

Department of Physics and Astronomy
University of Heidelberg

Bachelor Thesis in Physics
submitted by

Jan Maintok

born in Lucerne (Switzerland)

July 18, 2018

Measurement of the mixing frequency Δm_s of the B_s meson using the decay $B_s \rightarrow D_s^-(K^-K^+\pi^-)\pi^+\pi^-\pi^+$ with the LHCb detector

This Bachelor Thesis has been carried out by Jan Maintok at the
Physikalische Institute in Heidelberg
under the supervision of
Prof. Dr. Ulrich Uwer

Abstract

In this thesis the measurement of the B_s mixing frequency Δm_s using the decay $B_s \rightarrow D_s^- \pi^+ \pi^- \pi^+$ is presented using the D_s^- decay channel $D_s^- \rightarrow K^- K^+ \pi^-$. The data that was used in the analysis was collected using proton proton collisions at the LHCb detector in the years 2011 and 2012 at center-of-mass energies of $\sqrt{s} = 7\text{TeV}$ and 8TeV respectively. This corresponds to an integrated luminosity of 3fb^{-1} . The mixing frequency was measured to be

$$\Delta m_s/\hbar = (17.746 \pm 0.027_{\text{stat}} \pm 0.012_{\text{sys}})\text{ps}^{-1},$$

which is the second most precise measurement of this quantity. It agrees with the current world average within its errors.

Kurzfassung

In dieser Arbeit wird die Messung der B_s Mischfrequenz im Zerfall $B_s \rightarrow D_s^- \pi^+ \pi^- \pi^+$ präsentiert, wobei der D_s^- Zerfallskanal $D_s^- \rightarrow K^- K^+ \pi^-$ verwendet wurde. Die verwendeten Daten wurden bei Proton-Proton Kollisionen am LHCb Detektor in den Jahren 2011 und 2012, bei Schwerpunktsenergien von jeweils $\sqrt{s} = 7\text{TeV}$ und 8TeV aufgenommen und entsprechen einer integrierten Luminosität von 3fb^{-1} . Für die Mischfrequenz wurde ein Wert von

$$\Delta m_s/\hbar = (17.746 \pm 0.027_{\text{stat}} \pm 0.012_{\text{sys}})\text{ps}^{-1},$$

gemessen, was die bisher zweit genaueste Messung dieser Größe ist. Die Messung ist innerhalb der Fehler gut mit dem aktuellen Weltmittelwert verträglich.

Contents

1	Introduction	1
2	Theory	3
2.1	Standard Model of particle physics	3
2.2	Physics of particle-anti-particle oscillations	5
3	The LHCb experiment	7
3.1	LHC	8
3.2	Vertexdetector	9
3.3	Tracking	9
3.4	Magnet	10
3.5	Particle Identification	10
3.6	Calorimeters	11
3.7	Muon system	12
3.8	Trigger and Data collection	12
4	Analysis strategy	13
5	Signal selection	16
5.1	Variables	16
5.2	Preselection	18
5.3	Multivariate classification	21
5.3.1	BDT performance	22
5.4	Background subtraction	24
5.4.1	Comparison MC-Data	25
5.4.2	Mass dependency of the MVA-variables	27
6	Analysis	29
6.1	Time resolution	29
6.2	Acceptance function	30
6.3	Identification of B_s flavor at production	31
6.4	Final fit	33
7	Systematic uncertainties	35
7.1	Various fit systematics	35
7.2	Z-scale uncertainty	36
7.3	Momentum scale uncertainty	37
7.4	Fit bias	37
7.5	Summary	38
8	Discussion and summary	39
9	References and links	41
10	Appendix	44
11	Acknowledgment	48

1 Introduction

The Standard Model of Particle Physics describes the fundamental particles and their behavior. The theory has been extremely successful in predicting experimental results at small scales and high energies. However some questions about the universe still remain a mystery and searches to find physics beyond the Standard Model are ongoing.

There are two fundamentally different approaches to search for physics that is not described by the Standard Model, direct and indirect searches. Direct searches try to produce unknown particles in high energy particle collisions and directly measure their decay products. Indirect searches focus on the precise measurement of observables dominated by quantum loop processes, where unknown particles could give contributions that cannot be described by the Standard Model.

The mixing of the heavy B_s meson, which is the spontaneous change into its anti-particle \bar{B}_s , can only occur through a loop process, since flavor changing neutral currents are forbidden in tree level processes by the Standard Model. A Feynman diagram for the possible second order mixing processes can be seen in Fig.1.

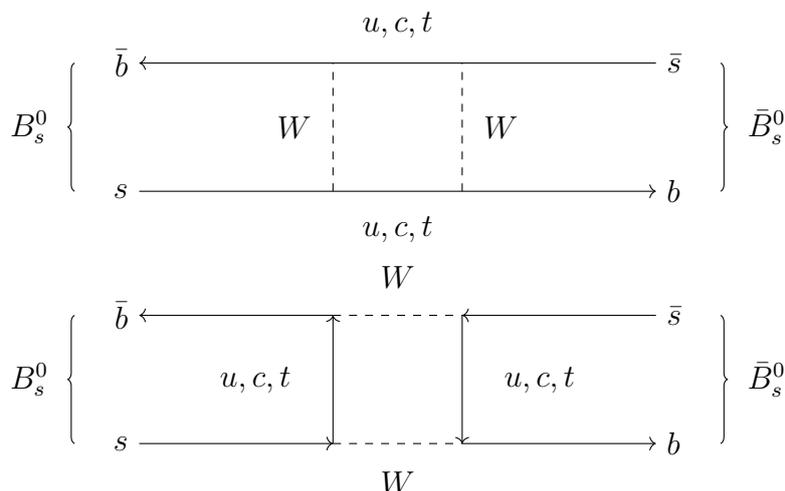


Fig. 1: Feynman diagram of mixing process

The mixing process of neutral B mesons is susceptible to contributions from physics beyond the Standard Model, for example from new heavy particles in the loop. Historically it was possible to gain information about the top quark through the analysis of the equivalent mixing process in the B_d^0 system, before a direct measurement of the top quark was possible.

Unfortunately the theory prediction for the B_s mixing frequency [1]

$$\Delta m_s / \hbar = (17.3 \pm 2.6) \text{ps}^{-1} \quad (1.1)$$

has quite a large uncertainty. This means that the measurement of the mixing frequency Δm_s has only a limited potential to find new physics, which would result in a measurement that deviates from the theory prediction.

However, measurements of \mathcal{CP} violating observables like the weak mixing phase ϕ_s [2] and the CKM angle γ in the decay $B_s^0 \rightarrow D_s^\pm K^\mp$ [3] in the B_s system have a lot smaller theory

uncertainties and therefore have a higher sensitivity to find contributions from physics beyond the Standard Model. The measurement of time dependent \mathcal{CP} violation relies on knowing the mixing frequency of this system precisely.

An analysis published in 2013 by the LHCb collaboration [4] [5] gave the most precise ever measurement of the mixing frequency Δm_s of

$$\Delta m_s/\hbar = (17.768 \pm 0.0023_{\text{stat}} \pm 0.006_{\text{sys}})\text{ps}^{-1}. \quad (1.2)$$

The LHCb measurement mentioned above measured the mixing frequency in $B_s^0 \rightarrow D_s^- \pi^+$ decays. In the analysis presented here, the measurement of the mixing frequency is performed in the decay $B_s^0 \rightarrow D_s^- \pi^+ \pi^- \pi^+$, where the D_s^- decays into $K^+ K^- \pi^-$, either directly or through the ϕ or K^{*0} resonances. A Feynman diagram of this decay can be seen in Fig.2.

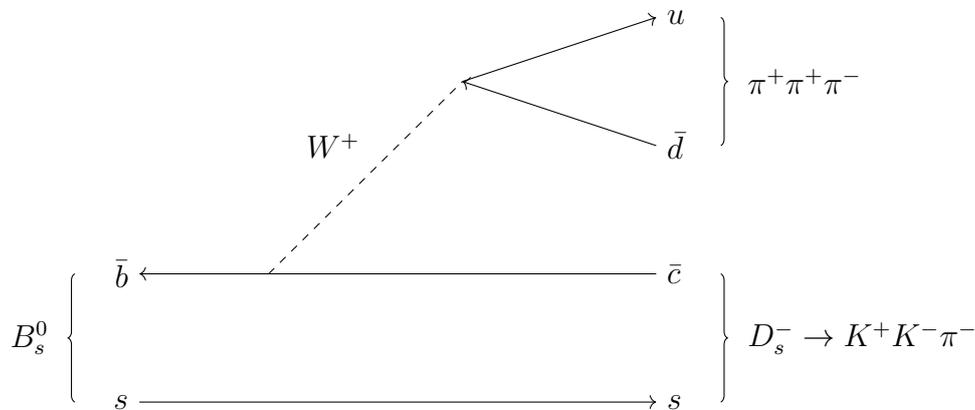


Fig. 2: Feynman diagram of the $B_s \rightarrow D_s^- \pi^+ \pi^- \pi^+$ decay

This decay is interesting for multiple reasons. Most importantly it is flavor specific, meaning that it is possible to imply the B_s flavor at decay through the charge of the decay products because the decays $B_s^0 \rightarrow D_s^- \pi^+ \pi^- \pi^+$ and $\bar{B}_s^0 \rightarrow D_s^+ \pi^- \pi^+ \pi^-$ are possible, while $B_s^0 \rightarrow D_s^+ \pi^- \pi^+ \pi^-$ and $\bar{B}_s^0 \rightarrow D_s^- \pi^+ \pi^- \pi^+$ are not. Another advantage is, that the decay contains only charged decay products which makes it more likely for the event to be fully reconstructed. Also the fact, that there are six particles in the final state, two more than in the analysis of the decay $B_s^0 \rightarrow D_s^- \pi^+$ makes vertex reconstruction and therefore the lifetime resolution a little bit more precise. On the other hand the selection is more challenging due to an enhanced background level.

2 Theory

2.1 Standard Model of particle physics

The Standard Model of particle physics is a theory that unifies the electroweak theory by Glashow, Weinberg and Salam with Quantum chromodynamics (QCD), thus describing all the known fundamental forces (electromagnetic, weak and strong) except gravity. It consists of several elementary particles as well as mediators of the fundamental interactions. Most of the matter that surrounds us in everyday life is made up of just three type of particles, electrons, up-quarks and down-quarks, which are held together and interact with each other mainly by photons and gluons. These are some of the fundamental particles, there are however more that play an important role in understanding the physics at the smallest scales and at large energies. Besides the up- and down-quark there are four more quarks: strange, charm, bottom (sometimes called beauty) and top quarks which have higher masses than the up- and down quarks. These are not stable and will decay into lighter particles over time. Up, charm and top quarks have an electric charge of $+2/3e$ while down, strange and bottom quarks have an electric charge of $-1/3$. Quarks experience all the known fundamental interactions. To interact via the strong interaction, they need to have a color charge, of which there are three types, usually called red, green and blue, as well as their anti-colors (anti-red, anti-green and anti-blue).

The electron is part of a group of particles called leptons, that do not experience the strong interaction (they do not have a color charge) but the other two remaining interactions. Its heavier partners are called muons and tauons, which all carry an electric charge of $1e$. Muons and tauons also decay into lighter particles over time.

Each lepton has a corresponding neutrino. These are very light particles that have no charge (neither electric nor color) and therefore only interact via the weak interaction.

All the particles mentioned so far have a corresponding anti-particle that mostly behaves the same as the particle but has opposite electric charge. Most matter around us is made up of particles while anti-particles are only formed in high energy interactions, even though particles can only be created from vacuum in pairs with their anti-particle partners. Why there is such an abundance of matter in the universe and no significant amount of anti-matter is not understood so far.

The fundamental interactions are mediated by so called gauge bosons. The electromagnetic interaction is mediated by photons, the strong interaction by gluons which are both massless, and the weak interaction by the electrically charged W^\pm and the neutral Z bosons which have a quite significant mass. This implies a very short range for the weak interaction.

To explain the mass of these gauge bosons it is necessary to introduce a mechanism called the Higgs mechanism. By spontaneous symmetry breaking the Higgs field couples to the bosons and thus gives them their mass. This mechanism was fundamental for the Standard Model to be consistent, though the Higgs particle, which is the quantum excitation of this Higgs field, was not observed until recently, when ATLAS and CMS, two experiments at CERN published the discovery of a new particle with properties consistent with the expected Higgs particle [6][7].

A summary of all the fundamental particles and some of their properties can be seen in Fig.3.

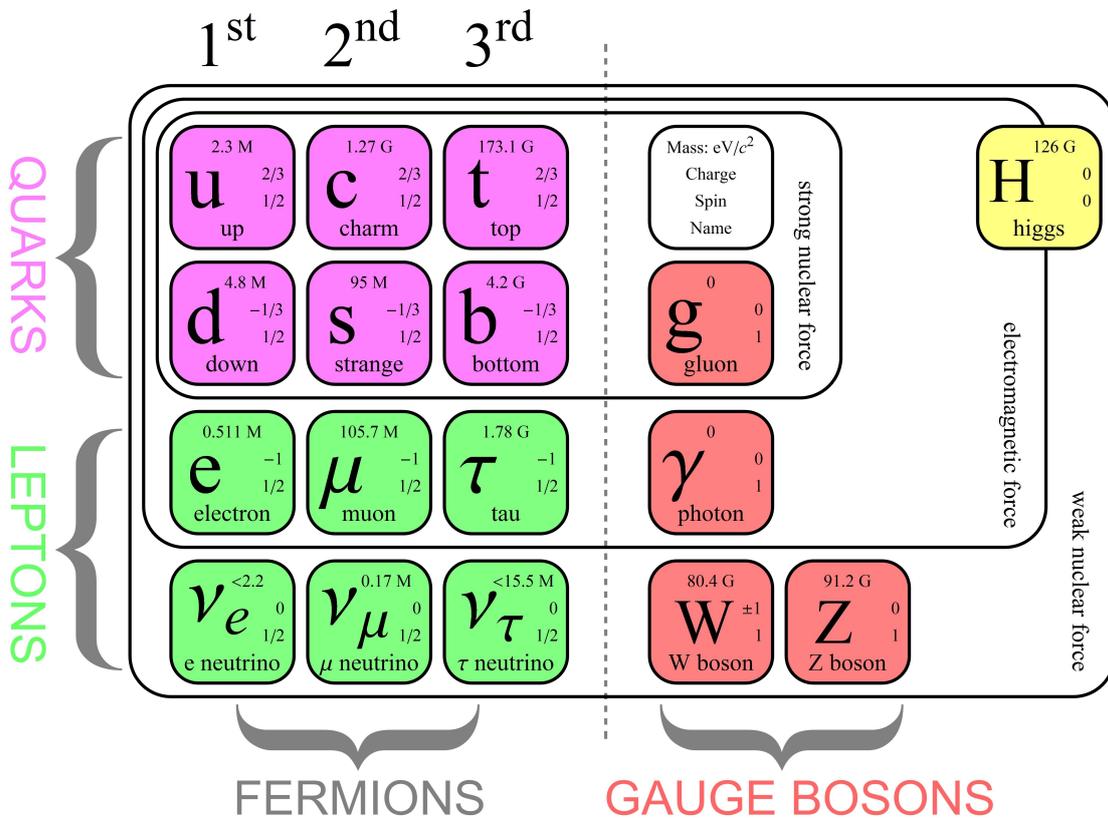


Fig. 3: Standard Model of particle physics, taken from [8]

Many of these fundamental particles can form bound states. Particles made up of multiple quarks are called hadrons. As a matter of fact quarks can only exist in bound hadrons due to an effect in QCD called confinement, which requires free particles to be colorless (a state with all three colors combines to be colorless as well as a state of one color and its anti-color). Hadrons that are made up of two quarks are called mesons, while hadrons made up of three quarks are called baryons. Recently, bound states of four and five quarks called tetra- and penta-quarks have also been observed at LHCb.

The Standard Model has been hugely successful so far and many effects have been predicted by it before there was experimental evidence. These include the heavy top quark, the tau-neutrino as well as the recently discovered Higgs boson.

There are however also some effects that are not explained by the Standard Model and theorists and experimentalists are working heavily to understand what is commonly called physics beyond the Standard Model. For example the Standard Model does not explain gravitation, the matter-anti-matter asymmetry that can be observed in the universe cannot be described by the Standard Model and dark matter and dark energy that are predicted in cosmological theories are also not part of the Standard Model.

2.2 Physics of particle-anti-particle oscillations

The flavor eigenstate $|B_s^0\rangle = |\bar{b}s\rangle$ can change into $|\bar{B}_s^0\rangle = |b\bar{s}\rangle$ through a box diagram as can be seen in Fig.1.

This means an initially produced B_s^0 meson will become a superposition of the two flavor eigenstates $|B_s^0\rangle$ and $|\bar{B}_s^0\rangle$. The time evolution of this state is described by the Schrödinger equation

$$i\frac{\partial}{\partial t} \begin{pmatrix} |B_s^0\rangle \\ |\bar{B}_s^0\rangle \end{pmatrix} = \left(\mathbf{M} - i\frac{\mathbf{\Gamma}}{2} \right) \begin{pmatrix} |B_s^0\rangle \\ |\bar{B}_s^0\rangle \end{pmatrix} \quad (2.1)$$

$$= \begin{pmatrix} M_{11} - i\frac{\Gamma_{11}}{2} & M_{12} - i\frac{\Gamma_{12}}{2} \\ M_{21} - i\frac{\Gamma_{21}}{2} & M_{22} - i\frac{\Gamma_{22}}{2} \end{pmatrix} \begin{pmatrix} |B_s^0\rangle \\ |\bar{B}_s^0\rangle \end{pmatrix} \quad (2.2)$$

$$= \begin{pmatrix} M - i\frac{\Gamma}{2} & M_{12} - i\frac{\Gamma_{12}}{2} \\ M_{12}^* - i\frac{\Gamma_{12}^*}{2} & M - i\frac{\Gamma}{2} \end{pmatrix} \begin{pmatrix} |B_s^0\rangle \\ |\bar{B}_s^0\rangle \end{pmatrix} \quad (2.3)$$

where \mathbf{M} describes the mass matrix and $\mathbf{\Gamma}$ the decay width matrix. These two are hermitian, which was used in the last step ($M_{21} = M_{12}^*$ and $\Gamma_{21} = \Gamma_{12}^*$). Also CPT invariance implies $M_{11} = M_{22} = M$ and $\Gamma_{11} = \Gamma_{22} = \Gamma$.

