Department of Physics and Astronomy

University of Heidelberg

Master thesis

in Physics

submitted by

Jascha Grabowski

born in Coburg

2016

Branching Fraction Measurement of the Decay $B^+ \to J/\psi \rho^+$ with the LHCb Experiment

This Master thesis has been carried out by

Jascha Grabowski

at the

Physikalisches Institut

under the supervision of

Prof. Dr. Ulrich Uwer

Dr. Michel De Cian

Dr. Sevda Esen

Messung des Verzweigungsverhältnisses des Zerfalls $\mathbf{B^+ \to J/\psi\,\rho^+}$ mit dem LHCb Experiment:

In dieser Arbeit wird die Messung des Verzweigungsverhältnisses des Zerfalls $B^+ \to J/\psi\,\rho^+$ vorgestellt. Es wurden Daten vom LHCb Experiment verwendet, die einer integrierten Luminosität von $3\,\mathrm{fb}^{-1}$ entsprechen, aufgenommen bei Schwerpunktsenergien von 7 TeV und 8 TeV. Für diese Analyse wurden ρ^+ Zerfälle in $\pi^+\pi^0$ und im Folgenden $\pi^0 \to \gamma\gamma$ verwendet. Die Messung erfolgte relativ zum Zerfall $B^0 \to J/\psi\,K^{*0}$ mit $K^{*0} \to K^+\pi^-$. Die Anzahl von Zerfallsereignissen wurde mit ungebinnten Maximum-Likelihood-Fits auf das Massenspektrum der B-Kandidaten bestimmt, woraus sich für das Verzweigungsverhältnis der folgende Wert bestimmt:

$$\mathcal{B}(B^+ \to J/\psi\,\rho^+) = (4.31 \pm 0.22(\mathrm{stat})\ ^{+0.49}_{-0.88}(\mathrm{syst})) \times 10^{-5}.$$

Dieses Ergebnis steht im Einklang mit der einzigen bisherigen Messung durch die BaBar Kollaboration (siehe Ref. [1]).

Branching Fraction Measurement of the Decay $\mathbf{B^+ \to J/\psi\,\rho^+}$ with the LHCb Experiment:

This thesis presents the measurement of the branching fraction of the decay $B^+ \to J/\psi\,\rho^+$ using data from the LHCb experiment corresponding to an integrated luminosity of $3\,\mathrm{fb}^{-1}$ taken in proton-proton collisions at center-of-mass-energies of 7 TeV and 8 TeV. For this analysis ρ^+ decays into $\pi^+\pi^0$ and subsequently $\pi^0 \to \gamma\gamma$ were used. The measurement was performed relative to the decay $B^0 \to J/\psi\,K^{*0}$, where $K^{*0} \to K^+\pi^-$. The yields were determined from unbinned maximum likelihood fits to the mass spectrum of the B-candidates. The branching ratio is found to be

$$\mathcal{B}(B^+ \to J/\psi\,\rho^+) = (4.31 \pm 0.22(\mathrm{stat})\ ^{+0.49}_{-0.88}(\mathrm{syst})) \times 10^{-5}.$$

This result is consistent with the only previous measurement by the BaBar collaboration (see Ref. [1]).

Contents

1 Introduction **1**

2 Theoretical Overview **3**
2.1 The Standard Model of Particle Physics 3
2.2 Flavour Physics 5
2.3 CP-violation 6

3 The LHCb Experiment **9**
3.1 The Large Hadron Collider 9
3.2 The LHCb Detector 9
3.2.1 Magnet 9
3.2.2 Tracking System 10
3.2.3 Particle Identification System 13
3.2.4 Trigger 17

4 Analysis Strategy **19**
4.1 Fitting Procedure 21
4.2 Variables 21

5 Signal Monte Carlo Calibration **25**
5.1 PID Resampling 25
5.2 $B^0 \to J/\psi K^{*0}$ Normalisation Signal Extraction 26
5.3 $B^+ \to J/\psi \rho^+$ Monte Carlo Reweighting 28
5.4 $B^+ \to J/\psi K^{*+}$ Control Signal Extraction 29

6 Event Selection **37**
6.1 Background Sources 37
6.1.1 Combinatoric Background 37
6.1.2 Peaking Background 37
6.1.3 Partially reconstructed Background 37
6.1.4 Non-resonant $B^+ \to J/\psi \pi^+\pi^0$ 39
6.2 Trigger Selection 39
6.3 Stripping and Preselection 40
6.4 Boosted Decision Tree 41
6.4.1 BDT Against Combinatoric Background 43
6.4.2 BDT Against Partially Reconstructed Background 46
6.5 Efficiencies 47
6.6 Cut Optimisation 48
6.7 Fit Description 53

7 Systematic Uncertainties **57**

8 Branching Fraction Determination **61**

9 Summary and Outlook **63**

A Appendix **64**
A.1 ${}_s\mathcal{P}lot$ Technique 64
A.2 Comparison between BDT Output for Signal Simulation with Normalisation Channel Data 64
A.3 Efficiency Tables 65
A.4 Detailed Fit Results 65

References **70**

1 Introduction

Modern physics is very successful at describing the interactions of particles using a theoretical framework called Standard Model. But this model is not the final answer, there are certain phenomena it cannot describe. A prominent example is the observed asymmetry of matter over anti-matter in the universe. The Standard Model includes so called CP-violation, which leads to a different treatment of matter over antimatter. This can in principle account for such an asymmetry, but the amount of asymmetry predicted from the model is too small to describe the observed universe. The main contribution must come from not-yet understood processes called New Physics.

In the last decades it was possible to deepen the understanding of CP-violation with precision measurements and new theory results on strange and beauty hadrons. Today it is possible to extract the Standard Model parameters responsible for CP-violation from various independent sources. One can therefore overconstrain the model to see, if an inconsistency occurs, which would show, where the model has to be improved and New Physics could come into play. Until now, all measurements are consistent with each other, which requires effects from New Physics to be so small that they could not be resolved yet. Measurements with higher precision are therefore necessary to get on to the track of New Physics.

The LHCb experiment is one of the four big physics experiments at the large Hadron Collider (LHC) near Geneva. It was designed for the study of physics with hadrons containing a b-quark. $B^+ \to J/\psi\,\rho^+$ with $J/\psi \to \mu^+\mu^-$ and $\rho^+ \to \pi^+\pi^0$, the decay considered in this thesis is an example thereof. It has a clean signature from the two muons and the Standard Model predicts a small direct CP violation, which would show itself in a non-zero charge asymmetry $A_{\mathrm{CP}} \lesssim 1\%$ (see Ref. [2]). To access this value one first measures the raw charge asymmetry in the decay:

$$A_{\text{raw CP}} = \frac{N(B^+ \to J/\psi\,\rho^+) - N(B^- \to J/\psi\,\rho^-)}{N(B^+ \to J/\psi\,\rho^+) + N(B^- \to J/\psi\,\rho^-)}$$

This value has then to be corrected for differences in material interactions of differently charged pions and production asymmetries between B^+ and B^-. A measurement by the BaBar collaboration resulted in $A_{\mathrm{CP}} = -0.11 \pm 0.12 \pm 0.08$ (see Ref. [1]), but with the larger sample of B-mesons from Run I at LHCb this precision can be exceeded, tightening the constraints on the Standard Model.

This thesis is the starting point for these measurements: It provides the signal selection procedure and the extraction of the branching ratio of this decay, which was established in only one previous measurement by BaBar to be $\mathcal{B}(B^+ \to J/\psi\,\rho^+) = (5.0 \pm 0.7_{\mathrm{stat}} \pm 0.3_{\mathrm{syst}}) \times 10^{-5}$ (see Ref. [1]). It will be shown that the precision on this value can be exceeded as well.

The thesis is structured as follows: After the theory is outlined in Sec. 2 and

the LHCb experiment is introduced in Sec. 3, the signal simulation validation and reweighting is explained in Sec. 5. The selection procedure of the signal candidates is detailed in Sec. 6. Here also the yield extraction by a fit to the B-candidate mass spectrum is described. Systematic uncertainties are discussed in Sec. 7 and the results for the branching ratio are presented in Sec. 8.

2 Theoretical Overview

In this section the basic concepts of particle physics are presented. The Standard Model of particle physics (SM) is introduced with a focus on flavour physics. A detailed introduction into the experimental side of the Standard Model can be found in Ref. [3], for an introduction into the underlying field theory see Ref. [4].

2.1 The Standard Model of Particle Physics

Throughout the second half of the 20th century the theory of particle physics was advanced into a common framework that is self-consistent and could describe the experimental data. This so-called Standard Model of particle physics is a Quantum Field Theory (QFT), where particles and interactions are described with fields and their quantised excitations.

There are 12 fundamental spin $\frac{1}{2}$ particles (fermions) in the SM: The 6 quarks (up, down, charm, strange, top, bottom) and 6 leptons (electron, electron neutrino, muon, muon neutrino, tauon, tauon neutrino), each having its own antiparticle with opposite-sign charges (see Fig. 1). The particles can be distinguished by the charges they carry with respect to the 3 SM forces (weak, strong, electromagnetic).

Quarks carry color charge, which means they interact via the strong interaction. Free colored objects have never been observed (colour confinement) so quarks always form composite objects (hadrons), which contain either one quark and one antiquark (mesons), or 3 quarks (baryons) or – more exotically– four or five quarks, like the so-called pentaquarks, whose existence was finally established in 2015 (see Ref [6]). They either have an electric charge of $Q = +\frac{2}{3}$ and weak isospin $T_3 = +\frac{1}{2}$ (up-type quarks) or $Q = -\frac{1}{3}$ and weak isospin $T_3 = -\frac{1}{2}$ (down-type quarks). They are arranged in 3 families with one up-type quark and one down-type quark each (more about quark families in Sec 2.2).

Ordinary matter is mostly made up of first generation quarks, the up-quark (u) and the down-quark (d), which have masses of a few MeV. Other quarks have higher masses, up to 173 GeV for the top quark (t). The bottom-quark (b) is especially important for this thesis, because it can bond with the much lighter u and d quarks to form the mesons B^+ and B^0. Due to its high mass of about 4 GeV it dominates the interaction inside the meson (u and d only have masses of a few MeV). This facilitates the use of approximations and symmetries in theoretical calculations of the B-meson system that are not valid for mesons with lower masses. As b quarks cannot decay inside their family (t quarks are heavier than b-mesons), b-mesons are relatively long-lived ($\approx 10^{-12}$ s) and exhibit interesting features like CP-violation (see Sec.2.3). B-mesons are therefore interesting both from a theoretical and experimental perspective.

The remaining fermions are called leptons, they do not carry colour charge and can be split into two groups: The so-called neutrinos do not have an electric

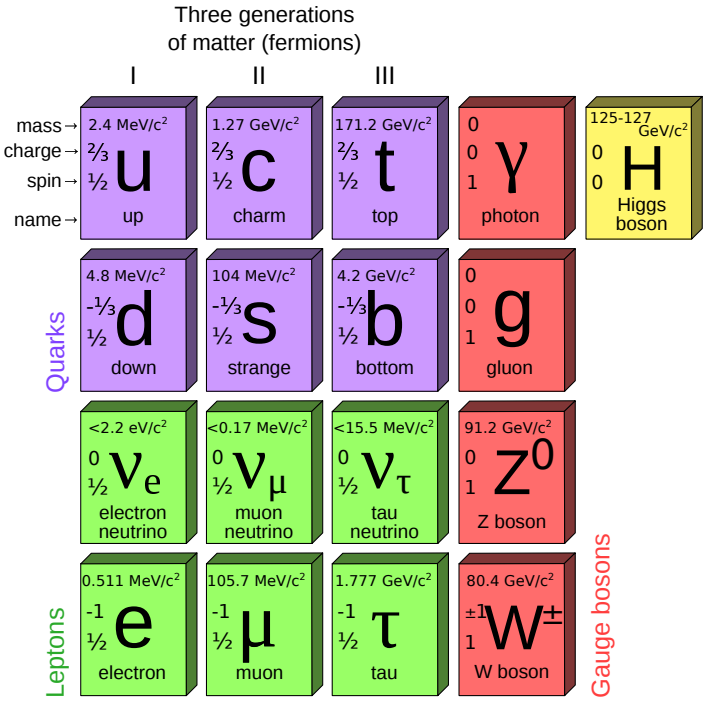

Figure 1: Fundamental particles in the Standard Model of Particle Physics. Figure taken from Ref. [5]

charge either, so they only interact via the weak interaction. The electron (e), muon (μ), and tauon (τ) on the other hand have an electric charge of $Q = -1$.

Interaction between the fermions is mediated by the exchange of spin 1 particles called gauge-bosons, which correspond to generators of the local gauge transformations under which the SM is invariant. The electromagnetic force is realized by the emission and absorption of massless photons (γ) generated by a U(1)-symmetry. The weak force employs the heavy W^+, W^-, and Z bosons generating the $\mathrm{SU}(2)_L$ (and right-handed antiparticles). The W-bosons couple only to left-handed particles and carry charges themselves, so they change the charges of the particle that has emitted them and can interact with the gauge bosons of weak and electromagnetic interaction themselves. The strong interaction is introduced into the SM by symmetry under the $\mathrm{SU}(3)_C$ group. It is described by the exchange of 8 massless bosons called gluons, which themselves carry color charge and therefore not only interact with quarks but with each other as well.

The last fundamental particle of the SM being discovered, is the so-called Higgs boson, a spin 0 charge 0 excitation of a component of the Higgs field. This field is used to explain the masses of all massive fundamental SM particles

(i.e. all but photons, gluons and possibly neutrinos[1]) and the structure of the electromagnetic and weak interaction via the Higgs mechanism, a form of spontaneous symmetry breaking as broached in Sec. 2.2.

Although the SM is very successful in describing the experimental data, it cannot be a final theory. For example it does not describe gravitational interaction and does not contain particles, which could make up the observed amount of Dark Matter in the universe. It is therefore necessary to probe the SM, to see at which point its predictions break down and how a new model has to look like that does not exhibit these flaws.

2.2 Flavour Physics

Through electroweak symmetry breaking the quarks acquire masses due to their Yukawa coupling to the Higgs field:

$$\mathcal{L}_{\text{Yukawa}} = y^d_{ij}\bar{Q}^i_L\phi D^j_R + y^u_{ij}Q^j_L\tilde{\phi}U^j_R + \text{h. c.} \tag{1}$$

with the Higgs field denoted as ϕ, the left-handed quark doublet

$$Q^i_L = \begin{pmatrix} u^i_L \\ d^i_L \end{pmatrix} \tag{2}$$

with i denoting the 3 different quark generations and the right-handed quark singlets U^j_R and D^j_R. y^d_{ij} and y^u_{ij} represent the coupling constants. ϕ has a constant component called vacuum expectation value (vev), which results in effective mass terms for the quarks. To write down these terms properly, one needs to diagonalize the matrices $y^{ij}_{d,u}$, but this is not possible for both matrices at the same time. By convention one chooses to have a diagonal up-type quark matrix. For the down-type quarks the mass basis (d, s, b) is then rotated from the interaction basis (d', s', b') by the Cabibbo-Kobayashi-Maskawa (CKM) matrix V_{CKM}:

$$d'^i = V^i_j d^j \tag{3}$$

$$u'^i = \delta^i_j u^j; \tag{4}$$

This rotation modifies the coupling of the W-bosons to quarks:

$$\mathcal{L}_{W\,qq} = \frac{ig}{2\sqrt{2}}\bar{u}'^i_L\gamma_\mu(1-\gamma^5)d'^j_L W^\mu + \text{h.c.} = \frac{ig}{\sqrt{2}}V_{ij}\bar{u}^i_L\gamma_\mu(1-\gamma^5)d^j_L W^\mu + \text{h.c.} \tag{5}$$

g is the coupling constant of the charged weak interaction, γ^μ and γ^5 are the Dirac matrices and $\frac{1}{2}(1-\gamma^5)$ is the chiral projection operator that results

[1] If neutrinos are Dirac particles, they could get their masses through the Higgs mechanism, implying the existence of sterile right-handed neutrinos. If they are Majorana particles, their masses could not come from their coupling to the Higgs field. This topic is subject of ongoing discussion and research.

in the W-boson W^μ coupling exclusively to left-handed particles. For $i \neq j$ the coupling to the W-bosons transforms up-type quarks of one generation into down-type quarks of another generation and the other way around. The coupling strengths of these flavour changing processes depend on the value of the corresponding element of V_{CKM}.

V_{CKM} is a complex unitary matrix with 4 free parameters. The hierarchical order of the elements can best be seen in the so-called *Wolfenstein* parametrization, which expands V_{CKM} in the parameter $\lambda \equiv \frac{V_{us}}{\sqrt{V_{us}^2 + V_{ud}^2}} \approx 0.2$:

$$V_{\mathrm{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{6}$$

$$= \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \tag{7}$$

The other 3 parameters are of order 1 ($A \approx 0.8$, $\rho \approx 0.4$, $\eta \approx 0.34$). These values result in diagonal elements, which are approximately unity, and because the transition probability is proportional to the absolute square of the matrix element $|V_{ij}|^2$, transitions inside quark families are therefore favoured with respect to transitions between different families. For this reason the latter are called Cabibbo-suppressed.

2.3 CP-violation

The Standard Model is a quantum field theory and as such invariant under combined CPT-transformation. C, P, and T are not conserved individually, and combined CP violation is one of the necessary conditions for the observed matter-antimatter asymmetry in the universe (see Ref. [7]) making their study not only interesting for particle physics but also for cosmology.

The weak interaction Lagrangian is the sum of vectorial objects ($\bar{u}\gamma_\mu d$) and axial-vectorial objects ($\bar{u}\gamma_\mu\gamma^5 d$). Parity transformation P flips the sign of all spacial components of these objects, which results in a relative minus sign between vectorial and axial vectorial objects. Looking at eq. 5 one sees that this leads to a violation of parity symmetry in the weak sector of the SM.

Charge conjugation C flips the sign of all charge-like quantum numbers resulting in a relative minus sign between vectorial and axial vectorial part as well as Hermitian conjugation:

$$\bar{u}\gamma_\mu d \xrightarrow{C} \bar{d}\gamma_\mu u \tag{8}$$

$$\bar{u}\gamma_\mu\gamma^5 d \xrightarrow{C} -\bar{d}\gamma_\mu\gamma^5 u \tag{9}$$

Plugging this into eq. 5 one sees that indeed the weak interaction violates symmetry under charge conjugation.

A combination of both C and P transformation cancels the relative minus sign between axial and vector objects. But CP does not transform the CKM matrix element:

$$V_{ij}\bar{u}^i\gamma_\mu(1-\gamma^5)d^j + V_{ij}^\dagger\bar{d}^j\gamma_\mu(1-\gamma^5)u^i \xrightarrow{CP} V_{ij}\bar{d}^j\gamma_\mu(1-\gamma^5)u^i + V_{ij}^\dagger\bar{u}^i\gamma_\mu(1-\gamma^5)d^j \tag{10}$$

The Lagrangian would therefore only be CP conserving, if $V_{ij} = V_{ij}^\dagger$, which could only be realized, if the CKM matrix had no complex entries. This is not the case: In this parametrization up to $\mathcal{O}(\lambda^4)$ there are two complex entries in the matrix: V_{ub} and V_{td} From these therefore the biggest contributions to CP violation are expected (see eq. 7).

CP violation leads to three different effects:

- CP violation in decay (direct CP violation) comes from interference between decay amplitudes and leads to different branching ratios for CP conjugated decays, for example $\mathcal{B}(B^+ \to \eta K^+) \neq \mathcal{B}(B^- \to \eta K^-)$.
- CP violation in mixing (indirect CP violation), for example neutral mesons like the B^0 can oscillate to their anti-partner $\overline{B^0}$, but the oscillation probability is different for the process $\overline{B^0} \to B^0$ than for $B^0 \to \overline{B^0}$.
- CP violation in interference between decays with and without mixing comes from interference between the direct decay $B \to f$ and the decay after an oscillation into the same final state $B \to \overline{B} \to f$.

The decay in this analysis $B^+ \to J/\psi\,\rho^+$ proceeds mainly via the two Feynman diagrams shown in Fig. 2. The left diagram shows the tree level decay containing V_{cb} and V_{cd}, which do not have an imaginary part at $\mathcal{O}(\lambda^4)$. This decay amplitude interferes with the penguin diagram on the right. It has a virtual quark loop, in which V_{ub} and V_{td}, the most significant complex components of the CKM-matrix, can contribute. This interference can result in direct CP-violation and would be experimentally accessible by different branching fractions for the charge conjugated decays: $\mathcal{B}(B^+ \to J/\psi\,\rho^+) \neq \mathcal{B}(B^- \to J/\psi\,\rho^-)$. One measures the raw charge asymmetry

$$A_{\text{raw CP}} = \frac{N(B^+ \to J/\psi\,\rho^+) - N(B^- \to J/\psi\,\rho^-)}{N(B^+ \to J/\psi\,\rho^+) + N(B^- \to J/\psi\,\rho^-)} \tag{11}$$

This value has then to be corrected for differences in material interactions of differently charged pions and production asymmetries between B^+ and B^- resulting in:

$$A_{\text{CP}} = \frac{\mathcal{B}(B^+ \to J/\psi\,\rho^+) - \mathcal{B}(B^- \to J/\psi\,\rho^-)}{\mathcal{B}(B^+ \to J/\psi\,\rho^+) + \mathcal{B}(B^- \to J/\psi\,\rho^-)}$$

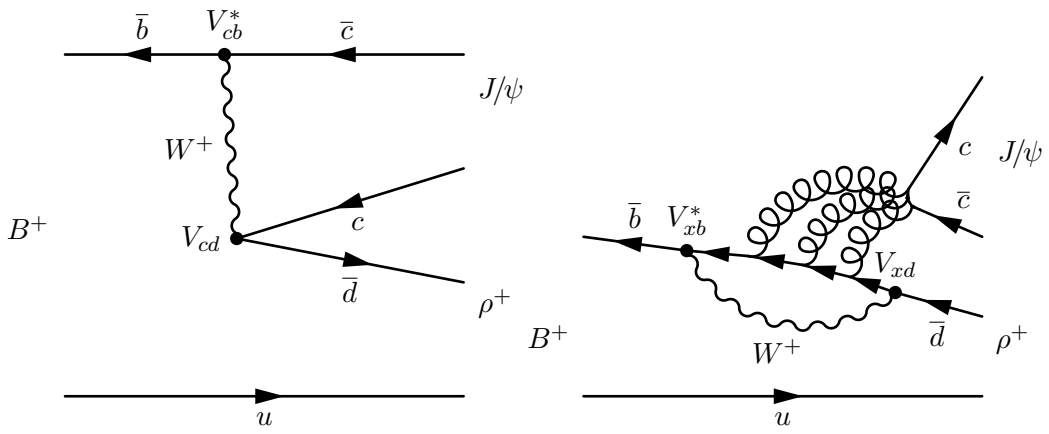

Figure 2: Left: tree level diagram of $B^+ \to J/\psi\,\rho^+$ with a $b \to c$–transition and a $c \to d$–transition. Right: penguin diagram of $B^+ \to J/\psi\,\rho^+$ with an up–type quark loop.

