

Department of Physics and Astronomy  
University of Heidelberg

Bachelor Thesis in Physics  
submitted by

**Daniel Baitinger**

born in Sinsheim (Germany)

**2016**

Measurement of the branching ratio of the decay  
 $B^+ \rightarrow J/\Psi K^{*+}$  at the LHCb experiment

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

## Abstract

This thesis presents a determination of the branching ratio of  $B^+ \rightarrow J/\Psi K^{*+}$ . The used data was recorded at the LHCb experiment with center-of-mass energies of 7 TeV and 8 TeV, corresponding to an integrated luminosity of  $3 \text{ fb}^{-1}$ . For this analysis, decays with  $K^{*+} \rightarrow K^+ \pi^0$  followed by  $\pi^0 \rightarrow \gamma\gamma$  are used. The branching ratio is determined relative to the decay  $B^0 \rightarrow J/\Psi K^{*0}$  with  $K^{*0} \rightarrow K^+ \pi^-$ . Unbinned maximum likelihood fits to the mass spectrum of the  $B$  candidates are used to obtain the number of events. This results in a branching ratio of

$$\mathcal{B}(B^+ \rightarrow J/\Psi K^{*+}) = (1.64 \pm 0.03_{stat.} \pm 0.21_{syst.}) \cdot 10^{-3}.$$

The result is dominated by the systematic uncertainty, which is caused by an insufficient description of partially reconstructed  $B$  decays which form significant backgrounds. The statistical uncertainty is smaller than in the previous measurements of the BaBar or Belle experiments.

## Zusammenfassung

In der vorliegenden Arbeit wird das Verzweigungsverhältnis des Zerfalls  $B^+ \rightarrow J/\Psi K^{*+}$  bestimmt. Die verwendeten Daten wurden am LHCb Experiment bei Schwerpunktsenergien von 7 TeV und 8 TeV aufgenommen und entsprechen einer integrierten Luminosität von  $3 \text{ fb}^{-1}$ . Für diese Analyse werden Zerfälle mit  $K^{*+} \rightarrow K^+ \pi^0$  und  $\pi^0 \rightarrow \gamma\gamma$  untersucht. Das Verzweigungsverhältnis wird relativ zum Zerfall  $B^0 \rightarrow J/\Psi K^{*0}$  mit  $K^{*0} \rightarrow K^+ \pi^-$  bestimmt. Durch eine Maximum-Likelihood-Fit-Methode wird die Anzahl der Ereignisse aus der Massenverteilung der  $B$ -Kandidaten extrahiert. Dadurch ergibt sich das Verzweigungsverhältnis

$$\mathcal{B}(B^+ \rightarrow J/\Psi K^{*+}) = (1.64 \pm 0.03_{stat.} \pm 0.21_{syst.}) \cdot 10^{-3}.$$

Das Ergebnis wird dominiert vom systematischen Fehler, der aus einer unzureichende Beschreibung von teilweise rekonstruierten  $B$ -Zerfällen hervorgeht. Der statistische Fehler ist kleiner als bei früheren Messungen der BaBar oder Belle Experimente.

# Contents

<b>1</b>	<b>Introduction</b>	<b>6</b>
<b>2</b>	<b>Theoretical overview</b>	<b>8</b>
2.1	The Standard Model of particle physics . . . . .	8
2.2	Flavour physics . . . . .	10
<b>3</b>	<b>The LHCb experiment</b>	<b>12</b>
3.1	The Large Hadron Collider . . . . .	12
3.2	The LHCb detector . . . . .	12
3.2.1	Magnet . . . . .	12
3.2.2	Tracking system . . . . .	13
3.2.3	Particle identification . . . . .	16
3.2.4	Trigger . . . . .	19
3.2.5	Data processing software . . . . .	19
3.3	Simulated Monte Carlo samples . . . . .	20
3.4	B meson physics . . . . .	20
3.4.1	The decay $B^+ \rightarrow J/\Psi K^{*+}$ . . . . .	21
<b>4</b>	<b>Data analysis</b>	<b>22</b>
4.1	Background sources . . . . .	22
4.1.1	Combinatorial background . . . . .	22
4.1.2	Partially reconstructed background . . . . .	22
4.1.3	Peaking background . . . . .	22
4.2	Analysis strategy . . . . .	23
4.2.1	Fitting procedure . . . . .	24
4.3	Data preparation and selection variables . . . . .	24
4.3.1	Selection variables . . . . .	25
4.3.2	Data set . . . . .	27
4.4	Preselection . . . . .	29
4.5	Signal Monte Carlo calibration . . . . .	29
4.5.1	Unfolding a pure signal sample . . . . .	31
4.5.2	Monte Carlo reweighting . . . . .	32
4.6	Multivariate analysis . . . . .	34
4.6.1	BDT against combinatorial background . . . . .	36
4.7	Fit model and remaining background contributions . . . . .	37
<b>5</b>	<b>Determining the branching ratio</b>	<b>44</b>
5.1	Efficiencies . . . . .	44
5.2	Branching ratio results . . . . .	45
<b>6</b>	<b>Systematic uncertainties</b>	<b>47</b>
<b>7</b>	<b>Summary and outlook</b>	<b>49</b>

<b>A</b>	<b>Appendix</b>	<b>50</b>
A.1	Trigger configuration . . . . .	50
A.2	$J/\Psi K^+$ veto . . . . .	50
A.3	Fits of the normalisation channel . . . . .	51
<b>B</b>	<b>Bibliography</b>	<b>53</b>
<b>C</b>	<b>Acknowledgement</b>	<b>55</b>

# 1 Introduction

The field of particle physics aims to find a correct description of matter and its interactions. With the discovery of the Higgs boson in 2013, the last missing piece of the so-called Standard Model (SM) was found. However, although this model accurately describes all of the phenomena of particle physics observed in laboratory experiments, some open questions remain. For example the Standard Model assumes that neutrinos are massless, whereas recent experimental results show that they have a finite mass. Moreover, according to the theory of dark matter and dark energy, only about 4% of the energy of the observable universe is described by the Standard Model. Even though particles and antiparticles are always produced in pairs, our universe mostly consists of matter. The small amount of CP violation (CPV) incorporated into the SM, which leads to a different treatment of particles over antiparticles, cannot make up for a difference this profound.

With the intention of explaining these inconsistencies, the Large Hadron Collider (LHC) at CERN was built. The main goal of the LHC is the search for new particles, and therefore *Physics beyond the Standard Model* (also called New Physics). These particles can either be directly detected at high energies, or indirectly enter in quantum corrections. One of LHC's four major experiments, the LHCb experiment, focuses on the study of  $B$  hadron decays. Decays containing the heavy  $b$  quark are theoretically well described and very sensitive to quantum corrections. Significant deviations of measurements and theoretical predictions would be an unambiguous hint of New Physics. Until now, no deviations have been found. Therefore, effects of New Physics must be small. By measuring the branching fraction of the decay  $B^+ \rightarrow J/\Psi K^{*+}$ , the measurement of direct CP-violation in this decay is prepared. In the Standard Model, CPV is introduced via quantum loops. Thus, these kind of measurements are sensitive to potential New Physics. Additionally, the signal selection developed here can be used for an angular analysis of higher order decays, such as  $B^+ \rightarrow \mu^+ \mu^- K^{*+}$ . These types of analyses have the advantage that new observables can be constructed so that uncertainties of the SM predictions largely cancel. Therefore, influences of physics beyond the SM can be studied more accurately.

Hence, it is important to provide a good signal selection and branching fraction measurement of this decay. The current global average of this branching fraction, provided by the Particle Data Group (PDG, [1]), is  $\mathcal{B}(B^+ \rightarrow J/\Psi K^{*+}) = (1.43 \pm 0.08) \cdot 10^{-3}$ .

The dataset used in this analysis was collected at LHCb in the years 2011 and 2012 with center-of-mass energies of 7 TeV and 8 TeV respectively, corresponding to a total integrated luminosity of  $\mathcal{L}_{int} \approx 3 \text{ fb}^{-1}$ .

In order to understand this analysis, the basic theoretical concepts of particle physics as well as the experimental setup are introduced in Section 2 and Section 3, respectively. A detailed description of the analysis procedure is given in Section 4: Section 4.2 gives a brief overview of the analysis strategy, followed by an introduction to the selection variables and the used data sets in Section 4.3. After

making sure that simulated signal distributions agree with those from real data and correcting possible differences in Section 4.5, a multivariate classifier is trained to reduce combinatorial background in Section 4.6. The signal yield is extracted from a fit to the data in Section 4.7, which is then used to calculate the branching fraction in Section 5. After assigning systematic uncertainties in Section 6, the final result is given in Section 7.

## 2 Theoretical overview

This chapter gives an outline of the basic concepts of particle physics that are important to understand this thesis. It will give an overview of the Standard Model of particle physics (SM) with a focus on flavour physics and the weak interaction, since this is the force responsible for the decay of the  $B$  meson. A detailed introductory of particle physics or the SM can be found in [2] or [3].

### 2.1 The Standard Model of particle physics

The Standard Model is a Quantum Field Theory, which describes the fundamental particles and the forces between them. Particles are described as fields and the forces are described by the exchange of mediator particles. Altogether, there are 12 fermions, each having their own antiparticle with opposite electrical charge, 5 gauge bosons and the Higgs boson (see Fig. 2.1).

The group of fermions, which are spin- $\frac{1}{2}$  particles, can be further distinguished by their charges with respect to the three SM forces (electromagnetic, strong, weak). There are 6 quarks (up, down, charm, strange, top, bottom) and 6 leptons (electron, electron neutrino, muon, muon neutrino, tau, tau neutrino), where the type of each particle is called *flavour*. The forces acting between these fermions are mediated by the gauge bosons, which only couple to particles carrying the corresponding charge (see table 2.1).

**Strong interaction** The charge of the strong interaction is the color charge. There are three types: red, green blue and the corresponding anti-colors. While (anti-) quarks are charged with one (anti-) color, the force carrying particles, the so-called gluons, form an octet of states that always carry a combination of color and anti-color, e.g. red-anti-green, green-anti-blue or a superposition of color-anti-color states. They only couple to color charged objects. Therefore, only quarks and the gluons itself participate in the strong interaction. As gluons can interact with each other, the so-called phenomenon of *confinement* occurs. When two quarks are separated from each other, the potential energy between them rises, until it is energetically more favorable to create new quark-anti-quark pairs. Hence quarks only form bound states. These states have to be neutral<sup>1</sup> in color. They either consist of three (anti-) quarks, called (anti-) *baryon*, or of a quark-antiquark pair, called *meson*. Recently, even states with four or five quarks were found [5]. Particles consisting of quarks are called *hadrons*.

**Electromagnetic and weak interaction** The electromagnetic interaction is mediated by the massless photon and couples to all electrical charged objects, i.e. to all fermions except the neutrinos. It has an infinite range, although the strength

---

<sup>1</sup>Bound states with equal amounts of red, blue and green charge or color-anticolor states are called color neutral.

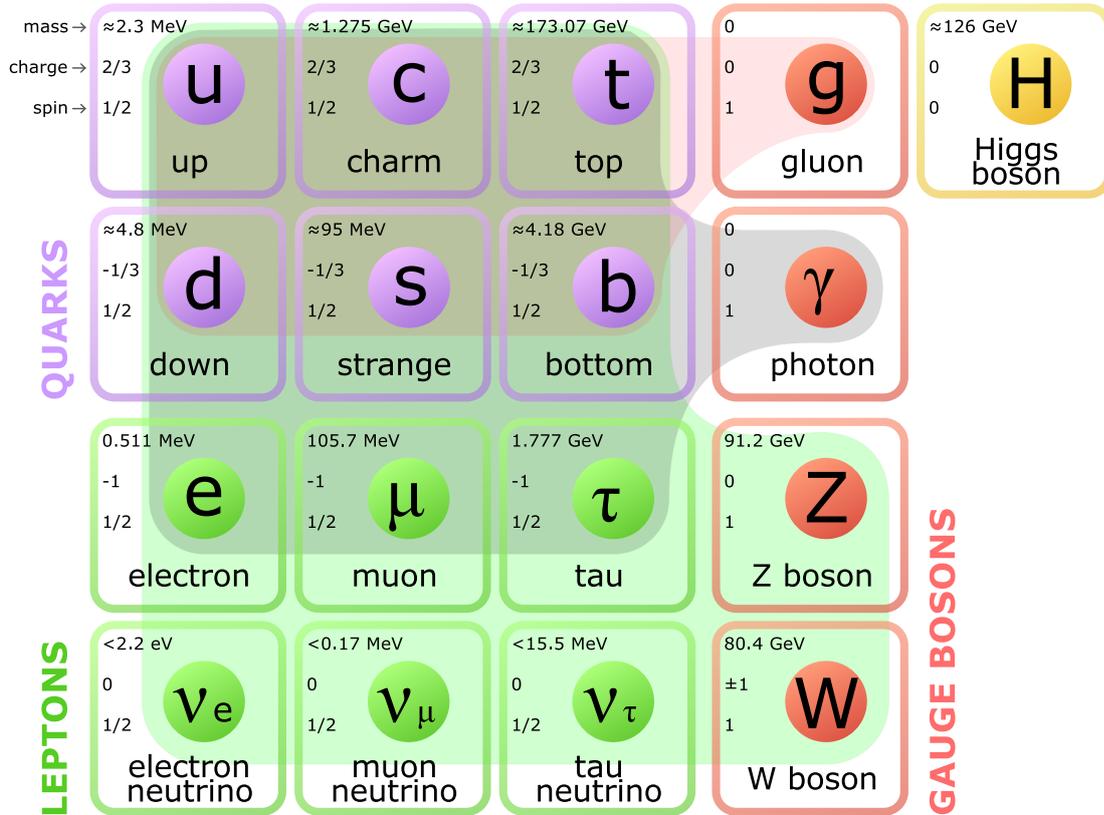


Figure 2.1: Elementary particles of the Standard Model and their interactions. Strong interaction in red, electromagnetic interaction in grey and weak interaction in green. Figure taken from Ref. [4]

decreases with distance.

In contrast to the strong and electromagnetic interaction, the mediator particles of the weak interaction ( $W^\pm, Z$ ) couple to all fermions and have a non-zero mass. This leads to highly suppressed interaction rates in comparison to the electromagnetic interaction for typical energies of a few GeV. For higher energies, both interactions have similar strength. Because of the  $W^\pm$  and  $Z$ 's high masses, they have a relatively short lifetime, ultimately leading to a short range of the weak interaction of about  $10^{-3}$  fm. The weak force is of special interest for the decay investigated in this thesis, as it is the only one that can change the flavour of particles. This will be discussed in more detail later.

It should be mentioned that electromagnetic and weak interaction can be unified to the so called electroweak interaction. As this differentiation is of no importance for this thesis, only the weak force will be discussed from now on. A unification of all forces is yet to be found.

The last missing component of the Standard Model is the Higgs boson. It is a spin and charge 0 excitation of the Higgs field and is responsible for the masses of the massive elementary fermions and gauge bosons of the SM.

interaction	relative strength	couples to	gauge boson	mass [GeV] <sup>2</sup>
strong	1	color charge	8 gluons (g)	0
electromagnetic	10 <sup>-3</sup>	electric charge	photon ( $\gamma$ )	0
weak	10 <sup>-8</sup>	weak charge	$W^\pm$	80.4
			$Z$	91.2

Table 2.1: List of the Standard Model forces. The relative strengths are approximate values for two fundamental particles at a distance  $d \approx 1$  fm. This table is taken from [2]

## 2.2 Flavour physics

In the weak interaction, quark flavor is not conserved. By coupling with a charged  $W^\pm$  boson, up-type<sup>3</sup> quarks can be transformed to down-type quarks and vice versa. This phenomenon can be described by the Cabibbo-Kobayashi-Maskawa (CKM) mechanism. As the weak eigenstates ( $d', s', b'$ ) are not equal to the mass eigenstates ( $d, s, b$ ), it is possible for quarks to change flavor. This behaviour can be described by the unitary  $V_{CKM}$  matrix

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{CKM} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$

The probability of a transition from mass eigenstate  $i$  to  $j$  is proportional to  $|V_{ij}|^2$ , where  $i$  is an up-type and  $j$  a down-type quark flavor. There are different ways to parameterize the CKM matrix, but as this is not of interest here, only the absolute values [2] are provided:

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} \approx \begin{pmatrix} 0.974 & 0.225 & 0.004 \\ 0.225 & 0.973 & 0.041 \\ 0.009 & 0.040 & 0.999 \end{pmatrix}$$

It can easily be seen that transitions within the same generation (diagonal elements) are more likely to happen. Still, it is possible to transform into a quark of another generation. In lowest order Feynman diagrams with only one vertex (tree level), transitions are realized by the emission of a  $W^\pm$  (Fig. 2.2). Since the mediator particle of this decay is charged, the initial and final state particles must have an electric charge difference of  $\Delta q = \pm 1$ . Therefore, only transitions from up- to down-type quark or vice versa happen. These processes are called *flavour changing charged currents*.

Transitions among up- or down-type quarks (*flavour changing neutral currents*, FCNC) can only occur when higher order Feynman diagrams are considered (see Fig. 2.3).

<sup>2</sup>In this thesis the convention  $c = \hbar = 1$  is used.

<sup>3</sup>Quarks with electrical charge  $+\frac{2}{3}$  ( $-\frac{1}{3}$ ) are referred to as up-type (down-type) quark.

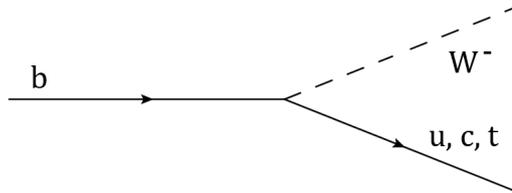


Figure 2.2: Flavour changing charged current: Emission of a  $W^-$ .

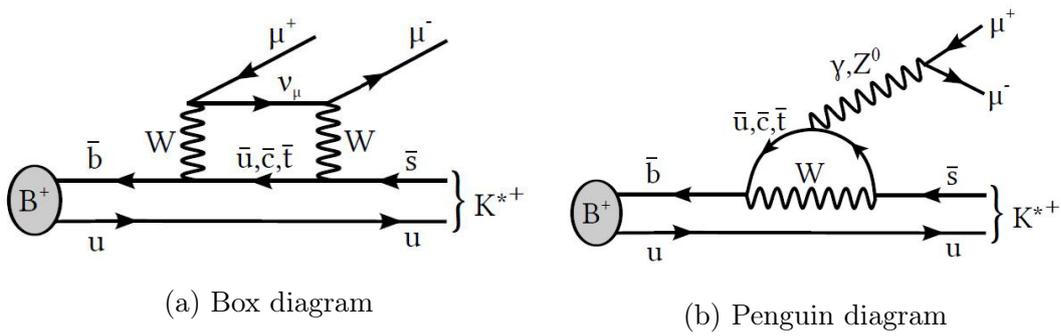


Figure 2.3: Feynman diagrams of the FCNC decay  $B^+ \rightarrow \mu^+ \mu^- K^{*+}$ , taken from [6].

