Department of Physics and Astronomy
University of Heidelberg

Bachelor Thesis in Physics
submitted by

Penelope Hoffmann

born in Heidelberg (Germany)

2023

Measurement of the branching fraction of the rare decay $B_s^0 \to \phi\gamma$

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

Abstract

This analysis presents a measurement of the branching fraction of the rare loop-level decay $B_s^0 \to \phi\gamma$ relative to the tree-level decay $B_s^0 \to J/\psi\phi$. Possible deviations in the value of the branching fraction from theoretical Standard Model predictions could indicate contributions from New Physics. The used dataset was collected by the LHCb experiment during Run 1 and Run 2, corresponding to an integrated luminosity of $\mathcal{L}_{int} = 9\,\mathrm{fb}^{-1}$ of pp collisions.
To efficiently select signal decays, a multivariate analysis was carried out. In total, 636 ± 27 $B_s^0 \to \phi\gamma$ decays are reconstructed. Using efficiencies determined from Monte Carlo simulations as well as the world-average of the branching fraction of the normalization channel, the total branching fraction is measured to

$$\mathcal{B}(B_s^0 \to \phi\gamma) = \left(xxx \pm 0.17_{stat.} \pm 0.14_{sys.}\right) \times 10^{-5},$$

where the uncertainties are of statistical and systematic nature, respectively. Since this analysis is in a preliminary stage, the results are blinded by omitting the central values.

Kurzfassung

Diese Analyse bestimmt das Verzweigungsverhältnis des seltenen Zerfalls $B_s^0 \to \phi\gamma$ relativ zu dem resonanten Zerfall $B_s^0 \to J/\psi\phi$. Abweichungen des gemessenen Wertes gegenüber den theoretischen Vorhersagen des Standardmodells könnten auf Beiträge Neuer Physik hinweisen. Die ausgewerteten Daten wurden während Run 1 und Run 2 des LHCb Experiments aufgenommen und entsprechen einer integrierten Luminosität von $\mathcal{L}_{int} = 9\,\mathrm{fb}^{-1}$ von pp-Kollisionen.

Zur Selektion des Zerfallssignals wurde eine multivariate Analyse durchgeführt. Insgesamt wurden 636 ± 27 $B_s^0 \to \phi\gamma$ Zerfälle rekonstruiert. Unter Verwendung von auf Monte Carlo Simulationen bestimmten Effizienzen und dem Weltdurchschnitt des Verzweigungsverhältnisses des Vergleichskanals, wurde das totale Verzweigungsverhältnis von $B_s^0 \to \phi\gamma$ zu

$$\mathcal{B}(B_s^0 \to \phi\gamma) = \left(xxx \pm 0.17_{stat.} \pm 0.14_{sys.}\right) \times 10^{-5}$$

bestimmt. Die angegebenen Fehler sind statistischer respektive systematischer Natur. Da sich die Analyse in einem vorläufigen Stadium befindet, werden die endgültigen Werte nicht angegeben.

Contents

1 Introduction

Our modern-day understanding of particle physics is based on one fundamental model: the Standard Model (SM). This model was developed in the 1960s when Glashow, Salem, and Weinberg proposed the electroweak theory that unified the electromagnetic and the weak force in a single theory. Their theory was first confirmed when neutral weak currents were found at the Gargamelle Bubble Chamber at CERN in 1973. Since then, various experiments have been carried out in order to test the predictions of the Standard Model as well as to determine its parameters, e.g. particle masses and coupling constants. While no significant deviations from the theoretical predictions were found, the Standard Model cannot explain phenomena such as matter-antimatter asymmetry and needs to be expanded as the observation of neutrino oscillations shows. Oscillations are only possible if the particles involved have mass eigenstates - until the discovery of neutrino oscillations, neutrinos were assumed to be massless in the SM.

The search for New Physics (NP) - physics beyond the Standard Model - is divided into two different approaches: the direct approach tries to directly produce yet unknown particles not included in the SM through high-energy collisions, while the indirect approach searches for virtual contributions of such new particles at loop level. Rare decays are an excellent possibility for searching for such contributions. The transition of a b quark to a s quark is suppressed in the SM as it is only allowed at loop-level and therefore is susceptible to contributions from heavy particles. The LHCb experiment at CERN is designed for the study of hadrons involving a b quark and is ideally suited for precision measurements of $b \to s$ decays. In this thesis, the rare decay $B_s^0 \to \phi\gamma$ and the normalization channel $B_s^0 \to J/\psi\phi$ are studied using the dataset collected during Run 1 and Run 2 at the LHC, corresponding to an integrated luminosity of $3\,\mathrm{fb}^{-1}$ and $6\,\mathrm{fb}^{-1}$, respectively.

This thesis is structured as follows: First, an introduction to the Standard Model and the physics behind rare decays is given in Section 2. The LHCb detector is presented in Section 4. The analysis strategy of the data, which is prepared via selection cuts (Section 5) and a multivariate analysis (Section 6), is outlined in Section 3. The process of fitting the reconstructed mass distribution of the B meson is documented in detail in Section 7. The determination of the cut efficiencies and subsequent calculation of the branching fraction is described in Section 8.1 and Section 8.2, respectively.

2 Theoretical Background

This section gives a brief overview of the physics underlying rare decays and the motivation for this research field, starting with an introduction to the Standard Model largely based on [1].

2.1 The Standard Model

The Standard Model of particle physics is an effective gauge quantum field theory (QFT) that incorporates the electroweak theory and Quantum Chromodynamics (QCD). It has been tested extensively through experimental searches and has successfully predicted the elementary particles including the Higgs-Boson discovered in 2012. With this latest discovery, the search for New Physics beyond the Standard Model has become one of the main foci of particle physics.

2.2 Fundamental Particles

The fundamental particles in the Standard Model are the twelve fermions, the five gauge bosons, and the Higgs boson. All fermions have a spin of $\frac{1}{2}$ and follow Fermi-Dirac statistics. As the description of their dynamics is subject to the Dirac equation of relativistic quantum mechanics, each fermion has a corresponding anti-particle of the same mass but of opposite quantum numbers. The fermions can be divided further into the leptons and the quarks which have differing physical properties and are listed in Table 1. Leptons interact weakly and in the case of the three particles e, μ and τ, that possess an electric charge of $1\mathrm{e}$, also via the electromagnetic force. Quark interactions are dominated by the strong force, but can also occur via the weak and the electromagnetic force as quarks carry electric and colour charge. Quarks with an electric charge of $+\frac{2}{3}\mathrm{e}$ are referred to as up-type quarks, while quarks with an electric charge of $-\frac{1}{3}\mathrm{e}$ are called down-type quarks. The fermions are categorized into three generations, where each succeeding generation has a higher mass but the same fundamental interactions. The first and only stable generation consists of the up and down quark as well as the electron and the electron neutrino. Due to colour confinement, quarks cannot exist as free particles but only as colourless mesons or baryons depending on whether the hadrons consist of a combination of a quark and an anti-quark or of three quarks, respectively.

	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 1: The fundamental fermions sorted into the three generations and into quarks and leptons. The values are taken from [2]. The neutrino upper mass bounds are based on the sum of all flavours.

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

Table 2: The fundamental gauge bosons and the Higgs boson [2].

The three forces, electromagnetic, strong, and weak, are mediated by the exchange of the five gauge bosons. The gauge bosons have spin 1 and therefore follow Einstein-Bose statistics. The Higgs Boson is the only scalar boson (spin 0) in the SM and gives the fundamental particles their masses via the mechanism of spontaneous symmetry breaking. A summary of the boson properties is given in Table 2.

2.2.1 Fundamental Forces

The three fundamental forces in the Standard Model are all described by their respective field theory and can be characterized by their gauge bosons, strength and range. The field theory corresponding to the electromagnetic force is called Quantum Electrodynamics (QED). Due to the massless photon serving as a mediator, the interaction range is spatially infinite, although the strength decreases with the inverse of the distance. In interactions via the electromagnetic force, which is charge-conserving, only charged particles participate.

This is not the case for the weak interaction, in which all fundamental particles can participate. Since it is mediated by massive bosons, its interaction range is limited. The W

bosons, responsible for the flavour-changing weak-charged current, couple to fermions with a difference of one unit in the electric charge, while the electrically neutral Z boson mediates the weak-neutral currents. All three bosons carry weak charge and can therefore couple to each other, while the W bosons due to their electric charge can also couple to the photon. The weak and the electromagnetic force are unified in the electroweak theory.

The strong force, described by Quantum Chromodynamics (QCD), is limited to interactions between particles carrying colour charge, which is solely a property of gluons and quarks. It is mediated by an octet of charged gluons that carry a combination of colour and anti-colour of which there are three (red, green, blue and anti-red, anti-green, anti-blue). Since coulour-charged particles cannot exist independently, quarks coupled with gluons form colour-neutral hadron states. For a hadron to be colour-neutral, it either has to be composed of a quark carrying a colour together with an anti-quark carrying the respective anti-colour or of quarks that together cover all three different colours.

2.3 Flavour Physics in the Standard Model

Flavour physics is the research field dealing with the weak interaction of quarks and leptons. Weak-charged currents change both flavour and charge, while weak neutral currents conserve both charge and flavour. In contrast to the universal coupling strength of the weak interaction of charged leptons, the coupling strengths vary for different quark flavours. This experimental observation motivates the notion that the flavour eigenstates of quarks are not equal to their mass eigenstates. The relation between the different eigenstates is given by the unitary Cabibbo-Kobayashi-Maskawa (CKM) matrix

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

The probability of an eigenstate i transitioning into an eigenstate j is proportional to $\left|V_{ij}\right|^2$. At tree-level, only the transition from an up-type quark flavour to a down-type quark flavour via a W-Boson and vice versa is allowed. Additionally, the transition of flavours within a generation is favoured over the transition between different generations resulting in a near-diagonal form of the CKM matrix [1]. This hierarchy is emphasized by the Wolfenstein parametrization that writes the CKM matrix as a Taylor expansion in a newly introduced parameter λ. In total, the Wolfenstein parametrization has four physical parameters: λ, A, ρ and the complex phase η [3].

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

Up to $\mathcal{O}\left(\lambda^3\right)$, the complex components of the CKM Matrix are included solely in the matrix elements V_{ub} and V_{td}. In order for CP violation to occur, the complex phase η in these two elements has to be non-zero. The Wolfenstein parameters are determined by a global fit on all available experimental measurements to [4]:

$$\lambda = 0.225\,00 \pm 0.000\,67, \quad A = 0.826^{+0.018}_{-0.015},$$
$$\bar{\rho} = \rho(1-\frac{\lambda^2}{2}) = 0.159 \pm 0.010, \quad \bar{\eta} = \eta(1-\frac{\lambda^2}{2}) = 0.348 \pm 0.010.$$

2.4 Search for New Physics

The indirect approach to searches for New Physics focuses on processes that could possibly be influenced by new particles - such as weak interactions. In both charged current (CC) and flavour-changing neutral current (FCNC) processes the quark flavour changes. However, the electric charge is conserved for the FCNCs. Consequently, FCNCs can only occur at loop-level. At loop-level, these processes are suppressed by their CKM matrix elements and by the mass difference of the quarks in the loop (GIM mechanism). Interactions via charged currents can take place at tree-level and therefore also dominate in weak decays [5]. Processes that transition via FCNCs, such as rare decays or neutral meson mixing, are a good search target for particles beyond the Standard Model, as New Physics can give significant contributions that would affect observables like the branching ratio of rare decays.

2.4.1 Rare Decay $B_s^0 \to \phi\gamma$

One advantage of using rare decays involving the decay of a B meson lies in the clean signature of a well-observable displaced secondary vertex due to the relatively long lifetimes of the B mesons. Additionally, attributable to the large production cross-section

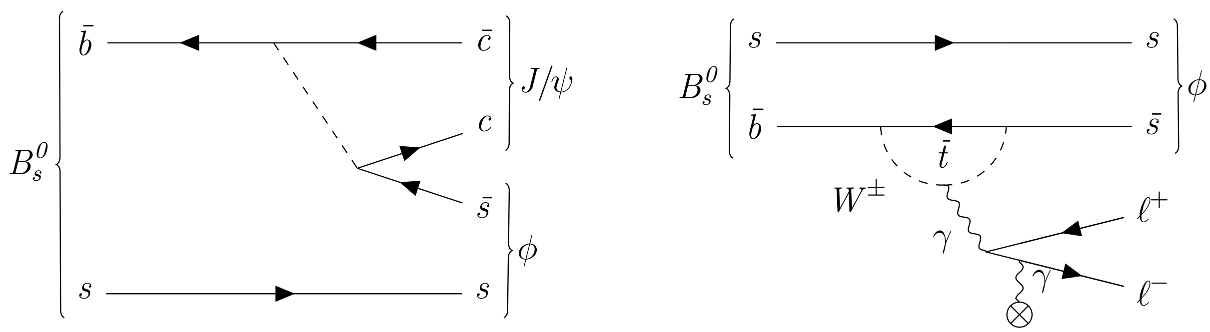


Figure 1: Lowest order Feynman graphs of the decays $B_s^0 \to J/\psi\phi$ (left) and $B_s^0 \to \phi\gamma$ (right). The J/ψ can decay further into $J/\psi \to e^+e^-$, the ϕ via $\phi \to K^+K^-$. For $B_s^0 \to \phi\gamma$, the photon is reconstructed via the displayed $\gamma \to e^+e^-$ conversions in material.

of $b\bar{b}$ quark pairs at high energies, statistically powerful datasets can be recorded. Inclusive radiative rare decays of the type $B \to X_s\gamma$ that involve the transition of a *b* quark to a *s* quark are well suited for the search for potential New Physics contributions. They are easier to predict theoretically than fully hadronic rare decays and are very susceptible to New Physics. A Feynman graph of the studied radiative rare decay $B_s^0 \to \phi\gamma$, where the ϕ meson can further decay via $\phi \to K^+K^-$, is displayed in Figure 1. Using $\gamma \to e^+e^-$ conversions in material, the γ is reconstructed from the electron and positron tracks. The decay can be used to determine the CKM element V_{ts}, as the top quark is the heaviest quark in the loop and therefore dominates the transition amplitude. The amplitude is proportional to the corresponding CKM elements, the couplings, and the ratio $\frac{m_t}{m_W}$. Here, m_t is the mass of the top quark and m_W is the mass of the *W* boson. While transitions of the type $b \to s\gamma$ were first discovered by the CLEO experiment, $B_s^0 \to \phi\gamma$ was first observed by the B-factory Belle [6] and has already been studied at the LHCb using Run 1 data [7]. In both cases, the photon was reconstructed from energy deposits in the Electromagnetic Calorimeter. In this analysis, using the complete Run 1 and Run 2 data, the branching fraction is determined relative to the dominant tree-level decay $B_s^0 \to J/\psi\phi$ (Figure 1), where J/ψ and ϕ can further decay via $J/\psi \to e^+e^-$ and $\phi \to K^+K^-$. This decay, with the same final state particles as the rare decay, was chosen in order to maximize the cancellation of experimental systematic uncertainties. The current theoretical prediction as well as the experimental world-average of the branching fraction of $B_s^0 \to \phi\gamma$ are listed in Table 3, the world-average of the branching fraction of $B_s^0 \to J/\psi\phi$ can be found in Table 4.

