the stripping. Figure 5.3 shows the effect of all pre-selection cuts for $\Lambda_b^0 \to \Lambda_c^+\pi^-$ and $\Lambda_c^+ \to pK^-\pi^+$ candidates.

As explained above, it is possible that the proton coming from the Λ_c^+ is actually a pion or a kaon due to mis-assignment. Hence, instead of $\Lambda_c^+ \to pK^-\pi^+$, there are actually D^+ or D_s^+ candidates decaying via $D^+ \to \pi^+K^-\pi^+$ or $D_s^+ \to K^+K^-\pi^+$. Other possible scenarios of mis-assignments are investigated as well, but the resulting contributions are found to be negligible. The channels and the respective veto range are listed in Table 5.2. The effect of the vetos is shown in Figure 5.4.

Figure 5.5 shows the invariant mass distributions of $\overline{B}_s^0 \to D_s^+\pi^-$ and $D_s^+ \to K^+K^-\pi^+$

Table 5.2: *Vetos in the selection of* $\Lambda_b^0 \to \Lambda_c^+\pi^-$ *candidates.*

Combination	Particle	Mass [MeV]	Veto range [MeV]
$K^+K^-\pi^+$	D_s^+	1968.27	1938–1998
$K^-\pi^+\pi^+$	D^+	1869.58	1839–1899

Table 5.3: *Vetos in the pre-selection of* $\overline{B}_s^0 \to D_s^+\pi^-$ *candidates.*

Combination	Particle	Mass [MeV]	Veto range [MeV]
$pK^-\pi^+$	Λ_c^+	2286.46	2266–2306
$K^-\pi^+\pi^+$	D^+	1869.58	1849–1889

candidates, which come directly from the stripping. Figure 5.6 shows the same distributions after all pre-selection cuts.

Similar to the selection of $\Lambda_b^0 \to \Lambda_c^+\pi^-$ candidates, no PID information has been used for the selection of $\overline{B}_s^0 \to D_s^+\pi^-$ candidates yet. Hence, it is possible that the kaon coming from the D_s^+ is actually a proton or a pion due to mis-assignment. Hence, instead of D_s^+, there are actually Λ_c^+ or D^+ candidates due to $\Lambda_c^+ \to pK^-\pi^+$ or $D^+ \to K^-\pi^+\pi^+$. The channels and the respective veto range are listed in Table 5.3. The effect of the vetos is shown in Figure 5.7.

Figure 5.8 shows the invariant mass distributions of $B^- \to D^0\pi^-$ and $D^0 \to K^-\pi^+$ candidates, which come directly from the stripping. Since the stripping already provides a very pure signal it was not necessary to apply any further cuts.

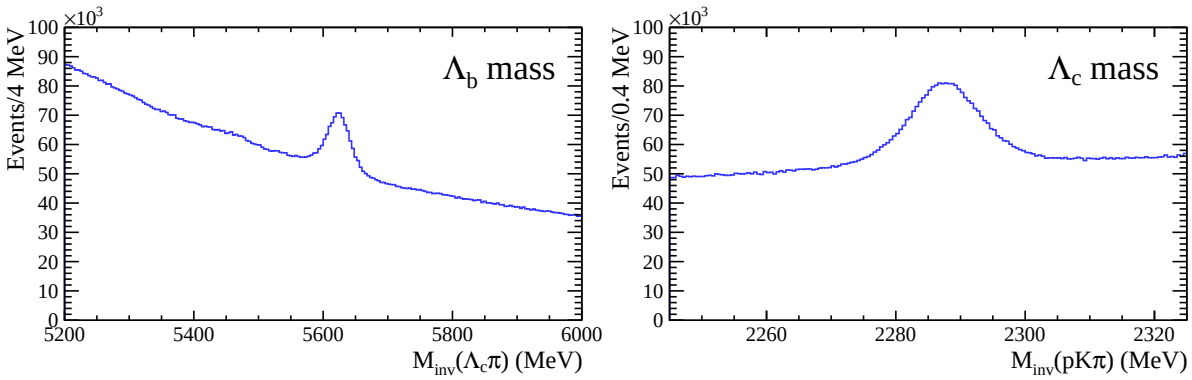

Figure 5.2: *Invariant mass distributions of* $\Lambda_b^0 \to \Lambda_c^+\pi^-$ *and* $\Lambda_c^+ \to pK^-\pi^+$ *candidates based on data which comes directly from the stripping.*

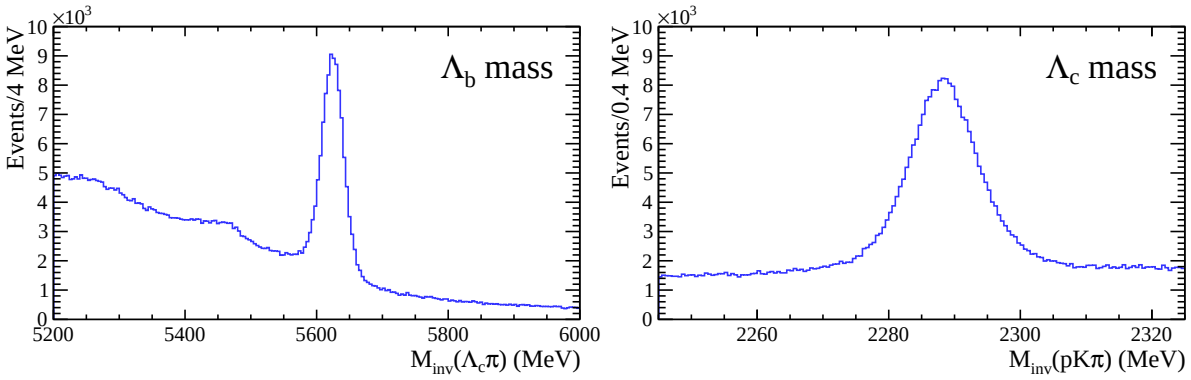

Figure 5.3: *Invariant mass distributions of $\Lambda_b^0 \rightarrow \Lambda_c^+\pi^-$ and $\Lambda_c^+ \rightarrow pK^-\pi^+$ candidates after the pre-selection cuts were applied.*

5.2 Signal selection

The training of a BDT requires pure signal and background samples. Based on these two samples, the BDT will learn how to discriminate between signal and background. It is a statistical problem to unfold the composition of a given data sample, which is solved in this thesis by the ${}_s\mathcal{P}lot$ technique [21]. This statistical tool starts from a *discriminating variable* for which the signal and background composition is already known. The discriminating variable is then used to deduce the sample composition with respect to other so-called *control variables* in which the sources are not known yet. The underlying assumption of this technique is that all the control variables are uncorrelated with the discriminating variable. The ${}_s\mathcal{P}lot$ provides a weight function, the so-called ${}_s\mathcal{W}eight$, which is used for the reconstruction of the true distributions. The invariant mass of the b and c hadrons are chosen as the two dimensional discriminating variable to take contributions from charmless decays into account, which is described below.

