

Department of Physics and Astronomy

University of Heidelberg

Master thesis

in Physics

submitted by

Nikolai Fomin

born in Saint Petersburg

October 2014

**Measurement of the differential  
branching fraction of the decay  $B_s^0 \rightarrow \phi\mu^+\mu^-$   
at the LHCb experiment**

This Master thesis has been carried out by Nikolai Fomin

at the

Physikalisches Institute Heidelberg

under the supervision of

Prof. Dr. Ulrich Uwer

## Zusammenfassung

Diese Arbeit stellt eine Messung der Differentialverzweigungsverhältnis des Zerfalls  $B_s \rightarrow \phi\mu^+\mu^-$ . Das Datensatz verwendet in dieser Analyse wurde im Jahr 2011 und im Jahr 2012 vom LHCb Experiment aufgezeichnet und entspricht einer integrierten Luminosität von  $3 \text{ fb}^{-1}$ . Der Differentialverzweigungsverhältnis ist entschlossen im Kasten von  $q^2$ , die Invariante dimuon Massen. Ein Integration über den gesamten  $q^2$  Bereich wird durchgeführt, um die Gesamtverzweigungsverhältnis von  $Br(B_s \rightarrow \phi\mu^+\mu^-) = (7.954_{-0.445}^{+0.458} \pm 1.13 \pm 6.69) \times 10^{-7}$  wobei die erste Unsicherheit statistische ist, die zweite systematischen ist und die dritte systematischen ist aufgrund der Unsicherheit der Normalisierung Kanal zu bestimmen. Das Ergebnis ist in guter Übereinstimmung mit der aktuellen Welt-durchschnitt.

## Abstract

This thesis presents a measurement of the differential branching fraction of the decay  $B_s \rightarrow \phi\mu^+\mu^-$ . The dataset used in this analysis was collected by the LHCb experiment in 2011 and 2012 and corresponds to an integrated luminosity of  $3 \text{ fb}^{-1}$ . The differential branching fraction is determined in bins of  $q^2$ , the invariant dimuon mass. An integration over the full  $q^2$  range is performed to determine the total branching ratio of  $Br(B_s \rightarrow \phi\mu^+\mu^-) = (7.954_{-0.445}^{+0.458} \pm 1.13 \pm 6.69) \times 10^{-7}$  where the first uncertainty is statistical, the second systematic and the third is due to the uncertainty of the normalization channel. The result is in good agreement with the current world average.

# Contents

<b>1</b>	<b>Introduction</b>	<b>6</b>
<b>2</b>	<b>Theoretical Introduction</b>	<b>8</b>
2.1	Standard Model of Particle Physics . . . . .	8
2.2	The Electroweak Standard Model Lagrangian . . . . .	9
2.3	CKM Mechanism . . . . .	10
2.4	Flavour Changing Neutral Current Decays . . . . .	13
<b>3</b>	<b>The LHCb Experiment</b>	<b>15</b>
3.1	The Large Hadron Collider . . . . .	15
3.2	The LHCb Experiment . . . . .	16
3.2.1	Track Reconstruction System . . . . .	17
3.2.2	Particle Identification System . . . . .	19
3.2.3	The LHCb Trigger System . . . . .	24
3.3	Data Samples . . . . .	27
<b>4</b>	<b>Analysis Strategy</b>	<b>29</b>
<b>5</b>	<b>Data Selection</b>	<b>32</b>
5.1	Trigger Selection . . . . .	32
5.2	Centralized Event Selection . . . . .	33
5.3	Cutbased Selection Procedure . . . . .	34
5.3.1	Vetoos . . . . .	35
5.4	Multivariate Selection Procedure . . . . .	35
5.4.1	Boosted Decision Trees . . . . .	36
<b>6</b>	<b>Efficiency determination using simulated events</b>	<b>40</b>
6.1	Track Efficiency . . . . .	41
6.2	Muon ID Efficiency . . . . .	42
6.3	PID Resampling . . . . .	42
6.4	Further Corrections . . . . .	46
6.5	PID Variables . . . . .	46
<b>7</b>	<b>Background Contamination From B Decays</b>	<b>50</b>
7.1	$B_s \rightarrow J/\psi\phi$ . . . . .	50
7.2	$B_s \rightarrow J/\psi\phi$ with final state particles misidentified. . . . .	51
7.3	$B_d^0 \rightarrow K^*\mu\mu$ . . . . .	52
7.4	$B_d^0 \rightarrow K^*J/\psi$ . . . . .	54

7.5	$\Lambda_b \rightarrow \Lambda(1520)\mu^+\mu^-$	55
7.6	$\Lambda_b \rightarrow pK\mu^+\mu^-$	56
7.7	$B_s^0 \rightarrow D_s\pi$	57
7.8	Semileptonic $b \rightarrow c\mu^-\bar{\nu}_\mu, c \rightarrow s\mu^+\nu_\mu$	57
<b>8</b>	<b>Differential Branching Fractions</b>	<b>60</b>
8.1	General Strategy	60
8.2	Extended maximum-likelihood fit	61
8.3	$B_s \rightarrow J/\psi\phi$ fit	62
8.4	Scaling Factors	64
8.5	Signal Yields	66
8.6	Selection Efficiencies	66
8.7	Ratio of Branching Fractions	68
<b>9</b>	<b>Systematic Uncertainties</b>	<b>69</b>
9.1	Peaking Background	69
9.2	Branching Fraction of the Decay $J/\psi \rightarrow \mu^+\mu^-$	69
9.3	Relative Efficiencies Precision	69
9.4	$q^2$ binning	70
9.5	Tracking Efficiency	70
9.6	Muon ID Efficiency	70
9.7	Explicit Reweighting	71
9.8	Particle Identification Resampling	71
9.9	Signal Mass Model	71
9.10	Branching Fraction of the Normalization Channel	71
<b>10</b>	<b>Results and Conclusion</b>	<b>73</b>
10.1	Result	73
10.2	Conclusion	73
<b>A</b>	<b>Appendix</b>	<b>76</b>
<b>B</b>	<b>Lists</b>	<b>81</b>
B.1	List of Figures	81
B.2	List of Tables	83
<b>C</b>	<b>Bibliography</b>	<b>85</b>

# 1 Introduction

The Standard Model of Particle Physics describes elementary particles and their interactions. In the last fifty years it has been tested in many experiments and in 2012 the last particle predicted by the Standard Model, the Higgs boson, has been discovered. However, while the Standard Model is extremely successful it still has shortcomings. The Standard Model doesn't include gravitational force and doesn't have explanations for the dark matter or for baryogenesis. This among other things means that the Standard Model cannot explain the Big Bang and the following evolution of the Universe. Many models based on the Standard Model have been developed in the last decades in attempt to describe these phenomena. Most such models predict existence of new, yet undiscovered particles and the experimental search for physics beyond the Standard Model is an important goal in modern particle physics.

While the search for the Higgs boson has been the main reason behind LHC, the largest particle accelerator built, the high luminosity and energy make the searches for new physics possible. ATLAS and CMS experiments are general purpose experiments designed for direct searches of undiscovered particles. The LHCb experiment is in contrast designed to measure processes related to decays of b and c hadrons with high precision. Many models predict new particles to be very heavy and direct searches might require very big energies. However such particles should also appear in quantum loops as virtual particles. Existence of new particles would change the amplitudes of various processes compared to the Standard Model predictions. Decay rates of rare decays and CP violating asymmetries are particularly sensitive to such changes. The LHCb experiment uses decays of B hadrons to investigate these quantities.

In this thesis the measurement of the differential branching fraction of the rare flavour-changing neutral current  $B_s \rightarrow \phi \mu^+ \mu^-$  decay is presented. The dataset used in the analysis consists of data collected by LHCb during 2011 and 2012 runs. It corresponds to an integrated luminosity of  $3.0 \text{ fb}^{-1}$ . The analysis is part of the effort to measure angular observables of the final states of the  $B_s \rightarrow \phi \mu^+ \mu^-$  decay.

The thesis is structured as follows: Chapter 2 gives an introduction to the Standard Model and the flavour-changing processes such as  $B_s \rightarrow \phi\mu^+\mu^-$ . In Chapter 3 the LHCb experiment is introduced. Reliable particle identification is crucial for the analysis and the LHCb PID system is described in detail. Chapter 4 outlines the strategy of the analysis. Chapters 5 - 8 describe the analysis in details. Chapter 9 discusses the systematic uncertainties and Chapter 10 presents the results of the analysis.

## 2 Theoretical Introduction

In this chapter a brief overview of the Standard Model is given and the CKM mechanism is described. Furthermore the possibilities of flavour changing neutral currents are discussed.

### 2.1 Standard Model of Particle Physics

The Standard Model (SM) of particle physics is a renormalizable quantum field theory that combines the electromagnetic force, the weak force and the strong force. The Standard Model Lagrangian is gauge invariant under local transformation of the  $SU(3) \times SU(2)_L \times U(1)_Y$  symmetry group. Gravity is negligible at the scales below the Planck scale ( $1.22 \cdot 10^{19} \text{GeV}$ ) which is far above energies that are currently possible to reach with experiments. The Standard Model is very successful in providing precise experimental predictions. There is a lot of literature available on the subject, for example [13] and [18]. In the following the main focus will be on the Glashow-Salam-Weinberg electroweak theory that unifies electromagnetism and weak force as it is most relevant to the analysis performed.

The fundamental particles that build up the matter as well as the carriers of the forces are described in the Standard Model as fields. Fermions, fields of spin  $1/2$ , are quarks and leptons. Quarks have both color and electroweak charge and leptons have only electroweak charge. Both quarks and leptons are divided in three generations with different mass scales, but the same quantum numbers. Each fermion has a corresponding antiparticle with the same mass, but the opposite quantum numbers. The strong force is mediated by 8 massless gluons, the electroweak force by massless photons  $\gamma$  and three massive gauge bosons  $W^+$ ,  $W^-$  and  $Z^0$ . Force carriers have spin 1. The masses of gauge bosons are generated by spontaneous symmetry breaking of the electroweak symmetry group,  $SU(2)_L \times U(1)_Y$ . A mechanism suggested by Higgs et. al. [10], [19], [14] predicts existence of the Higgs field that breaks the  $SU(2)$  symmetry without disturbing the gauge invariance which leads to a massive spin 0 particle, the Higgs boson. It has remained the last undiscovered particle, but

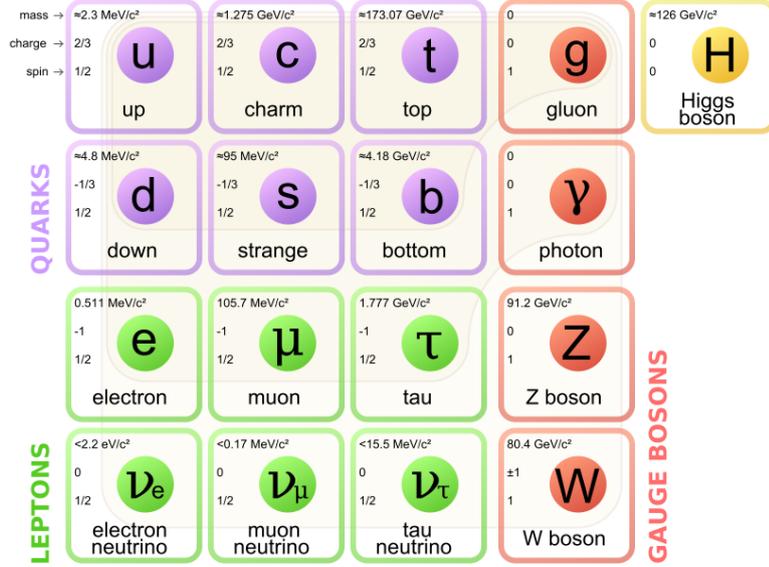


Figure 2.1: Constituents of the Standard Model [3]

its discovery has been announced in 2012 further validating the Standard Model.

Figure 2.1 shows all the fundamental particles of the Standard Model with their measured mass, spin and charge. For each of the fermions an antiparticle exists with the same mass and spin, but opposite charge. Antiparticles are not treated separately as they mostly behave like their counterpart.

## 2.2 The Electroweak Standard Model Lagrangian

The electroweak Standard Model Lagrangian is

$$\mathcal{L}_{SM} = \mathcal{L}_{gauge} + \mathcal{L}_{fermion} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa}.$$

The gauge term is

$$\mathcal{L}_{gauge} = -\frac{1}{4}W_{\mu\nu}^i W^{\mu\nu i} - \frac{1}{4}B_{\mu\nu} B^{\mu\nu}$$

where  $W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i - g\epsilon^{ijk}W_\mu^j W_\nu^k$  and  $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$  are the field strength tensors for SU(2) and U(1) respectively. This term includes the gauge bosons kinetic energy terms and describes the self-interactions of  $W_{\mu\nu}$  gauge bosons. The abelian U(1) gauge bosons  $B_{\mu\nu}$  have no self-interactions.

The fermion part of the Standard Model is

$$\mathcal{L}_{fermion} = \sum_j^3 (\bar{q}'_{jL} i \not{D} q'_{jL} + \bar{l}'_{jL} i \not{D} l'_{jL} + \bar{u}'_{jR} i \not{D} u'_{jR} + \bar{d}'_{jR} i \not{D} d'_{jR} + \bar{e}'_{jR} i \not{D} e'_{jR} + \bar{\nu}'_{jR} i \not{D} \nu'_{jR} + h.c.)$$

where  $q'_{jL} = \begin{pmatrix} u'_j \\ d'_j \end{pmatrix}_L$  represents the left-handed doublets of quarks,  $l'_{jL} = \begin{pmatrix} e'_j \\ \nu'_j \end{pmatrix}_L$  are the left-handed doublets of leptons,  $q'_{jR}, d'_{jR}$  are the right-handed quark singlets and  $e'_{jR}, \nu'_{jR}$  are the right-handed lepton singlets. This term describes the interactions of fermions with gauge bosons.

The Higgs part of the Lagrangian is given by

$$\mathcal{L}_{Higgs} = (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi)$$

where  $\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$  is a complex Higgs scalar and  $V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2$  is the Higgs potential. This term describes the Higgs interactions with the gauge bosons and the Higgs self-interactions.

The Yukawa Lagrangian is given by:

$$\mathcal{L}_{Yukawa} = -\sum_{i,j} \left( \bar{q}'_{iL} g_{ij}^{(d)} \phi d'_{jR} + \bar{q}'_{iL} g_{ij}^{(u)} \tilde{\phi} u'_{jR} + \bar{l}'_{iL} g_{ij}^{(e)} \phi e'_{jR} + \bar{l}'_{iL} g_{ij}^{(\nu)} \tilde{\phi} \nu'_{jR} \right) - h.c.$$

where  $g^{(u)}, g^{(d)}, g^{(e)}, g^{(\nu)}$  are  $3 \times 3$  matrices that describe the so called Yukawa couplings between the single Higgs doublet  $\phi$  and the fermions.

A schematic overview of the electroweak Standard Model is given in Figure 2.2.

## 2.3 CKM Mechanism

Suppose that the Higgs potential has a minimum at  $\phi = v$  where  $v$  is the vacuum expectation value of  $\phi$ . After the symmetry breaking  $\phi = \begin{pmatrix} 0 \\ v+h(x) \end{pmatrix} / \sqrt{2}$ . The Yukawa Lagrangian can be rewritten as:

$$\mathcal{L}_{Yukawa} \subset -\sum_{i,j} \left( \bar{d}'_{iL} \frac{g_{ij}^{(d)} v}{\sqrt{2}} d'_{jR} + \bar{u}'_{iL} \frac{g_{ij}^{(u)} v}{\sqrt{2}} u'_{jR} \right)$$

The  $M_{ij}^{(u,d)} = \frac{g_{ij}^{(u,d)} v}{\sqrt{2}}$  can be interpreted as the quark mass matrix.

$$\mathcal{L}_{Yukawa} \subset -(\bar{u}'_L M^{(u)} u'_R + \bar{d}'_L M^{(d)} d'_R)$$

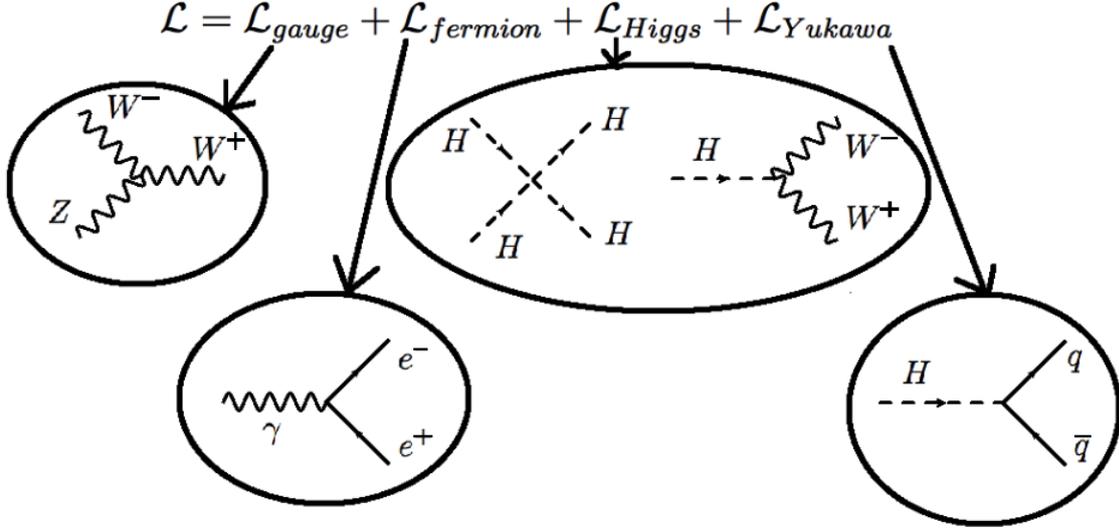


Figure 2.2: A schematic overview of different types of interactions described by each term of the electroweak Standard Model Lagrangian.

where  $u'_L = \begin{pmatrix} u'_{1L} \\ u'_{2L} \\ u'_{3L} \end{pmatrix}$ ,  $d'_L = \begin{pmatrix} d'_{1L} \\ d'_{2L} \\ d'_{3L} \end{pmatrix}$  and  $M^{(u)}$  and  $M^{(d)}$  are the non-diagonal mass matrices. The mass matrices can be diagonalized to find the physical "mass states" (e.g.  $\bar{u}_L m_u u_R$ ):

$$U_d^\dagger M^{(d)} V_d = D^d = \begin{pmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{pmatrix}$$

$$U_u^\dagger M^{(u)} V_u = D^u = \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{pmatrix}$$

where  $U_u U_u^\dagger = V_u V_u^\dagger = U_d U_d^\dagger = V_d V_d^\dagger = \mathbb{1}$ .

With this the Lagrangian describing the quark masses can be rewritten as:

$$-\mathcal{L} \subset \bar{d}_L D^d d_R + \bar{u}_L D^u u_R$$

where  $u_L$ ,  $u_R$ ,  $d_L$ ,  $d_R$  are the mass eigenstates now.

The Lagrangian describing the weak charged current interaction of quarks

$$\mathcal{L}_{CC} = -\frac{g}{\sqrt{2}} W_\mu^\dagger \bar{u}'_{Lj} [U_u U_u^\dagger] \gamma^\mu [U_d U_d^\dagger] d'_{Lj} + h.c.$$

can be expressed through the quarks mass eigenstates.

$$\mathcal{L}_{CC} = -\frac{g}{\sqrt{2}}W_\mu^\dagger \bar{u}_{Lj}[U_u^\dagger]\gamma^\mu[U_d]d_{Lj} + h.c.$$

Introducing the weak eigenstates of the down-type quarks  $d_L^W$ ,  $s_L^W$ ,  $b_L^W$  one can rewrite the weak charged current interaction

$$\mathcal{L}_{CC} \subset -\frac{g}{\sqrt{2}}W_\mu[\bar{u}_L\gamma^\mu d_L^W + \bar{c}_L\gamma^\mu s_L^W + \bar{t}_L\gamma^\mu b_L^W] + h.c.$$

where the weak eigenstates are defined by

$$\begin{pmatrix} d_L^W \\ s_L^W \\ b_L^W \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix}$$

and the matrix  $V$  is given by  $V = [U_u U_d^\dagger]$  and is called Cabibbo-Kobayashi-Maskawa (CKM) matrix. It is unitary, but not diagonal. The CKM matrix has four free parameters, three real amplitudes and a phase. There are several ways to parametrize it, a usual choice is the Wolfenstein parameterisation:

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

The CKM matrix contains information on the strength of flavour-changing weak decays. The Wolfenstein parameterisation shows that transition from up-type quark to down-type in the same generation is not suppressed, transitions between neighboring generations are suppressed by factors  $\lambda$  and  $\lambda^2$  while the transition from first to third generation is suppressed by the factor of  $\lambda^3$ . For the parameter  $\lambda$  one finds  $\lambda \approx 0.22$  Other parameters in the Wolfenstein parameterisation have values between 0.1 and 1 and they do not change the order of magnitude.

## 2.4 Flavour Changing Neutral Current Decays

The part of the Lagrangian describing the weak neutral current interactions is given by

$$\mathcal{L}_{NC} \subset -(u_{Lj}, d_{Lj})\gamma^\mu(g\tau_3 W_\mu^3/2 + g'Y B_\mu/2)\begin{pmatrix} u_{Lj} \\ b_{Lj} \end{pmatrix}.$$

Since  $\tau_3$  and  $Y$  are diagonal (in contrast to  $\tau_1$  and  $\tau_2$ ) it can be concluded that the neutral current part of the Lagrangian is diagonal in the mass basis. As a consequence there are no Flavour Changing Neutral Currents (FCNC) in Standard Model at tree level.

While FCNC processes are forbidden at tree level, they can still proceed through loops. The most common diagrams that describe such decays are penguin diagrams and box diagrams, see Figure 2.3. If new particles exist even if they are very heavy they can still appear in quantum loops. Processes to which the tree contributions are forbidden are very sensitive to the appearances of new particles in the loops.

The aim of this analysis is to measure the branching fraction of the decay  $B_s \rightarrow \phi\mu^+\mu^-$ . As can be seen in Figure 2.3 this decay involves a  $\bar{b} \rightarrow \bar{s}$  transition, it is thus a FCNC process. That means that  $B_s \rightarrow \phi\mu^+\mu^-$  is suppressed compared to e.g.  $B_s \rightarrow J/\psi(\rightarrow \mu^+\mu^-)\phi$  that is not forbidden at the tree level. The yield of  $B_s \rightarrow J/\psi(\rightarrow \mu^+\mu^-)\phi$  is expected to be roughly 150 times larger than that of  $B_s \rightarrow \phi\mu^+\mu^-$ . In extensions of the Standard Model contributions from new particles can change the branching ratio. For example in Figure 2.4 the dependence of branching ratio on the mass of the charged Higgs for a two Higgs doublet model. The theoretical prediction for the SM branching ratio of the  $B_s \rightarrow \phi\mu^+\mu^-$  decay range from  $14.6 \cdot 10^{-7}$  to  $19.2 \cdot 10^{-7}$  [15], [8], [27].

To conclude  $B_s \rightarrow \phi\mu^+\mu^-$  and analogous decays such as  $B_d \rightarrow K^*\mu^+\mu^-$  and  $\Lambda_b \rightarrow \Lambda\mu^+\mu^-$  are interesting to analyze since they are sensitive to the gauge structure and various extensions of the SM. Such decays can be used not only for testing the Standard Model, but also for probing new physics.

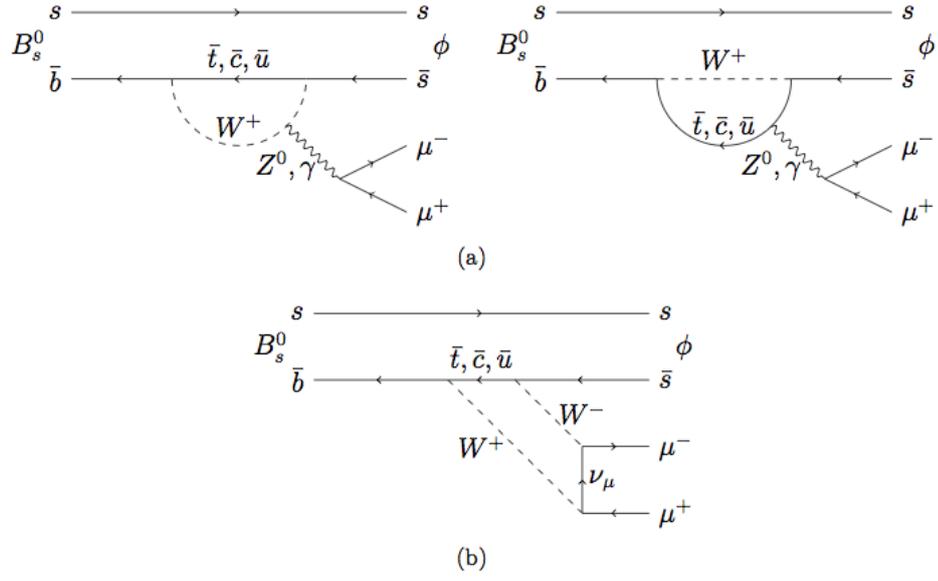


Figure 2.3: Decay of  $B_s \rightarrow \phi \mu^+ \mu^-$  mediated via (a) penguin diagrams, (b) box diagram. Figure taken from [22].

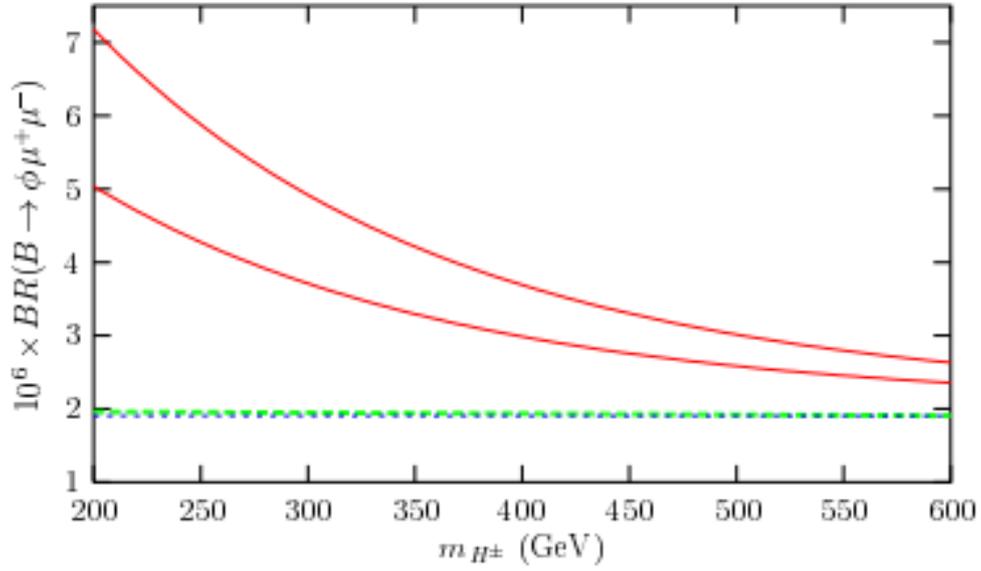


Figure 2.4: Dependence of  $BR(B_s \rightarrow \phi \mu^+ \mu^-)$  on the mass of charged Higgs for a two Higgs doublet model. Solid lines show the prediction for the model, dotted lines show prediction for the SM. Figure taken from [15].

## 3 The LHCb Experiment

The LHCb experiment (Large Hadron Collider beauty experiment) is one of the four main experiments at the Large Hadron Collider (LHC) at CERN. It is primarily dedicated to the measurements of b- and c-hadron decays and searches for CP violation and rare B and D meson decays. The following chapter will briefly introduce the LHC and LHCb experiment and discuss the detector components important for this analysis.

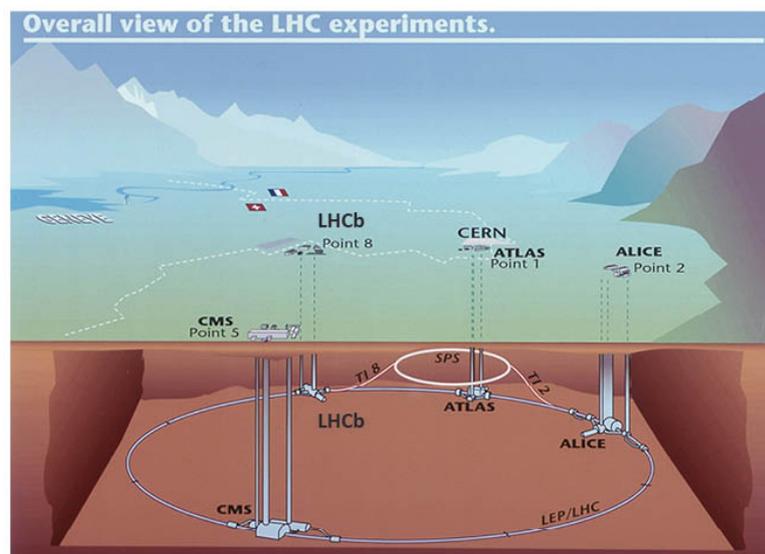


Figure 3.1: Schematic overview of the LHC main experiments and underground accelerator ring. Taken from [17]

### 3.1 The Large Hadron Collider

The Large Hadron Collider is a proton-proton (pp) collider located at CERN in Geneva, Switzerland. Two proton beams are accelerated in a 27 km long under-

ground tunnel and brought to collision at four interaction points where the main LHC experiments are placed: ATLAS, CMS, ALICE and LHCb. Figure 3.1 shows the tunnel and the position of the four experiments. ATLAS and CMS are general purpose detectors, ALICE is designed to observe heavy ion collisions and LHCb focuses on b physics.