The mass eigenstates are given as the eigenvectors of the Hamiltonian ($\mathbf{H} = \mathbf{M} - i\mathbf{\Gamma}/2$) and can be constructed from the flavor eigenstates as

$$|B_L\rangle = p|B_s^0\rangle + q|\bar{B}_s^0\rangle \quad (2.4)$$

$$|B_H\rangle = p|B_s^0\rangle - q|\bar{B}_s^0\rangle \quad (2.5)$$

where L indicates the lighter eigenstate and H the heavier one. p and q are complex numbers, that have to fulfill $|p|^2 + |q|^2 = 1$. These mass eigenstates have the mass eigenvalues $(M_L + i\frac{\Gamma_L}{2})$ and $(M_H + i\frac{\Gamma_H}{2})$, which leads to a time evolution of

$$|B_L(t)\rangle = e^{-(iM_L + \frac{\Gamma_L}{2})t} |B_L\rangle \quad (2.6)$$

$$|B_H(t)\rangle = e^{-(iM_H + \frac{\Gamma_H}{2})t} |B_H\rangle. \quad (2.7)$$

Using eq.2.5 one gets the time evolution of the flavor eigenstates

$$|B_s^0(t)\rangle = \frac{1}{2p} (|B_L(t)\rangle + |B_H(t)\rangle) = g_+(t) |B_s^0\rangle + \frac{q}{p} g_-(t) |\bar{B}_s^0\rangle \quad (2.8)$$

$$|\bar{B}_s^0(t)\rangle = \frac{1}{2q} (|B_L(t)\rangle - |B_H(t)\rangle) = \frac{p}{q} g_-(t) |B_s^0\rangle + g_+(t) |\bar{B}_s^0\rangle \quad (2.9)$$

with the definition

$$g_{\pm}(t) := \frac{1}{2} \left(e^{-(iM_L + \frac{\Gamma_L}{2})t} \pm e^{-(iM_H + \frac{\Gamma_H}{2})t} \right). \quad (2.10)$$

The decay rate for a B_s^0 meson into a final state f is given by

$$\frac{d\Gamma(B_s^0 \rightarrow f)}{dt} = |\langle f | B_s^0(t) \rangle|^2. \quad (2.11)$$

In the case of this analysis a flavor specific decay is analyzed. That means that the decay into the final state $D^-\pi^+\pi^-\pi^+$ can only come from a B_s^0 meson, but not from a \bar{B}_s^0 . The same is true for the charge conjugate (from now on when it is not specifically stated

otherwise the charge conjugate of any process is also implied). This simplifies the problem significantly because in this case we get

$$\frac{d\Gamma(B_s^0 \rightarrow D_s^- \pi^+ \pi^- \pi^+)}{dt} = |\langle D_s^- \pi^+ \pi^- \pi^+ | B_s^0(t) \rangle|^2 \quad (2.12)$$

$$\propto |\langle B_s^0 | B_s^0(t) \rangle|^2 \quad (2.13)$$

$$= |g_+|^2 \quad (2.14)$$

$$= \frac{1}{2} e^{-\Gamma t} \left(\cosh \left(\frac{\Delta\Gamma}{2} t \right) + \cos(\Delta m t) \right) \quad (2.15)$$

$$\frac{d\Gamma(\bar{B}_s^0 \rightarrow D_s^+ \pi^- \pi^+ \pi^-)}{dt} = |\langle D_s^+ \pi^- \pi^+ \pi^- | \bar{B}_s^0(t) \rangle|^2 \quad (2.16)$$

$$\propto |\langle \bar{B}_s^0 | \bar{B}_s^0(t) \rangle|^2 \quad (2.17)$$

$$= |g_+|^2 \quad (2.18)$$

$$= \frac{1}{2} e^{-\Gamma t} \left(\cosh \left(\frac{\Delta\Gamma}{2} t \right) + \cos(\Delta m t) \right) \quad (2.19)$$

Similarly one gets

$$\frac{d\Gamma(B_s^0 \rightarrow \bar{B}_s^0 \rightarrow D_s^+ \pi^- \pi^+ \pi^-)}{dt} = |\langle D_s^+ \pi^- \pi^+ \pi^- | B_s^0(t) \rangle|^2 \quad (2.20)$$

$$\propto |\langle \bar{B}_s^0 | B_s^0(t) \rangle|^2 \quad (2.21)$$

$$= \left| \frac{q}{p} g_- \right|^2 \quad (2.22)$$

$$= \frac{1}{2} e^{-\Gamma t} \left(\cosh \left(\frac{\Delta\Gamma}{2} t \right) - \cos(\Delta m t) \right) \quad (2.23)$$

$$\frac{d\Gamma(\bar{B}_s^0 \rightarrow B_s^0 \rightarrow D_s^- \pi^+ \pi^- \pi^+)}{dt} = |\langle D_s^- \pi^+ \pi^- \pi^+ | \bar{B}_s^0(t) \rangle|^2 \quad (2.24)$$

$$\propto |\langle B_s^0 | \bar{B}_s^0(t) \rangle|^2 \quad (2.25)$$

$$= \left| \frac{p}{q} g_- \right|^2 \quad (2.26)$$

$$= \frac{1}{2} e^{-\Gamma t} \left(\cosh \left(\frac{\Delta\Gamma}{2} t \right) - \cos(\Delta m t) \right) \quad (2.27)$$

here, the definitions

$$\Gamma = \frac{\Gamma_L + \Gamma_H}{2} \quad (2.28)$$

$$\Delta\Gamma = \Gamma_L - \Gamma_H \quad (2.29)$$

$$\Delta m = M_H - M_L \quad (2.30)$$

as well as the fact that $|p/q| \approx 1$ where used. The last approximation is due to negligible effects of CP-violation in the mixing process of B_s^0 mesons. Meaning, that the two mixing processes $B_s \rightarrow \bar{B}_s$ and $\bar{B}_s \rightarrow B_s$ happen with the same probability. [9]

A more indepth description of the Standard Model as well as the theory of B_s mixing can be found in [5], [10]–[14] on which this description is based.

3 The LHCb experiment

LHCb is a specialized detector to perform precision measurements of hadrons containing the heavy beauty and charm quarks, such as B- and D-Mesons which are used in this analysis. These type of events are strongly forward boosted in hadron colliders which is why LHCb is, different to the other large Experiments (ATLAS, CMS and ALICE) at LHC, asymmetric and "only" covers a pseudorapidity range from $2 < \eta < 5$. The pseudorapidity η is commonly used in particle physics and is defined as

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] \quad (3.1)$$

where θ describes the angle between the particles three-momentum and the beam axis. It is a useful quantity to describe the particle distribution relative to the beam axis since particle production is roughly constant as a function of the pseudorapidity.

The detector is made up of different subsystems with different purposes. Tracking detectors are able to measure the path of particles, thus giving positional information. High resolution tracking detectors close to the interaction point can give information about possible displaced vertices and thus the lifetime of certain short-lived particles. Installing a strong magnetic field results in the possibility to also measure the particles momentum and charge using the tracking systems. Particle identification detectors can give information about the velocity of particles and thus identify their type using the momentum measurement from the tracking devices. The calorimeters give information about the energy of hadrons, electrons and photons. Because of the high number of events in the detector it is not possible to save all the information and data that is collected. The trigger reduces the total number of possible events to a reasonable amount of interesting ones that can be saved to long-term storage.

A sketch of the LHCb detector can be seen in Fig.4

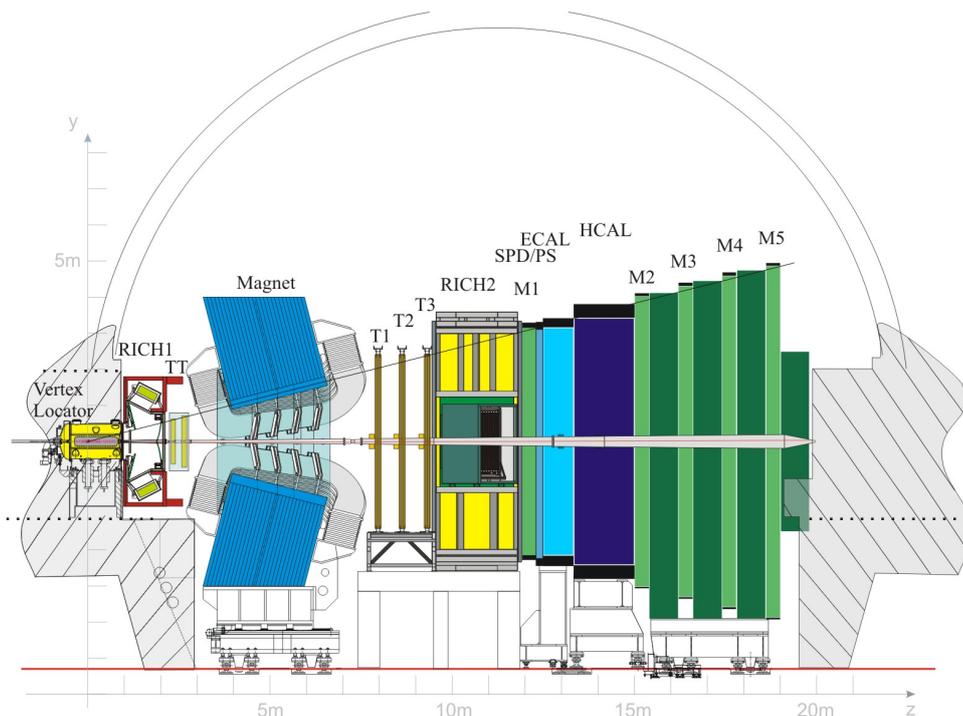


Fig. 4: Sketch of the LHCb Experiment, taken from [15]

The interaction takes place inside the **Vertex Locator** (VELO), which is a high resolution tracking device, specifically for vertex reconstruction. It is followed by the first **Ring Imaging Cherenkov** detector (RICH1), a subdetector for particle identification. The **Tracker Turicensis** (TT) is a tracking detector, followed by the big magnet responsible for bending the tracks of charged particles. The **Tracking stations** T1-T3 will give more positional information and, through the bending by the magnet, momentum measurements. The RICH2 will also give information about the identity of particles and is followed by the first **muon** chamber (M1), responsible for measuring muons. The **Electromagnetic Calorimeter** (ECAL) and the **Hadron Calorimeter** (HCAL) will measure the energies of particles. At the very rear of the detector four more **muon** stations (M2-M5) are placed for muon measurements. These subdetectors will be described in more detail in the following section.

The description in this section is mostly based on [16] and [17].

3.1 LHC

The Large Hadron Collider (LHC) is the largest and most powerful particle accelerator in the world. It is located at CERN (named after the **Conseil européen pour la recherche nucléaire**), the European Organization for Nuclear Research close to the Swiss city of Geneva. Starting construction in 1998, it replaced the **Large Electron-Positron Collider** (LEP) in its 27km long tunnel roughly 100m beneath the earth. Hadrons, mostly protons but also heavy ions such as lead, are accelerated close to the speed of light up to a center of mass energy of 14TeV. 1232 superconducting dipole magnets of 15m in length are keeping the particles on their circular path, while 392 quadrupole magnets are used to focus the beam. Two beams are accelerated in opposite direction in two separate beam pipes before they are made to collide in the experiments such as the LHCb experiment.

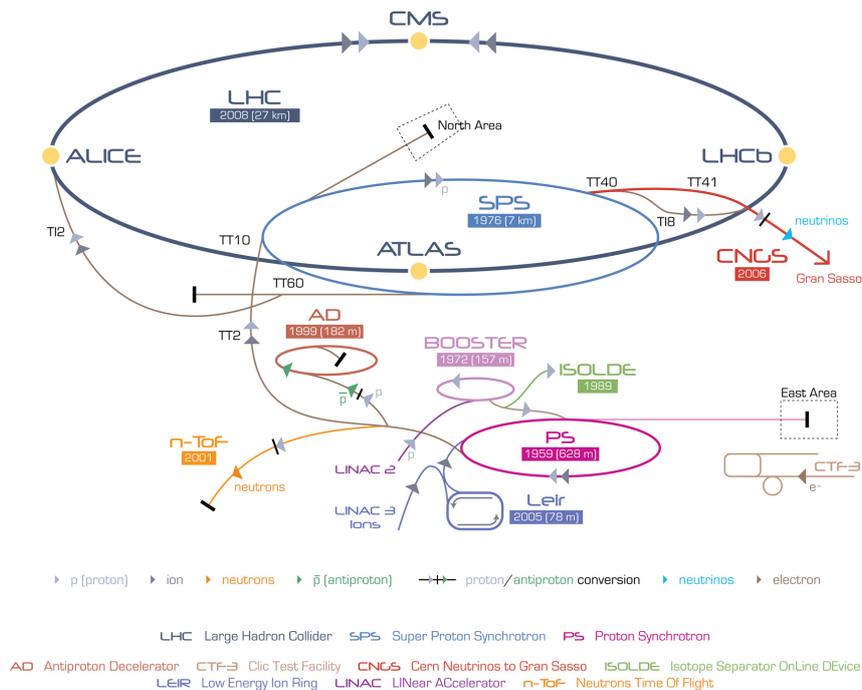


Fig. 5: Sketch of the CERN site with all the pre-accelerators as well as the LHC, Taken from [18]

3.2 Vertexdetector

To be able to reconstruct secondary vertices, that occur in the decay of charm- and beauty-hadron decays, it is crucial to precisely measure the position of the particles coming out of the interaction. The Vertex Locator (VELO) is a detector made up of 42 silicon modules which are placed along the beam axis. During measurements when the beam is focused it is placed just 5mm from the beam to give the best possible vertex reconstruction. When the beam is still unstable the detector is physically retracted from the beam axis to prevent any damage from the high energy beam. The VELO is made up of silicon tracking detectors, which have a positional resolution of $10\mu\text{m}$ which makes it possible to reconstruct vertices at a resolution of less than $50\mu\text{m}$. Charged tracks will lead to ionization of the silicon. Through an applied voltage the electron-ion pairs that are produced will drift to the read out electronics and register a signal.

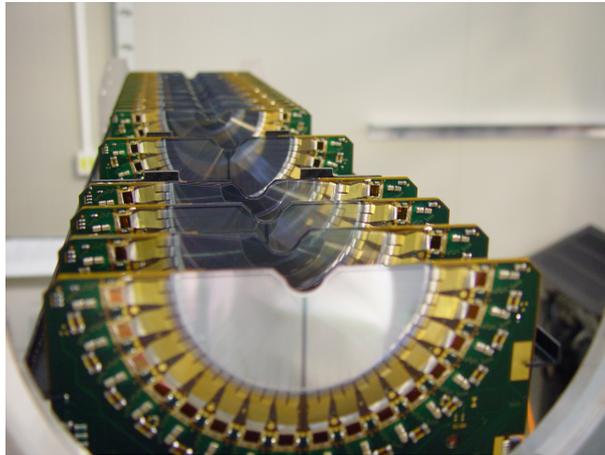


Fig. 6: One half of the VELO, Taken from [19]

3.3 Tracking

The main task of the tracking system is to provide accurate information about the position of passing charged particles. This information can be used to reconstruct particles tracks as well as their momentum (thanks to the magnet described in sec.3.4). The tracking system consists of the Tracker Turicensis (TT) which is situated before the magnet as well as three layers of inner (IT) and outer trackers (OT) behind the magnet, called the tracking stations T1-T3. The TT and the inner tracker are made up of silicon tracking detectors, that function as described above. These give excellent positional resolution but are quite expensive, which is why they are used in areas of high track density such as close to the interaction point in the VELO and TT but also close to the beam pipe in the later tracking stations. They are produced with a strip pitch of $200\mu\text{m}$ giving a positional resolution of $\sigma_x \approx 60\mu\text{m}$ for just one detector layer. The TT covers the whole acceptance of the detector at a size of $150 \cdot 130\text{cm}^2$ while the IT covers only an area of 120cm width and 40cm height. The outer tracker covers the rest of the detector acceptance. It is made up of straw-tube drift chambers, which have a diameter of 5mm and are filled with a gas mixture of 70% argon and 30% carbon dioxide. Two layers of straw tubes are shifted by the radius of the tubes for each layer to guarantee the best coverage. They give a positional resolution of $\sigma_x \approx 0.2\text{mm}$, thanks to the additional information that can

be obtained by measuring the drift time of the ionization. Even though the resolution is significantly less than it is in the silicon detectors, they are much cheaper to build and therefore excellent to cover the large area of up to $600 \cdot 490\text{cm}^2$ in T3.

3.4 Magnet

A magnet is used to bend the charged particle tracks to measure their momentum and charge. For the required momentum resolution of $\sigma_p < 0.004p$ for momenta up to $200\text{GeV}/c$ an integrated magnetic field of 4Tm is required. The magnet is made up of two massive aluminum coils containing almost 3000m of aluminum cable to achieve this goal.

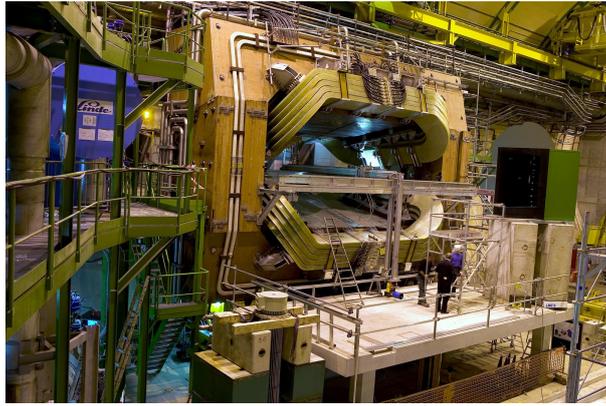


Fig. 7: Image of the LHCb magnet, Taken from [20]

3.5 Particle Identification

Particles can be identified by combining information about their charge, mass and lifetime. The charge can be found by the tracking systems, while the lifetime will be calculated from the reconstructed vertices. Getting information about the mass of the particle requires an indirect approach in LHCb. This information can either be gathered by measuring the energy of a particle or its velocity and combining it with the momentum measurement from the tracking system. The **Ring Imaging Cherenkov** (RICH) detectors are subdetectors specifically made to measure the velocity of particles. One of them, RICH-1, is placed directly behind the VELO to identify low momentum particles ($\sim 1\text{-}50\text{ GeV}/c$) that will be bent outside the detectors acceptance by the magnet, while RICH-2 is placed behind the tracking stations T1-T3 to identify particles of momenta up to $150\text{GeV}/c$.

RICH detectors rely on the fact that charged particles that pass an optically transparent medium faster than the phasevelocity of light in that medium will emit Cherenkov radiation. This radiation will be emitted on a cone with an opening angle of

$$\cos \theta = \frac{1}{n\beta} \quad (3.2)$$

where n is the refractive index of the medium, while $\beta = \frac{v}{c}$ describes the velocity of the particle in units of the speed of light. Measuring this opening angle will therefore give information about the velocity of the passing particle.

RICH-1 uses silica aerogel as well as a Perfluorobutane (C_4F_{10}) gas radiator to produce the Cherenkov radiation, while RICH-2 uses a Tetrafluoromethane (CF_4) gas radiator.

The light cones produced by these radiators are then reflected into photon detectors by spherical mirrors, that are made especially light in order to reduce scattering and energy loss of the passing particles.

The photon detectors are made up of **Hybrid Photon Detectors** (HPDs). They have the advantage that they combine the high gains as in traditional photo multipliers with an excellent position resolution like in a silicon photomultiplier. A photon will hit a photocathode and produce an electron via the photoeffect. This electron is then accelerated by a high voltage to hit a silicon pixel detector that has 1024 pixels, each $500 \cdot 500 \mu m^2$ in size. RICH-1 contains 196 and RICH-2 288 of these HPDs.