5.2.1 Treatment of charmless decays

Considering $\Lambda_b^0 \rightarrow \Lambda_c^+\pi^-$ for instance, it is possible that Λ_b^0 will directly decay via $\Lambda_b^0 \rightarrow pK^-\pi^+\pi^-$. This so-called *charmless decay* will contaminate the signal region of the Λ_b^0 mass which is, however, used to identify a real Λ_c^+ coming from a Λ_b^0. The charmless decay is expected to be suppressed relative to the charm decay. Nevertheless, in order to avoid this confusion, the mass spectrum of $\Lambda_c^+\pi^-$ (showing the Λ_b^0 peak) is plotted against the mass spectrum of $pK^-\pi^+$ (showing the Λ_c^+ peak). A two-dimensional fit to both the Λ_b^0 and Λ_c^+ mass spectrum then allows to distinguish the signal channel $\Lambda_b^0 \rightarrow \Lambda_c^+\pi^-$ from the charmless decay. Signal events simultaneously lie in both the Λ_b^0 and Λ_c^+ signal region. The ${}_s\mathcal{P}lot$ technique allows within this region to discriminate between signal events and contributions from other events. Similarly for the extraction

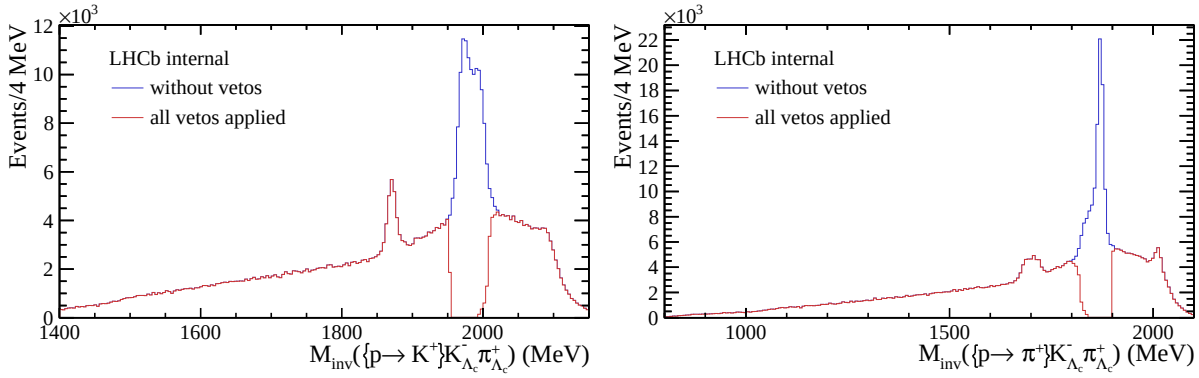

Figure 5.4: *Effect of vetos for $\Lambda_c^+ \to pK^-\pi^+$. Left: A D_s^+ signal in $K^+K^-\pi^+$ invariant mass distribution where the kaon mass is assigned to the proton candidate. The left peak is due to the $c \to s$ decay $D^- \to K^+K^-\pi^-$. Right: A D^+ signal in $K^-\pi^+\pi^+$ invariant mass distribution where the pion mass is assigned to the proton candidate. The peak to the right of D^+ is due to $D^{*+} \to D^0(K^-\pi^+)\pi^-$ whilst the other peak to the left of D^+ is due to the partially reconstructed decay $D^+ \to K^-\pi^+\pi^+\pi^0$, where the π^0 is not reconstructed.*

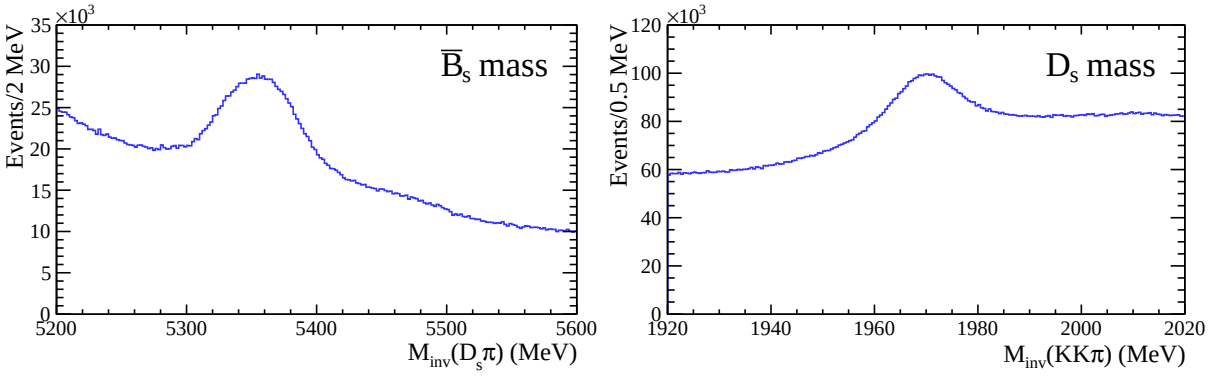

Figure 5.5: *Invariant mass distributions of $\overline{B}_s^0 \to D_s^+\pi^-$ and $D_s^+ \to K^+K^-\pi^+$ candidates based on data which comes directly from the stripping.*