The design centre-of-mass energy of the proton collisions of LHC is  $\sqrt{s} = 14$  TeV. The proton beams are separated into 2808 bunches to reach the design luminosity of  $\mathcal{L} = 10^{34} \text{cm}^{-2} \text{s}^{-1}$ , each bunch containing  $\approx 10^{11}$  protons. The time separation between the bunches is 50 ns, the resulting interaction rate is 15 MHz. In 2011 during the data-taking the center of mass energy was  $\sqrt{s} = 7$  TeV and in 2012 it was increased to  $\sqrt{s} = 8$  TeV.

## 3.2 The LHCb Experiment

The LHCb detector is a single-arm forward spectrometer designed to cover the acceptance for b-hadrons. It covers a pseudo rapidity range of  $1.8 < \eta < 4.9$ . 25% of  $b\bar{b}$  quark pairs are produced inside the LHCb acceptance, see Figure 3.2. The LHCb detector cannot operate at the (maximum) instantaneous luminosity of LHC, so the proton beams are slightly defocused before collision. Furthermore a hardware trigger is used to reduce the event rate to about 1MHz, at which the entire detector can be read out.

Many events in the LHCb acceptance are background events containing no B meson. Furthermore not all B meson decays are equally interesting for e.g. CP violation studies. A combination of hardware and software triggers is used to achieve the highest efficiency for the events interesting to offline analyses while rejecting as much background as possible. The detector itself consists of several subdetectors used either for track reconstruction or particle identification (PID). Calorimeters measure the energy of the decay products. A dipole magnet produces an approximately homogeneous vertical field to separate positively and negatively charged particles. Figure 3.3 shows a vertical cross-section of the whole detector.

The LHCb coordinate system has the origin at the interaction point. The z-axis runs along the beam-line from with the detector being in the positive direction. The positive y-axis points upwards. This fixes the positive x-axis to point toward the cavern access.

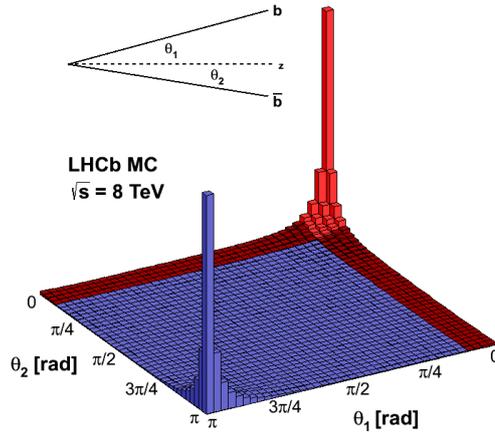


Figure 3.2: Simulated distribution of polar angles  $\Theta_1$  and  $\Theta_2$  of produced  $b\bar{b}$  pairs. Most of the  $b$  and  $\bar{b}$  are produced in either forward or backward direction. The LHCb acceptance is colored red. Taken from [17]

### 3.2.1 Track Reconstruction System

The LHCb tracking system consists of the vertex locator system (VELO) and the Tracker Turicensis stations (TT) upstream of the magnet and three tracking stations (T1-T3) downstream of the magnet. The purpose of the tracking system is to combine measurements of position before the magnet and after the magnet to form trajectories of the charged particles and subsequently measure the momentum of the particles.

**Dipole Magnet** The dipole magnet is designed such that the magnetic field is nearly homogenous with a large component in y-direction ("upwards") and as small as possible component in x- and z-direction. Therefore particles moving in the z-direction will bend mostly in x-z plane. The integrated magnetic field over distance of 10 m is

$$\int B dl = 4 \text{ Tm}.$$

The layout of the magnet and the strength of the field as function of the distance along the beam line can be seen in Figures 3.4a and 3.4b. The polarity of the magnet can be flipped so that positively (negatively) charged particles are bent in the direction negatively (positively) charged particles were bent before. During the data taking the polarity is changed periodically and the physics analyses are

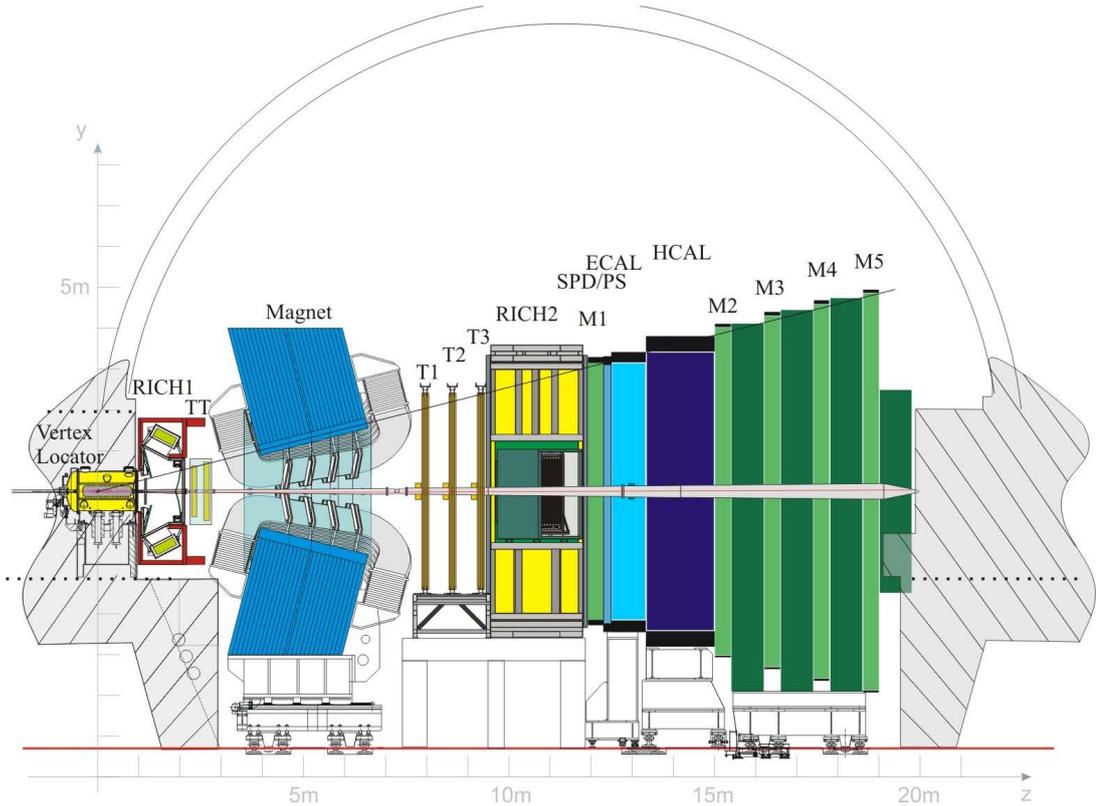
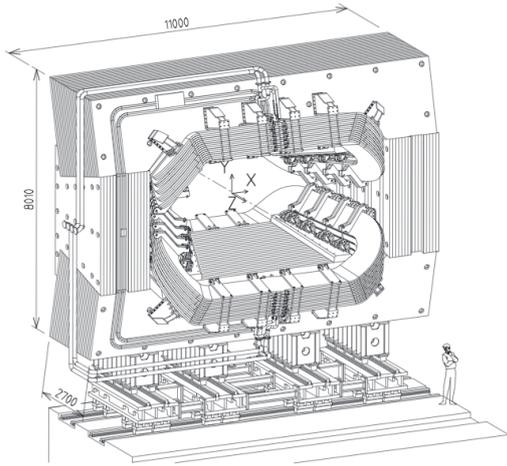


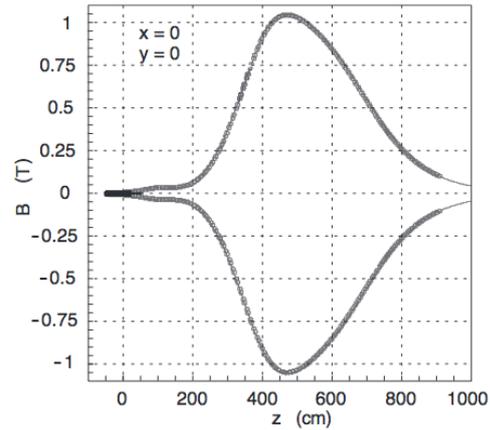
Figure 3.3: Vertical cross-section of the LHCb detector. Taken from [17]

performed on the combined data. This approach allows to cancel many detector asymmetries.

**Vertex Locator** The Vertex Locator (VELO) is a silicon strip detector positioned around the interaction point. The purpose of the VELO is to perform precise measurements of the primary and secondary vertices and to distinguish particles produced from secondary and primary vertices (prompt particles). Figure 3.4 shows the location of VELO stations. Each VELO station measures the distance from the beam axis  $r$  and the polar angle  $\phi$  of the tracks. Such cylindrical coordinates are chosen since they allow fast track reconstruction in the software trigger. Additionally two pile-up stations are placed upstream of the VELO. The hardware trigger uses them to detect beam-gas interactions. To achieve high sensitivity the VELO stations are placed at 8 mm distance from the beam axis. During the injection of the beam the VELO sub sensors are retracted to prevent radiation damage.



(a) Schematic view of LHCb dipole magnet (units in mm). Taken from [11]



(b) Magnetic field along the z axis. Taken from [11]

**Tracker Turicensis** The Tracker Turicensis (TT) is a silicon detector positioned upstream of the dipole magnet and downstream of the RICH1 detector. It consists of two stations with two layers of silicon strips each. The distance between the stations is 27 cm. In the two inner layers the strips are rotated by  $5^\circ$  to also perform measurements in the y-plane.

**Inner Tracker** The main tracking stations (T1-T3) are divided into Inner Tracker (IT) and Outer Tracker (OT). The IT is a silicon strip detector positioned in the centre of the main tracking stations. Each station consists of four layers with geometry similar to that of TT and the inner layers are also rotated by  $5^\circ$ .

**Outer Tracker** In contrast to other tracking detectors the OT is a straw tube drift-time detector. It covers the acceptance outside of the IT acceptance.

### 3.2.2 Particle Identification System

Particle identification is the process of identifying the type of the particle passing the detector using the information left by the particle. It is an essential requirement for the reconstruction of B meson decays. At LHCb an especially difficult problem is separating pions from other charged particle because pions are much more abundant.

**RICH detectors** The Ring Imaging Cherenkov detectors (RICH) are used to identify charged particles over a wide momentum range. The RICH detectors measure

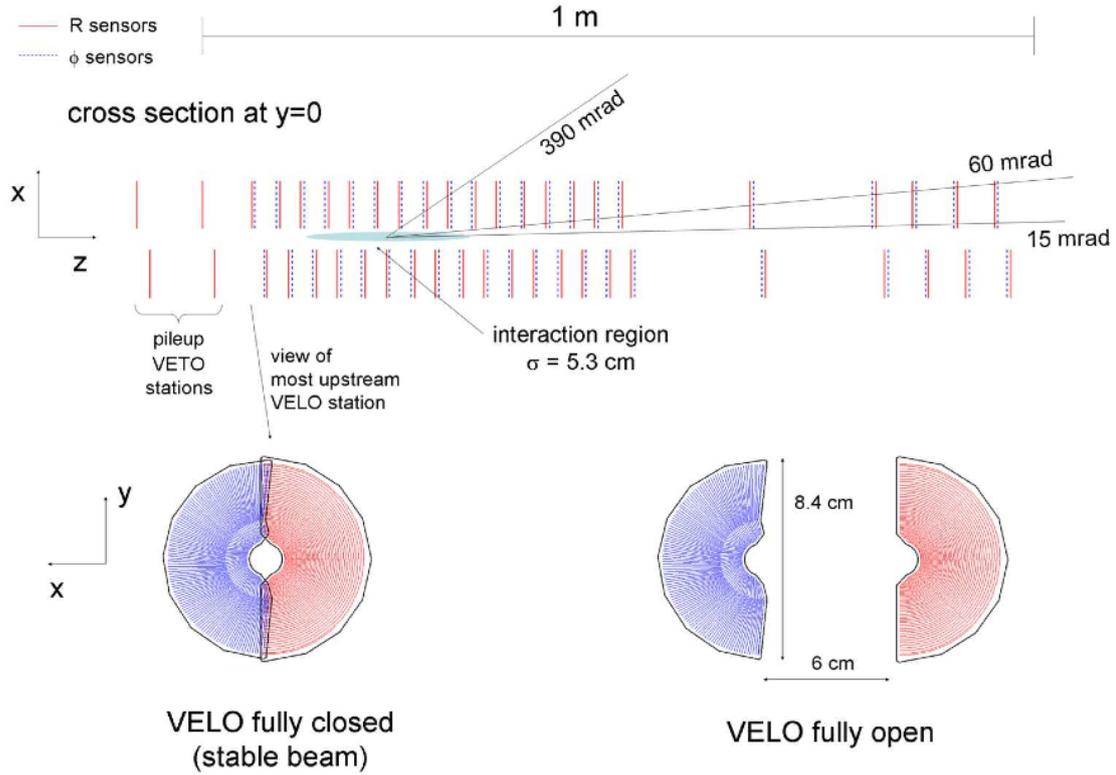


Figure 3.4: Cross section in the x-z plane of the VELO sensors. The VELO consists of 21 stations along the beam pipe. Until the beam stabilizes the modules are kept open. Taken from [11]

so-called Cherenkov radiation: photons emitted when a particle travels through a medium with refractive index  $n$  (also called radiator) with velocity higher than  $c' = \frac{c}{n}$ , the speed of light in this medium. Cherenkov radiation is emitted in a cone, in analogy to sound waves generated by a supersonic body propagating through a medium. The opening angle of the cone is given by:  $\cos \Theta_C = \frac{c'}{v} = \frac{1}{n\beta}$ .

If in addition one knows the particle's momentum (from the tracking system) and refractive index  $n$  one can predict  $v$  and subsequently  $\Theta_C$ . It can then be compared to the measured  $\Theta_C$ . Different hypotheses (particle being proton, kaon, pion, electron or muon) are tested and the likelihood of each hypotheses is returned. There are two RICH detectors installed at LHCb to cover a larger momentum range and to better separate charged particles. Figure 3.5 shows the Cherenkov angles for different particle species over the momentum range. From the plot it can be seen that the aerogel radiator offers good separation at lower momentum while  $CF_4$  gas is better for separation of high momentum particles.

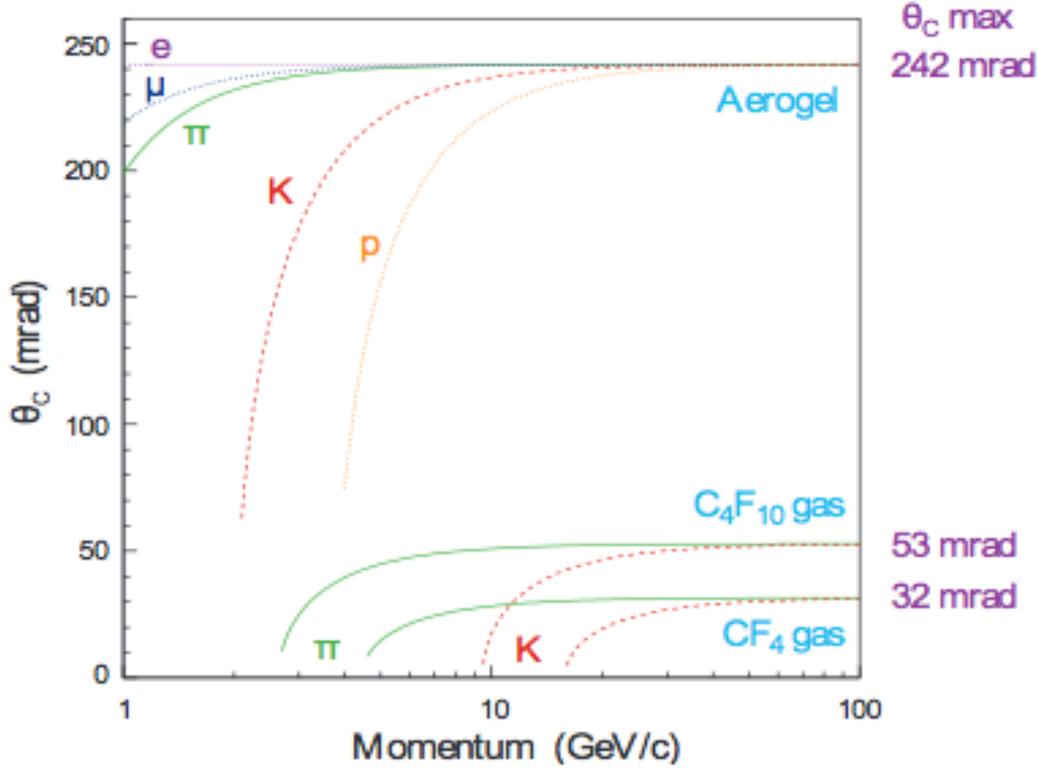
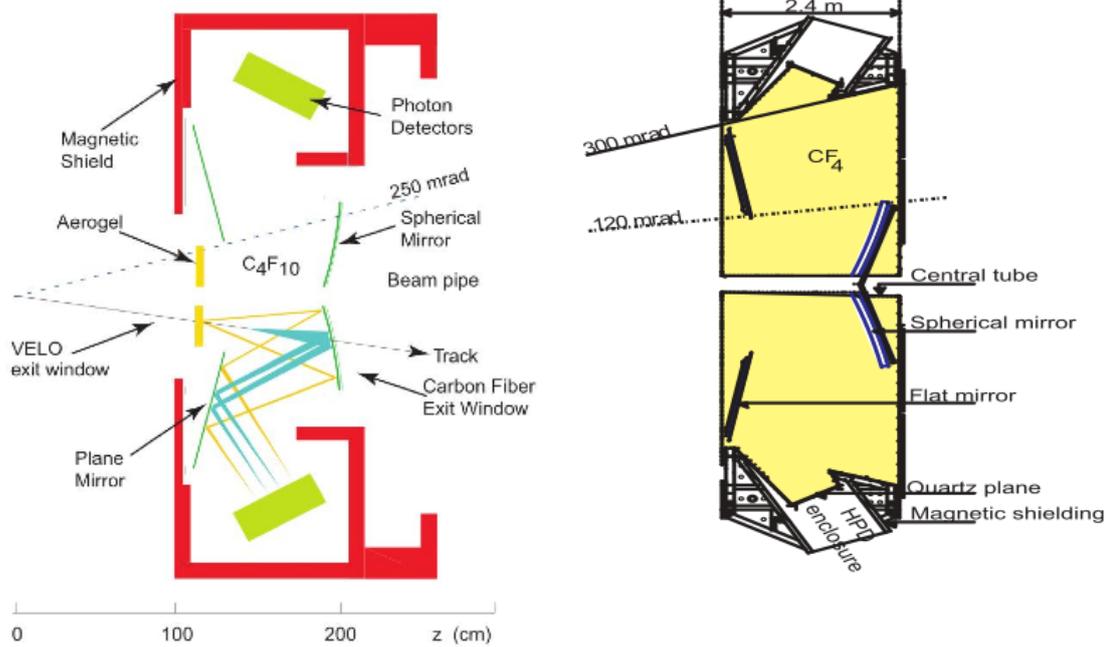


Figure 3.5: Cherenkov angle versus particle momentum for the RICH radiation. Taken from [11]

The upstream RICH1 is placed close to the VELO. RICH1 covers the full LHCb acceptance ( $\pm 25$  mrad to  $\pm 300$  mrad in the horizontal and  $\pm 250$  mrad in the vertical planes). It uses aerogel and  $C_4F_{10}$  gas as radiators and covers the low-momentum range of 1-60 GeV. A schematic layout of the RICH1 can be seen in Figure 3.6a.

RICH2 is located between the T-stations and the electromagnetic calorimeter and has a smaller acceptance than RICH1 ( $\pm 15$  mrad to  $\pm 120$  mrad the horizontal and  $\pm 100$  mrad in the vertical planes). It uses  $CF_4$  as radiator and covers the higher-momentum range of 15-100 GeV. RICH2 does not need to cover the whole LHCb acceptance since it is primarily used for separation of high-momentum particles that pass the detector with small opening angle. A schematic layout of the RICH2 is presented in Figure 3.6b.

**Muon System** The LHCb muon system is composed of five stations (M1-M5) placed along the beam axis, see Figure 3.7. The muon system provides data used



(a) Side view schematic layout of RICH1. Taken from [11]  
 (b) Top view schematic layout of RICH2. Taken from [11]

Figure 3.6

for identification and reconstruction of muons as well as for muon triggering. Also it provides muons  $p_T$  to the hardware trigger for the online triggering.

The first station M1 is placed before the calorimeter, its purpose is to improve the measurement of muon transverse momentum  $p_T$  of the trigger before scattering in the calorimeter takes place. Stations M2-M5 are placed downstream of the calorimeter with 80 cm thick iron absorbers between them to absorb all non-muons. The total thickness corresponds to roughly 20 hadronic interaction lengths, the minimum momentum of the muon required to cross all five stations is around 6 GeV.

The muon stations are partitioned in four regions with increasing distance from the beam axis. The linear ratio is 1 : 2 : 4 : 8 and it is chosen such that the particle flux is approximately the same in each quadrant see Figure 3.8 for further information.

**Calorimeters** The LHCb calorimeter system selects high transverse energy ( $E_T$ ) hadrons, photons and electrons for the hardware trigger and provides measurements of particles' energies and positions. It is also the only system that is sensitive to neutral particles so it is essential for reconstruction of  $\pi^0$  and prompt photons.

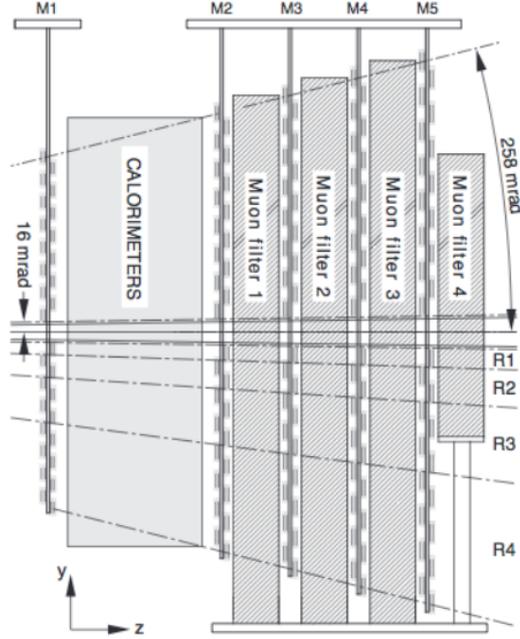


Figure 3.7: Side view schematic layout of LHCb muon system. Taken from [11]

The calorimeter system consists of several sub-detectors. The electromagnetic calorimeter (ECAL) is followed by the hadronic calorimeter (HCAL). For rejection of background a scintillator pad detector (SPD) and a pre shower detector (PS) are used. All detectors use a lateral segmentation, with granularity being finer closer to the beam. All calorimeters work by the same principle - first particles traverse absorber material to induce particle showers. The showers pass scintillating material emitting photons. The scintillation light is transmitted by wavelength-shifting fibers to photo-multipliers. The energy of a particle is measured from the photomultiplier signal.

**SPD** and **PS** detectors are located behind the M1 muon station. The SPD/PS detector consists of two layers of scintillator pads separated by 15 mm of lead. The SPD is used to separate electrons from photons (photons do not create a signal in the scintillator). The purpose of the PS detector is to suppress  $\pi^\pm$  background. Particles passing through lead create showers, the difference in energy deposition between  $\pi^\pm$  and electrons of 50 GeV momentum is shown in Figure 3.9a.

The electromagnetic calorimeter, **ECAL**, is used to detect particle showers from photons and electrons. It is build by alternating 2 mm thick layers of lead and 4mm

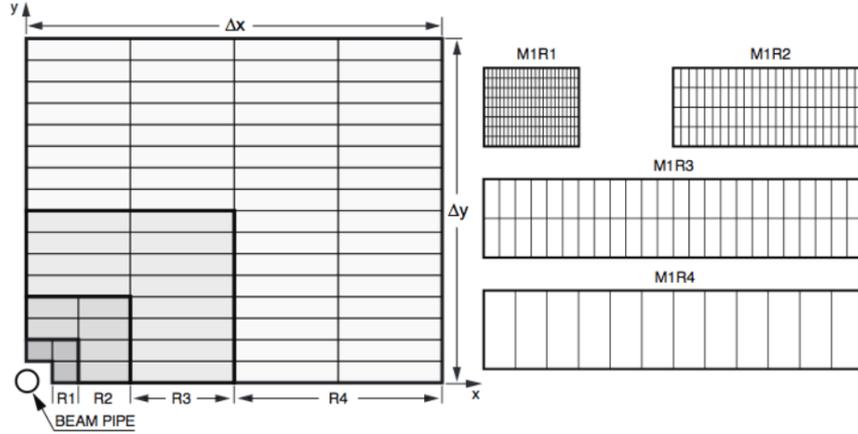


Figure 3.8: Left: four regions of one quadrant, R1-R4. Each rectangle represents a chamber. Right: division into logical pads of a chamber in each region. Taken from [11]

thick scintillator tiles for the shower detection. The design energy resolution is

$$\frac{\theta_E}{E} = \frac{10\%}{\sqrt{E[GeV]}} \oplus 1\%$$

where  $\oplus$  means the summation in quadrature.

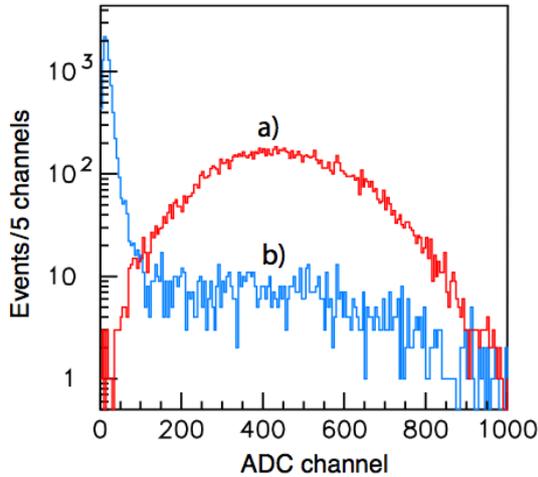
The hadronic calorimeter, **HCAL**, is used to detect showers from hadronic particles. The special feature of the detector is that the orientation of the scintillating tiles is parallel to the beam line. The tiles are intersected with iron absorbers, with a length equal to one hadronic interaction length in steel in longitudinal direction. In the transverse direction the iron absorbers are 1 cm thick. See Figure 3.9b for schematic view of the HCAL. The scintillation light is guided by fibers to the PMTs at the back of the detector. The design energy resolution is

$$\frac{\theta_E}{E} = \frac{80\%}{\sqrt{E[GeV]}} \oplus 10\%$$

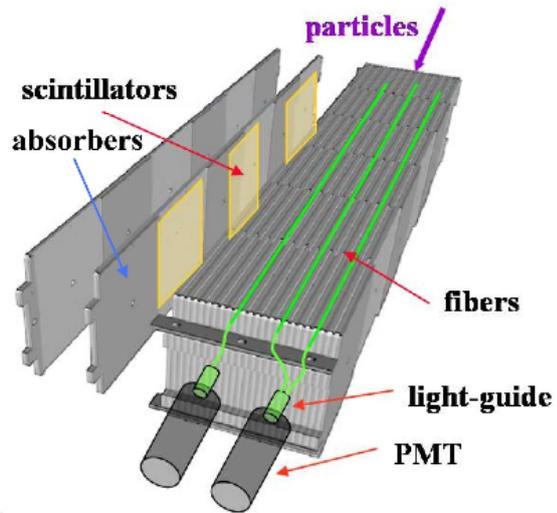
where  $\oplus$  denotes the summation in quadrature.

