

Department of Physics and Astronomy

University of Heidelberg

Diploma Thesis

in Physics

submitted by

Christopher Otto Alfred Vahl

born in Heidelberg

2017

**First Observation of the Decay $\Lambda_b^0 \rightarrow J/\psi \Lambda^0 \phi$ with
the LHCb Experiment**

This diploma thesis has been carried out by

Christopher Otto Alfred Vahl

at the

Physikalisches Institut

under the supervision of

Prof. Dr. Ulrich Uwer

Zusammenfassung

In dieser Arbeit wird zum ersten Mal das Verzweigungsverhältnis des seltenen Zerfalls $\Lambda_b^0 \rightarrow J/\psi\Lambda^0\phi$ relativ zu $\Lambda_b^0 \rightarrow J/\psi\Lambda^0$ gemessen. Der zugrundeliegende Datensatz hat eine integrierte Luminosität von 3 fb^{-1} und wurde mit dem LHCb-Experiment in den Jahren 2011 und 2012 aus Proton-Proton-Kollisionen mit einer Schwerpunktsenergie von 6.5 TeV (2011) beziehungsweise 7 TeV (2012) gewonnen. Die Anzahl der aus dem Datensatz extrahierten $\Lambda_b^0 \rightarrow J/\psi\Lambda^0\phi$ -Zerfälle beträgt 81.8 ± 11.1 (stat.). Unter Einbezug von Rekonstruktions- und Selektionseffizienzen ergibt sich das relative Verzweigungsverhältnis zu

$$\frac{\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi\Lambda^0\phi)}{\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi\Lambda^0)} = \{5.29 \pm 0.72 \text{ (stat.)} \pm 0.18 \text{ (syst.)}\} \times 10^{-2}.$$

Abstract

This thesis presents the first branching fraction measurement of the rare decay $\Lambda_b^0 \rightarrow J/\psi\Lambda^0\phi$ relative to $\Lambda_b^0 \rightarrow J/\psi\Lambda^0$, using a data set with an integrated luminosity of 3 fb^{-1} collected by the LHCb experiment from proton-proton collisions in 2011 and 2012 with a center-of-mass energy of 6.5 TeV and 7 TeV, respectively. The number of $\Lambda_b^0 \rightarrow J/\psi\Lambda^0\phi$ decays extracted from the data set is 81.8 ± 11.1 (stat.). After correcting for reconstruction and selection efficiencies, the branching fraction ratio is found to be

$$\frac{\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi\Lambda^0\phi)}{\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi\Lambda^0)} = \{5.29 \pm 0.72 \text{ (stat.)} \pm 0.18 \text{ (syst.)}\} \times 10^{-2}.$$

Contents

I	Introduction	11
1	Introduction and Motivation	12
1.1	QCD, Color Confinement, and Exotic Hadrons	12
1.2	Observation of Pentaquark Candidates	13
1.3	Strange Pentaquarks?	14
1.4	Pentaquark Models	15
2	The LHCb Experiment	18
2.1	The Large Hadron Collider	18
2.2	The LHCb Detector	18
2.2.1	Tracking System	20
2.2.2	Particle Identification System	25
2.2.3	Trigger System	28
2.2.4	Monte Carlo Simulation	30
II	Analysis	33
3	Analysis Strategy	34
3.1	Formula for the Branching Fraction	34
3.2	Decay Topology	34
3.3	Reference Channel	36
3.4	Steps of the Analysis	36
4	Data Selection	38
4.1	Important Selection Variables	38
4.2	Trigger Strategy	39
4.3	Cut-Based Selection	40
4.3.1	Stripping	40
4.3.2	Preselection	40
4.4	Background Sources	43
4.4.1	Combinatorial Background	43
4.4.2	Background from Misidentified Particles	43

4.4.3	Partially Reconstructed Decays as Background	47
4.4.4	Non-Resonant Decays as Background	47
4.5	Multiple Candidates	47
4.6	Multivariate Selection	49
4.6.1	Training of Boosted Decision Trees	50
4.6.2	Working-Point Optimization	52
5	Signal Extraction	54
5.1	Fit Method	54
5.2	Decay Tree Fitter	55
5.3	Signal Channel Fits	55
5.4	Reference Channel Fits	56
5.5	Sweights	59
6	Efficiencies	60
6.1	Correcting Discrepancies Between Data and MC	60
6.2	Efficiencies as a Function of the Dalitz Plot	61
6.3	Dealing with Discrepancies in p_T and η of the Λ_b^0	61
6.3.1	Reweighting the Signal Channel MC	62
6.4	Acceptance Cut Efficiencies	65
6.5	Reconstruction, Trigger, and Selection Efficiencies	68
6.5.1	Uncertainties on Efficiencies and Adaptive Binning	68
6.6	Average Efficiencies	70
6.7	BDT Efficiencies	70
7	Determination of the Branching Fraction	72
7.1	Efficiency-Corrected Yields	72
7.2	Determination of the Branching Fraction for LL and DD Samples	73
8	Systematic Studies	74
8.1	LL/DD Discrepancies	74
8.2	Track Reconstruction Efficiency of the Kaons	75
8.3	Helicity Angle of the ϕ	77
8.4	Fit Model	77
8.5	Summary of Systematic Uncertainties	78
9	Results	80
9.1	Weighted Mean Branching Fraction	80
9.2	Conclusion	81

III Appendix	83
A BDT Input Variables	84
A.1 Data versus MC	84
A.2 Training Samples	87
B Efficiency Plots	89
C Bibliography	91

Part I

Introduction

1 Introduction and Motivation

In this analysis, the branching fraction of the decay $\Lambda_b^0 \rightarrow J/\psi\Lambda^0\phi$ is measured relative to that of the decay $\Lambda_b^0 \rightarrow J/\psi\Lambda^0$ (charge conjugation is implied throughout the text). This is the first observation of the decay mode $\Lambda_b^0 \rightarrow J/\psi\Lambda^0\phi$ and establishing this channel is a first step in the search for resonances in the $J/\psi\Lambda^0$ subsystem. These resonances would have a minimal quark content of $c\bar{c}uds$ and thus would be candidates for pentaquarks containing an s quark. The search for strange pentaquarks is motivated by the recent discovery of pentaquark candidates in the subsystem $J/\psi p$ of the decay $\Lambda_b^0 \rightarrow J/\psi p K^-$ by the LHCb Collaboration [1, 2]. These candidates have the minimal quark content $c\bar{c}uud$.

The current chapter begins with a very brief introduction to QCD and color confinement, with the latter giving a motivation to search for exotic hadrons with quark content $qq\bar{q}\bar{q}$ or $qqqq\bar{q}$. This is followed by a review of the observed pentaquark states and several approaches trying to explain the nature of these states.

1.1 QCD, Color Confinement, and Exotic Hadrons

The Standard Model of particle physics describes three of the four known fundamental forces of interaction between elementary particles: the electromagnetic, the weak, and the strong force, with gravity not yet being incorporated into the model.

The theory describing the strong force is called quantum chromodynamics (QCD). QCD is a non-abelian gauge theory with symmetry group $SU(3)$. The charge of QCD is color, analogous to the electric charge of QED. Gluons are the force-mediating particles of QCD; they are massless and electrically neutral particles with spin 1. There exist eight types of gluons corresponding to the eight generators of the $SU(3)$ group, and three colors (red, green, and blue denoted as r , g , and b) corresponding to the three orthogonal states of the $SU(3)$ color space. The fundamental fermions of QCD are the quarks, carrying color charge; antiquarks carry anticolor (\bar{r} , \bar{g} , or \bar{b}). Quarks and antiquarks only exist in color-neutral bound states, called hadrons. The phenomenon that only color-neutral states are observed is called *color confinement*.

In the current state of theory, color confinement is explained as follows [3]: Gluons carry color charge, opposed to the photon carrying no electric charge. This charge leads to the existence of an attractive force between gluons, which is non-existent in case of photons. When two quarks are pulled apart, the energy density stored in the gluon field between them becomes constant at relatively large distances due to the attractive force described above. This constant energy density means it would

take an infinite amount of energy to completely separate two quarks. Thus, at a certain distance, the creation of new quarks in the gluon field, which then combine to color-neutral hadrons, becomes energetically favorable to a further increase of energy stored in the field. This process of forming new hadrons is called *hadronization*. The exact mechanism of color confinement has not yet been derived from the theory.

There exist two experimentally established types of hadrons, differentiated by their valence quark content (the quarks responsible for the quantum numbers of the hadron): baryons, consisting of three quarks qqq (also anti-baryons with $\bar{q}\bar{q}\bar{q}$) and mesons, consisting of a quark and an anti-quark $q\bar{q}$. In a simplified picture¹, the three quarks of a baryon have color charge rgb , the quark-antiquark pair of a meson has $r\bar{r}$, $g\bar{g}$, or $b\bar{b}$, all resulting in color-neutral hadrons. However, QCD also allows for the existence of hadrons with a higher number of valence quarks, so-called tetraquarks ($qq\bar{q}\bar{q}$) and pentaquarks ($qqqq\bar{q}$). In Section 1.4, two models are presented which differ in the explanation of why pentaquarks are color-neutral. The term *pentaquark* was coined by Lipkin [4] in 1987, but the possibility of the existence of these exotic states was already considered in 1964 by Gell-Mann [5] and Zweig [6].

1.2 Observation of Pentaquark Candidates

In the 2000s, several experiments presented signs of evidence for the existence of a pentaquark comprised of four light quarks and a strange antiquark, termed Θ^+ . A large number of publications explored the nature of this state. However, later experiments, producing higher statistics, could not confirm the existence of the Θ^+ . For an overview of this episode of particle physics, see [7].

In 2015, LHCb published evidence for resonances in the $J/\psi p$ subsystem of the decay $\Lambda_b^0 \rightarrow J/\psi p K^-$ [1]. These were found by performing an amplitude analysis and they have a minimal quark content of $c\bar{c}uud$ and thus are pentaquark candidates.

An amplitude analysis is a tool to obtain information about intermediate resonances, which are not directly observable. Such an analysis consists of the following steps: First, a model describing the decay amplitude (e.g. of $\Lambda_b^0 \rightarrow J/\psi p K^-$) is built, which includes intermediate resonances and is parameterized in terms of the properties of these resonances, such as their masses, widths and spins. From this model, observable distributions of decay angles and invariant masses of subsystems are derived, which then depend on the number and properties of the intermediate resonances. These distributions are fitted to the observed distributions, thereby inferring the presence of intermediate resonances and the values of their properties.

In case of $\Lambda_b^0 \rightarrow J/\psi p K^-$, five decay angles and the invariant mass of the pK^- subsystem were used. A first fit model, which included 14 Λ^* -resonances listed by

¹The actual color wave functions are superpositions of several color states, which can be constructed with the help of the ladder operators of the SU(3) group.

Table 1.1: Masses and widths of the states observed in [1].

State	Mass [MeV/c ²]	Width [MeV/c ²]
$P_c^+(4380)$	4380 ± 8 (stat.) ± 29 (syst.)	205 ± 18 (stat.) ± 86 (syst.)
$P_c^+(4450)$	4449.8 ± 1.7 (stat.) ± 2.5 (syst.)	39 ± 5 (stat.) ± 19 (syst.)

the Particle Data Group [8], did not result in a satisfactory fit quality. It was found that adding two resonances in the $J/\psi p$ system leads to a satisfactory fit. These resonances were labeled $P_c^+(4380)$ and $P_c^+(4450)$; their masses and widths are listed in Table 1.1. Several spin configurations gave satisfactory fit results, but it could be concluded that the two resonances have opposite parities. The obtained significances of the $P_c^+(4380)$ and $P_c^+(4450)$ are 9σ and 12σ , respectively.

A follow-up analysis [2] employed a model-independent approach, making minimal assumptions about the spin and mass of the pK^- contributions and none about their number and resonant or nonresonant nature. The hypothesis that the $\Lambda_b^0 \rightarrow J/\psi p K^-$ decays only involve intermediate resonances in pK^- could be excluded with a significance of more than 9σ , thus giving strong evidence for the need of resonances in $J/\psi p$.

In a third analysis [9], an amplitude analysis was performed on $\Lambda_b^0 \rightarrow J/\psi p \pi^-$, which differs from the channel in which the two P_c^+ states were first observed by a π^- instead of a K^- in the final state. It was found that adding the $P_c^+(4380)$ and $P_c^+(4450)$ significantly improves the fit quality.

1.3 Strange Pentaquarks?

If the above P_c^+ states indeed are pentaquarks, similar resonances with different quark contents should exist. One such possible resonance would differ from the P_c^+ by a unit of strangeness, having the quark content $c\bar{c}uds$ instead of $c\bar{c}uud$. This resonance may be called P_{cs}^0 . One channel in which the P_{cs}^0 might be observed is $\Lambda_b^0 \rightarrow P_{cs}^0 \phi$ with $P_{cs}^0 \rightarrow J/\psi \Lambda^0$. Feynman diagrams of the decay $\Lambda_b^0 \rightarrow J/\psi p K^-$, in which the P_c^+ was found, and of $\Lambda_b^0 \rightarrow J/\psi \Lambda^0 \phi$ are shown in Figure 1.1. The latter decay is obtained from the former by replacing the $u\bar{u}$ pair with an $s\bar{s}$ pair.

Prior to this analysis, the decay $\Lambda_b^0 \rightarrow J/\psi \Lambda^0 \phi$ has not been observed. Consequently, performing a branching fraction measurement is the first step toward an amplitude analysis in this decay mode, which might reveal the P_{cs}^0 . Notice that resonances in each subsystem of this channel would be exotic, with a minimal quark content of $c\bar{c}uds$, $c\bar{c}s\bar{s}$, or $udss\bar{s}$ for the subsystems $J/\psi \Lambda^0$, $J/\psi \phi$, and $\Lambda^0 \phi$, respectively.

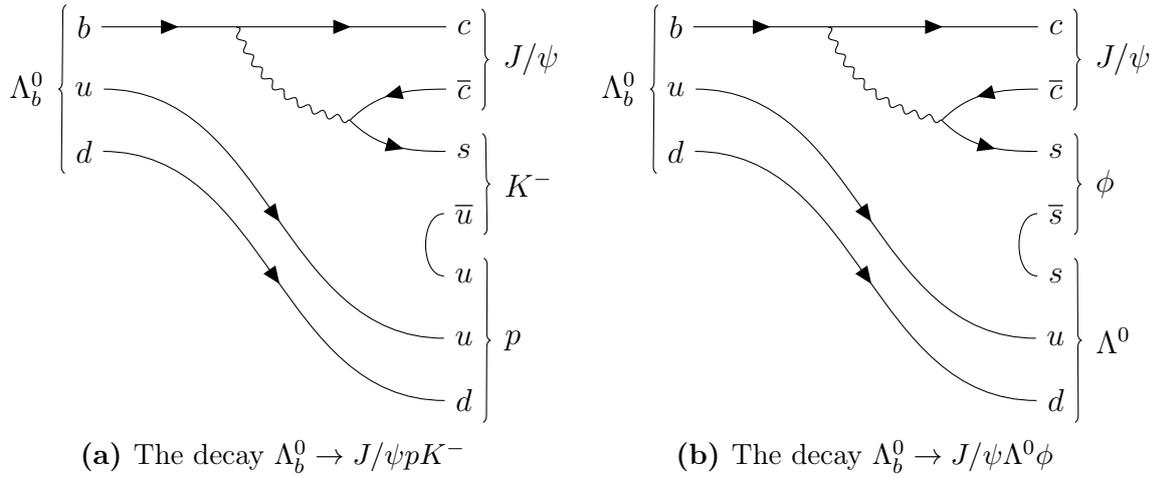


Figure 1.1: Feynman diagrams of decays with possible pentaquark contributions.

1.4 Pentaquark Models

The observation of the $P_c^+(4380)$ and $P_c^+(4450)$ resulted in a large response by the theory community. A review of several models trying to explain the observed states can be found in [10]. Here, only a brief description of three such models is given.

Rescattering

Some authors, e.g. [11], attribute the peaks in the invariant mass spectrum of the $J/\psi p$ system to purely kinematic *rescattering* effects. Such effects can lead to singularities in the scattering amplitude, but, opposed to resonances, have no pole in the imaginary part. In case of the $P_c^+(4450)$, the rescattering could occur via $\chi_{c1} p \rightarrow J/\psi p$.

Molecular Models

In the *molecular model*, pentaquarks are described as molecules consisting of lightly bound, color-neutral hadrons, specifically a baryon and a meson. A well-known system with such a binding mechanism is the deuteron, composed of a proton and a neutron. If a resonance's mass is close to the threshold of a two-particle system, this points toward the resonance being of molecular nature. The $P_c^+(4380)$ is close to $\Sigma_c^+(2520)\bar{D}^0$ with $m(\Sigma_c^+(2520)) = (2517.5 \pm 2.3) \text{ MeV}/c^2$ and $m(\bar{D}^0) = (1864.83 \pm 0.05) \text{ MeV}/c^2$; the $P_c^+(4450)$ is close to a number of thresholds, including $\Lambda_c^+(2595)\bar{D}^0$ with $m(\Lambda_c^+(2595)) = (2592.25 \pm 0.28) \text{ MeV}/c^2$. The attractive force between the baryon and meson is mediated via meson exchange. For a review of molecular models, see [12]. All models presented in this publication predict the existence of neutral partners to the two observed states. The J^P values of

these two states are limited by the partial waves contributing to the binding. Opposite parities can only be explained with P-wave or higher contributions. Decays of molecules with such bindings are suppressed by the centrifugal barrier, leading to a broader resonance.

Tightly Bound Pentaquarks

In the model of *tightly bound pentaquarks*, the two P_c^+ states are hypothesized to consist of colored subsystems, opposed to the colorless subsystems in the molecular model. There exists a wide variety of possible combinations of such colored subsystems, one of which is presented here.

Maiani, Polosa, and Riquer [13] consider the P_c^+ in the *diquark model*. Taking the preferred J^P configurations of the fit from the first LHCb publication [1], they assume the pentaquarks to have the following constituents:

$$\begin{aligned} P_c^+(4380, 3/2^-) &= \bar{c}[cq]_{s=1}[qq]_{s=1}, L = 0, \\ P_c^+(4450, 5/2^+) &= \bar{c}[cq]_{s=1}[qq]_{s=0}, L = 1, \end{aligned}$$

where $[cq]$ and $[qq]$ are heavy and light diquarks, respectively.

The above antiquark-diquark-diquark configuration results in a colorless state. This can be seen as follows: Mesons are colorless because their valence quark carries color and their valence antiquark carries anticolor. Baryons are also colorless, so two of their three valence quarks must be combinable to an anticolor-carrying diquark. Consequently, a pentaquark, having the constituents antiquark-diquark-diquark, has the color configuration anticolor-anticolor-anticolor, which is equivalent to the configuration of a (colorless) anti-baryon.

The opposite parities are explained by the additional angular momentum unit of $P_c(4450)$. The mass difference of the two pentaquarks results from the following two summands: the orbital excitation and the mass difference between diquarks with $s = 0$ and $s = 1$, which are estimated to be of the order of $280 \text{ MeV}/c^2$ and $200 \text{ MeV}/c^2$, respectively, giving a total mass difference of about $80 \text{ MeV}/c^2$.

Can Strange Pentaquarks Be Observed?

Several studies considered the potential to observe strange partners of the P_c^+ . These would have the quark content $c\bar{c}uds$. Feijoo et al. [14] predicted the signature of strange pentaquarks in the decay channel $\Lambda_b^0 \rightarrow J/\psi\Lambda^0\eta$. The method they used is independent of the binding mechanism of the pentaquarks. Feijoo et al. concluded that strange pentaquarks could leave an observable signature in the invariant mass spectrum of $J/\psi\Lambda^0$, independent of variations in such model parameters that are not precisely known.

Lu et al. [15] performed a similar analysis for the channel $\Lambda_b^0 \rightarrow J/\psi \Lambda^0 K^0$ and also found that the strange pentaquarks could leave an observable signature, with the signature not being sensitive to sensible variations in model parameters.

The decay channel of this analysis is very similar to the above two channels, the only difference being a ϕ instead of an η or K^0 in the final state. This similarity lets us conclude that establishing the $J/\psi \Lambda^0 \phi$ decay mode of the Λ_b^0 with the ultimate goal of an amplitude analysis is a worthwhile endeavor.

Chapter 2 introduces the LHCb detector, an outline of the analysis is given in Chapter 3, and the subsequent chapters describe the analysis in detail.

2 The LHCb Experiment

This chapter introduces the LHC accelerator environment and the layout and functionality of the LHCb detector.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [16], located near Geneva at the France-Switzerland border, is the largest particle accelerator in the world, with a circumference of 27 km. In LHC's main mode of operation, protons are collided with protons, but the machine is also capable of accelerating heavy ions. During the data-taking periods of 2011, 2012, and 2015 onward, the pp collisions had a center-of-mass energy of 7 TeV, 8 TeV, and 13 TeV, respectively. This analysis uses data from 2011 and 2012.

Before entering the LHC, the protons pass through a succession of preaccelerators, reaching a beam energy of 450 GeV at injection. They then are further accelerated inside the LHC by radio-frequency cavities and kept on their circular path by superconducting magnets cooled with superfluid helium. At a beam energy of 6.5 TeV, the field created by the magnets has a strength of 7.7 T. The two proton beams are contained within two separate vacuum pipes with a distance of only 19 cm between the pipes. The protons are accelerated in bunches with a maximum bunch frequency of 40 MHz and approximately 10^{11} protons per bunch.

The beams are brought to collision at four interaction points around the ring. The four large experiments located at these points are ATLAS, CMS, ALICE, and LHCb. ATLAS and CMS are general-purpose detectors, ALICE specializes in heavy-ion collisions, and LHCb was designed to study b hadron physics.

2.2 The LHCb Detector

The LHCb detector [17] is a single-arm forward-spectrometer covering an angle of approximately 10 mrad to 300 mrad in the plane in which the particles are bent by the magnet, and 10 mrad to 250 mrad in the non-bending plane. This corresponds to a pseudorapidity range of approximately $2 < \eta < 5$ with η being related to the polar angle θ between the beam direction and the particle's trajectory by $\eta = -2 \ln(\tan \theta/2)$.

The detector only covers this relatively narrow range because b hadrons are predominantly produced with a small angle with respect to the beam axis. Figure 2.1

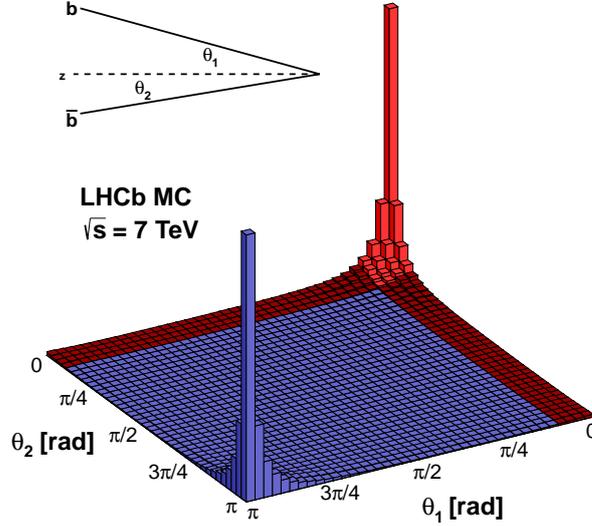


Figure 2.1: Simulated polar angle distribution of $b\bar{b}$ pairs produced in pp collisions at a center-of-mass energy of 7 TeV. The LHCb acceptance is marked in red. Figure taken from [19].

shows that a large part of the $b\bar{b}$ pairs lie within the acceptance of the detector. The reason for this is the following: At LHC, $b\bar{b}$ pairs are dominantly produced by quark anti-quark annihilation and gluon fusion. The small threshold of $b\bar{b}$ production compared to the beam energy leads to the quarks or gluons producing the $b\bar{b}$ pair to cover a wide p_z range, where p_z is the momentum parallel to the beam axis. This spread in p_z makes it very likely for the two quarks or gluons to have very different p_z values, resulting in a boost in the beam direction.

The main research areas of LHCb are CP violation and rare decays of B and D mesons. To perform the precision measurements necessary in these fields, a low number of pp collisions per bunch crossing is desirable, making the events easier to analyze. An average of 1.7 pp collisions per bunch crossing is achieved by slightly offsetting the proton beams relative to each other, resulting in a modest luminosity of $2.8 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$ for 2011 and $4.0 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$ for 2012 [18], and an integrated luminosity of 1 fb^{-1} and 2 fb^{-1} , respectively.

Figure 2.2 shows a profile of the detector. The coordinate system on this profile, which will be used throughout the text, is right-handed Cartesian with the z -axis pointing in beam direction toward the detector, the y -axis pointing upward and the x -axis pointing to the left side of the detector.

The components of the detector can be divided into two categories: the tracking

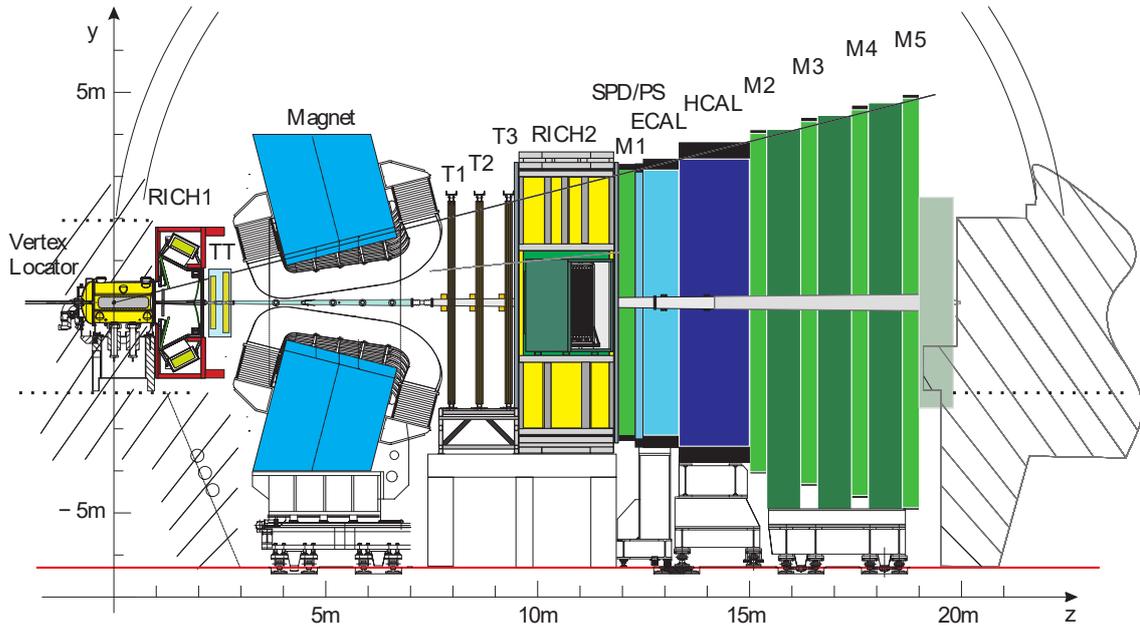


Figure 2.2: Profile of the LHCb detector in the yz -plane with Vertex Locator (VELO), ring-imaging Cherenkov detectors (RICH1 and RICH2), Tracker Turicensis (TT), magnet, tracking stations (T1-T3), Scintillating Pad Detector (SPD), Pre-Shower Detector (PS), Electromagnetic Calorimeter (ECAL), Hadronic Calorimeter (HCAL), and Muon System (M1-M5). Image taken from [17].

system and the particle identification system. The former consists of the Vertex Locator (VELO), the Tracker Turicensis (TT), the magnet and the tracking stations T1-T3, while the latter is comprised of two ring-imaging Cherenkov detectors (RICH1 and RICH2), the Scintillating Pad Detector (SPD), the Pre-Shower Detector (PS), the Electromagnetic and Hadronic Calorimeters (ECAL and HCAL), and the Muon System (M1-M5). The following sections briefly describe the individual components, details can be found in [17].

