Department of Physics and Astronomy
Heidelberg University

Bachelor Thesis in Physics
submitted by

Laura Luisa Scholl

born in Weinheim (Germany)

2024

Measurement of the branching fraction of the rare decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$

This Bachelor Thesis has been carried out by Laura Luisa Scholl at the
Physikalisches Institut in Heidelberg
under the supervision of
Dr. Christoph Langenbruch

Abstract

This thesis presents a measurement of the branching fraction of the rare B-meson decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ relative to the resonant decay $B^+ \to J/\psi K^+\pi^+\pi^-$. The analysis is performed on data collected by the LHCb Collaboration in the years 2011 and 2012.
Using data corresponding to 3 fb^{-1}, 16425 $\pm$ 1358 $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ decays and 2042318 $\pm$ 7376 $B^+ \to J/\psi K^+\pi^+\pi^-$ decays are reconstructed, resulting in a relative branching fraction of

$$\frac{\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-)}{\mathcal{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^+\pi^+\pi^-)} = (8.13 \pm 0.67\ (\text{stat})) \cdot 10^{-3}.$$

The total branching fraction is determined to be

$$\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-) = (4.55 \pm 0.38(\text{stat}) \pm 0.74(\text{syst})) \cdot 10^{-7}.$$

Kurzfassung

Diese Arbeit präsentiert eine Messung des Verzweigungsverhältnisses des seltenen B-Mesonzerfalls $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ relativ zum resonanten Zerfall $B^+ \to J/\psi K^+\pi^+\pi^-$. Die Analyse wird auf von der LHCb Kollaboration aufgenommenen Daten aus den Jahren 2011 und 2012 durchgeführt.
Mit Daten, die einer integrierten Luminosität von 3 fb^{-1} entsprechen, werden 16425 $\pm$ 1358 $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ Zerfälle und 2042318 $\pm$ 7376 $B^+ \to J/\psi K^+\pi^+\pi^-$ Zerfälle rekonstruiert, woraus das relative Verzweigungsverhältnis

$$\frac{\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-)}{\mathcal{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^+\pi^+\pi^-)} = (8.13 \pm 0.67\ (\text{stat})) \cdot 10^{-3}.$$

folgt. Das absolute Verzweigungsverhältnis wird zu

$$\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-) = (4.55 \pm 0.38(\text{stat}) \pm 0.74(\text{syst})) \cdot 10^{-7}$$

berechnet.

Contents

1 Introduction

The development of the Standard Model (SM) has been one of the greatest achievements of modern physics. It summarises all known elementary particles and three out of four fundamental forces that govern interactions between them. This theory provides a description of a wide variety of experimental results and has successfully predicted phenomena before they have been observed, the most prominent example being the Higgs boson, which was theorised by Peter Higgs in 1964 [1] and discovered almost 40 years later in 2012 [2].

However, the SM fails to explain some phenomena such as baryon asymmetry or dark matter. Therefore, the search for New Physics, beyond the SM, is of great interest for particle physicists. One approach to this pursuit is the search for virtual contributions of new particles in loop-level processes. Since flavour changing neutral currents (FCNC) are forbidden on tree-level and may only occur as higher-order loop processes within the SM, they provide an interesting opportunity for investigating New Physics contributions as the relative contribution from New Physics could be sizeable. One instance of a rare decay proceeding via $b \to s$ FCNC is the $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ decay, which will be studied in this thesis. The tree-level decay $B^+ \to J/\psi K^+\pi^+\pi^-$ is used as a normalisation channel.

An introduction of the Standard Model and the physics underlying rare decays is presented in Section 2. The LHCb detector is described in Section 3. The analysis strategy is summarised in Section 4. Datasets, selection cuts and the multivariate analysis are documented in detail in Section 5. The reweighting procedure is explained in Section 6. The extraction of signal yields is disclosed in Section 7, followed by the calculation of the branching fraction, which is presented in Section 8.

2 Theoretical Background

This section provides an outline of the Standard Model based on [3] and [4], followed by a brief introduction to flavour physics and rare decays.

2.1 The Standard Model

The Standard Model is a quantum field theory (QFT) describing all known elementary particles and three out of four known fundamental forces - the electromagnetic, weak and strong force - as well as the force mediators. The fourth force, gravity, is not included and since the gravitational force between particles is particularly small compared to the other forces, it is negligible in particle physics.

2.1.1 Fundamental Particles

The SM contains twelve fermions, five gauge bosons and the Higgs boson. Fermions are particles with a half-integer spin and therefore obey the Fermi-Dirac statistics. An essential concept of the SM is the idea that for every fermionic particle, there is a corresponding antiparticle with the same mass but opposite charge. Each of these twelve fermions represents a certain species of particles, also known as flavour. Fermions are divided into leptons and quarks based on the type of interactions they experience as they have different physical properties, which are depicted in Table 2.1.
Fermions can be further categorized into three generations, where each succeeding generation has a higher mass. Each generation consists of two leptons and two quarks. One differentiates between up-type quarks (up, charm and top) that possess an electric charge of $+\frac{2}{3}e$ and down-type quarks (down, strange and bottom) with an electric charge of $-\frac{1}{3}e$. Quarks additionally carry a colour charge, a quantum-chromodynamic equivalent to electric charge, and interact via the strong force, but also through the electromagnetic and weak force. In contrast to leptons, quarks cannot exist individually due to colour confinement. Thus, they only exist in bound color-neutral states either consisting of a quark-antiquark pair hadronising to a meson or three quarks or antiquarks forming a baryon.
Leptons do not carry a colour charge and can therefore not take part in strong interactions. Charged leptons, namely electrons, muons and tauons, may interact through the electromagnetic and weak force. Neutral leptons, the neutrinos, do not carry an electric charge and therefore only interact weakly.
The three forces are mediated through gauge bosons, the gluon, photon, $W^{\pm}$ and Z^0

Table 2.1: The three generations of fundamental fermions divided into leptons and quarks. Values are taken from [5].

	Leptons				Quarks			
	Particle		Q[e]	Mass[GeV]	Particle		Q[e]	Mass[GeV]
First	electron	e^-	-1	0.0005	up	u	$+2/3$	0.002
generation	neutrino	ν_e	0	$< 10^{-10}$	down	d	$-1/3$	0.005
Second	muon	μ^-	-1	0.106	charm	c	$+2/3$	1.3
generation	neutrino	ν_μ	0	$< 10^{-10}$	strange	s	$-1/3$	0.09
Third	tau	τ^-	-1	1.78	top	t	$+2/3$	173
generation	neutrino	ν_τ	0	$< 10^{-10}$	bottom	b	$-1/3$	4.2

Table 2.2: The fundamental gauge bosons and the Higgs boson [5].

Boson	Q[e]	Mass [GeV]	coupling
Gluon g	0	0	strong
Photon γ	0	0	electromagnetic
$W^\pm$	$\pm$ 1	80.4	weak, electromagnetic
Z^0	0	91.2	weak
Higgs H^0	0	125	mass

bosons. These are spin 1 particles, hence obey Bose-Einstein statistics. Their properties are listed in Table 2.2. The latest addition to the Standard Model's elementary particles is the Higgs boson with a spin of 0, making it the only scalar boson in the SM.

2.1.2 Fundamental Forces

Each relevant force in particle physics is described by a quantum field theory characterised by their corresponding gauge bosons. The first gauge boson to be discovered is the photon, which mediates electromagnetic interactions, i.e. interactions between charged particles. The photon is electrically neutral and massless, such that the interaction range is infinite. The gauge bosons mediating strong interactions are called gluons. They are massless and electrically neutral like the photon, but additionally carry a combination of colour and anti-colour, forming an octet of coloured states that mediates interactions between quarks.

Weak interactions, which all fermions can participate in, are mediated by massive bosons. While the W^+ and W^- bosons couple fermions differing in one unit of electric charge, the Z^0 mediates weak neutral-current interactions.

2.2 Flavour Physics

Flavour physics refers to the research field revolving around the weak interactions of quarks and leptons. Within the SM, one differentiates between weak-charged currents where flavour and charge both change, and weak-neutral currents where both properties are conserved. Both processes occur on tree-level.
One phenomenon of particular interest is the variation of coupling strengths between different generations. For example, the coupling between u- and s-quarks is weaker than the one between u- and d-quarks. A mathematical formulation of this observation was presented in 1973 via the Cabibbo-Kobayashi-Maskawa (CKM) formalism. It proposes that the weak eigenstates of quarks differ from their mass eigenstates. The unitary CKM-matrix yields a direct relation of the weak eigenstates q' to the mass eigenstates q by

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.1}$$

The probability of a flavour eigenstate i transitioning to an eigenstate j through the coupling to a W-boson is proportional to $|V_{ij}|^2$. Since electrical charge is conserved, i must be an up-type quark flavour while j is a down-type.
The structure of the CKM-matrix also provides an explanation of why the transition of flavours between two generations is suppressed compared to those of the same generation, as the off-diagonal terms are much smaller than the diagonal ones, making the matrix nearly diagonal. Therefore, it is beneficial to express the matrix as a Taylor expansion in the parameter λ, leading to the Wolfenstein parametrisation [6]. The CKM-matrix may now be rewritten in terms of four real parameters λ, A, ρ and η as

$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4). \tag{2.2}$$

Note that to $\mathcal{O}(\lambda^3)$ complex components reside in V_{ub} and V_{tb}, which are necessary for the occurence of CP-violation. This leads to η being non-zero. The Wolfenstein parameters are determined experimentally through a global fit using all available measurements [7]:

$$\lambda = 0.22500 \pm 0.00067, \qquad A = 0.826^{+0.018}_{-0.015}$$
$$\bar{\rho} = \rho\left(1 - \tfrac{\lambda^2}{2}\right) = 0.159 \pm 0.010, \qquad \bar{\eta} = \eta\left(1 - \tfrac{\lambda^2}{2}\right) = 0.348 \pm 0.010.$$

2.3 Search for New Physics

Since the Standard Model fails to correctly account for observations such as matter-antimatter asymmetry and is incompatible with general relativity as of now, it is crucial to research physics beyond the Standard Model, New Physics (NP). One approach to the search for New Physics are precision measurements of processes that could be influenced by virtual new particles, such as flavour-changing neutral currents (FCNC). Within these processes, the flavour of a quark changes but the electric charge is conserved, which is why they cannot occur at tree-level within the SM and only as higher-order electroweak loop-processes. However, these processes are loop- and CKM suppressed and therefore rare.
One instance of interactions proceeding through FCNC is rare decays. This makes them susceptible to New Physics contributions since contributions from NP could be sizeable and observables like branching fractions or amplitudes would be affected by such.

2.3.1 The Rare Decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$

The rare decay of interest in this thesis is $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$, which was first observed in 2014 by the LHCb collaboration at CERN using Run 1 data from the LHCb experiment [8]. It occurs through the transition of a b quark into an s quark, making it receptive to New Physics contributions, c.f. Figure 2.1. A significant contribution to the $K^+\pi^+\pi^-\mu^-\mu^-$ system emerges from the rare decay $B^+ \to K_1(1270)\mu^+\mu^-$. Since the resonance structure of the $K^+\pi^+\pi^-$ system is unknown, there are no theoretical predictions for the branching fraction of the rare decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$. The branching fraction of this rare decay as it was determined by the LHCb collaboration is listed in Table 2.3.
In this thesis, the branching fraction is determined relative to the normalisation channel, the tree-level decay $B^+ \to J/\psi K^+\pi^+\pi^-$. The J/ψ is reconstructed in its decay via $J/\psi \to \mu^+\mu^-$, such that the final state particles are the same as in the rare decay which cancels many systematic uncertainties at leading order. Another decay with the same final state particles which is studied in this thesis is the resonant decay $B^+ \to \psi(2S)K^+$. The $\psi(2S)$ then decays via $\psi(2S) \to J/\psi(\to \mu^+\mu^-)\pi^+\pi^-$. The world-averages for both decay channels are denoted in Table 2.4.

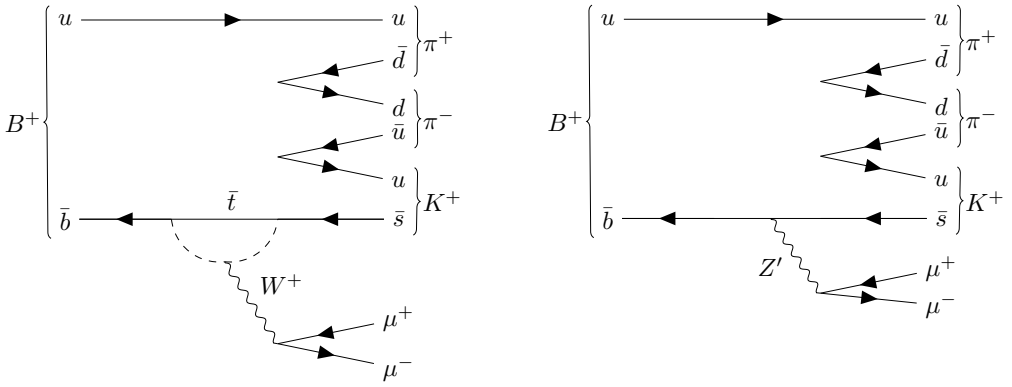


Figure 2.1: Higher order Feynman graph of the $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ decay in accordance with the SM (left) and the $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ decay considering a new exchange particle Z' (right).

Table 2.3: Experimental value for the branching fraction of the rare decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ by the LHCb experiment [8]. The first and second uncertainties are statistical and systematical respectively, the third stemming from uncertainties of the normalisation channel.

Value	Branching fraction
Experimental (LHCb)	$(4.36^{+0.29}_{-0.27} \pm 0.21 \pm 0.18) \cdot 10^{-7}$

Table 2.4: World-average experimental values for the branching fraction of the resonant decay $B^+ \to \psi(2S)K^+$ and the tree-level decay $B^+ \to J/\psi K^+\pi^+\pi^-$ as well as relevant subsequent decays [5].