The small branching fraction of $\mathcal{B} = (5.0 \pm 0.8) \times 10^{-5}$ together with a low π^0 reconstruction efficiency in the subsequent decay $\rho^+ \to \pi^+\pi^0$ make this value hard to measure experimentally, the only measurement of A_{CP} by the BaBar collaboration determined a value of $A_{\mathrm{CP}} = 0.11 \pm 0.14$ (see Ref. [1]). A value of $A_{\mathrm{CP}} \lesssim 1\%$ is predicted theoretically (see Ref. [2]).

At LHCb it is possible to measure this value more precisely due to the larger sample of B^+. The actual determination of A_{CP} exceeds the scope of this thesis. The thesis prepares the ground for the actual measurement of A_{CP}, and focuses on the signal selection and branching ratio determination.

3 The LHCb Experiment

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a synchrotron located at the European Organization for Nuclear Research laboratory (CERN) near Geneva, Switzerland [8]. Protons are bunched and accelerated inside two distinct vacuum pipes and then brought to collision at four interaction points, where the major physics experiments are situated: ALICE [9], ATLAS [10], CMS [11], and LHCb [12]. In Run 1 the collisions took place at center of mass energies of 7 TeV in 2011 and 8 TeV in 2012. 1374 bunches per beam were cycling simultaneously with a separation of 50 ns, each consisting of 1.6×10^{11} protons. The data taken during that time was used for this analysis corresponding to an integrated luminosity of $3\,\text{fb}^{-1}$. Since 2015 the collisions take place at a center of mass energy of 13 TeV.

3.2 The LHCb Detector

The LHCb detector is designed to detect hadrons containing b or c quarks, which are produced mainly by gluon-gluon interaction. The two gluons usually carry very different fractions of their respective proton's momentum, which results in a strong boost of the created particles in the direction of the beams. LHCb exploits this fact by only covering the pseudorapidity range $2 < \eta < 5$, where a large fraction of the boosted particles are emitted. Fig. 3 shows a cross section of the LHCb detector. The z-axis is defined in the direction of the beam, the y-axis points in the vertical direction, while the x-axis horizontally completes a righthanded coordinate system. The main components of the detector will be explained in the following based on information from Ref. [12].

3.2.1 Magnet

LHCb's magnet consists of two identical saddle-shaped normal-conducting aluminium coils placed mirror-symmetrically above and below the beam pipe, which results in a bending of charged particles mostly in the x–z–plane (see Fig 3). Its magnetic field integral along the z-axis is approximately 4 Tm for a typical flight path of 10 m through the magnet. The track of a charged particle traversing is bent perpendicular to the field lines. The resulting kink between a particle's track before the magnet and behind it is used to calculate the particle's momentum p. The relative momentum resolution for particles traversing the whole tracking system (see Sec 3.2.2) is $\Delta p/p = 0.4\,\%$ at $p = 2\,\text{GeV}$ and $\Delta p/p = 0.6\,\%$ at $p = 100\,\text{GeV}$. The polarity of the magnet is switched periodically during data taking to decrease systematic uncertainties caused by detector asymmetries.

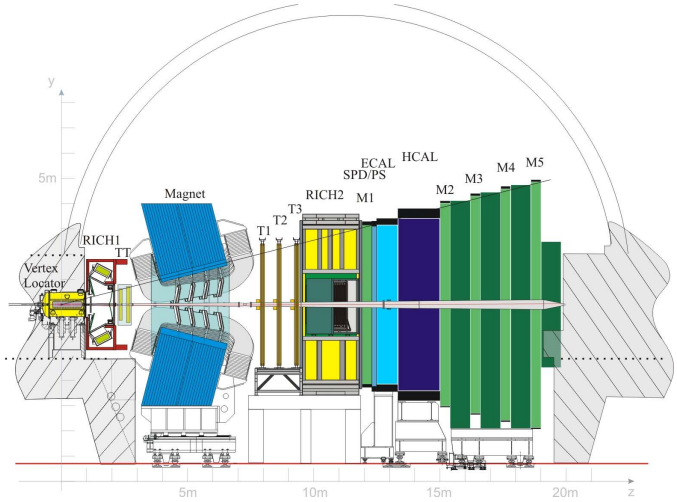

Figure 3: Cross section of the LHCb detector. Figure from [13]

3.2.2 Tracking System

The tracking is performed with information from different subsystems: The vertex locator system (VELO) around the interaction region, the Tracker Turicensis (TT) before the magnet section, and the Tracking stations (T1-T3) behind the magnets. VELO, TT, as well as the inner section of the T stations use silicon microstrip detectors for the detection of the traversion of charged particles. In the outer parts of the T stations so-called straw-tubes are used.

Vertex Locator The Vertex Locator (VELO) measures the position of tracks near the interaction point with the intention to identify the primary vertices (PV) and possible displaced secondary vertices. Many tracks intersect, where the *pp*-collision takes place, because a large number of charged particles is created there. *b*-hadrons produced in the collision usually travel about 1 cm before they decay and thereby form a secondary vertex, which can then be used to calculate the decay time of the particle and distinguish its daughter particles from background particles created directly at the interaction point. The VELO consists of 21 pairs of modules placed concentrically along the beam axis measuring the radial and azimuthal position in the x-y-plane of traversing charged particles (see Fig 4). They are moved near the beam axis during each fill of the LHC as soon as the beam is stable and not likely to damage the detector. Their active region then starts at about 8 mm from the beam axis and is able to measure tracks and vertices near the interaction point up to a precision of 4 µm [14].

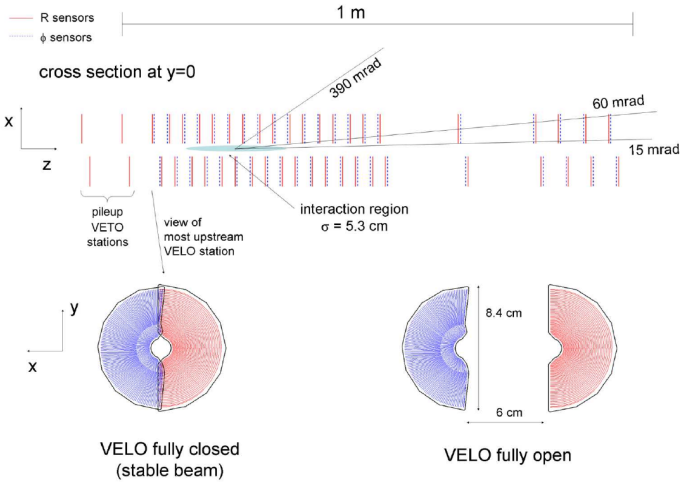

Figure 4: Top: Schematic of the VELO modules relative to interaction region. Bottom: Front view of module pair in closed (left) and open position (right). Figure from [12]

Silicon Tracker The Tracker Turicensis (TT) is located downstream of the VELO in front of the magnet and consists of two stations with two silicon microstrip detector layers each. The stations are separated by about 27 cm from each other. They have a strip pitch of about 200 µm, resulting in a hit resolution of about 50 µm. The strips are arranged vertically in the first and last layer, the layers in the middle are tilted by $\mp 5°$ in the x–y-plane to allow for spacial resolution in y-direction as well (see Fig 5).

The Inner Tracker (IT) is located behind the magnets in the inner region of the tracking stations T1-T3, where the particle flux is too high for the straw tube technology used for the outer part of the stations (see Sec 3.2.2). Each of the three tracking stations contains four detector layers. The layers are tilted in the same scheme as the 4 layers of the TT. The IT uses a similar technology as the TT and has approximately the same hit resolution.

Outer Tracker The outer part of the T stations consists of straw tube detectors. They are made up from parallel thin straw tubes 2.4 m long and with inner diameter of 4.9 mm. Through each tube, a 25 µm thin tungsten anode wire is guided, set to a voltage of +1550 V, the cathode is made of a 40 µm foil of conducting carbon doped polyimide (Kapton-XC). The tubes are filled with a gas mixture of Argon, CO_2, and oxygen (70 %, 28 %, 2 %) which reduces the drift time to below 50 ns. The hit position of a charged particle traversing the tube is determined by measuring the time the ionized particles take to drift to the anode wire. Knowing the collision time this drift time can be converted into a radial distance to the wire. The effective resolution of the OT is approximately

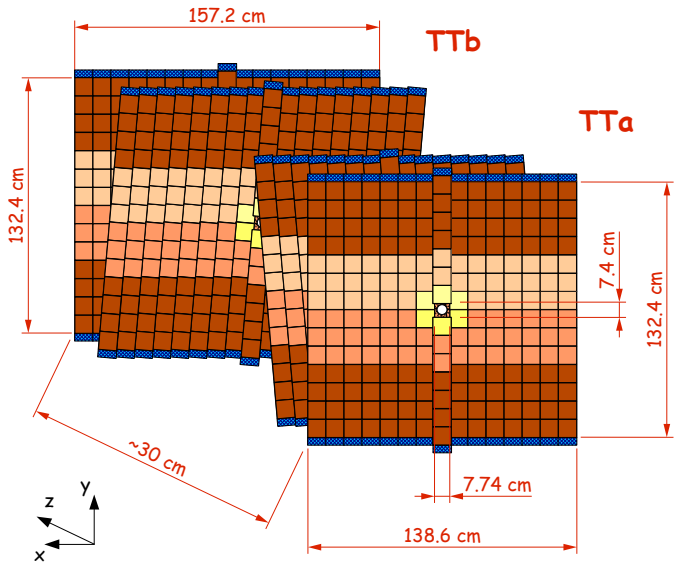

Figure 5: Layout of the Tracker Turicensis. Figure from [15]

3 ns, which translates to a hit position resolution of 200 µm. Tubes are put into two staggerd layers (see Fig 6 left) (monolayers). Each tracking station contains 4 monolayers oriented after the same scheme as IT and TT with the inner layers tilted with respect to the vertical axis to allow for resolution in y-direction. A schematic drawing of TT, IT, and OT can be found in Fig 6 on the right.

Track Reconstruction Hits of all the introduced detectors are combined to reconstruct the trajectories of traversing charged particles. In this analysis only so-called long tracks were used, which require the track to contain hits in both VELO and tracking stations. The reconstruction starts with searching the VELO for hits along straight lines and then combining them with hits in the T stations. The magnetic field points mainly in y-direction, so the flight direction in the y–z-plane does not change much between VELO and T station. If hits in the TT can be associated with the track, these are included as well to enhance the momentum estimation (a detailed description of the track reconstruction can be found in Ref. [16]). A Kalman filter algorithm is used to fit the hits to a track including corrections for multiple scattering in the detector material (see Ref. [17]). The χ^2/n_{dof} of the fit is used as a quality criterion of the track.

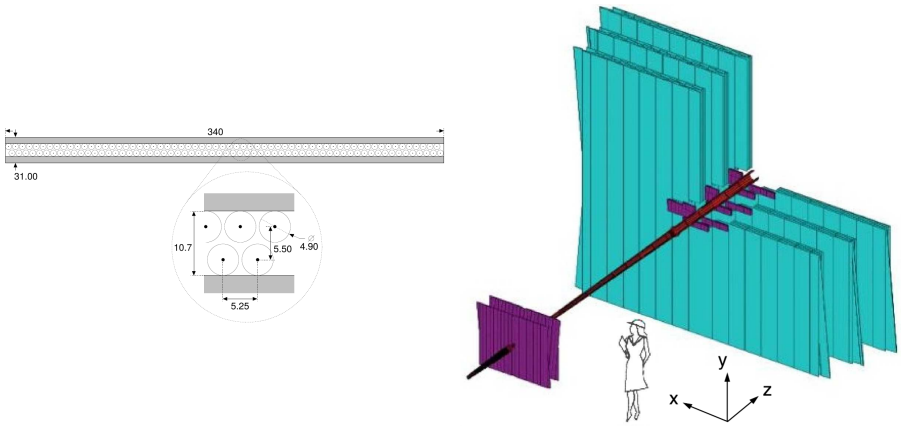

Figure 6: Left: Layout of the monolayer consisting of two stacked layers of straw tubes. Taken from [12]. Right: Layout of the Tracking Stations (OT in cyan and IT in purple) and TT in purple. Taken from [12].

3.2.3 Particle Identification System

To establish a mass hypothesis of final state particles, they need to be identified. This is done using information from various subsystems of the detector: muon chambers to identify particles as muons, two Ring Imaging Cherenkov (RICH) detectors to determine the species of charged hadrons, and the calorimeter system, which distinguishes electrons, photons and hadrons. They are collectively called particle identification (PID) system. For every possible particle hypothesis a likelihood is calculated from data.

The most abundant particles at LHCb are charged pions. Therefore the calculated likelihoods are all divided by the likelihood for the pion hypothesis. Since the possible likelihood ratios are very spread, the logarithm of the ratio is taken: $DLL_{X\pi} = \log \mathcal{L}_X - \log \mathcal{L}_\pi$. In the following the subsystems of the PID system are introduced.

RICH The RICH system uses the Cherenkov effect to distinguish between charged particles of different masses, mainly between pions and kaons, which are produced abundantly in the pp-collision. If a charged particle with a velocity v moves through a medium of a refractive index n and v is higher than the speed of light in that medium ($c' = c/n$), photons are emitted from the particle. Together with the momentum p measured in the tracking system (see 3.2.2) the emission angle of the Cherenkov photons $\theta = \arccos(c/vn)$ with respect to the particle's flight direction can be used to establish a mass hypothesis (see Fig. 7 on the right). It shows that for a given refractive index the difference in Cherenkov angle between particle species diminishes with growing momentum

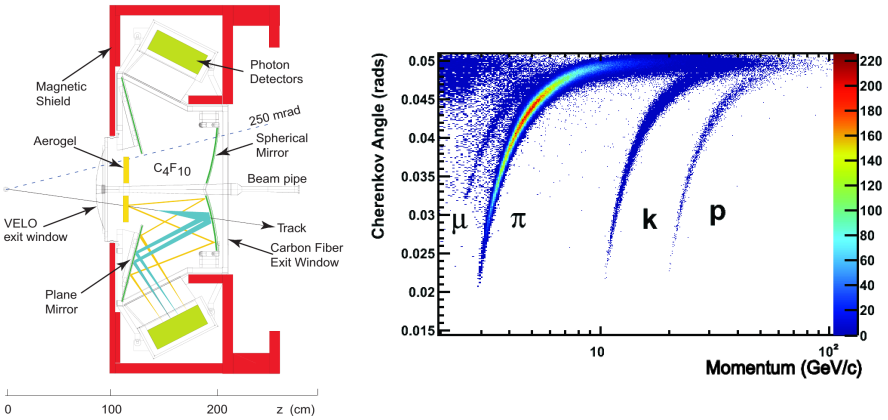

Figure 7: Left: vertical view of RICH 1 detector. Taken from [14]. Right: Reconstructed Cherenkov angle over track momentum in RICH 1. Taken from [18].

and makes it harder to distinguish between them. For this reason two RICH detectors are used in LHCb: RICH 1 covers the low and intermediate momentum region of 2–40 GeV in the full geometric acceptance region of LHCb and RICH 2 covers the region of 15–100 GeV in a smaller geometric window around the beampipe. The smaller geometric acceptance of RICH 2 is justified with the higher momenta of the particles it is designed to detect. Those have higher boosts and are deflected less by the magnetic field.

Different media are in use inside the systems to make them sensitive to their respective momentum regions: in RICH 1, silica aerogel ($n = 1.03$) and gaseous C_4F_{10} ($n = 1.0014$) and in RICH 2 gaseous CF_4 ($n = 1.0005$). The optical systems in both RICH detectors are similar: The photons emitted from the particle in a cone are reflected out of the detector acceptance region with focusing tilted spherical mirrors and secondary flat mirrors (see Fig 7 on the left). This is done to reduce the material the particles have to traverse. Less material means smaller probability for the particles to scatter, which leads to better momentum resolution. In the focusing plane, hybrid photon detectors (HPD) are positioned, which measure the hit-positions of the Cherenkov photons, which are then converted into the emission angle.

Calorimeter Two calorimeters are used in LHCb, an electromagnetic calorimeter (ECAL) in front of a hadronic calorimeter (HCAL). They follow the same principle: Traversing particles interact with the calorimeter material which results in pair creation forming particle showers. These produce scintillation light in the active area of the calorimeter that is guided through wavelength shifting fibres to photomultiplier tubes (PMT), converting it into a voltage that is read out.

A Scintillating Pad Detector (SPD) is positioned before the calorimeters. It consists of scintillating plates, which only trigger a signal, if a charged particle travels through it.

Behind it, the PreShower (PS) detector is set which contains 15 mm of lead (2.5 radiation lenghts X_0) to create electromagnetic showers from electrons. Charged hadrons are less likely to produce a shower, which makes the PS an effective tool to distinguish electrons from charged pions. Behind the lead about 12 000 rectangular scintillating pads are placed to detect developed showers with a good enough resolution to map them to tracks.

The ECAL is built up from alternating layers of active scintillator material and absorptive layers to produce showers. The material sums up to 25 radiation lengths, so that electrons and photons deposit all their energy inside the ECAL. The energy resolution is $\sigma_E/E = \frac{10\,\%}{\sqrt{E}} \oplus 1\,\%$ for E in GeV where the varying term comes from statistical fluctuations while the constant term depends on systematic effects e.g. the stability of the calibration or loss of energy in the non-active regions of the calorimeter.

The HCAL has a thickness of 5.6 interaction lengths due to spacial limitations and consists of a shashlik structure alternating between iron and scintillating tiles. Its energy resolution is worse than that of the electromagnetic calorimeter due to the higher fluctuation of energy deposition of hadronic showers and the low thickness: $\sigma_E/E = \frac{(69\pm5)\,\%}{\sqrt{E}} \oplus (9 \pm 2)\,\%$ for E in GeV where the varying term again comes from fluctuations and the constant term from calibration. For angular resolution the calorimeters are divided into cells with finer granularity near the beampipe, where the particle flux is higher.

Reconstruction of π^0 The decay inspected in this analysis contains a neutral pion in the decay chain, which needs to be reconstructed. π^0 are very abundant at LHCb, more than 20 are produced in every *pp*-collision. They decay after about 10^{-16} s with a branching ratio of $\mathcal{B} = 98.8\,\%$ into two photons. The photons then produce electromagnetic showers in the ECAL, which are measured as energy deposits in the calorimeter cells. These cells are classified using a cluster algorithm to determine the energy and direction of the photons from the size and energy deposition inside the clusters (see Ref. [19]).

The photons coming from π^0 with less than 2 GeV transversal momentum are mostly reconstructed as two distinct clusters in the ECAL. These so-called resolved π^0 are reconstructed by first creating photon candidates of all clusters and then loop over the candidates. If the reconstructed invariant mass of a pair of photon candidates lies inside the π^0 mass region ($m(\gamma\gamma) \in [105, 165]$ MeV), they are combined into a π^0 candidate. To reduce combinatoric background from random photon combinations coming directly from the PV or two different π^0's, a p_T cut of 200 MeV is applied on every photon candidate. This decreases the reconstruction efficiency for π^0 with p_T below 1 GeV (see blue-lined histogram in Fig 8 left), because in this region it becomes likely that one of the photon p_T

is below the cut.

For π^0 with momenta $> 2\,\text{GeV}$ the two photons have a bigger boost, which can result in their clusters' overlapping (merged π^0). This leads to a decrease in resolved π^0 reconstruction efficiency at higher p_T. Another algorithm is used to disentangle the energy contribution from the two photons on the cluster cells that are shared between clusters. The two clusters are then used as photon candidates and it is tested if their combined invariant mass lies within the π^0 mass window.

The resulting reconstruction efficiency for these merged π^0 is shown as red-dotted histogram in Fig. 8 on the left.

Most of the π^0 in the signal decay $B^+ \to J/\psi\,\rho^+$ have p_T below 2 GeV (Fig 8 on the right), so only a small fraction of the simulated π^0 is reconstructed as merged π^0. Additionally the mass resolution for the merged reconstruction is far worse than for resolved π^0. Hence only resolved π^0 are used in this analysis.

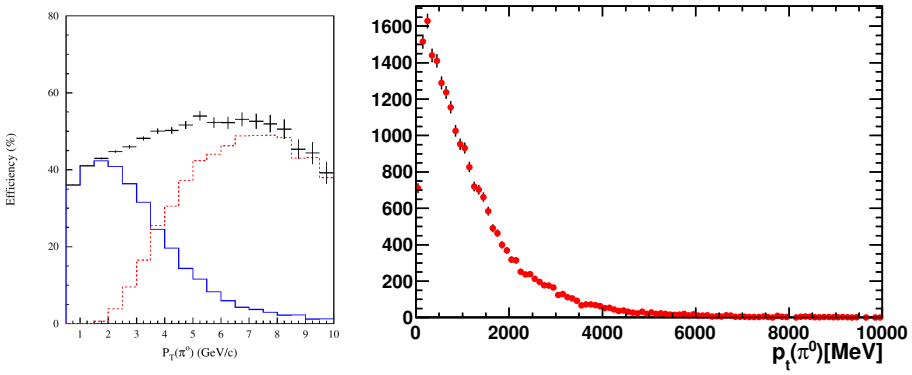

Figure 8: Left: π^0 reconstruction efficiency over transversal momentum. Blue for resolved π^0, red for merged π^0 (for explanation see Sec 3.2.3), black for total efficiency. Taken from [19]. Right: Generator level MC of π^0 transversal momentum distribution for the decay $B^+ \to J/\psi\,\rho^+$. Most π^0 lie in the momentum region, where they would be reconstructed as resolved π^0.