2.2.1 Tracking System

The tracking system provides the information necessary to reconstruct trajectories and momenta of charged particles. It does so by measuring the positions of particles in the xy -plane at different z -positions. Correspondingly, all tracking detectors (VELO, TT, and T1-T3) are arranged in layers orthogonal to the proton beams. The VELO, the TT and the parts of T1-T3 close to the beam, called the Inner Tracker, are comprised of silicon microstrip detectors, while the outer parts of T1-T3, called the Outer Tracker, consist of straw tubes.

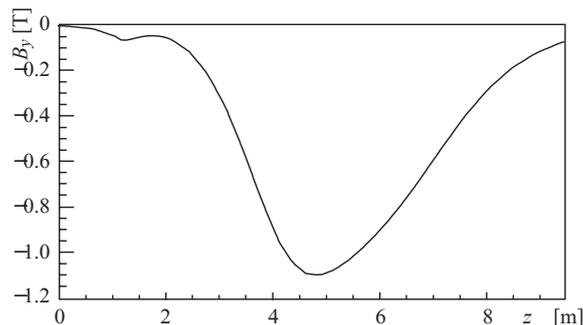


Figure 2.3: Field strength of the magnet. Figure taken from [20].

Magnet

The magnet is located downstream of the Tracker Turicensis and upstream of the tracking stations T1-T3 at a distance of 5 m from the interaction region. The magnetic field, shown in Figure 2.3, is nearly homogeneous in y -direction with very small components in x - and z -direction, resulting in the particles being bent in the xz -plane. The field strength is $\int B dl \approx 4 \text{ Tm}$ for a particle traveling 10 m in z -direction and is very small at the location of the VELO, the TT and T1-T3, resulting in approximately straight tracks in these subdetectors. The magnet polarity is switched regularly during data taking, so positively charged particles are either bent into the left or right half of the detector, making it possible to study systematic effects due to detector asymmetries with data sets for individual polarities, or to use data sets with combined polarities, in which these uncertainties should cancel.

Vertex Locator (VELO)

The VELO [21] is a silicon microstrip detector surrounding the interaction region of the proton beams. It precisely measures track coordinates near the interaction region, which are used to separate vertices of the pp interactions (primary vertices, PV) from vertices of b hadron decays (secondary vertices, SV), and also from vertices of decays of b hadron daughters (tertiary vertices, TV). Hadrons containing a b quark can only decay via the weak force, resulting in a lifetime of the order of 1 ps and a flight distance of several mm. With a resolution of approximately $25 \mu\text{m}$, the VELO easily can resolve the SV from the PV. Additional to the vertex separation, the VELO provides a first measurement of a particle's trajectory, which is used as an independent segment in the track reconstruction (see Section 2.2.1).

The VELO, shown in Figure 2.4, consists of 21 stations with each station being comprised of an R-sensor with a circular strip structure, measuring the radial distance of the particles to the beam axis, and a ϕ -sensor with a radial strip structure, measuring the azimuthal angle. The third component is provided by the station's z -position. This cylindrical setup, opposed to a rectilinear one, leads to a faster

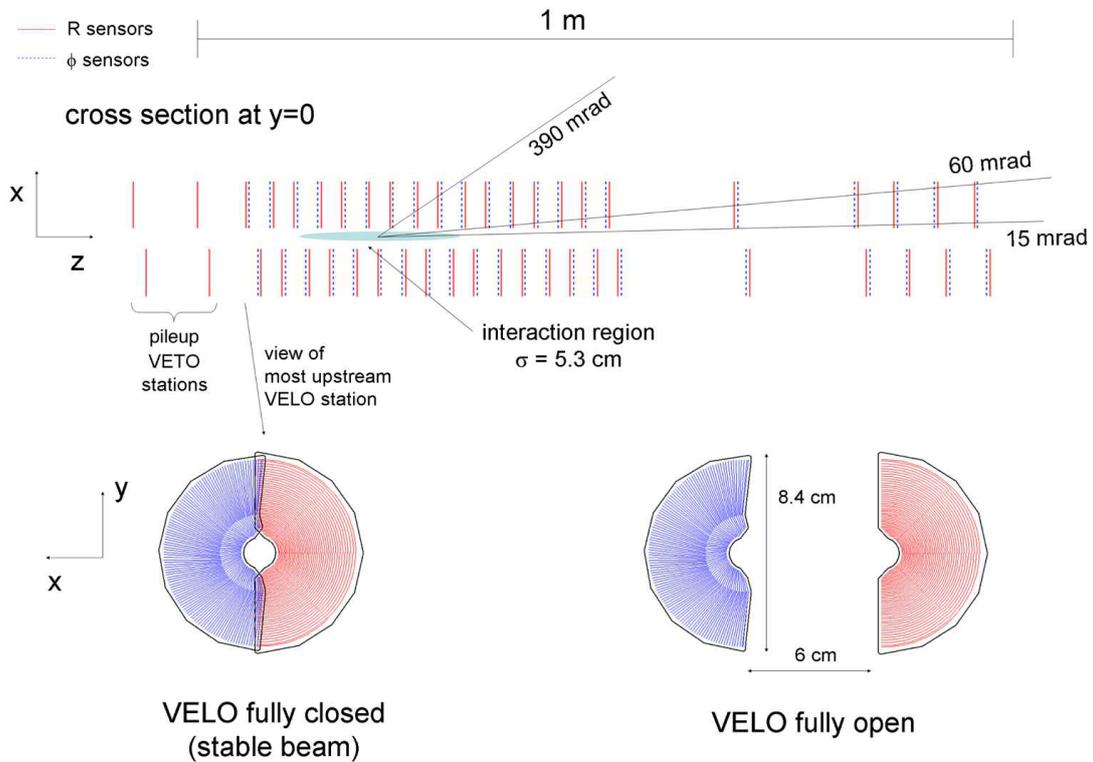


Figure 2.4: Top: Cross section of the complete VELO. Bottom: Cross section of the most upstream VELO station. Figure taken from [17].

track and vertex reconstruction. The sensors each are $300 \mu\text{m}$ thick and have a strip pitch varying from $40 \mu\text{m}$ to $100 \mu\text{m}$ with finer granularity near the beam.

Upstream of the interaction point, two pile-up stations are placed, which are used in the hardware trigger to veto events with multiple pp collisions.

In operation, the VELO has a distance of 8 mm from the beam axis, but during the injection phase can be retracted to prevent radiation damage. Inside the VELO, the proton beams are only separated from the detector by a radio-frequency shielding foil made of an aluminium alloy. This drastically reduces the material traversed by particles before reaching the VELO stations, minimizing undesirable scattering effects. The foil is necessary to protect the ultra-high vacuum of the LHC from outgassing of the detector, and to shield the detector from electromagnetic effects from the beam.

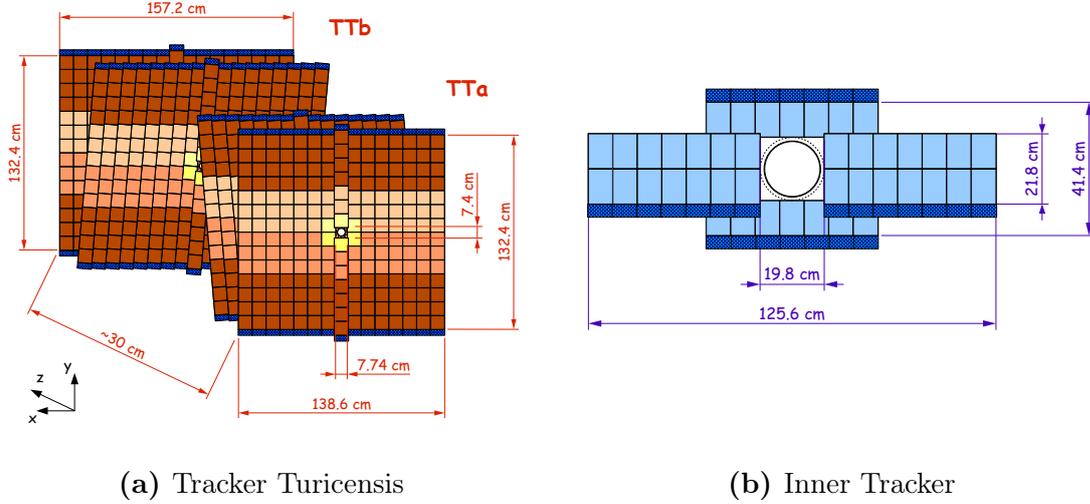


Figure 2.5: Cross section of the Tracker Turicensis and one module of the Inner Tracker. Figures taken from [22].

Tracker Turicensis (TT)

The TT is a silicon microstrip detector, located upstream in front of the magnet. The purpose of the TT is to provide position measurements for long-lived, neutral particles decaying into charged daughters outside of the VELO and consequently having no tracks in the VELO, and also for particles with low momentum, which may be bent out of the acceptance by the magnet. In both cases, without the TT there would exist insufficient information to infer the particle's momentum.

To measure the momentum of the long-lived, neutral Λ^0 , one of the Λ_b^0 daughters of the decay mode analyzed in this thesis, the TT is a crucial component, since a large part of the Λ^0 baryons decay outside of the VELO.

The TT, whose layout is shown in Figure 2.5a, is comprised of two stations separated by 27 cm, each consisting of two layers of silicon strips. The first and fourth layer have strips parallel to the y-axis, while the strips of the second and third layer are rotated by a stereo angle of ± 5 degrees with respect to the y-axis. This setup improves the spacial resolution in the y-direction. The layers each have a width of 150 cm and a height of 130 cm, covering the full acceptance. The strip pitch is $183 \mu\text{m}$ and the TT's resolution is approximately $50 \mu\text{m}$.

Tracking Stations T1-T3

The tracking stations T1-T3, also called T-stations, are comprised of two detector types. Close to the beam axis, the Inner Tracker is a silicon microstrip detector, while the Outer Tracker is a straw tube drift time detector. This layout was chosen to

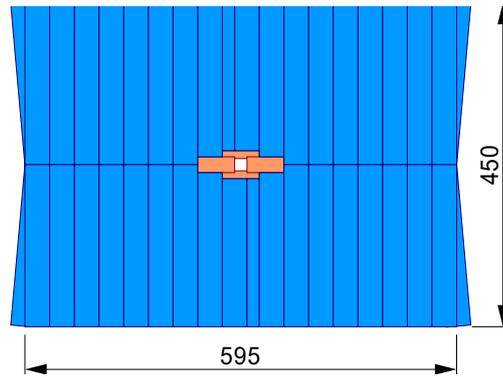


Figure 2.6: Cross section of a T-station with Inner Tracker (orange) and Outer Tracker (blue). Image taken from [23].

achieve a high spacial resolution in the inner area, which has a high particle flux, and to economically cover the outer acceptance region. Figure 2.6 shows the dimensions of the T-stations with the Inner Tracker covering only 1.3% of the station's area, but receiving 20% of the incoming particles.

The Inner Tracker of each T-station has a layout similar to that of the TT, consisting of four cross-shaped layers with the strips of the inner two being rotated by ± 5 degrees. The cross section of an outer layer with horizontal strip alignment is shown in Figure 2.5b. The tracker has a strip pitch of $197 \mu\text{m}$ and a resolution of approximately $50 \mu\text{m}$.

The Outer Tracker's gas tubes are filled with a mixture of Argon (70%) and CO_2 (30%), have a diameter of 4.9 mm, and are arranged in four layers per station with the tubes of the inner two layers being rotated by ± 5 degrees. The detector's drift time is below 50 ns and its resolution is approximately $200 \mu\text{m}$.

Track Reconstruction

A particle's trajectory can be reconstructed by several different algorithms. The two most common ones form *long tracks*, which are tracks with hits in the VELO and the T-stations. They both start by combining hits in the VELO to a track segment. In the *forward tracking* algorithm [24], this segment is then combined with matching hits in one of the stations T1-T3. With a VELO track and a hit in one of the T-stations, the momentum of the particle and its trajectory through the whole detector can be determined. Hits in other T-stations can be used to further strengthen the track hypothesis. The *track matching* algorithm [25] first forms separate track segments in the VELO and the T-stations and then matches these segments to form a complete track.

Neutral, long-lived particles decaying downstream of the VELO leave no hits in the VELO and thus need to be reconstructed from hits in the TT and the T-stations.

Details on the algorithm used to form these *downstream tracks* can be found in [26].

All algorithms benefit from the negligible magnetic field strength at the tracking stations, resulting in an approximately straight trajectory.

After the track reconstruction, a particle's momentum in most cases is obtained by measuring the dislocation of its track by the magnet. An exception are particles with low momentum, which may be bent out of the acceptance by the magnet. Consequently, their momentum needs to be measured with position information from the VELO and the TT, which both are located upstream of the magnet.

The combined tracking system has a momentum resolution between 0.4% at 5 GeV/c and 0.6% at 100 GeV/c for a track traversing the whole detector. The resolution is dominated by multiple scattering on the detector material over almost the entire range of particle momenta.

2.2.2 Particle Identification System

The particle identification system consists of the two RICH detectors, the calorimeters, and the Muon System. Electrons and photons are primarily identified by the calorimeters, charged hadrons by the RICH detectors, and muons by the Muon System. Neutral hadrons, such as the π^0 , are identified by the calorimeter system.

RICH1+2

The RICH detectors are based on the phenomenon of Cherenkov radiation, which is electromagnetic radiation emitted by a charged particle when it passes through a dielectric medium at a speed greater than the phase velocity of the speed of light in that medium. The phenomenon is analogous to the sonic boom of an aircraft flying with a speed faster than that of sound. Cherenkov radiation is emitted under the *Cherenkov angle* $\cos\theta_C = 1/(n\beta)$ with respect to the particle's direction of motion, with n being the refractive index of the traversed medium, called the radiator, and $\beta = v/c$.

The RICH detectors measure θ_C as follows: The light cones emitted by the particles traversing the radiator are projected onto photodetectors by a system of mirrors. The radius of the projected ring is proportional to the Cherenkov angle, which in turn directly gives the particle's speed. Combining this information with the momentum measured by the tracking system, an estimate of the particle's mass is obtained. The actual particle identification is performed by comparing the pattern observed by the photodetector to the pattern expected for the reconstructed track under several different mass hypotheses. Each mass hypothesis, corresponding to a particle-type hypothesis, then is assigned a likelihood value. These values are saved and can be accessed at later stages of the analysis.

To cover a large momentum range, LHCb employs two RICH detectors at different positions and using different radiator materials.

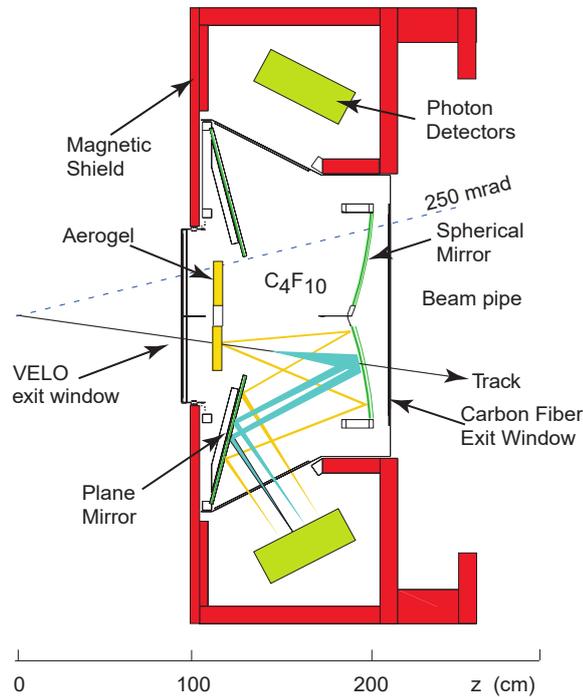


Figure 2.7: Side view schematic layout of the RICH1 detector. Image taken from [17].

RICH1, located upstream of the magnet between the VELO and the TT, identifies particles with a momentum of $1 \text{ GeV}/c$ to $60 \text{ GeV}/c$, including low-momentum particles, which may be bent out of the detector by the magnet. The radiators used are aerogel and C_4F_{10} gas. The layout of RICH1 is shown in Figure 2.7.

RICH2 is located between the last tracking station and the first muon station and is filled with gaseous CF_4 , which has a lower refractive index than the C_4F_{10} of RICH1. The detector identifies particles with momenta ranging from $15 \text{ GeV}/c$ to $100 \text{ GeV}/c$. These high-momentum particles most likely have a small polar angle, thus RICH2 does not cover the whole acceptance of LHCb, contrary to RICH1.

Calorimeters

The calorimeter system has several functions: It provides fast detection of electrons, photons, and hadrons with high transverse energy (E_T) for the hardware trigger. (*Transverse* in this text always means *transverse to the beam axis* and refers to the xy-plane of the coordinate system defined above.) Additionally, it distinguishes between electrons, photons and hadrons and measures their energies and positions. The identification of the hadron type, especially the separation of kaons from pions, is performed by the RICH detectors. The calorimeter is the only system capable of

reconstructing and identifying neutral particles, such as prompt photons and neutral pions.

All calorimeters consist of alternating layers of scintillating and absorbing materials. Traversing particles produce showers, which lead to the emission of photons in the scintillator. These photons are guided into photomultipliers by wavelength-shifting fibers. All subsystems have finer granularity close to the beam, with three segments in case of the SPD/PS and ECAL, and two segments in case of the HCAL.

Scintillator Pad Detector (SPD) and Pre-Shower Detector (PS) The SPD and the PS are the most upstream parts of the calorimeter system, located in front of the Electromagnetic Calorimeter. They are separated by 15 mm of lead, corresponding to 2.5 radiation lengths, and both consist of scintillating pads.

The SPD separates charged particles from neutral ones, mainly electrons from photons. It does so by measuring the energy deposition of the traversing particle. Opposed to electrons, photons deposit little to no energy. The PS distinguishes between electrons and charged hadrons, mainly charged pions, which differ in their deposited energy.

Electromagnetic Calorimeter (ECAL) The ECAL has a width of 7.8 m and a height of 6.3 m, is located at 12.5 m downstream from the interaction point, and is comprised of alternating 2 mm thick lead and 4 mm thick plastic scintillator tiles. These tiles amount to a total length of 42 cm, corresponding to 25 radiation lengths, but only 1.1 hadronic interaction lengths. Consequently, electrons and photons hardly ever reach the Hadronic Calorimeter downstream of the ECAL. This fact can be used additionally to the information of the PS to distinguish between electrons/photons and hadrons.

The ECAL's energy resolution is

$$\frac{\Delta E}{E} \approx \frac{10\%}{\sqrt{E[\text{GeV}]}} \oplus 1\%,$$

with \oplus denoting quadratic summation. The first term is related to statistical fluctuations and the second the result of the readout.

Hadronic Calorimeter (HCAL) The HCAL's alternating scintillator and iron absorber tiles are arranged parallel to the beam, with the iron tiles having a lateral thickness of 1 cm, corresponding to one radiation length. The length of the HCAL in beam direction is 1.2 m and corresponds to 5.6 hadronic interaction lengths. The HCAL's energy resolution is

$$\frac{\Delta E}{E} \approx \frac{80\%}{\sqrt{E[\text{GeV}]}} \oplus 10\%,$$

Muon System

Muons are present in the final states of many decays relevant to the research areas of the LHCb collaboration, including the decay analyzed in this thesis. The Muon System provides information for the identification and reconstruction of muons, as well as for both the hardware and the software trigger. The transverse momentum (p_T) resolution of the Muon System is approximately 20%.

The Muon System consists of five stations M1-M5, with M1 positioned upstream of the calorimeter system and M2-M5 comprising the most downstream parts of the detector. M1 is important for the measurement of muon p_T before scattering can occur in the calorimeters. This information is used in the hardware trigger. M1-M3 have a high spacial resolution to precisely measure muon tracks and p_T , while M4-M5 are mainly used for muon identification by detecting traversing particles. M2-M5 are each separated by 80 cm thick iron absorbers. These absorbers and the calorimeter system (ECAL and HCAL) have a combined thickness of approximately 20 nuclear interaction lengths. Consequently, muons with a minimum momentum of 6 GeV/c are the only particles capable of traversing the whole detector to reach the last muon station. This results in muons being the particles identified with the least ambiguity. An additional iron absorber is located downstream of M5 to ensure only muons originating in LHCb's interaction point are detected.

A side view of the Muon System is shown in Figure 2.8. Each station is divided vertically into four regions R1-R4 with R1 being the inner most and R4 the outer most. The height of the regions increases by a factor of two from inner to outer regions, resulting in a ratio of 1:2:4:8. The granularity of the detector decreases from R1 to R4, as does the particle flux, resulting in an approximately constant occupancy over all regions.

All regions except the inner most of M1 are covered with multi-wire proportional chambers, which measure the trajectories of the muons by gaseous ionization. In the inner region of M1, gas electron multipliers (GEMs) are used, because in an environment with high particle flux, they were found to have better aging properties than proportional chambers.

2.2.3 Trigger System

The trigger system decides whether an event is recorded or discarded. It is comprised of a hardware-based trigger (Level-0 or L0) and a software-based one (High Level Trigger), the latter consisting of two stages (HLT1 and HLT2). An event is recorded if all trigger stages give positive decisions. The overall aim of the trigger system is to reduce the event rate from the bunch crossing rate of 40 MHz to a manageable rate of 5 kHz, while retaining as many of the interesting events as possible. Figure 2.9 shows the schematic design of the trigger setup for 2012. Details of the trigger performance in 2012 can be found in [28].

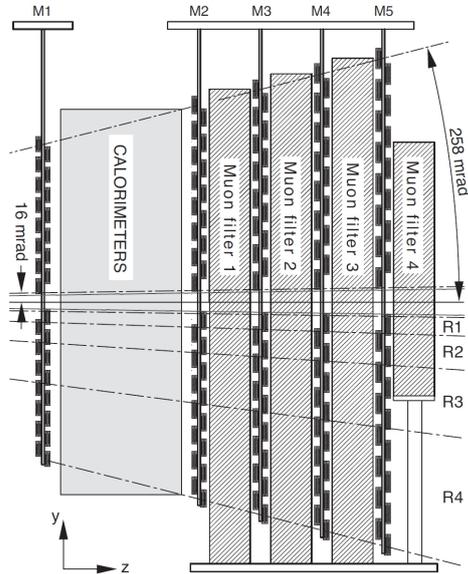


Figure 2.8: Side view of the Muon System. Image taken from [27].

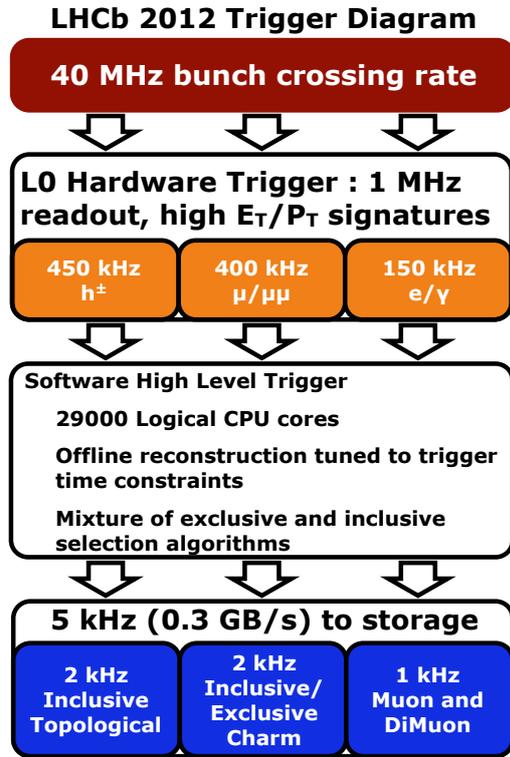


Figure 2.9: Schematic design of the trigger setup for 2012. Image taken from [19].

L0 Trigger

The L0 trigger is implemented in custom-made electronics and reduces the event rate from the bunch crossing rate of 40 MHz to 1 MHz, at which the whole detector can be read out. At the L0 stage, only information from the calorimeters, the Muon System, and the pile-up stations is used. Due to the reduced luminosity of LHCb and its bunch structure, the effective bunch crossing rate during data-taking periods of 2011 and 2012 was approximately 15 MHz. Nevertheless, the L0 runs synchronous to the nominal 40 MHz rate.

The hardware trigger attempts to reconstruct the hadron, electron and photon with the highest E_T in the calorimeters and the two muons with the highest p_T in the Muon System. Events with multiple collisions and a large number of tracks may be rejected, using information from the pile-up stations of the VELO and the calorimeters. Corresponding to the necessary information, the L0 is divided into a pile-up trigger, a calorimeter trigger and a muon trigger. These three components are connected to the Level-0 Decision Unit, which derives the final trigger decision. In 2012, the thresholds for an L0 decision were 1.76 GeV/c for single muons,

$(1.6 \text{ GeV}/c)^2$ for dimuon $p_{T_1} \times p_{T_2}$, 3.7 GeV for hadrons, and 3 GeV for electrons and photons [28].

High Level Trigger (HLT1+2)

The High Level Trigger is completely software-based, running on standard CPUs. Consequently, it is much more flexible than L0 and can be optimized with increasing knowledge of the experimental conditions.

HLT1 reduces the event rate from 1 MHz to 80 kHz. A partial event reconstruction is performed, starting by reconstructing track segments in the VELO. Segments with a high impact parameter or with matching hits in the Muon System are then extrapolated into T1-T3 by the *forward tracking* algorithm (see Section 2.2.1), which is also used during offline reconstruction. The trigger decisions are grouped into several lines, each line requiring specific criteria for a positive trigger decision. There exists a line for beauty and charm decays triggering on good quality tracks with a p_T threshold and a displacement from the primary vertex. In this analysis, lines triggering on tracks matched with hits in the Muon System are used. For details on these lines, see Section 4.2.

HLT2 performs a full event reconstruction for events selected by HLT1. It reconstructs all tracks with a minimum p_T of 300 MeV/c with the reconstruction differing only in minor aspects from the offline reconstruction to ensure the input rate of 80 kHz can be handled. After the full reconstruction, events are filtered by beauty, charm, and muon trigger lines to reduce the rate of stored events to 5 kHz. These events can then be further analyzed in the offline selection. Details on the HLT2 line used in this analysis can be found in Section 4.2.

2.2.4 Monte Carlo Simulation

Simulating pp collisions, the subsequent production and decay processes, and the interaction of the decay products with the detector is an essential part of the experiment. In this analysis, the simulation, called Monte Carlo (MC), is used for the training of a multivariate classifier (Section 4.6) and to obtain the efficiencies to trigger, reconstruct and select relevant decays (Chapter 6).

The MC is produced in three stages: the generation of the event, the tracking of particles through the detector, and the response of the detector to these particles. For the event generation, the PYTHIA program [29] is used, employing the EvtGen package [30] to simulate b hadron decays. The tracking of particles through the detector and the hadronic and electromagnetic interactions between particles and detector material are simulated by the GEANT4 package [31]. The detector model used in this process also includes passive elements, such as the beam pipes, frames and shielding material. Low-energy particles produced in secondary interactions are also simulated. The detector response is simulated by LHCb-specific software,

whose output is similar to the output obtained from real events and which can be processed the same way.