Decay channel	Branching fraction
$B^+ \to \psi(2S)K^+$	$(6.24 \pm 0.20) \cdot 10^{-4}$
$B^+ \to J/\psi K^+\pi^+\pi^-$	$(8.1 \pm 1.3) \cdot 10^{-4}$
$\psi(2S) \to J/\psi\pi^+\pi^-$	$(34.68 \pm 0.30) \cdot 10^{-2}$
$J/\psi \to \mu^+\mu^-$	$(5.971 \pm 0.033) \cdot 10^{-2}$

3 The LHCb Experiment

The data analysed in this thesis were recorded by the LHCb collaboration in 2011 and 2012 during Run 1 of the Large Hadron Collider beauty (LHCb) experiment. It is one of the four major experiments at the Large Hadron Collider (LHC) operated by CERN and is designed to investigate decays involving bottom and charm quarks.
This section provides an overview of the accelerator and the LHCb detector as well as the subdetectors.

3.1 The Large Hadron Collider

The LHC is a particle-particle collider and accelerator near Geneva on the French-Swiss border. It consists of a 26.7 km long ring where two counter-rotating beams of protons or heavy ions are accelerated. The beams then collide at four interaction points corresponding to the locations of the main experiments - ALICE, ATLAS, CMS, and LHCb.
During the first phase of operation in 2011 and 2012, referred to as Run 1, it operated at centre-of-mass (CMS) energies of $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV respectively. For Run 2, the CMS energy was increased to $\sqrt{s} = 13$ TeV in the years 2015 - 2018.

3.2 The LHCb Detector

The LHCb detector is a single-arm forward spectrometer covering the pseudorapidity[1] range of $2 < \eta < 5$, corresponding to an angular range of 10 mrad to 300 (250) mrad in the bending (non-bending) plane [9]. This particular detector geometry was chosen since $b\bar{b}$ pairs are mainly produced on the same forward or backward cones at high energies. The polar angles of b and $\bar{b}$ quarks simulated for $\sqrt{s} = 14$ TeV are shown in Figure 3.1. The detector layout including all subdetectors during Run 1 is shown in Figure 3.2.

[1]The pseudorapidity η is defined as $\eta = -\ln(\tan(\frac{\theta}{2}))$ where θ is the polar angle with respect to the beam.

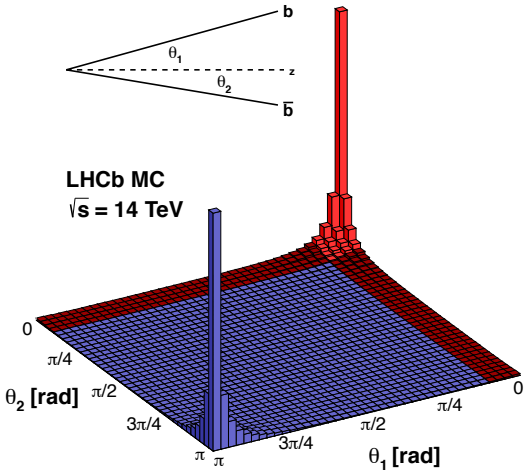


Figure 3.1: Simulation of the polar angles of b and $\bar{b}$ quarks at $\sqrt{s} = 14$ using PYTHIA8 and CTEQ6 NLO [10].

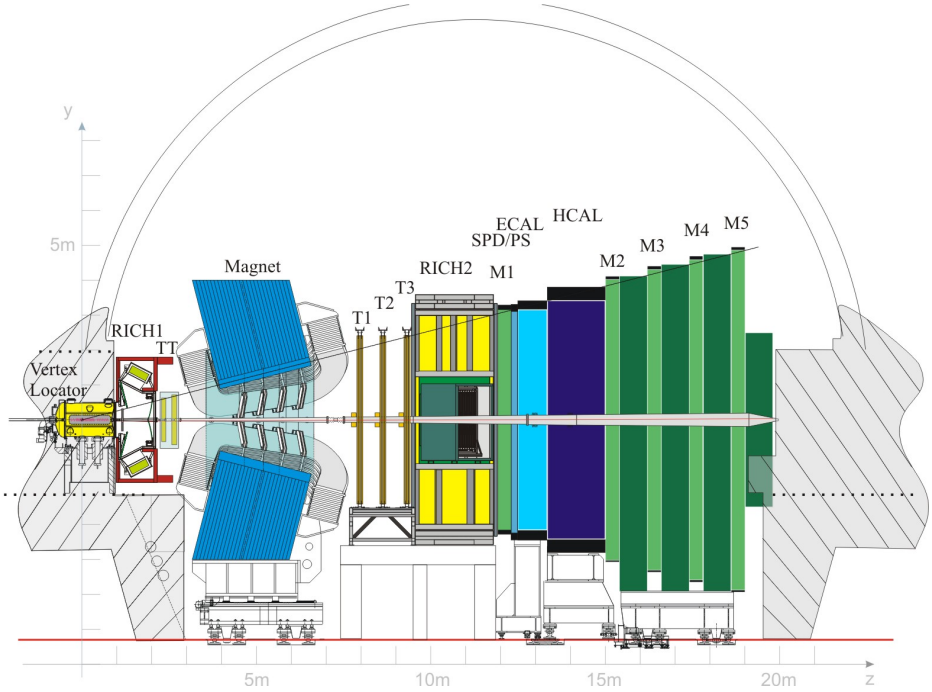


Figure 3.2: Cross section drawing of the LHCb detector in the y-z-plane [11].

3.2.1 Magnet

The momentum of charged particles traversing the detector is measured through the curvature of their tracks inside a magnetic field. The magnet used to provide the magnetic

field consists of two saddle shaped coils in a window-frame yoke with an increasing pole gap to meet the detector acceptance. The magnet generates a vertical magnetic field with an integrated magnetic field of 4 Tm along the z-axis. The magnet's polarity is inverted regularly during data taking, denoted as `MagUp` for positive and `MagDown` for negative polarity, to minimise effects arising from detector asymmetries.

3.2.2 Tracking System

To accurately reconstruct a particle's momentum, charge and vertex position, LHCb utilises silicon microstrip detectors and straw-tubes in the tracking system, consisting of the Vertex Locator and four tracking stations labeled as TT, T1, T2, T3 in Figure 3.2.

Vertex Locator: The Vertex Locator (VELO) is a silicon microstrip detector located around the proton-proton interaction point used to identify primary vertices (PV) and displaced secondary vertices. The 42 semi-circular sensors are placed perpendicularly to the beam at a radial distance of 7 mm to the beam axis. Each module contains one sensor for the measurement of the radial distance R to the beam and one for the azimuthal angle ϕ, providing a best hit resolution of 4 μm [12].

Silicon Tracker: The Silicon Tracker (ST) refers to two further silicon subdetectors, the Tracker Turicensis (TT) and Inner Tracker (IT). While the TT is located upstream of the magnet and covers the full detector acceptance, the IT only covers the centres of the tracking stations T_1 - T_3 downstream of the magnet and close to the beam. Each ST station has four detection layers in an (x-u-v-x) arrangement, i.e. vertical strips in the first and fourth layer, and strips rotated by a stereo angle of $-5°$ and $+5°$ in the second and third layer respectively. Both detectors have a single-hit resolution of about 50 μm.

Outer Tracker: The outer regions of the tracking stations T_1 - T_3 are covered by the Outer Tracker (OT) where the track multiplicity is lower than in the IT. The OT is a drift-time detector for the tracking of charged particles. It is an array of gas-tight straw-tube modules providing a spacial resolution in x of 200 μm [13]. Similarly to the ST, there are four detection layers in an (x-u-v-x) arrangement per station.

3.2.3 Particle Identification System

An important part of the reconstruction of a particle is correctly assigning its mass, thus identifying the particle type. The LHCb Particle Identification (PID) system utilises two Ring Imaging Cherenkov detectors, two calorimeter systems and muon chambers for that

task.

Ring-Imaging Cherenkov system: The Ring-Imaging Cherenkov (RICH) system is applied to discriminate charged hadrons from one another, mainly pions and kaons since they are commonly produced in B or D meson decays. This system measures the velocity-dependent emission angle of Cherenkov photons, which are emitted when a charged particle traverses a medium with a speed faster than the speed of light in that medium. Having determined the momentum and velocity, the mass of a particle can be reconstructed. The RICH system consists of two detectors, RICH1 and RICH2. The former is stationed upstream of the magnet covering the full angular range and the lower momentum range of 2 - 40 GeV/c using aerogel and C_4F_{10} radiators. RICH2 is positioned downstream of the magnet covering a higher momentum range of 15 -100 GeV/c using CF_4, with an angular range of 15 - 120 mrad [14].

Calorimeter system: The calorimeter system measures the energy and momentum of hadrons, electrons and photons through hadronic and electromagnetic (EM) showers. The calorimeters are located downstream of the RICH2. The first layer passed by a particle is the Scintillating Pad Detector (SPD) selecting charged particles, and allowing a distinction between electrons and photons. It is followed by the PreShower detector (PS), which is also a scintillating pad detector but with an additional 15 mm of lead in front of the scintillators, such that EM showers can occur in that layer, hence allowing a discrimination of hadrons and electrons. Photons and electrons are detected in the electromagnetic calorimeter (ECAL), while hadronic showers take place in the hadronic calorimeter (HCAL). Both are composed of alternating layers of scintillators and absorbers. Optimal energy resolution for the ECAL is achieved by containing the whole EM shower in the calorimeter, resulting in a thickness of 25 radiation lengths. Hadronic showers are reconstructed in the HCAL, which has a thickness of 5.6 interaction lengths.

Muon system: The muon system consists of five stations. M1 is located in front of the calorimeter system to improve the transverse momentum p_T measurement in the L0 trigger, M2 - M5 downstream. In between each station, an 80 cm thick iron absorber is placed to filter out hadrons, resulting in only muons with a momentum larger than 6 GeV/c passing all five stations [9]. Thus, M4 and M5 serve to identify particles with high p_T. M1 - M3 have a high spatial resolution in x, allowing a definition of the track direction and a calculation of p_T with a resolution of 20%.
Since the L0-muon trigger, which is discussed in Section 3.2.4, requires a five-fold coincidence among all stations, the efficiency of each station must be $\geq 99\%$ [15]. The average

muon identification efficiency is 97%, while the misidentification efficiency for pions is 1-3% [16].

3.2.4 Trigger System

The main objective of the trigger system is to reduce the data volume produced by proton-proton collisions to a manageable size, from a total event rate from about 40 MHz to 2 kHz [9]. That is achieved by using two trigger levels, the hardware-based Level-0 trigger and the software-based High Level Trigger.

Level-0 trigger: The Level-0 trigger (L0) reduces the beam crossing rate of 40 MHz to 1 MHz by using information from the calorimeters and the muon system [17]. The L0-calorimeter trigger processes information from the calorimeter system and computes the transverse energy E_T for clusters in 2x2 cells in the ECAL and the HCAL:

$$E_T = \sum_{i=1}^{4} E_i \sin\theta_i. \tag{3.1}$$

E_i denotes the energy deposited in cell i, θ_i the angle between the z-axis and the line connecting the mean position of the pp collision to the cell center. Events with high E_T hadrons, photons or electrons are accepted if their energy passes a certain fixed threshold using information from the SPD and PS.
For this analysis, the muon trigger is of particular interest. The L0-muon trigger searches for events with hits that define a straight line through all five muon stations and with an origin close to the interaction point in the $y-z$ plane. In order for events to be accepted, either the p_T of the muon with the highest p_T has to exceed the `L0Muon` p_T threshold, or the product of the p_T of the muons with the largest and second largest transverse momenta has to exceed the `L0DiMuon` p_T threshold.

High Level Trigger: The High Level Trigger (HLT) is divided into two stages, HLT1 and HLT2. HLT1 partially reconstructs events that passed the L0 trigger stage using information from the VELO and the T-stations, either confirming or rejecting them. The remaining event rate of about 30 kHz then allows a full pattern recognition in HLT2, such that the data can be written to storage for further analysis.

4 Analysis Strategy

The aim of this thesis is to calculate the branching fraction of the rare decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ using the tree-level decay $B^+ \to J/\psi K^+\pi^+\pi^-$ as a normalisation channel. The J/ψ subsequently decays via $J/\psi \to \mu^+\mu^-$, such that the final state particles are the same as in the rare decay. This strategy is beneficial since most systematic uncertainties are expected to cancel each other out at leading order, for example muon identification and reconstruction. Moreover, the branching fraction of the normalisation channel has been established [5]. The rare decay has already been observed as well by the LHCb Collaboration [8], which is why this analysis is performed without blinding.

The analysis strategy consists of the following steps:

1. Dataset stripping: A series of loose preselection cuts is applied to the datasets produced by the LHCb experiment to reduce the computing effort of the analyses.

2. Signal preselection: Another set of loose cuts is applied to extract the signal and reduce combinatorial background, i.e. background arising from combinations of particle pairs that do not originate from the same decay.

3. Signal selection: The combinatorial background is further reduced by performing a multivariate analysis. A Boosted Decision Tree (BDT) is trained on a data sample from the high B mass sideband as signal proxy and simulated Monte Carlo samples serving as a signal proxy.

4. Reweighting: Signal candidates for $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to J/\psi K^+\pi^+\pi^-$ events are reweighted with their inverse efficiency to account for acceptance effects. The BDT reweighter is trained on Monte Carlo simulations at generator-level and reconstructed and selected candidates.

5. Signal yield determination: The reweighted mass distributions of $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to J/\psi K^+\pi^+\pi^-$ are fitted by using extended weighted unbinned maximum likelihood fits.