.

Muon System Muons with momenta above a few hundred MeV are minimum ionising. This means their stopping power dE/dX (energy loss over distance travelled through material) is very low and most of them are not stopped inside the calorimeters. Consequently, particles can be identified as muons by bringing additional material into their way and check whether they traverse it.

The muon system consists of 5 stations which use Multiwire Proportional Chambers (MWPC) to detect ionizing particles. The first station (M1), which sits upstream of the calorimeters, has additionally a GEM detector in the high-

flux region near the beampipe, which is more radiation resistant than MWPCs. M1 is only used for the first trigger stage L0 (see Sec 3.2.4) to detect high-p_{T} muons. The other four stations are located downstream of the calorimeters, each separated from the next one by 80 cm thick iron absorbers to stop other particles that have higher (dE/dX). The muon identification efficiency is about 95 % while the misidentification of other particles as muons lies below 2 %.

3.2.4 Trigger

The data used in this analysis was taken with a bunch crossing rate of 20 MHz, but only 5 kHz of event rate could be written to storage, so only physically interesting events were saved. To detect these events a multi-level trigger system is used in LHCb, which will be presented in the following. The exact trigger configuration used for this analysis can be found in Sec 6.2.

L0 The first trigger stage consists of custom designed hardware working synchronously with the bunch crossings. It reconstructs the highest-E_{T} candidate per particle species in the calorimeters and the two highest-p_{T} muon candidates in the muon chambers, assuming the muons originate from the PV and get a single kink from the magnets. Only if one or more of these values lie above their respective threshold, the full detector gets read out and the event gets processed further. This reduces the rate to below 1 MHz.

High Level Trigger (HLT) The events that were triggered in L0 are fed into the Event Filter Farm (EFF), where they are analysed by the HLT. The HLT itself is split into two stages: at stage 1 only information from the VELO and the T-stations is used to reconstruct first tracks, and estimate their momenta and confirm the absence of charged tracks at the positions of photon candidates in the ECAL. Additional requirements on tracks can be applied, like their impact parameter with respect to the primary vertex. This reduces the event rate to below 30 kHz at which the final stage of the HLT can perform nearly offline-quality reconstruction using information from all subsystems of the detector to decrease the event rate to a recordable value of 5 kHz.

4 Analysis Strategy

The following section provides an overview of the analysis, which is detailed in the main part of this thesis. Goal of this analysis is the extraction of the Branching fraction of the decay

$$B^+ \to (J/\psi \to \mu^+\mu^-)(\rho^+ \to \pi^+\pi^0)$$

relative to the decay

$$B^0 \to (J/\psi \to \mu^+\mu^-)(K^{*0} \to K^+\pi^-).$$

B^+ and B^0 have a mass difference of only 0.3 MeV ($m(B^+) = 5279.25\,\text{MeV}$, $m(B^0) = 5279.58\,\text{MeV}$) and both channels contain a vector meson and a J/ψ ($m(\rho^+) = 770\,\text{MeV}$, $m(K^*) = 896\,\text{MeV}$, $m(J/\psi) = 3097\,\text{MeV}$). The latter decays into two muons with very similar kinematics. Systematic uncertainties, e.g. on the muon efficiency, cancel therefore largely in a relative measurement between the two channels. The branching fraction of $B^0 \to J/\psi K^{*0}$ is known with a relative uncertainty of only 4.5 % and a clean signal (see Sec. 5.2) can be extracted from data due to its high branching ratio ($\mathcal{B}(B^0 \to J/\psi K^{*0}) = 1.32 \times 10^{-3}$, $\mathcal{B}(K^* \to K^+\pi^-) = 2/3$)). This is about 17 times the branching ratio of the signal channel ($\mathcal{B}(B^+ \to J/\psi \rho^+) = (5.0 \pm 0.8) \times 10^{-5}$, $\mathcal{B}(\pi^0 \to \gamma\gamma = 99.8\,\%)$), which was measured by the `BaBar` collaboration (see Ref. [1]).

The data sample analysed in this thesis was collected with the LHCb detector in the years 2011 and 2012 (Run 1), and corresponds to an integrated luminosity of $3\,\text{fb}^{-1}$. Collisions took place at a center-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$ in 2011 and $\sqrt{s} = 8\,\text{TeV}$ in 2012. Candidate events were recorded, if they fired at least one of the trigger lines that were implemented for decays of B-mesons into final states with muons (see Sec. 6.2). These events were reconstructed using the Brunel software (see Ref. [20]) and analysed with the DaVinci software (see Ref. [21]). The particles for MC simulations were generated by `Pythia8` (see Ref. [22]), their detector response was simulated with Geant4 (see Ref. [23], and [24]). The simulated events were reconstructed with the same software and the same selection as the data samples.

The Analysis was performed in the following steps:

- MC simulation of the signal decay was examined to check if it described the data well and corrected if needed (see Sec. 5).

- A so-called stripping was performed on the recorded candidates, in which the decay chain was reconstructed and a loose selection was applied. At this stage many trigger lines were allowed to tag an event to be written to storage. In the trigger selection the final trigger configuration for this analysis was determined (see Sec. 6.2).

- In the preselection cuts on kinematic variables against combinatoric background and cuts on PID-variables against peaking background from incorrectly identified particles were applied (see Sec. 6.3).

- To further reduce the combinatoric background, a boosted decision tree (BDT) was trained with sWeighted (see Ref. [25] $B^0 \to J/\psi K^{*0}$ data as signal sample and data from the upper sideband of the reconstructed B^+-mass as background sample (see Sec. 6.4.1).

- Against partially reconstructed background a second BDT was trained with simulated $B^+ \to J/\psi \rho^+$-events as signal and a sample of simulated inclusive $B_{u/d/s} \to J/\psi X$ decays as background (see Sec. 6.4.2).

- The cuts on the first and second BDT were optimized using simulated $B^+ \to J/\psi \rho^+$-events and signal channel upper sideband data (see Sec. 6.4.2).

- After selection $B^+ \to J/\psi \rho^+$ and $B^0 \to J/\psi K^*$ were separately fitted in the B-candidate invariant mass to obtain their respective yields. The signal branching fraction was then determined from the relative yields:

$$\mathcal{B}_{sig} = \frac{\mathcal{B}_{\text{norm}}}{\mathcal{B}_{\pi^0 \to \gamma\gamma}} \times \frac{N_{sig}}{N_{norm}} \times \frac{\xi_{norm}}{\xi_{sig}} \times \frac{\epsilon_{norm}^{\text{reco, presel}}}{\epsilon_{sig}^{\text{reco, presel}} \times \epsilon^{bdtc} \times \epsilon^{bdtp}}$$

  For the branching ratio of the normalisation channel $\mathcal{B}_{norm}$ the PDG value (see Ref. [26]) was taken, and for the subsequent decay $\mathcal{B}(K^{*0} \to K^+\pi^- = 2/3$ isospin symmetry was assumed[2]. $\mathcal{B}_{\pi^0 \to \gamma\gamma} = 0.98823 \pm 0.00034$ is the fraction of π^0 decaying into two photons, N_i are the fit yields of the signal decay $B^+ \to J/\psi \rho^+$ respectively the normalisation decay $B^0 \to J/\psi K^{*0}$. $\epsilon_i^{\text{reco, presel}}$ are the efficiencies of the generation, reconstruction and preselection taken form MC simulated events. ξ_i are the geometric acceptances, which are determined using MC simulation: The number of all MC simulated events that have all the particles from the decay chain emitted in a region of the detector, where they could be reconstructed, divided by the total number of generated MC events containing this decay chain. ϵ^{bdtc} is the efficiency of the cut on the output of the BDT against combinatorics, and ϵ^{bdtp} is the efficiency of the cut on the output of the BDT against peaking background, both were taken from Monte Carlo simulated signal events (see Sec. 6.5).

[2]The branching fraction of $K^{*0} \to K\pi$ is $\mathcal{B} = 100\%$, which is split into two possible modes: $K^{*0} \to K^-\pi^+$ and $K^{*0} \to K^0\pi^0$. K^{*0} is an isospin state (1/2, 1/2), K^- is a (1/2, −1/2) state, and π^+ is a (1, 1) state. The projection of the isospin state of K^{*0} on the isospin state of the daughter particles therefore gives a factor of 2/3. Assuming isospin symmetry this is the fraction of K^{*0} decaying into K^- and π^+

4.1 Fitting Procedure

All fits were performed using unbinned maximum likelihood estimators (see e.g. Ref. [27]): The likelihood function $\mathcal{L}$ gives the probability of measuring a given dataset with specific parameters:

$$\mathcal{L} = \prod_{i=1}^{N} f(\vec{x}_i|\vec{a}) \tag{12}$$

Here N is the number of events, each event is described by one vector of measured variables $\vec{x}_i$. It has a probability given by the probability distribution function $f(\vec{x}_i|\vec{a})$ for a certain choice of parameter values $\vec{a}$. The choice of $\vec{a}$ for which $\mathcal{L}$ is maximized (the one under which the measured data are most plausible) is then chosen as the fit result. For computational reasons $-\log(\mathcal{L}) = -\sum_{i=1}^{N} \log(f(\vec{x}_i|\vec{a}))$ is minimized instead.

For all fits the RooFit toolkit was employed (see Ref. [28]) using `Minuit`. It first applied the `MIGRAD` method, which minimizes $\mathcal{L}$ with a multidimensional quasi-Newton method using only an approximation of the Hessian matrix to reduce computing cost (see Ref. [29]). After `MIGRAD` had converged, the `HESSE` method was invoked using the exact Hessian matrix to calculate a more accurate error matrix for the parameter estimates.

4.2 Variables

This section explains the variables used during the signal selection:

- **FD** stands for flight distance and is the distance between the reconstructed decay vertex of the B-candidate and the primary vertex that was closest to its reconstructed flight direction.

- **DLS** (Decay Length Significance) measures how well separated a particle's decay vertex is from its vertex of origin: $DLS = \frac{FD^2}{\chi^2(FD)}$

- *Vertex* $\boldsymbol{\chi^2}$ is a measure of the quality of the vertex reconstruction. For a given reconstructed vertex $\boldsymbol{\Delta\chi^2_{add-track}}$ gives the minimum χ^2-difference, when an additional track is added to the reconstruction of the vertex under consideration. If other charged particles than those reconstructed were near the decay vertex, this value is small. It is then likely that the decay included the other particle as well. Hence this variable is supposed to have lower values for partially reconstructed background than for signal.

- **IP** stands for the so-called impact parameter. It is the minimal distance between the primary vertex (PV) and the extrapolation of a particle's track. If this value is small, the particle was likely produced in the PV. In the triggers and preselection a cut on IP is performed on all B-candidates.

Their lifetime is $c\tau \approx 0.5\,\text{mm}$ and they are boosted, so they travel a few cm in the detector system. $\chi^2(IPD) = (IP/\sigma(IP))^2$ gives the significance of the value of IP.

- $\boldsymbol{\chi^2}$**(IP)** is the value, by which the χ^2 of the PV-fit changes, when the track is added. If the track really comes from a B-meson, that flew a distance before it decayed, then this value should be higher than for particles that were created directly in the pp-collision.

- $\mathbf{A^{cc}_{p_T}}$ measures the p_T asymmetry inside a cone originating from the decay vertex of the B-candidate. The tip of the cone points to the PV. The transversal momentum p_T with respect to the z-axis of all tracks that lie in this cone and are not associated to the B-candidate are summed up and compared to the p_T of the B-candidate.

$$A^{cc}_{p_\text{T}} = \frac{p_{\text{T}\,B\,cand} - \sum p_{\text{T}\,otherTracks}}{p_{\text{T}\,B\,cand} + \sum p_{\text{T}\,otherTracks}}$$

  If there are only reconstructed tracks from the B-candidate's decay inside of the cone, then $A^{cc}_{p_\text{T}}$ becomes unity. For a background event, in which an additional charged particle was produced in the decay, this particle's track is likely to be inside the cone as well and $A^{cc}_{p_\text{T}}$ is smaller.

- The Decay Tree Fitter (**DTF**) is a tool that performs a fit of the whole decay chain with the option to constrain particle masses to their PDG-values and the flight direction from primary particles to the PV (see Ref. [30]). In this analysis the flight direction of the B-candidate was constrained to point from the PV to its reconstructed decay vertex. Additionally the masses of π^0 and J/ψ-candidates were set to their PDG-masses in the DTF. In contrast to a so-called bottom-up-approach the DTF fits all 4-momenta and vertex positions of the decay chain simultaneously, using all available information about correlations. 4-momentum conservations at each vertex are used as internal constraints. The measured 4-momenta of the final state particles, the PV-position, and enforced resonance masses of π^0 and J/ψ are used as external constraints (see Ref. [31]). $\boldsymbol{\chi^2}$**(DTF)** is the χ^2 of this fit. DTF-fits of combinatoric background events, which do not contain common decay vertices, usually result in higher χ^2 values than for signal events.

- **DIRA** stands for DIRection Angle and represents the cosine of the angle between the reconstructed momentum vector of the B-candidate and the vector between the PV and its decay vertex. If the reconstruction misses a daughter particle or the B-candidate is put together from random particles the angle is usually large, whereas for signal both vectors align and the angle is small, resulting in $DIRA \approx 1$.

- **CL**(π^0) is the confidence level for the π^0-candidate. It is the output of a neural net that uses information from the ECAL and SPD about the cluster shape, cluster size, and energy deposit of the photons to distinguish between real π^0 and random combinations of 2 photons whose combined energy lies in the right mass window.

- **DOCA** (Distance of Closest approach) is the minimum distance that two given tracks have from each other. $\chi^2(DOCA) = DOCA/\sigma^2(DOCA)$ gives the significance of that value.

- $\mathrm{DLL}_{K\pi}$ (Delta Log Likelihood) is the log likelihood difference $\Delta log\,\mathcal{L}(K - \pi) = log(\mathcal{L}(K)/\,\mathcal{L}(\pi))$. For each traversing particle the information from the PID-system can be used to calculate likelihoods for different mass hypotheses (electron, muon, pion, kaon, proton). $\mathrm{DLL}_{K\pi}$ describes how likely the hypothesis 'kaon' is with respect to the default hypothesis 'π' (for information about the PID-system see Sec 3.2.3).

- A reconstructed track that does not come from a particle (or is combined from hits of more than one particle) is called a ghost. **ghostprob** is the output of a neural net that calculates from various tracking information (e.g. detector occupancy), how probable it is that the reconstructed track is actually a ghost.

- **IsMuon** is a Boolean value. It requires that the respective muon has hits in at least two (three) of the Muon stations, if it has low (high) momentum (see Ref. [32]).

5 Signal Monte Carlo Calibration

Due to imperfect simulation, the distributions of the variables introduced in Sec. 4.2 can differ between real events and MC generated ones. For example, the number of tracks *nTracks* in an event is underestimated in simulation. This analysis uses distributions of MC events to train classifiers against background events. If the classifier is trained on "wrong" distributions, it does not perform well on real data. Additionally the efficiencies of the cuts on the BDT outputs were determined from MC simulated samples, therefore the accuracy of the simulation of the signal MC directly affected the measured branching ratio. Therefore only those variables were used in the classifier that were confirmed to be similar in MC and data.

For the kinematic variables this was done by comparing the distributions in $B^+ \to J/\psi\,\rho^+$ MC simulation with the distributions in the normalization channel $B^0 \to (J/\psi \to \mu^+\mu^-)(K^{*0} \to K^+\pi^-)$ (see Sec. 5.2). This was done in two steps: First it was verified that the distributions in $B^+ \to J/\psi\,\rho^+$ MC simulation agreed with the corresponding distributions in a MC simulation of the normalisation channel. In a second step, the distributions in $B^+ \to J/\psi\,\rho^+$ MC simulation were compared to a clean data-sample of the normalisation channel $B^0 \to (J/\psi \to \mu^+\mu^-)(K^{*0} \to K^+\pi^-)$.

For variables that are influenced by the number of charged tracks in the decay, or are directly deduced from the reconstructed π^0, the charged B-decay $B^+ \to J/\psi\,K^{*+}$ was used. It also has a π^0 in the final state (see Sec. 5.4) and therefore has the same number of reconstructed tracks from charged particles as the signal channel.

To obtain correct particle identification variables in MC a PID resampling is performed (see Sec. 5.1).

5.1 PID Resampling

PID variables are known to be simulated inaccurately in MC, so they were not used for cuts. New PID variables were created instead with a data-driven method using a tool called PID-Calib. It comprises of a number of clean data samples, coming from easy-to-select channels like $K^0_S \to \pi^+\pi^-$, where no cuts on PID-variables were necessary for selection. The samples therefore exhibit the real distributions of PID-variables for a given particle species.

The distributions of PID variables are correlated to kinematic quantities of the particle and the number of tracks in the event. These correlations need to be taken into account, because the kinematic quantities of the particle differ between the signal channel $B^+ \to J/\psi\,\rho^+$ and the samples in PID-Calib. For given MC simulated events, each containing for example a truthmatched π^+, one can therefore obtain correctly distributed values of $\mathrm{DLL}_{K\pi}$: The PID-Calib datasample for pions gets binned in η, *nTracks*, and p_T. Because of the

correlations, each bin then contains a different distribution of $\mathrm{DLL}_{K\pi}$. A π^+ from MC-simulation is assigned to one of these bins according to its values for η, $nTracks$, and p_T and a value of $\mathrm{DLL}_{K\pi}$ is sampled from the $\mathrm{DLL}_{K\pi}$ distribution inside this bin. This is done for every simulated pion separately, which results in a $\mathrm{DLL}_{K\pi}$ distribution similar to data including correlations with η, $nTracks$, and p_T of the π^+. $\mathrm{DLL}_{K\pi}$ can then be used instead of the originally simulated one.

This of course requires a correct description of η, $nTracks$, and p_T in the MC sample. The distribution of MC events in η is simulated well, while $nTracks$ and $p_\mathrm{T}(B^+)$ need to be reweighted (see Sec. 5.3) to align their distributions with data.

5.2 $B^0 \to J/\psi\, K^{*0}$ Normalisation Signal Extraction

The channel $B^0 \to J/\psi\, K^{*0}$ was chosen as normalization and control channel due to its high branching fraction of $\mathcal{B} = (1.32 \pm 0.06) \times 10^{-3}$ and its kinematics being similar to the signal channel: In $B^+ \to J/\psi\, \rho^+$, ρ^+ and J/ψ are emitted in the B^+ rest frame with a momentum of 1611 MeV while in $B^0 \to J/\psi\, K^{*0}$, $K^*(892)$ and J/ψ get a momentum of 1571 MeV. These values are averages due to the width of the daughter particles ($\Gamma(\rho^+) = 150\,\mathrm{MeV}$, $\Gamma(K^{*+}) = 49\,\mathrm{MeV}$) and together with the finite detector resolution, the distributions in p_T of K^{*+} and ρ^+ are similar (see Fig. 9 on the left).

Normalisation channel candidates ($B^0 \to J/\psi\, K^{*0}$) were selected using the same stripping and trigger decisions as for signal ($B^+ \to J/\psi\, \rho^+$, see Sec. 6). Additional cuts had to be adjusted to the difference in final states:

- $DLL_{K\pi}(K^+) > 0$
- $DLL_{K\pi}(\pi^-) < 0$
- $p_\mathrm{T}(K^+) > 800\,\mathrm{MeV}$ (same as $p_\mathrm{T}(\pi^0)$-cut)
- $p_\mathrm{T}(K^{*0}) > 800\,\mathrm{MeV}$ (same as $p_\mathrm{T}(\rho^+)$-cut)

The events remaining after this selection were then fitted. For the 2012 magnet polarity down sample this resulted in 125 790 events from $B^0 \to J/\psi\, K^{*0}$ and 1070 $B_s^0 \to J/\psi\, K^{*0}$ events ($\mathcal{B} = 4.4 \times 10^{-5}$). This is the expected ratio for the two channels considering that about 4 times as many B^0 as B_s^0 are produced (see Fig. 9 on the right).

Besides extraction of the number of $B^0 \to J/\psi\, K^{*0}$ events, the fit was used to create a clean $B^0 \to J/\psi\, K^{*0}$ sample. This was done using the ${}_s\mathcal{P}lot$ technique (see Ref [25]) with the mass of the B-candidates as discriminating variable. A short description of the ${}_s\mathcal{P}lot$-technique is given in Sec. A.1.

The clean control channel data sample could then be used for 3 purposes:

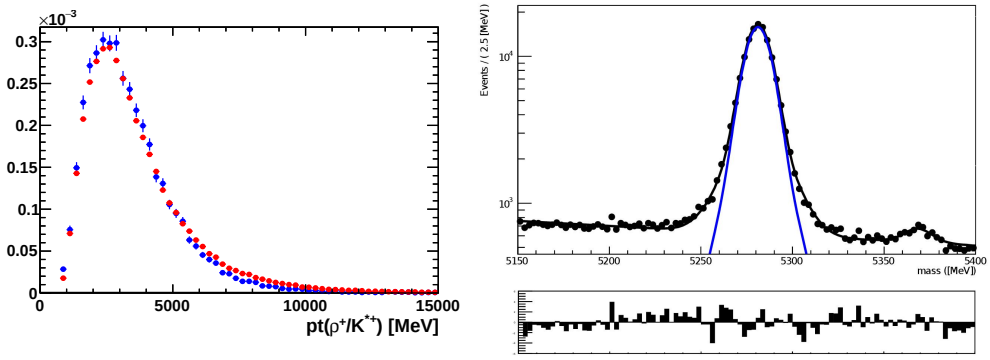

Figure 9: Left: p_{T} distribution of ρ^+ from simulated $B^+ \to J/\psi\,\rho^+$ events in blue and p_{T} distribution of K^* from simulated $B^0 \to J/\psi\,K^{*0}$ events in red (normalized). They are very similar. Right: Fit to mass of selected B^0-candidates from $B^0 \to J/\psi\,K^{*0}$ performed on 2012 Magnet Down data sample. Modelled with a double Gaussian in blue, combined signal and background pdf in black. The shape parameters for the signal were left to float. The smaller peak from $B_s^0 \to J/\psi\,K^{*0}$ around 5370 MeV was fitted using the same shape as the normalization channel signal shifted by $m(B_s^0) - m(B^0)$. The combinatoric background was fitted with an exponential distribution. The y-axis is logarithmic.