### 3.2.3 The LHCb Trigger System

LHCb is designed to work at an instantaneous luminosity much lower than the maximum LHC luminosity. This means that there will be fewer events per beam collision, but the LHCb keeps the instantaneous luminosity constant during the data-



(a) Energy deposition of (a) 50 GeV electrons and (b) pions in the PS detector (shown in arbitrary units). Taken from [11]



(b) Schematic view of the HCAL detector cell structure. Taken from [11]

Figure 3.9

taking. To achieve the lower and constant instantaneous luminosity, the proton beams are slightly defocused before interaction point. The LHCb trigger system further reduces the number of events written to storage for offline analysis. The main purpose of the trigger system is to reduce the event rate from  $40\text{MHz}^1$  to about  $5\text{ kHz}$ . The trigger system consists of a Level-0 hardware trigger and two software based High Level Trigger stages. The process is schematically shown in Figure 3.10.

**Level-0 trigger** The Level-0 (L0) trigger reduces the event rate from  $40\text{ MHz}$  to  $1\text{ MHz}$ . The L0 trigger system is fully synchronous with the  $40\text{ MHz}$  interacting rate and is implemented in hardware. The latency of the L0 trigger, i.e. the time between the pp interaction and the arrival of the trigger decision is fixed to be  $4\ \mu\text{s}$ . This time interval includes time-of-flight, cable delays and delays in front-end electronics leaving  $2\ \mu\text{s}$  for the processing of data in the L0 trigger to arrive to a decision.

Level-0 is subdivided into three components - the pile-up system, the Level-0 calorimeter trigger and the Level-0 muon trigger. Each component is connected to

<sup>1</sup> $40\text{ MHz}$  is the nominal machine bunch crossing frequency, for most of 2011-2012 data taking the bunch crossing rate was  $15\text{ MHz}$ .

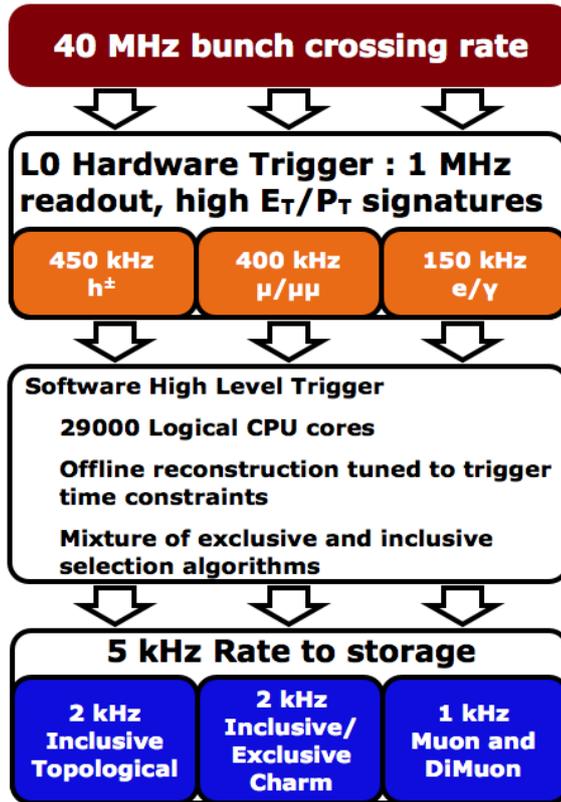


Figure 3.10: Overview over LHCb trigger system. Taken from [17]

the Level-0 decision unit (DU) which collects all the information and evaluates the final decision. The pile-up trigger uses information from the VELO to distinguish events with one of multiple interactions per bunch. This trigger is not used for selection of events in flavour physics analysis. The L0 calorimeter looks for high  $E_T$  particles. The  $E_T$  is measured using  $2 \times 2$  calorimeter cells, high  $E_T$  events are selected and a particle hypothesis ( $\gamma$ , electron or hadron) is assigned. The L0 muon trigger selects two muons with highest  $p_T$  for each quadrant of the detector. The L0 muon processor searches for hits defining a curve through all five stations and pointing towards the interaction point taking the magnet field into account. The  $p_T$  is determined by the position of the track in the first two stations. The event is accepted if either the  $p_T$  of a single track or the combination of the  $p_T$  of two tracks are higher than a threshold (L0Muon and L0DiMuon correspondingly).

**High Level Trigger** The purpose of the High Level Trigger (HLT) is to further reduce the event rate from the 1 MHz output of the hardware L0 stage. The HLT is divided in two stages. The first stage, HLT1, attempts partial event reconstruction

of the events passing the L0 decision. This reduces the rate to about 80 kHz. At this rate HLT2 can perform a full reconstruction similar to the offline reconstruction. Several inclusive and exclusive selections are performed to reduce the rate to 5 kHz. Several different sequences of reconstruction algorithms and selection called trigger lines or alleys are used. Each of these trigger line returns a decision. An event is accepted if it is accepted by at least one trigger line. HLT is completely software based and therefore much more flexible than the L0 trigger.

### 3.3 Data Samples

The data used in the presented analysis was taken by the LHCb detector during 2011 and 2012. During 2011, the LHC was running at a center mass energy of  $s = \sqrt{7}$  and during 2012 at  $s = \sqrt{8}$ . The data corresponds to  $3 \text{ fb}^{-1}$  of integrated luminosity. The data consists of all events that have passed the trigger and reconstruction as well as the centralized LHCb selection called stripping. Chapter 5 describes further the selection that is applied to the data.

**Monte Carlo simulated events** Besides the real data sample the analysis uses samples of simulated events to determine the efficiency of the selection and to study systematic effects due to peaking backgrounds. The LHCb simulation program is called Gauss [9]. Gauss generates the pp collisions and the decay particles produced and tracks the particles through the detector. The simulated events are further discussed in Chapter 6. Table 3.1 shows the Monte Carlo simulated samples (MC) used for the analysis.

Sample	Decay File	Events	Simulation + Processing
$B_s \rightarrow \phi\mu^+\mu^-$	13114006	546322	2012 Sim08c Pythia8, Reco14a-Strip20Filtered
$B_s \rightarrow \phi\mu^+\mu^-$	13114006	535403	2012 Sim08c Pythia6, Reco14a-Strip20Filtered
$B_s \rightarrow J/\psi\phi$	13114001	5114480	2012 Sim08a Pythia8, Reco14a-Strip20Flagged
$B_s \rightarrow J/\psi\phi$	13114001	5129989	2012 Sim08a Pythia6, Reco14a-Strip20Flagged
$B_d \rightarrow K^*J/\psi$	11144001	3014491	2012 Sim08c Pythia8, Reco14a-Strip20Flagged
$B_d \rightarrow K^*\mu\mu$	11114001	517748	2012 Sim08b Pythia8, Reco14a-Strip20Flagged
$B_s \rightarrow D_s\pi$	13264021	5009988	2012 Sim08c Pythia8, Reco14a-Strip20Flagged
$B_s \rightarrow D_s^- \rightarrow (\phi\mu^-\bar{\nu}_\mu)\mu^+\nu_\mu$	11174000	2M	2012 Sim08c Pythia8, generator level
$B_d \rightarrow D^- \rightarrow (K^*\mu^-\bar{\nu}_\mu)\mu^+\nu_\mu$	11174000	2M	2012 Sim08c Pythia8, generator level
$\Lambda_b \rightarrow pK\mu\mu$	15114011		
$\Lambda_b \rightarrow \Lambda(1520)\mu\mu$	15114000		
$B_s \rightarrow \phi\mu^+\mu^-$	13114005	10M	2012 Sim08c Pythia8, generator level

Table 3.1: Monte Carlo simulated samples used in this analysis for efficiencies, acceptances and background determination.

## 4 Analysis Strategy

The goal of the analysis is to determine the differential branching fraction of the decay  $B_s^0 \rightarrow \phi\mu^+\mu^-$  relative to the normalization channel  $B_s^0 \rightarrow J/\psi\phi$ . The branching ratio is determined in bins of the dimuon invariant mass squared,  $q^2$ . The bins are selected such that the signal yield is approximately equal in each bin, see Table 4.1. In addition the yield is determined in the  $1 \text{ GeV}^2 < q^2 < 6 \text{ GeV}^2$  bin as this region is best described by theory and can be used for comparison with theoretical models. As a cross-check the yield is also determined in the  $15 \text{ GeV}^2 < q^2 < 19 \text{ GeV}^2$  bin. The measurements are performed in the  $B_s$  mass peak window  $[5316.3, 5416.3] \text{ MeV}$  (signal region). The  $q^2$  bins and the signal region are shown in Figure 4.1. The analysis can be divided in several main stages:

- The data sample of  $B_0^s \rightarrow \phi\mu^+\mu^-$  is polluted with both physical and combinatorial background events. It is essential to reduce the number of background events as much as possible since the expected number of signal events is very small. A combination of cuts is used to strongly reduce the physical background while the combinatorial background is suppressed using boosted decision trees (BDT), a multivariate analysis tool. This is further discussed in Chapter 5.
- The analysis relies on the use of Monte Carlo simulated events, in particular to determine the efficiency of the signal selection. However generated events do not always accurately simulate distributions of some physical quantities that are required for the analysis. Therefore an explicit correction of the Monte Carlo data is performed so that the difference between simulated events and real events is minimized. This is further discussed in Chapter 6. Another issue that has to be addressed is the choice of the particle identification variables to use.
- After the Monte Carlo simulated events are corrected a study of possible peaking backgrounds is performed. Using Monte Carlo simulated data the amount

of peaking background present in the signal region is estimated and, when possible, reduced using veto cuts. This is further discussed in Chapter 7.

- After all veto cuts have been determined they are included in the selection procedure and the Monte Carlo correction procedure and the data samples are reprocessed.
- The differential branching ratio of the decay is determined using the data samples produced in the previous step. The signal yield is found using an unbinned extended maximum likelihood fit to the  $B_s^0$  mass. The parameters of the fit are determined from the fit of the  $B_s \rightarrow J/\psi\phi$  reference channel. The fit is performed in bins of  $q^2$ . For each bin a separate scaling factor is calculated to account for the dependence of the mass resolution on  $q^2$ . The selection efficiencies for signal and reference channels are calculated using corrected Monte Carlo generated events. Then the differential branching fraction is calculated. The full description of the procedure is given in Chapter 8.
- As the last part of the analysis possible systematical uncertainties are evaluated. This is discussed in Chapter 9.

Bin	Range [GeV <sup>2</sup> ]
1	0.1 – 2.0
2	2.0 – 5.0
3	5.0 – 8.0
4	11.0 – 12.5
5	15.0 – 17.0
6	17.0 – 19.0
7	1.0 – 6.0
8	15.0 – 19.0

Table 4.1:  $q^2$  binning.

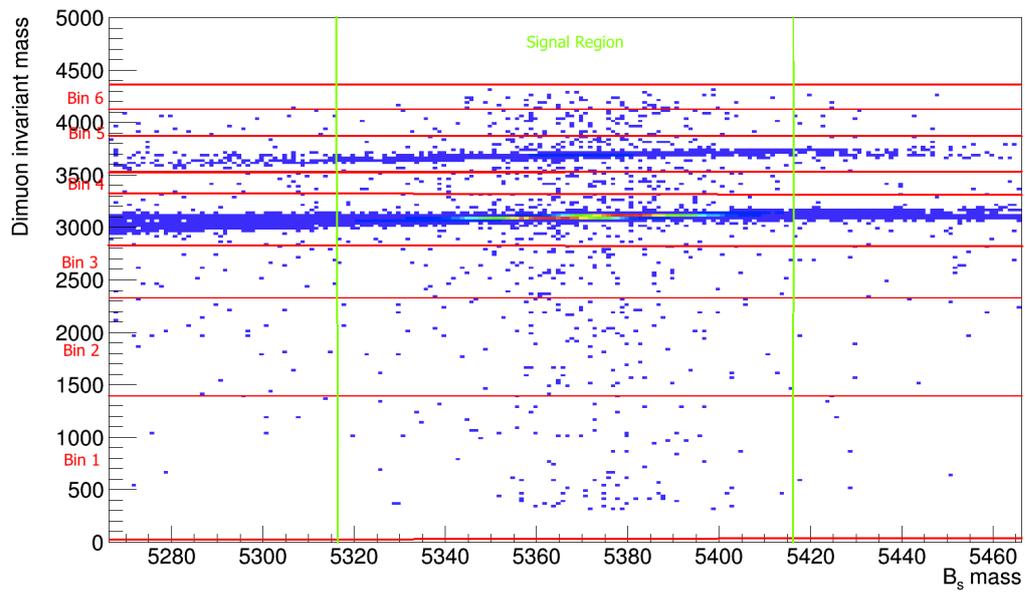


Figure 4.1: Two dimensional plot of the invariant  $B_s$  mass versus the invariant dimuon mass. The signal region in and the  $q^2$  bins in which measurements are performed are marked.

## 5 Data Selection

This chapter describes the selection procedure of  $B_s^0 \rightarrow \phi\mu^+\mu^-$  signal events that will be used for the determination of the differential branching ratio. The purpose of the selection is to separate data and background such that as many signal events remain as possible while as many background events as possible are removed. The selection consists of several steps. At first the event candidate has to pass the LHCb trigger system. Then a preselection is performed centrally for all LHCb data. After that the final selection optimized to achieve optimal signal significance is applied.

For the analysis the decay mode  $B_s \rightarrow \phi(\rightarrow K^+K^-)\mu^+\mu^-$  is used. The corresponding normalization (or reference) mode is  $B_s^0 \rightarrow \phi(\rightarrow K^+K^-)J/\psi(\rightarrow \mu^+\mu^-)$ . The final states of the signal and the reference channels are identical. The invariant dimuon mass is used to separate the muons originating from the resonant  $J/\psi \rightarrow \mu^+\mu^-$  from the non-resonant signal. The signal yield is not measured in the dimuon mass windows corresponding to the  $J/\psi$  and  $\psi(2S)$  resonances.

The final state particles are stable in the sense that they mostly pass the detector without decaying. The final state of the signal decay does not allow to determine the decay flavour of the B meson so events originating both from  $B_0$  and  $\bar{B}_0$  are selected.

### 5.1 Trigger Selection

The first step in the selection of the events is the trigger. Normally an event is accepted by LHCb if it satisfies at least one of each L0, Hlt1 and Hlt2 trigger lines. For this analysis these conditions are tightened. It is required that events trigger at least one of the lines given in Table 5.1 for each of L0, Hlt1 and Hlt2. Those trigger lines were selected since they provide the optimal signal efficiency. Most of the trigger lines used in the analysis select events based on the quality of the muon candidates. For L0 trigger lines the choice depends on the transverse momentum  $p_T$  of the track. For the Hlt1 choice the trigger lines chose events based on the invariant mass of dimuon system or on the  $p_T$  of the muons. The Hlt2 trigger lines

make the decision based on the topology of the tracks. The triggered on signal (TOS) requirement means that the lines are triggered explicitly by the decay product of the signal candidate.

## 5.2 Centralized Event Selection

All events at LHCb pass a centralized selection of events called stripping. A stripping line is a sequence of selections used to create candidates and select final events, similar to trigger lines. A generic stripping line used for many  $b \rightarrow s\mu\mu$  decays called B2XMuMu is also used for the presented analysis. This stripping line selects events with one primary vertex and two muons in the final state. Loose cuts are applied on kinematical and PID variables of both mother and daughter particles.

There are more than  $16 \cdot 10^6$  events that are contained in the B2XMuMu stripping line. All events are interpreted as  $B_s \rightarrow J/\psi(\rightarrow \mu^+\mu^+)\phi(\rightarrow K^+K^-)$ . It is however clear that a large fraction of these events are misidentified background events, also see Figure 5.1. Further cuts are needed to separate signal from background. It is important to mention that it is non-trivial to separate  $B_s \rightarrow J/\psi(\rightarrow \mu^+\mu^+)\phi(\rightarrow K^+K^-)$  from  $B_s \rightarrow \phi(\rightarrow K^+K^-)\mu^+\mu^-$  events. However at this stage it is not a problem since the reference channel is used for the determination of the branching fraction and has to be selected anyway.

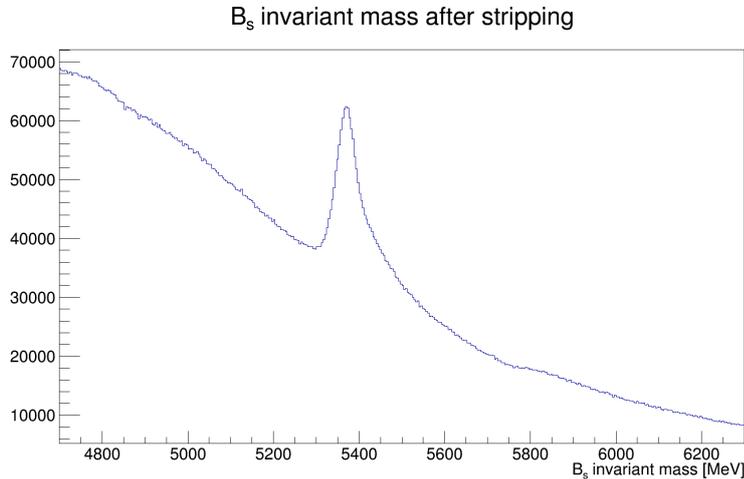


Figure 5.1: Distributions of the invariant mass of the  $B_s \rightarrow \phi\mu^+\mu^-$  candidates after the stripping.

L0	B0_L0MuonDecision_TOS B0_L0DiMuonDecision_TOS
Hlt1	B0_Hlt1TrackAllL0Decision_TOS B0_Hlt1TrackMuonDecision_TOS B0_Hlt1DiMuonLowMassDecision_TOS B0_Hlt1DiMuonHighMassDecision_TOS B0_Hlt1SingleMuonHighPTDecision_TOS
Hlt2	B0_Hlt2Topo(2,3,4)BodyBBDTDecision_TOS B0_Hlt2TopoMu(2,3,4)BodyBBDTDecision_TOS B0_Hlt2SingleMuonDecision_TOS B0_Hlt2DiMuonDetachedDecision_TOS B0_Hlt2DiMuonDetachedHeavyDecision_TOS

Table 5.1: Trigger line used for the selection of the signal decay

Variable	Cut
$K^-$ PIDK	$PIDK(K^-) > -3$
$K^+$ PIDK	$PIDK(K^+) > -3$
$m_\phi$	$1007.455\text{MeV} < m_\phi < 1031.455\text{MeV}$
$m_{B_s}$	$5100\text{MeV} < m_{B_s} < 5800\text{MeV}$
$\mu K$	$ m(K^\pm, \mu^\mp \text{ as } K^\mp) - 1019.455  > 8\text{MeV}$
$\mu K$	$ m(u^\pm, K^\mp \text{ as } \mu^\mp) - 3096.916  > 50\text{MeV}$
$KK$	$5575\text{MeV} < m(K^\pm, K^\mp \text{ as } p^\mp) < 5665\text{MeV}$ and $(PIDp(K^\mp) - PIDK(K^\mp)) > 10$

Table 5.2: Selection requirements for the cut based selection

### 5.3 Cutbased Selection Procedure

After the trigger and the preselection a significant amount of background events still remain. Another problem is the contribution from the non-resonant  $B_s \rightarrow J/\psi K^+ K^-$  decay channel (so called S-wave). To select only final states with a real  $\phi$  decay the fact that  $\phi$  has a very narrow mass peak is used. A very tight cut of  $\pm 12\text{MeV}$  on the  $\phi$  invariant mass allows keeps most of the signal events and reduces the amount of background by a few orders of magnitude. The S-wave contribution in the tight  $\phi$  mass window is expected to be  $\sim 2\%$  [23] for the  $B_s \rightarrow J/\psi \phi$  candidates. For the  $B_s \rightarrow \phi \mu \mu$  selection a similar contribution is assumed. A wide cut on the invariant mass of  $B_s$  is applied to reduce the amount of events and improve performance. The final set of cuts (including vetoes) is given in Table 5.2.

### 5.3.1 Vetoes

In addition to simple mass cuts there are additional vetoes applied against specific backgrounds. The cut on  $\text{PIDK(K)}$ <sup>1</sup> is done to reject pions misidentified as kaons from  $B_d^0 \rightarrow K^*(\rightarrow \pi K)\mu^+\mu^-$  decays. Another possible source of peaking background is if a muon from a  $B_s^0 \rightarrow J/\psi\phi$  is misidentified as kaon and a kaon as a muon. This would change the invariant mass of the dimuon candidates which is the only separation criterion between  $\phi\mu^+\mu^-$  and  $J/\psi\phi$  decays. To reject this misidentification background two mass hypothesis are tested. First the kaon mass is assumed for a reconstructed muon. If the invariant mass of the  $K^\pm$  and the  $\mu^\mp$  as  $K^\mp$  candidate is within 8MeV of the nominal  $\phi$  mass<sup>2</sup> the event is rejected. Then the muon mass is assumed for a reconstructed kaon candidate. If it lies within 50 MeV of the nominal  $J/\psi$  mass<sup>3</sup> the event is also rejected. The third possible peaking background comes from  $\Lambda_b \rightarrow \Lambda(1520)(\rightarrow pK)\mu^+\mu^-$  where the proton is misidentified as a kaon. To suppress this background a combination of mass cuts and PID cuts is used. The event is rejected if the invariant mass of the  $B_s$  candidate where one of the kaon candidates is assumed to have proton mass lies within 45MeV of the nominal  $\Lambda_b$  mass<sup>4</sup>. In addition it is required that the difference of  $\text{PIDp}$  and  $\text{PIDK}$  of the kaon candidate is larger than 10. The effect of the vetoes as well as the peaking background contribution of different decays are discussed in Chapter 7, they are included here for completeness.

## 5.4 Multivariate Selection Procedure

After the cut based selection (including the vetoes) around 90000 signal candidates remain. The invariant mass of the  $B_s$  candidates versus invariant mass of the dimuon system can be plotted to investigate how those events are distributed, see Figure 5.2. There is a clear signal visible everywhere outside of the *charmonium* bins, that is outside of  $J/\psi(1S)$  and  $\psi(2S)$  muons. There is still combinatorial background left, that is  $B_s \rightarrow \phi\mu^+\mu^-$  candidates reconstructed from final states of multiple different decays. The problem with such background is that there is no straightforward method to suppress it.

---

<sup>1</sup> $\text{PIDK(K)}$  is a rather abstract value that describes the probability of a kaon candidate to really be a kaon. Further explanations about the PID variables are given in Chapter 6.

<sup>2</sup> $1019.461 \pm 0.019$  MeV [12]

<sup>3</sup> $3096.916 \pm 0.011$  MeV [12]

<sup>4</sup> $5619.5 \pm 0.4$  MeV [12]

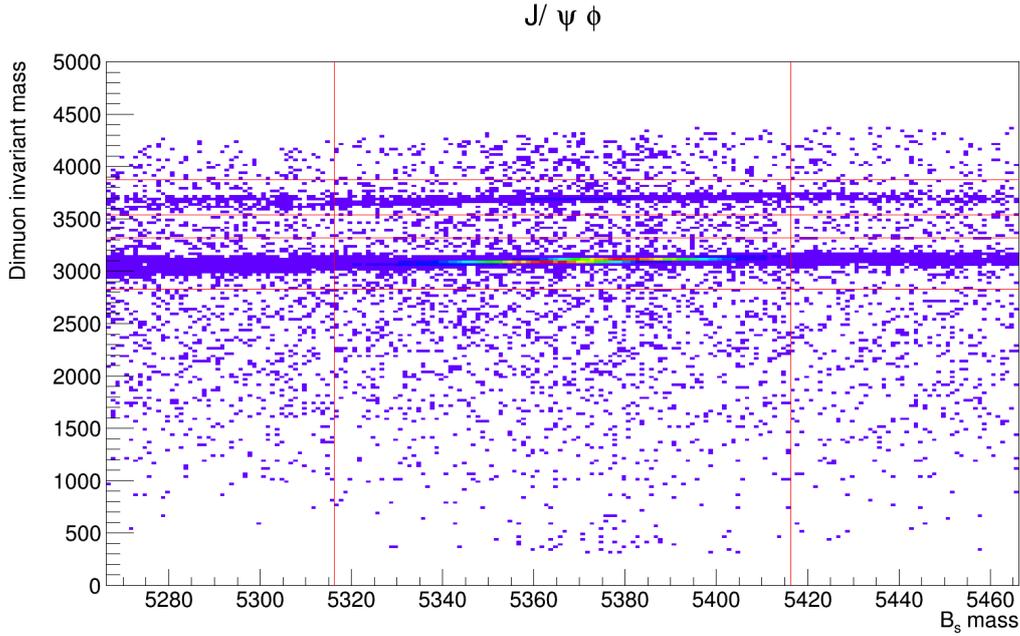


Figure 5.2: Distributions of invariant mass of  $B_s$  versus the invariant dimuon mass after the cutbased selection after the cutbased selection.

One wants to somehow separate signal events from background. A very simplistic approach would be to define some measure of goodness of separation, for example signal significance,  $S/\sqrt{S+B}$ , where  $S$  is the signal yield and  $B$  is the background yield). When this quantity is maximal the separation is the best. In order to find the best selection one changes one variable at a time until a local  $S/\sqrt{S+B}$  maximum is found. Then the next variable is changed until a local maximum is found etc. The problem is that many of the variables are correlated and in this case the method doesn't work reliably. In particle physics a usual way to handle such multidimensional problems is to use a multivariate analysis (MVA), a mathematical technique used for simultaneous observation and analysis of more than one outcome variable. A package implementing many different MVA techniques called Toolkit for Multivariate Data Analysis (TMVA) [20] is used in this analysis.

### 5.4.1 Boosted Decision Trees

In this analysis the suppression of the combinatorial background is performed by a boosted decision tree (BDT) [21]. A BDT is one of the methods implemented in the TMVA tool. A decision tree is a collection of if-statements that return 0 or 1. In the case of TMVA the decision trees are obtained by machine learning.

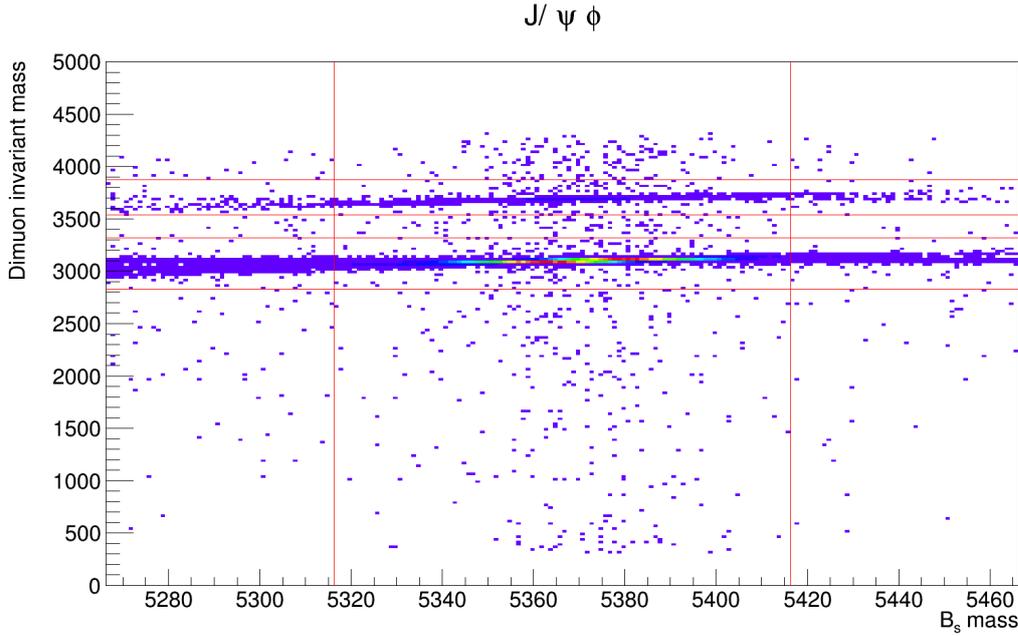


Figure 5.3: Distributions of invariant mass of  $B_s$  versus the invariant dimuon mass after the cutbased and the BDT selection has been applied.

A boosted decision tree is a collection of decision trees (forest) that is averaged over. The process works as follows - first one decision tree is created. This tree has limited precision. Now the data is reweighted - misidentified events gain weight and correctly identified events lose weight. A new tree is trained on the data, it now focuses on the problems that were not addressed by the first tree. A few hundred trees are trained in this way (400 is a default value in TMVA). This process is called boosting. For the analysis the AdaBoost algorithm [25] is used. Given two algorithms that make the signal-background separation more effective than a random choice the AdaBoost algorithm can always combine them to an algorithm that is more effective than either of the original algorithms. So even if all of the trees in the forest have very low efficiency the forest itself can have a high efficiency. This makes AdaBoost a very good out-of-the-box choice for MVA analysis as it requires very little adjustment.

