Department of Physics and Astronomy
University of Heidelberg

Bachelor Thesis in Physics
submitted by

Paul André Günther

born in Kassel (Germany)

2015

Measurement of the branching fraction of the rare decay $B^+ \rightarrow K^{*+}\mu^+\mu^-$ exploiting the $K^{*+} \rightarrow K^+\pi^0$ decay at the LHCb experiment

This Bachelor Thesis has been carried out by Paul André Günther at the
Physikalische Institut in Heidelberg
under the supervision of
Prof. Stephanie Hansmann-Menzemer

Abstract

In this thesis, a measurement of the branching fraction of the rare B meson decay $B^+ \to K^{*+}\mu^+\mu^-$ relative to the resonant decay $B^+ \to J/\psi K^{*+}$ is presented, whereby the K^{*+} decay mode $K^{*+} \to K^+\pi^0$ is chosen exclusively. The used data was collected by the LHCb experiment during the years 2011 and 2012 with centre-of-mass energies of $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV, respectively. In total, 81 ± 16 $B^+ \to K^{*+}(\to K^+\pi^0)\mu^+\mu^-$ decays with a statistical significance of 6σ are reconstructed. The relative branching fraction is found to be

$$\frac{\mathcal{B}(B^+ \to K^{*+}\mu^+\mu^-)}{\mathcal{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^{*+})} = (1.03 \pm 0.20_{stat.} \pm 0.04_{syst.}) \times 10^{-2}$$

where the uncertainties are statistical and systematic, respectively. The total branching fraction of $B^+ \to K^{*+}\mu^+\mu^-$ is calculated to be

$$\mathcal{B}(B^+ \to K^{*+}\mu^+\mu^-) = (0.88 \pm 0.17_{stat.} \pm 0.06_{syst.}) \times 10^{-6}.$$

Kurzfassung

In dieser Arbeit wird das Verzweigungsverhältnis des seltenen B-Mesonzerfalls $B^+ \to K^{*+}\mu^+\mu^-$ relativ zu dem resonanten Zerfall $B^+ \to J/\psi K^{*+}$ bestimmt, wobei ausschließlich $K^{*+} \to K^+\pi^0$ Zerfälle betrachtet werden. Die ausgewerteten Daten wurden in den Jahren 2011 und 2012 vom LHCb-Experiment bei Schwerpunktsenergien von $\sqrt{s} = 7$ TeV bzw. $\sqrt{s} = 8$ TeV aufgenommen. Es werden 81 ± 16 $B^+ \to K^{*+}\mu^+\mu^-$ Zerfälle mit einer statistischen Signifikanz von 6σ rekonstruiert. Das relative Verzweigungsverhältnis wird zu

$$\frac{\mathcal{B}(B^+ \to K^{*+}\mu^+\mu^-)}{\mathcal{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^{*+})} = (1.03 \pm 0.20_{stat.} \pm 0.04_{syst.}) \times 10^{-2}$$

bestimmt, wobei die Unsicherheiten statistisch bzw. systematisch sind. Das absolute Verzweigungsverhältnis von $B^+ \to K^{*+}\mu^+\mu^-$ kann dann zu

$$\mathcal{B}(B^+ \to K^{*+}\mu^+\mu^-) = (0.88 \pm 0.17_{stat.} \pm 0.06_{syst.}) \times 10^{-6}$$

berechnet werden.

Contents

1 Introduction

The domain of particle physics deals with the smallest components of our Universe. These components and the mechanisms holding them together, forming the stunning variety of Nature, are described by a single theory - the Standard Model of particle physics. The Standard Model was introduced in the 1960's and its predictions have been tested extensively ever since in order to either confirm them, or to disprove their validity. But no deviation has been observed by laboratory experiments so far. However, the Standard Model is only a theory of 'almost everything'. A multitude of questions remain unanswered, for example: What is dark matter? Why is there more matter than antimatter? Is there a link between the theory of gravity and the Standard Model? All these questions are addressed by the search for so-called New Physics. Yet unknown particles, and thus New Physics, are searched for in either direct or indirect measurements. The direct approach searches for new particles that are created directly in high-energy particle collisions. The LHCb experiment at CERN is devoted to the indirect searches in hadrons that accommodate a b quark. In the Standard Model of particle physics, the transition of a b quark to an s quark is forbidden on tree level. It only proceeds via higher order Feynman diagrams - loop or box diagrams - and is therefore rare on the one hand, but on the other hand it also offers a sensitive probe of physics that lies beyond the Standard Model, because new particles may hide in the loops. The decay[1] of a B meson into an excited kaon and a non-resonant muon pair, $B^+ \to K^{*+}(892)\mu^+\mu^-$, is induced by such a transition and is therefore analysed in this thesis with the objective to measure its branching fraction. While the branching fraction can in principle be used to distinguish between New Physics and the Standard Model, its Standard Model prediction, however, suffers from relatively large uncertainties. Hence, the determination of the branching fraction in this thesis is only the first step in exploiting measurable properties of the decay $B^+ \to K^{*+}(892)\mu^+\mu^-$ to test the Standard Model. From knowledge about the branching fractions, it is possible to construct new observables such as the isospin asymmetry, in which uncertainties of the predictions largely cancel. The excited kaon in this decay is of special interest because it enriches the decay topology by two more measurable angular distributions in comparison to a kaon in the ground state. If the statistics of the decay is sufficiently large to measure the branching fraction with satisfactory statistical precision, these angles may be studied in an angular analysis.

The K^* decays immediately into a kaon and a pion whereby the commonly used modes are $K^{*0} \to K^+\pi^-$ and $K^{*+} \to K_S^0(\to \pi^+\pi^-)\pi^+$, respectively. These modes are chosen in most analyses since their decay products are charged and it is therefore relatively simple to determine their momentum accurately. Though, a large amount of excited kaons decays also via emission of a neutral pion which then decays into two photons. The mode $K^{*+} \to K^+\pi^0(\to \gamma\gamma)$ has not been used so far by the LHCb experiment because the reconstruction of the π^0 is less efficient and less accurate than it is for its charged counterparts. In this thesis, the $K^{*+} \to K^+\pi^0(\to \gamma\gamma)$ mode is used exclusively in order to investigate the feasibility of reconstructing a number of events from the rare decay $B^+ \to K^{*+}\mu^+\mu^-$ that is sufficient to measure the branching fraction with acceptable precision. Future analyses may then exploit this mode as well for searches for New Physics. For this purpose, the 3 fb^{-1} dataset collected with the LHCb detector in the years 2011 and 2012 with centre-of-mass energy $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV, respectively, is used.

In Section 2 of this thesis, basic concepts of the Standard Model are highlighted with a focus on flavour physics. Section 3 describes experimental aspects of the search for physics that lies beyond the Standard Model and the decay $B^+ \to K^{*+}\mu^+\mu^-$ is discussed in this context.

In order to understand the procedure of the data analysis, it is vital to understand the detector system which is depicted in Section 4. The analysis strategy is outlined in Section 5.1, the dataset and important variables are introduced in Section 5.2 and the procedure of the signal selection using a multivariate

---

[1]Charge conjugation is implied throughout this thesis.

analysis technique (Section 5.3.5) is documented thoroughly in Section 5.3. Finally, from the results of the signal selection (Section 5.3.7), the branching fraction of $B^+ \to K^{*+} \mu^+ \mu^-$ is determined in Section 6. In Section 7, systematic uncertainties in the analysis procedure are studied before the whole analysis is reviewed and concluded in Section 8.

2 Theoretical background

This section shortly describes some basic theoretical concepts of particle physics. The Standard Model of particle physics (SM) is sketched-out with a focus on Flavour Physics and some of its unanswered questions are mentioned. The research on rare B meson decays is motivated in this context. For a detailed description of theoretical and experimental aspects of the SM, it is referred to [1] and [2].

2.1 The Standard Model of particle physics

Particle physics is devoted to the understanding of fundamental constituents of the Universe and the forces which interact between them. The SM is a Quantum Field Theory (QFT) that provides a unified picture where particles are described by fields and the forces are described by the exchange of particles. On the one hand, deviations from the SM are yet to be found in experimental searches, on the other hand the SM cannot account for all observed phenomena. Since the discovery of a Higgs boson candidate in 2012 and its later confirmation [3], the search for physics lying beyond the SM has outlined one of the main goals of particle physics.

2.1.1 Fundamental particles and forces

A principle idea of the SM is that every fermionic matter particle has an antimatter equivalent with the same mass but opposite electric charge. This pattern is first approached by Dirac, who proposed an electron with positive charge [4, 5] - today known as the positron.

All in all, the SM contains twelve fermions and twelve antifermions, five gauge bosons and the Higgs boson. Fermions are particles with half-integer spin that are therefore characterized by Fermi-Dirac statistics. There is a further distinction between quarks and leptons (both with spin $\frac{1}{2}$) - according to their dominant interaction. Both, quarks and leptons, occur in three generations with two particles each. The mass and electrical charge of the different fermions can be found in Table 1. Quarks carry the QCD (quantum chromodynamics) equivalent of electric charge, called colour charge, and interact dominantly via the strong force but also participate in weak and electromagnetic processes. Thus, quarks also carry weak and electric charge. A charge $+\frac{2}{3}$ and $-\frac{1}{3}$ in units of the electric charge of an electron defines the so called up-type quarks and down-type quarks, respectively[2]. Due to the hypothesis of colour confinement, quarks cannot exist individually, but hadronise on small timescales into colourless mesons (bound states of two quarks) and baryons (bound states of three quarks). This can be understood qualitatively by considering what happens when two free quarks are pulled apart. The interaction between the two quarks is mediated by the exchange of virtual[3] gluons. Because of the attractive interaction between these virtual gluons, the colour field between the quarks is squeezed, leading to a constant energy density between the quarks. Consequently, a separation of quarks results in increasing the field's energy enormously, following a potential of the form $V(\vec{r}) \propto \kappa r$, where r is distance between the quarks and κ an empirical constant[4]. It is therefore energetically more favourable for quarks to form colourless hadronic states, as well as it is for gluons to be confined in such objects, which explains the short range of the strong interaction.

Leptons interact via the weak force and in case of the electron, muon and tau particle, which are charged $\pm 1e$, also via the electromagnetic force. The neutrinos are colour neutral like the other leptons but carry no electric charge and thus interact via the weak interaction only.

The three forces of relevance to particle physics are mediated by gauge bosons (spin-1 particles, obeying Bose-Einstein statistics). The photon (γ) is the gauge boson of quantum electrodynamics (QED) and

---

[2]Charges are conjugated for the corresponding antiquarks.

[3]A virtual particle is a mathematical construct, it appears in neither the initial nor the final state of a particle interaction and is not bound to the Einstein energy-momentum relationship.

[4]experimentally $\kappa \sim 1$ GeV/fm [1]

	Quarks		
	Particle (flavour)	Q[e]	Mass [MeV/c^2]
first generation	up (u)	+2/3	$2.3^{+0.7}_{-0.5}$
	down (d)	−1/3	$4.8^{+0.5}_{-0.3}$
second generation	charm (c)	+2/3	1275 ± 25
	strange (s)	−1/3	95 ± 5
third generation	top (t)	+2/3	$(173.21 \pm 0.51 \pm 0.71) \times 10^3$
	bottom (b)	−1/3	$(4.18 \pm 0.03) \times 10^3$

(a) List of quarks

	Leptons		
	Particle	Q[e]	Mass [MeV/c^2]
first generation	electron (e^-)	−1	0.511
	electron neutrino (ν_e)	0	$< 2 \times 10^{-6}$
second generation	muon (μ^-)	−1	105.66
	muon neutrino (ν_μ)	0	< 0.19
third generation	tau (τ^-)	−1	1776.82 ± 0.16
	tau neutrino (ν_τ)	0	< 18.2

(b)List of leptons

Table 1: The twelve fundamental fermions divided into quarks (a) and leptons (b). The u, d and s quark masses are the "current-quark masses" at an energy scale $\mu \approx 2$ GeV. The c and b quark masses are the "running" masses. The t quark mass is taken from the direct measurements. The electron mass and the muon mass are rounded. All values are taken from [6].

mediates the electromagnetic force between electrically charged particles. It is massless and electrically neutral, which is why the electromagnetic force acts on an infinite range.

The force-carrying particles of the strong interaction are the gluons (g). They form an octet of coloured states, always carrying a combination of colour and anticolour, e.g. red-antigreen, blue-antired or a superposition of different colour-anticolour states. Hence, they are able to self-interact, which is the reason for colour confinement. Apart from that, gluons are massless and electrically neutral like photons. The gauge bosons of the weak charged-current interactions are the massive W^+ and W^- bosons. They couple fermions that differ in one unit of electric charge. The massive Z^0 boson mediates the weak neutral-current interaction. Since these bosons carry also weak charge and in case of the W's also electric charge, they couple to each other.

The last piece of the puzzle is the Higgs boson (H^0), which is a massive spin-0 scalar particle. It corresponds to the Higgs field that gives the other fundamental particles their mass. The properties of the bosons are summarised in Table 2.

	Bosons		
Boson	Charge [e]	Mass [GeV/c^2]	Coupling
Photon γ	$< 10^{-35}$	0	electromagnetic
Gluon g	0	0	strong
W	± 1	80.385 ± 0.015	weak, electromagnetic
Z	0	91.1876 ± 0.0021	weak
H^0	0	125.7 ± 0.4	mass

Table 2: The fundamental gauge bosons and the Higgs boson. All values are taken from [6].

Although the electromagnetic and weak interaction is treated separately here, they are unified by Salam and Weinberg in an electroweak theory. The feasibility of an unification of QCD and electroweak theory

in a grand unified theory (GUT) remains unclear [7].

2.1.2 Flavour physics

Flavour physics concentrates on the dynamics of the different quark and lepton flavours. The weak force can alter the flavour quantum number and accommodates CP violation[5]. A famous example of a flavour changing charged current is the nuclear β^--decay, where a down-quark decays into an up-quark by coupling to a W^- boson which then decays into an electron and a neutrino. However, it is found experimentally that the transitions between up-quarks and strange-quarks are weaker than expected for a universal weak coupling. This difference in the coupling strength motivates the idea that weak eigenstates of quarks differ from their mass eigenstates. The mixing of quark flavour is described by the unitary Cabibbo-Kobayashi-Maskawa (CKM) matrix. The weak eigenstates are denoted q' and the mass eigenstates q, where q is the replacement character for quark flavour. In a natural parameterisation, the weak eigenstates of down-type quarks are related to the mass eigenstates by

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

The choice of down-type quarks in this parametrisation is arbitrary. Since the CKM matrix is unitary, a definition with up-type quarks does not influence the physical essence.

The probability of a decay of a flavour eigenstate i into eigenstate j via the coupling to a W boson is proportional to $|V_{ij}|^2$, where i is always an up-type quark flavour and j a down-type quark flavour, due to the conservation of electrical charge in the SM. Transitions among up/down-type quarks are forbidden on tree level but possible though, when higher ordered terms are considered. These flavour changing neutral currents (FCNC) are predicted by the electroweak unification and appear by considering quantum loops in the Feynman diagrams, then called penguin diagram or box diagram, respectively (cf. Section 3).

The off-diagonal terms in the CKM matrix are relatively small compared to the on-diagonal ones. Consequently, the interaction of quarks of different generations is suppressed, leading to a near diagonal form of the CKM matrix. The Wolfenstein parameterisation uses this 'near-diagonality' to expand the matrix in a small parameter, λ. The CKM matrix is then written in terms of four real parameters, λ, A, ρ and η, to $\mathcal{O}(\lambda^4)$ as

$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4).$$

In this parameterisation, V_{ub} and V_{td} reside complex components which is necessary for CP to be violated in the quark sector. Hence, η must be non-zero. The CKM matrix elements are determined experimentally. A global fit that uses all available measurements gives the following Wolfenstein parameters [6]:

$$\lambda = 0.22537 \pm 0.00061, \qquad A = 0.814^{+0.023}_{-0.024}$$

$$\bar{\rho} = \rho\left(1 - \frac{\lambda^2}{2}\right) = 0.117 \pm 0.021, \qquad \bar{\eta} = \eta\left(1 - \frac{\lambda^2}{2}\right) = 0.353 \pm 0.013$$

---

[5]CP-symmetry violation, CP-symmetry is the combination of charge conjugation(C) and parity(P) symmetry. A CP transformation exchanges a particle with its antiparticle.

2.1.3 Open questions

The SM is a very successful model of Nature on small distance scales. However, it is mathematically complicated and arbitrary. It depends on a large free parameter set[6], does not include gravity and does not account for the patterns of fermion masses. Cosmological and astrophysical measurements provide evidence for the existence of dark matter that makes up 23% of the energy-matter density of the Universe and is not enclosed by the SM, while describable baryonic matter forms only 5% of the matter in the Universe. The majority of the energy-matter density is the dark energy (72%). Possible candidates for dark matter particles, e.g. particles from Supersymmetry (SUSY), have yet eluded detection.

The running of coupling constants could lead to a convergence on high energy scales suggesting the existence of a yet unknown unified theory of the three forces.

Furthermore, the observed CP violation in the SM seems to be insufficient to account for the observed matter-antimatter asymmetry.

A lot of theories for non-SM physics are developed, for example SUSY, large-scale extra dimensions and string theory. SUSY predicts the existence of heavy partners for each particle, that could be measured directly or through loop corrections in the next years at particle colliders such as the Large Hadron Collider (LHC) (see Section 3 and 4.1).

---

[6]For example the four parameters of the CKM matrix, the fermion masses and the coupling strengths of the three forces.

3 Experimental searches for New Physics

In order to test for contributions of phenomena that occur beyond the SM, two complementary approaches are common in particle physics; direct and indirect measurements. The direct search involves the direct creation of new particles in high energetic particle collisions and reconstructing their decay products. This approach is mainly limited by the centre-of-mass energy provided by the accelerator, and is therefore currently constricted to searches for new particles on the TeV-scale. The alternative indirect measurement of new physics offers the possibility of detecting heavy new particles at significantly higher scales. They are not created as 'real' particles, but as 'virtual' ones, existing in loops only and possibly affecting measurable quantities, such as amplitudes, angles and branching fractions.

3.1 B meson physics

B meson physics is an excellent place to study possible effects of New Physics. It involves the study of CP violation (CPV) in neutral B meson mixing[7] and CPV in the interference between decays to the same final state. Furthermore, the precise determination of observables in B decays is a meaningful test of the SM because the B meson has many loop-induced decays that have precisely predicted quantities. These quantities are more precise than the ones in other decays because the light quark in the B is negligible in comparison to the heavy b quark, which simplifies the corresponding calculations. Since B mesons contain the heavy b quark, they decay into many different final states with lighter quarks. The variety of these decays provides a solid basis for a multitude of analyses. The signature of B decays in the experiment is often clean because B mesons have a relatively long lifetime and it is therefore possible to distinguish their decays from primary collisions. At high energies, the production cross-section of $b\bar{b}$ quark pairs is large which leads to statistically powerful datasets.

Decays via FCNC (see Figure 1 left) are of special interest because they are suppressed by 2 - 3 orders of magnitude in comparison to the tree-level processes that give the same final state. New physics may enter in the FCNC loops and affect observables of these decays (see Figure 1 right).
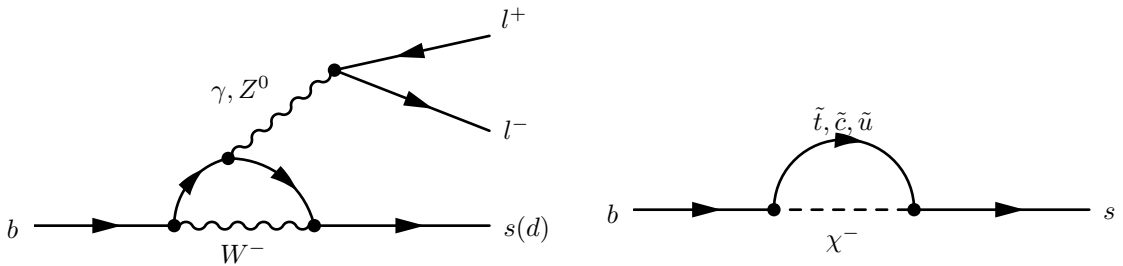

Figure 1: Feynman diagram showing the FCNC transition $b \rightarrow q l^+ l^-$ (left). Example of possible new physics loop contribution to $b \rightarrow s l^+ l^-$ with squarks ($\tilde{t}, \tilde{c}, \tilde{u}$) and chargino ($\chi^-$) (right), which are introduced by SUSY models.

---

[7]Oscillations of the $B^0(\bar{b}d) \leftrightarrow \bar{B}^0(b\bar{d})$ system, for example.

3.1.1 $B \to K^* \mu^+ \mu^-$

A well suited class of decays to test the SM are FCNC of B mesons. Figure 2 shows the FCNC penguin and box diagrams of the rare decay $B^+ \to K^{(*)+}\mu^+\mu^-$, which will be studied in this thesis. The dimuon (or dilepton in general) final state is experimentally advantageous because it gives a clean detector signature. In addition, the lepton pair can only be produced by a FCNC transition in the non-resonant case. The excited kaon is interesting in these decays because it allows to conduct an angular analysis with more angles due to its non-zero angular momentum quantum number. The measurable angles provide a set of observables which may be sensitive to non-SM physics.
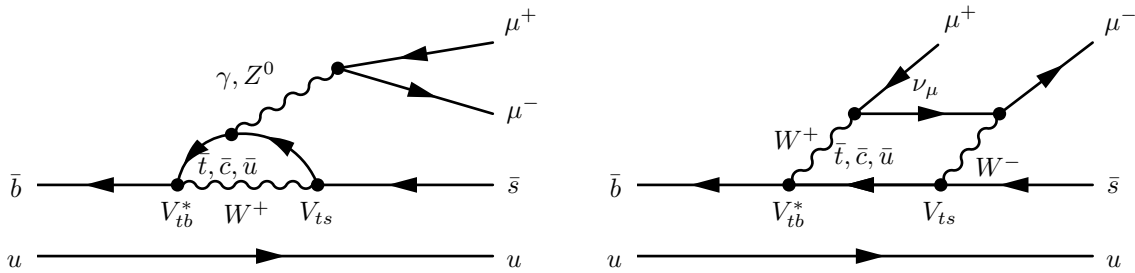
Figure 2: Lowest-order Feynman diagrams for the decay $B^+ \to K^{(*)+}\mu^+\mu^-$. Electroweak photon and Z penguin diagrams (left) and W^+W^- box diagram (right) are shown.

The amplitudes of the diagrams shown in Figure 2 are proportional to the CKM matrix elements of the different vertices, to the coupling constants and to $\frac{m_q}{m_W}$, where m_q is the mass of the virtual quark and m_W the mass of the W boson. Since the top quark is much heavier than the other flavours (see Table 1), it dominates the transition amplitude. For physics beyond the SM, new penguin and box diagrams can contribute with heavy[8] particles inside the loops (see Figure 1 right).

The decay $B^+ \to K^{*+}\mu^+\mu^-$ was observed for the first time at the B-factory Belle [8] and has been studied previously and subsequently by BaBar [9, 10]. The analysis performed here deals with the decay $B^+ \to K^{*+}\mu^+\mu^-$ where $K^{*+} \to K^+\pi^0$. This mode has not been studied at the LHCb experiment so far. As a first step towards an angular analysis of $B^+ \to K^{*+}\mu^+\mu^-$ and to assess its feasibility considering this mode, the branching fraction is determined relative to the dominating resonant tree level decay $B^+ \to K^{*+}J/\psi(\to \mu\mu)$ (see Figure 3). This normalisation is advantageous since most systematics cancel in the ratio. The world averages of the branching fractions of the signal and normalisation modes can be found in Table 3, together with other branching fractions that are important for this thesis. The results of an LHCb measurement of $B^+ \to K^{*+}\mu^+\mu^-$ where $K^{*+} \to K_S^0(\to \pi^+\pi^-)\pi^+$ are listed in Table 4. A theoretical prediction for the branching fraction of the decay can be found in Table 5.