		Branching Fraction
SM prediction	NNLO[1] method	$(4.3 \pm 1.4) \times 10^{-5}$
Experimental values	world-average	$(3.4 \pm 0.4) \times 10^{-5}$
	Belle	$(3.6 \pm 0.5 \pm 0.7) \times 10^{-5}$
	LHCb ($1\,\mathrm{fb}^{-1}$)	$(3.38 \pm 0.34 \pm 0.20) \times 10^{-5}$

Table 3: Theoretical and experimental literature values for the branching fraction of the rare decay $B_s^0 \to \phi\gamma$. The theoretical value is obtained from [8] and the world-average experimental value from [2]. The uncertainties on the experimental values found by Belle [6] and by the LHCb, using $1\,\mathrm{fb}^{-1}$ of data collected during Run 1 [7], are statistical and systematic uncertainties, respectively.

Value	Branching Fraction
Experimental (world-average)	$(1.04 \pm 0.04) \times 10^{-5}$

Table 4: World-average experimental literature value for the branching fraction of the tree-level decay $B_s^0 \to J/\psi\phi$ [2].

[1] next to next to leading order

3 Outline of the Analysis Strategy

The aim of this thesis is the determination of the total branching fraction of the rare loop-level B meson decay $B_s^0 \to \phi\gamma$ using the tree-level decay $B_s^0 \to J/\psi\phi$ as the normalization channel. The J/ψ further decays into e^+e^- and the ϕ into K^+K^-. The γ is reconstructed from $\gamma \to e^+e^-$ conversions in material. Due to the same final state particles as in the $B_s^0 \to \phi\gamma$ decay, the use of $B_s^0 \to J/\psi\phi$ as a normalization channel proves to be advantageous: Factors such as the PID efficiencies or the $\sigma_{b\bar{b}}$ production cross section as well as most systematic uncertainties cancel each other out, increasing the precision of the branching fraction measurement. Additionally, the total branching fraction of $B_s^0 \to J/\psi\phi$ has already been determined - also by the LHCb - with an uncertainty of $3.8\,\%$.

The analysis consists of the following steps:

1. Dataset Stripping: The dataset is produced by the LHCb Collaboration having undergone the stripping process, which is a loose preselection choosing the channels relevant to the analysis. This is done to reduce the data to a manageable size.
2. Signal Selection:
   a. Signal Preselection: In order to extract the signal and to minimize the combinatorial background, loose cuts are applied.
   b. Multivariate Analysis: A Multivariate Analysis is carried out to further reduce the combinatorial background. An XGBoost Classifier is trained on the preselected data sample and the classifier output is optimized for both the normalization channel and the rare mode.
3. Signal Yield Determination: On Monte Carlo simulations, a fit model is determined. Using Maximum Likelihood Fits the signal yields are obtained by fitting the mass distributions for $B_s^0 \to \phi\gamma(\to e^+e^-)$ and for $B_s^0 \to J/\psi(\to e^+e^-)\phi$ candidates.
4. Efficiency Determination: The signal selection cut efficiencies for both channels are calculated using Monte Carlo simulated samples.
5. Branching Fraction Calculation: Using the signal yields and the efficiencies, the relative branching fraction is calculated by

$$\frac{\mathcal{B}(B_s^0 \to \phi\gamma)}{\mathcal{B}(B_s^0 \to J/\psi(\to e^+e^-)\phi)} = \frac{N(B_s^0 \to \phi\gamma)}{N(B_s^0 \to J/\psi(\to e^+e^-)\phi)} \times \frac{\epsilon_{J/\psi\phi}}{\epsilon_{\phi\gamma}}. \tag{3.1}$$

N is the respective signal yield, and ϵ is the respective reconstruction cut selection efficiency for the normalization channel and the rare mode. The total branching fraction $\mathcal{B}(B_s^0 \to \phi\gamma)$ is obtained by multiplying with the total branching fraction $\mathcal{B}(B_s^0 \to J/\psi(\to e^+e^-)\phi)$.

Since this analysis uses the full available data taken by the LHCb detector and a full determination of the systematic uncertainties is beyond the scope of this thesis, the results of the efficiency and branching fraction calculation are blinded by omitting the calculated central value and only giving the associated uncertainties. This still allows the comparison with the theoretical and the present experimental prediction.

4 The LHCb Detector

The Large Hadron Collider is a proton-proton (pp) collider at CERN located near Geneva on the Swiss-French border. It operated at center-of-mass energies of $\sqrt{s} = 7$ TeV in 2011, $\sqrt{s} = 8$ TeV in 2012 and $\sqrt{s} = 13$ TeV from 2015 to 2018. The Large Hadron Collider beauty experiment (LHCb) is a detector at the LHC with the investigation of CP violation and rare decays as its main focus. It mainly studies decays involving bottom (hence the name) and charm quarks. In this section, a brief overview of the detector layout will be given, followed by an introduction to the tracking, particle identification, and trigger system.

4.1 Overview

The LHCb detector (Figure 3) is a single-arm forward spectrometer with a length of $20\,\mathrm{m}$ that is designed for forward tracking. This particular layout was chosen in order to accommodate for the characteristic production of $b\bar{b}$ quark pairs at small angles to the beamline in either the forward or the backward direction. As illustration, the polar angles of b and $\bar{b}$ quarks simulated for $\sqrt{s} = 14\,\mathrm{TeV}$ are given in Figure 2. The LHCb detector covers an angular range of 10 to 300(250) mrad in the horizontal magnetic bending (vertical non-bending) plane, which allows for approximately 20 % of all produced pairs to be reconstructed [9]. A right-handed coordinate system is used, with the z-axis in the direction of the beamline, y in the vertical direction, and x in the horizontal direction.

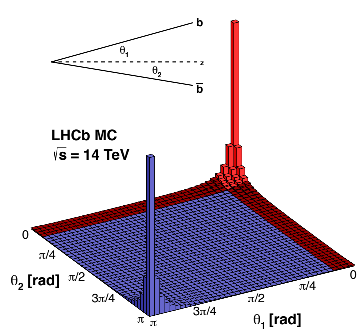


Figure 2: Simulation of the polar angles of b and $\bar{b}$ quarks at $\sqrt{s} = 14\,\mathrm{TeV}$ using PYTHIA8 and CTEQ6 NLO [10].

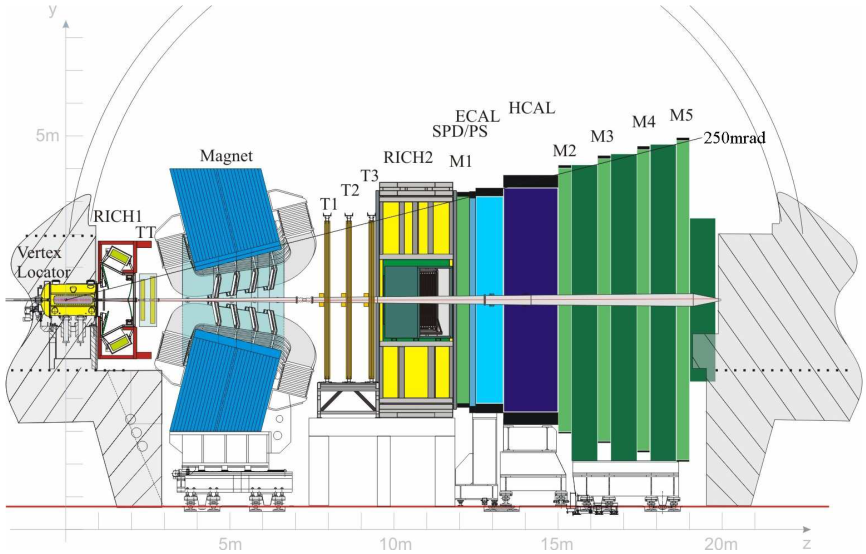


Figure 3: Side view of the LHCb detector in the y-z-plane [11].

4.2 Magnet

For the momentum measurement of charged particles traversing the detector, the curvature of the charged tracks inside a magnetic field is used. In LHCb's case, this is a warm dipole magnet made up of two saddle-shaped coils symmetrically mounted in a window-frame yoke with increasing pole gap (see Figure 4) in order to meet the required detector acceptance. It produces a vertical magnetic field with an integrated power of $4\,\mathrm{T\,m}$ for $10\,\mathrm{m}$ tracks in z-direction and extends into the Ring-Imaging-Cherenkov system with a field of $2\,\mathrm{mT}$ [11]. For data-taking, the polarity is reversed regularly (`MagUp` and `MagDown`), thereby minimizing effects due to detector and subsequent interaction asymmetries [12].

4.3 Tracking System

In order to determine the transverse momentum from the curvature of the track, the track has to first be reconstructed. Additionally, an excellent vertex resolution is required to determine the origin of particles and to achieve a good invariant mass resolution. At the LHCb, this is accomplished by the tracking system consisting of the Vertex Locator, the Silicon Tracker, and the Outer Tracker.

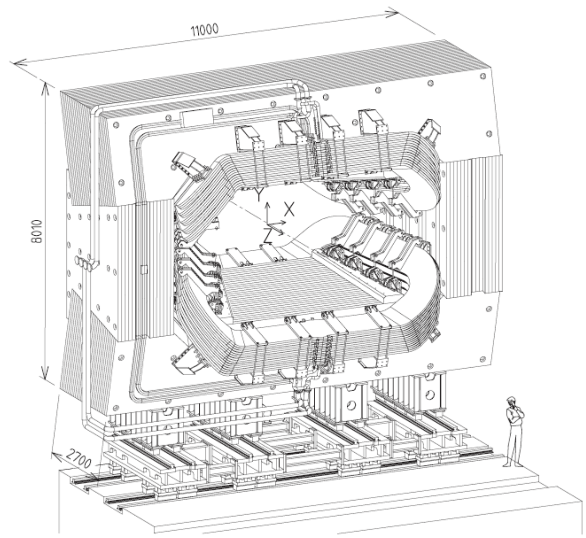


Figure 4: Perspective view of the LHCb dipole magnet [11].

Vertex Locator (VELO): The VELO is a silicon microstrip detector that is placed in close proximity (7 mm) to the beam at the interaction point [13]. It consists of two halves with slight overlap for full angular coverage that are retracted for safety purposes during beam injection. The semi-circular modules in each half are placed in the beam direction and contain one sensor for measuring the R-coordinate and one for the ϕ-coordinate. Its measurement of track coordinates is essential for the identification of the primary vertex and of the distinctive displaced secondary vertex for *b*-hadron decays.

Silicon Tracker (ST): The Silicon Tracker refers to two silicon microstrip detectors with 200 $\mu\mathrm{m}$ strip pitch and a single-hit resolution of 50 $\mu\mathrm{m}$, the Tracker Turicensis (TT) and the Inner Tracker (IT). The TT is positioned upstream of the magnet and covers the full detector acceptance, while the IT at the center of the Tracking stations T_1 - T_3 only covers the region close to the beam where track multiplicity is the highest. Each module of the two detectors is made up of 4 slightly overlapping detection layers with a x-u-v-x geometry. In the first and last layer, the strips are placed vertically, while in the second and third layer, the strips have a rotation of -5 and $+5$ $^\circ$, respectively, in order to measure the transverse momentum component [11].

Outer Tracker (OT): The outer regions of the T_1 - T_3 are covered by a gaseous straw tube detector - the Outer Tracker. The drift-time detector for charged particles has a x-coordinate resolution of 200 $\mu\mathrm{m}$ and its geometry is similar to IT's and TT's x-u-v-x geometry [14]. IT and OT combined fully cover the acceptance of LHCb.

4.4 Particle Identification System

Particle Identification is an essential requirement for this analysis as it allows the distinction between decays with the same decay topology and the reduction of backgrounds from misidentified decays.

Ring-Imaging-Cherenkov system (RICH): The Ring-Imaging-Cherenkov system aims to identify charged hadrons by using two detectors. RICH1, situated upstream of the magnet, covers the lower spectrum of momentum of $2\text{-}40\,\mathrm{GeV}$ and the complete angular acceptance. RICH2, downstream of the magnet only identifies particles with a high momentum of $15\text{-}100\,\mathrm{GeV}$ and therefore only covers the angular range of $15\text{-}120\,\mathrm{mrad}$. The detectors measure the velocity-dependent Cherenkov angle $\cos\theta = \frac{1}{\beta n}$, at which Cherenkov photons are emitted by particles traversing material with the refractive index n at speeds exceeding the speed of light in the radiator [15]. Figure 5 shows the Cherenkov angle for different charged hadrons. Together with the track momentum measured by the tracking system, the mass of the particles can be estimated.

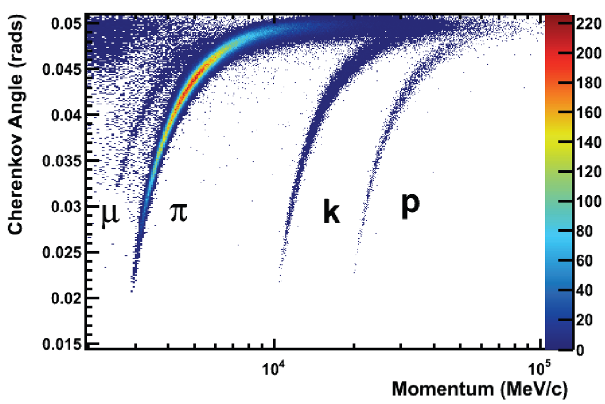


Figure 5: Cherenkov angles in dependency of the track momentum for different particles in RICH 1 [15].

Calorimeter system: The calorimeter system's task is the identification of hadrons, electrons, and photons as well as the measurement of their energy [16]. Energy is deposited either by hadronic showers in the case of hadrons or by EM-showers induced by bremsstrahlung and pair production in the case of charged particles and photons.

- **Scintillating Pad Detector (SPD):** The SPD is the first layer of the calorimeter in which a hit indicates the presence of a charged particle and thus allows the separation of electrons and photons.

- **PreShower Detector (PS):** The PS is also a scintillating pad detector that has a 15 mm thick layer of lead in front of its scintillators. This layer marks the start of an EM-shower thereby allowing to distinguish between charged hadrons and electrons or photons.
- **Electromagnetic Calorimeter (ECAL):** The ECAL measures the transverse energy of particles as well as their shower position and consists of 4 mm thick scintillator plates that alternate with 2 mm thick lead absorbers. To contain the complete EM shower and thereby optimize the energy resolution the calorimeter covers a total of 25 radiation lengths.
- **Hadronic Calorimeter (HCAL):** The HCAL is a succession of iron absorbers with active scintillator material and is placed parallel to the beam axis. Even though the HCAL only covers 5.6 interaction lengths and therefore does not contain the full hadronic shower, the length suffices in order to give an estimate of the measured transverse energy and shower shape.

The increase in particle density for the regions close to the beamline is reflected in the segmentation of the calorimeters as seen in Figure 6. Smaller calorimeter cells are placed closer to the beam line - the granularity decreases with the distance from the beamline.

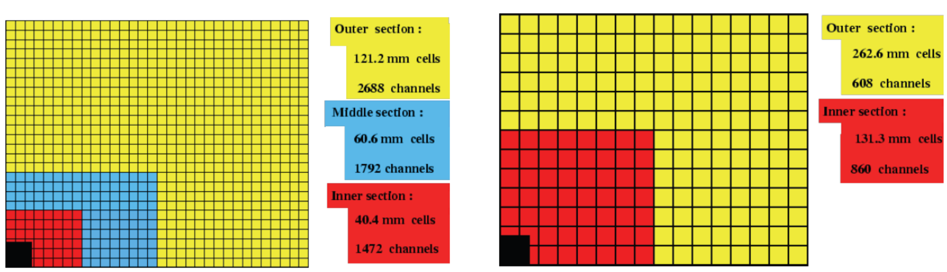


Figure 6: Segmentation of calorimeter cells of LHCb's SPD, PS, ECAL (left) and HCAL (right) [16].