For machine learning one needs a sample of signal events and background events. Normally Monte Carlo simulated events are used as signal sample, but in case of the  $B_s \rightarrow \phi\mu^+\mu^-$  signal decay there is an easier alternative - real  $B_s \rightarrow J/\psi\phi$  events are used as signal proxy. There are slight kinematical differences between the reference and the signal decays, but they are negligible for the purpose of training the BDT.

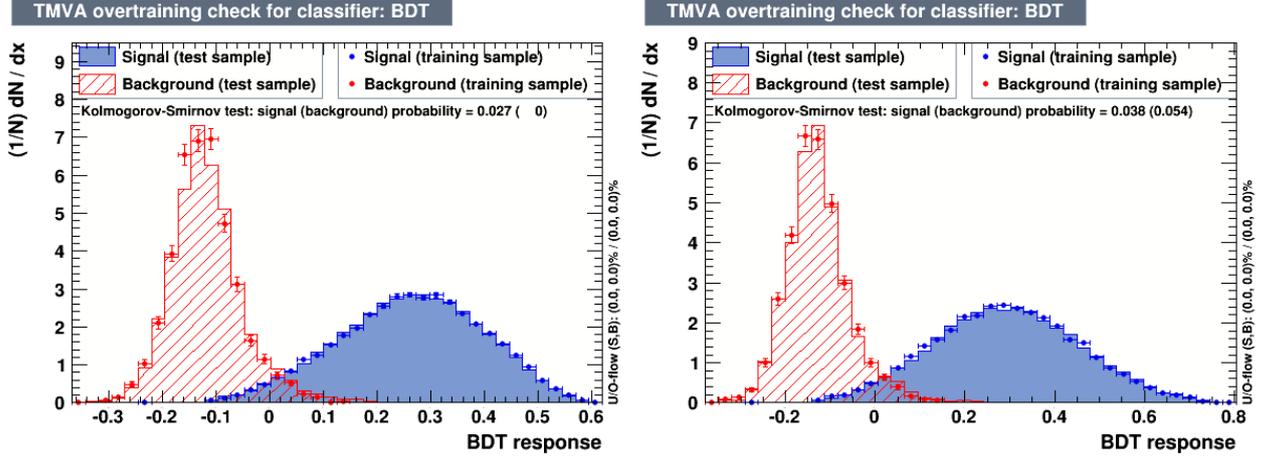


Figure 5.4: BDT response distributions for the signal and the background training and control samples.

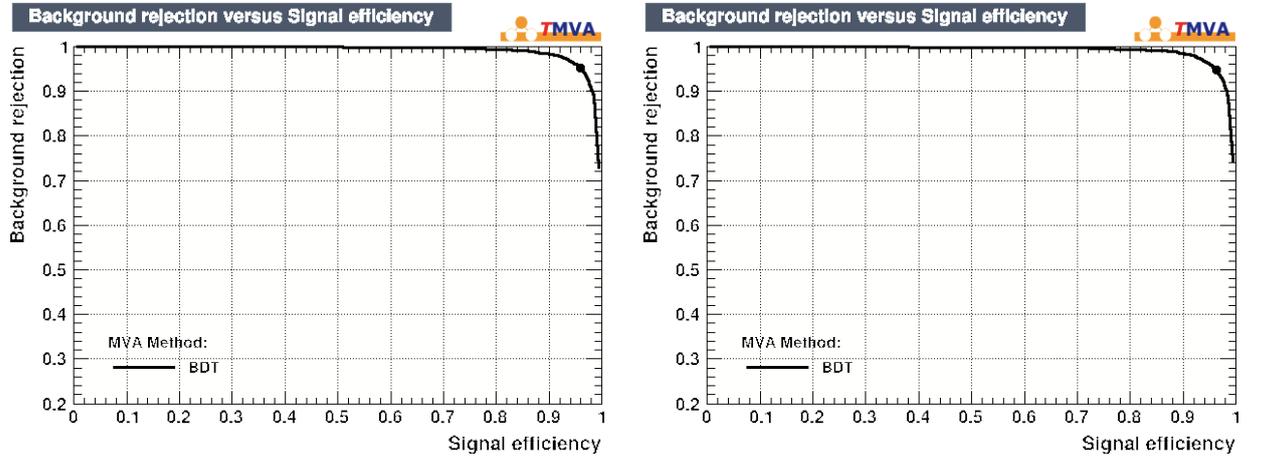


Figure 5.5: Signal efficiency vs background rejection ROC curve. The BDT cut point is given as black dot.

As the signal for the BDT machine learning the sWeighted  $^5 B_s \rightarrow J/\psi\phi$  reference decay is used. For the background the mass sidebands are used, that are events that lie outside the  $B_s$  mass window  $m_{B_s} \in [5166.3 - 5566.3]$ . Those events are not used anywhere else in the analysis. To get more background events the  $\phi$  mass cut is relaxed to  $|m_\phi - 1019.455| < 50\text{MeV}$ .

<sup>5</sup>sWeighting technique [24] is a statistical technique that is designed to explore data samples consisting of several sources of events (signal and background in our case) merged into a single sample. This is further discussed in Chapter 6. Using variables for which distributions of all the sources of events are known (e.g.  $B_s$  invariant mass) sWeighting allows to reconstruct distributions for other variables independently for each source of events without any knowledge of the variables as long as the variables are uncorrelated.

Particle	Variable
$K^\pm$	$\min(PIDK(K^+), PIDK(K^-))$
$K^\pm$	$\max(PIDK(K^+), PIDK(K^-))$
$K^\pm$	IP $\chi^2$
$\mu^\pm$	$\min(PIDmu(\mu^+), PIDmu(\mu^-))$
$\mu^\pm$	IP $\chi^2$
$B_s$	Vertex $\chi^2$
$B_s$	IP $\chi^2$
$B_s$	FD $\chi^2$
$B_s$	DIRA
$B_s$	$p_T$

Table 5.3: Input variables used in BDT

Variables that were used for the BDT training are listed in Table 5.3. Figures A.1 - A.2 shows the distributions of the training variables for the signal and background samples. The data sample is divided in two equal parts. The BDT is trained on one sample and applied to the other sample. This is called cross-training. The reason to do so is that if the BDT is trained on the whole sample it would be very effective for the sample on which it was trained. However there wouldn't be any independent sample to determine the signal efficiency and to cross-check the BDT. The use of cross-trained BDTs solves this problem at the cost of slightly decreased efficiency.

In TMVA the BDT selection assigns an abstract value called BDT response to every event. As a rule the background events have lower BDT response values than the signal events. Figure 5.4 shows the BDT response of the training sample and the control sample for each of the BDTs trained. After the BDTs are trained and applied to the data samples a BDT response cut value needs to be selected. All values with the BDT response below the cut value are rejected as background. The optimization of  $S/\sqrt{S+B}$  in the signal region is performed to determine the cut value. The number of expected signal events is calculated using the relative branching ratio  $\frac{Br(B_s \rightarrow \phi\mu\mu)}{Br(B_s \rightarrow J/\psi\phi)} = 0.00113$  [7]. The expected background is determined by interpolation from sidebands. The corresponding receiver operating characteristic (ROC) curves are shown in Figure 5.5. The resulting cuts that maximize  $S/\sqrt{S+B}$  are determined for each BDT separately and applied to the data samples. That is the last step of the selection procedure. The distribution of events after the full selection including BDT can be seen in the Figure 5.3. Comparing Figure 5.3 with Figure 5.2 it can be seen that the combinatorial background has been decreased by the BDT selection.

## 6 Efficiency determination using simulated events

To determine the efficiency of the selection Monte Carlo (MC) simulated events are used. However Monte Carlo simulations are not perfect, not all distributions of variables match those of the real data. So corrections have to be performed in order to make the difference between real data and MC minimal. The variables that have to be checked (and corrected if needed) are the ones used in the selection, see Table 5.3. Since only a small number of real  $B_s \rightarrow \phi\mu^+\mu^-$  events are available for comparison  $B_s \rightarrow J/\psi\phi$  decay channel and corresponding Monte Carlo simulations are used instead.

**LHCb Monte Carlo simulation** The LHCb experiment simulation software is Gauss [9]. It mimics the processes that occur in LHCb. That is done in two phases. In the generator phase pp collisions are generated and then the subsequent hadronisation is simulated. In the subsequent simulation phase the particles are tracked through the detector and the detector response is simulated. This is done using the Geant4 toolkit [1]. For the generation different generators are available. For the MC samples used in the analysis Pythia8 [2] was used in the generation stage. The simulation conditions used are Sim08 Beam4000GeV-2012-Nu2.5. That means that Sim08 version of simulation conditions is used, collision of 4TeV protons with average 2.5 pp interactions per bunch are simulated.

**sWeighting** To compare Monte Carlo simulated events with real data one needs to extract pure signal distributions from the real data. The technique used for that in the analysis is called sPlot [24]. It assumes that events from different sources are mixed together in one sample (signal and background). Furthermore it is assumed that those events are characterized by two sets of variables, discriminating variables that are known and can be used to separate the two samples and control variables which are not known. If discriminating variables are uncorrelated with the control variables one can infer the distributions of control variables for the individual sources

of events. In the case of  $B_s \rightarrow J/\psi\phi$  decay the invariant  $B_s$  mass is used as the discriminating variable, a model using a double crystal ball shape for signal and an exponential for the background is used, and every event is assigned a weight. The weighted events represent the contribution from signal only.

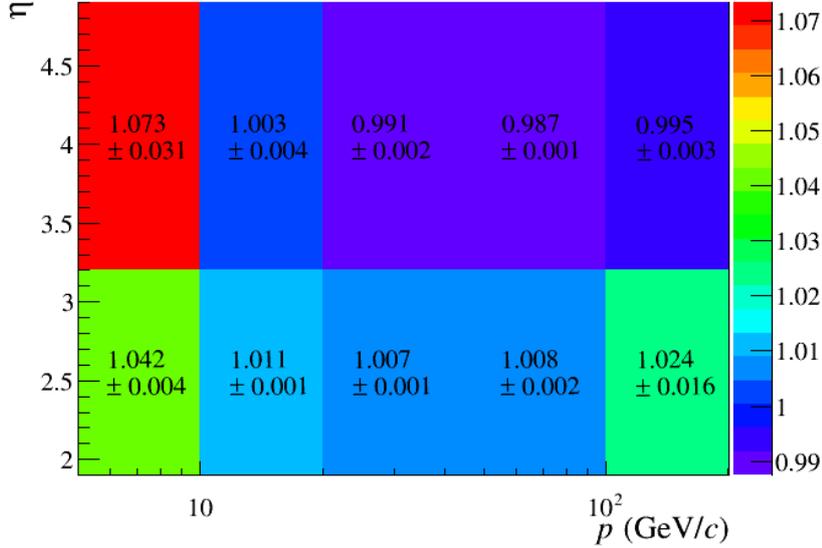


Figure 6.1: Ratios of muon ID efficiencies in bins of momentum and transverse momentum.

## 6.1 Track Efficiency

The Monte Carlo simulations do not always perfectly describe the tracking efficiency. The tracking efficiency is dependent on momentum and pseudorapidity of the particle. The tracking efficiency for  $J/\psi \rightarrow \mu^+\mu^-$  events is measured in bins of momentum and pseudo rapidity for both real data and MC and the ratio of the efficiencies is determined. Figure 6.1 shows the resulting ratios of tracking efficiency of data over Monte Carlo,  $\frac{\epsilon_{track}(data)}{\epsilon_{track}(MC)}$  that are used as weights. Those weights are assigned to the Monte Carlo tracks that lie in the corresponding bins. To correct each  $B_s \rightarrow \phi\mu^+\mu^-$  Monte Carlo simulated event the product of weights of the four tracks is used for each event.

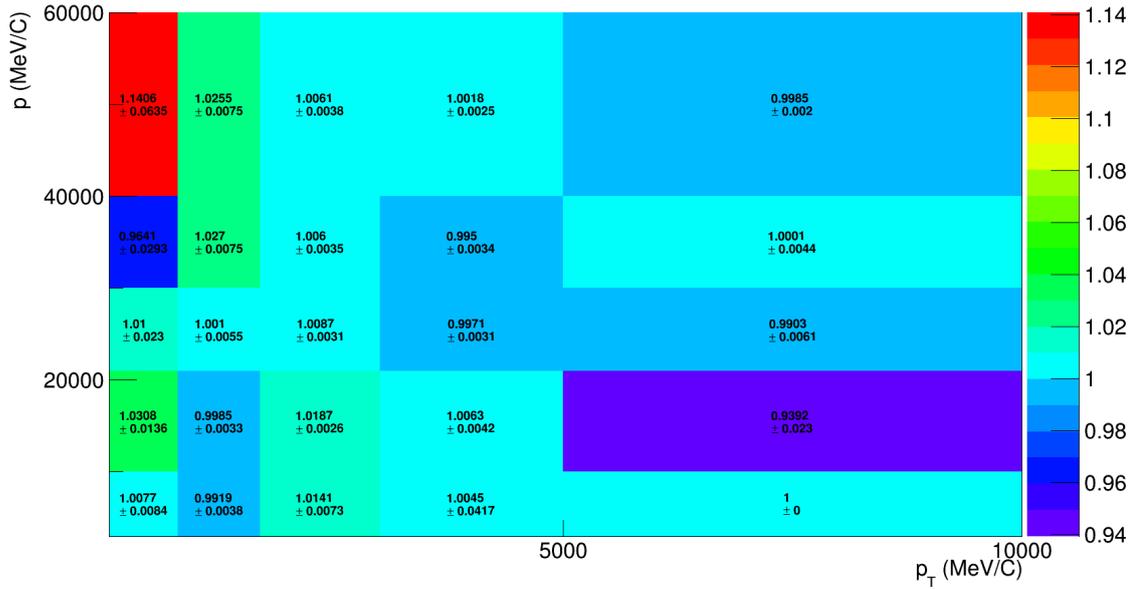


Figure 6.2: Ratios of tracking efficiencies in bins of momentum and pseudorapidity.

## 6.2 Muon ID Efficiency

The muon identification algorithm at LHCb consists of two parts. First a boolean decision called  $IsMu\mu on$  is applied to the tracks. For a given track and field of interest (FOI)  $IsMu\mu on$  is true if the track has at least one hit in FOI in a certain number of muon stations (depending on momentum of the track). In the second step the proper PID variable is calculated. The B2XMuMu line requires  $IsMu\mu on = 1$  so it is a relevant variable to correct for inefficiencies. The  $IsMu\mu on$  efficiency is determined in bins of momentum and transverse momentum. Figure 6.2 shows the ratios of  $\frac{\epsilon_{IsMu\mu on}(data)}{\epsilon_{IsMu\mu on}(MC)}$  that are assigned as weights to the Monte Carlo simulated muons. The weight assigned to each event is once again the product of weights assigned to the two muons.

## 6.3 PID Resampling

The simulation of the particle identification performance, in particular that of RICH detectors, is not accurate. There are two different types of particle identification (PID) variables, CombDLL and ProbNN, available for LHCb events. For PID particle likelihoods are calculated separately by RICH, CALO and the muon system.

A combined likelihood  $L$  can be determined for each particle type by multiplying the RICH, CALO and muon likelihoods. The probability for a particle to be e.g a kaon is related to the combined  $\Delta \log L$  (CombDLL) PID variable defined as

$$\Delta \ln L_{K\pi} = \ln L(K) - \ln L(\pi).$$

A shorthand notation for this value is PIDK. This is a straightforward way to define PID variables. The problem with such definition is however that the value of the variable by itself doesn't have any meaning. One would prefer to have PID variables that could be interpreted as probability for a track to be the corresponding particle. The ProbNN PID variables are created using neural networks, one for each particle hypothesis and the output is a value from 0 to 1 that can be interpreted as Bayesian probability.

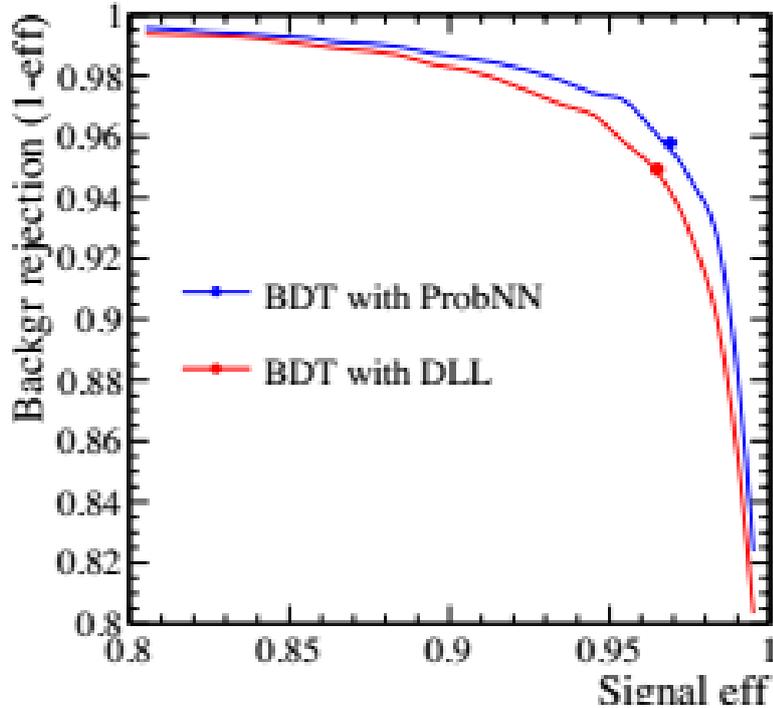


Figure 6.3: Comparison of BDTs trained using ProbNN and CombDLL PID variables. The BDT cut points are given as dots.

The ProbNN PID variables are expected to behave better than the CombDLL on data. In case of this analysis the selection procedure has been performed using either ProbNN or CombDLL and the ProbNN PID variables were indeed found to behave marginally better. Figure 6.3 shows the signal efficiency and the background

rejection efficiency of the BDTs trained using the ProbNN and the CombDLL PID variables. The BDT trained using the ProbNN PID variables has higher signal efficiency and background rejection than the BDT trained using the CombDLL variables. One therefore would ideally use the ProbNN variables, however simulating them is challenging.

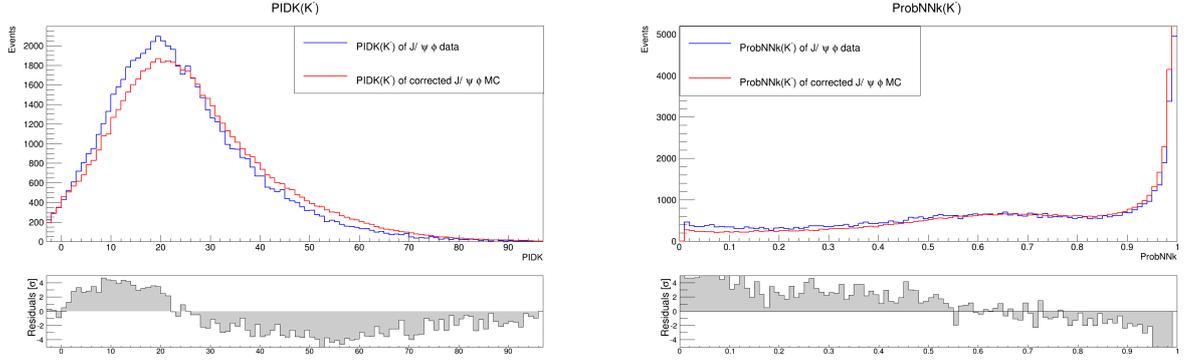
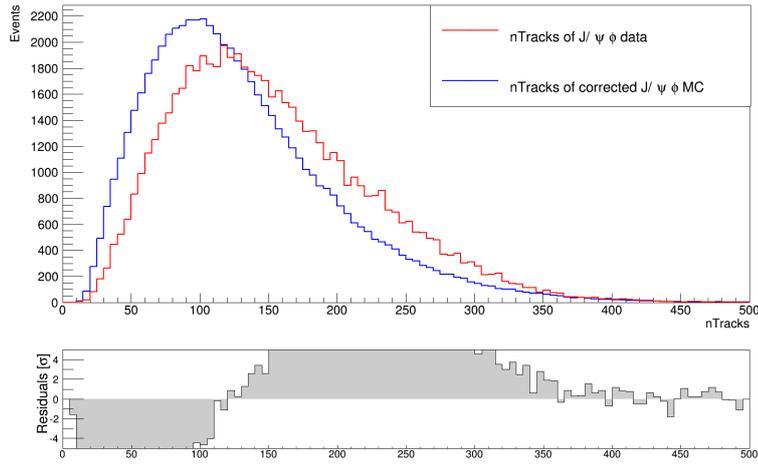
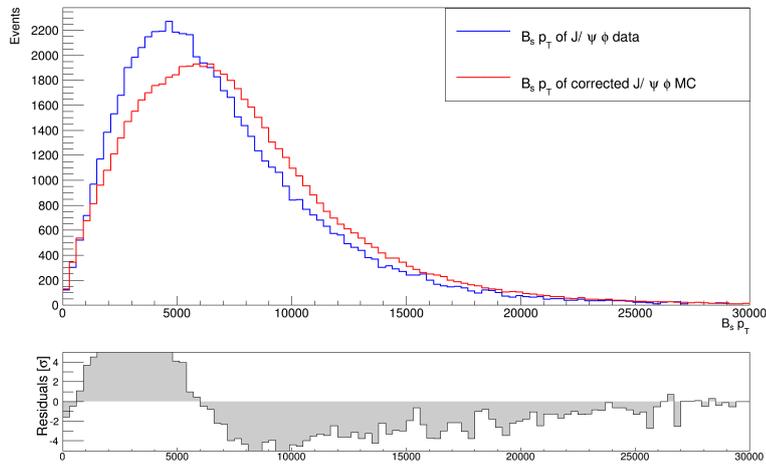


Figure 6.4: Distributions of PID variables of  $K^-$  in the real data and in the Monte Carlo generated sample after the PID resampling. To the left is the distribution of the  $\text{PIDK}(K^-)$  variable, to the right is the distribution of the  $\text{ProbNNk}(K^-)$  variable.

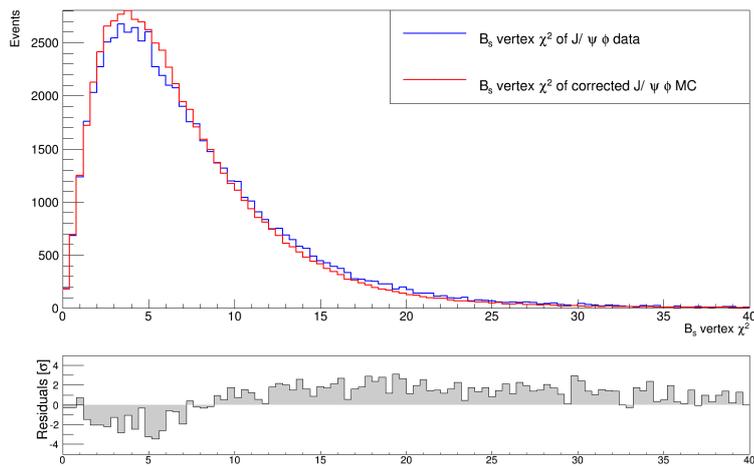
For the Monte Carlo simulated events a correction of PID variables to account for differences between data and Monte Carlo simulation is applied. To perform the correction the PIDCalib tool [4] is used. The main idea behind the tool is the assumption that the PID variables are strongly dependent on momentum and pseudorapidity ( $\eta$ ) of the particle and on the track multiplicity. One can use abundant (and relatively clean) decay modes such as  $D^*(2010)^+ \rightarrow D(\rightarrow K^- \pi^+) \pi_s^+$  and  $J/\psi \rightarrow \mu^+ \mu^-$  to extract distributions of PID variables in bins of momentum,  $\eta$  and event multiplicity. Then one can resample the PID variables in the Monte Carlo data, that is to completely discard the old values and to assign new values chosen at random from the corresponding bins. For the analysis 20 bins in momentum, 4 bins in  $\eta$  and 4 bins in track multiplicity are used. If a particle is outside of the bins range the old PID values are kept. However even after the corrections are applied there are significant differences in PID variables for the data and the Monte Carlo simulation. A comparison of  $\text{PIDK}(K)$  and  $\text{ProbNNk}(K)$  for sWeighted  $B_s \rightarrow J/\psi \phi$  reference events in data and for Monte Carlo generated events are shown in Figure 6.4.



(a) Distributions of track multiplicity ( $n_{\text{Tracks}}$ ) in the real data and MC generated data.



(b) Distributions of  $B_s p_T$  in the real data and MC generated data.



(c) Distributions of  $B_s$  vertex  $\chi^2$  in the real data and MC generated data.

Figure 6.5

## 6.4 Further Corrections

After the reweighting some of the variables used in the BDT and for the resampling of the PID variables still show significant disagreement between data and Monte Carlo simulation. Figures 6.5a - 6.5c show transverse momentum of  $B_s$ , vertex  $\chi^2$  of  $B_s$  and event multiplicity from data and MC. To correct for these discrepancies each of the variables is reweighted. The distributions of each of the variables for the simulations are divided in bins and assigned weights to match the distributions of the corresponding variable in the corresponding bins for the data. Since the variables are not highly correlated such an approach gives better precision than a simultaneous reweighting in three dimensions which is limited by the statistical significance. It is also expected that the reweighting of the track multiplicity is going to improve the distributions of PID variables since they are assigned in bins of the track multiplicity.

## 6.5 PID Variables

After the described reweighting most of the variables relevant for the analysis look reasonable in Monte Carlo simulated data. However the PID variables are still not perfect. Figure 6.6 shows the distributions of the ProbNNk and CombDLLk of the  $K^+$  candidate for both data and simulation. Distributions of CombDLL(K) do look like they reproduce the general shape of the data, but for ProbNNk(K) even that is not true. The PID variables are used for the BDT training so the question is how is the BDT selection affected. Figure 6.7 shows output of the two different BDTs, trained using the CombDLLs or ProbNNs as the PID variables. The distributions of the BDTs trained using CombDLLs is much more consistent. However training the BDTs using ProbNNs gives a somewhat better efficiency (as was already shown in Figure 6.3) so it would be preferential to be able to use the ProbNN PID variables.

The main difference between MC and real data seems to be in the lower values of the ProbNNk(K). So the first thing to check is whether there is any background contribution that causes the difference between MC and real data. The cut on the invariant  $\phi$  mass can be used to suppress the possible background. The largest contribution to the background is expected to be from  $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*(\rightarrow \pi K)$  where pion is misidentified as kaon. Applying the full selection procedure to the  $B^0 \rightarrow J/\psi(\rightarrow \mu\mu)K^*(\rightarrow \pi K)$  MC around 230 events are estimated that pass

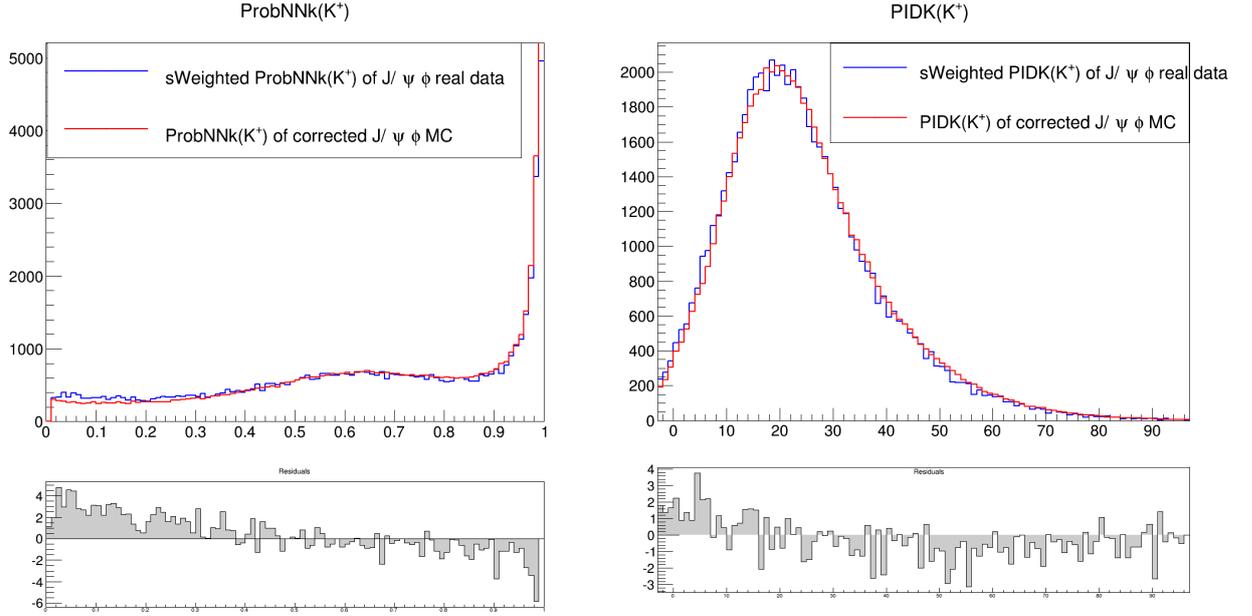


Figure 6.6: Distributions of PID variables of  $K^+$  in the real data and MC generated data. To the left are the distributions of ProbNNk, to the right are distributions of PIDK.

the selection of data. Introducing stronger cuts of  $m_\phi \pm 4\text{MeV}$  and  $m_{B_s} \pm 50\text{MeV}$  decreases that the background yield by a factor of three while the signal yield is only decreased by a factor of 1.33. So if the unsatisfactory performance of the ProbNNs was due to background an improvement should be seen. However there doesn't seem to be a large difference after the cuts have been introduced, see Figure 6.8a.