## 3 The LHCb experiment

The Large Hadron Collider beauty (LHCb) experiment at the European Organization for Nuclear Research (CERN) near Geneva, Switzerland is one of the four major experiments of the Large Hadron Collider (LHC). It aims at finding indirect evidence of physics beyond the Standard Model in CP violation and rare decays of bottom and charm hadrons [7].

After a brief overview of the accelerator, the detector and its subsystems are discussed, followed by an introduction to  $B$  meson physics.

### 3.1 The Large Hadron Collider

The Large Hadron Collider is the world's most powerful particle accelerator. In two distinct vacuum pipes with 26.7km circumference, lying 40 to 170 meters deep underground, two proton bunches are accelerated in opposite direction and then brought to collision at one of four interaction points. The major experiments ATLAS, CMS, ALICE and LHCb are located there. ATLAS and CMS are general-purpose experiments, covering a large range of high energy physics topics. They are dedicated to the direct search for new physics. ALICE focuses on heavy-ion (i.e. Pb nuclei) collisions and LHCb is dedicated to flavor physics [8, 9].

The data collected at LHCb during Run I in 2011 and 2012 is used in this analysis. The collisions took place at center-of-mass energies of  $\sqrt{s} = 7$  TeV and  $\sqrt{s} = 8$  TeV, respectively. Since 2015, the accelerator is running at a center-of-mass energy of  $\sqrt{s} = 13$  TeV.

### 3.2 The LHCb detector

This section will give a brief overview of the LHCb detector and its main components. The following information is taken from [7].

In high energy proton-proton collisions, not the proton as a whole, but its constituents, the so-called partons (quarks and gluons), collide with each other. The momenta of both partons are different, leading to a strong forward boost of light particles such as the  $b$  quark. The LHCb detector exploits this fact. It covers an angle of approximately 10 mrad to 300 (250) mrad in the bending (non-bending) plane. Figure 3.1 shows a cross section of the detector along the bending plane. The chosen coordinate system is a right-handed system with the  $z$ -axis defined along the direction of the beam pipe into the detector (downstream),  $x$  horizontal and  $y$  vertical.

#### 3.2.1 Magnet

The magnet used at the LHCb experiment consists of two identical, water cooled, saddle shaped aluminium coils. They are mirror-symmetrically placed above and

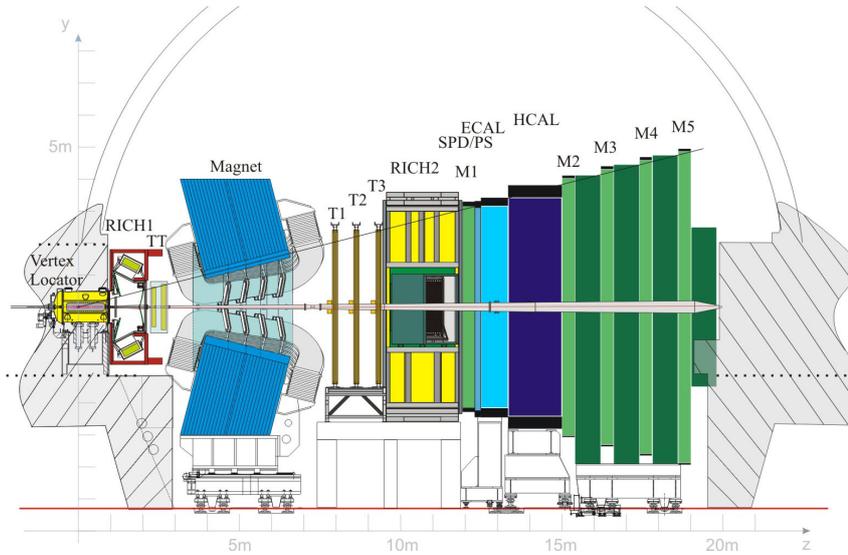


Figure 3.1: Cross section of the LHCb detector along the bending plane. Figure taken from [10].

below the beam pipe, tilted towards the interaction point (cf. Fig. 3.1). Its integrated magnetic field is approximately  $4\text{Tm}$  along the  $z$ -axis for tracks of  $10\text{m}$ . Charged particles traversing the magnet are bent mostly in the  $x$ - $z$ -plane, perpendicular to the magnetic field lines. The charged particles momentum  $p$  is determined by measuring the kink between the track before and behind the magnet. To reduce the influence of detector asymmetries, the polarity of the magnet, referred to as *Magnet Up* and *Magnet Down*, is reversed periodically.

### 3.2.2 Tracking system

The tracking system of LHCb is made up of several subsystems: The Vertex Locator (VELO) around the interaction point followed by the Trigger Tracker (TT) before the magnet and the three tracking stations (T1-T3) downstream of the magnet. Silicon microstrip detectors are used for VELO, TT and the inner sections of T1-T3, where the particle flux is high. The outer parts of the tracking stations use straw-tube detectors.

**Vertex Locator** A large number of particles are created directly at the point of the proton-proton collision. Many tracks intersect at this so-called *primary vertex* (PV). Long living particles such as the  $B$  mesons travel a distance before they decay (see also Sec. 3.4), which allows to distinguish between this decay vertex, called *secondary vertex* (SV), and the PV. The Vertex Locator (VELO) is designed to measure the position of primary and secondary vertices. It consists of 21 circular pairs of silicon modules placed around the beam pipe, measuring the distance to the beam line and the azimuthal angle in the  $x$ - $y$ -plane (Fig. 3.2). During the acceleration process, the modules are opened to prevent radiation damage. They are moved near the beam axis as soon as the beam is stable. The spatial hit resolution is around  $4\mu\text{m}$ .

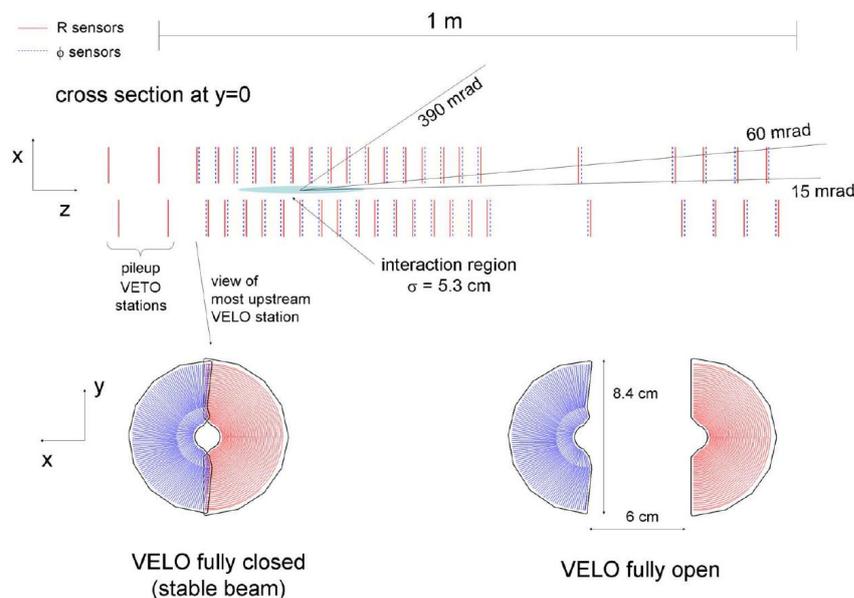


Figure 3.2: *Top*: Cross section in the  $x$ - $z$ -plane of the VELO. *Bottom*: Front view of a closed (left) and opened (right) module. Figure from [7].

**Silicon Tracker** The Trigger Tracker (TT) is located between the VELO and magnet and consists of two stations separated by 27 cm from each other, where each station is further split into four silicon microstrip detector layers. The first and last layer are arranged vertically, while the layers in between are tilted by  $\pm 5^\circ$  in the  $x$ - $y$ -plane (see Fig. 3.3). The full detector acceptance is covered by the TT. The center of the tracking stations T1-T3, where the track density is too high for the straw tube technology, is also covered with the silicon microstrip technology (Inner Tracker, Fig. 3.4). Each tracking station is made up of four silicon detector layers that are arranged analogously to the TT. Both TT and IT have a strip pitch of about  $200 \mu\text{m}$ , corresponding to a spatial resolution of about  $50 \mu\text{m}$ .

**Outer Tracker** The outer region of the tracking stations is covered with an array of gas-tight straw-tube modules. Two layers of drift-tubes (monolayer) with inner diameters of 4.9 mm make up each module (see Fig. 3.4). The tubes are filled with an Argon (70 %) and  $\text{CO}_2$  (30 %) mixture which reduces the drift time of ionized particles to below 50 ns. The time it takes an ionized particle to drift to the anode wire is used to determine the position of charged particles traversing the tube. A spatial resolution of about 200 nm is achieved.

Similar to the TT and IT, each station contains four monolayers of which the two inner layers are tilted by  $\pm 5^\circ$ .

A schematic view of the tracking system is given in Fig. 3.4.

**Reconstruction** Trajectories of charged particles traversing the detector are reconstructed from hits in the previously introduced detectors. This analysis uses only long tracks which require at least hits in the VELO as well as in the tracking stations T1-T3 (T stations). First, the VELO is searched for lines that can be

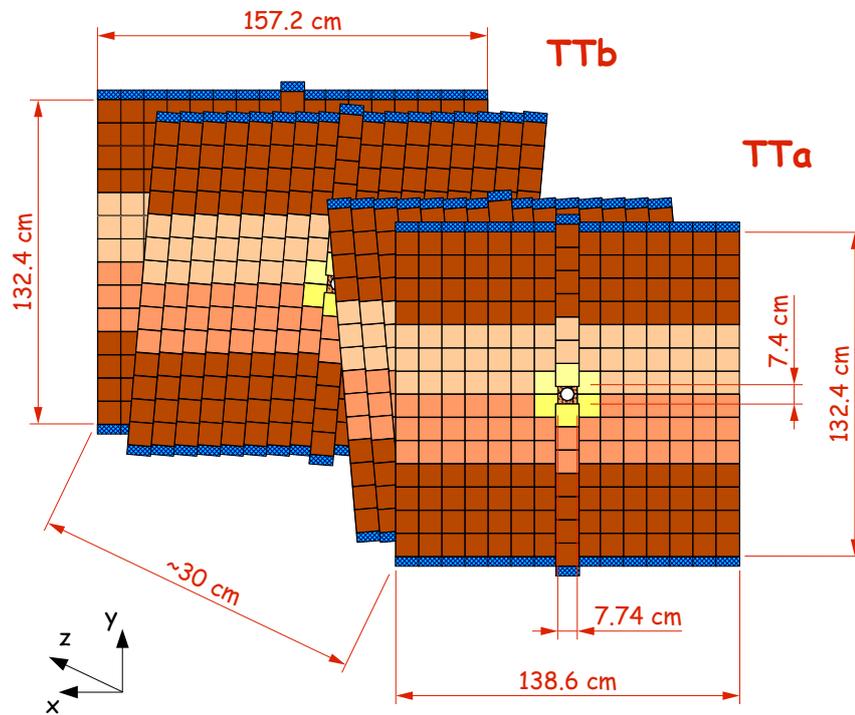


Figure 3.3: Layout of the TT layers. Taken from [11].

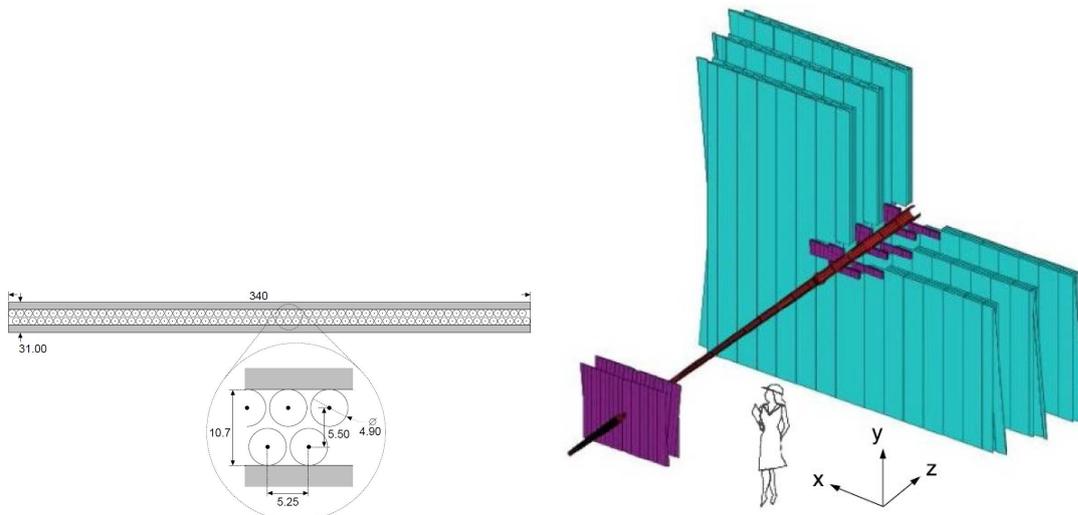


Figure 3.4: *Left*: Cross section of a monolayer consisting of two layers of straw tubes. *Right*: Schematic view of the TT and the Tracking stations T1-T3. Silicon tracker in purple, OT in turquoise. Figures from [7].

combined with hits in the T stations. If possible, hits of the TT are also included to improve the precision of the momentum estimation. After finding the tracks, a so-called Kalman filter is used to fit the trajectories while taking multiple scattering and energy loss through ionization into account.

### 3.2.3 Particle identification

To obtain full information about the particles, their type has to be identified. A mass hypothesis for final state particles can be claimed with information from the particle identification (PID) system: Two Ring Imaging Cherenkov (RICH) detectors are used to determine the species of charged hadrons, two calorimeter systems distinguish between electrons, photons and hadrons and several muon stations are used to identify particles as muons.

A likelihood is calculated for every possible particle hypothesis. Due to the large amount of charged pions at LHC, the likelihoods are divided by the likelihood of the pion hypothesis. They are denoted as  $PID_{X\pi}$ , where  $X$  can either be a proton, kaon, electron or muon. To smooth the distributions of the calculated likelihoods, the logarithm is taken.

In the following, the subsystems of the PID system are introduced.

**RICH** The RICH system is, in combination with the tracking stations, able to determine the masses of charged hadrons ( $K, \pi, p$ ). This is especially important to distinguish between kaons and pions, which are produced abundantly in proton-proton collisions. When charged particles travel with a velocity  $v$  through a medium with refraction index  $n$ , so that the particle is faster than the speed of light  $c' = c/n$  in that medium, photons are emitted at a Cherenkov angle  $\theta = \arccos(c/(vn))$ . This allows to determine the velocity, and with the momentum estimation from the tracking system, the mass of the particle.

Because the difference in Cherenkov angles reduces for high momentum (see Fig. 3.5), two detectors with different radiators are used. RICH1 is located upstream of the magnet and filled with silica aerogel ( $n = 1.03$ ) and gaseous  $C_4F_{10}$  ( $n = 1.0014$ ), covering a momentum region of 2-40 GeV over the full detector acceptance. RICH2 is placed downstream of the magnet, after the tracking stations. It uses gaseous  $CF_4$  ( $n = 1.0005$ ) to distinguish particles with momenta between 15-100 GeV. As high momentum particles have a higher boost and are deflected less by the magnetic field, a smaller geometric acceptance is covered.

**Calorimeter** The calorimeter system is composed of an electromagnetic calorimeter (ECAL) in front of a hadronic calorimeter (HCAL). Each calorimeter is split into many cells of scintillating material. If a charged particle traverses through the calorimeter, they interact in such a way that particle showers are formed, creating scintillation light. Photomultiplier tubes convert this light to a voltage that contains information about the amount of energy the particle disposed in the corresponding cell. Electrons and photons produce electromagnetic showers via bremsstrahlung and pair production. Hadrons with high energy create hadronic showers that also create photons in the scintillating material. Calorimeters are

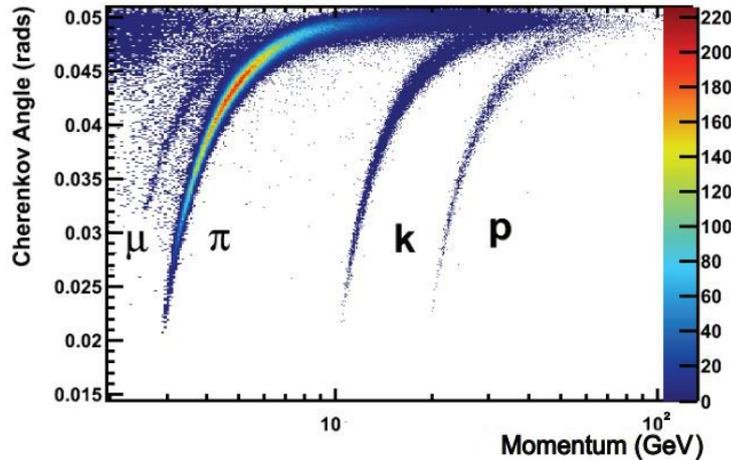


Figure 3.5: Reconstructed Cherenkov angle over track momentum in RICH1. Taken from [13]

divided into more cells near the beam pipe to allow for sufficient resolution in the regions of higher particle density.

A Scintillating Pad Detector (SPD) is placed in front of the calorimeters. It consists of a small scintillating plate which provides a trigger signal if a charged particle traverses it. Hence, background from neutral particles is reduced.

The PreShower detector (PS) consists of two layers of scintillating pads with a 15 mm lead wall in between, corresponding to 2.5 electron radiation lengths. Since the interaction length of hadrons is much higher, most of the showers in this thin layer of lead are caused by electrons. Combined with a fine granularity, a good distinction between electrons and hadrons as well as a good spatial resolution is achieved.

The ECAL is composed of alternating layers of absorption and scintillating material, corresponding to 25 radiation lengths. That way, electrons and photons deposit all of their energy in the ECAL. The Energy resolution is  $\sigma_E/E = \frac{10\%}{\sqrt{E}} \oplus 1\%$  for energies in the GeV scale. The first term describes statistical fluctuations whereas the second term describes systematic uncertainties such as incorrect calibration.

The HCAL is build similar as the ECAL, but has a thickness of 5.6 interaction lengths because of spatial limitations. Hadronic showers are more complex than electromagnetic showers, which leads to higher fluctuations in the deposited energy. This leads to an energy resolution of  $\sigma_E/E = \frac{(69 \pm 5)\%}{\sqrt{E}} \oplus (9 \pm 2)\%$  ( $E$  in GeV).

**Reconstruction of  $\pi^0$**  Neutral pions decay into two photons with a branching ratio of  $\mathcal{B} = 98.8\%$ . These photons are reconstructed in the ECAL, where they produce electromagnetic showers. A cluster algorithm uses the energy deposit and shower shape from neutral clusters to determine energy and direction of the photons [14].