6. Branching fraction calculation: The relative branching fraction of the rare decay to the resonant decay is calculated using

$$\frac{\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-)}{\mathcal{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^+\pi^+\pi^-)} = \frac{N_{K\pi\pi\mu\mu}}{N_{J/\psi K\pi\pi}} \cdot \frac{\epsilon^{J/\psi K\pi\pi}_{DecProd}}{\epsilon^{K\pi\pi\mu\mu}_{DecProd}}. \tag{4.1}$$

N denotes the yields from the weighted fits and $\epsilon_{DecProd}$ the geometrical efficiency. The geometrical efficiency accounts for the limited angular range of the detector, as only a certain fraction of events ends up in the detector's acceptance.

5 Data Analysis

This section documents the variables and datasets used for the analysis.

5.1 Definition of Variables

Variables that are used for calculations or training the BDT classifier are defined and explained below.

Invariant mass: The invariant mass of a particle X that is combined from e.g. tracks is calculated from its four-momentum vector $p = (E, p_x, p_y, p_z)^T$ provided by the tracking system, with E being the energy and $p_x,\ p_y,\ p_z$ the momentum components in $x,\ y$ and z direction respectively, by

$$m_X = \sqrt{p_X^2}. \tag{5.1}$$

Transverse momentum: The transverse momentum is defined as the momentum component perpendicular to the beam line (the z-axis)

$$p_T = \sqrt{p_x^2 + p_y^2}. \tag{5.2}$$

DLL: The Delta Log-Likelihood (DLL) of a particle X is defined as the difference of the logarithmic likelihood of X to the logarithmic likelihood of a pion

$$\Delta \ln(\mathcal{L}(X - \pi)) = \ln \mathcal{L}(X) - \ln \mathcal{L}(\pi) = \ln \left(\frac{\mathcal{L}(X)}{\mathcal{L}(\pi)} \right). \tag{5.3}$$

The likelihood function $\mathcal{L}$ is computed by the PID system.

Track and vertex fit quality: The χ^2 of a track or vertex fit is a measure for its fit quality.

Impact parameter: The impact parameter (IP) of a particle is the minimal distance between the particle's reconstructed track and the primary vertex (PV). The χ^2_{IP} is defined as the difference of the PV's χ^2 before and after the track was added, thus providing a measure for the likelihood of the particle track originating from said PV.

Flight distance: The flight distance is the distance a particle has travelled before decaying.

Direction angle: The direction angle (DIRA) is defined as the cosine of the angle between the reconstructed momentum vector and flight direction. The flight direction is determined by the vector connecting the PV and the secondary vertex.

Ghost probability: A particle's ghost probability P_{ghost} is the probability of the particle track being a fake track from combinations of random hits in the tracking system.

ISMUON: Binary variable classifying whether candidate is compatible with being a muon based on muon station and calorimeter information.

HASMUON: Binary variable classifying whether muon candidate has information from the muon stations assigned.

5.2 Datasets

The dataset used in this analysis was recorded by the LHCb in 2011 and 2012 with a centre-of-mass energy of $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV respectively with both magnet polarities `MagUp` and `MagDown`. Data correspond to an integrated luminosity of 3 fb^{-1} for Run 1.
In addition to real data, phase-space (PHSP) Monte Carlo simulations are used for both the rare and normalisation channel during several stages of the analysis. Simulations are used for training the BDT classifier on a pure signal sample as well as unfolding efficiency distributions which are used to reweight data. They have to pass the same reconstruction and preselection criteria as the real datasets, thus making it possible to extract pure, "truthmatched" signal samples. The simulations with the corresponding geometrical efficiencies used for later calculations are listed in Table 5.1.

5.3 Stripping

Reconstructed candidates for $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to J/\psi K^+\pi^+\pi^-$ decays must pass a series of loose selection requirements based on their decay topology, the process being commonly referred to as stripping. The used stripping line is `Bu2KpipiMM`, considering only decays with the requested initial and final state particles. The requirements are listed in Table 5.2.

5.4 Preselection

In order to reduce background contributions from resonant decays and separate the different decay channels from one another, cuts on the invariant mass of the dimuon-system, referred to as q^2 in the following, are applied, ruling out that the muons originate from charmonium decays (J/ψ or $\psi(2S)$), as depicted in Table 5.3. Mass requirements for the normalisation and control channel are listed in Table 5.4. The q^2 of normalisation channel

candidates is required to be within 50 MeV/c^2 of the known J/ψ mass. For the control channel, the invariant mass of the $\pi^+\pi^-\mu^+\mu^-$ system must lie within 60 MeV/c^2 of the known $\psi(2S)$ mass. Additional preselection cuts to reduce combinatorial background are listed in Table 5.5.
While the B mass peaks are clearly visible in the normalisation and control channels (see Figures 5.1d and 5.1c), the rare channel is still heavily dominated by combinatorial background (see Figure 5.1b).

Table 5.1: Monte Carlo simulated samples used in this analysis. Up and down refer to the polarity of the magnet. N_{gen} is the number of generated events, N_{reco} the number of reconstructed and selected ones.

Year	Decay	N_{gen}	N_{reco}	$\epsilon_{\text{DecProd}}$
2011	$B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ up	501308	44975	0.14549 ± 0.00049
2011	$B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ down	500599	45265	0.14629 ± 0.00049
2012	$B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ up	948205	78917	0.14865 ± 0.00051
2012	$B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ down	939668	78256	0.14959 ± 0.00051
2011	$B^+ \to J/\psi K^+\pi^+\pi^-$ up	501187	43531	0.14739 ± 0.00049
2011	$B^+ \to J/\psi K^+\pi^+\pi^-$ down	500246	43632	0.14749 ± 0.00051
2012	$B^+ \to J/\psi K^+\pi^+\pi^-$ up	919153	74272	0.15084 ± 0.00054
2012	$B^+ \to J/\psi K^+\pi^+\pi^-$ down	929547	75274	0.15047 ± 0.00052

Table 5.2: Cuts applied during stripping on the `Bu2KpipiMM` stripping line [18].

Particle or event	Variable	Cut
Event	n_{SPD}	< 600
B	$\lvert m(K\pi\pi\mu\mu) - m_{\text{PDG}}\rvert$	$< 1500\text{MeV}/c^2$
	χ^2_{vtx} / dof	< 9
	χ^2_{FD} wrt. PV	> 100
	χ^2_{IP}wrt. PV	< 25
	DIRA	> 0.9995
K	$\chi^2_{\text{trk}}/n_{\text{dof}}$	< 3
	P_{ghost}	< 0.4
	p_T	$> 250\text{MeV}/c$
	χ^2_{IP}	> 4
π	$\chi^2_{\text{trk}}/n_{\text{dof}}$	< 3
	P_{ghost}	< 0.4
	p_T	$> 250\text{MeV}/c$
	χ^2_{IP}	> 4
K_1	$m(K\pi\pi)$	$> 0\text{MeV}/c^2$
	$m(K\pi\pi)$	$< 6000\text{MeV}/c^2$
	$\sum_{K,\pi,\pi} p_T$	$800\text{MeV}/c$
	χ^2_{vtx}	< 12
	$\sum_{K,\pi,\pi} \chi^2_{\text{IP}}$	> 48
μ	p_T	$> 300\text{MeV}/c$
	χ^2_{IP}	> 9
	ISMUON	True
	HASMUON	True
Dimuon	p_T	$> 0\text{MeV}/c$
	m	$< 5500\text{MeV}/c^2$
	χ^2_{vtx} / dof	< 9
	χ^2_{FD} wrt. PV	> 16
	χ^2_{IP}wrt. PV	> 0

Table 5.3: Charmonium veto regions defined through the q^2 of the dimuon system. $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ event candidates must not lie within these regions.

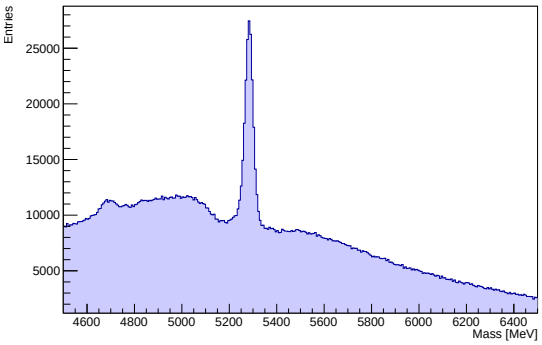
Veto region	q^2 [GeV^2/c^4]
J/ψ	$[8.0, 11.0]$
$\psi(2S)$	$[12.5, 15.0]$

Table 5.4: Mass requirements for $B^+ \to J/\psi K^+\pi^+\pi^-$ and $B^+ \to \psi(2S)K^+$ channel.

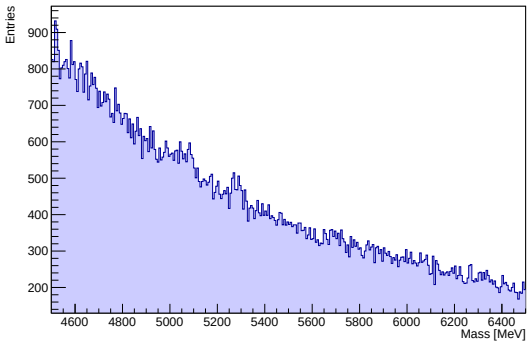
Decay channel	Mass requirement
$B^+ \to J/\psi K^+\pi^+\pi^-$	$3046.9 < m(\mu^+\mu^-) < 3146.9$ MeV/c^2
$B^+ \to \psi(2S)K^+$	$3626.1 < m(\pi^+\pi^-\mu^+\mu^-) < 3746.1$ MeV/c^2

Table 5.5: Overview of the preselection cuts applied to rare channel and normalisation channel.

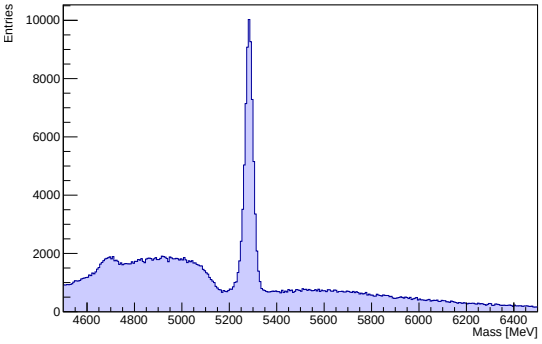
Variable	Cut
$K^\pm$ PIDK	> -5
$p_T(K^\pm)$	> 500 MeV/c
$p_T(\pi^\pm)$	> 500 MeV/c



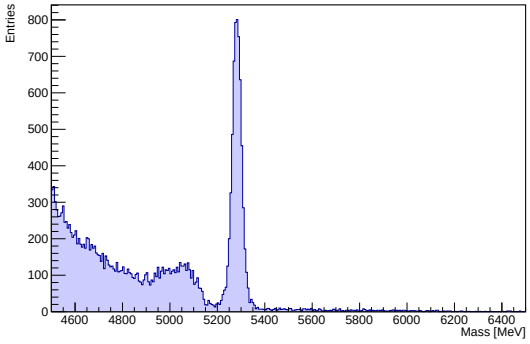
(a) B^+ mass distribution after stripping.



(b) B^+ mass distribution with q^2 cut for rare channel.



(c) B^+ mass distribution with q^2 cut for normalisation channel.



(d) B^+ mass distribution with q^2 cut for control channel.

Figure 5.1: B^+ mass distributions before and after preselection and cuts on q^2.

5.5 Multivariate Analysis

To separate background and signal events more efficiently, the multivariate analysis (MVA) method known as the Boosted Decision Tree (BDT) model is applied. This section documents the training of the BDT, input variables and optimising the output using the ROOT library TMVA [19].

5.5.1 Introduction to Boosted Decision Trees

A single decision tree is a binary decision structure used to classify events either as signal or background. Repeated either/or decisions are performed on a variable from a set training sample in one node until a stop criterion is fulfilled, such that the space of variables is split into regions that can be eventually classified as signal or background. However, a single tree is prone to learning statistical fluctuations, which is why it is beneficial to use a multitude of trees, referred to as forest. These trees are also derived from the training sample, but the event weights are subsequently subjected to *boosting*, a process where misclassified events are assigned increased weights before being passed on to the next tree. For training the BDT, an adaptive boost algorithm (AdaBoost) is implemented. Misclassified events from the training sample are repeatedly assigned an increased boost weight α determined by the previous tree. α is derived from the misclassification rate, err, via

$$\alpha = \frac{1 - \text{err}}{\text{err}}. \tag{5.4}$$

Defining the result of a single classifier as $h(\mathbf{x})$ with $\mathbf{x}$ being the tuple of input variables as $h(\mathbf{x}) = +1$ for signal and $h(\mathbf{x}) = -1$ for background, the output value y_{Boost} is given by

$$y_{\text{Boost}} = \frac{1}{N_{\text{Trees}}} \cdot \sum_{i}^{N_{\text{Trees}}} \ln(\alpha_i) \cdot h_i(\mathbf{x}). \tag{5.5}$$

5.5.2 Creating a Training Sample

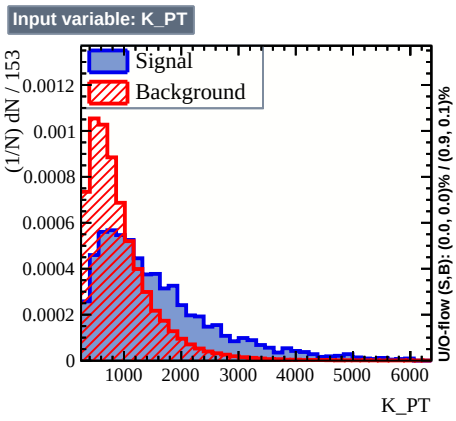
The BDT used for suppressing the combinatorial background is trained on pure signal and background samples. The signal sample is extracted from MC simulations where the normalisation channel serves as a proxy for the rare decay channel. Furthermore, only truthmatched events are taken into account.
The background sample is extracted from the high B mass sideband where $m(B^+) > 5600$ MeV/c^2. The lower mass sideband is not included since it contains partially reconstructed events.