Part II

Analysis

3 Analysis Strategy

This chapter introduces the formulas relevant to calculate the branching fraction, the decay topology, the reference channel, and gives a brief overview of the analysis steps.

3.1 Formula for the Branching Fraction

The branching fraction for a particular decay mode i of a particle is defined as $\mathcal{B}(i) = \Gamma_i/\Gamma$, where Γ is the total decay width of the particle, related to its mean lifetime τ by $\tau = \hbar/\Gamma$. The partial decay width Γ_i gives the decay rate for the decay mode i and the sum over all partial decay widths is equal to the total decay width: $\Gamma = \sum_i \Gamma_i$.

The branching fraction for the signal channel as a function of measurable quantities is given by

$$\begin{aligned} \mathcal{B}(\Lambda_b^0 \rightarrow J/\psi \Lambda^0 \phi) &= \\ &= \frac{N(\Lambda_b^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) \Lambda^0(\rightarrow p \pi^-) \phi(\rightarrow K^+ K^-))}{\int \mathcal{L} dt \cdot 2\sigma_{b\bar{b}} \cdot f_{\Lambda_b^0} \cdot \epsilon_{J/\psi \Lambda^0 \phi} \cdot \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-) \cdot \mathcal{B}(\Lambda^0 \rightarrow p \pi^-) \cdot \mathcal{B}(\phi \rightarrow K^+ K^-)} \end{aligned}$$

with the following quantities: $N(\Lambda_b^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) \Lambda^0(\rightarrow p \pi^-) \phi(\rightarrow K^+ K^-))$ is the number of signal events in the data set, detected using the listed decay modes of the Λ_b^0 daughter particles; $\int \mathcal{L} dt$ is the luminosity integrated over the time during which the data was taken; $\sigma_{b\bar{b}}$ is the production cross section of $b\bar{b}$ pairs with the factor 2 taking into account that $\sigma_{b\bar{b}}$ is the cross section for a b quark *pair*; $f_{\Lambda_b^0}$ is the b quark fragmentation fraction (= probability) to form a Λ_b^0 ; $\epsilon_{J/\psi \Lambda^0 \phi}$ is the complete efficiency of detecting the decay, giving the probability of a decay to be triggered, reconstructed, selected, and finally end up in the data set which is used to determine $N(\Lambda_b^0 \rightarrow J/\psi \Lambda^0 \phi)$; finally, $\mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)$, $\mathcal{B}(\Lambda^0 \rightarrow p \pi^-)$, and $\mathcal{B}(\phi \rightarrow K^+ K^-)$ are the branching fractions of the Λ_b^0 daughter decay modes used to detect these daughters.

3.2 Decay Topology

Figure 3.1 shows the topology of the signal channel decay ($\Lambda_b^0 \rightarrow J/\psi \Lambda^0 \phi$) including the decay modes of the Λ_b^0 daughter particles used in this analysis. The channel is

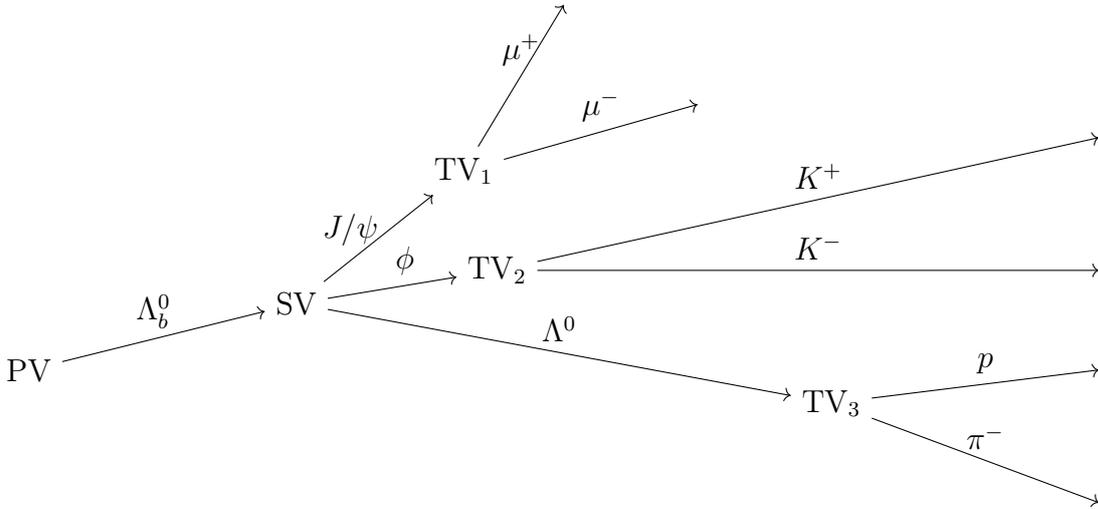


Figure 3.1: Decay topology of the signal channel with primary (PV), secondary (SV), and tertiary (TV) vertices.

called *signal* channel to distinguish it from the *reference* channel (see Section 3.3). The pp collision in which the Λ_b^0 is produced takes place at the primary vertex (PV). Due to its long lifetime of $(1.466 \pm 0.010) \times 10^{-12}$ s [8] and the boost it gets by being produced in the pp collisions, the Λ_b^0 travels a distance of the order of several mm before decaying at the secondary vertex (SV). The resolution of the Vertex Locator is fine enough to give a clear separation between the PV and the SV, making it possible to differentiate between the Λ_b^0 daughters and particles originating directly in the pp collision. The Λ_b^0 daughters J/ψ , Λ^0 , and ϕ decay at the tertiary vertices (TV) into $\mu^+\mu^-$, $p\pi^-$, and K^+K^- , respectively, resulting in a total of six charged final-state tracks, which all need to be reconstructed.

Of the three Λ_b^0 daughters, the J/ψ ($\tau_{J/\psi} = (7.1 \pm 0.2) \times 10^{-21}$ s [8]) and the ϕ ($\tau_\phi = (1.54 \pm 0.01) \times 10^{-22}$ s [8]) decay almost instantaneously, while the Λ^0 ($\tau_{\Lambda^0} = (2.63 \pm 0.02) \times 10^{-10}$ s [8]) travels a significant distance before decaying.

The decay channels of the Λ_b^0 daughters used in this analysis and their relative branching fractions (BF, quoted from [8]) are as follows: For the J/ψ , the mode $J/\psi \rightarrow \mu^+\mu^-$ with a BF of $(5.961 \pm 0.033)\%$ was chosen due to the very clean signature of the two muons, which are the particles most easily separated from other particles, since only muons are capable of reaching the last stations of the Muon System of the detector. For the Λ^0 , the decay $\Lambda^0 \rightarrow p\pi^-$ with a BF of $(63.9 \pm 0.5)\%$ was used and for the ϕ , the channel $\phi \rightarrow K^+K^-$ with a BF of $(48.9 \pm 0.5)\%$ was chosen.

Long and Downstream Tracks

Due to its long lifetime, the majority of Λ^0 baryons decay outside of the Vertex Locator or close to its edge. This leads to a significant amount of Λ^0 daughters forming *downstream tracks*, meaning that, opposed to *long tracks*, they have no tracks in the Vertex Locator, but only in the TT and T1-T3. In this analysis, Λ^0 candidates were either formed from p and π^- candidates both having *long tracks* (denoted as LL) or both having *downstream tracks* (denoted as DD). The probability of one particle having a *long track* and the other a *downstream track* is negligible.

3.3 Reference Channel

The branching fraction of the signal channel is measured relative to the branching fraction of a so-called reference channel, which was chosen to be $\Lambda_b^0 \rightarrow J/\psi \Lambda^0$, with the decay modes of the J/ψ and Λ^0 being identical to those of the signal channel. Measuring this branching fraction *ratio* has the following advantages: (1) In the formula for the relative branching fraction several quantities cancel, which could only be measured with large uncertainties. These quantities are $\int \mathcal{L} dt$, $\sigma_{b\bar{b}}$, and $f_{\Lambda_b^0}$. (2) Systematic uncertainties in the reconstruction of the J/ψ and the Λ^0 will cancel and therefore need not be taken into account explicitly.

The relative branching fraction is given by

$$\begin{aligned} \frac{\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi \Lambda^0 \phi)}{\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi \Lambda^0)} &= \\ &= \frac{N(\Lambda_b^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) \Lambda^0(\rightarrow p \pi^-) \phi(\rightarrow K^+ K^-)) \cdot \epsilon_{J/\psi \Lambda^0}}{N(\Lambda_b^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) \Lambda^0(\rightarrow p \pi^-)) \cdot \epsilon_{J/\psi \Lambda^0 \phi} \cdot \mathcal{B}(\phi \rightarrow K^+ K^-)}. \end{aligned}$$

The reference channel quantities $N(\Lambda_b^0 \rightarrow J/\psi \Lambda^0)$ and $\epsilon_{J/\psi \Lambda^0}$ are defined analogously to the corresponding quantities of the signal channel.

3.4 Steps of the Analysis

The analysis consists of the following steps:

1. In the *selection* (Chapter 4), a subset of all events recorded by LHCb is selected, which contains as many of the relevant events as possible, while rejecting as many background events as possible. This, in essence, is done by placing requirements on event variables obtained from the detector.
2. After the selection, the signal is separated from the remaining background by fitting two probability density functions (PDFs), modeling signal and background, to the data (Chapter 5). The number of signal events present in the data set, called the yield, is given by the integral of the signal PDF.

3. This yield has to be corrected for the efficiency (Chapter 6) of the detector hardware and software, which is the probability to actually reconstruct and select an event.
4. The relative branching fraction is calculated from the efficiency-corrected yields of the signal and reference channel in Chapter 7.
5. Systematic studies are presented in Chapter 8.
6. Chapter 9 gives the final result for the branching fraction ratio with statistical and systematic uncertainties.

4 Data Selection

In the first step of the analysis, the *selection*, the data collected by LHCb was filtered for events which have a high probability of containing the relevant decay chains.

4.1 Important Selection Variables

The essential part of the selection consists of placing requirements on variables obtained from the detector. The following list gives an overview of important variables used in this analysis.

- p_T (transverse momentum): Relatively heavy particles, such as the Λ_b^0 , tend to boost their decay products, resulting in these products having a high p_T . Thus, excluding particles with low p_T reduces background from the primary vertex (PV) of the pp collisions.
- $\chi_{\text{vtx}}^2, \chi_{\text{track}}^2$: Vertices and tracks are fitted from hits in the detector. From these fits, χ^2 values, indicating the fit quality, can be obtained. These values can also be given per degrees of freedom, χ^2/ndf , with ndf being the number of degrees of freedom.
- χ_{IP}^2 : The impact parameter (IP) is the distance between the PV and the particle's trajectory. The χ_{IP}^2 value of a track is the change of χ_{vtx}^2 when this track is added to the vertex. A high χ_{IP}^2 value ensures a displacement of the trajectory from the PV, giving a high probability that the particle was not produced at the PV.
- χ_{DOCA}^2 : χ^2 value of the distance of closest approach between two tracks.
- DIRA (direction angle): Cosine of the angle between the reconstructed momentum vector of a particle and the vector pointing from its vertex of origin to its decay vertex. With a perfect detector resolution, these vectors would have the same direction, giving a DIRA of 1.
- FD (flight distance): Distance between the reconstructed decay vertex of the particle and its primary vertex.
- DLS (decay length significance): Significance of the separation between the decay vertex and the primary vertex, $\text{DLS} = \text{FD}^2/\chi_{\text{FD}}^2$.

Table 4.1: Trigger lines employed in the selection, both for signal and reference channel. Only *trigger on signal* (TOS) decisions were used.

L0	HLT1	HLT2
L0Muon L0DiMuon	Hlt1DiMuonHighMass Hlt1TrackMuon Hlt1TrackAllL0	Hlt2DiMuonDetachedJPsi

- $DLL_{x\pi}$ (Delta Log Likelihood): For each stable particle ($\pi^{0,+,-}$, $K^{+,-}$, $p^{(-)}$, $e^{+,-}$, γ , $\mu^{+,-}$), information from the particle identification system is used to calculate a likelihood for it being a certain type of particle, which amounts to placing a mass hypothesis on it. The variable $DLL_{x\pi}$ gives the likelihood of a particle being of type x , relative to it being a pion (which is the most common particle in the detector), with $DLL_{x\pi} = \log \mathcal{L}(x - \pi) = \log \mathcal{L}(x) / \log \mathcal{L}(\pi)$.
- Track ghost probability: Ghost tracks consist of hits from multiple particles, falsely combined to form a track of a single *ghost* particle.

4.2 Trigger Strategy

The trigger system decides whether an event is recorded or discarded. Thus, it is the first step in the selection process. For a general description of the LHCb trigger system, see Section 2.2.3. The trigger system consists of three stages (L0, HLT1, and HLT2) and at each stage, the event has to fulfill certain criteria to result in a positive trigger decision. These criteria are organized into so-called *trigger lines*, with one line triggering, for instance, on muon pairs and another line triggering on B mesons. An event is recorded if it passes one of the existing trigger lines at all three trigger stages.

The trigger lines used in this analysis are listed in Table 4.1, a detailed description follows in the next paragraph. The lines were chosen as *trigger on signal* (TOS), meaning the particles in the signal decay chain are sufficient for a positive trigger decision. In this case, only events triggered by the muon pair associated with the Λ_b^0 candidate¹ were chosen. In contrast, TIS (trigger independent of signal) means that the decision is triggered by a particle not belonging to the signal decay chain. Since the muon pair is present in both the signal and the reference channel, identical trigger lines could be used for both channels. This has the advantage of canceling systematic uncertainties related to the trigger.

The L0Muon (L0DiMuon) line requires one track (two tracks) reconstructed in the Muon System with a certain minimum value of p_T . HLT1 matches reconstructed

¹ Since the identity of reconstructed particles can never be determined with absolute certainty, they are often called *candidates*.

VELO tracks with tracks in the Muon System. If a match is found, Hlt1TrackMuon places requirements on $\chi_{\text{track}}^2/\text{ndf}$, p_T and χ_{IP}^2 . Hlt1DiMuonHighMass requires a minimum value of the invariant mass of the muon pair. Hlt1TrackAllL0 takes events that have been accepted by any L0 line and requires at least one track with certain track quality criteria, a displacement from all reconstructed primary vertices, and a minimum p_T . HLT2 performs a full track reconstruction. The most important cuts for Hlt2DiMuonDetachedJPsi are on the decay length significance of the dimuon vertex (DLS) and the dimuon mass ($|m(J/\psi) - m(\mu^+\mu^-)| < 120 \text{ MeV}/c^2$). For details on muon trigger lines, see [32].

4.3 Cut-Based Selection

Events passing the trigger next undergo a cut-based selection in two steps. The first one, called the *stripping*, is a central selection process starting with all data recorded at LHCb. This process is organized into several stripping lines, each of which has selection requirements hard enough to significantly decrease the size of the data set, but loose enough to be useful for a variety of analyses. In a second step, this relatively loose selection is tightened according to the needs of the individual analysis.

4.3.1 Stripping

The stripping line used in this analysis is called DiMuonJpsi2MuMuDetached and its requirements are listed in Table 4.2. *Detached* here means the J/ψ should not come directly from the primary vertex. The line takes two oppositely charged muons and combines them to form a J/ψ candidate. The Λ^0 and ϕ candidates need to be added after the stripping. This is possible since the stripping line saves the complete event information, not only the information required to form the J/ψ candidate.

The cut on the p_T of the muons reduces background from the primary vertex. Additionally, requirements on the muon track quality and the probability of the candidate being a muon were made. The invariant mass of the muon pair was forced to be within $\pm 100 \text{ MeV}/c^2$ of the J/ψ . A certain quality was required of the separation of the decay vertex of the J/ψ from the primary vertex ($\text{DLS}(J/\psi) > 3$) and of the vertex formed by the muons and the J/ψ ($\chi_{\text{vtx}}^2/\text{ndf}(J/\psi) < 20$).

4.3.2 Preselection

Up to this point, the selections for the signal and reference channel were completely identical. In the cut-based selection particular to this analysis, the *preselection*, a Λ^0 candidate was formed in both channels by combining a proton and a π^- . In the signal channel, a ϕ was added by combining two oppositely charged kaons. Finally,

Table 4.2: Requirements of the stripping line DiMuon.Jpsi2MuMuDetached, used for both signal and reference channel.

Particle	Requirement
J/ψ	DLS > 3 $m \in (2996.916, 3196.916) \text{ MeV}/c^2$ $\chi_{\text{vtx}}^2/\text{ndf} < 20$
μ^+, μ^-	$p_T > 500 \text{ MeV}/c$ $\chi_{\text{track}}^2/\text{ndf} < 5$ DLL $_{\mu\pi} > 0$

a Λ_b^0 candidate was formed from the J/ψ , the Λ^0 , and, in case of the signal channel, the ϕ .

LL and DD samples Due to the Λ^0 being neutral and long-lived, its daughters can either form *long* or *downstream tracks* (Section 3.2). Starting with the preselection, the data sets of the signal and reference channel were each split into two separate samples: one consisting of Λ_b^0 candidates with *long track* Λ^0 daughters, the *LL sample*, and one with *downstream track* Λ^0 daughters, the *DD sample*. This was done for the following reasons: (1) The distributions of many variables are different for the LL and DD sample, leading to different selection criteria, separately trained boosted decision trees (BDTs) (Section 4.6.1), and different fit parameters (Chapter 5). (2) The efficiency to reconstruct a decay is vastly different for events of the LL and the DD sample. Consequently, efficiencies were calculated separately for each sample and yields were corrected separately (for details, see Section 6.5 and the footnote on page 68).

The preselection cuts are listed in Table 4.3. For a general explanation of the variables used, see Section 4.1. Here only a few noticeable cuts are mentioned: $\chi_{\text{DOCA}}^2(\Lambda^0)$ ensures the two daughters (p, π^-) come from the same mother. Putting a lower limit on $\chi_{\text{IP}}^2(p)$, $\chi_{\text{IP}}^2(\pi^-)$ and $\chi_{\text{IP}}^2(K^{+/-})$ further reduces background from the primary vertex in addition to the p_T cuts. Contrarily, the upper limit on the $\chi_{\text{IP}}^2(\Lambda_b^0)$ makes sure the Λ_b^0 originated in the primary vertex, while the flight distance cut guarantees it travels at least 1 mm, which is typical for b hadrons.

Fiducial cuts on $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$ were introduced for the following reason: To get valid efficiencies from MC, MC and data must agree in relevant variables; $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$ are not well represented in MC for the channels used in this analysis. In case of the signal channel, these discrepancies were fixed by reweighting the MC with weights taken from data. These weights can only be obtained in a $(p_T(\Lambda_b^0), \eta(\Lambda_b^0))$ region in which the detector is reliable; the ranges of the fiducial cuts have been chosen accordingly. For details on the reweighting process, see Section 6.3.1.

Table 4.3: Cut-based preselection for the signal channel. The same selection was applied to the reference channel, except for the missing ϕ and kaons.

Particle	Requirement
J/ψ	$m \in (3040, 3160) \text{ MeV}/c^2$
Λ^0 LL (DD)	$m \in (1112, 1120) \text{ MeV}/c^2$ ($m \in (1108, 1124) \text{ MeV}/c^2$) $\chi_{\text{DOCA}}^2 < 30$ (25) $\chi_{\text{vtx}}^2 < 30$ (25)
p, π^- LL (DD)	$p > 2 \text{ GeV}/c$ $\chi_{\text{IP}}^2 > 9$ (> 4) $p_T > 250 \text{ MeV}/c$
ϕ	$m \in (1005, 1035) \text{ MeV}/c^2$ $p_T > 250 \text{ MeV}/c$ $\chi_{\text{vtx}}^2/\text{ndf} < 12$
K^+, K^-	$p_T > 250 \text{ MeV}/c$ $\chi_{\text{IP}}^2/\text{ndf} > 10$ $\chi_{\text{track}}^2/\text{ndf} < 3$
Λ_b^0	$m \in (5000, 7000) \text{ MeV}/c^2$ $\chi_{\text{vtx}}^2/\text{ndf} < 16$ $\chi_{\text{IP}}^2 < 25$ DIRA > 0.999 FD $> 1 \text{ mm}$
Λ_b^0 (fiducial cuts)	$p_T \in (0, 20) \text{ GeV}/c$ $\eta \in (2.2, 4.5)$
K^+, K^-, p, π^-	Track ghost prob LL (DD) < 0.4 (< 1.0)
K_S veto	$m_{\text{inv}}((p \rightarrow \pi^-)\pi^-) \notin (487.6, 507.6) \text{ MeV}/c^2$ only in LL sample of the reference channel, see Section 4.4.2

A final remark on the TOS trigger lines used in this analysis: To evaluate whether the trigger decision was TIS or TOS, the full decay chain up to the Λ_b^0 has to be reconstructed. This directly follows from the definition of a TOS decision (Section 4.2). The Λ_b^0 candidate was formed in the preselection; thus, the selection of events with a TOS decision needs to take place after the preselection. The stripping line used in this analysis requires no specific trigger lines, but takes all recorded events, that is, all events with positive trigger decisions, as input. Consequently, the choice of specific trigger lines needs to take place after the stripping. In this analysis, both the choice of trigger lines and TOS decisions took place after the preselection. It was already mentioned that only events triggered by the muon pair associated with the Λ_b^0 candidate were selected. This association took place in the preselection, by associating the Λ_b^0 candidate with the J/ψ candidate, which in turn is formed from the muon pair during the stripping. This results in the J/ψ candidate required

by the trigger line `Hlt2DiMuonDetachedJPsi` being identical to the one required by the stripping line `DiMuonJpsi2MuMuDetached`.

4.4 Background Sources

After the cut-based selection, several possible sources of background, contaminating the signal, remained. This section describes these sources and the steps taken to eliminate them.

4.4.1 Combinatorial Background

Combinatorial background consists of candidates formed by random combinations of particles not sharing a common ancestor, accidentally passing the selection. It is almost always present and usually follows an exponential distribution in the invariant mass spectrum of the candidate. Thus, it is easy to separate from the signal. In this analysis, a multivariate classifier was applied to further reduce combinatorial background remaining after the cut-based selection (Section 4.6).

4.4.2 Background from Misidentified Particles

Candidates being identified as the wrong type of particle can be a source of *peaking background*, which in turn is background forming a peak close to or overlapping with the signal peak in the invariant mass spectrum of the Λ_b^0 candidate or its daughters. This overlap can make distinguishing signal from background difficult, so a careful investigation and minimization of peaking backgrounds is crucial.

Misidentified particles can lead to peaking backgrounds via the following mechanism: The identification of a final state particle by the detector results in a mass hypothesis on this particle; using this hypothesis and the measured three-momentum, the four-momentum is obtained. The invariant mass of a mother particle is calculated from the four-momenta of its daughters. A misidentification of a particle thus leads to a false mass hypothesis, which in turn changes the invariant mass of the mother of the misidentified particle. This way, mother particles of type x could end up in the mass window of particle y and form a peak in its invariant mass spectrum.

One example of a possible background from misidentified particles (mis-ID background) in this analysis is the misidentification of the π^+ in $K_S^0 \rightarrow \pi^+\pi^-$ as a proton, leading to a higher invariant mass of the K_S , so it could mimic a $\Lambda^0 \rightarrow p\pi^-$ decay. The similar lifetimes of the particles ($\tau(K_S) = (8.954 \pm 0.004) \times 10^{-11}$ s, $\tau_{\Lambda^0} = (2.63 \pm 0.02) \times 10^{-10}$ s [8]) make it difficult to separate the two by placing requirements on the flight distance, for instance.

One way to check if mis-ID backgrounds exist in a data set is to undo a hypothetical misidentification by replacing the mass hypothesis of a final state particle

with a different one and then recalculate the invariant mass of the mother with this new hypothesis. Misidentified particles should end up as peaks in the resulting mass spectrum at their true mass.

In case of the K_S , the mass of the proton in $\Lambda^0 \rightarrow p\pi^-$ was changed to that of a pion. In the $m_{inv}((p \rightarrow \pi^+)\pi^-)$ spectrum, where $p \rightarrow \pi^+$ denotes the changed mass hypothesis, a peak should then be visible at $m(K_S) = 497.6 \text{ MeV}/c^2$. The hypothesis of this peak being a K_S is strengthened if it is enhanced by applying a cut like $\text{PProbNNPI} > 0.9$. The variable PProbNNPI here is the probability of the proton candidate being a pion, calculated by a neural network taking various detector variables as input. A small K_S contribution was found in the LL sample of the reference channel. For plots showing the effect of the PProbNNPI cut, see Figure 4.1.

Check for Mis-ID Backgrounds in Λ^0 Candidates

A check for mis-ID backgrounds was performed as described above in $m_{inv}(p\pi^-)$ both in the signal and reference channel, for LL and DD samples separately. Pions, kaons and protons were considered as sources for misidentification. Muons were neglected due to their small probability of being misidentified as a hadron. This lead to the following combinations of possible misidentifications: (1) π^+ or K^+ being misidentified as p , π^- correctly identified; (2) K^- or p^- being misidentified as π^- , p correctly identified; or (3) combinations of π^+ or K^+ being misidentified as p , and K^- or p^- being misidentified as π^- . All these combinations were checked and the only background that was found is the above mentioned K_S contribution in the LL sample of the reference channel.

Check for Mis-ID Backgrounds in ϕ Candidates

Analogously, $m_{inv}(K^+K^-)$ in the signal channel was checked, using the combined LL and DD samples, since they only differ in the nature of the Λ^0 candidate, not in the ϕ candidate. Again, all possible combinations of misidentifications between pions, kaons, and protons were checked and no peaks were found. The spectra $m_{inv}(K^{+/-}p)$ and $m_{inv}(K^{+/-}\pi^-)$ were not checked, since they do not correspond to a mother particle of this analysis.