- reweighting the simulated kinematic distributions of the signal channel $B^+ \to J/\psi\,\rho^+$ (see Sec. 5.3)
- as signal proxy in the training of the BDT to suppress combinatoric background (see Sec.6.4.1)
- as normalisation channel, with respect to which the branching ratio of $B^+ \to J/\psi\,\rho^+$ was determined (see Sec. 8)

The $B^0 \to J/\psi\,K^{*0}$ distribution was fitted with two Gaussian distributions with shared mean. A Gaussian distribution is given by

$$G(m|\mu,\sigma) = A \exp\left(-\frac{(m-\mu)^2}{2\sigma}\right) \tag{13}$$

with A a normalisation constant, μ the mean of the distribution and σ its width. The model for $B^0 \to J/\psi\,K^{*0}$ was therefore given by:

$$\mathcal{P}_{\mathrm{sig}} = f_1 G(m|\mu,\sigma_1) + (1-f_1) G(m|\mu,\sigma_2) \tag{14}$$

with a floating parameter f_i denoting the relative fractions of the components such that $\mathcal{P}$ is normalized.

The decay $B_s^0 \to J/\psi\,K^{*0}$ proceeds kinematically very similar to $B^0 \to J/\psi\,K^{*0}$, therefore its mass distribution was fitted with a sum of two gaussians

as well, using the same shapes as for $B^0 \to J/\psi K^{*0}$ only shifted by the mass difference $\Delta m \equiv m(B_s^0) - m(B^0) = 87.35\,\text{MeV}$:

$$\mathcal{P}_{B_s^0} = f_1 G(m|\mu + \Delta m, \sigma_1) + (1 - f_1) G(m|\mu + \Delta m, \sigma_2) \tag{15}$$

The combinatoric background was fitted with an exponential distribution with a slope parameter C that was left to vary:

$$\mathcal{P}_{\text{cbg}} \equiv BG_{comb}(m|C) \propto \exp(Cm)\,. \tag{16}$$

For the combined PDF this resulted in

$$\mathcal{P} = f_{\text{sig}}\mathcal{P}_{\text{sig}} + f_{B_s^0}\mathcal{P}_{B_s^0} + (1 - f_{\text{sig}} - f_{B_s^0})\mathcal{P}_{\text{cbg}}. \tag{17}$$

With the result of the fit to the data (see Fig 9 on the right), each event was weighted with the ${}_s\mathcal{P}lot$ technique such that the true $B^0 \to J/\psi K^{*0}$ kinematic distributions were unfolded in the dataset. This was done using the reconstructed mass as discriminating variable.

5.3 $B^+ \to J/\psi \rho^+$ Monte Carlo Reweighting

The normalisation channel was used to check if the distributions of selection-variables agreed between MC simulation and data, or if they needed to be corrected. This was particularly necessary in this analysis, because MC generated events were used to determine the selection efficiency. A difference between data and simulation in the distribution of selection-variables affects the branching ratio measurement.

Normalisation channel and signal channel are kinematically very similar. It is therefore expected that many variables had similar distributions in both channels. On the other hand it is known that in MC especially $p_{\text{T}}(B)$ and the number of reconstructed long tracks in an event (*nTracks*) are simulated poorly. It was checked that distributions for these variables are simulated similarly for signal and normalization channel (see Fig. 10). With the normalisation channel data it was then possible to confirm the discrepancy between simulated events and real data by comparing the distributions between signal $B^+ \to J/\psi \rho^+$ MC and sWeighted $B^0 \to J/\psi K^{*0}$ data (see Fig. 11 before and after reweighting). Two one-dimensional weightings (one in $p_{\text{T}}(B)$ and one in $nTracks$) were performed to match MC to data as good as possible: *nTracks* and $p_{\text{T}}(B)$ are not correlated, therefore each signal MC event was reweighted in bins of $p_{\text{T}}(B)$ and *nTracks* separately and the two weights could simply be multiplied. The combined weights of the signal MC were all positive and additionally had no outliers with very big weights (see Fig. 12 on the left).

Distributions like $A^{cc}_{p_{\text{T}}}$ that are correlated to both $p_{\text{T}}(B)$ and *nTracks* were assimilated to the distribution in $B^0 \to J/\psi K^{*0}$ data as well by the reweighting and their use in the BDT is thereby justified (see Fig. 11).

Due to the higher beam energy, the number of particles created in collisions was higher in 2012 than in 2011, so the reconstructed number of tracks tended to be higher as well (see Fig. 12 right). Thus the reweighting was performed separately for samples simulating 2011 conditions and samples simulating 2012 conditions.

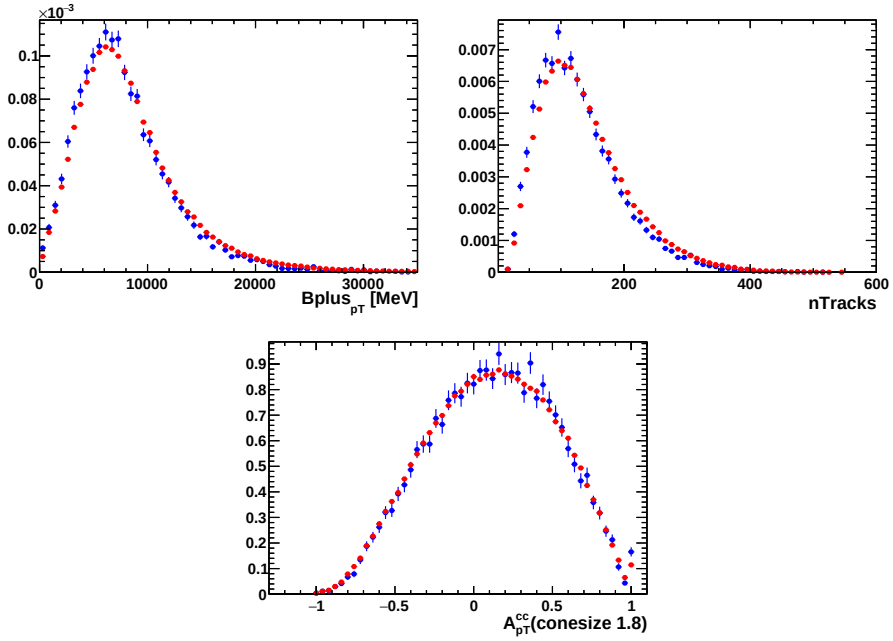

Figure 10: Normalized distributions before reweighting for simulated $B^+ \to J/\psi \rho^+$ events in blue and for simulated $B^0 \to J/\psi K^{*0}$ events in red. Left: $p_{\mathrm{T}}(B)$. Right: *nTracks*. Bottom: $A^{cc}_{p_{\mathrm{T}}}$. The distributions are similar to each other.

5.4 $B^+ \to J/\psi K^{*+}$ Control Signal Extraction

The normalisation channel contains a K^+ instead of a π^0 in the decay chain and therefore has an additional track in the final state. Hence it could not be used to check π^0-related quantities like $CL(\pi^0)$ or quantities that differ with the number of tracks in the decay like χ^2_{DTF} or $DIRA$ (for a description of these variables see Sec. 4.2). In these cases the channel $B^+ \to J/\psi K^{*+}$ with $K^{*+} \to \pi^0 K^+$ was consulted.

A short study had been performed to check, if this channel was feasible as a normalisation channel. It turned out negative due to differences in the kinematic distributions of K^{*+} and the quality of the B-decay-vertex $Vertex\chi^2$ with respect to ρ^+ and $Vertex\chi^2$ from $B^+ \to J/\psi \rho^+$. Additionally the sample is not as clean as and the event yield is smaller than that of $B^0 \to J/\psi K^{*0}$.

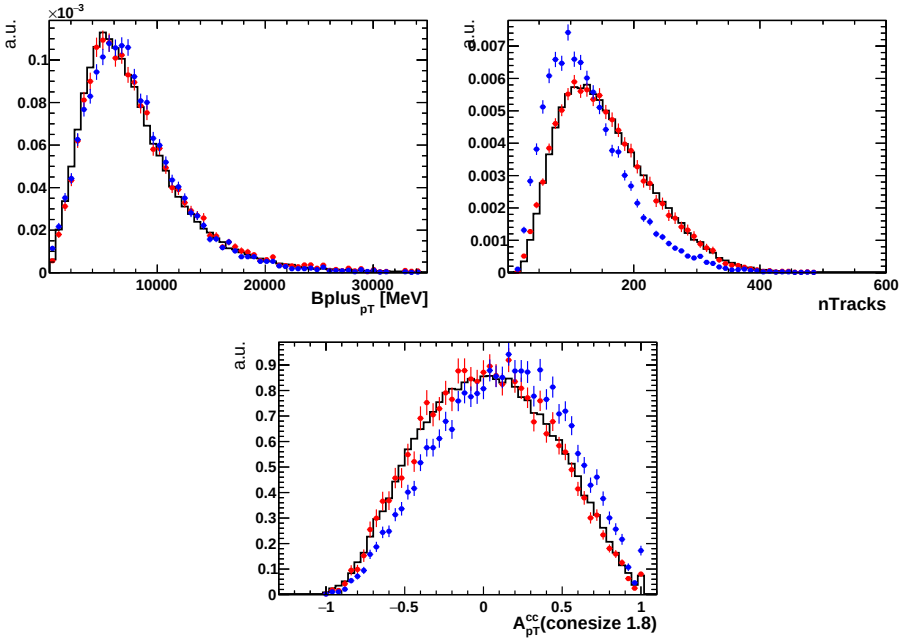

Figure 11: Normalized distributions for $B^+ \to J/\psi\,\rho^+$ simulation before reweighting as blue dots, after reweighting in *nTracks* and $p_{\mathrm{T}}(B)$ as red dots, $B^0 \to J/\psi\,K^{*0}$ *sWeighted* data as black line. Left: $p_{\mathrm{T}}(B)$. Right: *nTracks*. Bottom: $A^{cc}_{p_{\mathrm{T}}}$. Before the reweighting the distributions differ significantly, after the reweighting they are more similar.

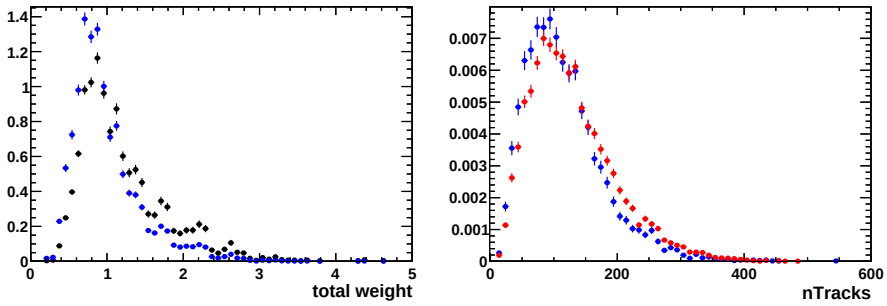

Figure 12: Left: Distribution of product of weights for $B^+ \to J/\psi\,\rho^+$ simulation from weighting in $p_{\mathrm{T}}(B)$ and *nTracks* (normalized). Right: Distribution of *nTracks*, 2011 (2012) Magnet Down signal simulation in blue (red).

The stripping line `FullDSTDiMuonJpsi2MuMuDetachedLine` (see Sec. 6.2) was not available for the $B^+ \to J/\psi\,K^{*+}$ samples but `StrippingB2XMuMu` instead. It was therefore compared to a correspondent `StrippingB2XMuMu` $B^+ \to J/\psi\,\rho^+$-sample. There is no cut to explicitly select J/ψ by the invariant mass of

the two muons in this line. To reduce the rate from background, the other cuts are harder, as can be seen in table 1 (in comparison with the cuts from `FullDSTDiMuonJpsi2MuMuDetachedLine` in Table 5). Its signal was then ex-

Candidate	Selection
B^+	$\chi^2/ndf(vertex) < 8.0$
	$\chi^2(IP) > 16.0$
	$DIRA > 0.9999$
	$\chi^2(FD) > 121.0$
J/ψ	$\chi^2/ndf(vertex) < 12.0$
	$m < 5050.0$
$\mu^\pm$	$\chi^2(IP) > 9.0$
	$\chi^2(track) < 4.0$
	$DLL_{\mu\pi} > -3.0$
	$P(ghost) < 0.5$
K^{*+}	$\Delta m < 300.0$
K^+	$\chi^2(IP) > 6.0$
	$HASRICH = TRUE$
	$\chi^2/ndf(vertex) < 9.0$
	$\chi^2(track) < 4.0$
	$P(ghost) < 0.5$
π^0	$p_\mathrm{T} > 700.0$
	$\Delta m < 30\,\mathrm{MeV}$

Table 1: Stripping cuts in the `B2XMuMu` stripping line for Stripping 21.

tracted using a selection of cuts taken from Ref. [33] listed in table 2 and a fit to the B^+-candidate mass was performed. Its results were used to calculate sWeights, which statistically identify the signal contribution to the sample.

The fit model consists of an exponential distribution for the background and two crystal ball (CB) distributions with common mean for the description of $B^+ \to J/\psi K^{*+}$-events.

The CB shape is an empirical distribution developed by the Crystal Ball collaboration to describe the asymmetric shape of mass distributions with energy loss involved (see Ref. [34]). It consists of a Gaussian core and a power law tail on one side beyond a threshold. Its shape is given by:

$$f(m|\alpha, n, \mu, \sigma) = N \cdot \begin{cases} \exp(-\frac{(m-\mu)^2}{2\sigma^2}), & \text{for } \frac{m-\mu}{\sigma} > -\alpha \\ A \cdot (B - \frac{m-\mu}{\sigma})^{-n}, & \text{for } \frac{m-\mu}{\sigma} \leqslant -\alpha \end{cases} \tag{18}$$

with

Candidate	Selection
B^+	$p_{\mathrm{T}} > 2000$
	$\chi^2_{vertex} < 12$
	$\eta < 4.9$
	$DIRA > 0.99996$
	$m > 4900$
	$\chi^2(FD) > 121$
	$\chi^2(IP) < 16$
	$A_{cc}^{p_{\mathrm{T}}}(1.0) > -0.5$
J/ψ	$m \in [2780, 3250]$
μ	$\chi^2(IP) > 9$
K^{*+}	$m \in [792, 1050]$
	$\chi^2(FD) > 9$
K^+	$\chi^2/ndf(track) < 2$
	$p_{\mathrm{T}} > 300$
	$DLL_{K\pi} > 0$
π^0	$p_{\mathrm{T}} > 800$
	$CL > 0.15$

Table 2: Selection cuts for $B^+ \to J/\psi\, K^{*+}$ as in Ref. [33].

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

$$B = \frac{n}{|\alpha|} - |\alpha|$$

$$N = \frac{1}{\sigma(C+D)}$$

$$C = \frac{n}{|\alpha|} \cdot \frac{1}{n-1} \cdot \exp\left(-\frac{|\alpha|^2}{2}\right)$$

$$D = \sqrt{\frac{\pi}{2}}\left(1 + \mathrm{erf}\left(\frac{|\alpha|}{\sqrt{2}}\right)\right)$$

The parameters are μ and σ, the mean and width of the gaussian core of the distribution, α the distance in multiples of σ from the mean, where the power law tail starts. The shape of the tail is described by n. In this decay the energy loss calling for the use of a CB shape comes from the muons emitting bremsstrahlung, which is not corrected for in the reconstruction. Additionally an overestimation of the energy of the π^0 due to additional photons in the vicinity

of the clusters belonging to the photons from the π^0 is possible, which calls for a high-mass-tail. Therefore the fit model for this decay contains two CB:

$$\mathcal{P} = f_{\mathrm{dCB}}\mathcal{P}_{\mathrm{dCB}} + f_{\mathrm{cbg}}\mathcal{P}_{\mathrm{cbg}} \tag{19}$$

with

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

The fit result can be seen in the in Fig. 13, all parameters had been left free to float.

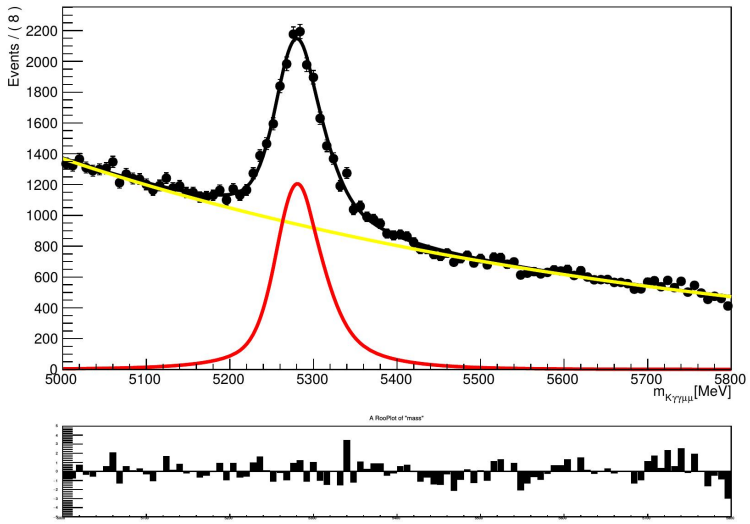

Figure 13: sFit of B^+-candidate mass distribution from 2012 Magnet Down data after selection. $B^+ \to J/\psi K^{*+}$ modelled with a double CB with common mean in red, background modelled with exponential distribution in yellow, combined signal and background distribution in black.

The fit result, i.e. the background and the signal contributions were used to calculate sWeights to statistically identify the signal events. The resulting statistically separated signal event sample could then be employed to test the correct simulation of selection-variables in $B^+ \to J/\psi \rho^+$ MC the same way as with the normalization channel $B^0 \to J/\psi K^{*0}$. The distributions of $DIRA$, $\chi^2/ndf(DTF)$ and $CL(\pi^0)$ agree between $B^+ \to J/\psi \rho^+$-simulation and $B^+ \to J/\psi K^{*+}$-simulation (see Fig. 14). Additionally there is little difference between $B^+ \to J/\psi \rho^+$-simulation and $B^+ \to J/\psi K^{*+}$ sWeighted data (see Fig. 15). This justifies their use in the BDT against partially reconstructed background (see Sec. 6.4.2).

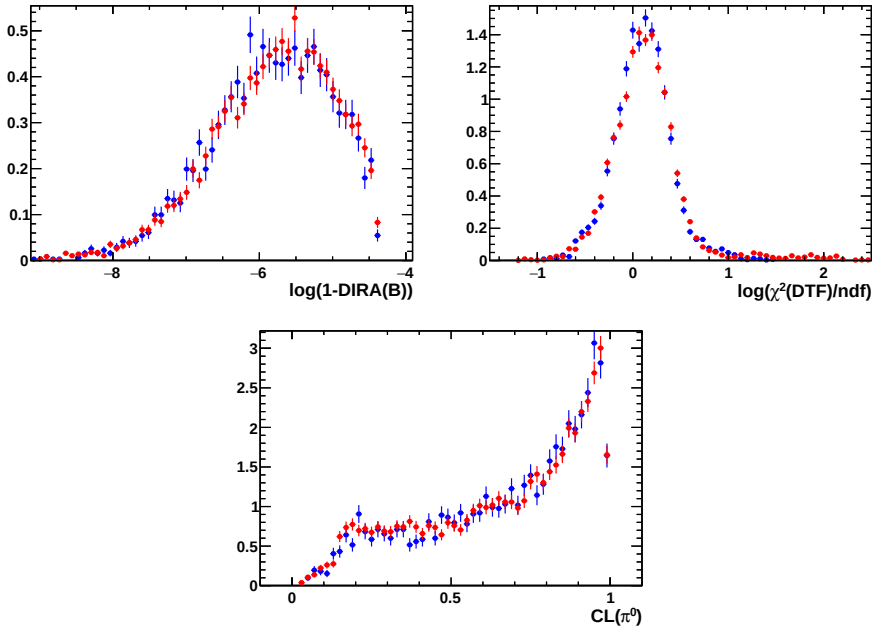

Figure 14: Normalized distributions of selected variables from $B^+ \to J/\psi\,\rho^+$ simulation in blue and $B^+ \to J/\psi\,K^{*+}$ simulation in red after selection given in Table 1 and Table 2. Left: $\log(1 - DIRA)$, Right: $\chi^2/ndf(DTF)$, Bottom: $CL(\pi^0)$

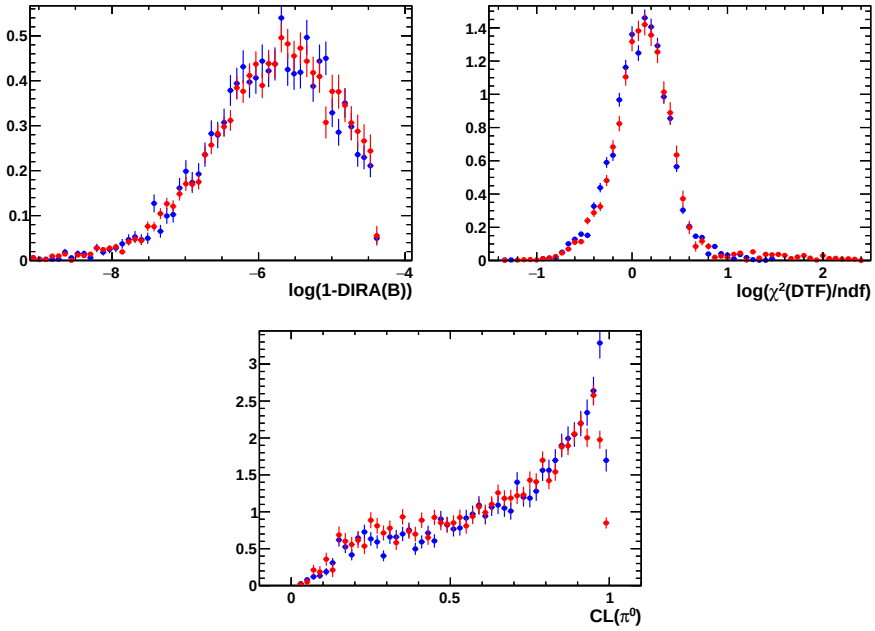

Figure 15: Normalized distributions of selected variables from $B^+ \to J/\psi \rho^+$ simulation in blue and $B^+ \to J/\psi K^{*+}$ $sWeight$ed data in red after selection given in Table 1 and Table 2. Left: $\log(1 - DIRA)$, Right: $\chi^2/ndf(DTF)$, Bottom:$CL(\pi^0)$

6 Event Selection

This section details the signal selection as it was outlined in Sec. 4. First, a characterisation of the different background contributions expected to be found in data is given, then the different selection steps are introduced.