Muon system: The muon system is comprised of 5 detector stations placed along the beam axis. The first station M1 is located upstream of the calorimeter in order to improve the p_T measurement in the L0 trigger. M2-5 are placed downstream of the calorimeter with 80 cm thick iron absorbers placed in between each station and are mainly used for identifying and tracing the muons. As such, potential muon tracks are first verified with M2-5. M1 and M2-3 have a high spatial resolution along the x-axis and are therefore used for the measurement of the transverse momentum, p_T, of the muons. M4-5 mainly serve to identify particles with high p_T, as only muons with a minimum momentum of 6 GeV can pass through all 5 stations [17].

4.5 Trigger System

In order to process the magnitude of data produced in the pp collisions, a trigger system is required. At the LHCb, it is designed to differentiate between b and c decays and backgrounds from light quarks. The trigger system is split into two levels, the hardware-based Level-0 Trigger and the software-based High-Level Trigger (HLT).

4.5.1 Level-0 Trigger

The goal of the L0 Trigger level is to reduce the bunch crossing rate of $40\,\mathrm{MHz}$ to 1 MHz with which the full detector can be read out within the fixed latency time of 4 µs. Information from the calorimeter system is processed in the L0-calorimeter trigger, where the transverse energy E_T is calculated for clusters in 2x2 cells in the ECAL and the HCAL:

$$E_T = \sum_{i=1}^{4} E_i \sin\theta_i \tag{4.1}$$

θ_i corresponds to the angle between the z-axis and the line connecting the cell center to the mean position of the pp collision. Using additional information from the SPD and the PS, events with high momentum hadrons, electrons, and photons are identified and accepted if their energy passes a certain threshold [18].

With the data from the muon system the muon trigger searches for events with hits in a straight line through the five stations and with origin in proximity to the interaction point in the y-z-plane. For a candidate to be accepted, either the p_T from the muon with the highest p_T has to exceed the `L0Muon` p_T threshold or the product of the p_T from the muons with the largest and second-largest value surpasses the `L0DiMuon` threshold.

4.5.2 High Level Trigger

The High Level Trigger (HLT) is a software trigger that runs on an Event Filter Farm and possesses two stages. Due to time constraints, only a partial event reconstruction is performed during the first trigger stage HLT1. For this, track segments are reconstructed from the VELO and are then extended in the tracking stations. Only detached

tracks with significant transverse momentum p_T pass the HLT1 stage.
The second stage HLT2 runs the particle identification using the full information from the RICH and calorimeter system. The HLT reduces the data rate from 1 MHz to 3-5 kHz in Run 1 [19] and to 12.5 kHz in Run 2 [20] which is then saved to storage. Since Run 2, the two HLT stages are able to run asynchronously which allows for the data from HLT1 to be used for direct calibration and alignment of the detector thereby improving the reconstruction [18, 20].

4.6 Recovery of Bremsstrahlung Photons at LHCb

In interactions with the detector material, electrons can lose energy via the emission of a photon thereby leading to a reduction in momentum resolution and tracking efficiency for the electrons. A distinction is drawn between two kinds of bremsstrahlung photons: If the photon is emitted after the electron has passed the magnetic field, the electron's momentum determination is not affected, since the curvature of the track remains unchanged. However, if the photon emission occurs before the electron traverses the magnetic field, the momentum measurement is biased. The loss in energy due to emitted bremsstrahlung radiation reduces the momentum of the electron, thereby impacting the curvature of the electron track. This reflects itself in a poorer resolution of the invariant mass distribution. Thus, bremsstrahlung reconstruction is necessary.

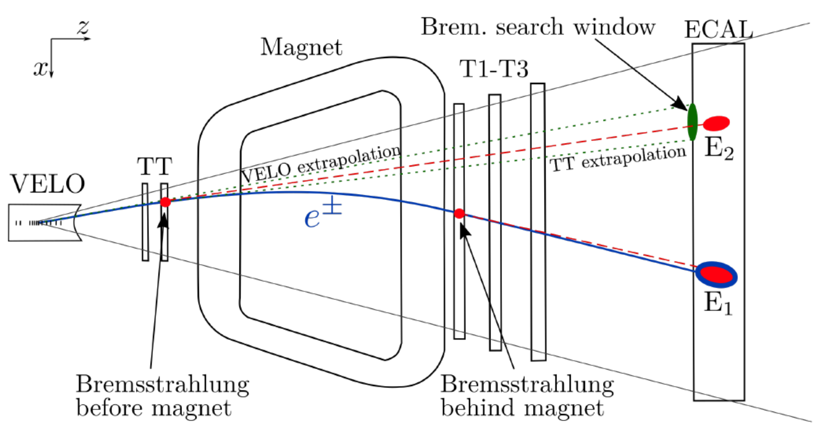


Figure 7: Photon reconstruction process as seen from the top of the detector [21]. Photons emitted by bremsstrahlung are marked with red dashed lines. The green region denoted Brems. search window corresponds to the extrapolated window in which a photon cluster is associated to the corresponding electron.

If the emitted photon is out of the ECAL or does not have the required energy to start a shower, it cannot be reconstructed. In the case of the production of a shower with its energy exceeding a certain energy threshold, the shower cluster can be reconstructed. In order to associate a photon cluster to its electron, the photon cluster has to be located in the ECAL area between the assumed electron track extrapolated solely using VELO information and the assumed electron track extrapolated using VELO and TT information [22]. This process is outlined in Figure 7. The momentum measurement is corrected by adding the energy of the photon cluster in the ECAL to the momentum of the associated electron.

5 Data Analysis

This section documents the dataset and the variables in this analysis. Initial stripping and preselection cuts are discussed.

5.1 Dataset and Monte Carlo Simulation

The data used in this thesis was collected by the LHCb experiment during the years 2011 to 2012 (Run 1) and 2015 to 2018 (Run 2) with two different magnet polarities (`MagUp` and `MagDown`). During those years, the center-of-mass energy increased from $\sqrt{s}$ = 7 TeV in 2011 to $\sqrt{s}$ = 8 TeV in 2012 to $\sqrt{s}$ = 13 TeV from 2015 on. The corresponding integrated luminosity for Run 1 is $3\,\mathrm{fb}^{-1}$ and $6\,\mathrm{fb}^{-1}$ for Run 2. Due to the changes in the center-of-mass energies and in the detector configurations, the full dataset is split into three smaller sub-samples. The Run 1 dataset is made up of the data recorded in the years 2011 and 2012. Run 2, part 1, summarizes 2015 and 2016 data, while Run 2, part 2, includes the 2017 and 2018 data.

In addition to the recorded data, Monte Carlo (MC) simulations for both the normalization channel and the rare decay are used. Before utilizing the simulations, the events have to pass the same reconstruction and selection as the dataset. Each data sub-sample has its corresponding Monte Carlo simulation - MC 2012 corresponds to the Run 1 dataset, MC 2016 to Run 2 part 1 and MC 2018 to Run 2 part 2. Monte Carlo simulations serve as a proxy for the signal of the corresponding dataset and have the advantage that even rare decays can be produced with high statistics. They simulate signal events by taking theoretical predictions and experimental observations as well as the detector geometry and its acceptance into account. In Table 5, the Monte Carlo simulations along with the number of generated events are listed.

5.1.1 Definition of Variables

The meaning of the variables used during the course of this analysis is explained below.

Mass of particle X: The invariant mass of a particle X is calculated from the four-momentum $\mathrm{p} = \left(E, p_x, p_y, p_z\right)^T$, where E is the energy and $\vec{p} = \left(p_x, p_y, p_z\right)^T$ is the momentum vector, by:

$$m_X = \sqrt{\mathrm{p}_X^2}.$$

year	Decay	# Simulated Events	ϵ (Gen. level)
2012	$J/\psi\phi$ up	1008410	0.16614
	$J/\psi\phi$ down	1004107	0.16614
	$\phi\gamma$ up	500901	0.26147
	$\phi\gamma$ down	502669	0.26147
2016	$J/\psi\phi$ up	1988734	0.16614
	$J/\psi\phi$ down	2091446	0.16614
	$\phi\gamma$ up	2002026	0.270145
	$\phi\gamma$ down	2004513	0.270145
2018	$J/\psi\phi$ up	5002902	0.177295
	$J/\psi\phi$ down	4991002	0.177295
	$\phi\gamma$ up	2037824	0.2700
	$\phi\gamma$ down	2006224	0.2700

Table 5: Number of simulated Monte Carlo events and their corresponding generation efficiency for samples used in this analysis. Up and down refer to the polarity of the magnet.

Transverse Momentum: The momentum component perpendicular to the beam line, z-direction in LHCb's case, is defined as:

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

Impact Parameter: The impact parameter (IP) is given by the minimal distance between a reconstructed particle track and the primary vertex (PV). Its $\chi^2_{\texttt{IP}}$ measures the likelihood of the track originating from the vertex by taking the difference of the χ^2 of the vertex before and after the reconstructed track was added.

Vertex $\boldsymbol{\chi^2}$: The χ^2 fit of the decay vertex is a measure for its fit quality.

Direction Angle: The direction angle (`DIRA`) is the cosine of the angle between the reconstructed momentum vector and the particle's flight direction that is described by the vector pointing from the primary vertex to the decay vertex.

Flight Distance: The flight distance is the distance the particle has traveled before decaying - e.g. the distance between production and decay vertex.

PID/DLL: The PID-system generates a Likelihood function $\mathcal{L}$ for each particle. The difference of the logarithmic likelihood of a given particle X to the logarithmic likelihood of a pion

$$\Delta \ln(\mathcal{L}(X - \pi)) = \ln \mathcal{L}(X) - \ln \mathcal{L}(\pi) = \ln \left(\frac{\mathcal{L}(X)}{\mathcal{L}(\pi)} \right) \tag{5.1}$$

Type	Requirement
Global Event Cut	$\texttt{nSPDHits} < 600$
B	$\|m - m_B^{\texttt{PDG}}\| < 1500\,\mathrm{MeV}$ $\texttt{DIRA} > 0.9995$ $\chi^2_{\texttt{IP}}(\mathrm{PV}) < 25$ end vertex $\chi^2/\mathrm{ndf} < 9$ PV χ^2 separation > 100
ϕ	$985\,\mathrm{MeV} < m_\phi < 1100\,\mathrm{MeV}$ $p_T > 400\,\mathrm{MeV}$ origin vertex $\chi^2/\mathrm{ndf} < 25$
K	$\texttt{DLL}_{K\pi} > -5$ (only data) $\chi^2_{\texttt{IP}}(\mathrm{PV}) > 9$
ee	$m < 5000\,\mathrm{MeV}$ end vertex $\chi^2/\mathrm{ndf} < 9$ origin vertex χ^2 separation > 16
e	$\texttt{DLL}_{e\pi} > 0$ (only data) $p_T > 300\,\mathrm{MeV}$ $\chi^2_{\texttt{IP}}(\mathrm{PV}) > 9$

Table 6: Overview of the cuts applied during the stripping process. The utilized stripping line is `Bu2LLKeeLine2`.

is referred to as Delta Log Likelihood $\mathrm{DLL}_{X\pi}$ or PIDX, where X denotes the particle hypothesis.

5.2 Stripping Process

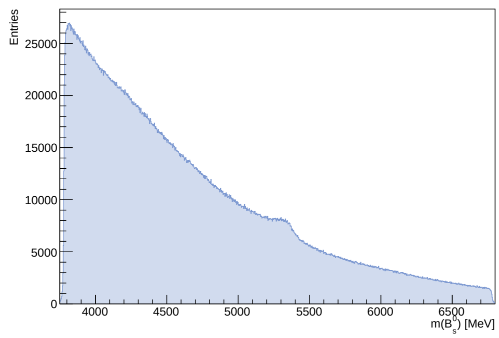
After event reconstruction, the decay candidates have to pass a first selection also called *stripping*, which selects events based on their decay topology. This reduces the dataset to only the decays with the correct final states. In this analysis, the stripping line `Bu2LLKeeLine2` is used, which imposes the requirements listed in Table 6 on the decay particles.

5.3 Trigger Configuration

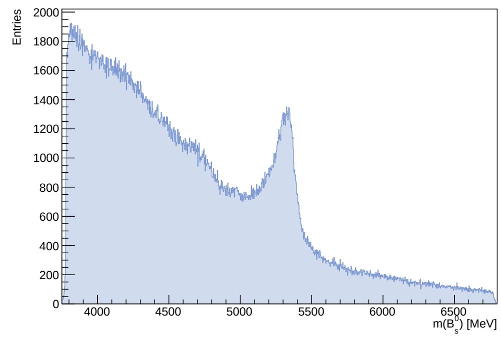
The signal candidates are additionally required to pass the trigger system (described in Section 4.5). The specific trigger lines used in this analysis are listed in Appendix A. An event is stored as *Triggered On Signal (TOS)* with respect to the fired trigger line if a final state particle of the signal candidate is responsible for firing the trigger. If an event is triggered independently of the final state particles, e.g. by another particle in the event unassociated with the signal candidate, this event is referred to as *Triggered Independent of Signal (TIS)* with respect to the trigger line. The `L0` trigger lines select events with the respective particle having a high momentum. `HLT1TrackAllL0` requires events to have triggered at least one `L0` trigger line and additionally imposes higher p_T and $\chi^2_{\mathtt{IP}}$ criteria. `HLT2`'s topological lines trigger on partially reconstructed b decays, where at least two charged daughter particles exist. Input particles that have passed preliminary cuts on the track quality are reconstructed to two- or three-body objects with the input particles required to pass a cut on the distance of closest approach [23]. Finally, a multivariate analysis (BDT in Run 1) trained on kinematic and topological variables is carried out on these objects and a cut on the MVA output is applied.

5.4 Preselection

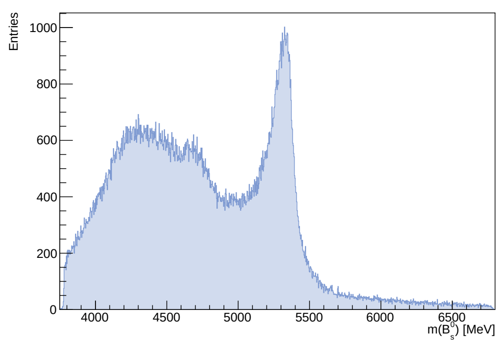
After passing the stripping selection the reconstructed B_s^0 meson mass distribution is still dominated by combinatorial background (see Figure 8(a)), e.g. the random combination of measured particle tracks, thus forming a false signal. This outlines the need for a proper signal selection starting with the application of preselection cuts. The chosen preselection cuts are outlined in Table 7. Here, advantage is taken of the narrow ϕ resonance, since this allows the placement of a tight cut around the peak location. The spectrum of the B_s^0 mass after preselection is portrayed in Figure 8(b). While there is still a large amount of combinatorial background, a peak is now visible in the mass region of 5100 - $5500\,\mathrm{MeV}$ corresponding to the dominating $B_s^0 \to J/\psi(\to e^+e^-)\phi$ decay. The mass distributions with the q^2 cuts applied for the normalization channel and the rare mode are displayed in Figure 8(c) and Figure 8(d), respectively. The mass distribution for the rare mode, especially, still shows a large amount of combinatorial background that has to be further reduced.