These detectors allow for excellent particle identification, which is crucial in the experiments at LHCb.

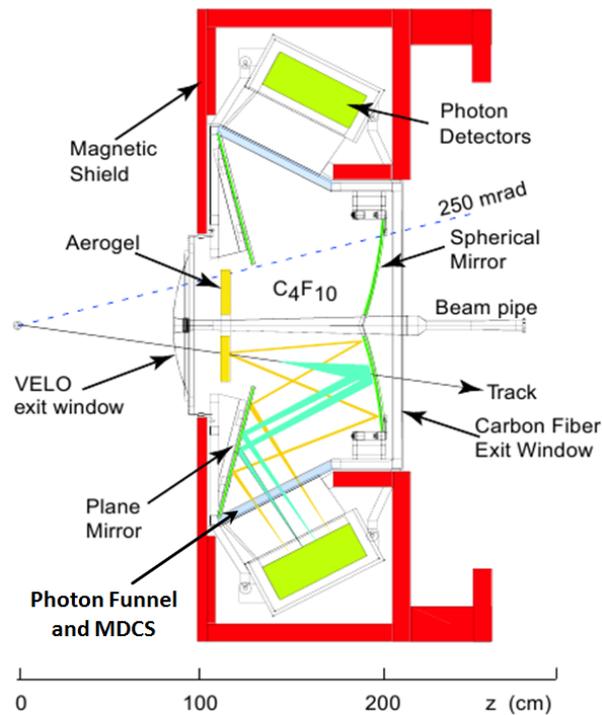


Fig. 8: Sketch of RICH1, Taken from [21]

3.6 Calorimeters

The **Electromagnetic Calorimeter** (ECAL) and the **Hadron Calorimeter** (HCAL) will mainly provide energy measurements of passing particles but also give some positional information. They are also important to trigger on electrons, photons and hadrons. Except for the triggering of the hadrons present in the decay, they will not play such a crucial role for the analysis presented in this thesis, because the decay does not contain any neutral hadrons, electrons or photons. The Calorimeters will therefore only be described very briefly. The ECAL measures the energy of lighter particles, such as electrons or photons, while the HCAL will measure the energies of heavier particles, such as protons, neutrons or other hadrons. They are both sampling shower detectors, meaning that between the active

layers, which actually detect something, there will be layers of inactive absorber. This reduces the energy resolution by a bit but has the advantage that it makes the detector more compact, since the absorber has more stopping power than the active detector layers.

3.7 Muon system

The muon system is also not of such fundamental importance for this analysis. It will only be useful to distinguish the pions in the decay from muons, which have quite a similar mass. At the very rear of the detector five layers (the first is positioned before the calorimeters, the other behind it) of **M**ulti **W**ire **P**roportional **C**hambers (MWPCs) measure the position of passing particles. Since most other particles are stopped in the earlier parts of the detectors the probability that a particle that registers here is a muon is quite high.

3.8 Trigger and Data collection

The rate of collisions at LHCb is roughly 40MHz. Saving all the information from these interactions is unfortunately not possible and the events that are actually interesting have to be selected to be saved. During the years 2011 and 2012 about 5000 events could be saved at LHCb per second. To reduce the number of events from 40MHz to just 5kHz two levels of triggers are implemented.

The Level-0 trigger is implemented in hardware electronics, which makes it fast enough to cope with the high event rates. It reduces the rate to 1MHz by analyzing the transverse momentum, using information from the calorimeters and the muon systems. These are fast enough to be able to give information in the short time available. Events that contain B decays typically have particles with high transverse momentum.

The **H**igh **L**evel **T**rigger (HLT) is a software trigger, that is further divided into HLT1 and HLT2.

HLT1 adds information of the tracking system and the VELO. Especially it is based on the fact that B hadrons typically have quite a long life time which means some of the particles will be produced in secondary or tertiary vertices. If there are particles present in the event that are unlikely to come from the primary vertex, the event is more likely to contain a B decay. HLT1 further reduces the event rate to just a few tens of kHz.

HLT2 finally fully reconstructs the events and selects those which are most likely useful for further analysis and calibration. These events are then written to storage to be further analyzed.

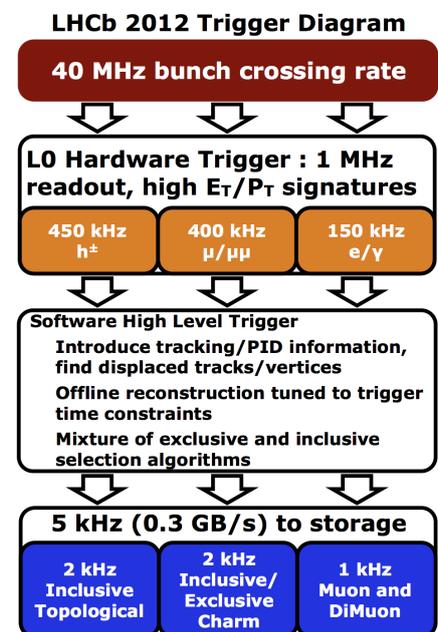


Fig. 9: Sketch of data flow in the LHCb Trigger, Taken from [22]

4 Analysis strategy

To measure the mixing frequency Δm_s of the B_s -Mesons several steps are necessary.

Figure 10 shows a typical signature for a B_s^0 decay in the decay channel that is analyzed in this thesis. A B_s^0 meson will be produced in the primary vertex (PV) in a pair with another b hadron that has to be of opposite flavor than the signal B_s^0 . The information from the decay of this second \bar{b} hadron will be used to get information about the flavor of the B_s^0 at production, because of the mixing this cannot be found from the decay products of the B_s^0 . This process is called flavor tagging and will be described in more detail later on. The lifetime τ of the B_s^0 is high enough for it to fly a significant distance (in case of the events used in this analysis this distance is of the order of 10mm) in the detector during which it will change flavor on average 9 times before it decays. The decay into a D_s^- (assuming its flavor is B_s^0 at decay) and three pions will create a secondary vertex (SV), which is displaced by the flight distance (FD) from the primary vertex. The primary vertex has to be reconstructed from other particles coming from the primary interaction. The three pions formed in the secondary vertex are stable enough not to decay before they reach the end of the detector or are absorbed. The D_s^- however will decay after flying for another bit (typically also in the order of 10mm) before it decays into two kaons and a pion in the tertiary vertex (TV). These will also live long enough to detect them directly. From these six particle tracks the full decay, meaning the position of the secondary and tertiary vertex and thus the lifetime of the B_s^0 and D_s^- as well as the flavor of the B_s^0 at decay, can be reconstructed.

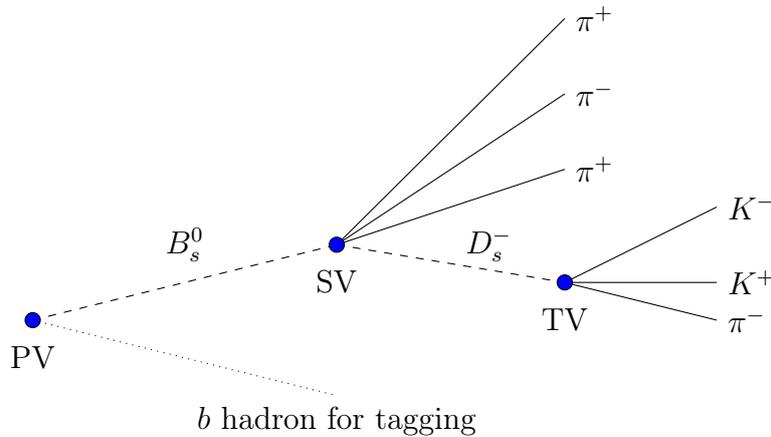


Fig. 10: Typical B_s^0 decay signature

The problem is that similar signatures can be created from other decays or just prompt (meaning coming from the primary interaction) particles, that might be falsely identified as signal decays and influence the result. Therefore the first step of the analysis must be to best identify the properties of signal and background events to select those events that most likely come from actual B_s^0 decays in the decay channel stated above.

From theory one would expect a decay time distribution of

$$\mathcal{P}(t, \Delta m_s, q) \propto \frac{1}{2} e^{-\Gamma t} \left(\cosh \left(\frac{\Delta \Gamma}{2} t \right) - q \cos(\Delta m_s t) \right) \quad (4.1)$$

where q is 1 for mixed events, meaning the flavor at production is not the same as the flavor at decay, and -1 for unmixed events, where the flavor at decay is the same as at decay. However, there are some effects that play a role in the actual measurement that have to be taken into account in the final fit.

First of all, the tagging is unfortunately not 100% efficient and with a certain probability events will be mistagged (more on that in section 6.3). Because of this, a tagging dilution factor \mathcal{D} has to be introduced to account for this inefficiency. This factor reduces the amplitude of oscillating term in the mixing distribution.

Another effect that has to be accounted for is the time resolution of the detector. The decay time is reconstructed from the position of the vertices as well as the particles momentum as

$$ct = \frac{\text{FD}}{\beta\gamma} \quad (4.2)$$

where t is the decay time, FD is the flight distance (distance between production and decay vertex) and $\beta = \frac{v}{c}$ and $\gamma = \frac{1}{\sqrt{1-\beta^2}}$ are relativistic quantities that describe the Lorentz boost. These will have uncertainties that will result in a non-ideal time resolution of the detector. Therefore a model $\mathcal{G}(t_{res})$ has to be found that describes the decay time resolution. In the decay time dependent mixing distribution, this model has to be convolved with the decay time dependent model.

Through the selection process described above, not all decay times of the B_s will have the same chance of making it through the trigger and selection process, since events with very short decay times will have a signature that could also easily be created by non-signal events. This is why a detector acceptance function $\epsilon(t)$ has to be found, that describes how efficient the selection is for different decay times. To describe the data, this acceptance function has to be multiplied with the model.

The full model that describes how the mixing process will behave as a function of the decay time is given by

$$\mathcal{P}(t, \Delta m_s) \propto \left[e^{-\Gamma_s t} \left(\cosh \left(\frac{\Delta\Gamma_s t}{2} \right) - q \cdot \mathcal{D} \cdot \cos(\Delta m_s t) \right) \otimes \mathcal{G}(t_{res}) \right] \cdot \epsilon(t) \cdot \mathcal{N}_{tag} \quad (4.3)$$

where Γ_s describes the decay width, $\Delta\Gamma_s$ the difference in decay width for the two mass eigenstates, q the mixing information (it is 1 for mixed events and -1 for unmixed events), $\mathcal{D} = 1 - 2\omega$ the tagging dilution factor, with ω being the mistag probability. The mixing frequency is given by the mass difference of the two mass eigenstates Δm_s , $\mathcal{G}(t_{res})$ describes the decay time resolution and $\epsilon(t)$ the detector acceptance function.

To find the parameters that appear in this model, the following steps will be taken

- The decay time resolution $\mathcal{G}(t_{res})$ and the detector acceptance function $\epsilon(t)$ are estimated from simulated data to serve as an input to the decay time fit. This is discussed in sections 6.1 and 6.2
- The tagging performance of the detector is then analyzed in section 6.3
- With all the information, it is finally possible to fit for the mixing frequency. This can be found in section 6.4

To fit the distribution to the data (and all other fits in this analysis where binning is not specifically mentioned) an unbinned maximum likelihood fit is performed. How such a fit is performed is described in the Appendix (10).

The simulated data is obtained from a series of programs that are specialised on simulating different parts of the event. They are based on the Monte Carlo technique. Pythia [23] is used to generate the particles as well as the momenta that come out of the primary proton-proton interaction. EvtGen [24] is specialised on simulating the decays of heavy flavor particles such as B mesons. How these particles interact with the detector is then simulated by Geant4 [25]. The events are then reconstructed using the same algorithms that are used for the real data.

5 Signal selection

The dataset that has been used for this analysis was taken in the years 2011 and 2012 by the LHCb detector. This corresponds to an integrated luminosity of $\mathcal{L} \approx 3\text{fb}^{-1}$ at an energy of 7 and 8 TeV respectively. The event reconstruction is performed centrally at LHCb and B_s candidates consistent with the desired topology are selected by applying some loose cuts. (These are not listed here in detail, because the additional cuts that will be applied for the analysis are much tighter) The invariant B_s^0 mass spectrum of this data, calculated from the four momenta of the B_s^0 decay products, can be seen in Fig. 11. Here one can clearly see how the signal events (small peak at the B_s mass $m_{B_s} \approx 5367\text{MeV}/c^2$) are buried in a large amount of background events.

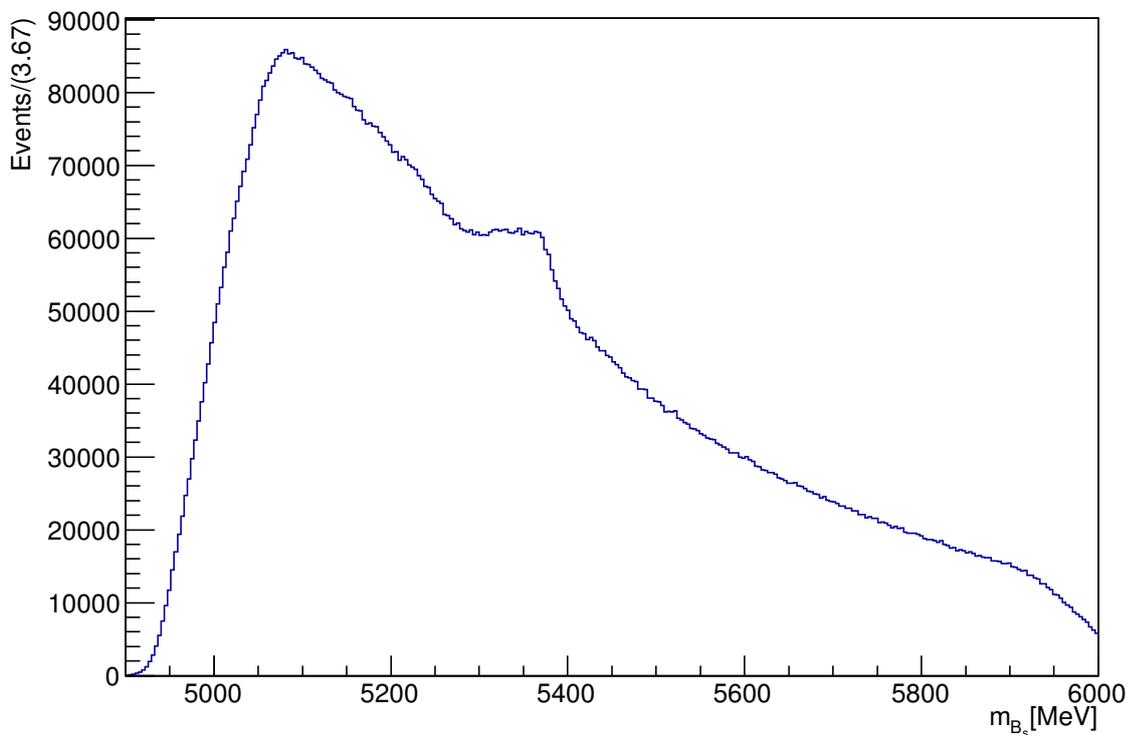


Fig. 11: Reconstructed B_s -mass spectrum

5.1 Variables

Several variables that can be used to distinguish between signal and background events have already been mentioned in the previous section. Further variables that were used in this analysis will be introduced in more detail here.

Invariant mass (\mathbf{m}): The invariant mass of a particle is reconstructed from the measured four momentum $p = (E, p_x, p_y, p_z)^T$ where E is the energy of the particle and $p_{x,y,z}$ the momentum in x, y and z direction. How these are measured is described in section 3. The invariant mass is then calculated as

$$m = \sqrt{p^2} \tag{5.1}$$

where the Minkowsky metric ($p^2 = E^2 - \mathbf{p}^2$) is used to calculate the scalar product of the four vectors.

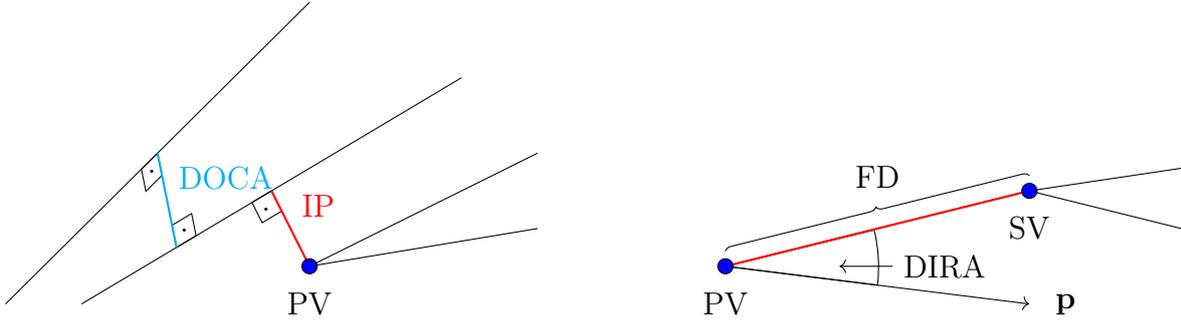


Fig. 12: Sketches of some of the variables

Flight distance (FD): The flight distance is the distance a particle travels from its production vertex (in this case: for the B_s that is the primary vertex (PV), for the D_s^- it is the secondary vertex (SV)) to its decay vertex.

Decay time (t): The decay time is defined as the time from production of a particle to its decay in its rest frame. It can be obtained from the flight distance as

$$ct = \frac{\text{FD}}{\beta\gamma} \quad (5.2)$$

where $\beta = \frac{v}{c}$ and $\gamma = \frac{1}{\sqrt{1-\beta^2}}$ are relativistic quantities that describe the Lorentz boost.

Impact parameter (IP): The impact parameter describes the minimal distance between a reconstructed track and the primary vertex. A track that originated from the PV should have a small impact parameter. Particles from a secondary vertex are more likely to have a larger impact parameter.

Distance of closest approach (DOCA): Is defined as the closest distance between two reconstructed tracks. Particles that come from the same vertex should have a small DOCA while particles from different vertices are more likely to have a larger DOCA.

Direction angle (DIRA): Is the cosine of the angle between the reconstructed momentum of a particle and the vector from its production to decay vertex. It can give information about how well the vertices are reconstructed and whether the assumptions about the decay channel of the particles makes sense.

PID variables: To distinguish between different types of particles, information from the RICH detectors, the calorimeters and the muon system are combined. A likelihood for each type of particle is then calculated. Since pions are the most common particles in the detector the likelihood for a certain particle type is always given relative to the pion hypothesis as a difference in the logarithm of the two likelihoods

$$\Delta \ln \mathcal{L}_{X\pi} = (\ln \mathcal{L}(X) - \ln \mathcal{L}(\pi)). \quad (5.3)$$

The *isMuon* flag is a Boolean variable that is 1 if a particle is likely to be a muon and 0 if not. Since no muons should be in the decay analyzed here, it should be 0 for all particles.

Fit quality: The variables χ_{DTF}^2 and χ_{vtx}^2 describe the quality of the fits performed to reconstruct the decay topology and the vertex.