6.1 Background Sources

6.1.1 Combinatoric Background

During event reconstruction it can happen that random particles that do not share an ancestor are selected to form the B^+-candidate. This so-called combinatoric background is featureless and can be fitted in the B^+-mass using an exponential distribution.

6.1.2 Peaking Background

Peaking background events are formed from final state particles that derive from a common ancestor, but are incorrectly reconstructed. This can happen, if one or more of the daughter particles are misidentified as another particle species. These events can have reconstructed B^+-masses showing peaks that lie in the same range as the signal's reconstructed B^+-mass.

For the decay analysed in this thesis, the most prominent peaking background forming channel is $B^+ \to J/\psi K^{*+}$ ($\mathcal{B} = (1.44 \pm 0.08) \times 10^{-3}$, $M = 892\,\text{MeV}$, $\Gamma \approx 50\,\text{MeV}$) with $K^{*+} \to K^+\pi^0$ ($\mathcal{B} = 2/3$), where the K^+ gets misidentified as a π^+. The K^{*+} resonance lies inside the mass range of the ρ^+ resonance ($M = 770\,\text{MeV}$, $\Gamma = 150\,\text{MeV}$) and the branching ratio of this decay is approximately 20 times higher than that of $B^+ \to J/\psi\,\rho^+$. Hence a strict cut on $\text{DLL}_{K\pi}$ of the π^+-candidate is used to reduce its contribution (see Sec. 6.3).

6.1.3 Partially reconstructed Background

Partially reconstructed background is formed from decays of which not all daughter particles were reconstructed. The species of the reconstructed daughter particles can be exactly the same as those from the signal decay, which renders PID variables useless against this background component. Due to the missing mass and momentum of the particles, their shape in the mass spectrum of the reconstructed mother particles is shifted to lower values (which is why the cut on $M(\rho^+)$ in the preselection is asymmetric). Additionally, quantities like the direction angle *DIRA* and the $A_{cc}^{p_\text{T}}$ are less signal-like. If the missing particle was charged, then the variable $\Delta\chi^2_{add-track}$ could be efficient as well, which is why a cut on this variable was applied in the preselection. In the following, the most important partially reconstructed backgrounds for this analysis will be introduced.

- The channel $B^+ \to J/\psi\, K_1(1270)^+$ ($\mathcal{B} = (1.8 \pm 0.5) \times 10^{-3}$) with $K_1(1270)^+ \to \rho^+ \not{K}^0$ ($\mathcal{B} = (21 \pm 3)\,\%$) can form partially reconstructed background, if the K^0 is not reconstructed. Due to the missing Kaon mass, events of this type have a lower reconstructed B^+-candidate mass, where it overlaps with the tails of other partially reconstructed background channels, which only miss a π^0 or a π^+.

- $B^+ \to J/\psi\, K^{*+}$($\mathcal{B} = 1.43 \times 10^{-3}$) with $K^{*+} \to K^0\pi^+$($\mathcal{B} = 2/3$ assuming isospin symmetry) and $K^0_S \to \pi^0 \not{\pi}^0$($\mathcal{B} = 0.307$) has the same final state particles as the signal channel, if the reconstruction is missing a π^0. This makes this channel peak below the signal region in B-candidate mass. The missing π^0 does not produce a track, which renders the preselection cut on $\Delta\chi^2_{add-track}$ useless for separation from the signal channel. Its branching ratio is more than 3 times the signal branching ratio, if one considers that 50 % of K^0 are K^0_S. Its distributions in *DIRA* and $A_{cc}^{p_T}$ are different to signal's.

- $B^0_s \to J/\psi\,\phi$ ($\mathcal{B} = 1.09 \times 10^{-3}$) with $\phi \to \rho^+ \not{\pi}^-$ or $\phi \to \pi^+\pi^0 \not{\pi}^-$ (both resonant and non-resonant combined $\mathcal{B} = 15.32\,\%$) has the same final state particles as the signal, if one π^+ is not reconstructed. This shifts the peak in the B^+-mass spectrum down, but the B^0_s-mass is 87 MeV higher than the B^+-mass, which partly compensates the missing mass from the π^+ and results in a peak that overlaps with the signal region. The expected number of produced events from this channel before any selection is approximately 0.8 times the number of signal events (about four times as many B^+ are produced as B^0_s). The missing π^+ creates an additional track, which results in a smaller $\Delta\chi^2_{add-track}$ than in signal, which can be used together with *DIRA* and $A_{cc}^{p_T}$ to diminish this background's contribution further.

- The decay $B^0 \to J/\psi\,\omega$ ($\mathcal{B} = (2.3 \pm 0.6) \times 10^{-5}$) with $\omega \to \pi^+\pi^0\pi^-$ ($\mathcal{B} = 89.2\,\%$) is a candidate for partially reconstructed background as well. Its low branching ratio, the additional charged track and the low reconstructed ρ^+-candidate mass in the decay diminish its contribution to the lower sideband.

- π^0 don't produce tracks, so if the original π^0 is not reconstructed, another π^0 could take its place. As pions from the PV are very abundant at LHCb, this is very likely to happen. Thus for every decay introduced a version with a random π^0 is possible as well. These events would on average have lower p_T as well as worse *DIRA* and $\chi^2(DTF)$ distributions than signal events. Additionally their mass distributions would not show clear peaks for the B-candidate and the ρ^+-candidate.

- No significant background contribution is expected from fake J/ψ, as the two μ give a clean signal and other possible $c\bar{c}$ resonances are suppressed with a cut on the J/ψ-mass.
- Decays with two or more missing pions are suppressed by constraining the fit window for the B-candidate mass to values above 4950 MeV.
- Other possible partially reconstructed decays result in additional charged tracks, which make it possible to reduce their contributions with $\Delta\chi^2_{add-track}$ and/or they have a branching ratio that is too low to make them contribute significantly. This was tested using an inclusive $B_{u/d/s} \to J/\psi X$ MC simulation and an inclusive $X \to J/\psi Y$ MC simulation. Only the channels mentioned before showed significant contribution in these samples after signal selection had been applied. Their shapes were then extracted from exclusive Monte Carlo simulations for the fit to data.
- A possible additional source for partially reconstructed background could be the decay $B^+ \to J/\psi\, \mathrm{a1}(1260)^+$ ($\mathcal{B} < 1.2 \times 10^{-3}$ @ 90 %cl, see Ref. [35]) with $\mathrm{a1}(1260)^+ \to \rho^+ \not{\pi}^0$. As this channel is not yet established, it was not included in the inclusive $B_u \to J/\psi X$ simulation. If its branching ratio is of the same order as the signal branching ratio, a contribution to the lower sideband should be formed from this decay.

6.1.4 Non-resonant $B^+ \to J/\psi \pi^+\pi^0$

The signal decay can proceed without an intermediate ρ^+ resonance as well: $B^+ \to J/\psi \pi^+\pi^0$ ($\mathcal{B} < 7.3 \times 10^{-6}$@ 90 % CL). It shows the same distribution as signal in the B-candidate mass spectrum, but is suppressed. Its contribution to the B^+ peak could be disentangled from signal by a fit to the ρ^+ candidate mass spectrum. Here this background would show a pure phase space distribution while the ρ^+ resonance could be described by a relativistic Breit-Wigner distribution.

6.2 Trigger Selection

Candidate events are required to be triggered on signal (`TOS`), which means that one or more of the signal's signature-particles must have triggered. The triggers used are listed in Table 3. For a description of the trigger mechanism see subsection 3.2.4. The triggers apply cuts on muon momentum, mass of the J/ψ-candidate, the fit quality of its decay vertex ($vertex\,\chi^2$), how well separated the decay vertex of the J/ψ-candidate is from the PV (χ^2_{IP}), and the track quality of the muons ($track\,\chi^2/ndf$). These requirements are only on the J/ψ-side of the decay, which is very similar for the signal decay $B^+ \to J/\psi\,\rho^+$ and the normalisation decay $B^0 \to J/\psi\, K^{*0}$ resulting in reduced systematic uncertainty on the trigger efficiency. The trigger lines are described in detail

in [36] and [37] for the muon triggers and in [38] for the topological trigger `Hlt2TopoMu2BodyBBDTDecision`.

Stage	Triggers
L0	`L0MuonDecision`
HLT1	`Hlt1TrackAllL0Decision`or `Hlt1TrackMuonDecision` or `Hlt1DiMuonHighMassDecision`
HLT2	`Hlt2TopoMu2BodyBBDTDecision` or `Hlt2DiMuonDetachedJpsiDecision`

Table 3: Trigger requirements on candidate events.

6.3 Stripping and Preselection

A so-called `Stripping` is a central selection applied on recorded events to reduce the datasets that are eventually used for analyses to a manageable size. Depending on the decays of interest, different lines are in use; for this analysis the line `DiMuonJpsi2MuMuDetachedLine` (Stripping 21, `Reco14`) was utilized, which has the requirements listed in table 4.

Candidate	Selection
B^+	$DLS > 3$
J/ψ	$2996.916 < M < 3196.916$
	$vertex\chi^2/ndf < 20$
$\mu^\pm$	$DOCA\chi^2 < 30$
	$p_T > 500\,\mathrm{MeV}$
	$track\chi^2 < 5$
	$DLL_{\mu\pi} > 0$
	$isMuon$ True

Table 4: Stripping cuts in `DiMuonJpsi2MuMuDetachedLine` for Stripping 21

Events that passed these criteria needed to fulfil additional preselection cuts listed in table 5. These were inspired by the angular analysis of $B^0 \to J/\psi\, K^{*0}$ (see Ref. [39]). Due to the π^0 in $B^+ \to J/\psi\, \rho^+$, the angular resolution is worse, so the *DIRA*-cut had to be relaxed. Combinatoric background from random combinations of two photons into π^0-candidates was additionally suppressed with the cut on $CL(\pi^0)$. Other combinatoric background was reduced with the cuts on p_T and the requirements on the B decay vertex separation from the PV. To reduce peaking background from $B^+ \to J/\psi\, K^{*+}$ with the K^+ misidentified

as π^+ (see Sec. 6.1) a hard cut on $\mathrm{DLL}_{K\pi}$ was applied. Asymmetric cuts on the mass of the ρ^+-candidate served against partially reconstructed background not forming a ρ^+-resonance. Against decays containing more charged particles than the signal decay, a cut on $\Delta\chi^2_{add-track}$ was applied.

Candidate	Selection
B^+	$4600\,\mathrm{MeV} < M < 7000\,\mathrm{MeV}$
	$IP\chi^2 < 16$
	$PV\chi^2_{dist} > 64$
	$DIRA > 0.9995$
	$\Delta\chi^2_{add-track} > 3$
	$\chi^2/n_{\mathrm{dof}} < 10$
ρ^+	$p_{\mathrm{T}} > 800\,\mathrm{MeV}$
	$570\,\mathrm{MeV} < M < 995 B_{\mathrm{MeV}}$
π^0	$p_{\mathrm{T}} > 800\,\mathrm{MeV}$
	$CL > 0.02$
π^+	$ghost\ prob < 0.5$
	$hasRich$ True
	$IPD\chi^2 > 6$
	$\mathrm{DLL}_{K\pi} < -5$
$\mu^\pm$	$ghost\mathrm{prob} < 0.5$
	$IPD\chi^2 > 9$
	$DLL_{\mu\pi} > 0$

Table 5: Preselection cuts

6.4 Boosted Decision Tree

For further selection of signal candidates a multivariate classifier called `Boosted Decision Tree` (BDT) was applied (see Ref. [40], [41]). The following explanation orients itself at Ref. [42]. A schematic view of the structure of a decision tree can be found in Fig 16.

Decision Trees are a common method in data analysis to create a model that predicts a target value from several input variables. In this thesis the target value is the signal-likeness of an event and the input variables are the selection-variables that are used to separate signal from background. Beginning from the top, an event is sorted at each node with a binary cut on one of the input variables into the left or the right branch. At each node different variables and cut values can be used. Depending on which outermost node (leaf) this event ended up, it gets classified as either signal-like or background-like. With this method the phase space, which is spanned by the input variables, is split

into many distinct hypercubes, whereas independent (orthogonal) cuts on the input variables only select one possible hypercube.

The Decision Tree has first to be trained with two data sets: one that works as proxy for signal, so its input distributions should be similar to those of the signal in the dataset that later is to be classified, and one that has similar input-variable distributions as the expected background in that dataset.

Starting from the top, at each node a cut on one of the input variables is performed on the training sample, such that the portion of the data set that has reached this node is split into a background-like part and a signal-like part. Variable and cut at each node are chosen such that signal and background are separated optimally by minimising a metric. The metric quantizing variable- and cut-quality was chosen to be the so called Gini-impurity (see Ref. [40]) $p(1-p)$, where p is the fraction of signal in the node. It is minimal for a sample containing only signal or only background and maximal for a sample consisting of 50 % signal. The splitting is repeated at each node until a stopping criterion is reached, which could be perfect separation between signal and background: $p(1-p) = 0$. This is always possible, because with enough cuts, there could finally be nodes with only one event from the training sample. Another, more robust criterion is pruning: nodes are not allowed to contain less than a certain fraction of the training sample.

In most cases, pruning is used as stopping criterion to reduce so-called overtraining of the classifier, where the tree becomes sensitive to statistical fluctuations of the distributions of the datasample. This leads to a good performance on the training sample, but tested on another sample it would perform worse than a classifier that is not overtrained. One can check, if a classifier is overtrained by comparing the distribution of the BDT-output for the training sample with that of a similar test sample. If the two distributions differ significantly, this is a strong sign of overtraining.

To make up for the necessary pruning, boosting is used, which extends the decision tree to a whole forest of trees. Events that were misclassified in one tree get a higher weight in the next tree, such that it is more sensitive to those events. Responses of all the trees are then combined into a single weighted response between -1 (background) and 1 (signal).

Additionally, one can use only a fraction (typically $40\,\% - 60\,\%$) of all events in the training dataset in each tree. This so called Bootstrap AGGregatING (Bagging) is used so that each tree sees different statistical fluctuations, which helps to reduce overtraining.

The training and testing of the BDT was performed with the `ROOT` package `TMVA` (see Ref. [42]) using the `AdaBoost`-algorithm (see Ref. [43]).

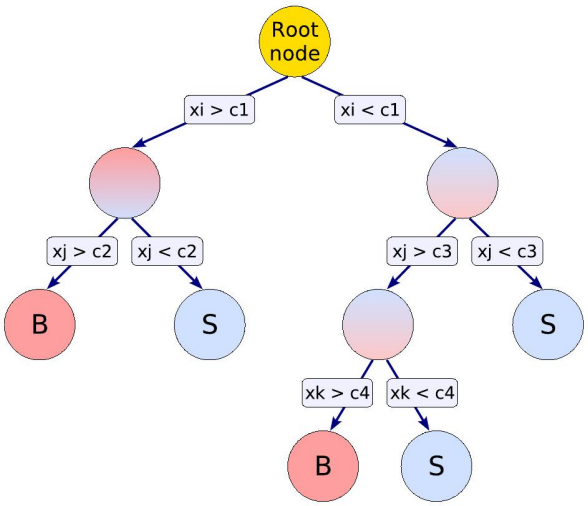

Figure 16: Scheme of a Decision Tree. At each node a one-dimensional cut ci in that input-variable xi is performed, which optimally discriminates signal from background. Leaf nodes are labelled S for signal or B for background depending on which event type comprises the majority of its entries. Taken from [42].

6.4.1 BDT Against Combinatoric Background

A BDT was trained against combinatoric background using a sample of sWeighted $B^0 \to J/\psi K^{*0}$ candidates as signal sample. The distributions of the BDT input variables were verified to be the same in both $B^0 \to J/\psi K^{*0}$ and $B^+ \to J/\psi \rho^+$ (see Sec. 5). The $B^0 \to J/\psi K^{*0}$ sample was used, because it contains approximately a factor of 10 more events than the sample of simulated $B^+ \to J/\psi \rho^+$ events. A bigger sample size allows for decision trees with more nodes, where for example correlations between input variables can be exploited further to separate signal from background.

$B^+ \to J/\psi \rho^+$ candidates from the upper sideband in the B^+-candidate's mass distribution ($M(B^+) \in [5700, 6000]\,\text{MeV}$) from 2012 data with magnet polarities Up and Down were used as background sample. Only a very small contribution of non-combinatoric background is expected there. The lower sideband was not used due to partially reconstructed background components in this mass region (see Sec 6.1). These were tackled with an additional BDT (see Sec 6.4.2).

Table 6 shows the variables that were chosen as BDT input. They were required to show a good discrimination power between signal and background (Fig. 17) and to be confirmed to be similar in data and simulation (see Sec. 5). For a definition of the quantities see Sec. 4.2.

The output distribution of the BDT against combinatoric background, in

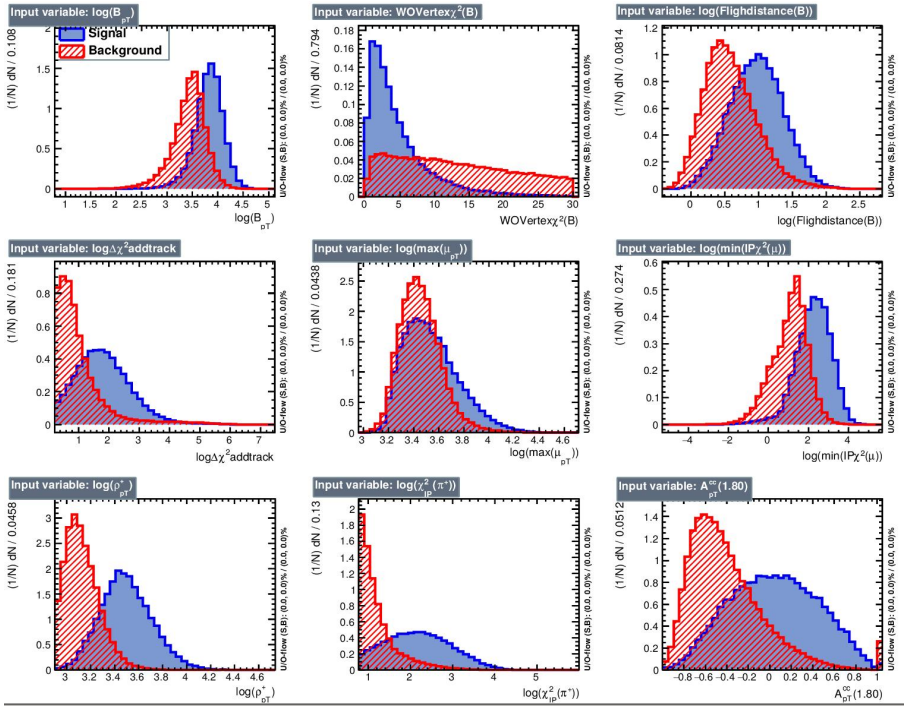

Figure 17: Input variable distributions for BDT against combinatorics. Sample of sWeighted $B^0 \to J/\psi K^{*0}$ candidates from 2012 Down data in blue (signal proxy), $B^+ \to J/\psi \rho^+$ candidates from upper sideband in 2012 Up and Down data in red (background).

$p_t(B^+)$	$\min(IP_{\chi^2}(\mu^+), IP_{\chi^2}(\mu^-))$
$p_t(\rho^+)$	$Vertex\chi^2(B^+)$
$\max(p_t(\mu^+), p_t(\mu^-))$	$\Delta\chi^2_{add-track}$
$A^{cone}_{p_t}(B^+)$	$FD(B^+)$
$IP_{\chi^2}(\pi^+)$	

Table 6: Input variables of the BDT against combinatoric background. For explanation of the quantities see Sec 4.2.

the following called BDTc, shows good separation (see Fig 18 on the left). The output distributions are very similar for test-sample and training sample. This implies the classifier is not very overtrained.

A small correlation between the output of the BDT and the reconstructed mass of the B-candidate from the background sample is visible (see Fig 18 on the right).

If the BDT was really sensitive to the mass of the reconstructed B-candidates,

it would simply use the fact that most of the signal events lie in the mass region $m(B) \in [5150, 5450]$, while the background sample comes from the mass region $m(B) \in [5700, 6000]$. The BDT would then classify every event outside $m(B) \in [5700, 6000]$ as signal, not only the "real" signal events. A cut on the BDT-output distribution would then distort the background distribution in the B-candidate mass spectrum such that background is more suppressed outside the signal region than inside. This would bias the result for the signal fit yield.

In this case the background sample is not totally clean: The correlation could come from partially reconstructed B-meson decays that have a tail into the upper sideband region. The BDT output of those events is usually higher than for combinatoric background. The correlation in the plot could just describe the mass distribution of those events inside the sideband. Additionally the correlation is very small and its influence on the branching fraction measurement is negligible.

As a cross-check the output distributions of the BDT against combinatoric background for a sample of simulated $B^+ \to J/\psi\,\rho^+$ events and a sample of sWeighted $B^0 \to J/\psi\,K^{*0}$ candidates were compared (see Fig. 27 in the appendix). The corresponding distributions of the BDT input-variables had been checked to be similar in the two samples, but different correlations between the variables inside the two samples could lead to different output distributions. The output distributions of the two samples are similar, which justifies the use of the sample of sWeighted $B^0 \to J/\psi\,K^{*0}$ candidates as signal proxy in the BDT.

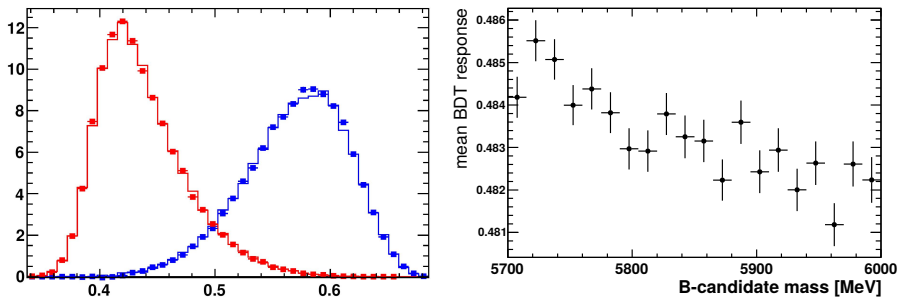

Figure 18: Left: BDT output distribution of BDT against combinatorics. sWeighted $B^0 \to J/\psi\,K^{*0}$ candidates from 2012 Down data in blue (signal proxy), $B^+ \to J/\psi\,\rho^+$ upper sideband 2012 Down data in red (histograms denote the test samples, squares denote training samples). Training and testing sample have very similar distributions both for signal and for background. This implies little overtraining. Right: mean BDT output for $B^+ \to J/\psi\,\rho^+$ candidates from upper sideband in 2012 Up data (background proxy) in bins of B-candidate mass with $BDTc > 0.45$. A small correlation is visible, probably coming from partially reconstructed background.