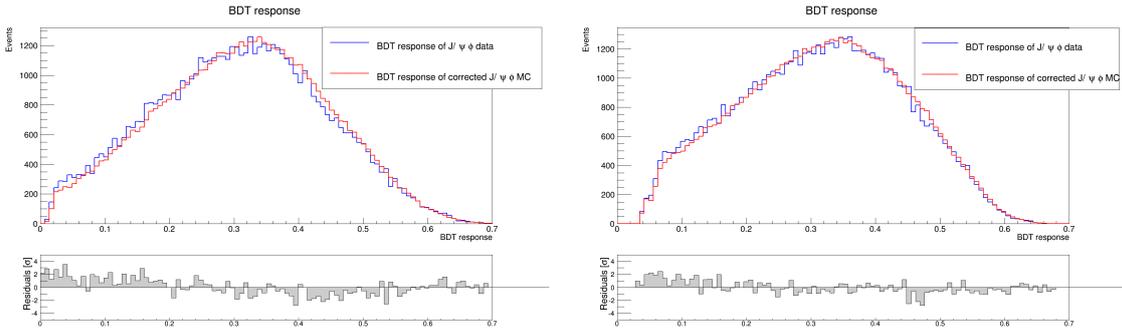
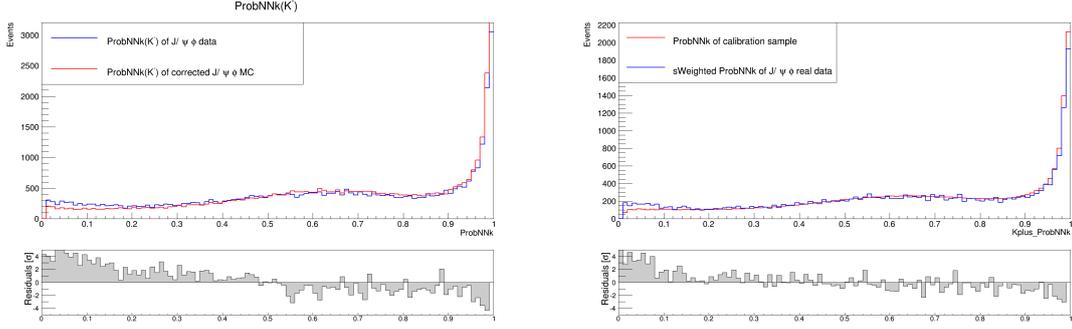


Figure 6.7: BDT output after processing of MC simulated data compared to BDT output of real data.

Next step is to check whether the tool itself works properly, i.e. whether the distributions that are produced by the tool really are similar to those extracted



(a) Distributions of ProbNNk( $K^-$ ) in the real data and MC generated data after suppressing the possible background. (b) Distributions of ProbNNk( $K^+$ ) in the real data and calibrations samples reweighted in  $p$ ,  $\eta$ , ntracks.

Figure 6.8

from calibration samples. For this the calibration samples of  $D^*(2010)^+ \rightarrow D(\rightarrow K^-\pi^+)\pi_s^+$  are reconstructed, sWeighted to separate signal to background and then a 3d reweighing in momentum, pseudo rapidity and track multiplicity is performed to compare calibration samples in data to the ProbNNk(K) for  $B_s \rightarrow \phi\mu^+\mu^-$  data. However the resulting shape of ProbNNk(K) for the calibration sample shows exactly the same discrepancies from the data as the ProbNNk(K) for the resampled MC. Figure 6.8b shows the comparison of the ProbNNk( $K^+$ ) distributions for the calibration sample and the  $B_s \rightarrow J/\psi\phi$  data. It can be seen that the discrepancies are similar to those in Figure 6.6. The conclusion is that the tool works properly.

Another possibility that is worth investigating is that ProbNNk(K) should also be reweighted in other variables, not only in  $p$ ,  $\eta$  and number of tracks for the K. First the track  $\chi^2$  of kaons is checked as it is easily accessible. The distribution of track  $\chi^2$  for the  $B_s \rightarrow \phi\mu^+\mu^-$  data and for the calibration samples is given in Figure 6.9. There is indeed some difference in the two distributions. It was therefore tried to reweight in track  $\chi^2$  and to see whether that changes anything. Figure 6.9 shows that reweighting in track  $\chi^2$  for the K does not have any effect on the shape of the ProbNNk(K).

At this point it was decided to use CombDLL PID variables. The resampling tool works well with CombDLL based variables and the resulting distributions of PIDK(K) and PIDmu( $\mu$ ) are shown in Figures 6.10a - 6.10d. The combinations of PID variables that are used to train the BDTs are given in Figure A.5.

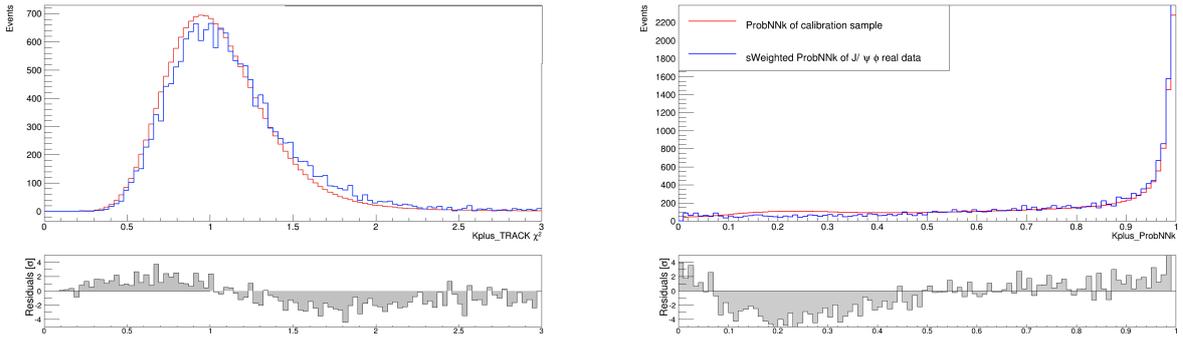
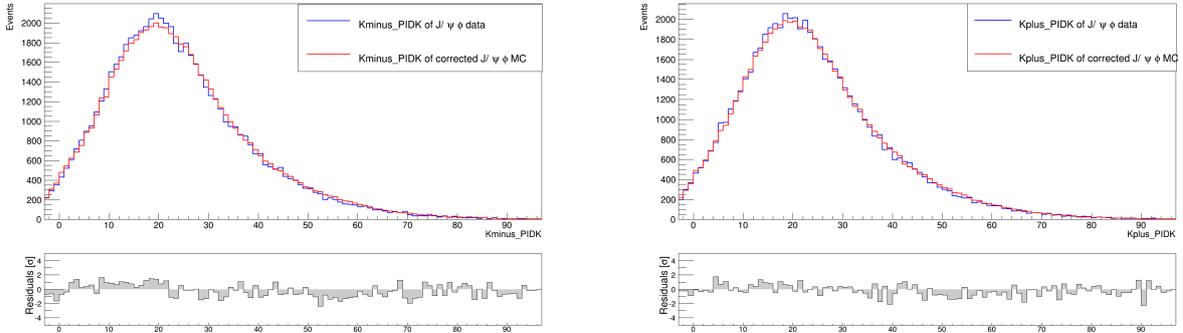
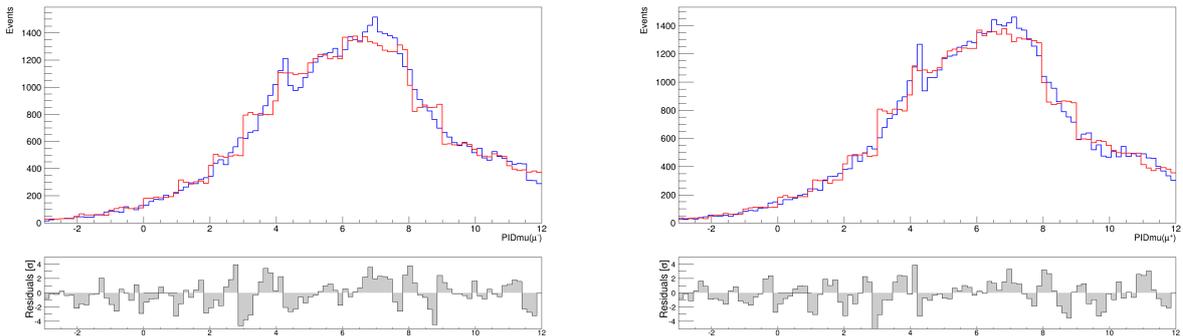


Figure 6.9: To the left - comparison of track  $\chi^2$  of  $K^+$  of the  $B_s \rightarrow \phi\mu^+\mu^-$  data and the calibration samples. To the right - comparison of the ProbNNK( $K^+$ ) distributions for the  $B_s \rightarrow \phi\mu^+\mu^-$  data and the calibration samples with the track  $\chi^2$  weights applied.



(a) Distributions of PIDK( $K^-$ ) in the real data and Monte Carlo generated data after all corrections have been applied. (b) Distributions of PIDK( $K^+$ ) in the real data and Monte Carlo generated data after all corrections have been applied.



(c) Distributions of PIDmu( $\mu^-$ ) in the real data and Monte Carlo generated data after all corrections have been applied. (d) Distributions of PIDmu( $\mu^+$ ) in the real data and Monte Carlo generated data after all corrections have been applied.

# 7 Background Contamination From B Decays

In this chapter sources of possible peaking background are discussed. The number of background events in the signal mass window is estimated and the effects of various vetoes are evaluated.

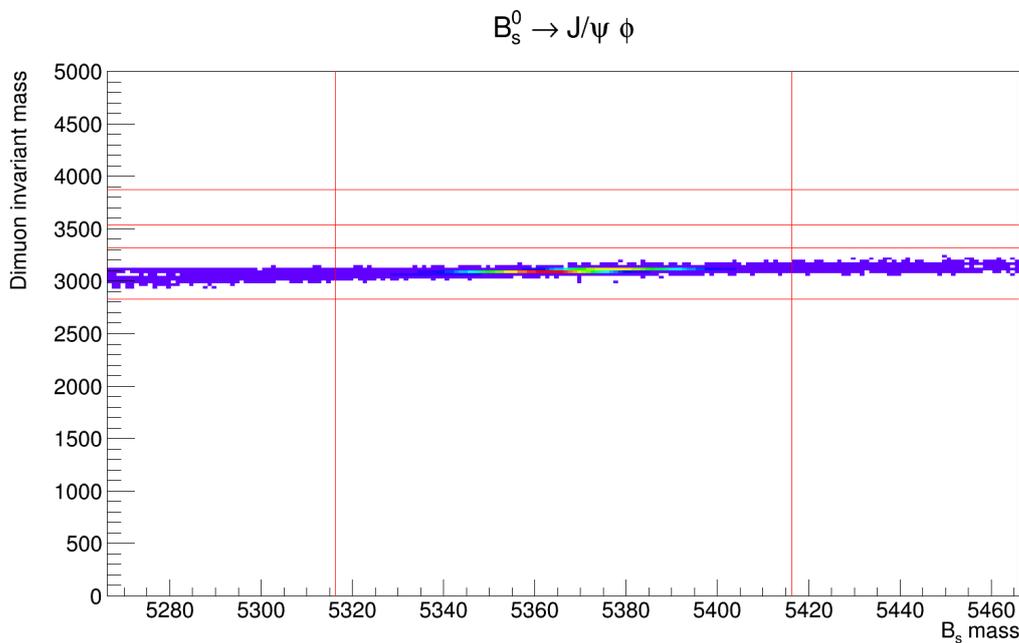


Figure 7.1: Distribution of  $B_s \rightarrow J/\psi(\rightarrow \mu^+\mu^-)\phi(\rightarrow K^+K^-)$  Monte Carlo events after full selection.

## 7.1 $B_s \rightarrow J/\psi\phi$

The largest background source present in  $B_s \rightarrow \phi\mu^+\mu^-$  is obviously the  $B_s \rightarrow J/\psi\phi$  reference decay. Cuts on the invariant dimuon mass are used to suppress this background. To evaluate how efficient those cuts are the  $B_s \rightarrow J/\psi\phi$  Monte Carlo sample is used. The  $B_s \rightarrow J/\psi\phi$  simulated events are selected in the same way

as the real signal events and are also corrected as described in Chapter 6. Figure 7.1 shows two dimensional distribution of the invariant dimuon mass ( $m_{\mu\mu}$ ) versus the invariant  $B_s$  mass ( $m_{B_s}$ ) for  $B_s \rightarrow J/\psi(\rightarrow \mu^+\mu^-)\phi(\rightarrow K^+K^-)$  Monte Carlo generated events. The charmonium mass window is large enough to include all  $B_s \rightarrow J/\psi\phi$  events.

It is also important to estimate the efficiency of the selection. The efficiency of selection is defined as  $\epsilon(x) = \frac{\text{selected events}}{\text{generated events}}$ . 376594  $B_s \rightarrow J/\psi\phi$  events remain after selection in the signal region  $5316.3 < m_{B_s} < 5413.6$ . 29897085 events have been generated when producing the Monte Carlo sample. The efficiency is then  $\epsilon(J/\psi\phi) = 0.012596 \pm 0.0000021$ .

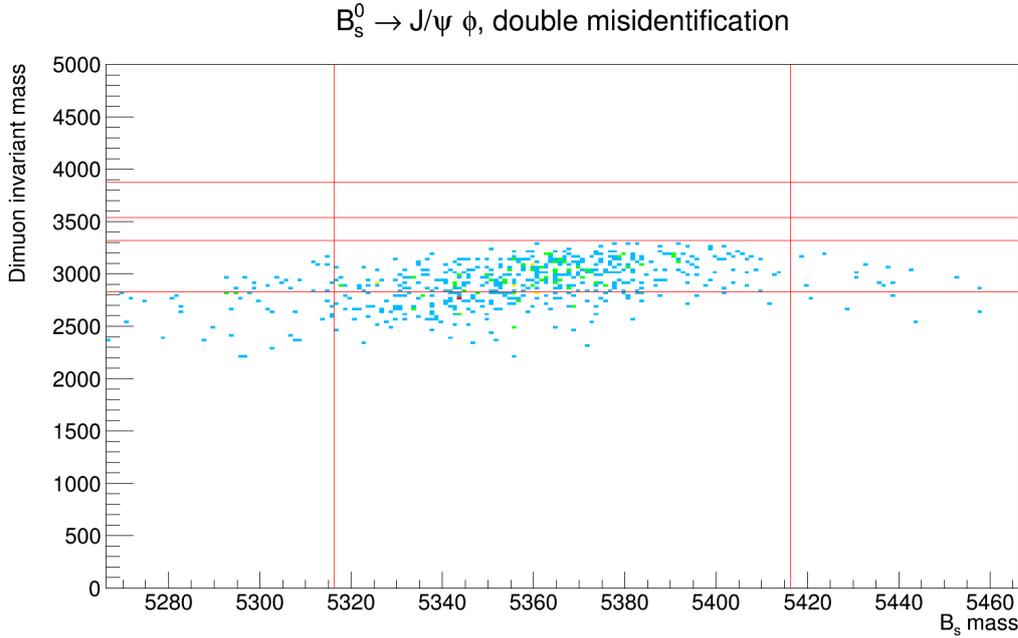


Figure 7.2: Distribution of  $B_s \rightarrow J/\psi\phi$  Monte Carlo events after full selection where a  $\mu$  is misidentified as  $K$ .  $K \mu$  vetoes were not applied in this plot.

## 7.2 $B_s \rightarrow J/\psi\phi$ with final state particles misidentified.

With the charmonium mass window chosen wide enough all  $B_s \rightarrow J/\psi\phi$  events are excluded from the range in which the branching ratio is determined. However if a  $K$  is swapped with a  $\mu$  of the same charge the reconstructed invariant dimuon mass of the event can end up outside the charmonium mass window. Therefore such misidentifications are a source of a peaking background for  $B_s \rightarrow \phi\mu^+\mu^-$ . In

this case two particles have been misidentified, a true  $K^\pm$  as  $\mu^\pm$  and a true  $\mu^\pm$  as a  $K^\pm$ . Figure 7.2 shows two dimensional distribution of the  $m_{\mu\mu}$  versus  $m_{B_s}$  for  $B_s \rightarrow J/\psi\phi$  Monte Carlo generated events where a  $\mu$  was swapped with a K of the same charge during the reconstruction.

To suppress this source of background two vetoes are applied on the invariant mass of  $K\mu$  system ( $K\mu$  vetoes). Assuming K mass for a  $\mu$ , the event is rejected if the invariant mass of  $K\mu \rightarrow K$  system lies within 8 MeV of the nominal  $\phi$  mass. Effectively one is undoing the swap and checks if the event is then consistent with the a  $J/\psi\phi$  decay. The second veto assumes  $\mu$  mass for K and rejects events if the invariant mass of  $\mu K \rightarrow \mu$  system lies within 50 MeV of nominal  $J/\psi\phi$  mass. After the veto is applied no  $B_s \rightarrow J/\psi\phi$  events where a  $\mu$  was swapped with a K pass the selection. Therefore this background is neglected in the analysis.

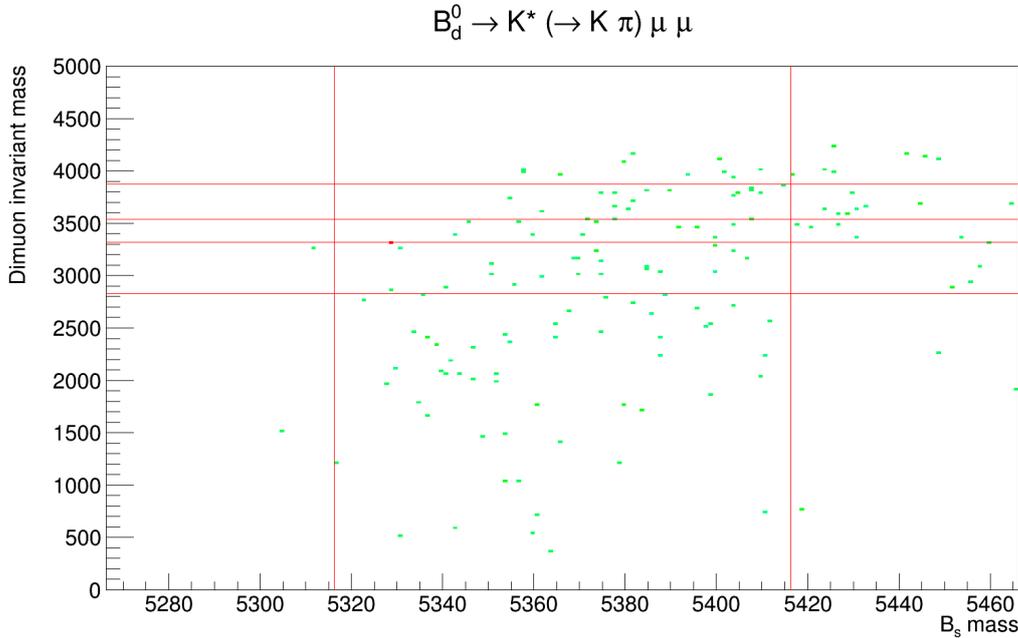


Figure 7.3: Distribution of  $B_d^0 \rightarrow K^*\mu\mu$  Monte Carlo events after full selection without the PID cut applied.

### 7.3 $B_d^0 \rightarrow K^*\mu\mu$

If the pion in the final state of  $B_d^0 \rightarrow K^*(\rightarrow K\pi)\mu^+\mu^-$  decay is misidentified as a K the final state mimics a  $B_s \rightarrow \phi(\rightarrow K^+K^-)\mu^+\mu^-$  decay. Since  $K^*(892)$  is a broad resonance a fraction of events will pass the  $\phi$  mass cut after the  $\pi$  is misidentified as

K. To suppress this background a cut on the  $PIDK(K)$  is introduced. To estimate the efficiency of the PID cut as well as the effect of the remaining background, Monte Carlo simulated  $B_d^0 \rightarrow K^* \mu^+ \mu^-$  events are used. The same selection procedure as described in Chapter 5 is applied and the simulated events are corrected as described in Chapter 6. Using Monte Carlo simulated  $B_d^0 \rightarrow K^* \mu^+ \mu^-$  events it is found that requiring  $PIDK(K^\pm) > -3$  reduces the background by a factor of 6 in the signal region. Figure 7.3 shows the distribution of  $B_d^0 \rightarrow K^* \mu^+ \mu^-$  Monte Carlo events before the PID cut is applied and in Figure 7.4 the PID cut has been applied.

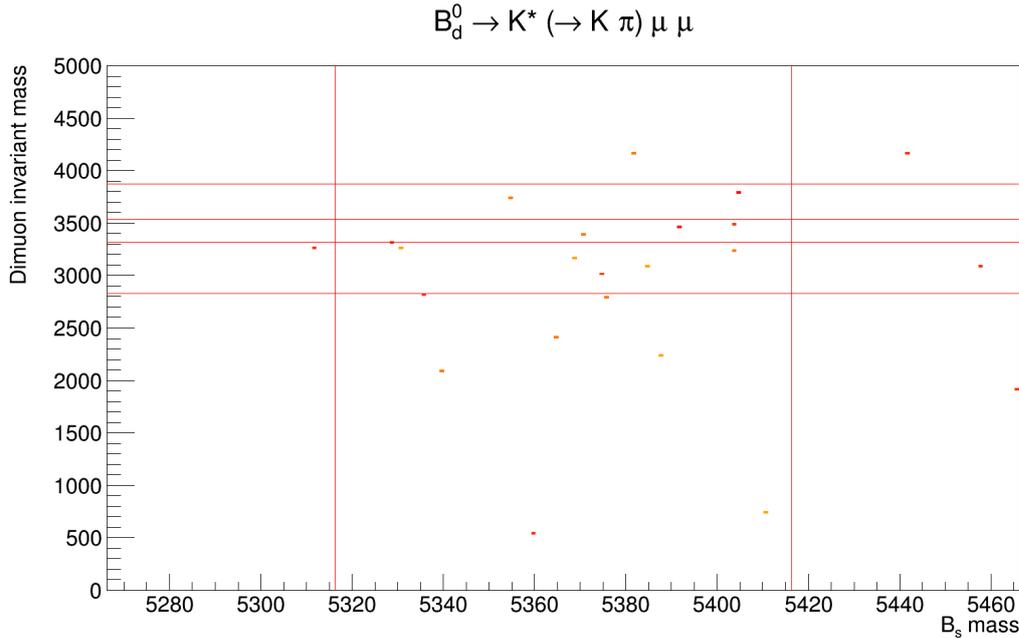


Figure 7.4: Distribution of  $B_d^0 \rightarrow K^* \mu \mu$  Monte Carlo events after full selection with the PID cut applied.

The  $PIDK(K) > -3$  is effective in the sense that it suppresses the peaking background. However it does not suppress the background completely and it is important to estimate how many background events are expected in the real data. This number is calculated by the following formula:

$$N_x = N_{J/\psi\phi} \cdot \frac{\epsilon(x) \cdot BR(x) \cdot f_x}{\epsilon(J/\psi\phi) \cdot BR(J/\psi\phi) \cdot f_s}$$

where  $BR(x)$  is the branching fraction of  $x$ ,  $\frac{f_x}{f_s}$  is the ratio of production of  $x$  primary particle, in this case  $B_d$ , and  $B_s$  and  $N_{J/\psi\phi} = 60844 \pm 245.2$  is the number of  $B_s \rightarrow J/\psi\phi$  events obtained by a fit (for more information about the fit see Chapter

8).

The relevant numbers are:

- 12.847 Monte Carlo simulated  $B_d \rightarrow K^* \mu^+ \mu^-$  events are observed in the signal mass window.
- 2518808 Monte Carlo events  $B_d \rightarrow K^* \mu^+ \mu^-$  have been generated.
- $\epsilon(B_d \rightarrow K^* \mu^+ \mu^-) = \frac{12.847}{2518808} = (5.1 \pm 1.42) \cdot 10^{-6}$
- $BR(B_d \rightarrow K^* \mu \mu) = (1.06 \pm 0.10) \cdot 10^{-6}$  [12]
- $BR(K^* \rightarrow \pi K) = 1$  [12]
- $BR(J/\psi \rightarrow \mu \mu) = 0.05961 \pm 0.00033$  [12]
- $BR(B_s \rightarrow J/\psi \phi) = (1.07 \pm 0.09) \cdot 10^{-3}$  [12]
- $BR(\phi \rightarrow K^+ K^-) = 0.489 \pm 0.005$  [12]
- $\frac{f_d}{f_s} = 3.79 \pm 0.29$  [16]

The result is  $N_{K^* \mu \mu} = 3.17 \pm 0.86$ . So 3.2 background events are expected among the signal  $B_s \rightarrow \phi \mu^+ \mu^-$  events. With a rough approximation of 400 signal events expected this is  $\approx 1.5\%$  of the signal. However since the signal is measured in bins of invariant dimuon mass the concentration of  $B_d \rightarrow K^* \mu^+ \mu^-$  background events might be higher for some of the bins. This is further investigated in Chapter 9.

## 7.4 $B_d^0 \rightarrow K^* J/\psi$

In analogy to  $B_s \rightarrow J/\psi \phi$  the resonant decay  $B_d^0 \rightarrow K^*(\rightarrow \pi K) J/\psi(\rightarrow \mu^+ \mu^-)$  can mimic the  $B_s \rightarrow \phi \mu^+ \mu^-$  decay if the  $\pi$  is misidentified as the  $\mu$  of the same charge and the  $\mu$  is misidentified as a K during the reconstruction. After the full selection and correction of the corresponding Monte Carlo sample 182 events remain, also shown in Figure 7.5. All events are contained in the charmonium mass window, so this background is neglected for the selection of  $B_s \rightarrow \phi \mu^+ \mu^-$ .

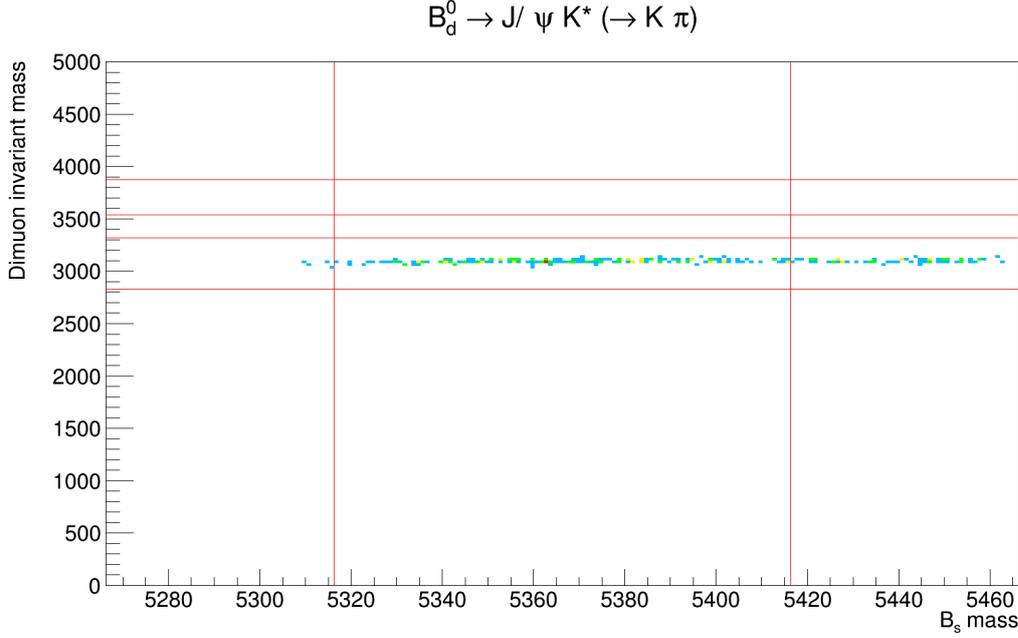


Figure 7.5: Distribution of  $B_d^0 \rightarrow J/\psi K^*$  Monte Carlo events after full selection.