---

[8]The possible new particles are assumed to be heavy because otherwise, they probably would have already been found.
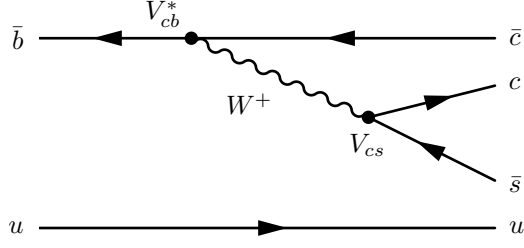
Figure 3: Lowest-order Feynman diagram for the decay $B^+ \to J/\psi(1S)K^{*+}$. The J/ψ can decay further via $J/\psi \to \mu\mu$ into the same final state as the processes in Fig. 2 and is therefore a resonant version of this decay.

Decay channel	Branching fraction
$B^+ \to K^{*+}l^+l^-$	$(1.29 \pm 0.21) \times 10^{-6}$
$\to K^{*+}\mu^+\mu^-$	$(1.12 \pm 0.15) \times 10^{-6}$
$B^+ \to J/\psi(1S)K^{*+}$	$(1.44 \pm 0.08) \times 10^{-3}$
$B^+ \to \psi(2S)K^{*+}$	$(6.7 \pm 1.4) \times 10^{-4}$
$K^{*+} \to K^+\pi^0$	$\sim 1/3$
$\pi^0 \to \gamma\gamma$	$(98.823 \pm 0.034)\%$
$J/\psi(1S) \to \mu^+\mu^-$	$(5.961 \pm 0.033)\%$
$\psi(2S) \to \mu^+\mu^-$	$(7.9 \pm 0.9) \times 10^{-3}$

Table 3: World averages of the branching fractions for B^+ and daughter particle decays used in this analysis. The values are taken from [6].

Decay channel	Branching fraction	Integrated luminosity
$B^+ \to K^{*+}\mu^+\mu^-$	$(1.16 \pm 0.19) \times 10^{-6}$	$1~\text{fb}^{-1}$
$B^+ \to K^{*+}\mu^+\mu^-$	$(0.924 \pm 0.093_{stat.} \pm 0.067_{syst.}) \times 10^{-6}$	$3~\text{fb}^{-1}$

Table 4: Results of the LHCb measurements for $B^+ \to K^{*+}\mu^+\mu^-$ where $K^{*+} \to K_S^0(\to \pi^+\pi^-)\pi^+$. In these analyses the differential branching fraction is measured; the total branching fractions is obtained by integration over the q^2-range. [11, 12]

Model	Branching fraction prediction	Reference
NNLO[9] QCD	$(1.19 \pm 0.39) \times 10^{-6}$	[13]

Table 5: SM predictions for the branching fraction of $B^+ \to K^{*+}\mu^+\mu^-$.

---

[9]next-to-next-to-leading order

4 The LHCb experiment

This section describes the apparatus of the Large Hadron Collider beauty (LHCb) experiment at the Conseil Européen pour la Recherche Nucléaire (CERN), which recorded the data analysed in this thesis. It is one of seven experiments at the Large Hadron Collider (LHC). The purpose of LHCb is to look for indirect evidence of new physics in decays of beauty and charm hadrons [14].

After a brief overview of the accelerator and the LHCb detector, the main sub-detectors are discussed. For more details, see for example [14–19].

4.1 The LHC

The LHC at CERN near Geneva close to the France-Switzerland border is a particle-particle collider and accelerator composed of two rings with superconducting magnets and counter-rotating beams. The rings are approximately circular and of 26.7 km circumference, lying in a depth between 45 m and 170 m under the ground of France and Switzerland. There are eight interaction points in which particles can collide, four of them correspond to the positions of the four large experiments - ATLAS, CMS, ALICE and LHCb. Aside from these experiments, there are three smaller experiments - TOTEM (near CMS), LHCf (near ATLAS) and MoEDAL (near LHCb). ATLAS and CMS use general-purpose detectors designed independently to investigate a large range of physics at high energies. They are - among other duties - devoted to direct searches for New Physics. ALICE focuses on heavy-ion physics, while LHCb is a dedicated B physics experiment [17, 18].

The LHC is designed for maximum centre-of-mass energies of up to $\sqrt{s} = 14$ TeV at a maximum luminosity of $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1} at ATLAS and CMS for proton-proton[10] (pp) collisions [18]. During Run I of operation of the LHC in 2011 and 2012, the maximum centre-of-mass energy was $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV, respectively. Run II started in June 2015 with $\sqrt{s} = 13$ TeV.
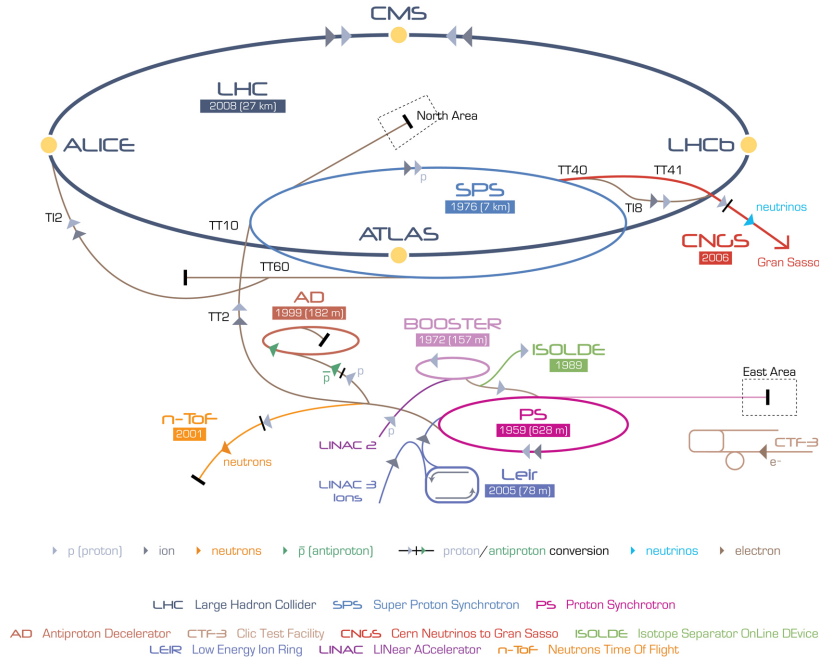

Figure 4: *Accelerator complex at CERN [17].*

---

[10]For heavy-ion physics at ALICE, also collisions of lead ions are possible, with much less luminosity though.

The accelerator complex is shown in Figure 4. Protons are taken from ionised hydrogen and are injected in bunches into the initial linear accelerator (LINAC2), which accelerates them to a kinetic energy of 50 MeV before bringing them in the Proton Synchrotron Booster (BOOSTER). The BOOSTER brings the protons up to an energy of 1.4 GeV and transfers them into the Proton Synchrotron (PS) that accelerates them to 25 GeV. From the PS the protons are injected into the Super Proton Synchrotron (SPS) that speeds them up to 450 GeV before inserting them in the LHC, where the protons are finally accelerated to multi-TeV energies (6.5 TeV in 2015).

4.2   The LHCb detector
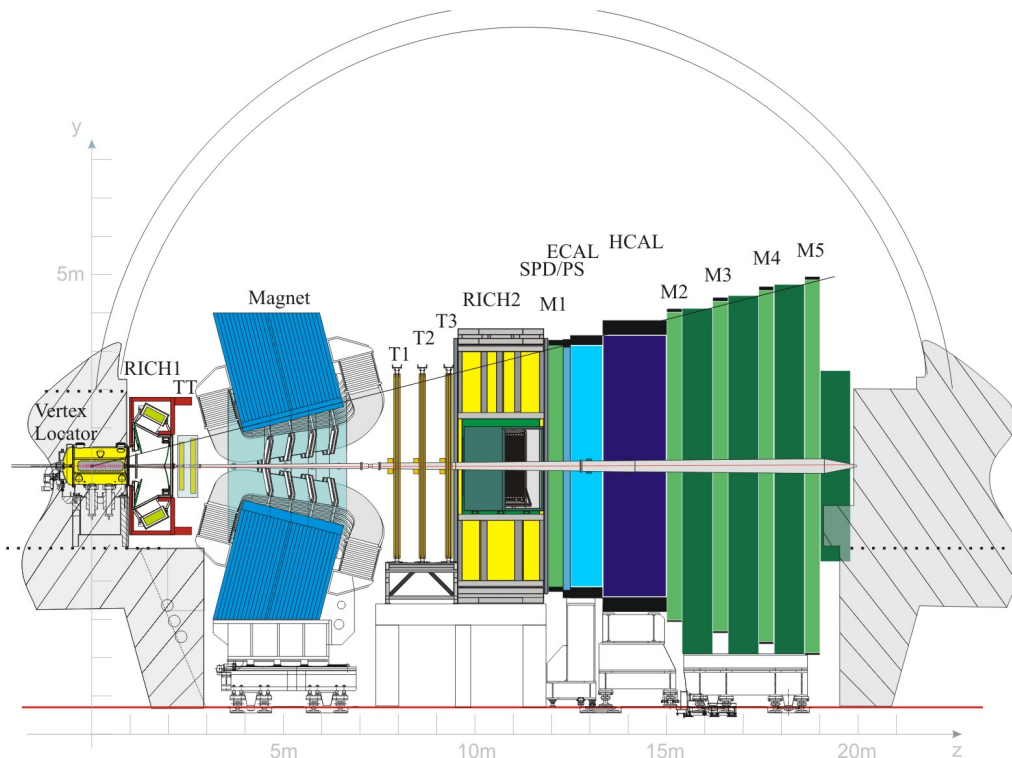
Figure 5: Layout of the LHCb spectrometer shown from the side. The Vertex Locator (VELO) is located around the collision point ($z = 0$). RICH1, RICH2 are ring imaging Cherenkov detectors. TT and T1-3 are the tracking stations. M1-5 are the muon stations and SPD/PS, ECAL, HCAL are parts of the calorimeter systems [14].

4.2.1   LHCb overview

LHCb is a single-arm spectrometer covering approximately 10 mrad to 300 (250)[11] mrad of the angle in forward direction in the bending (non-bending) plane (see Figure 5). This asymmetric orientation is favourable because the production of $b\bar{b}$ quark pairs in the pp collisions is forward-biased due to the different momentum of the partons[12] in the protons resulting in a high boost of the b quarks along the beam axis at energies that are large in comparison to the b quark mass.

In order to conduct precise measurements of B meson decay times, it is vital to reconstruct primary and secondary vertices of the processes accurately and in particular to distinguish between particles from pp interactions and those from B meson decays. At the LHC, the beauty and charm production

---

[11]The detector covers $1.6 < \eta < 4.9$ [14], where $\eta \equiv -\ln(\tan(\frac{\theta}{2}))$ is the pseudorapidity; θ is the angle between a particle's momentum and the beam axis.

[12]Parton refers to the parton model and means constituents of hadrons, i.e. quarks and gluons.

cross-section is large[13] and in 2011 and 2012 approximately 10^{12} heavy flavour decays were collected. However, the c and b cross-sections are 10 and 200 times smaller than the total inelastic pp cross-section[14], respectively [19]. Therefore, a very efficient tracking and thus precise momentum determination is required and provided by the interplay of the different detector components. The functionality in a nutshell: a $b\bar{b}$ pair is created in a pp collision among other hadrons and leptons and hadronises into a B meson. The B meson decays within the Vertex Locator and the daughter particles pass the first Cherenkov detector that detects characteristic Cherenkov radiation. Entering the magnet, the charged particles are deflected differently (in the x-z plane) depending on their momentum before leaving tracks and hits in the tracking stations and calorimeters. With these tracks the whole trajectory of a particle through the detector is reconstructed and matched to the corresponding vertices. The information on the momentum of the particles is then used - together with the information from the Cherenkov detector - to provide a mass hypothesis.

LHCb uses a right-handed coordinate system with z defined along the beam axis into the detector (downstream), y vertical and x horizontal - pointing towards the centre of the accelerator ring (see Figure 5).

4.2.2 Magnet

LHCb's dipole magnet consists of two saddle-shaped aluminium coils in a window-frame steel yoke having a total weight of ~ 1600 tons, see Figure 6. The magnet poles are tilted towards the interaction point, following the acceptance of LHCb (cf. Figure 5). It is a water cooled warm magnet[15] with a field integral along the z-axis (10m) of $\int B \mathrm{d}l \approx 4$ Tm [14]. The deflection of charged particles takes place mostly in the x-z-plane.
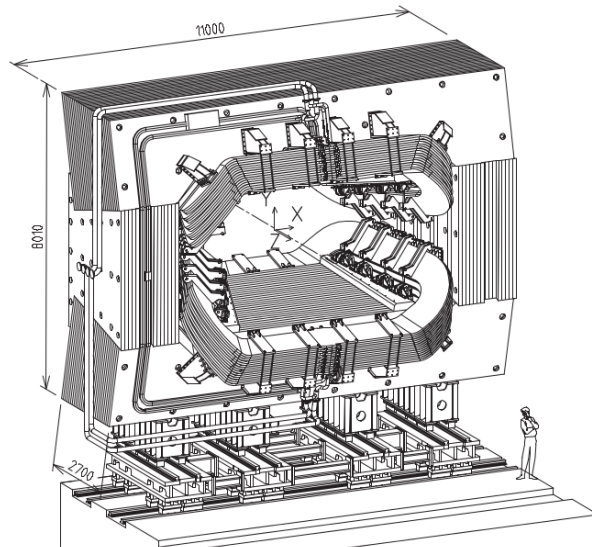

Figure 6: The LHCb magnet viewed from the larger aperture side [14].

The two coils are identical and placed mirror-symmetrically to each other. Their polarity is reversed periodically during data taking, which on average prevents the measurement of detector asymmetries. The polarity is denoted 'MagUp' and 'MagDown' for positive and negative polarity, respectively.

---

[13]$\sigma(pp \to b\bar{b}X) = (284 \pm 20 \pm 49)\mu$b at $\sqrt{s} = 7$ TeV [20].

[14]$\sigma_{inel} = 66.9 \pm 2.9 \pm 4.4$ mb at $\sqrt{s} = 7$ TeV after extrapolation to the full phase space. The first uncertainty is experimental and the second is due to the extrapolation. [21]

[15]A superconducting magnet was first proposed but omitted because of financial issues.

4.2.3 Tracking

The tracking system of LHCb uses two different detector technologies - silicon microstrips and straw tubes - in the Vertex Locator (VELO) and four planar tracking stations, of which three are located downstream of the magnet and one is placed upstream. In Figure 5 they are labelled TT, T1, T2, T3.

The Vertex Locator The Vertex Locator (VELO) is arranged around the proton-proton interaction point and is therefore the first detector that is traversed by particles. Its purpose is to identify the primary vertices (PV) and the displaced secondary vertices that are distinctive for b and c-hadron decays. It consists of 42 circular silicon modules that are composed of two parts measuring either the distance R to the beamline or the azimuthal angle ϕ in the x-y plane with silicon strips (see Figure 7). The full diameter of a module is 90.5 mm, 300μm in thickness and the minimum pitch between the strips in the inner region is around 40μm. The sensors are placed at a radial distance of about 5mm from the beam axis at a known position on the z-axis. The best hit resolution is around 4 μm [19].
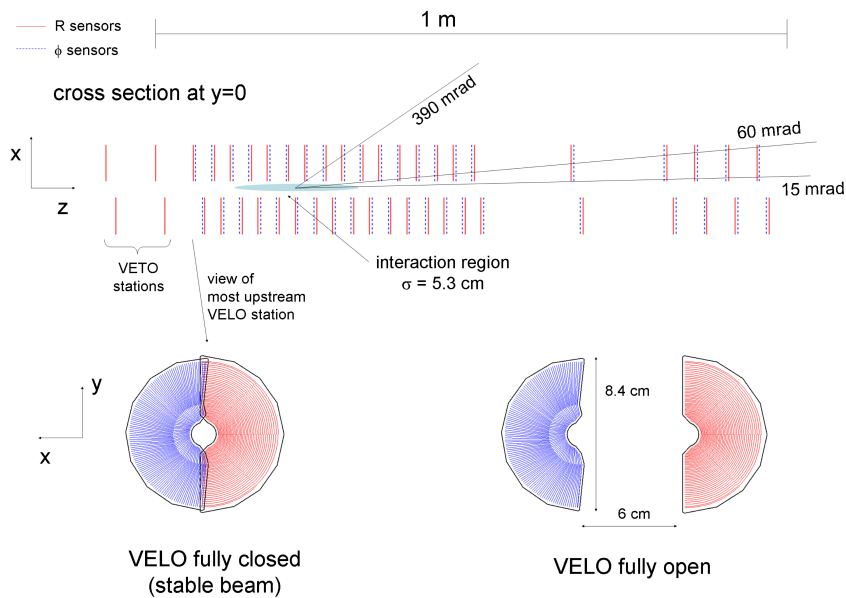

Figure 7: Setup of VELO silicon modules along the beam axis. Also shown are the angles for which at least three VELO stations are crossed, which is the minimum requirement to reconstruct a particle trajectory. [14]

The Silicon Tracker In addition to the VELO, there are two more silicon detector modules that are defined as the Silicon Tracker (ST): the Trigger Tracker[16] (TT) and the Inner Tracker (IT) (see Figure 8). Both TT and IT use silicon microstrip sensors with a strip pitch of about $200\,\mu$m, resulting in a spatial resolution of about $50\,\mu$m.

The TT is located upstream in front of the magnet. It is a 150 cm wide and 130 cm high planar tracking station, covering the full acceptance of the experiment. Both ST stations are composed of four layers arranged in an (x-u-v-x) configuration: vertical strips in the first and the last layer, and strips rotated by a stereo angle of -5° and +5° in the x-y-plane in the second and the third layer, respectively.

The IT is installed downstream of the magnet in the centre of the tracking stations T1-T3 (see Figure 8), close to the beam pipe. It is a 120 cm wide and 40 cm high silicon microstrip detector that does not cover the entire acceptance of LHCb but is chosen to be in the inner region because of high track density.

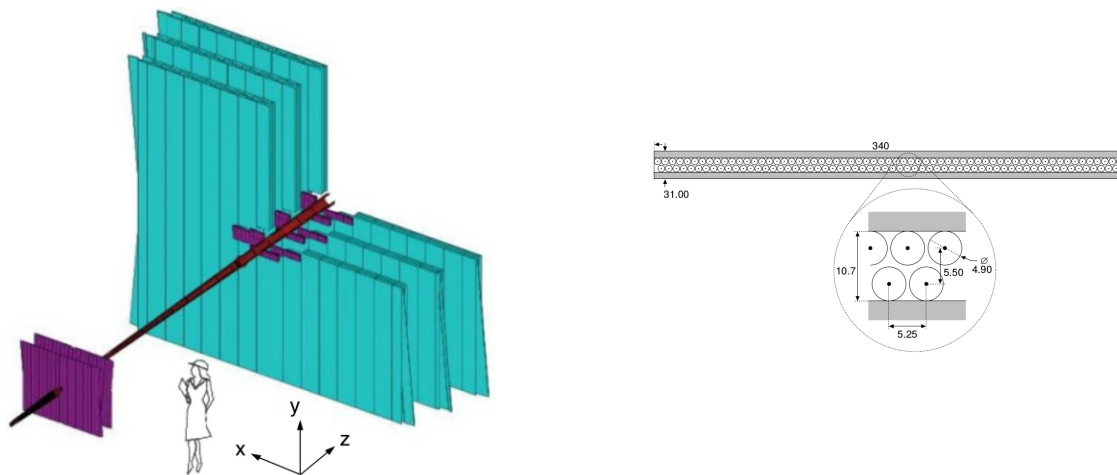[16]The Trigger Tracker is also known as Tracker Turicensis.

Figure 8: Left: The LHCb tracking system, on the left front is the TT. Back right are T1, T2 and T3 stations. In purple are the ST, in turquoise the straw tube OT. Right: Sketch of the cross section of a straw-tubes module in the OT.

The Outer Tracker The Outer Tracker (OT) covers the outer region of the three tracking stations T1-T3 (see Figure 8 left) where the particle flux is lower than in the IT. It is a drift-time detector, for the tracking of charged particles. The gas-tight straw-tubes modules that are employed here contain two layers of drift-tubes with inner diameters of 4.9 mm (see Figure 8 right). The spatial resolution of a single cell is about $200\,\mu$m. The layout of the OT is similar to that of the IT and TT, with four layers in an (x-u-v-x) arrangement. With an active area of 5971×4850 mm^2, the outer boundary corresponds to the full acceptance of LHCb.

Track reconstruction The trajectories of charged particles produced in a collision are reconstructed from hits in the VELO, TT, IT and OT detectors. The most important tracks for physics analyses are the long tracks, defined as tracks with hits in both, the VELO and the T stations, and optionally in the TT. These tracks have the most precise momentum estimate. They are mostly reconstructed by first searching in the VELO for straight line trajectories and then combining them with information from the T stations. This set of tracks is checked for consistency with TT hits to improve the momentum determination. Finally, the tracks are fitted using a Kalman filter algorithm, which takes into account multiple scattering and energy loss due to ionisation. The quality of the fit and the track is determined by the χ^2 per degree of freedom (χ^2/ndf). An example of reconstructed tracks in a typical event is shown in Figure 9.
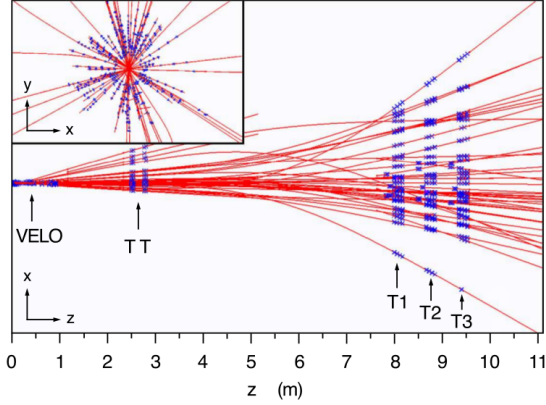
Figure 9: Display of reconstructed tracks and assigned hits in an event (simulation without noise). The insert shows a zoom in the VELO region. [14]
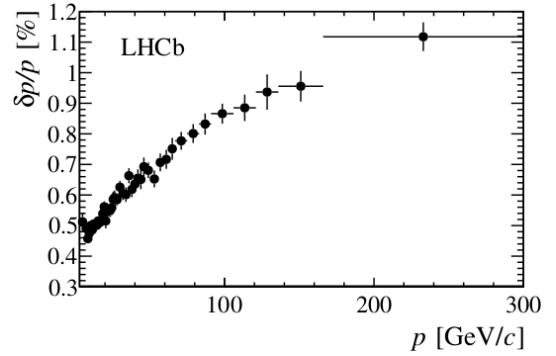

Figure 10: Relative momentum resolution versus momentum for long tracks in data obtained using J/ψ decays. [19]

With the reconstructed trajectory of the particle and the known magnetic field, the momentum of the particle is determined. The momentum resolution is about 5 per mille for particles below 20 GeV/c and about 8 per mille for particles around 100 GeV/c (see Figure 10).