Cone p_t asymmetry: To calculate the cone p_t asymmetry, a cone is constructed with the tip at the production vertex of this particle and an opening angle such that all the tracks of its decay products are contained within the cone. From all the decay products as well as all other tracks contained within the cone the p_t asymmetry is calculated as

$$A_{pt}^{cone} = \frac{p_{t,\text{decay products}} - \sum p_{t,\text{other tracks}}}{p_{t,\text{decay products}} + \sum p_{t,\text{other tracks}}}. \quad (5.4)$$

In proton-proton collisions a large number of particles are created at small angles to the beam axis. In these regions of high particle densities it is more likely for some combination of particles to fulfill the selection criteria and create a fake B_s candidate. However the cone p_t asymmetry is also more likely to be low for such background events. For events that are clearly isolated from any background particles the probability is higher for it to be a signal decay and the cone p_t asymmetry is close to 1.

5.2 Preselection

As a first step, several linear cuts are performed on the data to reduce the number of background events. These cuts can be seen in Tab.1 and 2. These already reduce the number of events from around 12.000.000 to just around 200.000 events and isolate the signal peak quite well, as can be seen in Fig.13. The broad peak in the mass range below the signal peak can be assigned to partially reconstructed background events. If the B_s meson decays through the $B_s \rightarrow D_s^{*-} \pi^+ \pi^- \pi^+$ decay channel, where the D_s^{*-} decays into $D_s^{*-} \rightarrow D_s^-(\gamma, \pi^0)$ and the photon or the pion are not reconstructed, the B_s mass will be reconstructed a bit below the actual B_s mass. However this background is not expected to spill over into the region of the signal region.

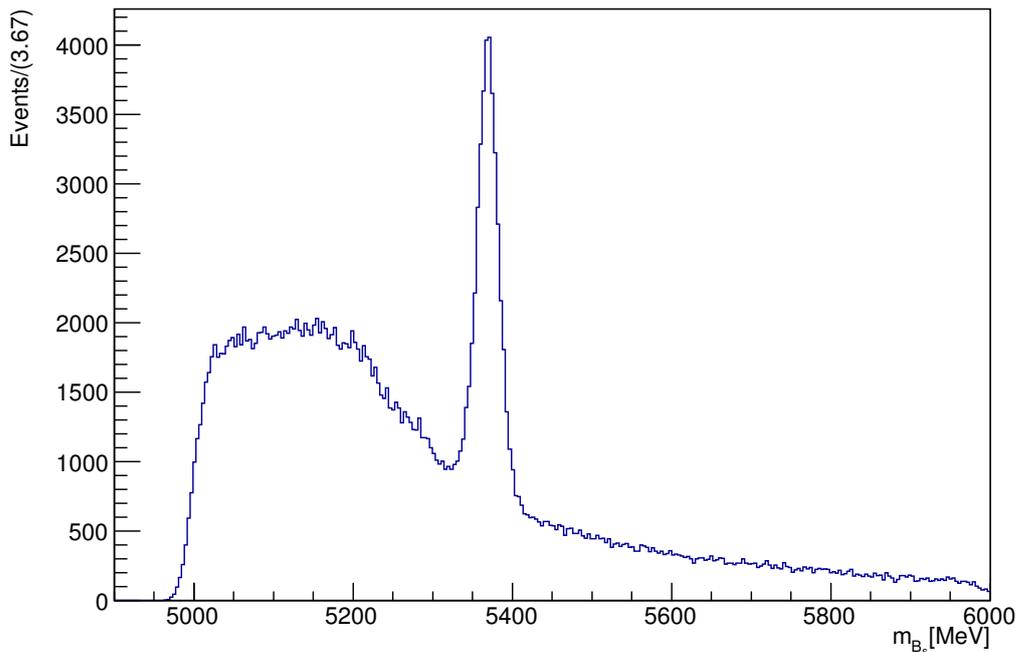


Fig. 13: B_s mass distribution of preselected data

Cuts 1-12 are general cuts, cuts 13 and 14 are cuts on the D_s^- daughters, cuts 15-22 are cuts on the decay $D_s^- \rightarrow \phi\pi^-$, cuts 23-31 are cuts on the decay $D_s^- \rightarrow K^*(892)K^-$ and cuts 32-40 are cuts on the nonresonant decay $D_s^- \rightarrow K^+K^-\pi^-$. The reason why these cuts are useful will now be discussed.

The invariant mass of the B_s daughters should be in the region of the B_s mass. By cutting away events that have a lower mass, a lot of combinatorical background is rejected in cut 1.

Cuts 2 and 6 cut away events where the model that has been used to fit the decay tree and the vertices does not describe the topology of the event very well, which makes the event more likely to be background but also the fit results less meaningful.

As has been discussed above the cosine of the direction angle (DIRA) should be close 1, the flight distance significant and the impact parameter of the B_s small. This is done by cuts 3-5.

Very small decay times mean, that the flight distance is very short. This means the secondary vertex is not far from the primary vertex and particles coming from the primary vertex are more likely to be identified as part of the decay chain, which is why they are cut away by cut 7.

The pions coming from the secondary vertex most likely decay through the $a_1(1260)$ resonance ($a_1(1260)^+ \rightarrow \pi^+\pi^-\pi^+$), which means that much higher invariant masses are more likely to come from background events. (Cuts 8 and 9)

The particles in the decay should be well identified, which is why cuts on the PID variables are implemented. (Cuts 10-12)

The invariant mass of the D_s^- daughters is supposed to be in the region of the nominal D_s^- mass. To filter events with a $D^0 \rightarrow KK$ that looks like a D_s^- decay with a random pion the invariant mass computed from the two kaons is required to be below 1840MeV. (Cuts 13 and 14)

Some additional cuts are then optimized for the different decay channels of the D_s^- . The ϕ resonance is so narrow, that events that fulfill $m(KK) = m_\phi \pm 20\text{MeV}$ are dominated by signal events and the additional cuts are therefore more loose to make sure that not too many signal events will be filtered out. The $K^*(892)$ resonance is a bit broader which means that more background events will fall into the reconstructed invariant mass region of the $K^*(892)$. The cuts therefore have to be a bit tighter for an optimal signal significance. Non-resonant decays however contain a lot of background which is why the tightest cuts have to be applied here.

The cuts 15, 23, 24, 32 and 33 divide the events into these three categories. Cuts 16-18, 25-27 and 34-36 are additional PID cuts which vary in tightness. Cuts 19, 20, 28, 29, 37 and 38 are cuts on the flight distance, that should be significant and in forward direction. Small flight distances could come from background events, where the D_s^- daughter candidates actually do not come from a tertiary vertex as is expected.

Finally it is possible that $B^0 \rightarrow D^-\pi^+\pi^-\pi^+$ decays with $D^- \rightarrow K^+\pi^-\pi^+$ could be identified as B_s^0 decays if the π^- passes the kaon requirements. Therefore events, where the invariant mass calculated with the changed particle hypothesis is within $\pm 30\text{MeV}$ of the nominal D^- mass, have to fulfill a stricter PID cut. Also the decay $\Lambda_b \rightarrow \Lambda_c(\rightarrow pK^-\pi^+)K\pi\pi$ could falsely be identified as a signal event if the proton is identified as a kaon. Events where

the recomputed invariant mass of the D_s^- daughters lies within $\pm 30\text{MeV}$ of the nominal Λ_c mass have to fulfill a tighter PID cut on the kaon.

	Nr.	Description	Requirement	
$B_s \rightarrow D_s \pi \pi \pi$	1	$m(D_s \pi \pi \pi)$	$> 5200 \text{ MeV}$	
	2	χ_{vtx}^2/ndof	< 8	
	3	DIRA	> 0.99994	
	4	χ_{FD}^2	> 100	
	5	χ_{IP}^2	< 20	
	6	χ_{DTF}^2/ndof	< 15	
	7	t	$> 0.4\text{ps}$	
	8	$m(\pi \pi \pi)$	$< 1.95\text{GeV}$	
	9	$m(\pi^+ \pi^-)$	$< 1.2\text{GeV}$	
	10	PIDK(π^+)	< 5	
	11	PIDK(π^-)	< 10	
	12	isMuon(π^+)	$= 0$	
$D_s^- \rightarrow KK\pi^-$	13	$m(KK\pi^-)$	$= m_{D_s} \pm 20\text{MeV}$	
	14	$m(KK)$	$< 1840\text{MeV}$	
$D_s^- \rightarrow \phi\pi^-$	15	$m(KK)$	$= m_\phi \pm 20\text{MeV}$	
	16	PIDK(K^+)	> -10	
	17	PIDK(K^-)	> -10	
	18	PIDK(π^-)	< 20	
	19	χ_{FD}^2	> 0	
	20	FD in z	$> -1 \text{ mm}$	
	21	D^- veto	$m(K^+ K_\pi^- \pi^-) \neq m(D^-) \pm 30\text{MeV} \parallel$ $\text{PIDK}(K^-) > 0$	
	22	Λ_c veto	$m(K^+ K_p^- \pi^-) \neq m(\Lambda_c) \pm 30\text{MeV} \parallel$ $\text{PIDK}(K^-) - \text{PIDp}(K^-) > 0$	
	$D_s^- \rightarrow K^*(892)K^-$	23	$m(KK)$	$\neq m_\phi \pm 20\text{MeV}$
		24	$m(K^+ \pi^-)$	$= m_{K^*(892)} \pm 75\text{MeV}$
25		PIDK(K^+)	> -10	
26		PIDK(K^-)	> -5	
27		PIDK(π^-)	< 10	
28		χ_{FD}^2	> 2	
29		FD in z	$> 0 \text{ mm}$	
30		D^- veto	$m(K^+ K_\pi^- \pi^-) \neq m(D^-) \pm 30\text{MeV} \parallel$ $\text{PIDK}(K^-) > 5$	
31		Λ_c veto	$m(K^+ K_p^- \pi^-) \neq m(\Lambda_c) \pm 30\text{MeV} \parallel$ $\text{PIDK}(K^-) - \text{PIDp}(K^-) > 5$	

Tab. 1: Preselection Cuts

	Nr.	Description	Requirement
$D_s^- \rightarrow (KK\pi^-)_{\text{NR}}$	32	$m(KK)$	$\neq m_\phi \pm 20\text{MeV}$
	33	$m(K^+\pi^-)$	$\neq m_{K^*(892)} \pm 75\text{MeV}$
	34	PIDK(K^+)	> 5
	35	PIDK(K^-)	> 5
	36	PIDK(π^-)	< 10
	37	χ_{FD}^2	> 5
	38	FD in z	> 0 mm
	39	D^- veto	$m(K^+K\pi^-) \neq m(D^-) \pm 30\text{MeV} \parallel$ PIDK(K^-) > 20
	40	Λ_c veto	$m(K^+K\pi^-) \neq m(\Lambda_c) \pm 30\text{MeV} \parallel$ PIDK(K^-)–PIDp(K^-) > 5

Tab. 2: Preselection Cuts

5.3 Multivariate classification

To increase the signal significance by reducing the combinatorial background contribution even further a multi variate analysis [26] was performed. Several classifiers were tested, but a boosted decision tree (BDT) with gradient boosting showed the best separation power.

So far, only linear cuts on some variables were used to classify events as signal like or background like. However, combining the information about more than one variable might be able to classify events better. A decision tree is able to perform such a multivariate analysis. It starts with the full sample and tries to find a variable and a corresponding cut to maximize the number of signal like events in one category and the number of background like events in the other category. The two resulting sets are called leaves of the decision tree. Each leaf is split into two more daughter leaves until a leaf is created that has a purity that is high enough to confidently categorize the event as either signal like or background like. Figure 14 shows an example of such a decision tree. Visually speaking a decision tree cuts the hyperspace of all the input variables into many hypercubes, that are categorized as signal like and background like.

Boosting is a technique used to improve the performance of such a tree, where not just one such tree is generated but a whole forest of trees, which will generate a combined output value, that is close to -1 for background like events and 1 for signal like events. In every new tree, events that are wrongly categorized will get a higher weight.

Such a boosted decision tree has to be trained on samples where it is known whether they are signal or background events. In this case signal events are taken from a Monte Carlo simulation whereas background events are taken from the upper mass sideband of the preselected data ($m_{B_s} > 5600\text{MeV}$).

One problem that can occur in boosted decision trees is an effect called overtraining. If no two events have the exact same variable values it will always be possible to create a tree that correctly categorizes all of the events of a training sample by just making the hypercubes around the events very small. To counteract this effect a minimum number of events per leaf can be used to guarantee that the hypercubes will stay big enough to be able to contain multiple events. Overtraining can also be checked for by training the BDT

on one half of the training sample (with randomly selected events) and checking if the other half of the sample, that has not been used for the training, shows a similar output distribution.

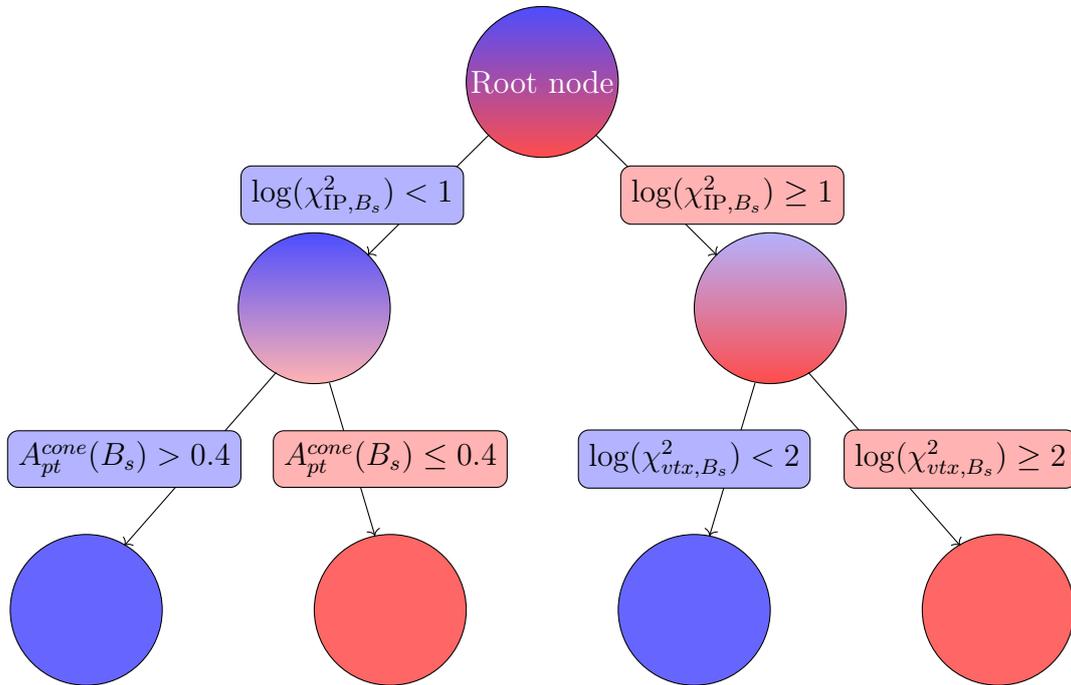


Fig. 14: Sketch of a decision tree, blue: signal like, red: background like

5.3.1 BDT performance

The BDT was trained using the following classification variables

- B_s and D_s^- cone p_t asymmetry
- the minimum IP significance χ_{IP}^2 of the three pions coming from the secondary vertex
- minimum IP significance χ_{IP}^2 of the D_s daughters
- IP significance χ_{IP}^2 of the B_s
- B_s and D_s^- DIRA
- maximum DOCA from the three pions coming from the secondary vertex
- FD significance χ_{FD}^2 of the B_s and D_s^-
- quality of the SV vertex fit χ_{vtx}^2
- quality of the decay tree fit χ_{DTF}^2
- the maximum angle between the D_s^- and another track i in the plane transverse to the beam $\max(\cos(D_s h_i))$

The reason why these variables have a separation power has already been discussed above. The distribution of these variables for signal and background events can be found in the appendix (10). The output distribution for the background and the signal sample can be seen in Fig.15. One can clearly see that signal events give an output value close to 1,

while background events give an output value close to -1. One can also see, that there is no evidence for significant overtraining, since the test sample (solid histogram) shows a similar distribution of the output value as the training sample (points with error bars). If overtraining occurred one would expect a different distribution for training and test sample.

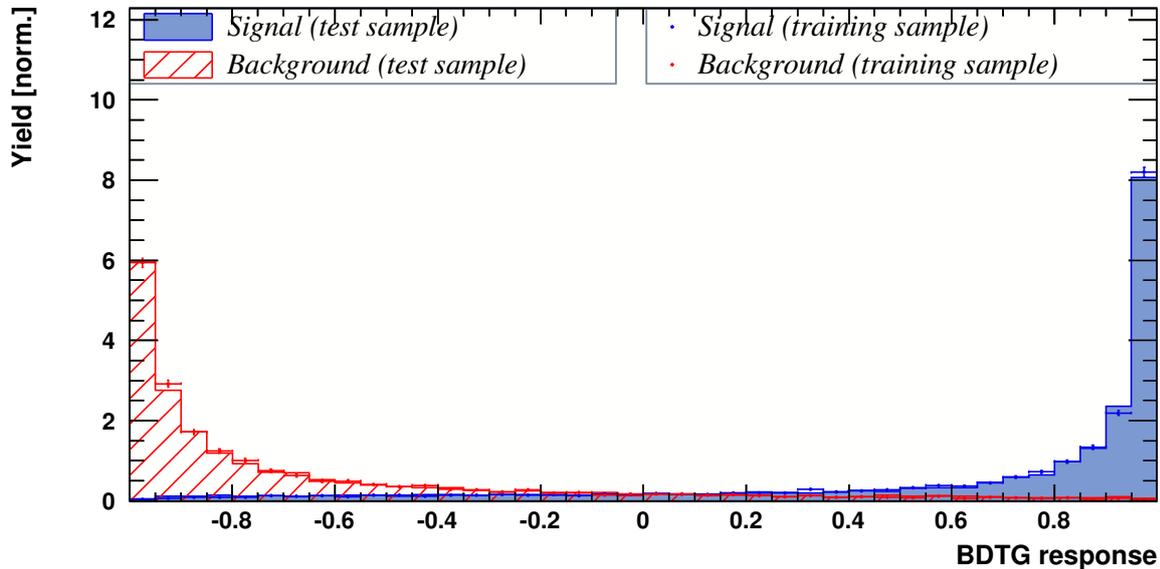


Fig. 15: BDT output distributions for signal (blue) and background (red) events

In Fig.16 one can see the signal and background efficiencies depending on the value of the cut on the BDT classifier. Having a cut at 0 means, that every event that gives a BDT output value smaller than 0 will be filtered out. To find the optimal cut one tries to maximize the signal significance

$$\text{signal significance} = \frac{S}{\sqrt{S+B}} \quad (5.5)$$

where S is the number of signal events and B is the number of background events inside the signal region that remain after the cut. In data it is not possible to know for sure which events are signal and which are background events. To estimate the number of signal and background events one therefore relies on analyzing the mass distribution. If one assumes an exponential background mass distribution and a double gaussian signal distribution, a fit to the mass distribution of the form

$$\mathcal{F}(m_{reco}) = S \cdot \mathcal{S}(m_{reco}) + B \cdot \mathcal{B}(m_{reco}) \quad (5.6)$$

where

$$\mathcal{S}(m_{reco}, m_{B_s}) = \frac{f}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(m_{B_s}-m_{reco})^2}{2\sigma_1^2}} + \frac{(1-f)}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(m_{B_s}-m_{reco})^2}{2\sigma_2^2}} \quad (5.7)$$

is the normalized signal distribution, with m_{B_s} being the nominal B_s mass, σ_1 and σ_2 the widths of the two gaussians and f the fraction of the first gaussian.