Mis-ID Background from $\Lambda_b^0 \rightarrow \psi(2S)\Lambda^0$

The decay $\Lambda_b^0 \rightarrow \psi(2S)\Lambda^0$ with $\psi(2S) \rightarrow J/\psi\pi^+\pi^-$ and $J/\psi \rightarrow \mu^+\mu^-$ could lead to a peak in $m_{inv}(\mu^+\mu^-K^+K^-)$ in the signal channel if both pions were misidentified as kaons. To study this potential background, $m_{inv}(\mu^+\mu^-(K^+ \rightarrow \pi^+)(K^- \rightarrow \pi^-))$ was plotted after the cut-based selection. The change of mass hypotheses is denoted by $K^{+/-} \rightarrow \pi^{+/-}$. The result for the DD sample is shown in Figure 4.2a. Notice that this plot was created with the loose cut on the Λ_b^0 mass listed in Table 4.3,

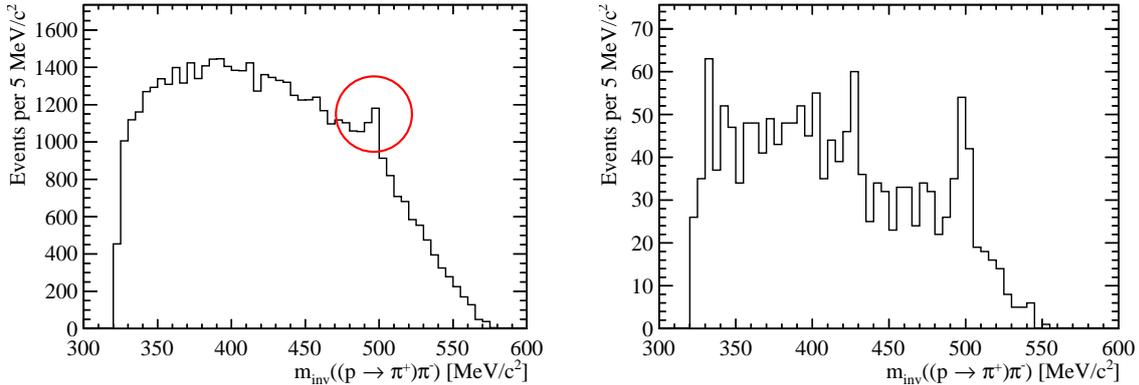
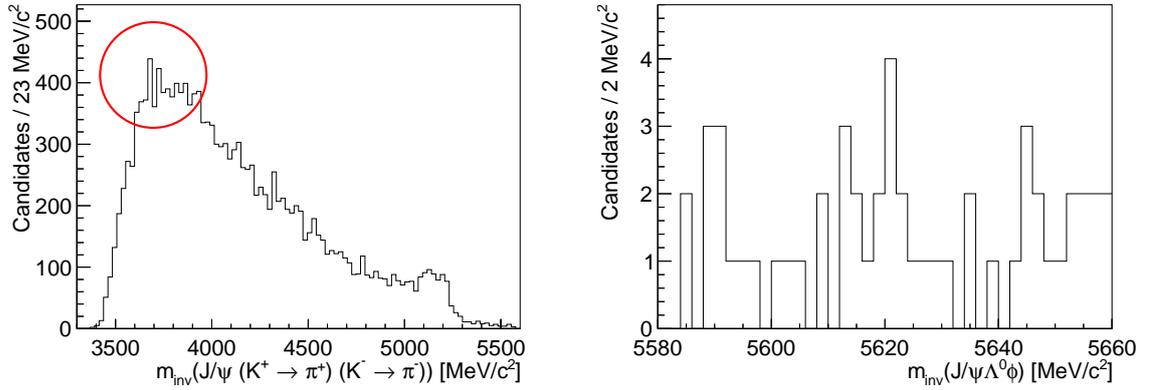


Figure 4.1: Invariant mass spectrum of $p\pi^-$ from the reference channel LL sample with the mass hypothesis of the p replaced by that of a π^+ , after the cut-based selection without K_S veto. Left: Without PPROBNNPI cut, K_S peak circled in red. Right: With PPROBNNPI > 0.9 , PPROBNNPI is the probability of the p candidate being a π^+ . The K_S peak at $497.6 \text{ MeV}/c^2$ is enhanced by the PPROBNNPI cut.

because a tighter cut would result in insufficient statistics to judge whether a peak near the mass of the $\psi(2S)$ is present. A small peak is seen at the region around $m(\psi(2S)) = (3686.097 \pm 0.025) \text{ MeV}/c^2$ [8].

To investigate whether events forming this peak are close to the Λ_b^0 peak in the mass spectrum of the Λ_b^0 daughters, this spectrum is shown in Figure 4.2b for events with $m_{inv}(\mu^+\mu^-(K^+ \rightarrow \pi^+)(K^- \rightarrow \pi^-)) \in (m(\psi(2S)) - 20 \text{ MeV}/c^2, m(\psi(2S)) + 20 \text{ MeV}/c^2)$, again for the DD sample after the cut-based selection. A small peak can be seen in the region around $m(\Lambda_b^0) = (5619.51 \pm 0.23) \text{ MeV}/c^2$ [8]. However, the $\psi(2S)$ peak in Figure 4.2a only makes up a very small fraction of the total events in the region around $m(\psi(2S))$.

As a next step, the $\psi(2S)$ peak in the $m_{inv}(\mu^+\mu^-(K^+ \rightarrow \pi^+)(K^- \rightarrow \pi^-))$ spectrum was removed by placing requirements on the PROBNNPI and PROBNNK variables of the kaon candidates, which represent the likelihood of a candidate being a pion or a kaon. The peak could be removed without losing a significant amount of events not contributing to the peak. The same PROBNNPI and PROBNNK requirements were then applied to the Λ_b^0 daughter spectrum, resulting in a spectrum without contributions from the misidentified $\psi(2S)$ daughters. No significant change of the distribution around the mass of the Λ_b^0 was observed compared to the spectrum without the PROBNN requirements. It was concluded that mis-ID background from $\Lambda_b^0 \rightarrow \psi(2S)\Lambda^0$ does not significantly contribute to the Λ_b^0 peak and thus can be neglected.



(a) Invariant mass spectrum $m_{inv}(\mu^+\mu^-(K^+ \rightarrow \pi^+)(K^- \rightarrow \pi^-))$ for the signal channel DD sample after the cut-based selection. The peak of the $\psi(2S)$ is circled in red.

(b) Invariant mass spectrum of the Λ_b^0 daughters for events with $m_{inv}(\mu^+\mu^-(K^+ \rightarrow \pi^+)(K^- \rightarrow \pi^-)) \in (m(\psi(2S)) - 20 \text{ MeV}/c^2, m(\psi(2S)) + 20 \text{ MeV}/c^2)$ for the signal channel DD sample after the cut-based selection.

Figure 4.2: Plots to investigate potential mis-ID background from $\Lambda_b^0 \rightarrow \psi(2S)\Lambda^0$.

Elimination of the K_S Background

Mis-ID backgrounds can be eliminated by either excluding the mass region of their peak in the invariant mass spectrum with the appropriately changed mass hypothesis – in the case of K_S , $m_{inv}((p \rightarrow \pi^+)\pi^-)$ – a so-called *veto cut*, or by placing requirements on particle identification (PID) variables such as PPROBNNPI.

The problem of using PID variables is that they are poorly modeled by the MC, leading to incorrect efficiencies. To avoid this, the PID variables in the MC have to be recalibrated using data samples of decays with a simple topology, in which the final state particles are easy to identify and from which the correct PID distributions can be obtained. This procedure is not trivial and introduces a systematic uncertainty.

The disadvantage of veto cuts in the invariant mass spectrum is that not only the peak from the misidentified particle is filtered out, but also real events lying in the region of the veto. Since the only relevant mis-ID background found in this analysis is in the reference channel, which has about 200 times as much statistics as the signal channel, losing statistics by a veto was not an issue, so a cut of $\pm 10 \text{ MeV}/c^2$ around the K_S mass was applied and no PID variables were used. Due to the exceptionally clean signal in both channels, PID variables were also not needed at any other stage of the analysis, making a recalibration of these variables unnecessary.

4.4.3 Partially Reconstructed Decays as Background

If a b baryon with a mass higher than that of the Λ_b^0 has a final state consisting of the final state particles of the signal or reference channel and one or more additional particles, it could peak close to the Λ_b^0 mass if the additional particles were not reconstructed.

In the two channels considered in this analysis, no final state particle is a direct descendant of the Λ_b^0 . Cuts on χ_{vtx}^2 of the vertex formed by the Λ_b^0 daughters and their daughters, and on m_{inv} of the Λ_b^0 daughters reduce the probability of backgrounds sharing the final state particles ($\mu^{+/-}$, p , π^- , $K^{+/-}$), but not the Λ_b^0 daughters (J/ψ , Λ^0 , and ϕ).

Since the invariant mass spectrum of the Λ_b^0 candidates is exceptionally clean both for the signal and the reference channel, no partially reconstructed decays were considered as possible backgrounds. Partially reconstructed decays of b baryons, which form a Λ_b^0 candidate when combined with random particles, were treated as combinatorial background.

4.4.4 Non-Resonant Decays as Background

Non-resonant decays of $\Lambda_b^0 \rightarrow J/\psi p \pi^-$ without forming an intermediate Λ^0 and $\Lambda_b^0 \rightarrow \Lambda^0 \mu^+ \mu^-$ without forming a J/ψ were observed at LHCb. These decays could contribute to the signal of the reference channel, showing a flat distribution in $m_{\text{inv}}(p\pi^-)$ and $m_{\text{inv}}(\mu^+\mu^-)$, respectively, opposed to the peak of resonant decays.

To check for possible non-resonant contributions in both channels, the background-subtracted mass spectrum of the Λ_b^0 daughters after application of the BDT (described in Section 4.6) is shown in Figure 4.3. The background subtraction was achieved via the sPlot-technique (see Section 5.5 and [33]). It was concluded that non-resonant, flat contributions are negligible.

4.5 Multiple Candidates

There may exist more than one Λ_b^0 candidate in an event fulfilling all the selection criteria. Since the branching fraction of the signal channel is assumed to be very small, it is unlikely that these *multiple candidates* truly are separate Λ_b^0 baryons. More probable is one Λ_b^0 , but multiple daughter or granddaughter candidates passing the selection. For instance, there could exist two Λ^0 candidates, which result in two Λ_b^0 candidates when combined with the same J/ψ and ϕ . In such a case, all but one of the multiple candidates were eliminated. To prevent any bias in this choice, one of the candidates was chosen at random. This elimination was done between the preselection and the multivariate selection.

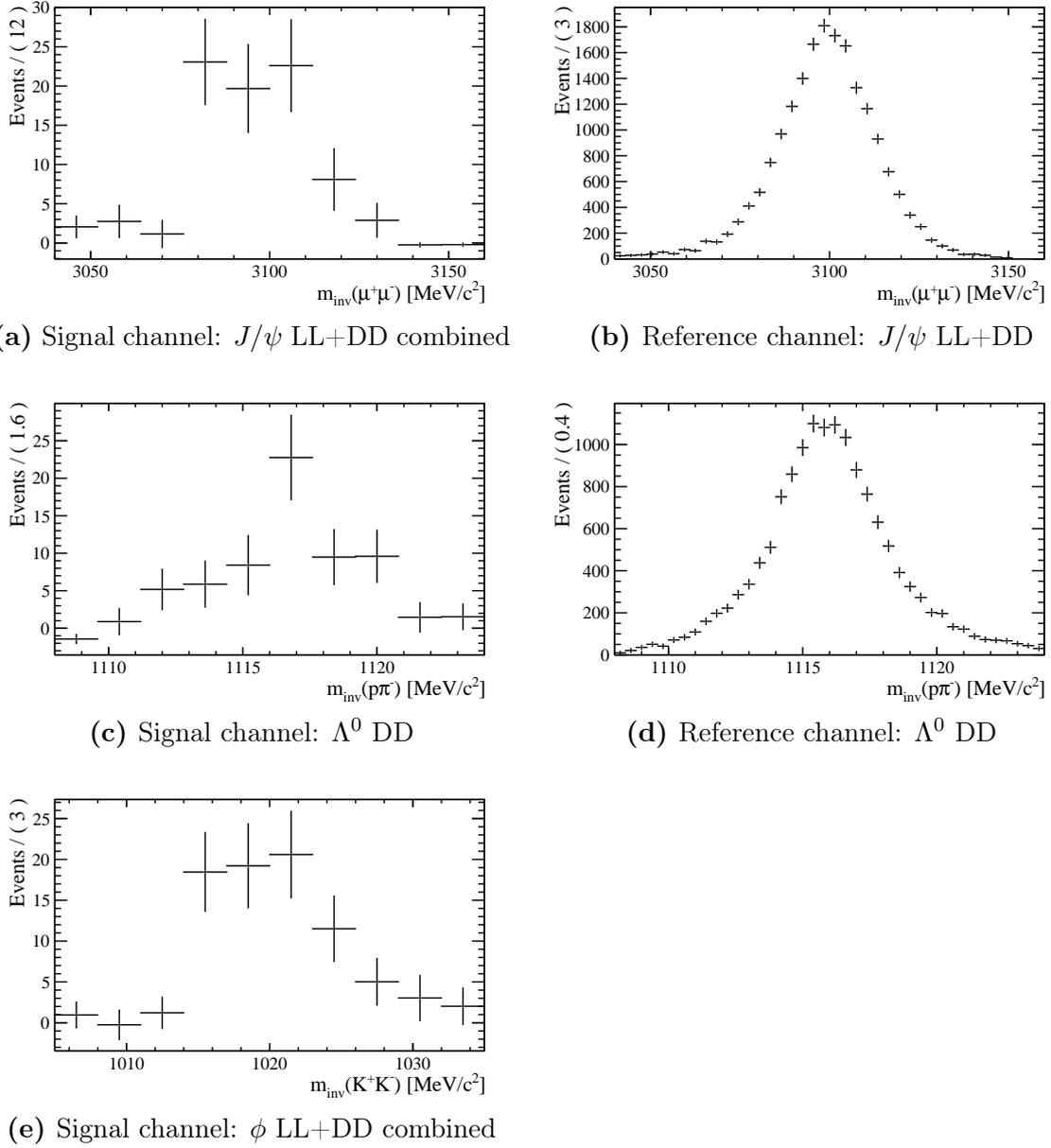


Figure 4.3: Background-subtracted mass spectrum of Λ_b^0 daughters after application of the BDT. For J/ψ and ϕ , combined LL+DD samples are shown, since the samples only differ in the nature of the Λ^0 candidate.

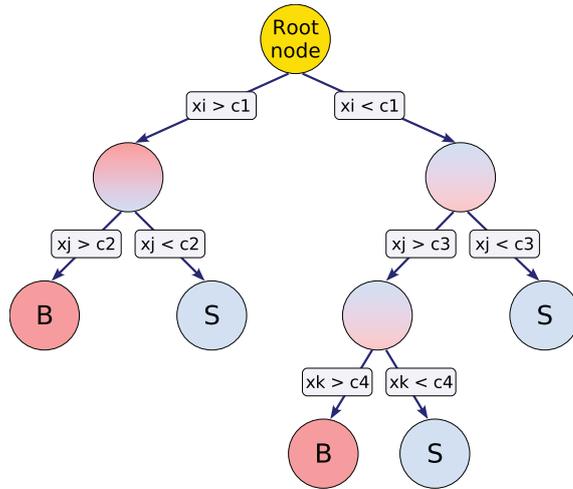


Figure 4.4: Example of a small BDT, explained in the text. Image taken from [34].

4.6 Multivariate Selection

To further reduce combinatorial background after the cut-based selection, a multivariate classifier was applied. These classifiers take several event variables as input and as output give a value which can be interpreted as a probability of the event belonging to one of several classes. In this case the inputs are detector variables like those described in Section 4.1 and the two classes are signal and background events. The advantage of a multivariate classifier is the possibility of selecting multiple hypercubes in the space spanned by the input variables, while a cut-based selection can only select one. Put differently, a multivariate classifier can exploit correlations between input variables and the classification of the event as signal or background in a more fine-grained way than a cut-based selection.

Boosted Decision Trees

The type of classifier applied in this analysis is a boosted decision tree (BDT). The following description is oriented on [34]. A decision tree consists of nodes representing binary decisions of the form $x_i > c_i$ or $x_i < c_i$, where x_i is the value of an input variable i and c_i is a constant obtained by the BDT training (see below). Depending on whether $x_i > c_i$ or $x_i < c_i$, one of two connected nodes is reached, where a new decision $x_j < / > c_j$ with a different input variable j is made. This process is repeated until the end of the tree is reached, which consists of a node representing the decision of whether the event is classified as signal or background. An example of a small decision tree is shown in Figure 4.4.

The process of determining the c values is called *training*. For the training, two samples ideally consisting solely of signal events and of background events, respec-

Table 4.4: BDT training sample sizes and sideband ranges. $*(x, y)$ means lower sideband $(m(\Lambda_b^0) - y, m(\Lambda_b^0) - x)$ and upper sideband $(m(\Lambda_b^0) + x, m(\Lambda_b^0) + y)$ were used, units in MeV/c^2 .

Sample	Sig chan LL	Sig chan DD	Ref chan LL	Ref chan DD
Signal sample	769	3298	2325	9992
Background sample	373	3778	3048	9656
Sidebands used*	(30, 1230)	(30, 630)	(80, 240)	(80, 130)

tively, are needed. At the first node of the tree, the decision-making variable and its value are chosen such that the decision gives the best separation between signal and background for the combined signal+background sample. The sample is split accordingly into two subsamples and the process of choosing a discriminating variable and its value is repeated for each subsample. The subsamples are again split and new nodes are created, until a subsample reaches a desired minimum number of events, after which all events in this subsample are classified as either signal or background.

During training, a problematic effect can arise when two variables have comparable separation powers. One of the two variables might be chosen due to statistical fluctuations in the sample, possibly leading to a suboptimal performance of the tree. This problem is overcome by a process called *boosting*: A large amount of trees is trained with varying weights for the events of the training samples. After training, an event will be classified as signal by some trees of the BDT and as background by others. The output of the BDT is the (possibly weighted) average of the decisions of the individual trees, usually with signal being +1 and background being -1, so the output of the BDT has values in $(-1, 1)$. The process of choosing an optimal cut-value on this output is described in Section 4.6.2.

4.6.1 Training of Boosted Decision Trees

For the training and application of the BDT, the TMVA framework [34] was used. Separate BDTs were trained on LL and DD samples due to different distributions of variables related to the Λ_b^0 and its daughters. Signal samples were taken from MC and background samples from both upper and lower sidebands of the data. The distance between the sidebands and the mean of the Λ_b^0 peak was chosen to be approximately 4σ to 5σ of the Λ_b^0 peak to make sure no significant amount of signal is polluting the background sample (for fit values, see Chapter 5). The lower limit of the lower sideband and the upper limit of the upper sideband were then chosen such that the number of events in the background sample is of the same order as that of the signal sample. The number of events used for training and the values of the sideband intervals are shown in Table 4.4.

Table 4.5: BDT input variables, used both for LL and DD samples, same for signal and reference channel except ϕ variables.

$\chi_{\text{FD}}^2(J/\psi), \min(\log(p_T(\mu^+)), \log(p_T(\mu^-))), \max(\log(p_T(\mu^+)), \log(p_T(\mu^-)))$ $\chi_{\text{FD}}^2(\Lambda^0), \log(p_T(\Lambda^0))$ $\chi_{\text{IP}}^2(\phi), \log(p_T(\phi))$ $\log(p_T(\Lambda_b^0)), \chi_{\tau}^2(\Lambda_b^0)$
--

BDT Input Variables

The variables used as BDT inputs are listed in Table 4.5. Distributions of transverse momenta commonly have a long tail with few events; to obtain a more evenly distributed input, the logarithm of p_T was used. A larger set of variables has been investigated and the ones listed were chosen for their separation power between signal and background in the signal channel. Additionally, the linear correlation matrices of these variables were checked for signal and background samples. In case of highly correlated variables, the one with the highest separation power was chosen. Comparisons between the input variable distributions of signal and background training samples are shown in Appendix A.2 for the LL and DD samples of the signal channel.

Finally, in the reference channel, the agreement between the distributions on data and MC was checked. This agreement is important for a good performance of the BDT, but crucial for getting reliable BDT efficiencies, which are estimated on MC. Comparisons of the BDT input variables between background-subtracted data and reweighted MC are shown in Appendix A for the reference channel. In the signal channel, statistics are too low for a meaningful comparison. For details on reweighting and BDT efficiencies, see Chapter 6.

Since the error on the branching fraction is dominated by the statistical error in the signal channel, a suboptimal performance of the BDT in the reference channel is unproblematic. Thus, the input variables from the signal channel were also used for the reference channel, except for variables related to the ϕ .

BDT Performance

The output of the BDT for LL and DD samples of the signal channel is shown in Figure 4.5. A clear separation between signal and background can be seen. In addition, these plots are useful to check for *overtraining*. A BDT is overtrained when it shows sensitivity for statistical fluctuations in the training sample, often due to the minimum number of events at a node being set too low. To check for overtraining, the signal and background samples were split in half and only one half, the *training sample*, was used to train the BDT. After training, the BDT was applied to both the training sample and the other half, the *test sample*, separately.

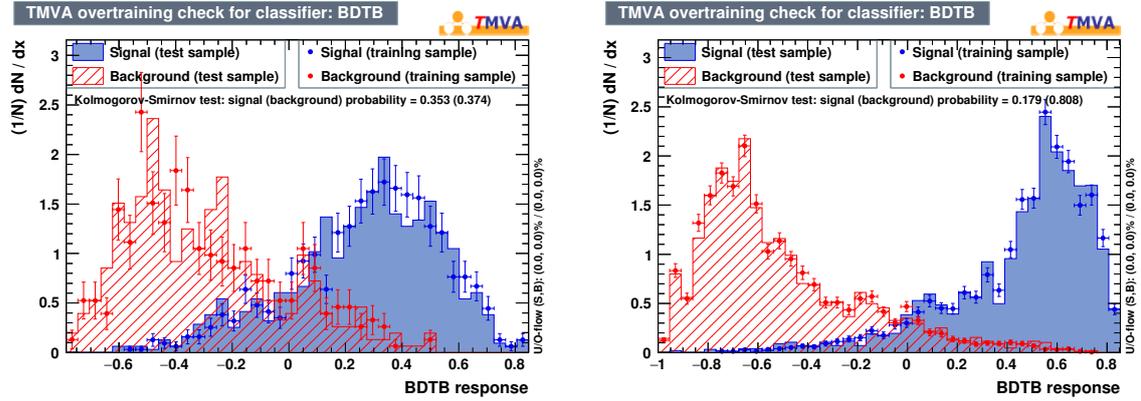


Figure 4.5: BDT performance and overtraining check for signal channel LL sample (left) and DD sample (right).

If no or little overtraining occurred, the BDT responses to the two samples should be in good agreement, which here is the case.

4.6.2 Working-Point Optimization

To apply the BDT on data, the cut value of the BDT response, the so-called *working-point*, has to be chosen. Figure 4.6 shows the background rejection versus the signal efficiency of the BDTs trained on the LL and DD samples of the signal channel. Such a plot is called a ROC curve (receiver operating characteristic curve) and each point of the curve corresponds to a specific working-point value. Background rejection is defined as the fraction of background the BDT filters out, while signal efficiency is defined as the fraction of signal the BDT retains. A perfect BDT would retain 100% of signal events, while rejecting 100% of background events. In reality, a higher signal efficiency corresponds to a lower background rejection, as can be seen on the ROC curves. An optimal compromise between these two quantities should be found. This compromise can be achieved by maximizing a figure of merit (FoM), defined as $\text{FoM} = N_{\text{sig}} / \sqrt{N_{\text{sig}} + N_{\text{bg}}}$, where N_{sig} is the number of signal events and N_{bg} is the number of background events in the relevant mass range after application of the BDT. The denominator, $\sqrt{N_{\text{sig}} + N_{\text{bg}}}$, can be interpreted as a Poissonian error, so the FoM is the number of signal events over the error of the total events in a mass range.

In principle, the maximum value of the FoM could be obtained by evaluating N_{sig} and N_{bg} for a range of working-points directly on data, but this process is sensitive to statistical fluctuations in the data and could bias the number of signal events. To prevent this, the maximum value of the FoM was estimated by the following procedure:

First, a working-point w_0 far from the optimum was chosen and a fit to the

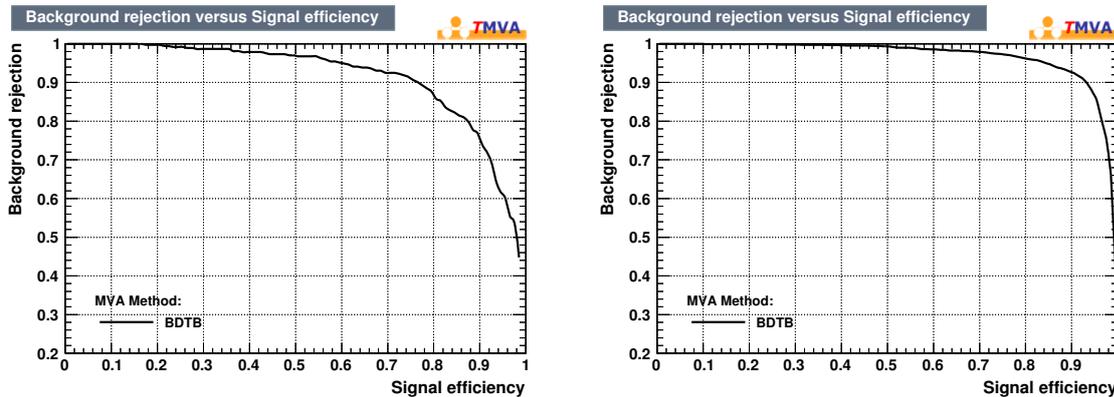


Figure 4.6: Background rejection versus signal efficiency for the BDTs trained on LL samples (left) and DD samples (right) of the signal channel.

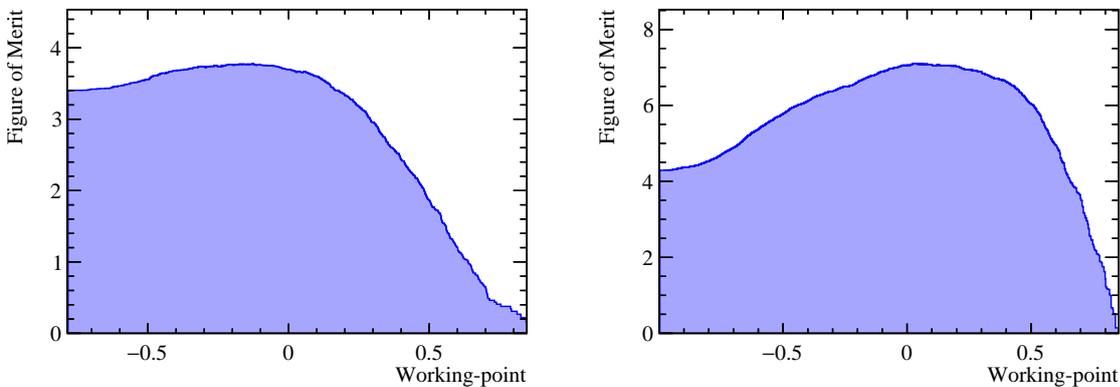


Figure 4.7: Figure of merit as a function of the working-point for the signal channel LL sample (left) and DD sample (right).

data was performed to obtain N_{sig} and N_{bg} at w_0 . The mass range in which these values were obtained is $\pm 2\sigma$ around the mean of the Λ_b^0 peak. The fit method and probability density functions for signal and background were the same as described in Chapter 5.

For the second step, the efficiency of the BDT was calculated on the MC test sample mentioned above as a function of the working point for signal ($\epsilon(w)_{\text{sig}}$) and background ($\epsilon(w)_{\text{bg}}$). The variable $\epsilon(w)$ is defined as $N(w)/N_{\text{tot}}$ with $N(w)$ being the number of events after application of the BDT at the working-point w and N_{tot} being the total number of events before the application of the BDT. Using these efficiencies, N_{sig} , N_{bg} and the FoM were extrapolated from w_0 to other working-points. The FoM as a function of the working-point is shown in Figure 4.7. From these plots, the estimated maximum FoM value was taken. Since the FoM was estimated using efficiencies calculated on MC, it is not sensitive to statistical fluctuations in the data.

5 Signal Extraction

After application of the BDTs, there still remained some combinatorial background in the data. To separate this background from the signal and extract the total number of signal events in each channel, fits were performed on the data.

The first step of the fitting procedure is a suitable choice of probability density functions (PDFs) following the shape of the signal and background. A simple choice would be a Gaussian for signal and an exponential for background. These PDFs have free parameters (mean and width in case of the Gaussian), whose values are inferred by fitting the combined signal and background PDF to the data.