5.5.3 Training and Testing the BDT Classifier

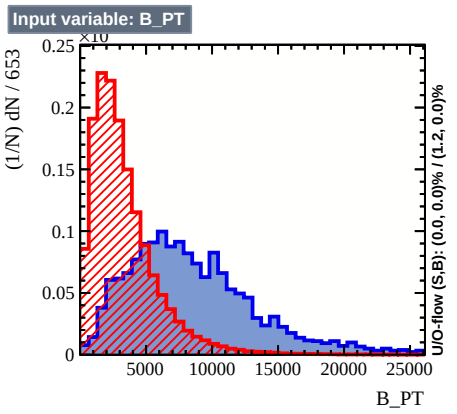
The BDT is trained on several geometric and kinematic variables as well as the fit quality of the B decay vertex and its displacement. The variables are chosen based on their background-discriminating power, i.e. the signal and background distributions must exhibit a significant difference in shape. Distributions of several variables used are depicted in Figure 5.2. As expected, the p_T of kaons and B mesons tend to be higher for signal events (see Figures 5.2a and 5.2b), which is characteristic for heavy particle decays. Since B mesons are produced at the primary vertex, their impact parameter and the corresponding χ^2 are expected to be small for signal candidates, as illustrated in Figure 5.2c. The kaons, however, must come from a detached vertex and thus have a higher IP with regard to the PV (see Figure 5.2d). The entirety of variables used to train the BDT is listed in Table 5.6. Remaining distributions can be found in Appendix A.
After training, the BDT is applied to a testing sample in order to check for overtraining (see Figure 5.3a), which occurs when the algorithm learns statistical fluctuations from the training sample and is unable to adapt to new data. A measure for the performance is provided by the Receiver Operations Characteristics (ROC) curve by plotting the background rejection (1 - background efficiency) against signal efficiency. The area under the curve (AUC) then describes how well the classifier separates signal and background. An AUC score of 1 indicates a perfect performance, whereas a score of 0.5 would correspond to random guessing [20]. The BDT used for this analysis indeed seems to perform quite well with an AUC score of 0.996, as can be seen in Figure 5.3b.

Table 5.6: Variables used to train the BDT.

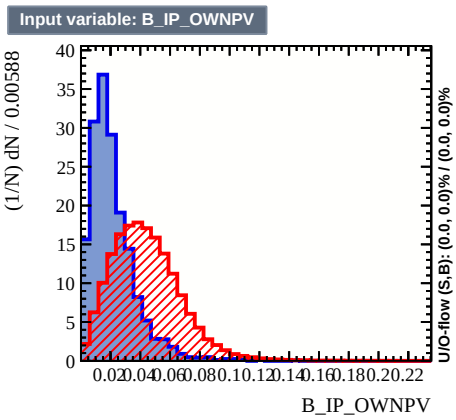
IP	p_T	PID(μ)	DIRA	χ^2(Endvertex)	χ^2(FD)
B^+	B^+		B^+	B^+	B^+
K^+	K^+				
π^+	π^+				
π^-	π^-				
μ^+		μ^+			
μ^-		μ^-			



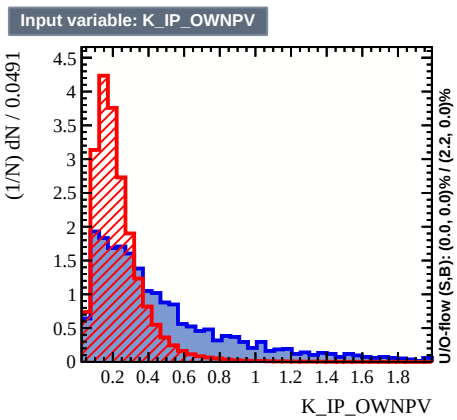
(a) Transverse momentum of the kaon.



(b) Transverse momentum of the B meson.

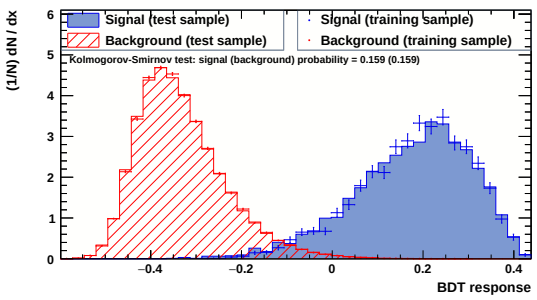


(c) Impact parameter of the B meson.

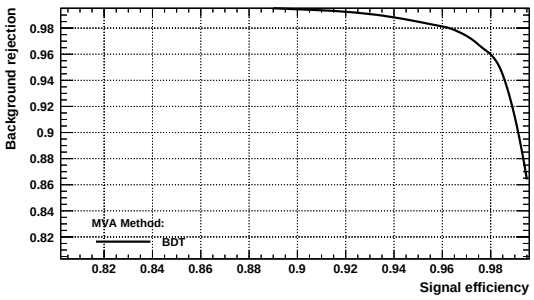


(d) Impact parameter of the kaon.

Figure 5.2: Distributions of signal samples (blue) and background samples (red) for different variables used to train the BDT.



(a) BDT output for training and testing samples.



(b) ROC curve.

Figure 5.3: BDT classifier output and ROC curve.

5.5.4 Optimising the BDT Output Cut

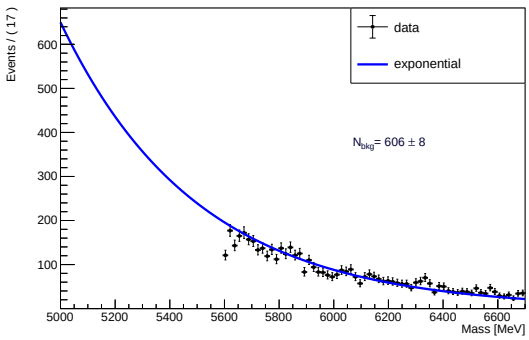
In order to maximise background rejection and minimise signal loss, the optimal cut on the BDT response must be determined. This is done by maximising the Figure of Merit (FoM)

$$\text{FoM} = \frac{S}{\sqrt{S+B}}, \tag{5.6}$$

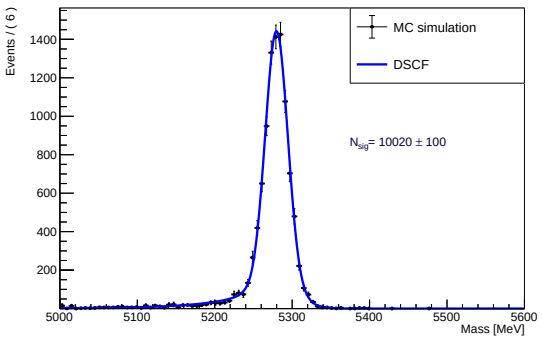
where S and B denote the expected signal and background yields, respectively. The background yield is estimated by fitting an exponential to the high B mass sideband and extrapolating the yield into the signal region $m(B) \pm 50$ MeV (see Figure 5.4a).
The expected signal yield is extracted by scaling the yield of the normalisation channel by the ratio of selection efficiencies and the branching fraction ratio

$$S_{rare} = S_{norm} \cdot \frac{\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-)}{\mathcal{B}(B \to J/\psi(\to \mu^+\mu^-)K^+\pi^+\pi^-)} \cdot \frac{\epsilon_{sel}^{rare}}{\epsilon_{sel}^{norm}}. \tag{5.7}$$

The selection efficiencies are determined through Monte Carlo simulations as well as the signal yield for the normalisation channel, where a double Crystal Ball (CB) function is fitted to the MC simulation via an unbinned maximum likelihood fit. This fit is shown in Figure 5.4b.
The FoM may now be described as a function of the BDT response as depicted in Figure 5.5. The local maximum of this distribution is at -0.08, thus corresponding to the optimal BDT cut value.



(a) Exponential fit to the high B mass sideband.



(b) Double CB fit to the MC simulation for the normalisation channel.

Figure 5.4: Fits to obtain the expected signal and background yields in order to optimise the BDT output cut.

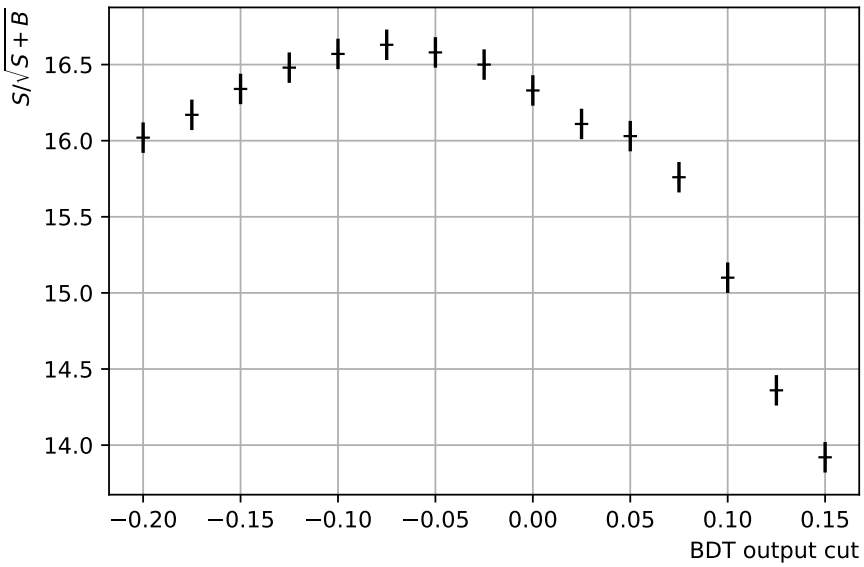


Figure 5.5: FoM as a function of the BDT output cut.

6 Reweighting

To correct for acceptance effects, a BDT reweighter [21] is used. This is done by weighting each event by the inverse selection and reconstruction efficiency. The reweighter is trained on generator level and reconstructed and selected MC samples and then applied to the data that have passed the BDT classifier as signal candidates. This section outlines the machine learning algorithm, testing on a toy model and its application to real data.

6.1 Introduction to BDT Reweighters

Usually, reweighting is applied to Monte Carlo simulations to minimise differences between real data (RD) and simulations stemming from inaccuracies of the simulation, such that the MC distribution (original) coincides with real data (target). The easiest approach to reweighting is splitting the space of variables into bins and multiplying the MC events in a given bin by

$$\text{multiplier}_{\text{bin}} = \frac{w_{\text{bin, RD}}}{w_{\text{bin, MC}}}, \tag{6.1}$$

$w_{\text{bin, RD}}$ and $w_{\text{bin, MC}}$ being the total weights of the real data and MC distributions. However, this method proves to be flawed, especially when reweighting multidimensional distributions, since the amount of data needed to reliably estimate density functions grows exponentially with an increasing number of variables.
One solution to this problem is offered by a BDT reweighter. The space of variables is split into regions by the BDT in accordance to the given problem rather than each variable being simply divided into several bins. Those regions are determined by maximising the metric χ^2.

$$\chi^2 = \sum_{\text{leaf}} \frac{(w_{\text{leaf, MC}} - w_{\text{leaf, RD}})^2}{w_{\text{leaf, MC}} + w_{\text{leaf, RD}}}. \tag{6.2}$$

$w_{\text{leaf, MC}}$ denotes the amount of MC events within one region (leaf), $w_{\text{leaf, RD}}$ the amount of RD events. Regions of high importance for the reweighting process are characterised by their high corresponding summands in χ^2, for example when $w_{\text{leaf, MC}}$ is much higher than $w_{\text{leaf, RD}}$.
Analogous to the previously described BDT classifier (see Section 5.5.1), multiple trees are trained successively, each one correcting for discrepancies found in previous trees.

6.2 Implementation of BDT Reweighter

In contrast to the example above, the aim of using the reweighter in this thesis is reweighting real data events, assigning each of them a weight w_i, corresponding to the per-event efficiency ϵ_i^{-1}. The efficiency description is obtained by reweighting generated MC distributions to match the reconstructed and selected ones, resulting in efficiency distributions in dependence of given variables. Based on those variables, the algorithm then assigns event weights for real data events.

6.2.1 Testing BDT Reweighter on Toy Model

Before the reweighter is applied to the multi-dimensional distribution, it was tested on a toy sample reduced to one dimension, such that the structure of the BDT decisions could be made visible. The variable $m(K^+\pi^+\pi^-)$ was chosen for this. There are to be more events in the generator level sample in the lower mass regions, and less in the higher mass region (see Figure 6.1). Therefore, the weights in the lower mass regions are expected to be < 1, while they should be > 1 for higher invariant masses.
This hypothesis was tested for various numbers of decision trees, each with a depth of one, such that each BDT cut is visualised through the change in predicted weights, as illustrated in Figure 6.2: One tree implies one cut, i.e. two plateaus in the $m(K^+\pi^+\pi^-)$ vs. weight distribution. Five trees produce six plateaus, etc. Indeed, the reweighter predicts the weights for different mass regions according to expectations.

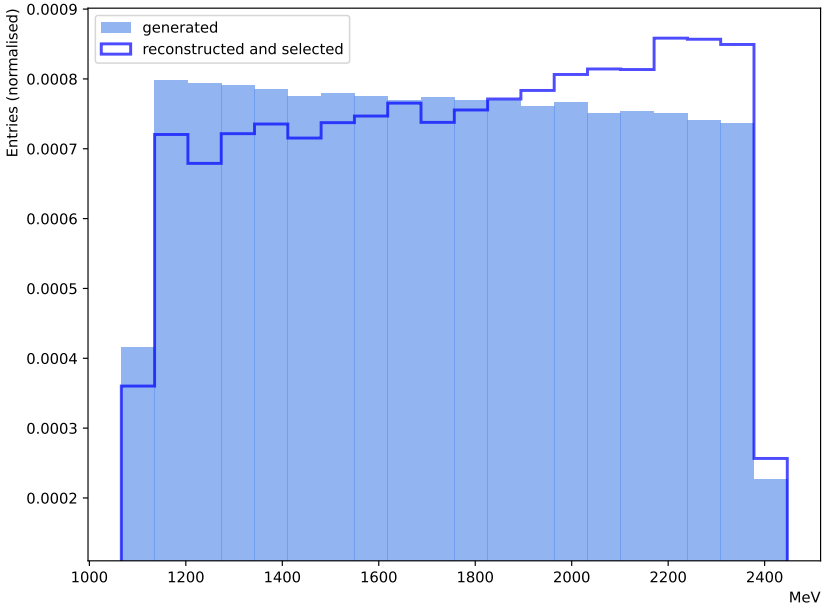


Figure 6.1: $m(K^+\pi^+\pi^-)$ distributions on generator and reconstructed and selected level.