Low and high  $p_T$  neutral pions are reconstructed differently. Photons emerging

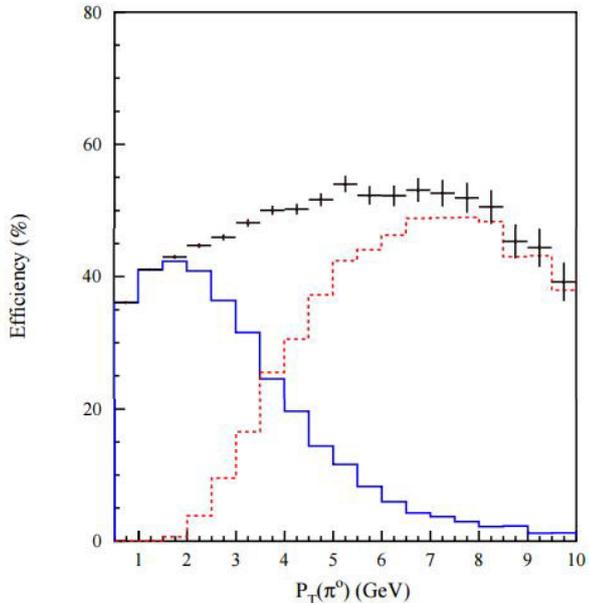


Figure 3.6: Reconstruction efficiency of resolved (blue) and merged (red)  $\pi^0$  over transverse momentum. Total efficiency in black. Taken from [14].

from pions with  $p_T < 2$  GeV are mostly reconstructed from two separate clusters in the ECAL (resolved  $\pi^0$ ). Since there are many  $\pi^0$ s at LHCb, many different photons can be combined. Therefore, it is looped over all possible photon-photon combinations. Only photon pairs with an invariant mass in the  $\pi^0$  mass region  $m(\gamma\gamma) \in [105, 165]$  MeV are combined to pion candidates. However, only photons with transverse momentum larger than 200 MeV are kept to reduce combinatorial background and wrongly combined  $\pi^0$ s. This also reduces the reconstruction efficiency for pions with  $p_T < 1$  GeV (see Fig. 3.6), since it then is likely that one photon will not pass the threshold.

For high momentum pions the two photons have a higher boost along the  $z$ -axis, which often leads to overlapping ECAL clusters (merged  $\pi^0$ ). Another algorithm is used to disentangle the two photons in shared clusters. If the invariant mass of these two photons lies inside the pion mass window, they are combined to a  $\pi^0$  candidate. As this algorithm is not as accurate as for resolved  $\pi^0$ s, and pions are more likely to have smaller transverse momenta for  $B^+ \rightarrow J/\Psi K^{*+} (\rightarrow K^+ \pi^0)$  decays, only resolved  $\pi^0$ s were used for this analysis.

**Muon System** Muons are the only particles that easily pass through the calorimeter system. This is due to their low energy loss per distance travelled in a material ( $dE/dx$ ). They are also produced in many interesting  $B$  decays. Therefore, LHCb uses an efficient system for muon identification. In total, there are five muon stations (M1-M5). M1 is placed upstream of the calorimeter system and used for trigger purposes (see Sec. 3.2.4). M2-M5 are placed downstream of the calorimeter system (see Fig. 3.1). An 80 cm thick iron absorber is placed between each station to stop hadrons, which have higher  $dE/dx$  than muons, that travelled through the HCAL. Particles that traverse the muon stations can then be identified as muons. While M1-M3 have high spatial resolutions along the  $x$ -axis to improve the  $p_T$  resolution, M4 and M5 are only used for identification purposes. The selection efficiency is approximately 95%. Less than 2% of other particles are misidentified

as muons.

All muon stations use Multiwire Proportional Chambers (MWPC) to detect ionizing particles. Only the inner region of M1, where the particle flux is too high, uses a triple-GEM (gas electron multiplier) technology [7].

### 3.2.4 Trigger

At LHCb, the frequency of bunch crossings with events in the detector acceptance is around 10 MHz. The trigger system reduces the event rate to about 5 kHz, which is low enough for the data to be written to storage. A low level hardware trigger (Level 0, L0) and high level software triggers (HLT1, HLT2) are used to pick physically interesting events. The trigger configuration of this analysis can be found in appendix A.1.

**Level 0 (L0)** The L0 trigger is a pure hardware trigger with custom electronics. It uses information from the calorimeters and muon stations, which are fast to read out. Particles from heavy  $B$  decays have high transverse momenta. Therefore the highest  $p_T$  muon from the muon stations and highest  $E_T$  particle from the calorimeters are required to pass a certain threshold. This reduces the rate from 10 MHz to 1 MHz.

**High Level Trigger (HLT)** Events that passed L0 are further analysed by the HLT algorithm, which runs on a farm of 1000 16-core computers. Stage 1 uses information of VELO and tracking stations only. Candidates are required to emerge from a detached vertex or to have high momentum. This reduces the event rate to the order of 10 kHz. The rate is now low enough for the second stage of the HLT. Techniques similar to those used in offline analyses combined with information from all detector subsystems reduce the rate to around 5 kHz.

### 3.2.5 Data processing software

LHCb uses many different software packages to make the raw data of the detector accessible for end users. The high level trigger is handled by the `Moore` software [15]. The triggered, raw data is then given to the `Brunel` application [16] which creates objects such as tracks from the detector hits. At this stage, an analysis would theoretically be possible but is not effective due to computational restrictions. A so-called *stripping* (see also Sec. 4.3.2) is applied to reduce the data to a manageable amount and pick interesting events only. As soon as a hypothesis on the decay chain is made, in this case  $B^+ \rightarrow J/\Psi(\rightarrow \mu^+\mu^-)K^{*+}(\rightarrow K^+\pi^0(\rightarrow \gamma\gamma))^4$ , the so-called *Decay Tree Fitter Tool* is used (see Ref. [12]). It performs a fit on the whole decay chain, taking four-momenta, vertex positions and possible correlations or external constraints (e.g. mass constraints for daughter particles) into account. For this analysis, the masses of  $\pi^0$  and  $J/\Psi$  candidates are set to their PDG values. Additionally, the flight direction of the B-candidate is constraint to the primary vertex. This is done by the `DaVinci` software [17]. The remaining data is then analysed with software such as `ROOT`.

---

<sup>4</sup>The decays in brackets indicate the subsequent decays of the  $J/\Psi$ ,  $K^{*+}$  and  $\pi^0$ , respectively.

### 3.3 Simulated Monte Carlo samples

Simulated data samples are an important part of many analyses. In general, the distributions of a decay are unknown and not visible in unselected, real data. However, knowledge about the distributions is required to be able to effectively select events of the decay of interest. Additionally, the number of signal events in data samples is not known. This makes it impossible to obtain the selection efficiencies, which are essential to determine the branching ratio. Since simulated samples only consist of the decay channel, both of these problems are solved. Nonetheless, events can be incorrectly reconstructed by the detector. Here, the big advantage of simulated samples comes into play: the so-called truthmatching. In every step of the simulation, the true identity of a particle is known. This means that even if particles are misidentified by the detector, the true signal distribution can be obtained by looking only at correctly identified events.

The LHCb `Gauss` application supports many so-called Monte Carlo (MC) generators and manages different packages to simulate the generated events. For the simulations used in this analysis, `Pythia 8` was used to simulate  $p$ - $p$  collisions. The decay of the simulated particles is then managed by `EvtGen`. To propagate the particles through and simulate their interactions with the detector the `Geant4` software is used. Documentations of the `Gauss` project and its packages can be found on the web [18]. After the simulated events are propagated through the detector, the same software is used for trigger, reconstruction and stripping as for real data (see Sec. 3.2.5).

In this analysis, several simulated MC samples are used:

- A simulated  $B^+ \rightarrow J/\Psi K^{*+} (\rightarrow K^+ \pi^0 (\rightarrow \gamma\gamma))$  sample is investigated to find efficient preselection cuts (Sec. 4.4). Additionally, it is used as signal proxy for a multivariate classifier (Sec. 4.6) and to determine the selection efficiencies (Sec. 5.1).
- Simulated  $B^0 \rightarrow J/\Psi K^{*0} (\rightarrow K^+ \pi^-)$  events are used to determine the selection efficiencies of the normalization channel.
- To find possible background contributions, a simulated  $B^+ \rightarrow J/\Psi X$  sample is investigated, where  $X$  can be any allowed decay product.
- Several fit models are tested to describe possible background shapes by using simulated  $B^+ \rightarrow J/\Psi K_1 (\rightarrow K^{*+} \pi^0)$  and  $B^0 \rightarrow J/\Psi K^{*0} (\rightarrow K^+ \pi^-)$  samples.

### 3.4 B meson physics

B meson physics offers a lot of advantages for studying possible physics beyond the SM. Studies of CP violation in decay ( $\mathcal{B}(B^+ \rightarrow J/\Psi K^{*+}) \neq \mathcal{B}(B^- \rightarrow J/\Psi K^{*-})$ ), in mixing ( $\mathcal{B}(B^0 \rightarrow \bar{B}^0) \neq \mathcal{B}(\bar{B}^0 \rightarrow B^0)$ ) and in interference between decay and mixing ( $\mathcal{B}(B^0 \rightarrow f) \neq \mathcal{B}(B^0 \rightarrow \bar{B}^0 \rightarrow f)$ ) are possible. Moreover, there are many loop-induced decays, e.g. FCNC, in which flavour physics beyond the SM could enter. Since the amplitudes of weak interactions are proportional to the  $V_{CKM}$

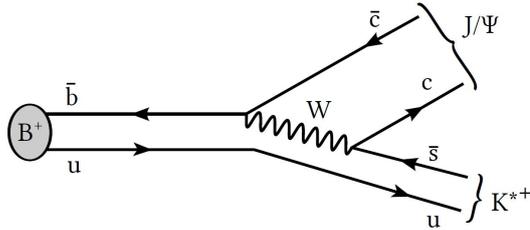


Figure 3.7: Feynman diagram of the decay  $B^+ \rightarrow J/\Psi K^{*+}$  [6].

matrix elements, the coupling constants and  $\frac{m_q}{m_W}$ , where  $m_q$  is the mass of the virtual quark, decays with virtual heavy  $t$  quarks dominate. Additionally, the corresponding matrix element is nearly unity ( $|V_{tb}| \approx 0.999$ ). This leads to a large amount of higher order  $B$  decays. With a high  $b\bar{b}$  production rate of  $\sim 10^{11}$  pairs per  $1 \text{ fb}^{-1}$ , high statistics are available for analyses.

Furthermore, the  $B$  meson lives long enough to travel a short distance from the collision point (PV). This makes it possible to distinguish between particles that are produced at the PV, so-called background, and final state particles that originate from the point of the  $B$  decay. Such a clean experimental signature is easy to select.

As  $B$  mesons consist of a heavy  $b$  and a light up-type quark (a so-called spectator quark), the up-type quark can be neglected in theoretical computations. Therefore, precise predictions of observables can be made. Deviations from these predictions would indicate New Physics.

### 3.4.1 The decay $B^+ \rightarrow J/\Psi K^{*+}$

This section will briefly introduce the here investigated decay channel  $B^+ \rightarrow J/\Psi(\rightarrow \mu^+\mu^-)K^{*+}(\rightarrow K^+\pi^0)$ , where the  $\pi^0$  decays into two photons. The corresponding Feynman diagram is shown in Fig. 3.7. Both the  $J/\Psi$  and  $K^{*+}$  decay almost instantly, which makes it impossible to spatially resolve the travelled distance. Effectively, the muons and kaons emerge from the same vertex. As already discussed in Sec. 3.2, the muons leave a clean signal in the detector.

The  $\pi^0$  is not visible in the tracking system. Instead, the energy deposit of the  $\pi^0 \rightarrow \gamma\gamma$  decay in the ECAL (see Sec. 3.2.3) is measured to reconstruct the  $\pi^0$  energy and momentum. As this cannot provide high accuracy, the reconstruction efficiency of the  $K^{*+}$  suffers. Therefore, a proper selection is required.

Measuring the branching ratio of  $B^+ \rightarrow J/\Psi K^{*+}$  sets up a direct CP violation measurement:

$$A_{CP} = \frac{\mathcal{B}(B^+ \rightarrow J/\Psi K^{*+}) - \mathcal{B}(B^- \rightarrow J/\Psi K^{*-})}{\mathcal{B}(B^+ \rightarrow J/\Psi K^{*+}) + \mathcal{B}(B^- \rightarrow J/\Psi K^{*-})}. \quad (3.1)$$

## 4 Data analysis

The following chapter gives an outline of the analysis as well as an introduction to the used datasets and variables which describe the decay, followed by a detailed description of the signal selection. First, loose cuts are applied to reduce some of the combinatorial background. The agreement of topological and kinematical observables in simulated and real decays is verified using a more abundant control channel. Afterwards a multivariate classifier is used to reduce combinatorial background even further.

### 4.1 Background sources

In order to obtain the branching fraction, signal and background distributions have to be described accurately. Different sources of background events exist, which will be explained in this section. Contributing decay channels and fitting shapes are explained in Sec. 4.7. A good set of selection variables is introduced in Sec. 4.3.1.

#### 4.1.1 Combinatorial background

$B$  candidates that are combined from random particles that do not share an ancestor are called combinatorial background. These particles are mostly produced at the primary vertex. Combinatorial background is featureless and can be described with an exponential or polynomial function.

#### 4.1.2 Partially reconstructed background

Partially reconstructed background is formed by events similar to  $B^+ \rightarrow J/\Psi K^{*+} (\rightarrow K^+ \pi^0)$ , where an additional particle such as a  $\pi$  or  $\gamma$  is missed in the reconstruction. It then looks identical to the signal decay channel. Because of the missing particle, the masses of these candidates are shifted to lower regions. An example for this is the decay  $B^+ \rightarrow J/\Psi K_1^+$ , where the  $K_1^+$  decays to  $K^{*+} \pi^0$ , but the  $\pi^0$  is not reconstructed.

It is important to describe these contributions, even if they lie outside of the signal region, because they still influence the shape for combinatorial background and therefore have a direct influence on the signal yield.

#### 4.1.3 Peaking background

If particles are misidentified by the detector, the possibility exists that they look exactly like the signal channel. This mostly occurs when kaons are misidentified as pions or vice versa. These decays then form a peaking contribution in the invariant mass spectrum of the  $B$  candidate. For example, the decay  $B^+ \rightarrow J/\Psi \rho^+ (\rightarrow \pi^+ \pi^0)$ , where the  $\pi^+$  is misidentified as  $K^+$ , looks like a  $B^+ \rightarrow J/\Psi K^{*+} (\rightarrow K^+ \pi^0)$  decay. As this has a branching fraction that is two orders of magnitude smaller than

the signal channel, contributions from this channel could be neglected. However, a significant contribution could be found from  $B^0 \rightarrow J/\Psi K^{*0} (\rightarrow K^+ \pi^-)$ , where the  $\pi^-$  is not reconstructed and instead a random  $\pi^0$  is added (see also Sec. 4.7).

## 4.2 Analysis strategy

The goal of this thesis is to determine the branching fraction of

$$B^+ \rightarrow J/\Psi K^{*+}$$

relative to the branching ratio of the decay

$$B^0 \rightarrow J/\Psi K^{*0},$$

which is further referred to as *control channel* or *normalisation channel*<sup>5</sup>. To determine the ratio, decays with  $J/\Psi \rightarrow \mu^+ \mu^-$ ,  $K^{*+} \rightarrow K^+ \pi^0 (\rightarrow \gamma \gamma)$  and  $K^{*0} \rightarrow K^+ \pi^-$  are used. The control channel is chosen for its similar kinematic properties as the signal channel. Taking the ratio then leads to a cancellation of systematic uncertainties, of which the muon efficiencies are an example. Moreover, all final state particles of the normalisation channel are charged, which means they give a clean signal in the detector. By extracting a signal sample (see Sec. 4.5.1) in data, simulated events can be compared to real data. The simulation can be corrected to the data if necessary. Because of the well known branching ratio of the normalisation channel, which is measured with a relative uncertainty of 4.5%, it is used to calculate the absolute branching fraction of the signal channel  $B^+ \rightarrow J/\Psi K^{*+}$ .

The analysis is executed in the following steps:

1. Data preparation. The data is recorded at the LHCb-experiment. A loose selection, the so-called *stripping*, is already applied centrally by the collaboration to reduce the amount of data.
2. Preselection. Loose cuts on different kinematic variables, as well as trigger decisions, are applied to reduce background. Cuts on *PID*-variables are used to reduce peaking background from misidentified particles (Sec. 4.4).
3. Comparison of simulation and data. It is checked if distributions in the simulation agree with the data and if any corrections are needed (Sec. 4.5).
4. Multivariate Analysis. To suppress combinatorial background even further, a *Boosted Decision Tree* (BDT) is trained with data from the upper sideband<sup>6</sup> of the reconstructed  $B^+$  mass as background sample and truthmatched MC samples as signal sample (Sec. 4.6).

---

<sup>5</sup>Control and normalisation channel are not synonyms. The control channel is used to compare the signal channels distributions to clean, "real" distributions, while the normalisation channel is used to determine the branching fraction. In this analysis, the same channel is used for both.

<sup>6</sup>Events with  $B$  masses above the signal window are called upper sideband. In this thesis, it refers to  $m(B^+) > 5700$  MeV. Analogously, the lower sideband can be defined.

5. Signal fit. After applying all selections, a fit is performed on the mass distributions of signal and control channel to obtain the signal yields (Sec. 4.7).
6. Efficiency determination. The efficiencies of each selection step are determined by using simulated samples (Sec. 5.1).
7. Determining the branching ratio. Using the previously determined yields and efficiencies, the branching ratio could be obtained by

$$\mathcal{B}_{sig} = \frac{\mathcal{B}_{norm}}{\mathcal{B}_{corr}} \times \frac{N_{sig}}{N_{norm}} \times \frac{\xi_{norm}}{\xi_{sig}} \times \frac{\epsilon_{norm}^{reco,presele}}{\epsilon_{sig}^{reco,presele} \cdot \epsilon_{sig}^{BDT}}$$

where  $\mathcal{B}_{sig} = \mathcal{B}(B^+ \rightarrow J/\Psi K^{*+})$  is the branching fraction of interest,  $\mathcal{B}_{norm} = \mathcal{B}(B^0 \rightarrow J/\Psi K^{*0}(\rightarrow K^+\pi^-))$  is the branching fraction of the normalisation channel and  $\mathcal{B}_{corr} = \mathcal{B}(K^{*+} \rightarrow K^+\pi^0) \cdot \mathcal{B}(\pi^0 \rightarrow \gamma\gamma)$  accounts for the rest of the decay chain used in this analysis. The previously determined efficiencies are denoted by  $\epsilon_i^{reco,presele}$  for the reconstruction and preselection efficiency, where  $i = sig, norm$  denotes the signal and normalisation channel, and by  $\epsilon^{BDT}$  for the used BDT cut. The geometrical acceptances of the detector are denoted by  $\xi_i$ . In particle collisions, the decay products are emitted in all directions, whereas the detector only covers a small region. Therefore, only a small fraction of the produced particles can be detected. A more detailed description is given in section 5.2.