$$\mathcal{B}(m_{reco}, a) = \mathcal{N} e^{-a \cdot m_{reco}} \quad (5.8)$$

describes the background shape where \mathcal{N} is a normalization factor that normalizes the distribution to 1 in the region of the fit and a is a parameter that describes the shape of the background. Fitting the distribution \mathcal{F} will yield the number of signal and background events. The fit was performed in the mass region from 5340-5700 MeV while the signal and background numbers were only taken from the region $m_{B_s, \text{pdg}} \pm 40 \text{MeV}$. This fit is now performed for different values of the BDT cut, from which the optimal cut can be found. As already stated above the result of this can be seen in Fig.16

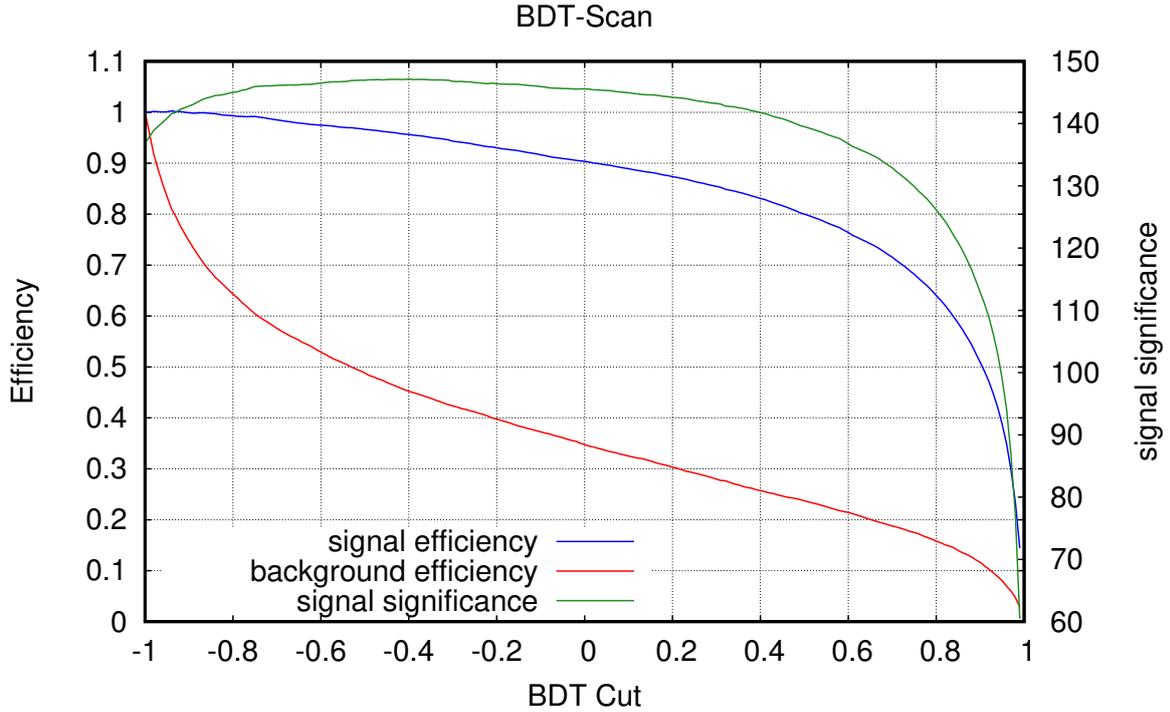


Fig. 16: BDT scan

The optimal cut was found to be at -0.43, for which 137392 signal candidates remain. The signal and background efficiencies are normalized to the total number of signal and background events without any cut.

5.4 Background subtraction

The B_s -mass distribution has some additional information about how likely an event is a signal or a background event. Events far from the signal peak are most likely background events, while events close to the signal peak are more likely to be signal events. From the knowledge of the mass distribution and the expected signal (double-Gaussian) and background (exponential) shapes it is possible to calculate so-called sWeights, a technique called the sPlot method. An in-dept view at the mathematics of this technique can be found in [27]. Applying these weights will statistically positively weigh events that are likely to be signal events and negatively weigh events that are likely to be background events. Using this technique a sideband subtraction is effectively performed. If the decay time distribution is independent of the mass the whole background will be subtracted from the sample on a statistical basis.

In Fig.17 one can see the fit to the mass distribution that is used to calculate these weights. Of all the events that pass the BDT cut, 29150 lie in the signal peak.

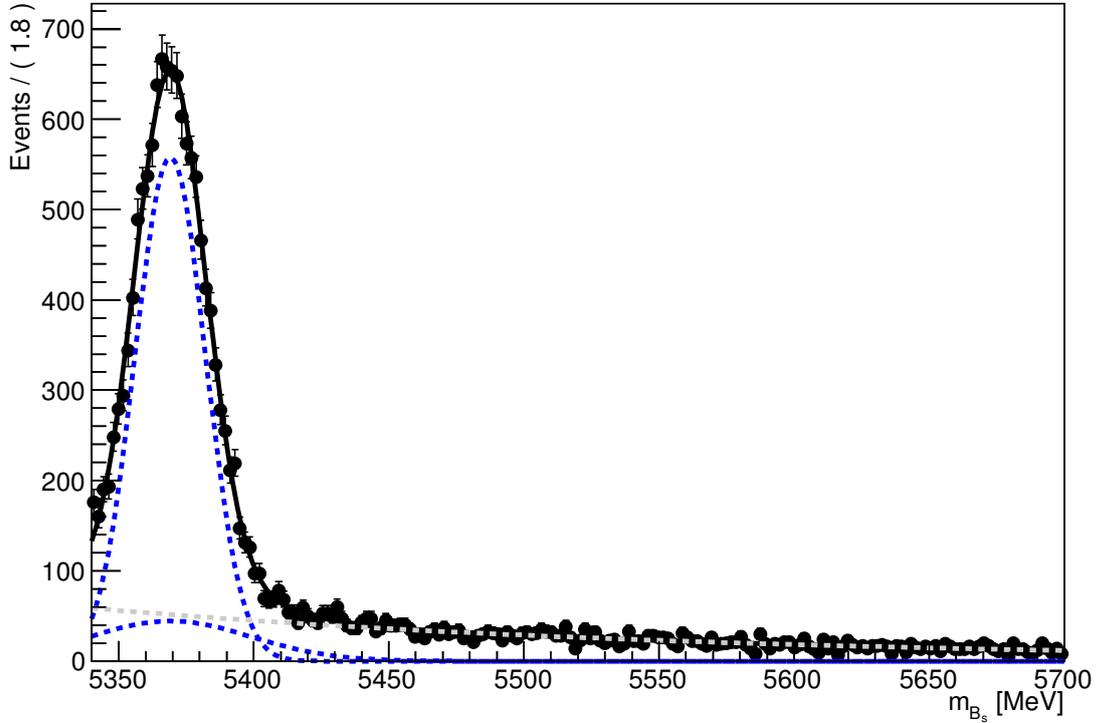


Fig. 17: B_s mass distribution, Blue: Signal component (sum of two gaussians), Gray: Exponential background component, Black: Combined fit

5.4.1 Comparison MC-Data

Training the BDT with simulated data and applying it to real data makes it necessary that the variables used for the training are properly reproduced. The distribution of Monte Carlo events and signal events should therefore be the same for the variables that have been used. In Fig.18 and 19 one can see that this assumption is reasonable.

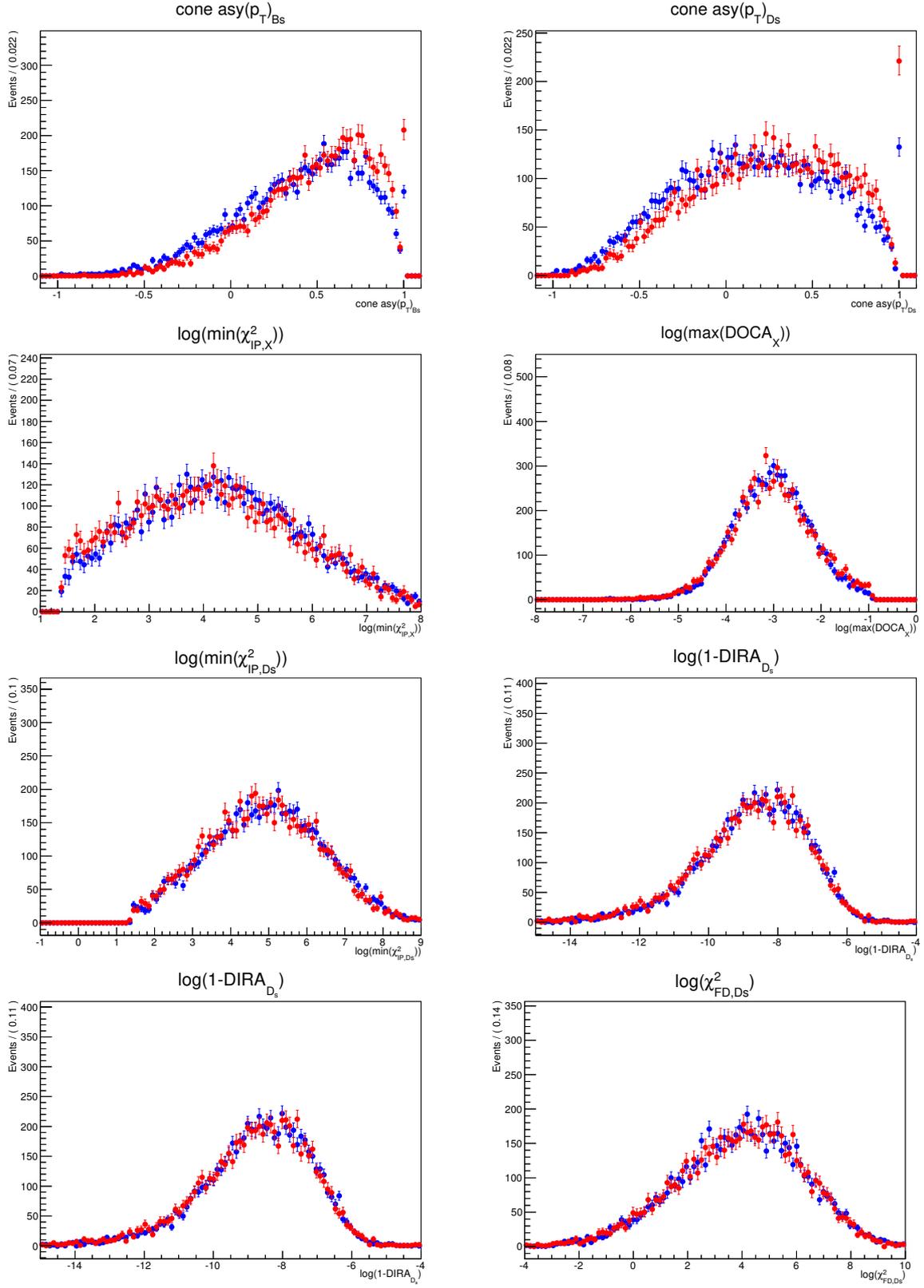


Fig. 18: Distribution of training variables in data (blue) and Monte Carlo (red)

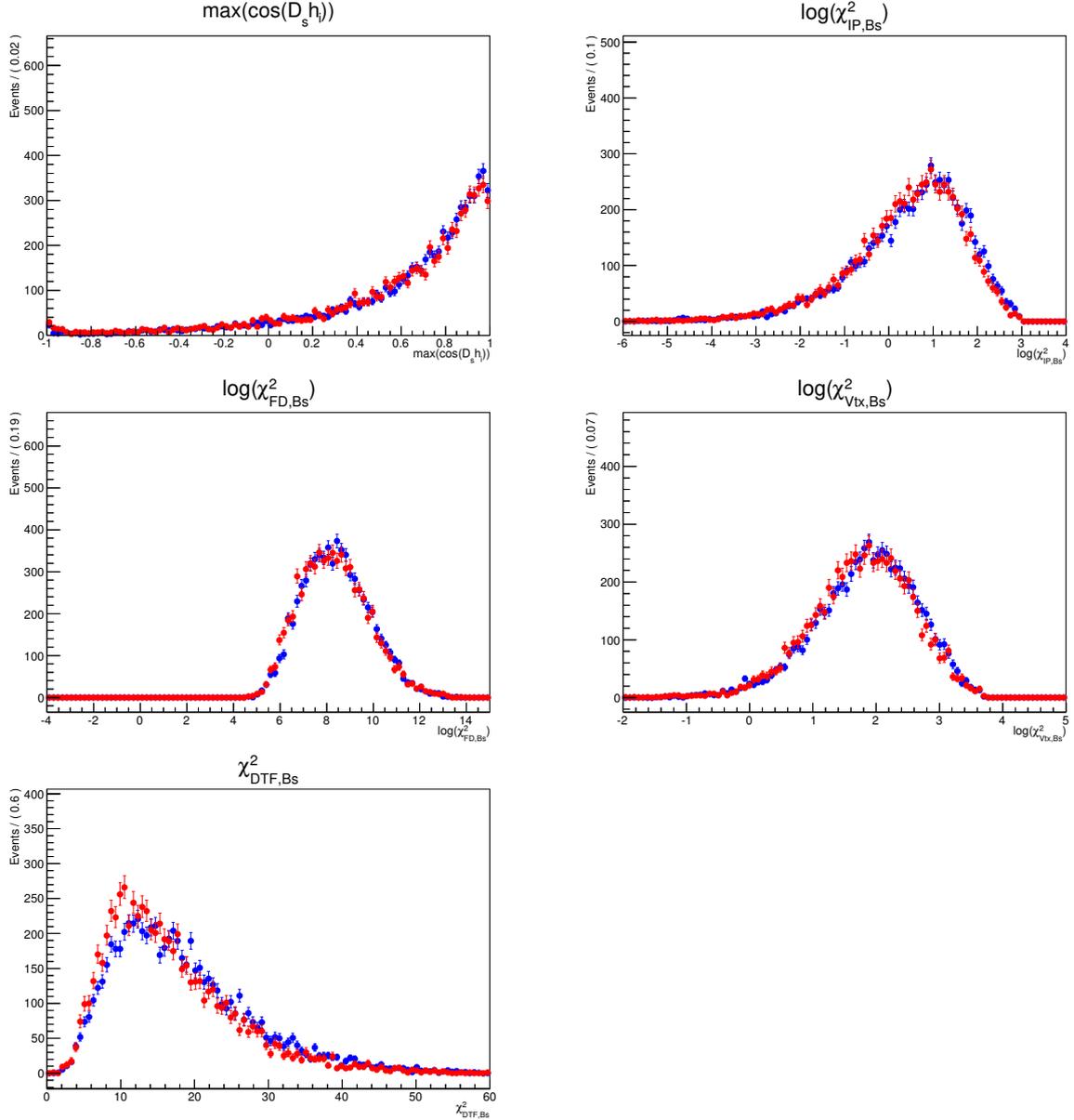


Fig. 19: Distribution of training variables in data (blue) and Monte Carlo (red)

5.4.2 Mass dependency of the MVA-variables

The performance of the BDT is optimal if the variables used are not correlated with the B_s mass, since events from the upper mass sideband ($m_{B_s} > 5600\text{MeV}$) are used, while it is actually applied to events with a B_s mass in the region around the nominal B_s mass. In Fig.20 the mean value of the BDT output is shown for different bins of the B_s mass in the region above the signal peak. One can clearly see that there is a mass dependency in the BDT output. Also in Fig.21 one can see the same plot for some of the BDT variables. (The same plots for the other BDT variables can be found in the appendix 10) The variables that have been used were chosen in a way that there should not be such an effect. The reason for this dependency has not been found in this analysis and it would be interesting to analyze this further in future analyses. It should however not influence the result of this thesis, the only possible effect would be a suboptimal BDT performance.

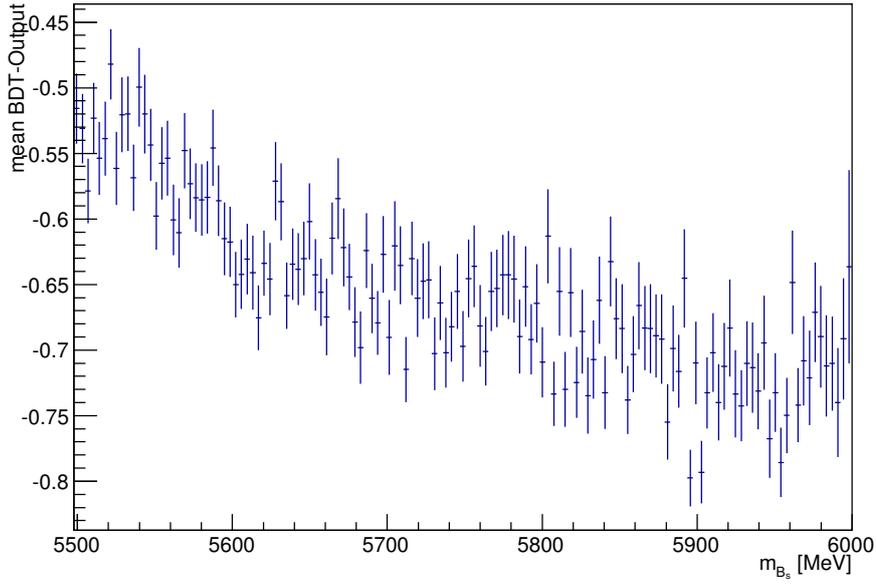


Fig. 20: Mass dependency of BDT-Classifier

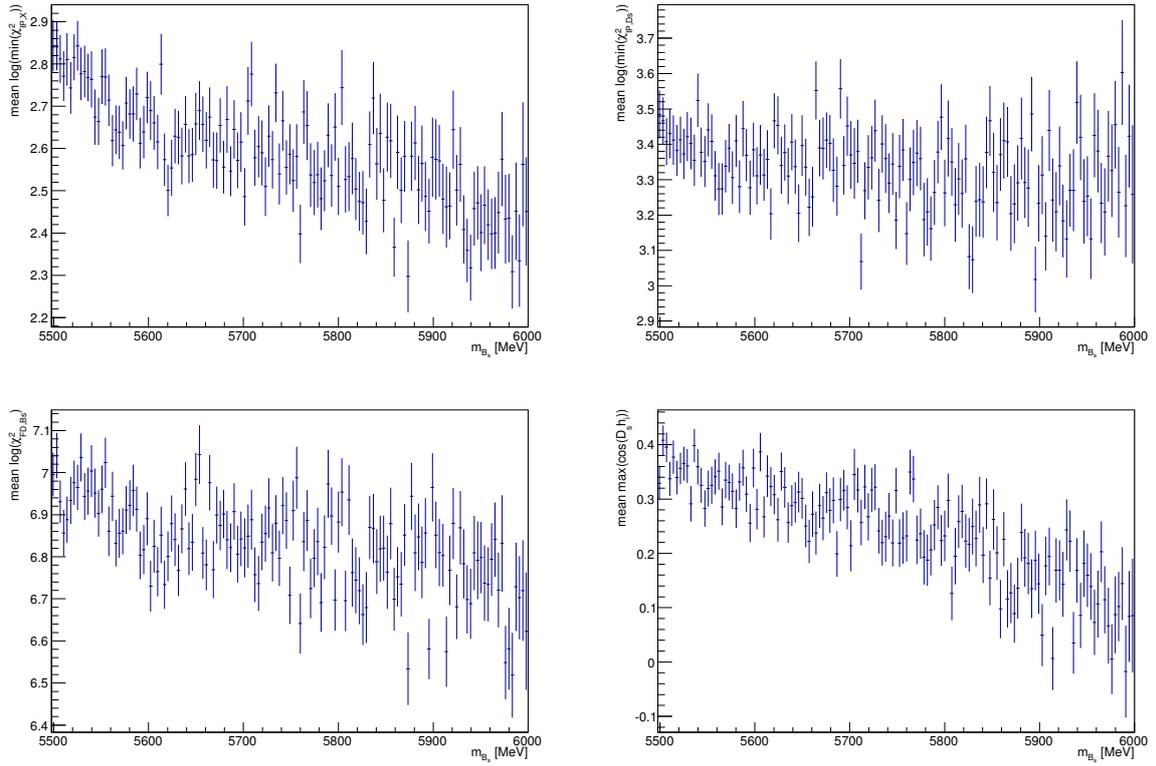


Fig. 21: mass dependencies of *top left:* $\log(\min(\chi^2_{IP,X}))$, *top right:* $\log(\min(\chi^2_{IP,D_s}))$, *bottom left:* $\log(\chi^2_{FD,B_s})$, *bottom right:* $\max(\cos(D_s h_i))$

6 Analysis

As stated above the analysis is done decay time dependent. A fit function of the form

$$\mathcal{P}(t, \Delta m_s) \propto \left[e^{-\Gamma_s t} \left(\cosh \left(\frac{\Delta \Gamma_s t}{2} \right) - q \cdot \mathcal{D} \cdot \cos(\Delta m_s t) \right) \otimes \mathcal{G}(t_{res}) \right] \cdot \epsilon(t) \cdot \mathcal{N}_{tag} \quad (6.1)$$

will be used to describe the mixing process. There are several parameters that have to be determined before the fit can be performed.