5.1 Fit Method

The fitting method employed here is called *maximum likelihood estimation* [35]. Essentially, this method consists of maximizing a function which represents the likelihood of the PDF to describe the data. This function can be derived as follows: Assume the data is described by the PDF $f(x|\theta)$ with x being the measured quantity (in this case, the invariant mass of the Λ_b^0) and θ being a set of parameters. The probability of a measured value x_i lying in $(x_i, x_i + dx_i)$ then is $f(x_i|\theta)dx_i$ and the probability of a set of N measured values lying in the corresponding intervals is proportional to $\prod_{i=1}^N f(x_i|\theta)dx_i$. Since we are interested in θ and dx_i does not depend on it, the latter can be dropped and the *likelihood function* is defined as

$$L(\theta) = \prod_{i=1}^N f(x_i|\theta).$$

If the PDF describes the data well, $\prod_{i=1}^N f(x_i|\theta)$ should be high and at the optimal value of θ , $L(\theta)$ should have a maximum. Since the numerical maximization of a sum is easier than that of a product, usually the logarithm of L ,

$$\log L(\theta) = \sum_{i=1}^N \log f(x_i|\theta),$$

is maximized. This results in equations of the form $\partial \log f / \partial \theta = 0$, which can be evaluated relatively easily.

In this analysis, several parameters were constrained to values taken from MC. Constraining a parameter opposed to fixing it to an exact value leaves the parameter

floating around that value c_v , but introduces a penalty in L for large deviations from c_v . This penalty is realized by multiplying L by a Gaussian with mean c_v and width c_e , where c_v is the fit value obtained from MC and c_e is the error on it. Thus L is multiplied by a factor which has its maximum at $\theta_i = c_v$ for the corresponding i -th constrained parameter.

5.2 Decay Tree Fitter

The fits were performed on the invariant mass spectrum of the Λ_b^0 candidates. This invariant mass is obtained via the four-momenta of the daughters, which in turn are calculated from the four-momenta of the granddaughters ($\mu^{+/-}$, p , π^- , $K^{+/-}$). Traditionally, these calculations are done using a bottom-up approach consisting of two steps: First, the four-momenta of the daughters are calculated separately for each daughter, and second, the Λ_b^0 four-momentum is calculated.

In this analysis, a *decay tree fitter* (DTF) was used, which evaluates all four-momenta at once, while taking into account kinematic constraints. This evaluation is done by performing a fit on the complete decay chain, parameterized in terms of vertex positions and four-momenta. The fit is constrained by all available relevant information, such as the measured parameters of the final-state particles and four-momentum conservation at each vertex. Taking into account this information all at once instead of using a bottom-up approach improves the accuracy of the invariant mass spectrum of the Λ_b^0 candidates. The DTF uses a Kalman fitter, details on which can be found in [36].

5.3 Signal Channel Fits

For all fits, the RooFit toolkit was used [37]. The low statistics in the signal channel result in a high ambiguity regarding the choice of a suitable signal PDF. Looking only at the data, a simple Gaussian would suffice to achieve a satisfactory fit quality. However, more information about the signal shape can be obtained from a fit to the MC samples, since they are much larger than the available data samples. It was found that a double Gaussian best describes the MC and consequently was chosen as the signal PDF for the data, as well. The shape of the double Gaussian fitted to the data was constrained to the shape obtained in a fit to the MC. As a systematic study, the fits were performed with an alternative set of PDFs in Section 8.4.

A double Gaussian consists of the sum of two Gaussian functions G_1 and G_2 with shared mean μ , but different widths σ_1 and σ_2 . The PDF then has the following form:

$$PDF_{MC}(x) = N_{\text{sig}} \{ f \cdot G_1(x|\mu, \sigma_1) + G_2(x|\mu, \sigma_2) \},$$

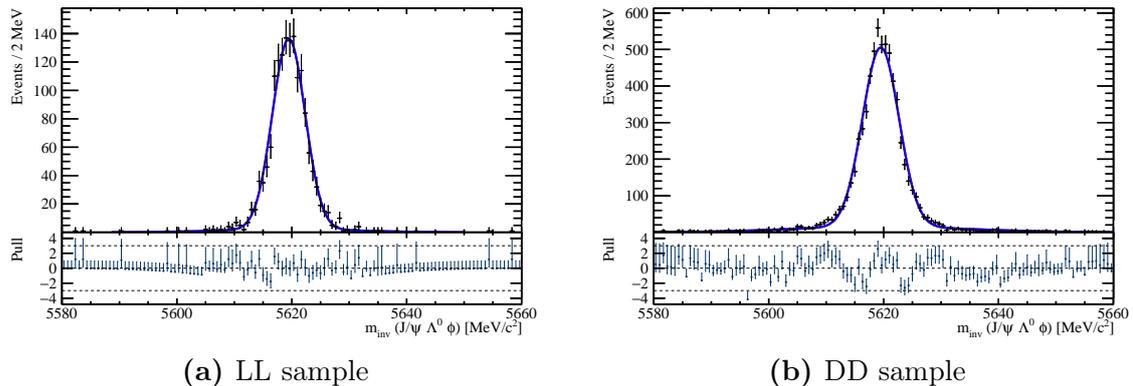


Figure 5.1: Fits to the mass spectrum of the Λ_b^0 candidates in the signal channel on MC.

where f is a scaling factor for the first Gaussian relative to the second and N_{sig} is the yield of signal events.

The fits on MC were performed after the cut-based selection and before the application of the BDT, since the constraints are needed for the estimation of the figure of merit (Section 4.6.2). The fits were done separately for the LL and DD sample, because the different Λ^0 resolutions lead to different Λ_b^0 resolutions. The resulting plots are shown in Figure 5.1 and the values of the fit parameters are listed in Table 5.1.

For the fit to the data, the shape of the signal PDF was constrained to that of the MC. Specifically, σ_1 , σ_2 , and f were constrained, μ and the yield were left floating freely. For the combinatorial background, an exponential function was used, leading to the complete PDF

$$PDF_{\text{data}}(x) = N_{\text{sig}} \{ f \cdot G_1(x|\mu, \sigma_1) + G_2(x|\mu, \sigma_2) \} + N_{\text{bg}} \cdot \exp(c_{\text{bg}} \cdot x).$$

Plots of the fits are shown in Figure 5.2 and fit parameters are listed in Table 5.1. The value of c_{bg} is compatible with 0, which is equivalent to a flat background. As a cross-check, a fit was also performed on the combined LL+DD sample. The shape for this fit was constrained to values obtained from a fit to the combined LL+DD MC sample. The yield obtained from this fit, $N_{\text{sig LL+DD}}$, is in excellent agreement with the sum of the individual fits, $N_{\text{sig LL}} + N_{\text{sig DD}}$ (Table 5.1).

5.4 Reference Channel Fits

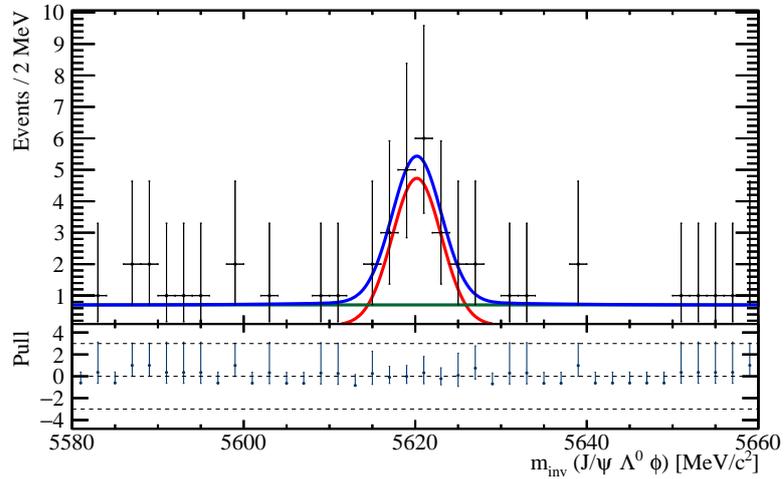
The statistics in the reference channel were sufficiently high to float all fit parameters. Again, a double Gaussian was used for signal and an exponential for background. Plots of the fits are shown in Figure 5.3 and the fit values are listed in Table 5.2.

Table 5.1: Fit results for the signal channel.

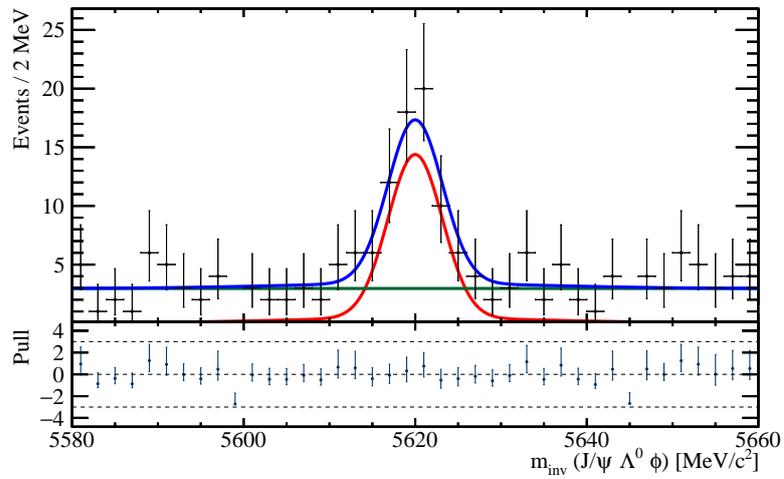
Parameter	LL sample		DD sample	
Signal channel: MC				
μ [MeV/c ²]	5619.47	± 0.08	5619.49	± 0.05
σ_1 [MeV/c ²]	12.08	± 0.98	15.27	± 0.48
σ_2 [MeV/c ²]	2.86	± 0.04	3.18	± 0.04
f	0.06	± 0.01	0.13	± 0.01
Signal channel: data				
μ [MeV/c ²]	5620.19	± 0.91	5620.00	± 0.53
σ_1 [MeV/c ²]	12.08	± 0.98	15.27	± 0.48
σ_2 [MeV/c ²]	2.86	± 0.06	3.18	± 0.04
f	0.06	± 0.01	0.13	± 0.01
$\sigma := \frac{f \cdot \sigma_1 + \sigma_2}{f+1}$	3.42	\pm	4.55	\pm
N_{sig} [Events]	17.84	± 4.96	63.78	± 9.87
$N_{\text{sig LL+DD}}$ [Events]	81.79 ± 11.15			
N_{bg} [Events]	28.17	± 5.91	118.23	± 2.29
c_{bg}	5.3085×10^{-9}	± 0.0133	2.90×10^{-8}	± 0.02

Table 5.2: Fit results for the reference channel.

Parameter	LL sample		DD sample	
Reference channel: data				
μ [MeV/c ²]	5621.21	± 0.14	5621.33	± 0.10
σ_1 [MeV/c ²]	5.77	± 0.48	6.31	± 0.44
σ_2 [MeV/c ²]	12.13	± 1.67	11.95	± 0.96
f	0.62	± 0.11	0.51	± 0.10
$\sigma := \frac{f \cdot \sigma_1 + \sigma_2}{f+1}$	9.70	\pm	10.05	\pm
N_{sig} [Events]	4485.45	± 85.20	14341.57	± 173.68
$N_{\text{sig LL+DD}}$ [Events]	18811.89 ± 192.57			
N_{bg} [Events]	3814.49	± 81.15	25693.69	± 203.68
c_{bg}	-0.00100	± 0.00011	-0.00100	$\pm 1.61091 \times 10^{-5}$



(a) LL sample



(b) DD sample

Figure 5.2: Fits to the mass spectrum of the Λ_b^0 candidates in the signal channel on data. Signal in red, background in green. This is the first time this decay has been observed.

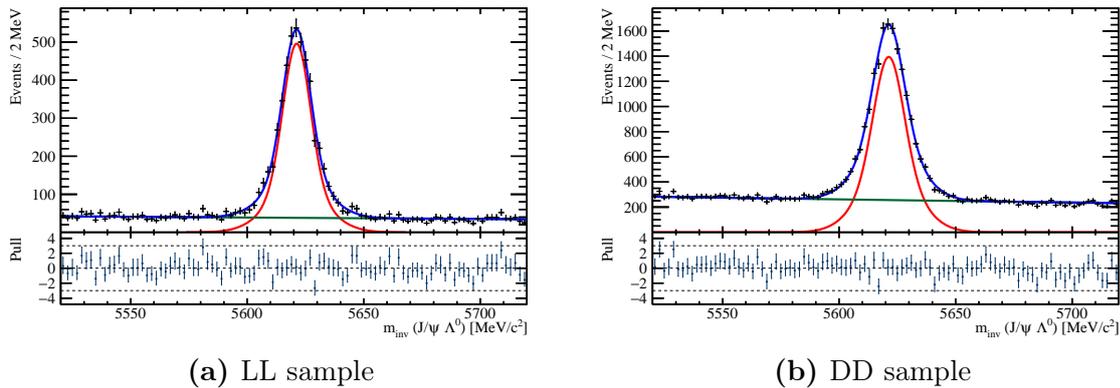


Figure 5.3: Fits to the mass spectrum of the Λ_b^0 candidates in the reference channel on data. Signal in red, background in green.

5.5 Sweights

Sweights [33] are per-event-weights, which can be obtained after performing a maximum likelihood fit on a data set, and which are defined for each event category; in this case, signal and background. Events with higher sweight values have a higher probability of belonging to the category the sweight was defined for. However, sweights are not a probability in the strictest sense, since they can be negative. The sweight of an event i defined in terms of the category *signal* is denoted as $sw_{\text{sig}}(i)$. Sweights have the following property: $\sum_i sw_{\text{sig}}(i) = N_{\text{sig}}$; the sum over the sweights of all events in the data set gives the yield of the category for which they were defined. From this point on, the sig-index will be dropped, since in this analysis, only the signal category is of interest. These signal sweights were used to calculate efficiency-corrected yields event-by-event (Section 7.1).

Sweights are a form of statistical background subtraction; they can be used to extract the distribution of variables for signal events which otherwise might be difficult to obtain, because the distributions of the signal events are mixed with the distributions of background events. For this procedure to give valid results, the variable over which the sweights are calculated (here, the invariant mass of the Λ_b^0 candidate) must be uncorrelated with the variable for which the distribution is to be extracted (e.g. $p_T(\Lambda_b^0)$). For a detailed description of the so-called sPlot-technique, with which the sweights are obtained, see [33].

6 Efficiencies

The signal yields obtained in the previous chapter are far smaller than the actual amount of decays which took place during data taking. This loss of events is taken into account by introducing efficiencies by which the yields are corrected. There exist several reasons for the loss of events: The trigger system does not give a positive decision for every relevant decay, not every trajectory is reconstructed successfully nor is every particle identified correctly. In the offline analysis, signal events are lost during the selection, which always is a trade-off between retaining signal and rejecting background.

6.1 Correcting Discrepancies Between Data and MC

In this analysis, all efficiencies were determined from simulation. The MC production (Section 2.2.4) results in two samples: the full set of generated events and the reconstructed events.

To get valid efficiencies, the simulation has to agree with real data in the aspects relevant for efficiency calculations. Unfortunately, there are discrepancies between data and MC. The main reason for these discrepancies is insufficient knowledge of the QCD processes taking place during pp collisions, leading to the inaccurate representation of distributions of variables in the MC. Two such variables, which are highly correlated with the efficiency to reconstruct the decay, are p_T and η of the Λ_b^0 .

There are two methods available to correct for effects coming from these discrepancies: The MC distributions can be reweighted to fit the data (for details, see Section 6.3.1), or the efficiencies can be calculated as a function of the inaccurately represented variables. The latter is usually done by evaluating the efficiencies in a histogram binned in these variables. For the two methods to be valid, it has to be assumed that the efficiencies are simulated correctly for specific values of the inaccurately represented variables, and that only the distribution of these variables is incorrect.

In this analysis, the efficiencies of the signal channel were binned in variables corresponding to the Dalitz plot of the decay; the efficiencies of the reference channel were binned in $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$. Additionally, the signal channel MC was reweighted in $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$. The following sections explain the reasons for these choices.

6.2 Efficiencies as a Function of the Dalitz Plot

Dalitz plots are a way to analyze three-body decays, first proposed and applied in [38]. A classic Dalitz plot is a 2D-histogram binned in two variables describing the decay kinematics (sometimes, similar plots with higher dimensions are also called Dalitz plots). In case of all three decay products and the mother particle having spin 0, two variables are sufficient to fully describe the decay kinematics. Notice that this is not the case for the signal decay in this analysis, where both the Λ_b^0 and Λ^0 have spin 1/2, and both the J/ψ and ϕ have spin 1. Nonetheless, the Dalitz plot still contains relevant information about the decay kinematics.

The two axes of the Dalitz plot are usually chosen to be the invariant masses squared of two pairs of the three decay products. Three such pairs are available ($m_{inv}^2(12), m_{inv}^2(23), m_{inv}^2(13)$) and two of these can be chosen freely.

If all three decay products and their mother have spin 0 and there are no resonances present, the events are flatly distributed on the Dalitz plot. A resonance in $m_{inv}(J/\psi\Lambda^0)$, for instance, would appear in the Dalitz plot as a band structure with higher event density perpendicular to the $m_{inv}^2(J/\psi\Lambda^0)$ axis at the $m_{inv}^2(J/\psi\Lambda^0)$ coordinate corresponding to the resonances' mass. Interference effects between multiple resonances can lead to more complex structures.

Since the decay products in the signal channel do not have spin 0 and the ultimate goal of the analysis is to search for a resonance corresponding to a pentaquark in $m_{inv}(J/\psi\Lambda^0)$, the distribution of events on the Dalitz plot is not expected to be flat. The simulation used to obtain the efficiencies, however, was created with a flat phase space distribution and no polarization effects or resonances taken into account. Thus, the signal channel efficiencies were binned in $m_{inv}^2(J/\psi\Lambda^0)$ and $m_{inv}^2(\Lambda^0\phi)$. Binning was chosen over reweighting, since for the latter to give valid results, the real distributions of the Dalitz variables would have to be known with sufficient precision, which here is not the case due to low statistics.

6.3 Dealing with Discrepancies in p_T and η of the Λ_b^0

The distributions of $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$ are not well represented in the simulation due to insufficient knowledge about the QCD processes taking place during pp collisions. In the reference channel, these discrepancies were taken into account by binning the efficiencies in these two variables. In the signal channel, there is insufficient MC statistics for a binning in these variables additional to the binning on the Dalitz plot, which, when combined, would lead to a four-dimensional binning. Instead, the simulation was reweighted.

6.3.1 Reweighting the Signal Channel MC

The purpose of the reweighting is to correct the MC distributions so that they more closely resemble the data distributions. This correction is achieved by applying correction factors, called *weights* in this context, to the simulation. These weights depend on the reweighting variables and can be obtained as follows:

Suppose the MC is to be reweighted in variable x (generalizing the following steps to multiple dimensions is straightforward): First, two histograms binned in x are filled with the number of MC and background-subtracted data events, respectively. These histograms need to have an identical binning scheme. Second, the weights are calculated per bin and are defined as the normalized number of background-subtracted data events in this bin divided by the normalized number of MC events in this bin:

$$w(x_i) = \frac{N_{\text{data}}(x_i)/N_{\text{data tot}}}{N_{\text{MC}}(x_i)/N_{\text{MC tot}}},$$

where $N(x_i)$ is the number of events in the i -th bin and $N_{\text{data/MC tot}}$ is the total number of data or MC events, respectively. Reweighting the simulation then consists of assigning each simulated event the weight found in the bin corresponding to the event's x value. In case of efficiencies, both the samples in the numerator and the denominator have to be reweighted. Notice that the MC from which the weights are obtained and the one to which they are applied need not necessarily be the same.

Reweighting $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$ of the signal channel with weights taken from the same channel is not sensible, since the number of signal events in data is too low to reliably infer the needed distributions. Instead, the signal channel MC was reweighted with weights taken from the reference channel. The weights were calculated using an equidistant 10x10 binning with histogram boundaries being equal to the fiducial cuts applied ($p_T(\Lambda_b^0) \in (0, 20)$ GeV/c and $\eta(\Lambda_b^0) \in (2.2, 4.5)$).

Weights were obtained separately for LL and DD samples after the cut-based selection and before the BDT to avoid possible distortions of the distributions by the BDT. The background in the data samples was already sufficiently low at this stage to perform reliable fits. In bins containing less than 0.5% of the sample, the weights were recalculated as the mean of nearest neighbor bins with more than 0.5% of the sample. Diagonal neighbors were included in the mean with a weight of $1/\sqrt{2}$. This procedure was repeated until all bins, including those with no initial nearest neighbors, were filled. Figure 6.1 shows the resulting weights. Distributions of the signal channel MC after BDTs, to which the weights were applied, are shown in Figure 6.2.

Ideally, the p_T and η distributions of the Λ_b^0 for signal and reference channel would be very similar for MC on the one and for background-subtracted data on the other hand, because it then would be obvious that the weights from the reference channel also improve the MC distributions of the signal channel. Figure 6.3 shows comparisons of $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$ between the signal and the reference channel for

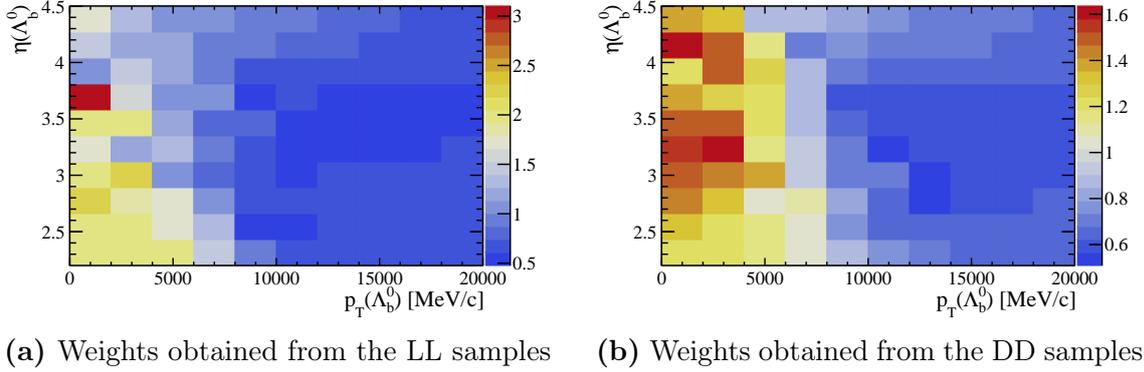


Figure 6.1: Weights obtained from the comparison of sweighted data and MC of the reference channel.

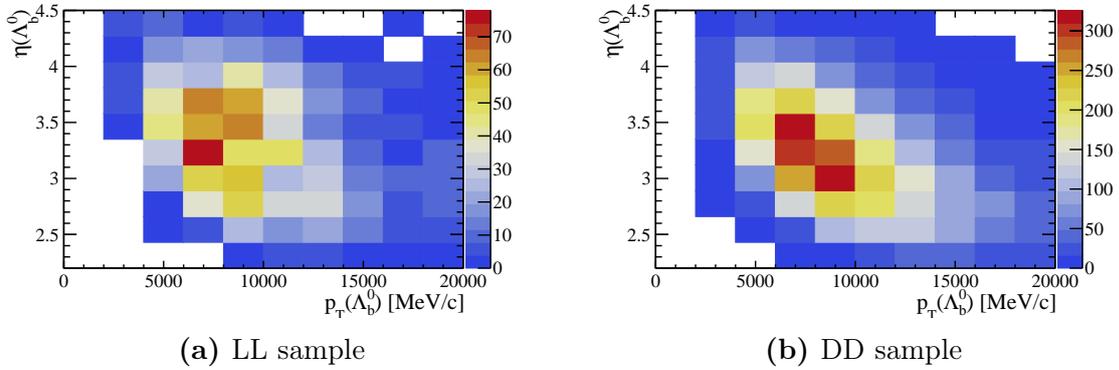


Figure 6.2: (p_T, η) distributions of the Λ_b^0 for the signal channel MC after application of the BDTs.

data and MC. Due to low statistics in the signal channel, for the comparison of data, the LL and DD samples have been combined. The distributions of the signal and the reference channel clearly are very different, with the difference in $p_T(\Lambda_b^0)$ being noticeably larger between signal and reference MC than between reference MC and reference data.

Despite these differences, reweighting the signal channel simulation with weights taken from the reference channel has been attempted. The weights were not taken directly from the histograms shown in Figure 6.1, but calculated for each point in the (p_T, η) plane by bilinear interpolation based on the four nearest bin centers. Figure 6.4 and Figure 6.5 show the results for LL and DD samples, respectively, with background-subtracted data being compared to unweighted and reweighted MC. Again, due to the low statistics in the signal channel, the combined LL and DD samples are also shown (Figure 6.6). These combined samples were reweighted with weights taken from the combined LL+DD sample of the reference channel. To quantify the compatibility between the compared distributions, Kolmogorov-

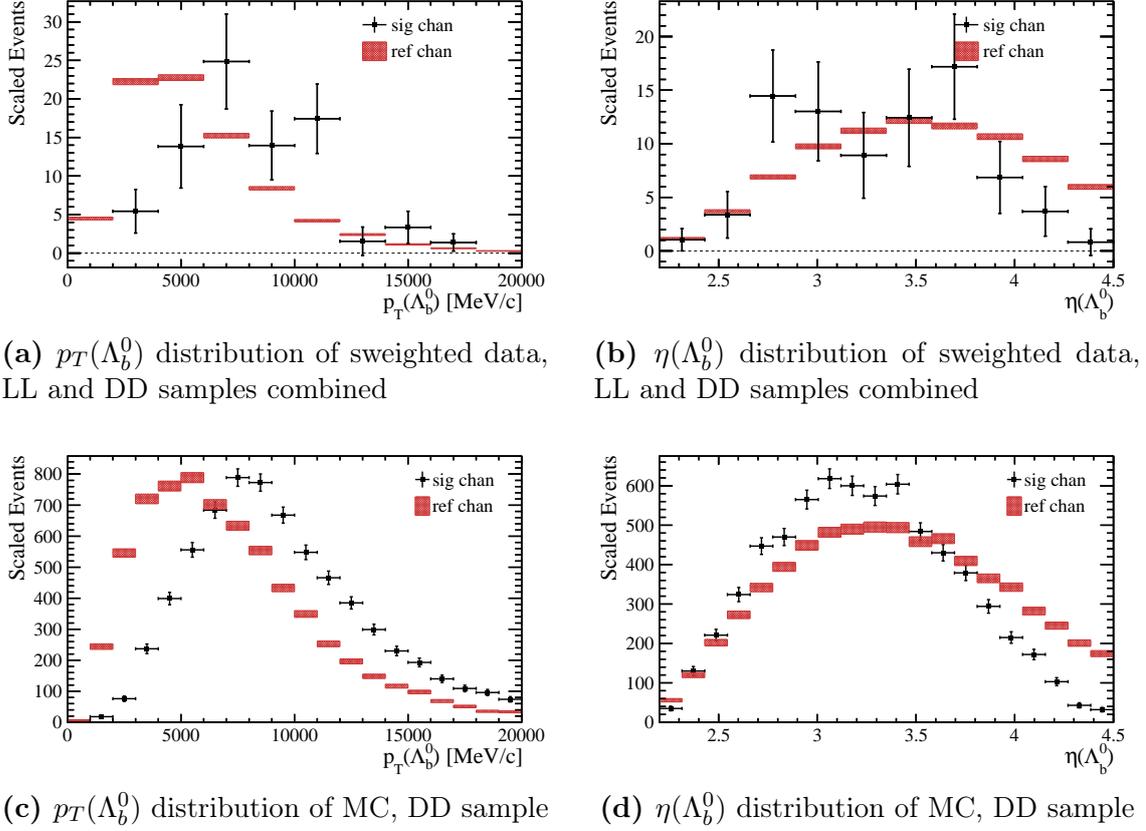


Figure 6.3: $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$ compared between signal and reference channel on data and MC.