6.4.2 BDT Against Partially Reconstructed Background

A second BDT was trained to separate partially reconstructed background. A list of the variables that were used can be found in Table 7. The variables in the left column had already been used in the BDT against combinatoric background, while the variables on the right could not be used before: they have different distributions in $B^+ \to J/\psi\,\rho^+$ and $B^0 \to J/\psi\,K^{*0}$. Due to the extra track, the *DIRA* can be measured more precisely in $B^0 \to J/\psi\,K^{*0}$ and the extra track does also change the $\frac{\chi^2_{DTF}}{nDOF}(B)$-distribution. Additionally the variable CL(π^0), which shows good separating power (see Fig. 19 in the second row on the right), has no counterpart in $B^0 \to J/\psi\,K^{*0}$, because the channel does not contain a π^0. The sample of sWeighted $B^0 \to J/\psi\,K^{*0}$ candidates could therefore not be used as signal proxy in this BDT and simulated $B^+ \to J/\psi\,\rho^+$ events with settings for 2012 Magnet Down and Up were used instead.

The variables in the right column of Table 7 could be included in this BDT after it had been verified that they were distributed correctly by comparing with the channel $B^+ \to J/\psi\,K^{*+}$ (see Sec 5).

Simulated inclusive $B_{u/d/s} \to J/\psi\,X$ events were used as background proxy. The daughters of the J/ψ were required to go into the detector acceptance, the π^+ and J/ψ-candidates were truthmatched. Additionally it was required that they descend from a $B_{u/d/s}$-meson. To increase the background sample size, $B_{u/d/s} \to J/\psi\,X$ samples for 2011 and 2012 for both magnet polarities were combined.

Due to the large number of different decay channels included in the inclusive $B_{u/d/s} \to J/\psi\,X$-sample, a reweighting of variables, which would have to be done separately for each channel, was not possible. To avoid a reduced performance of the BDT from weighting only the signal sample, the $B^+ \to J/\psi\,\rho^+$-sample was not weighted either for the training of the BDT to separate partially reconstructed background. For the evaluation of the signal efficiency, a weighted signal MC was used later. As working point a cut on the output of the BDT against combinatoric background at 95 % background rejection was applied before the training ($\epsilon_{sig} = 83.6\,\%$).

$p_t(\rho^+)$	$DIRA(B^+)$
$\Delta\chi^2_{add-track}$	$\frac{\chi^2_{DTF}}{nDOF}(B^+)$
$\min(IP_{\chi^2}(\mu^+), IP_{\chi^2}(\mu^-))$	$CL(\pi^0)$
$IP_{\chi^2}(\pi^+)$	

Table 7: Input variables of the BDT against non-combinatoric background. For explanation of the quantities see Sec 4.2.

The distributions of the input variables for signal and background samples can be seen in Fig. 19.

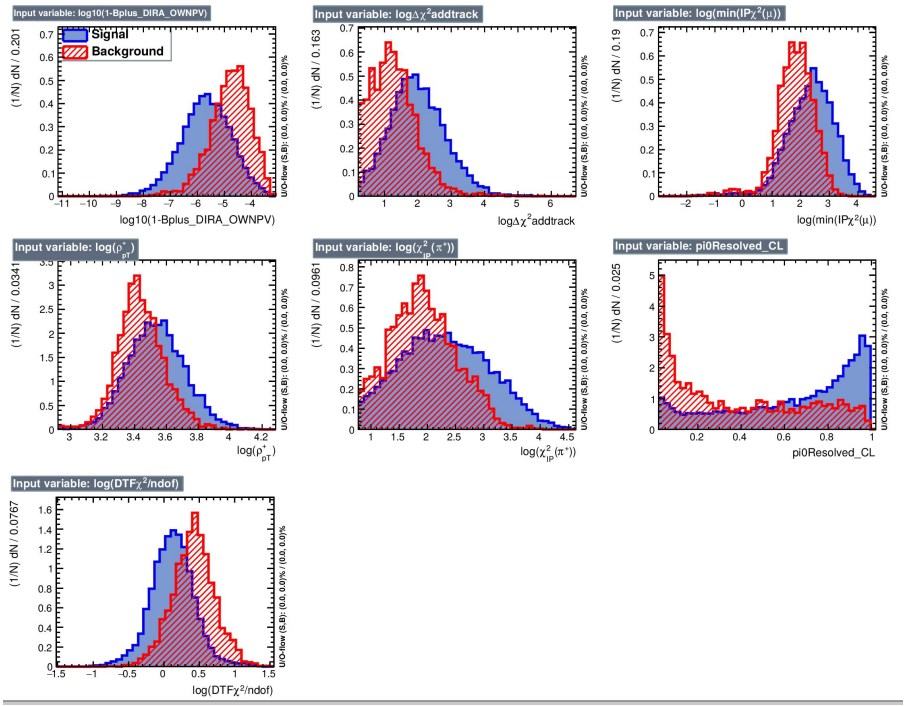

Figure 19: Input variable distributions for BDT against partially reconstructed background. Simulated $B^+ \to J/\psi\,\rho^+$ events with 2012 Magnet Up and Down settings in blue (signal proxy), simulated inclusive $Bu/d/s \to J/\psi\,X$ events with 2011+2012 Magnet Up+Down settings in red (background proxy).

The BDT output, in the following called BDTp, shows very similar distributions for test-sample and training-sample. This implies little overtraining (see Fig 20 on the left). Additionally BDTp shows only a slight correlation with the B-candidate mass in the upper sideband of the 2012 Magnet Down data sample. This is probably due to partially reconstructed background contribution as for the BDT against combinatoric background (see Fig 20). The effect is negligibly small.

6.5 Efficiencies

To extract the branching ratio of the signal decay from the determined yields, it is necessary to know the reconstruction efficiency for both the signal decay and the normalisation decay. The geometric acceptance of the detector and the efficiencies of the preselection were taken from MC simulated events with no additional weights applied.

The PID-cut efficiency was determined with the `PIDCalib` package. During the stripping loose cuts had already been applied on the inaccurately simulated

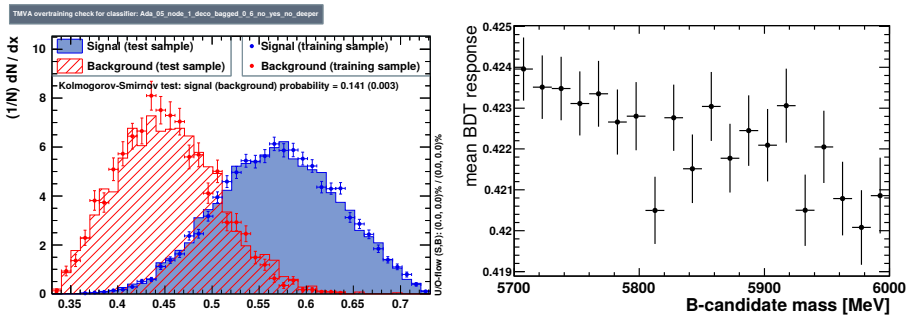

Figure 20: Left: Output distribution of BDT against partially reconstructed background. simulated $B^+ \to J/\psi\,\rho^+$ events with 2012 Magnet Up and Down settings in blue (signal proxy), simulated inclusive $Bu/d/s \to J/\psi\,X$ events with 2011+2012 Magnet Up+Down settings in red (background proxy). Histograms denote the test samples, points denote the training samples. Training and testing sample have very similar distributions, both for signal and for background samples. This implies little overtraining. Right: mean BDT output for $B^+ \to J/\psi\,\rho^+$ candidates from upper sideband in 2012 Magnet Up data (background proxy in the BDT to separate combinatoric background) in bins of B-candidate mass with $BDTc > 0.45$. A small correlation is visible, probably coming from partially reconstructed background.

PID-variables and PID efficiency could only be determined afterwards. This introduced a small systematic uncertainty.

Both BDT contain variables that had been reweighted, hence the efficiencies of both BDT output cuts were determined with reweighted simulated signal events. The determination of the actual cut values is described in Sec. 6.6.

All these efficiencies were determined separately for samples with 2012 and 2011 Magnet polarity Up and Down configuration (Table 8 for 2012 Magnet Down, other configurations Table 11, 12, 13 in appendix).

6.6 Cut Optimisation

The cuts on the outputs of the two BDT needed to be chosen such that signal and background were well distinguishable from each other in the mass spectrum of the B-candidates. The following subsection details the selection of these cuts.

The optimisation was performed maximizing a Figure of Merit $FoM = \frac{N_{\mathrm{sig}}}{\sqrt{N_{\mathrm{sig}}+N_{\mathrm{bg}}}}$, where N_{sig} denotes the expected number of signal events and N_{bg} the expected number of background events inside the signal region.

These numbers could not be taken from a fit to data including the signal region, because this would have given rise to a bias: The maximum FoM could then be caused by a statistical fluctuation giving a higher signal yield from the fit, which would directly translate into a higher branching ratio result. Therefore

Selection	$B^+ \to J/\psi\,\rho^+$ Candidates	ϵ_{cut}	$B^0 \to J/\psi\,K^{*0}$ Candidates	ϵ_{cut}
Events in Acceptance	996 620	15.2%	4 435 958	16.1%
Stripping Selection	22 058	2.2%	416 104	9.4%
Trigger Cuts	16 853	76.4%	332 415	79.9%
Offline Cuts	12 490	74.1%	271 429	77.0%
PID Cuts	8889	71.2%	213 641	83.5%
Weighted Events	7515			
BDTc & BDTp Cut	3699	49.2%	–	–
wrt Stripping	–	19.8%	–	51.3%
Total Efficiency		0.439%		4.82%

Table 8: Number of remaining events and resulting efficiency for the simulated 2012 Down signal channel sample and simulated 2012 Down normalization channel sample after each selection step. Efficiencies are relative to number of events passing all prior cuts. PID efficiency was taken from `PIDCalib` (see Sec. 5.1), BDT efficiencies were taken from weighted MC simulation.

the numbers were extracted separately.

For N_{bg} a fit was performed in the sidebands of the B-candidate mass distributions, where signal contribution is small. To avoid contamination by partially reconstructed background with more than one missing pion and still allow for picking up the shape of partially reconstructed background with one missing pion, the lower sideband was chosen to start at 4900 MeV. For the signal region $M(B) \in [5150, 5450]$ MeV was chosen, which contained approximately 96% of the simulated signal events after preselection. This left for the sidebands the ranges $m(B) \in [4900, 5150]$ MeV and $m(B) \in [5450, 6000]$ MeV. The background PDFs were then interpolated into the signal mass region to determine N_{bg}. To increase the fit stability, the whole dataset was used for the sideband fits, both from 2011 and 2012 and Magnet polarity Up and Down.

To avoid the aforementioned bias, the signal yield N_{sig} was calculated from the branching fraction $\mathcal{B}(B^+ \to J/\psi\,\rho^+) = 5.0 \pm 0.8$ from the previous measurement by BaBar (see Ref. [1]) by restructuring the equation for the branching ratio determination:

$$N_{\text{sig}} = \frac{\mathcal{B}_{\text{sig}}}{\mathcal{B}_{\text{norm}}} \times N_{\text{norm}} \times \frac{\xi_{\text{sig}}}{\xi_{\text{norm}}} \times \frac{\epsilon_{\text{sig}}^{\text{reco, presel}}}{\epsilon_{\text{norm}}^{\text{reco, presel}}} \times \epsilon^{BDTc} \times \epsilon^{BDTp}$$

The variables are explained in Sec 4.

The contribution to the data sample from the different partially reconstructed decays introduced in Sec. 6.1 was estimated from simulated samples of the respective decay chains. Of the background channels introduced in Sec. 6 only

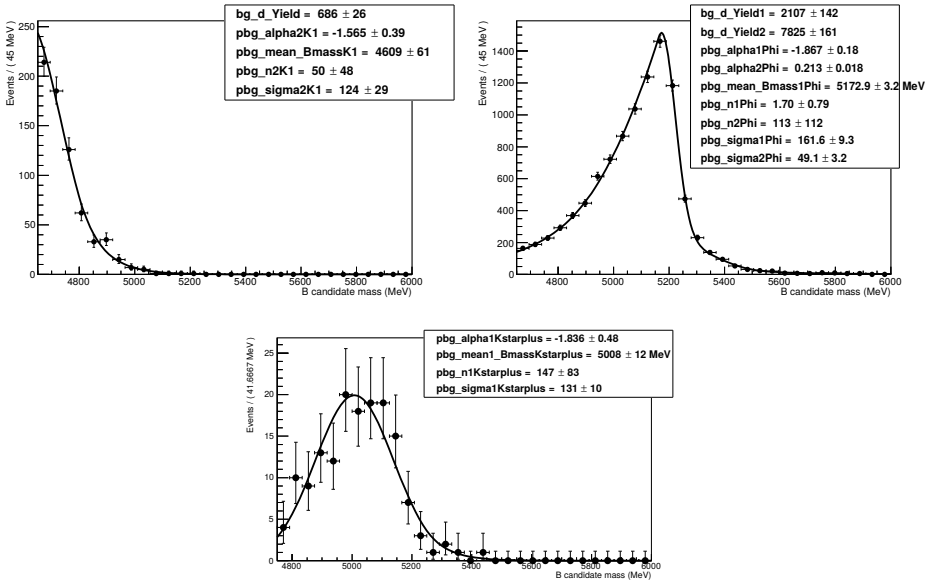

Figure 21: Partially reconstructed background simulations after selection ($BDTc > 0.53$, $BDTp > 0.52$). For the fit to data only events with masses above 4900 MeV were considered (see Fig. 22). The number of events depicted are not to scale with each other, they are not normalized either to give an impression of the statistical significance of the extracted shapes. Left: $B^+ \to J/\psi K_1(1270)^+$ simulation, only a small fraction of events lies in the fit region. Right: $B_s^0 \to J/\psi\,\phi$ simulation fitted with double CB, Bottom: $B^+ \to J/\psi K^{*+}$ simulation fitted with CB, because due to a small sample size only few events are left after selection.

the following channels were expected to contribute to the dataset with more than 5% of the signal yield for the chosen cut ranges in $BDTc$ and $BDTp$:

- $B^+ \to J/\psi K_1(1270)^+$: the restriction to masses above 4900 MeV in the fits reduced its contribution to a negligible amount (see Fig. 21 on the left), it was therefore not used as a fit component.

- $B_s^0 \to J/\psi\phi$: it was described with two CB with common mean. A definition of the CB pdf is given in Sec 5. The simulated sample is large enough to distinguish the two shapes. For the sideband fits the shapes were fixed to those obtained in the fit to simulation (see Fig. 21 on the right).

- $B^+ \to J/\psi K^{*+}$: after preselection only 500 events were left from the simulated sample of this decay. To reduce overfitting, it was described by only one CB function (see Fig. 21 on the bottom). Due to their big uncertainties, its fit parameters were allowed to vary in the sideband fit

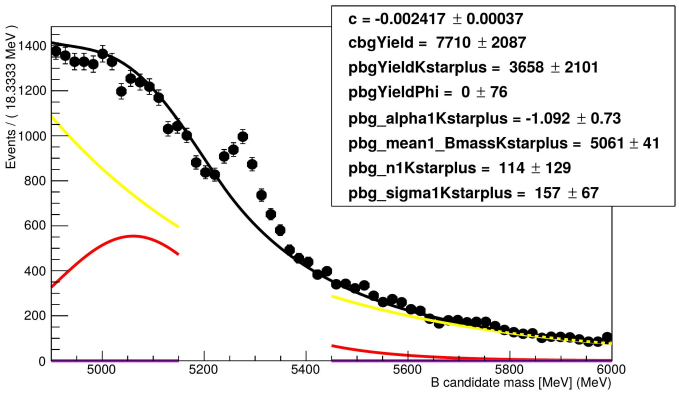

Figure 22: Sideband fit to full dataset after selection ($BDTc > 0.53$, $BDTp > 0.52$). Combinatoric background in yellow, $B_s^0 \to J/\psi\,\phi$ in violet, $B^+ \to J/\psi\,K^{*+}$ in red, and combined pdf interpolated into signal region in black. Due to the loose constraints on the shape of the contribution from $B^+ \to J/\psi\,K^{*+}$, the background from $B_s^0 \to J/\psi\,\phi$ was not picked up by the fit. Nevertheless the expected background contribution inside the signal region is described reasonably well.

within Gaussian constraints[3].

The shapes of the simulated partially reconstructed backgrounds were extracted for every tested *BDTc* and *BDTp* cut combination, because their shapes varied with the cuts. Combinatoric background was described with an exponential distribution with free decay constant. An example of a sideband fit can be seen in Fig. 22. The flexibility of the CB describing background from $B^+ \to J/\psi\,K^{*+}$ resulted for some cut-combinations in the fit not picking up the contribution from $B_s^0 \to J/\psi\,\phi$. Nevertheless for an estimation of the *FoM* the background in the signal region is described well enough.

Combinations of cuts on the output of the BDT against combinatoric background (*BDTc*) and output of the BDT against partially reconstructed background (*BDTp*) were tested on a grid between $BDTc \in [0.5, 0.56]$, $BDTp \in [0.52, 0.62]$ (efficiency on reweighted simulated $B^+ \to J/\psi\,\rho^+$ events with respect to no BDT cuts: 71% resp. 20%).

The resulting *FoM* had a large relative uncertainty from the uncertainties of the background fit yields, effectively allowing a wide range of cuts (see Fig. 23). The output cut combination, which gave the maximum *FoM* was then chosen for the yield-extraction fit:

$BDTc > 0.53$, $BDTp > 0.56$

[3]The constraints were realized in Roofit by introducing an additional factor to the likelihood $\mathcal{L}$: a multi-variate Gaussian with mean, width, and linear correlations from the parameter estimates of a fit to simulation.

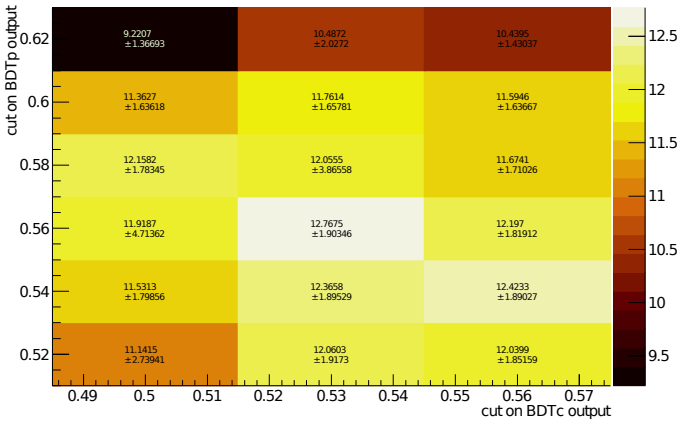

Figure 23: Figure of Merit for different BDTc and BDTp cut combinations. Uncertainties from statistical uncertainties on fit yields and uncertainty on $\mathcal{B}(B^+ \to J/\psi\,\rho^+)$ only.

For comparison between data and simulation, the mass distributions after selection are shown in Fig. 24 for B-candidates and J/ψ-candidates and in Fig. 25 for ρ^+-candidates and π^0-candidates. Depicted are signal simulation and data from 2012.

- The B-candidate mass distribution shows a clear peak in the signal region and an elevated lower sideband, which is constituted from partially reconstructed background and combinatoric background together (see Fig. 24 on the left). The upper sideband shows no significant features.

- The mass distribution of the J/ψ-candidates is very similar in data and simulation and very little background is visible (see Fig. 24 on the right).

- The mass distribution of ρ^+ in data is taken from only the signal B-mass region $M(B^+) \in [5150, 5450]$ to increase the visibility of the ρ^+-mass peak. A significant background contribution is visible in the ρ^+-candidate mass spectrum (see Fig. 25 on the left).

- The mass distribution of π^0-candidates shows a significant background contribution as well (see Fig. 25 on the right). One must note that the measured mass of π^0 was not used in the calculation of the mass of B^+, but the PDG mass was used instead. This is explained in Sec. 4.2 in the introduction of the Decay Tree Fitter.

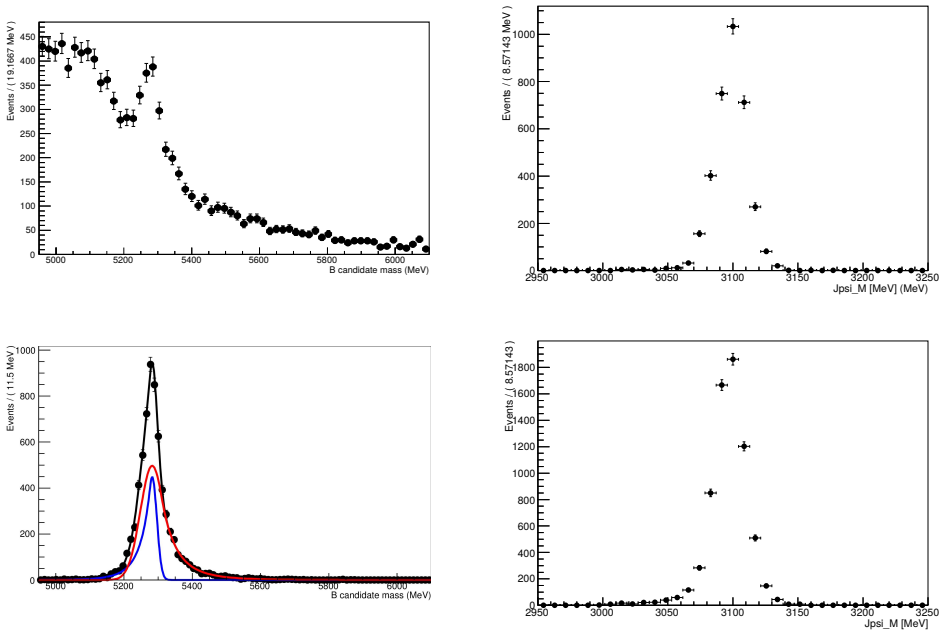

Figure 24: Top Left: B-candidate mass distribution in 2012 data sample after selection. Bottom Left: B-candidate mass distribution from simulated $B^+ \to J/\psi\,\rho^+$ events after selection. The fit pdf is shown as well. Top Right: J/ψ-candidate mass distribution in 2012 data sample after selection. Bottom Right: J/ψ-candidate mass distribution from simulated $B^+ \to J/\psi\,\rho^+$ events after selection.