4.2.4 Particle identification

In order to provide complete information on the four-momentum, the mass (and therefore identity) of the particles has to be determined. The particle identification (PID) in LHCb is provided by four different detectors: the two RICH detectors, the calorimeter system and the muon stations.

The RICH system   The primary role of the Ring Imaging Cherenkov detector (RICH) system is the identification of charged hadrons (π, K, p). Especially the K-π discrimination is crucial, since these particles are often produced in decays of B and D hadrons. This task is performed by the two RICH detectors RICH1 and RICH2 (see Figure 5).

When a charged particle passes through a dielectric medium faster than the local speed of light, Cherenkov photons are produced at an angle that depends on the particle's speed. This angle is measured by detecting the photons. Then, knowing the momentum of the particle (Section 4.2.3), the mass can be deduced.

Three different radiators of different refractive indices are used to cover a reasonable momentum range. RICH1 is located upstream of the LHCb magnet and covers the momentum range 1-60 GeV/c using a 5 cm thick layer of aerogel and C_4F_{10} as radiators. RICH2, downstream of the magnet, uses CF_6 gas to identify particles with momentum from 15 GeV/c to above 100 GeV/c. In Figure 11 the Cherenkov angle versus the momentum is shown for isolated tracks.

The calorimeter system   The calorimeter system of LHCb measures positions and energies of hadrons, electrons and photons. This information is crucial for PID and the trigger. The calorimeter is located downstream of RICH2 between the first and the second muon stations (M1, M2 in Figure 5). The calorimeter comprises four components:

- **Scintillating Pad Detector**
  A Scintillating Pad Detector (SPD) forms the first layer of the calorimeter. It provides a trigger signal for charged particles, which interact in the scintillator material, in order to reduce background from neutral particles. This is crucial because no track requirements are set which could distinguish charged from neutral particles.
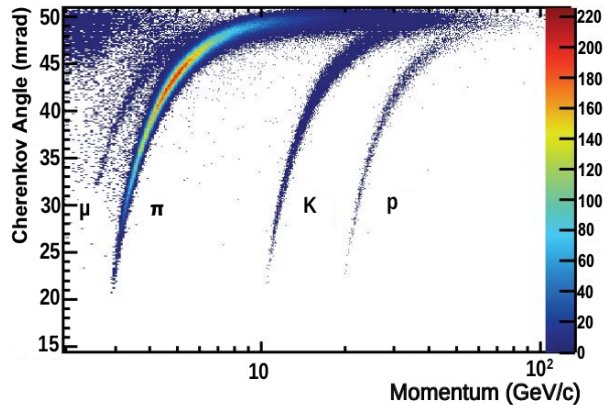
23

Figure 11: Cherenkov angle as a function of momentum in the C_4F_{10} radiator for isolated tracks in data. The curved bands clearly show the different types of particles. [19]

- **PreShower detector**
  The PreShower detector (PS) with its 15 mm of lead causes electrons to shower. Since hadrons have a longer interaction length than electrons, the PS provides the longitudinal segmentation that is required to distinguish the electrons from charged hadrons. In addition, the PS has a finer granularity than the ECAL which provides a better localisation of electrons and photons in combination of both detectors.

- **Electromagnetic calorimeter**
  The electromagnetic calorimeter (ECAL) is composed of several absorption layers (2 mm lead, 120 μm white paper), each followed by a scintillator, forming a 42 cm deep stack corresponding to 25 radiation lengths. Its purpose is the detection of electrons and photons. The energy resolution by design is $\sigma_E/E = \frac{10\%}{\sqrt{E}} \oplus 1\%$ (E in GeV).

- **Hadronic calorimeter**
  The hadronic calorimeter (HCAL) is made from iron and scintillating tiles, as absorber and active material respectively. The thickness of the iron is chosen to be 5.6 interaction lengths. Its purpose is the detection of hadrons. The energy resolution was determined to be $\sigma_E/E = \frac{(69\pm5)\%}{\sqrt{E}} \oplus (9\pm2)\%$ (E in GeV).

Electrons and photons create electromagnetic showers in the ECAL's absorber material by bremsstrahlung and pair production. The charged particles in these showers create photons in the scintillator, which are then detected in photomultipliers. The same principle is used in the HCAL, high energetic hadrons produce hadronic showers in the absorber, creating charged particles that are detected.

Neutral pion reconstruction   The ECAL is used to identify electrons and photons. Neutral pions with low transverse momenta are mostly reconstructed as pairs of separated photons (resolved π^0 candidates, see Figure 12) with a mass resolution of 8 MeV/c^2 [19]. Due to the finite granularity of the ECAL, photon pairs with high transverse momentum (>2 GeV/c) cannot be resolved as individual clusters. These pairs of photons are reconstructed as 'merged' π^0 candidates using an algorithm that further splits each single ECAL cluster into sub-clusters. The resolution of merged π^0s is worse than the resolution for resolved ones.

The procedure that is used to reconstruct a π^0 is as follows: the photon candidates are reconstructed assuming them to come from the PV and their direction pointing to the 3D barycentre of the produced shower in the calorimeter. It is then looped over all photon candidates, pairing them and comparing the

24

corresponding invariant masses with the nominal π^0 mass. Among the photon candidates, only those with a transverse momentum greater than 200 MeV/c are kept and paired to reconstruct a neutral pion. Hence, the resolved π^0 identification efficiency depends strongly on p_T because at low p_T, one of the two γ's is more likely to not pass the minimum p_T of 200 MeV/c. The cut is nevertheless necessary to reduce combinatorial background. The efficiency of the π^0 reconstruction is shown in Figure 13 and is overall around 50%. It has to be mentioned that this efficiency accounts only for the π^0s in the detector acceptance, but the calorimeter has a large hole at the position of the beam line which leads to a large loss of photons in this area. For more detailed information on photon and π^0 reconstruction at LHCb see for example [22].

The muon system Since muons are present in the final states of many interesting B decays and therefore vital to the physics programme of LHCb, an efficient identification is a crucial requirement. The muons are detected in five muon stations, marked M1-M5 in Figure 5, that provide efficient muon triggering and offline identification. M1 is located upstream in front of PS calorimeter, to improve p_T measurement in the trigger. M2 to M5 are placed downstream the calorimeters. Between these stations, there are iron absorbers, each 80 cm thick, to reduce the hadronic background. The minimum momentum of a muon to cross all five stations including the calorimeter is approximately 6 GeV/c. M1-M5 have a relatively high spatial resolution along the x coordinate to define the track direction and to calculate the p_T of the muon candidate with a resolution of 20%.

The muon stations have an inner and outer acceptance of 20 (16) mrad and 306 (258) mrad in the bending (non-bending) plane, respectively. The stations comprise mainly Multi Wire Proportional Chambers, only the inner region of M1 is made of triple-GEM (gas electron multiplier) chambers, due to a higher particle flux.

The muon selection efficiency depends on p_T and is measured to be greater than 92% for muons with transverse momentum $0.8 < p_T < 1.7$ [GeV/c] and greater than 96% for those with $p_T > 1.7$ [GeV/c]. The misidentification rates are 1-2% for protons, pions and kaons with $p_T > 1.7$ [GeV/c] [19].

The PID information obtained from the muon, RICH, and calorimeter systems is combined into a likelihood function. The log likelihood difference, $\Delta \log \mathcal{L}(X - \pi) \equiv \log[\mathcal{L}(X)/\mathcal{L}(\pi)]$, compares the calculated likelihoods between two mass hypotheses, where the default hypothesis is 'pion' because it is the most abundant particle. The compared particle X is either a kaon, proton, electron or muon (see Figure 11). The $\Delta \log \mathcal{L}$ variable reflects how likely a mass hypothesis is.

4.2.5 Trigger

The LHCb trigger system is subdivided in a low level pure hardware part (Level 0, L0) and a high level software part (High Level Triggers, HLT1 and HLT2). The trigger has to decide which events are kept for further analysis which is essential as the crossing frequency with visible[17] interactions is about 10 MHz. This rate is reduced to approximately 5 kHz by the trigger and can then be written to storage for offline analysis.

L0 trigger The L0 is a hardware trigger that reduces the 10 MHz collision rate down to 1 MHz using information from the muon and calorimeter system. It detects muons with high p_T and objects with high E_T indicating the decay products of heavy meson. At 1 MHz the full detector can be read out.

---

[17]An interaction is defined to be visible if it produces at least two charged particles with sufficient hits in the VELO and T1–T3 to allow them to be reconstructible.

HLT The HLT is a software trigger consisting of a C++ application that runs on ~ 2000 multicore processors in the Event Filter Farm. The HLT1 is fed with the detector output at 1 MHz and reduces the rate to 11 kHz by reconstructing tracks with the information from the VELO and the tracking stations. It confirms or rejects candidates from the L0 trigger.

The HLT2 further reduces the event rate to about 5 kHz using global track reconstruction. It uses techniques similar to those used in offline analyses.
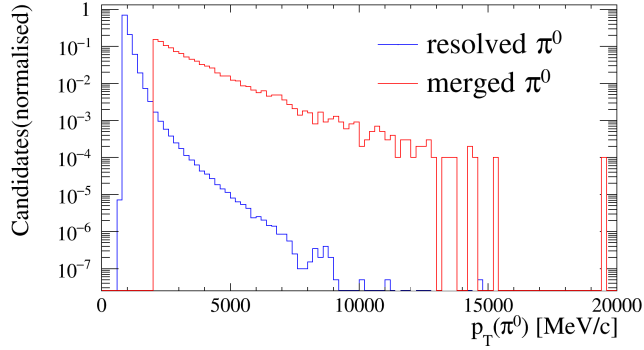
Figure 12: Transverse momentum of reconstructed neutral pions from the decay $B^+ \to K^{*+}(\to K^+\pi^0(\to \gamma\gamma))\mu\mu$ (data). The contribution of resolved and merged π^0 is indicated with blue and red histograms, respectively.
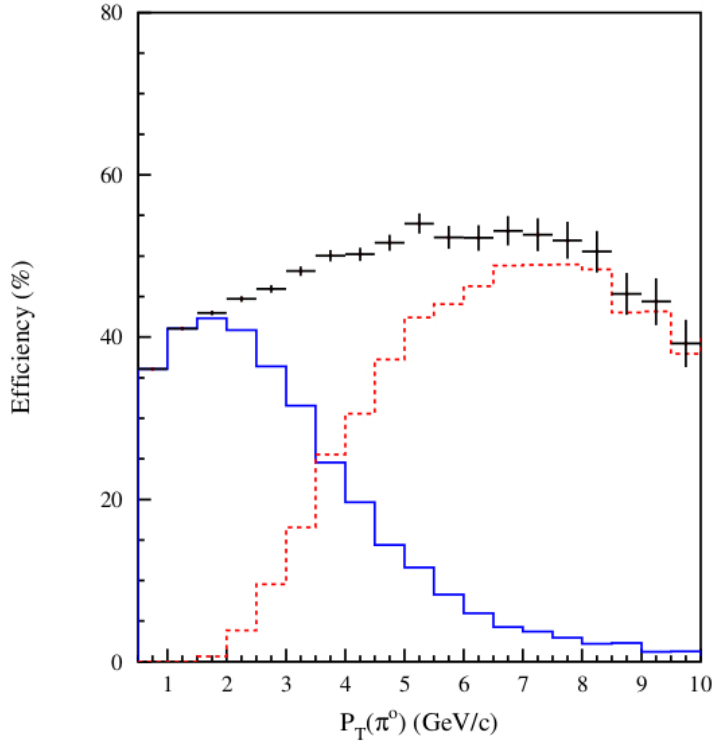
Figure 13: Overall π^0 efficiency (number of $\pi^0 \to \gamma\gamma$ identified in the mass window over the number of $\pi^0 \to \gamma\gamma$ in the detector acceptance with $p_T(\gamma) > 200$ MeV/c) (simulation). The separate contributions from resolved and merged π^0 are indicated by the solid and dashed histograms, respectively. [22]

26

5 Data analysis

This section depicts the procedure of the data analysis. The dataset, variables and methods used in the analysis are described, and the signal selection is documented.

5.1 Analysis strategy

The goal of this thesis is to determine the branching fraction of the rare B meson decay

$$B^+ \to K^{*+}\mu^+\mu^-$$

relative to the tree-level decay

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

In both cases the decay is followed by $K^{*+} \to K^+\pi^0$. The normalisation to the tree-level decay channel - in the following referred to as normalisation channel - is used since the decay products in each channel are expected to have similar kinematic properties and thus most systematic uncertainties cancel in the ratio, in particular, those related to the muon and π^0 reconstruction and identification. Furthermore, the branching fraction of the normalisation channel is well known (cf. Table 3) with a relative uncertainty of $\sim 6\%$ and is therefore used to calculate the absolute branching fraction of $B^+ \to K^{*+}\mu^+\mu^-$. This decay of a charged B meson into an excited kaon and a non-resonant muon pair has already been observed and studied by various collaborations, including LHCb. For this reason, the analysis is done without blinding[18] the signal region.

The analysis strategy consists of several steps:

0. Dataset preparation. Datasets with a loose preselection[19] are produced centrally by the LHCb collaboration to minimise the computing effort of the individual analyses (see Section 5.2.2).

1. Signal preselection. Loose cuts are applied to the dataset (see Section 5.2.2) in order to reduce combinatorial background and to make the signal of the normalisation channel visible.

2. Signal selection. The combinatorial background is further suppressed by a multivariate selection. The multivariate classifier is trained on data and the selection is optimised using the normalisation channel.

3. Signal fit. After the selection, the mass distributions of signal candidates for $B^+ \to K^{*+}\mu^+\mu^-$ and $B^+ \to J/\psi(\to \mu^+\mu^-)K^{*+}$ are fitted using unbinned maximum likelihood fits (see Section 5.1.1) to obtain the signal yields.

4. Efficiency determination. The efficiency of the signal selection for both channels is determined using simulated[20] samples.

5. Branching fraction determination. The relative branching fraction is determined by

$$\frac{\mathcal{B}(B^+ \to K^{*+}\mu^+\mu^-)}{\mathcal{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^{*+})} = \frac{N(B^+ \to K^{*+}\mu^+\mu^-)}{N(B^+ \to J/\psi(\to \mu^+\mu^-)K^{*+})} \times \epsilon' \times \zeta'$$

where N denotes the signal yield obtained by the fits, ϵ' is the relative efficiency of the signal selection and ζ' denotes the relative geometrical detector acceptance of simulated events.

---

[18]Blinding means the covering or hiding of the signal region and signal related results during the analysis in order to prevent a bias by expectations.

[19]Also referred to as 'stripping'.

[20]Simulated samples are generated with the so-called Monte Carlo (MC) method.

5.1.1 Maximum likelihood method

All fits in this thesis are performed with the unbinned maximum likelihood method. Considering n measurements of a vector $\vec{x}$ of random variables with measured values $\vec{x}_1, \vec{x}_2, ..., \vec{x}_n$ and known probability density function (PDF) $f(\vec{x}|\vec{a})$, where $\vec{a}$ is a vector of unknown parameters, the likelihood function is defined as

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

with $\int_\Omega f(\vec{x}|\vec{a})\mathrm{d}\vec{x} = 1$ for all $\vec{a}$. The likelihood function gives the probability of measuring the dataset with a given parameter set. The estimate of the parameter $\vec{a}$ is obtained by maximising the likelihood function [23].

In practice, the negative log-likelihood function is often used, which then has to be minimised:

$$F(\vec{a}) = -\ln\mathcal{L}(\vec{a}) = -\sum_{i=1}^{n} \ln f(\vec{x}_i|\vec{a}).$$

5.2 Dataset and variables

This section defines variables exploited in the analysis and gives an overview of the used dataset.

5.2.1 Definition of variables

Many variables used in this analysis are deduced directly from the measured four-momenta of the final state particles: K^+, γ_1, γ_2, μ^+, μ^-. The four-momentum $\mathbf{P}$ is a conserved quantity and its square is invariant under Lorentz transformation. The four-momentum of a decaying particle is calculated by adding the four-momenta of its decay products, e.g. :

$$\mathbf{P}_{B^+} = \mathbf{P}_{K^+} + \mathbf{P}_{\gamma_1} + \mathbf{P}_{\gamma_2} + \mathbf{P}_{\mu^+} + \mathbf{P}_{\mu^-}.$$

Mass variable   The invariant mass of the B^+ meson can then be calculated from its four-momentum by:

$$m_{B^+} = \sqrt{\mathbf{P}_{B^+}^2}$$

Transverse momentum   The transverse momentum is defined as the momentum component transverse to the z axis in the experimental frame (see Section 4.2.1):

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

Pseudorapidity   The pseudorapidity is defined by the angle θ between the particle momentum and the beam axis (see Figure 14 left):

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

Opening angle   The opening angle is the angle between the particle trajectories coming from the same vertex. In this thesis, it is used only for the angle between the two photons coming from the π^0 (see Figure 14 right).

$$OpenAngle = \cos\beta = \frac{\vec{p}_{\gamma_1}\vec{p}_{\gamma_2}}{|\vec{p}_{\gamma_1}||\vec{p}_{\gamma_2}|}$$

Cone p_T asymmetry   The cone p_T asymmetry uses the p_T of all particle tracks inside a cone, the cone end of which points to the corresponding vertex. The variable is only used for the B^+ candidates in this thesis. This means that the cone is spanned around the trajectory of the B^+ candidate so that the cone end points to the PV (see Figure 15). The asymmetry is calculated by

$$A_{p_T}^{Cone} = \frac{p_{T,B^+cand.} - \sum p_{T,otherTrack}}{p_{T,B^+cand.} + \sum p_{T,otherTrack}}$$

where $p_{T,B^+cand.}$ is the transverse momentum of the B^+ candidate and $\sum p_{T,otherTrack}$ the sum of all transverse momenta of other tracks in this cone. Hence, when there is only the B^+ candidate track inside the cone, the asymmetry is unity.

Photon confidence level   The confidence level (CL) of the photon is calculated from PID variables (see Section 4.2.4) values and additionally takes into account information about cluster size and shower shape in the calorimeter, and energy deposit in the SPD.

Impact parameter   The impact parameter (IP) is the minimal distance of a reconstructed track to any PV (see Figure 16a). The $IP\chi^2$ is defined as the difference between the χ^2 values of the fit of the primary vertex that is reconstructed with and without the considered track. Since daughter particles of the B meson come from a detached vertex, their IP and $IP\chi^2$ must be larger than zero, in contrast to background particles which are produced at the PV in many cases.

Direction angle   The direction angle (DIRA) of the mother particle is defined as the cosine of the angle α between its flight direction and its reconstructed momentum (see Figure 16b). The flight direction is obtained by connecting the PV and the secondary vertex (SV). For well reconstructed momenta and vertices, and if no particles are missed in the reconstruction of the SV, this angle is small.

Track and vertex quality   The variables corresponding to the track and vertex quality are $Track\chi^2$ and $Vertex\chi^2$, respectively. They are computed by the track or decay vertex fit.
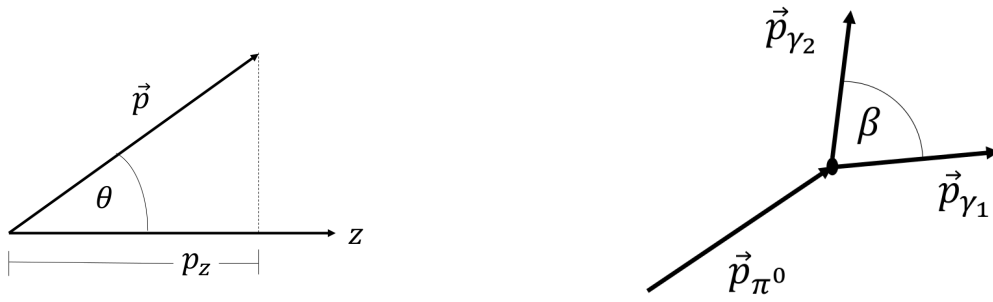
Figure 14: Schematic depiction of the components of the pseudorapidity (left) and of the opening angle (right).
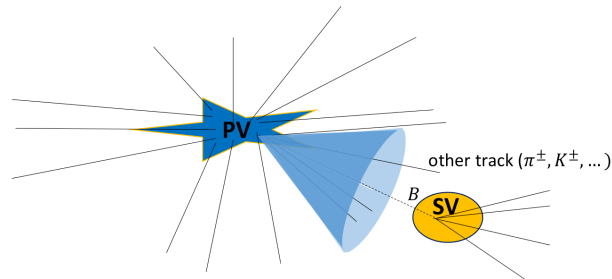

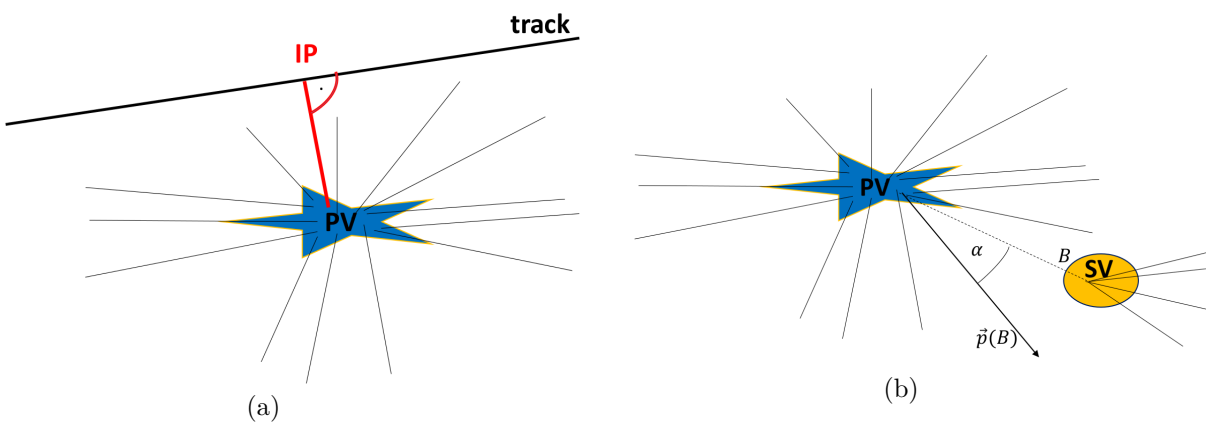**Figure 15:** Schematic depiction of the 'Cone p_T asymmetry'.

Figure 16: Schematic depiction of the impact parameter (a) and the direction angle (b).

5.2.2 Dataset

The data collected with the LHCb detector in the years 2011 and 2012 with $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV, respectively, is used in this analysis (MagUp and MagDown for both years, see Section 4.2.2) whereby only resolved π^0s have been taken into account because the merged π^0s contribute considerably less statistics. The integrated luminosity of both years combined is approximately 3 fb^{-1}.

The $B^+ \to K^{*+}\mu^+\mu^-$ and $B^+ \to J/\psi K^{*+}$ candidate events are required to pass the two-stage trigger system (see Section 4.2.5). The initial hardware stage selects events with one or two muons in the final state with sufficient transverse momentum. The subsequent software stage requires final-state particles to originate from a vertex that is displaced from a PV. The fit quality of the tracks is required to be high. The software stage therefore applies cuts on the IP and $IP\chi^2$ variables, as well as on the transverse momenta and the $Track\chi^2/ndf$ of the particles. A detailed description of the different trigger decisions is given in [24].