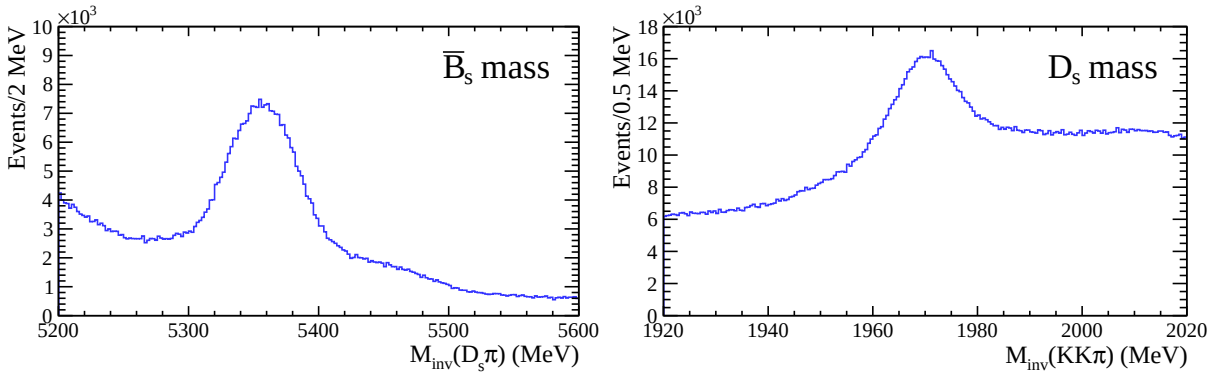

Figure 5.6: *Invariant mass distributions of* $\overline{B}{}^0_s \to D^+_s\pi^-$ *and* $D^+_s \to K^+K^-\pi^+$ *candidates after the pre-selection cuts were applied.*

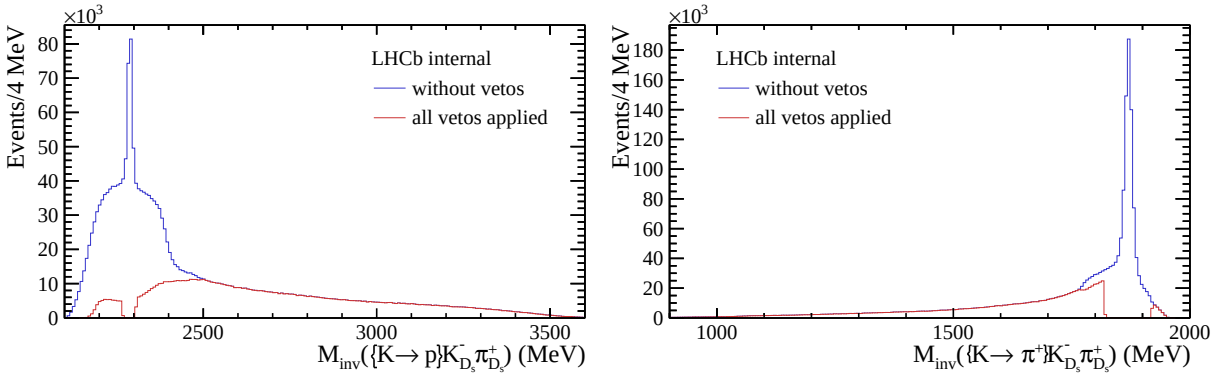

Figure 5.7: *Effect of vetos for* $D^+_s \to K^+K^-\pi^+$*. Left: A* Λ^+_c *signal in* $pK^-\pi^+$ *invariant mass distribution where the pion mass is assigned to the kaon candidate. Right: A* D^+ *signal in* $K^-2\pi^+$ *invariant mass distribution where the pion mass is assigned to the kaon and combined with the* K^- *and* π^+ *from the* D^+_s *candidate.*

of the yields in the $\overline{B}{}^0_s \to D^+_s\pi^-$ and $B^- \to D^0\pi^-$ channels charmelss contributions need to be taken into account as well.

The first step for performing a two-dimensional fit is to parametrise both signal and background shapes. Several shape parametrisations have been examined. The signal shapes of the beauty and charm mass distributions are well described by the sum of two Gaussian functions with a common peak position μ^b, μ^c respectively and different widths σ^b_1, σ^b_2 and σ^c_1, σ^c_2. Combinations of Chebyshev polynomials to different orders provide a good description of the background of both the beauty (first order) and charm (third order) mass distributions. These combinations $P_n(x)$ can be written as

$$P_n(x) = 1 + \sum_{k=1}^{n} a_k T_k(x), \tag{5.2}$$

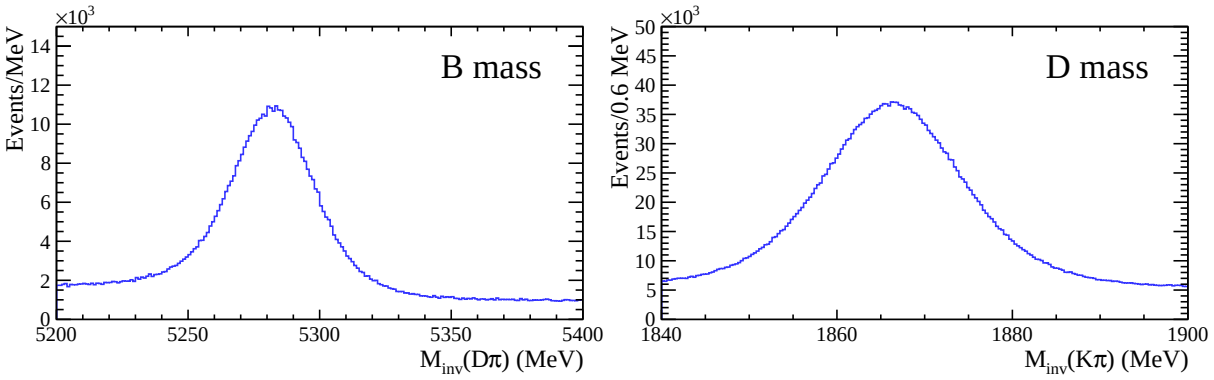

Figure 5.8: *Invariant mass distributions of $D^0\pi^-$ and $K^-\pi^+$ based on data which comes directly from the stripping.*