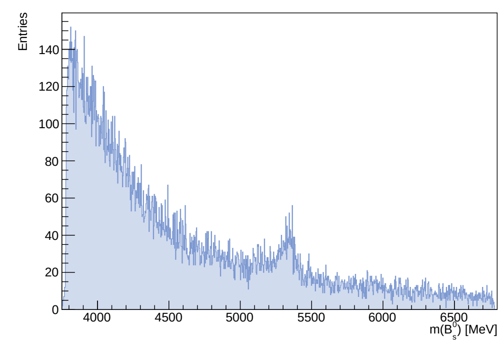
(a) B_s^0 meson mass distribution after stripping.



(b) B_s^0 meson mass distribution after preselection without additional cut on q^2.



(c) B_s^0 meson mass distribution after preselection with the q^2 cut (6-11 GeV^2) for the normalization channel applied.



(d) B_s^0 meson mass distribution after preselection with the q^2 cut (< $0.0001\,\mathrm{GeV}^2$) for the rare mode applied.

Figure 8: B_s^0 meson mass distributions before and after preselection.

variable	normalization channel	rare mode
m(ϕ)	$(1019.461 \pm 12.000)\,\mathrm{MeV}$	
$e^{\pm}$ PIDe	> 2	
$K^{\pm}$ PIDK	> 0	
q^2	6-11 GeV^2	< $0.0001\,\mathrm{GeV}^2$

Table 7: Overview of the preselection cuts and the additional q^2 cut applied to separate normalization channel and rare mode.

6 Multivariate Analysis

In order to further and more efficiently reject combinatorial background, an Extreme Gradient Boosting (XGBoost) model is trained. The model uses data with multiple correlated features x_i to give a prediction of the target variable y_i.

6.1 Introduction to Extreme Gradient Boosting (XGBoost)

The basis of the model are the Classification and Regression Trees (CART). In a tree, the training data sample is split into subsamples at every node using binary variable decisions, e.g. $((x_j)_i > 0)$. This process is repeated at every node until a stop criterion such as the maximum depth of the tree is reached. The final sub-samples are stored in leaves in the tree and are assigned a score with the real value w.
CARTs and Boosted Decision Trees (BDTs) are at risk of learning specific statistical fluctuations in the training sample resulting in a very good classification performance on the training data - but not on another independent training sample. In this case, the model has lost its predictive power since it trains on noise. The effect of overtraining is alleviated by using the method of *Boosting*. A forest of trees, all derived from the same training sample, is grown, where events that were misclassified in one tree have their weights increased and are then given to the next new tree. The total prediction of an ensemble of *K* CARTs f_k is then given by

$$\hat{y_i} = \sum_{k=1}^{K} f_k(\vec{x_i}). \tag{6.1}$$

The training is done additively. Only one tree is added at a time and the already learned trees remain fixed. Consequently, the same applies for optimizing the trees. The optimization of a tree occurs by minimizing the objective function

$$\mathcal{L} = L(\vec{\theta}) + \Omega(\vec{\theta}) \tag{6.2}$$

where $L(\vec{\theta})$ denotes the loss function measuring the difference between the prediction $\hat{y_i}$ and the real value y_i and $\Omega(\vec{\theta})$ denotes the weight w-dependent regularization term that limits the complexity of the model to avoid overfitting. $\vec{\theta}$ are the parameters of the algorithm e.g. the model trains to find the parameters $\vec{\theta}$ that best describe the data x_i

and the target variable y_i. The exact minimization process is presented in detail in [24].

6.2 Creating the Training Sample

The data sample used to train an XGBoost Model is made up of a pure signal and background sample. Since the signal of the rare decay is not yet visible on the data, the signal of the normalization channel is used as a proxy for training the BDT. An overview of the cuts applied to the total dataset from Run 1 and Run 2 in order to produce the signal and background sample is given in Table 8. For the signal sample, the cuts vary little from the preselection cuts - with the exception of the new variable cut on the B_s^0 mass distribution calculated with a constraint on the J/ψ mass and the tightened cut on the q^2 region. These last cuts were introduced to ensure a signal sample with negligible background contribution. The J/ψ mass constraint allows for a more precise B_s^0 mass reconstruction since placing this constraint corrects the momentum and bremsstrahlung radiation losses of the leptonic final state particles. In particular, the cut on $m(B_{s,J/\psi_{DTF}})$ ensures that there is no contribution from partially reconstructed events, where one or several final state particles of a B meson decay are not reconstructed.
The background sample is taken from the upper mass side band where $m(B_s^0) > 5500\,\mathrm{MeV}$. The lower mass sideband is not included in the background sample since it contains partially reconstructed events.

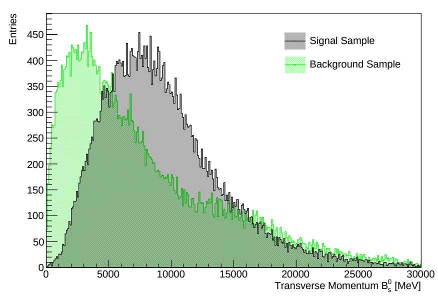
variable	signal sample	background sample
m(ϕ)	$(1019.461 \pm 12.000)\,\mathrm{MeV}$	
$e^{\pm}$ PIDe	> 2	
$K^{\pm}$ PIDK	> 0	
q^2	$(9.36 \pm 0.60)\,\mathrm{GeV}^2$	$< 8\,\mathrm{GeV}^2$
m($B_{s,J/\psi_{DTF}}$)	5300 - 5450 MeV	-
m(B_s^0)	-	> 5500 MeV

Table 8: Overview of the cuts applied to render the signal and background sample for the multivariate analysis. "-" means no cuts were applied on the variable for this sample.

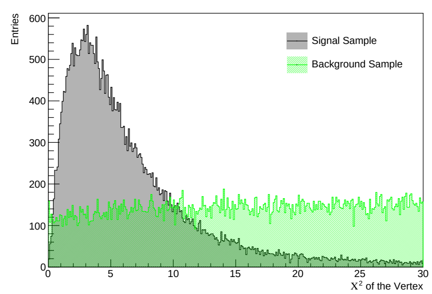
6.3 BDT Input Features

Besides a signal and a background sample, the BDT needs variables to train on. In order to achieve a good suppression of background, those variables should show different be-

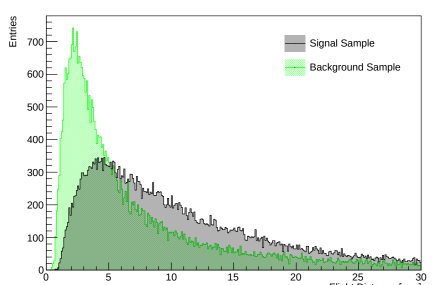
haviour in their signal and background distributions.
Typically this is to be expected for vertex quality and impact parameter variables. B_s^0 mesons are produced at the primary vertex, which means the B_s^0 impact parameter and its χ^2 value should be small for signal events. The opposite is the case for the kaons and electrons coming from the detached B_s^0 vertices. The expected behaviour of high impact parameter and high $\chi^2_{\mathtt{IP}}$ values for the signal sample is shown in Figure 9(d). The cut-off for lower $\chi^2_{\mathtt{IP}}$ values is a result of the stripping selection.
Another category of variables with separated signal and background distributions are kinematic variables based on momentum information. As is visible in Figure 9(a), events from heavy particles like the B_s^0 meson have higher transverse momenta than events from the upper mass sideband. All variables used in the training of the BDT have similar well-distinguished signal and background distributions (see Appendix B.1) and are listed in Table 9.



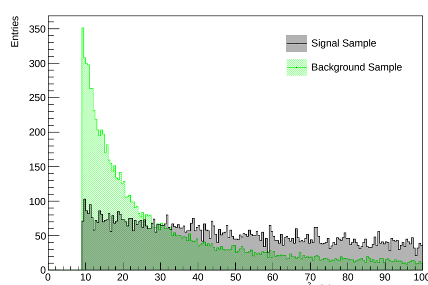
(a) Transverse Momentum of the B_s^0 meson.



(b) χ^2 of the B_s^0 decay vertex fit.



(c) Flight Distance of the B_s^0 meson.



(d) χ^2 of the Impact Parameter of the electron E1.

Figure 9: Signal and Background Distributions for select variables used in training.

Impact Parameter χ^2 ($\chi^2_{\mathtt{IP}}$)	p_T	Vertex χ^2	Flight Distance (FD)	FD χ^2 ($\chi^2_{\mathtt{FD}}$)
B_s^0	B_s^0	B_s^0	B_s^0	B_s^0
$K^\pm$	ϕ			
$e^\pm$				

Table 9: Training variables used in the multivariate analysis.

6.4 k-Folding

For proper training and validation, three statistically independent data samples are needed. One data sample is used for parameter optimization (training), one for performance validation (testing), and one for application (overtraining detection) [25]. For implementation on the dataset defined in Section 6.2 a k-Fold approach is implemented. The dataset is divided into $k = 10$ sub-samples (also referred to as folds) of approximately the same size. For training, 8 out of those 10 folds are used, 1 fold is used for testing, and 1 for application. The process of training, testing, and application is repeated k-times by rotating the folds making up the three needed samples so that every sample serves once as a test and once as an application sample.
The output of the BDT Classifier is a continuous spectrum from 0 to 1 (see Figure 10) for all events, where 0 signifies background-like and 1 signifies signal-like. The cut for suppression of the combinatorial background is placed on the BDT Classifier output; the process of finding the optimal cut will be discussed in Section 6.6.

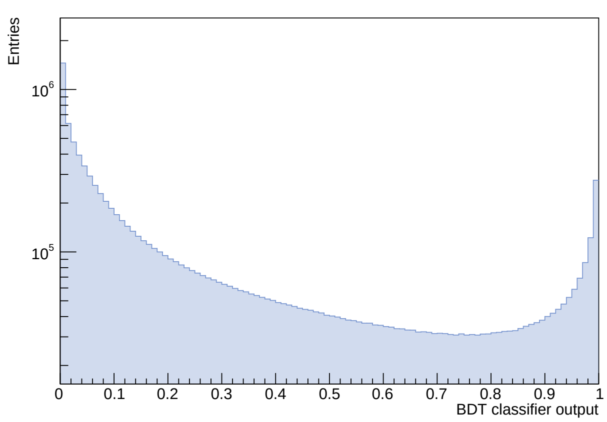
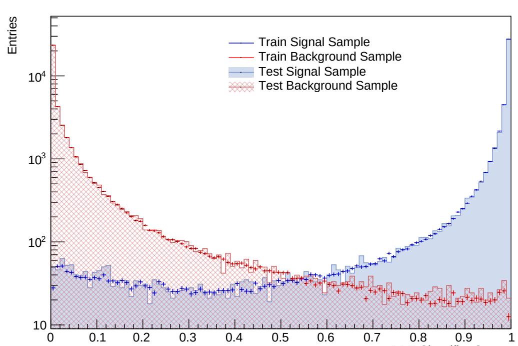


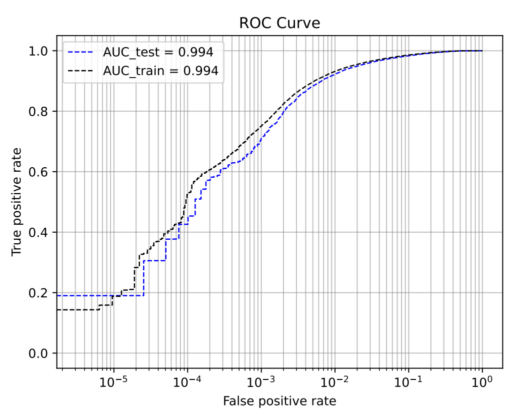
Figure 10: Plot of the BDT Classifier Output for Run 1 and Run 2 Data.

6.5 Evaluating Model Performance

To evaluate the BDT's performance and check for possible overtraining, the *Receiver Operating Characteristic* (ROC) curve, a performance measurement for classification problems, is used (see Figure 11(b)). The curve plots the false positive rate (also 1-specificity), e.g. the ratio of events that were wrongly classified as signal over all background events, against the true positive rate (also called sensitivity), e.g. the ratio of events that were correctly classified as signal over all signal events. The area under the ROC curve (AUC) is a measure of separability and gives an estimate of how good the model is at distinguishing between two classes. At an AUC of 1, a model separates perfectly between signal and background [26].
While the ROC curve of the test sample at times lies slightly below the ROC curve of the train sample, there is not too much loss in the model's prediction power as can be seen by the very good AUC score of 0.994, which is the same for train and test sample. Similar conclusions can be drawn from comparing the signal and background distributions of the train sample (markers with errorbars) and the test sample in Figure 11(a).



(a) BDT Classifier Output showing signal and background distribution for train and test sample.



(b) ROC Curve.

Figure 11: BDT Classifier Output for the XGBoost Model and its ROC Curve.

6.6 Optimization of the BDT Cut

To further reduce the combinatorial background a selection cut is placed on the output of the BDT for Run 1 and Run 2 data. The optimal BDT cut is first determined for the normalization channel and later, using the upper mass sideband for the B_s^0 mass, also for the rare mode. This is done by maximizing the Figure-of-Merit (FoM)

$$\mathrm{FoM} = \frac{S}{\sqrt{S+B}} \tag{6.3}$$

with S being the signal yield and B the background yield.

Normalization Channel: For the normalization channel, which only takes events in a q^2 range of 6 - $11\,\mathrm{GeV}^2$ into consideration, the overall signal and background yields are estimated by fitting the B_s^0 mass distribution in the range of 4.6 - $6.2\,\mathrm{GeV}$. The signal and background yields used for the calculation of the FoM are extrapolated from the 2-σ region around the location of the peak. The fit model of a double-sided Crystalball, two Gaussians, and an exponential term used to describe the invariant mass distribution of the B_s^0 meson is explained in detail in Section 7.3.1.

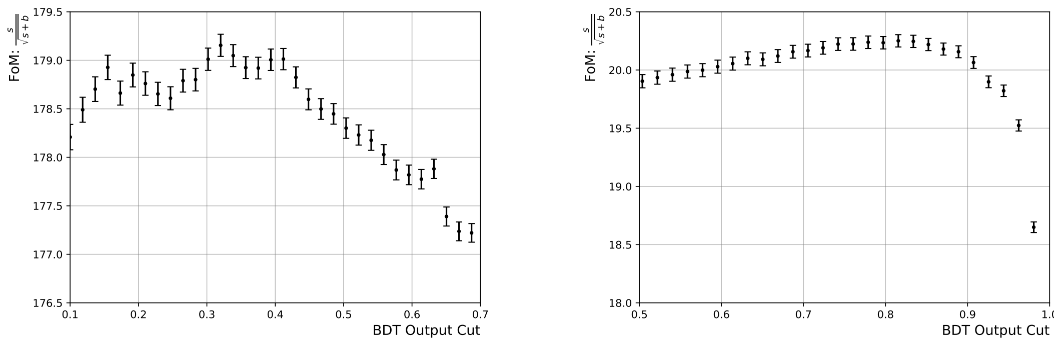
Rare Mode: The expected signal yield on the rare mode is estimated by scaling the yield of the normalization channel with the relative branching fraction of $B_s^0 \to \phi\gamma(\to e^+e^-)$ and $B_s^0 \to J/\psi(\to e^+e^-)\phi$:

$$S_{\phi\gamma} = S_{J/\psi\phi} \times \frac{\mathcal{B}(B_s^0 \to \phi\gamma)}{\mathcal{B}(B_s^0 \to J/\psi\phi) \cdot \mathcal{B}(J/\psi \to e^+e^-)} \times \frac{\epsilon_{\phi\gamma}^{MC}}{\epsilon_{J/\psi\phi}^{MC}}. \tag{6.4}$$

The efficiencies are estimations based on the Monte Carlo simulations. A more accurate calculation, taking every selection cut into account, is presented in Section 8.1. For the background yield, the number of events in the upper mass sideband (m(B_s^0) > $5.5\,\mathrm{GeV}$) independent of the BDT output cut in the normalization channel ($N_{J/\psi\phi}$) and in the rare mode ($N_{\phi\gamma}$) is determined. The background yield on the rare mode is then calculated by

$$B_{\phi\gamma} = B_{J/\psi\phi} \times \frac{N_{\phi\gamma}(m(B_s^0) > 5.5\,\mathrm{GeV})}{N_{J/\psi\phi}(m(B_s^0) > 5.5\,\mathrm{GeV})}. \tag{6.5}$$

Figure 12 shows the Figure-of-Merit in dependency of the BDT output cut for the normalization channel and the rare mode.