Smirnov tests (KS tests) were performed and their results are shown on the plots. Higher KS test values imply better compatibility between the two shown distributions. The reweighting lead to a significant improvement in the KS test values both for $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$ in the DD sample and the combined LL+DD sample. The value only worsened noticeably for $p_T(\Lambda_b^0)$ in the LL sample, where the KS test is not reliable due to the very low statistics and the resulting large fluctuations in the data.

The improvements introduced by the reweighting imply that the relative differences between data and MC per (p_T, η) bin are similar between signal and reference channel. This can be explained by the origin of the discrepancies between data and MC mainly being the inaccurate simulation of the Λ_b^0 production, because then the deviation of the simulation from real data in an individual (p_T, η) bin is independent of the decay chain of the Λ_b^0 , which is simulated after the Λ_b^0 production. The noticeable improvements in the compatibility of data and MC in the signal channel were taken as sufficient cause to reweight the MC.

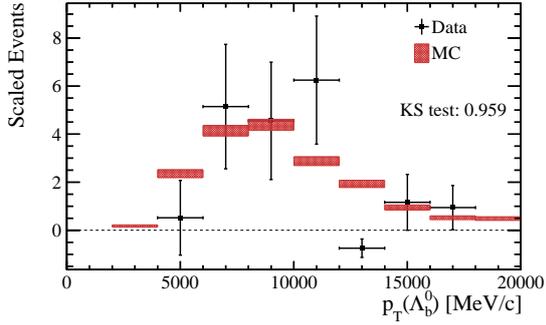
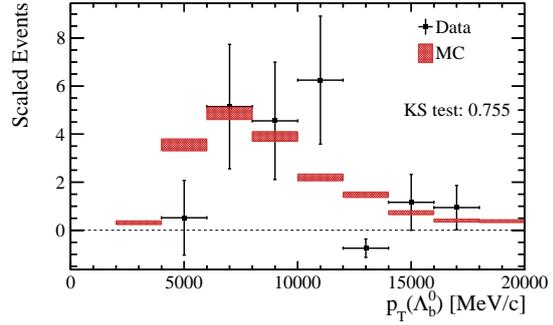
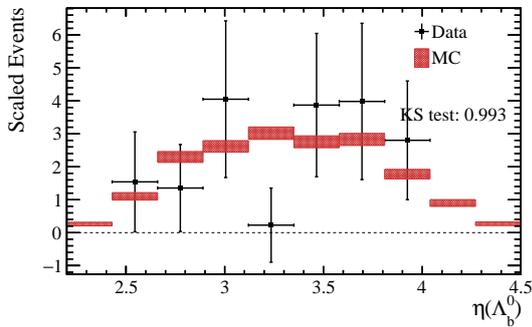
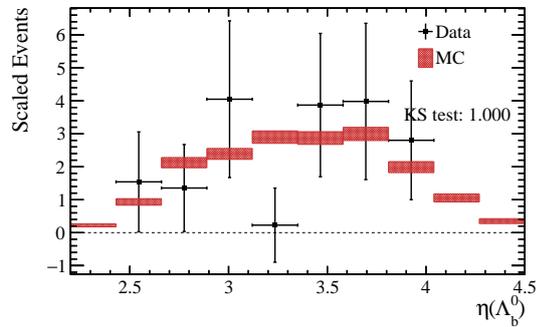
(a) Signal channel $p_T(\Lambda_b^0)$ for LL, unweighted(b) Signal channel $p_T(\Lambda_b^0)$ for LL, reweighted(c) Signal channel $\eta(\Lambda_b^0)$ for LL, unweighted(d) Signal channel $\eta(\Lambda_b^0)$ for LL, reweighted

Figure 6.4: $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$ for the signal channel LL samples, comparisons between background-subtracted data and unweighted/reweighted MC. KS test values quantify the compatibility between the two distributions shown in a plot. The effect of the reweighting can be seen by comparing KS test values for two plots in the same row.

6.4 Acceptance Cut Efficiencies

The detector simulation is the most time-consuming part of the MC production process. To save time, the detector is only simulated if all final-state particles of the decay lie in a certain acceptance region. The efficiency of this *acceptance cut* is defined as $\epsilon_{\text{Acc}} = N_{\text{Acc}}/N_{\text{Gen}}$ with N_{Acc} being the number of events after the acceptance cut and N_{Gen} the number of generated events before the acceptance cut, but with the fiducial cuts in $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$ (Section 4.3.2) applied.

To evaluate the acceptance cut efficiency, two MC samples were produced with the detector simulation turned off; one with and one without the acceptance cut applied. The former gives N_{Acc} and the latter N_{Gen} . Notice that at this stage there is no distinction between *long* and *downstream tracks*, since this distinction requires

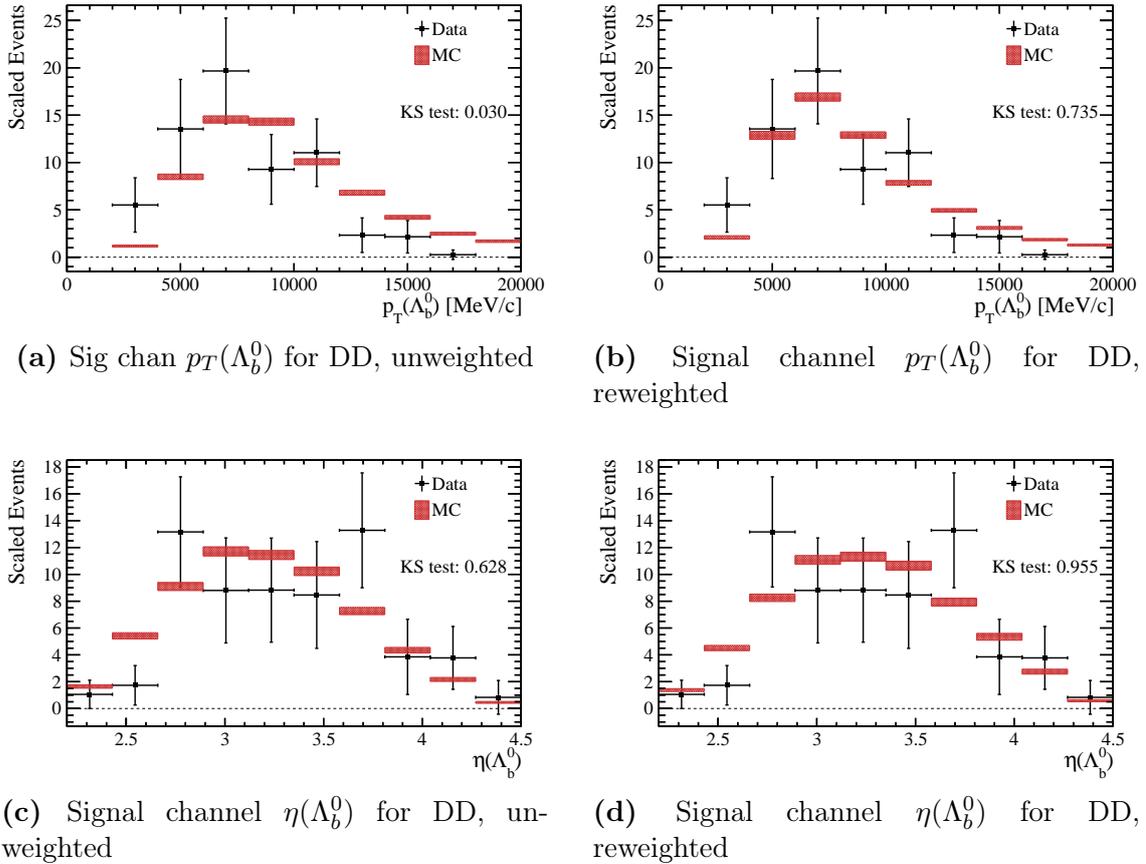


Figure 6.5: $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$ for the signal channel DD samples, comparisons between background-subtracted data and unweighted/reweighted MC. KS test values quantify the compatibility between the two distributions shown in a plot. The effect of the reweighting can be seen by comparing KS test values for two plots in the same row.

the simulation of the detector, and consequently the samples are not split into LL and DD (see the footnote on page 68 for details).

The acceptance cut efficiencies were calculated without binning for the following reason: Binned efficiencies are introduced to account for discrepancies between data and MC. Here, the relevant discrepancies are the ones in p_T and η of the Λ_b^0 , since these variables are highly correlated with the acceptance cut efficiency, because both are defined in terms of the decay geometry relative to the detector geometry. To estimate the effect of the discrepancy on the acceptance cut efficiency, one would need to know the real $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$ distributions outside the acceptance region, which obviously is not possible. Thus, a binning in these variables does not improve the accuracy of the acceptance cut efficiency in a relevant way and can be omitted.

A binning on the Dalitz plot is unnecessary, since the acceptance cut efficiency

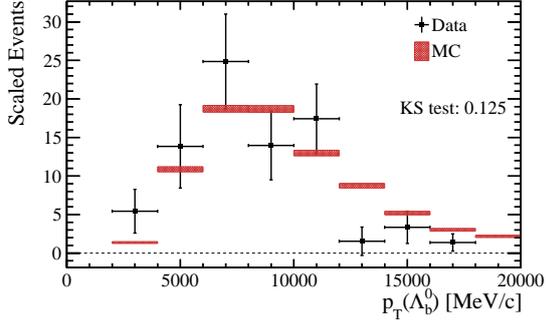
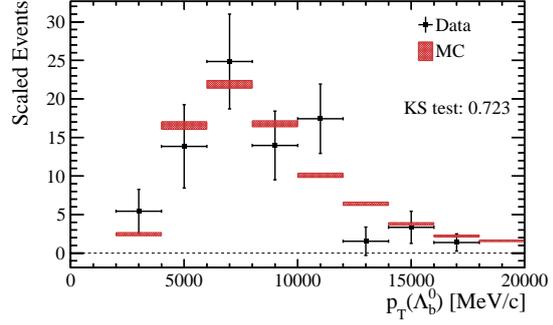
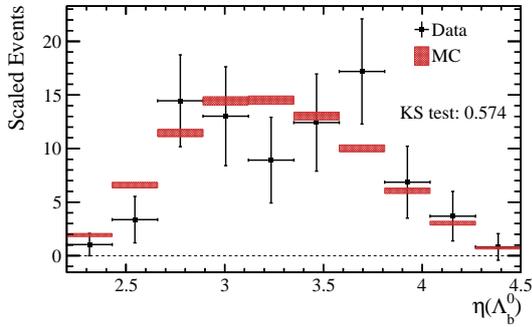
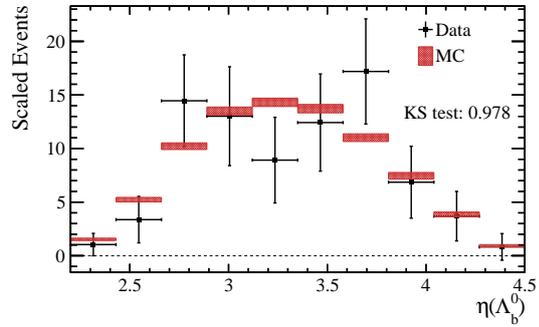
(a) Signal channel $p_T(\Lambda_b^0)$ for LL+DD, unweighted(b) Signal channel $p_T(\Lambda_b^0)$ for LL+DD, reweighted(c) Signal channel $\eta(\Lambda_b^0)$ for LL+DD, unweighted(d) Signal channel $\eta(\Lambda_b^0)$ for LL+DD, reweighted

Figure 6.6: $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$ for the signal channel LL and DD samples combined, comparisons between background-subtracted data and unweighted/reweighted MC. KS test values quantify the compatibility between the two distributions shown in a plot. The effect of the reweighting can be seen by comparing KS test values for two plots in the same row.

does not depend on the position of the event on the Dalitz plot. This can be seen as follows: ϵ_{Acc} depends on the decay geometry relative to the detector geometry, specifically, the angle of the final-state particles ($\mu^{+/-}$, p , π^- , $K^{+/-}$) with respect to the beam axis. Contrarily, the position of the event on the Dalitz plot only depends on the internal decay kinematics of the Λ_b^0 daughters (J/ψ , Λ^0 , ϕ) and is independent of the decay's orientation relative to the detector. Thus, ϵ_{Acc} and the event's position on the Dalitz plot are uncorrelated and ϵ_{Acc} is expected to be flat across the Dalitz plot. Values of the unbinned acceptance cut efficiencies are listed in Table 6.1.

6.5 Reconstruction, Trigger, and Selection Efficiencies

After the acceptance cut efficiency, the remaining efficiencies were calculated in a single step, defining $\epsilon_{\text{RTS}}^{\text{TT}} = N_{\text{Sel}}^{\text{TT}}/N_{\text{Acc}}$ with N_{Acc} being the number of events after the acceptance cut and $N_{\text{Sel}}^{\text{TT}}$ the number of events with a specific track type (denoted as TT, it can be LL or DD) after the complete selection, including the BDT. *RTS* here stands for *reconstruction, trigger and selection*. Defined this way, ϵ_{RTS} gives the efficiency of an event being detected as either LL or DD, not the efficiency of it being detected at all. This definition, in the end, results in two separate efficiency-corrected yields (one for LL and one for DD), both giving the total number of decays (N_{tot}) which took place during data taking.¹ For the formula of the efficiency-corrected yield, see Section 7.1.

The efficiency ϵ_{RTS} of the signal channel was binned in $m_{\text{inv}}(J/\psi\Lambda^0)$ and $m_{\text{inv}}(\Lambda^0\phi)$ after reweighting the MC in p_T and η of the Λ_b^0 . Weights were calculated and applied separately for LL and DD samples. ϵ_{RTS} of the reference channel was binned in p_T and η of the Λ_b^0 .

6.5.1 Uncertainties on Efficiencies and Adaptive Binning

The error on the efficiency was calculated per bin and is derived from the error on N_{Sel} and N_{Acc} per bin; these are $\sqrt{N_{\text{Sel}}}$ and $\sqrt{N_{\text{Acc}}}$, respectively. The error is then given by

$$\text{err}(\epsilon_{\text{RTS}}) = \left(\frac{N_{\text{Sel}}^2}{N_{\text{Acc}}^3} + \frac{N_{\text{Sel}}}{N_{\text{Acc}}^2} \right)^{1/2}$$

and the relative error is given by

$$\frac{\text{err}(\epsilon_{\text{RTS}})}{\epsilon_{\text{RTS}}} = \left(\frac{1}{N_{\text{Acc}}} + \frac{1}{N_{\text{Sel}}} \right)^{1/2}.$$

Since $N_{\text{Sel}} \ll N_{\text{Acc}}$, the relative error is dominated by N_{Sel} and decreases with increasing N_{Sel} per bin. The latter is equivalent to a coarser binning, which decreases the accuracy of the correction of inaccurately represented variables, for which the binning was introduced in the first place. To optimize the trade-off between a decreasing relative error and a decreasing correction accuracy, while making optimal

¹ At this point, the question may arise why efficiencies were not defined as $\epsilon_{\text{RTS}}^{\text{TT}} = N_{\text{Sel}}^{\text{TT}}/N_{\text{Acc}}^{\text{TT}}$ (with $N_{\text{Acc}}^{\text{TT}}$ instead of N_{Acc}). With this definition, the efficiency-corrected yields of the LL and DD samples would need to be summed to give N_{tot} . It is, however, not possible to determine at the generator stage (at which N_{Acc} is obtained) which particle will be detected with a *long track* and which with a *downstream track*. The track type depends on the number and quality of hits in the Vertex Locator and tracking stations, and both are a matter of probabilities, which are only evaluated during simulation of the detector, that is, after the generator stage.

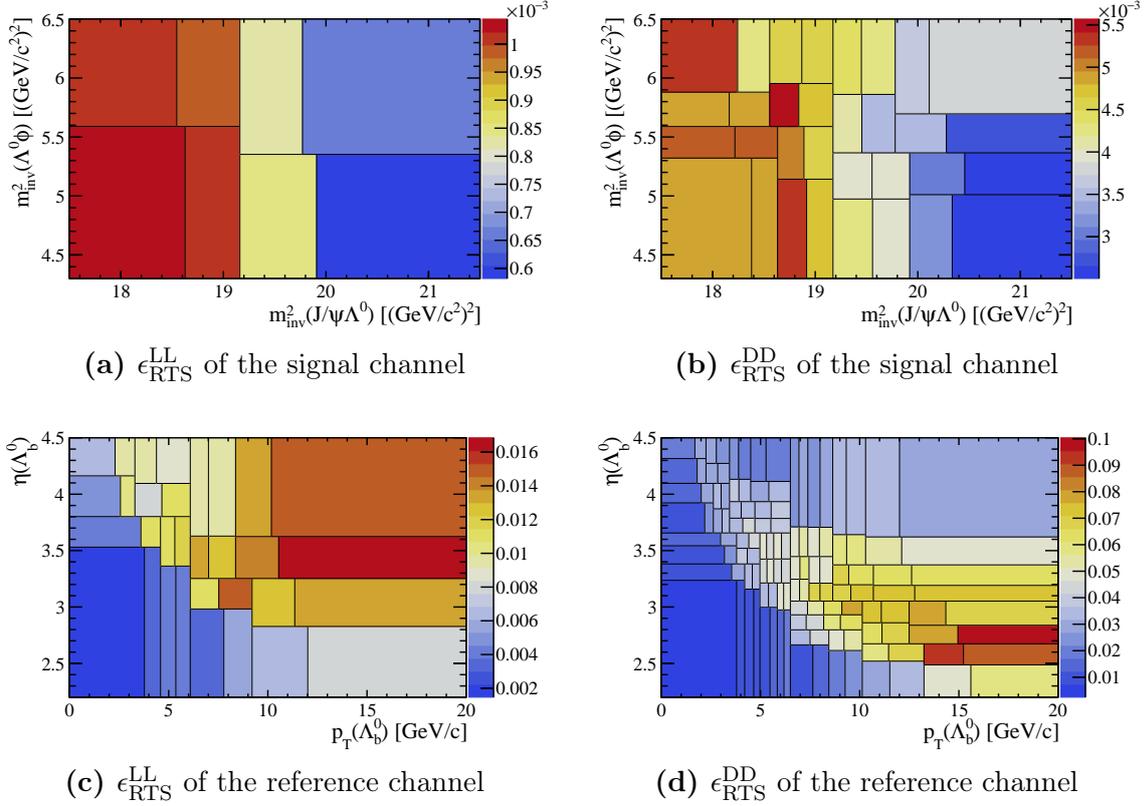


Figure 6.7: ϵ_{RTS} with adaptive binning for track type LL and DD for the signal channel (as a function of the Dalitz plot) and for the reference channel (as a function of $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$). Notice the 10^{-3} scale in case of the signal channel.

use of the limited available MC statistics, an adaptive binning was implemented. The adaptive binning algorithm takes a minimum number of events per bin (N_{min}) as input and as output gives a binning scheme with all bins having numbers of events between N_{min} and $2N_{\text{min}}$. In the case of efficiencies, it is sufficient to derive an adaptive binning scheme for the events in the numerator of the efficiency, since by definition these events are a subset of the events in the denominator.

The algorithm to determine the binning scheme consists of the following steps: (1) Start with 1x1 binning and split the x-axis at the median to get a 2x1 binning, so there are two bins with equal numbers of events. (2) Split the y-axis of each bin at the median only if the bin has more than $2N_{\text{min}}$ entries. (3) Split the x-axis of each bin at the median only if the bin has more than $2N_{\text{min}}$ entries. (4) Repeat the previous two steps until there are only bins left with $N_{\text{min}} < N < 2N_{\text{min}}$ entries.

Figure 6.7 shows $\epsilon_{\text{RTS}}^{\text{LL/DD}}$ for the signal and reference channels. Bins are larger in areas with low density of events after selection. The minimum number of events for the adaptive binning was set to 100 in all cases.

6.6 Average Efficiencies

To calculate the efficiency-corrected yield, the efficiency was looked up for each event individually. An average efficiency $\langle\epsilon\rangle$ was calculated with

$$\langle\epsilon\rangle = \frac{\sum_i sw_i \epsilon_i(x)}{\sum_i sw_i}$$

where $\epsilon_i(x)$ is the efficiency of event i depending on a set of variables x ; in this case $m_{inv}(J/\psi\Lambda^0)$ and $m_{inv}(\Lambda^0\phi)$ or $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$. sw_i is the sweight of the event. Defining $\langle\epsilon\rangle$ this way, the individual efficiencies are weighted by the sweight of the corresponding event, giving events with a higher *signal likeness* a higher weight.

Table 6.1 shows the average efficiencies for signal and reference channel. The combined efficiencies in ϵ_{RTS} here were broken down into three steps with $\epsilon_{\text{RTS}} = \epsilon_{\text{RecoStripCutsel}} \cdot \epsilon_{\text{Trig}} \cdot \epsilon_{\text{BDT}}$. To evaluate these stepwise average efficiencies, adaptively binned histograms were created analogously to the ones for ϵ_{RTS} . These histograms are shown in Appendix B. A definition of the stepwise efficiencies follows:

- $\epsilon_{\text{RecoStripCutsel}} = N_{\text{Cutsel}}/N_{\text{Acc}}$ (with N_{Cutsel} being the number of events after the cut-based selection and N_{Acc} the number of events after the acceptance cut) is the combined efficiency of the decay reconstruction, the stripping and the cut-based selection (excluding the BDT).
- $\epsilon_{\text{Trig}} = N_{\text{Trig}}/N_{\text{Cutsel}}$ is the trigger efficiency, with N_{Trig} being the number of events after specific trigger lines were chosen. This choice was made after the cut-based selection, since the stripping line used in this analysis takes all existing trigger lines as input (see the last paragraph of Section 4.3.2). Thus, ϵ_{Trig} here is the efficiency of the chosen trigger lines with respect to all available trigger lines.
- $\epsilon_{\text{BDT}} = N_{\text{BDT}}/N_{\text{Trig}}$ (with N_{BDT} being the number of events after application of the BDT) is the efficiency of the BDT. See Section 6.7 for a cross-check of this efficiency.
- $\epsilon_{\text{tot}} = N_{\text{BDT}}/N_{\text{Gen}}$ is the total combined efficiency of all the above efficiencies, including the acceptance cut efficiency. Notice that the numbers for ϵ_{tot} in Table 6.1 are not necessarily equivalent to the product $\epsilon_{\text{Acc}} \cdot \epsilon_{\text{RecoStripCutsel}} \cdot \epsilon_{\text{Trig}} \cdot \epsilon_{\text{BDT}}$, but to $\epsilon_{\text{Acc}} \cdot \epsilon_{\text{RTS}}$, where ϵ_{RTS} was calculated in a single step.

6.7 BDT Efficiencies

The validity of the BDT efficiencies depends on the agreement of the input variable distributions between data and MC. Comparisons of these distributions can be

Table 6.1: Average efficiencies with the exception of ϵ_{Acc} , which was calculated as a single number.

Efficiency	Sig chan LL	Sig chan DD	Ref chan LL	Ref chan DD
ϵ_{Acc}	0.6302	0.6302	0.6508	0.6508
$\langle \epsilon_{\text{RecoStripCutsel}} \rangle$	0.0011	0.0061	0.0115	0.0539
$\langle \epsilon_{\text{Trig}} \rangle$	0.8317	0.7813	0.7115	0.7003
$\langle \epsilon_{\text{BDT}} \rangle$	0.9671	0.9187	1	0.9668
$\langle \epsilon_{\text{tot}} \rangle$	0.0006	0.0027	0.0054	0.0240

Table 6.2: BDT efficiencies evaluated on MC and data.

Channel + Sample	$\langle \epsilon_{\text{BDT MC}} \rangle$	$\epsilon_{\text{BDT data}}$
Ref chan LL	1	1.001 ± 0.027
Ref chan DD	0.967	0.975 ± 0.017

found in Appendix A for the reference channel. For this comparison, data is shown reweighted with sweights obtained from a fit before the application of the BDT. The MC has been reweighted with the weights shown in Figure 6.1 to mimic the effect the binning in $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$ has on the efficiencies. In the signal channel, statistics are too low for a meaningful comparison.

As a cross-check, the BDT efficiency was evaluated on data by performing fits to the invariant mass spectrum of the Λ_b^0 daughters before and after the application of the BDT and dividing the yields; Table 6.2 shows the results, compared to the average BDT efficiencies evaluated on MC as described in Section 6.6. The uncertainties of the efficiencies evaluated on data are derived from the uncertainties on the yields. Again, the comparison is only shown for the reference channel due to the amount of statistics. The values are in good agreement.

7 Determination of the Branching Fraction

7.1 Efficiency-Corrected Yields

After obtaining all the necessary efficiencies (see Chapter 6), the efficiency-corrected yields were calculated. The yield per channel and track type (denoted as TT, it can be LL or DD) is given by

$$N_{\text{cor}}^{\text{TT}} = \sum_{i=1}^{N^{\text{TT}}} \frac{sw_i}{\epsilon_{\text{Acc}} \cdot \epsilon_{\text{RTS}}^{\text{TT}}(x)},$$

with the sum running over all data events of the LL or DD sample left after the BDT. The sweight sw_i of the i -th event was obtained in Chapter 5. The combined reconstruction, trigger, and selection efficiency $\epsilon_{\text{RTS}}^{\text{TT}}(x)$ depends on the set of event variables x , being either the event's position on the Dalitz plot or $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$. The acceptance cut efficiency ϵ_{Acc} is independent of the track type and of x .

By weighting events with their sweight, the background still present in the data set is statistically subtracted. Also notice that the sum over all sweights in a data set is equal to the total signal yield obtained from the fit from which the sweights were determined.

Table 7.1 shows the efficiency-corrected yields for each channel and track type. Statistical uncertainties are calculated by scaling the uncertainties on the uncorrected yields obtained from the fits in Chapter 5. The scaled statistical uncertainty is given by $\sigma_{\text{cor}} = \sigma \cdot \frac{N_{\text{cor}}}{N_{\text{fit}}}$ with N_{fit} being the uncorrected yield. Uncertainties related to efficiencies, originating in limited MC statistics (see Section 6.5.1), are also listed in Table 7.1.