### 4.2.1 Fitting procedure

The *unbinned maximum likelihood method* is used for all fits in this thesis. For a measured dataset  $\vec{x}_i$  and a given *probability density function* (PDF)  $f(\vec{x}_i|\vec{a})$ , the likelihood function is defined as

$$\mathcal{L}(\vec{a}) = \prod_{i=1}^N f(\vec{x}_i|\vec{a})$$

where  $N$  is the number of events and  $\vec{a}$  is a vector of unknown parameters. For a certain  $\vec{a}$ , and therefore a chosen probability distribution model, the likelihood function expresses the probability to obtain the given dataset  $\vec{x}_i$ . The best estimate of  $\vec{a}$  is found by maximising  $\mathcal{L}$ . Further information can be found in [19].

In order to reduce the computational effort, the negative log likelihood function  $-\log(\mathcal{L}) = -\sum_{i=1}^N \log(f(\vec{x}_i|\vec{a}))$  is minimized instead.

## 4.3 Data preparation and selection variables

This section gives an overview of the used dataset and its preparation. All variables used in this analysis are explained.

### 4.3.1 Selection variables

A commonly used variable in particle physics is the four-momentum  $p = (E, p_x, p_y, p_z)^T$ . It is a conserved quantity and the product of two four-momenta  $p_1 \cdot p_2 = E_1 \cdot E_2 - \sum_i p_{1,i} \cdot p_{2,i}$  is invariant under Lorentz transformation, where  $i = x, y, z$ . Many of the variables used here are derived from the measured four-momenta of the final state particles  $K^+, \mu^+, \mu^-, \gamma_1, \gamma_2$ . To obtain the four-momentum of a mother-particle<sup>7</sup>, all four-momenta of the decay products are added. A description of how the four-momenta are measured is given in Sec. 3.2

Other important variables to describe the decay and discriminate it from background originate from the particle identification system and reconstruction procedure.

In the following, a short explanation of the used variables is given:

**Mass ( $m$ )** The square of the invariant mass of a particle, for example of the  $B$  meson, can be calculated by taking the square of its four-momentum  $p$ :

$$m^2 = p^2 \Rightarrow m = \sqrt{p^2}.$$

**Transverse momentum ( $p_T$ )** The momentum component transversal to the  $z$ -axis (the direction of the beam pipe) can be computed in the following way

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

For decay products of heavy particles like the  $B$  meson, the transverse momentum is usually large. This can be exploited to separate combinatorial background from signal events. The used coordinate system is introduced in section 3.2.

**Pseudorapidity ( $\eta$ )** relates to the angle  $\theta$  between the beam axis and the particles momentum:

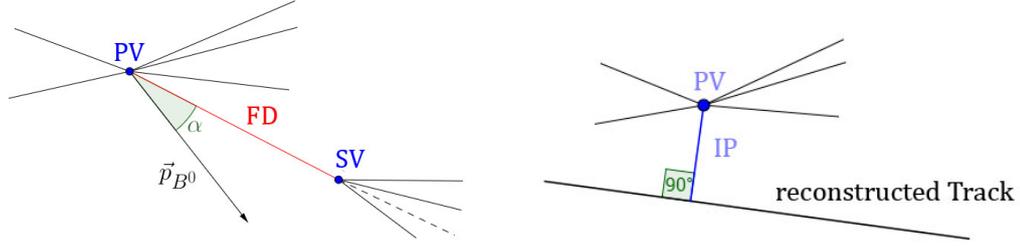
$$\eta = -\ln \left( \tan \frac{\theta}{2} \right) = \operatorname{artanh} \left( \frac{p_z}{|\vec{p}|} \right).$$

**Charged cone  $p_T$  asymmetry ( $A_{p_T}^{cc}$ )** By spanning a cone around the trajectory of the  $B^+$  candidate with its tip at the PV, an asymmetry can be calculated by

$$A_{p_T}^{cc} = \frac{p_{T,B^+ \text{ cand.}} - \sum p_{T, \text{ other-charged-tracks}}}{p_{T,B^+ \text{ cand.}} + \sum p_{T, \text{ other-charged-tracks}}}$$

where  $\sum p_{T, \text{ other-charged-tracks}}$  takes all transverse momenta of charged particles into account that do not emerge from the  $B$  decay and lie inside the cone. Therefore, if only reconstructed tracks lie inside the cone,  $A_{p_T}^{cc}$  becomes unity. In proton-proton collisions, a vast amount of background is produced at the primary vertex. Most of these particles have low masses, and therefore low  $p_T$ . This leads to a high spatial density of tracks for low angles. Therefore it is more likely that background events have additional tracks inside of the cone, which leads to a smaller  $A_{p_T}^{cc}$ . Hence, this variable can be used to discriminate signal and background events.

<sup>7</sup>A decaying particle is called mother-particle, while its decay products are called daughter-particles



(a) Flight distance ( $FD$ ) and direction angle ( $DIRA$ ).

(b) Impact parameter ( $IP$ ).

Figure 4.1: Visualisation of the variables  $FD$ ,  $DIRA = \cos(\alpha)$  and  $IP$ .

**Direction Angle ( $DIRA$ )** The cosine of the angle  $\alpha$  between the reconstructed momentum of the  $B^+$  candidate and the vector connecting PV and SV is called Direction Angle (Fig. 4.1a). If particles are missing or are randomly combined in the reconstruction, this angle tends towards larger values, whereas well reconstructed signal candidates have rather small angles and therefore  $DIRA \approx 1$ .

**Flight Distance ( $FD$ )** The flight distance is the distance between primary vertex and the reconstructed decay vertex of the  $B^+$  (Fig. 4.1a).

**Impact Parameter ( $IP$ )** The impact parameter is defined as the minimal distance between the PV and the trajectory of the considered particle (Fig. 4.1b). Therefore, this value should be close to zero for  $B$  candidates and big for daughter-particles as the  $B^+$  travels some distance before decaying. Combinatorial background is produced in many places, so  $IP$  will have very different values. The  $\chi^2$  of the impact parameter measures the difference of  $\chi^2$  of the PV-fit before and after the considered track is added. The values behave like  $IP^2/\sigma^2$ , where  $\sigma$  denotes the uncertainty.

**Distance of closest approach ( $DOCA$ )** The distance of closest approach measures the minimal distance of two given tracks. The variable  $\chi^2(DOCA)$  measures the likelihood of both fitted tracks under the assumption that the two tracks cross each other. It behaves almost like  $\chi^2(DOCA) = DOCA^2/\sigma^2(DOCA)$ .

**Decay Length Significance ( $DLS$ )** is a measure of the separation power of a particle's decay and creation vertex:

$$DLS = \frac{FD}{\sigma(FD)}.$$

**Vertex- $\chi^2$**  measures the quality of the vertex reconstruction. For a good reconstruction, the  $\chi^2$  per *number of degrees of freedom* ( $ndof$ ) is close to 1 for large  $ndof$ , while bad reconstructions tend to have higher values.

The minimal  $\chi^2$ -difference when an additional track is added to the already reconstructed track is called  $\Delta\chi^2_{add-track}$ . This value is small when additional charged

particles are near the decay vertex. In that case it is likely that the additional particle originates from the same decay. Hence, this variable discriminates between signal and partially reconstructed background.

**Track- $\chi^2$**  This variable measures the quality of the reconstruction of a track.

**Ghost probability (*ghostprob*)** Tracks that do not come from a particle or are combined from hits of several particles are called *ghosts*. The probability that a track actually is a ghost is denoted by *ghostprob*.

***hasRICH*** is a boolean variable that denotes if the RICH subsystem registered a track or not.

***isMuon*** is a boolean value. When a muon candidate hits the first muon station, a *Field of Interest (FOI)* is extrapolated. The candidate is required to hit at least two or three more muon stations in the FOI, if it has low or high momentum, respectively [20].

**Confidence Level (*CL*)** The confidence that a neutral particle's ID is correctly assigned is given by *CL*. It is calculated by taking the ratio of the likelihood of a particle hypothesis and the sum of the likelihoods of all hypotheses (see also Sec. 3.2.3).

The confidence level of the  $\pi^0$  is calculated by combining both CL from its daughter particles, i.e. the photons.

### 4.3.2 Data set

The dataset collected at LHCb in 2011 and 2012 with a center-of-mass energy of  $\sqrt{s} = 7 \text{ TeV}$  and  $\sqrt{s} = 8 \text{ TeV}$ , respectively, is used in this analysis. Both Magnet polarities (Up and Down) are used, corresponding to a total integrated Luminosity of  $\approx 3 \text{ fb}^{-1}$ . Because merged  $\pi^0$ s have a worse resolution (see Sec. 3.2.3) and contribute far less statistics, only resolved  $\pi^0$ s are used.

Many different trigger configurations were allowed during data recording to facilitate a variety of different analyses. In this analysis, candidate events are *triggered on signal* (TOS). That means that daughter particles of the signal signature must have fulfilled the trigger requirements. Events are required to have one or two muons with high transverse momentum ( $p_T$ ). The decay vertex of the  $J/\Psi$ -candidate has to be well separated from the PV ( $IP\chi^2$ ) and the fit quality of the decay vertex ( $Vertex\text{-}\chi^2$ ) and track ( $Track\text{-}\chi^2/ndof$ ) is required to be good. The chosen trigger configuration for this thesis is shown in appendix A.1, table A.1. A detailed description of the trigger lines is given in [21].

In order to reduce the recorded data to a manageable amount, a so called *stripping* is applied offline<sup>8</sup> afterwards. Only events that fulfill the specified conditions are used for an analysis. Many different versions exist, depending on the analysis type and decay channel.

---

<sup>8</sup>That means after the data is already recorded.

For this analysis, the stripping line `FullDSTDiMuonJpsi2MuMuDetachedLine` version Stripping 21 (see Table 4.1) was manually expanded by the cuts in Table 4.2. In the following, the combination of both is referred to as *stripping*.

Candidate	Selection
$B^+$	$DLS > 3$
$J/\Psi$	$2996.916 \text{ MeV} < M < 3196.916 \text{ MeV}$ $Vertex-\chi^2/ndof < 20$
$\mu^\pm$	$DOCA\chi^2 < 30$ $p_T > 500 \text{ MeV}$ $Track-\chi^2/ndof < 5$

Table 4.1: Cuts applied in `FullDSTDiMuonJpsi2MuMuDetachedLine` for Stripping 21.

Candidate	Selection
$B^+$	$4000 \text{ MeV} < M < 6000 \text{ MeV}$ $IP\chi^2 < 16$ $DIRA > 0.9995$ $\chi^2(FD) > 64$ $\min(IP\chi_{daughters}^2)^9 > 9$ $Vertex-\chi^2 < 10$
$K^+$	$ghostprob < 0.5$ $\min(IP\chi_{any-PV}^2)^{10} > 6$ $hasRICH = true$ $PID_{K\pi} > 0$
$K^{*+}$	$p_T > 800 \text{ MeV}$ $492 \text{ MeV} < M < 1292 \text{ MeV}$
$\mu^\pm$	$ghostprob < 0.5$ $\min(IP\chi_{any-PV}^2) > 9$ $PID_{\mu\pi} > 0$ $isMuon = true$
$\pi^0$	$p_T > 500 \text{ MeV}$ $CL > 0.02$

Table 4.2: Additional cuts to `FullDSTDiMuonJpsi2MuMuDetachedLine` (Stripping 21) to pick the decay into  $K^{*+}$ .

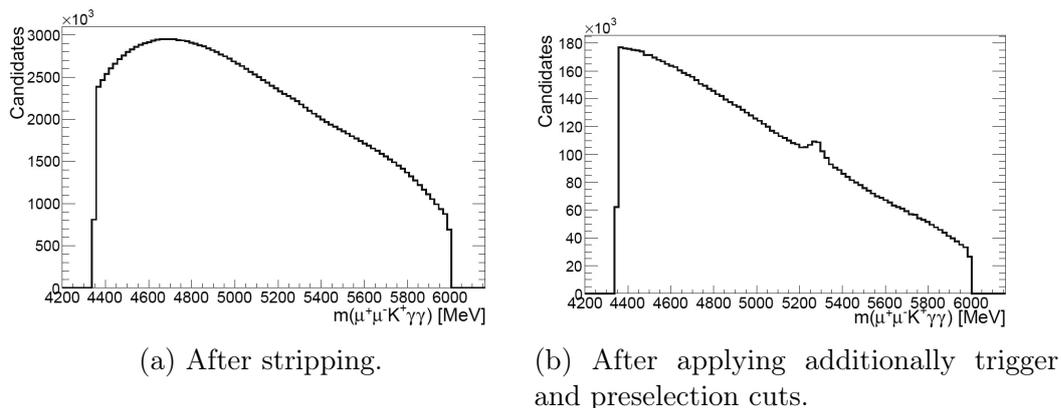


Figure 4.2: Reconstructed mass distribution of the  $B^+$  candidate in data.

## 4.4 Preselection

In order to obtain a good set of variables to discriminate signal and background, a sample of simulated  $B^+ \rightarrow J/\Psi K^{*+}$  decays is compared to background events from the upper sideband<sup>11</sup> of the  $B$  candidate mass distribution in data. The lower sideband is not used for this comparison because partially reconstructed background contributions are found at lower masses. These events have similar characteristics than the signal event. This would distort the distributions from combinatorial background and make the preselection less efficient.

Because we want to keep as much signal as possible at this stage of the analysis, only highly efficient cuts are applied. That means that after applying one cut, only a few percent of the signal events are lost according to simulations, while a significantly higher amount of background events is removed. A comparison of the mass distribution of data before and after the preselection can be found in Fig. 4.2.

Besides combinatorial background, a peaking contribution in the upper sideband is found. This is due to random  $\pi^0$ s added to  $B^+ \rightarrow J/\Psi K^+$  decays. Therefore a veto for events with an invariant mass within  $3\sigma$  of the nominal  $B^+$  is applied (Fig. 4.3).

A summary of all cuts can be found in Table 4.3.

## 4.5 Signal Monte Carlo calibration

Because Monte Carlo samples are used to train a multivariate classifier (Sec. 4.6.1) and to determine the selection efficiencies (Sec. 5.1), it is important that all used variables are simulated correctly. To check this, the channel  $B^0 \rightarrow J/\Psi K^{*0} (\rightarrow K^+\pi^-)$  is chosen as control channel. With similar particles as in the signal channel

<sup>9</sup>This variable checks every daughter particle for their  $IP\chi^2$  and puts a lower limit on it.

<sup>10</sup> $\min(IP\chi^2_{any-PV})$  measures the minimal  $IP\chi^2$  to any primary vertex in the recorded data.

<sup>11</sup>Combinatorial background (see Sec.4.1) is featureless and has on average the same characteristics at any reconstructed  $B$  mass. Therefore it is expected that the chosen cuts reduce background uniformly.

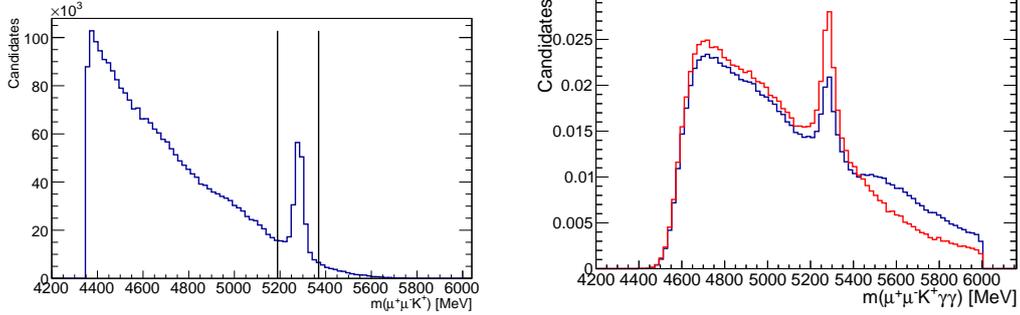


Figure 4.3: *Left*: Invariant mass of  $J/\Psi$  and  $K^+$  combined in data. A clear peak in the  $B^+$  mass region can be seen. A veto for the  $B^+ \rightarrow J/\Psi K^+$  decay is applied. The cut window is indicated with black lines. *Right*: Normalised  $B^+ \rightarrow J/\Psi K^{*+}$  mass distribution in data. Only a loose selection was applied to make the contribution of  $B^+ \rightarrow J/\Psi K^+$  better visible. Distribution before (after) applying the veto on  $m(J/\Psi K^+)$  in blue (red). The distributions are normalised to make the shoulder better visible. A not normalised comparison can be found in appendix A.1

Candidate	Selection
$B^+$	$Vertex\chi^2/ndof < 10$ $DTF\_ \chi^2 < 500$ $\eta < 5.5$ $\log(\Delta\chi^2_{add-track}) > 2$
$J/\Psi$	$3051.916 \text{ MeV} < M < 3141.916 \text{ MeV}$
$K^{*+}$	$p_T > 1000 \text{ MeV}$ $792 \text{ MeV} < M < 992 \text{ MeV}$
$K^+$	$p_T > 400 \text{ MeV}$ $PID_{K\pi} > 0$
$J/\Psi K^+$	$M < 5189 \text{ MeV}$ or $M > 5369 \text{ MeV}$
$\pi^0_{resolved}$	$104.98 \text{ MeV} < M < 164.98 \text{ MeV}$ $CL > 0.05$
$\gamma$	$max(CL(\gamma_1), CL(\gamma_2)) > 0.05$ $max(p_T(\gamma_1), p_T(\gamma_2)) > 300 \text{ MeV}$

Table 4.3: Preselection cuts for the signal channel  $B^+ \rightarrow J/\Psi K^{*+}$ .

Candidate	Selection
$K^{*+}$	$p_T > 800 \text{ MeV}$
$K^+$	$p_T > 800 \text{ MeV}$
	$PID_{K\pi} > 0$
$\pi^-$	$PID_{K\pi} < 0$

Table 4.4: Preselection cuts for the control channel  $B^0 \rightarrow J/\Psi K^{*0}$  as in [22].

$B^+ \rightarrow J/\Psi K^{*+} (\rightarrow K^+ \pi^0)$ , it is supposed to have similar kinematics, but also to leave a clean signal in the detector as all final state particles are charged. This channel is used as a control and normalisation channel in [22]. The selection is adopted and therefore the fit and selection efficiencies are the same.

### 4.5.1 Unfolding a pure signal sample

$B^0$  candidates are selected using the same stripping and trigger configuration as in Sec. 4.3.2. The preselection for the control channel is shown in Table 4.4. A distinct peak in the invariant mass spectrum can already be seen with this selection. To unfold the pure signal from the background polluted distributions, the so-called *sPlot* technique is used. It is a statistical tool that uses a set of variables with known distributions of all sources (i.e. background and signal) of events, the so-called *discriminating* variables, to unfold the pure distributions of variables with unknown distributions, the *control* variables. It is essential that discriminating and control variables are uncorrelated with each other.