6.1 Time resolution

One of these parameters is the time resolution of the detector. This can be obtained from the Monte Carlo simulation. The decay time will only be reconstructed with a certain precision. Because the true decay time as well as the reconstructed decay time is known for the Monte Carlo sample, the residuals

$$\Delta t = t_{reco} - t_{true} \quad (6.2)$$

can be calculated. If the reconstruction algorithm is calibrated correctly one expects a distribution around zero. The width of which gives the decay time resolution. The distribution that was obtained like this can be seen in Fig.22. A double gaussian distribution

$$\mathcal{D}(\mu, \sigma_1, \sigma_2) = \frac{f}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{\Delta t^2}{2\sigma_1^2}} + \frac{(1-f)}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{\Delta t^2}{2\sigma_2^2}} \quad (6.3)$$

with the same mean was fitted to this distribution.

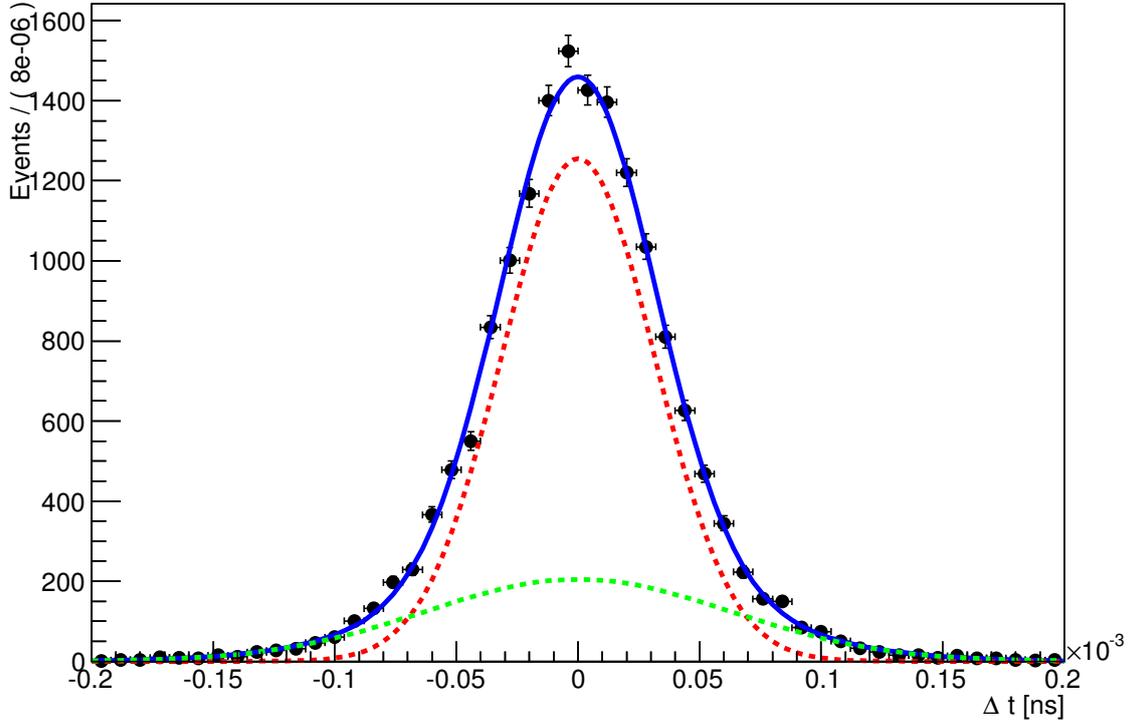


Fig. 22: Decay time resolution of the detector for simulated events

The decay time resolution can now be obtained as

$$\sigma_t = \sqrt{f \cdot \sigma_1^2 + (1 - f) \cdot \sigma_2^2} \quad (6.4)$$

which yields

$$\sigma_t = (41.5 \pm 2.6)\text{fs} \quad (6.5)$$

which is consistent with the expected decay time resolution of the LHCb-detector. [28]

6.2 Acceptance function

The selection process is not equally efficient for all decay times. Therefore an acceptance function has to be found. We determine it from the Monte Carlo sample and assume it to be of the form

$$a(t) = 1 - e^{\lambda(t-\delta)} \cdot (1 - \beta t) \quad (6.6)$$

Because the lifetime that has been used to simulate the Monte Carlo sample is known to be $\tau = 1.51\text{ps}$, the parameters of the acceptance function can be obtained from fitting the decay time distribution (eq.6.1, with $q = 0$) with a fixed value for τ . This yields

$$\lambda = (-1.360 \pm 0.160) \cdot 10^3 \text{ns}^{-1} \quad (6.7)$$

$$\beta = (30 \pm 11) \text{ns}^{-1} \quad (6.8)$$

$$\delta = (2.70 \pm 0.23) \cdot 10^{-4} \text{ns} \quad (6.9)$$

In Fig.23 one can see the acceptance function.

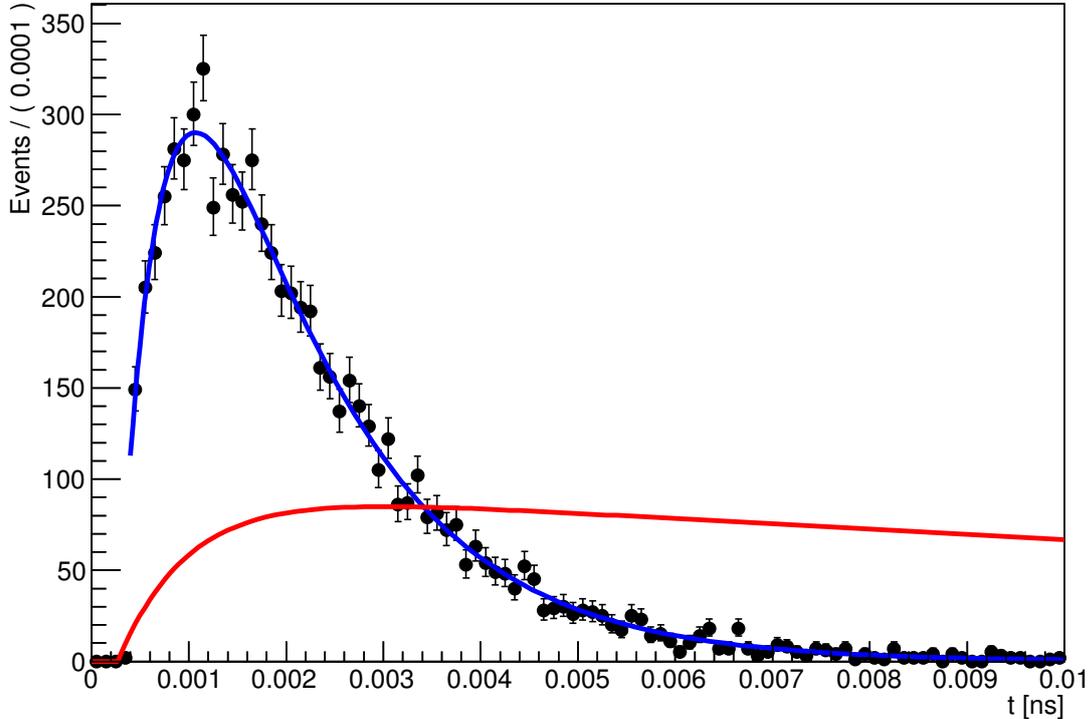


Fig. 23: Decay time distribution (blue) and acceptance function (red)

6.3 Identification of B_s flavor at production

To find out whether an event has mixed or not it is crucial to determine the flavor of the B_s meson produced in the primary vertex (B_s or \bar{B}_s). This information cannot be obtained from the decay products of the B_s but has to be found by analyzing what other particles were created in the event. The algorithms that determine this flavor are called flavor taggers. There are two types of flavor taggers that can be used for this task, that use different properties of the proton-proton collisions.

Opposite Side Taggers (OST) [29] rely on the fact that almost all b quarks are produced in $b\bar{b}$ pairs. Therefore information about the flavour of one of them can be obtained from analyzing the decay of the other one. Depending on the way the second b quark hadronises the information that can be gained has different significance. Hadrons containing b quarks typically decay quite fast, which means they cannot be directly measured and their flavor has to be reconstructed from their decay products. One therefore has to assume that the second b meson has not changed its flavor before its decay (by mixing). If the second b quark forms a second B_s meson, the decay products will not hold significant information because of the high mixing frequency. Charged b Hadrons, that do not change flavor however will contain more information about the flavor of the b meson. The flavor tagger tries to use this information to give a decision about the flavor of the first b meson.

Same Side Kaon Taggers (SST) [30] rely on the fact that in a proton-proton collision the net strangeness is zero. The s quark in the B_s meson therefore has to be created from the vacuum and further produce another particle with strangeness. In most cases this is a kaon. From this kaon the flavor of the B_s can be obtained. The problem with this method is that in proton-proton collisions kaons are not uncommon and the right kaon has to be found to get information about the B_s meson.

These algorithms are trained on calibration channels, where no mixing occurs and the flavor of the decay products is directly correlated with the charge of the decaying particle.

As might be clear by now these algorithms will not work perfectly every time. The algorithms are not fully efficient and will create mistags. To quantify this we define the tagging efficiency as

$$\epsilon_{tag} = \frac{N_R + N_W}{N_R + N_W + N_U} \quad (6.10)$$

where N_R is the number of events that are correctly tagged, N_W is the number of events that are wrongly tagged and N_U is the number of events where the algorithm does not produce a decision, these events are called untagged. Untagged events do not hold any information about the mixing of events.

The mistag probability is defined as

$$\omega = \frac{N_W}{N_R + N_W} \quad (6.11)$$

To compare two algorithms the tagging power or effective tagging efficiency can be defined as

$$\epsilon_{eff} = \epsilon_{tag}(1 - 2\omega)^2 = \epsilon\mathcal{D}^2 \quad (6.12)$$

where \mathcal{D} is the dilution factor that was already introduced in eq.4 To find out these quantities for the two taggers, an analysis was made using the Monte Carlo samples. Using the Monte Carlo truth it can be analyzed which, and more importantly how many events

were correctly and incorrectly tagged by the two different taggers and thus the mistag probability and the effective tagging efficiency can be calculated. This gives

$$\epsilon_{eff,OS} = (2.27 \pm 0.44)\% \quad (6.13)$$

$$\epsilon_{eff,SS} = (1.61 \pm 0.59)\% \quad (6.14)$$

which is consistent with the results in [5] of $\epsilon_{eff,OS} = (2.4 \pm 0.4)\%$ and $\epsilon_{eff,SS} = (1.2 \pm 0.3)\%$ within the errors.

Because the Opposite Side Tagger showed a better performance all further analysis was done using this tagger only. Using the Opposite Side Tagger 11570 events remain, that pass all the selection criteria and are tagged. Untagged events hold no information about the mixing.

The tagging algorithm also provides a predicted mistag probability ω , which indicates how confident the tagger is about its decision. If it is low, it is more likely that the decision is correct, while for $\omega = 0.5$ the decision is basically just a guess. This predicted mistag probability ω is not identical with the actually observed mistag probability η , which is why the taggers were calibrated using Monte Carlo data. For bins of ω the actual mistag probability was calculated using the Monte Carlo truth information and Eq.6.11. This can be seen in Fig.24.

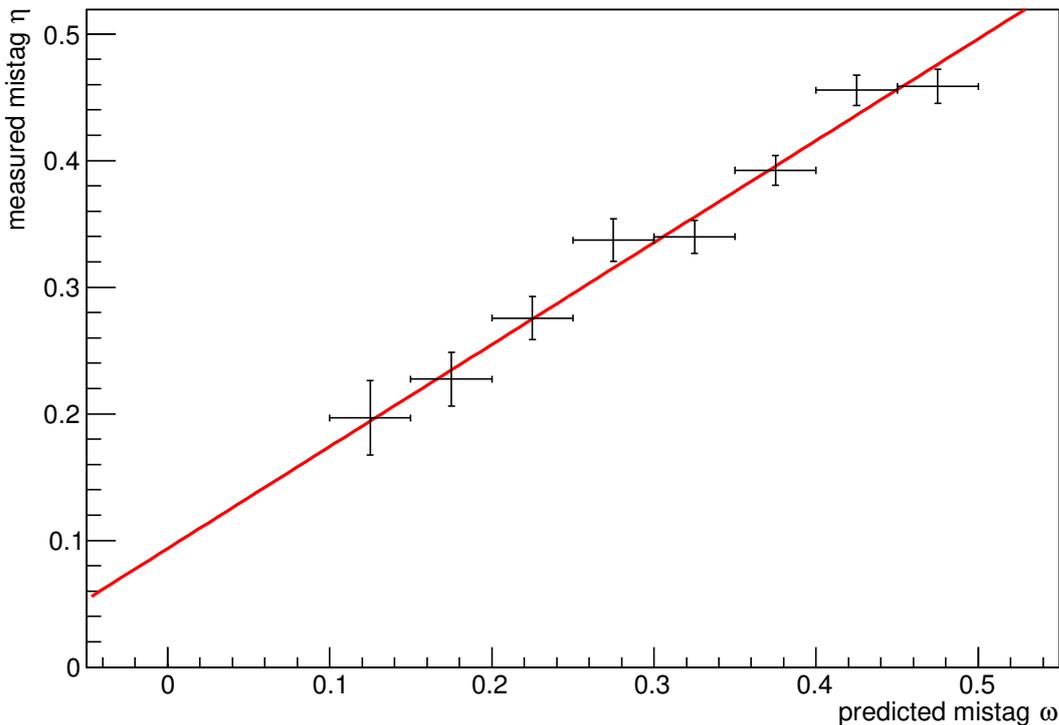


Fig. 24: Tagging calibration for the OS tagger

A linear fit gave a calibrated η ,

$$\eta = 0.805 \cdot \omega + 0.093. \quad (6.15)$$

In the following section the calibrated mistag probability will be used.

6.4 Final fit

Because the events that have a low predicted mistag probability contain more information about the mixing than those with high predicted mistag probabilities, the final fit was done in three bins ($(0 < \eta < 0.37)$, $(0.37 \leq \eta < 0.435)$, $(0.435 \leq \eta \leq 0.5)$) of the calibrated mistag probability, where the mixing frequency was simultaneously fitted to the three bins. These bins were chosen to contain an equal number of signal events each. Using the tagging information, the events were classified into mixed and unmixed events and the distributions

$$\mathcal{P}_{unmixed}(t, \Delta m_s) \propto \left[e^{-\Gamma_s t} \left(\cosh \left(\frac{\Delta \Gamma_s t}{2} \right) + \mathcal{D} \cdot \cos(\Delta m_s t) \right) \otimes \mathcal{G}(t_{res}) \right] \cdot \epsilon(t) \quad (6.16)$$

$$\mathcal{P}_{mixed}(t, \Delta m_s) \propto \left[e^{-\Gamma_s t} \left(\cosh \left(\frac{\Delta \Gamma_s t}{2} \right) - \mathcal{D} \cdot \cos(\Delta m_s t) \right) \otimes \mathcal{G}(t_{res}) \right] \cdot \epsilon(t) \quad (6.17)$$

where simultaneously fitted to the data. All parameters were fixed to the pdg-values [31] except for the mixing frequency Δm_s as well as the mistag probabilities for each of the three bins. This gave

$$\Delta m_s / \hbar = (17.746 \pm 0.027_{\text{stat}}) \text{ps}^{-1} \quad (6.18)$$

with $\omega_1 = 0.306 \pm 0.017$, $\omega_2 = 0.416 \pm 0.018$ and $\omega_3 = 0.473 \pm 0.017$.

In Fig.25 the fit of the mixing distribution and the signal events can be seen for each of the three bins of the mistag probability. In the plots not the whole fit range (from 0 to 10ps) is presented such that the mixing can be easier seen despite the high mixing frequency. The effect of the tagging dilution can clearly be seen.