## 7.5 $\Lambda_b \rightarrow \Lambda(1520)\mu^+\mu^-$

$\Lambda_b \rightarrow \lambda(1520)(\rightarrow pK)\mu^+\mu^-$  can mimic the final state of  $B_s \rightarrow \phi\mu^+\mu^-$  if the proton is misidentified as a K. As for all other potential backgrounds the Monte Carlo simulated events are selected and corrected. Figure 7.6 shows the distribution of these events. For such events under assumption of a K having p mass of 938.3 MeV the invariant mass  $m_{(p\rightarrow K)K\mu^+\mu^-}$  should be consistent with the  $\Lambda_b$  mass peak [5575, 5665]MeV. However since the  $B_s$  mass peak is rather wide many  $B_s \rightarrow \phi\mu^+\mu^-$  also satisfy this condition. The contribution from  $\Lambda_b \rightarrow \Lambda(1520)\mu^+\mu^-$  is removed by vetoing events with  $m_{(p\rightarrow K)K\mu^+\mu^-}$  around the nominal  $\Lambda_b$  mass. In addition it is required that the proton candidate has a proton-like  $DLL_{pK} > 10$ . This veto was adjusted to reduce the expected background contribution to below 2% of expected signal yield. Figure 7.7 shows the  $\Lambda_b \rightarrow \lambda(1520)(\rightarrow pK)\mu^+\mu^-$  Monte Carlo events after the veto is applied.

There are no measurements of the branching ratio of the  $\Lambda_b \rightarrow \Lambda(1520)\mu^+\mu^-$  so as a conservative estimate the branching ratio of  $\Lambda_b \rightarrow \Lambda\mu\mu$  is used. The relevant numbers are:

- 70.41 Monte Carlo simulated  $B_d \rightarrow K^*\mu^+\mu^-$  events are observed in the signal mass window.

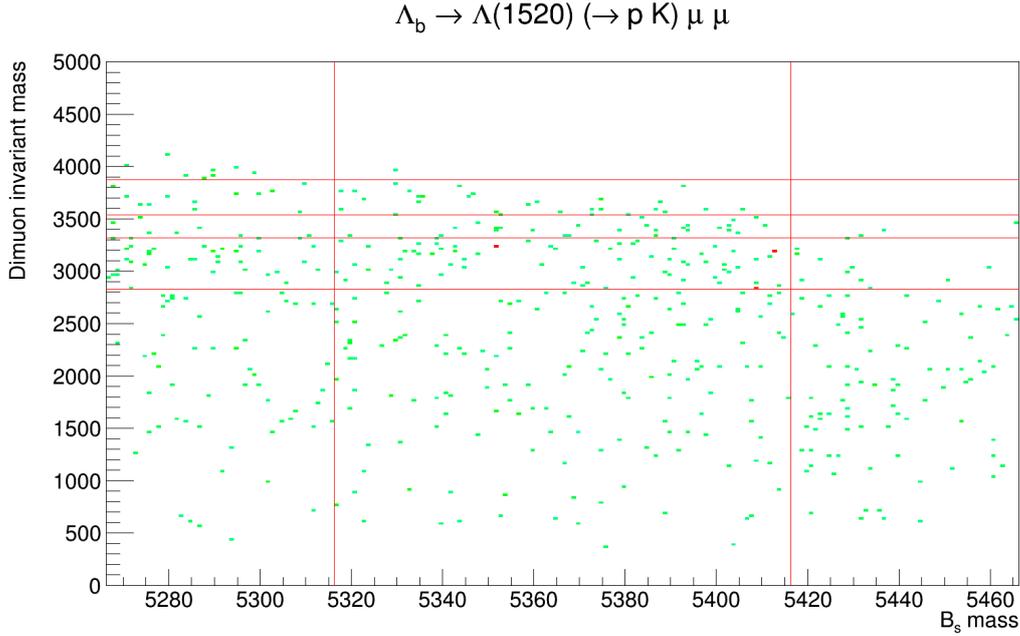


Figure 7.6: Distribution of  $\Lambda_b \rightarrow \Lambda(1520)(\rightarrow pK)\mu\mu$  Monte Carlo events after full selection without the  $K \rightarrow p$  veto applied.

- The Monte Carlo sample consists of 1511006 events.
- $\epsilon(\Lambda_b \rightarrow \Lambda\mu\mu) = \frac{70.41}{1511006} = (0.47 \pm 0.05) \cdot 10^{-4}$
- $BR(\Lambda_b \rightarrow \Lambda\mu\mu) = (0.96 \pm 0.31) \cdot 10^{-6}$  [6]
- $BR(\Lambda(1520) \rightarrow Kp) = 0.225 \pm 0.00225$  [12]
- $\frac{f_{\Lambda_b}}{f_s} = 2.31 \pm 0.24$  [16]

Using these numbers one finds  $N_{\Lambda(1520)\mu\mu} = 3.63 \pm 1.68$  events expected outside of the charmonium mass window. The effect this background has on the  $q^2$  bins is discussed in Chapter 9.

## 7.6 $\Lambda_b \rightarrow pK\mu^+\mu^-$

The products of the non-resonant decay of  $\Lambda_b \rightarrow pK\mu^+\mu^-$  can also mimic the final state of  $B_s \rightarrow \phi\mu^+\mu^-$  if  $p$  is misidentified as  $K$ . However in comparison with the resonant decay described in the previous section the invariant mass of  $pK$  is very wide and therefore heavily constrained by the cut on the  $\phi$  mass. Using the  $\Lambda_b \rightarrow \Lambda\mu^+\mu^-$  branching ratio as the conservative estimate again one finds that  $N = 0.3$

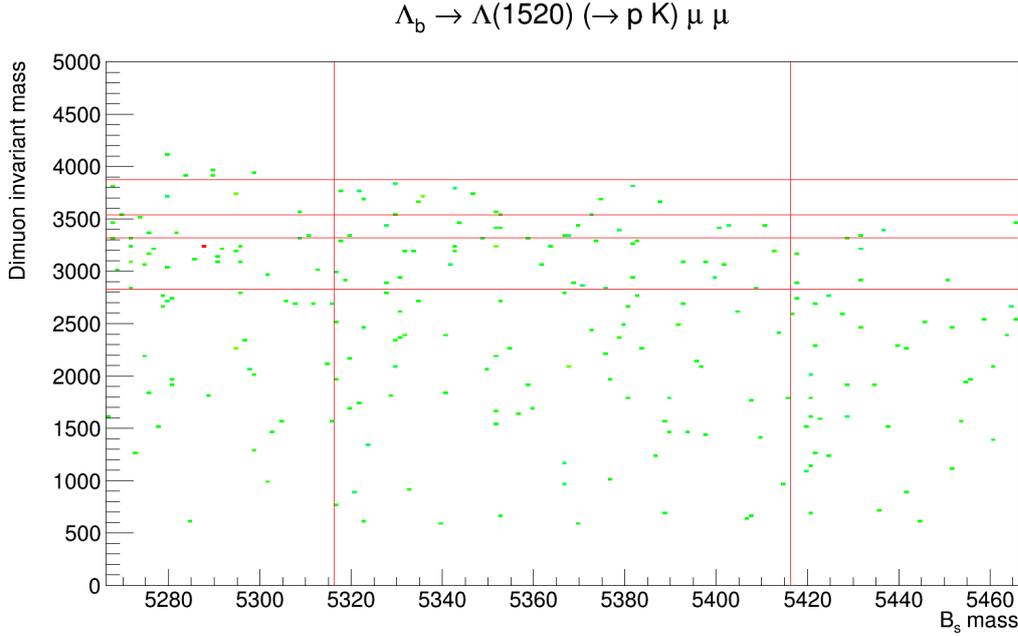


Figure 7.7: Distribution of  $\Lambda_b \rightarrow \lambda(1520)(\rightarrow pK)\mu\mu$  Monte Carlo events after full selection with the  $K \rightarrow p$  veto applied.

events are expected outside of the charmonium mass window. This background is ignored in the further analysis.

## 7.7 $B_s^0 \rightarrow D_s \pi$

The decay  $B_s^0 \rightarrow D_s \rightarrow (\phi \rightarrow (K^+ K^-) \pi) \pi$  can mimic the final state of  $B_s \rightarrow \phi \mu^+ \mu^-$  if both  $\pi$  are misidentified as muons. After full selection no events of the Monte Carlo remain in the signal region. So this background is ignored.

## 7.8 Semileptonic $b \rightarrow c \mu^- \bar{\nu}_\mu, c \rightarrow s \mu^+ \nu_\mu$

Cascade decays like  $B_s \rightarrow D_s^- \rightarrow (\phi \mu^- \bar{\nu}_\mu) \mu^+ \nu_\mu$  and  $B_d \rightarrow D^- \rightarrow (K^* \mu^- \bar{\nu}_\mu) \mu^+ \nu_\mu$  can mimic the final state of the  $B_s \rightarrow \phi \mu \mu$ . These decays are much more abundant than the signal decays. The branching fractions are  $2 \cdot 10^{-3}$  and  $8 \cdot 10^{-4}$  correspondingly. However due to the two neutrinos in the final state the reconstructed mass of the final state of these semileptonic decays peaks below the nominal  $B_s$  mass. For each decay channel  $2 \cdot 10^6$  generator level events have been generated. A conservative momentum resolution of 1% is assumed. Figures 7.8 and 7.9 show the resulting

$$B_d \rightarrow D \mu \nu$$

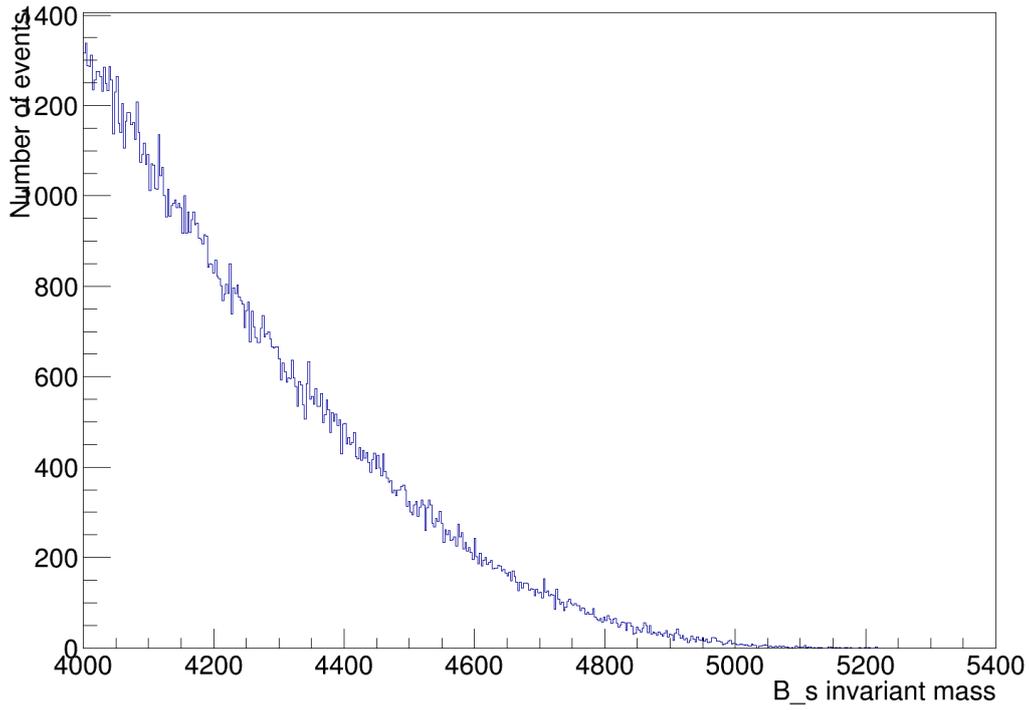


Figure 7.8: Distribution of 2 million  $B_s \rightarrow D_s^- \rightarrow (\phi\mu^-\bar{\nu}_\mu)\mu^+\nu_\mu$  generator level events. No events found in the signal region  $5316.3 \text{ MeV} < m_{B_s} < 5416.3 \text{ MeV}$ .

distributions of reconstructed  $B_s$  mass. No events are found in the  $B_s$  signal region so these backgrounds are ignored in the analysis.

$B_s \rightarrow D_s \mu \nu$

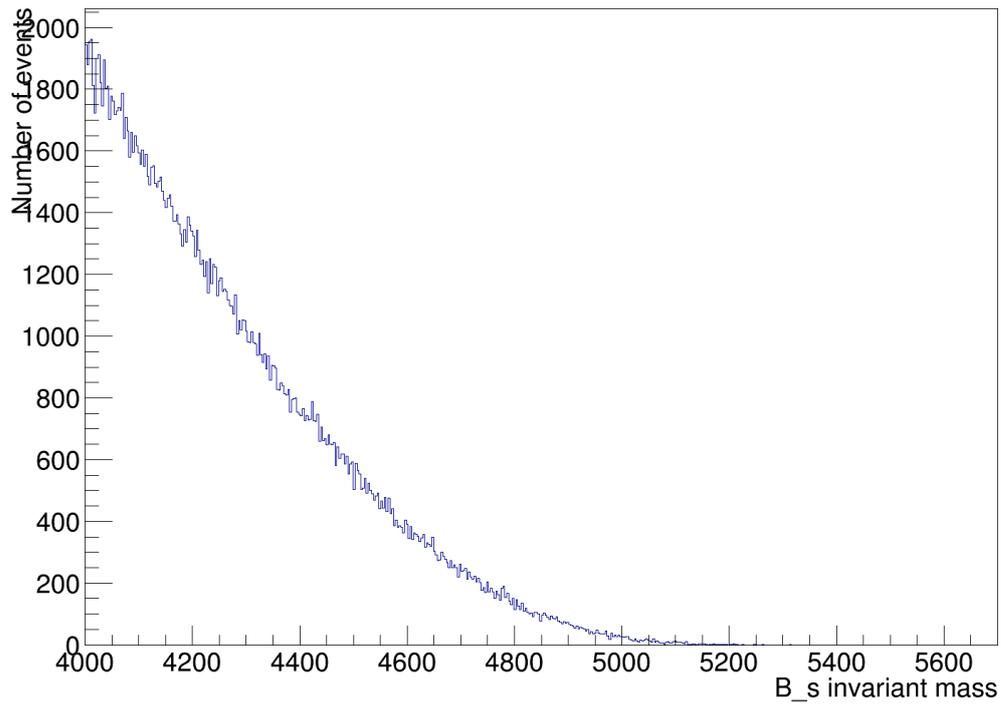


Figure 7.9: Distribution of 2 million  $B_s \rightarrow D_s^- \rightarrow (\phi \mu^- \bar{\nu}_\mu) \mu^+ \nu_\mu$  generator level events. No events found in the signal region  $5316.3 \text{ MeV} < m_{B_s} < 5416.3 \text{ MeV}$ .

## 8 Differential Branching Fractions

The aim of the analysis presented in this thesis is to measure the differential branching fraction of the decay channel  $B_s \rightarrow \phi\mu^+\mu^-$  as function of the invariant dimuon mass. To determine the branching fraction three numbers are required. First the signal yield has to be determined from the selected dataset. The yield of the reference channel  $B_s \rightarrow J/\psi\phi$  is also required to determine the relative branching ratio  $\frac{Br(B_s \rightarrow \phi\mu^+\mu^-)}{Br(B_s \rightarrow J/\psi\phi)}$ . To be able compare the signal and the reference channel the relative signal/reference efficiency has to be calculated as well. This chapter describes the method used to determine the signal and the reference yields and the calculation of the relative efficiency.

### 8.1 General Strategy

To determine the reference channel yield an extended maximum likelihood fit to the invariant  $B_s$  mass distribution is performed in the range [5166.3, 5566.3]MeV. This corresponds to a window of  $\pm 200$  MeV around the world average of the  $B_s$  mass measurements.

The shape of the invariant  $B_s$  mass observed is due to detector resolution and can be reasonably described by a combination of two Gaussians with the same mean. There are however several smaller contributions (such as radiative decays) that change the signal shape, especially the tails. A useful function to describe such shapes is a Crystal Ball function [26]. It consists of a Gaussian core with a power-law tail, so that various tails can be described. For the signal fit the shape parameters are fixed by the fit to the reference channel and only the absolute yield is left free.

**Crystal Ball function** The Crystal Ball function is a probability density function that is often used in high-energy physics to account for energy loss due to photon

radiation. The function is given by

$$f(x) = N \cdot \begin{cases} \exp(-\frac{(x-\bar{x})^2}{2\theta^2}), & \text{for } \frac{x-\bar{x}}{\theta} > -\alpha \\ A \cdot (B - \frac{x-\bar{x}}{\theta})^{-n}, & \text{for } \frac{x-\bar{x}}{\theta} \leq -\alpha \end{cases} \quad (8.1)$$

where  $N$  is a normalization factor,  $A = (\frac{n}{|\alpha|})^n \cdot \exp(-\frac{|\alpha|^2}{2})$ , and  $B = \frac{n}{|\alpha|} - |\alpha|$ . In a Crystal Ball function parameters  $\bar{x}$  and  $\theta$  play a role similar to the mean and width of a Gaussian. The  $\alpha$  parameter determines at which point the Gaussian transitions into the power-law tail. The  $n$  parameter determines the shape of the power-law tail.

## 8.2 Extended maximum-likelihood fit

Let  $P(X|\alpha)$  be the probability of measuring a value  $X$  for a given variable on an event.  $\alpha$  is a set of parameters that  $P$  depends on and  $X$  is some measured variable. For  $N$  independent measurements the likelihood function  $L$  is defined as

$$L(\alpha) = \prod_{j=1}^N P(X_j|\alpha)$$

It is important to note that  $L$  is a function of  $\alpha$ , not of  $X_j$ . The value of  $\alpha$  that maximizes  $L$  is called Maximum Likelihood Estimator (MLE). It is the set of parameters for which  $P$  describes  $\{X_j\}$  the best. A convenient method of finding  $\alpha$  is working with  $\ln(L)$  instead of just  $L$ . The product of probabilities becomes a sum and working with derivatives becomes much easier.

Suppose that the invariant  $B_s$  mass distribution is described by some normalized function  $S(\alpha)$  and the background by some normalized function  $B(\alpha)$ . Then the total distribution  $p$  is described by:

$$p = f_s S(\alpha) + (1 - f_s) B(\alpha)$$

where  $f_s$  is the fraction of signal events. Signal yield  $N_s$  is given by  $f_s N$  where  $N$  is the total number of events. However the statistical uncertainty  $\Delta N_s$  is not described by  $\Delta f_s N$ . One would like to describe the distribution  $P$  using  $N_s$  instead of  $f_s$  to easily access  $\Delta N_s$ . Then one has to work with some unnormalized distribution  $P$  instead of the normalized  $p$ . The number of observed events  $N$  is a Poisson random

variable with the mean value  $(N_s + N_b)$  where  $N_s$  and  $N_b$  are the expected number of signal and background events. The likelihood function has to be modified by the Poisson probability  $e^{-(N_s+N_b)} \frac{(N_s+N_b)^N}{N!}$  to account for that.

$$\begin{aligned} L(\alpha) &= e^{-(N_s+N_b)} \frac{(N_s + N_b)^N}{N!} \cdot \prod_{j=1}^N \left( \frac{N_s}{N_s + N_b} S(\alpha) + \frac{N_b}{N_s + N_b} B(\alpha) \right) = \\ &= \frac{e^{-(N_s+N_b)}}{N!} \cdot \prod_{j=1}^N (N_s S(\alpha) + N_b B(\alpha)). \end{aligned} \quad (8.2)$$

This is called extended likelihood function. After a maximum -likelihood fit is applied to this function one gets a value for  $N_s$  and can easily calculate the  $\Delta N_s$ . This is called extended maximum-likelihood fit, It is usually the most powerful fitting method (compared to a binned likelihood and  $\chi^2$  fit), but there is no goodness-of-fit estimation. In the analysis the estimation of the errors of  $N_s$  is important so extended maximum likelihood is used.

### 8.3 $B_s \rightarrow J/\psi\phi$ fit

The initial assumption is that the invariant  $B_s$  mass is well described by a combination of two Gaussians with exponential background:

$$\text{PDF} = f^{sig}(f^{struc} \cdot G_1(x) + (1 - f^{1struc})G_2(x)) + (1 - f^{sig}) \cdot exp(x)$$

where  $G_1$  and  $G_2$  are Gaussians with same mean, but different widths,  $exp(x)$  is an exponential function,  $f^{struc}$  is the ratio of events described by the Gaussian  $G_1$  to the number of events described by both Gaussians and  $f^{sig}$  is the ratio of events described by the two Gaussians to the total number of events. Both  $f^{struc}$  and  $f^{sig}$  have values between 0 and 1. For the reference channel the fit is performed in the mass window corresponding to dimuon mass [3021.9, 3171.9]MeV, the nominal  $J/\psi$  mass  $\pm 75$ MeV. Figure 8.1 shows the results of the fit. The shape parameters determined in the fit are given in Table 8.1. The double Gaussian model seems to struggle with the description of the tails. This effect is likely due to photon radiation.

The tails are better described by Crystal Ball based models so the PDF is changed to:

$$\text{PDF} = f^{sig}(f^{struc} \cdot CBS_1(x) + (1 - f^{struc})CBS_2(x)) + (1 - f^{sig}) \cdot exp(x)$$

parameter	result	$\sigma_{\text{down}}$	$\sigma_{\text{up}}$
$N_{\text{sig}}$	$62961.4 \pm 254.1$	-264.27	+ 263.6
$N_{\text{bkg}}$	$4111.0 \pm 81.3$	-102.95	+ 105.0
$m_{B^0}$	$5370.759 \pm 0.078$	-0.08	+ 0.078
$f_{m,1}^{\text{sig}}$	$0.8073 \pm 0.0036$	-0.01	+ 0.0096
$\sigma_{m,1}$	$15.875 \pm 0.068$	-0.13	+ 0.13
$\sigma_{m,2}$	$37.28 \pm 0.40$	-0.94	+ 0.97
$\alpha_m$	$0.00410 \pm 0.00016$	-0.000172	+ 0.00017

Table 8.1: Results of the double Gaussian fit to the  $B_s$  invariant mass of the reference channel.

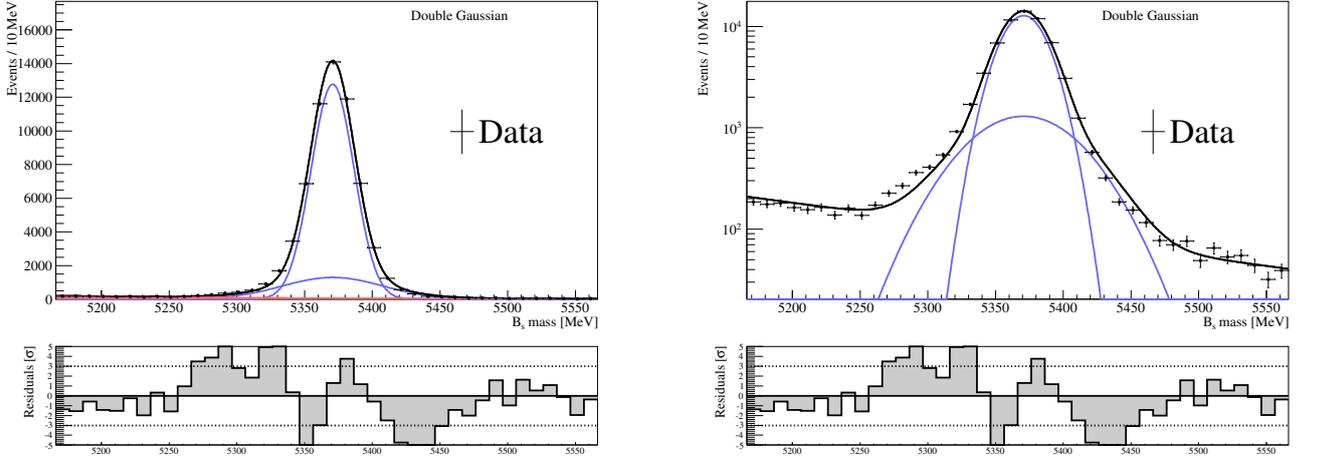


Figure 8.1: Fit of the invariant mass spectrum of the reference channel  $B_s \rightarrow J/\psi\phi$  using double Gaussian to model the signal and an exponential to model the background. Linear scale to the left, log-scale to the right. Results of the fit are given in Table 8.1.

where  $CBS_1$  and  $CBS_2$  are Crystal Ball functions with common means,  $n$  and  $\alpha$ , but with different widths. The resulting fit for the reference channel is shown in Figure 8.2 and the extracted fit parameters are given in the Table 8.2. This model performs better than the double Gaussian and is used for all further fits.

$N_{\text{sig}}$  gives the number of reference channel  $B_s \rightarrow J/\psi\phi$  events in the whole  $B_s$  mass range of  $[5166.3, 5566.3]$  MeV. For the analysis however, the yield in the mass range  $[5316.3, 5416.3]$  MeV is interesting since the  $B_s \rightarrow \phi\mu^+\mu^-$  signal events are also measured in this mass window. To estimate this yield the double Crystal Ball function with the parameters used to perform the fit is integrated twice, once over the whole range and once only over the signal region. The fraction of these two

parameter	result	$\sigma_{\text{down}}$	$\sigma_{\text{up}}$
$N_{\text{sig}}$	$63192.6 \pm 254.7$	-266.67	+ 269.3
$N_{\text{bkg}}$	$3879.8 \pm 81.5$	-112.35	+ 111.0
$m_{B^0}$	$5371.347 \pm 0.077$	-0.095	+ 0.097
$f^{\text{sig}}$	$0.8496 \pm 0.0040$	-0.017	+ 0.015
$\sigma_{\text{m},1}$	$15.849 \pm 0.069$	-0.171	+ 0.16
$\sigma_{\text{m},2}$	$34.76 \pm 0.51$	-1.57	+ 1.7
$\alpha_{\text{CB}}$	$1.498 \pm 0.018$	-0.069	+ 0.070
$n_{\text{CB}}$	$8.82 \pm 0.88$	-2.47	+ 5.3
$\alpha_{\text{m}}$	$0.00347 \pm 0.00016$	-0.00014	+ 0.00020

Table 8.2: Results of the double Crystal Ball fit to the  $B_s$  invariant mass of the reference channel.

integrals multiplied by  $N_{\text{sig}}$  gives the signal yield of  $N_{J/\psi\phi} = 60844.0 \pm 245.2$ .

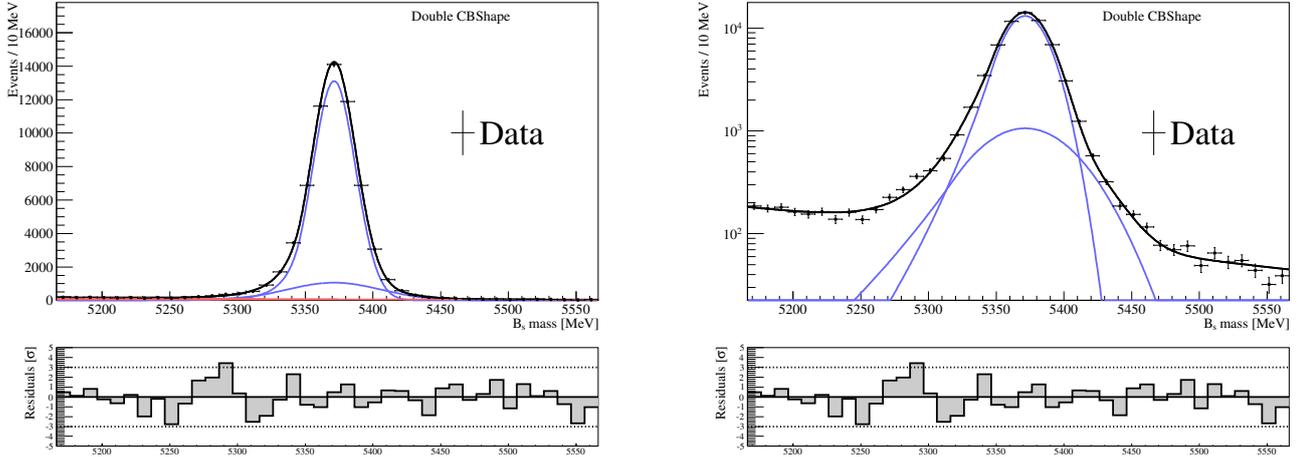


Figure 8.2: Fit of the invariant mass spectrum of the reference channel  $B_s \rightarrow J/\psi\phi$  using using double Crystal Ball function to model the signal and an exponential to model the background. Normal scale to the left, log scale to the right. Results of the fit are given in Table 8.2.