Furthermore, the candidates for $B^+ \to K^{*+}\mu^+\mu^-$ and $B^+ \to J/\psi K^{*+}$ are required to pass the selection `StrippingB2XMuMu` version Stripping21, which explicitly selects the decay of a B meson into K^{*+} and two muons in the final state. These requirements are summarised in Table 6. Candidates that pass the stripping line are then preselected additionally in order to reduce a large amount of combinatorial background. These preselection cuts are chosen because their background-discriminating power is found to be strong. The signal efficiency of the preselection is around 70% according to simulations.

Candidate	Selection
B meson	4900 MeV/c^2 < M < 7000 MeV/c^2
	$IP\chi^2$ < 16
	$Vertex\chi^2/ndf$ < 8
	$DIRA$ > 0.9999
	flight distance (FD) χ^2 > 121
K^{*+}	$|m(K^+\pi^0) - M(K^{*+})|$ < 300 MeV/c^2
	$Vertex\chi^2/ndf$ < 9
	FD χ^2 > 9
$\mu^+\mu^-$	$m(\mu^+\mu^-)$ < 7100 MeV/c^2
	$FD\chi^2$ > 9
π^0	p_T > 800 MeV/c
muon	$IP\chi^2$ > 9
	`IsMuon == True`
	$PID_{\mu\pi}$ > −3
tracks	ghost prob.< 0.35
Global Event Cut (GEC)	nSPDHits< 600

Table 6: Stripping selection criteria in `B2XMuMu` for Stripping 21.

Variable	Selection
K^{*+} mass	$792 \text{ MeV}/c^2 < m(K^+\pi^0) < 1050 \text{ MeV}/c^2$
$B^+ \, A_{P_T}^{Cone}$	> -0.5
$B^+ \, p_T$	$> 2000 \text{ MeV}/c$
$B^+ \, Vertex\chi^2$	< 12
$B^+ \, \eta$	< 4.9
$B^+ \, DIRA$	> 0.99996
$K^+ \, Track\chi^2/ndf$	< 2
$K^+ \, p_T$	> 300
$K^+ \, PID_{K\pi}$	> 0
$\gamma_{1/2} \, CL$	> 0.15

Table 7: Preselection cuts applied to stripped candidates.

The cuts are checked to be uncorrelated to the mass of the B candidate. These additional criteria are listed in Table 7.

In addition to the genuine data from the LHCb detector, two simulated samples are used in the analysis (see Table 8). These samples have the advantage that rare decays can be produced with high occurrence and that all real (i.e. true) values of the variables are known as well as the reconstructed ones. It is therefore possible to 'truthmatch' these samples, which means that a pure signal sample can be extracted from these simulated events by requiring correct identification and matching of reconstructed mother particles and their daughters to the simulated decay. A pure signal sample is crucial for the determination of the selection efficiencies. The simulated data taking conditions correspond to the year 2012.

Decay	#simulated events in detector acceptance	#truthmatched signal candidates
$B^+ \rightarrow K^{*+}\mu^+\mu^-$	144672	16284
$B^+ \rightarrow J/\psi K^{*+}$	146321	19313

Table 8: Simulated samples used in the analysis. These samples passed the same stripping as data but the given numbers are without preselection.

5.2.3 Mass distributions in data

In Figure 17a and Figure 18a the mass distribution for the B^+ candidates and the $\mu\mu$ invariant mass are shown after the stripping, respectively. The $\mu\mu$ spectrum comprises two clearly visible peaks at around 3100 MeV/c^2 and 3700 MeV/c^2, respectively. The background in this spectrum is small compared to the peak size since the two muons give a relatively clean detector signal. The resonances that correspond to the peaks are the $J/\psi(1S)$ with $m_{J/\psi}^{PDG} = 3096.916 \pm 0.011$ MeV/c^2 and the $\psi(2S)$ with $m_{\psi}^{PDG} = 3686.109^{+0.012}_{-0.014}$ MeV/c^2. The spectrum of the B^+ candidates is dominated by combinatorial background mainly caused by the vast amount of neutral pions coming from the PV and from hadronic interaction of particles with the detector.

After the preselection, the shape of the $\mu\mu$ spectrum has not changed (see Figure 18b). However, in the spectrum of the B^+ candidates (see Figure 17b) a peak at around 5300 MeV/c^2 is visible which is caused dominantly by the decay $B^+ \to J/\psi(\to \mu^+\mu^-)K^{*+}$.

(a) (b)
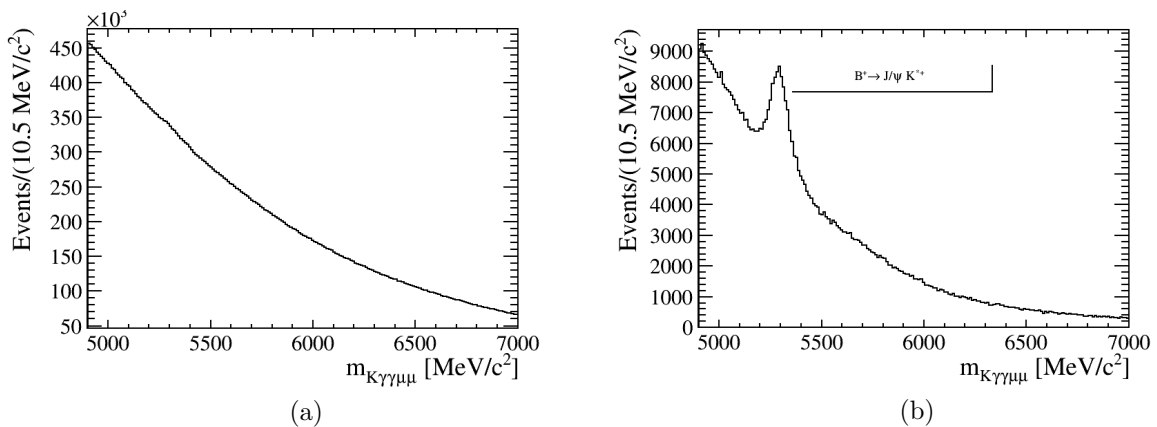
Figure 17: Mass spectrum of B^+ candidates after stripping (a) and additional preselection (b) for 2011 and 2012 data combined.

(a) (b)
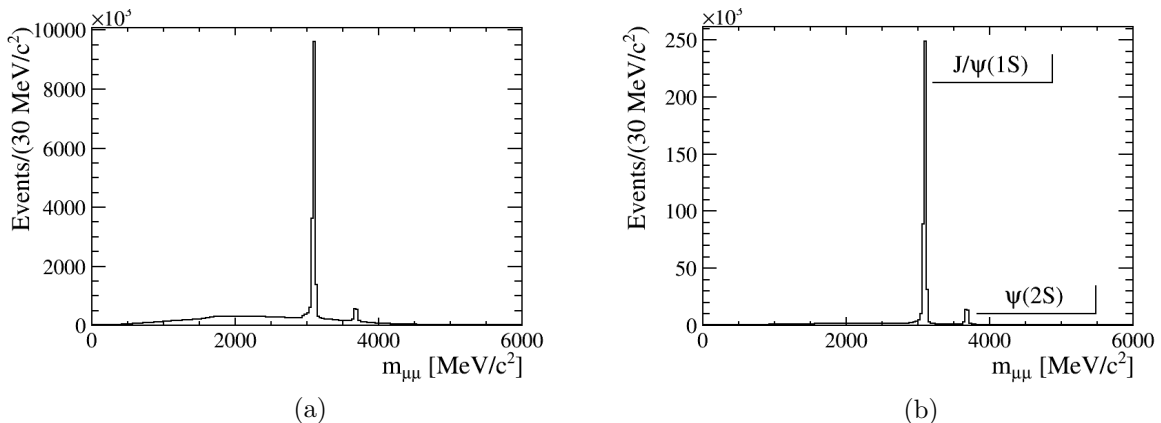
Figure 18: $\mu\mu$ mass spectrum after stripping (a) and additional preselection (b) for 2011 and 2012 data combined.

33

5.3 Signal selection

This section documents the signal selection process after the preselection. The objective is to achieve an effective separation of signal and background in the $B^+ \to K^{*+}\mu^+\mu^-$ and $B^+ \to J/\psi K^{*+}$ decay channel. The preselection has already reduced a certain amount of background using variables that allow a distinction between signal and background. In these variables, knowledge of the kinematics and vertex properties of the decay is exploited. This approach is followed up by using well-separating variables in a multivariate analysis method - the boosted decision tree (BDT). The BDT needs pure signal and background samples as input. There are two common methods to produce pure signal samples: using a truthmatched decay sample in simulation and correct for possible deviations between data and simulation, or using event weights to unfold a pure signal distribution from background on data. Both approaches are expected to perform similarly. However, the correction for possible deviations between data and simulation is difficult or even impossible. Especially deviations that are not well understood and multiple correlations between variables may complicate the correction procedure. Hence, the latter approach is chosen to ensure a proper performance of the BDT.

The selected signal will later be visible in the B^+ mass spectrum. The width of the resulting peak is mostly governed by the momentum resolution of the daughter particles and, as already has been mentioned, the momentum resolution of the π^0 is relatively low. In order to investigate a possible workaround for this problem, also an invariant mass distribution of the B^+ is used where the neutral pion mass is constrained to its PDG value in the fit of the B^+ decay [25]. Since this constraint changes the shape of the B^+ mass distribution significantly, the signal selection with constrained and unconstrained π^0 mass is conducted separately. It is expected that the mass constraint has an influence on the significance of the signal at the end of the selection because the constraint narrows the peak in the B^+ mass spectrum. In order to document the effect of the constraint in comparison to the B^+ mass distribution without constraint, both spectra are shown throughout the signal selection.

5.3.1 Unfolding a pure signal sample

The unfolding of the signal and the background contribution on data is done via the $_s\mathcal{P}lot$ technique. The $_s\mathcal{P}lot$ is a statistical tool that uses a so-called *discriminating* variable with known sources of events (i.e. signal and background) to infer the behaviour of the individual sources of events with respect to so-called *control* variables. It is important to note that these control variables are assumed to be uncorrelated with the discriminating variable. The knowledge about the discriminating variable is obtained using a maximum Likelihood fit, where the individual sources of events are parameterised by yield parameters. The result of an $_s\mathcal{P}lot$ is the so-called *sWeight*, which is an event weight that reproduces the true distribution of a source on average by summing over all events. For a more detailed description of the $_s\mathcal{P}lot$ technique see [26].

The discriminating variable used here is the invariant mass of the B meson. This is convenient because it is known from simulated data how the signal events are distributed and thus a model assumption for signal and background can be made. Furthermore, the multivariate analysis that is conducted later on also requires variables that are uncorrelated to the B mass.

Since the signal of the rare $B^+ \to K^{*+}\mu^+\mu^-$ decay is not visible on data yet, the signal of the normalisation channel is used as a proxy. On the preselected data (Figure 17b), an additional cut is therefore applied on the $\mu\mu$ invariant mass spectrum around the J/ψ mass ($m_{\mu\mu} \in [2780, 3250]$ MeV/c^2). The mass distributions with the fits are shown in Figure 20 and 21, respectively. The fit model is a double Crystal Ball[21] (CB) function for the $B^+ \to J/\psi K^{*+}$ component and an exponential function for the

---

[21]Named after the Crystal Ball Collaboration - more precisely after the Crystal Ball detector - at the Stanford Linear Accelerator Center [27].

combinatorial background.

The CB function consists of a Gaussian core and a power-law tail. The tail takes into account that particles can lose energy in radiative processes like bremstrahlung. The CB function is continuous, it has four parameters and it is defined as

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

where

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

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

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

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

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

N is a normalisation factor and erf is the error function. The parameter α marks the transition between Gaussian and tail, n determines the shape of the power-law, μ is the peak position and σ is the width of the Gaussian.

The double CB is chosen in order to take detector resolution into account and to provide a tail to the left as well as to the right (negative α). The tail to the left is chosen to describe the bremsstrahlung of the muons. The tail to the right models a structure that comes from overestimated photon energies in the calorimeter. This effect is attributed to the non-perfect treatment of the cases where the calorimeter clusters associated with photons are overlapped in space and the energy release in the shared calorimeter cells is not taken into account properly. The parameters of the power-laws are fixed fitting the corresponding mass distribution on simulated data (see exemplary Figure 19) and in case of the fit of the invariant B mass with constrained pion mass, the ratio of the widths σ_2/σ_1 is fixed additionally. The fixed parameters are listed in Table 9. The model for the signal is given by:

$$F_S = f \cdot CB(m; \alpha_1, n_1, \mu, \sigma_1) + (1 - f) \cdot CB(m; \alpha_2, n_2, \mu, \sigma_2)$$

where f denotes the fraction of the first CB and m is the mass.

The background model is given by:

$$F_B = \frac{1}{N}e^{\tau m}$$

where N is a normalisation and τ the slope of the exponential.

The overall model for the distribution is then given by:

$$F_{S+B} = N_{sig}F_S + N_{bkg}F_B$$

where N_{sig} and N_{bkg} denotes the number of signal and background events, respectively.

Parameter	Fixed value
α_1	0.87
α_2	-0.879
n_1	2.7
n_2	8.3

(a)

Parameter	Fixed value
α_1	0.73
α_2	-0.7715
n_1	3.33
n_2	5.4
σ_2	$1.12\sigma_1$

(b)

Table 9: Parameters that were fixed on simulated data for the fit of B mass without π^0 mass constraint (a) and with constrained π^0 mass (b).
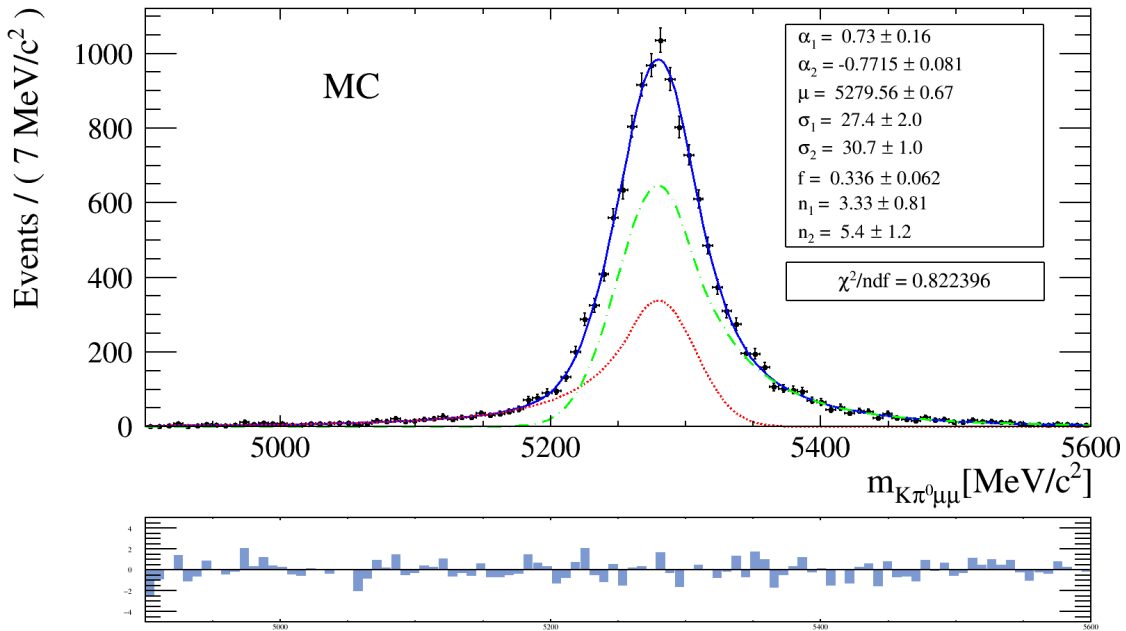

Figure 19: Exemplary fit of the B mass distribution with π^0 mass constraint from $B^+ \rightarrow J/\psi K^{*+}$ on simulated data. The fit model is a double CB where CB_1 is the dotted-dashed green line and CB_2 is the dashed red line.
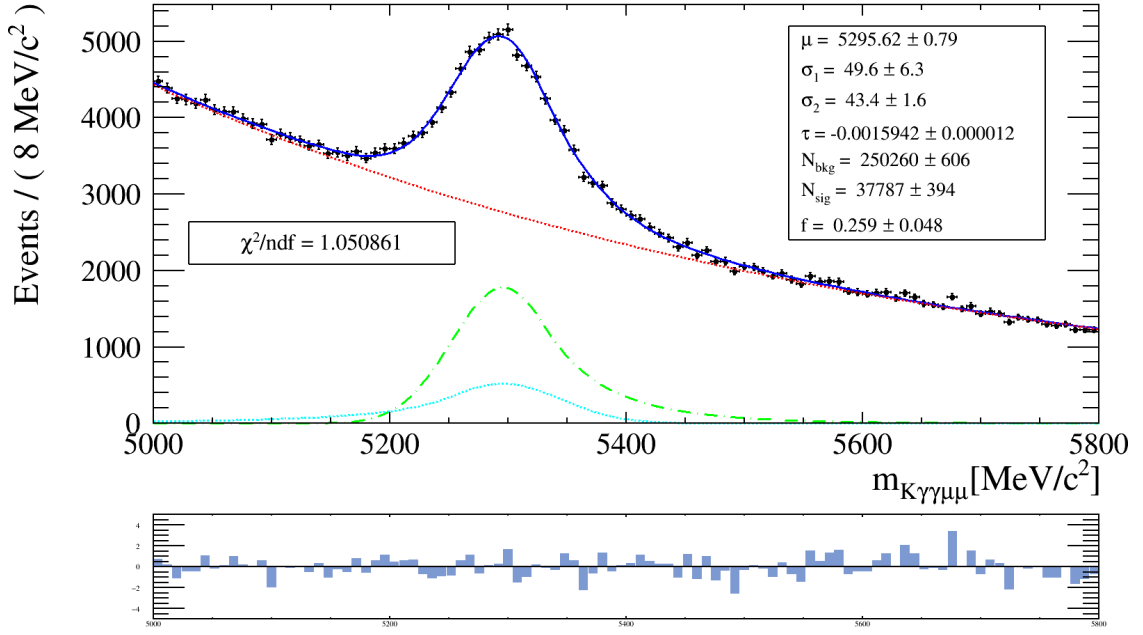
Figure 20: Fit of the B mass spectrum for the normalisation channel with a double CB function and an exponential. The two CB components are drawn in green (dotted-dashed) and cyan (dashed), respectively, the exponential function in red.

The data is described well by the fit models in both cases and thus the sWeights calculated by the $_s\mathcal{P}lot$ technique (see Figure 22) can be used in the further signal selection. The comparison between the widths of the B mass without and the one with constrained π^0 mass already shows the narrowing of the peak due to the constraint.
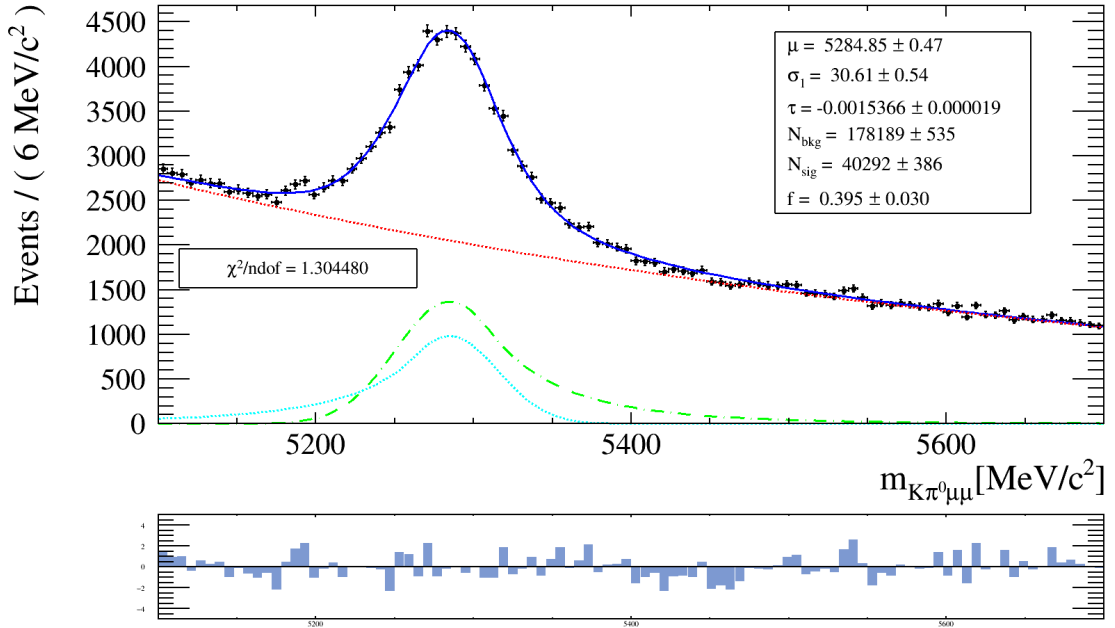
Figure 21: Fit of the B mass spectrum with constrained π^0 mass for the normalisation channel with a double CB function and an exponential. The two CB components are drawn in green (dotted-dashed) and cyan (dashed), respectively, the exponential function in red.
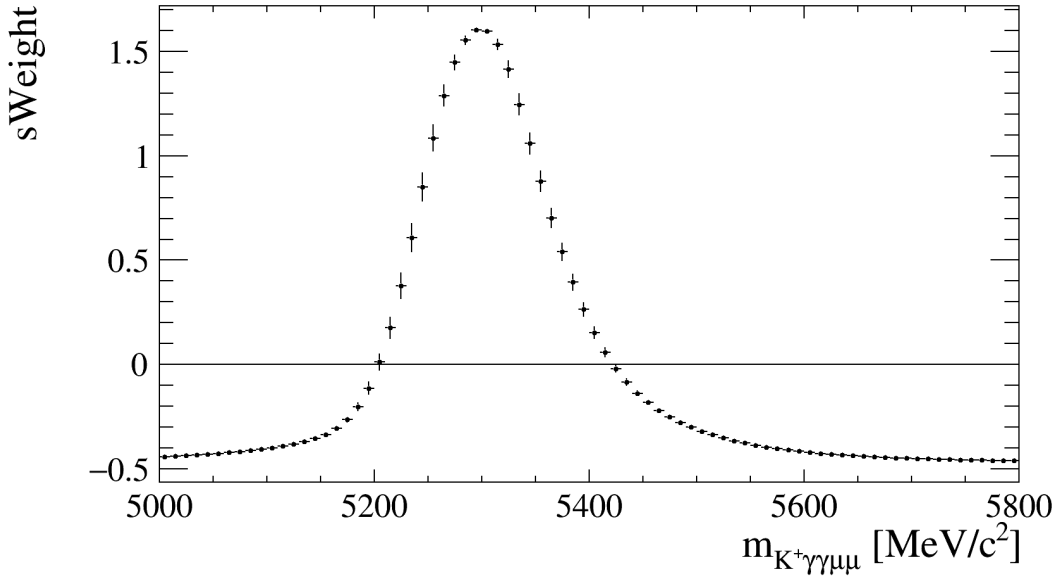
Figure 22: Profile of sWeights calculated from Figure 20 versus the invariant B^+ mass without π^0 mass constraint.