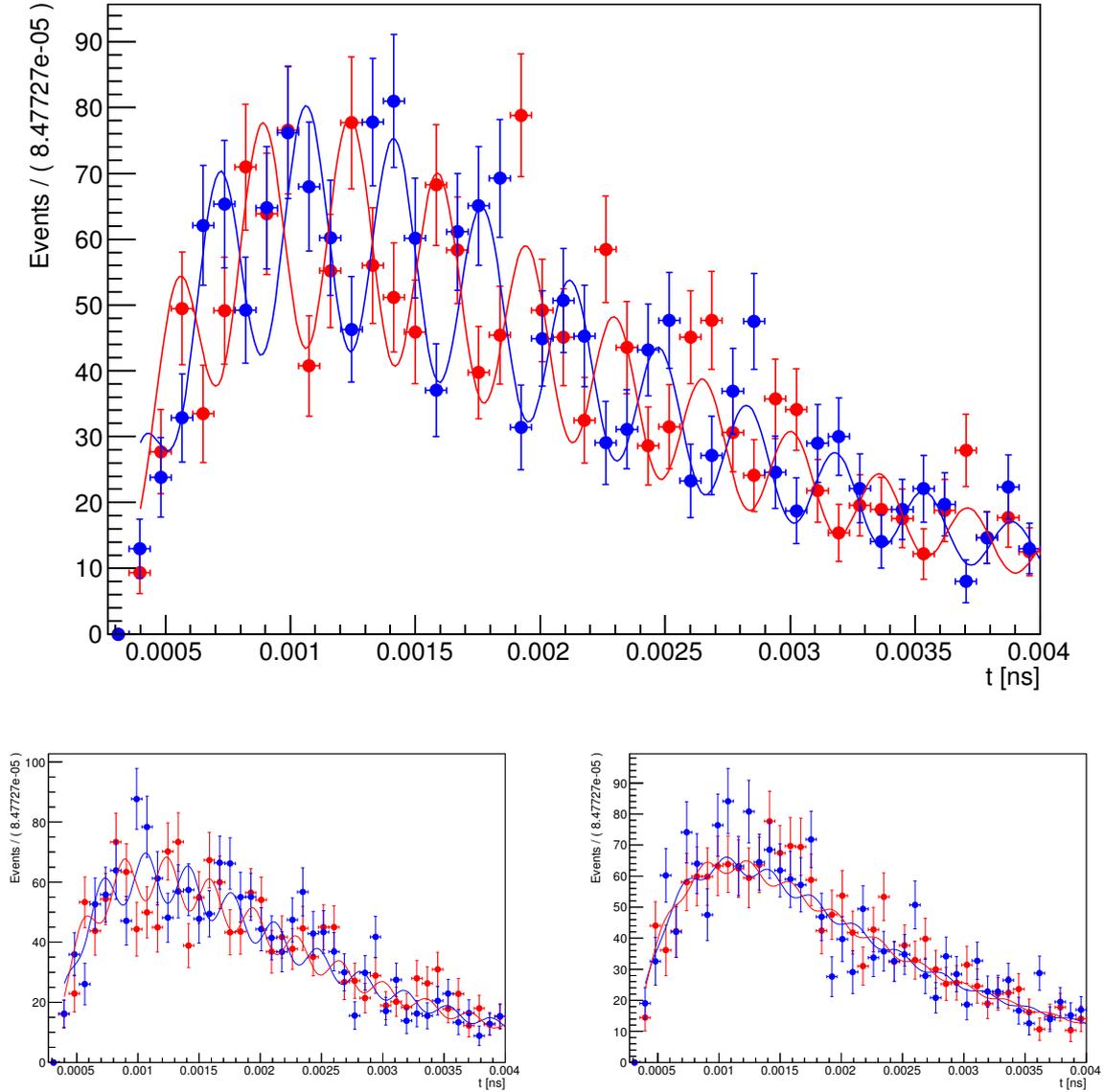


Fig. 25: Mixing distribution for the three bins of the mistag probability, *Top:* lowest mistag probability, *Bottom left:* medium mistag probability, *Bottom right:* highest mistag probability

7 Systematic uncertainties

Systematic biases on the final result, that cannot be explained by the statistical limitations of the sample, will be analyzed in this section.

7.1 Various fit systematics

First it was checked whether changing any of the parameters that were obtained from simulation would have an effect on the final fit result.

The parameters of the acceptance function were each (one by one) changed to plus and minus 3σ of the value that was obtained from simulation. This did not have a significant effect on the fit result for the mixing frequency. The parameters were then also released and directly fitted to the data distribution. This also did not change the result significantly. The same was done for the value of the time resolution as well as the values for the B_s decay time τ and decay width difference (of the two mass eigenstates) $\Delta\Gamma$. For the last two the values were varied by 3σ of the current world average, given by the particle data group [31].

It was also checked if the mass model for signal and background might influence the fit. A linear as well as a constant model instead of an exponential one was checked for the background distribution. Instead of a double gaussian signal shape a single gaussian was also checked.

The influence of the chosen binning of the predicted mistag probability was investigated as well. Fits have been performed using the two and three bin combinations

- [(0,0.405), (0.405,0.5)]
- [(0,0.37), (0.37,0.5)]
- [(0,0.435), (0.435,0.5)]
- [(0,0.35), (0.35,0.405), (0.405,0.5)]
- [(0,0.405), (0.405,0.45), (0.45,0.5)]

These bins were chosen because the intervals (0,0.405),(0.405,0.5) contain roughly 50% of data each, while the intervals (0,0.37), (0.435,0.5) contain roughly one third of all data and (0,0.35), (0.35,0.405), (0.405,0.45), (0.45,0.5) roughly one quarter each. This means the bins that were chosen correspond to a 1:1, 1:2, 2:1, 1:1:2 and 2:1:1 split of the data.

The exact influences that these effects had on the mixing frequency can be found in Tab.3. To combine these effects to an overall systematic error, they are added in quadrature. If several measurements are done for one effect ($x \pm 3\sigma_x$, multiple background models, multiple ways of binning η) the largest of these values was taken. This results in a systematic error of

$$\sigma_{\Delta m_s}/\hbar = 10.79\text{ns}^{-1} \quad (7.1)$$

for these error sources.

Description	Change	Δm_s	$\Delta(\Delta m_s)$	$\sigma(\Delta m_s)$
Acceptance function	$\lambda = \lambda_0 + 3\sigma_{\lambda_0}$	17746.22	0	26.58
	$\lambda = \lambda_0 - 3\sigma_{\lambda_0}$	17746.22	0	26.58
	$\beta = \beta_0 + 3\sigma_{\beta_0}$	17746.19	-0.03	26.58
	$\beta = \beta_0 - 3\sigma_{\beta_0}$	17746.32	0.10	26.63
	$\delta = \delta_0 + 3\sigma_{\delta_0}$	17746.22	0	26.72
	$\delta = \delta_0 - 3\sigma_{\delta_0}$	17746.22	0	26.58
	λ, β, δ free to fit	17746.10	-0.12	26.57
Time resolution	$\sigma_\tau = \sigma_{\tau,0} + 3\sigma_{\sigma_{\tau,0}}$	17748.86	2.64	26.60
	$\sigma_\tau = \sigma_{\tau,0} - 3\sigma_{\sigma_{\tau,0}}$	17743.92	-2.30	26.57
B_s decay time	$\tau = \tau_0 + 3\sigma_{\tau_0}$	17746.17	-0.05	26.47
	$\tau = \tau_0 - 3\sigma_{\tau_0}$	17746.27	0.05	26.54
B_s decay width difference	$\Delta\Gamma = \Delta\Gamma_0 + 3\sigma_{\Delta\Gamma}$	17746.39	0.17	26.54
	$\Delta\Gamma = \Delta\Gamma_0 - 3\sigma_{\Delta\Gamma}$	17746.08	-0.14	26.52
Bg model for B_s mass dist.	const. Bg	17741.12	-5.10	26.85
	linear Bg	17744.98	-1.24	26.69
signal model for B_s mass dist.	single gaussian	17746.64	0.42	26.57
Binning of η	$[(0,0.405),(0.405,0.5)]$	17755.35	9.13	28.46
	$[(0,0.37),(0.37,0.5)]$	17739.33	-6.89	27.12
	$[(0,0.435),(0.435,0.5)]$	17752.94	6.72	29.10
	$[(0,0.35),(0.35,0.405),$ $(0.405,0.5)]$	17754.30	8.08	27.11
	$[(0,0.405),(0.405,0.45),$ $(0.45,0.5)]$	17754.54	8.32	28.07
Combined systematic error			10.79	

Tab. 3: Influence of various systematics, in units of $\hbar\text{ns}^{-1}$

7.2 Z-scale uncertainty

The decay time is proportional to the measured flight distance of the B_s meson. The position of the VELO modules along the beam direction is only known with a certain precision which could bias decay times. The effect of this uncertainty was studied in [5] and found to be

$$\sigma_{z\text{-scale}} = 0.022\% \quad (7.2)$$

resulting in a systematic uncertainty on $\Delta m_s/\hbar$ of $\pm 0.004\text{ps}^{-1}$.

7.3 Momentum scale uncertainty

The decay time is directly linked to the momentum of the B_s . An uncertainty on the momentum scale therefore results in an equal bias in the mixing frequency. The effect of the momentum uncertainty was found to be

$$\frac{\Delta t}{t} = 0.02\% \quad (7.3)$$

in [5], an uncertainty of $\pm 0.004\text{ps}^{-1}$ was therefore assigned on the mixing frequency Δm_s .

7.4 Fit bias

It was also checked whether the fit itself or the error would be biased. Using the fitted model all parameters were fixed to the fit results and 2700 Monte Carlo toy samples with the same number of signal events as in the real data set were generated. Each of the resulting distributions was then fitted with the model. All samples that converged reasonably, were then analyzed. The values of the pull $p = \frac{\Delta m_{s,fit} - \Delta m_{s,input}}{\sigma_{\Delta m,fit}}$ should be distributed as a Gaussian with the mean at 0 and a width of 1 if the fit is accurate. The distribution can be seen in Fig.26 and the mean ($\mu_p = -0.022 \pm 0.028$) and error ($\sigma_p = 0.961 \pm 0.026$) were compatible with 0 and 1 respectively. One can therefore conclude that the fit does not artificially alter the result.

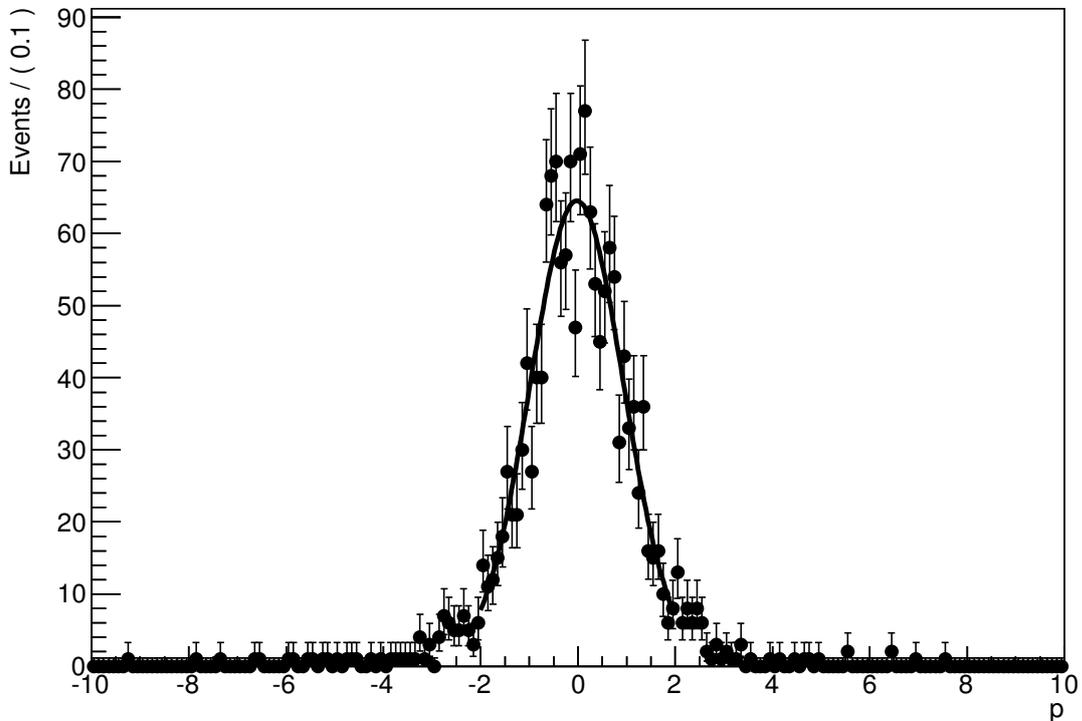


Fig. 26: dDistributions of pulls obtained from a series of monte carlo toy samples

7.5 Summary

Most of the systematics that have been analyzed showed no significant effect on the mixing frequency. The largest effect is due to the binning of the data into categories of the predicted mistag probability. The background model of the B_s mass distribution, the uncertainties of the z -scale as well as the momentum scale uncertainty also have some effect on the outcome of the measurement. Combined they give a systematic uncertainty of

$$\sigma_{\Delta m_s, \text{sys}}/\hbar = 0.012\text{ps}^{-1}. \quad (7.4)$$

8 Discussion and summary

In this analysis the mixing frequency Δm_s of the B_s^0 meson was measured in the decay $B_s \rightarrow D_s^- \pi^+ \pi^- \pi^+$. It was found to be

$$\Delta m_s / \hbar = (17.746 \pm 0.027_{\text{stat}} \pm 0.012_{\text{sys}}) \text{ps}^{-1}. \quad (8.1)$$

This is so far the second most precise measurement of this quantity. The value deviates from the current world average of $\Delta m_s / \hbar = (17.757 \pm 0.021) \text{ps}^{-1}$ by 0.30σ which means it is consistent within the uncertainties. Combining this measurement with the current world average and calculating a weighted average one gets

$$\Delta m_s = \frac{\sum_i^2 \frac{\Delta m_{s,i}}{\sigma_i^2}}{\sum_i^2 \frac{1}{\sigma_i^2}} \pm \sqrt{\frac{1}{\sum_i^2 \frac{1}{\sigma_i^2}}} = (17.753 \pm 0.017) \hbar \text{ps}^{-1} \quad (8.2)$$

which means an increase in precision of 20%.

To achieve this, a selection process with multiple steps has been used to filter out background events and isolate the signal events as well as possible. Several linear cuts as well as a multivariate classification have been performed on the data sample. Using information from the mass distribution, the remaining background was statistically subtracted. From a simulated data sample, the time resolution as well as the acceptance function of the detector have been found. With the information from flavor tagging algorithms it was possible to experimentally measure the mixing of B_s mesons and fit the mixing frequency in three bins of the predicted mistag probability. Finally systematic error sources have been estimated.

The analysis presented in this thesis is the first measurement of this quantity in the decay channel $B_s^0 \rightarrow D_s^- \pi^+ \pi^- \pi^+$ and is competitive with the most precise measurement of the B_s mixing frequency in the decay $B_s^0 \rightarrow D_s^- \pi^+$, where

$$\Delta m_s / \hbar = (17.768 \pm 0.023_{\text{stat}} \pm 0.006_{\text{sys}}) \text{ps}^{-1} \quad (8.3)$$

was measured.

The limiting factor on the precision of this measurement is the statistics. Using data from Run-II the measurement of the mixing frequency could therefore be improved. Also improving the selection process even more might result in better statistics. Due to the limited time available for this thesis, it was not possible to take the Same Side Tagging decision into account. Combining it with the information from the Opposite Side Tagger would give more information about the flavor of the B_s mesons at production, it is however not completely trivial to combine the two tagging decisions. Also the fit that was performed in this analysis was done in three bins of the predicted mistag probability, which was the biggest contribution to the systematic error. Actually using a per-event mistag probability, where every event would be weighted according to its predicted mistag probability without putting them in bins would also increase the precision of the measurement. The same could be done with the per-event predicted time resolution. Including these things in future analyses would likely give an even better measurement of this quantity, though the measurement presented in this analysis is already quite competitive.

9 References and links

- [1] A. Lenz and U. Nierste, “Numerical Updates of Lifetimes and Mixing Parameters of B Mesons,” in *CKM UNITARITY TRIANGLE. PROCEEDINGS, 6TH INTERNATIONAL WORKSHOP, CKM 2010, WARWICK, UK, SEPTEMBER 6-10, 2010*, 2011. arXiv: 1102.4274 [hep-ph]. [Online]. Available: <https://inspirehep.net/record/890169/files/arXiv:1102.4274.pdf>.
- [2] R. Aaij, C. Abellan Beteta, and B. Adeva, “Measurement of CP violation and the B_s^0 meson decay width difference with $B_s^0 \rightarrow J/\psi K^+ K^-$ and $B_s^0 \rightarrow J/\psi \pi^+ \pi^-$ decays,” *Phys. Rev. D*, vol. 87, p. 112010, 11 2013. DOI: 10.1103/PhysRevD.87.112010. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevD.87.112010>.
- [3] “Measurement of time-dependent CP -violation observables in $B_s^0 \rightarrow D_s^\mp K^\pm$,” 2012, Linked to LHCb-ANA-2012-068. [Online]. Available: <http://cds.cern.ch/record/1477943>.
- [4] R Aaij *et al.*, “Precision measurement of the B_s^0 - \bar{B}_s^0 oscillation frequency with the decay $B_s^0 \rightarrow D_s^- \pi^+$,” *New J. Phys.*, vol. 15, p. 053021, 2013. DOI: 10.1088/1367-2630/15/5/053021. arXiv: 1304.4741 [hep-ex].
- [5] S. Wandernoth, *Measurement of the B_s^0 - \bar{B}_s^0 Oscillation Frequency at LHCb using $1fb^{-1}$ of data taken in 2011*, https://www.physi.uni-heidelberg.de/Publications/wandernoth_dissertation.pdf, 2014.
- [6] G. Aad *et al.*, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys. Lett.*, vol. B716, pp. 1–29, 2012. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].
- [7] S. Chatrchyan *et al.*, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Phys. Lett.*, vol. B716, pp. 30–61, 2012. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].
- [8] *Standard Model*, <http://www.physik.uzh.ch/groups/serra/images/SM1.png>, Accessed: 11.7.2018.
- [9] M. Ciuchini, E. Franco, V. Lubicz, F. Mescia, and C. Tarantino, “Lifetime differences and CP violation parameters of neutral B mesons at the next-to-leading order in QCD,” *JHEP*, vol. 08, p. 031, 2003. DOI: 10.1088/1126-6708/2003/08/031. arXiv: hep-ph/0308029 [hep-ph].
- [10] B. Povh, K. Rith, C. Scholz, F. Zetsche, and W. Rodejohann, *Teilchen und Kerne, Eine einföhrung in die physikalischen konzepte*, ger, 9. Aufl. 2014, ser. SpringerLink : Bücher. Berlin, Heidelberg: Springer Spektrum, 2014, Online-Ressource (XII, 467 S. 160 Abb, online resource), ISBN: 978-364-23782-2-5. DOI: 10.1007/978-3-642-37822-5. [Online]. Available: <http://dx.doi.org/10.1007/978-3-642-37822-5>.
- [11] G. Ecker, *Teilchen, Felder, Quanten, Von der quantenmechanik zum standardmodell der teilchenphysik*, ger, ser. SpringerLink : Bücher. Berlin, Heidelberg: Springer Spektrum, 2017, Online-Ressource (XIV, 151 S. 33 Abb., 13 Abb. in Farbe, online resource), ISBN: 978-3-662-54550-8. DOI: 10.1007/978-3-662-54550-8. [Online]. Available: <http://dx.doi.org/10.1007/978-3-662-54550-8>.