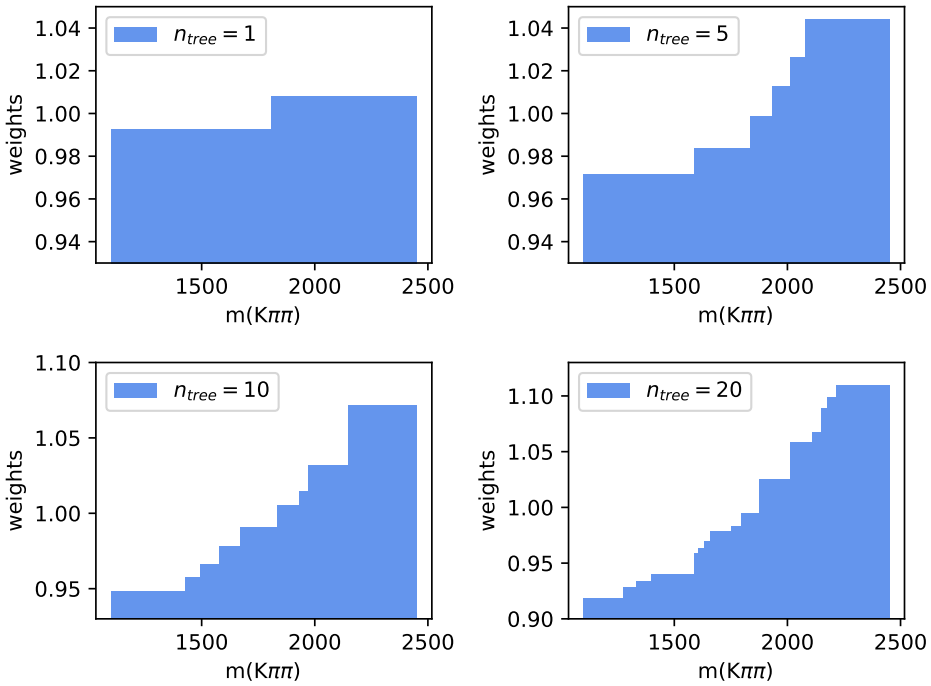


Figure 6.2: Weights predicted by BDT reweighter for $m(K^+\pi^+\pi^-)$ toy sample for different numbers of trees in a forest.

6.2.2 Training BDT Reweighter on Multidimensional Distributions

The BDT reweighter is trained on seven-dimensional $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ MC simulated distributions, the original distribution being on generator-level, the target being reconstructed and selected level. The efficiency is parametrised as a function of $m^2(\pi\pi)$, $m^2(K\pi)$, $m(K\pi\pi)$, q^2 and the cosines of decay angles $\cos\theta_L$, $\cos\theta_K$ and $\cos\theta_V$. The decay angles are illustrated in Figure 6.3. Both distributions used for training are shown in Figure 6.4, as well as the reweighted generated distribution. While the distributions already coincide well in $m^2(\pi\pi)$ and $m^2(K\pi)$, implying an even selection efficiency, they differ significantly in shape for the decay angles, $m(K\pi\pi)$ and q^2. The acceptance seems to be quite low for small decay angles for example. The reweighter accounts for that difference by assigning those events a smaller weight, such that a good agreement between the generated reweighted distribution and the reconstructed and selected one is observed. The hyperparameters set for training the reweighter on those distributions are listed in Table 6.1.

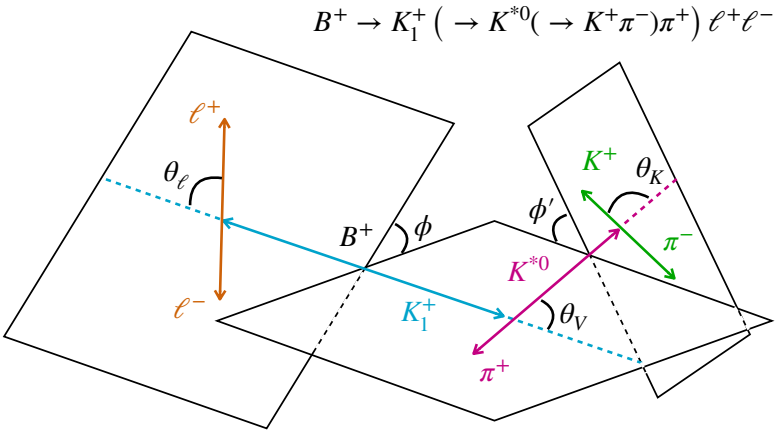


Figure 6.3: Decay angles for the rare decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$.

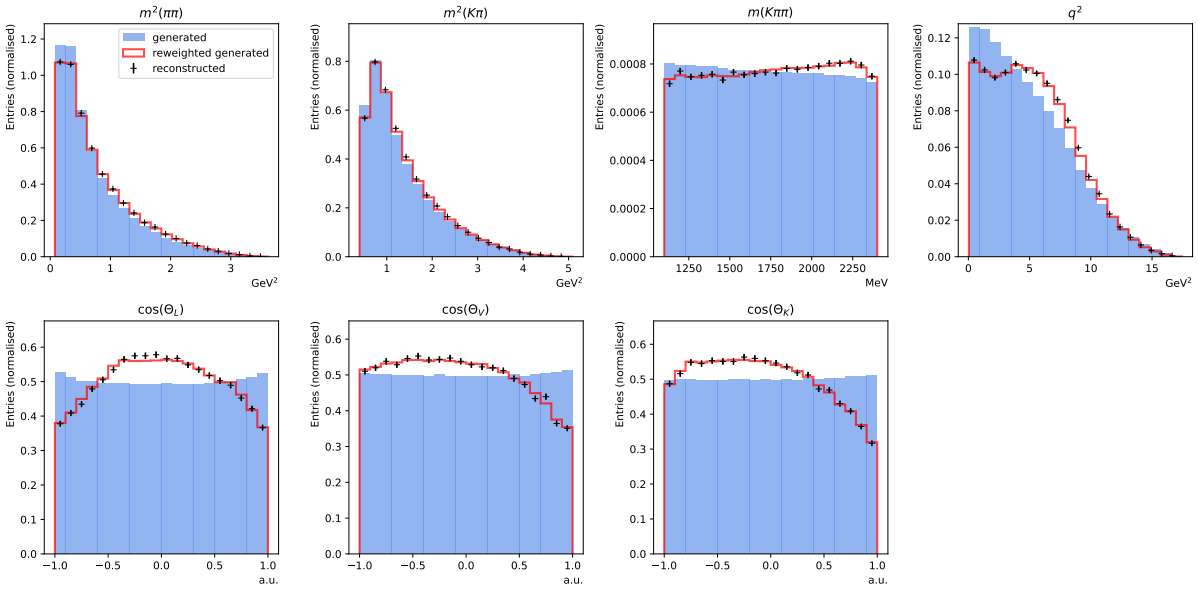


Figure 6.4: Distributions of generator-level, both unweighted and reweighted, and reconstructed and selected variables for $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ simulations. In order to compare the generated distribution to the reweighted generated one, all distributions are normalised.

Table 6.1: Parameters used for the BDT reweighter.

Parameter	Explanation	Setting
`n_estimators`	Number of trees in a forest. Large values increase stability of reweighting rule, but increase risk of overfitting.	200
`learning_rate`	Step size shrinkage. Lesser learning rate requires more trees but makes reweighting rule more stable.	0.1
`max_depth`	Maximal depth of trees. Large values increase complexity of the model, but also risk of overfitting.	3
`min_samples_leaf`	Minimal number of events remaining in a leaf. Large values decrease risk of overfitting.	400
`subsample`	Fraction of data that is randomly chosen to train one tree.	0.4

6.2.3 Cross-Checking Efficiencies Determined by BDT Reweighter in $B^+ \to J/\psi K^+\pi^+\pi^-$ channel

Since events in the normalisation channel are also reweighted by the reweighter trained on rare channel samples, the efficiencies determined by the reweighter have to be cross-checked with the actual reconstruction and selection efficiency. This is done by reweighting the MC generator-level distribution. The number of reweighted generated events, which is given by the sum of the weights

$$N_{\text{reweighted generated}} = \sum_{i=1}^{N_{\text{generated}}} w_i, \tag{6.3}$$

should approximately equal the number of reconstructed and selected events. This yields $N_{\text{reweighted generated}} = 78585$, which deviates by 1.8% from the number of reconstructed and selected events $N_{\text{reconstructed\&selected}} = 79997$.

The reweighted generated distribution as well as the reconstructed and selected one is shown in Figure 6.5. The q^2 distribution is not shown since q^2 is a sharp peak on generator level because of sharp J/ψ resonance, but a Gaussian on reconstructed and selected level due to limited detector resolution. Since the reweighter is trained on $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ simulations, it does not account for that effect.

The $\cos\theta_L$ distribution shows that porting to reweighting from the rare mode seems to be not perfect, nevertheless feasible to estimate the efficiency for the normalisation mode.

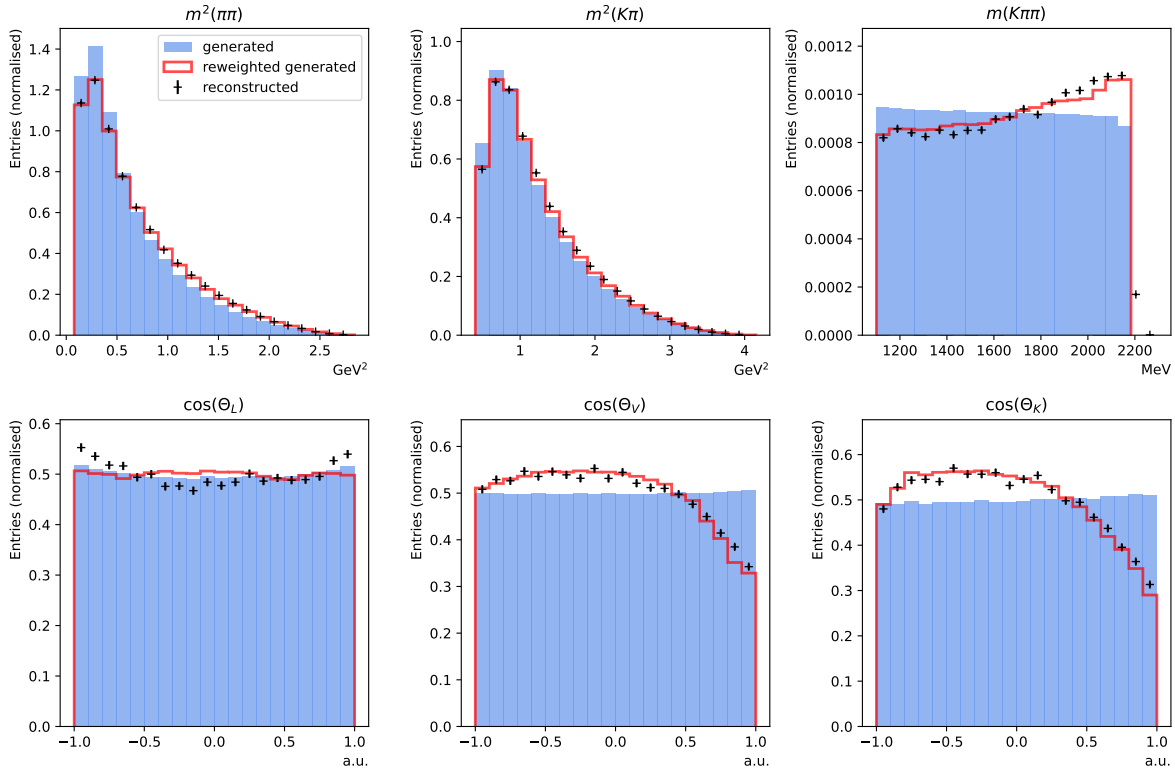


Figure 6.5: Distributions of generator-level, both unweighted and reweighted, and reconstructed and selected variables for $B^+ \to J/\psi K^+\pi^+\pi^-$ simulations. In order to compare the generated distribution to the reweighted generated one, all distributions are normalised.

6.2.4 Applying BDT Reweighter to Real Data

The BDT reweighter is now applied to real data events that have passed the multivariate selection and were classified as signal events. Both rare channel and normalisation channel data are reweighted, meaning that each event is assigned a per-event efficiency and thus weighted accordingly. The distributions of variables chosen to parametrise the efficiencies are shown in Figures 6.6 and 6.7. The corresponding weight distributions can be found in Appendix B.