5.3.2 Differences between data and simulated samples

At this stage, the agreement of the distributions of control variables on data and on simulation can be checked because s-weighting the control variables should now reflect a pure signal. It is known that differences between data and simulation can occur particularly for photons, which are reconstructed in the calorimeter system. They are caused by various factors like different photon conversion[22] probability due to incompletely known distribution of material in the LHCb detector, non-perfect calibration of ECAL, PS and SPD detectors, and the noise in the calorimeter system. However, Figure 23c, 23d and 45b, 45c (appendix) show non-serious deviations between photon-associated variables. Figure 23a on the contrary shows a large difference between simulation and s-weighted data in the transverse momentum of the B. This phenomenon is known from other LHCb analyses and it is uncomplicated to correct for the difference according to experience. The differences in other variables are relatively small. Because the multivariate classifier for the signal selection is trained on data only, these differences do not affect it. However, they will affect the efficiency determination that is conducted on simulation and are thus considered later on. It is also made sure that the signal of the normalisation channel is a good proxy for $B^+ \to K^{*+}\mu^+\mu^-$ by plotting the s-weighted data obtained from $B^+ \to J/\psi K^{*+}$ candidates together with the simulated sample of $B^+ \to K^{*+}\mu^+\mu^-$. These plots can be found in the appendix, Figure 44.

Another possible reason for discrepancies between data and simulation is the presence of an S-wave[23] contribution to the $K^+\pi^0$ system. This contribution comes from non-resonant $K^+\pi^0$ or $K_0^*(1430)$, which has a full width of $\Gamma = 270 \pm 80$ MeV and can therefore reach into the $K^*(892)$ spectrum where it is almost flat. These contributions are excluded in simulation by truthmatching. However, it is difficult to quantify the contribution to data. To minimise the possible impact, a cut on the $K^*(892)$ is applied in the preselection (see Table 7). Furthermore, the difference between the $K^+\pi^0$ invariant mass on simulation and data is investigated without the K^*-cut (see appendix Figure 46). No significant contribution is seen which conforms to the findings in [12] where the S-wave contribution is deemed negligible.

---

[22]Conversion means pair production $\gamma \to e^+ e^-$.

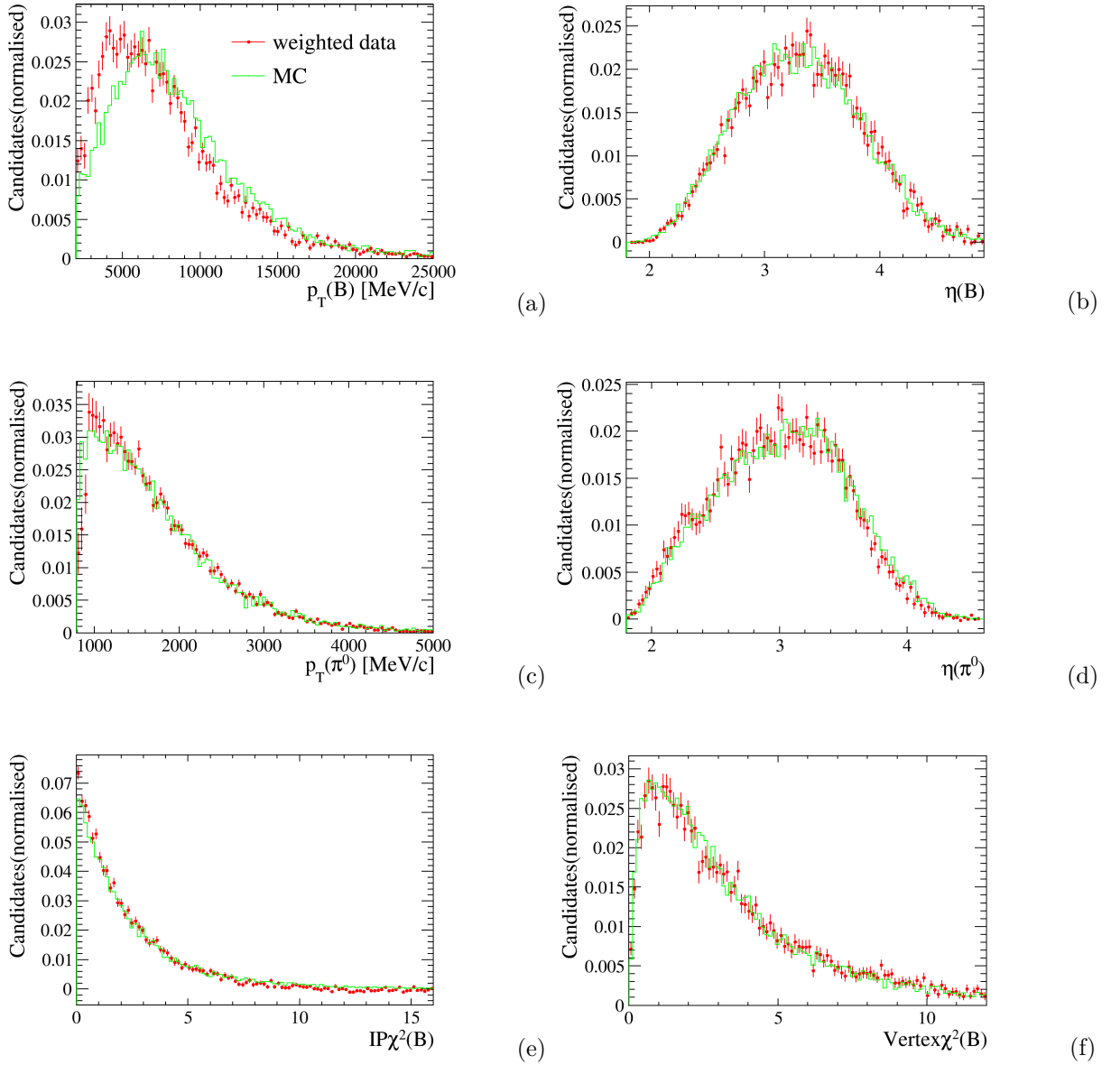[23]$K^+\pi^0$ contribution with zero angular momentum.

Figure 23: Distributions of $p_T(B)$, $\eta(B)$, $p_T(\pi^0)$, $\eta(\pi^0)$, $IP\chi^2(B)$ and $Vertex\chi^2(B)$ from simulation (green) and data weighted with sWeights (red, dotted) from $B^+ \to J/\psi K^{*+}$.

5.3.3 Charmonia resonances

The spectrum in Figure 18 shows the two charmonia states, $J/\psi(1S)$ and $\psi(2S)$, that are the main contribution to the resonant version of the decay $B^+ \to K^{*+}\mu^+\mu^-$. The branching fraction of the non-resonant decay is around two orders of magnitude smaller than the one of the resonant decay with $J/\psi \to \mu\mu$. Hence, it is crucial to sort out events corresponding to the resonant decay. This is done by cutting on the $\mu\mu$ invariant mass spectrum. For the rejection of $\psi(2S)$ candidates, the mass region $m_{\mu\mu} \in [3536, 3873]$ MeV/c^2 is removed, which is a common veto on this charmonium state within the LHCb collaboration [12]. The $J/\psi(1S)$ candidates are rejected by removing $m_{\mu\mu} \in [2780, 3250]$ MeV/c^2, which eliminates 99.7% of these candidates according to simulation (see Figure 24). The remaining $J/\psi(1S)$ candidates at lower or higher di-muon masses end up in lower and higher B mass regions, respectively, so that they are not considered as signal.

Figure 25 shows the corresponding 2D plot with data where both resonances are visible as horizontal bands along the mass of B^+ candidates. The black lines show the regions that are removed.
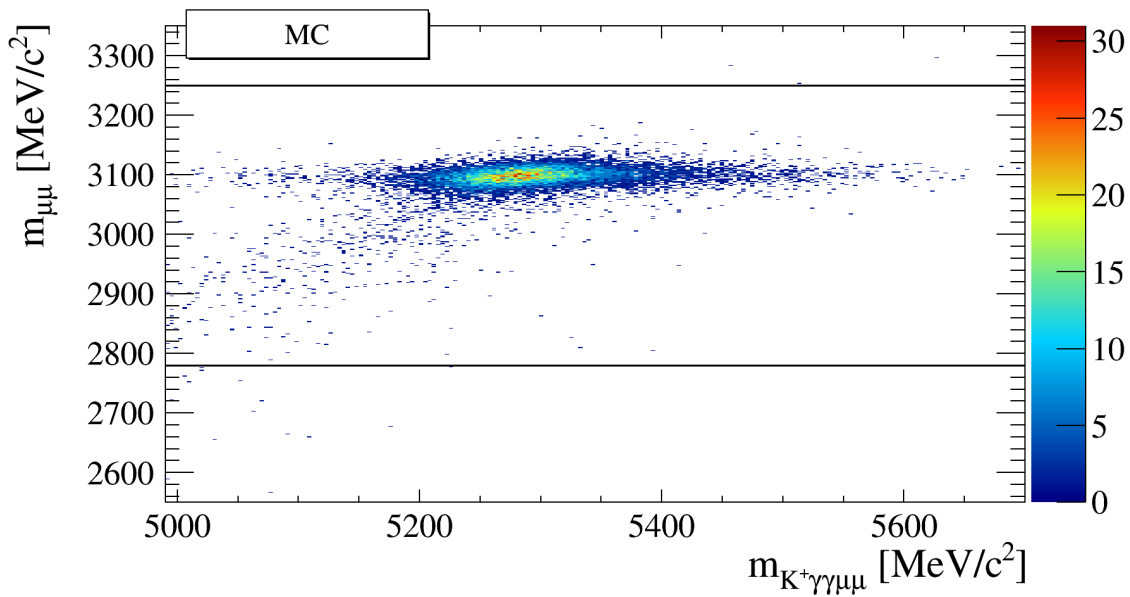
Figure 24: Mass of the di-muon system versus the mass of the $B^+ \to J/\psi(\to \mu\mu)K^{*+}$ candidates from simulation. The lines show the boundaries of the regions which are removed.

With these vetos, a large of amount of combinatorial background is also removed and obviously also a certain amount of signal events of $B^+ \to K^{*+}\mu^+\mu^-$.
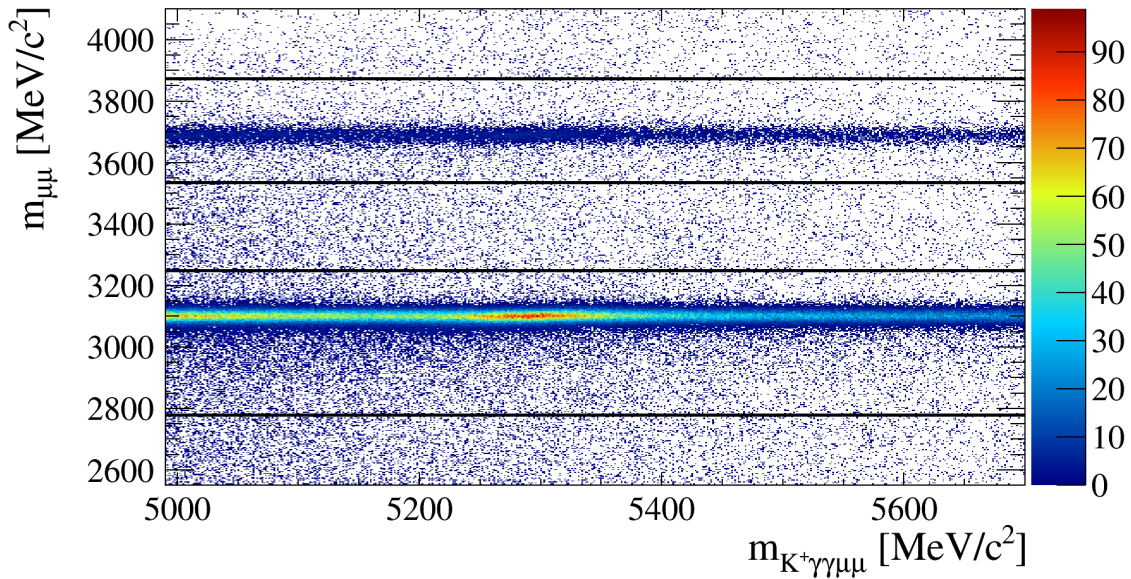
Figure 25: Mass of the di-muon system versus the mass of the B^+ candidates. Only the di-muon mass region close to the $J/\psi(1S)$ and $\psi(2S)$ masses is shown. The black lines show the boundaries of the regions which are removed.

5.3.4 Pure background sample

The BDT needs, as already mentioned, a pure background sample besides the pure signal sample to distinguish between these two sources of events. With the assumption that sources of peaking background[24] are negligible, every event that is not characterised as signal is characterised as combinatorial background. According to simulation, 98% of the signal for $B^+ \to K^{*+}\mu^+\mu^-$ reside in the mass region $m_{K^+\gamma\gamma\mu\mu} \in [5100, 5700]$ MeV/c^2. Therefore, the regions $m_{K^+\gamma\gamma\mu\mu} \in [5000, 5100]$ MeV/c^2 (called lower sideband) and $m_{K^+\gamma\gamma\mu\mu} \in [5700, 7000(6800 \text{ for } \pi^0 \text{ mass constraint})]$ MeV/c^2 (called upper sideband) are taken as samples for the combinatorial background. Figure 26a shows the lower and upper sideband for the B mass without π^0 mass constraint, Figure 26b for the B mass with constrained π^0 mass. The charmonia vetos are applied in these plots.

To ensure that the distributions of the variables, which are exploited in the BDT in Section 5.3.5, are comparable for the lower and upper sideband, the profiles of these variables in dependence of the B mass are investigated. Figure 27 shows the profile of the kaon transverse momentum, which is found to have the strongest but still negligible correlation to the B mass.

(a)   (b)
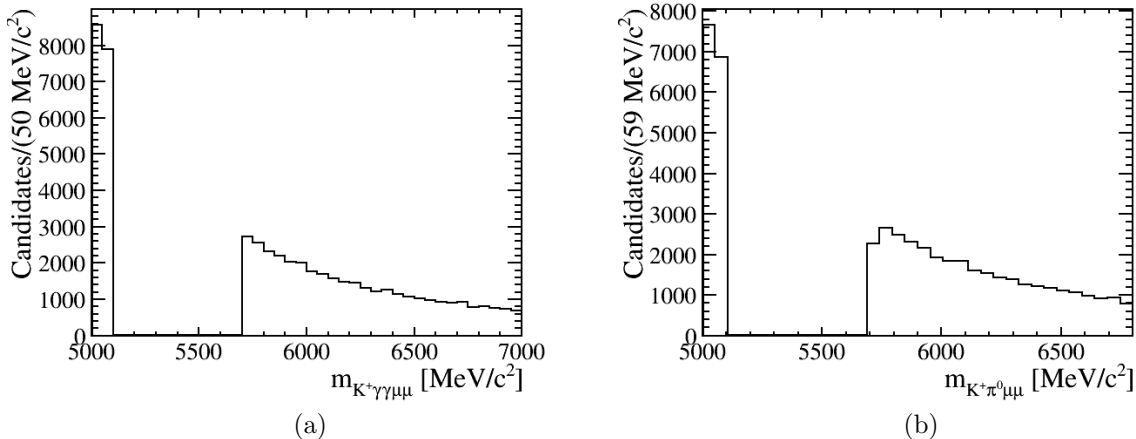
Figure 26: Upper and lower sideband of $B^+ \to K^{*+}\mu^+\mu^-$ candidates after charmonium vetos without (a) and with (b) π^0 mass constraint.
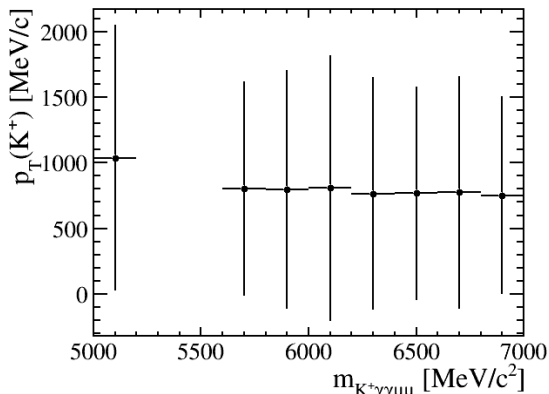
Figure 27: Profile of the mean (points) and RMS (error bars) of $p_T(K^+)$ in bins of the B mass.

---

[24]Peaking backgrounds can for example occur when decays are only reconstructed partially, or when particles from other decay channels are misidentified so that they appear to have the same final state.

5.3.5 Multivariate analysis

To reduce the background but keep as many signal candidates as possible is the aim of the signal selection. The suppression of background is best achieved by applying cuts on variables, which have well separated distributions for background and signal. To study these variables, the combinatorial background is taken from the lower and upper sidebands as described in Section 5.3.4, the signal sample is taken from the normalisation channel as described in Section 5.3.1. Several variables are tested for their usefulness in the analysis but only those with sufficient background-discriminating power are kept. Some of the well-discriminating variables are used already in the preselection and their utilisation is now justified *a posteriori* together with the other variables with background-discriminating power. The definitions of all variables are found in Section 5.2.1.

Since the kinematics of the decay is well known, variables containing momentum components often offer differences between signal and background. The transverse momentum is usually large in the decays of heavy particles like the B meson. Indeed, Figures 28a and 28b and 29a show that events from the sidebands tend to have low transverse momenta. In Figure 29a the higher transverse momentum of both photons is taken into account. It is useful to combine variables of the same kind for particles that are produced in the same decay step in order to minimise the total number of utilised variables. The logarithm is taken in order to smooth the distribution. The pseudorapidity of the B meson tends to be lower for signal events than for background events (see Figure 30a) which is due to the high particle flux density in regions of high pseudorapidity. It is then more likely to combine the final-state particles randomly. In Figure 30b the difference between the pseudorapidities of the K^+ and the π^0 ($|\eta(\pi^0) - \eta(K^+)|$) is shown. If these two particles are combined randomly, there is no correlation between their pseudorapidities and the difference is supposed to be rather large. Figure 29b shows the asymmetry of the transverse momenta of the B^+ candidates and all other tracks inside a cone ($A_{p_T}^{Cone}$). If there are more particle tracks inside the cone than the one of the B^+ candidate, the value of $A_{p_T}^{Cone}$ becomes smaller. With more tracks close to the B candidate, it is more likely to be combinatorial background. The distributions for the $DIRA$ of the B^+ candidate and the opening angle of the photons are found in Figure 32. These two variables also contribute to the separation because the angle between photons coming from a π^0 with high momentum is expected to be small as well as the angle between the momentum of the B^+ candidate and its flight direction, if it is reconstructed well.

Beside kinematic variables, also the vertex and track quality and the impact parameters play an important role in the rejection of combinatorial background. The B^+ candidate has to come from a PV and is therefore expected to have a small IP and especially an $IP\chi^2$ that is near unity (see Figure 31a). On the contrary, the interesting muons must come from a detached vertex and thus have a high IP and $IP\chi^2$, respectively. By taking into account the minimum of the $IP\chi^2$ of a muon pair (see Figure 31b), a more effective selection for the di-muon system of the decay is possible. The distributions for used variables concerning track and vertex quality are found in Figure 33, as well as the distributions for the CL of the photons.

After the identification of variables that accommodate a good separation of signal and combinatorial background, a naive approach would be to simply select the signal events by individual one-dimensional[25] cuts. However, for the selection of a rare decay, hard cuts need to be applied and because of correlations between the variables these cuts are less effective. It is then better to use a multivariate analysis technique - here the BDT from the Toolkit for Multivariate Data Analysis (TMVA)[28].

---

[25]A one-dimensional cut means that it takes into account only one dimension of the n-dimensional space that is spanned by n variables. A cut or a variable is orthogonal in this sense when the cut does not affect the other variables.
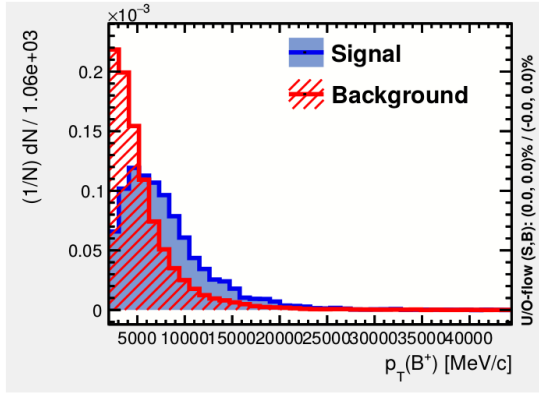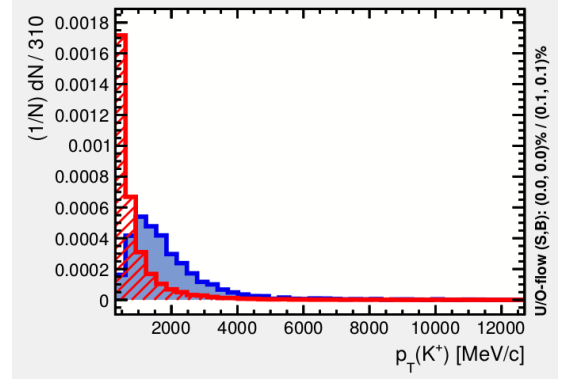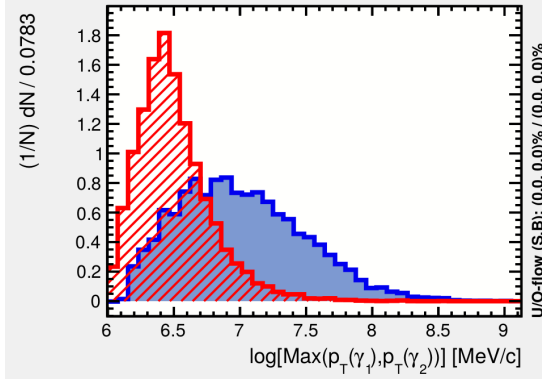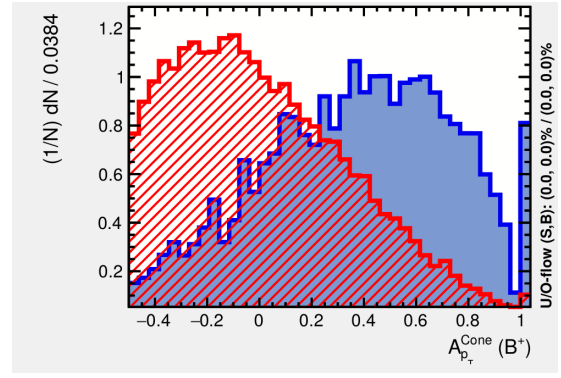
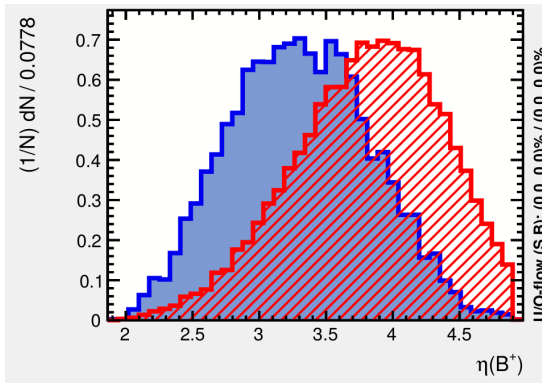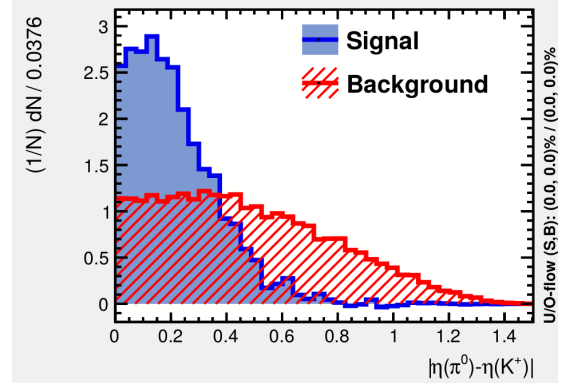**Figure 28:** Signal (blue) and background (red) distributions for $p_T(B^+)$ (a) and $p_T(K^+)$ (b).