-
- [12] C. Berger, *Elementarteilchenphysik, Von den Grundlagen zu den modernen Experimenten*, ger, 3. Aufl. 2014, ser. Springer-Lehrbuch. Berlin, Heidelberg: Springer Spektrum, 2014, Online-Ressource (XIV, 528 S. 217 Abb, online resource), ISBN: 978-364-24175-3-5. DOI: 10.1007/978-3-642-41753-5. [Online]. Available: <http://dx.doi.org/10.1007/978-3-642-41753-5>.
- [13] D. Boutigny *et al.*, “The BABAR physics book: Physics at an asymmetric B factory,” in *WORKSHOP ON PHYSICS AT AN ASYMMETRIC B FACTORY (BABAR COLLABORATION MEETING) PASADENA, CALIFORNIA, SEPTEMBER 22-24, 1997*, 1998. [Online]. Available: <http://www-public.slac.stanford.edu/sciDoc/docMeta.aspx?slacPubNumber=SLAC-R-504>.
- [14] K. Anikeev *et al.*, “ B physics at the Tevatron: Run II and beyond,” in *WORKSHOP ON B PHYSICS AT THE TEVATRON: RUN II AND BEYOND BATAVIA, ILLINOIS, SEPTEMBER 23-25, 1999*, 2001. arXiv: hep-ph/0201071 [hep-ph]. [Online]. Available: http://lss.fnal.gov/cgi-bin/find_paper.pl?pub-01-197.
- [15] H. Terrier, “Overview of LHCb,” in *PROCEEDINGS, 40TH RENCONTRES DE MORIOND ON QCD AND HIGH ENERGY HADRONIC INTERACTIONS: LA THUILE, AOSTA VALLEY, ITALY, MARCH 12-19, 2005*, 2005, pp. 85–88. arXiv: hep-ex/0506047 [hep-ex]. [Online]. Available: https://inspirehep.net/record/685746/files/arXiv:hep-ex_0506047.pdf.
- [16] *Description LHCb Detector*, <http://lhcb-public.web.cern.ch/lhcb-public/en/Detector/Detector-en.html>, Accessed: 11.7.2018.
- [17] A. A. Alves Jr. *et al.*, “The LHCb Detector at the LHC,” *JINST*, vol. 3, S08005, 2008. DOI: 10.1088/1748-0221/3/08/S08005.
- [18] *Sketch of LHC*, <http://te-epc-lpc.web.cern.ch/te-epc-lpc/machines/pagesources/Cern-Accelerator-Complex.jpg>, Accessed: 11.7.2018.
- [19] P. Collins, “Velo constraint system installation (site C). Velo constraint system installation (site C),” LHCb Collection., 2007, [Online]. Available: <http://cds.cern.ch/record/1017398>.
- [20] *Image of LHCb magnet*, https://www.weltmaschine.de/sites/sites_custom/site_weltmaschine/content/e28861/e36564/e36664/e36680/0411013_03-A4-at-144-dpi.jpg, Accessed: 11.7.2018.
- [21] A. Borgia *et al.*, “The Magnetic Distortion Calibration System of the LHCb RICH1 Detector,” *Nucl. Instrum. Meth.*, vol. A735, pp. 44–52, 2014. DOI: 10.1016/j.nima.2013.08.039. arXiv: 1206.0253 [physics.ins-det].
- [22] *LHCb Trigger Scheme*, <https://lhcb.web.cern.ch/lhcb/speakersbureau/html/TriggerScheme.html>, Accessed: 17.7.2018.
- [23] T. Sjostrand, S. Mrenna, and P. Z. Skands, “A Brief Introduction to PYTHIA 8.1,” *Comput. Phys. Commun.*, vol. 178, pp. 852–867, 2008. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].
- [24] D. J. Lange, “The evtgen particle decay simulation package,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 462, no. 1, pp. 152–155, 2001, BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(01\)00089-4](https://doi.org/10.1016/S0168-9002(01)00089-4). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0168900201000894>.
-

-
- [25] S. A. et al., “Geant4—a simulation toolkit,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250–303, 2003, ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [26] A. Hocker *et al.*, “TMVA - Toolkit for Multivariate Data Analysis,” *PoS*, vol. ACAT, p. 040, 2007. arXiv: physics/0703039 [PHYSICS].
- [27] M. Pivk and F. R. Le Diberder, “SPlot: A Statistical tool to unfold data distributions,” *Nucl. Instrum. Meth.*, vol. A555, pp. 356–369, 2005. DOI: 10.1016/j.nima.2005.08.106. arXiv: physics/0402083 [physics.data-an].
- [28] R. Aaij *et al.*, “LHCb Detector Performance,” *Int. J. Mod. Phys.*, vol. A30, no. 07, p. 1530022, 2015. DOI: 10.1142/S0217751X15300227. arXiv: 1412.6352 [hep-ex].
- [29] R. Aaij *et al.*, “Opposite-side flavour tagging of B mesons at the LHCb experiment,” *Eur. Phys. J.*, vol. C72, p. 2022, 2012. DOI: 10.1140/epjc/s10052-012-2022-1. arXiv: 1202.4979 [hep-ex].
- [30] “Optimization and calibration of the same-side kaon tagging algorithm using hadronic B_s^0 decays in 2011 data,” 2012, Linked to LHCb-ANA-2011-103. [Online]. Available: <https://cds.cern.ch/record/1484021>.
- [31] M. Tanabashi, “Review of Particle Physics,” *Phys. Rev. D* 98, 2018. [Online]. Available: <http://pdg.lbl.gov/>.

10 Appendix

Unbinned maximum likelihood fit

To fit the distribution to the data (and all other fits in this analysis where binning is not specifically mentioned) an unbinned maximum likelihood fit is performed. That means that the data is not put into bins of decay time but the actually reconstructed decay time is used for every event. Binning the data always leads to loss of information. For every event

$$f(x_i|a) \tag{10.1}$$

describes the probability density of getting a result x_i if x is distributed as $f(x|a)$ where a is a parameter to be estimated. To get the optimal value for a one has to maximise

$$L(a) = \prod_{i=1}^n f(x_i|a) \tag{10.2}$$

In the case of this analysis $f(x|a) \hat{=} \mathcal{P}(t|\Delta m_s)$.

Distributions of the MVA-variables for signal and background events

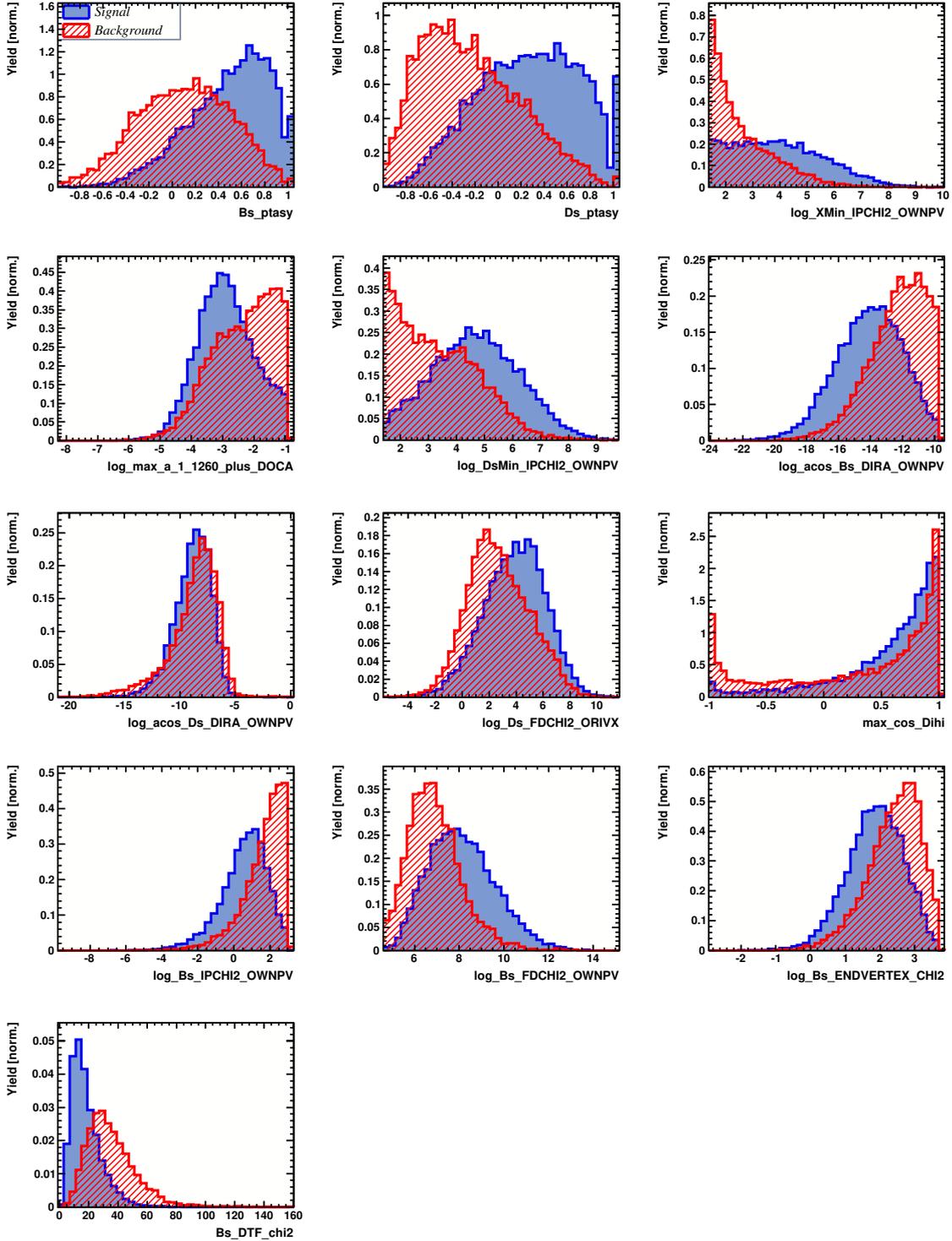


Fig. 27: Distributions of the MVA-variables for signal (blue) and background (red) events, from top left to bottom right: $A_{pt}^{cone}(B_s)$, $A_{pt}^{cone}(D_s^-)$, $\min(\chi_{IP,X}^2)$, $\max(\text{DOCA}_X)$, $\log(\min(\chi_{IP,D_s^-}^2))$, $\log(1 - \text{DIRA}_{B_s})$, $\log(1 - \text{DIRA}_{D_s^-})$, $\log(\chi_{FD,B_s}^2)$, $\max(\cos(D_i h_i))$, $\log(\chi_{IP,B_s}^2)$, $\log(\chi_{FD,B_s}^2)$, $\log(\chi_{vtx,B_s}^2)$, χ_{DTF,B_s}^2 , where X stands for the three pions from the B_s decay

Mass dependency of the MVA-variables

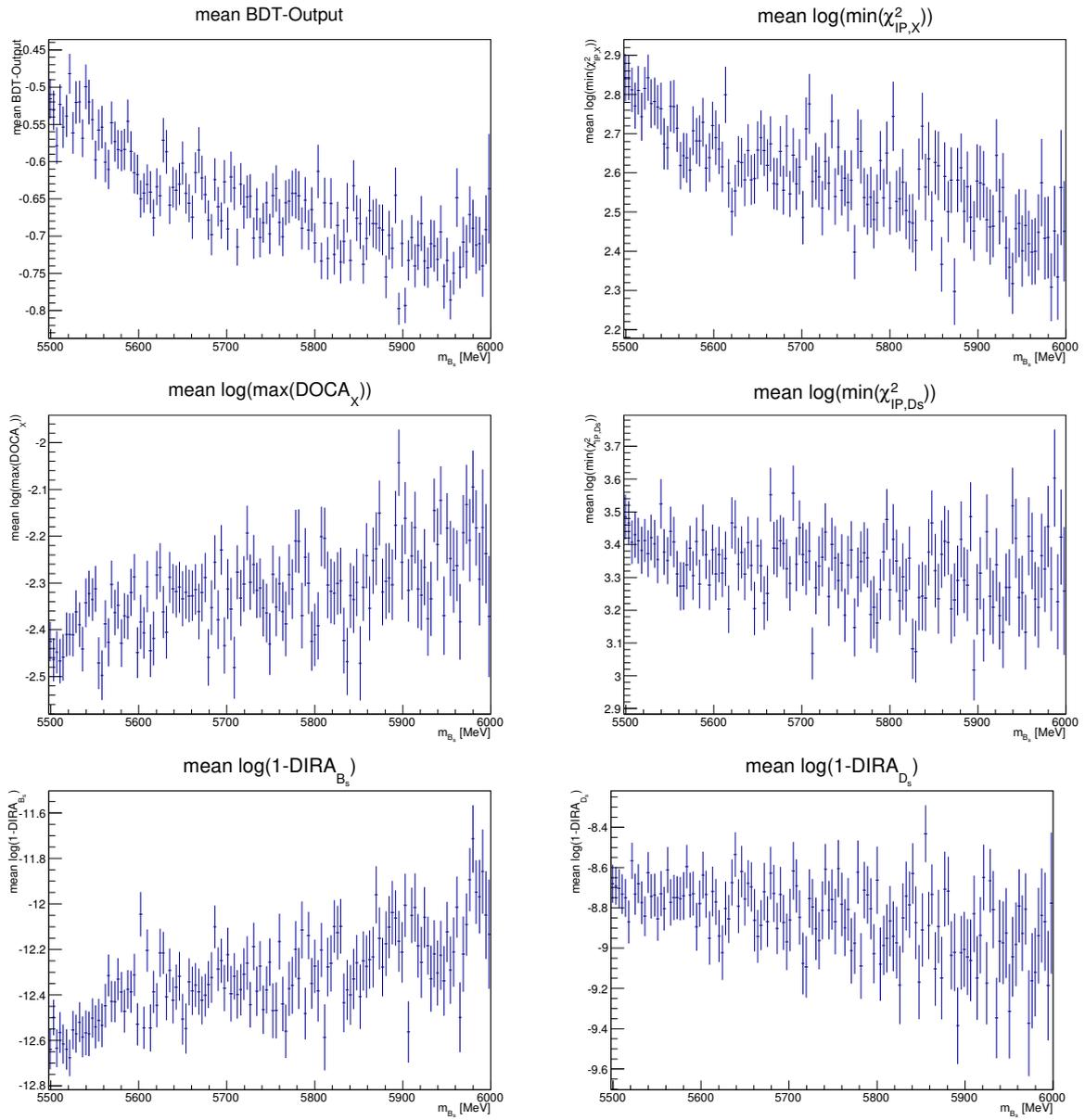


Fig. 28: mass dependency of the MVA-variables

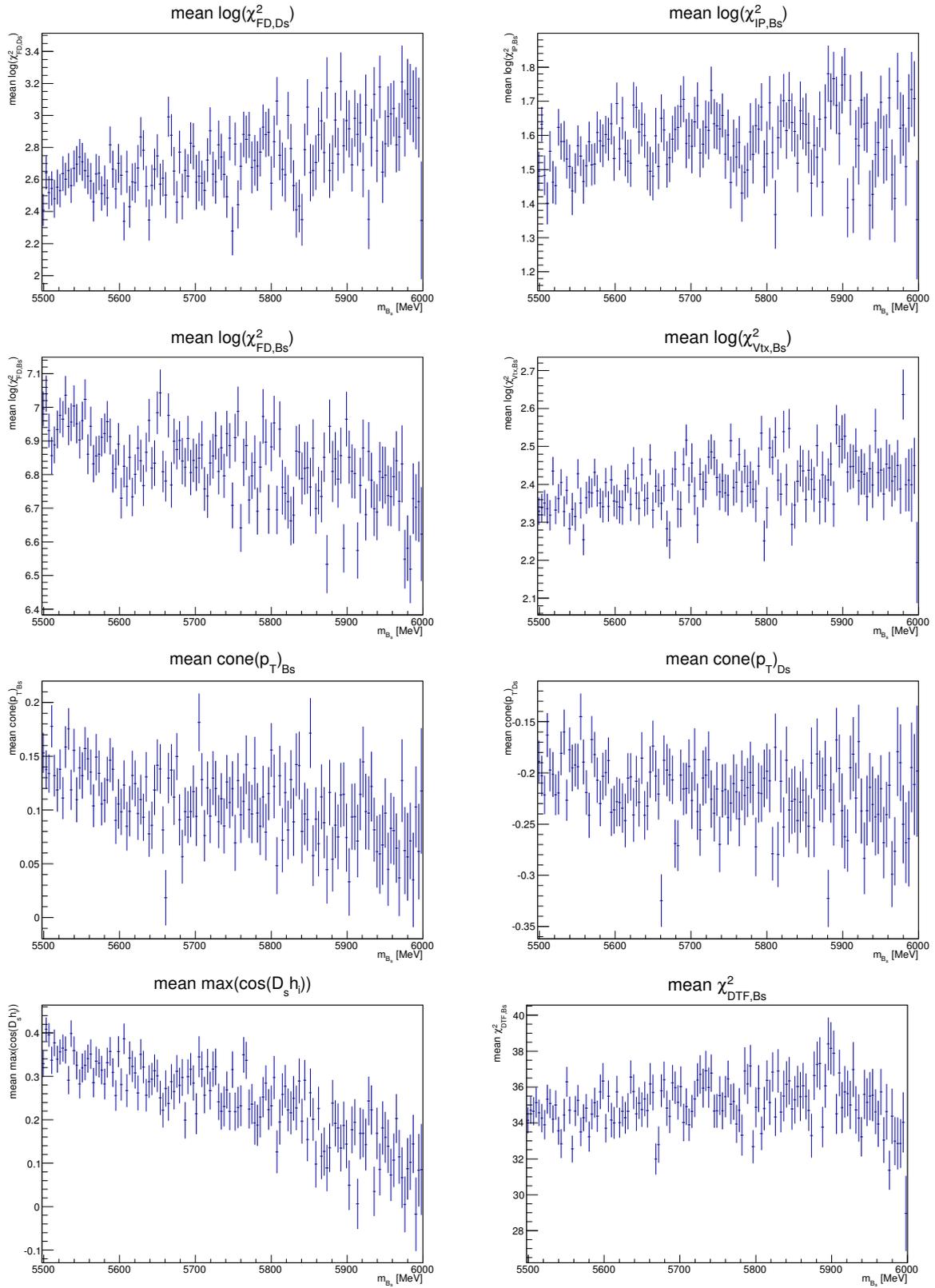


Fig. 29: mass dependency of the MVA-variables

11 Acknowledgment

At first I would like to thank Prof. Dr. Ulrich Uwer for giving me the opportunity to write my bachelor thesis working in the LHCb group and all the help I got writing my thesis. I would also like to thank Dr. Yvonne Pachmayer for agreeing to be the second referee for this thesis.

I am also especially grateful to Philippe d'Argent and Matthieu Kecke for being the advisers for this thesis. You helped me greatly with all my questions concerning the physics, but also the programming that was involved in this thesis. Discussing the contents of my work and the challenges that were present along the way, was always really interesting and also a lot of fun. It was a great pleasure to work with you, thank you for that!

I also want to thank the whole LHCb working group for welcoming me and the interesting topics I got to learn about in the weekly group meetings.

Erklärung

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

Heidelberg, den 20.7.2018,