## 8.4 Scaling Factors

The yield of  $B_s \rightarrow \phi\mu^+\mu^-$  signal events is determined in bins of  $q^2$ . However the mass resolution of the LHCb is  $q^2$  dependent. This means that the mass resolution for the reconstructed  $B_s \rightarrow \phi\mu^+\mu^-$  also varies with  $q^2$ . The double Crystal Ball model determined in the previous section for a single  $q^2$  bin cannot be used directly

for all  $q^2$  bins. A  $q^2$ -dependent scaling factor is introduced to scale the width of the Crystal Ball functions in accordance to the bin the fit is performed in. All the other fit parameters remain constrained to the values from the fit of the reference channel. The scaling factors are determined from the  $B_s \rightarrow \phi\mu^+\mu^-$  Monte Carlo simulated events. First a single Crystal Ball function is fit to the  $B_s \rightarrow J/\psi\phi$  Monte Carlo sample to determine the reference width  $\sigma_{J/\psi}$  and the best fit parameters. The fit itself is shown in Figure 8.3 and the parameters are given in Table 8.3. For each  $q^2$  bin a single Crystal Ball is fit to the  $B_s$  mass with all parameters, but width  $\sigma_i$  constrained. The scale factor is defined as the ratio  $\frac{\sigma_i}{\sigma_{J/\psi}}$ . The scaling factors of different bins as well as the widths  $\sigma_i$  are given in Table 8.4.

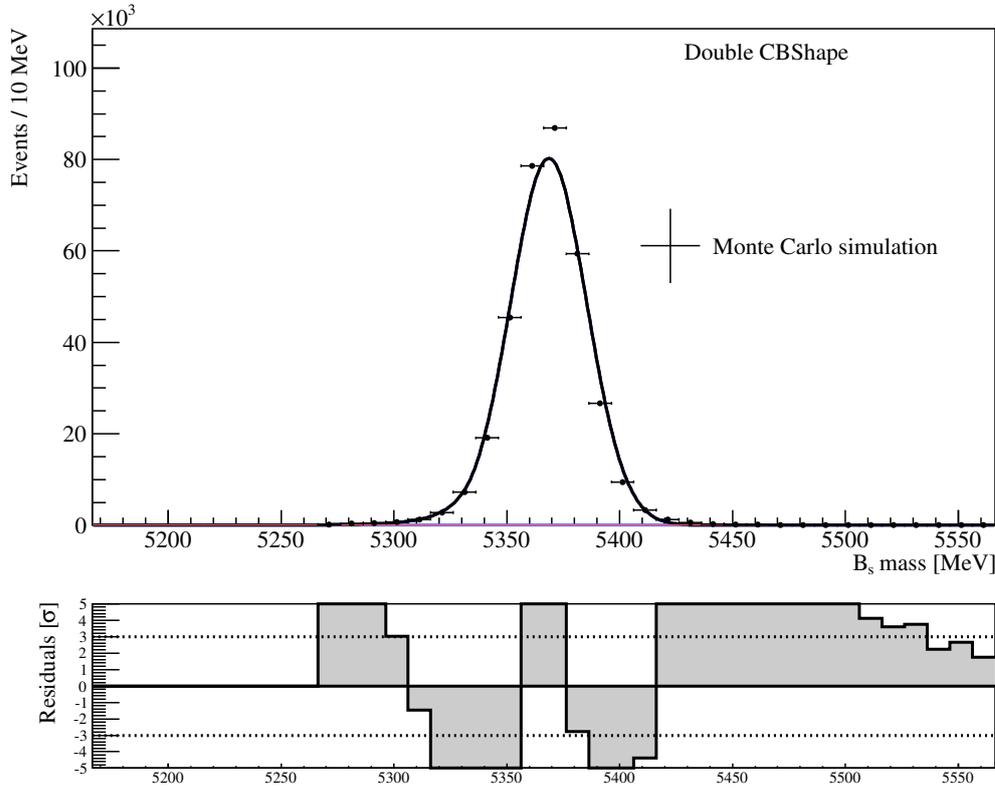


Figure 8.3: A single Crystal Ball fit to the invariant mass spectrum of the  $B_s \rightarrow J/\psi\phi$  Monte Carlo events.

parameter	result	$\sigma_{\text{down}}$	$\sigma_{\text{up}}$
$N_{\text{sig}}$	$344400.3 \pm 586.9$	-586.49	+ 587.2
$m_{B^0}$	$5368.559 \pm 0.031$	-0.031	+ 0.031
$\sigma_{\text{m},1}$	$16.907 \pm 0.024$	-0.024	+ 0.024
$\alpha_{\text{CB}}$	$1.8417 \pm 0.0084$	-0.0083	+ 0.0084
$n_{\text{CB}}$	$7.940 \pm 0.028$	- 0.028	+ 0.028

Table 8.3: Results of the single Crystal Ball function fit to the mass spectrum of the  $B_s \rightarrow J/\psi\phi$  Monte Carlo events.

Bin	bounds	width of MC	Scaling factor
Bin1	$0.1\text{GeV} < q^2 < 2\text{GeV}$	$17.28 \pm 0.10$	$1.02 \pm 0.01$
Bin 2	$2\text{GeV} < q^2 < 5\text{GeV}$	$16.75 \pm 0.10$	$0.99 \pm 0.01$
Bin 3	$5\text{GeV} < q^2 < 8\text{GeV}$	$16.779 \pm 0.08$	$0.99 \pm 0.01$
Bin 4	$11\text{GeV} < q^2 < 12.5\text{GeV}$	$17.02 \pm 0.10$	$1.01 \pm 0.01$
Bin 5	$15\text{GeV} < q^2 < 17\text{GeV}$	$17.86 \pm 0.10$	$1.06 \pm 0.01$
Bin 6	$17\text{GeV} < q^2 < 19\text{GeV}$	$18.65 \pm 0.16$	$1.10 \pm 0.01$
Bin 7	$1\text{GeV} < q^2 < 6\text{GeV}$	$16.77 \pm 0.07$	$0.99 \pm 0.01$
Bin 8	$15\text{GeV} < q^2 < 19\text{GeV}$	$18.08 \pm 0.09$	$1.07 \pm 0.01$
All bins	$\cup q^2$ bins	$17.25 \pm 0.04$	$1.02 \pm 0.01$

Table 8.4:  $q^2$ -dependent scaling factors for the widths of the Crystal Ball functions.

## 8.5 Signal Yields

The signal yield for the  $B_s \rightarrow \phi\mu^+\mu^-$  signal is determined in bins of  $q^2$  using an extended unbinned maximum likelihood fit. A double Crystal Ball function is used to describe the signal and an exponential is used to describe the background. The parameters used in the model are given in Table 8.2. All the parameters except  $N_s$  and  $N_b$  are constrained. The width  $\sigma_1$  and  $\sigma_2$  of the Crystal Ball functions are modified by the scaling factors given in Table 8.4. The resulting fits are given in Figures A.3 - A.4. The yields are given in Table 8.5.

## 8.6 Selection Efficiencies

In the presented analysis the branching ratio of the  $B_s \rightarrow \phi\mu^+\mu^-$  is determined relative to the branching ratio of the normalization decay  $B_s \rightarrow J/\psi\phi$ . To do that one has to know the relative efficiency of both selections. For both decays the efficiencies are determined using the corrected Monte Carlo events, see Chapter 6.

The efficiency is defined as  $\frac{\text{events selected}}{\text{events generated}}$  for the reference channel Monte Carlo

$q^2$ bin	$N_{sig}$
$0.1 < q^2 < 2$	$84.74^{+9.71}_{-10.57}$
$2 < q^2 < 5$	$58.94^{+9.45}_{-8.87}$
$5 < q^2 < 8$	$72.95^{+10.82}_{-10.30}$
$11 < q^2 < 12.5$	$69.14^{+10.33}_{-9.72}$
$15 < q^2 < 17$	$84.71^{+10.54}_{-10.01}$
$17 < q^2 < 19$	$49.11^{+7.88}_{-7.22}$
$1 < q^2 < 6$	$99.71^{+12.47}_{-11.83}$
$15 < q^2 < 19$	$134.49^{+13.05}_{-12.54}$
Reference $B_s \rightarrow J/\psi\phi$	$60844.0 \pm 245.2$

Table 8.5: The signal yields of the  $B_s \rightarrow \phi\mu\mu$  signal

simulated events. The efficiency for the reference channel  $J/\psi\phi$  was calculated in Chapterdatamc to be  $\epsilon(J/\psi\phi) = 1.26 \pm 0.02\%$ . The efficiency of the signal channel  $B_s \rightarrow \phi\mu\mu$  is calculated in bins of  $q^2$ . For that purpose 10 million generator level  $B_s \rightarrow \phi\mu\mu$  events have been generated.  $\epsilon(\phi\mu\mu)$  in a given bin  $i$  is determined as

$$\frac{\text{events selected}}{\text{events generated}} \cdot \frac{10^7}{\text{generator level events in bin } i}$$

Such a definition takes into account the possibly  $q^2$ -dependent preselection. As a crosscheck the separate efficiencies of the triggers  $\epsilon_{trig}$ , the cut based selection  $\epsilon_{del}$ , the BDT selection  $\epsilon_{bet}$  and the vetoes  $\epsilon_{veto}$  are calculated for all bins and for the reference channel. These efficiencies are given in Table 8.6. One sees that the efficiency of the triggers and of the vetoes are highly  $q^2$  dependent while the cut based selection and BDT selection have approximately the same efficiency over the whole  $q^2$  range.

As a crosscheck the efficiency of the trigger cuts is calculated from the  $B_s \rightarrow J/\psi\phi$  reference data sample. A way to extract this efficiency from real data is to compare the number of events that have been triggered independently of the  $B_s$  decay products with the number of events that were triggered both on the  $B_s$  decay products and independent of them. This ratio should also estimate the ratio of events events total to events triggered by the  $B_s$  decay products, i.e. 1/trigger efficiency. Calculating the trigger efficiency this way gives  $\epsilon_{TISTOS} = 85.0 \pm 5.85\%$ . From Table 8.6 the  $\epsilon_{trig}$  for the reference  $B_s \rightarrow J/\psi\phi$  Monte Carlo simulated events is  $\epsilon_{trig} = 81.8 \pm 0.2\%$  Within the statistical uncertainties the trigger efficiencies derived from Monte Carlo sample and real data are in agreement.

Bin	$\epsilon_{tot}$	$\epsilon_{trig}$	$\epsilon_{veto}$	$\epsilon_{bdt}$	$\epsilon_{sel}$
$0.1\text{GeV} < q^2 < 2\text{GeV}$	$0.009131 \pm 6.3 \cdot 10^{-5}$	$0.63 \pm 0.01$	$0.90 \pm 0.01$	$0.97 \pm 0.01$	$0.88 \pm 0.01$
$2\text{GeV} < q^2 < 5\text{GeV}$	$0.009481 \pm 6.6 \cdot 10^{-5}$	$0.67 \pm 0.01$	$0.93 \pm 0.01$	$0.97 \pm 0.01$	$0.88 \pm 0.01$
$5\text{GeV} < q^2 < 8\text{GeV}$	$0.010546 \pm 6.4 \cdot 10^{-5}$	$0.73 \pm 0.01$	$0.94 \pm 0.01$	$0.97 \pm 0.01$	$0.87 \pm 0.01$
$11\text{GeV} < q^2 < 12.5\text{GeV}$	$0.013184 \pm 9.2 \cdot 10^{-5}$	$0.84 \pm 0.01$	$0.99 \pm 0.01$	$0.97 \pm 0.01$	$0.88 \pm 0.01$
$15\text{GeV} < q^2 < 17\text{GeV}$	$0.012059 \pm 8.2 \cdot 10^{-5}$	$0.89 \pm 0.01$	$1 \pm 0.01$	$0.97 \pm 0.01$	$0.88 \pm 0.01$
$17\text{GeV} < q^2 < 19\text{GeV}$	$0.008602 \pm 9.5 \cdot 10^{-5}$	$0.9 \pm 0.02$	$1 \pm 0.02$	$0.97 \pm 0.026$	$0.88 \pm 0.01$
$1\text{GeV} < q^2 < 6\text{GeV}$	$0.009499 \pm 5.0 \cdot 10^{-5}$	$0.68 \pm 0.01$	$0.93 \pm 0.01$	$0.97 \pm 0.01$	$0.88 \pm 0.01$
$15\text{GeV} < q^2 < 19\text{GeV}$	$0.010861 \pm 6.3 \cdot 10^{-5}$	$0.89 \pm 0.01$	$1 \pm 0.01$	$0.97 \pm 0.01$	$0.88 \pm 0.01$
$J/\psi\phi$	$0.0126 \pm 0.0002$	$0.81 \pm 0.01$	$0.97 \pm 0.01$	$0.97 \pm 0.01$	$0.88 \pm 0.01$

Table 8.6: Table of efficiencies for  $B_s \rightarrow \phi\mu^+\mu^-$  Monte Carlo.

Bin	Branching ratio [ $10^{-4}$ ]
$0.1 < q^2 < 2$	$1.144^{+0.131}_{-0.143}$
$2 < q^2 < 5$	$0.767^{+0.123}_{-0.115}$
$5 < q^2 < 8$	$0.854^{+0.127}_{-0.121}$
$11 < q^2 < 12.5$	$0.647^{+0.097}_{-0.091}$
$15 < q^2 < 17$	$0.867^{+0.108}_{-0.102}$
$17 < q^2 < 19$	$0.705^{+0.113}_{-0.104}$
$1 < q^2 < 6$	$1.295^{+0.162}_{-0.154}$
$15 < q^2 < 19$	$1.5281^{+0.148}_{-0.143}$

Table 8.7: Relative branching ratios  $\frac{Br(B_s \rightarrow \phi\mu^+\mu^-)}{Br(B_s \rightarrow J/\psi\phi)}$ .

## 8.7 Ratio of Branching Fractions

The relative branching ratio of the signal  $B_s \rightarrow \phi\mu^+\mu^-$  to the reference channel  $B_s \rightarrow J/\psi\phi$  is determined using:

$$\frac{Br(B_s \rightarrow \phi\mu^+\mu^-)}{Br(B_s \rightarrow J/\psi\phi)} = \frac{N_{\phi\mu^+\mu^-}}{N_{J/\psi\phi}} \frac{Br(J/\psi\phi \rightarrow \mu^+\mu^-) \cdot \epsilon_{tot}(J/\psi\phi)}{\epsilon_{tot}(\phi\mu^+\mu^-)} \quad (8.3)$$

The branching ratio  $Br(J/\psi \rightarrow \mu^+\mu^-) = (5.961 \pm 0.033) \cdot 10^{-2}$ . The efficiencies  $\epsilon_{tot}$  have been calculated and are given in Table 8.6. The signal yields  $N_{\phi\mu^+\mu^-}$  and  $N_{J/\psi\phi}$  have been determined using maximum likelihood fit and are given in Table 8.5. The resulting relative branching ratios are calculated using Formula 8.3 and are given in Table 8.7.

## 9 Systematic Uncertainties

In this chapter a summary of all studied systematic uncertainties for the determination of the differential branching fraction ratio is presented. All the systematic uncertainties are gathered in Table 9.1.

### 9.1 Peaking Background

Products of various B decays, as discussed in Chapter 9.1, can mimic the final state of the signal  $B_s \rightarrow \phi\mu^+\mu^-$ . These peaking backgrounds are not included in the fit and can bias the signal yield. The expected number of background events for  $B_d \rightarrow K^*\mu^+\mu^-$  and  $\Lambda_b \rightarrow \Lambda(1520)\mu^+\mu^-$  are taken as the systematic uncertainty. The  $q^2$  dependence is extracted from the corresponding Monte Carlo sample by relaxing the PID cuts and the misID vetoes. The systematic uncertainty due to peaking backgrounds is given in Table 9.1.

### 9.2 Branching Fraction of the Decay $J/\psi \rightarrow \mu^+\mu^-$

The branching fraction of the decay  $J/\psi \rightarrow \mu^+\mu^-$  is  $Br(J/\psi \rightarrow \mu^+\mu^-) = 0.05961 \pm 0.00033$  [12]. This branching fraction is used in equation 8.3 to calculate the relative branching ratio of the signal to the reference channel. The uncertainty of the branching fraction causes a systematic uncertainty which is given in Table 9.1.

### 9.3 Relative Efficiencies Precision

The relative efficiency  $\frac{\epsilon(J/\psi\phi)}{\epsilon(\phi\mu\mu)}$  are determined from Monte Carlo simulated events. Due to limited statistics this relation can only be determined with a certain precision. The systematic uncertainty resulting from the limited statistics is given in Table 9.1.

## 9.4 $q^2$ binning

The determination of the relative branching fraction  $\frac{\epsilon(J/\psi\phi)}{\epsilon(\phi\mu\mu)}$  assumes that the  $B_s \rightarrow \phi\mu^+\mu^-$  efficiency is constant over the respective bin. To estimate the possible systematic effect of this averaging every  $q^2$  bin is divided in four sub-bins. Each event is assigned relative efficiency  $\epsilon(B_s \rightarrow J/\psi\phi)/\epsilon(B_s \rightarrow \phi\mu^+\mu^-)$  for the corresponding sub-bin and the fit is repeated using the weighted events. For the determination of the branching fraction ratio the fact that the fitted signal yield already contains the relative efficiency is taken into account. The resulting deviations are given in Table 9.1.

## 9.5 Tracking Efficiency

The tracking efficiency is only determined to a certain precision and can be a source of systematic deviations. To estimate the uncertainty the relative efficiency is recalculated using different weights. For tracks with momenta smaller than 10 GeV the efficiency is increased by the statistical uncertainty of the efficiency given in Figure 6.1 and by an additional systematic contribution of 1.5% for kaons and 0.4% for muons as recommended by the tracking group. For tracks with momenta larger than 10 GeV the efficiency is decreased by the the statistical uncertainty of the efficiency plus the additional systematic contribution. The difference of the tracking efficiencies is taken as systematic uncertainty. The systematic uncertainty due to varied tracking efficiency is given in Table 9.1.

## 9.6 Muon ID Efficiency

The corrections performed on the muon ID efficiency can lead to a systematic deviation. To estimate this systematic uncertainty the relative efficiency  $\frac{\epsilon(J/\psi\phi)}{\epsilon(\phi\mu\mu)}$  is recomputed with the varied efficiency. For muons with momenta smaller than 10 GeV the efficiency is increased by the uncertainty of the calibration procedure which is given in Figure 9.6. For muons with momenta larger than 10 GeV the efficiency is decreased by the uncertainty of the calibration procedure. The difference of the tracking efficiencies is taken as systematic uncertainty. The systematic uncertainty due to varied muon ID efficiency is given in Table 9.1.

## 9.7 Explicit Reweighting

The explicit reweighting in track multiplicity,  $B_s$  vertex  $\chi^2$  and  $B_s$   $p_T$  can have an effect on the relative selection efficiency  $\frac{\epsilon(J/\psi\phi)}{\epsilon(\phi\mu\mu)}$ . To estimate the effect on the relative branching ratio the relative efficiency is recomputed without the reweighting. The observed deviation from the nominal result is given in Table 9.1.

## 9.8 Particle Identification Resampling

The resampling of the PID variables as described in Chapter 6 can have an effect on the relative selection efficiency for the signal and the reference channels. To estimate the systematic effect the relative efficiency is recalculated using the original PID values of the Monte Carlo simulated events.

## 9.9 Signal Mass Model

To determine the influence of the mass model chosen for the fit of the signal yields the yields are also calculated for the double Gaussian mass model. The resulting systematic uncertainty is given in Table 9.1.

## 9.10 Branching Fraction of the Normalization Channel

The branching fraction of the reference channel is given by  $Br(B_s \rightarrow J\psi\phi) = (1.07 \pm 0.09) \cdot 10^{-3}$  [12]. The uncertainty of this value leads to an additional systematic uncertainty in the determination of the differential branching fraction  $Br(B_s \rightarrow \phi\mu^+\mu^-)$ .

Section	Systematic	Systematic uncertainties [ $10^{-5}$ ] in $q^2$ [GeV $^2$ ] bins							
		[0.1,2]	[2,5]	[5,8]	[11,12.5]	[15,17]	[17,19]	[1,6]	[15,19]
9.1	Peaking Bkg.	0.15	0.27	0.25	0.09	0.04	0.01	0.42	0.05
9.2	$Br(J/\psi \rightarrow \phi\mu\mu)$	0.06	0.04	0.05	0.04	0.05	0.04	0.07	0.08
9.3	Rel. efficiency	0.08	0.06	0.06	0.05	0.06	0.08	0.07	0.09
9.4	$q^2$ binning	0.23	0.13	0.15	0.10	0.06	0.19	0.23	0.13
9.5	Tracking	0.04	0.02	0.01	0.02	0.11	0.13	0.03	0.21
9.6	Muon ID	0.03	0.01	0.01	0.01	0.01	0.01	0.01	0.01
9.7	Reweighting	0.08	0.02	0.06	0.03	0.04	0.23	0.09	0.19
9.8	PID Resampling	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02
9.9	Sig. model	0.15	0.10	0.01	0.07	0.06	0.07	0.15	0.13
	$\Sigma$	0.34	0.33	0.31	0.17	0.17	0.35	0.52	0.36

Table 9.1: Systematic uncertainties on the differential branching fraction ratio  $Br(B_s \rightarrow \phi\mu^+\mu^-)/Br(B_s \rightarrow J/\psi\phi)$ .

# 10 Results and Conclusion

## 10.1 Result

The relative branching fraction  $\frac{Br(B_s \rightarrow \phi\mu^+\mu^-)}{Br(B_s \rightarrow J/\psi\phi)}$  has been determined in Chapter 8, the results are given in Table 8.5. The differential branching fraction of the signal decay in a  $q^2$  interval  $[q_{min}^2, q_{max}^2]$  is determined using

$$\frac{dBr(B_s \rightarrow \phi\mu^+\mu^-)}{dq^2} = \frac{1}{q_{max}^2 - q_{min}^2} \frac{N_{\phi\mu^+\mu^-}}{N_{J/\psi\phi}} \frac{\epsilon_{tot}(J/\psi\phi)}{\epsilon_{tot}(\phi\mu^+\mu^-)} \cdot Br(B_s \rightarrow J/\psi\phi) \cdot Br(J/\psi\phi \rightarrow \mu^+\mu^-).$$

The branching fractions used are  $Br(B_s \rightarrow J/\psi\phi) = (1.07 \pm 0.09) \cdot 10^{-3}$  [12] and  $Br(J/\psi\phi \rightarrow \mu^+\mu^-) = 0.05961 \pm 0.00033$ . The resulting differential branching fractions are given in Table 10.1.

Summing over the full  $q^2$  range, excluding the charmonium mass window, the relative branching fraction is  $\frac{Br(B_s \rightarrow \phi\mu^+\mu^-)}{Br(B_s \rightarrow J/\psi\phi)} = (49.84_{-2.79}^{+2.87} \pm 0.7) \times 10^{-5}$ . A simple estimation using the generator level  $B_s \rightarrow \phi\mu^+\mu^-$  Monte Carlo signal events determines the fraction of events rejected by the charmonium mass windows to be 33.0%. The relative branching fraction is extrapolated to the full  $q^2$  range and is determined to be  $(74.34_{-4.16}^{+4.28} \pm 1.06) \times 10^{-5}$ . This corresponds to a total branching fraction of  $Br(B_s \rightarrow \phi\mu^+\mu^-) = 79.54_{-4.45}^{+4.58} \pm 1.13 \pm 6.69 \times 10^{-8}$  where the first error is statistical, the second is systematic and the third comes from the uncertainty of the branching ratio of the reference channel  $B_s \rightarrow J/\psi\phi$ .

## 10.2 Conclusion

The differential branching fraction of rare flavour-changing neutral current  $B_s \rightarrow \phi\mu^+\mu^-$  decay has been determined in six nominal bins of  $q^2$ . The results are presented in Table 10.1. The relative branching fraction  $Br(B_s \rightarrow \phi\mu^+\mu^-)/Br(B_s \rightarrow$

$q^2$ bin (GeV <sup>2</sup> )	$N_{sig}$	$dBr/dq^2$ (10 <sup>-8</sup> GeV <sup>-2</sup> )
$0.1 < q^2 < 2$	$84.74^{+9.71}_{-10.57}$	$6.44^{+0.74}_{-0.80} \pm 0.19 \pm 0.54$
$2 < q^2 < 5$	$58.94^{+9.45}_{-8.87}$	$2.74^{+0.44}_{-0.41} \pm 0.18 \pm 0.23$
$5 < q^2 < 8$	$72.95^{+10.82}_{-10.30}$	$3.04^{+0.45}_{-0.43} \pm 0.11 \pm 0.26$
$11 < q^2 < 12.5$	$69.14^{+10.33}_{-9.72}$	$4.62^{+0.69}_{-0.65} \pm 0.12 \pm 0.39$
$15 < q^2 < 17$	$84.71^{+10.54}_{-10.01}$	$4.64^{+0.58}_{-0.55} \pm 0.09 \pm 0.39$
$17 < q^2 < 19$	$49.11^{+7.88}_{-7.22}$	$3.77^{+0.60}_{-0.55} \pm 0.19 \pm 0.32$
$1 < q^2 < 6$	$99.71^{+12.47}_{-11.83}$	$2.77^{+0.35}_{-0.33} \pm 0.11 \pm 0.23$
$15 < q^2 < 19$	$134.49^{+13.05}_{-12.54}$	$4.09^{+0.40}_{-0.38} \pm 0.10 \pm 0.34$

Table 10.1: The  $B_s \rightarrow \phi\mu^+\mu^-$  signal yields and the differential branching fractions  $\frac{dBr(B_s \rightarrow \phi\mu^+\mu^-)}{dq^2}$  in the  $q^2$  bins. The first error is statistical, the second is systematic and the third is due to the uncertainties of the branching fraction of the reference channel  $B_s \rightarrow J/\psi\phi$ .

$J/\psi\phi$ ) is extrapolated to the full  $q^2$  range to yield

$$\frac{Br(B_s \rightarrow \phi\mu^+\mu^-)}{Br(B_s \rightarrow J/\psi\phi)} = (74.34^{+4.28}_{-4.16} \pm 1.06) \times 10^{-5}.$$

The world average of the branching fraction of the reference channel,  $Br(B_s \rightarrow J/\psi\phi) = (1.07 \pm 0.09) \times 10^{-3}$ , is used to determine the total branching fraction

$$Br(B_s \rightarrow \phi\mu^+\mu^-) = (7.954^{+0.458}_{-0.445} \pm 1.13 \pm 6.69) \times 10^{-7}$$

where the first error is statistical, the second is systematic and the third is due to the uncertainty of the branching fraction of the reference channel. This value is compatible with the previous measurement at LHCb [5] of  $(7.07^{+0.64}_{-0.59} \pm 0.17 \pm 0.71) \times 10^{-7}$  and with the world average value of  $(7.6 \pm 1.5) \times 10^{-7}$  [12]. The measured branching fraction is lower than the Standard Model theory predictions that predict a range from  $14.6^{-7}$  [8] to  $19.2 \times 10^{-7}$  [27], [15], but such theoretical calculations have an uncertainty of 20% – 30%.

The dominating uncertainty is due to the uncertainty of the branching fraction of the  $B_s \rightarrow J/\psi\phi$  decay. The second largest uncertainty is the statistical uncertainty. There is definitely room for improvement and most likely after the new LHC run the branching fraction of the  $B_s \rightarrow \phi\mu^+\mu^-$  can be estimated with higher precision.

# Appendix

# A Appendix

In this Appendix a few plots that are relevant to the analysis are presented. Figures A.1 - A.2 show the separation between the background and the signal distributions for the variables that are used to train BDT. Figures A.3 - A.4 show the fits to the invariant  $B_s$  mass in bins of  $q^2$ . Figure A.5 shows the max and min of the Monte Carlo simulated PID variables that are used for the BDT training. The distributions of these PID variables is comparable with the distributions for the  $B_s \rightarrow J\psi\phi$  reference channel data.

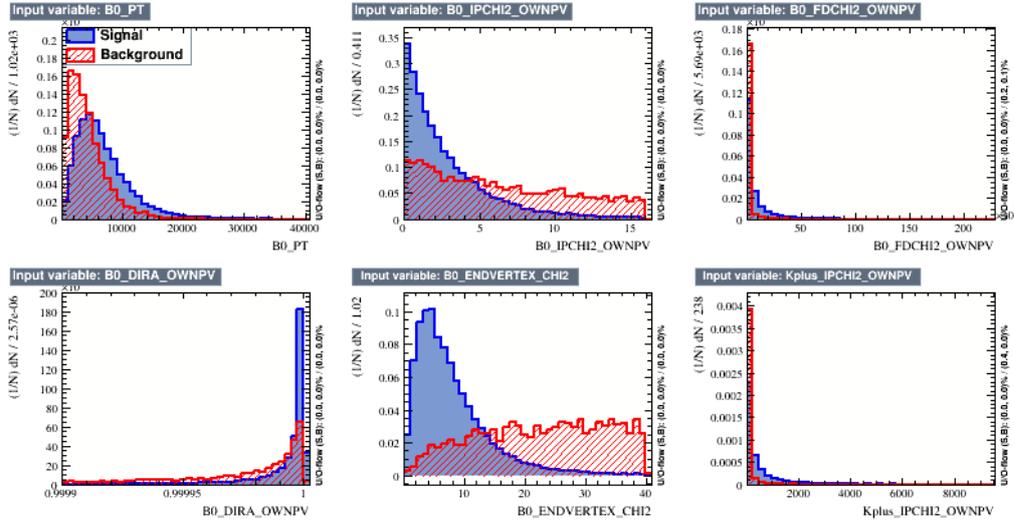


Figure A.1: Distributions of the BDT input variables. All input variables show large separation between signal and background.