6.7 Fit Description

The fit on the $B^+ \to J/\psi\,\rho^+$ data was performed on the mass spectrum of the B-candidate in the range $m(B) \in [4950, 6100]$. The lower bound was chosen to reduce contamination from partially reconstructed background with two missing pions, while no additional information for the fit was expected beyond 6100 MeV. A more reliable way of extracting the signal yield would have been a two-dimensional fit together with the mass distribution of the ρ^+-candidates. This could have decreased the statistical uncertainty on the yield, but due to the complexity of such a fit, there was not enough time to implement a model for the ρ^+-candidate mass distribution.

The fit model had the following components, only parameters left free to float in the fit are listed:

- Because it contains a π^0 in the decay chain, the signal channel $B^+ \to J/\psi\,\rho^+$ was described in the fit by a double crystal ball pdf (see Sec. 5.4). It was checked in the control channel $B^+ \to J/\psi\,K^{*+}$ that the shape parameters of a fit to the B mass peak agree within uncertainties between simulation

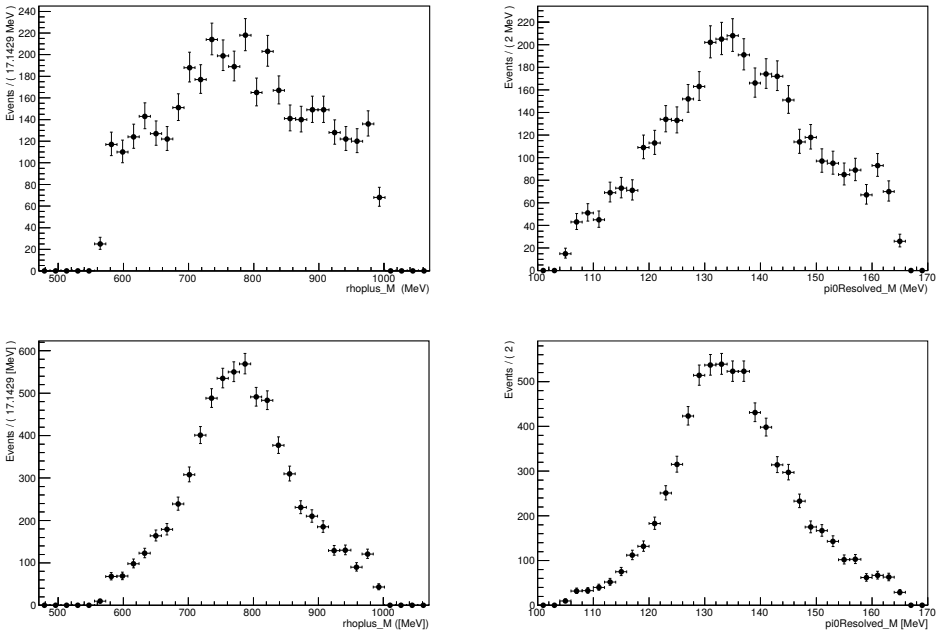

Figure 25: Top Left: ρ^+-candidate mass distribution in 2012 data sample after selection for $M(B) \in [5150, 5450]$. Bottom Left: ρ^+-candidate mass distribution from simulated $B^+ \to J/\psi\,\rho^+$ events after selection. Top Right: π^0-candidate mass distribution in 2012 data sample after selection. Bottom Right: π^0-candidate mass distribution from simulated $B^+ \to J/\psi\,\rho^+$ events after selection.

and data.

Because of the similarity between the two channels $B^+ \to J/\psi\,\rho^+$ and $B^+ \to J/\psi\,K^{*+}$, no disagreement in the B mass shapes between simulation and data in the channel $B^+ \to J/\psi\,\rho^+$ were expected either. The parameters of the signal pdf were therefore fixed to values obtained from a fit to simulated $B^+ \to J/\psi\,\rho^+$ events and only the common mean of the Gaussian cores was left to float: $\mathcal{P}_{\text{sig}}(m|\mu)$

- Combinatoric background was described with an exponential distribution with floating decay constant c: $\mathcal{P}_{\text{cbg}}(m|c)$.
- Background from $B_s^0 \to J/\psi\,\phi$ was described with a double CB pdf. Its shape and mean were fixed to a fit to simulation: $\mathcal{P}_\phi(m)$
- Background from $B^+ \to J/\psi\,K^{*+}$ was described with a single CB pdf, because of the little sample size of its simulation. From simulation it was known that the power-law tail of the distribution, described by α and n, only set in well below the fit region for the chosen BDT cut-combination

(see Fig. 21). These values were therefore not left to float. The mean and width of the Gaussian were tried to be left floating inside the constraints from the fit to simulation like in the fit to the sidebands (see Sec. 6.6). This resulted in unphysically large yields from this background channel, therefore all parameters of this background were fixed for the final fit: $\mathcal{P}_{K^{*+}}(m)$A larger sample of simulated events is in production at this writing, which should help to solve this problem.

- Combined partially reconstructed background from various sparsely contributing channels was described by a CB with all its parameters left to float: $\mathcal{P}_{\mathrm{CB}}(m|\alpha, n, \mu, \sigma)$

This combines to:

$$\mathcal{P}_{\mathrm{dCB}} = f_{\mathrm{sig}}\mathcal{P}_{\mathrm{sig}}(m|\mu)+f_{\mathrm{cbg}}\mathcal{P}_{\mathrm{cbg}}(m|c)+f_{\phi}\mathcal{P}_{\phi}(m)+f_{K^{*+}}\mathcal{P}_{K^{*+}}(m)+f_{\mathrm{CB}}\mathcal{P}_{\mathrm{CB}}(m|\alpha, n, \mu, \sigma)$$

The datasets of 2011 and 2012 were fitted simultaneously by optimizing a combined likelihood to help constrain the shape of the backgrounds, while the yields of the different components were floating independently.

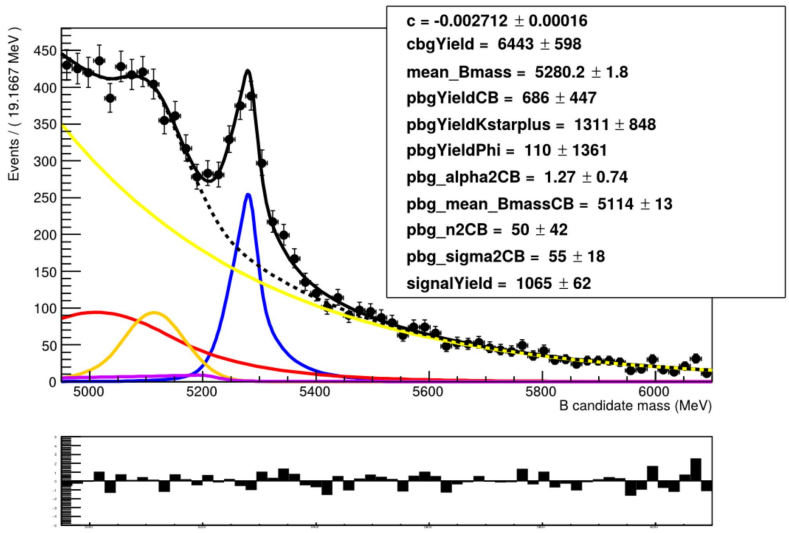

Figure 26: Fit of B-candidate mass after selection. Data sample from 2012 Magnet polarity Up and Down combined. $B^+ \to J/\psi\,\rho^+$ signal described by double CB pdf in blue, combinatoric background described by an exponential pdf in yellow, partially reconstructed background from $B_s^0 \to J/\psi\,\phi$ described by a double CB pdf in violet, partially reconstructed background from $B^+ \to J/\psi\,K^{*+}$ described by a CB pdf in red, combined partially reconstructed background described by a CB pdf in orange.

The fit for 2012 can be seen in Fig. 26, the one for 2011 is in the appendix Fig. 28. The derived signal yields are 358 ± 35 (1065 ± 62) for 2011 (2012).

7 Systematic Uncertainties

This section describes the different contributions to the systematic uncertainty on the branching ratio. The values are given for 2011 and 2012 results separately in case they differ between the samples. The branching fraction of $B^+ \to J/\psi\,\rho^+$ was determined relative to the normalisation channel $B^0 \to J/\psi\,K^{*0}$, which has very similar kinematics. In the ratio of the efficiencies, possible systematic uncertainties on the trigger efficiency or the luminosity cancel therefore largely.

The sources of systematic uncertainty studied are the following:

- Mass window: The fit was repeated in the following fit ranges: $M(B) \in [4900, 6100], [5000, 6100], [4950, 5900], [4950, 6200]$. The maximum difference in signal yield with respect to the actual fit was taken as systematic uncertainty.

- $B^+ \to J/\psi\,K^{*+}$ shape: In the fit the shape parameters of the partially reconstructed background contribution from $B^+ \to J/\psi\,K^{*+}$ were fixed to the values obtained from simulation. To estimate the effect of this on the measured branching ratio, the mean and width of the background from $B^+ \to J/\psi\,K^{*+}$ were left floating and the difference in the $B^+ \to J/\psi\,\rho^+$ yields with respect to the nominal fit was taken as systematic uncertainty (see Fig. 29).

- additional background shape: In the fit, a CB with all parameters free was used to describe additional partially reconstructed background in the lower sideband. To evaluate how the choice of using a CB influenced the result, the fit was repeated with a Gaussian with free parameters instead

Source	Uncertainty 11 (%)	Uncertainty 12 (%)
Mass window	± 3.4	± 2.5
$B^+ \to J/\psi\,K^{*+}$ shape	± 17.9	± 3.9
additional background shape	± 0.8	± 0.7
Simulated sample size	± 2.9	± 2.2
Geometric acceptance	± 0.3	± 2.2
Data-simulation difference	± 1.3	± 0.8
Normalisation sample size	± 0.7	± 0.5
non-resonant $B^+ \to J/\psi\,\pi^+\pi^0$	$^{+0}_{-19.4}$	$^{+0}_{-15.9}$
Sum	$^{+18.5}_{-26.8}$	$^{+5.7}_{-20.2}$

Table 9: Different contributions to systematic uncertainty for 2011 and 2012. The last row gives the quadratic sum of the entries.

of a CB. The difference between the resulting $B^+ \to J/\psi\,\rho^+$ yields with respect to the nominal fit was taken as uncertainty.

- Simulated sample size: The Simulation of the signal sample was used to determine the efficiency of reconstruction and selection $\epsilon_{\text{sig}}^{\text{reco, presel}}$, ϵ^{bdtc}, ϵ^{bdtp}. The uncertainty on $\epsilon_{\text{sig}}^{\text{reco, presel}}$ was calculated from the unweighted events while the uncertainty on $\epsilon^{\text{bdtc}} \times \epsilon^{\text{bdtp}}$ was calculated using the effective number of events after the reweighting.

- Geometric acceptance: The value for geometric acceptance is taken from simulation. Due to a finite sample size, it introduces a systematic uncertainty. The value given in the table is determined from the generator statistics.

- Data-simulation difference: Although the signal simulation had been reweighted, it could still be distributed differently with respect to data. The remaining differences between simulation and data are smaller than those from the reweighting. Therefore conservatively the difference between the branching ratio calculated using reweighted events and unweighted events was taken as systematic uncertainty.

- Normalisation sample size: The finite size of the normalisation sample is a source of systematic uncertainty as well. The uncertainty on the fit yield of the normalisation channel $B^0 \to J/\psi\,K^{*0}$ was taken for this contribution.

- non-resonant background: The channel $B^+ \to J/\psi\,\pi^+\pi^0$ can contribute to the signal yield. The upper boundary on this channel is $\mathcal{B} < 7.3 \times 10^{-6}$@ 90% CL (see Ref. [1]). In the mass spectrum of the B-candidates it can't be distinguished from signal, but with an additional fit to the ρ-candidate mass distribution one could disentangle the yield of this channel from the yield of $B^+ \to J/\psi\,\rho^+$.

  For this thesis the upper boundary is taken as estimate. This is very conservative, because a large fraction of this background is cut away by the mass constraints on the ρ^+ candidates $m(\rho^+) \in [570, 995]$. With a dedicated MC simulation of this channel one could determine the relative efficiencies for $B^+ \to J/\psi\,\rho^+$ and $B^+ \to J/\psi\,\pi^+\pi^0$, which would result in a less conservative estimate on this systematic. Such a simulation would be helpful for the description of this background in a potential fit to the ρ^+-candidate-mass as well.

Values for the different contributions are given in Table 9. Large contributions come from the fit-shape of $B^+ \to J/\psi\,K^{*+}$, especially for the 2011 sample. The fit can be seen in Fig. 29 in the appendix. The shapes and relative yields of the partially reconstructed background contributions differ significantly with

respect to the nominal fit (see Fig. 28 and 26). Due to the lower sample size in the 2011 fit, this has a bigger effect on the $B^+ \to J/\psi\,\rho^+$ yield there.

A less conservative approach to this systematic could have been a pseudo-experiment, where the shape of the background from $B^+ \to J/\psi\,K^{*+}$ would be fixed to a random value inside the range given by the errors on the fit to simulation. This fit would then be repeated a few hundred times and the spread in yield results could be used to estimate the actual dependency of the branching ratio on the background shape. Due to limited time, this is not possible within the scope of this thesis. Additionally at the time of this writing a larger simulation of this background is in production, which should help decrease the uncertainty from its shape.

The normalisation channel branching ratio is known with an accuracy of 4.5 %, which directly translates into an uncertainty on the branching ratio of $B^+ \to J/\psi\,\rho^+$.

This list is not complete, due to time constraints not all contributions could be investigated in this thesis:

Although the signal model fits the simulated events nicely, another model like *e.g.* a Bukin distribution (see Ref. [44]) for the B mass peak of $B^+ \to J/\psi\,\rho^+$ needs to be tested to evaluate a possible dependence of the signal yield on the model. The uncertainty from the choice of model usually lies below 5 % (see e.g. Ref. [45], [46])

The procedure of resampling the particle identification variable for simulation leads to a systematic uncertainty. In other LHCb analyses with a pion in the signal channel and a Kaon in the normalisation channel they are found to be around 3 % (see e.g. Ref. [45]).

Uncertainties on the trigger efficiency in the same analysis are found to be 1.1 % (see Ref. [45]).

Uncertainties on the reconstruction and preselection efficiency have to be included as well. The differences between efficiencies from data and simulation for each track of muons are found to be 0.8 % and for each hadron of 1.4 % (see Ref. [47]). The value for π^0 is supposed to be a little higher than for charged hadrons. The quadratic sum should be below 5 % as well, because in the ratio with the efficiency of the normalisation channel the uncertainties from the track simulation largely cancel.

Added up quadratically these additional contributions result in less than 8 % uncertainty.

8 Branching Fraction Determination

This section describes the extraction of the branching fraction $\mathcal{B}_{\text{sig}}$, which is done using the formula introduced in Sec. 4:

$$\mathcal{B}_{\text{sig}} = \frac{\mathcal{B}_{\text{norm}}}{\mathcal{B}_{\pi^0\to\gamma\gamma}} \times \frac{N_{\text{sig}}}{N_{\text{norm}}} \times \frac{\xi_{\text{norm}}}{\xi_{\text{sig}}} \times \frac{\epsilon_{\text{norm}}^{\text{reco, presel}}}{\epsilon_{\text{sig}}^{\text{reco, presel}} \times \epsilon^{\text{bdtc}} \times \epsilon^{\text{bdtp}}},$$

where the values for the branching fractions were taken from PDG (see Ref. [26]). The branching fraction of the normalisation channel $B^0 \to J/\psi K^{*0}$ with $\mathcal{B}(K^* \to K^+\pi^-) = 2/3$ is described by $\mathcal{B}_{\text{norm}} = (\frac{2}{3} \times 1.32 \pm 0.06) \times 10^{-3}$. Here isospin symmetry was assumed (see footnote Sec. 4). $\mathcal{B}_{\pi^0\to\gamma\gamma} = 0.98823 \pm 0.00034$ is the percentage of π^0 decaying into two photons.

In Sec. 6.7 the fit yields were determined to be $N_{\text{sig}} = 358 \pm 33$ (1065 ± 63) for the data sample from year 2011 (2012).

The yield from a fit to the normalisation channel, N_{norm} is shown in Fig. 9 on the right for the 2012 Down data sample. The values for different years and magnet polarities are given in Table 10.

ξ are the geometric acceptances taken from the generator statistics. The efficiencies $\epsilon^{\text{reco, presel}}$ of reconstruction and preselection were taken from unweighted simulation, while ϵ^{bdtc} and ϵ^{bdtp} are the efficiencies of the cuts on BDTp and BDTc, taken from weighted simulation. In Sec. 6.5 is described, how ξ and ϵ were determined.

The fit to data was performed on samples containing both polarities to increase the fit stability. The integrated luminosity, efficiencies, and geometric acceptances differed between the polarities, which is why a weighted mean between Up and Down was taken for the acceptance and efficiency. Here for example for the efficiency:

$$\epsilon = \frac{(\mathscr{L}^{\text{U}}\epsilon^{\text{U}}\xi^{\text{U}}) \times \epsilon^{\text{U}} + (\mathscr{L}^{\text{D}}\epsilon^{\text{D}}\xi^{\text{D}}) \times \epsilon^{\text{D}}}{\mathscr{L}^{\text{U}}\epsilon^{\text{U}}\xi^{\text{U}} + \mathscr{L}^{\text{D}}\epsilon^{\text{D}}\xi^{\text{D}}}$$

This gives the following results for the branching fraction:

	N_{norm}	Luminosity $\mathscr{L}[\,\text{pb}^{-1}]$
2011 polarity Up	$45\,741 \pm 253$	417
2011 polarity Down	$63\,375 \pm 300$	559
2012 polarity Up	$125\,056 \pm 430$	999
2012 polarity Down	$125\,789 \pm 253$	988

Table 10: Yields of fits to B-candidate mass distribution for data samples from different years and magnet polarities and integrated luminosity in the respective data sample.

$$2011:\ \mathcal{B}(B^+ \to J/\psi\,\rho^+) = (3.76 \pm 0.37(\text{stat})^{+0.70}_{-1.01}(\text{syst}) \pm 0.17(\text{norm})) \times 10^{-5}$$

$$2012:\ \mathcal{B}(B^+ \to J/\psi\,\rho^+) = (4.59 \pm 0.27(\text{stat})^{+0.26}_{-0.77}(\text{syst}) \pm 0.21(\text{norm})) \times 10^{-5}$$

The first uncertainty is statistical, the second is systematic, as given from the quadratic sum of the entries in Table 9, the third comes from the uncertainty on the normalisation channel branching fraction.

The statistical uncertainties are uncorrelated between 2011 and 2012 results, the systematic uncertainties are partly correlated due to the contribution from the non-resonant decay $B^+ \to J/\psi\,\pi^+\pi^0$, and the uncertainty from the normalisation channel is 100 % correlated. For the combined result this gives

$$\mathcal{B}(B^+ \to J/\psi\,\rho^+) = (4.31 \pm 0.22(\text{stat})\ ^{+0.35}_{-0.81}(\text{syst})) \times 10^{-5}.$$

With the preliminary estimates for systematic uncertainties from not-yet-studied sources, which were described in Sec. 7 included, this results in

$$\mathcal{B}(B^+ \to J/\psi\,\rho^+) = (4.31 \pm 0.22(\text{stat})\ ^{+0.49}_{-0.88}(\text{syst})) \times 10^{-5}.$$

The accuracy of this value is already at the level of the only previous measurement of this decay (see Ref. [1]):

$$\mathcal{B}(B^+ \to J/\psi\,\rho^+)_{\text{BaBar}} = (5.0 \pm 0.7(\text{stat}) \pm 0.3(\text{syst})) \times 10^{-5}$$

The BaBar measurement has a much higher statistical uncertainty due to the smaller sample of B-mesons, which resulted in an $B^+ \to J/\psi\,\rho^+$ yield of 219 after selection. Its systematic error is much lower, mostly because they estimated the contribution from non-resonant $B^+ \to J/\psi\,\pi^+\pi^0$ to their signal yield through a fit to the ρ^+-candidate mass. Both measurements agree within their uncertainties.

For easier comparison statistical and systematic combined:

$$\mathcal{B}(B^+ \to J/\psi\,\rho^+) = 4.31^{+0.54}_{-0.91} \times 10^{-5}$$

$$\mathcal{B}(B^+ \to J/\psi\,\rho^+)_{\text{BaBar}} = (5.0 \pm 0.8) \times 10^{-5}$$

The uncertainty from the non-resonant decay in this measurement leads to a higher uncertainty towards lower values with respect to the BaBar measurement. The uncertainty towards higher values is smaller by one third with respect to the old measurement.

In the ongoing analysis the systematic uncertainties will diminish, especially the lower bound is likely to decrease drastically by extracting the non-resonant yield from a simultaneous fit to the ρ^+-candidate mass distribution (see Sec. 7).

9 Summary and Outlook

The branching fraction of the decay $B^+ \to J/\psi\,\rho^+$ with $\rho^+ \to \pi^+\pi^0$ was measured using the Run I data sample of the LHCb experiment, corresponding to an integrated luminosity of $\mathscr{L} = 3\,\mathrm{fb}^{-1}$.

It resulted in the following value:

$$\mathcal{B}(B^+ \to J/\psi\,\rho^+) = (4.31 \pm 0.22(\mathrm{stat})\ ^{+0.49}_{-0.88}(\mathrm{syst})) \times 10^{-5}.$$

The asymmetric systematic uncertainty comes from possible non-resonant background from $B^+ \to J/\psi\,\pi^+\pi^0$ for which only an upper limit is known. This measurement for $\mathcal{B}(B^+ \to J/\psi\,\rho^+)$ is consistent with and of a similar precision as the only previous measurement by the BaBar collaboration (see Ref. [1]). Combining the two measurements into a new world-average reduces the 1-σ-region of the value by more than 30 %.