(a) Normalization Channel.

(b) Rare Mode.

Figure 12: Figure-of-Merit (FoM) in dependency of the cut on the BDT output for both normalization channel and rare mode.

In the normalization channel, the distribution exhibits a visible decrease in the value of the FoM for higher BDT output cuts. This decrease is shifted to larger BDT cut values in the rare mode. In Figure 12(b) only the statistical uncertainties on the Figure-of-Merit are displayed. However, the measurement is also affected by the uncertainties on the branching fractions and the efficiencies which are not shown since they correlate for each measurement. The optimal BDT cuts are chosen to be the cuts with a local maximum in the distribution. This is 0.32 for the normalization channel and 0.82 for the rare mode.

7 Extraction of Signal Yields

This section documents the determination of the signal yields on both rare mode and normalization channel via Unbinned Maximum Likelihood fits.

7.1 Unbinned Maximum Likelihood Method

All fits in this thesis are implemented using the Unbinned Maximum Likelihood Method. The probability of describing a set of n independently distributed observations $x_1 ... x_n$ with a set of unknown parameters $a_1 ... a_m$is given by the Likelihood Function

$$\mathcal{L} = \prod_{i=0}^{n} p(x_i; a_1 ... a_m). \tag{7.1}$$

Here, p is the known probability density function (*pdf*) normalized to unity: $\int p(x; a_1 ... a_m)dx = 1$. The optimal parameters $\hat{a}_1 ... \hat{a}_m$ are estimated by maximizing the Likelihood [27]. In order to avoid numerical problems during the calculation, the negative Log Likelihood

$$-\ln \mathcal{L} = -\sum_{i=1}^{n} \ln p(x_i; a_1 ... a_m). \tag{7.2}$$

is minimized. The uncertainty on the estimate of a parameter is given by the square root of its corresponding diagonal element in the covariance matrix. The covariance matrix, in turn, is computed by taking the inverse of the Hessian matrix which is the matrix of the second derivatives of the objective function, the negative Log Likelihood, evaluated at the parameters estimate $\hat{a}_i$. The required minimization of the negative Log Likelihood and the subsequent uncertainty calculation is carried out by Python's `iMinuit` package.

7.2 ${}_s\mathcal{P}lot$ Method

The ${}_s\mathcal{P}lot$ method is a statistical tool for analyzing datasets consisting of different sources of events (i.e. signal and background). This technique uses a so-called *discriminating* variable to unfold the distribution of the so-called *control* variables with respect to the different event sources. The term *discriminating* refers to a variable for which both signal and background distributions are known. In order to carry out the ${}_s\mathcal{P}lot$ method, the control variable has to be uncorrelated with the discriminating variable. To obtain knowledge about the event types in the discriminating variable - here the invariant mass of the B_s^0 meson - a Maximum Likelihood Fit for the event yields is performed. The result of the ${}_s\mathcal{P}lot$ technique, the ${}_s$*Weights*, are then calculated by

$$ {}_s\mathcal{P}(m) = \frac{V_{ss}P_s(m) + V_{sb}P_b(m)}{N_sP_s(m) + N_bP_b(m)}, \tag{7.3}$$

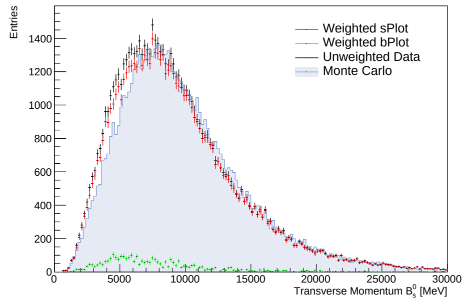
where N_i denotes the respective yield, $P_i(m)$ the respective pdf describing the two event sources, and m is the discriminating variable. V_{ss} and V_{sb} are elements of the covariance matrix and are determined by inverting the inverted covariance matrix where the elements are given by

$$V_{ij}^{-1} = \sum_m \frac{P_i(m)P_j(m)}{(N_sP_s(m) + N_bP_b(m))^2}. \tag{7.4}$$

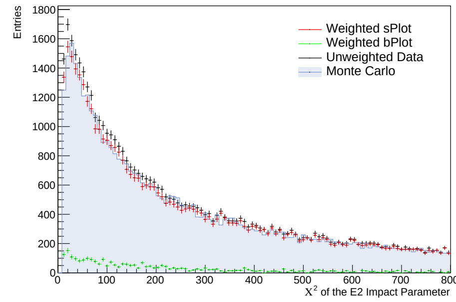
A more generalized description can be found in [28]. ${}_s\mathcal{P}lots$ can be utilized as a means to test the agreement between signal distributions on data and on MC simulation. For this, the input variables of the multivariate analysis chosen in Section 6.3 are used as control variables. The signal and background yields required for the calculation of the weights are extracted using the fit procedure described in Section 7.3.1. Signal and background distribution were obtained by using ${}_s$*Weights* and ${}_b$*Weights*, respectively, and along with the simulated events and the unweighted data of each variable are shown in Figure 13 and Appendix B.2. For the kinematic variables of the B_s^0 and the ϕ meson, deviations of the simulated events to the s-weighted data can be observed. This is a known phenomenon that most likely occurs because the B meson production is not modeled perfectly. The impact parameters, especially for the electrons and kaons, along with the fit quality of the flight distance only show very small differences. While these deviations do not affect the training of the BDT due to the utilization of data, they might influence the efficiency calculation.

7.3 Signal Fit on MC and Data

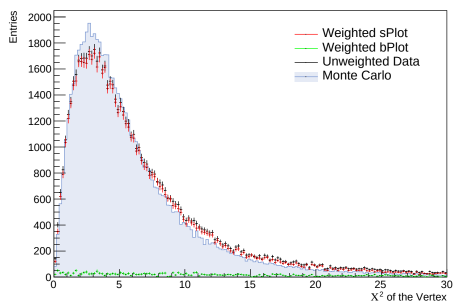
The fit analysis uses both the Monte Carlo simulated events and the datasets. The Monte Carlo simulations are utilized to determine the majority of the fit model parameters,



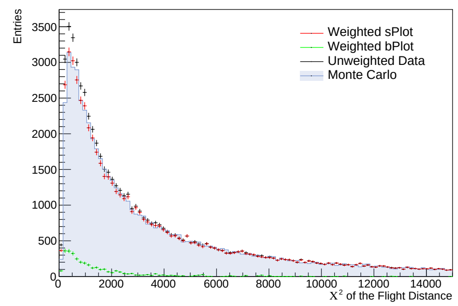
(a) Transverse Momentum p_T of the B_s^0 meson.



(b) χ^2 of the E2 Impact Parameter.



(c) χ^2 of the B_s^0 decay vertex fit.



(d) χ^2 of the flight distance of the B_s^0 meson.

Figure 13: Select BDT variable distributions from MC simulations (blue), unweighted data (black), data weighted with $_s$Weights (red) and data weighted with $_b$Weights.

which are then applied to the data to fit the signal yields for both the normalization channel and the rare mode. The model from the MC 2012 is transferred to data from 2011 and 2012, the model from the MC 2016 to data from 2015 and 2016, and the model from the MC 2018 to data from 2017 and 2018. This is done to take changes in the detector configuration and in the center-of-mass energies into account. It is important to note that the simulated events have to pass the same stripping, preselection, and BDT classifier output selection as the data.

In addition, the MC simulated and the dataset events are further separated into three subsets depending on the number of reconstructed bremsstrahlung photons (0, 1, or 2). This is necessary since the emission of photons by bremsstrahlung leads to smearing in the mass distribution, resulting in each subset having a slightly different mass shape. Most noticeably, the distribution is less asymmetric for a larger amount of recovered bremsstrahlung photons. The fit procedure as described in Section 7.3.1 and 7.3.2 is inspired by the analysis in [29].

7.3.1 Signal Fit on Normalization Channel

On both Monte Carlo simulated events and data of the normalization channel an additional cut on the J/ψ mass constrained B_s^0 mass of $m(B^0_{s,J/\psi_{DTF}}) > 5200\,\mathrm{MeV}$ is applied, thereby reducing the lower mass tail through the exclusion of partially-reconstructed background. The normalization channel is selected by cutting on q^2 in the range of 6 to $11\,\mathrm{GeV}^2$.
A central aspect of all performed mass fits is the double-sided Crystalball[2] (DSCB) function which models the experimental resolution of the mass distribution taking the detector resolution into account [30]. It is composed of a Gaussian core distribution and two power-law tails to either side of the core:

$$f(x;\beta_l,m_l,\beta_h,m_h,\mu,\sigma) = \mathcal{N}\cdot\begin{cases} A_l\cdot(B_l - \frac{x-\mu}{\sigma})^{-m_l}, & \text{for } \frac{x-\mu}{\sigma} \leq -\beta_l \\ \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), & \text{for } -\beta_l < \frac{x-\mu}{\sigma} < \beta_r \\ A_h\cdot(B_h + \frac{x-\mu}{\sigma})^{-m_h}, & \text{for } \frac{x-\mu}{\sigma} \geq \beta_r \end{cases} \tag{7.5}$$

with

$$A_{l,h} = \left(\frac{m_{l,h}}{\beta_{l,h}}\right)\cdot\exp\left(-\frac{|\beta|^2}{2}\right), \tag{7.6}$$

$$B_{l,h} = \frac{m_{l,h}}{|\beta_{l,h}|} - |\beta_{l,h}|. \tag{7.7}$$

$\mathcal{N}$ is a numerically determined normalization factor, that normalizes the DSCB in the fit range and is simply the inverse of the integrated DSCB in the fit range. The parameter μ marks the location of the peak of the mass distribution and σ gives the width of the Gaussian distribution. For the tails, m_l and m_h denote the exponents of the power-law tails to the left and the right side, respectively, while β_l and β_h determine the mass value where the Gaussian core distribution ends and the respective power-law tail sets in. The left tail models the energy loss due to bremsstrahlung photons produced by the electrons. The right tail describes the overestimation of reconstructed photon energies by the detector.

[2] named after the Crystal Ball Collaboration, which first introduced this function

The fit model for the signal is completed by a Gaussian on the left side of the Gaussian core distribution as well as an additional Gaussian on the right side that help model the mass distribution tails. The right-handed Gaussian is only implemented for the subsets containing events with one or two reconstructed photons as the right tail is more pronounced for these datasets. The mass distribution of the simulated events in the fit range $M(B_s^0) \in [4.5, 5.9]\mathrm{GeV}$ is thus described by

$$\begin{aligned} F_s = & (1 - f_{gr} - f_{gl} \cdot (1\text{-}f_{gr})) \cdot \mathtt{DSCB}(M; \beta_l, m_l, \beta_h, m_h, \mu, \sigma) \\ & + f_{gl} \cdot (1 - f_{gr}) \cdot \text{Gauss}(M; \mu - \mu_{l,\text{shift}}, \sigma_{gl}) + f_{gr} \cdot \text{Gauss}(M; \mu + \mu_{r,\text{shift}}, \sigma_{gr}), \end{aligned} \tag{7.8}$$

where M is the mass, f_{gl} (f_{gr}) is the fraction of the left(right)-handed Gaussian, σ_{gl} (σ_{gr}) its width and $\mu_{l,\text{shift}}$ ($\mu_{r,\text{shift}}$) the difference between the peak position of the Gaussian and the core distribution. The fit results on the individual subsets of the Monte Carlo simulated events are displayed in Appendix C.1. The datasets containing one or two reconstructed photons are described well by the chosen fit model. Though there are slight deviations visible at the beginning of the right tail, these are not significant. However, these differences grow more pronounced for the datasets without a reconstructed photon and therefore overestimate the data of Run 2 part 2 significantly in the mass region around $5.4\,\mathrm{GeV}$. Nevertheless, this deviation is expected to only have a small effect on the determination of the signal yields.

For modeling the signal on data, the fit parameters derived from the Monte Carlo simulation are either fixed or, in the case of the tail parameters, constrained by adding a Gaussian penalty term. A mass shift with the mass peak on data then described by $M_{\text{Data}} = M_{\text{MC}} + M_{\text{shift}}$ is applied to the peak position and the width of the core distribution is scaled by p_{scale}: $\sigma_{\text{Data}} = \sigma_{\text{MC}} \cdot p_{\text{scale}}$. Additionally, the fraction of the right-handed Gaussian is scaled with a correction term f_g: $f_{\text{gr, Data}} = f_{\text{gr, MC}} \cdot f_g$.

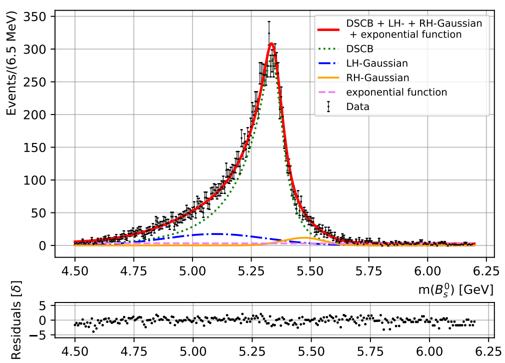
The combinatorial background is modeled by a simple exponential function, where λ is the decay constant and $\mathcal{N}_{\text{exp}}$ the corresponding normalization factor:

$$F_b = \frac{1}{\mathcal{N}_{\text{exp}}} \cdot \exp\left(-\lambda \cdot M\right). \tag{7.9}$$

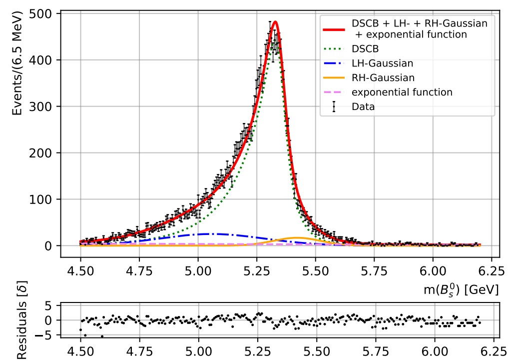
Thus, the complete fit model for both signal and background from 4.5 to $6.2\,\mathrm{GeV}$ is defined as

$$F_{J/\psi\phi} = \frac{N_s}{N_s + N_b} \cdot F_s(M; \beta_l, m_l, \beta_h, m_h, m_{\text{shift}}, p_{\text{scale}}, f_g) + \frac{N_b}{N_s + N_b} \cdot F_b(M; \lambda). \tag{7.10}$$

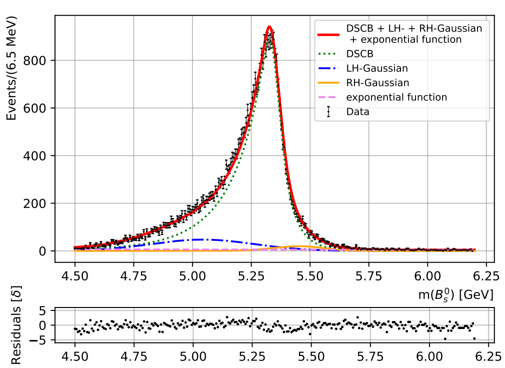
N_s and N_b denote the signal and background yield respectively. For the subset with 0 reconstructed photons, the right-handed Gaussian and its correction term on data is omitted. Each individual fit on data can be found in Appendix C.2. Interestingly enough, the previously observed overestimation of the peak on the Monte Carlo simulation and the subsequent decrease in the right tail on the simulated events have significantly decreased in the case of the Run 2 part 2 sample and is hardly visible for Run 1 and Run 2 part 1 data.