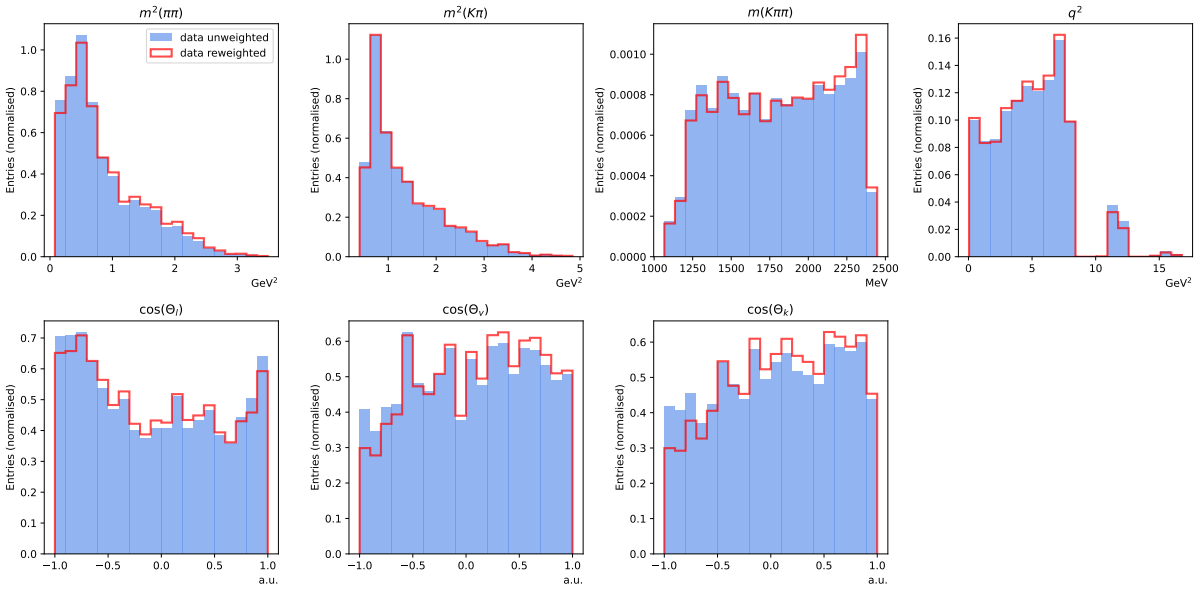


Figure 6.6: Normalised distributions of variables for efficiency parametrisation for rare channel real data, unweighted and reweighted.

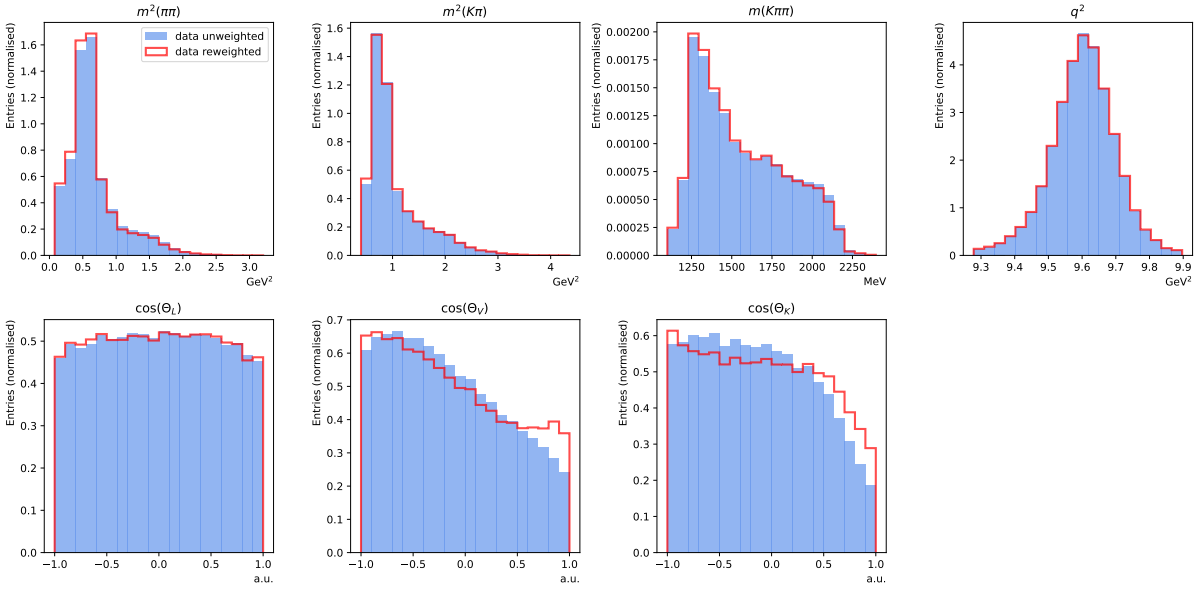


Figure 6.7: Normalised distributions of variables for efficiency parametrisation for normalisation channel real data, unweighted and reweighted.

7 Signal Yields

The signal yields for rare channel and normalisation channel are obtained by performing weighted unbinned extended maximum likelihood fits. This section outlines this fit method as well as the used fit models and correct error determinations.

7.1 Extended Maximum Likelihood Method

Considering N measurements $x_1, \ldots, x_N$ and a known probability density function (PDF) $p(x_i; \boldsymbol{a})$, where $\boldsymbol{a} = \{a_1, \ldots, a_m\}$ is a set of unknown parameters, the likelihood function $\mathcal{L}$ is given by

$$\mathcal{L} = \prod_{i=1}^{N} p(x_i; \boldsymbol{a}). \tag{7.1}$$

In contrast to the standard method of maximum likelihood, where the PDF is normalised to unity, the extended maximum likelihood method does not require that constraint on the normalisation. The likelihood function becomes

$$\mathcal{L} = \frac{\mathcal{N}^N e^{-\mathcal{N}}}{N!} \prod_{i=1}^{N} p(x_i; \boldsymbol{a}), \tag{7.2}$$

where $\mathcal{N} = \mathcal{N}(\boldsymbol{a})$ denotes the predicted number of events and N the observed number of events. The optimal parameters $\hat{\boldsymbol{a}}$ are estimated by maximising the likelihood. In practice, a modified log-likelihood function[2] is used to avoid numerical problems:

$$\ln \tilde{\mathcal{L}} = \sum_{i=1}^{N} \ln p(x_i; \boldsymbol{a}) - \mathcal{N}(\boldsymbol{a}). \tag{7.3}$$

The maximisation as well as uncertainty calculations are typically performed numerically using software packages like `Minuit` [22].
However, when considering per-event weights as it is done in this thesis, the expression for the logarithmic likelihood becomes

$$\ln \mathcal{L} = \sum_{i}^{\hat{N}} w_i \ln p(x_i; \boldsymbol{a}) + \hat{N}_{\text{tot}} \ln N_{\text{tot}} - N_{\text{tot}}, \quad \hat{N}_{\text{tot}} = \sum_{i}^{\hat{N}} w_i, \tag{7.4}$$

[2] The log-likelihood function is $\ln \mathcal{L} = \sum_{i=1}^{N} \ln p(x_i; \boldsymbol{a}) - \mathcal{N}(\boldsymbol{a}) - \ln(N!)$, but since $\ln(N!)$ does not depend on the parameters, $\ln \tilde{\mathcal{L}} = \ln \mathcal{L} + \ln(N!)$ is maximised.

N being the number of detected events, N_{tot} the weighted number of events. Furthermore, parameter uncertainties cannot simply be determined by inverting the Hessian matrix of the negative log-likelihood anymore since this approach does not yield asymptotically correct errors. The asymptotically correct approach for weighted maximum likelihood fits is derived in [23].

7.2 Signal Fit to Data

All fit models used for the mass fits in this thesis consist of double Crystalball (CB) functions [24] for the signal component with an exponential as the combinatorial background component. The CB is continuous and composed of a Gaussian core and a power-law tail:

$$CB(x;\alpha,n,\mu,\sigma) = \mathcal{N}\cdot\begin{cases}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), & \text{for } \frac{x-\mu}{\sigma} > -\alpha\\ A\cdot\left(B-\frac{x-\mu}{\sigma}\right)^{-n}, & \text{for } \frac{x-\mu}{\sigma} \leq -\alpha\end{cases} \tag{7.5}$$

with

$$\begin{aligned} A &= \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right),\\ B &= \frac{n}{|\alpha|} - |\alpha|. \end{aligned} \tag{7.6}$$

$\mathcal{N}$ denotes the normalisation factor. μ and σ are the mean and resolution of the Gaussian core distribution. The parameter α marks the transition of the Gaussian into the power-law tail and n is the exponent of said tail.
To account for differences between the left and right side of the invariant mass spectrum, a linear combination of two CBs with a shared mean is used. The fit model for the signal component is thus given by:

$$F_S = f\cdot CB(m;\alpha_1,n_1,\mu,\sigma_1) + (1-f)\cdot CB(m;\alpha_2,n_2,\mu,\sigma_2), \tag{7.7}$$

where f is the fraction of the first CB and m the mass.
The background component is modelled by an exponential:

$$F_B = \frac{1}{\mathcal{N}_{exp}}\cdot\exp(-\lambda m), \tag{7.8}$$

$\mathcal{N}_{exp}$ being the normalisation and λ the slope of the exponential.
The model for the entire mass distribution in the signal range is therefore given by:

$$F = \frac{N_S}{N_S+N_B}\cdot F_S(m;\alpha_1,\alpha_2,n_1,n_1,\mu,\sigma_1,\sigma_2) + \frac{N_B}{N_S+N_B}\cdot F_B(m;\tau), \tag{7.9}$$

where N_S and N_B denote the signal and background yields, respectively. The signal range is constrained through the B mass $m(B^+) \in [5170; 5600]$ MeV/c^2 to reject background from partially reconstructed decays accumulating in the lower mass range. While all parameters are allowed to vary in the fit for the $B^+ \to J/\psi K^+\pi^+\pi^-$ channel, mass shape parameters are fixed for the rare channel, using the ones from the normalisation channel, see Table 7.1. The fits for both decay channels are shown in Figures 7.1 and 7.2. The signal yields for both channels are summarised in Table 7.2.

Table 7.1: Fixed fit parameters for rare channel that were previously determined through fit on normalisation channel.

Parameter	Fixed value
α_1	3.94
α_2	1.30
n_1	4.36
n_2	2.23
σ_1	20.64
σ_2	13.96

Table 7.2: Signal yields of fits to $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to J/\psi K^+\pi^+\pi^-$.

Channel	N_S
$B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$	16425 $\pm$ 1358
$B^+ \to J/\psi K^+\pi^+\pi^-$	2042318 $\pm$ 7376

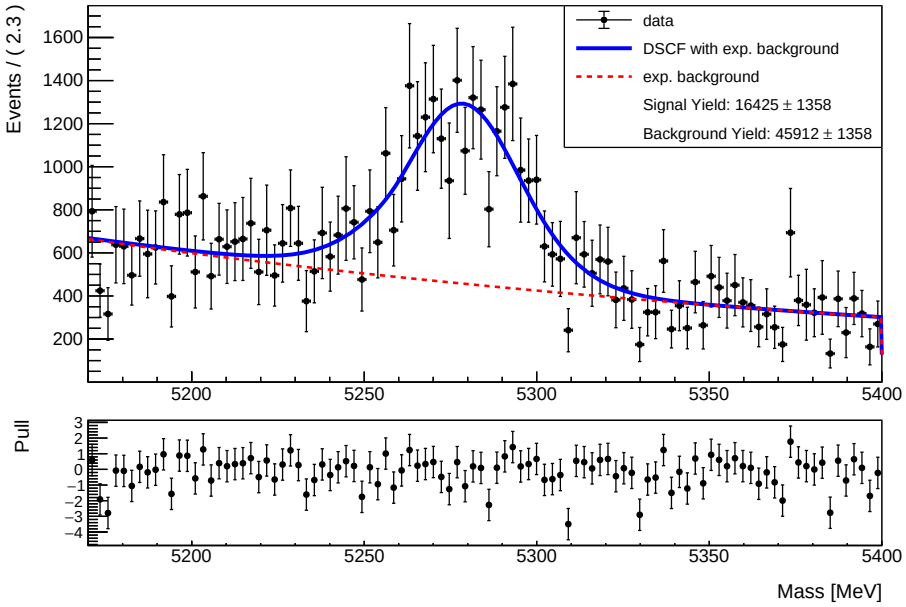


Figure 7.1: Weighted mass fit to rare channel using a double CB as signal component and exponential as background.

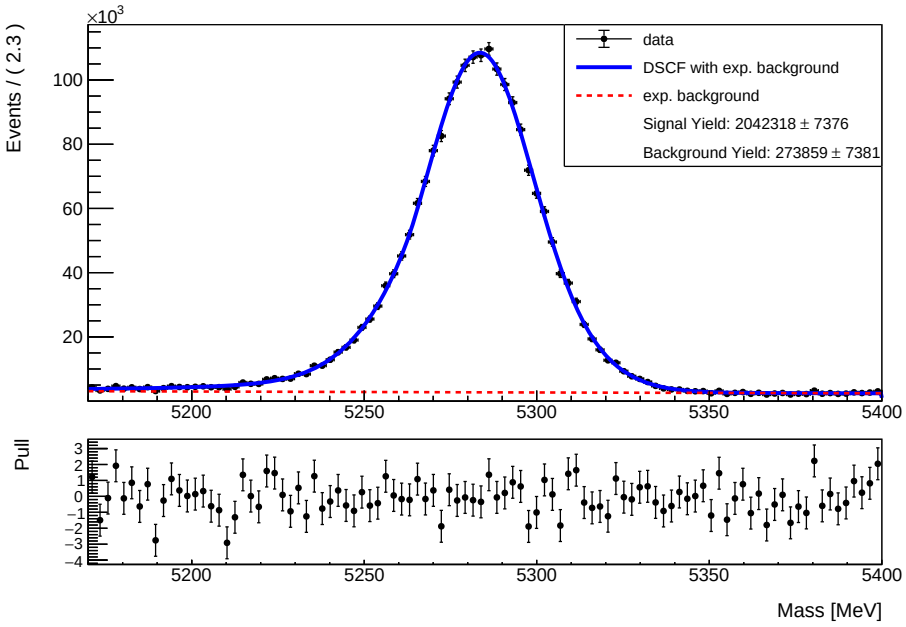


Figure 7.2: Weighted mass fit to normalisation channel using a double CB as signal component and exponential as background.

8 Determination of the Branching Fraction

This section describes the calculation of the branching fraction of the rare decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ relative to the normalisation channel $B^+ \to J/\psi K^+\pi^+\pi^-$, using the signal yields from Section 7.2.