Figure 29: Signal (blue) and background (red) distributions for $\log[Max(p_T(\gamma_1), p_T(\gamma_2))]$ (a) and $A_{p_T}^{Cone}$ (b).
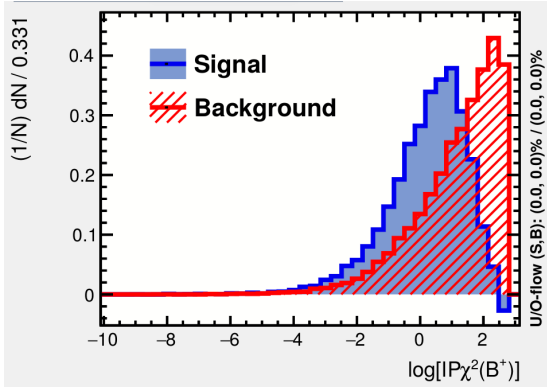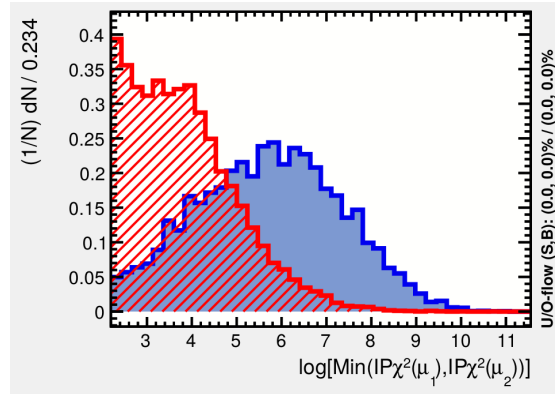

Figure 30: Signal (blue) and background (red) distributions for $\eta(B^+)$ (a) and $|\eta(\pi^0) - \eta(K^+)|$ (b).

(a)

(b)
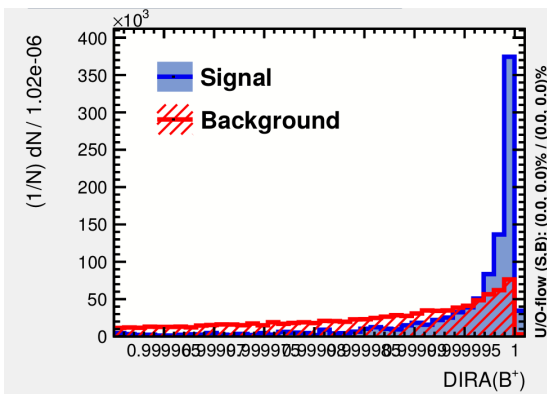
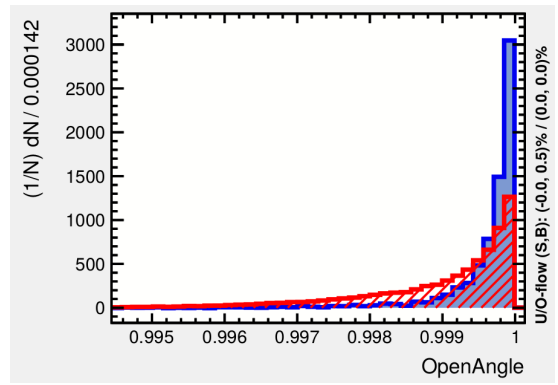**Figure 31:** Signal (blue) and background (red) distributions for $\log[IP\chi^2(B^+)]$ (a) and $\log[Min(IP\chi^2(\mu^+), IP\chi^2(\mu^-))]$ (b). In (a), the bin with negative content comes from a statistical fluctuation due to the sWeights.
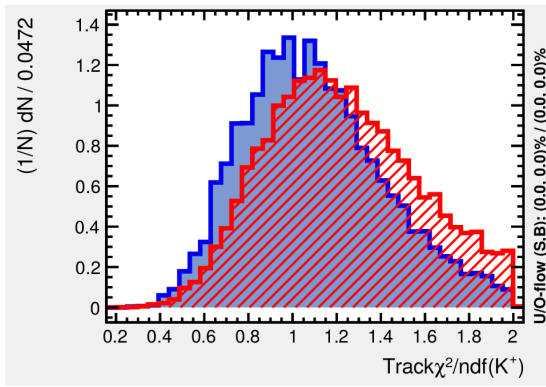

(a)

(b)

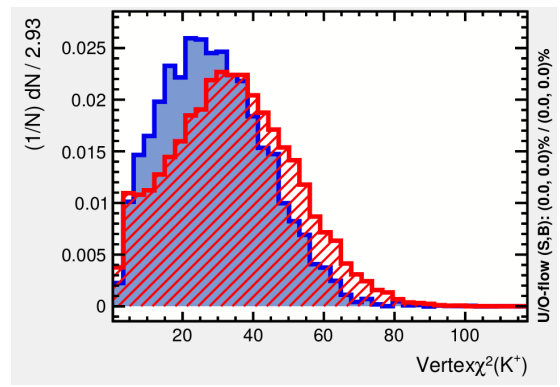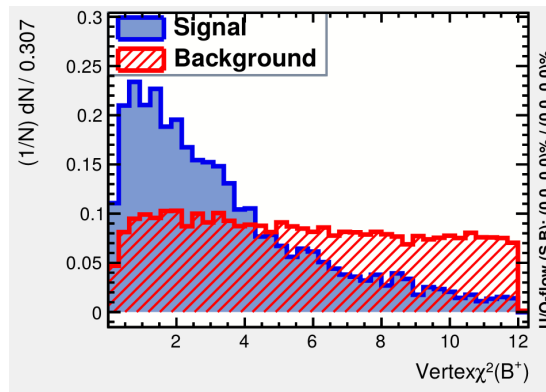**Figure 32:** Signal (blue) and background (red) distributions for $DIRA$ (a) and $OpenAngle$ (b).
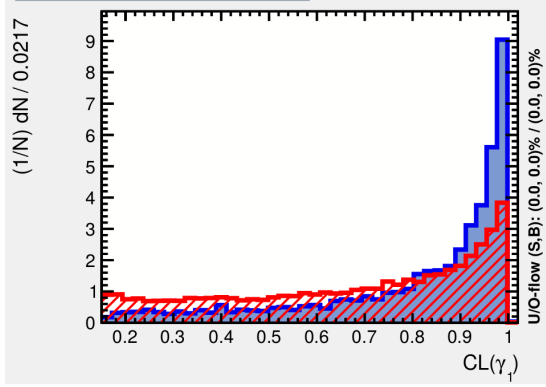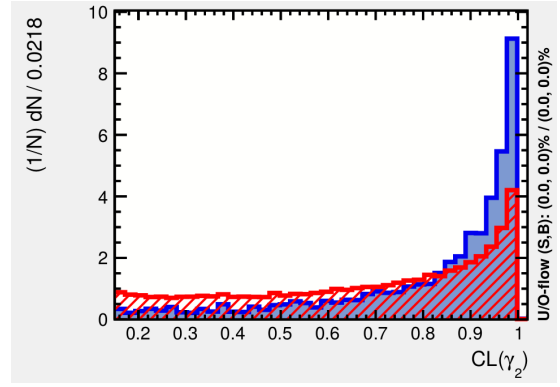
Figure 33: Signal (blue) and background (red) distributions for $CL(\gamma_1)$ **(a)**, $CL(\gamma_2)$ **(b)**, $Track\chi^2/ndf(K^+)$ **(c)**, $Vertex\chi^2(K^+)$ **(d)** and $Vertex\chi^2(B^+)$ **(e)**.

It is important to note that variables used for the signal selection have to fulfill an additional criterion. They have to be orthogonal to the mass of the B meson because otherwise it is possible to raise an artificial peak in the B mass by cutting on correlated variables. Therefore, the selection variables were all chosen to have negligible correlations to the B mass.

The variables described above and listed in Table 10 are then used by TMVA to determine signal-like and background-like events.

B^+	K^+	μ	γ
p_T	p_T	$\log[Min(IP\chi^2(\mu^+), IP\chi^2(\mu^-))]$	CL_1
$DIRA$	$Track\chi^2/ndf$		CL_2
$\log[IP\chi^2]$	$Vertex\chi^2$		$\log[Max(p_T(\gamma_1), p_T(\gamma_2))]$
$Vertex\chi^2$	$|\eta(\pi^0) - \eta(K^+)|$		$OpenAngle$
$A_{p_T}^{Cone}$			
η			

Table 10: Variables used in the multivariate analysis.

The algorithm that is used to classify the events is the *boosted decision tree*. A decision tree is a binary tree (see Figure 34) that is built on binary decisions at each node. The tree is build until a certain stop criterion is fulfilled. The decisions are made by considering one variable of the those mentioned above per node. Events that run through the tree pass several nodes and are classified as more background-like or more signal-like by the best variable at each one of them. If the stop criterion is fulfilled, e.g. a certain size of the tree is reached or all events have been classified unambiguously, every event resides in a final *leaf* node. Depending on the majority of events of one kind in the leaf, the leaf itself is eventually classified as background or signal, respectively, splitting the phase space of all variables into many regions according to this classification. These regions are hypercubes in phase space. With simple one-dimensional cuts only one hypercube is selected. The process that defines the splitting criteria for each node of the tree is the so-called *training* of the classifier.

A decision tree has the disadvantage that it is instable with respect to statistical fluctuations in the training sample which determines the tree structure. Assuming that two variables have a similar separation power (e.g. $p_T(\gamma_1)$ and $p_T(\gamma_2)$), a statistical fluctuation may cause the algorithm to select one variable, while the other one could have been chosen without the fluctuation resulting in a different response of the classifier. The impact of statistical fluctuations is attenuated by growing a forest of decision trees whereby an event is classified on a majority vote of the classification done by each tree in the forest. All trees in the forest are derived from the same training sample and the events that run through them are subsequently treated with a so-called *boosting*. After passing through one tree, the events in the leaves that were misclassified have their weights modified (boosted) before they are given to the next tree. The boosting algorithm used in this analysis is called Adaptive Boost (AdaBoost) which gives misclassified events a higher weight in the training of the following tree. The event weight depends on the misclassification rate of the previous tree. The gain of statistical stability from the boosting also increases the performance of the classifier at the cost that the single decision tree no longer allows a straightforward interpretation of the cuts.

It is interesting to think about what happens when events have negative sWeights as it is the case in this thesis. These events tend to receive increasingly stronger boosts because the separation gain is lowered by the boost instead of raised. It is as if background events were selected as signal and vice versa. The events with negative sWeight are therefore ignored in the training. However, this is not the best approach to handle negative weights. The best way is to pair every negative weight with the positive weight
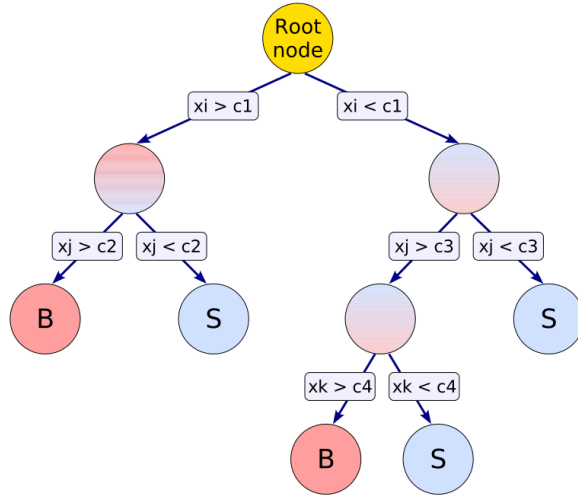
Figure 34: Schematic view of a decision tree. At each node a decision is made with respect to the discriminating variable x_i that gives the best separation between signal and background at this node. Variables can be used several times or not at all. The leaf nodes are denoted 'S' if the majority of events in this leaf are signal-like and 'B' if background-like, respectively.

that is located closest in phase space, so that they 'annihilate' each other [29]. Unfortunately, the time is not sufficient to investigate this approach, so the events with negative sWeight are ignored in the training.

Every event in the dataset runs through the decision forest and obtains a +1 if it is classified as signal-like or a −1 if it is classified as background-like. Eventually, the sum of these numbers is normalised to the number of trees in the forest (850 are used in this analysis). The BDT then gives a continuous response between -1 and +1 for all events where -1 means very background-like and +1 very signal-like. The result is a new variable called 'BDT response'. A single cut on this variable is comparable with a large set of complex cuts on all variables the BDT is provided with, taking into account correlations between the variables.

From the samples provided to the BDT (signal and background sample), half of the events - randomly selected - are used for training of the BDT and the other half for testing. By comparing the performance results between these two[26] statistically independent samples a possible overtraining can be detected. An overtraining occurs when the BDT algorithm has too many parameters (i.e. number of nodes) in comparison to the number of data points. The performance of the classifier is then overestimated in the training sample because of statistical fluctuations which presents itself in a worse performance on the test sample.

With the variables in Table 10 and the data samples from Sections 5.3.1 and 5.3.4 the BDT gives the responses shown in Figure 35a and 35b for the B^+ mass without and B^+ mass with π^0 mass constraint, respectively.

---

[26]Proper training and validation requires three statistically independent data sets. One for parameter optimisation, another for overtraining detection and the last for performance validation. The latter two are merged in order to increase statistics. The resulting bias is insignificant [28].
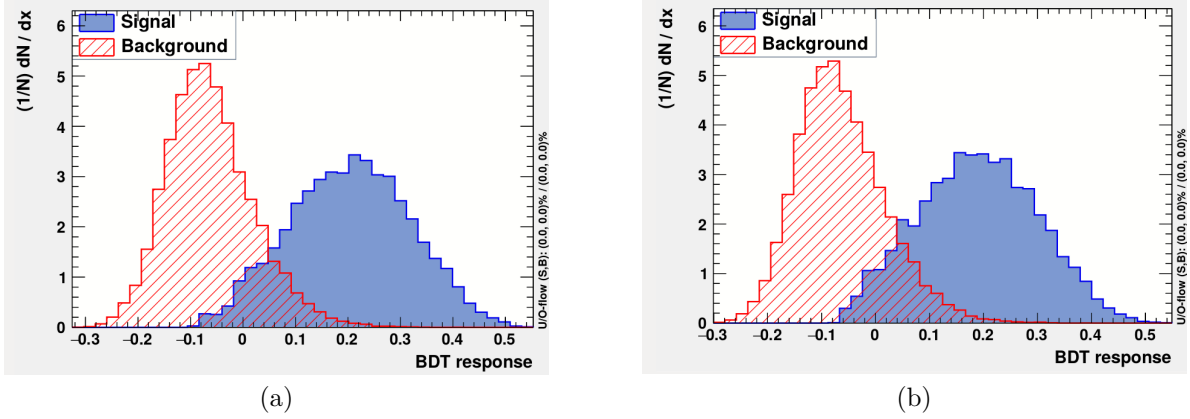
(a)   (b)

Figure 35: Result of the BDT training for B mass without (a) and B mass with constrained π^0 mass (b) evaluated on the test sample with events from the background sample in red and events from the signal sample in blue.

The correlation of the BDT response with the B^+ mass is expected to be negligible because the training variables show negligible correlation to the B^+. Figure 36 shows the profile histogram of the mean BDT response in bins of $m_{K^+\gamma\gamma\mu\mu}$ in the region where sWeights are applied. A small but negligible correlation is apparent. The signal region was removed due to the obvious correlation in this region.
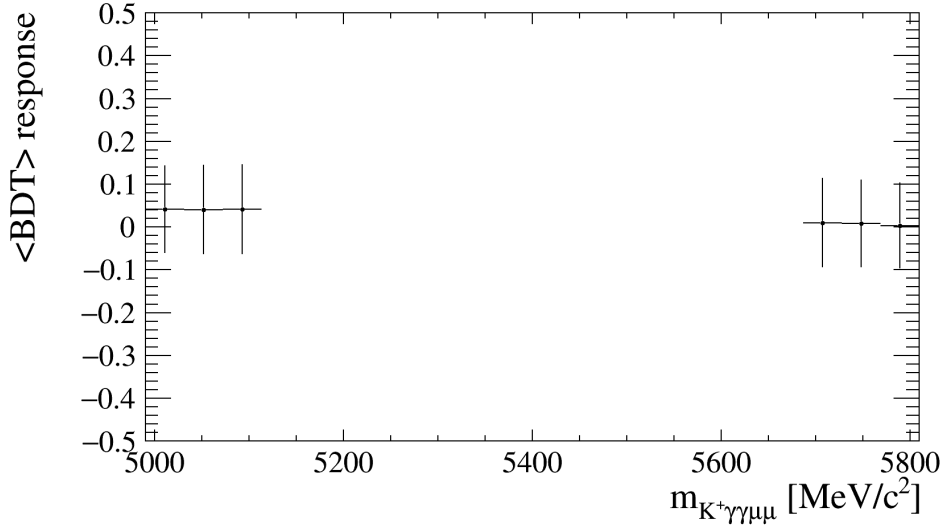

Figure 36: Profile of the mean (points) and RMS (error bars) of the BDT response in bins of $m_{K^+\gamma\gamma\mu\mu}$

5.3.6 Optimal BDT cut

The BDT response is now the variable with the most effective suppression of combinatorial background. However, not every cut on this variable is as efficient for the signal selection as the other. In Figure 35, a cut at -0.1 would maintain all of the signal but also rejects less than half of the combinatorial background, while a cut at 0.3 would eliminate almost the whole combinatorial background at the cost of more than half of the signal. The optimal value for the BDT cut lies somewhere between and is determined in the following.

The BDT cut is optimised using the normalisation channel and the sidebands in the B^+ mass. It is the goal to maximise a so-called figure of merit which is a measure for the significance of the resulting signal. For each BDT cut, the signal yield in the window $m_{K^+\gamma\gamma\mu\mu} \in [5100, 5700]$ is estimated by fitting $B^+ \rightarrow J/\psi K^{*+}$ events. To obtain the expected yield of $B^+ \rightarrow K^{*+}\mu^+\mu^-$, the yield of the

48

normalisation channel has to be scaled by the ratio of the total selection efficiency (ϵ^{MC}) - apart from the BDT cut - obtained from simulation, and the ratio of branching fractions between $B^+ \to K^{*+}\mu^+\mu^-$ and $B^+ \to J/\psi K^{*+}$ taken from [**PDG**]:

$$nSig_{\mu\mu}^{BDT} = nSig_{J/\psi}^{BDT} \times \frac{\epsilon_{\mu\mu}^{MC}}{\epsilon_{J/\psi}^{MC}} \times \frac{\mathcal{B}(B^+ \to K^{*+}\mu^+\mu^-)}{\mathcal{B}(B^+ \to J/\psi(\to \mu\mu)K^{*+})}.$$

The efficiencies are roughly estimated at this stage. The exact determination is explained in Section 6.1. The background yield ($nBkg_{\mu\mu}$) is estimated by fitting the upper and lower sideband and extrapolating the yield into the signal region ($m_{K^+\gamma\gamma\mu\mu} \in [5130, 5600]$). For the fits to the sidebands a simple exponential is used and for the normalisation channel a double CB with fixed power-law parameters from simulation (see Table 9) is fitted. An example of these fits is shown in the appendix in Figure 47. The chosen figure of merit (FoM) is

$$FoM = \frac{nSig_{\mu\mu}}{\sqrt{nSig_{\mu\mu} + nBkg_{\mu\mu}}}.$$

Figure 37 shows this parameter as a function of the BDT cut. The large error of the FoM results mainly from the fit errors.
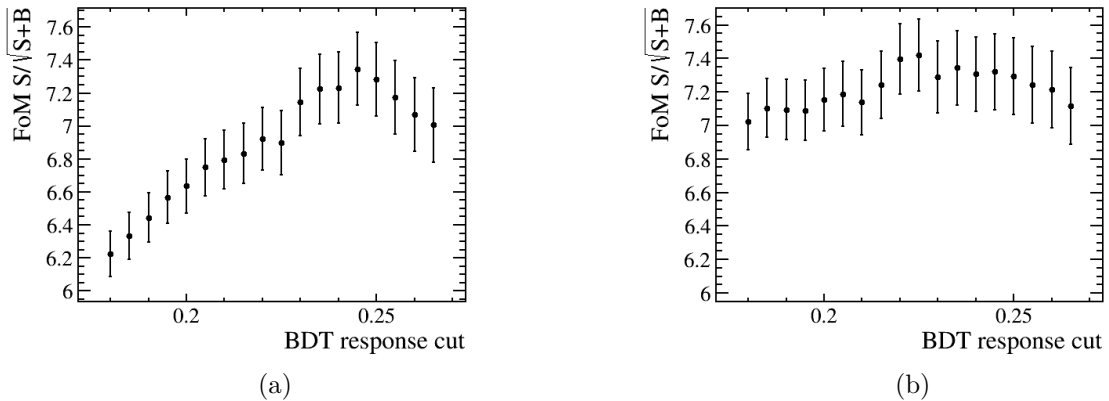

Figure 37: Significance as a function of the cut on the BDT variable for B^+ mass without (a) and for B^+ mass with π^0 mass constraint (b). The FoM values are highly correlated for each BDT cut.

Both distributions have a local maximum that indicates the optimal BDT cut value. However, the uncertainty of the efficiency is large and a multitude of cuts show comparable values for the FoM. The incomplete knowledge of the efficiency leads an incomplete knowledge of the position of the maximum. Since the significance shows a steeper decrease for harder BDT cuts than for weaker ones, a conservative cut at the low end of the maximum is applied. For the B^+ mass without π^0 mass constraint this cut value is 0.23 and for the B^+ mass with constrained π^0 mass the cut value 0.22 is chosen.

5.3.7 Result of the signal selection

After applying the BDT cuts from Section 5.3.6, the signal yields of $B^+ \to J/\psi K^{*+}$ and $B^+ \to K^{*+}\mu^+\mu^-$ are obtained by fitting the B^+ mass distribution in the range $m_{K^+\gamma\gamma\mu\mu} \in [5100, 5700]$. The fit model is composed of a double CB as signal model and an exponential as background model. The fit parameters for $B^+ \to K^{*+}\mu^+\mu^-$ are fixed to the values of the fit parameters for $B^+ \to J/\psi K^{*+}$ except for the yields and the slope of the exponential. The power-law parameters are fixed on simulation (see Table 9).

The statistical significance of the signal peak is determined using Wilk's theorem and is calculated by

$$S = \sqrt{2 \cdot Min(\log[\mathcal{L}_{bkg}]) - 2 \cdot Min(\log[\mathcal{L}_{sig+bkg}])}$$

where $Min(\log[\mathcal{L}_{bkg}])$ denotes the minimum of the negative log-likelihood value (cf. Section 5.1.1) of a fit with the background model only and $Min(\log[\mathcal{L}_{sig+bkg}])$ the minimum of the negative log-likelihood value of a fit with the full model.

Figures 38 - 41 show the final fit results after the BDT cut. The signal yield corresponds to the fit parameter $nSig$ which values are summarised in Table 11.

As expected, the significance of the signal peak in the B^+ mass distribution is larger than the one of the signal peak in the mass distribution without this constraint. The signal yields do not deviate significantly. Therefore, only the signal yields of the B^+ mass fit with π^0 mass constraint are considered in the following sections.