Additionally it is expected that in the ongoing analysis two major sources of uncertainty are going to diminish further:

The uncertainty from the shape of the background contribution from $B^+ \to J/\psi\,K^{*+}$ should decrease as soon as a new simulated sample is ready. With this a more reliable fit to its B-candidate mass distribution can be obtained.

Contribution from non-resonant $B^+ \to J/\psi\,\pi^+\pi^0$ could be measured by simultaneously fitting B-candidate and ρ- candidate mass. This would also help distinguish signal from partially reconstructed backgrounds like $B^+ \to J/\psi\,K^{*+}$ with $K^* \to K^0_S\pi^+$, where the π^+ and π^0 do not form a resonance.

After the branching fraction measurement is completed, a measurement of the direct CP-asymmetry $A_{\mathrm{CP}}(B^+ \to J/\psi\,\rho^+)$ is planned. The only previous measurement by the BaBar collaboration $A_{\mathrm{CP}} = 0.11 \pm 0.14$ (see Ref. [1]) was statistically limited, which implies that a measurement using the larger LHCb data sample can increase the precision on that value significantly.

The Standard Model predicts a value of $A_{\mathrm{CP}} \lesssim 1\%$ (see Ref. [2]), a precision measurement of this quantity would therefore put additional constraints on the model and help refine theory predictions for other channels as well.

This analysis is of course only a small part of the extensive physics programme at LHCb (see Ref. [48]). The experiment is now in Run II, collecting data at 13 TeV center-of-mass-energy. With the larger data sample, it will be possible to measure both $\mathcal{B}(B^+ \to J/\psi\,\rho^+)$, $A_{\mathrm{CP}}(B^+ \to J/\psi\,\rho^+)$ and the many other parameters constraining the Standard Model with an even better precision in the future.

A Appendix

A.1 ${}_s\mathcal{P}lot$ Technique

The sPlot technique is a tool to statistically unfold a signal component from a background polluted sample through a fit to a variable, for which one can describe signal and background (see Ref. [25]). The resulting weighted sample then has the same distributions as the signal in every variable not correlated to the variable, which was used for fitting:

A sample consisting of a signal contribution $f_\mathrm{S} = f_\mathrm{S}(x,y) = f_\mathrm{S}(x)f_\mathrm{S}(y)$ and a background contribution $f_\mathrm{B} = f_\mathrm{B}(x,y) = f_\mathrm{B}(x)f_\mathrm{B}(y)$, dependent on two uncorrelated variables x and y each, can be written as

$$f_\mathrm{tot}(x,y) = N_\mathrm{S} f_\mathrm{S}(x) f_\mathrm{S}(y) + N_\mathrm{B} f_\mathrm{B}(x) f_\mathrm{B}(y).$$

One is interested in the distribution

$$N_S f_S(y) = \int dx f_\mathrm{tot}(x,y) w(x)$$

with according weight $w(x)$ such that

$$\int dx f_\mathrm{S}(x) w(x) = 1$$

$$\int dx f_\mathrm{B}(x) w(x) = 0.$$

A function that fulfils these requirements and additionaly minimises the statistical uncertainty on the resulting distribution is the following:

$$w(x) = \frac{V_\mathrm{SS} f_\mathrm{S}(x) + V_\mathrm{SB} f_\mathrm{B}(x)}{N_\mathrm{S} f_\mathrm{S}(x) + N_\mathrm{B} f_\mathrm{B}(x)},$$

where V_SS and V_SB are the respective entries of the inverse covariance matrix of a fit in the variable x to the data sample. x is then called the discriminating variable.

Applying these weights on the data sample results in a sample that exhibits a clean distribution of signal in the variable y.

A.2 Comparison between BDT Output for Signal Simulation with Normalisation Channel Data

The BDT output distribution against combinatoric background (see Sec. 6.4.1) was compared between simulated $B^+ \to J/\psi\,\rho^+$ events with 2012 Up settings and sWeighted data of the normalisation channel $B^0 \to J/\psi\,K^{*0}$ for the 2012 Down samples. The distributions are similar, which implies that the distributions of input variables as well as their correlations are similar. This justifies a posteriori the use of the sWeighted normalisation channel sample as signal proxy in the training of the BDT against combinatoric background.

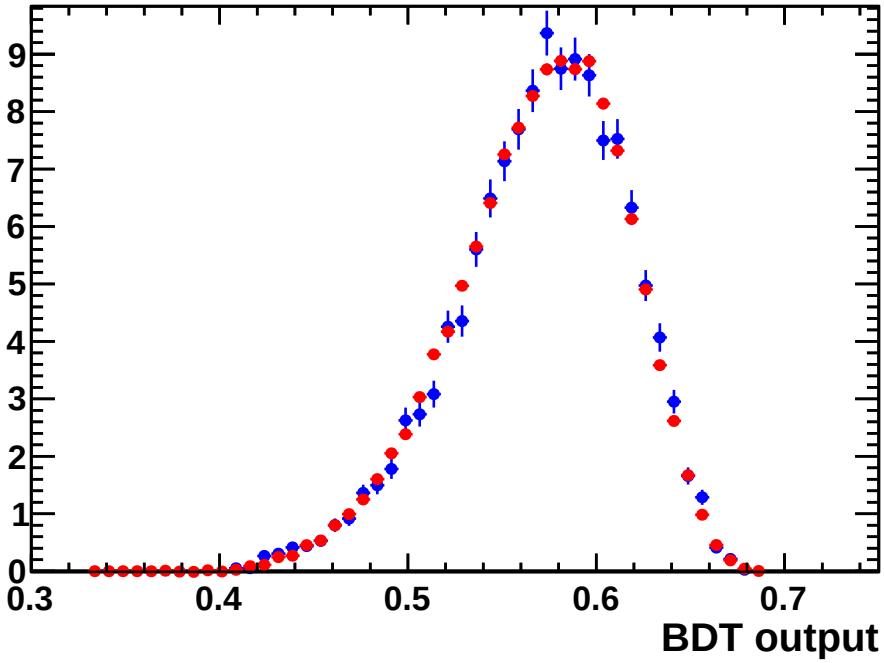

Figure 27: Output of BDT against combinatoric background. Simulated $B^+ \to J/\psi\,\rho^+$ events with 2012 Up settings in blue, $B^0 \to J/\psi\,K^{*0}$ sWeighted data from 2012 Down in red. The two distributions are similar.

A.3 Efficiency Tables

In the following the efficiency tables for the configurations 12 Magnet Up (Table 11), 11 Magnet Up (Table 12), and 11 Magnet Down (Table 13) are given. A description of the entries and the table for 2012 Magnet Down, Table 8, can be found in Sec. 6.5.

A.4 Detailed Fit Results

The fit to the 2011 data sample as introduced in Sec. 6.7 is shown in Fig. 28.

Fits to 2011 and 2012 data samples as introduced in Sec. 6.7, but with floating mean and width of the CB describing background from $B^+ \to J/\psi\,K^{*+}$ are shown in Fig. 29. This was done to access the systematic uncertainty from the choice of shape of this background component as explained in Sec. 7.

Selection	$B^+ \to J/\psi\,\rho^+$ Candidates	ϵ_{cut}	$B^0 \to J/\psi\,K^{*0}$ Candidates	ϵ_{cut}
Events in Acceptance	1 038 326	14.9 %	4 425 822	16.1 %
Stripping Selection	22 827	2.2 %	414 671	9.4 %
Trigger Cuts	17 611	77.1 %	331 199	79.9 %
Offline Cuts	12 982	73.7 %	270 565	81.7 %
PID Cuts	9271	71.4 %	226 066	83.6 %
Weighted Events	7936			
BDTc & BDTp Cut	3855	48.6 %	—	—
wrt. Stripping	—	19.7 %	—	54.5 %
Total Efficiency		0.434 %		5.11 %

Table 11: Number of remaining events and resulting efficiency for the simulated signal channel sample and simulated normalization channel sample after each selection step. Both for 2012 Up. Efficiencies are relative to number of events passing all prior cuts. PID efficiency was taken from `PIDCalib` (see Sec. 5.1), while BDT efficiencies were taken from weighted MC simulation.

Selection	$B^+ \to J/\psi\,\rho^+$ Candidates	ϵ_{cut}	$B^0 \to J/\psi\,K^{*0}$ Candidates	ϵ_{cut}
Events in Acceptance	509 842	14.6 %	3 076 566	15.8 %
Stripping Selection	12 608	2.5 %	318 849	10.4 %
Trigger Cuts	9 938	78.8 %	259 234	81.3 %
Offline Cuts	7 406	74.5 %	211 106	81.4 %
PID Cuts	5251	70.9 %	176 207	83.5 %
Weighted Events	4602			
BDTc & BDTp Cut	2291	49.8 %	—	—
wrt. Stripping	—	20.7 %	—	55.3 %
Total Efficiency		0.513 %		5.72 %

Table 12: Number of remaining events and resulting efficiency for the simulated signal channel sample and simulated normalization channel sample after each selection step. Both for 2011 Up. Efficiencies are relative to number of events passing all prior cuts. PID efficiency was taken from `PIDCalib` (see Sec. 5.1), while BDT efficiencies were taken from weighted MC simulation.

Selection	$B^+ \to J/\psi\,\rho^+$ Candidates	ϵ_{cut}	$B^0 \to J/\psi\,K^{*0}$ Candidates	ϵ_{cut}
Events in Acceptance	551 939	14.6 %	3 081 998	15.8 %
Stripping Selection	13 509	2.5 %	319 121	10.4 %
Trigger Cuts	10 598	78.5 %	260 527	81.6 %
Offline Cuts	7 727	72.9 %	212 420	81.5 %
PID Cuts	5249	71.2 %	177 137	83.4 %
Weighted Events	4733			
BDTc & BDTp Cut	2311	48.8 %	—	—
wrt. Stripping	—	19.0 %	—	55.5 %
Total Efficiency		0.464 %		5.75 %

Table 13: Number of remaining events and resulting efficiency for the simulated signal channel sample and simulated normalization channel sample after each selection step. Both for 2011 Down. Efficiencies are relative to number of events passing all prior cuts. PID efficiency was taken from `PIDCalib` (see Sec. 5.1), while BDT efficiencies were taken from weighted MC simulation.

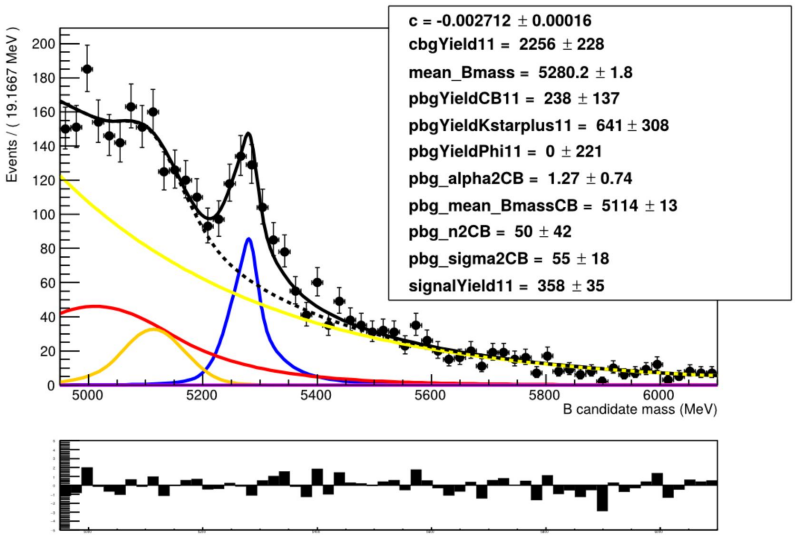

Figure 28: Fit of B-candidate mass after selection. Data sample from 2011 Magnet polarity Up and Down combined. $B^+ \to J/\psi\,\rho^+$ signal described by double CB pdf in blue, combinatoric background described by exponential distribution in yellow, partially reconstructed background from $B_s^0 \to J/\psi\,\phi$ described by double CB pdf in violet, partially reconstructed background from $B^+ \to J/\psi\,K^{*+}$ described by CB pdf in red, additional partially reconstructed background described by CB pdf in orange.

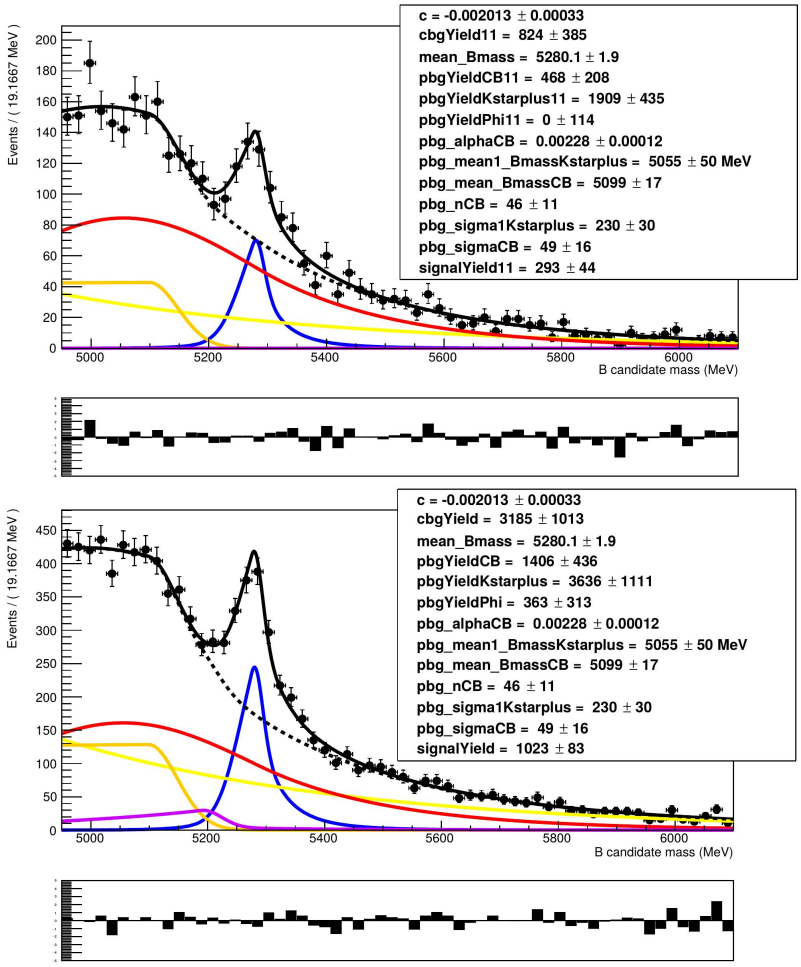

Figure 29: Fit of B-candidate mass after selection. $B^+ \to J/\psi\,\rho^+$ signal described by double CB pdf in blue, combinatoric background described by exponential distribution in yellow, partially reconstructed background from $B_s^0 \to J/\psi\,\phi$ described by double CB pdf in violet, partially reconstructed background from $B^+ \to J/\psi\,K^{*+}$ described by CB pdf in red, additional partially reconstructed background described by CB pdf in orange. In contrast to the real fit model, mean and width of the CB pdf describing the background component from $B^+ \to J/\psi\,K^{*+}$ were left to float (pbg_mean1_BmassKstarplus and pbg_sigma1Kstarplus). Top: Data sample from 2011 Magnet polarity Up and Down combined. Bottom: Data sample from 2012 Magnet polarity Up and Down combined. Shape and relative contributions of the floating background components have changed significantly between the nominal fit with fixed shape of the CB describing background from $B^+ \to J/\psi\,K^{*+}$ and this one (compare Fig. 28 and 26).

References

[1] BaBar, B. Aubert *et al.*, *Branching fraction and charge asymmetry measurements in $B \to J/\psi\pi\pi$ decays*, Phys. Rev. **D76** (2007) 031101, arXiv:0704.1266.

[2] I. Dunietz and J. M. Soares, *"Direct CP violation in $b \to dJ/\psi$ decays"*, Phys. Rev. **D49** (1994) 5904, arXiv:hep-ph/9312233.

[3] M. Thomson, *Modern particle physics*, Cambridge University Press, New York, 2013.

[4] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*, 1995.

[5] Dsperlich, *Standard Model of Elementary Particles*, https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg, last checked 05-18-2016.

[6] LHCb, R. Aaij *et al.*, *Observation of J/ÏĹp Resonances Consistent with Pentaquark States in $\hat{I}\dot{Z}_b^0$ âĘŠ J/ÏĹK$^-$p Decays*, Phys. Rev. Lett. **115** (2015) 072001, arXiv:1507.03414.

[7] A. D. Sakharov, *Violation of cp invariance, c asymmetry, and baryon asymmetry of the universe*, Soviet Physics Uspekhi **34** (1991), no. 5 392.

[8] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[9] ALICE, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[10] ATLAS, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[11] CMS, S. Chatrchyan *et al.*, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.

[12] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

[13] R. Lindner, *LHCb layout _2. LHCb schema_2*, LHCb Collection., http://cds.cern.ch/record/1087860, Feb, 2008.

[14] "LHCb collaboration", *LHCb detector performance*, International Journal of Modern Physics A **30** (2015), no. 07 1530022, arXiv:http://www.worldscientific.com/doi/pdf/10.1142/S0217751X15300227.

[15] LHCb collaboration, *TT detector*, http://lhcb.physik.uzh.ch/ST/public/material/index.php, last checked 04-29-2016.

[16] O. Callot and S. Hansmann-Menzemer, *The Forward Tracking: Algorithm and Performance Studies*, Tech. Rep. LHCb-2007-015. CERN-LHCb-2007-015, CERN, Geneva, May, 2007.

[17] R. E. Kalman, *A new approach to linear filtering and prediction problems*, Transactions of the ASME–Journal of Basic Engineering **82** (1960), no. Series D 35.

[18] LHCb RICH Group, M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, `arXiv:1211.6759`.

[19] O. Deschamps *et al.*, *Photon and neutral pion reconstruction*, Tech. Rep. LHCb-2003-091, CERN, Geneva, Sep, 2003.

[20] LHCb collaboration, *The LHCb reconstruction program*, http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/brunel/, last checked 04-29-2016.

[21] "LHCb collaboration", *The LHCb analysis program*, http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci/, last checked 04-29-2016.

[22] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, J. Phys. Conf. Ser. **331** (2011) 032047.

[23] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

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

[25] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, `arXiv:physics/0402083`.

[26] Particle Data Group, K. A. Olive *et al.*, *Review of particle physics*, Chin. Phys. **C38** (2014) 090001, and 2015 update.

[27] NIST/SEMATECH, *e-Handbook of Statistical Methods*, `http://www.itl.nist.gov/div898/handbook/`, 2015.

[28] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, ArXiv Physics e-prints (2003) `arXiv:physics/0306116`.

[29] F. James and M. Winkler, *MINUIT User's Guide*, `www.cern.ch/minuit`, 2004.

[30] Particle Data Group, K. A. Olive *et al.*, *Review of Particle Physics*, Chin. Phys. **C38** (2014) 090001.

[31] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nucl. Instrum. Meth. **A552** (2005) 566, arXiv:physics/0503191.

[32] A. Sarti, S. Furcas, G. Lanfranchi, and M. Palutan, *Calibration Strategy and Efficiency measurement of the Muon Identification procedure at LHCb*, .

[33] A. Guenther, *Measurement of the branching fraction of the rare decay $B^+ \to K^{*+}\mu^+\mu^-$ âĹŠ exploiting the $K^{*+} \to K^+\pi^0$ decay at the lhcb experiment*, Bachelor's thesis, Heidelberg University, 2015.

[34] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.

[35] M. Bishai *et al.*, *Study of b âĘŠ ĬĹĬś*, Physics Letters B **369** (1996), no. 2 186 .

[36] R. Aaij and J. Albrecht, *Muon triggers in the High Level Trigger of LHCb*, Tech. Rep. LHCb-PUB-2011-017. CERN-LHCb-PUB-2011-017, CERN, Geneva, Sep, 2011.

[37] V. Gligorov, C. Thomas, and M. Williams, *The HLT inclusive B triggers*, .

[38] M. Williams *et al.*, *The HLT2 Topological Lines*, .

[39] LHCb, R. Aaij *et al.*, *Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity*, JHEP **02** (2016) 104, arXiv:1512.04442.

[40] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, Wadsworth international group, Belmont, California, USA, 1984.

[41] B. P. Roe *et al.*, *Boosted decision trees as an alternative to artificial neural networks for particle identification*, Nucl. Instrum. Meth. **A543** (2005) 577, arXiv:physics/0408124.

[42] A. Hoecker *et al.*, *TMVA: Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040, arXiv:physics/0703039.

[43] R. E. Schapire and Y. Freund, *A decision-theoretic generalization of on-line learning and an application to boosting*, J. Comput. Syst. Sci. **55** (1997) 119.

[44] A. D. Bukin, *Fitting function for asymmetric peaks*, ArXiv e-prints (2007) arXiv:0711.4449.

[45] LHCb collaboration, R. Aaij *et al.*, *Branching fraction and CP asymmetry of the decays $B^+ \to K^0_S\pi^+$ and $B^+ \to K^0_S K^+$*, Phys. Lett. B **726** (2013) 646, Comments: 13 pages, 2 figures.

[46] LHCb collaboration, R. Aaij *et al.*, *Measurement of the $B^0_s \to J/\psi K^0_S$ branching fraction*, Phys. Lett. B **713** (2012) 172.

[47] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, J. Instrum. **10** (2014) P02007. 20 p.

[48] A. Hicheur, *Review of recent LHCb results and prospects for Run II*, in *LISHEP-International School on High Energy Physics- Session C MANAUS, Amazonas, Brazil, August 2-9, 2015*, 2015. arXiv:1509.07708.

Acknowledgements

I would like to thank Prof. Uwer for his helpful advice and the opportunity to write my thesis in his working group in the first place. My thanks go to Prof. Reygers as well for agreeing to be my second referee.

I'd also like to thank Dr. Sevda Esen and Dr. Michel De Cian, who guided me out of many problems I ran into during my thesis and always found time to "look at yet another bunch of plots" with me.

Thanks go to all other members of the LHCb group at Heidelberg University as well, especially my office mate Thomas Nikodem, whose expertise I could confide in on many occasions.

Zu guter Letzt geht ein großer Dank an meine Familie und meine Freunde für ihre durchgehende Unterstützung, nicht nur während dieser Arbeit.