8.1 Branching Fraction Results

Having implicitly included the selection efficiencies by weighting the events and known geometrical efficiencies, the relative branching fraction can be calculated as follows:

$$\frac{\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-)}{\mathcal{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^+\pi^+\pi^-)} = \frac{N_{K\pi\pi\mu\mu}}{N_{J/\psi K\pi\pi}} \cdot \frac{\epsilon^{J/\psi K\pi\pi}_{DecProd}}{\epsilon^{K\pi\pi\mu\mu}_{DecProd}}. \tag{8.1}$$

Using the weighted signal yields from Table 7.2, the relative branching fraction is determined as

$$\frac{\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-)}{\mathcal{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^+\pi^+\pi^-)} = (8.13 \pm 0.67\ (\text{stat})) \cdot 10^{-3}.$$

The total branching fraction of the normalisation channel is calculated from multiplying the branching fractions of $B^+ \to J/\psi K^+\pi^+\pi^-$ and $J/\psi \to \mu^+\mu^-$ [5]

$$\mathcal{B}(B^+ \to J/\psi K^+\pi^+\pi^-) = (8.1 \pm 1.3) \cdot 10^{-4}$$
$$\mathcal{B}(J/\psi \to \mu^+\mu^-) = (5.971 \pm 0.033) \cdot 10^{-2}.$$

The branching fraction of the rare decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ with the charmonium veto regions removed is determined to be

$$\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-) = (3.93 \pm 0.33(\text{stat})) \cdot 10^{-7}.$$

Extrapolating the current signal yield into the q^2 bins of the previous analysis[3] [8] yields a branching fraction of

$$\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-) = (4.25 \pm 0.35(\text{stat})) \cdot 10^{-7}.$$

The fraction of events removed by the charmonium veto cuts is determined from MC simulations of the $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ decay and yields 15.9%. Taking this correction

[3] The q^2 bins of the previous analysis are set to $[0.10, 8.68]$ GeV$^2/c^4$, $[10.09, 12.86]$ GeV$^2/c^4$, $[14.18, 19.00]$ GeV$^2/c^4$.

into account, the total branching fraction of the rare decay is found to be

$$\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-) = (4.55 \pm 0.38(\text{stat})) \cdot 10^{-7}.$$

The samples used to determine the branching fraction in this analysis and the previous one are not statistically independent, but the exact statistical overlap is not known. Therefore, the branching fractions cannot simply be compared through their σ-deviation from one another. One major difference between this analysis and the previous one is the selection of the normalisation channel: The previous analysis used $B^+ \to \psi(2S)K^+$ as the normalisation channel. This decay channel was chosen since its branching fraction has a relative uncertainty of 3%.
In this analysis, $B^+ \to J/\psi K^+\pi^+\pi^-$ is used as the normalisation channel because the efficiencies unfolded by the reweighter could be cross-checked with MC simulations. This is not the case for the $B^+ \to \psi(2S)K^+$ channel since simulations for that decay do not exist, thus one could not verify the efficiencies predicted by the reweighter.

8.2 Systematic Uncertainties

While a complete investigation of systematic uncertainties is out of the scope of this thesis, some suspected dominant sources are discussed briefly.
The most prominent uncertainty stems from the uncertainty of the branching fraction of the normalisation channel, having a relative error of 16%. Taking this error into account yields an uncertainty on the rare decay branching fraction of $0.73 \cdot 10^{-7}$, which is much higher than the previously measured one ($0.18 \cdot 10^{-7}$). However, this is to be expected using the $B^+ \to J/\psi K^+\pi^+\pi^-$ decay as a normalisation channel. Since the previous analysis uses the $B^+ \to \psi(2S)K^+$ mode as the normalisation channel, which has a much smaller relative uncertainty, the systematic uncertainty stemming from the normalisation is considerably smaller.
Other quantifiable sources of uncertainties are the statistical uncertainties of geometrical acceptances $\epsilon_{DecProd}$, which have been determined through Monte Carlo simulations for both relevant decay channels. Another systematic uncertainty emerges from the deviation between the number of reweighted generated events and the number of reconstructed and selected events for the normalisation channel, as described in Section 6.2.3. The contribution from those systematic uncertainties is determined to be $0.11 \cdot 10^{-7}$.
An additional source of uncertainties in this analysis that is not included is the model assumption for the fits of the B mass distributions. The uncertainty arising from the fit model could be estimated by using other fit models, such as a Student's t-distribution for

the signal component, since its shape resembles the double Crystalball, and a polynomial for the background instead of an exponential.
Major systematic uncertainties that are not quantified in this thesis originate from the efficiency calculations on PHSP Monte Carlo simulations, i.e. mismodeling in simulations. The fraction of events removed by the charmonium vetoes, which is determined on those PHSP simulations, has an uncertainty that is also not taken into account. This yields an underestimation of the total systematic uncertainties.

9 Conclusions

A measurement of the branching fraction of the rare decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ relative to the tree-level decay $B^+ \to J/\psi K^+\pi^+\pi^-$ is presented in this thesis. The analysis is carried out on Run 1 data recorded by the LHCb experiment at $\sqrt{s} = 7$ TeV in 2011 and $\sqrt{s} = 8$ TeV in 2012, corresponding to an integrated luminosity of 3 fb^{-1}.
Signal candidates for the rare decay are selected by applying a series of loose preselection cuts to remove combinatorial background. To isolate the signal peak, combinatorial background has to be further reduced, which is done by performing a multivariate analysis. Event candidates for the normalisation channel are chosen analogously.
To account for acceptance effects and unfold the selection and reconstruction efficiency, both decay channel mass distributions are reweighted by applying a machine learning algorithm trained on Monte Carlo simulations on generator and reconstructed and selected level. Signal yields are obtained by performing extended maximum likelihood fits on the unbinned, reweighted data. 16425 ± 1358 $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ events and 2042318 ± 7376 $B^+ \to J/\psi K^+\pi^+\pi^-$ events are reconstructed, yielding a relative branching fraction of

$$\frac{\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-)}{\mathcal{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^+\pi^+\pi^-)} = (8.13 \pm 0.67\ (\text{stat})) \cdot 10^{-3}.$$

The total branching fraction is determined to be

$$\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-) = (4.55 \pm 0.38(\text{stat}) \pm 0.11(\text{syst}) \pm 0.73(\text{norm})) \cdot 10^{-7}.$$

The total branching fraction determined in the previous analysis [8] by the LHCb Collaboration in 2014, also using 3 fb^{-1} of data, is

$$\mathcal{B}_{lit}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-) = (4.36^{+0.29}_{-0.27}(\text{stat}) \pm 0.21(\text{syst}) \pm 0.18(\text{norm})) \cdot 10^{-7}.$$

The statistic uncertainty as well as the uncertainty arising from the measurement of the branching fraction of the normalisation channel are higher in this analysis, the latter stemming from the choice of the current normalisation channel, which has a relative uncertainty of 16%. This uncertainty could be reduced by using the $B^+ \to \psi(2S)K^+$ as the normalisation channel, which has a much lower uncertainty than the currently chosen normalisation channel. However, this decay channel was not chosen in this thesis since MC simulations for that channel are not available, hence the efficiencies determined by the reweighter could not be cross-checked with the actual reconstruction and selection

efficiency.
The systematic uncertainty only takes the statistical uncertainty of the geometrical efficiencies and the deviation between the number of reweighted generated events and the number of reconstructed and selected events into account, therefore poses an underestimation of the total systematic uncertainties.

List of Tables

List of Figures

Bibliography

[1] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Physical Review Letters* 13.16 (Oct. 19, 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[2] The ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 1–29. arXiv: 1207.7214[hep-ex]. (Visited on 03/19/2024).

[3] M. Thomson. *Modern Particle Physics*. Cambridge University Press, 2013. DOI: 10.1017/CBO9781139525367.

[4] D. Griffiths. *Introduction to elementary particles*. 2nd ed. Wiley-VCH, 2008. ISBN: 978-3-527-40601-2.

[5] R. L. Workman et al. "Review of Particle Physics". In: *Prog. Theor. Exp. Phys.* 2022 (8 2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[6] L. Wolfenstein. "Parametrization of the Kobayashi-Maskawa Matrix". In: *Phys. Rev. Lett.* 51 (21 Nov. 1983), pp. 1945–1947. DOI: 10.1103/PhysRevLett.51.1945.

[7] R. L. Workman et al. In: *Prog. Theor. Exp. Phys.* 2022 (2022), p. 083C01.

[8] LHCb Collaboration et al. "First observations of the rare decays $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to \phi K^+\mu^+\mu^-$". In: *Journal of High Energy Physics* 2014 (Oct. 2014). arXiv: 1408.1137.

[9] The LHCb Collaboration et al. "The LHCb Detector at the LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[10] C. Elsasser. *$b\bar{b}$ production angle plots*. URL: https://lhcb.web.cern.ch/speakersbureau/html/bb_ProductionAngles.html (visited on 03/24/2024).

[11] LHCb Collaboration et al. *The LHCb Upgrade I*. Tech. rep. Geneva: CERN, May 2023. arXiv: 2305.10515.

[12] R. Aaij et al. "Performance of the LHCb Vertex Locator". In: *Journal of Instrumentation* 9.9 (Sept. 2014), P09007. DOI: 10.1088/1748-0221/9/09/P09007.

[13] R. Arink et al. "Performance of the LHCb Outer Tracker". In: *Journal of Instrumentation* 9.1 (Jan. 2014), P01002. DOI: 10.1088/1748-0221/9/01/P01002.

[14] M. Adinolfi et al. "Performance of the LHCb RICH detector at the LHC". In: *The European Physical Journal C* 73.5 (May 15, 2013). DOI: 10.1140/epjc/s10052-013-2431-9.

[15] A. A. Alves et al. "Performance of the LHCb muon system". In: *Journal of Instrumentation* 8.2 (Feb. 27, 2013), P02022–P02022. DOI: 10.1088/1748-0221/8/02/P02022.

[16] CERN. *Standard set of performance numbers*. URL: https://lhcb.web.cern.ch/speakersbureau/html/PerformanceNumbers.html (visited on 04/05/2024).

[17] R. Aaij et al. "The LHCb trigger and its performance in 2011". In: *Journal of Instrumentation* 8.4 (Apr. 22, 2013), P04022–P04022. DOI: 10.1088/1748-0221/8/04/P04022.

[18] J. Albrecht et al. *Test of lepton flavour universality with* $B^+ \to K^+\pi^+\pi^-\ell^+\ell^-$ *decays*. LHCb-ANA-2022-018. Geneva: CERN, 2023.

[19] A. Hoecker et al. *TMVA - Toolkit for Multivariate Data Analysis*. July 7, 2009. arXiv: physics/0703039.

[20] A. P. Bradley. "The use of the area under the ROC curve in the evaluation of machine learning algorithms". In: *Pattern Recognition* 30.7 (July 1, 1997), pp. 1145–1159. DOI: 10.1016/S0031-3203(96)00142-2.

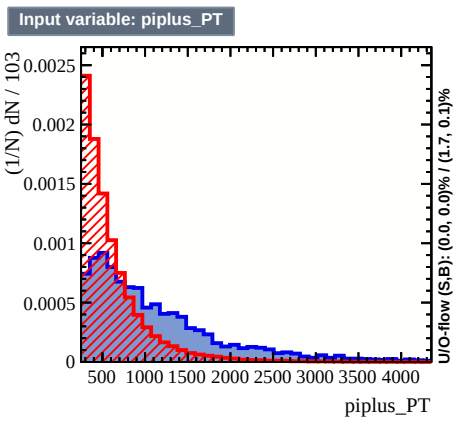
[21] A. Rogozhnikov. "Reweighting with Boosted Decision Trees". In: *Journal of Physics: Conference Series* 762 (Oct. 2016). arXiv: 1608.05806.

[22] F. James et al. "Minuit - a system for function minimization and analysis of the parameter errors and correlations". In: *Computer Physics Communications* 10.6 (Dec. 1, 1975), pp. 343–367. DOI: 10.1016/0010-4655(75)90039-9.

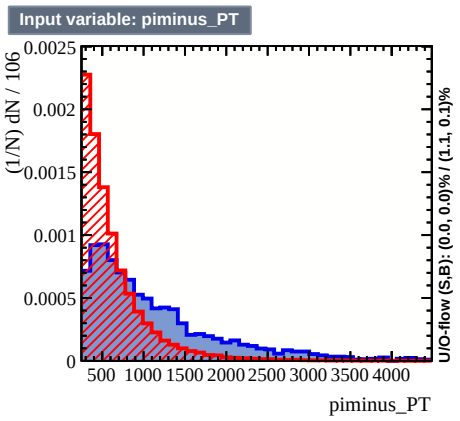
[23] C. Langenbruch. "Parameter uncertainties in weighted unbinned maximum likelihood fits". In: *The European Physical Journal C* 82.5 (May 2022), p. 393. arXiv: 1911.01303.

[24] M. Aaboud et al. "Search for resonances in diphoton events at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Journal of High Energy Physics* 2016.9 (Sept. 1, 2016), p. 1. DOI: 10.1007/JHEP09(2016)001.

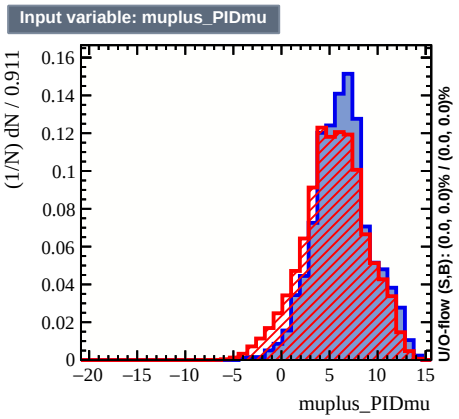
A Plots of BDT Training Variables



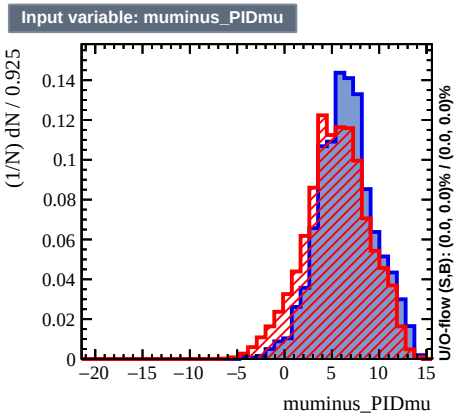
(a) Transverse momentum of the π^+.