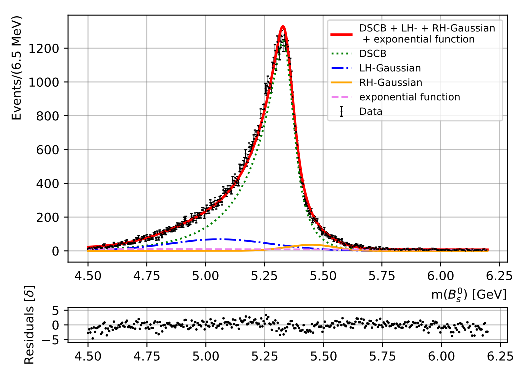
(a) 2011 and 2012 data.



(b) 2015 and 2016 data.



(c) 2017 and 2018 data.



(d) Total data.

Figure 14: Mass Fits on normalization channel for 2011/12, 2015/16, 2017/18 Data and total mass fit on all data.

The mass fits for all three time periods (see Figure 14) are obtained by the addition of the separate fits of the subsets split by the amount of reconstructed photons. The fit on all available data is then the addition of the fits of the three time periods. In Table 10 the extrapolated signal and background yields are listed for the three time periods and for the complete dataset.

7.3.2 Signal Fit on Rare Mode

For the fits on the rare mode, only events within the q^2-region of $q^2 < 0.0001\,\mathrm{GeV}^2$ are considered. Due to the negligible contribution of events with two reconstructed photons, the data was only split into two subsets with either no or a single reconstructed photon. The low amount of events with two reconstructed photons is explained by the near collinear production of the positron and electron in $\gamma \to e^+e^-$ conversions and the subsequent emission of bremsstrahlung in the same direction. Consequently, both photons produce a hit in the same ECAL cluster and are reconstructed as a single photon.
In the rare mode, the signal fit consists solely of a double-sided Crystalball function. The tail parameters $\beta_l, m_l, \beta_h, m_h$, the peak position μ and the distribution width σ are determined using Monte Carlo simulated events in the mass range of $M(B_s^0) \in [4.7, 5.75]\mathrm{GeV}$. All fits on simulated events can be seen in Figure 24 in Appendix D.1. Except for outliers due to the low statistics, the fit models the simulated signal events well. The fit procedure on data of the rare mode is quite similar to that of the normalization channel: The tail parameters are constrained using a Gaussian penalty term in the negative Log-Likelihood function and a shift parameter M_{shift} is added for the mass. The scaling of the width of the Gaussian distribution is calculated using the results from the normalization channel:

$$\sigma_{\phi\gamma}^{\text{Data}} = \frac{\sigma_{\phi\gamma}^{Data}}{\sigma_{J/\psi\phi}^{MC}} \cdot \sigma_{J/\psi\phi}^{Data}. \tag{7.11}$$

The same exponential term is fit to the combinatorial background. The total fit model applied on the data in a mass range of 4.9 to $6.2\,\mathrm{GeV}$ is simplified to

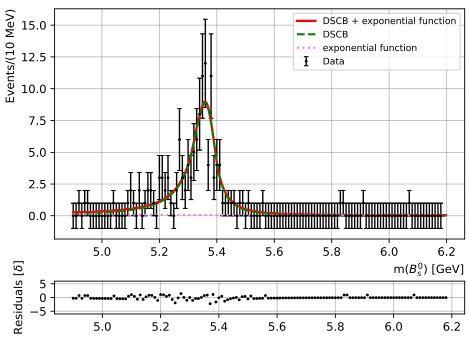
$$F_{\phi\gamma} = \frac{N_s}{N_s + N_b} \cdot \mathtt{DSCB}(M; \beta_l, m_l, \beta_h, m_h, M_{\text{MC}} + M_{\text{shift}}, \sigma_{\phi\gamma}^{\text{Data}}) + \frac{N_b}{N_s + N_b} \cdot F_b(M, \lambda). \tag{7.12}$$

The data fits are shown in Appendix D.2, the total fits for the individual time periods independent of the amount of reconstructed photons in Figure 15. Compared to the normalization channel, the fraction of background events in the signal range is higher.

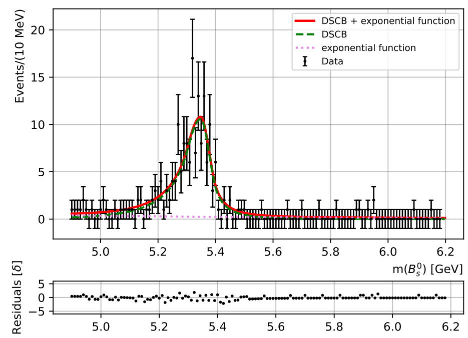
dataset	decay	N_{signal}	$N_{\text{background}}$
Run 1	$B_s^0 \to \phi\gamma$	120 ± 11	9 ± 5
	$B_s^0 \to J/\psi\phi$	$11\,985 \pm 115$	810 ± 46
Run 2, part 1	$B_s^0 \to \phi\gamma$	160 ± 14	29 ± 8
	$B_s^0 \to J/\psi\phi$	$18\,879 \pm 143$	786 ± 50
Run 2, part 2	$B_s^0 \to \phi\gamma$	356 ± 20	26 ± 10
	$B_s^0 \to J/\psi\phi$	$36\,664 \pm 198$	1403 ± 63
$\sum$	$B_s^0 \to \phi\gamma$	636 ± 27	64 ± 14
	$B_s^0 \to J/\psi\phi$	$67\,528 \pm 270$	2999 ± 92

Table 10: Signal and background yields for all datasets for normalization channel and rare mode. The uncertainties are the uncertainties on the fit results.

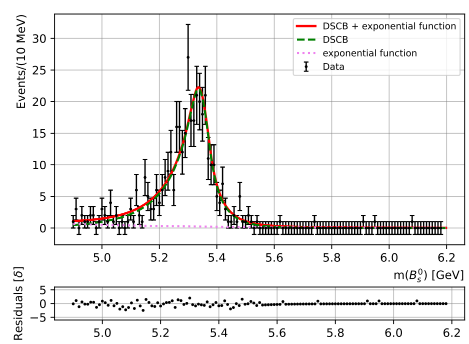
Nevertheless, the description of the mass distribution of the data via the fit model is similar to the description of the simulated events. Here, low statistics play their part in the good agreement between data and fit. The signal and background yields of the rare mode are listed together with the results of the normalization channel in Table 10.



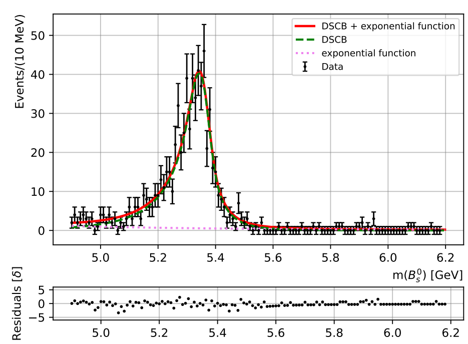
(a) 2011 and 2012 data.



(b) 2015 and 2016 data.



(c) 2017 and 2018 data.



(d) Total data.

Figure 15: Mass Fits on rare mode for 2011/12, 2015/16, 2017/18 Data and total mass fit on all data.

8 Determination of the Branching Fraction

This section covers the calculation of the branching fraction $B_s^0 \to \phi\gamma$ using the signal yields from Section 7.3 and the signal selection efficiencies, which have yet to be determined. The relative branching fraction is obtained first and with the help of the normalization channel $B_s^0 \to J/\psi(\to e^+e^-)\phi$ the total branching fraction is calculated.

8.1 Signal Efficiencies

In order to obtain a clean sample on the data and determine the signal yields from this clean sample, selection cuts are applied. The signal efficiency describes the amount of signal events that are reconstructed and selected relative to the total number of signal decays initially produced. It is convenient to obtain the efficiency from Monte Carlo simulations where the overall number of signal events is simply the number of simulated events. The efficiencies for the rare decay and the normalization channel are calculated by

$$\epsilon_{\phi\gamma} = \frac{N_{sel.}^{MC}(B_s^0 \to \phi\gamma(\to e^+e^-))}{N_{sim.}^{MC}(B_s^0 \to \phi\gamma(\to e^+e^-))} \tag{8.1}$$

and

$$\epsilon_{J/\psi\phi} = \frac{N_{sel.}^{MC}(B_s^0 \to J/\psi(\to e^+e^-)\phi)}{N_{sim.}^{MC}(B_s^0 \to J/\psi(\to e^+e^-)\phi)} \tag{8.2}$$

where $N_{sel.}^{MC}$ stands for the number of events that have passed reconstruction and selection and $N_{sim.}^{MC}$ for the simulated number of events in the MC simulation. The results are listed in Table 11 with the central values blinded as discussed in Section 3.

	Decay Channel	efficiency
2012	$B_s^0 \to \phi\gamma$	$(xxx \pm 0.33) \times 10^{-5}$
	$B_s^0 \to J/\psi\phi$	$(xxxx \pm 0.015) \times 10^{-3}$
2016	$B_s^0 \to \phi\gamma$	$(xxx \pm 0.24) \times 10^{-5}$
	$B_s^0 \to J/\psi\phi$	$(xxxx \pm 0.014) \times 10^{-3}$
2018	$B_s^0 \to \phi\gamma$	$(xxx \pm 0.25) \times 10^{-5}$
	$B_s^0 \to J/\psi\phi$	$(xxxx \pm 0.010) \times 10^{-3}$

Table 11: Calculated efficiencies for rare decay and normalization channel for all three Monte Carlo simulations (blinded).

8.2 Branching Fraction Fit

With the signal yields determined in Section 7.3 and the derived efficiencies from Table 11, the relative branching fraction of $B_s^0 \to \phi\gamma$ to $B_s^0 \to J/\psi(\to e^+e^-)\phi$ is expressed as

$$\frac{\mathcal{B}(B_s^0 \to \phi\gamma)}{\mathcal{B}(B_s^0 \to J/\psi(\to e^+e^-)\phi)} = \frac{N(B_s^0 \to \phi\gamma)}{N(B_s^0 \to J/\psi(\to e^+e^-)\phi)} \times \frac{\epsilon_{J/\psi\phi}}{\epsilon_{\phi\gamma}}. \tag{8.3}$$

In order to combine the three individual measurements of the relative branching fraction obtained from Run 1, Run 2, part 1, and Run 2, part 2, an Unbinned Maximum Likelihood Fit is performed using

$$pdf = \sum_{i=1}^{3} \frac{1}{\sqrt{2\pi\sigma_i^2}} \cdot \exp\left(-\frac{(\mu - x_i^2)}{2\sigma_i^2}\right), \tag{8.4}$$

where x_i denotes the relative branching fraction and σ_i the respective error on the relative branching fraction in the respective data sub-sample. The fit returns a relative branching fraction of

$$\frac{\mathcal{B}(B_s^0 \to \phi\gamma)}{\mathcal{B}(B_s^0 \to J/\psi(\to e^+e^-)\phi)} = (xxx \pm 0.28) \times 10^{-1}.$$

The total branching fraction of $B_s^0 \to \phi\gamma$ can then be determined by multiplying the relative branching fraction with the total branching fraction of $B_s^0 \to J/\psi(\to e^+e^-)\phi$. For this, the values of the branching fractions of $B_s^0 \to J/\psi\phi$ and $J/\psi \to e^+e^-$are needed [2].

$$\mathcal{B}(B_s^0 \to J/\psi\phi) = (1.04 \pm 0.04) \times 10^{-3}$$
$$\mathcal{B}(J/\psi \to e^+e^-) = (5.971 \pm 0.032) \times 10^{-2}$$

The total branching fraction of $B_s^0 \to \phi\gamma$ is calculated to:

$$\mathcal{B}(B_s^0 \to \phi\gamma) = (xxx \pm 0.17_{stat.}) \times 10^{-5}.$$

The high statistics of Run 1 and Run 2 combined are reflected in the low statistical uncertainty. With 0.17×10^{-5} it is competitive with previously measured uncertainties by Belle (0.5×10^{-5}) and by LHCb (0.34×10^{-5}) using $1\,\mathrm{fb}^{-1}$ of Run 1 data. A more thorough analysis of the complete LHCb data may be able to further improve the results.

8.3 Systematic Uncertainties

The branching fraction is subject to not only statistical errors but also systematics. However, the time frame of this thesis does not allow for an in-depth analysis of systematic uncertainties. As such, the main focus lies on the uncertainty stemming from the total branching fraction measurement of the normalization channel $B_s^0 \to J/\psi(\to e^+e^-)\phi$. The inclusion of this error returns a systematic uncertainty of 0.14×10^{-5}.
Since other sources of systematics are not taken into account, this represents an underestimation of the total systematic uncertainties. Other sources in this analysis would include the description of the invariant mass distribution via the chosen fit model. Alternative fit models such as a polynomial function for the combinatorial background or a Johnson SU distribution for the signal could provide an estimate for this uncertainty. Additionally, a study of potential background processes could describe the systematic uncertainty due to background pollution in the signal distribution. Another systematic uncertainty comes from the efficiency calculation on the uncorrected Monte Carlo simulations that include the preselection, signal selection, and trigger efficiencies. These efficiencies could be determined individually. The trigger efficiency can be calculated from data using the so-called TISTOS method [31]. Finally, the Monte Carlo simulations could be further corrected for the distinct differences between data and simulation as seen in Section 7.2.

9 Conclusion and Outlook

A measurement of the branching fraction of the rare radiative decay $B_s^0 \to \phi\gamma$ relative to the tree-level decay $B_s^0 \to J/\psi\phi$, where $J/\psi \to e^+e^-$, is presented. This rare decay is strongly suppressed in the Standard Model and is therefore a sensitive probe for New Physics. Data taken by the LHCb experiment in the years 2011 at $\sqrt{s} = 7\,\mathrm{TeV}$, 2012 at $\sqrt{s} = 8\,\mathrm{TeV}$ and 2015 to 2018 at $\sqrt{s} = 13\,\mathrm{TeV}$ corresponding to a total integrated luminosity of $\mathcal{L}_{int} = 9\,\mathrm{fb}^{-1}$ is used. Since the full available dataset is utilized for this preliminary analysis, the results for the efficiencies and the overall branching fraction are blinded.