Table 7.1: Efficiency corrected yields with errors

Channel+Sample	N_{cor}	Error stat.	Error eff.
Sig chan LL	$32\,254 \pm$	$8961 \pm$	771
Sig chan DD	$24\,711 \pm$	$3825 \pm$	316
Ref chan LL	$1\,170\,154 \pm$	$22\,227 \pm$	2011
Ref chan DD	$913\,015 \pm$	$11\,057 \pm$	1107

7.2 Determination of the Branching Fraction for LL and DD Samples

The general formula for the branching fraction of $\Lambda_b^0 \rightarrow J/\psi\Lambda^0\phi$ relative to $\Lambda_b^0 \rightarrow J/\psi\Lambda^0$ was given in Section 3.3. Expressed in terms of the quantities measured in this analysis, it is

$$\begin{aligned} \mathcal{R}^{\text{TT}} &:= \frac{\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi\Lambda^0\phi)}{\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi\Lambda^0)} \\ &= \frac{N_{\text{sig}}^{\text{TT}} \cdot \epsilon_{\text{ref}}^{\text{TT}}}{N_{\text{ref}}^{\text{TT}} \cdot \epsilon_{\text{sig}}^{\text{TT}} \cdot \mathcal{B}(\phi \rightarrow K^+K^-)} = \frac{N_{\text{sig cor}}^{\text{TT}}}{N_{\text{ref cor}}^{\text{TT}} \cdot \mathcal{B}(\phi \rightarrow K^+K^-)}, \end{aligned}$$

with TT denoting the LL or DD sample, $N_{\text{sig/ref}}$ being the yields for signal and reference channel, respectively, obtained by the signal extraction, $\epsilon_{\text{sig/ref}}$ being the corresponding efficiencies, and $N_{\text{sig/ref cor}}$ being the efficiency-corrected yields listed in Table 7.1. The branching fraction of the decay mode chosen for the ϕ in this analysis is $\mathcal{B}(\phi \rightarrow K^+K^-) = 0.489 \pm 0.005$ [8]. The uncertainty on this value is taken as a systematic uncertainty. Using these quantities, the branching fraction ratio was calculated separately for the LL and DD samples and was found to be

$$\begin{aligned} \mathcal{R}^{\text{LL}} &= 0.0537 \pm 0.0149 \text{ (stat.) and} \\ \mathcal{R}^{\text{DD}} &= 0.0527 \pm 0.0083 \text{ (stat.),} \end{aligned}$$

respectively. The values agree within statistical uncertainties. Systematic uncertainties are studied in Chapter 8 and added to the values of the branching fractions in Chapter 9. A weighted mean of \mathcal{R}^{LL} and \mathcal{R}^{DD} is also calculated in the latter chapter.

A note on the efficiency-corrected yields: The way the efficiencies were defined, the efficiency-corrected yields for the LL and DD samples of a specific channel, listed in Table 7.1, ideally would agree within uncertainties. For the signal channel, the level of agreement is hard to judge, but the values for the reference channel certainly do not agree. The ratios between the efficiency-corrected LL and DD yields for the signal and reference channel are $N_{\text{sig cor}}^{\text{LL}}/N_{\text{sig cor}}^{\text{DD}} = 1.31$ and $N_{\text{ref cor}}^{\text{LL}}/N_{\text{ref cor}}^{\text{DD}} = 1.28$, respectively. The similarity of these ratios points toward a misrepresented ratio of LL versus DD events in the simulation, which seems to be of the same order in the signal and the reference channel. Consequently, the discrepancies in the efficiency-corrected yields between the LL and DD samples cancel to a large degree when calculating the branching fraction ratio. This can be seen by forming the ratio $\mathcal{R}^{\text{LL}}/\mathcal{R}^{\text{DD}} = \frac{N_{\text{sig cor}}^{\text{LL}}/N_{\text{sig cor}}^{\text{DD}}}{N_{\text{ref cor}}^{\text{LL}}/N_{\text{ref cor}}^{\text{DD}}} = 1.02$. A systematic study on this topic can be found in Section 8.1.

8 Systematic Studies

8.1 LL/DD Discrepancies

As was observed in Section 7.2, there seem to exist discrepancies in the ratio of the number of LL versus DD events between data and simulation. In the branching fraction ratio, these discrepancies cancel to a large degree:

$$\frac{\mathcal{R}^{\text{LL}}}{\mathcal{R}^{\text{DD}}} = \frac{N_{\text{sig cor}}^{\text{LL}}/N_{\text{sig cor}}^{\text{DD}}}{N_{\text{ref cor}}^{\text{LL}}/N_{\text{ref cor}}^{\text{DD}}} = \frac{1.31}{1.28} = 1.02$$

To further investigate the origin of this discrepancy, the fraction of the raw, non-efficiency-corrected yield of the LL sample relative to the LL+DD yield was calculated for data and MC. This only was done in the reference channel, since the statistical uncertainties in the signal channel are too high for a meaningful comparison. On data, the ratio is $N_{\text{ref non-cor}}^{\text{LL}}/N_{\text{ref non-cor}}^{\text{tot}} = 0.24$ and on MC it is $N_{\text{ref MC non-cor}}^{\text{LL}}/N_{\text{ref MC non-cor}}^{\text{tot}} = 0.19$. The discrepancy between these ratios, which were obtained previous to the efficiency correction, strengthens the hypothesis of a misrepresented ratio of LL versus DD events in the MC.

Recall that the LL and DD samples are defined by the track types of the Λ^0 daughter particles. In case of the LL sample, both p and π^- have *long tracks*; in case of the DD sample, both have *downstream tracks*. *Long tracks* have track segments both in the VELO and the T-stations T1-T3, whereas *downstream tracks* have no track segments in the VELO, but only in the TT and the T-stations T1-T3. Consequently, the z-position of the Λ^0 decay vertex is a quantity which will differ between LL and DD events. Figure 8.1 shows the distribution of this position for the LL, the DD and the combined LL+DD sample on data and MC in the reference channel. The data and MC distributions are in relatively good agreement in case of the individual LL and DD sample, respectively. The agreement is noticeably worse for the combined LL+DD sample due to the misrepresented ratio of LL and DD events. The relatively good agreement for the individual samples lets us conclude that the z-position of the Λ^0 decay vertex is not one of the relevant misrepresented variables leading to the misrepresented LL/DD ratio. A more relevant variable might be the z-position of the first hit associated with a track. However, whether an event ends up as a *long* or *downstream track* most likely depends on a large number of variables and the details of the detector geometry¹. Due to time constraints, it was not possible to find the exact cause of the misrepresentation in the simulation.

¹Incidentally, this is the reason why the denominator in the reconstruction efficiency was defined

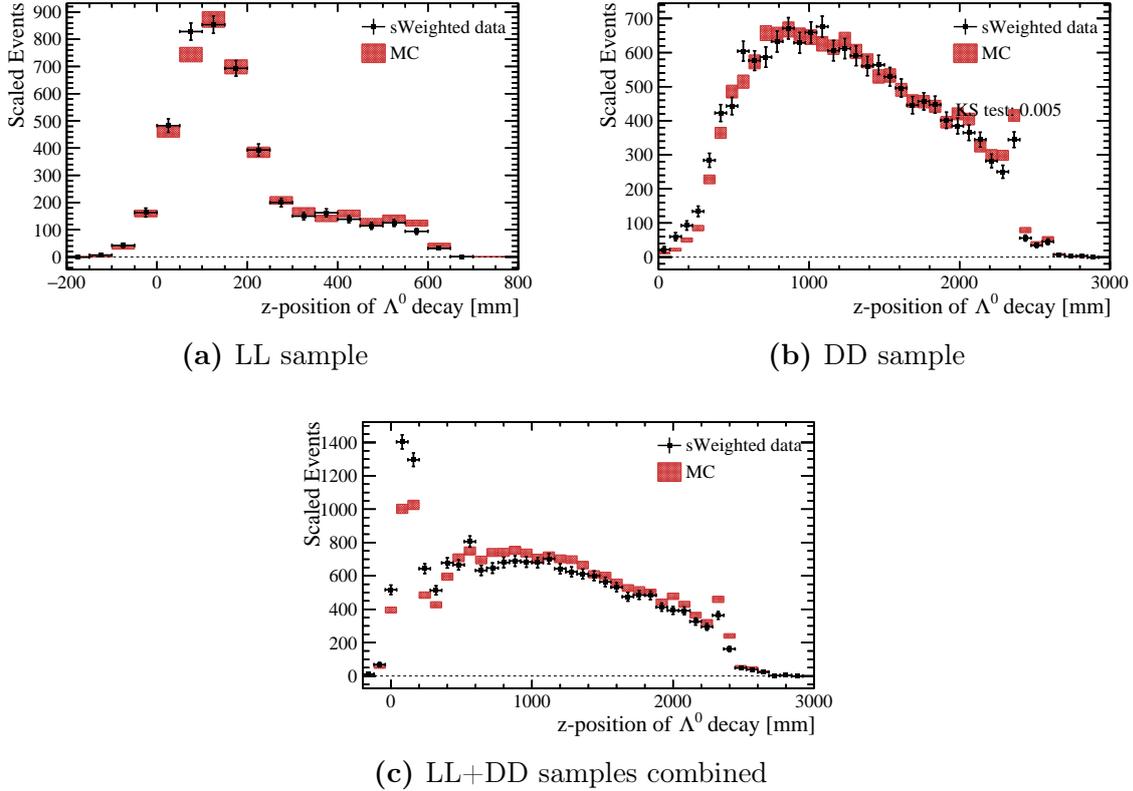


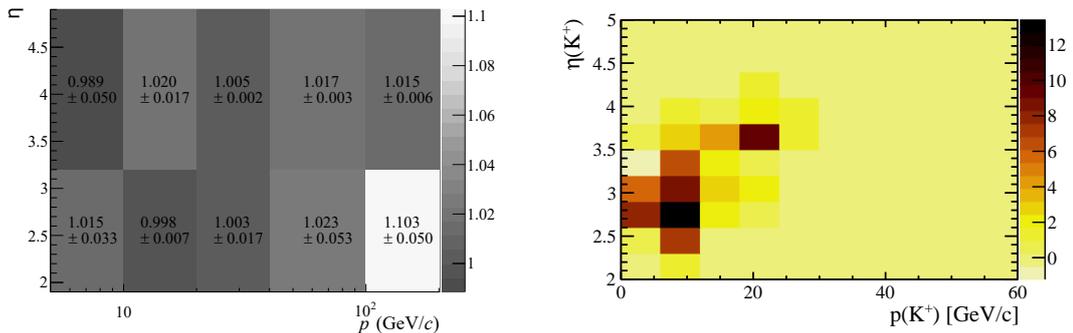
Figure 8.1: z-position of the decay vertex of the Λ^0 , comparison between data and MC in the reference channel. The location of the peak at $z \approx 2300$ mm corresponds to the position of the TT. The last VELO station is located at $z \approx 700$ mm.

To take the LL/DD discrepancy into account, a systematic uncertainty of $|\mathcal{R}^{\text{LL}} - \mathcal{R}^{\text{DD}}| = 0.0010$ was introduced. This uncertainty is added to the systematic uncertainty of the weighted mean branching fraction ratio $\overline{\mathcal{R}}$, calculated in Chapter 9.

8.2 Track Reconstruction Efficiency of the Kaons

The track reconstruction efficiency is the efficiency to reconstruct trajectories of charged particles, which have passed through the whole detector. This efficiency is taken into account in the MC used to obtain the efficiencies in Chapter 6. However, the tracking efficiency is not perfectly modeled by the simulation. The discrepancies between data and MC cancel in case of the daughter particles of the J/ψ and Λ^0 , since these are present both in the signal and the reference channel. Effects due

as the *total* number of generated events, not the number of generated events ending up as LL or DD events, if reconstructed.



(a) Tracking efficiency ratio between 2011 data and MC ($\epsilon_{\text{data}}/\epsilon_{\text{MC}}$). Figure taken from [39].

(b) Background subtracted distribution of (p, η) of the K^+ in the signal channel, LL and DD samples combined.

Figure 8.2: Left: Tracking efficiency discrepancy. Right: Kaon distribution.

to small differences between the kinematics of these particles in the signal and the reference channel, respectively, are negligible. The ϕ is only present in the signal channel. Consequently, discrepancies in the tracking efficiency of the K^+ and K^- do not cancel and have to be taken into account.

In [39], the tracking efficiency for particles traversing the whole detector and forming *long tracks* (hits in VELO and T1-T3, see Section 3.2) was measured with tag-and-probe methods and using the decay $J/\psi \rightarrow \mu^+\mu^-$. Tag-and-probe methods work as follows: Such decays of the J/ψ are used in which one muon, called the tag, is well reconstructed, forming a long track, and the other muon, called the probe, is only partially reconstructed, but leaves sufficient information to reconstruct the J/ψ . Since long tracks are formed from hits in the VELO and T1-T3, probes have missing tracks in either one or both of these systems. If the partially reconstructed probe track can be matched to a *long track*, the reconstruction is counted as efficient. For details, see [39].

After measuring the tracking efficiency on data, it was compared to the efficiency obtained from MC. Figure 8.2a shows the tracking efficiency ratio between data and MC in bins of p and η of the probe muon. The background subtracted (p, η) distribution of the K^+ from the signal channel is shown in Figure 8.2b. Since the K^+ has *long tracks* in both LL and DD samples, they were combined for this plot. The distribution of the K^- is not shown, but as expected, is very similar to that of the K^+ . The (p, η) regions with relevant kaon contributions show only minor discrepancies between data and MC. Consequently, and owing to time constraints, a systematic uncertainty of 1% for each kaon was assigned to the efficiency-corrected yield of the signal channel, resulting in a combined uncertainty of $\sqrt{2} \cdot 1\%$. This uncertainty was assigned to both the LL and DD sample. Values are listed in Table 8.1.

8.3 Helicity Angle of the ϕ

The ϕ is a vector meson ($S=1$), resulting in three possible helicity configurations $h = \pm 1, 0$, with helicity being the normalized component of a particle's spin along its flight direction: $h = \vec{S} \cdot \vec{p}/p$. An important, related quantity is the so-called helicity angle θ_h . In the $\phi \rightarrow K^+K^-$ system, this is the angle between (1) the flight direction of the ϕ in the laboratory frame and (2) the flight direction of one of the kaons in the rest frame of the ϕ . Since the kaons move back-to-back in the ϕ rest frame, the angle is independent of the choice of K^+ or K^- .

If the ϕ is produced without polarization, meaning that all three helicity configurations equally contribute to the particle's state, the distribution of θ_h is flat. If, however, the ϕ has a non-zero polarization, the θ_h distribution will acquire a dependence on $\cos^2 \theta_h$, resulting in a distribution described by the PDF $A + B \cdot \cos^2 \theta_h$ with A and B being constants. The values of A and B depend on the contributions of the individual helicity states to the polarization. The $\cos^2 \theta_h$ dependence is a consequence of a spin 1 particle (ϕ) decaying into two spin 0 particles (K^+, K^-). For details on the formalism with which these angular distributions can be derived, see [40].

The signal channel MC samples used in this analysis were produced with a flat θ_h distribution. Consequently, it is necessary to check whether this distribution is indeed flat in the data. Figure 8.3 shows the cosine of the helicity angle θ_h of the ϕ for background subtracted data and MC. The LL and DD samples were combined, since they do not differ in the nature of the ϕ .

The distribution on data clearly is not flat and some $\cos^2 \theta_h$ dependence seems to be present, which indicates a non-zero polarization of the ϕ . To account for this polarization, the signal channel MC ideally should be reweighted in $\cos \theta_h$. To obtain the necessary weights, a PDF of the form $A + B \cdot \cos^2 \theta_h$ should first be fitted to the data, to better define the shape of the distribution. Reweighting the MC in $\theta_h(\phi)$ would be a worthwhile future improvement to the analysis.

8.4 Fit Model

The effect of choosing PDFs different from the ones used for the fits described in Chapter 5 was studied. The same alternative PDFs were chosen for the signal and the reference channel. Specifically, a double Crystal Ball function replaced the double Gaussian modeling the signal, and a first order polynomial replaced the exponential function modeling the background.

A Crystal Ball (CB) function, named for the Crystal Ball collaboration, which first made use of it [41], is a Gaussian with a power-law tail on the low side to model

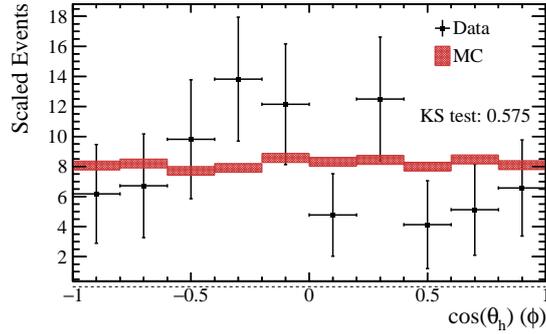


Figure 8.3: Cosine of the helicity angle θ_h of the ϕ . Background subtracted data compared to MC, combined LL+DD signal channel samples.

radiative energy losses. The function is defined as

$$f_{\text{CB}}(x; \alpha, n, \mu, \sigma) = N \cdot \begin{cases} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), & \text{for } \frac{x-\mu}{\sigma} > -\alpha \\ A \cdot (B - \frac{x-\mu}{\sigma})^{-n}, & \text{for } \frac{x-\mu}{\sigma} \leq -\alpha \end{cases}$$

with

$$A = \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right), \quad B = \frac{n}{|\alpha|} - |\alpha|,$$

and N being a normalization factor.

Fits with the alternative PDFs were performed on the final data sets after application of the BDTs. In case of the signal channel, all parameters of the double Crystal Ball except the shared mean were constrained to values obtained from fits to the MC. These parameters are α_i , n_i , and σ_i , with $i = 1, 2$ denoting the two CB functions. Signal channel fits with the alternative PDFs are shown in Figure 8.4 and reference channel fits in Figure 8.5.

Efficiency-corrected yields were calculated analogously to the main analysis, with the sweights taken from the fits described above. From the resulting yields, branching fraction ratios $\mathcal{R}_{\text{alt. PDFs}}^{\text{LL/DD}}$ were obtained. The absolute difference between the branching fraction ratio values obtained with the main and the alternative PDFs, were taken as a systematic uncertainty. Values are listed in Table 8.1.

8.5 Summary of Systematic Uncertainties

In Table 8.1, the individual systematic uncertainties are listed. The uncertainty on the efficiency is a consequence of limited MC statistics and is described in Section 6.5.1. The total systematic uncertainty is given by $\sqrt{\sum \text{Eff}_i^2}$ with Eff_i being the individual uncertainties.

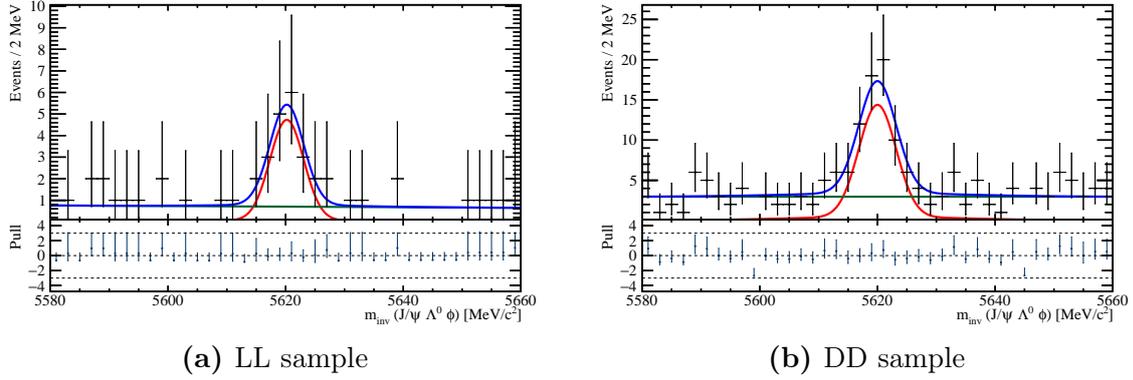


Figure 8.4: Fits to the mass spectrum of the Λ_b^0 candidates in the signal channel with alternative PDFs.

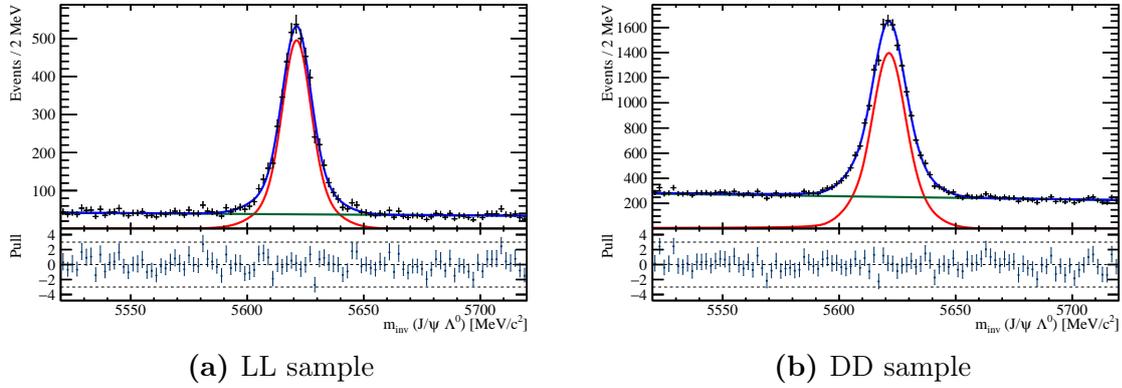


Figure 8.5: Fits to the mass spectrum of the Λ_b^0 candidates in the reference channel with alternative PDFs.

Table 8.1: Individual systematic uncertainties on the branching fraction and their quadratic sum. In the last row, the branching fraction values are shown for reference. *The systematic uncertainty on the LL/DD discrepancy will be added to the weighted mean of the branching fraction ratio calculated in Chapter 9.

Systematic	LL	DD
Efficiency	0.0014	0.0007
$\mathcal{B}(\phi \rightarrow K^+ K^-)$	0.0005	0.0005
ϕ tracking eff.	0.0008	0.0007
Fit model	0.0004	0.0015
$(\sum \text{Eff}_i^2)^{1/2}$	0.0017	0.0019
LL/DD discrepancy*	0.0010	
\mathcal{R}	0.0537	0.0527

9 Results

9.1 Weighted Mean Branching Fraction

The branching fraction ratios for the LL and DD samples, including the systematic uncertainties listed in Table 8.1, are

$$\begin{aligned}\mathcal{R}^{\text{LL}} &= 0.0537 \pm 0.0149 \text{ (stat.)} \pm 0.0017 \text{ (syst.)} \text{ and} \\ \mathcal{R}^{\text{DD}} &= 0.0527 \pm 0.0083 \text{ (stat.)} \pm 0.0019 \text{ (syst.)},\end{aligned}$$

respectively. The two results were combined into a weighted mean value with the data samples being independent of each other and consequently having uncorrelated statistical uncertainties. With these assumptions, the statistical uncertainty on the weighted mean value is minimized by taking the weight w^{TT} (TT = LL or DD) for the LL and DD values as the inverse squared of the statistical uncertainties σ^{TT} on the LL and DD value, respectively: $w^{\text{TT}} = (\sigma^{\text{TT}})^{-2}$. The weighted mean branching fraction ratio $\overline{\mathcal{R}}$ is then given by

$$\overline{\mathcal{R}} = \frac{(\sigma^{\text{LL}})^{-2} \cdot \mathcal{R}^{\text{LL}} + (\sigma^{\text{DD}})^{-2} \cdot \mathcal{R}^{\text{DD}}}{(\sigma^{\text{LL}})^{-2} + (\sigma^{\text{DD}})^{-2}},$$

and the statistical uncertainty on $\overline{\mathcal{R}}$ is

$$\sigma(\overline{\mathcal{R}}) = \sqrt{\frac{1}{(\sigma^{\text{LL}})^{-2} + (\sigma^{\text{DD}})^{-2}}}.$$

The systematic uncertainty on $\overline{\mathcal{R}}$ was obtained by first combining the systematic uncertainties of \mathcal{R}^{LL} and \mathcal{R}^{DD} into

$$Err_{\text{sys.}}^{\text{LL+DD}} = \left(\frac{(w^{\text{LL}} \cdot Err_{\text{sys.}}^{\text{LL}})^2 + (w^{\text{DD}} \cdot Err_{\text{sys.}}^{\text{DD}})^2}{(w^{\text{LL}} + w^{\text{DD}})^2} \right)^{1/2}$$

and finally combining this uncertainty with the uncertainty due to the discrepancies between the LL/DD ratio in data and MC (Section 8.1):

$$Err_{\text{sys.}}(\overline{\mathcal{R}}) = \sqrt{(Err_{\text{sys.}}^{\text{LL+DD}})^2 + (Err_{\text{sys.}}^{\text{LL/DD discr}})^2}.$$

9.2 Conclusion

The final result for the branching fraction ratio was found to be

$$\overline{\mathcal{R}} = \frac{\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi \Lambda^0 \phi)}{\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi \Lambda^0)} = 0.0529 \pm 0.0072 \text{ (stat.)} \pm 0.0018 \text{ (syst.)}$$

This is the first observation of the decay $\Lambda_b^0 \rightarrow J/\psi \Lambda^0 \phi$. The branching fraction ratio was obtained by analyzing a data set with an integrated luminosity of 3 fb^{-1} , collected in 2011 and 2012. Since 2015, LHCb is taking data with increased center-of-mass energy and luminosity. Consequently, it is only a matter of time until sufficient data is available to perform an amplitude analysis and search for strange pentaquarks in $\Lambda_b^0 \rightarrow J/\psi \Lambda^0 \phi$.

Part III

Appendix

A BDT Input Variables

A.1 Data versus MC

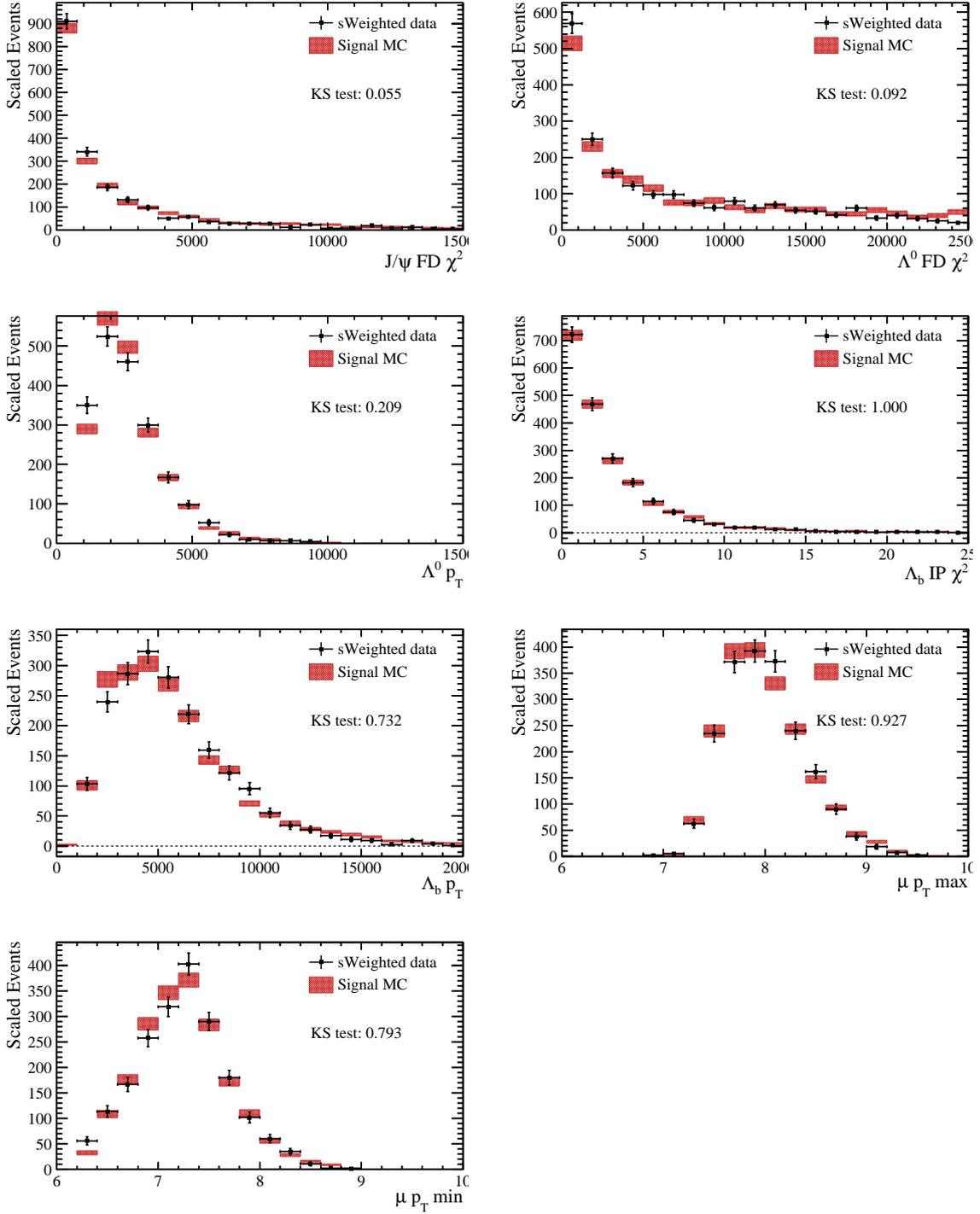


Figure A.1: Input variables of the reference channel LL sample, MC is reweighted in $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$.