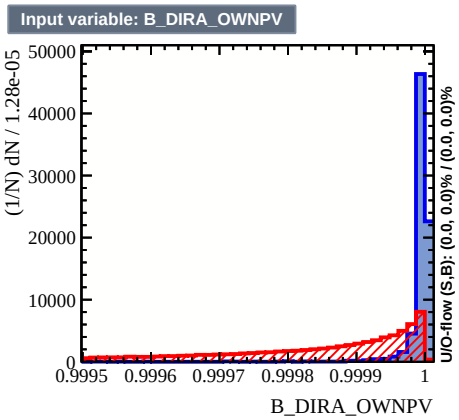
(b) Transverse momentum of the π^-.



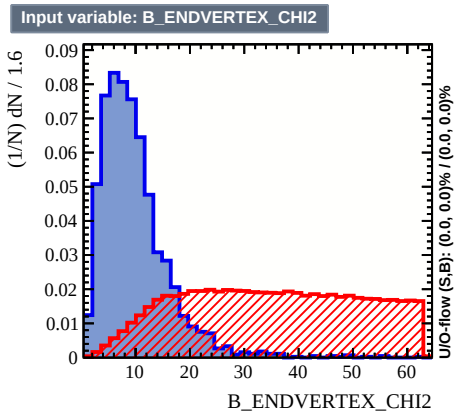
(c) PID of the μ^+.



(d) PID of the μ^-.

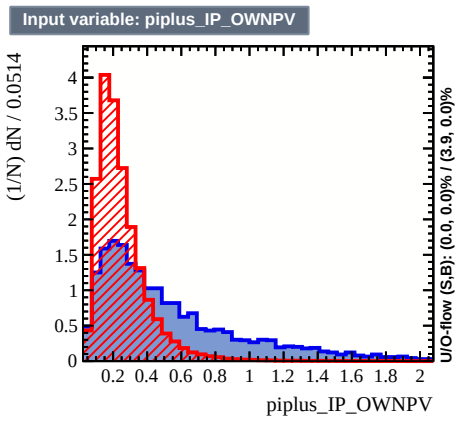


(e) DIRA of the B meson.

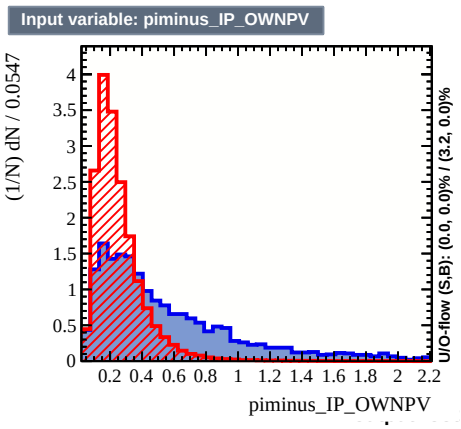


(f) χ^2 of the B meson endvertex.

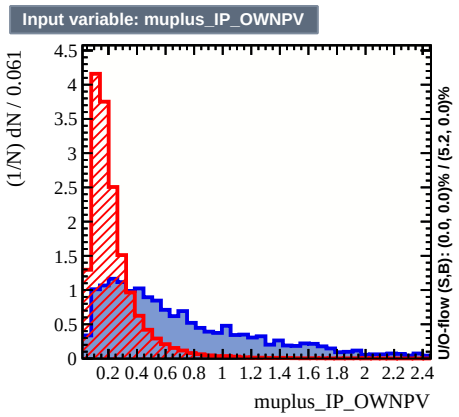
Figure A.1: Distributions of signal samples (blue) and background samples (red) for different variables used to train the BDT (1).



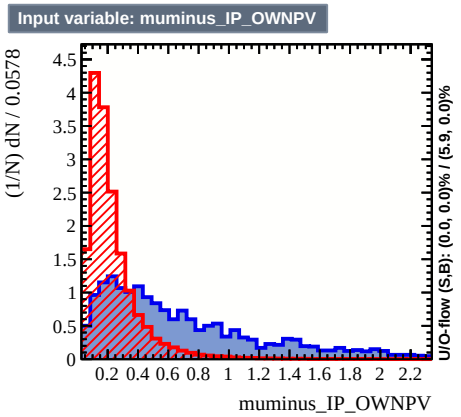
(a) Impact parameter of the π^+.



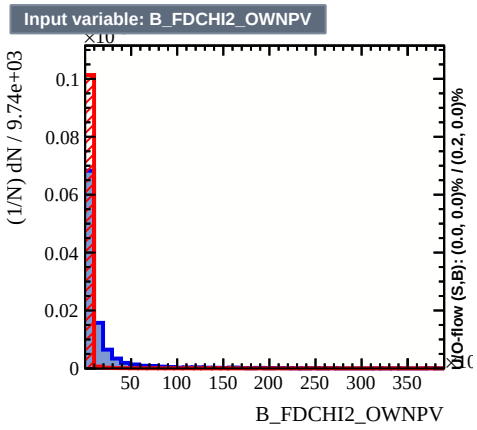
(b) Impact parameter of the π^-.



(c) Impact parameter of the μ^+.



(d) Impact parameter of the μ^+.



(e) Flight distance χ^2 of the B meson.

Figure A.2: Distributions of signal samples (blue) and background samples (red) for different variables used to train the BDT (2).

B Weight Distributions for Real Data

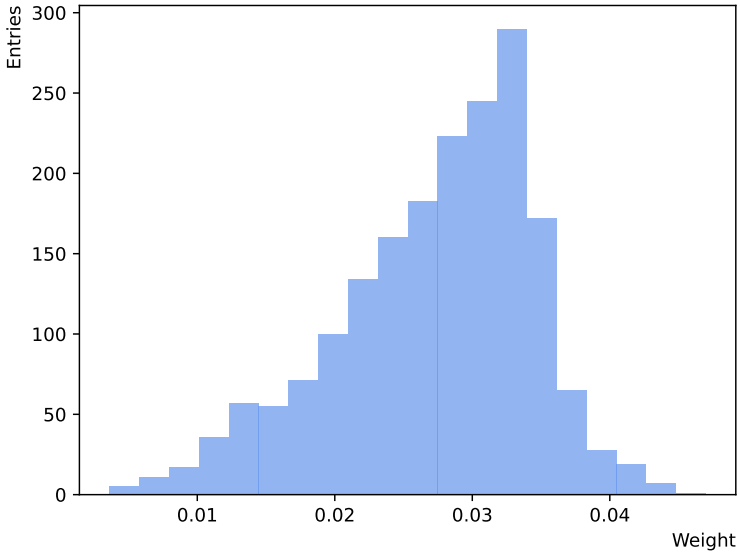


Figure B.1: Weight distribution for rare channel.

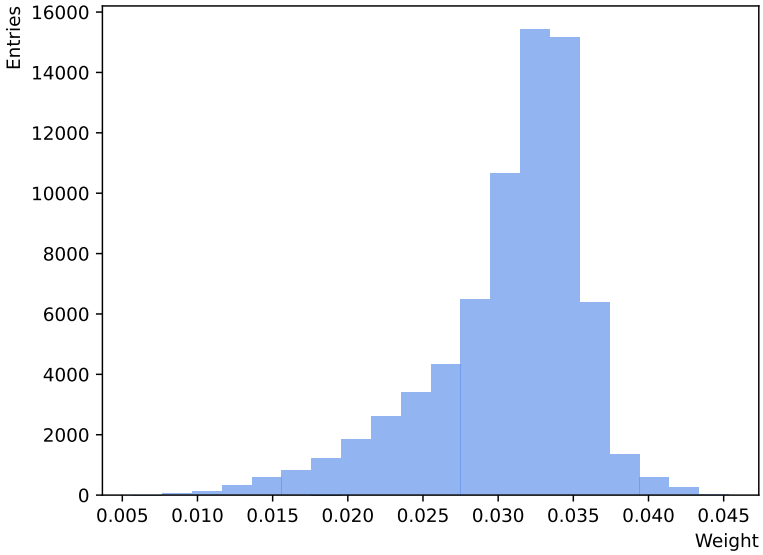


Figure B.2: Weight distribution for normalisation channel.

C Unweighted Mass Distribution Fits

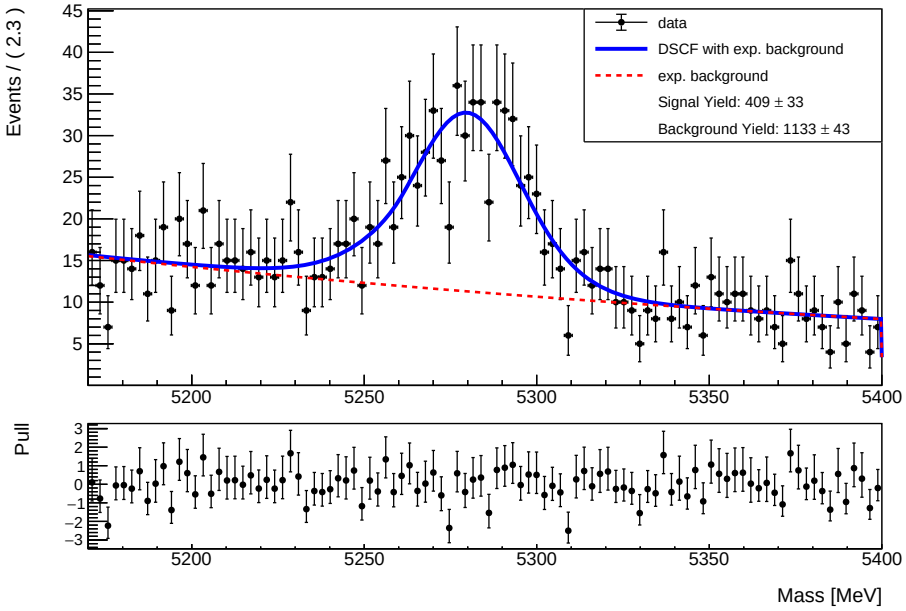


Figure C.1: Unweighted mass fit to rare channel using a double CB as signal component and exponential as background. Shape parameters have previously been determined on normalisation channel and fixed, as described in Section 7.2.

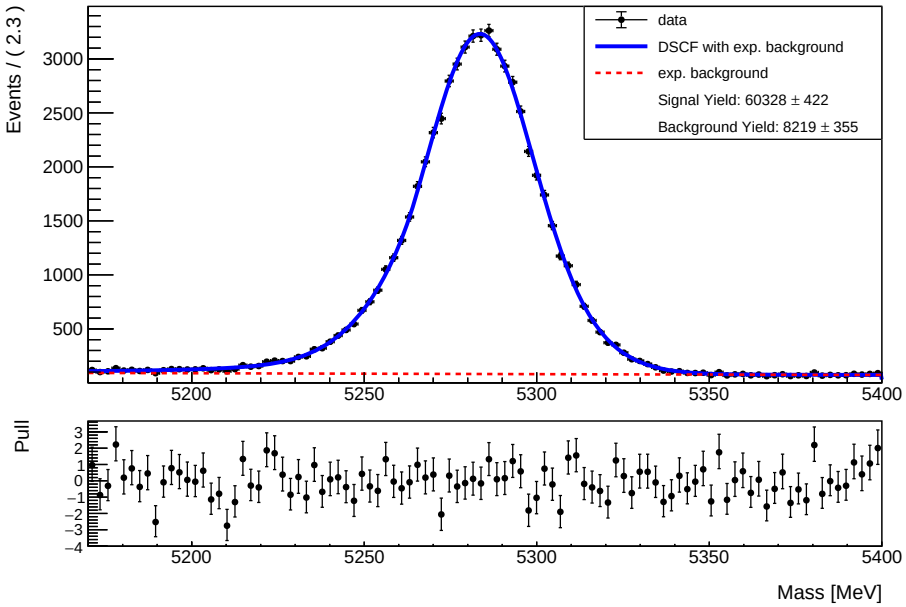


Figure C.2: Unweighted mass fit to normalisation channel using a double CB as signal component and exponential as background.

Table C.1: Fixed fit parameters for rare channel unweighted fit that were previously determined through fit on normalisation channel.

Parameter	Fixed value
α_1	3.00
α_2	1.33
n_1	5.00
n_2	2.56
σ_1	20.75
σ_2	14.06

Acknowledgements

First and foremost, I thank Dr. Christoph Langenbruch for supervising my thesis. Furthermore, I thank Prof. Dr. Klaus Reygers for agreeing to be the second referee.
I thank Prof. Dr. Ulrich Uwer for giving me the opportunity to work in such a kind and welcoming group, as well as the LHCb group in Heidelberg itself. Special thanks go to Rowina Caspary for proofreading my thesis.
I am grateful for my friends and fellow office mates, who have continuously provided IT- and emotional support. Thanks to Marius for never failing to brighten my day over a cup of coffee, and to Furkan for the gym sessions and Clone Wars watch parties that provided balance in my life (and, of course, in the force).
I thank my friends at home, Klein Kiki and Groß Kiki, for offering me a safe haven away from physics on afternoons filled with cappuccinos and gossip.
Der größte Dank gilt meiner Familie, Mama, Oma und Tini, die mich jederzeit bedingslos unterstützt hat.

Erklärung

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

Heidelberg, den 08.04.2024,

Laura Luisa Scholl