Table 5.4: *Components of the two-dimensional fit.*

Component	Name	Definition
1	Signal	$b_{\text{sig}}c_{\text{sig}}$
2	Charmless	$b_{\text{sig}}c_{\text{bkg}}$
3	Combinatorial charm	$b_{\text{bkg}}c_{\text{sig}}$
4	Combinatorial	$b_{\text{bkg}}c_{\text{bkg}}$

where T^k denotes the Chebyshev polynomials. The first few used in the following fit are

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_2(x) &= 2x^2 - 1 \\ T_3(x) &= 4x^3 - 3x \end{aligned}$$

The respective probability density functions (PDF) can thus be written

$$b_{\text{sig}}(x;\mu^b,\sigma_1^b,\sigma_2^b) = \frac{f_1}{\sqrt{2\pi}\sigma_1^b}\exp\left[\frac{(x-\mu^b)^2}{2\sigma_1^b}\right] + \frac{1-f_1}{\sqrt{2\pi}\sigma_2}\exp\left[\frac{(x-\mu^b)^2}{2\sigma_2^b}\right] \tag{5.3}$$

$$c_{\text{sig}}(x;\mu^c,\sigma_1^c,\sigma_2^c) = \frac{f_2}{\sqrt{2\pi}\sigma_1^c}\exp\left[\frac{(x-\mu^c)^2}{2\sigma_1^c}\right] + \frac{1-f_2}{\sqrt{2\pi}\sigma_2^c}\exp\left[\frac{(x-\mu^c)^2}{2\sigma_2^c}\right] \tag{5.4}$$

$$b_{\text{bkg}}(x;a^b) = P_1(x) = 1 + a^b T_1(x) \tag{5.5}$$

$$c_{\text{bkg}}(x;a^c,b^c,c^c) = P_3(x) = 1 + a^c T_1(x) + b^c T_2(x) + c^c T_3(x) \tag{5.6}$$

The two-dimensional fit to both the beauty and charm mass distribution consists of four components, which are defined as products of the PDFs in 5.3–5.6 respectively. Table 5.4 shows the components. The overall fit model represented by the PDF $f(x;\theta)$ is then obtained by the sum of all components presented above and can be written

$$f(x;\theta) = N_{\text{sig}}b_{\text{sig}}c_{\text{sig}} + N_{\text{cless}}b_{\text{sig}}c_{\text{bkg}} + N_{\text{ccmobbkg}}b_{\text{bkg}}c_{\text{sig}} + N_{\text{bkg}}b_{\text{bkg}}c_{\text{bkg}}, \tag{5.7}$$

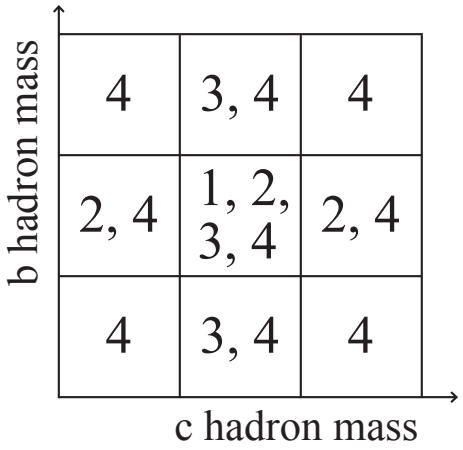

Figure 5.9: *Components from the two-dimensional fit to the mass distributions.*

Table 5.5: *Yields from the two-dimensional fit to the mass distributions.*

	Yield		
Component	$\Lambda_b^0 \to \Lambda_c^+\pi^-$	$\overline{B}_s^0 \to D_s^+\pi^-$	$B^- \to D^0\pi^-$
N_{sig}	90 333	69 744	372 889
N_{cless}	1359	2302	19 630
N_{ccombbkg}	7185	8348	80 130
N_{comb}	106 986	307 254	185 701

where $\theta = (\mu^b, \sigma_1^b, \sigma_2^b, \mu^c, \sigma_1^c, \sigma_2^c, a^b, a^c, b^c, c^c)$ denotes the vector representing all parameters. The components of the two-dimensional fit are displayed in Fig. 5.9. This table can be seen as a visualisation of the two-dimensional plot of both the Λ_b^0 (column) and Λ_c^+ (row) mass distribution. The yields from the respective fit are displayed in Table 5.5. The two-dimensional fit to both the Λ_b^0 and Λ_c^+ mass has shown that the charmless contribution is very small for all training channels. The projections of the two-dimensional fits are shown in Figure 5.10. The mass window of $\overline{B}_s^0 \to D_s^+\pi^-$ candidates (Figure 5.10, middle) is chosen to reject events related to the partially reconstructed decay $D_s^{*+} \to D_s^+\gamma$, which directly lie to the left of the $\overline{B}_s^0$ peak.

5.3 Training of Boosted Decision Trees

The two-dimensional fit and the calculation of the ${}_s\mathcal{W}eights$ enables a clear statistical separation of the signal and background sample, which are necessary for training a boosted decision tree. Before the training is documented including the chosen configuration and training variables, a short introduction to the functionality of BDTs is given.