By using information from fits of signal and background events of the discrimination variable, a so-called *sWeight* is calculated for each event. Applying these *sWeights* on the data sample results in clean signal distributions of the control variables. The discriminating variable is chosen to be the invariant mass of the  $B$  candidate. A detailed explanation of the *sPlot* technique is given in [23].

To describe the signal peak of  $B^0 \rightarrow J/\Psi K^{*0}$  in the discriminating variable, two Gaussian distributions with shared mean were used. The Gauss function is defined as

$$G(m|\mu, \sigma) = A \cdot \exp\left(-\frac{(m - \mu)^2}{2\sigma^2}\right)$$

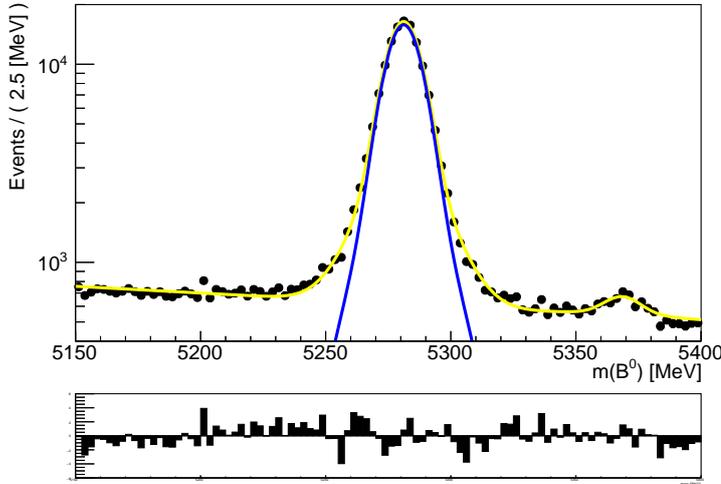
where  $A$  is a normalisation constant,  $\mu$  is the mean and  $\sigma$  the width of the distribution. Therefore, the double Gaussian is given by

$$F_s = f_1 \cdot G(m|\mu, \sigma_1) + (1 - f_1) \cdot G(m|\mu, \sigma_2)$$

with  $f_1$  describing the fraction of the two components. Mean, width and fraction are left to float in the fit of the data sample.

To account for the kinematically similar decay  $B_s^0 \rightarrow J/\Psi K^{*0}$ , which visibly contributes to the decay, the same fit model was used, but with a mean shifted by the mass difference  $\Delta m = m(B_s^0) - m(B^0) = 87.35 \text{ MeV}$ :

$$F_{B_s^0} = f_1 \cdot G(m|\mu + \Delta m, \sigma_1) + (1 - f_1) \cdot G(m|\mu + \Delta m, \sigma_2).$$



$$\begin{aligned}
 f_1 &= 0.249 \pm 0.014 \\
 \mu_1 &= (5281.149 \pm 0.026) \text{ MeV} \\
 N_{B^0} &= 125789 \pm 424 \\
 \sigma_1 &= 14.99 \pm 0.44 \text{ MeV} \\
 \sigma_2 &= 6.860 \pm 0.66 \text{ MeV}
 \end{aligned}$$

Figure 4.4: Fit to the mass spectrum of selected  $B^0 \rightarrow J/\Psi K^{*0}$  candidates on 2012 Magnet Down data as in [22]. A double Gaussian is used for the signal contribution (blue). Combinatorial background was fitted with an exponential. The y-axis is logarithmic to show the small contribution of  $B_s^0 \rightarrow J/\Psi K^{*0}$ . It is fitted with the same model as for  $B^0 \rightarrow J/\Psi K^{*0}$ , but with a shifted mean of  $\Delta m = m(B_s^0) - m(B^0) = 87.35 \text{ MeV}$ . The full fit model is shown in yellow. The obtained fit values for the signal are shown on the right.

Combinatorial background is fitted with an exponential distribution

$$F_{bkg} = B \cdot \exp(-\lambda \cdot m)$$

where  $B$  is the normalisation constant and  $\lambda$  describes the slope. Combining all distributions, we get the full PDF:

$$F_{full} = f_{sig} F_{sig} + f_{B_s^0} F_{B_s^0} + (1 - f_{sig} - f_{B_s^0}) F_{bkg}.$$

The fit of the 2012 Magnet Down sample is shown in Fig. 4.4, while fits of all other samples of the control channel can be found in appendix A.3. This fit was also used to obtain the number of events, which was later used to determine the branching fraction (see Sec. 5.2).

## 4.5.2 Monte Carlo reweighting

At this point, the distributions of truthmatched events of the signal channel could be compared to the sWeighted  $B^0 \rightarrow J/\Psi K^{*0}$  data, which now reflects the true distribution of signal events. While most distributions are in reasonable agreement with each other, the distributions of  $p_T(B^+)$  and  $nTracks$ <sup>12</sup> show significant deviations (Fig. 4.5).

As these variables have direct influence on other parameters (e.g.  $A_{PT}^{cc}$ ) that are used in the selection and BDT, they have to be corrected. This can easily be done

<sup>12</sup>The number of unique tracks in an event is denoted by  $nTracks$ .

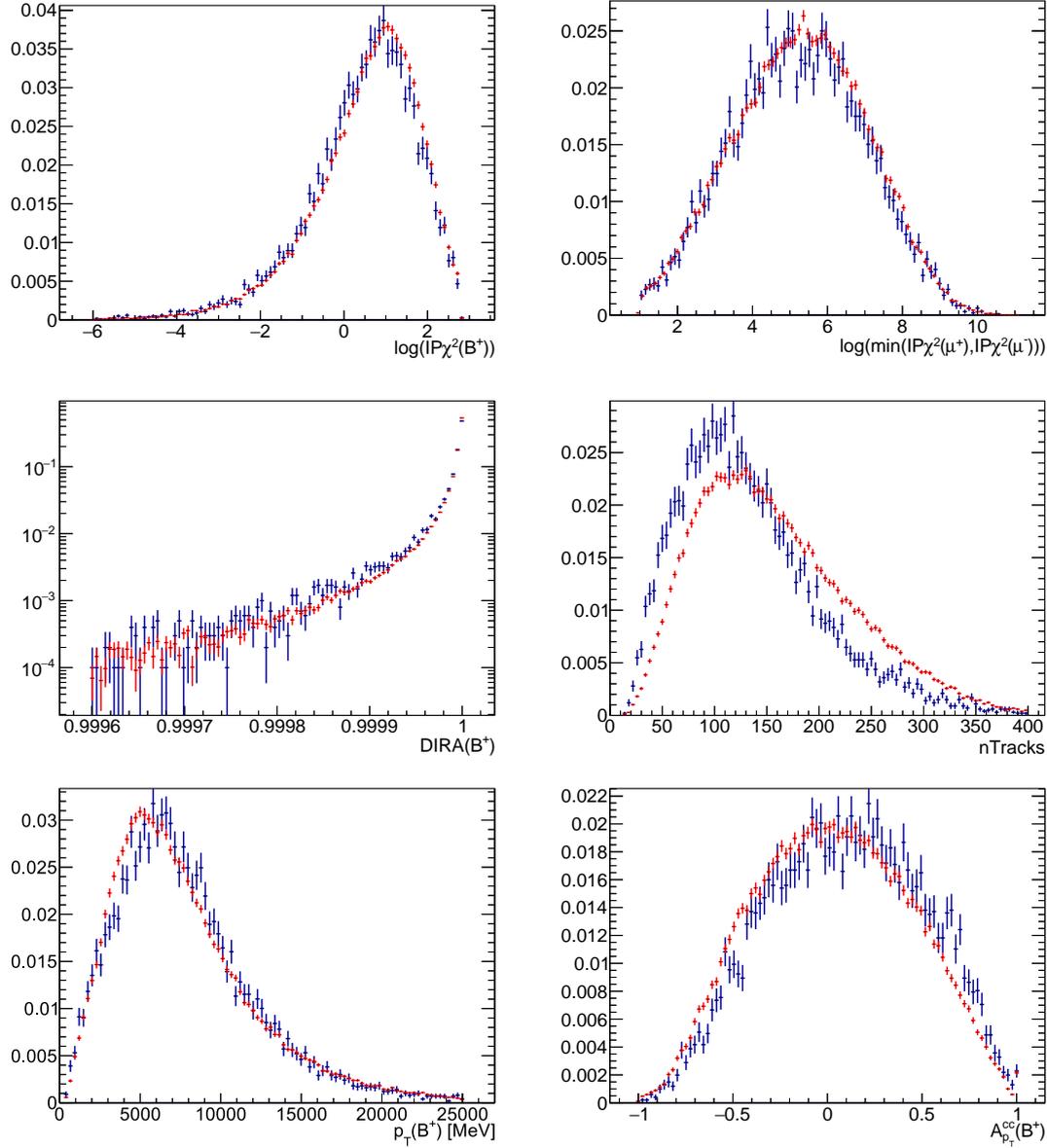


Figure 4.5: Comparison of simulated, truthmatched  $B^+ \rightarrow J/\Psi K^{*+}$  distributions (blue) with sWeighted  $B^0 \rightarrow J/\Psi K^{*0}$  distributions (red). Only the distributions of  $p_T(B^+)$ ,  $n\text{Tracks}$  and  $A_{p_T}^{cc}$  shows significant deviations. For illustration only some of the agreeing distributions are shown. Same stripping and trigger decisions are applied for both channels. The used preselection is explained in Sec. 4.4.

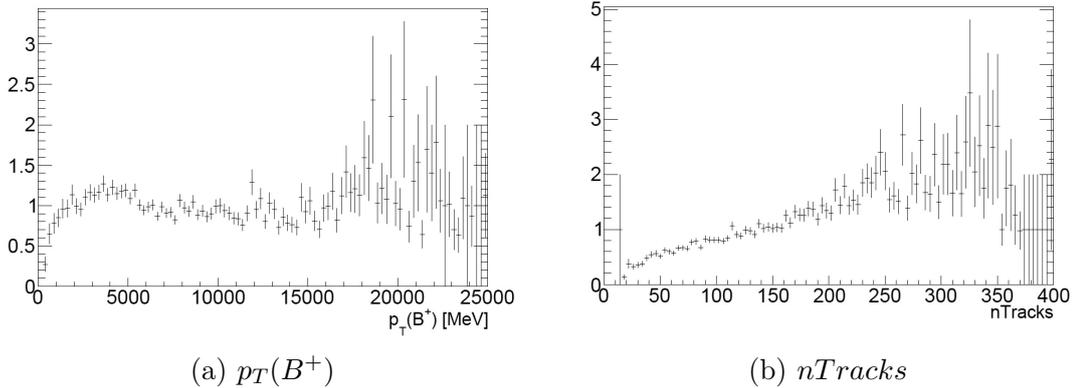


Figure 4.6: Weights calculated to reweight the signal Monte Carlo sample. The normalised, *s*Weighted distribution of  $B^0 \rightarrow J/\Psi K^{*0}$  was divided by the normalised and truthmatched distribution of simulated  $B^+ \rightarrow J/\Psi K^{*+}$  events. To prevent the usage of wrong weights, those with relative errors greater than 50% are set to 1, as can be seen in the last few bins of the right figure.

by reweighting the distributions. This means that, for a chosen binning, the normalised histogram of *s*Weighted control channel is divided by the truthmatched MC signal distribution, resulting in the weight distributions shown in Fig. 4.6. Because some bins contained very few events, outliers with large errors can be found. This happens especially for high values, where the distributions become flat (see Fig. 4.7). To prevent the usage of wrong weights, those with relative uncertainties greater than 50% were set to 1.

The transverse momentum  $p_T(B^+)$  and  $nTracks$  are assumed to be uncorrelated with each other, correcting for both parameters is thus done by simply multiplying the two weights. After applying the combined weights to each event in simulation, the discrepancies of all distributions are reduced to an insignificant amount (Fig. 4.7). The weighted  $B^+ \rightarrow J/\Psi K^{*+}$  simulated sample now reflects the true signal distributions. Therefore it can be used as signal proxy for a multivariate classifier (see next section).

## 4.6 Multivariate analysis

Now that we have reduced the data to a manageable amount and have identified important background discriminating variables (see Sec. 4.3.1), stricter cuts need to be applied to obtain a well separated signal. A first attempt would be to simply apply linear cuts on different parameters. However, this does not include possible correlations between the parameters and is therefore less efficient. A better selection can be applied by using a multivariate classifier. The *Toolkit for Multivariate Data Analysis* (TMVA, [24]) provides many algorithms that can be used. For this analysis, the so-called *Boosted Decision Tree* (BDT) is chosen. It uses information of inputs for signal and background distributions to create a single variable output, which classifies all events. A single cut on this *BDT output* is then the most efficient one, taking all correlations into account. The following description

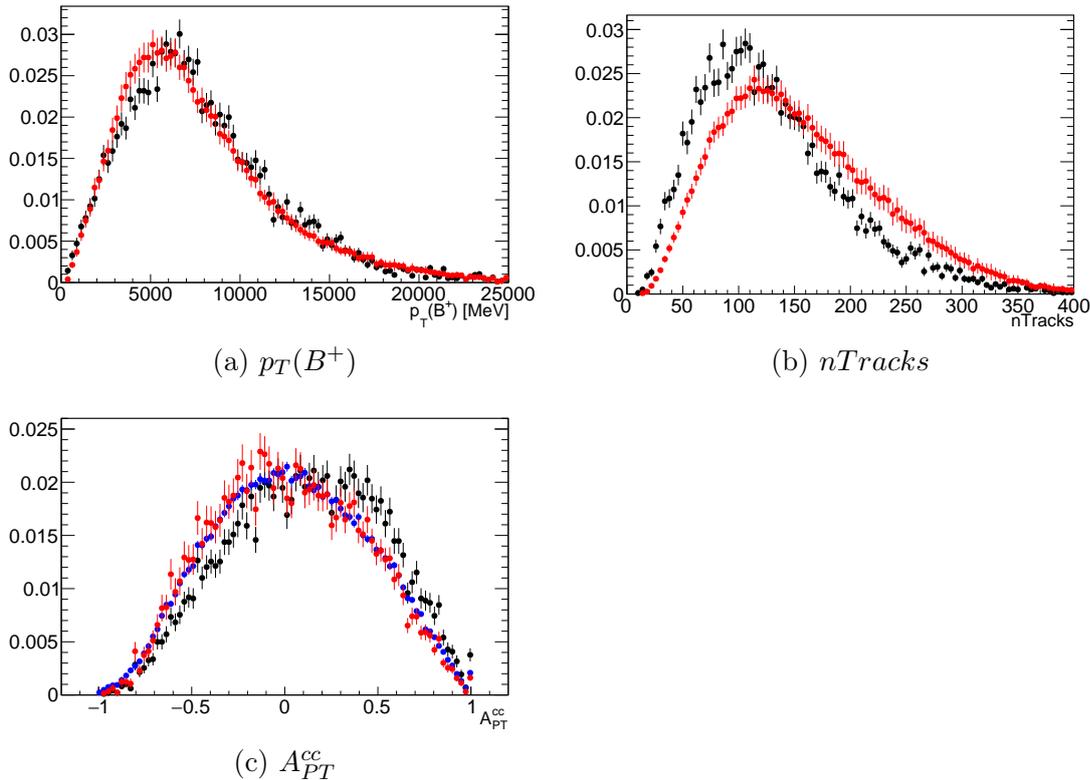


Figure 4.7: Comparison between simulated signal samples before and after reweighting and sWeighted data of the control channel . Simulated  $B^+ \rightarrow J/\Psi K^{*+}$  sample in black, reweighted MC distribution in red and sWeighted  $B^0 \rightarrow J/\Psi K^{*0}$  sample in blue. All distributions show good agreement between reweighted and sWeighted samples.  $A_{PT}^{cc}$  was chosen to represent the results of the reweighting.

orientates itself by Ref. [24].

Before a BDT can be used, it needs to be trained in how to separate signal and background events from each other. Therefore, samples displaying the corresponding distributions have to be provided. One half of the provided samples is used for training and one is used for testing.

A single decision tree (see Fig. 4.8) applies a binary cut on one of the input variables of the training sample at each node, resulting in the best separation into a background-like and signal-like part. In this case, the cut is chosen to minimise the *Gini-impurity*  $p(1-p)$ , with  $p$  being the the fraction of signal in the node. The Gini-impurity is constructed such that it is minimal for total separation of background and signal and maximal for 50% signal. This scheme will be repeated until a certain stopping criterion is fulfilled. In most cases, only a maximum number of nodes are allowed in order to reduce overtraining<sup>13</sup>. To check for overtraining, the distributions of the BDT output of training and test sample are compared to each other. Large deviations indicate overtraining.

In order to increase the separation power and make up for the now reduced amount

<sup>13</sup>A decision tree that becomes sensitive to statistical fluctuations is called overtrained. It shows good performance on the training sample, but will perform worse on other samples.

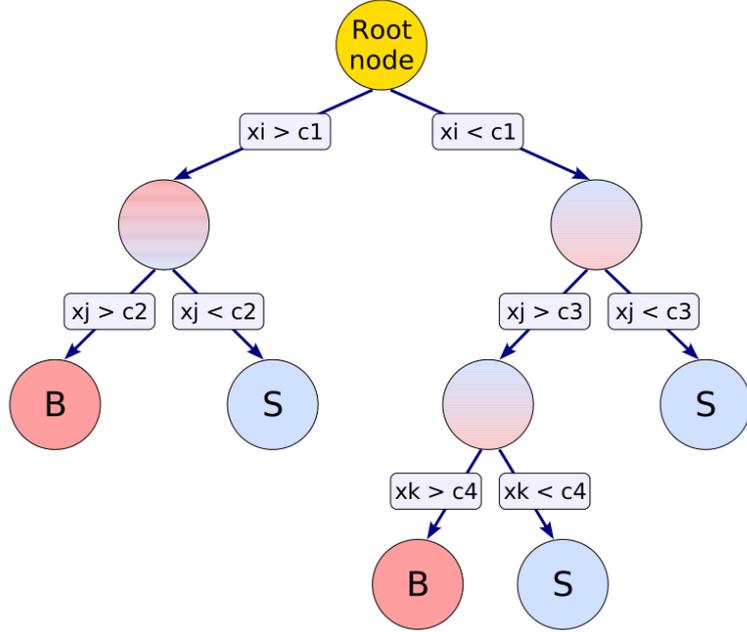


Figure 4.8: Scheme of a single Decision Tree. A one-dimensional cut  $c_i$  is applied on input-variable  $x_i$  such that the best discrimination between background and signal is achieved. Leaf nodes are labelled B for background and S for signal, depending on the majority of it entries. Figure taken from [24]

of nodes in the tree, boosting is used. Instead of just one decision tree, a whole forest of trees is trained. A weight is assigned to each event in the training sample. If events are incorrectly classified, their weight will be increased, so that the next tree is more sensitive to those events. As a final result, each event gets classified by a majority vote of all trees in the forest. A single response variable is created with values between  $-1$  (background-like) and  $+1$  (signal-like). For this analysis, the *adaptive boost algorithm* (AdaBoost, see [24]) was chosen.