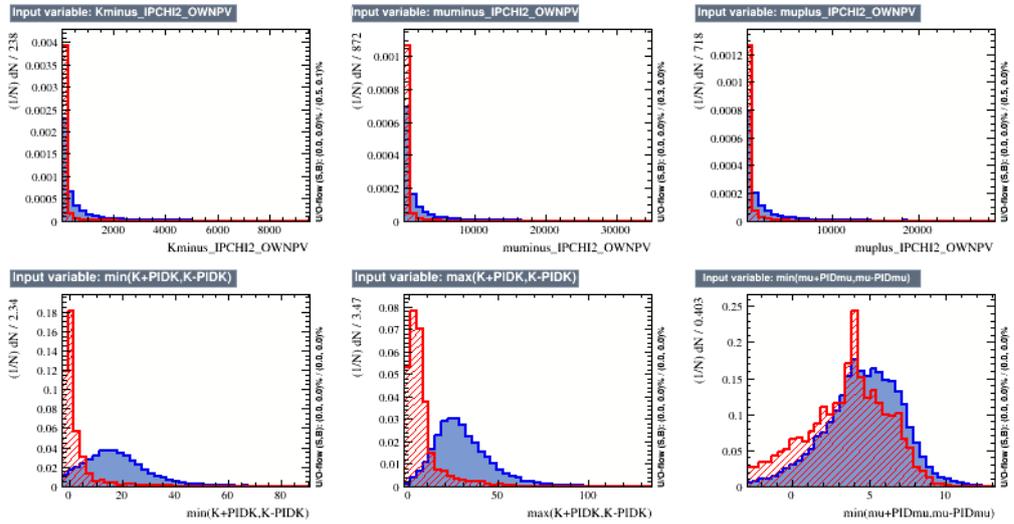


Figure A.2: Distributions of the BDT input variables. All input variables show large separation between signal and background.

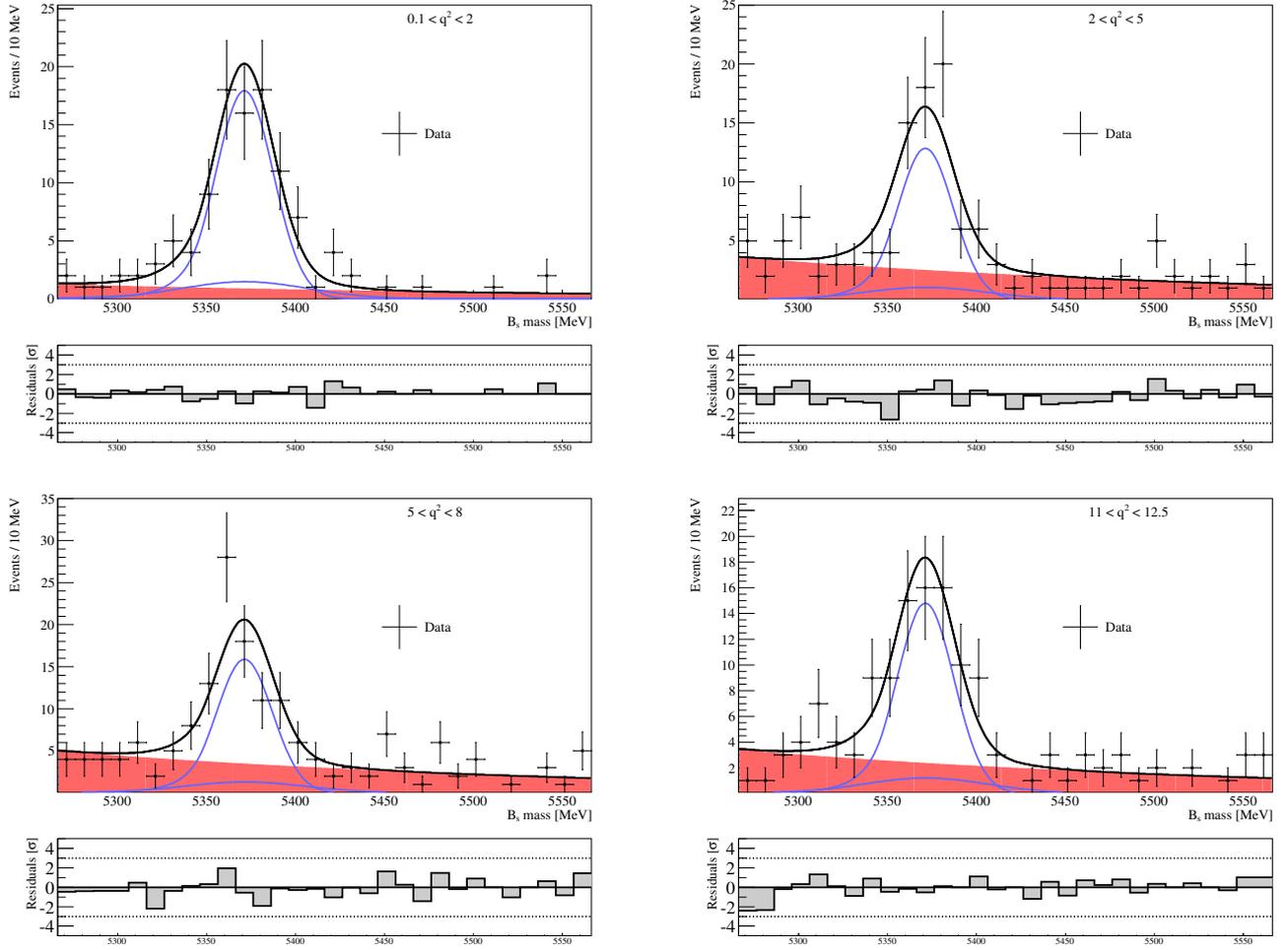


Figure A.3: Mass spectrum fits of the  $B_s \rightarrow \phi \mu^+ \mu^-$  signal channel in bins of  $q^2$ .

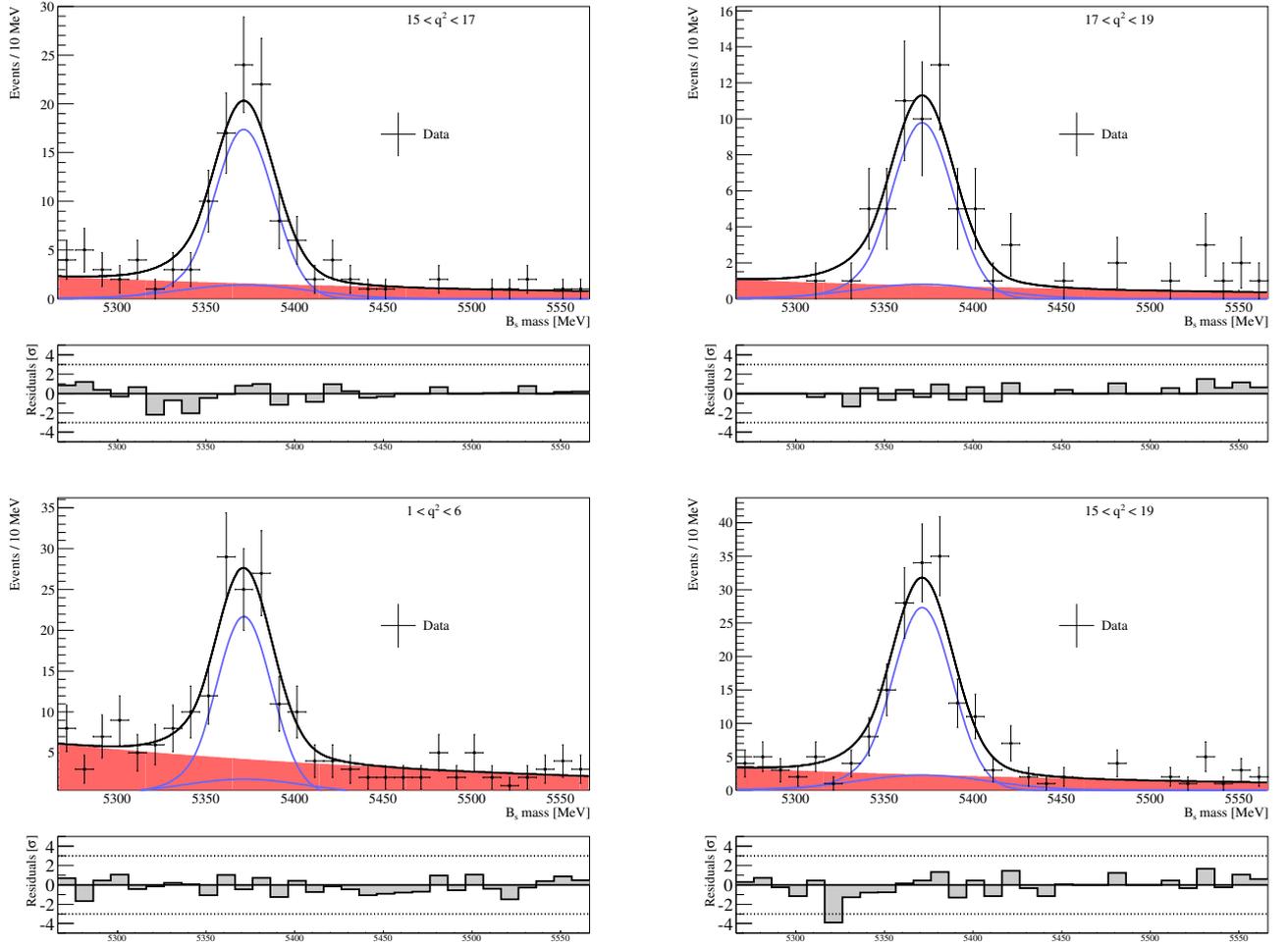


Figure A.4: Mass spectrum fits of the  $B_s \rightarrow \phi \mu^+ \mu^-$  signal channel in bins of  $q^2$ .

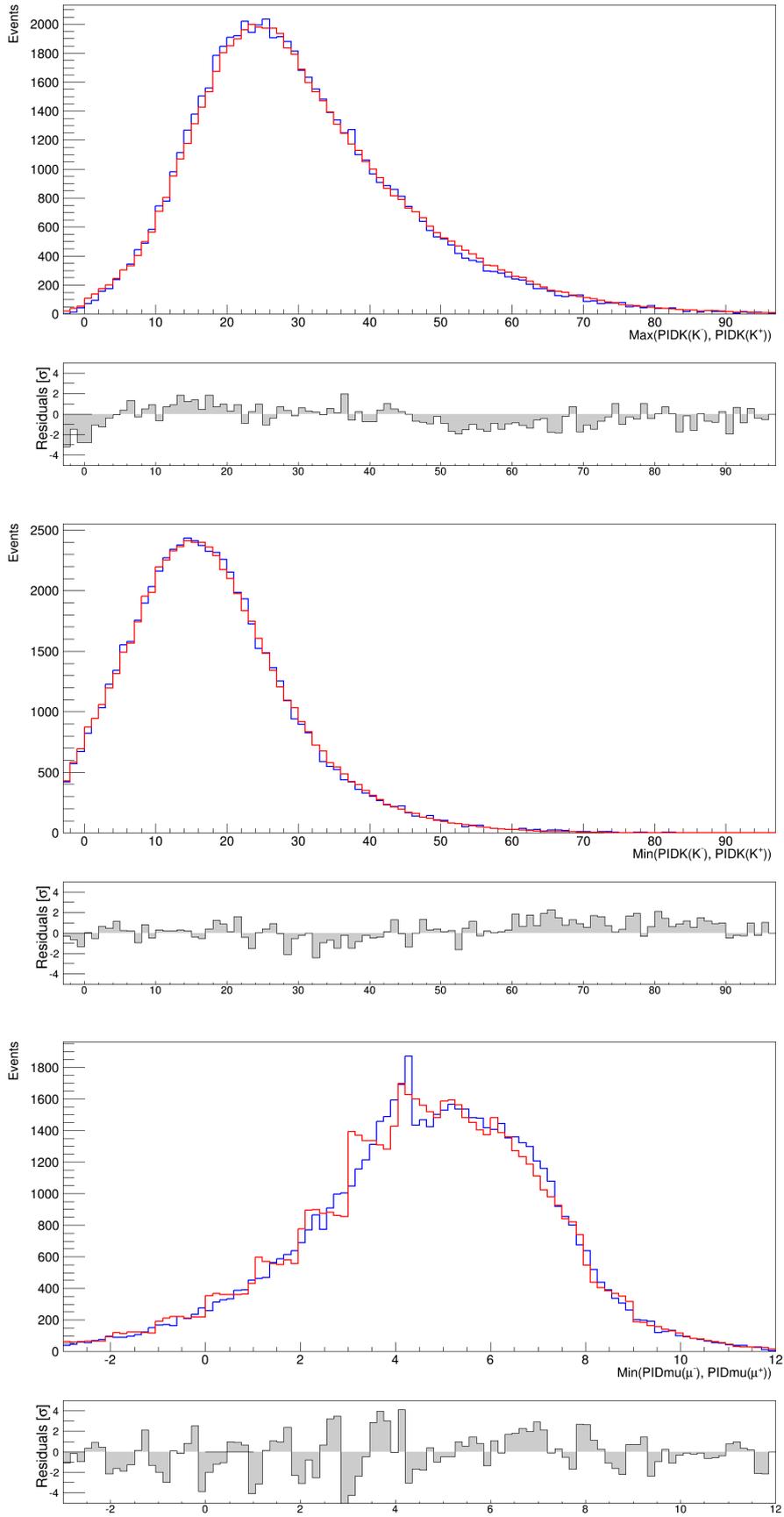


Figure A.5: Distributions of the PID BDT input variables from the  $B_s \rightarrow J/\psi\phi$  Monte Carlo simulated samples after all corrections described in Chapter 6 compared to the data. The distributions agree nicely.

# B Lists

## B.1 List of Figures

2.1	Constituents of the Standard Model [3] . . . . .	9
2.2	A schematic overview of different types of interactions described by each term of the electroweak Standard Model Lagrangian. . . . .	11
2.3	Decay of $B_s \rightarrow \phi\mu^+\mu^-$ mediated via (a) penguin diagrams, (b) box diagram. Figure taken from [22]. . . . .	14
2.4	Dependence of $BR(B_s \rightarrow \phi\mu^+\mu^-)$ on the mass of charged Higgs for a two Higgs doublet model. Solid lines show the prediction for the model, dotted lines show prediction for the SM . Figure taken from [15]. . . . .	14
3.1	Schematic overview of the LHC main experiments and underground accelerator ring. Taken from [17] . . . . .	15
3.2	Simulated distribution of polar angles $\Theta_1$ and $\Theta_2$ of produced $b\bar{b}$ pairs. Most of the $b$ and $\bar{b}$ are produced in either forward or backward direction. The LHCb acceptance is colored red. Taken from [17] . . . . .	17
3.3	Vertical cross-section of the LHCb detector. Taken from [17] . . . . .	18
3.4	Cross section in the x-z plane of the VELO sensors. The VELO consists of 21 stations along the beam pipe. Until the beam stabilizes the modules are kept open. Taken from [11] . . . . .	20
3.5	Cherenkov angle versus particle momentum for the RICH radiation. Taken from [11] . . . . .	21
3.6	. . . . .	22
3.7	Side view schematic layout of LHCb muon system. Taken from [11] . . . . .	23
3.8	Left: four regions of one quadrant, R1-R4. Each rectangle represents a chamber. Right: division into logical pads of a chamber in each region. Taken from [11] . . . . .	24
3.9	. . . . .	25
3.10	Overview over LHCb trigger system. Taken from [17] . . . . .	26
4.1	Two dimensional plot of the invariant $B_s$ mass versus the invariant dimuon mass. The signal region in and the $q^2$ bins in which measurements are performed are marked. . . . .	31
5.1	Distributions of the invariant mass of the $B_s \rightarrow \phi\mu^+\mu^-$ candidates after the stripping. . . . .	33

5.2	Distributions of invariant mass of $B_s$ versus the invariant dimuon mass after the cutbased selection after the cutbased selection. . . . .	36
5.3	Distributions of invariant mass of $B_s$ versus the invariant dimuon mass after the cutbased and the BDT selection has been applied. . .	37
5.4	BDT response distributions for the signal and the background training and control samples. . . . .	38
5.5	Signal efficiency vs background rejection ROC curve. The BDT cut point is given as black dot. . . . .	38
6.1	Ratios of muon ID efficiencies in bins of momentum and transverse momentum. . . . .	41
6.2	Ratios of tracking efficiencies in bins of momentum and pseudorapidity.	42
6.3	Comparison of BDTs trained using ProbNN and CombDLL PID variables. The BDT cut points are given as dots. . . . .	43
6.4	Distributions of PID variables of $K^-$ in the real data and in the Monte Carlo generated sample after the PID resampling. To the left is the distribution of the PIDK( $K^-$ ) variable, to the right is the distribution of the ProbNNk( $K^-$ ) variable. . . . .	44
6.5	. . . . .	45
6.6	Distributions of PID variables of $K^+$ in the real data and MC generated data. To the left are the distributions of ProbNNk, to the right are distributions of PIDK. . . . .	47
6.7	BDT output after processing of MC simulated data compared to BDT output of real data. . . . .	47
6.8	. . . . .	48
6.9	To the left - comparison of track $\chi^2$ of $K^+$ of the $B_s \rightarrow \phi\mu^+\mu^-$ data and the calibration samples. To the right - comparison of the ProbNNK( $K^+$ ) distributions for the $B_s \rightarrow \phi\mu^+\mu^-$ data and the calibration samples with the track $\chi^2$ weights applied. . . . .	49
7.1	Distribution of $B_s \rightarrow J/\psi(\rightarrow \mu^+\mu^-)\phi(\rightarrow K^+K^-)$ Monte Carlo events after full selection. . . . .	50
7.2	Distribution of $B_s \rightarrow J/\psi\phi$ Monte Carlo events after full selection where a $\mu$ is misidentified as $K$ . $K$ $\mu$ vetoes were not applied in this plot. . . . .	51
7.3	Distribution of $B_d^0 \rightarrow K^*\mu\mu$ Monte Carlo events after full selection without the PID cut applied. . . . .	52
7.4	Distribution of $B_d^0 \rightarrow K^*\mu\mu$ Monte Carlo events after full selection with the PID cut applied. . . . .	53
7.5	Distribution of $B_d^0 \rightarrow J/\psi K^*$ Monte Carlo events after full selection. .	55
7.6	Distribution of $\Lambda_b \rightarrow \lambda(1520)(\rightarrow pK)\mu\mu$ Monte Carlo events after full selection without the $K \rightarrow p$ veto applied. . . . .	56
7.7	Distribution of $\Lambda_b \rightarrow \lambda(1520)(\rightarrow pK)\mu\mu$ Monte Carlo events after full selection with the $K \rightarrow p$ veto applied. . . . .	57

7.8	Distribution of 2 million $B_s \rightarrow D_s^- \rightarrow (\phi\mu^-\bar{\nu}_\mu)\mu^+\nu_\mu$ generator level events. No events found in the signal region $5316.3 \text{ MeV} < m_{B_s} < 5416.3 \text{ MeV}$ . . . . .	58
7.9	Distribution of 2 million $B_s \rightarrow D_s^- \rightarrow (\phi\mu^-\bar{\nu}_\mu)\mu^+\nu_\mu$ generator level events. No events found in the signal region $5316.3 \text{ MeV} < m_{B_s} < 5416.3 \text{ MeV}$ . . . . .	59
8.1	Fit of the invariant mass spectrum of the reference channel $B_s \rightarrow J/\psi\phi$ using double Gaussian to model the signal and an exponential to model the background. Linear scale to the left, log-scale to the right. Results of the fit are given in Table 8.1. . . . .	63
8.2	Fit of the invariant mass spectrum of the reference channel $B_s \rightarrow J/\psi\phi$ using using double Crystal Ball function to model the signal and an exponential to model the background. Normal scale to the left, log scale to the right. Results of the fit are given in Table 8.2. . . . .	64
8.3	A single Crystal Ball fit to the invariant mass spectrum of the $B_s \rightarrow J/\psi\phi$ Monte Carlo events. . . . .	65
A.1	Distributions of the BDT input variables. All input variables show large separation between signal and background. . . . .	77
A.2	Distributions of the BDT input variables. All input variables show large separation between signal and background. . . . .	77
A.3	Mass spectrum fits of the $B_s \rightarrow \phi\mu^+\mu^-$ signal channel in bins of $q^2$ . . . . .	78
A.4	Mass spectrum fits of the $B_s \rightarrow \phi\mu^+\mu^-$ signal channel in bins of $q^2$ . . . . .	79
A.5	Distributions of the PID BDT input variables from the $B_s \rightarrow J/\psi\phi$ Monte Carlo simulated samples after all corrections described in Chapter 6 compared to the data. The distributions agree nicely. . . . .	80

## B.2 List of Tables

3.1	Monte Carlo simulated samples used in this analysis for efficiencies, acceptances and background determination. . . . .	28
4.1	$q^2$ binning. . . . .	30
5.1	Trigger line used for the selection of the signal decay . . . . .	34
5.2	Selection requirements for the cut based selection . . . . .	34
5.3	Input variables used in BDT . . . . .	39
8.1	Results of the double Gaussian fit to the $B_s$ invariant mass of the reference channel. . . . .	63
8.2	Results of the double Crystal Ball fit to the $B_s$ invariant mass of the reference channel. . . . .	64
8.3	Results of the single Crystal Ball function fit to the mass spectrum of the $B_s \rightarrow J/\psi\phi$ Monte Carlo events. . . . .	66

8.4	$q^2$ -dependent scaling factors for the widths of the Crystal Ball functions.	66
8.5	The signal yields of the $B_s \rightarrow \phi\mu\mu$ signal . . . . .	67
8.6	Table of efficiencies for $B_s \rightarrow \phi\mu^+\mu^-$ Monte Carlo. . . . .	68
8.7	Relative branching ratios $\frac{Br(B_s \rightarrow \phi\mu^+\mu^-)}{Br(B_s \rightarrow J/\psi\phi)}$ . . . . .	68
9.1	Systematic uncertainties on the differential branching fraction ratio $Br(B_s \rightarrow \phi\mu^+\mu^-)/Br(B_s \rightarrow J/\psi\phi)$ . . . . .	72
10.1	The $B_s \rightarrow \phi\mu^+\mu^-$ signal yields and the differential branching fractions $\frac{dBr(B_s \rightarrow \phi\mu^+\mu^-)}{dq^2}$ in the $q^2$ bins. The first error is statistical, the second is systematic and the third is due to the uncertainties of the branching fraction of the reference channel $B_s \rightarrow J/\psi\phi$ . . . . .	74

## C Bibliography

- [1] Geant4? a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003.
- [2] A brief introduction to {PYTHIA} 8.1. *Computer Physics Communications*, 178(11):852 – 867, 2008.
- [3] <http://upload.wikimedia.org/wikipedia/>, 2014.
- [4] <https://twiki.cern.ch/twiki/bin/view/LHCb/PIDCalibPackage>, 2014.
- [5] R. Aaij, C. Abellan Beteta, B. Adeva, M. Adinolfi, C. Adrover, A. Affolder, Z. Ajaltouni, J. Albrecht, F. Alessio, M. Alexander, S. Ali, G. Alkhazov, P. Alvarez Cartelle, A. A. Alves Jr, S. Amato, S. Amerio, Y. Amhis, L. Anderlini, J. Anderson, R. Andreassen, R. B. Appleby, O. Aquines Gutierrez, F. Archilli, and A. Zvyagin. Differential branching fraction and angular analysis of the decay  $B_s^0 \rightarrow \phi \mu^+ \mu^-$ . *J. High Energy Phys.*, 07(arXiv:1305.2168. CERN-PH-EP-2013-078. LHCb-PAPER-2013-017):084. 19 p, May 2013. Comments: Updated to correspond to proof from JHEP. Results unchanged.
- [6] R. Aaij, B. Adeva, M. Adinolfi, C. Adrover, A. Affolder, Z. Ajaltouni, J. Albrecht, F. Alessio, M. Alexander, S. Ali, and et al. Measurement of the differential branching fraction of the decay  $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ . *Physics Letters B*, 725:25–35, August 2013.
- [7] T. Aaltonen et al. Observation of the Baryonic Flavor-Changing Neutral Current Decay  $\Lambda_b \rightarrow \Lambda \mu^+ \mu^-$ . *Phys.Rev.Lett.*, 107:201802, 2011.
- [8] Q. Chang and Y.-H. Gao. Probe a family non-universal  $Z^?$  boson effects in  $B \rightarrow \phi \mu^+ \mu^-$  decay. *Nuclear Physics B*, 845:179–189, April 2011.
- [9] LHCb collaboration. The gauss project. <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/gauss/>, 2014.
- [10] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.
- [11] A. A. Alver Jr. et al. The lhcb detector at the lhc. [https://cds.cern.ch/record/1129809/files/jinst8\\_08\\_s08005.pdf](https://cds.cern.ch/record/1129809/files/jinst8_08_s08005.pdf), 2014.
- [12] K.A. Olive et al. (Particle Data Group). The review of particle physics. *Chin. Phys. C*, 38, 090001, 2014.

- [13] Alan D. Martin Francis Halzen. *Quarks and Leptons*. John Wiley and Sons, 1984.
- [14] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964.
- [15] Gursevil Turan Guray Erkol. The exclusive  $b_s \rightarrow \phi\mu\mu$  decay in the two higgs doublet models. *Eur.Phys.J. C25 (2002) 575-581*, 2002.
- [16] Heavy Flavor Averaging Group, Y. Amhis, S. Banerjee, R. Bernhard, S. Blyth, A. Bozek, C. Bozzi, A. Carbone, A. Oyanguren Campos, R. Chistov, G. Cibinetto, J. Coleman, J. Dingfelder, W. Dungel, M. Gersabeck, T. J. Gershon, L. Gibbons, B. Golob, R. Harr, K. Hayasaka, H. Hayashii, O. Leroy, D. Lopes Pegna, R. Louvot, A. Lusiani, V. Luth, B. Meadows, S. Nishida, M. Patel, D. Pedrini, M. Rama, M. Roney, M. Rotondo, O. Schneider, C. Schwanda, A. J. Schwartz, B. Shwartz, J. G. Smith, R. Tesarek, D. Tonelli, K. Trabelsi, P. Urquijo, and R. Van Kooten. Averages of b-hadron, c-hadron, and tau-lepton properties as of early 2012. *ArXiv e-prints*, July 2012.
- [17] [http://lhcb.web.cern.ch/lhcb/speakersbureau/html/Material\\_for\\_Presentations.html](http://lhcb.web.cern.ch/lhcb/speakersbureau/html/Material_for_Presentations.html), 2014.
- [18] Ian Aitchinson Anthony Hey. *Gauge Theories in Particle Physics*. Institute of Physics Publishin, 2004.
- [19] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.
- [20] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, and Helge Voss. TMVA: Toolkit for Multivariate Data Analysis. *PoS, ACAT:040*, 2007.
- [21] R. A. Olshen L. Breiman, J. H. Friedman and C. J. Stone. *Classification and regression trees*. Wadsworth international group, Belmont, California, USA, 1984.
- [22] C. Langenbruch. Differential branching fraction and angular analysis of the decay  $b_s \rightarrow \phi\mu\mu$ .
- [23] T. L. collaboration LHCb collaboration. Tagged time-dependent analysis of  $B_s^0 \rightarrow J/\psi K^+ K^-$  and  $B_s^0 \rightarrow J/\psi \pi^+ \pi^-$  decays with  $1fb^{-1}$ , 2014. LHCb-ANA-2012-067.
- [24] M. Pivk and F. R. Le Diberder. Plots: A statistical tool to unfold data distributions. *Nuclear Instruments and Methods in Physics Research A*, 555:356, 2005.

- [25] R. E. Schapire and Y. Freund. A decision-theoretic generalization of on-line learning and an application to boosting. *Jour. Comp. and Syst Sc.*, 55:119, 1997.
- [26] T. Skwarnicki. *A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances*, *Ph.D Thesis*. PhD thesis, DESY, 1986.
- [27] U. O. Yilmaz. Analysis of  $B_s \rightarrow \phi^+ l^-$  decay with new-physics effects. *European Physical Journal C*, 58:555–568, December 2008.

Erklärung:

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

Heidelberg, den (Datum) .....