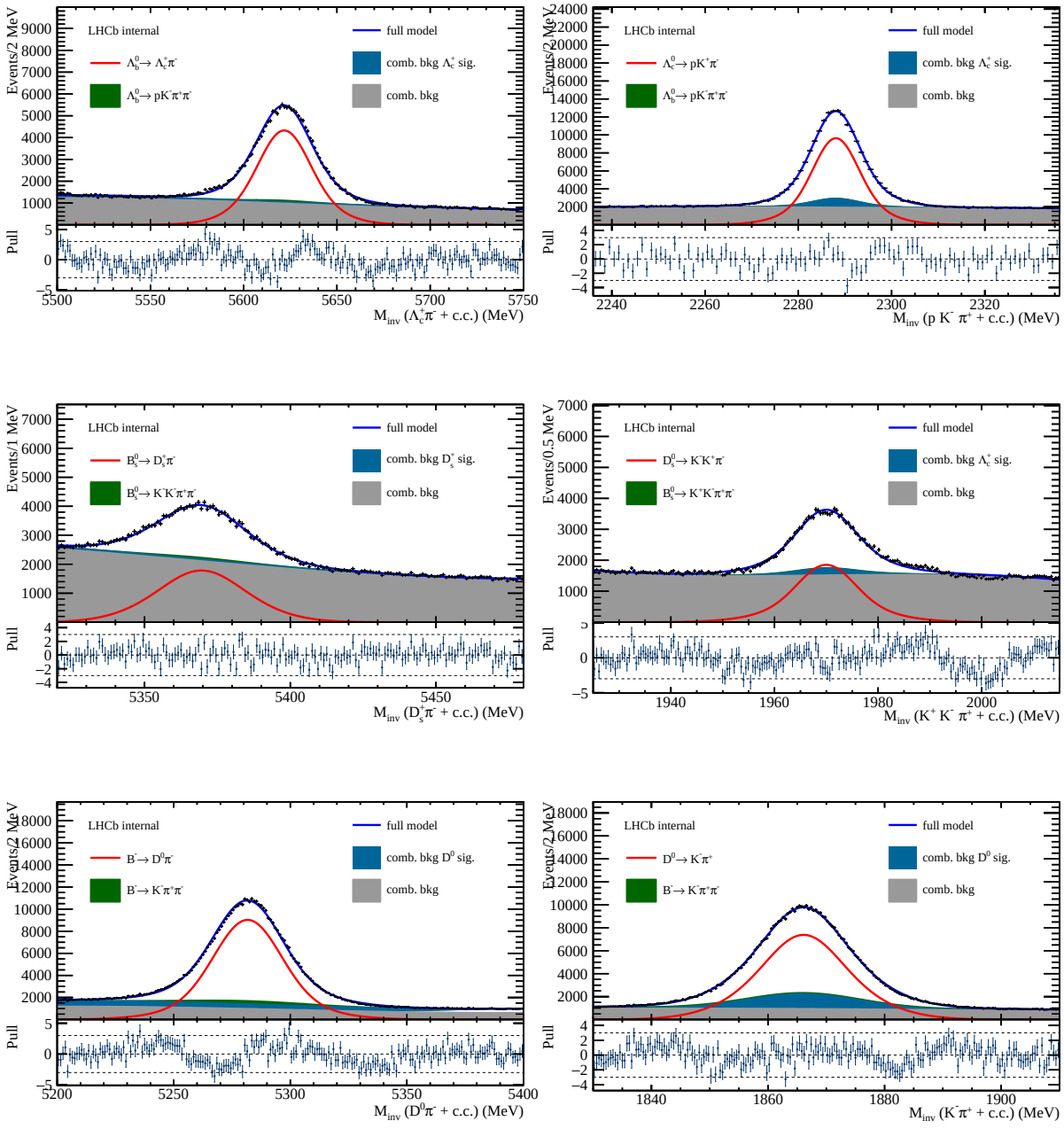

Figure 5.10: *Fit to the invariant mass distributions of* $\Lambda_b^0 \to \Lambda_c^+\pi^-$, $\overline{B}_s^0 \to D_s^+\pi^-$ *and* $B^- \to D^0\pi^-$ *(left) and* $\Lambda_c^+ \to pK^-\pi^+$, $D_s^+ \to K^+K^-\pi^+$ *and* $D^0 \to K^-\pi^+$ *(right) candidates.*

5.3.1 Functionality of Boosted Decision Trees

A *decision tree* is a tree-structured classification algorithm based on a series of binary splits [17, 22]. Figure 5.11 shows the schematic view of a decision tree. It starts from the original sample, so-called *root node*, and is followed up by repeating splits and nodes. Each node is splitted into two daughter nodes representing either signal or background label by applying a cut on a certain variable. The final node classifies a given event according to its label. The variable and cut value, which provide the best separation, are obtained in this thesis by minimizing the *Gini index* $p(1-p)$, which is a measure of the mixture of a given sample (p denotes the signal purity). Therefore, the same variable can be used at several nodes, while other variables may never be used. This procedure of sequentially splitting the training sample is repeated until a certain convergence criterion is fulfilled. In contrast to a pure cut-based analysis, which is able to only select one hypercube in the multidimensional observable space, the decision tree will split that space in many signal-like and background-like hypercubes making it a powerful classification method.

By construction, decision trees are fairly sensitive to statistical fluctuations in the training sample. Considering two variables with same separation power (in terms of Gini indices) a fluctuation would bias the decision on what variable the cut should be applied. The usage of a *forest* of many boosted decision trees, which are all derived from the same training sample, will overcome this problem. Boosting is a way of enhancing the decision tree. The training of *boosted* decision tree starts with the original sample as well but gets a modified training sample in the next training step where the previously misclassified events are given a larger weight. The classification of a given event is then a majority vote among all classifications done by each decision tree.

Overtraining refers to the fact that too many model parameters were adjusted to too few data points. This results in a increase of classification performance when measured on the training sample, and a decrease when measured on the independent test sample. To determine any indications for overtraining, the MVA variable distributions of training and testing samples for both signal and background are simultaneously plotted in order to compare the classification performance between training and test samples.