#### 4.6.1 BDT against combinatorial background

The discriminating variables chosen as input for the BDT are shown in table 4.5. A good separation power is required (Fig. 4.9) as well as a good agreement between simulation and data (Sec. 4.5). Although there are strong background discriminating photon associated quantities, they are not used for the multivariate analysis. The reason for this is that our control channel has only charged final state particles, which means it is not possible to check if the quantities are correctly simulated.

For the training samples, a truthmatched and reweighted MC sample of  $B^+ \rightarrow J/\Psi K^{*+}$  of each magnet polarity is used as signal proxy. They need to be combined because a single sample provides too little statistics. Magnet Up data from the upper sideband is used as background sample. No significant deviations between the used distributions of up and down polarity could be found, therefore combining both samples is justified. 2011 and 2012 data are separately trained.

$A_{p_T}^{cc}(B^+)$	$\log(\Delta\chi_{add-track}^2(B^+))$
$DIRA(B^+)$	$\log(\min(IP\chi^2(\mu^+), IP\chi^2(\mu^-)))$
$\log(Vertex-\chi^2(B^+))$	$\log(\max(p_T(\mu^+), p_T(\mu^-)))$
$\log(IP\chi^2(B^+))$	$p_T(K^+)$
$p_T(B^+)$	

Table 4.5: Input variables of the BDT. An explanation of these quantities is given in Sec. 4.3.1.

The *BDToutput* shows good separation between signal and background events (Fig. 4.10). Test nor training samples show significant deviations, which implies that the classifier is not overtrained.

The next step would be to find the optimal cut value of the *BDToutput* by maximising a *Figure of Merit*  $FOM = \frac{S}{\sqrt{S+B}}$ , where  $S$  ( $B$ ) is the number of expected signal (background) events for a chosen *BDToutput* cut. Trying to obtain these numbers from a fit to data would give rise to a bias. Statistical fluctuations could influence the maximum *FOM*, leading to a higher signal yield and therefore a higher branching ratio. Hence, one has to obtain these numbers separately. The number of expected signal events can be obtained from rescaling the MC sample size to the data and extract the number of events after truthmatching and applying the *BDToutput* cut.

By excluding the signal region and fitting only background events from the upper and lower sideband, the amount of background events  $B$  in the signal region can be extrapolated. This should have been done for many different *BDToutput* cut values. Unfortunately, the time was not sufficient to do this. Instead, a cut value was chosen by simply looking at the *BDToutput*. It is chosen to be *BDToutput*  $> 0.3$ .

## 4.7 Fit model and remaining background contributions

To extract the signal yield needed for the branching ratio determination, a fit has to be performed on the data. To do this, partially reconstructed and peaking backgrounds have to be taken into account. This section lists all found contributions as well as the chosen PDFs to describe them.

To fit  $B^+ \rightarrow J/\Psi K^{*+} (\rightarrow K^+\pi^0)$  candidates, a PDF consisting of two Crystal Balls (CB) with shared mean, also called double Crystal Ball, is chosen. The CB function consists of a Gaussian with a power tail on one side. It is a continuous function that is defined as follows:

$$\mathcal{P}_{CB}(m|\alpha, n, \mu, \sigma) = \begin{cases} \frac{\left(\frac{n}{|\alpha|}\right)^n \exp\left(-\frac{1}{2}\alpha^2\right)}{\left(\frac{n}{|\alpha|} - |\alpha| - t\right)^n}, & x \leq -|\alpha| \\ \exp\left(-\frac{1}{2}t^2\right), & x > -|\alpha| \end{cases}$$

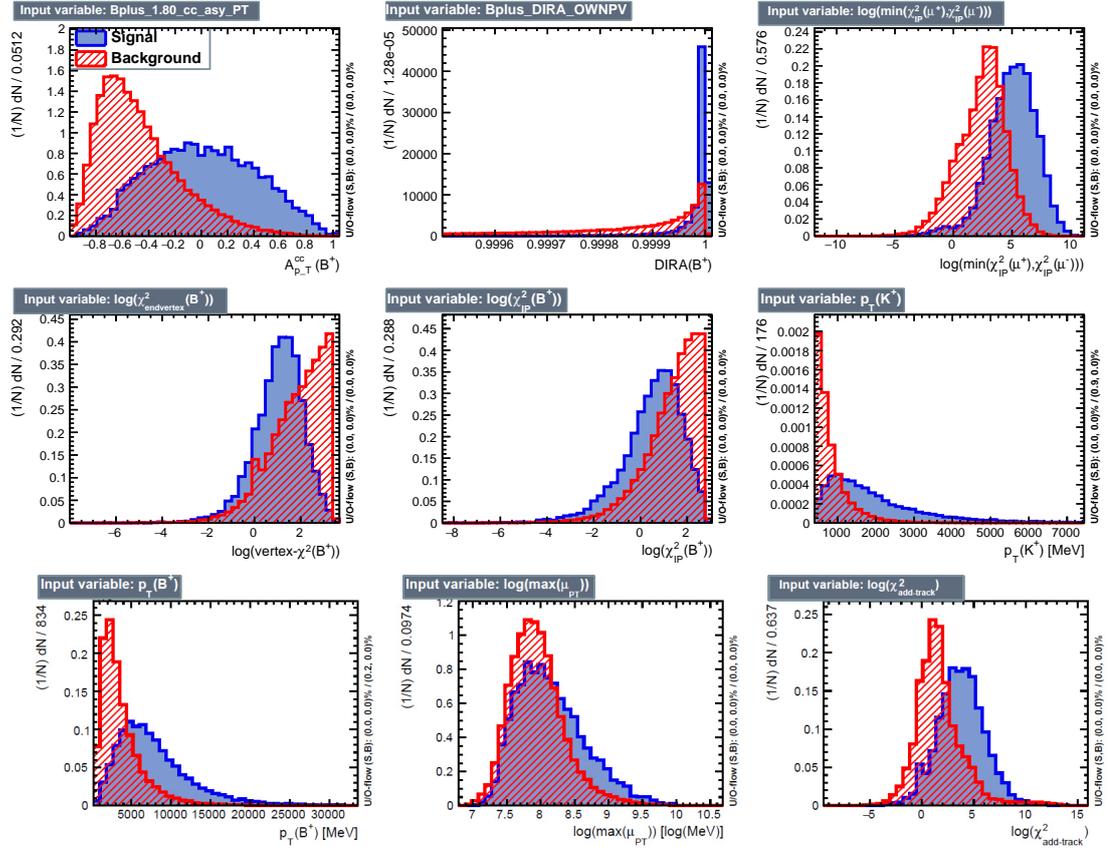


Figure 4.9: Comparison of signal and background distributions for BDT input variables. Truthmatched and reweighted  $B^+ \rightarrow J/\Psi K^{*+}$  MC sample of 2012 in blue (signal proxy) and data from the upper sideband of the 2012 Magnet Up data sample in red (background proxy).

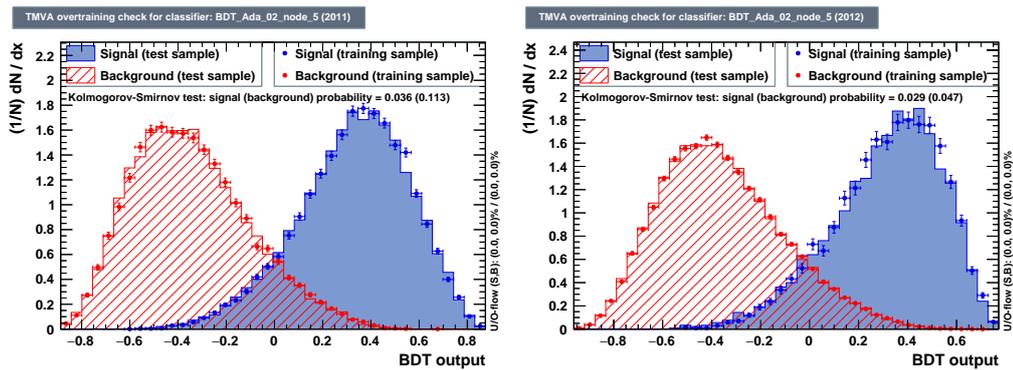


Figure 4.10:  $BDT_{output}$  for the years 2011 (left) and 2012 (right). Simulated, truthmatched  $B^+ \rightarrow J/\Psi K^{*+}$  events of Magnet Up and Down polarity are used as signal proxy. Events from the upper mass sideband of  $B^+ \rightarrow J/\Psi K^{*+}$  Magnet Up data is used as background proxy. No deviations between Magnet Up and Down polarities could be found, therefore combining them is justified. The training samples are displayed as dots, the test samples as filled areas. Both have similar distributions. This implies little overtraining.

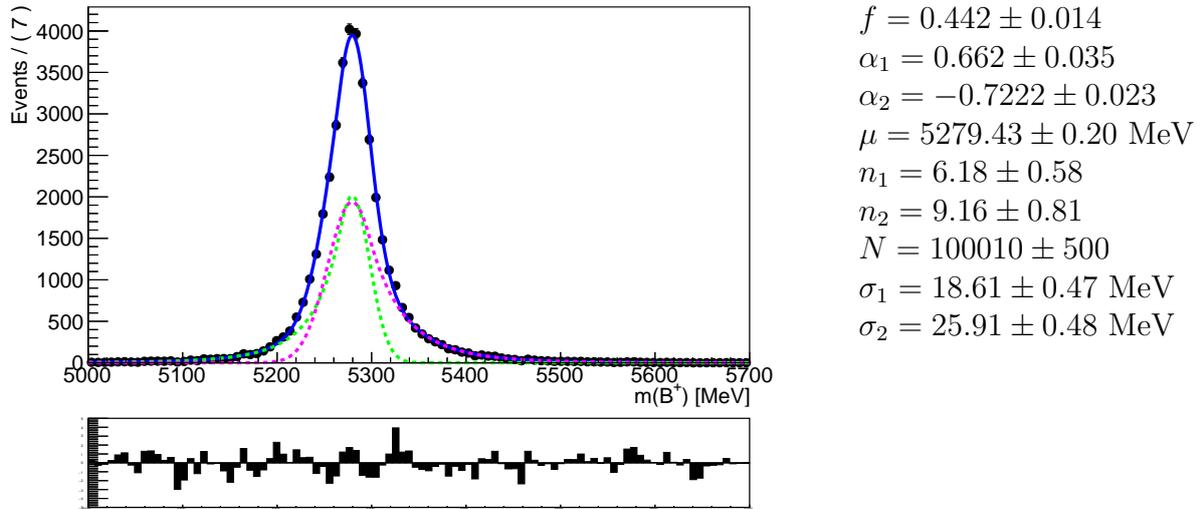


Figure 4.11: Simulated  $B^+ \rightarrow J/\Psi K^{*+}(\rightarrow K^+\pi^0)$  distribution after truthmatching. Full selection is applied. All samples (2011, 2012, Up and Down) were used. A double Crystal Ball is chosen to describe the signal shape. The fit result is shown in blue, the single Crystal Ball contributions in green and magenta, the fit parameters are shown on the right.

with  $t = \frac{m-\mu}{\sigma}$ . This formula displays a CB with the power-tail on its left side. The transition between Gaussian and power-tail takes place at  $\alpha \cdot \sigma$ ;  $n$  describes the slope of the tail,  $\mu$  is the peak position and  $\sigma$  the width. The left tail describes the energy loss from radiative processes like bremsstrahlung. A Crystal Ball with tail to the right is defined by  $t = -t$ ,  $\alpha < 0$ . The right tail is needed because energies of photons inside the calorimeters are overestimated if photon associated clusters overlap in space. The released energy is then not correctly measured. The PDF of the double Crystal Ball is then:

$$\mathcal{P}_{sig} = f \cdot \mathcal{P}_{CB}(m|\alpha_1, n_1, \mu, \sigma_1) + (1 - f) \cdot \mathcal{P}_{CB}(m|\alpha_2, n_2, \mu, \sigma_2)$$

with  $\alpha_1 > 0$  and  $\alpha_2 < 0$ .

An attempt is made to fix all parameters to the values obtained from simulation (see Fig. 4.11), but it could not describe the signal shape properly. Therefore, the peak position  $m$ , the widths  $\sigma_1, \sigma_2$  and the fraction  $f$  are left to float during the signal fit.

To describe the background shape, several PDFs have to be constructed. In the following, a short explanation of each background type, as well as the contributing decays, is given. A more detailed explanation of background sources can be found in Sec. 4.1.

**Combinatorial Background** describes  $B^+$  candidates that are reconstructed from random particles that do not share an ancestor. It is featureless and was fitted

with a Chebyshev polynomial ([25]) of third order:

$$\mathcal{P}_{cbg}(m|a_1, a_2, a_3) = 1 + \sum_{i=1}^3 a_i T_i(m)$$

with

$$\begin{aligned} T_1 &= m \\ T_2 &= 2m^2 - 1 \\ T_3 &= 4m^3 - 3m. \end{aligned}$$

Chebyshev polynomials can describe the same shapes as regular polynomials, but are more stable in fits because of clever reorganization of the terms, leading to less correlations between the coefficients  $a_i$ .

**Partially reconstructed background** comes from events similar to  $B^+ \rightarrow J/\Psi K^{*+}$  ( $\rightarrow K^+ \pi^0$ ), where an additional particle such as a  $\pi$  or  $\gamma$  is not reconstructed. The following contributions are expected:

- $B^+ \rightarrow J/\Psi K_1^+$  with  $K_1^+ \rightarrow K^{*+} \cancel{\pi}^{14}$  has the same final state particles as the signal decay. The expected branching fraction is 6-7% relative to  $B^+ \rightarrow J/\Psi K^{*+}$ . A PDF consisting of a convolution<sup>15</sup> of a Gaussian and an Argus function was fitted to simulated events (see Fig. 4.12). The Gaussian function is given as

$$\mathcal{P}_{Gauss}(m|\mu, \sigma) = \exp\left(-0.5 \left(\frac{m - \mu}{\sigma}\right)^2\right)$$

with mean  $\mu$  and width  $\sigma$ .

The Argus function is an empirical function given by

$$\mathcal{P}_{Argus}(m|\mu, c, p) = x \left(1 - \left(\frac{m}{\mu}\right)^2\right)^p \cdot \exp\left(c \left(1 - \left(\frac{m}{\mu}\right)^2\right)\right),$$

where  $\mu$  is the cut-off value and  $c, p$  are shape parameters. When fitting the data, the cut-off  $\mu$  as well as the yield of this PDF were left to float to account for the  $\chi_{c1}$ -decay (see next bullet point). All other parameters are fixed to the values obtained from the fit to simulated data.

- The decay  $B^+ \rightarrow \chi_{c1}(\rightarrow J/\Psi \gamma) K^{*+}$  is expected to form partially reconstructed background. It has a branching fraction of  $\approx 7\%$  relative to the signal channel. Unfortunately, no simulation is available for this channel. Hence, no PDF could be constructed.
- Another contribution from  $B^+ \rightarrow \Psi(2S)(\rightarrow J/\Psi \cancel{\pi} \cancel{\pi}) K^{*+}$  can be seen in the data. Since two particles have to be missed for this contribution, it ends up far below the signal mass window and contributes little, although

<sup>14</sup>Crossed out particles imply that these were not reconstructed.

<sup>15</sup>The convolution  $(f * g)(t)$  of two functions  $f$  and  $g$  is defined as  $(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau$  and can be seen as "blending" one function  $f$  with another function  $g$ .

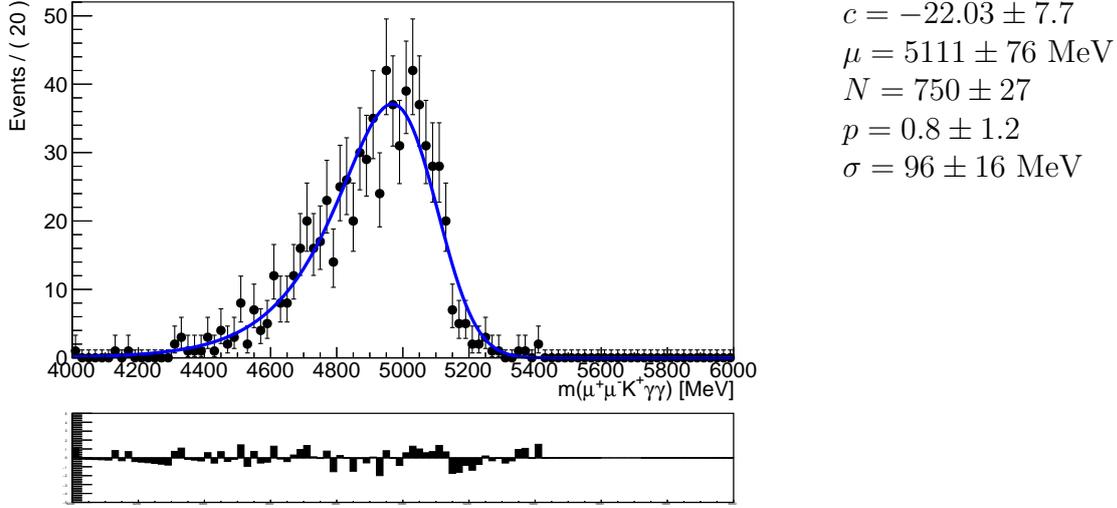


Figure 4.12: Simulated  $B^+ \rightarrow J/\Psi K_1(\rightarrow \pi^0 K^{*+}(\rightarrow K^{*+}\pi^0))$  distribution after truthmatching. Full selection is applied. A convolution of Gaussian and Argus function is chosen to describe the shape. The fit result is shown in blue, the fit values are shown on the right.

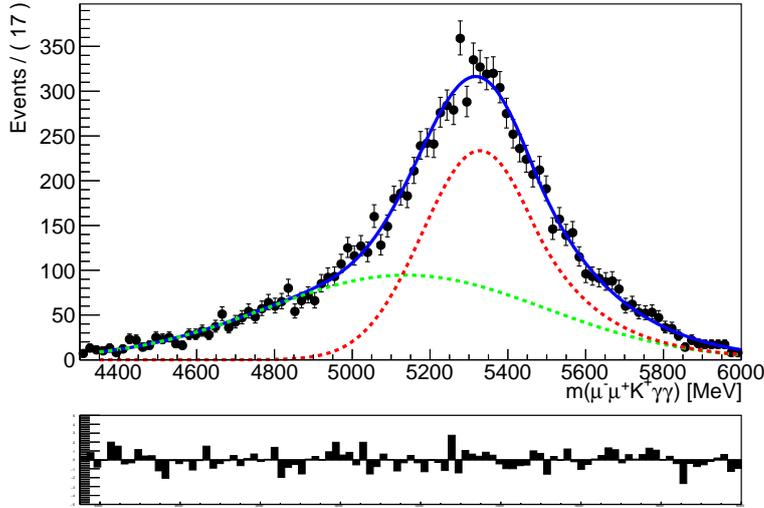
enough to influence the polynomial PDF of the combinatorial background. No simulation is available for this either. A single Gaussian PDF, with all parameters left floating, is chosen to describe this decay.