To isolate the signal of the rare decay and suppress the vast combinatorial background, two stages of selection cuts, a loose preselection, and a multivariate analysis, using normalization channel data as a proxy for the signal, have been implemented. Monte Carlo simulated events are used to determine the selection efficiencies and to obtain a fit model for the B_s^0 mass distribution. On data, the signal yields for both the rare mode and the normalization channel are obtained via unbinned maximum likelihood fits. A total of 636 ± 27 $B_s^0 \to \phi\gamma$ decays are reconstructed. The total branching fraction is measured to

$$\mathcal{B}(B_s^0 \to \phi\gamma) = \left(xxx \pm 0.17_{stat.} \pm 0.14_{sys.}\right) \times 10^{-5}.$$

The statistical uncertainty is comparable to the uncertainties determined in previous studies of this decay, by Belle and, using $1\,\mathrm{fb}^{-1}$ of data, also by LHCb. As systematic uncertainty, only the uncertainty on the branching fraction of the normalization channel is considered leading to an underestimation of the overall systematic uncertainty. Therefore, a full study of systematic uncertainties is necessary but beyond the scope of this thesis. Additionally, a study of potential background processes, e.g. partially reconstructed background, with which the fit model and the selection process could be improved would be advantageous. Using the ${}_s\mathcal{P}lot$ technique and plotting the variables used for training the BDT, some deviations between Monte Carlo simulated events and signal distribution are observed. Correcting the Monte Carlo simulations, especially in regard to B_s^0 meson and kinematics-related variables, could improve the selection efficiency determination and potentially the fit parameters derived from simulated events.

List of Figures

List of Tables

References

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

[2] Particle Data Group et al. In: *Prog.Theor.Exp.Phys.* (2022). 083C01 (2022) and 2023 update. URL: https://pdglive.lbl.gov/ (visited on 08/16/2023).

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

[4] Particle Data Group et al. In: *Progress of Theoretical and Experimental Physics* 2022.8 (Aug. 2022), p. 083C01. URL: https://pdg.lbl.gov/2022/web/viewer.html?file=../reviews/rpp2022-rev-ckm-matrix.pdf.

[5] Jure Zupan. "Introduction to flavour physics". In: *CERN Yellow Rep. School Proc.* 6 (2019). Ed. by M. Mulders and C. Duhr, pp. 181–212. DOI: 10.23730/CYRSP-2019-006.181. arXiv: 1903.05062 [hep-ph].

[6] Jean Wicht. "Study of the radiative penguin $B_s^0 \to \phi\gamma$ and $B_s^0 \to \gamma\gamma$ decays and the resonant $B^+ \to K^+\gamma\gamma$ decays at Belle". In: (2008), p. 149. DOI: https://doi.org/10.5075/epfl-thesis-4057.

[7] R Aaij et al. "Measurement of the ratio of branching fractions $BR(B_0 \to K^{\star 0}\gamma)/BR(B_{s0} \to \phi\gamma)$ and the direct CP asymmetry in $B_0 \to K^{\star 0}\gamma$". In: *Nucl. Phys. B* 867 (2013), pp. 1–18. DOI: 10.1016/j.nuclphysb.2012.09.013. arXiv: 1209.0313 [hep-ex].

[8] A. Ali, B.D. Pecjak, and C. Greub. "Towards B→Vγ decays at NNLO in SCET". In: *The European Physical Journal C* 55.4 (May 2008), pp. 577–595. DOI: 10.1140/epjc/s10052-008-0623-5.

[9] I. Belyaev et al. "The history of LHCb". In: *The European Physical Journal H* 46.1 (Mar. 2021). DOI: 10.1140/epjh/s13129-021-00002-z.

[10] Christian Elsasser. *$\bar{b}b$ production angle plots*. URL: https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb%5C_ProductionAngles.html (visited on 08/15/2023).

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

[12] R. Aaij et al. *The LHCb upgrade I*. May 2023. arXiv: `2305.10515 [hep-ex]`.

[13] R Aaij et al. "Performance of the LHCb Vertex Locator". In: *Journal of Instrumentation* 9.09 (Sept. 2014), P09007–P09007. DOI: `10.1088/1748-0221/9/09/p09007`.

[14] R. Arink et al. "Performance of the LHCb Outer Tracker". In: *Journal of Instrumentation* 9.01 (Jan. 2014), P01002–P01002. DOI: `10.1088/1748-0221/9/01/p01002`.

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

[16] C. Abellán Beteta et al. *Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC*. 2020. arXiv: `2008.11556 [physics.ins-det]`.

[17] A A Alves et al. "Performance of the LHCb muon system". In: *Journal of Instrumentation* 8.02 (Feb. 2013), P02022–P02022. DOI: `10.1088/1748-0221/8/02/p02022`.

[18] R Aaij et al. "The LHCb trigger and its performance in 2011". In: *Journal of Instrumentation* 8.04 (Apr. 2013), P04022–P04022. DOI: `10.1088/1748-0221/8/04/p04022`.

[19] T Head. "The LHCb trigger system". In: *Journal of Instrumentation* 9.09 (Sept. 2014), p. C09015. DOI: `10.1088/1748-0221/9/09/C09015`.

[20] R. Aaij et al. "Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC". In: *Journal of Instrumentation* 14.04 (Apr. 2019), P04013–P04013. DOI: `10.1088/1748-0221/14/04/p04013`.

[21] Stephan Escher. "Test of Lepton Universality with $b \to s\ell^+\ell^-$ Decays at LHCb". Presented 24 Nov 2022. 2022. URL: `https://cds.cern.ch/record/2846313`.

[22] Titus Mombacher. "Beautiful leptons - setting limits to New Physics with the LHCb experiment". Presented 25 Nov 2020. 2020. URL: `https://cds.cern.ch/record/2748430`.

[23] Vladimir V Gligorov, Christopher Thomas, and Michael Williams. *The HLT inclusive B triggers*. Tech. rep. LHCb-INT-2011-030. Geneva: CERN, 2011. URL: `https://cds.cern.ch/record/1384380`.

[24] Tianqi Chen and Carlos Guestrin. "XGBoost". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Aug. 2016. DOI: 10.1145/2939672.2939785.

[25] A. Hoecker et al. *TMVA - Toolkit for Multivariate Data Analysis*. 2009. arXiv: physics/0703039 [physics.data-an].

[26] Andrew P. Bradley. "The use of the area under the ROC curve in the evaluation of machine learning algorithms". In: *Pattern Recognition* 30.7 (1997), pp. 1145–1159. ISSN: 0031-3203. DOI: 10.1016/S0031-3203(96)00142-2.

[27] Roger Barlow. "Extended maximum likelihood". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 297.3 (1990), pp. 496–506. ISSN: 0168-9002. DOI: 10.1016/0168-9002(90)91334-8.

[28] M. Pivk and F.R. Le Diberder. "sPlots: A statistical tool to unfold data distributions". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 555.1-2 (Dec. 2005), pp. 356–369. DOI: 10.1016/j.nima.2005.08.106.

[29] R. Aaij et al. "Measurement of lepton universality parameters in $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ decays". In: *Phys. Rev. D* 108.3 (2023), p. 032002. DOI: 10.1103/PhysRevD.108.032002. arXiv: 2212.09153.

[30] 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. 2016). DOI: 10.1007/jhep09(2016)001.

[31] S Tolk et al. *Data driven trigger efficiency determination at LHCb*. Tech. rep. Geneva: CERN, 2014. URL: https://cds.cern.ch/record/1701134.

A Trigger Lines

Trigger Level	Trigger Line
L0	Bs_L0ElectronDecision_TIS or Bs_L0MuonDecision_TIS or Bs_L0HadronDecision_TIS or E1_L0ElectronDecision_TOS or E2_L0ElectronDecision_TOS
Hlt1	Bs_Hlt1TrackMVADecision_TOS
Hlt2	Bs_Hlt2Topo2BodyDecision_TOS or Bs_Hlt2Topo3BodyDecision_TOS or Bs_Hlt2TopoE2BodyDecision_TOS [3] or Bs_Hlt2TopoE3BodyDecision_TOS[3]

(a)

L0 Trigger Lines for Run 1 and 2 and HLT Trigger Lines for Run 2.

Trigger Level	Trigger Line
Hlt1	Bs_Hlt1TrackAllL0Decision_TOS
Hlt2	Bs_Hlt2Topo2BodyBBDTDecision_TOS or Bs_Hlt2Topo3BodyBBDTDecision_TOS or Bs_Hlt2TopoE2BodyBBDTDecision_TOS or Bs_Hlt2TopoE3BodyBBDTDecision_TOS

(b)

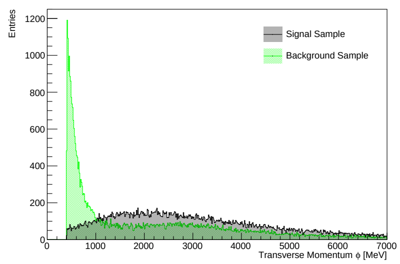
HLT Trigger Lines for Run 1.

Table 12: Applied Trigger Configurations.

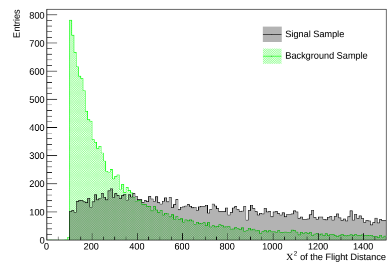
[3] These trigger lines were not used for the 2015 data due to unavailability.

B Plots of BDT Variables

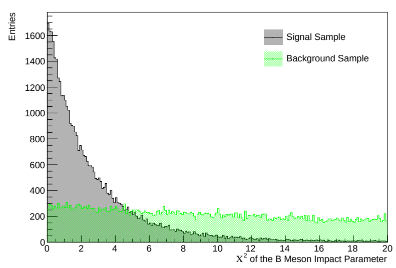
B.1 Signal and Background Distribution of BDT Variables



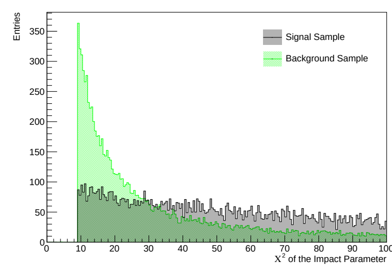
(a) Transverse momentum of the ϕ meson.



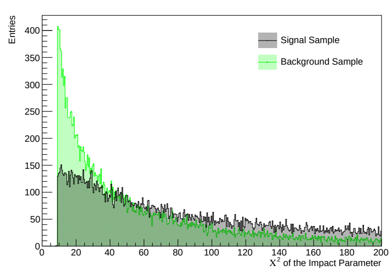
(b) χ^2 of the Flight Distance of the B_s^0 meson.



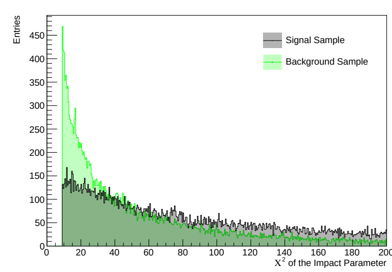
(c) χ^2 of the Impact Parameter of the B_s^0 meson.



(d) χ^2 of the Impact Parameter of the electron E2.



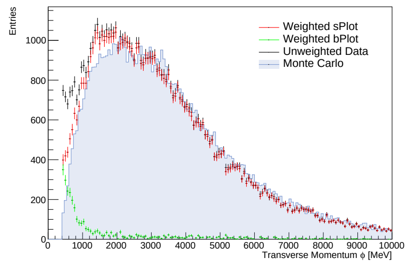
(e) χ^2 of the Impact Parameter of the kaon K1.



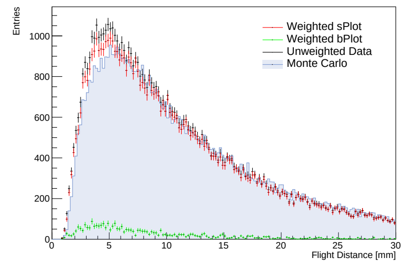
(f) χ^2 of the Impact Parameter of the kaon K2.

Figure 16: Signal and Background Distributions for variables of mesons and leptons used in training the XGBoost Model.

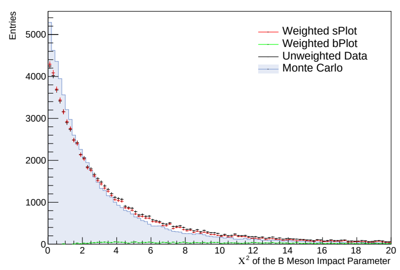
B.2 $_s\mathcal{P}lots$ of BDT Variables



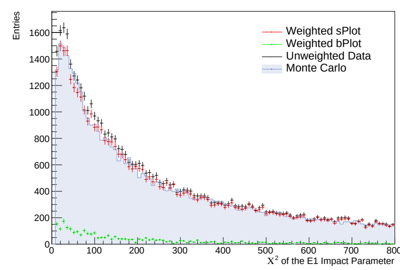
(a) Transverse momentum of the ϕ meson.



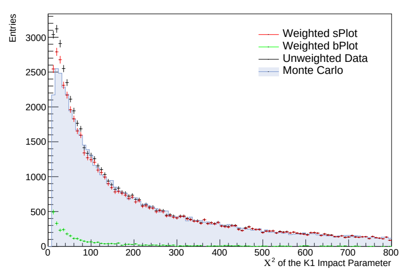
(b) Flight Distance of the B_s^0 meson.



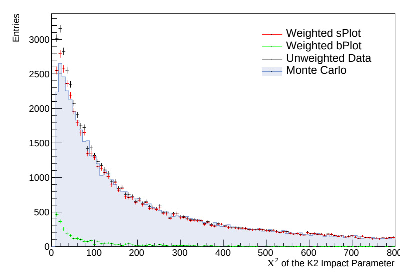
(c) χ^2 of the Impact Parameter of the B_s^0 meson.



(d) χ^2 of the Impact Parameter of the electron E1.



(e) χ^2 of the Impact Parameter of the kaon K1.

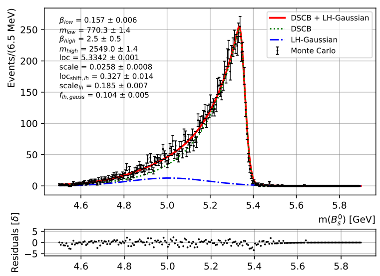


(f) χ^2 of the Impact Parameter of the kaon K2.

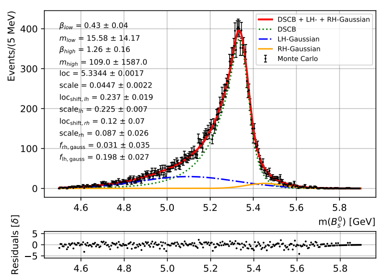
Figure 17: BDT variable distributions from MC simulations (blue), unweighted data (black), data weighted with $_s$Weights (red), and data weighted with $_b$Weights.

C $B_s^0 \to J/\psi(\to e^+e^-)\phi$ Mass Distribution Fits

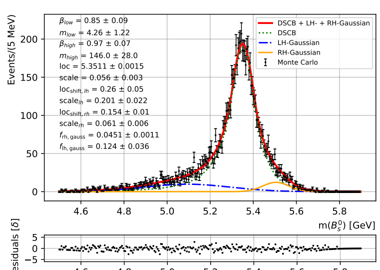
C.1 Fits for Monte Carlo Simulations



(a) 2012: 0 γ.

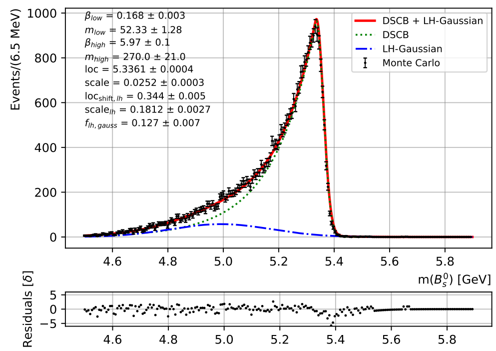


(b) 2012: 1 γ.