5.3.2 Training configuration

For the training of each BDT, the dataset is randomly split into equal sized training and validation sample. The validation sample will be used to estimate the efficiencies of the BDT. The boosting type is choosen to be *Adaptive Boost*. The number of decision trees in one forest is set to 800. The number of cuts tested during the minimization of the Gini index is set to 200. The convergence is reached when a node containing less than 20 % of the total training sample is created. The maximal depth of each decision tree is set to 3. All these TMVA options providing the best result are manually optimized and

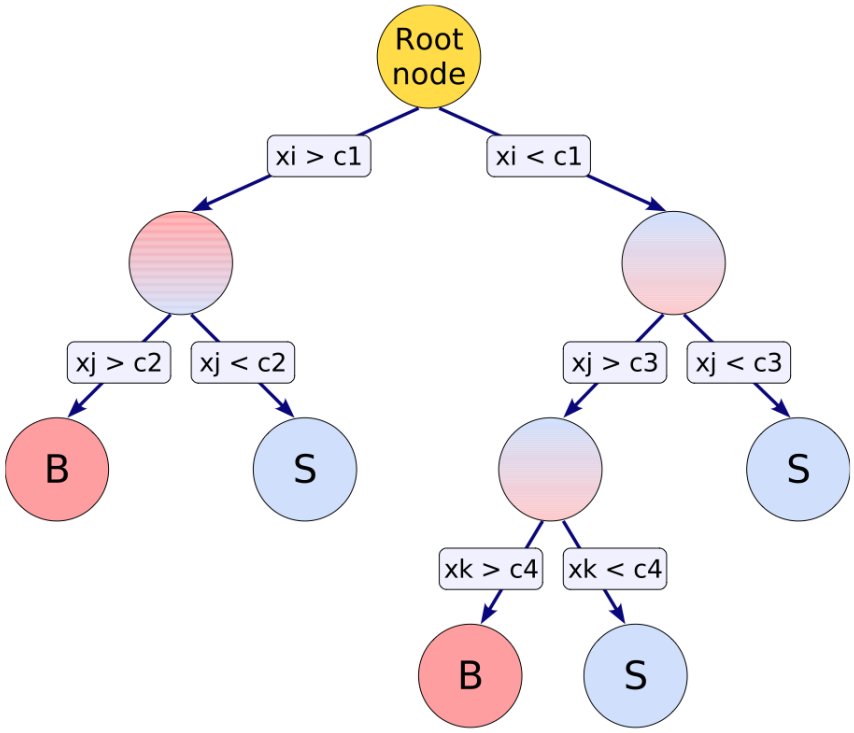

Figure 5.11: *Schematic View of a Decision Tree* [17].

summarized in Table 5.6.

5.3.3 Training variables

The variables used in the training are listed in Table 5.7. The BDT uses several kinematic properties of the charm hadron and its daughters. In addition, there is a large number of PID variables related to the charm daughters, which is included in the training. About 70 variables are used in the training of $\Lambda_b^0 \to \Lambda_c^+\pi^-$ and $\overline{B}_s^0 \to D_s^+\pi^-$, while about 50 variables are used for $B^- \to \overline{D}^0\pi$. Good agreement of all these variables between all training channels and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ has been found.

Since the BDT can handle flat distribution of variables better than spiking ones, several variables are rescaled to bring them into a more appropriate form. Instead of the original variable, log or arctan of this variables is used. The usage of these transformations during the retraining phase was found to provide a significant improvement in the performance of the BDTs.

5.3.4 Training results

The result of the multivariate analysis is a PID-like variable provided by the trained BDT. The training results are shown in Figure 5.12. The BDT response plots displayed below show the distribution of this BDT variable. In particular, the distributions of training and testing for both signal and background can be seen. Indications for overtraining can be

Table 5.6: *TMVA options.*

Option	Value
BoostType	AdaBoost
NTrees	800
UseBaggedBoost	true
BaggedSampleFraction	0.6
SeparationType	GiniIndex
MinNodeSize	20 %
nCuts	200
PruneMethod	nopruning
MaxDepth	3
NegWeightTreatment	InverseBoostNegWeights

Table 5.7: *Training variables.*

Particle	Variable
Λ_c^+, D_s^+, D	p, $\log p_T$, χ^2_{EV}, χ^2_{FD}
Daughters	p, $\log p_T$, $\log \eta$, $\log \chi^2_{\mathrm{FD}}$, $\log \chi^2_{\mathrm{IP}}$, $\arctan \log \mathcal{L}_{\mathrm{Muon}}(\mathrm{bkg})$, $\arctan \log \mathcal{L}_{\mathrm{Muon}}(\mu)$, $N_{\mathrm{shared}}(\mu)$ $b_{\mathrm{thresh}}(\mathrm{K})$, $b_{\mathrm{thresh}}(\pi)$, $b_{\mathrm{thresh}}(p)$, $\arctan \log \mathcal{L}_{\mathrm{RICH}}(\mathrm{bt})$, $\arctan \mathcal{L}_{\mathrm{RICH}}(e-\pi)$, $\arctan \mathcal{L}_{\mathrm{RICH}}(\mathrm{K}-\pi)$, $\arctan \mathcal{L}_{\mathrm{RICH}}(\mu-\pi)$, $\arctan \mathcal{L}_{\mathrm{RICH}}(p-\pi)$, $\log \mathrm{Prob}_{\mathrm{NN}}(\mathrm{Ghost})$, $\log \chi^2_{\mathrm{Match}}$, $\log \chi^2_{\mathrm{T,NDOF}}$, $\log \chi^2_{\mathrm{Velo,NDOF}}$, $b_{\mathrm{RICH},1}$, $b_{\mathrm{RICH},2}$, $b_{\mathrm{RICH}}(\mathrm{aerogel})$, $\log Q_{\mathrm{Velo}}$, $b_{\mathrm{PID}}(\mu)$

determined by comparing the training and testing distributions. There are no significant indications for overtraining in the BDT response plots displayed below.

A typical problem is to find the optimal working point of this variable or, in other words, to find the optimal cut value. This value can be found by studying the *Receiver-Operating-Characteristic* (ROC) curve. The ROC curve indicates the performance of a classification algorithm by plotting the relation between signal efficiency ε_{sig} and background rejection $1 - \varepsilon_{\text{bkg}}$ for all possible cut values. A particular value of the MVA variable corresponds to one point of the ROC curve. The background rejection becomes worse for higher signal efficiencies and vice versa because the distributions of MVA variable for signal and background usually overlap. The method with the largest area underneath the curve will provide on average the best performance. The areas are displayed in the ROC curves below for each BDT.