**Peaking Background** comes from decays where one particle is misidentified so that the decay looks like the one investigated. In this analysis,  $B^0 \rightarrow J/\Psi K^{*0}(\rightarrow K^+\pi^-)$  can form peaking background if the  $\pi^-$  is missed in the reconstruction and replaced by a random  $\pi^0$ . To describe this decay, a Gaussian and Crystal Ball with all parameters left to float are fitted to simulated MC events (see Fig. 4.13). In order to estimate the contribution of these events, a simulated sample was compared to the control channel. The yield for misidentified  $B^0 \rightarrow J/\Psi K^{*0}(\rightarrow K^+\pi^-)$  events is determined to be  $10416 \pm 138$ . All parameters obtained from simulation are fixed in the data fit, including the yield, because floating this parameter leads to unphysical high yields.

Combining all PDFs, the final fit function becomes:

$$P_{full} = N_{sig}\mathcal{P}_{sig}(m|\mu, \sigma_1, \sigma_2, f) + N_{cbg}\mathcal{P}_{cbg}(m|a_1, a_2, a_3) + N_{K_1}\mathcal{P}_{K_1}(m|\mu) + N_{\Psi(2S)}\mathcal{P}_{\Psi(2S)}(m|\mu, \sigma) + N_{B^0}\mathcal{P}_{B^0}(m)$$

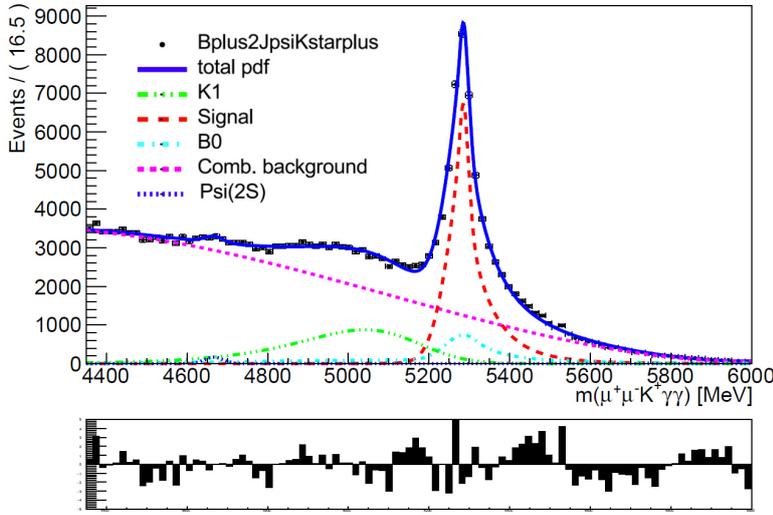
where  $N_i$  is the yield of every contribution. The yield  $N_{B^0}$  was fixed to  $N_{B^0} = 10416$ . A systematic uncertainty is introduced in chapter 6 to account for the uncertainty of this yield. All other yields are left to float. The datasets of 2011 and 2012 are combined to increase the stability of the fit. Figure 4.14 shows the final fit result, where the signal yield is determined to be  $(37005 \pm 627)$ .



$$\begin{aligned}
 \alpha &= -0.918 \pm 0.10 \\
 \mu_{CB} &= 5328.1 \pm 4.5 \text{ MeV} \\
 \mu_{Gauss} &= 5137 \pm 32 \text{ MeV} \\
 N_{CB} &= 5605 \pm 473 \\
 N_{Gauss} &= 4925 \pm 472 \\
 n &= 123 \pm 243 \\
 \sigma_{CB} &= 148.3 \pm 7.4 \text{ MeV} \\
 \sigma_{Gauss} &= 360 \pm 13 \text{ MeV}
 \end{aligned}$$

Figure 4.13: Simulated  $B^0 \rightarrow J/\Psi K^{*+}(\rightarrow K^{*+}\pi^-)$  events misidentified as  $B^+ \rightarrow J/\Psi K^{*+}(\rightarrow K^{*+}\pi^0)$ . Full selection was applied. A combination of Gaussian and CB was chosen to describe the shape. All parameters are left to float. The Gaussian is shown in green, the CB in red and the combination of both in blue. Fit values are shown on the right.

The chosen fit model does not accurately describe the data. Especially the transition zone from signal to partially reconstructed background cannot be described well. This is most likely due to an incomplete description of background contributions. A simulated sample of  $B^+ \rightarrow \chi_{c1}(\rightarrow J/\Psi \gamma)K^{*+}$  as well as a bigger sample of  $B^+ \rightarrow J\Psi K_1(\rightarrow K^{*+}\pi^0)$  could have improved the description. The yield of these two decays combined is expected to be around 14% relative to the signal yield, whereas we measure around 50%. Therefore, the contributions of partially reconstructed or peaking backgrounds have to be further investigated. To account for this issue, a systematic uncertainty is assigned (see Sec. 6).



$$\begin{aligned}
 f_{dCB} &= 0.215 \pm 0.012 \\
 a_1 &= -1.12639 \pm 0.0059 \\
 a_2 &= 0.056 \pm 0.011 \\
 a_3 &= 0.1029 \pm 0.0063 \\
 \mu_{sig} &= 5285.21 \pm 0.32 \text{ MeV} \\
 \mu_{K1} &= 51585.8 \pm 5.7 \text{ MeV} \\
 \mu_{\Psi(2S)} &= 4666.1 \pm 7.0 \text{ MeV} \\
 N_{sig} &= 37005 \pm 627 \\
 N_{cbg} &= 152284 \pm 1621 \\
 N_{K1} &= 21391 \pm 947 \\
 N_{\Psi(2S)} &= 577 \pm 240 \\
 \sigma_{CB1} &= 12.17 \pm 0.51 \text{ MeV} \\
 \sigma_{CB2} &= 44.5 \pm 1.1 \text{ MeV} \\
 \sigma_{\Psi(2S)} &= 23 \pm 12 \text{ MeV}
 \end{aligned}$$

Figure 4.14: Fit of the  $B^+$  candidate mass after selection. The data samples of 2011 and 2012 are combined. Signal events from  $B^+ \rightarrow J/\Psi K^{*+} (\rightarrow K^+ \pi^0)$  are described with a double CB (red, dashed), combinatorial background is described by a Chebyshev polynomial of third order (pink, dotted), partially reconstructed background from  $B^+ \rightarrow J/\Psi K_1 (\rightarrow K^{*+} \pi^0)$  and  $B^+ \rightarrow \chi_{c1} (\rightarrow J/\Psi \gamma) K^{*+}$  by a convolution of Gaussian and Argus in green (dash-dotted with 3 dots between each dash) and  $B^+ \rightarrow \Psi(2S) (\rightarrow J/\Psi \pi^0 \pi^0) K^{*+}$  by a Gaussian in thin-dashed blue. Peaking background from  $B^0 \rightarrow J/\Psi K^{*0} (\rightarrow K^+ \pi^-)$  is described by a Gaussian and Crystal Ball obtained from MC (light blue, dash-dotted). Its yield is fixed to  $N_{B^0} = 10416$ .

## 5 Determining the branching ratio

This chapter gives a detailed explanation of the branching ratio determination of  $B^+ \rightarrow J/\Psi K^{*+}$  relative to  $B^0 \rightarrow J/\Psi K^{*0}$ . The signal yield obtained in section 4.7 only reflects the number of events passing a certain selection. Therefore, the selection efficiencies have to be determined before the relative branching fraction can be calculated.

### 5.1 Efficiencies

Selection efficiencies represent what fraction of signal candidates pass a certain selection. It is crucial to know these numbers for both channels as they directly influences the branching ratio measurement (see Sec. 5.2). They are obtained from simulated samples since the number of signal events is known at every step from truthmatching. The efficiency of a certain selection is then calculated by

$$\epsilon_{sel}^{MC} = \frac{N_{passed}^{MC}}{N^{MC}}$$

where  $N_{passed}^{MC}$  ( $N^{MC}$ ) denotes the number of signal events after (before) the selection.

For this to be correct, the distributions of signal and data have to show good agreement with each other. This is true for most quantities used in the preselection, but could not be checked for photon associated variables since the control channel  $B^+ \rightarrow J/\Psi K^{*0} (\rightarrow K^+ \pi^-)$  has charged final state particles only. Because no tools exists yet to get possible correction factors for the efficiencies of such cuts and time was not sufficient for further investigations, no strict cuts on photon associated variables are applied. For a more precise measurement possible deviations have to be investigated and corrected.

Moreover, *PID*-variables are known to be poorly simulated. With the `PIDCalib` tool [26], the true efficiency of a specified cut in data can be determined. It uses a number of clean data samples like  $K_s^0 \rightarrow \pi^+ \pi^-$ , which can be selected easily. The efficiencies for these samples can be determined precisely. The *PID* cut efficiency is chosen to be the average efficiency of all these samples.

As already seen in section 4.5, the signal MC samples have to be reweighted before using them as BDT input. The number of events  $N'$  after weighting is given by the sum of the weights  $w_i$ :

$$N' = \sum_i^N w_i,$$

where  $N$  is the total number of events before weighting. Because weighting can change the statistical contribution of single events, but cannot increase the total statistical power of the sample, a normalisation factor

$$w_{eff} = \frac{\sum_i^N w_i}{\sum_i^N w_i^2}$$

selection	$\epsilon(12U)$ [%]	$\epsilon(12D)$ [%]	$\epsilon(11U)$ [%]	$\epsilon(11D)$ [%]
preselection	1.50	1.50	1.56	1.53
<i>PID</i> cuts	87.6	87.8	87.6	87.6
<i>BDT</i> cut	69.2	68.3	69.7	71.6
total efficiency	0.909	0.900	0.952	0.960

Table 5.1: Selection efficiencies after each step for the simulated  $B^+ \rightarrow J/\Psi K^{*+}$  samples. Efficiencies are relative to the previous step. Stripping and trigger cuts are included in the preselection efficiency. The total efficiency is calculated by multiplying all efficiencies. PID efficiencies are obtained from `PIDCalib`, BDT efficiencies are taken from reweighted MC samples. An explanation of both methods is given in Sec. 5.1. Sample names are abbreviated, e.g. "12U" stands for "2012 Magnet Up".

selection	$\epsilon(12U)$ [%]	$\epsilon(12D)$ [%]	$\epsilon(11U)$ [%]	$\epsilon(11D)$ [%]
preselection	6.1	5.8	6.9	6.9
<i>PID</i> cuts	83.6	83.5	83.5	83.4
total efficiency	5.11	4.82	5.72	5.75

Table 5.2: Selection efficiencies after each step for the simulated  $B^0 \rightarrow J/\Psi K^{*0}$  samples. Efficiencies are relative to the previous step. Stripping and trigger cuts are included in the preselection efficiency. The total efficiency is calculated by multiplying all efficiencies. PID efficiencies were obtained from `PIDCalib`. An explanation of this method is given in Sec. 5.1. Sample names are abbreviated, e.g. "12U" stands for "2012 Magnet Up". Efficiencies are taken from [22].

is introduced [27]. The effective number of events is then given by

$$N_{eff} = w_{eff} \sum_i^N w_i.$$

This number corresponds to a non-weighted sample size with the same statistical power as the reweighted sample and is used to calculate the efficiencies.

The efficiencies were separately determined for every used MC sample. Table 5.1 shows the efficiencies after each step for all simulated signal samples. The numbers for the normalisation channel can be found in Table 5.2.

## 5.2 Branching ratio results

With the signal yield obtained in Section 4.7 and the efficiencies from Section 5.1, the relative branching fraction of  $B^+ \rightarrow J/\Psi K^{*+}$  can be calculated by

$$\frac{\mathcal{B}(B^+ \rightarrow J/\Psi K^{*+}(\rightarrow K^+ \pi^0(\rightarrow \gamma\gamma)))}{\mathcal{B}(B^0 \rightarrow J/\Psi K^{*0}(\rightarrow K^+ \pi^-))} = \frac{N_{sig}}{N_{norm}} \times \frac{\xi_{norm}}{\xi_{sig}} \times \frac{\epsilon_{norm}^{total}}{\epsilon_{sig}^{total}}.$$

Data sample	$N_{norm}$	Luminosity $\mathcal{L}$ [pb $^{-1}$ ]
2012 Up	$125789 \pm 424$	999
2012 Down	$125056 \pm 430$	988
2011 Up	$45741 \pm 253$	417
2011 Down	$64475 \pm 300$	559

Table 5.3: Signal yields of the normalisation channel  $B^0 \rightarrow J/\Psi K^{*0}$  and Luminosities for the different samples. An explanation of the yield determination method is given in Sec. 4.5.1.

The measured yields are denoted as  $N_i$ , where the signal yield is determined to be  $N_{sig} = 37005 \pm 627$  in Sec. 4.7. Table 5.3 shows the yields of the normalisation channel. The geometrical acceptances are taken from Monte Carlo simulations:  $\xi_{sig} = 15.2\%$ ,  $\xi_{norm} = 16.2\%$ . Because all data samples are combined for the fit, although they have different efficiencies, the efficiencies need to be combined too. Therefore, a weighted mean is used:

$$\epsilon_{norm}^{total} = \frac{\sum_i (\mathcal{L}_i \times \epsilon_i)}{\sum_i \mathcal{L}_i},$$

where  $i$  denotes the data samples of 2011 and 2012 of both magnet polarities. The luminosities for each data sample are given in Table 5.3. It should be mentioned that the efficiency also needs to be corrected for the different  $b\bar{b}$  cross-sections of 2011 and 2012 data. In comparison to the uncertainties, this can be neglected.

Using these values, the relative branching fraction is determined to be:

$$\frac{\mathcal{B}(B^+ \rightarrow J/\Psi K^{*+}(\rightarrow K^+\pi^0(\rightarrow \gamma\gamma)))}{\mathcal{B}(B^+ \rightarrow J/\Psi K^{*0}(\rightarrow K^+\pi^-))} = 0.62 \pm 0.01.$$

With the well-known branching ratios

$$\begin{aligned} \mathcal{B}(\pi^0 \rightarrow \gamma\gamma) &= (99.824 \pm 0.034)\% \\ \mathcal{B}(B^0 \rightarrow J/\Psi K^{*0}) &= (1.32 \pm 0.06) \cdot 10^{-3} \end{aligned}$$

and, assuming isospin symmetry,

$$\begin{aligned} \mathcal{B}(K^{*+} \rightarrow K^+\pi^0) &= 1/3 \\ \mathcal{B}(K^{*0} \rightarrow K^+\pi^-) &= 2/3 \end{aligned}$$

the total branching fraction of  $B^+ \rightarrow J/\Psi K^{*+}$  results in:

$$\mathcal{B}(B^+ \rightarrow J/\Psi K^{*+}) = (0.62 \pm 0.01) \cdot \frac{2/3 \cdot \mathcal{B}(B^0 \rightarrow J/\Psi K^{*0})}{1/3 \cdot \mathcal{B}(\pi^0 \rightarrow \gamma\gamma)} = (1.64 \pm 0.03) \cdot 10^{-3}.$$

Only the statistical uncertainty obtained from the fit result is shown here. It could be further decreased by optimising the  $BDT_{output}$  cut (see also Sec. 4.6.1), taking more data or improving the selection. Errors from the branching fractions of other decays or efficiencies are generally viewed as systematic uncertainty and will be investigated in the next section.

## 6 Systematic uncertainties

This section describes possible systematic uncertainties of the branching ratio measurement. Some uncertainties, for example from the luminosity, cancel by taking the ratio of  $B^+ \rightarrow J/\Psi K^{*+}$  and  $B^0 \rightarrow J/\Psi K^{*0}$ . Only a few systematic uncertainties could be studied in the scope of the thesis. These are expected to contribute the most. At the end of this chapter, a short outlook on other possible uncertainties is given. The quantified systematic uncertainties are the following:

- **Fit model:** The most obvious and biggest uncertainty is due to an insufficient description of the data (Fig. 4.14). It was not possible to study all found contributions from partially reconstructed backgrounds and other possible sources cannot be ruled out. To quantify the impact of the fit model, all fit parameters except the width of the gaussian from the  $K_1$  distributions and the shape from misidentified  $B^0$  events are left to float. The difference in the branching fraction results in an relative uncertainty of 10.1%. Detailed studies of (not yet produced) Monte Carlo simulations could help to reduce this error significantly.
- **Fit range:** The influence of the chosen mass window is determined by fitting the mass ranges: [4300,6100] MeV, [4300,5900] MeV, [4400,6100] MeV, [4400,5900] MeV and [4350,5800] MeV. The maximum difference in the signal yield is taken as uncertainty. This results in a difference of 6.0%.
- **Normalisation channel:** The uncertainty on the branching ratio of the normalisation channel directly translates to the branching ratio of  $B^+ \rightarrow J/\Psi K^{*+}$  and is accounted with 4.5%.
- **Simulated sample size:** Due to a finite number of events in simulated samples, a statistical uncertainty on the simulated number of events is introduced. Relative uncertainties of the efficiencies of signal and normalisation channel combine to a contribution of 3.4%.
- **Data-simulation differences:** Additional to the finite size of simulated events, possible deviations of simulated and real data have to be taken into account. Only corrections to the transverse momentum of  $B^+$  as well as  $nTracks$  are applied. Differences in other variables are assumed to be small enough to neglect them in the efficiency determination. Therefore, the impact of those variables is not assumed to exceed the impact of the  $p_T(B^+)$  and  $nTracks$  corrections. A rather conservative estimate of the systematical uncertainty is made by taking the difference of efficiencies with weighted and unweighted  $p_T(B^+)$  and  $nTracks$  distribution. This results in an relative uncertainty of 1.3%.
- **$B^0$  background:** To account for the uncertainty of the estimated  $B^0$  yield, the data is fitted with fixed yields of  $\pm 1\sigma$  difference. The average difference to the signal yield is 1% and taken as systematic uncertainty.

Source	Relative uncertainty [%]
Fit model	$\pm 10.1$
Fit range	$\pm 6.0$
Normalisation channel	$\pm 4.5$
Simulated sample size	$\pm 3.4$
Data-simulation differences	$\pm 1.3$
$B^0$ background	$\pm 1.0$
Quadratic sum	$\pm 13.1$

Table 6.1: Summary of the different contributions to the systematic uncertainty.

The total systematic uncertainty is determined to be 13.1 % by taking the quadratic sum of all contributions. A summary of all contributions is given in Table 6.1.

Although an uncertainty is accounted for the differences of simulated and data distributions, the weighting itself could contribute too. Calculating the weights is done in bins, where the size of the bins influences the relative error of the weights. A smaller bin size does not necessarily reduce this error, so a study of different bin sizes could be done to quantify the influence on the efficiencies.