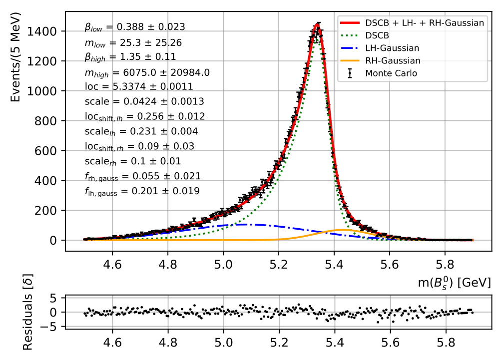


(c) 2012: 2 γ.

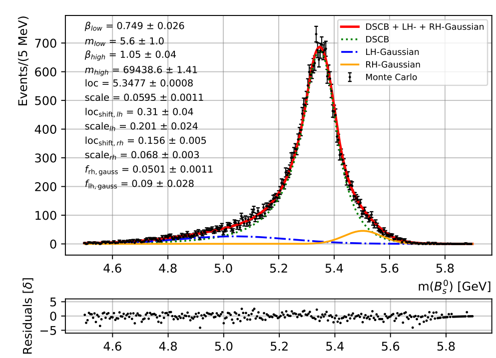
Figure 18: Fits to the mass distribution of the normalization channel for 2012 Monte Carlo simulations



(a) 2016: 0 γ.

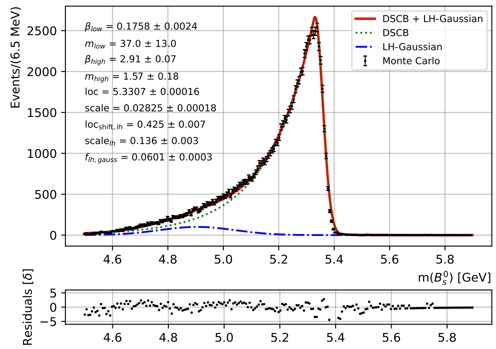


(b) 2016: 1 γ.

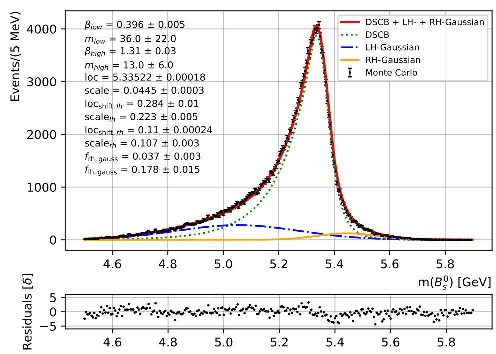


(c) 2016: 2 γ.

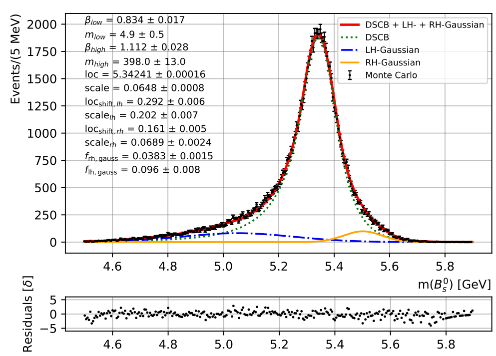
Figure 19: Fits to the mass distribution of the normalization channel for 2016 Monte Carlo simulations.



(a) 2018: 0 γ.



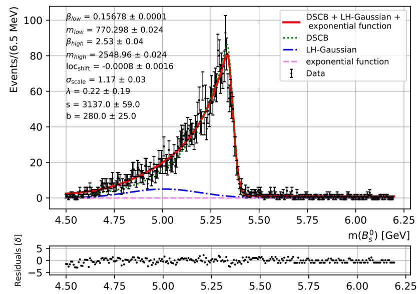
(b) 2018: 1 γ.



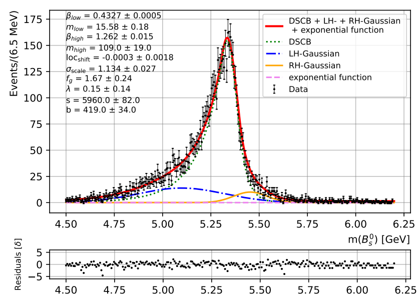
(c) 2018: 2 γ.

Figure 20: Fits to the mass distribution of the normalization channel for 2018 Monte Carlo simulations.

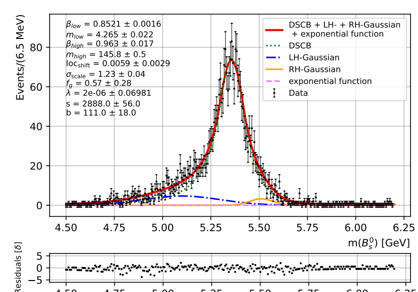
C.2 Fits for Data



(a) Data: 0 γ.

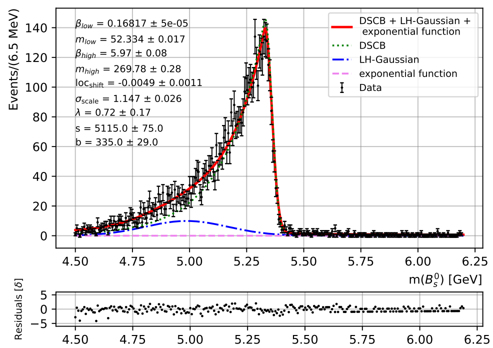


(b) Data: 1 γ.

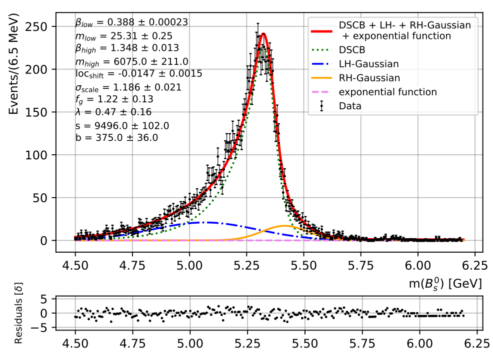


(c) Data: 2 γ.

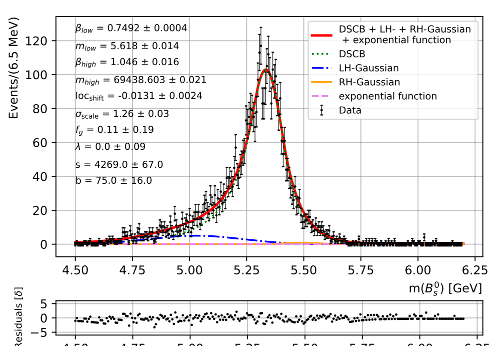
Figure 21: Fits to the mass distribution of the normalization channel for Run 1 Data.



(a) Data: 0 γ.

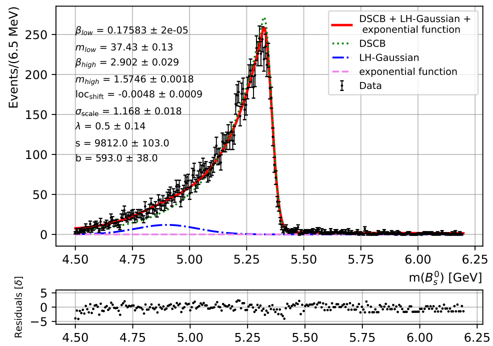


(b) Data: 1 γ.

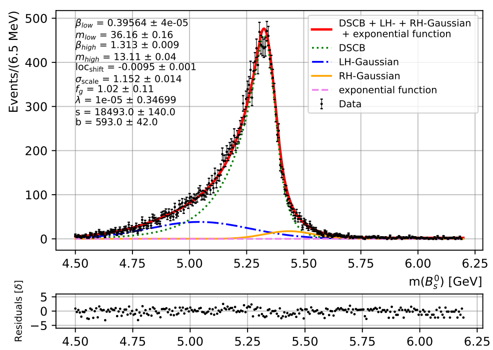


(c) Data: 2 γ.

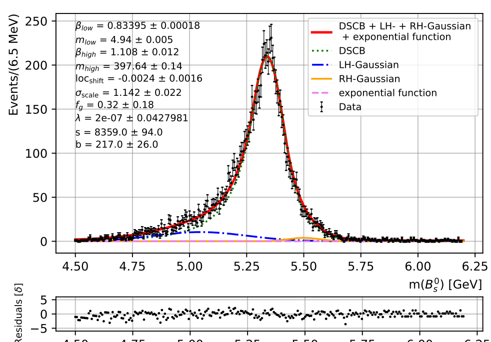
Figure 22: Fits to the mass distribution of the normalization channel for Run 2 Part 1 Data.



(a) Data: 0 γ.



(b) Data: 1 γ.

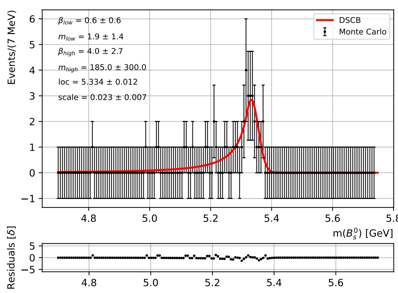


(c) Data: 2 γ.

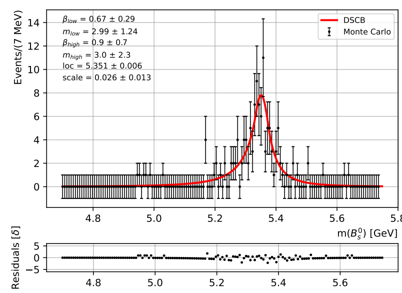
Figure 23: Fits to the mass distribution of the normalization channel for Run 2 Part 2 Data.

D $B_s^0 \to \phi\gamma(\to e^+e^-)$ Mass Distribution Fits

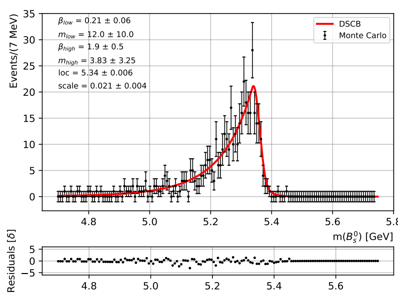
D.1 Fits for Monte Carlo Simulations



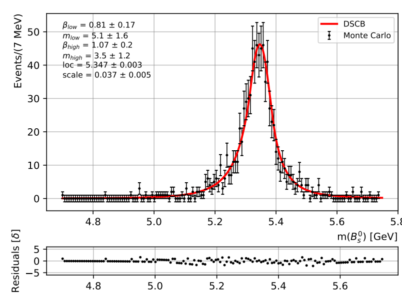
(a) 2012: 0 γ.



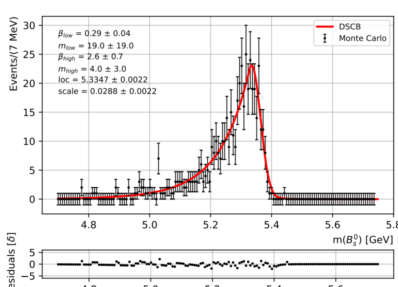
(b) 2012: 1 γ.



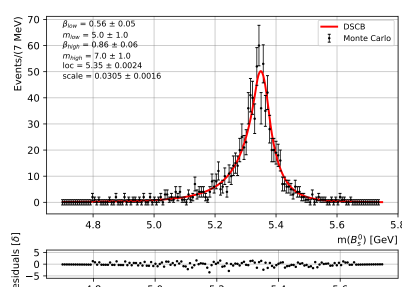
(c) 2016: 0 γ.



(d) 2016: 1 γ.



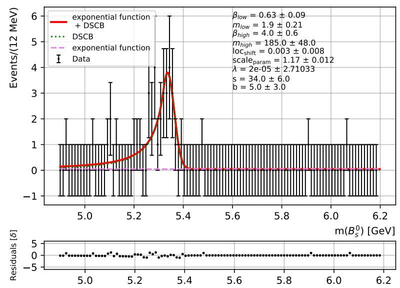
(e) 2018: 0 γ.



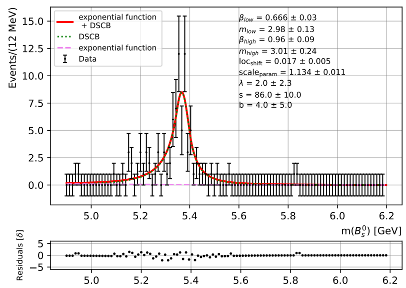
(f) 2018: 1 γ.

Figure 24: Fits to the mass distribution of the rare mode for Monte Carlo simulations split into the respective years and the number of reconstructed bremsstrahlung photons.

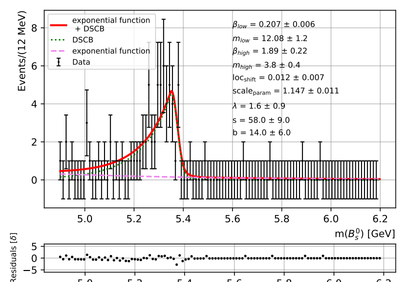
D.2 Fits for Data



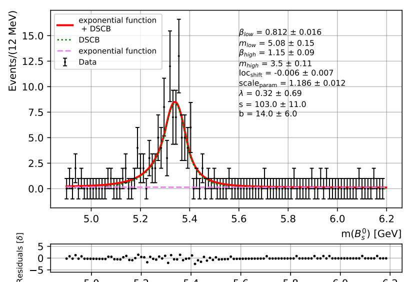
(a) 2012: 0 γ.



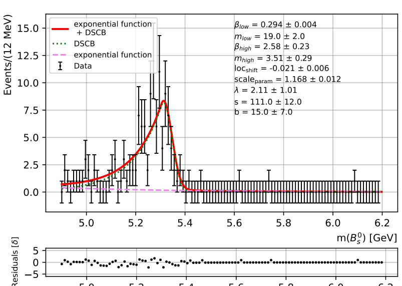
(b) 2012: 1 γ.



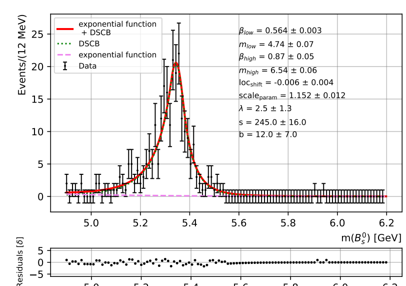
(c) 2016: 0 γ.



(d) 2016: 1 γ.



(e) 2018: 0 γ.



(f) 2018: 1 γ.

Figure 25: Fits to the mass distribution of the rare mode on Data split into the respective years and the number of reconstructed bremsstrahlung photons.

Acknowledgements

I am extremely grateful to my supervisor Dr. Christoph Langenbruch for his excellent guidance and support throughout this thesis. His advice and feedback both on the analysis and in the writing of the thesis were invaluable. I also want to thank Prof. Dr. Ulrich Uwer for providing me with the opportunity to write my thesis in the LHCb group and for agreeing to be the second referee.
Additionally, I wish to express my gratitude to the LHCb group, especially to my fellow office mates. Our engaging conversations covered a wide range of topics, both related to physics and unrelated.
I am deeply appreciative of my partner, friends, and family for supporting and believing in me. Thank you for our daily table football games, proofreading my thesis, and bringing me cake to lift my spirits.