5.3.5 BDT efficiencies

The final question is of course whether the method by training several BDTs to provide PID-like variables has solved the problem of not well-understood efficiencies as mentioned at the beginning. A possible factorisation of the BDT efficiencies has been tested. Assuming that the efficiencies $\varepsilon_{\Lambda_c^+}, \varepsilon_{D_s^+}$ of the BDTs for Λ_c^+ and D_s^+ factorise, $\varepsilon_{\text{tot}} = \varepsilon_{\Lambda_c^+} \times \varepsilon_{D_s^+}$, the efficiency-corrected yields are calculated similar as described in 4.1. Figure 5.13 shows the efficiency-corrected yields in comparison with the yields without PID cuts clearly demonstrating that the BDT efficiencies for the c hadrons do factorise. A more detailed description of the correlation of the efficiencies can be found in [23].

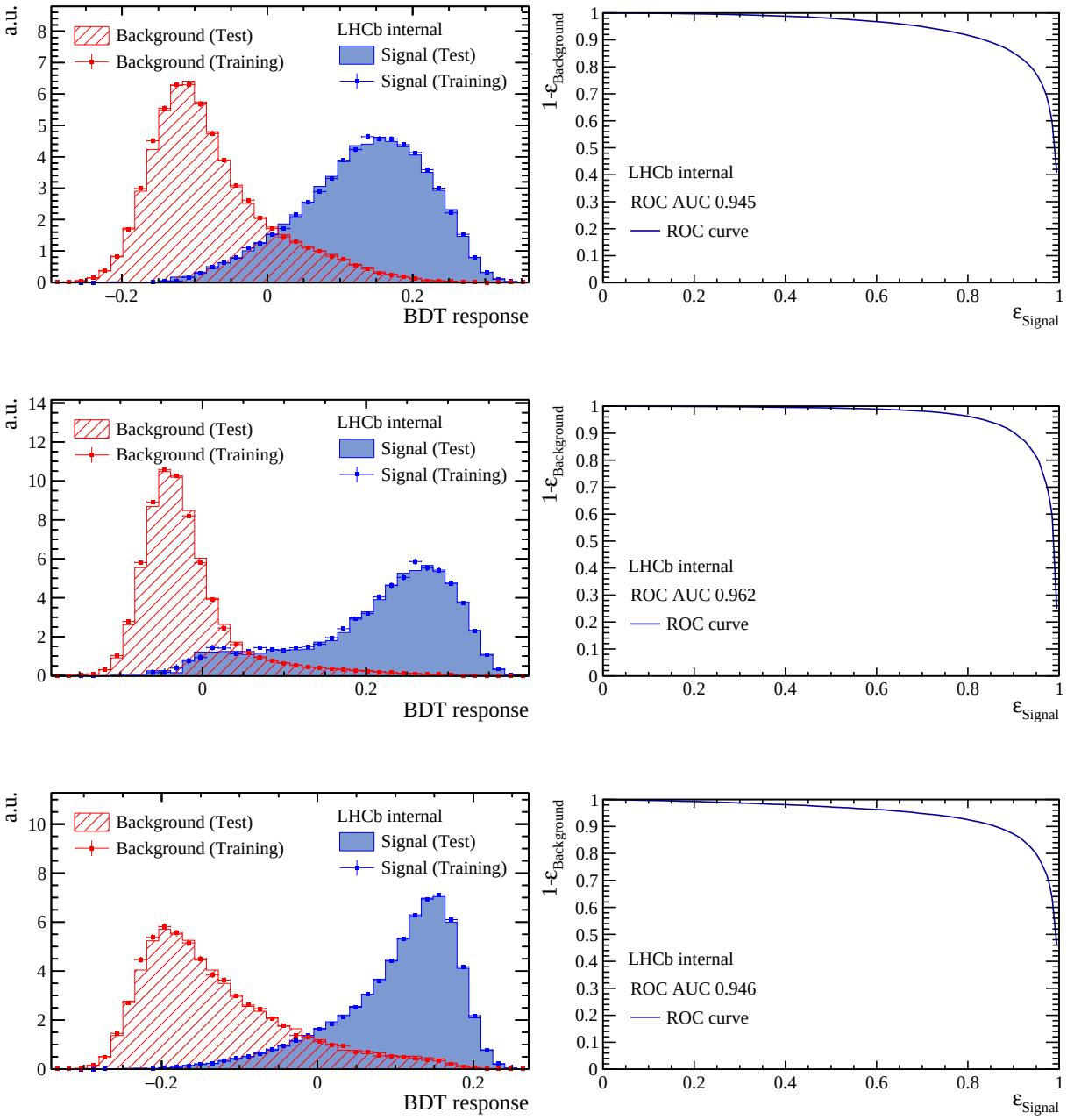

Figure 5.12: *Output plots from TMVA after training a BDT for* $\Lambda_c^+ \to pK^+\pi^-, D_s^+ \to K^+K^-\pi^-$ *and* $D^0 \to K^+\pi^-$. *Left: Signal vs. background for training and testing samples. Right: ROC curve.*

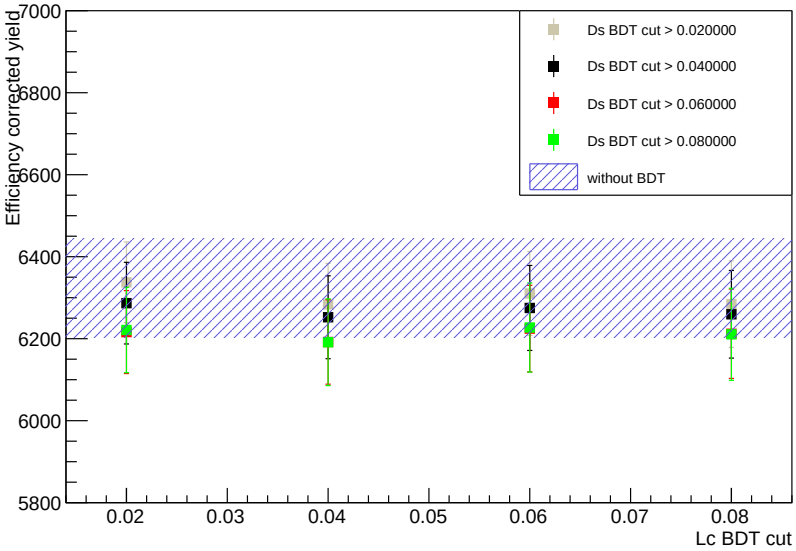

Figure 5.13: *Efficiency-corrected yields for the decay $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ showing the comparison with and without BDT cuts applied.*