Moreover, the corrections of the  $BDT_{output}$  cut efficiencies were done after striping, trigger- and preselection. Ideally, possible differences between data and simulation should be corrected before any cut is applied.

The chosen models to describe the shapes of each contribution to the data show good agreement with MC simulations, but do not describe the data well. This is tried to be solved by floating some parameters. Other models, e.g. a Bukin distribution for asymmetric peaks [28], could be tested to see if they fit better.

## 7 Summary and outlook

A measurement of the branching ratio of  $B^+ \rightarrow J/\Psi K^{*+}$ , using data samples with  $K^{*+} \rightarrow K^+ \pi^0$  and  $\pi^0 \rightarrow \gamma\gamma$  from Run I of the LHCb experiment, is presented in this thesis.

The branching fraction is measured relative to the decay  $B^0 \rightarrow J/\Psi K^{*0}$ . Because of the vast amount of combinatorial background produced in proton-proton collisions, this analysis focused on reducing combinatorial background. A pure signal sample of the control channel is obtained via the *sPlot* technique. It is used to assure a proper signal selection of a boosted decision tree. After this step, large contributions of partially reconstructed and peaking backgrounds are found. An attempt to describe these contributions is made by adding them to the final fit of the data, which is used to obtain the signal yield.

This results in an absolute branching ratio of

$$\mathcal{B}(B^+ \rightarrow J/\Psi K^{*+}) = (1.64 \pm 0.03_{stat.} \pm 0.21_{syst.}) \cdot 10^{-3}$$

and is within  $1\sigma$  agreement with the current world average

$$\mathcal{B}(B^+ \rightarrow J/\Psi K^{*+}) = (1.43 \pm 0.08) \cdot 10^{-3}.$$

The measurement as was performed in this thesis is dominated by the systematic uncertainty. Table 6.1 clearly shows that the uncertainties emerging from the fit result dominate the uncertainty. To further investigate this channel, e.g. for a direct CPV measurement, possible background sources have to be studied further. Additional MC data samples for the individual contributions can help to increase the accuracy of the fit, but also to find possible background and signal discriminating variables. This way, an additional BDT could be trained to reduce partially reconstructed and/or peaking background contributions.

The statistical uncertainty is already low, but can be further decreased by optimising the *BDToutput* cut. With data from Run II of LHCb, which started in June 2015, the statistical uncertainty will decrease even further.

# A Appendix

## A.1 Trigger configuration

Stage	Trigger
L0	L0MuonDecision
HLT1	Hlt1TrackAllL0Decision or Hlt1TrackMuonDecision or Hlt1DiMuonHighMassDecision
HLT2	Hlt2TopoMu2BodyBBDTDecision or Hlt2DiMuonDetachedJpsiDecision

Table A.1: Chosen trigger configuration for candidate events. A detailed description of the trigger lines is given in [21].

## A.2 $J/\Psi K^+$ veto

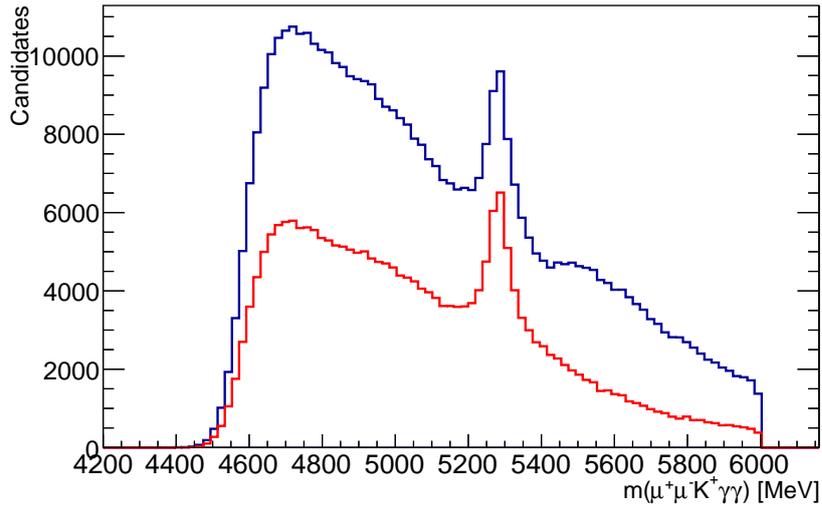


Figure A.1: Comparison of the  $B^+ \rightarrow J/\Psi K^{*+}$  mass distribution in 2012 Magnet Up data before and after the  $J/\Psi K^+$  veto. Only a loose selection was applied to make the contribution of  $B^+ \rightarrow J/\Psi K^+$  better visible. Distribution before (after) applying the veto on  $m(J/\Psi K^+)$  in blue (red).

### A.3 Fits of the normalisation channel

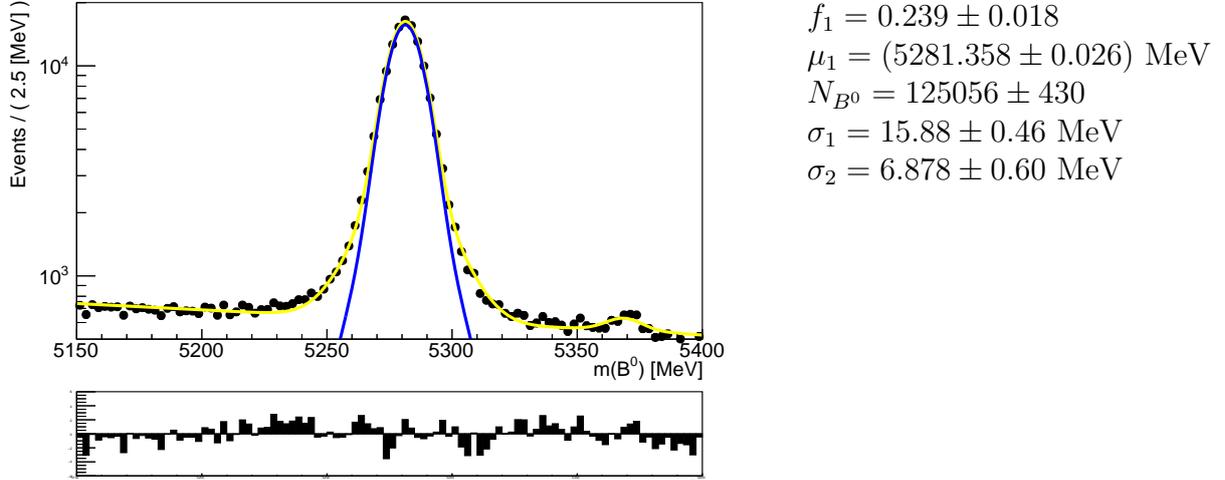
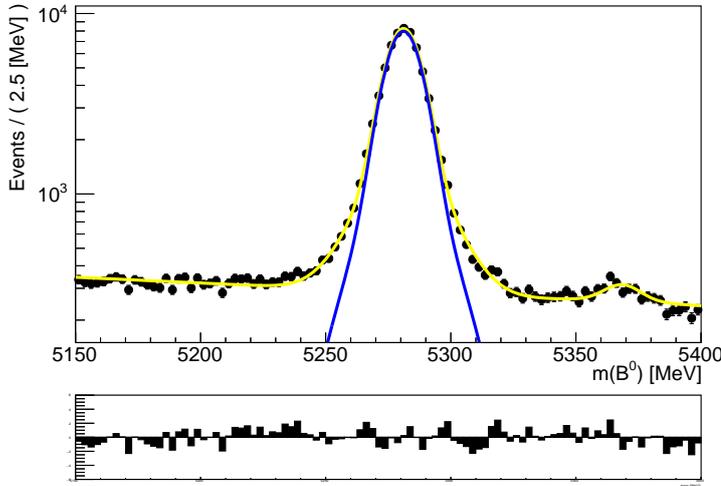
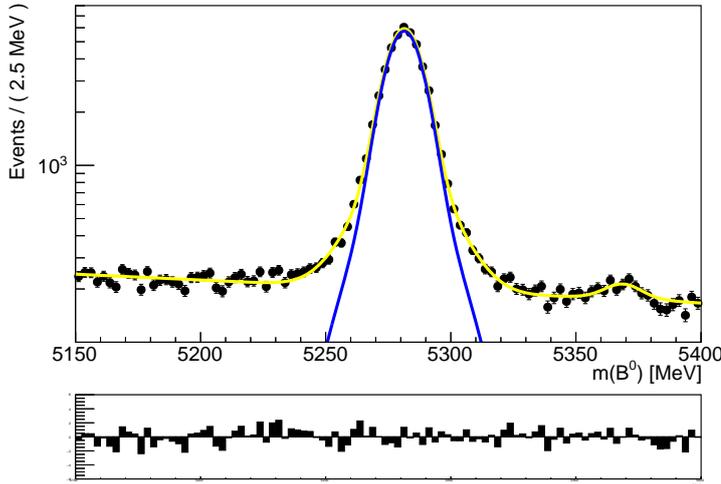


Figure A.2: Fit to the mass spectrum of selected  $B^0 \rightarrow J/\Psi K^{*0}$  candidates on 2012 Magnet Up data as in [22]. A double Gaussian is used for the signal contribution (blue). Combinatorial background is fitted with an exponential. The y-axis is logarithmic to show the small contribution of  $B_s^0 \rightarrow J/\Psi K^{*0}$ . It is fitted with the same model as for  $B^0 \rightarrow J/\Psi K^{*0}$ , but with a shifted mean of  $\Delta m = m(B_s^0) - m(B^0) = 87.35 \text{ MeV}$ . The full fit model is shown in yellow. The obtained fit values for the signal are shown on the right.



$$\begin{aligned}
 f_1 &= 0.756 \pm 0.017 \\
 \mu_1 &= (5281.081 \pm 0.037) \text{ MeV} \\
 N_{B^0} &= 64475 \pm 300 \\
 \sigma_1 &= 6.844 \pm 0.083 \text{ MeV} \\
 \sigma_2 &= 15.61 \pm 0.59 \text{ MeV}
 \end{aligned}$$

Figure A.3: Fit to the mass spectrum of selected  $B^0 \rightarrow J/\Psi K^{*0}$  candidates on 2011 Magnet Down data as in [22]. A double Gaussian is used for the signal contribution (blue). Combinatorial background is fitted with an exponential. The y-axis is logarithmic to show the small contribution of  $B_s^0 \rightarrow J/\Psi K^{*0}$ . It is fitted with the same model as for  $B^0 \rightarrow J/\Psi K^{*0}$ , but with a shifted mean of  $\Delta m = m(B_s^0) - m(B^0) = 87.35$  MeV. The full fit model is shown in yellow. The obtained fit values for the signal are shown on the right.



$$\begin{aligned}
 f_1 &= 0.236 \pm 0.019 \\
 \mu_1 &= (5281.390 \pm 0.043) \text{ MeV} \\
 N_{B^0} &= 45741 \pm 253 \\
 \sigma_1 &= 15.66 \pm 0.71 \text{ MeV} \\
 \sigma_2 &= 6.904 \pm 0.096 \text{ MeV}
 \end{aligned}$$

Figure A.4: Fit to the mass spectrum of selected  $B^0 \rightarrow J/\Psi K^{*0}$  candidates on 2011 Magnet Up data as in [22]. A double Gaussian is used for the signal contribution (blue). Combinatorial background is fitted with an exponential. The y-axis is logarithmic to show the small contribution of  $B_s^0 \rightarrow J/\Psi K^{*0}$ . It is fitted with the same model as for  $B^0 \rightarrow J/\Psi K^{*0}$ , but with a shifted mean of  $\Delta m = m(B_s^0) - m(B^0) = 87.35$  MeV. The full fit model is shown in yellow. The obtained fit values for the signal are shown on the right.

## B Bibliography

- [1] K. A. Olive et al. “Review of Particle Physics”. In: *Chin. Phys.* C38 (2014), p. 090001. DOI: 10.1088/1674-1137/38/9/090001. URL: <http://pdg.lbl.gov/>.
- [2] M. Thomson. *Modern particle physics*. Cambridge University Press, 2013.
- [3] W.N. Cottingham. *An Introduction to the Standard Model of Particle Physics*. Cambridge University Press, 2007.
- [4] MissMJ, Nasfarley88, et al., eds. *The Standard Model of Elementary Particles*. URL: <https://commons.wikimedia.org/w/index.php?curid=36335876> (visited on 07/30/2016).
- [5] R. Aaij et al. “Observation of  $J/\Psi p$  resonances consistent with pentaquark states in  $\Lambda_b^0 \rightarrow J/\Psi K^- p$  decays”. In: *Phys. Rev. Lett.* 115, 072001 (2015). eprint: <https://arxiv.org/abs/1507.03414>.
- [6] P. Kuehnl. “Determination of the branching ratio of the decay  $B^+ \rightarrow \mu\mu K^{*+} (\rightarrow K_s^0 \pi^+)$ ”. Bachelor thesis. University of Heidelberg, 2015.
- [7] LHCb Collaboration, Augusto Alves Jr, et al. “The LHCb Detector at the LHC”. In: *Journal of Instrumentation* 3.08 (2008), S08005. URL: <http://stacks.iop.org/1748-0221/3/i=08/a=S08005>.
- [8] *CERN public webpage*. URL: <http://home.cern/> (visited on 08/26/2016).
- [9] Lyndon Evans and Philip Bryant. “LHC machine”. In: *Journal of Instrumentation* 3.08 (2008), S08001.
- [10] Rolf Lindner. “LHCb layout\_2. LHCb schema\_2”. LHCb-PHO-GENE-2008-002. LHCb Collection. Feb. 2008. URL: <http://cds.cern.ch/record/1087860>.
- [11] LHCb Collaboration. *LHCb Silicon Tracker*. URL: <http://lhcb.physik.uzh.ch/ST/public/material/index.php> (visited on 08/29/2016).
- [12] Wouter D. Hulsbergen. “Decay chain fitting with a Kalman filter”. In: *Nuclear Instruments and Methods in Physics Research A* 552 (Apr. 2005). DOI: 10.1016/j.nima.2005.06.078. eprint: <http://arxiv.org/abs/physics/0503191>.
- [13] M Adinolfi et al. “Performance of the LHCb RICH detector at the LHC”. In: *The European Physical Journal C* 73.5 (2013), pp. 1–17. DOI: 10.1140/epjc/s10052-013-2431-9.
- [14] O Deschamps et al. *Photon and neutral pion reconstruction*. Tech. rep. CERN-LHCb-2003-091, 2003.
- [15] The LHCb Collaboration. *The LHCb High Level Trigger application*. URL: <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/moore/> (visited on 08/31/2016).

- [16] The LHCb Collaboration. *The LHCb event reconstruction application*. URL: <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/brunel/> (visited on 08/31/2016).
- [17] The LHCb Collaboration. *The LHCb physics analysis software*. URL: <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci/> (visited on 08/31/2016).
- [18] The LHCb Collaboration. *The LHCb simulation software*. URL: <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/gauss/> (visited on 08/31/2016).
- [19] NIST/SEMATECH. *e-Handbook of Statistical Methods*. URL: <http://www.itl.nist.gov/div898/handbook/> (visited on 08/15/2016).
- [20] G. Lanfranchi et al. *The Muon Identification Procedure of the LHCb Experiment for the First Data*. LHCb-PUB-2009-013, CERN-LHCb-PUB-2009-013. 2009.
- [21] R Aaij et al. “The LHCb trigger and its performance in 2011”. In: *Journal of Instrumentation* 8 (Apr. 2013). DOI: 10.1088/1748-0221/8/04/P04022. eprint: <http://arxiv.org/abs/1211.3055>.
- [22] J. Grabowski. “Branching Fraction Measurement of the Decay  $B^+ \rightarrow J/\Psi \rho^+$  with the LHCb Experiment”. Master thesis. University of Heidelberg, 2016.
- [23] Muriel Pivk and Francois R. Le Diberder. “sPlot: a statistical tool to unfold data distributions”. In: *Nuclear Instruments and Methods in Physics Research A* 555 (Dec. 2005). DOI: 10.1016/j.nima.2005.08.106. eprint: <https://arxiv.org/abs/physics/0402083>.
- [24] A. Hoecker et al. “TMVA: Toolkit for Multivariate Data Analysis”. In: *PoS ACAT* 040 (2007). eprint: <https://arxiv.org/abs/physics/0703039>.
- [25] Eric W. Weisstein. *Chebyshev Polynomial of the First Kind*. URL: <http://mathworld.wolfram.com/ChebyshevPolynomialoftheFirstKind.html>.
- [26] Wenbin Qian. *PIDCalib Packages*. URL: <https://twiki.cern.ch/twiki/bin/view/LHCb/PIDCalibPackage>.
- [27] S. Stahl. “Measurement of CP asymmetry in muon-tagged  $D^0 \rightarrow K^- K^+$  and  $D^0 \rightarrow \pi^- \pi^+$  decays at LHCb”. Dissertation. University of Heidelberg, 2014.
- [28] AD Bukin. “Fitting function for asymmetric peaks”. In: (2007). eprint: <http://arxiv.org/abs/0711.4449>.

## C Acknowledgement

First, I want to thank everyone who supported me during the last few months. Many thanks to my supervisor Prof. Dr. Stephanie Hansmann-Menzemer not only for giving me the opportunity to work in the LHCb group at the University of Heidelberg, but also for being a great help with the writing of this thesis. I also want to thank Prof. Dr. Klaus Reygers for agreeing to be the second referee of my thesis.

Special thanks go to Dr. Sevda Esen and Dr. Michel De Cian, the advisers for my thesis. Whenever I needed some guidance or was stuck, you took some time to help me out. I learned a lot during my time here and really appreciate having you as advisers. Thank you! Apart from challenging physic topics, programming caused quite a few difficulties. Thank you Jascha Grabowski for helping me out with these kinds of problems.

Last but not least, I want to thank my family and friends for their support during my work on this thesis, but also in all other aspects of life. Caroline Bender, Joy Lauter, Lukas Herrmann, Nadine Abendschein, Jannis Rott and Nils Schmidgruber, I still remember how we became friends in school, and even though everyone took their own path, not much has changed. Thank you for making the past years very special!

Impossible to forget in these acknowledgements is Jan Trautmann. We met each other at the first days of University and went through all of this together, studying together, sharing notes and anxiety before exams, but also enjoying some well-deserved beers together afterwards. You became a good friend, thank you for the great time. Here's to the next two years!

Moreover, I want to thank Anne Winkens. You are not only one of the best friends one could have, you also put some great effort in reading my thesis and giving me valuable tips. Dank je wel!

Am meisten möchte ich jedoch meinen Eltern danken. Ihr habt mich bis jetzt in jedem Aspekt meines Lebens unterstützt und an mich geglaubt. Eure Unterstützung hat es mir ermöglicht meinen Träumen nachzugehen. Dafür bin ich euch unendlich dankbar!

# Erklärung

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

Heidelberg, den 13.09.2016

Daniel Baitinger