Channel	$nSig$
$B^+ \to K^{*+}\mu^+\mu^-$	80 ± 16
m_{π^0} constr.	81 ± 16
$B^+ \to J/\psi K^{*+}$	15235 ± 208
m_{π^0} constr.	15846 ± 196

Table 11: Signal yields of fits to $B^+ \to K^{*+}\mu^+\mu^-$ and $B^+ \to J/\psi K^{*+}$. The uncertainty is statistical only.
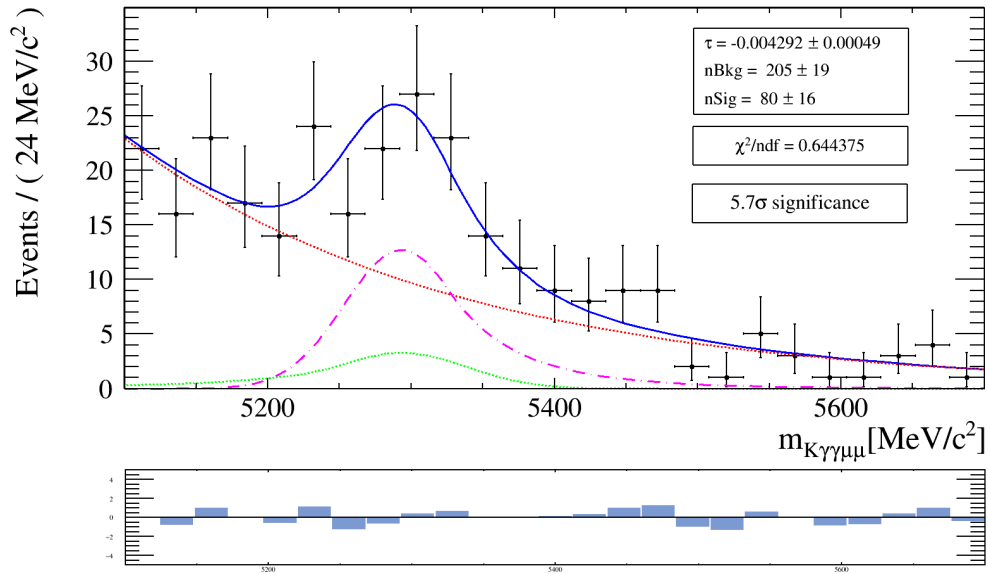
Figure 38: B^+ mass spectrum after BDT cut at 0.23 and charmonia vetos. The fit function is a double CB with an exponential. All fit parameters are fixed by the fit in Figure 39, except for the yields and the exponential slope. The components of the fit model are drawn as dashed lines: CB$_1$ green, CB$_2$ magenta (dotted-dashed) and exponential red.
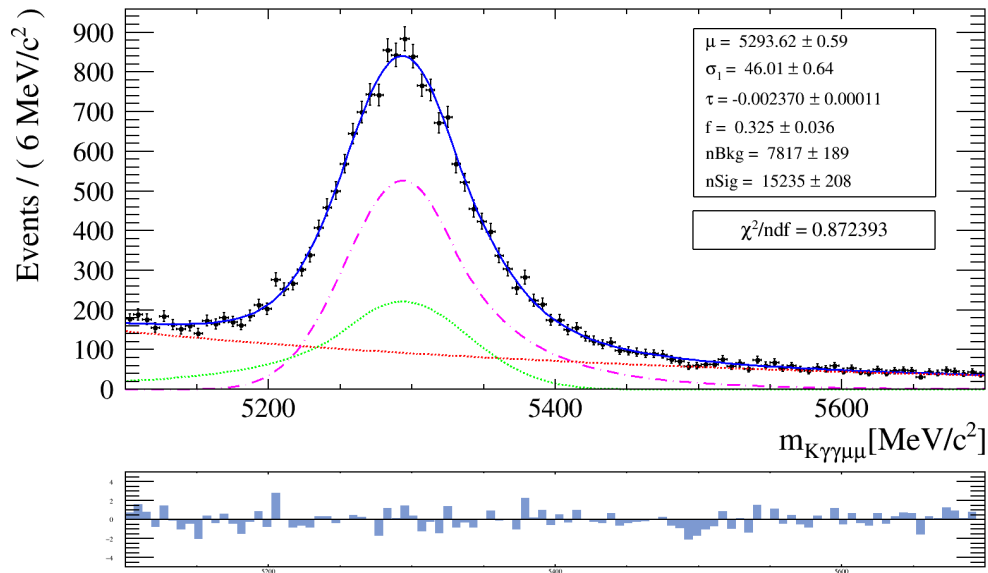

Figure 39: B^+ mass spectrum after BDT cut at 0.23 and inverse $J/\psi(1S)$ veto. The fit function is a double CB with an exponential. Power-law parameters are found in Table 9. The components of the fit model are drawn as dashed lines: CB$_1$ green, CB$_2$ magenta (dotted-dashed) and exponential red.
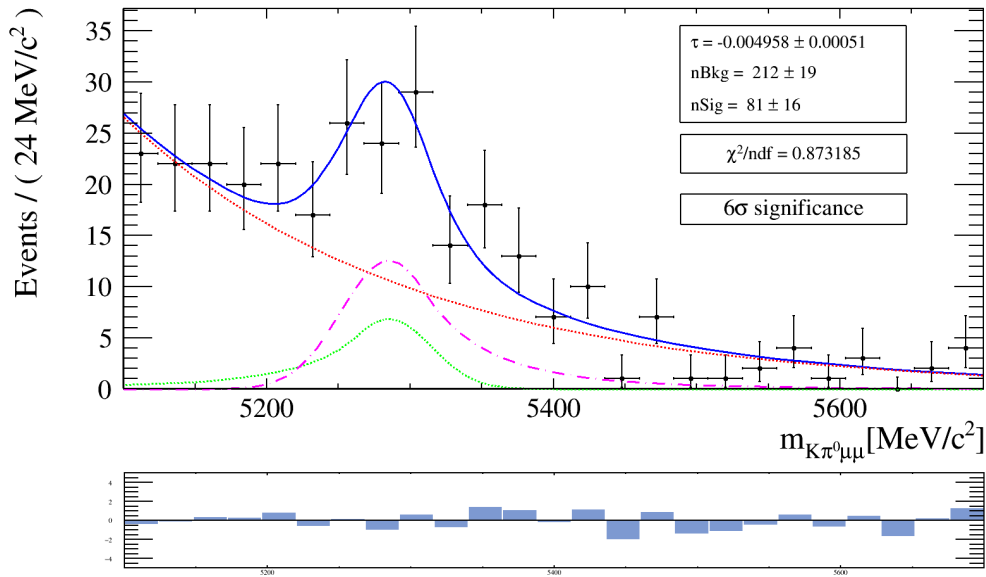
Figure 40: B^+ mass spectrum, with constrained π^0 mass, after BDT cut at 0.23 and charmonia vetos. All fit parameters are fixed by the fit in Figure 41, except for the yields and the exponential slope. The components of the fit model are drawn as dashed lines: CB_1 green, CB_2 magenta (dotted-dashed) and exponential red.
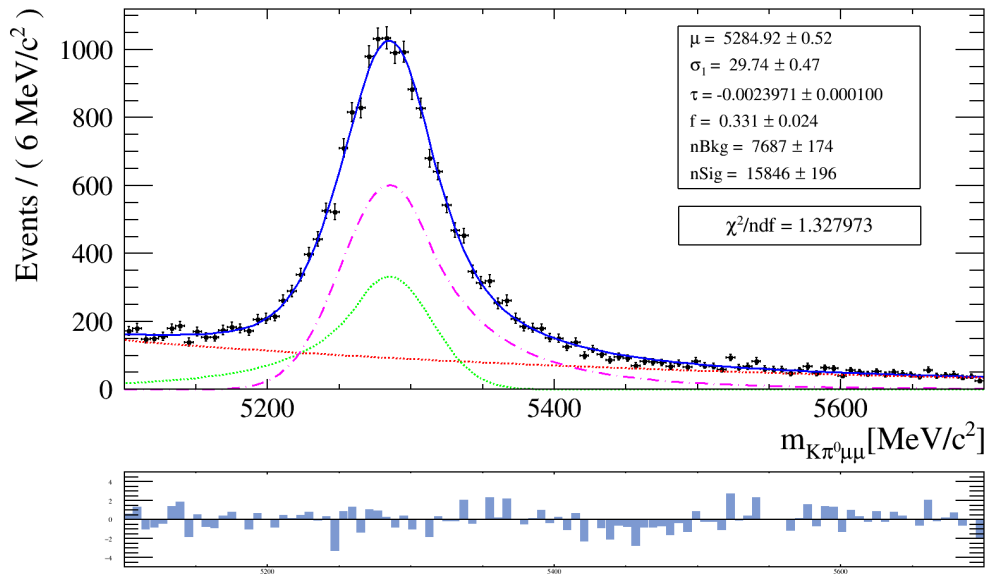

Figure 41: B^+ mass spectrum, with constrained π^0 mass, after BDT cut at 0.23 and inverse $J/\psi(1S)$ veto. The fit function is a double CB with an exponential. Power-law parameters are found in Table 9. The components of the fit model are drawn as dashed lines: CB_1 green, CB_2 magenta (dotted-dashed) and exponential red.

6 Determination of the branching fraction

This section describes the calculation of the branching fraction of the decay $B^+ \to K^{*+}\mu^+\mu^-$ relative to the decay $B^+ \to J/\psi(\to \mu^+\mu^-)K^{*+}$ using the signal yields obtained in Section 5.3. However, these yields only reflect the number of signal events which pass a certain selection. The according selection efficiency is determined first. Then the total branching fraction of the decay $B^+ \to K^{*+}\mu^+\mu^-$ is calculated using the known branching fraction of the normalisation channel.

6.1 Selection efficiencies

The selection efficiency specifies the number of signal candidates that are reconstructed and which pass the the signal selection relative to all present signal candidates. It is convenient to obtain this number from simulation because the number of simulated decays is known and it is simple to only consider signal events by truthmatching. The selection efficiency for both decay channels is then calculated by

$$\epsilon_{\mu\mu}^{MC} = \frac{N^{MC}(B^+ \to K^{*+}\mu\mu)}{N_{sim.}^{MC}(B^+ \to K^{*+}\mu\mu)}$$

and

$$\epsilon_{J/\psi}^{MC} = \frac{N^{MC}(B^+ \to J/\psi(\to \mu\mu)K^{*+})}{N_{sim.}^{MC}(B^+ \to J/\psi(\to \mu\mu)K^{*+})},$$

respectively. N^{MC} denotes the number of reconstructed signal candidates of the corresponding decay in simulation and $N_{sim.}^{MC}$ the corresponding number of simulated events.

The applicative determination of the selection efficiencies requires an overall good agreement between data and simulation. As is seen in Section 5.3.2, the distributions of variables in simulation and data do not match exactly. The difference is small in most variables but it is e.g. not negligible in $p_T(B^+)$. Therefore, the distribution of this variable is weighted to match the data distribution. The weight distribution (Figure 42) is obtained by dividing the normalised distributions of the simulated sample and s-weighted data (Figure 23a and 44a). After multiplying the simulated distribution with these weights, it matches the data distribution exactly in the chosen binning (Figure 43).
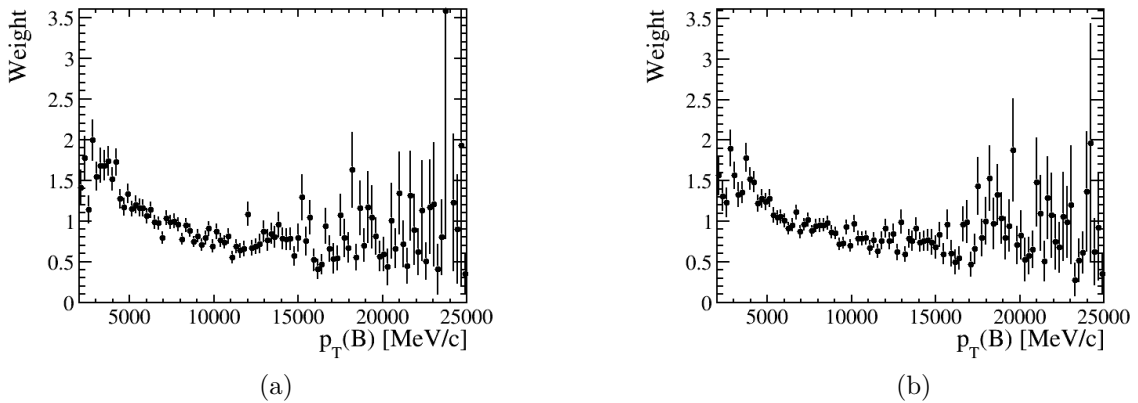

Figure 42: Weights for adaption of simulation to data. (a) shows the weights for adaption of the simulated $B^+ \to K^{*+}\mu^+\mu^-$ decay and (b) the weights for the simulated $B^+ \to J/\psi K^{*+}$ decay.

The weights are now applied to the truthmatched simulation sample after all cuts (i.e. trigger, stripping, preselection, vetos and BDT). It has to be remarked that the correction for differences between simulation and data in fact would have to be done before any cuts are applied, resulting in different efficiencies for each selection stage. This cannot be done offhand and therefore only a correction in the
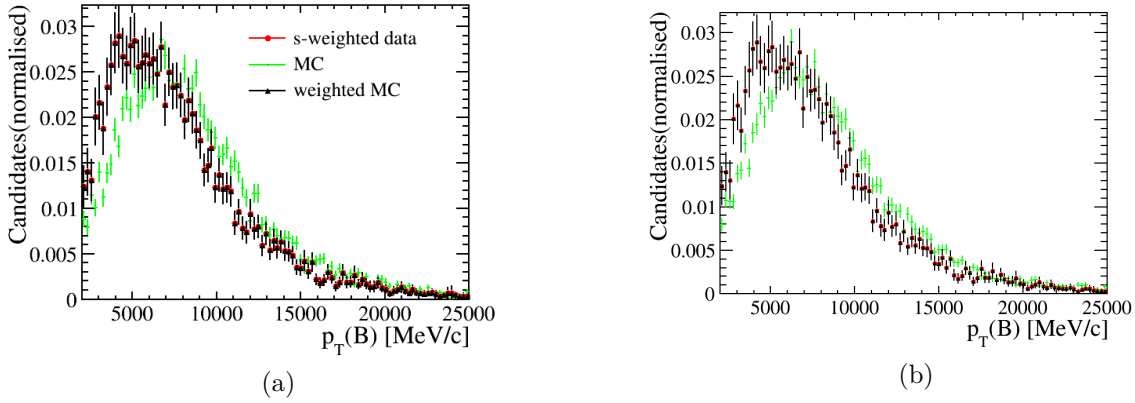
Figure 43: (a): Simulated distribution of $p_T(B^+)$ from $B^+ \to K^{*+}\mu^+\mu^-$ (green), s-weighted data (red) and weighted simulation (black). (b): Simulated distribution of $p_T(B^+)$ from $B^+ \to J/\psi K^{*+}$ (green), s-weighted data (red) and weighted simulation (black).

last selection step, i.e. for the BDT cut and the charmonia vetos, is taken into account and a systematic uncertainty is introduced to account for the remaining difference between simulation and data.

The number of remaining events after the selection and weighting is obtained by summing over the weights:

$$N' = \sum_{i=1}^{N} w_i$$

where N is number of events after the selection without weighting and w_i is the weight for event i.

The number of signal events that remains after the selection on simulation is listed in Table 12 together with the number of simulated MC events.

Decay channel	$N'^{MC}_{(sim.)}$
$B^+ \to K^{*+}\mu^+\mu^-$	2649 ± 64
simulated	1055086
$B^+ \to J/\psi K^{*+}$	5203 ± 101
simulated	1025698

Table 12: Number of events remaining after the selection for both channels and for the selection with constrained mass of the neutral pion. Also shown are the numbers of generated MC events.

Because of the weighting the standard Poissonian error is no longer valid for the number of events. The weighting can in general decrease or increase the number of events and thus change the relative statistical uncertainty without changing the real statistical power of the sample. This is avoided by normalising every weight with an effective weight w_{eff} [30]. The uncertainty of the number of events is then calculated by:

$$\Delta N' = N' \frac{\sqrt{w_{eff} \sum_i w_i}}{w_{eff} \sum_i w_i}$$

where

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

The selection efficiencies are determined with the results in Table 12:

$$\epsilon_{\mu\mu}^{MC} = (2.51 \pm 0.06) \times 10^{-3}$$
$$\epsilon_{J/\psi}^{MC} = (5.07 \pm 0.10) \times 10^{-3}.$$

6.2 Branching fraction results

With known selection efficiencies ϵ^{MC} the relative branching fraction of $B^+ \to K^{*+}\mu^+\mu^-$ and $B^+ \to J/\psi(\to \mu^+\mu^-)K^{*+}$ is given by

$$\frac{\mathcal{B}(B^+ \to K^{*+}\mu^+\mu^-)}{\mathcal{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^{*+})} = \frac{N(B^+ \to K^{*+}\mu^+\mu^-)}{N(B^+ \to J/\psi(\to \mu^+\mu^-)K^{*+})} \times \epsilon' \times \zeta',$$

where N is the signal yield from Table 11 and ϵ' is the relative efficiency

$$\epsilon' = \frac{\epsilon_{J/\psi}^{MC}}{\epsilon_{\mu\mu}^{MC}}.$$

$\mathcal{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^{*+})$ means that the branching fraction of $B^+ \to J/\psi K^{*+}$ is multiplied by the branching fraction of $J/\psi \to \mu\mu$.

ζ' is the relative geometrical acceptance of both decay channels in simulation. The acceptance takes into account that only a certain fraction of all decays end up in the detector acceptance. These numbers are known from the generation of the simulated sample:

$$\zeta' = \frac{\zeta_{B^+ \to J/\psi K^{*+}}}{\zeta_{B^+ \to K^{*+}\mu^+\mu^-}} = \frac{0.154}{0.1547}.$$

The relative branching fraction is then determined to be

$$\frac{\mathcal{B}(B^+ \to K^{*+}\mu^+\mu^-)}{\mathcal{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^{*+})} = (1.03 \pm 0.20_{stat.}) \times 10^{-2}.$$

The total branching fractions of $B^+ \to J/\psi K^{*+}$ and $J/\psi \to \mu\mu$ are well-known [**PDG**]:

$$\mathcal{B}(B^+ \to J/\psi K^{*+}) = (1.44 \pm 0.08) \times 10^{-3}$$
$$\mathcal{B}(J/\psi \to \mu\mu) = (5.961 \pm 0.033)\%$$

and are used eventually to calculate the total branching fraction of $B^+ \to K^{*+}\mu^+\mu^-$:

$$\mathcal{B}(B^+ \to K^{*+}\mu^+\mu^-) = (0.88 \pm 0.17_{stat.}) \times 10^{-6}.$$

7 Systematic uncertainties

A complete investigation of systematic uncertainties is out of the scope of this thesis. Nevertheless, estimates on dominating sources of systematic uncertainties are made and sources that are expected to be small based on the experience from other analyses are discussed briefly. A summary of quantified systematics can be found in Table 13.

The most obvious and dominating source of systematic uncertainty is the one from the branching fraction measurement of the normalisation channel taken from [**PDG**]. The normalisation leads to a relative uncertainty for the total branching fraction of $B^+ \to K^{*+}\mu^+\mu^-$ of about 5.5%.

The used simulated samples have a finite number of simulated decays and therefore a statistical uncertainty is introduced for the efficiencies (see Section 6.1). This uncertainty is commonly considered as a systematic one and it is accounted for with 3%.

In addition to the uncertainty due to the finite number of simulated events, a systematic uncertainty regarding the correction of the simulation to determine the efficiency has to be considered. Only the transverse momentum of the B meson is corrected for the differences between simulation and data (Figure 43), although other variables also show discrepancies to simulation (Figure 23). These discrepancies are assumed to be sufficiently small - in comparison to the difference in $p_T(B^+)$ - to neglect them in the efficiency determination. In order to estimate the uncertainty due to this negligence, it is assumed that the impact of the discrepancies on the total branching fraction is at most of the same order as the impact from the $p_T(B^+)$ correction. The relative uncertainty results in 2.2%, which is certainly a conservative estimate.

Source	Impact on branching fraction
Norm. to $B^+ \to J/\psi(\to \mu\mu)K^{*+}$	5.5%
Finite number of sim. events	3%
Differences sim. and data	2.2%
Quadratic sum	6.6%

Table 13: Summary of the quantified systematic uncertainties relative to the branching fraction of $B^+ \to K^{*+}\mu^+\mu^-$.

Even though the transverse momentum of the B meson from simulation is matched to the corresponding data distribution, a small systematic error remains. This is due to the fact that the weighting is conducted in bins. By increasing the number of bins this uncertainty is not necessarily reduced because the smaller bin size results in larger relative errors of the weights. In order to quantify the systematic uncertainty arising from the binning, the weights or the binning itself could be varied to investigate the influence on the efficiencies.

Furthermore, only the efficiency regarding the vetos and the BDT cut (i.e. signal selection) is corrected for the differences between simulation and data. But the overall efficiency, which is determined in Section 6.1, is composed of separate efficiencies for the trigger, stripping, preselection and signal selection, which leads to a systematic uncertainty if the simulation is not corrected before the different stages or the efficiency is not determined differently. The trigger efficiency can be determined e.g. with the so-called 'tag and probe method'[27].

The pollution from decays that have the same final state is expected to be small as well as that from final states where a particle is misidentified. The charmonia resonances $J/\psi(1S)$ and $\psi(2S)$ are removed

---

[27]In the decay $J/\psi \to \mu\mu$, for example, one muon is detected (i.e. the trigger fires) and it is used as tag for the other muon. From the rates how often the trigger fires for the second muon the trigger efficiency (of the single muon trigger) is determined.

carefully so that the uncertainty because of remaining events is negligible in comparison to other uncertainties. No veto is applied to remove the decay $B^+ \to \phi(1020)K^{*+}$ with $\phi \to \mu^+\mu^-$ because the contribution is about 0.1% relative to $B^+ \to K^{*+}\mu^+\mu^-$. A contribution to the $B^+ \to J/\psi K^{*+}$ signal yield from $B^+ \to J/\psi \rho^+(\to \pi^+\pi^0)$, when the pion is misidentified as kaon, is also negligible. No significant contribution of partially reconstructed decays is seen in the invariant mass spectrum of the B meson. The impact of S-wave contribution is minimised by the cut on the K^* mass and is also assumed to be small [12].

A model assumption is made for the fits of the B meson invariant mass which leads to a systematic uncertainty related to the fit model. This uncertainty could be estimated using alternative fit models for the description of the invariant mass distributions. A Student's t-distribution or a Novosibirsk function [31] may be considered for the signal model because they resemble the shape of the double Crystal Ball function. The background could be modeled by polynomial instead of an exponential. For each fit model, the ratio of event yields can be calculated and the maximum deviation may be taken as systematic uncertainty.

The overall systematic uncertainty on the branching fraction is 6.6% and thus small compared to the relative statistical uncertainty of $\sim 19\%$.

8 Conclusion

A measurement of the branching fraction of $B^+ \to K^{*+}\mu^+\mu^-$, where $K^{*+} \to K^+\pi^0$, is presented in this thesis, by using data collected by the LHCb experiment.

The branching fraction of the rare decay $B^+ \to K^{*+}\mu^+\mu^-$ is measured relative to the tree-level decay $B^+ \to J/\psi K^{*+}$ using the Run I dataset recorded by the LHCb experiment at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV in 2011 and 2012, respectively, with an integrated luminosity of 3 fb^{-1} .

The main challenges of this analysis are to reduce the vast amount of combinatorial background that is caused by neutral pions and to carve out the rare signal of $B^+ \to K^{*+}\mu^+\mu^-$. A loose preselection of normalisation channel candidates is performed in order to unfold a pure signal distribution via the $_s\mathcal{P}lot$ technique. The actual signal selection is based on a multivariate analysis. The efficiency of the signal selection is evaluated using simulated samples that are corrected for differences between data and simulation. To obtain the yields of $B^+ \to K^{*+}\mu^+\mu^-$ and $B^+ \to J/\psi K^{*+}$ the mass distribution of the B^+ is described using unbinned extended maximum likelihood fits.