6 Conclusion and Outlook

The goal of this thesis was to provide a classification algorithm having the ability to identify charmed hadrons from b decays. Such a classification was motivated by the search for exotic hadrons composed of five quarks in the decay $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$. It was found that the efficiencies provided by the standard PID variables are not well-understood for such a complicated decay involving six final state hadrons. They do not simply factorise and the correlation between them seems to be more complex. Instead of obtaining the real correlation of these PID efficiencies, which is not a trivial problem, the usage of these variables for the classification purpose at hand is completely avoided. By using multivariate algorithms, a PID-like variable was defined and calibrated on certain control channels, namely $\Lambda_b^0 \to \Lambda_c^+ \pi^-$, $\overline{B}_s^0 \to D_s^+ \pi^-$ and $B^- \to D^0 \pi^-$. This was achieved by training several Boosted Decision Trees for each type of charm hadron needed for the study of potential 5-quark resonances in $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ and related control channels. Those are Λ_c^+, $\overline{D}^0$ and D_s^+.

The BDTs provides a PID-like variable for the selection of a charmed hadron coming from a b decay. The BDTs optimally combine the information from several tracks into one variable. Thus, all possible correlations between the variables related to the charm daughters are taken into account automatically. This procedure results in a PID-like variable, which is not defined per track (as the standard PID variables) but per c hadron.

This thesis reports the entire process to develop and validate these PID-like BDT based variables. Good signal and background separation was achieved for all three c hadrons under consideration. Furthermore, the correlation of these BDTs was tested and found to be negligible for the reference channel $\Lambda_b^0 \to \Lambda_c^+ D_s^+$. The next step is to check if this is as well true when the BDTs are applied in the selection of the decay $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$.

Bibliography

[1] R. Aaij et al. Observation of $J/\psi p$ resonances consistent with pentaquark states in $\Lambda_b^0 \to J/\psi p K^-$ decays. *Phys. Rev. Lett.*, 115:072001, 2015.

[2] Yong hui Lin et al. Decay behaviors of Pc hadronic molecules, 2017.

[3] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.

[4] M. Gell-Mann. A Schematic Model of Baryons and Mesons. *Phys. Lett.*, 8, 1964.

[5] R. Aaij et al. LHCb detector performance. *Int. J. Mod. Phys.*, A30:1530022, 2015.

[6] Rolf Lindner. URL http://cds.cern.ch/record/1087860.

[7] R. Aaij et al. Measurement of the track reconstruction efficiency at LHCb. *JINST*, 10:P02007, 2015.

[8] R. Aaij et al. Performance of the LHCb Vertex Locator. *JINST*, 9:P09007, 2014.

[9] M. Adinolfi et al. Performance of the LHCb RICH detector at the LHC. *Eur. Phys. J.*, C73:2431, 2013.

[10] R. Aaij et al. Performance of the LHCb calorimeters. 2014.

[11] A. A. Alves Jr. et al. Performance of the LHCb muon system. *JINST*, 8:P02022, 2013.

[12] F. Archilli et al. Performance of the muon identification at LHCb. *JINST*, 8:P10020, 2013.

[13] R. Aaij et al. The PIDCalib package, 2016.

[14] R. Aaij et al. The LHCb trigger and its performance in 2011. *JINST*, 8:P04022, 2013.

[15] URL https://lhcb-public.web.cern.ch/lhcb-public/en/DataCollection/Triggers-en.html.

[16] Software Carpentry. The LHCb data flow. URL https://lhcb.github.io/first-analysis-steps/dataflow.html.

[17] A. Hoecker et al. TMVA – Toolkit for Multivariate Data Analysis. 2013.

[18] M. Williams. Generic D from B Selections. Technical Report LHCb-INT-2012-002, CERN, 2012.

[19] S. Hall and S. Neubert. Generic BDTs implemented in a TupleTool. (LHCb-INT-2013-016), 2013.

[20] P. Guenther. Measurement of the branching fraction of the rare decay $B^+ \to K^{*+}\mu^+\mu^-$ exploiting the decay $K^{*+} \to K^+\pi^0$ decay at the LHCb experiment, 2015. Bachelor Thesis.

[21] M Pivk and F. R. Le Diberder. splot: a statistical tool to unfold data distributions. *Nucl.Instrum.Meth.A555:356-369,2005*, 2004.

[22] O. Behnke et al. *Data Analysis in High Energy Physics.* Wiley-VCH, 2013.

[23] F. Reiss. . Master's thesis, University Heidelberg, 2017. in preparation.

Acknowledgement

At first, I would like to express my gratitude to my supervisor Prof. Dr. Stephanie Hansmann-Menzemer for giving me the opportunity to work and to write my bachelor thesis in the LHCb group. I am also grateful to Prof. Dr. Klaus Reygers for being the second referee of this thesis.

I would particularly like to thank Dr. Sebastian Neubert, Marian Stahl, Alessio Piucci and Florian Reiss for inspiring discussions, valuable tips to fix software problems, helping me to write this thesis and always answering my questions. It was a great pleasure to work with you. I felt very comfortable in this group. Thank you!

Finally, I want to thank my parents and my siblings for their never-ending moral support during the last years, especially the last months.

A Appendix

A.1 Stripping lines

The stripping lines used in this thesis are listed in Table A.1. The information on all strippings can be found in the *stripping project website* including all the algorithms run and cuts applied in each line:
http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/stripping/

Table A.1: *Stripping lines.*

Decay	Stripping line
$\Lambda_b^0 \to \Lambda_c^+ \pi^-$	`Lb2LcPiNoIPLc2PKPiBeauty2CharmLine`
$B_s^0 \to D_s^+ \pi^-$	`B02DPiD2HHHBeauty2CharmLine`
$B^- \to D^0 \pi^-$	`B2DPiPiD2HHHCFPIDBeauty2CharmLine`

Erklärung

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

Heidelberg, den 13. März 2017