A BDT Input Variables

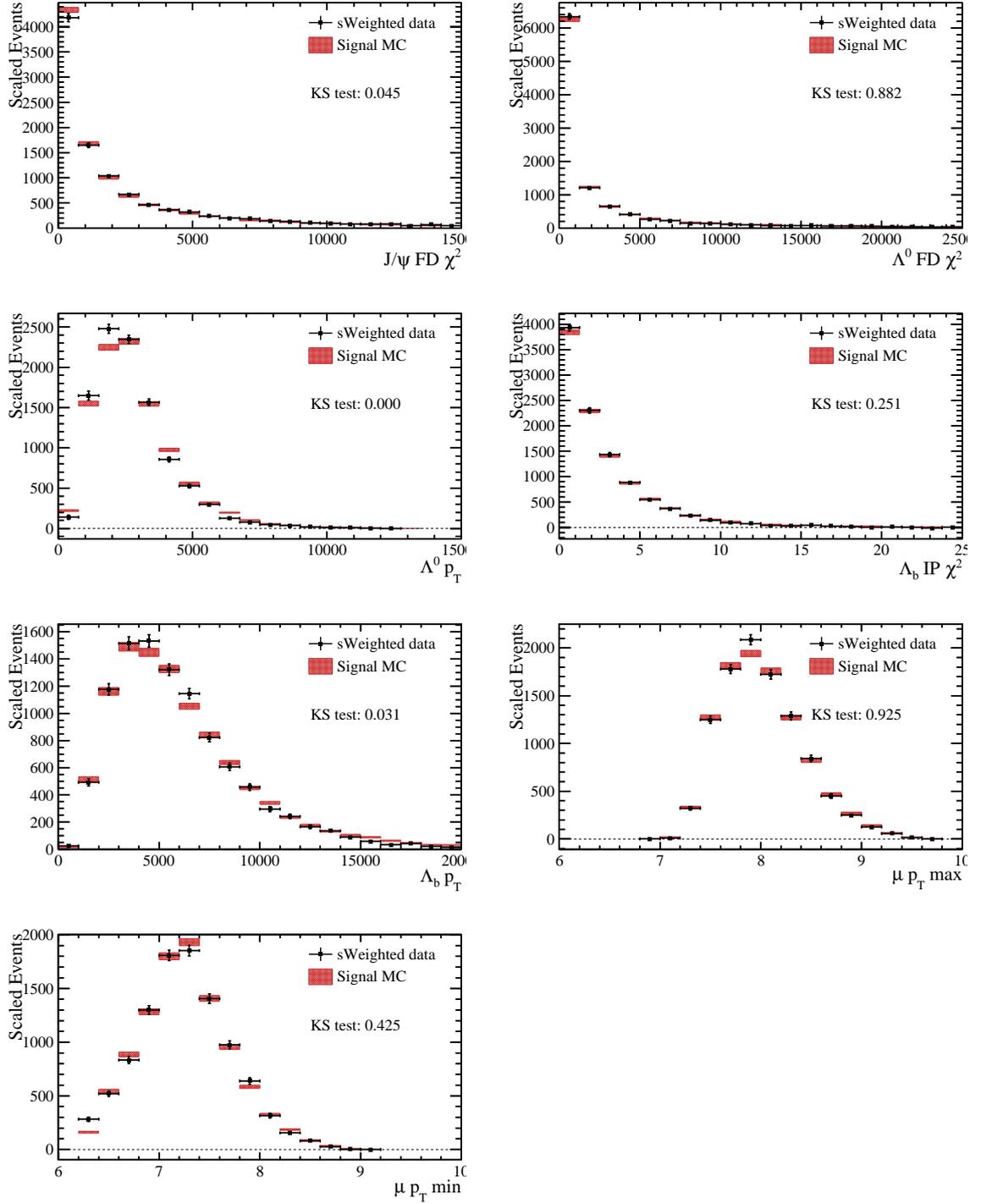
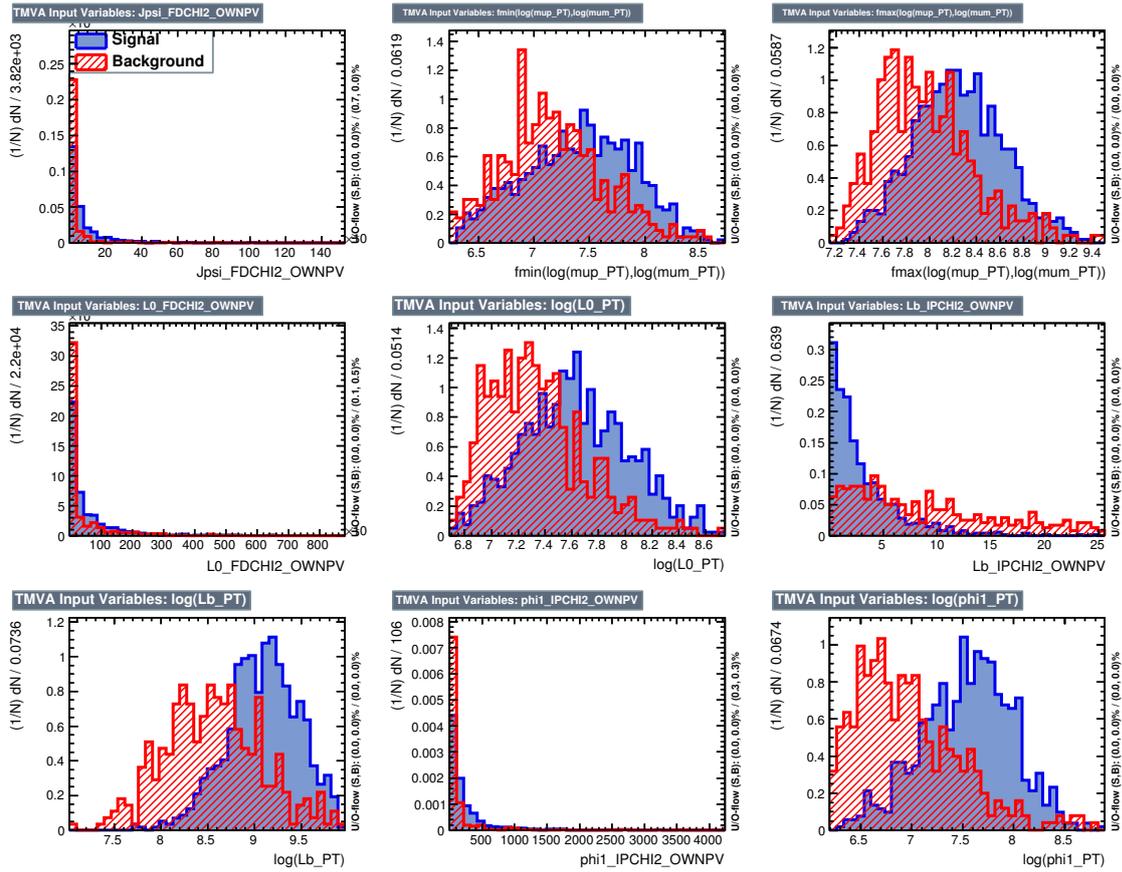


Figure A.2: Input variables of the reference channel DD sample, MC is reweighted in $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$.

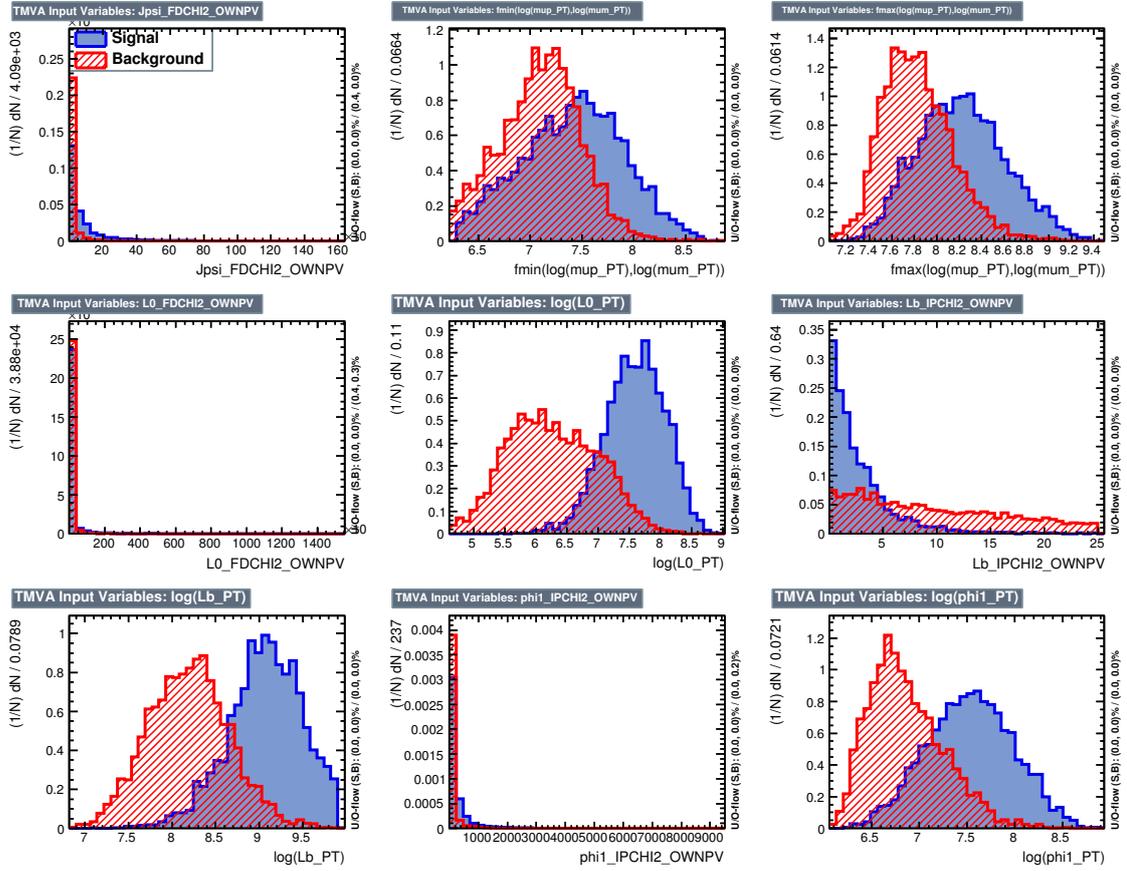
A.2 Training Samples

Shown are the distributions of the BDT input variables of the signal and background training samples.

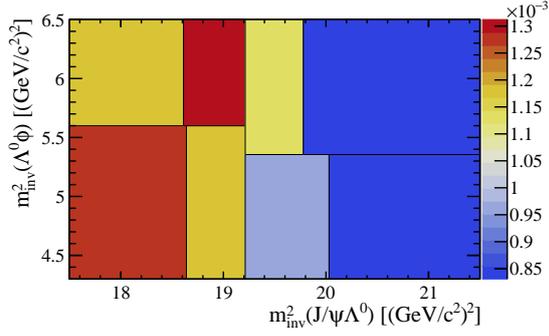
Signal Channel LL Samples



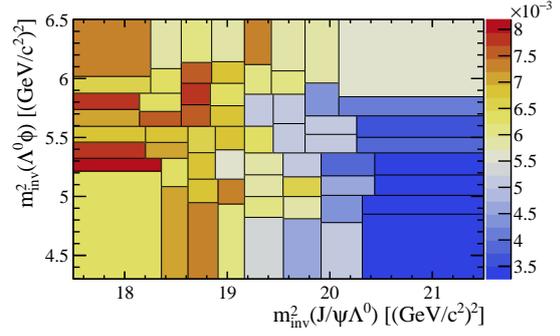
Signal Channel DD Samples



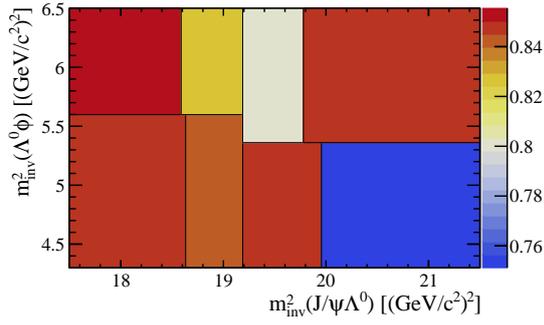
B Efficiency Plots



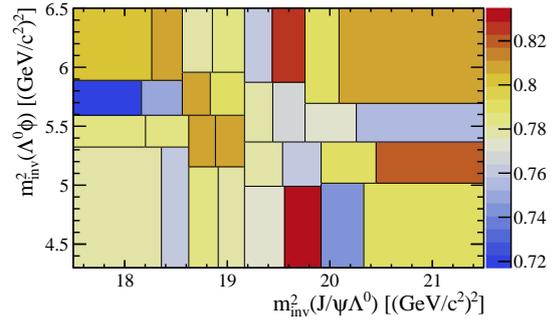
(a) $\epsilon_{\text{RecoStripCutsel}}^{\text{LL}}$ of the signal channel



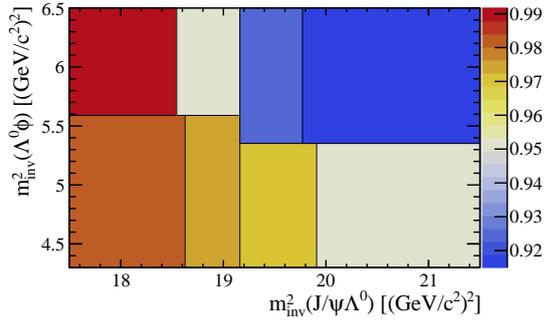
(b) $\epsilon_{\text{RecoStripCutsel}}^{\text{DD}}$ of the signal channel



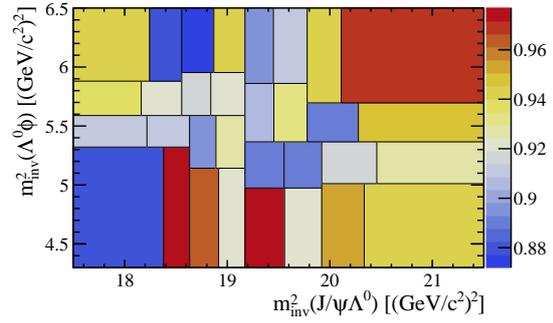
(c) $\epsilon_{\text{Trigg}}^{\text{LL}}$ of the signal channel



(d) $\epsilon_{\text{Trigg}}^{\text{DD}}$ of the signal channel



(e) $\epsilon_{\text{BDT}}^{\text{LL}}$ of the signal channel



(f) $\epsilon_{\text{BDT}}^{\text{DD}}$ of the signal channel

Figure B.1: Stepwise efficiencies for the signal channel. Average efficiencies in Section 6.6 are based on the above histograms.

B Efficiency Plots

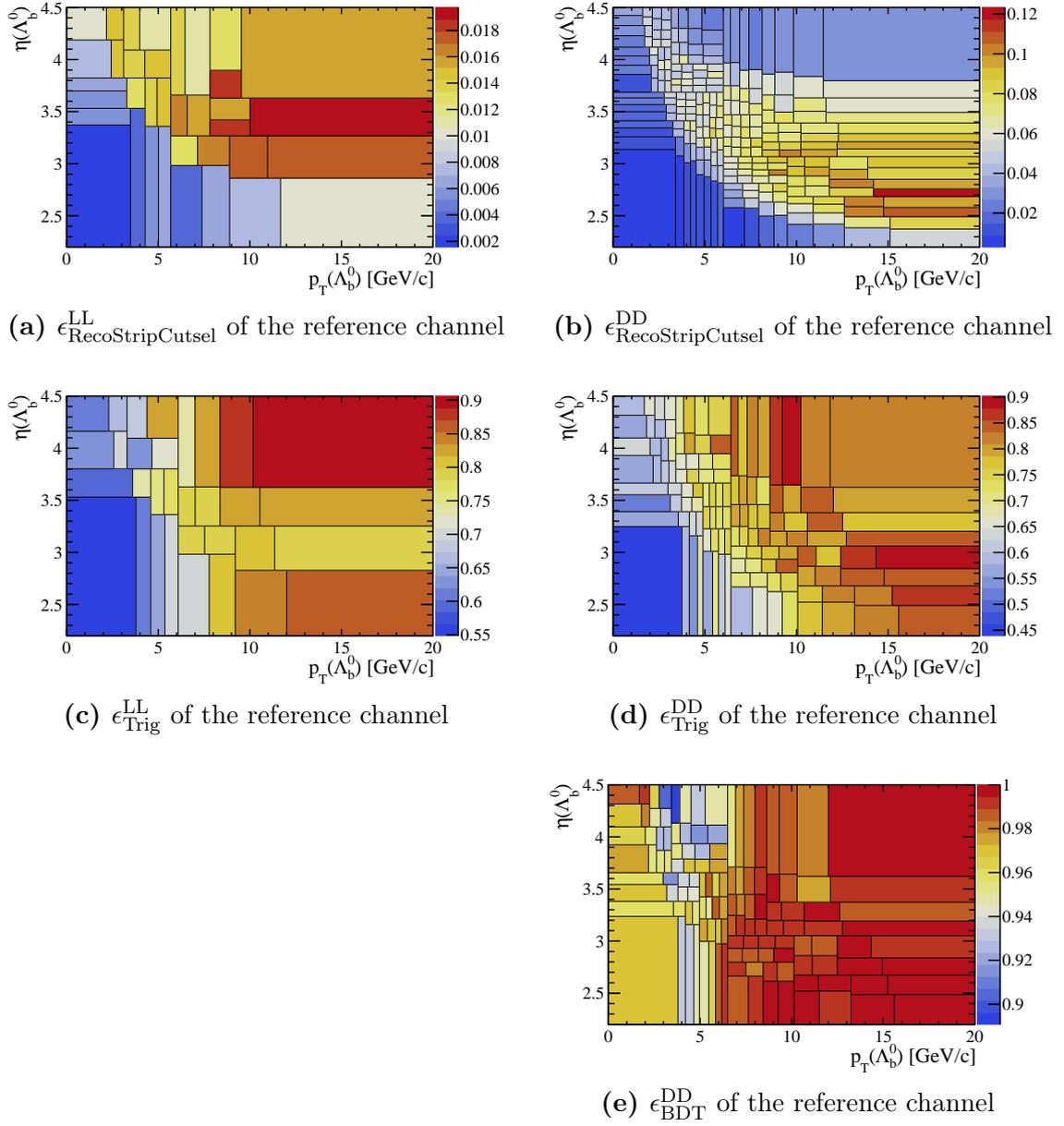


Figure B.2: Stepwise efficiencies for the reference channel. Average efficiencies in Section 6.6 are based on the above histograms. The histogram for $\epsilon_{\text{BDT}}^{\text{LL}}$ is not shown due to a faulty plotting function. The value of $\langle \epsilon_{\text{BDT}}^{\text{LL}} \rangle$, however, is correctly obtained from this histogram (see the efficiency obtained from data, described in Section 6.7).

C Bibliography

- [1] R. Aaij *et al.*, “Observation of $J/\psi p$ resonances consistent with pentaquark states in $\Lambda_b^0 \rightarrow J/\psi K^- p$ decays,” *Phys. Rev. Lett.*, vol. 115, p. 072001, 2015.
- [2] R. Aaij *et al.*, “Model-independent evidence for $J/\psi p$ contributions to $\Lambda_b^0 \rightarrow J/\psi p K^-$ decays,” *Phys. Rev. Lett.*, vol. 117, p. 082002, 2016.
- [3] M. Thomson, *Modern particle physics*. New York: Cambridge University Press, 2013.
- [4] H. J. Lipkin, “New possibilities for exotic hadrons - anticharmed strange baryons,” *Physics Letters B*, vol. 195, pp. 484–488, 1987.
- [5] M. Gell-Mann, “A schematic model of baryons and mesons,” *Physics Letters*, vol. 8, no. 3, pp. 214–215, 1964.
- [6] G. Zweig, “An SU_3 model for strong interaction symmetry and its breaking; Version 2,” Tech. Rep. CERN-TH-412, CERN, 1964.
- [7] K. H. Hicks, “On the conundrum of the pentaquark,” *The European Physical Journal H*, vol. 37, no. 1, pp. 1–31, 2012.
- [8] C. Patrignani *et al.*, “Review of Particle Physics,” *Chin. Phys.*, vol. C40, no. 10, 2016.
- [9] R. Aaij *et al.*, “Evidence for exotic hadron contributions to $\Lambda_b^0 \rightarrow J/\psi p \pi^-$ decays,” *Phys. Rev. Lett.*, vol. 117, p. 082003, 2016.
- [10] H.-X. Chen, W. Chen, X. Liu, and S.-L. Zhu, “The hidden-charm pentaquark and tetraquark states,” *Phys. Rept.*, vol. 639, pp. 1–121, 2016.
- [11] F.-K. Guo, U.-G. Meißner, W. Wang, and Z. Yang, “How to reveal the exotic nature of the $P_c(4450)$,” *PRD*, vol. 92, no. 7, p. 071502, 2015.
- [12] T. J. Burns, “Phenomenology of $p_c(4380)^+$, $p_c(4450)^+$ and related states,” *The European Physical Journal A*, vol. 51, no. 11, p. 152, 2015.
- [13] L. Maiani, A. Polosa, and V. Riquer, “The new pentaquarks in the diquark model,” *Physics Letters B*, vol. 749, pp. 289–291, 2015.

- [14] A. Feijoo, V. K. Magas, A. Ramos, and E. Oset, “A hidden-charm $S = -1$ pentaquark from the decay of Λ_b^0 into $J/\psi\eta\Lambda^0$ states,” *Eur. Phys. J.*, vol. C76, no. 8, p. 446, 2016.
- [15] J.-X. Lu, E. Wang, J.-J. Xie, L.-S. Geng, and E. Oset, “The $\Lambda_b \rightarrow J/\psi K^0\Lambda$ reaction and a hidden-charm pentaquark state with strangeness,” *Phys. Rev.*, vol. D93, p. 094009, 2016.
- [16] L. Evans and P. Bryant, “LHC Machine,” *Journal of Instrumentation*, vol. 3, no. 08, p. S08001, 2008.
- [17] A. A. Alves, Jr. *et al.*, “The LHCb Detector at the LHC,” *Journal of Instrumentation*, vol. 3, no. 08, 2008.
- [18] V. Gligorov, “Performance and upgrade plans of the LHCb trigger system,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 718, pp. 26–29, 2013. Proceedings of the 12th Pisa Meeting on Advanced Detectors La Biodola, Isola d’Elba, Italy, May 20-26, 2012.
- [19] LHCb Collaboration, “Material for Presentations.” http://lhcb.web.cern.ch/lhcb/speakersbureau/html/Material_for_Presentations.html, 2017.
- [20] R. Antunes-Nobrega *et al.*, “LHCb reoptimized detector design and performance: Technical design report,” tech. rep., CERN, 2003.
- [21] LHCb Collaboration, “LHCb VELO TDR: Vertex Locator. Technical design report,” Tech. Rep. CERN-LHCC-2001-011, CERN, 2001.
- [22] LHCb Collaboration, “LHCb Silicon Tracker - Material for Publications.” <http://lhcb.physik.uzh.ch/ST/public/material/index.php>, 2006.
- [23] LHCb Collaboration, “LHCb: Inner Tracker technical design report,” Tech. Rep. CERN-LHCC-2002-029, CERN, 2002.
- [24] O. Callot and S. Hansmann-Menzemer, “The Forward Tracking,” tech. rep., CERN, 2007.
- [25] M. Needham and J. Van Tilburg, “Performance of the track matching,” Tech. Rep. CERN-LHCb-2007-020, CERN, 2007.
- [26] O. Callot, “Downstream pattern recognition,” Tech. Rep. CERN-LHCB-2007-026, CERN, 2007.
- [27] LHCb Collaboration, “LHCb Muon System Technical Design Report.” <http://lhcb-muon.web.cern.ch/lhcb-muon/results/tdr.pdf>, 2001.

-
- [28] J. Albrecht, V. V. Gligorov, G. Raven, and S. Tolk, “Performance of the LHCb High Level Trigger in 2012,” *J. Phys.: Conf. Ser.*, vol. 513, no. CERN-LHCb-PROC-2014-005, p. 012001. 8 p, 2014. arXiv:1310.8544.
- [29] T. Sjostrand, L. Lonnblad, and S. Mrenna, “PYTHIA 6.2: Physics and manual,” 2001.
- [30] D. J. Lange, “The EvtGen particle decay simulation package,” *Nucl. Instrum. Meth.*, vol. A462, pp. 152–155, 2001.
- [31] S. Agostinelli *et al.*, “GEANT4: A Simulation toolkit,” *Nucl. Instrum. Meth.*, vol. A506, no. SLAC-PUB-9350, FERMILAB-PUB-03-339, pp. 250–303, 2003.
- [32] R. Aaij and J. Albrecht, “Muon triggers in the High Level Trigger of LHCb,” Tech. Rep. CERN-LHCb-PUB-2011-017, CERN, 2011.
- [33] M. Pivk and F. R. Le Diberder, “sPlot: A statistical tool to unfold data distributions,” *Nucl. Instrum. Meth.*, vol. A555, pp. 356–369, 2005.
- [34] A. Höcker *et al.*, “TMVA - Toolkit for multivariate data analysis with ROOT: Users Guide,” Tech. Rep. CERN-OPEN-2007-007, CERN, 2007.
- [35] G. Cowan, *Statistical Data Analysis*. Oxford science publications, Clarendon Press, 1998.
- [36] W. D. Hulsbergen, “Decay chain fitting with a Kalman filter,” *Nucl. Instrum. Meth.*, vol. A552, pp. 566–575, 2005.
- [37] W. Verkerke and D. Kirkby, “RooFit Users Manual v2.91.” https://root.cern.ch/download/doc/RooFit_Users_Manual_2.91-33.pdf, 2008.
- [38] R. H. Dalitz, “On the analysis of τ -meson data and the nature of the τ -meson,” *Philosophical Magazine*, 1953.
- [39] M. De Cian, S. Hansmann-Menzemer, A. Jaeger, P. Seyfert, J. van Tilburg, and S. Wandernoth, “Measurement of the track finding efficiency,” Tech. Rep. CERN-LHCb-PUB-2011-025, CERN, 2012.
- [40] S. U. Chung, “Spin formalisms,” tech. rep., Brookhaven Nat. Lab., 2008.
- [41] J. E. Gaiser, *Charmonium spectroscopy from radiative decays of the J/ψ and ψ'* . PhD thesis, SLAC, 1982.

Erklärung:

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

Heidelberg, den 19.05.2017

.....