The relative branching fraction is found to be

$$\frac{\mathcal{B}(B^+ \to K^{*+}\mu^+\mu^-)}{\mathcal{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^{*+})} = (1.03 \pm 0.20_{stat.} \pm 0.04_{syst.}) \times 10^{-2},$$

where the uncertainties are statistical and systematic, respectively. With the well-known branching fraction of the normalisation channel, the total branching fraction of $B^+ \to K^{*+}\mu^+\mu^-$ is determined to be

$$\mathcal{B}(B^+ \to K^{*+}\mu^+\mu^-) = (0.88 \pm 0.17_{stat.} \pm 0.06_{syst.}) \times 10^{-6}.$$

The number of reconstructed $B^+ \to K^{*+}\mu^+\mu^-$ signal decays is

$$nSig_{\mu\mu} = 81 \pm 16,$$

which corresponds to a significance of 6σ. The result of the determination of the total branching fraction is in agreement with the SM prediction from Table 5 and with the latest LHCb measurement of this decay (see Table 4). Although the statistical uncertainty is high with 19% in comparison to 10% in the latest LHCb analysis, the number of reconstructed signal candidates indicates that it is definitely possible to take the $K^{*+} \to K^+\pi^0$ mode into account in future analyses of $B^+ \to K^{*+}\mu^+\mu^-$, for example in an angular analysis. Especially with data from LHC Run II that was launched in June 2015 the statistical uncertainty will further decrease. The systematic uncertainty is dominated by the uncertainty of the normalisation to the resonant decay $B^+ \to J/\psi K^{*+}$ at the moment. It is surely possible to improve the value of the branching fraction of $B^+ \to J/\psi K^{*+}$ with the statistics provided by LHCb. For further improvement of the systematic uncertainty, the simulation would have to be emended, which could improve the fixation of fit model parameters and the signal efficiency determination.

A   Differences between data and simulation
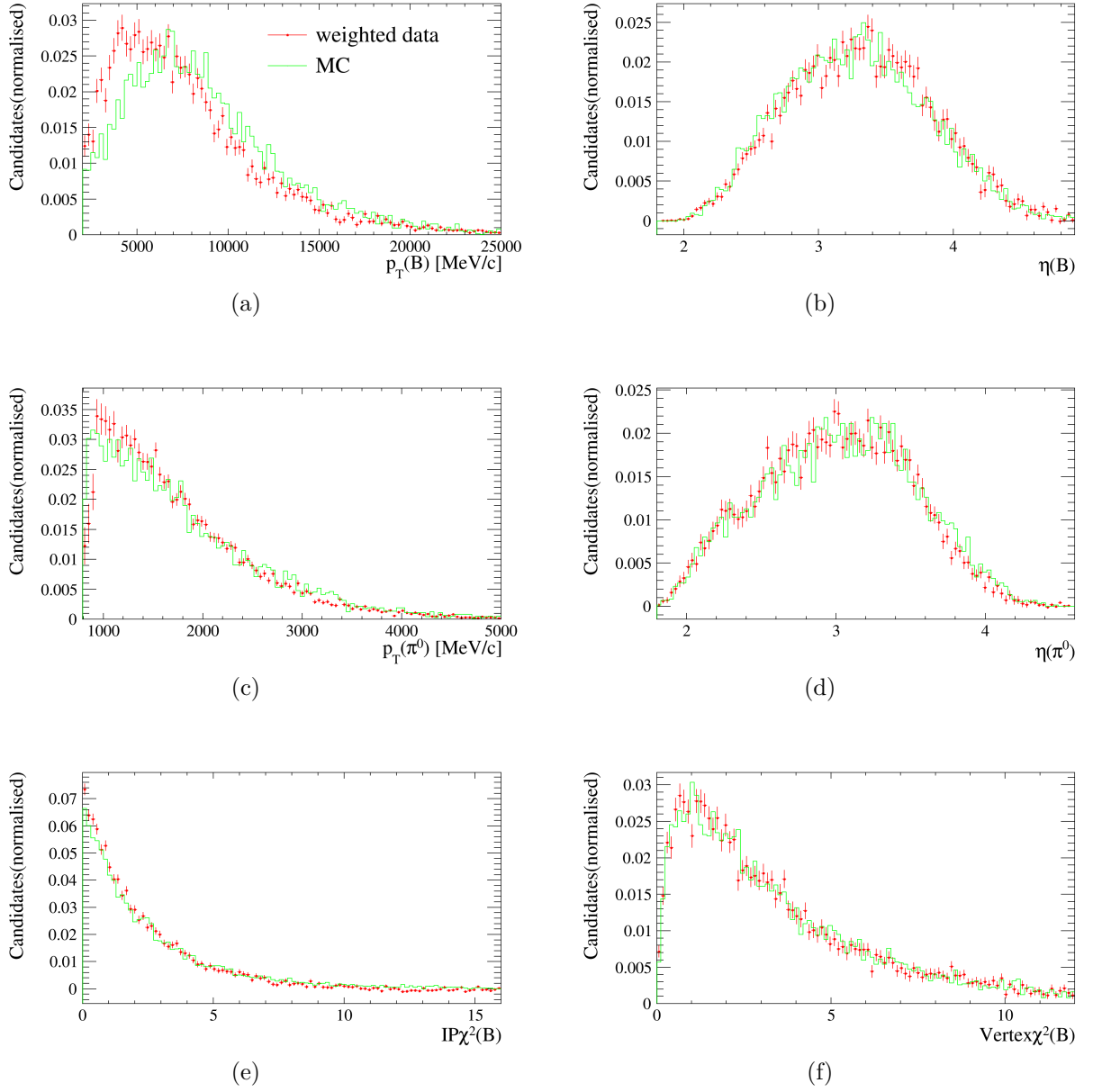

(a)

(b)

(c)

(d)

(e)

(f)

Figure 44: Distributions of $p_T(B)$, $\eta(B)$, $p_T(\pi^0)$, $\eta(\pi^0)$, $IP\chi^2(B)$ and $Vertex\chi^2(B)$ for simulation (green) and data weighted with sWeights (red, dotted). Here, the simulation sample for $B^+ \to K^{*+}\mu^+\mu^-$ was used. The distributions are very similar to those in Figure 23.
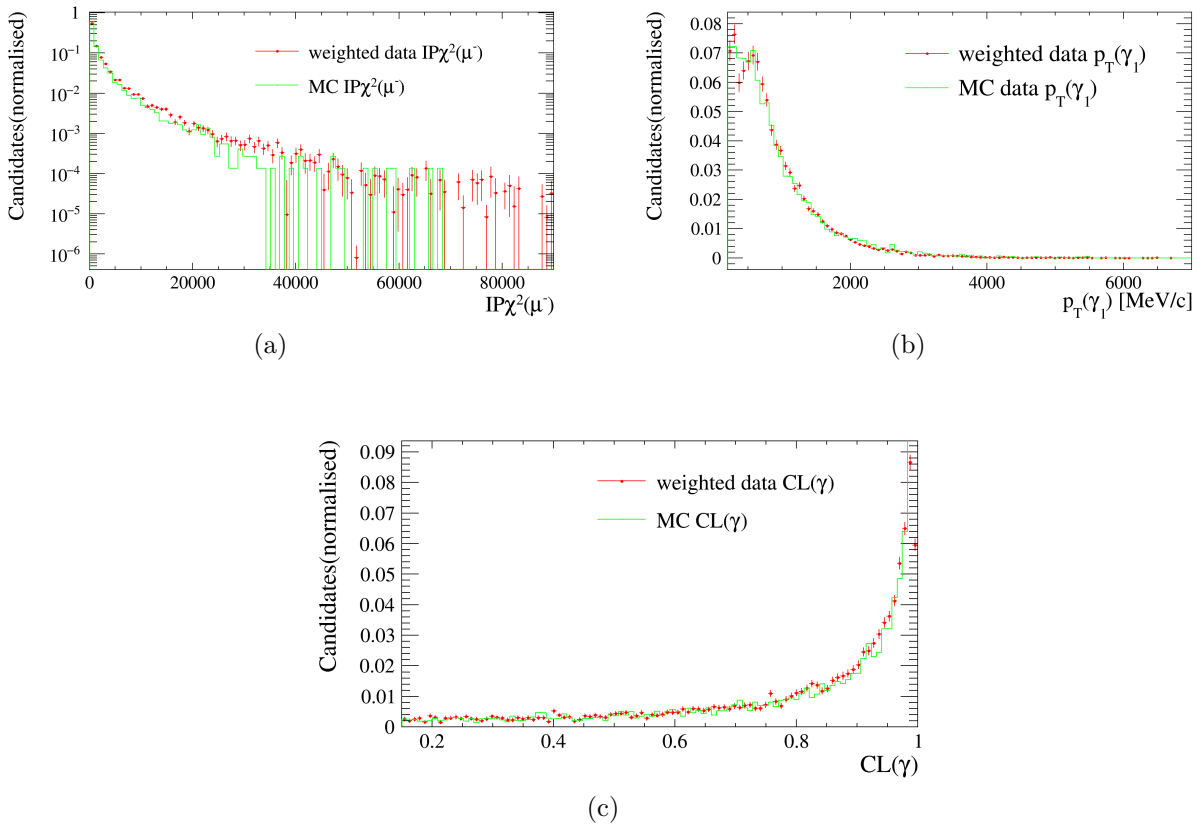
(a)

(b)

(c)

Figure 45: Distributions of $IP\chi^2(\mu^-)$ (a), $p_T(\gamma_1)$ (b) and $CL(\gamma)$ (c) for simulation (green) and data weighted with sWeights (red, dotted). The simulation sample for $B^+ \to K^{*+}\mu^+\mu^-$ was used.
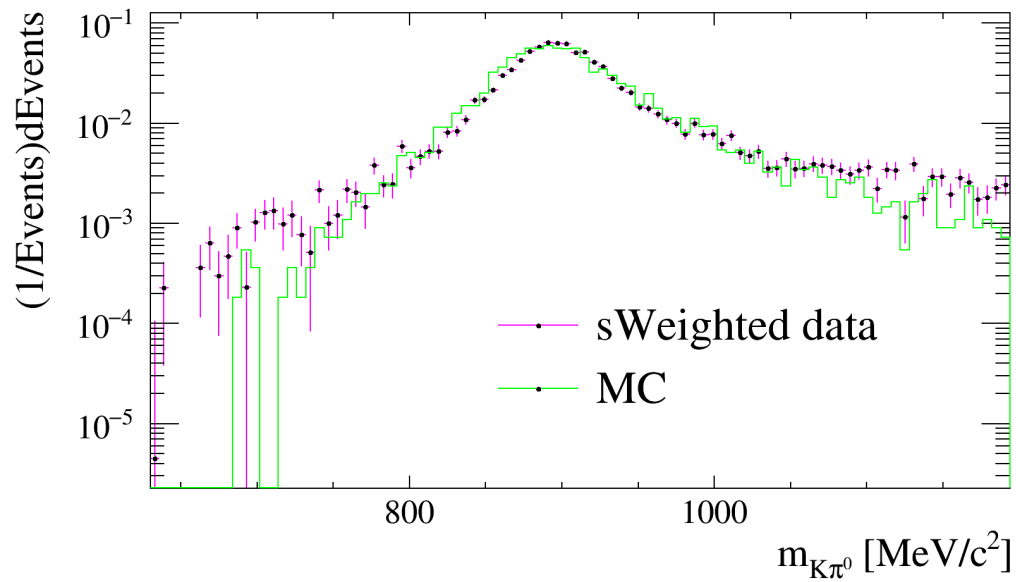
Figure 46: Distribution of the $K^+\pi^0$ invariant mass from simulation (green) and sWeighted data (magenta) after a separately trained BDT and BDT cut, without cut on the $K^+\pi^0$ invariant mass. The cut applied in the analysis is $m_{K^+\pi^0} \in [792, 1050]$.
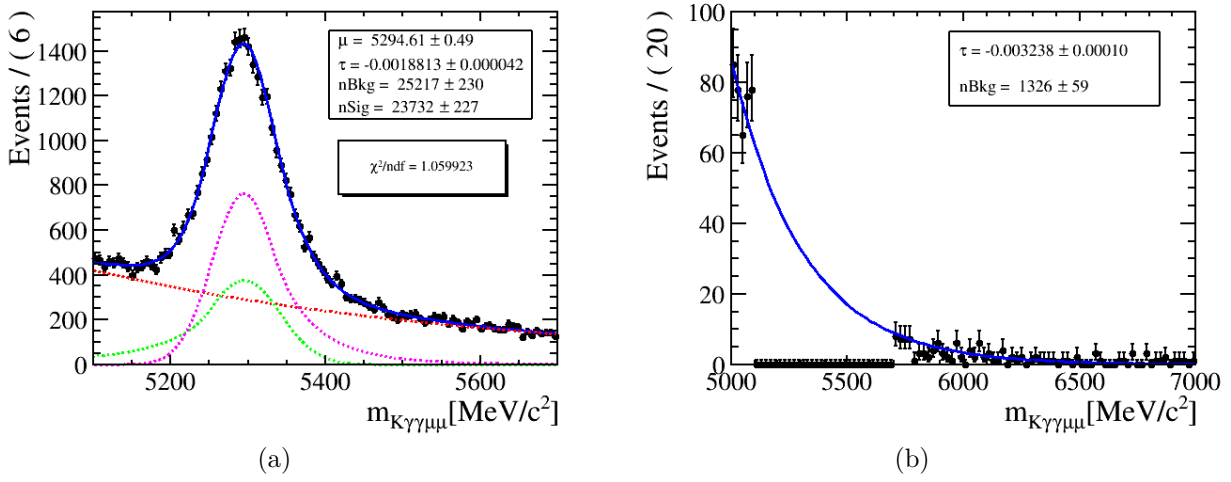
B Fit example for BDT optimisation

Figure 47: (a): Signal fit example for the BDT cut optimisation. The normalisation channel is fitted with a double CB (CB$_1$ dashed green, CB$_2$ dashed magenta) and an exponential (dashed red). Some fit parameters have been fixed to the values noted in the analysis part. (b): Background fit example for the BDT cut optimisation. The upper and lower sideband is fitted after BDT cut and charmonia vetos. The signal region was removed.

Bibliography

[1] M. D. Thomson. *Modern particle physics.* Cambridge University Press, 2013, XVI, 554 S. ISBN: 978-1-107-03426-6 (cit. on p. 11).

[2] P. Langacker. *The standard model and beyond.* Series in high energy physics, cosmology, and gravitation. Boca Raton [u.a.]: CRC Press, 2010, XII, 663 S. ISBN: 978-1-4200-7906-7 ; 1-4200-7906-9 (cit. on p. 11).

[3] G. Aad et al., ATLAS, CMS. "Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments". In: *Phys.Rev.Lett.* 114 (2015), p. 191803. DOI: 10.1103/PhysRevLett.114.191803. arXiv:1503.07589 [hep-ex] (cit. on p. 11).

[4] P. A. M. Dirac. "The Quantum Theory of the Electron". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 117.778 (1928), pp. 610–624. ISSN: 0950-1207. DOI: 10.1098/rspa.1928.0023 (cit. on p. 11).

[5] P. A. M. Dirac. "A Theory of Electrons and Protons". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 126.801 (1930), pp. 360–365. ISSN: 0950-1207. DOI: 10.1098/rspa.1930.0013 (cit. on p. 11).

[6] K. A. Olive et al., (Particle Data Group). "Review of Particle Physics". In: *Chin.Phys.* C38 (2014), p. 090001. DOI: 10.1088/1674-1137/38/9/090001 (cit. on pp. 12, 13, 17).

[7] B. Povh et al. *Teilchen und Kerne. Eine Einführung in die physikalischen Konzepte.* 8th ed. Springer-Lehrbuch ; SpringerLink : Bücher. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, p. 419. ISBN: 978-3-540-68075-8. DOI: 10.1007/978-3-540-68080-2 (cit. on p. 13).

[8] A. Ishikawa et al., Belle. "Observation of B —> K* l+ l-". In: *Phys.Rev.Lett.* 91 (2003), p. 261601. DOI: 10.1103/PhysRevLett.91.261601. arXiv:hep-ex/0308044 [hep-ex] (cit. on p. 16).

[9] J. Lees et al., BaBar. "Measurement of Branching Fractions and Rate Asymmetries in the Rare Decays $B \to K^{(*)} l^+ l^-$". In: *Phys.Rev.* D86 (2012), p. 032012. DOI: 10.1103/PhysRevD.86.032012. arXiv:1204.3933 [hep-ex] (cit. on p. 16).

[10] B. Aubert et al. "Evidence for the rare decay $B \to K^* \ell^+ \ell^-$ and measurement of the $B \to K \ell^+ \ell^-$ branching fraction". In: *Phys.Rev.Lett.* 91 (2003), p. 221802. DOI: 10.1103/PhysRevLett.91.221802. arXiv:hep-ex/0308042 [hep-ex] (cit. on p. 16).

[11] R Aaij et al., LHCb. "Measurement of the isospin asymmetry in $B \to K^{(*)} \mu^+ \mu^-$ decays". In: *JHEP* 1207 (2012), p. 133. DOI: 10.1007/JHEP07(2012)133. arXiv:1205.3422 [hep-ex] (cit. on p. 17).

[12] R. Aaij et al., LHCb. "Differential branching fractions and isospin asymmetries of $B \to K^{(*)} \mu^+ \mu^-$ decays". In: *JHEP* 1406 (2014), p. 133. DOI: 10.1007/JHEP06(2014)133. arXiv:1403.8044 [hep-ex] (cit. on pp. 17, 38, 39, 57).

[13] A. Ali et al. "Improved model independent analysis of semileptonic and radiative rare B decays". In: *Phys.Rev.* D66 (2002), p. 034002. DOI: 10.1103/PhysRevD.66.034002. arXiv:hep-ph/0112300 [hep-ph] (cit. on p. 17).

[14] A. Alves et al., The LHCb Collaboration. "The LHCb Detector at the LHC". In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005 (cit. on pp. 18–21, 23).

[15] J. Hauptman. *Particle physics experiments at high energy colliders.* eng. Weinheim: Wiley-VCH, 2011, XIII, 210 S. ISBN: 978-3-527-40825-2 (cit. on p. 18).

[16] C. Grupen and B. A. Shwartz. *Particle detectors.* eng. 2. ed. 26. Previous ed.: published as by Claus Grupen with the cooperation of Armin Böhrer and Ludek Smolík. 1996. New York, NY: Cambridge University Press, 2008, XXIII, 651 S. ISBN: 0-521-84006-6 ; 978-0-521-84006-4 (cit. on p. 18).

[17] CERN public webpage. 2015. URL: http://http://home.web.cern.ch (cit. on p. 18).

[18] L. Evans and P. Bryant, (editors). "LHC Machine". In: *JINST* 3 (2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001 (cit. on p. 18).

[19] R. Aaij et al., LHCb. "LHCb Detector Performance". In: *Int.J.Mod.Phys.* A30.07 (2015), p. 1530022. DOI: 10.1142/S0217751X15300227. arXiv:1412.6352 [hep-ex] (cit. on pp. 18, 20, 21, 23–25).

[20] R. Aaij et al., LHCb. "Measurement of $\sigma(pp \to b\bar{b}X)$ at $\sqrt{s} = 7$ TeV in the forward region". In: *Phys.Lett.* B694 (2010), pp. 209–216. DOI: 10.1016/j.physletb.2010.10.010. arXiv:1009.2731 [hep-ex] (cit. on p. 20).

[21] R. Aaij et al., LHCb. "Measurement of the inelastic pp cross-section at a centre-of-mass energy of $\sqrt{s} = 7$ TeV". In: *JHEP* 02 (2015), p. 129. DOI: 10.1007/JHEP02(2015)129. arXiv:1412.2500 [hep-ex] (cit. on p. 20).

[22] O Deschamps et al. *Photon and neutral pion reconstruction*. Tech. rep. LHCb-2003-091. Geneva: CERN, 2003. URL: http://cds.cern.ch/record/691634 (cit. on pp. 25, 26).

[23] V. Blobel and E. Lohrmann. *Statistische und numerische Methoden der Datenanalyse*. ger. Teubner-Studienbücher : Physik. Stuttgart ; Leipzig: Teubner, 2012. ISBN: 978-3-935702-66-9 (e-Buch) (cit. on p. 28).

[24] R Aaij et al. "The LHCb Trigger and its Performance in 2011". In: *JINST* 8 (2013), P04022. DOI: 10.1088/1748-0221/8/04/P04022. arXiv:1211.3055 [hep-ex] (cit. on p. 31).

[25] W. D. Hulsbergen. "Decay chain fitting with a Kalman filter". In: *Nuclear Instruments and Methods in Physics Research A* 552 (Nov. 2005), pp. 566–575. DOI: 10.1016/j.nima.2005.06.078. eprint: physics/0503191 (cit. on p. 34).

[26] M. Pivk and F. R. Le Diberder. "Plots: A statistical tool to unfold data distributions". In: *Nuclear Instruments and Methods in Physics Research A* 555 (Dec. 2005), pp. 356–369. DOI: 10.1016/j.nima.2005.08.106. eprint: physics/0402083 (cit. on p. 34).

[27] R. Partridge et al. "Decay $\frac{J}{\psi} \to 3\gamma$ and a Search for the η_c". In: *Phys. Rev. Lett.* 44 (11 1980), pp. 712–716. DOI: 10.1103/PhysRevLett.44.712 (cit. on p. 34).

[28] A. Hoecker et al. "TMVA: Toolkit for Multivariate Data Analysis". In: *PoS* ACAT (2007), p. 040. arXiv:physics/0703039 (cit. on pp. 42, 47).

[29] H. Voss, TMVA author. personal communication. 2015 (cit. on p. 47).

[30] S. Stahl. "Measurement of CP asymmetry in muon-tagged $D^0 \to K^-K^+$ and $D^0 \to \pi^-\pi^+$ decays at LHCb". PhD thesis. Combined Faculties for the Natural Sciences and Mathematics, Ruperto-Carola-University of Heidelberg, 2014 (cit. on p. 54).

[31] J. P. Lees et al., BaBar. "Branching Fraction Measurements of the Color-Suppressed Decays $\bar{B}^0 \to D^{(*)0}\pi^0$, $D^{(*)0}\eta$, $D^{(*)0}\omega$, and $D^{(*)0}\eta'$ and Measurement of the Polarization in the Decay $\bar{B}^0 \to D^{*0}\omega$". In: *Phys. Rev.* D84 (2011). [Erratum: Phys. Rev.D87,no.3,039901(2013)], p. 112007. DOI: 10.1103/PhysRevD.84.112007, 10.1103/PhysRevD.87.039901. arXiv:1107.5751 [hep-ex] (cit. on p. 57).

[32] R. Klemt. "Bestimmung des Verzweigungsverhältnisses des Zerfalls $B_s^0 \to \mu\mu f_0$ am LHCb-Experiment". Bachelorarbeit. Universität Heidelberg, 2014.

[33] S. Redford. "The branching fraction and CP asymmetry of $B^\pm \to \psi\pi^\pm$ and $B^\pm \to \pi^\pm\mu^+\mu^-$ decays". PhD thesis. St Anne's College, University of Oxford, 2012.

Acknowledgement

Erklärung

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

Heidelberg, den 06.08.2015

Paul André Günther